

Effectiveness and subjectivity of visual inspection as a method to assess bending stiffness and strength of chestnut elements

Hélder S. Sousa^{1, a}, Jorge M. Branco^{1, b} and Paulo B. Lourenço^{1, c}

¹University of Minho, ISISE - Dept. of Civil Engineering, Azurém, 4800-058 Guimarães, Portugal

^ahssousa@civil.uminho.pt, ^bjbranco@civil.uminho.pt, ^cpbl@civil.uminho.pt

Keywords: chestnut timber; visual inspection; visual strength classes; variation analysis; subjectivity evaluation.

Abstract. Besides the difficulty of assessing an existing timber structure on site, the efficiency and accuracy of visual inspection is often compromised by its subjective nature inherited by the level of expertise of the inspector. This often leads to conservative predictions of the mechanical properties, even with the use of specific visual grading norms.

The main objectives of this work are to assess the effectiveness of visual inspection as to define different classes of strength and stiffness and to provide a statistical analysis on its subjectivity. For that aim, visual inspection using Italian standard UNI 11119:2004 and bending tests of 20 old chestnut beams (*Castanea sativa* Mill.) at different scales, were carried out. Comparisons and effectiveness of visual inspection is analyzed within and between different scales, and also regarding different level of expertise of inspectors. The results evidence similar percentages of segments classified with higher and lower visual grades and proved to be a good qualitative indicator of bending strength between sawn beams. For the most experienced inspectors, an overall 42% accuracy was found with a better differentiation between visual classes, whereas lower level inspectors scored approximately 5% less. Lower level inspectors evidenced higher concentration of values around a higher mean for each class denoting a more conservative approach. Regardless of the inspector level, knot size was the main limiting visual parameter with higher influence in small scales of the timber elements. When studying the characterization of a single knot, coefficients of variation of 15.7% and 21.8% were found for measuring the minimum and maximum diameter. Bayesian probability networks were considered as to individually assess the accuracy in stiffness prediction of different level of inspectors, and by combination of their information, evidencing that parallel combination for prior information may allow the increase in visual inspection accuracy.

Introduction

The result of visual strength grading is the attribution of indicative values or allowable stress levels, for prediction of key properties, which allow for the mechanical characterization of the material. However, besides the difficulty of assessing an existing timber structure on site, the efficiency and accuracy of visual inspection is often compromised by its subjective nature inherited by the different levels of expertise of inspectors. Measurement uncertainty arises from random or systematic errors made by different inspectors when identifying and characterizing defects that limit the performance of the timber element. To overcome those errors, guidelines for visual grading have been proposed in the past decades with subsequent national standards. Comparative studies for visual strength grading standards of different countries has been addressed (e.g. [1,2]). However without analysis of uncertainty between inspectors, with different expertise, when using the same standard. Visual inspection is subjective and may be inaccurate as evidenced in [3], when inspectors achieved only 68% of the expected result in recognizing, locating and identifying defects in grading red oak lumber. Also in [4], the quality yield for the automatic system was found to be between 52% and 75%, while manual grader only reached between 31% and 61%, comparing to the optimal grade defined by decision of several manual graders. Human visual inspection rarely achieves better than 70% performance in grading lumber [5]. While in [6] it was found that when grading boards into four grades, only 60% were assigned the same grade by two different expert inspectors.

The main objectives of this work are to assess the effectiveness of visual inspection as a method to define different classes of bending strength and stiffness and to provide a statistical analysis on the evaluation of the subjectivity of this method, based on the effective wood performance. For that aim, visual inspection following Italian standard UNI 11119:2004 [7] and bending tests of 20 old chestnut beams (*Castanea sativa* Mill.) at different scales were carried out. Inspectors with different levels of expertise were considered in the subjectivity analysis and its accuracy was measured.

Experimental Campaign

Aiming to assess bending stiffness and strength using visual inspection, 20 chestnut (*Castanea sativa* Mill.) beams were visually graded and the results were compared to 4-point bending test results. The more than a century old timber beams were rescued from a building in Northern Portugal, where they served as structural floor beams. The experimental campaign was divided into three main phases, corresponding to different scales of the timber members. From one to the next phase, the timber elements were sawn into smaller sizes in order to isolate the influence and location of defects, and also to provide a better definition of the distribution of stiffness and strength along the length and height of the beam. The main experimental phases correspond to the members: (i) in the initial state of conservation as they were in the building; (ii) after being sawn to beams with $7 \times 15 \times 300 \text{ cm}^3$ dimension; and (iii) after being sawn to boards with $7 \times 4 \times 300 \text{ cm}^3$ dimension. In each phase, the members were visually inspected and graded on each 40 cm segment, using UNI 11119:2004 [7]. This standard establishes objectives, procedures and requirements for the diagnosis of the state of conservation and estimates nominal stiffness and strength values for structural wood elements. For strength grading of a single element, it considers three classes (I, II and III) regarding on site diagnosis. The wood element is from a given class if it fulfills all the imposed requirements, otherwise, in this study, graded as non-classifiable (NC). Although the standard bases its grading on the evaluation of the critical zone considered by the effective loading of the element, which would correspond in bent members to the tension side of the beam's central third, in this study all segments were considered as to provide a larger size sample with broader variety of defects. Therefore, visual grading was obtained by segment, rather than on an element level. Visual grading was performed by several inspectors divided by level of expertise and experience. Level 1 inspectors (L1), have prior experience in visual grading, both to softwood and hardwood species, with different grading norms and also experience in mechanical testing. Level 2 inspectors (L2) have experience in mechanical testing but have less practice in visual grading of hardwoods. Level 3 inspectors (L3) correspond to individuals with low previous experience in both visual grading and mechanical testing. A total of 336, 616 and 63 inspections to segments were made, respectively by L1, L2 and L3 inspectors.

The sawn beams and boards were submitted to 4-point bending test according to EN408:2003 [8], obtaining local ($E_{m,l}$) and global ($E_{m,g}$) modulus of elasticity in bending and bending strength (f_m). Further detail on the experimental campaign sequence and results has been presented in [9].

Effectiveness and Accuracy

Effectiveness and accuracy of visual grading was considered by analysis of the percentage of segments attributed to a given visual grade comparing to the mechanical tests results. This analysis compiled the comparative study between visual inspection of L1 inspectors in different scales (test phases) and the visual inspection of L1 and L2 inspectors made to a sample of sawn boards.

Different Scales. For the old beams, visual inspection was essential for determining the most damaged zones which were mainly concentrated in the extremities of the beams (with biological decay and cracks), and/or in the central span zones (knots with relevant size and drying cracks). With the purpose of a more adequate comparison with the following test phases, the possibility to discard the parameters influenced by external damage due to structural or environmental reasons was considered and a residual cross section was assumed. This allowed to obtain the mechanical properties related to the material itself and its defects rather than to the state of conservation. The results of visual grading in different scales are presented in Fig. 1, with consideration to the

percentage of segments classified in each visual class. The disregard of external damage in old beams led to a significant higher percentage of class I segments and decrease of class III and NC segments, evidencing the need to define properly the residual cross section. The following phases present similar percentages for intermediate classes, however different in extreme classes (lower class I percentage and higher NC percentage). This indicates that decreasing the cross-section size produced a decrease in visual grading, especially when the size of knots was the limiting parameter.

Based in a qualitative analysis, visual inspection proved to assess f_m efficiently in sawn beams. Considering the sample of sawn beams taken to failure, a higher f_m was obtained for the beam with higher percentage of class I and II segments, while a lower f_m was obtained for the beam with higher percentage of NC segments. In all cases, beam failure initiated in a segment previously identified as a critical section where higher percentage of defects were accounted (see also [9]).

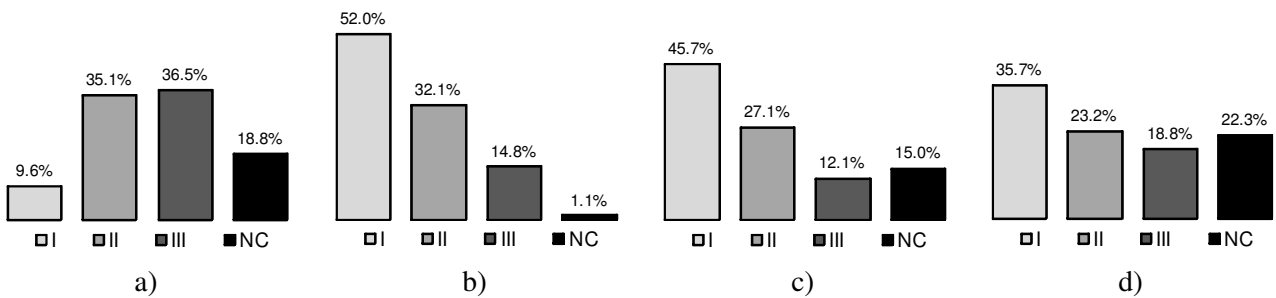


Figure 1: Visual grading classes percentage for full sample in different scale: a) old beams; b) old beams with no consideration of external damage; c) sawn beams; d) sawn boards.

Different Level Inspectors. A random sample of 22 sawn boards was considered for visual inspection by L1 and L2 inspectors, corresponding to a total of 154 segments. For each segment $E_{m,l}$ and $E_{m,g}$ were determined by mechanical testing, and 52 segments, representative of different distribution of defects, were tested for f_m quantification. Visual inspection was accounted only for bottom and top faces, as the height of the lateral faces was too small compared to the knots size and often led to NC segments which, however, still performed efficiently in mechanical tests, and accounting that, in bending, a defect on the side under tension is more important for the analysis.

Sample Results. For the analyzed sample, L1 inspectors evidences a higher percentage of class I than L2 inspectors, while the latter present more III and NC segments. According to the different inspectors grading, the results of mechanical tests were sorted and averaged by grading classes (see Fig. 2, Fig. 3). Results from L1 inspector evidence a larger variation between mean values of visual classes for both stiffness and strength compared to L2 inspectors, and also for all cases a decrease is visible from higher to lower visual grades. The variation in intermediate classes for L2 inspectors is less defined both in bending stiffness and strength. This evidences that L2 inspectors can more easily identify and group clear segments (or with small defects) and segments with significant defects, however encountering difficulties in assessing medium size defects. The coefficients of variation for stiffness and strength within each visual class, for both L1 and L2 inspectors, are high (> 20%) with lower values for classes with minor presence of defects (I and II) and higher for classes where defects were more relevant (III and NC).

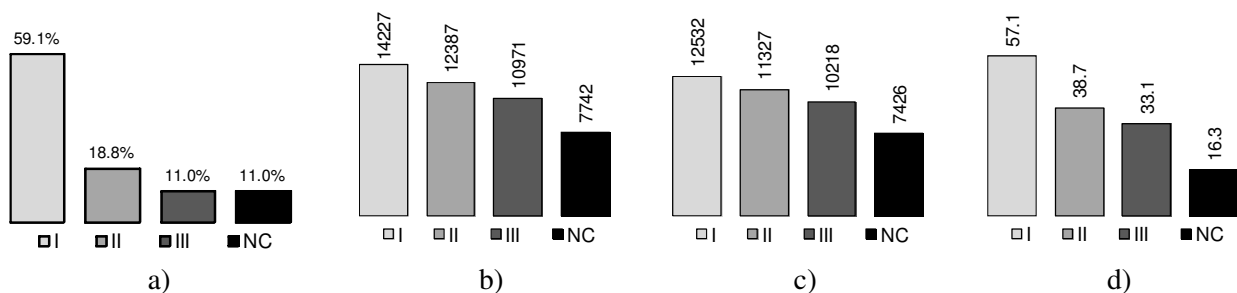


Figure 2: Level 1 inspector: a) visual grading classes percentage for sample of sawn boards; and mean value (N/mm^2) by classes for: b) $E_{m,l}$; c) $E_{m,g}$; d) f_m .

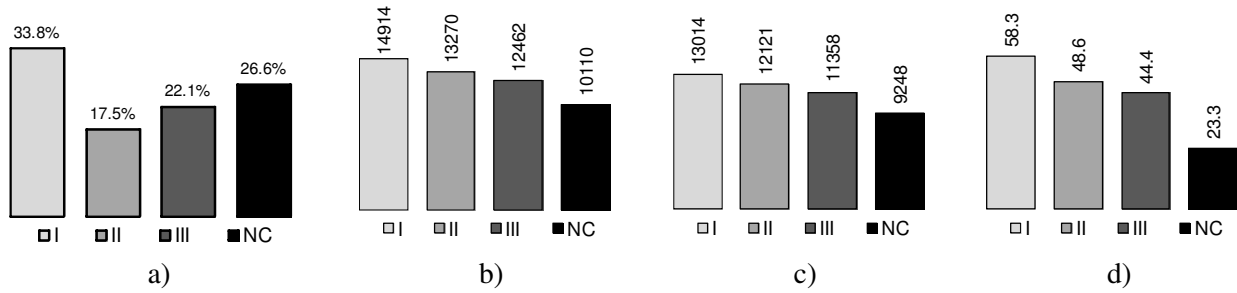


Figure 3: Level 2 inspector: a) visual grading classes percentage for sample of sawn boards; and mean value (N/mm^2) by classes for: b) $E_{m,l}$; c) $E_{m,g}$; d) f_m .

Attribution of Visual Grading Classes. As visual inspection was made for the whole length of each segment, the results of $E_{m,g}$ were considered for the reference interval limit. By consideration of nominal stiffness values in [7] as lower bounds for each class, an average of 59% segments were correctly identified by L1 inspectors, whereas L2 reached 54%. If the same analysis is conducted by class it is shown that L1 inspectors reached an accuracy of 68%, 18%, 22% and 47%, respectively for classes I, II, III and NC, while L2 reached 64%, 27%, 19% and 32% for the same classes.

Taking into consideration that the overall mean value of the mechanical tests evidenced significantly higher results than the nominal values given by [7], the interval bounds for each class were adjusted. For that purpose, the total sample of segments in sawn boards was analyzed in terms of cumulative frequency, and divided into four distinct classes. The lower bounds were calculated at each 25% cumulative frequency with values of 13500 N/mm^2 , 12100 N/mm^2 and 10300 N/mm^2 found respectively for classes I, II and III (see Fig. 4). The reduction factors between mean values of each class are similar to those of the norm, where a 10% and a 20% reduction factors are found for downgrading from class I, respectively, to class II and III [7]. For this case, an average of 42% segments were correctly identified by L1 inspectors, whereas L2 reached 38%. If the same analysis is conducted by class it is shown that L1 inspectors reached an accuracy of 94%, 28%, 17% and 31%, respectively for classes I, II, III and NC, while L2 reached 84%, 23%, 21% and 27% for the same classes. With consideration to the adjusted class limits, a decrease in the overall accuracy is found in both L1 and L2 but an increase at evaluating class I segments is obtained. In both cases, L1 inspectors had approximately a 5% more accurate grading.

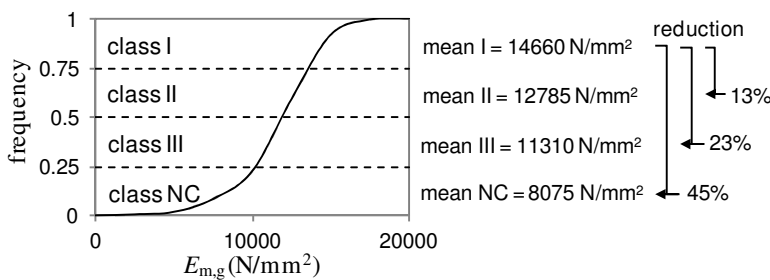


Figure 4: Division of the cumulative distribution function of full sample in visual grading classes, with reduction factors between class I mean value and remaining classes.

Subjectivity and Variation

Distribution Modeling. Considering each sample of values given by L1 and L2 for different visual grading classes, probability distribution parameters were obtained by maximum likelihood estimates (MLE). Probability distributions were selected by χ^2 goodness-of-fit tests. For both $E_{m,l}$ and $E_{m,g}$, 2-parameter Weibull distributions were considered, while a lognormal distribution was selected for f_m . With consideration to those distributions and to the MLE, a large sample was simulated to obtain cumulative distribution functions (cdf) (see Fig. 5). From its analysis is shown that L1 present a wider interval between extreme classes, with better differentiation between intermediate classes than L2. With respect to f_m , an intersection of class II and III cdf's is visible, evidencing the lower sensibility of inspectors for definition of intermediate classes. From this analysis is found that a more experienced inspector will tend to assume similar parameters for defect characterization within classes and better differentiate between visual classes.

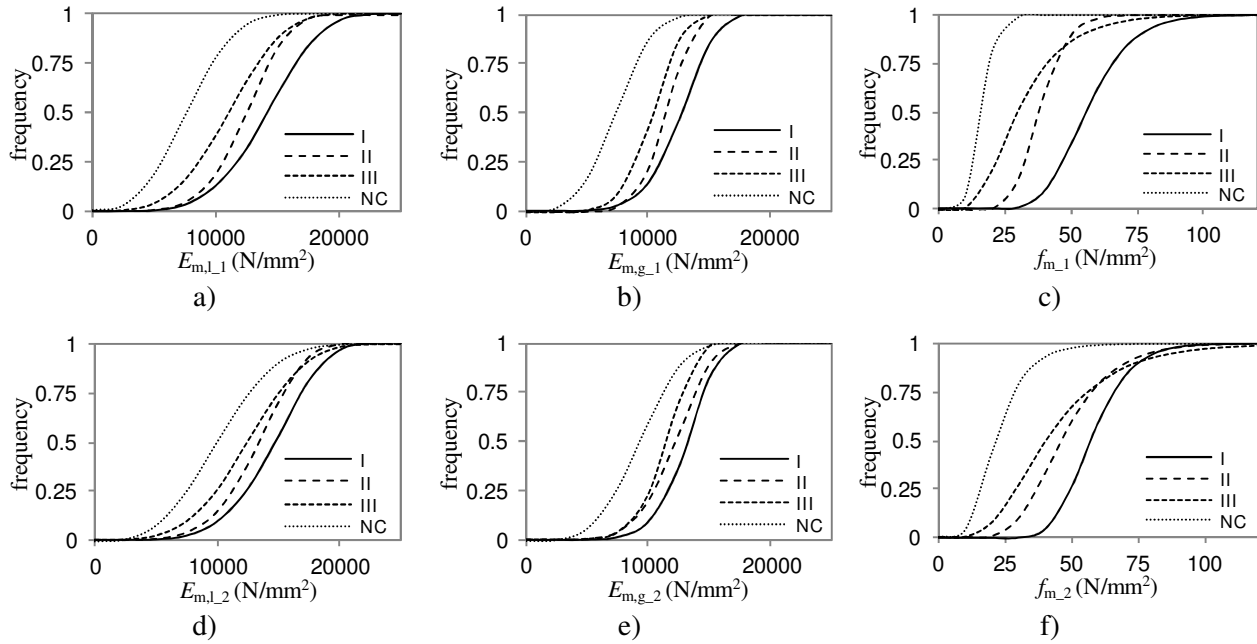


Figure 5: Cumulative distribution function of sawn boards for different visual grading and different level inspectors, level 1 (L1): a) $E_{m,l}$; b) $E_{m,g}$; c) f_m , and level 2 (L2): d) $E_{m,l}$; e) $E_{m,g}$; f) f_m .

Analysis of Variation. Aiming to analyze the difference between results of mechanical tests in different samples given by visual classes, single factor analysis of variation (ANOVA) tests, with confidence level of 95%, were considered. In all sub-samples, a significant variation was found, evidencing that the mean value of stiffness or strength of segments with different visual classes may be considered to have different statistical origins. A significant variance, in this case, is indicative that the visual grading allowed to differentiate between segments with different stiffness or strength. The results also indicate that L1 have a better differentiation between classes for stiffness and strength analysis than L2, therefore consistent with the results of the cdf modeling with MLE.

Analysis of Dispersion. For dispersion analysis of the results, box plots were considered for the results of the mechanical tests differentiated by visual grading of L1 and L2 inspectors (Fig. 6).

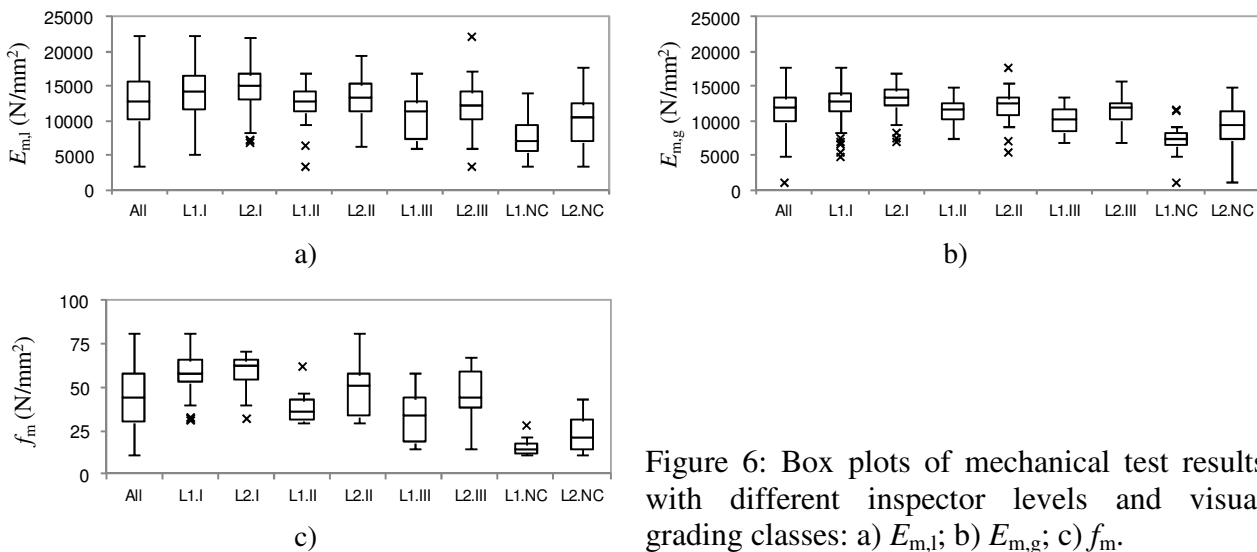


Figure 6: Box plots of mechanical test results with different inspector levels and visual grading classes: a) $E_{m,l}$; b) $E_{m,g}$; c) f_m .

In both stiffness and strength, L1 inspector evidenced a larger dispersion of values for class I than L2 inspectors, however, for the remaining classes a smaller dispersion in the sub-samples is found for L1 inspectors. L2 inspectors also evidence higher concentration of values around a higher mean for each class denoting a more conservative approach.

For this study, the analysis of lower bound outliers is of special interest, as they are indicative of

segments that should have been attributed a lower visual grade. Although existing outliers for both L1 and L2 inspectors, the most relevant feature in this analysis is the interval between minimum and maximum non outliers values, evidencing large dispersion and subjectivity in each sub-sample. **Defect Detection and Measurement.** In order to analyze the subjectivity in detection and measurement of defects between different inspectors, all level inspectors (L1, L2 and L3) performed a visual inspection to all segments of a single sawn board. That board was selected since all segments that composed it performed, in terms of stiffness, as class I in the mechanical tests and therefore, lower differences between inspectors could be expected. The average accuracy of visual inspection was 86%, 68% and 19% respectively for L1, L2 and L3 inspectors. The percentage of different visual grades given to a specific segment is presented in Fig. 7, where is visible that three of the seven segments were given all visual grades. The main limiting parameter that led to the consideration of lower visual classes was the presence of isolated knots, which was also mentioned in the case of [10]. Also in [11] is stated that visual inspection in hardwoods may be significantly harder than in softwoods due to their smaller difference in density between knots and clear wood and also the knot's less defined appearance. In order to analyze the subjectivity in measurement of single knots, the biggest isolated knot in the segment with higher percentage of NC was considered. Mean values for the minimum and maximum diameter of 20.9 mm and 29.9 mm were found with respectively, 15.7% and 21.8% coefficient of variation.

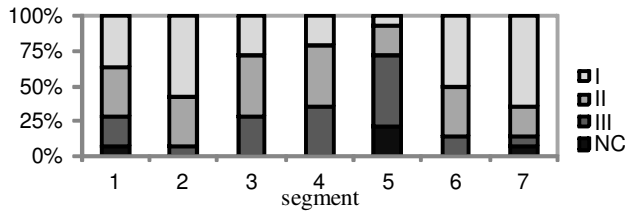


Figure 7: Visual grading classes percentage for a single sawn board, according the combination of all inspectors.

Application to Bayesian Probability Networks

In order to identify the relevance of different expertise of inspectors and the use for the assessment of stiffness properties by visual inspection, different hypothesis of configurations and of prior information were implemented in Bayesian Probability Networks (BPN). By this process, the probability of a given segment being from a specific class is given by information of different levels of visual inspectors.

Stiffness Grading by Single Level Inspectors. In order to assess the individual accuracy of each inspector (L1 and L2) with respect to $E_{m,g}$, a BPN was considered where prior information was given solely by one of the levels of inspectors. The results are presented in Table 1, where is visible that L1 inspectors performed better than L2 inspectors.

Table 1: Conditional probability table for $E_{m,g}$ regarding different visual classes, given L1 or L2.



BPN	$P(E_{m,g} L1 \text{ or } L2)$ [%]	L1 or L2			
		I	II	III	NC
	$E_{m,g}^*$ L1	40.00	37.93	47.06	88.24
	$E_{m,g}^*$ L2	37.80	26.05	32.50	58.44

* Probability of $E_{m,g}$ is given for the same class of that given by the prior information in L1 or L2.

Stiffness Grading by Combined Level Inspectors. The combination of information of L1 and L2 inspectors was considered for the prediction of conditional probability of $E_{m,g}$. The first assumption, considers prior information in a series combination (Table 2). In this case, the second node of prior information controls the analysis if information is given in that node, however in both cases (either being the second node L2 or L1) the accuracy of the visual inspection will be given by the lower

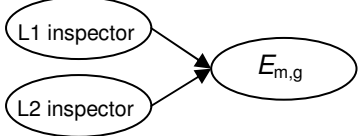
values of accuracy of both inspectors. The second assumption, considered a parallel combination for prior information (Table 3). In this case, an improvement of accuracy in classes I and III were found, supporting the significance in combining information by several inspectors even with similar or inferior individual accuracy. Therefore, due to the large variability in a visual grading, a second opinion may improve the global efficiency, even if provided by a less experienced inspector.

Table 2: Conditional probability table for $E_{m,g}$ regarding different visual classes, given L1 and L2 in a series combination.

BPN		P($E_{m,g} L1,L2$) [%]	L1 and L2			
			I	II	III	NC
	I	$E_{m,g}$	37.80	21.01	7.50	2.60
	II		27.98	26.05	20.00	9.09
III	22.32		26.89	32.50	29.87	
NC	11.90		26.05	40.00	58.44	
						

Shadowed values correspond to the probability of $E_{m,g}$ for the same class given by the prior information in L1 and L2. L1 and L2 are given the same visual grading class as prior information.

Table 3: Conditional probability table for $E_{m,g}$ regarding different visual classes, given L1 and L2 in a parallel combination.

BPN		P($E_{m,g} L1,L2$) [%]	L1 and L2			
			I	II	III	NC
	I	$E_{m,g}$	43.68	4.26	4.55	0.00
	II		27.44	36.17	4.55	0.00
	III		19.13	42.55	54.55	15.91
	NC		9.75	17.02	36.36	84.09

Shadowed values correspond to the probability of $E_{m,g}$ for the same class given by the prior information in L1 and L2. L1 and L2 are given the same visual grading class as prior information.

Conclusions

The efficiency and accuracy of visual inspection is often compromised by its subjective nature inherited by the different levels of expertise of different inspectors. Often the accuracy of grading is based on expert inspectors, which provide a higher confidence degree in their decisions, and are therefore chosen to perform the grading process and/or to confirm the grading of other inspectors. However, in this work, timber beams were visually graded by inspectors with different levels of expertise, and the results compared to 4-point bending tests made in different scales. Effectiveness and accuracy of visual grading was considered by analysis of the percentage of segments attributed to a given visual grade class in comparison to the results obtained in the mechanical tests.

Based in a qualitative analysis, visual inspection proved to assess bending strength efficiently in sawn beams. Results from the most experienced inspectors evidence a larger variation between mean values of visual classes for both stiffness and strength, whereas for less experienced inspectors the variation in intermediate classes is less defined.

With respect to the standard values, an overall 59% accuracy was found. However, since bending stiffness test results were significantly higher than the indicated by the standard, an increase of the lower stiffness bounds values was considered, accounting to the results of mechanical tests to this specific sample. Similar reduction factors as given by the used visual inspection standard were also obtained. Regarding that the origin of timber is important for its mechanical characterization, for this sample an average update of the indicative values, per visual grade, of [7] by increasing 43.4% in the modulus of elasticity in bending is proposed, accounting the assumed grading division and increase of the lower stiffness bounds values. With this grading division, an overall 42% accuracy of the most experienced inspectors was found with approximately less 5% accuracy for lower level

inspectors. However, higher accuracies were found for extreme visual classes (I and NC). From this analysis is also found that a more experienced inspector will tend to assume similar parameters for defect characterization within classes and better differentiate between them. A significant variance between visual grades was found, indicative that visual inspection allowed to differentiate between segments with different stiffness or strength, however, the most experienced inspectors evidenced a larger dispersion of values for the higher class but lower for the remaining classes. Less experienced inspectors also evidenced higher concentration of values around a higher mean for each class denoting a more conservative approach. Regardless of the inspector level, knot size was considered the main limiting visual parameter with higher influence in small scales of the timber elements. When studying the characterization of a single knot, coefficients of variation of 15.7% and 21.8% were found for measuring, respectively, the minimum and maximum diameter.

The use of Bayesian probability networks permitted to individually assess the accuracy in stiffness prediction of different level of inspectors, and also by combination of their information. By use of a parallel combination for prior information, it was concluded that a second opinion may improve the global efficiency, even if provided by a less experienced inspector.

Acknowledgement

The first author acknowledges the financial support of the Portuguese Science Foundation (Fundação de Ciência e Tecnologia, FCT), through project SFRH/BD/62326/2009.

References

- [1] F.J.A. Almazán, E.H. Prieto, F.A. Martitegui, C. Richter, Comparison of the Spanish visual strength grading standard for structural sawn timber (UNE 56544) with the German one (DIN 4074) for Scots pine (*Pinus sylvestris* L.) from Germany. *Holz Roh Werkst* 66(4) (2008) 253-258.
- [2] G.R. Muñoz, A.R. Gete, F.P. Saavedra, Implications in the design of a method for visual grading and mechanical testing of hardwood structural timber for assignation within the European strength classes. *Forest Systems INIA* 20(2) (2011) 235-244.
- [3] H.A. Huber, C.W. McMillin, J.P. McKinney, Lumber defect detection abilities of furniture rough mill employees. *For Prod J* 35(11/12) (1985) 79-82.
- [4] A. Lycken, Comparison between automatic and manual quality grading of sawn softwood. *For Prod J* 56(4) (2006) 13-18.
- [5] O. Silvén, M. Niskanen, H. Kauppinen, Wood inspection with non-supervised clustering. *Mach Vision Appl* 13 (2003) 275-285.
- [6] U. Grönlund, Quality improvements in forest products industry: classification of biological materials with inherent variations. Doctoral thesis. Luleå University of Technology, Sweden (1995).
- [7] UNI 11119, On site inspections for the diagnosis of timber members, ENIU (2004).
- [8] EN 408, Structural Timber and Glued Laminated Timber. Determination of Some Physical and Mechanical Properties, CEN (2003).
- [9] H.S. Sousa, J.M. Branco, P.B. Lourenço, Assessment of strength and stiffness variation within old timber beams. In Proc. SAHC2012, Wroclaw, Poland (2012).
- [10] M. Piazza, M. Riggio, Visual strength grading and NDT of timber in traditional structures. *J Build Appraisal* 3(4) (2008) 267-296.
- [11] M. Nocetti, M. Bacher, M. Brunetti, A. Crivellaro, J.W.G. Kuilen, Machine Grading of Italian Structural Timber Preliminary Results on Different Wood Species. In Proc. 11th WCTE, Trentino, Italy (2010).