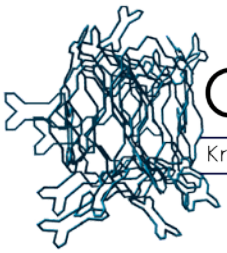

C O M P L E X

Knowledge Based Climate Mitigation Systems for a Low Carbon Economy

Software package 2: data-driven stochastic modelling

Date	**Tuesday, 31 May 2016**
Report Number	**D6.11**
VERSION NUMBER:	1.0
Main Author:	Anna Shchiptsova

DIFFUSION LEVEL – PU

PU	**PUBLIC**
RIP	**RESTRICTED INTERNAL AND PARTNERS**
RI	**RESTRICTED INTERNAL**
CO	**CONFIDENTIAL**

Coordinator: Nick Winder, nick.winder@ncl.ac.uk

INFORMATION ON THE DOCUMENT

Title	D6.11 Software package 2: data-driven stochastic modelling
Authors	Anna Shchiptsova

DEVELOPMENT OF THE DOCUMENT

Date	Version	Prepared by	Institution	Approved by	Note
30.05.2016	1.0	AS	IIASA		

CONTENTS

1 Overview

URL:
http://www.iiasa.ac.at/web/home/research/researchPrograms/AdvancedSystemsAnalysis/land-use-spatial-analysis.html

Available data for download contains jar files, source code in zip archives, example datasets in zip archives.

Components:
Package 1. lu-preprocessing-1.0.0-standalone.jar
Package 2. lu-rescaling-1.0.0-standalone.jar
Package 3. lu-regression-1.0.0-standalone.jar
Package 4. lu-approximation-1.0.0-standalone.jar

Installation:
No installation needed. Packages are supplied as standalone java applications.
Requires JRE 1.8 on the target machine.

Author and maintainer:
Anna Shchiptsova, International Institute for Applied Systems Analysis (IIASA)

Repository:
github.iiasa.ac.at/LandUseSpatialDynamics.git (*not publicly available*)

Development environment:
All components are developed using Clojure v.1.8.0 and Incanter v.1.5.7, a Clojure-based, R-like statistical computing and graphics environment for the JVM. Both are distributed under an open source software license EPL.

2 Description

Software is supplied in four packages, which should be executed sequentially. Input data is supplied either in asc format used for the GIS-based variables (lattice resolution) or in csv format used for census data. Table 1 summarizes execution steps.

Table 1. Description of execution steps.

STEP 1. Data cleaning goal: preprocessing of the GIS based variables. file: lu-preprocessing-1.0.0-standalone.jar input: .asc files output: .asc files	STEP 2. Rescaling goal: upscaling of the GIS based variables file: lu-rescaling-1.0.0-standalone.jar input: asc files output: csv files
STEP 3. Regression analysis goal: model selection and evaluation file: lu-regression-1.0.0-standalone.jar input: csv file output: csv files	STEP 4. Approximation goal: lattice approximation and accuracy estimation file: lu-approximation-1.0.0-standalone.jar input: csv files output: csv files

Detailed algorithm is described as follows:
STEP 1. Exclude cells, which either belong to the border of the studied area or have undefined values in their neighbourhood in any of the given maps in the dataset. Drop cells, which fall into protected natural areas or cells whose neighbourhood contains cells belonging to these areas.
STEP 2. Upscale the GIS-based variables to the level of administrative division by averaging the values on the lattice.
STEP 3. Run multiple linear regression with the unknown joint distribution of error terms using OLS and resampling. Regression coefficients are estimated using ordinary least squares. Perform hypothesis testing using the permutations method to assess overall model significance and individual coefficient's significance. Estimate confidence intervals for different model statistics using bootstrapping.
STEP 4. Approximate lattice data by the multiple linear regression model with resampling and supply estimates of the approximation accuracy. Regression coefficients are estimated using ordinary least squares. Accuracy estimates are calculated for cells grouped by the number of urban cells in the Moore neighbourhood (comprises the eight cells surrounding a central cell on a two-dimensional square lattice).

References:
[1] Anderson, M. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. Canadian Journal of Fisheries and Aquatic Sciences, 58(3): 626-639. DOI: 10.1139/f01-004.
[2] Freedman, D., & Lane, D. (1983). A Nonstochastic Interpretation of Reported Significance Levels. Journal of Business & Economic Statistics, 1(4): 292-298. DOI: 10.2307/1391660.

[3] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics, 7(1): 1-26. DOI:10.1214/aos/1176344552.
[4] Efron, B., & Tibshirani, R. (1993). An Introduction to the Bootstrap. New York: Chapman and Hall.
[5] Geary, R. (1954). The Contiguity Ratio and Statistical Mapping. The Incorporated Statistician, 5(3): 115-145. DOI: 10.2307/2986645.
[6] Moran, P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika, 37(1-2): 17-23. DOI: 10.2307/2332142.
[7] Lin, K.-P., Long, Z.-H., & Ou, B. (2011). The Size and Power of Bootstrap Tests for Spatial Dependence in a Linear Regression Model. Computational Economics, 38(2): 153-171. DOI: 10.1007/s10614-010-9224-0.

3 Packages

Package 1. Preprocessing of the GIS-based Variables	
Name	lu-preprocessing-1.0.0-standalone.jar
Summary	Standalone application for cleaning data
Version	1.0.0
License	MIT http://opensource.org/licenses/MIT
Description	Excludes sample points with undefined values, generates additional asc maps
Imports	Clojure 1.8.0, https://clojure.org/

Package 2. Rescaling and Transformations	
Name	lu-rescaling-1.0.0-standalone.jar
Summary	Standalone application for rescaling asc data
Version	1.0.0
License	MIT http://opensource.org/licenses/MIT
Description	Supports upscaling of GIS-based data to the level of administrative division
Imports	Clojure 1.8.0, https://clojure.org/

Package 3. Multiple Regression with Resampling	
Name	lu-regression-1.0.0-standalone.jar
Summary	Standalone application for regression analysis with resampling
Version	1.0.0
License	MIT
	http://opensource.org/licenses/MIT
Description	Runs multiple linear regression with the unknown joint distribution of error terms using OLS and resampling
Imports	Clojure 1.8.0, https://clojure.org/
	Incanter 1.5.7, http://incanter.org/

Package 4. Approximation	
Name	lu-approximation-1.0.0-standalone.jar
Summary	Standalone application for approximation
Version	1.0.0
License	MIT
	http://opensource.org/licenses/MIT
Description	Approximates data by the multiple linear regression model and supplies estimates of the approximation accuracy
Imports	Clojure 1.8.0, https://clojure.org/
	Incanter 1.5.7, http://incanter.org/