Innovative Food Science and Emerging Technologies 96 (2024) 103771

Contents lists available at ScienceDirect

Innovative Food Science and Emerging Technologies

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

Diagnosis based on sensory data: Application to wheat grading quality

Mélanie Munch [a,c,*], Cédric Baudrit [b], Hubert Chiron [a], Benoît Méléard [d], Luc Saulnier [a], Kamal Kansou [a]

[a] INRAE UR BIA, Imp. Yvette Cauchois, Nantes 44300, France
[b] INRAE, Bordeaux University-I2M, 351 Cours de la Libération, Talence 33405, France
[c] INRAE UMR STLO, 65 Rue de Saint-Brieuc, Rennes 35000, France
[d] Arvalis-Institut du Végétal, Station Expérimentale, Boigneville 91720, France

A R T I C L E I N F O

Keywords:
Food quality
Sensory data
Expert knowledge
Wheat grading

A B S T R A C T

Sensory evaluation is an important aspect of food quality and control. However, even when carried out by a group of experts, it is generally difficult to link the results of a sensory evaluation to physico-chemical or technological measurements. This study is based on the premise that formalising the interpretation of sensory observations in terms of the physical state of the product can help to link together sensory and physical properties. The main proposal of this paper is a methodological framework adapted from a diagnostic approach to capture the relationships between sensory evaluations of a type of product, here wheat dough, and its physical states called quality profiles. A probabilistic analysis is proposed to identify the quality profiles and their signatures, i.e. the corresponding sensory observations that result from grouping the probabilities of the observations. This work is supported by the analysis of a large historical sensory evaluation dataset from the routine application of the French baking standard to estimate the baking value of common wheat (*Triticum aestivum* L.) flour. Application of the method to this dataset revealed two defective quality profiles for wheat dough, Slackening (due to weakness of the gluten network) and Resistant (excessive strength of the gluten network), along with their signatures in terms of sensory observations of the dough. Promising relationships were found between the quality profiles attributed to the wheat samples and usual technological criteria of the wheat flour quality: gluten index, (Ie) elasticity index and (W) dough strength. This methodological framework applied to food opens up interesting perspectives for the use of sensory data for crop and food quality assessment using computational approaches.

1. Introduction

In agri-food systems, the evaluation of crop or product quality often includes a dose of sensory evaluation, in addition to measurements from technological tests and compositional analysis. Sensory evaluation is used, for example, for grading crops before primary transformation, as in the Speciality Coffee Association of America (Lingle, 2011) standards for grading coffee beans, or the Bread Baking Assay standard (NF V03–716) for grading soft/common wheat in France. Among the various types of measurements used to assess the quality of crop or food, sensory evaluation is generally close to the perception of the end user but difficult to relate to the physical properties and composition of the product.

Ruan and Zeng (2004) distinguish two types of sensory evaluations on industrial products, the one performed by trained experts for product design and development (B2B) using analytical and neutral descriptors and the one used for consumer and marketing research (B2C) by untrained consumer panels using analytical and hedonic descriptors. The sensory evaluation of agricultural product quality using standards belongs clearly to the first type. The involvement of trained experts and the use of neutral sensory descriptors, generally associated to the technological knowledge on the product, should reduce the intra-individual variability and produce consistent evaluation over time compared to hedonic evaluations. Even in these controlled conditions, the inter-individual variability may be significant and the relations between physical/chemical or technological features of the product and the sensory criteria are complex (e.g. nonlinear) (Ruan & Zeng, 2004). Moreover training expert to sensory evaluation is time and resource consuming; not surprisingly, food technology has sought to replace sensory evaluation with technological measurements (see (Bourne,

* Corresponding author at: INRAE UMR STLO, 65 Rue de Saint-Brieuc, Rennes 35000, France.
E-mail address: melanie.munch@inrae.fr (M. Munch).

https://doi.org/10.1016/j.ifset.2024.103771
Received 3 April 2024; Received in revised form 26 July 2024; Accepted 27 July 2024
Available online 31 July 2024
1466-8564/

2002) for a reference book on this topic). Establishing the relationships between sensory data and the physical conditions of a food product is critical for decision support, for example for monitoring physical changes in the product (Baudrit, Sicard, Wuillemin, & Perrot, 2010; Curt, Trystram, Nogueira-Terrones, & Hossenlopp, 2004). However, this requires an appropriate method to exploit the data obtained from the sensory evaluations of the food products, which is an additional difficulty.

Classical statistics and factorial analysis methods have been used for a long time to address the problem of physical interpretation of sensory observations, but they are sometimes insufficient (Ruan & Zeng, 2004), as for example MCA, which provides a low-dimensional representation of the data that can lead to a loss of information. Intelligent computing techniques (especially fuzzy logic) have been applied in the field of food quality and control to mimic the reasoning of experts on sensory criteria (Allais, Perrot, Curt, & Trystram, 2007a; Birle, Hussein, & Becker, 2013; Mavani et al., 2022; Nunes et al., 2023). In particular, they have been applied to sensory control of food transformations processes such as cheese ripening (Baudrit et al., 2010; Curto et al., 2020; Perrot et al., 2004), sausage drying and biscuit aeration (Allais, Perrot, Curt, & Trystram, 2007b), or the prediction of coffee bean sorting (Livio & Hodhod, 2018) or wheat dough condition (Kansou, Chiron, Della Valle, Ndiaye, & Roussel, 2014; Ndiaye, Valle, & Roussel, 2009). Most of these models were designed to compute sensory observations in order to support decision in an industrial context, they did not integrate the physical interpretation of these observations. An illustration of model relating sensory observations to the product physical/chemical state is provided by Baudrit et al. (2010)'s dynamic Bayesian network of the cheese ripening. This model combines a model of the physical/chemical processes and expert-based sensory indicators of the ripening phases. The approach used to capture the sensory indicators is presented in Sicard, Baudrit, Leclerc-Perlat, Wuillemin, and Perrot (2011); it is based on the understanding of operator's cognition during the process control. Roughly experts learn with experience to relate characteristic groups of sensory observations, called "chunks, with the product condition which helps them to diagnose a process drift or a defective product for example; main aspects of this theory are reported in Sect.2.1. Thus, chunks helps experts to make the link between sensory observations and the product physical state, therefore our first hypothesis is that the identification of chunks can be used to relate the sensory observations with the physical/chemical or technological measurements. However in the food domain the relation between the product defective states and the corresponding chunk of sensory observations is largely tacit knowledge held by domain experts, making it difficult to capture (Kansou et al., 2022). Our second hypothesis is that chunks can be approximated by clusters of sensory observations obtained from an appropriate analysis of sensory data. To assess these two hypotheses this work is grounded on the analysis of sensory data collected through the routine application of the bread baking test standard (NF V03–716) to characterise the quality of soft wheat grain grown in France over the last two decades, which amounts to more than 10,000 sensory evaluations.

This works led up to two contributions, for both the sensory evaluation in food industry (Section 2) and the wheat quality (3). In Section 2 this article presents a novel method for modelling expert interpretation of product quality based on a probabilistic analysis of historical sensory data. In Section 3, a real-world application, which aims to better integrate sensory evaluation in the determination of wheat quality, is fully addressed. This includes the assignment of wheat quality classes and the confrontation with the usual technological measurements of wheat quality. Section 4 discusses the various results and possible extensions that this work opens up for sensory analysis in general.

2. A diagnosis approach for the sensory evaluation of food

This section presents the formalisation of the diagnostic approach for the use of sensory data. By way of illustration, the example of the wheat dough kneading sensory evaluation is presented in the context of the French Breadmaking Assay. The application to the other unit operations of the bread-making assay is given in section.3.

2.1. Sensory-based diagnosis for food product

The act of diagnosis is the act of proposing a defect (the output) as a plausible cause for a set of observations of a product (the input) (Schreiber et al., 2001), denoted **signature** (Cordier, Travé-Massuyès, & Pucel, 2006). In the case of a food quality, the state of the product, denoted **quality profile**, is the output diagnosed from the analysis of observations that are the results of technological tests or sensory evaluations (inputs).

Intuitively, the quality profile is based on the knowledge and understanding of the product's behaviour, while the observation is based on human perception. For Raufaste et al. (Raufaste, Eyrolle, & Mariné, 1998), medical diagnosis involves two processes, cognitive (knowledge of the possible inputs and outputs of the diagnosis) and perceptual (how the inputs are perceived). Perceptual processing in particular plays an important role in sensory analysis, while cognitive knowledge is useful in formulating a plausible causal explanation for the observations. Experts in particular are able to associate a set of observations with a piece of knowledge: Ballester et al. (Ballester, Patris, Symoneaux, & Valentin, 2008) shows for example that trained experts are able to recognise and characterise a wine better than novices. To explain this ability, Chase et al. (Chase & Simon, 1973) introduce the concept of chunks through the study of chess, where expert players identify patterns that allow faster mobilisation of knowledge. Thus, chunks are typical configurations of situations acquired by experts during their practice. Sicard et al. (Sicard et al., 2011) apply this theory to capture expertise in the cheese ripening process. In this study, chunks of sensory observations used by experts to determine the main stages of ripening were made explicit, and this knowledge was integrated into a predictive model of cheese ripening dynamics (Baudrit et al., 2010).

As shown in Fig.1, baking experts associate quality profiles (the output) with chunks (patterns within the observations, the inputs) when making a diagnosis from sensory evaluations. Therefore, a food diagnostic model must rely on domain expertise to define the quality profiles and then to associate the chunks of observations with the quality profiles. To elicit domain expertise, Sicard et al. (Sicard et al., 2011) conducted interviews with experts in cheese making, but this approach is time consuming and exhausting for domain experts because the know-how of food experts, built up through experience, is often tacit knowledge that is particularly difficult to put into words (Kansou et al., 2022; Wooten & Rowley, 1995). In addition, human perceptions of probability are known to be biased by prior knowledge or expectations (Kahneman

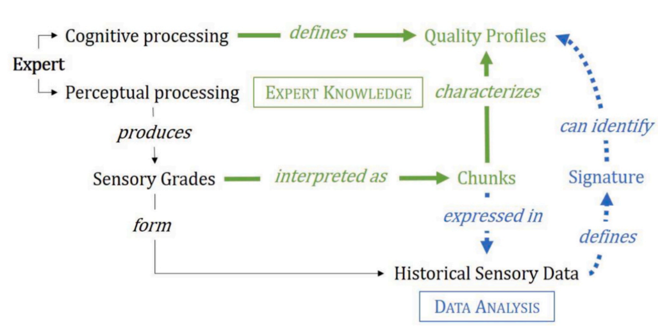

Fig. 1. Principle of food quality diagnosis based on sensory data. In green (Expert Knowledge) the reasoning steps of a diagnostic process through chunks (i.e. expert-based group of observations). In blue (Data Analysis) the approach followed in this work, which uses the historical sensory data set to establish the diagnostic process steps through the identification of signatures (i.e. data based group of observations). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

& Tversky, 1973; Raufaste, da Silva Neves, & Mariné, 2003), making it difficult to objectively associate quality profiles with observations.

As shown in Fig.1, our working hypothesis is that historical sensory data (i.e. a large collection of test results) can be used to identify signatures and bypass expert chunk elicitation using data analysis. This hypothesis assumes that expert chunks are reflected by patterns in the data set. The approach presented in this paper shows how expert knowledge and data analysis of sensory data can be combined to build a diagnostic model for the sensory evaluation of a food product.

Example 2.1. A simple test to characterise a sample of bread during kneading is considered. It consists of five sensory attributes (**Dough Stickiness**, **Slackening**, **Consistency**, **Extensibility** and **Elasticity**) extracted from the sensory Bread Making test (Sec.3 details more in-depth results). Three values are defined: **N**ormal and either **I**nsufficient or **E**xcessive, the two latter being problematic. In this small example, only five defects are considered: Excessive **Dough Stickiness**, **Slackening**, **Consistency** and **Elasticity**, and Insufficient **Extensibility**. According to experts, three dough quality profiles can be diagnosed: a gold standard one (i.e. no defect is observed) and two defective dough behaviours, slackening dough due to weak gluten network (`WG`) and resistant dough due to strong gluten network (`SG`). The test principle is to assess deviations from the **N**ormal value for the five attributes. The signature of the gold standard profile is the normal value for the five attributes. The quality profiles `WG` and `SG` have no defined signatures. The aim of this example is to show how the signatures can be identified from sensory data.

2.2. Probabilistic analysis of the observations

In order to reason on a probability basis, this section introduces the formal notation of sensory tests. Be a test with n attributes denoted X_i ($i \in [1;n]$), and $X_i(v)$ an observation of this attribute taking the value v. Following the probabilistic notation, each observation $X_i(v)$ is associated to its probability of happenstance, denoted $\mathrm{P}(X_i(v))$. For instance, $\mathrm{P}(X_i(v)) = 0.4$ means that X_i has a probability of 0.4 to take the value v. By extension, given $i \neq j$, $\mathrm{P}(X_i(v_i)|X_j(v_j))$ denotes the probability of observing $X_i(v_i)$ knowing that $X_j(v_j)$ is known.

Example 2.2. Table 1 shows the conditional probability table between the **Consistency** and **Elasticity** defects. The first table shows the observation of **Consistency** when **Elasticity** is known, the second the observation of **Elasticity** when **Consistency** is known. These two examples show the non symmetry of conditional probabilities: P(**Consistency(I)**|**Elasticity(I)**) = 0.11 while P(**Elasticity(I)**|**Consistency(I)**) = 0.79. In other words, this means that an insufficiency in **Elasticity** is highly probable ($p = 0.79$) when an insufficiency in **Consistency** is known; but knowing there is an insufficiency in **Elasticity** doesn't guarantee ($p = 0.11$) to observe an insufficiency in **Consistency**.

Table 1
Example of a conditional probability tables describing relationships between **Consistency** and **Elasticity** defects.

P(**Consistency**|**Elasticity**)	**Consistency(I)**	**Consistency(E)**
Elasticity(I)	**0.11**	0
Elasticity(E)	0	**0.03**

P(**Elasticity**|**Consistency**)	**Elasticity(I)**	**Elasticity(E)**
Consistency(I)	**0.79**	0
Consistency(E)	0.17	**0.39**

For instance, P(**Consistency(I)**|**Elasticity(I)**) (0.11) represents the probability of having an insufficiency of **Consistency** when an insufficiency in **Elasticity** is already known. Conversely, P(**Elasticity(I)**|**Consistency(I)**) (0.79) represents the probability of having an insufficiency of **Elasticity** when an insufficiency in **Consistency** is known. Bolded probabilities represent the maximum likelihood.

While this approach is not causal (one cannot say from $\mathrm{P}(X_i(v_i)|X_j(v_j))$ that observation $X_j(v_j)$ caused $X_i(v_i)$), it gives an overview of the correspondences between pairs of observations: "When $X_j(v_j)$ is known, the probability of observing $X_i(v_i)$ is low/high". This approach helps to discover patterns (groups of observations that occur together) that could constitute potential signatures.

2.3. Signatures identification

This section shows how naive Bayes network models (Zhang, 2004) are used to identify patterns and associate them with signatures. We consider a food product with different quality profiles d, and a test with m boolean observations o_k ($k \in [1,m]$), each representing whether or not an attribute took a specific value. A naive Bayes model is proposed (Fig.2), whose simple structure makes it easy to understand how each feature contributes to the final diagnosis. It is composed of (a) each observation o_k (boolean variables indicating whether a value is observed or not) and (b) the variable **Quality Profile** (taking its value within the set of all possible quality profiles d). The relations are such that:

1. Knowing the value d of **Quality Profile** gives total information about each observation o_k.
2. Each observation o_k is independent of the others when **Quality Profile** is known. This is due to the independence property of naive Bayes: when **Quality Profile** is known, information does not flow between observations: knowing o_i has no effect on o_j if their common cause **Quality Profile** is defined.

To uncover potential signatures, a conditional probability table, i.e.mxm matrix is constructed so that, given a row i and a column j, the probability $\mathrm{P}(o_j|o_i)$ is given. The matrix is then clustered in row and column, using the Ward pairwise distance method:

1. Horizontal clusters (Behaviour clusters) between observations when their value is Known, to study how they affect the remaining ones. This cluster describes groups of observations that can be used as potential signatures for quality profiles. If observations $X_i(v)$ and $X_j(k)$ are clustered together, it suggests that other defects will have the same probability of happening when one of the both is known;
2. Vertical clusters (Accuracy clusters) between the remaining Observed defects when another one is known, to study their frequency. This cluster helps refine potential signatures by trimming observations

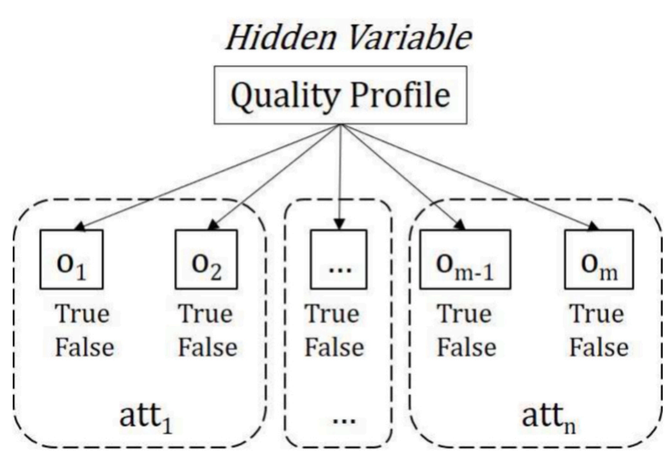

Fig. 2. Naive Bayes model for diagnosis. The value of quality profile has influence over the different attributes' observations. Knowing the quality profile value guarantees independence between observations. Observations o_k ($k \in [1,m]$) are booleans representing whether or not an attribute takes a value.

that are not sufficiently characteristic of a quality profile. If observations $X_i(v)$ and $X_j(k)$ are clustered together, it suggests that $X_i(v)$ and $X_j(k)$ have similar behaviour when different defects are known;

Results are presented in a clustermap.

Example 2.3. Fig.3 presents the relations between different observations. It shows for instance that when **Slackening(E)** is KNOWN, the OBSERVED attributes **Dough stickiness(E)** and **Extensibility(I)** both happen with probabilities of respectfully 0.91 and 0.42. Looking at their Behaviour cluster C_1 shows that **Dough Stickiness(E)** has a similar impact on OBSERVED defects. Being in the same Accuracy cluster C_A, they also roughly have the same probability of appearing when another defect is KNOWN: for instance, **Dough stickiness(E)** and **Extensibility(I)** have respectively probabilities of 0.09 and 0.05 when **Consistency(E)** is known.

2.3.1. Behaviour clusters

In the map reported in Fig.3, rows are grouped as the **Behaviour clusters**. Behaviour clusters group by similarity the observations that condition the likelihood of the other observations, namely $X_j(v_j)$ in P$(X_i(v_i)|X_j(v_j))$, describing behaviours exhibited by the sample. Behaviour clusters thus represent potential signatures for quality profiles.

Example 2.4. The clusters of the example are shown in Fig.3. The observations in C_1 describe a dough with **Stickiness(E)** and **Slackening (E)** happening together; while those in C_2 describe a dough with **Consistency(E)**, **Extensibility(I)** and **Elasticity(E)** happening together. This clustering is compatible with the two quality profiles expected by the experts, where C_1 describes a dough with a weak gluten network and C_2 a strong gluten network. Thus, in a first approach, both C_1 and C_2 are assigned with signatures s_{WG} and s_{SG} for the weak and strong gluten network defect profiles.

2.3.2. Accuracy clusters

In the map reported in Fig.3, columns are grouped as the accuracy clusters. Accuracy clusters are used to trim potential signatures described in Behaviour clusters; they convey information about the probability of the observations, namely $X_i(v_i)$ in P$(X_i(v_i)|X_j(v_j))$. More formally, given a known quality profile d, a good signature s_d is one with.

- The highest sensitivity $P(s_d|d)$, which is the probability of observing s_d knowing that d is true (true positive).
- The lowest fall-out $P(s_d|\neg d)$, which is the probability that s_d is observed knowing d is false (false positive).

The tolerance for both values depends on the expectations experts have of the diagnostic tool. For example, a sensitivity of 0.5 may not be high enough, as it means its associated quality profile, when true, would be detected on average once out of two times. On the other hand, a low fallout guarantees the absence of false alarms: a fallout of 0.5 might be too high, as the signature would be observed once out of two times, even if the quality profile is not true.

However, **Quality Profile** d is a hidden variable: its probability cannot be computed directly and must be approximated to estimate

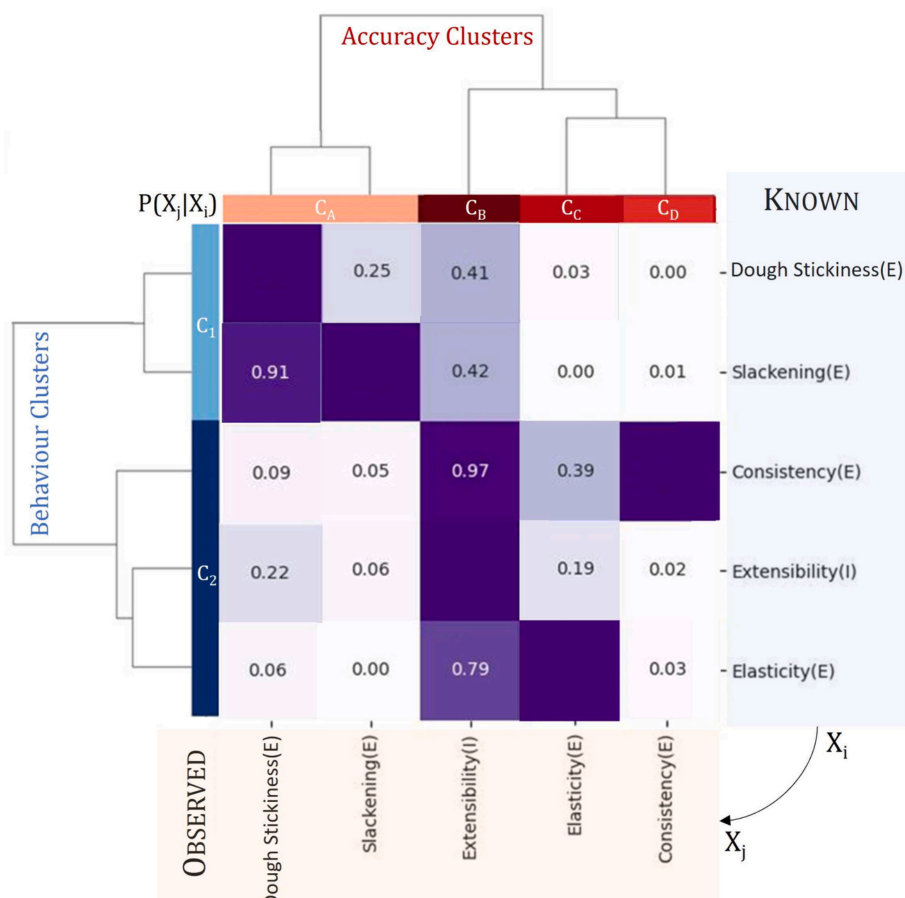

Fig. 3. Clustermap of probabilities for the reduced example. Each case $(i; j)$ (row i, column j) indicates the probability P$(j|i)$, *i.e.* the probability of the OBSERVED defect j when defect i is KNOWN. Probabilities between two identical observations are not indicated, since their value is 1 by definition. Behaviour clusters define C_1 and C_2. Accuracy clusters define C_A, C_B, C_C and C_D.

sensitivities and fallout. If d is true and s_d is its signature (which has been identified by Behaviour clusters), then the observations $o \in \{o_1, \ldots o_m\}$ that are true in s_d must be observed. Given an observation o, we approximate $P(o|d)$ as

$$P(o|s_d) = \frac{1}{|s_d/o|} \sum_{o_j \in s_d/o} P(o|o_j) \tag{1}$$

Note that this approximation does not take into account the probability of the signature occurring, nor the possible interactions between the observations. The aim of this calculation is only to highlight whether o has a high (or low) average probability of occurring along s_d's observations.

Example 2.5. Table 2 shows for each accuracy cluster the average probability of its observations with respect to s_{WG} and s_{SG}. For example, $P(\textbf{Stickiness(E)}|C_2)$ is the average between 0.09, 0.22 and 0.06. Since C_A has high sensitivities and low fallout for s_{WG} (0.91 versus 0.12 for **Stickiness(E)**, 0.25 versus 0.03 for **Slackening(E)**), and $C_1 = C_A$, we keep $s_{\text{WG}} = C_1 = \{$**Stickiness(E)**, **Slackening(E)**$\}$.

On the other hand, the observations of C_2 are split between C_B, C_C and C_D. Both C_C and C_D have a low sensitivity (0.29 and 0.03), and C_B has a much too high fallout (0.42). The definition $s_{\text{SG}} = C_2 \cap (C_C \cup C_D)$ means that the diagnosis would be hard to get (since the highest sensitivity is 0.29), but could be trusted (since the highest fallout is 0. 02); $s_{\text{SG}} = C_2 \cap (C_B \cup C_C \cup C_D)$, on the other hand, guarantees a better detection of the associated quality profile (since **Extensibility(I)** has a sensitivity of 0.92, making it certain to be observed if $d = SG$), but with a higher risk of error if only **Extensibility(I)** is observed.

2.4. Conclusion

In conclusion, this approach aims at proposing a new way to look at sensory data, and a method to actually determine the components of a diagnostic model, summarised in Fig.4.

- **Behaviour clusters** highlight groups of Known defects which have a similar impact on the other Observed defects conditional probabilities. Associated with quality profiles, they represent potential signatures for diagnosis.
- **Accuracy clusters** make it possible to visualise the sensitivity (the odd of detecting a given profile) and the fallout (the odd of false detections) of Observed defects when a Known defects is fixed. They give an indication of the diagnostic performance of the signature. To be noted however, the signature composition itself is a matter of choice, one could choose to increase the sensitivity or, on the contrary, to keep low fallout to reduce false positive.

Example 2.6. Signature truncation can occur for a number of reasons. In the example Table 2, **Consistency(E)** is a very rare observation, but it is slightly more likely to be associated with a SG quality profile. For this reason it could be included in the s_{SG} signature, while keeping in mind that the absence of **Consistency(E)** for a dough does not exclude a strong gluten quality profile.

Table 2
Average probability of each defect to appear along s_{WG} and s_{SG} (**maximum likelihood**).

Accuracy	Defect	$s_{\text{WG}} = C_1$	$s_{\text{SG}} = C_2$
C_A	**Stickiness(E)**	**0.91**	0.12
	Slackening(E)	**0.25**	0.03
C_B	**Extensibility(I)**	0.42	**0.92**
C_C	**Elasticity(E)**	0.02	**0.29**
C_D	**Consistency(E)**	0.01	**0.03**

Finally, it is important to note that signatures only reflect how close a food is to a quality profile: the more observations it shares with a signature, the more confident the diagnosis is. Confidence can be expressed in different ways: the next section presents an application with an ad hoc score that illustrates how much a bread sample expresses (or does not express) a quality profile.

3. Application to the bread making test

In this section, an application is developed to show how observation clustering was used to determine signatures for new quality profiles. It is composed of several steps:

1. Experts define the expected quality profiles (in this application, slackening and resistant) they use to qualify wheat quality.
2. From a database of sensory observation, a matrix of conditional probabilities is computed to reflect the defects relations. A clustering in row and column is then applied to obtain the Behaviour and Accuracy clusters.
3. Signatures are built from the clustering to reflect the two quality profiles expected by experts. From these signatures a score is derived to compute the closeness of a given sample and the two behaviours.
4. From the two scores derived from signatures seven new quality profiles are designed using expert's inputs. These profiles illustrate the existing diversity of dough's behaviours.

The seven new quality profiles are then evaluated by comparing their correlation with technological measures.

3.1. Database description

The database used in this work has been developed to store the measurements commonly used for wheat quality. It consists of two parts:

1. **Breadmaking test results.** This test, described in AFNOR standard NF-V03–716, is widely used in the French context to assess the baking value of common wheat. The dataset contains the results for 11,184 batches of wheat collected over the period 2002–2022 by ARVALIS, the French applied agricultural research organization dedicated to arable crops. The size and time range guarantee a good representation of the different quality profiles likely to be encountered during the breadmaking test.
2. **Technological measurements.** In addition to the baking tests, the assessment of wheat quality is also based on technological measurements carried out on wheat (Hajšelová & Alldrick, 2003). For 10,897 wheat samples the measurements are non-systematic and changing, for the 287 wheat lots measured during the period 2021–2022 the measurements are exhaustive and complete. In contrast to the baking test, the technological measurements were not performed systematically over the twenty years. The technological data will be used to validate the method.

The remainder of this section presents in more detail (1) the characteristics of the bread-making test and (2) the four technical measures selected: protein content, gluten index and two measures from Chopin's alveograph, Ie and W.

3.1.1. Bread making test

The French Bread Baking Test describes the different operation units of bread making and the associated sensory measurements used to characterise the process and bread quality. It includes a rating scale with a maximum of seven values (from **I**nsufficiency to **E**xcess, $1 \prec 4 \prec 7 \prec 10 \prec 7 \prec 4 \prec 1$), with the reference value for a standard French bread making process as the central value equal to 10. By definition, a defect is characterized by a rating $<$10. Some attributes have only one type of defect and are then rated on a four-point scale, e.g. stickiness can

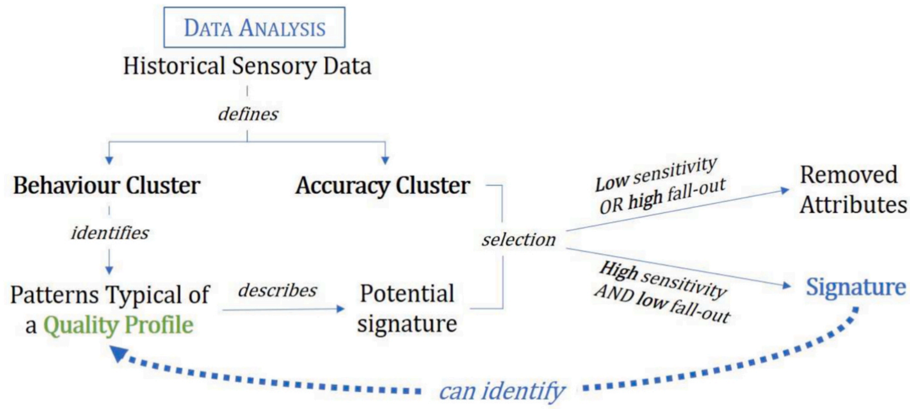

Fig. 4. Overview of the historical sensory data analysis that allows signatures to be linked to quality profiles. Behaviour clusters are used to identify patterns describing potential signatures; accuracy clusters trim the potential signatures to keep only observations with high sensitivity and low fallout for their assigned quality profile.

only be excessive, never insufficient. The test assesses the main unit operations of the breadmaking process, from kneading to baked bread (including crumb and crust evaluation). In this article we focus only on the diagnostic of the dough which involves the following operation units:

- **Kneading (K).** After the mixing of the ingredients of common french bread, six attributes are measured (such as the dough smoothing, the stickiness, …).
- **First Rising (FR).** The dough is left to rise a first time. Only one attribute (the slackening) is measured.
- **Dividing (D).** Once the dough has risen, dough pieces are formed to shape the future breads. Four attributes (lengthening, tearing, …) are measured.
- **Second Rising (SR).** Dough pieces are left to rise a second time. Two attributes (rising level, tearing) are measured.
- **Baking (B).** Dough pieces are baked. Before being put in the oven, two attributes are measured (stickiness and free standing).
- **Bread analysis (Ba).** While not directly about dough, experts have underlined that dough's behaviour and the size of the bread's section lengthwise (after being cut) could be strongly linked. That is why this attribute is also added in the study.

The results are presented in a specific evaluation grid from which three scores are calculated as a weighted sum on a scale from 0 to 100, one for the dough (denoted dough grade), one for the bread aspect and one for the crumb. The sum of the three scores gives the overall score of the baking test, widely used in the wheat production sector to distinguish between gold standard and defective quality products. A maximum value (300) is reached when all the attributes score 10. As such, the global score is a relevant and widely used indicator of the distance between a given dough characteristic and gold standard quality. However, it does not provide any information about the defective properties. Indeed, insufficiencies or excesses are penalized in roughly the same way. To trace back the defects at the origins of a low score, it is necessary to look at the sensory observations reported in the grid. Diagnosing the physical state at the origin of the defects requires specific knowledge of the baking test and the physical behaviour of the dough.

3.1.2. Technological measurements

The **protein content** is a widely accepted criterion used worldwide for wheat quality. Roughly speaking, the protein content is a proxy for the insoluble protein content, which is directly related to the ability to form the gluten network of the dough. In the database, protein content is measured using a near-infrared spectrometer (standard method NF EN 15948).

The **Gluten Index** is a measure of the strength of the gluten network. It is determined by weighing the remaining quantity of gluten from a dough after washing, using the Perten Glutomatic (NF EN ISO 21415-2).

The Chopin alveograph method (NF EN ISO 27971) is a wheat flour characterisation method based on dough rheology (Dubois, Dubat, & Launay, 2008) to provide information on dough extensibility and elasticity. This study focuses on two measurements: W, widely used for flour evaluation worldwide (Dobraszczyk, 2004), and Ie, a good predictor of dough behaviour according to domain experts and highly correlated with a critical rheological property (Strain Hardening Index) (Jødal & Larsen, 2021).

3.2. Quality profiles signatures determination

In addition to the gold standard quality profile (no defect), two consensual quality profiles for French bread making were elicited from four domain experts. Based on their explanations, they can be defined as follows:

(a) **Slackening Profile** `SP` describes a dough that is too sticky, lacks consistency and elasticity and tends to flow. This quality profile is associated with a weak gluten network.
(b) **Resistant Profile** `RP` describes a strong dough able to resist intensive mechanical action. For French bread making, it is difficult to handle as it retracts too much. In particular, it presents an insufficient lengthening. This quality profile is associated with a excessively elastic gluten network.

Similar to the example developed in section.2, the focus is on defects, namely when the attribute value is either **E**xcess or **I**nsufficiency. Given the quality profiles descriptions, the clustering is shown in Fig.5 for all the selected breadmaking steps. The clusters C_1 and C_2 are easily assigned to the quality profiles from the definitions provided by the experts as follows: $s_{\mathrm{SP}} = C_1$ and $s_{\mathrm{RP}} = C_2$. In addition, the cluster map shows a third cluster C_3 which, according to the experts, does not correspond to any known defect profile nor typical dough behaviour.

From the results shown in Fig.5, sensitivities are calculated using Eq.1 and shown in Table 3. It shows that $C_D \cup C_E$ has the best sensitivities (0.60 and 0.29) and fall-out (0.06 and 0.02) for s_{SP}. Since $(C_D \cup C_E) \subset C_1$, this gives $s_{\mathrm{SP}} = C_D \cup C_E$. The remaining observations from C_1, **SR.Dough Tearing(E)** and **K.Smoothing(E)**, both in C_B, are rare events with low probability of occurrence and can be left out of S_{SP}. Similarly, only C_A has a significant sensitivity (0.43) and fallout (0.04 and 0.18) for s_{RP}. The only observation not in $C_A \cap C_2$, **K.Consistency (E)**, is also too rare an event to be informative. The signature of `RP` is thus defined as $s_{\mathrm{RP}} = C_2 \cap C_A$.

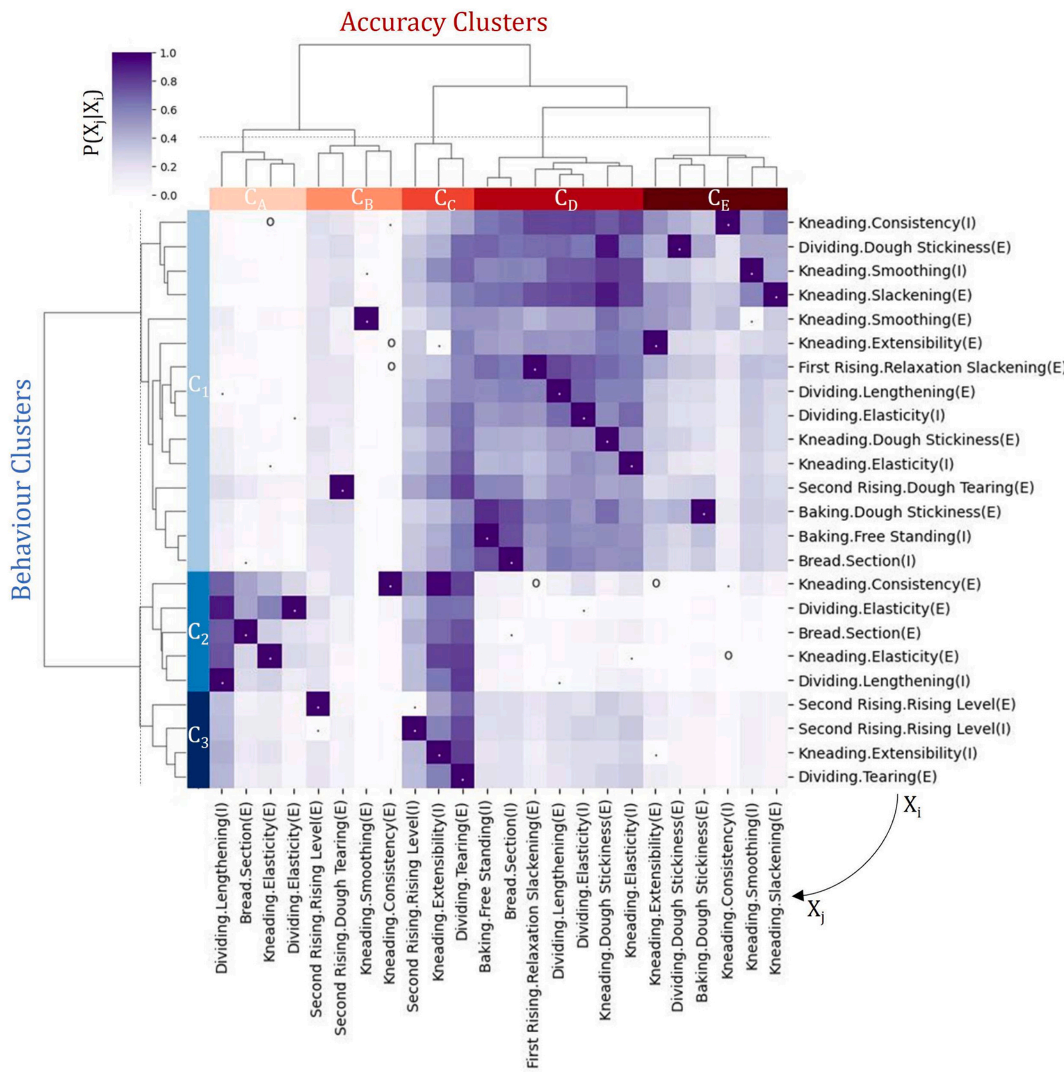

Fig. 5. Clustering over the probabilities $P(X_j|X_i)$, given two defects X_i (row) and X_j (column). Comparisons between identical or opposite defect are marked by a (.), comparisons never observed in the database are marked by a (o).

Table 3
Average probability for each intersection of behaviours and accuracy clusters (95% confidence interval computed over the set of all probabilities included in the intersection). Bolded values represent the **maximum likelihood**.

	C_A	C_B	C_C	C_D	C_E
s_{SP} =	0.04	**0.10**	0.45	**0.60**	**0.29**
C_1	(±0.01)	(±0.02)	(±0.05)	(±0.02)	(±0.03)
s_{RP} =	**0.43**	0.08	**0.64**	0.06	0.02
C_2	(±0.13)	(±0.04)	(±0.11)	(±0.02)	(±0.01)
C_3	0.18	0.05	0.52	0.22	0.07
	(±0.07)	(±0.03)	(±0.19)	(±0.02)	(±0.02)

The remaining observations of C_3 are either in C_B or C_C, two clusters whose average probabilities are not significantly different in S_{SP}, S_{RP} or C_3. In a first approach, C_3 is thus not a signature of a particular quality profile, but rather a group of observations that are not typical for a quality profile, either because they never happen (C_B) or because they happen too often (C_C). However, C_C has a high confidence interval (around 0.1) compared to the other clusters. Looking at each individual observation of C_C, it appears that **K.Extensibility(I)** has an average probability significantly higher for s_{RP} (0.75±0.17) than for s_{SP} (0.4±0.07). This could be due to distinct evaluation practices, as already noted for this assay in a previous paper (Kansou et al., 2014). Since Table 3 shows that s_{RP} has a low average sensitivity (0.43), including **K. Extensibility(I)** to s_{RP} improves the detection level of RP; while the risk of false positives mirrored by the high fall out for s_{SP} is compensated by the relatively high number of observations that composes S_{SP} (15 observations).

A signature-based scoring method is presented in the next section to validate how signatures are mapped to quality profiles.

3.3. Computing quality profiles scores

In total three quality profiles have been identified so far for wheat dough from the baking test dataset: the slackening, the resistant and the gold-standard (no defect) profiles. However the three profiles are rather extreme. In practice the sensory evaluation of an actual wheat dough is somewhere in-between the three profiles signatures, and samples could display observations belonging to distinct signatures. For example, a dough might behave mostly like gold-standard with only some slackening traits. In addition, the severity of the defect is expressed by the seven-level rating scale of the evaluation grid; an example of dough evaluation using the baking test grid is provided in Fig.6 for illustration. This is also important information to include in a diagnostic model applied to the baking test.

To account for these two aspects, three scores are computed to

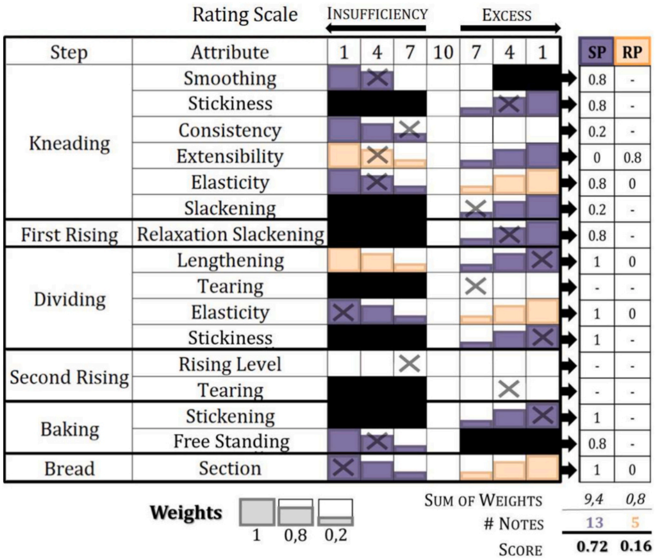

Fig. 6. Distribution of Bread Making Test observations between Slackening Profile SP and Resistant Profile RP, and an example of score calculation. The size of the patches (weight) indicates the relative importance of the intensity of a defect in the evaluation of a score. Less weight was given to 7, as experts often perceive this grade as not too problematic: when asked, they tend to classify as gold standard quality a sample with no 1 or 4 but several 7 s. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

quantify the distance between a sensory evaluation and the three signatures. The score associated to the gold standard signature (all criteria are rated 10, no defect is observed) is simply the dough grade obtained from the application of the baking test procedure. It is roughly a weighted sum giving a score between 0 and 100, which measures the growing similarity between the sensory observations and the gold standard. The dough grade involves all the attributes of the evaluation grid. Each sensory observation affects more or less the dough grade depending on the seriousness of the defects, e.g. a sticky dough is strongly penalized.

Two similar scores are developed to quantify the similarity between an evaluation and s_{SP} and s_{RP}. As shown in Fig.6, each signatures observation are assigned the following value which reflects the "severity" of the defect: 0 no defect, 0.2 mild, 0.8 strong, 1 extreme. This corresponds to the rating scale, where 10 indicates no defect, 7 mild, 4 strong and 1 extreme. The final score for s_{SP} (resp. s_{RP}) is the mean of the values assigned to all the observations of the signature. Hence 0 indicates total dissimilarity and 1 observations that certainly corresponds to the quality profile signature. A score of 0.5 indicates that about half of the observations matches the quality profile signature. In the example provided in Fig.6, the wheat produces a dough with strong defects such as **D.Lengthening(E)**, **D.Elasticity(E)**, **D.Stickiness(E)**, associated to the Slackening profile, resulting in a score of 0.72 for s_{SP}. The score of 0.16 for s_{RP} is due to the defect **K.Extensibility(I)**. However, as noted previously this defect is prone to false positives.

To obtain a more progressive qualification of the wheat in-between the three extreme profiles, four intermediary quality profiles are defined, giving a total of seven quality profiles. These were defined using expert inputs to reflect the transition between the different extreme behaviours, so that each one is a unique combination of slackening and resistant scores with respect to the dough grade. Fig. 7 shows this distribution among the quality profiles:

- **Slackening** (Sl.) and **Resistant** (Re.) profiles obtains respectively the highest slackening and resistant average scores (0.6 and 0.4) for the lowest average dough grades (respectively 25 and 60). The difference between the low dough grade value for slackening is due to the higher penalties for the observations of the slackening signature.
- **Slackening-No Defect** (Sl.-ND) and **Resistant-No Defect** (Re.-Sl.) obtains lower average slackening and resistant scores (0.35 and 0.38), for higher average dough grades (55 and 78). Compared to the slackening and resistant profiles this indicates that certain defects are less severe or missing, hence the higher dough grade.
- **No Defect-Slackening** (ND-Sl.) and **No Defect-Resistant** (ND-Re.) have an average low score (0.1 and 0.18) and high average dough grades (75 and 80). They still show a tendency towards SP or RP, but the severity is less pronounced.
- **No Defect** (ND) is the gold-standard quality profile. It has the highest possible average dough grade (90). Some observations from s_{SP} or s_{RP} may be observed, but never with a high severity (e.g. no rating equal to 1 in the evaluation grid).

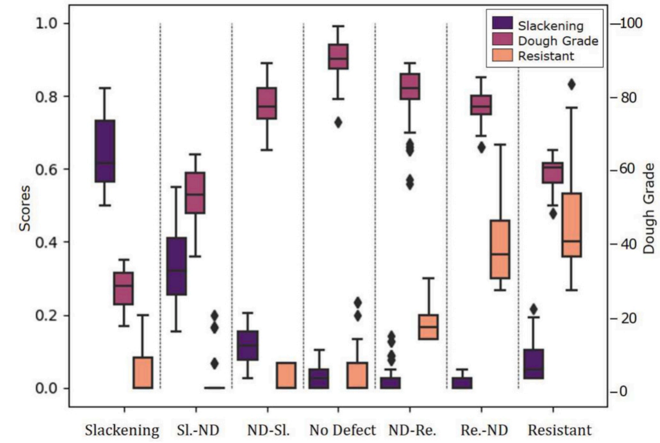

Fig. 7. Box plots of the distribution of slackening and resistant scores and dough grade across the quality profiles. Quality profiles (Slackening, Sl.-ND, …) are defined using experts inputs. Each is defined by an average slackening score, dough grade and resistant score. For instance, the Slackening profile is characterized by a mean slackening score of 0.6, a mean dough grade of 25 and a mean resistant score of 0. (11,184 samples).

3.4. *Quality profiles evaluation*

Fig.8 shows a detail of the average value of the evaluation grid attributes across the seven quality profiles defined above. This first result highlights the accuracy of some attributes over others. For example, **D. Lengthening** is significantly better associated with the quality profiles than **K.Extensibility**: this shows that, within the database, **D.Lengthening** is the most informative attributes of the evaluation grid. This analysis also shows that the seven quality profiles depict a progressive and consistent evolution of several dough properties (stickiness, elasticity, lengthening free standing, section) which demonstrates the consistency of the profiles definition and assignment to the wheat lots.

Fig.9 shows the distribution of values for the four selected technical attributes across the seven quality profiles. The figure shows that the lowest mean values of Gluten Index are associated with quality profiles with slackening tendencies. Similarly, both Ie and W mean values have a tendency to increase along with the dough's resistant behaviour: the more resistant the dough is, the higher the mean values are.

According to Migliorini et al. (Migliorini et al., 2016), a Gluten Index between 65 and 80 reflects an optimal gluten network condition. This range corresponds mainly to the ND-Sl. quality profile in Fig.9, which is slightly lower than the average value for the No Defect quality profile, which is above 80. Furthermore, prior research has demonstrated a positive correlation between the Gluten Index and Ie (Baudouin, 2012). Additionally, W has been validated as an indicator of gluten strength, as

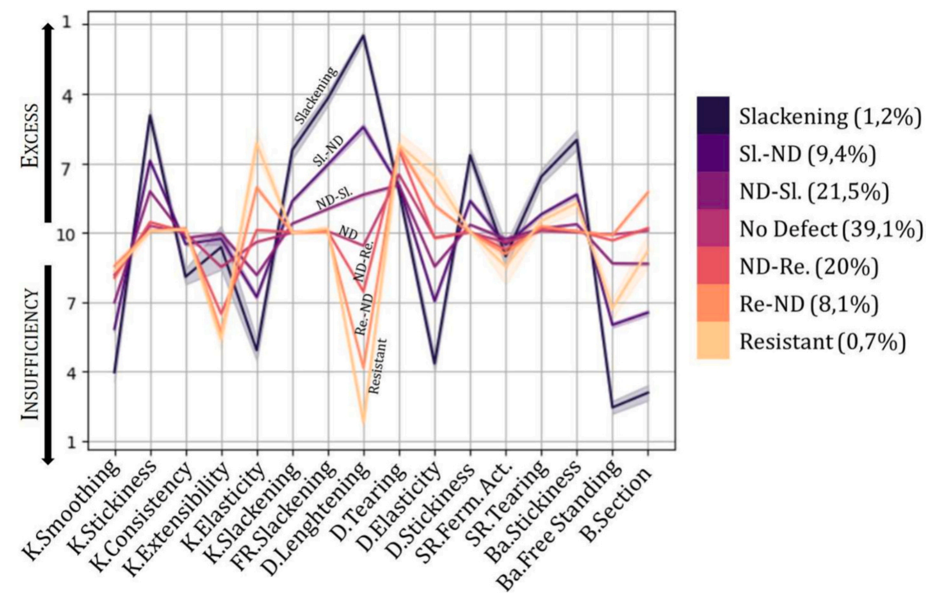

Fig. 8. Distribution of mean attributes across quality profiles. Shaded areas represent 95% confidence intervals. The percentage in the legend indicates the percentage of profiles observed in the dataset (11,184 samples).

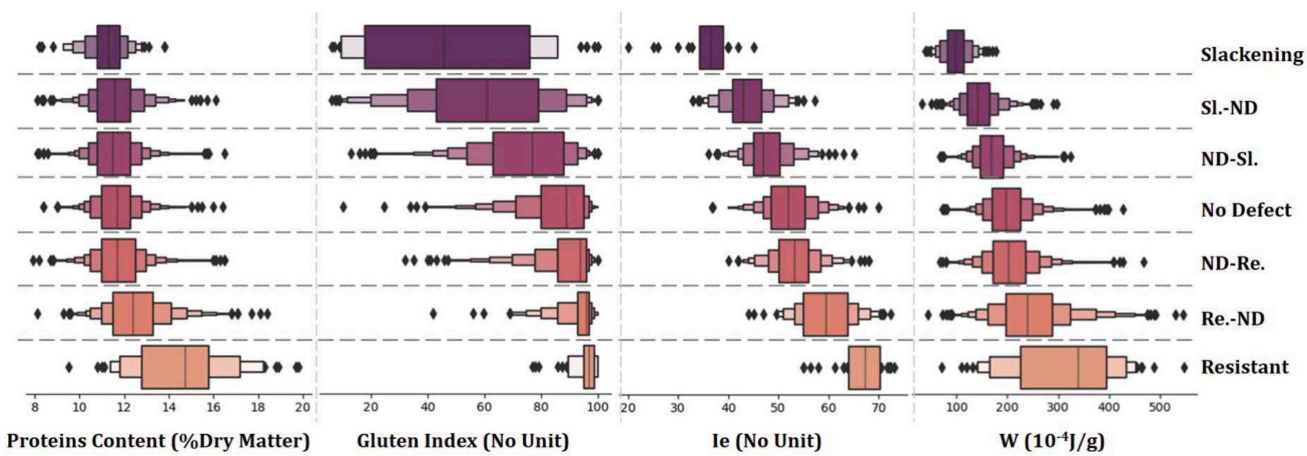

Fig. 9. Distribution of values the four technical attributes for each profile. Number of samples: Proteins Content (10566), Gluten Index (3733), Ie (1452), W (10395).

it correlates with both the quantity and quality of gluten present in the dough (Jødal & Larsen, 2021). Finally, Fig.9 shows that, apart from the Re.ND and Resistant quality profiles, the protein content is on average the same for all profiles. However, the protein content gives an indirect indication of the quantity of insoluble proteins involved in the gluten network and, unlike the other measurements, does not give an

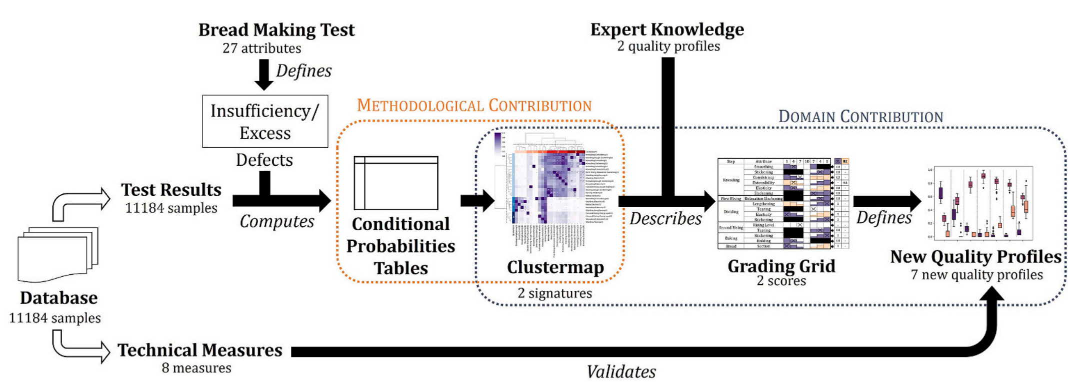

Fig. 10. Article contributions. Methodological Contribution includes the clustering of conditional probabilities approach presented in Sec.2. Conditional Probability tables are presented in Sec.2.2; clustermap in Sec.2.3. Domain Contribution includes the application to the wheat quality evaluation presented in Sec.3. Database is presented in Sec.3.1; clustermap applied to the domain in Sec.3.2; grading grid in Sec.3.3; new quality profiles in Sec.3.4.

indication of the quality of the gluten network, which can vary considerably for the same protein content value.

The quality profiles, which represent a gradation of the expert's assessment of the strength of the gluten network, correspond to the technical measures of the dough: slackening quality profiles are associated with weak gluten quality (low gluten index, low Ie and low W) and resistant quality profiles with strong gluten quality (high proteins content, high gluten index, high Ie and high W). In conclusion, this validates the signatures for identifying quality profiles and opens up new ways of defining the quality of a wheat.

4. Conclusion

This article's main contributions are summarised in Fig.10.

In this paper, we present a methodological framework for capturing and modelling relationships between sensory evaluations of food products and their interpretations in term of product physical states, called quality profiles. Based on the clustering of conditional probabilities table, the method can reveal implicit expert practices reflected by the sensory data and support the identification of the relationships between the sensory observations and quality profiles, specific to the context. In particular, in contrast to a more classical clustering method, this method compute the accuracy of the defects, i.e. the amount of information they convey. It also distinguishes rare from very frequent observed defects. For stakeholders in the food industry, this methodological contribution offers several key benefits:

- Clarification of the context of a sensory dataset by highlighting expert practices;
- Integration of the tacit knowledge of the experts about the physical interpretation of the sensory observations;
- Establishment of quality profiles that can be correlated with analytical or technological measurements;
- Evaluation of the effectiveness of sensory testing in identifying quality profiles.

In the case of wheat grain quality grading, the domain contribution is useful for.

- Formulating new quality profiles based on expert knowledge of dough behaviour;
- Linking these quality profiles to established analytical/technological measurements, which validates our hypothesis;
- Computing new features for wheat grading that can be used to feed machine learning models.

A limit of this approach is is that it depends on the representativeness of the dataset and on the clustering method chosen. Different sources of sensory data or clustering algorithms will affect the quality profiles signatures. Moreover, an untracked change in the context of the sensory evaluation (e.g. a change in the sensory evaluation process, or the introduction of a new evaluator) can hinder the clusters identification. Finally, this approach relies on expert knowledge to identify quality profiles. If a quality profile is not described in advance, then it cannot be used to analyse the clusters.

Thus, this work presents a new method for exploiting sensory data through the identification of food quality profiles. In particular, the good correlations between wheat flour quality profiles and technological measurements describing the wheat dough behaviour obtained for the real case application of the method open up interesting perspectives for development of a predictive machine learning model.

CRediT authorship contribution statement

Mélanie Munch: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Cédric Baudrit:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Hubert Chiron:** Writing – review & editing, Validation. **Benoît Méléard:** Resources, Validation. **Luc Saulnier:** Writing – review & editing, Validation, Project administration, Funding acquisition, Conceptualization. **Kamal Kansou:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was funded by the ANR Evagrain (ANR-20-CE21-0008). The authors would like to thank the ARVALIS technical institute for providing the data set and the experts from the AXIANE MEUNERIE and LIMAGRAIN milling industries for their input on the quality profiles.

References

Allais, I., Perrot, N., Curt, C., & Trystram, G. (2007a). Modelling the operator know-how to control sensory quality in traditional processes. *Journal of Food Engineering, 83*, 156–166. https://doi.org/10.1016/j.jfoodeng.2007.02.016. eFFoST 2005 Annual Meeting: Innovations in Traditional Foods.

Allais, I., Perrot, N., Curt, C., & Trystram, G. (2007b). Modelling the operator know-how to control sensory quality in traditional processes. *Journal of Food Engineering, 83*, 156–166.

Ballester, J., Patris, B., Symoneaux, R., & Valentin, D. (2008). Conceptual vs. perceptual wine spaces: Does expertise matter? *Food Quality and Preference, 19*, 267–276. https://doi.org/10.1016/j.foodqual.2007.08.001

Baudouin, F. (2012). *Influence des paramètres du procédé et des composants de la farine de blé sur la formation du réseau de gluten et son extraction*. Ph.D. thesis. Montpellier SupAgro

Baudrit, C., Sicard, M., Wuillemin, P. H., & Perrot, N. (2010). Towards a global modelling of the camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic bayesian networks. *Journal of Food Engineering, 98*, 283–293.

Birle, S., Hussein, M., & Becker, T. (2013). Fuzzy logic control and soft sensing applications in food and beverage processes. *Food Control, 29*, 254–269. https://doi.org/10.1016/j.foodcont.2012.06.011

Bourne, M. C. (Ed.). (2002). *Food texture and viscosity* (2nd ed.). Academic Press, London: Food Science and Technology. second edition ed. https://doi.org/10.1016/B978-012119062-0/50000-0.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Cordier, M. O., Travé-Massuyès, L., & Pucel, X. (2006). *Comparing diagnosability in continuous discrete-event systems, in: Safeprocess'2006, Beijing, China*. URL: https://inria.hal.science/inria-00463798.

Curt, C., Trystram, G., Nogueira-Terrones, H., & Hossenlopp, J. (2004). A method for the analysis and control of sensory properties during processing—Application to the dry sausage process. *Food Control, 15*, 341–349. URL: https://www.sciencedirect.com/science/article/pii/S0956713503000847.

Curto, B., Moreno, V., García-Esteban, J. A., Blanco, F. J., González, I., Vivar, A., & Revilla, I. (2020). Accurate prediction of sensory attributes of cheese using near-infrared spectroscopy based on artificial neural network. *Sensors, 20*, 3566. https://doi.org/10.3390/s20123566. number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Dobraszczyk, B. (2004). Wheat | dough rheology. In C. Wrigley (Ed.), *Encyclopedia of grain science* (pp. 400–416). Oxford: Elsevier. https://doi.org/10.1016/B0-12-765490-9/00003-3.

Dubois, M., Dubat, A., & Launay, B. (2008). *The AlveoConsistograph handbook* (2nd ed.). AACC International.

Hajšelová, M., & Alldrick, A. (2003). 9 - analysing wheat and flour. In S. P. Cauvain (Ed.), *Bread making* (pp. 187–199). Woodhead Publishing. Woodhead Publishing Series in Food Science, Technology and Nutrition. https://doi.org/10.1533/9781855737129.1.187.

Jødal, A. S. S., & Larsen, K. L. (2021). Investigation of the relationships between the alveograph parameters. *Scientific Reports, 11*. https://doi.org/10.1038/s41598-021-84959-3

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251. https://doi.org/10.1037/h0034747. place: US Publisher: American Psychological Association.

Kansou, K., Chiron, H., Della Valle, G., Ndiaye, A., & Roussel, P. (2014). Predicting the quality of wheat flour dough at mixing using an expert system. *Food Research International, 64*, 772–782. https://doi.org/10.1016/j.foodres.2014.08.007

Kansou, K., Laurier, W., Charalambides, M. N., Della-Valle, G., Djekic, I., Feyissa, A. H., Marra, F., Thomopoulos, R., & Bredeweg, B. (2022). Food modelling strategies and approaches for knowledge transfer. *Trends in Food Science & Technology, 120*, 363–373. https://doi.org/10.1016/j.tifs.2022.01.021

Lingle, T. R. (2011). *The coffee Cupper's handbook systematic guide to the sensory evaluation of Coffee's flavor* (4th ed.). Long Beach, California: Specialty Coffee Association of America.

Livio, J., & Hodhod, R. (2018). Ai cupper: A fuzzy expert system for sensorial evaluation of coffee bean attributes to derive quality scoring. *IEEE Transactions on Fuzzy Systems, 26*, 3418–3427.

Mavani, N. R., Ali, J. M., Othman, S., Hussain, M. A., Hashim, H., & Rahman, N. A. (2022). Application of artificial intelligence in food industry—A guideline. *Food Engineering Reviews, 14*, 134–175. https://doi.org/10.1007/s12393-021-09290-z

Migliorini, P., Spagnolo, S., Torri, L., Arnoulet, M., Lazzerini, G., & Ceccarelli, S. (2016). Agronomic and quality characteristics of old, modern and mixture wheat varieties and landraces for organic bread chain in diverse environments of northern Italy. *European Journal of Agronomy, 79*, 131–141. URL: https://www.sciencedirect.com/science/article/pii/S1161030116301009.

Ndiaye, A., Valle, G. D., & Roussel, P. (2009). Qualitative modelling of a multi-step process: The case of french breadmaking. *Expert Systems with Applications, 36*, 1020–1038. https://doi.org/10.1016/j.eswa.2007.11.006

Nunes, C. A., Ribeiro, M. N., de Carvalho, T. C., Ferreira, D. D., de Oliveira, L. L., & Pinheiro, A. C. (2023). Artificial intelligence in sensory and consumer studies of food products. *Current Opinion in Food Science, 50*, Article 101002. https://doi.org/10.1016/j.cofs.2023.101002

Perrot, N., Agioux, L., Ioannou, I., Mauris, G., Corrieu, G., & Trystram, G. (2004). Decision support system design using the operator skill to control cheese ripening—application of the fuzzy symbolic approach. *Journal of Food Engineering, 64*, 321–333.

Raufaste, E., da Silva Neves, R., & Mariné, C. (2003). Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence, 148*, 197–218. https://doi.org/10.1016/S0004-3702(03)00021-3

Raufaste, E., Eyrolle, H., & Mariné, C. (1998). Pertinence generation in radiological diagnosis: Spreading activation and the nature of expertise. *Cognitive Science, 22*, 517–546. https://doi.org/10.1207/s15516709cog2204_4

Ruan, D., & Zeng, X. (2004). *Intelligent sensory evaluation: An introduction* (pp. 1–9). Berlin Heidelberg, Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-07950-8_1

Schreiber, G., Akkermans, H., Anjewierden, A., Hoog, R., Shadbolt, N., Velde, W., & Wielinga, B. (2001). *Knowledge engineering and management - the CommonKADS methodology* (volume 24). MIT Press. https://doi.org/10.7551/mitpress/4073.001.0001

Sicard, M., Baudrit, C., Leclerc-Perlat, M. N., Wuillemin, P. H., & Perrot, N. (2011). Expert knowledge integration to model complex food processes. Application on the camembert cheese ripening process. *Expert Systems with Applications, 38*, 11804–11812. https://doi.org/10.1016/j.eswa.2011.03.068

Wooten, T. C., & Rowley, T. H. (1995). Using anthropological interview strategies to enhance knowledge acquisition. *Expert Systems with Applications, 9*, 469–482. https://doi.org/10.1016/0957-4174(95)00017-8

Zhang, H. (2004). The optimality of naive bayes. *Proc. Seventeenth Int. Florida Artif. Intell*, 562–567.