Ricerca finanziata dal Ministero della Salute
Dipartimento per la Sanità Pubblica Veterinaria,
la Nutrizione e la Sicurezza degli Alimenti

IZSAM G.CAPORALE
TERAMO

Caratterizzazione dell'origine geografica della mozzarella di bufala mediante Risonanza Magnetica Nucleare

Salvatore Milone – Reparto Bromatologia e Residui

MOZZARELLA DI BUFALA

La mozzarella di bufala è un prodotto di derivazione del latte tipicamente italiano, le cui caratteristiche organolettiche sono legate alla provenienza, alla qualità delle materie prime e alle tecniche produttive utilizzate.

In particolare, la "Mozzarella di Bufala Campana", prodotto molto apprezzato dai consumatori, è ottenuta esclusivamente da latte di bufala prodotto in definite aree del Sud Italia.

Il Regolamento (CE) N. 1107/96 ha attribuito alla "Mozzarella di Bufala Campana" il **marchio DOP.** Il disciplinare di produzione definisce le zone di provenienza e trasformazione del latte e le modalità di lavorazione del formaggio.

Iscrizione nel Registro delle denominazioni di origine protette (aggiornato al 24 maggio 2013)

N	Denominazione	Cat.	Tipologia	Regione	Provincia
133	Mozzarella di Bufala Campana	D.O.P.	Formaggi	Campania, Lazio, Molise, Puglia	Benevento, Caserta, Napoli, Salerno, Frosinone, Latina, Roma, Foggia, Isernia

OBIETTIVO

Messa a punto di metodi di analisi basati sulla spettroscopia di Risonanza Magnetica Nucleare (NMR) per **avviare lo sviluppo di un modello** per la caratterizzazione della **mozzarella di bufala** in grado in futuro di verificare la provenienza di campioni incogniti.

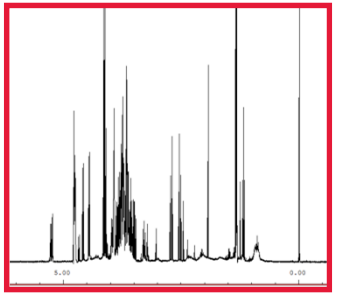
IL METODO

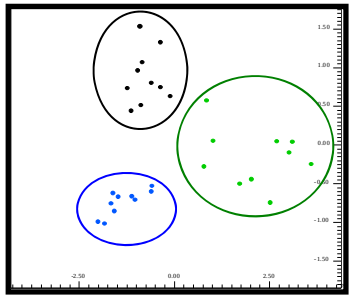
-Caseifici con allevamenti propri

-Caseifici situati in zone interne ed esterne all'area DOP

-Tecnica di analisi in soluzione: ^{1}H HR-NMR

-Tecniche di analisi statistica multivariata
- -Unsupervised (outliers)
- -Supervised (modello)

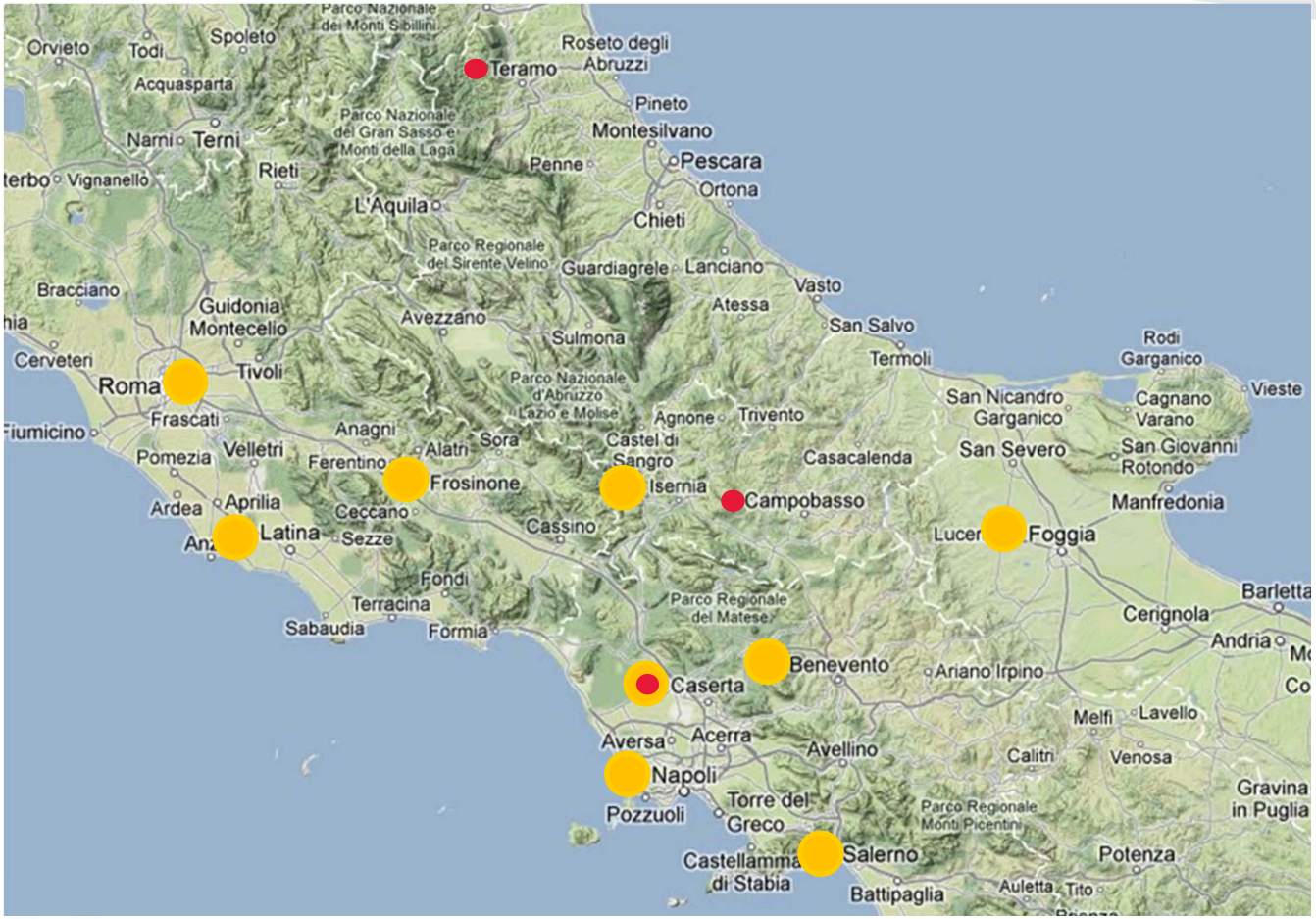




IL PIANO DI CAMPIONAMENTO

		Caserta	Campobasso	Teramo
I campionamento Estate 2009	caseifici	1	1	1
	numero di campioni prelevati	10	10	10
II campionamento Primavera 2010	caseifici	4	1	1
	numero di campioni prelevati	40	10	10
III campionamento Inverno 2011-12	caseifici	7	1	0
	numero di campioni prelevati	66	9	0
Totale	caseifici	7	2	1
	numero di campioni prelevati	116	29	20

IL PIANO DI CAMPIONAMENTO



- *Zone definite nel disciplinare di produzione*
- ***Campionamento effettuato nella zona del Casertano, in provincia di Campobasso e Teramo negli anni 2009-2012.***

IL METODO DI ANALISI

omogeneizzazione

aliquota di omogeneizzato
+
standard interno in D_2O

filtrazione

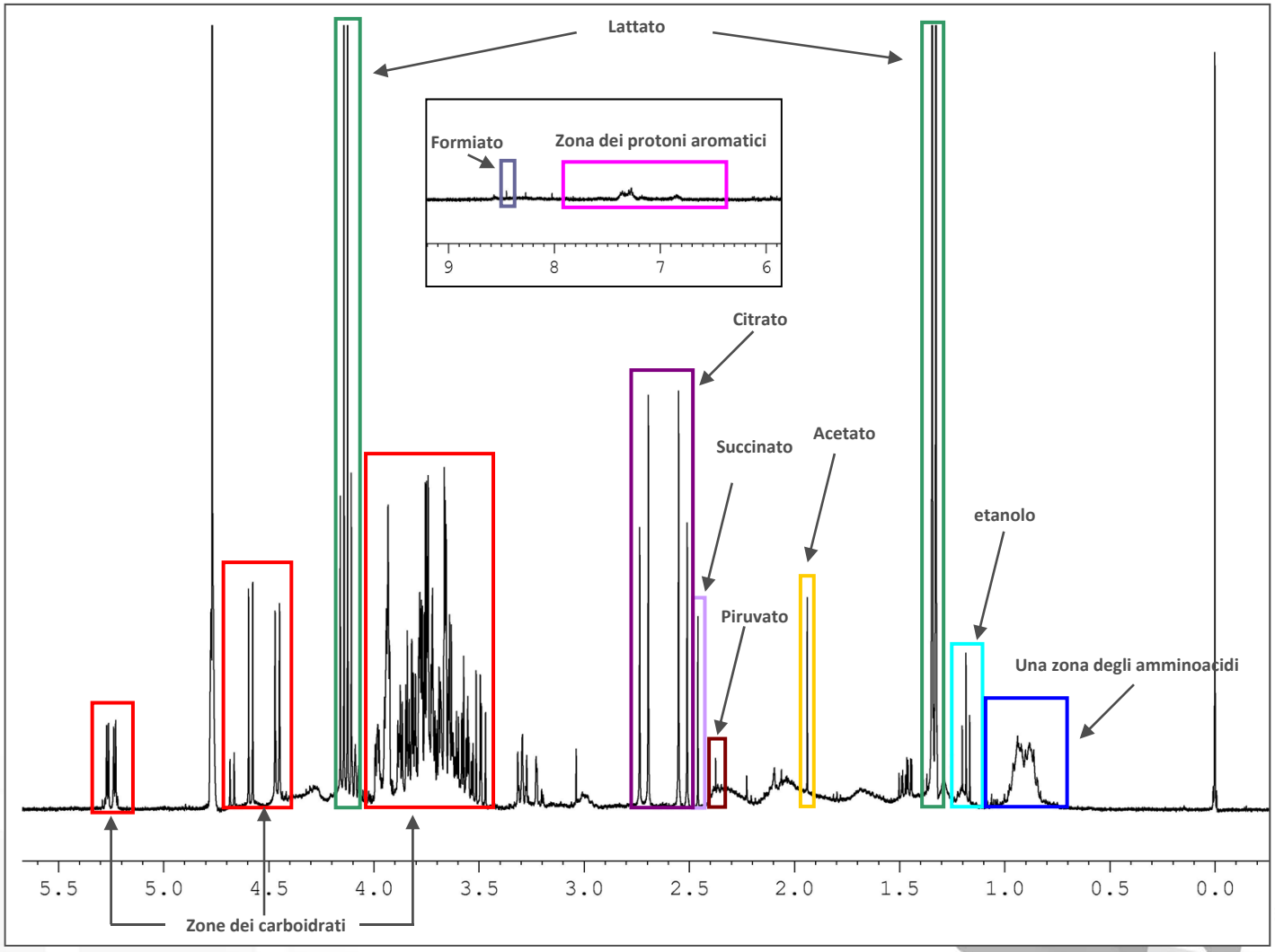
centrifugazione

agitazione

LO SPETTROMETRO NMR

NMR Bruker AVANCE III 400 Ultrashield con probe BBI-z (Broadband Inverse Probe-z gradient) operante a 400,13 MHz sulla frequenza del protone.
Sistema dotato di autocampionatore (B-ACS 60).

SPETTRO ^{1}H HR-NMR DI UN ESTRATTO DI MOZZARELLA



Spettro ^{1}H HR-NMR* in D_2O con presaturazione sul segnale residuo dell'acqua (*noesygppr1d*)

*TopSpin 3.0, Bruker

ANALISI STATISTICA MULTIVARIATA: LINEAR DISCRIMINANT ANALYSIS (LDA)

-Preprocessamento dei dati spettrali*

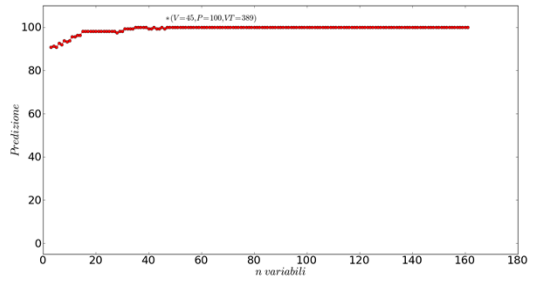
-selezione delle regioni

-rumore spettrale

-normalizzazione PQN

-bucketing

-Selezione delle variabili (ANOVA)**



-Costruzione del modello di classificazione**

-Validazione interna del modello**

-submodello

-cross validation

* R. Lamanna, home made tnmr 1.0.3 software available from the author under GNU license.

** R. Lamanna, home made statnmr 0.1 software available from the author under GNU license.

ANALISI STATISTICA MULTIVARIATA: LINEAR DISCRIMINANT ANALYSIS (LDA)

Normalizzazione degli spettri:

Importante per l'eliminazione di variabili sperimentali indesiderate e di bias.

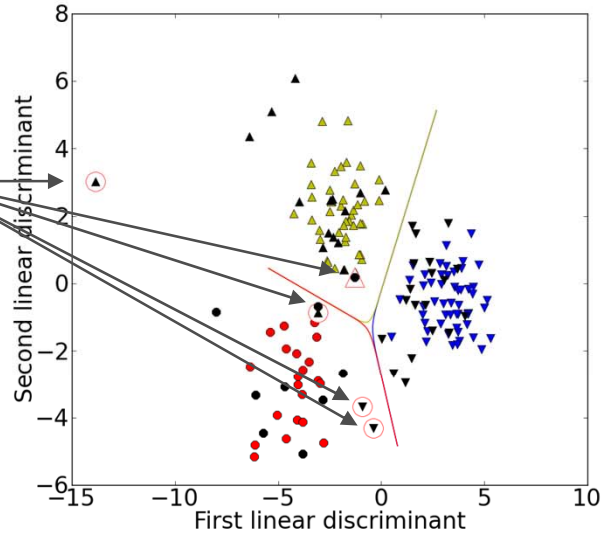
LDA effettuata:

- senza normalizzazione
- con normalizzazione all'integrale
-con l'algoritmo di normalizzazione PQN (Probabilistic Quotient Normalisation).

VALIDAZIONE DEL MODELLO DI CLASSIFICAZIONE

Per poter "visualizzare" le prestazioni del modello ottenuto, è stato utilizzato un processo di validazione in cui 1/3 del numero totale di campioni viene selezionato casualmente e eliminato dal modello.
I campioni rimanenti vengono utilizzati per generare un sub-modello, attraverso il quale i campioni eliminati, e quindi non assegnati a nessuna classe, vengono predetti.

Esempi di errori di predizione



Le linee, che separano le classi, sono le *separation boundaries* e rappresentano le zone in cui un campione ha la probabilità del 50% di essere predetto correttamente nella propria classe.

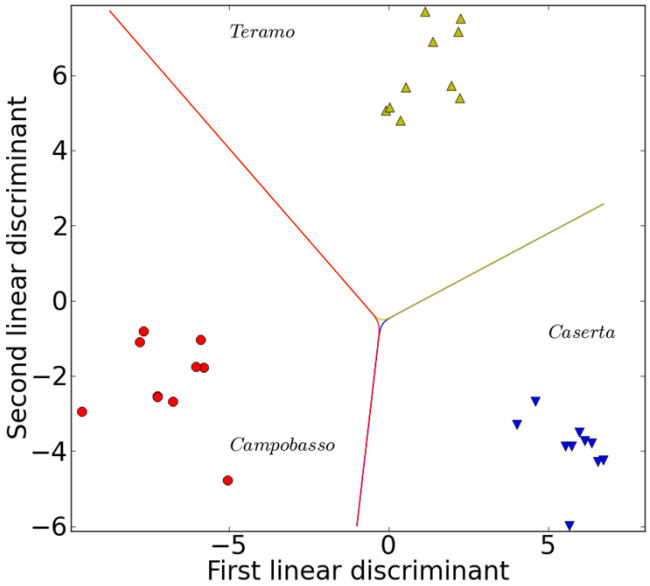
VALIDAZIONE DEL MODELLO DI CLASSIFICAZIONE CROSS-VALIDATION

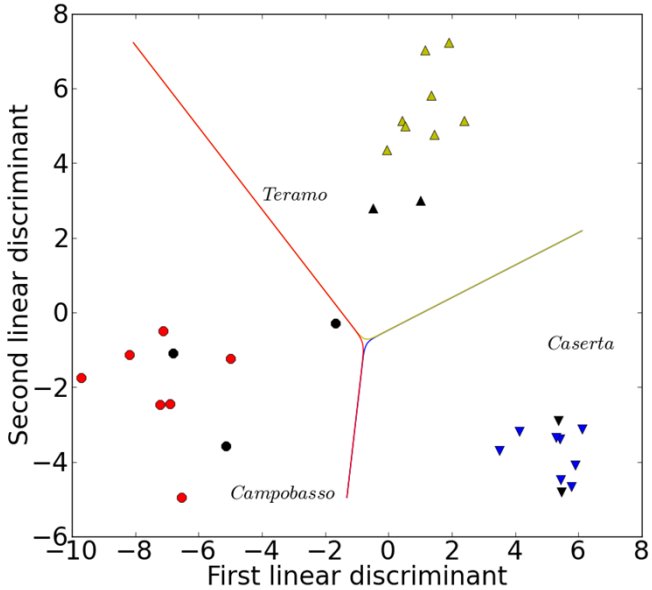
Per passare da un modello visivo al calcolo numerico della performance del modello ottenuto si è utilizzata una cross validation interna (K-fold).

L'approccio consiste nell'assegnare casualmente i campioni a un training set e a un test set. Quest'ultimo contiene circa 1/k dei campioni (k-fold = 10 per classe). Tale divisione permette di avere un numero di campioni sufficiente nel training set e rappresentativo nel test set. Tale processo è ripetuto k volte variando così la popolazione delle classi in modo da assicurare che tutti i campioni abbiano la possibilità di essere inclusi nella valutazione del set almeno una volta.

LINEAR DISCRIMINANT ANALYSIS (LDA)
I CAMPIONAMENTO – ESTATE 2009

30 campioni



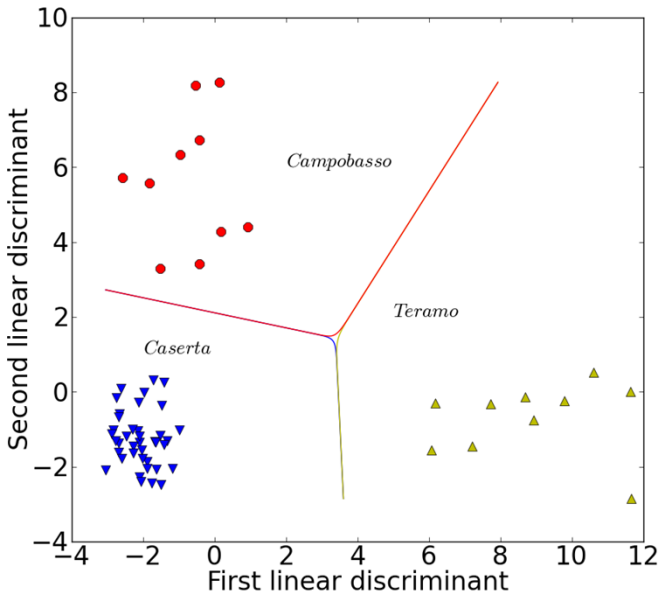


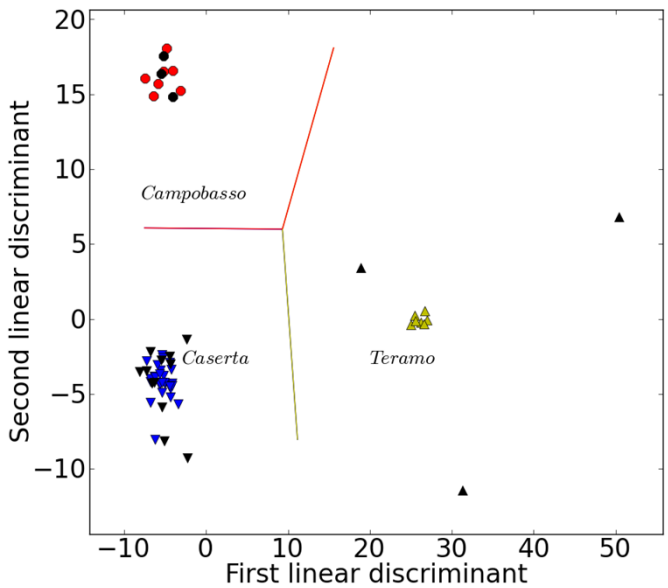
Cross validation con
7 variabili selezionate da ANOVA

Media (%)	**99**
SD (%)	1.704
MAX (%)	100
MIN (%)	97

LINEAR DISCRIMINANT ANALYSIS (LDA)
II CAMPIONAMENTO – PRIMAVERA 2010

59 campioni



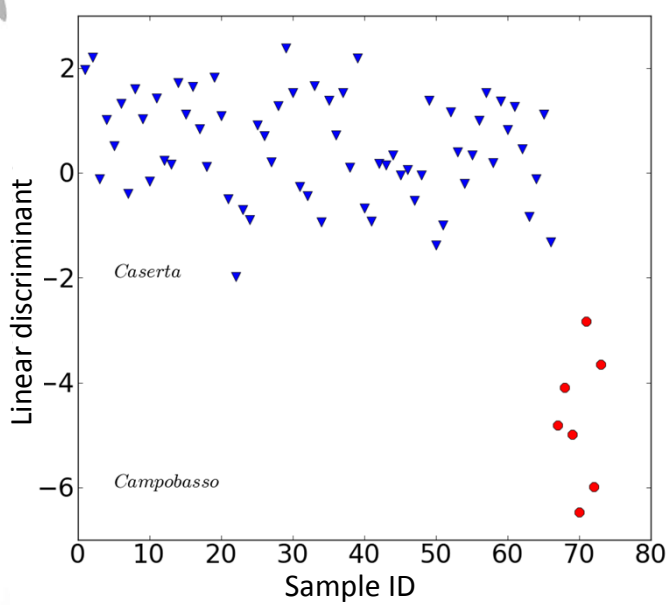


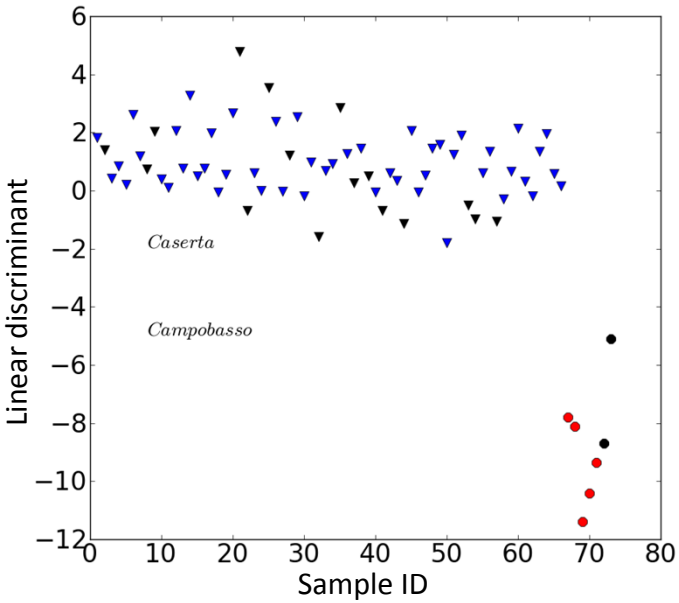
Cross validation con
10 variabili selezionate da ANOVA

Media (%)	**99**
SD (%)	1.414
MAX (%)	100
MIN (%)	96

LINEAR DISCRIMINANT ANALYSIS (LDA)
III CAMPIONAMENTO – INVERNO 2011-12

73 campioni





Cross validation con
23 variabili selezionate da ANOVA

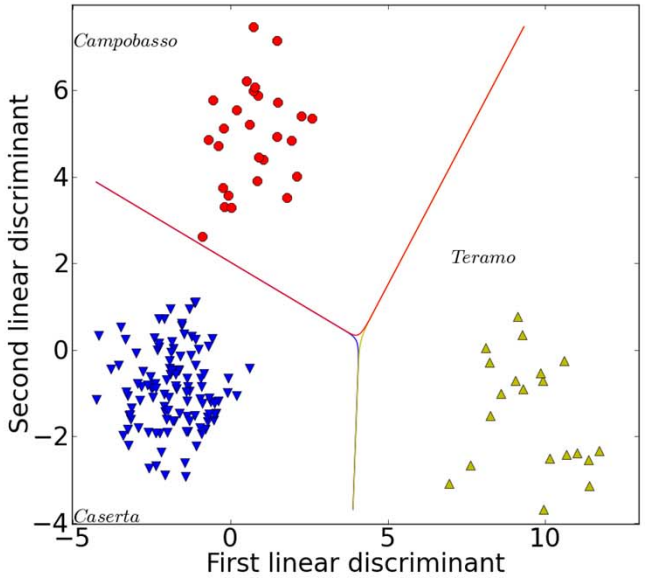
Media (%)	**99**
SD (%)	1.084
MAX (%)	100
MIN (%)	97

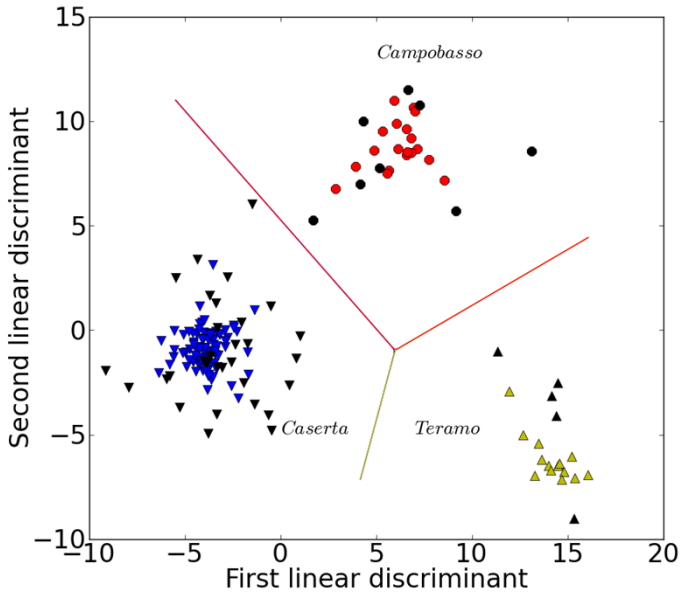
Obiettivo dello studio:

Il modello di classificazione
per origine geografica

LINEAR DISCRIMINANT ANALYSIS (LDA) CASERTA,CAMPOBASSO,TERAMO

162 campioni

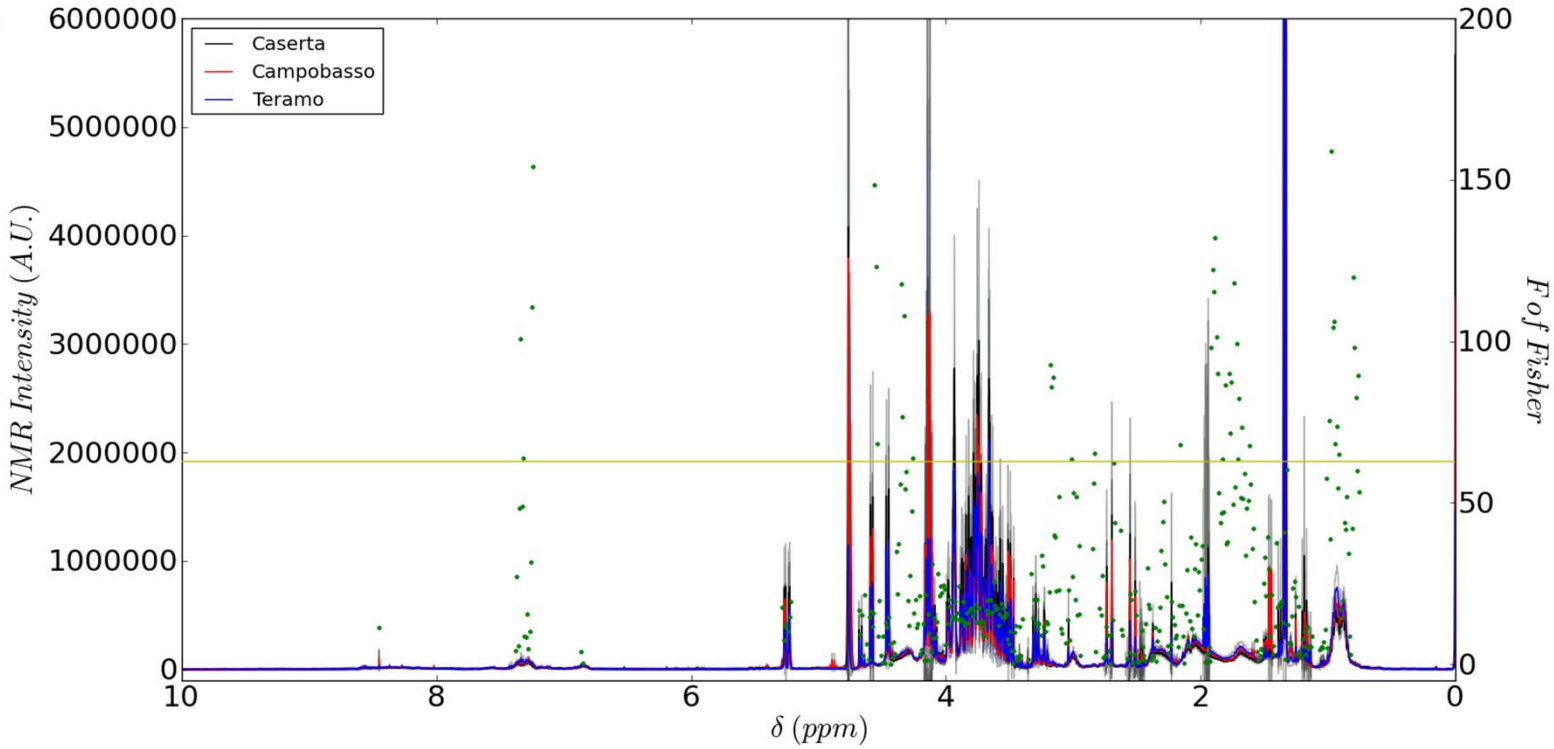




Cross validation con
45 variabili selezionate da ANOVA

Media (%)	**99**
SD (%)	0,846
MAX (%)	100
MIN (%)	97

LINEAR DISCRIMINANT ANALYSIS (LDA) CASERTA,CAMPOBASSO,TERAMO



Sovrapposizione degli spettri medi delle tre classi con tutte le variabili presenti e riduzione di queste al numero minimo per le quali è possibile ottenere una predizione stabile del 100% (45 variabili).

Ampliamento dello studio:

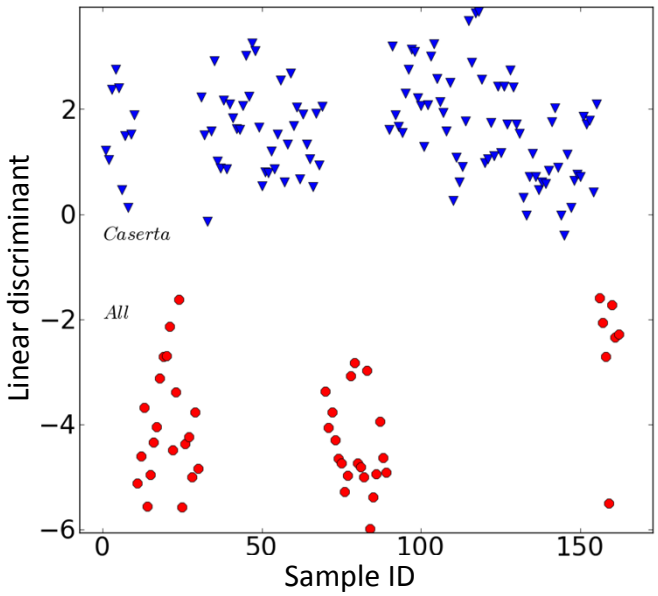
- modelli LDA a due classi
- stagionalità
- confronti
- PLS-DA

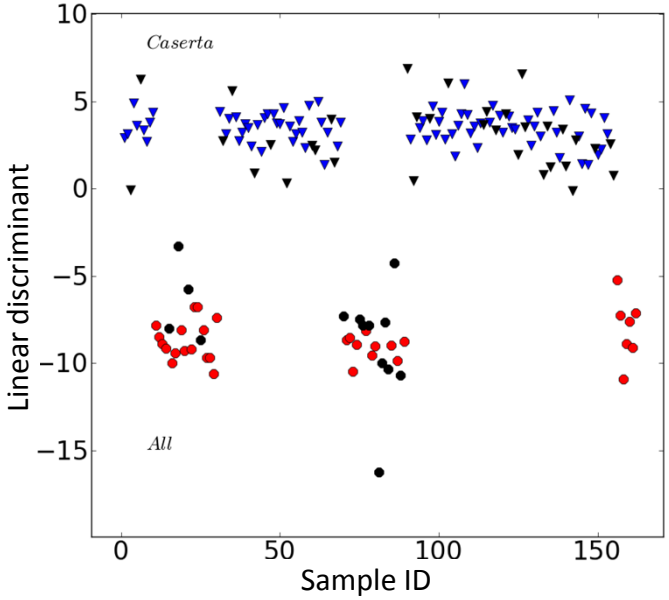
Ampliamento dello studio:

modelli LDA a due classi

LINEAR DISCRIMINANT ANALYSIS (LDA)
CASERTA VS ALL

162 campioni



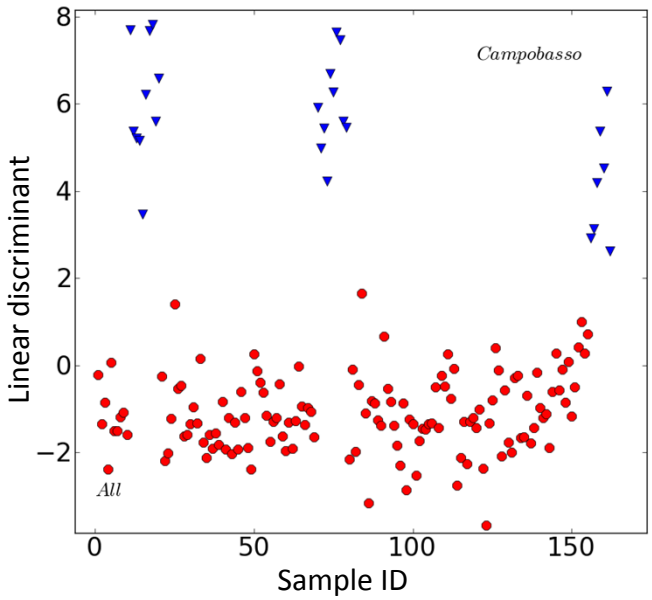


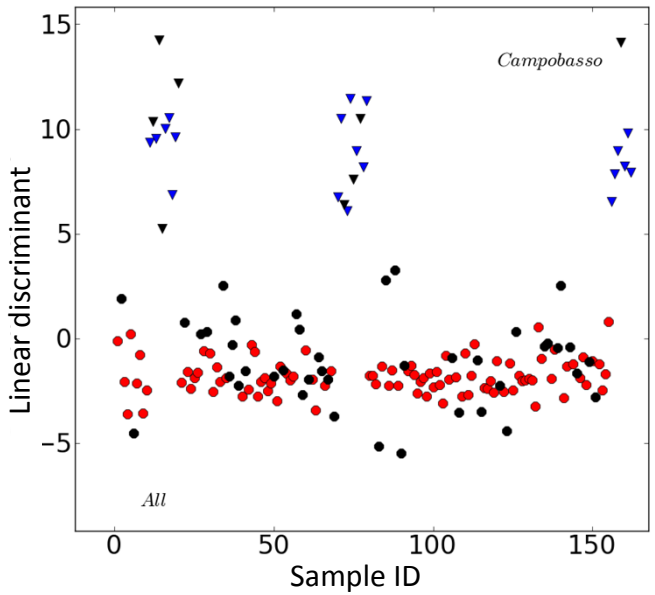
Cross validation con
41 variabili selezionate da ANOVA

Media (%)	**99**
SD (%)	1.137
MAX (%)	100
MIN (%)	97

LINEAR DISCRIMINANT ANALYSIS (LDA)
CAMPOBASSO VS ALL

162 campioni



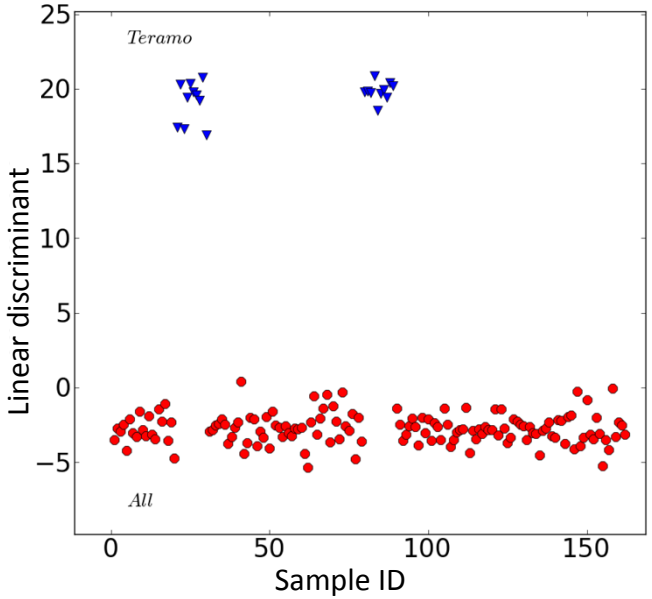


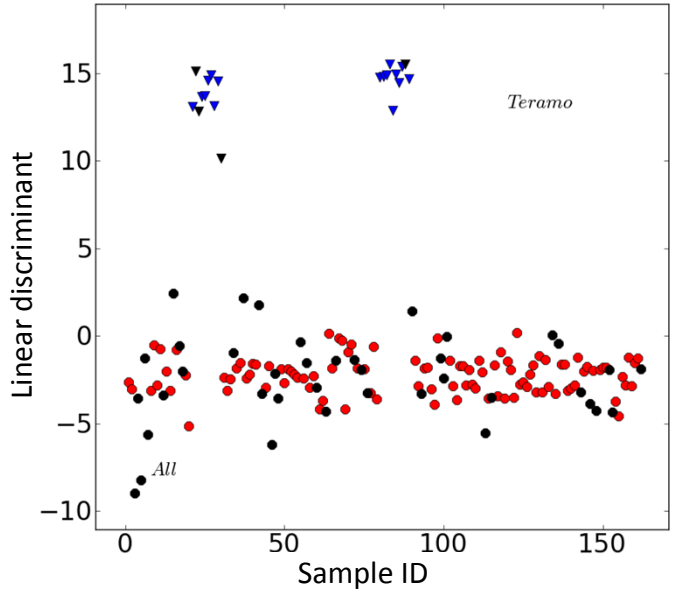
Cross validation con
60 variabili selezionate da ANOVA

Media (%)	**98**
SD (%)	1.418
MAX (%)	100
MIN (%)	95

LINEAR DISCRIMINANT ANALYSIS (LDA)
TERAMO VS ALL

162 campioni





Cross validation con
14 variabili selezionate da ANOVA

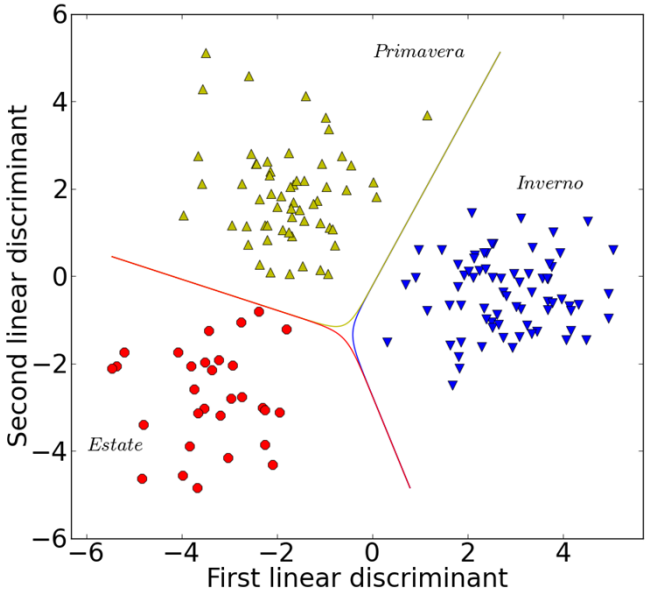
Media (%)	**99**
SD (%)	0.834
MAX (%)	100
MIN (%)	98

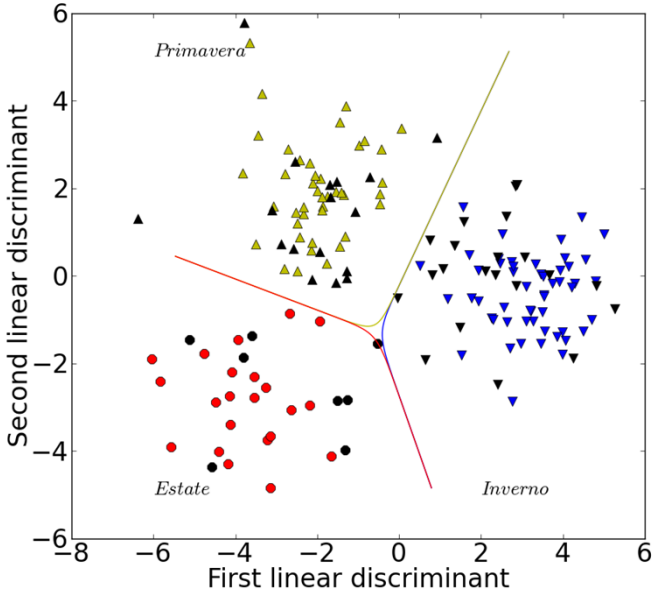
Ampliamento dello studio:

stagionalità

LINEAR DISCRIMINANT ANALYSIS (LDA) STAGIONALITA'

162 campioni



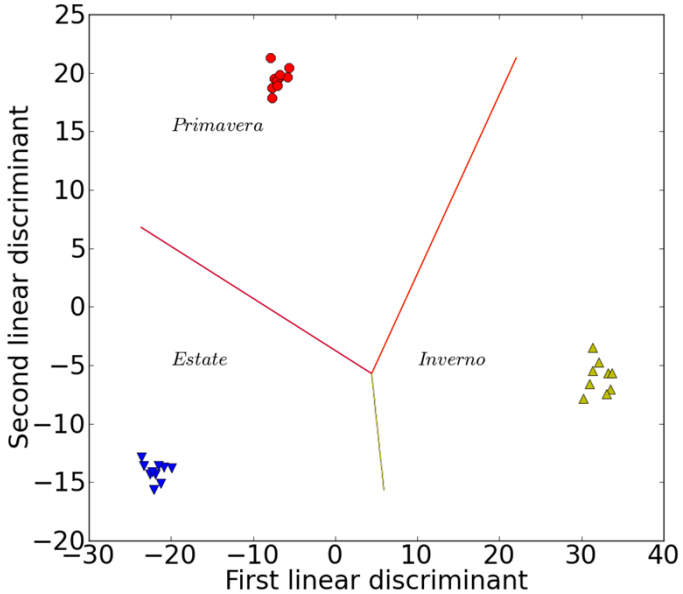


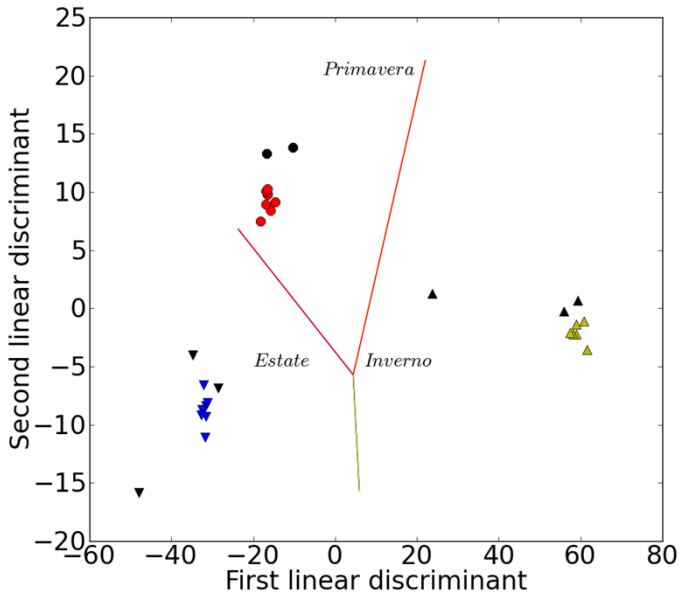
Cross validation con
38 variabili selezionate da ANOVA

Media (%)	**98**
SD (%)	2.113
MAX (%)	100
MIN (%)	93

LINEAR DISCRIMINANT ANALYSIS (LDA)
UN CASEIFICIO DIVERSE STAGIONI

30 campioni





45 variabili selezionate da ANOVA

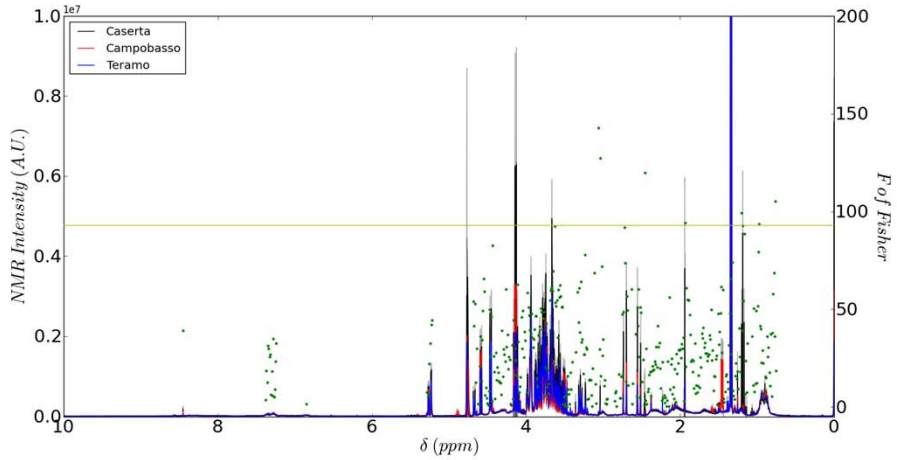
Ampliamento dello studio:

confronti

Sovrapposizione degli spettri medi delle tre classi con tutte le variabili presenti e riduzione di queste al numero minimo per le quali è possibile ottenere una predizione stabile del 100%.

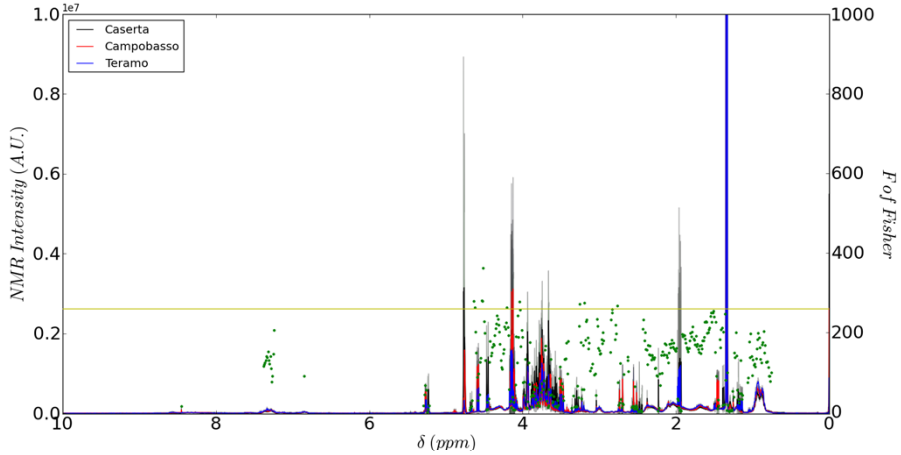
I campionamento

7 variabili



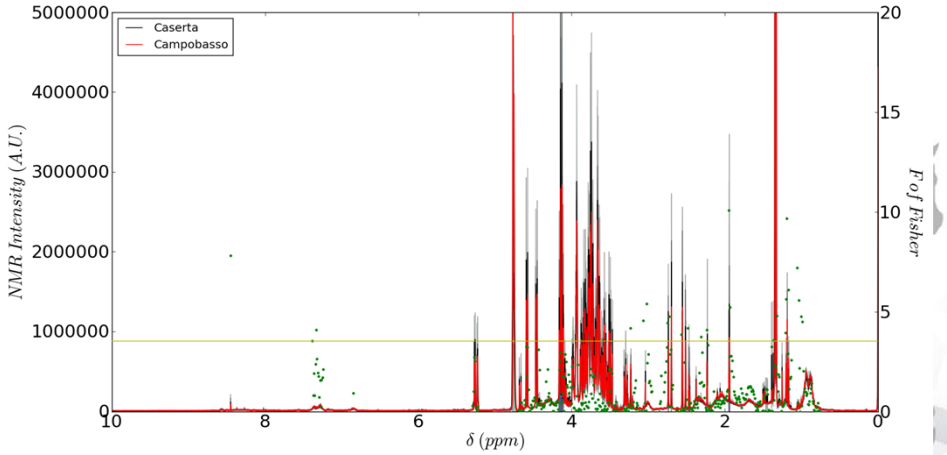
II campionamento

10 variabili



III campionamento

23 variabili

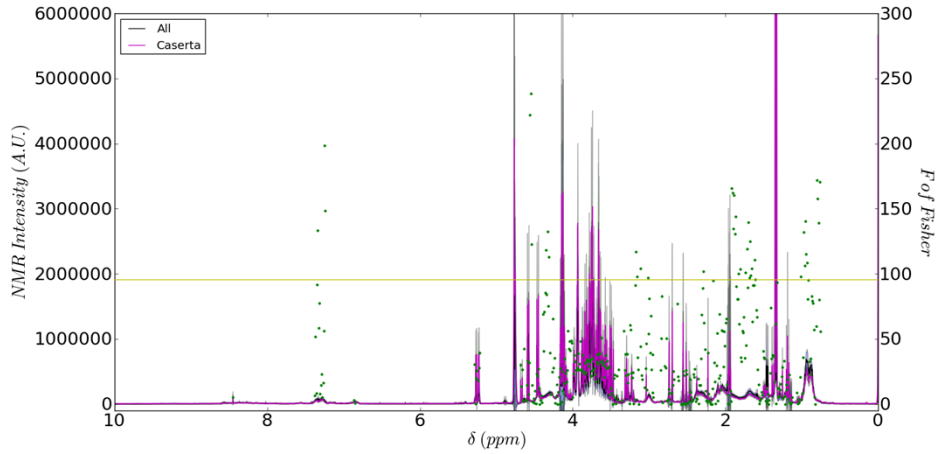


IZSAM G.CAPORALE TERAMO

162 campioni

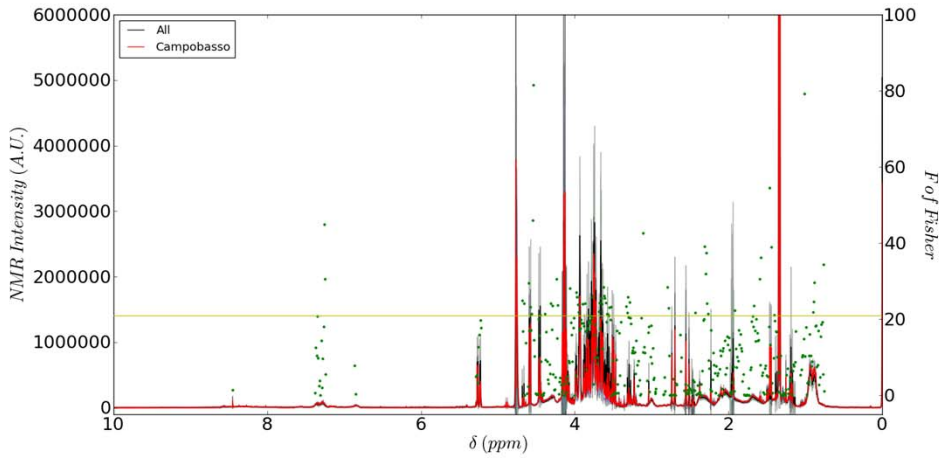
Caserta vs all

41 variabili



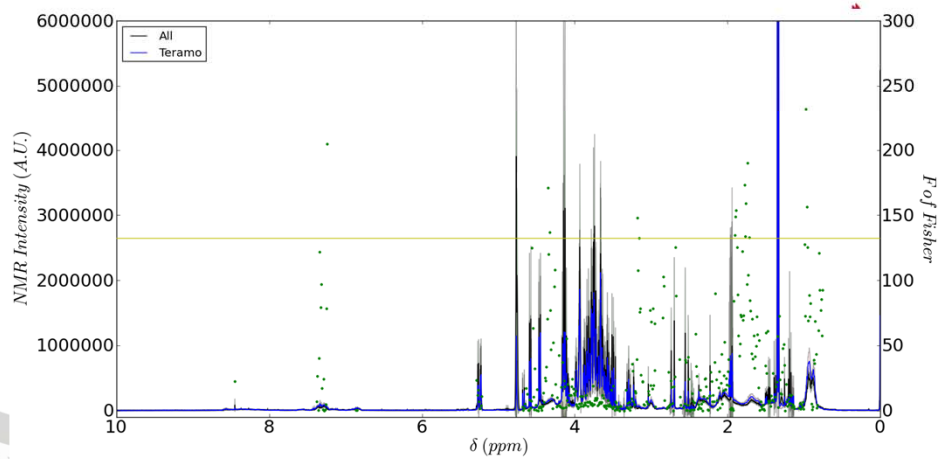
Campobasso vs all

60 variabili



Teramo vs all

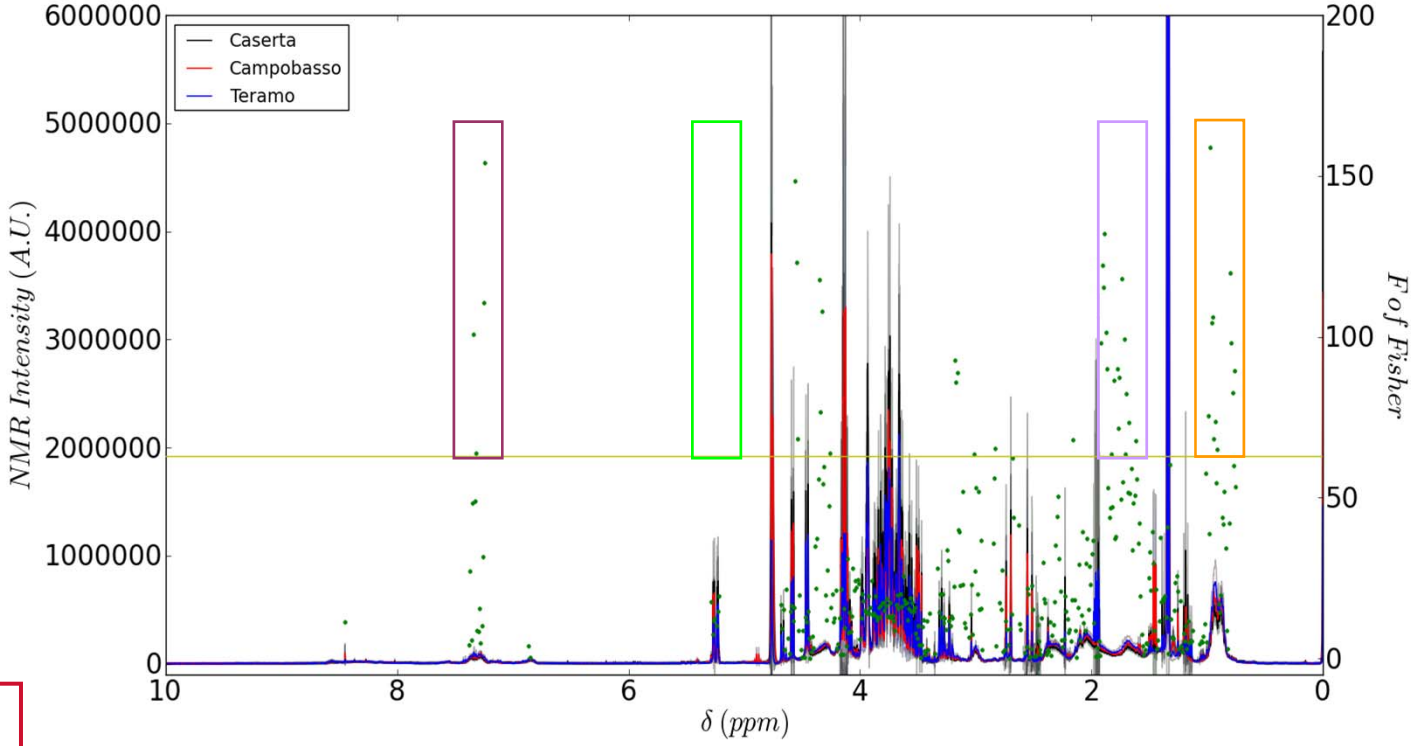
14 variabili



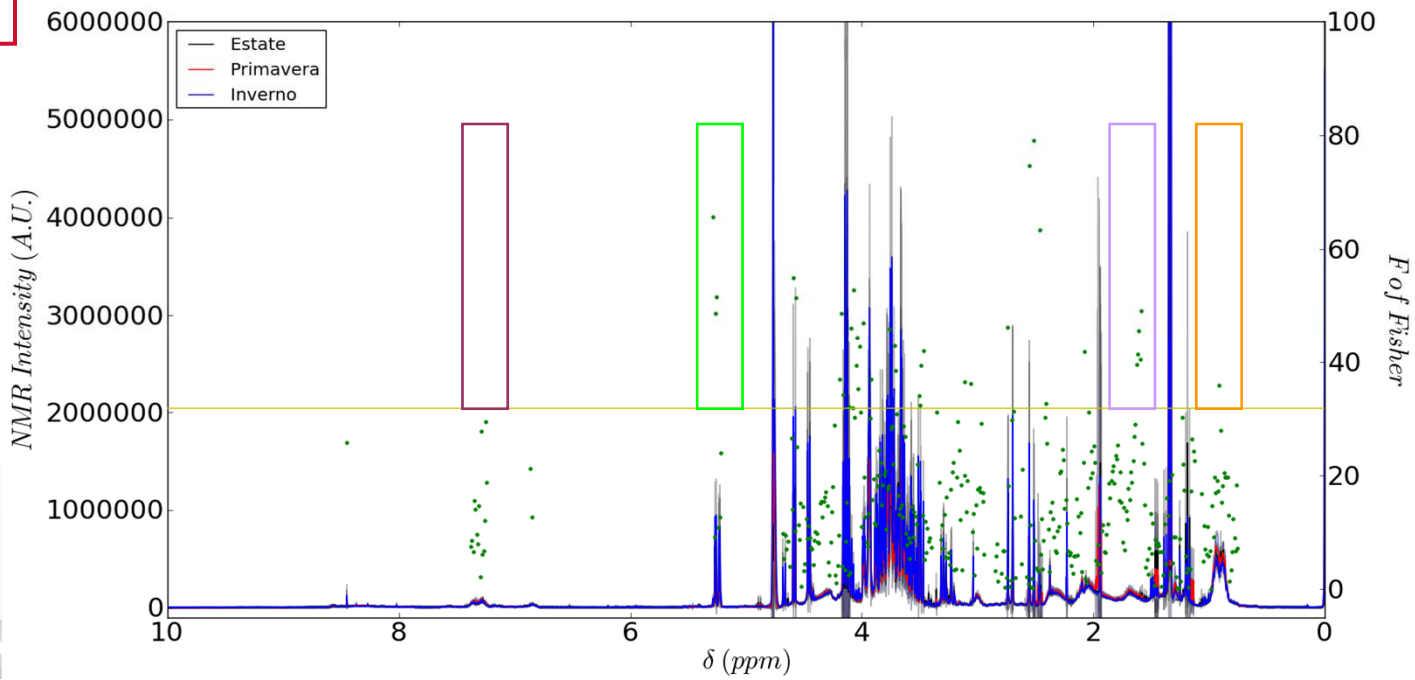
IZSAM G.CAPORALE TERAMO

Origine Geografica
(45 variabili)

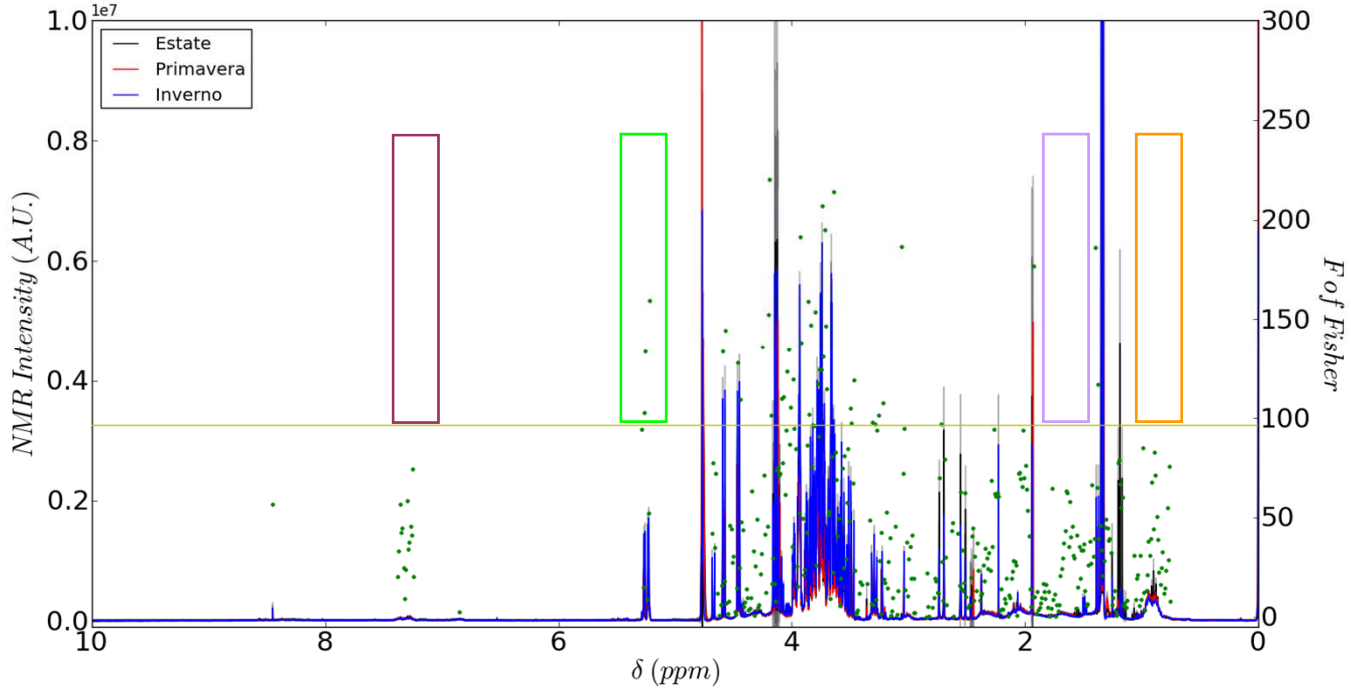
Stessi 162 spettri



Stagionalità
(38 variabili)

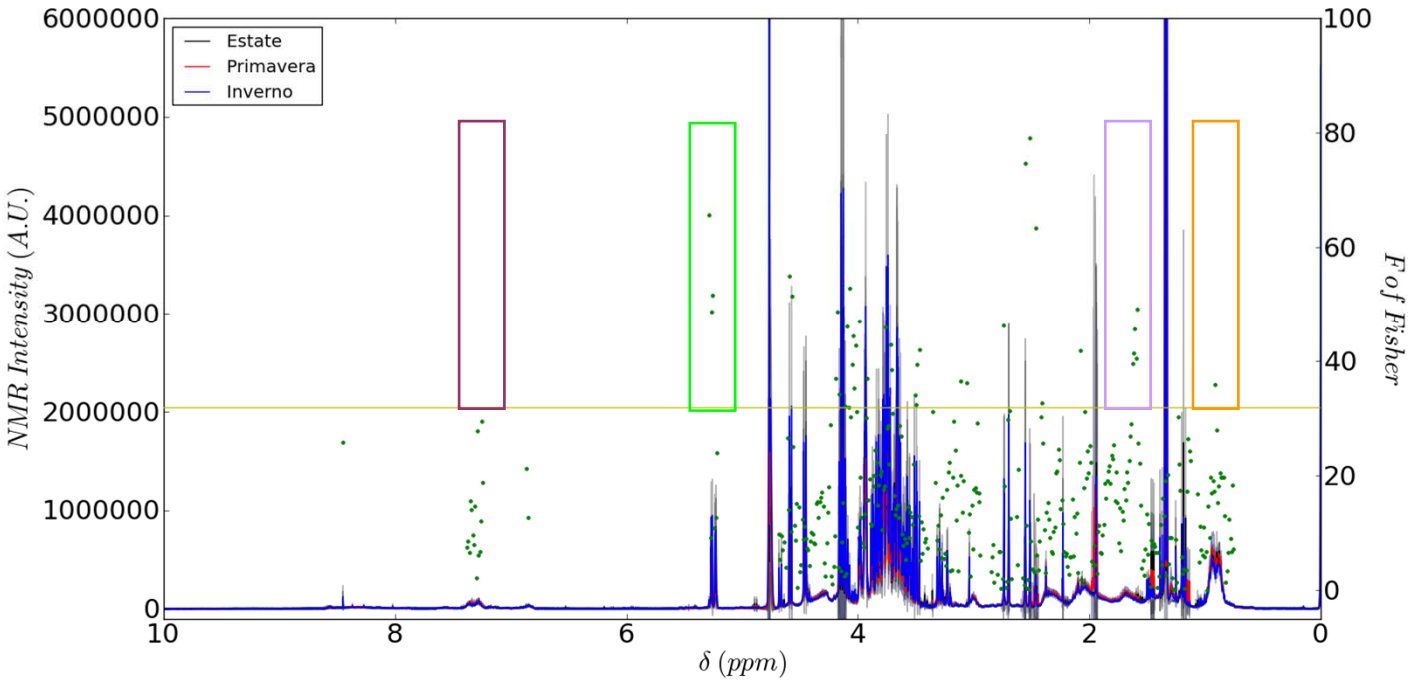


Un caseificio
30 spettri
(45 variabili)



Stagionalità

Totale
162 spettri
(38 variabili)



Ampliamento dello studio:

PLS-DA

PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS
PLS-DA*

-Preprocessamento dei dati spettrali
- -selezione delle regioni
- -integrazione e scalatura
- -bucketing

-Selezione delle componenti PLS

-Costruzione del modello di classificazione

-Validazione interna del modello
- -cross validation
- -submodelli e predizione

*AMIX Statistics version 3.9.12, Bruker

PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS PLS-DA*

Origine geografica

162 campioni

Cross validation con 52 variabili latenti:

assegnazioni corrette	161
assegnazioni errate	1
campioni correttamente assegnati (%)	**99.4**

$$Q^2 = \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y})^2} = 0.92$$

CONCLUSIONI

Questo lavoro effettuato con l'utilizzo della tecnica NMR in combinazione con l'analisi statistica multivariata, mostra la possibilità di differenziare i campioni di mozzarella di bufala oggetto di studio in funzione dell'origine geografica.

Elenco dei collaboratori

Dr. Giampiero Scortichini, Responsabile Scientifico

Dr.ssa Alessandra Micozzi, Chimico
Dr.ssa Arianna Stramenga, Chimico
Dr. Salvatore Milone, Chimico

Ringraziamenti

Si ringraziano:

Personale *Bruker BioSpin*

Dr. Raffaele Lamanna (*ENEA-Trisaia)*

Dr. Vincenzo D'Andrea *(ASL Caserta)*

Dr. Lucio Marino (*IZSAM Campobasso*)

I proprietari dei caseifici

Grazie per l’attenzione

Per contatti:

g.scortichini@izs.it

a.micozzi@izs.it

s.milone@izs.it