Volatile Compounds Selection via Quantile Correlation and Composite Quantile Correlation: A Whiting Case Study

Ibrahim Sidi Zakari1*, Assi N’guessan2, Alexandre Dehaut3, Guillaume Duflos3

1Department of Mathematics and Computer Science, Abdou Moumouni University, Niamey, Niger
2Paul Painlevé Laboratory, UMR CNRS 8524, Lille 1 University, Lille, France
3ANSES, Laboratoire de Sécurité des Aliments Département des Produits de la Pêche et de l’Aquaculture, Boulevard du Bassin Napoléon, Boulogne-Sur-Mer, France
Email: *isidizakari@refer.ne, assi.nguessan@polytech-lille.fr, alexandre.dehaut@anses.fr, guillaume.duflos@anses.fr

Abstract

The freshness and quality indices of whiting (Merlangius merlangus) influenced by a large number of chemical volatile compounds, are here analyzed in order to select the most relevant compounds as predictors for these indices. The selection process was performed by means of recent statistical variable selection methods, namely robust model-free feature screening, based on quantile correlation and composite quantile correlation. On the one hand, compounds 2-Methyl-1-butanol, 3-Methyl-1-butanol, Ethanol, Trimethylamine, 3-Methyl butanal, 2-Methyl-1-propanol, Ethylacetate, 1-Butanol and 2,3-Butanedione were identified as major predictors for the freshness index and on the other hand, compounds 3-Methyl-1-butanol, 2-Methyl-1-butanol, Ethanol, Trimethylamine, 3-Methyl butanal, 2-Methyl-1-propanol, Ethylacetate, 1-Butanol, 2,3-Butanedione, 3-Pentanol, 3-Pentanone and 2-Methyl-1-propanol were identified as major predictors for the quality index.

Keywords

Volatile Compounds, Freshness and Spoilage Indices, Quantile Correlation, Composite Quantile Correlation, Sure Independence Screening

1. Introduction

Fish freshness is a key attribute for the quality of fish, which is a highly perishable product. The fishing industry is an important contributor to many economies in the world. One of the senses used by consumers to determine the freshness of fish is the smell. Indeed, the volatilome of fish changes rapidly according to the product degree of
freshness, and that is why sensory analysis are used by consumers and industrialists to assess fish quality. Then, the key volatile compounds that contribute to this characteristic odor can be measured and used as quality indicators [1] [2] [3]. These characteristic aromatic volatile compounds are generated by different biological pathways including the lipid autoxidation, the action of spoilage organisms and autolytic enzymes.

Recently, Duflos et al. [1] studied the spoilage of whiting at five stages of ice storage by comparing the analysis of volatile compounds obtained by solid phase microextraction (SPME) coupled to the combination of gas chromatography/mass spectrometry (GC/MS) and SPME with two sensory analysis methods. Two separate steps of statistical multidimensional approaches were used to identify volatile compounds and characterize fish freshness assessed by two different indices. In the first step, control charts were used to control the daily progression of freshness and spoilage indices. The second step begins by reducing the dimension of the data set (excluding the two indices variables) to two principal components via the application of Principal Component Analysis (PCA) method.

Then, a hierarchical clustering approach and a heuristic variable selection were used for clustering the fish samples on three classes and to identify the volatile compounds that respectively characterize these classes. However, the indices (or response variables) were not directly taken into account in the later procedure.

Recently, Sidi et al. [4] applied stability selection and randomization techniques in \( L_1 \) norm penalized quantile regression on the same data set. These approaches highlighted volatile compounds that are more relevant for the evaluation of fish freshness throughout its storage, so, are assumed to influence more the fish freshness and quality.

Using penalized quantile regression approaches on whiting data set is motivated by the fact that consumers, generally faced different categories of fish freshness. The interest of quantile regression approach is its ability to provide a model for each level of quality. More details on quantile regression and penalized quantile regression can be found in the following references [5]-[12].

This paper aims at using Ma and Zhang [13] approach to select a reduced subset of volatile compounds which can be used to explain whiting spoilage during its conservation. This approach allows a robust and model-free feature screening based on quantile correlation proposed by Li et al. [14].

The lines below are organized as follow: The methodology is briefly presented in section 2 and section 3 is dedicated to the experimental framework. Finally, the results are discussed in section 4 followed by concluding remarks in section 5.

2. Methodology

This section is dedicated to the following methods: Quantile Correlation, Sure Independence Screening via Quantile Correlation, Composite Quantile Correlation and Sure Independence Screening via Composite Quantile Correlation.

2.1. Quantile Correlation

As advocated in Li et al. [14], quantile correlation is a novel measure used to examine
the linear relationship between any two random variables \( Y \) and \( X \) for a given quantile \( \tau \in (0,1) \). So,

\[
qcor_\tau(Y, X) = \frac{qcov_\tau(Y, X)}{\sqrt{\text{var}_\tau(Y) \cdot \text{var}(X)}} = \frac{E(\psi_s(Y - Q_{\tau,Y}))(X - E(X))}{\sqrt{\text{var}_\tau(Y) \cdot \text{var}(X)}}
\]

(1)

where

\[
qcov_\tau(Y, X) = \text{cov}(I(Y - Q_{\tau,Y} > 0), X) = E(\psi_s(Y - Q_{\tau,Y}))(X - E(X))
\]

(2)

\( Q_{\tau,Y} \) is the \( \tau^{th} \) conditional quantile of \( Y \) and

\[
\psi_s(s) = \tau - I(s < 0)
\]

(3)

Moreover, if \( X \) is independent of \( Y \), the \( qcor_\tau(Y, X) = 0 \); else (\( X \) and \( Y \) are correlated), the \( qcor_\tau(Y, X) \neq 0 \).

2.2. Sure Independence Screening via Quantile Correlation

Consider \( Y \) as the dependent variable and \( X = (X_1, \ldots, X_p)^T \) be the p-dimensional independent variables. Let \( \omega = (\omega_1, \ldots, \omega_p)^T \) with

\[
\omega_k = E(qcor^2(Y, X_k)), k = 1, \ldots, p
\]

(4)

Sure Independent Screening via Quantile Correlation method selects the first \( d \) independent variables with largest \( \hat{\omega}_k \); where

\[
\hat{\omega}_k = \frac{n}{n(n-1)(n-2)} \bar{\omega}_k
\]

(5)

with \( \bar{\omega}_k \) the sample estimate of \( \omega_k \).

2.3. Composite Quantile Correlation

Composite Quantile Correlation(CQC) is motivated by the fact that previous quantile correlation cannot characterize the entire relationship between \( X \) and \( Y \). So, the composite quantile correlation is defined by:

\[
cqcor(Y, X) = E(qcor_\tau(Y, X)) = \int_0^1 qcor_\tau(Y, X) \, d\tau
\]

(6)

2.4. Sure Independence Screening via Composite Quantile Correlation

The CQC screening is based on the vector \( \omega = (\omega_1, \ldots, \omega_p)^T \) with components

\[
\omega_k = |cqcor(Y, X_k)|, k = 1, \ldots, p.
\]

(7)

Sub models are selected based on decreasing values of \( \omega_k \). Furthermore, as advocated by Ma and Zhang [13], when using screening techniques, the number of selected variables is often set to be \( n - 1 \) or the integer part of \( \frac{n}{\log(n)} \).

3. Materials and Methods

The sample preparation and sensory evaluation methods are briefly presented in this.
I. S. Zakari et al.

section. More details about the full experimental procedure can be found in [1].

3.1. Sample Preparation

As advocated by Duflos et al. [1], the sample considered is based on two different catches of respectively 20 and 15 fish. These catches were stored in crushed ice at 4°C in self-draining polystyrene boxes for 7 days. Fresh crushed ice was added daily. Sensory evaluation and volatile analysis were performed on seven different fish on days 1, 2, 3, 4 and 7.

3.2. Sensory Evaluation

According to Duflos et al. [1], two methods were used for the sensory evaluation of fish. These methods lead to freshness and quality indices which represent two response variables for our selection process.

4. Results and Discussion

The empirical results of the analysis of freshness and quality indices influenced by a great number of volatile compounds are presented below.

The sample size and the number of predictors (volatile compounds) are respectively \( n = 35 \) and \( p = 55 \). So, the number of predictors is higher than the sample size.

In order to perform variables selection, screening methods are applied on whiting data set using QC-SIS package available for R software.

The tuning parameter \( d \) used to select covariates with significant effect on each response variable can be set to \( n - 1 = 34 \) or the integer part of \( \frac{n}{\log(n)} = 9 \).

The results for \( d = 9 \) are presented in Table 1 and Table 2.

Furthermore, Figure 1 and Figure 2 display the Pearson correlation matrix through bivariate scatter plots for each index with corresponding selected compounds. These figures have been made using Performance Analytics package available for R software.

Table 1. Selected volatile compounds ranked by decreasing weights for freshness index

<table>
<thead>
<tr>
<th>Selected Volatile Compounds</th>
<th>QC SIS</th>
<th>CQC SIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Methyl-1-butanol</td>
<td>0.314</td>
<td>0.515</td>
</tr>
<tr>
<td>3-Methyl-1-butanol</td>
<td>0.292</td>
<td>0.492</td>
</tr>
<tr>
<td>Ethanol</td>
<td>0.264</td>
<td>0.478</td>
</tr>
<tr>
<td>Trimethylamine</td>
<td>0.140</td>
<td>0.327</td>
</tr>
<tr>
<td>3-Methyl butanal</td>
<td>0.132</td>
<td>0.322</td>
</tr>
<tr>
<td>2-Methyl-1-propanol</td>
<td>0.119</td>
<td>0.305</td>
</tr>
<tr>
<td>Ethylacetate</td>
<td>0.106</td>
<td>0.302</td>
</tr>
<tr>
<td>1-Butanol</td>
<td>0.102</td>
<td>0.291</td>
</tr>
<tr>
<td>2,3-Butanedione</td>
<td>0.099</td>
<td>0.278</td>
</tr>
</tbody>
</table>
Table 2. Selected volatile compounds ranked by decreasing weights for quality index.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>QC SIS Weights</th>
<th>Compounds</th>
<th>CQC SIS Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Methyl-1-butanol</td>
<td>0.302</td>
<td>3-Methyl-1-butanol</td>
<td>0.501</td>
</tr>
<tr>
<td>2-Methyl-1-butanol</td>
<td>0.288</td>
<td>2-Methyl-1-butanol</td>
<td>0.499</td>
</tr>
<tr>
<td>Ethanol</td>
<td>0.224</td>
<td>Ethanol</td>
<td>0.450</td>
</tr>
<tr>
<td>3-Methyl butanal</td>
<td>0.158</td>
<td>3-Methyl butanal</td>
<td>0.363</td>
</tr>
<tr>
<td>3-Hydroxy-2-butanone</td>
<td>0.122</td>
<td>1-Butanol</td>
<td>0.327</td>
</tr>
<tr>
<td>1-Butanol</td>
<td>0.119</td>
<td>3-Hydroxy-2-butanone</td>
<td>0.286</td>
</tr>
<tr>
<td>2,3-Butanediene</td>
<td>0.097</td>
<td>3-Pentanol</td>
<td>0.272</td>
</tr>
<tr>
<td>3-Pentanol</td>
<td>0.094</td>
<td>2,3-Butanediene</td>
<td>0.271</td>
</tr>
<tr>
<td>3 Pentanone</td>
<td>0.090</td>
<td>2-Methyl-1-propanol</td>
<td>0.245</td>
</tr>
</tbody>
</table>

Figure 1. Correlation matrix chart for freshness index and nine(9) related compounds: On top; the (absolute) value of the correlation with significance levels. The distribution of each variable is represented on the diagonal and at bottom, the bivariate scatterplots, with a fitted line. Components of the vector are respectively tagged with symbols corresponding to the associated p-values: “***” (p-value ≤ 0.001), “**” (p-value ≤ 0.01), “*” (p-value ≤ 0.05), “.” (p-value ≤ 0.1), “ ” (p-value ≤ 1).
4.1. Results for Freshness Index

According to Table 1, Quantile Correlation and Composite Quantile Correlation Sure Independent Screening methods select the same subset of volatile compounds for freshness index.

These compounds have been previously identified as spoilage markers.

For example, the compounds Ethanol, 3-Methyl-1-butanol, 2,3-Butanedione, 2-Methyl-1-butanol, 3-Methyl butanal, 2-Methyl-1-propanol and Ethylacetate are identified as correlated to the second principal component axis in Duflos et al. [1].

The previous seven compounds were included in the eight compounds (with limonene) that characterized category 2/3 (intermediate category between freshness and spoilage) in Duflos et al. [1].
Moreover, compounds Trimethylamine and 1-Butanol were not identified as correlated to the first principal component axis in Duflos et al. [1].

Finally, considering Freshness index, only Trimethylamine, 1-Butanol and 2-Methyl-1-butanol were not selected by \( L_1 \) QR and other randomization approaches highlighted in Sidi et al. [4].

### 4.2. Results for Quality Index

According to Table 2, Quantile Correlation and Composite Quantile Correlation Sure Independent Screening methods do not select the same subset of volatile compounds for quality index.

The compounds Ethanol, 3-Methyl-1-butanol, 2,3-Butanedione, 2-Methyl-1-butanol, 3-Methyl butanal and 2-Methyl-1-propanol are identified as correlated to the second principal component axis in Duflos et al. [1].

The previous six compounds were included in the eight compounds (with limonene) that characterized category 2/3 (intermediate category between freshness and spoilage) in Duflos et al. [1].

The compounds 1-Butanol, 3-Pentanol, 3-Pentanone and 3-Hydroxy-2-butanol are not identified as correlated to the first principal component axis in Duflos et al. [1].

For the quality index, only Ethanol, 3-Methyl butanal, 3-Methyl-1-butanol and 2-Methyl-1-butanol were selected in Sidi et al. [4].

Choosing \( d = n - 1 = 34 \) includes freshness markers like Propanal, Hexanal, 1-Penten-3-ol, Pentanal, 2,3-Pentanedione, 1-Penten-3-one, Heptanal, (E)-2-Pentenal, (Z)-2-Penten-1-ol, 1-Pentanol, Butanal, Octanal, 1-hexanol and 4,4-Dimethyl-1,3-dioxane.

### 5. Concluding Remarks

Sure Independence Screening via Quantile Correlation and Composite Quantile Correlation methods highlighted relevant volatile compounds influencing freshness and quality indices during whiting conservation.

The selected compounds include Ethanol, 3-Methyl-1-butanol, 2,3-Butanedione, 2-Methyl-1-butanol, 3-Methyl butanal, 2-Methyl-1-propanol and Ethylacetate, previously identified as spoilage markers.

For future investigation on whiting data, it will be very interesting to explore the following issues:

2) Selection of groups of highly correlated compounds [8]
3) Quantile regression models and inference processes based on [15] and [16].

### Acknowledgements

The authors wish to thank Pr Abdallah Mkhadri (Cadi Ayyad University of Morocco), François Leduc (ANSES), The Nord-Pas-de-Calais Regional Council (France) and ANSES (France). This support is greatly appreciated.
References


Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/
Or contact ojs@scirp.org