Food traceability and authenticity assessment by application of NMR spectroscopy with a non-targeted approach

Abstract

In recent years, the interest of the research community and industry in investigation methods that ensure traceability and enable the valorisation of quality products has increased considerably. The development of such methods becomes essential to protect the commercial value of agri-food products and, at the same time, safeguard consumers' health. Under this scenario, there has been an increasing use of the so-called non-targeted approach. The application of this approach to food matrices allows to define their metabolic profile, providing a unique fingerprint for each matrix. This advantage has led the non-targeted method to become a viable strategy to address complex food quality control issues, including authenticity and traceability. Nuclear magnetic resonance (NMR) spectroscopy is a promising tool to produce fingerprint food in a non-targeted analysis. The generated NMR spectrum represents a unique signature for every food product, which is impossible to modify artificially. The quality of a food product can be, thus, assessed quickly and reliably, tracing back all those factors (presence of adulterants, agronomic practices, geographical origin) that may influence the metabolic profile of the product. The applications of the non-targeted NMR-based method reported in the PhD thesis demonstrate the potential of this approach in ensuring the traceability and assessing the authenticity of agri-food products. The high reliability of this approach would enhance the certification systems currently in use, making the latter even more robust and resistant to fraud attempts, also supported by blockchain-based information technologies.

DICATECh

2023

Marica Antonicelli

Food traceability and authenticity assessment by application of NMR spectroscopy with a non-targeted approach

03

D.R.S.S.

Doctor of Philosophy in Environmental and Building Risk and Development

Coordinator: Prof. Eng. Michele Mossa

XXXV CYCLE
CHIM/07- Chemical fundamentals of technology

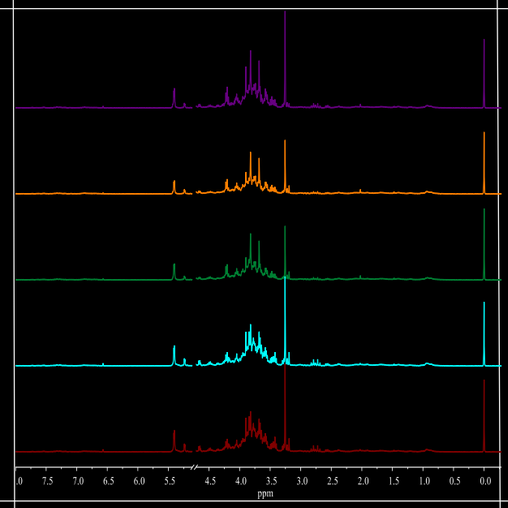
DICATECh
Department of Civil, Environmental, Building Engineering and Chemistry

POLITECNICO DI BARI

Marica Antonicelli

Food traceability and authenticity assessment by application of NMR spectroscopy with a non-targeted approach

Prof. Piero Mastrorilli
Prof. Mario Latronico
Dott.ssa Biagia Musio
DICATECh - Department of Civil, Environmental, Building Engineering and Chemistry



03

2023

Politecnico di Bari

LIBERATORIA PER L'ARCHIVIAZIONE DELLA TESI DI DOTTORATO

Al Magnifico Rettore
del Politecnico di Bari

La sottoscritta Marica Antonicelli nata a Mola di Bari il 04/06/1991 residente a Rutigliano in via Madonna delle Graie 12/E e-mail marica.antonicelli@poliba.it iscritta al 3° anno di Corso di Dottorato di Ricerca in Rischio e Sviluppo Ambientale, Territoriale ed Edilizio ciclo XXXV ed essendo stata ammessa a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

"Food traceability and authenticity assessment by application of NMR spectroscopy with a non-targeted approach"

DICHIARA

1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
2) di essere iscritto al Corso di Dottorato di ricerca in Rischio e Sviluppo Ambientale, Territoriale ed Edilizio ciclo XXXV, corso attivato ai sensi del "*Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari"*, emanato con D.R. n.286 del 01.07.2013;
3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archivierà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito http://www.creativecommons.it/Licenze), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviate/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dal sottoscritto tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

Luogo e data: Bari, 26/04/2023 Firma 

Il/La sottoscritto, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

Luogo e data: Bari, 26/04/2023 Firma 

D.R.S.S.	**POLITECNICO DI BARI**	03
Doctor of Philosophy in Environmental and Building Risk and Development		2023
Coordinator: Prof. Michele Mossa		
XXXV CYCLE CHIM/07-Chemical fundamentals of technology		
DICATECh Department of Civil, Environmental, Building Engineering and Chemistry		
	Marica Antonicelli	
	Food traceability and authenticity assessment by application of NMR spectroscopy with a non-targeted approach	
	Prof. Piero Mastrorilli Department of Civil, Environmental, Land, Building Engineering and Chemistrv Polytechnic of Bari	
	Prof. Mario Latronico Department of Civil, Environmental, Land, Building Engineering and Chemistry Polytechnic of Bari Dott.ssa Biagia Musio Department of Civil, Environmental, Land, Building Engineering and Chemistry Polytechnic of Bari	

D.R.S.S.

POLITECNICO DI BARI

03

2023

Dottorato di Ricera in Rischio e Sviluppo Ambientale, Territoriale ed Edilizio

Coordinatore: Prof. Michele Mossa

XXXV CICLO
CHIM/07- Fondamenti chimici delle tecnologie

DICATECh
Dipartimento di Ingegneria Civile, Ambientale, Edile e Chimica

Marica Antonicelli

Valutazione della tracciabilità e dell'autenticità degli alimenti mediante applicazione della spettroscopia NMR utilizzata in un approccio non-targeted

Prof. Piero Mastrorilli
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica
Politecnico di Bari

Prof. Mario Latronico
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica
Politecnico di Bari

Dott.ssa Biagia Musio
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica
Politecnico di Bari

EXTENDED ABSTRACT *(eng)*

In recent years, the interest of the research community and industry in investigation methods that ensure traceability and enable the valorisation of quality products has increased considerably. The development of such methods becomes essential to protect the commercial value of agri-food products and, at the same time, safeguard consumers' health. Under this scenario, there has been increasing use of the so-called non-targeted approach. The application of this approach to food matrices makes it possible to characterise their entire metabolic profile and obtain what is known as a fingerprint. This advantage has led the non-targeted method to become a viable strategy to address complex food quality control issues, including authenticity and traceability. Nuclear magnetic resonance (NMR) spectroscopy is a promising tool to produce fingerprint food in a non-targeted analysis. The generated NMR spectrum represents a unique signature for that food product, impossible to modify artificially, through which its quality can be assessed quickly and reliably, tracing back all those factors (presence of adulterants, agronomic practices, geographical origin) that may influence the metabolic profile of the product. The applications of the non-targeted NMR-based method reported in this thesis demonstrate the potential of this approach in ensuring the traceability and assessing the authenticity of agri-food products. The high reliability of this approach would enhance the certification systems currently in use, making the latter even more robust and resistant to fraud attempts, also supported by blockchain-based information technologies.

key words: quality control; traceability; food fraud; fingerprint; nuclear magnetic resonance

EXTENDED ABSTRACT *(ita)*

Negli ultimi anni, l'interesse del mondo della ricerca e dell'industria verso metodi d'indagine che garantiscano la tracciabilità e permettano la valorizzazione delle produzioni di qualità è aumentato in modo considerevole. Lo sviluppo di tali metodi diventa essenziale per proteggere il valore commerciale dei prodotti agroalimentari e, al tempo stesso, tutelare la salute dei consumatori. In questo scenario, si è assistito ad un crescente impiego del metodo definito *non-targeted*. L'applicazione di tale approccio alle matrici alimentari permette di caratterizzarne l'intero profilo metabolico ed ottenere quella che viene definita impronta digitale o *fingerprint*. Tale vantaggio ha portato il metodo non-targeted a diventare una valida strategia per far fronte alle complesse problematiche relative al controllo qualità degli alimenti, tra cui l'autenticità e la tracciabilità. La spettroscopia di risonanza magnetica nucleare (NMR) è un promettente strumento per ottenere il fingerprint degli alimenti in un'analisi non-targeted. Lo spettro NMR generato rappresenta una firma unica per quel prodotto, impossibile da modificare artificialmente, attraverso il quale è possibile valutarne la qualità in maniera rapida e affidabile, risalendo a tutti quei fattori (presenza di adulteranti, pratiche agronomiche, origine geografica) che possono influenzarne il profilo metabolico. Le applicazioni del metodo non-targeted basato sull'impiego dell'NMR riportate in questa tesi dimostrano la potenzialità di tale approccio nel garantire la tracciabilità e autenticità dei prodotti agroalimentari. L'elevata affidabilità di tale approccio potenzierebbe i sistemi di certificazione attualmente in uso, rendendo quest'ultimo ancora più robusto e resistente ai tentativi di frode, supportato anche con le tecnologie informatiche basate sulla blockchain.

key words: **controllo qualità; tracciabilità, frode alimentare, impronta digitale; risonanza magnetica nucleare**

INDEX

INTRODUCTION

1.0 Introduction

The traceability of agri-food products is a key aspect to provide their quality and value, along with the safety of the consumers. Traceability is defined as the set of all the procedures applied along the food supply chain, from crop to the end-user, which, by guaranteeing the authenticity of the product, protects not only consumers but also the producers themselves. Therefore, it has become one of the most important tools required for ensuring quality and food safety.

In the last years, the development and optimisation of accurate and reliable analytical methods able to assess the authenticity of agri-food products, have become fundamental research activities in the industry as well as in the scientific community involved in food control.

Besides classical document-based systems, the official analytical methods currently used to assess the authenticity of agri-food products are based on the identification and subsequent quantification of known chemical compounds present within the food matrix under examination. Such methods referred to as *targeted*, comprise a series of laboratory techniques that allow the isolation/identification of a specific compound or the corresponding chemical class, according to specific legal limits. Therefore, the application of targeted methods requires *a priori* knowledge of the food matrix to be analysed. However, this approach may be disadvantageous if it is not possible to know a priori the composition of the food matrix or, since it focuses on specific compounds, the targeted approach may not highlight unknown adulterations and/or additives. In such cases, especially if the adulterants are present below the legal limits detectable by the targeted approach, they would be missed during the analysis. Moreover, the increasing use of multiple and sophisticated adulterants has made difficult their detection with these established methods (Hong *et al.*, 2017).

The *non-targeted* (or *fingerprinting*) approach has been reported to be efficient in the field of quality control and authenticity assessment of food products (Gallo *et al.*, 2020). The non-targeted approach provides a comprehensive description of the complex matrix under investigation, allowing for the detection of known

and unknown compounds in the mixture. All these detected compounds comprise the fingerprint of the food matrix that may be exploited as a valuable tool for quality control and authenticity assessment (Ballin and Laursen, 2019). Indeed, by means of food fingerprint, it is possible to trace the agronomic practices, and the geographical origin, and to detect unknown contaminants present in the investigated product.

One of the main drawbacks of the non-targeted approach is the current lack of official guidelines governing the standardization and harmonization of the method development and validation (Riedl, *et al.*, 2015).

However, some first guidelines for the development and validation of non-targeted methods recently released by United States Pharmacopeia (USP) have been a go-to resource in the absence of other harmonized guidelines or standards. Specifically, such a document refers to non-targeted classification methods for detecting economically motivated adulteration (EMA) in food, generally known as food fraud, regardless of the analytical techniques used.

Non-targeted analysis of food matrices can be efficiently performed using Nuclear magnetic resonance spectroscopy (NMR), which can provide the fingerprint of the examined matrix by analysing relatively small amounts of sample in a non-destructive and rapid way.

This analytical technique allows for the simultaneous identification and quantification of the metabolites contained in the matrix under investigation, without the necessity of *a priori* knowledge of its composition and the need for elaborate sample preparation (Markley *et al.*, 2017).

In the past, the sensitivity of NMR was considered a main limitation, but the continuous progress of hardware resulted in the high sensitivity of this technique. Therefore, NMR spectroscopy enables a collection of comprehensive metabolic profiles that can be reliably used to achieve information on various aspects of food analysis (Godelmann *et al.*, 2013).

The large potentiality of the NMR-based non-targeted approach, for the quality and authentication assessment of foodstuffs, has been widely reported in the

latest literature (Spyros, 2016; Cagliani, *et al.*, 2017; Hatzakis, 2019). Such an approach has been applied to food products to identify the crop year (Consonni *et al.*, 2010), the agricultural practices (Gallo *et al.*, 2014), the geographical origin (Bachmann *et al.*, 2018), and to detect food adulterations (Musio *et al.*, 2022).

Moreover, the non-targeted NMR method is gaining increasing interest in this field due to its capability to generate highly reliable instrumental responses. It was demonstrated that NMR spectroscopy can provide statistically equivalent signals when the same sample is analysed by spectrometers that are different in terms of magnetic field strength, manufacturer, hardware configurations and age (Gallo *et al.*, 2015; Gallo *et al.*, 2020; Ragone *et al.*, 2020).

Despite the high reliability and reproducibility of the NMR non-targeted method, its applicability is still hampered due to the lack of harmonized and standardised agreed procedures, as mentioned above.

Thus, my PhD thesis project aims to develop and validate a non-targeted NMR testing method applied for traceability and authenticity assessment of different food products, namely saffron, lentils and durum wheat.

Specifically, a NMR-based non-targeted method was applied to establish the authenticity of saffron samples, with the purpose of differentiating pure saffron samples from those adulterated using turmeric and safflower at three different concentration levels. i.e., 5, 10 and 20 wt%. Furthermore, the performed analysis also revealed the agronomic practice adopted during the production of this precious spice, distinguishing between organic and conventional cultivation (Musio *et al.*, 2022).

Furthermore, the same analytical approach was applied to Italian lentils and Italian durum wheat, with the aim of being able to assess the authenticity and ensure geographical origin in their supply chain, by exploring the related NMR fingerprint generated.

CHAPTER I

2.1 Food quality and safety

Food quality and safety are two important terms describing aspects of food products and they have become topics of considerable interest. The consumer of the 21st century is a highly demanding one, exhibiting greater concern about the quality and health benefits of the products to buy (Sajdakowska *et al.*, 2018). All the steps in the food production chain have a potentially significant influence on the quality of the final product, that the consumer places on the table (Stewart *et al.*, 2018). There is an abundance of ways to define the term quality, both in food and otherwise (Grunert, 2005).

The concept of food quality should be considered on a much broader basis. Indeed, changes in consumer expectations, legislative needs, and instrumental analysis developments are included into an ideal concept of food quality (Müller and Steinhart, 2007). Thus, quality can be described as the requirements necessary to satisfy the needs and expectations of the consumer.

Several aspects contribute to defining the quality of a food product. These include intrinsic qualities such as taste and other organoleptic properties and external factors such as origin and labelling (Grunert, 2002; Jover *et al.*, 2004). Along with an objective definition, there is a strongly subjective component in the concept of quality, which is the quality as perceived by consumers, and it is influenced by the various characteristics of the product (Sadílek, 2018).

Food quality could not be separate from the concept of food safety, at least to the extent that consumers believe food safety to be a desirable property (Anklam and Battaglia, 2001). Safety is an important component of quality since a lack of safety can result in serious health issues for the consumer. A product can appear to be of high quality (i.e. well coloured, appetizing and flavourful, etc.), but it can be unsafe because it might be contaminated with undetected pathogenic organisms, toxic chemicals, or physical hazards.

Food safety can be defined as the opposite of food risk. This concept means the assurance that the food to be consumed is non-toxic, harmless, compliant with reasonable nutritional requirements, and will not cause any potential hazards or

diseases to human health (Liu, *et al.*, 2020). Food safety hazards may occur at a variety of points in the food chain. Therefore, the mentioned requirements should be met in all the activities included in the food supply chain, and shared by producers, processors, distributors, retailers, and consumers. (Liu, *et al.*, 2020).

With the expansion of international trade, food safety is becoming a global health issue, and the requirement for food quality increases constantly (Liu, *et al.*, 2020). Indeed, the globalization of food production and distribution has intensified the risks of food pollution caused by environmental and anthropogenic sources including microbial organisms, organic and inorganic chemical pollutants, drugs, and physical hazards (Akrami-Mohajeri *et al.*, 2018; Derakhshan *et al.*, 2018).

Both quality and safety aspects are interrelated and linked to the trust/confidence of consumers. These two are seen to be the consequence of taste, control, and origin, while resulting in health and a feeling of calm (Rijswijk *et al.*, 2008).

Today food control systems include procedures that consider the chemical composition of foodstuffs, hygienic aspects, nutritional adequacy, and authenticity. Nevertheless, the methods used in routine analysis focus mainly on known chemicals contained in the food matrices.

Therefore, endless efforts are required to improve the efficiency of analytical methods for ensuring public safety and quality control. In this manner, more sensitive, robust, efficient, and cost-effective methods should be developed to ensure the safety, quality, and traceability of foods without compromising nutritional, functional, and sensory characteristics according to legislation and consumer demands (Li *et al.*, 2021).

2.1.1 Traceability in the food supply chain

Food traceability is a relevant topic in food analysis and is closely linked to food quality, food safety, and human health (Kaufmann, 2014).

The several food scandals encountered in the last two decades, such as Bovine Spongiform Encephalopathy (BSE) or Mad Cow Disease, Dioxin in chicken feed,

Food-and-Mouth Disease (FMD) and incidents such as the use of Genetically Modified (GM) crops in foods (Aung and Chang, 2014) have drawn the attention of consumers to the importance of food's traceability (Ingrassia *et al.*, 2017). Nowadays, it is ascertained that consumers consider traceability as an important further attribute to evaluate the quality of food products before purchase. As reported in the previous paragraph, to define food quality and safety, consumers evaluate both intrinsic and external features of the products, such as traceability, origin geographical indications and certification (Jover, *et al.*, 2004).

As a result of its importance, traceability can be considered an effective safety and quality-monitoring system within food chains, as well as to increase consumer confidence and connect them to the producers (Kher *et al.*, 2010). Therefore, the design and implementation of traceability have become necessary tools to ensure food quality and safety in the whole supply chain.

The International Standard EN ISO 22005:2007 defines traceability as "the ability to follow the movement of a food through specified stage(s) of production, processing, and distribution" (ISO 22005:2007). Therefore, traceability makes it possible to capture, store and transmit all the information about a foodstuff at all stages in its supply chain, from the origin to the destination of the product. In this way, it can be traced at any time and checked for safety and quality control. A conceptual framework of traceability in the whole food supply chain is reported in fig. 1. The safety and quality regulations require all the participants to apply safety and quality assurance systems that comply with regulations and to manage all their operations in an efficient and standard manner.

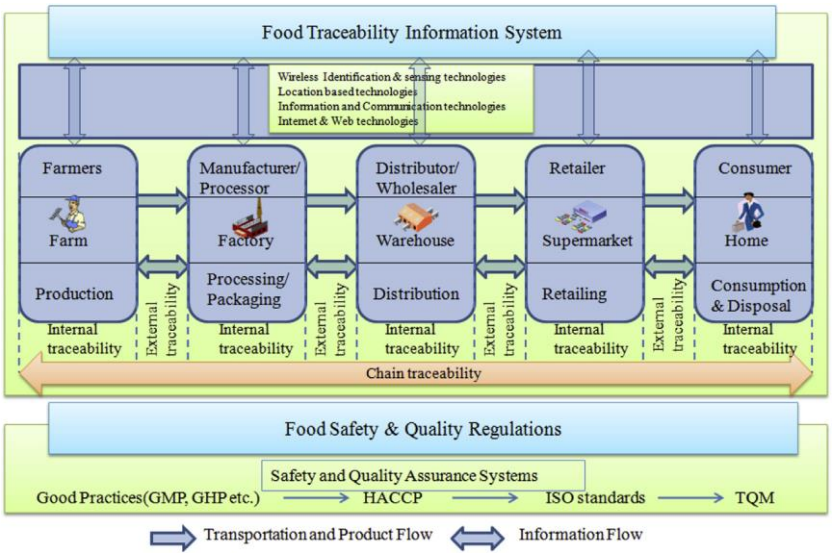


Fig.1- Conceptual framework of food traceability system (Aung and Chang, 2014)

In the growing global market, food supply chains are becoming longer and more complex. As a result, traceability tools become necessary to offer transparency to consumers and food manufacturers. Currently, to avoid food quality and safety issues resulting from asymmetric information conveyed between suppliers and consumers, and enhance customer confidence, traceability requirements in the food chain rely on regulations/standards or certifications (Aung and Chang, 2014).

For this purpose, the European Union has established a harmonised set of regulations and associated certifications for protecting food products and easily communicating to consumers correct information about the specific character of the products, the agricultural practices, and the place of origin. The adopted certification, namely PDO (Protected Designation of Origin), PGI (Protected Geographical Identification) and TSG (Traditional Specialty Guaranteed), protect product names from misuse and imitation and give a higher quality value to foodstuffs (Council Regulation, EEC N° 2081/92).

PDO identifies food products that are produced, processed, and prepared in a defined geographical area. Hence, “designation of origin” means the name of a

region, a specific place or a country used to describe a foodstuff originating in that specific place and whit exclusively quality due to the characteristic geographical environment. PGI identifies agricultural products and foods linked to a defined geographical area. Nevertheless, differently from PDO which describes a food product produced, processed, and prepared in a given geographical area, in the "geographical indication" case, the geographical link must occur in at least one of the stages of production, processing, or preparation. TSG (Traditional Specialties Guaranteed) identify and protects traditional methods of production. Thus, a TSG does not refer to the origin but highlights traditional character, either in the composition or means of production. (Danezis *et al.*, 2016). Examples of typical Italian food products which have gained these trademarks are Prosciutto di Parma (PDO), Lenticchie di Altamura (PGI), Cappellacci di zucca ferraresi (PGI), Culurgionis d'Ogliastra (PGI), Pasta di Gragnano (PGI), Pizzoccheri della Valtellina (PGI), Maccheroncini di Campofilone (PGI), and Mozzarella cheese (TSG).

Furthermore, optional quality terms such as OQT, "mountain product" and "product of island farming" were defined (1151/2012 EU Regulation). The information includes a characterization of the geographic region and reinforces the consumer perception of special quality attributed to mountain and island products.

All these specific regulations have implemented food safety and quality monitoring systems, protecting each country's culture, history, and the local economy at the same time. The management of food supply chains via these certifications represent a current methodology to achieve traceability requirements.

Another methodology records logistics operations and production processes via the design of a food traceability system that provides transparent trace back and tracks forward information (Hong *et al.*, 2011).

The main objectives of a traceability system are to give information on food product origin, processing, and retailing, safeguard food in transit, and decrease the associated time and cost of food recalls. It represents a useful tool to assure safe

and good quality food. A more-in depth presentation of this traceability system is reported in paragraph 2.1.1.1.

2.1.1.1 ISO 22005:2007

ISO is the world's largest developer and publisher of international standards. ISO standards are used to achieve uniformity and prevent technical barriers to trade worldwide. The requirements for food safety and traceability are added in ISO 22005 series with a major focus on traceability in the 2007 version.

ISO 22005 (2007) "*Traceability in the feed and food chain — General principles and basic requirements for system design and implementation*" defined the principles and objectives of traceability and also specifies basic requirements for the design and implementation of a feed and food traceability system. It can be applied by an organisation operating at any step on the feed and food chain. The principles given in this ISO are intended to be flexible enough to allow the feed and food organisations to achieve identified objectives.

The work of preparing International Standards is normally carried out through ISO technical committees. Each member interested in a subject for which a technical committee has been established has the right to be represented on that committee. ISO 22005 was prepared by the Technical Committee ISO/TC 34 *Food products*. The proposed traceability system is a technical tool to assist an organisation operating within a feed and food chain to conform with its defined objectives and it is useful when necessary to determine the history, or location of a product or its relevant components.

The choice of a traceability system is influenced by regulations, product characteristics and customer expectations. Hence, the design of the traceability system by an organisation can vary depending on:

- features of the product (i.e., nature of the raw materials, size of lots);
- technical limits inherent to the organisation (i.e., collection and transport procedure, processing and packaging method);
- objective to be achieved;

- cost and benefits of applying such a traceability system.

The document ISO 22005 gives the guidelines for the design and implementation of a traceability system (normative references, terms and definitions, principles, and objectives of traceability) which are indispensable for its following application. In this manner, the organisation operating within a feed and food chain should be able to improve its effectiveness and productivity by conforming to its management system.

As reported in the document, a traceability system is the totality of data and operations that can maintain desired information about a product and its components through all or part of its production and utilization chain.

The mentioned traceability system should be able to record the history of the product and/or locate it in the feed and food chain. It can establish the nonconformity and if necessary, allow for the withdrawal and/or recall of products.

The principles identified in ISO 22005 state that the traceability system should be:

- verifiable,
- applied consistently and equitably,
- results-oriented,
- cost-effective,
- practical to apply,
- compliant with any applicable regulations or policy, and with defined accuracy requirements.

These principles should take into consideration in the design of a feed and food chain traceability system. In addition, in its development, the following shall be included:

a) objectives;

b) regulatory and policy requirements relevant to traceability;

c) products and/or ingredients;

d) position in the feed and food chain;

e) flow of materials;

f) information requirements;
g) procedures;
h) documentation;
i) feed and food chain coordination.
Each organisation shall establish a traceability plan that includes all the identified requirements; moreover, the organisation shall review and implement the traceability system at specific intervals or whenever changes are made to the established objectives and/or the product or processes (ISO 22005:2007).

2.1.2 Authenticity of food

Food authentication is a rapidly growing field due to increasing public awareness concerning food quality and safety (Danezis *et al.*, 2016). The main objective of food quality is to guarantee the appropriate organoleptic and appreciated characteristics of the food product, whereas food authenticity covers many several aspects. These latter ones are related to the mislabelling process, adulterations of precious foodstuffs, and misleading claims about the origin (species, geographical or genetic), production method (conventional, organic, traditional procedures, free range), or processing technologies (irradiation, freezing, microwave heating).
Food authenticity is a topic of great interest to the food industry, consumers, and scientific researchers. Moreover, this topic is attracting a high level of attention from the authorities and media worldwide, due to the production of counterfeit food, illegal food trade, and frequent fraudulent labeling attempts targeted at valuable products by unscrupulous suppliers (Dimitrakopoulou and Vantarakis, 2021). Authentic food is defined as food that conforms to the description provided by the producer or processor and includes the process history of a product or ingredient, its geographical region of origin, or the species or variety of ingredients. The standardised procedures described in section 2.1.1 establish norms in the production process and control of origin of foodstuffs. The production of consumer goods according to these standardized procedures normally results in

better products and is rewarded with higher prices at the point of sale. Unfortunately, these financial benefits attract the production of counterfeit food and illegal food trades.

In the last years, many counterfeiting and adulterating practices in food trace have occurred, as evidenced, by the use of the banned dyes Sudan I and Para Red in Worcester Sauce in the United Kingdom (Di Anibal *et al.*, 2009; Hu *et al.*, 2017), the addition of melamine to milk in China in 2008 (Pei *et al.*, 2011), the fraudulent use of slaughterhouse waste in meat products in Germany, the detection of dioxin in meat and eggs, and the so-called horse meat scandal that scattered Europe in 2013. Furthermore, recently, the broad-spectrum insecticide fipronil, whose use is not permitted in livestock, has been found in eggs (Schieber, 2018).

The term food fraud represents a collective one used to encompass the deliberate and intentional substitution, addition, tampering, or misrepresentation of food, food ingredients, or food packaging; or false or misleading statements made about a product for economic gain (Spink and Moyer, 2011). Hence, there is consensus that food fraud is an intentional act carried out for economic gain, whereas a food defense incident is an intentional contamination with intentional harm, in contrast to a food safety incident which is an unintentional act with unintentional harm.

Within the category of food fraud committed for economic purposes, the following can be distinguished:

- adulteration, which occurs when there is a change in the natural composition of the food with the use of lower-quality substances in the product;
- sophistication is committed with the addition of elements which are not useful to the composition of the food, but which improve its external appearance, covering up its defects;
- alteration occurs when there is a transformation of chemical components that change the physical and chemical properties of the food itself;
- falsification is a substitution of one food product for another;

- counterfeiting consists in substituting one element of that food with another of lesser quality than the first, along with misleading the purchaser as to the quality and purity of the product.

Recently, the widely used term for food counterfeiting is economically motivated adulteration (EMA). EMA has been defined as the fraudulent, intentional substitution or addition of a substance in a product for the purpose of increasing the apparent value of the product or reducing the cost of its production. Among these fraudulent procedures, the usual one is the substitution of one ingredient for a similar cheaper one which is difficult to recognise by the consumer and to detect by current analytical techniques (Cubero-Leon, *et al.*, 2014).

Although different fraudulent practices motivated by economic profit may exist, these could also pose a direct threat to public health. Apart from the adverse public health impact, food fraud negatively affects the food industry business. It is estimated that the cost of food fraud in the global food industry is approximately EUR 30 billion every year (FAO, 2021). Detecting food fraud is a challenge because consumers alone cannot detect them, and food fraudsters are usually innovative in the ways they avoid detection. Therefore, the scientific community is working hard to develop new technologies and new reliable analytical approaches for quality control and authenticity assessment in the food supply chain.

Currently, food authentication methods have traditionally concerned the measurement of single markers or a small set of markers with comparison to set thresholds. This type of approach, known as a *targeted* method, is thereby based on the detection and, eventually, quantification of compounds known *a priori*. However, this approach may fail when unknown substances are indicative of either accidental or intentional alteration of the food product. In this contest, the *non-targeted* analytical method is becoming a more valuable tool for quality control and authenticity assessment (Ballin and Laursen, 2019). Indeed, the non-targeted method offers the possibility to extract a larger amount of information

which can be advantageously used to unveil the compounds that may affect the authenticity of the food sample under investigation.

CHAPTER II

2.2 Non-targeted method for food authenticity assessment

To date, traditional approaches applied to food authentication focus on one or more pre-defined analyte(s). These *targeted* methods detect and quantify only one compound at a time or a set of specific markers, which are usually indicative of a peculiar property of the examined food product. Such analytical targets are chemically or biologically characterised and annotated with established importance before data acquisition. The results of the analysis are compared to legal limits to determine if the compound tested is in excess (McGrath *et al.*, 2018).

The performance of the targeted method to detect specific compounds and certain residues in a food product is covered by EU legislation (Commission Decision 2002/657/EC).

Nevertheless, as the targeted approach is based on the detection and quantification of known compounds in a complex matrix, it often provides insufficient information to reliably verify the authenticity of a food product. Indeed, food fraudsters know about such testing programmes and how to avoid fraud detection (McGrath *et al.*, 2018).

The increasing attention to highly complex authentication issues, including geographical origin, agricultural production methods, and food adulterations, has driven the necessity for the development of novel analytical methods (Ballin and Laursen, 2019).

In recent years, the non-targeted method has emerged as a powerful approach to tackling problems of food authenticity, food quality and safety. Using analytical instrumentation techniques, non-targeted methods have been developed and applied in many food and agricultural situations. This development is evident from the increasing number of scientific papers based on non-targeted methods for food authentications during the past decade (Ballin and Laursen, 2019).

The non-targeted methods are not focused on specific target compounds and try to give a picture as complete as possible of the chemical composition of the investigated sample (Sobolev *et al.*, 2015). Indeed, the non-targeted analysis is also

referred to as “fingerprinting”, due to the advantage of simultaneously displaying numerous unspecified targets detected in the sample.

Although these reliable methods are described in a growing body of literature, their uptake in terms of routine surveillance has so far been limited (Riedl, *et al.*, 2015). The application of non-targeted analytical tools for food surveillance to ensure consumer protection is needed because as laboratories improve targeted detection methods specific to particular markers, fraudulent producers might use unknown adulterants that can evade detection with the targeted approach (Karunathilaka *et al.*, 2018).

However, how the non-targeted method should be applied and validated is still confusing and lacks standard guidelines, since there is no legislation equivalent to the aforementioned targeted approach. This lack has resulted since the non-targeted approach has been developed by scientists working in many disciplines, and each seems to have developed its own unique set of definitions and terminologies. This has led to inconsistency and confusion in the way scientists talk about non-targeted methods and hinders a broader understanding of the techniques (Gao *et al.*, 2019). In 2017 the “Guidance on developing and validating non-targeted methods for adulteration detection” has been released by the United States Pharmacopeia (USP).

This guidance document is intended to standardize some of the key processes related to developing and validating non-targeted methods of analysis. Moreover, it would assist users in supply chain management by providing information that can generally be applied to the testing and authentication of raw materials with a variety of analytical techniques.

These guidelines identify and assess the following criteria for non-targeted fraud detection in food commodities:

- the objective of methodology;
- adulterant investigated;
- sample related criteria;
- instrumentation methodology;

- chemometric methodology;
-model performance criteria.

It is noteworthy that the prescribed procedure in the USP guideline is independent of either the principle of the analytical method used or the food type (Nichani *et al.*, 2023).

Among the possible analytical techniques available, NMR is regarded as a versatile and useful tool to perform a non-targeted detection of food authentications (Consonni and Cagliani, 2010; Fan and Zhang, 2019).

Using NMR, it is possible to rapidly and simultaneously detect a great number of compounds comprising the whole composition (fingerprint) of the sample while giving qualitative and reproducible quantitative results with minimal sample pretreatment (Naz *et al.*, 2014). Thus, the abundant chemical information provided by the NMR-based non-targeted approach may be used to reliably assess the authenticity of foodstuff, including the identification of geographical origin, and the detection of unknown adulterants.

The setting up of a non-targeted NMR method involves several stages:

(i) the collection of authentic reference samples to build up a database in which unknown samples can later be included.

(ii) The sample preparation should be limited but may include milling, homogenization, and extraction procedure.

While targeted analysis often requires a very complex and specific extraction process, the sample extraction for non-targeted analysis is usually kept as simple as possible (Creydt and Fischer, 2020). One of the advantages of NMR is the simplicity of the sample extraction by using deuterated solvents.

(iii) Acquisition of spectral data by performing NMR experiment of authentic samples and the following collection of NMR data in a suitable database.

(iv) Training and validation of classification algorithms.

One of the fundamental strategies of non-targeted detection is to solve a classification problem, by using both multivariate and univariate modelling methods.

(v) NMR experiment of unknown samples (test) according to the developed method to ensure the correct classification ability of the model.

The applications of the non-targeted NMR method reported in this thesis (chapters IV, V and V) may represent suitable studies within the validation and harmonisation framework of this new method.

CHAPTER III

2.3 Introduction to Nuclear Magnetic Resonance (NMR)

During the last two decades, Nuclear Magnetic Resonance (NMR) spectroscopy has been widely recognized as an important tool for food analysis (Sobolev *et al.*, 2019). This technique is rapid and rich in structural and quantitative information and allows the metabolites to be analysed simultaneously (Wang *et al.*, 2003).

NMR spectra of food products can act as "fingerprints" that can be used to compare, discriminate, or classify samples. In this way, the NMR fingerprint allows evaluating, in one shot, the entire metabolites' pattern variation due to external factors such as agronomical practices, geographical origin, environmental conditions and the presence of contaminants.

NMR offers a unique view of the entire set of most abundant, free small molecules in a biological sample that can be detected and measured in a single spectrum with minimal sample preparation. Indeed, the NMR sample does not require any physical or chemical treatment before the analysis, but only the solution conditions such as the temperature, pH and salt concentration must be adjusted. Since, in many cases, nearly no sample pre-treatment is required in NMR spectroscopy, the inherent properties of the sample are well-kept. In general, NMR can offer several unparalleled advantages, including selective isotope detection in complex mixtures, highly reproducible data, de novo determination of crucial structural parameters of unknown metabolites, accurate quantification without standards, and in situ analysis of pathway dynamics from cells to whole organisms without

any destructive effects (Markley *et al.*, 2017).

The NMR data collected are studied using multivariate statistical analysis based on mathematical and statistical procedures. Thus, multivariate analysis becomes an excellent tool for exploring all data contained in the NMR spectra without a deep knowledge of the metabolites that compose the sample.

2.3.1 NMR basic principles

Nuclear Magnetic Resonance Spectroscopy is one of the most valuable tools to observe and understand phenomena of different natures, in the field of chemistry, physics, biology and medicine. It is a powerful analytical technique, which gives detailed information concerning the molecular structure of the biological material observed, reflecting at the end the metabolic status of a biological living system, without losing important information on the system. This spectroscopic technique is generally used to detect hydrogen atoms in metabolites (^{1}H NMR). Thus, in a typical sample (biological fluid in medical research or organic extract in food research) all hydrogen-containing molecules (almost all metabolites) will give a ^{1}H-NMR spectrum, as long as they are present in concentrations above the detection limit.

NMR spectroscopy is based on interaction between an electromagnetic radiation and the atomic nucleus immersed in an external magnetic field. Based on the number of protons and neutrons that constitutes the internal structure of an atomic nucleus, a nucleus can have a spin state that is described through *I*, *I* is called the nuclear spin quantum number. Depending on the cases *I* can assume: zero value and in this case the nucleus is said to be NMR inactive and therefore cannot be observed by NMR spectroscopy; or semi-integer or integer values and in these cases the nucleus is defined as active and therefore observable NMR.

The nuclear spin quantum number *I* may assume values: -I; -I+1; ..... I+1; I.

So, for a nucleus with spin ½ only two orientations are possible: +1/2 and -1/2.

In normal condition these two states are energetically degenerate.

In the classical description, the atomic spin is regarded as a sphere in rotation around an axis, so every nucleus having a non-zero quantum number spin *I* has an intrinsic angular momentum P:

$$P = \frac{h}{2\pi}\sqrt{I(I+1)} \quad (1)$$

where *I* is the spin angular momentum and *h* is the Plank constant.

The value of nuclear spin I depends on the atomic number and atomic mass:
- if the atomic mass is odd, I is half of an integer number and the nucleus is NMR active (^{1}H, ^{13}C, ^{15}N, ^{31}P);
- if the atomic number is odd and the atomic mass is even, I is an integer and the nucleus is NMR active (^{14}N, ^{2}H);
- if the atomic number and atomic mass are even, I is zero and the nucleus is NMR silent (^{12}C, ^{16}O).

Elements can have more than one isotope with non-zero nuclear spin, each one with its natural abundance. In the following table, several active NMR nuclei are listed along with their natural abundance and *I*.

Table 1. Spin nuclear and abundance of some common NMR nucleus

Nucleus	**Percentage abundance (%)**	**I**
^{1}H	99.98	1/2
^{13}C	1.108	1/2
^{15}N	0.366	-1/2
^{19}F	100	1/2
^{31}P	100	1/2
^{29}Si	4.67	-1/2
^{195}Pt	33	1/2

Since a nucleus has a positive electric charge, and every electric charge in movement produces a magnetic field, then the nucleus will also have a magnetic moment *m*, i.e. the nucleus can be considered like a small magnet.
The magnetic moment has the same direction as the spin angular momentum and is directly proportional through the relationship:

$$\vec{\mu}=\gamma \vec{P} \quad (2)$$

The constant of direct proportionality between the angular momentum and the magnetic moment is typical of each nucleus and is indicated with γ and is called the gyromagnetic ratio.

Considering a population of nuclei with spin ½ in absence of an external magnetic field, the orientation of the magnetic moments of atomic nuclei is random and equally distributed between the two degenerate states +1/2 and -1/2.

In the presence of an external field B_0, conventionally oriented along the z-axis, the magnetic moments of spins interact with it and the population of nuclei distributes in two different orientations (Figure 2): slightly more than half will reside in the "+1/2" state (parallel, α state, lower energy) and slightly less than half will reside in the "-1/2" state (antiparallel, β state, higher energy).

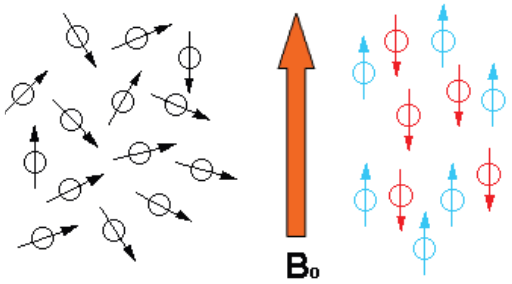


Fig. 2 The possible orientations of a nucleus with I=½

The spin nuclei are not perfectly aligned but they are only oriented along the direction of applied magnetic field, this causes in a circular motion called precession. The rate of this precession is directly proportional to external applied magnetic field and gyromagnetic ratio and is called Larmor Frequency:

$$v_0 = \frac{\gamma}{2\pi} B_0 \quad (3)$$

As for all spectroscopies, it is possible to promote a nucleus from the state of lower energy to the higher energy state by absorbing an amount of energy equal to:

$$\Delta E = h \times v_0 = \frac{h}{2\pi} \gamma B_0$$
(4)

with γ being the gyromagnetic ratio and h/2π is the quantum of angular momentum. The dependence of ΔE by applied magnetic field strength is shown in figure 3 for nuclei with I = 1/2.

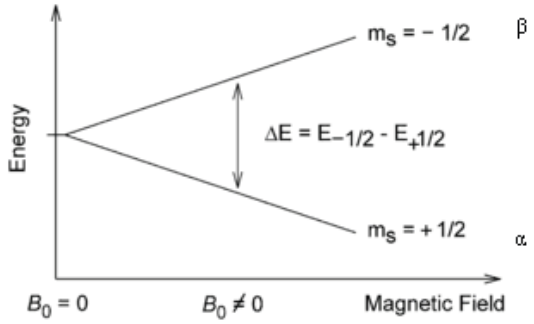


Fig. 3- The splitting of the spin states of the nucleus with I=½ upon application of a magnetic field

The population difference between the two spin states is given by the Boltzmann distribution function:

$$N_\beta / N_\alpha \propto e^{-\Delta E/kT}$$
(5)

where N_α and N_β represent the spin states populations, k is the Boltzmann constant and T the temperature. Since the energy difference between the two spin states is very small, the corresponding population differences are similarly small. Anyway, the slight excess of the nuclei in the α state is responsible for the net magnetization that is fundamental in NMR spectroscopy. According to the classic representation, a nucleus with spin 1/2 moves around the z-axis, coincident with the axis of the applied magnetic field on the surface of a double cone (Figure 4).

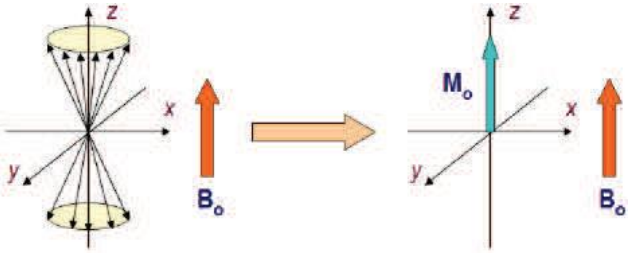


Fig. 4- Representation of the magnetization vector

The vector sum of all the magnetic moments of the nuclei contained in the sample gives the macroscopic magnetization M_0. This vector has a direction coincident with that of the magnetic field B_0. When electromagnetic radiation, generated by a sinusoidal current flow, is given to the system, the system is perturbed. The electromagnetic radiation can be considered as an applied magnetic field B_1 ($|B_1| << |B_0|$), orthogonal to B_0, because it originates from a coil wrapped in the plane xy (Sanders, 1988; Breitmaier, 2003). The interaction between the magnetization vector M_0 and the applied magnetic field B_1 occurs when the radiofrequency generating B_1 is equal to the Larmor frequency. In this way, the transition between the two energy levels α and β occurs through the absorption of the radiofrequency ν. For magnetic fields applied in common NMR spectrometers (from 1 to 20 Tesla), the frequency of precession of nuclei varies from tens to hundreds of MHz. Perturbation of equilibrium condition, necessary for experiment execution, consists of a short (few microseconds) and intense pulse of energy able to generate the desired radiofrequency. Such a pulse excites simultaneously all nuclei resonating within a continuous band of frequencies which are symmetrical concerning the frequency centre ν_1 (Keeler, 2010).

Pulse application has the effect of the abovementioned applied magnetic field B_1. The interaction between B_1 and magnetization M_0 produces a torque that moves the magnetization vector in the xy plane with intensity M_{xy}. The pulse generating this shift from the z-axis to the xy plane is named 90° pulse (Figure 5).

After the perturbation, the system tends to restore the equilibrium condition where the energy levels are repopulated according to Boltzmann distribution.

Such a process is denoted as the relaxation of the nuclei. During relaxation, magnetization M_0 returns to its initial value aligned along the z-axis, while magnetization M_{xy} loses progressively its intensity to reach a net value equal to 0.

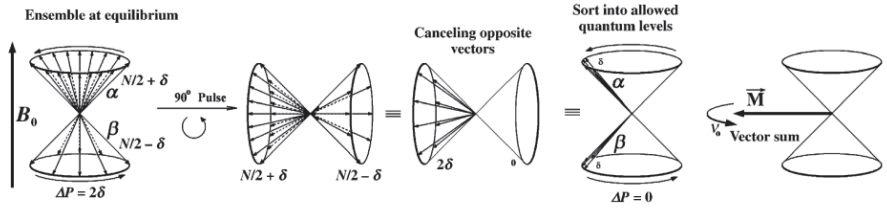


Fig. 5 - Representation of magnetization vector and a 90° pulse (Jacobsen, 2007)

2.3.2 Relaxation processes and NMR signal generation

The relaxation process allowing vector M_0 to return to its initial position along the z is due to two different relaxation sub-processes:

- longitudinal or spin-lattice relaxation,
- transverse or spin-spin relaxation.

The first sub-process considers the energy exchanges taking place between the spin system and the molecular environment and produces an increment of the magnetization along the z axis to equilibrium value M_0 (Figure 6). In this type of relaxation, in homogeneous systems, the magnetization recovery usually follows a trend fitted by the exponential and characterized by a constant T_1 denoted as longitudinal relaxation constant.

$$M_z = M_0 [1-e^{(-t/T1)}] \quad (6)$$

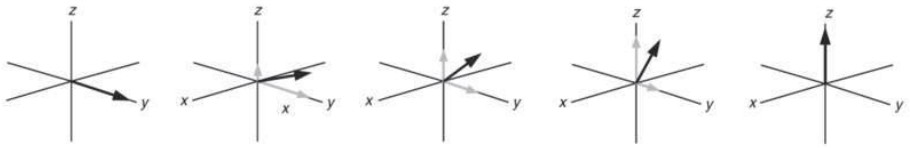


Fig. 6- Longitudinal or spin-lattice relaxation

Transverse relaxation consists of the loss of phase coherence among the resonances of the nuclear spins (Figure 7) with the concomitant loss of magnetization intensity in the xy plane resulting in the cancellation of the magnetization in the same xy plane. Such a magnetization decay usually follows a trend fitted by the exponential and characterized by a constant T_2 denoted as transverse relaxation constant:

$$M_{xy} = M_0 [1-e^{(-t/T2)}] \quad (7)$$

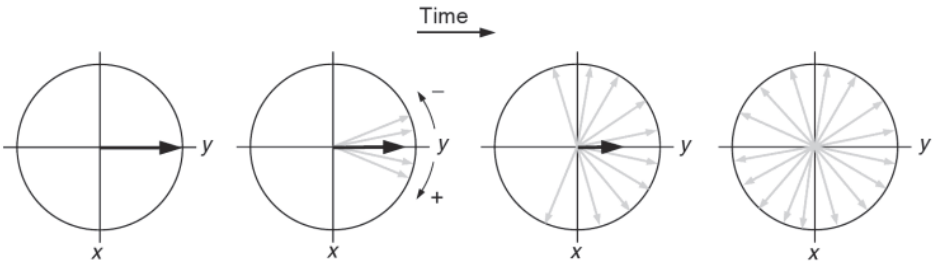


Fig. 7- Transverse or spin-spin relaxation

The intensity of M_{xy} is recorded by a detector and analysed to extract chemical information.

The detector consists of a receiver coil placed perpendicularly to the z axis. It receives a combination of frequencies forming output signal called FID (Free Induction Decay).

This is a signal in the time domain that is converted into a signal in the frequency domain by Fourier transformation (FT) (Figure 8). Commonly, the Fourier transformation of a single FID generates a low intensity signal with a low signal-to-noise ratio. To increase the intensity of the output signal, and then the signal-to-noise ratio (by a factor equal to $\sqrt{n}$), n measurements are executed. In this case, a delay (named recycling delay or relaxation delay) must separate the experiments to allow M_0 recovery before the successive application of the 90° pulse.

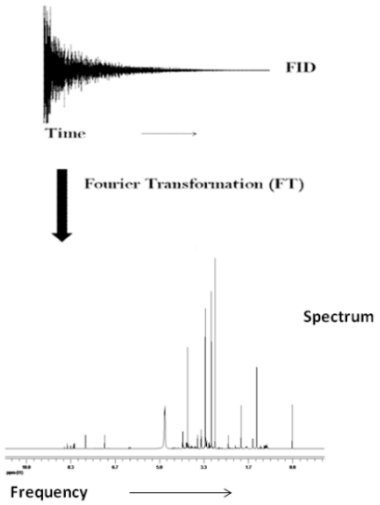


Fig. 8- Fourier transformation (Keeler, 2010)

When NMR signals are used for quantitative methods, acquisition, and processing parameters settings for the Fourier transform (choice of zero-filling and apodization function) are equal for all samples to be evaluated, since they may influence the finer details in the spectra. In real samples, nuclei are never isolated each other but are affected by local magnetic fields generated by the surrounding electrons. Such magnetic fields are less intense than the applied field but cannot be ignored. The interaction between each nucleus and these small magnetic fields causes little variations in the energy levels and then little differences in the NMR signals. The magnetic field at the nucleus (the effective field B_{eff}) is generally lower than the applied one by a fraction σ denoted as nuclear magnetic shielding according to the following equation:

$$B_{eff} = B_0 (1 - \sigma) \quad (8)$$

Consequently, the Larmor frequency ν_L, for a given nucleus must be rewritten as:

$$\nu_L = \frac{\gamma B\ eff}{2\pi} = \frac{\gamma B0}{2\pi}(1-\sigma) \quad (9)$$

The dependence of ν from the applied magnetic field B_0 prevents a direct comparison between signals generated by different spectrometers. To render the resonating frequency independent of B_0 and indicative of specific nuclei, the chemical shift δ was introduced as the difference between the resonance frequency of a given nucleus ν and that of a frequency ν_0 related to a standard compound:

$$\delta = (\nu - \nu_0)\ 10^6 / \nu_0 \quad (10)$$

For ^{1}H NMR spectroscopy, the most common reference compound is $Si(CH_3)_4$ (tetramethylsilane, TMS) which is soluble in many organic solvents, is inert and possesses twelve equivalent protons generating a single signal. Such a signal is conventionally fixed to 0.00 ppm. Similarly, the sodium 3- (trimethylsilyl) propionate-2,2,3,3-d_4 (TSP) is widely used in aqueous media.

In solution, when the local field of a given nucleus is affected by a nearby set of nuclei, a split of the energy levels (and then of the signals) occurs depending on the number of the surrounding nuclei and their nuclear spin I. This phenomenon is called scalar spin-spin coupling and represents the indirect magnetic interaction of nuclear spins with each other through the involvement of the electrons. Scalar coupling manifests itself by signal splitting and is characterized by a coupling constant J accounting for the specific chemical features of the molecule. The chemical shift and the coupling constant are important parameters linking NMR spectroscopy and chemistry. The chemical shift indicates the local electronic environment, and the J coupling provides a direct spectral manifestation of the chemical bonds.

The most important features of a NMR spectrum are:

- the chemical shift, which indicates the chemical group contained in the molecule;

- the multiplicity of the signal and the spin-spin coupling constant J related to the structure of the molecule;
- the intensity of the signal which is directly related to the number of nuclei in the sample.

To obtain a well-resolved NMR spectrum, the applied magnetic field B_0 must be long lasting stable in intensity and homogeneous within the sample. Concerning the field stability, even though spectrometers ensure very high performance, little field fluctuations are unavoidable, and this results in signal broadening. Such a problem is efficiently overcome by a lock procedure on deuterium contained in a reference compound. That is the reason for the use of deuterated solvents in the preparation of the NMR samples. The homogeneity of the magnetic field nearby of the sample is controlled and adjusted by shimming procedures consisting of the application of very weak additional fields.

2.3.3 Instrumentations

A typical NMR spectrometer consists of a magnet, a radio frequency generator, a receiver unit and a probe. Currently, magnets are based on superconducting materials cooled in liquid helium. Spectrometers work with pulsed techniques which can stimulate all the nuclei contained in the sample by a radio frequency pulse. The excited nuclei return to the lowest energy level and produce FID, the ensemble of signals recorded in the time domain by the receiver unit. FID is converted in frequency domain spectrum by Fourier transformation giving NMR signals easy to interpret. Figure 9 shows the functional scheme of a generic NMR spectrometer.

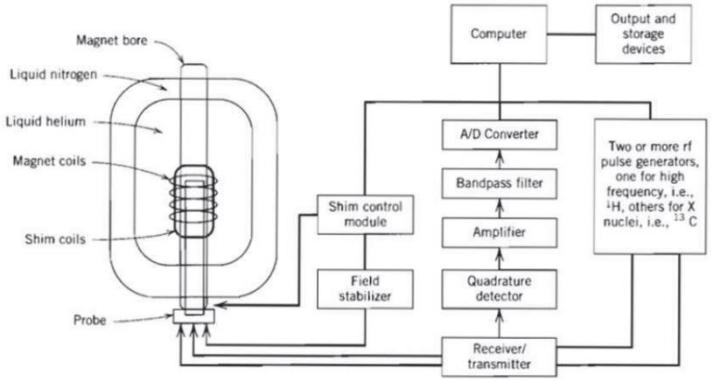


Fig. 9- Scheme of a generic NMR spectrometer

A very important component of a NMR spectrometer is the probe which represents the interface between a sample and the instrument allowing for excitation and detection of the nuclear spins. The probe is located into the centre of the magnetic field and is equipped with a temperature controller. It contains radiofrequency (R_f) coils tuned at specific frequencies. Often, probes are constructed with two coils, the inner coil closest to the sample, and the outer coil more distant from the sample. The nuclei excited by the inner coil are detected with the highest sensitivity. When the inner-coil is tuneable in a wide range of frequencies (broadband coil) the probe is denoted as a "broadband observe" probe and allows for excitation/detection of many nuclei having their Larmor frequencies in a well-defined range (i.e. frequencies between ^{31}P and ^{15}N). When the inner-coil is tuned to ^{1}H Larmor frequency, the probe is denoted as a "broadband inverse" or "indirect detection" probe and gives the highest proton sensitivity. Modern NMR probes also include an actively-shielded Pulsed Field Gradient (PFG) coil which allows for the execution of powerful and rapid experiments by application of field-gradient pulses.

Fig. 10- Bruker Avance 400 spectrometer used for the experiments reported in this thesis

CHAPTER IV

2.4 Multivariate Statistical Analysis

NMR spectroscopy enables to detection of all metabolites in a sample. The resulting NMR spectrum is full of information (variables), part of which most of the time results to be redundant.

For this reason, it is important to compress these variables to obtain only those that contain useful information. This goal is usually achieved through a bucketing procedure, where the NMR spectrum is subdivided into several bins, usually 0.02-0.04 ppm width each, which are then integrated to obtain the total intensity of each bin. This kind of procedure reduces the total number of variables to be used in the subsequent statistical analyses. Multivariate statistical analysis can be performed by unsupervised or supervised techniques. The unsupervised methods seek discriminating factors between the independent variables to obtain a graphical representation as the result of the maximization of variance. On the contrary, supervised methods find the best-fitting relationship between independent and dependent variables and are often applied to obtain sample classification among groups of interest and to build predictive models (Takis *et al.*, 2019).

2.4.1 Principal Component Analysis

Among the multivariate analysis techniques, Principal component analysis (PCA) is the most frequently used because it represents the starting point of data mining (Leardi, 2002; Siebert, 2001). Such an unsupervised method aims at data exploration, by reducing large complex data set into a series of optimized and interpretable views. These views emphasize the natural groupings in the data and show which variables greatly influence those patterns. Therefore, the goals of PCA are to

1) extract the most important information from the data matrix;
2) compress the size of the data set by keeping only this important information;
3) simplify the description of the data set;

4) analyse the structure of the observations and the variables (Abdi and Williams, 2010).

To achieve these goals, PCA is designed to extract and display the systematic variation on a data matrix X. Hence, the starting point for PCA is a matrix X of data with *n* rows (observations) and *K* columns (variables). The observations include analytical samples, chemical compounds or reactions, biological individuals, etc. The variables (columns) might be of spectral origin (NMR, NIR, IR…), chromatographic origin or other kinds of measurements.

The variables describing the data are transformed into new variables, called principal components (PCs), which are linear combinations of the original variables (Bro and Smilde, 2014).

The PCs have two interesting properties:

- they are extracted in decreasing order of importance. The first PC always contains more information than the second, the second more than the third and so on.
- they are orthogonal to each other. There is no correlation between the information contained in different PCs.

Usually, 2 to 5 PCs are sufficient to approximate a data table well.

Before performing PCA, data are typically pre-processed to make them into a form suitable for the analysis. The typical pre-treatment of data includes *scaling* and *mean-centering*. Then, PCA works by decomposing the X-matrix into the product of two matrices, scores (T) and loadings (P) matrix:

$$X=1*x' +T*P'+ E \quad (11)$$

In the expression above, the first term, 1*x', represents the variable averages and originates from the pre-processing step. The second term, the matrix product T*P', models the structure of the PCA, and the third term, the residual matrix, E, contains the noise (Krzanowski and Jackson, 1992).

The scores matrix (T) contains information about the observations, which are described in terms of projections onto the PCs; the loading matrix (P') contains information about the variables. The information not contained in these matrices remains as "unexplained X-variance" in a residual matrix (E) which has the same dimensionality as the original X-matrix (Figure 11).

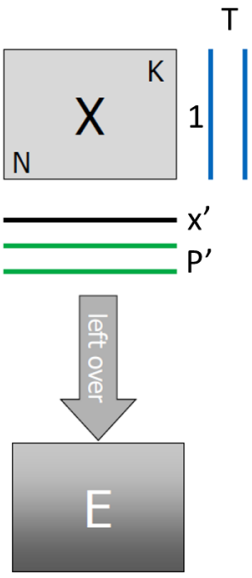


Fig. 11- A matrix representation of how a data table X is modelled by PCA.

PCA provides a graphical representation of the observations, using the *scores plot*, and the variables in the *loadings plot*. Scores are linear combinations of original variables. The scores plot is useful to analyse objects' behaviour, highlighting their similarities based on the considered components. In this way, it is possible to identify groups of similar objects (clusters), ascertain the presence of outliers, to evaluate the regularity and distributions of the observations.

The meaning of the scores is given by the loadings. The loadings plot allows us to analyse the role of the variables in the different components, their importance and their direct or inverse correlations.

Variables that appear close together in the loadings carry information common or similar (i.e. they are directly related). When variables are inversely correlated, they are positioned on opposite sides of the plot origin.

PCA also forms the basis for another exploring technique, namely hierarchical cluster analysis (HCA). With the hierarchical approach, score variables derived from multiple PC models at the lower modelling level, are concatenated to form new upper modeling level data matrices that are again analysed by PCA. The upper level provides a superficial overview of the relationships among the different blocks of variables. The lower level allows zooming into the most influential variables in the important variable blocks.

2.4.2 Partial least squares discriminant analysis

In contrast to PCA, Partial least squares discriminant analysis (PLS-DA) and Orthogonal partial least squares discriminant analysis (OPLS-DA) are supervised techniques. This kind of analysis enhances the separation between groups of observations by rotating PCA components so that a maximum separation among classes is obtained (Chevallier *et al.*, 2006).

One of the discriminant techniques most widely applied is the so-called partial least squares discriminant analysis (PLS-DA) (Oliveri and Downey, 2013).

PLS-DA is used to optimise separation between different groups of samples, which is accomplished by linking two data matrices *X* (i.e., raw data) and *Y* (i.e., groups, class membership etc.) (Gromski et al., 2015).

The method applies multivariate regression by partial least squares (PLS) using binary class-membership indices (0 and 1) for each class as response variables (Oliveri and Downey, 2013).

PLS is a method for relating two data matrices, X and Y, to each other by a linear multivariate model (Wold, *et al.*, 2001). When fitting PLS, the model finds the linear relationship between the matrix Y (dependent variables) and the matrix X (predictor variables) expressed as:

$$Y=f(X)+E \quad (12)$$

In PLS, also the Y matrix is decomposed into principal components and the principal components of X are rotated in the rough direction correlation concerning the principal component of Y.

PLS modelling consists of the simultaneous projection of both X and Y spaces on low dimensional hyper planes. The coordinates of the points on these hyper planes constitute the elements of the matrices T (matrix of scores that summarizes the X variables) and U (matrix of scores that summarizes the Y variables).

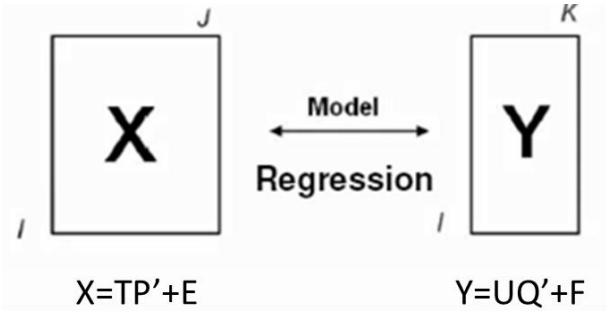


Fig. 12- The matrix relationships in PLS

In the PLS algorithm, there are additional loadings, W, called weights. These express the correlation between U and X and are used to calculate T.

PLS was contrived to model continuous responses, but it can be applied even for classification purposes by establishing an appropriate Y related to each sample belonging to a class. In this case, it is called Partial Least Squares – Discriminant Analysis (PLS-DA). The extension of PLS from a regression to a classification method was one such important avenue pursued by the chemometrics community (Brereton and Lloyd, 2018).

PLS-DA make it possible to accomplish a rotation of the projection to give latent variables that focus on class separation (referred to as discrimination). The method offers a convenient way of explicitly considering the class membership of observations. Thus, the objective of PLS-DA is to find a model that separates

classes of observations on the bases of their X-variables description. This model is developed from a training set of observations of known class membership.
In PLS-DA, the X-matrix consists of the data related to the observations; the Y-matrix consists of “dummy” variables (a dummy variable is an artificial variable that assumes a discrete numerical value in the class description), which describes the class membership of each observation.
The results of PLS-DA are interpreted by scores and weight or loadings plots. The scores plot shows the class discriminant ability of the developed model; the weight plot shows which variables contribute to the class separation.
Typical applications of PLS-DA in chemistry include the classification of samples according to their origin, their dating, or the classification of molecules according to their properties or their reaction mechanisms.

2.4.3 Orthogonal Partial least squares discriminant analysis

Orthogonal partial least squares discriminant analysis (OPLS-DA) is a powerful tool used for dimension reduction and identification of specific features that drive group separation. It is a supervised modelling method and may be used instead of PLS-DA due to its capabilities to distinguish between variations in a dataset relevant to predicting group labels and variations irrelevant to predicting group labels. In this sense, OPLS-DA tends to make models that are less complex and more insightful than PLS-DA.
OPLS-DA is especially suited for highlighting class discriminating variables in the two-class problem. The reason for this is that in a two-class discriminant problem, there can, by theory, only be one single predictive component in the OPL-DA mode. Hence, one single predictive loading vector is construed to identify the class discriminating variables. The OPLS-DA is easiest to interpret with 2 classes but is extendable to more ones.
The OPLS algorithm is an extension of the PLS regression method which integrates an orthogonal signal correction filter to distinguish the variations in the data that are useful for the prediction of a quantitative response from the

variations that are orthogonal to the prediction (Trygg and Wold, 2002; Boccard and Rutledge, 2013). Its discriminant analysis counterpart, namely OPLS-DA was demonstrated as a powerful tool for the analysis of qualitative data structures (Bylesjö *et al.*, 2006).

OPLS-DA uses information in the categorical response matrix Y to decompose the X matrix into three distinct parts as described in equation (13), where T_P denotes the predictive score matrix for X, P_P denotes the predictive loading matrix for X, T_O denotes the corresponding Y-orthogonal score matrix, P_O denotes the loading matrix of Y-orthogonal components and E denotes the residual matrix of X:

$$X = T_P P_P^T + T_O P_O^T + E \quad (13)$$

OPLS-DA also relies on a projection of X data as PCA does, but in this case, a maximum variance model becomes the corresponding one of a maximum separation model. This separation is then possible because the projection of the X data will be guided by known class information.

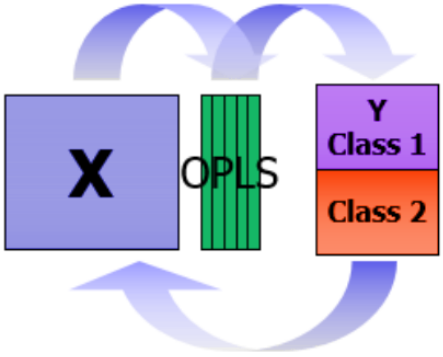


Fig. 13- The matrix relationships in OPLS-DA

The main advantage of OPLS-DA is an easier interpretation of the models, as it focuses on the predictive, i.e. discriminant, information that is summarised in the predictive component(s). The corresponding predictive scores and loading vectors are less subject to orthogonal variation; whereas the orthogonal

component(s) describe the variation that is unrelated to the class response. However, this kind of information can be useful to highlight unexpected systematic variability related to an experimental bias or biological variations (Trygg and Wold, 2002). The main benefit of interpretation using OPLS-DA compared to PLS-DA thus lies in the ability of OPLS-DA to separate predictive from non-predictive (orthogonal) variation.

OPLS-DA is suitable for diagnosing differences between two groups or systems. It indicates which variables have the largest discriminatory power; how the variables are correlated and which are the driving forces among the variables. Scores plot makes it possible to visualize the differences if they exist and the loading plot indicates the variables that express these differences.

OPLS-DA strategy is widely used for the analysis of “omics” data to identify potential biomarkers or product authentication.

The supervised techniques described in this chapter are powerful in classifying the samples according to their group membership; however, they are prone to data overfitting. Thus, they could work very well for the samples inside the model but could be unreliable for predicting unknown samples. For this reason, it is important to validate the statistical model using cross-validation techniques (i.e., CV-ANOVA, and permutation tests) to ensure model reliability.

CHAPTER V

2.5 Saffron

Saffron, one of the costliest herbal spices across the world, is produced from the dried stigmas of *Crocus sativus* L., a blue-purple flower of the Iridaceae family (Figure 14).

Fig. 14– Crocus Sativus L. flower. The red stigmas are used to obtain saffron

This valuable spice is used as a food flavouring and colouring agent, due to its unique aroma, flavour, and distinctive colour. However, saffron possesses also pharmacological and therapeutic properties, being a traditional herbal medicine (Xi and Qian, 2006). Among its chemical compounds, the characteristic ones are crocins, picrocrocin and safranal. The crocins, a family of water-soluble carotenoids, contribute to the rich red colour of saffron; the colourless monoterpene glucoside picrocrocin (4-[β-d-glucopyranosyloxy]-2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde (referred to as picrocrocin) is responsible for the typical bitter taste of saffron (Valle García-Rodríguez *et al.*, 2014); the safranal, released through hydrolysis and dehydration of picrocrocin during the drying and preservation of saffron, is responsible for its characteristic aroma. In addition, other several constituents have been identified in saffron composition, such as flavonoids (such as glycosides of kaempferol and quercetin), sugars, vitamins, amino acids, and aromatic compounds (Winterhalter and Straubinger, 2000).

Currently, Iran ranks first among the saffron producing countries and accounts for around 90% of the global production (Ghorbani, 2007). Moreover, among the major producing countries there are Spain, Morocco, Greece and Italy, where the

Protected Designation of Origin (PDO) trademark guarantees the authenticity of Italian saffron from Sardinia, L'Aquila, and S. Gimignano (Cagliani *et al.*, 2015).
The quality of saffron is greatly influenced by several factors including geographical conditions, harvesting time, drying process utilized, and temperature and oxygen exposure during storage (Caballero-Ortega *et al.*, 2004). Currently, the quality of saffron is determined by specifications described within the International Organization for Standardization (ISO) 3632 1, 2:2010 through the application of a typical targeted analytical method (ISO 3632-2:2010). According to ISO 3632, saffron can be classified into three categories of quality (I, II, III) depending on crocin, picrocrocin and safranal content. Such metabolites are quantified by a spectrophotometric analytical protocol.
Saffron represents one of the most expensive spices on the market (up to 20.000 €/kg). This very high price, although justified by the hard manual labour together with its limited production (around 150,000 flowers are needed to produce 1 kg of saffron), makes saffron very susceptible to economically motivated adulterations (Kumar *et al.*, 2009). There are several ways for adulterating saffron, such as safflower or marigold (floral species of similar morphological features to saffron) are mixed with saffron filaments; the addition of turmeric, paprika, salts and synthetic dyes of familiar appearance to powdered saffron; dipping of saffron with glycerine, honey or syrups to increase its weight; mixing of floral parts of saffron flower including styles and stamens; mislabelling of saffron either belonging to inferior quality or of different origin than that of claimed (Kumari, Jaiswal and Tripathy, 2021). However, safflower (*Carthamus tinctorius* L.) and turmeric (*Curcuma longa* L.) are the major frequent adulterants used to counterfeit saffron (Hagh-Nazari and Keifi, 2007).

2.5.1 Non-targeted NMR method to assess the authenticity of saffron

The current analytical procedure applied for saffron quality control (ISO 3632 1,2:2010) may not be able to detect amounts of safflower or turmeric lower than 20 wt% (Sabatino *et al.*, 2011), as well as other kinds of adulterants in low

concentrations (Marieschi, *et al.*, 2012). Consequently, there is a significant interest in developing accurate analytical methods for saffron quality characterization that could be applied to prevent adulteration or false labelling (Cagliani *et al.*, 2015). In this work, it is reported the application of a non-targeted NMR method toward the detection of adulterated saffron. We explored the capability of such a method to detect contamination of saffron samples with a threshold level of 5 wt% of safflower (*Carthamus tinctorius* L.) and turmeric (*Curcuma longa* L.). The latter represent the two most frequent contaminants found in adulterated saffron on the market (Hagh-Nazari and Keifi, 2007). NMR spectroscopy coupled with chemometric analysis has been reported as a suitable approach for the quality assessment and authentication of this precious spice, being able to detect a wide range of adulterants, from synthetic dyes to herbal materials (Kumari, *et al.*, 2021; Kuballa *et al.*, 2018; Ordoudi *et al.*, 2015; Petrakis *et al.*, 2017; Sobolev *et al.*, 2014; Yilmaz *et al.*, 2010). However, to date, no official NMR analytical protocol is available to detect levels of contaminants in saffron below the threshold level of 20 wt%. Aiming at identifying the metabolites which can be exploited as biomarkers to detect the adulteration taking place, a study was conducted on saffron samples extracted both in D_2O and in DMSO-d_6. In this Chapter the results obtained from this project, conducted during my PhD, are presented, and described in the article below, published on the 2nd of March 2022 in the journal *Applied Sciences*.

Role of the PhD candidate: NMR analysis and spectra elaboration; writing the section of sample preparation.

2.5.2 Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production

applied sciences

MDPI

Article

Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production

Biagia Musio [1,*], Stefano Todisco [1], Marica Antonicelli [1], Cristiano Garino [2], Marco Arlorio [2], Piero Mastrorilli [1,3], Mario Latronico [1,3] and Vito Gallo [1,3,*]

[1] Department of Civil, Environmental, Land, Building and Chemical Engineering, DICATECh, Politecnico di Bari, Via Edoardo Orabona 4, 70125 Bari, Italy; stefano.todisco@poliba.it (S.T.); marica.antonicelli@poliba.it (M.A.); piero.mastrorilli@poliba.it (P.M.); mario.latronico@poliba.it (M.L.)
[2] Department of Pharmaceutical Science—Food Chemistry, Biotechnology, and Nutrition Unit, Università del Piemonte Orientale, Largo Donegani 2/3, 28100 Novara, Italy; cristiano.garino@bfr.bund.de (C.G.); marco.arlorio@uniupo.it (M.A.)
[3] Innovative Solutions S.r.l., Zona H 150/B, 70015 Noci, Italy
* Correspondence: biagia.musio@poliba.it (B.M.); vito.gallo@poliba.it (V.G.)

Featured Application: The analytical approach described in the present paper could find applications in the omics analytical tools devoted to food traceability, quality control, and authenticity assessment by reliable methods.

Abstract: The development of analytical methods aimed at tracing agri-food products and assessing their authenticity is essential to protect food commercial value and human health. An NMR-based non-targeted method is applied here to establish the authenticity of saffron samples. Specifically, 40 authentic saffron samples were compared with 18 samples intentionally adulterated by using turmeric and safflower at three different concentration levels, i.e., 5, 10, and 20 wt%. Statistical processing of NMR data furnished useful information about the main biomarkers contained in aqueous and dimethyl sulfoxide extracts, which are indicative of the presence of adulterants within the analyzed matrix. Furthermore, a discrimination model was developed capable of revealing the type of agronomic practice adopted during the production of this precious spice, distinguishing between organic and conventional cultivation. The main objective of this work was to provide the scientific community involved in the quality control of agri-food products with an analytical methodology able to extract useful information quickly and reliably for traceability and authenticity purposes. The proposed methodology turned out to be sensitive to minor variations in the metabolic composition of saffron that occur in the presence of the two adulterants studied. Both adulterants can be detected in aqueous extracts at a concentration of 5 wt%. A lower limit of detection was observed for safflower contained in organic extracts in which case the lowest detectable concentration was 20%.

Keywords: food counterfeit; traceability; quality control; spectroscopy; organic farming; conventional farming; metabolic composition; fingerprint; chemometrics; biomarker

Citation: Musio, B.; Todisco, S.; Antonicelli, M.; Garino, C.; Arlorio, M.; Mastrorilli, P.; Latronico, M.; Gallo, V. Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production. *Appl. Sci.* **2022**, *12*, 2583. https://doi.org/10.3390/app12052583

Academic Editor: Roberto Romaniello

Received: 28 January 2022
Accepted: 24 February 2022
Published: 2 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

In recent years, globalization of the food supply chain has occurred at a considerable level, and accompanying it, also an escalation of food fraud events. According to the last technical report drawn up by the Food and Agriculture Organization (FAO), it is estimated that the cost of food fraud in the global food industry is approximately 30 billion EUR every year [1]. To face the high economic damage and to prevent potential high risk for human health, the scientific community is working hard to develop new technologies and new analytical approaches to achieve complete traceability of the food chain. In this

Fig. 15– Page 1/15 of the published article

context, the combination of blockchain and artificial intelligence is becoming a valuable tool to trace the several stages within the food chain, including all processes from the acquisition of raw materials to production, consumption, and disposal [2]. Nevertheless, one of the main challenges is to assess food features by detecting variations in the metabolic composition of the product under investigation. The "traditional" approach relies on a targeted analytical method, by which a few a priori known compounds are detected and, eventually, quantified. However, this approach may fail when unknown substances are indicative of either accidental or intentional alteration of the food product. In this contest, the non-targeted analytical method is becoming a more valuable tool for quality control and authenticity assessment [3].

Over the last decades, the use of an NMR-based analytical approach for food quality and authentication has significantly increased [4–6]. The distinctive ability of NMR spectroscopy compared to other analytical techniques is that of providing reproducible spectral data in terms of relative signal intensity, regardless of the spectrometer features [7,8]. Recently, our research group contributed to the harmonization of NMR data production by differently configured spectrometers [9,10]. A further attractive feature of the NMR-based analytical approach relies on the possibility of collecting a large amount of data related to the same sample. Such an NMR feature allows the entire metabolic profile (fingerprint) of a food product to be obtained rapidly and reproducibly [11–15]. Thus, variations of food fingerprints derived from possible alterations can be easily unveiled by NMR analysis at any stage of the control process in the supply chain. As a consequence, NMR data can be exploited, through Distributed Ledger Technologies such as blockchain, to trace food products and certify their quality, origin, and authenticity.

In this work, we report on the application of the non-targeted NMR method toward the detection of adulterated saffron, which represents one of the most expensive spices on the market (up to 20.000 €/kg). This very high price, although justified by the hard manual labor together with its limited production (around 150,000 flowers are needed to produce 1 kg of saffron) [16], makes saffron very susceptible to economically motivated adulterations (EMAs). One of the most common strategies for adulterating saffron is to add natural phytochemicals able to confer organoleptic and/or commercial traits like those naturally belonging to saffron. Safflower (*Carthamus tinctorius* L.) and turmeric (*Curcuma longa* L.) represent the two most frequent contaminants found in adulterated saffron on the market [17].

Currently, the quality of saffron is determined by specifications described within the ISO 3632-2:2011 [18] through the application of a typical targeted analytical method. According to ISO 3632, saffron can be classified into three categories of quality (I, II, III) depending on the concentration of the three known metabolites, namely crocins, picrocrocin, and safranal, contained in this spice and responsible for its typical taste and color. Such metabolites are quantified by a spectrophotometric analytical protocol. Nevertheless, it has been reported that this analytical protocol may not be able to detect amounts of safflower or turmeric lower than 20 wt% [19], as well as other kinds of adulterants in low concentration [20]. Further reported studies explored the application of alternative analytical methods to ascertain the presence of contaminants in saffron under the threshold value of 20% through the use of MALDI mass spectrometry [21], diffuse reflectance infrared Fourier transform spectroscopy [22], Raman spectroscopy [23], and electronic nose [24]. In this context, NMR spectroscopy coupled with chemometric analysis has been reported as a suitable approach for the quality assessment and authentication of this precious spice, being able to detect a wide range of adulterants, from synthetic dyes to herbal materials [25–41]. However, to date, no official NMR analytical protocol is available to detect levels of contaminants in saffron below the threshold level of 20 wt%. In the present work, we explored the capability of the NMR-based non-targeted method to detect contamination of saffron samples with a threshold level of 5 wt% of safflower (*Carthamus tinctorius* L.) and turmeric (*Curcuma longa* L.). Aiming at identifying the metabolites which can be exploited as biomarkers to detect the adulteration taken place, the study was

Fig. 16- Page 2/15 of the published article

conducted on saffron samples extracted both in water and in DMSO-d_6. As a result, a pool of metabolites indicative of the contamination of the saffron samples was identified. In addition, during the present study, the developed non-targeted method was exploited to obtain useful insights into the agronomical practice adopted during the cultivation of the saffron under investigation, allowing discrimination between organic saffron and the conventional type.

2. Materials and Methods

2.1. Chemicals

3-(Trimethylsilyl)-2,2,3,3-tetradeutero-propionic acid sodium salt (TSP, CAS N. 24493-21-8, 99% D, Armar Chemicals, Döttingen, Switzerland), hydrochloric acid (HCl, 37%, CAS N. 7647-01-0; ≥99.5%, Sigma-Aldrich, Milan, Italy), sodium oxalate ($Na_2C_2O_4$, CAS N. 62-76-0; ≥99.5%, Sigma-Aldrich, Milan, Italy) sodium azide (NaN_3, CAS N. 26628-22-8; ≥99.5%, Sigma-Aldrich, Milan, Italy), deuterium oxide (D_2O, CAS. N. 7789-20-0, 99.86% D, Eurisotop, Saclay, France) hexadeutero dimethylsulfoxide (DMSO-d_6, CAS. N. 2206-27-1, 99.5% D, Sigma-Aldrich, Milan, Italy) were used for sample preparation. NMR tubes (Norell 509-UP 7) were provided by Norell, Landisville NJ, United States. The NMR samples were prepared using an automated system for liquid handling (SamplePro Tube, Bruker BioSpin, Bruker, Billerica, MA, USA).

2.2. Plant Material

The saffron stigmas (*Crocus sativus* L.) submitted for investigation were provided by four Italian farms. Table 1 summarizes the number of samples (n) provided by each farm together with all the information on the geographical origin, and the agronomic practice applied for their cultivation.

Table 1. Summary of the main features of the pure saffron samples provided by certified Italian farms.

Producer	*n*	Geographical Origin	Agronomic Practice
Farm A	16	Florence	Organic
Farm B	8	Grosseto	Conventional
Farm C	4	Perugia	Organic
Farm D	12	Padua	Conventional

2.3. Sample Preparation

2.3.1. Artificial Adulteration of Pure Saffron Samples

Saffron samples were adulterated by mixing the pure saffron powder with turmeric (*Curcuma longa*) or safflower (*Carthamus tinctorius*) powders to get turmeric-adulterated samples and safflower-adulterated samples, respectively. Three levels of adulteration were obtained: 20 wt%, 10 wt%, and 5 wt%. Three replicates were prepared for each level of adulteration, obtaining a total of 18 samples of adulterated saffron.

2.3.2. Preparation of Aqueous Extracts

An aliquot of 100 mg of each powdered sample was added to 2.0 mL of buffer solution [$Na_2C_2O_4$ (0.25 M), NaN_3 (2.5 mM)] at pH 4.2 [adjusted by addition of HCl (37%)], vortexed (Advanced Vortex Mixer ZX3, VELP Scientifica Srl, Usmate Velate, Italy) for 5 min at 2500 rpm, pasteurized at 80 °C for 10 min, and centrifuged for 10 min at 6000 rpm (ROTOFIX 32 A, Hettich, Milan, Italy). The pasteurization phase proved essential to maintaining sample stability over 24 h (see Figure S1a,b in Supplementary Materials for further details). NMR tubes were filled in by an automated system for liquid handling (SamplePro Tube, Bruker BioSpin, Bruker, Billerica, MA, USA) according to the following proportion: 540 µL of the saffron extract and 60 µL of TSP/D_2O solution [3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt in D_2O (0.20%)].

Fig. 17- Page 3/15 of the published article

2.3.3. Preparation of Organic Extracts in DMSO-d_6

An aliquot of 10 mg of each powdered sample was added to 600 µL of DMSO-d_6, vortexed (Advanced Vortex Mixer ZX3, VELP Scientifica Srl, Usmate Velate, Italy) for 3 min at 2500 rpm, and centrifuged for 10 min at 6000 rpm (ROTOFIX 32 A, Hettich, Milan, Italy). The DMSO extracts proved to be stable within 24 h even without pasteurization (see Figure S1c in Supplementary Materials for further details). NMR tubes were filled in by an automated system for liquid handling (SamplePro Tube, Bruker BioSpin, Bruker, Billerica, MA, USA) according to the following proportion: 540 µL of the obtained extract and 60 µL of TSP/D_2O solution [3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt in D_2O (0.20%)].

2.4. NMR Experiment

One-dimensional ^{1}H NOESY NMR and ^{1}H NMR spectra were recorded through a Bruker Avance 400 MHz spectrometer equipped with a 5 mm inverse probe and with an autosampler (Bruker, Billerica, MA, USA). The following acquisition parameters were used to record ^{1}H NOESY NMR: pulse program = noesygppr1d; size of fid (TD) = 64 K; spectral width (SW) = 20 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; power level for pre-saturation (pl9) = 62.80 dB; dummy scans (ds) = 4; number of scans (ns) = 32; acquisition time = 4.09 s; mixing time (d8) = 0.01 s; recycle delay (d1) = 3 s. The following acquisition parameters were used to record ^{1}H NMR: pulse program = zg; size of fid (TD) = 64 K; spectral width (SW) = 20 ppm; transmitter offset = 5.00 ppm; 90° hard pulse (p1) = 8.16 µs; power level (pl1) = 1.00 dB; dummy scans (ds) = 0; number of scans (ns) = 128; acquisition time = 4.08 s; recycle delay (d1) = 10.00 s. Bidimensional spectra (TOCSY, and COSY) were acquired with 4096 and 256 data points. The following acquisition parameters were used to record COSY on the aqueous extracts: pulse program = cosygppqf; size of fid (TD) = 4 K; spectral width (SW) = 15 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; power level for pre-saturation (pl9) = 70.00 dB; dummy scans (ds) = 4; number of scans (ns) = 32; acquisition time = 0.34 s; recycle delay (d1) = 9 s; gradient pulse (p16) = 1.00 ms. The following acquisition parameters were used to record COSY on the DMSO-d_6 extracts: pulse program = cosygpqf; size of fid (TD) = 2 K; spectral width (SW) = 15 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; dummy scans (ds) = 4; number of scans (ns) = 16; acquisition time = 0.18 s; recycle delay (d1) = 6 s; gradient pulse (p16) = 1.00 ms. The following acquisition parameters were used to record TOCSY on the aqueous extracts: pulse program = mlevphpr; size of fid (TD) = 4 K; spectral width (SW) = 20 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; power level for pre-saturation (pl9) = 62.80 dB; dummy scans (ds) = 4; number of scans (ns) = 32; acquisition time = 0.26 s; recycle delay (d1) = 7 s; spin lock (d9) = 40.00 ms. The following acquisition parameters were used to record TOCSY on the DMSO-d_6 extracts: pulse program = mlevph; size of fid (TD) = 4 K; spectral width (SW) = 15 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; dummy scans (ds) = 4; number of scans (ns) = 8; acquisition time = 0.26 s; recycle delay (d1) = 2 s; spin lock (d9) = 75.00 ms.

Each spectrum was acquired using TOPSPIN 2.1 software (Bruker BioSpin GmbH, Rheinstetten, Germany) under an automatic process that lasted ca. 15 min and encompassed sample loading, temperature stabilization for 5 min, tuning, matching, shimming, and 90° pulse calibration.

NMR raw data (Free Induction Decays, FIDs) were processed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spain). The FIDs were zero-filled to 128 K number of points and then underwent Fourier transformation by applying an exponential multiplication function with a line broadening of 0.1 Hz. Phase and baseline were automatically corrected, and the TSP singlet signal set at δ = 0.00 ppm was used as a chemical shift reference.

Fig. 18- Page 4/15 of the published article

2.5. Pre-Treatment of Raw Data for the Statistical Analysis

The raw data (FIDs) relative to the 1D ^{1}H NOESY NMR and ^{1}H NMR: experiments were processed by a single operator using Mestrelab and segmented into regular-sized (0.04 ppm) intervals (buckets) in the range of [10.50, 0.50] ppm. The underlying area of each bucket was calculated and normalized to the total intensity. The areas of the buckets in the region [5.13, 4.69] ppm, corresponding to the residual water signal, were set to 0. The data matrices were imported into MetaboAnalyst 5.0, and buckets were subjected to mean-centering and divided by the standard deviation of each variable (Unit Variance scaling). Multivariate statistical analyses such as Principal Component Analysis (PCA), Hierarchical Clustering Dendrogram (HCD), and Partial Least Square-Discriminant Analysis (PLS-DA) were performed. PCA and HCD were used as unsupervised approaches to get an overview of the data. PLS-DA was used as a supervised method that uses multivariate regression techniques to extract via a linear combination of original variables (X) the information that can predict the class membership (Y). The PLS regression was performed using the *plsr* function provided by *R pls* package [42]. The classification and cross-validation were performed using the corresponding wrapper function offered by the *caret* package [43]. To assess the significance of class discrimination, a permutation test was performed. In each permutation, a PLS-DA model was built between the data (X) and the permuted class labels (Y) using the optimal number of components determined by cross-validation for the model based on the original class assignment [44].

3. Results and Discussion

3.1. Fingerprinting for Saffron Authenticity

3.1.1. Metabolic Analysis of Authentic Saffron Aqueous Extracts

The NMR analysis (1D ^{1}H NOESY, COSY, and TOCSY) performed on the aqueous extracts of the authentic pure saffron enabled the identification of a pool of metabolites (see Table S1 in Supplementary Materials for the full list of metabolites). The main representative classes of metabolites detectable in the aqueous extracts of authentic saffron (see Figure S2 in Supplementary Materials for further details) include organic acids (lactic, acetic, malic, fumaric, and formic acids), amino acids (alanine, asparagine, glycine), and carbohydrates (D-glucose, fructose, D-gentiobiose, and D-sucrose). Additionally, the colorless monoterpene glucosidepicrocrocin (4-[β-d-glucopyranosyloxy]-2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde (hereafter referred to as picrocrocin) has been reported as abundantly present in the aqueous extracts of saffron. It represents one of the metabolites majorly responsible for the typical bitter taste of saffron [45]. However, it has been previously reported as occurring from degradation of picrocrocin in aqueous saffron spice extracts upon thermal treatment [46]. Therefore, since the analyzed samples were subjected to a pasteurization process at 80 °C to keep the samples stable for 24 h (see Figure S1 in Supplementary Materials for stability studies), aglycone 4-hydroxy-2,6,6-trimethylcyclohex-1-enecarbaldehyde (HTCC) was expected to be detected as a degradation product. Indeed, characteristic signals as singlets ascribable to this compound were detected at 1.20, 1.23, and 9.97 ppm. Correspondingly, the saccharidic moiety, namely β-D-glucose, derived from the degradation of picrocrocin was identified by the well-resolved doublets at 5.24 and 4.65 ppm, and the doublet of doublets at 3.26 ppm. Moreover, it has been reported that the analyzed aqueous extracts contain significant amounts of a group of water-soluble carotenoids derived from crocetin (8,8′-diapo-Ψ,Ψ′-carotenedioic acid) which are responsible for the coloring capacity of saffron and the use of this spice as food colorant [47]. While a pool of carotenoids derived from crocetin (8,8′-diapo-Ψ,Ψ′-carotenedioic acid) has been identified through LC-ESI-MS using water (or acidified with 0.25% formic acid) and acetonitrile [48], the 1D ^{1}H NOESY spectrum in D_2O, in agreement with the literature [26], contains predominantly the signals related to one of the identified derivatives, namely *trans* crocetin (β-D-gentibiosyl) ester, commonly appointed as crocin with the *trans* isomer significantly more abundant than the *cis* one, which has not been detected. Indeed, in our case, characteristic signals ascribable to the polyconjugated olefinic protons of *trans* crocetin (β-

Fig. 19- Page 5/15 of the published article

D-gentibiosyl) ester are present at 7.26, 6.42, and 6.12 ppm, while the broad singlets related to the methyl moieties can be found at 1.87 and 1.77 ppm. In addition, kaempferol, which is a flavonol under the class of flavonoids, is contained in the saffron aqueous extracts. While stigma samples have been reported to contain high levels of crocin, picrocrocin, and their derivatives, the stamens and sepals of saffron have been reported as rich in flavonols which are present under three forms: kaempferol-3-sophoroside, kaempferol-3-sophoroside-7-glucoside, and kaempferol-3,7,4′-triglucoside [49]. Kaempferol can be detected in the 1D ^{1}H NOESY NMR spectrum of aqueous extracts of stigmas as a mixture of two glycosylated forms, which have been characterized as 7-*O*-glucopyranoside-3-*O*-sophoroside, and the 7-*O*-sophoroside glycosides of kaempferol [49]. In agreement with the literature [26], two doublets at 8.03 and 7.94 ppm, and, two further doublets at 6.97 and 6.89 ppm can be ascribed to these two forms of kaempferol.

3.1.2. Statistical Analysis to Reveal Adulterated Saffron Aqueous Extracts

Univariate Analysis

The binned 1D ^{1}H NOESY NMR spectra data constituted 58 (samples) by 227 (spectra bins) data matrix. Upon normalization to constant sum and auto-scaling, the obtained data were subjected to a univariate analysis method providing a preliminary overview of features that are potentially significant in discriminating the samples under study (see Figure S3 in Supplementary Materials for a visual inspection of 1D ^{1}H NOESY NMR spectra of authentic saffron aqueous extracts and adulterated ones). The unpaired two-sample Wilcoxon test (also known as the Wilcoxon rank-sum test or Mann–Whitney test) as a non-parametric alternative to the unpaired two-samples *t*-test was used to compare two independent groups of samples [50]. As a result, 72 significant features selected by *t*-tests with a *p*-value threshold of 0.05 were detected (see Table S2 in Supplementary Materials for the full list of the top 50 features identified by the Wilcoxon rank-sum test). Among these, the most important variables were related to the spectral regions containing signals ascribed to crocin at 1.79, 6.15, 6.47, and 5.63 ppm, glucose at 3.67 ppm, and formic acid at 8.43 ppm (Figure 1). While variables related to crocin and formic acid were positively correlated to the pure saffron samples (see Supplementary Materials, Figure S4a–e), the variable related to glucose was positively related to the adulterated samples (see Supplementary Materials, Figure S4f).

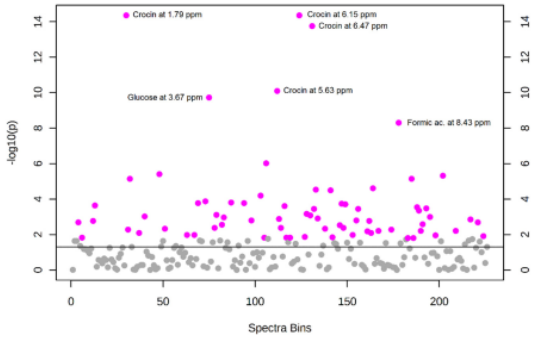


Figure 1. Results of the Wilcoxon rank-sum test. Important features selected by *t*-tests with threshold $p = 0.05$. The red circles represent features above the threshold. The p values were transformed by $-\log10$ so that the more significant features (with smaller p values) are plotted higher on the graph.

Fig. 20- Page 6/15 of the published article

Hierarchical Clustering

An agglomerative hierarchical cluster analysis was performed using the *hclust* function in package *stat*, to detect similarities among the samples under investigation. Selecting the opportune similarity measure and clustering algorithm, each sample begins as a separate cluster and the algorithm proceeds to combine them until all samples belong to one cluster. Specifically, using *Pearson* as distance measure and *average* as a clustering algorithm (clustering uses the centroids of the observations), it was possible to obtain useful information on the metabolites that most characterized the authentic saffron samples compared to the adulterated ones, and vice versa. As a result, the investigated samples clustered into two main groups as observed in the obtained dendrogram (see Figure S5 in Supplementary Materials for further details). The first cluster was composed of pure saffron samples, while the second group contained adulterated samples. Interestingly, the samples which were adulterated with safflower (indicated with S) clustered separately from the samples adulterated with turmeric (indicated with T). Moreover, within each group of adulterated samples, a clear similarity according to the level of adulteration (5, 10, and 20 wt%) was noticeable.

When the results of the hierarchical analysis were analyzed through a heatmap, further information could be extrapolated (Figure 2). In agreement with the results provided by the univariate analysis, the adulterated samples were characterized by higher levels of saccharides, like sucrose (3.59 ppm), fructose (buckets at 3.99 and 4.03 ppm, respectively), and glucose (buckets at 3.67 and 5.39 ppm, respectively, containing also the overlapping signals related to the gentiobiose bound to crocetin). In particular, the samples of saffron previously adulterated with turmeric presented a higher content of these saccharides. On the other hand, the samples adulterated with safflower were characterized by a higher level of succinic acid (bucket at 2.55 ppm). The samples of authentic saffron presented chiefly predominance of crocins (buckets at 1.79, 1.87, 5.63, 6.15, 6.43, and 6.47 ppm) and a shift of the formic acid signal (buckets at 8.43 ppm).

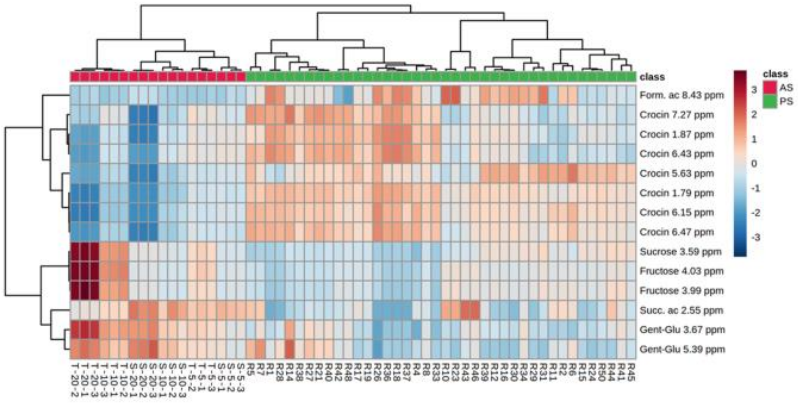


Figure 2. Results of the hierarchical clustering heat map performed on the normalized data (distance measure using *Correlation*, and clustering algorithm using *Complete*). Each colored cell on the map corresponds to a concentration value in the data table, with samples in columns and metabolites in rows. The belonging class of samples is indicated as red and green squares for adulterated saffron (AS) and authentic one (PS), respectively.

Fig. 21- Page 7/15 of the published article

Supervised Multivariate Analysis: Partial Least Squares Discriminant Analysis (PLS-DA)

Aiming at developing a model to discriminate the adulterated saffron samples from the authentic ones, PLS-DA was applied to the data matrix obtained from the binned and pre-processed spectral data. The model presented eight components, though the first seven components were selected as the optimal number of components for classification according to a 10-fold cross-validation method. To estimate the predictive ability of the model, the performance measure, Q2, was calculated via cross-validation (CV). In each CV, the predicted data were compared with the original data, and the sum of squared errors was calculated. The prediction error was then summed over all samples (Predicted Residual Sum of Squares or PRESS). For convenience, the PRESS was divided by the initial sum of squares and subtracted from 1 to resemble the scale of the R2. A model characterized by good predictions and without overfitting should have low PRESS or high Q2, as in the case of the PLS-DA model developed here (see Figure S6 in Supplementary Materials for further details). The PLS-DA model was validated through a permutation test upon passing a set of 1000 permutations and testing the prediction accuracy during training. As a result, the observed statistic *p*-value was lower than 0.001, confirming the good predictive performance of the developed model (see Figure S7 in Supplementary Materials for further details). The first four components exerted the highest contribution with an explained variance of 62.5% (see Figure S8 in Supplementary Materials for further details). Indeed, the PLS-DA scores plot between Component 1 and Component 2 clearly shows two clusters along Component 1 with an explained variance of 14.4% (Figure 3a). The analysis of the Variable Importance in Projection (VIP) was performed to obtain insights into the variables, and, thus, the metabolites, which mostly contributed to the observed distribution of saffron samples in the PLS-DA scores plot. Indeed, VIP represents a weighted sum of squares of the PLS loadings considering the amount of explained Y-variation in each dimension. As a result, a pool of 66 variables with VIP > 1 on the Component 1 was identified (see Table S3 in Supplementary Materials for the full list). A deeper look into the top 15 VIPs (Figure 3b), revealed that crocin represented the main metabolite contributing to the clustering of the authentic saffron samples which were characterized by a higher content of this compound. Such a result agrees with the information obtained from the univariate data analysis of the same aforementioned samples. On the other hand, the group of adulterated saffron samples presented a higher content of saccharides, like fructose and glucose.

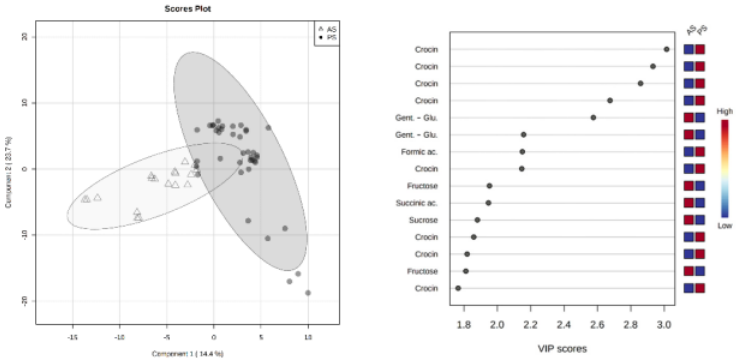
A visual inspection of the spectra of authentic and adulterated saffron samples confirmed the results of the class discrimination since some noticeable variations were observed in the spectral regions containing the signals ascribed to the aforementioned metabolites. Specifically, the content of crocin was relatively higher in the samples of authentic saffron (Figure 4, blue spectrum), while fructose was more abundant in the samples of saffron adulterated with safflower (Figure 4, red spectrum). Significant variations were observed also for the spectral regions containing the signals assigned to the organic acids like formic acid and succinic acid. While for formic acid a slight shift of the corresponding signal was detected, in the case of succinic acid a remarkable abundance of this metabolite in both the adulterated samples (Figure 4, red and green spectra) was observed compared to the authentic ones.

3.1.3. Statistical Analysis to Reveal Adulterated Saffron Organic Extracts

To reveal the adulteration through the study of the less hydrophilic metabolites, a spectroscopic non-targeted method was applied also to the saffron samples obtained by extraction with DMSO-d_6 (see Section 2.3.3 for further details). The determination of the metabolic composition of authentic saffron was achieved by NMR analysis (1D ^{1}H NOESY, COSY, and TOCSY), and the results were compared to spectral data reported in the literature [27,32]. Less hydrophilic metabolites could be detected, like safranal, glycosylated forms of kaempferol, linoleic acid, and further fatty acids (see Figure S9 and Table S4 in Supplementary Materials for further details). A data matrix (57 samples by 245 spectra bins) was obtained from the processing (normalization to the sum and auto-

Fig.22- Page 8/15 of the published article

scaling) of the recorded ^{1}H NMR spectra (see Figure S10 in Supplementary Materials for a visual inspection of ^{1}H NMR spectra of extracts in DMSO-d_6 of authentic and adulterated saffron samples) and was, next, subjected to univariate analysis (*t*-test) and unsupervised multivariate analysis (PCA) in analogy to the statistical elaboration performed on the corresponding aqueous extracts. As a result, a less noticeable separation between the authentic samples and the adulterated ones could be extrapolated. Indeed, discrimination between these groups of samples could be detected only along the PC5 with a relatively low (4.5%) explained variance (see Figure S11 in Supplementary Materials for further details). A more noticeable clustering of these two classes of samples could be detected through supervised multivariate data analysis, namely OPLS-DA, as depicted in the corresponding scores plot (Figure 5a). Such a model was characterized by one predictive component and four orthogonal ones, and it was successfully validated upon passing the permutation test (see Figures S12 and S13 in Supplementary Materials for further details). As illustrated in Figure 5b, the top ten variables with VIP[**t**] > 1.70 on Component 1 (*t* score) included buckets related to the compounds belonging to the class of curcuminoids (indicated in Figure 5b as Cur), namely curcumin, demethoxycurcumin, and bisdemethoxycurcumin. These compounds are natural phenols and they are often employed to color foods and medicines, because of the typical yellow color derived from them [51]. They were previously isolated from turmeric and characterized via NMR [52]. The adulterated samples presented a relatively higher content of curcuminoids (Cur), along with linoleic acid (Lin. Ac.), picrocrocin (Picr), and safranal (Safr).



(**a**) (**b**)

Figure 3. Results of PLS-DA performed on the spectral data obtained from the aqueous extracts of authentic (PS, represented as circles) and adulterated saffron (AS, represented as triangles): (**a**) Scores plot between the selected PCs. The explained variances are shown in brackets; (**b**) Analysis of the Variable Importance in Projection (VIP) identified by PLS-DA. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study.

The analysis of the hierarchical clustering heat map performed on the normalized data (distance measurement, Euclidean; clustering method, Ward) helped to identify the spectral regions most susceptible to changes in the metabolic composition induced by the two different types of adulteration, namely by addition of turmeric vs. safflower (Figure 6). It is interesting to note that the adulteration through turmeric could be spotted even at a very low concentration, i.e., 5 wt%. Indeed, diagnostic buckets were those related to crocin (1.97 ppm), picrocrocin (1.76 ppm), curcuminoids (7.07, 7.01, 7.33, and 6.05 ppm),

Fig.23- Page 9/15 of the published article

and safranal (5.97 ppm). On the other hand, the samples adulterated with safflower could be detected only at higher levels of concentration, like 20 wt%. In this case, the spectral regions containing the signals of linoleic acid (at 0.85 ppm) and crocin (at 7.33 ppm which overlaps the curcuminoid signals) could be adopted as diagnostic metabolites to detect the adulteration with safflower.

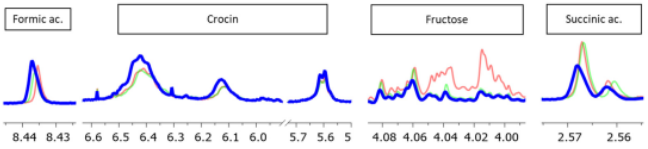
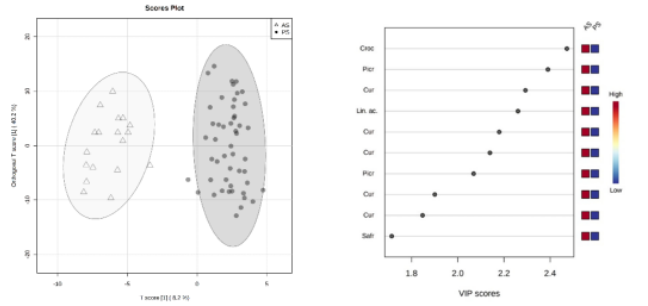


Figure 4. Comparison of typical 1D ^{1}H NOESY NMR spectra (Bruker Avance 400 MHz, D_2O) of aqueous extracts of authentic saffron (blue spectrum), adulterated samples with 20 wt% of safflower (red spectrum), and adulterated samples with 20 wt% of turmeric (green spectrum).



(**a**) (**b**)

Figure 5. Results of OPLS-DA performed on the spectral data obtained from the extracts of authentic (PS, represented as circles) and adulterated (AS, represented as triangles) saffron in DMSO-d_6: (**a**) Scores plot between the selected PCs. The explained variances are shown in brackets. (**b**) Analysis of the Variable Importance in Projection (VIP) identified by OPLS-DA. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study.

3.2. Fingerprinting for Determining the Agronomic Practice Adopted for Saffron Cultivation

The non-targeted NMR approach applied to the spectroscopic data derived from the analysis of the aqueous extracts of pure saffron allowed extrapolation of insightful information related to the agronomic practice adopted for this spice cultivation. In the PC1/PC2 scores plot (see Supplementary Materials, Figure S11), a clear clustering of the samples was observed when the regular-sized (0.04 ppm) spectral buckets were subjected to the unsupervised PCA. Through a more in-depth examination of the characteristics of the sample, including the geographical origin, the producer, and the agronomic practice adopted for cultivation (see Table S5 in Supplementary Materials for the complete list of available information relating to the investigated samples), the agronomic practices were found to be the most influencing factor on samples distribution. The supervised PLS-DA (see Figures S15 and S16 in Supplementary Materials for further details on the model overview and model validation) was conducted to obtain more information regarding changes in metabolic composition associated with the agronomic practice employed. The analysis of

Fig.24- Page 10/15 of the published article

the Variable Importance in Projection, VIP (see Table S5 in Supplementary Materials for the full list of 78 variables with VIP > 1), along with the visual inspection of the 1D ^{1}H NOESY NMR spectra allowed the selection of a pool of 15 spectral regions acting as diagnostic variables towards the discrimination between the saffron samples obtained through conventional cultivation compared to those derived from organic cultivation (Figure 7b).

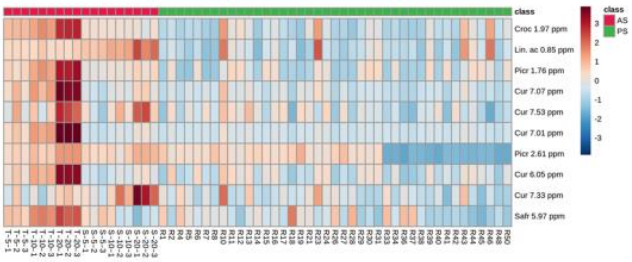
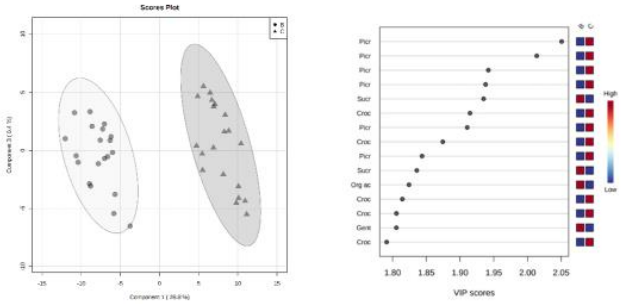


Figure 6. Results of the hierarchical clustering heat map performed on the normalized data (distance measure using *Euclidean*, and clustering algorithm using *Complete*). Each colored cell on the map corresponds to a concentration value in the data table, with samples in columns and metabolites in rows. The belonging class of samples is indicated as red and green squares for adulterated saffron (AS) and authentic one (PS), respectively.



(**a**) (**b**)

Figure 7. Results of PLS-DA performed on the spectral data obtained from the aqueous extracts of authentic saffron cultivated through conventional cultivation (C, represented as triangles) and organic one (B, represented as circles): (**a**) Scores plot between the selected PCs. The explained variances are shown in brackets. (**b**) Analysis of the Variable Importance in Projection (VIP) identified by PLS-DA. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study.

The hierarchical clustering heatmap performed on the normalized data (distance measure, Euclidean; Clustering method, Ward) clearly shows that the aqueous extracts of saffron cultivated according to an organic agronomic practice were characterized by a relatively higher content of sugars, like sucrose and gentiobiose, along with organic acids, as succinic and malic acids. On the other hand, it was found that the two most typical metabolites of saffron, namely picrocrocin and crocin, were contained majorly in those samples derived from a conventional agronomic practice (Figure 8).

Fig.25- Page 11/15 of the published article

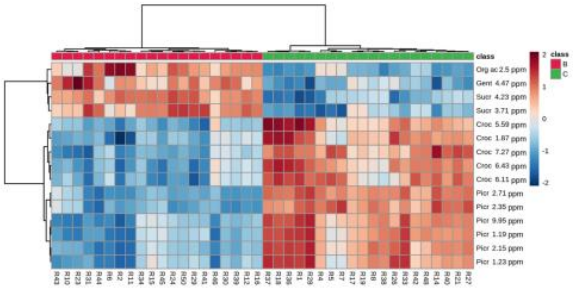


Figure 8. Results of the hierarchical clustering heat map performed on the normalized data (distance measure using *Euclidean*, and clustering algorithm using *Ward*). Each colored cell on the map corresponds to a concentration value in the data table, with samples in columns and metabolites in rows. The belonging class of samples is indicated as red and green squares for saffron cultivated according to organic farming (B) and conventional (C), respectively.

4. Conclusions

In this study, a non-targeted NMR screening was performed on a group of authentic saffron samples and on a group of saffron samples intentionally adulterated by the addition of safflower and turmeric at different concentrations. The elaboration of NMR data of aqueous extracts by chemometrics allowed the discrimination between authentic and adulterated saffron samples, detecting the presence of both adulterants at a concentration value equal to 5 wt%. When a similar study was performed on the same samples extracted in DMSO-d_6, the presence of turmeric was detected at a concentration level of 5%, while it was possible to identify the presence of safflower only at higher levels of concentration, i.e., 20 wt%. To the best of our knowledge, to date, no non-targeted NMR-based method has been reported in the literature capable of detecting the presence of contaminants at these concentration levels. Furthermore, through the same analytical approach, it was possible to obtain crucial information on the type of cultivation practice adopted for the production of saffron. The results reported in this study can help to enhance the application of non-targeted NMR-based metabolomics in the control processes of the agri-food supply chain with the final goal to valorize sustainable farming practices and ensure traceability and authenticity of food products.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app12052583/s1. Figure S1: Effect of pasteurization step on the stability of aqueous extracts of pure saffron. Table S1: List of metabolites contained in the aqueous extracts of authentic saffron samples and identified by 1D 1H NOESY NMR. Figure S2: Typical 1D ^{1}H NOESY NMR spectrum of aqueous extracts of authentic saffron. Figure S3: 1D ^{1}H NOESY NMR spectra of aqueous extracts of authentic saffron, authentic turmeric, authentic safflower, saffron adulterated with turmeric, and saffron adulterated with safflower. Table S2: List of the top 50 features identified by the Wilcoxon rank-sum test applied to the aqueous extracts of authentic and adulterated saffron samples. Figure S4: Trend of the top six most significant variables according to the Wilcoxon rank sum test applied to the aqueous extracts of authentic and adulterated saffron samples. Figure S5: Results of the hierarchical analysis applied to the spectral data obtained from the aqueous extracts of authentic and adulterated saffron samples. Figure S6: 10-fold cross-validation method applied to the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples. Figure S7: Permutation test applied to the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples. Figure S8: Overview of the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating

Fig.26- Page 12/15 of the published article

between authentic and adulterated saffron samples. Table S3: List of the top 66 variables importance in projection (VIPs) related to the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples. Figure S9: Typical ^{1}H NMR spectrum of extracts of authentic saffron in DMSO-d_6. Figure S10: ^{1}H NMR spectra of extracts of authentic saffron, authentic turmeric, authentic safflower, saffron adulterated with turmeric, and saffron adulterated with safflower in DMSO-d_6. Table S4: List of metabolites contained extracts of authentic saffron samples in DMSO-d_6 and identified by ^{1}H NMR. Figure S11: Results of PCA applied to the spectral data obtained from extracts of authentic saffron and adulterated saffron in DMSO-d_6. Figure S12: Overview of the OPLS-DA model built on the spectral data obtained from the extracts in DMSO-d_6 for discriminating between authentic and adulterated saffron samples. Figure S13: Permutation test applied to the OPLS-DA model built on the spectral data obtained from the extracts in DMSO-d_6 for discriminating between authentic and adulterated saffron samples. Table S5: List of features characteristic of each sample of authentic saffron under investigation. Figure S14: Results of PCA performed on the spectral data obtained from the aqueous extracts of authentic saffron. Figure S15: Overview of PLS-DA built on the spectral data of the aqueous extracts of authentic saffron samples for discriminating according to the agronomic practice: organic vs. conventional. Figure S16: 10-fold cross-validation method applied to the PLS-DA model built on spectral data of the aqueous extract of authentic saffron samples for discriminating according to the agronomic practice: organic vs. conventional. Table S6: List of the top 78 Variables importance in projection (VIPs) related to the PLS-DA model built the spectral data of the aqueous extracts of authentic saffron samples for discriminating according to the agronomic practice: organic vs. conventional.

Author Contributions: Conceptualization, V.G. and B.M.; methodology, B.M.; validation, B.M.; formal analysis, S.T. and M.A. (Marica Antonicelli); investigation, S.T. and M.A. (Marica Antonicelli); resources, M.A. (Marco Arlorio), and C.G.; writing—original draft preparation, B.M.; writing—review and editing, B.M., V.G., P.M., M.L., C.G., and M.A. (Marco Arlorio); visualization, M.A. (Marica Antonicelli) and V.G.; supervision, B.M. and V.G.; project administration, V.G.; funding acquisition, V.G. and M.A. (Marco Arlorio). All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by The European Union's Seventh Framework Program for research, technological development, and demonstration, grant number 613688—FOODINTEGRITY.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the Italian farms which provided the saffron samples subjected to the present study: Starace Filomena (Grosseto, Tuscany), Poggio al Sole (Fiesole, Tuscany), La saggezza della Terra (Monselice, Veneto), and Alessandro Mazzuoli (Città della Pieve, Umbria).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. FAO Food Fraud-Intention. *Detection, and Management: Food Safety Technical Toolkit for Asia and the Pacific*; FAO: Bangkok, Thailand, 2021.
2. Xu, J.; Guo, S.; Xie, D.; Yan, Y. Blockchain: A new safeguard for agri-foods. *Artif. Intell. Agric.* **2020**, *4*, 153–161. [CrossRef]
3. Ballin, N.Z.; Laursen, K.H. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends Food Sci. Technol.* **2019**, *86*, 537–543. [CrossRef]
4. Hatzakis, E. Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 189–220. [CrossRef]
5. Sobolev, A.P.; Thomas, F.; Donarski, J.; Ingallina, C.; Circi, S.; Cesare Marincola, F.; Capitani, D.; Mannina, L. Use of NMR applications to tackle future food fraud issues. *Trends Food Sci. Technol.* **2019**, *91*, 347–353. [CrossRef]
6. Consonni, R.; Cagliani, L.R. The potentiality of NMR-based metabolomics in food science and food authentication assessment. *Magn. Reson. Chem.* **2019**, *57*, 558–578. [CrossRef]
7. Musio, B.; Ragone, R.; Todisco, S.; Rizzuti, A.; Latronico, M.; Mastrorilli, P.; Pontrelli, S.; Intini, N.; Scapicchio, P.; Triggiani, M.; et al. A community-built calibration system: The case study of quantification of metabolites in grape juice by qNMR spectroscopy. *Talanta* **2020**, *214*, 120855. [CrossRef]

Fig.27- Page 13/15 of the published article

8. Gallo, V.; Intini, N.; Mastrorilli, P.; Latronico, M.; Scapicchio, P.; Triggiani, M.; Bevilacqua, V.; Fanizzi, P.; Acquotti, D.; Airoldi, C.; et al. Performance Assessment in Fingerprinting and Multi Component Quantitative NMR Analyses. *Anal. Chem.* **2015**, *87*, 6709–6717. [CrossRef]
9. Gallo, V.; Ragone, R.; Musio, B.; Todisco, S.; Rizzuti, A.; Mastrorilli, P.; Pontrelli, S.; Intini, N.; Scapicchio, P.; Triggiani, M.; et al. A Contribution to the Harmonization of Non-targeted NMR Methods for Data-Driven Food Authenticity Assessment. *Food Anal. Methods* **2020**, *13*, 530–541. [CrossRef]
10. Ragone, R.; Todisco, S.; Triggiani, M.; Pontrelli, S.; Latronico, M.; Mastrorilli, P.; Intini, N.; Ferroni, C.; Musio, B.; Gallo, V. Development of a food class-discrimination system by non-targeted NMR analyses using different magnetic field strengths. *Food Chem.* **2020**, 127339. [CrossRef]
11. Kim, S.B.; Bisson, J.; Friesen, J.B.; Bucchini, L.; Gafner, S.; Lankin, D.C.; Chen, S.-N.; Pauli, G.F.; Mcalpine, J.B. *The Untargeted Capability of NMR Recognizes Adulterated Natural Products*; ChemRxiv: Cambridge, UK, 2020. [CrossRef]
12. Sobolev, A.P.; Mannina, L.; Proietti, N.; Carradori, S.; Daglia, M.; Giusti, A.M.; Antiochia, R.; Capitani, D. Untargeted NMR-Based Methodology in the Study of Fruit Metabolites. *Molecules* **2015**, *20*, 4088–4108. [CrossRef] [PubMed]
13. Gao, B.; Holroyd, S.E.; Moore, J.C.; Laurvick, K.; Gendel, S.M.; Xie, Z. Opportunities and Challenges Using Non-targeted Methods for Food Fraud Detection. *J. Agric. Food Chem.* **2019**, *67*, 8425–8430. [CrossRef]
14. Medina, S.; Pereira, J.A.; Silva, P.; Perestrelo, R.; Câmara, J.S. Food fingerprints—A valuable tool to monitor food authenticity and safety. *Food Chem.* **2019**, *278*, 144–162. [CrossRef] [PubMed]
15. McGrath, T.F.; Haughey, S.A.; Patterson, J.; Fauhl-Hassek, C.; Donarski, J.; Alewijn, M.; van Ruth, S.; Elliott, C.T. What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed?—Spectroscopy case study. *Trends Food Sci. Technol.* **2018**, *76*, 38–55. [CrossRef]
16. Kumar, R.; Singh, V.; Devi, K.; Sharma, M.; Singh, M.K.; Ahuja, P.S. State of art of saffron (*Crocus sativus* L.) agronomy: A comprehensive review. *Food Rev. Int.* **2009**, *25*, 44–85. [CrossRef]
17. Nazari, S.H.; Keifi, N. Saffron and various fraud manners in its production and trades. *Acta Hortic.* **2007**, *739*, 411–416. [CrossRef]
18. SIST ISO 3632-2:2011—Spices—Saffron (*Crocus sativus* L.)—Part 2: Test Methods. Available online: https://standards.iteh.ai/catalog/standards/sist/5236fc90-8d64-4cb1-839f-3cd687309147/sist-iso-3632-2-2011 (accessed on 30 December 2021).
19. Sabatino, L.; Scordino, M.; Gargano, M.; Belligno, A.; Traulo, P.; Gagliano, G. HPLC/PDA/ESI-MS evaluation of saffron (Crocus sativus L.) adulteration. *Nat. Prod. Commun.* **2011**, *6*, 1873–1876. [CrossRef] [PubMed]
20. Marieschi, M.; Torelli, A.; Bruni, R. Quality control of saffron (Crocus sativus L.): Development of SCAR markers for the detection of plant adulterants used as bulking agents. *J. Agric. Food Chem.* **2012**, *60*, 10998–11004. [CrossRef]
21. Aiello, D.; Siciliano, C.; Mazzotti, F.; Di Donna, L.; Athanassopoulos, C.M.; Napoli, A. Molecular species fingerprinting and quantitative analysis of saffron (*Crocus sativus* L.) for quality control by MALDI mass spectrometry. *RSC Adv.* **2018**, *8*, 36104–36113. [CrossRef]
22. Petrakis, E.A.; Polissiou, M.G. Assessing saffron (Crocus sativus L.) adulteration with plant-derived adulterants by diffuse reflectance infrared Fourier transform spectroscopy coupled with chemometrics. *Talanta* **2017**, *162*, 558–566. [CrossRef]
23. Dai, H.; Gao, Q.; He, L. Rapid Determination of Saffron Grade and Adulteration by Thin-Layer Chromatography Coupled with Raman Spectroscopy. *Food Anal. Methods* **2020**, *13*, 2128–2137. [CrossRef]
24. Heidarbeigi, K.; Mohtasebi, S.S.; Foroughirad, A.; Ghasemi-Varnamkhasti, M.; Rafiee, S.; Rezaei, K. Detection of adulteration in saffron samples using electronic nose. *Int. J. Food Prop.* **2015**, *18*, 1391–1401. [CrossRef]
25. Kumari, L.; Jaiswal, P.; Tripathy, S.S. Various techniques useful for determination of adulterants in valuable saffron: A review. *Trends Food Sci. Technol.* **2021**, *111*, 301–321. [CrossRef]
26. Dowlatabadi, R.; Farshidfar, F.; Zare, Z.; Pirali, M.; Rabiei, M.; Khoshayand, M.R.; Vogel, H.J. Detection of adulteration in Iranian saffron samples by 1H NMR spectroscopy and multivariate data analysis techniques. *Metabolomics* **2017**, *13*, 1–11. [CrossRef]
27. Cusano, E.; Consonni, R.; Petrakis, E.A.; Astraka, K.; Cagliani, L.R.; Polissiou, M.G. Integrated analytical methodology to investigate bioactive compounds in Crocus sativus L. flowers. *Phytochem. Anal.* **2018**, *29*, 476–486. [CrossRef] [PubMed]
28. Petrakis, E.A.; Cagliani, L.R.; Tarantilis, P.A.; Polissiou, M.G.; Consonni, R. Sudan dyes in adulterated saffron (*Crocus sativus* L.): Identification and quantification by 1H NMR. *Food Chem.* **2017**, *217*, 418–424. [CrossRef] [PubMed]
29. Petrakis, E.A.; Cagliani, L.R.; Polissiou, M.G.; Consonni, R. Evaluation of saffron (*Crocus sativus* L.) adulteration with plant adulterants by1H NMR metabolite fingerprinting. *Food Chem.* **2015**, *173*, 890–896. [CrossRef]
30. Sobolev, A.P.; Carradori, S.; Capitani, D.; Vista, S.; Trella, A.; Marini, F.; Mannina, L. Saffron samples of different origin: An nmr study of microwave-assisted extracts. *Foods* **2014**, *3*, 403–419. [CrossRef] [PubMed]
31. Ordoudi, S.A.; Cagliani, L.R.; Lalou, S.; Naziri, E.; Tsimidou, M.Z.; Consonni, R. 1H NMR-based metabolomics of saffron reveals markers for its quality deterioration. *Food Res. Int.* **2015**, *70*, 1–6. [CrossRef]
32. Cagliani, L.R.; Culeddu, N.; Chessa, M.; Consonni, R. NMR investigations for a quality assessment of Italian PDO saffron (*Crocus sativus* L.). *Food Control* **2015**, *50*, 342–348. [CrossRef]
33. Righi, V.; Parenti, F.; Tugnoli, V.; Schenetti, L.; Mucci, A. Crocus sativus Petals: Waste or Valuable Resource? The Answer of High-Resolution and High-Resolution Magic Angle Spinning Nuclear Magnetic Resonance. *J. Agric. Food Chem.* **2015**, *63*, 8439–8444. [CrossRef]
34. Varliklioz Er, S.; Eksi-Kocak, H.; Yetim, H.; Boyaci, I.H. Novel Spectroscopic Method for Determination and Quantification of Saffron Adulteration. *Food Anal. Methods* **2017**, *10*, 1547–1555. [CrossRef]

Fig.28- Page 14/15 of the published article

35. Schumacher, S.; Mayer, S.; Sproll, C.; Lachenmeier, D.; Kuballa, T. Authentication of saffron (*Crocus sativus* L.) using 1H nuclear magnetic resonance (NMR) spectroscopy. In Proceedings of the XIII Intern. Conf. on the Applications of Magnetic Resonance in Food Science, Karlsruhe, Germany, 7–10 June 2016; p. 13. [CrossRef]
36. Yilmaz, A.; Nyberg, N.T.; Mølgaard, P.; Asili, J.; Jaroszewski, J.W. 1 H NMR metabolic fingerprinting of saffron extracts. *Metabolomics* **2010**, *6*, 511–517. [CrossRef]
37. Wan, K.W.; Lee, M.W.; Choi, H.K. Metabolic discrimination of safflower petals of various origins using 1H NMR spectroscopy and multivariate statistical analysis. *Bull. Korean Chem. Soc.* **2007**, *28*, 557–560. [CrossRef]
38. Ordoudi, S.A.; Cagliani, L.R.; Melidou, D.; Tsimidou, M.Z.; Consonni, R. Uncovering a challenging case of adulterated commercial saffron. *Food Control* **2017**, *81*, 147–155. [CrossRef]
39. Gad, H.A.; Bouzabata, A. Application of chemometrics in quality control of Turmeric (*Curcuma longa*) based on Ultra-violet, Fourier transform-infrared and 1H NMR spectroscopy. *Food Chem.* **2017**, *237*, 857–864. [CrossRef]
40. Belmonte-Sánchez, E.; Romero-González, R.; Garrido Frenich, A. Applicability of high-resolution NMR in combination with chemometrics for the compositional analysis and quality control of spices and plant-derived condiments. *J. Sci. Food Agric.* **2021**, *101*, 3541–3550. [CrossRef] [PubMed]
41. Kuballa, T.; Brunner, T.S.; Thongpanchang, T.; Walch, S.G.; Lachenmeier, D.W. Application of NMR for authentication of honey, beer and spices. *Curr. Opin. Food Sci.* **2018**, *19*, 57–62. [CrossRef]
42. Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Softw.* **2007**, *18*, 1–23. [CrossRef]
43. The Caret Package. Available online: https://topepo.github.io/caret/ (accessed on 30 December 2021).
44. Bijlsma, S.; Bobeldijk, I.; Verheij, E.R.; Ramaker, R.; Kochhar, S.; Macdonald, I.A.; Van Ommen, B.; Smilde, A.K. Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation. *Anal. Chem.* **2006**, *78*, 567–574. [CrossRef]
45. Chrysanthou, A.; Pouliou, E.; Kyriakoudi, A.; Tsimidou, M.Z. Sensory Threshold Studies of Picrocrocin, the Major Bitter Compound of Saffron. *J. Food Sci.* **2016**, *81*, S189–S198. [CrossRef]
46. Sánchez, A.M.; Carmona, M.; Jarén-Galán, M.; Mínguez Mosquera, M.I.; Alonso, G.L. Picrocrocin kinetics in aqueous saffron spice extracts (Crocus sativus L.) upon thermal treatment. *J. Agric. Food Chem.* **2011**, *59*, 249–255. [CrossRef]
47. Bathaie, S.Z.; Farajzade, A.; Hoshyar, R. A review of the chemistry and uses of crocins and crocetin, the carotenoid natural dyes in saffron, with particular emphasis on applications as colorants including their use as biological stains. *Biotech. Histochem.* **2014**, *89*, 401–411. [CrossRef] [PubMed]
48. Carmona, M.; Zalacain, A.; Sánchez, A.M.; Novella, J.L.; Alonso, G.L. Crocetin esters, picrocrocin and its related compounds present in Crocus sativus stigmas and Gardenia jasminoides fruits. Tentative identification of seven new compounds by LC-ESI-MS. *J. Agric. Food Chem.* **2006**, *54*, 973–979. [CrossRef] [PubMed]
49. Agrawal, P.K. Characterization of By-products of Saffron (Crocus sativus L.) Production. *Nat. Prod. Commun.* **2014**, *9*. [CrossRef]
50. Kassambara, A. *Practical Statistics in R II—Comparing Groups: Numerical Variables*; Independently Published; The Design Themes Inc.: Sydney, Australia, 2019.
51. Giuliani, A.; Cerretani, L.; Cichelli, A. Colors: Properties and Determination of Natural Pigments. *Encycl. Food Health* **2016**, 273–283. [CrossRef]
52. Praveen, A.; Prasad, D.; Mishra, S.; Nagarajan, S.; Chaudhari, S.R. Facile NMR approach for profiling curcuminoids present in turmeric. *Food Chem.* **2021**, *341*, 128646. [CrossRef] [PubMed]

Fig.29- Page 15/15 of the published article

2.5.3 Supplementary Materials: Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production

Supplementary Materials: Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production

Biagia Musio [1,*], Stefano Todisco [1], Marica Antonicelli [1], Cristiano Garino [2], Marco Arlorio [2], Piero Mastrorilli [1,3], Mario Latronico [1,3] and Vito Gallo [1,3,*]

Summary

Fig. 30- Page 1/31 of Supplementary Materials

Fig. 31- Page 2/31 of Supplementary Materials

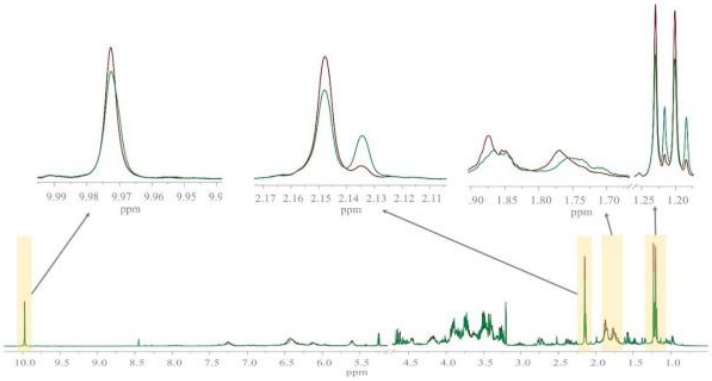


Figure 1. Effect of pasteurization step on the stability of aqueous extracts of pure saffron.

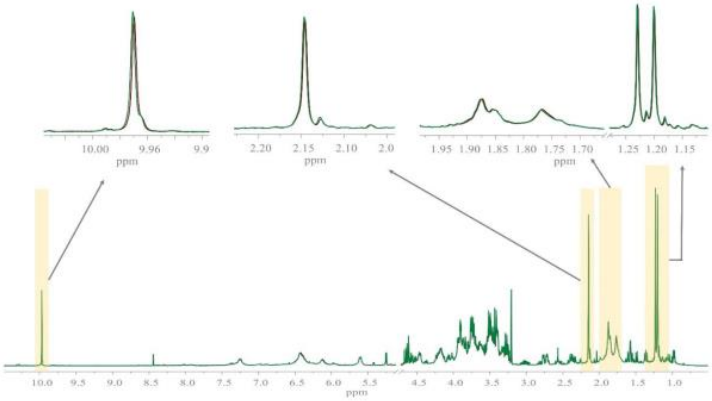


Figure S1 a. Comparison of typical 1D ^{1}H NOESY NMR spectra of aqueous extracts of pure saffron not subjected to pasteurization (80°C). The red spectrum was recorded on the freshly prepared sample; the green spectrum was recorded on the same sample after 24h. Selected regions of spectra (indicated with yellow rectangles and expanded for clarity) show noticeable variations in the metabolic composition over 24 h.

Fig. 32- Page 3/31 of Supplementary Materials

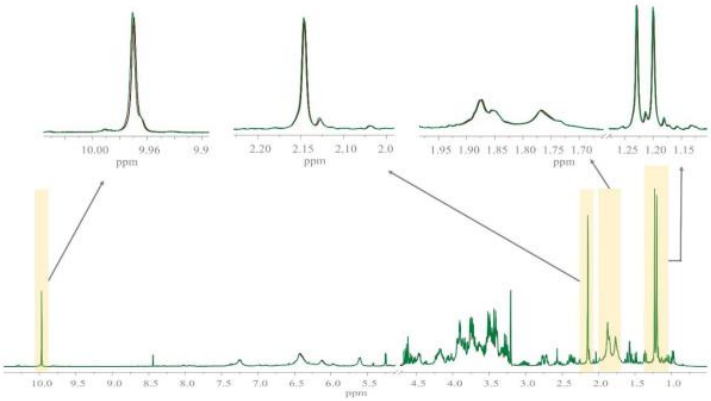


Figure S1 b. Comparison of typical 1D 1H NOESY NMR spectra of aqueous extracts of pure saffron subjected to pasteurization (80°C). Red spectrum recorded on the freshly prepared sample; green spectrum recorded on the same sample after 24h. Selected regions of spectra (indicated with yellow rectangles and expanded for clarity) that do not show noticeable variations in the metabolic composition over 24 h. The metabolic composition resulted stable over 24 h.

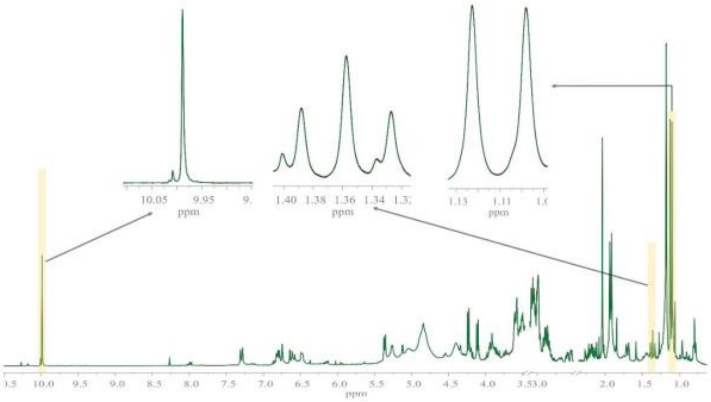


Figure S1 c. Comparison of typical ^{1}H NMR spectra of extracts of pure saffron in DMSO-d_6 not subjected to pasteurization (80°C). Red spectrum recorded on the freshly prepared sample; green spectrum recorded on the same sample after 24h. The metabolic composition resulted stable over 24 h.

Fig. 33- Page 4/31 of Supplementary Materials

Table S1. List of metabolites contained in the aqueous extracts of authentic saffron samples and identified by 1D ^{1}H NOESY NMR. Bruker Avance 400 MHz. The assignment of NMR signals has been obtained by comparison with standard compounds, except for the metabolites indicated with a star. For these metabolites the assignment was obtained referring to spectral data reported in the literature.[1]

Compound	^{1}H ppm	Multeplicity	J (Hz)
Monoterpenes			
Picrocrocin*	1.20	s	
	1.23	s	
	1.57	t	12.1
	1.85	brs	
	2.14	s	
	2.36	dd	18.8;9.4
	2.72	ddd overlapped	
	4.17	m	
	9.97	s	
Carotenoids			
***trans*-crocetin(β-D-gentibiosyl) ester**	1.77	brs	
	1.87	brs	
	6.12	brs	
	6.42	brs	
	5.60	d	7.7
	7.26	d	11.2
Organic acids			
Lactic acid	1.35	d	6.9
	4.15	q	6.9
Acetic acid*	2.03	s	
Malic acid	2.58	dd	16;8.5
	2.74	dd overlapped	16; 4
	4.37	dd	8.5; 4
Succinic acid	2.57	s	
Fumaric acid	6.58	s	
Formic acid*	8.44	s	
Flavonoids			
Kaempferol-3-sophoroside*	6.90	d	8.4
	7.94	d	8.7
Kaempferol-3-sophoroside-7-β-glucoside*	6.90	d	8.4
	6.97	d	8.6
	8.03	d	8.6
Carbohydrates			
D-glucose	3.25	dd	9.1; 7.9
	3.42	m	
	3.47	m	
	3.55	dd	9.8; 3.7
	3.74	m	
	3.84	m	
	3.90	dd overlapped	12.3; 2.1
	4.65	d	7.9
	5.24	dd	3.7
D-sucrose	3.48	t	9.2

Fig. 34- Page 5/31 of Supplementary Materials

	3.57	dd	9.9;3.7
	3.69	s	
	3.78	t	9
	3.83	m	
	3.87	m	
	3.91	dd overlapped	6.2; 3.5
	4.06	t	8.5
	4.22	d	8.7
	5.42	d	3.8
Fructose	3.57	m	
	3.72	m	
	3.81	m	
	3.91	dd	9.9; 3.4
	4.03	m	
	4.12	m	
Amino Acids			
Alanine	1.48	d	7.3
	3.79	q	7.3
Asparagine	3.02	m	
Glycine*	3.54	s	
Quaternary ammonium salts			
Choline	3.2	s	
	3.51	m	
	4.07	m	

Fig. 35- Page 6/31 of Supplementary Materials

Figure S2. Typical 1D ^{1}H NOESY NMR spectrum of aqueous extracts of authentic saffron. Bruker Avance 400 MHz.

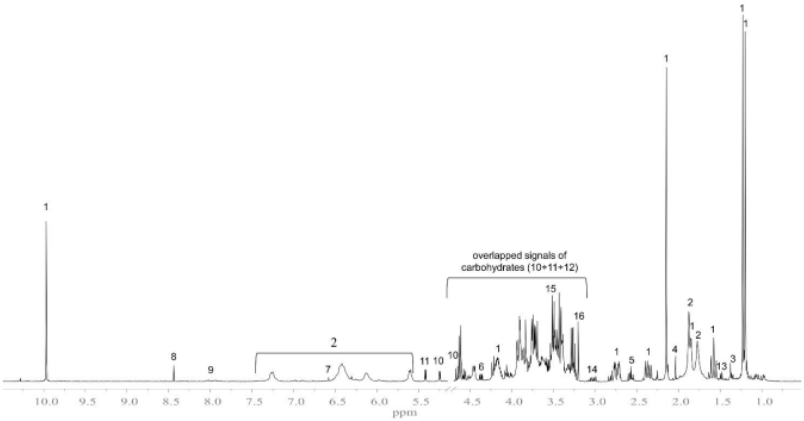


Figure S2 a. ^{1}H NOESY NMR spectrum. Residual water signal (4.78 ppm) is hidden. Signals are indicated as follow: 1, picrocrocin; 2, trans-crocetin (β-D-gentibiosyl) ester; 3, lactic acid; 4, acetic acid; 5, succinic acid; 6, malic acid; 7, fumaric acid; 8, formic acid; 9, glycosylated isoforms of kaempferol; 10, D-glucose; 11, D-sucrose; 12; Fructose; 13, alanine; 14, asparagine; 15, glycine; 16, choline.

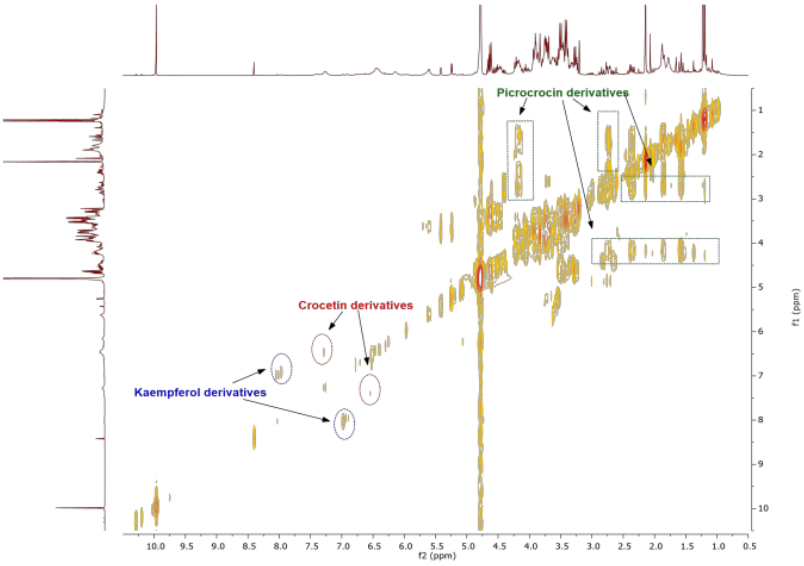


Figure S2. b. COSY spectrum.

Fig. 36- Page 7/31 of Supplementary Materials

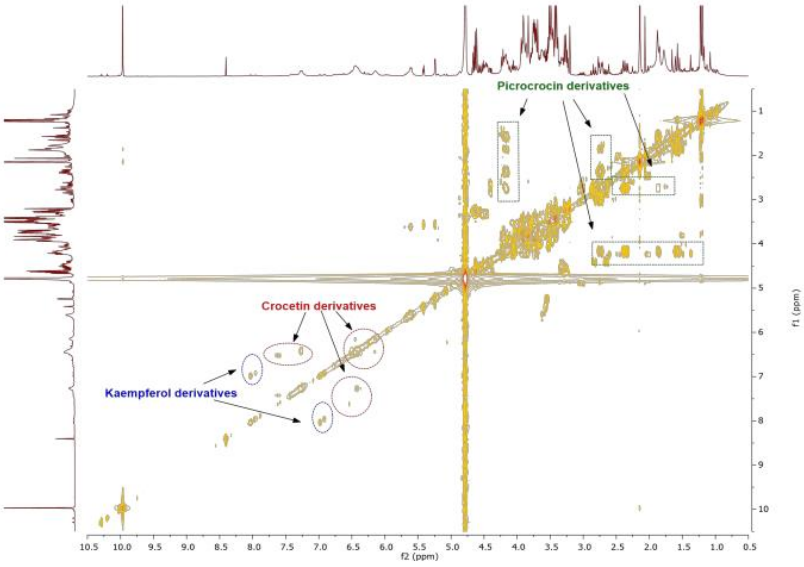


Figure S2. c. TOCSY spectrum.

Fig. 37- Page 8/31 of Supplementary Materials

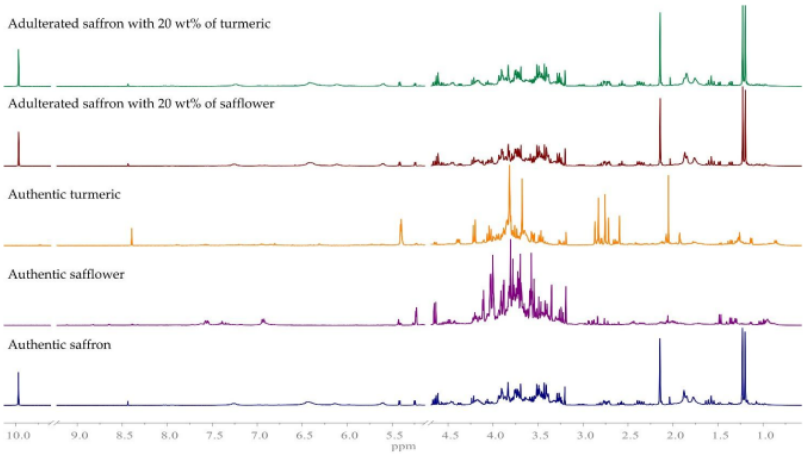


Figure S3. 1D ^{1}H NOESY NMR spectra of aqueous extracts of authentic saffron, authentic turmeric, authentic safflower, saffron adulterated with turmeric, and saffron adulterated with safflower. Comparison of typical 1D ^{1}H NOESY NMR spectra (Bruker Avance 400 MHz) of aqueous extracts (100 mg of the corresponding powder in 2.0 mL oxalate buffer pH 4.2) of authentic saffron (blue spectrum,), authentic safflower (purple spectrum), authentic turmeric (orange spectrum), adulterated saffron with 20 wt% of safflower (red spectrum), and adulterated saffron with 20 wt% of turmeric (green spectrum). Residual water signal (4.78 ppm) is hidden.

Fig. 38- Page 9/31 of Supplementary Materials

Table S2. List of the top 50 features identified by Wilcoxon rank-sum test applied to the aqueous extracts of authentic and adulterated saffron samples.

Range (ppm)	V	p.value	-log10(p)	FDR
1.77272 - 1.81272	0	4.4447e-15	14.352	5.0225e-13
6.13272 - 6.17272	0	4.4447e-15	14.352	5.0225e-13
6.45272 - 6.49272	2	1.7779e-14	13.75	1.3393e-12
5.61272 - 5.65272	28	8.0787e-11	10.093	4.5645e-09
3.65272 - 3.69272	688	1.8915e-10	9.7232	8.5498e-09
8.41272 - 8.45272	50	4.9928e-09	8.3017	1.8806e-07
5.37272 - 5.41272	631	9.6195e-07	6.0168	3.1057e-05
2.53272 - 2.57272	618	3.9254e-06	5.4061	0.00011089
9.49272 - 9.53272	616	4.8169e-06	5.3172	0.00012096
1.85272 - 1.89272	108	7.1909e-06	5.1432	0.00014774
8.77272 - 8.81272	612	7.1909e-06	5.1432	0.00014774
7.85272 - 7.89272	599	2.4547e-05	4.61	0.0004623
6.61272 - 6.65272	123	2.9371e-05	4.5321	0.00051061
6.93272 - 6.97272	596	3.2099e-05	4.4935	0.00051817
5.25272 - 5.29272	132	6.3974e-05	4.194	0.00096387
3.57272 - 3.61272	579	0.00013326	3.8753	0.0018822
4.17272 - 4.21272	577	0.00015596	3.807	0.0020054
3.41272 - 3.45272	144	0.0001686	3.7732	0.0020054
4.45272 - 4.49272	144	0.0001686	3.7732	0.0020054
7.17272 - 7.21272	575	0.00018216	3.7395	0.0020585
7.25272 - 7.29272	146	0.00019673	3.7061	0.0021172
1.01272 - 1.05272	572	0.0002291	3.64	0.0023535
5.77272 - 5.81272	149	0.00024706	3.6072	0.0024276
8.89272 - 8.93272	569	0.00028688	3.5423	0.0027015
9.09272 - 9.13272	567	0.00033249	3.4782	0.0029941
6.57272 - 6.61272	154	0.0003577	3.4465	0.0029941
7.53272 - 7.57272	566	0.0003577	3.4465	0.0029941
8.93272 - 8.97272	563	0.00044413	3.3525	0.0035848
6.33272 - 6.37272	557	0.00067642	3.1698	0.0052714
3.81272 - 3.85272	555	0.0007755	3.1104	0.0058421
6.41272 - 6.45272	166	0.00082982	3.081	0.0060496
2.21272 - 2.25272	168	0.00094892	3.0228	0.0067017
9.17272 - 9.21272	551	0.0010141	2.9939	0.006945
4.01272 - 4.05272	550	0.0010833	2.9653	0.0072006
6.65272 - 6.69272	172	0.0012346	2.9085	0.007972
5.65272 - 5.69272	173	0.0013172	2.8804	0.008269
10.0927 - 10.1327	546	0.0014047	2.8524	0.0085803
4.61272 - 4.65272	176	0.0015957	2.797	0.0092469
7.49272 - 7.53272	544	0.0015957	2.797	0.0092469
0.972723 - 1.01272	543	0.0016997	2.7696	0.009369
7.77272 - 7.81272	543	0.0016997	2.7696	0.009369
0.612723 - 0.652723	540	0.0020492	2.6884	0.01077
10.2527 - 10.2927	540	0.0020492	2.6884	0.01077
9.01272 - 9.05272	536	0.0026149	2.5825	0.013431

Fig. 39- Page 10/31 of Supplementary Materials

3.97272 - 4.01272	535	0.0027766	2.5565	0.013945
7.13272 - 7.17272	534	0.0029471	2.5306	0.014479
3.77272 - 3.81272	528	0.0041808	2.3787	0.019283
5.69272 - 5.73272	192	0.0041808	2.3787	0.019283
7.21272 - 7.25272	528	0.0041808	2.3787	0.019283
2.65272 - 2.69272	526	0.0046838	2.3294	0.020756
6.81272 - 6.85272	526	0.0046838	2.3294	0.020756
1.81272 - 1.85272	524	0.0052395	2.2807	0.022342
8.25272 - 8.29272	524	0.0052395	2.2807	0.022342
7.97272 - 8.01272	521	0.0061825	2.2088	0.024951
8.97272 - 9.01272	521	0.0061825	2.2088	0.024951
9.77272 - 9.81272	521	0.0061825	2.2088	0.024951
7.73272 - 7.77272	520	0.0065285	2.1852	0.025885
2.09272 - 2.13272	204	0.0080886	2.0921	0.030983
7.81272 - 7.85272	516	0.0080886	2.0921	0.030983
3.13272 - 3.17272	511	0.010491	1.9792	0.038241
3.33272 - 3.37272	511	0.010491	1.9792	0.038241
7.41272 - 7.45272	511	0.010491	1.9792	0.038241
9.29272 - 9.33272	510	0.01104	1.957	0.039602
10.3727 - 10.4127	508	0.012212	1.9132	0.043125
6.29272 - 6.33272	506	0.013492	1.8699	0.04691
6.97272 - 7.01272	505	0.014174	1.8485	0.048059
0.692723 - 0.732723	504	0.014886	1.8272	0.048059
5.33272 - 5.37272	504	0.014886	1.8272	0.048059
5.81272 - 5.85272	504	0.014886	1.8272	0.048059
5.89272 - 5.93272	504	0.014886	1.8272	0.048059
8.61272 - 8.65272	503	0.015628	1.8061	0.049054
8.81272 - 8.85272	503	0.015628	1.8061	0.049054

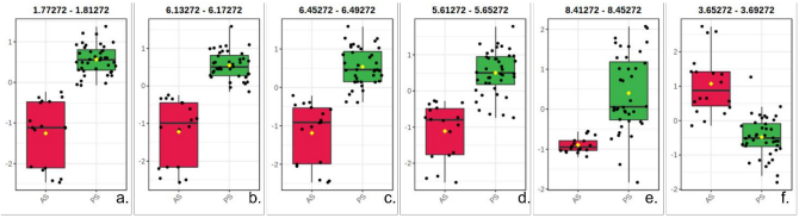


Figure S4. Trend of the top six most significant variables according to the Wilcoxon rank sum test applied to the aqueous extracts of authentic and adulterated saffron samples.Plots of the trend of the buckets containing the signals related to the following metabolites: a) crocin at 1.79 ppm; b) crocin at 6.13 ppm; c) crocin at 6.47 ppm; d) crocin at 5.63 ppm; e) formic acid at 8.43 ppm; f) glucose at 3.67 ppm.

Fig. 40- Page 11/31 of Supplementary Materials

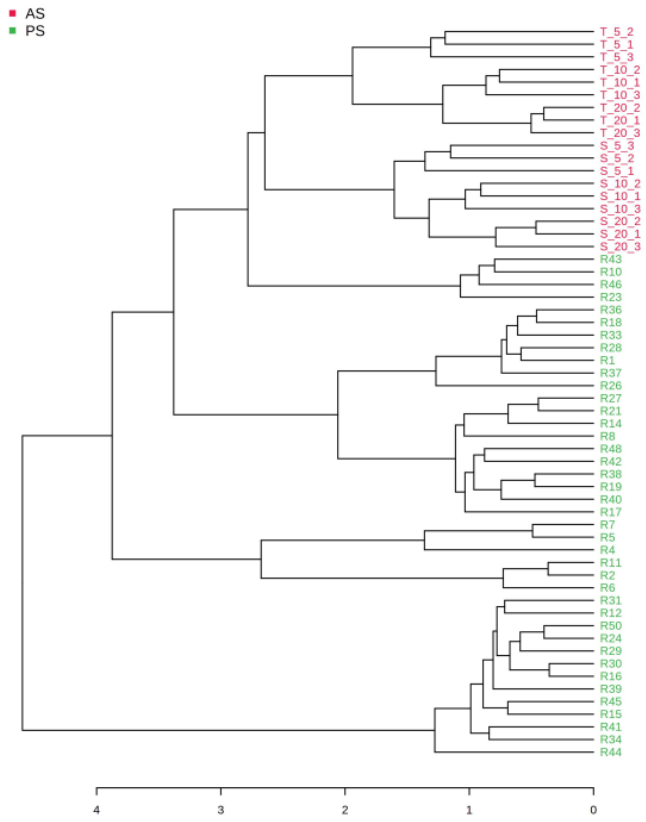


Figure S5. Results of the hierarchical analysis applied to the spectral data obtained from the aqueous extracts of authentic and adulterated saffron samples.Clustering uses the centroids of the observations. Distance measure: *Pearson*; clustering algorithm: average. The samples are indicated in red and green for adulterated samples (AS) and authentic ones (PS), respectively.

Fig. 41- Page 12/31 of Supplementary Materials

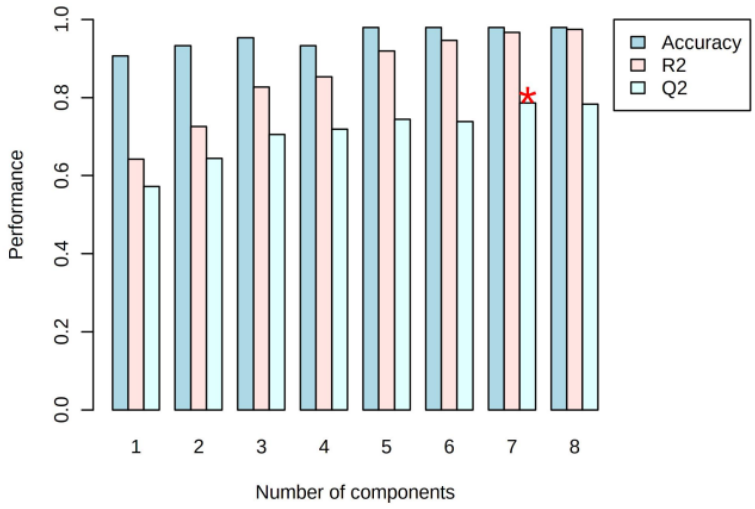


Figure S6. 10-fold cross-validation method applied to the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples. Application of 10-fold cross-validation to PLS-DA model to select the optimal number of components for classification and estimate the predictive performance of the classification model. The red star indicates the best classifier.

Fig. 42- Page 13/31 of Supplementary Materials

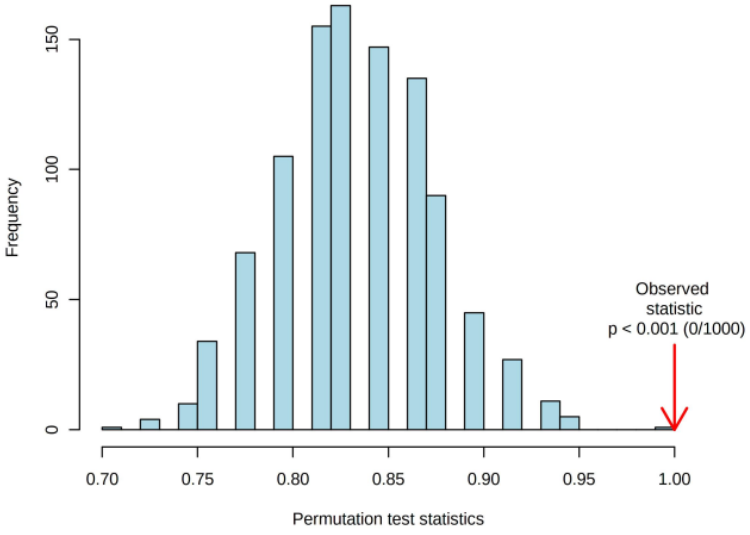


Figure S7. Permutation test applied to the PLS-DA model built on spectral data built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples. The permutation test consists of a set of 1000 permutations and testing the prediction accuracy during training. The *p*-value based on permutation is $p < 0.001$ (0/1000).

Fig. 43- Page 14/31 of Supplementary Materials

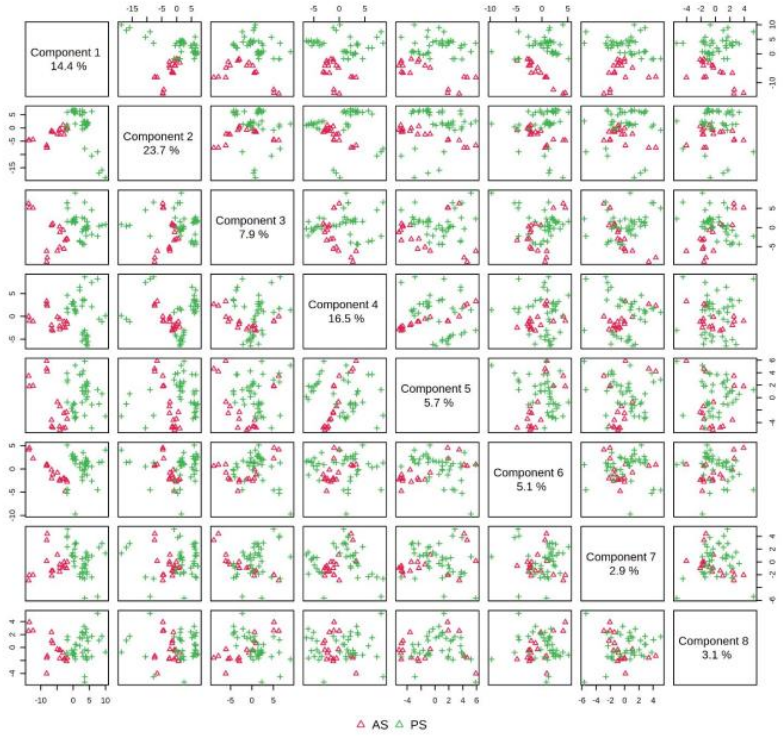


Figure S8. Overview of the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples. Pairwise scores plots between the selected components, where the scores are indicated as follows: green crosses and red triangles for authentic samples (PS) and adulterated ones (AS), respectively. The explained variance of each component is shown in the corresponding diagonal cell.

Fig. 44- Page 15/31 of Supplementary Materials

Table S3. List of the top 66 variables importance in projection (VIPs) related to the PLS-DA model built on the spectral data obtained from the aqueous extracts for discriminating between authentic and adulterated saffron samples.List of the top 66 variables identified as VIP > 1 on Component 1.

Range (ppm)	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
1.77272 - 1.81272	3.0151	2.8929	2.7209	2.6788	2.5834
6.13272 - 6.17272	2.9333	2.7799	2.6072	2.5658	2.482
6.45272 - 6.49272	2.8587	2.7254	2.5622	2.5215	2.4415
5.61272 - 5.65272	2.6757	2.5212	2.3627	2.3422	2.2601
3.65272 - 3.69272	2.576	2.4225	2.2789	2.2438	2.1683
5.37272 - 5.41272	2.158	2.0333	1.9231	1.914	1.8456
8.41272 - 8.45272	2.1505	2.0958	2.0563	2.0856	2.0631
1.85272 - 1.89272	2.1473	2.0798	1.951	1.9255	1.9084
4.01272 - 4.05272	1.9534	1.8372	1.7624	1.7475	1.6843
2.53272 - 2.57272	1.9475	1.9033	1.7889	1.7831	1.7525
3.57272 - 3.61272	1.8813	1.7692	1.68	1.6623	1.6128
7.25272 - 7.29272	1.8593	1.83	1.7234	1.6988	1.7087
6.41272 - 6.45272	1.8197	1.7489	1.6445	1.6205	1.6488
3.97272 - 4.01272	1.8121	1.7047	1.6611	1.6414	1.5863
1.81272 - 1.85272	1.7657	1.6813	1.6235	1.6693	1.6264
7.17272 - 7.21272	1.7632	1.762	1.7	1.7406	1.7025
3.77272 - 3.81272	1.7494	1.6452	1.5822	1.5753	1.5297
4.09272 - 4.13272	1.743	1.6481	1.6521	1.6331	1.5743
7.53272 - 7.57272	1.6824	1.5822	1.5491	1.5264	1.4714
4.17272 - 4.21272	1.6593	1.6282	1.5357	1.5175	1.4626
4.45272 - 4.49272	1.6569	1.5715	1.4991	1.501	1.4495
6.33272 - 6.37272	1.6453	1.5901	1.5203	1.5691	1.5136
7.85272 - 7.89272	1.6126	1.5447	1.4607	1.4421	1.3899
5.25272 - 5.29272	1.6085	1.5316	1.4369	1.4254	1.3892
4.61272 - 4.65272	1.6004	1.537	1.5111	1.4898	1.4746
9.05272 - 9.09272	1.5578	1.617	1.5943	1.5916	1.5563
3.41272 - 3.45272	1.5519	1.46	1.4231	1.4036	1.4129
8.61272 - 8.65272	1.5285	1.4636	1.4394	1.4221	1.3709
3.81272 - 3.85272	1.5057	1.4161	1.3776	1.358	1.3623
6.93272 - 6.97272	1.5014	1.4894	1.3985	1.3897	1.3568
5.65272 - 5.69272	1.4869	1.446	1.3641	1.345	1.3139
9.09272 - 9.13272	1.479	1.4477	1.5738	1.5579	1.5243
8.81272 - 8.85272	1.4688	1.4147	1.4385	1.4164	1.3811
1.93272 - 1.97272	1.451	1.3651	1.3092	1.296	1.2975
9.49272 - 9.53272	1.4393	1.3703	1.3022	1.2854	1.241
7.21272 - 7.25272	1.3818	1.3504	1.4218	1.4214	1.3746
3.73272 - 3.77272	1.3632	1.2963	1.2267	1.2393	1.2133
2.61272 - 2.65272	1.335	1.3057	1.2651	1.2451	1.2047
2.21272 - 2.25272	1.3279	1.2788	1.2141	1.2012	1.1612
8.97272 - 9.01272	1.3132	1.2349	1.2048	1.2181	1.1806
7.97272 - 8.01272	1.3129	1.2627	1.2192	1.2668	1.3242
7.57272 - 7.61272	1.304	1.2288	1.264	1.2444	1.2004
3.33272 - 3.37272	1.296	1.2194	1.178	1.1894	1.1586
1.61272 - 1.65272	1.2718	1.2428	1.1665	1.1487	1.2279

Fig. 45- Page 16/31 of Supplementary Materials

4.13272 - 4.17272	1.2605	1.2285	1.151	1.1364	1.1668
6.29272 - 6.33272	1.2574	1.2902	1.2263	1.2981	1.2575
2.57272 - 2.61272	1.2559	1.2223	1.1605	1.1422	1.1609
7.73272 - 7.77272	1.2366	1.1665	1.1783	1.1599	1.1206
7.41272 - 7.45272	1.2206	1.2463	1.1678	1.18	1.156
6.57272 - 6.61272	1.2189	1.206	1.1401	1.1222	1.123
7.01272 - 7.05272	1.1968	1.1405	1.1403	1.134	1.1334
9.77272 - 9.81272	1.1844	1.1693	1.1044	1.0873	1.1014
3.45272 - 3.49272	1.1765	1.1138	1.1029	1.1097	1.1066
3.01272 - 3.05272	1.1732	1.1033	1.0902	1.0734	1.1454
3.49272 - 3.53272	1.1694	1.1189	1.1327	1.1361	1.1575
2.09272 - 2.13272	1.1548	1.0901	1.0219	1.0078	0.99676
1.57272 - 1.61272	1.1075	1.0428	1.1431	1.1495	1.1165
6.61272 - 6.65272	1.091	1.1275	1.0644	1.0475	1.0335
1.25272 - 1.29272	1.0816	1.0181	1.0039	0.98803	1.0543
4.53272 - 4.57272	1.0679	1.0043	1.1189	1.1078	1.0688
4.05272 - 4.09272	1.0589	1.0119	0.97142	0.98594	0.99046
9.97272 - 10.0127	1.0447	0.98788	1.0181	1.0048	0.96931
9.57272 - 9.61272	1.0438	1.0596	1.0456	1.0291	0.99289
4.41272 - 4.45272	1.0418	1.0141	0.97588	1.001	1.0187
5.69272 - 5.73272	1.0044	1.0688	1.0261	1.01	0.97542
3.93272 - 3.97272	1.0033	0.94716	0.88779	0.9336	0.90987

Fig. 46- Page 17/31 of Supplementary Materials

Figure S9. NMR analysis of extracts of authentic saffron in DMSO-d_6.Bruker Avance 400 MHz.

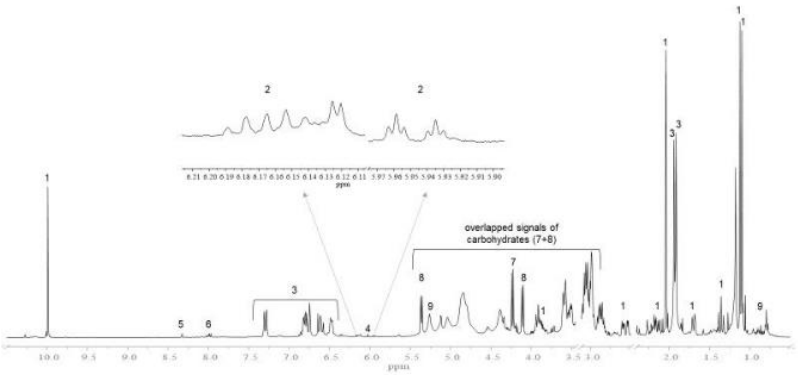


Figure S9. a. ^{1}H NMR. Residual solvent signals (2.44 and 3.27 ppm) are hidden. Signals are indicated as follow: 1, picrocrocin; 2,safranal; 3, trans-crocetin (β-D-gentibiosyl) ester ;4,fumaric acid; 5,formic acid; 6,glycosylated isoforms of kaempferol; 7,β-D-glucosyl of picrocrocin; 8,β-D-gentiobiosyl of trans-crocetin ester; 9,linoleic acid.

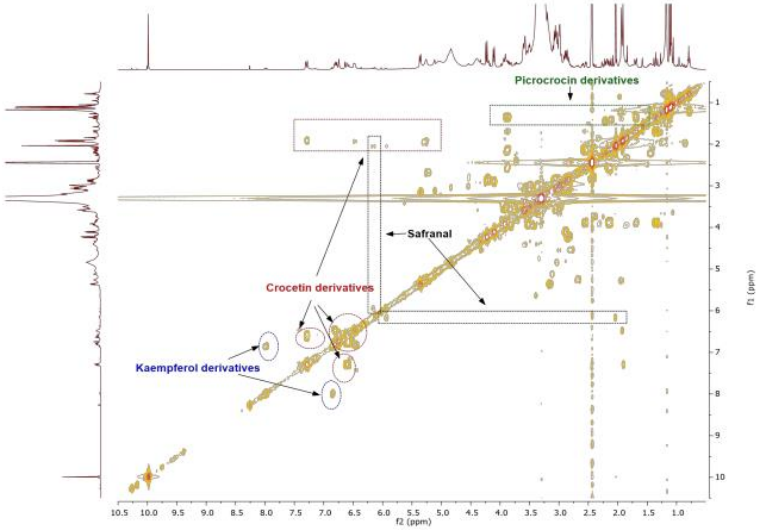


Figure S9. b. COSY spectrum.

Fig. 47- Page 18/31 of Supplementary Materials

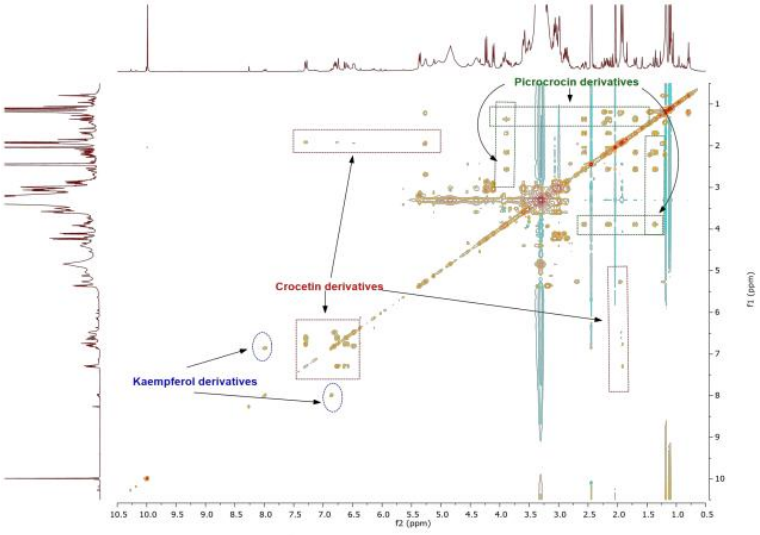


Figure S9. c. TOCSY spectrum.

Fig. 48- Page 19/31 of Supplementary Materials

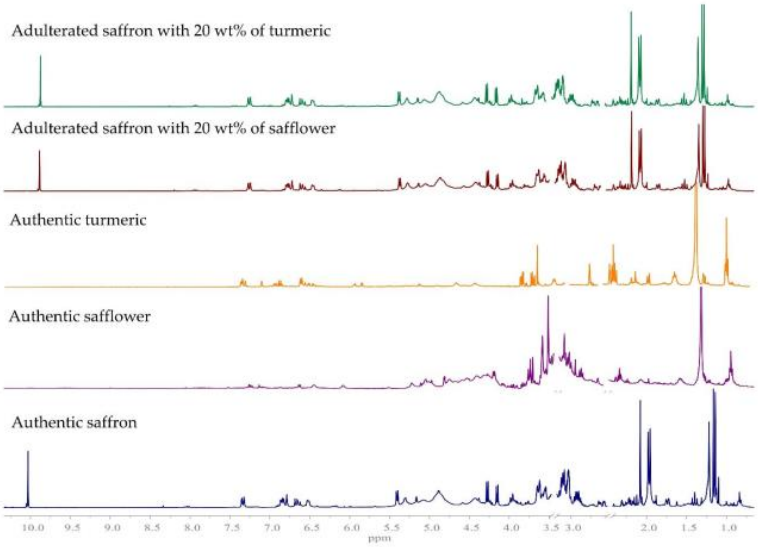


Figure S10. ^{1}H NMR spectra of extracts of authentic saffron, authentic turmeric, authentic safflower, saffron adulterated with turmeric, and saffron adulterated with safflower **in DMSO-d_6.** (Bruker Avance 400 MHz) Comparison of typical 1H NMR spectra of extracts in DMSO-d6 of authentic saffron (blue spectrum), authentic safflower (purple spectrum), authentic turmeric (orange spectrum), adulterated saffron with 20 wt% of safflower (red spectrum), and adulterated saffron with 20 wt% of turmeric (green spectrum). Residual solvent signals (2.44 and 3.27 ppm) are hidden.

Fig. 49- Page 20/31 of Supplementary Materials

Table S4. List of metabolites contained extracts of authentic saffron samples in DMSO-d_6 and identified by ^{1}H NMR. The assignment of NMR signals has been obtained by comparison with standard compounds, except for the metabolites indicated with a star. For these metabolites the assignment was obtained referring to spectral data reported in the literature. [2,3].

Compound	^{1}H ppm	Multeplicity	J (Hz)
Monoterpenes			
Picrocrocin*	1.10	s	
	1.12	s	
	1.36	t	12.2
	1.70	ddd	12.2; 2.5
	2.04	s	
	2.15	dd	18.8;9.2
	2.56	ddd	18.9; 5.3; 1.8
	3.88	m	
	9.99	s	
Safranal*	2.02	m	
	5.95	dt	9.4; 1.8
	6.16	dd	9.6; 5.1
Carotenoids			
trans-crocetin(β-D-gentibiosyl) ester *	1.91	s	
	1.94	s	
	6.48	dd	7.7;2.6
	6.61	dd	14.9; 11.5
	6.76	d	14.9
	6.81	dd	7.9;2.6
	7.29	d	11.2
Organic acids			
Fumaric acid	6.02	s	
Formic acid	8.33	s	
Flavonoids			
Kaempferol-3-sophoroside*	6.11	d	2.1
	6.37	d	2.1
	6.85	d	8.8
	7.98	d	8.9
Kaempferol-3-sophoroside-7-β-glucoside*	6.87	d	8.8
	8.01	d	8.9
Carbohydrates			
β-D-glucosyl of picrocrocin*	4.23	d	7.8
β-D-gentiobiosyl of trans-crocetin ester*	5.24	d	7.7
	4.11	d	7.7
Fatty acid			
Linoleic acid*	0.86	t	7.5

Fig. 50- Page 21/31 of Supplementary Materials

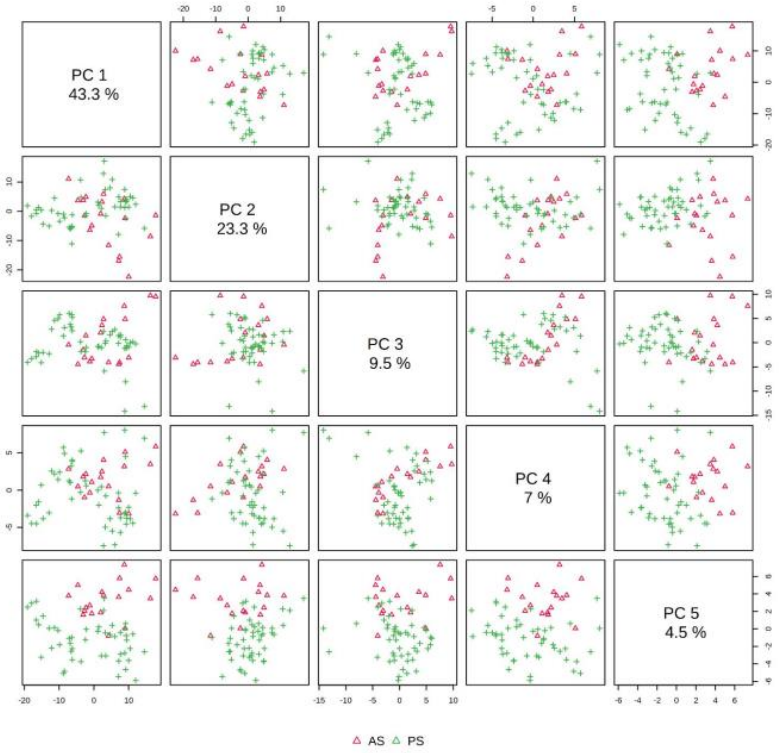


Figure S11. Results of PCA applied to the spectral data obtained from extracts of authentic saffron and adulterated saffron in DMSO-d_6. PCA was applied to the 67 spectra of authentic saffron and adulterated saffron extracted in DMSO-d_6 by using UV-scaled 0.04 ppm-sized bucketing. Pairwise scores plots between the selected components where the scores are indicated as follows: green crosses and red triangles for authentic samples and adulterated ones, respectively. The explained variance of each component is shown in the corresponding diagonal cell.

Fig. 51- Page 22/31 of Supplementary Materials

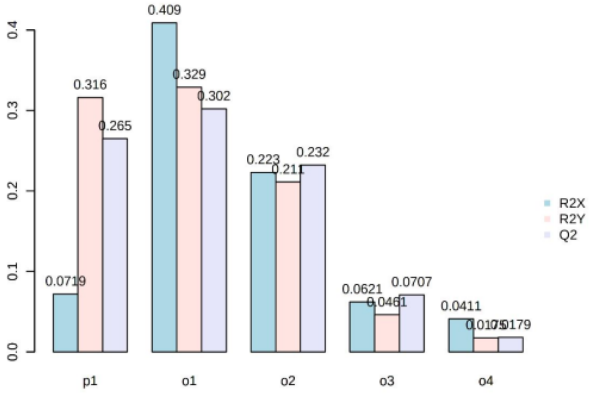


Figure S12. Overview of the OPLS-DA model built on the spectral data obtained from the extracts in DMSO-d_6 for discriminating between authentic and adulterated saffron samples.

Fig. 52- Page 23/31 of Supplementary Materials

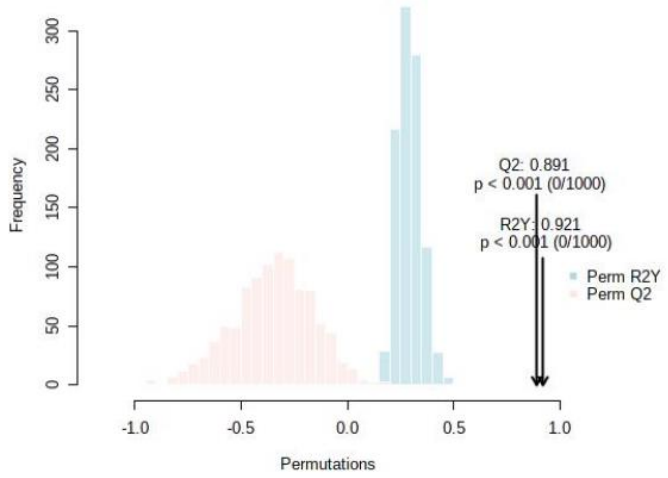


Figure S13. Permutation test applied to the OPLS-DA model built on the spectral data obtained from the extracts in DMSO-d_6 for discriminating between authentic and adulterated saffron samples. The permutation test consists of a set of 1000 permutations and testing the prediction accuracy during training. The *p*-value based on permutation is $p < 0.001$ (0/1000).

Fig. 53- Page 24/31 of Supplementary Materials

Table S5. List of features characteristic of each sample of authentic saffron under investigation.

Primary Id	Producer	Region	Town	Agronomic Practice
R1	Starace Filomena	Tuscany	GR	C
R2	Poggio al Sole	Tuscany	FI	B
R4	Starace Filomena	Tuscany	GR	C
R5	La saggezza della Terra	Veneto	PD	C
R6	Poggio al Sole	Tuscany	FI	B
R7	La saggezza della Terra	Veneto	PD	C
R8	La saggezza della Terra	Veneto	PD	C
R10	Alessandro Mazzuoli	Umbria	PG	B
R11	Poggio al Sole	Tuscany	FI	B
R12	Poggio al Sole	Tuscany	FI	B
R14	La saggezza della Terra	Veneto	PD	C
R15	Poggio al Sole	Tuscany	FI	B
R16	Poggio al Sole	Tuscany	FI	B
R17	La saggezza della Terra	Veneto	PD	C
R18	Starace Filomena	Tuscany	GR	C
R19	La saggezza della Terra	Veneto	PD	C
R21	La saggezza della Terra	Veneto	PD	C
R23	Alessandro Mazzuoli	Umbria	PG	B
R24	Poggio al Sole	Tuscany	FI	B
R26	Starace Filomena	Tuscany	GR	C
R27	La saggezza della Terra	Veneto	PD	C
R28	Starace Filomena	Tuscany	GR	C
R29	Poggio al Sole	Tuscany	FI	B
R30	Poggio al Sole	Tuscany	FI	B
R31	Poggio al Sole	Tuscany	FI	B
R33	Starace Filomena	Tuscany	GR	C
R34	Poggio al Sole	Tuscany	FI	B
R36	Starace Filomena	Tuscany	GR	C
R37	Starace Filomena	Tuscany	GR	C
R38	La saggezza della Terra	Veneto	PD	C
R39	Poggio al Sole	Tuscany	FI	B
R40	La saggezza della Terra	Veneto	PD	C
R41	Poggio al Sole	Tuscany	FI	B
R42	La saggezza della Terra	Veneto	PD	C
R43	Alessandro Mazzuoli	Umbria	PG	B
R44	Poggio al Sole	Tuscany	FI	B
R45	Poggio al Sole	Tuscany	FI	B
R46	Alessandro Mazzuoli	Umbria	PG	B
R48	La saggezza della Terra	Veneto	PD	C
R50	Poggio al Sole	Tuscany	FI	B

Fig. 54- Page 25/31 of Supplementary Materials

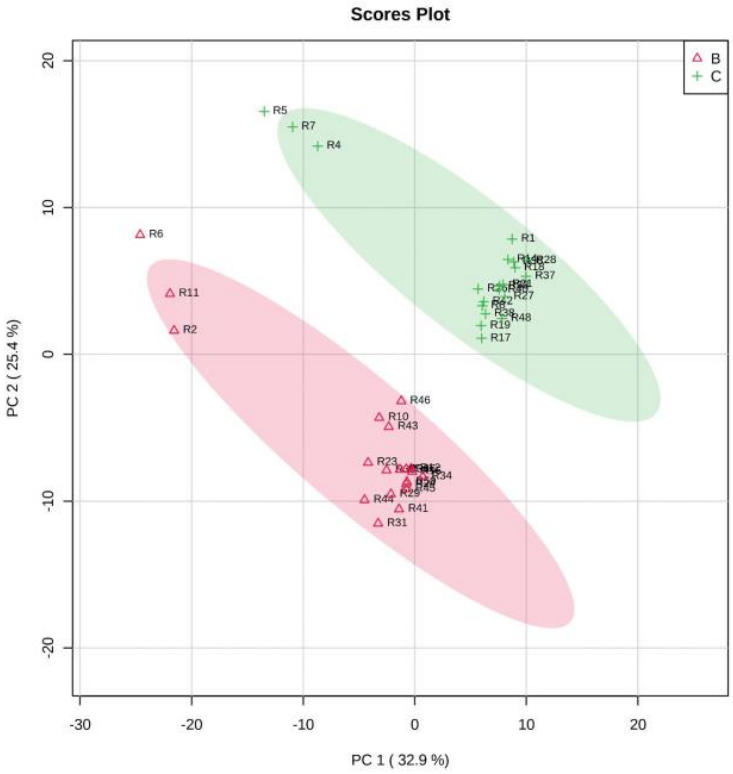


Figure S14. Results of PCA performed on the spectral data obtained from the aqueous extracts of authentic saffron. Scores plot between the selected PCs. The explained variances are shown in brackets. The samples are indicated according to the agronomic farming: conventional cultivation (C, represented as red crosses) and organic one (B, represented as red triangles).

Fig. 55- Page 26/31 of Supplementary Materials

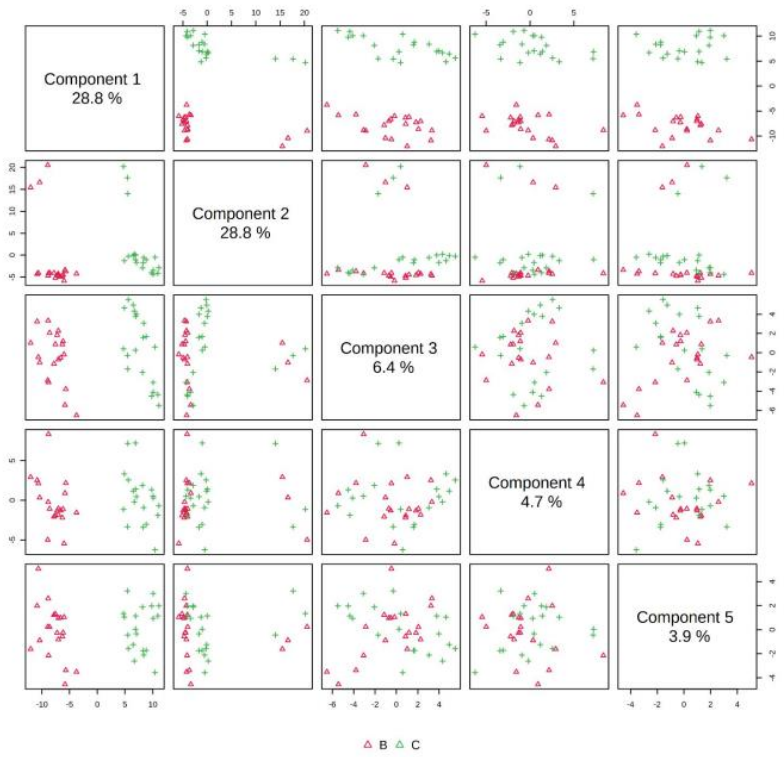


Figure S15. Overview of PLS-DA built on the spectral data of the aqueous extracts of authentic saffron samples for discriminating according to the agronomic practice: organic *vs* conventional. Pairwise scores plots between the selected components, where the scores are indicated as follows: green crosses and red triangles for authentic saffron cultivated through conventional cultivation (C), and organic one (B). The explained variance of each component is shown in the corresponding diagonal cell.

Fig. 56- Page 27/31 of Supplementary Materials

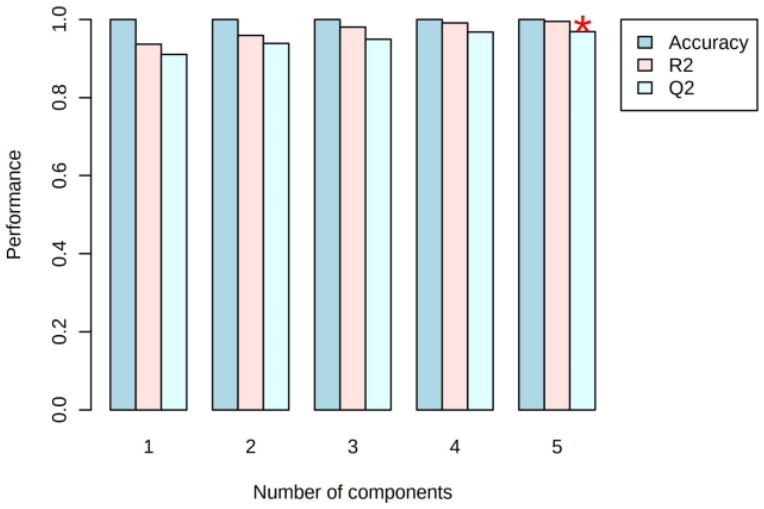


Figure S16. 10-fold cross-validation method applied to the PLS-DA model built on spectral data of the aqueous extract of authentic saffron samples for discriminating according to the agronomic practice: organic *vs* conventional. Application of 10-fold cross-validation to PLS-DA model to select the optimal number of components for classification according to the agronomic practice: organic (B) *vs* conventional (C), and estimate the predictive performance of the classification model. The red star indicates the best classifier.

Fig. 57- Page 28/31 of Supplementary Materials

Table S6. List of the top 78 Variables importance in projection (VIPs) related to the PLS-DA model built the spectral data of the aqueous extracts of authentic saffron samples for discriminating according to the agronomic practice: organic *vs* conventional. List of the top 78 variables identified as VIP > 1 on Component 1.

Range (ppm)	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
2.69272 - 2.73272	2.0506	2.0195	1.9956	1.9827	1.9794
2.33272 - 2.37272	2.0137	1.9847	1.9613	1.9483	1.9452
2.13272 - 2.17272	1.942	1.912	1.8897	1.8775	1.8745
9.93272 - 9.97272	1.9379	1.9079	1.8859	1.8734	1.8706
4.21272 - 4.25272	1.935	1.9052	1.8824	1.8783	1.8772
7.25272 - 7.29272	1.9148	1.8854	1.8637	1.8534	1.8508
1.17272 - 1.21272	1.9107	1.8811	1.8624	1.8511	1.8486
6.41272 - 6.45272	1.8745	1.846	1.8292	1.8171	1.8141
1.21272 - 1.25272	1.8437	1.8154	1.7955	1.7848	1.7819
3.69272 - 3.73272	1.8359	1.8092	1.7932	1.7814	1.7787
2.57272 - 2.61272	1.8243	1.7971	1.7765	1.7651	1.7623
6.09272 - 6.13272	1.8145	1.7864	1.774	1.7632	1.7604
5.57272 - 5.61272	1.8057	1.7805	1.7673	1.756	1.7532
4.45272 - 4.49272	1.8056	1.7777	1.7613	1.7565	1.7536
1.85272 - 1.89272	1.7912	1.7663	1.7452	1.7343	1.7315
3.97272 - 4.01272	1.7908	1.7641	1.7512	1.7395	1.7367
2.05272 - 2.09272	1.7896	1.7621	1.7471	1.7361	1.7333
3.89272 - 3.93272	1.7847	1.7605	1.7445	1.7332	1.7306
3.49272 - 3.53272	1.7847	1.7604	1.7411	1.7304	1.7277
2.37272 - 2.41272	1.773	1.748	1.7304	1.7249	1.7236
4.01272 - 4.05272	1.7586	1.7317	1.7159	1.7052	1.7028
1.49272 - 1.53272	1.7464	1.7218	1.7045	1.6931	1.6906
5.21272 - 5.25272	1.7431	1.7161	1.6957	1.6852	1.6838
3.93272 - 3.97272	1.7191	1.6969	1.6766	1.6675	1.6669
6.37272 - 6.41272	1.7076	1.6829	1.6801	1.6694	1.6669
1.45272 - 1.49272	1.6995	1.6768	1.6626	1.6525	1.6503
3.73272 - 3.77272	1.6797	1.6612	1.6457	1.6347	1.6322
4.13272 - 4.17272	1.6692	1.6437	1.6242	1.6134	1.6117
1.89272 - 1.93272	1.6315	1.6062	1.5983	1.5981	1.5962
2.17272 - 2.21272	1.6189	1.5938	1.5878	1.5773	1.5771
3.61272 - 3.65272	1.6065	1.5824	1.5791	1.5686	1.567
4.65272 - 4.69272	1.6058	1.5816	1.5632	1.5534	1.555
2.29272 - 2.33272	1.6045	1.5935	1.5744	1.5665	1.5643
2.49272 - 2.53272	1.5825	1.562	1.5491	1.5389	1.5365
4.05272 - 4.09272	1.5615	1.5389	1.5274	1.5325	1.5304
1.77272 - 1.81272	1.5509	1.5297	1.5141	1.5042	1.5017
1.65272 - 1.69272	1.5488	1.5293	1.5561	1.5476	1.5453
1.57272 - 1.61272	1.548	1.5245	1.5063	1.5023	1.4998
5.25272 - 5.29272	1.5463	1.5251	1.5074	1.4987	1.4965
6.45272 - 6.49272	1.5417	1.5199	1.5073	1.4976	1.4952
3.81272 - 3.85272	1.5348	1.5112	1.4947	1.512	1.5135
4.41272 - 4.45272	1.5127	1.4893	1.4805	1.4724	1.47
5.65272 - 5.69272	1.5123	1.4984	1.4806	1.4707	1.4685
3.41272 - 3.45272	1.4844	1.4686	1.4551	1.4486	1.4466

Fig. 58- Page 29/31 of Supplementary Materials

9.61272 - 9.65272	1.4814	1.4744	1.4589	1.4514	1.4498
4.37272 - 4.41272	1.447	1.4292	1.4137	1.4046	1.4027
8.57272 - 8.61272	1.4345	1.4284	1.4479	1.4389	1.437
10.0127 - 10.0527	1.4113	1.4053	1.3885	1.3803	1.3781
6.05272 - 6.09272	1.3924	1.3709	1.3915	1.3824	1.3802
2.45272 - 2.49272	1.388	1.3758	1.3638	1.3548	1.3526
6.33272 - 6.37272	1.3855	1.3649	1.3809	1.3741	1.3718
0.972723 - 1.01272	1.3709	1.3535	1.3549	1.3541	1.352
4.61272 - 4.65272	1.3677	1.3468	1.3335	1.3365	1.3361
1.81272 - 1.85272	1.3558	1.3447	1.3313	1.3327	1.3306
5.41272 - 5.45272	1.3541	1.3359	1.3227	1.3387	1.3417
2.25272 - 2.29272	1.3425	1.3218	1.3241	1.3286	1.3279
10.1727 - 10.2127	1.3289	1.332	1.3164	1.3095	1.3077
0.892723 - 0.932723	1.3231	1.3102	1.2958	1.2885	1.2864
1.73272 - 1.77272	1.3118	1.3043	1.3115	1.3029	1.3014
1.37272 - 1.41272	1.2619	1.2484	1.2659	1.2789	1.278
9.81272 - 9.85272	1.2616	1.243	1.2348	1.2267	1.226
2.21272 - 2.25272	1.2451	1.2261	1.2349	1.2381	1.2361
3.17272 - 3.21272	1.2302	1.2117	1.1976	1.1916	1.1968
7.89272 - 7.93272	1.2178	1.2098	1.228	1.2199	1.2187
1.01272 - 1.05272	1.2162	1.1991	1.1979	1.1952	1.1933
7.21272 - 7.25272	1.2102	1.2028	1.2157	1.2169	1.2149
0.932723 - 0.972723	1.1751	1.1589	1.1657	1.1662	1.1643
6.13272 - 6.17272	1.1588	1.1472	1.1393	1.1323	1.1307
2.61272 - 2.65272	1.1588	1.1582	1.1455	1.1388	1.1376
9.73272 - 9.77272	1.1351	1.1237	1.1183	1.1119	1.1107
10.2527 - 10.2927	1.1329	1.1406	1.1359	1.1404	1.1396
3.01272 - 3.05272	1.1176	1.1267	1.1156	1.1092	1.1082
7.61272 - 7.65272	1.1082	1.1066	1.0948	1.0884	1.0923
5.61272 - 5.65272	1.1046	1.0997	1.1098	1.1025	1.1029
9.53272 - 9.57272	1.0579	1.0416	1.0453	1.039	1.0407
7.93272 - 7.97272	1.0428	1.0272	1.0178	1.0396	1.0416
9.69272 - 9.73272	1.0144	1.0073	1.0083	1.0066	1.005

Fig. 59- Page 30/31 of Supplementary Materials

References

1. Dowlatabadi, R.; Farshidfar, F.; Zare, Z.; Pirali, M.; Rabiei, M.; Khoshayand, M.R.; Vogel, H.J. Detection of adulteration in Iranian saffron samples by 1H NMR spectroscopy and multivariate data analysis techniques. *Metabolomics* **2017**, *13*, 1–11, doi:10.1007/s11306-016-1155-x.
2. Cagliani, L.R.; Culeddu, N.; Chessa, M.; Consonni, R. NMR investigations for a quality assessment of Italian PDO saffron (Crocus sativus L.). *Food Control* **2015**, *50*, 342–348, doi:10.1016/j.foodcont.2014.09.017.
3. Cusano, E.; Consonni, R.; Petrakis, E.A.; Astraka, K.; Cagliani, L.R.; Polissiou, M.G. Integrated analytical methodology to investigate bioactive compounds in Crocus sativus L. flowers. *Phytochem. Anal.* **2018**, *29*, 476–486, doi:10.1002/pca.2753.

Fig. 60- Page 31/31 of Supplementary Materials

CHAPTER VI

2.6 Lentils

The lentil (*Lens culinaris* Medik.) is a low-fat, high-protein, and high-fibre pulse crop belonging to the legume family (Jarpa-Parra, 2018). Nowadays, lentils constitute one of the most important traditional dietary and healthy food, with potential anticarcinogenic, blood pressure-lowering, hypocholesterolemic and hypoglycaemic effects (Mo'ez Al-Islam Ezzat Faris, *et al.*, 2013). Indeed, due to unique nutritional and functional characteristics, lentils have gained much importance in the field of developing healthy and functional foods (Campos-Vega *et al.*, 2010; Roy, *et al.*, 2010; Dilis and Trichopoulou, 2009).

Lentils are characterised by high protein content, becoming a significant food source of these macronutrients for developing countries and low-income people (Hoover *et al.*, 2010). Moreover, carbohydrates also represent one of the major components of lentil seeds, mainly sucrose and α-galactosides or raffinose series of oligosaccharides (Tahir, *et al.*, 2011). These latter include raffinose, stachyose, and verbascose, sharing a common sucrose subunit.

Lentils are an important source of a plethora of minerals, especially Iron is present in significant quantities, and vitamins.

The human healthy and nutritional benefits of lentils are attributed to the bioactive phytochemicals available in this pulse. In particular, the beneficial activities of polyphenolic compounds in lentils, such as flavonoids, anthocyanins, and phenolic acid, are well documented (Singh *et al.*, 2017; *et al.*, 2002).

The lentil is an annual leguminous plant, and flowering occurs between May and July. Its typical fruit is a flattened pod, containing two seeds with a characteristic lens shape and the colour varies according to the cultivar.

To date, several lentil cultivars are found in the world, distinguished by their colour (green, yellow, red, and brown) and size (small, medium, large). The crop is best adapted to temperate climates but now lentils are produced in over 40 countries of the world, with Canada, India, Australia, Turkey, and the United States being the leading producers.

Canada accounts for the most global production of lentils and ranks as the world's largest export producer of this legume *(Kaale, et al.*, 2022).

2.6.1 Non-targeted NMR method to assess the authenticity of Italian lentils

In Italy, lentils have been cultivated since ancient times, thanks to some regions' optimal combination of climate, soil, and moisture. The growing lentil cultivation that occurred in the first half of the 20th century led the country among the major producers of these pulses (Piergiovanni, 2000).

Over the last few years, a drastic decline in lentil production has been recorded in Italy. One of the primary causes of this decrease has been the scarcity of genetic improvement activities of lentils (Sellami, *et al.*, 2021). Other reasons for the limitations to the expansion of lentil cultivation in Italy have been represented by the low and variable yield of this crop, the high production cost due to the lack of modernization of agricultural systems, and the depopulation of farmed lands (Bacchi *et al.*, 2010). As consequence, Italy has become the largest importer of lentils, mainly from Canada, the USA, Turkey, and China. Currently, in Italy, lentils are still cultivated in specific geographical localities mainly in the central and southern regions. As a result, Italian lentils cultivated within these specific marginal areas gained recognizable national trademarks, such as ‘‘protected geographical indication” (PGI), ‘‘traditional agricultural food products” (PAT) and Slow Food Presidium. Such denominations ensure the commercial values of Italian lentils, by guaranteeing their high-quality level and protecting their typicality.

Some Italian lentil cultivars considered typical products which have received designations of origin are *Altamura Lentil*, *Lentil of Castelluccio di Norcia* with Protected Geographical Indication (PGI), *Lentil of Santo Stefano di Sessanio* as Traditional Agri-food Product and Slow Food Presidium and *Colfiorito Lentil*, recognized by Traditional Food Product.

Fig. 61- Typical lentil landraces of Castelluccio di Norcia

Nevertheless, unscrupulous producers, driven by economic illicit profits, often sell Italian lentils blended or substituted with foreign ones compromising the quality and commercial value of this product.

Therefore, the development of rapid and robust analytical methods capable of establishing the authenticity of lentils and, thus, preserving the cultivation of local varieties is becoming an urgent task.

Among these methods, the non-targeted Nuclear Magnetic Resonance (NMR) approach is receiving increasing attention as a rapid tool for determining food quality and geographical origin, due to its ability to unveil specific product features that can be used by data-driven classifiers to establish the quality of the food (Ragone *et al.*, 2020; Gallo *et al.*, 2020).

In the present work, a non-targeted NMR approach was applied to assess the authenticity of Italian lentils and to establish their geographical origin, in a way to discriminate this class of product from foreign ones. The collection and analysis of authentic samples represent the starting point for developing and validating the proposed method.

For this purpose, authentic lentil samples from three different countries, i.e., Italy, Canada, and Turkey were used to build a model able to classify the lentils according to their geographical origin. The authenticity of these samples was guaranteed by the Department of Central Inspectorate for Quality Protection and Fraud Repression of Agri-Food Products (ICQRF). ICQRF, one of Europe's largest control authorities in the agri-food sector, protect consumers and producers from unfair competition in the Italian territory.

2.6.2 Materials and Methods

2.6.2.1 Samples collection

Samples of authentic lentils provided by ICQRF were used to develop a classifier model able to certify their geographical origin, in a way to discriminate the national product from the foreign one.

ICQRF inspectors carried out sampling and controls according to standard procedure EC Reg. 882/2004. The corresponding collection card was attached to each provided sample.

Lentils' samples were sent to the Chemistry Laboratory of the Polytechnic of Bari and stored at room temperature until their preparation for the NMR experiment. The pool of authenticated origin lentils taken into consideration was 93; of these 86 were authentic Italian lentils, whereas the other 7 were authentic foreign samples from Canada and Turkey. Table 2 reports the geographical origin and the harvest year of the provided lentils' samples:

Table 2. Authentic lentils' samples provided by ICQRF

Number of samples	Geographical origin	Harvest year
6	Italy	2017
6	Italy	2018
14	Italy	2019
31	Italy	2020
29	Italy	2021
4	Canada	2019
1	Turkey	2018
2	Turkey	2019

2.6.2.2 Optimization of the NMR sample preparation procedure

The lentils 'samples were analysed with an optimised procedure. Firstly, the quantity of lentil powdery sample obtained by grinding at 1 minute, 3 minutes, and 5 minutes was evaluated. Therefore, an aliquot of 50 g of a random sample of lentils was ground through a blender at the three times indicated. The resulting powdery samples were sieved by a laboratory sieve (mesh size of 0.5 mm) and weighed. The different amounts of the grounded and sieved samples are reported in table 3.

Table 3. Data obtained from milling optimisation of lentils 'samples

Grinding time	Percentage of the sieved sample
1 minute	52.50%
3 minutes	75.56 %
5 minutes	87.28 %

The largest amount of the finely ground and sieved sample was obtained by milling lentils for 5 minutes (87.28% of sieved sample).

Therefore, all the 93 lentils' samples were ground for 5 minutes, and the corresponding powder obtained was subjected to an aqueous extraction protocol.

The extraction protocol for NMR analysis of lentils was also optimised by evaluating the use of buffers with different pH values. In particular, different buffers at pH=4.2 and pH=7 were tested on a lentil sample.

An aliquot of 100 mg of ground and sieved lentil sample was dissolved in 1.5 ml volume of a buffer solution [$(HC_2O_4)^-/(C_2O_4)^{2-}$ = 0.25 M] at pH= 4.2 [pH adjusted by addition of HCl (37%)] and containing sodium azide (NaN_3) as a bacteriostatic agent (NaN_3=2.5 x 10^{-3} M).

An additional sample was prepared by adding 1.5 ml of a phosphate buffer solution at pH=7 to 100 mg of the lentil powder.

The two distinct mixtures were sonicated for 10 min at 25°C, vortexed for 10 minutes at 2500 rpm, and centrifuged for 5 minutes at 2500 rpm by using a

centrifuge filter in polytetrafluoroethylene (PTFE, with a pore size of 0.22 µm). During the centrifugation step, the two mixtures were also filtered. The filtrate was used to prepare two NMR tubes, filled with 630 µL of the obtained extract and 70 µL of a 0.2 wt% TSP-d_4/D_2O solution [3-(trimethylsilyl) propionic-2,2,3,3-d4 acid sodium salt in D_2O]. The TSP-d_4 was used as the reference internal compound.

The related 1D ^1H NOESY spectra were acquired by Bruker Avance 400 MHz spectrometer, and the metabolic profiles were examined.

Figure 62 shows the red NMR spectrum recorded on the lentils' sample solubilized in the buffer at pH=4.2; whereas the blue one is the NMR spectrum of the sample prepared with the buffer at pH=7.

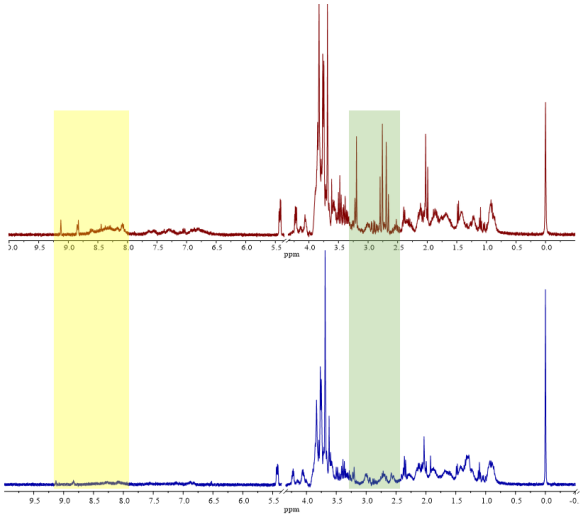


Fig. 62– 1D ^1H NOESY spectra recorded from the extracts of lentils sample at pH=4.2 (red) and pH=7 (blue).

The NMR spectra showed a variation in the chemical shift of some signals but also significant differences in the extracted metabolites.

The green and yellow rectangles delimit the spectral regions affected by these significant differences.

In particular, the green rectangle marks the citric acid signals observed in the NMR spectrum of the sample prepared with the buffer at pH=4.2 and completely

absent at pH=7. Similarly, the yellow rectangle highlights signals related to the trigonelline, a metabolite revealed in the spectral profile of the lentils' sample prepared at pH=4.2 and not at pH=7. However, other differences are also observed in the spectral region characteristic of aromatic protons [6.00, 9.00 ppm]. Considering the variations of the metabolic profile of the lentils' sample prepared with the different buffer solutions, the extraction protocol at pH=4.2 was considered as the most informative and was selected as the optimal one in the preparation procedure of lentils' samples for NMR analysis.

2.6.2.3 Extraction procedure

50 g of each sample of lentils was ground using a blender for 5 minutes and sieved using a laboratory sieve (mesh size of 0.5 mm).

An aliquot of 100 mg of the resulting powder was dissolved in 1.5 mL of the buffer solution [$(HC_2O_4)^-$/ $(C_2O_4)^{2-}$ (0.25 M) and NaN_3 (2.5 mM)] at pH 4.2, sonicated for 10 min at 25°C, and vortexed for 10 minutes at 2500 rpm. The resulting mixture was filtered through a centrifuge filter in polytetrafluoroethylene (PTFE, with a pore size of 0.22 µm). Therefore, the mixture was centrifuged for 5 minutes at 2500 rpm. The filtrate was used to fill the NMR tubes. NMR tubes were filled in by an automated system for liquid handling (SamplePro Tube, Bruker BioSpin) according to the following method: 630 µL of the obtained extract and 70 µL of the 0.2 wt% TSP-d_4/D_2O solution [3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt in D_2O].

Two replicates were prepared for each sample of lentils, obtaining 186 NMR tubes labelled with a unique barcode.

2.6.2.4 NMR experiment

Ten replicates of lentil samples were prepared according to the reported extraction procedure and were subjected to two experiments: one-dimensional ^{1}H NOESY and one-dimensional ^{1}H CPMG.

^{1}H NOESY spectra were acquired under automation at 298.1° K using a Bruker Avance 400 MHz spectrometer equipped with a 5 mm inverse probe and with an autosampler (Bruker, Billerica, MA, USA). The following acquisition parameters were used: pulse program = noesygppr1d; size of fid (TD) = 64 K; spectral width (SW) = 14 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 11.65 µs; number of scans (ns) = 128; acquisition time = 5.85 s; mixing time (d8) = 0.01 s.

The proton spectra recorded by a Carr-Purcell-Meiboom-Gill (CPMG) pulse program (cpmgpr1d) were acquired with the following acquisition parameters: pulse program = cpmgpr1d; size of fid (TD) = 64 K; spectral width (SW) = 14 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 11.65 µs; number of scans (ns) = 128; acquisition time = 5.85 s; loop for T2 filter=128.

The NMR spectra were acquired using TOPSPIN 2.1 software (Bruker BioSpin GmbH, Rheinstetten, Germany) under an automatic process that lasted ca. 22 min and encompassed sample loading, temperature stabilization for 5 min, tuning, matching, shimming, and 90◦ pulse calibration.

The spectra acquired by ^{1}H NOESY experiment are the red ones in figure 63; the blue spectra are that acquired by ^{1}H CPMG experiment.

From the spectra acquired using the two experiments, seven signals (shown in figure 63 S1-S7) were selected and the coefficient of variation percentage (CV%) of the area subtended by each chosen signal was calculated for the evaluation of the repeatability (the CV% value is indicated below the corresponding signal in figure 63).

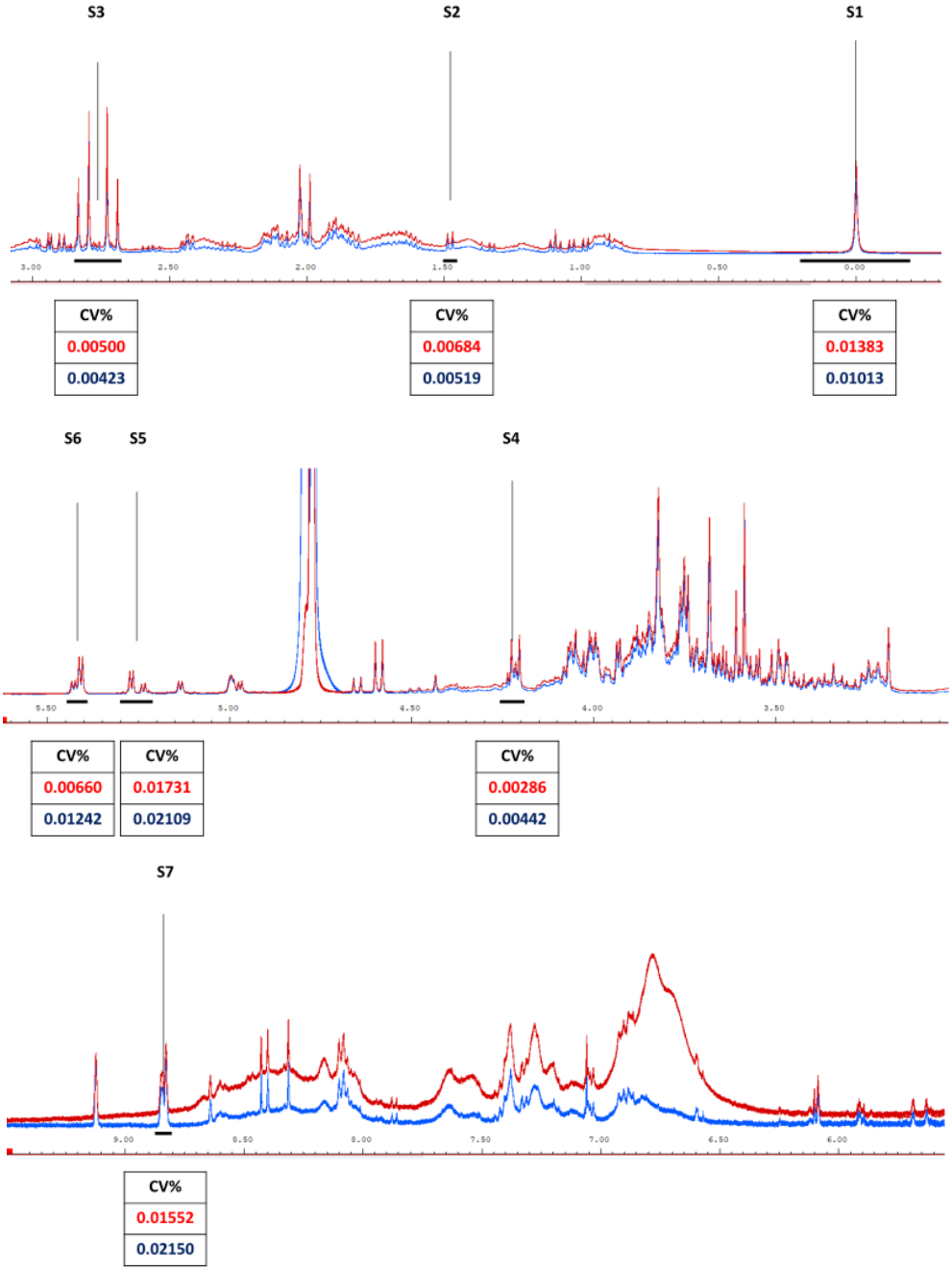


Fig. 63– Expanded regions of the spectra recorded from the same NMR tube of aqueous extract of lentils acquired by ^{1}H NOESY experiments (red) and ^{1}H CPMG experiments (blue). The signals selected for the evaluation of the repeatability are indicated.

These preliminary analyses on lentil samples showed that the ^{1}H NOESY and ^{1}H CPMG experiments guarantee the acquisition of an optimal metabolic profile, in which all signals can be observed clearly. However, with the ^{1}H NOESY experiment, a better CV% value was obtained for four signals, especially for those close to the water suppressed signal, thus proving to be more repeatable. Indeed, the ^{1}H NOESY experiment resulted in a better presaturation of the water signal as shown in figure 64.

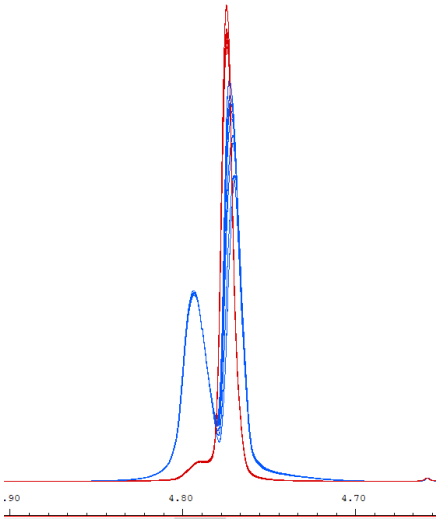


Fig. 64– The residual water signal in ^{1}H NOESY experiment (red) and ^{1}H CPMG experiment (blue).

Therefore, analyses of all lentil samples were performed by means of ^{1}H NOESY experiments. ^{1}H NOESY spectra were acquired under automation at 298.1° K using a Bruker Avance 400 MHz spectrometer equipped with a 5 mm inverse probe and with an autosampler (Bruker, Billerica, MA, USA).

The following acquisition parameters were used: pulse program = noesygppr1d; size of fid (TD) = 64 K; spectral width (SW) = 20 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; power level for pre-saturation (pl9) = 62.77 dB; dummy scans (ds) = 4; number of scans (ns) = 64; acquisition time = 4.09 s; mixing time (d8) = 0.01 s; recycle delay (d1) = 10 s. Each spectrum was acquired

using TOPSPIN 2.1 software (Bruker BioSpin GmbH, Rheinstetten, Germany) under an automatic process that lasted ca. 22 min and encompassed sample loading, temperature stabilization for 5 min, tuning, matching, shimming, and 90◦ pulse calibration. NMR raw data (Free Induction Decays, FIDs) were processed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spain).

The 1D ^{1}H NOESY spectra were automatically Fourier transformed, using an exponential multiplication function with a line broadening of 0.1 Hz, manually phased, and automatic baseline corrected (by a 1st order Polynomial Fit).

Chemical shifts were referenced to the TSP-d_4 peak at δ = 0.00 ppm.

2.6.2.5 Processing of spectral data for statistical analysis

The processed 1D ^{1}H NOESY spectra were reduced to a numerical matrix called bucket-table. The bucket table was obtained by dividing the entire overlapped spectra in the range of [10.5, -0.50] ppm into rectangular intervals (buckets) of 0.04 ppm in width, excluding the region [4.90,4.70] ppm corresponding to the residual water signal. The underlying area of each bucket was normalized to the total intensity. Thus, in the bucket table, the NMR spectra recorded from samples constituted the observation, and the buckets constituted the *x*-variables.

The bucket table generated was imported into SIMCA 17.0.2 software (Umetrics, Umea, Sweden) to perform multivariate statistical analyses (MVA).

2.6.3 Validation of the method applied

The analytical procedure applied to samples of lentils was validated, i.e., repeatability, and intermediate precision. A total of 10 NMR tubes were prepared from the identical lentils' sample, and the respective spectra were acquired, according to the procedure described in paragraphs 2.6.2.3 and 2.6.2.4.

In this way, the variability induced by the operator's skills and the degree of sample homogeneity were evaluated (intermediate precision).

In addition, 10 NMR spectra were acquired for the same NMR tube to measure the variability induced by the instrument (repeatability). For this purpose, the NMR tube was pulled out and inserted ten times into the instrument.
Three signals were opportunely chosen for the evaluation of repeatability, and the integrals of their subtended area were determined. The three selected signals were the doublet of the alanine (1.48 ppm), the signal of citric acid (2.76 ppm), and the singlet of the trigonelline (9.12 ppm). These signals were denoted S1, S2, and S3 respectively. The integrals range considered for the selected signals were: [1.50,1.45 ppm] for S1 (alanine); [2.86,2.66 ppm] for S2 (citric acid), and [9.13,9.10 ppm] for S3 (trigonelline).
From the resulting integral values, the coefficient of variation percentage (CV%) was calculated. The CV% [(μ/σ) *100], given by the ratio of the mean (μ) to the standard deviation (σ), is an index of the dispersion of the data; thus, CV% is associated with the measurement of the uncertainty of the procedure applied.
Therefore, the CV% calculated for the selected signals was used to indicate the large and close repeatability.
In addition, the stability of the investigated matrix was studied to assess any changes over time in the metabolic profile.

2.6.3.1 Evaluation study of repeatability

A single NMR tube of aqueous extract of lentils was analysed ten times for the evaluation of repeatability. This NMR tube was removed and reinserted ten times inside the spectrometer, in an automated way, and ten 1D ^{1}H NOESY spectra were recorded. The acquisition parameters of NMR spectra are indicated in paragraph 2.6.2.4.
The acquired ten 1D ^{1}H NOESY spectra were processed by the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spagna). The processing consisted of Fourier transformation, manual phase, automatic baseline correction, and alignment of the spectra to the signal of TSP-d_4 at 0 ppm. Figure

65 shows the 10 spectra recorded from the same NMR tube of aqueous extract of lentils, chosen for the evaluation of the repeatability.

The spectral areas selected for the calculation of integrals are highlighted in a yellow rectangle for S1 (1.48 ppm), and green and blue ones for S2 (2.76 ppm), and S3 (9.12 ppm) respectively.

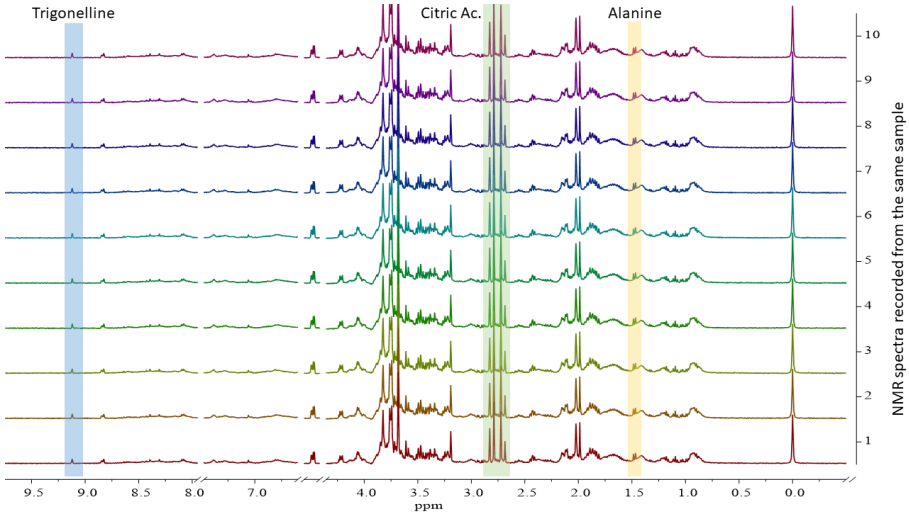


Fig. 65– 1D ^{1}H NOESY spectra recorded from the same NMR tube of aqueous extract of lentils prepared for the repeatability evaluation.

The integration of the selected signals was carried out using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spagna). The integrals obtained were expressed in absolute values. From these values, the mean (μ), the standard deviation (σ) and, the coefficient of variation percentage (CV%) were calculated. The data obtained from those calculations are collected in table 4.

Table 4. Results of the repeatability study of lentils

	S1	S2	S3
Integration range (ppm)	**[1.50,1.45]**	**[2.86,2.66]**	**[9.13,9.10]**
1D ^{1}H NOESY spectra	**Integral values**		
1	10097.1	61841.4	2042.8
2	10318.2	64199.8	1979.0
3	9904.2	61736.1	2006.9
4	10156.8	63426.4	1869.2
5	10195.4	63555.7	2007.4
6	10272	63587.7	2026.4
7	9794.5	60893.8	1978.0
8	10211.7	63519.6	1991.6
9	10362	64290.7	2042.8
10	10242.3	63670.6	2005.2
Standard deviation	180	1153.9	49.8
Mean	10155.4	63072.1	1994.9
CV%	**1.7**	**1.8**	**2.5**

The calculated CV% values for the evaluation of repeatability (1.7, 1.8, and 2.5 for alanine, citric acid, and trigonelline signals, respectively), were found lower than 5%. Thus, the NMR procedure applied for the analysis of lentils' samples was not affected by significant errors occurring because of the spectrometer and/or the operator's skills.

2.6.3.2 Evaluation study of intermediate precision

From the same lentils' sample, 10 distinct aqueous extracts were prepared, according to the protocol described in paragraph 2.6.2.3. Each aqueous extract was used to prepare the relative NMR tube.

The 10 NMR tubes were filled in by an automated system for liquid handling (SamplePro Tube, Bruker BioSpin) using these volumes: 630 µL of the filtrate and 70 µL of TSP-d_4/D_2O solution [3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt in D_2O (0.2% w/w)]. The 10 NMR tubes were submitted to NMR analysis

(Bruker Avance 400 MHz), according to the acquisition parameters indicated in paragraph 2.6.2.4. The acquired 1D ^{1}H NOESY spectra were processed by the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spagna).

The processing consisted of Fourier transformation, manual phase, automatic baseline correction, and alignment of the spectra to the signal of TSP-d_4 at 0 ppm. The spectra recorded from the 10 NMR tubes are illustrated in figure 66. The signal of residual water (4.78 ppm) was hidden.

The yellow rectangle highlights the spectral area of S1 (1.48 ppm); the green and the blue ones delimitate the spectral area of S2 (2.76 ppm), and S3 (9.12 ppm) respectively. These areas were selected for the calculation of integrals.

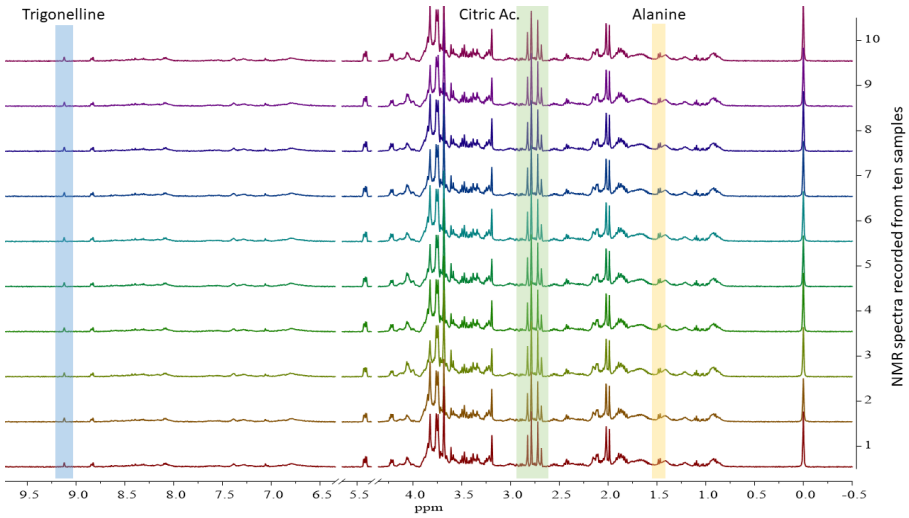


Fig. 66– 1D ^{1}H NOESY spectra recorded from the ten NMR tubes of aqueous extracts of lentils prepared for the intermediate precision evaluation.

The integration of the areas subtended by S1 (yellow rectangle in figure 30), S2 (green rectangle in figure 30), and S3 (blue rectangle in figure 30), were performed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spagna).

The integrals were expressed in absolute values. From these values, the mean (μ), the standard deviation (σ) and, the coefficient of variation percentage (CV%) were calculated. The data obtained from those calculations are collected in table 5.

Table 5. Results of the intermediate precision study of lentils

	S1	**S2**	**S3**
Integration range (ppm)	**[1.50,1.45]**	**[2.86,2.66]**	**[9.13,9.10]**
NMR tube	**Integral values**		
1	10763.3	61240	2152.7
2	10807.3	59099	2035.8
3	11750.3	67557	2140.2
4	10782.9	60126.9	2098.2
5	10782.9	63050.8	2093.4
6	11506.3	63403.3	2096.2
7	10761.2	60490.6	2092.8
8	10950.1	60338.8	2232.1
9	10860.1	63734.9	2158.2
10	10821	59695.7	2038.3
Standard deviation	351.7	2574.5	59
Mean	10978.5	61873.7	2113.8
CV%	**3.2**	**4.2**	**2.7**

The CV% values obtained in the study of intermediate precision (3.2 for alanine, 4.2 for citric acid, and 2.7 for trigonelline) were below 5% and were considered to be acceptable as an error including sample preparation and sample analysis.

2.6.3.3 Study of the matrix stability over time

A NMR tube of aqueous extract of lentils was used to check the stability of the investigated matrix over time. Therefore, the 1D ^{1}H NOESY spectrum was recorded on the freshly prepared sample; then, the NMR tube was left at room temperature. NMR spectra were acquired after 48 hours and, one week later from

its preparation, according to the experimental parameters reported in paragraph 2.6.2.4.

Figure 67 shows the recorded NMR spectra on the same sample as follows: the red spectrum is related to the freshly NMR tube prepared (0 h); the green one corresponds to the NMR spectrum recorded after 48 hours; finally, the blue spectrum represents the NMR spectrum acquired after one week.

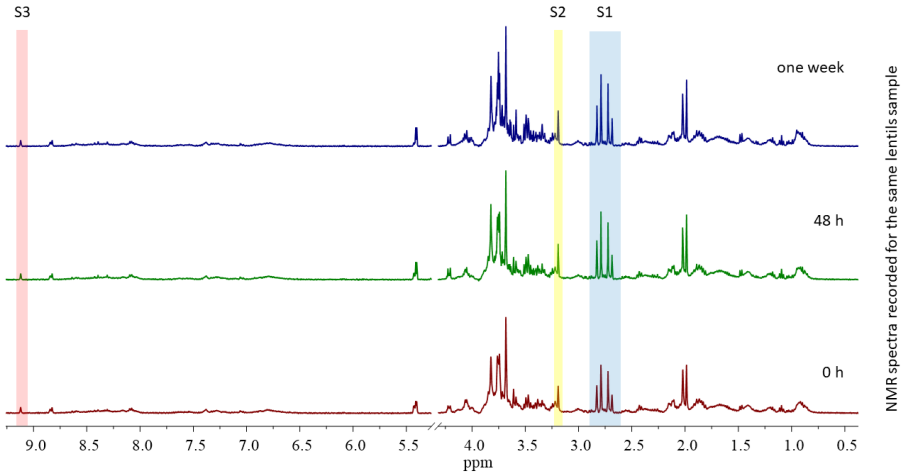


Fig. 67– 1D ^{1}H NOESY spectra of the same lentils sample recorded at different time points for the stability study.

The stability of the lentils' sample was assessed by integrating the areas of three signals (S1, S2, and S3 in fig. 5), selected in the NMR spectra recorded at the time points indicated. The signal S1 was related to citric acid, S2 to the singlet of choline, and S3 to the trigonelline. The integrals of the areas subtending the signal were calculated in the following regions: [2.88,2.66 ppm] for citric acid (S1); [3.21,3.18 ppm] for choline (S2); [9.13,9.10 ppm] for trigonelline (S3).

The integration procedure was performed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spagna). The integrals were expressed in absolute values.

Then, the following statistical parameters were calculated: mean (μ), standard deviation (σ) and, the coefficient of variation percentage (CV%).

The variation percentage coefficient was considered an index to validate the measurements of stability assessment.

The data obtained from those calculations are reported in table 6.

Table 6. Results of the stability of lentils' sample

	S1	S2	S3
Integration range (ppm)	**[2.86,2.66]**	**[3.21,3.18]**	**[9.13,9.10]**
Time points	**Integral values**		
0 h	61418	11022.4	2525.7
48 h	63823.4	11955.5	2696.4
1 week	66390.7	12002.7	2771.2
Standard deviation	2486.7	552.8	125.8
Mean	63877.3	11660.2	2664.4
CV%	**3.8**	**4.7**	**4.7**

The calculated integrals showed a moderate increase in the area subtended by the three selected signals over time.

To illustrate this variation, the expansions of the signals of citric acid (S1), choline (S2), and trigonelline (S3) in the three recorded NMR spectra are shown in figure 68. In particular, the intensity of the selected signals increased the NMR spectrum recorded after one week (blue in fig. 68), compared to the NMR spectrum recorded on the freshly NMR tube prepared (red in fig. 68).

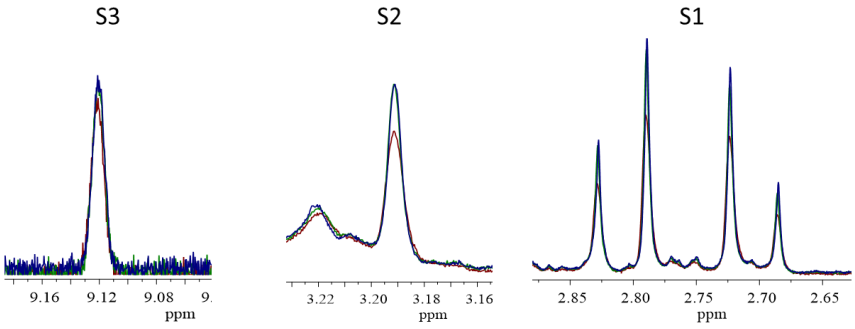


Fig. 68– The expanded signals were chosen for the stability study of the lentils' sample

Despite a moderate increase in the integrals calculated for the selected signals, the CV% values were less than 5% (3.8 for S1, 4.7 for S2, and 4.7 for S3). Thus, the lentils' sample suffered no significant changes in its metabolic profile over one week. However, it was considered advisable to conduct the NMR analysis of the lentils' samples immediately after their extraction procedure, to avoid any possible conversion processes, especially involving the carbohydrates.
The characteristic metabolic profile of the investigated matrix is described in the following paragraph.

2.6.4 Metabolites characterization of Italian authentic lentil

The characteristic metabolites contained in lentils were identified by exploring literature data and comparing a library of 1D ^{1}H NOESY spectra of aqueous extract (pH=4.2) of standard compounds.
The main metabolites found in authentic Italian lentils, listed in table 7, belong to the class of amino acids, organic acids, and carbohydrates. Moreover, signals related to choline, and trigonelline were identified.

Table 7. Characteristic metabolites of aqueous extract of authentic Italian lentils

Compound	**Chemical shift (ppm)**	**Multiplicity: J (Hz)**
Amino acids		
Isoleucine (1)	0.92	t (7.4)
	0.99	d (7)
	1.24	m
	1.44	m
Valine (2)	0.98	d (7.0)
	1.04	d (7.0)
	2.27	m
Alanine (3)	1.47	d (7.3)
	3.76	q overlapped
Arginine (4)	1.69	m
	1.91	m
	3.25	m overlapped

	3.78	m overlapped
	6.78	bs
	7.29	bs
Glutamic acid (5)	2.10	m
	2.41	m
	3.76	dd overlapped
Asparagine (6)	2.87	dd (16.8; 7.3)
	2.95	dd (16.9; 4.4)
Lysine (7)	1.72	m
	1.90	m
	3.03	t (7.5)
Phenylalanine (8)	3.13	dd (14.4; 7.7)
	7.33	m
	7.42	m
Tyrosine (9)	6.90	d (8.5)
	7.19	d (8.5)
Organic acids		
Acetic acid (10)	2.03	s
Citric acid (11)	2.67	d (15.2)
	2.81	d (15.2)
Formic acid (12)	8.45	s
Carbohydrates		
Sucrose (13)	3.48	t (9.2)
	3.69	s
	3.75	m
	3.83	m
	4.06	t (8.4)
	4.21	d (8.7)
	5.41	d (3.9)
RFOs (14)	4.23	d (8.8)
	5.43	d (3.7)
Quaternary ammonium compounds		
Choline (15)	3.20	s

Alkaloids		
	8.08	m
Trigonelline (16)	8.83	m
	9.12	s

In figure 69 a typical 1D ^{1}H NOESY spectrum of an aqueous extract of authentic Italian lentils is reported showing the assignment of the identified metabolites signals.

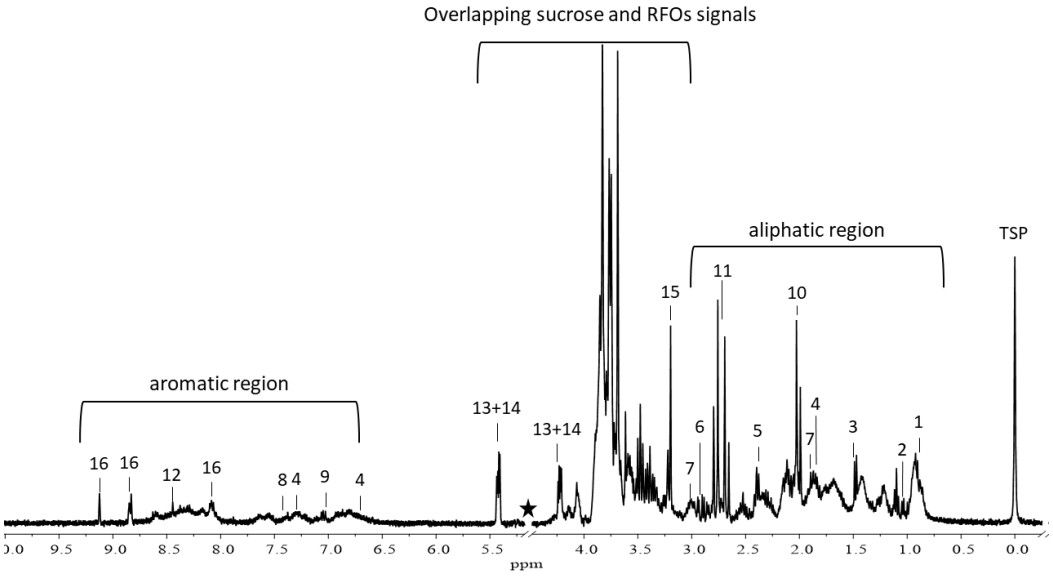


Fig. 69– Typical 1D ^{1}H NOESY spectrum of an aqueous extract of authentic Italian lentils. The HDO signal is hidden; numbers indicate the metabolites reported in table 7.

In the aliphatic region of the ^{1}H NMR spectrum [0.8-3.00 ppm], the signals mainly identified related to isoleucine, valine, alanine, arginine, glutamic acid, asparagine, lysine, and acetic, and citric acids.

In the central area of the spectrum [3.00-6.00 ppm], the intense overlapping signals of carbohydrates were observed, principally sucrose, along with its α-1,6-galactosyl extensions, namely raffinose family oligosaccharides (RFOs). These assignments were in agreement with the literature data (Longobardi *et al.*, 2017).

L. culinaris is known to contain large amounts of the raffinose family oligosaccharides, including raffinose, stachyose, and verbascose (Gulewicz *et al.*, 2000). Moreover, the singlet attributed to choline was detected.

The aromatic region [6.00-9.00 ppm] displays the typical signal of the amino acids phenylalanine, and tyrosine, along with the signal of the fumaric acid and the ones referred to trigonelline, as reported in the literature (Longobardi *et al.*, 2017). The resonances assignments were confirmed by bidimensional experiment (^{1}H COSY). The ^{1}H COSY spectrum acquired from an aqueous extract of lentils is shown in figure 70. The signals related to isoleucine, valine arginine, asparagine, sucrose, phenylalanine and trigonelline, which are included in the main metabolites identified in lentil's samples, are indicated in the figure 70.

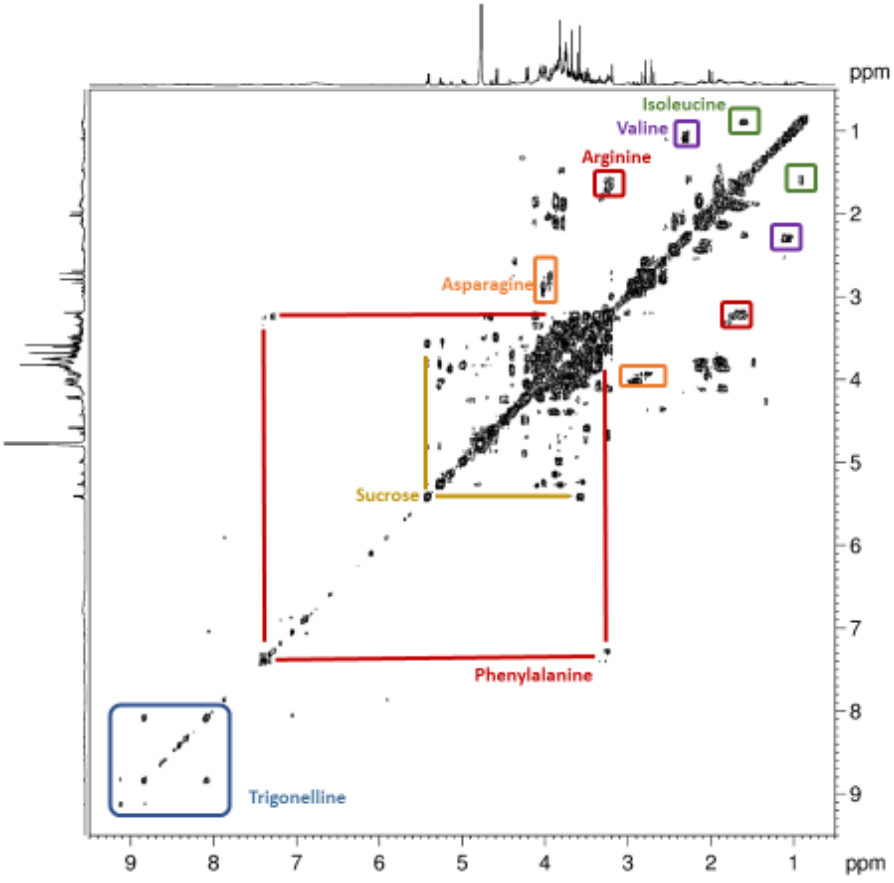


Fig. 70–^{1}H COSY spectrum of aqueous extracts of lentils

2.6.5 Results

2.6.5.1 Principal Component Analysis of NMR data

The 186 processed 1D 1H NOESY spectra (93 samples x 2 replicates) recorded on authentic lentils' samples were reduced to a bucket-table, according to the procedure explained in paragraph 2.6.2.5. The obtained data matrix was constituted of 186 observations (samples) and 244 x-variables (spectra buckets). The bucket-table was imported into SIMCA 17.0.2 software (Umetris, Umea, Sweden) to carry out multivariate statistical analyses. Buckets were subjected to Unit Variance scaling (mean-centred and divided by the standard deviation of each variable) to avoid noise inflation.

First, an unsupervised method, the Principal Component Analysis (PCA) was applied for obtaining an overview of the NMR data. The quality of the PCA model was evaluated based on the R2 (goodness-of-ft) and Q2 (goodness-of-prediction) parameters. Such values obtained were R2=0.824 and Q2=0.729.

Figure 68 shows the scores plot t[1]/t[2] relating to the first two principal components (PCs). No trend of NMR data of authentic lentils'samples according to their geographical origin was illustrated in PCA scores plot (figure 71a). Moreover, no significant separation based on geographical origin was observed in the PCA scores plot even at higher components.

However, considering the same analysis and colouring the NMR data according to the harvest year, two groups of observations were observed in the PCA scores plot t[1]/t[2] (fig. 71b).

In particular, the lentils' samples harvested in 2021 (light blue triangles in fig. 71b) were contained in a separate cluster from those collected in previous years (2020,2019,2018, and 2017 in fig. 71b). This separation was clearly visible along the first component (PC1), which explained 43% of the variance (R^2X [1] =0.43).

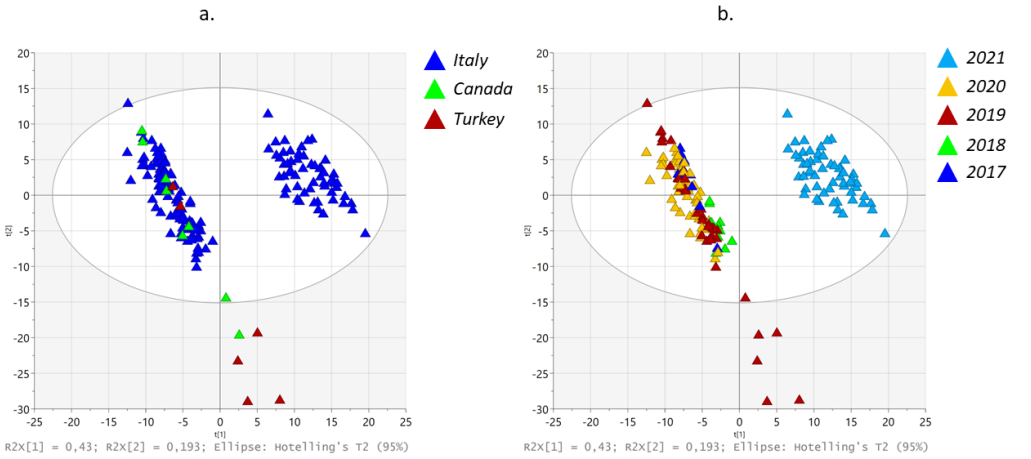


Fig. 71– Scores plot t[1]/t[2] obtained from the PCA applied to NMR data of authentic lentils' samples. a) The observations are coloured according to the geographical origin. b) The observations are coloured according to the harvest year.

Two factors were taken into account to explain the observed behaviour, namely the storage stability over the time and the influence of the agroclimatic conditions.

In the literature it was reported that the factor "year" had a influence on the metabolic composition of lentils, with a particular reference to their polyphenol content, leading to a clustering of the lentils grown in the years 2019 and 2020 (Irakli et al., 2021). The reported differences may be ascribed to the different agroclimatic conditions, such as temperature and amount of rainfall before harvest (Irakli et al., 2021).

Therefore, to obtain more information about the possible effect of the agroclimatic conditions which can vary from one year to another one, a study was conducted on samples of lentils harvested in 2020 and 2021.

Analysing the PCA scores plot (fig. 72a) related to the NMR data of lentils harvested in 2020 (blue circles in fig. 72a) and 2021 (green circles in fig.72a), a clear separation of the observations between the different years was observed. The two groupings were visible along the PC1, which explained 31% of the variance (R2X [1] =0.31).

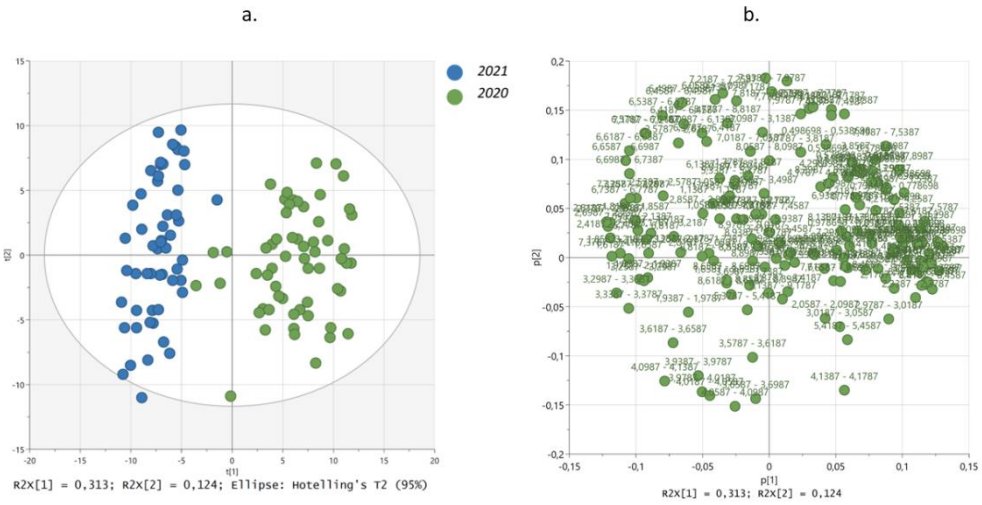


Fig. 72– a) Scores plot t[1]/t[2] obtained from the PCA applied to NMR data of authentic lentils' samples harvested in the years 2020 and 2021. The observations are coloured according to the harvest year.b) Loadings plot for the first and second principal components t[1]/ t[2]

The bucket at [7.05-7.09 ppm] was identified as one of the main variables (bucket) which most contributed to the observed separation of lentils' samples. In particular, the signal at 7.06 ppm was probably attributable to gallic acid, which is reported to be one of the main phenolic acids found in lentils (Ganesan and Xu, 2017). Gallic acid, a polyphenol natural product, is known to have anti-oxidant, anti-inflammatory, anti-allergic, anti-microbial, and radical scavenging activities (Giannakoula et al., 2012). This potent antioxidant even at a lower concentration may contribute significantly to the antioxidant activity of lentils (Alshikh, et al., 2015). Analogously to other polyphenols, gallic acid can undergo to chemical transformations over the time and, thus, represent a marker for the freshness of the lentil product.

Therefore, the signal assigned to gallic acid has been monitored over time to also investigate the storage aspect of lentils. Therefore, a lentils' sample of the 2020 was analysed in March 2021 (red spectrum in fig. 73). The same sample, stored at room temperature as grounded and sieved powder, was analysed in July 2022 (green spectrum in fig 73). The two 1D ^{1}H NOESY spectra showed the intense

signal of gallic acid at 7.06 ppm (highlighted in the yellow rectangle in fig. 73); thus, the metabolite remained stable after more than one year. Data reported in the literature have demonstrated the stability of the gallic acid in grape stem under different conditions of light and temperature and throughout six months of storage of the analysed samples (Esparza et al., 2020).

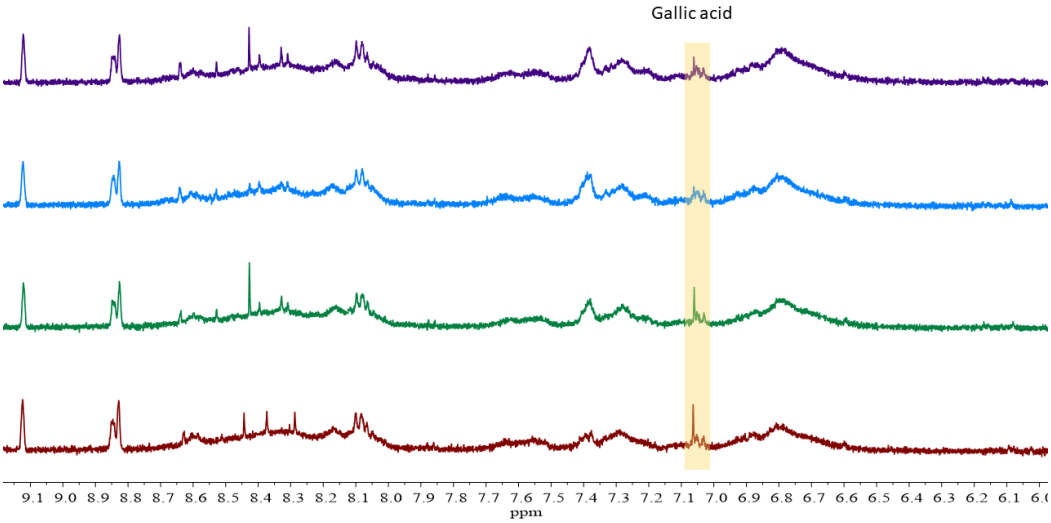


Fig. 73–1D ^{1}H NOESY spectra of a lentils' sample harvested in 2020 and analysed in March 2021 (red), and July 2022 (green). 1D ^{1}H NOESY spectra of a lentils' sample harvested in 2021 and analysed in September 2021 (blue), and July 2022 (violet). The gallic acid signal is highlighted in the yellow rectangle.

A lentils' sample of the 2021 was analysed soon after harvesting and the relative 1D ^{1}H NOESY spectrum (blue in fig. 73) showed no gallic acid signal. The same sample, after ten months of storage at room temperature as grounded and sieved powder, was analysed in July 2022 and the acquired NMR spectra (violet in fig.70) showed the signal at 7.06 ppm.

One possible explanation of the observed behaviour can be ascribed to the formation of gallic acid from some natural derivatives which are contained in the lentils, including gallocatechin, catechin gallate, epicatechin gallate (Ganesan and Xu, 2017), digallate procyanidin dimer, and procyanidin gallate. The last ones

were previously identified by HPLC-PAD and HPLC-ESI-MS in a red lentil acetonic extract (Amarowicz *et al.*, 2009). Likely, once gallic acid is formed, it remains reasonably stable over time. However, the formation of gallic acid a few months after harvesting can constitute a biomarker for evaluating the freshness of the analysed lentils.

However, based on the current findings, it cannot be established which of the two factors (harvest year and storage condition of the samples) may have most affected the distribution of authentic lentils' samples observed in the PCA analysis (fig. 71b).

2.6.5.2 Development of a classifier to ascertain the geographical origin of lentils

The described work aimed to develop a reliable classifier to establish the geographical origin of lentils' samples.

To discriminate authentic lentils' samples according to their geographical origin, the supervised Orthogonal Partial Least Squares discriminant analysis (OPLS-DA) was applied to the previous NMR data. In the OPLS-DA analysis, the differentiation of assigned classes of interest (in this case Italy, Canada, and Turkey classes) is stressed. For OPLS-DA, the quality of the model was evaluated considering the parameters R^2, and Q^2, and permutation tests were performed to avoid data overfitting.

The NMR spectra acquired from the authentic lentils' samples were used as the *training set*, to "train" and optimise the developed discriminant model to identify the classes of the lentils' samples.

The obtained OPLS-DA model presented two predictive components (P), and five orthogonal ones (O), with R^2= 0.735 and Q^2= 0.501.

The parameters characterizing the OPLS-DA model built by using the NMR data are detailed in table 8.

Table 8. The parameters of the OPLS-DA analysis performed on the spectral data of the authentic lentils' samples

Component	R^2X	R^2X(cum)	R^2	R^2(cum)	Q^2	Q^2(cum)	R^2Y	R^2Y(cum)
Model		**0.64**		**0.735**		**0.501**		**1**
Predictive		0.0855		0.735		0.501		1
P1	0.0624	0.0624	0.509	0.509	0.392	0.392	0.632	0.632
P2	0.0231	0.0855	0.226	0.735	0.109	0.501	0.368	1
Orthogonal in X (OPLS)		0.554						
O1	0.359	0.359						
O2	0.0977	0.456						
O3	0.0495	0.506						
O4	0.0279	0.534						
O5	0.0202	0.554						

In figure 74, the P1 versus P2 (t[1] *vs.* t[2]) scores plot shows the clear separation of the observations, according to the investigated classes: Italy (blue triangles in fig.74), Canada (green triangles in fig.74) and Turkey (red triangles in fig.74).

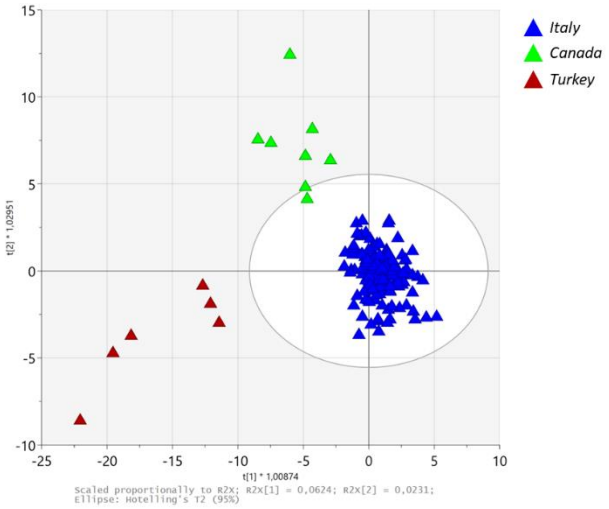


Fig. 74– Scores plot (t[1]/t[2]) of the OPLS-DA analysis performed on spectral data of authentic lentils' samples

The robustness of the OPLS-DA model was ascertained by validation with a 100-permutation test (y-intercept values for R^2 and Q^2 were acceptable for all classes). The permutation tests performed are shown in figure 75.

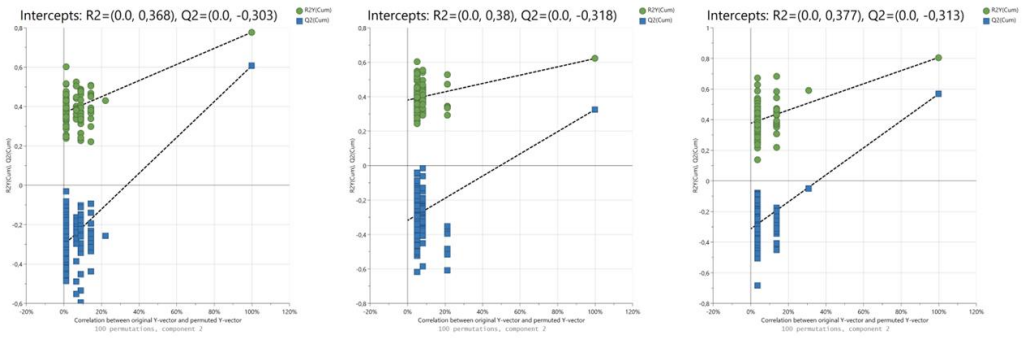


Fig. 75– Permutation test applied to the OPLS-DA model built on the spectral data obtained from authentic lentils' samples. The values of the R^2 and Q^2 coefficients are indicated.

2.6.5.3 Validation of the developed classifier

The OPLS-DA model (fig. 74) developed using NMR data of lentils' samples of certain origin was validated. Therefore, the capability of this model to predict and classify the geographical origin of 31 lentils' samples not supplied by ICQRF (referred to as "commercial" lentils) was evaluated. The geographical origin of these latter ones was unknown. Hence, the spectral data recorded from these lentil's samples were included in the statistical analysis. These new observations constituted the *test set*, used to test the performance of the developed model, and thus establish its predictive ability. The results of the conducted prediction analysis are shown in the misclassification table (table 9).

In particular, the authentic Italian and Turkish lentils' samples, included in the training set, were classified in the correct class at 100%. The authentic Canadian lentils' samples (also part of the training set) were classified in the "Canada" class with a correctness of 87.5%; the lower value obtained may be due to the reduced number of foreign lentils' samples provided.

The results of the prediction of the test set samples showed that 20 samples of lentils were classified as being of Italian origin, and 11 were classified as Canadian ones. The developed model was capable to predict the geographical origin of the lentil's samples 99.41 % correctly (red value in table 9), proving to give high prediction ability and reliability.

Table 9. Misclassification table for training and test samples

	Members	Correct	"Italy" class	"Canada" class	"Turkey" class
Authentic Italian lentils	161	100%	161	0	0
Authentic Canadian lentils	8	87.5%	1	7	0
Authentic Turkish lentils	6	100%	0	0	6
Test set	62		39	23	
Total	231	**99.41%**	195	30	6
Fisher's prob.	5,1e-22				

The results reported in the table 9 showed that 3 samples of authentic Canadian lentils were classified correctly, but for the fourth sample, one replicate was classified as Canadian while the other replicate was classified as Italian. This different classification could be explained by the data reported in the *classification list*, that is complementary to the misclassification table. However, the misclassification table has tougher rules for class-assignment compared with the classification list. In the classification list, when the obtained probability value is close to 1 and it is indicated on a green background, it means that the observation is considered as a member of that class. When the probability value is below 0.5, it is displayed on a yellow background and the observation's membership in a particular class is questionable.

When the probability value is zero or less, the observation does not belong to that class and appears on a white background.

To illustrate this point, the classification list scores of a correctly classified Canadian sample and the one with the different classification of its two replicates are shown in table 10.

Table 10. A section of the classification list of the predicted observations

NMR ID	Class	Prediction Class ITALY	Prediction Class CANADA	Prediction Class TURKEY
L3740	Canada	0.477321	0.535723	-0.0130445
L3741	Canada	0.0356654	1.00863	-0.0442933
L3742	Canada	0.156002	0.692011	0.151988
L3743	Canada	0.0886753	0.723695	0.18763

The two replicates of the same sample of lentils of Canadian origin with a different classification were marked by the dashed lines in table 10. One replicate received a score of 1 and it was classified correctly as Canadian (green in table 10); the other replicate (yellow in table 10) received a probability value lower than 1, but similar between Italy and Canada classes. Therefore, this replicate was not correctly assigned at the Canada class and was classified as Italian in the misclassification table (table 9). The different classification of the two replicates of the same sample could be explained by the similar probability score obtained between the Canada and Italy classes (0.53 and 0.47, respectively), likely related to the limited number of Canadian lentils included in the statistical model or even the sample preparation.

It was also possible to extend the same explanation to one of the samples included in the test set. One replicate was classified as Italian (green in table 11) and the other one was classified uncertainly between Italy and Canada, since it obtained probability values around 0.5 (yellow in table 11).

Table 11. A section of the classification list of the predicted observations

NMR ID	Class	Prediction Class ITALY	Prediction Class CANADA	Prediction Class TURKEY
L3773		0.605439	0.456337	-0.0617758
L3774		0.655283	0.343963	0.000753701

The results obtained in this work demonstrate the potential and the beneficial application of non-targeted NMR in assessing and preserving the authenticity of Italian lentils. Future developments concern the possibility of analysing balanced pool of samples by including more lentils of Canadian and Turkish origin. In this way, the reliability and robustness of the developed model would improve.

Although a more balanced data would improve the robustness of the developed model, the analysis of the VIPs (Variable Importance for the Projection), along with the visual inspection of the spectra, allowed for the identification of the main metabolites responsible for the separation of the investigated classes (Italy, Canada, and Turkey). In particular, the Canadian and Turkish lentils samples (green and red spectra, respectively, in fig. 75) presented a higher content of some amino acids, like valine, alanine and phenylalanine, as displayed in figure 75. While the Italian lentils samples (blue spectrum in fig. 75) showed a more intense signal of citric acid.

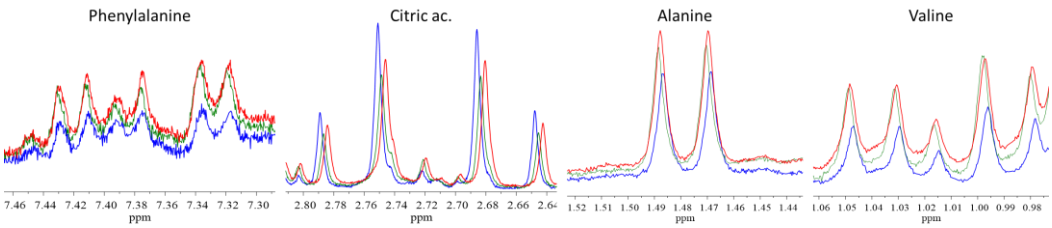


Fig. 75– Main spectral regions of authentic Italian lentils (blue spectrum), Canadian lentils (green spectrum), and Turkish lentils (red spectrum) responsible for the separation of the classes according to the geographical origin.

CHAPTER VII

2.7.1 Wheat

Wheat is one of the world's oldest and most widely crops, and it is likely to remain one of the mainstays of human nutrition for the foreseeable future (Le Gouis, *et al*., 2020). Wheat (*Triticum* spp.) is a versatile cereal for the preparation of different foods, originally from the Near East's Fertile Crescent (more than 10,000 years ago), and it is now the most widely cultivated crop in the world (Matsuoka, 2011).

Worldwide wheat production is about 765 million tons annually, depending on agro-climatic conditions, with an average yield of 3.5 t ha^{-1}. Asia contributes the most to global wheat production, followed by Europe and America. China is the first world-producing country, while in Europe, the biggest producers of wheat are (in order) France, Germany, the UK, Italy, Romania, Spain, and Hungary.

In America, the largest producers are Canada, the United States of America and Argentina (FAOSTAT 2019).

Wheat plays a crucial role in global food security. This cereal supplies most of the daily calories and proteins for human nutrition, especially for people living in developing countries, who derive a significant proportion of their nutrient needs from cereal-based foods (Shiferaw *et al.*, 2013). In particular, it is estimated to provide up to 45% of the calories and proteins for daily human consumption in North Africa and Central Asia (Peña-Bautista, *et al*., 2017).

In addition, wheat is also a source of essential amino acids, minerals and vitamins, beneficial phytochemicals, and dietary fibre components in the human diet (Shewry and Hey, 2015). It can be claimed that wheat has become the most traded crop in the international agricultural food market (Pequeno *et al.*, 2021).

2.7.1.1 Modern wheat species

Modern cultivated wheat originates through an evolutionary process in which hybridizations, polyploidizations, domestication, and mutations occurred.

Over the centuries, these selection processes have contributed to several changes in the genomic structure of wheat (Rahman et al., 2020).

Currently, the modern species that dominate wheat corps production are *Triticum aestivum* and *Triticum durum*, are two closely related species characterized by different grain hardness or texture, which confer their distinct technological properties (Bhave and Morris, 2008).

Indeed, the two wheat species are traditionally used for different end-use products.

Triticum aestivum L. var. *aestivum* is the "common "or "bread" wheat, with a hexaploidy genome (AABBDD, 6n= 42). The soft texture is well suited to produce bread and biscuits, typical food end-products (Pauly *et al.*, 2013).

Triticum turgidum L. subsp. *durum (Desf.),* called durum wheat, has a tetraploid genome (AABB, 4n= 28).

The very hard kernel texture, together with its intense yellow colour and nutty taste, has made durum wheat ideal for making pasta (Sissons, 2008). More details about durum wheat are given in the following paragraph.

Nowadays, bread wheat accounts for most of the global wheat crop, and its cultivation is widespread worldwide, from locations as far north as Norway, Finland, and Russia to southern countries, such as Argentina (Levy and Feldman, 2022).

Durum wheat represents about 5% of the al wheat production and is grown in a wide range of agricultural regions.

The Mediterranean region represents the major world's wheat growing area because durum wheat is well adapted to their climatic conditions.

Although bread wheat is grown globally, the most important in terms of spread and economic impact is durum wheat (Saia *et al.*, 2019).

2.7.1.2 Durum wheat

Durum wheat (*Triticum durum* Desf.) is free-threshing naked wheat and is easy to harvest (Peng, et al., 2011).

This cereal originates from domesticated emmer wheat strains, and as its primitive parent, the durum wheat plant is awned (with bristles).

Fig. 76 – Typical spikes and kernels of durum wheat

The wheat plant is composed of: stem, normally from 60 to 90 cm long; primary roots that originate directly from the seeds and secondary roots that originate from the base of the stem; the leaves, usually 5-8 per plant, and the inflorescences, called spikes. The wheat kernel or seed (caryopsis) is enclosed within the spikes. In brief, the kernel is constituted of three distinct parts: the germ (2-3% of dry matter), the bran (13-17% of dry matter), and the endosperm, the main part of the caryopsis (80-85% of dry matter).

The detailed structure of caryopsis is illustrated in figure 77.

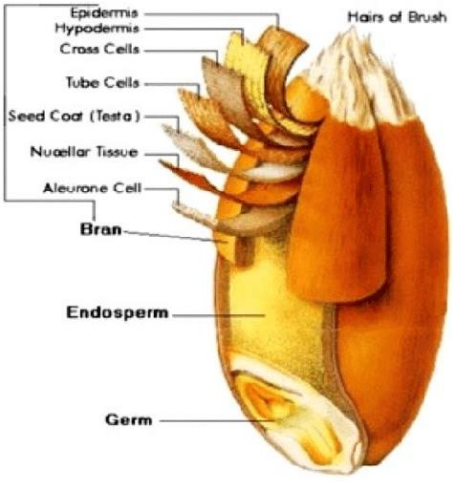


Fig. 77 – Structure of durum wheat kernel (Anderssen and Haraszi, 2009)

The wheat endosperm, from which semolina derives, contains mainly starch (60-70%) and storage proteins (8-18%).

The bran, one of the main fractions obtained by milling, along with germ, are rich in dietary fibres, minerals, and phytochemicals, which play important roles in nutrition and are beneficial to human health (Manai–Djebali *et al.*, 2021).

The detailed chemical composition of the durum wheat caryopsis is shown in table 12.

Table 12. Chemical composition and energy value of durum wheat grain (relative to 100 g). INRAN data

Water (g)	11.5
Proteins (g)	13
Lipids (g)	2.9
Starch (g)	53.9
Sugars (g)	3.2
Fibre (g)	9.8
Energy (Kcal)	312
Potassium (mg)	494
Iron (mg)	3.6
Calcium (mg)	30
Phosphorus (mg)	330
Magnesium (mg)	160
Zinc (mg)	2.9
Copper (mg)	0.40
Selenium (µg)	3.8
Thiamine (mg)	0.43
Riboflavin (mg)	0.15
Niacin (mg)	5.70
Vitamin A (µg)	2

Of great interest is the composition and function of the endosperm proteins, which have been extensively investigated in relation to the technological quality of durum wheat.

In particular, the gliadins and glutenins, the compounds of gluten, confer to durum wheat the essential viscoelastic properties for the required quality of its flour

(Shewry and Halford, 2002). Furthermore, durum wheat endosperm possesses carotenoid pigments (predominantly lutein), which confer it an intense bright yellow colour desirable for its derived products, mainly pasta (Saia *et al.*, 2019). The characteristic yellow pigment content, the vitrousness aspect, the major hardness of endosperm, the protein content, and the gluten strength are the key quality traits of technological properties of durum wheat.

Indeed, these characteristics determine the suitability of durum wheat for milling into semolina, used for pasta production.

However, other semolina-based products can be made (couscous, bourghul, and unleavened bread, etc.), also due to the very large amount of genetic diversity existing for the durum wheat crop (Sall, 2019).

The European Union (EU) accounts for the largest production of durum wheat, along with Canada, the United States, Turkey, and Algeria.

Among European Union (EU) countries, the main producers of durum wheat are Italy, Spain, France, and Greece.

The cultivation of this cereal is more concentrated in the Mediterranean since durum wheat is very well adapted to the dry climatic conditions of this area. Moreover, the countries of the Mediterranean basin are the major importers and the largest consumers of durum wheat products (Xynias *et al.*, 2020).

The appreciation of durum wheat in those areas is deeply connected with the Mediterranean history and culinary traditions. (Martínez-Moreno *et al.*, 2020).

2.7.2 Non-targeted NMR method applied to ensure the traceability of Italian durum wheat

Italy has been the first country within the Mediterranean area to begin durum wheat breeding, and to date is one of the biggest producers of this cereal, with an average production of 4.26 million tonnes in the last decade (Xynias *et al.*, 2020). The leading role of Italy in durum wheat production is partly attributable to the economic importance of its pasta industry, which has fuelled the intense

breeding work carried out in the state since the beginning of the 20th century (Mefleh *et al.*, 2019).

Within the national country, Southern Italy (Apulia, Basilicata, and Sicily) accounts for the largest production of durum wheat (Flagella, 2006).

Food authentication, in terms of geographical and varietal/animal origin, is considered of primary importance at all levels of the production process.

In the last year, assurance requirements (certifications and supply chain traceability) have been introduced in the durum wheat sector, to guarantee product quality to the consumers. These quality assurance systems, for example, have led to gaining the PDO mark for Altamura bread.

Thus, certifying the geographical origin of Italian durum wheat is of great commercial importance, particularly for wheat grown in given geographical areas, which can affect its quality and subsequently the commercial value of this cereal. To counter false declaration of origin or the creation of fraudulent mixtures of Italian durum wheat with foreign for economical gain, the development of methods able to assess the geographical origin of wheat is needed (Liu *et al.*, 2021). Among the latter, the non-targeted method based on Nuclear Magnetic Resonance (NMR) has received increasing attention as a reliable tool for tracing and assessing the authenticity of food products. Non-targeted NMR spectroscopy approach allows the extraction of the fingerprint of the food sample under investigation in a rapid and non-destructive way.

The fingerprint provided by this analysis may be primarily intended not for identifying the chemical compounds, but for the definition of a pattern to trace the sample and identify it by suitable classification tools (Gallo *et al.*, 2020).

Furthermore, the successfully used non-targeted NMR method for assessing the authenticity of foods is due to its ability to generate highly reliable instrumental responses. Indeed, this approach provides statistically equivalent NMR spectra when a single sample is analysed by different instruments.

This aspect suggested the potential creation of a community-built system containing NMR spectra comparable and exploitable, to empower the assessment of

food authenticity and traceability (Gallo *et al.*, 2020; Ragone *et al.*, 2020). Thereby, a non-targeted NMR method was applied to authentic durum wheat, whose geographical origin was assured by the Italian Central Inspectorate for Quality Protection and Fraud Repression of Agri-Food Products (ICQRF). Inspectorate, operating in the Italian territory, protects consumers and producers from unfair competition.

The investigated method, in collaboration with ICQRF, aims to ensure the identification of the geographical origin of authentic durum wheat samples.

The need emerges to avoid the possibility that breeders/operators might resell foreign products as Italian. The collection of authentic samples is the first effort in the building of an official classifier able to establish the geographical origin of unknown durum wheat by exploiting the associated NMR spectra. (Fig. 78).

In this work, the NMR spectra of the authentic durum wheat samples were employed for the development and validation of a discriminant model, able to differentiate the Italian product from the foreign ones.

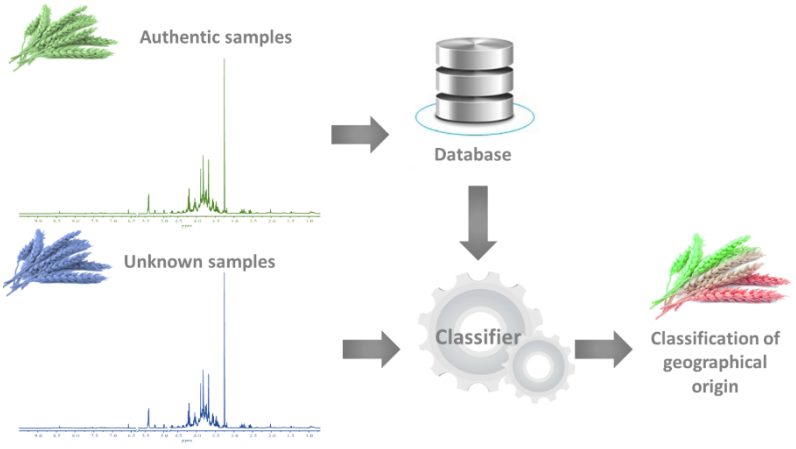


Fig. 78 – Flowchart of the development of a classifier for durum wheat traceability by application of non-targeted NMR method

2.7.3 Materials and Methods

2.7.3.1 Samples collection

Samples of authentic durum wheat provided by ICQRF – *Central Inspectorate for Quality Protection and Fraud Repression of Agri-Food Products*-were used to develop and validate the classifier model able to certify their geographical origin, in a way to discriminate the national product from the foreign one. The Inspectorate carried out sampling and controls according to standard procedure EC Reg. 882/2004. The corresponding collection card was attached to each provided sample.

Wheat samples were sent to the Chemistry Laboratory of the Polytechnic of Bari and stored at room temperature until their preparation for NMR analysis.

The pool used to develop and validate the method consisted of 147 samples of durum wheat of certified origin. The geographical origin of the latter is reported in table 13:

Table 13. Authentic wheat samples provided by ICQRF

Number of samples	Geographical origin	Harvest year
18	Italy	2020
85	Italy	2021
22	Italy	2022
6	USA	not specified
11	Canada	not specified
1	Croatia	not specified
1	France	not specified
1	Mexico	not specified

1	Spain	not specified
1	Hungary	not specified

2.7.3.2 Optimization of the sample preparation procedure

The procedure of sample preparation is very important for NMR analysis. Therefore, before conducting the spectroscopic analysis, it is necessary to optimise the sample preparation protocol to obtain reliable and meaningful results.

In optimising the preparation of the durum wheat samples, the matrix concentration to be used for the extraction step was evaluated. For this purpose, 50 g of wheat was ground using a blender for 1 minute and the resulting flour was then sieved using a laboratory sieve (sieve size of 0.5 mm).

Different samples were prepared for the aqueous extraction step, using a buffer solution [$(HC_2O_4)^-/(C_2O_4)^{2-}$ = 0.25 M] at pH= 4.2 [pH adjusted by addition of HCl (37%)] and containing sodium azide (NaN_3) as a bacteriostatic agent (NaN_3=2.5 x 10^{-3} M):

- sample 1: 1 g of flour was added to 4 mL of the buffer solution;
- sample 2: a volume of 3 mL of the buffer solution was added to 100 mg of wheat flour;
- sample 3: a volume of 3 mL of the buffer solution was added to 200 mg of wheat flour;
- sample 4: a volume of 3 mL of the buffer solution was added to 300 mg of wheat flour.

Samples were put in an ultrasonic bath at 60° C for 10 minutes.

After being vortexed for 5 minutes at 2500 rpm, samples were then centrifuged at 6000 rpm for 10 minutes. Then, the supernatants were transferred to new tubes and pasteurized at 90° C for 10 minutes to inhibit enzyme activity.

Each solution was filtered through a polytetrafluoroethylene syringe filter (PTFE, 0.2 µm, LLG-Labware) and the filtrates were used to fill the NMR tube (Norell 509-UP 7). 630 µL of each extract and 70 µL of 0.2 % w/w TSP/D_2O solution (3-

(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt in D_2O) were added in NMR tubes. The TSP was used as a chemical shift reference.

One-dimensional ^1H NOESY spectra were recorded through a Bruker Avance 400 MHz spectrometer equipped with a 5 mm inverse probe and with an autosampler. The following acquisition parameters were used: pulse program = noesygppr1d; size of fid (TD) = 64 K; spectral width (SW) = 20 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 µs; power level for pre-saturation (pl9) = 62.77 dB; dummy scans (ds) = 4; number of scans (ns) = 64; acquisition time = 4.09 s; mixing time (d8) = 0.01 s; recycle delay (d1) = 10 s.

Free induction decays (FIDs) were Fourier transformed and processed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spain).

The phase was manually corrected, the baseline was automatically corrected, and the spectra were aligned to the TSP singlet, fixed at 0 ppm. Figure 79 displays the 1D ^1H NOESY spectra recorded from the aqueous wheat extracts prepared with the different concentrations, as described above.

The red spectrum is relative to sample 1 (1 g of wheat flour in 4 mL of buffer); the green spectrum relates to sample 2 (100 mg of wheat flour in 3 mL of buffer); the blue spectrum is related to sample 3 (200 mg of wheat flour in 3 mL of buffer); the violet spectrum relates to sample 4 (300 mg of wheat flour in 3 mL of buffer).

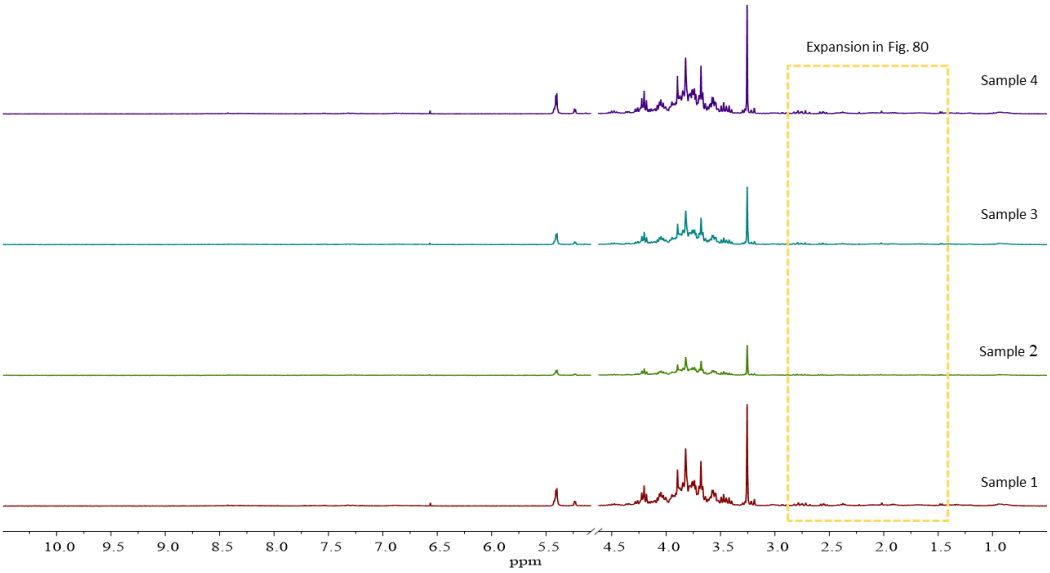


Fig. 79 – Comparison of 1D ^{1}H NOESY spectra recorded from the aqueous wheat extracts prepared from different concentrations.

From the comparison of the 1D ^{1}H NOESY spectra, illustrated in figure 79, it was possible to observe certain differences in the metabolic profile between the different samples. Looking at the spectra related to sample 2 (green spectrum in figure 79) and sample 3 (blue spectrum in figure 79), it was noted a lower intensity of the signals relating to amino acids (the spectral region between 1-3 ppm) and sugars (the spectral region between 3-6 ppm). By contrast, the spectra recorded from samples 1 and 4 (red and violet spectra, respectively) did not show visible variations in metabolic composition. For example, figure 80 shows a spectral region in which clear variations in the metabolic profile of the samples prepared with different concentrations of wheat flour were observed.

In particular, the figure illustrates expanded signals of threonine (1.32 ppm), arginine (1.91 ppm) and γ-Aminobutyric acid (2.38 ppm), delimitated from yellow rectangles.

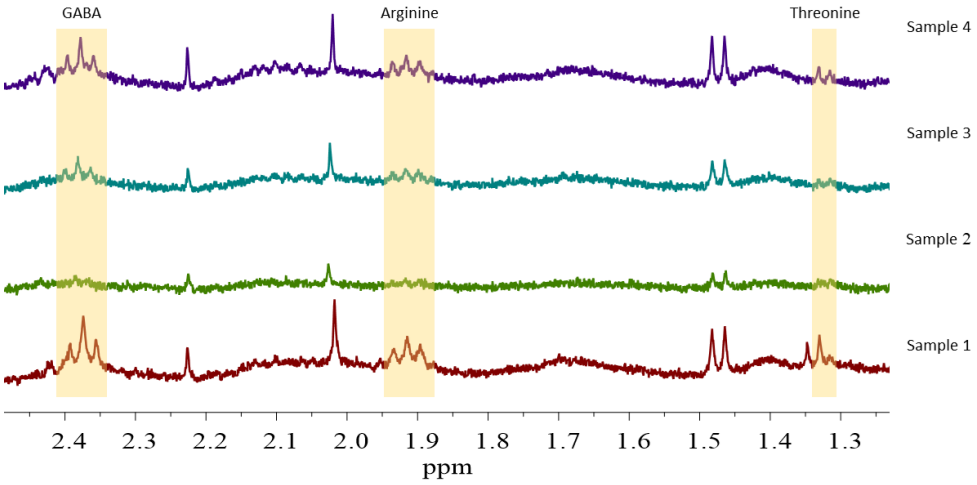


Fig. 80 – Signals displaying an intensity variation in 1D ^{1}H NOESY spectra recorded from aqueous wheat extracts prepared with different concentrations.

The half-height width of the TSP signal ("hwcal") was estimated, verifying that its value was less than 1.5 Hz. In addition, the ratio of a representatively chosen signal (S) to the noise background (N) was evaluated. The signal chosen was at 3.25 ppm, related to the betaine compound. The metabolites characterization of aqueous durum wheat extract is described below.

The data obtained from the measurement of the half-width of the TSP signal (hwcal) and the signal-to-noise ratio (S/N) are collected in Table 14.

Table 14. Parameters evaluated from the recorded spectra

Sample	hwcal	S/N
1	1.69 Hz	662.50
2	1.55 Hz	214.22
3	1.51 Hz	389.94
4	1.49 Hz	737.79

Sample 4 has obtained the lowest half-width value of the TSP signal (1.49 Hz) and the highest signal-to-noise ratio (S/N) value.
Hence, the concentration of 300 mg of wheat flour in 3 ml of the buffer solution was favourable for the extraction procedure of durum wheat samples.

2.7.3.3 Extraction procedure

50 g of each durum wheat sample was ground using a blender for 1 minute and sieved using a laboratory sieve (sieve size of 0.5 mm). An aliquot of 300 mg of the resulting wheat flour was dissolved in 3 mL of buffer solution [$(HC_2O_4)^-/(C_2O_4)^{2-}$ (0.25 M) and NaN_3 (2.5 mM)] at pH 4.2, sonicated for 10 min at 60°C, vortexed for 5 min at 2500 rpm and, centrifuged for 10 min at 6000 rpm. The resulting supernatant was drawn off and pasteurized for 10 min at 90°C.
After pasteurization, the sample was filtered through a syringe filter equipped with a polytetrafluoroethylene (PTFE) hydrophilic membrane with a pore size of 0.22 µm, to gain a clear sample without any suspended particles that may distort the magnetic field homogeneity. The filtrate was used to fill the NMR tubes.
NMR tubes were filled in by an automated system for liquid handling (SamplePro Tube, Bruker BioSpin) according to the following method: 630 µL of the obtained extract and 70 µL of TSP/D_2O solution [3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt in D_2O (0.2% w/w)]. Two replicates were prepared for each sample.

2.7.3.4 NMR analysis

One-dimensional ^{1}H NOESY spectra were acquired under automaton at 298.1° K using a Bruker Avance 400 MHz spectrometer equipped with a 5 mm inverse probe and with an autosampler (Bruker, Billerica, MA, USA).

The following acquisition parameters were used: pulse program = noesygppr1d; size of fid (TD) = 64 K; spectral width (SW) = 20 ppm; transmitter offset = 4.70 ppm; 90° hard pulse (p1) = 8.16 μs; power level for pre-saturation (pl9) = 62.77 dB; dummy scans (ds) = 4; number of scans (ns) = 64; acquisition time = 4.09 s; mixing time (d8) = 0.01 s; recycle delay (d1) = 10 s. Each spectrum was acquired using TOPSPIN 2.1 software (Bruker BioSpin GmbH, Rheinstetten, Germany) under an automatic process that lasted ca. 22 min and encompassed sample loading, temperature stabilization for 5 min, tuning, matching, shimming, and 90◦ pulse calibration.

NMR raw data (Free Induction Decays, FIDs) were processed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spain).

The 1D ^{1}H NOESY spectra were automatically Fourier transformed, manually phased, and automatic baseline corrected (by a 1st order Bernstein polynomial equation). Chemical shifts were referenced to the TSP-d_4 peak at δ = 0.00 ppm.

2.7.3.5 Processing of spectral data for statistical analysis

The processed 1D ^{1}H NOESY spectra were reduced to a numerical matrix called bucket-table. The bucket table was obtained by dividing the entire overlapped spectra in the range of [-0.5, 10.5] ppm into rectangular intervals (buckets) of 0.04 ppm in width, excluding the region [4.30,5.10] ppm corresponding to the residual water signal. The underlying area of each bucket was normalized to the total intensity. Thus, in the bucket table, the NMR spectra recorded from samples constituted the observation, and the buckets constituted the *x*-variables.

The numerical matrix generated for the durum wheat samples provided by ICQRF is described in the paragraphs of the results.

The bucket table was imported into SIMCA 13.0.3 software (Umetrics, Umea, Sweden) to perform multivariate statistical analyses (MVA).

2.7.4 Validation of the method applied

The analytical procedure applied to durum wheat samples was validated, i.e. the repeatability, intermediate precision, and reproducibility was were evaluated.

For the evaluation of reproducibility, ten weighings were carried out from the same sample of durum wheat and the respective ten aliquots were prepared for NMR analysis, according to extraction protocol described in the paragraph 2.7.3.3**.** In this way, it was possible to determine the variability induced by operator skill and degree of sample homogeneity.

For the evaluation of intermediate precision and repeatability, a bulk solution of the examined matrix was prepared. The bulk solution was made to prepare 10 NMR tubes and the subsequent acquisition of the respective NMR spectra (intermediate precision). In addition, 10 NMR spectra were acquired for the same NMR tube to measure the variability induced by the instrument (repeatability). For this purpose, the NMR tube was pulled out and inserted ten times into the instrument.

The betaine signal (3.25 ppm) was selected for the evaluation of the repeatability. Therefore, the integrals of the spectral area [3.24,3.26] from the recorded 1D ^{1}H NOESY spectra were determined.

From the resulting integral values, the coefficient of variation percentage (CV%) was calculated. Therefore, the CV% calculated for the selected signal of betaine was used as an indicator of the repeatability study.

In addition, the stability of the investigated matrix was studied to assess any changes over time in the metabolic profile.

2.7.4.1 Evaluation study of reproducibility

From a ground and sieved sample of durum wheat, 10 different weighings (each of 300 mg) were made and the 10 aliquots were subjected to the preparation

protocol described in in the paragraph 2.7.3.3. The prepared NMR tubes were subjected to spectroscopic analysis, following the instructions for the acquisition of 1D ^{1}H NOESY proton spectra given in the paragraph 2.7.3.4.

The processing consisted of Fourier transformation, manual phase and automatic baseline correction, and alignment of the spectra to the signal of TSP-d_4 at 0 ppm. The spectra recorded from the ten NMR replicates are illustrated in figure 81. The signal of residual water (4.78 ppm) was hidden.

The yellow rectangle highlights the spectral area of the betaine signal (3.25 ppm), selected for the calculation of integrals.

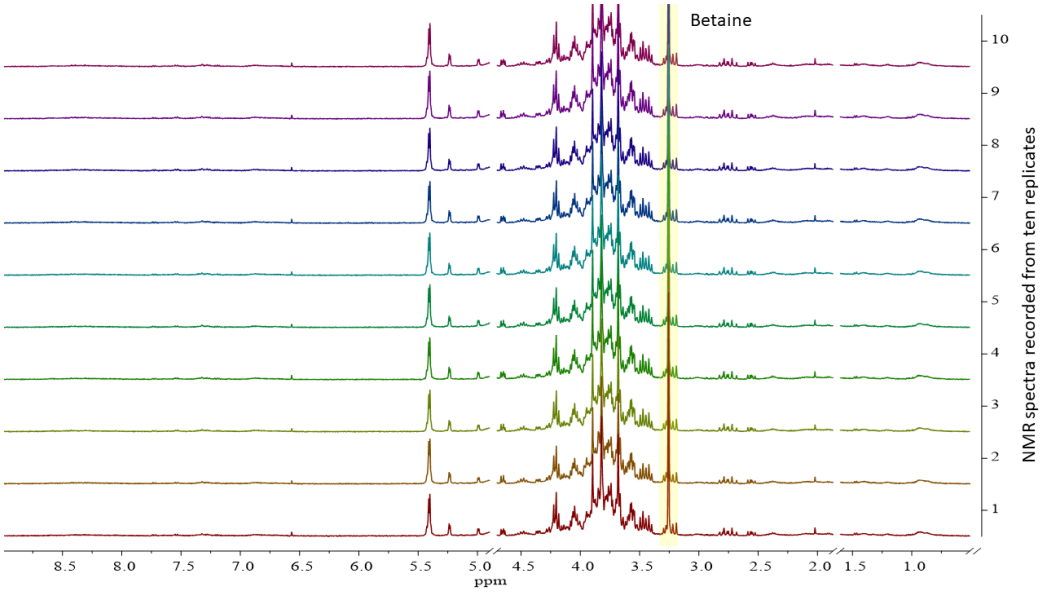


Fig. 81– 1D ^{1}H NOESY spectra of durum wheat recorded from the ten NMR tubes prepared for the reproducibility evaluation.

The integration of the area subtended by the betaine signal (yellow rectangle in figure 42) was performed using the software MestReNova 11.0 (Mestrelab Research SL, Santiago de Compostela, Spagna). The data obtained from those calculations are collected in table 15.

Table 15. Results of the reproducibility study of durum wheat

NMR replicates	The integral value of the betaine signal
1	12401
2	11274.5
3	11797.9
4	12424.3
5	11364.7
6	12020.3
7	12706.1
8	12058.5
9	13562.4
10	12863.2
Standard deviation	698.3
Mean	12247.2
CV%	**5.7**

2.7.4.2 Preparation of a bulk solution

The bulk solution was prepared by mixing 3 g of the wheat flour (the wheat flour resulted from the ground and sieved durum wheat sample) and 30 ml of the buffer solution [$(HC_2O_4)^-/(C_2O_4)^{2-}$ (0.25 M) and NaN_3 (2.5 mM) at pH 4.2]. Then, the bulk solution was sonicated for 10 minutes at 60°C, vortexed for 5 minutes at 2500 rpm and, centrifuged for 10 minutes at 6000 rpm.

The supernatant was pasteurized for 10 min at 90°C; then the bulk solution was filtered through a centrifuged filter in in PVDF (polyvinylidenefluoride, with pore size of 0.2 um) for 5 minutes at 4000 rpm.

2.7.4.3 Evaluation study of intermediate precision

The filtrated bulk solution was distributed into 10 vials for the preparation of the respective NMR tubes. NMR tubes were filled in by an automated system for liquid handling (SamplePro Tube, Bruker BioSpin) according to these volumes: 630

µL of the filtrate and 70 µL of TSP-d_4/D_2O solution [3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt in D_2O (0.2% w/w)]. The 10 NMR tubes were submitted to NMR analysis (Bruker Avance 400 MHz), according to the acquisition parameters indicated in paragraph 2.7.3.4.

The spectra recorded from the ten NMR tubes, prepared from the bulk solution, are illustrated in figure 82. The signal of residual water (4.78 ppm) was hidden. The yellow rectangle highlights the spectral area of the betaine signal (3.25 ppm), selected for the calculation of integrals.

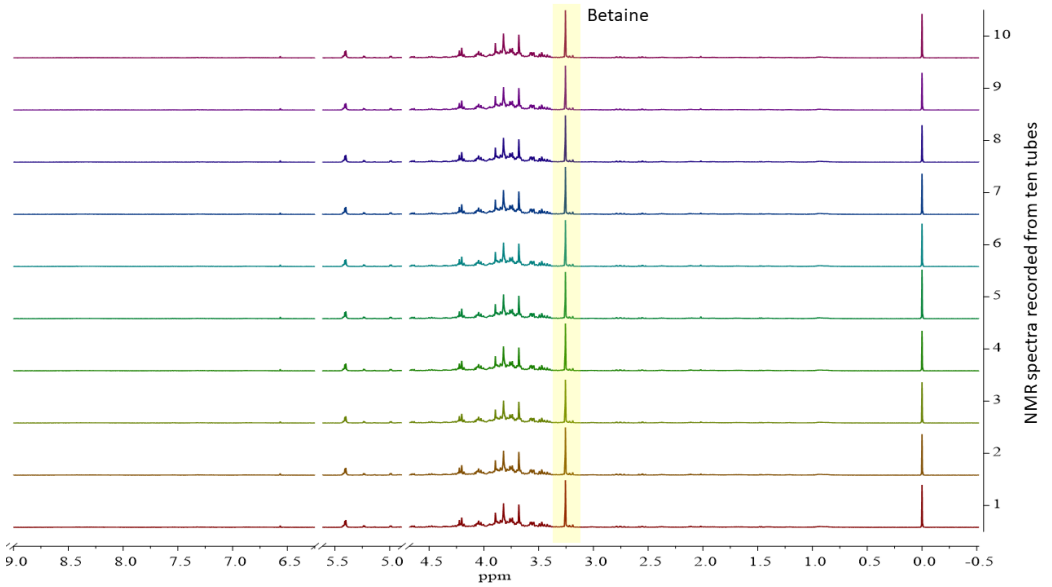


Fig. 82– 1D ^1H NOESY spectra of durum wheat recorded from the ten NMR tubes prepared for the intermediate precision evaluation.

The data obtained from those calculations are collected in table 16.

Table 16. Results of the intermediate precision study of durum wheat

NMR tubes	The integral value of the betaine signal
1	11815.8
2	12007.1
3	10955.9
4	11913.9
5	11670
6	11769.7
7	11815.1
8	11868.2
9	11210.5
10	11890
Standard deviation	338.2
Mean	11691.6
CV%	**2.8**

2.7.4.4 Evaluation study of repeatability

An NMR tube prepared from the bulk solution of wheat (previously described) was analysed ten times for the evaluation of repeatability. This NMR tube was removed and reinserted ten times inside the spectrometer, in an automated way. Thus, ten 1D ^{1}H NOESY spectra of the same tube were recorded. The acquisition parameters are indicated in paragraph 2.7.3.4.

Figure 83 shows the 10 spectra recorded from the same NMR tube, chosen for the evaluation of repeatability. The yellow rectangle highlights the spectral area of the betaine signal (3.25 ppm), selected for the calculation of integrals.

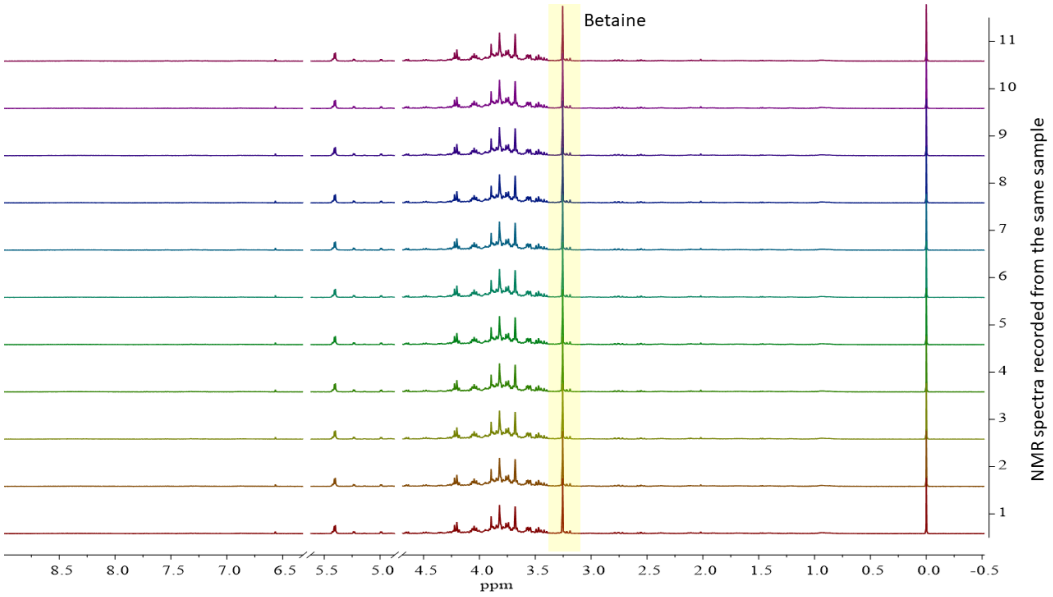


Fig. 83– 1D ^{1}H NOESY spectra of durum wheat recorded from the same NMR tubes prepared for the repeatability evaluation.

The data obtained from those calculations are collected in table 17.

Table 17. Results of the repeatability study of durum wheat

1D ^{1}H NOESY spectra	**The integral value of the betaine signal**
1	11739.3
2	11706.5
3	11704.5
4	11695.5
5	11687.2
6	11706.5
7	11738.5
8	11676.1
9	11704.2
10	11718.3
Standard deviation	20.1
Mean	11707.6
CV%	**0.2**

The calculated CV% values for the evaluation of intermediate precision and the repeatability (2.8, and 0.2), were found inferior to 5%. In contrast, the CV% value obtained in the evaluation of reproducibility was just over 5% (5.7), but it was higher than the intermediate precision and repeatability ones. This may result because the operator performs different weighings, so the possibility of an experimental error is greater.
Thus, due to the good data obtained, the applied analytical procedure resulted reliable for the analysis of durum wheat samples.

2.7.4.5 Study of the matrix stability over time

A NMR tube of aqueous extract of durum wheat was used to check the stability of the investigated matrix over time. Therefore, the 1D ^{1}H NOESY spectrum was recorded on the freshly prepared sample; then, the NMR tube was left at room temperature. NMR spectra were acquired after 24 hours, five and six week later from its preparation, according to the experimental parameters reported in paragraph 2.7.3.4.
Figure 84 shows the recorded NMR spectra on the NMR tube used to evaluate the stability as follows: the pink spectrum is related to the freshly NMR tube prepared (0 h); the green one corresponds to the NMR spectrum recorded after 24 hours; the orange spectrum represents the NMR spectrum acquired after five weeks; finally, the blue spectrum represents the NMR spectrum acquired after six week from its preparation.

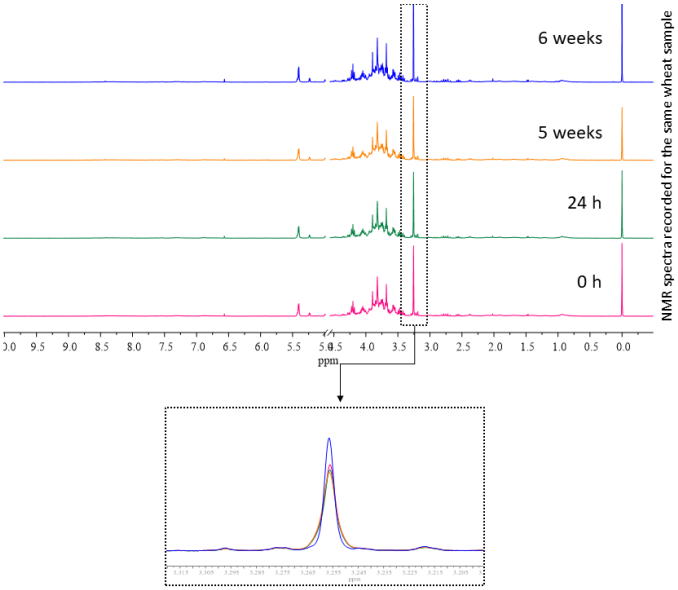


Fig. 84– 1D ^{1}H NOESY spectra of durum wheat sample recorded from the same NMR tube for the stability study. The signal of betaine (3.25 ppm) chosen to evaluate the stability of the sample is expanded.

The stability of the sample was assessed by integrating the areas of the signal of betaine (3.25 ppm) selected in the NMR spectra recorded at the time points indicated (the expansion of the signal is shown in fig.84). The integral of the areas subtending the signal was calculated in the spectral regions [3.27,3.24 ppm]. The data obtained are reported in table 18.

Table 18. Results of the stability of durum wheat sample

	Betaine
Integration range (ppm)	**[3.27,3.24]**
Time points	**Integral values**
0 h	14179.3
24 h	13249.4
5 weeks	13893.8
6 weeks	14197.5
Standard deviation	442.8
Mean	13880
CV%	**3.2**

The durum wheat proved to be very stable over time, showing no obvious changes in its metabolic profile even six weeks after the preparation. An indication of that is indicated by the very good CV% value calculated for the selected signal of betaine (3.2 %).

2.7.5 Metabolites characterization of Italian durum wheat

NMR analysis (1D ^{1}H NOESY experiments) allowed for the identification of hydrosoluble metabolites contained in typical aqueous extracts (pH = 4.2) of durum wheat samples.

Metabolite identification was performed employing a library of 1D ^{1}H NOESY spectra of pure compounds prepared and recorded with the procedure described for durum wheat samples. The main metabolites identified are listed in table 19.

Table 19. Characteristic metabolites of aqueous extract of durum wheat

Compound	**Chemical shift (ppm)**	**Multiplicity: J (Hz)**
Amino acids		
Isoleucine (1)	0.93	t (7.3)
	1.00	d (7.0)
	1.26	m
	1.46	m
	1.96	m
Valine (2)	0.98	d (7.0)
	1.03	d (7.0)
	2.27	m
	3.64	d overlapped
Alanine (4)	1.47	d (7.2)
	3.81	q overlapped
Arginine (5)	1.69	m
	1.91	m
	3.25	m overlapped
	3.73	m overlapped

Proline (6)	2.00	m
	2.07	m
	2.34	m
Glutamic acid (7)	2.18	m
	2.47	m
	3.78	dd overlapped
γ-Aminobutyric acid (8)	1.92	m
	2.38	t
	3.04	bs
Asparagine (9)	2.86	dd (16.8; 7.3)
	2.95	dd (16.9; 4.4)
	6.89	bs
	7.62	bs
Tyrosine (10)	6.89	m
	7.18	m
Tryptophan (11)	7.19	t (7.4)
	7.28	t (7.5)
	7.32	d (2.2)
	7.54	d (8.1)
	7.73	d (8.0)
Phenylalanine (24)	7.33	m overlapped
	7.42	m overlapped
Organic acids		
Lactic acid (3)	1.32	d (6.8)
	4.16	q overlapped
Acetic acid (12)	2.05	s
Succinic acid (13)	2.59	s
Malic acid (14)	2.62	dd (15.8; 7.8)
	2.81	dd (15.8; 4.3)
	4.38	dd (8.1; 4.3)
Citric acid (15)	2.73	d (15.3)
	2.84	d (15.3)
Fumaric acid (16)	6.60	s

Formic acid (17)	8.40	s
Lactic acid (25)	1.35	
	4.14	
Carbohydrates		
Sucrose (18)	3.47	t (9.4)
	3.56	dd (9.8; 3.7)
	3.68	s
	3.76	m
	3.82	m
	4.05	t (8.6)
	4.21	d (8.7)
	5.40	d (3.8)
Glucose (19)	3.24	dd overlapped
	4.64	d (7.9)
	5.23	d (3.7)
Galactose (20)	4.59	d (7.9)
	5.26	d (3.9)
RFOs (21)	4.99	d
	5.14	d
	5.42	d
Quaternary ammonium compounds		
Betaine (22)	3.26	s
	3.90	s
Choline (23)	3.19	s

The typical 1D ^{1}H NOESY spectrum of an aqueous extract of durum wheat is shown in figure 85. The spectrum can be divided into three regions, which are expanded for clarity in figure 86.

In the aliphatic region of the ^{1}H NMR spectrum [0.8-3.00 ppm], the most abundant signals were related to the amino acid detected: isoleucine, valine, alanine, arginine, glutamic acid, glutamine, asparagine. The aliphatic region also included signals arising from the organic acids identified: lactic, acetic, succinic, malic, and citric acids.

The central area of the spectrum [3.00-6.00 ppm], was dominated by overlapping signals of carbohydrates corresponding to sucrose, glucose, and galactose; in addition, the characteristic peaks corresponding to raffinose family oligosaccharides (RFOs) were detected. The most evident signal at 3.26 ppm was the singlet of betaine (3.25 ppm). This is to the finding reported in previous studies, indicating that betaine is one of the predominant metabolites of wheat (Graham, et al., 2009; Shewry, et al., 2017). Betaine plays a significant role in nutrition (Zeisel, et al., 2003; Likes, et al., 2007), and wheat resistance to pathogenic attack (Strange, et al., 1974; Browne & Brindle, 2007).

The aromatic region [6.00-9.00 ppm] included signals relating to aromatic amino acids such as tyrosine, phenylalanine and tryptophan. Moreover, signals referred to fumaric and formic acid were identified as well.

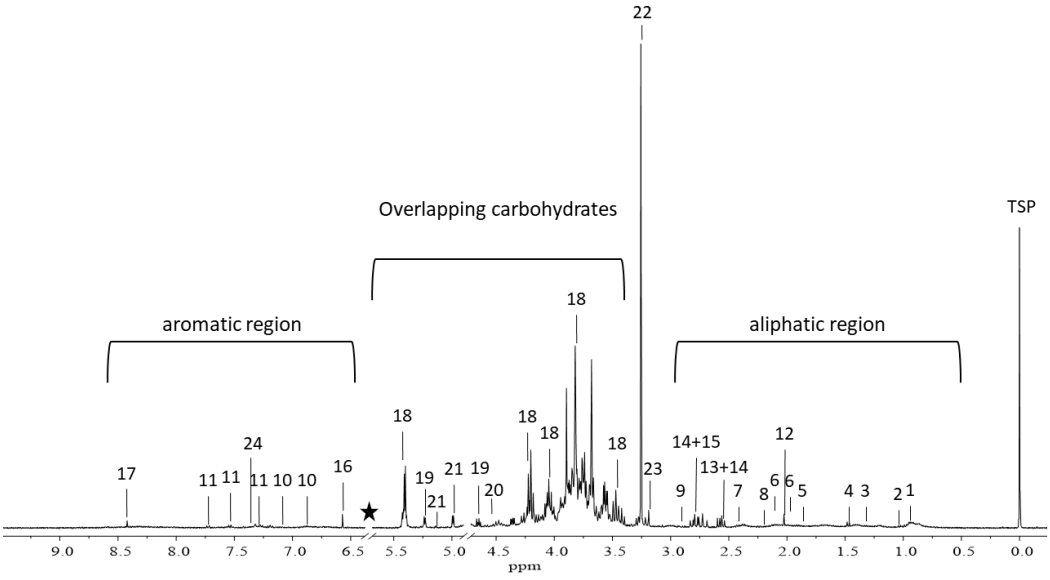


Fig. 85– 1D ^{1}H NOESY spectrum of a typical aqueous extract of durum wheat. The HDO signal is hidden; numbers indicate the metabolites reported in table 17.

The following figure shows the expansion of the three described spectral area: a is the aliphatic region; b is the spectral region characteristics of the carbohydrates; c is the aromatic region.

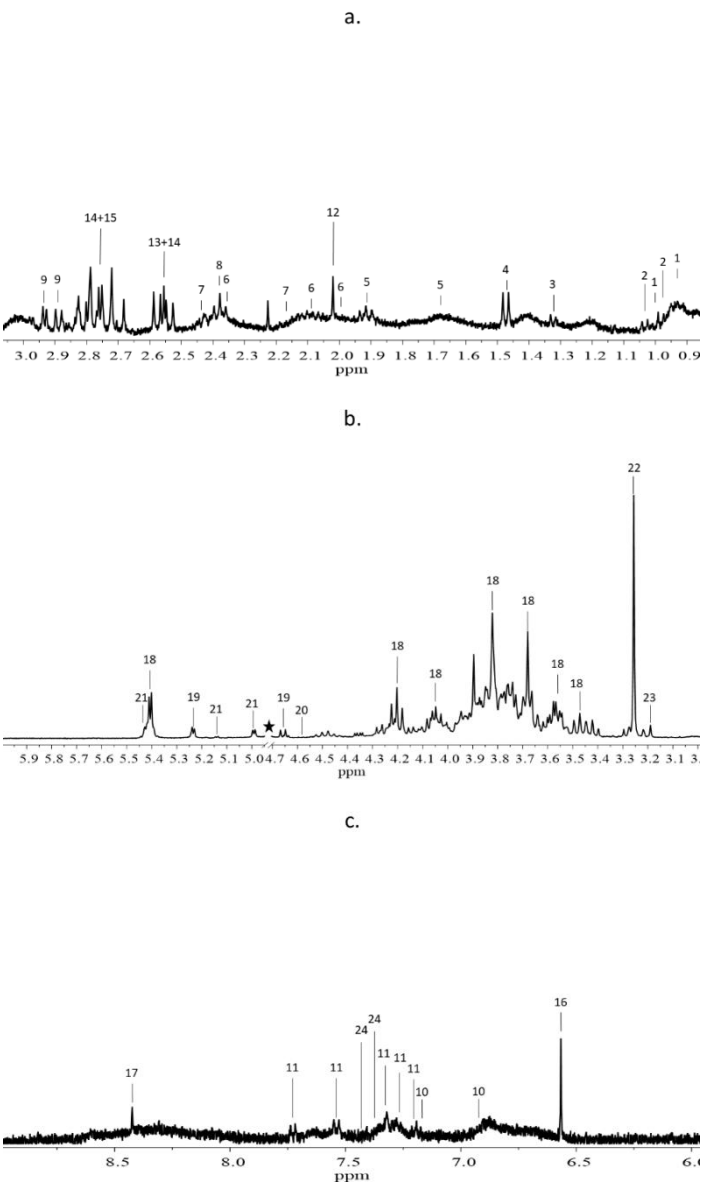


Fig. 86– Expansion of a) aliphatic region, b) carbohydrates region and c)aromatic region of the typical NMR spectrum of durum wheat. The HDO signal is hidden; numbers indicate the metabolites reported in table 17.

The resonances assignments were confirmed by bidimensional experiment (^{1}H COSY). The ^{1}H COSY spectrum acquired from durum wheat sample is shown in

figure 87. Signals of isoleucine, lactic acid, arginine, malic acid, glucose, tyrosine, tryptophan and phenylalanine are highlighted in the figure 87.

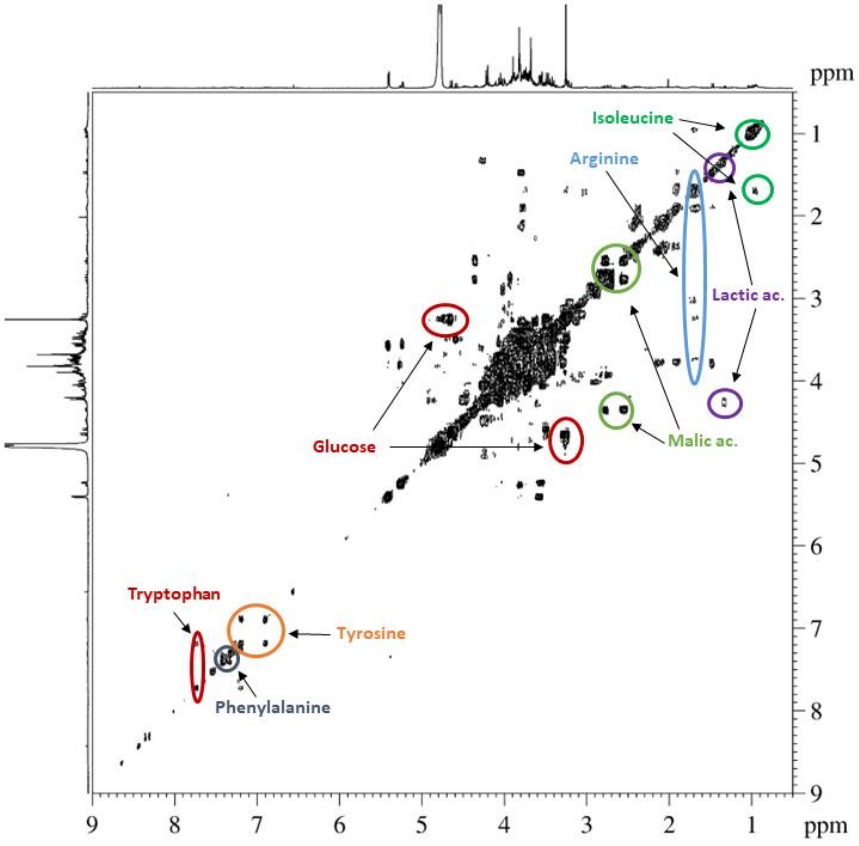


Fig. 87–^{1}H COSY spectrum of durum wheat

2.7.6 Results

2.7.6.1 Development of a discriminant model to ascertain the geographical origin of durum wheat

The 294 processed 1D ^{1}H NOESY spectra (147 samples x 2 replicates) recorded on the durum wheat samples provided by ICQRF were reduced to bucket-table. The bucket table was obtained as mentioned in paragraph 2.7.3.5: the entire over-lapped spectra were segmented in the range of [-0.5, 10.5] ppm into buckets of 0.04 ppm in width, excluding the region [4.30,5.10] ppm corresponding to the

residual water signal. The underlying area of each bucket was normalized to the total intensity. The bucket table obtained was composed of 198 *x*-variables (bucket) and 294 observations (spectra recorded), and was imported into SIMCA 13.0.3 software (Umetrics, Umea, Sweden). Buckets were subjected to Unit Variance scaling (mean-centred and divided by the standard deviation of each variable) to avoid noise inflation. An Orthogonal Partial Least Squares discriminant analysis (OPLS-DA) was performed to create a discriminant model to ascertain the geographical origin of durum wheat samples. OPLS-DA is an excellent tool in which the differentiation of assigned classes (in this case Italy vs foreign origin) is stressed. For the development of this model, the NMR spectra recorded on the authentic samples provided by ICQRF were separated into a *training set* and a *test set*. The training set was used to "train" and optimise the discriminant model to identify the classes of the samples, known *a priori.* It consisted of a set of 125 authentic durum wheat samples divided as follows:

- 103 samples of durum wheat of Italian origin,
- 11 samples of durum wheat of Canadian origin,
- 6 samples of durum wheat of American origin,
- 1 sample from France, Hungary, Spain, Mexico, and Croatia.

The purpose was to unequivocally differentiate the product of Italian origin from foreign ones. Therefore, at the pool of authentic samples got from Italy, was assigned the class "Italy". Instead, the pool of authentic samples of foreign origin, regardless of their provenance (Canada, America, France, etc…), was assigned the class "Foreign origin".

The quality of the model developed was assessed based on the parameters R^2 and Q^2, which represent the descriptiveness (goodness-of-fit) and the predictivity in cross-validation (goodness-of-prediction in cross-validation), respectively. Moreover, a permutation test was carried out to verify the absence of model overfitting**.**

The OPLS-DA model was characterized by one predictive component (t[1]), and four orthogonal components (to), with R^2= 0.825 and Q^2= 0.784.

The parameters characterizing the OPLS-DA model, built by using the training set samples are detailed in table 20.

Table 20. The parameters of the OPLS-DA analysis performed on the spectral data of the authentic durum wheat samples included in the training set

Component	R^2X	R^2X(cum)	R^2	R^2(cum)	Q^2	Q^2(cum)	R^2Y	R^2Y(cum)
Model		**0.684**		**0.825**		**0.784**		**1**
Predictive		0.0949		0.825		0.784		1
P1	0.0949	0.0949	0.825	0.825	0.784	0.784	1	1
Orthogonal in X (OPLS)		0.589						
O1	0.31	0.31						
O2	0.116	0.426						
O3	0.127	0.554						
O4	0.0355	0.589						

The scores plot (t[1] vs to[1]) obtained from the analysis is shown in figure 88. As expected, a clear separation of the two classes, “Italy” (blue circles in fig. 88) and “Foreign origin” (green circles in fig.88) was observed on the scores plot based on the first predictive (t[1]) and orthogonal (to[1]) components.

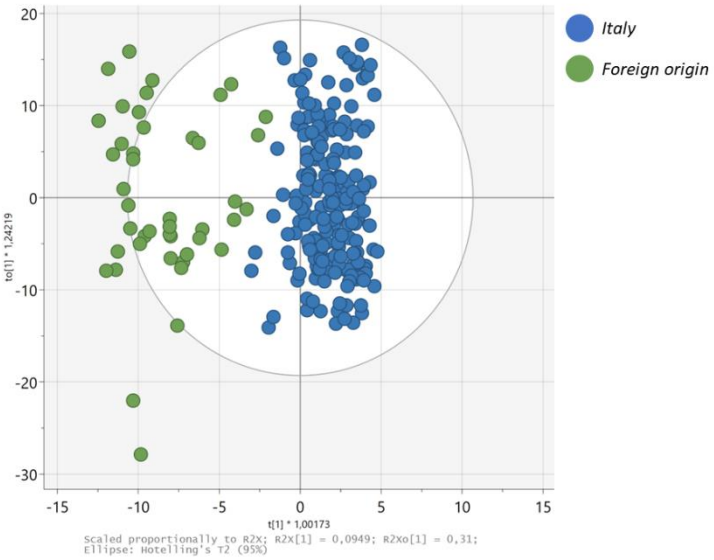


Fig. 88– Scores plot (t[1] vs to[1]) of the OPLS-DA analysis performed to spectral data of authentic durum wheat samples selected as the training set.

The permutation test with 100 iterations for the OPLS-DA model is shown in figure 89.

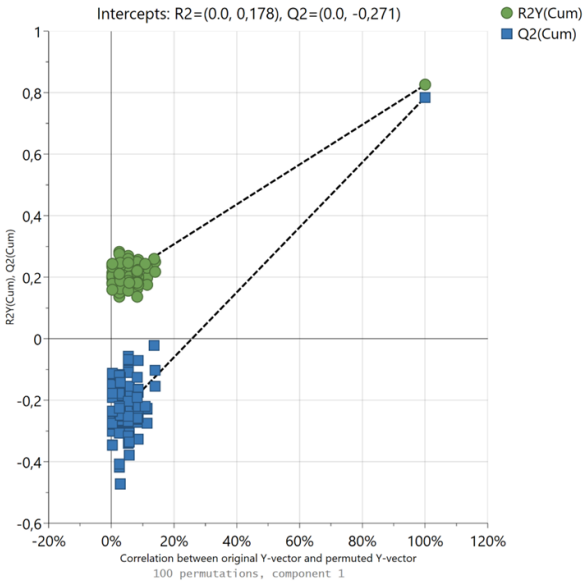


Fig. 89– Permutation test applied to the OPLS-DA model built on the spectral data obtained from authentic durum wheat samples included in the training set. The permutation test consists of a set of 100 permutations. The values of the R^2 and Q^2 coefficients are indicated.

The R^2 and Q^2 values resulted were 0.178 and -0.271, respectively. Since these values agreed with the recommended limits ($R^2 < 0.3$ and $Q^2 < 0.05$), the permutation test ensured the statistical significance of the model developed.

2.7.6.2 Validation of the developed discriminant model

The discriminant model built on the NMR data of the training set (125 durum wheat samples with certified origin) was able to distinguish samples efficiently according to geographical origin ("Italy" *vs* "Foreign origin"). Such a model was validated, i.e., the ability to predict the geographical origin of a test set was evaluated. The test set was composed of the remaining 22 Italian authentic durum wheat samples, not included in the development of the discriminant model. The OPLS-DA model was used to predict the test set samples with the declared origin, but blind for the discriminant model. Therefore, a prediction analysis of the test samples was performed, using the model previously developed with the training

set samples. The predicted scores plot resulting from the analysis is shown in figure 90.

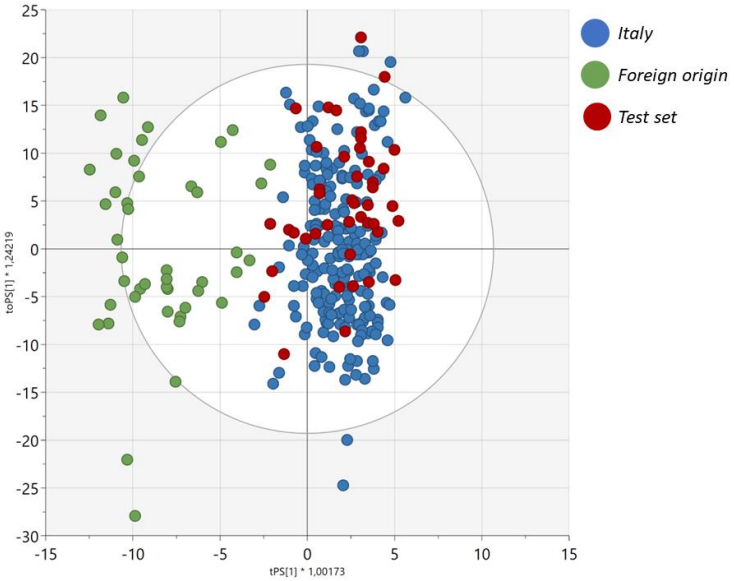


Fig. 90– OPLS-DA predicted scores plot for durum wheat samples included in the test set. The predicted samples are indicated as red circles.

The OPLS-DA scores plot shows that the test samples (red circles in fig.90) were classified in the “Italy” class.

The misclassification table (table 21) showed the number of correctly classified observations in the prediction analysis. In particular, the authentic Italian durum wheat samples (part of the training set) were classified in the “Italy” class 100% correctly. The authentic foreign durum wheat samples (part of the training set) were classified as the foreign origin with a correctness of 93.18%. It is important to note that the lower value obtained may be due to the reduced amount of foreign durum wheat samples provided for model development. The model showed an ability to correctly predict the geographical origin of the test set of 98.8 % (red value in table 19). The authentic durum wheat samples constituting the test set were predicted within the “Italy” class (red circles in fig.85). The classification of

these samples was consistent with their declared Italian origin, guaranteed by ICQRF. This result also supported the reliability of the developed model.

Table 21. Misclassification table for training and test durum wheat samples

	Members	Correct	"Foreign origin" class	"Italy" class
Training set				
Authentic foreign durum wheat	44	93,18%	41	3
Authentic Italian durum wheat	206	100%	0	206
Test set				
No class	44		0	44
Total	294	**98,8%**		
Fisher's prob.	6,9e-44			

The obtained results have demonstrated the efficiency of the non-targeted NMR method for certain the geographical origin of durum wheat. The statistical elaboration of the corresponding NMR spectra has led to the development of a model capable of discriminating authentic Italian durum wheat against foreign ones.
The pool of samples of foreign origin should be enlarged to improve the robustness of the discriminant model developed.
The proposed discrimination model should promote the build of an official classifier system based on NMR spectroscopy to assess the authenticity of durum wheat, as shown in the flowchart in figure 78.

2.7.7 Cultivars innovation, quality, and traceability of Apulian durum wheat

The quality traits of durum wheat grain directly impact the value of its derived products. Hence, the protein content, gluten strength and colour of durum wheat are important factors affecting pasta-making technology characteristics (Blanco, et al., 2006). These factors are under genetic control but may be affected by the different varieties and the agronomical practices applied during the cultivation of durum wheat. Nitrogen fertilisation is one of the main agronomic techniques

that exerts great influence on the technological quality of durum wheat grain, mainly raising the protein content. Moreover, Nitrogen fertilizers application is related to an increasing wheat grain yield (Ehdaie and Waines, 2001).

According to good practices, the most used dose of Nitrogen fertilizers for durum wheat cultivation is 120 kg ha^{-1}. Nevertheless, Nitrogen fertilizers pose a threat to the environment and human health by polluting the water and soil (Gadaleta et al., 2022). In recent years, growing attention among consumers on healthy food and environmental benefits has emerged. As a result, the adoption of organic cultivation has begun to increase.

The lowest level of protein content in durum wheat acceptable to the processing industry is 12%. The protein content of organically grown durum wheat is almost always lower than conventionally one, at or below 12% (Flagella, 2006).

Therefore, a good varietal choice and appropriate use of organic fertilisers could result in a protein content like that obtained by conventional agronomic practices. Based on these considerations, a project was designed

In particular, the strongest challenges of this project are the selection of higher-yielding durum wheat varieties and the improvement of agronomic practices towards more sustainable techniques.

The first project's task was to identify new and traditional durum wheat varieties to produce a high-quality wheat grain, suitable for bread and pasta with enhanced added value. The durum wheat varieties exploited were the Apulian ones, *Saragolla*, *Marco Aurelio*, *Maciste*, *Antalis* and *Iride*.

These autochthonous varieties should provide a high-quality wheat grain, together with an extensive crop yield. In this way, local farmers could benefit economically from the cultivation of these productive varieties.

The crop yield depends not only on the cultivars of durum wheat used but also on the agronomic practices applied during their cultivation.

To shift the application of conventional agronomic techniques toward the use of rational sustainable ones, the second task of the project was the identification of

durum wheat varieties (among *Saragolla*, *Marco Aurelio*, *Maciste*, *Antalis*, and *Iride*) suitable for low-input conventional cultivation and organic one.
An organic regime and different doses of nitrogen fertilisation (120 kg ha^{-1}, and 80 Kg ha^{-1}) applied for the cultivation of the interested durum wheat varieties, were investigated.
The expected results would lead to a reduction in the doses of nitrogen fertilisers to be used for durum wheat cultivation, and the achievement of economic and environmental benefits.
Finally, the specificity of the durum wheat grain obtained and the traceability in the supply chain of these Apulian varieties (from field durum wheat to its derived products), was supported through the application of the developed non-target NMR spectroscopy method (described in the previously paragraphs). Such an approach would develop a model able to classify durum wheat, bread, and pasta samples of the investigated varieties *Saragolla*, *Marco Aurelio*, *Maciste*, *Antalis*, and *Iride*.
An NMR interlaboratory comparison (ILC) according to ISO/IEC 17043:2010 (Gallo et al., 2013) was organized for the validation of the non-targeted NMR method developed.

2.7.7.1 Samples collection

Local farmers "Cooperativa la Piramide" (Torremaggiore), "Azienda Agricola Dipalma" (Gravina), "Azienda Agricola Parisi" (Gravina-Poggiorsini), and "Società Agricola Denora" (Gravina-Matera) provided durum wheat samples. These samples belong to the five varieties under investigation, cultivated with the three agronomic practices to be examined: nitrogen regime applied at the dose of 120 kg ha^{-1}, 80 kg ha^{-1}, and organic treatment. Samples were provided during 2021 and 2022.
In addition to the wheat grain of each variety supplied, the corresponding spikes were also harvested directly from the field.

The company PanBiscò (Altamura) was responsible for the supply of monovarietal bread samples. Casa Prencipe (Monte Sant'Angelo) was engaged to produce monovarietal pasta, and delivery the sample; however, several circumstances did not make possible the production of *Saragolla* pasta, excluding these samples in the analysis presented in this work. Table 22 reports the details of all the samples provided and analysed for the project.

Table 22. Summary of samples

Cultivar	**Durum wheat**	**Spikes**	**Bread**	**Pasta**
Saragolla	14 of the 2021 year 8 of the 2022 year	6 of 2021	2	0
Marco Aurelio	14 of the 2021 year 8 of the 2022 year	6 of 2021	2	6 [production batch 071/22] 10 [production batch 069/22]
Maciste	14 of the 2021 year 8 of the 2022 year	6 of 2021	2	11 [production batch 298/21]
Antalis	14 of the 2021 year 8 of the 2022 year	6 of 2021	0	20 [production batch 060/22] 10 [production batch 059/22]
Iride	14 of the 2021 year 8 of the 2022 year	6 of 2021	2	11 [production batch 298/21]

Samples were sent to the Chemistry Laboratory of the Polytechnic of Bari and stored at room temperature until their preparation for NMR analysis. Only the bread samples were subjected to a drying process and stored at -18°C before to perform extraction procedure.
The extraction protocol and the experimental conditions of the NMR analysis described in paragraph 2.7.3. were applied for durum wheat, pasta and bread samples.

2.7.7.2 Evaluation of repeatability of pasta samples

The analytical procedure applied to durum wheat samples has resulted reliable for the analysis of this type of matrix, according to the obtained data reported in paragraph 2.7.4 ("Validation of the method applied"). A repeatability study was also conducted for pasta samples.
For the evaluation of reproducibility, 10 NMR tubes were prepared from the same pasta samples, according to the protocol described in paragraph 2.7.3. The respective NMR spectra are shown in figure 91.
For the evaluation of intermediate precision and repeatability, a bulk solution of pasta was prepared by mixing 3 g of ground and sieved pasta in 30 ml of buffer solution, and then the extraction protocol described in the sub-paragraph 2.7. 4.2 was applied. An NMR tube among those prepared from the bulk solution of pasta was analysed ten times, for the evaluation of repeatability.
This NMR tube was removed and reinserted ten times inside the spectrometer, in an automated way. Thus, ten 1D ^{1}H NOESY spectra of the same tube were recorded. The acquisition parameters of the NMR experiment are indicated in paragraph 2.7.3.4.
The betaine signal (3.25 ppm) was selected for the evaluation of the repeatability. Therefore, the integrals of the spectral area [3.26,3.24 ppm] from the recorded NMR spectra were determined.

The 1D ^{1}H NOESY spectra recorded on the 10 NMR tubes prepared for the reproducibility are shown in figure 91. The integral of the betaine signal (3.25 ppm) is delimited within the yellow rectangle.

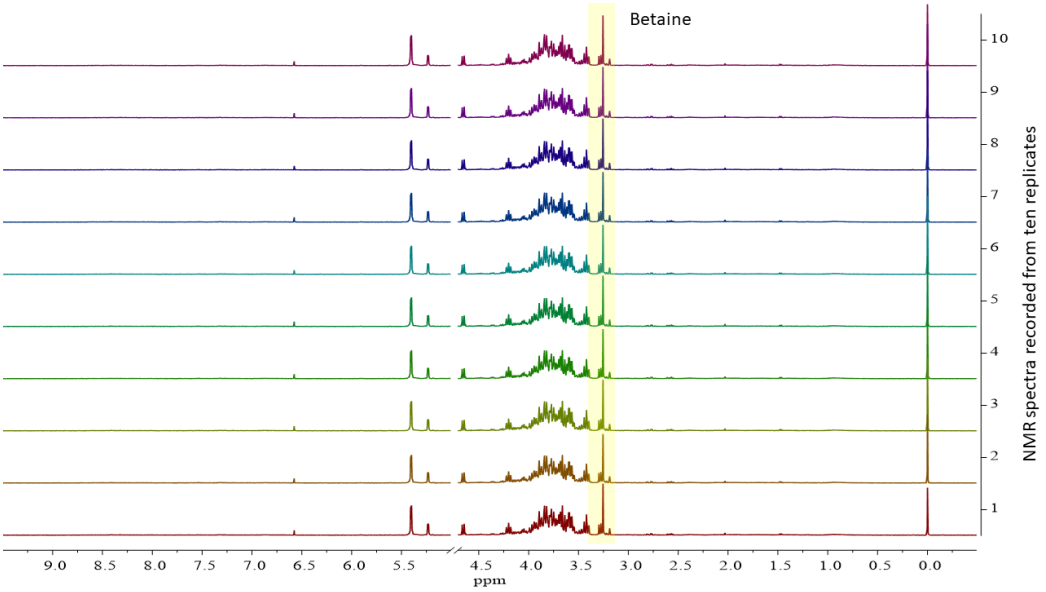


Fig. 91– 1D ^{1}H NOESY spectra of pasta recorded from the ten NMR tubes prepared for the reproducibility evaluation

The data obtained from the integration analysis are collected in table 23.

Table 23. Results of the reproducibility study of pasta

NMR replicates	The integral value of the betaine signal
1	10072.7
2	9508.1
3	10162.4
4	9711.8
5	9830.7
6	9936.6
7	9985.5
8	10172.9
9	9965.9
10	9647.4
Standard deviation	222
Mean	9899.4
CV%	**2.2**

The 1D ^{1}H NOESY spectra recorded acquired on the ten NMR tubes prepared from the bulk solution of pasta are illustrated in figure 92. The integral of the betaine signal (3.25 ppm) is delimited within the yellow rectangle.

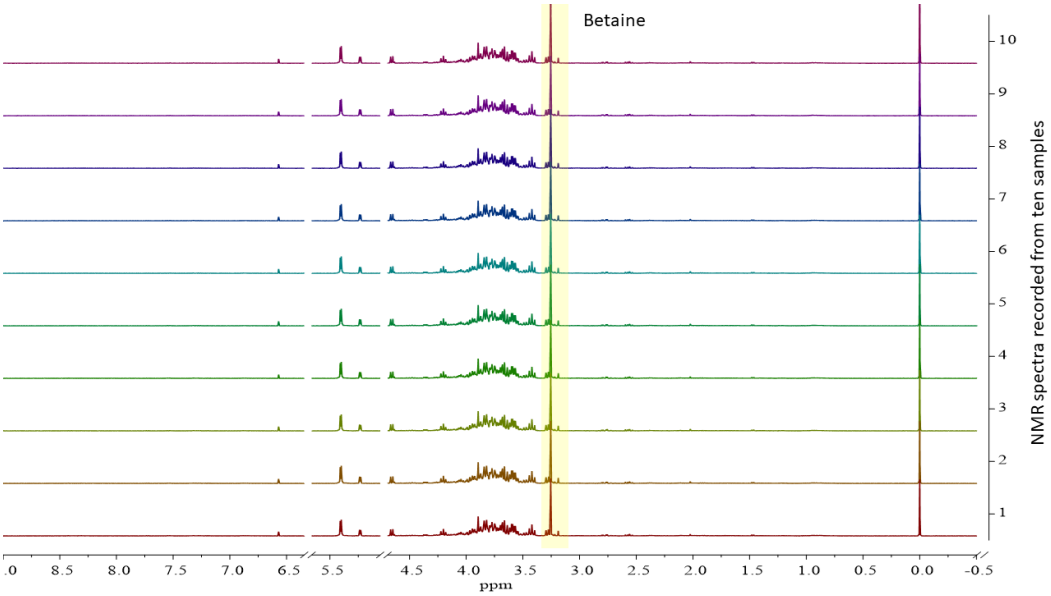


Fig. 92– 1D ^{1}H NOESY spectra of pasta recorded from the same NMR tubes prepared for the intermediate precision evaluation.

The data obtained from the integration analysis are collected in table 24.

Table 24. Results of the intermediate precision study of pasta

NMR tubes	The integral value of the betaine signal
1	11220.2
2	12261.1
3	11340.9
4	11482.6
5	11963.5
6	11757.3
7	11740.8
8	11593.1
9	11647.5
10	11921.1
Standard deviation	308.7
Mean	11692.8
CV%	**2.6**

Figure 93 shows the 10 spectra recorded from the same NMR tube of pasta sample used for the evaluation of repeatability. The yellow rectangle highlights the spectral area of the betaine signal (3.25 ppm), selected for the calculation of integrals.

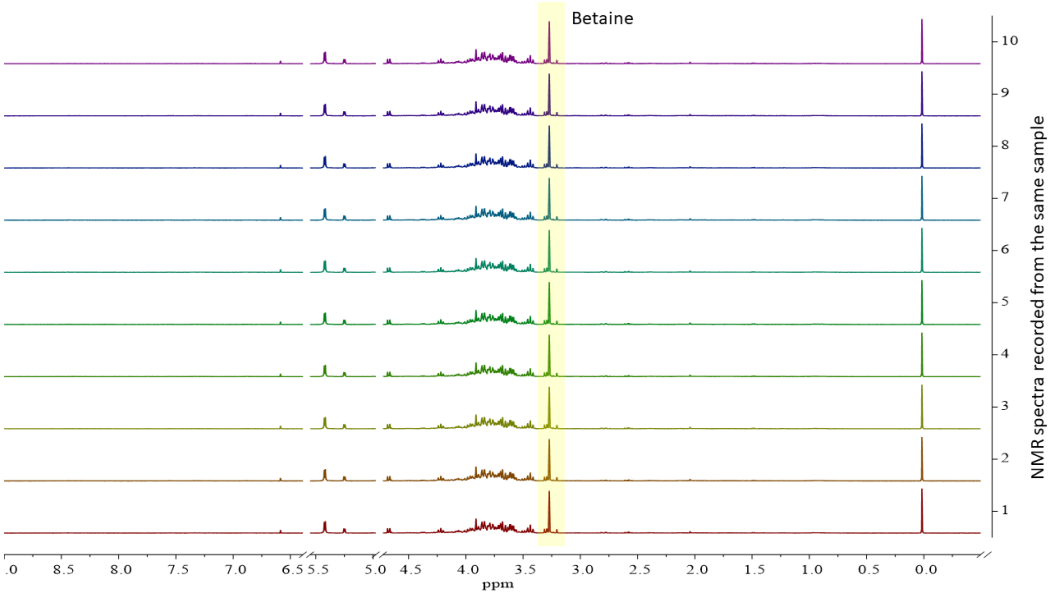


Fig. 93– 1D ^{1}H NOESY spectra of pasta recorded from the same NMR tubes prepared for the repeatability evaluation.

The data obtained from the integration analysis are collected in table 25.

Table 25. Results of the repeatability study of pasta

1D ^{1}H NOESY spectra	The integral value of the betaine signal
1	11587.6
2	11468.5
3	11488.3
4	11481.1
5	11578.8
6	11606.2
7	11588.3
8	11599.8
9	11599.7
10	11620.5
Standard deviation	58.2
Mean	11561.8
CV%	**0.5**

In the evaluation of the reproducibility, intermediate precision and repeatability of the analytical procedure applied to the pasta sample, CV% values of less than 5% were obtained. In particular, the CV% value resulting from the reproducibility study was 2.2%; that of intermediate precision was 2.6%; for repeatability the CV% value calculated was 0.5%. Therefore, the analytical procedure applied to pasta samples resulted not affected by a significative systematic bias.

2.7.7.3 Metabolites characterization of pasta

The NMR analysis allowed for the identification of the main metabolites contained in typical aqueous extracts (pH = 4.2) of pasta samples.

Metabolite identification was performed employing a library of 1D ^{1}H NOESY spectra of pure compounds prepared and recorded with the procedure described for durum wheat samples. The typical 1D ^{1}H NOESY spectrum of an aqueous extract of pasta is shown in figure 94.

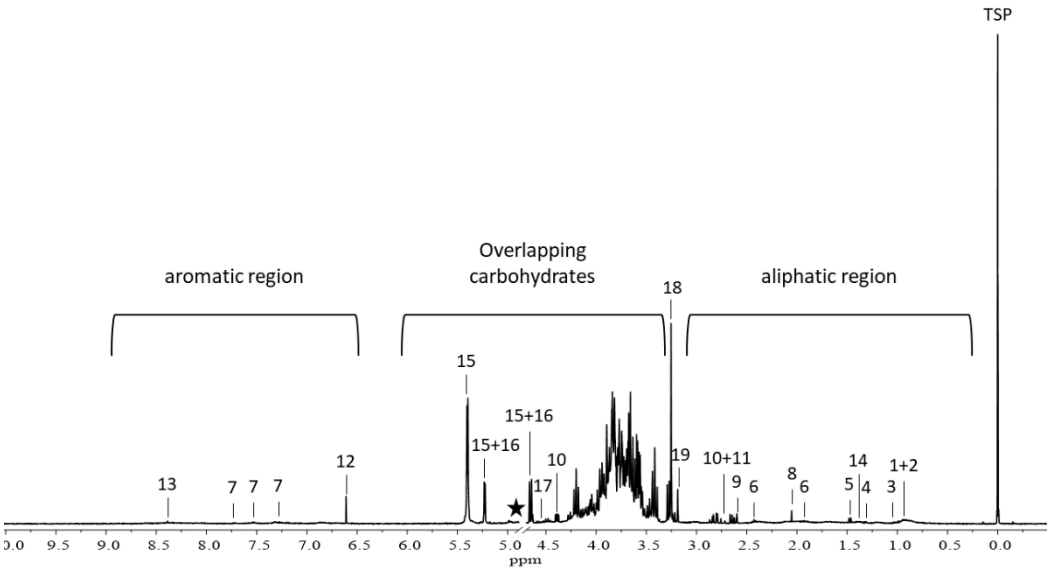


Fig. 94– 1D ^{1}H NOESY spectra of a typical aqueous extract of pasta. The HDO signal is hidden; numbers indicate the metabolites reported in table 24.

The main representative classes of metabolites included organic acids (acetic, succinic, malic, citric, fumaric, formic, and lactic acids), amino acids (isoleucine, leucine, valine, alanine, γ-aminobutyric acid, and Tryptophan), and carbohydrates (maltose, glucose, and galactose). Moreover, 1D ^{1}H NOESY spectra contained also signals related to betaine and choline. The main metabolites identified are listed in table 26.

Table 26. Characteristic metabolites of aqueous extract of pasta

Compound	Chemical shift (ppm)	Multiplicity: J (Hz)
Amino acids		
Isoleucine (1)	0.93	t overlapped
	1.00	d (7.0)
Leucine (2)	0.94	d overlapped
	0.96	d
	1.72	m

Valine (3)	0.98	d (7.0)
	1.03	d (7.0)
	2.27	m
	3.64	d overlapped
Threonine (4)	1.32	d (6.8)
Alanine (5)	1.47	d (7.1)
	3.81	q overlapped
γ-Aminobutyric acid (6)	1.92	m
	2.43	t (7.3)
Tryptophan (7)	7.19	t (7.3)
	7.27	t (7.6)
	7.32	d (2.4)
	7.54	d (8.2)
	7.73	d (8.1)
Organic acids		
Acetic acid (8)	2.05	s
Succinic acid (9)	2.60	s
Malic acid (10)	2.63	dd (15.9; 7.9)
	2.82	dd (15.9; 4.3)
	4.39	dd (7.9; 4.3)
Citric acid (11)	2.74	d (15.3)
	2.85	d (15.3)
Fumaric acid (12)	6.61	s
Formic acid (13)	8.39	s
Lactic acid (14)	1.35	d (7.0)
Carbohydrates		
Maltose (15)	3.27	dd (9.3; 7.9)
	3.42	t (9.4)
	3.58	m
	3.62	m
	3.66	m
	3.70	m
	3.76	m

	3.89	m
	4.65	d (7.9)
	5.23	d (3.7)
	5.40	d (3.8)
Glucose (16)	3.24	dd overlapped
	4.64	d (7.9)
	5.23	d (3.7)
Galactose (17)	4.58	d (7.9)
	5.26	d (3.6)
Quaternary ammonium compounds		
Betaine (18)	3.26	s
	3.90	s
Choline (19)	3.19	s

2.7.8 Interlaboratory Comparison (ILC) organisation

The PhD candidate actively took part in the drawing up the guidelines for Inter-laboratory Comparison, that are here reported.

The international standard UNI CEI EN ISO/IEC 17043:2010 "*Conformity assessment-General requirements for proficiency testing"* specifies general requirements for the competence of the organizers of interlaboratory, applicable to all types of ring test. This document has been prepared by Technical Committee ISO/CASCO "Committee on conformity assessment" in collaboration with Technical Committee CEN/CLC/TC 1 "Criteria for conformity assessment bodies".

Interlaboratory comparisons are widely used for a number of purposes and their use is increasing internationally.

An interlaboratory comparison (ILC) was proposed in this project to validate the non-targeted NMR method developed to classify samples belonging to *Marco Aurelio* and *Iride* varieties, along with the determination of the betaine in durum wheat and pasta using a ^{1}H-qNMR procedure.

Despite the advantages deriving from the use of qNMR and despite the plethora of well documented validation processes (e.g., precision, accuracy, linearity, reproducibility, robustness, selectivity, and specificity), very few official NMR based

quantification methods are known. Validation by ILC is the starting point for official recognition of such analytical method. The durum wheat varieties were chosen based on the distinct intensity of the betaine signal at 3.25 ppm in their recorded NMR spectra. Therefore, this metabolite was selected for the quantification procedure in durum wheat and pasta samples, by using not only an internal reference compound (TSP-d_4) but also the betaine as a pure external standard. The external standard of betaine was added to a mixture of the TSP-d_4, fructose and Dimethylsulphone ($DMSO_2$), a qNMR Certified Reference Material.
The detailed preparation of this mixture solution is described in the section "sample preparation" of the paragraph 2.7.8.2.

2.7.8.1 The scheme of ILC

The ILC scheme was organized according to the following step:

a) call;
b) registration;
c) sample preparation;
d) publication of the experimental instructions for participants;
e) sample delivery;
f) NMR analysis;
g) submission of the results;
h) data elaboration;
i) report publication.

Table 27 shows the timetable for each phase of this ILC.

Table 27. ILC timetable

Period	Description
2022/07/15-2022/10/05	Call open and registration
2022/10/06-2022/11/05	Recognition of the spectrometers available for the ILC and publication of the "Guidelines and contract terms"; sample preparation; NMR sample homogeneity assessment and sample stability control
2022/11/06-2022/11/30	Sample delivery
2022/12/01-2023/01/15	NMR experiment registration and data submission
2023/01/16-2023/02/10	Data elaboration
2023/02/11-2023/02/28	Publication of the report

a. Call

The call made known the aim of the ILC, the scientific committee, the sample matrix and the NMR experiment selected for the ILC. Technical requirements of the participants were also indicated.

The call of the proposed ILC was published on 2022/07/15 and the deadline for registration was on 2022/10/05.

b. Registration

For the present ILC, registration was free and after the deadline, the number of registered participants was 40 for a total of 47 spectrometers.

Table 28.

Magnetic Field	NMR spectrometers: 47
700 MHz	5
600 MHz	15
500 MHz	4
400 MHz	18
300 MHz	1
F80	3

c. Samples preparation

The NMR tubes prepared were eight for each participant. Tubes were labelled with T, A, B, C, E, F, and QR.

The aqueous solution of the standard was prepared by dissolving 19.355 mg of TSP-d_4 (CAS No. 24493-21-8, 99 %D, Armar Chemicals, Döttingen, Switzerland),19.281 mg of certified powder of Betaine (CAS: 107-43-7; Lot. no: LRAC 5297 EXP: Feb/24), 19.719 mg of certified powder of the standard $DMSO_2$ (CAS: 107-43-7; Lot. no: LRAC 5297 EXP: Feb/24), and 1.00011 g of D (-)-Fructose (Fructose 99% CAS: 57-48-7 Lot: A0282904 Product of Belgium) in 100 ml of [$HC_2O_4^-/C_2O4_2^-$] = 0.25 M buffer pH=4.2 containing sodium azide (NaN_3 = $2.5 \cdot 10^{-3}$ M, CAS No. 26628-22-8; ≥99.5%, Sigma-Aldrich, Milan, Italy).

The aqueous solution was used to fill the QR tube.

To prepare the bulk solution of durum wheat and pasta (*Marco Aurelio* and *Iride varieties*), 6 g of each ground and sieved matrix was solubilized to 60 ml of buffer [$HC_2O_4^-/C_2O4_2^-$] = 0.25 M buffer pH=4.2 containing sodium azide (NaN_3 = $2.5 \cdot 10^{-3}$ M, CAS No. 26628-22-8; ≥99.5%, Sigma-Aldrich, Milan, Italy).

Then, sonication was conducted through an ultrasonic bath at 40 KHz at 60°C for 10 minutes; after being vortexed at 2500 rpm for 5 minutes, each solution was centrifuged at 4000 rpm for 10 minutes.

Supernatants were collected, transferred into two new bottles, and pasteurised at 90°C for 10 minutes. After that, the solutions were centrifuged and filtered by centrifuged filter in PVDF (polyvinylidenefluoride, with a pore size of 0.2 um) at 4000 rpm for 5 minutes. Each filtered solution was stored overnight at 4°C.
The aqueous solutions of durum wheat and pasta obtained from the *Marco Aurelio* variety were numbered with 1, whereas the ones produced from the *Iride* variety were denoted by the number 2.
The four filtered solutions were used to prepare NMR tubes A, B, C, D, E, and F. The sample composition of each tube is specified in table 29.

Table 29. List of NMR tubes prepared for ILC along with the corresponding sample composition

Tube	Solutes
T	600 µl of Methanol-d_4 (≥ 99.8 atom % D)
QR	630 µl of aqueous solution of standard + 70 µl of D_2O (99.86 % D)
A	630 µl of pasta 2 aqueous solution + 70 µl of 0.2 wt% of TSP-d_4/D_2O solution
B	630 µl of pasta 1 aqueous solution + 70 µl of 0.2 wt% of TSP-d_4/D_2O solution
C	630 µl of pasta 1 aqueous solution + 70 µl of D_2O (99.86 % D)
D	630 µl of durum wheat 1 aqueous solution + 70 µl of D_2O (99.86 % D)
E	630 µl of durum wheat 1 aqueous solution + 70 µl of 0.2 wt% of TSP-d_4/D_2O solution
F	630 µl of durum wheat 2 aqueous solution + 70 µl of 0.2 wt% of TSP-d_4/D_2O solution

It was observed that the signal intensity of the TSP may be affected within the matrix analysed, probably due to the potential bounds with protein. Hence, to

investigate the behaviour of the TSP, it was not included in all the prepared NMR tubes. The NMR tubes A, B, E, and F (filled by aqueous extract of pasta 2, pasta 1, durum wheat 1, durum wheat 2, respectively), contained the 0.2 wt% TSP-d_4/ D_2O solution, whereas in NMR tube C, and D (filled by aqueous extract of pasta 1, and durum wheat 1 respectively), TSP-d_4 was not used as a reference compound.

The homogeneity of the entire batch of samples, was verified before sending them to the participants.

d. Publication of the experimental instructions for participants

Participants acquired all the spectra under manual or automatic procedure by following the steps listed below.

Step 1: Temperature calibration using the tube T

Use T tube as a nuclear magnetic resonance thermometer to evaluate the correct calibration of the temperature control unit (TCU). It was suggested to use the following procedure reported in the literature (Findeisen, *et al.* 2007).

On the Temperature Control Unit (TCU) of the spectrometer, set the sample temperature at 298 K. Record a routine ^1H spectrum by 90° single excitation pulse sequence ("s2pul" for Agilent/Varian; "zg" for Bruker; "proton" for Jeol).

Consider the two signals at approximately 4.8 and 3.3 ppm and calculate the difference in chemical shifts (Δδ). Use the Δδ value to calculate the sample temperature T [K] according to the following equation:

$$T = -16.7467 \cdot (\Delta\delta)^2 - 52.5130 \cdot \Delta\delta + 419.1381$$

If the calculated T is different from 298 K, adjust the TCU to obtain a value of Δδ suitable to reach T = 298.0 ± 0.1 K. Once optimized, use the TCU setup for all further measurements described in the next paragraph.

Temperature optimization is not required for every session.

Step 2: Experiment setup

Experiment parameters were setting according to the following procedures, adapted to the technical specifications of the different spectrometers.

Each participant set up the experiment using the most suitable set of parameters for the spectrometer in use, as indicated in table 30. In this way, all FIDs acquired by all participants had the same duration (acquisition time) and digital resolution

Table 30. List of acquisition parameters suggested according to the magnetic field of the spectrometer in use.

Field (MHz)	**80**	**300**	**400**	**500**	**600**	**700**
Time Domain (points)	10240	38400	51200	64000	76800	89600
Spectra Width (ppm)	15	15	15	15	15	15
Spectra Width (Hz)	1200	4500	6000	7500	9000	10500
Dwell time (µs)	416.7	111.1	83.3	66.7	55.6	47.6
Acquisition time (s)	4.267	4.267	4.267	4.267	4.267	4.267
Digital Resolution (Hz)	0.1172	0.1172	0.1172	0.1172	0.1172	0.1172

Samples were measured in accordance with the operating manual of the NMR spectrometer in use. In case of measurements obtained in different sessions, the QR tube was analysed before each session.

Step 3: Naming the individual samples

Each sample was named according to this scheme: “spectrometer code assigned to lab_sample ID_repetition number_date”. The spectrometer code was assigned to each participant by Innovative Solutions and communicated along with a username and password. The sample ID was indicated on the label applied to each NMR tube contained in the delivered package.

For instance, for the spectrometer numbered “Lab01”, the file/dataset names of the five repetitions of the tube A01 (the label was on the tube) recorded on 5th December 2022, was named as follows:

Lab01_A01_R01_20221205 (repetition 1)

Lab01_A01_R02_20221205 (repetition 2)
Lab01_A01_R03_20221205 (repetition 3)
Lab01_A01_R04_20221205 (repetition 4)
Lab01_A01_R05_20221205 (repetition 5)
The same procedure was followed for all A-F, and QR tubes.

Step 4: Spectra registration (tubes A-F, and QR)
For each NMR tube, 5 spectra were recorded to comply with conditions for repeatability, i.e. same NMR tube, same spectrometer, same user, consecutive repetitions. For each tube, participants followed the instructions reported below:
- load the sample in the magnet;
- wait for the temperature stabilization at 298.0 ± 0.1 K for at least 5 min;
- create a new dataset using the experiment setup saved in step 2;
- lock on solvent signal: H_2O:D_2O (90:10 V/V) for tubes A-F and QR; CD_3OD for tube T;
- perform the tuning and matching of the probe;
- adjust the field homogeneity (shimming) so that the width of the 3.25 ppm signal (betaine) at half-height was < 2.5 Hz;
- perform 5 consecutive measurements of the same sample to obtain five repetitions.

Step 5. Processing parameters for 1D ^{1}H NOESY experiments (tubes A-F, and QR)
Double the FID size by zero filling. Fourier transform by applying exponential multiplication function with a line broadening of 0.1 Hz. Chemical shift scale was referenced to the betaine resonance (3.25 ppm). Correct phase and baseline (user chose either manual or automatic procedures without any limitations provided that the same procedure was applied to all NMR spectra).
For tubes A, B, E, F, and QR, the integral area of the signals listed below were calculated. Integration consisted of the calculation of peak area in the range reported in the following table.

Table 31. Integral range of the interested signals

Signal label	Integration range	
	Left limit (ppm)	Right limit (ppm)
TSP	0.10	-0.10
Betaine	3.26	3.24

Step 6. Data and spectra submission on website https://nmr.istracer.it

Please, log in with their user name and password on the https://nmr.istracer.it website. Click on 'Data input' and select 'NMR Interlaboratory Comparison IS-NMR-ILC 01_2022 - Tubes A-F and QR'. On the “Data input” page (figure 95), upload a single zip folder containing all recorded NMR spectra named according to step 2 (see above). The total number of spectra must be 36 [5 spectra * 7 tubes (A-F, and QR) + 1 spectrum for tube T].

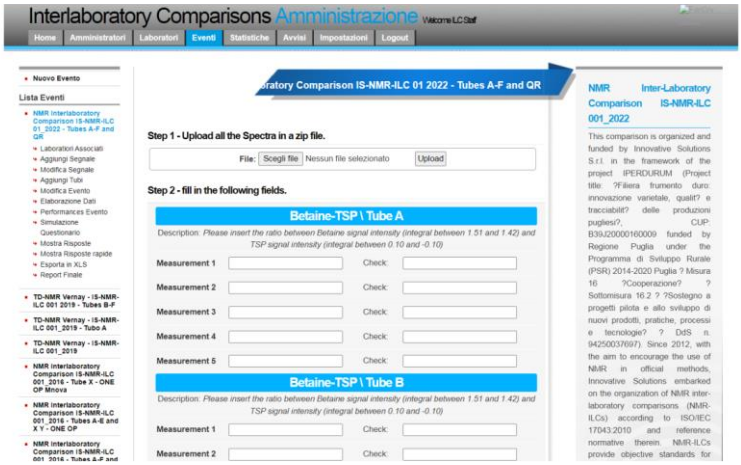


Fig.-95

Next, fill in the form with the values of the integral ratio $I_{Betaine}I_{TSP}$ where $I_{Betaine}$ is the area under the signal in the range 3.26 ÷ 3.24 ppm and I_{TSP} is the area under the signal in the range 0.10 ÷ -0.10 ppm. Use at least 8 decimal places. Please, note that “Measurement” stands for “Repetition”. In the “Check” field re-enter the value of the corresponding measurement (see example in Figure 96).

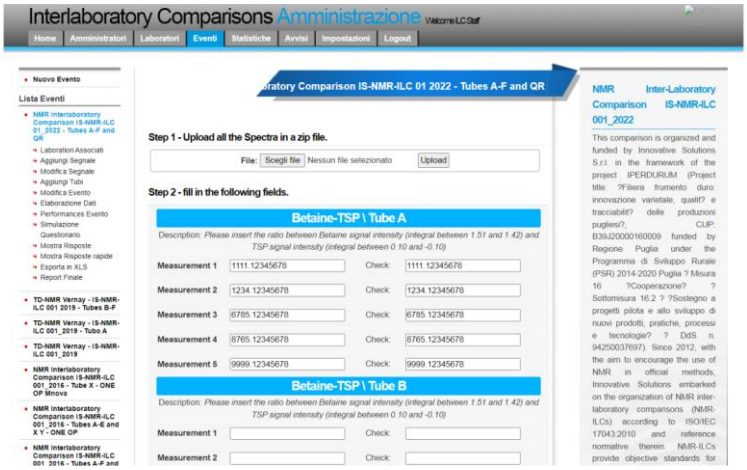


Fig.-96

Consider proceeding with partial data input. Once you have filled in a block of data (e.g. “Betaine-TSP \ Tube A), go to the bottom of the page (Figure 97), select “I agree with contract terms” and click on “Save and submit”. In this way, you will save your work. If you do not save frequently, you may lose your data due to webpage timeout limitations and you have to restart data entry.

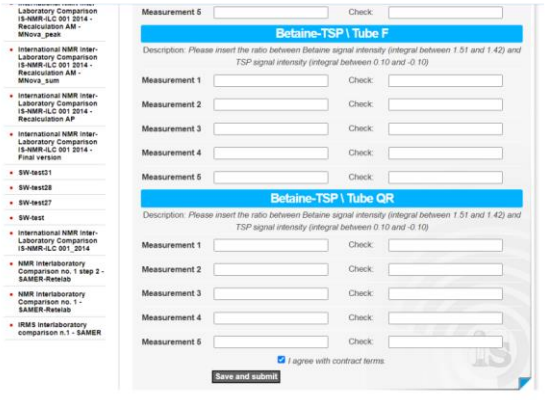


Fig.97

e. Sample delivery

The eight NMR tubes (Norell 509-UP 7) labelled as A, B, C, D, E, F, T and QR were sealed and placed within a fitting package, to guarantee the integrity of materials during transport. The homogeneity of the entire batch of samples was verified before sending them to the participants.
Each package was sent by an express courier.
The participants received the package and discovered the secret code assigned not related to the order of their registration.

f. NMR analysis
Samples will be analyzed by participants, in accordance with methods and time-tables provided by protocol and instructions. Each deviation from instructions, if any, must be declared to coordinator before the closure of comparison.

g. Submission of the results
Submission of the results will be possible by online procedure using the form on website. After deadline, result submission is no longer possible.

h. Data elaboration
NMR data are submitted to statistical processing, in accordance with the appropriate statistical principles of reference normative.

i. Report publication
Results of the statistical elaboration of the NMR will be published on the website within 28/02/2023.

2.7.8.2 Homogeneity and stability test

The PhD candidate was involved in the homogenisation of the ILC and conducted the homogeneity and stability tests reported below.
According to technical requirements of ISO 17043:2010 , the homogeneity of the samples and the stability of the investigated matrices was checked. This

procedure is aimed to ascertain, before sample delivery, the negligibility of the variation in composition with respect to variation introduced by the measurements conducted by participants.

An NMR tube prepared from the bulk solution of durum wheat (the preparation of the bulk solution is described in the section c. of the sub-paragraph 2.7.8.1) was sealed and left at room temperature, to record 1D ^{1}H NOESY spectra after 48 hours, 4 days, 6 days, and 13 days. The processed NMR spectra recorded at the time points specified are shown in figure 98. The NMR spectrum acquired on the freshly NMR tube of wheat (0 h) is represented by the red spectrum. The spectrum recorded after 48 hours, is represented by the blue one; those acquired after 4,6, and 13 days are indicated by the green, orange, and violet spectra, respectively.

The stability of the sample was assessed by integrating the areas of three signals (S1, S2, and S3 in fig.98), selected in the series of NMR spectra recorded in the time points indicated.

The chosen signals were: S1, relative to the doublet of the alanine (1.48 ppm); S2, related to the singlet of betaine (3.25 ppm); S3 related to a doublet of the tryptophan signals pattern (7.52 ppm).

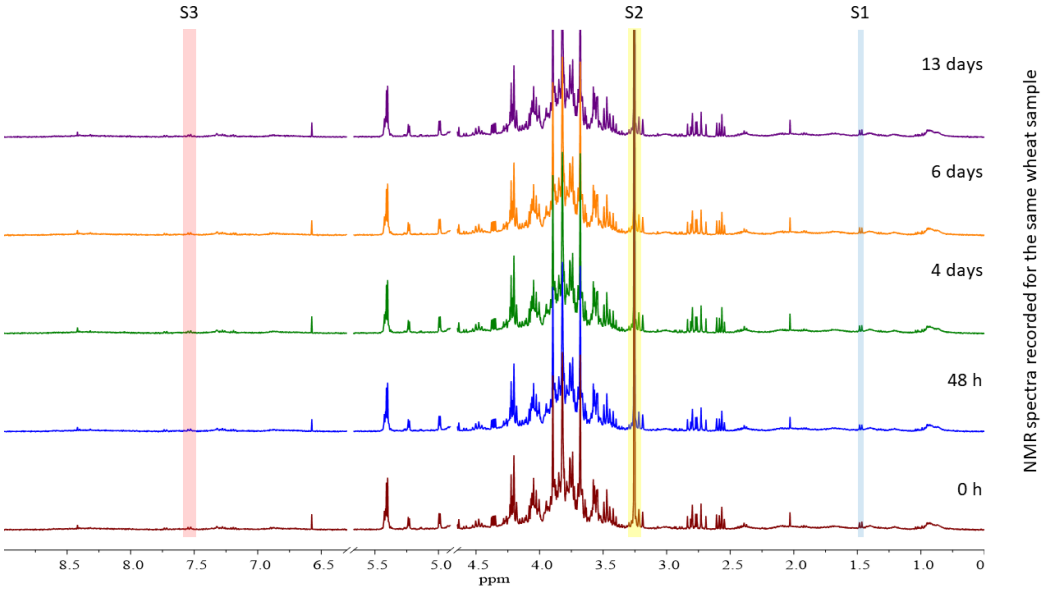


Fig. 98– 1D ^{1}H NOESY spectra of the same durum wheat sample recorded at the different time points for the stability evaluation.

For signal S1, the spectral area considered for the analysis of the integrals was in the range [1.45,1.50 ppm]; for signal S2, the integration range was [3.24,3.26 ppm]; for signal S3, the integration range was [7.51,7.55 ppm]. The integrals were expressed in absolute values. Then, the following statistical parameters were calculated: mean (μ), standard deviation (σ) and, the coefficient of variation percentage (CV%). The data obtained from those calculations are collected in table 32.

Table 32. Results of the stability check of durum wheat sample

	S1 integrals	S2 integrals	S3 integrals
Integration range (ppm)	**[1.50,1.45]**	**[3.26,3.24]**	**[7.55,7.51]**
Time points	**Integral values**		
0 h	873.7	17264	315.6
48 h	845.6	16159.6	325.3
4 days	926.2	16959.6	342.8
6 days	911.6	16989.5	338.5
13 days	875.9	17070.3	330.4
Standard deviation	32.2	424.4	10.8
Mean	886.6	1688.6	330.5
CV%	**3.6**	**2.5**	**3.3**

The variation percentage coefficient was considered an index to validate the measurements of stability assessment. The calculated CV% values for the three selected signals were below 5% (3.6 for S1, 2.5 for S2, and 3.3 for S3). As a result, durum wheat can be considered a very stable matrix for NMR analysis, without noticeable changes in its chemical composition for up to 13 days.

Moreover, the stability of the pasta samples over two weeks was evaluated.

An NMR tube prepared from the bulk solution of pasta (the preparation is described in the section c. of the sub-paragraph 2.7.8.1) was sealed and left at room temperature, to record 1D ^{1}H NOESY spectra after 3, 7, and 14 days.

NMR spectra were acquired according to the experimental parameters reported in paragraph 2.7.3.4. The processed NMR spectra recorded at the time points specified are shown in figure 99.

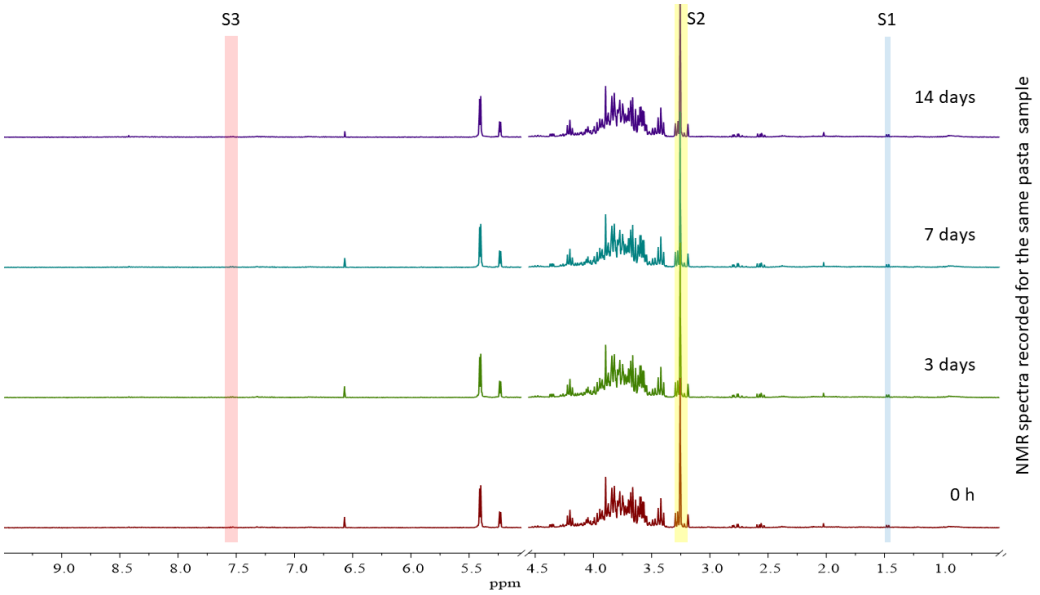


Fig. 99 1D ^{1}H NOESY spectra of the same pasta sample recorded at the different time points for the stability evaluation.

The NMR spectrum acquired on the freshly NMR tube of pasta (0 h) is represented by the red spectrum. The spectrum recorded after 3 days, is represented by the green one; those acquired after 7, and 14 days are indicated respectively by the blue and violet ones.

For the evaluation of the stability of the pasta sample, the same signals selected for wheat stability were integrated (S1, S2, and S3 in fig.99).

For signal S1 (doublet of alanine), the spectral area considered for the analysis of the integrals was in the range [1.45,1.50 ppm]; for signal S2 (singlet of betaine), the integration range was [3.24,3.26 ppm]; for signal S3 (doublet of tryptophan), the integration range was [7.51,7.55 ppm]. The integrals were expressed in absolute values.

The data obtained from those calculations are collected in table 33.

Table 33. Results of the stability of the pasta sample

	S1 integrals	S2 integrals	S3 integrals
Integration range (ppm)	**[1.50,1.45]**	**[3.26,3.24]**	**[7.55,7.51]**
Time points	**Integral values**		
0 h	541.7	11408.7	215.9
3 days	550.2	11831.6	217.3
7 days	592.16	11855.4	193.9
14 days	627.9	12394.4	193.7
Standard deviation	39.9	403.9	13.2
Mean	577.9	11872.5	205.2
CV%	**6.9**	**3.4**	**6.4**

The CV% values calculated for signals S1 and S3 resulted in being higher than 5%. On the other hand, the CV% of signal S2 of betaine was below 5%.

These procedures have guaranteed the stability control of the samples before batch delivery to ILC participants.

CONCLUSIONS

3.0 Conclusions

In this thesis, effective applications of a non-targeted NMR approach in food chemistry were described. Such an approach was applied to a group of authentic saffron samples and on a group of saffron samples intentionally adulterated by the addition of safflower and turmeric at different concentrations. The elaboration of NMR data by chemometrics allowed the discrimination between authentic and adulterated saffron samples, detecting the presence of both adulterants at a concentration value equal to 5 wt% in aqueous extracts, while on the same samples in DMSO-d6, the presence of turmeric was detected at a concentration level of 5%, and it was possible to identify the presence of safflower only at higher levels of concentration, i.e., 20 wt%. Furthermore, through the same analytical approach, it was possible to obtain crucial information on the type of cultivation practice adopted for the production of saffron.

A non-targeted NMR approach was successfully applied to assess the authenticity of Italian lentils and to ensure the geographical origin, in a way to discriminate this class of product from the imported ones. The developed model was capable to predict the geographical origin of blind lentils' samples 99.41 % correctly.

The same analysis was performed on samples of durum wheat of authentic origin. The obtained results have demonstrated the efficiency of the non-targeted NMR method for certain the geographical origin of durum wheat. The developed model has shown an ability to discriminate the authentic Italian samples from the foreign ones was, and to correctly predict the geographical origin of a test set by 98.8 %.

Moreover, an Interlaboratory Comparison has been organised to validate the method applied for the classification of durum wheat samples, as well as the validation of a non-targeted NMR method for the classification of pasta samples and the validation of a qNMR method using internal and external standards.

The results of these studies can help to enhance the application of non-targeted NMR methods in the control processes of the agri-food supply chain with the final goal to ensure traceability and authenticity of food products.

REFERENCES

4.0 References

Abdi, H. and Williams, L. J. (2010) 'Principal component analysis', *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), pp. 433–459. doi: 10.1002/wics.101.

Akrami-Mohajeri, F. *et al.* (2018) 'The prevalence and antimicrobial resistance of Listeria spp in raw milk and traditional dairy products delivered in Yazd, central Iran (2016)', *Food and Chemical Toxicology*. doi: 10.1016/j.fct.2018.02.006.

Alshikh, N., de Camargo, A. C. and Shahidi, F. (2015) 'Phenolics of selected lentil cultivars: Antioxidant activities and inhibition of low-density lipoprotein and DNA damage', *Journal of Functional Foods*. doi: 10.1016/j.jff.2015.05.018.

Amarowicz, R. *et al.* (2009) 'Antioxidant activity of a red lentil extract and its fractions', *International Journal of Molecular Sciences*. doi: 10.3390/ijms10125513.

Anderssen, R. S. and Haraszi, R. (2009) 'Characterizing and exploiting the rheology of wheat hardness', *European Food Research and Technology*, 229(1), pp. 159–174. doi: 10.1007/s00217-009-1037-9.

Di Anibal, C. V. *et al.* (2009) 'Determining the adulteration of spices with Sudan I-II-II-IV dyes by UV-visible spectroscopy and multivariate classification techniques', *Talanta*. doi: 10.1016/j.talanta.2009.05.023.

Anklam, E. and Battaglia, R. (2001) 'Food analysis and consumer protection', *Trends in Food Science and Technology*, 12(5–6), pp. 197–202. doi: 10.1016/S0924-2244(01)00071-1.

Aung, M. M. and Chang, Y. S. (2014) 'Traceability in a food supply chain: Safety and quality perspectives', *Food Control*. doi: 10.1016/j.foodcont.2013.11.007.

Bacchi, M. *et al.* (2010) 'Agronomic evaluation and genetic characterization of different accessions in lentil (Lens culinaris Medik.)', *Italian Journal of Agronomy*, 5(4), pp. 303–314. doi: 10.4081/ija.2010.303.

Bachmann, R. *et al.* (2018) '1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts', *Journal of Agricultural and Food Chemistry*. doi: 10.1021/acs.jafc.8b03724.

Ballin, N. Z. and Laursen, K. H. (2019) 'To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication', *Trends in Food Science and Technology*. Elsevier Ltd, pp. 537–543. doi: 10.1016/j.tifs.2018.09.025.

Bhave, M. and Morris, C. F. (2008) 'Molecular genetics of puroindolines and related genes: Allelic diversity in wheat and other grasses', *Plant Molecular Biology*, 66(3), pp. 205–219. doi: 10.1007/s11103-007-9263-7.

Blanco, A., Simeone, R. and Gadaleta, A. (2006) 'Detection of QTLs for grain protein content in durum wheat', *Theoretical and Applied Genetics*, 112(7), pp. 1195–1204. doi: 10.1007/s00122-006-0221-6.

Boccard, J. and Rutledge, D. N. (2013) 'A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion', *Analytica Chimica Acta*. doi: 10.1016/j.aca.2013.01.022.

Breitmaier, E. (2003) *Short Introduction to Basic Principles and Methods*, *Structure Elucidation by NMR in Organic Chemistry*. doi: 10.1002/0470853069.ch1.

Brereton, R. G. and Lloyd, G. R. (2018) 'Partial least squares discriminant analysis for chemometrics and metabolomics: How scores, loadings, and weights differ according to two common algorithms', *Journal of Chemometrics*. doi: 10.1002/cem.3028.

Bro, R. and Smilde, A. K. (2014) 'Principal component analysis', *Analytical Methods*. doi: 10.1039/c3ay41907j.

Browne, R. A. and Brindle, K. M. (2007) '1H NMR-based metabolite profiling as a potential selection tool for breeding passive resistance against Fusarium head blight (FHB) in wheat', *Molecular Plant Pathology*, 8(4), pp. 401–410. doi: 10.1111/j.1364-3703.2007.00400.x.

Bylesjö, M. *et al.* (2006) 'OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification', *Journal of Chemometrics*. doi: 10.1002/cem.1006.

Caballero-Ortega, H. *et al.* (2004) 'Chemical composition of saffron (Crocus sativus L.) from four countries', *Acta Horticulturae*, 650(October 2014), pp. 321–326. doi: 10.17660/ActaHortic.2004.650.39.

Cagliani, L. R. *et al.* (2015) 'NMR investigations for a quality assessment of Italian PDO saffron (Crocus sativus L.)', *Food Control*. Elsevier Ltd, 50, pp. 342–348. doi: 10.1016/j.foodcont.2014.09.017.

Cagliani, L. R., Scano, P. and Consonni, R. (2017) 'NMR Spectroscopy', in *Spectroscopic Methods in Food Analysis*. doi: 10.5005/jp/books/11222_6.

Campos-Vega, R., Loarca-Piña, G. and Oomah, B. D. (2010) 'Minor components of pulses and their potential impact on human health', *Food Research International*. Elsevier Ltd, 43(2), pp. 461–482. doi: 10.1016/j.foodres.2009.09.004.

Chevallier, S. *et al.* (2006) 'Application of PLS-DA in multivariate image analysis', in *Journal of Chemometrics*. doi: 10.1002/cem.994.

Consonni, R. *et al.* (2010) 'Evaluation of the production year in Italian and Chinese tomato paste for geographical determination using O2PLS models', *Journal of Agricultural and Food Chemistry*. doi: 10.1021/jf100949k.

Consonni, R. and Cagliani, L. R. (2010) 'Nuclear Magnetic Resonance and Chemometrics to Assess Geographical Origin and Quality of Traditional Food Products', in *Advances in Food and Nutrition Research*. doi: 10.1016/S1043-4526(10)59004-1.

Creydt, M. and Fischer, M. (2020) 'Food authentication in real life: How to link nontargeted approaches with routine analytics?', *Electrophoresis*, pp. 1–15. doi: 10.1002/elps.202000030.

Cubero-Leon, E., Peñalver, R. and Maquet, A. (2014) 'Review on metabolomics for food authentication', *Food Research International*. doi: 10.1016/j.foodres.2013.11.041.

Danezis, G. P. *et al.* (2016) 'Food authentication: Techniques, trends & emerging approaches', *TrAC - Trends in Analytical Chemistry*. doi: 10.1016/j.trac.2016.02.026.

Derakhshan, Z. *et al.* (2018) 'Survey on the effects of electron beam irradiation on chemical quality and sensory properties on quail meat', *Food and Chemical Toxicology*. doi: 10.1016/j.fct.2017.12.015.

Dimitrakopoulou, M. E. and Vantarakis, A. (2021) 'Does Traceability Lead to Food Authentication? A Systematic Review from A European Perspective', *Food Reviews International*. doi: 10.1080/87559129.2021.1923028.

Dueñas, M., Hernández, T. and Estrella, I. (2002) 'Phenolic composition of the cotyledon and the seed coat of lentils (Lens culinaris L.)', *European Food Research and Technology*, 215(6), pp. 478–483. doi: 10.1007/s00217-002-0603-1.

Ehdaie, B. and Waines, J. G. (2001) 'Sowing date and nitrogen rate effects on dry matter and nitrogen partitioning in bread and durum wheat', *Field Crops Research*, 73(1), pp. 47–61. doi: 10.1016/S0378-4290(01)00181-2.

Esparza, I. *et al.* (2020) 'Stability of phenolic compounds in grape stem extracts', *Antioxidants*. doi: 10.3390/antiox9080720.

Fan, K. and Zhang, M. (2019) 'Recent developments in the food quality detected by non-invasive nuclear magnetic resonance technology', *Critical Reviews in Food Science and Nutrition*. doi: 10.1080/10408398.2018.1441124.

FAO (2021) 'Food fraud-Intention, detection and management. Food safety technical toolkit for Asia and the Pacific No. 5. Bangkok', pp. 1–44. Available at: https://www.fao.org/3/cb2863en/cb2863en.pdf.

Findeisen, M. (2007) 'Related compounds', *Magnetic Resonance in Chemistry*, 45(1), pp. 175–178. doi: 10.1002/mrc.

Flagella, Z. (2006) 'Qualità nutrizionale e tecnologica del frumento duro', *Ital. J. Agron*, 1(January), pp. 203–239.

Gadaleta, A. *et al.* (2022) 'Durum Wheat Response to Organic and Mineral Fertilization with Application of Different Levels and Types of Phosphorus-Based Fertilizers', *Agronomy*, 12(8). doi: 10.3390/agronomy12081861.

Gallo, V. *et al.* (2014) 'Effects of agronomical practices on chemical composition of table grapes evaluated by NMR spectroscopy', *Journal of Food Composition and Analysis*. Academic Press Inc., 35(1), pp. 44–52. doi: 10.1016/j.jfca.2014.04.004.

Gallo, V. *et al.* (2015) 'Performance Assessment in Fingerprinting and Multi Component Quantitative NMR Analyses', *Analytical Chemistry*. American Chemical Society, 87(13), pp. 6709–6717. doi: 10.1021/acs.analchem.5b00919.

Gallo, V. *et al.* (2020) 'A Contribution to the Harmonization of Non-targeted NMR Methods for Data-Driven Food Authenticity Assessment', *Food Analytical Methods*. Food Analytical Methods, 13(2), pp. 530–541. doi: 10.1007/s12161-019-01664-8.

Ganesan, K. and Xu, B. (2017) 'Polyphenol-rich lentils and their health promoting effects', *International Journal of Molecular Sciences*. doi: 10.3390/ijms18112390.

Gao, B. *et al.* (2019) 'Opportunities and Challenges Using Non-targeted Methods for Food Fraud Detection', *Journal of Agricultural and Food Chemistry*, 67(31), pp. 8425–8430. doi: 10.1021/acs.jafc.9b03085.

Ghorbani, M. (2007) 'The economics of Saffron in Iran', in *Acta Horticulturae*. doi: 10.17660/ActaHortic.2007.739.42.

Giannakoula, A. E. *et al.* (2012) 'The effects of plant growth regulators on growth, yield, and phenolic profile of lentil plants', *Journal of Food Composition and Analysis*. doi: 10.1016/j.jfca.2012.06.005.

Godelmann, R. *et al.* (2013) 'Targeted and nontargeted wine analysis by 1H NMR spectroscopy combined with multivariate statistical analysis. differentiation of important parameters: Grape variety, geographical origin, year of vintage', *Journal of Agricultural and Food Chemistry*. doi: 10.1021/jf400800d.

Le Gouis, J., Oury, F. X. and Charmet, G. (2020) 'How changes in climate and agricultural practices influenced wheat production in Western Europe', *Journal of Cereal Science*. Elsevier Ltd, 93(March), p. 102960. doi: 10.1016/j.jcs.2020.102960.

Graham, S. F. *et al.* (2009) 'Application of NMR based metabolomics for mapping metabolite variation in European wheat', *Metabolomics*, 5(3), pp. 302–306. doi: 10.1007/s11306-008-0154-y.

Gromski, P. S. *et al.* (2015) 'A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding', *Analytica Chimica Acta*. doi: 10.1016/j.aca.2015.02.012.

Grunert, K. G. (2002) 'Current issues in the understanding of consumer food choice', in *Trends in Food Science and Technology*. doi: 10.1016/S0924-2244(02)00137-1.

Grunert, K. G. (2005) 'Food quality and safety: Consumer perception and demand', *European Review of Agricultural Economics*, 32(3), pp. 369–391. doi: 10.1093/eurrag/jbi011.

Gulewicz, P. *et al.* (2000) 'Simple method of isolation and purification of α-galactosides from legumes', *Journal of Agricultural and Food Chemistry*, 48(8), pp. 3120–3123. doi: 10.1021/jf000210v.

Hagh-Nazari, S. and Keifi, N. (2007) 'SAFFRON AND VARIOUS FRAUD MANNERS IN ITS PRODUCTION AND TRADES', *Acta Horticulturae*, (739), pp. 411–416. doi: 10.17660/ActaHortic.2007.739.54.

Hatzakis, E. (2019) 'Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review', *Comprehensive Reviews in Food Science and Food Safety*. doi: 10.1111/1541-4337.12408.

Hong, E. *et al.* (2017) 'Modern analytical methods for the detection of food fraud and adulteration by food category', *Journal of the Science of Food and Agriculture*, 97(12), pp. 3877–3896. doi: 10.1002/jsfa.8364.

Hong, I. H. *et al.* (2011) 'An RFID application in the food supply chain: A case study of convenience stores in Taiwan', *Journal of Food Engineering*. doi: 10.1016/j.jfoodeng.2011.04.014.

Hoover, R. *et al.* (2010) 'Composition, molecular structure, properties, and modification of pulse starches: A review', *Food Research International*. Elsevier Ltd, 43(2), pp. 399–413. doi: 10.1016/j.foodres.2009.09.001.

Hu, Y. *et al.* (2017) 'Application of nuclear magnetic resonance spectroscopy in food adulteration determination: The example of Sudan dye i in paprika powder', *Scientific Reports*. doi: 10.1038/s41598-017-02921-8.

Ingrassia, M. *et al.* (2017) 'Traceability and labelling of food products from the consumer perspective', *Chemical Engineering Transactions*, 58, pp. 865–870. doi: 10.3303/CET1758145.

Irakli, M. *et al.* (2021) 'Genotypic and environmental effect on the concentration of phytochemical contents of lentil (Lens culinaris l.)', *Agronomy*. doi: 10.3390/agronomy11061154.

ISO 22005:2007 (2007) 'International Standard International Standard', p. 13.

ISO 527 - 3 (2015) 'INTERNATIONAL STANDARD iTeh STANDARD iTeh STANDARD PREVIEW', *International Organization for Standardization*, 10406–1:20, pp. 3–6.

Jacobsen, N. E. (2007) *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. doi: 10.1002/9780470173350.

Jarpa-Parra, M. (no date) 'Lentil protein a review of functional properties and food application An overview of lentil protein functionality', *International Journal of Food Science and Technology*.

Karunathilaka, S. R. *et al.* (2018) 'Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: A study using eleven potential adulterants', *Heliyon*. doi: 10.1016/j.heliyon.2018.e00806.

Kaufmann, A. (2014) 'Combining UHPLC and high-resolution MS: A viable approach for the analysis of complex samples?', *TrAC - Trends in Analytical Chemistry*. doi: 10.1016/j.trac.2014.06.025.

Keeler, J. (2010) 'Understanding NMR Spectroscopy, 2nd Edition | Wiley', *Wiley*.

Kher, S. V. *et al.* (2010) 'Experts' perspectives on the implementation of traceability in Europe', *British Food Journal*. doi: 10.1108/00070701011029138.

Krzanowski, W. J. and Jackson, J. E. (1992) 'A User's Guide to Principal Components.', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. doi: 10.2307/2982678.

Kuballa, T. *et al.* (2018) 'Application of NMR for authentication of honey, beer and spices', *Current Opinion in Food Science*. Elsevier Ltd, 19, pp. 57–62. doi: 10.1016/j.cofs.2018.01.007.

Kumar, R. *et al.* (2009) *State of art of saffron (Crocus sativus L.) agronomy: A comprehensive review*, *Food Reviews International*. doi: 10.1080/87559120802458503.

Kumari, L., Jaiswal, P. and Tripathy, S. S. (2021) 'Various techniques useful for determination of adulterants in valuable saffron: A review', *Trends in Food Science and Technology*. Elsevier Ltd, 111(April 2020), pp. 301–321. doi: 10.1016/j.tifs.2021.02.061.

Leardi, R. (2002) 'Chemometrics: From classical to genetic algorithms', *Grasas y Aceites*. doi: 10.3989/gya.2002.v53.i1.294.

Levy, A. A. and Feldman, M. (2022) 'Evolution and origin of bread wheat', *The Plant cell*, 34(7), pp. 2549–2567. doi: 10.1093/plcell/koac130.

Li, S. *et al.* (2021) 'Recent advances in the application of metabolomics for food safety control and food quality analyses', *Critical Reviews in Food Science and Nutrition*. Taylor & Francis, 61(9), pp. 1448–1469. doi: 10.1080/10408398.2020.1761287.

Likes, R. *et al.* (2007) 'The betaine and choline content of a whole wheat flour compared to other mill streams', *Journal of Cereal Science*, 46(1), pp. 93–95. doi: 10.1016/j.jcs.2006.11.002.

LilianD. Kaale, Muhammad Siddiq, S. H. (no date) 'Lentil Lens culinaris Medik as nutrient-rich and versatile food legume A review', *Legume Science*.

Liu, H. Y. *et al.* (2021) 'Wheat authentication : An overview on different techniques and chemometric methods', *Critical Reviews in Food Science and Nutrition*. Taylor & Francis, 0(0), pp. 1–24. doi: 10.1080/10408398.2021.1942783.

Liu, S. J., Wu, Y. N. and Chan, L. (2020) 'Application of Metabonomics Approach in Food Safety Research-A Review', *Food Reviews International*. Taylor & Francis, 36(6), pp. 547–558. doi: 10.1080/87559129.2019.1655571.

Longobardi, F. *et al.* (2017) 'Geographical origin discrimination of lentils (Lens culinaris Medik.) using1H NMR fingerprinting and multivariate statistical analyses', *Food Chemistry*. Elsevier Ltd, 237, pp. 443–448. doi: 10.1016/j.foodchem.2017.05.159.

Manai–Djebali, H. *et al.* (2021) 'Chemical composition of durum wheat kernels: impact of the growing location', *Euro-Mediterranean Journal for Environmental Integration*. Springer International Publishing, 6(1), pp. 1–11. doi: 10.1007/s41207-020-00239-y.

Marieschi, M., Torelli, A. and Bruni, R. (2012) 'Markers for the Detection of Plant Adulterants Used as Bulking', *Journal of Agricultural and Food Chemistry*, (60), pp. 10998–11004.

Markley, J. L. *et al.* (2017) 'The future of NMR-based metabolomics', *Current Opinion in Biotechnology*. doi: 10.1016/j.copbio.2016.08.001.

Martínez-Moreno, F. *et al.* (2020) 'Durum wheat in the Mediterranean Rim: historical evolution and genetic resources', *Genetic Resources and Crop Evolution*, 67(6), pp. 1415–1436. doi: 10.1007/s10722-020-00913-8.

Matsuoka, Y. (2011) 'Evolution of polyploid triticum wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification', *Plant and Cell Physiology*, 52(5), pp. 750–764. doi: 10.1093/pcp/pcr018.

McGrath, T. F. *et al.* (2018) 'What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – Spectroscopy case study', *Trends in Food Science and Technology*. Elsevier Ltd, 76, pp. 38–55. doi: 10.1016/j.tifs.2018.04.001.

Mefleh, M. *et al.* (2019) 'From ancient to old and modern durum wheat varieties: interaction among cultivar traits, management, and technological quality', *Journal of the Science of Food and Agriculture*, 99(5), pp. 2059–2067. doi: 10.1002/jsfa.9388.

Mike Sissons (2008) 'Role of Durum Wheat Composition on the Quality of Pasta and Bread', *Food*, 2(2), pp. 75–90.

Mo'ez Al-Islam Ezzat Faris, Hamed Rabah Takruri, A. Y. I. (no date) 'Role of lentils (Lens culinaris L.) in human health and nutrition: a review - IOS Press', *Mediterranean Journal of Nutrition and Metabolism*. Available at: https://content.iospress.com/articles/mediterranean-journal-of-nutrition-and-metabolism/mnm6-1-02 (Accessed: 8 January 2021).

Musio, B. *et al.* (2022) 'Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production', *Applied Sciences (Switzerland)*. doi: 10.3390/app12052583.

Naz, S. *et al.* (2014) 'Method validation strategies involved in non-targeted metabolomics', *Journal of Chromatography A*. doi: 10.1016/j.chroma.2014.04.071.

Nichani, K. *et al.* (2023) 'Food Chemistry : X Essential terminology and considerations for validation of non-targeted methods', 17(November 2022). doi: 10.1016/j.fochx.2022.100538.

'Nutritional and health properties of pulses' (2009) *Mediterranean Journal of Nutrition and Metabolism*, 1(3), pp. 149–157. doi: 10.1007/s12349-008-0023-2.

Oliveri, P. and Downey, G. (2013) 'Discriminant and class-modelling chemometric techniques for food PDO verification', in *Comprehensive Analytical Chemistry*. doi: 10.1016/B978-0-444-59562-1.00013-X.

Ordoudi, S. A. *et al.* (2015) '1H NMR-based metabolomics of saffron reveals markers for its quality deterioration', *Food Research International*, 70, pp. 1–6. doi: 10.1016/j.foodres.2015.01.021.

Pauly, A. *et al.* (2013) 'Wheat (Triticum aestivum L. and T. turgidum L. ssp. durum) Kernel Hardness: II. Implications for End-Product Quality and Role of Puroindolines Therein', *Comprehensive Reviews in Food Science and Food Safety*, 12(4), pp. 427–438. doi: 10.1111/1541-4337.12018.

Pei, X. *et al.* (2011) 'The China melamine milk scandal and its implications for food safety regulation', *Food Policy*. doi: 10.1016/j.foodpol.2011.03.008.

Peña-Bautista, R. J. *et al.* (2017) ' CIMMYT Series on Carbohydrates, Wheat, Grains, and Health: Wheat-Based Foods: Their Global and Regional Importance in the Food Supply, Nutrition, and Health , ', *Cereal Foods World*, 62(5), pp. 231–249. doi: 10.1094/cfw-62-5-0231.

Peng, J. H., Sun, D. and Nevo, E. (2011) 'Domestication evolution, genetics and genomics in wheat', *Molecular Breeding*, 28(3), pp. 281–301. doi: 10.1007/s11032-011-9608-4.

Pequeno, D. N. L. *et al.* (2021) 'Climate impact and adaptation to heat and drought stress of regional and global wheat production', *Environmental Research Letters*, 16(5). doi: 10.1088/1748-9326/abd970.

Petrakis, E. A. *et al.* (2017) 'Sudan dyes in adulterated saffron (Crocus sativus L.): Identification and quantification by 1H NMR', *Food Chemistry*, 217, pp. 418–424. doi: 10.1016/j.foodchem.2016.08.078.

Piergiovanni, A. R. (2000) 'The evolution of lentil (Lens culinaris Medik.) cultivation in Italy and its effects on the survival of autochthonous populations', *Genetic Resources and Crop Evolution*, 47(3), pp. 305–314. doi: 10.1023/A:1008789614680.

Ragone, R. *et al.* (2020) 'Development of a food class-discrimination system by non-targeted NMR analyses using different magnetic field strengths', *Food Chemistry*. Elsevier Ltd, 332. doi: 10.1016/j.foodchem.2020.127339.

Rahman, S. *et al.* (2020) 'Current progress in understanding and recovering the wheat genes lost in evolution and domestication', *International Journal of Molecular Sciences*, 21(16), pp. 1–19. doi: 10.3390/ijms21165836.

Riedl, J., Esslinger, S. and Fauhl-Hassek, C. (2015) 'Review of validation and reporting of non-targeted fingerprinting approaches for food authentication', *Analytica Chimica Acta*. Elsevier B.V., 885, pp. 17–32. doi: 10.1016/j.aca.2015.06.003.

van Rijswijk, W. *et al.* (2008) 'Consumer perceptions of traceability: A cross-national comparison of the associated benefits', *Food Quality and Preference*. doi: 10.1016/j.foodqual.2008.02.001.

Roy, F., Boye, J. I. and Simpson, B. K. (2010) 'Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil', *Food Research International*. Elsevier Ltd, 43(2), pp. 432–442. doi: 10.1016/j.foodres.2009.09.002.

Sabatino, L. *et al.* (2011) 'HPLC/PDA/ESI-MS evaluation of saffron (Crocus sativus L.) adulteration', *Natural Product Communications*, 6(12), pp. 1873–1876. doi: 10.1177/1934578x1100601220.

Sadílek, T. (2018) 'Perception of food quality by Czech consumers - Literature review', *Proceedings of the 31st International Business Information Management Association Conference, IBIMA 2018: Innovation Management and Education Excellence through Vision 2020*, XXII(1), pp. 6183–6191. doi: 10.35808/ersj/1407.

Saia, S. *et al.* (2019) 'Metabolomics Provides Valuable Insight for the Study of Durum Wheat: A Review', *Journal of Agricultural and Food Chemistry*, 67(11), pp. 3069–3085. doi: 10.1021/acs.jafc.8b07097.

Sajdakowska, M. *et al.* (2018) 'Importance of health aspects in Polish consumer choices of dairy products', *Nutrients*. doi: 10.3390/nu10081007.

Sall (2019) 'Cultivation and Potential Expansion in', *Durum Wheat (Triticum durum Desf.): Origin, Cultivation and Potential Expansion in Sub-Saharan Africa*, 9, pp. 1–20.

Sanders, J. K. M. (1988) 'Modern NMR Spectroscopy: a guide for chemist.' doi: 10.1177/1757913910379198.

Schieber, A. (2018) 'Introduction to Food Authentication', in *Modern Techniques for Food Authentication*. doi: 10.1016/b978-0-12-814264-6.00001-3.

Sellami, M. H., Pulvento, C. and Lavini, A. (2021) 'Selection of suitable genotypes of lentil (Lens culinarismedik.) under rainfed conditions in south italy usingmulti-trait stability index (mtsi)', *Agronomy*, 11(9). doi: 10.3390/agronomy11091807.

Shewry, P. R. *et al.* (2017) 'Defining genetic and chemical diversity in wheat grain by 1H-NMR spectroscopy of polar metabolites', *Molecular Nutrition and Food Research*, 61(7), pp. 1–9. doi: 10.1002/mnfr.201600807.

Shewry, P. R. and Halford, N. G. (2002) 'Cereal seed storage proteins: Structures, properties and role in grain utilization', *Journal of Experimental Botany*, 53(370), pp. 947–958. doi: 10.1093/jexbot/53.370.947.

Shewry, P. R. and Hey, S. J. (2015) 'The contribution of wheat to human diet and health', *Food and Energy Security*, 4(3), pp. 178–202. doi: 10.1002/FES3.64.

Shiferaw, B. *et al.* (2013) 'Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security', *Food Security*, 5(3), pp. 291–317. doi: 10.1007/s12571-013-0263-y.

Siebert, K. J. (2001) 'Chemometrics in brewing - A review', *Journal of the American Society of Brewing Chemists*. doi: 10.1094/asbcj-59-0147.

Singh, B. *et al.* (2017) 'Phenolic composition and antioxidant potential of grain legume seeds: A review', *Food Research International*. Elsevier, 101(June), pp. 1–16. doi: 10.1016/j.foodres.2017.09.026.

Sobolev, A. *et al.* (2014) 'Saffron Samples of Different Origin: An NMR Study of Microwave-Assisted Extracts', *Foods*, 3(3), pp. 403–419. doi: 10.3390/foods3030403.

Sobolev, A. P. *et al.* (2015) 'Untargeted NMR-based methodology in the study of fruit metabolites', *Molecules*. doi: 10.3390/molecules20034088.

Sobolev, A. P. *et al.* (2019) 'Use of NMR applications to tackle future food fraud issues', *Trends in Food Science and Technology*. Elsevier, 91(March 2018), pp. 347–353. doi: 10.1016/j.tifs.2019.07.035.

Spink, J. and Moyer, D. C. (2011) 'Defining the Public Health Threat of Food Fraud', *Journal of Food Science*. doi: 10.1111/j.1750-3841.2011.02417.x.

Spyros, A. (2016) 'Application of NMR in food analysis', *Nuclear Magnetic Resonance*. doi: 10.1039/9781782624103-00269.

Stewart, D. *et al.* (2018) *Crops and Tasty, Nutritious Food - How Can Metabolomics Help?*, *Annual Plant Reviews online*. doi: 10.1002/9781119312994.apr0467.

Strange, R. N., Majer, J. R. and Smith, H. (1974) 'The isolation and identification of choline and betaine as the two major components in anthers and wheat germ that stimulate Fusarium graminearum in vitro', *Physiological Plant Pathology*, 4(2), pp. 277–290. doi: 10.1016/0048-4059(74)90015-0.

Tahir, M., Vandenberg, A. and Chibbar, R. N. (2011) 'Influence of environment on seed soluble carbohydrates in selected lentil cultivars', *Journal of Food Composition and Analysis*. Elsevier Inc., 24(4–5), pp. 596–602. doi: 10.1016/j.jfca.2010.04.007.

Takis, P. G. *et al.* (2019) 'Uniqueness of the NMR approach to metabolomics', *TrAC - Trends in Analytical Chemistry*. doi: 10.1016/j.trac.2018.10.036.

Trygg, J. and Wold, S. (2002) 'Orthogonal projections to latent structures (O-PLS)', *Journal of Chemometrics*. doi: 10.1002/cem.695.

Valle García-Rodríguez, M. *et al.* (2014) 'Determination of saffron quality by high-performance liquid chromatography', *Journal of Agricultural and Food Chemistry*. American Chemical Society, 62(32), pp. 8068–8074. doi: 10.1021/jf5019356.

Verdú Jover, A. J., Lloréns Montes, F. J. and Fuentes Fuentes, M. del M. (2004) 'Measuring perceptions of quality in food products: The case of red wine', *Food Quality and Preference*. doi: 10.1016/j.foodqual.2003.08.002.

Wang, Y. *et al.* (2003) 'Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues', *Analytical Biochemistry*. doi: 10.1016/j.ab.2003.07.026.

Winterhalter, P. and Straubinger, M. (2000) 'Saffron - Renewed interest in an ancient spice', *Food Reviews International*. doi: 10.1081/FRI-100100281.

Wold, S., Sjöström, M. and Eriksson, L. (2001) 'PLS-regression: A basic tool of chemometrics', in *Chemometrics and Intelligent Laboratory Systems*. doi: 10.1016/S0169-7439(01)00155-1.

Xi, L. and Qian, Z. (2006) 'Pharmacological properties of crocetin and crocin (digentiobiosyl ester of crocetin) from saffron', *Natural Product Communications*, 1(1), pp. 65–75. doi: 10.1177/1934578x0600100112.

Xynias, I. N. *et al.* (2020) 'Durum wheat breeding in the Mediterranean region: Current status and future prospects', *Agronomy*, 10(3), pp. 1–27. doi: 10.3390/agronomy10030432.

Yilmaz, A. *et al.* (2010) '1 H NMR metabolic fingerprinting of saffron extracts', *Metabolomics*, 6(4), pp. 511–517. doi: 10.1007/s11306-010-0221-z.

Zeisel, S. H. *et al.* (2003) 'Erratum: Concentrations of choline-containing compounds and betaine in common foods (Journal of Nutrition (2003) 133 (1302-1307))', *Journal of Nutrition*, 133(9), pp. 2918–2919.

ACKNOWLEDGEMENTS

5.0 Acknowledgements

My sincere thanks go to my tutors, Prof. Piero Mastrorilli, Prof. Mario Latronico and Dr. Biagia Musio.
I would express my gratitude to Prof. Vito Gallo.
I would like to thank Rosa Ragone, Stefano Todisco, Maurizio Triggiani, Antonino Rizzuti, Alessandra Ciampa, Stefania Pontrelli, Angelica Sicilia e Giuseppe D'Orazio.
Thanks also to the support of Valentina, Serena, Ambra, Elhussein, Maria and Annamaria.

CURRICULUM

6.0 Curriculum

MARICA ANTONICELLI

Nationality
Italian
Date of birth
04/06/1991
E-mail
marica.antonicelli@poliba.it

EDUCATION

Politecnico di Bari, PhD in Risk and environmental, territorial and building development
01/11/2019 – 31/01/2023
The PhD research activity aimed to optimise and validate non-targeted methods based on Nuclear Magnetic Resonance spectroscopy in quality control, traceability, and authenticity assessment of agri-food products.
Università degli studi di Bari "A. Moro", Master's Degree in Medical Biotechnology and Molecular Medicine
21/03/2019-110 cum laude
Università degli studi di Bari "A. Moro", Bachelor's Degree in Medical and Pharmaceutical Biotechnology
22/07/2015-110 cum laude

SCIENTIFIC PUBLICATIONS

- Musio, B.; Ahmed E.; Antonicelli, M.; Chiapperini, D.; Dursi, O.; Grieco, F.; Latronico, M.; Mastrorilli, P.; Ragone, R.; Settanni, R.; Triggiani, M.; Gallo, V. A spectroscopic study to assess heavy metals absorption by a combined hemp/spirulina system from contaminated soil. Environmental Advances 2022, 7.
https://doi.org/10.1016/j.envadv.2021.100144.

- Musio, B.; Todisco S.; Antonicelli M.; Garino C.; Arlorio M.; Mastrorilli P.; Latronico M.; Gallo V. Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production. Applied Sciences 2022, 12(5), 2583. https://doi.org/10.3390/app12052583