Givaudan

ENGAGING
THE
SENSES

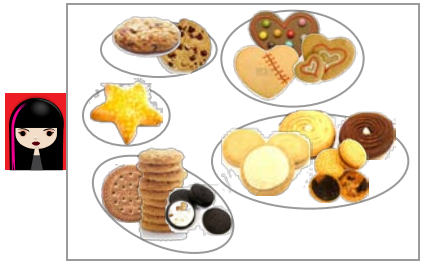
A method to investigate the stability of a Sorting map

Guillaume Blancher, Benoît Clavier, Carine Egoroff, Jason Parcon

July 2010 – Sensometrics 2010, Rotterdam

Background: what is sorting?

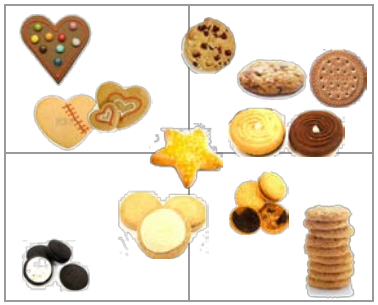
Stimuli & panelists → **Sensory task - categorization** → **Statistical tools** → **Sensory distances, sensory maps, …**

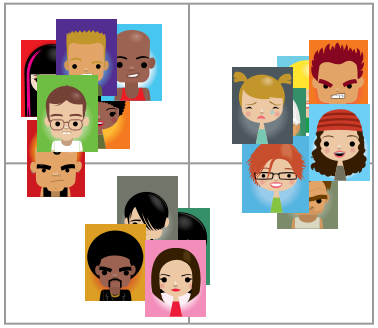


…

…

INDSCAL, DISTATIS, (…)





Renewed & growing interest for Sorting

- Sensory & consumer world [1]
- Statistical world: MDS, DISTATIS [2], FAST [3], SORT-CC [4]

Givaudan

Checking the quality of results in sensory analysis

	Descriptive analysis	Sorting
Panelist level	▪ Repeatability ▪ Discrimination power ▪ Aligment	▪ Rand index & Hierarchical clustering [1,8]
Panel level	▪ Repeatability ▪ Discrimination ▪ Homogeneity	▪ Sensory-based approaches • Repeated products, Repeated sessions, Comparison of different types of panelists, Comparison of Sorting with other methods ▪ Statistical-based approaches • Bootstrap - Confidence ellipses on the sensory maps ([5], [6], [7]) • Bootstrap - RV coefficient: pionneered by Faye et al. [8] • Global analysis of the similarities/dissimiliarities between the panelists ([2], [4])

Givaudan

Research questions





- Need for a simple statistical approach to assess the quality of Sorting results
- Need to change the focus away from the graphical representation of the products, and to focus on the differences between panelists

- **Can we develop a simple quantified indicator of the stability of Sorting results?**

- **Can we try and understand why in some cases we get stable results and in others we don't?**

Givaudan

Data sets

MiniVAS

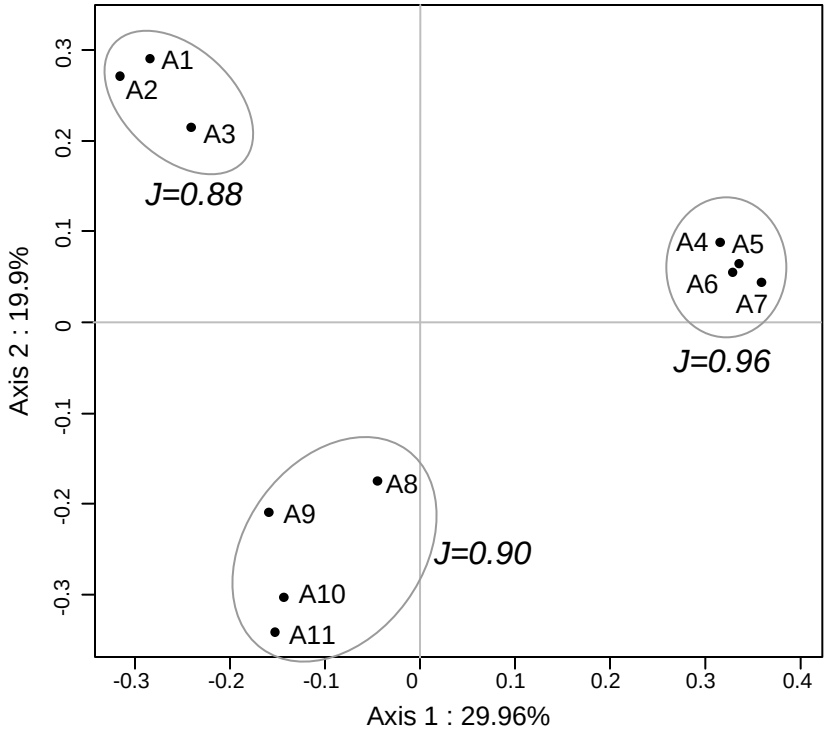
Data set	Stimuli	Type of panelists	Number of evaluations
DS1	11 chocolate aromas 	▪Half panelists were experienced in QFP™ [9], half panelists were internal employees ▪Not familiar with the stimuli	37
DS2	8 beers 	▪ *"beer consumers but did not have any formal training in sensory evaluation of beers"* [2]	10
DS3	12 market yogurts 	▪ Panel experienced in QFP™ ▪ But not familiar with the stimuli	25
DS4	14 vanilla aromas 	▪ Panel experienced in QFP™ ▪ Familiar with the stimuli	3*12 = 36
DS5	14 vanilla aromas 	▪ Internal employees ▪ Not all specifically familiar with the stimuli	2*59 = 118

▪Same stimuli
▪Different panelists

Evaluation = one sorting task by one panelist
MiniVAS = device to release aroma to the panelists with a controlled intensity

Givaudan

Data set 1



Stimuli put alone more than 80% of the times	Number of stimuli clusters and cluster stability	PC1+PC2 (%)
-	3 clusters, all very stable	49.9

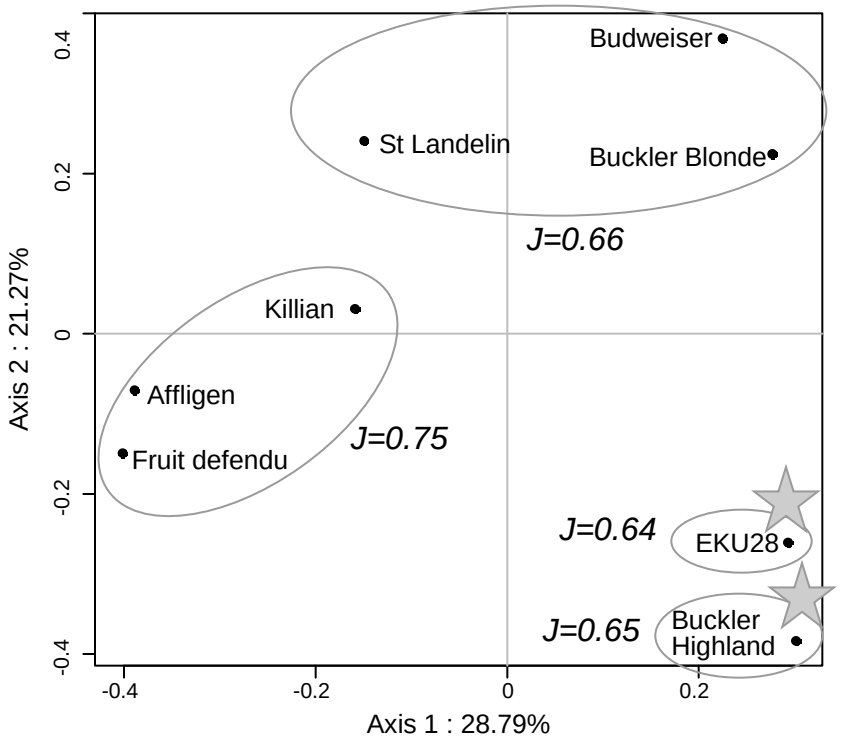
➢ **Overall, quite straightforward sensory space**

Cluster stability

- Bootstrapped Jaccard coefficient (J) and Hierarchical Cluster Analysis with Ward's criterion [10].
- A cluster was judged stable if J ≥ 0.75.

Givaudan

Data set 2



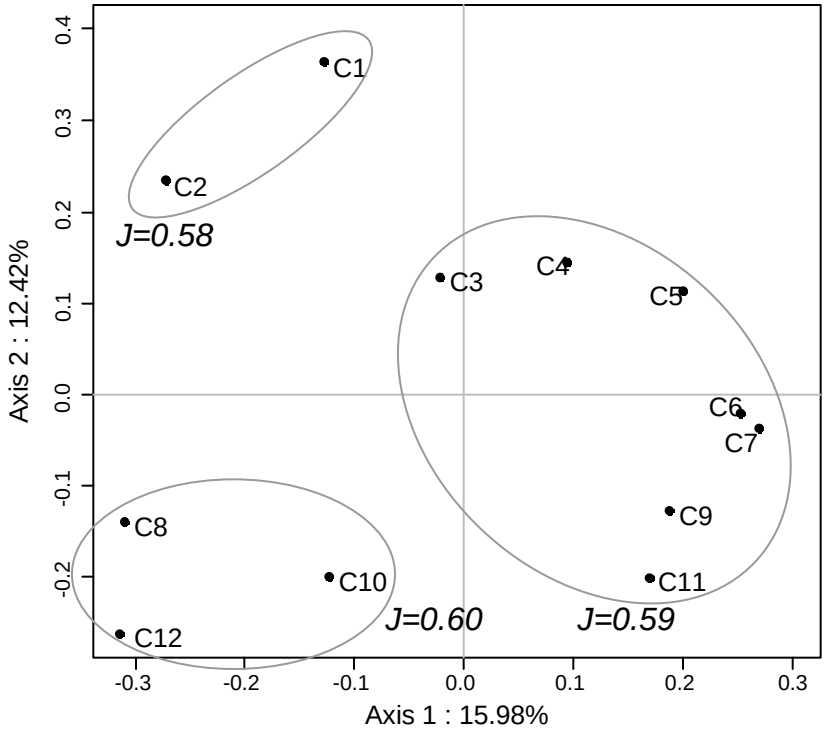
Product put alone 80% of the times or more

Stimuli put alone more than 80% of the times	Number of stimuli clusters and cluster stability	PC1+PC2 (%)
EKU28 and Buckler Highland	4 clusters, not all stable	50.1

- **More complicated product space than DS1**

Givaudan

Data set 3



Stimuli put alone more than 80% of the times	Number of stimuli clusters and cluster stability	PC1+PC2 (%)
-	3 clusters, not of them stable	28.4

- **Overall, a complicated product space compared to both DS1 and DS2**

Givaudan

Data sets 4 & 5
Composition of the stimuli

Flavor	Anise	Spicy	Creamy	Vanillin	Phenol	Caramel	Vanilla	Main aroma characters
D1	o	o	L	L	**H**	o	L	Low flavor intensity
D2	o	o	L	**H**	L	o	L	Low flavor intensity
D3	**H**	o	L	L	L	o	L	Anise at a high level
D4	**H**	o	**H**	**H**	L	o	L	
D5	o	o	**H**	L	L	**H**	**H**	Vanilla and caramel key notes at a high level
D6	**H**	o	**H**	L	**H**	**H**	**H**	
D7	**H**	o	**H**	**H**	L	**H**	**H**	
D7.bis	**H**	o	**H**	**H**	L	**H**	**H**	
D8	M	M	M	M	M	M	M	All aroma key notes present at a medium level
D8.bis	M	M	M	M	M	M	M	
D9	**H**	**H**	**H**	**H**	**H**	o	L	Spicy key note at a high level
D10	o	**H**	L	**H**	**H**	o	L	
D11	**H**	**H**	**H**	**H**	L	**H**	L	
D12	o	**H**	**H**	L	**H**	**H**	**H**	
D13	o	**H**	L	**H**	**H**	**H**	**H**	
D14	**H**	**H**	L	**H**	**H**	**H**	L	
D14.bis	**H**	**H**	L	**H**	**H**	**H**	L	

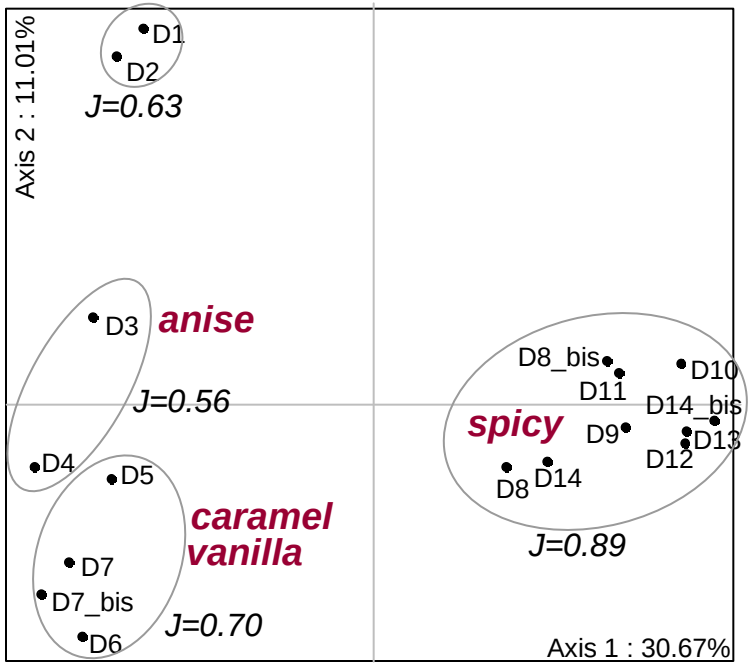
repeated

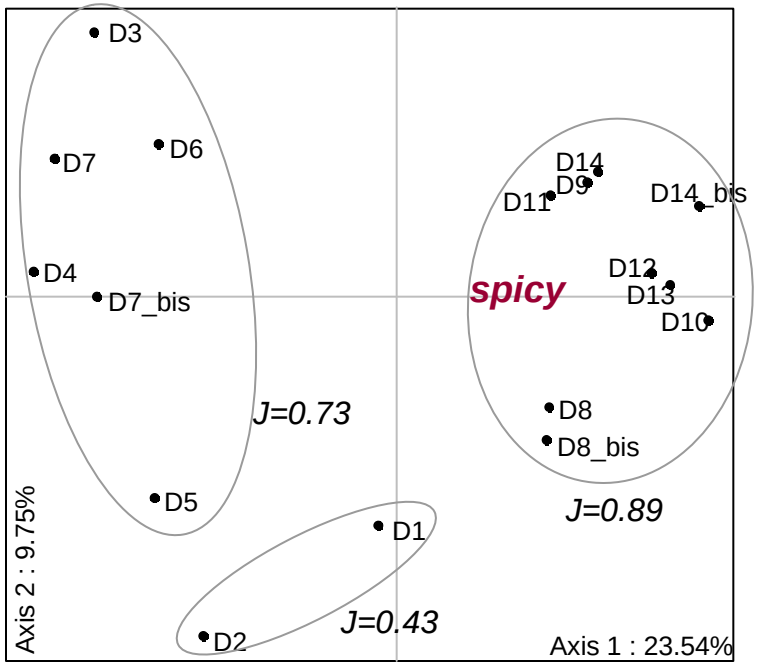
repeated

repeated

Level	Meaning
o	Not present
L	Low
M	Medium
H	High

Givaudan



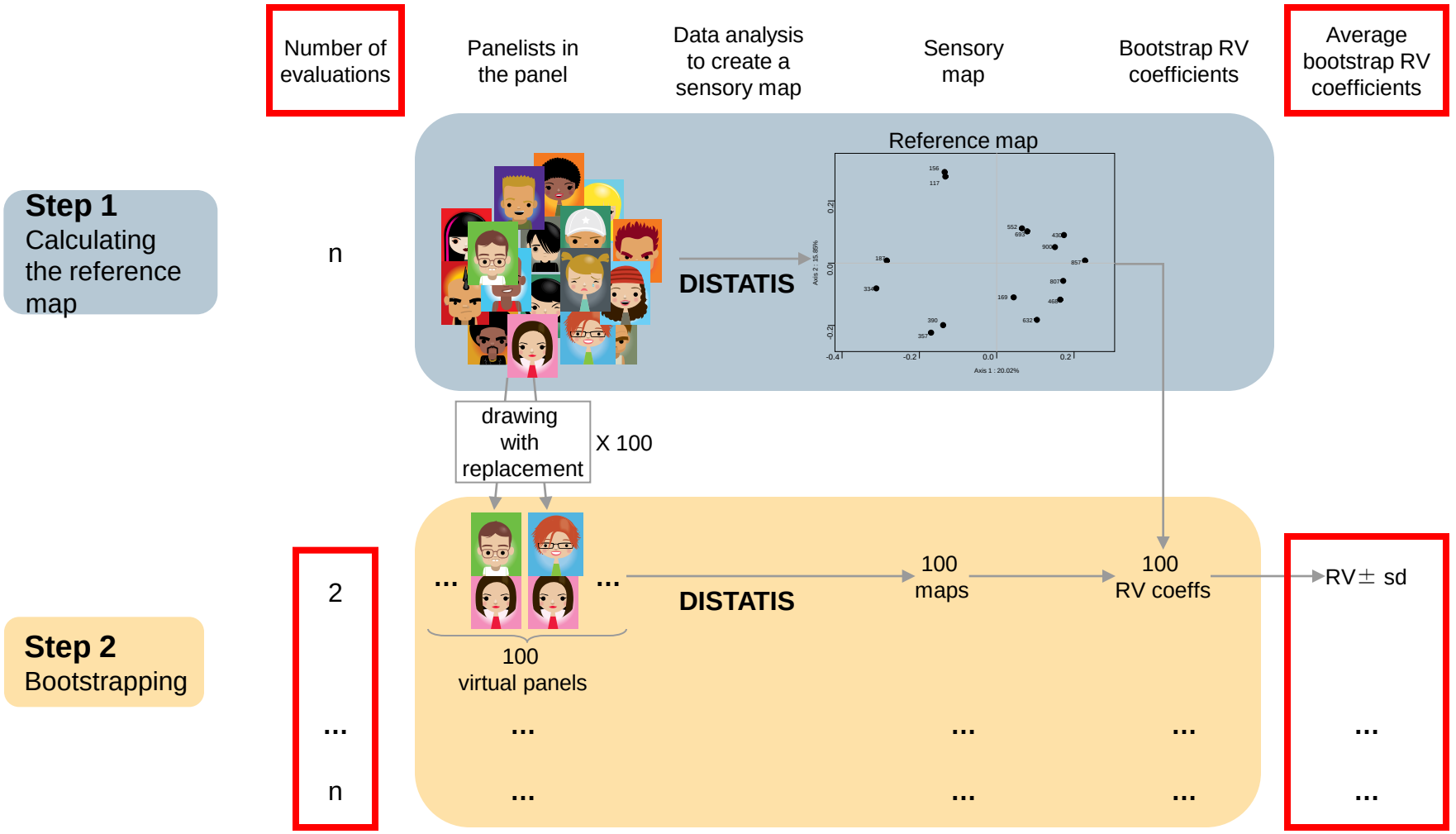


Data sets 4 & 5

	Stimuli put alone more than 80% of the times	Number of stimuli clusters and cluster stability	PC1+PC2 (%)
DS4	-	4 clusters, one of them very stable	41.7
DS5	-	3 clusters, one of them very stable	33.3

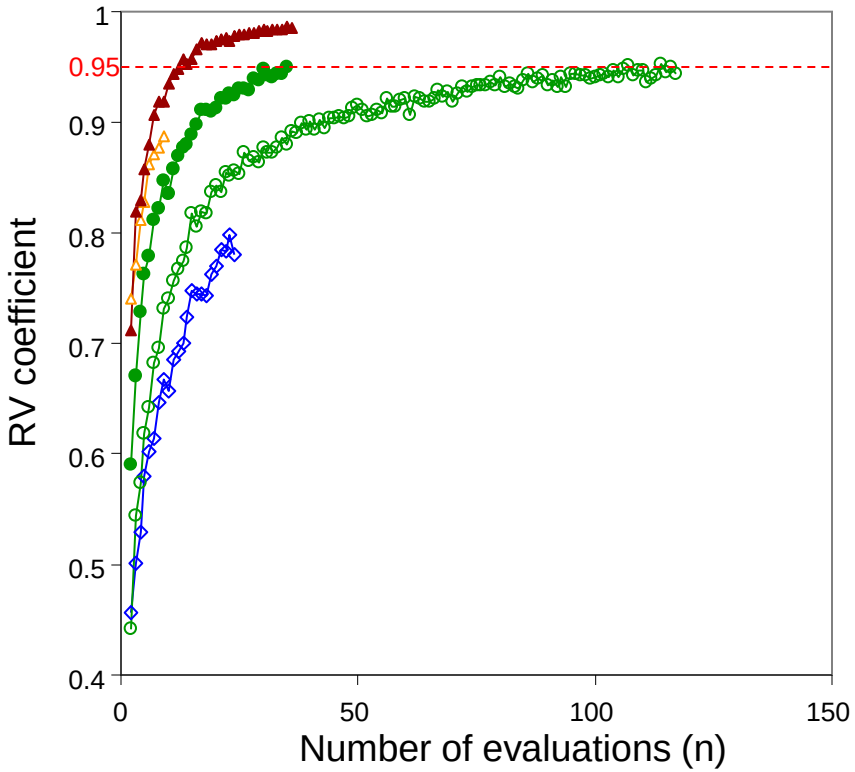
- **The level of complexity of DS4 and DS5 seems to be intermediate between DS1 and DS3.**
- **The sensory task at hand was more difficult for DS5 than for DS4.**

Bootstrapping approach



Givaudan

Evolution of the stability of the Sorting maps as a function of number of evaluations



▲ **DS1** ✓ **Stable**

△ **DS2** ***Not completely stable***

◊ **DS3** ***Unstable***

● **DS4**

○ **DS5**

✓ **Stable**

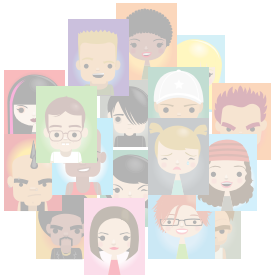
➢ **The RV coefficient reached with all available evaluations is a good indicator of the stability of the Sorting map.**

Givaudan

Research questions







- Need for a simple statistical approach to assess the quality of Sorting results
- Need to change the focus away from the graphical representation of the products, and to focus on the differences between panelists

✓ **Can we develop a simple quantified indicator of the stability of Sorting results?**

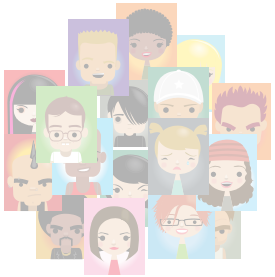
- **Can we try and understand why in some cases we get stable results and in others we don't?**

Givaudan

Research questions







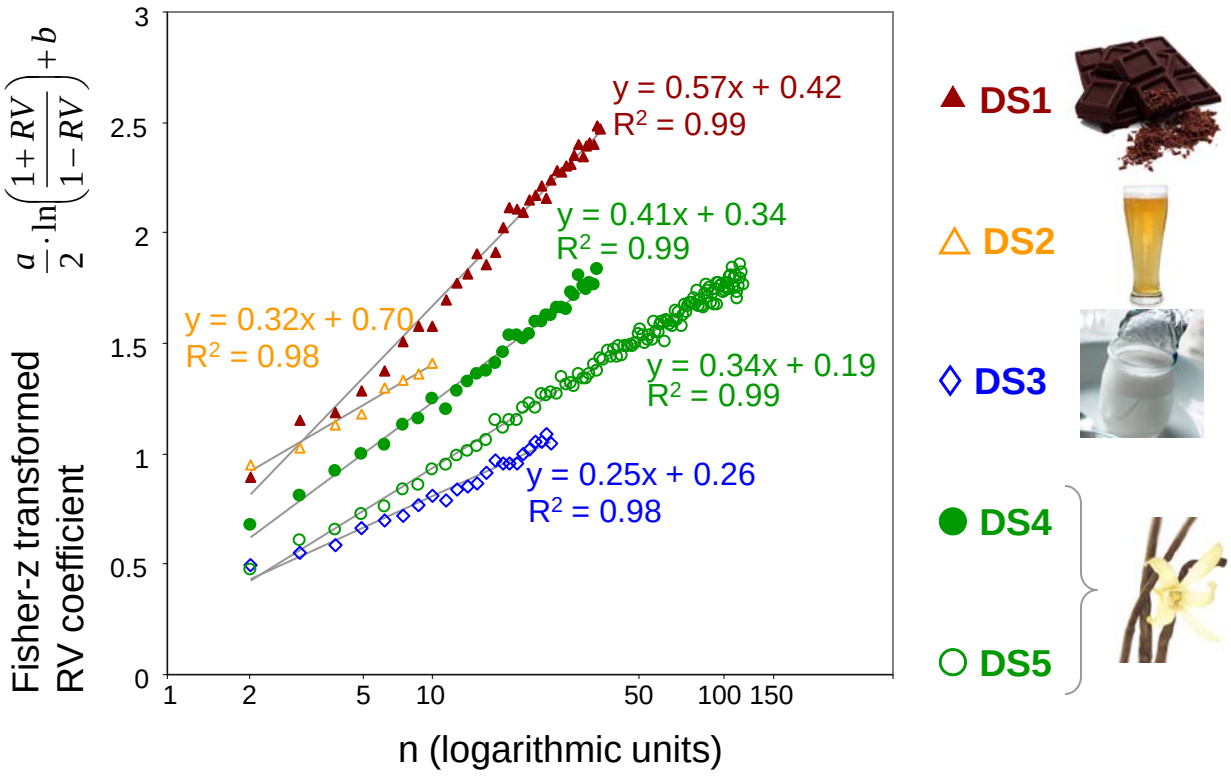
- Need for a simple statistical approach to assess the quality of Sorting results
- Need to change the focus away from the graphical representation of the products, and to focus on the differences between panelists

✓ **Can we develop a simple quantified indicator of the stability of Sorting results?**

- **Can we try and understand why in some cases we get stable results and in others we don't?**

Givaudan

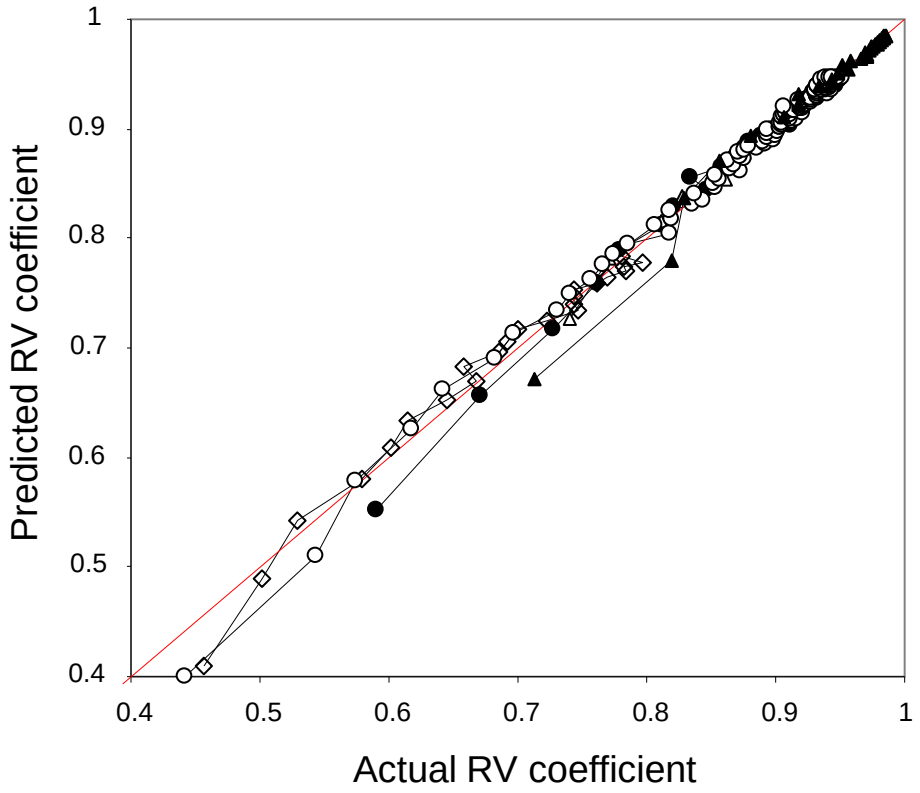
Modeling the stability of a Sorting map as a function of the number of evaluations



- **Slope "a"** = general level of agreement between the panelists
- **Intercept "b"** = average level of agreement of the panelists with the consensus

Givaudan

Checking the quality of the model



- Model:

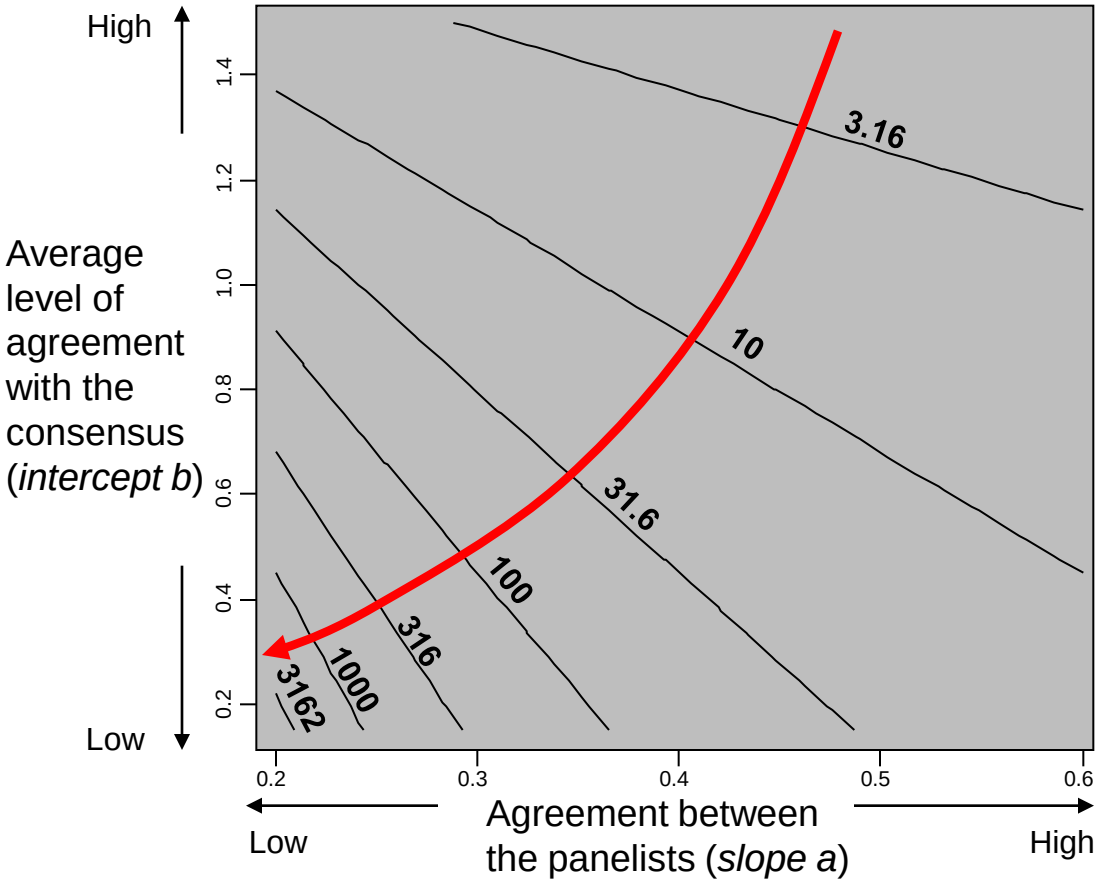
$$RV = \frac{(e^{b} \cdot Evaluations^{a})^{2} - 1}{(e^{b} \cdot Evaluations^{a})^{2} + 1}$$

- Slight underestimation of the RV coefficient at low values

➢ **Overall, very satisfactory model**

Givaudan

Predicted number of evaluations necessary to reach a stable map



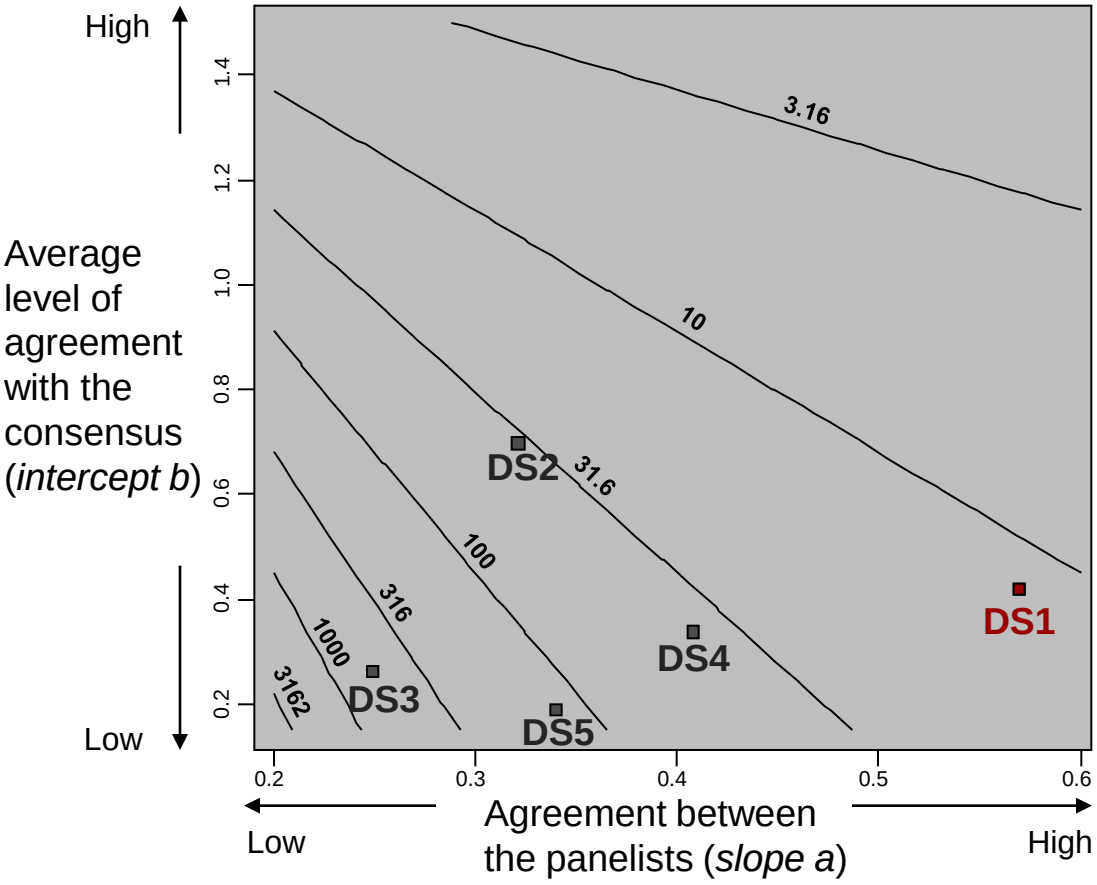
- Each data set is characterized by:
  - a = agreement
  - b = discrimination
- Predicted number of evaluations necessary to reach a stable map (RV = 0.95):

$$Evaluations = \left(\frac{1 + RV}{(1 - RV) \cdot e^{2b}} \right)^{\frac{1}{2a}}$$

- **The number of evaluations necessary to yield a fully stable map vary depending on the characteristics of the panelists and of the stimuli.**
- **Starting off with 30 evaluations seems a reasonable number.**

Givaudan

Predicted number of evaluations necessary to reach a stable map

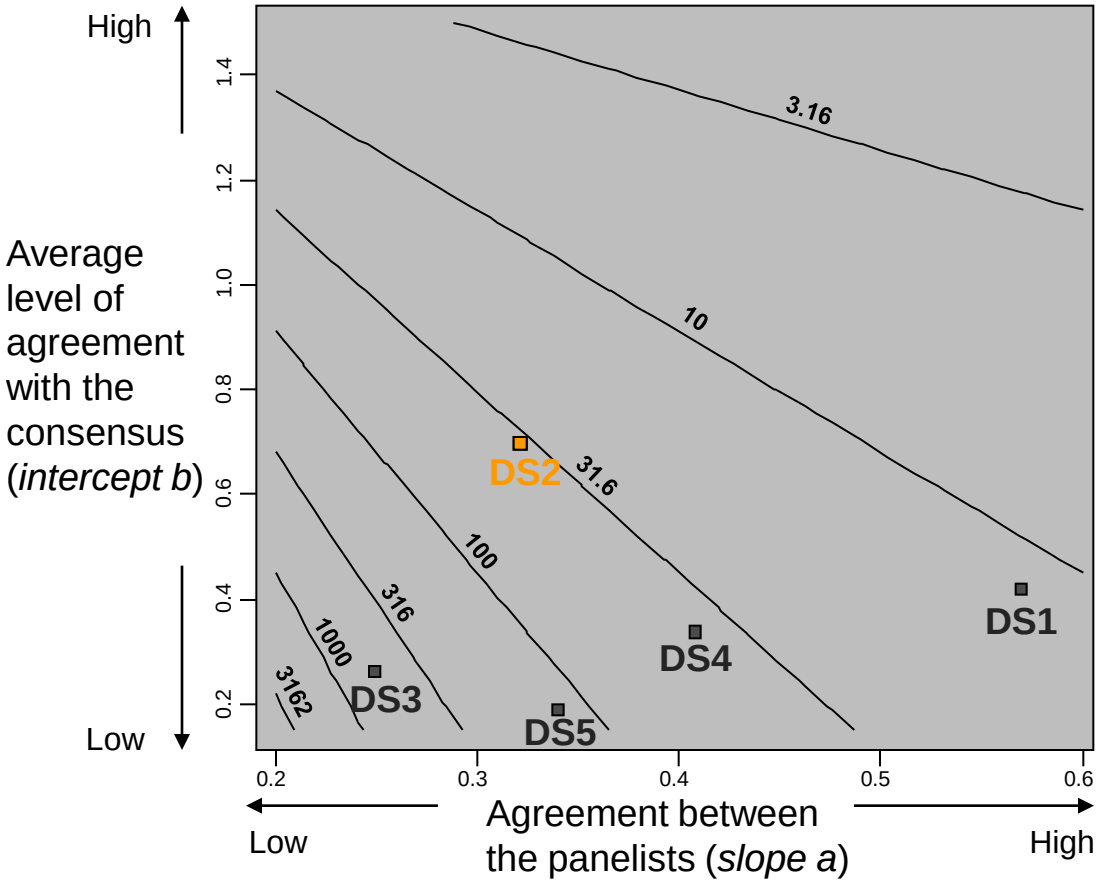


DS1

- 37 evaluations were available
- According to the model, a stable map could have been reached after only 12 evaluations
- There was a very high level of agreement between the panelists, despite the fact that they were not all specifically trained in sensory analysis and were not familiar with the stimuli.
- The high agreement is probably due to the fact that there were 3 straightforward clusters of stimuli.
- Besides the 3 clusters, not really possible to further discriminate between the stimuli.

Givaudan

Predicted number of evaluations necessary to reach a stable map

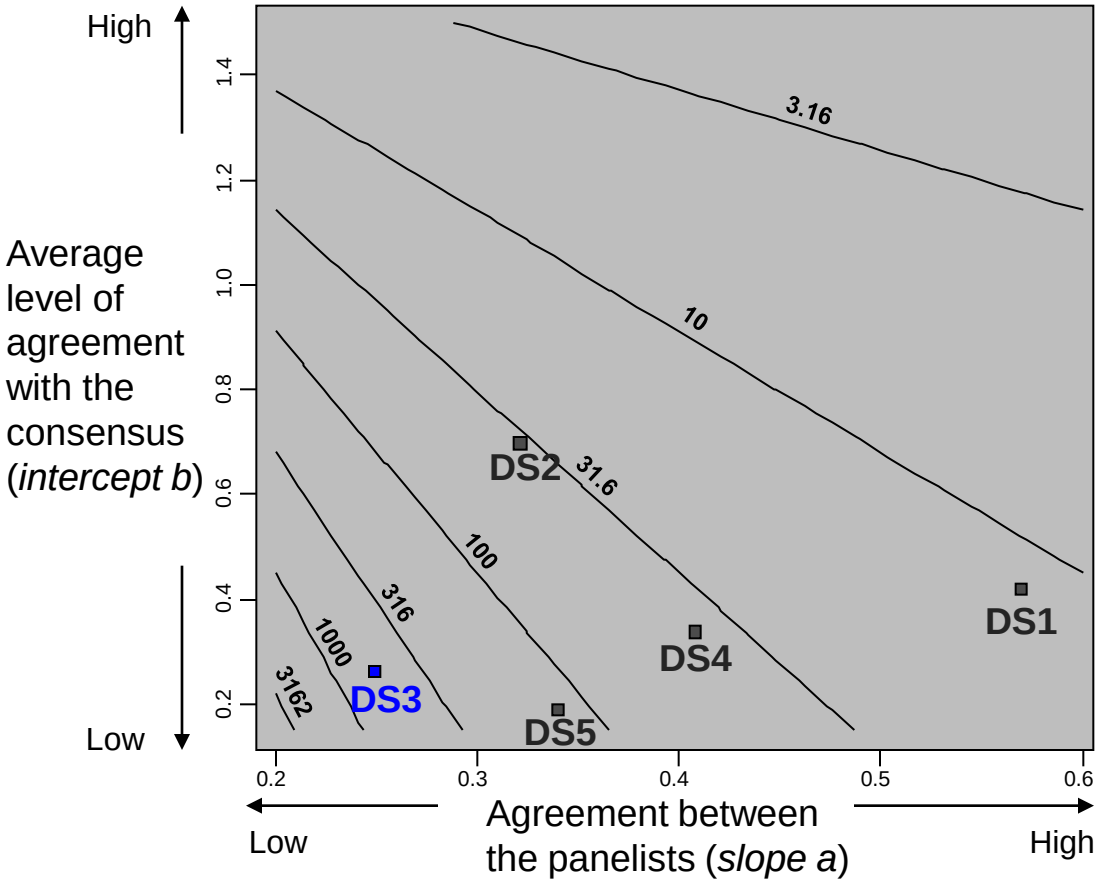


DS2

- Only 10 evaluations were available.
- A fully stable map could have been reached after about 34 evaluations.
- There was a medium level of agreement between the panelists, probably due to the absence of training of the panelists.
- The relatively high average agreement with the consensus is probably due to the fact that 2 products were rarely or never put with the other ones

Givaudan

Predicted number of evaluations necessary to reach a stable map



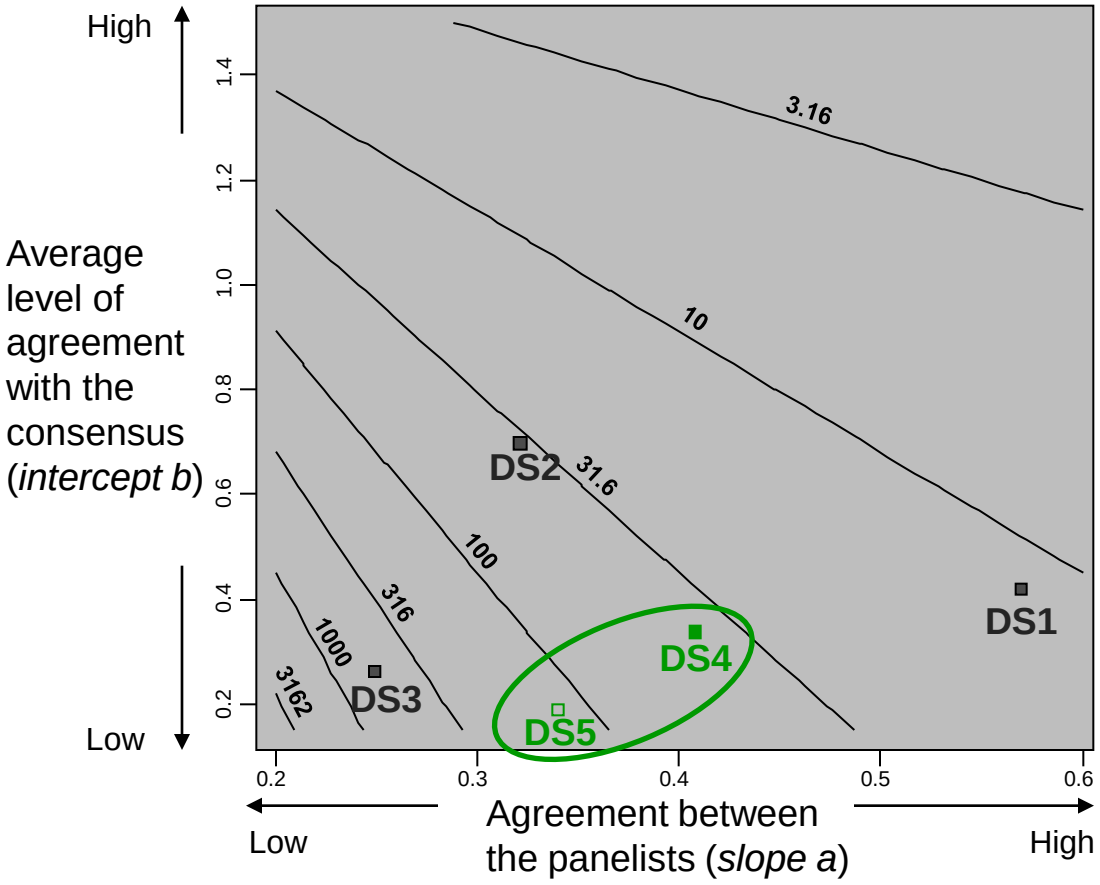
DS3

- "Only" 25 evaluations were available.
- A fully stable map could have been reached only after about 550 evaluations… *(not reasonable)*
- Low level of agreement between the panelists and with the consensus, probably due to a complex product set (multiple categorization criteria).

➢ *Further train the panel, or ask the panelists to focus only on one given attribute at a time? Sorting task not suited?*

Givaudan

Predicted number of evaluations necessary to reach a stable map

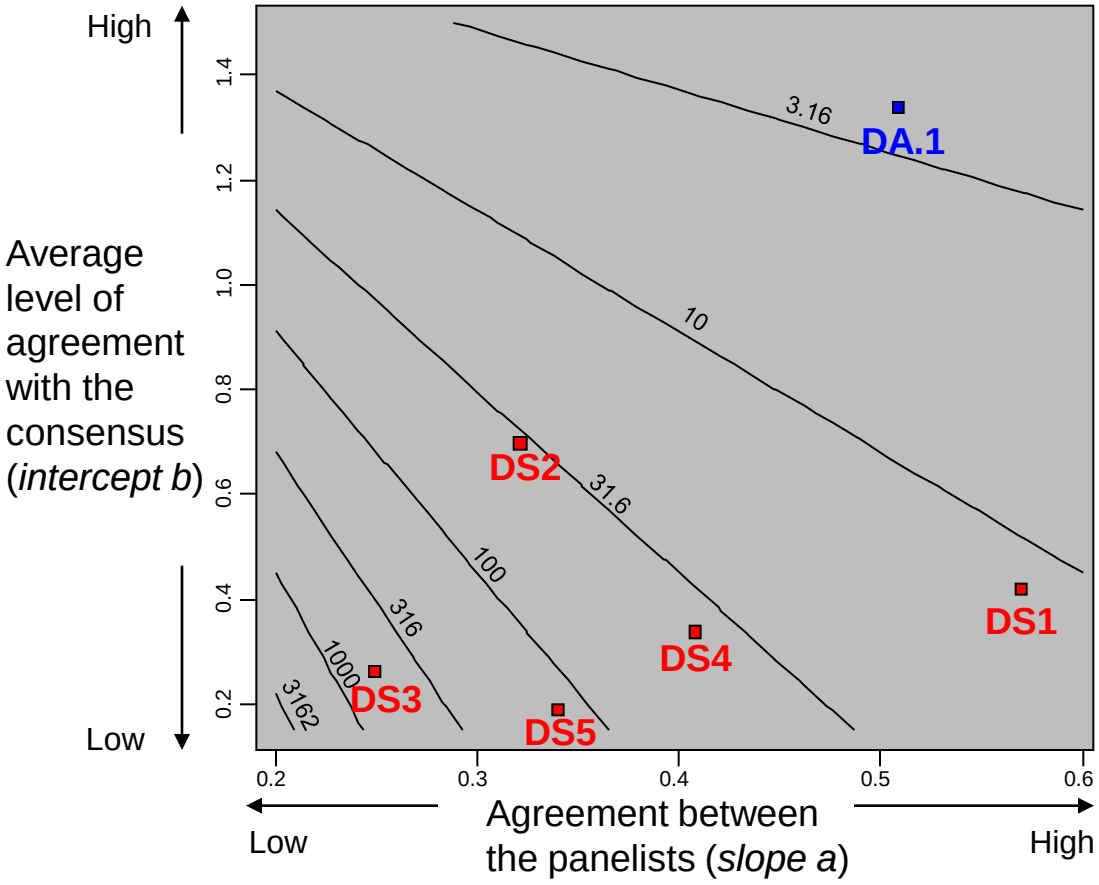


DS4 and DS5

- Same stimuli, different panelists.
- DS4:
  - panelists highly trained to QFP™ and familiar with the stimuli; used Sense It™, a common descriptive language
  - a stable map (RV=0.95) was reached after 36 evaluations
- DS5:
  - Internal employees, not trained to Sensory analysis and not all familiar with the stimuli; no common descriptive language
  - an almost stable map (RV=0.94) was reached after 118 evaluations

Givaudan

This approach can also be applied to Descriptive Analysis



- **DA.1** [11] corresponds to a QDA® performed on the texture of jellies by a highly trained:
  - 28 evaluations were available (14 subjects, 2 reps)
  - A stable map (RV=0.95) could have been reached after only about 3 evaluations.
  - High level of agreement between the panelists due to an extensive training,
  - High ability to discriminate between the products due to the training and to the sensory method.

Givaudan

Summary

	Objectives	Conclusions
Objective 1	▪To develop a simple quantified indicator of the stability of sorting results.	▪The RV coefficient reached with all available evaluations is a good indicator of the stability of a sorting map.
Objective 2	▪ Try and understand why in some cases we get stable results, and in others we don't.	▪Each sorting test is unique, no definite number of evaluations can be given a priori to reach a stable map. ▪The stability of a sorting map depends on: • the general level of agreement between the panelists • the average level of agreement of the panelists with the consensus

- This rather universal approach could be extended to:
  - Other types of sensory tests (Flash Profiling [12], QDA®, projective mapping [13,14], etc.)
  - Other statistical methods which outputs are sensory maps (PCA, GPA, MDS, etc.)
- The two indicators of panel "performance" that were developed are valid at the panel level. What about the panelist level?

Givaudan

Acknowledgements

- Christel Adam for supporting this research
- Alexis Luco for running part of the sensory studies
- Kees Duineveld for very helpful insights

Givaudan

Bibliography

- [1] Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality and Preference, 15*(7-8), 781-791.
- [2] Abdi, H., Valentin, D., Chollet, S., & Chrea, C. (2007). Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality and Preference, 18*(4), 627-640.
- [3] Cadoret, M., Lê, S., & Pagès, J. (2009). A Factorial Approach for Sorting Task data (FAST). *Food Quality and Preference, 20*(6), 410-417.
- [4] Qannari, E. M., Cariou, V., Teillet, E., & Schlich, P. (2010). SORT-CC: A procedure for the statistical treatment of free sorting data. *Food Quality and Preference, 21*(3), 302-308.
- [5] Blancher, G., Mattei, B., Oelhafen, N., & Adam, C. (2008). A comparison of free sorting and hierarchical sorting tasks, A case study with sauces. In, *Agro-Industrie et Méthodes Statistiques - 10èmes Journées Européennes, January 2008*. Louvain-La-Neuve, Belgium.
- [6] Abdi, H., Dunlop, J. P., & Williams, L. J. (2009). How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage, 45*(1), 89-95.
- [7] Santosa, M., Abdi, H., & Guinard, J.-X. (2010). A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality and Preference, In Press, Accepted Manuscript*.
- [8] Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A., & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference, 17*(7-8), 604-614.
- [9] Stampanoni, C. R. (1993a). Quantitative flavour profiling: an effective tool in flavour perception. *Food and Marketing Technology*, 4-8.
- [10] Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis, 52*, 258-271.
- [11] Blancher, G., Lê, S., Sieffermann, J.-M., & Chollet, S. (2008). Comparison of visual appearance and texture profiles of jellies in France and Vietnam and validation of attribute transfer between the two countries. *Food Quality and Preference, 19*(2), 185-196.
- [12] Sieffermann, J.-M. (2000). Le profil flash - un outil rapide et innovant de valuation sensorielle descriptive. In, *AGORAL 2000, XIIèmes rencontres "L'innovation : de l'idée au succès"*. Montpellier, France.
- [13] Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference, 16*, 642–649.
- [14] Risvik, E., McEwan, J. A., Colwill, J. S., Rogers, R., & Lyon, D. H. (1994). Projective mapping: A tool for sensory analysis and consumer research. *Food Quality and Preference, 5*, 263–269.

Givaudan

Givaudan

ENGAGING
THE
SENSES