Doctoral School in

Agrifood and Environmental Sciences

Michele Ricci

Applications of advanced data analysis procedures in food quality control

Supervisor
prof. Eugenio Aprea
Tutor
prof. Flavia Gasperi

XXXV Cycle, 2023

ACKNOWLEDGEMENTS

The PhD program in Agrifood and Environmental Sciences is managed by the Center Agriculture Food Environment (C3A, University of Trento) in cooperation with the Edmund Mach Foundation (FEM).

The research was financially supported by the Research and Innovation Center of Fondazione E. Mach (as part of contracts with Trentingrana Consorzio dei Caseifici Sociali Trentini, the ADP funding by the Autonomous Province of Trento, Italy and the TRENTINGRANA project - RDP 201-202, CUP C66D17000180008).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ASCA	ANOVA Simultaneous Component Analysis
DMC	Double Milk Collection
FAO	Food and Agriculture Administration
FDA	Food And Drugs Administration
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceutical for Human Use.
LMM	Linear Mixed Models
MFA	Multiple Factors Analysis
MMC	Multiple Milk Collection
MSE	Mean Square Error
O-PLS-(DA)	Orthogonal Partial Least Squares Regression (Discriminant Analysis)
PARAFAC	Parallel Factor Analysis
PARAFASCA	Parallel Factor ANOVA simultaneous Component Analysis
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PDO	Product Designation of Origin
PLS-(DA)	Partial Least Squares Regression Discriminant Analysis
QbD	Quality by Design
SMAPE	Symmetric Mean Absolute Percentage Error
SMC	Singular Milk Collection
SPME/GC-MS	Solid Phase Micro-extraction Gas Chromatography Mass Spectrometry
VOCs	Volatile Organic Compounds

ABSTRACT

In food manufacturing, the quality control procedure is a critical activity that consists in organizing, measuring, tracking, and filing the conditions of the production process and the final product, with the goal of guaranteeing the designed quality standard. During the last 30 years, due to a mounting concern by both consumers and lawmakers, the definition of quality and the application of quality control improved drastically, and new methodologies have been developed to ensure better control of food production and to understand the effect of raw materials and the process condition on the final quality of the food product.

This thesis discusses the approaches to quality control procedures in food manufacture, focusing on the relationship between the conditions of the process and the quality profile of the final product, testing in a real-case scenario of a complex production process advanced data analysis procedures.

The statistical and analytical procedures proposed have been applied in a real case studio from Trentingrana cheese production, a dairy consortium in the northeast region of Italy producing a ripened semi-artisanal hard cheese under the Protected Denomination of Origin (PDO) of Grana Padano. The aim is developing tailored statistical procedures that infer the effect of the critical factors of production on quality properties of this PDO product considering its semi-artisanal production process and the presence of multiple confounding factors. The statistical analyses were applied to a dataset of measurements of physical, sensory, and chemical properties collected on cheese wheels sampled systematically to represent the variability of the production of the Trentingrana wheels over two years of production.

In the first introductory chapter, after a review of the different definitions of quality, the most important quality parameters for a food product and the standard measurement techniques adopted in

quality control are presented. Then, in the chapter 2, the standard procedures of data analysis are reviewed, as well as the new approaches derived from the context of the foodomic sciences and machine learning models for the analysis of quality control data in food manufacturing.

Two implemented and tested practical statistical procedures in the context of the Trentingrana consortium are reported: the results are discussed according to the objectives of the quality control process, the type of data, and the organization of food production. In the first case, reported in chapter 3, Linear Mixed Model ANOVA Simultaneous Component Analysis (LMM-ASCA) was developed to investigate the effect of the dairy factory, the bimester of production, and the variability within a cheese wheel using colorimetric and textural measurements. In the second case, reported in chapter 4, a standard ASCA model with the addition of a blocking factor to include systematic error was developed to investigate the relationship between the dairy factory and bimester of production and the volatile organic compounds (VOCs) profile of Trentingrana cheese wheels.

In addition, in chapter 5, an approach to relate physical measurements on Trentingrana samples with sensory evaluations of texture by a trained panel is presented. The objective of this procedure is to incorporate the quality control procedure information from different quality parameters. The development of the Partial Least Squares (PLS) predictive model, its validation, and the evaluation of its performances are discussed.

In the last section (chapter 6), the development of an image analysis procedure to measure the visual quality of the rind thickness of cheese wheels is reported, comparing the performances of two different algorithms.

The data analysis tools proposed in this thesis have been proved to be useful for exploring, inferring, and plotting the process quality properties and suitable for analyzing complex and unbalanced experimental designs. Furthermore, the data analysis procedures proposed improve quality control activity both at the process level and at the product level, increasing the information that is possible to extract from the measurement collected in a context where standard statistical approaches cannot infer significant information.

CHAPTER 1. INTRODUCTION

The present chapter reports a brief discussion of the definition of food quality, reporting how it was defined throughout the last century and why this concept is important in food manufacturing. In the rest of the chapter, the definition of quality control procedures in food manufacture is discussed, focusing on the quality protocol called "Quality by Design". The objectives and the procedures of this protocol are reported, and its applications are discussed. In the last section, the case study is presented. The case consists of the development of systematic quality control procedures to estimate the quality profile of hard-seasoned cheese wheels produced by the Trentingrana consortium, a consortium of dairy factories producing a grana-type cheese under the Product Designation of Origin (PDO) of Grana Padano. In the present section, there will be a brief excursus on the context of the manufacturing sector of hard seasoned cheese production in Italy and of Trentingrana consortium.

1.1 Quality control in food production

1.1.1 Definition of food quality

In the European agrifood sector, food quality is a determining factor in consumer choices and food intake (*Eurobarometer, 2014*).

Food quality represents the sum of all properties and assessable attributes of a food item, and it is defined by different parameters according to the approach adopted toward the evaluation of the food product. It is influenced by a wide range of situational and contextual factors.

The meaning of the term food quality has not always been the same, and the shifts in the interpretation of this term are aligned with the most important concerns for consumers and stakeholders in the food chain.

The first reference for a suitable definition of quality is the one proposed by FAO (2003), which states that:" Food safety refers to all those hazards […] that may make food injurious to the health of the consumer. […] Quality includes all other attributes influencing a product's value to the consumer. This includes negative attributes such as spoilage, contamination with filth, discoloration, off-odors, and positive attributes such as the food's origin, color, flavor, texture, and processing method.". This definition considers food quality adopting the point of view of food and nutrition sciences.

Klaus Grunert in its review (Grunert, 2005) presented another definition of food quality that adopts a wider perspective, including in the evaluation ~~of~~ the approach of sensory and consumer sciences. According to this interpretation, food quality is defined both in objective and subjective ways: one refers to the properties that are measurable with instrumental procedures and the other is related to the estimation of the perceptions of the user.

This approach represents a shift of perspective from the previous definition of quality, including the evaluation of food products adopting methodologies from different scientific fields, such as sensory science, consumer science, and behavioral psychology.

For the specific case of sensory quality control, the application of a systematic approach for its estimation in food industries started in the 1990s, proposing more scientific and repeatable procedures (Muñoz, 2002). Sensory sciences applied to food analysis is an established method for understanding the intrinsic properties of the food product and gaining insight into consumers' opinions (Yang & Lee, 2019). This definition of food quality overcomes the evaluation of the safety of food products as the only objective of the properties of the final product and adds more points of view on the product and the production process, defining multiple aspects that determine quality.

A more explanatory definition of food quality was reported by Leitzmann (Leitzmann, 1993), who states that to be estimated, food quality requires the adoption of multiple approaches to evaluate the different parameters that affect the choice of food consumption, from the contents of nutrients to the sensory properties, also including the analysis of the economic and the ecological impact of food production. This wider approach defines food quality as the combination of different dimensions, each having its quality parameters, that may be correlated or not. These dimensions are related to physical properties, subjective perception (sensory, preferences, psychological needs), and socio-environmental impact (economic, cultural, and ecological importance).

A functional summarization of this definition of quality as a collection of many parameters estimated from different disciplines was presented by Giusti and coauthors (Giusti *et al.* 2008), adopting the term Total Food Quality (TFQ), as a summary of the evaluations from multiple scientific disciplines, which are summarized in table 1.

Table 1 – Summary of quality parameters considered in the total food quality approach and their definition.

PARAMETER	**DEFINITION**
SENSORY	*Color, appearance, texture, juiciness, taste, astringency and aroma*
SAFETY	*Presence of toxic compounds normally contained in foods, contaminants, mycotoxins, pathogen, and toxigenic microorganisms*
NUTRITIONAL VALUE	*Calories content and macronutrients composition, as well as non-nutrients with high biological activity, compounds from technological processes, digestibility, and bioavailability*
FUNCTIONAL PROPERTIES	*Ease of use of several ingredients used for processing and transformation*
SERVICE AND STABILITY	*Resistance to rapid deterioration (processing, storage, transportation, and shelf-life conditions)*
HEALTHINESS	*Capacity of some food components to exert beneficial effects on consumers' health (e.g. probiotics, vitamins)*
PSYCHOLOGICAL	*Convenience, price, ease of use, novelty, psycho-active effects of food*

1.1.2 *Quality protocols*

In the agrifood industry, the meaning of quality determines most of the goals for the industrial activity in terms of product and process standards. For example, defining quality as compliance with safety standards (as the first definition proposed) corresponds to a producer setting the goal that the food products comply with safety standards. To achieve this, it is necessary to develop a production process that ensures the absence of health risks and a control procedure to monitor product safety.

For food manufacturers, the adoption of the Total Food Quality approach corresponds to the development of a production process that is optimized to comply with a wider range of quality standards.

To comply with those quality standards many quality procedures have been developed over the years. A quality procedure is a standardized protocol proposed for agri-food industries that consists of a series of activities and recordings that are necessary to reach the desired standards.

There are many categories of quality protocol, according to their legal significance, they can be briefly summarized in mandatory protocols such as (HACCP, HARCP, Fielding *et al.* 2011), food quality certifications from International Standard Organizations (ISO 9001, ISO 22000, Psomas & Fotopoulos 2010), European Union quality label (PDO, PGI, TSG, Grunert & Aachmann 2016), and Private Label Certifications (GlobalGAP, BRC, Tey *et al.* 2016).

This thesis focuses on the quality protocol called Quality by Design, because of its compatibility with the concept of total food quality and because it allows estimating properly the relationships between the production process in food manufacturers and the properties of the final product.

1.2 Quality by Design

1.2.1 Definition

The Quality by Design approach has been defined in the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceutical for Human Use (shortened in ICH) Q9 guideline as ‘a systematic approach to quality development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management’ (ICH, 2023).

This approach consists of several steps of process evaluation, identification of key parameters, development of systematic control procedures, and monitoring of the performance of the control procedures (Rathore, 2014, Rathore & Winkle, 2009). It is neither always appropriate nor always necessary to use a formal risk management process (using recognized tools and/ or internal procedures e.g., standard operating procedures). The use of informal risk management processes (using empirical tools and/ or internal procedures) can also be considered acceptable (ICH, 2023).

The Quality by Design approach saves time and resources of post-manufacturing quality testing to assure that the product is compliant. In other words, the quality is ‘built into’ the process, as opposed to ‘tested on’ the products.

This approach to the quality control procedures is based on the idea that the improvement of the quality properties of the food product must be associated with the conditions of the process by deliberate design, analyzing the effect of the parameters on the final properties of the product.

It is possible to guarantee the desired quality level by monitoring the conditions of the production process to maintain the optimal parameters once the relationship between the quality profile of the final product and the condition of the production process is inferred.

To gain this result, it is necessary to develop a process analytical protocol capable of measuring as fast as possible the conditions of the process and alerting when the values measured are outside the optimal level. This procedure is defined Process analytical technology (PAT).

The concept of quality by design was determined by Food and Drug Administration in 2004 to encourage in the manufacturing sector this new approach, defining a set of scientific principles and tools supporting innovation and a strategy for regulatory implementation that would encourage innovation (FDA, 2004).

To implement it practically in the agrifood production context, it has been summarized in seven points (Savitha & Devi, 2022):

1. Identify quality target product profile (QTPP):

The Quality target product profile is, as defined in ICH guidelines, “A prospective summary of the quality characteristics of a drug product that ideally will be achieved to ensure the desired quality, taking into account safety and efficacy of the drug product”. In this step the quality profile of the product is defined, according to the quality approach adopted by the producer.

2. Identify Critical quality attributes (CQA):

Once QTPP has been identified, the next step is to identify the relevant CQA. A CQA has been defined as “a physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality”. Identification of

CQA is performed through risk assessment as per the ICH guidance Q9 (ICH, 2023). Prior product knowledge, data from literature references, and laboratory tests are needed to establish the CQA for the given product. This step consists in defining measurable characteristics of the product matching the quality standards defined.

3. Defining the product design space

The product design space is a multidimensional matrix containing the realistic combinations of CQA values intervals based on the data from the laboratory, consumer studies on the product, published literature, and process capabilities considering the variability observed in the manufactured lots (Rathore *et al.* 2010).

4. Defining the process design space

Once the product design space has been defined, process characterization studies are carried out to estimate the relevant variables in process conditions for final product quality. Then, the realistic combinations of process parameters are estimated and reported. The available combinations are reported considering the process capabilities and the properties requested in the final product (van Hoek *et al.* 2009).

5. Defining a control strategy

Control strategy has been defined as "a planned set of controls, derived from current product and process understanding that assures process performance and product quality" (FDA 2004). The control strategy in the Quality by Design paradigm is established via risk assessment that integrates standard safety procedures such as HACCP with the critical values estimated for CQA and process parameters. According to the food manufacturer considered and the quality parameters requested the control strategy could include procedural controls, in-process controls, lot release testing, process monitoring, characterization testing, comparability testing, and stability testing.

6. Process validation and regulatory filings

Process validation consists in scaling the defined process procedure to factory conditions, to demonstrate that the process can deliver a product in the desired quality parameters operating in the established design space and that the reduced and/or pilot scale systems are representative of the manufacturing scale process. Regulatory filing procedures include the final acceptable ranges for all key process parameters in addition to a more restricted operating space. From the regulatory standpoint, a key parameter is an adjustable parameter of the process that, when maintained within a narrow range, ensures operational reliability whereas a critical parameter is an adjustable parameter of the process that should be maintained within a narrow range to not affect a critical product quality attribute (Parenteral Drug Association, 2012).

7. Process monitoring, life-cycle management, and continuous improvement

Process monitoring is required during the product's lifecycle to keep track of changes and manufacturing deviations to avoid declines in product quality. To monitor the robustness of the quality system is necessary to consider the following four elements: process performance/product-quality monitoring; preventive/corrective action; management of changes in the process; management review of process performance and product quality.

1.2.2 Development of Quality by Design approach in food manufacture

The adoption of a Quality by Design protocol in a food manufacturing context is analogous to the approach of other protocols. Because Quality by Design does not correspond to an official commercial label, such as IFS, BRC, GlobalGAP or International Standard Organization certifications such as ISO 22000 or ISO 9001, its implementation consists of internal activities that do not require to be confirmed by an external audit. Rather, QbD activities are a code for an efficient collection and interpretation of data for the estimation of the effect of the process conditions on the properties of the final product (Moran *et al.* 2017).

Many measurement protocols adopted for the evaluations of the process in Quality by Design are already collected and filled according to mandatory HACCP procedures (Quinn & Marriott 2002). Usually, standard hazard analysis does not include enough measurements to estimate the quality profile of the final product.

Another difference between HACCP and Quality by Design is the viewpoint shift: HACCP protocol is focused on hazard monitoring and a quality control approach that aims to avoid risks for the consumer and to guarantee the absence of non-conformities in the product. Consequently, the procedure controls critical points of the process to ensure that critical levels do not exit from a critical area. Differently, Quality by Design aims to optimize the conditions of the process to obtain the best available properties of the final product, deploying a systematic measurement procedure to estimate the quality profile of the final product and the key variables of the production process.

To estimate the quality profile according to all the different quality dimensions (according to the Total Food Quality approach) it is necessary to develop a systematic sampling procedure of the final product to infer the results of the measurements to the real-scale production process. A well-planned sampling procedure must consider all the systematic sources of variability that may affect the product quality and the natural variability of the product itself: the first issue requires defining sampling from different levels of process factors, and the other to estimate the number of samples needed to reach enough statistical power. Those topics will be discussed in chapter 2. Another critical issue in the sampling procedure is the handling of the food samples: to ensure that the measurement would not be biased by spoilage, it is necessary to define a regular protocol for storing food samples to avoid unwanted modifications in chemical properties and especially in the sensory profile.

To estimate the Total Food Quality of a product it is necessary to define analytical procedures adopting different scientific approaches according to the quality dimension measured.

The main scientific approaches are physical, chemical, biological, sensory, and consumer analysis. Each consists of measurement techniques based on different principles, the adoption of different instruments, and the development of different statistical procedures (Grunert 2005).

A summary of the quality parameter of each analytical approach is reported in table 2.

Table 2: *Summary of analytical approaches adopted to estimate quality parameters in food.*

Analytical approaches	Quality parameter related	Main analytical techniques
Physical	*Color, structure*	*Cairone et al. 2020, Banks 2007, Drake et al. 1999*
Chemical	*Gross composition, volatile compounds, micronutrients, contaminants*	*Qin et al. 2022*
Biological	*Microbial count, enzymatic activity*	*Erkmen 2022*
Sensory	*Sensory profile, defects*	*Drake 2007*
Consumer	*Preference, emotion, affection*	*Ruiz Capillas et al. 2021*

To develop a Quality by Design procedure, the safety-related parameters that are usually measured to ensure many standard HACCP procedures can be also related to sensory quality parameters (Andrews et al. 2021). In table 3 are reported many examples of measurements related to standard safety measurements that are also related to standard measurements.

Table 3: *Summary of analytical measurements that affect quality at multiple levels*.

Property measured	Implication		Reference
	Primary	Sensory	
Microbial counts	*Food safety, nutritional value*	*Atypical flavour, etc. (rotten)*	*McAuliffe et al.(2019)*
pH/acidity	*Functionality, composition*	*Sourness*	*Barnes et al. (1991)*
Peroxide value (of fats)	*Nutritional value*	*Atypical flavour, etc. (oxidised)*	*Mehta et al.(2018)*
Proximate analyses	*Nutritional value, cost*	*Atypical texture, taste, etc.*	*Barnes et al. (1991)*
Free fatty acids	*Functionality (of fat)*	*Atypical flavour (lipolytic)*	*Mannion et al. (2016)*
Rheological properties	*Functionality*	*Texture, etc.*	*Laguna et al.(2017)*
Tribology	*Food quality*	*Mouthfeel*	*Sarkar & Krop (2019)*
'Thermal' properties (SFC, DSC, etc.)	*Functionality*	*Mouthfeel (e.g., ice cream melt)*	*Roland et al. (1999)*
Particulates (particle size, insoluble particles, foreign particles, etc.	*Food safety, food quality*	*Texture (e.g., grittiness, creaminess)*	*Shewan et al. (2020)*

It is worth saying, even if it is not discussed in the present thesis, that also the quality dimension related to the influence of socio-economic features such as price, claims, availability, needs to be studied according to the principles of marketing related disciplines.

1.3 Presentation of the case study

1.3.1 The importance of hard-seasoned cheese in Italy

In the Italian market, the production of hard-seasoned cheese is a critical sector of the agri-food chain. The average yearly production of DOP hard-seasoned cheese in Italy is 358.050 tonnes, and 41.07% of these were exported (CLAL 2023a). During the last two years 2020 and 2021, the trade balance for dairy products registered a positive score between the importation and exportation of dairy products. DOP hard seasoned cheese has an important part in this trade balance, with an overall value for exportation of 1143.3 million in 2021 (increasing of 10.9% from the previous year), and it is the ninth category of product for the value of exportations (CREA 2022).

1.3.2 Trentingrana Consortium

Trentingrana cheese is an extra-hard-seasoned cheese type with an Italian Protected Designation of Origin (PDO) falling under the Grana Padano PDO (Eur lex. 1996), and it is produced by dairy factories located in the Autonomous Province of Trento in the north-east region of Italy. In 2021, the Trentingrana consortium produced 126.781 cheese wheels, an increase of 2.0% compared to the previous year (CLAL 2023b).

The Trentingrana trademark, embossed on the wheel near the “Grana Padano” label, emphasizes the distinctive properties of this cheese (Endrizzi et al. 2013) and its production process that has many specific differences from the product specification of Grana Padano: the use of raw cow milk only from livestock on mountain terrains in a delimited area (Autonomous the Province of Trento, Northeast Italy), the application of restricted cattle feeding, and the removal of lysozyme and silage from the cow’s feeding (MiPAAF, 2022).

1.3.3 Production Process of Trentingrana cheese

The Trentingrana consortium is a second-level consortium formed by 15 dairy factories, each of which is organized as a primary level consortium formed by the farmers producing milk.

The production process of Trentingrana cheese consists of several steps:

1. Milk is produced in farms by cows bred according to official guidelines in farms located in the Trentino region.
2. Milk is collected and brought to the dairy, the milk may be brought all in the evening (single milk collection procedure) or partially in the evening and partially in the morning after (double milk collection procedure).
3. During the night, milk is skimmed adopting the gravity separation procedure, and skimmed milk is moved to a separate repository.
4. After this, milk is placed in specific copper vats, that are heated at a temperature between 30 and 36 °C and added with rennet for a duration from 7 to 12 minutes. The presence of rennet will start the coagulation procedure in the vat of the milk while it is agitated, and operators provide to spike the milk adopting specific instruments to make sure that the curd will coagulate in flakes of the dimension of rice.
5. After the coagulation procedure is complete, the curd is cooked at a temperature ranging from 52 to 57 °C degrees during an amount of time from 15 to 27 minutes.
6. The curd granules are left to rest in the copper cauldron, immersed in the whey, for a maximum of 70 minutes from the end of the heating phase, so that they aggregate, to form a compact mass.
7. Using a sort of wooden shovel ("*pala*") and a linen cloth ("*schiavino*"), the curd mass is raised from the bottom of the cauldron and cut into two equal parts, to create two twin wheels.
8. Each of the two wheels is removed from the copper cauldron, wrapped in linen cloths, and placed on a shelf ("*spersola*").
9. Each wheel is placed into a special mould ("*fascera*") made of suitable plastic material. A heavy object of the same material is then placed on top of the cheese to place pressure.

10. After about 12 hours, a piece of plastic engraved with the Marks of Origins is inserted.
11. Then a casein plate, with a specific ID code is placed on the top face of the wheel, this is crucial when identifying the traceability of each wheel.
12. Two days later, the process of salting ("*salatura*") starts: the cheese wheels are soaked in brine, a solution of water and salt. This step can take from 14 to 30 days, depending on the saline solution and the size of the wheel.
13. Once the salting is finished, the wheels are taken into a "hot room" ("*camera calda*") where they will dry for a few hours.
14. Finally, the cheese wheels will be taken to a specific maturing warehouse, where they will be left to age for a minimum of 9 months.
15. After 9 months, the cheese wheels are moved from the dairy factory to another repository of the consortium where they spent the last part of the ripening phase, which generally last 7-9 months.

This production process presents several critical production points (Suherman *et al.* 2021), which are reported there:

- Milk collection inside the dairy factory: microbial activity, enzymatic activity, and gross composition of the raw milk.
- Milk coagulation in the vat: time and temperature of coagulation and cooking phases.
- Breaking of the curd: lowering speed of the pH of the curd.
- Salting: the brine solution must maintain saturation.
- Ripening: the condition of temperature and humidity needs to be maintained at acceptable levels.
- Final ripening: after 3 months, it is necessary to check if there are no present defects of holes or cracks inside the cheese wheel due to unwanted microbial activity. To control this a

traditional procedure called “*battitura*” is adopted: trained operators listen to the sound of cheese wheels after hitting them with a small hammer. According to the sound, it is estimated if there are present cracks or gas bubbles inside the cheese wheels.

1.3.4. Critical Issues of the production process

In the present section are summarized the features of the Trentingrana production process that needs to be considered to develop a quality control procedure that represents the overall production process. All the issues reported are relevant sources of variability that affect the final properties of the cheese or causes the large variability between each cheese wheel.

1.3.4.1. Semi artisanal process

A semi-artisanal process is food manufacture that is not completely standardized, and part of the operative decision relies on the experience of the operators working there. In Trentingrana cheese production each dairy factory has important sources of variability of the quality of the final product: the composition of the milk may vary significantly due to the seasonal effect on cows’ lactation; each cheese wheel may vary due to the variations in the conditions of time and temperature during coagulation, cooking and rest phase in the copper vat.

The preparation and the addition of the rennet in the vats is done manually by the operators, and the correct development of the coagulation procedure and the development of the textural properties of the cheese are influenced by the handling of the curd, which often is still done by hand, or adopting simple mechanical devices.

Furthermore, variations in the heating procedure affect the microbial growth, the occurrence of Maillard’s reaction, and the water dispersion inside the cheese wheels. Furthermore, during the ripening phase, it is necessary to maintain the right environmental conditions of heat and humidity, but it is not possible to monitor the exact microbial activity inside the cheese wheel, nor to intervene to change it.

Overall, the production process has got some steps that can't be monitored properly because of the intrinsic properties of the food product and the production process. There are many sources of variation that may cause differences between the sensory, physical, and chemical properties of cheese wheels produced in similar conditions.

1.3.4.2. Structure of the production process

The cheese wheels labeled with Trentingrana mark are produced in 15 different dairy factories from different areas of Trentino region, in the north-east area of Italy.

Each dairy factory operates inside the regulation of the consortium doing all these steps of the production process, but the product specification allows producers to collect milk from different farms that differ according to cow's breed, altitude, and use of unifeed mixer wagons or traditional feeding procedures (Bittante *et al.*, 2011). Additionally, the product specification allows applying slight changes in rennet, whey starters, and heating/storing machinery used in the dairy factory. Altogether, this can reasonably affect the peculiar physical and sensory properties of the final product (Ricci *et al.*, 2022).

Only the final steps of the ripening phase are completed in a comprehensive repository for the whole consortium.

1.3.4.3. Discontinuous process

The production of hard-seasoned cheese is discontinuous, which means that the raw milk is treated in regular intervals to recreate the final product, not in a continuous production process.

The discontinuity is then exacerbated by the presence of a long ripening phase that does not permit to intervene in a critical step for the correct development of the final physical and sensory characteristics of the product (Khattab *et al.* 2019).

The discontinuous structure of the process affects the quality control procedure in the following ways: greater storage space is needed for in-between production stages (such as curing and ripening), non-conformities in critical control points lead to greater waste and production cost because most of the defects of grana cheese can be detected only in the final stages of the production process. There are also practical issues as increased employee downtime due to waiting between processes and meticulous quality control, improperly planned batch processes can lead to bottlenecks that limit production.

1.4. Development of quality control procedure

Because of the issues of the agrifood context, monitoring the quality of a product from a discontinuous production process dislocated in many different small plants, with many uncontrolled sources of variability, requires a quality control procedure designed specifically to deal with those issues.

The Quality by Design approach needs to be designed carefully considering the estimations done during the previous steps of the projects. The quality control procedure needs to be designed considering the quality traits of interest, the properties of the final product, and the structure of the production process, to develop a representative procedure to collect useful information on the properties of the product.

The case study of Trentingrana cheese represents an interesting example of a quality control procedure developed according to the features of the product and the process and can be presented as the application of a tailored Quality by Design approach to understand the relationship between the condition of the process and the quality properties of the food product.

1.4.1. Sampling procedure

Every 2 months, 1 to 3 cheese wheels were randomly sampled from the 18-month-ripened “first-quality” wheels produced by each dairy of the Trentingrana consortium during the considered sampling period of 2 years (Endrizzi *et al.* 2013, Bittante *et al.* 2011).

The number of cheese wheels sampled from each dairy factory was determined according to its volume of production during the two months: one wheel for each dairy delivering up to 1000 wheels, two wheels for 1001 to 1500 delivered, and three for more than 1500 delivered cheese wheels.

Due to the internal organization procedures of the Trentingrana consortium, the cheese wheels sampled during the last couple of months had a ripening period of three weeks less than the others.

All the measurements were acquired on a weekly basis: a subset of 6 cheese wheels was brought from the storehouse to the laboratory of the Edmund Mach Foundation, where each wheel was opened and visually evaluated by a panel of experts. Then, each cheese wheel was portioned. One portion was evaluated by the quality control panel of the Trentingrana consortium, and the other one was directed to instrumental analysis.

A balanced design of cheese wheels from different dairies in each session day was introduced to balance the data collection procedure, avoiding the effect of the session on quality evaluation. The cheese wheels from the dairies that produced a lower amount of first-class wheels due to the insurgence of defects in the previous semester were submitted to the expert panel before the others, to collect first data from the most problematic sections of the production process.

CHAPTER 2. DATA STRUCTURES AND STRATEGIES FOR DATA ANALYSIS

In the present chapter the procedures to manage data from quality control processes are reported. The first step is the definition of an experimental design. Then it is necessary to report an analysis of the factors that affect the conditions of the production process and the validation procedure of the hypothesis related to these factors. In the last section of this chapter, a list of advanced statistical models and machine learning algorithms that could be adopted to overcome these issues is reported.

2.1 Experimental Designs in quality control

The approach of the design of experiments (DOE) was formalized between the second and third decades of the twentieth century (Fisher, 1935) but has been applied in food manufacture for quality control procedures involving also sensory parameters on a larger scale at the beginning of the twenty-first century. The aim of this procedure is to select the optimal number of measurements necessary to estimate the effect of one or more input variables and their interaction on the final conditions if there could exist a causal relationship between them (Montgomery, 2013).

Furthermore, the experimental design requires to be designed according to the objective of the research. In the context of quality control procedure, it is very important to choose a design that represents the overall variability of the values and that reports the variability associated with the data considering all the issues reported previously. The most common experimental designs are factorial designs, fractional factorial designs, factorial design containing blocking factors, split-plot designs, nested designs, and multi-way design.

2.1.1 Factorial Design

The factorial design is the basic experimental design, consisting of n^k observation applied, where k is the number of factors present in the experimental design, and n is the number of levels in the experimental design if each factor has got the same number of levels. This design is widely used in experiments involving two or more factors where it is necessary to study both the effect of factors and their interaction.

This kind of study is particularly useful when it is necessary to analyze the effect of many different factors at the same time to identify if there is an overall effect of the considered factors. In this phase is preferable to adopt a low number of levels for each factor, to not collect an excessive number of measurements when there is not enough information available to estimate whether some factors are significant.

This experimental procedure can detect the presence of interaction and avoiding misleading conclusions due to the presence of interactions if all the possible combinations are not considered. The main downside of this experimental design is that it requires a vast number of measurements, and consequently a vast amount of time and resources to collect all the information. For example, an experiment with 6 factors, each of them having 2 levels, requires $2^6= 64$ combinations, each of them with at least two repetitions. This amount of measurement could easily outrun the resources available in most experiments, because of the large number of measurements requested.

2.1.2 Fractional Factorial Design

To handle a reasonable amount of measurement and consider the hypothesis that the most important effects are related to singular factors and second-level interactions, the construction of a specific experimental design, called **fractional factorial design**, is proposed.

This procedure is the most widely adopted for product development, process improvement, and experimentation in industrial contexts because it allows a functional procedure for screening factors for large effects by adopting the minimum amount of measurement available.

The experimental design requires n^{k-1} combinations, which are the one-half fraction of the measurement estimated by the complete factorial design. The combinations are chosen to adopt a generator procedure that allows the experimenter to estimate the most effects and interactions with a smaller number of measures.

2.1.3 Factorial designs containing block factor

In experimental design applied to an industrial context it is often necessary to include process-related factors that could or could not affect the overall response of the other factors considered in the experiment. To include those measurements properly in the experimental procedure, it is necessary to adopt a blocking factor structure for the factorial design, which means that the potential confounding factor is integrated with the experiment as a fixed blocking factor and the construction of the experimental design needs to be planned accordingly.

The first procedure consists in creating many sub-experimental designs similar in each blocking factor, which could be a reasonable solution if the experimenter knows that there are important differences between each block. This procedure is mandatory for every control related to the HACCP procedure for safety control of food manufacturing in European countries.

If the analyses are not critical, and the blocks considered are not too different, it is possible to include the blocking factor in the experimental design, applying inside each block a different set of n^{k-2} combinations of measurements, to make sure that considering every block together every combination is applied. In this way, both the information related to the effect of the factors and the information related to the effect of the blocking factor can be estimated.

2.1.4 Nested design

To include nested factors in an experimental design it is necessary to develop a hierarchical experimental design, which consists of an experimental design optimized to identify the major source of variability considering the structure of data. The procedure to optimize the collection of data is called general m-stage nested design and consists in estimating each variable nested inside another variable as part of the experimental design without including the interactions term in the estimation of all the combinations, due to the structure of the experiment. Hence, the nesting factors cannot interact with the levels nested in the other levels, differently from the estimations adopted in designs adopting blocking factors.

If in an experimental design, both nested and factorial factors are present, it is possible to apply a factorial experimental design inside each nesting factor.

2.1.5 Split-plot design

Split-plot designs are constructed to extrapolate the largest amount of information from factorial experiments which cannot apply complete randomization of the order of the measurements due to practical reasons. This could be related to practical issues related to the organization of the plant or the presence of seasonal effects.

This experimental design consists of estimating the main batches representing one factor, measuring multiple times, and then changing the condition of the same levels according to the same order of levels for the other factors. To explain more clearly, the split-plot design could consist in preparing of the same batch of raw material and treating different portions of it in different time conditions, instead of repeating the whole procedure for each interaction, or taking measurements in the same fields in a different part of the year, instead of randomizing the fields and applying a sampling procedure during different years.

2.1.6 Multi-way design

Multi-way data are characterized by several sets of variables that are measured in a crossed fashion. Chemical examples could be fluorescence emission spectra measured at several excitation wavelengths for several samples, fluorescence lifetime measured at several excitation and emission wavelengths, or any kind of spectrum measured chromatographically for several samples. Determining such variables will give rise to three-way data, i.e., the data can be arranged in a cube instead of a matrix as in standard multivariate data sets (Coppi 1994).

Differently from two-way multi factors datasets, multi-way datasets contain increasing and decreasing trends due for example in peak structures in spectral data and to the internal correlation structure of time series analysis, where two values near in time cannot be too different and usually are differentiated.

2.2 Factors affecting data structure

Data collected by Quality by Design procedures need to be collected to be representative of the structure of the production process. To be sure of that, it is necessary to collect enough measurements from each combination of process variables.

The experimental design needs to be built to be as representative as possible of the process analyzed. Many critical issues that characterize the quality control procedures from an industrial process that need to be included in the experimental design structure are reported.

2.2.1 Type of variable in project design space

To create a proper sampling procedure, it is necessary to know which are the variables of interest and their levels. According to the Quality by Design procedure, this corresponds to point 4 of the procedure: "defining the process design space". The variables of interest can be of two types:

- **Condition related**: These are variables that affect directly the desired and undesired reactions occurring in food products during their transformation. Those parameters are continuous, such as the conditions of the time and temperature during the heating process, or the increased rate of the microbial population during fermentation.
- **Process Related**: These variables are directly related to the physical structure of the production process and to factors affecting the process. They are related to practical issues of the organization of the production process, such as the production plant and the machinery used, the season, and the production batch. Those factors are prevalently discrete, and they need to be included in the model because they represent random sources of variability that cannot be controlled directly.

To build a reliable experimental design and consequently a reliable sampling procedure, it is important to properly address these two different kinds of sources of variability. Generally physicochemical related factors can be treated as fixed factors, as they are defined in Montgomery (2013), and in the construction of an experimental design they need to be addressed in that way, recognizing the levels of interest and how the experimental design can be arranged consequently, for example selecting which of all the possible interactions needs to be observed from time to time. On the other side, process related factors can be intended like blocking factors or random factors, and considered accordingly in the estimation of the experimental design, because they are uncontrolled sources of variations that need to be included in the measurements to understand properly the condition of the process.

Obviously, not including process related variables in the estimation of the sampling procedure can only lead to an incomplete and not representative sampling procedure, and consequently to an impossible detection of the defects in the products and of the anomalies in the process. At the same time, to properly include physicochemical variables in the experimental design it is necessary to both prior knowledge on their effect on the final features of the product and how they may influence each other, to choose the right levels, and to estimate which interaction of interest needs to be inserted in the experimental design. This knowledge is partially available from literature, especially on heating process (Fryers & Robbins, 2005).

2.2.2 Dimension of the process

The dimension of the process is related to the volume of production of the manufacture analyzed, which will require more samples to reach enough statistical power to infer significantly informative information from the data, without occurring in type I and type II statistical errors while investigating the presence of significant variations in the products and the effect of different process conditions in the final product (Dumicic & Zmuk, 2013).

Obtaining a representative sample of a large volume of production is one of the most difficult issues related to Quality by Design procedure. This procedure is especially complicated in food manufacturing, where the individual variability of each sample may be significant and is always present.

To manage this issue, available solutions are establishing a regular sampling procedure or sampling the products after bottleneck passages to detect overall variation in a unit of time. This is particularly functional in discontinuous processes, where the food is produced in batches that are affected by similar conditions. Two key concepts in sampling procedures are design balance and randomization. The first indicates that the sample needs to be representative of the number of elements in the original population for each level of the factor considered, permitting a representative inferring of the condition of the project. Randomization is a cornerstone concept in statistical data analysis and consists in randomly determining the singular samples from the batches and the order in which they are analyzed, to guarantee the assumption of the independent distribution of the measurements.

Generally, this practical issue requires a cost/benefit evaluation, considering first the properties that need constant monitoring because of specific requirements of food manufacturers (Johnsen 2014) or known issues (Duan et al. 2023). Then it is necessary to consider all the practical issues related to the availability of time, personnel, and instruments.

2.2.3 Multivariate Structure

To develop a quality control system, it is necessary to consider that food ~~food~~ products cannot be characterized adopting only one measurement, but to measure properly each quality parameter it is necessary to consider it from a multivariate point of view. Many properties of food cannot be measured using only one variable, such as volatile organic profile and textural properties, and many properties are measured by obtaining more than one variable, such as color. Sensory analysis is multivariate in nature too and requires that all the properties collected are considered both in univariate and multivariate ways at the same time to convey useful information.

At the same time, the food production manufacture must be monitored considering more than one parameter at a time. Even simple processes are affected by multiple variables at the same time, such as time, temperature, and humidity, that interact between them.

The presence of multivariate data in quality control requires a multivariate approach to detect outliers in standard quality control procedure: first, to detect outliers it is necessary to consider multivariate variance instead of univariate, and consequently adopt a different set of statistical tests and visualization techniques to inspect this feature. Multivariate datasets such as spectral measurements, time series, and mass spectrometry have also an issue due to their internal correlation: multiple variables collected may be correlated or anti-correlated with each other, leading to a loss of information and a redundancy of the datasets that many multivariate approaches can handle properly.

Furthermore, to estimate a reliable association between variables of the process and properties of the product it is necessary to deploy a reliable statistical procedure adopting multivariate regression algorithms, such as Partial Least Squares regression (PLS). The algorithm available will be discussed in a further section of the chapter. Multivariate data not only requires specific algorithms, but also a different validation Also, multivariate data requires some attention at the experimental design level: different properties measured simultaneously may imply multivariate statistical error added to each value, which needs to be addressed considering repeated measurements.

2.2.4 Confounding Factors

To infer information about the effect of the condition of the process, quality control procedures associate measurements with known process conditions. It is necessary to consider that all the information is collected with measurements applied in real-time from a real scale production process, hence the variability of the products is related to multiple factors at the same time. Estimating a causal relationship between food properties and process condition requires an accurate knowledge of the underlying phenomenon occurring during food manufacture, a strong familiarity with the production process with the possibility to check directly in the plant the conditions if required, and a good understanding of the measurement techniques adopted. It is always necessary to investigate if the variability detected in the plant is related to external or accidental sources of variability, to not estimate wrong relationships between products that are not actually related together.

Different process conditions may affect the properties of food, so statistical analysis of the data collected from quality control procedures may be easily affected by the presence of multiple confounding factors at the same time. For this reason, it is necessary to have a good knowledge of the underlying phenomenon to directly inspect the food processing facilities, and directly investigate the presence of an external or accidental source of variability present that needs to be included.

It is necessary, after the collection of the measurements the application of preliminary exploratory analysis, such as residual analysis of ANOVA or PCA for multivariate data. If the effect of confounding factors is detected it is necessary to evaluate if either do not consider completely the measurements or to include in further analysis the effect of the confounding factors. If the decision is to use the data, proper statistical analysis includes pre-treatment as mean-centering for each level of a confounding factor, and the inclusion of a blocking factor in the definitive ANOVA or linear models.

2.2.5 Nested factors

During quality control procedures in the food production process, it is possible that the condition of a factor of interest is not always identical according to the different conditions of another factor, such as the effect of the different suppliers or machinery adopted according to different production plants.

This arrangement of an experimental design is called nested or hierarchical design (Montgomery, 2013).

The nested and the nesting factors can be both random and fixed, this depends on the condition of the experiment. Considering the previous examples, the variability associated with different machinery in different plants can be defined as nested fixed factors, while the effect of a different operator inside the same plant at each machine could be defined as a random factor.

Considering nested structure is necessary to correctly interpret the relationships between different factors and to interpret them properly because it is necessary to state that it is not possible to measure the effect of a factor level outside of another factor's level. Thus, not defining properly the exact structure can conduct in the integration of non-existing interaction factors and different calculations for the estimation of the degrees of freedom associated with each factor and interaction and consequently with significant results in the analysis.

2.3 Advanced statistical models

Models for the analysis of a multivariate experimental design (Buvé *et al.* 2022)

2.3.1 Principal Component Analysis

PCA is the most popular unsupervised multivariate analytical technique to extract information out of large multivariate data sets among food scientists (Wold *et al.* 1987). Besides quality control, it is used for data exploration, to look for overall trends, for outlier detection, and to recognize patterns (Abdi and Williams, 2010). Thanks to its powerful visualization tools of scores and loadings, PCA is applied to almost all data at least for exploration purposes (Abdi and Williams, 2010, Jolliffe, 2002, van den Berg *et al.*, 2006).

In the context of quality control, the most common application of Principal Component Analysis is outlier detection, which consists in estimating a multivariate confidence interval on the values of the principal components and assuming that values outside those intervals are significantly different from average values at a multivariate level. This principle was then implemented to create SIMCA classifiers, that are essentially decision trees based on the projection of new measurements onto Principal components estimated from the categories that need to be tested.

Another application in food quality control is the estimation of the importance of the variables from loading values. This procedure is used to estimate which variables contain more information in terms of variance, and to estimate the correlation or anticorrelation structure between variables.

For the analysis of product quality, PCA can highlight the presence of sub-groups that are significantly different at multivariate levels, but it does not provide a clear view of the effect of the differences detected. This is related to the fact that PCA is an unsupervised technique, so the information about the Y-variables needed to be investigated is not explicitly included (in contrast to PLS). Nevertheless, the trends in the data set related to process conditions would be detected by PCA if the multivariate distribution of the parameters is strongly affected by the process conditions.

2.3.2 Parallel Factors Analysis (PARAFAC)

Parallel Factor Analysis consists of a multivariate analysis of multi-way datasets estimating the tensor structure of the dataset (Rasmus, 1997). In the case of a three-way data set, PARAFAC algorithm decomposes it into a sum of triple vector products. Each element in the three-way array estimated by the algorithm, x_{ijk}, can be described as a function of these loadings as presented in Eq. 1

$$x_{ijk} = \sum^{R} a_{ir} + b_{jr} + c_{kr} + e_{ijk}$$

where x_{ijk} is (or represents) the value of the *i*-th sample for the *j*-th and *k*-th variable, R the number of latent components chosen in the PARAFAC model, a_{ir}, b_{jr} and c_{kr} the loadings of the first, second and third dimension for each PARAFAC component r and e_{ijk} the residual. The PARAFAC model is built using an iterative Alternating Least Squares method based on tensor algebra. The PARAFAC model minimizes the sum of squares of the residual e_{ijk} (Rasmus, 1997).

PARAFAC can be applied to the multi-way data set avoiding the unfolding procedure. Unfolding can result in a loss of information, due to the absence of correspondence between the original data and the applied model (Bro, 1997). In comparison with PCA unfolding, PARAFAC algorithm avoids overfitting issues and is more robust with multi-way datasets, but generally fits worse the data than PCA algorithm. It is interesting to consider that, being robust, PARAFAC models are less sensitive to noise, so if the model is able to describe the data well enough, the larger variance explained by PCA algorithm will include probably noise (Rasmus, 1997, Smilde et al., 2004).

It should be noted that until now PARAFAC was almost exclusively applied to fluorescence spectroscopy data due to its ability to model physical phenomena directly (i.e. to split the fluorescence signal into signals of pure constituents). For this reason, the application of those algorithms in other types of multi-way data needs to be discussed and validated before starting to analyze its applicability. There are interesting contexts where PARAFAC procedure could be useful for inferring and summarizing interesting information, such as temporal sensory data (TDS and TCATA), and complete stress/strain curves from textural measurements.

2.3.3 Multivariate factors Analysis (MFA)

Multiple Factors analysis (MFA) (Becue-Bertaut & Pages, 2008) is a multivariate data analysis method for summarizing and visualizing a complex data table in which individuals are described by several variables (quantitative and/or qualitative) structured into groups. It considers the contribution of all active groups of variables to define the distance between individuals.

The number of variables in each group may differ and the nature of the variables (qualitative or quantitative) can vary from one group to another, but the variables should be the same nature in each group (Abdi & Williams, 2010). MFA is a general factor analysis, based on PCA when variables are quantitative and MCA (multiple correspondence analysis) when variables are qualitative.

The main principle of MFA is the application of factor analysis to the whole set of variables in which each group of variables is weighted respectively by PCA or MCA transformation. This procedure represents individuals and variables as in any factor analysis, adopting score plots, loading plots, and biplots. This statistical instrument is capable of displaying at the same time the effect of each group of variables on individuals and comparing them at individual and aggregated level (Escofier & Pagés, 1994).

2.3.4 ANOVA Simultaneous Component Analysis (ASCA)

Anova Simultaneous Component Analysis (ASCA) consists of applying a multivariate matrix decomposition (based on SVD) on the matrices of the expected values estimated from a univariate analysis of the experimental variables. In other words, the ASCA approach analyzes at a multivariate level the effect of each factor, and uses the powerful visualization instruments of PCA, such as score and loading plots, to highlight the latent factors that are contributing to each design factor.

The first step consists in applying to every variable a parametric statistical model which must apply the same decomposition to every variable. The aim of this procedure is to estimate the effect on each singular variable of the levels of each factor included in the model. To infer significant information, the model applied must apply a decomposition representative of the experimental design of the data, including factors and interaction that correspond to the real experimental design. To properly represent more complex experimental design different versions of ASCA have been proposed to include nested and random factors, such as LMM-ASCA, and to consider multidimensional correlation structures, such as PARAFASCA.

Once the models are estimated, it is possible to infer for every level of every factor including the effect on each variable, obtaining multiple matrices reporting the variation from the overall mean for every variable. The considerable outcome from this passage consists in obtaining a simplification and a representative summarization of the individuals of the experiments, capable of proceeding with further multivariate analysis with a restricted dataset that consists in the estimation of the tendencies of the dataset.

The significance of these values is estimated by the validation of the models at the univariate level, selecting only the values significantly different for the factor considered. This step is added to ensure that the analysis uses only variables affected by the factors investigated. It is worth saying that the ASCA procedure applies a standard multivariate analysis based on Single Value Decomposition, which estimates principal components to maximize the variance between individuals, therefore the values estimated by non-significant models do not add significant difference between subjects.

All the significant effects for each level of each factor are estimated using this equation:

$$X_{aj} = T_{aj} * P_{aj}$$

where X_{aj} is the effect matrix of the factor a in the j-th variable, T is the contrast matrix, and P is the vector of the effects. The effect matrix is grouped for all the parameters to obtain a complete effect matrix that represents the decomposition of the effect estimated for each level of a factor for every variable.

To observe the distribution of the samples and to interpret their relations with each variable, Principal Component Analysis (PCA) is adopted on the matrix of effects.

Before the estimation of principal components, to compare the effects estimated for different variables, the matrices of the effects of each factor were centered for each variable.

The results can be reported by adapting the graphical tools proper of PCA, such as bi-plots and scree plots, allowing ~~of the~~ summarization of large pieces of information in a clear and deliverable way.

For the interpretation of ASCA graphical tools, it is necessary to consider that the single value decomposition is not applied to the overall variance of the dataset, but to the variance related to a single factor or interaction at the time, consequently, the percentage of explained variance for each component represents the variance associated to the sum of squares of the considered factors. To associate the value of the explained variance to the overall variance of the process it is necessary to estimate the contribution of each factor to the multivariate variance by estimating the overall sum of every sum of squares for the considered factor for each model and consider its proportion toward the overall sum of the total sum of squares for each model (Thiel *et al.* 2017). A complete calculation will lead to the determination of the explained variance of each factor and interaction included in the decomposition procedure, including the residuals, and the percentage of multivariate variance explained by this procedure. The results can be properly represented by a bar plot. This plot, along with the results of the univariate or multivariate permutation tests can lead to the interpretation of the significance of the multivariate decomposition of ASCA. If the multivariate variance is explained mostly by the residuals and there is no multivariate significant difference and not even a large amount of univariate significant parameters, the factor considered can be considered as not significant.

Subsequently, the presence of unincluded multivariate effects can be included in analyzing the multivariate structure of residuals adopting biplots. This visual analysis helps detect the presence of significant multivariate effects that are not represented in the model by estimating if the multivariate analysis of the residual is different from a multivariate normal distribution.

2.3.5 Linear Mixed Models-ASCA

Linear mixed models (LMM) are an extension of standard linear models for regression analysis of experimental designs containing observations that cannot be assumed to be independent of each other, such as repeated measurements or measurements from the same sample. This method considers the error structure in the data, estimating at the same time the effect of fixed factors, which are the factors of interest, and random factors, which represent the individual variability caused by non-measurable sources of variation. This univariate statistical analysis can be integrated into a multivariate framework extending the principle of ANOVA simultaneous component analysis (ASCA).

The application of linear mixed models with ASCA decomposition in an experimental design with nested and random factors and unbalanced levels is still an uncommon procedure in data analysis (Martin & Govaerts 2020). Therefore, the issues related to the evaluation of this type of statistical model and its application are still not widely addressed in the context of food manufacturing quality control. In the last decade, the interest in this methodology has been increasing, due to the growing availability of vast multivariate datasets from stratified experimental designs acquired in ecological and industrial studies.

2.3.6 Parallel Factor ANOVA simultaneous component analysis (PARAFASCA)

PARAFASCA consists of the application of Parallel Factor Decomposition to the effect matrix estimated from the ANOVA decomposition of a multi-way multivariate experimental design. This procedure is proposed for a multi-dimensional dataset, which means datasets that contains at least three crossed design factors. These factors affect independently the response and all the levels of one factor are crossed with all the levels of the other factor, for example a dataset of spectral data collected along time where all the measurement were estimated for the same frequencies (Jansen et al. 2008). These data contain two simultaneous correlation structures within each factor, and an interaction between these two the other factors considered (Acar & Yener 2009).

The first part of the analytical procedure is the same as ASCA, consisting of the decomposition of each variable according to the experimental design considered, but the decomposition is applied to each variable at each level of the third-dimensional factor.

PARAFASCA consists of the combination between PARAFAC and ASCA by modeling the estimate values obtained by ANOVA decomposition of relevant design contributions to the variation of parameters using PARAFAC instead of PCA. The interpretation of this novel technique is more straightforward for multiway datasets than standard ASCA because PARAFAC can disentangle factors and interactions in the experimental design.

2.3.7 Partial Least Squares regression (PLS)

PLS is a widespread algorithm that was proposed initially by Wold (1966) that consists, in its simplest form, in a method to associate two data matrices using a linear multivariate model that estimates a series of mathematical objects that maximizes the covariance between the two datasets. These mathematical objects are latent variables calculated iteratively to estimate the coefficients of each variable to maximize the covariance between the two matrices. The iteration consists in applying the calculation of the next principal components adopting the residuals from the previous component adopted, assuring that augmenting the number of components will improve the fitting of the overall model (Vinzi *et al.* 2010).

The iterative estimation of these components allows the researcher to estimate multiple coefficients for each variable, allowing the representation of complex correlation structures between predictors and predicted values. Those values can also have an interesting informative value, allowing to represent ~~of~~ this correlation structure with the correlation plot, which is a specific application of the bi-plot graph, adapted for the information inferred from PLS models.

2.3.8 Orthogonal Partial Least Squares Regression (O-PLS)

O-PLS consists of an alternative to the PLS algorithm that has certain properties rendering the model more tractable, and the mathematical procedure allows modeling separately the variations of the predictors correlated and orthogonal to the response (Thevenot et al., 2015). This model improves the interpretation of the effect of the predictors and their systematic variation compared to standard PLS (Pinto et al.,2012).

The O-PLS algorithm divides the overall variance in the X-block into two model parts, one part which models the correlations between X and Y and another part that expresses the variation that is not related (orthogonal) to Y. Components that are correlated to Y are here called predictive. Components that are uncorrelated to Y are here called orthogonal. It is worth noting that if the model is trained to predict a single Y-variable, the O-PLS algorithm will estimate one predictive component and any number of orthogonal (no relation to Y) or pseudo-orthogonal (very little relation to Y; only when missing values in X) components (Eriksson et al. 2013). Therefore, differently from standard PLS, only one set of predictors will be estimated.

The O-PLS algorithm structure represents in a more operative way the performance of the model and the significance of the variables because it employs only one principal component, so only one coefficient is estimated in the model, therefore the importance of variables is reported in an easily interpretable way. The importance of variables is reported by adopting tailored graphical solutions such as S-plots, which are scatterplots representing the correlation of each variable or bar plots comparing the index of importance available.

2.3.9 Classifiers for process monitoring

Classifiers are machine learning algorithms that instead of predicting continuous values they assign a category to each subject according to the values of the dataset they were trained to analyze. Those algorithms have got wide application in the context of quality control because they can allow a fast and effective procedure for assisting or automatizing the decision-making process in quality evaluation. There are multiple different algorithms that are adopted for these procedures. A good example is Support Vector Machine (SVM). SVM is a widely adopted algorithm for classification, then improved also for regression (Cervantes *et al.* 2020), that is based on the definition of variables spaces for classification with a machine learning procedure based on a hyperplane that maximizes the margin between classes. The training of the parameters is not based on minimizing the mean squared error, but on maximizing the existing distance between the hyperplane estimated and the closest observation. This index is called “margin”.. The method of training based on maximizing margin is necessary to find a unique solution for the definition of the hyperplane because otherwise there would be an infinite number of solutions in most situations.

The SVM algorithm is used mainly in the context of bioinformatics and image analysis, but there are also applications in food science (Zhu & Spachos 2021; Astuti *et al.* 2018).

The downsides of applying the Support Vector Machine algorithm are the high computational cost due to the algorithm's complexity and the fact that it is designed for binary classification problems. Furthermore, the performance of standard SVM algorithms is severely affected by unbalanced datasets. For those reasons, support vector machines are adopted mostly for the development of sensors for quality control and not for higher-level classification procedures (Cervantes *et al.* 2020).Random Forest is a machine learning algorithm used for classification and regression (Fawagreh *et al.* 2014). Developed by Breiman (2001), the method combines the bagging sampling approach (Breiman 1996), and the random selection of features, (Amit & Geman 1997), to construct a collection of decision trees (a "forest") with a controlled variation. The bagging sampling consists in constructing each decision tree in the ensemble using a sample with replacement from the training data, while the random selection of features consists in selecting a subset of variables. After the estimation of each tree, the ensemble determines the class label of an unlabeled instance. This is done via majority voting where each classifier casts one vote for its predicted class label, then the most voted label is used to classify the instance.

The random forest algorithm has been applied virtually in every field, in food quality, there are many applications for the relationship between the effect of the process and the chemical properties of the final product (Granitto *et al.* 2006; Fabris *et al.* 2010; Meoni *et al.*, 2021).

Because of the bagging procedure, each random forest model can estimate the importance of each variable for the classification in one or another group, performing a simple and easily interpretable variable selection procedure. Another important advantage related to the random forest algorithm is the reiterated principle of estimating the classification with multiple decision trees, which allows us to estimate the average amount of attribution of a certain class rather than another. Another interesting feature of the random forest algorithm is that is a non-parametric model, so it is not necessary that the variables of the models respond to a specific distribution of values, because the random forest algorithm relies on a structure of multiple decision trees.

The downfalls of random forest are the risk of overfitting, which can be avoided by adopting many samples to develop a reliable predictive model. Because of its non-parametric structure, it is important to develop a good pre-selection procedure for the variables, removing variables that do perfect separation in classification procedures (Belgiu & Dragut, 2016).

2.4 Validation Procedures

For the estimation of the significance of a statistical model validation is a mandatory step in statistical analysis. Validation can be defined as an established scientific and statistical approach proving that a set of measurements is sufficiently reliable to prove fitness for purpose. In the context of statistical analysis, validation consists of the statistical procedure to estimate the probability that the result inferred by the model is significantly different from the results that could be inferred if the data acquired were not representative of any distribution.

2.4.1 Outlier estimation

This procedure consists of the estimation of the importance of each measurement in the final model, by comparing the performance obtained in a model without these measurements (Cousineau & Chartier, 2010). The objective of the comparison between these two models is to check if the prediction changes significantly after the removal of a singular value. If a model is significantly affected by a singular measurement, this model is not representative of the overall variability of the population, it rather summarizes the variance between the overall population and the outlier(s). Once the outlier is detected, it is reported and investigated as a possible signal of an anomaly inside the production process.

2.4.2 Univariate Monte Carlo Simulation

As stated above the null hypothesis consists in the absence of significant effect of the parameters of interest in the experiment, except for the random error and the presence of effects related to random factors. To estimate the null distribution usually a Montecarlo simulation (n = 1000) is adopted to estimate the null hypothesis' distribution of the sum of squares decomposition and to compare the results to the decomposition of the variables applied considering the factors of interest of the experiement, as discussed in Stamirova et al. (2013). To perform a reliable permutation test, it is necessary to maintain the structure of the dataset, including all the nested and random factor structures in the random reassigning procedures, to ensure that the null hypothesis' distribution estimated is related only to the effect of the factors investigated and their interactions. Furthermore, the α-values needs to be chosen, in a proportion related to the number of iterations. The null hypothesis is rejected when the permuted sum of squares decompositions is lower than the values obtained by the real model for less than 5% of the permutations. After the permutation, for each factor, it is necessary to adjust the estimated p values using the Bonferroni correction.

Using multiple permutations at the univariate level, after an appropriate adjustment for multiple comparisons, instead of a singular permutation test at the multivariate level, allows for avoiding the effect of internal correlation when evaluating the significance of each parameter.

2.4.3 Multivariate Monte Carlo Simulation

The procedure of validation can be performed also at a multivariate level, validating at once the significance of the data matrix estimated by the multivariate model adopted, such as ASCA, LMM-ASCA, PLS, or PCA, comparing the F norm with a null distribution of F norm estimated from models inferred from permutated data matrices.

This multivariate validation procedure is considered preferable: from a computational point of view, because it requires less calculation power, consisting of only one permutation procedure, while from a statistical point of view, because the internal correlation is included in the matrix analyzed so the overall structure of the data is considered in the procedure of estimating the null distribution. This procedure also is more easily interpretable, because the hypothesis verified is the significance of the whole analysis and not the significance of a singular variable in the overall model. An example of the representation of the results of this statistical procedure is reported in Figure 1.

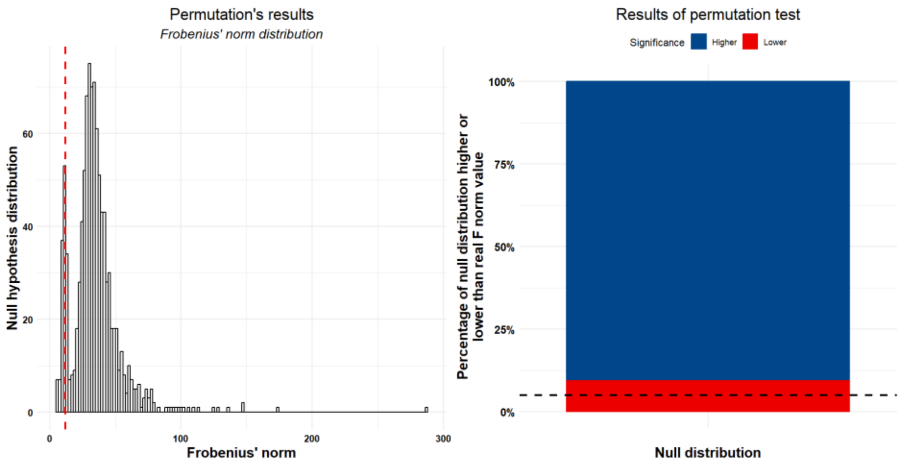

Figure 1: *Graphical representation of multivariate validation via permutation test. The histogram on the left represents the null distribution of the F norm estimated via permutation tests, while the red vertical thinned line represents the F norm estimated from the actual matrix of data. On the right the fill column on the right represents the proportions between the values of the null distribution lower than the actual F value (red) and the values of the null distribution higher than the actual F value (blue). The black thinned horizontal line represents the confidence limit of 5%.*

2.5 Estimation of the importance of variables

Once a model has been developed and validated, the estimation of the importance of the variables is a procedure that consists in determining which variable of the dataset composes the significant effect detected. This procedure is functional to interpret multivariate models. The importance of variables is reported as a singular value adopting different numeric indexes. The estimation of the variable importance could be executed by adopting different procedures: according to the algorithm adopted there may be a different value determining the importance of the variable, but algorithms such as Random Forest, PLS, and O-PLS-DA can determine the importance of each variable. In many cases, research relies on the value of these variables estimated by these methods, but it is preferable to validate these indices by estimating confidence intervals by bootstrap procedure: Random Forest algorithm performs already an internal bootstrapduring the validation step, so a well-trained algorithm already performs a complete and informative validation process.

The indices adopted are the Variable Importance Parameter (VIP), the regression coefficient, or the loading values (Galindo-Prieto *et al.* 2014).

The bootstrap procedure consists of the estimation of the confidence intervals by estimating multiple models removing a small set of samples inside the dataset and replacing the removed samples by repeating other measurements. Adopting the bootstrap procedure guarantees that each estimation contains the same values of mean and median, to guarantee a similar data structure with different values. The values estimated are used to estimate a confidence interval for each variable. The estimation of the confidence interval is necessary to estimate and compare the importance of the variables.

According to the algorithm adopted and the structure of the data, the results can be represented using a bar plot, with error bars to represent confidence intervals, or adopting loading plots if the model adopted is composed of multiple mathematical objects, such as PLS or Support Vector Machine. Another plot often adopted is the S-plot, which is a scatterplot that represents at the same time the covariance and the correlation structure of the predictor's variables in PLS and O-PLS models. More exactly, in the x-axis of the S plot is reported the covariance of the variable for the principal component considered (this is particularly useful when it is adopted by the O-PLS algorithm, which comprises only one principal component), and the correlation between each variable and the score vector of the predictive component.

CHAPTER 3. STATISTICAL MODELS FOR PHYSICAL MEASUREMENTS FROM QUALITY CONTROL PROCEDURES

In the present chapter, the application of advanced statistical workflows for the interpretation of data from quality control procedures in the Trentingrana case study are reported. The importance of physical measurement is presented in the context of the case study. An LMM-ASCA statistical model is proposed to infer the effect of the condition of the project considering the experimental design available. The reported work was already presented in Ricci *et al.* (2022), published in the open access journal Foods.

3.1. Measurement procedures to deal with quality control issues

3.1.1. Instrumental textural measurements

There is a wide range of instrumental techniques to assess food texture in both research and industry (Oraguzie *et al.*, 2009, Zdunek *et al.*, 2010). Those measurement methods are based on different mechanical principles, they interact in a different way with the sample, and they differ on the amount of information reported, some estimate only a single value, while others provide broader information on the history of deformation, such as time-series data on texture measurement (Derington *et al.*, 2011).

The majority of the textural measurement procedures estimate the force exerted by the instrument on the sample in a limited and repeatable interval, which can be until a rupture in the structure of the sample until a defined length of deformation or until the amount necessary to maintain a uniform speed during the deformation of the sample is reached (Chaunier *et al.*, 2007, Greve *et al.*, 2010). Thus, it is difficult to compare the results obtained by using different instruments without knowing properly the context of the product analyzed and the meaning of the values estimated from the specific measurement procedures adopted.

The first characterization of the typologies of mechanical measurements of food texture consists of two macro-categories, one of the destructive methods and the other of the non-destructive methods. Destructive measurements analyze the properties associated with the micro-structural and molecular structures of the product. This group comprehends the three-point bending test, single-edge notched bend (SENB) test, puncture, penetration, and cutting tests.

Non-destructive textural measurements do not cause visible damage on the matrix and are possible to apply this procedure online, but they are still destructive on a micro-scale and at the same time the information obtained from experiments is not comprehensive.

Instrumental texture analysis is not directly related to mouthfeel, so there cannot be a direct association between the sensory properties estimated via sensory analysis, but there are multiple studies about the relationships between instrumental and sensory properties.

3.1.2. Colorimetric measurements

The color is a property that is defined both from a physical point of view and a sensory point of view. According to the first point of view color appearance is the response of retina rods and cones to the reflected radiation in the so-called visible region of the electromagnetic spectrum, that is, the range between 400 and 700 nm, which is due to the interaction between a light source and pigments in the food sample.

On the other hand, color is also a perception mediated by neuronal answers, to complex external stimuli, by which everyone gives a personal interpretation, mediated by different external influences (Cairone *et al.* 2020).

Color represents the first characteristic of a food product noticed by the consumer and plays a dominant role in the decision-making process.

When a consumer decides to acquire foodstuff, they use the color appearance as the first sensory evaluation, matching the color appearance to other food proprieties such as ripening, freshness, and absence of defects. The evaluation of color perception by consumers may be or not be conscious, according to the individual and to the food product, but the stimuli perceived by the eyes are interpreted by the brain and could be influenced by other information.

Nevertheless, the energy associated with the light's reflection process depends on the pigment type and the content that can be measured. Hence, the strict measure of material quality could be an objective physical measure of chemical parameters and an indicator associated with the sensory quality parameters, depending on the structure ~~and the structure~~ and the properties of the food product (Pathare *et al.* 2013).

There are multiple systems to unambiguously define color (Schanda 2007), the CIEL*a*b* color space is the standard method proposed by the Commission Internationale de l'Eclairage with the aim to answer the human perception of colors. According to this method, each color is defined using three parameters: L*, which represents the luminance, between black and white, perceived by the retina rods; a* which is the expression of the greenness for negative values or of the redness for positive values, and b* which represents the blueness for negative values or the yellowness for positive values.

3.2. Application 1 – Study of the effect of the production process on the physical properties of Trentingrana cheese

3.2.1. Introduction

The texture properties of extra-hard cheese affect how the cheese is portioned and packed. Texture also affects the behavior of the cheese when it is subjected to shredding or grating, and how the cheese retains gas and hence, its predisposition to form eyes, cracks, or swell. The color of cheese significantly contributes to sensory responses and plays an important role in the anticipation phase of the selection and consumption of food materials (Fox & Cogan 2004). Consumer expectations are influenced both by cheese color itself and its homogeneity. Both colorimetric and textural properties are critical for the commercial value of hard-seasoned cheese, such as Trentingrana cheese. The measurement of those properties defines the product from a technological and commercial point of view.

The quality of the raw milk (casein content, casein micelle structure, and fat content, Bittante *et al.* 2011, McDermott *et al.* 2016) and conditions of the cheese-making process (for example, pre-acidification of milk, type and quantity of rennet, cooking temperature, acidification of the cheese mass, and temperature and humidity during seasoning, Mucchetti *et al.* 2014) are fundamental factors that influence the textural and colorimetric properties of cheese (Banks 2007).

The production chain of hard-seasoned cheese with a Protected Designation of Origin (PDO) usually consists of many individual dairy factories belonging to the same producer cooperative that transforms raw milk conferred daily from many small farms. This fragmentation of the process suggests that there may be significant differences in the process and the characteristics of the raw material (Cipolat-Gotet *et al.* 2013) despite the presence of a consortium that regulates the production. At a supply chain level, the effect of process and raw materials conditions on the physical characteristics of the final product are mostly attributable to the dairy factory and the time of the year when the fresh milk is delivered.

Due to the large dimension of the Trentingrana cheese wheel (a height from 20 to 26 cm and a diameter from 35 to 45 cm), these physical properties can vary according to the position in the cheese wheel, since water content and temperature in the early stage of the process, as well as microbial activity, depends on the distance from the center. Thus, for a more comprehensive description of the physical properties of a single cheese wheel, it is necessary to take multiple samples from the same cheese wheel, addressing the distance from the central position.

To date, there are few studies that have investigated the effect of the dairy factory and the time of the year on the colorimetric and textural properties of hard cheeses; nevertheless, Bellesia *et al.* (2003) highlight a large variability among dairies for volatile components of Parmigiano Reggiano cheese while Careri *et al.* (1996) report a much lower variability on the same type of cheese in relation to chemical parameters and non-volatile fractions. Franceschi *et al.* (2019) notice how the month of the year and the dairy can determine the efficiency of the cheese-making process. According to our knowledge, no study has tried to estimate the effect of different dairies and different months of the year on the textural and colorimetric properties of cheese.

To estimate the effect of the dairy factory and the time of the year on both colorimetric and textural properties while considering the natural variability of artisanal cheese production, it is necessary to characterize a real-scale production process. From a statistical point of view, the optimal data analysis strategy has to be able: (a) to take into account the multilevel nature of the experimental design separating the contribution of the different study factors from the variability arising during the cheese-wheel making process; (b) to estimate the correlation among colorimetric and textural properties.

Linear mixed models (LMM) are an extension of standard linear models for regression analysis of experimental designs containing observations that cannot be assumed to be independent of each other, such as repeated measurements or measurements from the same sample (Bats *et al.* 2013). This method considers the error structure in the data, estimating at the same time the effect of fixed factors, which are the factors of interest, and random factors, which represent the individual variability caused by non-measurable sources of variation. In our specific scenario, linear mixed models can handle information better than ANOVA, because they take into consideration repeated measurements in the same cheese wheel, and they describe in the models the variability related to uncontrolled production parameters.

LMM is a univariate approach that can be integrated into a multivariate framework extending the principle of ANOVA simultaneous component analysis (ASCA) (Smilde *et al.* 2012), which consists of applying a multivariate matrix decomposition (based on SVD) on the matrices of the expected values estimated from a univariate analysis of the experimental variables. In other words, the ASCA approach analyzes at a multivariate level the effect of each factor, and uses the powerful visualization instruments of PCA, such as score and loading plots, to highlight the latent factors that are contributing to each design factor.

The application of linear mixed models with ASCA decomposition in an experimental design with nested and random factors and unbalanced levels is still an uncommon procedure in data analysis. Therefore, the issues related to the evaluation of this type of statistical model and its application are still not widely addressed in the context of food manufacturing quality control. In the last decade, the interest in this methodology has been increasing, due to the growing availability of vast multivariate datasets from stratified experimental designs acquired in ecological and industrial studies.

Stamirova et al. (2013) applied a linear mixed model to evaluate differences in food manufacturing, presenting a statistically robust procedure for analyzing data on the effect of different herbal tea pasteurization treatments from three different years of production. The effect of many confounding factors in the chemical properties of a herbal product is estimated using a statistical model to describe the complex interactions between the main fixed and random factors, obtaining consistent results.

Martin and Govaerts (2020) reviewed several different applications of linear mixed models with ASCA decomposition for different datasets, ranging from metabolomics to sensory science. A step-by-step procedure to develop and evaluate a statistical model with both random and fixed effects is explained, and the procedures to estimate models' outcomes, such as statistical significance and effect size are compared. A suitable procedure for matrix decomposition and reconstruction is presented, with a focus on the individuation of the most important effects after the transformation of data.

This study aims to analyze the variation of textural and colorimetric parameters of a semi-artisanal PDO product according to dairy, sampling position, and time of the year using LMM-ASCA analysis, which has proven to be a valid statistical procedure to evaluate a large-scale dataset of measurements from Trentingrana industrial quality control process.

3.2.2. Materials and methods

3.2.2.1. Sampling procedure

During the years 2017 and 2018, a total of 317 Trentingrana cheese wheels were collected from the 15 dairies belonging to Trentingrana Consortium, according to the sampling procedure of the Trentingrana project.

Subsequently, from each wheel, 24 blocks of cheese were sampled, each block with a length of 3 cm, a width of 1.5 cm, and a height of 1.5 cm. Each set of blocks from the same product was cut at the same distance from the center of the wheel and assigned to one of six different categories according to the distance from the central position. The sampling position is illustrated in Figure 2. Colorimetric and textural analyses were carried out on each block.

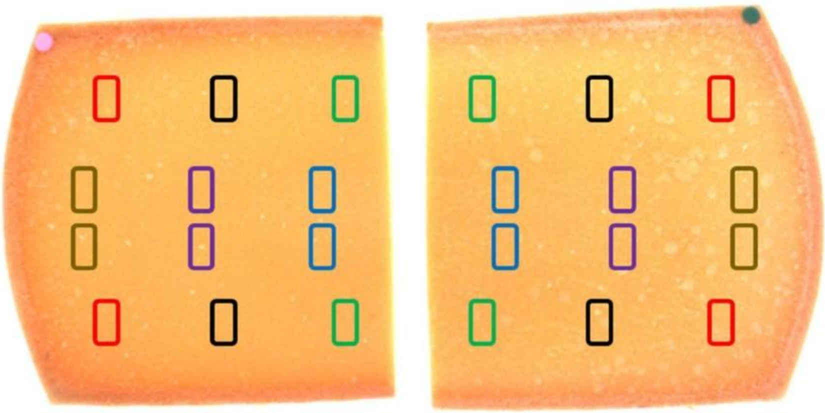

Figure 2: *Example of two slices of Trentingrana cheese. Each box corresponds to one sampling position, each color to a sampling zone: round (RND, red); round central (RNDs, brown); external plate (RNDp1, black); intermediate zone (Int, purple); internal plate (RNDp2, green); center (CNT, blue).*

Globally, 317 cheese wheels were sampled from 12 sampling sessions from each dairy on a bi-monthly basis and 7608 measurements were acquired for 6 different parameters during 54 analytical sessions.

3.2.2.2. Analytic Determinations

L*a*b* components from the CIELAB color space model (Schanda 2007) were measured once on one of the wider surfaces of each cheese block sample using a CR-400 colorimeter (Konica Minolta Sensing Inc., Tokyo, Japan) using the D65 illuminant source, an observation angle of 2°, and previously calibrated with a reference white standard ceramic tile. Data were acquired using the CM-S100w SpectraMagicTM color data software (Konica Minolta Sensing Inc., Tokyo, Japan).

Texture properties were measured on each cheese block by a TA-XT texture analyzer (Stable MicroSystem Ltd., Godalming, UK) applying a uniaxial compression/penetration on one of the wider sides of the cheese block sample. Following the method described by Noël *et al.* (1996), a 4 mm probe was used with a speed of 1.67 mm/s, a trigger force of 5 N, setting the endpoint of the measurement when a maximum strain of 90% of the height of the sample was obtained, and three mechanical parameters were calculated on the recorded curves. Those parameters are reported in Table 4.

Table 4. Parameters extrapolated from the stress/strain curve estimated from uniaxial compression.

Parameter	Description	Measure Unit
Maximum Force (Fmax)	*The maximum amount of force applied by the uniaxial probe to the sample.*	*N*
Area under the curve (Ac)	*The whole area under the force/strain curve during the compression of the sample until the endpoint.*	*N*mm*
Elastic modulus (El)	*The slope of the linear part of the stress-strain curve.*	*N/mm*

3.2.2.3. Statistical analysis

Due to the various sources of variability that affect the final product, the results are highly structured data. Before each analysis, each variable has been centered and scaled to unit variance, to obtain comparable results from each model. Each variable was checked for the assumption of normality using QQ plots in Figure 3.

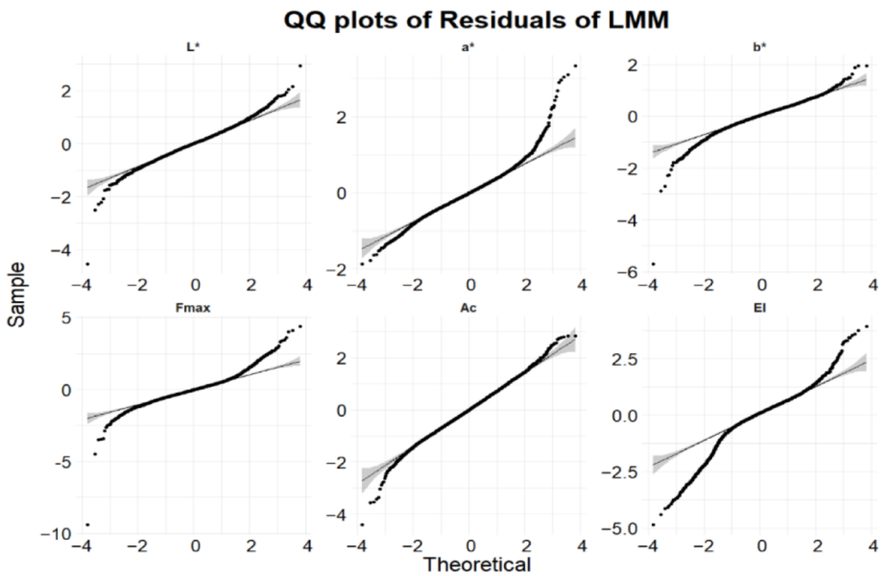

Figure 3: *Q-Q plot of the normal distribution of each parameter of instrumental measurements*.

The experimental design was built to consider the many specific sources of variation inside the supply chain of a PDO product (Endrizzi et al. 2013). Here we wanted to estimate the effect of the dairy factory and the time of the year ("Time" factor, from now). The position in the wheel was also included in the model to estimate the natural variability present in each cheese wheel.

To describe the design of the sampling campaign, we used an unbalanced experimental design with four factors and their interactions, represented by this Equation:

$$x_{iklgt} = \mu + \alpha_k + \beta_l + \gamma_g + (\alpha\beta)_{kl} + (\alpha\gamma)_{kg} + (\beta\gamma)_{lg} + \delta_t + \epsilon_{klgti}$$

Where α corresponds to the fixed factor "sampling position" with level k, β corresponds to the fixed factor "Time" with level l, γ to the fixed factor "dairy factory" with level g, and δ to the random factor “cheese wheel” at level t. Level i represents the repeated measurements conducted in the same sampling position of the same cheese wheel from different blocks. This model is applied to each colorimetric and textural parameter estimated.

Because each couple of months the cheese wheel has been analyzed from each dairy, the cheese wheel factor is nested inside the other fixed factors, so there is no interaction effect that involves this factor.

For each variable, a linear mixed model was estimated to assess the significance of the factor “Time”, “dairy factory”, and “position”. The cheese wheel factor was defined as random. For each factor and each interaction, the matrix of contrasts was set to obtain a sum to zero estimation. This procedure ensures that each level would be tested with the grand mean of the dataset.

As previously stated, the aim of the experiment was to determine the effect of the dairy factory (n = 15) and the time when cheese is produced (n = 6).

Each model was validated by a permutation test. The null hypothesis for each variable was that there is no significant effect for any of the parameters except for the random factor of the cheese wheel. Montecarlo simulation (n = 1000) was used to estimate the null hypothesis’ distribution of the sum of squares decomposition and to compare the results to the effective decomposition, as discussed in Stamirova *et al.* (2013). Each permutation test was performed, maintaining the nested structure of the factor of the sampling position inside the other two factors, to make sure that the null hypothesis’ distribution estimated was attributable only to the effect of the factors investigated and their interactions. The α-values chosen are 1/1000 and the null hypothesis is rejected when the permuted sum of squares decompositions is lower than the values obtained by the real model for less than 5% of the permutations. After the permutation, for each factor, the estimated p values were adjusted using the Bonferroni correction. Each analysis have been performed using R statistical language (R core Team, 2021).

3.2.3. Results and discussion

According to the results of the permutation test, the factors “Dairy Factory” and “Sampling Position” are significant in the models estimated for all parameters, and the factor “Time” is significant for each parameter except for colorimetric index a* and the textural parameter “Area under the curve”. The interactions between “Dairy Factory” and “Time” are not significant for all the parameters, the “Time” and “Sampling Position” interaction is significant only for the colorimetric parameters, and the interaction between “Dairy Factory” and “Sampling Position” is significant for each parameter except the elastic modulus.

Those results state that the color and the structure of the cheese wheels change when there are changes in the production process, in the sampling position, and, to a lower amount, the season of the year when milk was produced.

There are no important differences between the variations of each dairy at different seasons, but the overall value of a sampling position varies according to the dairy and the time, at least for colorimetric properties. Results are shown in Figure 4.

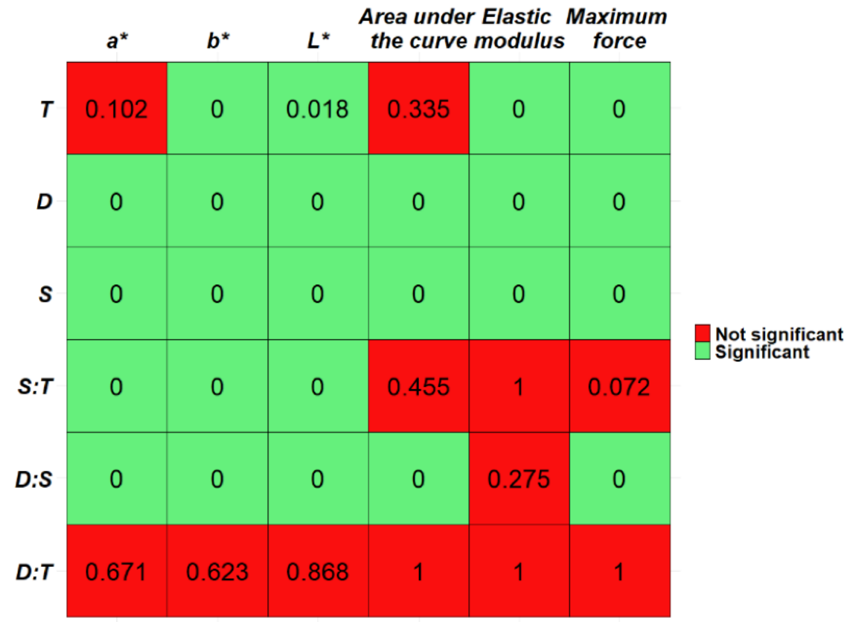

Figure 4: *Tile plot reporting the p values estimated from the permutation tests for the significance of each factor ("T": Time, "D": Dairy Factory, "S": Sampling Position, "S:T": interaction between Sampling Position and Time, "D:S": interaction between Dairy Factory and Sampling Position, "D:T": interaction between Dairy Factory and Time) of the linear mixed models.*

3.2.3.1.<u>Simultaneous Component Analysis: LMM-ASCA Results</u>

The first features of the model evaluated were the estimations of the contribution of each parameter to the explained variance for each different factor, as reported in Figure 5. The contribution of each parameter to the multivariate decomposition of each factor was estimated according to Kassambara (2017a). The contribution of explained variance was estimated for all the dimensions considered for each factor: four for "Dairy Factory" and two for "Sampling Position" and "Time". The number of dimensions taken into consideration was estimated to obtain at least 85% of explained variance for each factor; the values are represented in the scree plots in Figure 6.

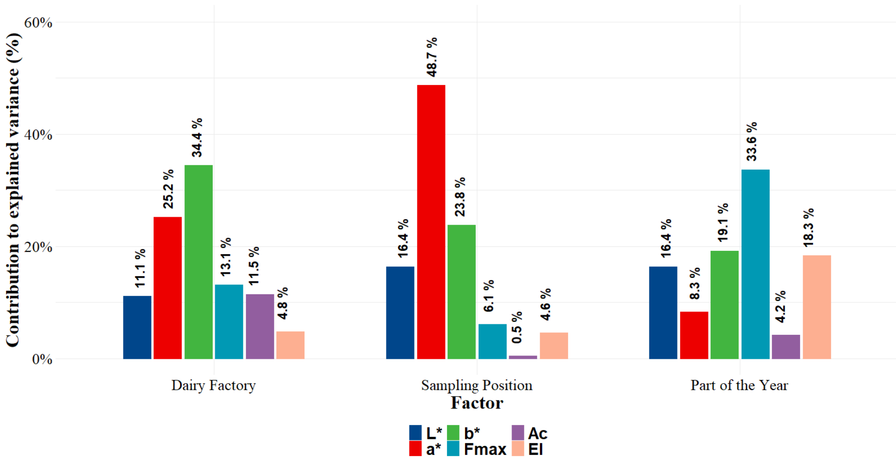

Figure 5: *Barplots showing the contribution of each parameter to the explained variance for each LMM-ASCA decomposition of the model.*

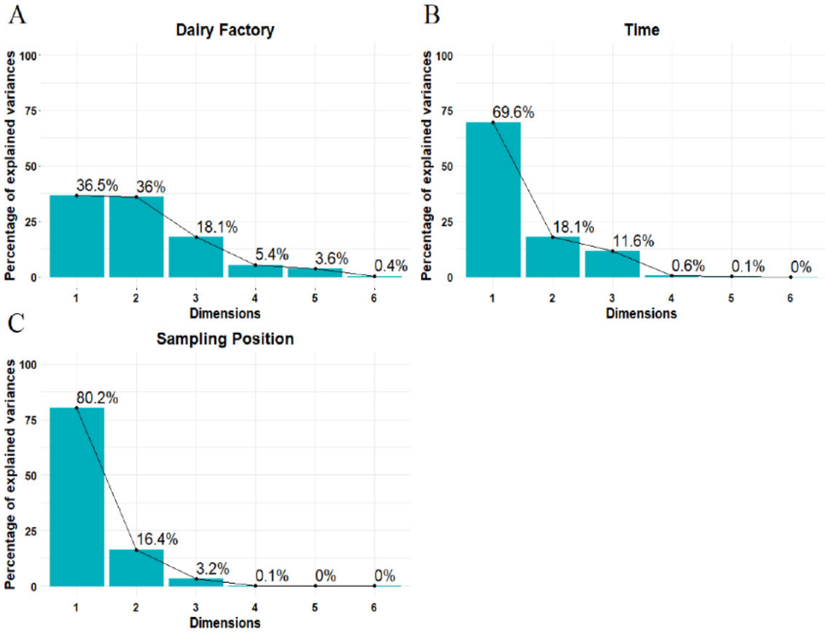

Figure 6: *Scree plots for the ASCA decomposition of the factors "Dairy Factory" (A), "Time" (B), and "Sampling Position" (C).*

In the LMM-ASCA model, the contributions of the experimental variables represent the effective fraction of the variance considered for each factor due to the preprocessing step of unit scale normalization of each parameter before the estimation of all the models, and the estimation of the effect matrix from all models together.

Different from PCA, principal components from ASCA decompositions represent the effective percentage of explained variance quantitatively, and the contribution of each parameter is quantitative information of which the parameter effectively varies the most between the levels of each factor considering the effect of all the other factors in the experimental design.

The parameters containing the largest amount of variance in ASCA decomposition for the factor "Dairy Factory" are the colorimetric parameters and, for a less amount, the textural parameter Area Maximum Force (Fmax); for the factor "Sampling Position", the colorimetric indices contribute more, and for the factor "Time", the textural parameters are the most important in determining the overall variance together colorimetric index b*.

For the factor "Dairy Factory", the results of the ASCA multivariate analysis show that 96% of explained variance is described by the first four principal components, as shown in the scree plots in Figure 6.

The first component describes 36.5% of the total variation. This component describes mostly the variation from the overall mean for the colorimetric component a*, then the variation for the component L*, and then the anticorrelated variation to those colorimetric properties of all the textural parameters.

The biplot in Figure 7, shows that according to the first principal the dairy factories are divided according to the overall value of the textural properties and the value of the colorimetric parameters L* and a* of their cheese wheels. The dairies labeled C-14 and C-11, which produced cheese ~~are~~ characterized by overall higher values for all the textural properties and less bright color of the grain, are located in the left side of the plot; The dairies C-1, C-2, C-3, C-6, C-7, C-8, C-9, C-13, and C-15, which produced wheels with textural parameters and L* and a* colorimetric parameters not significantly different from the overall mean, are placed in the central part of the plot; the remaining dairies (C-4, C-5, C-10, C-12) produced cheese wheels having textural parameters lower than the overall mean and higher values of L* and a* colorimetric values than the overall mean.

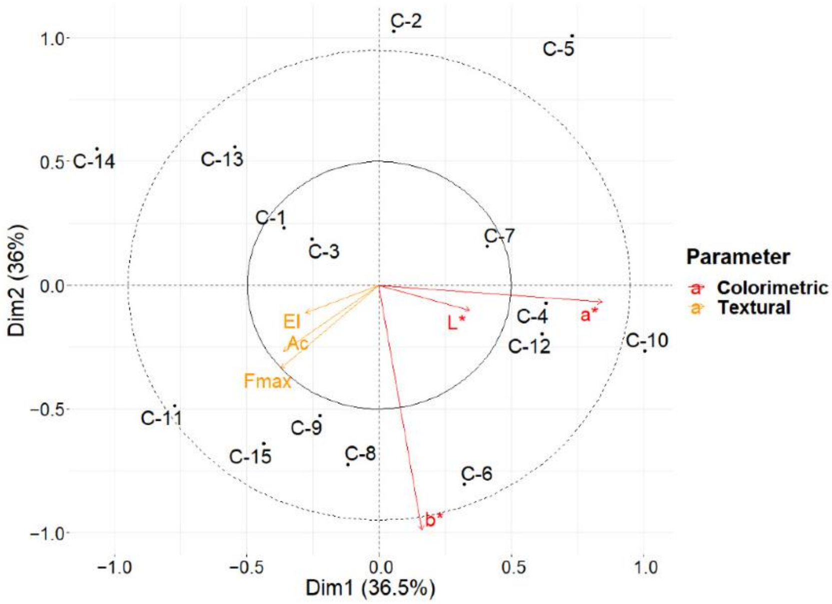

Figure 7: *Biplot for ASCA showing the values for the first two principal components for the LMM-ASCA decomposition to the factor "Dairy Factory". Score values for each level are represented by black dots and labeled from C-1 to C-15 for each dairy factory. Loading values are represented using arrows for each parameter acquired and colored according to the type of the measurement: "El" stands for elastic modulus; "Ac" for the area under the curve; "Fmax" for maximum force; "L*", "a*", and "b*" labels are referred to the coordinates of the Lab color spaces.*

The variation related to the colorimetric index b* is represented by the second component, which accounts for 36% of the total variance. The dairies C-2, C-5, C-13, and C-14 are those producing cheese wheels with lower values of the colorimetric parameter b* than the overall mean, consequently they are placed on the upper side of the plot; in the central are of the plot, between the score values of −0.55 and 0.55, there are the dairies C-1, C-3, C-4, C-7, C-9, C-10, C-11, and C-12, which produced cheese wheels having an average value of b* not significantly different from the overall mean; the remaining dairies, labeled C-6, C-8, and C-15, produced cheese wheels having an average value of b* higher than the overall mean, and are locatd in the lower part of the graph.

The first two principal components show that there are four dairy factories that differ from the overall mean by more than 95% of the explained variance, which are C-2, C-5, C-10, and C-14.

The third principal component describes the 18.1% of the overall variance (Figure 8) and is determined mostly by the values of all the physical parameters, with the colorimetric b* index in anticorrelated position with respect to all the other parameters. This component divides three dairies (C-1, C-3, C-7) from all the others, for their overall lower values of textural and colorimetric parameters, except b*.

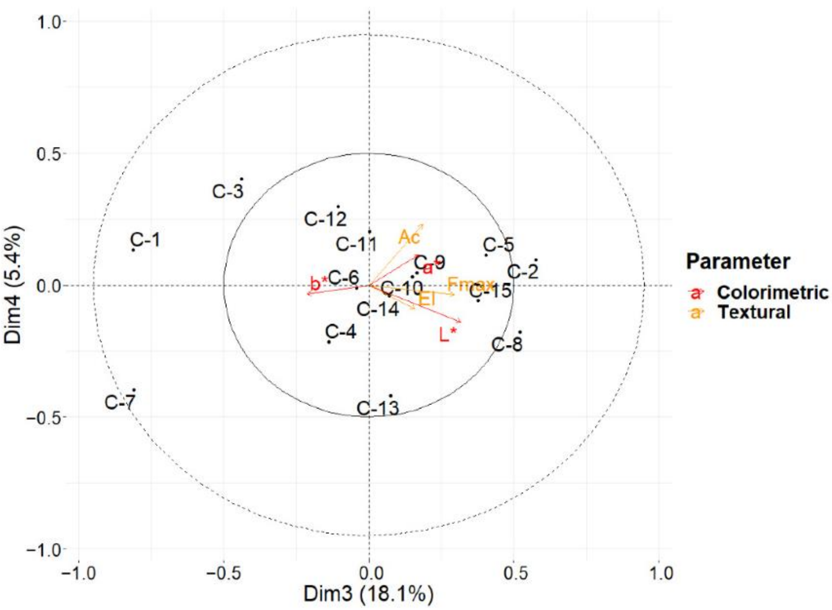

Figure 8: *Biplot for ASCA showing the values for the third and fourth principal components for the LMM-ASCA decomposition to the factor "Dairy Factory". Score values for each level are represented by black dots and labeled from C-1 to C-15 for each dairy factory. Loading values are represented using arrows for each parameter acquired and colored according to the type of the measurement: "El" stands for elastic modulus; "Ac" for the area under the curve; "Fmax" for maximum force; "L*", "a*", and "b*" labels are referred to the coordinates of the Lab color spaces.*

The fourth principal component explains only 5.4% of the total variance, and it describes the variation mostly related to the textural index of the area below the curve and, in a lower amount, to the a* colorimetric index. It divides mostly the dairies C-2 and C-5 from the overall mean. These latter two principal components explain 23.5% of the overall variance and highlight another interesting cluster including four dairies C-6, C-9, C-10, and C-14.

To estimate the presence of cluster inside the ASCA decomposition for this factor, which, differently than the other two analyzed, has many different levels and more structured results, a k-means cluster analysis has been performed on the score value matrix of the LMM-ASCA decomposition. The clustering procedure was performed using the Hartigan-Wong algorithm reported by Hartigan & Wong (1979); the optimal number of clusters was estimated using the Silhouette method as reported by Kassambara (2017b). Results are summarized in Figure 9.

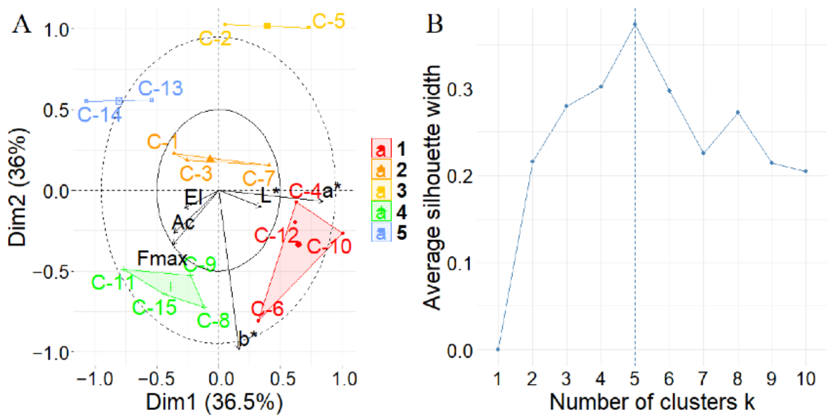

Figure 9: *Results of the clustering analysis on LMM-ASCA decomposition of the "Dairy Factory" factor; (A) representation on clusters; (B) results of the Elbow method for the estimation of the optimal number of clusters. Score values for each level are represented by black dots and labeled from C-1 to C-15 for each dairy factory, and colored according to the cluster estimated. Loading values are represented using arrows for each parameter acquired: "El" stands for elastic modulus; "Ac" for area under the curve; "Fmax" for maximum force; "L*", "a*", and "b*" labels are referred to the coordinates of the lab color spaces.*

Cluster analysis highlights the presence of two big clusters, named 1 and 4 both containing four dairies. Then it is estimated that cluster 2, which contains three different dairies with values near to the center of the plot (C-1, C-3, and C-7), and two small clusters, that represents the presence of the dairy factories have values very different from the overall mean (C-2 and C-5 in cluster 3; C-13 and C-14 in cluster 5).

Differences in the properties of the final product can be explained by differences in the raw material and the technological process: together with the quality of raw milk used, the type and quantity of rennet and natural whey starters, the handling of the curd, the operations of acidification, heating, cooling, and the ripening condition and time, all these factors are able to affect the quality of the final product. Even though the production of Trentingrana cheese follows a general standardized procedure and the Trentingrana cheese product specification reports the crucial technological steps, it is a semi-artisanal production, with a significant internal variability related to the different dairy.

The clusters estimated can be partially explained because of important similarities in the technology of the process: The dairy C-1 and C-3 use both the same kind and the same quantity of rennet, and they have similar values of time and temperature in the procedure of the heating of the curd (data not shown). Unfortunately, many other similarities are quite complex to interpret, because they do not correspond to other know properties of the dairy factories.

Differences between dairy factories belonging to the same consortium have already been detected by Franceschi *et al.* (2019) for the efficiency of the process, and Mucchetti *et al.* (2014) noticed the effect of slight variation inside the production process of extra-hard cheese as an important source of variation for the appearance and the structure of the cheese; our research also highlighted differences in the physical properties of the cheese wheels that can be used as quality indices.

For the factor "Time", ASCA decomposition represents 87.1% of total variance with the first two principal components. As shown in the biplot in Figure 10, the first PC describes variations in all the colorimetric parameters, with b* in an anticorrelated position with a* and L* ~~and b*~~, and, with less significance, the textural parameters elastic modulus and maximum force applied. The second principal component describes variation mostly due to the colorimetric index b*, the area below the curve (Ac) and the maximum force applied (Fmax). Along with the first component, the wheels sampled in May and June are characterized by brighter and thicker grain than the average; the wheels sampled from January to April, in November and December, and in July and August are characterized by medium values of textural and colorimetric properties, and the wheels sampled from September to October show higher values of the textural parameters and more yellow color than the average. The second component separates the wheels produced from November to December from all the others because of a significantly higher value of the colorimetric index b* that can be interpreted as a more yellowish color than the average.

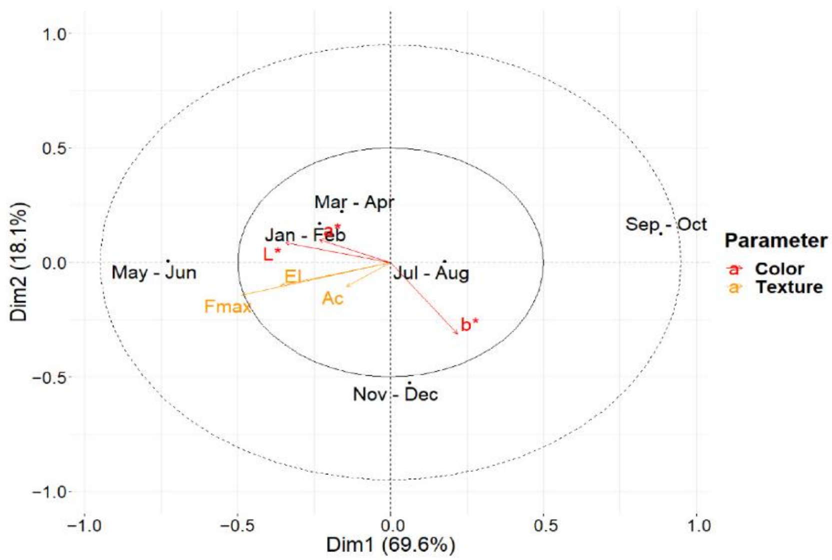

Figure 10: *Biplot for ASCA showing the values for the first two principal components for the LMM-ASCA decomposition to the factor "Time". Score values for each level are represented by black dots, labeled using the first three letters of the two months of each couple of months when the cheese wheels were sampled. Loading values are represented using arrows for each parameter acquired and colored according to the type of the measurement: "El" stands for elastic modulus; "Ac" for area under the curve; "Fmax" for maximum force; "L*", "a*", and "b*" labels are referred to the coordinates of the Lab color spaces.*

Differences and similarities found in this analysis are likely to highlight the effect of seasonal variation in the raw milk used, which causes a change in the overall content of protein (data not shown). The overall lower values of density of the last two couples of months (November and December) may partly be due to the design of the sampling procedure. Indeed, the cheese wheels produced over that period were seasoned 3 weeks less than the others to ease the logistic organization of the sampling procedure within the consortium.

For the factor "Sampling Position", multivariate ASCA estimated 96.4% of the explained variance with two principal components. The first component, as reported in the biplot in Figure 11, describes the overall variation related to all textural and colorimetric parameters, with a strong anticorrelation between the colorimetric parameter L* and all the other measurements.

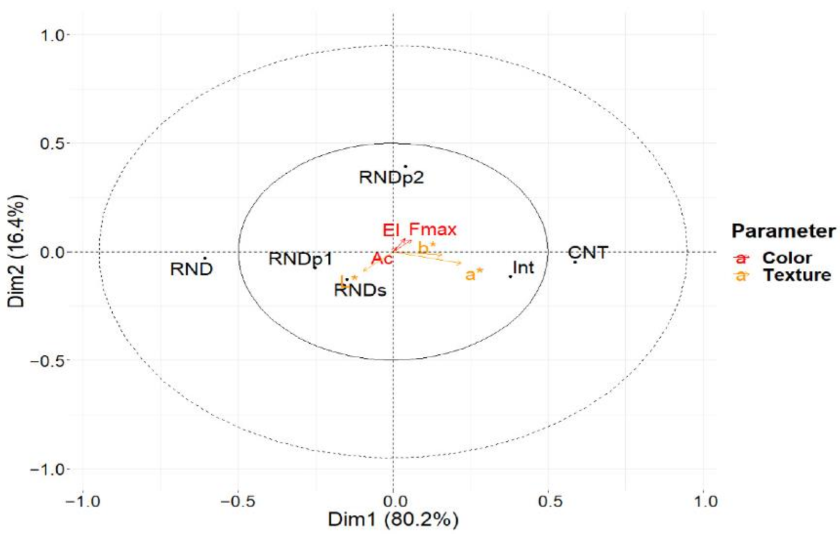

Figure 11: *Biplot for ASCA showing the values for the first two principal components for the LMM-ASCA decomposition of the factor "Sampling Position". Score values for each level are represented*

by black dots, each label corresponds to a different sampling position: "RND": round; "RNDs": round central; "RNDp1": external plate; "Int": intermediate zone; "RNDp2": internal plate; "CNT": center. Loading values are represented using arrows for each parameter acquired and colored according to the type of the measurement: "El" stands for elastic modulus; "Ac" for area under the curve; "Fmax" for maximum force; "L", "a*", and "b*" labels are referred to the coordinates of the Lab color spaces.*

In this component, the sampling position levels are placed from left to right according to higher values of texture and a darker color. This disposition corresponds to a decreasing distance from the center of the cheese wheel.

The second component describes mostly variation in the colorimetric indices and the area under the curve. According to this component, all the levels but one do not have a significant variation from the grand mean, except for the intermediate position between the central zone and the round of the wheel, which results in higher values for each index. This disposition in the biplot suggests that the gradient highlighted by the first component is not linear and that the intermediate position has higher values than those expected according to the variation of the distance from the center.

Differences due to the sampling position inside the cheese wheels are well known (Prentice 1992), they are due to the inhomogeneous content of water, protein, and fat in the casein structure, related to the diffusion phenomenon during curing and seasoning of the cheese wheel. Water, which spreads from the center to the external zone of the wheel when the evaporation phenomenon begins, acts as a low-viscosity lubricant, reducing overall texture properties and as a solvent for coloring molecules, reducing their concentration, hence the color brightness decrease (Milovanovic *et al.* 2020). The color difference is mostly related to the concentration of products of the Maillard reaction naturally occurring in cheese (Bley *et al.* 1985): the central zone of the cheese wheel maintains a temperature above 50 °C longer than the external zones, therefore more chromophore molecules are formed during and after the heating of the curd (Adachi et al. 2020).

3.2.4. Conclusions

Significant differences have been detected among the dairy factories of the Trentingrana consortium, the time of the year, and the sampling positions. Analysis of significant differences in the context of real-scale quality control at an aggregate level allows for detecting which are the phenomena that may modify the physical properties of cheese, to allow improvements in the production process, and most importantly, to focus on what are the features that affect the production process and to understand them better.

Those results are related to the effect of variations of the process of production on the physical properties of the final product and can be detected when changes in the three factors considered (dairy factory, time, and sampling position) have an overall significant effect. The differences in the percentage of explained variance by each different factor evidenced the importance of the sampling position as a critical point in the evaluation of cheese wheels: the results highlighted a large effect due to sampling position, especially on color values. Research on these products should consider the importance of the geometry of the cheese wheel to characterize its overall properties.

Differences detected for the dairy factory factor are more difficult to interpret. Results show that there are similarities and differences related to the production process, as resulted from information given by the consortium. At the same time, raw material properties are not always easily related to the final product, and this may also be explained by the importance of the technologies applied for the caseification process, as, for example, the use of curd in liquid or in powder form, or a mixed combination, its coagulation activity, and the gradient of temperature during the heating of the paste.

Differences related to the time when the cheese wheels were produced are likely related to the seasonal variation of the raw milk, which tends to have different protein content and a different concentration of beta-carotene depending on the season and to a different diet, causing variations in the textural and colorimetric properties.

Quality evaluation of hard seasoned cheese requires many different measurements to define all its properties, from physical to sensory. In this study, LMM-ASCA procedures allow for analyzing the effect of multiple different factors comparing many measurements at the same time, giving a statistically valid comparison of the whole profile of the product between different subgroups in complex or nested experimental designs.

The results of those analyses are functional to the development of classification models for the quality monitoring of the production of hard seasoned cheese and to evaluate the effect of the process of production on the commercial quality of the product Trentingrana Cheese.

At this moment, further research is needed to better understand the relationship between the variation in the production process and the physical quality of hard seasoned cheese.

CHAPTER 4. STATISTICAL ANALYSIS FOR QUALITY CONTROL OF CHEMICAL PROFILE

In the present chapter, the importance of chemometrics measurements is presented and discussed, and a statistical model to interpret VOCs data from quality control procedure is presented and discussed. The analytical procedure proposed was published in Ricci *et al.* (2023).

4.1.Importance of Chemical measurements

4.1.1. Chemical measurements

Chemical properties of food are the most common procedures to estimate the properties of a food product and to develop quality control procedures. From traditional chemical essays, since the 2000s the application of new technologies like mass spectrometry and near infra-red spectroscopy allowed to collect a large amount of information that are related to different chemical properties at the same time. The adoption of these procedures represented a significant improvement for the quality control procedure and helped to develop a new paradigm of fingerprinting of the products. Fingerprinting consists in collecting at the same time multiple measurements related to different properties of the product and developing a multivariate approach to create a reliable procedure to analyze them. For a broader investigation on the adoption of Volatile organic compound profile for quality control, we recommend the work of Lin *et al.* (2022).

4.2. Application 2 – Study of the effect of the production process on the Volatile Organic Compounds composition

4.2.1. Introduction

Volatile organic compounds (VOCs) are molecules characterized by high vapor pressure at room temperature and low water solubility. Several VOCs in food contribute to odors and flavors that play a key role in sensory quality perception and liking responses (Khattab *et al.* 2019, Liaw *et al.* 2011). In cheese, VOCs are produced to a great extent during ripening by the catabolic activity of microorganisms on carbohydrates, lipids, and proteins naturally present in milk and rennet (Kilcawley *et al.* 2018; Marilley & Casey, 2004; McSweeney & Sousa, 2000; McSweeney, 2004). The metabolic pathways responsible for the synthesis of VOCs are affected by the properties of raw milk and the conditions of the production process. For this reason, VOCs are considered reliable markers of process quality and traceability of cheese products (Pisano *et al.*, 2016; Suh, 2022).

Trentingrana dairy factories differ in the adoption of a double or a single milk collection procedure (Endrizzi *et al.*, 2012), which determines differences in storage time, temperature conditions of the raw milk before transformation, and intensity of the milk skimming process. This latter process decreases bacterial and somatic cell counts by natural gravity separation of fat, thus standardizing the properties of fat and casein/fat ratio (McSweeney *et al.*, 2004). The effects of the milk collection procedure, the skimming process, and storage temperature on the sensory and chemical properties of Trentingrana cheese have already been studied. Endrizzi *et al.* (2012) found differences in physical properties and sensory quality attributes in cheese wheels produced in pilot plants with different milk collection procedures and in different seasons, showing significant differences in commercial quality, colorimetric properties, and the VOC profile. Franciosi et al. (2011) reported the effect of the milk collection procedure on chemical and VOCs composition in Trentingrana cheese showing a higher content of free fatty acids and related esters in cheese wheels produced using double collection without refrigeration. In a similar study, Fabris *et al.* (2010) trained a random forest classifier to recognize the milk collection procedure from the VOC content in Trentingrana cheese and to highlight which molecules are determinants for discriminating the cheese wheels produced in different seasons.

Monitoring the VOC profile of Trentingrana in its real-scale production process is functional to understanding how to operatively improve its quality: associating the presence of chemical compounds to a production process condition or a feature of the final product allows to develop a faster quality control procedure and to estimate how the issue studied is related to the chemical properties of food (Ellis & Mayhew, 2014).

Previous studies were done using a restricted batch of samples from pilot plants with a balanced experimental design to study the factors of interest, and thus excluding at multivariate level the effect of the other factors that may influence the final quality of the product in a real production context. Overall, previous results highlighted that there exists a need to develop a large-scale monitoring of the chemical properties of Trentingrana cheese, to estimate the significance and the importance of the factors investigated in previous experiments in the real context of the production process. Because of the presence of many factors and the multivariate structure of VOC data, there is also the need to develop a functional and reliable statistical procedure to infer the effect of the process-related factors at multivariate level removing the nuisance due to confounding factors.

The objective of this work is to test the suitability of ANOVA-Simultaneous Component Analysis (ASCA) to extract useful information from volatile organic compound data measured in Trentingrana cheese in a real production context where several confounding factors are present, and no experimental design is designed focusing on predefined a priori factors. To this end, within the collaboration with Trentingrana – Consorzio dei Caseifici Sociali Trentini (Italy), the VOC profile of Trentingrana cheese over two years of production was analyzed by SPME/GC-MS, sampling a representative selection of cheese wheels from its real-scale production process. We estimated how VOC profiles are related to distinct raw materials, different processing of the different cheese factories, the enzymatic activity of the rennet adopted, and different parts of the year when the milk is produced.

A two-step analytical process is presented: first, for the estimation of the significance and the effect size of two process-related factors (Dairy Factory and Time of the year when the wheels are produced), an ANOVA-Simultaneous Component Analysis was adopted. Analyzing the results, an overall tendency in the VOC content between different dairy factories was detected. The effect of the milk collection procedure adopted by the dairy factory is proposed as an interpretation for this tendency. To test this hypothesis and to estimate its effect, an O-PLS-DA predictive model was trained and validated (Smilde *et al.* 2012, Trygg & Wold, 2002).

4.2.2. Materials and methods

4.2.2.1 Sampling procedure

A total of 317 cheese wheels were bi-monthly sampled from the Trentingrana Consortium repository (located in the Autonomous Province of Trento, Northeast Italy) during the years 2017 and 2018. The wheels were produced from November 2015 to October 2017 by 15 different dairy factories located in the Province of Trento (Italy), ripened for 18 ± 2 months, and labeled progressively from C-1 to C-15 (production traits of each dairy are summarized in Table 5). For each dairy factory, the number of samples varied in proportion to the volume of cheese wheels delivered from a minimum of 1 cheese wheel and a maximum of 3. For more details on sampling criteria, see Ricci *et al.* (2022). The samples for the analysis of VOCs were prepared by taking 24 parallelepipeds of cheese (3 × 1.5 × 1.5 cm) from various positions of the freshly cut half part of the wheel then finely grounded and well mixed.

Approximately 3 g of grounded cheese were weighted inside a 20 mL GC-MS vial (Supelco, Bellefonte, CA, USA), capped with PTFE/silicone septa (Supelco, Bellefonte, CA, USA), and stored at −80 °C. For each cheese wheel, three vials from the same mixing were prepared. Before the analysis, samples were thawed for 1 h at room temperature, then each vial was spiked with the internal standard, just before the beginning of the analysis. Each sample was classified according to the dairy factory where it was produced, the part of the year when the milk was collected, and the milk collection procedure adopted for its production. At the beginning of the sampling procedure, from one of the first cheese wheels, 100 vials were prepared with the same grounded cheese mix and stored at −80 °C than used as quality control (QC sample) during GC-MS analysis over time.

Table 5: *Production traits across dairy factories of the Trentingrana consortium.*

Dairy Factory	Number of farms associated to the dairy	Percentage of farms using unifeed alimentation procedure	Number of cheese wheels produced during production year 2015/16	Number of cheese wheels produced during production year 2016/17	Milk collection procedure adopted
C-1	*17*	*58.8*	*14476*	*14929*	*Double*
C-2	*18*	*16.7*	*1207*	*1534*	*Mixed*
C-3	*51*	*3.9*	*9882*	*10415*	*Double*
C-4	*38*	*26.3*	*11881*	*10537*	*Single*
C-5	*10*	*30*	*6145*	*6296*	*Double*
C-6	*59*	*6.8*	*3010*	*2387*	*Single*
C-7	*38*	*10.5*	*6695*	*7390*	*Single*
C-8	*47*	*4.3*	*5286*	*5900*	*Double*
C-9	*46*	*0*	*5119*	*5407*	*Double*
C-10	*12*	*8.3*	*1985*	*3721*	*Double*
C-11	*78*	*1.3*	*7072*	*7795*	*Double*
C-12	*95*	*1.1*	*7961*	*8609*	*Double*
C-13	*9*	*0*	*5137*	*5028*	*Single*
C-14	*72*	*2.8*	*8287*	*9183*	*Double*
C-15	*27*	*0*	*2635*	*1801*	*Single*

In dairy factories where double milk collection (DMC) procedure is adopted, the full-fat milk of the evening milking is delivered to the cheese factory and undergoes a gravity separation process overnight in large vats (Ma & Barbano, 2000). After that, the milk of the morning milking is added to the semi-skimmed milk and used to produce cheese according to the standard cheese-making procedure of the Trentingrana. The single milk collection (SMC), instead, consists in storing the milk of the morning milking at the dairy farm in controlled conditions and then adding the evening milk, before moving the raw milk to the cheese factory, where the skimming procedure takes place overnight. Samples produced in the dairy factory labeled C-2 were classified in a third-class called mixed milk collection (MMC) because in that specific dairy factory both collection procedures are used according to the farm where the milk is collected.

4.2.2.2 SPME/GC-MS analysis

The procedure for Headspace solid-phase microextraction coupled with gas chromatography-mass spectrometry (SPME/GC-MS) was performed according to Endrizzi et al. (2012) with a few amendments. The samples were equilibrated at 40 °C for 30 min, and then in the headspace environment, a fused silica fiber coated with 2 cm of 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS, Supelco, Bellefonte, PA, USA) was inserted and exposed for 30 min without changing the temperature. The desorption of the volatile compounds from the SPME fiber was performed at 250 °C for 5 min in the injector port of a GC-MS operating in electron ionization mode (EI, internal ionization source; 70 eV). The control of the procedure phases was managed using an auto-sampling system (CTC combiPAL, CTC Analysis AG, Zwingen, Switzerland) equipped with a cooling system that kept the vials at 4 °C before the start of the analysis. Separation was conducted on an HP-Innowax fused silica capillary column (30 m, 0.32 mm ID, 0.5 μm film thickness; Agilent Technologies, Palo Alto, CA, USA). Separation conditions were as follows: carrier gas was helium at a constant flow rate of 2 mL/min; oven temperature programming was 40 °C for 3 min, an increase from 40 to 180 °C at 4 °C/min, stationary at 180 °C for 6 min, then another increase from 180 to 220 °C at 5 °C/min and finally, 220 °C for 3 min. The mass spectrometer operated a mass scan range from 33 to 300 m/z (GC Clarus 500, PerkinElmer, Norwalk CT, USA).

Compound identification was based on mass spectra matching with those present in the standard NIST14 (NIST/EPA/NIH, 2014) library and linear retention times calculated injecting C7–C30 n-hydrocarbon series under the same chromatographic conditions. Compounds were semi-quantified spiking samples with 4-methyl-2-pentanone (Sigma-Aldrich) as I.S. 0.05 g/L in aqueous solution. Amount of VOCs in the samples were expressed as μg/kg equivalent of the I.S.

The analytical measurements were performed over a period of two months and required four different batches of SPME fibers to overcome the decline of the performances due to the deterioration of the fiber itself. The repeatability of the method was assessed for each batch of SPME fiber, analyzing twelve replicates of a reference cheese on the same day. The observed average variation, estimated for the classes of acids, esters, ketones, and aldehydes agreed with the literature for SPME analysis with this type of matrix (Barbieri *et al.*, 1994; Bellesia *et al.*, 2003, results in table 6). Furthermore, a QC sample was measured every ten cheese samples over all the period of measurements.

Table 6: *Repeatability index for different classes of molecules for each SPME fiber adopted in SPME/GC-MS analysis. Values are estimated from 12 repeated analysis on a reference cheese sample.*

SPME fiber	Acids	Aldehydes	esters	ketones
1	14%	12%	6%	8%
2	27%	26%	15%	25%
3	37%	28%	10%	34%
4	22%	25%	6%	20%

4.2.2.3 Statistical analysis

VOCs data were analyzed using ANOVA simultaneous component analysis (ASCA, Smilde et al., 2012) to identify multivariate patterns significantly associated with the different study factors: Dairy Factories, Time of the year, and their interaction, and the effect of the batch of SPME fiber. The ASCA model decomposes the signal of each volatile x in the following form:

$$x_{jkni} = \mu + \alpha_j + \beta_k + \gamma_n + (\alpha\beta)_{jk} + \epsilon_i$$

Where μ represent the overall mean of the volatile compound, α_j the expected value for the jth Dairy Factory, β_k the expected value for the kth Time of the year, γ_n the expected value for the nth batch of SPME fibers, $(\alpha\beta)_{jk}$ the interaction between the Dairy Factory and Time and ϵ_i the residual error for the ith cheese wheel representing the natural variability of each cheese wheel. The effect of the batch of SPME fibers is considered a known and controllable nuisance source of variability and it is integrated into the model as a blocking factor (Montgomery, 2013).

A permutation test (n = 1000) was applied to assess the univariate statistical significance of each factor for each volatile compound estimating empirical null distributions for the univariate sum of squares ($\alpha = 0.05$).

An Orthogonal Partial Least Squares – Discriminant Analysis (O-PLS-DA, Trygg & Wold, 2002) classifier was developed to deeply investigate the Milk Collection procedure adopted to produce each cheese wheel from its VOC profile.

The model was built using a restricted dataset of the 14 dairy factories that use single or double milk collection (n = 306, prevalence of DMC = 70.9%), to analyze only the two most prominent modalities of the milk collection process. Data were partitioned into train and test sets (train/test ratio = 0.8) and a cross-validation procedure was performed in the train set to estimate the optimal number of orthogonal projections. To assess the predictive capacity of the model, the results from repeated partitions were compared with a null distribution obtained by a permutation test (n = 1000) in terms of sensitivity, specificity, and overall accuracy (using Cohen's Kappa index, Ferri *et al.* 2009).

To identify which VOCs were related to the variation of the Milk Collection Procedure, a bootstrap procedure (n = 1000) was employed to estimate the confidence intervals and the significance of the regression coefficient of every VOC included in the model (Lazraq et al. 2003). The validation of the coefficients of the model identified the VOCs affected by the different milk collection procedure. This procedure of testing for significance considers the dimensionality and the structure of the data as it is modeled by O-PLS-DA and does not require standard statistical assumptions.

4.2.3 Results and discussions

4.2.3.1 <u>Qualitative VOCs assessment</u>

A total of 75 volatile organic compounds have been identified by SPME/GC-MS analysis. These compounds belong to the following chemical classes: esters (n = 17), alcohols (n = 13), ketones (n = 11), acids (n = 9), aldehydes (n = 8), sulfurs (n = 5), hydrocarbons (n = 4), phenols (n = 3), lactones (n = 2), terpenes (n = 2) and pyrazines (n = 1). Overall results are summarized in table 7.

Table 7: *Overall values for SPME/GC-MS measurements for each volatile organic compound detected in Trentingrana Cheese expressed in µg/kg equivalent of i.s., each measurement has been carried out in triplicate. The values of the linear retention index are obtained from NIST 14 database (NIST/EPA/NIH 2014).*

Compound Category	Compound Name	Minimum	Mean	Maximum	Retention Index Estimated	Retention Index NIST
Acids	*acetic acid*	30.21	222.45	633.28	1529	1449
	propionic acid	0.00	50.76	876.65	1594	1535
	butanoic acid	85.66	682.33	5040.80	1689	1625
	2-methyl butanoic acid	0.00	0.72	12.11	1759	1662
	hexanoic acid	78.00	870.81	9791.60	1914	1846
	heptanoic acid	0.00	5.44	85.49	2049	1950
	octanoic acid	25.84	232.10	3846.18	2127	2060
	nonanoic acid	0.00	7.74	51.94	2211	2171
	decanoic acid	5.87	48.29	678.66	2339	2276
	Total Acids mean	235.63				
	Total acids Standard deviation	472.30				
Alcohols	*2-propanol*	1.14	6.66	43.26	932	927
	ethanol	6.79	1030.07	4455.16	940	932
	2-butanol	0.00	6.75	230.39	1035	1025
	2-methyl, 1-butanol	6.19	35.68	159.91	1142	1119
	1-butanol	0.00	9.11	69.39	1164	1142
	3-methyl 1 butanol	0.00	13.35	72.21	1222	1209
	1-pentanol	0.00	1.36	6.85	1261	1250
	3-methyl, 3-buten 1-ol	2.11	9.14	24.43	1261	1248
	2-heptanol	0.00	10.08	70.88	1330	1320
	prenol	1.27	5.06	11.31	1332	1320
	hexanol	0.00	6.48	66.12	1364	1355
	2-ethyl hexanol	0.00	1.00	10.40	1501	1491
	2-nonanol	0.00	0.74	35.89	1531	1521
	Total alcohols mean	87.34				
	Total alcohols Standard deviation	522.54				
Aldehydes	*2-methyl butanal*	1.47	6.92	22.06	919	914
	3-methyl butanal	5.45	23.81	61.35	922	918

Compound Category	Compound Name	Minimum	Mean	Maximum	Retention Index Estimated	Retention Index NIST
	hexanal	0.00	2.20	7.31	1100	1083
	3-methyl 2-butenal	0.00	0.29	2.06	1212	1215
	nonanal	0.00	1.71	19.78	1402	1391
	decanal	0.00	0.38	3.29	1510	1498
	benzaldehyde	0.00	3.96	36.74	1533	1520
	phenyl acetaldehyde	0.73	4.24	14.55	1654	1640
	Total Aldehydes mean	5.44				
	Total aldehydes standard deviation	8.61				
Esters	*ethyl acetate*	1.27	21.13	110.65	900	888
	ethyl propanoate	0.00	4.70	145.03	965	953
	isopropyl isobutanoate	0.00	32.12	188.59	970	959
	ethyl butanoate	15.96	299.17	1668.76	1049	1035
	2-methyl ethyl butanoate	0.00	0.01	0.50	1062	1051
	butyl acetate	0.00	0.87	12.00	1093	1074
	ethyl valerate	0.00	1.53	9.76	1147	1134
	butyl butanoate	0.00	0.95	20.41	1229	1220
	ethyl hexanoate	6.04	240.55	1777.48	1244	1233
	isoamyl butanoate	0.00	0.77	9.86	1273	1259
	butyl pentanoate	0.00	0.00	0.19	1324	1310
	propyl hexanoate	0.00	0.13	8.98	1329	1316
	isopentyl hexanoate	0.00	0.00	0.31	1423	1451
	butyl hexanoate	0.00	0.07	2.79	1423	1408
	ethyl octanoate	0.44	21.28	203.50	1444	1435
	2-hydroxy, 4-methyl, methyl pentanoate	0.00	0.36	9.16	1481	1513
	ethyl decanoate	0.00	4.17	32.61	1649	1638
	Total esters mean	26.16				
	Total esters standard deviation	107.10				
Hydrocarbons	*(E) 3-octene*	0.00	0.97	4.65	885	850

Compound Category	Compound Name	Minimum	Mean	Maximum	Retention Index Estimated	Retention Index NIST
	ethyl benzene	0.00	0.94	18.64	1135	1129
	p-xylene	0.00	0.36	5.35	1143	1138
	m-xylene	0.00	0.88	14.04	1148	1143
	Total hydrocarbons mean	0.63				
	Total hydrocarbons standard deviation	1.39				
Ketones	*2-propanone*	3.30	29.61	80.99	882	819
	2-butanone	1.42	13.73	448.25	909	907
	2-pentanone	20.72	117.10	460.19	986	981
	2-hexanone	0.00	6.18	15.84	1099	1083
	3-heptanone	0.00	2.56	12.87	1165	1161
	2-heptanone	87.21	240.30	545.57	1194	1182
	2-octanone	0.00	0.54	6.68	1294	1287
	acetoin	0.00	5.19	60.71	1295	1284
	2-nonanone	7.64	27.49	69.51	1397	1390
	2-undecanone	1.47	5.37	11.76	1608	1598
	acetophenone	0.00	0.36	6.24	1661	1647
	Total ketones mean	37.37				
	Total ketones standard deviation	77.97				
Lactones	*butanolactone*	0.00	0.71	7.28	1637	1632
	delta decalactone	0.96	3.12	6.88	2162	2194
	Total lactones mean	1.92				
	Total lactones standard deviation	1.64				
Phenols	*phenol*	0.54	1.37	2.63	2021	2000
	4-methyl phenol	0.23	1.03	10.60	2095	2080
	3-methyl phenol	0.75	5.24	18.66	2103	2091
	Total phenols mean	2.55				
	Total phenols standard deviation	2.74				

Compound Category	Compound Name	Minimum	Mean	Maximum	Retention Index Estimated	Retention Index NIST
Pyrazines	*2, 6 dimethyl pyrazine*	0.00	5.61	21.84	1336	1328
Sulfurate Compounds	*methanthiol*	0.00	1.24	6.97	866	692
	carbon disulfide	0.00	2.86	31.88	870	735
	dimethyl sulfide	0.53	3.77	12.90	872	754
	dimethyl disulfide	0.00	1.14	5.84	1089	1077
	dimethyl sulfone	0.00	1.75	6.43	1911	1903
	Total sulfurate mean	2.15				
	Total sulfurate standard deviation	2.20				
Terpenes	*α thujene*	0.00	1.39	23.56	1026	1028
	limonene	0.00	4.02	189.39	1201	1200
	Total terpenes mean	1.80				
	Total terpenes standard deviation	10.08				

Identified compounds agreed with the literature on VOCs in grana cheese (Qian & Reineccius, 2002). The most prominent compound type by overall relative concentration is organic acids, followed by ketones and alcohols. Those classes contain several molecules that are directly related to the natural content of raw milk, such as medium-chain fatty acids, and they are naturally occurring in many milk-based products due to lipid catabolism by endogenous enzymes and microbial activity (Collins, McSweeney, & Wilkinson, 2004).

Esters were characterized by a high overall mean but also high variability. These compounds are synthetized from the lipidic fraction by the microbial activity in milk during ripening and they are often associated with positive sensory descriptors in hard seasoned cheese (Liu *et al.* 2004, Qian & Reineccius 2002). High levels of variance in ester compounds were already detected in previous works on Trentingrana cheese (Fabris *et al.*, 2010). Conversely, both terpenes and hydrocarbons were present at low levels with high variability because most of them were not detected in all the samples.

According to literature, these compounds are related to the cows' diet and the seasonal effect and are not naturally occurring in ripened cheeses (Kilcawley *et al.*, 2018). Lastly, phenols, 3-methyl phenol and 4-methyl phenol are mostly related to amino-acid metabolism, however their presence may also be related to the diet and the external environment (Curtin & McSweeney 2004; Panseri *et al.* 2014).

4.2.3.2 <u>ANOVA simultaneous component analysis</u>

The percentage of total variance explained by each factor and interaction was estimated according to Bertinetto *et al.* (2020) by calculating the percentages for each factor of the sum of squares, as reported in Figure 12.

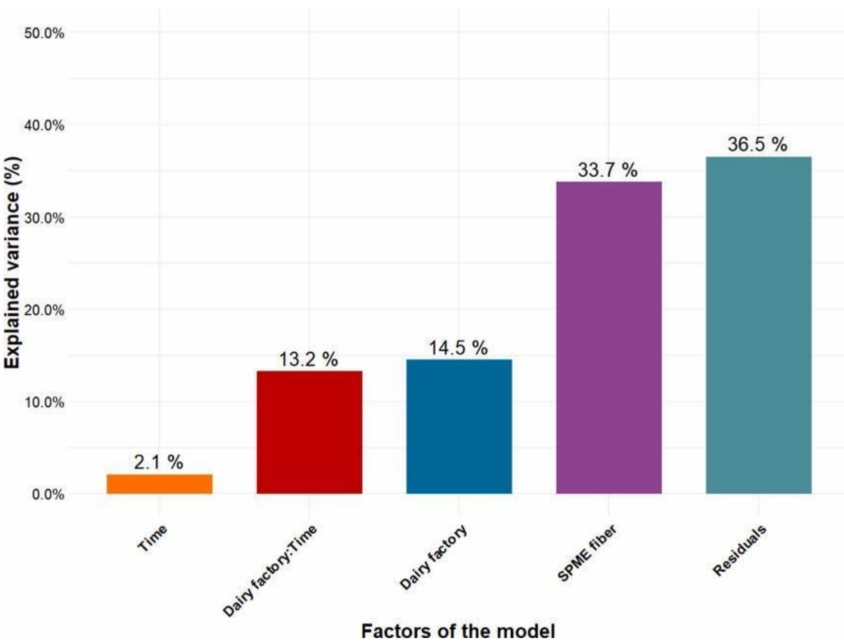

Figure 12: *Barplot reporting the percentage of the overall variance explained by each factor of the ASCA model.*

The high percentage of explained variance related to the effect of the SPME fibers highlighted that there is an important systematic error related to the 4 different batches of fibers used. Even if the effect of a measurement-related bias is important, the ASCA framework allows to analyze the effect of Dairy Factory and Time removing the effect of a potential confounding factor such as the variation of the SPME fiber.

Results of ASCA permutation test showed no significant effects at a univariate level for the interaction of Dairy Factory and Time, while 3 molecules were significantly responding to the factor Time. The Dairy Factory factor was significant for 46 molecules out of 75 and the SPME fiber factor was significant for 55 out of 75 compounds (Figure 13).

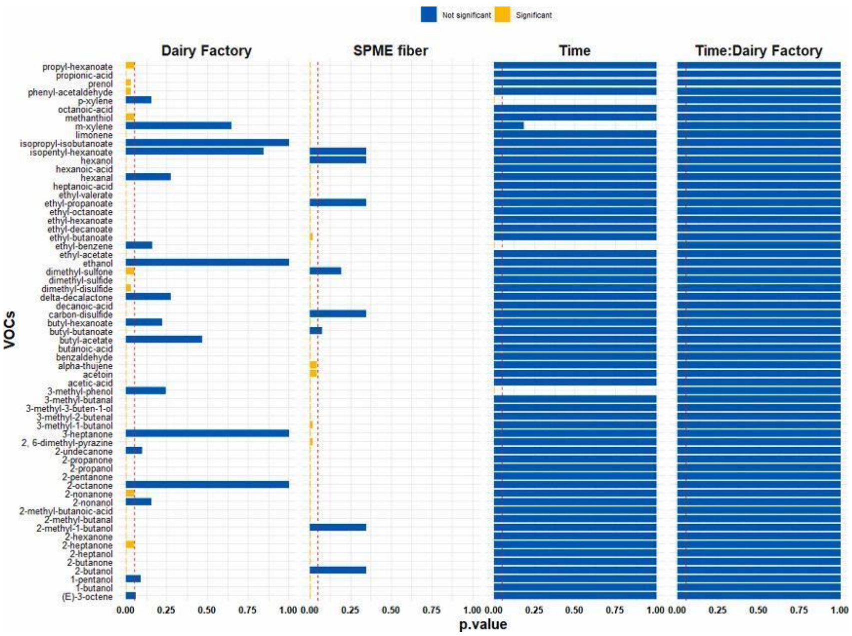

Figure 13: Barplot showing the *significance of the factors of ASCA decomposition reported using p.values estimated using a permutation test for each VOC. The red line indicates the significance threshold of 0.05, results were adjusted using Bonferroni correction.*

Considering that none of the compounds was significantly responding to the interaction factor (Time: Dairy Factory), the multivariate decomposition of this term was not considered.

The Time factor has been included in the model, but it has been analyzed also at a univariate level.

The permutation test demonstrated that the ASCA decomposition with the model proposed is representative of the overall data structure. Results are discussed further below for the factor Dairy Factory and Time of the year, respectively.

To verify the presence of significant factors not included in the model of the multivariate decomposition, the PCA biplot of the residuals was analyzed to estimate the presence of effects not represented in the model (Figure 14).

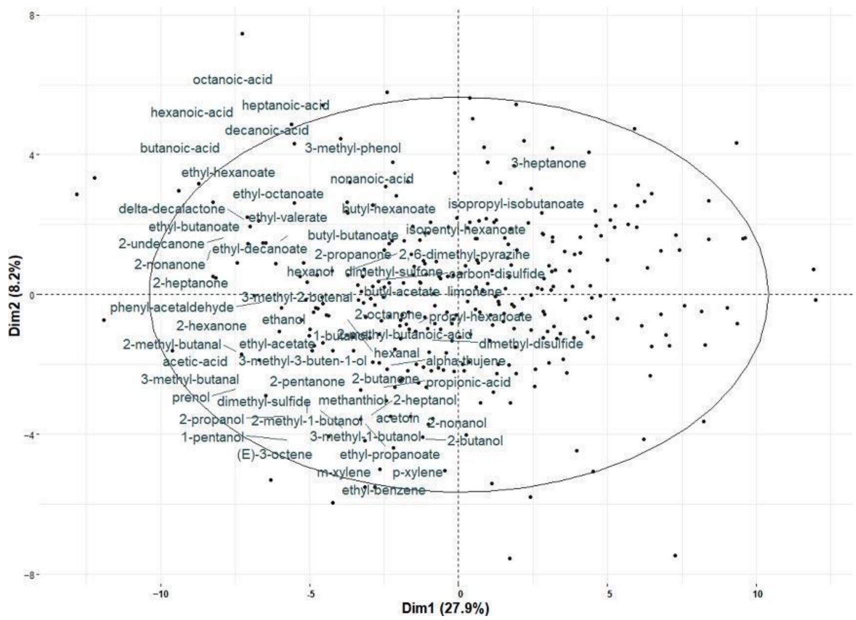

Figure 14: *Multivariate distribution of residuals of ASCA decomposition. Each point corresponds to a single measurement, the ellipsis indicates the interval of confidence at 95%.*

In ASCA decomposition, the Dairy Factory factor was the factor related to the production process that described the largest percentage of explained variance (14.5%, Figure 12), excluding the blocking factor SPME fiber. The variations in the production process adopted in the dairy factory significantly affect the VOCs profile of Trentingrana cheese more than the other factors included in the model.

The results of the multivariate decomposition of the Dairy Factory term of the ASCA decomposition are shown in Figure 15. The biplot indicates that the first two components account for 51.3% of the overall variability. The first principal component separated from left to right all the dairies for the content of organic acids, their esterified form, and 1- and 2- butanol. Along with this component, the samples were separated from the right to left for the content of two ketones (propanone and pentanone), and from left to right for the increasing content of free fatty acids and their esterified forms.

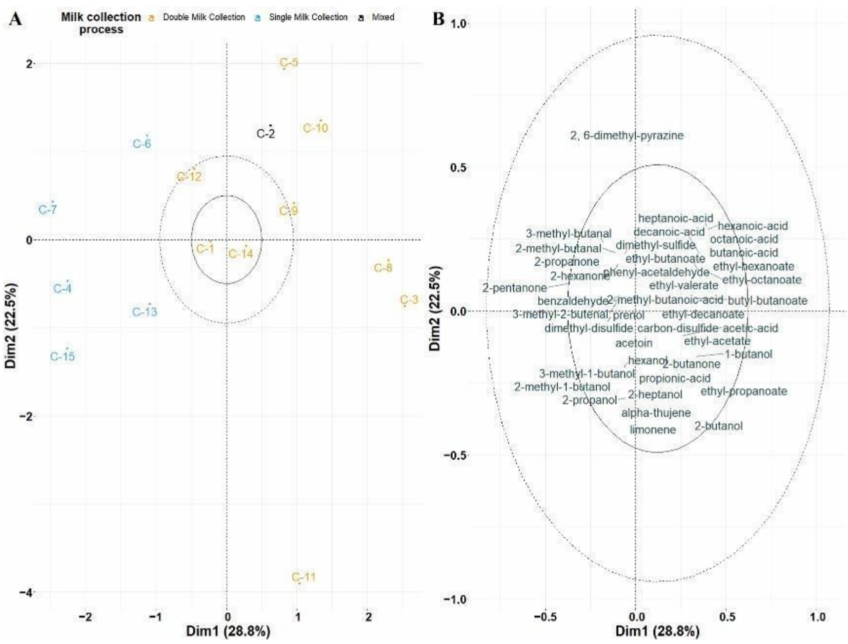

Figure 15: *Results of ASCA decomposition for the factor Dairy Factory.*

In the plot A the score values for each level labeled from C-1 to C-15 for each dairy factory are reported and they are represented by blue, orange, or black dots according to the milk collection procedure they adopt (respectively blue for DMC, yellow for SMC, and black for MMC). In plot B loading values are reported and represented using dark-gray text, only VOCs significantly different are represented. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

On the first component, which explained 28.8% of the overall variance, the dairy factories were distributed in different groups, with three dairies (C-7, C-4, and C-15) placed on the far left, C-6 and C-13 placed on the left side, C-1, C-2, C-5, C-9, C-12, and C-14 placed in the central position of the axis, and C-3, C-8, C-10 and C-11 in the right side of the plot.

The second principal component explained 22.5% of the overall variance and was related to the variation among dairies due to propionic acid, limonene, α-thujene, 2,6-dimethyl pyrazine, and alcohols such as 2-propanol, 2-butanol, and 2-heptanol. Along with this component, the dairy factory C-11 was separated from all the others in the lower part of the graph. This was due mostly to the higher content of limonene, and to the lower content of 2,6-dimethyl pyrazine. Higher levels of limonene have previously been related to the cow's diet (Kilcawley *et al.*, 2018) or to process related contaminants from industrial detergents. Conversely, the formation of 2,6-dimethyl pyrazine is related to the Maillard reaction occurring during milk cooking during cheese production (Divine *et al.* 2012), and to the higher content of propionic acid, which is related to the activity of contaminant microbes, which are associated to the handling of the raw milk and the condition of the production process (Giraffa 2021). The presence of propionic bacteria in Trentingrana was already reported by Rossi *et al.* (2012), who also found significant differences in the microbial activity during ripening between dairies and between different parts of the year.

The formation of free fatty acids and their esterified form is related to the catabolism of triglycerides during ripening (Collins *et al.* 2004). The distinct levels of these molecules between dairies along the first principal component suggest that the concentration is related to the process of the milk collection procedure.

The effect of the dairy factory on the textural and colorimetric properties of Trentingrana cheese was reported in a previous work (Ricci *et al.* 2022). Comparing the results, the dairy factories that were similar for the overall physical properties of their cheese were not similar for the overall VOCs profile. This could be due to the fact that the factors that affect the physical properties and the factors that affect the formation of VOCs in Grana cheese are different: color and texture of cheese are mostly affected by the properties of the raw milk and the treatments of the curd, while VOCs formation is affected also in a large scale by the microbial activity during the ripening process (Divine *et al.*, 2012; Fox *et al.* 2017; Kilcawley *et al.*, 2018).

Noteworthily, the dairy factories differ in the farm producing the raw milk, the heating and storing machinery used, the properties of the whey starter, and the milk collection procedures, which can be double or single. Comparing the results to the information about the dairies available in Table 5, it should be noted that the factor that affects the overall variation of VOCs between dairy factories is the milk collection procedure. Conversely, the average volume of production for year, the number of farms delivering the milk, and the adoption of unifeed alimentation system in the farms are not directly related to the formation of volatile organic compounds in hard seasoned cheese. These results highlight the importance of the milk collection procedure on the overall chemical profile of Trentingrana cheese, coherently with the results reported by Endrizzi et al. (2012).

Overall, the ASCA model highlighted the importance of the milk collection procedure on the VOC profile of Trentingrana cheese and separated the dairy C-11 from the others in the second principal component, detecting that the production process differs in that specific plant, modifying the concentration of other volatile organic compounds. These results indicated that ASCA is a valuable data analysis procedure to recognize significant differences at process levels in large-scale sampling procedures, such as routinary sampling procedures. The relations between volatile compounds and the milk collection procedure is further discussed in section 4.2.3.3.

For the effect of the time of the year, as shown in Figure 12, only 2.6% of the multivariate variance was explained by the factor Time. Hence, considering the small number of significantly different molecules for this factor according to the permutation test (Figure 13), the multivariate ASCA decomposition could be misleading, however it was reported in Figure 16.

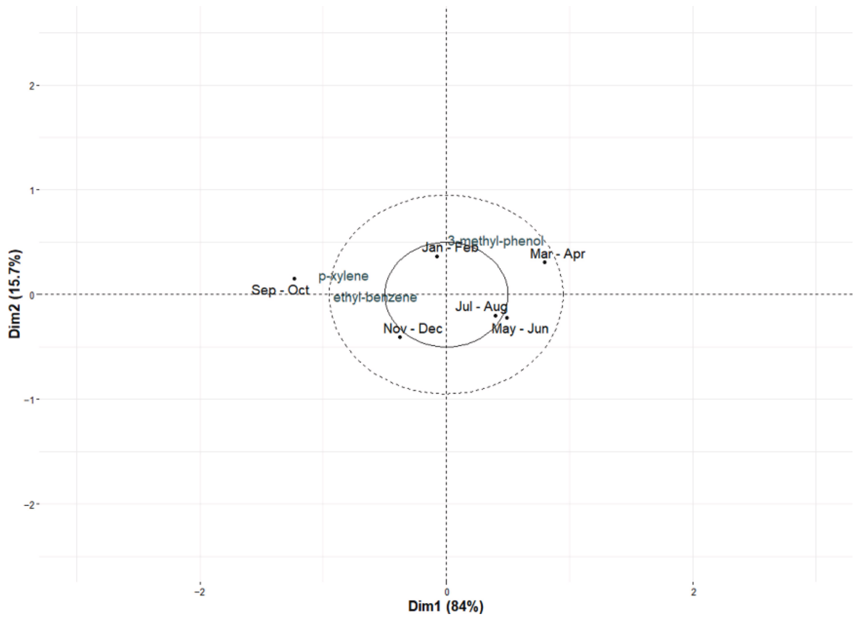

Figure 16: *ASCA decomposition related to the factor Time. Score values for each level are represented by black dots, labeled using the first three letters of the two months of each couple of months when the cheese wheels were sampled.*

The permutation test reports that the three molecules that vary significatively by the production time of the year included 3-methyl phenol (m-cresol), p-xylene, and ethylbenzene.

Ethylbenzene is classified as a pollutant (Panseri *et al.*, 2014), and hence was not included in further analysis.

The effect of the time of the year on these single compounds is reported in the boxplot in Figure 17. The formation of p-xylene and 3-methyl-phenol is associated with the metabolism of aromatic amino acids (Curtin & McSweeney, 2004), and a seasonal effect in their concentrations was reported in raw milk used for cheesemaking by Faustini *et al.* (2019). Post-hoc analysis highlighted that the cheese wheels produced from September to December have a lower content of 3-methyl-phenol than those produced from January to April, and higher levels of ethyl benzene and p-xylene than those produced from March to June.

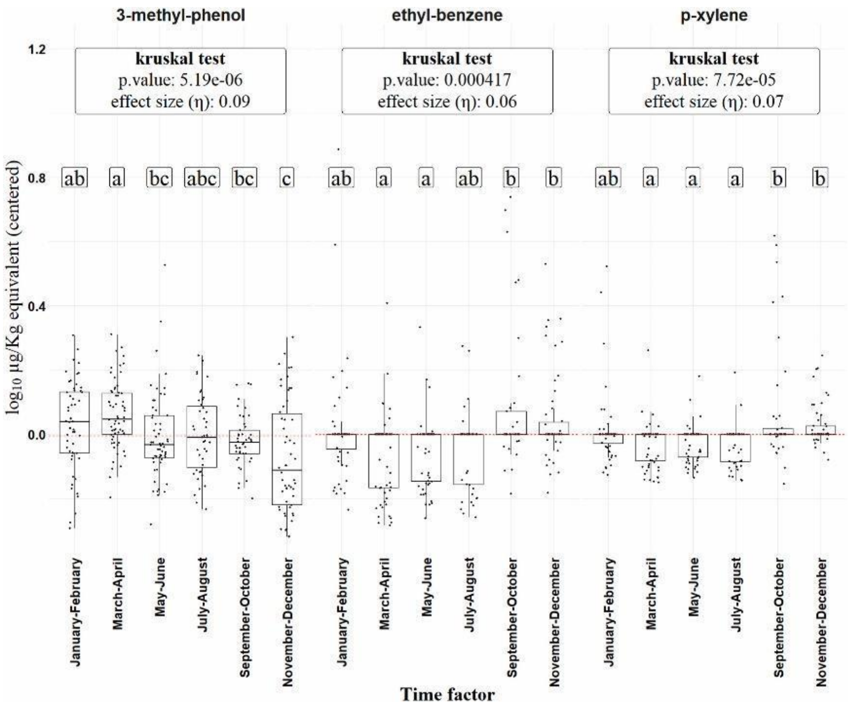

Figure 17: *Boxplot reporting values of the three significantly different volatile compounds for the factor "Time".*

Values are centered for the mean value of each batch of SPME fiber adopted. Each dot corresponds to a single measurement, the bold black line represents the median value, upper and lower margins of the blocks indicate the limit of the second and third quartile respectively, and whiskers indicate upper confidence intervals at 95%. The dashed red line represents the overall median for each compound. Letters show the groups estimated from post-hoc pairwise comparison tests.

The estimated effect of time of the year when the wheels were produced on the content of VOCs was low at the multivariate level. Interestingly, Fabris *et al.* (2010), detected 8 molecules that changed significantly according to different part of the year when the milk is produced, analyzing with PTR-MS a small sample of cheese wheels from a pilot plant. The molecules were tentatively identified as medium-chain organic acids and ketones. The different results could be interpreted considering the smaller analytical power of the previous research due to the limited number of samples analyzed.

This could also be explained considering the variance added by the natural variability of the product: the seasonal effect may influence the volatile compounds in raw milk, but there are no indications that it could also modify the conditions during the production process and the ripening phase, thus a transformed product could be less affected by seasonal conditions.

4.2.3.3 O-PLS-DA predictive model

The O-PLS-DA algorithm allows modeling separately the variations of the predictors correlated and orthogonal to the response. This model improves the explication of the effect of the predictors and their systematic variation compared to standard PLS (Pinto et al. 2012). To estimate the effective presence of the effect of the milk collection procedure on the VOC content, the predictive O-PLS-DA model was validated using a permutation test. The significance was tested by comparing the performance indices of the models trained in the permutation test, considered as the null distribution, with the indices estimated from multiple partitions of the real data set. The comparisons between the replicates and the null distributions are shown in the violin plots in Figure 18.

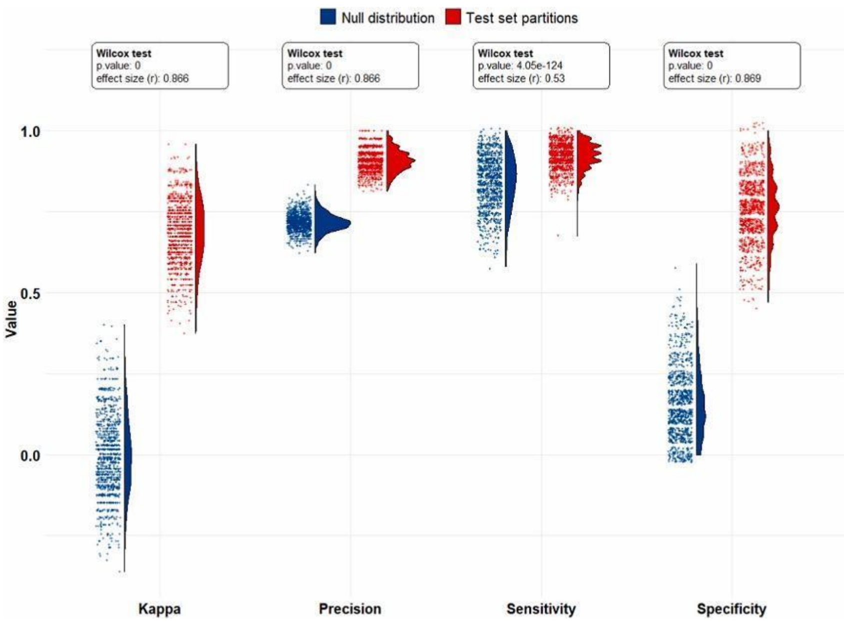

Figure 18: *Violin plot reporting the comparisons between the distributions of the model performance on repeated partitions and the null distribution of the permutation test.*

Results of the Wilcoxon statistical test are reported in the box of text above each parameter, each dot represents a single measurement, null distribution and multiple partitions distribution are colored in blue and red respectively.

Results of the comparison between multiple partitions and null distribution demonstrated that the final O-PLS-DA model had significant predictive capacities, reported by the significant difference from the null distribution of the kappa index estimated from repeated partitions. Moreover, it efficiently separated the two groups, as reported by the high values of sensitivity and specificity and by their significant difference from null distribution. This model demonstrated that at multivariate level the content of VOCs in a single cheese wheel could be associated with a different milk collection procedure adopted by the dairy factory. The test adopted demonstrated that the model was representative of the underlying data structure and that its performances were reliable independently from the single data partition.

The confidence interval of each regression coefficient of the model was estimated using a bootstrap approach, and the molecules whose confidence interval included 0 value were labeled as non-significant. The validated coefficient absolute values of the significant molecules are reported in the barplot in Figure 19.

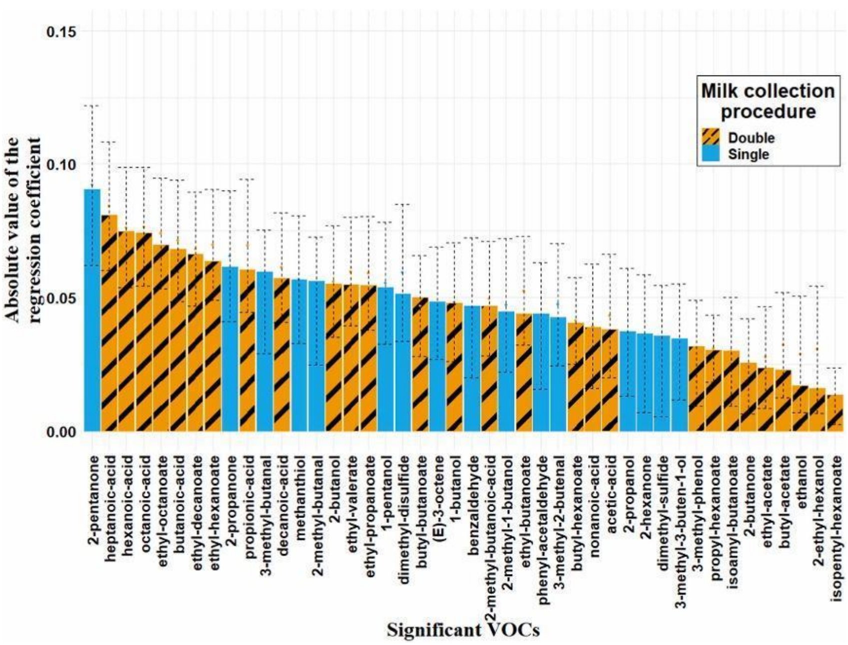

Figure 19: *Barplot showing the absolute value of the significant regression coefficients of the O-PLS-DA model validated using bootstrap.*

Values reported in light-blue bars are referred to the Single Milk Collection procedure (SMC) and values in orange bars are referred to the Double Milk Collection procedure (DMC). Bars show the absolute values of the coefficients of the model, dashed error bars report the estimated confidence intervals.

The bootstrap test of the O-PLS-DA model determined that the concentration of 44 volatile compounds is related to a different milk collection procedure. Results showed that the content of volatile compounds produced by the catabolism of fat in cheese, such as medium-chained free fatty acids, their esterified forms, and secondary alcohols, is related to the different milk delivery procedures.

The formation of 2-pentanone was related to the different pasture techniques of the cows by different studies (Kilcawley *et al.*, 2018; Villeneuve *et al.*, 2013). However, as the product specification of production regulates the food intake of the cows, the most reasonable production pathway is the oxidative pathway of fatty acids by microbial activity (Collins *et al.*, 2004). Instead, 3-methyl butanal and 2-methyl-1-butanol are transitory compounds of branched amino acids' catabolism during ripening (Bovolenta *et al.*, 2014).

Medium-chain fatty acids, such as hexanoic, heptanoic, octanoic, nonanoic, and decanoic acids were significantly higher in cheese wheels produced by a double delivery procedure of milk. The different quantities could be explained by a higher lipolytic activity due to the storage and collecting procedure, as suggested by Franciosi *et al.* (2012).

The formation of ethyl hexanoate, ethyl octanoate, and ethyl decanoate is related to the esterification of an organic acid with ethanol due to microbial activity, and their presence is related to the availability of free fatty acids (Kilcawley *et al.*, 2018). Interestingly, such compounds were significantly higher in cheese from double milk collection procedures, thus reasonably suggesting these molecules as other reliable markers of the process. Moreover, these results are coherent with those highlighted by the ASCA model for the factor Dairy Factory.

According to Collins *et al.* (2004), lipase activity in cheese is affected by process conditions and microbial and enzymatic activity, and it is a critical step for the synthesis of secondary products of lipid metabolism during ripening. Wang and Randolph (1978) reported a reduction of the lipase activity in skim milk after temperature inactivation of the lipase naturally present in milk in conditions similar to the milk collection procedures reported in the present work. Eugster *et al.* (2019), instead, reported that Non-Starter-Lactic-Acid-Bacteria (NSLAB) can produce acetoin, 2-butanone, and 2-butanol at high levels from pyruvate in hard and semi-hard cheeses, affecting the catabolism of amino acids positively. Altogether, an explanation for the variations related to the milk collection process could be that the different duration of the skimming process may affect the activity of the endogenous lipase enzyme in milk and of the non-starter lactic bacteria (NSLAB) naturally present in the raw milk, which grows better in raw milk collected using the double delivery procedure (Franciosi *et al.* 2011, Giraffa 2021). The routinary analysis of the total microbial population in raw milk sampled from vats done by Trentingrana Consortium reported a significant difference between dairy factories according to the different milk collection procedure (data not shown).

The results reported on the effect of milk collection procedure are coherent with previous research in a single dairy factory with controlled conditions of milk collection procedures (Endrizzi et al. 2012), which reported significant differences in the content of organic acids and esters between cheese wheels produced adopting a different milk collection procedure. This confirms the validity of the adopted multivariate strategy based on ANOVA-Simultaneous Component Analysis (ASCA) to extract useful information from volatile organic compound data in a real production context where several confounding factors are present.

4.2.4 Conclusions

The analysis of food products directly sampled from the production process allowed measuring the relation between the production process' variables and the properties of the food product at a real scale semi-industrial level.

ANOVA simultaneous component analysis showed the effect of the season and the production plant on the content of volatile compounds in cheese, highlighting the differences between dairy factories due to milk collection and the sub-products of Maillard reaction at low temperatures.

These results highlight how the Trentingrana cheese chemical profile is affected by the first steps of the production process: raw milk storing and skimming. For this reason, the quality control procedure to produce hard-seasoned cheeses needs also to monitor and uniform the conditions of the process in those early stages to ensure the same properties.

In conclusion, the proposed analytical framework can be applied in other research related to large-scale food production processes to highlight the factors responsible reliably and effectively for the differences observed when the latter are masked by several confounding cofactors.

Further research is needed to estimate the underlying mechanism at the chemical and microbial level of technological variations in the production process and their effect on the quality of cheese at the technological and sensory levels. The association of microbial and VOCs data could help to understand the synthesis of VOCs in Trentingrana cheese. Furthermore, if associated with process conditions data, they could provide useful information for the optimization of the production process.

CHAPTER 5. FROM INSTRUMENTAL TO SENSORY

The relationship between the instrumental and the sensory quality parameters is a topic of great concern in food technology because the estimation of those relationships gives insights into the interpretations of the causes behind the mechanism of perceptions and consumer preference. Furthermore, understanding those relationships is necessary to associate mechanical measurements with sensory defects to develop automatic quality control procedures. In the present chapter, the procedure for the association between instrumental and sensory measurement is presented and discussed and an application of the Trentingrana case study is reported. The results of the analysis proposed here were published in poster format and are reported in the abstract book of the conference EUROSENSE 2022: A Sense of Earth 10th European Conference on Sensory and Consumer Research, Turku, Finland, 13 - 16 September 2022, written by Ricci M., Gasperi F., Menghi L., Endrizzi I., Cliceri D., Aprea E., in the context of the TRENTINGRANA project by Edmund Mach Foundation.

5.1 The association between sensory and instrumental measurements

5.1.1 Sensory Measurements

Sensory quality control is adopted in a large amount of food manufactures, with a slight variation to adapt the analytical procedure to the specific production processes and sensory properties (Munoz 2002).

The application of sensory quality control is a fundamental tool to detect many defects that cannot be detected by adopting instrumental procedures. Many sensory properties are critical for consumer acceptance, and they usually cannot be measured efficiently with the instrumental procedure but can be easily detected by humans developing an adequate sensory analytical framework.

The adoption of sensory quality procedures represents an important change in perspective in the specification of the product because it requires a definition of sensory parameters to be measured and to define which are the optimal levels of sensory properties.

The procedures of sensory quality control may be applied in agrifood manufacture in many ways, ISO 20613 (ISO 2019) reports some of these measurements.

5.1.2. Relating instrumental to sensory measurements

During the last 30 years, the association between sensory properties and instrumental properties of food has been studied at different levels (Andrews *et al.* 2021, Spence 2021, Smyth & Cozzolino 2013, Seisonen *et al.* 2016, Foegeding & Drake 2007). Despite the vast number of experiments, a lot of work remains to be done for a detailed association between many different chemical and physical properties of food products and for many sensory properties.

In the context of quality control procedures, instrumental analysis can be used to quickly detect sources of sensory defects. This is possible for issues where the chemistry or the structure of the product is related to the hypotheses about the cause of the quality traits detected through sensory analysis. There are many factors that can impact the integrity of the supply chain and lead to sensory defects: an interesting example in the Trentingrana production process is the incorrect handling of the temperature during the cooking of the curd, which can cause excessive retention of water inside the cheese wheel and consequently the formation of stirred paste.

Sensory analysis is necessary for determining the presence of sensory defects. The association between sensory and instrumental analysis consists in validating a statistical procedure that estimates the range of the values from instrumental measurements that are associated with the presence of defects. It is worth saying that for sensory defects instrumental (and other non-sensory) analyses can be useful, but not infallible: false negatives may happen because may incur problems that are related to other causes that do not affect the properties measured by the instrument.

To associate chemical instrumental analysis with human sensory analysis to infer the associations between flavor-related sensory properties and physical-chemical properties, first of all, it is necessary to understand how the sensory information analyzed is acquired and how it is interpreted, to ensure that the associations inferred are not spurious (Andrews *et al.* 2021). Then, it is necessary to have clearly defined terms and definitions for the sensory attributes adopted (Drake & Civille 2003). Defined terms are functional tools for panel training and minimizing variability but also for choosing and interpreting instrumental measurements of the sensory attribute.

The comparison of chemical or physical instrumental analysis versus human sensory analysis in the determination of flavor- or texture- related sensory properties requires knowing how the sensory properties are acquired by the human and how it is interpreted.

In the specific case of the analysis of texture sensory properties in cheese products, the perceptions are affected by the food structure, but the presence of taste components can modify the perception, due to the phenomenon of cross-modal interaction (Foegeding *et al.* 2007).

Thus, human perception of the sensory profile of a food depends on multiple factors rather than the simple presence of a range of individual chemical compounds and their absolute levels or of the several mechanical properties of material structure. Regarding texture and flavour properties of cheese, we need to acknowledge the events that happen during mastication: the temperature increases, particles are mixed with saliva, and the crushing process is between jagged surfaces of teeth. Furthermore, because cheese is a viscoelastic material the mechanical properties sensed during chewing are affected by the rate of jaw closure. In most instrumental textural measurements those conditions are not present, surfaces are plain, and analyses are done isothermally.

The property of cheese texture that is most easily measured by instrumental and sensory analysis is Hardness (Wium *et al.* 1997, Drake *et al.* 1999). This is because both the human mouth and mechanical property-testing instruments are capable of measuring force and are more easily correlated. However, Hardness is not the only defining textural term for cheese. As will become evident in the case study, properties that are assessed by chewing, such as cohesiveness, adhesiveness, and smoothness, require more analysis.

The correlation structure between sensory texture and instrumental textural profile in cheese was estimated differently in different research (Bryant *et al.* 1995, Drake *et al.* 1999). There are various factors that may explain the variations, such as the mechanics of the test, the individual differences between the panelists, the adoption of slightly different tasting procedures, or non-coinciding definitions for the same terms.

5.2 Statistical Models

For the analysis of the relations between sensory and instrumental parameters is available an array of traditional statistical methods, such as experimental designs combined with response surface methodologies and linear regression methods (comprising multiple linear regression, partial least squares, preference mapping). Also, more advanced procedures have been proposed, such as artificial neural networks (Yu et al. 2018).

Experimental designs are denoted by a matrix, that represents the independent variables analyzed with the columns, and the samples or experimental runs with rows, divided by the factors of interest.

Because of the multivariate structure of sensory data, standard univariate statistical procedure such as one-factor analysis of variance (ANOVA) cannot estimate the correlation structure between food physicochemical attributes, sensory profiles, and hedonic properties (Zielinski *et al.* 2014). Furthermore, the presence of masking and synergistic effects, especially between different sensory properties of food, determines relationships between these factors that cannot be properly represented by adopting simple univariate statistical procedures (Noble & Ebeler, 2002). A single physical property such as texture or flavor may be related to multiple sensory attributes in different ways, as perceived by the human brain (Perrot *et al.* 2006). Multivariate statistical methods are essential for these evaluations. These methods are categorized into linear and nonlinear, and in the next paragraphs some of the most used procedures will be presented.

Multiple linear regression is a statistical model that predicts a single dependent variable from multiple independent variables. As explained in the previous paragraph, linear regression using a single independent variable is often insufficient and is unlikely to yield a satisfactory explanation of the relationship between the independent and dependent variables.

Stepwise multiple linear regression is a linear regression technique adopting a selection algorithm of the variables in building a regression model. This regression procedure is divided into steps, consisting in adding or removing a model term from the model. Model terms with the most significant change in the designed index of good fitness will be added or removed, to maintain only the most relevant features.

Partial Least Squares Regression (PLSR) is a linear regression method that can be thought of as a combination between multiple linear regression and principal component analysis. For further discussion, we recommend chapter 2 of this thesis and Noble & Elber (2002).

Preference mapping is a further application of PCA and other related methods (principal component regression, PLSR, etc.). It is used in producing a visual representation of sensory and consumer data from which significant trends and observations can be easily deduced, for an extensive discussion on the background of preference mapping, readers can be referred to Greenhoff and MacFie (1994).

Artificial Neural Networks (ANN) are machine learning models that are very efficient in correlating process variables with non-linear relationships, which are very common in food processing and sensory analysis. ANN has been used in food areas associated with sensory analysis, sensory quality-based food process control, and control set points of processes (Kupongsak & Tan 2006).

ANN algorithm consists of a network of many neurons that carry non-linear function, and a different unit that contains different function is weighted, connected, and combined to produce overall output.

Before applying ANN, the input and output variables should be established. Next, the type of ANN should be chosen, such as radial basis function and other critical parameters. Input variables are inserted in the first layer, and the calculation steps proceed layer by layer until the last one.

ANNs have been used for many purposes in food technology (Marini *et al.* 2008, Cevoli *et al.* 2011, Bona *et al.* 2011, Baykal & Yildrim 2013).

5.3 Application 3: prediction of sensory textural properties of Trentingrana cheese

5.3.1 Introduction

The distinctive production process of Trentingrana cheese ensures the peculiar sensory quality and its characteristic granular texture.

Texture properties are critical quality traits for ripened cheese, affecting its overall sensory quality and acceptability. Because monitoring texture sensory properties using traditional sensory evaluations is highly expensive and time-consuming, predictive models could be promising tools that lower the costs and produce effective results.

The present study developed Partial Least Squares Regression models to estimate textural sensory attributes from instrumental texture measurements and gross composition in 64 cheese wheels sampled from the Trentingrana consortium dairies (n = 15).

The same cheeses were analyzed for mechanical properties using uniaxial penetration in 24 replicates in the same portion of cheese evaluated by the panel, sampled in different positions considering the variability inside the product. Cheese gross composition was estimated by Near Infra-Red (NIR) spectroscopy.

The objectives of the work were to estimate textural sensory attributes from instrumental texture measurements and gross composition, and to infer which physical properties affect the textural sensory properties of hard-ripened cheese. This relationship was estimated and validated for each sensory descriptor.

The validation procedure consists in evaluating if the results estimated by the models are significantly more accurate than results estimated by random values, confirming the hypothesis of an association between instrumental and sensory measurements. The validation step verifies if the model can give more information than a random estimation of the results or add only more nuisance and cannot estimate associations between sensory and instrumental data.

5.3.2. Materials and methods

5.3.2.1.Instrumental measurements

Texture properties were measured on each cheese block by a TA-XT texture analyzer (Stable MicroSystem Ltd., Godalming, UK) applying a uniaxial compression/penetration on one of the wider sides of the cheese block sample. Following the method described by Noël *et al.* (1996), a 4 mm probe was used with a speed of 1.67 mm/s, a trigger force of 5 N, setting the endpoint of the measurement when a maximum strain of 90% of the height of the sample was obtained, and three mechanical parameters were calculated on the recorded curves. Those parameters are the same adopted in application 1 and are reported in Table 4.

5.3.2.2.Sensory measurements

A trained panel (n = 14; 71 % males; average age = 40 years old) analyzed Trentingrana samples in duplicate according to conventional descriptive sensory analysis. The texture attributes evaluated were Hardness, Friability, Humidity, Crystals, Microstructure, and Solubility, on a continuous scale from 0 to 100. The panel evaluated 3 samples in duplicate in each session and a total of 20 weekly session were performed.

The panel was previously trained on each descriptor with reference products to define the minimum and maximum levels. The sensory scores assigned to each sample were the average value of each panelist's assessments. The list of sensory descriptors is reported in table 8.

Table 8: *Summary of texture descriptors, their definition, and how the judges are trained to recognize and estimate them.*

Descriptor	Definition	Evaluation procedure	Reference for lowest intensity	Reference for maximum intensity
Hardness	*Resistance of the sample to a small pressure exerted using the molars.*	*Put the whole sample between the molars, close the jaw uniformly, and estimate the resistance of the sample before being deformed.*	*Wurstel*	*Steamed carrot*
Friability	*Product's tendency to form multiple fragments at the beginning of the mastication.*	*Bite the sample using molars from 2 to 4 times and estimate the increment of the number of fragments before they solve in the saliva.*	*Egg yolk*	*Canestrello biscuit*
Humidity	*The perceived humidity/dryness during mastication.*	*Put the sample in the mouth, masticate it 4 times and evaluate the level of humidity/dryness.*	*Madeleine*	*Wurstel*
Crystals	*The number of crystals perceived in the sample during mastication.*	*Masticate the sample (8 times until it turns into bolus) considering the sound produced by the compression of crystals by the teeth.*	*Ricotta cheese*	*Ricotta cheese + granulated sugar*
Microstructure	*Particle size perception in the bolus after the mastication of the sample.*	*Masticate the sample (4-8 times until bolus formation) and evaluate the shape and the perception of the particles in contact with tongues, cheeks and gums (from thin to coarse).*	*Ricotta cheese*	*Cous cous*
Solubility	*Perception of fast solving of the sample in the mouth's saliva.*	*Masticate the sample from 4 to 8 times with molars and estimate the speed of the dissolution in the saliva of the whole sample or of part of it.*	*Almond*	*Egg yolk*

5.3.2.3. Statistical analysis

For each sensory descriptor a Partial Least Squares Regression model was developed by adopting a Kernel PLS algorithm. The regressor adopted were all the instrumental measurements adopted and of their second-level interactions. To estimate the optimal number of components for each model, a cross-validation procedure with a multi-fold balanced partition was applied.

In the present context, because a second set of measurements of sensory and instrumental data to test the model was not available, a permutation procedure is proposed (n = 1000). A permutation test consists of multiple estimations of predictive models with a dataset with randomized predicted values for each model and estimating the predictive capacity of each model estimated.

The predictive capacity of the model is summarized by the Random Mean Square Error (RMSE) values between the predicted values and the values of the test set. Then, the RMSE index distribution of all the values estimated by the predictive models trained with the real dataset is estimated, using each time a different balanced randomized partition.

Each predictive model estimated for each permutation is trained after a balanced random partition between train and test subgroups, followed by a cross validation procedure inside the train set to estimate the optimal number of components for that specific partition. Subsequently, the model is applied in the test partition.

This procedure estimates two distributions, one representing the null distribution of the random error once the data are completely mislabeled (the null error), and the other representing the predictive capacity of the model including the effect of the different partition on the significance of the model. The difference between the two distributions is verified by a Wilcox pairwise test considering both the significance of the test and the *r* value for the effect size (Tomczak & Tomczak, 2014).

To estimate the predictive power of the trained model, a multiple partition procedure was performed to estimate the average estimation for every value of the dataset by the final model, to estimate the predictive power of the PLS structure.

5.3.3. Results and discussion

5.3.3.1. Validation of the models

The results of the validation test were reported adopting a violin plot, to show the comparison between the null distribution of the error and the distribution of the error from multiple partitions. To interpret correctly the test, the results of an unpaired Wilcoxon test statistic were also reported on the top of each parameter. The interpretation of the results, to overcome the effect of the high number of measurements due to the recursive structure of the permutation, was done considering both the p value of the test and the r index reporting the size of the effect detected by the test. The results are shown in Figure 20.

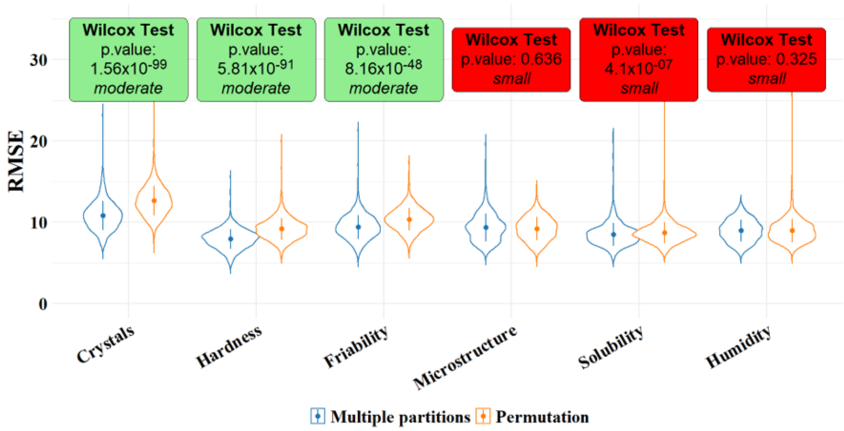

Figure 20: *Violin plot representing the results of the permutation test. The box above each descriptor reports the results of the Wilcox test. The point inside each violin represents the mean of the distribution, and the line represents the standard deviation.*

The results reported in the image show clearly that only three descriptors were significantly predicted more accurately than the null distribution, so it is legit to assume that the PLS model can effectively predict three sensory descriptors: Crystals, Friability and Hardness. It is generally acknowledged

(Foedeging & Drake 2007) that Hardness and Friability are two sensory descriptors associated with perceptions that can be related more easily than others with the results of instrumental measurements capable of estimating the force necessary to break the structure, while the descriptor for crystal has been much less investigated in the literature, being a descriptor of a quality trait related to the quality profile of grana cheese (Kindstedt & Polowsky 2021). The validation test also showed that there are no significant relations between the instrumental measurement of texture and gross composition and the perception of the Microstructure, Solubility, and Humidity. These descriptors are related to properties that probably are not measured by the analytical techniques adopted, consequently were not considered in the rest of the work.

After the first validation step, a cross validation procedure was performed for the three significantly different descriptors, to estimate the optimal number of components that the models must adopt to represent optimally the results. The results of the cross-validation procedure were reported in the scatter plots in Figure 21.

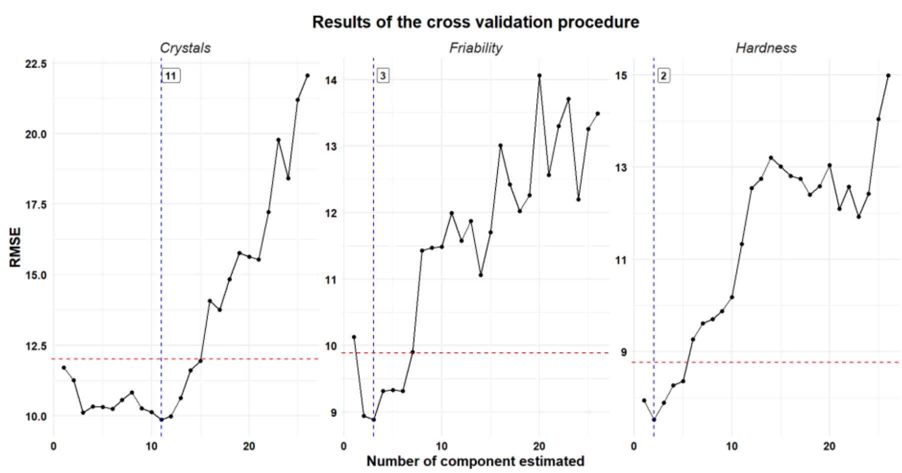

Figure 21: *Scatterplots representing the results of the cross-validation procedure of the PLS models for Hardness, Friability, and Crystals. The vertical blue dotted line represents the number of components with the lower value of RMSE, and the horizontal red dotted line represents the value of the standard deviation of the observed values.*

The results of the cross-validation procedures show that for Hardness and Friability respectively two and three principal components can represent optimally the overall structure between sensory and instrumental measurements, showing that there are no excessively complex correlation structures between them. For the descriptor Crystal were elected 11 principal components to represent the correlation structure. The number of components was chosen because the lowest Random Mean Square Error (RMSE) estimated using cross validation was the one corresponding to eleven principal components. This result may be related to the presence to a complex correlation structure, which needs a larger number of principal components than the amount adopted for other sensory descriptors, because it is not a simple linear relation.

5.3.3.2. Predictive power of the models

Adopting a multiple partition procedure (n=1000), the overall predictive capacity of the PLS models was estimated, and the overall explained variance was measured. The predicted values were summarized by the means for each observed value, to estimate the average estimation of the models compared to the real value. The comparison between the observed values and the mean predicted values represented the average predictive power of the model independently from the partition adopted to train it. The results were represented by adopting scatter plots comparing the predicted versus the observed values reported in Figure 22.

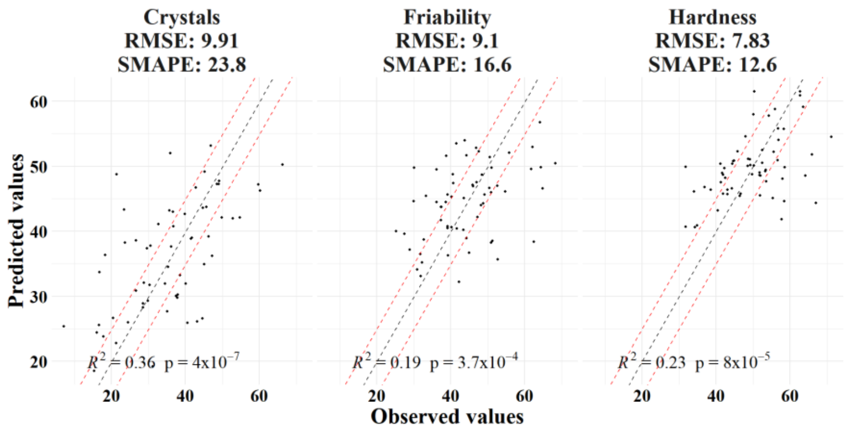

Figure 22: *Scatterplots reporting the comparison between observed and predicted values for the three sensory descriptors which have significant PLS predictive models. The black dotted lines represent the bisector passing by the zero values of both axes, while the two red dotted lines represent the bisector passing through the +5 and -5 values of Y axis, to help interpret the difference between predicted and observed values. The RMSE and SMAPE values for each descriptor are reported above each scatterplot, while the values of R^2 are reported in the lower part of each plot.*

As we can see in Figure 22, the results show that even if they are significant, the PLS predictive models can represent only a small part of the overall variance without overfitting. The overall error of this procedure is lower than the overall standard deviation of the observed values, thus the model can estimate significant differences. However, the low values of the R^2 for each sensory descriptor indicate that only a fraction of the overall variance of the sensory properties of Trentingrana cheese can be detected by the actual models.

5.3.3.3. Correlation structure

To estimate which are the variables that the PLS models estimated adopted mostly to predict the sensory properties of cheese, the loadings values of the validated models were studied. To represent the correlation structure estimated by PLS models, biplots representing the loading values for each variable of the matrix X, and the loading values for the Y values were used. For the sake of interpretation, only the individual variables without interactions were represented.

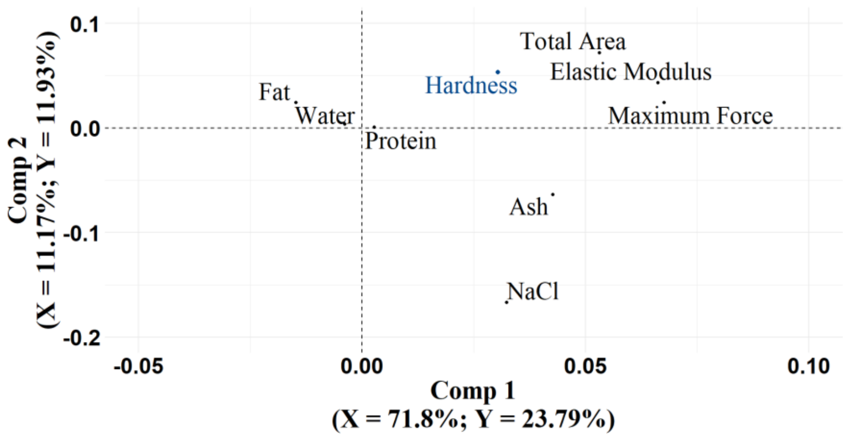

Figure 23: *Biplot representing the loading values for the X and Y matrix of the PLS model estimated for predicting the descriptor "Hardness". The Y loading value is reported in blue, the x loading values are reported in black.*

In the model predicting "Hardness" sensory descriptor, as reported in Figure 23, the two principal components explain 82.9% of the variance of the matrix X and 35.7% of the variance of Y. Those values mean that a large amount of the variance of the instrumental measurement is correlated with a fraction of the overall variation of Hardness.

The correlation structure shows that the values estimated by the texture analyzer are all positively correlated with Hardness, as well as the content of ashes too. Furthermore, we can see that the content of fat and the content of NaCl are anti-correlated with the intensity of Hardness.

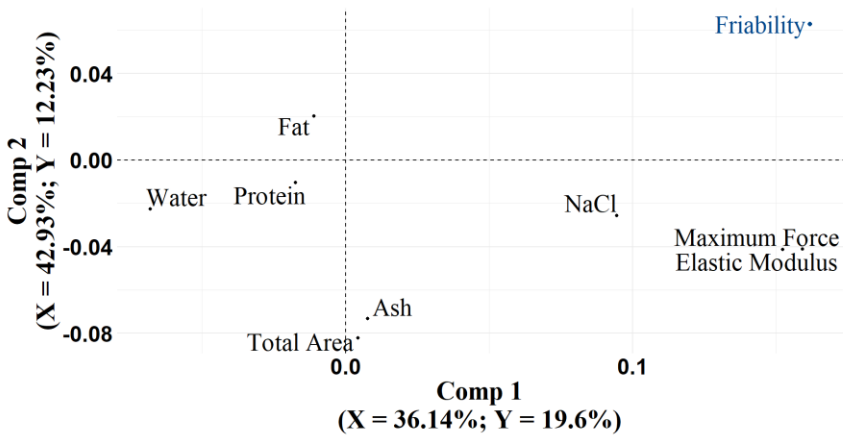

Figure 24: *Biplot representing the loading values for the X and Y matrix of the PLS model estimated for predicting the descriptor "Friability". The Y loading value is reported in blue, the x loading values are reported in black.*

For the PLS model predicting Friability, as we can see in the biplot in Figure 24, the first two principal components of the PLS model represent cumulatively the 79.1% of the variance of the matrix X and the 31.8% of the Y value. This represents, similarly to the previous PLS model, that the model associates a large percentage of the X matrix with a small percentage of the Y matrix.

The distribution of the loadings values highlights that the instrumental parameters of Maximum force and elastic modulus by the texture analyzer are positively correlated with the perception of friability ~~firmness~~, as well as the content of NaCl and fat. The plot also shows that the textural parameter "Total Area" (that represents the integral sum of the force applied during the complete penetration of the probe in the sample) and the content of water and ash are negatively correlated with the perception of friability ~~Firmness~~.

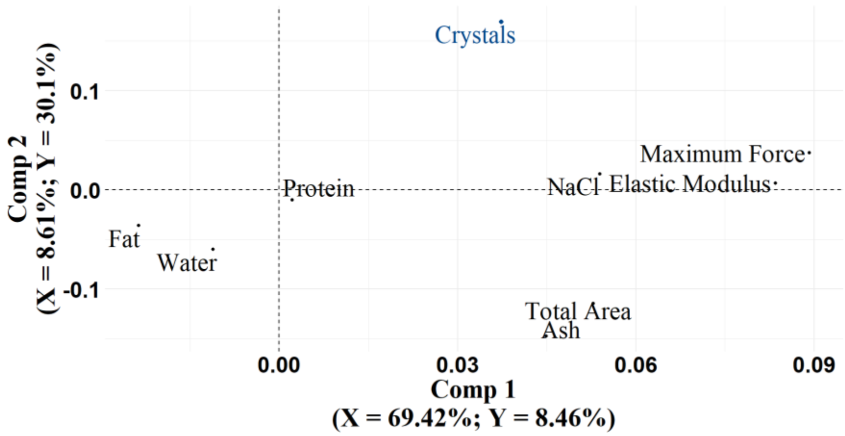

Figure 25: *Biplot representing the loading values for the X and Y matrix of the PLS model estimated for predicting the descriptor "Crystals". The Y loading value is reported in blue, the x loading values are reported in black.*

The first two components of the PLS predictive model for the descriptor "Crystals" represent 78% of the X matrix and 38.6% of the Y matrix.

The loadings values highlight that the value of Maximum Force exerted by the texture analyzer and the content of protein and NaCl are positively correlated with the perception of Crystals, while the value of Total Area, water, Fat, and Ash are anticorrelated.

5.3.4. Conclusions

The predictive models had performances significantly different from the null distribution for three descriptors: Hardness, Friability, and Crystals. The coefficients of the models were bootstrap validated, thus estimating the relationships between oral perception and the physical properties of cheese. Significant correlations between perceived Hardness and the maximum force applied and between perceived Friability and the presence of NaCl were estimated.

The low R^2 values are due to the absence of critical parameters of the tasters and of other sensory properties that couldn't be included in these models (Andrewes *et al.* 2021). Thus, the PLS models inferred the correlation structures between physical and sensory properties, highlighting the anticorrelation structure between Maximum force and Total area for descriptors Friability and Crystals and the different effects of the combination of NaCl and Fat content with sensory textural properties.

CHAPTER 6. IMAGE ANALYSIS FOR AUTOMATIC QUALITY CONTROL

In the present chapter the use of image analysis in food quality control procedures, reporting its main applications in food quality control is discussed. The rest of the chapter discusses the development of a quality control procedure for the case study, reporting the development of two different image analysis algorithms and comparing their performance. The analyses proposed here are reported in the proceedings of the Improve 2023, 3rd International Conference on Image Processing and Vision Engineering, written by Caraffa A., Ricci M., Lecca M., Modena C.M., Aprea E., Gasperi F., Messelodi S., in the context of a collaboration between Edmund Mach Foundation and Bruno Kessler Foundation (Caraffa et al., 2023).

6.1 Applications of image analysis procedures

Image analysis is a well-established tool in process analytical technology, and it is applied in many different food productions, such as fruit sorting, spoilage recognition, and detection of sensory eligibility of the product (Meenu *et al.* 2021). Nevertheless, for several food manufacturers, quality control still relies on the evaluation of human operators, who judge the quality of food by their perception of quality features such as visual appearance, texture, hardness, crispness, smell, and taste (Munoz *et al.* 2002). Trained operators can elaborate a large variety of sensory perceptions at the same time and summarize them in a single evaluation, with eventually the aid of simple visual comparison (Brosnan & Sun, 2004), but their work requires a large amount of time, it is affected by sensory fatigue and cannot be applied to many samples on a regular basis (Laddi *et al.* 2013). Furthermore, visual quality evaluation is affected by subjective differences between judges and an overall inter-individual variability, which is managed by adopting a well-designed experimental protocol, the basic rules of good practice of sensory analysis, and carefully selecting, training, and monitoring judges. For this reason, many instrumental measurement procedures are included in process analytical technology to replace human evaluation.

Visual inspection plays an important role in the quality assessment of many food products, and it is of great concern to develop many interesting analytical procedures. Visual analysis can collect information about the color, structure, and composition of food products. Optical analysis of food products can be classified according to the scale of the visual data collected, dividing the analysis between microscopic and macroscopic scale measurements. Microscopic scale defines all the analyses performed using technologically advanced instruments, e.g., near-infrared imaging systems, spectroscopy and hyperspectral imaging systems, and X-ray imaging sensors, that provide objective measures regarding microstructures (Russ, 2015; Lei and Sun, 2019).

Macroscopic analysis comprehends the analysis of images at a scale not significantly different from the perception of the human eye, such as Computer vision techniques, that analyze the visual properties of food through image processing algorithms. Those techniques are promising tools for objective, effective, sustainable, cheaper, and faster assessment of food quality (Du and Sun 2004, Turgut *et al.* 2014, Ma *et al.* 2016, Jackman and Sun 2013). An interesting application in food manufacturing is the assessment of the degree of conformity to the quality standards of a protected cheese brand (Bosakova-Ardenska *et al.* 2021, Badarò *et al.* 2021).

In practice, image analysis procedure consists of the development and the application of algorithms on the matrix corresponding to the image, after preprocessing procedures, such as cropping, filtering, noise removal, and the extraction of contours.

The next step is the extraction of the features, consisting in measuring the properties of the image, to summarize its properties. The features extracted are the measurements of the color content, the presence of patterns, and the morphological/geometrical properties in the image.

Then, the values extracted from the previous procedures are processed by adopting the tailored algorithm, according to the objectives of the analysis. First, a selection procedure or a dimension reducing procedure is applied, then the effective data analysis procedure, which could be a statistical model, a fuzzy logic model, or a machine learning model, according to the fitness for a purpose is applied to the data and the results are validated.

The application of artificial neural networks models or another black-box model on image analysis and computer vision may require only simple pre-processing steps at most, while all the other steps of feature extraction and selection are bypassed, because during the training of artificial neural networks for image analysis there are many procedures that proceed to detect the importance of variables.

An overview of computer vision methods applied to quality control procedures of dairy products, together with a discussion about the limitations of adopting standard image analysis algorithms, can be found in (Lukinac *et al.* 2018).

6.2 Visual quality of hard seasoned cheese

6.2.1 Visual quality in Trentingrana

The Trentingrana Consortium included in the quality control procedure a visual evaluation of the cheese wheels after they were cut in half. The visual examination considered three quality parameters: the color of the grain, its structure, and the rind thickness.

The color of the grain is evaluated considering the hue, the intensity, the homogeneity and the presence of spots and defects in the surface of the grain right after the opening of the cheese wheel.

The structure of the grain is an overall evaluation of the morphological properties analyzing the surface of the grain ready after the opening of the cheese wheel. The properties of the grain surface considered are the granularity, the presence of cracks or holes, crystals, and the overall homogeneity. Also, the presence of defects, such as stretched, chalky, or foamy.

The last parameter, the rind thickness is the focus of the rest of the chapter and will be thoroughly discussed in the next section.

6.2.2 Importance of rind thickness

The rind is the external region of cheese, it has a more compact texture and darker color compared to the grain in the internal regions. Rind thickness may change from one side to another in the same wheel and may be larger at the edges and in the round sides. The color of the grain below the rind fades gradually to the overall color of the grain. Differently from fresh, soft cheese, hard seasoned cheese such as Trentingrana cheese have a rind section characterized by a darker color and higher Hardness of the grain. The formation of the rind is due to the variation in the water content in the external part of the cheese wheel that is caused by the water dispersion due to salting and ripening procedures (Fox & Cogan 2004). The correct exposition in terms of time, temperature, and humidity during the ripening phase, along with systematic removal of the external water exuded from the wheel guarantees the correct development of the ripening procedures in terms of physical phenomena. For this reason, the thickness of the rind is directly related to the intensity of the water dispersion inside the cheese wheel and represents an indicator of the condition of the ripening phase. If the rind is too thick, it is due to excessive water dispersion, while a thin rind is related to a too low amount of water dispersion, which can cause the growth of undesired microbial flora. It is also of great concern for quality control the identification of rind defects such as calcium lactate crystal formation and excessive rind halo, because they are caused by the supersaturation of the cheese serum phase with calcium and lactate ions due to bad water diffusion, which then forms crystals producing white hazes on the surface of the cheese (Khattab *et al.* 2019, Rajbhandari & Kindstd 2008).

Another important parameter of the properties of the rind is its uniformity, which is directly related to the handling of the cheese wheels during ripening: difference in the thickness of the rind between the upper and the lower plate is directly related to the position of the rind during the ripening and the amount of time spent after being rotated in the stall. A difference in uniformity is inevitable, but important differences between the upper and the lower plate are related to the exposition to air, guaranteed by regular reversals of the wheel.

Because of these relations between the thickness of the rind and the condition of the ripening process, along with the indication of the internal product specification of the Grana Padano Consortium, and the concern of the consumers, the Trentingrana consortium included in the quality control procedure ~~of~~ the evaluation of the rind thickness as one of the visual parameters that the quality jury must evaluate.

6.3 Application 4 – Evaluation of rind thickness in Trentingrana cheese

6.3.1 Introduction

The objective of this work is to assess the visual quality of Trentingrana cheese, a local variety of Protected Designation of Origin (PDO) Grana Padano cheese, adopting advanced machine vision techniques. The cheese wheels produced were sampled and rated by a panel of experts assessing multiple quality parameters after tasting and visual evaluation. Some wheels are opened along their diameter and visually screened by the experts.

For the estimation of visual quality, one key parameter is the rind thickness, because ungrounded cheese with a thick or uneven rind isn't appreciated on the market, and it is an index of anomalies in the water content due to ripening conditions. For this reason, this parameter needs to be estimated, with other visual and sensory parameters, to evaluate the best preparation of the final product. In the present work, a deep learning method was developed and tested to estimate the rind thickness of Trentingrana cheese in six different positions from high-definition images.

Using an industrial camera several images of Trentingrana cheese slices were obtained from a half-wheel by wire-cutting a piece and then dividing it into left and right slices. To estimate the human perception of rind thickness, a panel of 12 experts examined each picture and measured the rind thickness in the upper and lower faces of the cheese slice and the heel. To improve the assessment of the overall rind quality, for each cheese wheel the image of two slices was collected and measured and for each image, the thickness of the rind was measured in five different positions. In this stage, the jury was asked to evaluate the rind thickness from digital images of the cheese slices instead of the real ones. This assured that the models were trained using estimation that were inferred exactly from the same visual parameters that will be used to train the model.

The data collected were used to train and test an artificial neural network that estimates the thickness of the rind reported by the experts from the images. Results show that the model predicts accurately the panel estimation of the rind thickness. As a reference, a purpose-driven algorithm that relies on the traditional computer vision procedure of detecting the contrast between rind and paste areas was developed. Compared to the mean thicknesses reported by the experts, the Mean Absolute Error

(MAE) for the hand-crafted algorithm stands at 1.10 mm, while the MAE for the deep learning method is 0.51 mm.

6.3.2 Materials and methods

6.3.2.1 Image acquisition

We considered 45 cheese wheels from 15 dairy factories related to Trentingrana Consortium. Each wheel was opened along its diameter with a special knife. A piece about 2.5 cm thick was wire-cut and divided into a left and a right slice.

Using a visual analyzer (Iris, AlphaMos, Tolouse France) under the top and bottom lighting (D65 compliant, 6700°K color temperature), we acquired N=90 images $I_1, \ldots, I_N$ with dimension 1024x768 pixels, each depicting a slice of cheese.

A calibration step was performed to estimate the conversion factor from pixels to millimeters. For this purpose, an image was captured depicting an object of known dimensions (in millimeters), and its size in the image (in pixels) was extracted, resulting in 1 pixel corresponding to 0.29 mm.

6.3.2.2 Image annotation

The visual evaluation has been performed through an online questionnaire created using EyeQuestion® (EyeQuestion, 2022), submitting all the images in randomized order to a jury of K = 12 experts $E_1, \ldots, E_K$. The experts quantitatively estimated the rind thickness of each cheese slice in three regions: for each image, each expert provided an estimate of the rind thickness in centimeters measured on the upper face A, heel B, and lower face C. A set of three rulers were superimposed on each image to assist the task of the experts.

Each region received K subjective measurements of rind thickness. The analysis of these estimates shows the presence of outliers and different evaluators provided different thickness values because of the smooth transition from rind to paste.

6.3.2.3 Region extraction

The Region Extractor receives as input an image I. It identifies three parts on the cheese border and three regions where the experts performed their measurements.

To this end, it computes the foreground mask F corresponding to the cheese area in I and localizes the upper and lower faces A, C, and the heel B by analyzing the concavities along its boundary partial F. One-thick-pixel parts P_A, $P_{B,}$ and P_C are extracted from partial F around the middle point of the three sides. Three rectangular regions (R_A, R_B, and R_C) around these parts are cropped and used as input to the Thickness Estimator. The size of the rectangles was chosen to roughly match the area observed by the experts to provide their measurements (280x150 pixels). Regions R_B and R_C were rotated to appear as R_A, i.e., with the white background at the top of the region.

6.3.2.4 Ground truth estimation

We defined a method to associate each region $R^{i,p}$, where i=1, …, N and p ∈ {A, B, C}, with a thickness measure taking into account the intrinsic variability of the annotations provided by different experts. Given the set of measures $T_k^{i,p}$, where k=1, …, K, we first normalized them with respect to a central position as follows:

$$M^{i,p} = \frac{1}{K-2}(\sum_{k-1}^{K} T_k^{i,p} - T_k^{i,p} - T_k^{i,p})$$

$$M = \frac{1}{3N}\sum_{i=1}^{N}\sum_{p\in\{A,B,C\}} M^{i,p}$$

$$M_k = \frac{1}{3N}\sum_{i=1}^{N}\sum_{p\in\{A,B,C\}} T_k^{i,p}$$

Then, we defined the ground-truth for $R^{i,p}$ as:

$$T^{i,p} = \frac{1}{K}\sum_{k=1}^{K} T_k^{i,p}\frac{M}{M_k}$$

The dataset was organized to keep track of the slice/wheel from which each region came.

6.3.2.5 Standard image analysis algorithm

The first computer vision method implemented for detecting the rind consisted in a hand-crafted algorithm (from here it will be called HCA). This algorithm was based on the empirical evidence that the rind of the cheese is darker than the interior because of its higher density and that the paste has a uniform coloring, although not constant. Thus, the rind thickness can be detected by searching for a color variation in the image by analyzing adjacent regions close to the three parts of interest.

The algorithm presented takes as input the image of a slice and extracts the 1-pixel-thick parts P_A, P_B, P_C as done in the data preparation step for the training of the deep learning method.

To highlight as much as possible the variation between rind and paste, and to attenuate an illumination gradient due to the acquisition device, the image is pre-processed by means of a color edge-preserving smoothing followed by an intensity normalization step. HCA works on this last image, named G. Let P be an element in $\{P_A, P_B, P_C\}$ and R the corresponding rectangular region in $\{R_A, R_B, R_C\}$. HCA computes n + 1 regions $P_0 = P$, $P_1 = S(P, 1)$, . . ., $P_n = S(P, n)$, where P_i is obtained by shifting P_0 by i pixels towards the cheese interior. n was fixed in such a way to ensure the scanning of a sufficiently large area to include both rind and paste, i.e., approximately 5 cm expressed in pixels. For each i, HCA computes the median of G's values along Pi and plots it with respect to *i* (Figure 26. In this way, HCA builds up the projection function f: $\{0, . . ., n\} \rightarrow [0, 255]$ such that f_i is the median value of the G values over Pi. We choose to compute the median as it is not affected by outlier values due, for example, to the presence of crystals in the paste.

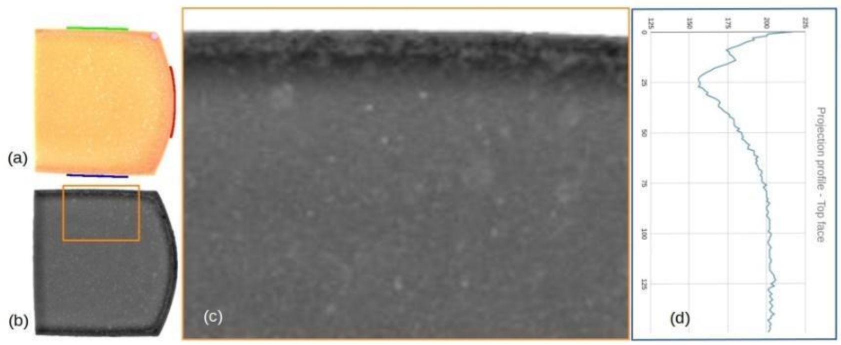

Figure 26: *Procedure of image cropping and parameter extraction for Handmade algorithm. In (a) a right slice, where the parts A, B and C are marked in color. In (b) an intensity normalization of (a). In (c), the intensity values of a region adjacent to part A are considered to build the plot in (d), representing the mean intensity value of the designed set of pixels.*

In Figure 26, the plot of the average intensity value shows for most of the images a shape with low values at the very first part of the rind (darker), increasing inside the rind and sub-rind area, followed by a flat area generated when scanning the paste region. HCA estimates the rind thickness for each part of interest P by analyzing the generated intensity distribution *f* as follows.

1. Compute the value V_j of average intensity *f* in the plateau zone. V_j is estimated as the median of all the pixels belonging to the zone obtained by a strong shrink of the mask F (corresponding to 2 cm, to overpass the rind) and the rectangle R of the part P, to capture a paste-only zone.

2. Compute the local minima of S_f and the depth of their basins, where S_f indicates a Gaussian smoothing of the average intensity. Discard the minima with low depth and those with ordinate too close to V_j; select the minimum M = (Mi, Mj) having the greater abscissa i.

3. Determine the transition point $U = (U_i, U_j)$ with $U_i > M_i$. The transition point should define the end of the rind zone. Depending on the cheese slice this can be marked: this is expressed by the slope f between M and the starting point of plateau V. In the current implementation of HCA U_j is defined as:

$$U_j = (1 - \mu)M_j + \mu V_j$$

where μ is a real-valued coefficient ranging between 0 and 1. In this implementation, μ has been set empirically to 0.3.

The end of the rind region is determined in HCA by selecting in the set $f - 1(U_j)$ the point in the interval (M_i, V_i) with greater abscissa. We observe that the value of μ has little influence on the position of Ui if the ramp between M and V is steep, i.e., when the separation between rind and paste is quite clear. The position of U_i can be influenced more significantly by the choice of μ in the case where the transition between rind and paste is very smooth. This agrees with the uncertainty of different annotators in cheese slices that exhibit a gradual transition. A graphical representation of the estimated values by this algorithm are reported in the plot in Figure 27.

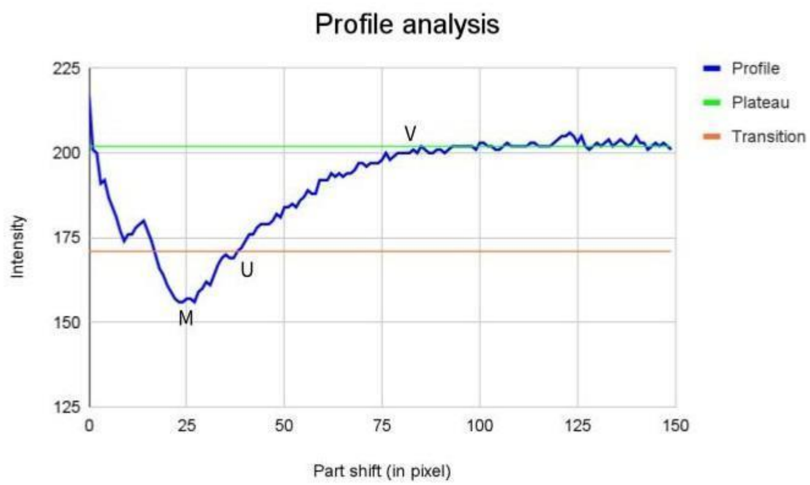

Figure 27: *Score plot representing the profile analysis: the green line highlights the plateau, while the estimated transition line, between the plateau and the selected minimum **M** of the intensity, is depicted in orange. **U** indicates the value used to compute the rind thickness which corresponds to its abscissa.*

6.3.2.6 Artificial Neural networks

Initially, as a pre-processing step, from each image were extracted three parts called P_A, P_B, P_C, respectively containing the upper plate, the heel, and the lower plate.

A ResNet18 model, a well-known architecture for image recognition tasks (He *et al.* 2016), was developed. A fully connected layer replaced the final layer, returning a single value. The whole network was trained from scratch testing with a lower performance by fine-tuning a backbone pre-trained on ImageNet (Russakovsky *et al.* 2015). The deeper network ResNet50 was tried without achieving better results. As ResNet18 resulted superior to the deeper ResNet50, a shallower network, i.e., a ResNet variant with only 10 layers was tested, but did not boost the performance. Other architecture and approaches tested were ShuffleNet (Zhang *et al.* 2018), RegNet (Radosavovic *et al.* 2020), ConvNeXt (Li *et al.* 2022), and CLIP (Radford *et al.* 2021). However, there were no significant differences between their performances and the previous one. A more in-depth investigation on the reasons for this behavior is left for future work.

6.3.3 Results and discussion

6.3.3.1 Performances of the models

The performance of the deep learning-based method is compared with the performance of the HCA and the original dataset of averaged quality evaluations by the panel of human experts. Figure 28 shows the performance of the proposed deep learning method compared with the HCA algorithm in terms of Mean Absolute Error (MAE) for the three slice regions and for the overall mean. The human evaluation consists of the error of the individual evaluations of the panel of 12 experts used to estimate the ground truth. The MAE of the average value of the three regions of the deep learning method is 0.51 mm, while the MAE of the HCA is 1.10 mm, which is similar to the MAE for a generic human evaluator, about 1.24 mm.

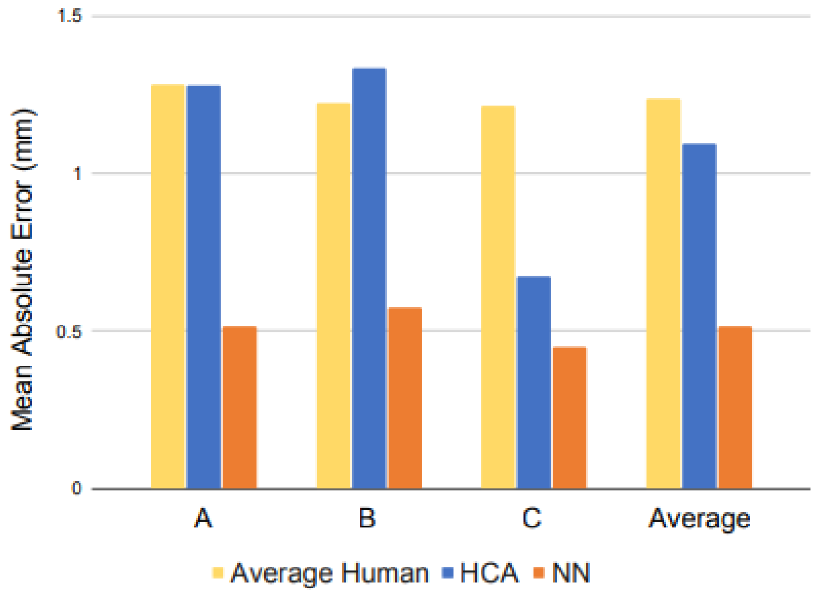

Figure 28: *Bar plot showing the comparison of the error produced by a virtual average expert (yellow), the HCA algorithm (blue), and the proposed Neural Network (NN) approach (orange). Mean Absolute Error (in millimeters) is reported separately for regions A, B, C, and for the overall mean***.**

Compared to a classical image analysis method based on manual feature selection, the deep learning approach achieves significantly better performance and avoids the critical tasks of selecting features typical of traditional methods.

To validate further the network, the areas of the input images that most influenced the final estimate were plotted using a Grad-CAM variant for visual explanations of network decisions in regression problems (Selvaraju et al.2017). In Figure 29 the reasonable areas of the input images are used by the model for predicting the thickness of the image.

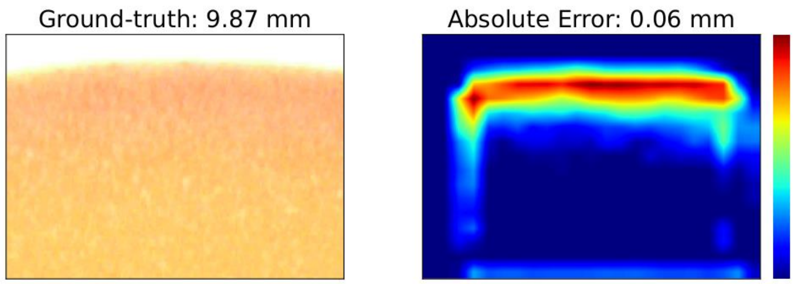

Figure 29: *Comparison of the original image crop for the estimation of the upper segment extraction and the estimation of the importance of each pixel in the estimation of the result.*

In Figure 30, the values estimated by the network are compared with the sorted values of the ground-truth thickness of the regions. We can observe that the network maintains, in trend, the ordering of measures.

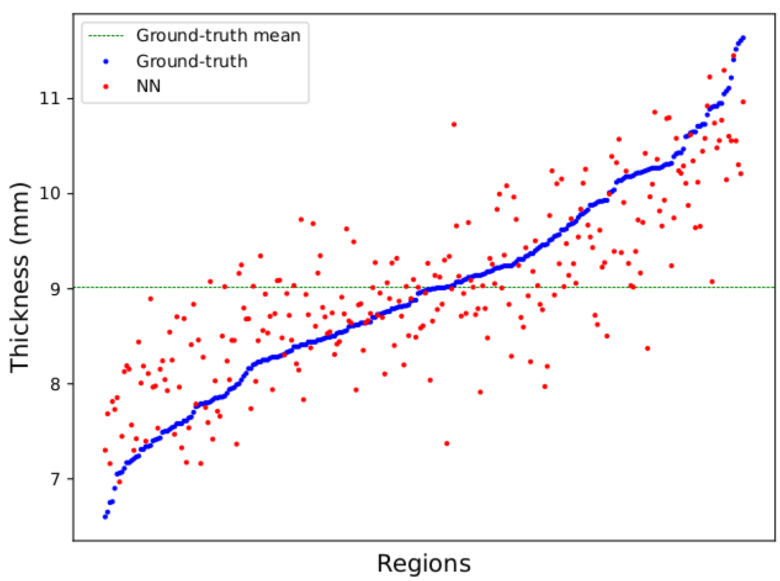

Figure 30: *Plot reporting the values predicted by the ANN model compared with the ground truth values. Ground truth values are reported in blue and are in increasing order, while predicted values are reported in red. The continuous green line represents the mean of the observed values.*

As reported in Figure 30, the artificial neural network algorithm overestimates the measurements when the wheel has a thicker rind and underestimates it for regions with a thinner rind. The R^2-score of 0.68 means that the MSE of the model is less than a third of the standard deviation of the ground truth values.

6.3.4 Conclusion

The research objective was the quality assessment of PDO Trentingrana cheese proposing a deep learning-based method to estimate the thickness of the rind, which currently is estimated by a pool of experts, but their work requires time and is rather expensive.

The automation of this procedure is functional to support panels in visual quality evaluation. The artificial neural network algorithm proposed implements a regression technique that learns the cheese thickness from a set of measurements deduced by the experts' annotations.

It is worth saying that in the present experiment, the amount of data collected is quite low compared to the number of observations that a deep learning model requires to be properly trained. A data augmentation procedure was applied to reach enough measurements, but the results were not related to the final measurements. A larger dataset for the training procedure of the model is required to improve the overall performance of the model and its adaptability to more real-case samples.

The Mean Absolute Error (MAE) reported for the Neural Network model is 0.51mm, i.e., less than half the MAE of a generic human evaluator. Moreover, the approach enabled better results than a hand-crafted method, which was specifically implemented to localize accurately the transition between cheese rind and paste.

From the present results, the next activity required is the estimation of the rind thickness using images depicting rock-cracked cheese slices instead of wire-cut slices. Analyzing these images, standard image analysis algorithms based on color change, like the one presented in this work, cannot work properly, because there are fewer visual changes in the transition from the rind to the grain.

Further objectives consist in inferring a quality evaluation from the physical parameters estimated by the judges. Assuming that the evaluation of the quality is a comprehensive evaluation of the dimension, color, uniformity, and distribution of the rind along with the presence of defects (such as the accumulation of dehydrated grain behind the rind), it is reasonable that an algorithm can recognize the presence of those visual parameters, and estimate if are present excessive anomalies and if the image can be associated to a positive or a negative evaluation.

Finally, rind thickness is only one of the visual quality features considered by the quality panel of the Trentingrana Consortium, hence the automatic estimation of other visual characteristics, such as paste color and texture, could be a topic for future research to provide more comprehensive support to cheese producers.

CHAPTER 7. CONCLUSIONS

In the last years, a general improvement in measurement technologies and an increasing concern about many different quality parameters caused an overall increase in the amount of measurement collected systematically in the context of food production (Rathore 2014).

The collection of large amounts of data from the complex experimental design of quality control requires the adoption of improved statistical analysis to correctly interpret those data and to develop reliable procedures that are functional for an automated analytical procedure.

The statistical procedures proposed in chapters 3 and 4 are procedures capable of inferring the correct information from complex experimental designs which are present in many productions process.

The parameters that are discussed in the two works reported in chapters 3 and 4 are related to the production processes, the season effect, and the treatment of raw material, such as the milk collection procedure. The proposed statistical framework can be easily adapted to different food production considering for example different kinds of measurement, related to other quality parameters, or to different experimental designs.

In the present work, the objective was to present and discuss a reliable statistical procedure to estimate the significance of predictive models adopting a permutation test, to guarantee the availability of information from statistical or machine learning models only after a reliable exclusion of the null hypothesis. Furthermore, the estimation of the predictive power can measure the effectiveness of the models developed and how they can really explain the relationship between the sensory and the instrumental data. The adoption of these procedures allows a quantitative estimation of the effectiveness of the models adopted and consequently for the interpretation of the correlation structure between sensory properties and instrumental measurements.

The activity in chapter 6 consists of a first step into the development of a rapid quality control procedure functional for the improvement of the Process Analytical Technology of the Trentingrana case study. Further research aims to estimate the quality evaluation of Trentingrana cheese on the parameter of "rind thickness" from the estimation of the dimension of the rind and its distribution in the whole cheese wheel. This represents an interesting example of the optimization of quality control procedures, simplifying the activity of human quality evaluators adding a previous step of evaluation by a predictive model, and proposing to human evaluation only cheese wheels that have been detected as problematic, or not fitting all the automatized standard set.

Interesting improvements for the procedure proposed here consist in developing statistical procedures capable of more accurate clustering procedures after multivariate estimation of the effect of factors. The development of an improved ASCA structure, adopting the analytical approach of Multiple Factor Analysis to the estimation of multiple series of Analysis of Variance is a possible direction that needs to be considered. Another interesting improvement could be the application of the PARAFASCA algorithm in the context of time-dependent sensory analysis: data from Temporal Dominant Sensation (TDS) and Temporal Check All That Apply (TCATA) still have got nuances due to their multi-way structure (Meyners & Castura 2018). The PARAFAC decomposition of the values of multiple analyses of variance applied to all the multi-way structures could propose an overall estimation of the effects of multiple confounding factors in this experimental design.

Another possible application of the ASCA-based statistical procedures proposed in this thesis is the analysis of data collected from quality control procedures collected in different food manufacturers, to test the inferring capacity of ASCA and ASCA-based models in other real-case frameworks and their effectiveness in comparing data from different manufacturers.

It is also worth proposing the development of a statistical tool equipped with a user-friendly interface to increase the adoption of those statistical tools in a wider range of analytical contexts.

REFERENCES

Abdi H., & Williams, L.J., 2010. Principal component analysis. *WIREs Comp Stat*, 2. 433-459. https://doi.org/10.1002/wics.101

Acar E. & Yener B., 2009. Unsupervised Multiway Data Analysis: A Literature Survey. *Knowledge and Data Engineering, IEEE Transactions,* on. 21. 6 - 20. https://doi.org/10.1109/TKDE.2008.112

Adachi K., Igoshi A., Murata M., 2020. Analyses of factors affecting the browning of model processed cheese during storage. *Journal Nutritional Science and Vitaminology*, 66, 364–369. https://doi.org/10.3177/jnsv.66.364

Amit Y. & Geman D., 1997. Shape Quantization And Recognition With Randomized Trees. *Neural Computation*. 9, 1545-1588. https://doi.org/10.1162/neco.1997.9.7.1545

Andrewes P., Bullock S., Turnbull R., Coolbear T., 2021. Chemical instrumental analysis versus human evaluation to measure sensory properties of dairy products: What is fit for purpose?, *International Dairy Journal*, 121, https://doi.org/10.1016/j.idairyj.2021.105098Astuti W., Dewanto S., Edi K., Soebandrija N., Tan S., 2018 Automatic fruit classification using support vector machines: a comparison with artificial neural network, *IOP Conf. Ser.: Earth Environmental Sciences* 195, https://doi.org/10.1088/1755-1315/195/1/012047

Badaró A.T., Hebling e Tavares J.P., Blasco J., Aleixos-Borrás N., Barbin D.F., 2022. Near infrared techniques applied to analysis of wheat-based products: Recent advances and future trends. *Food Control*, 140, 109115, https://doi.org/10.1016/j.foodcont.2022.109115

Banks, J.M., 2007. What general factors affect the texture of hard and semihard cheeses? In Cheese Problems Solved; McSweeney, P.L.H., Ed.; Woodhead Publishing Limited: Sawston, UK, pp. 200–201.

Barbieri G., Bolzoni L., Careri M., Mangia A., Parolari G., Spagnoli S., Virgili R., 2014. Study of the Volatile Fraction of Parmesan Cheese. *Journal of Agricultural and Food Chemistry*. 42(5), 1170-1176. https://doi.org/10.1021/jf00041a023

Barnes D. L., Harper S. J., Bodyfelt F. W., McDaniel M. R., 1991. Prediction of Consumer Acceptability of Yogurt by Sensory and Analytical Measures of Sweetness and Sourness. *Journal of Dairy Science*, 74(11), 3746-3754. https://doi.org/10.3168/jds.S0022-0302(91)78566-4Baykal H., & Yildirim H. K., 2013. Application of artificial neural networks (ANNs) in wine technology. *Critical reviews in food science and nutrition*, 53(5), 415–421. https://doi.org/10.1080/10408398.2010.540359

Becue-Bertaut, M. and Pages, J., 2008. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. Computational Statistic and Data Analysis, 52, 3255-3268, https://doi.org/10.1016/j.csda.2007.09.023

Belgiu M., Drăguţ L., 2016. Random Forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31, https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bellesia, F., Pinetti, A., Pagnoni, U.M., Rinaldi, R., Zucchi, C., Caglioti, L., Palyi, G., 2003. Volatile components of Grana Parmigiano-Reggiano type hard cheese. *Food Chemistry*, 83, 55–61. https://doi.org/10.1016/S0308-8146(03)00037-2

Bertinetto C., Engel J., Jansen J., 2020. ANOVA simultaneous component analysis: A tutorial review. *Analytica Chimica Acta*, X, p. 6, https://doi.org/10.1016/j.acax.2020.100061

Bittante G., Cologna N., Cecchinato A., De Marchi M., Penasa M., Tiezzi F., Endrizzi I., Gasperi F., 2011, Monitoring of sensory attributes used in the quality payment system of Trentingrana cheese, *Journal of Dairy Science*, 94(11), 5699-5709, https://doi.org/10.3168/jds.2011-4319

Bley M.E., Johnson M.E., Olson N.F., 1985. Factors Affecting Nonenzymatic Browning of Process Cheese. Journal of Dairy Science. 68, 555–561. https://doi.org/10.3168/jds.S0022-0302(85)80858-4

Bona E., Silva R., Borsato D., Bassoli, D., 2011. Optimized Neural Network for Instant Coffee Classification through an Electronic Nose. *International Journal of Food Engineering*. 7. https://doi.org/10.2202/1556-3758.2002

Bosakova-Ardenska A., Andreeva H., Danev A., Panayotov P., Boyanova P., 2021. Cut surface of Bulgarian white brined cheese evaluation by image analysis in HSI color space. *IOP Conference Series: Materials Science and Engineering*. 1031. 012114. https://doi.org/10.1088/1757-899X/1031/1/012114

Bovolenta S., Romanzin A., Corazzin M., Spanghero M., Aprea E., Gasperi F., Piasentier E., 2014.Volatile compounds and sensory properties of Montasio cheese made from the milk of Simmental cows grazing on alpine pastures, *Journal of Dairy Science*, 97(12), 7373-7385, https://doi.org/10.3168/jds.2014-8396

Breiman L., 1996. Bagging predictors. *Machine Learning*, 24, 123–140 (1996). https://doi.org/10.1007/BF00058655

Breiman L., 2001. Random Forests. *Machine Learning* 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brosnan T., Sun D., 2004. Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering*, 61(1), 3-16, https://doi.org/10.1016/S0260-8774(03)00183-3

Bryant A., Ustunol Z., Steffe J., 1995. Texture of Cheddar cheese as influenced by fat reduction, *Journal of Food Science*, 60, 1216-1219, 1236. https://doi.org/10.1111/j.1365-2621.1995.tb04559.x

Buvé C., Saeys W., Rasmussen M. A., Neckebroeck B., Hendrickx M., Grauwet T., Van Loey A., 2022. Application of multivariate data analysis for food quality investigations: An example-based review, *Food Research International*, 151, https://doi.org/10.1016/j.foodres.2021.110878

Cairone F., Carradori S., Locatelli M., Casadei M., Cesa S., 2020. Reflectance colorimetry: a mirror for food quality. A mini review. *European Food Research and Technology*. 246. 259-272. https://doi.org/10.1007/s00217-019-03345-6

Caraffa A., Ricci M., Lecca M., Modena C., Aprea E., Gasperi F., Messelodi S., 2023. A Deep Learning Approach for Estimating the Rind Thickness of Trentingrana Cheese from Images. In *Proceedings of the 3rd International Conference on Image Processing and Vision Engineering - IMPROVE*, SciTePress, 76-83. https://doi.org/10.5220/0011830000003497

Careri M., Spagnoli S., Panari G., Zannoni M., Barbieri, G., 1996. Chemical parameters of the non-volatile fraction of ripened Parmigiano-Reggiano cheese. *International Dairy Journal*, 6, 147–155. https://doi.org/10.1016/0958-6946(94)00056-5

Cervantes J., Garcia-Lamont F., Rodríguez-Mazahua L., Lopez A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends, *Neurocomputing*, 408, 189-215, https://doi.org/10.1016/j.neucom.2019.10.118

Field Code Changed

Cevoli C., Cerretani L., Gori A., Caboni M.F., Gallina Toschi T., Fabbri A., 2011. Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC–MS analysis of volatile compounds, *Food Chemistry*, 129(3), 1315-1319, https://doi.org/10.1016/j.foodchem.2011.05.126

Chaunier L. & Della Valle G. & Lourdin, D., 2007. Relationships between texture, mechanical properties and structure of cornflakes. *Food Research International*. 40. 493-503. https://doi.org/10.1016/j.foodres.2006.07.014

Cipolat-Gotet, C., Cecchinato, A., de Marchi, M., Bittante, G., 2013. Factors affecting the variation of different measures of cheese yield and milk nutrient recovery from an individual model cheese-manufacturing process. Journal Dairy Science. 96, 7952–7965. https://doi.org/10.3168/jds.2012-6516

[Dataset] CLAL, 2023a. *Italia: Rapporto fra produzione ed export di Formaggi* https://www.clal.it/?section=prod_export (Accessed 19 February 2023).

[Dataset] CLAL, 2023b. *Italia: Produzioni Trentingrana DOP*, https://www.clal.it/?section=produzioni_trentingrana (Accessed 19 February 2023).

Collins Y.F., McSweeney P.L.H., Wilkinson M.G., 2004. Lipolysis and catabolism of fatty acids in cheese, in: P.F. Fox, L. Paul, H. McSweeney, T.M. Cogan, T.P. Guinee (Eds.), Cheese: Chemistry, physics and microbiology, Academic Press, pp. 373-389, https://doi.org/10.1016/S1874-558X(04)80075-7

Coppi R., 1994. An introduction to multiway data and their analysis, *Computational Statistics & Data Analysis*, 18(1), 3-13, https://doi.org/10.1016/0167-9473(94)90130-9

Cousineau, D. & Chartier, S., 2010. Outliers detection and treatment: A review. *International Journal of Psychological Research*. 3. https://doi.org/10.21500/20112084.844

CREA, 2022. Commercio con l’estero dei prodotti agroalimentari 2021, https://www.crea.gov.it/documents/20126/0/Rapporto+commercio+estero+2021.pdf/8bb8ac3b-2bf6-5ea8-e1ee-accc68211602?t=1658218039620 (Accessed 19 February 2023).

Curtin Á.C., McSweeney P.L.H., 2004. Catabolism of amino acids in cheese during ripening. In: P.F. Fox, P.L.H. McSweeney, T.M. Cogan, T.P. Guinee (Eds.), Cheese: Chemistry, physics and microbiology, Academic Press, pp. 435-454, https://doi.org/10.1016/S1874-558X(04)80077-0

Derington A.J., Brooks J.C., Garmyn A.J., Thompson L.D., Wester D.B., Miller, M.F., 2011. Relationships of slice shear force and Warner-Bratzler shear force of beef strip loin steaks as related to the tenderness gradient of the strip loin, *Meat Science*, 88(1), 203-208, https://doi.org/10.1016/j.meatsci.2010.12.030

Divine R.D., Sommer D., Lopez-Hernandez A., Rankin S.A., 2012. Short communication: Evidence for methylglyoxal-mediated browning of Parmesan cheese during low-temperature storage. *Journal of Dairy Science*, 95(5), 2347-2354, https://doi.org/10.3168/jds.2011-4828

Drake M.A., Gerard P.D., Truong V.D., Daubert C.R., 1999. Relationship between instrumental and sensory measurements of cheese texture. Journal Texture Studies. 30, 451-476. https://doi.org/10.1111/j.1745-4603.1999.tb00230.x

Drake M.A., Civille G.V., 2003. Flavor Lexicons. Comprehensive Review Food Science, 2, 33-40. https://doi.org/10.1111/j.1541-4337.2003.tb00013.x

Drake M.A., 2007. Invited Review: Sensory Analysis of Dairy Foods, Journal of Dairy Science, 90(11), 4925-4937, https://doi.org/10.3168/jds.2007-0332

Du, C., & Sun D., 2004. Recent Developments in the Applications of Image Processing Techniques for Food Quality Evaluation. *Trends in Food Science & Technology*. 15:230–249. https://doi.org/10.1016/j.tifs.2003.10.006

Duan S., Liu F., Qin Q., Jia Q., Cao X., Hua Z., Fan Y., Wang C., 2023. Implementation of the HACCP System for Apple Juice Concentrate Based on Patulin Prevention and Control. *Foods*. 12(4):786. https://doi.org/10.3390/foods12040786

Dumičić, K. & Žmuk, B., 2013. Use of Power Analysis in Choosing Appropriate Sample Size for Quality Inspection. *Poslovna Izvrsnost*. 147-160.

Ellis A.M., Mayhew C.A., 2014. PTR-MS in the food sciences, in: Ellis A.M., Mayhew C.A. (Eds.), Proton transfer reaction mass spectrometry, pp. 221-265, https://doi.org/10.1002/9781118682883.ch6

Endrizzi I., Fabris A., Biasioli F., Aprea E., Franciosi E., Poznanski E., Cavazza A., Gasperi F., 2012. The effect of milk collection and storage conditions on the final quality of Trentingrana cheese: Sensory and instrumental evaluation, *International Dairy Journal*, 23(2), 105-114, https://doi.org/10.1016/j.idairyj.2011.10.004.

Endrizzi, I.; Aprea, E.; Biasioli, F.; Corollaro, M.L.; Demattè, L.; Penasa, M.; Bittante, G.; Gasperi, F., 2013. Implementing sensory analysis principles in the quality control of PDO products: A critical evaluation of a real-world case study. *Journal of Sensory Studies*, 28, 14–24 https://doi.org/10.1111/joss.12018

Eriksson l., Byrne T., Johansson E., Trygg J., Vikstrom C., 2013. *Multi- and Megavariate data analysis, Basic principles and Applications, third revised edition*, Umetrics Academy

Erkmen, Osman. (2022). Microbiological Analysis of Foods and Food Processing Environments. Academic Press. https://doi.org/10.1016/C2021-0-01219-0

Escofier B., Pagès J., 1994. Multiple factor analysis (AFMULT package), *Computational Statistics & Data Analysis*, 18(1), 121-140, https://doi.org/10.1016/0167-9473(94)90135-X

Eugster E., Fuchsmann P., Schlichtherle-Cerny H., Bütikofer U., Irmler S., 2019. Formation of alanine, α-aminobutyrate, acetate, and 2-butanol during cheese ripening by Pediococcus acidilactici FAM18098. *International Dairy Journal*, 96, 21-28, https://doi.org/10.1016/j.idairyj.2019.04.001

EUR-lex 1996. Commission Regulation (EC) No 1107/96 of 12 June 1996. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:01996R1107-20081120 (Accessed 19 February 2023)

[Data set] European Commission, Directorate-General for Communications, 2014. *Special Eurobarometer 389: Europeans' attitudes towards food security, food quality and the countryside* (v1.00). http://data.europa.eu/88u/dataset/S1054_77_2_EBS389

Fabris, A., Biasioli, F., Granitto, P. M., Aprea, E., Cappellin, L., Schuhfried, E., Soukoulis, C., Märk, T. D., Gasperi, F., & Endrizzi, I., 2010. PTR-TOF-MS and data-mining methods for rapid characterization of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. *Journal of mass spectrometry: JMS*, 45(9), 1065–1074. https://doi.org/10.1002/jms.1797

FAO, 2003. Important Food Issues, in: FAO/WHO (Eds.) Assuring Food Safety And Quality: Guidelines For Strengthening National Food Control Systems. FAO Food And Nutrition Papers. https://www.fao.org/3/y8705e/y8705e.pdf

Faustini M, Quintavalle Pastorino G, Colombani C, Chiesa LM, Panseri S, Vigo D, Curone G., 2019. Volatilome in Milk for Grana Padano and Parmigiano Reggiano Cheeses: A First Survey. Veterinary Sciences. 6(2), 41. https://doi.org/10.3390/vetsci6020041

Fawagreh K., Gaber M. M. & Elyan E., 2014. Random forests: from early developments to recent advancements, *Systems Science & Control Engineering*, 2:1, 602-609, https://doi.org/10.1080/21642583.2014.956265

FDA, 2004. PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance Guidance for Industry https://www.fda.gov/media/71012/download (Accessed 19 February 2023)

Ferri C., Hernández-Orallo J., Modroiu R., 2009. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1), 27-38, https://doi.org/10.1016/j.patrec.2008.08.010

Fielding L., Ellis L., Clayton D., Peters A., 2011. An evaluation of process specific information resources, aimed at hazard analysis, in Small and Medium Enterprises in food manufacturing. *Food Control*, 22(8), 1171-1177, https://doi.org/10.1016/j.foodcont.2011.01.011

Fisher, R. A., 1935. The design of experiments. Oliver & Boyd. Edinburgh, United Kingdom

Foegeding E.A., Drake M.A., 2007. Invited Review: Sensory and Mechanical Properties of Cheese Texture, *Journal of Dairy Science*, 90(4), 1611-1624, https://doi.org/10.3168/jds.2006-703

Fox, P.F.; Cogan, T.M., 2004. Factors that Affect the Quality of Cheese. In Cheese: Chemistry, Physics and Microbiology, III ed.; Elsevier: Amsterdam, Netherlands, 1, 583–608 https://doi.org/10.1016/S1874-558X(04)80084-8

Fox P.F., Guinee T.P., Cogan T.M., McSweeney P.L.H., 2017. Cheese: Structure, rheology and texture. P.F. Fox, T.P. Guinee, T.M. Cogan, P.L.H. McSweeney (Eds.), Fundamentals of cheese science, Springer, Boston, MA, 475-532, https://doi.org/10.1007/978-1-4899-7681-9_14

Franceschi P., Malacarne M., Formaggioni P., Cipolat-Gotet C., Stocco G., Summer A., 2019. Effect of Season and Factory on Cheese-Making Efficiency in Parmigiano Reggiano Manufacture. *Foods* 2019, 8, 315 https://doi.org/10.3390/foods8080315

Franciosi E., De Sabbata G., Gardini F., Cavazza A., Poznanski E., 2011. Changes in psychrotrophic microbial populations during milk creaming to produce Grana Trentino cheese. *Food Microbiology*, 28(1), 43-51, https://doi.org/10.1016/j.fm.2010.08.003

Franciosi E., Gardini F., Monfredini L., Tabanelli G., Fabris A., Endrizzi I., Poznanski E., Gasperi F., Cavazza A., 2012. Does milk treatment before cheesemaking affect microbial and chemical traits of ripened cheese? Grana Trentino as a case study. *Journal of Dairy Science*, 95(10), 5485-5494, https://doi.org/10.3168/jds.2011-4693

Fryer P. J., Robbins P. T., 2005. Heat transfer in food processing: ensuring product quality and safety, *Applied Thermal Engineering*, 25(16), 2499-2510, https://doi.org/10.1016/j.applthermaleng.2004.11.021

Galindo-Prieto, B., Eriksson, L. and Trygg, J., 2014. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS), Journal of Chemometrics, 28, 623– 632, https://doi.org/10.1002/cem.2627

Giraffa G., 2021. The microbiota of grana Padano cheese. A review. *Foods*, 10(11), 2632, https://doi.org/10.3390/foods10112632

Giusti, A.M., Bignetti, E. & Cannella, C., 2008. Exploring New Frontiers in Total Food Quality Definition and Assessment: From Chemical to Neurochemical Properties, *Food Bioprocess Technology*, 1, 130–142. https://doi.org/10.1007/s11947-007-0043-9

Granitto, P.M., Furlanello, C., Biasioli, F., & Gasperi, F., 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemometrics and Intelligent Laboratory Systems, 83, 83-90, https://doi.org/10.1016/J.CHEMOLAB.2006.01.007

Greenhoff K., MacFie H.J.H., 1994. Preference mapping in practice. In: MacFie H.J.H., Thomson D.M.H. (Eds.), *Measurement of/ food preferences*, Springer US, pp. 137-166. https://doi.org/10.1007/978-1-4615-2171-6_6

Greve, P., Lee, Y.S., Meullenet, J.-F. Kunz, B., 2010. Improving the prediction for sensory texture attributes for multicomponent snack bars by optimizing instrumental test conditions. *Journal of Texture Studies*, 41, 358-380. https://doi.org/10.1111/j.1745-4603.2010.00229.x

Grunert, K. G., 2005. Food Quality and Safety: Consumer Perception and Demand. *European Review of Agricultural Economics*. 32. 369-391. https://doi.org/10.1093/eurrag/jbi011

Grunert K. G., Aachmann K., 2016. Consumer reactions to the use of EU quality labels on food products: A review of the literature, *Food Control*, 59, 178-187, https://doi.org/10.1016/j.foodcont.2015.05.021

Hartigan J.A., & Wong M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Journal Royal Statistical Society, 28, 100–108. https://doi.org/10.2307/2346830

ICH Q9 Pharmaceutical Development, 2023, Quality guidelines – Q9: Quality Risk Management, Concept Paper. https://database.ich.org/sites/default/files/Q9-R1_Concept%20Paper_2020_1113.pdf (Accessed 19 February 2023). Jackman, P., & Sun, D. 2013. Recent advances in image processing using image texture features for food quality assessment. *Trends in Food Science and Technology*, 29, 35-43. https://doi.org/10.1016/J.TIFS.2012.08.008

Jansen, J.J., Bro, R., Hoefsloot, H.C.J., van den Berg, F.W.J., Westerhuis, J.A. and Smilde, A.K., 2008. PARAFASCA: ASCA combined with PARAFAC for the analysis of metabolic fingerprinting data. *J. Chemometrics*, 22: 114-121. https://doi.org/10.1002/cem.1105

Johnsen, J., 2014. Food Industry Perspectives on the Implementation of a PAT Strategy. In: O'Donnell, C., Fagan, C., Cullen, P. (eds.) Process Analytical Technology for the Food Industry. Food Engineering Series. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-0311-5_12

Jolliffe I. T., 2002. Principal Component Analysis. Springer New York, NY. https://doi.org/10.1007/b98835

Kassambara A., 2017a. Practical Guide to Principal Component Methods in R; STHDA: Marseille, France.

Kassambara A. 2017b. Multivariate Analysis I: Practical Guide to Cluster Analysis in R. Unsupervised Machine Learning; STHDA: Marseille, France.

Kassambara A., 2020. Ggpubr: 'ggplot2' based publication ready plots. R Package Version 0.4.0. Available online: https://CRAN.R-project.org/package=ggpubr , (Accessed 19th February 2023)

Kassambara A., Mundt F., 2020. Factoextra: Extract and visualize the results of multivariate data analyses. R Package Version 1.0.7. https://CRAN.R-project.org/package=factoextra , (Accessed 19th February 2022)

Khattab A. R., Guirguis H. A., Tawfik S. M., Farag M. A., 2019. Cheese ripening: A review on modern technologies towards flavor enhancement, process acceleration and improved quality assessment, *Trends in Food Science & Technology*, 88, 343-360, https://doi.org/10.1016/j.tifs.2019.03.009

Kilcawley K.N., Faulkner H., Clarke H.J., O'Sullivan M.G., Kerry J.P., 2018. Factors influencing the flavour of Bovine milk and cheese from grass based versus non-grass based milk production systems. *Foods*, 7(3), 37, https://doi.org/10.3390/foods7030037

Kindstedt P.S., Polowsky P.J., 2021. Invited review: Crystals in cheese: More than a curiosity, *Journal of Dairy Science*, 104(2), 1233-1250, https://doi.org/10.3168/jds.2020-19119 Kupongsak S., Tan J., 2006. Application of fuzzy set and neural network techniques in determining food process control set points, Fuzzy Sets and Systems, 157(9), 1169-1178, https://doi.org/10.1016/j.fss.2005.12.015

He K., Zhang X., Ren S., Sun J., 2016. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 770-778, https://doi.org/10.1109/CVPR.2016.90

ISO, 2019. ISO 20613:2019 Sensory analysis — General guidance for the application of sensory analysis in quality control.

Laddi A., Sharma S., Kumar A., Kapur P., 2013. Classification of tea grains based upon image texture feature analysis under different illumination conditions. *Journal of Food Engineering*, 115(2), 226-231, https://doi.org/10.1016/j.jfoodeng.2012.10.018

Laguna L., Farrell G., Bryant M., Morina A., Sarkar A., 2017. Relating rheology and tribology of commercial dairy colloids to sensory perception. *Food & Function*, 8(2), 563-573. https://doi.org/10.1039/C6FO01010E

Lazraq A., Cléroux R., Gauchi J.P., 2003. Selecting both latent and explanatory variables in the PLS1 regression model. *Chemometrics and Intelligent Laboratory Systems*, 66(2), 117-126, https://doi.org/10.1016/s0169-7439(03)00027-3

Lei T., & Sun D.W., 2019. Developments of nondestructive techniques for evaluating quality attributes of cheeses: A review. *Trends in Food Science & Technology*, 88, 527-542, https://doi.org/10.1016/j.tifs.2019.04.013

Leitzmann, C., 1993. Food Quality—Definition and a Holistic View. *Safeguarding Food Quality*, 3–15. https://doi.org/10.1007/978-3-642-78025-7_2

Li Z., Gu T., Li B., Xu W., He X., Hui X., 2022. ConvNeXt-Based Fine-Grained Image Classification and Bilinear Attention Mechanism Model. *Applied Sciences*. 12(18). 9016. https://doi.org/10.3390/app12189016

Liaw I.W., Miracle R.E., Jervis S.M., Listiyani M.A.D., Drake M.A., 2011. Comparison of the flavor chemistry and flavor stability of Mozzarella and Cheddar wheys. *Journal of Food Science*, 76(8), 1188-1194, https://doi.org/10.1111/j.1750-3841.2011.02360.x

Lin H., Jiang H., Adade S. Y. S. S., Kang W., Xue Z., Zareef M., Chen Q., 2022. Overview of advanced technologies for volatile organic compounds measurement in food quality and safety, *Critical Reviews in Food Science and Nutrition*, https://doi.org/10.1080/10408398.2022.2056573

Liu S.Q., Holland R., Crow V.L., 2004. Esters and their biosynthesis in fermented dairy products: A review. *International Dairy Journal*, 14(11), 923-945, https://doi.org/10.1016/j.idairyj.2004.02.010

Lukinac J., Jukić M., Habschied K., Lučan M., 2018. Application of computer vision and image analysis method in cheese-quality evaluation: a review. *Ukrainian Food Journal*. 7. 192-214. https://doi.org/10.24263/2304-974X-2018-7-2-4

Ma Y., Barbano D.M., 2000. Gravity separation of raw Bovine milk: Fat globule size distribution and fat content of milk fractions. *Journal of Dairy Science*, 83(8), 1719-1727, https://doi.org/10.3168/jds.S0022-0302(00)75041-7

Ma J., Sun D. W., Pu H., 2016. Spectral absorption index in hyperspectral image analysis for predicting moisture contents in pork longissimus dorsi muscles. *Food Chemistry*, 197(A), 848-854, https://doi.org/10.1016/j.foodchem.2015.11.023

Mannion D.T., Furey A., Kilcawley K.N., 2016. Free fatty acids quantification in dairy products. *International Journal of Dairy Technology*, 69(1), 1-12

Marilley L., Casey M.G., 2004. Flavours of cheese products: Metabolic pathways, analytical tools and identification of producing strains. *International Journal of Food Microbiology*, 90(2), 139-159, https://doi.org/10.1016/S0168-1605(03)00304-0

Marini F., Bucci R., Magrì A.L., Magrì A.D., 2008. Artificial neural networks in chemometrics: History, examples and perspectives, *Microchemical Journal*, 88(2), 178-185, https://doi.org/10.1016/j.microc.2007.11.008

Martin, M, Govaerts, B., 2020. LiMM-PCA: Combining ASCA+ and linear mixed models to analyse high-dimensional designed data. *Journal of Chemometrics*. 34:e3232. https://doi.org/10.1002/cem.3232

McAuliffe O., Kilcawley K., Stefanovic E., 2019. Symposium review: Genomic investigations of flavor formation by dairy microbiota. *Journal of Dairy Science*, 102, 1, 909-922, https://doi.org/10.3168/jds.2018-15385.

McDermott, A., Visentin, G., McParland, S., Berry, D. P., Fenelon, M. A., & De Marchi, M., 2016. Effectiveness of mid-infrared spectroscopy to predict the color of bovine milk and the relationship between milk color and traditional milk quality traits. *Journal of dairy science*, 99(5), 3267–3273. https://doi.org/10.3168/jds.2015-10424

McSweeney P.L.H., Sousa M.J., 2000. Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Lait*, 80(3). 293-324, https://doi.org/10.1051/lait:2000127

McSweeney P.L.H., Ottogalli G., Fox P.F., 2004. Diversity of cheese varieties: An overview, in: Fox P.F., McSweeney P.L.H., Cogan T.M., Guinee T.P. (Eds.), Cheese: Chemistry, physics and microbiology, Academic Press, pp. 1-23, https://doi.org/10.1016/S1874-558X(04)80037-X

Meenu M., Kurade C., Neelapu B.C., Kalra S., Ramaswamy H. S., Yu Y., 2021. A concise review on food quality assessment using digital image processing. *Trends in Food Science & Technology*, 118(A), 106-124, https://doi.org/10.1016/j.tifs.2021.09.014

Mehta B.M., Kumar Jain A., Darji V. B., Aparnathi K. D., 2018. Evaluation of different methods to monitor primary stage of oxidation of heat clarified milk fat (ghee). *Journal of Food Process Preservation*. 42(8). https://doi.org/10.1111/jfpp.13688

Meoni, G., Luchinat, C., Gotti, E., Cadena, A., Tenori, L., 2021. Phenotyping Green and Roasted Beans of Nicaraguan Coffea Arabica Varieties Processed with Different Post-Harvest Practices, *Applied Sciences*, 11, 11779. https://doi.org/10.3390/app112411779

Meyners M., Castura J. C., 2018. The analysis of temporal check-all-that-apply (TCATA) data, Food Quality and Preference, 67, 67-76, https://doi.org/10.1016/j.foodqual.2017.02.003

MiPAAF (Ministero delle Politiche Agricole Alimentari e Forestali) 2022. Modifica del disciplinare di produzione della denominazione «Grana Padano» registrata in qualità di denominazione di origine protetta in forza al regolamento (CE) n. 1107/96 della Commissione del 12 giugno 1996. Gazzetta Ufficiale Della Repubblica Italiana, 187, 84-92. https://www.gazzettaufficiale.it/eli/gu/2022/08/11/187/sg/pdf (accessed on 25February 2023).

Milovanovic B., Djekic I., Miocinovic J., Djordjevic V., Lorenzo J.M., Barba F.J., Mörlein D., Tomasevic I., 2020. What Is the Color of Milk and Dairy Products and How Is It Measured? *Foods*, 9, 1629. https://doi.org/10.3390/foods9111629

Montgomery D.C., 2013. Randomized blocks, Latin squares, and related designs, in: Montgomery D. C. (Ed.), Design and analysis of experiments, John Wiley & Sons Inc., pp. 139-182

Moran F., Sullivan C., Keener K., Cullen P., 2017. Facilitating smart HACCP strategies with Process Analytical Technology, *Current Opinion in Food Science*, 17, 94-99, https://doi.org/10.1016/j.cofs.2017.11.003

Mucchetti, G.; Gatti, M.; Nocetti, M.; Reverberi, P.; Bianchi, A.; Galati, F.; Petroni, A., 2014. Segmentation of Parmigiano Reggiano dairies according to cheese-making technology and relationships with the aspect of the cheese curd surface at the moment of its extraction from the cheese vat. *Journal Dairy Science*. 97, 1202–1209. https://doi.org/10.3168/jds.2013-7022

Muñoz A. M., 2002. Sensory evaluation in quality control: an overview, new developments and future opportunities, *Food Quality and Preference*, 13(6), 329-339, https://doi.org/10.1016/S0950-3293(02)00014-9

NIST/EPA/NIH (National Institute of Standards and Technology/Environmental Protection Agency/National Institutes of Health), 2014. NIST mass spectral library. National Institute of Standards and Technology, US Secretary of Commerce, Washington, DC

Noble A. C., & Ebeler S. E., 2002. Use of multivariate statistics in understanding wine flavor. *Food Reviews International*, 18:1, 1-20, https://doi.org/10.1081/FRI-120003414

Noël Y., Zannoni M., Hunter E.A., 1996. Texture of Parmigiano Reggiano cheese: Statistical relationships between rheological and sensory variates. *Lait*, 76, 243–254. https://doi.org/10.1051/lait:1996320

Oraguzie N., Alspach P., Volz R., Whitworth C., Ranatunga C., Weskett R., Harker R., 2009. Postharvest assessment of fruit quality parameters in apple using both instruments and an expert panel, *Postharvest Biology and Technology*, 52(3), 279-287, https://doi.org/10.1016/j.postharvbio.2009.01.004

Panseri S., Luca M., Zecconi A., Soncini G., Noni I., 2014. Determination of volatile organic compounds (VOCs) from wrapping films and wrapped PDO Italian cheeses by using HS-SPME and GC/MS. *Molecules*, 19(7), 8707-8724. https://doi.org/10.3390/molecules19078707

Parenteral Drug Association, 2012. Technical Report No. 54 Implementation of quality risk management for pharmaceutical and biotechnology manufacturing operations. Ed. Parenteral Drug association, Betheseda, USA.

Pathare, P.B., Opara, U.L. & Al-Said, F.AJ., 2013. Colour Measurement and Analysis in Fresh and Processed Foods: A Review. *Food Bioprocessing Technology* 6, 36–60. https://doi.org/10.1007/s11947-012-0867-9

Perrot N., Ioannou I., Allais I., Curt C., Hossenlopp J., Trystram G., 2006. Fuzzy concepts applied to food product quality control: A review. *Fuzzy Sets and Systems*, 157(9), 1145-1154, https://doi.org/10.1016/j.fss.2005.12.013

Pinto R.C., Trygg J., Gottfries J., 2012 Advantages of orthogonal inspection in chemometrics *Journal of Chemometrics*, 26, 231-235, https://doi.org/10.1002/cem.2441

Pisano M.B., Scano P., Murgia A., Cosentino S., Caboni P., 2016. Metabolomics and microbiological profile of Italian mozzarella cheese produced with buffalo and cow milk. *Food Chemistry*, 192, 618-624, https://doi.org/10.1016/j.foodchem.2015.07.061

Prentice J.H., 1992. Dairy Rheology—A Concise Guide; VCH: Weinheim, Germany.

Psomas E., & Fotopoulos C., 2010. Total quality management practices and results in food companies. *International Journal of Productivity and Performance Management*. 59. 668-687. https://doi.org/10.1108/17410401011075657

Qian M., Reineccius G., 2002. Identification of aroma compounds in Parmigiano-Reggiano cheese by gas chromatography/Olfactometry. *Journal of Dairy Science*, 85(6), 1362-1369, https://doi.org/10.3168/jds.S0022-0302(02)74202-1

Qin C., Liu L., Wang Y., Leng T., Zhu M., Gan B., Xie J., Yu Q., Chen Y., 2022. Advancement of omics techniques for chemical profile analysis and authentication of milk, *Trends in Food Science & Technology*, 127, 114-128, https://doi.org/10.1016/j.tifs.2022.06.001

Quinn B.P., & Marriott N.G., 2002. HACCP plan development and assessment: A review. *Journal of Muscle Foods*. 13. 313 - 330. https://doi.org/10.1111/j.1745-4573.2002.tb00339.x

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/ , (Accessed 19th February 2023)

Radford A., Kim J. W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I., 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020. https://doi.org/10.48550/arXiv.2103.00020

Radosavovic, I., Kosaraju R. P., Girshick R., He K., Doll´ar P., 2020. Designing Network Design Spaces. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 10428-10436. https://doi.org/10.48550/arXiv.2003.13678

Rajbhandari P., Kindstedt P.S., 2008. Characterization of Calcium Lactate Crystals on Cheddar Cheese by Image Analysis. *Journal of Dairy Science*, 91(6), 2190-2195, https://doi.org/10.3168/jds.2008-1077

Rasmus B., 1997. PARAFAC. Tutorial and applications, *Chemometrics and Intelligent Laboratory Systems*, 38(2), 149-171, https://doi.org/10.1016/S0169-7439(97)00032-4

Rathore, A., Winkle, H., 2009. Quality by design for biopharmaceuticals. *Nature Biotechnology* 27, 26–34. https://doi.org/10.1038/nbt0109-26

Rathore, A. S., Bhambure, R., & Ghare, V., 2010. Process analytical technology (PAT) for biopharmaceutical products. Analytical and bioanalytical chemistry, 398(1), 137–154. https://doi.org/10.1007/s00216-010-3781-x

Rathore A., 2014. QbD/PAT for bioprocessing: moving from theory to implementation, *Current Opinion in Chemical Engineering*, 6, 1-8, https://doi.org/10.1016/j.coche.2014.05.006

Ricci M, Gasperi F, Endrizzi I, Menghi L, Cliceri D, Franceschi P, Aprea E., 2022. Effect of Dairy, Season, and Sampling Position on Physical Properties of Trentingrana Cheese: Application of an LMM-ASCA Model. *Foods*. 11(1):127. https://doi.org/10.3390/foods11010127

Ricci M., Gasperi F., Betta E., Menghi L., Endrizzi I., Cliceri D., Franceschi P., Aprea E., 2023. Multivariate data analysis strategy to monitor Trentingrana cheese real-scale production through volatile organic compounds profiling, *LWT*, 173, 114364, https://doi.org/10.1016/j.lwt.2022.114364

Roland A. M., Phillips L. G., Boor K. J., 1999. Effects of Fat Content on the Sensory Properties, Melting, Color, and Hardness of Ice Cream. *Journal of Dairy Science*, 82(1), 32-38. https://doi.org/10.3168/jds.S0022-0302(99)75205-7

Rossi F., Gatto V., Sabattini G., Torriani S., 2012. An assessment of factors characterising the microbiology of Grana Trentino cheese, a Grana-type cheese. *International Journal of Dairy Technology*, 65(3), 401-409, https://doi.org/10.1111/j.1471-0307.2012.00844.x

Ruiz-Capillas C., Herrero A.M., Pintado T., Delgado-Pando G., 2021. Sensory Analysis and Consumer Research in New Meat Products Development. *Foods*, 16;10(2):429. https://doi.org/10.3390/foods10020429.

Russ, J.C. 2015. Image Analysis of Foods. *Journal of Food Science*. 80, E1974-E1987. https://doi.org/10.1111/1750-3841.12987

Russakovsky, O., Deng, J., Su, H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A. C., Fei-Fei. L., 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal Computer Vision.* 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sarkar A., Krop E. M., 2019. Marrying oral tribology to sensory perception: a systematic review. *Current Opinion in Food Science*, 27, 64-73. https://doi.org/10.1016/j.cofs.2019.05.007

Savitha, S. & Devi, K., 2022. Quality By Design: A Review. *Journal of Drug Delivery and Therapeutics*. 12. 234-239. https://doi.org/10.22270/jddt.v12i2-S.5451

Schanda, J. CIE Colorimetry in Colorimetry: Understanding the CIE System; Scahnda, J., Ed.; Wiley-Interscience: Hoboken, NJ, USA, 2007.

Shewan H. M., Stokes J. R., Smyth H. E., 2020. Influence of particle modulus (softness) and matrix rheology on the sensory experience of 'grittiness' and 'smoothness'. *Food Hydrocolloids*, 103, 105662. https://doi.org/10.1016/j.foodhyd.2020.105662

Smilde A.K., Timmerman M.E., Hendriks M.M.W.B., Jansen J.J., Hoefsloot H.C.J., 2012. Generic framework for high-dimensional fixed-effects ANOVA. *Briefing Bioinformatic*, 13, 524–535. https://doi.org/10.1093/bib/bbr071

Seisonen S., Vene K., Koppel K., 2016. The current practice in the application of chemometrics for correlation of sensory and gas chromatographic data. *Food Chemistry*, 210, 530-540, https://doi.org/10.1016/j.foodchem.2016.04.134

Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D., 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017*, 618-626, https://doi.org/10.1109/ICCV.2017.74

Smyth H., & Cozzolino D., 2013. Instrumental Methods (Spectroscopy, Electronic Nose, and Tongue) As Tools To Predict Taste and Aroma in Beverages: Advantages and Limitations. *Chemical Reviews*. 113(3), 1429-1440. https://doi.org/10.1021/cr300076c

Spence C., 2021. What Is the Relationship between the Presence of Volatile Organic Compounds in Food and Drink Products and Multisensory Flavour Perception? *Foods*. 10(7). 1570. https://doi.org/10.3390/foods10071570

Stamirova I., Kazura M., De Beer D., Joubert E., Schulze A.E., Beelders T., de Villiers A., Walczak B., 2013. High-dimensional nested analysis of variance to assess the effect of production season, quality grade and steam pasteurization on the phenolic composition of fermented rooibos herbal tea. *Talanta*. 115, 590–599. https://doi.org/10.1016/j.talanta.2013.06.023

Suh J.H., 2022. Critical review: Metabolomics in dairy science – evaluation of milk and milk product quality. *Food Research International*, 154, https://doi.org/10.1016/j.foodres.2022.110984

Suherman S., Janitra A. A., Budhiary K., Pratiwi W., Idris F., 2021. Review on hazard analysis and critical control point (HACCP) in the dairy product: Cheese. *IOP Conference Series: Materials Science and Engineering*. 1053. 012081. https://doi.org/10.1088/1757-899X/1053/1/012081

Tey Y. S., Rajendran N., Brindal M., Sidique S., Shamsudin M. N., Radam A., Abdul H., Ahmad H., 2016. A Review of an International Sustainability Standard (GlobalGAP) and its Local Replica (MyGAP). *Outlook on Agriculture*. 45. 67-72. https://doi.org/10.5367/oa.2016.0230 Thevenot E.A., Roux A., Xu Y., Ezan E., Junot C., 2015. Analysis of the human adult urinary metabolome variations with age, body mass index and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *Journal of Proteome Research*, 14(8), 3322-3335, https://doi.org/10.1021/acs.jproteome.5b00354

Thiel, M, Féraud, B, Govaerts, B., 2017. ASCA+ and APCA+: extensions of ASCA and APCA in the analysis of unbalanced multifactorial designs. Journal of Chemometrics. 31:e2895. https://doi.org/10.1002/cem.2895

Tomczak M., & Tomczak E., 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size, *Trends in sport science*, 21, 19-25.

Trygg J., Wold S., 2002. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, 16, 119-128, https://doi.org/10.1002/cem.695

Turgut S. S., Karacabey E., Kucukoner E., 2014. Potential of image analysis based systems in food quality assessments and classifications. Foodbalt 2014 - 9th Baltic Conference on Food Science and Technology “Food for Consumer Well-Being”, Conference Proceedings. 8-12.

van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A. et al., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7, 142. https://doi.org/10.1186/1471-2164-7-142

van Hoek P., Harms J., Wang X., Rathore A. S., 2009. Case study on definition of Process Design Space for Microbial Fermentation Step, in: Rathore, A.S., Mhatre R. (Eds.), Quality by Design for Biopharmaceuticals: Principles and Case Studies, Wiley Series in Biotechnology and Bioengineering, New York, pp.85-110.

Villeneuve M.-P., Lebeuf Y., Gervais R., Tremblay G.F., Vuillemard J.C., Fortin J., Chouinard P.Y., 2013. Milk volatile organic compounds and fatty acid profile in cows fed timothy as hay, pasture, or silage. *Journal of Dairy Science*, 96(11), 7181-7194, https://doi.org/10.3168/jds.2013-6785

Vinzi E. V., Trinchera L., & Amato S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. *Handbook of Partial Least Squares*. https://doi.org/10.1007/978-3-540-32827-8_3

Wang L., & Randolph H.E., 1978. Activation of lipolysis. I. Distribution of lipase activity in temperature activated milk. *Journal of Dairy Science*, 61(7), 874-880, https://doi.org/10.3168/jds.S0022-0302(78)83664-9

Wium H., Qvist K.B., Gross, M., 1997. Uniaxial compression of uf-feta cheese related to sensory texture analysis. *Journal of Texture Studies*, 28: 455-476. https://doi.org/10.1111/j.1745-4603.1997.tb00128.x

Wold H., 1966. Nonlinear estimation by Iterative Least Squares Procedures. In: F. N. David and J. Neyman, (Eds.), Research Papers in Statistics, Festschrift for J. Neyman, Wiley, London, New York.Wold S., Esbensen K., Geladi P., 1987. Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37-52 https://doi.org/10.1016/0169-7439(87)80084-9

Yang, Jiyun & Lee Jeehyun, 2019. Application of Sensory Descriptive Analysis and Consumer Studies to Investigate Traditional and Authentic Foods: A Review. *Foods*. 8(2):54. Doi: https://doi.org/10.3390/foods8020054

Yu P., Low M.Y., Zhou W., 2018. Design of experiments and regression modelling in food flavour and sensory analysis: A review. *Trends in Food Science & Technology*, 71, 202-215, https://doi.org/10.1016/j.tifs.2017.11.013

Zdunek, A., Konopacka, D. and Jesionkowska, K., 2010. Crispness and crunchiness judgment of apples based on contact acoustic emission. Journal of Texture Studies, 41. 75-91. https://doi.org/10.1111/j.1745-4603.2009.00214.x

Zhang X., Zhou X., Lin M., Sun J., 2018. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. Conference: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 6848-6856. https://doi.org/10.1109/CVPR.2018.00716

Zielinski A.A.F., Haminiuk C.W.I., Nunes C.A., Schnitzler E., van Ruth S.M., Granato D., 2014. Chemical Composition, Sensory Properties, Provenance, and Bioactivity of Fruit Juices as Assessed by Chemometrics: A Critical Review and Guideline. *Comprehensive Reviews in Food Science and Food Safety*, 13: 300-316. https://doi.org/10.1111/1541-4337.12060

Zhu L., Spachos P., 2021. Support vector machine and YOLO for a mobile food grading system, *Internet of Things*, 13, https://doi.org/10.1016/j.iot.2021.100359