

# ModF User's Guide

## Version 2.0 “for Matlab users”, 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) **in Matlab**.

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.

Retrieve a copy of the file ModF\_Matlab.zip on your computer and unzip its content to a folder on the Matlab path.

In Matlab, call the **ModF** function:

$$[output] = \text{ModF}(data, varname, casename, boxcox, standardize, 'outputprefix')$$

where *output* is a *structure array*; *data* is a *numerical array*; *varname* and *casename* are *character arrays*; *boxcox* and *standardize* are either “1” or “0”; and ‘*outputprefix*’ is a *string*. The names of the input and output files are defined by the user.

### Input files

#### data:

First and second columns: Spatial coordinates of the  $N$  sampling locations in 2-D space

Third column: The corresponding observations for  $Y$ , denoted  $\{y(\mathbf{u}_1), \dots, y(\mathbf{u}_N)\}$  in Dutilleul *et al.* (2008).

Fourth and following columns: The associated observations for  $X_1, \dots, X_q$ , with  $q \geq 1$ .

#### varname:

Contains the names (identifiers) of spatial coordinates and  $1 + q$  variables.

#### casename:

Contains the indices from 1 to  $N$ , where  $N$  is the total number of sampling locations.

#### boxcox:

Use “1” when you want to apply the Box-Cox power transformation to your data, and “0” otherwise. 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  $\lambda$  is selected among values from -2.5 to 2.5, by steps of 0.1.

standardize:

Use “1” when you want to standardize the variables to a zero mean and a variance of one, and “0” otherwise. Note that when the Box-Cox transformation is used, variables are automatically standardized.

'outputprefix':

Will be the name used to identify the set of output files (\*.txt) generated by the program and saved in the Matlab folder.

Note: If you do not want to save the results as text files, simply omit this input file and type

$$[output] = \text{ModF}(data, varname, casename, boxcox, standardize)$$

## Output file

output:

Name of the output file generated in the workspace of Matlab and containing the results of the data analysis. The output file is made of the following 19 fields, which can be viewed by typing *output* on the command line.

R2_classic	Classical $R^2$
Fstatistic_classic	Classical, unmodified $F$ -test statistic
Ftest_numer_df	Number of degrees of freedom of the numerator
Ftest_denom_df	Number of degrees of freedom of the denominator
pvalue_classic	Probability of significance for the classical $F$ -test
effective_sample_size_modelfree	Effective sample size for the model-free modified $F$ -test
Fstatistic_modelfree	Model-free modified $F$ -test statistic
pvalue_modelfree	Probability of significance for the model-free modified $F$ -test
effective_sample_size_modelbased	Effective sample size for the model-based modified $F$ -test
Fstatistic_modelbased	Model-based modified $F$ -test statistic
pvalue_modelbased	Probability of significance for the model-based modified $F$ -test
nugget_observedY	Estimated nugget effect for the dependent variable $Y$ in the model-based modified $F$ -test
sill_spherical_observedY	Estimated sill of the spherical model for the dependent variable $Y$ in the model-based modified $F$ -test
range_observedY	Estimated range of the spherical model for the dependent variable $Y$ in the model-based modified $F$ -test
nugget_predictedY	Estimated nugget effect for the OLS predictor of $Y$ in the model-based modified $F$ -test
sill_spherical_predictedY	Estimated sill for the OLS predictor of $Y$ in the model-based modified $F$ -test

range_predictedY	Estimated range for the OLS predictor of $Y$ in the model-based modified $F$ -test
Variograms	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
Boxcox	A column with the value of coefficient $\lambda$ used for the Box-Cox transformation of each variable

## Output files

In addition to being generated in the workspace of Matlab, the results of the data analysis can be saved as text files in the Matlab folder, using ‘*outputprefix*’ as prefix. Recall that no text file is generated if ‘*outputprefix*’ is omitted when calling the ModF function; see above. The four text files that can be saved in the Matlab folder are described below.

*Outputprefix\_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.

*Outputprefix\_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  $X_1, \dots, X_q$ . 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.

*Outputprefix\_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.

*Outputprefix\_boxcox.txt*: Comprises two columns with the coefficient  $\lambda$  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*). It can also be directly loaded into the Matlab workspace using the command `load('datatest.mat')`. 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.