

## USING THE TOTAL EXPLORATORY FACTOR ANALYSIS (T-EFA) AS A CALIBRATION TECHNIQUE FOR SLEUTH MODEL

I. E. Ayazli<sup>1,\*</sup>

<sup>1</sup> Sivas Cumhuriyet University, Engineering Faculty, 58140 Sivas, Turkey - eayazli@cumhuriyet.edu.tr

### Commission IV, WG IV/8

**KEY WORDS:** Geo-simulation, Cellular Automata, Exploratory Factor Analysis, Urban Growth

#### ABSTRACT:

Developments in information technologies (IT) allow to modelling dynamic and complex form of cities and several studies have been implemented since 1990s. The cellular automata based urban growth simulation model, SLEUTH is the most well-known one among the simulation models. Calibration is the most important stage of the model created in three stages such as test, calibration, prediction. The more precise the calibration is completed, the more accurate the model generates. Several methods have been developed for the calibration step in which growth coefficients values are calculated by metrics. The study aims to investigate success of the Total Exploratory Factor Analysis (T-EFA) technique, which provides using the 13 metrics all together, in rapid grown settlement areas using high resolution data. In this context, the Sancaktepe district of Istanbul was selected as the study area and a simulation model was generated for the year 2050. The obtained results are promising to apply the T-EFA method in different studies.

### 1. INTRODUCTION

Today's half of the world population lives in cities and reaching this rate to 67% is expected in 2050. The increase in population leads to physical expansion of urban space, called as urban growth. If urban growth is not controlled, some environmental, physical, social and economic problems (low air and water quality, insufficiency of natural resources, energy consumption, pollution etc.) are emerged. In metropolitan cities, where urban growth is intense, natural and environmental resources are polluted and garbage collection and storage problems threaten public health.

Urban growth not only causes destructions of natural resources but also triggers social problems and segregations between luxury residential areas, rural and slum areas, located in urban fringes. In terms of sustainable urban management, urban growth should be monitored and prevented its negative effects.

Contemporary urban management is needed determining direction and speed of urban growth so that feasible plans can be made. Furthermore, the generated urban growth simulation models provide main data of smart cities components, which are smart economy, smart transportation/mobility, smart environment, smart people, smart living and smart governance.

In line with the 2030 target of "Sustainable City and Living Areas" published at the 70th General Assembly of the United Nations (UN) in New York in September 2015 and adopted by 193 UN member countries for the solution of problems caused by urban growth/sprawl, this growth should be brought under control (UN, 2015). In other words, to manage cities in a contemporary sense, the speed and direction of change should be determined in the urban fabric so that applicable plan decisions should be taken. In this context, urban growth simulation models are needed to understand the problems of

cities all over the world better and to produce solutions and use natural resources more efficiently.

The first studies that can be considered as the ancestors of urban growth simulation models generally have a radial form and linear structure. After the 1950s, complexity system modelling approach has been used in planning. Therefore, simulation methods are used instead of traditional analysis methods to determine complex system behaviors (Junfeng, 2003; Benenson, et al., 2004; Batty, 2007). Cellular Automata (CA) are among the first models that come to mind in modeling complex system behaviors (Clarke & Gaydos, 1998).

The concept of automaton emerged with the universal Turing machine, which was attributed to Alan M. Turing, who introduced the basic working principles of computers. In the 1940s, the first CA was designed by John von Neumann and Stanislaw Ulam. A simple CA has five basic components, including lattice (cell space), state, neighbourhood, transition rules, and time. The lattice component, which can be one- or two-dimensional, is used to define a space in which automata are present (Torrens, 2000). Each cell in the lattice, which is a grid network, has an automaton, and this cell contains the state of the automaton. Automaton behaviours changing according to the state of the neighbouring cells over time are determined by transformation rules (Benenson, et al., 2004; Torrens, 2000).

Geospatial analysis studies with CA started in the late 1970s (Tobler, 1979; Couclelis, 1985; Phipps, 1989) and have been used in land use/cover modelling projects since the 1990s (Clarke, et al., 1997; White, et al., 1993; Batty, et al., 1994; Portugali, et al., 1995; Wu, 1998), and many models have been developed. The most popular ones among these are GeoDynamica and METRONAMICA developed in the Netherlands, DUEM developed in the UK, and the SLEUTH

\* Corresponding author

model developed in the USA. One of the most widely used models among them is the SLEUTH urban growth model developed by Keith C. Clarke (Clarke, et al., 1997). This coupled model called SLEUTH with the integration of the Deltatron Land Use/ Land Cover model with the UGM, enabled the investigation of land use/cover changes with urban growth. Since the 1990s, it has been used in urban growth analysis studies in a total of 25 cities in 15 states in the United States of America and in more than 15 countries worldwide (URL 1). In SLEUTH software, the settlement area created in four different time zones and the land use/cover created in two different time zones, slope, transportation network, and excluded area data are required. Furthermore, the second software named Cygwin is needed as operating system to run the model in Windows. The parameters required for the model are defined by changes made in the scenario file. Thanks to four growth rules and five growth coefficients that affect the application of these rules, an urban growth model is created as a result of the test, calibration, and prediction stages by using the Monte Carlo iteration method.

The data to be used at the test stage are checked, and if this stage is completed without any problems, the second step, the calibration stage, is proceeded to. It is aimed to calculate the optimal coefficients at the calibration stage, which is completed in four different steps: coarse, fine, final, and forecast. The urban growth/sprawl simulation model is generated by using the calculated coefficients at the prediction stage.

Although a number of calibration methods have been used since the 1990s, there is still no consensus on which one should be used to determine coefficients. The aim of this paper is to generate an urban growth simulation model using Total Exploratory Factor Analysis (T-EFA) method in Sancaktepe district of Istanbul (Figure 1). The population of the district, which was 229,093 at the time of its establishment, increased by 81% in the last years and reached 414,143 (TSI, 2019). The main data source of the model is high-resolution cadastral maps in four periods, such as 1961, 1992, 2002 and 2014. Transitions in land cover since 1961 were examined, and a CA-based simulation model were generated for the years 2050. Results of the T-EFA technique, is led to create high accuracy urban growth simulation model.

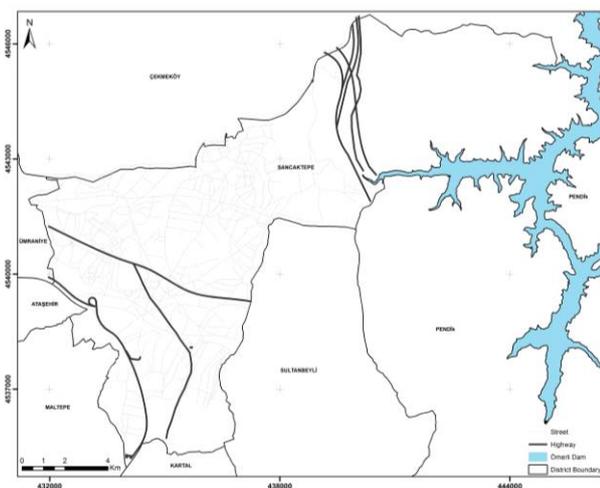


Figure 1. Study Area

## 2. METHODOLOGY

A multivariate statistical technique T-EFA aims to calculate a few conceptually significant new variables (factors, dimensions) by using the relationships between variables (Tabachnick & Fidell, 2013; Buyukozturk, 2002). The most important problem in the technique is the sample size. It should be at least five times the number of the observed variables, but if there are robust, reliable relationships and a small number of factors, the sample size can be decided to be at least 50, provided that it is greater than the number of variables (Tabachnick & Fidell, 2013). In order to interpret the adequacy of the sample size, the KMO and Bartlett's test of sphericity hypothesis should also be applied to the model. If the KMO value is greater than 0.45 (Balanza, et al., 2007; Li, et al., 2017) and Bartlett's test of sphericity significance level value is less than 0.05 (Tabachnick & Fidell, 2013), the sample size is determined as acceptable.

The eigenvalue coefficient should also be calculated to find number of factors. It should be greater than 1 to determine the number of important factors.

At the final step of the T-EFA technique, rotation component matrix (RCM) should be calculated so that the factors find variables and can be interpreted more easily (Tabachnick & Fidell, 2013). If there is no correlation between the factors, the orthogonal rotation algorithm is applied and varimax method is frequently used for this rotation in the literature.

In this paper, four periods of land cover, three periods of transportation and three periods of building stock, the administrative boundaries produced by Istanbul Metropolitan Municipality (IMM), two periods of transportation, one period of building stock, and the digital elevation model (DEM) produced by the General Directorate of Mapping data were used. The building stock, land covers, and accessibility data used in the study were produced from cadastral maps and zoning plans. It was found that lots which are smaller than 100 m<sup>2</sup> were not suitable for building (Ayazlı & Başlık, 2016). For this reason, the pixel size was selected as 10 m x 10 m for input data.

T-EFA was used for coarse, fine and final calibration of the urban growth simulation model and calculations were performed using IBM SPSS software. According to KMO and Bartlett's significance values (p), the suitability of the sample size was investigated. After that, the eigenvalue-based Total Variance Explained (TVE) values were computed for each factor, and at the final step of T-EFA, the RCM was generated. Finally, the factor clusters in which the intercorrelated SLEUTH metrics were classified.

## 3. RESULTS

In this part of the paper, the urban growth in high populated district, Sancaktepe, was analyzed by using high-resolution cadastral data, and the obtained findings were presented. After checking the compatibility of the input data, the calibration stage was started and completed in four steps but the T-EFA was applied coarse, fine and final calibrations.

In order to calculate the best fit values of the coefficients using the Brute Force Calibration (BFC) method, the input data with a pixel size of 40 m were used in the first step called coarse calibration. In the scenario file, the initial value was set to 0, the last value was set to 100. The calculations were made with 5 iterations, and the 13 r<sup>2</sup> scores were produced for SLEUTH

metrics. The similar computations were made for fine and final calibration steps.

The T-EFA method was used for range selection in calibration step. Due to the fact that the KMO score is greater than 0.45 and the p-value is less than 0.05 (Table 1), therefore, the sample size is suitable.

Calibration Step	KMO	p	TVE (%)
Coarse	0.757	0.000	70
Fine	0.809	0.000	82
Final	0.872	0.000	90

Table 1. T-EFA results

After the calibration, the best fit scores of growth coefficients were calculated shown in Table 2, used for prediction step in order to create simulation model. According to TVE scores, cumulatively 70% of total variance was explained in coarse calibration, 82% in fine calibration and 90% in final calibration.

Growth Coefficients	Best Fit Scores
Diffusion	68
Breed	68
Spread	68
Slope	45
Road gravity	91

Table 2. The best fit scores of growth coefficients

After the prediction step urban growth simulation model was generated for Sancaktepe and probable threaten forest and agricultural areas were detected (Figure 2). According to change detection analysis between 2014-2050, while there is a decrease in 70% agricultural and 74% forest areas, residential areas rate increases 2.6 folds.

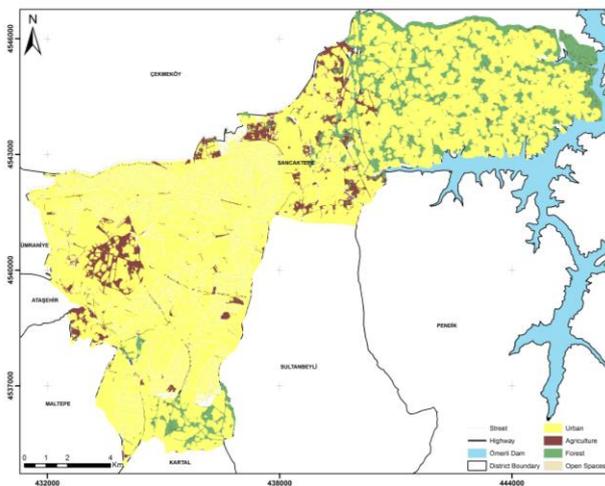


Figure 2. Land cover for 2050

#### 4. DISCUSSION AND CONCLUSIONS

In this study, an urban growth simulation model was created for the year 2050 in Sancaktepe, Istanbul. The model provides a foresight about the future land cover of the district and the more accurate models bring about the more realistic urban planning. In Sancaktepe, where population growth has rapidly increased in the last 10 years, high-resolution raster data from cadastral data has been used in order to generate a precise model. The simulation results showed us the necessary precautions should

be immediately taken in order to control urban growth and protect the forest and agricultural areas in Sancaktepe, Istanbul. In the T-EFA method, all  $r^2$  scores can be handled at the same time, and the inter-correlated metrics are grouped in the same cluster, and growth coefficient ranges can be narrowed using highly correlated variables. Although promising results were obtained using high resolution data, success of the T-EFA should be tested with low resolution data.

#### ACKNOWLEDGEMENTS

This study was supported by TUBITAK with project number 112K469.

#### REFERENCES

- Ayazlı, I. E., & Başlık, S. (2016). *Mülkiyet Deseni İle Kentsel Büyüme İlişkisinin Simülasyon Modelinin Oluşturulması Projesi Sonuç Raporu*. Sivas: Proje Sonuç Raporu, TÜBİTAK.
- Balanza, R., Garcı'a-Lorda, P., Pe' rez-Rodrigo, C., Aranceta, J., Bonet, M. B., & Salas-Salvado', J. (2007). Trends in food availability determined by the Food and Agriculture Organization's food balance sheets in Mediterranean Europe in comparison with other European areas. *Public Health Nutrition*, 10(2), 168–176.
- Batty , M., & Xie, Y. (1994). From cells to cities. *Environment and planning B: Planning and design*, 21(7), 31-48.
- Batty, M. (2007). *Cities and Complexity*. Cambridge, London: MIT Press.
- Benenson, I., & Torrens, P. M. (2004). *Geosimulation. Automata-based modeling of urban phenomena*. John Wiley & Sons Ltd.
- Buyukozturk, S. (2002). Factor Analysis: Basic Concepts and Using to Development Scale. *Educational Administration in Theory & Practise*, 32, 470-483.
- Clarke, K., Gaydos, L., & Hoppen, S. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area. *Environment and Planning B: Planning and Design*, 24, 247-261.
- Couclelis, H. (1985). Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A*, 17, 585-596.
- Junfeng, J. (2003). *Transition Rule Elicitation for Urban Cellular Automata models* (Doctoral dissertation, tesi di dottorato (non pubblicata). ITC.
- Li, S., Yang, Z., & Li, H. (2017). Statistical evaluation of no-reference image quality assessment metrics for remote sensing images. *ISPRS International Journal of Geo-Information*, 6(5), 133. doi:https://doi.org/10.1016/j.optcom.2018.08.032
- Phipps, M. (1989). Dynamical behavior of cellular automata under the constraint of neighborhood coherence. *Geographical analysis*, 21(3), 197-215.
- Portugali, J., & Benenson, I. (1995). Artificial planning experience by means of a heuristic cellspace model: Simulating international migration in the urban process. *Environment and Planning A*, 27, 1647–1665.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics (Sixth Edition)*. Boston: Pearson.

Tobler, W. R. (1979). Cellular geography. In *Philosophy in geography* (pp. 379-386). Springer, Dordrecht.

Torrens, P. M. (2000). *How cellular models of urban systems work* (1. Theory). from <http://discovery.ucl.ac.uk/1371/1/paper28.pdf>

UN, United Nations. 2015, from <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>

URL 1, Project Gigalopolis. 2019, from <http://www.ncgia.ucsb.edu/projects/gig/Repository/SLEUTHapplications.htm>

TSI, Turkish Statistical Institute. 2019, from <https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr>

White, R., & Engelen, G. (1993). Cellular dynamics and GIS: Modelling spatial complexity. *Geographical Systems*, 1(237–253).

Wu, F. (1998). An experiment on the generic polycentricity of urban growth in a cellular automatic city. *Environment and Planning B: Planning and design*, 25(5), 731-752.