Introduction to Exploratory Spatial Analysis of the non-Temporal GIS Data

Anna Shchiptsova

September 30, 2016

International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria

This document outlines the use of software components in conducting exploratory spatial analysis of the non-temporal GIS data. The goal is to study association between variables over space using resampling methods in regression analysis.

CONTENTS
1 Installation
2 Overview
3 Preprocessing of the GIS-based variables
   3.1 Background
   3.2 Implementation in 'lu-preprocessing-1.0.0-standalone.jar'
4 Upscaling the GIS-based variables
   4.1 Background
   4.2 Implementation in 'lu-rescaling-1.0.0-standalone.jar'
5 Regression analysis with resampling
   5.1 Background
   5.2 Implementation in 'lu-regression-1.0.0-standalone.jar'
6 Approximation to the GIS lattice
   6.1 Background
   6.2 Implementation in 'lu-approximation-1.0.0-standalone.jar'
7 Example Data
Acknowledgements
References
1 Installation

No installation needed. All packages are standalone java applications. Requires JRE 1.8 installed on the target machine.

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

Available data contains jar files, source code 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'
- Additional package: 'regression-tests-1.0.0-standalone.jar'

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

Development environment: All components were 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 Overview

The software is supplied in four jar packages, which should be executed sequentially. Execution steps are carried out as follows:

STEP 1: Preprocessing of the GIS-based variables (using 'lu-preprocessing-1.0.0-standalone.jar' package).
The procedure includes filtering of cells with undefined values, reclassification of cell values in the land use map, logarithmic and unit rescaling transformations of the cell values in other maps, creation of spatial proximity matrix of contiguous administrative units.

STEP 2: Upscaling the GIS-based variables (using 'lu-rescaling-1.0.0-standalone.jar' package).
The GIS-based variables are upscaled to the level of administrative subdivision by averaging the cell values on the GIS lattice.

STEP 3: Regression analysis with resampling (using 'lu-regression-1.0.0-standalone.jar' package).
Regression coefficients are estimated using the ordinary least squares method. Hypothesis testing on overall model significance and individual coefficient significance are conducted using the method of permutations (Fisher, 1935). Percentile bootstrap scheme (Efron, 1979; Efron and Tibshirani, 1993) is applied for parameter estimation.
STEP 4: Approximation to the GIS lattice (using 'lu-approximation-1.0.0-standalone.jar' package).

Regression model estimated at the level of administrative subdivision is treated as a stochastic approximation to the GIS lattice. We apply bootstrapping to estimate the accuracy of model approximation to cell values.

Additionally, spatial autocorrelation in the regression residuals is checked using 'regression-tests-1.0.0-standalone.jar' package. Details of this procedure can be found in the "Testing Spatial Autocorrelation with the Bootstrap" guide (Shchiptsova, 2016).

Software components include:

<table>
<thead>
<tr>
<th>Name</th>
<th>lu-preprocessing-1.0.0-standalone.jar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>jar package</td>
</tr>
<tr>
<td>Summary</td>
<td>Standalone application for preprocessing of the GIS-based variables.</td>
</tr>
<tr>
<td>Version</td>
<td>1.0.0</td>
</tr>
<tr>
<td>License</td>
<td>MIT, <a href="http://opensource.org/licenses/MIT">http://opensource.org/licenses/MIT</a></td>
</tr>
<tr>
<td>Imports</td>
<td>Clojure 1.8.0, <a href="https://clojure.org/">https://clojure.org/</a></td>
</tr>
<tr>
<td>Command line options</td>
<td>-t, --trace  Print stack trace</td>
</tr>
<tr>
<td></td>
<td>-h, --help   Print command help</td>
</tr>
<tr>
<td>Author and maintainer</td>
<td>Anna Shchiptsova, <a href="mailto:shchipts@iiasa.ac.at">shchipts@iiasa.ac.at</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>lu-rescaling-tests-1.0.0-standalone.jar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>jar package</td>
</tr>
<tr>
<td>Summary</td>
<td>Standalone application for rescaling .asc data</td>
</tr>
<tr>
<td>Version</td>
<td>1.0.0</td>
</tr>
<tr>
<td>License</td>
<td>MIT, <a href="http://opensource.org/licenses/MIT">http://opensource.org/licenses/MIT</a></td>
</tr>
<tr>
<td>Imports</td>
<td>Clojure 1.8.0, <a href="https://clojure.org/">https://clojure.org/</a></td>
</tr>
<tr>
<td>Command line options</td>
<td>-t, --trace  Print stack trace</td>
</tr>
<tr>
<td></td>
<td>-h, --help   Print command help</td>
</tr>
<tr>
<td>Author and maintainer</td>
<td>Anna Shchiptsova, <a href="mailto:shchipts@iiasa.ac.at">shchipts@iiasa.ac.at</a></td>
</tr>
</tbody>
</table>
### Name
lu-regression-1.0.0-standalone.jar

### Type
directory

### Summary
Standalone application for regression analysis with resampling

### Version
1.0.0

### License
MIT, http://opensource.org/licenses/MIT

### Imports

### Command line options
- `-t, --trace` Print stack trace
- `-h, --help` Print command help

### Author and maintainer
Anna Shchiptsova, shchipts@iiasa.ac.at

---

### Name
lu-approximation-1.0.0-standalone.jar

### Type
directory

### Summary
Standalone application for approximation

### Version
1.0.0

### License
MIT, http://opensource.org/licenses/MIT

### Imports

### Command line options
- `-t, --trace` Print stack trace
- `-h, --help` Print command help

### Author and maintainer
Anna Shchiptsova, shchipts@iiasa.ac.at

---

### Name
regression-tests-1.0.0-standalone.jar

### Type
directory

### Summary
Library for spatial statistical analysis with resampling

### Version
1.0.0

### License
MIT, http://opensource.org/licenses/MIT

### Imports

### Command line options
- `-t, --trace` Print stack trace
- `-h, --help` Print command help

### Author and maintainer
Anna Shchiptsova, shchipts@iiasa.ac.at

---

### 3 Preprocessing of the GIS-based variables

#### 3.1 Background

Before statistical modeling, we clean and rescale the original GIS-based maps. At first, we exclude those cells, which either contain an undefined value or have undefined values in their Moore neighborhood (which comprises the eight cells surrounding a central cell on a two-dimensional square lattice) in any of the given maps in the dataset.

Next, we apply normalization to all GIS-based variables except the land use by bringing the cell values of an individual variable into the range \([0,1]\):
\[ x' = \frac{x - x_{min}}{x_{max} - x_{min}}, \]

where \( x \) is an original cell value, \( x_{min} \) and \( x_{max} \) denote the minimum and maximum values among all cell values of the variable on a regional GIS lattice.

After that, we drop cells, which fall into the masked areas in a "black list" GIS-based map.

Finally, we reclassify the original map with land use classification. By convention, original map uses the following classification scheme: 1- urban land use class (including all artificial surfaces) and 0 - non-urban land class (including vegetation, wetlands, agricultural land and water). We take a number of cells with urban land use in a cell neighborhood, including the cell itself, as a cell value in the land use map.

In order to represent location of administrative units with respect to one another, we create a \( n \times n \) matrix of spatial proximity \( W \) such that \( w_{ij} = 0 \), if administrative units \( i \) and \( j \) have no common boundary, i.e., they are not contiguous; otherwise, \( w_{ij} \) is inverse proportional to the total number of contiguous administrative units to the administrative unit \( i \). By convention, \( w_{ii} = 0 \) for any \( i = 1 \ldots n \).

### 3.2 Implementation in 'lu-preprocessing-1.0.0-standalone.jar'

#### Usage

$ java -jar lu-preprocessing-1.0.0-standalone.jar [options] settings-path

Arguments:

- settings-path Path to the file with settings

Options:

- -t, --trace Print stack trace
- -h, --help Print command help

#### Input

```xml
<?xml version="1.0"?>
<files>
  <raster path = "land-use.asc" group = "land-use" />
  <raster path = "sections.asc" group = "region" />
  <raster path = "zoning.asc" group = "mask" />
  <raster path = "density.asc" transform = "log" />
  <raster path = "distance_roads.asc" transform = "unit-rescaling" />
  <raster path = "distance_industrial_commercial.asc" transform = "unit-rescaling" />
  <raster path = "distance_airports.asc" transform = "unit-rescaling" />
</files>
```
As an input argument, the jar package receives an XML file with settings. Land use map is specified with 'land-use' value of the 'group' XML-attribute. The map should contain land use classification with values equal either 1 (including all artificial surfaces) or 0 (including vegetation, wetlands, agricultural land and water).

The 'region' value of the 'group' attribute defines a GIS-based map with regional administrative division. Only integer values are supported as identifiers of administrative units.

The 'mask' value of the 'group' attribute indicates a "black list" map with mask cell values. The mask cell value should be installed to 1.

It is expected that settings xml file includes one and only one raster XML element from the 'land-use' group, one and only one raster element from the 'region' group, one and only one raster element from the 'mask' group.

## Output

Results are saved to the 'lu-preprocessing' folder in the root execution directory.

Output includes cleaned and transformed .asc maps, files with Moore neighborhood statistics and a file with the matrix of spatial proximity (e.g., 'sections-neighbours.csv').
4 Upscaling the GIS-based variables

4.1 Background

The cleaned and transformed GIS-based maps are rescaled to the level of administrative subdivision. In particular, an average of cell values in an administrative unit is taken as a single value for the GIS-based variable.

4.2 Implementation in 'lu-rescaling-1.0.0-standalone.jar'

## Usage

```
$ java -jar lu-rescaling-1.0.0-standalone.jar [options] map-folder region-path
```

Arguments:
- map-folder: Path to the folder with original GIS-based maps
- region-path: Path to the file with administrative units

Options:
- -t, --trace: Print stack trace
- -h, --help: Print command help

## Input

'map-folder' contains .asc files. E.g.,
- "land_use.asc"
- "distance_roads.asc"
- "distance_industrial_commercial.asc"
- "distance_airports.asc"
- "distance_waterfront.asc"
- "distance_forest.asc"
- "distance_10ths_city.asc"
- "distance_50ths_city.asc"

The 'region-path' argument defines a file with administrative subdivision, e.g., 'sections.asc'.

## Output

Results are saved to the 'lu-preprocessing/sample.csv' file in the root execution directory.

<table>
<thead>
<tr>
<th>sections_density</th>
<th>distance_roads</th>
<th>distance_industrial_commercial</th>
<th>distance_airports</th>
<th>distance_waterfront</th>
<th>distance_forest</th>
<th>distance_10ths_city</th>
<th>distance_50ths_city</th>
<th>land_use</th>
</tr>
</thead>
<tbody>
<tr>
<td>920</td>
<td>7.157735</td>
<td>0</td>
<td>0</td>
<td>0.073039</td>
<td>0.313935</td>
<td>0.03761</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>558</td>
<td>6.326746</td>
<td>0</td>
<td>0.103834</td>
<td>0.259699</td>
<td>0.312151</td>
<td>0.032535</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>584</td>
<td>-1.97735</td>
<td>0.336714</td>
<td>0.384921</td>
<td>0.34229</td>
<td>0.024488</td>
<td>0.244743</td>
<td>0.154481</td>
<td>0.116898</td>
</tr>
<tr>
<td>487</td>
<td>5.305789</td>
<td>0.001239</td>
<td>0.351523</td>
<td>0.387961</td>
<td>0.164373</td>
<td>0.008174</td>
<td>0</td>
<td>0.184289</td>
</tr>
</tbody>
</table>

... ... ... ... ... ... ... ... ...
5 Regression analysis with resampling

5.1 Background

Let us consider a geographic region consisting of \( n \) administrative units. Suppose that we have panel data \((X, y)\) collected in every administrative unit of the region. Here, \( X \) is a \( n \times (p + 1) \) matrix of the explanatory variables and \( y \) is a \( n \times 1 \) observable vector of the response. Each column \( X^i \) consists of the sample observations on a single explanatory variable.

In general, we want to relate the response variable to available explanatory factors in an administrative unit based on the reported spatial panel data. For this purpose, we put forward a multiple regression model in the following form

\[
y = X\beta + \varepsilon
\]

\[
\varepsilon_1, \ldots, \varepsilon_n \sim F(0, \sigma^2)
\]

where \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^T \) is a \((p + 1) \times 1\) vector of the unknown model parameters to be estimated from the data using the ordinary least squares method. By assumption, \( X^1 \) is identically 1, so that the regression equation has an intercept \( \beta_0 \). For a GIS-based variable, an average value over cells belonging to the same administrative unit is considered to be an observation in this unit included in \( X \) (section 4). The error term \( \varepsilon \) is a \( n \times 1 \) vector of independent identically distributed errors with common distribution \( F \) having mean 0 and finite variance \( \sigma^2 \). Both \( F \) and \( \sigma^2 \) are unknown.

Hypothesis testing in model (1) is conducted using the method of permutations (Fisher, 1935). We run a permutation test with the test statistic \( R^2 \) to assess the overall model significance. The significance of each individual coefficient is tested using the Freedman and Lane procedure (Freedman and Lane, 1983).

Once the statistically significant combination of the explanatory variables is identified with the permutation tests, we assume that model (1) is a correct one; it is a more likely explanation of the data than the random assignment. Next, we use the percentile bootstrap method (Efron and Tibshirani, 1993) for parameter estimation. The bootstrap estimates are computed for \( \beta, R^2 \) and MSE (mean square error) parameters of the regression model. Further details on resampling procedures can be found in the "Permutation Hypothesis Testing and Bootstrapping in Regression Model" guide (Shchiptsova, 2016).

The method of permutations does not avoid the assumption on the observations to be independent and identically distributed. Additionally, we can verify the absence of spatial autocorrelation in the error terms of model (1); that is, we check whether the errors are determined and assigned to the neighboring administrative units independently and at random.
Details of this procedure using 'regression-tests-1.0.0-standalone.jar' package can be found in the "Testing Spatial Autocorrelation with the Bootstrap" guide (Shchiptsova, 2016).

5.2 Implementation in 'lu-regression-1.0.0-standalone.jar'

## Usage

```
$ java -jar lu-regression-1.0.0-standalone.jar [options] path n-rep
```

Arguments:

- `path` Path to the csv file with an original sample
- `n-rep` Number of permutations and bootstrap replications

Options:

- `-t, --trace` Print stack trace
- `-h, --help` Print command help

## Input

;;;;; 'sample-x1-x2-x3.csv'

<table>
<thead>
<tr>
<th>density.asc</th>
<th>land_use.asc</th>
<th>distance_roads.asc</th>
<th>distance_industrial_commercial.asc</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.157735</td>
<td>9</td>
<td>0</td>
<td>0.03761</td>
</tr>
<tr>
<td>6.326746</td>
<td>9</td>
<td>0</td>
<td>0.032535</td>
</tr>
<tr>
<td>-1.977347</td>
<td>0.019694</td>
<td>0.154481</td>
<td>0.244743</td>
</tr>
<tr>
<td>5.305789</td>
<td>6.25</td>
<td>0</td>
<td>0.008174</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The 'path' argument defines a file with sample values, e.g., 'sample-x1-x2-x3.csv'. It is expected that the first row contains variable labels. The first column should contain values of the response $y$.

## Output

Results are saved to the 'lu-regression' folder in the root execution directory. The 'lu-regression/permutation_tests.csv' file contains results of permutation testing. Bootstrap estimates are saved to 'lu-regression/regression-stat-bootstrap.csv'.

;;;;; 'permutation_tests.csv'

<table>
<thead>
<tr>
<th>test</th>
<th>p-value</th>
<th>lower-bound-ci</th>
<th>upper-bound-ci</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall-test-r2</td>
<td>0.0001</td>
<td>-0.000096</td>
<td>0.000296</td>
</tr>
<tr>
<td>land_use.asc-test-t-stat</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>distance_roads.asc-test-t-stat</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>distance_industrial_commercial.asc-test-t-stat</td>
<td>0.007699</td>
<td>0.005986</td>
<td>0.009412</td>
</tr>
</tbody>
</table>
6 Approximation to the GIS lattice

6.1 Background

We apply bootstrapping to estimate the accuracy of model approximation to the values on a GIS lattice. Suppose that we have data \((X', y')\), where \(X'\) is a \(N \times (p + 1)\) matrix of the explanatory variables and \(y'\) is a \(N \times 1\) vector of the response at the cell level. \(N\) is a total number of cells on a GIS lattice. For the given cell \(j\), we define accuracy \(\rho_j\) of the response value from model (1) to the true value \(y'_j\), observed in this cell, as a distance between the expected model response and the true value. That is

\[
\rho_j = |y'_j - \hat{y}'_j|,
\]

where \(\hat{y}'_j = X'_j \beta\) is the fitted value of the response in model (1) for the observed values \(X'_j = (x'_j1, ..., x'_jp)\) of explanatory variables in cell \(j\) \((j = 1, ..., N)\).

For the given subset of cells \(S\), we measure the \((100 \times k)\)-th percentile of the accuracy of the response values from model (1) to the true values \(\{y'_j\}_{j \in S}\), observed in the cells belonging to this subset, as a minimum value below which at least the \(k\)-th fraction of the cell accuracy values fall, and denote it by \(\rho(k, S)\). To summarize the sample of accuracy values in subset \(S\), we examine maximum accuracy value \(\rho(1, S)\) and the sample quartiles, which coincide with the values of \(\rho(0.25, S)\), \(\rho(0.5, S)\) and \(\rho(0.75, S)\). By convention, these descriptive statistics are denoted as \(\rho_{\text{max}}(S), \rho(Q_1,S), \rho(Q_2,S)\) and \(\rho(Q_3,S)\) respectively.

Since distribution \(F\) is unknown in model (1), we bootstrap data \((y, X)\) to get the reference distribution for an accuracy statistic following the percentile bootstrap scheme (section 5). Specifically, we calculate accuracy values for an estimated vector of coefficients \(\hat{\beta}_y\) and summarize the sample by the selected percentiles in every bootstrap replication \(\gamma\). Thus, we obtain separate bootstrap samples for \(\rho_{\text{max}}(S), \rho(Q_1,S), \rho(Q_2,S)\) and \(\rho(Q_3,S)\), and find \(100\times \alpha\)-% percentile confidence intervals for these statistics.
## 6.2 Implementation in 'lu-approximation-1.0.0-standalone.jar'

### Usage

```
$ java -jar lu-approximation-1.0.0-standalone.jar [options] sample-path values-path n-rep
```

Arguments:
- **sample-path**: Path to the csv file with an original sample
- **values-path**: Path to the csv file with target values for approximation
- **n-rep**: Number of bootstrap replications

Options:
- `-t, --trace` Print stack trace
- `-h, --help` Print command help

### Input

```
;;;;; 'sample-x1-x2-x3.csv'

<table>
<thead>
<tr>
<th>density.asc</th>
<th>land_use.asc</th>
<th>distance_roads</th>
<th>distance_indus</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.157735</td>
<td>6.326746</td>
<td>-1.977347</td>
<td>5.305789</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>0.019694</td>
<td>6.25</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.154481</td>
<td>0.008174</td>
</tr>
<tr>
<td>0.03761</td>
<td>0.032535</td>
<td>0.244743</td>
<td></td>
</tr>
</tbody>
</table>
```

```
;;;;; 'cells-x1-x2-x3.csv'

<table>
<thead>
<tr>
<th>density.asc</th>
<th>group</th>
<th>land_use.asc</th>
<th>distance</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.69062</td>
<td>0</td>
<td>0</td>
<td>0.406045</td>
<td>0.550763</td>
</tr>
<tr>
<td>-1.69062</td>
<td>0</td>
<td>0</td>
<td>0.416017</td>
<td>0.555664</td>
</tr>
<tr>
<td>-1.69062</td>
<td>0</td>
<td>0</td>
<td>0.42604</td>
<td>0.560662</td>
</tr>
<tr>
<td>-1.69062</td>
<td>0</td>
<td>0</td>
<td>0.436063</td>
<td>0.565755</td>
</tr>
</tbody>
</table>
```

The 'sample-path' argument defines a file with the sample values in administrative units, e.g., 'sample-x1-x2-x3.csv'. It is expected that the first row contains variable labels. The first column should contain values of the response $y$.

The 'values-path' argument defines a file with the target cell values, e.g., 'cells-x1-x2-x3.csv'. It is expected that the first row contains variable labels. The first column should contain values of the response $y$. The second column contains the group id for the given cell value.

### Output

Results are saved to the 'lu-approximation/accuracy-bootstrap.csv' file in the root execution directory.
<table>
<thead>
<tr>
<th>id</th>
<th>group-id</th>
<th>95-percent-ci-1</th>
<th>95-percent-ci-2</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-25-percent-0</td>
<td>0</td>
<td>0.91972</td>
<td>1.154213</td>
<td>1.034962</td>
</tr>
<tr>
<td>p-25-percent-1</td>
<td>1</td>
<td>1.648701</td>
<td>1.834848</td>
<td>1.740372</td>
</tr>
<tr>
<td>p-25-percent-all</td>
<td>all</td>
<td>0.949766</td>
<td>1.181227</td>
<td>1.063125</td>
</tr>
<tr>
<td>p-50-percent-0</td>
<td>0</td>
<td>1.802652</td>
<td>2.270311</td>
<td>2.02227</td>
</tr>
<tr>
<td>p-50-percent-1</td>
<td>1</td>
<td>3.189866</td>
<td>3.481749</td>
<td>3.337134</td>
</tr>
<tr>
<td>p-50-percent-all</td>
<td>all</td>
<td>1.875895</td>
<td>2.342561</td>
<td>2.094395</td>
</tr>
<tr>
<td>p-75-percent-0</td>
<td>0</td>
<td>2.844491</td>
<td>3.490305</td>
<td>3.134313</td>
</tr>
<tr>
<td>p-75-percent-1</td>
<td>1</td>
<td>4.402656</td>
<td>4.712864</td>
<td>4.555963</td>
</tr>
<tr>
<td>p-75-percent-all</td>
<td>all</td>
<td>2.976079</td>
<td>3.602863</td>
<td>3.256569</td>
</tr>
<tr>
<td>p-max-0</td>
<td>0</td>
<td>11.925989</td>
<td>20.971695</td>
<td>15.92601</td>
</tr>
<tr>
<td>p-max-1</td>
<td>1</td>
<td>9.958119</td>
<td>14.244063</td>
<td>10.7058</td>
</tr>
<tr>
<td>p-max-all</td>
<td>all</td>
<td>11.925989</td>
<td>20.971695</td>
<td>15.92626</td>
</tr>
<tr>
<td>p-min-0</td>
<td>0</td>
<td>0.000001</td>
<td>0.000076</td>
<td>0.00002</td>
</tr>
<tr>
<td>p-min-1</td>
<td>1</td>
<td>0.000009</td>
<td>0.001301</td>
<td>0.000361</td>
</tr>
<tr>
<td>p-min-all</td>
<td>all</td>
<td>0</td>
<td>0.000072</td>
<td>0.000019</td>
</tr>
</tbody>
</table>
7 Example Data

Table 1 includes description of source files, which are used in examples.

<table>
<thead>
<tr>
<th>File reference</th>
<th>Type of source data</th>
<th>Description</th>
<th>Year</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>land-use.asc</td>
<td>GIS map</td>
<td>Land use classification: urban (e.g., all artificial surfaces) and non-urban (e.g., vegetation, wetlands, agricultural land and water)</td>
<td>2003</td>
<td>REDIAM, 2015</td>
</tr>
<tr>
<td>sections.asc</td>
<td>Panel data</td>
<td>Administrative division of the Province of Seville (sections)</td>
<td>2003</td>
<td>REDIAM, 2015</td>
</tr>
<tr>
<td>zoning.asc</td>
<td>GIS map</td>
<td>protected natural areas</td>
<td>2015</td>
<td>REDIAM, 2015</td>
</tr>
<tr>
<td>density.asc</td>
<td>Panel data</td>
<td>population density (people per cell)</td>
<td>2001</td>
<td>INE, 2015</td>
</tr>
<tr>
<td>distance_roads.asc</td>
<td>GIS map</td>
<td>distance to the nearest road (km)</td>
<td>2005</td>
<td>CNIG, 2015</td>
</tr>
<tr>
<td>distance_industrial_commercial.asc</td>
<td>GIS map</td>
<td>distance to the nearest area of commercial or industrial land use (km)</td>
<td>2006</td>
<td>EEA, 2015</td>
</tr>
<tr>
<td>distance_airports.asc</td>
<td>GIS map</td>
<td>distance to the nearest airport (km)</td>
<td>2006</td>
<td>EEA, 2015</td>
</tr>
<tr>
<td>distance_waterfront.asc</td>
<td>GIS map</td>
<td>distance to the nearest waterfront (km)</td>
<td>2005</td>
<td>CNIG, 2015</td>
</tr>
<tr>
<td>distance_forest.asc</td>
<td>GIS map</td>
<td>distance to the nearest area of forest (km)</td>
<td>2006</td>
<td>EEA, 2015</td>
</tr>
<tr>
<td>distance_10ths_city.asc</td>
<td>GIS map</td>
<td>proximity to a city center with more than 10,000 inhabitants (km)</td>
<td>2011</td>
<td>INE, 2015</td>
</tr>
<tr>
<td>distance_50ths_city.asc</td>
<td>GIS map</td>
<td>proximity to a city center with more than 50,000 inhabitants (km)</td>
<td>2011</td>
<td>INE, 2015</td>
</tr>
</tbody>
</table>
Acknowledgments

The author would like to acknowledge DG research for funding through the FP7-funded COMPLEX project #308601, www.complex.ac.uk.

Views or opinions expressed herein do not necessarily represent those of the International Institute for Applied Systems Analysis, its National Member Organizations, or other organizations supporting the work.

References


Web References


