ModF User's Guide

Version 2.0 "for stand-alone application", January 2011

This documentation explains how to use the computer program called "ModF", which implements the model-free modified *F*-test and the model-based modified *F*-test with estimated ranges from Dutilleul *et al.* (2008, Section 5).

Reference:

Dutilleul, P., Pelletier, B., and Alpargu, G. 2008. Modified *F* tests for assessing the multiple correlation between one spatial process and several others. *Journal of Statistical Planning and Inference* 138:1402–1415.

Installation

1- Following the instructions posted at http://environmetricslab.mcgill.ca, download **mcrinstaller.exe** from our ftp server and install it on your computer.

2- Retrieve a copy of the file **ModF_standalone.zip** on your computer and unzip its content to a folder of your choice.

3- Launch the program by clicking on **ModFui.exe**. Two windows should then appear: a user interface, to run the ModF program, and a command window. Note that the first time the program is launched, a folder containing encrypted Matlabfiles will be generated in your folder.

In the user interface, you are asked to perform the following tasks:

1- Data file: Write the name of a tab-delimited text file containing

- On the first row, the names (identifiers) of the spatial coordinates and the $1 + q$ variables.

  Followed (starting on the second row) by

- In the first column, the indices from 1 to N, where N is the total number of sampling locations.
- In the second and third columns, the spatial coordinates of the N sampling locations in 2-D space.
- In the fourth column, the corresponding observations for Y, denoted $\{y(\mathbf{u}_1), \ldots, y(\mathbf{u}_N)\}$ in Dutilleul *et al.* (2008).
- In the fifth and following columns, the associated observations for $X_1, \ldots, X_q$, with $q \geq 1$.

2- Box-Cox transformation: Check the box if you want to perform a Box-Cox transformation of your data. This procedure is aimed at improving the normal distribution assumption on the transformed data. This power transformation is defined by $y' = (y^{\lambda} - 1)/\lambda$ when $\lambda \neq 0$ and $y' = \ln(y)$ when $\lambda = 0$. In ModF, the coefficient λ is selected among values from -2.5 to 2.5, by steps of 0.1.
Reference:
Box, G.E.P and Cox, D.R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, series B* 26:211–243.

3- Standardization. Check the box if you want to standardize the variables to a zero mean and a variance of one. Note that when the Box-Cox transformation is used, variables are automatically standardized and this option is no longer available.

4- Write a set of characters (see *Userdefinedprefix* below) that will be used to identify the four output files (*.txt) generated by the program and saved in the Matlab folder.

Output files

The content of the four output files can be described as follows.

*Userdefinedprefix*_results.txt: Comprises three columns corresponding to the classical, unmodified *F*-test and the model-free and model-based modified *F*-tests, respectively. The classical R^2 (the same for the three tests) is on the first row. The three *F*-test statistics are reported on the second row, the sample size used in each test on the third row, and the corresponding probabilities of significance on the fourth row.

*Userdefinedprefix*_parameter_estimates.txt: Is specific to the model-based modified *F*-test. It contains the variogram model parameter estimates obtained by fitting linear models of regionalization to the direct variograms of the dependent variable *Y* (first column) and its OLS predictor (second column) defined from $X_1, \ldots, X_q$.The estimated nugget effect is on the first row, and the estimated sill and range of the spherical structure on the second and third rows, respectively.

*Userdefinedprefix*_variograms.txt: Comprises four columns with (1) the number of pairs of observations per distance class, (2) the mean distance value per class, and (3–4) the corresponding values of the experimental variograms computed for *Y* and its OLS predictor, respectively.

*Userdefinedprefix*_boxcox.txt: Comprises two columns with the coefficient λ used in the Box-Cox transformation, i.e., between -2.5 and 2.5, and an indicator of whether or not the variables were standardized to a zero mean and a variance of one, i.e., standardized = 1; non-standardized = 0.

Example

The dataset used in the example of Dutilleul *et al*. (2008) is provided here as a text file (*datatest.txt*). Since the definition of distance classes in the computation of experimental variograms in the ModF program differs from that in Dutilleul *et al.* (2008) (see below), the results are slightly different from those published. Note also that Dutilleul *et al.* (2008) used the Box-Cox transformation.

Complementary note 1

This note in four parts is about the definition of distance classes in the computation of experimental variograms. First, the area covered by the sampling grid is estimated by convex hull, using the Matlab function "convhull". Second, half the side length of the square with same area is used as maximum lag distance. Third, this maximum lag distance is divided by 8 to obtain the minimum lag distance, which is also used as the increment between distance classes. If there are less than 100 pairs of observations in at least one distance class, the maximum lag distance is divided by 7, 6, etc., until each distance class has at least 100 pairs of observations or the number of distance classes is four. Fourth and last, the mean distances of classes are used to plot experimental variograms and fit variogram models.

Complementary note 2

The square of a "Student" *t*–distribution with *k* degrees of freedom being a Fisher-Snedecor *F*–distribution with 1 and *k* degrees of freedom, it follows that ModF can also be used to perform Dutilleul's (1993) modified *t*-test for assessing the significance of the simple correlation between two spatial processes.

Reference:

Dutilleul, P. 1993a. Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics* 49:305–314.