Useful indicators for Spatial Microsimulation

Although microsimulation is an increasingly used tool to create or project (spatial explicit) datasets, the evolution of the both the initial dataset and the final results is still under exposed.

Apart from evaluating the usefulness of IPF, this paper also aims to use a new set of (spatial) indicators.

# Often used evaluation techniques in MSM studies

## Evaluation

Generally, to verify the integrity of any model, it is necessary to validate the model outputs, using both internal and external validation methods (Edwards et al. 2010). Internal validation is the process whereby the variables that were used in the estimation of the output data are compared, so the input dataset is compared with the output dataset for those variables. This process examines the data from which the simulated dataset is drawn. External validation is the process whereby the variables that are being estimated are compared to data from another source, external to the estimation process, so the output dataset is compared with another known dataset for those variables.

### Simple (Spearman) correlations expressed by R2;

To do a simple linear scatter plot, converting the data to percentages can be a useful technique (Ballas et al. 2005; Edwards and Clarke 2009). These results can be presented as scatter plots, with the simulated proportion on the x-axis and the actual (census) data on the y-axis with a trend line drawn through the datapoints. The R-squared statistic (the coefficient of determination) (R2) is an indicator that ranges in value from 0 to 1, and it reveals how closely the simulated values for the regression trend line fit the actual (census) data. A trend line is most reliable when its R 2 is at or close to 1. Thus, with this technique, we would expect to see a high coefficient of determination for the constraint/benchmark variables (i.e. variables used in the input dataset for the model) and, if known data is available, for the variable(s) being estimated.

However, regression analysis does not give any information about the fit of the simulated data to the ‘ideal’ line (i.e. where y  =  x and the simulated data is the same as the actual data). Rather, R 2 expresses the fit of the data to the ‘best fit’ line through that data. That is, the coefficient of determination is providing information about precision, not accuracy (Edwards and Tanton, 2013).

One way to do this is with a t test. With a spatial microsimulation model validation, the data are paired (given we are comparing simulated with actual data), thus an equal variance 2-tailed t test can be used to determine if there is any significant difference between the two datasets (i.e. simulated and actual). Thus, if the simulation is robust, we would expect to see no significant differences between the simulated and actual values for the input variables (and estimated/output variables, if known data are available). This enables the model accuracy to be assessed, as opposed to simply its precision.

### Standard Error around Identity (SEI)

where SEI is the standard error around identity, y est are the estimated values for each area, y rel are the reliable estimates for each area from a census or other data source and y rel is the mean estimate for all areas where reliable data are available. This estimate has been used in validation by both Ballas and Tanton (see Ballas et al. 2007; Tanton et al. 2011).

### Total Absolute or Standardized Error (TAE/SAE)

The SAE addresses this issue by using the population size as the denominator, but generally, authors seem to use total population, rather than the population for that categorization of the variable, which may be deemed to understate the size of the errors. Also, this measure does not provide any information on whether any differences are statistically significant (Edwards and Tanton, 2013).

### Z-score

Here, the evaluation mainly takes place at the cell level (for a single constraint in a single area). However, there should also be attention for the overall picture like spatial concentration.

# Often used evaluation techniques in other studies

## General indicators

### Abundance

Clearly, the abundance index *AB* can be most simply calculated as

AB = (1)

in which I() is the indicator function which returns a value 1 when the expression within parentheses is true and 0 otherwise. In words, AB measures the number of distinct groups that are present in the population. Hence 1 ≤ AB ≤ G.

Of course we would expect that the larger the population we consider, the more likely it is that even rare groups are present. It is therefore useful to also consider indexes of relative abundance such as the Margalev *MD* index (Margalef, 1958):

MD = (2)

or the Menhinick index *MI* (e.g. Whittaker, 1977):

MD = (3)

Both measures have been developed in the context of ecological diversity where sampling is common. These relative abundance measures are useful in socio-economic contexts where indexes are compared across populations of quite different sizes.

### Diversity

A common diversity measure based on information theory is the Shannon index (also referred to as the Shannon-Weaver, Shannon-Wiener or entropy index, see Desmet et al. 2009). The index is given by

(6)

This index can only be calculated when each group has at least one member. The index varies between 0 (when there is only one group) and a maximum of ln *G* when all groups have an equal number of members.

A related index is the Hoover index *HI* (also referred to as the Robin Hood index) which calculates the proportion of the population of each group that would have to be redistributed in order to achieve an even distribution, with each group having *P•* / *G* members. The Hoover index is a special case of the Duncan and Duncan dissimilarity index *DD* in which the distribution of individuals across a classification is compared for two populations (Duncan and Duncan, 1955). Section 4.1 discusses this index when the classification refers to spatial areas.

### Socio-cultural distance weighted measures

Recall that we defined *τgh* as the normalised similarity between groups *g* and *h*, such that *τgg* = 1 for all *g* (maximum similarity) and *τgh* = 0 (minimum similarity) when the socio-cultural distance between groups *g* and *h* is large. We can then reconsider the fractionalization index (5) but consider weighted products of fractions of groups *g* and *h*. We define the similarity-weighted diversity index *WD* as

WD = (7)

Like the fractionalization index, *WD* varies between 0 (everyone belongs to one group, or everyone is spread evenly across all groups that have maximum similarity) and 1 − 1/*G* (the population is spread evenly across *G* groups with minimum similarity). For a given distribution of population across group, the more similar the groups are, the lower *WD*. The double summation expression is a special case of what Desmet et al. (2009) refer to as “social effective antagonism”. Desmet et al. (2009) also introduce measures that vary with the scale of the population.

## Spatial Indicators

Some of these are referred to as *global* measures in that they indicate an average spatial pattern across all areas, whereas others are *local* measures in that they are calculated for each area.

One of the most common global spatial diversity measures cited in the literature is the *dissimilarity index*. The index is a measure of displacement – the proportion of people in group one which would have to relocate in order to make their distribution identical to that of group two (Duncan & Duncan, 1955). When the dissimilarity index is computed between one group and all other groups combined, it is known as the *segregation index*. The group segregation index for group *g* across area units *a* is

GSg = (8)

Being a global index, the group segregation index will only reveal an average situation for the group.