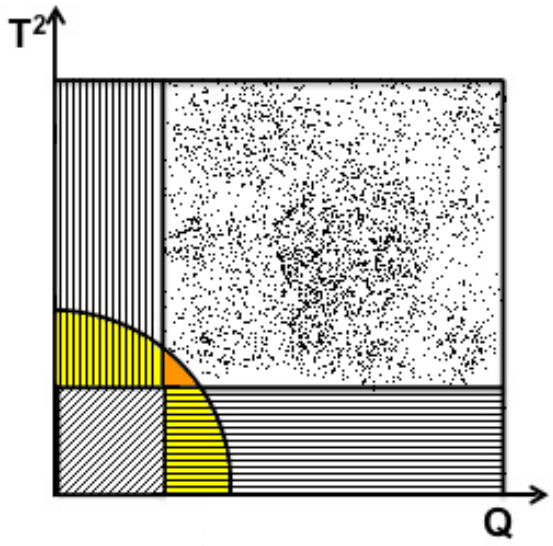
APPENDIX I

The differences between original and PLS toolbox SIMCA could be summarized through the following Figure.



Schematic representation of the differences among original SIMCA formulation and PLS toolbox one.

Given a generic set of objects, their T^2 and Q values are reported in the figure. The horizontal and vertical lines represent the T^2 and Q limit respectively. The semi-circle represents the SIMCA-PLS toolbox classification boundary. If it is assumed that the Q_{lim} is the same in both the SIMCA formulations (or it is assumed that the same objects are in or out with respect to both Q_{lim} and F-test with respect to $s_0^{(q)}$), the classification achieved by the two methods is the same for all the objects belonging to the shaded area, i.e in the small square down left Figure.

In the shaded area by horizontal lines original SIMCA considers the objects as not belonging to the class, since they are outside the Q_{lim}, thus being more severe than the SIMCA-PLS toolbox that accepts those in the green shaded area. The same holds for the orange area. In the dotted region $Q<Q_{lim}$ and

$T^2 > T_{lim}$, thus SIMCA-PLS toolbox will accept the objects in the green area. The same cannot be so easy to say in the original SIMCA case because of the normalization factor in the $d_p^{(q)}$ formula. As far as the building of class limits is concerned, some authors suggest using cross-validated scores instead of the fitted ones. This should reduce the number of false outliers and give more reliable classification. Furthermore, all initial and future scores are projected and can be assumed coming from the same distribution (See as reference the Section 2.1.3.2).

APPENDIX II

NSIMCA Toolbox, How to use

Since there are not papers regarding the practical use of the routine NSIMCA, in this section is reported a manual for the use of the NSIMCA toolbox written in MATLAB™ code developed by Cocchi et al. It will be soon available on the website http://www.models.kvl.dk/.

The classification method is composed by six different parts of code:

1. **NSIMCA.mat**
2. **NClass.mat**
3. **Toplot2D.mat**
4. **NSIMCAQTplots.mat**
5. **Make_coomans_SIMCA.mat**
6. **Toplot_coomans.mat**

1. **NSIMCA** performs the decomposition of the data set:

```
>>[XfitResults,XvalResults,XtestResults]=NSIMCA('method',[FacMax],[FacMin],'Xnam','class',['pretMod1';'pretMod2';'pretMod3'],[ordpret],[indblk],alpha);
```

INPUT INFORMATION

Method: PARAFAC `'par'` or Tucker3 `'tuc'`.

FacMax/FacMin: maximum and minimum number of factors for each class modeled, if the method chosen is PARAFAC (e.g. for two classes "…`,[9,10],[3,3],`…"); maximum and minimum combination of factors for the modes, for each class, if the method is Tucker3 (e.g. for two classes and a three-way data set "…`,[10 9 8;9 9 9],[4 3 4;2 4 4],`…").

Xnam/Class: data set and vector of classification, where in both are included test set samples, if any, with a value of zero in class, while the calibration set objects of each class are numbered from one to the number of classes respectively.

['pretMod1';'pretMod2';'pretMod3']: different pretreatment are implemented, mean-centering 'mnc', block-scaling 'blk', pareto-scaling 'psc', standard deviation scaling 'aut', for each mode a pretreatment must be insert (e.g. for a 3-way data set …,['mnc', 'psc','aut'],…), 'non' is used for no pretreatments to be applied in a given mode.

[ordpret]: Order of the pretreatment: gives the possibility to chose the order of pretreatment on each mode (e.g. …,[3 2 1],…).

[indblk]: only in blockscaling pretreatment it is applied and contains the vector for the application of the scaling, it is a structured array indicized with respect of the number of modes.

'alpha': it is the significance level for T^2 and Q distribution.

OUTPUT INFORMATION

It saves in the directory of work three structured array of dimension number of classes per number of factors tested in PARAFAC (or combination of factors tested if Tucker3 decomposition is used), containing fields with H, Q, Hlimit, Qlimit, RMSEC, R^2 for calibration [XfitResults], validation [XvalResults] and prediction [XtestResults].

2. **NClass** make class assignment by using ten different classification rules and calculat sensitivity and specificity for each class and for each factor:

```
>>Nclass_0(XfitResults,XvalResults,Xval_ncd_Results,Method,
rtype)
```

INPUT INFORMATION

Method: `'par'` or `'tuc'` like previously.

rtype: this parameter indicates the type of range for scores to be used in SIMCA original calculation; `rtype=1` refers to the original publication of Wold with $t_{max}+1/2*t_{student}*std(t)$, `rtype=2` normal range with t_{max}, `rtype=3` $t_{max}+std(t)$.

OUTPUT INFORMATION

`ClassResults`: It is a structure containing `SENSf SPECf` and `classtestf` containing sensitivity and specificity in cross validation for calibration and prediction (if any test set was given) `SENS_minpress SPEC_minpress` and `CLASStest_minpress` containing sensitivity for the different model dimensionalities.

Each column of the output corresponds to a different classification rules as follows:

1. H_fit/Q_fit (limF_hTfit Qlim_fit according to Forina)
2. H_CV/Q_CV (Hrlim_CV Qrlim_CV)
3. H_fit/Q_fit (hT_lim_fitAP Qlim_fitAP) Pomerentsev fit
4. H_CV/Q_CV(limAP_hT_CV limAP_QCV) Pomerentsev CV
5. SIMCA Wold fit
6. SIMCA Wold CV
7. SIMCA Wold fp (not defined for 2D data)

`CVResults`: It is a mat file containing `sensCV` and `specCV` with sensitivity and specificity calculated in cross validation. The geometric mean among these two features, called efficiency, is used to chose of the optimal dimensionality of the model.

3. **`Toplot2D`** makes a figure with as many subplots as the classes modeled, in which sensitivity (solid line) and specificity (dashed line)

are reported in function of the number of factors (PARAFAC) or combination of Factors (Tucker3) tested:

```
>> toplot2D(nc,SENSf,SPECf,classtestf,ci_test)
```

INPUT INFORMATION

`nc`: is the number of classes modeled

`SENSf`/`SPECf`/`classtestf`: are output from NClass, they are saved in the directory of work as `ClassResults` so they must be loaded on the workspace before the command is given.

`ci_test`: is the vector with class assignation for the test set.

OUTPUT INFORMATION

This routine makes plot of SENS and SPEC, from which the best dimensionality can be chosen and selected for each class, this information is given as input on the command line, for instance for three class the command is then given in this form `[4,12,62]` where the numbers represent the row in which the factor combination is calculated.

The efficiency calculated as geometric mean between SENS and SPEC is orbained from the following command:

```
>> geom=sqrt(SENSf(:,:,:).*SPECf(:,:,:));.
```

This solution sometimes can be more easy to interpret that the SENS/SPEC plot and will be implemented in the routine.

4. **`NSIMCAQTplots`** makes Q T^2 plots with all the criteria previously listed:

```
>>NSIMCA_QTplots(XfitResults,XvalResults,XtestResults,Xval_ncd
_Results,'method',[bestcomp],classid,ci_test)
```

INPUT INFORMATION

`XfitResults`, `XvalResults`, `XtestResults`: from `NSIMCA` like previously.

[bestcomp]: insert the best combination of factor selected in toplot.mat, the input must be a row vector containing best component for each class (e.g. for 3 class par: "…,[9 8 8],…", tuc : "…,[6 9 7;7 5 4;7 6 7],…"). Classid and ci_test are the class vector of training and test set.

OUTPUT INFORMATION

For each class it gives as response a plot of the object (training and test) in the Q vs. T^2 space reporting all the limits listed in the previous section, the values are calculated for the best component chosen in toplot2D. Figures are named and saved on the current folder.

5. **Makecoomans_SIMCA** calculates distances from the class boundaries and from the class plane in the original SIMCA framework:

```
>>makeCoomans_SIMCA(XfitResults,XvalResults,XtestResults,
'method',rtype,[bestcomp]);
```

INPUT INFORMATION

XfitResults, XvalResults, XtestResults: from NSIMCA like previously.

Method: 'par' or 'tuc' like previously. Like in Nclass_0 select 1 – 2 – 3 for the different SIMCA used. [Bestcomp]: insert the best combination of factors selected in toplot.mat, the input must be a row vector containing best component for each class (e.g. for 3 class par: "…,[9 8 8],…", tuc : "…,[6 9 7,7 5 4,7 6 7],…").

OUTPUT INFORMATION

It saves in the current directory two files distSIMCA and fSIMCA with the distances among classes needed for Coomans plot.

6. **Toplot_coomans** makes the Coomans plot for pairs of classes and plot the bisector:

```
>>toplot_coomans(distSIMCA,fSIMCA,icx,icy,classid,ci_test,'fit
')
```

INPUT INFORMATION

`DistSIMCA` and `fSIMCA` are the file (structure arrays) from the previous routine `makeCoomans_SIMCA`; `icx` is an integer specifying the number of the class whose distance you want to plot on the x-axis, `icy` is the same for the y-axis. `classid` and `ci_test` are the vectors of class for training and test.

OUTPUT INFORMATION

It makes Coomans plot of one class against another.

APPENDIX III

Concentrations of Lambrusco data set samples.

	Phenolic compounds						
SAMPLES	*Syrigic acid*	*Caffeic acid*	*Myrecetin*	*Quercetin*	*Gallic acid*	*(+)-Catechin*	*p-Coumaric*
w0001GRA	1,57	1,16	3,65	5,99	3,77	3,27	1,08
w0002GRA	1,20	0,92	3,46	6,17	3,38	3,12	1,37
w0003GRA	0,92	1,06	2,76	3,71	3,15	3,18	2,01
w0004GRA	1,59	1,51	5,12	9,29	4,16	3,36	4,46
w0005GRA	1,48	2,44	2,33	3,36	3,30	3,46	4,02
w0006GRA	1,12	1,13	2,12	4,00	3,00	3,12	1,65
w0007GRA	1,29	1,08	1,49	2,45	3,65	3,40	1,92
w0008GRA	1,25	0,98	2,41	5,78	3,27	3,22	2,13
w0009GRA	1,41	1,66	2,32	4,49	3,25	3,66	1,91
w0010GRA	0,83	1,86	5,00	9,40	3,40	3,06	6,10
w0011GRA	1,46	2,78	0,76	5,28	6,78	6,30	3,67
w0012GRA	1,02	1,10	2,29	4,21	3,46	3,27	1,77
w0013GRA	0,97	1,54	2,47	3,37	2,99	3,08	3,17
w0014GRA	2,07	2,60	6,44	11,58	6,11	4,45	3,96
w0015GRA	1,36	1,07	4,11	5,30	3,13	3,13	2,61
w0016GRA	1,38	1,53	2,53	3,32	3,22	3,37	2,79
w0017GRA	1,52	1,35	1,70	2,35	3,35	3,36	1,70
w0018GRA	1,27	1,55	3,06	3,14	3,84	3,58	2,54
w0019GRA	1,39	1,56	7,23	13,64	4,24	3,50	2,23
w0020GRA	0,97	1,19	2,48	3,33	3,30	3,34	2,09
w0021GRA	0,97	1,13	3,60	4,98	3,33	3,23	1,75
w0022GRA	1,05	0,97	1,57	2,90	3,72	3,21	1,68
w0023GRA	1,20	1,02	3,03	4,85	3,97	3,42	1,12
w0024GRA	1,11	0,84	2,84	4,71	3,46	3,18	1,02
w0025GRA	1,00	0,60	2,00	4,35	3,14	2,99	1,33
w0026GRA	1,62	1,00	2,65	3,05	3,31	3,26	1,94
w0027GRA	2,05	1,65	2,73	3,48	3,22	3,49	2,75
w0028GRA	0,99	0,77	3,18	4,46	3,15	3,13	1,49
w0029GRA	0,95	1,08	2,56	3,25	3,21	3,27	2,34
w0030GRA	1,62	1,39	3,63	4,67	3,34	3,28	1,99
w0031GRA	1,77	1,02	4,12	4,77	3,54	3,41	1,59
w0032GRA	1,34	1,29	3,30	5,32	4,40	3,29	1,10
w0033GRA	1,55	1,70	2,58	3,53	4,81	3,98	1,03
w0034GRA	1,43	1,34	3,03	5,27	3,72	3,41	2,43
w0035GRA	1,04	1,37	2,05	2,67	3,35	3,36	2,30
w0036GRA	0,96	1,20	2,37	3,31	3,24	3,22	2,28

w0037GRA	1,37	2,48	2,73	4,07	3,50	3,34	3,75
w0038GRA	1,21	1,72	2,80	3,24	3,91	3,60	0,72
w0001SAL	0,82	1,41	2,38	4,27	3,14	3,42	1,81
w0002SAL	1,44	2,44	6,23	10,59	4,62	3,85	1,81
w0003SAL	1,25	2,14	4,84	7,82	3,63	3,62	3,15
w0004SAL	0,94	2,14	2,02	3,15	8,70	6,06	2,80
w0005SAL	0,77	1,71	2,62	3,68	3,33	3,02	2,78
w0006SAL	1,00	1,70	4,03	6,66	4,19	3,74	2,24
w0007SAL	2,01	3,12	2,56	4,20	6,44	5,80	4,10
w0008SAL	1,06	1,52	2,94	4,58	3,80	3,40	2,65
w0009SAL	1,60	2,97	7,25	13,75	4,42	3,95	3,79
w0010SAL	0,92	1,80	3,17	4,92	5,37	4,70	2,63
w0011SAL	0,88	1,59	2,44	5,88	3,99	4,01	2,87
w0012SAL	0,92	1,98	2,86	6,35	6,05	4,30	2,83
w0013SAL	0,71	0,99	3,02	5,87	3,12	3,13	1,80
w0014SAL	0,68	1,51	4,77	8,56	4,53	3,50	1,83
w0015SAL	0,84	1,83	5,57	9,36	3,40	3,21	2,85
w0016SAL	0,64	1,71	2,66	4,30	3,37	3,23	3,03
w0017SAL	0,85	1,81	1,52	2,28	4,49	4,21	1,93
w0018SAL	0,63	0,91	2,81	3,96	3,10	3,07	1,28
w0019SAL	0,18	0,40	0,84	1,60	3,22	3,33	0,28
w0020SAL	0,88	2,44	3,72	5,55	3,80	3,79	0,38
w0021SAL	1,89	1,97	4,04	5,37	4,17	3,82	2,75
w0022SAL	1,02	0,93	3,85	4,26	3,02	3,11	0,87
w0023SAL	1,00	1,51	4,94	8,09	3,55	3,51	2,92
w0024SAL	1,15	2,35	2,80	6,36	8,87	6,03	3,45
w0025SAL	0,96	1,41	2,89	5,50	3,40	3,42	2,84
w0026SAL	0,91	1,49	5,58	7,73	3,38	3,34	3,26
w0027SAL	1,63	1,81	3,28	5,59	3,59	3,38	3,48
w0028SAL	0,73	0,86	1,61	2,56	3,07	3,05	1,74
w0029SAL	0,96	1,29	4,19	7,16	3,60	3,37	2,27
w0030SAL	0,63	1,20	2,40	4,30	3,07	3,16	2,41
w0031SAL	1,29	1,21	3,55	4,15	3,22	3,30	1,28
w0032SAL	1,00	1,61	2,57	4,48	4,75	3,42	2,99
w0033SAL	0,69	1,48	3,78	6,78	3,33	3,37	2,76
w0034SAL	0,61	1,47	2,81	5,38	3,38	3,40	2,80
w0035SAL	0,97	2,11	2,09	2,79	8,32	6,91	2,09
w0036SAL	0,85	1,38	1,98	3,95	3,27	3,40	1,88
w0037SAL	1,31	2,99	5,10	8,31	8,62	6,84	4,71
w0038SAL	0,88	1,47	3,38	6,32	5,03	4,12	2,49
w0001SOR	0,51	1,70	1,70	3,22	3,09	5,39	2,06
w0002SOR	1,00	1,75	1,66	2,80	3,35	3,39	3,05
w0003SOR	0,32	1,07	0,96	1,98	3,15	3,29	1,34

w0004SOR	0,92	1,78	0,93	1,91	3,31	3,54	2,36
w0005SOR	1,00	2,56	2,46	4,98	3,74	4,70	3,51
w0006SOR	0,66	1,31	1,50	2,04	3,05	3,13	2,63
w0007SOR	0,94	2,98	3,06	4,90	4,25	5,42	3,40
w0008SOR	1,05	1,87	5,48	9,33	3,39	3,32	3,56
w0009SOR	0,79	1,84	2,34	4,34	3,31	3,60	3,35
w0010SOR	0,40	1,12	0,89	1,91	3,00	3,28	1,31
w0011SOR	0,36	1,17	1,66	4,12	3,07	3,17	2,69
w0012SOR	0,55	2,77	1,19	2,76	7,32	7,52	4,74
w0013SOR	0,84	1,33	2,02	4,35	3,56	3,66	2,35
w0014SOR	0,37	0,79	1,10	1,95	3,08	3,14	0,90
w0015SOR	0,57	1,71	1,74	3,27	3,14	3,14	3,44
w0016SOR	1,30	1,63	3,18	5,65	8,06	4,78	2,41
w0017SOR	0,78	1,60	1,24	2,31	3,12	3,31	3,65
w0018SOR	0,69	1,75	1,28	2,31	3,29	4,42	1,89
w0019SOR	0,36	1,07	0,93	1,80	3,00	3,15	1,57
w0020SOR	0,76	1,52	1,67	2,22	4,39	4,51	2,23
w0021SOR	0,46	0,97	1,39	2,25	3,07	3,16	2,32
w0022SOR	0,66	1,33	1,65	2,54	3,17	3,25	3,22
w0023SOR	0,96	2,30	1,16	1,99	4,15	4,94	2,69
w0024SOR	0,43	1,76	1,30	1,99	3,18	3,13	3,86
w0025SOR	0,46	2,18	1,52	2,23	3,34	6,14	2,18
w0026SOR	0,86	1,71	1,10	2,57	3,09	3,52	3,57
w0027SOR	0,97	1,52	2,20	3,82	3,37	3,42	3,29
w0028SOR	0,36	1,31	1,09	2,10	3,08	3,38	1,79
w0029SOR	0,49	1,13	1,87	3,26	3,14	3,39	1,82
w0030SOR	0,45	1,35	1,55	2,37	3,17	3,35	2,74
w0031SOR	0,78	2,08	1,38	2,14	3,59	5,03	2,73
w0032SOR	0,81	1,53	2,09	3,17	3,32	3,53	3,68
w0033SOR	0,32	3,27	1,46	2,64	7,80	5,00	5,69
w0034SOR	0,64	1,46	1,80	2,37	3,41	3,15	3,03

Legend	
[Con]	<LOQ
	<LOD

APPENDIX IV

List of pubblications:

E. Salvatore, M. Cocchi, A. Marchetti, F. Marini, A. de Juan; *Determination of phenolic compounds and authentication of PDO Lambrusco wines by HPLC–DAD and chemometric techniques*. Analytica Chimica Acta (2013) 761 34–45.

E. Salvatore, M. Bevilacqua, R. Bro, F. Marini, M. Cocchi; *Classification methods for multiway arrays as a basic tool for food PDO authentication*. In M. De La Guardia and A. Gonzalvez Illueca (Eds.) *Food protected designation of origin: methodologies & applications.* Wilson & Wilson's Comprehensive Analytical Chemistry, Volume 60, 2013. Elsevier B.V., in press.

E. Salvatore, M. Cocchi, F. Marini, A. de Juan; *Multi-way and Multiset approaches for classification. Applications on food analysis*. (In preparation)

Other publications not included in the thesis:

M. Silvestri, L. Bertacchini, C. Durante, A. Marchetti, E. Salvatore; Application of data fusion techniques to direct geographical traceability indicators. Analytica Chimica Acta (2013) in press.

L. Bertacchini, M. Cocchi, M. Li Vigni, A. Marchetti, E. Salvatore, S. Sighinolfi, M. Silvestri, C. Durante, "Chpt. 10 The impact of chemometrics on food traceability". In F. Marini (Ed.) Chemometrics in Food Chemistry Vol. 28 Data Handling in Science and Thecnology series, Elsevier 2013, in press.

M. Franco, R. Jasionowska, E. Salvatore; Application of CZE methods in routine analysis for determination of B-complex vitamins in pharmaceutical and veterinary preparations. International Journal of Analytical Chemistry (2012).

ACKNOWLEDGEMENTS

When I embarked on this experience, three years ago, writing this page was the least of my worries, but everything has an end fortunately and unfortunately, and it is my desire to express my gratitude to all the people who encouraged, supported and inspired me.

It is a must starting with the working group of UNIMORE. Prof. Marina Cocchi, for much more than I could write here (really thank you!). Prof. Lorenzo, Tassi, Prof. Andrea Marchetti and Dr. Simona Sighinolfi for making their immense experience available. Dr. Caterina Durante, Dr. Mario Li Vigni, Dr. Carlo Baschieri, Dr. Lucia Bertacchini and Dr. Michele Silvestri, for their experience, but also for the pleasure of having their friendship.
My 'second house' Sapienza, represented by Dr. Federico Marini, Dr. Antonio Daloise, Dr. Riccardo Nescatelli, Dr. Marta Bevilacqua, Dr. Raffaele Vitale and Prof. Remo Bucci.
The chemometrics research group of Universitat de Barcelona, where I spent my period as a visiting student: Prof. Anna de Juan, Dr. Silvia Mas, Dr. Sara Piqueras, Dr. Daniel Serrano.
My parents, my family, my friends. You are my backbone.
Stefano Liguori, just to remember you that, if we started with the beginning of my PhD, this does not mean that now it is time to stop.

There is only a person I would like to dedicate this work, for two reasons. The first is that, although this thesis might not be worth a Noble Prize, I know you would appreciate it, perhaps even more. The second is that nobody told you that March 15th is the Ides of March, in Roman calendar the beginning of Spring.
Don't forget it!

To Vincenzo Pierini