SAISIR: A New Generalist Chemometric Toolbox

Christophe B.Y. Cordella (a)*, Dominique Bertrand (b)

(a) Institut National de la Recherche Agronomique, UMR1145 GENIAL, Equipe Ingénierie Analytique pour la Qualité des Aliments/Chimiométrie, 16 rue Claude Bernard 75005 PARIS – France

(b) Institut National de la Recherche Agronomique, Unité Biopolymères Interactions Assemblages, Equipe Bioinformatique, Rue de la Géraudière-BP 71627, 44316 Nantes cedex 3 - France

The SAISIR Toolbox for Matlab®, Octave and Scilab is a freely available collection of functions and algorithms for modeling physicochemical, sensorial data and multidimensional data by a large range of bilinear and multilinear models. The **SAISIR** toolbox can be requested for a free of charge CD-ROM shipping via internet from http:// www.chimiometrie.fr/saisir_request.html.

In most of the chemometric works, it is needed to manipulate large data files and to have the capability to make complex chain of processing steps. For example, a simple data processing may consist in loading data such as spectra, numeric images, sensory data …, pre-treating them by specific relevant methods, sampling the data, and eventually processing them using PCA, PLS, or any adapted methods (multiblock table methods, discrimination …). Moreover, it is essential for the user to keep the trace of the identifiers of the rows and the columns of the data matrix. For example, when processing spectra, it is important to identify the particular role of given variables (labeled, for example, by a wavenumber or a wavelength value). In the same way, the rows of the data matrices are generally logically identified by *names making sense to the chemometrician*. When the chain of procedures is complex, it is not easy to make such a work in graphically interfaced environments. One may want to keep the trace of the applied procedures and to be able to re-use them on other data sets, or in cross-validation procedures. The **SAISIR** (*statistics applied to the interpretation of InfraRed spectra*) environment has been developed in order to cope with such situations. Initially devoted to spectral data, it as been extended as a generalist environment coping with many kinds of data. **SAISIR** contains many functions for loading, saving, manipulating or displaying data (See CD-ROM presentation in figure 1). Contrary to windowed environments, it has the great advantage to allow batch procedures. It also makes it possible to mix data of any origin (chemical and physical data, spectroscopic data, numeric images ...). Due to its simple way of working, it is always possible to find a practical solution, and the user is never at complete standstill. Over 200 functions are available in **SAISIR** involved in data processing field such as graphics (biplot, curves, confidence ellipse, barycenter representation, dendrogram...), factorial methods (PCA, correspondence analysis, factorial distance analysis...), discriminative methods (linear, factorial, quadratic , PLS, stepwise...), regression methods (PLS1, PLS2, PCR, ridge, latent root , stepwise...). SAISIR is particularly rich in function dealing with multiblock tables (STATIS, ACOM, ACCPS, Multiple factor analysis, canonical analysis...) and more...**Keywords**: Chemometric toolbox, MATLAB, OCTAVE, SCILAB, physicochemical, sensorial data, regression, multidimensional analysis, discrimination, multiblock analysis