Training Course on Multivariate Calibration in R

We are pleased to inform you that the Natural Resources Cluster, from the Centre for Development and Environment, University of Bern, announces an **R based** training course focusing on **soil spectroscopy**.

"The R project for Statistical Computing: An ideal tool to build bridges between soil routine analysis and Reflectance Spectroscopy"

Date: 12- 14th November, 2012

Venue: 12th November: **University main building, Room 331**
                   13th & 14th November: **University main building, Room 304**

Target Audience
The course is aimed at scientists and PhD students in soil science and physical geography with background in statistics and ideally basic knowledge of R. Own laptop required with R 2.15.1 and RStudio installed.

Objectives of the course
This training course is pursuing two goals: (1) Provide sound introduction in the elaboration of a multivariate calibration model for soil spectroscopy, and (2) promote the use of R as a first-choice research tool.
The aim is to provide a focused and concentrated training course on statistical procedures required for soil spectral analysis and modeling using data-mining and machine learning techniques as well as programing in R environment. However, such procedures are widely applied in natural sciences and therefore the training course is certainly also interesting and instructive for participants from other fields of natural science.
The course will be conducted in English by Leonardo Ramirez Lopez and Antoine Stevens from the Georges Lemaître Centre for Earth and Climate Research, Université catholique de Louvain.

Kindly find the course details on the next page. Please send your application to Christian Hergarten: christian.hergarten@cde.unibe.ch

Please note that the **deadline for applications is November 8, 2012 and the number of participants is limited to 15.** Participants from CDE and GIUB take priority; from other participants a fee of 300 CHF will be demanded.

Weblinks:

R Statistical Software: www.r-project.org/
RStudio: http://rstudio.org/

Training Programme:

Day 1

Course Introduction

Basics on soil spectroscopy

Why using R in soil spectroscopy?

Terminology and notation

RStudio, an IDE for R

Basic linear algebra review in R (optional):

- Matrices and vectors
- Addition and scalar multiplication
- Matrix vector multiplication
- Inverse and transpose

Basics of R:

- R objects (list, data.frame, factor, vector, matrix, …)
- Read/write data from file
- Basic operations (arithmetics, …)
- Data handling (data subsetting, checking data structure,…)
- Computing on data (column and row operations, apply family,…)
- Programming and looping with for, while and repeat
- Function creations
- Graphical procedures for representing (spectral) data

Day 2

Signal processing:

- Linear filters (smoothing, derivatives, etc.)
- Centering and scaling
- Continuum removal
- Principal component analysis for spectral de-noising
-

Pattern recognition:

- Methods for spectral dimensionality reduction
- Principal components analysis
- Supervised and unsupervised algorithms for pattern recognition
- Computing reliable similarity/dissimilarity measures between soil spectra
- Clustering and cluster validation procedures

Multivariate calibration (Part 1):

- Principles and terminology
- Data preparation for modeling (outlier detection, …)
- Calibration sampling algorithms (Kennard–Stone, Fuzzy c-means sampling, conditioned latin hypercube, etc)
- Feature transformation, selection, extraction and compression (wavelet transform, recursive feature elimination, …)
- Multivariate calibration for regression and classification
  - Multiple linear regression and classification
  - Partial least squares
  - Support vector machines
  - Regression trees and random forest
  - Multivariate Adaptive Regression Splines
  - Artificial neural networks

Day 3
Multivariate Calibration (Part 2)

- Writing your own local models (locally weighted regression, LOCAL, CARNAC, etc)
- Validation strategies (autoprediction, cross-validation, boosting, independent validations, etc)
- Extract knowledge from your models: analyzing model performance (RPD, R2 ,…) and variable importance for modeling
- improving soil predictions by incorporating spatial information (regression kriging, definition of proper sample domains, etc.)
- Other useful data mining and machine learning tools (Cubist, …)
- An unified interface for predictive models: the caret packageUseful R packages and functions for soil spectroscopy (hyperSpec, soil.spec, kernlab, cluster, signal,…)
- Speeding up you R code: Parallelization with the foreach package