Yet another MAUP: how agent-based model uncertainty and sensitivity are space-dependent

Arika Ligmann-Zielinska[1]

[1] Department of Geography, 673 Auditorium Road, Michigan State University, East Lansing, MI 48824, U.S., ligmannz@msu.edu

Abstract

This paper evaluates the spatial dimension of agent-based model uncertainty and sensitivity. Monte Carlo simulations are carried out, leading to distributions of results, recorded for each agent. The results are aggregated into zones of varying shape and size. Output variance decomposition is performed on the zones in order to quantify model sensitivity to every input. For every zone, inputs that have maximum impact on output variability are identified and rendered using maps of dominant factors. The maps are then compared among different aggregation schemes to demonstrate how model sensitivity changes with the change in shape and size of the zones.

Keywords: sensitivity analysis, uncertainty analysis, MAUP, agent-based modelling.

1. Introduction

Geographic modeling is imbued with uncertainty resulting from its spatially-dependent model input factors including parameters, data, and functions. While the spatial dimension of model uncertainty has been recognized and studied for decades, few efforts have been made to evaluate the geographic nature of model sensitivity, that is, the spatial heterogeneity of the influence of factors on the spatially heterogeneous uncertain results (Ligmann-Zielinska, 2013). In this paper, I use an agent-based model of obesity dynamics (oABM), to *evaluate how model sensitivity changes with the change of output data aggregation scheme,* and determine whether oABM outcome sensitivity is prone to the modifiable areal unit problem - MAUP (Openshaw, 1983). To address this question, I perform Monte Carlo simulations to obtain a distribution of results, which are the body mass index (BMI) measures recorded for every agent at the end of model execution. For every zoning scheme, I aggregate the BMI values and calculate their variance (uncertainty analysis, UA) and then use variance decomposition (Lilburne and Tarantola, 2009) to quantify the contribution of inputs to the variability of BMI per zone (sensitivity analysis, SA).

2. Materials and Methods

2.1. Agent-Based Model and Computational Experiments

The oABM provides a platform for computational experimentation, in which a synthetic population of heterogeneous human agents occupies a GIS-based urban environment. The oABM allows for exploration of obesity prevalence by incorporating empirical health and geographic data collected for a selected neighborhood in San Diego, CA, and integrated into a model that simulates weight change (measured using BMI) due to the combined impact of health behavior and the built environment.

The oABM runs for five years with daily increments. The agents use a rule of energy intake and energy expenditure to imitate weight dynamics. Caloric intake is set to a constant value calibrated based on secondary data. The energy balance model requires five factors estimated from BRFSS surveys for the area (http://www.cdc.gov/brfss/): agent's *AGE*, *WEIGHT*, and *HEIGHT* to calculate its basal metabolism, and length of *WORKOUT* plus calories *BURNED* to calculate excess burned energy. If the agent is easily *INFLUENCED*, it can employ a lifestyle change leading to weight loss by invoking rules based on the transtheoretical model of behavior change (Prochaska and DiClemente 1992). After a number of days of *CONTEMPLATION*, the agent starts a new exercise regimen in which it performs *EXTRA_WORKOUT* on a daily basis. The workout is divided into strength training (dependent on accessibility to physical activity centers i.e., *GYM*) and cardio (dependent on neighborhood walkability and safety i.e., *WALK*). The duration of the exercise regimen depends on agent's *PERSISTENCE* and social *SUPPORT*. Occasionally, an agent can *RELAPSE* into its old lifestyle.

The output of this oABM is the BMI value recorded on an individual basis. The model is used to perform Monte Carlo experiments (N=1920), in which the values drawn from probability distributions of thirteen model factors (*UPPER CASE ITALIC* above) are variously configured, rendering a distribution of BMI per individual at the end of model simulation. The number of agents in the census tracts was scaled to match the population distribution based on Census 2010. The model was implemented in Python programming language (https://www.python.org/) and run using the computing resources in the High Performance Computer Center at Michigan State University (http://icer.msu.edu/).

2.2. Variance Decomposition

To evaluate the influence of inputs on the resulting BMI distribution, I employ SA based on variance decomposition (Lilburne and Tarantola, 2009). This model-independent method of SA partitions the variance of model outcome distribution and apportions it into model input factors represented either singly or in combinations to account for the interaction effects. In this paper, the results of SA are presented as total effects sensitivity indices (ST) that quantify the overall relative contribution of factors to BMI variability. This information is instrumental in simplifying the model, where factors with low ST values can be set to constant values, and improving model accuracy, by prioritizing input data collection efforts to factors that notably affect the BMI variance. SA was performed using SimLab (http://ipsc.jrc.ec.europa.eu/?id=756).

3. Results

Four zoning schemes were applied. Census tracts (TRACTS) served as the baseline. After calculating the average (AVG) area of the census tracts, I created a fishnet of squares of AVG size. The third and fourth zoning schemes was a square fishnet with a size of AVG minus/plus the standard deviation (std) of census tract area (AVG_M_STD and AVG_P_STD, respectively). All fishnets were aligned with the lower left corner of the minimum bounding box of the census tracts.

3.1. Uncertainty Analysis

BMI distributions were summarized using a mean and a std of BMI for agents that fall within a particular zone. Figure 1 shows the results for all four zonings. As hypothesized, the aggregation scheme has an observable effect on both the average BMI and its variance. With the increase in zone size (from AVG_M_STD to AVG_P_STD) there is an increase in spatial homogeneity of both statistics, with variance (the critical statistics for SA) approaching a stable value of about one unit of BMI in the AVG_P_STD map. Consequently, I used the two fine-scale schemes (Figure 1.2. and 1.3) to identify a cluster of the highest std = 4.9 BMI units (marked with an oval in Figure 1).

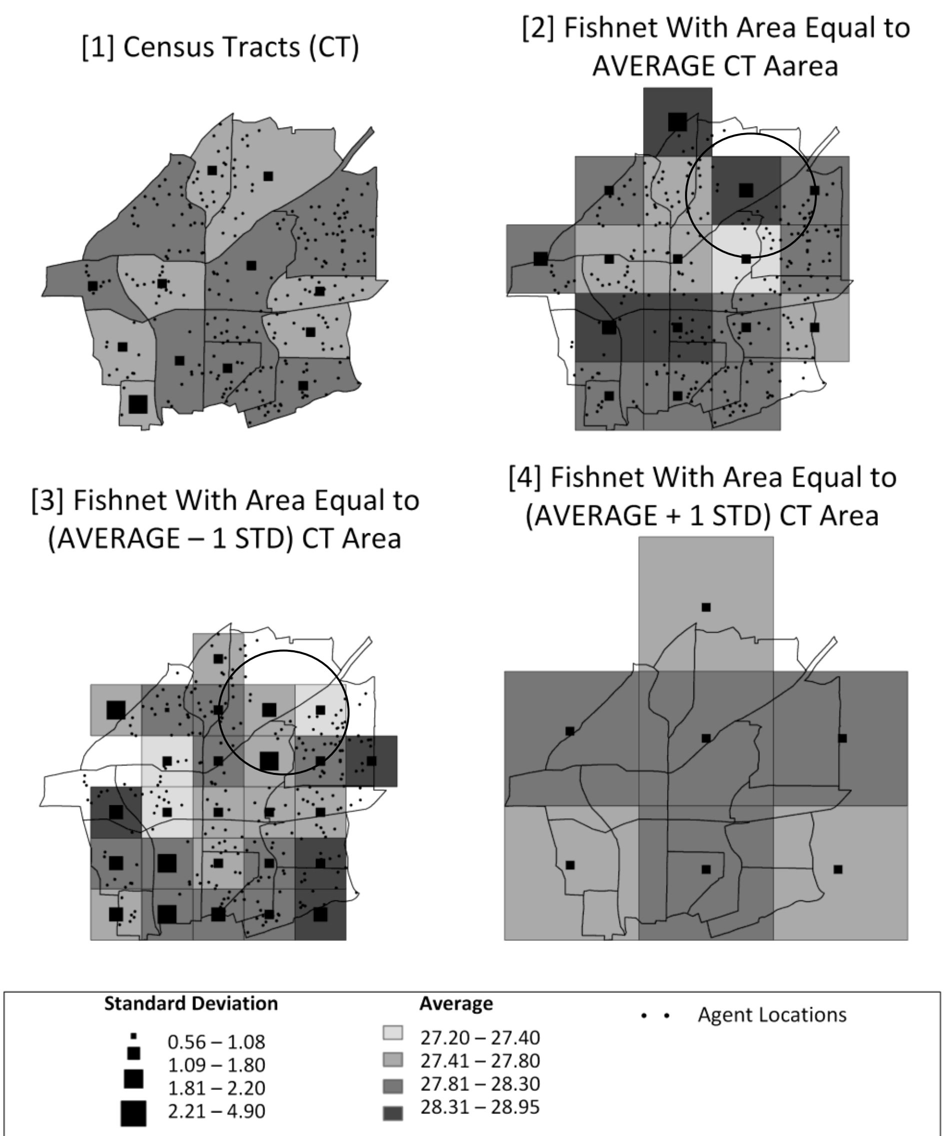

Figure 1: Uncertainty analysis maps of agent BMI values.

3.2. Sensitivity Analysis

The ST values were calculated for every unit of every zoning scheme, resulting in thirteen ST indices per square/tract. Most of the factors proved unimportant in shaping the variance of BMI across all zones and all aggregation schemes. Consequently, these factors could be set to constant values in future model versions (reducing the dimensionality of input space). Five factors were influential: AGE, HEIGHT, WORKOUT, GYM and WALK. To reduce the amount of information obtained through SA, for every spatial unit I identified inputs that scored highest on the ST values, creating maps of dominant factors (Figure 2). The values for these factors ranged from 0.2 to 0.4. In some cases, more than one factor scored high. Consequently, the dominant factors were established within up to 4% of the max ST per zone. For example, if HEIGHT ST=0.23 and AGE ST=0.2 they both share the dominant factor category.

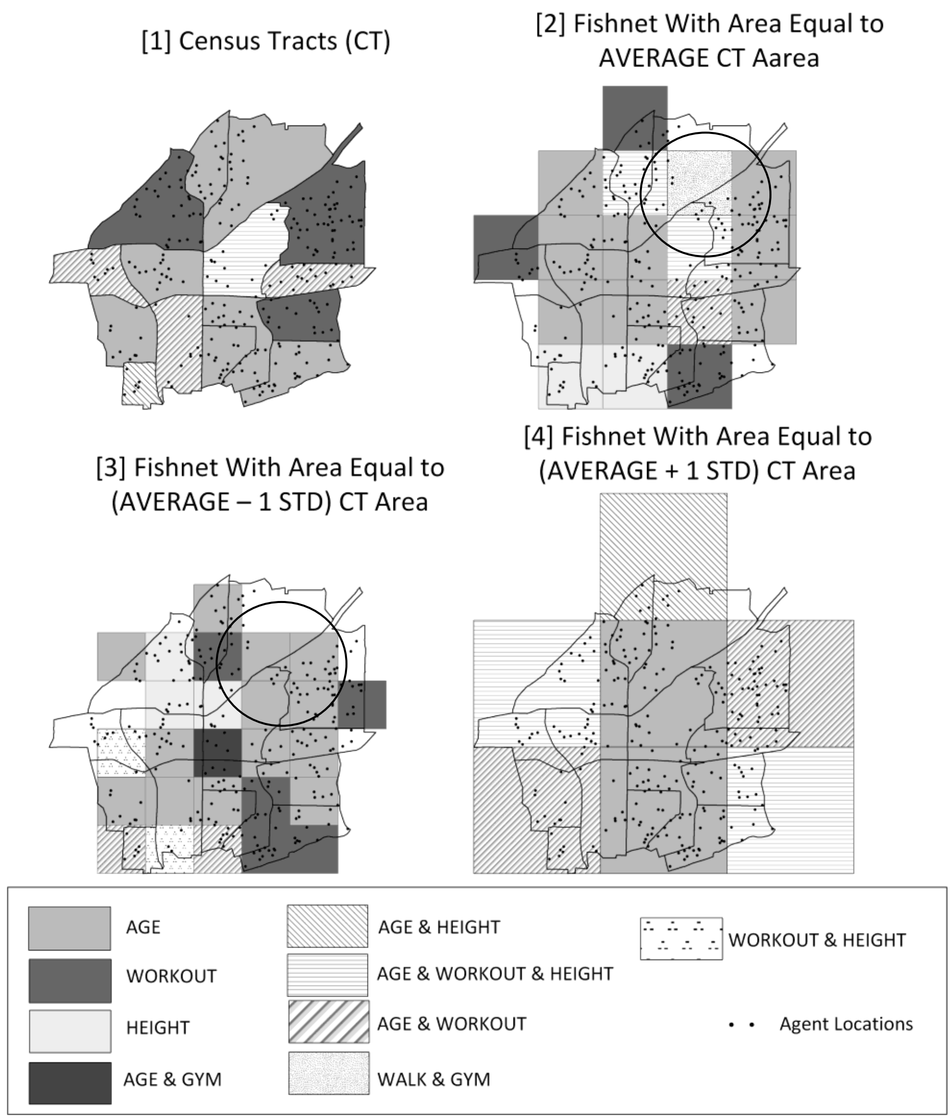

Figure 2: Sensitivity analysis maps with the dominant factors per zone.

The dominant factor maps strengthen the hypothesis that SA is prone to MAUP. While AGE dominates all factors in the vast majority of spatial units (Figure 2), other (combinations of) factors become influential with the change of aggregation. For the cluster of high variance (the center of the oval in Figure 1.2 and 1.3), the AVERAGE aggregation renders WALK and GYM as the most influential on BMI variance, whereas AVG_M_STD results in AGE as the factor driving the variance. Assuming that SA is performed to prioritize input data collection efforts for model refinement, a conservative approach would require data improvement on all three factors in the cluster.

4. Conclusion

The recent proliferation of process-based spatial models calls for reevaluation of their UA and SA. In this paper, I demonstrated that one the potential challenges of UA and SA of spatially-explicit models in the modifiable areal unit problem. Based on the results of an agent-based model of obesity dynamics, I conclude that noticeable differences exist among aggregation schemes in terms of result variability (measured using variance) and the factors that drive this variability (identified through variance decomposition). Future research will focus on comprehensive statistical analysis of MAUP in UA and SA and identification of guidelines addressing this problem.

Acknowledgments

Financial support for this project was provided by the Faculty Initiatives Fund, College of Social Science, Michigan State University. I would also like to thank three anonymous reviewers for their feedback on the previous draft of this manuscript.

References

Ligmann-Zielinska, A., (2013). "Spatially-explicit sensitivity analysis of an agent-based model of land use change." *International Journal of Geographical Information Science*, Vol. 27 (9):1764-1781.

Lilburne, L., Tarantola, S. (2009). "Sensitivity analysis of spatial models." *International Journal of Geographical Information Science*, Vol. 23 (2):151-168.

Openshaw, S. (1983), *The modifiable areal unit problem*, Geo books, Norwich.

Prochaska, J. O., DiClemente C.C. (1992). "The transtheoretical approach." *In:* Norcross J. C., , Goldfried M.R. (eds.) *Handbook of Psychotherapy Integration*: Oxford University Press, pp. 147-171.