

APIMONDIA 2013
KIEV



IDENTIFYING THE BOTANICAL ORIGIN
OF HONEYS WITH
BOOSTED DECISION TREES

CARI ASBL

Marie WARNIER

Laure-Anne MINSART

Etienne BRUNEAU

Honey origine Identification



- Identification for consumers
- Information for beekeepers on monofloral honey production
- Protected Labelisation





CARI ASBL
4, Place Croix du Sud
B - 1348 Louvain-la-Neuve
BELGIQUE
Tél. : +32 (0)1047 34 19
Fax. : +32 (0)1047 34 94
e-mail : info@cari.be
http://www.cari.be

Rapport d'essai

Louvain-la-Neuve, le 16/07/2013



Associé sous la norme
ISO 17025
Certifié 1719-TEST

Informations transmises

Vos références: 06/201

Lieu de production: Lt

1. EXAMEN PHY

Méthodes adoptées de :
"Normes et méthodes de
Norme Régale - AR relatif

a) Essais accrédités

- Humidité (%)

- pH et acidité

pH
pH au point d'équivalence
Acidité libre (AL) (mg/kg)

- Conductivité (mS)

- Indice de saccharose

- HMF (mg/kg)

* Incertitude évaluée sur un n

● Sucres (% matières fraîches)

Monosaccharides

Fructose **38,85** ± 0,32
Glucose **42,31** ± 2,14
Fructose/Glucose **0,92**

Disaccharides

Maltose + indét. **1,23** ± 1,32
Turanose + indét. **0,48** ± 0,84
Mélbiose et isomaltose **0,05** ± 0,38
Saccharose **0,00** ± 0,10
Tréhalose **0,00** ± 0,10
Gentobiose **0,00**
Palatose **0,00** ± 0,08

Chromatographie en phase gazeuse

Analysé le 04/07/2013

Trisaccharides

Raffinose **0,00** ± 0,12
Eriose **0,00** ± 0,16
Mélézirose **0,00** ± 0,40
Maltotriose **0,00** ± 0,32
Panose **0,00** ± 0,58
Isomaltotriose **0,00** ± 0,09

2. Examen pollinique (non accrédité)

● Analyse pollinique

Densité générale

Pollens dominants

Pollens d'accompagnement
(de 10 à 40 %)

Pollens isolés
(<10%)

Pollens isolés significatifs

Eléments figurés

Moyenne

Humidité : Humidité normale

HMF:

Enzymes : Teneur en saccharase normale

Commentaire :

Acidoglycose selon Erdtman G. 1968. Handbook of
Pollenology.
Munksgaard, Copenhagen, 488 p.

Analysé le 01/07/2013

3. Examen organolept

3.1. Présentation

Examen visuel

Couleur:

miel liquide (Pfund)

miel cristallisé (Pantone)

3.2. Profil odorant et goût

Légende: Contribution à l'inte

"Les notes 'chimiques' ou 'avancées'"

ODEURS

Intensité

moyenne

chaude chimique* exogène

flor./fruit. boisée

fraîche avancée*

ARÔMES

Intensité

moyenne

Chaud **2** ⇒ Doux

Floral/fruité **1** ⇒ Florale

Frais

Chimique* **1** ⇒ Jéjé

Boisé **1** ⇒ Soufflé

Avancé* **1** ⇒ Soufflé

Les résultats ne sont représentatifs

FE LAB 8.1 Révis. II

Seule la mg

Chaud ⇒ doux

Floral/Fruité ⇒ floral - fruité

Frais

Chimique*

Boisé ⇒ végétal sec

Avancé* ⇒ soufflé

"Les notes 'chimiques' ou 'avancées' sont liées à la flore butinée par les abeilles,

mais ne résultent en aucun cas d'une contamination exogène

SAVEURS ET SENSATIONS

Intensité

Sucrée

Acide

Amère

Astringente

Froid

Piquant

ARÔMES, SENSATIONS EXOGÈNES

⇒ Néant

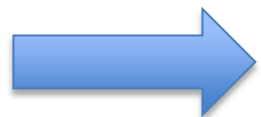
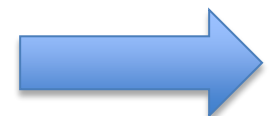
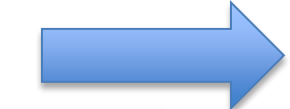
PERSISTANCE

physicochemical parameters

Sugars

Pollens

Organoleptic Properties



Intensité des odeurs **moyenne**

Intensité des arômes **moyenne**

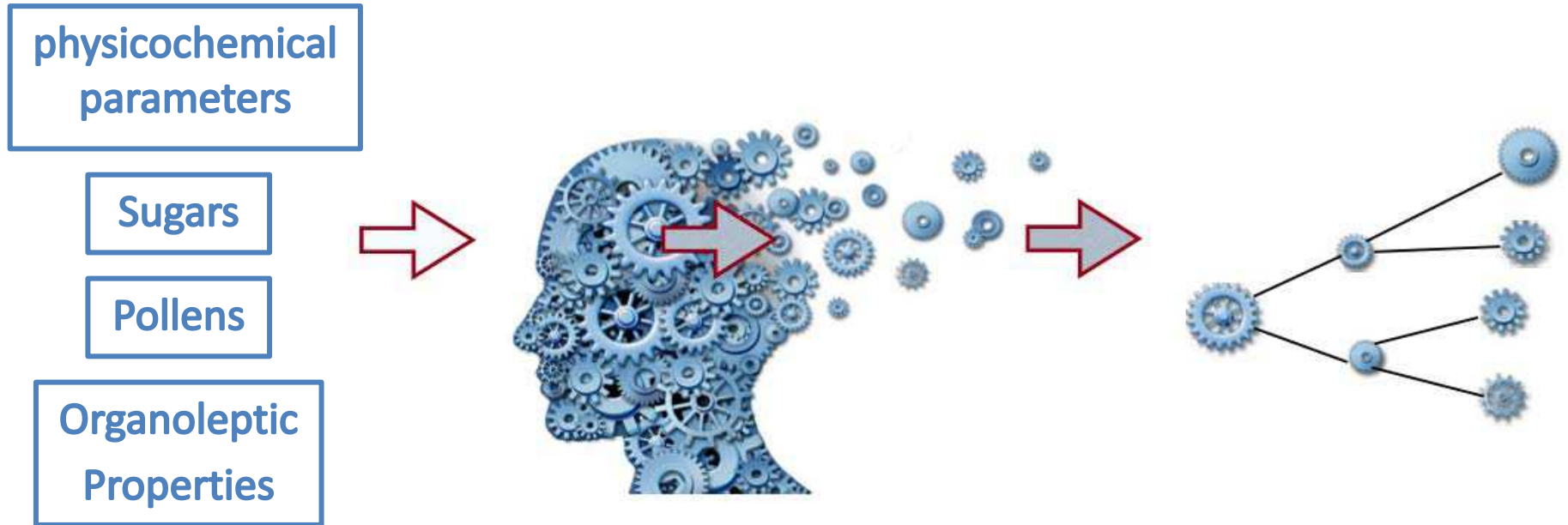
Saveurs et arômes : **Miel doux aux notes végétales, florales et légèrement fruitées**

Intensité des saveurs et sensations **discrète**

L'échantillon analysé répond aux normes légales et aux critères de qualité conseillés

L'échantillon analysé répond aux spécifications

Typical human analysis

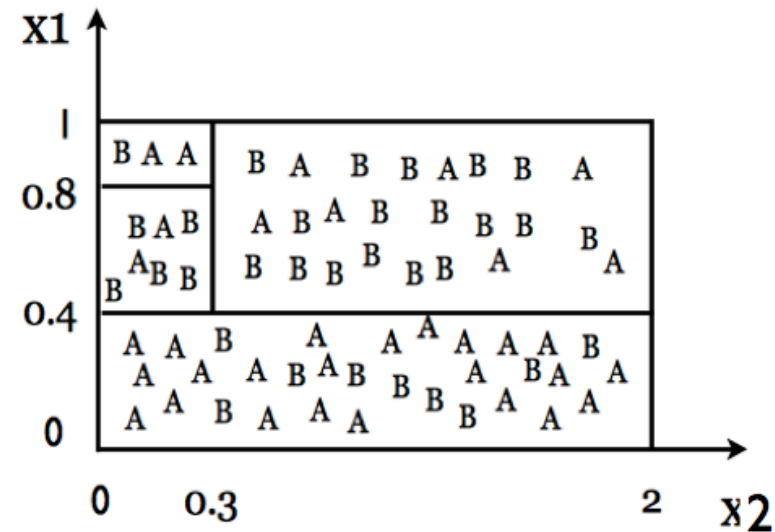
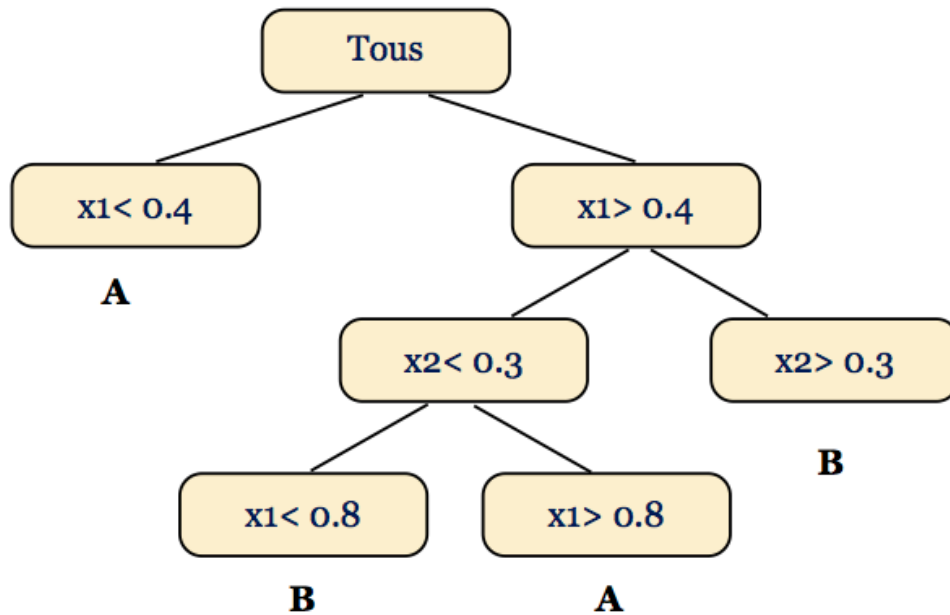


GOAL OF BOOSTED DECISION TREES :

STATISTICAL DECISION SUPPORT

Decision Trees

Example of a 2-class Decision Tree



Separation of the space using the decision tree
 Elements are classified in boxes depending on the variables and the nodes chosen

Decision Trees



Variable linked to each node :

P-factor
$$P = \frac{\text{nb}(A)}{\text{nb}(A) + \text{nb}(B)}$$
 with nb : numbers of individuals

Gini
$$\text{Gini} = P(1 - P)$$

Gini factor measures the lack of purity on each subgroup (Gini = 0 for a perfect classification)

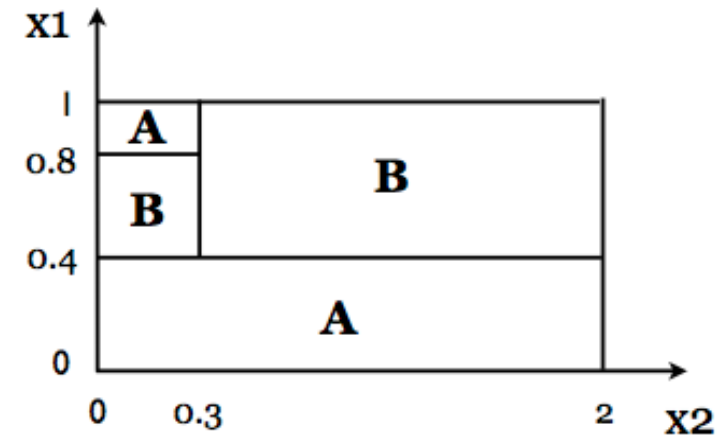
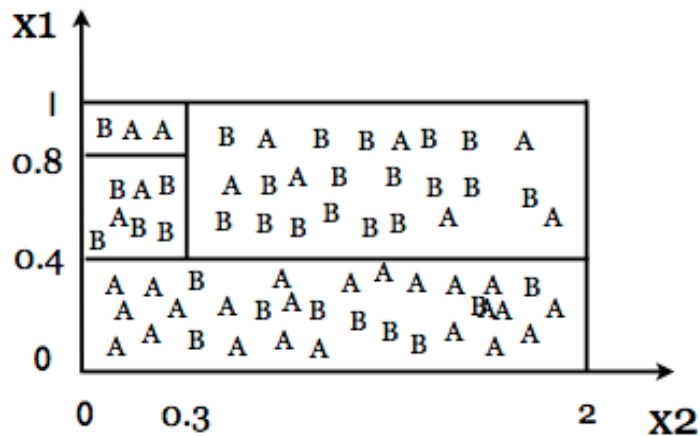
On each node → Gini (left) + Gini (Right) is calculated

Program choose the variables and the level which minimizes sum of Gini

Decision Trees



Tree n°1 :



Each partition of the space gets the name of the most represented population

Decision Trees



Function and errors associated to each partition:

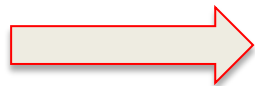
Associated error :
$$error = \frac{nb \text{ misclassified individuals}}{nb \text{ total individuals}}$$

Quality index α of tree n°1 is linked to the associated error : $\alpha \nearrow$ if error \searrow

Before building a 2nd tree, weights are associated to individuals :

$w_i = C$ if individual is well classified

$w_i = Ce^{\alpha}$ if individual is misclassified

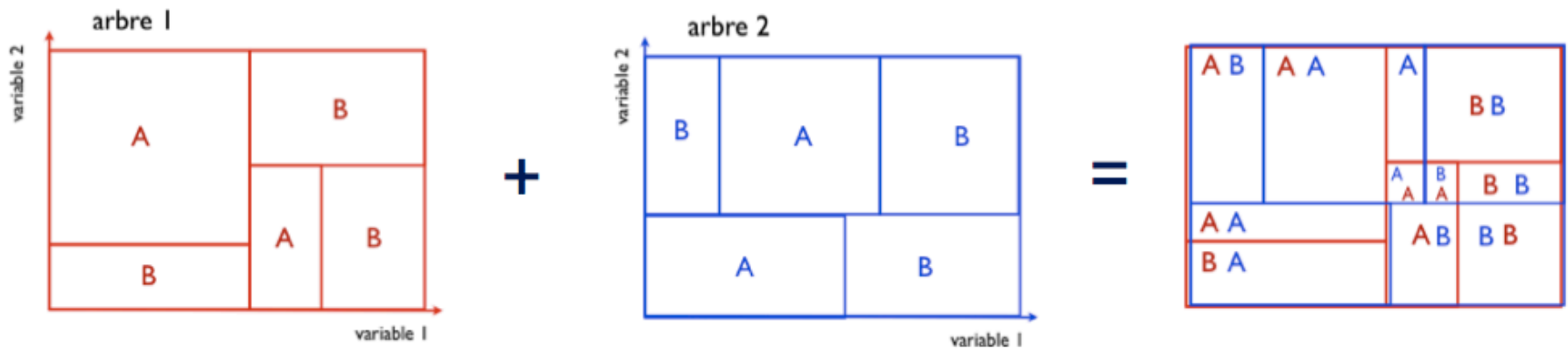


Misclassified individuals get a higher weight

Boosted Decision Trees



Boosting is an iteration of several trees



All the decision trees obtained are summed for a better classification and for a better stability of the answer

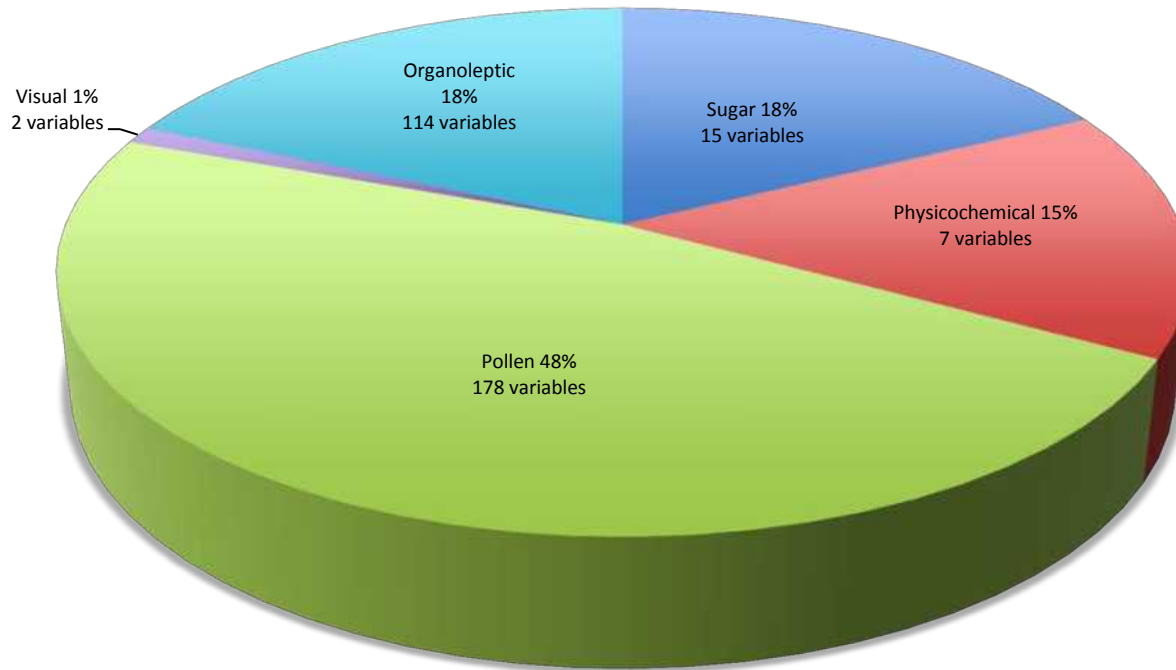
Application to honey



- Database : analysis results of honey samples collected from 2009 to 2012 (4 years)
- Database was « cleaned » :
 - Monofloral honeys
 - Min. 3 individuals / class
 - Qualitative variables transformed in quantitative variables (1 – 0)
- Number of iterations : 50

Number of Individuals = 406
Number of botanical origins = 26
Number of variables = 316

Variable importance





condu	9.68033932008605
po_Ronces	4.91564305467384
s_glu	4.30028352910444
s_fru	4.18556897405754
po_Trèfle	4.02418925384346
po_Tilleul	3.8534426975245
po_Romarin	3.68676181992028
po_Saule	3.12293576746948
ph	2.98954958970574
po_Fruitiers	2.87774381211671
po_Baies.roses	2.43555950049976
po_Tournesol	2.18436792407647
ga_ref_flor_Cassis	2.10244705042988
s_eri	2.08217203762341
s_sac	1.96346539740271
po_Brassicacées	1.88888152398904
ga_ref_flor_Litchis	1.87283503352997
po_Elaeis	1.69952432311766
po_Sarrasin	1.6336184774648
po_Eucalyptus	1.24243421871157
ga_ref_chaud_Pâte.d.amandes	1.24018738241987
po_Bourdaine	1.21565769621518
s_mele	1.19213536732895
aci_l	1.17384436354893
po_Myrthacées	1.12617408011203
po_Lamiacées	1.09330397512452
po_Cistacées	1.04310433947906
go_intensite	0.970268126059644
po_Châtaignier	0.953367348163582
s_mal	0.923780317790425
is	0.8939099348904
ga_ref_flor_Fleur.d.oranger	0.886836168682136
po_Citrus	0.848990716224256
po_Bouleau	0.828962323890527
s_meli	0.793827616264473
po_Lavande	0.747385155038932
ga_sous_bois_Néant	0.742714186327937
ga_intensite_boises	0.696402039856156
densite	0.695741773315093
pfund	0.582441275256517
ga_intensite_chim	0.57772321718869
po_Centaurées	0.553367812566357
po_Plantain	0.548322329075191
po_Vipérine	0.547620093484327
ga_sous_chaud_Néant	0.53400962969548
go_chaude	0.521728718496157
ph_pe	0.476416485554065
pantone	0.468920969954908
gs_intensite_amere	0.465014315981447
po_Pissenlit	0.450901778232765

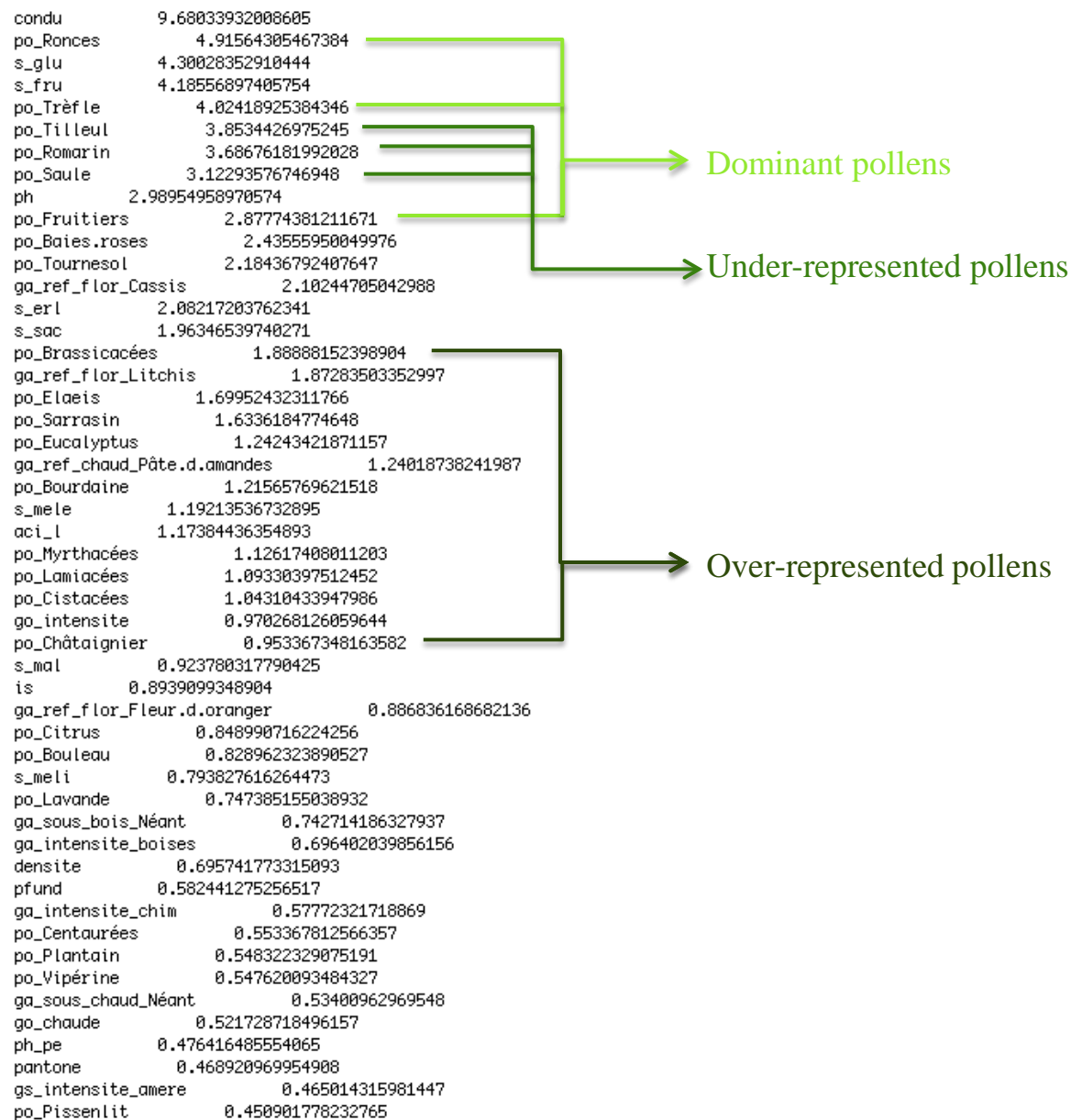
Conductivity & pH : most discriminating physico-chemical parameters

Glucose, fructose, erlose & saccharose : Most discriminating sugars

Typical tastes

Color parameters

⋮



Outcomes for 2013 honeys



- Prediction realised on 75 honeys from 2013 with the Boosted Decision Tree (2009-2012)
- 13 monofloral honeys
 - 12 well classified (rapeseed, rosemary, palm)
 - 1 misclassified (jujube, class not yet in the database)
 - 100% good identification
- 62 polyfloral honeys

Outcomes for 2013 honeys



1) Monofloral honey

```
individu 14684 , class Colza , bien classe = TRUE
```

```
-----  
type : Colza , proba: 0.614675107965911  
type : Tournesol , proba: 0.12143056361069  
type : Agrumes , proba: 0.0629587490456149  
type : Acacia , proba: 0.0591173906844131
```

Laboratory Analysis	Boosted Decision Tree Analysis
COLZA (RAPESEED) Reference honey	COLZA (RAPESEED)

Application to honey



2) Polyfloral honey

```
individu 14672 , class Fruitiers dominant , bien classe = FALSE
```

```
-----  
type : Fruitiers , proba: 0.234889248178218  
type : Pissenlit , proba: 0.1977279975379  
type : Fruitiers & Saule , proba: 0.122222570157609  
type : Bourdaine , proba: 0.0824073914285493  
type : Colza , proba: 0.0818284402434684  
type : Acacia , proba: 0.080111747020175  
type : Saule , proba: 0.0792822741775859
```

Laboratory Analysis	Boosted Decision Tree Analysis
FRUITIERS DOMINANT (FRUIT PREVAILING) & SAULE (WILLOW)	FRUITIERS (FRUIT) & PISSENLIT (DANDELION) & FRUITIERS & SAULE (FRUIT & WILLOW)

High conductivity
Dandelion pollens

Application to honey



individu 14653 , class Toutes fleurs , bien classe = FALSE

type : Colza , proba: 0.155861185724068
type : Agrumes , proba: 0.146318352854727
type : Acacia , proba: 0.138102209903404
type : Lavande , proba: 0.103182696553883
type : Trèfles , proba: 0.0814418172783824
type : Thym , proba: 0.0600130038971022
type : Sarrasin , proba: 0.0421238033479969
type : Pissenlit , proba: 0.0406941713287997
type : Fruitiers , proba: 0.0393152333730134

Laboratory Analysis	Boosted Decision Tree Analysis
TOUTES FLEURS (FLOWER HONEY) Romarin (Rosemary) & Type Rosacées (Rosaceae) & Fruitiers (Fruit)	COLZA (RAPESEED) & AGRUMES (CITRUS) & ROBINIER (ROBINIA) & LAVANDE (LAVENDER)

No confidence on
BDT Analysis

Conclusions



- Powerful tool
- Should be used in addition to human analysis



Boosted Decision Tree should be carried out after typical analysis → no influence

Pay attention to the results proposed

- Perspectives in the future :
 - Enrichment of the database needed
 - Model should be adapted to polyfloral honeys, prevailing types and honeydew