Modeling Urban Sprinkling with Cellular Automata

Lucia Saganeiti[1], Ahmed Mustafa[2], Jacques Teller[3], Beniamino Murgante[1]

[1] School of Engineering, University of Basilicata, Potenza, Italy

[2] Urban Systems Lab, The New School, New York, USA

[3] LEMA, Urban and Environmental Engineering Department, Liège University, Belgium

Abstract

This paper presents a spatiotemporal analysis to simulate and project urban sprinkling with coupled cellular automata (CA) and multinomial logistic regression (MLR) model. Our case study, the Basilicata region, south of Italy, is characterized by urban sprinkling - literally "a small amount of urban territory distributed in scattered particles". The region is witnessing a decoupled growth between demographic trend and urban expansion. We applied a coupled approach based on CA and MLR for urban sprinkling modeling and simulation. From three regional building datasets (1989, 1998 and 2013) building density maps were created and used to calibrate and validate the model and to project future urban expansion. Built-up causative factors were identified through an analysis of 19 articles that were compared and discussed according to their main features (methods, case studies, drivers, urbanization dynamics and demographic growth). The transition probability for the first period (1989-1998) was calibrated with MLR for built-up causative factors and with the multi-objective genetic algorithm (MOGA) for CA neighborhood effects. The calibrated model was used to simulate the 2013 urban pattern which was compared with the actual map of 2013 (validation). We then used our calibrated model to simulate urban expansion in 2030. The results of the 2030 forecast show the largest variations in class 1 (low density built-up patches) that correspond to urban sprinkling.

Keywords: Urban Sprinkling, Low-density, Cellular Automata, Multinomial Logistic Regression, Basilicata Region.

1. Introduction

The quantitative and qualitative analysis and prediction of land use transformation dynamics plays an increasingly central role in earth system science (Feng, 2017; Feng & Tong, 2020). Among these transformation dynamics, urban sprawl is undoubtedly amongst the most studied dynamics. Urban sprawl is usually associated with the development of mono-functional, low-density urban settlements occupying a large territory around medium- and large-sized cities (Galster et al., 2001; Hasse & Lathrop, 2003; Nechyba & Walsh, 2004). A similar, though somehow distinct, phenomenon of territorial transformation is observed in the proximity of smaller cities. This has been defined as urban sprinkling (Romano, Zullo, Fiorini, Ciabò, et al., 2017). Literally speaking, according to Merriam-Webster's definition, sprinkling is defined as: "a small amount falling into drops or scattered particles". Urban sprinkling is typical of rural areas and involves the construction of low-density, small-scale residential settlements, scattered throughout the territory, far from existing public services and characterized by low levels of accessibility. It is characterized by very low population density and building density rates (Romano, Zullo, Fiorini, Marucci, et al., 2017). It is usually associated with territorial fragmentation (Saganeiti et al., 2018). The configuration of urban sprinkling is extremely dispersed. It develops

through heterogeneous and very small-sized aggregates, composed of a single building or a small group of buildings. Aggregates may welcome different uses, from residential to industrial, agricultural or commercial uses, which are inserted within the rural matrix. These aggregates typically develop along the existing road networks, which are usually characterized by an organic scheme. While existing cities tend to attract urban sprawl, which progressively consolidates over time, newly built areas would rather tend to develop farther away from existing urban settlements in urban sprinkling.

On the Italian territory, urban sprinkling is usually associated with weak urban planning. In some cases, it is the consequence of abusive initiatives, encouraged by amnesties for the violation of building regulations. In Belgium it is largely related to deficiencies of land planning documents adopted in the 1970's, which provided room for the so-called "ribbon development" along secondary roads without much congruency with existing settlements (Halleux et al., 2012; Mustafa, Van Rompaey, et al., 2018; Verbeek et al., 2014). Uncontrolled, poorly managed and fragmented urban transformation generates an economic and social impact on the population also defined as a social cost. The social cost includes both direct costs arising from the construction of new roads and new infrastructure services, as well as indirect costs arising from health costs owing to increased pollution and therefore increased travel, costs on account of loss of landscape quality and other costs associated with the daily life of the local population (Jan K. Brueckner, 2001; Carruthers & Ulfarsson, 2003; Freilich & Peshoff, 1997).

It is of great interest to study the dynamics of territorial transformation, in particular those linked to urban sprinkling, so as to understand which factors (drivers) have most contributed to the development of the phenomenon. Moreover, the projection into the future (even if not a very distant one) of the transformation dynamics may help control the social costs that the population bears because of the urban sprinkling phenomenon. As demonstrated in Manganelli et al., 2020, the social costs increase, as urban sprinkling increases.

A fundamental prerequisite for developing or applying an urban transformation prediction model is certainly the knowledge of the factors that influence the process, i.e. the causative factors (drivers). Many studies (G. Li et al., 2018; Mustafa et al., 2017; Puertas et al., 2014; Rienow & Goetzke, 2015; J. Yang et al., 2020) are dedicated to the analysis of the drivers that regulate the transformation dynamics of a territory, i.e. the urbanization processes. Table A in the appendix A shows a collection of 19 scientific articles from 2014 to 2020. For each one, in addition to other information discussed in the next paragraph, the drivers taken into consideration have been reported. The processes of expansion are influenced by geophysical, socio-economic and legislative conditions. Most research shows that economic factors, including population growth, income and value of agricultural land, are of primary importance in setting the rules for urban expansion (J. K. Brueckner & Fansler, 1983; McGrath, 2005). As a matter of fact, urban prediction models are generally applied in contexts where the population growth rate is positive and settlement density is relatively high. Among the reviewed articles, only Rienow et al. (Rienow & Goetzke, 2015) analyses a rural context with scattered settlements and a negative demographic trend in a German region.

Land use models allow to project and simulate future urban patterns in order to act on the dynamics and mechanisms of urban expansion (Wahyudi & Liu, 2016). Urban land use change models are generally analyzed and applied to provide a decision support to policy makers in the implementation of new urbanization plans (Allen & Lu, 2003; Nasiri et al., 2019; Puertas et al., 2014). Furthermore, these models can support planning policies such as flood risk mitigation (Mustafa, Bruwier, et al., 2018), regulation of climate change

and the provision of ecosystem services (Shoyama & Yamagata, 2014) as well as the development of scenarios for environmental impact assessment (Poelmans & Van Rompaey, 2009). In this study, our focus is to cross the literature gap concerning the correlation between demographic trends and urban expansion in order to arrive at the conclusion that a predictive model can be used to control urban transformations where they are not necessary and to preserve those areas with a particular environmental value.

In this research, we have applied a modeling approach that integrates multinomial logistic regression (MLR) and cellular automata (CA) to analyze and project urban sprinkling. The model was proposed by Mustafa et al. 2018 (Mustafa, Heppenstall, et al., 2018) and used to simulate urbanization scenarios in Belgium.

Our case study, the Basilicata region in southern Italy, is affected by the urban sprinkling phenomenon and a decoupled growth between demographics and built-up areas in the last 30 years (Saganeiti et al., 2018). This phenomenon is representative of a number of internal Western areas which, even though their population is in decline and/or static, have the necessary resources to continue to invest in urbanization processes (Caselli et al., 2020; Martinez-Fernandez et al., 2012; Wiechmann & Pallagst, 2012). A simulation and projection model of urban expansion based on a multi-density approach (4 urban density classes) will be used. This approach appears to be fundamental and novel for a context characterised by urban sprinkling or an urban expansion in the absence of population growth. Therefore, for urban expansion modeling, , built-up density maps were generated on the basis of three regional building datasets (1989, 1998 and 2013) with four density classes: non built-up, low density, medium density and high density and used for the calibration, validation, and simulation phases. The transition probability for calibration (1989-1998) was calculated with MLR for the built-up causative factors and with multi-objective genetic algorithm (MOGA) for CA neighborhood interactions. Among the causative factors considered, are those concerning the physical, socio-economic, proximity and constraints factors. The calibrated parameters were used for the simulation of the 2013 map which was compared with the actual map of 2013 (validation). The final objective is to simulate business as usual urban pattern in 2030.

2. Relevant features of urban expansion models

Why model future urban expansion scenarios? Undoubtedly, to understand the spatial processes of urban development dynamics over time and to serve as tools to project future policies focused on the principles of sustainable development. Geographic Information System (GIS) and remote sensing techniques are used to feed and construct urban expansion models. Among the common approaches: cellular automata, logistic regression, geographic weighted regression, Markov Chain integrated with cellular automata methods and agent-based. The ability of the cellular automata approach to simulate and predict land use changes assumes that any previous urban expansion, influences future characteristics through local and regional interactions between different types of land use.

CA models are easy to apply, can simulate complex models, have an open structure, integrate with other models and can simulate spatial and temporal models (Aburas et al., 2016). They overcome their limit of not being able to include drivers for urban expansion simulation, by integrating with other quantitative and space-time methods such as Analytic Hierarchy Process (AHP) and MLR. Table A lists 19 scientific articles on urban expansion modeling from 2014 to 2020. Due to the presence of several recent reviews of prediction models (Aburas et al., 2016; Dang & Kawasaki, 2016; Poelmans & Van Rompaey, 2010; van Vliet et al., 2016; Wahyudi & Liu, 2016), articles published before 2014 have not been considered in this summary. The search was conducted on February 2020, in the Scopus website and was carried out considering one of these keywords:

"Urban growth", "Model prevision", "Prediction", "Cellular Automata", "Logistic regression", "Scenario", "Urban expansion". Only articles from scientific journals in the disciplines of urban and territorial sciences, environmental engineering and earth sciences were selected. A further selection was made by reading the abstracts and finally 119 papers were collected.

In conclusion, for each year, the papers with the most citations were analyzed in detail and synthetized. Table 1 shows a summary of the main elements collected in Table A in the appendix where the following elements are reported:

- Authors: reference is made to the first author of the paper and the year of publication;
- Study case: the area of study analyzed is reported;
- Application: the type of application that was carried out in the paper such as: simulation and forecasting of urban expansion, study of the drivers that regulate the dynamics of transformation, flood risk mitigation, modelling of urban densification processes and others;
- models and techniques: used in research. Among the most common are certainly the models with cellular automata, logistic regression and artificial neural networks;
- predictive variables: this includes all variables that affect the dynamics of urban transformation. They have been divided into 6 categories: physical factors, proximity factors, social factors, neighborhood interactions and urban policies. For each article the box corresponding to the driver taken into consideration is highlighted in grey and, in most cases, the drivers are specified (with text);
- population trends and urban dynamics: in this column the demographic growth and urban expansion dynamics of the various case studies have been included. Population growth rates and urban expansion dynamics are highlighted.

Table 1 Summary table of the 19 collected papers.

Reference	Models and techniques	Physical factors	Proximity factors	Social factors	Economic factors	Neighborhood interaction	Urban policies or constraints	Population trend**
(Hua et al., 2014)	SLEUTH	■				■	■	++
(Puertas et al., 2014)	CA, Markov and LR	■	■				■	+
(Basse et al., 2014)	CA and ANN	■	■			■		+
(Rienow & Goetzke, 2015)	SLEUTH, BLR and SVM	■	■	■		■		+ -
(Halmy et al., 2015)	Markov - CA						■	na
(Han et al., 2015)	CLUE-S and Markov	■	■	■	■			+
(Chen et al., 2016)	SA-Patch-CA	■	■			■		+
(Ku, 2016)	SR and CA		■	■	■	■		++
(Vermeiren et al., 2016)	ABM	■	■					+
(Mustafa et al., 2017)	LR, CA and ABM	■	■			■		ns
(Shafizadeh-Moghadam et al., 2017)	ANN, SVR, RF and other	■	■			■		++
(Aburas et al., 2017)	CA and Markov based on AHP and FR	■	■					+
(C. Li et al., 2017)	GWR and CA	■	■			■	■	++
(Martellozzo et al., 2018)	SLEUTH and AHP	■				■	■	+

Reference	Method							Symbol
(G. Li et al., 2018)	SEM - Probit and SAR – Probit	■	■	■	■	■	■	++
(Mustafa, Heppenstall, et al., 2018)	MLR, CA and GA	■	■		■	■	■	na
(Y. Yang et al., 2019)	SSP and LUSD-urban	■	■	■	■		■	++
(J. Yang et al., 2020)	CA and TMC	■	■		■	■	■	+
(Feng & Tong, 2020)	CA, statistical analysis and heuristics analysis	■	■	■	■			na

[*]Acronyms: Slope, Land Use, Exclusion, Urban Extent, Transportation and Hillshade (SLEUTH), Cellular Automata (CA), Logistic regression (LR), Artificial Neural Network (ANN), Binomial logistic regression (BLR), Support Vector Machines (SVM), Conversion of Land Use and its Effects at Small regional extent (CLUE – S), Survival Analysis (SA), Spatial regression (SR), Agent Based Model (ABM), Support Vector Regression (SVR), Random Forest (RF), Analytical Hierarchy Process (AHP), Frequency Ratio (FR), Geographical Weighted Regression (GWR), Spatial Error Probit Model (SEM - Probit), Spatial Log Probit Model (SAR – Probit), Genetic algorithm (GA), Multinomial Logistic Regression (MLR), Shared Socioeconomic Pathways (SSP), Land Use Scenario Dynamics (LUSD), Time Monte Carlo (TMC).

[**]Symbol: ++ very high positive demographic trend; + positive demographic trend; - negative demographic trend; ns not significant demographic trend; na not available data on demographic trend.

The articles analyzed (16 papers out of 19) show a positive demographic trend and high population density indices, which confirm that urban expansion is usually associated with population growth. Only in the article by Rienow et al. (Rienow & Goetzke, 2015) is there a negative growth rate, here the case study was analyzed to be compared with another case study with positive population growth in the same article. Other cases of non-significant population growth concern the work of Mustafa et al. (Mustafa et al., 2017) which analyzes a case study with a low population growth rate and the article by Halmy et al. (Halmy et al., 2015) which contains no information on population growth. Concerning drivers influencing urban transformation dynamics, physical factors have been considered in all papers except in the Ku (Ku, 2016) and Halmy et al. who consider only proximity factors and urban planning policies. Elevation and slope are the most used physical factors. The case of Yang et al. (J. Yang et al., 2020) is particular as it also considers soil quality (in terms of habitat quality) among the physical factors. With the exception of Hua et al. (LIU & MA, 2011) and Martellozzo et. al (Martellozzo et al., 2018) who used the SLEUTH model (Slope, Land Use, Exclusion, Urban Extent, Transportation and Hillshade), all the others considered proximity as a factor. The proximity to road networks is certainly amongst the most utilized. Among those used less, one can find the proximity to highway entrances (J. Yang et al., 2020) to irrigation canals (Halmy et al., 2015). Social and economic factors are not always considered as drivers for urban transformation dynamics. Among those most frequently used are population density, access to jobs, gross domestic product (GDP) and richness rates. Eleven papers in all consider interactions between neighbors from different neighborhood sizes ranging from 3x3 to 5x5. Only 10 papers consider urban planning policies as a driving force. On the basis of this data collection, the study area was modelled using a logistic regression with driving forces belonging to each identified category of the predictive variables.

3. Materials and Methods

3.1 Study area

The urban expansion model was applied to the entire territory of the Basilicata region in the south of Italy. It is a region characterized by a low population density and settlement. With 56 inhabitants per square kilometer it is the second to last Italian region, followed only by the mountainous Aosta Valley region with 39 inhabitants

per square kilometer (data from the 15th national census of National Institute of Statistics -ISTAT - 2011 (*Istat.it*, 2011)). The territory is affected by a decoupling phenomenon (Li et al., 2019; Song & Liu, 2014). As shown in a previous study (Manganelli et al., 2020; Saganeiti et al., 2018; Scorza et al., 2020), the demographic rate has been negative since the 90s and, in contrast, the expansion rate of new residential buildings is substantially increasing. To give an example, in the time interval from 2006 to 2013 there is a decrease in population of 2.2% and an increase of 4.5% in the number of buildings for residential use.

Figure 1 shows: on the left, a schematic map of Italy with the identification of the study area in black, and on the right the Basilicata region. The territory of the Basilicata region is mainly mountainous, the highest peaks reach 2000 meters and the percentage of mountainous land is 46.8% with an altitude higher than 700m above sea level. The remaining 45.1% of the territory is hilly and only 8% is lowland. There are only two larger urban agglomerations: the Potenza and Matera municipalities (the two provincial capitals). The urban areas are indicated in red. The other agglomerations are small and dispersed throughout the whole territory.
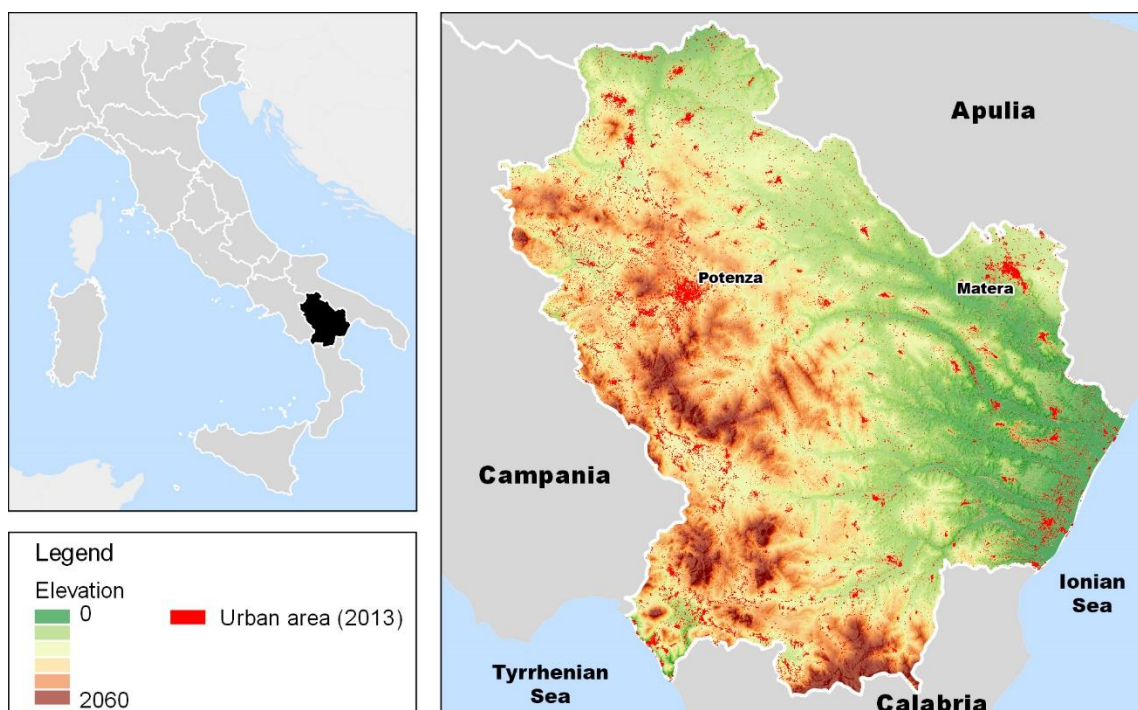
Figure 1 On the left, the Italian territory with the study area in black. On the right, the Basilicata region with the Digital Elevation Model and urban areas as at 2013.

This diffusion and expansion of the urban and rural areas occurred according to an urban sprinkling transformation dynamic, a characteristic of those areas with a low settlement density, with the expansion of small urban agglomerations far from the existing ones and scattered in the territory without following specific planning rules.

3.2 Building density maps

In order to generate built-up density maps, we used the regional technical map, made available by the regional geo portal (*RSDI – Geoportale Basilicata*, n.d.) in a vector format. This dataset includes the footprint of all buildings, whatever their use (residential, industrial, commercial or other) as at 2013. Following a comparison between the vector of the buildings in 2013 and the orthophotos of several years, the vector of buildings in 1998 and 1989 were obtained. These three vector maps (1989, 1998 and 2013) were then rasterized at a 2x2

m pixel resolution and have been aggregated to a 100x100 pixel resolution in which every pixel has a range density value from 0 to 2500 (counting the 2x2 m pixels within each 100x100 m pixel). We then classified the urban density continuum into four classes: non built-up, low density, medium density and high density. The geometric classification was used to identify the ranges between the density classes. It is a classification method to represent/display data that is not normally distributed. Compared to other classifications (for quantiles, logarithmic or equal intervals), it shows more variations on the data due to class breaks that occur at a constant geometric interval than the interval before the breaks (Arlinghaus & Kerski, 2013). Table 2 summarizes the ranges of values for the building density classes. The density 18.75 is the limit below which the pixel is considered "not-built-up". This value corresponds to a building of 75 square meters which is the minimum size (established by Italian legislation) for a building to be declared inhabitable by 4 people.

Table 2 Buildings density classes.

Class	Density classes	Density values	Built surface/ha	Coverage ratio [%]	% of cells in the study area (1989)
Not-built up	0	from 0 to 18.74	Less than 75 m^2	-	93.62%
Low density	1	from 18.75 to 54	Less than 220 m^2	from 0.75% to 2.1%	2.82%
Medium density	2	from 55 to 329	Less than 1320 m^2	from 2.2% to 13.19%	2.91%
High density	3	from 330 to 2500	Until 10,000 m^2	from 13.20% to 100%	0.65%

Density values must always be related to the context under analysis. If we compare these values with the ones obtained in Mustafa, Van Rompaey, et al., 2018 we notice that there is a strong disparity: the first three density classes of this work are positioned between the intervals corresponding to the first two classes of Mustafà's work (Mustafa, Van Rompaey, et al., 2018). This is highlighted by the lower limit of the high-density class which corresponds to an area of 1320 sq.m of buildings per hectare and a coverage ratio of only 13%, in the reference area of 1 ha. The percentage of cells in the high density class over the entire study area is 0.65% and in comparing them with the urban aggregates of Saganeiti et al., 2018, we observe that they correspond to the largest urban aggregates of each municipality in the region, i.e. compact urban centers, for a percentage equal to 66%.

Table 3 shows the amount of change that occurred in the study area between the various density classes and during the two temporal periods: 1989-1998 and 1998-2013.

Table 3 Cells changed from one class to another and percentage of change for each class based on total transformation.

		1998			
		Class 0	Class 1	Class2	Class 3
1989	Class 0	-	5895 (42.76%)	3725 (27.02%)	532 (3.86%)
	Class 1	-	-	2850 (20.67%)	54 (0.39%)
	Class 2	-	-	-	731 (5.30%)

		2013			
		Class 0	Class 1	Class 2	Class 3
1998	Class 0	-	5353 (40.49%)	3451 (26.11%)	630 (4.76%)
	Class 1	-	-	2831 (21.41%)	75 (0.57%)
	Class 2	-	-	-	880 (6.66%)

The changes from density class 0 to classes 1, 2 and 3 are considered as expansion processes and, in the two phases examined, are dominant: 73.64% and 71,36% respectively for the periods 1989/1998 and

1998/2013. Differently, changes from density class 1 to classes 2 and 3 and from class 2 to 3 are considered as densification processes and correspond, for the two temporal phases, to 26.36% and 28.64% respectively on the basis of total changes. Since the changes from density class 1 to density class 3 interest an irrelevant number of cells (0.39% and 0.57% for the two temporal phases) they will not be considered in the further analysis.

Dominant over all the changes, for both time phases, is the expansion of the low-density class (greater than 40%). The low-density class has coverage percentages that correspond to the sprinkling ones, which therefore, can be considered the dominant transformation in the study area.

3.3 Built – up causative factors

The drivers considered to have an impact on the settlement system are listed in Table 4, and will be used to model the transition rules, with the MLR. The drivers were chosen following the literature review and the principal characteristics of the study area. The expected results of the MLR are, for example, a negative correlation between the proximity driver from cities and the urban transformation, since the general trend is to build new urban agglomerations close to existing ones so as to minimize/contain the costs of infrastructures and public services (Allen & Lu, 2003; G. Li et al., 2018; Poelmans & Van Rompaey, 2010; Wu et al., 2006). A negative correlation with the proximity drivers from roads is also expected, because the probability of expansion in the vicinity of existing roads, is generally higher (Hu & Lo, 2007; Luo & Wei, 2009; Mustafa, Rienow, et al., 2018).

Drivers related to social and economic factors are continuous data that is distributed by the municipalities and is available through ISTAT. The other drivers are derived from data available on the regional geo portal. In particular, the slope was obtained from the elevation raster. Proximity factors from roads and railway stations were derived from a vector dataset and were calculated using the Euclidean distance method. Roads were classified into three categories: highways, suburban roads, and local roads. Proximity factors also included distances to large (X_7) and medium-sized cities (X_8). Considering the characteristics of the study area, large cities are those with a population > 50,000 inhabitants, i.e. the two provincial capitals (Potenza and Matera); medium-sized cities are those with a population between 10,000 and 50,000 inhabitants, cities that cover an important role at regional level from the point of view of economy, industry, agriculture or tourism. For the urban policy factors, all strict constraints on possible urbanization existing in the territory were considered, such as mountains above 1200m, buffers for rivers, lakes and coastal territories, archaeological heritage, forests, road and rail buffers and hydrogeological risk (0: non-urbanizable area, 1: urbanizable area).

Table 4 Built-up causative factors.

Driver	Name	Datum	Unit	Predictive variables
X_1	Elevation	continuous	meter	Physical factors
X_2	Slope	continuous	percent	
X_3	DISTANCE to Railway Station	continuous	meter	Proximity factors
X_4	DISTANCE to road 01-highways	continuous	meter	
X_5	DISTANCE to road 02-suburban roads	continuous	meter	

X_6	DISTANCE to road 03-local roads	continuous	meter	
X_7	DISTANCE to large city	continuous	meter	
X_8	DISTANCE to medium city	continuous	meter	
X_9	Population density	continuous	percent	Social factors
X_{10}	Employment rate	continuous	percent	Economic factors
X_{11}	Zoning	categorical	Binary (0-1)	Urban policies

All drivers were rasterized with a 100x100 pixel resolution, and standardized, except for the zoning driver which is a binary map.

3.4 Methods

Built-up density maps were used for the calibration phase (1989-1998) and validation phase (1998-2013). Two components were considered for the calculation of the transition potential from one density class to another for the calibration phase. The first component concerns the built-up development causative factors, calibrated with the MLR. The second component was the CA neighborhood effects that were calibrated using a MOGA as in Mustafa et al. (Mustafa, Heppenstall, et al., 2018). The calibrated parameters were used to simulate the built-up pattern of 2013. We then validated our model by comparing the simulated 2013 with the actual 2013.

3.4.1 Transition potential and calibration process

The general goal of the calibration process is to obtain the best set of parameters that are well adapted to the urban expansion of a specific area.

The estimation of the transition potential *(P)* of a cell *(ij)* changing its state from not-built up to one of the other density classes (or from one density class to another) is carried out by Mustafa, Heppenstall, et al., 2018 as follows:

Eq a

$$P_{ij} = \sqrt{(P_c)_{ij}\text{x}(P_n)_{ij}}$$

where $(P_c)_{ij}$ represents the transition probability based on the drivers affecting the expansion process (see Table 4), and $(P_n)_{ij}$ represent the neighborhood effect on the *ij* cells for each density class. The $(P_c)_{ij}$ will be determined using MLR. The latter, is an extension of the binary logistic regression and it returns more than two response categories that are considered simultaneously to describe the relationship between one or more independent variables (Aravkin et al., 2018). In this study MLR was performed for class 0, class 1 and class 2; the dependent variables (*Y*) represent, respectively for the classes analyzed, the change from class 0 to classes 1, 2 and 3, the change from class 1 to class 2 and the change from class 2 to 3. Considering the general form of MLR, the probability of change for each *k* class is obtained by the Eq b:

Eq b

$$((P_c)_{ij}, Y = k_0) = \frac{1}{1 + \exp\left(\alpha_{k_1} + \beta_{k_11}X_1 + \cdots + \beta_{k_1v}X_v\right) + \cdots + \exp\left(\alpha_{k_n} + \beta_{k_n1}X_1 + \cdots + \beta_{k_nv}X_v\right)}$$

...

$$((P_c)_{ij}, Y = k_n) = \frac{\exp\left(\alpha_{k_n} + \beta_{k_n 1}X_1 + \cdots + \beta_{k_n v}X_v\right)}{1 + \exp\left(\alpha_{k_1} + \beta_{k_1 1}X_1 + \cdots + \beta_{k_1 v}X_v\right) + \cdots + \exp\left(\alpha_{k_n} + \beta_{k_n 1}X_1 + \cdots + \beta_{k_n v}X_v\right)}$$

Where: $(P_c)_{ij}$ represents the probability that in a *ij* cell, the dependent variable *Y* changes from the reference class to the specific class k_n; $(X_1 \dots X_v)$ constitutes the set of independent variables; β is the regression coefficient for each independent variable; α is a coefficient representing the intercept between a specific class *kn* and the reference class. In order to obtain the best coefficients, MLR uses the maximum likelihood estimation method.

The MLR provides a set of coefficients β which expresses the relationship between the independent variables and the expansion and densification processes. Using these coefficients, we generated a probability map for each class according to eq b. The goodness of the MLR fit is expressed by the relative operating characteristic (ROC) coefficient, according to which values of the coefficient are between 0.5 (random fit) and 1 (perfect fit) (Pontius Jr & Batchu, 2003; Pontius & Schneider, 2001). A multicollinearity test, variance inflation factor (VIF), was performed to verify that there are no driving forces measuring the same phenomenon among all the independent variables (Belsley, 1991; Montgomery & Runger, 1994; Salem et al., 2019). The presence of multicollinearity between the variables is expressed by VIF values greater than 4. Figure 2 shows all the predictive variables (standardized) for the temporal step 1989-1998 and used in the MLR.
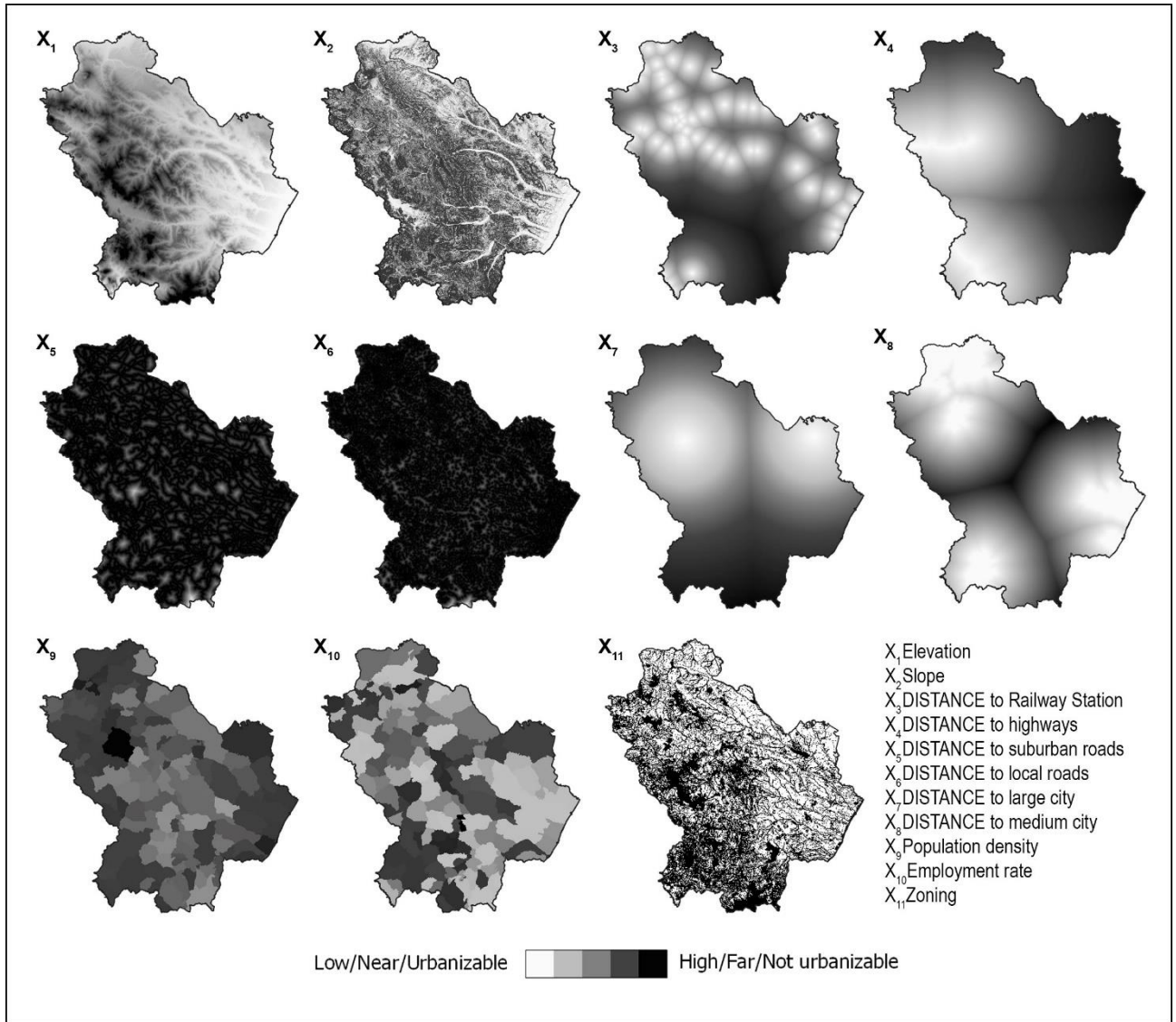
Figure 2 Maps of predictive variables.

The $(P_n)_{ij}$, the neighborhood effect, was evaluated through a 3x3 moving window. It will be determined using the CA approach, with the method proposed by White & Engelen (1997, and already applied in Mustafa, Heppenstall, et al. (2018), expressed with Eq c. Mustafa, Rienow, et al., (2018), Chen et al., (2014) and Poelmans & Van Rompaey, (2009) examined several square sizes and found that the model runs with the 3x3 neighborhood moving window and produced a land-use pattern that best fits the actual pattern. Furthermore, Mustafa, Rienow, et al., (2018) used a 100 m resolution, similar to that of this research, and found that a 3x3 neighborhood moving window outperformed other window sizes (5x5, 7x7, 9x9, and 11x11).

Eq c

$$(P_n)_{ij} = \sum_k \sum_x \sum_d W_{kxd} * I_{kxd}$$

Where: W_{kxd} is the weighting parameter assigned to a cell *ij* with class *k*, which represents one of the building density classes, at position *x* at distance zone *d*; I_{kxd} is 1 if a cell in distance zone d is occupied by class *k* or 0 otherwise. Eq. c parameters are calibrated with multi-objective genetic algorithm (MOGA) for expansion and a genetic algorithm (GA) for densification. The output of the optimization process consists of a set of solutions

defined Pareto front (Sun & Shen, 2008). It is a set of optimal solutions consisting of all the points that are not dominated. The optimal pareto font solution is achieved when the allocation of resources is such that it is impossible to make improvements to the system (Deb, 2001).

For both MOGA and GA the parameters used to set the optimization process are described below. At the initialization process, the algorithm creates a random population of 200 individuals. Each new generation is initiated by selecting parents from the former generation. The algorithm randomly picks 2 individuals, and then selects the best individual according to its fitness score. A new individual is then generated by crossing over a couple of parents according to a binary tournament at 0.8. The Gaussian function was chosen to mutate every new individual by adding a random number to each vector entry of an individual. This random number is taken from a Gaussian distribution centered on zero. The standard deviation of this distribution is controlled with two parameters: the scale parameter (2.00) and the shrink parameter (1.00). The number of generations is flexible. However, we set a convergence stopping criterion when the fitness score for the succeeding 25 generations is less than 0.001. The parameter values that maximize the objective function will be selected as the best calibration outcome. The objective function for the calibration is a fuzzy membership function.

3.4.2   Model performance and fitness function

The model's ability to predict the transition from one density class to another is verified (tested) by comparing the 2013 map obtained with the calibration process (simulated map) with the 2013 actual map. The comparison considers only the new built-up transition from 1998 to 2013 and it is calculated with a fuzziness similarity index calculated as follows (Eq d) (Mustafa, Heppenstall, et al., 2018).

Eq d

$$A_k = \frac{\sum_{x_k \in X_{k,sim}} \left| I_{x_{k0}} \left(\frac{1}{2}\right)^{0/2}, I_{x_{k1}} \left(\frac{1}{2}\right)^{1/2}, \dots, I_{x_{kd}} \left(\frac{1}{2}\right)^{d/2} \right|_{max}}{X_{k,actual}} * 100$$

A_k is the average fuzziness index (between 0 and 100) for each class k; I_{x_k} is 1 if cell i_k in the simulated map at zone d (between 0 and 4) has a similar building density class to one cell at zone d in the actual map, or else 0. $X_{k,sim}$ corresponds to the number of class k changes in the simulated map and $X_{k,actual}$ in the actual map.

4.   Results

In this section the results of the MLR, calibration and validation processes are listed and discussed. The VIF index was used for the multicollinearity test among the independent variables, with values < 2.58 implying no collinearities, so that all the 11 variables presented in Table 4 were introduced in MLR. The X_9 and X_{10} variables refer respectively to the population density and employment rate for 2001. This data is collected at a municipal level and, in particular, the employment rate was calculated by dividing the employed population by the resident population in each municipality.

Figure 3 shows the MLR coefficients with the respective level of significance for each transition analyzed (expansion and densification process) for the period 1989-1998. Since the amount of change was marginal, the densification process from class 1 to 3 has not been considered.

For the expansion process, the most influential factor is the distance to local roads for transition to density class 0 to class 1 and 2 and, the slope for the transition from class 0 to class 3. The most important factors for the densification process, are the slope for transition from class 1 to class 2 and proximity to local roads for

transition from class 2 to class 3. As regards the correlation between the dependent variable and the predictive variables, the correlation with the proximity factors from roads and stations was negative for the expansion processes. Therefore, as anticipated in section 3.3, the probability of cells changing their state from non-urban to urban, increases by decreasing distance to roads and railway stations. This trend, i.e. more urban transformation in the vicinity of train stations and major roads, is usually considered a key component of a sustainable settlement system. Only for the transition from class 1 to class 2 of the densification process, the correlation with the proximity of medium-sized cities was negative. Interestingly, for all other transitions the correlation between urban transformations and proximity variables from cities was positive. Therefore, instead of building new urban agglomerations close to the existing ones to minimize/contain the costs of infrastructure and public services, in this case study the trend is to build further away from existing cities, away from roads and train stations, in an ever increasing scattered pattern that dramatically increases social costs. This is precisely what we defined as urban sprinkling, as opposed to urban sprawl that rather tends to develop and consolidate along existing urban cores even though with lower densities than in central areas.
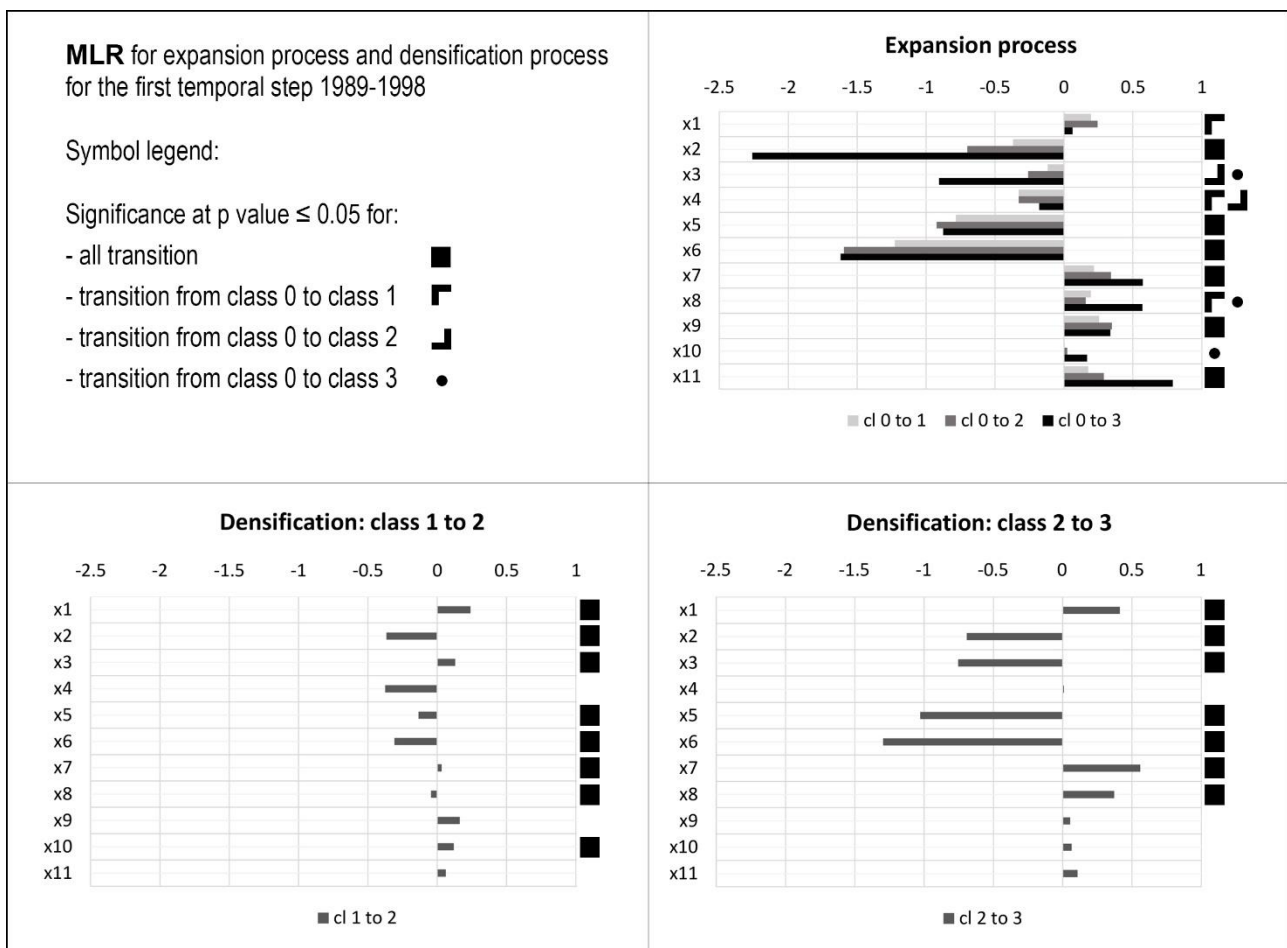
Figure 3 MLR coefficients for the expansion and densification process for the first temporal step 1989-1998.

The correlation with variables X_9 and X_{10}, which represent socioeconomic factors, is positive but poorly correlated for almost all transitions. This is in contrast to other research which has shown that, in both high (Han et al., 2015; Hua et al., 2014) and low settlement density contexts (Vermeiren et al., 2016), the drivers that most influence urban expansion are those related to population density and employment rate.

The correlation between the urban transformation processes and the physical factor of elevation is positive. This is due to the high altimetry in which most of the municipalities of the Basilicata region reside, as highlighted

in the research conducted by (Cillis et al., 2019), rural buildings (abandoned or not) in the Basilicata region are found at an average altitude of 500 meters above sea level.

The ROC values of MLR for the expansion process are: 0.89 for transition from class 0 to class 1 and 2, and 0.85 for transition from class 0 to class 3; for the densification process, the ROC values are 0.67 for densification from class 1 to class 2 and 0.73 for densification from class 2 to 3.

The optimization for the expansion process, stopped at generation 122 and returned a set of Pareto front solutions from which the best solution was selected. The calibrated map of 1989-1998 was used for the validation of the 1998-2013 map. The accuracy rates for the validation process are 55.80%, 34.14% and 23.75% for class 0 to 1, 0 to 2 and 0 to 3 respectively. The GA optimization for densification from 1 to 2 stopped at 58 generation and for densification from 2 to 3 at 61 generation with an accuracy rate for the validation process of 47.64% and 40.47% respectively. Figure 4 shows the optimal weights obtained for all transformation processes that define the neighborhood interaction. The results show that the probability of expansion of the low-density class increases with the expansion of the existing urbanized cells (density classes 1,2 and 3) in particularly class 1. The probability of expansion of class 2 is remarkably correlated with the presence of high-density cells. The expansion of class 3 is strongly correlated with the presence of high-density cells. The probability of transition from class 1 to 2 is positively linked with the existence of low-density cells and negatively with the others. The densification process from class 2 to 3 shows a strong negative relation with non-urbanized and low-density cells and a positive link with medium density cells.
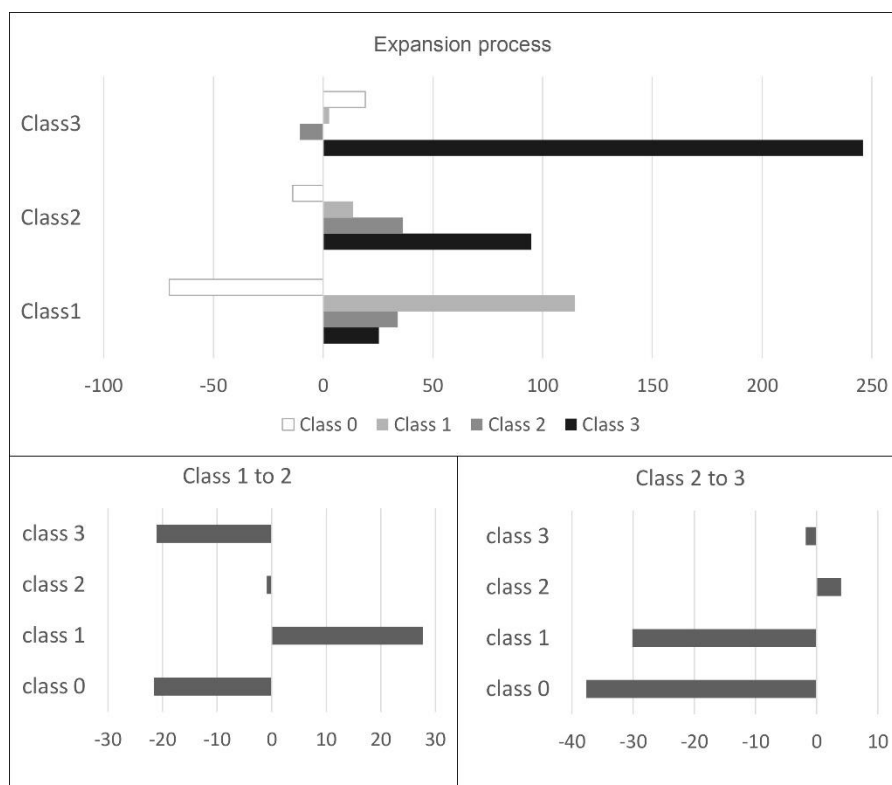
Figure 4 Weighting values of neighborhood parameter values for all transformation processes

These results show that the trend is to develop new buildings near low-density areas increasing fragmentation and fueling urban sprinkling. Furthermore, densification processes take place near low and medium density cells leading to the development of settlements that are located far from high density areas (medium and large cities), and that will require new services and infrastructures.

The results of the 2030 projections show a loss of 7649 hectares of non-urbanized land. This is almost half the administrative surface of the city of Potenza (17543 hectares (*Istat.it*, n.d.)). In the study area, the estimated loss of area between 2013 and 2030 corresponds to 450 hectares per year. This area is scattered over the territory because major changes mainly concern the low-density class - which correspond to urban sprinkling – whose variation amounts to 73% of the total change. Low-density cells increased 17% compared to 2013. In the 2030 prediction, the medium density class is affected by a percentage of change of 22% of the total variation, while the remaining part (5%) concerns class 3. Figure 5 shows some of the most significant places in the study area through a comparison between the actual 2013 maps and those of 2030 as predicted by the model.
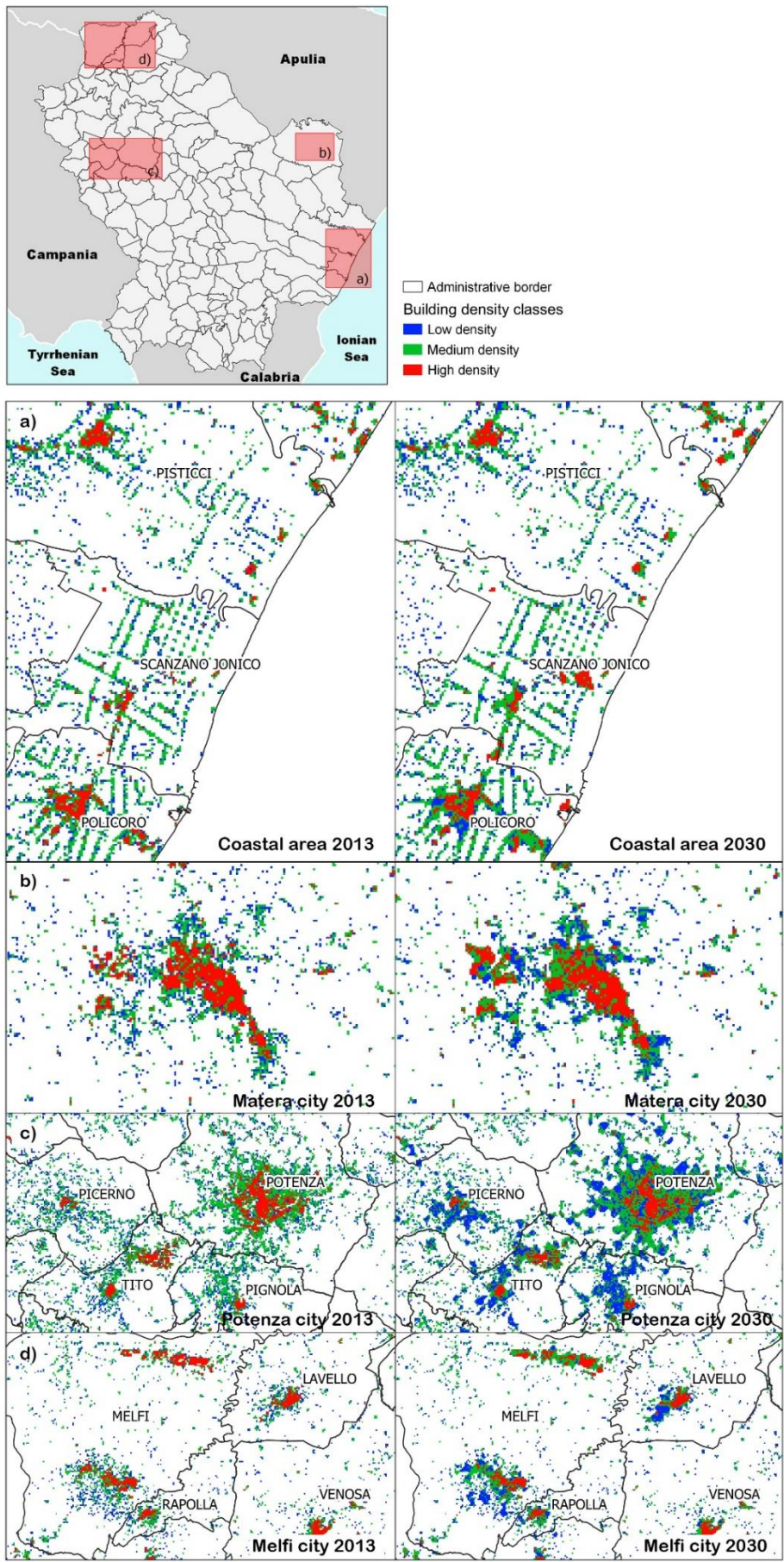
Figure 5 Some examples of the study area. On the left, the maps of 2013 (actual) and on the right those of 2030 (predicted).

A disaggregation of the data on the basis of the two provincial territories (province of Potenza and province of Matera) shows that the province of Potenza undergoes a greater change in terms of absolute values equal to 83% of the total change. The biggest changes concern: (i) the city of Potenza, which in 2030 increased its low-density area by 925 hectares and three near municipalities: Tito, Picerno and Pignola, with an increase in each municipality, of 180, 207 and 139 hectares respectively, of low-density areas (see Figure 5 c); (ii) the municipality of Melfi, located in the extreme north-west of the province of Potenza (see Figure 5 d), affected by a substantial increase (compared to other municipalities) of the medium density areas of 185 hectares (75% of the total changes in the municipality).

For the territory of the province of Matera the total change in 2030 is equal to 17% of the total regional variation. Most changes are located along the coast and mainly concern density classes 2 and 3. With reference to Figure 5a, the major changes from 2013 to 2030 are recorded in the municipality of Policoro and concern the medium density class for an area of 164 hectares. The municipality of Scanzano Jonico, registers an increase in the medium density area of 59 hectares (54% of the total) and an increase in the high-density area of 45 hectares; only 5% of the total change affects the low-density class. In the municipality of Pisticci, instead, the transformations concern, in the same quantity, density class 1 and class 3 for a total of 54 hectares. As regards the city of Matera (Figure 5 b), the provincial capital, the biggest changes concern the low-density classes (211 hectares) and medium density classes (200 hectares), with a total change of 1.19% since 2013,

5. Discussion

The model applied in this study addresses the dynamics of urban expansion through a multi-density approach. It was possible to assess the different relationships between built-up causative factors and built-up density classes for each transition, with specific factors derived from MLR and CA coefficients for each density class. Specifically, it allows to read and analyze the different impacts of drivers on the transformation modalities, especially on the expansion and densification processes.

One of the limitations of this study is that built-up density does not consider building heights and use. This could lead to misreading the expansion process. An example concerns the predicted expansion along the coastal area (figure 5a) which has certainly undergone an increase in expansion due to the development of the tourism sector. In fact, even if it is not the dominant development sector in the Basilicata region, the tourism sector has greatly influenced the expansion of the Ionian coast with the development of numerous tourist resorts and holiday flats (especially in the municipalities of Policoro and Pisticci ((Amato et al., 2014) - figure 5a). However, in the municipality of Scanzano Jonico a strong expansion of the high-density class emerges in the projection map of 2030, resulting from the expansion over the years of agricultural structures (greenhouses, barns and warehouses). In this case a distinction of use of the buildings would have allowed to discriminate the buildings of purely industrial use and would have improved the final result of the 2030 projection. Considering the variable height of the buildings, this would have made it possible to make further considerations on the number of floors per building, on the population density within the building complexes and therefore assess the final projection not only on the shape and size of the building but on its effective capacity.

A possible future research in the construction of the model could consider only residential buildings. In addition, the variables concerning population density and employment rate were used at a municipal level (due to the unavailability of other data). Certainly, more detailed data in the cadastral sections or on the working compartments would give inclusive results. Since the results of the forecast for 2030 mostly concern density

class 1, i.e. the one corresponding to sprinkling, possible developments in this study could concern the application of the model to predict the variation in the sprinkling index.

6. Conclusions

Simulating future urban expansion scenarios is fundamental to understand the spatial model in action in a given territory. It can help policy makers adopt urban policies aimed at containing expansion. The aim of this work has been to project future business-as-usual scenario in urban sprinkling contexts in order to control and regulate urban transformation processes.

An innovative aspect of this study consists in considering urban expansion not in a dual way (built-up/not built-up) but along four density classes (not built-up, low, medium and high density). This multi-density approach allowed to evaluate, through the MLR, the different correlations of each density class with the built-up causative factors taken into consideration. Both the expansion and densification processes were explored. From the results of the MLR a scarce significance of the socioeconomic variables on all the expansion processes has emerged. The observed urban expansion is inversely proportional to the demographic growth. This trend is typical at a national scale not only in the regional context of Basilicata (Amato et al., 2016). This phenomenon leads to new urban expansion in the absence of real estate demand and therefore to an unnecessary waste of land with all its consequences. The results of the MLR show that the most important factor influencing urban expansion is the proximity to roads. In particular, the correlation between the proximity variable from local roads and urban expansion, is most significant for the expansion process (transition from class 0 to all other classes). This corresponds to the typical phenomenon of urban sprinkling in which new settlements tend to be built close to local roads (with less availability of services) and far from the main public service centers with a subsequent increase in social costs (Manganelli et al., 2020; Romano, Zullo, Fiorini, Marucci, et al., 2017).

The results of the CA neighborhood effects, show that the general trend is to construct new buildings (in the low-density class) in the vicinity of low-density areas, which further increases the sprinkling effect. In addition, densification processes take place close to low and medium density areas. In a sustainable urban development, it is appropriate to encourage densification processes close to medium and high-density cells where more services are concentrated, thus limiting further dispersion of urbanized cells on the territory.

From the 2030 simulation, it emerges that the greatest change occurs in the class corresponding to sprinkling (density class 1) and mainly the territory of the province of Potenza. In the territory of the province of Matera, most changes are concentrated along the coast and does not primarily concern sprinkling but, on the contrary, density classes 2 and 3. For this reason, and for other factors linked to the different physical characteristics of the two provincial territories, in terms of altitude, road infrastructure, number of railway stations and territorial extension, it would be appropriate to carry out the analysis on the two territories considering different built-up causative factors. By distributing the total area lost - i.e. the area that has changed from "not built-up" to the other classes of density - for the total regional population in 2019 (562,869 inhabitants (*Istat.it*, n.d.)), we will obtain a loss of about 136 square meters per capita in 2030. This area is subtracted from its natural use at the expense of a non-existent demographic demand.

Starting from the assumption that in the context of the Basilicata region the main transformation dynamics correspond to urban sprinkling and the demographic trend is constantly decreasing - completely eliminating the real demand for the real estate market - we ask the following question: why build a prediction model?

Certainly, a prediction model in this context could be a valid instrument of the decision-making support system to identify, on the basis of prediction: (i) the most vulnerable areas to transform and where to develop ad hoc

regulations to limit transformation or to transfer development in the proximity of areas suitable for transformation (those with a greater presence of public services); (ii) the areas that are susceptible and threatened by anthropic transformations on which to implement measures for the protection and conservation of biodiversity and the naturalness of lands.

7. References

Aburas, M. M., Ho, Y. M., Ramli, M. F., & Ash'aari, Z. H. (2016). The simulation and prediction of spatio-temporal urban growth trends using cellular automata models: A review. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 380–389. https://doi.org/10.1016/J.JAG.2016.07.007

Aburas, M. M., Ho, Y. M., Ramli, M. F., & Ash'aari, Z. H. (2017). Improving the capability of an integrated CA-Markov model to simulate spatio-temporal urban growth trends using an Analytical Hierarchy Process and Frequency Ratio. *International Journal of Applied Earth Observation and Geoinformation*, *59*, 65–78. https://doi.org/10.1016/j.jag.2017.03.006

Allen, J., & Lu, K. (2003). Modeling and prediction of future urban growth in the Charleston region of South Carolina: A GIS-based integrated approach. *Ecology and Society*, *8*(2). https://doi.org/10.5751/es-00595-080202

Amato, F., Maimone, B., Martellozzo, F., Nolè, G., & Murgante, B. (2016). The Effects of Urban Policies on the Development of Urban Areas. *Sustainability*, *8*(4), 297. https://doi.org/10.3390/su8040297

Amato, F., Pontrandolfi, P., & Murgante, B. (2014). *Using Spatiotemporal Analysis in Urban Sprawl Assessment and Prediction* (pp. 758–773). Springer, Cham. https://doi.org/10.1007/978-3-319-09129-7_55

Aravkin, A., Deng, L., Heigold, G., Jebara, T., Kanevski, D., & Wright, S. J. (2018). *Log-linear models, extensions, and applications*. MIT Press.

Arlinghaus, S. L., & Kerski, J. J. (2013). Spatial mathematics: Theory and practice through mapping. In *Spatial Mathematics: Theory and Practice through Mapping*. CRC PRESS. https://doi.org/10.1080/2325548x.2014.901863

Basse, R. M., Omrani, H., Charif, O., Gerber, P., & Bódis, K. (2014). Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography*, *53*, 160–171. https://doi.org/10.1016/j.apgeog.2014.06.016

Belsley, D. A. (1991). A Guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, *4*(1), 33–50. https://doi.org/10.1007/BF00426854

Brueckner, J. K., & Fansler, D. A. (1983). The economics of urban sprawl: theory and evidence on the spatial sizes of cities. *Review of Economics & Statistics*, *65*(3), 479–482. https://doi.org/10.2307/1924193

Brueckner, Jan K. (2001). Urban Sprawl: Lessons from Urban Economics. *Brookings-Wharton Papers on Urban Affairs*, *2001*(1), 65–97. https://doi.org/10.1353/urb.2001.0003

Carruthers, J. I., & Ulfarsson, G. F. (2003). Urban Sprawl and the Cost of Public Services. *Environment and Planning B: Planning and Design*, *30*(4), 503–522. https://doi.org/10.1068/b12847

Caselli, B., Ventura, P., & Zazzi, M. (2020). Performance-based spatial monitoring. An interpretative model for long-term shrinking medium-small Italian towns. *Sustainable Cities and Society*, *53*, 101924. https://doi.org/10.1016/j.scs.2019.101924

Chen, Y., Li, X., Liu, X., & Ai, B. (2014). Modeling urban land-use dynamics in a fast developing city using the

modified logistic cellular automaton with a patch-based simulation strategy. *International Journal of Geographical Information Science*, *28*(2), 234–255. https://doi.org/10.1080/13658816.2013.831868

Chen, Y., Li, X., Liu, X., Ai, B., & Li, S. (2016). Capturing the varying effects of driving forces over time for the simulation of urban growth by using survival analysis and cellular automata. *Landscape and Urban Planning*, *152*, 59–71. https://doi.org/10.1016/j.landurbplan.2016.03.011

Cillis, G., Statuto, D., & Picuno, P. (2019). Vernacular Farm Buildings and Rural Landscape: A Geospatial Approach for Their Integrated Management. *Sustainability*, *12*(1), 4. https://doi.org/10.3390/su12010004

Dang, A. N., & Kawasaki, A. (2016). A Review of Methodological Integration in Land-Use Change Models. *International Journal of Agricultural and Environmental Information Systems*, *7*(2), 1–25. https://doi.org/10.4018/IJAEIS.2016040101

Deb, K. (2001). Multi-objective optimization using evolutionary algorithms. In *IEEE Transactions on Evolutionary Computation* (Vol. 16). John Wiley & Sons. https://doi.org/10.1109/TEVC.2002.804322

Feng, Y. (2017). Modeling dynamic urban land-use change with geographical cellular automata and generalized pattern search-optimized rules. *International Journal of Geographical Information Science*, *31*(6), 1198–1219. https://doi.org/10.1080/13658816.2017.1287368

Feng, Y., & Tong, X. (2020). A new cellular automata framework of urban growth modeling by incorporating statistical and heuristic methods. *International Journal of Geographical Information Science*, *34*(1), 74–97. https://doi.org/10.1080/13658816.2019.1648813

Freilich, R. H., & Peshoff, B. G. (1997). The Social Costs of Sprawl. *The Urban Lawyer*, *29*(2), 183–198. https://doi.org/10.2307/27895056

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Housing Policy Debate*, *12*(4), 681–717. https://doi.org/10.1080/10511482.2001.9521426

Halleux, J. M., Marcinczak, S., & van der Krabben, E. (2012). The adaptive efficiency of land use planning measured by the control of urban sprawl. The cases of the Netherlands, Belgium and Poland. *Land Use Policy*, *29*(4), 887–898. https://doi.org/10.1016/j.landusepol.2012.01.008

Halmy, M. W. A., Gessler, P. E., Hicke, J. A., & Salem, B. B. (2015). Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. *Applied Geography*, *63*, 101–112. https://doi.org/10.1016/j.apgeog.2015.06.015

Han, H., Yang, C., & Song, J. (2015). Scenario Simulation and the Prediction of Land Use and Land Cover Change in Beijing, China. *Sustainability*, *7*(4), 4260–4279. https://doi.org/10.3390/su7044260

Hasse, J. E., & Lathrop, R. G. (2003). Land resource impact indicators of urban sprawl. *Applied Geography*, *23*(2–3), 159–175. https://doi.org/10.1016/J.APGEOG.2003.08.002

Hu, Z., & Lo, C. P. (2007). Modeling urban growth in Atlanta using logistic regression. *Computers, Environment and Urban Systems*, *31*(6), 667–688. https://doi.org/10.1016/J.COMPENVURBSYS.2006.11.001

Hua, L., Tang, L., Cui, S., & Yin, K. (2014). Simulating Urban Growth Using the SLEUTH Model in a Coastal Peri-Urban District in China. *Sustainability*, *6*(6), 3899–3914. https://doi.org/10.3390/su6063899

Istat.it. (n.d.). Retrieved April 5, 2019, from https://www.istat.it/

Ku, C. A. (2016). Incorporating spatial regression model into cellular automata for simulating land use change. *Applied Geography*, *69*, 1–9. https://doi.org/10.1016/j.apgeog.2016.02.005

Li, C., Zhao, J., & Xu, Y. (2017). Examining spatiotemporally varying effects of urban expansion and the

underlying driving factors. *Sustainable Cities and Society*, *28*, 307–320. https://doi.org/10.1016/j.scs.2016.10.005

Li, G., Sun, S., & Fang, C. (2018). The varying driving forces of urban expansion in China: Insights from a spatial-temporal analysis. *Landscape and Urban Planning*, *174*, 63–77. https://doi.org/10.1016/J.LANDURBPLAN.2018.03.004

Li, Shi, Duan, Chen, Wang, & Hao. (2019). Spatiotemporal Decoupling of Population, Economy and Construction Land Changes in Hebei Province. *Sustainability*, *11*(23), 6794. https://doi.org/10.3390/su11236794

LIU, C., & MA, X. (2011). Analysis to driving forces of land use change in Lu'an mining area. *Transactions of Nonferrous Metals Society of China*, *21*, s727–s732. https://doi.org/10.1016/S1003-6326(12)61670-7

Luo, J., & Wei, Y. H. D. (2009). Modeling spatial variations of urban growth patterns in Chinese cities: The case of Nanjing. *Landscape and Urban Planning*, *91*(2), 51–64. https://doi.org/10.1016/J.LANDURBPLAN.2008.11.010

Manganelli, B., Murgante, B., & Saganeiti, L. (2020). The Social Cost of Urban Sprinkling. *Sustainability 2020, Vol. 12, Page 2236*, *12*(6), 2236. https://doi.org/10.3390/SU12062236

Martellozzo, F., Amato, F., Murgante, B., & Clarke, K. C. (2018). Modelling the impact of urban growth on agriculture and natural land in Italy to 2030. *Applied Geography*, *91*, 156–167. https://doi.org/10.1016/J.APGEOG.2017.12.004

Martinez-Fernandez, C., Audirac, I., Fol, S., & Cunningham-Sabot, E. (2012). Shrinking Cities: Urban Challenges of Globalization. *International Journal of Urban and Regional Research*, *36*(2), 213–225. https://doi.org/10.1111/j.1468-2427.2011.01092.x

McGrath, D. T. (2005). More evidence on the spatial scale of cities. *Journal of Urban Economics*, *58*(1), 1–10. https://doi.org/10.1016/j.jue.2005.01.003

Montgomery, D. C., & Runger, G. C. (1994). Applied Statistics and Probability for Engineers. In *European Journal of Engineering Education* (Vol. 19, Issue 3). https://doi.org/10.1080/03043799408928333

Mustafa, A., Bruwier, M., Archambeau, P., Erpicum, S., Pirotton, M., Dewals, B., & Teller, J. (2018). Effects of spatial planning on future flood risks in urban environments. *Journal of Environmental Management*, *225*, 193–204. https://doi.org/10.1016/J.JENVMAN.2018.07.090

Mustafa, A., Cools, M., Saadi, I., & Teller, J. (2017). Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM). *Land Use Policy*, *69*, 529–540. https://doi.org/10.1016/j.landusepol.2017.10.009

Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., & Teller, J. (2018). Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computers, Environment and Urban Systems*, *67*, 147–156. https://doi.org/10.1016/J.COMPENVURBSYS.2017.09.009

Mustafa, A., Rienow, A., Saadi, I., Cools, M., & Teller, J. (2018). Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. *European Journal of Remote Sensing*, *51*(1), 391–401. https://doi.org/10.1080/22797254.2018.1442179

Mustafa, A., Van Rompaey, A., Cools, M., Saadi, I., & Teller, J. (2018). Addressing the determinants of built-up expansion and densification processes at the regional scale. *Urban Studies*, *55*(15), 3279–3298. https://doi.org/10.1177/0042098017749176

Nasiri, V., Darvishsefat, A. A., Rafiee, R., Shirvany, A., & Hemat, M. A. (2019). Land use change modeling through an integrated Multi-Layer Perceptron Neural Network and Markov Chain analysis (case study: Arasbaran region, Iran). *Journal of Forestry Research*, *30*(3), 943–957. https://doi.org/10.1007/s11676-018-0659-9

Nechyba, T. J., & Walsh, R. P. (2004). Urban sprawl. *Journal of Economic Perspectives*, *18*(4), 177–200. https://doi.org/10.1257/0895330042632681

Poelmans, L., & Van Rompaey, A. (2009). Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A case study in the Flanders–Brussels region. *Landscape and Urban Planning*, *93*(1), 10–19. https://doi.org/10.1016/J.LANDURBPLAN.2009.05.018

Poelmans, L., & Van Rompaey, A. (2010). Complexity and performance of urban expansion models. *Computers, Environment and Urban Systems*, *34*(1), 17–27. https://doi.org/10.1016/J.COMPENVURBSYS.2009.06.001

Pontius Jr, R. G., & Batchu, K. (2003). Using the Relative Operating Characteristic to Quantify Certainty in Prediction of Location of Land Cover Change in India. *Transactions in GIS*, *7*(4), 467–484. https://doi.org/10.1111/1467-9671.00159

Pontius, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, *85*(1–3), 239–248. https://doi.org/10.1016/S0167-8809(01)00187-6

Puertas, O. L., Henríquez, C., & Meza, F. J. (2014). Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago Metropolitan Area, 2010-2045. *Land Use Policy*, *38*, 415–425. https://doi.org/10.1016/j.landusepol.2013.11.024

Rienow, A., & Goetzke, R. (2015). Supporting SLEUTH - Enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, *49*, 66–81. https://doi.org/10.1016/j.compenvurbsys.2014.05.001

Romano, B., Zullo, F., Fiorini, L., Ciabò, S., & Marucci, A. (2017). Sprinkling: An Approach to Describe Urbanization Dynamics in Italy. *Sustainability*, *9*(1), 97. https://doi.org/10.3390/su9010097

Romano, B., Zullo, F., Fiorini, L., Marucci, A., & Ciabò, S. (2017). Land transformation of Italy due to half a century of urbanization. *Land Use Policy*, *67*, 387–400. https://doi.org/10.1016/j.landusepol.2017.06.006

RSDI – Geoportale Basilicata. (n.d.). Retrieved March 6, 2019, from https://rsdi.regione.basilicata.it/

Saganeiti, L., Favale, A., Pilogallo, A., Scorza, F., Murgante, B., Saganeiti, L., Favale, A., Pilogallo, A., Scorza, F., & Murgante, B. (2018). Assessing Urban Fragmentation at Regional Scale Using Sprinkling Indexes. *Sustainability*, *10*(9), 3274. https://doi.org/10.3390/su10093274

Salem, M., Tsurusaki, N., & Divigalpitiya, P. (2019). Analyzing the Driving Factors Causing Urban Expansion in the Peri-Urban Areas Using Logistic Regression: A Case Study of the Greater Cairo Region. *Infrastructures*, *4*(1), 4. https://doi.org/10.3390/infrastructures4010004

Scorza, F., Saganeiti, L., Pilogallo, A., & Murgante, B. (2020). Ghost planning: the inefficiency of energy sector policies in a low population density region. *Archivio Di Studi Urbani e Regionali*, *127*, 34–55. https://doi.org/10.3280/ASUR2020-127-S1003.

Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., & Taleai, M. (2017). Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, *64*, 297–308. https://doi.org/10.1016/j.compenvurbsys.2017.04.002

Shoyama, K., & Yamagata, Y. (2014). Predicting land-use change for biodiversity conservation and climate-change mitigation and its effect on ecosystem services in a watershed in Japan. *Ecosystem Services*, *8*, 25–34. https://doi.org/10.1016/J.ECOSER.2014.02.004

Song, W., & Liu, M. (2014). Assessment of decoupling between rural settlement area and rural population in China. *Land Use Policy*, *39*, 331–341. https://doi.org/10.1016/J.LANDUSEPOL.2014.02.002

Sun, Y., & Shen, G. (2008). Improved NSGA-II multi-objective genetic algorithm based on hybridization-encouraged mechanism. *Chinese Journal of Aeronautics*, *21*(6), 540–549. https://doi.org/10.1016/S1000-9361(08)60172-7

van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S., & Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software*, *82*, 174–182. https://doi.org/10.1016/J.ENVSOFT.2016.04.017

Verbeek, T., Boussauw, K., & Pisman, A. (2014). Presence and trends of linear sprawl: Explaining ribbon development in the north of Belgium. *Landscape and Urban Planning*, *128*, 48–59. https://doi.org/10.1016/j.landurbplan.2014.04.022

Vermeiren, K., Vanmaercke, M., Beckers, J., & Van Rompaey, A. (2016). ASSURE: a model for the simulation of urban expansion and intra-urban social segregation. *International Journal of Geographical Information Science*, *30*(12), 2377–2400. https://doi.org/10.1080/13658816.2016.1177641

Wahyudi, A., & Liu, Y. (2016). Cellular Automata for Urban Growth Modelling: *International Review for Spatial Planning and Sustainable Development*, *4*(2), 60–75. https://doi.org/10.14246/irspsd.4.2_60

White, R., & Engelen, G. (1997). Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B: Planning and Design*, *24*(2), 235–246. https://doi.org/10.1068/b240235

Wiechmann, T., & Pallagst, K. M. (2012). Urban shrinkage in Germany and the USA: A Comparison of Transformation Patterns and Local Strategies. *International Journal of Urban and Regional Research*, *36*(2), 261–280. https://doi.org/10.1111/j.1468-2427.2011.01095.x

Wu, Q., Li, H., Wang, R., Paulussen, J., He, Y., Wang, M., Wang, B., & Wang, Z. (2006). Monitoring and predicting land use change in Beijing using remote sensing and GIS. *Landscape and Urban Planning*, *78*(4), 322–333. https://doi.org/10.1016/J.LANDURBPLAN.2005.10.002

Yang, J., Gong, J., Tang, W., & Liu, C. (2020). Patch-based cellular automata model of urban growth simulation: Integrating feedback between quantitative composition and spatial configuration. *Computers, Environment and Urban Systems*, *79*, 101402. https://doi.org/10.1016/j.compenvurbsys.2019.101402

Yang, Y., Zhang, D., Nan, Y., Liu, Z., & Zheng, W. (2019). Modeling urban expansion in the transnational area of Changbai Mountain: A scenario analysis based on the zoned Land Use Scenario Dynamics-urban model. *Sustainable Cities and Society*, *50*, 101622. https://doi.org/10.1016/j.scs.2019.101622

Appendix A

Table A Collection of scientific articles on urban expansion models from 2014 to 2020. The symbol ■ indicate the symbol indicates the presence of the considered predictive variable.

Authors	Study case	Application	Models and techniques	Predictive variables						Population trend and urban dynamics
				Physical factors	Proximity factors	Social factors	Economic factors	Neighborhood interaction	Urban policies or constraints	
(Hua et al., 2014)	South-east coast of China	Simulate urban growth and historical urban expansion in a coastal city.	SLEUTH	■				■	■	Population increase (duplicated in 12 years – from 2000 to 2012)
(Puertas et al., 2014)	Santiago metropolitan area – Latin America	Spatial dynamics of urban expansion and prediction	Cellular automata, Markov chain and Logistic regression	■ Elevation	■ Proximity to: main roads, population centers and commercial centers	■ Population variation	■ Real estate factors		■ Availability land for urbanization	Population increase Annual average growth rate: 1,01%

Reference	Study area	Aim	Method								
(Basse et al., 2014)	Luxemburg with a greater area of border	Simulation of Land Use Land Cover change	Cellular automata and artificial neural networks	■ Slope	■ Proximity to: highway, bus station, train station			■ Neighborhood (3x3): water urban, industrial, forest, commercial, agricultural.			Population increases (not available specific information)
(Rienow & Goetzke, 2015)	Germany, 2 areas with high density urbanization (Cologne and Düsseldorf); 1 rural area with dispersed settlement patterns (Bergisches Land)	Combine the simulation strengths of SLEUTH and SVM	SLEUTH, binomial logistic regression and Support Vector Machines (SVM)	■ Elevation, slope	■ Proximity to: airport, big city, highway river	■ Population density, variation population density, migration rate	■ Land price, unemployment, number of jobs in different sector, income, per capita net dwelling area and change, cars	■	■		Increasing population in Cologne and Düsseldorf. Population decline in Bergisches Land
(Halmy et al., 2015)	Egypt	Prediction of land use change in 2011 and project changes in 2023 by extrapolating current trends.	Markov – Cellular automata		■ Proximity to: road, coast, residential centers, irrigation canal system					■	

(Han et al., 2015)	Beijing, China	Predict and simulate the evolution tendency of land use spatial patterns under two different scenarios: development and protection.	CLUE-S (Conversion of Land Use and its Effects at Small regional extent) and Markov models	■ Elevation, slope	■ Proximity to: river, urban center, local road, railway, highway	■ Population density, urbanization	■ GDP			Built up area expansion of about 3 time from 1980 to 2013. The population rate is continuously increase.
(Chen et al., 2016)	Shenzhen, China	Simulation (2013) and prediction (2018) of the urban expansion.	SA-Patch-CA (Survival Analysis-Patc-Cellular Automata)	■ Slope	■ Proximity to: airport, city center, highways, railway station road networks, town centers, port					Rapid urbanization and urban expansion

(Ku, 2016)	Tainan City, Taiwan	Model simulation through the integration of spatial regression into the CA model	Spatial regression and CA		■ Proximity to: roads, schools, existing settlement	■ Population density		■ Neighborhood (5x5)	■ Land use policies	Population growth with a density of 4375,58 inhabitants per square kilometers
(Vermeiren et al., 2016)	Kampala, Uganda (Africa)	Simulate urban expansion and social segregation which aims to serve as a tool for decision support in the development of rapidly expanding cities. ASSURE: Agent-based simulation of social segregation and urban expansion)	Agent-based model (on the basis of 4 group: extremely poor, poor, middle income, rich)	■ Elevation, slope, wetland	■ Proximity to: roads, city centre	■ Population density	■ Middle income	■ Random		Horizontal city with relatively low population density and population increase.

Reference	Location	Aim	Models							Notes
(Mustafa et al., 2017)	Wallonia region, Belgium	Simulation of urban expansion between 1990 and 2000 through a hybrid urban expansion model (HUEM)	Logistic regression, Cellular Automata and Agent-Based model (on the basis of 3 group: developer, farmer and planning permission authority)	■ Slope	■ Proximity to: cities, roads (4 kinds),		■ Access to job		■ Zoning	Urban sprawl, not significative increase in population
(Shafizadeh-Moghadam et al., 2017)	Tehran, Iran	Simulate urban expansion through the comparison of six land use change models	Artificial Neural Networks, Support Vector Regression, Random Forest, Classification and Regression Trees, Logistic Regression and Multivariate adaptive Regression Splines.	■ Elevation, slope, Northing and Easting parameters	■ Proximity to: agriculture areas, open lands, built-up areas, green spaces, roads.			■ Neighborhood (5x5)		Very fast growing in population and artificial land. High density population.

Reference	Location	Aim	Method	■	■	■		■	■	
(Aburas et al., 2017)	Negeri, Malaysia	Simulating and predicting urban expansion trend	Cellular Automata Markov Chain based on Analytical Hierarchy Process (AHP) and on Frequency Ratio (FR)	Elevation, slope, soil texture, Land cover	Proximity to: commercial area, educational area, residential area, industrial area, roads highway, railway, power line and rivers	Population density				Given the geographical location of the study area (proximity to cities with great economic power) it is expected to become an incubator of urban development in Malaysian peninsular.
(C. Li et al., 2017)	Xuzhou, China	Analysis of the relationship between the driving factors and urban expansion.	Geographically weighted regression (GWR), and cellular automata (CA)	Slope	Proximity to: district centres, major roads, minor roads.	Population density		Neighborhood variable	Environmental protection areas, subsidence areas	Large population density and rapid growth of urban area

Reference	Country	Objective	Model / Method	Physical	Proximity	Population	Economic	Neighborhood	Policy / Exclusion	Results
(Martellozzo et al., 2018)	Italy	Investigate the loss of natural and agricultural land to urban areas.	SLEUTH and AHP	■				■	■ Exclusion layer building through multi criteria decision makers (MCDM) process	Population increase of about 4% (from 1990 to 2012).
(G. Li et al., 2018)	China	Study of drivers of urban expansion and their effects across different regions in China in different periods and at different scales.	Spatial Error Probit Model (SEM - Probit) and Spatial Log Probit Model (SAR – Probit)	■ Elevation, slope	■ Proximity to: lake, river, city center, country center, highway, national way, railway	■ Population density	■ GDP	■ Neighborhood (3x3): urban area		Rapid growth in population and in built-up area
(Mustafa, Heppenstall, et al., 2018)	Wallonia region, Belgium	Model to simulate built-up expansion and densification process.	Multinomial Logistic Regression, Cellular Automata and Genetic Algorithm	■ Elevation, slope	■ Proximity to: roads (4 kinds) railway station, large sized cities		■ Employment rate and Richness index	■ Neighborhood (3x3): for density classes	■ Zoning	Expansion rate of built-up area very low

(Y. Yang et al., 2019)	Changbai Mountain (TACM), ChinaSCS	Modelling urban expansion in different scenarios	Shared Socioeconomic Pathways (SSPs) and the zoned Land Use Scenario Dynamics-urban (LUSD-urban)	■ Elevation, slope	■ Proximity to: National roads, provincial roads, railways, city center	■ Urban population		■		Rapid growth of urban population and urban area
(J. Yang et al., 2020)	Ezhou, China	The model links the quantitative composition of a territory with the spatial configuration of urban expansion by incorporating a trade-off mechanism that adaptively regulates the combined suitability of the event for non-urban terrain types based on transition intensity analysis.	Cellular Automata and Time Monte Carlo (TMC)	■ Elevation, slope, soil quality	■ Proximity to: lake, city center, highway entrance, railway station, adjacent city, sub center, town center		■ Kernel density of Point of Interest (POI)	■ Neighborhood (3x3)	■	Very fast growth of urban land and with uncontrolled manner. Growth of population

(Feng & Tong, 2020)	Shanghai, China	Model to simulate urban epansion	Cellular automata, statistical analysis and heuristics analysis	■ Elevation	■ Proximity to: city centers, road networks, subway	■ Population density	■ GDP			Very rapid urban expansion (3,8% per year)