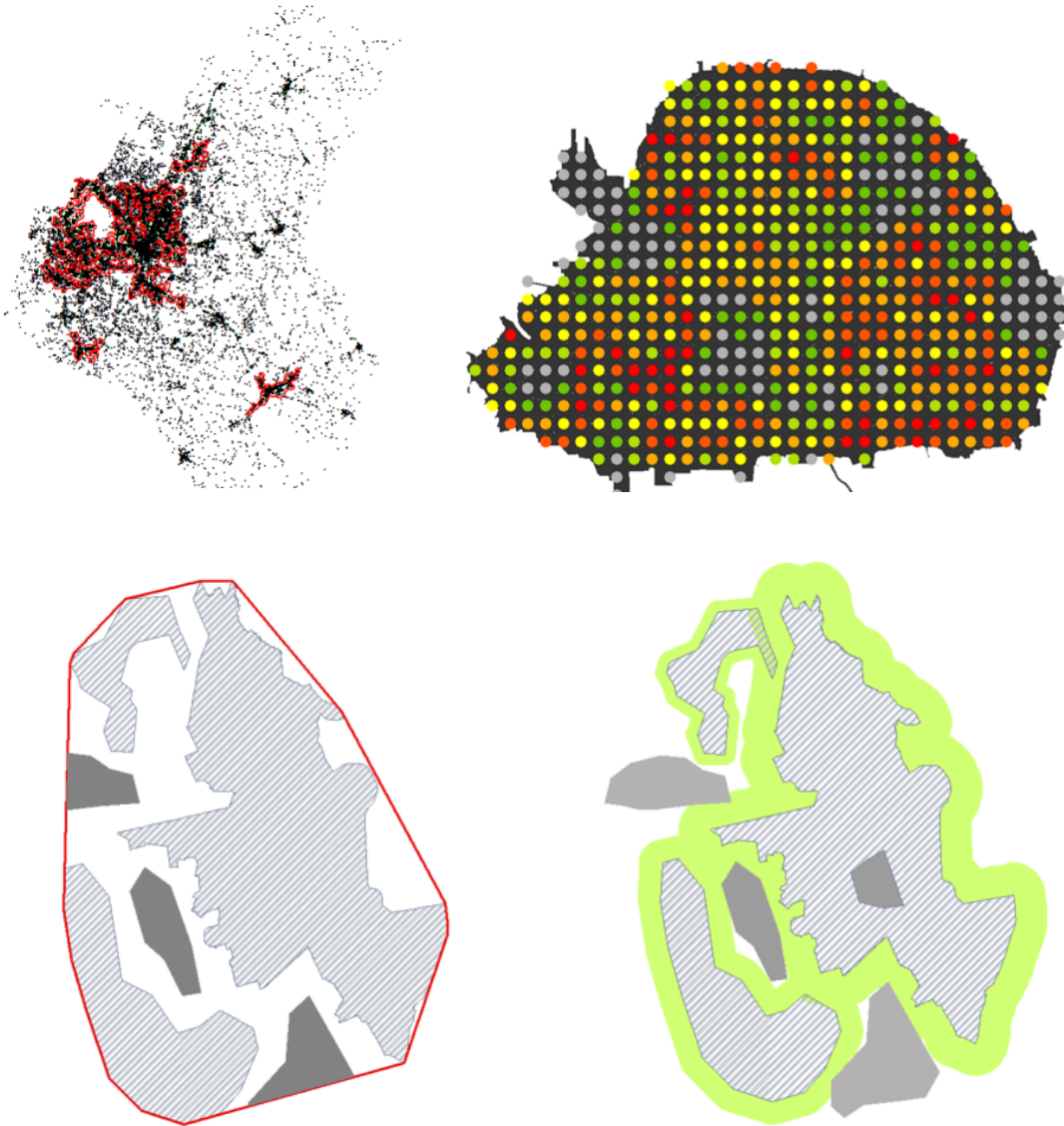# Metropolitan Form Analysis Toolbox

for ArcGIS v10.2 and v10.3.1

20150913

LICENSE

1. INTRODUCTION

Metropolitan Form Analysis (MFA) toolbox automates the computation of a series of metrics for describing the spatial footprint and land-use patterns of cities: Coverage, Compactness, Discontiguity, Land-use Mix, Polycentricity, and Expandability. The detailed specification of these metrics can be found in a paper that was presented in the Urban Affair Association's 2013 Conference in San Francisco, and in a preceding World Bank report.

2. INSTALLATION GUIDELINES

2.1. Download

Download the zip folder (*Toolbox.zip*) to a location of your choice on your computer and unzip its contents. You will find a toolbox file (*MFA.tbx*), six python scripts (.py), and four ArcGIS layer files (.lyr). The Metropolitan Form Analysis toolbox is distributed and can be used under the Creative Commons Licensee Attribution 4.

2.2. Installation

The Metropolitan Form Analysis toolbox doesn't require installation. As long as the python scripts, layer files and the toolbox file are in the same directory, you can use the MFA toolbox through the ArcCatalog window in ArcMap:

- In ArcMap 10.2 or 10.3.1, open Catalog window
- In Catalog window browse MFA.tbx, and expand the MFA.tbx folder.
- Under MFA.tbx, you will find six scripts
- Double click on any of the scripts you wish to use

Alternatively, you can add the Metropolitan Form Analysis toolbox to ArcToolbox and open the scripts from ArcToolbox:

- In ArcMap 10.2 or 10.3.1, open ArcToolbox
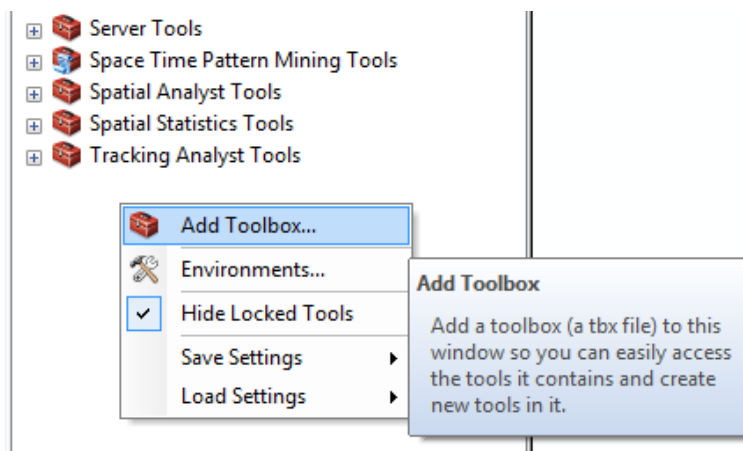- Right click within the ArcToolbox window and select "add toolbox"

Figure 1: Adding MFA toolbox to ArcToolbox

- Browse to MFA.tbx and add it to ArcToolbox
- Right click within the ArcToolbox window, and select "Save Setting" and after that "To Default"

3. COMPACTNESS TOOL

3.1 Specification

Compactness measure the degree to which the a geographic phenomena, such population, jobs, businesses, or buildings are spatially spread out. The proposed compactness metric borrows an analogous measure for transportation literature, "Gravity" (Hansen, 1959):

$$G_i = \sum_{j \in G - \{i\}} \frac{W[j]}{e^{\beta \cdot d[i,j]}}$$

where Gi is the gravity index for location i, W[j] is the weight of destination j, and d[i,j] is the distance between locations i and j. Beta is called decay rate, which controls the how fast the attraction of destination j drops when the distance between i and j increases.

MFA's compactness tool computes the gravity index for each feature in the dataset – point or polygon – and reports the weight average of gravity indices as the raw compactness for the entire study area. In order to make the compactness index comparable across cities of different size and scale, the compactness tool offers four normalization options (see the normalization parameter below).

3.2 Input and output datasets

Main input dataset: a point or polygon feature class – either in a shapefile (shp) or a geodatabase (gdb) – that stores desired information (in a numeric field) about the geographic phenomenon of interest; business locations and building footprint centroids are examples of point input datasets, while building footprint, census tracts and census blocks are examples of polygon input datasets for MFA's compactness tool.

Optional/special case input dataset: one or multiple polygon feature class(es) containing unbuildable areas used to constrain the reference case for normalization (see normalization options below for more details) ; unbuildable areas are geographical extent that cannot be developed: for example, bodies of water, wetlands, steep land , or persevered lands. Unbuildable area polygon feature class(es) can be either in a

geodatabase (gdb) or a shapefile(shp). Unbuildable area input dataset(es) are required only if the normalization option with geographic constraints is selected.

Output datasets: a normalization reference case polygon feature class and its equivalent point representation -- stored in a point feature classes if normalization by reference case is selected by the user (see below the resolution parameter for more detail). If both modes normalization by reference case is selected, there will be one polygon and one point feature class output for each mode. Polygon and point feature class output will both be stored in the output directory, while polygon feature class(es) will be also automatically added to the active data frame in ArcMap.
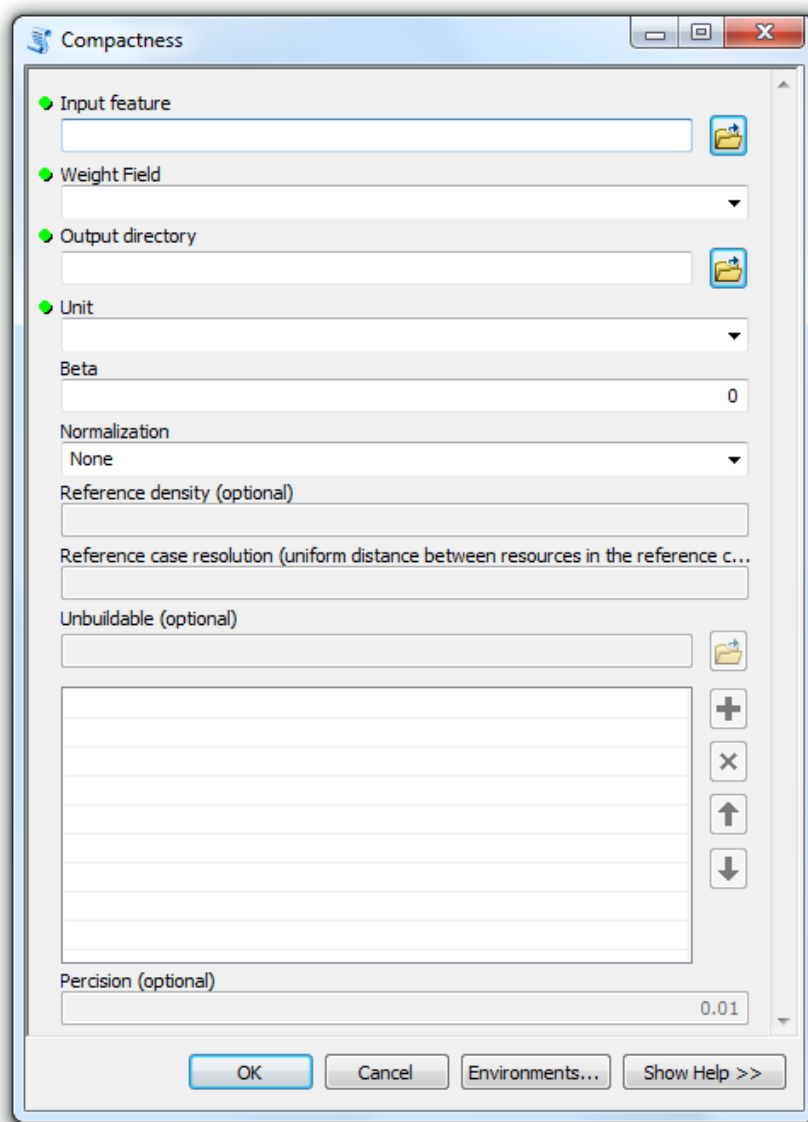
3.3.Setting Parameters

Figure 2: Graphic user interface of the compactness tool

Weight field: a float, double or integer field in the main input dataset that stores values of a property of interest, such as number of people, number of employees, building footprint area or building floor area.

Output directory: a directory where output dataset or intermediate files will be stored. Output directory can be a folder or a geodatabase.

Unit: the unit used for measuring distances in gravity metric computation (see the specification of compactness metric for more detail). It is also used as the unit for resolution parameter. It also determines the unit for the reference case density parameter; for example if unit is specified as *kilometer*, the unit of the reference case density parameter will be in weight (from weight field) per square *kilometer*.

Beta: is the decay rate in the gravity metric (see the specification of compactness metric for more detail).

Normalization: is used to make computed raw compactness index comparable across cities of different size and scale. There four options for normalization:

- *None*: Compactness tool returns raw compactness value and compactness normalized (i.e. divided) by the total weight (e.g. total population) in the input dataset

- *Without Geographic Constraints:* Compactness tool generates a circular shape reference case with the exact same total weight (e.g. total population) as the study area and a uniform density passed in by the user – its area is determined by dividing total weigh over density. It then computes the raw compactness in the reference case and reports the ratio between the raw compactness in the study area and the reference case.

- *With Geographic Constraints:* Compactness tool generates a reference case, constrained by unbuildable areas, which has the same geometric center and the exact same total weight (e.g. total population) as the study area and a uniform density passed in by the user. Since such polygon feature cannot be derived analytically in most cases, it is generated through an approximation iterative

process. It then computes the raw compactness in the reference case and reports the ratio between the raw compactness in the study area and the reference case.

To generate the reference case with geographic constraints, MFA's compactness tool first generates a circular shape polygon feature centered at the geometric center of the study area, area of which is derived by dividing total weigh (e.g. total population) in the study area over the reference density given by the user. Then it iteratively removes the unbuildable areas and expands the resulted polygon feature using the buffer tool to compensate for removed areas. The iteration of removing unbuildable areas and expanding the polygon feature continues until the ratio between the target area (total weigh over references density) and the area of resulted polygon feature reaches to an acceptable range from 1 – see precision parameter.

- *Both:* reports compactness normalized by total weight and both types of reference case – with and without geographic constraints.

Reference density: used to generate the reference case – see normalization parameter above.

Reference case resolution: used to convert the reference polygon feature to a set of points, used in computing raw compactness for the reference case. It determines the distance between points, which are uniformly distributed over the extended of the reference case. The fishnet tool is used to create the set of points, and the resolution parameter determines the cell size for the fishnet tool.

Precision: is required if the normalization with geographic constraint is selected by the user. It determines how far the ratio between the target area (total weigh over references density) and the area of the reference case polygon feature can be from 1 – see the normalization parameter above. Note that this two areas ideally should be equal.

4. DISCONTIGUITY TOOL

4.1. Specification

The proposed discontiguity metric takes into account the at the rank order and relative size difference between discontiguous urban clusters. MFA's discontiguity tool computes for

each polygon feature in the input dataset the ratio between its area and area of polygons that are smaller:

$$\sum_{n=1}^{N} \left(\frac{\sum_{i=n+1}^{N} A_i}{A_n} \right)$$

where N is the number of urbanized clusters, and A_n the area of cluster n. Note that $A_n \geq A_{n+1}$. It then reports the weighted average of these values as the discontiguity index of the input dataset:

$$DC = \frac{\sum_{n=1}^{N} \left(\frac{\sum_{i=n+1}^{N} A_i}{A_n} \cdot A_n \right)}{A_{total}}$$

which can be simplified to:

$$DC = \sum_{n=1}^{N} \left(\frac{\sum_{i=n+1}^{N} A_i}{A_{total}} \right)$$

Important note about the specification of discontiguity metric: the specification that MFA's discontiguity tool uses, is slightly different from the specification that was presented in the conference paper and the World Bank report, the following equation:

$$\sum_{n=1}^{N} \left(\frac{\sum_{i=n+1}^{N} A_i}{A_n} \right) \left(\frac{\sum_{i=n}^{N} A_i}{A_{total}} \right)$$

That only difference is in how the weighted average is calculated. The earlier form of the specification uses $\sum_{i=n}^{N} A_i$ as the weight of in the weighted average, which is for each feature the sum of its area and areas of smaller features. In the current specification that is used in the MFA toolbox, simply A_n , is used as the weight.
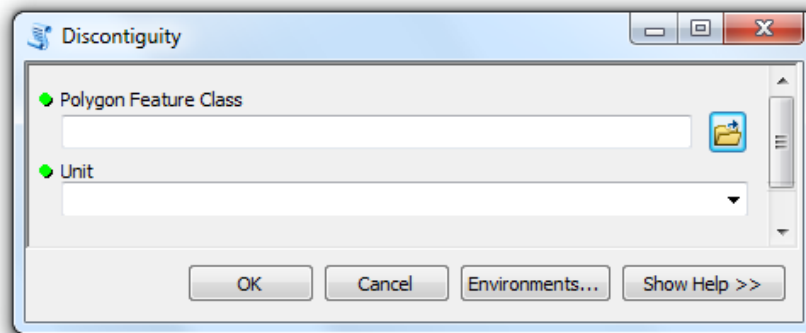

Figure 3: Graphic user interface of the discontiguity tool

4.2. Input and output datasets

Main input dataset: a polygon feature class of the urban land coverage in the study area, which can be either a shapefile (shp) or in geodatabase (gdb).

Output datasets: NA; the discontiguity tool just reports a discontiguity index and does not generate any output dataset.

4.3. Setting Parameters

Unit: it's used in the calculation of the area of polygon features. Discontiguity index will, however, be the same regardless of unit selected.

5. COVERAGE

5.1. Specification

The proposed coverage metric measures what percent of the buildable area within the convex hull around the existing built-up area, has already been developed. MFA's coverage tool first generates the convex hull around the entire built-up area, and then finds the buildable area within this convex by removing unbuildable areas. It finally reports the ratio of the built-up area to buildable area of the convex hull.
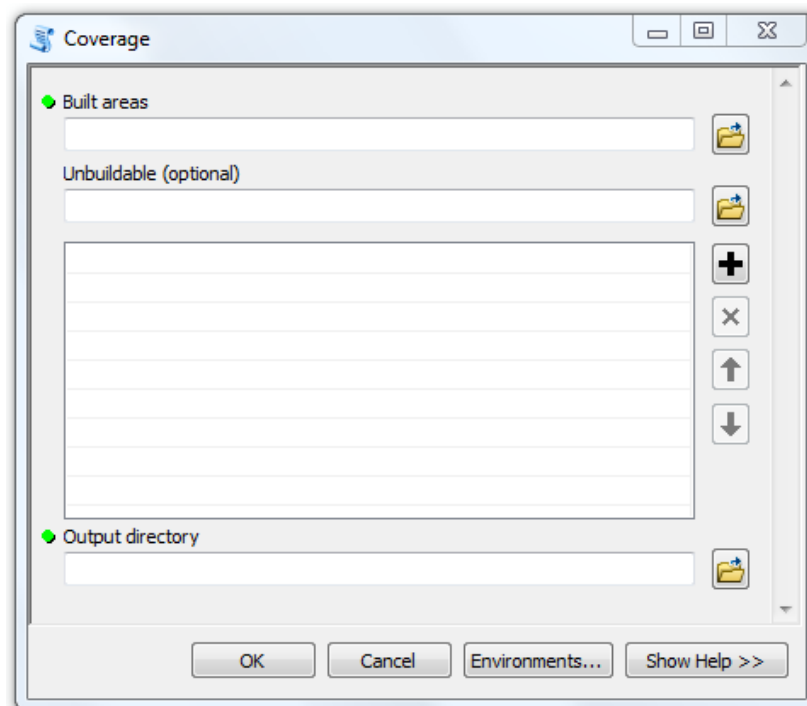

Figure 4: Graphic user interface of the coverage tool

5.2. Input and output datasets

Main input dataset: a polygon feature class of the urban land coverage in the study area, which can be either a shapefile (shp) or in geodatabase (gdb).

Optional input dataset: one or multiple polygon feature class(es) containing unbuildable areas; unbuildable areas are geographical extent that cannot be developed: for example, bodies of water, wetlands, steep land , or persevered lands. Unbuildable area polygon feature class(es) can be either in a geodatabase (gdb) or a shapefile(shp).

Output datasets: a polygon feature class containing the convex hull around the built-up areas (urban coverage), and a polygon feature class containing the unbuildable areas within the convex hull. Both output datasets will be automatically added to the active dataframe in ArcMap.

5.3. Setting Parameters

Output directory: a directory where output dataset or intermediate files will be stored. Output directory can be a folder or a geodatabase.

6. EXPANDABILITY

6.1 Specification

Expandability metric measures what percent of a buffer around the existing built-up area that has the same area as the existing built-up area, is buildable. MFA's expandability tool generates such buffer polygons for each feature in the input dataset and then by removing unbuildable areas, it finds the buildable area within the buffer. Since such buffer can be determined analytically in most case, expandability tool uses an iterative approximation method (see precision parameter below).
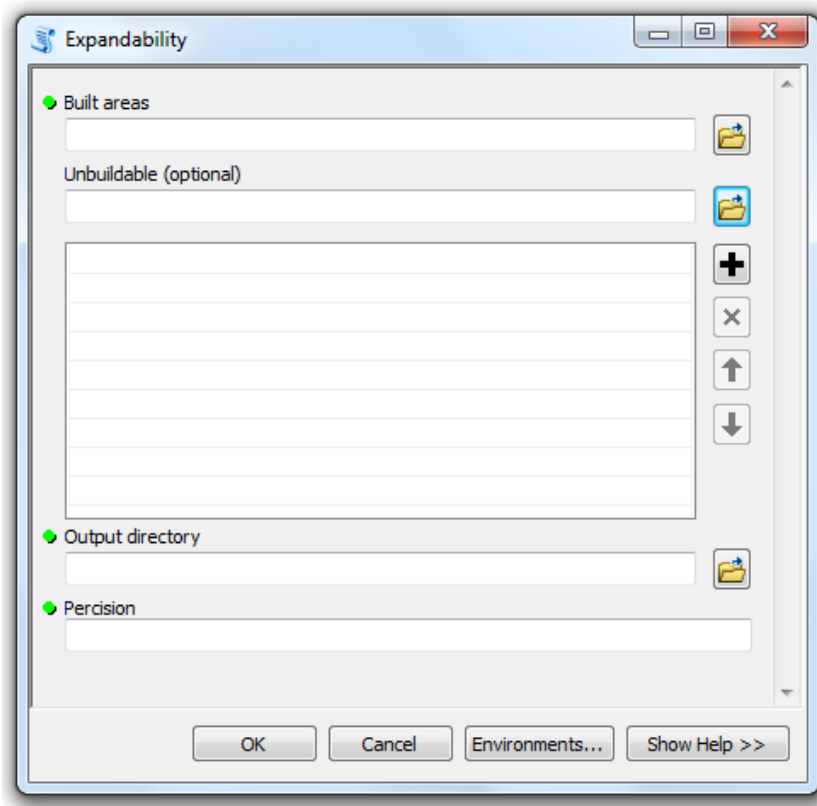
Figure 5: Graphic user interface of the expandability tool

6.2. Input and output datasets

Main input dataset (built-up areas): a polygon feature class of the urban land coverage in the study area, which can be either a shapefile (shp) or in geodatabase (gdb).

Optional input dataset (unbuildable areas): one or multiple polygon feature class(es) containing unbuildable areas; unbuildable areas are geographical extent that cannot be developed: for example, bodies of water, wetlands, steep land , or persevered lands. Unbuildable area polygon feature class(es) can be either in a geodatabase (gdb) or a shapefile(shp).

Output datasets: a polygon feature class that contains the 100% buffer around the built-up area – a buffer that has the same area as the current built-up area – and a polygon feature class that contains only the buildable areas of this buffer. Both output datasets will be stored in the output directory. The latter will be automatically added to the active dataframe in ArcMap.

6.3.Setting Parameters

Output directory: a directory where output dataset or intermediate files will be stored. Output directory can be a folder or a geodatabase.

Precision: MFA's expandability tool creates a polygon feature class containing a buffer around each built-up area polygon feature and then computes what percentage of this buffer is buildable. Since such polygon features cannot be analytically derived for most instances, they are generated through an iterative approximation process. The precision parameter specifies how far the ratio between the area of generated polygon features and their expected area – the same as the area of the existing built-up area – can be from 1.

7. POLYCENTRICITY

7.1. Specification

MFA's polycentricity tool, before computing the polycentricity index for the input dataset, identifies centers based on a series of predefined criteria (see the conference paper or the World Bank's report for more information):

1) A center's density – of the property of interest – has to be higher than two standard deviations from the city-wide mean density (**important:** in the earlier version of the MFA toolbox released before September 10, this criteria was defined as *a center's density has to be higher than city-wide mean density*)

2) In case of polygon input datasets, if several adjacent polygons pass the above cut-off threshold, then these neighboring polygons are grouped together.

3) A center must contain 10/√population% or more of all jobs in a city.

Important: Since what a center is a concept relative to the context of and geographic phenomena polycentricity of which is measure, centers identified based on the above criteria may not necessarily be meaningful to certain analysis. The above criteria are calibrated for population and job densities. If wished, these criteria can be modified in the polycentricity.py in lines 43, 76 and 126.

Based on identified centers, the polycentricity index is computed as follows:

$PC = HI \times N \times R_c$

where *PC* is the polycentricity index, *HI* the homogeneity index, *N* the number of centers, and *R_c* the ratio of the total weight (e.g. total jobs) found in identified centers to the total weight in the study area. *HI* measures the degree to which the sizes of centers are homogenous. *HI* is defined based on Limtanakool et al's (2009) Entropy Index:

$$HI = -\sum_{i=1}^{L} \frac{(Z_i) \ln (Z_i)}{\ln (N)}$$

, where Z_i is the ratio of the total weight in center i to the total weight of all centers. HI is constrained between 0 and 1; it is 1 if the total weight is equally distributed among identified centers and tends to zero if a center's share of the total weight tends to 100%. The index is undefined if all jobs are in a single center as the denominator would be zero.
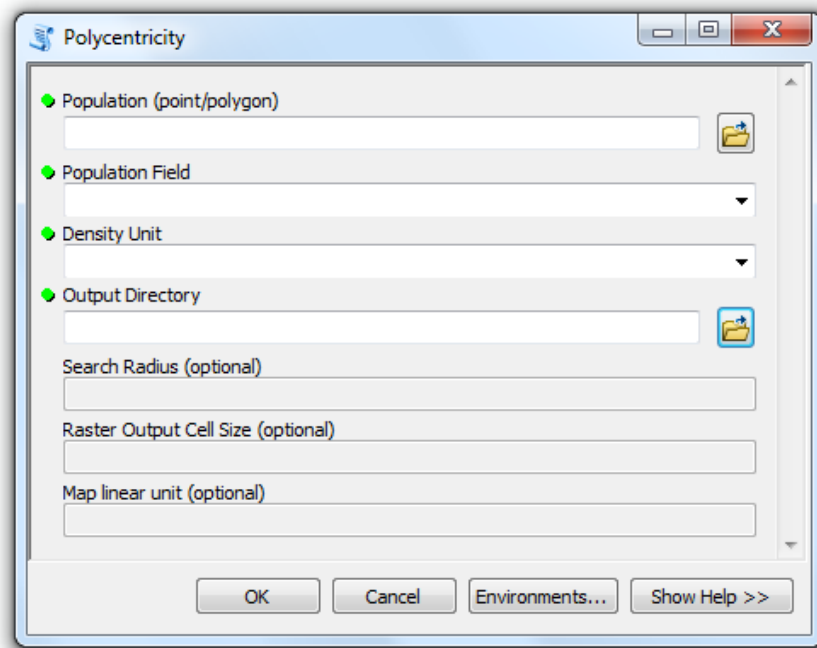

Figure 6: Graphic user interface of the polycentricity tool

7.2. Input and output datasets

Main input dataset: a point or polygon feature class – either in a shapefile (shp) or a geodatabase (gdb) – that stores desired information (in a numeric field) about the geographic phenomenon of interest; business locations and building footprint centroids are examples of point input datasets, while census tracts and census blocks are examples of polygon input datasets for MFA's polycentricity tool.

Important note on polygon input datasets: although this tool can carry out polycentricity computation for any polygon feature class that has at least one numeric field, without throwing errors, it should be noted that polygon features should cover the study area thoroughly – as for example census tracts do, otherwise the results may not be meaningful

or accurate. For datasets such as buildings footprint polygon datasets, which obviously do not cover the entire study area, their centroid can be used instead. If a building footprint dataset is used as input, either no center will be detected – since most likely no individual building has enough weight to pass the criteria as a center – or individual building will be detected as centers.

Optional/special case input dataset (Analysis area): a polygon feature class of the analysis area, which could be passed in by the user only if the main input dataset is a point feature class. If no polygon feature class is passed-in, the polycentricity tool, uses the convex hull around the point input dataset as the analysis area. In computing the mean and standard deviation of cells values over the kernel density raster, only cells that are within the analysis area, will be taken into account.

Output datasets: a polygon feature class containing detected centers, and if the input dataset is a point feature class, a raster kernel density dataset.

7.3. Setting Parameters

Population (weight) field: a float, double or integer field in the main input dataset that stores values of a property of interest, such as number of people, number of employees, building footprint area or building floor area.

Density unit: the unit in which density -- or kernel density in case of point input datasets – are calculated.

Output directory: a directory where output dataset or intermediate files will be stored. Output directory can be a folder or a geodatabase.

Search radius: is required if the input dataset is a point feature class and is used by the kernel density tool to create a raster density dataset from which centers will be derived.

Raster output cell size: is required if the input dataset is a point feature class and determines the cell size for the raster dataset that is generated by the kernel density tool.

8. LAND-USE MIX

8.1. Specification

The land-use mix tool generates a set of sample points that uniformly covers the entire study area, then computes for each location (sample point) a land-use mix metric and

reports the average of land-use indices of all locations as the land-use mix index of the entire study area. The density of sample points – that how close to each other they are located – is derived from the parameter for area of unit of analysis, which is specified by the user.

For each location (sample point), the proposed land-use mix metric measures how closely the distribution of observed uses in a given buffer around that location corresponds to an expected distribution, which is either passed in by the user or derived from city-wide distribution of uses (see reference case parameters below for more detail). The area of the buffer is specified by the user, e.g. as 1 sqkm. In order to avoid making the index vulnerable to small shifts in the boundaries of analysis units, the units of analysis overlap with each other, as shown in Figure 7 below.
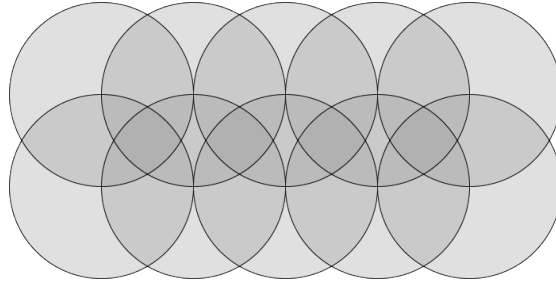

Figure 7: buffer areas overlap with each other, such that the orthogonal distances between centers equals the radius of a circle

Land-use mix index for each location i is defined as follows:

$$MX_i = S_i \cdot \prod_{n=1}^{N} M_{n:i}$$

in which S_i is the observed share of all land uses of interest from the buffer area around i. $M_{n:i}$ is the *matching index* for land-use *n* around location *i*, which shows how closely the observed share of land use *n* in all land uses of interest around location *i* matches its expected share, and is measured as:

$$M_{n:i} = 1 - |S_{n:i} - ES_{n:i}|$$

$S_{n:i}$ is the observed share of land use *n* in all land uses of interest around location *i*, and $ES_{n:i}$ is its expected share that is either passed in by the user or derived based on observed city-wide distribution of uses (see the reference type parameter for more detail on how the
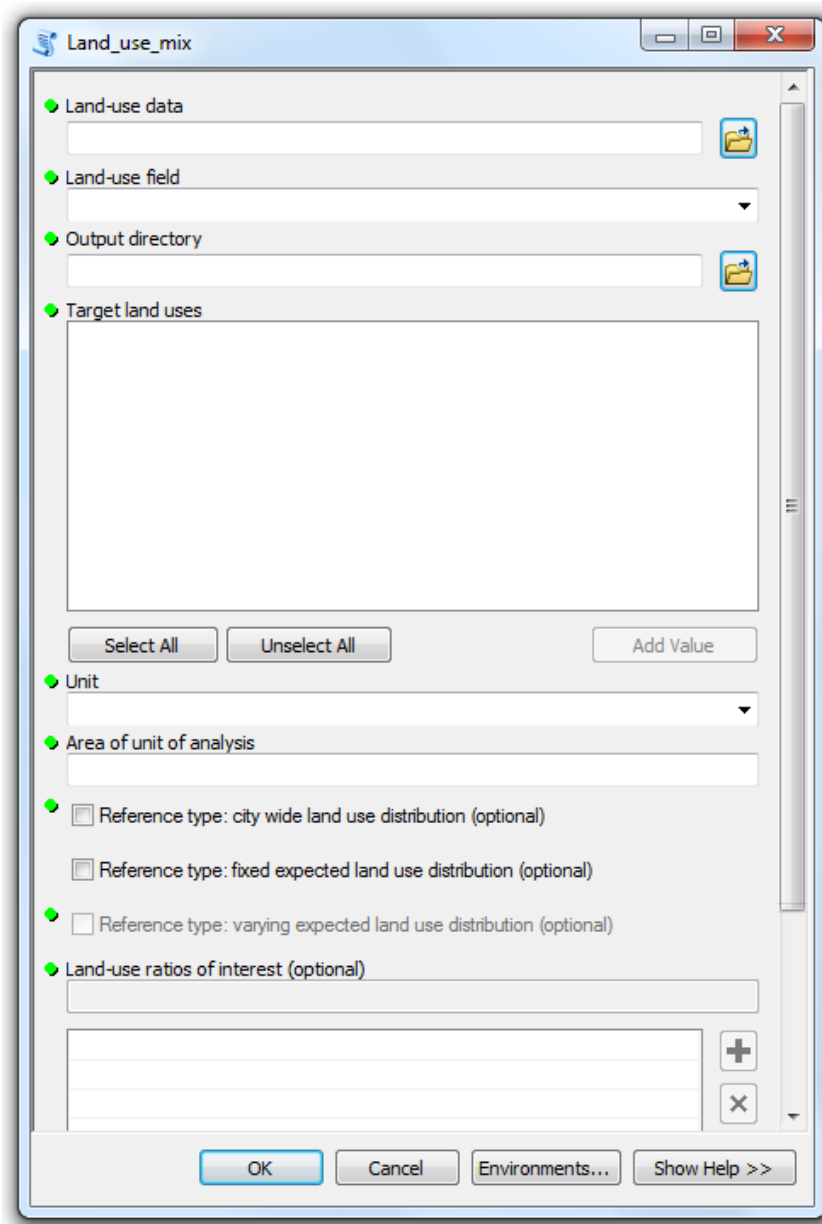
expected share could be specified).

Figure 8: Graphic user interface of the land-use mix tool

8.2. Input and output datasets

Main input dataset (land-use data): a polygon feature class – either in a shapefile (shp) or a geodatabase (gdb) – containing the land-use type of each polygon feature in one of its fields. Parcel/property datasets often contain land-use data.

Output dataset: a point feature class covering the study area extent, which contains computed land-use mix values for each samples.

8.3. Setting Parameters

Land-use field: a field in the land-use dataset that stores the land-use type of each polygon feature.

Output directory: a directory where output dataset or intermediate files will be stored. Output directory can be a folder or a geodatabase.

Target land-uses: as soon as the land-use filed is selected, it is automatically populated by unique values from land-use field. Users can select only land-uses that they wish to be taken into account in land-use mix computation.

Unit: is used to determine the unit for area calculations. It is also used as the unit for the *area of unit of analysis* parameter.

Area of unit of analysis: determines the area of the buffer around each sample point, within which the land-use mix index is computed. Its unit is specified in the *unit* parameter.

Reference type: land-use mix is measured against an expected distribution of land uses, which can be passed in by the user or derived from the city-wide distribution of land uses. Users can pass in fixed land-use ratio – as the expected distribution – for all sample points, regardless of what land use is dominant in the proximity of each point. Alternatively, they can pass in land-use ratios that vary depending on the dominant land-use around sample points (Figure 9).


```
land_use.txt - Notepad

File  Edit  Format  View  Help

-Replace #'s with respective ratios for each land use for each dominant use in a search area
-land-uses on left are the dominant use in a search area
-Note that the sum of ratios in each line cannot be greater than 1
-Changing output directory or any Change in target land uses will close and delete this file

RESIDENTIAL||RESIDENTIAL:#;GREEN:#;FARM:#
GREEN||RESIDENTIAL:#;GREEN:#;FARM:#
FARM||RESIDENTIAL:#;GREEN:#;FARM:#
```
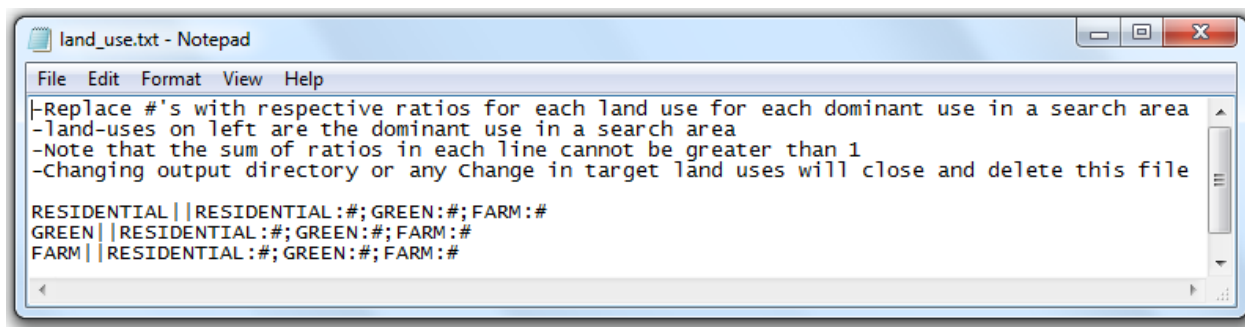
Figure 9: A Notepad window pops up as soon as the reference type with varying land-use distribution is selected by the user, which can be used to pass in expected land-use distributions around sample points based on the dominant land use around them.

9. REFERENCES

Amindarbari, R., Sevtsuk, A., 2013. Measuring Growth and Change in Metropolitan Form, presented at *the Urban Affairs Association Conference*, San Francisco

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Planning Association*, 25(2), 73–76.

Limtanakool, N., Schwanen, T., & Dijst M. (2009). Developments in the Dutch urban systems on the basis of flows. *Regional Studies*, 43(2), 179–196.