Comparison between automatic and manual quality grading of sawn softwood

Anders Lycken

Abstract
An investigation has been performed to determine the repeatability and the accuracy of automatic and manual grading in a sawmill's final grading station. In total, approximately 400 boards of pine (Pinus sylvestris L) and spruce (Picea abies) of different dimensions are included in the study. The most common automatic grading system for sawmills in the Nordic countries is studied. The study shows that in many cases it is possible to replace manual graders with an automatic system and, at the same time, make production more effective and improve grading, which results in higher value yield. The value yield produced by the automatic system is over 90 percent (91% to 98%) for both boards and planks for both species. The manual grader reached over 92 percent value yield for planks, but only 83 percent for boards. The quality yield for the automatic system is between 52 and 75 percent. The manual grader reached between 31 and 61 percent. The repeatability for the value yield varied between 85 and 96 percent for the manual grader and between 85 and 94 percent for the system. An overshadowing problem in this kind of study is how to obtain an irrefutably true grading result, as all grading of wood is subjective by nature.

In connection to the introduction of systems for automatic grading and sorting in sawmills, it is becoming more and more interesting to know how well human graders can stand up to the competition from automatic systems. In the Nordic countries, automatic systems for sorting and grading have become common in secondary processing, where they function as cutting optimizers, and in sawmills, where they function as green sorters. Functional systems for final grading in sawmills have not been commercially available until the last few years. Depending on wood species, type of defects, grading rules, user interface, and other demands, the time from installation to full production can take from a few days up to several months, as the problems of fine tuning the systems vary.

By tradition, there has normally not been any quality grading in the green sorting station, the sorting of undried sawn wood. The sorting performed on green timber has been based only on dimension, i.e., thickness, width, and sometimes length. Manual quality grading is very seldom performed on green timber. By quality grading green timber, it is possible to direct the lumber to the right kiln dryer and drying program according to quality. The final quality grading is performed in the final sorting station. The reasons for investing in automatic grading systems for final grading vary from sawmill to sawmill. The main reasons are to make grading more homogenous and to create a more flexible way to set and maintain the many complicated grading rules. The grading rules are mostly based on Nordic Timber rules (Anon. 1994) or the older rules, called the Green Book (Anon. 1980). Sawmills have their own interpretation of the rules according to customer needs. After grading, the boards are stamped with suitable shipping marks.

The final grading takes place after drying of the timber. Each board is fed in front of the grader, or into the grading system. The boards are inspected on all four sides, and the features such as knots, cracks, wane, etc., are recognized and classified. A decision on economical revenue in proportion to length and quality has to be taken before the next board is fed into view. The grader or the grading system has to decide on cutting position in both ends of the board and the quality of the remaining board to make the most economical choice for each board. For the manual grader, the time for inspection and decision is approximately 2 to 3 seconds per board. Some automatic systems are able to run at speeds of up to 180 boards per minute, which is up to 9 times faster than the human grader.

A number of studies have been performed on grading accuracy, but only a few have been published. There is only a little information on appearance grading of sawn softwood in the sawmill. The main part of the published material is about hardwood grading and optimization or secondary processing (Huber et al. 1985; Klinkhachorn et al. 1989; Huber et al. 1990; Regalado et al. 1992a, 1992b; Kline et al. 2000; Buehl-

The studies regarding grading accuracy, performed more or less regularly at the sawmills, are often very simple and mostly consist of checking a number of ready-made packages to see that the graders are at the same level and that not too many pieces are off grade. The Nordic Timber grading rules state, “at least 90% of the pieces in the batch shall not have wood features with values exceeding the maximum permitted values for a contracted GRADE” (Anon. 1994). By checking the lumber this way, it is impossible to see the real cost of grading errors. That is because the cutoff pieces are chipped and the possibility to see an alternative cutting position starting from the whole board is wasted. If the board is uncut at the time of judging, it is possible to evaluate whether the grader’s decision is accurate and optimal.

When discussing the matter of grading error with sawmill managers and personnel, the general impression is that a good grader is correct up to approximately 80 percent over time and a normal grader around 60 to 70 percent, which is in line with other studies, where rough mill employees performed at 68 percent of perfect (Huber et al. 1985) and where value recovery was between 62 and 78 percent (Regalado et al. 1992a).

As it is with almost every test situation, the problem with “correct” and subjectivity arises when grading lumber. Subjectivity has been a norm in the older Swedish grading rules, which stated that “A piece having a minor such defect as would place it in a lower grade may, nevertheless, remain in the higher grade if, in respect of other defects, it can be classed among the best pieces within that grade” (Anon. 1980). The problem with borderline defects has been reported in Kline et al. (2000) and Lycken (2000), while the problem with subjectivity and exactitude is expressed in Kline et al. (2003) and Huber et al. (1990).

**Objective**

The objective of the study was to examine a way to determine the accuracy and repeatability of automatic grading systems for final grading of dried lumber in sawmills and to compare one of the existing systems with manual graders. By doing that, it is possible to set the right demands when planning for investment in grading systems as well as to point out where system development needs an extra push. This paper does not cover the issue of investment costs, but it is intended to be a help to calculate the cost of grading differences between automatic and manual grading.

**Materials and methods**

In this study, one system for automatic grading was investigated and compared to a manual grader. The tests were performed at a Swedish sawmill during the spring of 2003. The system has transverse feeding and needs a turning device to be able to scan all four sides of the board. In one scanning station one face and the two edges are scanned; in the other scanning station the other face side is scanned. The system is installed as a final grading system in line with three manual graders as a complement. When the sawmill is confident in the function and all the different grading rules have been tested and tuned, the manual graders will be relieved.

**Material**

All material was taken from the running production. In the study, 100 boards of spruce (Picea abies) dimensioned 50 by 125 mm, 100 boards of spruce dimensioned 22 by 100 mm, 100 boards of pine (Pinus sylvestris) dimensioned 63 by 125 mm and 100 boards of pine dimensioned 25 by 125 mm were used. Due to some difficulties during the test, not all boards made it to the end. The exact number in each test group is shown in Table 1. In addition to that, 20 boards of pine 25 by 175 mm were used for repetition tests.

The moisture content was 18 percent for all boards.

**Grading rules**

The grading rules used in the test were rules for appearance grading similar to the Nordic Timber grading rules (Anon. 1994) in the following sense: the allowed defect sizes are, according to the dimension of the lumber, defined in millimeters or part of board width, thickness, or length, and the grading system is hierarchical (Lycken 2000), i.e., a lower grade allows defects the same size or larger than those of a better grade. The intention of Nordic Timber is that “Numerical limit values have been given to all measurable features” (Anon. 1994). The number of grades in Nordic Timber is up to seven. In the present case, four different grades were used. Generally, the more grades used, the harder it is to grade the pieces correctly (Lycken 2000). The grade of a board is “determined on the basis of the number, location, type, and the maximum permitted values of the wood features according to the tables” (Anon. 1994). In addition, the Nordic Timber rules contain the following directions (Anon. 1994, p. 21):

- “Each side of the piece shall be graded separately
- The maximum values of wood features which in each GRADE are permitted on the worst one meter of length are given in the tables
- The GRADE is decided on the basis of the outside face and both edges
- The inside face may be one GRADE lower”

Table 2 shows the allowed knot sizes for the different grades and dimensions in Nordic Timber. The rules for other features are defined in similar ways. The actual limits used in the test are company secrets. The boards may be cut at both ends to eliminate downgrading defects in order to increase the quality of the remaining board and maximize value. No actual cuts were made during the test, but the cutting positions were registered. Up to a 15-cm difference between runs and between actual and optimal cut is considered as acceptable.
The manual grader in operation was 20 boards per minute. With one board on every third carrier, the speed for manual grading chairs was approximately 60 carriers per minute. The grading rules were supposed to be the same for both the system and the manual grading, but the grades actually used by the system manufacturer, and the test leader. The decisions were taken unanimously. The use of this method opens possibilities to say if the system was right or wrong in its decision, and also why. In the attempt to establish the irrefutable truth, all help was used, such as written grading rules, caliper, measuring tape, etc. As stated earlier, grading is always subjective in some ways. On almost every board there is at least one defect that "might" downgrade the board, but other factors will affect that decision. In Lycken (2000), examples are shown of how borderline defects can change the quality yield by as much as 15 to 20 percent, depending on how they are treated. The value yield does not vary as much.

The assumption of the test was that borderline defects were to be treated in favor of the system, which was done in most cases. The manual grading did not enjoy that favor to the same extent. In test run 2, no favor was given.

Calculations

To calculate the repeatability of the runs, the following formulas were used:

The repeatability, $R_v$, for the value of the batch is calculated as:

$$ R_v = \frac{1}{m} \sum_{k=1}^{m} r_{vk} $$  \hspace{1cm} (1)

where $k = \text{number of the board}; m = \text{total amount of boards}; r_v = \text{repeatability for individual board}.$

$$ r_v = 1 - s/M $$  \hspace{1cm} (2)

where $s = \text{standard deviation for the value for each individual board}; M = \text{mean value for the value for each individual board}.$

The repeatability, $R_D$, for length, quality, and defect detection can be calculated as:

$$ R_D = \frac{1}{m} \sum_{l=1}^{m} R_{dl} $$  \hspace{1cm} (3)

where

$$ R_{dl} = \frac{1}{m(n-1)} \sum_{k=1}^{n} \sum_{j=1}^{n} \delta_{ijk} \times 100\% $$  \hspace{1cm} (4)

where $\delta_{ijk} = \begin{cases} 1 & \text{if } d_{jk} = d_{ij} \text{ and } j \neq k \\ 0 & \text{otherwise} \end{cases}$

where $R_{dl} = \text{repeatability for individual board}; d = \text{individual board}; d = \text{quality (length, defect, etc.); } f = \text{run number}; k = \text{run number}; l = \text{board number}; m = \text{total amount of boards}; n = \text{total amount of runs}.$

When calculating repeatability, the correctness of the grading decision is not considered. A 100 percent repeatability level can be reached if all boards in a test are given the same grade, length, and value, regardless of whether everything is totally wrong.

The value yield, $Y_v$, for each board is calculated as:

$$ Y_v = \frac{O - |V - O|}{O} $$  \hspace{1cm} (5)

which is the normal speed for a manual grader. The automatic system was able to run up to 140 boards per minute.

To get optimal grading, the boards were graded manually between run 1 and run 2 by the sawmill’s head grader, one other grader, the system manufacturer, and the test leader. The grading test can be divided into four components:

1. Grading accuracy for manual grading at production speed;
2. Grading accuracy for the automatic grading system;
3. Repeatability for the manual graders at production speed;
4. Repeatability for the automatic grading system.

All 420 pieces were numbered, fed through the system, and in front of the manual grader. This was repeated once for the planks, so they were graded twice. The 20 boards of pine 25 by 175 mm were run 3 times through the system and in front of the manual grader. This was repeated once for the planks, so they were graded twice. The 20 boards of pine in front of the system behaved as it did. The system is designed to recognize defects such as knots of different types, cracks, scar, wane, and bark.

The grading rules were supposed to be the same for both the system and the manual grading, but the rules actually used by the manual grader is impossible to know, as the ones used in practice always are interpretations of the written rules.

Table 2. — Allowed knot sizes in Nordic Timber grading. The allowed sizes in the test were similar, but not exactly the same.

<table>
<thead>
<tr>
<th>Type of knot</th>
<th>Reduced to percent of sound knot size</th>
<th>Other knot</th>
<th>Reduced to percent of sound knot size</th>
<th>Dead knot</th>
<th>Reduced to percent of sound knot size</th>
<th>Barkringed knot</th>
<th>Reduced to percent of sound knot size</th>
<th>Unsound knot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face 16 to 19</td>
<td>15</td>
<td>70</td>
<td>70</td>
<td>100</td>
<td>50</td>
<td>60</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Face 22 to 25</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>55</td>
<td>45</td>
<td>60</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Face 32 to 38</td>
<td>30</td>
<td>30</td>
<td>50</td>
<td>45</td>
<td>40</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Face 44 to 50</td>
<td>40</td>
<td>35</td>
<td>50</td>
<td>60</td>
<td>35</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Face 63 to 75</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Edges 16 to 19</td>
<td>15</td>
<td>70</td>
<td>70</td>
<td>100</td>
<td>50</td>
<td>60</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Edges 22 to 25</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>55</td>
<td>45</td>
<td>60</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Edges 32 to 38</td>
<td>30</td>
<td>30</td>
<td>50</td>
<td>45</td>
<td>40</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Edges 44 to 50</td>
<td>40</td>
<td>35</td>
<td>50</td>
<td>60</td>
<td>35</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Edges 63 to 75</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

- Asterisk indicates a knot size equal to the thickness of the timber.

Methods

The grading test can be divided into four components:

1. Grading accuracy for manual grading at production speed;
2. Grading accuracy for the automatic grading system;
3. Repeatability for the manual graders at production speed;
4. Repeatability for the automatic grading system.

As stated earlier, grading is always subjective. The grading test was divided into four components:

1. Grading accuracy for manual grading at production speed;
2. Grading accuracy for the automatic grading system;
3. Repeatability for the manual graders at production speed;
4. Repeatability for the automatic grading system.

All 420 pieces were numbered, fed through the system, and in front of the manual grader. This was repeated once for the planks, so they were graded twice. The 20 boards of pine 25 by 175 mm were run 3 times through the system and in front of the manual grader. "Free lengths" were used in the test, which means that cuts are allowed at any length. No actual cuts were made. In the system, all defect files were saved for all boards. By this means, it is possible to check the grading and see why the system behaved as it did. The system is designed to recognize defects such as knots of different types, cracks, scar, wane, and bark.

The grading rules were supposed to be the same for both the system and the manual grading, but the rules actually used by the manual grader are impossible to know, as the ones used in practice always are interpretations of the written rules.

The feed speed through the system and in front of the three manual grading chairs was approximately 60 carriers per minute. With one board on every third carrier, the speed for the manual grader in operation was 20 boards per minute.
where $O$ = optimal value; $V$ = assumed value of the board (automatic or manual grading); $|V-O|$ = the unsigned difference between the two values.

Every board that is classified different from the optimal receives a value yield less than optimal, regardless of whether the assumed value is higher or lower than optimal. The value yield can thus never be more than 100 percent, and the direction of the error is insignificant.

The value yield for the batch, $Y_b$, is calculated as the average of the boards' yields:

$$ Y_b = \frac{1}{m} \sum_{i=1}^{m} Y_i $$

where $Y_i$ = value yield for each board; $l$ = board number; $m$ = total amount of boards.

The quality yield for a batch is calculated as the amount of correctly graded boards divided by the total amount of boards. A board is regarded as correct quality if it is of the same grade and is the same length as or shorter than the optimal length. If a higher quality board is placed in a lower grade, it is considered incorrectly graded, as is the reverse. In this study, an approved board has to strictly be on grade.

Quality yield = Number of approach boards/ Total number of boards

### Results

As shown in Figure 1, the value yield for the automatic grading system in run 1 was between 91 and 98 percent, depending on species and dimension. For manual grading, the value yield varied between 83 and 92 percent. In run 2, the value yield was between 87 and 92 percent for the automatic grading system and between 82 and 88 percent for the manual grader.

Figure 2 shows that the quality yield is much lower for the manual grader than for the automatic system except for pine 63 by 125 mm in run 2. The quality yield is also lower than the value yield. In run 1 the quality yield for the automatic grading system varied between 66 percent and 93 percent. For manual grading, the quality yield varied between 36 percent and 60 percent. In run 2 the variation was 52 percent to 54 percent for the automatic grading system and 31 percent to 61 percent for the manual grader.

It turned out that there were disproportionately more oil stains on the boards in this study, compared to boards in normal production, due to handling procedures during the test. It also turned out that some boards had been graded wrong by the system due to rot, both actual rot and "ghost" rot. As the system was not fully trimmed to distinguish rot from other features, such as heartwood and annual rings, an adjustment was made in which the boards were graded without taking rot and extra oil into consideration. The results from the adjusted gradings are 2 to 3 percent better than the results presented in Figure 1.

Figure 3 shows the results of the repeatability test for different parameters for the automatic grading system and the manual grader. The value yield varied between 85 and 96 percent for the manual grader and between 85 and 94 percent for the automatic grading system. The automatic grading system had 80 percent of the 20 boards of pine 25 by 175 mm in the same quality all three runs; 90 percent were in the same quality in 2 out of 3 runs. The manual grader had 90 percent of the boards in the same quality all three runs; 100 percent were in the same quality 2 out of 3 runs.

### Discussion

This study shows that in many cases it is possible to replace the manual grader with an automatic system. The output will probably not be the same, but not necessarily wrong. This study is to be used as a basis for discussion about how good...
Automatic grading is compared to manual grading. During the test, it was obvious that the choice of test material matters to a great extent. A "difficult" material with a lot of hard-to-define knots will probably give other results compared to an "easy" material. In this test, we tried to get a representative sample of boards from the ordinary production, but found that due to drying at higher temperature than normal, the resin in the knots formed drops that could be mistaken for knots. This was to some extent taken into consideration in calculating the result. Another factor in run 2 is that the boards were mistreated during handling and "ghost" defects such as dirt, oil stains, and chain marks were introduced, which made the grading results appear worse than they really were. In real production, the handling is gentler and the number of ghost defects is smaller.

After the test, the manufacturer of the automatic grading system reprogrammed the defect-detection algorithm, and the boards were graded in a simulation. This time the value repeatability for pine 63 by 125 mm was close to 100 percent, and quality repeatability was 97 percent.

As seen in Figures 1 and 2, the value yield can differ considerably from the quality yield. That is because the quality yield is binary, 0 or 1, for each piece of lumber, whereas the value yield is analogue, 0 to 100 percent. The assumed grade for a board can be B of a certain length, while the optimum is grade A of a shorter length. For that board, the value yield can be well over 90 percent, but the quality yield is 0 percent.

When comparing systems for grading, it is important to choose the right grading rules and to be aware of what the system is supposed to manage. The test method must have defined whether all defects included in the rules are to be regarded, or only the ones the system is able to detect. If defects that the system is unable to detect are visible on the board, the test method must define whether such defects are to be ignored or recognized. It is possible to make two calculations of the performance: one "relative" that states the system performance in relation to what it is supposed to do, and one "absolute" that states the system performance without considering the system's limitations. As wood grading is "subjective" work, it is important that all persons involved in the test be aware of the meaning of the grading rules. One of the most crucial points in doing tests like these is to establish an indisputable truth, or at least a trustworthy solution, that all participants can agree upon.

The automatic system in this test is doing very well compared to the manual grader when compared in a production line with ordinary grading rules and normal material to grade. The manual grader is able to see and classify unusual "new" defects better than the system. The automatic system needs to have every defect type specified, whereas the human grader may make a decision based on common sense. If the lumber contains many boards with defects that are hard to define, automation can be a difficult task. If, on the other hand, the properties of the lumber are well defined and the grading rules are defined, tested, and approved, the automatic system can replace the often tiresome and monotonous work of manual grading with a very good result.

Problems might arise when a small temporary change in the rules has to be made. It is easy to tell a grader to "allow slightly larger knots on the edges but not so much wane" if you need a change just for a small amount of lumber. In some automatic systems it is necessary to make a definition of a new grade, which might take a long time if the user interface is not fully developed to be user friendly. On the other hand, grading rules are often very complicated and difficult to simplify. If all degrees of freedom are to be allowed, the system is hard to grasp thoroughly. A general simplification of the grading rules would also make the interface easier to simplify.

After a change in the grading rules, it is not certain that the human grader will stay on the right track. It is very easy to fall back into old routines and thoughts, so a change might not be in effect for very long. Industrial experience shows that human graders tend to produce similar quality distribution regardless of incoming material, which means that when the timber is of "good quality," the "lower" grades are better than when the timber has a lot of defects.

A change in the system's rules will stay in effect until the next change. By using automatic grading systems, it is possible to collect data and information about the material and processes for control of raw material, yield, and machine behavior.

As with all automatic systems, the need for human supervision is not excluded. There have to be routines for cleaning and maintenance. It is not possible to "set and forget" a system.

Conclusion

According to these results, it is possible to replace manual grading with an automatic grading system. In the tests, the automatic system slightly outperformed the human graders in value repeatability, quality yield, and value yield for spruce planks, while the manual grading was better for pine planks. The differences were larger for the boards, both spruce and pine, where the system was much more accurate than the manual grader.

It is easier to automatically grade spruce than pine. The study shows, however, that an adjustment of the defect detection algorithm can make the system almost 100 percent correct even for pine planks. What was not analyzed in this study was what happens if timber with a different appearance is to be graded. The robustness of the system was never tested. More tests need to be done to verify the accuracy and robustness over a longer time.

One of the main problems involved in grading tests is to establish the irrefutable grading truth. To accomplish that, great thoroughness is required.

Literature cited


Holz als Roh- und Werkstoff 53:413-420.

Hagman, P.O.G. 1996. On reflections of wood - Wood quality features modelled by means of multivariate image projections to latent struc-


