Elsevier Editorial System(tm) for Talanta
Manuscript Draft

Manuscript Number: TAL-D-16-01245R1

Title: A NIR spectroscopy-based efficient approach to detect fraudulent additions within mixtures of dried porcini mushrooms

Article Type: Research Paper

Keywords: NIR spectroscopy; class-modelling; partial least squares density modelling (PLS-DM); dried porcini mushrooms; Boletus edulis; data fusion

Corresponding Author: Dr. Paolo Oliveri, PhD

Corresponding Author's Institution: University of Genova

First Author: Monica Casale, PhD

Order of Authors: Monica Casale, PhD; Lucia Bagnasco, PhD; Mirca Zotti, PhD; Simone Di Piazza, PhD; Nicola Sitta, PhD; Paolo Oliveri, PhD

Abstract: Boletus edulis and Allied Species (BEAS), known as "porcini mushrooms", represent almost the totality of wild mushrooms placed on the Italian market, both fresh and dehydrated. Furthermore, considerable amounts of these dried fungi are imported from China. The presence of Tylopilus spp. and other extraneous species (i.e. species edible but not belonging to BEAS) within dried "porcini" mushrooms - mainly from those imported from China and sold in Italy - may represent an evaluable problem from a commercial point of view.
The purpose of the present study is to evaluate near-infrared spectroscopy (NIRS) as a rapid and effective alternative to classical methods for identifying extraneous species within dried porcini batches and detecting related commercial frauds.
To this goal, 80 dried fungi including BEAS, Tylopilus spp., and Boletus violaceofuscus were analysed by NIRS.
For each sample, 3 different parts of the pileus (pileipellis, flesh and hymenium) were analysed and a low-level strategy for data fusion, consisting of combining the signals obtained by the different parts before data processing, was applied.
Then, NIR spectra were used to develop reliable and efficient class-models using a novel method, partial least squares density modelling (PLS-DM), and the two most commonly used class-modelling techniques, UNEQ and SIMCA.
The results showed that NIR spectroscopy coupled with chemometric class-modelling technique can be suggested as an effective analytical strategy to check the authenticity of dried BEAS mushrooms.

Novelty Statement

A novel class-modelling method, called partial least squares density modelling (PLS-DM) has been applied in this study to build authentication models, and the outcomes have been critically compared with those of classical class-modelling techniques used in chemometrics (SIMCA and UNEQ).

Furthermore, the present work performs, for the first time, a particular data fusion strategy. In fact, data fusion is commonly applied to merge analytical data obtained by different analytical instruments. Differently, in this study, we propose merging – by suitable chemometric strategies – NIR spectra acquired in the different biological tissues (pileipellis, flesh and hymenium) that characterise mushrooms pileus, in order to thoroughly exploit spectroscopic information to be used in chemometric authentication models.

***Highlights (for review)**

- NIR spectroscopy was applied to detect extraneous species within dried *porcini* mushrooms
- A novel class-modelling method (PLS-DM) allowed to build efficient authentication models
- A peculiar data fusion strategy was performed to enhance model performances
- Outcomes were critically compared with classical approaches

A NIR spectroscopy-based efficient approach to detect fraudulent additions within mixtures of dried *porcini* mushrooms

Monica Casale[a], Lucia Bagnasco[a], Mirca Zotti[b], Simone Di Piazza[b], Nicola Sitta[c], Paolo Oliveri[a,*]

a. *Department of Pharmacy, University of Genoa, Viale Cembrano, 4, I-16148 Genoa, Italy.*
b. *Department of Earth, Environment and Life Sciences – Laboratory of Mycology, Corso Europa, 26, I-16132 Genoa, Italy.*
c. *Professional Consulting Mycologist, Loc. Farné, 32, I-40042 Lizzano in Belvedere, Italy.*

Abstract

Boletus edulis and Allied Species (BEAS), known as "*porcini* mushrooms", represent almost the totality of wild mushrooms placed on the Italian market, both fresh and dehydrated. Furthermore, considerable amounts of these dried fungi are imported from China. The presence of *Tylopilus* spp. and other extraneous species (*i.e.*, species edible but not belonging to BEAS) within dried *porcini* mushrooms – mainly from those imported from China and sold in Italy – may represent an evaluable problem from a commercial point of view.

The purpose of the present study is to evaluate near-infrared spectroscopy (NIRS) as a rapid and effective alternative to classical methods for identifying extraneous species within dried *porcini* batches and detecting related commercial frauds.

To this goal, 80 dried fungi including BEAS, *Tylopilus* spp., and *Boletus violaceofuscus* were analysed by NIRS.

For each sample, 3 different parts of the pileus (pileipellis, flesh and hymenium) were analysed and a low-level strategy for data fusion, consisting of combining the signals obtained by the different parts before data processing, was applied.

Then, NIR spectra were used to develop reliable and efficient class-models using a novel method, partial least squares density modelling (PLS-DM), and the two most commonly used class-modelling techniques, UNEQ and SIMCA.

The results showed that NIR spectroscopy coupled with chemometric class-modelling technique can be suggested as an effective analytical strategy to check the authenticity of dried BEAS mushrooms.

Keywords: NIR spectroscopy, class-modelling, partial least squares density modelling (PLS-DM), dried *porcini* mushrooms, *Boletus edulis*, data fusion.

Introduction

Out of all the wild mushrooms in the world, there are few as prised and sought after as *Boletus edulis* Bull. and allied/related species (BEAS) [1]. BEAS, commonly named *porcini*, stands for a set of fungal species including, in Europe, *Boletus edulis* Bull., *B. aereus* Bull., *B. aestivalis* (Paulet) Fr., and *B. pinophilus* Pilát & Dermek, and dozens of other species worldwide [2–5], among which several Chinese species recently described elsewhere [6]. BEAS belong to Section *Boletus* (ex Sect. *Edules* Fr.), which is today considered representative of all the species of genus *Boletus* L. s. str. [7] (*Boletaceae, Boletales, Basidiomycota*). The purplish-hued Asian species *Boletus violaceofuscus* W.F. Chiu can be considered controversial for its placement inside or outside the BEAS. During last decade *Boletus violaceofuscus* was considered, on a morphological basis, as belonging to the section *Boletus* [8]; then a molecular study by Mello et al. [9] showed that it clustered outside the section *Boletus*. Now it has been ascribed again within section *Boletus*, even if in a distinct lineage called *Alloboletus* by recent molecular revisions [2,6,10].

BEAS are among the edible mushrooms the most widely collected in the world [11]. Dentinger et al. [2] affirm that their economic value is clearly substantial since 20.000-100.000 metric tons are estimated to be consumed annually and the median wholesale price in the U.S. for fresh mushroom in 2009 was ca. US $60/kg and can reach US $200/kg. It is worth noting that also in Europe mushroom market is an important source of revenue for a number of rural regions areas [11]. In the Mediterranean area, BEAS are an essential component of the traditional culture and cuisine, especially in Italy [1].

A significant portion of *porcini* are dried, packaged and then distributed worldwide. However, most of the *porcini* available on the Italian market or exported by Italy are imported from Eastern Europe and China where they are collected, dried on site, and then subjected to a first selection.

Unfortunately, among imported fungi ascribed to BEAS there is the presence of different, less valuable fungi, some of them not edible or not marketable according to some national laws.

Analyses performed to define macrofungi eligible for sale are mainly based on naked eye inspection aimed at identifying extraneous species (species, edible or not, not belonging to BEAS) and/or macromorphologic alterations.

Several species of the genus *Tylopilus* can be intermixed with BEAS, especially in the batches of dried *porcini* imported from China [12]. Several of these Asian species belong to the *Tylopilus plumbeoviolaceus* complex and are notoriously difficult to individuate due to their morpho-chromatic and organoleptic affinities [12,13]. These species can be easily confused with BEAS by a trivial visual inspection, even if a simple taste test highlights an intense bitter flavour.

Identifying *Tylopilus* spp. intermixed with BEAS is very difficult, not only for workers employed in mushroom manual selection and packaging, but also for mycologists.

BEAS derived products may be adulterated even with *Boletus violaceofuscus*. From a commercial point of view, this species independently of molecular analysis, is quite distinct from European species of *porcini* for its purple basidiomata, and it is not considered belonging to BEAS and, consequently, it represents an "extraneous species" in the dried *porcini* batches. Anyway, the presence of small amounts of such a species among dried specimens could be not so straightforward.

Visual inspection of basidiomata performed by professional mycologists is the most adopted methods for dried mushroom identification: up to now, no instrumental analytical techniques have been proposed to identify dried BEAS.

The present study describes an original, rapid, efficient and non-destructive analytical method, based on near infrared spectroscopy (NIRS) coupled with chemometrics, to detect additions of lower-quality and/or non-European mushroom species intermixed with BEAS.

To this goal, 80 dried fungi (44 BEAS, 20 *Tylopilus* spp. and 16 *Boletus violaceofuscus*) were analysed by NIRS.

The study was focused on the pileus (the technical name for the cap of a sporoma or fungal fruiting body), which is one of the most characterising anatomical portion. In more detail, for each fungus, three parts of the pileus were considered and analysed: pileipellis (cortical layer of pileus), flesh (layer under the pileipellis and above the hymenium), and hymenium (spore-bearing layer of the sporoma).

According to the classification proposed by Durrant-Whyte [14], a complementary fusion was performed: "*the information provided by the input sources represents different parts of the scene and could thus be used to obtain more complete global information…*". In particular, in this study, a low-level fusion approach, consisting in combining the whole signals provided by the different fungus parts before data processing was tested.

Then, a class-modelling approach was followed, aimed at characterising the BEAS samples. In more detail, unequal dispersed classes (UNEQ) [15,16] and soft independent modelling of class analogy (SIMCA) [17,18], the class-modelling techniques most commonly applied in chemometrics. Model performances, evaluated in terms of efficiency (geometric mean of sensitivity and specificity), indicated that data structure was quite complex and would have gained a benefit from the application of a method more capable to model non-normal distributions.

For this reason, a novel PLS-based class-modelling strategy was applied, called partial least squares density modelling (PLS-DM), which combines the features of PLS and potential function methods

(PFM), together with Q statistics, to obtain highly efficient class models for the characterisation of BEAS fungi. This method was presented the first time in 2014 by Oliveri *et al*. [19]. The efficiency of PLS-DM models, fully validated by means of an external test set, showed that NIR spectroscopy can be used as a valid tool for the verification of "authenticity" of BEAS.

2. Experimental

2.1. Samples

80 samples of dried mushrooms were analysed. They were samples of dried BEAS, *Tylopilus* spp. and *Boletus violaceofuscus*. Specimens were provided by different suppliers from different origins. In more detail, BEAS originated from Europe (Poland, Hungary, Bulgaria, Romany, Serbia, Macedonia, and other Balkans), Russia, and China (Yunnan province), whereas *Tylopilus* spp. and *Boletus violaceofuscus* were from China. All of the specimens were collected in three different years. Since the goal of this study is the characterisation of BEAS fungi, the 80 samples were divided into two classes:

1) 44 BEAS (the target class to be modelled);
2) 36 Non-BEAS comprehending 20 *Tylopilus* spp. and 16 *Boletus violaceofuscus* (used to evaluate specificity of the *Boletus* class models).

2.2. Apparatus and Procedure

NIR Spectroscopy.

NIR measurements were performed by a FT Near-Infrared spectrometer, based on a polarisation interferometer (Buchi NIRFlex N-500), in the 4000–10,000 cm^{-1} range with a 4 cm^{-1} resolution and a total of 512 scans were averaged for every spectrum. The diameter of the circular surface analysed was reduced to 3.0 mm by using a specific adaptor.

2.3. Data analysis.

One spectra for each part (pileipellis, flesh, and hymenium) of the sample was recorded; then a segment of the signals, from 9000 to 10,000 cm^{-1}, was removed because not informative.

Thus 3 spectra were available for each fungus and 3 NIR data matrices having 80 rows (samples) and 1250 columns (variables, reflectance at different wave numbers) were built. For simplicity these data matrices will be referred to as: $\mathbf{P1}_{80,1250}$, $\mathbf{P2}_{80,1250,\text{ and }}$ $\mathbf{P3}_{80,1250,}$ respectively for pileipellis, flesh and hymenium. **P1**, **P2**, and **P3** were submitted separately to standard normal variate (SNV) transform and second derivative; derivative spectra were calculated with a Savitzky–Golay filter using a third-order polynomial and an 11-point window.

In order to extract useful complementary information from the different parts of mushrooms analysed, a strategy for data fusion was applied and the three pre-treated NIR data matrices were combined, forming new data matrices. In particular, two unified matrices were tested:

1) **U1-2-3**$_{80,\ 3750}$ obtained combining all three data matrices **P1**, **P2** and **P3**.
2) **U1-3**$_{80,\ 2500}$ obtained combining only the two mushroom parts found to be most informative, *i.e.*, pileipellis and hymenium.

As the first step, principal component analysis (PCA) [20] was applied on the three pre-treated separate matrices (**P1**, **P2**, and **P3**) and also on the unified matrices, as a display method in order to visualise data structure.

Then, a class-modelling approach was applied in order to build a model useful to verify authenticity of BEAS dried mushrooms on the basis of NIR spectra. Model performances were evaluated in terms of efficiency, computed as the geometric mean of sensitivity and specificity. Sensitivity is defined as the percentage of samples belonging to the modelled class which is correctly accepted by the class model. It can also be defined as the rate of true positives. Specificity is the percentage of samples not belonging to the modelled class which is correctly rejected by the model. It can also be defined as the rate of true negatives. [21].

As a first attempt, two class-modelling techniques widely used in chemometrics, UNEQ and SIMCA, were applied.

UNEQ [15]$^{-}$ [16] is the name currently used in chemometrics for the method originally developed by H. Hotelling for multivariate quality control based on the Hotelling T^2 statistics, and revived by Derde and Massart as a class-modelling method. UNEQ can also be considered as the class modelling version of quadratic discriminant analysis (QDA).

SIMCA is a powerful distance-based method, the first class-modelling technique to be introduced into chemometrics by Wold in 1977 [17]. SIMCA build a class model based on the principal components of the category, that are generally computed after separated category autoscaling or centering.

Principal components of autoscaled data are computed and K components (ideally, the significant ones) are used to build the model. Such K components define the inner space, the space of the structure. The subsequent components define the outer space, the space of noise. SIMCA model is a (hyper) parallelepiped in the space of the first components, delimited by the score range.

Then, PLS-DM – a novel class modelling method proposed by Oliveri et al. in 2014 [19] –was performed on the same data matrices **P** and **U**.

A PLS model is developed using the NIR spectral data as **X** predictor matrix and a density vector as the **y** response vector. The response value (y_i) – for each sample i of the training set of the class to

be modelled – is computed as an estimation of sample density (d_i), based on inter-sample distances in the multivariate space. In more detail, all the Euclidean distances from sample *i* to each of the other training samples are computed. Such distances are, therefore, ordered, and the density value (d_i), obtained as the sum of the *k* smallest (*i.e.*, lowest-order) distances, is studied varying *k*. Parameter *k* influences the smoothness of density function, which evolves from a sharper to a smoother shape while increasing *k*.

After PLS modelling, the PLS scores on the first *L* latent variables selected are used as an input to estimate the PFM probability density of the class, with different smoothing coefficients (*a*). Then, the critical value, fα, of the probability density distribution is computed, at a preselected confidence level ($1 - \alpha$). In addition, the PLS residuals are used to compute the critical value of *Q* statistics, Q_α, at the same confidence level. In this way, compliance of each object with the class model is granted when it satisfies both f_α and Q_α criteria.

The algorithm calculates models with all of the different parameter combinations – *i.e.*, distance (*k*), smoothing coefficient (*a*), and the number of latent variables (*L*), as well as the suitable **X**-block pre-processing. At the end, outcomes of all of the combinations are evaluated – in terms of sensitivity and specificity – by means of Pareto diagrams, useful to choose the optimal conditions.

For all of the methods, feature optimisation was performed by a 5-fold Venetian blind cross-validation scheme. Finally, prediction ability was computed on an external test set including a randomly selected 20% of total samples.

3. Results and Discussion

3.1. Display analysis

First of all, PCA was applied as a display method on the pretreated (SNV and second derivative) data matrices of separate mushroom parts (series **P**).

For the sake of conciseness, only the most explanatory results, obtained on **P3** are shown: in Figure 1, score plot on the two lowest-order components, explaining about 75% of total variance, a rather clear distinction between BEAS fungi (B) and the other samples can be observed; in particular *Tylopilus* spp. samples (T) fall further away from BEAS. Instead, a larger overlap with *Boletus violaceofuscus* (V) samples can be observed.

Score plots obtained from PCA performed on **P1** and **P2** (*not shown*) present a greater overlap between BEAS and other fungi, indicating that pileipellis and, even more, flesh are less important than hymenium for the characterisation of BEAS; such a conclusion completely agrees with what reported by mycologists.

PCA was also performed on the unified data matrices **U**. The related score plots (*not reported*) did not exhibit any better sample characterisation.

3.2. Class-modelling analysis

Models obtained with the three class-modelling methods (UNEQ, SIMCA, and PLS-DM) were optimised by the same cross-validation scheme and the final models were validated on the same external test sets, to guarantee a meaningful comparison of results.

The outcomes of all the optimal models developed for BEAS using NIR spectra of the three parts separately and the unified matrices are summarised in Table 1.

The results obtained with the unified matrix **U1-2-3** were, in all cases and with all class-modelling methods, less satisfactory than those obtained on the data matrices **P** separately; this is probably due to the inclusion of part 2 of the fungi, the flesh that also the mycologists consider as the less informative part for the characterisation of BEAS.

UNEQ is the method that provided the less satisfactory results. This can be motivated considering that, when samples are described by complex distributions that deviate from normality, which is the case for spectral data in this study, UNEQ – being based on multivariate normal distributions – is not appropriate[15].

SIMCA provided more satisfactory results: considering the three parts of fungi separately, efficiency on the test set ranging from 72.8% for **P2** (flesh) to 79.8% for **P1** (pileipellis). The best results were achieved with the unified data **U1-3**: a well-balanced model, with 81.8% sensitivity and 80.6% specificity on the external test set, was obtained.

As far as PLS-DM is concerned, according to the methodology described above, several parameters were settled in order to define the optimal model – namely, pre-processing, *k*, *L*, and *a*. In more detail, four possibilities were considered for variable pre-processing: no pre-processing, mean centring, scaling and autoscaling. As for the *k* parameter, integer values from 1 to 7 were considered. The number *L* of latent variables was varied from 1 to 20. Finally, the smoothing coefficient *a* was varied from 0.3 to 0.8, with 0.1 increments. These parameters were varied within the specified ranges and the results of all the combinations were tested by a cross-validation scheme to select the optimal model.

The optimal number of LVs was selected, for each condition, by considering the maximum efficiency of the resulting class model, evaluated by cross-validation. In a second step, models at fixed LVs were evaluated by examining a Pareto diagram, whose axes correspond to sensitivity and specificity of class models.

Pareto optimal solutions, which define the Pareto front, are represented by points connected by the black line in Figures 2.a-d. The final model was selected among the solutions laying on the front, looking for a balance between sensitivity and specificity.

As it can be noticed looking at the results in Table 1, PLS-DM was able to provide more efficient and more balanced models, in terms of cross-validation sensitivity and specificity, compared with those obtained by both UNEQ and SIMCA. Also considering validation on the test sets, overall higher efficiency value was obtained by PLS-DM for all of the three parts of fungi; in some cases, these efficiencies were significantly better than those of SIMCA (especially **P3**), in other cases (**P2** and **U1-3**), they were comparable to those of SIMCA.

Summarising all the observations and considerations:

1) All the results reported are in agreements with mycological considerations; in fact, the most efficient models were always obtained on matrix **P3**, that can be considered as the most informative part of the fungus from an histological and biochemical point of view. On the contrary, the less satisfactory class-modelling performances were obtained with NIR spectra of flesh that are indeed considerable as the less characterising part of fungi.

2) UNEQ is the method less suitable for this type of data, and provides the less satisfactory results; this can be justified by the fact that sample distributions in this study are very complex, with deviation from normality and, thus, do not fulfil UNEQ requirements. Moreover, UNEQ was the only method that provided optimal models on autoscaled spectral data. A reason for this could be identified in the fact that – differently from the other two methods – UNEQ does not take into account residuals and, therefore, may benefits from column autoscaling to enhance minor spectral features. Such minor – but useful – information is accounted for by residuals in the other two methods.

3) SIMCA provides quite satisfactory results, with efficiency on the external test set ranging from 72.8%, on matrix **P2** (flesh), to 81.2%, on the unified matrix **U1-3**. In the latter case, fusion of information provided by two different fungus parts (pileipellis and hymenium) allows obtaining more complete global information on the sample useful for its characterisation.

4) PLS-DM proves to be the most suitable class-modelling technique for the characterisation of BEAS based on NIR spectra. In fact, PLS-DM models are more balanced in terms of sensitivity and specificity and, except for part 2 (flesh, **P2**) that even by mycologists is described as the less characterising part for fungi, PLS-DM reaches efficiency higher than 80% on the external test set. Such models can be

considered absolutely suitable for screening purposes, considering the nature and complexity of the problem under study.

4. Conclusions

The present study showed that NIR spectroscopy, coupled with chemometric class-modelling technique, can be used as an effective analytical strategy for the verification of authenticity of dried *porcini* mushrooms (BEAS).

Deeper insights into results reveal that the different parts of fungi considered have a different potential for chemical characterisation of specimens. In particular, pileipellis and hymenium exhibit a greater potential than flesh. Such an observation, which is coherent with the biochemical characteristics of the respective parts, may be exploited as a useful indication for addressing more targeted control analyses.

From a chemometric point of view, it can be confirmed that class-modelling methods based on the normal distribution (such as UNEQ) are not suitable for describing complex class distributions, which may arise from complex spectroscopic data. Conversely, SIMCA and the recently PLS-DM method are able to efficiently manage such data and to supply satisfactory models. In more detail, PLS-DM proved to be able to provide models more efficient than SIMCA ones especially in the case of the most informative part (namely, those from hymenium). In the case of less informative data (pileipellis, and flesh) SIMCA and PLS-DM outcomes are comparable.

Fusion of data from the different parts slightly improves efficiency of the BEAS authentication models, probably not so significantly to suggest a joint analysis as a standard protocol. Mushrooms analysed were collected over a period of three years and came from non-completely known habitats: for this reason, a further model validation is convenient in order to take into account possible variability due the habitat of mushrooming, specifically type of wood and soil composition.

Efficiencies of models obtained (above 80%), together with favourable characteristic of NIR reflectance spectroscopy, such low operation costs, rapidity, null invasiveness towards samples and no requirement for highly skilled personnel, makes the analytical strategy proposed within the present study very suitable for screening purposes, on control systems. Given the versatility of the spectroscopic technique, even direct on-line implementations can be efficiently foreseen. Indeed, the present study demonstrates the way forward for the development of analytical strategies based on hyperspectral imaging in the NIR region, able

to fully control whole sample batches. Such strategies have the potential of providing real-time answers, identifying the eventual presence of different mushroom species in mixtures.

Acknowledgments

Financial support by the Italian Ministry of Education, Universities and Research (MIUR) is acknowledged – Research Project SIR 2014 "Advanced strategies in near infrared spectroscopy and multivariate data analysis for food safety and authentication", RBSI14CJHJ (CUP: D32I15000150008).

References

[1] N. Sitta, M. Floriani, Nationalization and Globalization Trends in the Wild Mushroom Commerce of Italy with Emphasis on Porcini (Boletus edulis and Allied Species), Econ. Bot. 62 (2008) 307–322. doi:10.1007/s12231-008-9037-4.

[2] B.T.M. Dentinger, J.F. Ammirati, E.E. Both, D.E. Desjardin, R.E. Halling, T.W. Henkel, et al., Molecular phylogenetics of porcini mushrooms (Boletus section Boletus), Mol. Phylogenet. Evol. 57 (2010) 1276–1292. doi:10.1016/j.ympev.2010.10.004.

[3] D. Arora, J. Frank, *Boletus rubriceps*, a new species of porcini from the southwestern USA, North Am. Fungi. 9 (2014) 1. doi:10.2509/naf2014.009.006.

[4] R.E. Halling, D.E. Desjardin, N. Fechner, D. Arora, K. Soytong, B.T.M. Dentinger, New Porcini (Boletus sect. Boletus) from Australia and Thailand, Mycologia. 106 (2014) 830–834. doi:10.3852/13-340.

[5] N.-K. Zeng, Z.-Q. Liang, Z.L. Yang, Boletus orientialbus, a new species with white basidioma from subtropical China, Mycoscience. 55 (2014) 159–163. doi:10.1016/j.myc.2013.07.004.

[6] Y.-Y. Cui, B. Feng, G. Wu, J. Xu, Z.L. Yang, Porcini mushrooms (Boletus sect. Boletus) from China, Fungal Divers. (2015). doi:10.1007/s13225-015-0336-7.

[7] G. Wu, B. Feng, J. Xu, X.-T. Zhu, Y.-C. Li, N.-K. Zeng, et al., Molecular phylogenetic analyses redefine seven major clades and reveal 22 new generic clades in the fungal family Boletaceae, Fungal Divers. 69 (2014) 93–115. doi:10.1007/s13225-014-0283-8.

[8] M. Floriani, G. Simonini, N. Sitta, Boletus violaceofuscus, una specie orientale commercializzata in Italia, Boll. Gr. Micol. G. Bres. 43 (2000) 4–15.

[9] A. Mello, S. Ghignone, A. Vizzini, C. Sechi, P. Ruiu, P. Bonfante, ITS primers for the identification of marketable boletes, J. Biotechnol. 121 (2006) 318–329. doi:10.1016/j.jbiotec.2005.08.022.

[10] B. Feng, J. Xu, G. Wu, N.-K. Zeng, Y.-C. Li, B. Tolgor, et al., DNA Sequence Analyses Reveal Abundant Diversity, Endemism and Evidence for Asian Origin of the Porcini Mushrooms, PLoS One. 7 (2012) e37567. doi:10.1371/journal.pone.0037567.

[11] E. Boa, Wild edible fungi: a global overview of their use and importance to people, Food and A, Rome, 2004.

[12] N. Sitta, Presence of Tylopilus into dried "Porcini" mushrooms from China, Micol. Ital. 29 (2000) 96–99.

[13] M. Gelardi, A. Vizzini, E. Ercole, E. Horak, Z. Ming, T. Li, Circumscription and Taxonomic Arrangement of Nigroboletus roseonigrescens Gen. Et Sp. Nov., a New Member of Boletaceae from Tropical South–Eastern China, PLoS One. 10 (2015) e0134295. doi:10.1371/journal.pone.0134295.

[14] H.F. Durrant-Whyte, Sensor Models and Multisensor Integration, Int. J. Rob. Res. 7 (1988) 97–113. doi:10.1177/027836498800700608.

[15] H.H. Hotelling, Multivariate Quality Control, in: M.H. and W.W. C Eisenhart (Ed.), Tech. Stat. Anal., New York: McGraw-Hill, 1947: p. 111.

[16] M.P. Derde, D.L. Massart, UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution, Anal. Chim. Acta. 184 (1986) 33–51. doi:10.1016/S0003-2670(00)86468-5.

[17] M.S. S. Wold, A method for analysing chemical data in terms of similarity and analogy, in: B.R Kowalski (Ed.), Chemom. Theory Appl., ACS Symposium Series 52; American Chemical Society, Washington, 1977: p. 243.

[18] S.L. M. Forina, Data Analysis in Food Chemistry, in: B.R. Kowalski (Ed.), Chemom. Math. Stat. Chem., NATO ASI S, Reidel Publ.Co, Dordrecht, 1984: p. 439.

[19] P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini, et al., Partial least squares density modeling (PLS-DM) - a new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy., Anal. Chim. Acta. 851 (2014) 30–6. doi:10.1016/j.aca.2014.09.013.

[20] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.L. Lewi, J. Smeyers-Verbeke, No Title, in: B.G.M. Vandeginste, S.C. Rutan (Eds.), Handb. Chemom. Qaulimetrics Part B, Elsevier, 1998: pp. 88–103.

[21] P. Oliveri, V. Di Egidio, T. Woodcock, G. Downey, Application of class-modelling techniques to near infrared data for food authentication purposes, Food Chem. 125 (2011) 1450–1456.

Table 01

Table 1. Optimal class models obtained by UNEQ, SIMCA and PLS-DM, respectively, on the different data matrices considered. Cross-validation (CV) and test set results.

		$P1_{80,1250}$			$P2_{80,1250}$			$P3_{80,1250}$			$U1\text{-}3_{80,2500}$		
	METHOD	**sensitivity**	**specificity**	**efficiency**	**sensitivity**	**specificity**	**efficiency**	**sensitivity**	**specificity**	**efficiency**	**sensitivity**	**specificity**	**efficiency**
CV	**UNEQ**	100.00	0.00	0.00	96.97	10.56	31.99	96.97	23.33	47.57	100.00	0.00	0.00
		autoscaling, 4 PC			*centering, 4 PC*			*centering, 3 PC*			*autoscaling, 3 PC*		
	SIMCA	69.70	77.78	73.63	54.55	83.89	67.64	60.61	96.67	76.54	66.67	81.11	73.54
		centering, 2 PC			*centering, 2 PC*			*centering, 2 PC*			*centering, 2 PC*		
	PLS-DM	75.76	77.78	76.76	81.82	67.78	74.47	78.79	87.22	82.90	84.85	78.89	81.81
		no scaling, k=1, a=0.5, 7 LV			*centering, k=1, a=0.5, 4 LV*			*centering, k=2, a=0.4, 4 LV*			*scaling, k=3, a=0.4, 3 LV*		
TEST SET	**UNEQ**	100.00	0.00	0.00	100.00	33.33	57.73	100.00	33.33	57.73	100.00	0.00	0.00
		autoscaling, 4 PC			*centering, 4 PC*			*centering, 3 PC*			*autoscaling, 3 PC*		
	SIMCA	81.82	77.78	79.77	63.64	83.33	72.82	63.64	94.44	77.52	81.82	80.56	81.18
		centering, 2 PC			*centering, 2 PC*			*centering, 2 PC*			*centering, 2 PC*		
	PLS-DM	90.91	72.22	81.03	81.82	66.67	73.85	81.82	83.33	82.57	81.82	80.56	81.18
		no scaling, k=1, a=0.5, 7 LV			*centering, k=1, a=0.5, 4 LV*			*centering, k=2, a=0.4, 4 LV*			*scaling, k=3, a=0.4, 3 LV*		

Figure Caption

Figure 1. PCA score plot of FT-NIR spectra of hymenium. Data pre-treated by SNV and second derivative.
Figure 2. Pareto diagrams – PLS-DM. Each point represents a model obtained varying PLS-DM parameters; optimal solutions are connected by the dashed lines (Pareto front). a) **P1** matrix; b) **P2** matrix; c) **P3** matrix; d) **U1-3**.

Figure 01

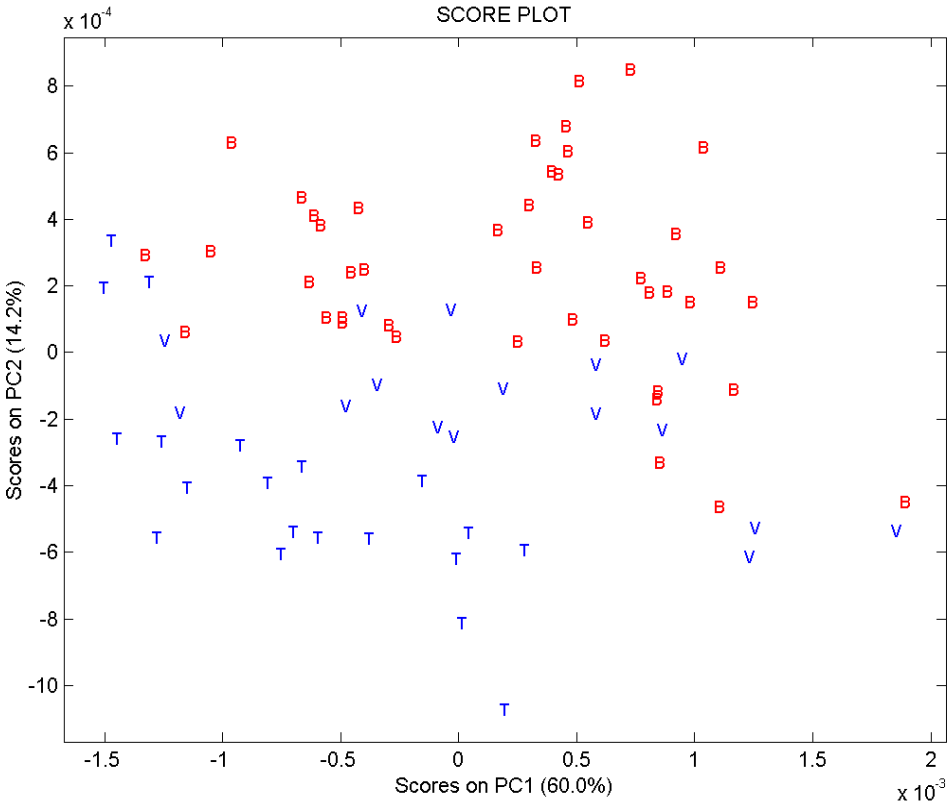
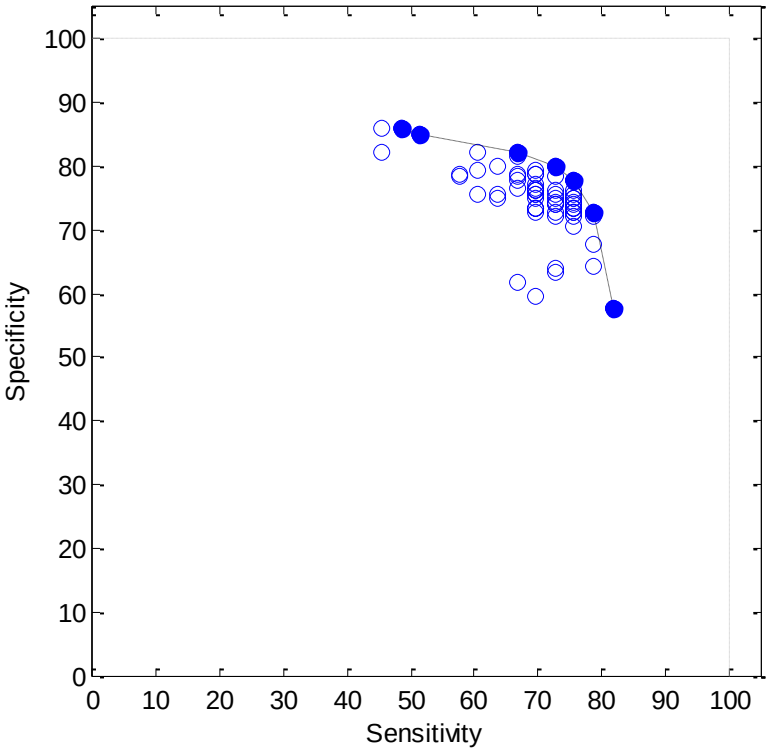
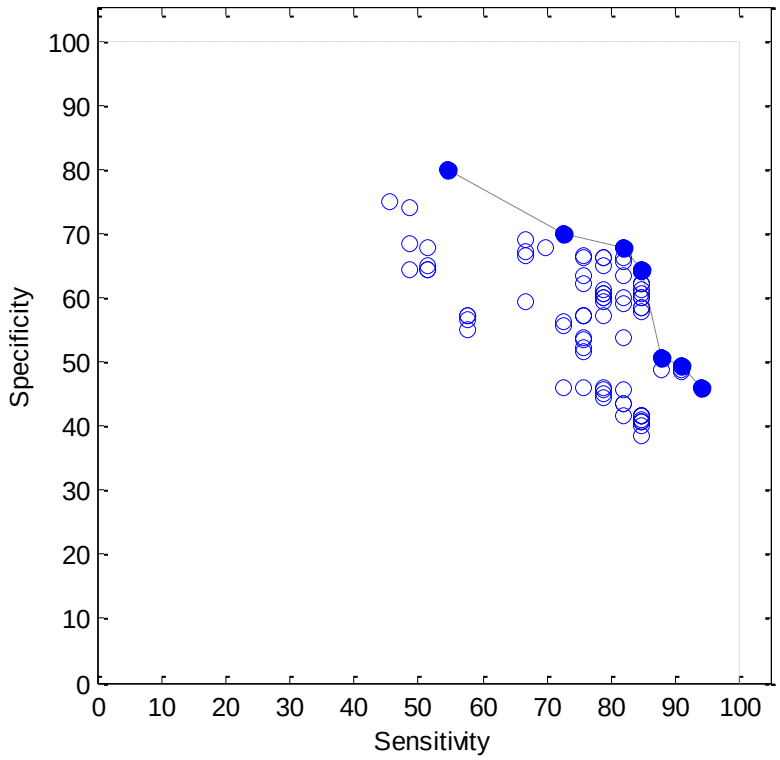

SCORE PLOT
x 10^-4
8
6
4
2
0
-2
-4
-6
-8
-10
Scores on PC2 (14.2%)
-1.5
-1
-0.5
0
0.5
1
1.5
2
Scores on PC1 (60.0%)
x 10^-3

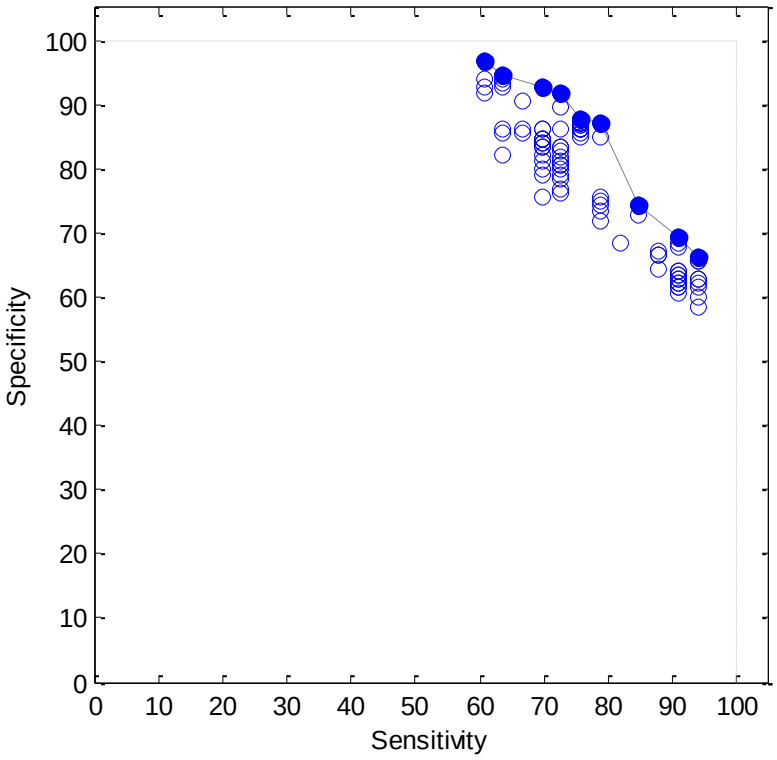
Figure 02


Specificity
Sensitivity
0
10
20
30
40
50
60
70
80
90
100

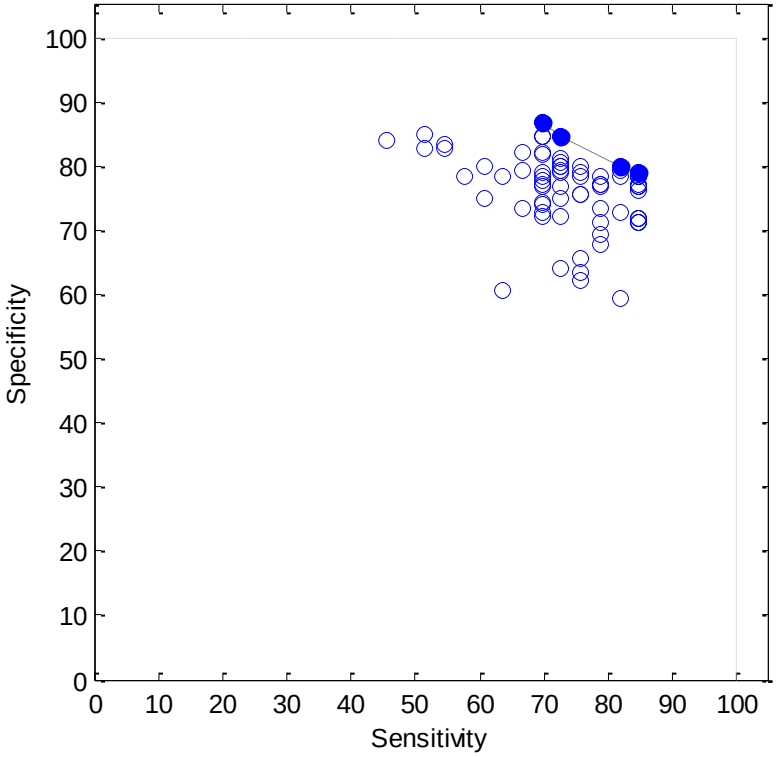
a


Specificity
Sensitivity
0
10
20
30
40
50
60
70
80
90
100

b


Specificity
Sensitivity
0
10
20
30
40
50
60
70
80
90
100

c


Specificity
Sensitivity
0
10
20
30
40
50
60
70
80
90
100

d

*Graphical Abstract (for review)

Click here to download high resolution image

Boletus edulis *Tylopilus felleus* *Boletus violaceofuscus*

Dried mushrooms

FT-NIR spectrometer

Multivariate Data Analysis