



ELSEVIER

Contents lists available at [ScienceDirect](#)

## Food and Bioproducts Processing

journal homepage: [www.elsevier.com/locate/fbp](http://www.elsevier.com/locate/fbp)

IChemE



# Classification modeling based on surface porosity for the grading of natural cork stoppers for quality wines

Vanda Oliveira\*, Sofia Knapic, Helena Pereira

Universidade de Lisboa, Instituto Superior de Agronomia, Centro de Estudos Florestais (CEF), Tapada da Ajuda, P-1349-017 Lisboa, Portugal

## ABSTRACT

The natural cork stoppers are commercially graded into quality classes according with the homogeneity of the external surface. The underlying criteria for this classification are subjective without quantified criteria and standards defined by cork industry or consumers. Image analysis was applied to premium, good and standard quality classes to characterize the surface of the cork stoppers and stepwise discriminant analysis (SDA) was used to build predictive classification models. The final goal is to analyze the contribution of each porosity feature and propose an algorithm for cork stoppers quality class classification. This study provides the knowledge based on a large sampling to an accurate grading of natural cork stoppers.

In average all the models presented accuracy in relation to the commercial classification over 68% with a higher mismatch in the mid-quality range. Color showed an important discriminating power, increasing the accuracy in 10%. The main discriminant features were porosity coefficient and color variables, calculated for the lateral surface. A quality classification algorithm was presented based on a simplified model with an accuracy of 75%. The classification based on color vision systems can ensure improved quality class uniformity and a higher transparency in trade.

© 2013 The Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

**Keywords:** Natural cork stoppers; Quality classes; Image analysis; Porosity; Discriminant analysis; Classification algorithm

## 1. Introduction

Natural cork is an outstanding material for the closure of wine bottles combining physical performance and durability, and allowing a balanced development of wine during bottle aging through its oxygen transfer characteristics (Lopes et al., 2005). Cork is the closure material preferred by wine consumers, as shown by recent surveys (Barber et al., 2009). Natural cork stoppers have also a very favorable footprint and are associated to cork oak forests, a sustainable ecosystem with high biodiversity richness.

Cork is a cellular material with chemical inertia and a set of specific physical and mechanical properties that provide an unmatched closure for bottles and for high performance insulation, with natural cork stoppers as the premium product of the cork industry (Fortes et al., 2004; Pereira, 2007).

Natural cork stoppers are graded into quality classes in function of the apparent homogeneity of their external surface, as seen by human or machine vision (Fortes et al., 2004; Pereira, 2007). The heterogeneity of the cork surface is given primarily by the presence of lenticular channels, as well as by woody inclusions, small fractures or other defects, that can be visually outsingled from the cork surface and are named as the porosity of cork (Gonzalez-Adrados and Pereira, 1996; Pereira et al., 1996). This evaluation is made using automated image-based inspection systems with high throughput rates based on line-scan cameras and a computer embedded in an industrial sorting machine capable of acquiring and processing in real-time the surface image of the stoppers (Lima and Costa, 2006). The systems allow an identification of surface defects and quantification of porosity features, e.g. total area, number or concentration of pores (Chang et al., 1997; Jordanov

\* Corresponding author. Tel.: +351 21 365 3491; fax: +351 21 365 3338.

E-mail address: [vandaoliveira@isa.ulisboa.pt](mailto:vandaoliveira@isa.ulisboa.pt) (V. Oliveira).

Received 18 January 2013; Received in revised form 5 November 2013; Accepted 15 November 2013

Available online 24 November 2013

0960-3085/\$ – see front matter © 2013 The Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

<http://dx.doi.org/10.1016/j.fbp.2013.11.004>

and Georgieva, 2009; Pereira et al., 1994; Radeva et al., 2002).

However, it is known that the underlying criteria for the quality classification of cork stoppers are subjective to some extent and no standards were defined by the cork industry or the consumers to grade sorting (Pereira, 2007). The classification is usually based on reference samples showing the range of quality variation that can be found in the consignment for a given client (Lopes and Pereira, 2000).

Several studies have been published on identifying the contribution of each porosity feature for the grading of cork stoppers (Costa and Pereira, 2005, 2006, 2007, 2009), as well as of cork discs (Lopes and Pereira, 2000) and cork planks (Benkirane et al., 2001; Gonzalez-Adrados et al., 2000; Gonzalez-Adrados and Pereira, 1996; Pereira et al., 1996). All the studies have shown that there is overlap between classes and that the role of non-quantified features, i.e. related to operator or industry is significant. However, the sampling used in these studies was limited, e.g. in number of cork stoppers and in the surface area covered by image analysis, and doubts may remain regarding the selected features since the natural variability in cork is high. A recent study analyzed in detail the porosity of a large sample of natural cork stoppers, showing that variation of some of the pores characteristics throughout the lateral surface of the stopper is important (Oliveira et al., 2012).

Other studies have been published focusing on modeling the classification of cork stoppers and discs. Chang et al. (1997) proposed a cork stopper quality classification system based on features extraction and a fuzzy neural network, with 6.7% of rejection after reevaluation of the results by human experts. Vega-Rodríguez et al. (2001) presented a system for image processing using reconfigurable hardware and an algorithm for the cork stoppers classification that uses a simplified set of porosity features (defect area, size of the biggest defect and area occupied by defects of different sizes) for the two tops. Vitrià et al. (2007) presented a cork stopper classification model based on feature extraction and class-conditional independent component analysis, reaching an average error rate of 2%. Paniagua et al. (2011) developed for cork discs a neurosystem to model the human cork quality classification. Other techniques have been applied to cork classification such as X-ray Compton tomography (Brunetti et al., 2002) or terahertz/millimeter wave spectroscopy (Hor et al., 2008) in order to refine the visual classification with inclusion of possible voids, cracks, and defects inside the stopper. Recently, Gómez-Sánchez et al. (2013) used colorimetric image analysis techniques with application of segmentation algorithms to measure the porosity of cork samples. They obtained the best NIRS calibrations by measuring the porosity into three classes of color, matching the results obtained by image analysis.

This paper analyses in detail the porosity features of cork stoppers in the grades used today by the quality wine markets: three major quality grades, of premium, good and standard stoppers. The study was made on the total external surface (lateral surface and tops) of the stoppers, and was based on a large sample in order to encompass the cork natural variability and to allow better confidence in the results. The contribution of each porosity feature to quality classification is analyzed using adequate statistical validation, the significant indicators for the grading selected and a classification algorithm is proposed for cork stoppers quality classification using today's quality grades. The final goal is to define a simple and objective classification that could be used by the industry in order to ensure improved quality class uniformity and a higher

transparency in trade, and additionally generate knowledge on the key features of the raw material that may be used in the research, development and enhancement of new products.

## 2. Materials and methods

### 2.1. Sampling

The natural cork stoppers (24 mm diameter × 45 mm length) used in this study were collected from one major Portuguese cork industrial unit. The stoppers were randomly sampled (before washing and surface treatment) after a first step grading by the automated vision system used routinely in the industrial production line. The criteria considered in such automated grading include total area and number of pores, area of the largest pores, pore concentration level, location of defects, vertical and horizontal projection of pores, and the presence of cracks. Subsequently they were inspected by skilled operators and graded into three references quality classes as required today by the wine market, coded as premium (including the traditional “flor” and extra commercial classes), good (superior and 1st commercial classes) and standard (2nd and 3rd commercial classes). After this manually validated classification, a sample of 200 cork stoppers of each quality class was randomly taken and used as the reference for the classification modeling.

### 2.2. Image acquisition

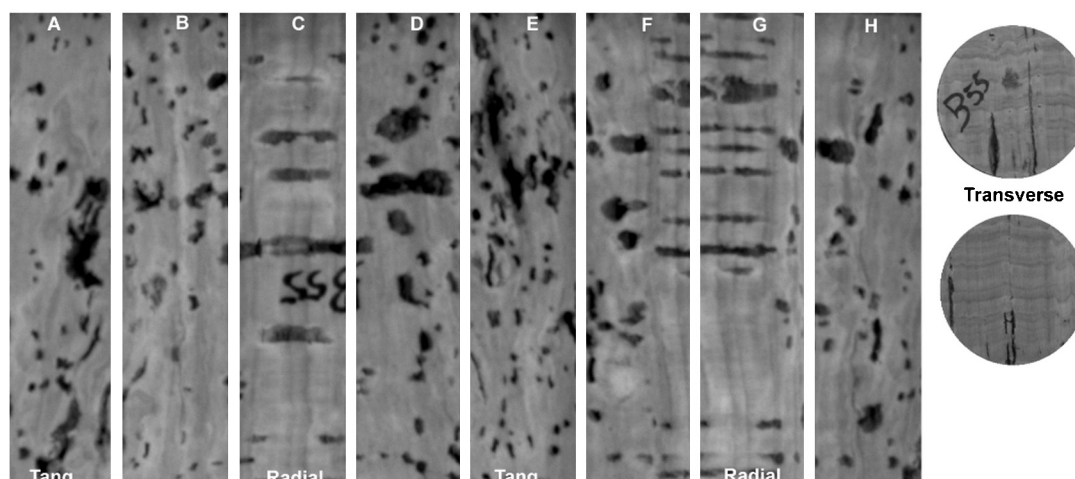
The natural cork stoppers were individually analyzed and their image surface (cylindrical lateral surface and circular bases) acquired with an image analysis system that included a digital 7 mega pixels in macro stand solution set on an acquisition Kaiser RS1 Board with a controlled illumination apparatus, connected to a computer using AnalySIS® image processing software (Analysis Soft Imaging System GmbH Münster, Germany, version 3.1).

The image acquisition covered 100% of the lateral area by using eight successive frames of the cylindrical lateral surface of the body. The first frame was acquired parallel to cork growth rings and the others subsequently taken by rotating the stopper 45° (Fig. 1). Two circular frames were acquired for the two tops corresponding to 96% of the total area. Due to the way stoppers are punched out from the cork strip, the tops correspond to transversal sections of cork while the lateral surface of the stopper includes tangential and radial sections of cork and all the in-between sections (Pereira, 2007; Pereira et al., 1987).

The object extraction was carried out inside two predefined regions of interest, one rectangular, 45.05 mm long and 9.41 mm wide (area 423.92 mm<sup>2</sup>), for the lateral surface, and another circular for the tops with 23.51 mm diameter (area 433.92 mm<sup>2</sup>). The image threshold was adjusted individually for each image and ranged in a RGB system from 65 to 135 for red, from 60 to 115 for green, and from 65 to 120 for blue.

### 2.3. Image data analysis

A set of variables was collected automatically for each pore: area (mm<sup>2</sup>), calculated by the number of pixels of the particle times the calibration factors; mean diameter (mm), defined as the arithmetic mean of all diameters of a particle (range angles between 0° and 179°, with step width of 1°); maximum diameter (mm), is the maximum diameter of all maximum



**Fig. 1 – Photographs of the cylindrical lateral surface (composition of eight successive pictures) and the two circular tops (transverse section). Frames A and E represent tangential section while frames C and G represents cork radial section.**

diameters determined at each angle (varies in  $1^\circ$  steps); mean rectangle, defined as the area of the mean rectangle which sides consist of tangents to the particle borders; maximum rectangle, defined as the area of the biggest rectangle which sides consist of tangents to the particle borders; shape factor, defined as  $(\text{area}/\text{perimeter}^2) \times 4\pi$ , measuring the roundness of the pores; convexity, defined as the fraction of the pore area and the area of its convex hull (ranging between 0 and 1); aspect ratio, as the maximum ratio between width and length of a bounding rectangle of the pore; sphericity, that describes the elongation of the pore by using central moments (value of 1 for a perfect circular particle); and mean value of all red, green and blue intensities in each pore.

These data were filtered and only pores with area equal or superior to  $0.1 \text{ mm}^2$  were considered for the analysis. The pixel size was  $0.05 \text{ mm}$  and therefore the lowest pore area that could be confidently resolved was  $0.02 \text{ mm}^2$  corresponding to eight contiguous pixels. Previous studies considered only porosity superior to  $0.5 \text{ mm}^2$  on the assumption that small porosity is functionally and esthetically irrelevant, and only brings higher variance and variability to the sample (Costa and Pereira, 2005, 2006, 2007, 2009; Gonzalez-Adrados et al., 2000; Lopes and Pereira, 2000; Pereira et al., 1996).

The variables collected for each pore were processed and originate several calculated variables for each frame: porosity coefficient (%), defined as the proportion of the area occupied by pores; total number of pores and number of pores by dimension classes; total area of pores ( $\text{mm}^2$ ), sum of the area of all pores in the frame; average pore area ( $\text{mm}^2$ ), calculated as the arithmetic mean of the area of all pores in the frame; maximum pore area ( $\text{mm}^2$ ), defined as the area of the biggest pore in the frame; mean diameter, calculated as the mean diameter average of all pores; maximum diameter, defined as the biggest of all pores maximum diameter; mean rectangle, calculated as the arithmetic mean of the mean rectangle of all pores in the frame; maximum rectangle, defined as biggest of all pores maximum rectangle; the shape-variables (shape factor, sphericity, aspect ratio and convexity) and the color-variables (red, green and blue) from pores were averaged into frame variables. Moreover, these variables were averaged and transformed into cork stopper transverse (two tops), tangential (frames A and E) and radial (frames C and G) section and cork stopper body variables (considering the eight frames of the stoppers lateral surface).

## 2.4. Statistical analysis

Several graphical and descriptive statistical analyses were carried out for the characterization of the quality classes and of the stoppers lateral surface.

In order to differentiate between quality classes and to predict the class of a future observation, several predictive classification models of stoppers were built based on their surface characteristics using stepwise discriminant analysis (SDA). Discriminant analysis is the appropriate statistical technique when the dependent variable is categorical and the independent variables are quantitative. Discriminant analysis assumes that data come from a multivariate normal distribution and that the covariance matrices of the groups are equal (Sharma, 1996). Another characteristic of the data that can affect the results is multicollinearity among the independent variables. The stepwise method can be useful when in the presence of many predictors by automatically selecting the “best” variables to use in the model. The stepwise selection begins with no variables in the discriminant function, and at each step a variable is either added or removed, if it adds the most discriminating power or if it does not significantly lower the discriminating power, as measured by the statistical criterion. The procedure stops when at a given step no variable is added or removed from the discriminant functions (Sharma, 1996). The statistical criterion used to measure variable discriminating power was Wilks’  $\Lambda$  value ( $p < 0.05$ ). All the statistical analysis was performed using SPSS<sup>®</sup> statistical software (version 19.0; SPSS Inc., Chicago IL).

The sample was randomly divided into two groups: 70% of the cork stoppers were used to estimate the discriminant functions and produced the classification models (modeling set), and 30% of the stoppers were used for external validation of the models (validation set).

These models were analyzed and compared, and a simplified model and a classification algorithm for cork stoppers quality class classification were proposed.

## 3. Results

### 3.1. Characterization of quality classes

The three quality classes of natural cork stoppers are characterized in Table 1 for the lateral surface.

**Table 1 – Mean ( $n = 200$ ) and standard deviation (in brackets) of the independent variables of the lateral surface for the three reference quality classes of natural cork stoppers.**

	Premium	Good	Standard
Porosity coefficient (>0.1 mm <sup>2</sup> ) (%)	2.4 (0.8)	4.0 (1.4)	5.6 (2.0)
Porosity coefficient (>0.5 mm <sup>2</sup> ) (%)	2.0 (0.8)	3.5 (1.4)	5.0 (2.0)
Number of pores			
Total	135 (48)	167 (58)	193 (54)
0.1–0.5 mm <sup>2</sup>	86	99	112
0.5–1.0 mm <sup>2</sup>	26	31	35
1.0–2.0 mm <sup>2</sup>	15	20	23
2.0–5.0 mm <sup>2</sup>	7	13	16
>5.0 mm <sup>2</sup>	1	4	7
Total area of pores (mm <sup>2</sup> )	82.7 (28.8)	135.1 (47.6)	191.2 (69.1)
Average pore area (mm <sup>2</sup> )	0.6 (0.2)	0.9 (0.3)	1.1 (0.4)
Maximum pore area (mm <sup>2</sup> )	6.5 (3.2)	11.3 (6.0)	18.0 (9.9)
Average pore shape factor	0.6 (0.1)	0.5 (0.0)	0.5 (0.1)
Average pore aspect ratio	2.0 (0.3)	2.1 (0.2)	2.1 (0.2)
Average pore <u>Red</u>	93 (29)	96 (21)	91 (30)
Average pore <u>Green</u>	71 (22)	77 (16)	69 (24)
Average pore <u>Blue</u>	76 (20)	80 (14)	73 (22)

The average porosity coefficient including all pores above 0.1 mm<sup>2</sup> was 2.4%, 4.0% and 5.6%, respectively, for premium, good and standard class. When considering only pores above 0.5 mm<sup>2</sup>, the values decreased to 2.0%, 3.5% and 5.0%, respectively.

The small porosity (between 0.1 and 0.5 mm<sup>2</sup>) represented on average an increase of 0.5% in the porosity coefficient and represented 63.9%, 59.4% and 58.1% of the total number of pores for the premium, good and standard quality classes, respectively.

The standard deviation of the porosity coefficient of the lateral surface was 0.8%, 1.4% and 2.0% for the premium, good and standard quality class, respectively, regardless of the minimum pore area considered for the porosity coefficient calculations (0.1 or 0.5 mm<sup>2</sup>).

The maximum pore area in the lateral surface of cork stoppers differed significantly between classes with 6.5 mm<sup>2</sup>, 11.3 mm<sup>2</sup> and 18.0 mm<sup>2</sup> for premium, good and standard classes, and the average pore area also increased from 0.6 mm<sup>2</sup> to 0.9 mm<sup>2</sup> and 1.1 mm<sup>2</sup>, respectively.

The shape variables presented identical values between quality classes. The RGB of the pores was, on average, 93 red, 71 green and 76 blue for premium class, 96 red, 77 green and 80 blue for good class and 91 red, 69 green and 73 blue for the standard class.

Fig. 2 presents the variation of the porosity coefficient and the average pore area in the transverse, tangential and radial sections for the three quality classes of natural cork stoppers. These sections correspond respectively to the tops (transverse), frames A and E (tangential) and frames C and G (radial).

The porosity coefficient ranged between 3.1% and 6.2% in the tangential section, respectively, for premium and standard quality class, and 1.9% and 5.3% in the radial section (Fig. 2a). The premium class had average pore areas of 0.6 mm<sup>2</sup>, 0.8 mm<sup>2</sup> and 1.3 mm<sup>2</sup> for the, tangential, radial and transverse section, respectively. The corresponding values for the standard quality class were 0.9, 1.4 and 1.7 mm<sup>2</sup> (Fig. 2b).

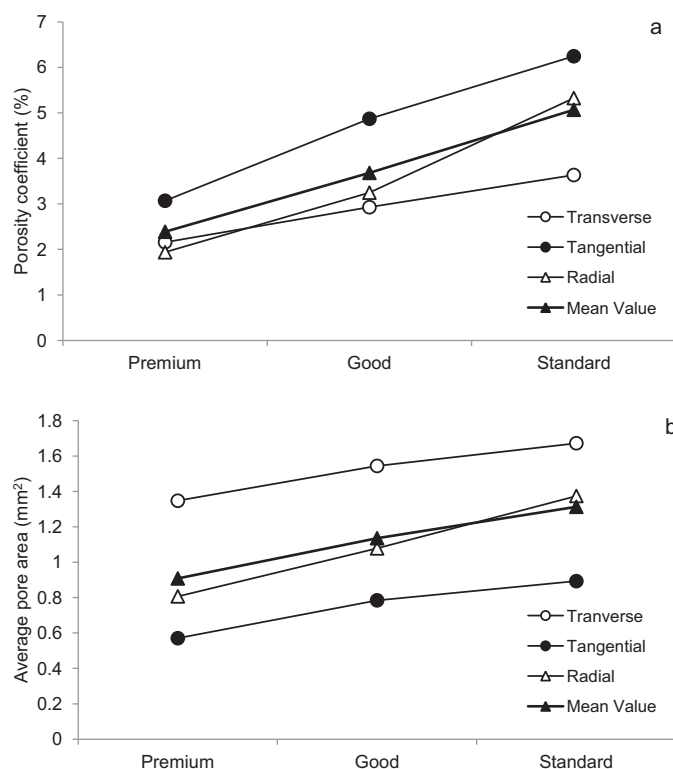
### 3.2. Quality classification models

Five predictive classification models of stoppers into quality classes were analyzed and their differences characterized. Table 2 shows the variables selected by SDA to build each

model and Table 3 presents the results of the classification of cork stoppers into the three quality classes considered in this study.

The multivariate normality assumption was investigated by the application of Kolmogorov–Smirnov test with Lilliefors significance correction to the independent variables and there was no reason to believe that was violated. Box's M test was used to verify if the variance–co-variance matrices are equivalent. Although M was significant ( $p < 0.00$ ), with large samples, a significant result is not regarded as too important.

Models 1 and 2 were built with the application of SDA to the variables used traditionally in the automated image-based inspection systems, i.e. dimension, concentration and shape



**Fig. 2 – Variation in the transverse, tangential and radial sections and mean value for the three quality classes of natural cork stoppers: (a) porosity coefficient and (b) average pore area.**

**Table 2 – Type and variables selected by SDA for each predictive classification model.**

Variables	Models							
	Model 1	Model 2		Model 3	Model 4		Model 5	
	Body	Body	Tops	Body	Body	Tops	Tang	Rad
<i>Concentration and dimension</i>								
Porosity coefficient	x	x	x	x	x		x	x
Total number of pores	x	x		x		x		x
Mean pore area						x		x
Maximum pore area	x	x		x	x			
Mean diameter	x	x		x	x		x	
Maximum diameter	x	x						
Mean rectangle							x	
Maximum rectangle	x	x						x
<i>Shape</i>								
Shape factor	x	x						
Sphericity	x	x			x		x	
Aspect ratio	x	x			x			
Convexity				x	x		x	
<i>Color</i>								
Average pore <u>Red</u>				x	x	x	x	
Average pore <u>Green</u>				x	x	x	x	
Average pore <u>Blue</u>				x	x	x	x	

variables averaged for the body lateral surface and tops. In both models the body variables selected by SDA were porosity coefficient, shape-variables (shape factor, sphericity and aspect ratio) and dimension-variables (maximum pore area, mean and maximum diameter, total number of pores and maximum rectangle). In both models the average accuracy for modeling and validation set was 69.5% and 68.3%, respectively. The good quality class showed the lowest accuracy of 60.0% and 62.9% for, respectively, models 1 and 2, while premium and standard quality classes had accuracy values above 70%.

Models 3 and 4 are comparable to models 1 and 2 with the addition of pores color-variables. The independent variables selected by SDA for model 4 were porosity coefficient, maximum pore area, mean diameter, total number of pores, convexity and average pore red, green and blue. These color variables were selected by SDA leading to an accuracy of 80.0% for model 3 and 84.1% for model 4. The good class accuracy increased to 70.7% and 78.6% for models 3 and 4, respectively. The validation set showed an average accuracy over 73% for both models with the highest value of 85.0% for the standard

**Table 3 – Classification of cork stoppers into three quality classes using discriminant analysis in percentage of the initial number of stoppers in each quality class (modeling set). In brackets is the classification of the validation set. Bold corresponds to the match of classifications.**

Original quality class	Predicted quality class			Mean accuracy
	Premium	Good	Standard	
<b>Model 1</b>				
Premium	75.7 (70.0)	20.7 (30.0)	3.6 (0.0)	69.5 (68.3)
Good	21.4 (20.0)	<b>60.0 (61.7)</b>	18.6 (18.3)	
Standard	5.7 (8.3)	21.4 (18.3)	<b>72.9 (73.3)</b>	
<b>Model 2</b>				
Premium	74.3 (70.0)	22.9 (30.0)	2.9 (0.0)	69.5 (68.3)
Good	20.0 (20.0)	<b>62.9 (60.0)</b>	17.1 (20.0)	
Standard	5.0 (6.7)	23.6 (18.3)	<b>71.4 (75.0)</b>	
<b>Model 3</b>				
Premium	<b>84.3 (73.3)</b>	13.6 (25.0)	2.1 (1.7)	80.0 (72.8)
Good	23.6 (23.3)	<b>70.7 (61.7)</b>	5.7 (15.0)	
Standard	10.7 (10.0)	4.3 (6.7)	<b>85.0 (83.3)</b>	
<b>Model 4</b>				
Premium	<b>85.0 (66.7)</b>	12.9 (30.0)	2.1 (3.3)	84.1 (73.3)
Good	17.1 (21.7)	<b>78.6 (68.3)</b>	4.3 (10.0)	
Standard	7.1 (8.3)	4.3 (6.7)	<b>88.6 (85.0)</b>	
<b>Model 5</b>				
Premium	<b>80.0 (61.7)</b>	15.7 (31.7)	4.3 (6.7)	78.8 (70.0)
Good	20.0 (20.0)	<b>73.6 (65.0)</b>	6.4 (15.0)	
Standard	13.6 (10.0)	3.6 (6.7)	<b>82.9 (83.3)</b>	
<b>S. model</b>				
Premium	<b>84.3 (70.0)</b>	13.6 (28.3)	2.1 (1.7)	76.2 (70.6)
Good	20.7 (25.0)	<b>68.6 (63.3)</b>	10.7 (11.7)	
Standard	14.3 (16.7)	10.0 (5.0)	<b>75.7 (78.3)</b>	

**Table 4 – Structure matrix.**

	Function 1	Function 2
Porosity coefficient (>0.1 mm <sup>2</sup> ) (%)	0.762 <sup>a</sup>	0.562
Average pore Red	−0.080	0.225 <sup>a</sup>
Average pore Green	−0.097	0.150 <sup>a</sup>
Average pore Blue	−0.041	0.078 <sup>a</sup>

<sup>a</sup> Largest absolute correlation between each variable and any discriminant function.

class in model 4. In model 3 the good quality class had the lowest accuracy of 70.7% and 61.7% for the validation set and the standard quality class had the highest accuracy for validation set with 83.3%.

Model 5 considered dimension, concentration, shape and color variables calculated for radial and tangential sections of cork. The variable with the highest discriminating power was the porosity coefficient from both sections. The other variables of the tangential section selected for the model by SDA were: color (RGB), shape (sphericity and convexity) and mean dimension variables (diameter and rectangle). For the radial section the variables selected were: total number of pores, average pore area and maximum rectangle. The average accuracy for modeling and validation set was 78.8% and 70.0%, respectively.

Two discriminating functions were fitted automatically for all the models, but nearly all the variance explained in terms of differences between classes by the models was due to the first discriminant function. In all the models both functions were considered relevant for the discriminating power.

### 3.3. Quality classification algorithm

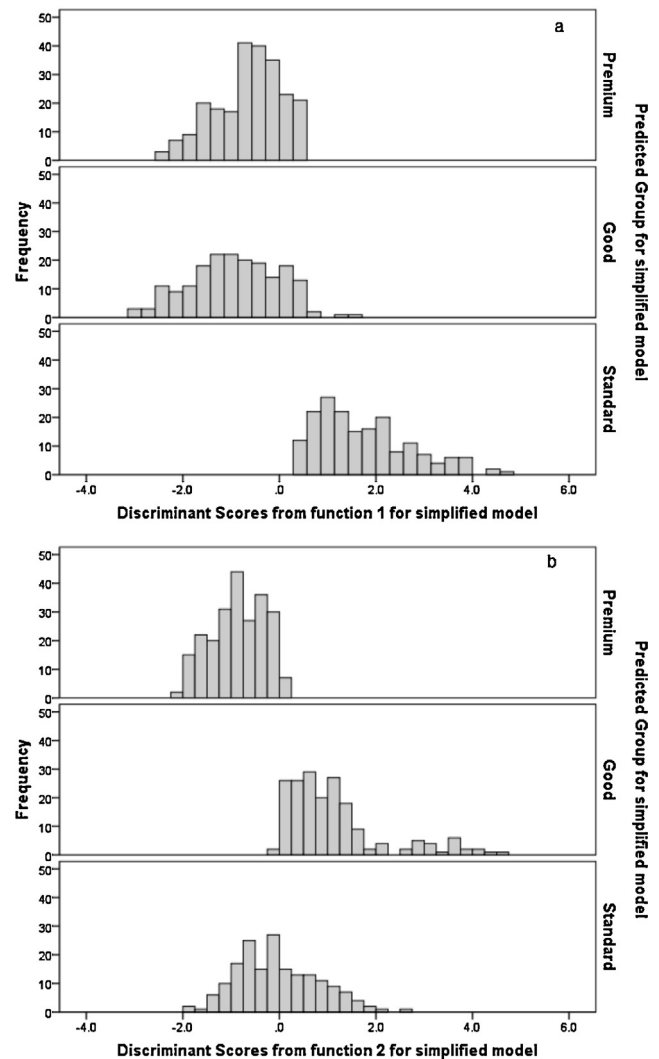
A simplified model was made using the main discriminant features, i.e. porosity coefficient and the RGB color-type variables calculated for the body lateral surface. The analysis of the structure matrix table (Table 4) reveals that porosity coefficient is highly related with function 1 while function 2 explains essentially the color variability.

Fig. 3 illustrates the distribution of the discriminant functions scores for each predicted group. Reading the scores ranges on the axes and the overlaps of the graphs, a substantial discrimination is revealed. Function 1 discriminates the standard quality class from premium and good stoppers (Fig. 3a), whereas function 2, that represents color variability, has discriminating power between premium and good stoppers (Fig. 3b).

Fisher's classification function coefficients were used to classify the cases between the quality classes. The coefficients of the independent variables shown in Table 5 were used to construct a discriminant function for each quality class. To

**Table 5 – Classification function coefficients (Fisher's linear discriminant functions).**

	Quality class		
	Premium	Good	Standard
Porosity coefficient (>0.1 mm <sup>2</sup> ) (%)	1.283	1.727	2.695
Average pore Red	−1.689	−1.981	−1.332
Average pore Green	−2.198	−1.854	−2.090
Average pore Blue	4.948	4.995	4.312
(Constant)	−34.033	−37.037	−32.853



**Fig. 3 – Histograms showing the distribution of discriminant scores for each predicted cork stoppers quality class by the simplified model: (a) function 1 and (b) function 2.**

perform classification the three functions are computed and an individual is assigned to the group with the highest score.

The classification results for the modeling set reveal that 76.2% of the cork stoppers were classified correctly into quality classes. Like the previous models, the good quality class had the lowest accuracy of 68.6%. It is important to highlight the accuracy of 70.6% obtained with the validation set (Table 3).

When comparing to a simplistic model with the porosity coefficient as the only variable, the inclusion of color-type variables increased the overall accuracy in more than 10%. The misclassification corresponded, on average, to 13% of the totality of the cork stoppers, with higher values for the good quality class. It is noteworthy that 14.3% of the cork stoppers originally classified as standard quality were predicted by the model as premium class stoppers.

## 4. Discussion

### 4.1. Characterization of stoppers quality classes

As expected, the surface of the cork stoppers is not homogeneous when comparing the transverse, tangential and radial sections. The lenticular channels appear with a different

aspect in the three sections: in the radial and transverse sections they look like elongated rectangular channels and in the tangential section they have a circular to elliptical form (Figs. 1 and 2).

Several porosity features such as porosity coefficient or total number of pores presented an increasing trend from premium to standard quality class, for all sections in accordance with the published data (Costa and Pereira, 2007).

The tangential section had the highest porosity coefficients (Fig. 2a) due to the high number of pores even if the average pore area was smaller (Fig. 2b). Pereira et al. (1996) had already stated that the average pore area was higher in the transverse/radial section than in the tangential section. Lopes and Pereira (2000) reported an average pore area of 0.25 mm<sup>2</sup> for cork discs corresponding to the tangential section of cork. Pores in the transverse section had a higher aspect ratio than in the radial and tangential sections, and therefore were more elongated or thinner.

The shape variables presented identical values between quality classes, as previously reported by Gonzalez-Adrados and Pereira (1996), and Costa and Pereira (2009). On the other hand, the color-type variables showed differences between classes although without a trend.

#### 4.2. Quality classification models

Several predictive classification models of stoppers into quality classes were built based on the presented surface characteristics using stepwise discriminant analysis (SDA) for selection of a specific set of variables (Table 2).

All the five classification models presented, on average, accuracy superior to 69.5% for the modeling set or superior to 68.3% for the validation set. The results achieved with the validation set are much better than those presented by Costa and Pereira (2006) for seven quality classes with the application of the established decisions rules for classification stopper by stopper, on an individual basis.

It is empirically known that the quality classification of stoppers is subjective to some extent and therefore varying between individual experts as demonstrated by Barros and Pereira (1987) who reported classification match values of 31% for two operators and higher classification difficulty in the mid-quality classes. This lack of objective criteria for visual classification contributes to the higher mismatch found for the mid-quality range; indeed it was in the good class of stoppers that the lowest accuracy was obtained (from 60.0% in model 1 to 78.6% for model 4).

In the analysis of models 1 and 2 it is important to notice that the unique variable considered by SDA with significant power from tops was the porosity coefficient. Gonzalez-Adrados et al. (2000) reported that the first variable to be entered into classification models of cork planks was the porosity coefficient in the transverse section. However, this variable does not have the same discriminant power in cork stoppers classification probably because the transverse area (of the tops) is smaller and visually less relevant when compared with that in cork planks.

The accuracy achieved with these models is comparable with the 65% and 72% of overall accuracy achieved by Costa and Pereira (2006) when tested a simplified classification using three quality classes and considering the significant variables selected by SDA. The evaluation of models 1 and 2 suggests that the observation of cork stoppers tops is irrelevant for the construction and accuracy of the classification model,

and therefore for the classification of cork stoppers in quality classes.

In models 3 and 4 the color variables were selected for the discriminating functions and lead to a higher accuracy in the classification when compared with models 1 and 2. The accuracy obtained for the validation set (72.8% and 73.3% for, respectively, models 3 and 4), despite being smaller than those obtained when building the model (modeling set), improve those published to date by Costa and Pereira (2006). The color variation of pores represents the color heterogeneity of the filling tissues of pores and allows a better separation of the quality classes. In general “darker” pores will lead to the allocation of the corresponding stoppers to worse quality classes.

In model 5 the different selection of discriminating variables that occurred for both sections reflects the anisotropy of porosity features shown by cork stoppers.

Overall the porosity coefficient is the most important variable with discriminant power for separation of cork stoppers quality classes and was selected in all the models.

Comparing the three “new” models (3, 4 and 5) that included the color type-variables with the “traditional” models represented by models 1 and 2, it seems that color has an important discriminating power conferring a better accuracy. The classification based on colored vision systems can therefore ensure improved quality class uniformity and a higher transparency in trade.

#### 4.3. Quality classification algorithm

A simplified model using the main discriminant features, i.e. porosity coefficient and the RGB color-type variables calculated for the body lateral surface were therefore proposed here. This model minimizes the processing of data by reducing the dimensionality of the problem and suppressing information redundancy. Moreover, the classification results were slightly better than the ones found for “traditional” models. Most published studies, like Paniagua et al. (2011), assume that the classification made by human experts is optimal/perfect and the aim is to obtain the most similar classification results. However the stoppers that were considered as misclassified should be analyzed, because of the human inspection subjectivity already referred. A blind belief in the accuracy of the present in-use classification is therefore an exaggeration. The 30% of “misclassification” should therefore be considered only as indicative, and the model classification probably translates more exactly the features and quantified appearance of the external surface of the stoppers.

The most common criteria considered in automated grading include variables like total area and number of pores, area of the largest pores and concentration level of pores, i.e. dimension and concentration variables. The inclusion of color-variables improves the classification and probably approximates more the model to the presently used classification related to visual appearance of cork stopper as perceived by the operator, thereby approximating the results from manual and automated grading. The proposed quality classification algorithm can be applied in the industry to ensure improved uniformity within each stoppers' quality class. By using measurable and quantified features of the external surface of the natural stoppers, this classification algorithm will allow a higher transparency in trade. Moreover, the highlight on key features of cork can be used in research, development and enhancement of new products that have cork as raw material.

## Acknowledgments

This work was supported by FEDER funds through the Operational Programme for Competitiveness Factors – COMPETE and by National Funds under the project FCOMP-01-0124-FEDER-005421. Centro de Estudos Florestais (Forest Research Center) is a research unit supported by the National Research funding of Fundação para a Ciência e Tecnologia (PEst-OE/AGR/UI0239/2011). Funding from FCT is acknowledged by Vanda Oliveira as a doctoral student (SFRH/BD/77550/2011), and Sofia Knapic as a post-doctoral researcher (SFRH/BPD/76101/2011). The authors acknowledge the collaboration of Amorim & Irmãos, S.A. in materials supply and the assistance of our colleague Lídia Silva in the image analysis.

## References

- Barber, N.A., Taylor, C.D., Dodd, T.H., 2009. [Twisting tradition: consumers' behavior toward alternative closures](#). *J. Food Prod. Market.* 15, 80–103.
- Barros, L., Pereira, H., 1987. [Influência do operador na classificação manual da cortiça por classes de qualidade](#). *Cortiça* 582, 103–105.
- Benkirane, H., Benslimane, R., Hachmi, M., Sesbou, A., 2001. [Possibilité de contrôle automatique de la qualité du liège par vision artificielle](#). *Ann. For. Sci.* 58, 455–465.
- Brunetti, A., Cesareo, R., Golosio, B., Luciano, P., Ruggero, A., 2002. [Cork quality estimation by using Compton tomography](#). *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. Atoms* 196, 161–168.
- Chang, J., Han, G., Valverde, J.M., Griswold, N.C., Duque-Carrillo, J.-F., Sanchez-Sinencio, E., 1997. [Cork quality classification system using a unified image processing and fuzzy-neural network methodology](#). *IEEE Trans. Neural Netw.* 8, 964–974.
- Costa, A., Pereira, H., 2009. [Computer vision applied to cork stoppers inspection](#). In: Zapata, S. (Ed.), *Cork Oak Woodlands and Cork Industry: Present, Past and Future*. Museu del Suro de Palafrugell Publications, Barcelona, pp. 394–405.
- Costa, A., Pereira, H., 2007. [Influence of vision systems, black and white, colored and visual digitalization, in natural cork stopper quality estimation](#). *J. Sci. Food Agric.* 87, 2222–2228.
- Costa, A., Pereira, H., 2006. [Decision rules for computer-vision quality classification of wine natural cork stoppers](#). *Am. J. Enol. Viticult.* 57, 210–219.
- Costa, A., Pereira, H., 2005. [Quality characterization of wine cork stoppers using computer vision](#). *J. Int. Sci. Vigne Vin* 39, 209–218.
- Fortes, M.A., Rosa, M.E., Pereira, H., 2004. *A Cortiça*, first ed. IST Press, Lisbon.
- Gómez-Sánchez, I., Ceca, J.L., García-Olmo, J., Lara-Buil, L., López-Luque, R., Prades, C., 2013. [Application of image analysis and NIRS technology to the evaluation of the porosity of planks, sheets and cork stoppers, and its relation with the industrial quality](#). *Maderas Ciencia Tecnol.* 15, 293–309.
- Gonzalez-Adrados, J., Pereira, H., 1996. [Classification of defects in cork planks using image analysis](#). *Wood Sci. Technol.* 30, 207–215.
- Gonzalez-Adrados, J., Lopes, F., Pereira, H., 2000. [Quality grading of cork planks with classification models based on defect characterisation](#). *Holz Roh Werkst.* 58, 39–45.
- Hor, Y.L., Federici, J.F., Wample, R.L., 2008. [Nondestructive evaluation of cork enclosures using terahertz/millimeter wave spectroscopy and imaging](#). *Appl. Opt.* 47, 72–78.
- Jordanov, I., Georgieva, A., 2009. [Neural network classification: a cork industry case](#). In: *IEEE International Symposium on Industrial Electronics*, Seoul. IEEE Computer Society, pp. 232–237.
- Lima, J., Costa, P., 2006. [Real-time cork classification method: a colour image processing approach](#). *Int. J. Fact. Autom. Rob. Soft Comput.* 2, 1–10.
- Lopes, F., Pereira, H., 2000. [Definition of quality classes for champagne cork stoppers in the high quality range](#). *Wood Sci. Technol.* 34, 3–10.
- Lopes, P., Saucier, C., Glories, Y., 2005. [Nondestructive colorimetric method to determine the oxygen diffusion rate through closures used in winemaking](#). *J. Agric. Food Chem.* 53, 6967–6973.
- Oliveira, V., Knapic, S., Pereira, H., 2012. [Natural variability of surface porosity of wine cork stoppers of different commercial classes](#). *J. Int. Sci. Vigne Vin* 46, 331–340.
- Paniagua, B., Vega-Rodríguez, M.A., Gómez-Pulido, J.A., Sánchez-Pérez, J.M., 2011. [Automatic texture characterization using Gabor filters and neurofuzzy computing](#). *Int. J. Adv. Manuf. Technol.* 52, 15–32.
- Pereira, H., Rosa, M.E., Fortes, M.A., 1987. [The cellular structure of cork from \*Quercus suber\* L. \*IAWA Bull.\* 8, 213–218.](#)
- Pereira, H., Melo, B., Pinto, R., 1994. [Yield and quality in the production of cork stoppers](#). *Holz Roh Werkst.* 52, 211–214.
- Pereira, H., Lopes, F., Graça, J., 1996. [The evaluation of the quality of cork planks by image analysis](#). *Holzforschung* 50, 111–115.
- Pereira, H., 2007. *Cork: Biology, Production and Uses*. Elsevier, Amsterdam.
- Radeva, P., Bressan, M., Tovar, A., Vitrià, J., 2002. [Bayesian classification for inspection of industrial products](#). In: Escrig, M.T., et al. (Eds.), *Proceedings of 5th Catalanian Conference Artificial Intelligence*. Springer-Verlag, Heidelberg, pp. 399–407.
- Sharma, S., 1996. *Applied Multivariate Techniques*. John Wiley & Sons, Inc.
- Vega-Rodríguez, M.A., Sánchez-Pérez, J.M., Gómez-Pulido, J.A., 2001. [Cork stopper classification using FPGAs and digital image processing techniques](#). In: *Proceedings of the Euromicro Symposium on Digital Systems Design*, Washington, DC. IEEE Computer Society, pp. 270–275.
- Vitrià, J., Bressan, M., Radeva, P., 2007. [Bayesian classification of cork stoppers using class-conditional independent component analysis](#). *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 37, 32–38.