Food Quality and Preference 59 (2017) 97–108

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Causal reasoning applied to sensory analysis: The case of the Italian wine

Silvia Golia *, Eugenio Brentari, Maurizio Carpita

Department of Economics and Management, University of Brescia, C.da S. Chiara 50, 25122 Brescia, Italy

ARTICLE INFO

Article history:
Received 21 July 2016
Received in revised form 12 February 2017
Accepted 14 February 2017
Available online 20 February 2017

Keywords:
Sensory analysis
Linear SEM
Bayesian networks
Wine quality

ABSTRACT

In this paper, a structure learning algorithm is applied to the sensory analysis field to study the factors that have an influence on the quality of Italian wines. Directed acyclic graphs, involving chemical as well as sensory variables, will be proposed to suggest hypotheses about causal connections between these variables and the Altroconsumo's Global Score of Quality, given by the Italian independent consumer's association Altroconsumo in its annual publication Guida Vini (Wines' Guide). The analysis is performed considering all types of wine included in the database, as well as red and white wines separately.

1. Introduction

Each year, the food and beverage segment is characterized by the presence of a high number of wine guides that trigger a lively discussion between the experts of the field, regarding their reliability. How the wines are chosen or what are the evaluation criteria of the wine, are common open questions. However, from the producers' side, the most relevant question concerns the most important aspects in wine making that have to be kept under control to obtain a good evaluation of the wine produced. Answering this question is, in general, extremely difficult, because the data used to rate the wines are not available to the public. In the present case, Altroconsumo, an Italian independent consumer's association, which publishes, since 2006, a yearly wine guide called *Guida Vini*, has made available to the Data Methods and Systems Statistical Laboratory (DMS StatLab) of the University of Brescia the data used to rate the wines considered in their guides from 2006 to 2013. Each year, Altroconsumo selects about 300 wines, measuring about fifty variables for each of them, including chemical and sensory variables, as well as variables of context, and then, making use of the results to establish a score, called here Global Score of Quality (GSQ), ranging from 0 to 100. Being able to access such a dataset constitutes a great opportunity for researchers to answer the producers? question regarding the most relevant aspects of a wine mainly contributing to the Altroconsumo's GSQ. The Altroconsumo database has been used toward determining wine pricing by Brentari and Zuccolotto (2011), Brentari et al. (2011) and Brentari and Levaggi (2014), as well as wine quality by Brentari et al. (2012) and Brentari and Vezzoli (2015).

The approach followed in this paper makes use of the causal reasoning (Spirtes et al., 2000) applied to sensory analysis (Lawless, 2013). Few studies have used this approach (Tenenhaus et al., 2005; Phan, 2012), but none of them has addressed the issue presented in this paper and, in this sense, this work represents a novel contribution. Causal reasoning relates to identifying the cause and effect relationships between the variables that describe an area of interest (for example the quality of a wine) to be able to provide reliable predictions about the effects of interventions. Causation is a stronger concept than correlation, which measures the existence of a linear association between two variables, and even if a strong correlation between two variables exists, it does not imply that one causes the other, given that both could have one or more common causes explaining their association. The causal reasoning process ends with the identification of a causal model, which highlights the causal relations between all the variables under study and allows one to measure the direct and indirect effects that the variables have on each other. Therefore, the aim of a causal model is to provide a system that represents the data-generating process and, through the evaluation of the causal effects, predicts how the system would respond to hypothetical interventions; as a results, the information that can be derived from it is wide. In the context of this paper, the application of the causal reasoning ends with the proposal of hypotheses about causal connections between the variables under study, highlighting the most relevant determiners of the quality of a wine, measured by the GSQ, among various chemical and sensory variables.

* Corresponding author.
E-mail address: silvia.golia@unibs.it (S. Golia).

http://dx.doi.org/10.1016/j.foodqual.2017.02.009
0950-3293/

Linear regression models are the most common example of association-based models applied to observational data used for testing causal hypothesis. However, given that they are based on association and not causation, they do not belong to the proper statistical methodology for testing causality. The justification of this practice relies on the fact that the theoretical context, represented by the particular model, itself provides the causality (Shmueli, 2010). In absence of a strong theory, the use of linear regression models as causal models is not correct. Moreover, linear regression models allow estimating only direct causal impacts, neglecting indirect effects, and this represents a limit to the amount of information that could be of interest for practitioners.

In this paper, Bayesian Networks (BNs) and Structural Equation Models (SEMs) are the specific methodologies used for discovery and testing causality. BNs (Pearl, 1988) specify, for each variable in the network, a density function as a function of the values of its causes, whereas SEMs (Bollen, 1989; Pearl, 2009) specify, for each variable, the values of the variable as a function of the values of its causes, including some unmeasured noise term. These two models are closely linked, as shown in Spirtes et al. (2000). BNs are mainly used with categorical variables; BNs,applied to both categorical and continuous variables, are uncommon and need extra constraints on their structure. SEMs involve continuous variables in their structural part and in this paper, one refers to SEM without latent variables.

As shown in the next section, BNs as well as SEMs are represented by a graph that describes the relations between the available variables which may be interpreted causally if strong assumptions are valid. In this study, this graph has been obtained merging the knowledge of oenologists with the output of the PC algorithm; the result of this merging process can be interpreted as a generation of hypotheses about the causal structure underlying the available data.

The rest of the paper is organized as follows. In Section 2 a short recall of causal modelling is given. Section 3 contains a description of the Altroconsumo database and the variables used in the study. Section 4 reports the results of the application of the theory described in Section 2 to chemical variables plus GSQ and chemical and sensory variables plus GSQ, considering all types of wine as well as red and white wines separately. Concluding remarks follow in Section 5.

2. Methods and materials

Following the definition of a causal model given in Spirtes (2010), a *causal model with free parameters* is a statistical model that specifies a set of probability densities over a given set of variables and, in addition, for each manipulation that can be performed on the population, it also specifies a set of post-manipulation probability densities over the variables. Manipulation is the result of an intervention that changes the state of the world in a specific way, for example forcing a specific value on a variable. In general, a causal model is defined in two parts: a statistical model and a causal graph, that describes the causal relations between the variables.

A graph $G = (\mathbf{V}, \mathbf{E})$ consists of a set $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ of vertices, or nodes, that represent the set of variables that define the state of the world, and a set $\mathbf{E}$ of edges, or links, that connect some pairs of vertices. The edges can be directed ($\rightarrow$) or undirected ($-$). If the graph contains only directed edges, it is called directed graph. The graph in Fig. 1 is an example of a directed graph. V1 is *parent* of V3, and V3 is *child* of V1. Moreover V1 and V3 are *adjacent* because there is an edge between them. A *path* in a graph G is a distinct sequence of vertices such that all successive pairs of vertices in the sequence are adjacent in G. A *directed path* is a sequence of adjacent edges all pointing in the same direction. For example, $V1 \rightarrow V3 \rightarrow V5$ is a directed path between V1 and V5, whereas $V1 \rightarrow V4 \leftarrow V3$ is a path between V1 and V3 but not a directed path. V5 is a *descendant* of V1 because there is a directed path from V1 to V5. If a directed graph does not contain cycles, that is, no directed paths from any vertex to itself are present, then it is acyclic and therefore, called Directed Acyclic Graph (DAG). The graph in Fig. 1 is the DAG that shows the relations between those five variables. The DAG G is a *causal* DAG for a population if it contains an edge $X \rightarrow Y$ if and only if X is a direct cause of Y in that population (Spirtes, 2010). For example, considering the DAG in Fig. 1, under the causal interpretation there is a directed edge from V1 to V3 if and only if V1 is a direct cause of V3 for the population under study.

The most frequently applied causal models belong to two broad families: the causal BNs and SEMs (Spirtes et al., 2004).

A BN (Pearl, 1988) is given by the pair (G, P), where G is a DAG defined over the set of variables $\mathbf{V}$, and P is the joint distribution of the variables in $\mathbf{V}$ that is factored according to the following rule:

$$P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P(V|Parents(G, V)) \tag{1}$$

where $Parents(G, V)$ denotes the set of parent variables of variable V in the DAG G. Fig. 1 reports the joint distribution of the five variables factorized according to the DAG shown in the right-hand side. If it is possible to give a causal interpretation to the DAG, the BN becomes a causal BN.

SEM specifies a set of structural equations, one for each variable in $\mathbf{V}$, and the distribution of the error terms. The structural equation for each variable $V_i \in \mathbf{V}$ is an equation with V_i on its left-hand side and the directed causes of V_i plus an error term, due to omitted factors, on its right-hand side. Each equation is

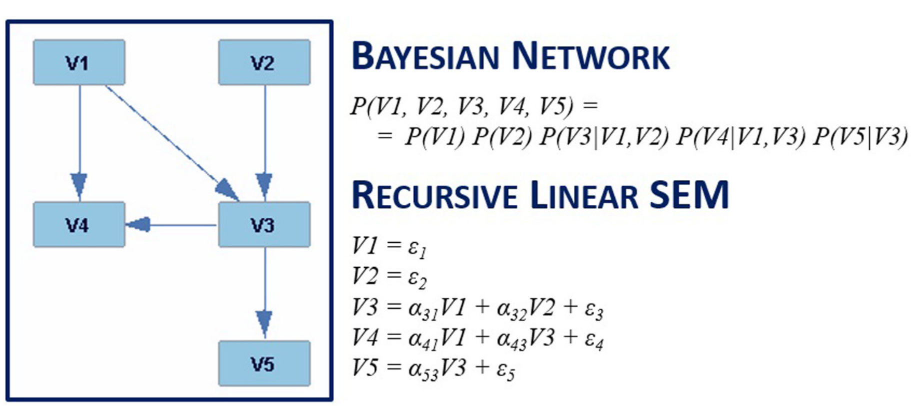

Fig. 1. Example of DAG for five variables and the corresponding BN and recursive linear SEM.

structural in the sense that it should be interpreted as an assignment process which expresses the causal relation between the dependent variable and its independent or explanatory variables which cause it (Pearl, 2009). The mathematical form of the equations can be anyone; the most common is the linear form (Bollen, 1989). Moreover, each SEM is associated with a graph that represents the causal structure of the model. If the graph is a DAG, then the SEM is said to be recursive; in this paper the recursive linear SEM will be used. Fig. 1 reports the set of linear structural equations for the five variables derived according to the DAG shown in the right-hand side.

In order to find causal relationships from data, the gold standards are interventional experiments that provide interventional data. However, there are situations in which it is not feasible to do interventional studies and the available data come from observational studies, as in the present paper. In the latter case, it is reasonable to wonder if it is possible to identify causal relationships by observation alone. In the last twenty years, theories were developed thanks to which, under strong and suitable assumptions, it is possible to recover the causal structure, or at least the equivalence class to which the true structure belongs (Complete Partially DAG (CPDAG), Chickering, 2002), that can be interpret as representing the causal relationships linking the available variables together (Kalisch and Bühlmann, 2014; Pearl, 2009; Spirtes et al., 2000; Spirtes, 2010). In this paper, the CPDAG is derived by applying the PC algorithm (Spirtes et al., 2000; Kalisch and Bühlmann, 2007) implemented in the free software TETRAD V (Spirtes et al., 2010). This algorithm is computational feasible and proofs of its asymptotic consistency have been given in Spirtes et al. (2000), generalized to the high-dimensional setting by Kalisch and Bühlmann (2007) and Harris and Drton (2013). The PC algorithm makes use of a series of conditional independence tests, whose choice relies on the nature of the variables considered. If the variables are continuous and normally distributed, it is possible to use the Fisher's z statistic, otherwise, in the case of non-normality, the conditional independence test for non-normal variables proposed by Ramsey (2014) can be used. When the variables are all categorical, the conditional independent tests can be based on the G^2 or the X^2 statistics. Nevertheless, in simulations, Spirtes et al. (2000) found that G^2 leads to the correct graph more often than X^2.

The PC algorithm is based on certain strong assumptions whose validities are difficult to check in practice, even if it is reasonable to assume their validities in most of the cases (Kalisch et al., 2010; Pearl, 2009). To explore the stability of the CPDAGs learnt in this paper, a bootstrap study has been performed and the number of times, in terms of percentage, that the edges present in the CPDAGs obtained from the original data were found in the bootstrap replications, was considered.

When possible, it is suitable to consider specific experts' knowledge during the searching step; this limits the possible causal graphs found by any searching algorithm, keeping it from exploring graphs that contain oriented edges showing unrealistic causal connections. On the other hand, the imposition of such constraints has to be done very carefully, given that the constraints condition the results of any searching algorithm considerably. In this paper, the experts' knowledge was formalized through the identification of forbidden edges, meaning that if a relationship between two variables connected by a forbidden edge exists, it is represented by an arrow with orientation opposite to the one expressed by the forbidden edge. The set of these forbidden edges can be originated from a tiers ordering, which illustrates an ordering in the variables, meaning that variables in higher-numbered tiers can cause, but not be caused by, the variables in lower-numbered tiers. Recently, Oates et al. (in press) have proposed a new approach that takes into account experts' knowledge in an explicit and strong way.

The statistical models used in the paper are the recursive linear SEM for chemical variables and the BN for chemical and sensory variables. The parameters in the linear SEM are estimated making use of the maximum likelihood method implemented in the R package lavaan (Rosseel, 2012). Given that the variables involved do not follow a multivariate normal distribution, the Satorra-Bentler scaling procedure is used in order to obtain robust standard errors and fit indices. The fit indices used to evaluate the goodness of fit of the model are the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), and the comparative fit index (CFI). The conditional probabilities contained into the factorization of the joint distribution given by the Eq. (1) for the BN, are estimated making use of the maximum likelihood method implemented in the free software TETRAD V (Spirtes et al., 2010). In order to evaluate the goodness of fit of the model, a Chi Square test (the null hypothesis implies a completely disconnected graph) is used. For all the BNs used in Section 4.2, the hypothesis of a completely disconnected graph is rejected at the usual significant level.

3. The variables of interest

This section contains a brief description of the Altroconsumo database and the variables used in this study. The dataset considered in this paper was created starting from the database produced by Altroconsumo for its publication Guida Vini from 2006 to 2013. Each year, about 300 wines were bought and some chemical and sensory characteristics, plus other variables of context, were measured. The wines were chosen to represent the variety of Italian vineyards, producers, and regions of origin. Additionally, for each year, different vineyards and producers were considered, so that the observations could be considered as independent. The database is composed of variables that measure chemical as well as sensory characteristics of a wine, two global score variables, and some exogenous variables. The dataset used in this paper comprises all the chemical variables, one global score and a selection of sensory and exogenous variables, chosen among the variables that are relevant for all types of wine (examples of excluded variables are type of cap or award gained).

The global score considered in this work is the Global Score of Quality (*Global.Score*); it is an indicator of the overall quality of the wine and assumes a score ranging from 0 to 100, with higher scores indicating better quality of the wine. In the analysis that follows, this variable will also be used in association with the sensory variables, which are categorical, so it is necessary to provide a categorization of it. The variable *Global.Score.Cat* is the categorical version of *Global.Score* and is generated according to the cut points reported in Table 1. This categorization is a finer version of Altroconsumo's classification, which labels wines with score ranging from 65 to 100 with a red point, wines with scores between 41 and 64 with a green point, whereas a black point is associated with wines with scores under 41. Given that the dataset contained only 15 wines tagged with a black dot, the first category (Very low) includes these very low scores plus half of the range of the scores labelled with a green dot.

Between the available exogenous variables, the ones considered are the type of wine, the designation of origin, and the region of production. The variable type of wine (*Type*) assumes two categories, still or sparkling white wine and still or sparkling red wine. The variable designation of origin (*Design*) assumes three categories, DOC (controlled designation of origin), DOCG (controlled and guaranteed designation of origin) and other designations (DOP, IGP, IGT), and the variable region of production (*Region*) assumes three categories, North, Center, South and Islands. Table 2 reports the distribution of the wines according to these three variables.

Table 1
Categorization of the GSQ

Global.Score	0–54	55–59	60–64	65–74	75–100
Global.Score.Cat	Very Low	Low	Medium	High	Very High

Table 2
Frequency distribution of the wines according to the variables type of wine, designation of origin and region of production.

Type of wine	Red 1184	White 785	
Designation of origin	DOC 1273	DOCG 269	Other 427
Region of production	North 900	Center 491	South and Islands 578

3.1. The chemical variables

The chemical variables evaluated by Altroconsumo were the total and the volatile acidity (*Acidity.Tot* and *Acidity.Vol*), the residual sugar (*Sugar*), the wine's verified alcoholic strength (*Verif.Alcohol*), the total sulfur dioxide (*SO2.Tot*), and the ratio between free and total sulfur dioxide, from which the free sulfur dioxide (*SO2.Free*) was obtained. These variables are the ones that are usually used in the evaluation of the chemical components of a wine and they are also the variables that can be manipulated by the oenologist. All these variables have a non-normal distribution (p-values of the Jarque–Bera normality test – Bera and Jarque (1980) – do not support the hypothesis of normal distribution) with the exception of the total sulfur dioxide. Table 3 reports the means and standard deviations of these variables found in the Altroconsumo sample, considering all the wines together and red and white separately. The two sulfur dioxides show a high degree of variability.

Acidity is a measurement of the quantity of organic acids present in a wine; these acids give wines their characteristic crisp, slightly tart taste. The acids are divided into two groups: volatile (primarily acetic acid but also lactic, formic, butyric, and propionic acids) and nonvolatile or fixed (mainly tartaric, malic, citric, and succinic acids) acids. The total acidity represents the total amount of acids (volatile and fixed) in wine and is expressed in terms of grams of tartaric acid per liter (g/L) of wine whereas the volatile acidity represents the amount of volatile acids in wine and is expressed in terms of grams of acetic acid per liter (g/L) of wine.

Residual sugar is the amount of sugar not converted to alcohol during fermentation and is usually measured in grams of sugar per liter (g/L) of wine. Residual sugar is indicative of a wine's relative sweetness.

The wine's verified alcoholic strength represents the actual alcoholic strength by volume of wine (expressed by the symbol % vol.) verified in the analysis made by Altroconsumo. The level of alcohol represents a natural protection against ageing and oxidation.

Sulfur dioxide (SO_2) is a multifaceted antiseptic and a powerful reducing agent that protects against oxidation. It is composed by a combined and a free form and its concentration is usually expressed in milligrams per liter (mg/L). The SO_2 in the free form is the fraction of SO_2 that is effective as an antioxidant and antiseptic. In order to have a balanced wine, it is necessary to have a sufficient quantity of sulfur present to maintain the free SO_2 levels needed to protect the wine. However, SO_2 total levels must not be high enough to be noticeable when the wine is consumed. Therefore, the total and free SO_2 contents of wine are key analytical parameters for must and wine quality control.

Both the alcohol and SO_2 content levels have a direct influence on the wine's resistance to shipping and temperature exposure during transit and storage.

Table 3
Mean and standard deviation, in brackets, of the chemical variables.

Variable	All	Red	White
Acidity.Tot	5.49 (0.48)	5.55 (0.46)	5.39 (0.51)
Acidity.Vol	0.36 (0.14)	0.45 (0.11)	0.23 (0.07)
Sugar	4.10 (2.38)	4.04 (2.32)	4.08 (2.21)
Verif.Alcohol	12.66 (0.82)	12.88 (0.81)	12.33 (0.71)
SO2.Tot	81.96 (27.39)	71.28 (25.29)	98.07 (21.95)
SO2.Free	17.01 (9.23)	14.73 (8.34)	20.55 (9.44)

3.1.1. Categorization of the chemical variables

To use the chemical variables in association with the sensory variables, which are categorical, it is necessary to transform them into categorical variables. The categorization task has to be performed with caution, because it may influence the quality of the learned causal structure. When possible, the discretization is made using an "expert" approach, meaning that the choice of the cut points is driven by the knowledge of experts, otherwise, a "statistical" approach is followed, choosing the cut points considering the characteristics of the probability distribution. The chemical variables discretized using an "expert" approach were the residual sugar, the alcohol content and the total and volatile acidities.

Wines are classified as dry, semi-dry, medium, or sweet, in order to give an idea of how sweet the wine tastes, and this classification depends on the wine's residual sugar content. Dry wines have typically up to 4 g/L residual sugar, semi-dry wines up to 12 g/L, medium wines up to 45 g/L and sweet wines possess over 45 g/L. These thresholds can be used as natural potential cut points in the discretization of the variable *Sugar*. Due to the distribution of this variable in the available dataset (1st Qu. = 2.5, Median = 3.4, 3rd Qu. = 5, Max. = 27.6), the new categorized variable *Sugar.Cat* will take modality "Dry Wine" if *Sugar* assumes values up to 4, and modality "Not Dry Wine" otherwise.

The alcohol content of a wine typically ranges between 8% and 17% vol. and a typical classification is the one shown in Table 4. Using this classification on the available dataset, the category "Very High" regroups only 37 observations, so it is merged with the category "High". The resulting categorized variable *Verif.Alcohol.Cat* will take three modalities that are "Very Low", "Moderately Low" and "High and Very High".

All wines have volatile acidity but it is important that it does not reach a detectable level. The so-called aroma threshold, which is the level over which the acidity is detectable, varies depending on the context of the wine and the sensitivity of the person sniffing, but is generally comprised between 0.6 and 0.9 g/L. Moreover, if the winemaker has done his/her work properly, most finished wines typically have acetic-acid levels of 0.3–0.5 g/L (Goode and Harrop, 2011). These considerations suggest the possibility to discretize the variable *Acidity.Vol* as reported in Table 5, creating the new variable *Acidity.Vol.Cat*.

Table 4
Categorization of the alcohol content of a wine.

% vol.	$\leqslant$ 12.5	13–13.5	14–14.5	> 14.5
Classification	Very Low	Moderately Low	High	Very High

Table 5
Categorization of volatile acidity.

g/L	< 0.3	[0.3–0.6)	$\geqslant$ 0.6
Classification	Low	Medium	High

The range for total acidity of a wine is in general comprised between 4 and 10 g/L. However, most people would find a wine with 10 g/L acidity too tart to drink and 4 g/L too flat. Moreover, a wine with total acidity close to 4 g/L is more susceptible to spoilage. A general range of total acidity for a balanced wine is 4.5–7.5 g/L (Altroconsumo, 2006–2013), and therefore, these values could be suitable cut points for this variable. Nevertheless, the distribution of this variable in the available dataset is summarized as follows: Min. = 4.01, 1st Qu. = 5.17, Median = 5.44, 3rd Qu. = 5.76, Max. = 8.29, and the use of the previous values would generate a categorical variable with almost all the observations concentrated in the central class, which would make this variable problematic for the analysis that will follow. Based on these considerations, the categorized version of the variable *Acidity.Tot*, *Acidity.Tot.Cat*, follows the rules reported in Table 6.

The European Union has set a legal limit for total SO_2 of 150 mg/L in red wines and 200 mg/L in white wines. However, because some individuals are sensitive to SO_2, it is mandatory to include 'contains sulfites' on the label if the total SO_2 is over 10 mg/L. Except for those bounds, there are no indications for convenient cut points for the total sulfur dioxide and, therefore, a statistical approach must be implemented in order to discretize this variable. The distribution of this variable is normal (p-value of the Jarque–Bera Normality Test = 0.2435), so a reasonable choice for the cut points are the tree quartiles (1st Qu. = 63, Median = 82, 3rd Qu. = 100), and the categories of the new variable *SO2.Tot.Cat* are reported in Table 7.

There are no indications from the experts regarding the selection of convenient cut points for the free sulfur dioxide. Therefore, a statistical approach must be implemented in order to discretize this variable. Even if the distribution of this variable is not normal, it is unimodal with skewness equals to 0.488 and kurtosis equal to 3.091. As a result, a reasonable choice for the cut points are the tree quartiles (1st Qu. = 10.08, Median = 16.48, 3rd Qu. = 22.89), and the categories of the new variable *SO2.Free.Cat* are reported in Table 8.

3.2. The sensory variables

The sensory characteristics considered in this study can be divided into four groups representing visual, olfactory, and gustatory characteristics of a wine and its intense aromatic persistence. The visual characteristics of a wine describe how a wine appears at a visual inspection, and they include the intensity of the color (*Color.Int*) and how pleasant the aspect of the wine is (*Attraency*). The olfactory characteristics are related to the wine aroma and can be represented by the intensity of the bouquet (*Olfact.Int*), that is a measure of quantity and not necessarily quality, several fragrances that can be perceived in a wine, such as floral (*Floral*), fruity (*Fruity*), spicy (*Spicy*) and vegetal (*Vegetal*), and the olfactory cleanness (*Olfact.Clean*) and quality (*Olfact.Qual*). The gustatory characteristics are connected to taste and mouthfeel of a wine which are described by its structure (*Structure*), the harmony of the different components measured by roundness (*Roundness*), gustatory harmony (*Gustatory.Harmony*), and the type of taste or mouthfeel sensation such as sourness (*Sourness*) and bitterness (*Bitterness*). The intense aromatic persistence is described by the persistence of aromas (*Persistence*), the aftertaste cleanness (*Aftertaste.Clean*), and quality (*Aftertaste.Qual*) and the aromatic richness (*Arom.Rich*).

These variables were evaluated with the help of Brescia's *Centro Studi Assaggiatori*, the most advanced unit of sensory analysis in Italy. About 21 judges, divided into three panels, evaluated the sensory characteristics of wines already described. They were all experienced judges with several specific qualifications, grouped into balanced panels in terms of age, sex, and experience. The tasting was blind with replication. The test used was the TrialTest developed by *Centro Studi Assaggiatori*, based on a form containing objective (parameters of evaluation) and hedonic (linked to the pleasure of the judges) describers divided following the canonical phases of the sensory evaluation, that is visual, olfactory, gustatory and retronasal evaluations of a wine. The judges, for each wine analyzed, were asked to score the perception of each sensory variable considered, using a 0–9 scale, where 0 denotes the absence and 9 the maximum perception; the median score was the final score recorded.

Due to the distribution of these sensory variables in the available dataset, it was necessary to properly merge the observed scores following the indications reported in Table 9. These cut points derive from the indication of experts in sensory analysis applied to wine, and they are commonly used as compact subdivisions of the original and finer scale.

4. Results and discussion

As stated previously, in order to generate hypotheses about the causal structure underlying the available data, the first step consists in searching the space of graphical structures, and the next step is in assigning a causal interpretation to that structure through a suitable model. The following two subsections report

Table 6
Categorization of total acidity.

g/L	⩽ 5	(5–6)	⩾ 6
Classification	Low	Medium	High

Table 7
Categorization of the total sulfur dioxide of a wine.

mg/L	⩽ 63	(63–82]	(82–100]	> 100
Classification	Very Low	Moderately Low	High	Very High

Table 8
Categorization of the free sulfur dioxide of a wine.

mg/L	⩽ 10.08	(10.08–16.48]	(16.48–22.89]	> 22.89
Classification	Very Low	Moderately Low	High	Very High

Table 9
Categorization of the sensory variables.

Variable	Low	Medium	High
Color.Int	0–4.5	5.0–7.0	7.5–9.0
Floral	0–2.5	3.0–5.5	6.0–9.0
Fruity	0–3.5	4.0–5.5	6.0–9.0
Spicy	0–2.0	2.5–4.5	5.0–9.0
Vegetal	0–1.5	2.0–4.0	4.5–9.0
Olfact.Clean	0–5.5	6.0–7.0	7.5–9.0
Olfact.Int	0–5.5	6.0–6.5	7.0–9.0
Olfact.Qual	0–5.5	6.0–7.0	7.5–9.0
Structure	0–4.5	5.0–7.0	7.5–9.0
Roundness	0–5.0	5.5–6.5	7.0–9.0
Sourness	0–3.5	4.0–5.0	5.5–9.0
Arom.Rich	0–4.5	5.0–7.0	7.5–9.0
Persistence	0–5.5	6.0–7.0	7.5–9.0
Aftertaste.Qual	0–5.5	6.0–7.0	7.5–9.0
Variable		Low	High
Attraency		0–6.5	7.0–9.0
Bitterness		0–3.5	4.0–9.0
Gustatory.Harmony		0–6.0	6.5–9.0
Aftertaste.Clean		0–6.5	7.0–9.0

the proposed CPDAGs and the corresponding models for the Altroconsumo database when the continuous chemical variables plus the GSQ Section 4.1, or the discretized chemical and sensory variables plus the discretized version of the GSQ and the three exogenous variables Section 4.2 are considered.

4.1. Chemical variables versus Global Score of Quality

In this subsection, the relations between the continuous chemical variables plus the GSQ are analyzed. The analysis will be conducted considering the red and white wines together, as well as separately.

Experts in the field entailed the constraints visualized in Fig. 2, where the edges in the right-hand side graph are forbidden edges originated from the tiers ordering shown in the left-hand side graph. The explanation of these constrains relies on the following reasoning. Alcohol is the main and most important product of the alcoholic fermentation, which converts the sugar contained in the grape juice into ethyl alcohol and carbon dioxide by means of yeast's enzymes. Before the beginning of the alcoholic fermentation, the oenologist is able to understand, analyzing the must, what the potential alcoholic strength of the must could be and, depending of what kind of wine he/she wants to achieve, he/she will act accordingly. Residual sugar is the natural grape sugar that is leftover after fermentation ceases (either intentionally or not), so it is inserted in tier 2. Among the most important byproducts of the alcoholic fermentation there is the acetic acid, which is the primary volatile acid in wine. Moreover, after the alcoholic fermentation, the malolactic fermentation, which converts malic acid into lactic acid, with the effect of lowering the total acidity, can occur. Therefore, the two acidities are located in tier 3. Free sulfur dioxide follows in the tier ordering due to its role as a wine protector.

Fig. 3 reports the result of the application of PC algorithm on the available data, considering red and white wines together. The bootstrap study found that the edges in this DAG were quite stable, given that all the edges were present at least in 80% of bootstrap samples, with the exception of the edge connecting the verified alcoholic strength to the total sulfur dioxide, whose presence was found in 52.5% of the bootstrap replications.

The graph proposed in Fig. 3 hypothesizes that the direct causes of the GSQ are the total sulfur dioxide and the verified alcoholic strength, whereas the residual sugar, the sulfur dioxide in the free form and the volatile acidity are indirect causes, whose effects on the GSQ are mediated by other variables. From this DAG, it is possible to derive a linear SEM that can be seen as a proposal of a causal model for the variables under study. Fig. 3 shows the estimates of the coefficients present in the linear SEM implied by the DAG, as edge weights. The values of the fit indices used to evaluate the goodness of fit of the model resulted as: SRMR = 0.032, RMSEA = 0.066, and CFI = 0.965; these values indicate an acceptable goodness of fit of the model to the data (West et al., 2012). From Fig. 3 it is possible to notice a negative direct effect of an increase in total sulfur dioxide on the GSQ, and a positive direct effect on the GSQ if the alcoholic strength is increased. From the graph, it is also possible to read the total effect of a variable on another one. This total effect defines the effect that a variable has, which can be direct only or direct as well as mediated by other variables. When a linear SEM is used, the total effect of X_i on X_j can be obtained multiplying the edge weights along each directed path from X_i to X_j, and then summing over the directed paths, if there is

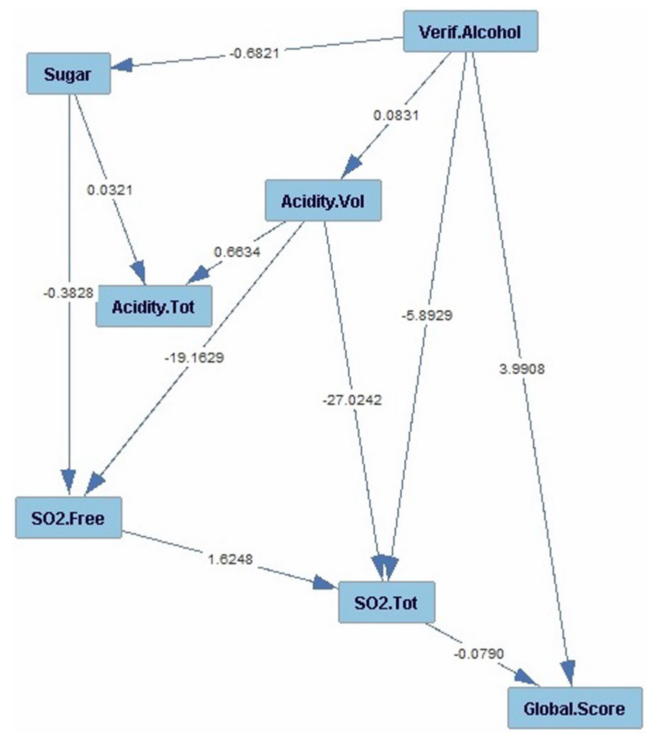

Fig. 3. Red and white wines. The causal DAG with the estimates of the linear SEM parameters as edge weights for the chemical variables and the GSQ.

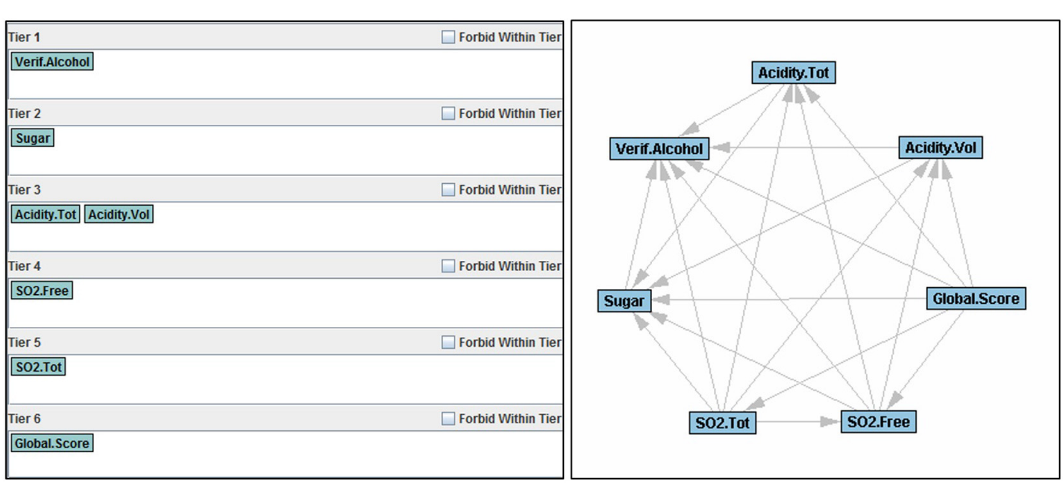

Fig. 2. Constraints due to domain knowledge for the chemical variables plus the GSQ.

Table 10
Total effect on the GSQ of the increase of 1 unit (unstandardized) or 1 standard deviation unit (standardized) of each variable.

Variable	Unstandardized	Standardized
Verif.Alcohol	4.805	0.452
SO2.Tot	−0.079	−0.246
SO2.Free	−0.128	−0.135
Acidity.Tot	0.000	0.000
Acidity.Vol	4.595	0.075
Sugar	0.049	0.013

more than one. Table 10 reports the total effect of the increase of one unit of each variable on the GSQ. To be able to compare the effects of the variables on the GSQ and to establish their relative importance, it is necessary to reduce the coefficients to a common unit (standard deviation) through standardization. In this case, the standardized effect represents the effect of the increase of one standard deviation unit of a variable on the target one. Table 10 reports the total effect calculated using the standardized coefficients. The variable for which one standard deviation unit increase has the biggest total effect on GSQ is the verified alcoholic strength, followed by the total sulfur dioxide.

Different graphs and results, that highlight the differences in making red and white wines, were obtained when the wines were considered separately depending on their type. Fig. 4 displays the result of the application of PC algorithm on the available data, considering red and white wines separately. The bootstrap study found that the edges in these DAGs were quite stable, given that all the edges were present in at least 95% and 85% of bootstrap samples generated considering red and white wines, respectively, with the exception of the edge connecting the verified alcoholic strength to the total sulfur dioxide, whose presence was found in 60% and 45% of the bootstrap replications generated considering red and white wines, respectively.

For both types of wine, the total sulfur dioxide and the verified alcoholic strength are proposed as the variables that exert a direct cause on the GSQ. Moreover, considering the graphs as a whole, the two DAGs highlight the difference in making red and white wine. For example, when making white wine, great attention has to be employed in controlling the volatile acidity in order to preserve the cleanness of the wine, whereas for red wine, the level of total acidity is the factor that will determine how clean the wine is.

Fig. 4 also shows the estimates of the coefficients present in the linear SEMs implied by the DAGs as edge weights. The fit indices for white wines were: SRMR = 0.047, RMSEA = 0.065, and CFI = 0.910, whereas for red wines they were: SRMR = 0.04, RMSEA = 0.078, and CFI = 0.923. In both cases, these values indicate an acceptable goodness of fit of the model to the data. From both graphs, it is possible to notice a negative direct effect of an increase in total sulfur dioxide on the GSQ, and a positive direct effect on the GSQ if the alcoholic strength is enhanced. From Fig. 4, it is also possible to read the total effect of the increase of one unit of each variable on the GSQ reported in Table 11. Table 11 also reports the total effect calculated using the standardized coefficients. For both white and red wines, the variable for which one standard deviation unit increase has the biggest total effect on GSQ is the verified alcoholic strength, followed by the total sulfur dioxide.

4.2. Chemical and sensory variables versus Global Score of Quality

This subsection contains the results of the analysis of possible causal relations between the discretized chemical variables, the sensory variables and three exogenous variables plus the discretized GSQ. The analysis was conducted considering red and white wines either together or separately.

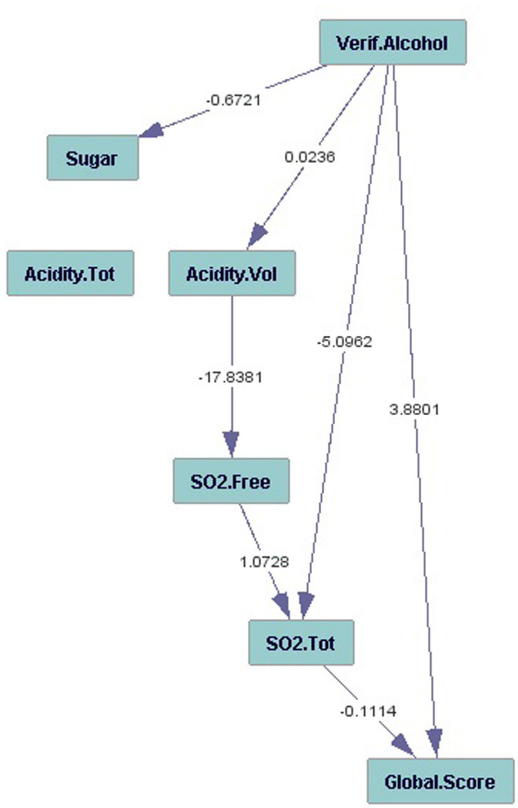

Fig. 4. Chemical variables and the GSQ: the causal DAG with the estimates of the linear SEM parameters as edge weights for white (left) and red (right) wines.

Table 11
Red and white wines separately. The total effect on the GSQ of the increase of 1 unit (unstandardized) or 1 standard deviation unit (standardized) of each variable.

	White wines		Red wines	
Variable	Unstandard	Standard	Unstandard	Standard
Verif.Alcohol	4.498	0.404	4.414	0.401
SO2.Tot	−0.111	−0.309	−0.064	−0.181
SO2.Free	−0.119	−0.142	−0.119	−0.111
Acidity.Tot	0.000	0.000	0.406	0.021
Acidity.Vol	2.132	0.019	0.000	0.000
Sugar	0.000	0.000	0.021	0.006

Background knowledge, in the form of the constraints derived from the tiers ordering shown in Fig. 5, was taken into account during the searching step. This tiers ordering originated from the fact that all the sensory variables can be caused by, but cannot cause, the chemical variables; therefore, they are contained in the lower tiers. The tiers ordering for the chemical variables is the one discussed in the previous subsection, whereas the tiers ordering of the sensory variables relies on the sequence of their evaluation. In fact, when a tasting section starts, judges first consider the visual characteristics of a wine, then, they evaluate the olfactory characteristics, followed by the gustatory characteristics, and finally, they rate aspects of a wine that describe its intense aromatic persistence. Moreover, the olfactory characteristics were split in two tiers; the several fragrances that can be perceived in a wine were inserted in the highest tier, whereas olfactory intensity, quality, and cleanness were placed in the lowest.

Fig. 6 shows the result obtained through the application of PC algorithm on the available data, considering red and white wines together. The bootstrap study found that the edges in this CPDAG were quite stable; 91.2% of the edges present in the CPDAG

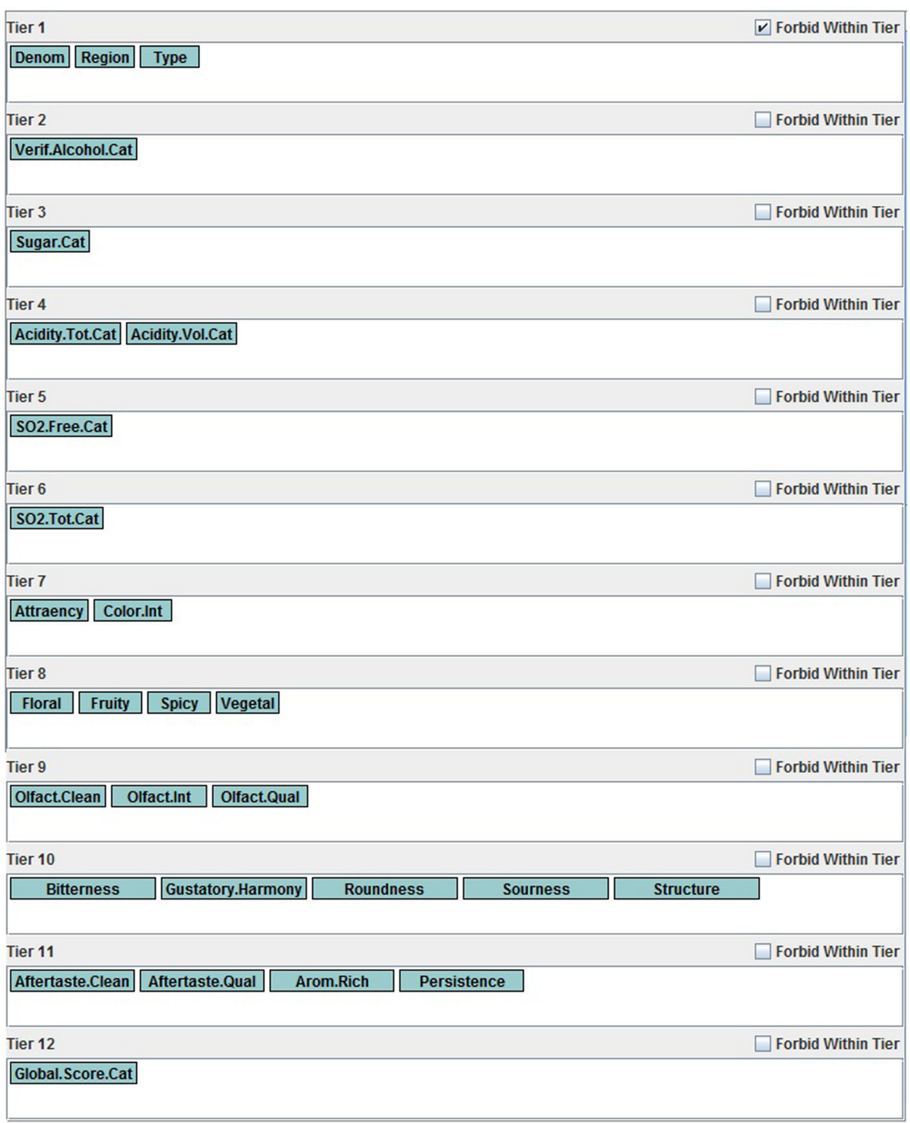

Fig. 5. Constraints due to domain knowledge for the discretized GSQ, chemical, sensory and exogenous variables.

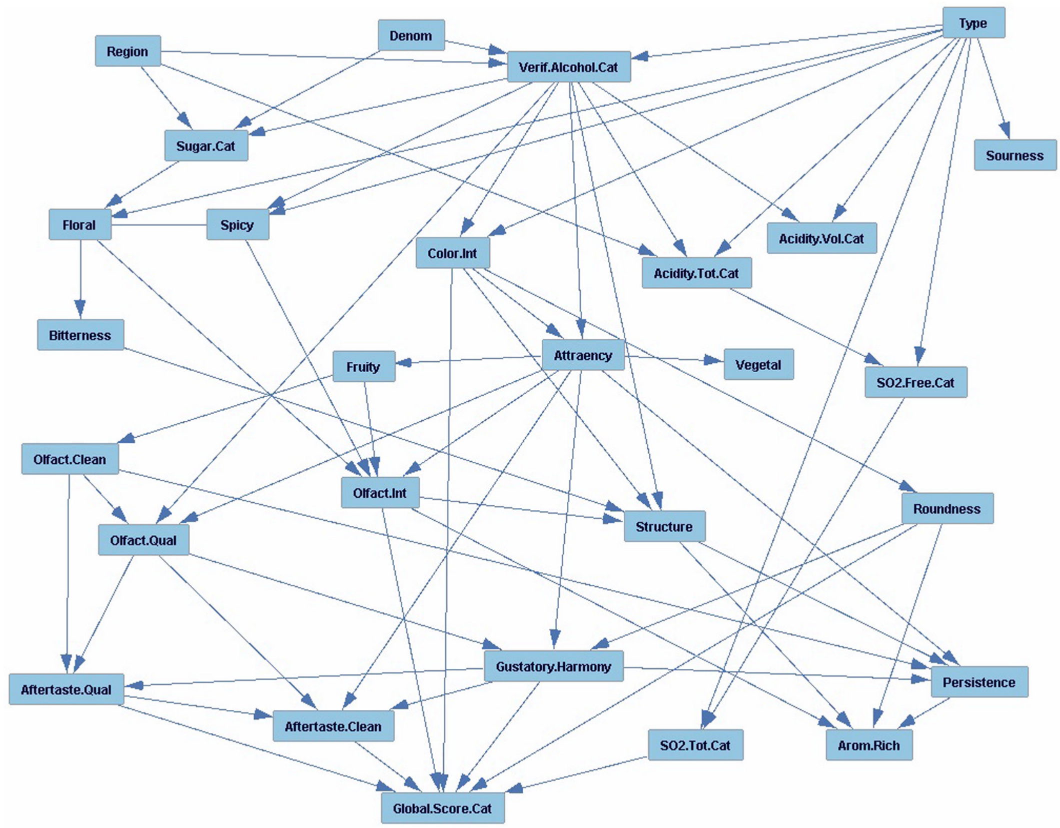

Fig. 6. Red and white wines. The causal CPDAG for chemical and sensory variables and the GSQ.

obtained from the original data, were included in at least 50% of the bootstrap samples, and among them, 90.3% appeared in at least 75% of the bootstrap replications.

From the resulting CPDAG, it is possible to notice that between the chemical variables, the total sulfur dioxide is the only one that exerts a direct effect on the GSQ. Considering the sensory variables, the variables that have a direct impact on the GSQ are related to all the characteristics of a wine (appearance, bouquet, flavor, and intense aromatic persistence). It is interesting to observe that between the olfactory characteristics of a wine considered here, the intensity of the bouquet is the only one with a directed link to the GSQ; this suggests that the intensity of the bouquet, and not just some particular perfumes, affects the GSQ.

The CPDAG in Fig. 6 contains only one non-oriented edge, the one between Floral and Spicy. As a result, from this graph, it is possible to read two BNs, which differ for the orientation of this arch.

The estimates of the conditional probabilities contained in Eq. (1) allow to identify the profiles of the variables with a direct impact on the GSQ that maximize the conditional probability of the GSQ to score very high. These profiles are reported in Table 12. The ordering shown in the table is based on the evaluation of the sensory aspects (best aspects contribute most) and the amount of sulfites (lowest amounts contribute most). The number of wines that received a score belonging to the class *Very High* of the variable *Global.Score.Cat* were 165. All these profiles have in common the maximum evaluation of the intensity of the bouquet and the aftertaste cleanness, and a medium or high score for the remaining sensory variables. Regarding the total sulfur dioxide, the results suggest that the amount of sulfites in a wine has a low impact on the evaluation of its quality given by Altroconsumo if the examination of the sensory variables ends with a good evaluation.

Table 12
Profiles of the variables that make the conditional probability of categorized version of the GSQ equal to *Very High*, bigger than 0.5.

SO2.Tot.Cat	Color.Int	Olfact.Int	Roundness	Gustatory.Harmony	Aftertaste.Clean	Aftertaste.Qual
VL	H	H	H	H	H	H
VL	H	H	M	H	H	H
VL	H	H	H	H	H	M
VL	M	H	M	H	H	H
ML	H	H	M	H	H	H
ML	M	H	M	H	H	H
ML	M	H	H	H	H	M
H	H	H	H	H	H	H
ML	M	H	M	L	H	H
H	H	H	M	H	H	H
H	M	H	M	H	H	H

VL = Very Low, ML = Moderately Low, L = Low, M = Medium, H = High.

When red and white wines are considered separately, the CPDAGs highlight some of their peculiarities; Fig. 7 displays the result of the application of PC algorithm on the available data, considering red and white wines separately. The bootstrap study found that the edges in these CPDAGs were quite stable; 90.2% and 90.7% of the edges present in the CPDAG obtained from the

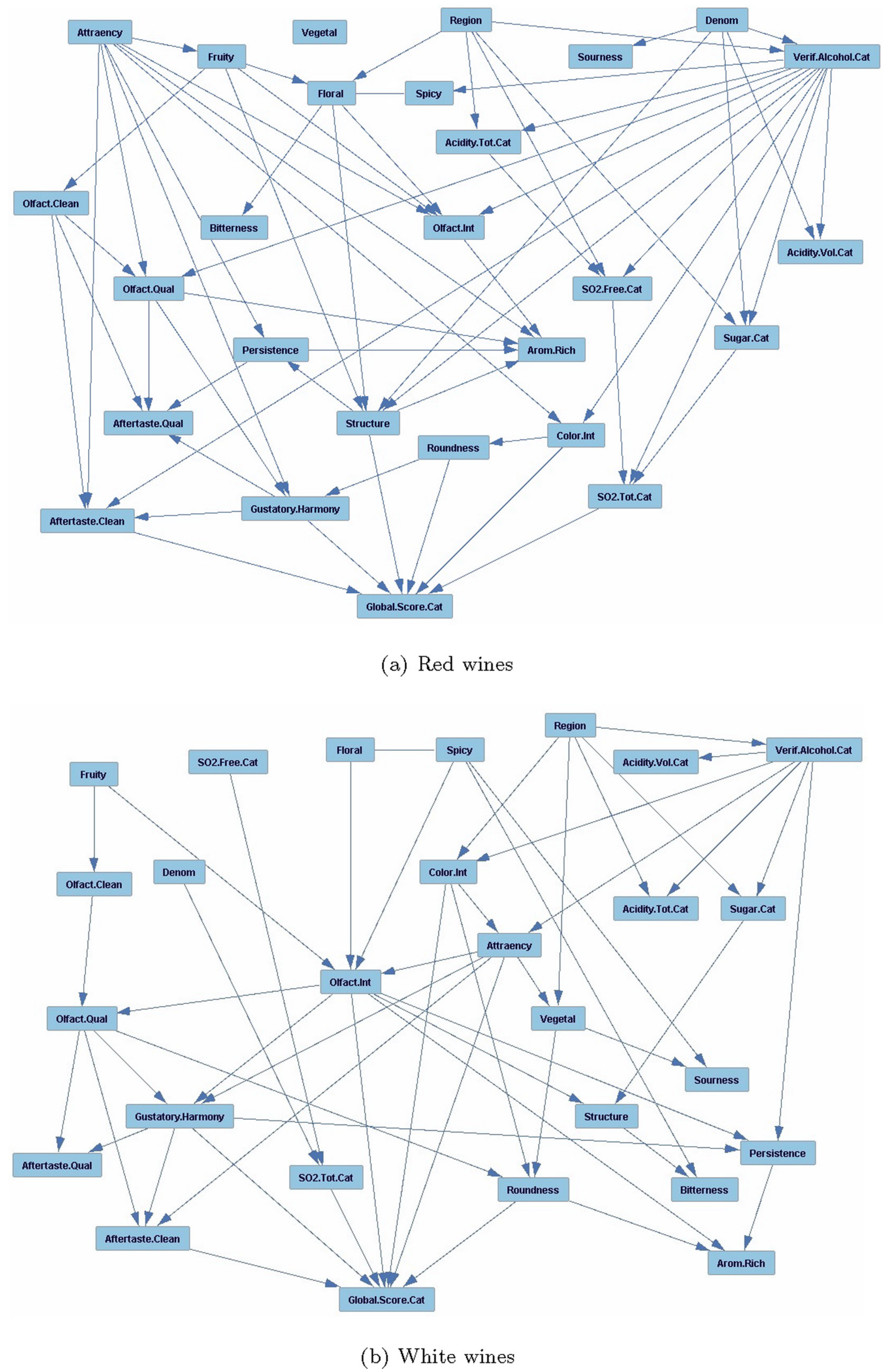

Fig. 7. Red and white wines separately. The causal DAG for chemical and sensory variables and the GSQ.

Table 13
Red wines. Profiles of the variables that make the conditional probability of categorized version of the GSQ equal to *Very High*, bigger than 0.5.

SO2.Tot.Cat	Color.Int	Structure	Roundness	Gustatory.Harmony	Aftertaste.Clean
VL	H	H	H	H	H
VL	H	M	H	H	H
VL	H	H	M	H	H
ML	H	H	H	H	H
ML	H	H	M	H	H
H	H	H	H	H	H
VL	H	H	M	L	L
H	H	H	M	H	H
ML	M	M	H	L	H
ML	M	H	M	L	H
ML	H	M	M	H	L

VL = Very Low, ML = Moderately Low, L = Low, M = Medium, H = High.

Table 14
White vines. Profiles of the variables that make the conditional probability of categorized version of the GSQ equal to *Very High*, bigger than 0.5.

SO2.Tot.Cat	Color.Int	Attraency	Olfact.Int	Roundness	Gustatory.Harmony	Aftertaste.Clean
ML	M	H	H	H	H	H
H	H	H	M	M	H	H

ML = Moderately Low, M = Medium, H = High.

original dataset for red and white wines, respectively, were included in at least 50% of the bootstrap samples, and among them, 89.1% and 83.3% appeared in at least 75% of the bootstrap replications obtained from the dataset for red and white wines, respectively.

The two proposed CPDAGs show that there are some common variables exerting a direct impact on the GSQ (total sulfur dioxide, roundness, gustatory harmony, aftertaste cleanness, and the intensity of the color of a wine) and other variables that are peculiar to red and white wines respectively and they highlight the differences among them. With respect to red wines, there is an extra sensory variable that has a direct impact on the determination of the GSQ, that is the structure. In general, red wines are more full-bodied than white wines, which are lighter wines and therefore, for those, the structure is less relevant. As a result, for red wine, the body of the wine appears as a discriminant factor on the quality of the wine. Considering white wines, the CPDAG reveals two extra sensory variables that have a direct impact on the GSQ, that is, how pleasant its aspect is, and the intensity of the bouquet. For white wines, the visual aspects are in general more pronounced than for red wines; for example, from a visual inspection it is possible to get an idea of the wine clarity and therefore, of its quality. Moreover, in general, white wines are less scented than red wines and therefore, being able to clearly perceive the fragrance in a white wine is an indication of quality.

The two CPDAGs in Fig. 7 contain only one non oriented edge, the one between Floral and Spicy. Hence, from each of them, it is possible to read two BNs, which differ for the orientation of this arch.

The estimates of the conditional probabilities contained in Eq. 1 allow to identify the profiles of the variables with a direct impact on the GSQ that maximize the conditional probability of the GSQ to score very high. These profiles are reported in Tables 13 (red wines) and 14 (white wines).

The number of red wines that received a score belonging to the class *Very High* of the variable *Global.Score.Cat* was 143. Almost all these profiles have in common the maximum evaluation of the intensity of the color, and medium or high evaluation of roundness and structure. Half profiles show a low rating of the gustatory harmony balanced by a high rating of aftertaste cleanness, or vice versa. Regarding the total sulfur dioxide, the results suggest that the amount of sulfites in a red wine has a low impact on the evaluation of its quality given by Altroconsumo if the examination of almost all the sensory variables ends with a good evaluation.

The number of white wines that received a score belonging to the class *Very High* of the variable *Global.Score.Cat* was very low (22 wines), explaining the low number of profiles in Table 14. These two profiles are in line with what was found previously, confirming that the wine gets a good rating if the evaluation of the sensory variables is medium or high, irrespectively of the level of total sulfur dioxide.

5. Conclusions

The aim of the study described in this paper was to try to provide an answer to the producers' question regarding what they should keep under control during the vinification to obtain a good product that can obtain a good evaluation in the wine guides. The data available are connected with the wine guide edited by Altroconsumo and, strictly speaking, the results obtained should be considered as consistent with the Altroconsumo policy in rating wines only. Nevertheless, it can be reasonable to presume that the wine guides follow common standards in their evaluation of a wine and therefore, the indications found in this paper could be generalizable.

The approach followed in the paper involved the use of the causal reasoning applied to sensory analysis; the obtained results can be of potential interest for those producers who want to score high, at least, in the Altroconsumo's Guida Vini. If a producer of red wines wants to obtain a good rating, at least in the Altroconsumo guide, she/he should mainly pay attention to the amount of sulfites, the roundness, the gustatory harmony, the structure, the aftertaste cleanness, and the intensity of the color of a wine. If she/he produces white wines, in order to obtain a good rating, at least in this guide, she/he should monitor the amount of sulfites, the roundness, the gustatory harmony, the aftertaste cleanness, the intensity of the color and the bouquet of a wine and how pleasant the aspect of the wine is.

The graphs and models were obtained considering data from seven years; it will be of interest to verify their endurance when new data, from future years, will be available.

Acknowledgements

The authors wish to thank S. Pagani, oenologist at Valentina Cubi winery (Fumane, Italy) and L. Odello, oenologist and president of Centro Studi Assaggiatori (Brescia, Italy), for their valuable indications and discussions, Altroconsumo for the permission to use their data and M. Sandri, the associate editor M. Meyners and the two anonymous reviewers for their suggestions and comments which improved the manuscript.

This study was conducted within the project of sensory analysis developed by the Data Methods and Systems Statistical Laboratory (DMS StatLab) of the University of Brescia.

References

Altroconsumo (2006–2013). Guida Vini. Milano: Altroconsumo Edizioni.

Bera, A. K., & Jarque, C. M. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255–259.

Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley.

Brentari, E., Carpita, M., & Vezzoli, M. (2012). CRAGGING: A novel approach for inspecting Italian wine quality. *Proceeding AGROSTAT 2012* (pp. 343–350).

Brentari, E., & Levaggi, R. (2014). The hedonic price for Italian Red Wine: Do chemical and sensory characteristics matter? *Agribusiness, 30*(4), 385–397.

Brentari, E., Levaggi, R., & Zuccolotto, P. (2011). Pricing strategies for Italian Red Wine. *Food Quality and Preference, 22*(8), 725–732.

Brentari, E., & Vezzoli, M. (2015). Evaluating Italian wine quality by cross-aggregating multiple regression trees. *Proceeding of the 143-rd Joint EAAE/AAEA Seminar on Consumer Behavior in a Changing World: Food, Culture and Society*. .

Brentari, E., & Zuccolotto, P. (2011). The impact of chemical and sensory characteristics on the market price of Italian red wines. *Electronic Journal of Applied Statistical Analysis, 4*(2), 265–276.

Chickering, D. M. (2002). Learning equivalence classes of Bayesian-network structures. *Journal of Machine Learning Research, 2*, 445–498.

Goode, J., & Harrop, S. (2011). *Authentic wine: Toward natural and sustainable winemaking*. University of California Press.

Harris, N., & Drton, M. (2013). PC algorithm for nonparanormal graphical models. *Journal of Machine Learning Research, 14*, 3365–3383.

Kalisch, M., & Bühlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research, 8*, 613–636.

Kalisch, M., & Bühlmann, P. (2014). Causal Structure Learning and inference: A selective review. *Quality Technology & Quantitative Management, 11*(1), 3–21.

Kalisch, M., Fellinghauer, B. A. G., Grill, E., Maathuis, M. H., Mansmann, U., Bühlmann, P., & Stucki, G. (2010). Understanding human functioning using graphical models. *Medical Research Methodology, 10*(14), 1–10.

Lawless, H. T. (2013). *Quantitative sensory analysis*. Wiley-Blackwell.

Oates, C.J., Kasza, J., Simpson, J.A. & Forbes, A.B. (in press). Repair of partly misspecified causal diagrams. Epidemiology. Accepted for publication.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann Publishers.

Pearl, J. (2009). *Causality: models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Phan, V. A., Ramaekers, M. G., Bolhuis, D. P., Garczarek, U., van Boekel, M. A. J. S., & Dekker, M. (2012). On the use of Bayesian networks to combine raw data from related studies on sensory satiation. *Food Quality and Preference, 26*, 119–127.

Ramsey, J.D. (2014). A scalable conditional independence test for nonlinear, non-gaussian data. arxiv.org/abs/1401.5031.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310.

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research, 11*, 1643–1662.

Spirtes, P., Scheines, R., Ramsey, J., Glymour, C. (2010). The TETRAD project: Causal models and statistical data. Available from http://www.phil.cmu.edu/tetrad/current.html.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: The MIT Press.

Spirtes, P., Scheines, R., Glymour, C., Richardson, T., & Meek, C. (2004). Causal inference. In D. Kaplan (Ed.). *The Sage handbook of quantitative methodology for the social sciences* (pp. 447–477). Sage Pubblications.

Tenenhaus, M., Pagès, J., Ambroisine, L., & Guinot, C. (2005). PLS methodology to study relationships between hedonic judgements and product characteristics. *Food Quality and Preference, 16*, 315–325.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: The Guildford Press.