DOCTORAL THESIS

Machine Learning for Appearance Grading of Sawn Timber using Cameras and X-ray Computed Tomography

Linus Olofsson

Wood Science and Technology

LULEÅ
UNIVERSITY
OF TECHNOLOGY

Machine Learning for Appearance Grading of Sawn Timber using Cameras and X-ray Computed Tomography

Linus Emanuel Olofsson

Division of Wood Science and Engineering
Department of Engineering Sciences and Mathematics
Luleå University of Technology, Skellefteå, Sweden

Supervisors:
Dick Sandberg, Olof Broman, Magnus Fredriksson, Johan Skog

Preface

Suppose any person should be singled out as responsible for the existence of this thesis. In that case, it is Thomas Lundmark, who is a close friend of my family. Ten years ago, Thomas reassured me on my educational career choice by coining the phrase ”Everybody should study Engineering Physics, the rest you will learn on the job”, and so I did. Five years later, just after my summer vacation, Thomas unexpectedly called me and asked if I wanted to become a PhD student – almost as if this was his plan all along – and so I did. Thank you, Thomas, for the small but significant nudges on my career path you have given me.

I want to extend a big thank you to my supervisors Dick Sandberg and Olof Broman – a dynamic duo of supervisors whose support complement each other perfectly. An extra thanks to Olof for his unrelenting support during work regarding data analysis, writing manuscripts, and camaraderie. Additionally, I would like to thank my supervisors Magnus Fredriksson and Johan Skog, for their support. Also, A big thank you to my wonderful colleagues at Luleå University of Technology for the many laughs and interesting debates we have shared. Finally, I am especially grateful to my fellow PhD students for their camaraderie.

Some notable people outside of Luleå University of Technology deserve praise for their contributions to this thesis: Johan Oja (Norra Timber), whom with his expertise on the subject of this thesis and his expertise in the sawmilling industry, enthusiastically contributed with vital insights and his analytical skills; Enrico Ursella (Microtec) for his technical expertise regarding the CT Log scanning system and assistance during data collection; Lars Juselius (FinScan) for his technical expertise regarding the Boardmaster scanning system and assistance during data collection.

To all the people I love, you are more important to me than I let you know, and I am forever grateful to have you in my life.

Skellefteå, September 2021

Abstract

This doctoral thesis deals with a new approach for the appearance grading of sawn timber adapted to the requirements of modern sawmilling industries and timber market situations. Appearance grading of sawn timber allows wood products to be made with a specific visual style due to wood features such as knots. Identifying and grading sawn timber by its visual style is a holistic-subjective task that is inherently suitable for humans. However, with the ever-increasing demand for a faster and more consistent grading operation, humans have been replaced by automatic systems during the past few decades. However, the human perception of the appearance of sawn timber is not something easily defined coherently and concisely for use in automatic systems, resulting in automatic systems struggling to perform appearance grading using conventional rule-based grading. As shown in this thesis, machine-learning methods can be used to teach an automatic system to perform holistic-subjective grading in a way that emulates manual grading while still performing the fast and consistent grading associated with automatic systems. This thesis introduced machine learning for product-adapted appearance grading of sawn timber and studied the use of machine learning to appearance grade sawn timber according to standardised quality grades, using an X-ray computed tomography (CT) scanner and a camera-based board scanner.

In the studies presented in this thesis, measurement data from the CT scanner and the board scanner was used to create a set of variables only regarding knots. The variable sets and the grades of the sawn timber were modelled by projection to latent structures (PLS) models. The grade of the sawn timber was determined in three ways; firstly, manual grading according to standardised quality grades; secondly, called the product grade, the sawn timber was delivered to a wall-panelling customer, and the grade of the sawn timber was determined by the quality yield at the customer; and thirdly, called the image grade, images were extracted from the board scanner and used to estimate the quality yield of the wall-panelling customer manually. The grading in each scanning system was performed using a machine-learning method and a conventional rule-based approach, and their performances were compared.

Seven data sets were collected in the studies presented in this thesis, each with a combination of variable sets from the scanners and quality grades as

described above. In each study, one or more PLS models were trained to model the relationship between a variable set and a quality grade and used to predict the quality of the sawn timber. A PLS model predicts a score for each piece of sawn timber, and if that score passes a classification threshold, the model assigns a quality grade. This classification threshold could be tuned manually to introduce a bias in the model and thereby change the sorting outcome.

When performing standardised appearance grading of dried sawn timber, both a PLS model and rule-based grading achieved about 80% grading accuracy, while a manual grader agreed to 95% with the PLS model and to 81% with the rule-based grading in a verification test. Furthermore, when performing customer-adapted grading of the standardised grades, a PLS model managed an 84% grading accuracy compared to 64% of the rule-based approach. These results show how a conventional rule-based approach struggled with performing customer-adapted grading compared to a PLS model. When performing standardised grading, however, both methods achieved similar grading accuracy, but only the grading performed by the PLS model could not be significantly distinguished from the targeted standardised grades.

Using a PLS model to perform product-adapted grading of dried sawn timber resulted in a grading accuracy of about 70%–80% for different scenarios. These gradings resulted in a quality yield, pass or fail, of about 80% for the wall-panelling customer. According to the customer, rule-based grading did not yield impressive product-adapted results, and no metric was given. Furthermore, this thesis showed that the image grade was as useful as the product grade for training the PLS models, which greatly simplifies the logistical process of creating a data set for training a product-adapted machine-learning model. Had a traceability method been used to collect the data from the scanners automatically, the image grade would allow for completely software-based data collection, which is very much in line with the industry 4.0 concept.

A CT scanner enables the appearance grading of virtual sawn timber in the 3D images of the scanned logs, which allows the logs to be sawn for maximum value or quality yield. The CT scanner was made to perform a primary product-adapted grading using either a PLS model or a rule-based approach.

In addition to this primary grading, the CT scanner and board scanner were programmed to perform a small secondary grading by limiting a small set of measurements that the CT scanner could not sufficiently account for. For example, large pith deviations were limited in the CT scanner, and rotten knots were forbidden by the board scanner, as these measurements were associated with a high risk of resulting in poor quality wall panels for the customer. With this setup, a dataset of 300 pieces of virtual sawn timber was studied. Using rule-based primary grading, the sawmill delivered about 200 pieces of sawn timber with a product yield of 77% for the customer, after the board scanner rejected 28 pieces (12%). Then, by controlling the classification threshold of a PLS model to make the primary grading very strict, meaning that the log was sawn to only yield very likely high-quality pieces of sawn timber, the sawmill could deliver 114 pieces of sawn timber with a product yield of 90%, after the board scanner rejected 9 pieces (7%). These results show that a PLS model achieved higher grading accuracy and higher quality yield than a rule-based approach. Furthermore, the classification threshold of the PLS model allows for easy and intuitive control over the sorting outcome, something that the rule-based approach does not support.

This thesis showed that a PLS-based machine-learning model could be used to perform holistic-subjective appearance grading by both a CT scanner and a board scanner, where a rule-based approach struggled in all but the most familiar case of standardised grading. Once a framework for a machine-learning method such as PLS has been implemented, this thesis showed the ease of customising and fine-tuning the grading performance to be in line with customers needs. A customer or product adaptation could conceivably be initiated and finalised completely in software by automatically collecting the data using a traceability method, collecting the reference grades needed for training by grading images of sawn timber, and using the intuitive classification threshold to fine-tune the sorting outcome.

List of Publications

The present thesis is based on the following appended Publications:

Publication I
Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2017). Customer Adapted Grading of Scots pine Sawn Timber – a Multivariate Method Approach. Zbiec, M. & Orlowski, K. (Eds.) 23rd International Wood Machining Seminar Proceedings: Warsaw, Poland, Warsaw University of Life Sciences, 360–361
URN: urn:nbn:se:ltu:diva-65840

Publication II
Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019). Multivariate Product Adapted Grading of Scots pine Sawn Timber for an Industrial Customer, part 1: Method Development. Wood Material Science & Engineering 14(6), 428–436
DOI: 10.1080/17480272.2019.1617779

Publication III
Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019). Multivariate Product Adapted Grading of Scots pine Sawn Timber for an Industrial Customer, part 2: Robustness to Disturbances. Wood Material Science & Engineering 14(6), 420–427
DOI: 10.1080/17480272.2019.1612944

Publication IV

Olofsson, L., Möller, C.-J., Wendel, C., Oja, J. and Broman, O. (2019). New Possibilities with CT scanning in the Forest Value Chain. Wang, X.; Sauter, U. H.; Ross, R. J., (Eds.) Proceedings, 21st international nondestructive testing and evaluation of wood symposium. General Technical Report FPL-GTR-272. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 569–576
URN: urn:nbn:se:ltu:diva-75506

Publication V

Olofsson, L., Broman, O. and Sandberg, D. (2019). Holistic-subjective Automatic Grading of Sawn Timber: Sensitivity to Systematic Changes. Schajer, G. S. (Ed.) 24th International Wood Machining Seminar Proceedings: Corvallis, Oregon, Oregon State University, 2019, 157-164
URN: urn:nbn:se:ltu:diva-74170

Publication VI

Olofsson, L., Broman, O., Oja, J. and Sandberg, D. (2021). The Effect of Class-balance and Class-overlap in the Training Set for Multivariate and Product-adapted Grading of Scots pine Sawn Timber. Wood Material Science & Engineering 16(1), 58–63
DOI: 10.1080/17480272.2020.1804996

Publication VII

Olofsson, L., Broman, O., Oja, J. and Sandberg, D. (2021). Product Adapted Grading of Virtual Scots pine Sawn Timber by an Industrial CT-scanner Using a Visually Trained Machine-Learning Method. Wood Material Science & Engineering 16(4), 279–286
DOI: 10.1080/17480272.2021.1955298

My Contributions to the Appended Publications:

Publication I: Participated in a key role of the data collection, machine learning engineer, and lead writer.

Publication II: Participated in a key role of the study planning and data collection, lead machine learning engineer, and lead writer.

Publication III: Participated in a key role of the study planning and data collection, lead machine learning engineer, and lead writer.

Publication IV: Lead study planner, lead machine learning engineer, and lead writer.

Publication V: Participated in a key role of the study planning and data collection, lead machine learning engineer, and lead writer.

Publication VI: Lead study planner, participated in a key role of the data collection, lead machine learning engineer, and lead writer.

Publication VII: Participated in a key role of the study planning and data collection, lead machine learning engineer, and lead writer.

Summary of Publications

Publication I *Can a machine-learning model be used to appearance grade sawn timber to be in line with standardised quality grades?*

Key takeaways:

- a multivariate PLS model could predict the standardised quality grades with slightly higher grading accuracy than the existing rule-based grading, and
- using a bigger dataset, this study validated a previous study showing similar results.

Publication II *Can a machine-learning model be used to appear grade sawn timber suitable for an industrial customer's product?*

Key takeaways:

- a multivariate PLS model could predict the product outcome of the sawn timber with 74% accuracy in a test set,
- the use of the PLS model's classification threshold could have been used to change the sorting outcome completely, and
- the performance is comparable to similar studies using PLS models to appearance grade sawn timber by different classes.

Publication III *Are machine-learning models robust to disturbances common in sawmills, for example, dust?*

Key takeaways:

- the PLS models tested were robust to, but not unaffected by, the disturbance from excessive amounts of dust on the sawn timber, and
- training should be performed on as large of a training set as possible since the grading accuracy improved with a larger training set.

Publication IV *When more than one grading system with different measurement techniques are used, does machine learning help perform consistent grading in each system?*

Key takeaways:

- introducing a PLS model in at least one of the of two studied systems increased the agreement between the two, and
- 92% of the data in this study could have been traced through the sawmill automatically.

Publication V *Depending on the type of sawn timber to be sorted, does a machine-learning model need to be trained to be task-specific?*

Key takeaways:

- automatically captured images of the sawn timber were successfully used to create a dataset with significant differences between sawn timber from butt logs or top logs, and
- a PLS model trained to grade sawn timber from the intended top logs performed better when grading sawn timber from top logs than from butt logs.

Publication VI *How do the aspects of class balance and class overlap in the training set affect the grading outcome of a machine-learning model, and can this be investigated using an image-based reference-grade?*

Key takeaways:

- when training using the product-grade reference, all available data should be used for maximum grading accuracy,
- training on the image-based reference-grade yields the same grading performance as training on the product-grade reference, and
- training on the image-based reference-grade where the quality expert was uncertain about the grade did not improve the grading accuracy of the model, and this data could be excluded from the training.

Publication VII *Is it possible to perform product-adapted grading of sawn timber before sawing using a CT scanning system and a machine-learning model, and can this be investigated using an image-based reference-grade?*

Key takeaways:

- a PLS model can grade sawn timber as suitable for a specific product before sawing and thereby allow for planning of the entire refinement process from log to product for each piece of sawn timber,
- the results indicated that in future studies, it could be possible to skip the dry-sorting process entirely and only perform the grading in the CT scanning system before sawing, and
- the image-based reference-grade was as useful for training as the product grade, indicating than product adaptation could be done entirely in software, had a traceability method been used.

Contents

Chapter 1

Introduction

Wood is a beautiful and a whimsical material

1.1 Thesis Overview

1.1.1 Background

The business of any sawmill includes refining round timber into sawn products such as sawn timber (planks and boards) and by-products. This is a complex task as the stems or logs come in a large variety of shapes and sizes, and their inherent features and composition decides the mechanical and chemical properties of the sawn products. Furthermore, customers of sawn timber have different needs depending on their businesses, for example, construction companies or furniture makers have wildly different requirements. Due to this complexity, most sawmills specialise in processing one or a few wood species, thereby greatly limiting the variety of both the incoming logs and the outgoing sawn timber, which limits the variety of suitable customers. On the other hand, such specialisations allow sawmills to use specialised machines and greatly increase production capacity.

Many sawmills in the Nordic countries have specialised to process the logs of the locally common softwoods species Scots pine (*Pinus sylvestris* L.) and Norway spruce (*Picea abies* L. Karst.) into sawn timber. Only in the southern regions of Sweden are small volumes of hardwoods sawn. The sawn timber is graded according to either strength characteristics or appearance according to standards after drying at a dry-sorting station before delivery to customers. This way, the customers are familiar with the sawn timber available for purchase from different sawmills. Strength graded sawn timber is only graded according to standards such that, for example, construction companies reliably can follow safety protocols. Appearance grading of sawn timber, on the other hand, is commonly graded according to some customer-specific criteria as different customers are trying to produce different wooden products with distinct appearance styles. Visual strength grading and appearance grading is commonly performed using rule-based grading, which limits what is allowed in a certain grade for relevant measurements, for example cracks, or the size of knots.

Some modern sawmills scan the logs using an X-ray computed tomography (CT) scanner to detect and measure features inside the logs before sawing. CT-scanning could allow for appearance grading of virtual sawn timber fitted inside the 3D CT image of a log and thereby for sawing optimisations based on this virtual estimation of the resulting sawn timber. In addition, grading (virtual) sawn timber prior to drying allows for control over what sawn timber is dried to customer-specific requirements. Such control could reduce the amount of sawn timber dried according to a customer's specifications and rejected at the dry-sorting station.

An overview of the sawmills collaborating with this thesis is shown in Figure 1.1, whose letters will be referenced throughout this thesis in soft parentheses. This figure will be repeated and "evolve" as the thesis progresses and the captions build upon previous captions.

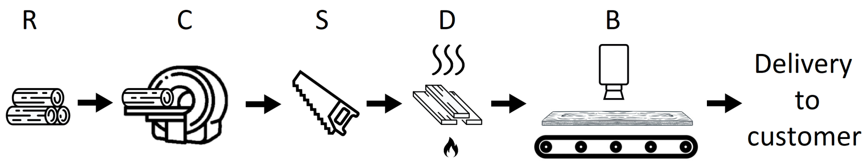


Figure 1.1: A modern sawmill's refinement process of round timber to sawn timber. R, round timber. C, CT-scanning system. S, sawing of the logs. D, drying of the sawn timber. B, camera-based board scanner dry-sorting system.

1.1.2 Motives for a new Grading Approach

Automatic appearance grading of sawn timber is difficult to customise for individual customers or products using today's common rule-based approach in two major ways. Firstly, rule-based appearance grading requires an intricate and coherent description of the desired sorting outcome in a way that is not natural to holistic-subjective appearance grading of sawn timber. Secondly, rule-based grading defines limits on a large number of individual features, which makes customisation troublesome. These problems can lead to a mismatch between customer requirements and sorting outcome, or prohibit customer-adapted grading entirely. Such a mismatch can lead to sawmills delivering, and customers processing, sawn timber not suitable for the end product, which indicates the need for a new method of appearance grading sawn timber that is easy to customise. For example, the collaborating sawmill in this thesis is using a customer-adapted version of standardised grading rules to better suit the needs of a particular customer producing wall panels. Without the customer-adapted grade, the customer might have been forced to purchase a more expensive sawn timber product to meet their product's quality requirements. The customer-adapted grading enables the customer to buy a less expensive sawn timber product with some modifications to the quality grade. Furthermore, if the sawmill can make such modifications to increase the value of the less expensive product, it would be a win-win situation for the sawmill and the customer. Additionally, the benefits of customer-adapted grading could be increased yields of the intended product by the customer, and therefore less volume of sawn timber unnecessarily delivered from the sawmill that could have been sold to a different customer. The benefits of customer-adapted grading could be taken one step further, as CT-scanning (C) allows for control over sawing and drying, which indicates that it is possible to avoid unnecessary drying of sawn timber by the same reasoning. Customer-adapted grading is partially achieved today using a customer-adapted set of rules at the dry-sorting station (B), but neither party is completely satisfied with the sorting outcome, due to the coherency and conciseness difficulties of customising rule-based grading.

1.1.3 Objective and Research Question

Objective:

Make the scanning systems in sawmills able to mimic the appearance grading performed by an experienced manual grader, but faster and more reliably, in a way that is easy to customise and easy to fine-tune.

Research Question:

Can a new machine-learning approach for appearance grading of sawn timber be designed to allow easy customer adaptability with a high agreement with customer requirements?

The goal of the works presented in the this thesis is to appearance grade sawn timber based on its knot structure in a way that is easily customisable for different customers. The core idea of implementing machine learning in sawmills for sorting sawn timber comes from this definition of machine learning – the ability for computer systems to learn a desirable behaviour without being explicitly programmed how to. For example, a customer is experienced with the type of sawn timber they would like to purchase and is good at appearance grading sawn timber as 'buy' or 'not buy'. The customer is using their experience to demonstrate their preferences of the appearance of sawn timber by grading to their needs. It is this skill of grading sawn timber that the this thesis aims to introduce to scanning systems in sawmills by introducing machine learning in a way that is easily customisable for different customers and achieves high grading accuracy.

This thesis introduces a machine-learning method that enables holistic-subjective grading of sawn timber to simplify the customisation process for individual products and customers (Figure 1.2). By using such a method, the utilisation of the forest's natural resources can be increased – sawmills can increase their profitability by only providing suitable material to the customer, and customers can increase their production capacity by only processing suitable material. The works presented in this thesis could be used to set in motion a transition to machine learning-based appearance grading of sawn timber that is beneficial for both customers and sawmills.

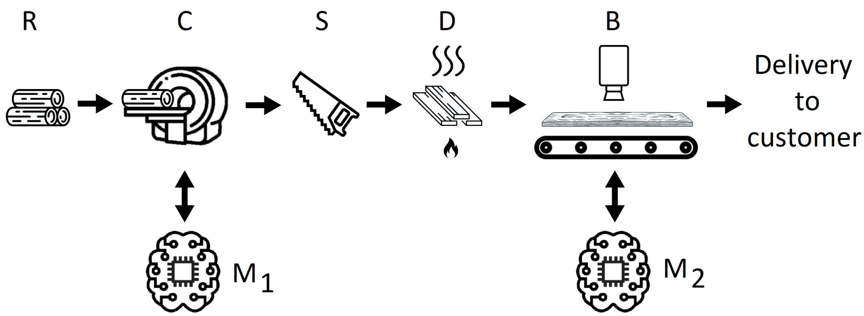


Figure 1.2: This thesis aims to introduce machine-learning models, M, into a modern sawmill's refinement process.

1.1.4 Limitations

Discolourations, rot, or cracks, for example, are arguably more important defects than knots for appearance graded sawn timber, but such defects are simpler to deal with in the sense that they are usually either allowed or not. For this reason, this thesis focuses on appearance grading the knot structure of sawn timber and ignores other defects.

The only material investigated was the centre yield of Scots pine round timber in industrial conditions. No information was gathered from the stands used during the felling of the trees or the bucking of the round timber. All data collection started after pre-sorting the logs as suitable for the wall-panelling customer or grading process, which was done according to standard practices by the sawmill.

1.2 Appearance Grading of Sawn Timber

1.2.1 Aesthetics of Wood Products

Wood products have an appearance we humans find naturally attractive. Its use is common in all parts of the world, particularly in the Scandinavian communities, and in our homes – from floor to ceiling. In part, the large variation and heterogeneity make the wooden material so intriguing to humans. The large number of tree species that exists comes with a large variety of appearances with different colours, wood grain appearances, and knot structures – each contributing to the unique appearance of wood products. While the colour of the wood and the texture represents important appearance aspects of, for example, furniture, which rarely show visible knots, the knot structure (or lack thereof) is commonly the most important aspect for wood products like flooring and wall cladding, where the natural appearance is often highlighted. For such products, the heterogeneous appearance of wood is mainly due to the knot structure, meaning there are often many knots with a variety of shapes and sizes on the wooden surface. If knots are not present, humans still often desire a heterogeneous appearance of, for example, a wooden floor and choose something like a fish-bone pattern. On knot-free surfaces, other features appear more visible to the human eye. In contrast, a homogeneous appearance of these surfaces would be, for example, a painted wall or wall-to-wall carpet. The knot structure of wooden surfaces is important for how humans subjectively interpret the appearance of the wood (Broman 2000), which is why most grading criteria of appearance graded sawn timber are regarding knots.

Humans subjectively perceive knots not only by their individual size and appearance but as a collective of many knots spread across the entire surface. This thesis, and some of its references, refers to this kind of perception as holistic and subjective and is inherent to manual appearance grading of sawn timber. The holistic-subjective perception of knots can be understood by pondering what makes, for example, a wall covered by wall panels constitute a particular style or not and, thereby, be subjectively beautiful. Note that since the appearance of sawn timber, and by extension each wall panel, has to be graded piece by piece at the sawmill, one should consider each wall panel individually and imagine the wall as being covered by similar pieces. Some example wooden surfaces can be found in Broman (2000). First, a wall covered with wall panels with no knots at all would certainly constitute a particular style. However, a wall with only one knot, regardless of size or appearance, would be frowned upon, and the single knot would be perceived as a defect and ruin the otherwise clean, homogeneous style. If each wall panel had several knots of the same size, and the wall had a single knot that was much larger, that knot would also be perceived as a defect, again ruining the otherwise homogeneous style. Lastly, if the wall panels each had several knots of a wide range of sizes, no knot would be considered too large or too small, and the wall would again have a homogeneous style due to the continuous ”messy” knot structure of the wall. By these arguments, it is not inherently the size of each knot that ’ruins’ a particular style or not, but rather the relative size of the knot compared to the holistic appearance of the wall and each wall panel. This illustration with a wall example only covers the size of knots, and the difficulties of appearance grading sawn timber will only become more difficult as more properties of knots are considered, such as shape and colour.

This holistic-subjective perception of wooden surfaces, particularly the ability to grade sawn timber as part of one appearance grade or the other, is the human skill this thesis aims to model using machine learning – this will be considered in section 1.3.

1.2.2 Appearance Grading

Appearance grading (classification) of sawn timber is a subjective process and, as discussed by Broman (2000), is no easy task. For example, in the study by Grönlund (1995), only 57% of 2045 pieces of sawn timber were given the same grade by two manual graders implementing rule-based grading, proving that there is an inconsistency between humans performing the same grading task. Since then, manual graders has been replaced by automatic systems in most sawmills. Such automatic grading systems are often camera-based and perform the final grading of the sawn and dried timber before delivery to a customer. The grading could be used to appearance grade dried sawn timber by a set of rules automatically. This automation allows for higher throughput at the grading station and higher value yield than manual grading (Lycken 2006).

By today's standards, conventional image processing, manipulation, and segmentation techniques have long been used to perform automatic grading. Such techniques have matured over a long period of time and have become the well-functioning systems used today (see section 1.2.3). These algorithms are used in the automatic grading systems studied in this thesis to detect and measure knots and other features and are used as the basis for machine learning implementation.

1.2.3 Rule-based Appearance Grading

In the Nordic countries, the Nordic timber grading rules (NTGR, Swedish Sawmill Managers Association (1994)) are often used to appearance grade Scots pine and Norway spruce sawn timber (Figures 1.3 and 1.4), or similarly by the EN 1611-1 standard (CEN 1999). The idea behind the NTGR was to standardise automatic appearance grading of sawn timber in a way that is in line with manual grading (see older manual grading rules such as Swedish Sawmill Managers Association (1982)), aiming to define the manual grades of sawn timber by a set of rules. The grades of sawn timber were carefully described by rules that define the limits of what features are allowed for each of the standardised grades. The rules govern the size, shape, status, position, number, and proximity to other features (clusters), often in relation to the dimensions of the sawn timber and measured on the worst metre of the piece. Different combinations of feature properties can be governed by

Figure 1.3: Example images of six standard NTGR grades of dimensions 50×150 mm. From www.traguiden.se

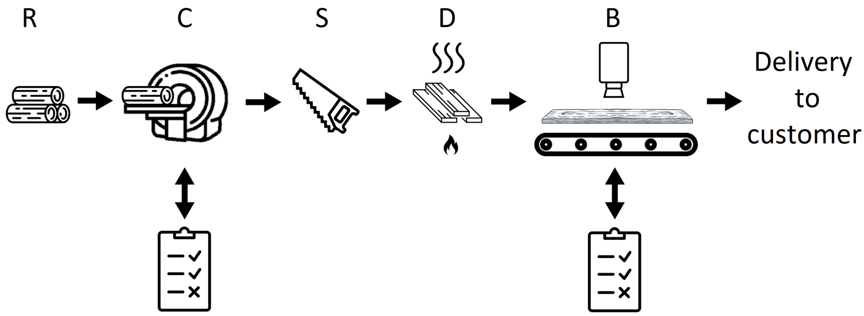


Figure 1.4: Today's use of rule-based appearance grading of sawn timber.

separate rules; for example, for a particular grade, the maximum allowed size of a knot is different for each of the faces of the sawn timber – inner face, outer face, or edges. The benefit of rule-based automatic grading is that it is objective, which means that a grade defined by a set of rules is the same no matter which system was used to grade the sawn timber, with some degree of measurement error. Therefore, rule-based grading is used as a common way to appearance grade sawn timber as standardised grades, such that the customers know what they are buying, indifferent of the sawmill. Nevertheless, each sawmill usually customises the standardised rules to some extent to better suit their material and their customers.

Lycken (2017) investigated rule-based automatic dry-sorting in 15 sawmills, using 4063 pieces of manually graded sawn timber from Scots pine and Norway spruce, and using two different grading standards to grade sawn timber as standardised grades. Each sawmill implemented its own customised sets of rules suitable for the material common to each respective sawmill, thereby allowing the performance of each sawmill to be measured according to their everyday grading operations. The study concluded that a sawmill's dry-sorting station operating under good conditions should be able to maintain a value yield of 98%, a quality yield of 95%, of the already sawn and dried timber, and that knots, in particular the status of the knots, was the reason for 44% of misclassifications (53% for pine and 32% for spruce). Furthermore, even though each sawmill got to calibrate and fine-tune their rules before the tests, some rules were observed to not be optimal in several cases and that some features were not assessed as intended, leading to acceptable but unexpected results. Nevertheless, these results show the strengths of rule-based grading of standardised grades that the sawmills are accustomed to working with.

There are, however, a few problems with rule-based appearance grading of sawn timber. Existing rule-based grades of sawn timber might not fully reflect the customer's needs, and some pieces of sawn timber can abide by the NTGR but can be perceived not to by humans. For example, if more than the usual number of knots on a piece of sawn timber are close to the maximum allowed size, the piece passes the quality rules but would appear excessively knotty to a human who might assign the piece a different grade. Therefore, customer-specific grades are sometimes desired, and a customer-specific set of rules needs to be defined, which need to be coherent and deliberate to result in a concise grading. Making a new set of rules, or customising existing rules, has two main problems (Lycken and Oja 2006), meaning the process is not easily made:

- *coherent* – it is difficult for a customer to describe their subjective view of the desired appearance of the sawn timber in a way that can easily be defined in objective grading rules, and
- *concise* – the number of variables that can be controlled to specify a grade is often more than enough to complicate customisation.

1.3 Machine Learning

Conventionally, programming problems are solved by algorithms – step by step instructions designed by programmers to achieve the desired behaviour. For many problems, algorithms are a reliable, interpretable, and relatively simple solution to implement. Some examples of tasks successfully solved by such algorithms are web pages and online shops, banking transactions, and physics simulations. However, consider the programming problem of writing a program that counts the number of cats in *any* image. Solving this problem by writing an algorithm is unfeasible, even though humans have no trouble counting the number of cats in an image. Humans can easily define the input data, images of cats, and the output data, the number of cats in a given image, but humans cannot translate the skill of counting cats into an algorithm. Enter, machine learning – the ability for computer systems to learn a desirable behaviour without being explicitly programmed. To implement machine learning is to specify the desired behaviour of a program by defining the input and the output data and let an optimisation process compute the algorithm that emulates the desired behaviour. For example, in the context of neural networks, humans specify the desirable behaviour by labelling images of cats with the number of cats in the image and define the architecture of the neural network, after which stochastic gradient descent can be used to compute the algorithm, the neural network weights, that emulates the behaviour of counting the number of cats in the image. The main strength of implementing machine learning for these tasks is to allow humans to perform the task that humans are good at in this example – counting cats – and let the machine-learning algorithms translate this skill into a program. This paragraph was inspired by the white paper by Karpathy (2017), which presents an extended discussion on this matter.

Machine learning has been used in the wood industry to help improve various tasks, see section 1.3.2. A particular interest in using neural networks (NN) and other learning approaches rose in popularity during the 1990s with varying levels of usability. As the machine-learning community matured alongside rapidly improving hardware and computational power, the possibility to effectively implement machine learning in the wood industry increased dramatically during the 2000s and 2010s.

1.3.1 Artificial Intelligence vs Machine Learning vs Statistics

The distinction between artificial intelligence (AI), machine learning, and statistics is not always apparent. For example, it is accepted that neural networks fall under the category of machine learning and that regression techniques fall under the category of statistics. However, both of these examples can be used to estimate an outcome from a new data point. Moreover, each of these methods can be made noise-resistant, making it seem that these models could, at least in theory, be equally capable of performing a task. Furthermore, as artificial intelligence is usually referred to as an agent that can act upon its environment based on the environment, one could argue that a linear regression model and an if-statement making some decision is enough to constitute an AI. To this point, there is a large overlap of words like AI, machine learning, and statistics, and any further debate to distinguish the difference between these words is futile and unwarranted within this thesis.

1.3.2 Applications of Machine Learning in the Wood Industry

There are several applications of machine learning in the wood industry and many more studies in the wood processing research sector. As this thesis focuses on appearance grading, this section only briefly introduces other machine-learning studies not related to this topic. Some studies of particular interest to this thesis are the works related to Giudiceandrea et al. (2011), Ursella et al. (2018) and Ursella (2021), that developed a CT-scanning system that used convolutional neural networks (CNNs) to detect knots in 3D CT-images of logs. Using CNNs, Ursella (2021) implemented a computationally efficient knot detection methodology that measured and detected the size and status of knots inside CT-scanned logs. Of particular importance was detecting of the knot status, as this is critical for many grading criteria regarding sawn timber. The CNN method of (Ursella 2021) outperformed older, algorithmic, knot measurement and status detection techniques by Oja (2000) and Johansson et al. (2013). The detected knots could be used to optimise to sawing of the logs according to value yield. These studies show the high fidelity information available using CT scanning, even before sawing the log. This type of information can greatly increase the value yield from conventional sawing practices like the so-called horns-down sawing (Fredriksson 2014), and the information can also be used for appearance grading of sawn timber, which will be covered later in this thesis.

Some interesting machine-learning studies are as follows, showcasing some of the progress made in the research area during recent times. Hagman (1996) performed extensive investigations on the reflections of wood, presenting a variety of techniques for a variety of tasks, including machine-learning approaches using multivariate image analysis. Understanding the reflective behaviour of wood and wood features is fundamental to developing and interpreting machine-learning techniques based on optical systems, particularly to understand for what tasks machine learning is suitable compared to using traditional techniques like image filtering and image manipulation algorithms. Gu et al. (2010) classified images of knots of different status with very high average accuracy (96.5%) and low false-positive rate (2.25%) in a very rigorous and detailed study using simple features and support vector machines (SVM). The study showed that it is possible to use machine learning for sub-tasks like knot classification, which is very important for subsequent processing or grading that relies on an accurate measurement of the knot structure to give accurate results. Rudakov et al. (2018) used a CNN to perform image segmentation of sawn timber to identify regions that suffered mechanical damage from feed rollers. Such mechanical damage is difficult for visual scanners to interpret as they are not expected and could lead to misinterpretations by, for example, a knot detection algorithm. Zolotarev et al. (2019) tested a few different CNN architectures to trace round timber sawn to boards by using a laser scanner to measure the surface of the logs and train a CNN to identify the boards sawn from that log. The CNNs were trained using CT data, and this study showed the potential of using a CT scanner to train a NN but then only require a surface laser scanner to be able to use the pre-trained NN. The approach indicated the possibility of investing in expensive machines, such as CT scanning systems with rich 3D information of the insides of a log, in one location and using derived machine-learning models in other locations using simpler and less expensive systems. Hu et al. (2019) used transfer learning of CNNs to classify a variety of sawn-timber defects and features with high accuracy. In particular, this study showed how techniques like transfer learning can be used to implement machine-learning techniques with limited sized data sets – a common problem for machine-learning studies. Batrakhanov (2021) used a generative adversarial network (GAN) to simulate sawing of logs and produce RGB images of the sawn timber from surface laser scanning of the logs. Such a technique could, for example, give a sawmill an estimate of sawing outcome to optimise sawing or calibrate the sawing equipment. Alternatively, such simulated sawing results could potentially be used to help visualise simulated results using the Swedish Pine Stem Bank (Grundberg et al. 1995), which has many studies regarding CT-scanning of logs and simulated sawing based upon it.

The few studies cited above shows only a small selection of the available machine learning studies in research areas not directly related to appearance grading of sawn timber. These studies show that machine learning is no longer a novel research area only for computer scientists studying benchmarking data sets but is an incredibly versatile skill that *needs* to be a part of a variety of research subject departments.

1.3.3 Appearance Grading of Sawn Timber using Machine Learning

The topic of appearance grading of sawn timber using machine learning is not covered to a great extent in the literature, and as such, this section is similar to a state of the arts on this narrow field of research. The existing literature introduced machine learning-based appearance grading of sawn timber which allowed for the "software 2.0" methodology described in section 1.3, meaning to manually define the desired grades of a data set and let a machine-learning model learn the skill of grading the sawn timber accordingly. The literature also introduces virtual grading of sawn timber, which means the grading of sawn timber either by simulation or, more commonly, the grading of sawn timber as they are simulated to appear inside the 3D image of a CT-scanned log. Specifically, the literature, as well as this thesis, used the machine-learning method projection to latent structures, also called partial least squares regression (PLS, Wold et al. (2001)). This particular technique is commonly used in the literature due to the possibility to perform classification using highly correlated variables with potentially a lot of noise using a variance-maximising technique (Eriksson et al. 2013). Furthermore, the fundamental concepts of PLS are derived from the common data analysis toolbox related to principal component analysis (PCA). As such, this family of machine learning and data analysis tools are easy to implement and makes the machine-learning models easily interpretable (Eriksson et al. 2013). Other machine-learning methods could be considered instead of PLS, in particular NNs. However, with the already established merits of PLS, and the extra complexity of implementing and interpreting NN models, such models were deemed over-complicated at this early stage of developing machine-learning models for appearance grading. Likewise, methods like logistic regression, k-means-clustering, or support-vector-machines (SVMs), were not chosen as these methods have trouble dealing with either highly correlated variables, outliers, or the high dimensionality of appearance grading sawn timber.

Nordmark and Oja (2004) studied the use of an optical 3D log scanner, and Oja et al. (2004) studied a two-axis X-ray log scanner to predict the potential quality of logs or the potential value of the logs depending on the applied sawing pattern, respectively, using PLS models. This was achieved with promising results based on variables regarding log characteristics, which were different depending on the scanning equipment used for each model. By modelling the value of the logs before sawing for different sawing patterns and similarly for positioning (Fredriksson 2014), the value yield of each log can be increased. In particular, the scanning systems used in the studies by Oja et al. (2004) and Nordmark and Oja (2004) are potential candidates for machine-learning methods. The systems both provide detailed information about the outer shape of the logs and the discrete X-ray log scanner can provide some information about the internal knot structure, such as knot whorls. Such information can be used to geometrically fit a sawing pattern to the log, but with limited information about the knot structure of the resulting sawn timber. As such, it is hard to define rules, as for rule-based dry-sorting of sawn timber, regarding how each log should be sawn. Instead, a machine-learning method, such as PLS, can be used to model the potential value of the log for different sawing patterns, as shown by Nordmark and Oja (2004) and Oja et al. (2004).

The appearance of the knot structure of sawn timber, or lack thereof, was studied by Broman (2000) who investigated the appearance of Scots pine sawn timber and how different characteristics (variables) correlated with customers' preferences, and what kind of appearance is considered desirable by the customers and why – what variables are important. By designing questionnaires regarding the appearance of wood products, the studies included in this thesis used both PCA and PLS to conclude that the impressions of the whole wooden surface was more important than the details – highlighting the importance of holistic-subjective appearance grading of sawn timber.

Breinig et al. (2015b) performed cluster analysis and classification of floorboards according to their holistic-subjective visual characteristics, allowing the grouping of floorboards into cluster grades of similar homogeneous appearance, with distinct differences between clusters. By dividing CT-scanned logs into three sections by the growth-ring orientation of the resulting sawn timber, three separate cluster analyses were performed on 300 floorboards. First, every knot on 300 RGB images of the floorboards was meticulously measured and classified as sound and dead, after which 29 knot variables and one pith-length variable were calculated. Then, based on these variables from 150 floorboards, the cluster analyses were tested for predictive classification capabilities on the remaining 150 floorboards, where some log sections showed acceptable results. Thus, this study showed indicative results for the potential to subjectively and holistically classifying entire faces of floorboards, and thereby sawn timber, by using a set of descriptive feature variables. Furthermore, the fact that the predictions were in part successful showed that enough information regarding the appearance of the floorboards was available in the 30 variables to have a clustering algorithm learn the differences between the holistic-subjective cluster grades.

In a continuation study to Breinig et al. (2015b), the study presented in Breinig et al. (2015a) used simulated sawing of CT-scanned logs to optimise the rotational angle of the logs to yield the most number of floorboards of cluster-grades defined as in the previous study. Using the same clustering technique as in the previous study, each cluster in each log section represented a cluster grade for classification using PLS models based on the CT-scanners measurements. Using several PLS models, the rotation of the log was optimised to result in the most number of floorboards of these cluster grades. This technique showed the PLS models managed to appearance grade the virtual sawn timber in a satisfactory way that showed a homogeneous appearance of the boards within a grade and clear distinctness between the grades. This study showed the potential to subjectively and holistically appearance grade virtual sawn timber inside CT-scanned logs before the sawing of the log. Fredriksson (2014) showed that the log's positioning is important for the value and volume yield of sawn timber, and now the works by Breinig et al. (2015a,b) studies showed that similar ideas could be used for specific products.

Using strict definitions of appearance grades, as opposed to Breinig et al. (2015a,b), and Broman (2000), PLS was used by Lycken and Oja (2006) to appearance grade sawn timber by quality grades according to rules based on the NTGR. The study implemented PLS using a similar set of variables to Breinig et al. (2015b) on each face of the sawn timber, mostly regarding knots, that was based on manual measurements from images of the sawn timber. This study showed the potential of using PLS models to appearance grade sawn timber in line with strictly defined grades and agreed to 81% with manual grading according to the NTGR. As the grading performed by a PLS model resembles manual grading in the sense that there is no strict limit of, for example, the maximum size of knots, it is expected that there would be some visual difference of a grade, depending on what method was used to perform the grading. However, no such study has been performed.

Berglund et al. (2015), a continuation of Lycken and Oja (2006), studied a customer-adapted sorting comparing the performance of a rule-based grading and a PLS-based grading methodology. A much larger set of knot variables were created by repeating 56 knot variables on each face of the sawn timber and in several different sections of the sawn timber, resulting in a total of 1566 highly correlated variables. This time, the variables' basis was measurement performed by a camera-based automatic grading system used for rule-based dry-sorting, which introduced measurement errors. The study concluded that the PLS-based grading performed better than the rule-based grading performing the same task compared to a manually defined true grade. However, the rules, which was the collaborating sawmill's active sorting rules for this large customer, suffered from the two problems of coherency and conciseness when customising a rule-based grade, as described in section 1.2.3. If the rule-based grading would have been replaced by the PLS-based grading, this study showed how the difficulties of customising rules for a particular customer could be replaced with one task of labelling a data set. This labelling task greatly simplified the process of creating new customised quality grades and ties back to the discussion in section 1.3 regarding how humans should be allowed to perform the task they are good at – grading sawn timber.

The choice of variables in Broman (2000), Lycken and Oja (2006), Berglund et al. (2015), and Breinig et al. (2015a,b), and in the studies presented in this thesis, is the seed for these kinds of studies and therefore needs to be chosen deliberately. However, why some variables are included and others are not is difficult to answer. The set of variables need to be comprehensive enough such that the necessary information to perform the task at hand is somehow available to the machine-learning model. This problem of defining a set of variables relates to the discussion in section 1.3 – humans have a difficult time intuitively understanding how a single variable is, or is not, important for the task at hand and to what degree. It is even more so when each variable is considered in the context of the selected machine-learning method with its particular strengths and weaknesses. Therefore, the task of defining a set of variables is not an easy task. It is, however, possible to create a set of variables that allows for clustering and grading of the appearance of sawn timber, as shown by the references above.

The choice of reference grades used in the aforementioned studies are in contrast with each other. Breinig et al. (2015a,b) and Broman (2000) used a subjective interpretation of the appearance of sawn timber to perform homogeneous appearance clustering while Lycken and Oja (2006) and Berglund et al. (2015) used a rule-based standardised grade or a customer-adapted rule-based grade as reference. These studies show the potential to appearance grade sawn timber according to conceptually different grades. As a continuation of these studies, this thesis will focus on product-adapted appearance grading of Scots pine sawn timber.

Chapter 2
The Machine-Learning Method

Machine learning is ”software 2.0”

To achieve the objective of sawmill scanning systems that understands the customers' preferences of sawn timber, this thesis implemented a machine-learning method. Implementing a machine-learning method consisted of collecting data sets with an input, $\mathbf{X}$, and an output, $\mathbf{Y}$, and the relationship (transformation) between the two was modelled using machine learning. To show the feasibility and benefits of using machine learning to holistically and subjectively appearance grade sawn timber, this thesis studied a modern sawmill that delivers sawn timber to a customer producing interior wall panels.

Here follows an overview of the Publications included in this thesis and the associated data sets. Two scanning systems were investigated – a CT scanning system and a board scanning system (sections 2.5.1 and 2.5.2). These scanning systems measure the knot structure of the sawn timber at different stages of the sawmill's refinement process. Based on their measurements, each system used a MATLAB® (The MathWorks, Inc. 2021a) script to compute a set of statistical variables, $\mathbf{X}$, that described the knot structure of the sawn timber (section 2.7.1). By tracking each piece of sawn timber from the point of being scanned by either system, to the point of becoming a wall panelling product, the final *product grade*, $\mathbf{Y}$, could be matched with the variables calculated from the scanners, $\mathbf{X}$. A machine-learning model could then be trained for each system (or, in theory, both simultaneously), giving the scanners the ability to appearance grade sawn timber as suitable or not for the final wall panelling product. Furthermore, instead of using the final product grade, which requires the processing of unsorted sawn timber, the automatically captured images from the board scanner were used as the basis for appearance grading. The images were carefully inspected and graded, similarly to traditional manual grading, as being suitable or not for a particular wall panelling product. Such a grading process requires an intricate understanding of the customers' requirements and the production process of the final product. This so-called *image grade* of the sawn timber was investigated as a substitute for the product grade but with a much simpler logistical process of collecting the reference grade, $\mathbf{Y}$.

In parallel to the studies regarding the wall-panelling customer (Publications II, III, V, VI, and VII) two studies were conducted on two separate unrelated, but similar, data sets using the NTGR as a quality reference (Publications I and IV). These two studies complemented the understanding of appearance grading sawn timber using machine learning gained during the studies directly related to the wall-panelling customer. For each of the studies presented in Publications I–VII, a data set was collected, and an overview of these studies is shown in Figure 2.1.

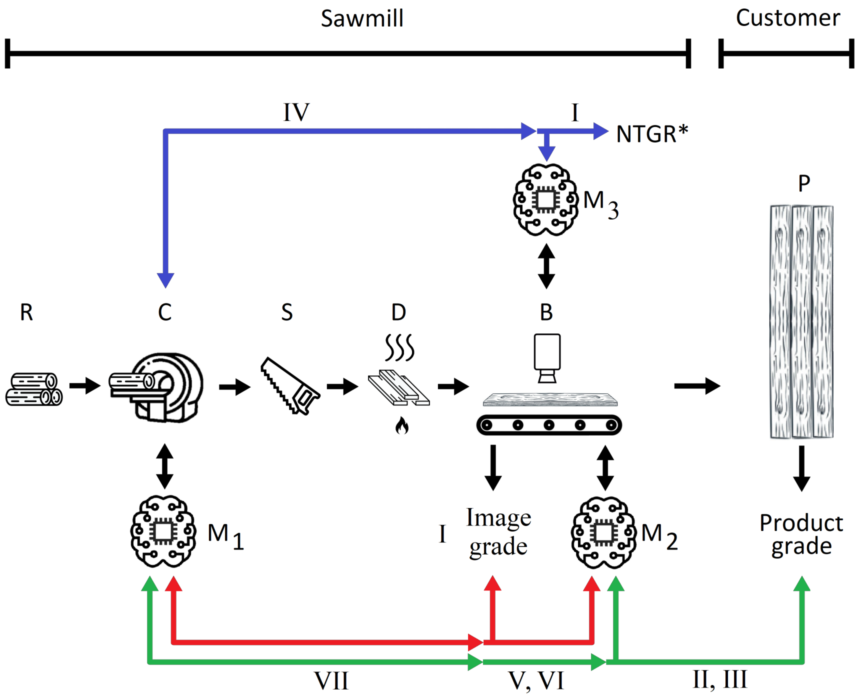


Figure 2.1: Thesis overview. Each coloured arrow indicated what type of data is available to the data set of each Publication I–VII. Each machine-learning model M is associated with a scanning system, which provides measurement inputs, $\mathbf{X}$. Each model is also connected to at least one grade, $\mathbf{Y}$, indicated by green for the product grade, red for the image grade, and blue for the NTGR grade. The roman numeral of each Publication (I–VII) is written next to one or two coloured arrows, indicating what data was available for that study. In Publications II, III, V, and VI, each Publication had access to all the previously collected data from (B) and (P) of these data sets. The letter I indicates the image grade referenced in the body text. The NTGR grade used was not related to the wall-panelling customer.

2.1 Collaborating Parties

The main collaborating industrial partners in the studies presented in this thesis were Norra Timber and Lundgrens hyvleri. Norra Timber owns the two sawmills Sävar Sawmill and Kåge Sawmill, that enthusiastically participated in the studies. Sävar Sawmill has a CT scanner and both sawmills uses the same dry-sorting system. This sawmilling company is located in northern Sweden and specialises in producing Scots pine and Norway spruce sawn timber (Figure 1.3). Lundgrens hyvleri is a customer of Norra Timber, also located in northern Sweden, producing wall panels from pine and spruce sawn timber (Figure 2.2), and is a prime candidate for holistic-subjective appearance grading of sawn timber. The parties will be referred to as "the sawmill" and ”the wall-panelling customer” throughout this thesis.

Figure 2.2: Three examples of Scots pine wall panels from Lundgrens hyvleri. Source: www.lundgrenshyvleriab.se

Additionally, Microtec, the producer of advanced scanning systems for the sawmill industry, such as the Microtec CT Log detailed in section 2.5.1, contributed with their expertise of CT scanning round timber during the data collection and data analysis.

2.2 Knot Properties

The most important property of a knot is its status – sound (alive), dead, rotten, etc. (Figure 2.3). When a knot dies, it stops growing and slowly darkens over time, and as the trunk of the tree continues to grow, it slowly covers the dead knot. This eventually leads to dead knots being encased in bark, appearing as a dark ring around the knot on the sawn timber. The distinction between a dead knot and a sound knot is by no means easy to make, not by human nor by machine. Even if a clear definition of a dead knot is referenced, humans and machines sometimes struggle to distinguish between a mostly-sound knot and a mostly-dead knot as they appear so similar. Therefore, the status of some knots is ambiguous, which is not always accurately represented by a binary sound or dead status measurement.

Figure 2.3: Examples of Scots pine knots: I – sound knot, II – dead knot, III – rotten knot, and IV – spike knot, that in this case reaches the edge of the sawn timber. From www.traguiden.se

The shape of the knot is also important for the appearance of sawn timber. As it appears on the surface of sawn timber, the shape of a knot depends on growth conditions and sawing pattern (Figure 2.3). Furthermore, the growth conditions can sometimes drastically change the appearance of the growth rings around the knot, possibly making the edge of the knot hard to detect. Knots are often cut across their lengthwise direction and appear as round or oval on the surface of the sawn timber. Spike knots refers to the shape of a knot as it appears on the surface of the sawn timber when the knots was sawn along the growth direction – along the pith.

The appearance of the knot structure on sawn timber depends on the knots' status, shape, size, and number. When combining these features, there exists a large variety in the individual appearance of each knot. When viewed as a knot structure on the surface of sawn timber, the large variety of knots leads so an even bigger variety of sawn timber appearances – which may or may not constitute a beautiful style, as discussed in section 1.2.1.

In the context of wall panels, dead knots, and especially dead spike knots encased in overgrown bark, can drastically affect the appearance of sawn timber as they appear pronounced. The appearance of dead knots may, or may not, be acceptable depending on the product, but dead knots are in general difficult to process as they tend to crack during splitting, profiling, or planing. As spike knots usually only appear on the inner face of sawn timber, the face closest to the pith, they can be hidden by using the inner face of the sawn timber as the backside of a wall panel (Figure 2.4).

2.3 Product based Quality Assessment

The sawn timber delivered to the wall-panelling customer was used to produce wall panels. Each piece of sawn timber was split into three boards – each profiled and planed (Figure 2.4) – which was the final product considered in this thesis (P). Neither the inner face nor the outer face of the sawn timber was used as a visible wall panel face. Each wall panel is then subjected to a manual holistic-subjective quality grading. The intended wall panelling product had two grades labelled as A, for the desired grade, and B, for the undesired grade. For the sake of the studies of this thesis, only knots were considered by the manual graders during this grading procedure. The most common reason for a wall panel being assigned the grade B was knot holes or cracked knots from the manufacturing process, which is most commonly a problem with dead knots but can happen with sound knots as well.

Since each piece of sawn timber results in three wall panels, and since the grading of the sawn timber happens at the sawmill, the quality grade of the three resulting wall panels had to be translated into a grade for the piece

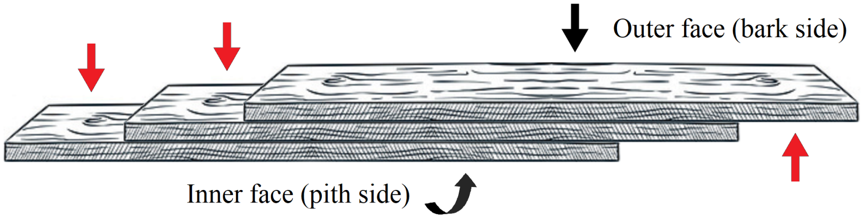


Figure 2.4: Image showing how the sawn timber is split into three wall panel boards. Red arrows indicate visible surfaces once the wall panels are installed.

of sawn timber. Therefore, the majority grade of the three wall panels was assigned to the piece of sawn timber and was called the *product grade* of that piece. For example, a piece of sawn timber resulting in two wall panels of grade A and one of grade B was given the grade A. The same labelling A and B was used both for the grade of the sawn timber (P) and the grade of each wall panel (P). Publication IV also uses the labels A and B but designates unrelated NTGR grades.

2.4 Image-based Quality Assessment

One way to assess the quality of sawn timber is by images of the sawn timber. The camera-based automatic dry-sorting system (B) could store the raw images of the sawn timber, and they were made presentable for easy human inspection. Depending on the grading process in mind, such a human inspection could be made with millimetre (pixel) precision to follow standards like the NTGR, or the grading could be made holistic-subjectively according to an intended quality grade. In the study presented in Publication V, such images of the sawn timber was used to identify timber sawn from Scots pine butt logs. In the studies presented in Publications VI and VII, a holistic-subjective grading based on four faces of the sawn timber was made to estimate the resulting product quality, A or B, described above using the software shown in Figure 2.5. This grade was called the *image grade* of the sawn timber (I). This visual grading process is relatable to the works by Broman (2000) and Breinig et al. (2015a,b).

Some amount of guesswork was required to determine the product grade of the sawn timber by images, since none of the surfaces of the sawn timber were the wall panels' front faces (Figure 2.4). As such, grading sawn timber as suitable or not for this product requires intuition about how knots appear on the surfaces of the sawn timber and how this translates to the appearance of the knots on the visible faces of the wall panels. Each knot is usually visible on at least two sawn timber faces, which helps assess if the knot is sound or dead through the piece. To alleviate this uncertainty, the image grade could be assigned a sub-grade of "certain" or "uncertain" (Figure 2.5), which was explicitly investigated in the study presented in Publication VI.

Images of the sawn timber were graded using a simple graphical user interface created as a MATLAB® app specifically for this purpose (Figure 2.5). With the help of two researchers, the wall-panelling customer's quality expert graded approximately 300 pieces of sawn timber to be used in the study presented in Publication VI. At this time, it was also agreed that the two researchers working on studies related to this product could perform a similar grading as the quality expert had done, which was done in the study presented in Publication VII. For consistency's sake, the images were graded as of grade A or B for easy comparison with the product grade, which was called the *image grade* of the sawn timber.

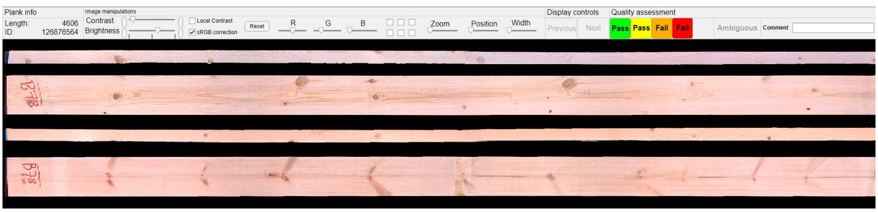


Figure 2.5: The software (a MATLAB® app) used to grade sawn timber by visual appearance. The software provides some simple visual manipulation tools to make the four images of the sawn timber appear easy to interpret and allows for grading the sawn timber piece as pass (buy) or fail (not buy) with optional "certain" or "uncertain" sub-grades. Only about half of length of the sawn timber images are currently visible and extends to the right.

2.5 Scanning Systems

Two scanning systems were investigated in the studies presented in Publications I–VII: a CT scanner and a board scanner. Other systems, like those used to pre-sort the logs at the log yard during data collection, were used according to standard sawmill practices but were not monitored in any way.

2.5.1 X-ray Computed Tomography Scanning

The X-ray computed tomography (CT) scanner used in this thesis was the Microtec CT Log (Giudiceandrea et al. 2011, Ursella et al. 2018) (Figure 2.6) installed at Sävar Sawmill in 2018. This CT scanner allows for 3D tomographic reconstruction and sawing optimisations at speeds up to 160 m/min, 24 hours per day. At this speed, the CT scanner measures the density inside each log in voxels $1 \times 1 \times 10$ mm in size (10 mm along the longitudinal direction of the log). These voxels are used to detect features inside the log, such as cracks or heartwood based on density measurements, and in parallel, the voxels are passed through a convolutional neural network that detects and classifies knots (Figure 2.7, centre). Once the internal features are detected and measured, a sawing optimisation algorithm iteratively creates virtual sawn timber according to available sawing patterns and the detected features are projected onto the 2D surface of virtual sawn timber (Figure 2.7, right). Finally, the virtual sawn timber is graded, and based on this grade the optimisation algorithm iterates the positioning of the log to find the position with maximum quality or value yield.

The CT Log installed at Sävar Sawmill was installed directly in front of the saw such that the CT scanning, reconstruction, feature detection, sawing optimisation, and sawing of a log takes place in a matter of seconds. Alternatively, a CT Log can be installed in the log yard, where each log is CT scanned and sorted before sawing according to some characteristics with fine-tuned sawing optimisation. This approach enables a wider range of optimisation options. Then, possibly days later, the logs are scanned again before sawing to retrieve the correct sawing optimisation from the CT Log using fingerprinting to identify each individual log. Ursella (2021, §3.10) presents a nice overview of the possibilities and benefits of using fingerprinting techniques to track sawn timber along the sawmill production line. This approach is used at Fiskarheden Sawmill in Sweden. Both alternatives of

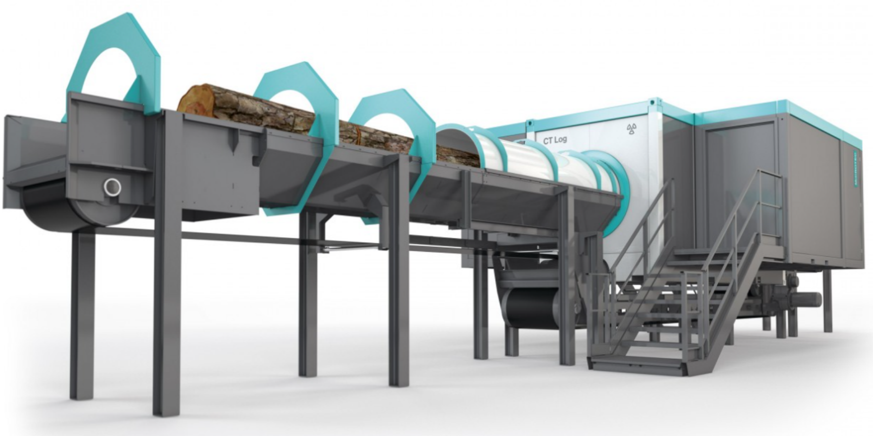


Figure 2.6: The Microtec CT Log – an example of an X-ray computed tomography log scanner.

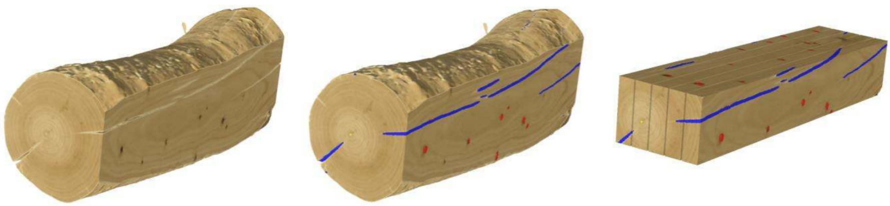


Figure 2.7: Example of a reconstructed log from the CT scanning (left), highlighted example features (centre), and creation of virtual sawn timber, as decided by the sawing optimisation (right). From Ursella et al. (2018).

installing a CT scanner are available depending on the sawmill's production strategies.

In this thesis, the data extracted from the CT Log was mainly the knot descriptions available for the virtual sawn timber (Figure 2.7); general density information was also extracted. To avoid excessive formalism, the use of "sawn timber" in this thesis will be used loosely and may sometimes refer to "virtual sawn timber" or "the resulting sawn timber from a CT scanned log", in the context of CT scanning.

The knots on each face of the virtual sawn timber were measured according to the NTGR and included:

- the position of the knot and its size in longitudinal and transversal direction,
- whether or not the knot is dead,
- whether or not the knot touches a corner of the sawn timber,
- whether or not the knot is a spike knot, and
- whether or not the knot is steep – growing at a high angle, close to the pith.

Based on these measurements, statistical variables were created (section 2.7.1).

2.5.2 Camera-based Dry-sorting Scanning

The camera-based dry-sorting board scanner used in the studies included in this thesis was the FinScan Boardmaster (Anon. 2018). The Boardmaster uses several cameras with a resolution of 0.25 mm^2 to scan the four faces (inner face, outer face, and edges) of the sawn timber and detect features at up to 240 boards per minute.

The Boardmaster measured features according to the definition by the NTGR and included:

- the position of the knot and its size in longitudinal and transversal direction,
- whether or not the knot is dead,
- whether or not the knot is enclosed in bark,
- whether or not the knot is rotten, and
- the position of bark inclusions and their size in the longitudinal and transversal direction.

Based on these measurements, statistical variables could be created (section 2.7.1).

2.6 Data Collection and Data Structure

2.6.1 The Material and Data Collection

The material consisted of 50×150 mm centre yield pieces of sawn timber from the Scots pine logs, 3.4–5.4 m in length. The sawn timber was kiln dried to approximately 14% moisture content.

As discussed in the introduction regarding the software 2.0 methodology, the important task for humans is to diligently collect and label data in a way that is representative of the real world. The labelling process demonstrates the skill of grading sawn timber in an industrial setting which a machine-learning method is then implemented to learn from the collected data. In the studies included in this thesis, data was collected from a few different sources with similar data structures (Figure 2.1). Each data set consisted of a set of descriptive variables, $\mathbf{X}$, almost exclusively related to knots, and a single response variable, $\mathbf{Y}$, which is the grade of the sawn timber as either the product grade, image grade, or NTGR grade.

2.6.2 ID-marking and Tracking

Referencing Figure 2.1, the studies presented in Publications IV and VII started the tracking of the material consisting of pre-sorted logs (R) at the CT scanning stage (C), while the studies presented in Publications I, II, III, V, and VI started the tracking of the material at the dry-sorting station (B). In the studies directly related to product adapted grading of sawn timber (Publications II, III, V, VI, and VII), the tracking of the material ended as profile-planed wall panels (P), while the studies presented in Publications I and IV ended the tracking at the dry-sorting station by assigning the sawn timber with an NTGR grade.

This thesis relied entirely on manually labelled and tracked material, detailed below, which was used, but not explicitly described in Publication VII. Similar, or at least effectively the same, methods of labelling and tracking were used in Publications I–VI but with different starting points of tracking. Publication IV discussed how these kinds of data collections could potentially be performed by a traceability method throughout the sawmilling process. However, traceability was not possible through the production process of the wall-panelling customer, which had to be done manually. For this reason, all of the data collections were done manually.

The butt-end of logs were colour coded with bright and distinct colours and ID-marked with an identification number before CT scanning and sawing (Figure 2.8a). The identification number was a two-digit number, giving a large range of unique ID markings in combination with the colour coding. The ID number followed a numbering scheme designed to be easy to read in any orientation without being misread, as logs are rotated, and sawn timber is often flipped. For example, the number 21 was not a valid ID number as this could be misread as the precursor number 12 if flipped upside down while moving on a conveyor belt (ᘔƖ). As this unique colour and numbering identifier was unique for each log, each piece of sawn timber from the same log shared the same markings, meaning both of the two centre yield pieces. To resolve this, an additional marking was added just after the sawing process to indicate the sawn timber's position inside the log. Studying two centre yield pieces, only the left or right position in the log needed to be tracked and could be achieved with a single marking on pieces on the right. At this point, each piece of centre yield had a unique marking on them that could identify from which log the timber was sawn and what position the virtual sawn timber had in the log when it was scanned by the CT scanning system. Tracking this unique marking, each wall panel was matched to the sawn timber it was produced from (Figure 2.8b).

In theory, the order in which the sawn timber exited the sawing machine would be the same throughout drying, dry-sorting, and further processing. However, as the sawn timber is likely to be scrambled on conveyor belts, the unique markings were moved (repeated) from the butt-end of the sawn timber to the flat faces of the sawn timber. This way, the unique markings were visible in the images captured by the camera-based automatic dry-sorting system used. This step was the first step during data collection for the studies only regarding the board scanner.

This data collection method (or traceability methods) could have be used to create a reference data set for the sawmill, and any data collected from any system involved in handling the sawn timber could have been collected. For example, in this thesis, data from the X-ray CT scanning and the dry-sorting visual scanning was collected; however, any number of scanning systems

(a) Logs at the sawmill's log yard colour coded and stamped with an ID number.

(b) Stacks of wall panels. Each patch of colour is one piece of sawn timber split into a triplet wall panels.

Figure 2.8: Images from start to finish of one data collection.

could have been used to collect more data for unrelated studies to this thesis. At this point, any number of different grades or outcomes could be appended to the collected data set, after which the data set could be used as a reference data set for any number of machine-learning studies. Of course, with only one destructive quality grade such as a bending strength test, or as in this thesis – producing wall panels.

This data collection could have been made using automatic tracking of the material. Pahlberg (2017) studied automatic tracking of sawn timber between camera-based systems, and Flodin et al. (2008) studied tracking between a CT scanner and a camera-based green sorting system. Both works are examples of the possibility to track individual pieces of sawn timber through different steps of the sawmilling process, and Möller (2019) tracked sawn timber from a CT scanner to the final dry sorting with varying accuracy depending on the implementation. By using such methods, a data collection like the one discussed above could be made automatically. Even if a tracking algorithm is not 100% accurate, Möller (2019) showed the possibility to track above 90% of the sawn timber with 100% accuracy, which over time could amount to any sized data set. As the data collected in this thesis regarding product-adapted grading needed to pass through the customer's refinement process, which did not allow for traceability, any data collection including the customer was done entirely manually. Publication IV explored the possibilities of using traceability from the CT scanning to the board scanning, based on Möller (2019).

To reiterate the introduction, the possibilities and benefits of traceability of material through a sawmill is nicely covered by Ursella (2021).

2.7 Machine-Learning Implementation and Analysis

Sections 2.3 and 2.4 introduced the quality grades A and B for the product grade and image grade, respectively. Figure 2.1 shows an overview of the available data showing the input, $\mathbf{X}$, from the CT scanner (C) and the board scanner (B), and the output, $\mathbf{Y}$, as the product grade (P), and the image grade (I), or the NTGR grades. The available data was collected as multiple data sets with either of the $\mathbf{X}$ and $\mathbf{Y}$, or both.

2.7.1 Available Data Sets and Variables

The attributes of the available data sets, variables, and grades will be detailed in Chapter 3, as the intricate combination of these parts and the respective results must be explained coherently and concisely. Therefore, the combination of attributes of each data set will only be covered in brief here, while the actual calculation of variables will be the main focus.

Five data sets were collected according to the tracking methodology described in section 2.6.2 – one for each of the studies presented in Publications II, III, V, VI, and VII. These five data set were considered primary, as they related to the wall-panelling customer. The two data sets associated with Publications I and IV were considered secondary as they did not relate to the wall-panelling customer. The secondary data sets were collected similarly to the primary data sets and were automatically attributed with grades according to the NTGR by the board scanner, as well as a manual quality grade reference used for training of the machine learning models.

Each of the data sets was tracked according to section 2.6.2, and was attributed with a visual quality assessment and product-based quality assessment according to Figure 2.1, section 2.3, and 2.4. One of the data sets were different from the other in that the sawn timber was partially covered in dust (see Publication III), while the remaining data sets were impossible to distinguish from each other by visual inspection. As all data sets passed through the dry-sorting station, posterior analysis of the entire data collection was made to confirm this statement. The data sets each consisted of approximately 250–300 pieces of sawn timber, which amounted to 1454 pieces in total.

Each data set was also attributed with variables calculated from measurements from each investigated scanning system related to the corresponding study. Out of the primary data sets, one data set had variables from both the CT scanner and the camera-based board scanner. The other primary data sets only had a set of variables from the board scanner. The secondary data sets had similar, but different, variables attributed to them from both or either of the CT scanner and the board scanner. These sets of variables were different for each secondary data set as different persons were in charge of calculating the variables due to mismatched formats that prevented the re-purposing of code-bases.

The board scanner's measurements of the knot structure on the surfaces of the sawn timber (see section 2.5.2) were used to calculate a large set of variables (Table 2.1). A collection of 22 variables, such as "total number of defects" or "maximum defect size", were calculated for 6 different defect types on 3 different faces of the sawn timber (edges together), and repeated in 9 different zones along the length of the sawn timber (Figure 2.9). This amounts to a total of $22 \times 6 \times 3 \times 9 = 3564$ highly correlated variables. The calculation of similar variables was originally done by Berglund et al. (2015) and again by Olofsson et al. (2017), which was the precursor to the set of variables detailed here.

The CT scanner's virtual sawn timber has similar information and data structure as the sawn timber scanned by the board scanner, namely a 2D measurement of the knot structure of the surfaces of the virtual sawn timber. The CT scanner's measurements of the virtual sawn timber were used to calculate a set of variables similar in aim and scope to the one detailed in Table 2.1. This CT-based variable set is not detailed here because it is similar enough to what is detailed in Table 2.1.

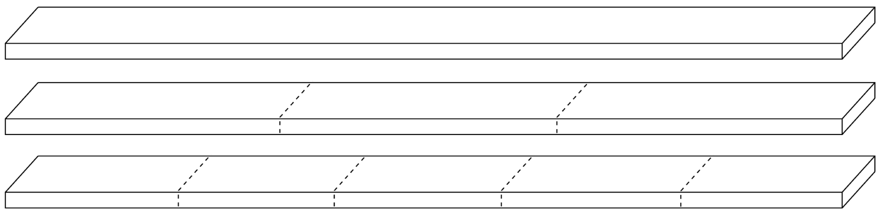


Figure 2.9: Illustration of the virtual copies and sectioning of each plank (Berglund et al. 2015).

Table 2.1: A complete description of all the variables calculated using the board scanner measurements. Similar variables were calculated based on the CT scanner measurements. Each variable in the first column was replicated for each entry in each of the other columns. The number of entries of each column is shown at the bottom, where the total number of variables is the multiple of the values in the bottom row.

#	Variables			Defect types		Plank faces		Sections
1	Total Number of defects			Sound knot		Inner face		1
2	Avg. defects size		(mm)	Dead knot		Outer face		3
3	Std. dev. of defects size		(mm)	Sound or dead knot		Both edges		5
4	Maximum defects size		(mm)	Bark-ringed knot				
5	Sum of defect area		(mm^2)	Rotten knot				
6	Ratio of $\frac{\text{defect area}}{\text{surface area}}$		(%)	Bark pocket				
7	Ratio of $\frac{\text{sum of defect area}}{\text{sum of all defect areas}}$		(%)					
8	Number of defects per meter		(m^{-1})					
	Ratio of defects of sizes:							
9	≤ 9 mm		(%)					
10	10–19 mm		(%)					
11	20–29 mm		(%)					
12	30–39 mm		(%)					
13	40–59 mm		(%)					
14	60–79 mm		(%)					
15	≥ 80 mm		(%)					
	Number of defects per meter:							
16	≤ 9 mm		(m^{-1})					
17	10–19 mm		(m^{-1})					
18	20–29 mm		(m^{-1})					
19	30–39 mm		(m^{-1})					
20	40–59 mm		(m^{-1})					
21	60–79 mm		(m^{-1})					
22	≥ 80 mm		(m^{-1})					
	22 variables = 3564 variables	$\times$		6 types	$\times$	3 sides	$\times$	9 zones

2.7.2 PCA, PLS, and Software

Multivariate principal component analysis (PCA) and projection to latent structures (PLS[1]) are commonly used tools for data analysis and regression modelling with predictions and is covered thoroughly in online sources and the literature, such as Wold et al. (1984, 2001) and Geladi and Kowalski (1986). In particular, the work of Eriksson et al. (2013), detailing computational implementations, analytical tools, and illustrative examples of PCA and PLS, which also is the basis of the Sartorius software SIMCA® (previously developed by Umetrics®) has been used throughout this thesis. Below is an overview of PCA and PLS for ease of reference. PCA was not explicitly used in the Publications in this thesis, but was extensively used as an analysis tool to further the development of the PLS models used.

PCA reduces the number of dimensions (number of variables) of a data set to easily describe it, while retaining as much information as possible and remove noise. The reduction of dimension is achieved by linearly combining the dimensions of the original data set into orthogonal so-called principal components (PC). By reducing the number of dimensions, underlying latent (hidden) structures and variable correlations can become apparent. The first such PC is computed as the eigenvector of the covariance matrix of the data set. For example, reducing a data set regarding cities in the USA from 9 dimensions to 2 greatly simplifies the interpretation of the data (based on the documentation presented by The MathWorks, Inc. (2021b)). This is particularly true for data sets with a very high number of dimensions (several hundred) which often can be described to a large extent with only a few PCs. A PC is computed to fit a data point cloud so that the variance of the residuals is minimised by least squares (Figure 2.10). Iteratively, more PCs can be added following the same procedure, where two components seems the most common – possibly because it allows for nice 2D plots of the data.

PLS regression also reduces the number of dimensions of a data set like PCA. Still, it is different from PCA because it is not meant to describe a data set, but to use the data to make predictions by regression or discriminant analysis (classification). Similarly to PCA, PLS regression iteratively

[1]PLS have in the literature been referred to as ”projection to latent structures”, ”partial least squares regression”, and ”projection to latent structures using partial least squares”.

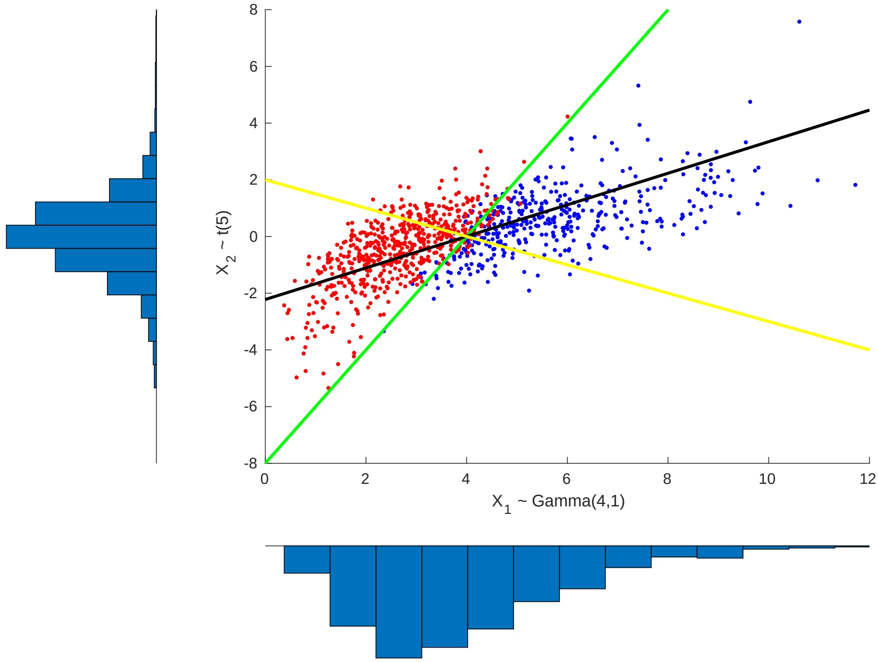


Figure 2.10: Simulated data used to illustrate the differences between PCA and PLS – for illustrative purposes only. The simulated data consists of two dependant and correlated variables, X_1 and X_2, following the Student-t distribution and Gamma distribution, respectively, as shown by the histograms. Red (class 0) and blue (class 1) data points indicate two classes separated by the green line and some noise. The black line is a principal component (PCA) along the direction of maximum variance in the data, regardless of class. The yellow line is a latent structure used for PLS discriminant analysis (PLS-DA), estimating the direction that maximally separates the classes. Based on the documentation presented by The MathWorks, Inc. (2021c)

computes latent structures to perform the regression or classification (Figure 2.10). This iterative process is repeatedly checked for overfitting using cross validation, meaning the model is only trained on relevant information and not noise, which should be removed from the regression. These two steps arguably differentiate PLS from ordinary regression and makes PLS part of the realm of machine learning.

To perform classification using PLS on the entire simulated data set shown in Figure 2.10, meaning self-prediction or predicting the training set in machine-learning terms, the entire data set is projected onto the latent structure (Figure 2.11). Then, by introducing a *classification threshold* of the predicted class of, for example, 0.1, all observations with a predicted class below 0.1 are classified as of class 0, while all observations above 0.1 are classified as of class 1. In this example, there is a large overlap of the classes along this latent structure, indicating the need for additional latent structures.

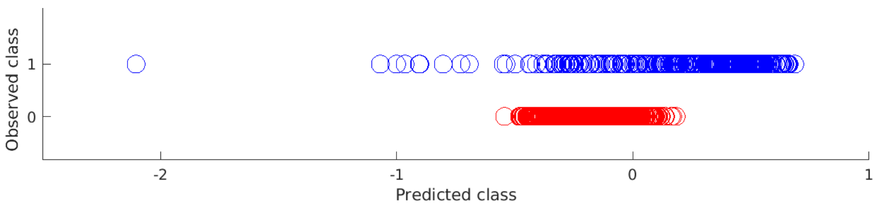


Figure 2.11: Observed vs predicted plot of the latent structure shown in Figure 2.10.

Chapter 3

Results and Discussion

Machine learning can model subjectivity

The objective of implementing a machine-learning model to perform holistic-subjective appearance grading of sawn timber in a way that is easy to customise and control was achieved using PLS models. The grading accuracy of the PLS models implemented was on par with, or exceeded, the performance of traditional rule-based grading. Furthermore, the sorting outcome was easy to control, giving the sawmill and customer the ability to balance high volume deliveries with high product yields. This was achieved by studying the implications of the training data on the grading performance of the PLS models.

3.1 PLS-based vs Rule-based Grading Accuracy

When considering the grading accuracy of appearance graded sawn timber, it is important to avoid unfair comparisons. Different grading tasks have different grading difficulties, and as such, the grading accuracy of one task should not be compared with the grading accuracy of another.

As the accuracy of standardised rule-based appearance grading of dried sawn timber was stated in the introduction to be good – maintaining a value yield of 98%, a quality yield of 95% (Lycken 2017) – one might question the need for a different grading method. These metrics are important for customers as a highly accurate quality assessment means that the customer knows what to expect of the delivered sawn timber. It is important to understand that these metrics are for standardised appearance grading – something the sawmills investigated are very familiar with, and their performances have improved over a long time.

A CT scanner allows for the grading of virtual sawn timber. The virtual grading performed in this thesis used rules to grade the virtual sawn timber for many sawing positioning alternatives to optimise the value yield of the log. This virtual grading was very similar to the grading performed by the board scanner at the dry-sorting station, with similar information available. However, due to technical reasons, the CT scanner was made to perform the PLS-based appearance grading of the virtual sawn timber after the rule-based method had been used to determine how the timber should be sawn.

The main focus of the studies presented in this thesis was regarding a specific wall panelling product of one customer, where the sawmill and customer have tried to calibrate a rule-based grading to functional but unimpressive results. There exists no metric today to describe this transaction other than this statement. The unimpressive results are the problems of coherency and conciseness of creating new rules, detailed in section 1.2.3. Using the machine learning methodology presented in this thesis allows for product-adapted appearance grading of sawn timber by the grade of the wall panelling product, which will be covered in section 3.1.2, while section 3.1.1 will be used to introduce PLS-based grading in a standardised use-case to establish context.

3.1.1 Standardised or Customised Rule-based Grading

Sorting of Dried Sawn Timber

An important consequence of the limitation to only consider knots in the studies presented in this thesis is that knots are the most difficult feature to base a quality assessment on. This is mainly due to the fact that the status of the knots are difficult to determine. For example, sawn timber with rot is easily detected and easily graded and therefore the grading accuracy of such pieces is high – regardless of the difficulty of assessing the knot structure of the sawn timber. In contrast, by limiting the grading process to only consider knots, the difficulty of the grading process increased. When the board scanner performed standardised appearance grading according to the NTGR and only considering knots, the grading accuracy was 81% according to a manual quality reference (Publication IV) – in contrast to the 95% found by Lycken (2017). Using PLS-based grading in the same test increased the grading accuracy marginally to 82%.

Studying appearance grading of dried sawn timber for an exporting customer using a customised rule-set based on the NTGR showed two major results (Publication I). Firstly, the grading accuracy of the customary rule-based approach calibrated for this task was 64%. In comparison, a multivariate PLS machine-learning method trained for the task managed 84%, compared to a manual reference grading. Secondly, in a verification test, the manual grader who performed the reference grading agreed with the grading of the rule-based system 81% of the time and with the PLS-based approach 95%

of the time. Thus, the verification test showed both approaches as having higher grading accuracy than initially found by the grading test, which indicates the difficulty of appearance grading. These results also showed that an appearance grading performed by a PLS-based system was difficult to distinguish from the manual quality reference defined to train the system on, even by the quality expert whom themselves performed the manual grading. Furthermore, the verification accuracy of the quality yield of the PLS-based approach was the same (95%) as the standardised appearance grading accuracy found by (Lycken 2017).

These results imply that a machine learning model, based on multivariate PLS, can outperform conventional rule-based grading when performing a customer-specific grading and that the performance is essentially equal when performing standardised quality grading. As described in section 1.2.3, the problems of coherency and conciseness of customising a set of rules make customising grading rules a very difficult task, even for highly skilled personnel (Lycken 2017). These difficulties make the grading accuracy lower when customising the grading rules, and the grading accuracy of a mature grading system at 95% (81% when only considering knots) can not be expected. As for the machine-learning approach, once the framework of a machine-learning approach has been implemented (automation of creating the variables, $\mathbf{X}$), the creation of a new customer-specific model only requires a new data set to be collected, or, better still, perform a manual grading $\mathbf{Y}$ on an already collected reference data set. The latter would be very easy to implement using images of the sawn timber as the basis for the quality assessment, as they can be easily distributed and re-used. The studies presented in Publications VI and VII showed the image grade to be equally as useful as the product grade for training the machine learning models for product-adapted grading, indicating that an image grade could be sufficient for other customers as well. Furthermore, the image grade was used to classify timber as sawn from butt logs or top logs (Publication V), which is also a type of appearance grade.

Grading of Virtual Sawn Timber by CT Scanning

The CT scanner graded the virtual sawn timber with a grading accuracy of 79% using rule-based grading, using manual grading of the sawn and dried timber as reference (Publication IV). In the same study, the board scanner used rule-based grading to achieve a grading accuracy of 81% – the same as the above-presented results of a rule-based board scanner. However, rule-based systems are sensitive to measurement errors since, for example, a single knot that is inaccurately measured too large could violate a rule and demote the sawn timber to a lower grade. Similarly, a rotational error during sawing could make a knot appear on another face of the sawn timber than intended and similarly violate a rule. These problems manifested themselves as an agreement between the two rule-based systems of 77% (Publication IV). When a PLS-based grading was used in both the CT scanner and the board scanner, a grading agreement of 82% was achieved. Furthermore, the board scanner using PLS-based grading agreed with manual grading to 82%. These results indicate that the soft modelling of features of PLS models, meaning that no single measurement determines the grade of the sawn timber, helped with the grading agreement between the two systems. This is due to the fact that PLS models linearly combines each variables and therefore the measurement of one variable can not solely determine the grade, which, for example, prevents measurement errors from altering the grade.

3.1.2 Product-adapted Grading

Sorting of Dried Sawn Timber

The studies presented in Publications II, III, V, and VI (Figure 2.1) were dedicated to studying the use of PLS-based machine-learning models in the board scanner at the dry-sorting station to perform product-adapted appearance grading. Four models were trained using the product grade, and one model was trained using the image grade. The models reached a grading accuracy of 74%–76%. More models were trained when investigating nuances of the training and testing conditions (Publication III, V, and VI), and the grading accuracy range expanded to 70%–77%, more on this in section 3.2. These grading accuracies should not be compared with the grading accuracies discussed in section 3.1.1, as the use-case is different, but should only be compared to the statement in the preamble of section 3.1 – unimpressive results. These studies resulted in the customer receiving a batch of 77%–82% grade A sawn timber (these number are not explicitly calculated in some of the Publications II, III, V, and VI). Worth remembering is that some grade B wall panels are always to be expected, as some knots can crack regardless of their status (section 2.3).

3.2 Implications of the Training Data

3.2.1 Measurement Disturbances

A key concept with machine-learning models is the generalisation of the model – the ability of a model to make predictions about things the model has never seen before (not been trained on). A good generalising PLS model is achieved by carefully selecting the training data to represent the intended use-case of the model, which in the studies presented in this thesis is an industrial sawmill. During the handling of sawn timber in a sawmill, things like dust and oil stains from conveyor chains can conceal or distort the appearance of the sawn timber in a way that is difficult to handle by camera-based systems. Such obstructions make knot detection more difficult and make the measurements of the board scanner less reliable. Effectively, the dust simulates excessive measurement errors. By training a PLS model on two data sets, one clean and one dusty, the grading accuracy of the model increased from 74% to 77% when grading clean sawn timber, and the grading accuracy increased from 70% to 73% when grading dusty sawn timber (Publication III). Essentially, the model became more robust to the adverse effect of the dust when this was accounted for in the training data – reinforcing the common idea that training on more data is better. It is unclear how a rule-based system would have been affected by this excessive amount of dust on the sawn timber. However, one can infer that a dust stain misinterpreted as a very large knot would negatively impact the grading accuracy, possibly to a large degree.

3.2.2 Class Balance and Class Overlap

Two important aspects of classification problems are the aspects of class balance and class overlap. Class balance indicates the proportions of each class in the classification problem. For example, a binary classification problem with 100 members of each class is a perfectly balanced classification problem. On the other hand, which is the case of, for example, cancer detection, the class of healthy people might have 98 members and the class of people with cancer only two. Training a machine learning classification model on such a data set would leave the model susceptible to a large bias, as predicting 100 out of 100 people as healthy would still leave the model with a "good" 98% classification accuracy, even though the model is completely useless.

In the context of product-adapted appearance grading of sawn timber, the data sets associated with Publications II, III, V, VI, and VII had a class balance of 62%–67% grade A pieces of sawn timber, which was considered a moderate class imbalance. However, the study results presented in Publication VI did not show a way to alleviate the class imbalance, nor did they show that class imbalance was a problem for the classification problem of the wall-panelling customer.

Class overlap indicates the difficulty of separating classes, meaning that according to some metric, an observation may appear to be part of several classes. For example, the classification problem of separating humans from gorillas by their height has a large class overlap (gorillas are approximately 1.6 m tall when standing upright), while separating the two by the span of their arms have a small class overlap (gorillas arms span approximately 2.4 m).

The data set associated with Publication VI was found to have 25% of the 251 pieces of sawn timber that the quality expert of the customer could not decisively assign an image grade. The quality grade of these 64 pieces was ambiguous to the degree that a model trained on the entire 251 pieces data set had the same grading accuracy as a model trained with these 64 pieces removed (Publication VI). Similar attempts were made in the same study to remove sawn timber with an ambiguous product grade, but the method used to remove pieces lowered the grading accuracy. The implication of removing parts of the training data yet retaining the same grading accuracy is that better data collection can be made. In the case of the image grade, any pieces of sawn timber inspected using the "plank-grader" app (Figure 2.5), where the grade can not be decisively determined, could be skipped to save time and effort as these does not contribute to the accuracy of the model.

The ambiguity of the sawn timber's image grade indicated that the true grade of the sawn timber was difficult to predict even for a quality expert, and the fact that the product grade does not yield stronger prediction models indicated that no additional information was to be gained by tracking the entire data set through the customer's production process (Publications VI and VII). Furthermore, more training data resulted in a higher grading accuracy compared to trying to compensate for the class balance or class overlap of the training data. Similar arguments was made about the effect of the dust discussed above.

These results indicate that it is possible that more training data would increase the grading accuracy of the models. It is also conceivable that the effects of class imbalance and class overlap were not a major concern in the studies presented in this thesis because of the fact that the data collections started as pre-sorted logs. The pre-sorted logs selected (Scots pine top logs) were considered suitable for this particular customer by the sawmill, which could have alleviated the effects of class imbalance and class overlap.

3.2.3 Variables

The creation of the variables used in the board scanner, detailed in Table 2.1, was based on previous work on the multivariate classification of sawn timber found in the literature (section 2.7.1). The scope of this set of variables was then reproduced for the CT scanner. Each variable used was selected to increase the available information in the variable set, partially based on the information the NTGR tries to convey. Furthermore, the sectioning of the sawn timber used (Figure 2.9) was implemented to include positional information of the knots into account. Admittedly, the decision to add more variables or not was difficult to argue either way. However, extensive testing outside of what is published in this thesis indicated that removing variables lowered the grading accuracy of the tested PLS models while adding more variables did not increase the grading accuracy. This would indicate that the number of variables in the finalised variable set detailed in Table 2.1 had reached a point of diminishing return such that no additional information could be provided with additional variables. Had a different approach to, for example, the sectioning of the sawn timber (Figure 2.9) been used, the diminishing return of the number of variables could have been experienced differently.

3.3 Sorting Outcome

Grading accuracy is only one of the metrics to consider when grading sawn timber. The sawmill wants to deliver as much of the sawn timber as possible to the intended customer, and the customer wants to process as much grade-A quality sawn timber as possible. Therefore, grading accuracy might not always be the best metric to judge a model's performance by, as grading accuracy only measures the accuracy of the grading system. PLS regression models use a classification threshold (section 2.7.2) to perform the classification. This threshold is by default selected to optimise grading accuracy, which in a normalised context would be 0.5, but can be set to any value (0–1 representing a continuous range from B–A). By setting the threshold to a larger value, making the model more strict (certain) about its classifications, a higher proportion of the sawn timber delivered from the sawmill would be of grade A, and vice versa (Publication II). This classification threshold gives very simple control over the sorting outcome, as only one setting needs to be controlled to change the balance between a high volume delivery and a high "purity" of grade A sawn timber in the delivery. However, changing the threshold (introducing a bias) inherently means that some pieces of sawn timber are misclassified, which would lower the grading accuracy on the total data set tested.

3.4 Product-adapted Sorting Before Sawing of the Log

CT scanning of logs gives the sawmill information about the internal knot structure of each log and thereby of the virtual sawn timber (Figure 2.7) prior to sawing. The virtual swan timber can be graded and, based on this grade, a decision about how to saw the log can be made. Effectively, grading the virtual sawn timber allows for green sorting of the sawn timber before the timber is sawn, as well as control over how each piece of sawn timber should be dried. For example, only allowing very likely grade A pieces of sawn timber to be dried according to the needs of the wall-panelling customer would greatly reduce the amount of sawn timber rejected at the dry-sorting station, which would save on drying costs. This section is based on the study presented in Publication VII and regards a data set consisting of 303 pieces of sawn timber.

No control over the sawing was given to any PLS model, instead, a rule-based control system decided how to saw the log, based on the above described process. The rule-based system used customer adapted rules to optimise the sawing of the log to maximise the yield of grade A sawn timber suitable for the wall-panelling customer. If no grade A virtual sawn timber was found in the log, it was sawn according to the needs of an unrelated customer. Once a decision had been made how to saw the log, the measurements of the virtual sawn timber was extracted and used by a PLS model to grade the virtual sawn timber. Furthermore, complementary rules was used to enforce upper limits to a few measurements in both the CT scanner and the board scanner.

The use of complementary rules was deemed necessary for this study, which measured, for example, abrupt pith deviations (CT scanner) and rotten knots (board scanner). The rules defined upper limits on six measurements in the CT scanner and 7 in the board scanner. The main reasoning behind this inclusion of rules was that these measurements were rare and ambiguous, meaning that very few data points showed whether or not these measurements correlated positively or negatively with the product grade. This made these measurements influence over the predicted grade weak, even though the few existing data points showed a high risk associated with these measurements.

Using the rule-based system in the CT scanner, and removing 28 pieces of sawn timber by complementary rules in the CT scanner and 28 pieces (12%) in the board scanner, resulted in a delivered batch of 200 pieces of sawn timber with 77% grade A pieces. By comparison, section 3.1.2 shows how a PLS-based dry-sorting without CT-based pre-sorting resulted in 77%–82% grade A sawn timber in the delivered batch. This means that the overall effectiveness of rule-based sorting is significantly lower than PLS-based grading in both systems.

Using a strict image-trained PLS model in the CT scanner, and removing 21 pieces of sawn timber by complementary rules in the CT scanner and 9 pieces (7%) in the board scanner, resulted in a delivered batch of 115 pieces of sawn timber with 90% grade A pieces. Additionally, if the dry-sorting station was skipped entirely, the delivered batch would have consisted of 85% grade A sawn timber. These results show that the image grade was useful to train a PLS model for grading of virtual sawn timber and how controlling a single setting (the classification threshold) changed the entire production strategy. With this approach, only virtual sawn timber with a very high probability of resulting in grade A wall panels can be dedicated to this customer, sawn and dried accordingly, and delivered with a high product yield for the customer. The number of delivered pieces of sawn timber is lower for the PLS-based approach, however, this can easily be tuned by changing the classification threshold, at the cost of lowering the "purity" of grade A pieces in the delivered batch.

3.5 Limitations of PLS and Machine-learning Methods

The use of variables as presented in this thesis is not in line with the software 2.0 methodology described in the introduction (section 1.3). Humans have an incredibly poor "feel" for what variables are important for a particular problem and to what degree – especially in the context of the specific machine-learning model used. Therefore, the creation of a variable set resembles a trial and error process. This trial and error approach, even with the careful analysis of the variable set using PCA and PLS models, makes it difficult to judge if the perceived maximum limit of the grading accuracy is due to the set of variables used, due to the PLS machine-learning method used, or due to a natural limit of this specific classification problem. In particular, since PLS is a multivariate linear regression method, it is difficult to tell how the linearity of the model affects the grading accuracy – if at all. For example, the number of knots is not perfectly linearly correlated with the grade of the sawn timber – too few knots leaves the appearance wanting for more, and too many knots makes the appearance appear messy. This means that there is, in some sense, an optimal number of knots, which can not be accurately described by a linear model. Figure 2.10 shows how a linear model fits moderately non-linear data.

Creating a set of variables is akin to defining, for example, the architecture of a neural net. The problem of knowing if the currently defined architecture is holding the performance of the neural net back or not is the same problem as defining the variable set is to a PLS model. Similar arguments can be made about any machine-learning method. In essence, machine-learning methods can appear to be a black-box plug-and-play solution to almost any problem, which it rarely is.

Chapter 4

Conclusions

I usually read this first

The objective of this thesis was to introduce machine learning to the process of the appearance grading of sawn timber in a way that is easy to customise and easy to fine-tune. To reach this objective, a machine-learning model needed to be implemented and trained to mimic the appearance grading performed by an experienced manual grader but faster and more reliably. Such an approach to appearance grading would allow for product-adapted grading and sorting of sawn timber, resulting in higher quality yield for the customer and lower refinement costs for the sawmill. This thesis only considered knots and detailed the implementation of PLS models in a CT scanner and a board scanner, and the implications of the training data on the grading performance and sorting outcome. Furthermore, to train a product-adapted machine-learning model, the use of the resulting product grade of the sawn timber or the image grade of the sawn timber was investigated.

To establish a baseline for comparison, appearance grading of dried sawn timber by standardised quality grades using a PLS model was studied. A PLS model trained on a manual quality reference achieved an 82% grading accuracy compared to 81% of a conventional rule-based approach. The manual grader agreed to 95% with the PLS model and 81% of the rule-based grading in a verification test. As this grading scenario was common practice for the sawmill, the set of rules used was considered fine-tuned and that the difficulties of coherently and concisely defining the rules were solved for this scenario. The PLS-based approach was deemed easy to customise for this grading scenario, as it could match the performance of the fine-tuned rule-based approach without implementation or calibration difficulties. The verification agreement of 95% shows that the PLS model graded the sawn timber according to how it was trained, as the manual grader could not significantly distinguish between the grade assigned by the PLS model from the quality grade they had defined for the training data.

When performing product-adapted grading of dried sawn timber, the set of rules used could not be coherently and concisely defined such that the grading outcome was impressive to the wall-panelling customer. Training a PLS model on either the product grade or the image grade resulted in a grading accuracy of $70\%--77\%$ in several testing scenarios. The scenarios investigated the aspects of class balance, class overlap, and measurement disturbances of the training and testing conditions for PLS models and found that training on all available data always yielded the highest grading accuracy. The grading accuracy resulted in the customer receiving 77%–82% grade A sawn timber batches. As the image grade was as useful for training a PLS model as the product grade, the logistical process of delivering unsorted sawn timber to the customer was unnecessary. Instead of performing a manual data collection, if a traceability method was implemented to collect the data from the board scanner automatically, the entire data collection needed for a new customer adaptation could be done in software, which would drastically simplify the data collection process, which is in line with the industry 4.0 concept.

Using a PLS model in the CT scanner for product-adapted grading of virtual sawn timber allowed easy control over the sorting outcome. For example, a PLS model trained using the image grade could be controlled such that the sorting outcome resulted in a delivered batch of 90% grade A pieces of sawn timber, following a dry sorting that only rejected 7% of the sawn timber. If the dry-sorting station had been skipped entirely, the delivered batch would have consisted of 85% grade A pieces of sawn timber. Furthermore, using a rule-based approach in the CT scanner, the dry-sorting station rejected 28 pieces (12%) and delivered a batch of 77% grade A pieces of sawn timber. These results show that PLS-based product-adapted grading is more efficient at separating grade A sawn timber from grade B than a rule-based grading in both the CT scanner and the board scanner. Furthermore, as a delivered batch of 85% grade A pieces of sawn timber is purer than the grading-accuracy optimised PLS-based dry-sorting of 77%–82% grade A pieces, it seems conceivable to perform CT-based sorting only.

Once a framework for a machine-learning method such as PLS has been implemented, this thesis shows the ease of customising and fine-tuning a PLS model. A PLS model was shown to be easy to customise for both a customer of appearance sorted sawn timber and a specific product. The use of a classification threshold makes the grading outcome easy to control intuitively, which allowed for new grading strategies in a way that rule-based grading does not support. Furthermore, the ability of the PLS models to retain their grading accuracy for different training and testing scenarios, as well as the use of an image grade as a quality reference for training, showed that the data collection necessary to implement machine-learning methods to be straightforward and not require an expert level understanding of machine learning to perform.

Chapter 5

Future Work

Someone should develop an app for this

Based on the findings of the present thesis, one can infer a few things about the future and potential opportunities for future studies. Although machine learning studies are extremely popular these days, this alone should not be the reason to publish articles related to machine learning. Careful consideration should be given during the design phase of new studies to the suitability of machine learning. Below follows examples of some areas of research potentially suitable to further machine-learning studies.

5.1 Qualitative Measurement of Grading Accuracy

The grading accuracy detailed in the present thesis was based on the product quality outcome of the sawn timber. However, as there is always some amount of grade B wall panels to be expected due to, for example, the random cracking of knots, grading accuracy in and of itself might not be a fair performance metric. Instead, the grading accuracy could be measured by a qualitative measurement of the quality of the sawn timber – essentially determine if each piece of sawn timber has a high chance of resulting in quality A wall panels.

5.2 CT Scanner Sorting Only

For some grading operations, the dry-sorting process of the sawn timber might not be necessary. With the intended wall-panelling customer in mind, the board scanner only removed 7% of the sawn timber following the sorting performed by the PLS-based CT scanner. Furthermore, the 7% of the sawn timber removed was largely due to visually very distinct features that the CT scanner could not easily detect, for example, rotten knots – something that is not suitable for wall panels. Such features are quite easy to detect (Figure 2.3) by the manual graders at the customer's quality control station. In practice, this could be used to skip the dry-sorting station entirely and deliver a batch of sawn timber directly after drying, which would save on cost for the sawmill, and the customer could get a cheaper price. The possibility of using only one grading operation in the CT scanner might be more applicable to other use cases.

5.3 Simulated Product Manufacturing in CT Images

CT scanning gives a high-resolution 3D image of the internals of each scanned log, which in the present thesis was used to create virtual sawn timber. Remembering that non of the sawn timber's outer faces were used as the wall panels' visible faces, therefore, it would be interesting to study simulated product manufacturing in the CT images by splitting, profiling, and planing the virtual sawn timber. This way, it would be possible to get information about the visible face of the wall panels and simulate how, for example, profiling of the virtual sawn timber would interact with knots.

5.4 Machine Learning-based Sawing Optimisation

As demonstrated by, for example, Ursella (2021), the accuracy of detecting knots and other features in CT images is getting higher and higher as the technology matures. With the accurate detection of features comes the ability to determine the optimal sawing conditions with high granularity. However, with the large number of features detected in CT images and with the large number of grades to choose from at the CT-scanning stage (a log can basically be sawn into any product), the task of determining the optimal sawing conditions is getting convoluted. This problem could be a candidate for a machine-learning solution, where a machine-learning model could determine the optimal sawing based on 3D CT images. Unfortunately, in the study presented in Publication VII, the PLS models were not in control of how the log was sawn, the implications of which should be the subject of future studies.

5.5 Traceability

Each log and the resulting sawn timber are not tracked throughout the refinement process in most sawmills. Assuming that it was, one would only have to perform one data collection from a particular sawmill to create a reference data set with available descriptive measurements, $\mathbf{X}$, from all available scanning systems. Any number of outputs or grades, $\mathbf{Y}$, could be appended to the reference data set later. At this stage, a large number of possible machine-learning studies would be possible.

Achieving 100% accurate traceability is close to reality, as 92% of the material investigated by Olofsson et al. (2019) was theoretically tracked from a CT-scanning system to a camera-based board scanning system, using the traceability algorithm developed by Möller (2019). Furthermore, Flodin et al. (2008) achieved above 95% when matching the centre yield sawn from CT-scanned logs with a camera-based green sorting system; Pahlberg (2017) achieved above 99% traceability of sawn products traced between camera-based systems, and; Ursella (2021) achieved 99.6% tracking accuracy between a CT Log scanner and another X-ray based scanning system (a Microtec Logeye). To be clear, traceability methods could bring many benefits even if they are not 100% accurate, especially since errors in matching are often due to pieces of sawn timber being so similar.

The use of traceability methods could enable co-operation between, for example, a CT scanner and a board scanner. The board scanner could complement the grading of the CT scanner with additional camera-based information, in contrast with today's setup where the systems might hinder each other (Publication IV). The addition of camera-based information regarding knots could be particularly useful as knots, possibly filled with resin, and bark, as they are difficult to distinguish from each other by their density profiles measured by the CT scanner. This difficulty could make the board scanner able to complement the information from the CT scanner with information about knots and bark and thereby increase the grading accuracy.

5.6 Product-adapted Appearance Grading à la Carte

A few puzzle pieces are needed for product-adapted grading à la carte. Firstly, a traceability method needs to be used to automatically collect a large database of sawn timber, consisting of many samples for each wood species and for each sawn timber dimension that a sawmill produces, depending on the use case. Each sample of sawn timber should be attributed with scanner data from each scanning system available at the sawmill. Secondly, images of the sawn timber need to be automatically collected by, for example, a board scanner, similar to those used in the present thesis. Thirdly, a website running a similar grading process as the one used in this thesis (Figure 2.5). Together, these three puzzle pieces would create an interface for a customer to select the species, dimensions, and appearance of sawn timber they would like to purchase. In the background, a machine-learning method could then simulate the sorting outcome and calculate a price for that sorting, which the customer can verify. Such an implementation could enable customers to finalise a customer or product-specific sorting grade by themselves before even contacting the sawmill.

References

Anon. (2018). Boardmaster. https://finscan.fi/products/boardmaster/?lang=en. (2018-06-12). Note: older predecessor to BoardmasterNOVA.

Batrakhanov, D. (2021). Virtual sawing using generative adversarial networks. Master's Thesis. Lappeenranta-Lahti University of Technology, Finland.

Berglund, A., Broman, O., Oja, J. and Grönlund, A. (2015). Customer adapted grading of Scots pine sawn timber using a multivariate method. Scandinavian Journal of Forest Research 30(1), 87–97.

Breinig, L., Broman, O., Brüchert, F. and Becker, G. (2015a). Optimization potential for perception-oriented appearance classification by simulated sawing of computed tomography-scanned logs of Norway spruce. Wood Material Science & Engineering 10(4), 319–334.

Breinig, L., Leonhart, R., Broman, O., Manuel, A., Brüchert, F. and Becker, G. (2015b). Classification of wood surfaces according to visual appearance by multivariate analysis of wood feature data. Journal of Wood Science 61(2), 89–112.

Broman, O. (2000). Means to measure the aesthetic properties of wood. PhD Thesis. Luleå University of Technology, Skellefteå, Sweden.

CEN (1999). European Committee for Standardization.

Eriksson, L., Byrne, T., Johansson, E., Trygg, J. and Vikström, C. (2013). Multi-and megavariate data analysis basic principles and applications. Volume 1. Umetrics Academy.

Flodin, J., Oja, J. and Grönlund, A. (2008). Fingerprint traceability of sawn products using industrial measurement systems for x-ray log scanning and sawn timber surface scanning. Forest Products Journal 58(11), 100–105.

Fredriksson, M. (2014). Log sawing position optimization using computed tomography scanning. Wood Material Science & Engineering 9(2), 110–119.

Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. Analytica Chemica Acta 185, 1–17.

Giudiceandrea, F., Ursella, E. and Vicario, E. (2011). A high speed CT scanner for the sawmill industry. Proceedings of the 17th international non destructive testing and evaluation of wood symposium. University of West Hungary Sopron, Hungary, 14–16.

Grönlund, U. (1995). Quality improvements in forest products industry: classification of biological materials with inherent variations. PhD Thesis. Luleå University of Technology, Skellefteå, Sweden.

Grundberg, S., Grönlund, A. and Grönlund, U. (1995). The Swedish stem bank: a database for different silvicultural and wood properties. Luleå University of Technology.

Gu, I. Y.-H., Andersson, H. and Vicen, R. (2010). Wood defect classification based on image analysis and support vector machines. Wood Science and Technology 44(4), 693–704.

Hagman, O. (1996). On reflections of wood: wood quality features modelled by means of multivariate image projections to latent structures in multispectral images. PhD Thesis. Luleå University of Technology, Skellefteå, Sweden.

Hu, J., Song, W., Zhang, W., Zhao, Y. and Yilmaz, A. (2019). Deep learning for use in lumber classification tasks. Wood Science and Technology 53(2), 505–517.

Johansson, E., Johansson, D., Skog, J. and Fredriksson, M. (2013). Automated knot detection for high speed computed tomography on *Pinus sylvestris* L. and *Picea abies* (L.) Karst. using ellipse fitting in concentric surfaces. Computers and Electronics in Agriculture 96, 238–245.

Karpathy, A. (2017). Software 2.0. https://karpathy.medium.com/software-2-0-a64152b37c35. Accessed: 2021-06-24.

Lycken, A. (2006). Comparison between automatic and manual quality grading of sawn softwood. Forest Products Journal 56(4), 13–18.

Lycken, A. (2017). Rapport efter studie av automatisk sortering i justerverk samt av regelverken Nordiskt trä och EN 1611-1 [Report after a study of automatic dry-sorting using the Nordic timber grading rules and EN 1611-1]. In Swedish with English Abstract.

Lycken, A. and Oja, J. (2006). A multivariate approach to automatic grading of *Pinus sylvestirs* sawn timber. Scandinavian Journal of Forest Research 21(2), 167–174.

Möller, C.-J. (2019). Development of Fingerprint Traceability in a Modern Sawmill. Master's Thesis. Umeå University, Sweden.

Nordmark, U. and Oja, J. (2004). Prediction of board values in *Pinus sylvestris* sawlogs using X-ray scanning and optical three-dimensional scanning of stems. Scandinavian journal of forest research 19(5), 473–480.

Oja, J. (2000). Evaluation of knot parameters measured automatically in CT-images of Norway spruce (*Picea abies* (L.) Karst.) Holz als Roh-und Werkstoff 58(5), 375–379.

Oja, J., Grundberg, S., Fredriksson, J. and Berg, P. (2004). Automatic grading of sawlogs: a comparison between X-ray scanning, optical three-dimensional scanning and combinations of both methods. Scandinavian Journal of Forest Research 19(1), 89–95.

Olofsson, L., Möller, C.-J., Wendel, C., Oja, J. and Broman, O. (2019). New possibilities with CT scanning in the forest value chain. International Nondestructive Testing and Evaluation of Wood Symposium.

Olofsson, L., Broman, O., Fredriksson, M., Skog, J. and Sandberg, D. (2017). Customer adapted grading of Scots pine sawn timber – a multivariate method approach. Proceedings of International Wood Machining Seminar 23, 360–371.

Pahlberg, T. (2017). Wood fingerprint recognition and detection of thin cracks. PhD Thesis. Luleå University of Technology, Skellefteå, Sweden.

Rudakov, N., Eerola, T., Lensu, L., Kälviäinen, H. and Haario, H. (2018). Detection of mechanical damages in sawn timber using convolutional neural networks. German Conference on Pattern Recognition. Springer, 115–126.

Swedish Sawmill Managers Association (1982). Sortering av sågat virke av furu och gran. Stockholm: Föreningen Svenska Sågverksmän.

Swedish Sawmill Managers Association (1994). Nordic timber: grading rules for pine (*Pinus sylvestris*) and spruce (*Picea abies*) sawn timber: commercial grading based on evaluation of the four sides of sawn timber. Markaryd: Föreningen Svenska Sågverksmän.

The MathWorks, Inc. (2021a). MATLAB®. Version R2021a. Natick, Massachusetts.

The MathWorks, Inc. (2021b). Analyze Quality of Life in U.S. Cities Using PCA. https://se.mathworks.com/help/stats/quality-of-life-in-u-s-cities.html (visited on 2021-09-05).

The MathWorks, Inc. (2021c). Simulating Dependent Random Variables Using Copulas. https://se.mathworks.com/help/stats/simulating-dependent-random-variables-using-copulas.html (visited on 2021-09-05).

Ursella, E. (2021). In-line industrial computed tomography applications and developments. PhD Thesis. Ca' Foscari University of Venice, Italy.

Ursella, E., Giudiceandrea, F. and Boschetti, M. (2018). A Fast and Continuous CT scanner for the optimization of logs in a sawmill. 8th Conference on Industrial Computed Tomography (iCT 2018) at Wels, Austria. Volume 2.

Wold, S., Albano, C., Dunn, W., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. (1984). Multivariate data analysis in chemistry. Chemometrics. Springer, 17–95.

Wold, S., Sjöström, M. and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems 58(2), 109–130.

Zolotarev, F., Eerola, T., Lensu, L., Kälviäinen, H., Haario, H., Heikkinen, J. and Kauppi, T. (2019). Timber tracing with multimodal encoder-decoder networks. International Conference on Computer Analysis of Images and Patterns. Springer, 342–353.

PUBLICATION I

Customer Adapted Grading of Scots pine Sawn Timber – a Multivariate Method Approach

Authors and year:
Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D.
2017

Publication originally published in:
Zbiec, M. & Orlowski, K. (Eds.) 23rd International Wood Machining Seminar Proceedings: Warsaw, Poland, Warsaw University of Life Sciences, 360–361
URN: urn:nbn:se:ltu:diva-65840

In: 23rd International Wood Machining Seminar (IWMS-23), Eds. Zbiec M & Orlowski K, Warsaw, Poland, May 28-31, pp. 360-371

Customer adapted grading of Scots pine sawn timber - a multivariate method approach

Linus Olofsson[a], Olof Broman[a], Magnus Fredriksson[a], Johan Skog[b], and Dick Sandberg[a]

[a]Wood Science and Engineering, Luleå Univ. of Technology, SE-931 87 Skellefteå, Sweden

[b]RISE Bioeconomy, Research Institutes of Sweden, SE-931 77 Skellefteå, Sweden

ABSTRACT

At Scandinavian softwood sawmills, the most common system for grading of sawn timber in dry conditions is optical scanning equipment together with a rule-based automatic grading system (RBAG). The procedure to define new grading rules towards a customer with specified requirements is a time-consuming work for sawmills and is rarely implemented in a satisfactory way neither for the customer nor for the sawmill. An important consequence is that sawmills will, in general, not be able to deliver products that utilize the full potential of the quality distribution of the sawn timber produced at the sawmill. Their customers will get products with mismatch in desired and delivered quality grades. Thus, there is a need for a methodology that facilitates time and cost effective grading toward specific customers' needs. The objective of the study was to further develop and validate a method that complements the RBAG by a holistic-subjective automatic grading (HSAG) approach - using multivariate regression models.

In the study, 790 Scots pine boards with cross-section dimensions of 38×150 mm and length between 3.4 m and 5.6 m were manually graded according to the preferences of a large-volume customer, and also scanned and graded by an RBAG system calibrated for the same customer. Multivariate models for prediction of board grade, based on aggregated knot variables obtained from the scanning, were calibrated using partial least squares regression.

The results show that prediction of board grades by the multivariate models were more correct than the grading by the RBAG system. The prediction of board grades based on multivariate models resulted in 84% of the boards graded correctly, according to the manual grading, while the corresponding number was 64% for the RBAG system. In a follow up grading test the accuracy of the two systems were 95% and 81%, respectively.

Keywords: automatic quality grading, multivariate model, rule-based model, sawmill, customer expert

In: 23rd International Wood Machining Seminar (IWMS-23), Eds. Zbiec M & Orlowski K, Warsaw, Poland, May 28-31, pp. 360-371

INTRODUCTION

The main problem in the manual grading of sawn timber is the high number of wood characteristics that the grader has to judge during the few seconds the grader has to spend on each piece of sawn timber. The grader needs to consider all visual defects on all sides of the piece and relate that information to the grading rules in use, often also related to a specific customer, and furthermore how any trimming options could be used to optimize the board value. Grönlund (1995) showed that only about 57% of the sawn timber is graded equally by separate graders using the same grading rules. On the other hand, this subjective grading method is sometimes desirable by customers as its rules are an attempt to describe the product that is demanded by the customer.

Overtaking this traditional manual grading method, since the last three decades, is the more efficient and consistent rule-based automatic grading (RBAG) systems that Lycken (2006) showed could replace manual graders for grading according to the Nordic timber grading rules (Swedish Sawmill Managers Association 1994). These RBAG systems scan individual boards from four directions and assign a grade using cameras, image processing and analysis, and per-customer customizable grading rules. However, as it would be logistically impossible to have an infinite number of grading rules, the Nordic timber grading rules were formulated in the Nordic countries, although most sawmills use their own set of rules adapted to different raw materials, board dimensions, markets, customers, and of course the fact of no sawmill being the other alike.

The problem with these RBAG systems is, actually, the increasing importance of customization of grading rules (European Confederation of Woodworking Industries 2004) in order to satisfy the customers. RBAG systems have a lot of settings describing each grade and customizing rules is a time-consuming process which often requires the customer's presence at the sawmill and hence quickly becomes complicated. This leads to RBAG rules being seldom changed (Lycken 2006; Lycken & Oja 2006). Another problem is that a board with good quality overall, but with a few larger than specified defects, could be tolerated by a customer even if the RBAG system deem it of a lesser grade and vice versa.

Berglund et al. (2015) combined the advantages of the subjective manual grading and RBAG and introduced a so-called holistic-subjective automatic grading (HSAG) system, which leads to the purpose of this study; a validation of the HSAG principle.

There has been several earlier attempts to industrialize an HSAG system, using neural networks, self-organizing maps, and fuzzy logics, to automatically detect and classify board defects, as well as to grade boards according to appearance (Labeda 1995; Kauppinen 1999; Kline et al. 2003; Niskanen 2003; Silvén et al. 2003; Breinig et al. 2015). One obstacle in reaching large scale applications for these systems is the complicated procedure of changing and formulating new grading rules as this require numerous strong examples of each new grade as well as for different defect types (Kline et al. 2003).

A different approach to HSAG, earlier introduced by Lycken & Oja (2006), is to use multivariate models. Berglund et al. (2015) used a partial least square (PLS) multivariate model which, just as neural networks, has to be calibrated using a training set of manually graded boards. A PLS based system is, however, easier to understand, validate and customize in comparison to a neural network (Esbensen 2002) while also being robust to noise and can, despite measurement errors, capture important systematics effects (Eriksson et al. 2006). A measurement error in one or more defects does not have a large effect on a PLS model, which is in contrast to an RBAG system which could decide the board grade based on a single measurement error.

Lycken & Oja (2006) showed that their PLS model was consistent with around 85% of manual gradings that used the Nordic timber grading rules. This is considered fairly good in comparison with manual grading. Lycken & Oja (2006) also concluded that if accurate PLS models are obtained it is easier to customize grading rules to satisfy customers' needs.

The study by Berglund et al. (2015), which is a continuation of Lycken & Oja (2006), found that a PLS-based HSAG system could outperform the traditional RBAG system in terms of board grading accuracy. The aim of this study was to validate the findings by Berglund et al. (2015) and to investigate a PLS-based HSAG system as a complement for an RBAG system.

MATERIALS AND METHODS

As this is a validation study, this study was made to mimic the study by Berglund et al. (2015) in order to make it possible to compare the results. The only difference between the two studies was that this one uses a new set of boards, higher in numbers.

Grading rules

The focus of this study was the appearance grading of boards according to grading rules based on the Nordic timber grading rules (Swedish Sawmill Managers Association 1994). These rules describe four different board grades, namely A, B, C, and D, where grade A is the best, and most valuable, and grade D the worst, and least valuable. Even though the Nordic timber grading rules consider most wood features on a board's surfaces, the most important rules are regarding knots, as the size and distribution of knots are the most descriptive of the general appearance of a board. This is the reason why this study only considered knots for the development of HSAG models to predict board grade. A consequence of this choice, was that grade D was not considered, as grade D is given to boards with one or several extreme wood features and is not necessarily related to knot characteristics.

Materials

790 randomly selected Scots pine (*Pinus sylvestris* L.) boards from a sawmill in northern Sweden were used. The board dimensions were 38×150 mm and originated from the two outer center boards in the sawing pattern $(38/50/50/38) \times 150$ mm. The boards varied between 3.4 m and 5.6 m in length.

Methods

The boards were manually graded into grades A, B, and C according to knots without trimming, by an expert in customer preferences for the North African market and this expert was the same person as in the study by Berglund et al. (2015). This was the reference grades for the boards.

All sides of the boards were scanned using one of the common RBAG systems in the market, a Finscan Boardmaster (Finscan 2017). As in the manual grading, only knot parameters was considered, and no trimming was allowed. The rules were, however, customized to the current sawmill's experience of customer preferences of the North African market.

Prediction models

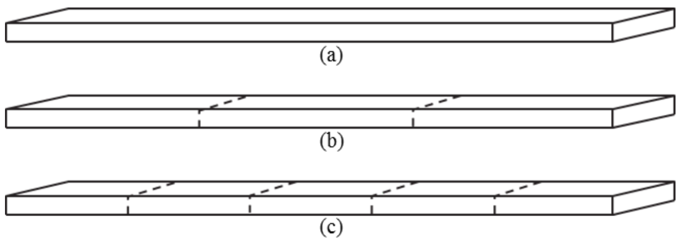
During the same scanning process as in the previous section a total of 58 variables regarding knots are extracted, using knot descriptive data from the RBAG system (Table 1). As knot characteristics in specific sections of a board can be more or less important to customers, each board was virtually divided into one, three, and five equally large sections (Figure 1), and the variables were extracted for each section. Furthermore, the variables were extracted from the pith and sapwood surfaces separately, but for the edge surfaces together. This resulted in $58 \times (1 + 3 + 5) \times 3 = 1566$ variables in total for each board. The variables were used to train multivariate prediction models for the three grades A, B, and C according to customer preferences of the North African market.

Table 1: The 58 variables (marked as x) related to knots that were extracted for each of the board surfaces.

Variable		Knot type		
		Sound and dead	Sound	Dead
Total no. of knots		x	x	x
Average knot size (mm)		x	x	x
St. dev. knot size (mm)		x	x	x
Maximum knot size (mm)			x	x
Ratio $\frac{\text{Knot area (mm}^2\text{)}}{\text{Surface area (mm}^2\text{)}}$ (%)		x		
Ratio sound knots (%)			x	
Ratio dead knots (%)				x
No. of sound knots per m (no./m)			x	
No. of dead knots per m (no./m)				x
Ratio of knots	≤9 mm	x	x	x
"	10–19 mm (%)	x	x	x
"	20–29 mm (%)	x	x	x
"	30–39 mm (%)	x	x	x
"	40–60 mm (%)	x	x	x
"	60–80 mm (%)	x	x	x
"	≥80 mm (%)	x	x	x
No. of knots	≤9 mm (no./m)	x	x	x
"	10–19 mm (no./m)	x	x	x
"	20–29 mm (no./m)	x	x	x
"	30–39 mm (no./m)	x	x	x
"	40–60 mm (no./m)	x	x	x
"	60–80 mm (no./m)	x	x	x
"	≥80 mm (no./m)	x	x	x

Partial Least Square (PLS) regression was used to correlate the predictor variables (the 1566 variables described above) to each other. It is easy to control and change settings in a PLS based HSAG system, thereby affect the quality and value yield. Such an HSAG system can be complemented by an RBAG system with grading rules for specific defects, such as wane, pitch pockets, or blue stain as desired.

Figure 1. Variables were extracted for (a) one, (b) three, and (c) five equally large sections along the board.



From the RBAG system, we extracted the 1566 predictor variables (X-variables) while dummy variables for each board grade were used as response variables (Y-variables). Two separate PLS regression models were then trained for board classification. The first, PLS model I, was used to separate boards of grade C from boards of grade A or B while the second, PLS model II, was used to separate boards of grade A from boards of grades B or C. This means that models I and

II were both trained on two classes, Class I and Class II, which are defined differently for each model. These classes for each PLS model is shown in Table 2.

Using two models is sufficient to distinguish between three classes, and since it was showed in preliminary tests to be more difficult to separate boards of grade B from boards of grade A or C such a model was not used.

Table 2. Predefined classes for PLS models I and II. A, B, and C are grades for sawn timber customized according to preferences of a customer in North Africa.

	Class I	Class II
PLS model I	C	A, B
PLS model II	A	B, C

Using the two PLS models to predict the grade of boards gives a probability value for each board to be of either grade A or C. A high probability value, i.e. $\gtrsim 1$, indicates a high probability of the board fitting that grade and vice versa. The natural delimiter for such a probability value would be to choose a limit of 0.5, but it's advantageous to choose a limit in dialogue with the customer. For example looking at PLS model I, choosing a high limit value would separate only the most distinct grade C boards from the rest, with a lot of grade C boards being classified as of grade A or B – false positives. On the other hand, choosing a low limit value would classify a lot of boards of grade A or B as boards of grade C – false negatives. For this reason, limit values for the two models should be chosen to result in acceptable numbers of false negatives, and false positives.

The two models were used to predict board grade according to the following decision tree, which is illustrated in Figure 2.

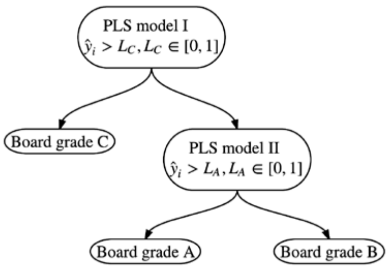
(1) Does the PLS model I estimate for Class I $\hat{y}_i$, where $i = 1, \dots, 790$, exceed a specified threshold limit, i.e. $\hat{y}_i > L_C,\ L_C \in [0,1]$, then the board is grade C, otherwise continue.

(2) Does the PLS model II estimate for Class I $\hat{y}_i$, exceed a specified threshold limit, i.e. $\hat{y}_i > L_A, L_A \in [0, 1]$, then the board is grade A, otherwise continue.

(3) If the board is neither grade A nor grade C, then it is grade B

Figure 2. Decision tree using two PLS models for automatic grading of boards into grades A, B, and C.

From both the sawmill and customer's perspective, grading a board of true grade C as a board



of grade A or B is worse than grading a board of true grade A as a board of grade B or C. This is the reason why prediction of grade C is performed first, as boards that are difficult to assess are more likely to be assigned grade C than grade A or B in this way. This is also controlled by adjusting the limits L_C and L_A for each model.

For any PLS regression model, the goodness of fit is given by the calculated coefficient of determination (R^2) and the goodness of prediction by the Q^2-value. The Q^2-value is based on cross-validation (Martens & Naes 1989). Cross-validation means that n-models are created, and each excluding $1/n$ of the observations when creating a training set to build the model on. Each model can then be tested on the observations that were excluded when building the model. These excluded observations are called the test set. The value of Q^2 represents the proportion of variance in y-values in the test sets that is explained by the model. This means that Q^2 is a measure of the model's ability to predict new observations, which are observations that were not included when building the model.

To analyze and verify the grading performance of the two PLS models, all 790 boards were graded according to the procedure described in Figure 2, using the limit values $L_A = 0.43$, and $L_C = 0.5$ – aiming at the same grade distribution found by the customer expert prior to the model establishment. These limits are used both for the study and follow-up grading test.

Follow-up grading

To simulate a customer receiving a package of boards and inspecting his or her purchase by validating the board grades, the same expert as before was instructed to do a follow-up grading test on six different groups of boards. These groups represent the three different grades A, B and C assigned when using the two different grading strategies RBAG and HSAG. To ensure the expert is not biased, the follow-up grading was performed without the expert knowing which grading strategy had been used for each group.

Out of the original 790 boards, 395 were available for the follow-up test and boards were selected to the six groups by the following procedure:

1. randomly select one board graded by RBAG as grade A and assigns it to group 1,
2. randomly select one board graded by HSAG as grade A and assigns it to group 2,
3. randomly select one board graded by RBAG as grade B and assigns it to group 3,
4. randomly select one board graded by HSAG as grade B and assigns it to group 4,
5. randomly select one board graded by RBAG as grade C and assigns it to group 5, and
6. randomly select one board graded by HSAG as grade C and assigns it to group 6.

As long as boards remained of each grade from each grading strategy, steps 1-6 were executed and repeated. Once there were no more boards of a specific grade from any of the two grading strategies the sorting procedure skipped steps including that grade while the remaining steps were executed before continuing. This was to ensure each grading strategy had equally many representatives for each board grade in the test.

The first group that was completely sorted according to the above-described procedure was the RBAG grade A group, which resulted in 16 boards in groups 1 and 2. Then procedure step 3-6 was repeated to fill up the groups 3-6 with 35 boards in each. The number of boards in each group is shown in Table 3.

Table 3. The amount of randomly selected boards for each grade and sorting system used in the follow-up grading test.

Group	Grading	Grade	No. of boards
1	RBAG	A	16
2	HSAG	A	16
3	RBAG	B	35
4	HSAG	B	35
5	RBAG	C	35
6	HSAG	C	35
7	HSAG	A	15*
Total			187

*All remaining boards in the test set that was assigned grade A by HSAG were added to an additional group 7.

As the RBAG resulted in few boards of grade A, groups 1 and 2 has fewer members, but as grade A is in general of great importance for sawmills, all remaining boards graded as grade A by the HSAG was added to an additional group 7.

The expert visually inspected all boards. To increase objectivity, five boards from each group were inspected in sequence before considering the next five boards from the next group, and so on until all boards had been inspected.

The expert's opinion on each board was documented by the opinions below, with one mandatory (1-3) and one optional complementary (4-5) opinion:

1. The board is graded correctly.
2. The board is of a lower grade.
3. The board is of a higher grade.
4. The board grade is a borderline case of being of a lower grade.
5. The board grade is a borderline case of being of a higher grade.

RESULTS AND DISCUSSION

Out of the 790 boards in this study, the expert graded 17% as of grade A, 51% as of grade B, and 32% as of grade C, Tables 4 and 5. This reference grade distribution is compared to the corresponding grade distribution in Berglund et al. (2015) of 24% as of grade A, 44% as of grade B, and 31% as of grade C. These different reference grade distributions will most likely not affect the comparison between the two studies.

When evaluating the customer satisfaction, a good measure is the share of boards that is graded correctly or where the grade has been underestimated. For a customer, a correct grade is the expected while an underestimated grade would mean that the customer purchases boards of a perceived higher grade to a lower price.

Rule-based automatic grading (RBAG)

The RBAG resulted in the grading distribution shown in Table 4, where the RBAG is compared to the expert's grading of the 790 boards. Out of these 790 boards, 64% were graded equally by the expert and the RBAG system, which is very consistent with the corresponding 63% in the study by Berglund et al. (2015). This is a good indication of the consistency of RBAG, which is one of the reasons mentioned in the introduction to why RBAG has replaced manual grading at sawmills.

In Table 4 it is seen that the RBAG system is underestimating the number of quality grade B boards by 17%, while consequently overestimating the number of boards of the other two grades, A and C. The same result was found in the study by Berglund et al. (2015). RBAG graded 86% of boards correctly or with an underestimated grade. However, of the remaining 14% of boards graded too high, 10 percentage points of those came from boards of grade B or C and were wrongly graded as grade A. As stated in the preamble of the results and discussion section, this is not appreciated by customers. On the other hand, of the 86% correctly or underestimated graded boards, 22 percentage points comes from underestimated boards, which hurts revenues at the sawmill.

When comparing with the RBAG case of the study by Berglund et al. (2015) it is clear that not much has changed. One difference is that in this study the proportion of boards graded correctly as of grade A has dropped 11 percentage points while the other measurements stayed basically the same. As every other grading accuracy measurement of the RBAG systems of the two studies are basically the same, it is interesting to note this 11% drop in accuracy when it comes to grade A. This could be for any number of reasons but is worth remembering for future comparisons.

Table 4 The number and percentage of the 790 boards in each grade by rule-based automatic grading (RBAG) column-wise and by the customer expert row-wise. For the different grades according to the automatic grading, the consistency with the customer expert was also calculated.

		RBAG			
	Grade	A	B	C	Total
Expert	A	93(12%)	38(5%)	4(1%)	135(17%)
	B	70(9%)	197(25%)	133(17%)	400(51%)
	C	8(1%)	35(4%)	212(27%)	255(32%)
	Total	171(22%)	270(34%)	349(44%)	790(100%)

Correct grade total:	(93+197+212)/790 = 64%
Correct or underestimated grade total:	(93+38+4+197+133+212)/790 = 86%
Correct grade A:	(93)/171 = 54%
Correct or underestimated grade B:	(197+38)/270 = 87%

Holistic-subjective automatic grading (HSAG)

For PLS model I, separating boards of grade C from grades A and B, the R^2-value was 0.61 and the Q^2-value was 0.49. This means that 61% of the variance in grade is explained by the model and 49% can be predicted according to cross-validation. The R^2-value for PLS model II was 0.55 and the Q^2-value was 0.44.

Figure 3 shows an observation score plot of the first 2 out of 3 principal components. Therein is a visual indication of the separation problem at hand, where the limit of PLS model I would be represented by a line trying to separate "triangles" from the rest, while the corresponding limit for PLS model II would be a line trying to separate "squares" from the rest.

Choosing threshold limits for board grade separation in an HSAG system is a trade-off between increased sawmill revenues and customer satisfaction. How much the share of higher grades can be increased to improve sawmill profitability depends also on the current market situation.

Table 5 show that the proportion of correctly graded boards in total for the HSAG was 84% with a delivered grade distribution of 16% grade A, 54% grade B, and 30% grade C, which was very close to the reference grading. HSAG produced a total board grade correctness of 84% which was in similar level of what Berglund et al. (2015) did present (78-87%) when two different sets of limits L_A and L_C were tested.

Out of all boards graded as of quality A, 81% were correctly classified and only 3 boards of grade C boards were given grade A.

As for the trade-off between sawmill and customer satisfaction, the HSAG of this study showed a correct or underestimated grade in 91% of the grades of which 7 percentage points of grades were underestimated. Similarly, for the HSAG model II by Berglund et al. (2015); correct or underestimated grade in 92% of the board grades of which 5 percentage points of board grades are being underestimated. Both studies (models) did overestimate 9% of board grades.

When comparing the RBAG model with the HSAG model, there was an increase in total grading accuracy from 64% to 84% with the most significant difference being the ability to distinguish boards of grade A. Out of all the boards graded as grade A by the HSAG model, 81% matched the reference grade A, while only 54% of boards graded as A by the RBAG model matched the reference grade A.

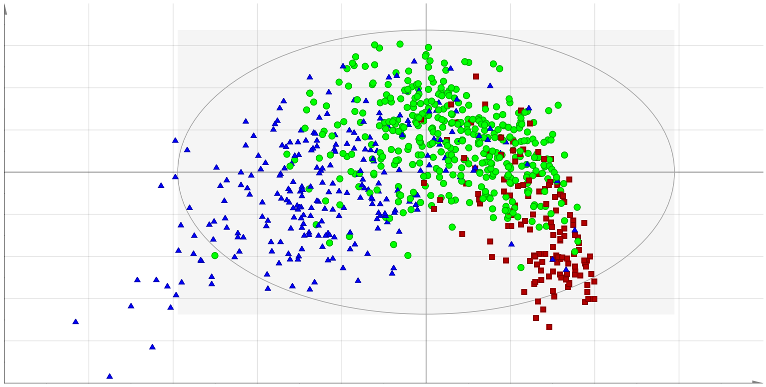


Figure 3. Score plot showing the different grades, according to the customer expert, and their formation in the coordinate system of the two first principal components, t[1], t[2]. Squares = grade A, circles = grade B, and triangles = grade C. Included is also a tolerance ellipse based on Hotelling's (1931) T2.

Table 5. The number and percentage of the 790 boards in each grade by holistic-subjective automatic grading (HSAG) column-wise and by the customer expert row-wise. The threshold limits used for separating board grades were $L_A = 0.43$ and $L_C = 0.5$. For the different grades according to HSAG, the consistency with the customer expert was also calculated.

Follow-up test

		HSAG			
	Grade	A	B	C	Total
Expert	A	105(13%)	30(4%)	0(0%)	135(17%)
	B	22(3%)	353(45%)	25(3%)	400(51%)
	C	3(0%)	44(6%)	208(26%)	255(32%)
	Total	130(16%)	427(54%)	233(30%)	790(100%)

Correct grade total:	(105+353+208)/790 = 84%
Correct or underestimated grade total:	(105+30+0+353+25+208)/790 = 91%
Correct grade A:	(105)/103 = 81%
Correct or underestimated grade B:	(353+30)/427 = 90%

According to Table 6, the RBAG method graded boards with 81% accuracy. Quality A grading was done 100% correct with 27% (4 boards) being borderline lower, while quality C and B grading were done with 76% and 77% correctness.

For the HSAG, Table 7 shows a grading correctness of 95%. Again, quality grade A was done 100% correct and with overall lower number of borderline cases.

When comparing Table 6 and 7, the follow-up test confirms the results earlier shown in Table 5, and the results of Berglund et al. (2015), the HSAG method outperforms the RBAG method.

To do an as thorough investigation of grade A as possible a seventh group was inspected in the same way. This group of boards contained all additional boards classified as grade A by the HSAG system. Of these total 15 additional boards only one was put in a lower class by the customer expert, which is consistent with the shown 95% grading accuracy. Correctly grading more boards of grade A is of course appreciated by both the sawmill and the customer.

Table 6. The result of a customer expert inspection of the three groups of randomly selected boards representing grades A, B, and C according to an RBAG grading system.

*One board each in grade A and C was rejected due to an uncertainty in board number identification.

Table 7. The result of a customer expert inspection of the three groups of randomly selected boards representing grades A, B, and C according to HSAG, using $L_A = 0.43$ and $L_C = 0.5$.

	RBAG Grade			
	A	B	C	Total
Correct	15(100%)	27(77%)	26(76%)	68(81%)
Lower	0(0%)	3(9%)	-	3(4%)
Higher	-	5(14%)	8(24%)	13(15%)
Total	15*	35	34*	84
Borderline lower	4(27%)	3(9%)	0(0%)	7(8%)
Borderline higher	0(0%)	2(6%)	2(6%)	4(5%)

	HSAG Grade			
	A	B	C	Total
Correct	16(100%)	34(97%)	32(91%)	82(95%)
Lower	0(0%)	1(3%)	-	1(1%)
Higher	-	0(0%)	3(9%)	3(4%)
Total	16	35	35	86
Borderline lower	1(6%)	1(3%)	0(0%)	2(2%)
Borderline higher	0(0%)	3(9%)	1(3%)	4(5%)

CONCLUSIONS

In this study, a partial least squares multivariate regression based holistic-subjective automatic grading (HSAG) method was used to grade sawn timber based on knot distribution. The method was compared with a conventional rule-based automatic grading (RBAG) system and both systems used an expert on board grade quality for the North African market as reference.

The HSAG model was able to grade boards with greater accuracy than the RBAG method. The RBAG system graded boards with about 65% accuracy and the HSAG model about 85% accuracy. This difference in grading accuracy also showed in the follow-up test where the RBAG managed a total correctness of about 80% while the HSAG managed 95%.

Other than being more accurate in its quality grading, an HSAG system separates quality grades by a single limit, which directly balances sawmill revenues vs. customer satisfaction of the grading results. This can be seen as a strength of an HSAG system because of the difficult work of defining new grades or changing existing ones of the conventional RBAG systems.

REFERENCES

Berglund, A., Broman, O., Oja, J. and Grönlund, A. (2015). *Customer adapted grading of Scots pine sawn timber using a multivariate method. Scand J For Res*. 30:1, pp 87-97.

Breinig, L., Leonhart, R., Broman, O., Manuel, A., Brüchert, F. and Becker, G. (2015) Classification of wood surfaces according to visual appearance by multivariate analysis of wood feature data. *J Wood Sci* 61:2, pp. 89-112.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C. and Wold, S. (2006). *Multi and megavariate data analysis – basic principles and applications*. 2nd ed. * Umeå (Sweden): Umetrics AB.

Esbensen, K. (2002). *Multivariate data analysis in practice*. *Oslo (Norway): CAMO.

European Confederation of Woodworking Industries. (2004). *Roadmap 2010 executive summary.* Brussels (Belgium).

Finscan. (2017). BoardMaster – automatic grading system. Available at: http://www.finscan.fi/en/ [Accessed 26 Jan. 2017].

Grönlund, U. (1995). *Quality improvements in forest products industry: classification of biological materials with inherent variations* [doctoral thesis]. Luleå (Sweden): Division of Quality Technology and Statistics, Luleå University of Technology.

Hotelling, H. (1931). The generalization of student's ratio. *Ann Math Stat*. 2, pp. 360–378.

Kauppinen, H. (1999). *Development of a color machine vision method for wood surface inspection* [doctoral thesis]. Oulu (Finland): Department of Electrical Engineering, Faculty of Technology, University of Oulu.

Kline, D. E., Surak, C. and Araman, P. A. (2003). Automated hardwood lumber grading utilizing a multiple sensor machine vision technology. *Comput Electron Agric*. 41, pp. 139–155.

Labeda, A. (1995). Practical experiences from industrial implementation of neurosorter in lumber grading. In: *2nd International Seminar/Workshop on Scanning Technology and Image Processing on Wood*, Skellefteå, Sweden: Luleå university of technology, p 18.

Lycken, A. (2006). Comparison between automatic and manual quality grading of sawn softwood. *For Prod J*. 56, pp. 13–18.

Lycken, A. and Oja, J. (2006). A multivariate approach to automatic grading of Pinus sylvestris sawn timber. *Scand J For Res*.21, pp.167–174.

Martens, H. and Naes, T. (1989). *Multivariate calibration*. New York (USA): Wiley.

Niskanen, M. (2003). *A visual training based approach to surface inspection* [doctoral thesis]. Oulu (Finland): Department of Electrical and Information Engineering, Faculty of Technology, University of Oulu.

Oja, J., Grundberg, S., Fredriksson, J. and Berg, P. (2004). Automatic grading of sawlogs: a comparison between X-ray scanning, optical three-dimensional scanning and combinations of both methods. *Scand J For Res*. 19, pp. 89–95.

Silvén, O., Niskanen, M. and Kauppinen, H. (2003). Wood inspection with non-supervised clustering. *Mach Vision Appl*. 13, pp. 275–285.

Swedish Sawmill Managers Association. (1994). Nordic timber: grading rules for pine (Pinus sylvestris) and spruce (Picea abies) sawn timber: commercial grading based on evaluation of the four sides of sawn timber. Umeå (Sweden).

PUBLICATION II

Multivariate Product Adapted Grading of Scots pine Sawn Timber for an Industrial Customer, part 1: Method Development

Authors and year:
Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D.
2019

Publication originally published in:
Wood Material Science & Engineering 14(6), 428–436
DOI: 10.1080/17480272.2019.1617779

WOOD MATERIAL SCIENCE & ENGINEERING
2019, VOL. 14, NO. 6, 428–436
https://doi.org/10.1080/17480272.2019.1617779

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Multivariate product adapted grading of Scots pine sawn timber for an industrial customer, part 1: Method development

Linus Olofsson [a], Olof Broman[a], Johan Skog[b], Magnus Fredriksson[a] and Dick Sandberg [a]

[a]Wood Science and Engineering, Luleå University of Technology, SE-931 87 Skellefteå, Sweden; [b]RISE Bioeconomy, Research Institutes of Sweden, SE-931 77 Skellefteå, Sweden

ABSTRACT
Rule-based automatic grading (RBAG) of sawn timber is a common type of sorting system used in sawmills, which is intricate to customise for specific customers. This study further develops an automatic grading method to grade sawn timber according to a customer's resulting product quality. A sawmill's automatic sorting system used cameras to scan the 308 planks included in the study. Each plank was split at a planing mill into three boards, each planed, milled, and manually graded as desirable or not. The plank grade was correlated by multivariate partial least squares regression to aggregated variables, created from the sorting system's measurements at the sawmill. Grading models were trained and tested independently using 5-fold cross-validation to evaluate the grading accuracy of the holistic-subjective automatic grading (HSAG), and compared with a re-substitution test. Results showed that using the HSAG method at the sawmill graded on average 74% of planks correctly, while 83% of desirable planks were correctly identified. Results implied that a sawmill sorting station could grade planks according to a customer's product quality grade with similar accuracy to HSAG conforming with manual grading of standardised sorting classes, even when the customer is processing the planks further.

ARTICLE HISTORY
Received 12 March 2019
Revised 18 April 2019
Accepted 25 April 2019

KEYWORDS
Sawn timber; visual grading; customer adoption; discriminant analysis

Introduction

During the last decades, automatic grading of sawn timber has replaced manual grading in many sawmills in Sweden, as well as in many other countries around the world, and it has improved the grading accuracy significantly compared to manual grading (Lycken 2006). The greater grading accuracy gives a sawmill greater control to make conscious decisions about the operations in the entire product line. The most common automatic grading approach is rule-based automatic grading (RBAG), which is a system that uses grading rules (limits) for detected features or measured variables, e.g. maximum size of dead knots. Rule-based systems have been used in a number of applications throughout the sawmill process, e.g. Lycken (2006) who used a straight-forward RBAG approach to grade sawn timber according to the Nordic Timber Grading Rules (Swedish Sawmill Managers Association 1994), and Kline *et al.* (2003) who used a multi-sensor set-up with fuzzy-logic for automatic grading of hardwood lumber. Several grading systems in the industry use some application of rule-based, or in general objective, grading, e.g. the automatic grading and sorting system by FinScan used in this study (Anon 2018).

RBAG is objective by nature and has the strength of being able to strictly follow standards, such as Swedish Sawmill Managers Association (1994), which in detail specifies the grade of sawn timber that the sawmill is selling – with of course some error margin. Individual customer's needs may however not at all be in line with these standardised grades with their corresponding pricing. Based on the standardised grading rules, industrial customers can sometimes request a custom-made grade that follows a different set of rules. However, such a custom order is complicated to produce (Lycken and Oja 2006), because:

(I) It is difficult for a customer to describe its subjective view of the desired plank quality in a way that can easily be defined in objective grading rules.
(II) The number of variables that can be controlled to specify a grade is often more than enough to make customisation complicated.

Since a rule-based system needs a conforming and deliberate set of rules to result in a concise grading, these problems make customisation troublesome. Furthermore, a fine-tuning process by trial and error can be costly as the results of a custom grading attempt cannot easily be validated before the customer receives a delivery or even before the customer produces further products from the sawn timber. Custom-made grading settings are therefore seldom made due to the complicated process of adjusting the grading rules according to the customer's needs, even

CONTACT Linus Olofsson linus.olofsson@ltu.se
This article has been republished with minor changes. These changes do not impact the academic content of the article.

though they could be beneficial to both sawmill and customer.

This customisation difficulty of RBAG indicated the need for a holistic-subjective automatic grading (HSAG) methodology; i.e. a grading that incorporates subjective grading of the entire piece. The grading method developed in this study is an extension of the work by Lycken and Oja (2006) and Berglund *et al.* (2015), the latter validated by Olofsson *et al.* (2017), both using a multivariate partial least squares regression (PLS Geladi and Kowalski 1986) based HSAG to grade sawn timber in conformity with manual grading. These studies showed that PLS-based discriminant analysis (PLS-DA) HSAG is superior to RBAG when conforming with manual grading. Multivariate based classification has similarly been used by Broman (2000) and Breinig *et al.* (2015) to classify wood surfaces by their appearance with regards to knots. Oja *et al.* (2003, 2004) also investigated PLS regression to grade logs based on the grade of the centre yield, and Hagman and Grundberg (1995) used multivariate image analysis to classify knots and wood surface features in X-ray computed tomography images. Using HSAG addresses the two specified difficulties with RBAG:

(I) The customer needs only to specify whether or not they want that piece of sawn timber – no objective description is necessary.
(II) The amount of available variables in the scanner is irrelevant as the HSAG method uses multivariate PLS-DA.

The customer will specify whether or not a piece of sawn timber is desirable by looking at the quality outcome of that piece to focus the automatic grading on product quality instead of the customary sawn timber quality. The use of PLS-DA frees the user from manually calibrating a large set of grading rules and instead rely on a computer to automatically determine the relationship between measured variables and the desired grading outcome.

The goal of the first part of this study was to investigate the possibilities of further expanding the use of PLS-based HSAG to predict not only a plank's standardised sorting grade at the sawmill but whether the plank will result in a product of desirable quality for the customer. As the collaborating industrial planning mill customer splits each plank into three boards before planing and milling, only two out of six flat faces of the boards are visible to the automatic scanner for grading at the sawmill. The splitting and further processing of the wood material might make the correlation between scanner measurements and product grade diffuse, especially for the centre plank of the splitting.

An automatic scanning system in a sawmill is a calibration-sensitive system located in a generally dusty environment, which raises the question of whether the variable-grade relationship can be found reliably. Accordingly, the goal of part two of this study was to investigate the robustness of PLS-based HSAG grading towards distortions of the measured input variables.

By scanning a set of planks, storing the measured feature variables, and training a PLS model against the resulting product grade from the collaborating industrial customer, the sawmill will be able to grade planks by its customer's product grade reliably, and the customers can purchase sawn timber that is suitable for its intended products.

Materials and methods

Material

A total of 308 Scots pine (*Pinus sylvestris* L.) planks were studied at Kåge sawmill and Lundgren's industrial planing mill in northern Sweden. The planks were sawn from top-logs from the sawmill's log sorting station. Each top-log was cant sawn and resulted in the centre yield of two planks of 50 by 150 mm in cross-section and between 3.5 and 5.7 m in length. The planks were dried to 14% moisture content before being scanned at the sawmill's dry sorting station. The scanner, grader, and sorter used was a Boardmaster by FinScan (Anon 2018) – the entire system will be referred to as Boardmaster.

Data collection at Kåge Sawmill

The test material came from logs that originated from different logging areas. The logs were sorted to a suitable top diameter class for the plank dimensions requested by the customer. Following conventional sawing and drying operations, two packages of approximately 150 dried planks were selected for the study. The planks were marked with an ID number for traceability throughout the study before being transported cross-wise through the Boardmaster at Kåge sawmill. The Boardmaster used cameras to scan, detect, and measure the planks' features. Each feature was described by size, position, and an attributed class – e.g. the size and position of a knot that was classified as a dead knot. All the planks were delivered to Lundgren's planing mill for processing and grading according to Lundgren's product requirements.

The feature variables measured by the Boardmaster were the foundation of the grading and they were therefore carefully selected based on the desired grading outcome. A Boardmaster can detect plank features such as knots, bark pockets, rot, discolouration, cracks, wane, warping, and so on. Knots are the most important features affecting the general appearance of sawn timber, which is why most of the grading rules in Swedish Sawmill Managers Association (1994) are related to knots. Bark pockets are also important to the planing mill due to the splitting and milling process. Therefore, features related to knots and bark pockets were specifically selected for this study, due to their importance for the planing mill and the difficulty in holistically describing them for the entire piece. All other features were ignored throughout this study, both at the sawmill and planing mill.

Data collection at Lundgren's planing mill

At the planing mill, each plank was split into three boards. Each board was planed, milled, and manually graded by the planing mill staff: grade A for the desired higher quality

Table 1. Complete description of all the *x*-variables used in the study, together forming the explanatory **X**-matrix.

#	Variables		Defect types	Plank faces	Sections
1	Total Number of defects		Sound knot	Inner face	1
2	Avg. defects size	(mm)	Dead knot	Outer face	3
3	Std. dev. of defects size	(mm)	Sound or dead knot	Both edges	5
4	Maximum defects size	(mm)	Bark-ringed knot		
5	Sum of defect area	(mm^2)	Rotten knot		
6	Ratio of $\frac{\text{defect area}}{\text{surface area}}$	(%)	Bark pocket		
7	Ratio of $\frac{\text{sum of defect area}}{\text{sum of all defect areas}}$	(%)			
8	Number of defects per metre	(m^{-1})			
	Ratio of defects of sizes:[a]				
9	≤ 9 mm	(%)			
10	10−19 mm	(%)			
11	20−29 mm	(%)			
12	30−39 mm	(%)			
13	40−59 mm	(%)			
14	60−79 mm	(%)			
15	≥ 80 mm	(%)			
	Number of defects per metre:[a]				
16	≤ 9 mm	(m^{-1})			
17	10−19 mm	(m^{-1})			
18	20−29 mm	(m^{-1})			
19	30−39 mm	(m^{-1})			
20	40−59 mm	(m^{-1})			
21	60−79 mm	(m^{-1})			
22	≥ 80 mm	(m^{-1})			
	22 variables	×	6 defect types ×	3 sides ×	9 zones = 3564

[a]This a header and does not count as a variable.

Notes: Each variable in the first column was replicated for each entry in each of the other columns. The number of entries of each column is shown at the bottom where the total number of variables is the multiple of the values in the bottom row, i.e. 3564 variables.

product or grade B for an undesired lesser quality product. This triplet of board grades from a plank defined the plank's grade by majority as A or B as desirable or undesirable by the planing mill. A desirable plank was a plank that produced the quality yield of two or three quality A boards and was given the grade A, and a plank that produced one or zero quality A boards, was given the grade B. Digital labels associated with each plank tracked the triplet of board grade results from each plank.

Construction of regression components

The outcome targeted in this study was a grading model that determines whether or not a plank is of grade A according to the planing mill, i.e. whether or not the planing mill will be able to produce a majority of quality A boards from the plank. Such a grading model was created using an HSAG method adapted specifically for the product produced at the planing mill. The HSAG was implemented as multivariate PLS regression discriminant analysis (PLS-DA[1]), and the explanatory **X**-matrix of plank measurements and response-grade **y**-vector were created from the data collection as follows.

Explanatory X-matrix

Since the size, position, and feature type measured by the Boardmaster are not sufficiently descriptive to objectively capture the subjective quality judgement of the customer, the Boardmaster's variable resolution was increased. Using knot and bark measurements from the Boardmaster, an additional software tool was used to create the set of 3564 variables, shown in Table 1 (see Berglund *et al.* 2015). The twenty-two variables in the first column were created for each defect type listed in the second column, measured separately for each face-entry in the third column, and replicated in each of the one, three, and five longitudinal sections in the fourth column (see Figure 1). An example of such a variable could read as follows, using the second entry of each column; "The average dead-knot size on the outer face of the plank in the longitudinal section two out of three". The three plank-face categories were created to capture the different importance of the different faces, with both edge-faces merged as they are similar and equally important in the milling process. Each plank was copied virtually three times and divided into equal longitudinal sections to capture the different importance of the plank ends or centre. The first plank had one section for the entire plank, the second plank had three sections, and the third plank had five sections, totalling to nine sections.

The matrix of all these variables for each plank in the investigation was the explanatory **X**-matrix used in the study. The construction process of the matrix was undertaken after the scanning procedure at the sawmill. Possible detection errors by the Boardmaster scanning system was not considered.

Response y-vector

The response variable *y* for each plank was a binary representation of the desirability for the planning mill. All planks formed the **y**-vector used for the PLS regression. A desirable plank (grade A) was represented by 1 and undesirable (grade B) by 0.

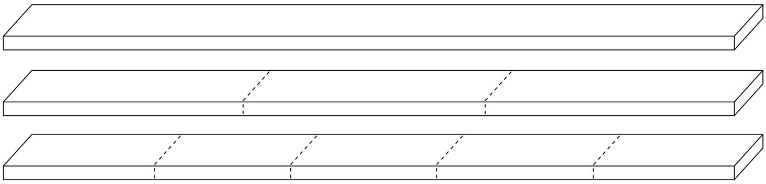


Figure 1. Illustration of the virtual copies of each plank (Berglund *et al.* 2015).

Table 2. The resulting numbers of planks with labels AAA, AA, BB, or BBB and their respective grade A or B.

	Grade A		Grade B	
Label	AAA	AA	BB	BBB
Number	126	73	43	66

Note: A total of 199 grade A planks and 109 grade B planks.

Since each plank resulted in three boards, each plank was digitally labelled AAA, AA, BB, or BBB based on the quality of boards produced – omitting B and A from the two mixed labels AAB and BBA to keep the labelling clean, as shown in row one of Table 2. This labelling makes it possible to investigate classification difficulties with borderline cases, e.g. whether a BB plank should be assigned grade B or grade A.

From the 308 planks, 924 boards were produced and graded. The resulting quality distribution of the complete data set is shown in Table 2, which shows the numbers of planks labelled AAA, AA, BB, and BBB, where the first two where given the grade A and the last two the grade B.

Implementation and evaluation of PLS classification models

With such a large explanatory **X**-matrix (308 by 3564 in size) and a single binary **y**-vector (**Y**-matrix 308 by 1 in size), PLS-DA was used using the SIMCA 14 software (Anon 2019). The strengths of PLS regression are in its ability to find the strongest correlation between the explanatory variables in **X** and separation of classes in the response variable(s) in **Y**. An automatic algorithm in the SIMCA 14 software was used to calculate one to several latent variables by linearly combining the explanatory variables in **X** that maximised the separation of classes in **Y**. PLS is resilient to noise and can handle many, possibly strongly correlated, variables, which is the case in this study. There is also a large toolbox for inspecting a PLS prediction model, making models easy to understand and adjust. For a more thorough overview of PLS Regression, see Geladi and Kowalski (1986) or Wold *et al.* (2001).

Prediction models can be evaluated by considering the results of using a trained model to predict a test set, which provides a measure of the grading accuracy and stability of the model. Ideally, when training and testing a prediction model, one data set is used for training (fitting) and another data set is used to test the prediction accuracy of the model. If only one data set is available, and it is large enough, it can be artificially split into two sets, in which case, the test set is usually a small sample of the whole data set, e.g. 20% of the observations. The benefit of training and testing the model on different data is that you can measure how the model copes with predicting previously unseen objects. The downside to such a split of the data set is that the model is trained on a smaller data set compared to a model trained on all available data and is therefor a weaker discriminator, as well as having only a relatively small test set which might not be completely representative of the entire data set (nor the entire population).

A common approach for dealing with data sets which are too small to be split into two sets is re-substitution (self-prediction); to train a prediction model based on the entire data set and then predict the entire data set. The benefit of this method is that the data set is re-purposed for both training and testing, artificially doubling the size of the data set. However, evaluating such a model is difficult as measurements of how such a model would cope with an unknown test set are only indicative. Re-substitution testing is generally known to be optimistic (biased) of a prediction models performance as it is optimised for the current test set. The classification performance of PLS-based HSAG by re-substitution testing will in this study serve as supplementary results, due to the somewhat small data set.

The data set used in this study was complex, with 3564 *x*-variables in an attempt to objectively describe the natural variation of knots and bark in a plank in relation to a customer's product quality by a binary response *y*-variable. Furthermore, the relationship between *x*-variables and plank grade was sometimes ambiguous, due to e.g. desirable sound knots cracking in the planing or milling process by chance. Because of this complex problem, and the continuous flow of planks through a sawmill sorting station, it was important to test the model on an unseen test set as this is its intended use. However, as the data set only contains roughly 100 grade B planks (Table 2), which is the estimated minimum of required observations per grade for PLS-based grading according to (Lycken and Oja 2006), and since the data is so complex, it can be difficult to evaluate the model based on a single test set due to the inherent variability of the data. To circumvent this complexity as much as possible, and to benefit from both of the methods mentioned above, 5-fold cross-validation was here adopted and substituted with a re-substitution test to evaluate the grading accuracy and stability of the PLS-based HSAG method.

Defining training and testing data sets

To make a comprehensive evaluation of the PLS-based HSAG method used in this study 5-fold cross-validation was used and compared with re-substitution. Based on the size of the data set, and the limited numbers of grade B planks (Table 2), the data was randomly re-sampled five times, which were called re-samplings 1–5. Each re-sampling resulted in a unique test set of 59 boards defined as the closest fit to one-fifth of the entire data set that followed the labelled quality distribution in Table 2, while 249 planks remained for use as the training set in each re-sampling. Theoretically, a completely random selection of test set members would follow the same quality distribution, but with the low number of grade B planks this enforcement was made in order to maintain a fair proportion between the training and the test sets for each label-class of planks during the cross-validation. This guided random selection is especially important with the ambiguity problem mentioned in the previous section. Each test set was unique in the sense that each observation (with some round-off error) was a part of one test set only. One PLS-DA model was made for each of the five re-samplings in addition to the model created from the entire data set for the re-substitution. The comparison between the prediction results of each of the five re-samplings, together with their average behaviour, and the re-substitution results gave measures of the prediction accuracy and stability of the methodology.

PLS-DA classification

A PLS-DA model was linearly fitted to maximise the correlation between the explanatory variables in **X** and the separation of classes in **Y** in the training data set. When **X** measures observations from a test set, the multivariate regression line is used to estimate the observations' response **Y**. These predicted values usually comes in the form of a value approximately in the range [0, 1][2]. This estimate can be interpreted as a probability of the new observation belonging to grade A, represented by 1. To complete the classification, a threshold is selected, e.g. 0.5, which defines any observation with an estimated response higher than 0.5 as grade A, otherwise B. The choice of threshold is in this study chosen to try to maximise the classification accuracy, although a different threshold might be desired by the sawmill and customer due to e.g. adequate delivery volume requirements. By investigating the PLS-DA regression of the training data, the threshold was chosen to separate the two grades optimally for the training sets of the cross-validation models and for the re-substitution model. For each prediction model, the choice of threshold is conceptually visualised by the choice of threshold for the re-substitution, shown in Figure 2. The resulting model and threshold from the training set are then used to grade the separate test set. This procedure was performed for all re-sampling models. As the re-substitution model was trained on the entire data set, there was no separate test set, and it is easy to see that the optimal threshold for the re-substitution might not be the optimal threshold for a new separate test set. Measurements of a model's classification performance, based on re-substitution, are hence optimistic (data set bias) to some degree. Such optimism (bias) can be extremely prominent in very small data sets with randomised class belonging (e.g. randomised plank grade) (Westerhuis *et al.* 2008) but becomes far less prominent for larger data sets (Lance *et al.* 2000) and less prominent still with non-randomised class belonging. For the complex data set of 308 planks in this study, the classification performance measured by re-substitution could approximate the upper limit for classification performance, using the presented scanner system and implementation of HSAG.

Table 3. Variables used to evaluate the grading outcome of the tested models.

Variable name	Description
A-correct	The proportion of quality A planks correctly graded as grade A.
B-correct	The proportion of quality B planks correctly graded as grade B.
Tot-correct	The proportion of correctly graded planks in total.
Delivered	The proportion of scanned planks delivered from the sawmill, i.e. planks graded A, correct or not.
A-purity	The proportion of the planks delivered to the planing mill that were of grade A.

Classification evaluation procedure

With six PLS models, one for each of the five re-samplings and one for the re-substitution, six test sets were graded. The

Table 4. Misclassification table (confusion matrix) of the average prediction results of 5-fold cross-validation, each model predicting their corresponding test sets at a threshold between 0.49–0.60, which on average was 0.55.

				Cross-validation Predicted			
		Label	Number	A	B	Correct	
Cross-validation Observed	A	AAA	24	21	3	87%	83%
		AA	14	11	3	76%	
	B	BB	8	5	3	40%	59%
		BBB	13	4	9	71%	
		Totals	59	41	18	74%	

Notes: Predicted grades were determined at the sawmill, and observed grades at the planing mill. Predicted number of planks are rounded to whole planks from the five re-samplings.

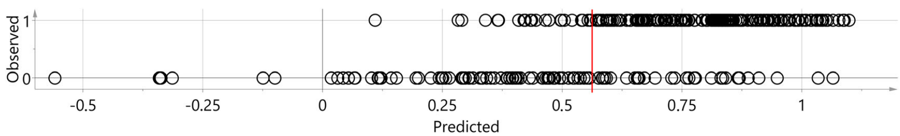


Figure 2. Observed vs predicted grade plot of the re-substitution test, showing the spread and separation of the predicted grade of all planks along the horizontal axis and the observed (true) grade on the vertical axis. Grade A planks are represented by 1 and grade B by 0. A threshold on the horizontal axis defines observations above the threshold as being of grade A and vice versa. The optimal class-separating threshold 0.56 is marked as a red vertical line. This figure is conceptually the same for each re-sampling model of the cross-validation.

Table 5. Misclassification table (confusion matrix) for the prediction results of the re-substitution test, classifying the entire data set with an optimal class-separating threshold of 0.56.

				Re-substitution Predicted			
		Label	Number	A	B	Correct	
Re-substitution Observed	A	AAA	126	118	8	94%	89%
		AA	73	59	14	81%	
	B	BB	43	17	26	61%	89%
		BBB	66	13	53	80%	
		Totals	308	207	101		83%

Note: Predicted grades were determined at the sawmill, and observed grades at the planing mill.

grading accuracy was investigated with the five measurements described in Table 3. Note that A-correct, B-correct, and Tot-correct all show different proportions of correctly graded planks, which is the correctness percentage to the right in the rightmost column in Tables 4 and 5.

Results

To evaluate the grading accuracy and reliability of multivariate PLS regression as a method to predict the yield of a customer's product quality, the prediction results of 5-fold cross-validation are presented, supplemented with results of a re-substitution test. For the five re-sampling models one or two (once) PLS components were used during modelling, and for the prediction model used for the re-substitution test one PLS component was used. The number of PLS components used was for all models determined automatically by cross-validation by the SIMCA 14 software used, to prevent over-fitting the prediction models to the training set.

Figure 2 shows the observed vs predicted grade for the re-substitution, where the prediction model was trained and tested on the entire data set. High-density groupings of observations are evident as black regions with the optimal class-separating threshold highlighted. Conceptually, this result was the same for each test set in the 5-fold cross-validation, apart from the thresholds which were chosen before testing.

Table 4 shows the average prediction results of the 5-fold cross-validation, with the average class-separating threshold 0.55 (0.49–0.60), which on average classified 74% (61–80%) of the planks correctly. Table 5 shows the re-substitution prediction results using the optimal class-separating threshold of 0.56, where 83% of the planks were correctly graded.

The measurements from Table 3, A-correct, B-correct, and Tot-correct, were measured vs the threshold value and compared in Figure 3, which show the average prediction results of the cross-validation, and the prediction results of the re-substitution. The average threshold of the cross-validation and the optimal threshold of the re-substitution are close to the same value, 0.55 and 0.56 respectively. The average threshold of the cross-validation resulted in 74% correctly graded planks, which is close to the peak Tot-correct value of 76% at the on average optimal threshold 0.52 (Figure 3).

Figure 4 shows Tot-correct vs threshold for each of the five re-samplings, including their average. All re-samplings, except re-sampling 3, follows the average trend quite closely. Each prediction model for each re-sampling had a threshold between 0.49 and 0.60 and classified between 61% and 80% of planks correctly, which on average classified 74% of planks correctly at the average threshold 0.55.

To follow up the implications of using PLS-based HSAG for the sawmill and planing mill, Figure 5 show prediction results using the measurements from Table 3 Delivered, and A-purity, for the average of the cross-validation models and the re-substitution respectively, and again they showed similar behaviour.

Discussion

The first part of this study indicated the potential at a multivariate PLS-regression-based HSAG method to an industrial sawmill's dry sorting station, even when the scanned sawn

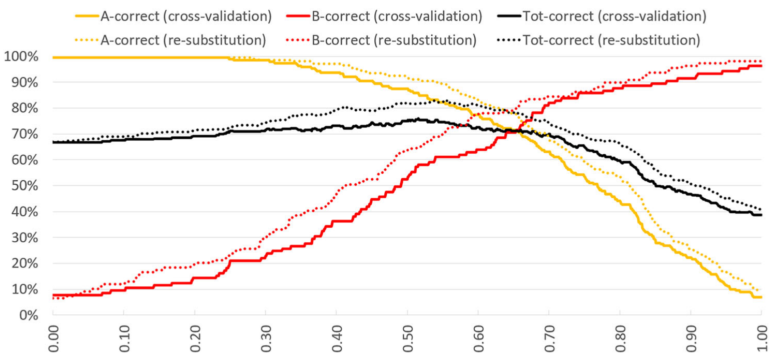


Figure 3. Average prediction results of the 5-fold cross-validation and for the re-substitution, showing the dependence of the measurements A-correct, B-correct, and Tot-correct on the threshold value. For the average of the cross-validation, the measurement Tot-correct showed 74% correctly classified planks at the threshold 0.55 (the optimal threshold was on average 0.52 where Tot-correct was 76%), while for the re-substitution the value of Tot-correct was 85% at the optimal threshold 0.56.

timber was subsequently split into three boards for further processing. On the basis of the average of the 5-fold cross-validation, 74% of all the planks were correctly graded while 86% of grade A planks were correctly identified at the average class-separating threshold of 0.55 (Table 4). Using the optimal class-separating threshold of 0.56 for the re-substitution model, 83% of all the planks were correctly graded and 89% of grade A planks were correctly identified (Table 5). Both the average cross-validation results and the re-substitution results showed in general a very similar behaviour for all measurements used (Table 3), as shown in Figures 3 and 5.

The similarity of the grading accuracy measurements at similar threshold values in the two investigations indicates that the product adapted HSAG method can grade planks according to customer product quality yield. This level of prediction accuracy is comparable to that achieved in previous work on multivariate-based HSAG grading; Berglund *et al.* (2015) showed, using re-substitution, a prediction accuracy of 76% and 87% for two different thresholds when grading sawn timber to conform to manual grading for a North African importer, and Lycken and Oja (2006) showed, using re-substitution, a prediction accuracy of 80% and 85% when grading planks to conform with manual grading of the standardised sorting grades Swedish Sawmill Managers Association (1994). The present study implies that using the current hardware a sawmill could grade sawn timber according to a customer's product quality with an accuracy similar to that achieved when grading sawn timber according to predefined sawn-timber-grades. This customer product adapted grading was possible even when the customer processes the wood material by both splitting it into three boards and further refining each piece to a finished product. The grading accuracies of the above presented multivariate grading results can be compared to the 80–90% grading accuracy of Nordic RBAG systems (Lycken and Oja 2006).

The re-substitution model achieved a separation of grade A and B planks to the extent shown in Figure 2, which is conceptually very similar to the separation achieved by each of the re-sampling models of the cross-validation. The result of these separations are shown in Tables 4 and 5 for the cross-validation and re-substitution at thresholds 0.55 and 0.56, respectively. For both tests, the grading accuracy of grade B planks was much lower than for grade A planks. This was due to the large spread of predicted grades of grade B planks, shown in Figure 2, with several observations with a predicted response grade above 0.75. These grade B planks were from a measurement standpoint very similar to distinct grade A planks, which indicated that ambiguous, or possibly even undetected defects caused the plank to receive grade B at the planing mill. Today's automatic sorting systems have a feature detection hit rate of around 70–80% (Lycken and Oja 2006) which could explain at least parts of this ambiguity problem. The higher grading accuracy of the re-substitution is mostly due to the increased grading accuracy of grade B planks. This could in part be due to the higher number of grade B planks in the training set: 109 grade B planks were in the training set for the re-substitution while 88 planks were in the training set in each of the models used for the cross-validation, which is slightly below the minimum recommended number of planks per grade suggested by Lycken and Oja (2006). The lower grading accuracy of grade B planks could also imply that grade B planks were less correctly graded in comparison to grade A planks due to the data set consisting of only one-third grade B planks. The low number of grade B planks was the primary reason re-substitution was used as supplementary testing to the cross-validation.

Figure 4 showed that the individual re-sampling models of the cross-validation graded planks with an accuracy between 61 and 80% at thresholds between 0.49 and 0.60, and on average 74% at the average class-separating threshold 0.55. This accuracy range, and the way the measurement Tot-correct for each re-sampling follow the average behaviour in Figure 4 with little variation, except for re-sampling 3, indicate that the data set used was large enough for reliable cross-validation testing. The same figure also show the need for a cross-validation, as a single randomised split of the data set into one training and one test set could have shown erratic results, e.g. re-sampling 3. The re-substitution test follows much the same behaviour as the average results of the cross-validation but with slightly higher grading accuracy, further supporting the grading accuracy of HSAG, especially in terms of the amount of planks delivered from the sawmill and the proportion of desirable planks received by the planing mill, shown in Figure 5. Both testing methods

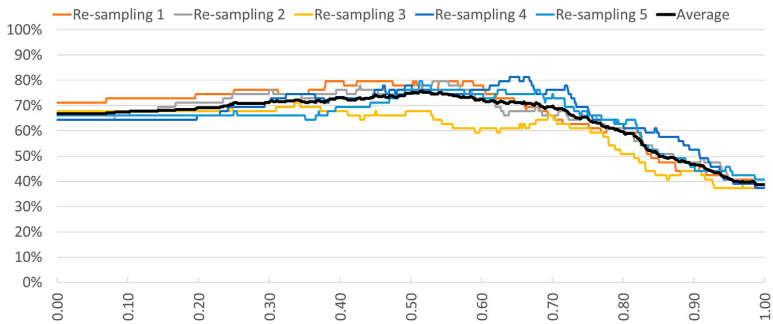


Figure 4. The prediction measurement Tot-correct vs threshold value for re-samplings 1–5, including the average. The value of Tot-correct for the different re-samplings at the pre-selected thresholds was in the range 61–80% at thresholds between 0.49 and 0.60, and on average Tot-correct was 74% at the average threshold 0.55.

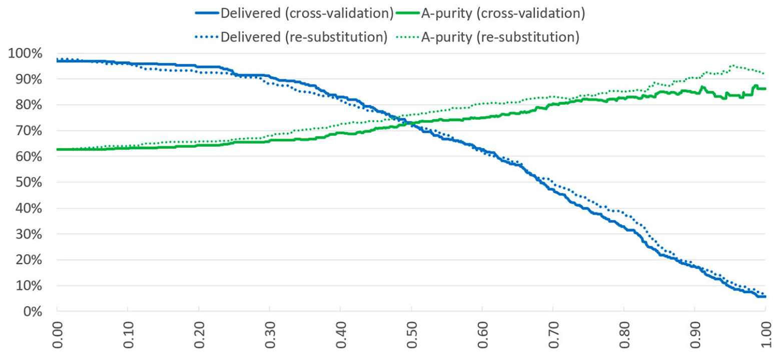


Figure 5. Prediction measurements (Table 3) Delivered and A-purity vs threshold of the average prediction results of the 5-fold cross-validation and for the re-substitution model.

indicate that PLS-based HSAG generate new predicted grades for sawn timber pieces with reliable accuracy. Furthermore, Figure 3 show that the prediction accuracy is slowly changing for practically applicable thresholds, say between 0.30–0.70, which in combination with Figure 5 suggests that the sawmill and customer can decide on a suitable and reliable threshold without great loss of prediction accuracy.

The measurements Delivered and A-purity were considered the general satisfaction of the grading outcome at the sawmill and planing mill respectively, i.e. the proportion of scanned planks graded and sold as grade A by the sawmill, and the proportion of grade A planks received in a batch purchased by the planing mill. The measurements Delivered and A-purity are paramount in a sorting discussion between sawmill and customer which for the average cross-validation results and the re-substitution results showed a very similar behaviour in Figure 5. Further simplifying the sorting discussion is the single threshold variable controlled in a PLS-based HSAG system, in contrast to RBAG systems which requires an objective description of a subjectively good quality sawn timber, using a large set of uncoupled variables.

Future research should investigate the methodology on larger data sets than in this study and preferably multiple and separate data sets simultaneous to further validate the robustness of the PLS-based HSAG methodology. A balance between the number of observations in each class could also be desired. A separate data set is especially needed because of the heterogeneous nature of wood and the continuous flow of sawn timber through a sawmill, never seeing the same piece twice. A large enough data set is needed to enable sufficient information to be obtained on all possible features of wood, as machine-learning techniques like PLS require large training sets that include every possible defect (categorically at least) to train models for future use.

The handling and processing of large sets of unsorted planks by both sawmill and customer to train a grading model can quickly become cumbersome, especially for the customer who has to process not only good quality reference planks but also obviously poor quality planks for the sake of having a bad quality reference for training. Future work should therefore investigate the possibility of using the cameras used by the scanning system to show the sawn timber and allow the customer to grade the images as being desirable or not, which would circumvent both transportation and processing of material for a training data set. This could allow for customer adapted grading.

Conclusions

This study showed that it is possible to use a multivariate PLS-regression-based HSAG to predict the product quality yield of an industrial planing mill, with regards to the performance of the methodology based on measurements made at a sawmill dry sorting station. The trained HSAG model can predict the final product grade at the planing mill based on measurement from the sawmill's scanner system, even though the scanned sawn timber is split into three boards at the planing mill before each board was further processed before final grading. This final product grade was the only input required from the customer, which was a simple holistic-subjective grading of their own product; removing the difficulties found in previous attempts to customise an RBAG system according to the customer's needs. The holistic-subjective automatic grading methodology simplifies the customisation process drastically for both sawmill and customer.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Financial support from the Swedish Innovation Agency (VINNOVA), project Sawmill 4.0 – Customised flexible sawmill production by integrating data driven models and decisions tools 2018-02749, is gratefully acknowledged.

Notes

1. PLS-DA is PLS regression implemented to distinguish between classes and is mathematically equivalent to PLS regression.
2. An estimate can lie outside this range e.g. if the observation is more extreme in **X** than any observation in the training set.

ORCID

Linus Olofsson http://orcid.org/0000-0002-5562-5142
Dick Sandberg http://orcid.org/0000-0002-4526-9391

References

Anon (2018) Boardmaster. https://finscan.fi/products/boardmaster/?lang=en. (2018-06-12). Note: older predecessor to BoardmasterNOVA.

Anon (2019) Simca. https://umetrics.com/products/simca. (2019-01-03).

Berglund, A., Broman, O., Oja, J. and Grönlund, A. (2015) Customer adapted grading of Scots pine sawn timber using a multivariate method. *Scandinavian Journal of Forest Research*, 30(1), 87–97.

Breinig, L., Leonhart, R., Broman, O., Manuel, A., Brüchert, F. and Becker, G. (2015) Classification of wood surfaces according to visual appearance by multivariate analysis of wood feature data. *Journal of Wood Science*, 61(2), 89–112.

Broman, O. (2000) *Means to measure the aesthetic properties of wood*. Thesis (PhD). Luleå tekniska universitet.

Geladi, P. and Kowalski, B. R. (1986) Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185, 1–17.

Hagman, O. and Grundberg, S. (1995) Classification of Scots pine (*Pinus sylvestris*) knots in density images from CT scanned logs. *Holz als Roh- und Werkstoff*, 53, 75–81.

Kline, D. E., Surak, C. and Araman, P. A. (2003) Automated hardwood lumber grading utilizing a multiple sensor machine vision technology. *Computers and Electronics in Agriculture*, 41, 139–155.

Lance, R. F., Kennedy, M. L. and Leberg, P. L. (2000) Classification bias in discriminant function analyses used to evaluate putatively different taxa. *Journal of Mammalogy*, 81(1), 245–249.

Lycken, A. (2006) Comparison between automatic and manual quality grading of sawn softwood. *Forest Products Journal*, 56(4), 13–18.

Lycken, A. and Oja, J. (2006) A multivariate approach to automatic grading of Pinus sylvestirs sawn timber. *Scandinavian Journal of Forest Research*, 21(2), 167–174.

Oja, J., Wallbäcks, L., Grundberg, S., Hägerdal, E. and Grönlund, A. (2003) Automatic grading of Scots pine (*Pinus sylvestris L.*) sawlogs using an industrial X-ray log scanner. *Computers and Electronics in Agriculture*, 41, 63–75.

Oja, J., Grundberg, S., Fredriksson, J. and Berg, P. (2004) Automatic grading of sawlogs: A comparison between X-ray scanning, optical three-dimensional scanning and combinations of both methods. *Scandinavian Journal of Forest Research*, 19, 89–95.

Olofsson, L., Broman, O., Fredriksson, M., Skog, J. and Sandberg, D. (2017) Customer adapted grading of Scots pine sawn timber - a multivariate method approach. *Proceedings of International Wood Machining Seminar*, 23(1), 360–371.

Swedish Sawmill Managers Association (NTGR) (1994) *Nordic Timber: Grading Rules for Pine (Pinus sylvestris) and Spruce (Picea abies) Sawn Timber: Comercial Grading Based on Evaluation of the Four Sides of Sawn Timber* (Markaryd: Föreningen Svenska Sågverksmän).

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J. J., van Duijnhoven, J. P. M. and van Dorsten, F. A. (2008) Assessment of plsda cross validation. *Metabolomics*, 4(1), 81–89.

Wold, S., Sjöström, M. and Eriksson, L. (2001) PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.

PUBLICATION III

Multivariate Product Adapted Grading of Scots pine Sawn Timber for an Industrial Customer, part 2: Robustness to Disturbances

Authors and year:
Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D.
2019

Publication originally published in:
Wood Material Science & Engineering 14(6), 420–427
DOI: 10.1080/17480272.2019.1612944

WOOD MATERIAL SCIENCE & ENGINEERING
2019, VOL. 14, NO. 6, 420–427
https://doi.org/10.1080/17480272.2019.1612944

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Multivariate product adapted grading of Scots Pine sawn timber for an industrial customer, part 2: Robustness to disturbances

Linus Olofsson [a], Olof Broman[a], Johan Skog[b], Magnus Fredriksson[a] and Dick Sandberg [a]

[a]Wood Science and Engineering, Luleå University of Technology, Skellefteå, Sweden; [b]RISE Bioeconomy, Research Institutes of Sweden, Skellefteå, Sweden

ABSTRACT
Holistic-subjective automatic grading (HSAG) of sawn timber by an industrial customer's product outcome is possible through the use of multivariate partial least squares discriminant analysis (PLS-DA), shown by part one of this two-part study. This second part of the study aimed at testing the robustness to disturbances of such an HSAG system when grading Scots Pine sawn timber partially covered in dust. The set of 308 clean planks from part one of this study, and a set of 310 dusty planks, that by being stored inside a sawmill accumulated a layer of dust, were used. Cameras scanned each plank in a sawmill's automatic sorting system that detected selected feature variables. The planks were then split and processed at a planing mill, and the product grade was correlated to the measured feature variables by partial least squares regression. Prediction models were tested using 5-fold cross-validation in four tests and compared to the reference result of part one of this study. The tests showed that the product adapted HSAG could grade dusty planks with similar or lower grading accuracy compared to grading clean planks. In tests grading dusty planks, the disturbing effect of the dust was difficult to capture through training.

ARTICLE HISTORY
Received 12 March 2019
Accepted 23 April 2019

KEYWORDS
Sawn timber; visual grading; customer adoption; discriminant analysis

Introduction

Part one of this two-part study showed that it was possible to predict a customer's product grade outcome by holistically and subjectively grade sawn timber using partial least squares discriminant analysis. However, given a sawmill's dusty environment, it is important to know how a visual grading system is affected by disturbances. This second part of the study aimed at testing if a holistic-subjective automatic grading system is robust to disturbance in the form of a layer of dusty on top of the scanned sawn timber.

Method development (part 1 summary)

Automatic dry sorting stations at sawmills in Scandinavia use mainly objective rule-based automatic grading (RBAG) to sort sawn timber into standardized visual quality grades, e.g. (NTGR, 1994). RBAG is objective by nature, i.e. it uses different measurement rules (limits) to define grades. The problem with using rules to separate grades that motivated part one of this study (Olofsson *et al.*, 2019), as well as previous work by Lycken and Oja (2006); Berglund *et al.* (2015), and Olofsson *et al.* (2017), was that a large set of correlated grading rules had to be manually created by some expert for a coherent grading. Sawmill customers often have a holistic-subjective view of sawn timber quality, meaning they judge the whole piece at the same time and rules can be overridden based on the overall appearance of the piece. The large set of grading rules required, and the difficulty in defining them to holistically describe a customer's subjective definition of desirable quality sawn timber means that attempts to customize grading rules are seldom made. This problem is prominent when a sawmill is trying to make customized quality grades for costumers whose needs are not at all in line with the standardized grades, as this requires that big changes to the standardized grading rules have to be made. The collaborating industrial planing mill in this study is an example of this. The problem with customizing an RBAG system for a new holistic-subjective grade manifests itself primarily in two ways defined by Lycken and Oja (2006) as: (1) it is difficult for a customer to describe their subjective view of the desired sawn timber quality in a way that can easily be defined in objective grading rules, (2) the number of variables that can be controlled to specify a grade is often large enough to make customization complicated.

The problems with RBAG for a customer were addressed in part one of this study by implementing a holistic-subjective automatic grading (HSAG) method of Scots pine sawn timber with a collaborating sawmill and planing mill. Using the current sawmill hardware, an automatic scanning, grading, and sorting system was used to show that it is possible to grade sawn timber with HSAG according to the planing mills quality grade outcome. The grading method used was multivariate partial least squares discriminant analysis (PLS-DA), where prediction models were trained on aggregated feature measurements of each plank and the manually determined quality grade yield. This was done even though the planing mill split, planed, and milled the sawn timber.

CONTACT Linus Olofsson linus.olofsson@ltu.se

The tested grading models gave an overview of the grading accuracy level, and sensitivity to the random selection of test sets. The prediction models showed a stable behaviour, indicating that the PLS-DA can be suitable for HSAG. The HSAG was in part tested by 5-fold cross-validation to correctly grade 74% of the planks, in a data set of 308 planks. This grading accuracy can be compared to the 76–85% grading accuracy achieved in re-substitution tests in previous work by Berglund *et al.* (2015) and Lycken and Oja (2006) when grading sawn timber according to visual standard grades from NTGR (1994).

Grading results from a PLS-based prediction model are highly dependent on how accurately the test set is represented by the training set. Since the input to a sawmill is naturally heterogeneous wood, it is highly probable that the input material to the sorting station has never been experienced by the currently active grading model. The problem of having a training set that is representative of the test set is especially important in the context of systematic differences, due to e.g. difference in visual scanning conditions. This gave rise to the question of how robust a PLS-based HSAG grading method is to unforeseen disturbances (changes) of the input. This was tested in part two of this study project.

Robustness to disturbances

Some automatic dry sorting stations in Scandinavian sawmills grade sawn timber by visual appearance by objective rule-based automatic grading (RBAG). This is done by cameras, advanced feature detection algorithms, and a set of well-defined grading rules. Such grading rules defines feature limits, e.g. the maximum size of dead knots or maximum number of sound knots, and are often created to follow standards like NTGR (1994). Lycken and Oja (2006) stated that the hit-rate for defect classification in Nordic visual automatic grading systems was around 70–80%, which resulted in approximately 80–90% plank-grade hit-rate, depending on grading rules and material. Part one of this study graded sawn timber using the same hardware and feature detection but with a different grading methodology. Instead of defining a set of grading rules, partial least squares discriminant analysis (PLS-DA) models were used to grade the sawn timber in a holistic and subjective way, i.e. grading the entire piece at the same time to match a subjective quality assessment. The subjective quality assessment of the sawn timber was the product quality yield of an industrial planing mill customer that manually graded the product outcome as desirable or undesirable. The PLA-DA model graded the sawn timber with a grading accuracy of 74%, based on 5-fold cross-validation. Because a PLS-based holistic-subjective automatic grading (HSAG) model has to be trained on a known data set before use, called training set, it is important to know how the system would grade sawn timber which is not systematically similar to the training set. Lycken and Oja (2006) studied the grading outcome of PLS grading models when grading sawn timber to conform with manual grading with grading models trained and tested on material of the same size, either 50 by 200 mm or 50 by 150 mm in cross-section dimensions. However, training a grading model on planks of dimensions 50 by 200 mm while using the model to grade planks of dimensions 50 by 150 mm resulted in a grading accuracy drop from 85% to 59%, and a drop from 80% to 70% in the reversed scenario. This change of plank dimensions was a systematic difference that drastically lowered the grading accuracy of the grading model. Systematic changes in input can also be unforeseen disturbances, e.g. sawdust on the camera lens. In the present study, a layer of dust on the sawn timber was the unforeseen disturbance in the grading process that was tested. The dust layer changed the input to the HSAG system since the condition for the feature detection was altered and the feature detection error rate was assumed to increased, which in turn should affect the grading outcome. Part two of this study aimed to investigate the robustness of PLS-DA models to disturbances in the detected features due to a layer of dusty on top of the sawn timber. The effect of the dust on the grading outcome was investigated to answer two questions.

(1) How a PLS model trained on clean planks react to the changes in input when grading dusty planks. The dust layer changed the visual conditions for the cameras which affected how, or if, features were detected and classified. The possibly misclassified features were in turn the altered input to the PLS model. This can be expected to affect the grading outcome based on how potential detection errors were interpreted by the PLS model. The first objective was tested by training a PLS model on clean planks and grading dusty planks.

(2) How the prediction accuracy of a PLS model trained on a data set including the dusty planks relates to the model trained only on clean planks. Trying to account for unforeseen disturbances by training the prediction model on a data set as diverse as possible could strengthen the ability of the PLS model to grade any kind of planks but with a potential consequence of losing grading accuracy when grading the undisturbed clean planks. The second objective was investigated by training PLS models on both clean planks and dusty planks, and then testing the models by grading test sets of mixed planks, clean planks, and dusty planks, respectively.

Apart from grading accuracy, the selected thresholds were compared to the respective optimal threshold values. Since the threshold was chosen based on the training set, the selected threshold might be far from its optimal value if the training set does not represent the test set due to the disturbance of the dust.

Materials and methods

The collection of scanner data from the sawmill, and customer response from the planing mill; the construction of regression components; the implementation of PLS prediction models; and the testing methodology followed the same procedure as in part one of this study (Olofsson *et al.*, 2019). The scanner, grader, and sorting system used was the same Boardmaster by FinScan (Anon, 2018) at Kåge Sawmill. Dialogue with FinScan ensured that the scanner data of the two batches were comparable. Lundgren's planing mill was again the industrial customer with the same product and production process as in part one of this study.

Material

Two data sets were used: (1) the 308 planks used in part one of this study, referred to as the clean data set; and (2) a new set of 310 planks with the unforeseen dusty disturbance, referred to as the dusty data set. Both sets of planks originated from the same large sawing batch, so that there were no differences in sawing or handling conditions, except for that the second, dusty, data set had been stored uncovered indoors at the sawmill for one year before being scanned.

The dusty data set consisted of 310 Scots pine (*Pinus sylvestris* L.) planks, sawn from top-logs from the sawmill's log yard sorting station. Each log was cant sawn, and the centre yield of two planks was used for the study. The planks measured 50 by 150 mm in cross-section and varied between 3.4 and 5.5 m in length. The planks were dried to 14% moisture content before being stored and eventually scanned at the sawmill.

During storage, the planks were exposed to the dusty environment which left dust stains especially on the ends of the planks, and with visible marks from the spacers. Example of planks from the two data sets are shown in Figure 1. Almost every plank in the second data set had one dusty side and one clean side, as the dust settled on top of each plank, but some few planks were clean altogether or dusty on both sides. The plank orientation was considered random, i.e. the dust settled randomly on either plank face. The Boardmaster was calibrated as in part one of the study as if the dust was not present.

Data collection

The ID marked sawn timber was scanned by the Boardmaster and feature variables regarding knots and bark were saved before the timber was delivered to the planing mill. Other feature variables detected by the Boardmaster, such as wane or cracks, were ignored in this study as knots and bark features are critical for the customer and suitable for holistic-subjective grading. The ignored variables were ignored by the planing mill as well. Each plank was split into three boards which were each planed, milled, and manually graded – each board given grade A for desirable or grade B for undesirable for the planing mill. Each plank ID was associated with a digital label according to the majority of board grades produced, to study borderline cases, i.e. planks were labelled AAA, AA, BB, or BBB – omitting a B or A from the mixed labels AAB and BBA to keep labelling clean. Each plank was given a grade A or grade B based on the majority of grade A or grade B boards produced, i.e. AAA and AA was given the grade A for desirable, and BB and BBB planks the grade B for undesirable.

The quality distributions of the two data sets are given in Table 1, which showed that the two data sets had very similar proportions of grade A planks, with a slightly larger proportion labelled AAA in the dusty data set.

Applying PLS regression

Below is a brief overview of the implementation of PLS described, for a more thorough explanation, see Olofsson *et al.* (2019) – especially Table 1 and Figure 1.

Using the knot and bark measurements from the Boardmaster, which provides size, position, and defect type, a detailed set of 3564 aggregated variables were created to expand the ability of the HSAG system to capture, in objective measurements, the subjective quality traits desired by the customer. Each of the 22 created variables was measured for 6 defect types, separately for each of the plank's 3 faces (edges together), and separately measured in 9 different sections of each plank which was divided into 1, 3, and 5 sections in the lengthwise direction of the plank. In total, $22 \cdot 6 \cdot 3 \cdot 9 = 3564$ variables were created.

Using the aggregated feature variables from the Boardmaster and the planing mill quality grade assessment, both data sets could be used to create an explanatory matrix $\mathbf{X}$ (up to 618 by 3564) and a response matrix $\mathbf{Y}$ (up to 618 by 1). Using the SIMCA 14 software (Anon, 2019), these matrices were used to train PLS regression models based on different training sets, which were then used to predict the quality outcome of a series of test sets.

Based on new measurements of a test plank, a PLS prediction model predicts a value approximately between 0 and 1

Table 1. The number of clean and dusty planks in each data set, shown with proportion of each label and plank grade.

			Clean set			Dusty set		
		Label	Number	Proportion		Number	Proportion	
Plank grade	A	AAA	126	41%	65%	147	47%	65%
		AA	73	24%		55	18%	
	B	BB	43	14%	35%	39	13%	35%
		BBB	66	21%		69	22%	
		Totals	308	100%		310	100%	

Figure 1. Examples of three clean planks of data set 1 (top), and three dusty sides of planks in data set 2 (bottom). The bottom planks show clear dust shades between the clear marks caused by the spacers. The number written on each plank is the unique ID-number of each data set and the numbers 8 and 9 repeats by chance.

which can be thought of as a probability of that plank being of grade A. A threshold value decides what the prediction model classifies as of grade A or not. The class-separating threshold of each prediction model is determined by choosing a threshold that optimally separates the classes A and B in the training set of that prediction model.

Testing specifications

From part one of this study, the average grading results of the 5-fold cross-validation of the clean data set served as the reference case. The grading accuracy of the PLS-based HSAG method was tested according to four user-case scenarios, and compared to the reference case (test 0), detailed in Table 2.

Six new PLS prediction models are created in part two of this study. One model was trained on the complete clean data set, and based on the training set the threshold 0.56 was selected. This model was used in test (1) to test the discriminant model's robustness to unforeseen disturbances (research question (1) presented in the introduction). Five more prediction models were created for 5-fold cross-validation of both the clean and the dusty data sets together. The benefit of cross-validation over a single training and test set is that the average behaviour is less test set sensitive, which is necessary due to the high variability and complexity of the data sets used in this study. Each of the five models was tested on a unique fifth of the combined data set, and trained on the rest. The test sets were proportionally randomly selected from both the clean and dusty data sets, as well as proportionally selected from each label class. For each model a threshold was determined from the training set, which on average was 0.56[1] and the average prediction result was calculated. The same average prediction results of the 5-fold cross-validation was used in tests (2–4), where tests (3–4) are a split of test sets from test (2). Tests (2–4) share the same prediction results and the same 5 prediction models for the cross-validation, the results of test (2) are separated into tests (3) and (4) to show differences in grading outcome between the different test sets. These tests were performed to test the average grading accuracy of the discriminant models when taking the dust into account by training the prediction models on a mixed set of planks (research question (2) in the introduction).

Table 2. Test setups and objectives.

#	Training set	Testing set	Test objective
0	Clean	Clean	The reference use-case where the sawmill graded clean planks by a prediction model trained on clean planks.
1	Clean	Dusty	To test the effect of the dust by grading dusty plank with a prediction model trained on clean planks.
2	Clean and dusty	Clean and dusty	To test the behaviour of a prediction model that was trained to take unforeseen disturbances into account by training the prediction model on both clean and dusty planks, here grading a set of mixed clean and dusty planks.
3	Clean and dusty	Clean	To test the behaviour of a prediction model that was trained unnecessarily to take unforeseen disturbances into account when grading only clean planks.
4	Clean and dusty	Dusty	To show the difference in grading outcome compared to test (1) when grading dusty planks when the prediction models were trained on a training set including dusty planks.

The robustness of a grading model to unforeseen disturbances was, apart from grading accuracy, also investigated by studying the used threshold compared to the optimal threshold value for the different tests. Selecting a threshold based on a training set that is not is not representative of the intended test set could result in a lowered grading accuracy and was therefore also investigated. The maximum grading accuracy of e.g. test (1) could be the same as for the reference model (test 0) but at a different threshold than for the reference case. In part one of this study, the average grading accuracy of the 5-fold cross-validation was 74% at the average class-separating threshold 0.55, compared to the average peak grading accuracy of 76% at the optimal threshold 0.52. Compared to the optimal, this is a loss of 2 percentage points of grading accuracy and a 0.03 points miss of the optimal threshold value. Similar comparisons will be made for each test in part two of this study. This loss of grading accuracy and miss of optimal threshold is due to the fact that the tested model was not optimized for the current test set. Furthermore, the dust layer was expected to lower the grading accuracy, especially in test (1).

Results

To evaluate the effects of unforeseen disturbances on the accuracy of PLS-based HSAG, the results of several prediction models were tested in tests 1–4. Prediction results were compared with the reference 5-fold cross-validation analysis from part one of this study, presented as test (0) in Table 3 for easy reference. The test specifications are detailed in Table 2 and the results of each test are presented in Table 3 using the corresponding selected threshold.

In test (1), a single prediction model was trained on the entire clean data set from part one of this study and was tested by grading the entire dusty data set from part 2 of this study. The prediction results are visualized in Figure 2 where the prediction results are shown for the selected class-separating threshold 0.56, and as test (1) in Table 3. In test (1), 73% of the dusty planks were correctly graded in comparison to the 74% correctly graded planks according to the reference case.

Figure 3 shows the grading accuracy for the four tests at thresholds between 0 and 1. The reference results from part one are shown as well. For test 1 the grading accuracy curve was close to constant (approximately 71%) for a wide range of thresholds (approximately 0.3–0.8) before dropping off, while the other tests all showed similar behaviour as the reference test for all thresholds.

Optimal grading accuracy and threshold

The optimal grading accuracy of a prediction model could in hindsight be found at some optimal class-separating threshold for each test. A small loss of grading accuracy

Table 3. Misclassification table (confusion matrix) of the reference prediction results (test 0, (Olofsson *et al.*, 2019)), and for all the tests performed (tests 1–4), showing the grading outcome of each respective test for each plank-label class, using the selected thresholds. Test (0) used the average class-separating threshold 0.55, test (1) used the class-separating threshold 0.56, and tests 2–4 used the average class-separating threshold 0.56. Predicted plank grades were determined at the sawmill, and the observed grades were the grade outcome of each plank at the planing mill. Tests 0, 2, 3, and 4 are based on 5-fold cross-validation and shows the average values. Predicted numner of planks are rounded to whole number of planks.

				Predicted		Grading accuracy		
		Label	Number	A	B			
Test 0 Observed	A	AAA	24	21	3	87%	83%	
		AA	14	11	3	76%		
	B	BB	8	5	3	40%	59%	
		BBB	13	4	9	71%		
		Totals	59	41	18			74%
Test 1 Observed	A	AAA	147	137	10	93%	83%	
		AA	55	31	24	56%		
	B	BB	39	21	18	46%	54%	
		BBB	69	29	40	58%		
		Totals	310	218	92			73%
Test 2 Observed	A	AAA	53	46	7	86%	81%	
		AA	25	17	8	69%		
	B	BB	16	9	7	45%	61%	
		BBB	26	8	18	71%		
		Totals	120	79	41			74%
Test 3 Observed	A	AAA	24	22	2	91%	85%	
		AA	14	10	4	74%		
	B	BB	8	4	4	48%	64%	
		BBB	13	3	10	74%		
		Totals	59	40	41			77%
Test 4 Observed	A	AAA	29	24	5	82%	77%	
		AA	11	7	4	62%		
	B	BB	8	5	3	43%	58%	
		BBB	13	4	9	68%		
		Totals	61	40	22			70%

compared to the optimal was expected as the threshold selection was not optimized for the test sets. The grading accuracy of the reference test (0) is shown in Table 4 to miss out on 2 percentage points of grading accuracy by using a threshold 0.03 points away from the optimal class-separating threshold. The grading accuracy of the prediction model(s) of each test, and their respective (average) thresholds, are compared to the optimal setting in Table 4.

Discussion

The present study showed that the HSAG system used was robust to, but not unaffected by, unforeseen disturbances in the form of a dust layer on the sawn timber. As in part one of this study, it was evident that when the sawmill and customer discuss which threshold to use to reach satisfactory grading outcome, they need to take grading accuracy as well as grading outcome into account. For example, the grading and sorting outcome from the sawmill using a very high class-separating threshold, say 0.8, would be a very small batch of almost entirely grade A planks with almost no incorrectly graded B-grade planks in the delivered batch to the planing mill, but at a great cost for the customer due to the low volume (see Figure 5 in part one of this study). For this reason, a threshold close to 0.5 is preferable for a balance of sawmill and customer satisfaction, as well as being close to the optimal grading accuracy. The prediction results of all prediction models are conceptually visualized by Figure 2.

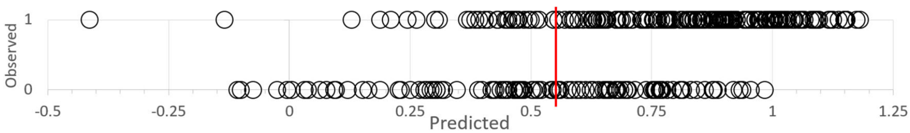


Figure 2. Observed vs predicted plot for test (1), where the dusty planks were predicted by the prediction model trained on the clean data set. The observed axis shows the plank grade from the planing mill while the predicted axis shows the predicted plank grade value from the prediction model used at the sawmill. Grade A planks are represented by 1 and grade B planks by 0. The grade separating threshold 0.56 is shown as a vertical line which defines planks with a predicted value above the threshold as of grade A, otherwise B. These prediction results are conceptually similar for all models used.

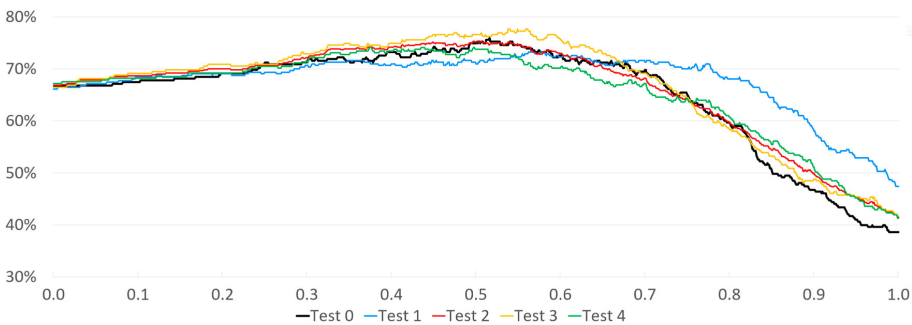


Figure 3. The grading accuracy as a function of threshold for the prediction models in tests; (0): trained on clean data and predicting clean data, from part one of this study; (1): trained on the entire clean data set and grading the entire dusty data set; (2): trained both data sets and predicting both data sets; (3): trained on both data sets and grading the clean data set; and (4): trained on both data sets and grading the dusty data set.

Table 4. Comparison of the used threshold and optimal threshold and the corresponding grading accuracy of each test, also showing the loss of grading accuracy and difference of threshold compared to the optimal setting. Tests 0, 2, 3, and 4 are based on 5-fold cross-validation and shows the average values.

Test	Accuracy	Optimal Accuracy	Used threshold	Optimal threshold	Accuracy loss (points)	Threshold difference
0	74%	76%	0.55	0.52	−2%	0.03
1	73%	73%	0.56	0.56	0%	0
2	74%	75%	0.56	0.50	−1%	0.06
3	77%	78%	0.56	0.54	−1%	0.02
4	70%	74%	0.56	0.50	−4%	0.06

In tests 1–4 (Table 3), the grading accuracy was similar to that of the reference-case test (0), but there was a tendency for the dust layer to affect the PLS-based HSAG results negatively. As in part one of this study, the label classes with higher number of planks in the data sets, shown in Table 1, were overall graded with a higher grading accuracy, which can be seen by comparing the number of planks in each label class with the corresponding grading accuracy in Table 3. Grade B planks were graded with the lowest grading accuracy, which could be attributed to the disproportionate number of grade B planks in the data sets. Table 1 show that 35% (approximately 110 pieces in each data set) of the planks are of grade B. A data set with a proportional number of planks per label-class might be desired, unless an increased grading accuracy of grade B planks comes with an undesirable trade-off of reduced grading accuracy of grade A planks. The lower grade B grading accuracy could also be a consequence of using a data set with few grade B planks, as Lycken and Oja (2006) estimated that a minimum of 100 planks per plank grade are required for PLS-based grading of sawn timber by visual grades, indicating that more grade B planks might be required for PLS-based product adapted grading.

In test (1), the prediction model had been trained only on clean planks, and the model graded dusty planks with similar accuracy to grading clean planks. The grading accuracy dropped 1% point from 74% to 73% at the selected thresholds 0.55 and 0.56, respectively (Table 3). This test simulated a real world scenario where, based on the clean data from part one of this study, a prediction model was created and used at the sawmill. The model was used to grade dusty plank and the PLS-based HSAG system would have functioned as expected but with a 1% point lower grading accuracy than the reference test (0). Due to the fact that the prediction model was used to grade planks with unforeseen disturbances, which it had not been trained for, this test was assumed prior to testing to show the worst grading performance, which was not the case. Test (1) showed the lowest grading accuracy for thresholds close to 0.5 but showed much higher grading accuracy than in the other tests for thresholds above 0.7. This could be because this is the only test that completely separates the two data sets for training and testing, making the separation of the classes more overlapping than in the reference test (0) (see Figure 2 in Olofsson *et al.* (2019)). The grading accuracy of test (1) means that the sawmill and customer could have decided on a very high threshold, up to say 0.8, without great loss of grading accuracy as in the other tests, however this test did not represent a common use-case.

In test (2), the prediction model was trained on a mixed set of both clean and dusty planks and graded a mixed data set by 5-fold cross-validation. The grading scenario in test (2) was close to the reference case, where the prediction model was trained and tested on similar data, i.e. clean, and mixed planks respectively. The grading accuracy curve for test (2) shown in Figure 3 was very similar to the curve for the reference test (0). This indicated that the sawmill should train the prediction model on data that is as representative of the intended grading batch as possible, dusty or not, as a mixed-trained model retained the grading accuracy compared to the reference test (0) when dusty planks are part of the test set. At the selected threshold 0.56, the grading accuracy increased from 73% to 74% compared to the reference, and since the prediction results are based on cross-validation on both the clean and the dusty planks, it was expected to see similar behaviour as the reference cross-validation of the clean data set.

In tests (3) and (4) the same grading models as in test (2) graded clean planks and dusty planks separately. Table 3 and Figure 3 showed that the grading accuracy was higher when grading clean planks than dusty planks. At the threshold 0.56, the mixed-trained model graded 77% of clean planks correctly, compared to 70% when grading dusty planks, compared to the 74% of correctly graded clean planks of the reference test (0). As expected, the model graded clean planks better than dusty planks. In Figure 3, the grading accuracy curves of test (3) and (4) behaved very much like expected, as they are the separation of test (2) into clean and dusty test sets separately; test (3) showed a higher grading accuracy than test (2) for almost all thresholds, while the opposite is true for test (4). Comparing tests (0) and (3) indicated that the mixed-trained model can be expected to perform equally or slightly better when grading clean planks than a model trained only on clean planks, and the mixed-trained model managed the highest prediction accuracy of all tests in test (3) (Table 4). One argument for the higher grading accuracy of the mixed-trained model when grading clean planks was the benefit of the larger training set, despite half of the training set being the dusty data set. However, as the dusty data set was not completely covered in dust, as almost all planks were only exposed to the settling dust on the top-face, there was a net positive effect of training on the dusty data set as well when grading clean planks. Comparing tests (1) and (4) indicated that the use of a mixed-trained model performed slightly worse when predicting the dusty planks, i.e. it was not possible in these tests to train a prediction model to take the dust into account for a higher grading accuracy of dusty planks. Comparing tests (0) and (3) showed that it was possible to train a prediction model on a larger data set and achieve higher grading accuracy when grading clean planks.

The grading accuracies were slightly higher for the models grading clean planks (tests 0 and 3) than for the models grading dusty planks (tests 1 and 4). Furthermore, prediction model trained on the combined data sets (test 2) did not grade planks more accurately than the reference model. This indicates that regardless of what data set(s) the prediction model is trained on, the dust on the planks introduces some difficulties. This was expected as the Boardmaster was calibrated for clean planks, and the layer of dust was bound to affect the feature detection in some way. For the sawmill, these tests indicated that the prediction model used should be trained on a data set that is as large as possible and as representative of the intended grading batch as possible. This is especially true when grading batches of sawn timber that is not expected to be dusty.

According to Table 4 the threshold for test (1) was by chance selected optimally, which can be explained by the flat grading accuracy curve in Figure 3 for thresholds between 0.4–0.6. In Figure 3 the maximum difference of the grading accuracy between test (0) and (1) was 4% points at the threshold 0.51, indicating that the loss of prediction accuracy due to the dusty could have been larger than the measured 1% point.

The robustness of PLS-based HSAG was, apart from the grading accuracy, investigated by measuring the optimal threshold stability, seen in Table 4. The threshold 0.56 was in both test (1) and tests (2–4) purposely chosen to imitate the use of a PLS-based HSAG system at the sawmill, with the goal of maximizing grading accuracy. However, as the prediction models tested were not optimized for the tests set there is an optimal threshold that can only be determined with known grade outcome at the planing mill. If the training data is representative of the test set, the threshold selected based on the training data should be close to the optimal threshold value. Any large changes in grading accuracy or large miss of optimal threshold might indicate that the prediction model is not well trained for the current test set, i.e. the training data is not fully representing the test data. Table 4 shows that in the reference case (test 0) the average threshold value of the 5-fold cross-validation is 0.03 points away from the optimal threshold, which resulted in a 2% points loss of grading accuracy. Given the size of the clean data set and complexity of the data used, this kind of variation was expected and is the reference case for optimal threshold stability. In test (1) the clean training data was not expected to be completely representative of the dusty testing data, but the threshold selected based purely on the clean training data was (surprisingly) the optimal threshold (Table 4), which can be explained by the approximately flat grading accuracy curve in Figure 3 for thresholds between 0.3 and 0.8. Anecdotally, in Figure 3 the maximum difference of the grading accuracy between test (0) and (1) was 4% points at the threshold 0.51, indicating that the loss of prediction accuracy due to the dusty could have been larger than the measured 1% point with a slightly different test set.

The threshold value difference between the used and optimal value for tests with dusty planks in the test set (tests 1, 2, and 4) showed higher differences (0.06 points), or the above mentioned optimal threshold ambiguity in test (1), compared to the tests with only clean plank in the test set. The larger threshold difference indicated that when predicting dusty planks the training data was not fully representative of the test set, no matter the training set. In tests (0) and (3) the threshold value difference was lower, at 0.03 and 0.02 points, which indicated that when grading clean planks the training sets were more representative of the test sets, no matter the training set. For a user of a PLS prediction model, it is very important that the training data is representative of the test data, otherwise grading outcome estimated from the training data might not transfer well to the actual grading outcome of a test or when grading for a customer. The optimal threshold stability measurements showed that when grading dusty planks, slight changes in grading outcome might occur unexpectedly.

Conclusions

The PLS-based HSAG of Scots Pine sawn timber investigated in this study was robust to unforeseen disturbances in the form of a layer of dusty on top of the sawn timber when grading planks for an industrial customer. Prediction models trained and tested on clean planks served as reference and predicted on average the plank grade of new clean planks correctly 74% of the time. The prediction accuracy dropped 1% point to 73% when a similar model was used to grade dusty planks, which indicated that PLS-based product adapted HSAG was robust to the disturbances of the dust. However, the class-separating threshold used (0.56) was seemingly selected optimally by chance for this test, and the loss of prediction accuracy could have been up to 4% points for a different threshold in the range 0.4–0.6.

Prediction models trained on both clean and dusty planks performed on average as good or better than the reference models when grading only clean planks, with on average 77% correctly graded clean planks. This indicated that the prediction model should be trained on a data set as large as possible consisting of planks as representative of the planks to be graded as possible, even if they are dusty on one side. It was, however, not possible in this study to improve the grading accuracy of dusty planks by training the prediction model on dusty planks as well as clean planks. Further research should investigate if separate grading methods can detect disturbances of the sawn timber measurements and handle the identified pieces accordingly.

The class-separating thresholds used in the study were selected based on the assumption that the training data was representative of the test data of each test. When the dusty planks were graded in any of the tests, the estimated grading outcome determined by the training data did not completely translate to the grading outcome of the tests, which indicated that the dust introduced some difficulty when selecting a class-separating threshold, as well as lowering the overall grading accuracy. Further research is required to fully understand how, and if, unexpected input to the PLS-based HSAG system should be handled. A larger data set with more extreme disturbances would be preferred for such a study.

Note

1. The threshold 0.56 is the same for test (1) and tests (2–4) by chance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Financial support from the Swedish Innovation Agency (Vinnova), project Sawmill 4.0 Customized flexible sawmill production by integrating data driven models and decisions tools 2018-02749, is gratefully acknowledged.

ORCID

Linus Olofsson http://orcid.org/0000-0002-5562-5142
Dick Sandberg http://orcid.org/0000-0002-4526-9391

References

Anon, (2018). Boardmaster. Accessed 12 June 2018, available at: https://finscan.fi/products/boardmaster/?lang=en. Note: older predecessor to BoardmasterNOVA.

Anon, (2019). Simca. Accessed 03 March 2019, available at: https://umetrics.com/products/simca.

Berglund, A., Broman, O., Oja, J. and Grönlund, A. (2015) Customer adapted grading of Scots pine sawn timber using a multivariate method. *Scandinavian Journal of Forest Research*, 30(1), 87–97.

Lycken, A. and Oja, J. (2006) A multivariate approach to automatic grading of Pinus sylvestirs sawn timber. *Scandinavian Journal of Forest Research*, 21(2), 167–174.

Olofsson, L., Broman, O., Fredriksson, M., Skog, J. and Sandberg, D. (2017) Customer adapted grading of Scots pine sawn timber – a multivariate method approach. *Proceedings of International Wood Machining Seminar*, 23(1), 360–371.

Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019) Multivariate Product Adapted Grading of Scots Pine Sawn Timber for an Industrial Customer, Part 1: Method Development. *Wood Material Science and Engineering*.

Swedish Sawmill Managers Association (NTGR). (1994). *Nordic timber: grading rules for pine (*Pinus sylvestris*) and spruce (*Picea abies*) sawn timber: comercial grading based on evaluation of the four sides of sawn timber* (Markaryd: Föreningen Svenska Sågverksmän).

Publication IV

New Possibilities with CT scanning in the Forest Value Chain

Authors and year:
Olofsson, L., Möller, C.-J., Wendel, C., Oja, J. and Broman, O.
2019

Publication originally published in:
Wang, X.; Sauter, U. H.; Ross, R. J., (Eds.) Proceedings, 21st international nondestructive testing and evaluation of wood symposium. General Technical Report FPL-GTR-272. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 569–576
URN: urn:nbn:se:ltu:diva-75506

New possibilities with CT scanning in the forest value chain

Linus Olofsson*
Wood Science and Engineering, Luleå University of Technology, Skellefteå, Västerbotten, Sweden, linus.olofsson@ltu.se

Carl-Johan Möller
Department of Physics, Umeå University, Umeå, Västerbotten, Sweden, mollercarlj@gmail.com

Charlotta Wendel
Department of Physics, Umeå University, Umeå, Västerbotten, Sweden, charlottawendel@hotmail.com

Johan Oja
Technical Manager, Norra Timber, johan.oja@norra.se

Olof Broman
Wood Science and Engineering, Luleå University of Technology, Skellefteå, Västerbotten, Sweden, olof.broman@ltu.se

* Corresponding author

Abstract

Industrial high-resolution X-ray computed tomography (CT) scanners have recently been installed at several sawmills worldwide for the description of roundwood interior features and external log shape. These CT scanners represent a technological advancement for sawmill businesses that open a way to higher volume and value yields and new production planning strategies. This paper will present an indicative study of innovative use of non-destructive CT log data in a Swedish softwood sawmill, linking high-quality information of the wood material along the wood-value chain. Sawn timber was observed throughout the sawmill process line, i.e. from the log yard through the sawmill process until grading after the timber was dried. Before sawing, the CT scanner scanned the logs and calculated knot measurements from the 3D CT log data of simulated value-optimized center yield. A corresponding set of knot measurements were later calculated from the camera-based grading of the dried timber. Only considering knots from the two sets of measurements, the sawn timber was automatically given a quality assessment based on CT data, by camera-based scanning data, and by manual visual grading for reference. Partial least squares regression was used to create prediction models by correlating the two sets of knot measurements with the automatically determined grade from the dry-sorting. The prediction models tested increased the grading consistency between the grading based on CT data of virtual planks and based on camera data of the same planks. Furthermore, a traceability algorithm was tested as a tool to generate large data sets for future studies.

Keywords: sawn timber grading, computed tomography, partial least squares

Introduction

Sawmills perform several quality assessments and product classifications along the entire sawline process—e.g. classifying the incoming logs (roundwood) by size, and the outgoing sawn timber by visual appearance or strength. A recent high-technological quality-assessment tool added to the sawline at Sävar Sawmill in northern Sweden is a high-resolution X-ray computed tomography (CT) scanner from Microtec (CT Log), which is used to create 3D CT images of the incoming logs. Giudiceandrea et al. (2011), and Longuetauda et al. (2012) presented both an overview and an insight in the difficulties and requirements of implementing an industrial CT-scanner in sawmills, capable of continuous high-speed operation, but also the potential benefits of increased value and volume yields. Fredriksson et al. (2017) presented a state of the art of the CT-scanning technology in the wood value chain, showing additional possibilities of using CT-scanner technology in modern sawmills apart from increased value and volume yields. Fredriksson (2014) calculated that fully utilizing CT-data of logs to determine optimal sawing parameters increased the value yield on average by 13% compared to sawing positioning based on outer shape only, and by 21% compared to the horns-down sawing position, based on 269 logs.

Once incoming logs to the sawmill have been CT-scanned, some algorithm is used to determine how the log should be sawn. In the case of Sävar sawmill, detected features, especially knot features, in the 3D CT-image of each log is used to determine optimal sawing conditions, i.e. a sawing pattern and positioning parameters. The detected features are projected onto sawn planks in the virtual 3D CT-image and are used to determine the grade and value of each plank. The grade and value are determined by rule-based automatic grading (RBAG), following the grading standard Nordic Timber Grading Rules (NTGR) (Swedish Sawmill Managers Association 1994) used at Sävar sawmill. By testing a large number of possible sawing patterns and positioning parameters the intended optimal sawing conditions can be found. The sawing pattern and positioning parameters are then sent to the sawing system that cuts the logs as decided by the CT Log.

After sawing and drying, the sawn timber is sorted at the sawmill's dry-sorting station where the final grade of the sawn timber is determined. In the case of Sävar sawmill, this is done using a camera-based automatic scanning and grading system called Boardmaster from Finscan. The Boardmaster uses cameras to detect features, e.g. knot features, cracks, wane, and so on, to determine the visual grade of each plank by RBAG in accordance with NTGR. The NTGR governs all possible wood features, but only knot features will be considered in this study, due to their importance to the final appearance grade and the difficulty of detecting and describing knot features in a concise way.

The functionality of the CT Log when grading a virtual plank inside a scanned log is by design very similar to the visual grading of the same plank performed by the Boardmaster at the dry sorting station, i.e. the two systems use the same grading rules in their respective grading operations. The difference is that the CT Log compares a lot of different grading outcomes to find the best sawing conditions while the Boardmaster assigns the final grade. In theory, given identical input parameters from a single plank, the two systems should give the plank the same grade. However, in practice, the CT Log and the Boardmaster will never see the same identical plank as the virtual plank selected by the CT Log to be sawn is separated from the actual plank, as seen by the Boardmaster's cameras, by the sawing process. The sawing process can introduce errors such as rotation or positioning errors during sawing, which becomes more difficult as the speed of the sawing increases. The separation between the virtual plank and the actual plank is made worse by possible detection errors. Longuetauda et al. (2012) presented an algorithm to detect knots in CT-images, as well as an overview of previous work regarding automatic knot detection using CT-imaging, showing an example of the difficulty of accurately describing knot features in CT-images. Using cameras to accurately describe plank features is no effortless task either for the Boardmaster. The differences between the virtual plank and the actual plank can result in a disagreement between the intended plank grade given by the CT Log and the grade assigned by the Boardmaster.

The goal of this study was to utilize the high-quality information extracted by the CT Log when calculating the intended grade of a virtual plank in a 3D CT-image in a way that better agree with the grade assigned by the Boardmaster, as to minimize the effect of detection errors and errors during sawing on the final value yield. Furthermore, this study will investigate the use of a traceability algorithm of individual planks between the CT Log and the Boardmaster as a means to create large data sets in the future for similar studies where information from both the CT Log and Boardmaster is important.

This study suggests the use of multivariate partial least-squares regression (PLS) models (Geladi and Kowalski 1986), similar to the ones used by Lycken and Oja (2006), Berglund et al. (2015), Olofsson et al. (2017, 2019a, 2019b), and in related studies by Broman (2000) and Breinig et al. (2015). PLS regression models have in previous studies been shown to be noise-resistant and capable of creating prediction (grading) models base on a large number of highly colinear variables using a relatively small number of observations.

Materials and methods

The material

A total of 238 pieces of Scots pine (*Pinus sylvestris* L.) sawn timber was studied at Sävar sawmill in northern Sweden. The planks originated from logs that had previously been scanned using X-ray computed tomography by the CT Log, from Microtec (Anon 2019a), installed directly before the sawing system at Sävar sawmill. The CT Log fitted a sawing pattern with two 75x125 mm center-yield planks and calculated value-optimized rotational and positioning parameters to be used during the following cant sawing process. After drying the sawn timber to 18% moisture content, the material was scanned and graded at the dry-sorting station by the camera-based and rule-based automatic grading (RBAG) system Boardmaster from Finscan (Anon 2019b). Manual grading of the material was also performed and throughout this study, only knot measurements were considered.

PLS modeling

The PLS modeling performed in this study, for both the CT Log and the Boardmaster, is very similar to PLS modeling performed by Wendel (2019), who presents a similar study to this one but with a different approach. Similar PLS modeling has been performed by the references at the end of the introduction, most notably by Olofsson et al. (2019a), Berglund et al. (2015), and Lycken and Oja (2006), where Olofsson et al. (2019a) includes a very detailed explanation of how this kind of PLS-based grading was implemented for sawn timber grading by a Boardmaster at Kåge sawmill (also a Boardmaster, using the same grading rules). Conceptually, the same process is performed here for the sawn timber grading by both the CT Log and Boardmaster and below follows a short description of the specifics about this implementation. For a more detailed description of a similar implementation, as well as performance evaluation by 5-fold cross-validation, see Olofsson et al. (2019a).

Knot variables

The aggregated knots variables used in this study, created by Wendel (2019), are 23 variables, detailed in Table 1, calculated in each of the 9 zones shown in Figure 1, for each of the 4 plank faces, resulting in $23\times4\times9=828$ variables. The 828 aggregated knot variables were calculated separately for both the CT Log and the Boardmaster, based on measurements from each system.

Table 1—Description of the 23 aggregated knot-variables used in this study. Each X marks one variable with the corresponding description and knot-type.

Variable description	All knots	Sound knots	Dead knots
Total number of knots	X	X	X
Number of knots per meter	X	X	X
Maximum size in longitudinal direction (mm)	X	X	X
Maximum size in axial direction (mm)	X	X	X
Mean size in longitudinal direction (mm)	X	X	X
Mean size in axial direction (mm)	X	X	X
Covered area (%)	X	X	X
The proportion of knots compared to all knots (%)		X	X

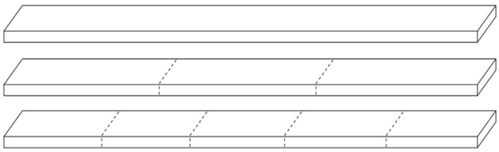


Figure 1—Zone partitioning of each plank. Each variable in Table 1 is calculated in each of the 9 zones shown.

Grading

Knot measurement data from both the CT Log and the Boardmaster was stored for each of the 238 planks. The grade of each virtual plank output by the CT Log—the intended sawing outcome grade—as well as the grade assigned to each plank by the Boardmaster was also stored. Manual grading of each plank was performed as a reference. Note that all grading was only based on knot features. Three different grades were used by each system and during the manual grading, called A, B, and C (Tabel 4, NTGR), where grade A represents a grade with few and small knots and high value and vice versa. Previous studies by Olofsson et al. (2019a), and Lycken and Oja (2006) estimated that approximately 100 planks per grade could be considered a minimum for training statistical grading models. To avoid underrepresenting the limiting number of 65 grade A planks (by manual grading) in a three-class classification problem, only the grades A and “other” were considered. In this study, the “other”-grade will be labeled grade B for a simple A vs B discussion, even though B here includes grade C.

To associate the correct CT-grade and Boardmaster-grade given to each plank with the corresponding ID number, a fingerprinting-traceability algorithm was used (Möller 2019). The study by Möller (2019) was performed in parallel to this study on 154 of the planks. The traceability algorithm used in the study by Möller (2019), called CPF, projected all knots detected by the CT Log and by the Boardmaster onto two one-dimensional binary vectors, creating a vector visually would look like a sparse bar-code. This bar-code-vector was then used as a fingerprint to match 96% of the 154 planks between the CT Log and the Boardmaster. This algorithm was used in this study to match 92% of the 238 planks to indicate the possibilities of using such a traceability algorithm, but in practice, all matches were manually checked to keep the CT Log-Boardmaster matching 100% for this study.

Both during the value-optimization process performed by the CT Log and during the final grading by the Boardmaster a customer-adapted implementation of the visual-appearance-grading-standard Nordic Timber Grading Rules (NTGR) (Swedish Sawmill Managers Association 1994) was used. This standard defines rules (limits) as the maximum number and size of plank features allowed in different grades. The standard defines rules for all types of wood features but this study will only use knot features, as knots are the main deciding factor that determines the visual grade of sawn timber. Knots are also of particular interest as a common belief in the sawmill industry is that customers tend to accept a few slightly larger than specified defects (knots) if the rest of the plank is better than the average of a particular grade (Berglund et al. 2015). The opposite is also true where customers might find a plank unsatisfactory if the overall appearance is not representative of a given grade, even though all features are within the specified limits.

PLS model construction and testing procedure

Two PLS models were created, called the CT-PLS model and the BM-PLS model for the CT Log and the Boardmaster, respectively. For each system, the corresponding set of aggregated knot variables were correlated to the A/B-grade determined by the Boardmaster by multivariate PLS regression. For the CT-PLS model, this means that the aggregated knot-variables from the CT Log was correlated to the A/B-grade set by the Boardmaster by traditional RBAG. For the BM-PLS model, this means that the aggregated knot-variables from the Boardmaster were correlated to the grade set by the Boardmaster by traditional RBAG. Once the two PLS-models were trained on the known data, the PLS models could be used to predict the grade of new planks with an unknown grade with some grading accuracy.

The performance of each PLS model was evaluated by 5-fold cross-validation, where the PLS model was trained on four-fifths of the available data and tested on the remaining fifth, using the RBAG grade from the Boardmaster. This is done to test the performance of the PLS model on data it has not been trained on. This procedure is repeated five times and the cumulative prediction results on the entire data set are presented as the PLS model's performance on this, limited, dataset.

Results

RBAG agreement comparison

To establish a baseline for comparison, the RBAG performed by the CT Log and the Boardmaster are compared to manual grading in Table 2, and to each other in Table 3.

Table 2—Misclassification table (confusion matrix) where the agreement of RBAG from both the Boardmaster (BM-RBAG) and the CT Log (CT-RBAG) is compared to manual grading, where the systems agree to 81% and 79% with the manual grading. Do note the Discussion section, regarding the low grading accuracy of grade A planks, and how these grading outcomes should not be interpreted as actual grading outcomes.

			BM-RBAG			**CT-RBAG**		
	Grade	Members	A	B	Correct	A	B	Correct
Manual	A	65	37	28	57%	24	41	37%
	B	173	17	156	90%	10	163	94%
	Total	238	54	184	**81%**	34	204	**79%**

Table 3—Misclassification table (confusion matrix) where the agreement of RBAG from the Boardmaster (BM-RBAG) is compared to the RBAG from the CT Log (CT-RBAG), showing a 77% agreement. Do note the Discussion section, regarding the low grading accuracy of grade A planks, and how these grading outcomes should not be interpreted as actual grading outcomes.

			CT-RBAG		
	Grade	Members	A	B	Correct
BM-RBAG	A	54	17	37	31%
	B	184	17	163	91%
	Total	238	34	204	**77%**

PLS-based grading agreement comparison

The grading results of introducing PLS-based grading in the CT Log is shown in Table 4, where the grading results are compared to the RBAG of the Boardmaster. Introducing PLS-based grading in both the CT Log and the Boardmaster resulted in the grading agreement shown in Table 5, where the grading results of the CT-PLS model is compared to the BM-PLS model, which is then compared to manual grading.

Table 4—Misclassification table (confusion matrix) where the agreement of RBAG from the Boardmaster (BM-RBAG) is compared to PLS-based grading from the CT Log (CT-PLS), showing an 82% agreement. The prediction results of the CT-PLS model is the cumulative prediction result of 5-fold cross-validation.

			CT-PLS		
	Grade	Members	A	B	Correct
BM-RBAG	A	54	41	13	76%
	B	184	30	155	84%
	Total	238	71	168	**82%**

Table 5—Misclassification table (confusion matrix) where the agreement of PLS-based grading from the Boardmaster (BM-PLS) is compared to the PLS-based grading from the CT Log (CT-PLS), and to manual grading, showing an agreement of 83% and 82%, respectively. The prediction results of the PLS models are the cumulative prediction result of 5-fold cross-validation.

			BM-PLS		
	Grade	Members	A	B	Correct
CT-PLS	A	70	54	16	77%
	B	168	25	143	85%
	Total	238	79	159	**83%**

			BM-PLS		
	Grade	Members	A	B	Correct
Manual	A	65	50	15	77%
	B	173	29	144	83%
	Total	238	79	159	**82%**

Discussion

This study has shown indicative results that the grading accuracy of the quality assessment systems CT Log and Boardmaster are high at their respective quality assessment segments of the sawline at Sävar sawmill, with 79% and 81% (Table 2) agreement with manual grading, respectively. However, due to e.g. detection errors and machining errors their grade assessment only agrees to 77% (Table 3), which can lead to problems in a few ways: if the CT Log misclassifies the grade of virtual planks in 3D CT-images the scanned log might not be sawn in an otherwise optimal way; if the Boardmaster misclassifies the sawn timber at the dry sorting station there can either be a loss of value or an unsatisfied customer; if both the CT Log and the Boardmaster classifies sawn timber correctly, a machining error in the sawing process might still lead to a reduced grade; which in each case shows the desirability of higher agreement between these two systems. By introducing PLS-based grading in the CT Log, the grading agreement with the Boardmaster increased from 77% to 82% (Table 4), and by introducing PLS-based grading in the Boardmaster as well, the agreements continue to increase to 83% (Table 5), while at the same time increasing the grading agreement between the Boardmaster and manual grading from 81% (Table 2) to 82% (Table 5). A high grading agreement between the two systems minimizes the effect of detection errors and machining errors on the total value yield of the grading process.

The results of this study are indicative for a few reasons. The small amount of available data limited the study to separate between two grades instead of three, and a combination of the two below described difficulties lowered the expected agreement between both the CT Log and the Boardmaster with the manual reference grade. Furthermore, the focus of this study was to maximize grading accuracy, which affects grading outcome.

The limitation introduced in this study to use only knot features is outside normal operations for the sawmill which affects the grading outcome. Furthermore, the sawn timber used in this study had to be handled by trucks and conveyors more than usual due to repeated runs through the dry sorting station to ensure intended functionality. The extra handing introduced more scuffs, dust, and oil-stains on the planks than usual, which resulted in a lower agreement between the Boardmaster's grading and the manual grade reference than expected. Möller (2019)—who developed the traceability algorithm used in this study, which involved partially the same material as in this study—showed in Figure 13 a set of planks that could be considered the worst-case scenario.

During the CT-scanning and sawing process of this study, an error occurred with the positioning mechanisms in the sawing system. This caused approximately 9% of the planks to be sawn with a rotational error above 10°, which is much higher than during normal operation. This rotational error could change on which face of the planks the knots inside the logs appear, which resulted in a lower agreement between the CT Log's grading and the manual grade reference than expected.

Conclusions

The indicative results of this study have shown that PLS-based grading increases the grading agreement compared to RBAG between the two grading systems CT Log and Boardmaster. This was achieved by introducing PLS-based grading in the CT Log, and by introducing PLS-based grading in the Boardmaster as well, the grading agreement between the two systems and the agreement between the Boardmaster and a manual reference grading either retained the same level of agreement or slightly increased it.

Each individual piece of sawn timber in this study required to be traced from the moment of the CT-scanning by the CT Log, throughout the entire sawline process, including drying, until the dry-sorting station where the Boardmaster scanned the same piece. This was done to 92% by a traceability algorithm and the remaining 8% was traced manually—showing the potential of creating similar studies like this one with the help of traceability algorithms.

Future studies should investigate the use of a traceability algorithm, like the one used in this study, to recreate this study on a new, larger, data set to further validate the findings made here.

Acknowledgments

Financial support from the Swedish Innovation Agency (VINNOVA), project Sawmill 4.0 – Customised flexible sawmill production by integrating data-driven models and decisions tools 2018-02749, is gratefully acknowledged. We thank all authors for their cooperation.

References

Anon (2019a) CT Log. Microtec. https://microtec.eu/ (Accessed: 2019-07-05).

Anon (2019b) Boardmaster. Finscan. https://finscan.fi (Accessed: 2019-07-05).

Anon (2019c) SIMCA. https://umetrics.com (Accessed: 2019-07-05).

Berglund, A.; Broman, O.; Oja, J.; Grönlund, A. 2015. Customer adapted grading of Scots pine sawn timber using a multivariate method. In: Scandinavian journal of forest research. 30(1): 87-97.

Breinig, L.; Leonhart, R.; Broman, O.; Manuel, A.; Brüchert, F.; & Becker, G. 2015. Classification of wood surfaces according to visual appearance by multivariate analysis of wood feature data. In: Journal of wood science. 61(2): 89-112.

Broman, O. 2000. Means to measure the aesthetic properties of wood. Doctoral dissertation, Luleå tekniska universitet. Available in DiVA diva2:989959 (Accessed: 2019-07-05).

Fredriksson, M.; Broman, O.; Sandberg, D. 2017. The use of CT-scanning technology in wood value-chain research and in wood industry: a state of the art. In: Pro Ligno. 13(4): 533-539.

Fredriksson, M. 2014. Log sawing position optimization using computed tomography scanning. In: Wood Material Science & Engineering. 9(2): 110-119.

Geladi, P.; Kowalski, B. R. 1986. Partial least-squares regression: a tutorial. In: Analytica chimica acta. (185): 1-17.

Giudiceandrea, F,; Ursella, E.; Vicario, E. 2011. A high speed CT scanner for the sawmill industry. In: Proceedings of the 17th International Nondestructive Testing and Evaluation of Wood Symposium, Sopron, Hungary.

Longuetauda, F.; Mothea, F.; Kerautretc, B.; Krähenbühlc, A.; Horyc, L.; Leband, J.M.; Debled-Rennesson, I. 2012. Automatic knot detection and measurements from X-ray CT images of wood: A review and validation of an improved algorithm on softwood samples. In: Computers and Electronics in Agriculture. 85(1): 77-89.

Lycken, A.; Oja, J. 2006. A multivariate approach to automatic grading of Pinus sylvestris sawn timber. In: Scandinavian Journal of Forest Research. (21)2: 167-174.

Möller, C.J. 2019. Development of Fingerprint Traceability in a Modern Sawmill. Master thesis, Umeå Universitet. Available in DiVA diva2:1305467 (Accessed: 2019-07-05).

Olofsson, L.; Broman, O.; Fredriksson, M.; Skog, J.; & Sandberg, D. 2017. Customer adapted grading of Scots pine sawn timber: a multivariate method approach. In: 23rd International Wood Machining Seminar (IWMS-23), Warsaw, Poland, May 28-31 2017 (pp. 360-361). Warsaw university of life sciences.

Olofsson, L.; Broman, O.; Skog, J.; Fredriksson, M.; Sandberg, D. 2019a. Multivariate product adapted grading of Scots pine sawn timber for an industrial customer, part 1: Method development. In: Wood Material Science & Engineering.

Olofsson, L.; Broman, O.; Skog, J.; Fredriksson, M.; & Sandberg, D. 2019b. Multivariate product adapted grading of Scots Pine sawn timber for an industrial customer, part 2: Robustness to disturbances. Wood Material Science & Engineering.

Swedish Sawmill Managers Association (NTGR). 1994. Nordic Timber: Grading Rules for Pine (Pinus sylvestris) and Spruce (Picea abies) Sawn Timber: Comercial Grading Based on Evaluation of the Four Sides of Sawn Timber. Markaryd: Föreningen Svenska Sågverksmän.

Wendel, C. 2019. Multivariate modeling improves quality grading of sawn timber. Master thesis, Umeå University. Available in DiVA diva2:1329008 (Accessed: 2019-07-07).

Publication V

Holistic-subjective Automatic Grading of Sawn Timber: Sensitivity to Systematic Changes

Authors and year:
Olofsson, L., Broman, O. and Sandberg, D.
2019

Publication originally published in:
Schajer, G. S. (Ed.) 24th International Wood Machining Seminar Proceedings: Corvallis, Oregon, Oregon State University, 2019, 157–164
URN: urn:nbn:se:ltu:diva-74170

Holistic-Subjective Automatic Grading of Sawn Timber: Sensitivity to Systematic Changes

Linus Olofsson[1], Olof Broman[1] and Dick Sandberg[1]

[1] Wood Science and Engineering, Luleå University of Technology, SE-931 87 Skellefteå, Sweden

ABSTRACT

Holistic-subjective automatic grading of sawn timber by a partial least squares (PLS) regression model required training of the model. This study tests the sensitivity towards systematic changes of a specialized PLS model, trained on a selected type of material suitable for a specific paneling product, when used to grade sawn timber systematically different than the material it was trained on. A sawmills automatic scanning system used cameras to measure knot and bark features on 900 planks. Each plank was split into three boards, and each board was shaped into an indoor paneling product and manually graded as desirable or undesirable at a planing mill. The plank grade was decided as the majority of the board-grade outcome. The knot and bark measurements were used to create a large set of feature variables for each plank that was correlated to the plank's grade by PLS regression. Of the 900 available planks, 434 planks sawn from top logs were used as a class-balanced specialized training set, with half of the planks resulting in a majority of desirable boards. The regression model trained on the class-balanced specialized training set was used to grade a test set of 282 planks, containing 64 planks that by manual classification of automatically captured images were determined to be sawn from butt logs and were systematically different from the training material. The PLS model's grading accuracy of the planks sawn from top logs was 76%, compared to 70% for the plank sawn from butt logs. The grading outcome resulted in a higher proportion of both delivered planks from the sawmill and received desirable planks by the planing mill when grading planks from top logs as compared to planks from butt logs. The results indicated that a specialized PLS model should not be used for a generalized use-case.

Keywords: sawmilling, wood, automatic grading, PLS-regression

INTRODUCTION

Automatic grading is the most common way of grading sawn timber in today's sawmills. For visual appearance grading of sawn timber, rule-based automatic grading (RBAG) by a camera-based system is a commonly used approach. This study is a continuation of the studies by Olofsson et al. [1, 2] and has the same parties involved, with the same industrial planing mill representing the customer. The collaborating sawmill uses visual RBAG to grade planks for delivery to the planing mill. For RBAG, a set of grading rules has to be defined to govern the grading outcome. Such rules are well defined to follow grading standards, such as the Nordic Timber Grading Rules [3], and controls e.g. the maximum allowed size and number of a specific feature for a specific grade. Appearance grading standards mostly specify rules regarding knots, as these are the main features affecting the general appearance of planks and will be the focus of this study. RBAG of standardized sawn timber has been used to a great extent so that standardized products are available at different sawmills. However, since not all customers are interested in the standardized grades, customization is sometimes required. As discussed in [1, 2, 4–6], RBAG can be intricate to customize due to mainly two difficulties, summarized by Olofsson et al. [1]:

(I) It is sometimes difficult for customers to describe their subjective view of the desired plank quality in a way that can easily be defined in objective grading rules.

(II) The number of feature variables that can be controlled to specify a grade is often more than enough to make customization complicated.

These customization difficulties incentivized the studies [1, 2, 4–6] to grade sawn timber by a holistic-subjective automatic grading (HSAG) method, i.e. a grading methodology that grades the entire piece subjectively.

In the works [1, 2, 4–6] and in this study, multivariate partial least-squares (PLS) regression [7, 8][1] was used to implement HSAG of Scots pine sawn timber. The works [4–6] used PLS regression based HSAG to grade sawn timber for the North-African market, while the studies [1] and [2] used the same method to grade sawn timber for an industrial planing mill customer in Sweden. The main benefit of using specifically PLS regression models for HSAG is its ease of use. The use of PLS-based HSAG also solves the two difficulties with customizing RBAG for a specific customer [1, 2, 4–6], since [1]:

(I) The customer needs only to specify whether or not they want each piece of sawn timber—no rule-based description is then necessary.

(II) The number of available parameters to be controlled is irrelevant as the HSAG uses multivariate PLS regression.

One introduced difficulty with using PLS-based HSAG is the requirement to train the multivariate regression model, and this requirement comes with some concerns regarding the material selection for training. The PLS model needs to find a balance between being specialized and generalized, where "specialized" meaning a PLS model trained on a specially selected type of material for a specific type of grading, and "generalized" meaning a PLS model trained on a diverse material meant for more robust general purpose grading. The problem with training a grading system to be either specialized or generalized is the difficulty of selecting training material that is as representative as possible of the material that the grading system is intended for. The planing mill customer that was the focus in [1] and [2], and will again be the focus in this study, preferably wants to purchase sawn material with mostly sound knots for its indented product. To provide suitable log material, the collaborating sawmill in the studies [1] and [2] pre-sorted the logs before sawing and trained a specialized PLS model to grade sawn material from top logs. A generalized alternative would have been to not pre-sort the logs and train a generalized PLS model on any sawn material of suitable dimensions. Test number 1 by Olofsson et al. [2] was made to test how a PLS model graded dusty plank when it was only trained on ordinary clean planks. Such a PLS model can be thought of as specialized since it was trained specifically to grade "clean" newly sawn and dried timber to a customer. The test exposed the specialized PLS model to a generalized grading test to grade dusty sawn timber. The test was meant to show how the PLS model performed when grading sawn timber with a disturbance, the dust, that the PLS model was not prepared for. The result was a slight reduction in grading accuracy that slightly worsened the sorting outcome. Similarly, this study will test how a specialized PLS model will grade sawn timber that it was not trained for by testing a PLS model on sawn timber systematically different from the sawn timber it was trained on. The difference between the two tests is that the test in this study is with regards to systematic differences while test number 1 by Olofsson et al. [2] was with regards to disturbances.

As the planing mill customer requires sawn timber originating from top logs with mostly sound knots for its intended product, sawn material from butt logs with mostly dead knots will be used as a systematically different test material. The objective of this study was to test the sensitivity of PLS-based HSAG towards systematic changes in the training and testing material, by training a PLS model on sawn timber from top logs and testing it on sawn timber from butt logs.

MATERIALS AND METHODS

The data collection, variables construction, PLS-grading implementation, and testing methodology in this study closely follows the procedures described in more detail by Olofsson et al. [1].

[1] See abstract of [7] for a quick overview of PLS regression and [1] for a more detailed and applied explanation.

The material

The material used in the study consisted of data from 900 Scots pine (*Pinus sylvestris* L.) planks. Selected logs were cant sawn at Kåge sawmill and resulted in a center yield of two planks of 50 by 150 mm in cross-section and between 3.4 and 5.7 m in length. The sawn timber was dried to 14% moisture content before being scanned at the sawmill's dry-sorting station, using a FinScan Boardmaster [9]. The sawn timber was delivered to Lundgren's planing mill where each plank was split into three boards and each board was shaped to an indoor paneling product, and manually graded as of a desirable grade A or an undesirable grade B. Each plank from the sawmill was given the average grade of its resulting boards after splitting and shaping, e.g. a plank resulting in two boards of grade A was given the grade A. All plank grades were stored in a response-grade vector y.

Data from the 900 planks was collected during three separate studies, 618 planks for the studies [1] and [2] and 282 for this study, resulting in three combined data sets. The 618 planks from studies [1] and [2] were sawn from pre-sorted top logs (with some error margin), based on 3D surface data and top diameter. Specifically, the log sorting was performed to select logs containing sound knots, which for Scots pine dominantly means top logs with some middle logs. For simplicity, these logs are generalized to all be top logs. The planks added to the total data set in this study consisted of 282 planks, of which 69 (24%) originated from butt logs. The data set collected for study [2] was the dusty data set mentioned in the introduction. Olofsson et al. [2] concluded that the potential negative effects of training on the dusty data set were outweighed by the benefits of training on a larger data set. Therefore, no further distinction will be made for the dusty data set.

Data collection

The 900 planks in this study were all ID marked before being scanned by a FinScan Boardmaster [9] at the dry-sorting station at Kåge sawmill. The Boardmaster automatic scanning and grading system can measure plank features such as knots, cracks, wane, bark inclusion, etc., and saves each feature's size, position, and classification. These measurements were the foundation of the HSAG and the feature variables used were therefore carefully selected with regards to the planing mill customer. Since knot features were important to the customer and especially difficult to describe subjectively by a set of grading rules, see points I and II in the introduction, features regarding knots were selected to be used for HSAG. As bark pockets were also important for the planing mill, bark features were also selected. The selected knot and bark features were stored and all other features were disregarded throughout the study—including during manual grading of the finished product at the planing mill. Since feature size, position, and classification is not sufficiently descriptive enough to describe the subjective quality outcome at the planing mill, the feature measurements were used to create a large set (matrix) of feature variables $\boldsymbol{X}$ by an additional software tool. Each row $\boldsymbol{x}$ represented a plank observation and each column represented a feature variable. The software tool calculated 22 variables, for 6 different defect types, on 3 faces of each plank—edges together, and in 9 different zones, for a total of $22 \cdot 6 \cdot 3 \cdot 9 = 3564$ feature variables, see Table 1 and Figure 1 by Olofsson et al. [1]. An example of such a feature variable could read "Sum of the defect area, of dead knots, on the inner face, in zone 3".

Butt log identification

Of the 900 planks available for this study, 618 planks were sawn from top logs and were available for training of prediction models. Of the remaining 282 planks, 69 planks were manually classified as sawn timber origination from butt logs. The butt-log origin of the sawn timber was determined by identifying planks with characteristics of butt logs, using images automatically captured by the Boardmaster during the data collection at the dry sorting station, see Figure 1. The images were created by an external software tool that combined 3 camera images per plank face. One typical knot-characteristic of butt-log planks is a large proportion of small dead knots. In this manual classification step, all available knowledge was used to identify butt-log material as accurately as possible.

Figure 1. Example of faces of four ID-marked planks, where the two on the top were determined to be sawn from butt logs and the two on the bottom were determined to be sawn from top logs. These images were not used for feature detection by the Boardmaster and were stitched together from three camera images per face after the data collection at the dry-sorting station using a Matlab script.

PLS implementation

Partial least-squares regression (PLS) was used to correlate the feature variables in $\boldsymbol{X}$ with the response vector $\boldsymbol{y}$ by multivariate linear regression, using the SIMCA 14 software [10]. The multidimensional regression line is then the prediction model trained on a sub-sample of $\boldsymbol{X}$, containing 217 planks of grade A and 217 of grade B from the 618 planks sawn from top logs. The reason for the use of a subsample of 434 planks during training is to avoid training a bias prediction model due to class imbalance [11]. The data of the superfluous planks of grade A was discarded. Since the PLS model was trained on sawn timber from top logs that are specifically suitable for the planing mill customer, the PLS model can be considered specialized.

By representing the plank grade A as 1, and the plank grade B as 0, a new observation can have its grade predicted by the regression line between 0 and 1[2]. In this way, the predicted grade value can be thought of as a probability that the plank's grade is A, e.g. if a new plank is measured as a new row $\boldsymbol{x}$, and the regression line predicts the grade-value y as 0.68, the plank should probably be classified as of grade A. However, a threshold has to be introduced somewhere between 0 and 1 (commonly 0.5) to decide the grade, which introduces a simple control over the grading outcome. For example, the sawmill could decide to use a threshold of 0.80 which would select planks for delivery to the customer that, by measurements standpoint, are very similar to distinct grade A planks. Such a selective prediction model would deliver grade A planks with very high grading accuracy, but to very low volumes, i.e a low yield of planks of grade A (see Figure 5 by Olofsson et al. [1] for more detail).

The focus of this study was on optimizing grading accuracy, where previous studies by Olofsson et al. [1, 2] have shown that an optimal threshold value is somewhere close to the range 0.50–0.60 for this specific implementation.

RESULTS

To test the sensitivity of product-adapted PLS-based HSAG towards systematic changes in the training and testing material, a PLS model was trained on the class-balanced set of 434 planks sawn from top logs and tested on the 282 planks sawn from both top logs and butt logs. The SIMCA 14 software used performed cross-validation to prevent overfitting and created two principal components for the prediction model. Based on the predicted grade-value of the class-balanced set of 434 planks from self-prediction, the optimal grade-separating threshold was estimated to be 0.488. This threshold was used to grade the set of 282 planks.

Table 1 shows the manual grading outcome at the planing mill, showing the disproportionate number of grade A planks sawn from top logs vs butt logs, where 67% of the top logs resulted in grade A planks compared to 36% of the butt logs.

---

[2] The predicted grade can lie outside the range 0–1 e.g. if the new observation is more extreme in $\boldsymbol{x}$ than any observation in the training set $\boldsymbol{X}$.

Table 1 Manual grading outcome at the planing mill of the 282 planks sawn from both top logs and butt logs, classified manually by staff at the planing mill.

Sawn timber from	Number of planks	Planks of grade A	Planks of grade B
All logs	282	168 (60%)	114 (40%)
Top logs only	213	143 (67%)	70 (33%)
Butt logs only	69	25 (36%)	44 (67%)

The PLS model trained on the 434 planks sawn from top logs was used to grade the 282 planks detailed in Table 1. The prediction outcome is shown in the confusion matrices in Table 2.

Table 2 Confusion matrices of the prediction results of the PLS model that was trained only on planks sawn from top logs and used to predict a set of planks from both top logs and butt logs. The prediction results are shown for all sawn material at the top and separately for the planks from top logs and butt logs below. Sorting outcome is also shown. Delivered planks shows the proportion of the scanned planks delivered from the sawmill, i.e. all planks graded as of grade A. Received A-planks shows the proportion of received planks by the planing mill that were correctly graded as of grade A.

			Predicted		
Sawn timber from	**Result**	**Number**	**Grade A**	**Grade B**	**Grading Accuracy**
All logs	**Grade A planks**	168	135	33	80%
	Grade B planks	114	40	74	65%
	Total	282	175	107	74%
	Delivered planks	175 (62%)			
	Received A-planks	135 (77%)			
Top logs	**Grade A planks**	143	117	26	82%
	Grade B planks	70	26	44	63%
	Total	213	143	70	76%
	Delivered planks	143 (67%)			
	Received A-planks	117 (82%)			
Butt logs	**Grade A planks**	25	18	7	72%
	Grade B planks	44	14	30	68%
	Total	69	32	37	70%
	Delivered planks	32 (46%)			
	Received A-planks	18 (56%)			

DISCUSSION

The results of this study show the possibilities of using automatically captured images to classify sawn material since images were the foundation of creating the distinctly different separation of sawn material from butt logs and top logs (Table 1).

The sawn timber from butt logs was systematically different and yielded much fewer planks of quality A for the planing mill customer than the sawn timber from top logs (Figure 1, Table 1), clearly showing why the planing mill purchases sawn timber from top logs. As the PLS model was trained specifically on sawn timber from top logs, as this material is suitable for the planing mill customer, the model was considered specialized. When the PLS model was used to grade sawn timber originating from top logs, as it was intended for, the grading accuracy was 76%, and, out of the scanned planks, 143 (67%) was delivered to the planing mill of which 117 (82%) yielded a majority of the desirable quality A boards (Table 2). Using the specialized PLS model for its indented use-case graded the sawn timber with similar grading accuracy as in the corresponding tests by Olofsson et al. [1, 2], showing that this implementation of product adapted HSAG is consistent over the course of three studies.

Table 2 shows that the specialized PLS model did not grade the sawn material from butt logs with the same level of grading accuracy as sawn material from top logs with 70% compared to 76% correctly graded planks, respectively. The lower grading accuracy is due to the systematically different butt log testing material. In test number 4 by Olofsson et al. [2], it was shown that the PLS model could not be trained to take the effect of the dust into account, in an attempt to increase the grading accuracy when grading dusty planks. This could be attributed to the random nature of the effect of the dust on the sawn material, as the dust seemingly randomly obscured or distorted the plank features during feature detection by the camera based Boardmaster. In this study, the systematically different material was sawn timber from butt logs, with a measurable systematic difference (Figure 1). In theory, this means that the sawn material from butt logs could be used for training to create a more generalized PLS model that could grade sawn timber from butt logs with similar accuracy as sawn timber from top logs. However, in this study, no generalized PLS model could be trained due to the lack of sawn material from butt logs. Lycken and Oja [4] estimated that a minimum of 100 planks per grade is required for training, which is supported by Olofsson et al. [1, 2]. This estimation of the required minimum number of planks could be extrapolated to each systematically different set of training material, which in the case of class-balanced sawn material from butt logs would require 100 planks of grade A and 100 planks of grade B, compared to the 25 and 44 planks available in this study. Future studies should investigate the possibilities of training PLS models on systematically different material to test the grading accuracy of a generalized PLS model vs a specialized PLS model in both a specialized and generalized use-case. This study can only conclude that a specialized PLS model is not suitable for a generalized use-case, as the grading accuracy is significantly lower.

The lower grading accuracy of sawn material from butt logs indicates that both the sawmill and planing mill will experience a poor sorting outcome (Table 2). Of the graded material from butt logs, the sawmill only gets to deliver 46% of the scanned planks, and, out of the delivered planks, the planing mill would only be able to use 56% of the planks to produce a majority of their desirable product of grade A.

CONCLUSIONS

The specialized multivariate PLS model, trained on sawn timber from top logs, graded the systematically different sawn timber from butt logs worse than the material the model was trained on. Theoretical arguments support the claim that sawn timber from butt logs could be included in the training set to create a generalized PLS model that could be used to grade sawn timber from either top log or butt logs without great loss of grading accuracy.

The result of this study is based on manual classification of sawn timber, as originating from top logs or butt logs, using automatically captured images. The possibility to automatically capture and use images of sawn timber opens up the opportunity of future work to train PLS models (or other machine learning implementations) based on image classification.

REFERENCES

1. Olofsson L, Broman O, Skog J, Fredriksson M, Sandberg D (2019) Multivariate Product Adapted Grading of Scots Pine Sawn Timber for an Industrial Customer, Part 1: Method Development. Wood Material Science & Engineering. In press. Available online DOI: 10.1080/17480272.2019.1617779
2. Olofsson L, Broman O, Skog J, Fredriksson M, Sandberg D (2019) Multivariate Product Adapted Grading of Scots Pine Sawn Timber for an Industrial Customer, Part 2: Robustness to Disturbances. Wood Material Science & Engineering. In press. Available online: DOI: 10.1080/17480272.2019.1612944
3. Swedish Sawmill Managers Association (1994) Nordic timber: grading rules for pine (Pinus sylvestris) and spruce (Picea abies) sawn timber: commercial grading based on evaluation of the four sides of sawn timber. Föreningen Svenska Sågverksmän, Stockholm.
4. Lycken A, Oja J (2006) A Multivariate Approach to Automatic Grading of Pinus Sylvestris Sawn Timber. Scandinavian Journal of Forest Research 21(2):167-174.
5. Berglund A, Broman O, Oja J, Grönlund A (2015) Customer Adapted Grading of Scots Pine Sawn Timber Using a Multivariate Method. Scandinavian Journal of Forest Research 30(1):87-97.
6. Olofsson L, Broman O, Fredriksson M, Skog J, Sandberg D (2017) Customer adapted grading of Scots Pine Sawn Timber – A Multivariate Method Approach. Proceedings of International Wood Machining Seminar 23(1):360-371.
7. Wold S, Sjöström M, Eriksson L (2001) PLS-Regression: A Basic Tool of Chemometrics. Chemometrics and Intelligent Laboratory Systems 58(2):109-130
8. Geladi P, Kowalski RB (1986) Partial Least-Squares Regression: a Tutorial. Analytica Chimica Acta 185(1):1-17.
9. Finscan (2019) Boardmaster. [Online]. Available at: https://finscan.fi/ [Accessed 3 May 2019]. Note: a predecessor to Boardmaster NOVA.
10. Sartorius Stedim Biotech (2019) SIMCA. [Online]. Available at: https://umetrics.com/ [Accessed 3 May 2019].
11. Japkowicz N, Stephen S (2002) The Class Imbalance Problem: A Systematic Study. Intelligent Data Analysis 6(5):429-499

PUBLICATION VI

The Effect of Class-balance and Class-overlap in the Training Set for Multivariate and Product-adapted Grading of Scots pine Sawn Timber

Authors and year:
Olofsson, L., Broman, O., Oja, J. and Sandberg, D.
2021

Publication originally published in:
Wood Material Science & Engineering 16(1), 58–63
DOI: 10.1080/17480272.2020.1804996

WOOD MATERIAL SCIENCE & ENGINEERING
2021, VOL. 16, NO. 1, 58–63
https://doi.org/10.1080/17480272.2020.1804996

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

The effect of class-balance and class-overlap in the training set for multivariate and product-adapted grading of Scots pine sawn timber

Linus Olofsson [a], Olof Broman[a], Johan Oja[a,b] and Dick Sandberg [a]

[a]Wood Science and Engineering, Luleå University of Technology, Skellefteå, Sweden; [b]Norra Timber, Skellefteå, Sweden

ABSTRACT

Using multivariate partial least squares regression (PLS) to perform visual quality grading of sawn timber requires a training set with known quality grades for the training of a grading model. This study evaluated the grading accuracy of an independent test set of sawn timber when changing the aspects of class-balance and class-overlap of the training set consisting of 251 planks. The study also compared two ways of expressing the reference-grade of the training set; by grading images picturing the planks, and by grading the product produced from the planks. Two grading models were trained using each reference-grade to establish a baseline for comparison. Both models achieved a 76% grading accuracy of the test set, indicating that both reference-grades can be used to train comparable models. To study the class-balance and class-overlap aspects of the training set, 25% of the training set was removed in two training scenarios. The models trained on class-balanced data indicated that class-imbalance of the training set was not a problem. The models trained on data with less class-overlap using the product-grade reference suffered a 4%-points grading accuracy loss due to the smaller training set, while the model trained using the image-grade reference retained its grading accuracy.

ARTICLE HISTORY
Received 12 March 2020
Revised 24 June 2020
Accepted 23 July 2020

KEYWORDS
Sawn timber; PLS regression; machine-learning; training aspects

1. Introduction

Automatic visual quality grading of kiln-dried Scots pine sawn timber can be implemented using machine learning. This paper uses multivariate partial-least squares regression (PLS Geladi and Kowalski 1986, Wold *et al.* 2001), which can mimic the holistic and subjective grading performed by manual graders. Using multivariate grading for automatic sorting of sawn timber for a specific product have been implemented in the studies by Olofsson *et al.* (2019b, 2019c), and for similar visual classification purposes by Broman (2000), Lycken and Oja (2006), Breinig *et al.* (2015), Berglund *et al.* (2015), and Olofsson *et al.* (2017). These studies showed the benefits of using multivariate PLS regression models in sawmills, but do not properly discuss the intricacies of selecting training data for machine-learning. Lycken and Oja (2006) tested how a PLS grading model trained on one dimension of planks performed on a different dimension of planks. Olofsson *et al.* (2019a) showed how a model trained on sawn timber from Scots pine top-logs performed when grading different material sawn from butt-logs (much fewer knots, smaller, and almost exclusively dead) – both providing some insight in how different training material affect the grading outcome.

Using machine learning for automatic sorting of kiln-dried sawn timber requires not only a data set with a large number of observations to train an accurate grading model on but also a high-quality dataset, i.e. a dataset with a high correlation between the automatically measured variables of the sawn timber and the assigned quality grade of every single piece of sawn timber.

This study focuses on understanding how different aspects of a training set affects the grading outcome of automatic sorting of kiln-dried sawn timber using multivariate PLS regression. More specifically, the aspects of class-imbalance and class-overlap. In the context of sorting sawn timber using machine-learning techniques, the aspect of class-imbalance means to use a dataset with two or more classes of sawn timber that are not equal in numbers. Uneven distribution of classes is the typical case of most sawing batches due to the way the grading standards used at the sawmill defines classes (Swedish Sawmill Managers Association 1994). The aspect of class-overlap means how closely related observations of different classes are by a set of measurements. A high class-overlap is the typical case of most sawing batches as the grading standards used separates classes by limits, e.g. 'maximum number of dead knots'. Using limits to separate between classes means two observations of different classes could possibly be distinguished only by a single knot, and since a large number of features can be used to describe a piece of sawn timber, the class-overlap is considerable. The class-overlap and class-balance aspects of classification problems are well known in the machine learning literature where Prati *et al.* (2004) explain these two aspects for synthetic data sets, using a k-nearest-neighbours approach.

Prati *et al.* (2004) showed that the problems of class-imbalance or class-overlap when training a machine-learning model is not the only obstacle for applying machine-learning algorithms to real-world problems, but rather the combination of the two. In the works by Olofsson *et al.* (2019b,

CONTACT Linus Olofsson linus.olofsson@ltu.se Wood Science and Engineering, Luleå University of Technology, Skellefteå SE-931 87, Sweden

2019c) the problems of class-imbalance and class-overlap was implicitly shown but neither highlighted nor appropriately discussed. The present study investigated how class-balance and class-overlap affect the training of a PLS grading model by testing different training scenarios. By using different training data with different properties, the performance of the PLS grading model is expected to change. Therefore, the present study compared different PLS models based on the same data, but with a different number of class-members and with different amount of class-overlap, and how this changes the grading accuracy and grading outcome.

The investigation of class-balance and class-overlap was performed with two different reference grades of the training material. (1) Like in the earlier cited works, the sawn-timber grade was assigned by a product manufacturer assessing the split, milled, and planed finished wall panel product produced by each piece of sawn timber. (2) The same product manufacturer graded images of each piece of sawn timber to assess their suitability for the intended product. The reasoning for investigating the second image-grade reference is that such training data is much easier and cheaper to acquire compared to having to process and grade large batches of unsorted sawn timber by a customer. Testing the effect of class-balance and class-overlap using these two different grading references for training may show if the subjective image-grade reference performs differently than the more objective product-grade reference.

2. Material and method

The implementation of PLS Discriminant Analysis in the present study followed the methodology applied by Olofsson *et al.* (2019b), where an almost identical implementation was presented in more detail. The study was executed in cooperation with Kåge Sawmill (Norra Timber) and Lundgren's planing mill, both located in northern Sweden. Since the available training data for this study was limited by the number of images graded by the customer, the only way to manipulate the different aspects of the training data was by removing observations. Since it is always desirable of a grading model to perform as accurately as possible, one should carefully consider the effect of removing training data before actual implementation. Furthermore, the specific implementation of PLS regression in this study makes the data unsuitable for re-sampling techniques, such as the well-known and otherwise useful SMOTE algorithm (Chawla *et al.* 2002). Other implementations of machine-learning techniques for automatic grading of sawn timber could benefit from such techniques.

2.1. Reference grades of the material

Three data sets from earlier studies were used (Olofsson *et al.* 2019b, 2019c, 2019a), consisting of nearly 300 Scots pine (*Pinus sylvestris* L.) planks each, for a total of 846 planks. All planks were cant-sawn from top logs or middle logs to the dimensions 50×150 mm and the target moisture content after drying was 14%. The planks had a length between 3.6 m and 5.4 m at the dry-sorting station where cameras automatically scanned the planks. No distinction was made between the data sets used in this paper. Sample images can be seen in Figure 1 in Olofsson *et al.* (2019c).

Table 1. Results of the product-grade reference and image-grade reference of the baseline training data set consisting of 251 planks.

	Grades				
Grade reference	A	Not-clearly A	Not-clearly B	B	Totals
Product	157	–	–	94	251
Image	123	46	18	64	251

The quality grade of each plank was determined in two ways. (1) The planks were delivered to Lundgren's planing mill where each plank was split into three boards (15×150 mm); each milled, planed, and manually graded as an accepted or rejected piece of wall panel. The resulting grade of each plank was the majority of the produced three pieces of boards, meaning a plank that produced two accepted and one rejected board would be given the grade A, meaning accepted, or vice versa the grade B, meaning rejected. This A or B grade of each plank was called the 'product-grade'. (2) Images of all plank faces of a sub-set of 251 planks, extracted from the automatic grading system at the sawmill's dry-sorting station, were presented to the quality expert at Lundgren's planing mill. The quality expert was instructed to only consider knot features when asked to try and predict if each plank would produce a majority of A quality boards, and in that case give that plank the grade A, and vice versa for grade B. This A or B grade was called the 'image-grade'. The quality expert graded the images of each plank with grade A or grade B, and also labelled 64 planks as 'not-clearly', meaning each plank was labelled with one of four labels: A, not-clearly A, not-clearly B, or B.

The focus of this study was to study different aspects of the training data when using a PLS regression model to grade sawn timber automatically. To be as consistent as possible when selecting training data for the comparison between models trained using the product-grade reference and the image-grade reference, the sub-set of 251 planks with both a product-grade reference and an image-grade reference was used as the so-called 'baseline' dataset for training. The outcome of the two grading processes of the baseline training set is presented in Table 1.

The remaining 595 planks, not part of the baseline training set, with only a product-grade, were used as a large test set for all tests. The grading accuracy of any model is how accurately it grades the test set to match the product grade reference, consisting of 386 planks of grade A and 209 planks of grade B.

2.2. PLS-DA implementation

Each plank was associated with a set of aggregated feature-variables regarding knots and a binary quality grade. At the sawmill's dry-sorting station, an automatic scanning and grading system by FinScan, called Boardmaster (Anon 2018), was used for detection of plank features (knots). The knots were classified as, e.g. sound or dead, and their size and position were determined. These measurements were used to create a set of 3564 aggregated variables, described like 'The total number of sounds knots on the inner face side of

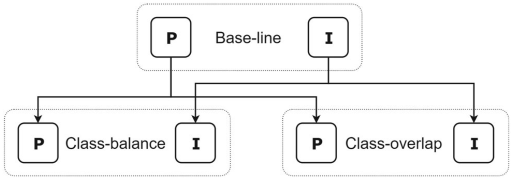


Figure 1. Flow-chart showing the different training scenarios and prediction models. Solid boxes represent models trained on the product-grade (P) or the image-grade (I), respectively. Dotted boxes show the different training scenarios where the two models are compared with the baseline training scenario.

the plank' (Olofsson *et al.* 2019b). These variables were created for each plank and stored as an **X**-matrix.

The product-grade and image-grade data were stored as two binary response **y**-vectors. Using the plank features stored in **X**, and the plank grade in **y** (product grade and image grade separately), the SIMCA 14 software (Anon 2019) was used to correlate **X** and **y** of a training set, using multivariate Partial Least Squares Discriminant Analysis (PLS-DA). Only one PLS-component was used to separate between the two grades in all tests as using two PLS-components showed signs of overfitting the training data. The trained model was then used to predict the grade of the test set.

2.3. Training and testing procedure

The baseline training set shown in Table 1 was used to train two PLS models, one using the product-grade reference and one using the image-grade reference (Figure 1). Once the baseline grading outcome was established, planks were removed from the baseline training dataset to change the class-overlap and class-balance aspects of the training data. For each of the new training scenarios, two new models were created for a total of six models. All trained models were used to predict the product-grade of the test set, i.e. the models trained on the image-grade were also used to predict the product-grade.

In order to be able to attribute the changes in the grading outcome, or lack thereof, to the changes in the training set in each training scenario, each model's class-separating threshold was calibrated to achieve a specific grading outcome (see Olofsson *et al.* 2019b for further details regarding the class-separating threshold). The grading outcome was controlled such that each grading model forcefully achieved the proportions of grade A and grade B planks of the test set; the test consisting of 595 planks had a ratio of grade A and grade B planks of 2:1 and the class-separating threshold was calibrated such that each model graded grade A and B planks in a 2:1 ratio. Furthermore, any grading errors made by the models will be evenly distributed between false-negatives and false-positives for a simpler comparison.

2.3.1. Class-overlap

Since the class-overlap aspect of the training data is different for the product-grade and image-grade references, the class-overlap aspect was manipulated separately for each reference. (1) For the image-grade reference, the decision of which planks to remove was made based on the image-grading performed by the quality expert at Lundgren's planing mill. To reduce the class-overlap (make the classes more distinct) of the training set, the 64 planks labelled 'not-clearly' in Table 1 were removed. (2) For the product-grade reference, the product-grade trained model from the baseline scenario was used to produce an observed-predicted plot of the baseline training data (predicting the data it was trained on), shown in Figure 2. The encircled observations in Figure 2 had a low correlation between their variables in **X** and product-grade in **y** since the model did not accurately capture their grade. To reduce the class-overlap of the training data, 64 observations like the encircled ones were removed in order of lowest correlation (left to right for grade A planks, and right to left for grade B).

Once 64 planks were removed from each reference-grade for the class-balance training scenario, two new models were trained on the remaining 187 planks, using the product-grade and the image-grade respectively. Note that different planks were removed for the training of each of the two models in this scenario, but the same number of planks were removed each time such that differently sized training sets do not influence the comparison of these two models, but only the difference of the class-balance aspect. Out of the removed 64 planks, 25 planks (36%) were removed from both reference grades while 39 unique planks (64%) were removed in each case.

2.3.2. Class-balance

For the training scenario where class-balance was investigated, the number of planks in the training set of each reference-grade was reduced to the largest possible number of observations that resulted in an equal number of grade A planks and grade B planks. For the product-grade reference, 63 planks were removed from the baseline training set, and the remaining 188 planks consisted of 94 planks of each

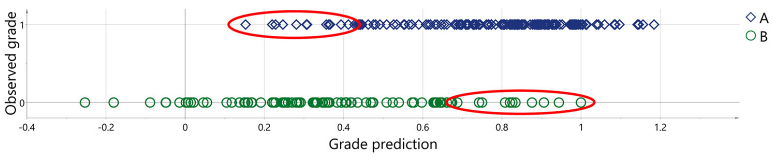


Figure 2. The observed-predicted plot of the baseline training data, using the model trained on the baseline training data with the product-grade reference. The upper observations (1) represents grade A, and the lower (0) represents grade B. The y-axis shows the grade of each plank as the actual binary grade, and the x-axis shows the continuous grade predicted by the model. The encircled observations have a weak correlation between their measured features and their assigned grade, i.e. an observation in the bottom right looks to the model as, and would have been predicted as, a plank of grade A (1) while the product-grade was grade B (0).

Table 2. Misclassification tables for the three training scenarios: baseline, class-overlap, and class-balance.

(a) Baseline training scenario

		Predicted by product-trained		Grading accuracy	Predicted by image-trained		Grading accuracy
Observed	Grade	A	B		A	B	
	A	314	72	81%	313	73	81%
	B	73	136	65%	72	137	66%
	Total			76%			76%

(b) Class-overlap training scenario

		Predicted by product-trained		Grading accuracy	Predicted by image-trained		Grading accuracy
Observed	Grade	A	B		A	B	
	A	302	84	78%	314	72	81%
	B	84	125	60%	73	136	65%
	Total			72%			76%

(c) Class-balance training scenario

		Predicted by product-trained		Grading accuracy	Predicted by image-trained		Grading accuracy
Observed	Grade	A	B		A	B	
	A	310	76	80%	307	79	80%
	B	77	132	63%	78	131	63%
	Total			74%			74%

Note: Each sub-table presents the grading results of the two models trained on the product-grade reference (product-trained) and image-grade reference (image-trained) for that training scenario. The test set contained 386 planks of grade A and 209 planks of grade B.

grade. For the image-grade reference, 87 planks were removed, and the remaining 164 planks consisted of 82 planks of each grade.

3. Results

To investigate the effect of the class-balance and class-overlap aspects of training data on the grading outcome relative to a baseline training scenario, using two different reference grades for training, six multivariate PLS-DA models were trained. The grading outcome of the same test set of each of the six models is shown in Table 2.

For further comparison of the grading outcome for different training scenarios, the grading agreement between the different models, i.e. the proportion of individual planks in the test set that was graded the same by two models, is presented in Table 3.

4. Discussion

The baseline training scenario (Table 2(a)) showed that the product-grade reference and image-grade reference resulted in the same grading accuracy, i.e. 76%. The two baseline models predicted the test set with an 88% agreement (Table 3). No qualitative investigation was made to investigate if any systematic difference could be detected. The reason for the higher grading accuracy of grade A planks for both models is partly due to the moderately overrepresented number of grade A planks in the training set, resulting in a slightly biased model with a higher grading accuracy for grade A planks. Furthermore, the higher grading accuracy of grade A planks can be attributed to the more homogeneous nature of grade A planks, as they are specifically graded for a wall panelling product for which a homogeneous knot-pattern is preferred, whereas grade B planks are much more heterogeneous in their knot-pattern. The similar total grading accuracy is strong evidence that a grading model using an image-based grading reference for training would perform as good as a model trained on the product-grade reference – which is much more tedious and costly to acquire due to having to process the training data into a finished product before reference grading.

Table 3. Grading agreement of all models, showing the proportion of the test set graded identically by two models measured in percent (%). The headers show the training scenario and the two corresponding models of that scenario, trained on the reference product-grade (P) and image-grade (I), respectively.

		Baseline		Class-overlap		Class-balance	
Scenario	Reference	P	I	P	I	P	I
baseline	P	100					
	I	88	100				
Class-overlap	P	93	86	100			
	I	88	96	85	100		
Class-balance	P	95	87	94	87	100	
	I	87	92	85	93	87	100

The class-overlap training scenario (Table 2(b)), where training observations were removed to increase the separation between the classes showed that the model trained on an image-grade reference retained a grading accuracy of 76%, and 96% of the test set was predicted identically as the corresponding baseline model (Table 3). This result implies that the 64 'not-clearly' labelled observations in Table 1 that was removed contributed very little to the model. For future studies using a similar image-grade reference this shows that if a quality-expert is not sure about the desirability of a specific plank, that plank might as well be removed from the experiment entirely, as it will probably not contribute much to a grading model. Future studies might also want to design the image-grading process in such a way that large quantities of distinct grade A and grade B planks can be processed efficiently, e.g. using images of automatically pre-sorted planks.

The model trained on the product-grade reference in the class-overlap training scenario surprisingly suffered a loss of 4% points of total grading accuracy, as the 64 observations that were removed, shown in Figure 2, were considered to be directly misleading during the training of the model. However, it is reasonable to assume that the objective nature of the product-grade reference contributes more to the grading performance of the model, however weak, than the observation labelled 'not-clearly' did for the image-grade reference trained model. Hence the results indicate that the more objective product-grading reference does not benefit from a smaller class-overlap at the cost of removing observations for training – at least not for this data set.

For the class-balance training scenario, both models performed again very similarly in terms of grading performance (Table 2), and 87% of the test set observations were predicted the same by both models (Table 3). Both models had a total grading accuracy of 74%, down from 76%. These results show that the performance of the baseline models is not hindered by an over-representation of grade A planks in the training set and that removing observations to achieve a class-balanced training set lowers the grading accuracy of the model in this study. The lower grading accuracy is assumed to be because of the smaller training set. Ideally, the training set should consist of an equal number of observations from each grade to remove any bias introduced when training on a class-imbalanced training set. However, due to the natural distribution of grade A and grade B observations found in the raw-material (2:1 ratio), and since removing grade A observations to achieve a class-balanced training set reduced the grading accuracy, the remaining alternative is to generate more observations of grade B artificially. However, due to the way the aggregated knot-feature variables were created, and due to their large number, using re-sampling techniques, like the SMOTE algorithm (Chawla *et al.* 2002), was considered to be unsuitable for this study. Future studies could instead try to simulate sawn timber from which the aggregated variables were created.

Overrepresentation of a class in a training set leads typically to a bias in the model that predicts the overrepresented class member more often than the underrepresented class (seen in the results in this study and in Olofsson *et al.* (2019b, 2019c, 2019a)). However, in the context of dry-sorting sawn timber, this is sometimes desirable since the unbalanced nature of the training set is usually reoccurring in the test set. Another reason for the lower grading accuracy of grade B planks than grade A planks is their more heterogeneous knot-pattern. Furthermore, the problem of class-balance was determined to be moderate in this study and the related cited works. In this study, the entire data available had roughly a 2:1 ratio of grade A planks to grade B planks according to the product-grade, in comparison to, e.g. medical studies dealing with cancer patients of a population with a vastly disproportionate number of healthy and sick subjects. For futures studies, these results show that the class-imbalance problem can be solved by changing the class-separating threshold, rather than sacrificing grading accuracy by removing training observation to achieve a class-balanced training set.

5. Conclusions

When training a PLS model for automatic grading of kiln-dried sawn timber for a customer's product, this study indicates that when using the objective product-grade reference, i.e. the product grade outcome of each piece of sawn timber, training on all available data will give the highest grading accuracy (76% correctly graded pieces of sawn timber). Removing observations from the training set to achieve a training set with less class-overlap, or a class-balanced training set, did not improve the grading accuracy (72% and 74%, respectively). When using the subjective image-grade reference for training, i.e. the grade given by the customer to images of each piece of sawn timber, removing training data to achieve a class-balanced training set did not improve the grading accuracy (74%). The results regarding class-balance indicate that the common problem of training machine-learning models on a class-imbalanced training set does not seem to be a problem in the context of dry-sorting sawn timber, as the class-imbalance of such data sets is only moderate, i.e. roughly 2:1 in this study.

Using the image-grade reference, training on all data achieved the same grading accuracy (76%) as the corresponding model trained with the product-grade reference. This shows that a product-adapted grading model could be trained using images extracted from the dry-sorting station of a sawmill, instead of having to process the training data into a split, milled, and planed product. Removing observations of the image-grade reference data, for which the customer had difficulty in determining the grade, in reducing the class-overlap retained the grading accuracy (76%). This retention in grading accuracy showed that there are benefits to be had with a more carefully structured image-grading procedure, as the removed training data had a net-zero effect on grading accuracy, i.e. a waste of time for the customer.

Acknowledgments

Financial support from the Swedish Innovation Agency (Vinnova), project Sawmill 4.0 – Customised flexible sawmill production by integrating data-driven models and decisions tools 2018-02749, is gratefully acknowledged. The authors also gratefully acknowledge the support of the CT-Wood – a centre of excellence at Luleå University of Technology for the development of X-ray computed tomography applications for use in the forest products industry.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Financial support from the Swedish Innovation Agency (VINNOVA), project Sawmill 4.0 – Customised flexible sawmill production by integrating data-driven models and decisions tools 2018-02749, is gratefully acknowledged.

ORCID

Linus Olofsson http://orcid.org/0000-0002-5562-5142

Dick Sandberg http://orcid.org/0000-0002-4526-9391

References

Anon (2018) Boardmaster. Accessed 12 June 2018, available at: https://finscan.fi/products/boardmaster/?lang=en. Note: Older predecessor to BoardmasterNOVA.

Anon (2019) Simca. Accessed 3 January 2019, available at: https://umetrics.com/products/simca.

Berglund, A., Broman, O., Oja, J. and Grönlund, A. (2015) Customer adapted grading of scots pine sawn timber using a multivariate method. *Scandinavian Journal of Forest Research*, *30*(1), 87–97.

Breinig, L., Leonhart, R., Broman, O., Manuel, A., Brüchert, F. and Becker, G. (2015) Classification of wood surfaces according to visual appearance by multivariate analysis of wood feature data. *Journal of Wood Science*, *61*(2), 89–112.

Broman, O. (2000) *Means to measure the aesthetic properties of wood*. PhD thesis, Luleå University of Technology.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002) SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Geladi, P. and Kowalski, B. R. (1986) Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, *185*, 1–17.

Lycken, A. and Oja, J. (2006) A multivariate approach to automatic grading of pinus sylvestirs sawn timber. *Scandinavian Journal of Forest Research*, *21*(2), 167–174.

Olofsson, L., Broman, O., Fredriksson, M., Skog, J. and Sandberg, D. (2017) Customer adapted grading of scots pine sawn timber – a multivariate method approach. *Proceedings of International Wood Machining Seminar*, *23*(1), 360–371.

Olofsson, L., Broman, O. and Sandberg, D. (2019a) Holistic-subjective automatic grading of sawn timber: Sensitivity to systematic changes. In *IWMS 24*. August 25–30, 2019, Corvallis, Oregon, USA, pp. 154–167. Oregon State University (OSU).

Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019b) Multivariate product adapted grading of scots pine sawn timber for an industrial customer, part 1: Method development. *Wood Material Science & Engineering*, *14*(6), 428–436.

Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019c) Multivariate product adapted grading of scots pine sawn timber for an industrial customer, part 2: Robustness to disturbances. *Wood Material Science & Engineering*, *14*(6), 420–427.

Prati, R. C., Batista, G. E. and Monard, M. C. (2004) Class imbalances versus class overlapping: An analysis of a learning system behavior. In *Mexican International Conference on Artificial Intelligence*, pp. 312–321. Springer.

Swedish Sawmill Managers Association (1994) *Nordic Timber: Grading Rules for Pine (Pinus Sylvestris*) and Spruce (*Picea Abies*) Sawn Timber: Comercial Grading Based on Evaluation of the Four Sides of Sawn Timber (Markaryd: Föreningen Svenska Sågverksmän).

Wold, S., Sjöström, M. and Eriksson, L. (2001) PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130.

Publication VII

Product Adapted Grading of Virtual Scots pine Sawn Timber by an Industrial CT-scanner Using a Visually Trained Machine-Learning Method

Authors and year:

Olofsson, L., Broman, O., Oja, J. and Sandberg, D.

2021

Reformatted version of publication submitted to journal:

Wood Material Science & Engineering 16(4), 279–286

DOI: 10.1080/17480272.2021.1955298

Product Adapted Grading of Virtual Scots pine Sawn Timber by an Industrial CT-scanner Using a Visually Trained Machine Learning Method

Linus Olofsson[a], Olof Broman[a], Johan Oja[a,b], and Dick Sandberg[a]

[a]Wood Science and Engineering, Luleå University of Technology, SE-931 87 Skellefteå, Sweden; [b]Norra Timber, Skellefteå, Sweden

ABSTRACT

Computed tomography (CT) scanning of logs allows for the appearance grading of virtual sawn timber before sawing the log. This study compares an existing rule-based approach with two partial least squares (PLS) regression machine-learning models when grading and deciding to dedicate virtual sawn timber for a specific customer's product or not. The data consisted of a 3D reproduction of the knot structure within each of the 156 CT scanned logs projected onto the surfaces of the virtual sawn timber. The virtual sawn timber's knot structure was graded as suitable for the intended customer or not suitable by the existing rule-based approach, and the log's position was optimised to maximise the yield of such sawn timber. Once the sawing was decided, the virtual sawn timber's knot structure was used to create a large set of descriptive statistical variables used by two machine learning models. The PLS models were trained on either of two quality references; the finished product grade, i.e. the quality grade of a wall panelling product produced by the customer, or the image-grade based on images of the sawn timber, extracted from the dry-sorting station's automatic grading system and graded by two experienced researchers. The results show that the two PLS models perform equally well at allocating suitable virtual sawn timber to the customer, indicating that the quality references are equally useful for training a PLS model. The PLS models delivered 93% of the dried sawn timber to the customer, leaving very little sawn timber with customer-specific properties at the sawmill, with 89% and 90% of the delivered material passing the intended product quality demands. The rule-based approach delivered 85% dried sawn timber with a 73% pass rate.

KEYWORDS

Sawn timber, CT, Machine Learning, PLS

1. Introduction

X-ray computed tomography (CT) scanning of logs before sawing allows for the earliest possible decision making in the sawmill process. CT-scanning a log gives information of the log's internal knot structure, which is the main feature to consider when grading sawn timber. Virtual sawn timber is fitted inside the CT scanned virtual representation of the log such that the knot structure on the surfaces and within the virtual sawn timber can be measured. The virtual sawn timber can then be graded according to some grading criteria. Once graded, each virtual piece of sawn timber can be dedicated to a specific customer, the intended *main product* (Sandberg and Teischinger, 2021), and the log can then be sawn. Based on virtual grading, the sawmill

CONTACT Linus Olofsson Email: linus.olofsson@ltu.se

can also plan the drying of the sawn timber to fulfil customer demands. One advantage of such planning is that sawn timber of the same dimension but with different drying requirements can be separated before drying and the traditional dry-sorting grading procedure. By performing the quality grading early in the production process, any subsequent grading can be used as a complementary grading process, if at all necessary. Such subsequent grading may be necessary, depending on the intended customer, to detect rotten or dead knots if they are filled with resin as CT scanning struggle to detect the small density variations for such features.

Appearance grading of sawn timber is usually performed using a rule-based approach. Rules (limits) are enforced such that, e.g. the knot structure of the sawn timber abide by standardised quality grades, such as the Nordic Timber Grading Rules (NTGR – Swedish Sawmill Managers Association (1994)). The sawmill participating in this study has two rule-based sawn timber grading systems; a CT-based system that optimises the log's positioning before sawing, and a camera-based system that performs the final grading at the dry-sorting station. Both systems implement the NTGR or, if needed, some customer adapted grading rules. Lycken and Oja (2006) argued that the NTGR rules are troublesome to customise for individual customers with special requirements. In particular, the difficulty in having the customer describe their subjective quality criteria of the entire face of the sawn timber in a way that can be described by a set of rules governing individual features was mentioned. Olofsson et al. (2019b) showed that a partial least squares regression (or projection to latent structures, PLS (Geladi and Kowalski, 1986; Wold et al., 2001)) model could, with benefits, replace the rule-based grading for a product-specific grading. In particular, only requiring the final product grade as input from the customer was mentioned. The benefits of PLS-based grading over rule-based grading suggested by Lycken and Oja (2006) and Olofsson et al. (2019b) could be summarised as the PLS grading model captures the customer's subjective and holistic grading criteria, which is difficult to describe using a rule-based approach.

One problem with using a machine learning approach for product or customer adapted grading is acquiring data to train and test the grading model. The data needs to consist of measured features, $\mathbf{X}$, and a reference-grade, $\mathbf{Y}$, of the sawn timber which the machine learning method can model, in this study by multivariate linear regression, PLS. To achieve a robust grading model, the data needs to represent all possible incoming data, i.e. it needs to include sawn timber resulting in both a good and a poor quality reference. In this study, acquiring such data requires tracking of each piece of sawn timber through the entire refinement process; from the point of being a virtual piece of sawn timber scanned by the CT scanner before sawing to the point of being a finished wall panel at the customer's final quality control. Such data collection is usually a troublesome and costly logistical process. Olofsson et al. (2021) showed that it is possible to acquire equally useful data for product-adapted dry-sorting in an easier way by having the customer grade images of the sawn timber as the basis for the quality grade reference. Based on images of the sawn timber, the customer can guess the product grade resulting from each piece of sawn timber with high certainty. The benefits of using images of the sawn timber as the basis for the quality grade reference is: that the images are easily acquired at the sawmill from a camera-based automatic grading system at the dry-sorting station; the images are easily delivered to the customer and graded by a human subjectively and holistically; the images are not customer-specific, only specie and dimension specific and can therefore be reused for different customers of that specie and dimension; the use of images as the basis for the quality reference greatly simplifies the data collection process as no material needs to be delivered and processed by the customer; (Olofsson et al., 2019a) showed the possibility to perform this data collection in future studies using a traceability algorithm to match sawn timber scanned by the CT scanner with the same sawn timber scanned by the camera-based automatic grading system at the dry-sorting station.

The purpose of the study was to investigate if it is possible to use a machine-learning PLS model to appearance grade virtual sawn timber inside a log before the log is sawn. A goal was to be able to use images of sawn timber to create a reference grade for the machine-learning system.

2. Material and methods

2.1. The material

The material investigated consists of 303 centre yield pieces, with a cross-section dimension of 50×150 mm, of sawn timber taken from 156 Scots pine (*Pinus sylvestris* L.) logs (9 pieces were lost or destroyed), between 3.6 m and 5.4 m in length, and dried to 14% moisture content. Two pieces of centre yield were sawn from each log.

2.2. The Method

An overview of the study is shown in Figure 1 – letters (A-J) throughout this paper refers to this figure. This study was performed at Sävar Sawmill ("the sawmill"), where log data from the on-site CT scanner (B) and sawn-timber data from the camera-based dry-sorting system (H), was collected. The sawn timber was then processed into wall panels at Lundgren's Planing Mill ("the customer"), where the quality grade of each piece of sawn timber (J) was determined based on the produced wall panels' quality grade.

All the 303 pieces of sawn timber were tracked from the virtual sawn timber from the CT-scanning of the log, the sawing and drying, the optical scanning in the dry-sorting station, the refinement process at the customer, and to the final quality-grading station at the customer. The data collected was used to train two PLS grading models using two different reference grades. Using 5-fold cross-validation, the 303 pieces of sawn timber were simulated to pass through the refinement process again as an independent test set. The summed results of all 5 tests were presented to compare the sorting performance of a rule-based approach and the two PLS-based models.

During the simulated sorting, the grading steps used in this study had the option to classify the sawn timber as the *main product* or not. The main product was dedicated to the customer and continues to the next step of the refinement process. The CT-scanning system can perform grading of virtual sawn timber before sawing the log. If the virtual sawn timber was deemed unsuitable for the customer, the CT-scanning system could decide to saw the log into a different main product for a different customer during the simulated sorting. The dried sawn timber deemed unsuitable for the customer is graded as a *consequence product* (Sandberg and Teischinger, 2021) and is not delivered to the customer during the simulated sorting.

Using either rule-based grading or PLS-based grading, the CT-scanning system is the main grading step in this study. The grading performed by the CT-scanning system is complemented by two steps, one before sawing (E) and one after drying (I), to remove outliers. The complementary steps are the same for both the rule-based approach and the PLS models.

2.3. Data Collection

The industrial customer Lundgren's Planing Mill aims to produce a specific wall panelling product. The sawn timber delivered to Lundgren's Planing Mill for this product is from Sävar Sawmill, and three wall panels are made from one piece of sawn timber by splitting and milling (K). The wall panels are manually graded as either the desirable grade A or the undesirable grade B. Each piece of sawn timber was given an A or B product-grade $\mathbf{Y_p}$ (J) as the majority wall panel grade produced from that piece of sawn timber. Both the wall panels and sawn timber share the A or B quality grade label, but as this study only deals with the grading of sawn timber, only the A or B grade of the sawn timber was used. This wall panelling product is sensitive to dead knots as they tend to crack or fall out, and sound knots can crack during splitting and milling.

The sawn timber delivered from the sawmill was scanned using a Boardmaster dry-sorting system (Anon., 2018) that uses cameras to measure externally visible features to describe the sawn timber in a way that a computer can automatically grade. Images of the sawn timber (Figure 2 was extracted from the Boardmaster and two experienced researchers graded the

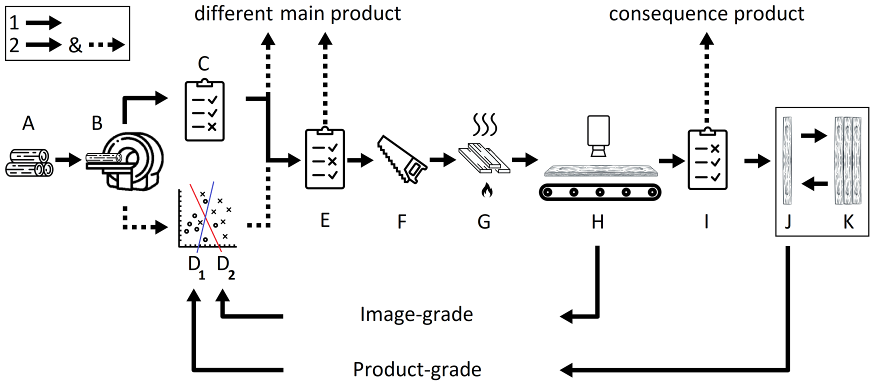


Figure 1. Material flow chart. 1, data collection of the entire dataset through the entire refinement process. 2, simulated sorting comparing the sorting outcome of the rule-based approach with the PLS-based approaches. A, the 156 logs selected for this study. B, CT-scanning all the logs performed by the CT Log. C, the rule-based approach optimising and deciding the sawing. (D_1) and (D_2), two PLS models correlating variables ($\mathbf{X}$) from the CT-scanning and the product-grade $\mathbf{Y_p}$, or image-grade ($\mathbf{Y_i}$), quality references, respectively. C and D can classify the virtual sawn timber as the main product intended for the wall panelling customer or, if not suitable for this customer, as a different main product for a different customer. E, additional rules applied to C and D to remove outliers. F, sawing the log. G, kiln drying the sawn timber. H, the Boardmaster at the dry-sorting station capturing images of the sawn material whose image-grade is used as a quality reference for (D_2). I, additional rules applied to remove outliers. J, a piece of sawn timber delivered to the customer receiving a quality product-grade ($\mathbf{Y_p}$) based on; K, split, milled, and quality graded by the customer triplet of wall panels.

images as suitable for the customer or not. This image-grade $\mathbf{Y_i}$ (H) classifies the dried sawn timber as the main product (suitable for Lundgren's Planing Mill) or as a consequence product (not suitable for Lundgren's Planing Mill) which, for easy comparison with the product-grade, will also have the label A or B, respectively.

The sawmill sorts the incoming logs as suitable for this customer based on external measurements such as top-log diameter. After that, the on-site CT-scanning system (B), called CT Log (Ursella et al., 2018), by Microtec, scans the logs. CT scanning of the logs before sawing allows for 3D measuring of internal features, most importantly knots. The CT Log transforms the 3D knot data into a 2D knot structure on the surfaces of the virtual sawn timber. A large set of variables $\mathbf{X}$ were created based on the 2D knot structure. Also, some log characteristics were included in the set of variables, e.g. pith deviation from a centre line.

Figure 2. Example of images used for image-based grading. All four faces of the sawn timber are shown.

2.3.1. Variable set created using the CT Log

A large set of variables were calculated based on the CT data, and a few specific measurements were used from the camera-based measurements of the Boardmaster. The CT-based variables $\mathbf{X}$ were used for PLS modelling (D), and the measurements used in the Boardmaster were used to create a few rules to remove outliers (I). The full $\mathbf{X}$ variable set, consisting of approximately 1800 variables, is similar in structure and scope to the set detailed in Olofsson et al. (2019b), but was based on the CT Log measurements described below.

The CT Log measured the scanned log's knots, geometric measurements, heartwood content, heartwood density, sapwood density, log spiral grain, and log pith deviation from a centre line. The CT Log describes each knot with a few different measurements and classes regarding shape, size, and sound or dead status. The measured knots were projected onto the surface of the virtual pieces of sawn timber as the CT Log had intended the sawing. The 2D knot structure described by a large set of various statistical measurements repeated in several sections of the sawn timber and on all the sawn timber faces. The log specific features, together with the 2D knot measurements' calculated variables, were used to calculate the **X** variables for each piece of virtual sawn timber.

Once the data was stored and the **X** variables were calculated, the log was sawn into sawn timber based on the instructions of the CT Log. Due to a shortcoming of the data collection, the data did not capture any inaccuracies, such as positioning or rotational errors, of the sawing process. Due to these errors, the variables based on the CT Log's knot description occasionally does not fully correlate with the material and, by extension, the grades associated with that piece. As such sawing errors are typically quite small, the effect of these errors is estimated to be small but does, however, introduce more noise in the data. Occasionally, a large rotational error, for example, would result in a piece of sawn timber that is not represented well in the CT Log data.

2.4. The Machine Learning Method - PLS

PLS is a simple but powerful machine learning method. PLS uses multivariate linear regression to correlate measured variables of a training dataset **X** with the grade of the same dataset **Y**. The training of the PLS model consists of finding the best weights w_i and biases b_i for maximum prediction accuracy on the training dataset of the true grade y by the model $\hat{y} = b_0 + w_1 x_1 + \ldots + w_N x_N$. For this study, the SIMCA 15 (Anon., 2019) software was used to train two grading models using the X-ray CT variables **X** and either of the product grade $\mathbf{Y_p}$ or image grade $\mathbf{Y_i}$ quality references, respectively. Both models were then used to predict the product grade $\mathbf{Y_p}$ of an independent test set. For a detailed description of a very similar implementation of PLS for product-specific grading in the dry-sorting station, see Olofsson et al. (2019b).

The PLS model outputs an estimated probability that a piece of sawn timber is of the higher grade A relative to the trained data. The actual classification is then a check if this probability passes a certain threshold (called PLS-DA, which is PLS regression implemented for discriminant analysis), by default 50%. If the probability passes the threshold, the sawn timber is assumed to be grade A and graded as the main product. To minimise the volume of grade B sawn timber dried to 14% moisture content, the classification threshold was raised such that only very likely grade A pieces of sawn timber were dedicated to the wall panel customer as the main product.

2.5. Sorting Procedure

Following the data collection and training of the two PLS models (D_1) and (D_2), a simulated sorting of the entire dataset was performed. Three independent simulated sortings were performed, one for the rule-based grading (C) and one for each of the (D_1) and (D_2) PLS models. The simulated sorting was performed using 5-fold cross-validation where the data set was split into 5 parts, and each of the parts took turns being the test set while the remaining 4 parts were the training set. The results presented are then the summed results of these 5 tests and will be referred to as the results of the entire dataset. The product grade was always considered the true grade.

The Scots pine logs selected for this study (A) were passed through the CT Log (B). Based on the 2D knot structure of the virtual sawn timber, a set of customer-adapted rules (C), similar in scope to the NTGR, determined the grade of the virtual sawn timber. Based on this rule-based grade (C), the CT Log optimises and decides the log's positioning during sawing to maximise the yield of sawn timber suitable for the customer – the main product. If the CT Log could not find a way to saw the log to yield sawn timber suitable for the customer, the log would be sawn into the main product for a different customer, i.e. sawn timber graded as B by the rule-based method was sawn with a different indented customer.

Once the sawing of the log was decided, the variables **X** of the virtual sawn timber was calculated, and the PLS model (D_1) or (D_2) predicted the grade of that piece of virtual sawn timber. Following the PLS-based grade, the complementary set of rules (E) was applied. For a fair comparison, the same complementary rules (E) was applied to the rule-based grade after the log had been sawn. The simulated sorting results are presented as if the grading steps were as shown in Figure 1 where the complementary rules help to optimise the sawing process. To reiterate, (C) was the only grading process used to optimise the log's positioning during sawing.

If the virtual sawn timber grade before sawing was determined to be B, i.e. a consequence product, that piece of sawn timber may be dedicated to a different customer as a different main product of the same dimensions. If the virtual sawn-timber grade was determined to be A, the log was sawn, and the sawn timber was dried to 14%. It was possible to dedicate one of the two centre yield pieces of sawn timber to Lundgren's Planing Mill and one piece to another customer.

Using the camera-based Boardmaster dry-sorting system, another set of 7 complementary rules was applied at the dry sorting station. These rules graded the dried sawn timber as the main product or as a consequence product and was the final grading step before delivery to the customer.

2.5.1. Complementary Rules

One of the drawbacks of the multivariate methods used for this study is that some important measurements can become too vague in the PLS model. This is not a problem with the number of variables, as the PLS model is optimised for predictive power and not optimised for the best regression fit of the **X** variables and the grade **Y**. The model is optimised for predictive power based on a training dataset, which is separate from the testing dataset and therefore not biased. An example of a vague variable is the measurement of log-pith deviation from a centre line measured by the CT Log. This measurement is correlated with the sawn-timber grade because a large deviation in the pith could indicate bark inclusion, which often results in wall panels of grade B. Still, since a large measurement of pith deviation does not guarantee bark inclusion, the model can classify the sawn timber as grade A even with a large log pith deviation measurement. A large pith deviation is considered high-risk, and therefore a rule (limit) for this variable, and similarly for a few others, were included for all tests in both the CT Log and the Boardmaster. These measurements are holistic in the sense that they measure e.g. the dead-knot area of the entire face of the sawn timber.

In the CT Log, 6 rules were enforced in all tests as an upper limit to:

- pith deviation.
- three measurements of the dead-knot area on the sapwood face of the sawn timber, and
- two measurements of the dead-knot area on the edge faces of the sawn timber.

In previous studies by Olofsson et al. (2019b,c,d) and Olofsson et al. (2021), an extensive variable set was calculated for similar studies regarding product adapted sawn timber grading using the Boardmaster. In this study, these variables were not needed, and only a very select few variables were used in the Boardmaster. If no customer-adapted sorting is performed by the CT Log, a much more strict set of customer-adapted rules is applied by the Boardmaster. As this study focuses on rule-based vs PLS-based grading in the CT Log, these rules were not active, and only the complementary rules were used.

The Boardmaster was set up to enforce the following 7 rules in all tests as an upper limit to:

- dead-knot area on the edges and sapwood face,

- bark-knot size on the pith face and sapwood face,
- bark-inclusion size of the pith face and sapwood face, and
- no rotten knots was allowed.

These rules for both systems were the same throughout the study for all tests. The limits were determined based on the entire dataset as a limit that would discard obviously poor-quality sawn timber or high-risk pieces. The rule-based grading and the PLS models also failed most of the sawn timber pieces that any of these rules would have failed.

3. Results

To investigate the possibility of dedicating virtual sawn timber to a specific product before sawing the log, using X-ray CT-scanning and machine-learning models, a simulated sawmill refinement process and sorting of 156 logs were investigated. A rule-based approach was compared with two PLS-based machine-learning models using the product-grade or the image-grade as quality reference for training, respectively, when performing the grading of the virtual sawn timber from the CT-scanning data.

The quality distribution of the 303 pieces of sawn timber of this study is shown in Table 1 and is repeated in Tables 2-4 for easy reference. The image grade and the product grade show a very different proportion of grade A sawn timber of the entire dataset.

Table 1. Material quality grade distribution.

	Product-grade	Image-grade
Grade A	189	114
Grade B	114	189
Total	303	303
Grade A (%)	62	38

Tables 2-4 show how the material is sorted throughout the sawmill, and each column shows the quality distribution of the main product dedicated to the customer *after* that step. Graded by the CT Log refers to (C+E) in Figure 1, and graded by the Boardmaster refers to (I). The product-grade A and B rows show the product-grade quality distribution of the main product. The total count shows the number of pieces delivered from the current step to the next step. The ratio of delivered sawn timber shows the ratio of the incoming sawn timber from the previous step that is graded as the main product and is delivered to the next step, which for the all-sawn-timber column always shows 100% as the entire dataset is delivered to the CT Log. The ratio of delivered sawn timber of grade A shows the proportion of the main product delivered to the next step that is of grade A, which for the log intake always show 62% as that is the proportion of product-grade A sawn timber found in the entire material (Table 1) delivered to the CT Log.

Table 2. Sorting outcome using the customer adapted rule-based grading in the CT Log and using the complementary rules in the CT Log and the Boardmaster.

		Graded as A by	
	All sawn timber	CT Log	Boardmaster
Product-grade A	189	159	153
Product-grade B	114	69	47
Total	303	228	200
Delivered (%)	100	75	88
Delivered grade A (%)	62	70	77

Table 3. Sorting outcome using the product-trained PLS model in the CT Log and using the complementary rules in the CT Log and the Boardmaster.

	All sawn timber	Graded as A by	
		CT Log	Boardmaster
Product-grade A	189	102	101
Product-grade B	114	21	13
Total	303	123	114
Delivered (%)	100	41	93
Delivered grade A (%)	62	83	89

Table 4. Sorting outcome using the image-trained PLS model in the CT Log and using the complementary rules in the CT Log and the Boardmaster.

	All sawn timber	Graded as A by	
		CT Log	Boardmaster
Product-grade A	189	105	104
Product-grade B	114	19	11
Total	303	124	115
Delivered (%)	100	41	93
Delivered grade A (%)	62	85	90

Table 5. Sorting outcome specific to the complementary rules (E) in the CT Log for the rule-based grading method and the two PLS-based methods.

	Rejected by complementary rules in		
	Rule-based	Product-trained	Image-trained
Product-grade A	6	3	5
Product-grade B	22	10	16
Total	28	13	21

The results using the product-trained (Table 3) and image-trained (Table 4) PLS-based CT Log sortings are very similar, with 41% of the scanned material delivered from the CT Log to the Boardmaster and both with 93% delivered from the Boardmaster to the customer. The proportion of grade A sawn timber by the image-trained PLS model managed 2%-points higher by the CT Log and 1%-point higher by the Boardmaster than the product-trained PLS model.

The rule-based sorting (Table 2) by the CT Log and the Boardmaster delivered 200 pieces of sawn timber to the customer, while the PLS-based CT Log approaches delivered 114 and 115 pieces of sawn timber, respectively. The higher amount of delivered sawn timber by the rule-based approach comes with a lower proportion of grade A sawn timber at 77% compared to 89% and 90%, and a lower delivered amount from the Boardmaster at 88% compared to 93% when compared to the PLS-based methods.

All of the pieces rejected by the complementary rules shown in Table 5 were considered obviously poor quality sawn timber or high-risk pieces based on visual inspection. The product-trained PLS model required less correction by the complementary rules (E) than the other methods, with only 10 pieces of grade B removed.

4. Discussion

When comparing the overall sorting performance of the current rule-based approach (Table 2) with the PLS-based sorting results (Tables 3 and 4), it is clear that the PLS-based methods would deliver sawn timber more in line with the requirements of the customer. The proportion of grade A sawn timber in the delivered batch from the Boardmaster and the dry-sorting station is about 90% for the PLS-based methods compared to 77% using the rule-based approach. A high proportion of grade A sawn timber is very important for the customer as this means the production process yields a higher proportion of grade A wall panels. Both of the PLS-based methods also managed to deliver a higher proportion of the material through the complementary rules in the Boardmaster than the rule-based approach, 93% vs 88%, respectively. Using either of the PLS-based methods would result in lower proportions of consequence products for the sawmill. The results for the PLS-based methods could be improved by allowing a PLS model to control the sawing optimisation instead of having to sort the sawn timber produced by the rule-based approach. Furthermore, the results could be improved by taking the log's positioning errors mentioned in Section 2.3.1 into account.

The higher proportion of grade A sawn timber achieved by the PLS-based methods comes at the consequence of a lower volume of sawn timber dedicated to the customer. The rule-based sorting in the CT Log dedicated 228 pieces of sawn timber to the customer during sawing compared to 123 and 124 pieces of sawn timber by the PLS-based approaches. This means that a higher amount of logs needs to be sawn using the PLS-based methods to reach the same delivery quota as the rule-based method. During discussions with the sawmill, it was agreed that a minimum of about one-third of the material scanned by the CT Log needs to be dedicated to the customer; otherwise, the drying kilns will not be filled during one working shift. As long as this requirement is met, the lower number of delivered pieces of sawn timber by a strict PLS-based approach in the CT Log is not a problem. Since the PLS models use a classification threshold, the balance of strict sorting vs high volume deliveries is a simple balancing problem controlled by a single parameter, see Olofsson et al. (2019b) for a more detailed discussion on this balancing act.

The use of complementary rules (both in the CT Log (E) (Table 5) and the Boardmaster at the dry-sorting (I) (Tables 2-4)), was deemed necessary as without these rules some obviously poor quality or high-risk sawn timber would have been graded as the main product and delivered to the customer. For the rule-based case, the need for complementary rules indicates how difficult rule-based appearance grading of sawn timber for this customer is. The difficulty comes from the fact that the rules used in (C) are on a per-feature basis, i.e. each knot is evaluated separately, which is one of the problems of customising rules for individual customers according to Lycken and Oja (2006) and Olofsson et al. (2019b). The complementary rules used in (E) use holistic variables, which is more suitable for this customer. The complementary rules are also necessary for the PLS-based methods. This could be due to the limited size of the dataset used in this study and that some of the important variables are difficult to model by linear regression. For example, the most important log feature apart from knots was the log pith deviation. A high value of log pith deviation can indicate top breakage and, therefore, a high risk of bark inclusion. Bark inclusion is almost guaranteed to result in grade B sawn timber. However, since a high value of log pith deviation does not guarantee bark inclusion, the variable is difficult to model by PLS.

One problem with using both the CT Log and the Boardmaster to grade the sawn timber is the possibility to introduce a disagreement between the two systems, as discussed in more detail by Olofsson et al. (2019a). A disagreement could result in sawn timber being dried to 14% and not be delivered to the customer, resulting in consequence products for the sawmill; while an agreement between the CT Log and the Boardmaster would mean that only sawn timber suitable for the customer would be dried 14% and delivered to the customer.

The similar sorting performance by the product-grade trained model (Table 3) and the image-grade trained model (Table 4) shows that an image-based reference for the PLS model training is at least as good as using the product-grade reference. This equivalence of reference grades was previously found for product-adapted dry-sorting by Olofsson et al. (2021). As the image-grade reference is much easier to acquire than the product-grade reference, the difficulty of gathering data for similar machine learning studies could potentially be drastically lower. The product grade and image grade were quite different (Table 1). The image grade determined by the two researchers was much stricter than the actual product grade. However, using a classification threshold (Olofsson et al., 2019b) allows for simple control over this bias with a single parameter. During discussions with the customer, the customer has stated that there are various subjective appearances within the grade A wall panel class and that acceptable and desirable are not always the same thing. This could explain the difference of opinion between the researchers and the product-grade – the researchers might have classified the images as desirable or not while the product-grade indicates acceptable or not. Another difficulty with using the sawn timber images as a reference is splitting the sawn timber into three panels, effectively leaving one wall panel hidden inside the piece. It is challenging to estimate how surface defects affect this hidden wall panel in the middle of the piece. The researchers were possibly overly cautious regarding such surface defects. The high proportion of grade A sawn timber delivered from the CT Log using the image-trained PLS model shows that even if subjective grading of the material is a bit ambiguous, and due to the splitting even incomplete, calibrating the classification threshold results in a sorting outcome on-par with the product-trained PLS model.

An important implication of the usefulness of images of sawn timber as a quality reference is their reusability. A database of sawn timber images could be used to create an appearance quality reference for many different customers independent of their drying requirements. The quality reference could be used to train individual machine learning models per customer, similarly to the image-trained PLS model used in this study. This could result in several models giving suggestions on how to saw a log after CT scanning. The sawmill can select the sawing pattern to be used based on material availability (supply), delivery requirements (demand), and price. Based on the 93% delivered material from the Boardmaster for the PLS-based methods, it seems plausible to only perform one appearance grading in the CT Log. Removing the dry-sorting entirely would for the image-trained PLS model mean that the CT Log column in Table 4 would be delivered directly to the customer after drying. The customer would then receive 83% grade A sawn timber in the delivery, which is still higher than the rule-based approach including the dry-sorting at 77% (Table 2).

Future studies should investigate a better way of handling rare but obviously poor quality or high-risk sawn timber. Their rarity makes them difficult to model by a machine learning method, and the use of complementary rules, as in this study, is a product-specific solution. Removing such high-risk pieces before sawing using CT-scanning could make it possible to skip the dry-sorting entirely.

Future studies should also investigate the possibility to simulate the splitting process by splitting the virtual sawn timber pieces into approximate wall panels to see if this would improve the grading results, thereby further utilising the 3D data available from the CT Log.

5. Conclusions

This study shows that a machine learning method (PLS) can be used to appearance grade sawn timber for a specific product before sawing using X-ray CT scanning and thereby plan the entire refinement process from the CT scanning of the log to the finished product. Furthermore, this machine learning approach's product-adapted aspect can be implemented without meticulous logistics and material processing to create a training dataset. The results show that a PLS model trained using images of the sawn timber acquired from the sawmill's dry-sorting station as the quality reference is as good as, or even slightly outperforms, an otherwise identical implementation using the finished product grade as training reference. The benefits of using a machine learning method trained using images as a reference are the simplicity of capturing the subjective and holistic quality requirements of a customer. This study indicates the future possibility for an image data bank of sawn timber being used by several customers to make product-specific quality references for the training of machine learning models. Such models could enable product-specific sawing and drying and possibly remove the need for dry-sorting for some customers.

Acknowledgements

Financial support from the Swedish Innovation Agency (Vinnova), project Sawmill 4.0 – Customised flexible sawmill production by integrating data-driven models and decisions tools 2018-02749, is gratefully acknowledged. The authors also gratefully acknowledge the support of the CT-Wood – a centre of excellence at Luleå University of Technology for the development of X-ray computed tomography applications for use in the forest products industry.

References

Anon. (2018). *Boardmaster.* https://finscan.fi/products/boardmaster/?lang=en. (2018-06-12). Note: older predecessor to BoardmasterNOVA.

Anon. (2019). *SIMCA.* https://umetrics.com/products/simca. (2019-01-03).

Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chemica Acta*, 185, 1–17.

Lycken, A. and Oja, J. (2006). A multivariate approach to automatic grading of *Pinus sylvestirs* sawn timber. *Scandinavian Journal of Forest Research*, 21(2), 167–174.

Olofsson, L., Broman, O., Oja, J. and Sandberg, D. (2021). The effect of class-balance and class-overlap in the training set for multivariate and product-adapted grading of Scots pine sawn timber. *Wood Material Science & Engineering*, 16(1), 58–63.

Olofsson, L., Möller, C.-J., Wendel, C., Oja, J. and Broman, O. (2019a). "New possibilities with CT scanning in the forest value chain". *International Nondestructive Testing and Evaluation of Wood Symposium.*

Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019b). Multivariate product adapted grading of Scots pine sawn timber for an industrial customer, part 1: Method development. *Wood Material Science & Engineering*, 428–436.

Olofsson, L., Broman, O., Skog, J., Fredriksson, M. and Sandberg, D. (2019c). Multivariate product adapted grading of Scots Pine sawn timber for an industrial customer, part 2: Robustness to disturbances. *Wood Material Science & Engineering*, 420–427.

Olofsson, L., Broman, O. and Sandberg, D. (2019d). “Holistic-subjective automatic grading of sawn timber: Sensitivity to systematic changes.” *IWMS 24. August 25-30, 2019, Corvallis, Oregon, USA*. Oregon State University (OSU), pp. 154–167.

Sandberg, D. and Teischinger, A. (2021). Chap. 22 Sawmilling. In: Springer Handbook in Wood Science and Technology. In print. Springer. Verlag, Heidelberg.

Swedish Sawmill Managers Association (1994). Nordic timber: grading rules for pine (*Pinus sylvestris*) and spruce (*Picea abies*) sawn timber: commercial grading based on evaluation of the four sides of sawn timber. Föreningen Svenska Sågverksmän: Markaryd.

Ursella, E., Giudiceandrea, F. and Boschetti, M. (2018). “A Fast and Continuous CT scanner for the optimization of logs in a sawmill”. *8th Conference on Industrial Computed Tomography (iCT 2018) at Wels, Austria*. Vol. 2.

Wold, S., Sjöström, M. and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), 109–130.

Department of Engineering Sciences and Mathematics
Division of Wood Science and Engineering

ISSN 1402-1544
ISBN 978-91-7790-874-6 (print)
ISBN 978-91-7790-875-3 (pdf)

Luleå University of Technology 2021

FSC
www.fsc.org
MIX
Papper från ansvarsfulla källor
FSC® C115359

SVANENMÄRKET
Miljömärkt trycksak
3041 0145

Print: Lenanders Grafiska, 136683

LULEÅ
UNIVERSITY
OF TECHNOLOGY