Article

The Constrained Median: A Way to Incorporate Side Information in the Assessment of Food Samples

Marc Sader [1], Raúl Pérez-Fernández [1,2,*] , Lotta Kuuliala [1,3], Frank Devlieghere [3] and Bernard De Baets [1]

[1]   KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, 9000 Gent, Belgium; marc.sader@ugent.be (M.S.); lotta.kuuliala@UGent.be (L.K.); bernard.debaets@ugent.be (B.D.B.)
[2]   Department of Statistics and O.R. and Mathematics Didactics, University of Oviedo, 33003 Oviedo, Spain
[3]   FMFP—Research Unit Food Microbiology and Food Preservation, Department of Food Technology, Safety and Health, Ghent University, 9000 Ghent, Belgium; frank.devlieghere@UGent.be
*   Correspondence: perezfernandez@uniovi.es

check for updates

Abstract: A classical problem in the field of food science concerns the consensus evaluation of food samples. Typically, several panelists are asked to provide scores describing the perceived quality of the samples, and subsequently, the overall (consensus) scores are determined. Unfortunately, gathering a large number of panelists is a challenging and very expensive way of collecting information. Interestingly, side information about the samples is often available. This paper describes a method that exploits such information with the aim of improving the assessment of the quality of multiple samples. The proposed method is illustrated by discussing an experiment on raw Atlantic salmon (*Salmo salar*), where the evolution of the overall score of each salmon sample is studied. The influence of incorporating knowledge of storage days, results of a clustering analysis, and information from additionally performed sensory evaluation tests is discussed. We provide guidelines for incorporating different types of information and discuss their benefits and potential risks.

1. Introduction

One of the most traditional ways of determining the quality of food is by performing sensory evaluation tests, such as asking panelists to provide absolute evaluations expressing the overall quality of a food sample [1–3]. Typically, scales with three to nine points are used to express the degree of spoilage/freshness [4–7], the appearance [8], or the flavor [9] of a food sample. In acceptance tests, the nine-point hedonic scale [10,11] is often used, where the points on the scale represent ordered categorical labels ranging from "dislike extremely" as a score of "1" to "like extremely" as a score of "9". These scores can be used to determine a consensus score of the overall quality of a food sample. Unfortunately, the availability of panelists is oftentimes limited. Generally, a small panel (usually less than 30 individuals) could hardly be a good representation of a target market [12], and the resulting consensus score might not be a good representation of the overall quality of a sample. Thus, incorporating other sources of information may help improve the assessment of the quality of a sample.

In most cases, side information about the samples is available. Typical examples of such information include: the storage days of the food sample, additionally performed chemical analyses, or additionally performed sensory evaluation tests, such as ranking, discrimination, and threshold tests. These types of information usually are of a relative nature and hint at some relations between the scores assigned to the food samples. Several studies have shown that learning with side information can be effective in machine learning [13,14]. In this paper, we develop a method that combines such

side information about several food samples and scores provided by panelists to find their overall score jointly.

This paper is organized as follows. In Section 2, the method to combine scores and other types of information is described and the experimental setup is provided. Section 2.1 describes the median, the most commonly used measure of central tendency of scores. Section 2.2 provides a non-exhaustive list of some potential real-life situations where side information about the samples could be available, and Section 2.3 describes an efficient method to combine scores and other types of information to assign an overall score to multiple samples. Section 3 illustrates the method by presenting an experiment on raw Atlantic salmon (*Salmo salar*). The results are shown and discussed in Section 4. The paper ends with some conclusions in Section 5.

2. Theory

2.1. The Median

We start with a description of the theory that provides the main building blocks of our approach. Denote by x_j the jth sample in a set $\mathcal{X} = \{x_1, \ldots, x_n\}$ of n food samples. Consider the setting where each of m panelists has assigned a score on a k-point scale to a given food sample, and the goal is to agree on the consensus score that should be assigned.

Since it will be the scale used in the experimental setup, the five-point scale illustrated in Figure 1 is considered. However, it is important to highlight that the results of this paper are straightforwardly extended to any other k-point scale. Scales with an odd number of points are typically preferred, allowing for a neutral response in case of bipolar scales [15].
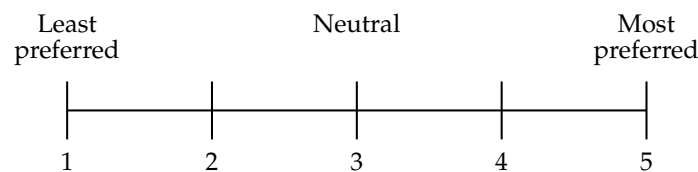
Figure 1. Example of a five-point scale, where the extreme scores of "1" and "5" represent least and most preferred, respectively, and the intermediate score of "3" represents a neutral preference.

Notably, the median is the most commonly used measure of the central tendency of scores found in studies on food quality [16–20]. This measure can be understood as the score that separates the lower half from the upper half of the scores provided by the panelists for a given sample [21]. As will be explained below, this procedure is equivalent to assigning the score that minimizes the sum of absolute differences.

Typically, gathering together several panelists is a challenging, time consuming, and expensive exercise. Therefore, it is common to provide the panelists with multiple food samples during the same experiment. In general, the scores assigned to each sample are considered to be independent, and the assessment of a consensus score to each sample is assumed to be an independent task. Note that it is often difficult to gather the same number of panelists for different experiments.

Consider the problem setting where several panelists provide scores for each of the n food samples. For example, consider a simple setting where nine panelists each assign a score to a given food sample on the five-point scale fixed in Figure 1. The scores provided by the panelists are $4, 1, 2, 1, 4, 3, 3, 4, 3$ and are represented in increasing order as $1, 1, 2, 3, 3, 3, 4, 4, 4$. Denote by m_j the number of scores assigned to the jth sample and by $\mathbf{s}^i(j)$ the ith lowest score assigned to sample x_j, where $j \in \{1, \ldots, n\}$ and $i \in \{1, \ldots, m_j\}$.

The goal is to agree on the consensus score that should be assigned to each sample in $\mathcal{X}$. Obtaining the median score for each sample is equivalent to directly computing the vector $\mathbf{s}^*$, as follows:

$$\mathbf{s}^* = \underset{\mathbf{s} \in \{1,\dots,5\}^n}{\arg\min} \sum_{j=1}^{n} \sum_{i=1}^{m_j} \left| \mathbf{s}(j) - \mathbf{s}^i(j) \right|, \tag{1}$$

where $\mathbf{s}^*(j)$ is referred to as the median of the jth sample. Note that in case m_j is odd for all j, the median (and thus, the minimizer of Equation (1)) is unique. In case m_j is even for at least one j, there can be multiple medians (and thus, multiple minimizers of Equation (1)).

2.2. Information about Samples

In most cases, side information about the different samples could be available. We provide a non-exhaustive list of some potential real-life situations hereafter.

2.2.1. Knowledge of Storage Days

Researchers are often interested in studying the temporal evolution of the attributes of perishable food. This is typically done by asking panelists to provide a score to a food sample that comes from different time spans of the shelf life of the same food product. In general, it is expected that food samples should be less fresh as time goes by. Thus, it is expected that the less fresh the sample is, the lower the score should be. One example is evaluating the freshness or tenderness of meats, where the score may only decrease with time [22].

Consider the setting where samples coming from the same food product are indexed in increasing order of storage days. Thus, the potential consensus scores should naturally reduce to those that satisfy the following constraints:

$$\mathbf{s}(1) \geq \cdots \geq \mathbf{s}(n).$$

Note that the overall trend of the scores should be decreasing; however, different decreasing patterns are possible. One possible pattern of scores is illustrated in Figure 2. Note that multiple consecutive samples could be assigned the same score, as illustrated in Figure 2 for Points 0, 1, and 2. Typically, this occurs when the number of points (k) on the scale is small.
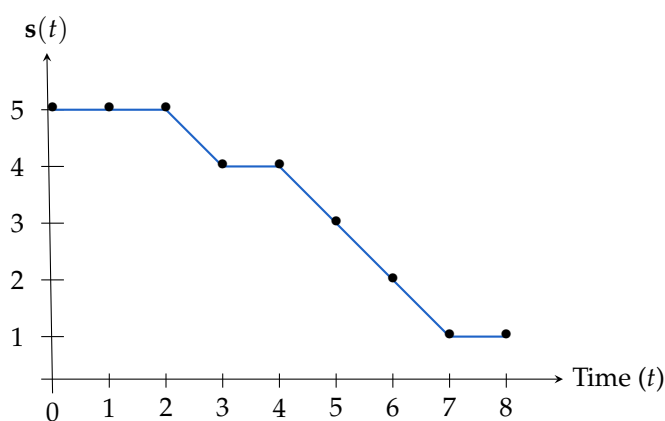

Figure 2. Example of scores describing the freshness of food decreasing over time.

In studies on the acceptability of beverages, the evolution of certain attributes of beverages is of interest. A common method used in this situation is the time-intensity (TI) method [21] (see Chapter 8). Typically, a beverage sample is first swallowed, and then, an attribute is evaluated by assigning a score at different time spans. For several types of beverages, such as beer, wine, and soda, it is expected that a beverage sample should have an increasing acceptance at first, eventually decreasing afterwards.

Typical examples include the evaluation of the astringency and flavor of beer and wine, where these attributes increase in intensity at first and eventually decrease with time [23,24].

Consider the setting where a beverage sample is evaluated by panelists over a period of time. Thus, the potential consensus scores should naturally reduce to those that satisfy the following constraints:

$$\mathbf{s}(1) \leq \cdots \leq \mathbf{s}(a) \text{ and } \mathbf{s}(a) \geq \cdots \geq \mathbf{s}(n),$$

for some $a \in \{1, \ldots, n\}$.

This means that there should be a unimodal pattern. One possible pattern of scores is illustrated in Figure 3. Note that if one considers a short duration, say $t \in \{0, \ldots, 3\}$, then the overall trend is only increasing. However, if one considers a long duration at a later time, say $t \in \{2, \ldots, 8\}$, then the overall trend is only decreasing.
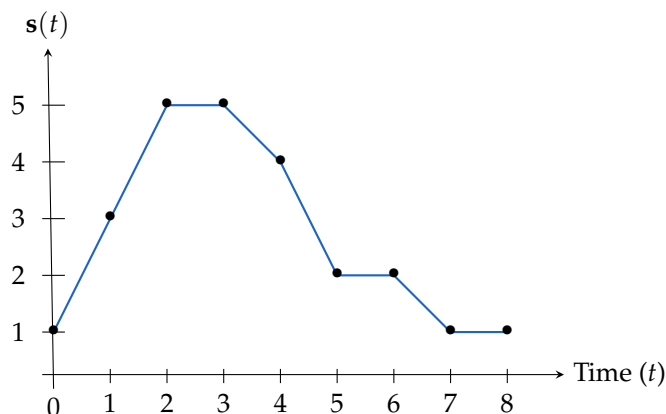

Figure 3. Example of scores describing the intensity of wine flavor showing a unimodal pattern.

2.2.2. Results of a Clustering Analysis

In many studies, food samples are stored at different (temperature and atmospheric) conditions or represent the same food product, but originating from a different initial batch, manufacturer, or season. In addition, the initial contamination of the food (i.e., initial microbial load) plays a big role in the spoilage rate of every sample, and thus, the decreasing pattern of the scores might not always hold. Thus, the storage days could not be used as the only tool to compare these samples. For instance, it is not always the case that samples that have been stored at different conditions for the same duration of time will be similar. Similarly, it is not always the case that a sample is always preferred over another sample that is stored at different conditions and has been stored for longer.

It is well established that microbial growth is the most important cause of food spoilage, particularly in meats [25], producing volatile organic compounds (VOCs) and, subsequently, off-odors and off-flavors. These odors and flavors result in an olfactory impact that is associated with the spoilage of food. Therefore, the relation between the VOC profiles and the quality of food has caught the attention of many researchers in food science. Several studies have successfully used the composition of the VOC profiles to evaluate the quality of food, such as seafood [26,27] and meat [28].

To establish a relation between the VOC profiles of the samples and their resulting consensus scores, clustering analysis, a method for merging similar groups of samples based on the similarity of their VOC profile, can be used. In general, it is expected that samples clustered together should be quite similar, and thus, their scores should not be very different. Therefore, the absolute difference of the scores of these samples should not exceed a certain threshold. Note that we prefer not to impose that samples in the same cluster should have strictly the same scores because this might be too restrictive. However, this is still a possibility, as will be further explained below.

The considered setting may naturally reduce the potential consensus scores to those that satisfy the following constraints:

$$|\mathbf{s}(i) - \mathbf{s}(j)| \leq \epsilon, \text{ for any } i, j \in I_b,$$

where I_b is the set of indices corresponding to the bth cluster and ϵ is a threshold on the absolute difference of the scores of samples in the same cluster. Note that the value of ϵ may depend on several considerations, such as the distance between clusters (different clustering analysis tools result in different distances between clusters) or the number of points (k) on the scale used for scoring. The special case where $\epsilon = 0$ amounts to restricting with equality constraints only. For instance, consider the case where samples $\{x_1, x_2\}$ are found in one cluster and samples $\{x_3, x_4, x_5\}$ are found in another cluster. It is expected that the absolute difference of the scores of every pair of samples in each cluster should be smaller than or equal to one. This process is illustrated in Figure 4. It can be seen that the absolute difference of scores for each couple of samples in the same cluster is smaller than or equal to one, thus satisfying the constraints.
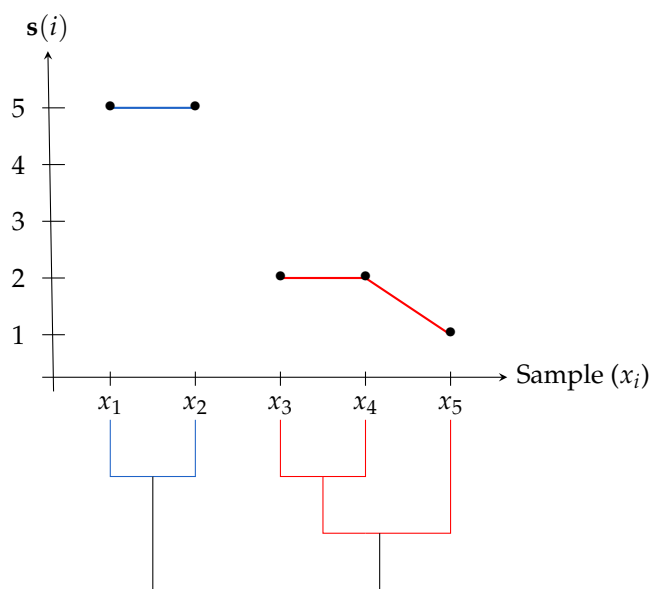

Figure 4. Example of clustering analysis, where samples $\{x_1, x_2\}$ are in the first (blue) cluster and samples $\{x_3, x_4, x_5\}$ are in the second (red) cluster. The scores are consistent with the clusters or, equivalently, satisfy the constraints.

2.2.3. Information from other Sensory Evaluation Tests

Ranking test

Recently, researchers have been adopting scoring methods for determining the quality of food samples along with ranking methods to order the samples according to their quality [29–32]. Ranking tests involve several panelists providing rankings (with ties) on samples. Typically, these rankings are aggregated to obtain a consensus ranking that describes an underlying order of the samples; thus, it is expected that the scores agree with this consensus ranking of the samples. For example, rankings have been previously used to study the desirability of different meats [33,34]. In this setting, these rankings can give useful information as a reference for the relative desirability of meats.

In general, it is expected that samples ranked higher are preferred over samples ranked lower; thus, it is expected that the higher the sample is ranked, the greater the score should be. Note that we do not impose that a sample ranked higher than another sample should have a strictly greater score; instead, we allow their scores to be equal as well. This is due to the fact that the considered scale is typically not rich enough for allowing to distinguish between similar samples. In case two samples are tied, it is expected that their scores should be similar. Note that the situation where samples are tied is

similar to that where samples are in the same cluster. For simplicity, the special case where $\epsilon = 0$ is considered. The considered setting may naturally reduce the potential consensus scores to those that satisfy the following constraints:

$$x_i \precsim x_j \Rightarrow \mathbf{s}(i) \leq \mathbf{s}(j)\,, \text{ for any } i, j \in \{1, \ldots, n\}\,.$$

For instance, consider the ranking with ties $x_1 \prec x_2 \sim x_3 \prec x_4$. It is expected that x_1 should be assigned a score smaller than or equal to that of sample x_2, which should be assigned a score equal to that of sample x_3, which should be assigned a score smaller than or equal to that of sample x_4. More formally, the resulting constraints are $\mathbf{s}(1) \leq \mathbf{s}(2) = \mathbf{s}(3) \leq \mathbf{s}(4)$. This process is illustrated in Figure 5.
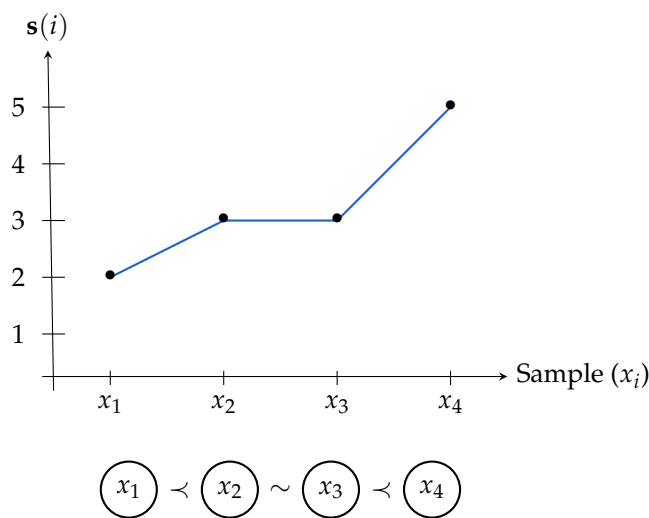

Figure 5. Example of scores describing the ranking $x_1 \prec x_2 \sim x_3 \prec x_4$.

Discrimination test

Many discrimination tests can be seen as a special case of a ranking test. For instance, in an A-notA test, panelists are provided with one sample and are asked whether or not it is similar to a reference sample A [35]. Based on the responses of the panelists, if there is no significant difference between the samples, then it is expected that they should be assigned a similar score. Therefore, the absolute difference of the scores of these samples should not exceed a certain threshold ϵ.

Another instance is a duo-trio test, where panelists are provided with two samples and a reference sample that is identical to one of the two samples and are asked to match one of the two samples to the reference sample [36] (see Chapter 4). It is expected that the reference sample and the sample identical to it should be scored equally. Moreover, if a large number of panelists are not able to distinguish the identical samples from the third sample, then it is expected that this third sample should be assigned a score similar to that of the identical samples. Therefore, the absolute difference of the scores of the non-identical samples should not exceed a certain threshold ϵ. This process is illustrated in Figure 6.

Another instance is a two-out-of-five test, where panelists are given five samples and are asked to distinguish two identical samples from the other three samples [37]. It is expected that the identical samples should be scored equally. Moreover, if a large number of panelists are not able to distinguish the identical samples from the other three, then it is expected that there is no significant difference among the five samples and that all the samples should be assigned similar scores. Therefore, the absolute difference of the scores of these samples should not exceed a certain threshold ϵ.
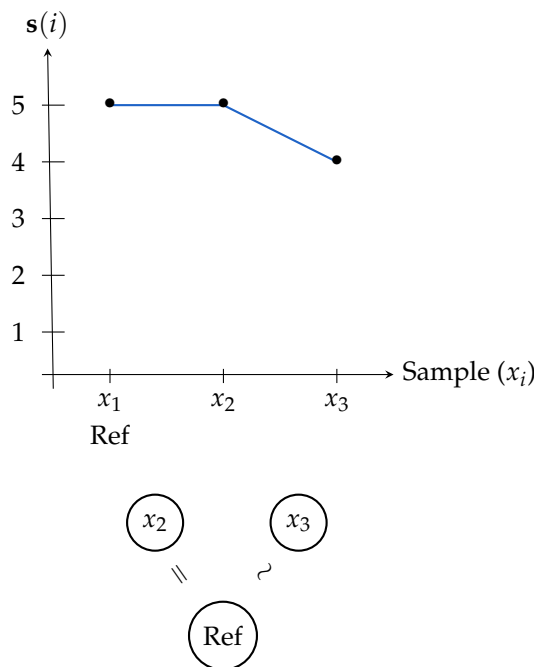
Figure 6. Example of scores describing that there is no significant difference between x_3 and the reference sample x_1 (equivalently, its identical to sample x_2) and where a threshold $\epsilon = 1$ is considered.

Threshold test

In threshold tests, panelists are asked to determine a threshold of noticing a certain stimulus [21] (see Chapter 6). Different versions of the threshold test have been proposed, the differential threshold test and the absolute threshold test being the most prominent examples. In the former, the aim is to determine the threshold at which an increase in a noticed stimulus can be perceived, whereas in the latter, the aim is to determine the lowest threshold at which a stimulus can be noticed. Note that the case in which there is a decrease in stimulus can also be considered. One example is determining the (consumer) rejection of chocolate bitterness [38].

In differential threshold tests, it is expected that the sample where an increase in stimulus is not noticed should have a quite similar score to the previous sample that has one increment less of the stimulus, and thus, their scores should not be very different. Therefore, the absolute difference of the scores of these samples should not exceed a certain threshold ϵ. However, it is expected that the samples where an increase in stimulus is noticed should have a score greater than or equal to the score of the previous sample. Therefore, the absolute difference of the scores of these samples should be greater than or equal to this threshold.

Note that samples arranged in increasing (or decreasing) order of stimulus should have scores that are either increasing $\mathbf{s}(i) \leq \mathbf{s}(i+1)$ or decreasing $\mathbf{s}(i) \geq \mathbf{s}(i+1)$ for any $i \in \{1, \ldots, n\}$. The considered setting may naturally reduce the potential consensus scores to those that satisfy the following additional constraints:

$$|\mathbf{s}(i) - \mathbf{s}(i+1)| \leq \epsilon, \quad \text{for any } i \in \xi_1,$$
$$|\mathbf{s}(i) - \mathbf{s}(i+1)| \geq \epsilon, \quad \text{for any } i \in \xi_2,$$

where ξ_1 is the set of indices corresponding to the samples where a stimulus is not noticed, ξ_2 is the set of indices corresponding to the samples where a stimulus is noticed, and ϵ is a threshold on the absolute difference of the scores of consecutive samples. For instance, consider that the stimulus is first noticed at sample x_4. It is expected that the absolute differences of the scores of consecutive samples x_1 and x_2 and samples x_2 and x_3 should be smaller than or equal to $\epsilon = 1$ and that the absolute difference

of the scores of samples x_3 and x_4 should be greater than or equal to $\epsilon = 1$. Similarly, given that the stimulus is noticed a second time at sample x_6, it is expected that the absolute difference of the scores of samples x_4 and x_5 should be smaller than or equal to one and that the absolute difference of the scores of samples x_5 and x_6 should be greater than or equal to one. This process is illustrated in Figure 7.
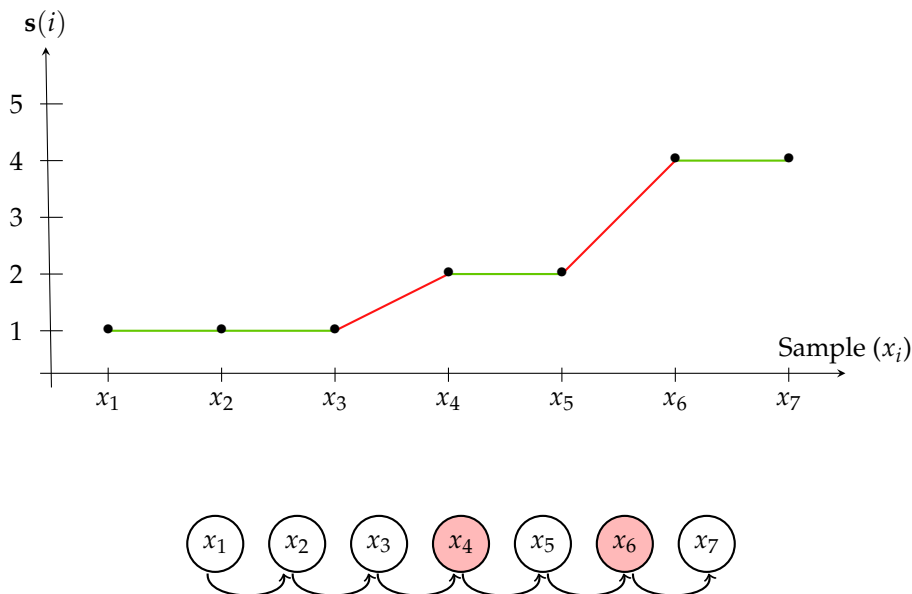

Figure 7. Example of scores in a differential detection test describing no detection of a stimulus (in green) and the detection of a stimulus (in red) at samples x_4 and x_6 and where a threshold $\epsilon = 1$ is considered.

In absolute threshold tests, samples with a stimulus are only compared to a reference sample. Thus, if an increase in stimulus is not noticed in a sample, then it is expected that this sample should be assigned a score similar to that of the reference sample. Therefore, the absolute difference of their scores should not exceed a certain threshold ϵ. However, it is expected that the samples where an increase in stimulus is noticed should be assigned a different score than that assigned to the reference sample. Therefore, the absolute difference of their scores should be greater than or equal to this threshold.

Note that the scores should be either increasing $\mathbf{s}(i) \leq \mathbf{s}(i+1)$ or decreasing $\mathbf{s}(i) \geq \mathbf{s}(i+1)$ for $i \in \{1, \ldots, n\}$. We consider the first sample ($i = 1$) to be the reference sample. The considered setting may naturally reduce the potential consensus scores to those that satisfy the following additional constraints:

$$|\mathbf{s}(1) - \mathbf{s}(i)| \leq \epsilon, \quad \text{for any } i \in \{1, \ldots, c-1\},$$
$$|\mathbf{s}(1) - \mathbf{s}(i)| \geq \epsilon, \quad \text{for any } i \in \{c, \ldots, n\},$$

where c is the sample at which an increase in stimulus is noticed and ϵ is a threshold on the absolute difference between the scores of the samples and that of the reference sample.

For instance, consider that a stimulus is first noticed at sample x_4. It is expected that the absolute difference of the scores of samples x_1 and x_4 should be greater than or equal to one. Now, consider that sample x_4, where the first stimulus is noticed, is the new reference and that a second stimulus is noticed at sample x_6. It is expected that the absolute difference of the scores of samples x_4 and x_6 should be greater than or equal to one. This process is illustrated in Figure 8.
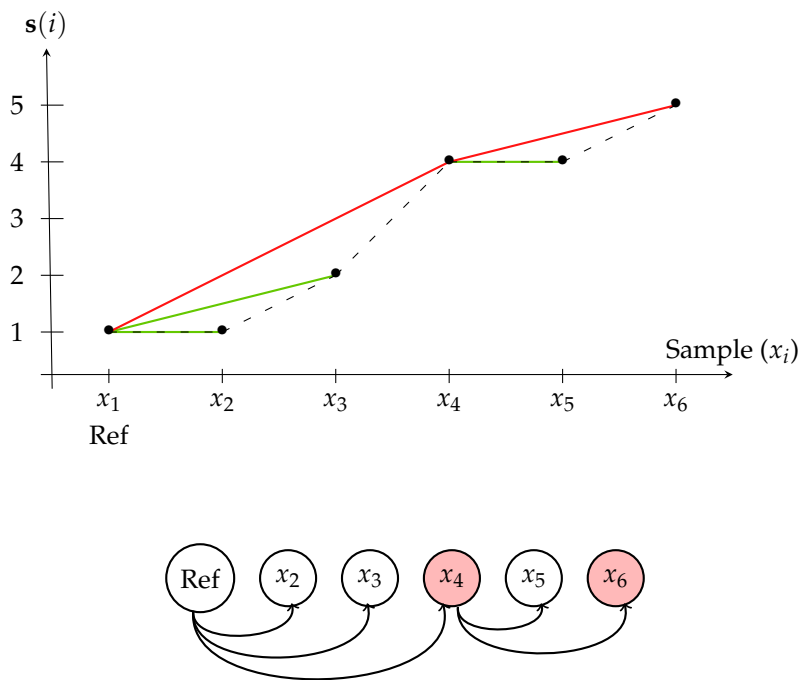
Figure 8. Example of scores that are increasing (dashed line) in an absolute detection test describing no detection of a stimulus (in green) and the detection of a first stimulus (in red) at sample x_4 when compared to the reference sample x_1 and a second stimulus (in red) at sample x_6 when compared to the new reference sample x_4 and where a threshold $\epsilon = 1$ is considered.

2.3. The Constrained Median

As we have previously discussed, the considered settings may naturally reduce the set of potential consensus scores from $\{1, \ldots, 5\}^n$ to a non-empty subset $\mathcal{S} \subseteq \{1, \ldots, 5\}^n$. We conjecture that, in most real-life situations, it seems natural for $\mathcal{S}$ to result from the conjunction of some (in)equality constraints on the components of **s**. However, this condition should not be a requirement if it does not comply with the characteristics of the considered problem.

Thus, the consensus scores should be the ones given by the vector that minimizes the sum of distances while satisfying the constraints of $\mathcal{S}$. Therefore, the problem defined by Equation (1) can now be restricted to $\mathbf{s} \in \mathcal{S}$, as follows:

$$\mathbf{s}^* = \underset{\mathbf{s} \in \mathcal{S}}{\arg\min} \sum_{j=1}^{n} \sum_{i=1}^{m_j} \left| \mathbf{s}(j) - \mathbf{s}^i(j) \right|, \tag{2}$$

where $\mathbf{s}^*$ is referred to as a constrained median. This concept is illustrated in the following example.

Example 1. *Consider a simple setting where nine panelists each assign a score to two given food samples on the five-point scale fixed in Figure 1. The scores assigned to each sample are represented in increasing order in Table 1.*

From Table 1, it can be clearly seen that the median for sample x_1 is three and the median for sample x_2 is four (i.e., the score in the middle, in this case for $i = 5$). Analogously, we consider the problem defined by Equation (1), and we compute the sum of distances between the scores provided by the panelists and every possible vector of scores (in general, for n samples and k scores, the number of possible vectors of scores is k^n). The results are illustrated in Table 2. We see that the minimizer (thus the median) is the vector of scores $(3, 4)$. Note that, as expected, this vector coincides with the result of computing the median for each of the different samples separately.

In the setting where it is known that the first sample is fresher than the second sample (they are samples from different time spans of the shelf life of the same food), it is expected that the score of the first sample should be greater than or equal to the score of the second sample. Finding a solution by simply looking at Table 1 is not an easy task. Thus, the problem defined by Equation (2) is considered. The set of constraints $\mathcal{S}$ is formed by the vectors of scores in which the first sample is assigned a score greater than or equal to the score of the second sample. Such vectors are highlighted in gray in Table 2. It can be seen that the minimizer (thus the constrained median) is the vector of scores $(4, 4)$. A conclusion is reached that the score assigned to sample x_1 should be greater than that originally assigned.

Table 1. The scores assigned by the panelists in Example 1.

The ith Lowest Score	1	2	3	4	5	6	7	8	9
Score of sample x_1 ($\mathbf{s}^i(1)$)	1	1	2	3	3	3	4	4	4
Score of sample x_2 ($\mathbf{s}^i(2)$)	2	3	4	4	4	5	5	5	5

Table 2. Sum of distances between the scores provided by the panelists and all possible vectors of scores. The minimizers are shown in bold, and the vectors of scores in which the first sample is assigned a score greater than or equal to the score of the second sample are highlighted in gray.

s	$\sum_{j=1}^{2} \sum_{i=1}^{9} \left\| \mathbf{s}(j) - \mathbf{s}^i(j) \right\|$	s	$\sum_{j=1}^{2} \sum_{i=1}^{9} \left\| \mathbf{s}(j) - \mathbf{s}^i(j) \right\|$
$(1,1)$	44	$(4,1)$	39
$(1,2)$	35	$(4,2)$	30
$(1,3)$	28	$(4,3)$	23
$(1,4)$	23	$(4,4)$	**18**
$(1,5)$	24	$(4,5)$	19
$(2,1)$	39	$(5,1)$	48
$(2,2)$	30	$(5,2)$	39
$(2,3)$	23	$(5,3)$	32
$(2,4)$	18	$(5,4)$	27
$(2,5)$	19	$(5,5)$	28
$(3,1)$	36		
$(3,2)$	27		
$(3,3)$	20		
$(3,4)$	**15**		
$(3,5)$	16		

Note that the difference between the number of panelists for each experiment can be extremely large in some instances. For instance, a very small number of panelists is gathered for one experiment, and a larger number of panelists is gathered for another experiment. Such a scenario can be approached from two different points of view: (a) each panelist is represented by one evaluation or (b) each sample is represented by one evaluation. The former approach is analogous to the problem defined by Equation (2), whereas the latter approach can be formalized as follows:

$$\mathbf{s}^* = \arg\min_{\mathbf{s} \in \mathcal{S}} \sum_{j=1}^{n} \sum_{i=1}^{m_j} \frac{\left| \mathbf{s}(j) - \mathbf{s}^i(j) \right|}{m_j}, \tag{3}$$

where the evaluations are averaged based on the number of panelists m_j that provide scores for the jth sample. It must be noted that both approaches are equivalent if all m_j's are equal. Moreover, both problems are also equivalent if there are no constraints (i.e., solving the problem defined by Equation (1) resulting in the median).

Example 2. *Consider a simple setting where one panelist assigns a score to a given food sample and nine panelists each assign a score to a second given food sample on the five-point scale fixed in Figure 1. The scores are represented in increasing order in Table 3.*

To determine the vector of consensus scores, the problem defined by Equation (1) is considered, and for each of the 25 possible vectors of scores, the sum of distances to the scores provided by the panelists is computed. The minimizer of this sum of distances (thus the median) is the vector of scores $(2, 3)$.

Consider now that we know that the first sample is fresher than the second sample. Based on the approach where each panelist is represented by one evaluation, the problem defined by Equation (2) is solved, resulting in the vector of scores $(3, 3)$ as the constrained median. Based on the approach where each sample is represented by one evaluation, the problem defined by Equation (3) is solved, resulting in the vector of scores $(2, 2)$ as the constrained median. If we consider that each panelist is represented by one evaluation (i.e., the score of each sample depends on the number of panelists), then it seems logical to change the median score of x_1 to three. However, if we consider that each sample is represented by one evaluation (i.e., the score of each sample depends on the proportion of panelists), then it seems logical to change the median score of x_2 to two.

Table 3. The scores assigned by the panelists in Example 2.

The ith Lowest Score	1	2	3	4	5	6	7	8	9
Score of sample x_1 ($\mathbf{s}^i(1)$)	2								
Score of sample x_2 ($\mathbf{s}^i(2)$)	2	2	3	3	3	5	5	5	5

3. Materials and Methods

The evolution of four fresh Atlantic salmon fillets (A, B, C, D) was studied over time. The data of this study originated from the same experiment performed by [39], who described the adopted materials and methods summarized hereafter.

Each fillet was (equally) divided into five samples that were stored under the same conditions for a specific number of storage days and then analyzed by selective-ion flow-tube mass spectrometry (SIFT-MS); a subindex was added to each salmon sample to indicate the corresponding storage day. Subsequently, after a sample was stored for a specific number of days, it was frozen at $-32\,°C$ under vacuum. The samples were thawed and grouped into five groups of four samples, one sample from each fillet, and were then provided to the panelists in a random order as shown in Table 4. The groups of samples were not provided to the panelists in chronological order to prevent the panelists from recognizing a pattern in the experiments that could affect their evaluations.

Table 4. The order of grouping the salmon samples from different storage days (represented by the corresponding subindex) and the day each group was provided to the panelists.

Group	Samples	Day
1	(A_1, B_2, C_3, D_4)	Tuesday
2	(A_2, B_3, C_4, D_5)	Thursday
3	(A_3, B_4, C_5, D_6)	Monday
4	(A_4, B_5, C_6, D_7)	Wednesday
5	(A_5, B_6, C_7, D_8)	Friday

Several panelists (nine or ten depending on the day) were recruited from the Department of Food Safety and Food Quality at the Faculty of Bioscience Engineering at Ghent University with prior experience in performing sensory evaluation of salmon. Sensory evaluation was based on olfactory evaluation and performed in individual booths under red light at SensoLab in Ghent University. Each

panelist was asked to assign to each sample a score on the 5-point scale described in Figure 1, where the scores "1", "3", and "5" instead represented spoiled, neither spoiled nor fresh, and fresh, respectively.

Additionally, several panelists (between 23 and 28 depending on the day) were recruited from multiple departments at the Faculty of Bioscience Engineering with no prior experience in sensory evaluation of salmon and were asked to express a ranking of the four samples of salmon by ordering them from most fresh to least fresh. For each group of samples, we computed the consensus ranking(s) r^* as described in [39], and we refer to our work for further details on the considered method of [40] for the aggregation of rankings.

4. Results and Discussion

4.1. Consensus Scores of Salmon

The scores assigned to each sample are represented in increasing order in Table 5. The median score for each sample can be directly determined as the scores in the middle of each column, shown in bold in Table 5. Note that for Groups 3, 4, and 5, there was an even number of scores provided by the panelists. Since the scores right before (i.e., $i = 5$) and right after (i.e., $i = 6$) the middle were the same, the median was unique.

Table 5. Scores assigned by the panelists on each day. The medians are shown in bold.

	Group 1				Group 2				Group 3				Group 4				Group 5			
	A_1	B_2	C_3	D_4	A_2	B_3	C_4	D_5	A_3	B_4	C_5	D_6	A_4	B_5	C_6	D_7	A_5	B_6	C_7	D_8
s^1	3	3	1	2	4	3	3	1	1	2	1	1	2	1	1	1	2	2	2	1
s^2	3	4	2	2	4	4	3	2	3	3	2	2	2	2	2	1	3	3	2	1
s^3	4	4	3	3	4	4	3	2	3	4	2	3	3	2	2	1	3	3	2	1
s^4	4	4	3	4	4	4	3	2	4	4	2	3	3	2	2	2	4	3	2	2
s^5	**5**	**4**	**3**	**4**	**5**	**4**	**4**	**2**	**4**	**4**	**3**	**4**	**4**	**3**	**2**	**2**	**4**	**3**	**2**	**2**
s^6	5	4	3	4	5	4	4	2	**4**	**4**	**3**	**4**	**4**	**3**	**2**	**2**	**4**	**3**	**2**	**2**
s^7	5	4	3	4	5	4	4	2	5	5	4	4	5	4	2	2	4	4	2	2
s^8	5	5	5	4	5	5	4	3	5	5	4	4	5	4	3	3	4	5	3	3
s^9	5	5	5	4	5	5	5	4	5	5	4	5	5	5	4	3	5	5	4	4
s^{10}									5	5	5	5	5	5	5	3	5	5	5	5

In what follows, the influence of incorporating knowledge of storage days, results of a clustering analysis, and information from additionally obtained rankings of the samples is illustrated, and the results are discussed.

4.2. Incorporating Knowledge of Storage Days

By considering the consensus scores of each sample separately, it was clear that the scores of the samples from Fillets A and B were decreasing over time. However, the scores of the samples from Fillets C and D were increasing at certain storage days. For instance, sample C_3 was assigned a score of three, and sample C_4 was assigned a higher score of four. Similarly, sample D_5 was assigned a score of two, and sample D_6 was assigned a higher score of four.

To determine the consensus vector of scores that should be assigned to the five samples of the same fillet, while incorporating the knowledge of storage days of the samples, the problem defined by Equation (2) was considered, where the additional constraints are summarized in Table 6. The median and the constrained median for the samples of each fillet are summarized in Table 9. After incorporating the knowledge of storage days of the samples from each of Fillets C and D, the results are seen in the upcoming Table 9 where there are no increasing values of evaluations for any fillet.

Table 6. The constraints based on the storage days of the samples.

Fillet	Constraints
A	$A_5 \prec A_4 \prec A_3 \prec A_2 \prec A_1$
B	$B_6 \prec B_5 \prec B_4 \prec B_3 \prec B_2$
C	$C_7 \prec C_6 \prec C_5 \prec C_4 \prec C_3$
D	$D_8 \prec D_7 \prec D_6 \prec D_5 \prec D_4$

Note that including the knowledge of storage days should only be considered when several factors of the initial conditions of the samples are similar, particularly their contamination, dimensions, composition, and packaging and storage conditions. Verifying the similarity of the initial conditions of the samples required measurements.

From Table 9, it was deduced that, in an ideal situation where the samples from the same fillet had similar initial conditions, incorporating the knowledge of storage days of the samples from Fillet C indicated that sample C_4 should be assigned a lower score and that samples C_3, C_4, and C_5 should be equally scored in terms of freshness. Moreover, it was deduced that incorporating the knowledge of storage days of the samples from Fillet D indicated that sample D_5 should be assigned a higher score and sample D_6 should be assigned a lower score, and these samples should be scored equally in terms of freshness.

There existed several potential risks when incorporating knowledge of storage days of samples that were not initially similar. The main concern was that an assumption was made that the samples had similar spoilage rates, and thus, their assigned scores might be incorrectly related to their storage days. Since microbiological analysis of each salmon sample was not performed, the storage days would not be used as the only tool to compare these samples.

4.3. Incorporating Results of a Clustering Analysis

Recently, the characterization of VOCs using selective-ion flow-tube mass spectrometry (SIFT-MS) has attracted the attention of many researchers and has been validated for fish metabolite research [26,27]. Additionally, researchers have used hierarchical agglomerative clustering [41], a commonly used clustering analysis tool, to establish a relation between the VOC profiles of the samples and their resulting consensus scores [42]. The methods of using SIFT-MS to quantify the VOC profiles and the results of the clustering were described by [39] and are summarized in Table 7.

Table 7. Clustered samples based on the similarity of their VOC profile as described by [39].

Clusters	Samples
Cluster 1	$\{A_1, A_2\}$
Cluster 2	$\{B_2\}$
Cluster 3	$\{A_3, A_4, A_5,$ $B_3, B_4, B_5, B_6,$ $C_3, C_4, C_6,$ $D_4, D_5, D_6\}$
Cluster 4	$\{C_5, C_7, D_7, D_8\}$

To determine the consensus vector of scores that should be assigned to each of the samples, while incorporating the results of a clustering analysis of the samples, the problem defined by Equation (2) with a threshold $\epsilon = 1$ on the absolute difference of the scores of the samples was considered.

The median and the constrained median for the samples of each fillet are gathered in the upcoming Table 9.

From Table 9, it was deduced on the basis of the scores provided by the panelists that, in Cluster 3, samples C_6 and D_5 were less fresh than all the other samples in the cluster. However, incorporating the results of a clustering analysis of all the samples indicated that samples C_6 and D_5 should each be assigned a higher score.

Note that including the results of a clustering analysis should only be considered when the measurements were accurate and the clusters were well defined. Otherwise, there existed the potential risk of inaccurately clustering samples and, thus, assigning incorrect scores. In many real-world datasets, there is no absolute optimal number of clusters. As a result, a balance between a clustering that reflects the data best and a parsimonious model should be determined. For instance, a very small number of clusters may result in a large number of samples in a single cluster, and thus, their assigned scores may be incorrectly set to a close value. Equivalently, a very large number of clusters may result in a small number of samples in each cluster, thus resulting in a small number of constraints.

4.4. Incorporating Consensus Rankings

The simultaneous adoption of different sensory evaluation tests, such as scoring tests for determining the quality of food samples and ranking tests for determining the existence of a significant difference between the samples, has attracted the attention of many researchers [29–32,39]. The additional data produced from the ranking tests could be incorporated to improve the assessment of the quality of the salmon samples. The methods to obtain a consensus ranking of these samples were described by [39], and the results are summarized in Table 8. As the method of [40] was considered for identifying the consensus ranking(s), it might be the case that for some groups, there were multiple consensus rankings (obtained as the minimizers of the Kemeny distance). In such a case, each of these consensus rankings was considered separately as a set of constraints.

Table 8. The consensus ranking(s) of salmon samples (A, B, C, D) in each group.

Group 1	Group 2	Group 3	Group 4	Group 5
$C_3 \prec D_4 \prec B_2 \sim A_1$	$D_5 \prec C_4 \prec B_3 \prec A_2$	$C_5 \prec D_6 \sim A_3 \prec B_4$	$D_7 \prec C_6 \sim B_5 \prec A_4$	$C_7 \prec D_8 \prec A_5 \prec B_6$
		$D_5 \prec C_4 \sim B_3 \prec A_2$	$C_5 \prec D_6 \prec B_4 \sim A_3$	
		$D_5 \prec C_4 \sim B_3 \sim A_2$		
		$D_5 \prec C_4 \prec B_3 \sim A_2$		

To determine the consensus vectors of scores that should be assigned to the four samples in each group, while incorporating the consensus rankings of the samples, the problem defined by Equation (2) was considered. The median and the constrained median for the samples in each group are summarized in the upcoming Table 9.

From Table 9, it was deduced that, in an ideal situation where a large number of panelists provided rankings of the samples, incorporating the knowledge of the consensus ranking of the samples in Group 4 indicated that sample B_5 was similar to sample C_6. Therefore, sample B_5 should be assigned a lower score, resulting in equal scores assigned to samples B_5, C_6, and D_7. Similar conclusions could be drawn for Groups 1, 2, 3, and 5.

Note that incorporating a consensus ranking should only be considered when the number of panelists providing a ranking was large enough. Otherwise, there existed the potential risk of inaccurately ordering samples and, thus, assigning incorrect scores. In this experiment, the number of panelists providing rankings on the salmon samples was between 23 and 28 depending on the group. This number may be considered to be large enough for obtaining consensus rankings; however, having more panelists would result in more reliable consensus rankings. Therefore, in this study, we did not use the consensus rankings as the only tool to compare these samples.

4.5. Comparing the Constrained Medians

To study the influence of incorporating the previously discussed information that invoked different constraints on the median of each sample, we summarize all the medians and the constrained medians for each setting in Table 9.

Table 9. The median and the constrained medians for each sample in every group after incorporating knowledge of storage days, results of a clustering analysis, and consensus rankings.

Method	Group 1				Group 2				Group 3				Group 4				Group 5			
	A_1	B_2	C_3	D_4	A_2	B_3	C_4	D_5	A_3	B_4	C_5	D_6	A_4	B_5	C_6	D_7	A_5	B_6	C_7	D_8
The median	5	4	3	4	5	4	4	2	4	4	3	4	4	3	2	2	4	3	2	2
The constrained median																				
Inc. knowledge of storage days	5	4	3	4	5	4	3	3	4	4	3	3	4	3	2	2	4	3	2	2
Inc. results of a clustering analysis	5	4	3	4	5	4	4	3	4	4	3	4	4	3	3	2	4	3	2	2
Inc. consensus rankings	4	4	3	4	5	4	4	2	4	4	3	4	4	2	2	2	4	4	2	2
								4	4	4	2									

It could be seen that the medians and the constrained medians for some of the samples were equal. A conclusion was reached that the scores assigned to these samples agreed with each type of side information on the samples. However, the medians and the constrained medians for the other samples differed. The additional constraints might provide a better understanding of the score that should be assigned to each sample.

Note that simultaneously incorporating the knowledge of storage days, the results of a clustering analysis and consensus rankings would result in many constraints. As a result, the constrained median of samples C_7, D_7, and D_8 stayed the same while the constrained median of all the other samples was either three or four. The reader should bear in mind that adding too many constraints might result in forcing the scores of all the samples to be the same or, even worse in the case of contradictory constraints, rendering the set $\mathcal{S}$ in Equation (2) empty.

It is important to note that choosing to incorporate side information depends on the reliability of the scores assigned to the samples. For instance, it is recommended to incorporate side information in settings where the assigned score might not be a good representation of the overall quality of a sample, such as when the number of panelists is very small. Moreover, choosing an optimal source of side information depends on the quality of that information. For instance, it is recommended that the initial conditions of the samples should be similar, the clustering analysis should be performed carefully, and the number of panelists performing ranking tests should be large enough.

In our case, since the measurements of the VOC profiles of our samples were accurate and we believed the chemical analysis of the VOC profile was our most reliable source of information, we restricted our attention to the constraints provided by the clustering analysis. We concluded that samples C_6 and D_5 should be assigned a higher score than was originally assigned to them by the classical median.

5. Conclusions

In this paper, we presented a new method allowing combining scores provided by panelists for given samples and different types of side information on these samples. The presented method was especially useful in the setting of sensory evaluation in case the number of panelists providing scores was very small, yet side information on the samples was available or could be easily obtained. It is noteworthy that this method was not limited by the size of the scale or the number of panelists and samples. Other examples of potential applications for this method include, but are not limited to, decision making problems [43,44], online evaluation [45], and recommender systems (e.g., social matching systems and gift, music, and movie recommenders) [46].

Here, we illustrated the method by means of an experiment concerning the freshness of raw Atlantic salmon. Moreover, we discussed the influence of incorporating the knowledge of storage days, the results of a clustering analysis, and the consensus ranking of the samples on the assignment of the consensus scores to these samples. We provided guidelines for incorporating the different types of side information and pointed out their benefits and potential risks.

We end by noting that, in the field of food science, researchers are not only interested in determining the quality of food samples, but also in understanding and identifying the reasons for the scores assigned to the samples. Thus, the resulting consensus scores can be of use in relating the characteristics of samples to their assigned scores.

References

1. Amerine, M.A.; Pangborn, R.N.; Poessler, E.B. *Principles of Sensory Evaluation of Food*; Academic Press: New York, NY, USA, 1965.
2. Meilgaard, M.C.; Civille, G.V.; Carr, B.T. *Sensory Evaluation Techniques*, 4th ed.; CRC Press: Boca Raton, FL, USA, 2006.
3. O'Mahony, M. *Sensory Evaluation of Food*; CRC Press: New York, NY, USA, 1986.
4. Heenan, S.P.; Hamid, N.; Dufour, J.P.; Harvey, W.; Delahunty, C.M. Consumer freshness perceptions of breads, biscuits and cakes. *Food Qual. Prefer.* **2009**, *20*, 380–390. [CrossRef]
5. Karlsen, A.M.; Aaby, K.; Sivertsen, H. Instrumental and sensory analysis of fresh Norwegian and imported apples. *Food Qual. Prefer.* **1999**, *10*, 305–314. [CrossRef]
6. Ouyang, Q.; Zhao, J.; Chen, Q. Instrumental intelligent test of food sensory quality as mimic of human panel test combining multiple cross-perception sensors and data fusion. *Anal. Chim. Acta* **2014**, *841*, 68–76. [CrossRef] [PubMed]
7. Papadopoulou, O.; Panagou, E.Z.; Tassou, C.C.; Nychas, G.J.E. Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage. *Food Res. Int.* **2011**, *44*, 3264–3271. [CrossRef]
8. Rogers, H.B.; Brooks, J.C.; Martin, J.N.; Tittor, A.; Miller, M.F.; Brashears, M.M. The impact of packaging system and temperature abuse on the shelf life characteristics of ground beef. *Meat Sci.* **2014**, *97*, 1–10. [CrossRef]
9. Hein, K.A.; Jaeger, S.R.; Tom Carr, B.; Delahunty, C.M. Comparison of five common acceptance and preference methods. *Food Qual. Prefer.* **2008**, *19*, 651–661. [CrossRef]
10. Lim, J. Hedonic scaling: A review of methods and theory. *Food Qual. Prefer.* **2011**, *22*, 733–747. [CrossRef]
11. Peryam, D.; Pilgrim, F. Hedonic scale method of measuring food preferences. *Food Technol.* **1957**, *11*, 9–14.
12. Stone, H.; Sidel, J.L. *Sensory Evaluation Practices*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2004; pp. 337–344.
13. De Bie, T.; Momma, M.; Cristianini, N. Efficiently learning the metric with side-information. In *Algorithmic Learning Theory*; Gavaldá, R., Jantke, K., Takimoto, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 175–189.
14. Jonschkowski, R.; Höfer, S.; Brock, O. Patterns for learning with side information. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 156–163.
15. Cox, E.P., III. The optimal number of response alternatives for a scale: A review. *J. Mark. Res.* **1980**, *17*, 407–422. [CrossRef]

16. Bendini, A.; Valli, E.; Barbieri, S.; Gallina, T. Sensory analysis of virgin olive oil. In *Olive Oil-Constituents, Quality, Health Properties and Bioconversions*; Boskou, D., Ed.; InTech: Rijeka, Croatia, 2012; Chapter 6, pp. 109–130.

17. Davis, B.M.; McEwan, M.J. Determination of olive oil oxidative status by selected ion flow tube mass spectrometry. *J. Agric. Food Chem.* **2007**, *55*, 3334–3338. [CrossRef]

18. Golia, S.; Brentari, E.; Carpita, M. Causal reasoning applied to sensory analysis: The case of the Italian wine. *Food Qual. Prefer.* **2017**, *59*, 97–108. [CrossRef]

19. Guillaume, S.; Charnomordic, B. Fuzzy inference systems to model sensory evaluation. In *Intelligent Sensory Evaluation*; Ruan, D., Zeng, X., Eds.; Springer: Berlin, Germany, 2004; Chapter 4, pp. 197–216.

20. Van Langeveld, A.W.B.; Gibbons, S.; Koelliker, Y.; Civille, G.V.; de Vries, J.H.M.; de Graaf, C.; Mars, M. The relationship between taste and nutrient content in commercially available foods from the United States. *Food Qual. Prefer.* **2017**, *57*, 1–7. [CrossRef]

21. Lawless, H.T.; Heymann, H. *Sensory Evaluation of Food*, 2nd ed.; Food Science Text Series; Springer: New York, NY, USA, 2010.

22. Butler, G.; Poste, L.M.; Mackie, D.A.; Jones, A. Time-intensity as a tool for the measurement of meat tenderness. *Food Qual. Prefer.* **1996**, *7*, 193–204. [CrossRef]

23. François, N.; Guyot-Declerck, C.; Hug, B.; Callemien, D.; Govaerts, B.; Collin, S. Beer astringency assessed by time-intensity and quantitative descriptive analysis: Influence of pH and accelerated aging. *Food Qual. Prefer.* **2006**, *17*, 445–452. [CrossRef]

24. Noble, A.C. Application of time-intensity procedures for the evaluation of taste and mouthfeel. *Am. J. Enol. Vitic.* **1995**, *46*, 128–133.

25. Gram, L.; Ravn, L.; Rasch, M.; Bruhn, J.B.; Christensen, A.B.; Givskov, M. Food spoilage-interactions between food spoilage bacteria. *Int. J. Food Microbiol.* **2002**, *78*, 79–97. [CrossRef]

26. Noseda, B.; Ragaert, P.; Pauwels, D.; Anthierens, T.; Van Langenhove, H.; Dewulf, J.; Devlieghere, F. Validation of selective ion flow tube mass spectrometry for fast quantification of volatile bases produced on atlantic cod (gadus morhua). *J. Agric. Food Chem.* **2010**, *58*, 5213–5219. [CrossRef] [PubMed]

27. Noseda, B.; Islam, M.T.; Eriksson, M.; Heyndrickx, M.; De Reu, K.; Van Langenhove, H.; Devlieghere, F. Microbiological spoilage of vacuum and modified atmosphere packaged Vietnamese Pangasius hypophthalmus fillets. *Food Microbiol.* **2012**, *30*, 408–419. [CrossRef]

28. Olivares, A.; Dryahina, K.; Spaněl, P.; Flores, M. Rapid detection of lipid oxidation in beef muscle packed under modified atmosphere by measuring volatile organic compounds using SIFT-MS. *Food Chem.* **2012**, *135*, 1801–1808. [CrossRef]

29. Bolhuis, D.P.; Costanzo, A.; Keast, R.S. Preference and perception of fat in salty and sweet foods. *Food Qual. Prefer.* **2017**, *64*, 131–137. [CrossRef]

30. Bowman, T.L.; Barringer, S. Analysis of factors affecting volatile compound formation in roasted pumpkin seeds with selected ion flow tube-mass spectrometry (SIFT-MS) and sensory analysis. *J. Food Sci.* **2012**, *77*, C51–C60. [CrossRef] [PubMed]

31. Kuuliala, L.; Al Hage, Y.; Ioannidis, A.G.; Sader, M.; Kerckhof, F.M.; Vanderroost, M.; Boon, N.; De Baets, B.; De Meulenaer, B.; Ragaert, P.; et al. Microbiological, chemical and sensory spoilage analysis of raw Atlantic cod (Gadus morhua) stored under modified atmospheres. *Food Microbiol.* **2018**, *70*, 232–244. [CrossRef] [PubMed]

32. Øvrum, A.; Alfnes, F.; Almli, V.L.; Rickertsen, K. Health information and diet choices: Results from a cheese experiment. *Food Policy* **2012**, *37*, 520–529. [CrossRef]

33. Calkins, C.R.; Hodgen, J.M. A fresh look at meat flavor. *Meat Sci.* **2007**, *77*, 63–80. [CrossRef] [PubMed]

34. Sullivan, G.A.; Calkins, C.R. Ranking beef muscles for Warner-Bratzler shear force and trained sensory panel ratings from published literature. *J. Food Qual.* **2011**, *34*, 195–203. [CrossRef]

35. Ennis, D.M. The power of sensory discrimination methods. *J. Sens. Stud.* **1993**, *8*, 353–370. [CrossRef]

36. Bi, J. *Sensory Discrimination Tests and Measurements: Sensometrics in Sensory Evaluation*, 2nd ed.; John Wiley & Sons: Chichester, UK, 2015; pp. 60–97.

37. Amoore, J.E.; Venstrom, D.; Davis, A.R. Measurement of specific anosmia. *Percept. Mot. Skills* **1968**, *26*, 143–164. [CrossRef] [PubMed]

38. Harwood, M.L.; Ziegler, G.R.; Hayes, J.E. Rejection thresholds in chocolate milk: Evidence for segmentation. *Food Qual. Prefer.* **2012**, *26*, 128–133. [CrossRef]

39. Sader, M.; Pérez-Fernández, R.; Kuuliala, L.; Devlieghere, F.; De Baets, B. A combined scoring and ranking approach for determining overall food quality. *Int. J. Approx. Reason.* **2018**, *100*, 161–176. [CrossRef]

40. Kemeny, J.G. Mathematics without numbers. *Daedalus* **1959**, *88*, 577–591.

41. Rokach, L.; Maimon, O. Clustering methods. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O.R., Ed.; Springer: Boston, MA, USA, 2005; Chapter 15, pp. 321–352.

42. Kuuliala, L.; Abatih, E.; Ioannidis, A.G.; Vanderroost, M.; De Meulenaer, B.; Ragaert, P.; Devlieghere, F. Multivariate statistical analysis for the identification of potential seafood spoilage indicators. *Food Control* **2018**, *84*, 49–60. [CrossRef]

43. Chiclana, F.; Herrera, F.; Herrera-Viedma, E. Integrating multiplicative preference relations in a multipurpose decision-making model based on fuzzy preference relations. *Fuzzy Sets Syst.* **2001**, *122*, 277–291. [CrossRef]

44. Ignacio, J.; Cabrera, F.E.; Vargas, L.G. Knowledge-based systems estimating the importance of consumer purchasing criteria in digital ecosystems. *Knowl. Based Syst.* **2018**, *162*, 252–264.

45. Peláez, J.I.; Bernal, R.; Karanik, M. Majority OWA operator for opinion rating in social media. *Soft Comput.* **2016**, *20*, 1047–1055. [CrossRef]

46. Nunes, M.A.S.; Hu, R. Personality-based recommender systems. In *Proceedings of the 6th ACM Conference on Recommender Systems*; ACM Press: Dublin, Ireland, 2012; pp. 5–6.