

# Utilizzo di una metodica “fingerprinting” per l’identificazione dell’origine botanica del miele

Scaramagli S.<sup>1</sup>, Gottardi F.<sup>1</sup>, Zappi A.<sup>2</sup>, Melucci D.<sup>2</sup>, Marcazzan GL.<sup>3</sup>

<sup>1</sup> COOP ITALIA Soc. Coop. – Direzione Qualità

<sup>2</sup> Dipartimento di Chimica – Università di Bologna

<sup>3</sup> CRA-API - Bologna



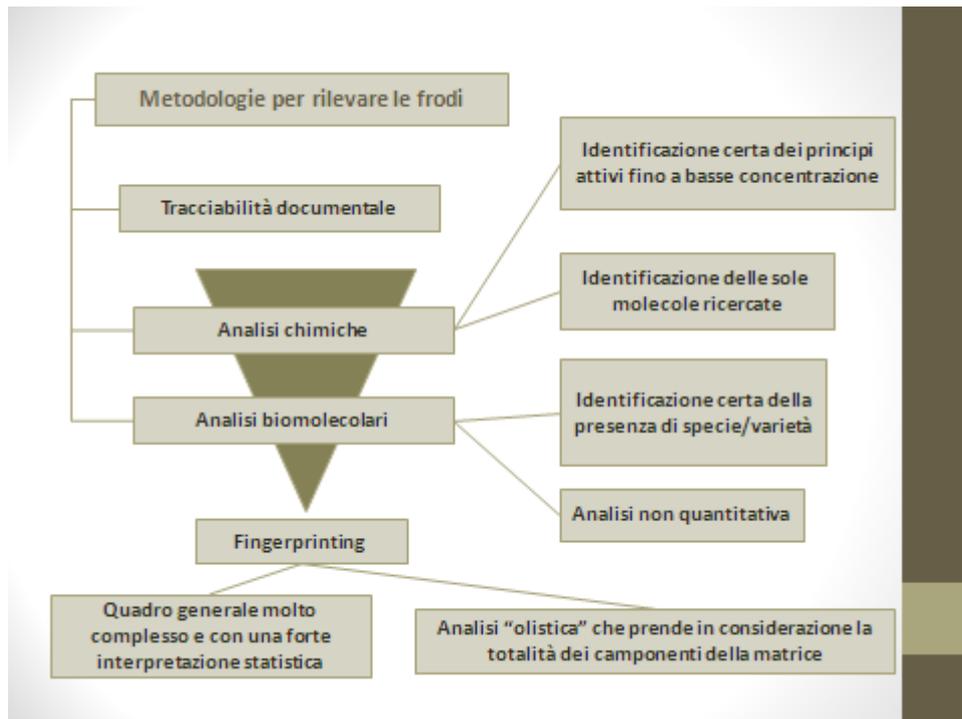
L’esperienza qui presentata è stata eseguita con un’analisi non convenzionale. Coop è impegnata ampiamente nel discorso della sicurezza e della autenticità alimentare.



Questa è una tabella presa da un articolo del 2010, non molto recente ma abbastanza attuale, da cui si evince che tra i prodotti alimentari più frodati, il miele è in terza posizione.

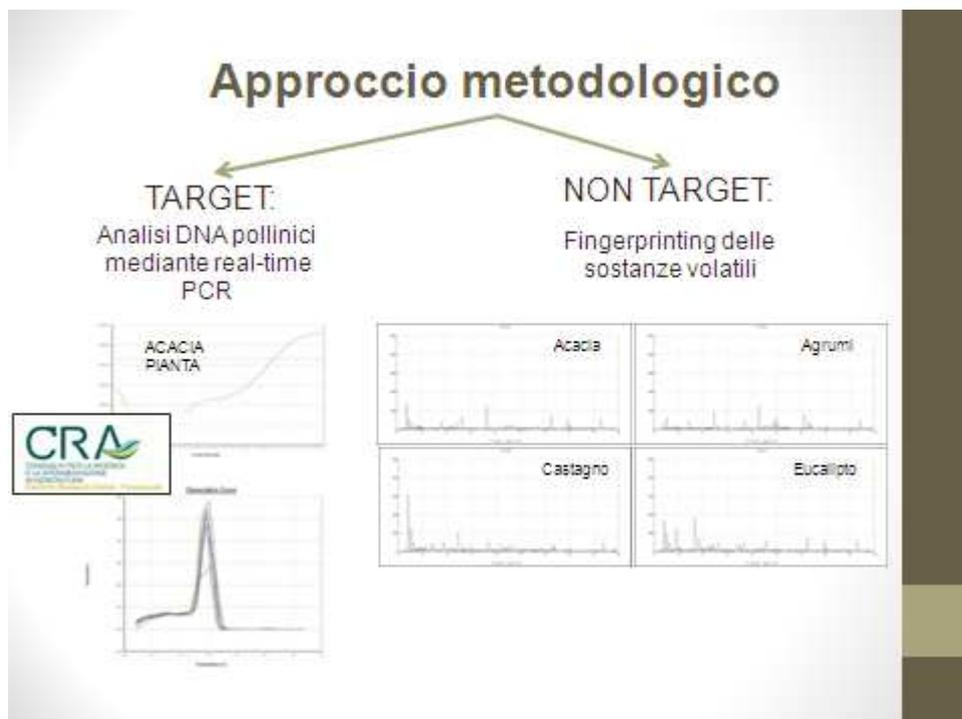
Questa frode ha provocato un danno di più di 400 milioni di euro nel 2013.

La tabella, nella quale sono classificati i vari tipi di frode, mostra che il 95% delle frodi registrate è legato allo scambio e in particolare alle false dichiarazioni di origine geografica, di specie botanica ecc..



La domanda che ci siamo posti è: «Quali sono le metodologie per rilevare le frodi?»

I nostri specialisti di prodotto hanno in mano diversi strumenti per poter rilevare le frodi: il primo è la tracciabilità documentale che deve poi essere supportata da prove analitiche di tipo chimico e di tipo biomolecolare. Le due tipologie di analisi sono state messe insieme perché la nostra esperienza di lavoro, che ci vede ogni giorno alle prese con tante matrici alimentari diverse, ci ha insegnato che non esiste l'analisi perfetta ma esiste un insieme di analisi, una sinergia di analisi, che comprende anche una sinergia di discipline. Infatti, se le analisi biologiche possono dare indicazioni precise sulla composizione qualitativa delle specie polliniche presenti in un miele, le analisi chimiche possono invece fornire dati precisi sulle quantità di principi nutrizionali. E' il quadro di insieme che infine bisogna leggere per capire la qualità di un prodotto.

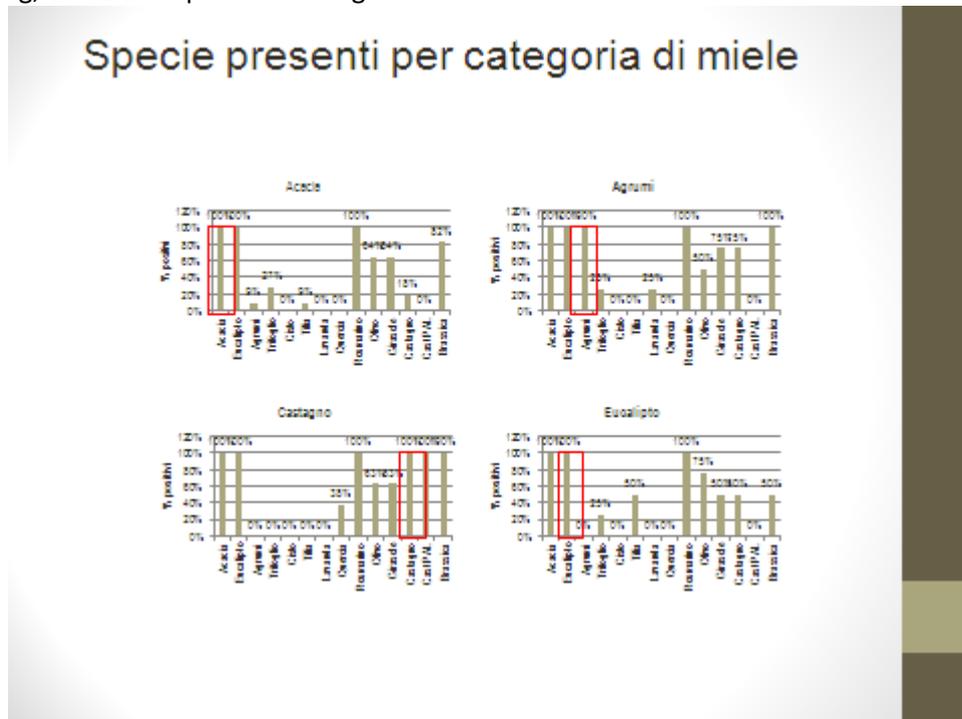


Negli ultimi anni sta prendendo piede un tipo di analisi untarget che si chiama fingerprinting. Di cosa si tratta? Il fingerprinting restituisce un quadro complessivo della matrice alimentare, quasi un quadro

olistico se si può dire, che però non è molto sensibile: riesco a vedere tutto quello che c'è nella matrice anche se non in grandi quantità.

Le analisi fingerprinting si possono affrontare in modo diverso a seconda dello strumento e della categoria di composti che vado ad analizzare.

La nostra esperienza sull'autenticità alimentare è cominciata 10 anni fa con le analisi del DNA, e anche il primo approccio con la matrice «miele» è stato biomolecolare, quindi con un tipo di analisi «target». In seguito abbiamo deciso di affrontare questa matrice con questo nuovo tipo di analisi, e cioè con il fingerprinting, senza uno specifico bersaglio.

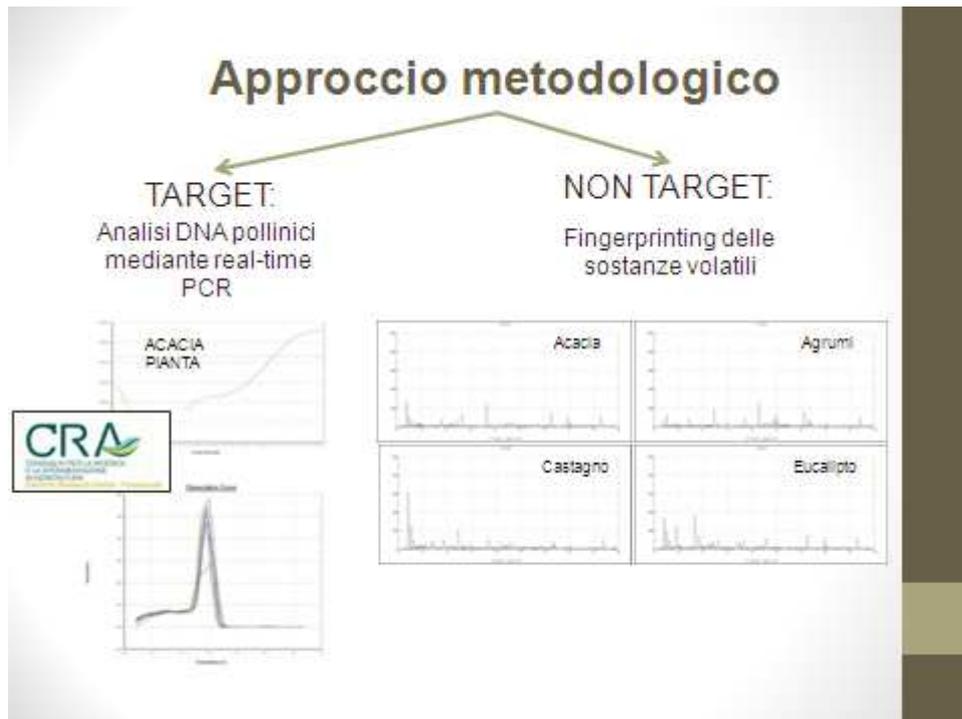


Brevemente vi illustro come abbiamo affrontato l'argomento e quali sono stati i suoi limiti.

Il primo approccio è stato quello di estrarre il DNA dal polline che si trova nel miele, e di capire attraverso l'analisi del DNA a quali piante appartenevano i granuli pollinici contenuti nel miele. Nel grafico vengono illustrati i risultati di queste analisi. Questa attività è stata eseguita con la collaborazione del CREA di Firenze, che ci ha aiutato dal punto di vista della genetica delle piante, e del CREA API, che ci ha fornito i mieli già identificati con analisi melissopalinochimica e quindi certi (è molto importante partire da un database certo).

Un esempio: tutti i mieli di acacia sono risultati contenere DNA di acacia al 100%, ma anche il 100% di eucalipto, c'era sempre del rosmarino, un po' meno delle altri tipi di piante. Gli esperti mi dicono: «bello fantastico, però ci aiuta poco perché in realtà l'analisi che dovresti fare per una classificazione deve essere quantitativa!» Purtroppo questo tipo di analisi al momento è solo qualitativa.

A questo punto abbiamo pensato di lasciare da parte questo tipo di analisi (target) e vedere cos'altro si può fare ad esempio con analisi di tipo non target.



Tra le tante possibilità di analisi non target visto che noi non lavoriamo solo sul miele ma abbiamo molteplici matrici alimentari da analizzare, abbiamo scelto di concentrarci sui composti volatili. Ci sono svariate ragioni che motivano questa scelta, una di queste è che non c'è niente di più caratteristico di un prodotto che la sua componente volatile cioè l'insieme di quei composti che vanno a dare il gusto e l'aroma. La nostra matrice di partenza è stata l'olio extravergine di oliva, il primo della "top ten" di alimenti frodati che avete visto prima, e sia nel caso dell'olio extravergine d'oliva sia nel caso del miele, la componente volatile è legata all'aspetto sensoriale che sappiamo essere molto importante per la classificazione e la qualità in questo tipo di matrici.

## Composti volatili

Mieli	Composti	Riferimenti
CASTAGNO	Alta percentuale di acido fenilacetico, bassa percentuale di acido benzoico	Jelencic and others 2007
	2-aminocaproline	Benjara and others 2009
	β-cedrene, fenilacetato, acido (11-β)-acetato di bromellane, pentadiene	Radovic and others 2001
	3-Metilciclopentano, α-acido omelibetico, α-omelibetico acetato	Quoddi and Vitali 1995 Guyot and others 1995
	3-Metilciclopentano, 1-Metilacetato	Zorvari and Coli 2002
AGRUMI	alcoli, acido ascorbico, acido citrico, α-pinene	Alsenzhofer and others 2007
	Cesce di metilacetato (2-Hinolo)	Wills and Dryen 1995
	Alcoli, isomeri	Castro-Vazquez and others 2009
	o-4-dimetil-2-ciclopentene-1-acetato Diole terpeni	Pisenzolo and others 2003 de la Fuente and others 2005
LAVANDA	β-cedrene, α-pinene, β-cedrene, cumarone, alte concentrazioni di acetato	Radovic and others 2001
	1-estrone, n-estrone, fenilacetato, cumarone, acetato, fenilacetato	Castro-Vazquez and others 2009
ACACIA	Cesce di β-pinene ad α-pinene	Radovic and others 2001
ROSMARINO	1-estrone acetato isomeri di acetato	Castro-Vazquez and others 2009

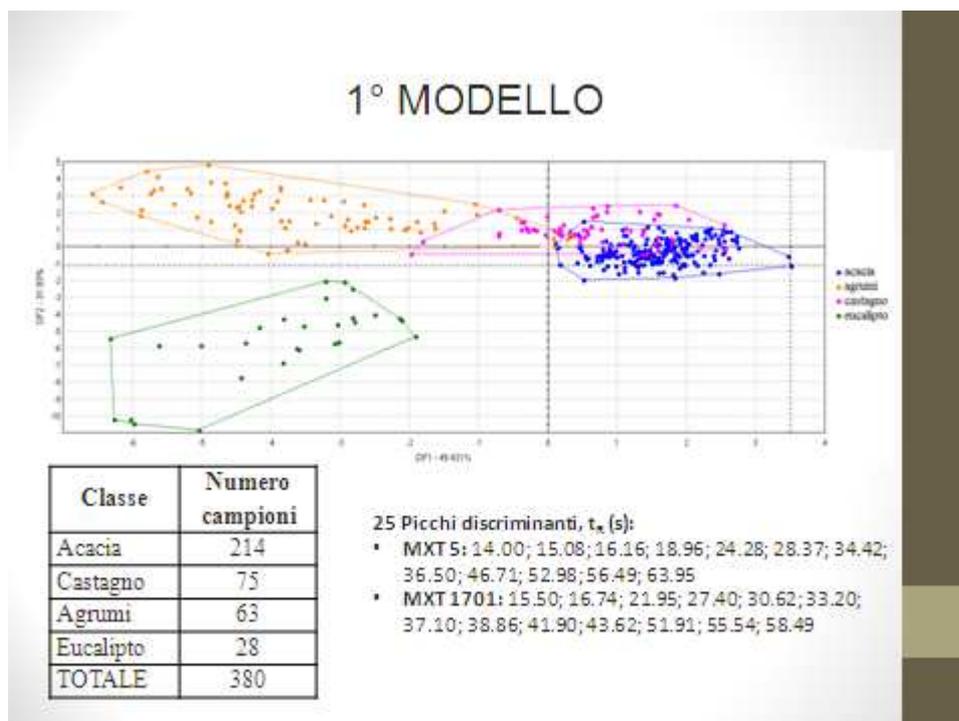
Questa tabella dimostra come l'analisi dei volatili nel miele è stata portata avanti già da parecchi anni da diversi autori .

Le analisi che vedete elencate qui però sono tutte analisi target e cioè i vari autori sono andati a ricercare nei diversi mieli dei composti ben definiti e caratteristici per le diverse tipologie.



noti. Grazie al CREA API siamo riusciti ad acquisire un certo numero di campioni noti. Alla fine abbiamo ottenuto un numero di dati veramente enorme e quindi ci siamo rivolti all'Università di Bologna, in particolare alla professoressa Dora Melucci, che è una chemiometra e abbiamo messo a punto una procedura chemiometrica che ci possa aiutare a classificare e a organizzare questi dati.

Ogni corsa, infatti, restituisce un profilo degli aromatici che può essere di circa 100-150 picchi perché i composti aromatici all'interno delle matrici alimentari sono veramente tanti; però io voglio dividere i miei campioni in base all'origine botanica quindi castagno, acacia ecc. La procedura chemiometrica prevede la ricerca, attraverso un processo matematico-statistico, di quei picchi che caratterizzano ciascuna origine e quindi opero una selezione di quelle che i chemiometri chiamano variabili discriminanti.



Quindi da 150 picchi si scende a 20-25 picchi caratteristici per il gruppo dell'acacia, degli agrumi ecc. poi si costruisce un modello LDA che è una visualizzazione grafica della distanza tra le classi, intendetele come categoria di miele, in cui ci sono i campioni che vengono raggruppati per le differenze e per le somiglianze. Quelli uguali, con lo stesso profilo, vengono messi insieme, quelli che hanno profili differenti vengono messi in un'altra classe. Questo tipo di modello statistico ha una possibilità interessante che è quello di poter «proiettare». Se il modello è robusto posso proiettare un campione incognito e cioè posso chiedere al sistema a quale classe appartiene, a quale il suo profilo è più simile.

Il primo modello che è stato costruito è quello che vedete nella slide. All'inizio il numero dei campioni era abbastanza limitato, ci eravamo limitati solo ai 4 mieli commerciali che noi vendiamo. Bisognerebbe superare i 100 campioni per categoria per riuscire ad avere un modello statistico robusto (questo è stato costruito con dei mieli che sono stati certificati e sono sicuri). Nella slide vedete anche i 25 picchi discriminanti, non hanno dei nomi perché stiamo lavorando in fingerprinting, non sappiamo a quale composto corrispondano questi picchi al momento.

## 1° MODELLO

Put into Group	True Group			
	Acacia	Agrumi	Castagno	Eucalipto
Acacia	206	5	4	0
Agrumi	3	69	2	2
Castagno	5	1	57	0
Eucalipto	0	0	0	26
Total N	214	75	63	28
N correct	206	69	57	26
Proportion	0.963	0.920	0.905	0.929
N = 380				
N Correct = 358				
Proportion Correct (NER) = 0.942				

Come si può vedere dalla tabella i gruppi sono abbastanza discriminati tra loro, infatti il test di cross-validazione indica che il modello da un 94% di risposte corrette.

## 1° MODELLO

Classe	Numero di oggetti	Numero di oggetti classificati correttamente dal modello
Acacia	68	42
Agrumi	4	4
Cardo	4	0
Castagno	32	22
Tiglio	16	0

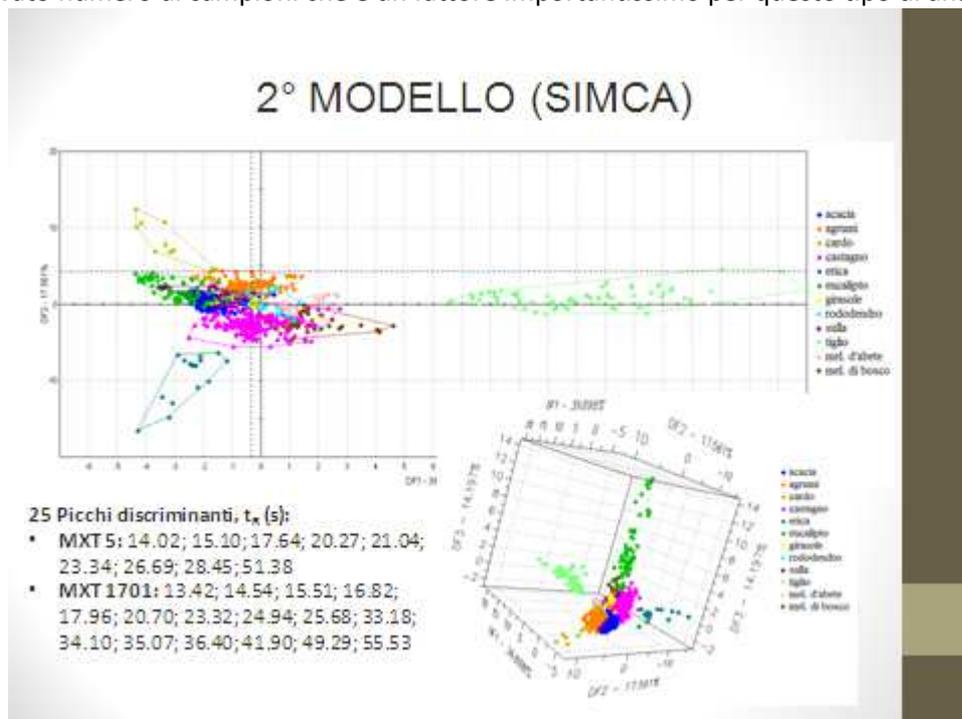
Classificazioni corrette: 71,0%

Il secondo test dal nome: “matrice di confusione” prende dei campioni estranei al modello, come i mieli di cardo e tiglio che non erano all’interno del modello e ci dice quanto è corretta la classificazione; il test restituisce un 71% di classificazioni corrette e tenendo conto che 2 gruppi non erano presenti nel modello il valore ottenuto è buono dal punto di vista di una analisi tipo predittivo.

## 2° MODELLO (SIMCA)

Classe	Numero oggetti
Acacia	333
Agrumi	183
Cardo	16
Castagno	254
Erica	16
Eucalipto	77
Girasole	20
Rododendro	19
Sulla	57
Tiglio	77
Melata d'abete	32
Melata di bosco	55
TOTALE	1139

Poi abbiamo costruito un secondo modello con più categorie e più campioni per categoria; grazie infatti alla collaborazione con il CREA API e in particolare con il dr Marcazzan abbiamo avuto la possibilità di analizzare un elevato numero di campioni che è un fattore importantissimo per questo tipo di analisi.



Questo è il modello sembra graficamente più confuso, con maggiore sovrapposizione tra i gruppi ma in realtà se si guardate il modello tridimensionale vedete che le classi sono più discriminate di quanto supposto in prima battuta.

## 2° MODELLO (SIMCA)

Put into Group	True Group											
	Abete	Acacia	Agrumi	Bosco	Cardo	Castagno	Erica	Bucalpto	Girasole	Rododend.	Sulla	Tiglio
Abete	32	0	0	2	0	0	0	0	0	0	0	1
Acacia	0	226	18	1	0	13	0	3	0	4	7	0
Agrumi	0	8	147	0	1	1	0	0	0	3	7	1
Bosco	0	0	1	43	0	26	0	0	0	1	0	0
Cardo	0	0	4	0	11	4	0	0	0	0	0	0
Castagno	0	0	1	7	0	202	0	0	0	0	0	0
Erica	0	0	0	0	0	0	18	0	0	0	0	0
Bucalpto	0	0	0	0	0	0	0	71	0	0	1	0
Girasole	0	3	1	0	0	3	0	1	18	0	7	0
Rododendro	0	0	1	0	0	2	0	0	0	11	8	0
Sulla	0	13	12	0	4	1	0	0	2	0	27	0
Tiglio	0	0	0	0	0	0	0	0	0	0	0	75
Total N	32	226	162	53	16	214	18	77	20	19	37	77
N correct	32	226	147	43	11	202	18	71	18	11	27	75
Proportion	1.000	0.889	0.803	0.815	0.688	0.795	1.000	0.922	0.900	0.579	0.474	0.974
N = 1139												
N Correct = 931												
Proportion Correct (NM) = 0.835												

Questo fatto viene confermato poi dai test. Infatti, seppure la capacità di dare risposte corrette del nuovo modello sia un po' scesa (83%) rispetto al modello precedente....

## 2° MODELLO (SIMCA)

Classe	Numero di oggetti	Numero di oggetti classificati correttamente dal modello
Acacia	68	38
Agrumi	4	3
Cardo	4	4
Castagno	32	29
Tiglio	16	12

Classificazioni corrette: 69,4%

.....Il test della matrice di confusione indica una capacità di classificazione del 69.4%, non molto dissimile da quello precedente. Quindi, nonostante il secondo modello sia molto più complesso come numero di classi, la sua capacità predittiva promette bene.

## CONCLUSIONI

- La cross-validation mostra che i modelli non hanno particolari criticità di classificazione
- Heracles II e l'elaborazione chemiometrica danno buoni prospettive future
- Validazione dei risultati anche con l'utilizzo di altre tecniche
- Assegnazione dei picchi alle molecole corrispondenti
- Ampliamento del data-set dei mieli e studio sull'origine geografica

In conclusione quindi:

- i test statistici mostrano che i modelli non hanno particolari criticità di classificazione.
- Questo è interessante perché sembrerebbe che questo tipo di approccio analitico possa essere sinergico con tutte le altre analisi consolidate, premesso la possibilità di crearsi un database di riferimento ad hoc, che possa essere alimentato con le nuove annualità;
- Un'altra cosa interessante dal punto di vista scientifico sarebbe andare ad indentificare i composti volatili che il sistema ha indicato come variabili discriminanti.
- Nel futuro inoltre potrebbe esserci la possibilità di migliorare altri tipi di analisi meno convenzionali come quella del DNA utilizzando nuove tecniche analitiche come la digital PCR e la Next Generation Sequencing.

Grazie per l'attenzione