

## Supplementary Material B

### Estimation of SIMCA PLS/OPLS parameter variability associated with row rearrangement.

In the SIMCA default procedure, the result of a row rearrangement is that validation and calibration subset selections are modified. As a consequence, the variability of the SIMCA parameters associated with row rearrangements is equivalent to the variability obtained when these subsets are randomly selected. By combining Microsoft Excel software and SIMCA it is possible to build various models for the same dataset with different selections of validation and calibration subsets. The variability of the SIMCA parameters can be estimated by comparing the values calculated with the various models. A simple procedure is proposed to compare M models of the same dataset.

- 1/ With Excel program insert M new columns in your dataset.
- 2/ Fill all the cells of these new columns with random numbers by using the rand() function.
- 3/ Create a new SIMCA project with the modified dataset. Define the new columns as “Secondary Data”.
- 4/ Select the PLS or OPLS analysis and push the Autofit button. A first model is built.
- 5/ In the Workset menu, create a new model as the previous one.
- 6/ Select the Model Options in the Workset menu.
- 7/ Select the CV-groups tab.
- 8/ Change the cross validation groups by selecting the button “Assign observation based on variable” and select the first column containing the random numbers.
- 9/ Build the model with Autofit.
- 10/ Repeat the steps 3/ to 9/ for the M-1 remaining columns inserted in the original dataset.