A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales

Dimitris Gounaridis, Ioannis Chorianopoulos, Elias Symeonakis, Sotirios Koukoulas

Science of the Total Environment, 646, DOI: 10.1016/j.scitotenv.2018.07.302

Accepted: July 2018

1. Introduction

Land use and land cover (LULC) changes are considered to be the most prominent influence of humans on the environment. Technological and medical advancements have brought about unprecedented increases in the human population and, consequently, in the need for access to resources. This need has in turn caused substantial and growing transformations to the Earth's surface (Vitousek et al., 1997) with often undesirable impacts and magnitudes that vary from local to global scales. The dual role of humans to actively contribute to LULC changes and, at the same time, be on the receiving end of experiencing the consequences of these changes, emphasizes the need for a better understanding of the human-LULC change nexus.

A wide variety of LULC change models have been developed to meet the scientific community needs for understanding how and why LULC evolves (Schrojenstein Lantman et al. 2011). Generally, LULC models are widely used to analyze the complex structure of linkages and feedbacks between drivers of change, determine their relevance to particular changes and project how much land is used where and for what purpose, under different predefined attributes and conditions. This type of information is then adopted in a meaningful way in order to support policy decision making related to land-use (Mallampalli et al. 2016). However, by definition, LULC models can not exactly replicate the complex interactions and nonlinear relations which are apparent in LULC systems. At a fundamental level, they are, rather, a process that provides a platform that, allows computer experiments to be undertaken (Brown et al. 2013). When the system in question is simple, the processes and interactions that characterize it can be easily determined and the results are somehow expected, while projections and other kinds of extrapolations are a straightforward task. When dealing, however, with inherently complex systems, as is the case with LULC changes, the models are able to represent and exemplify only a small fraction of the whole mechanism in order to highlight important processes.

The recent methodological and technological advancements have paved the way for more articulated LULC models which are able to answer more complex questions. Such questions could be in regard to what the possible outcomes would be if alternative pathways were followed, or which outcome is the most desirable from a list of alternatives, as well as a diverse range of other 'what-if' scenarios. Increasingly, scenario-based analysis is now being adopted by a range of disciplines pertaining to LULC change, as fruitful experiments for exploring the possible future trajectories of historical and current trends (Murray-Rust et al. 2013). Considering that the number of potential futures is actually infinite (Greeuw et al. 2000), scenarios are not used to predict the future in a precise manner, but to explore possible future directions and to consider a range of alternative pathways. To do so, the scenario-based analysis fully recognizes the infinity of potential futures and attempts to focus only to an understandable and manageable set of alternatives. This is achieved by delineating plausible, presumably coherent and internally consistent storylines of different socio-economic development trajectories (Rounsevell and Metzger, 2010).

When modelling LULC, the scale, the spatial resolution and the extent of the study area are important attributes of all spatially explicit models (Agarwal et al 2002). The term scale refers to the spatial, temporal, quantitative, or analytic dimension used to measure and study the processes that are modelled (Gibson et al., 2000). Scale also involves the terms extent and resolution: extent refers to the magnitude of a dimension used in measuring (e.g. study area boundaries on a map), whereas resolution refers to the precision used in this measurement (e.g. pixel size) (Gibson et al. 2000). Moreover, resolution refers not only to spatial resolution, but also to thematic, which is the level of precision in LULC categories. In addition, the term temporal resolution is used to refer to the time span and frequency of the analysis. Modelling LULC changes, therefore, requires a range of scales to be defined since it is a phenomenon that involves multiple processes that act over different scales. At each scale, different processes have a dominant influence on the outcome (Meentemeyer, 1989; Van Delden et al. 2011). Approaches that do not implement a multi-scale approach are prone to aggregation or oversimplification errors and thus fail to reproduce cross-scale interactions. This is due to the fact that features and processes that operate at local scales are not always observable when dealing with larger areas and coarser spatial resolution data (Verburg et al. 2004). On the other hand, studies that focus solely on the local level often fail to incorporate information about the general context which can only be derived from coarser spatial resolution data (Larondelle & Lauf, 2016). Given that all models are driven by their input data, studies focusing on specific LULC processes, considering only a single scale and using data that are particularly suitable only to a certain area, are not representative, transferable or reproducible to different scales. Therefore, such approaches are characterized by higher levels of uncertainty and depend on a number of critical assumptions (Kok and Veldkamp 2001, Van Delden et al. 2011; Veldkamp et al., 2001, Verburg et al. 2006). Moreover, it is a common assumption that the modelling results are highly affected by the quality and the technical details, such as the pixel size of inputs and the bias they entail (Kocabas & Dragicevic, 2006; Van Delden et al. 2011).

Models designed to analyze LULC dynamics can be divided into categories according to their perspective, their domain, the methodological framework they apply, their spatial or non-spatial nature etc (literature reviews by Agarwal et al., 2002, Briassoulis 2000; Schrojenstein Lantman et al. 2011). However, LULC models that solely rely on statistical approaches often suffer from limitations such as sensitivity to outliers and noise, collinearity issues and factors compatibility (Dormann et al. 2013; Eastman et al. 2005). On the other hand, more recently, a variety of models pertaining to artificial intelligence, such as agent-based models, have been successfully applied for addressing the complex, non-linear behavior of human-nature interactions and decision making. This type of models, however, are suitable to capture processes at the individual, household or neighborhood levels and when it comes to agent behavior they can be very complex and are often parametrized with qualitative social survey data and other types of participatory approaches (Zagaria et al. 2017)

Cellular automata (CA) consist of a dynamic simulation framework where space is represented as a grid of cells and time is considered as discrete unit. The basic principle of this type of LULC modeling framework is that the state of a given pixel is determined by taking into account its previous state, the spatial interactions with the surroundings in a given neighborhood and a set of defined transition rules. These elements dictate the possible change of a cell and can be expert-based or calculated from statistical analysis of historical LULC changes (White and Engelen, 2000). A growing body of the literature demonstrate that, although very simple, CA models have the strong ability to represent rich LULC patterns and handle nonlinear, stochastic and spatially explicit LULC processes (Sante et al 2010).

The biggest advantage of CA is that they are fully consistent with Geographic Information Systems (GIS) and remote sensing. Additionally, CA can be coupled with other types of models and thus they are flexible to allow the elaboration and extension of the methodological procedures according to the needs of a case study (Aburas et al. 2016). For instance, CA have been previously combined with a plethora of modeling frameworks such as Markov chains (Jokar et al. 2013), neural networks (Li and Yeh, 2002) support vector machines (Yang et al. 2008) and kernel-based methods (Liu et al. 2008) among others. More recently, CA have been successfully combined with Random Forests (RF) (Kamusoko and Gamba, 2015; Gounaridis et al. 2018a).

RF is a tree structured machine learning algorithm that generates a "forest" of randomized independent to each other and identically distributed decision trees. Each individual tree is composed with a random selection of the predictor variables and by searching across a randomly selected subset, it predicts the target response, casting a unit vote. This process is repeated until a user-defined number of trees has been built. The outputs are determined from the majority of votes by each individual tree. For a full detailed description of the RF algorithm, theory and applications, the reader is referred to Breiman (2001). The independency of each individual decision tree and the randomness in forming subsets of the input data makes RF insensitive to outliers, to noise and to overfitting (Chan and Paelinckx, 2008). Additionally, normal distribution of inputs is not a prerequisite and thus it can handle heterogenous data from various sources, units and scales (Gounaridis et al., 2014; Gounaridis et al., 2016; Gounaridis and Koukoulas, 2016). Another important advantage of RF is that it can handle large datasets with thousands of imputs being accurate and at the same time computationally faster (Rodriguez-Galiano et al. 2012).

The aim of this paper is, therefore, to explore potential future LULC dynamics in the Attica region, using a CA modelling approach with scenarios that reflect different economic performance realities and alternative planning options. The central premise is to simulate all categories of LULC change at the regional level and to evaluate the effects of different proximate and underlying causes. In order to spatially associate the spatial determinants (proxies) with the observed historical changes, a set of factors derived from multiple sources and expressed in different scales, units and resolutions are incorporated in the modelling framework. A multi-resolution sensitivity analysis is also carried out to assess the effect of spatial resolution of the input data to the model outputs. The results will quantify the importance of various spatial determinants (proxies) of change and shed light to the effect different economic performance realities and land-use planning choices can have on the landscape.

2. Study area

The study area is the region of Attica in mainland Greece, an example of the rapid socio-economic transformations that occurred in the country during the last decades, including the demographic dynamics and population redistribution. The region includes Athens, the capital city of Greece and the country's major economic hub. According to the latest census (2010), the region of Attica is inhabited by about 4 million people, or 35% of the country's total population. In more recent decades, economic and population growth triggered a persistent increase in housing demand and supply, and the redistribution of middle-class Athenians to the outskirts of Athens (Mantouvalou et al, 1995; Leontidou et al. 2007). Additionally, socio-economic conditions favoured a persistent amenity-driven trend for second homes along the coastal zone, albeit within a commuting distance from the city-centre (Arapoglou and Sayas, 2009). As a consequence, the landscape of peri-urban Athens has changed substantially. The urban growth trend was indirectly emboldened by the weak presence of land use planning checks and controls, which permitted the unhindered development at any environmental, social or long-term economic cost (Pagonis, 2013). Moreover, after successfully attracting national and foreign funds, and in preparation to host the Olympic Games of 2004, the demand for construction sites to accommodate commercial, industrial, transportation and recreational activities further increased the built-up transformation of the urban periphery (Chorianopoulos et al. 2010). After the phase of a rather stable economic growth, however, the area has recently been exposed to the negative consequences of the sovereign debt crisis and the succeeding economic recession (2010-2016). The decrease in purchasing power and a drastic drop in consumer demand affected both the housing and the construction industries (Gounaridis et al. 2018a).

Figure 1 about here

In terms of its topography, Attica also constitutes an interesting study case since it is characterised by an undulated morphology (Figure 1). Mount Parnitha (elevation 1413m), Pateras (elevation 1132m), Penteli (elevation 1109m), Hymettus (elevation 1026m) and Egaleo (elevation 468m) are the main mountain ranges. These geomorphological features separate the city of Athens from the adjacent flat districts of Thriasio, Messoghia and Marathonas (Figure 1), which are the only available areas to host residential and industrial settlements.

3. Material and methods

3.1 LULC Data

Five Landsat-based LULC maps spanning 25 years (1991, 1999, 2003, 2010, 2016) at 30m spatial resolution were used for the modelling. These maps were generated by devising a semi-automated sampling extraction based on a context that combined the no-change areas, spectral controlling, and prior knowledge of the area (Gounaridis et al. 2018b). Overall accuracy for all maps is above 90%. Most importantly, the maps come with a very high thematic resolution, achieved after disaggregating the urban-related LULC categories (Gounaridis & Koukoulas, 2016). Specifically, the maps depict eight land cover categories: i) continuous urban fabric, ii) discontinuous dense urban fabric, iii) discontinuous medium density urban fabric, iv) discontinuous low density urban fabric, v) industrial, commercial and transport units, vi) arable land and permanent crops, vii) forests, scrubs and other natural areas and viii) other (includes open spaces bare, mines and inland water bodies).

3.2 Transition probability modelling

Exploring future LULC patterns is a useful experiment for evaluating the causes and identifying the impact of LULC changes. The scenario-based simulations have been proven to be a useful way to sketch out how LULC patterns can evolve under different pathways with a level of plausibility (Greeuw et al. 2000; Rounsevell and Metzger, 2010). Scenario-based analyses involves a certain degree of uncertainty originating from the very nature of socioeconomic predictions that help define the scenarios. This is due to the inability to foresee any unexpected circumstances and integrate any emerging discontinuities or the data inputs for the models. Especially when dealing with complex systems, such as LULC changes, assumptions are unavoidable. The level of uncertainty can be minimized by combining an empirical analysis and sketching different scenarios, attributes and conditions that deviate from historic trends in LULC changes (Brown et al. 2013; Verburg et al. 2016)

3.2.1 Predictor variables

Taking into account previous LULC change modelling efforts (Gounaridis et al. 2018a), as well as data availability, a suite of 27 variables were chosen to best describe the LULC change processes that took place throughout Attica in the study period (1991-2016; Table 1). They are both categorical and continuous in nature and cover a broad spectrum of potential LULC change factors. They can act as spatial determinants of the changes that occurred during the last decades in Attica, and are derived from multiple sources, with different scales and resolutions.

Table 1 about here

During the study period, changes related to artificial surfaces were dominant in Attica and, therefore, the majority of the chosen variables represent factors that affect the decision-making process when selecting locations for the construction of new housing or infrastructure. Factors pertaining to social shifts, economic motives, inherent quality and attractiveness of a given place and proximity to basic needs and amenities were assumed to play a key role (Table 1). Variables related with the topography of the terrain, such as elevation, slope and aspect influence the inherent quality of a certain location and define the land suitability for built-up expansion. Proximity to the sea, to "blue-flag" beaches (Foundation for Environmental Education- http://www.fee.global/), as well as to natural reserves or urban green spaces are also perceived as added value in the pursuit for a better quality of life and aesthetics for both primary or secondary homes. Proximity to the city center of Athens or to nearest towns, to public transport, and the road density are proxies that reflect the commuting distance to work. Additionally, distance to social infrastructure including, among others, health provision, education and sports facilities, together with the density of private enterprises (kernel density of geo-tagged newly developed enterprises at 30m spatial resolution) serve as proxies to amenities. Demographic and socio-economic variables such as changes in population density, employment and unemployment rates provide insights on the shifts in the socio-economic profile of the area estimated at local authority (municipal) level.

It is worth noting that, variables available at a higher administrative level, that of the region, were not included since in Greece, implementation of local land use management policies falls under the remit of local municipal authorities. Factors expressed at the municipal level, therefore, were considered to represent an appropriate spatial unit for our analysis (Panori et al. 2016; Gounaridis et al. 2018a). All spatial and non-spatial datasets were collated in a GIS environment. Census data were mapped at the municipal level while distances were computed using the Euclidean distance function. The variables were then converted (resampled with bilinear interpolation) to 30m spatial resolution rasters to match the resolution of the Landsat-based land cover classifications (Gounaridis et al. 2018b).

3.2.2 Leap-frog development index (LFDI)

To enhance the accuracy of the model, and to ensure the accurate detection and representation of scattered development, the Leap-frog development index (LFDI), originally proposed by Xu et al. (2007) was calculated and included in the modelling scheme. Leap-frog development refers to the new urban patches that are formed spontaneously and have no direct spatial connection and shared boundaries with the existing urban patches. The index applies to artificial LULC types and has been proved to effectively delineate any type of scattered development, classifying the historical changes according to sharing boundaries properties (Gounaridis et al. 2018a).

Specifically, the index is calculated after dividing the length of the common boundaries between newly developed urban patches and already existing urban patches with the perimeter of the newly developed urban patches Xu et al. (2007). When the resulting value is higher than 0.5 the growth type is denoted as infilling. A resulting value lower than 0.5 denotes the edge growth while when the result is 0, it denotes the absence of a shared boundary, and the growth is identified as Leap-frog development. Therefore, following the approach by Gounaridis et al (2018a) the maps of 1991 and 2016 were converted to vector format and patches representing the four urban categories and industrial commercial and transport units were assigned values denoting which patches appeared in each date. Subsequently, using common functions in GIS, the length of common boundaries, their perimeter and the index were calculated. The last step was to convert the vector file to raster format at 30m spatial resolution.

3.2.3 Random Forests regressions

Following the approach adopted by Gounaridis et al. (2018a), the transition probability surfaces were generated by employing the machine learning regression algorithm of Random Forests (Breiman, 2001) using all variables, including the LFDI. Eighteen possible transitions were identified (Table 2), under three assumptions: (a) it is impossible for the urban fabric class to convert to any other land type as well as to decrease in density; b) the industrial, commercial and transport units cannot convert to any other land type, and c) the “other” category, that includes inland waters, bare land and mines, cannot interact with any of the other 7 classes. To train each of the 18 models, 5000 random points were dispersed throughout the study area. Two possible values were associated with these training points: 1 denotes change from any LULC class to any other class, and 0 denotes no change. The RF regressions were then implemented in R using the RandomForest package (Liaw & Wiener, 2002). To fine tune the RF regressions, five predictor variables (equal to the square root of the total number of 27 predictor variables) were used for each tree split and 700 trees for each run. The modelling process generated 18 transition probability surfaces, each indicating the degree of potential future LULC change.

Table 2 about here

RF also offer meaningful metrics about the importance of each predictor variable. To quantify the importance and contribution for each of the 27 predictor variables, two metrics, the Mean Decrease Gini and the Mean Decrease Accuracy were computed (Gounaridis and Koukoulas, 2016). The mean decrease in Gini coefficient informs about each variable's contribution to the impurity of the resulting random forest model. Variables with a high value in the decrease of Gini, tend to have nodes with high purity which is a measure of model's homogeneity. The mean decrease in accuracy provides information about how much the accuracy would decrease if a variable were excluded from the model. Therefore, the larger the value of mean decrease, the higher the importance of a variable is.

3.3 LULC change scenarios

Figure 3 shows the LULC trends in Attica between 1991 and 2016, based on the Landsat-based land cover maps. Three different phases of economic development and performance can clearly be identified and based on these, we devised the following three potential future scenarios:

Figure 3 about here

<u>Low development scenario</u>: this scenario reflects the 2010-2016 period, when urban expansion rates curtailed significantly as a consequence of economic recession and a drop in investment spending. Throughout this time, for instance, approximately 150,000 newly built houses in the region were left uninhabited (unsold), while over one

third of commercial facilities in the city of Athens closed down and remained shut (Serraos et al., 2016). Under this scenario, economic growth, as well as the population mobility would remain relatively stable.

<u>Medium development scenario:</u> this scenario reflects the period between 1991 and 1999, when the peri-urban areas of Athens conurbation, especially the uplands and the Messoghia plain, experienced significant population gains. Increase in demand for new houses boosted urban growth at the expense of other less profitable land uses, bringing gradually major changes in the peri-urban landscape. In fact, during this time, peri-urban Athens population had grown ten times faster than the Athens conurbation population, which remained relative stagnant (Petrakos and Mardakis, 1999). According to this scenario, peri-urban Athens experiences a steady population increase, fueled by the relocation choices of Athenians questing residence in lower density areas.

<u>High development scenario</u>: this scenario reflects the sharp urban expansion rates noted in the region in the 2000-2009 period, facilitated by stable economic growth and the continuation of a rather "loose" approach to land use planning controls. The era is chronologically framed by the effects of the 2008 global financial crisis, which were felt locally, however, in late 2009, in the form of an excessive budgetary deficit and a prolonged recession (Chorianopoulos and Tselepi, 2017). Under this scenario, population decentralization from Athens conurbation continues apace, further enhanced by labour migration from outside the country, fueling demand for new housing constructions. Following the development trends of the reference period, the spotlight of investment falls on the waterfront areas shifting further real estate dynamics towards tourism-related facilities and secondary homes. Similarly, spatial planning continues to play an important role in the "construction spree" by approving investment in transportation infrastructure, enhancing peri-urban accessibility to the city of Athens.

Consequently, major infrastructure works, private enterprises and shopping centres will keep on colonizing the Northern outskirts (Maranthon, Oropos, Messoghia and the Thriasio plains forming a 'suburban exploitation thesis' case (Pacione, 2009).

All three scenarios draw from clear reference periods and assume that profound social and political changes will not alter their traits. As far as the land use planning apparatus is concerned, it is assumed that it will continue to be rather permissive to development, following a political stance that approaches unregulated urban expansion as a "shortcut" to economic growth.

3.4 Model calibration

The CA model was designed and implemented using the Dinamica EGO platform (Soares-Filho et al. 2002). An important step, prior to the prediction phase, is model calibration. To calibrate the model and evaluate the goodness of fit, a comparison of simulated maps with reference maps is the most efficient way (Gounaridis et al. 2018a). Any CA modelling framework involves four components: the probability maps, the historical LULC maps, the transition rules and the neighborhood characteristics that define the parameters of the simulation.

The CA model was trained based on the 1991-2010 period, and the observed changes were used to predict the landscape structure and composition on 2016. To do so, the annual rates of change per LULC category between 1991 and 2010 were calculated generating a transition matrix. In order to replicate the actual structure and composition of the area, three landscape metrics were computed: (i) the mean patch size, (ii) the variance of patch size, and (iii) the patch isometry. In general, an increased patch size results in less fragmented landscapes, while the patch size variance denotes the diversity of newly developed patches. Isometry usually varies from 0 to 2 and thus, the greater the isometry, the more isometric (i.e. equal) the newly developed patches are. The first two metrics were computed for the input LULC map (2010) while the latter was adjusted through a trial-and-error process. Finally, the 18 transition probabilities were stacked together to drive the allocation of cells, based on the premise that the cells with the highest likelihood values should change first. The model was then set to run and predict LULC for 2016.

To evaluate the model's performance, the simulated LULC map of 2016 was compared with the observed LULC map of 2016 (i.e. the outcome of the Landsat-based classification; Gounaridis et al. 2018b) using the fuzzy similarity index at multiple resolutions (Hagen, 2003). This index evaluates the accuracy of simulation results considering the similarities of two maps (simulated versus observed) in a neighborhood context and within increasing window sizes (Mas et al. 2012). This involves the comparison of map fit and spatial agreement within a certain pixel vicinity allowing the comparisons of maps not only in a strict pixel-by-pixel basis but also considering the spatial similarity in multiple resolutions (Hagen, 2003). To gain insights about per class agreement we also computed the error matrix between the simulated and the observed maps of 2016. The sampling was based on 9399 samples holding LULC class values of 2016 (Gounaridis et al. 2018b). The samples come with relatively equal distribution among the LULC classes that ensures equal representation.

After calibration, the simulation of LULC changes under the three scenarios was implemented, taking 2016 as the initial year and 2040 as the final year, in a 5-year time step. The parameters used to calibrate the model were kept constant and only the quantity of LULC transitions per scenario were changed. A transition matrix was constructed for each epoch, i.e. 1991-1999, 1999-2010 and 2010-2016, to reveal the quantity of each possible transition per scenario (Table 2). Ideally, the predictor, and in turn, the transition probability surfaces, would also change per scenario, to better reflect the socio-economic conditions of each epoch. However, in our case, this option was not feasible due to data availability and temporal mismatch issues.

3.5 Multi-resolution sensitivity analysis

After completing the model simulations at 30m spatial resolution, a sensitivity analysis was also conducted at various spatial resolutions. It was hypothesized that when all other parameters of the model are held constant and only the spatial resolution of inputs changes, then the quantities, the spatial allocation and thus, the spatial patterns of outputs, can differ. The central premise behind this step was that the spatial resolution of the models' inputs can have important and substantial effects on the output. Thus, this parameter can limit or even enhance the ability of a model to project future scenarios of LULC change. Sensitivity analysis is a process that examines the variation in model outputs in response to variation in a set of model parameters, in this case the spatial resolution of input data. To do so, the 1991 and 2016 Landsat-based classifications were resampled (nearest neighbor) to 100m, 250m and 500m, respectively and change detection was performed for each case. Next, the transition probabilities were re-constructed through RF regression after resampling (bilinear interpolation) all predictors for each case. The calibration followed the same steps as aforementioned. The landscape metrics along with the transition quantities were re-calculated and introduced to the models for each case. After calibration, each scenario was simulated based on the transitions observed throughout each of the three epochs. Finally, all maps generated from each run were overlapped using rule-based cross classification in order to produce the final map per scenario. This step identified areas of change that are common regardless the spatial resolution of the inputs. To explore the influence of the spatial resolution on various consecutive steps of the modelling process, we compared the transition probability surfaces produced at the native resolution (30m) and at several coarser resolutions (100m, 250m, 500m). This was done after sampling the transition probability surfaces at 1000 random points, and computing the concordance correlation coefficient (Lin, 1989; 2000).

4. Results and discussion

4.1 Model calibration and performance

One common way to assess the level of model calibration and performance is to compare the simulated map for a given year versus the observed map, which is often derived from the classification of satellite data. Figure 4 depicts the resulting map of 2016 after calibration versus the reality (observed map of 2016). A visual comparison of these maps shows the relatively high similarity. This suggests that the RF-CA model was relatively accurate at allocating the LULC patterns of change in the study area. Table 3 reveals the level of agreement per class between the simulated map of 2016 and the observed map of 2016. Overall accuracy was acceptably high (88.36%) and the User and Producer accuracies for all classes ranged from 83.4% to 96.5%. Regarding the disagreements, confusion is evident between certain classes that are mostly spatially adjacent. For instance, between "discontinuous medium density urban fabric" and "discontinuous low density urban fabric", as well as between "discontinuous low density urban fabric" and "arable land and permanent crops".

Figure 4 about here

Table 3 about here

Figure 5 illustrates the fuzzy similarity index computed based on the overlay of the two maps. The accuracy assessment yielded a spatial fit of 85.18% within the 1x1 window size radius which improved to 95.08 % when widened to a 15x15 window size. The high scores in performance suggest that the suite of 27 predictor variables were used efficiently and the RF algorithm performed well with an adequate fit.

Figure 5 about here

Figure 6 depicts the components of agreement and disagreement between the simulated versus the observed maps. It reveals information about: (i) observed change simulated correctly as change (i.e. hits); (ii) observed persistence (i.e. LULC that remained unchanged) simulated correctly as persistence (i.e. null successes); (iii) observed change simulated incorrectly as persistence (i.e. misses), and (iv) observed persistence simulated incorrectly as change (i.e. false alarms). Most importantly, the model predicted accurately the leap-frog development and this proves the added value of the LFDI and the extensive training of the RF model.

Figure 6 about here

4.2 Variable importance

Figure 7 is the Mean Decrease in Gini coefficient which informs about each variable's contribution to the impurity of the resulting random forest model. Road density, enterprises density and elevation contributed the most for changes related to dense urban fabric. The same variables along with the distance to shoreline and education centers are the most related to discontinuous dense and medium density urban fabric. For the discontinuous low density urban fabric, which is a category broadly related to second homes, distance to shoreline, to blue flag beaches, elevation, road density and enterprises density were the most influential variables.

Figure 7 about here

Figure 8 is the mean Decrease in Accuracy which informs about how much the accuracy decreases if a variable would be excluded from the model. According to this, road density, distance to natural reserves, to prefecture center and to shoreline, as well as slope and elevation were the most influential variables for changes related to dense urban fabric. The same variables along with the distance to beaches, to urban green areas and to public buildings were the most influential to changes related to discontinuous dense and medium density urban fabric. For the discontinuous low density urban fabric, the elevation, slope, road density along with the distance to urban green, to shoreline, to natural reserves and to prefecture center contributed the most into the spatial changes description.

Figure 8 about here

4.3 Multi-resolution sensitivity analysis

The models yield similar patterns for each scenario but, as anticipated, as the resolution increases, the patterns tend to become more aggregated and smaller patches of change tend to be lost. Figure 9 depicts the concordance correlation coefficient (Lin, 1989, 2000) derived from transition probabilities for the continuous urban fabric class per different spatial resolution. The higher concordance value can be observed between the 30m and 100m pixel size. Gradually, as the difference in spatial resolution increases so is the distance of data's reduced major axis from the line of perfect concordance which reflects the concordance between the transition probability surfaces.

Figure 9 about here

The multi-resolution sensitivity analysis results provide evidence that the technical characteristics have substantial impact to the outputs of a model and thus to the observed patterns and to the conclusions drawn. Even if a model is rigorously calibrated, the predictability will decrease relative to the spatial resolution, and the patterns revealed in the results will become less informative.

4.4 Model predictions for 2040

Figures 10-12 depict the LULC changes projection under the three scenarios while Figure 13 provides a quantified insight to the final results.

Figure 10 about here

Under the medium economic development scenario (Figure 10), and with a pace of urban growth equivalent to that of 1991–1999, artificial surfaces are expected to expand predominantly at the expense of other, less profitable, land uses. Urban areas are anticipated to reach 41% of the region's surface, of which 17% will be discontinuous low density urban fabric. Industrial areas are expected to occupy almost 8% of the total area. At the same time agricultural land is expected to decline from 23.5% in 2016 to 10% in 2040 (Figure 13). Most changes will occur along the waterfront and in the periphery of Athens conurbation, effecting notable changes in Messoghia, the Thriassian plain, Marathonas, Oropos and Sounio. In these areas, pre-existing urban and industrial clusters portray a tendency to become denser and to expand considerably, ending up almost connected with Athens conurbation, especially in its northern parts. The region's coastline, especially in Messoghia, Marathonas, Oropos and Sounio, is also expected to exhibit remarkable changes. Existing towns display a tendency to become denser and to expand, transforming waterfront areas into a large and solid low density urban patch. Leap-frog development is also expected to increase sharply around road junctions of existing urban areas.

Figure 11 about here

Under the high economic development scenario (Figure 11), where the pace of urban growth reflects the traits of the 1999-2010 period, artificial surfaces are expected to increase remarkably. At the same time, they are expected to occupy more than half of the total surface of Attica region (56.7%). In more detail, urban uses, are expected to occupy an area of almost 48% in 2040, which can be translated to an increase of approximately 21%. In this land use category, discontinuous low density urban fabric will reach a high peak of almost 21% of the total area. The continuous dense, discontinuous high density urban fabric and discontinuous medium density urban fabric are expected to reach 9%, 10% and 8% respectively. At the same time agricultural areas are expected to decrease by 18%, occupying only 5.2% of the total area (Figure 13). All these accelerated landscape transformations are expected to occur throughout Attica region, leading to a mosaic of mixed land uses. Pre-existing urban and industrial clusters will become denser and expand considerably. In a similar fashion with the medium growth scenario, most changes are observed along the coastline and to the peri-urban zones of Athens conurbation. Changes are expected to be centered on the northern suburbs of Athens, the Messoghia and the Thriassian plain, Marathonas, Oropos and Sounio areas. Most notably, existing urban patches in the waterfront (Marathonas, Messoghia, Sounio, western Attica and Oropos), are expected to be linked with the conurbation forming an urban-rural continuum of low, and at places, medium density. In the western part of Attica, the Thriassian plain is expected to experience a considerable increase in industrial development and a notable increase in medium density urban use. Last but not least, the density of urban areas will increase sharply, especially in the northern and eastern suburbs of Athens.

Figure 12 about here

Under the low development scenario (Figure 12), an increase in artificial surfaces of approximately 6% is noted in the region; a rate, however, that is significantly lower if compared with the other two scenarios. Discontinuous low density urban fabric, for instance, is expected to occupy 15% of the total area by 2040, an increase of only 3% since 2016. Similarly, continuous dense and discontinuous high density urban fabric are expected to reach 6.7% and 5.6% respectively (Figure 13). Following the traits of the recession (2010-2016), urban expansion is observed throughout the region, yet at relatively moderate rates and in rather compact form. Foreseen changes will mostly occur around the road network and in the waterfront areas, particularly in the eastern and northern parts of Attica. Already existing urban areas appear to increase in density, rather than expanding, while leap-frog development is noted in areas of adequate transportation infrastructure, guaranteeing ease of access to Athens. Regarding urban density, slightly higher densities are expected in the northern suburbs of Athens.

Figure 13 about here

5 Discussion

5.1 RF-CA land use/cover modelling

The coupling of CA and RF proved to be a sound way to combine the advantages of each approach. Implementing the RF algorithm for transition potential modelling, allows the efficient combination of qualitative and quantitative data derived from multiple sources and with different nature in terms of scale and origin. In addition, RF proved insensitive to collinearity issues and normal distribution of data was not a prerequisite. The predictors incorporated in the models proved capable to spatially determine the phenomenon while the incorporation of the Leap-frog development index at the regional level, assisted the models in LULC prediction. In this approach, a total of 18 distinct transitions were identified and equal transition probability surfaces were generated. Their combination in a CA modelling environment seemed challenging and required intense training and calibration through trial and error. Currently, most LULC models can only simulate limited possible transitions due to complexity in definitions, attributes and transition rules (Liu et al. 2017). However, in reality, even in the same location, different LULC dynamics occur simultaneously and affect each other. Thus, a comprehensive outlook of these processes is much more effective in order to realistically determine the future trajectories. The interactions and competition among different types of LULC was explored by using a simple, yet effective competition mechanism. This approach allows the incorporation of various LULC transition probabilities as a single layer stack, containing all the probability surfaces. Each layer represented one single possible transition, while each cell contained values denoting the dominant LULC type and the likelihood to retain the current land type or transform to another type. The reproduction of LULC patterns and the calibration procedure, as a whole, improved considerably with the inclusion of the mean patch size, the variance of patch size, and patch isometry. Introducing these metrics to the CA framework, allows the models to take into account and to reproduce the actual parameters of the study area. The adoption of the fuzzy similarity index (Hagen, 2003) for assessing the model's spatial fit was another advantage of the approach, as it performs comparisons of simulated versus observed data within a neighborhood context, and not in a strict per-pixel context.

5.2 LULC predictions for 2040

We employed three socioeconomic and associated urban growth scenarios to explore potential LULC pathways to 2040. The 'low development' scenario draws from the current economic austerity and recession reality, framing a long-term setting in which economic downturn keeps on hindering urban expansion dynamics. Results obtained from the medium and high economic development scenarios, however, are multifaceted. Both scenarios shed light on the ways in which Attica would look like when the current economic crisis is reversed. Against this backdrop, they point to the critical role of land use planning in regulating urban expansion. Our results outline a future landscape shaped by the unmediated prerogatives of rapid economic development. They also underscore the significant socio-economic consequences such as enhanced residential segregation, high infrastructure investment costs, central areas underfunding. Moreover, the environmental impact such as increased car dependency and usage, loss of agricultural land and natural habitats are also evident. In light of the consequences, LULC changes that would occur locally are expected to create a maladaptive and nonfunctional setting, liable to undermine future economic development prospects (Chorianopoulos, et al. 2010).

Regarding the factors that contribute and the extent of this contribution to the different types of LULC change, our study incorporated a total of 27 variables into the modelling. By implementing 18 different models representing every permitted LULC transition, the contribution of each factor was quantified using the Mean Decrease Gini and the Mean Decrease Accuracy metrics. From the application of these models, three messages emerge:

(a) Firstly, the results demonstrate that depending on the LULC type, different factors play a key role in the spatial configuration of LULC change (Kizos et al. 2018). The interrelationships of urban related classes, for example, can clearly be distinguished according to their density, which translates to different residential use (e.g. secondary homes). In densely built urban areas, spatial factors, such as road network density, density of enterprises, proximity to social infrastructure (health services, educational institutions) and accessibility to the municipal centres, were the dominant determinants of change. In urban areas with lower density, distance to the shoreline and to "blue-flag" beaches were among the most important. The results are in agreement with the findings of other studies, especially with those related with the coastal zone of the Mediterranean (Boavida-Portugal et al. 2016; Houet et al. 2016; Lagarias, 2012; Marraccini et al. 2015; Petrov et al. 2009).

(b) Secondly, some factors that rank among the top determinants for a type of LULC change, may have a strong positive or negative correlation coefficient with the phenomenon. For instance, the slope and elevation variables, rank high in the urban categories with strong negative correlation coefficients, mostly due to the topography of Attica, limiting the majority of settlements within the plains.

(c) Finally, a possible important limitation that should be noted is that, all these patterns and numbers are case-specific and the conclusions drawn from the quantitative insights might not be transferable to other regions. This is mostly due to specificities present only in Attica, for example, the physical constraints related to the topography, the cultural choices for primary and secondary housing, or the presence or absence of a regulatory planning mechanism. Future research directions should include cross-cases comparison with areas that share common characteristics with Attica, e.g. coastal areas, Mediterranean administrative regions that include a big metropolitan area and areas with rapid socio-economic changes.

5.3 Multi-resolution sensitivity analysis

This paper demonstrated the importance of a multiple scales analysis, by incorporating in the modelling framework data derived from multiple sources, expressed at various scales and resolution. Given that, the data used as input in any model, affect the outcomes, and in turn the usefulness and the accuracy of the model, studies that utilize only data that concern a single scale or spatial resolution, fail to account for a wide range of information. Moreover, their transferability is limited (Veldkamp et al. 2001). Data expressed at coarse scales might hold information and patterns that are undetectable at finer scales and vice versa (Brown et al. 2013; Van Delden et al. 2011; Verburg et al. 2004). Furthermore, factors that determine a LULC change, might operate at a distance from the area of focus. Thus, when dealing with a system that involves multiple nonlinear relationships and various proximate and underlying factors, it is necessary to consider all available information (Larondelle & Lauf, 2016). Here, we exploited all possible resources and efficiently combined and integrated the available multi-scale and multi-resolution data.

Additionally, the simulation results were subjected to a multiple resolution sensitivity analysis. Since the modelling approaches generate outputs that are more or less driven by the parameters and characteristics of input data (Kocabas & Dragicevic, 2006; Van Delden et al. 2011), the results obtained by this approach are consistent to all pixel sizes and thus insensitive to the effect of pixel size.

4. Conclusions

This paper demonstrated an integrated approach to explore potential future LULC dynamics under different scenarios that reflect different economic performances and policy options. Our integrated framework was able to sufficiently: i) take into account socioeconomic, biophysical, legislative and land use factors spanning a broad spectrum of LULC change spatial determinants (proxies); ii) provide insights into hidden patterns by taking into account, not only the prominent changes between major LULC categories, but also changes in density; iii) take into account the multiple scales involved in LULC systems, and, v) provide results that are insensitive to the spatial resolution of the inputs.

Acknowledgments

The authors wish to thank the anonymous reviewers for their constructive feedback and suggestions on an earlier version of this manuscript.

References

Aburas, M. M., Ho, M. Y., Ramli, M. F., & Ashaari, 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.

Agarwal, C, Green, G M, Grove, J M, Evans, T P & Schweik, C M. (2002). A review and assessment of land-use change models: dynamics of space, time, and human choice. General Technical Report NE-297, Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA.

Arapoglou, V.P. and Sayas, J. (2009). New facets of urban segregation in southern Europe. *European Urban and Regional Studies*, 16(4), 345-362.

Batty, M., Couclelis, H., & Eichen, M. (1997). Urban systems as cellular automata. *Environment and Planning B*, 24(2), 159–164.

Boavida-Portugal, I., Rocha, J. & Ferreira, C.C. (2016). Exploring the impacts of future tourism development on land use/cover changes. *Applied Geography,* **77**, 82-91.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

Briassoulis, H. (2000). Analysis of Land Use Change: Theoretical and Modeling Approaches. In the Web Book of Regional Science (Loveridge S. Ed.), West Virginia University, Morgantown.

Brown, D.G., Goovaerts, P., Bumlckl, A. & Li, M-Y. (2002). Stochastic Simulation of Land-Cover Change Using Geostatistics and Generalized Additive Models. *Photogrammetric Engineering & Remote Sensing*, 68(10), 1051-1061.

Brown, D.G., Verburg, P.H., Pontius Jr, R.G. & Lange, M.D. (2013). Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Environmental Sustainability*, 5, 452–457.

Chan, J.C.W. & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost treebased ensemble classification and spectral band selection for ecotope mappingusing airborne hyperspectral imagery. *Remote Sensing of Environment*, 112 (6), 2999–3011.

Chorianopoulos, I., Pagonis, A., Koukoulas, S. and Drymoniti, S. (2010). Planning, competitiveness and sprawl in the Mediterranean city: the case of Athens. *Cities*, 27 (4), 249-259.

Chorianopoulos, I. and Tselepi, N. (2017) Austerity urbanism: Rescaling and collaborative governance policies in Athens. *European Urban and Regional Studies*, DOI: 10.1177/0969776417733309

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46.

Dongjie, G., Weijun, G., Kazuyuki, W. & Hidetoshi, F. (2008). Land use change of Kitakyushu based on landscape ecology and Markov model. *Journal of Geographical Science*, 18, 455–468.

Eastman, J.R., Solorzano, L.A., & van Fossen, M.E. (2005). Transition potential modeling for land-cover change. In *GIS, Spatial Analysis, and Modeling*; Maguire, D.J., Batty, M., Goodchild, M.F., Eds.; ESRI Press: California, UK, 357–385.

Gibson, C.C., Ostrom, E. & Ahn, T.K. (2000). The concept of scale and the human dimensions of global change: a survey. *Ecological Economics*, 32, 217–239.

Gounaridis, D., Zaimes, N.G. & Koukoulas, S. (2014). Quantifying spatio-temporal patterns of forest fragmentation in Hymettus Mountain, Greece. *Computers, Environment and Urban Systems*, 46, 35–44.

Gounaridis D., Apostolou A. & Koukoulas S. (2016). Land Cover of Greece, 2010: a semi-automated classification using Random Forests. *J Maps,* 12(5), 1055-1062.

Gounaridis, D. & Koukoulas, S. (2016). Urban land cover thematic disaggregation, employing datasets from multiple sources and RandomForests modeling. *International Journal of Applied Earth Observation and Geoinformation*, 51, 1–10.

Gounaridis, D., Chorianopoulos, I. & Koukoulas, S. (2018a). Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: The case of Athens. *Applied Geography* 90, 134-144.

Gounaridis, D., Symeonakis, E., Chorianopoulos, I. & Koukoulas, S. (2018b). Incorporating density in spatio-temporal land use/cover change patterns: the case of Attica, Greece. *Remote Sensing* 10(7), 1034.

Greeuw, S., van Asselt, M., Grosskurth, J., Storms, C., Klomp, N., Rothman, D. et al. (2000). Cloudy Crystal Balls: An Assessment of Recent European and Global Scenario Studies and Models: Experts' Corner Report. Office for Official Publications of the European Communities.

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3), 235–249.

He, Z. & Lo, C. (2007). Modeling urban growth in Atlanta using logistic regression. *Computers Environment and Urban Systems,* 31 (6), 667-688.

Houet T., Marchadier C., Bretagne G., Moine M.P., Aguejdad R., Viguié V. et al. (2016). Combining narratives and modelling approaches to simulate fine scale and long-term urban growth scenarios for climate adaptation. *Environmental Modelling and Software journal*, 86, 1-13

Jokar A, J., Helbich, M., Kainz, W. & Darvishi B, A. (2013). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observations and Geoinformation*, 21, 265–275.

Kamusoko, C. & Gamba, J. (2015). Simulating urban growth using a random forest cellular automata (RF-CA) model. *ISPRS International Journal of Geo-Information*, 4, 447–470.

Kizos, T., Verburg, P.H., Bürgi, M., Gounaridis, D., Plieninger, T. Bieling, C. & Balatsos, T. (2018). From concepts to practice: combining different approaches to understand drivers of landscape change. *Ecology and Society,* 23, 25-36.

Kocabas, V. & Dragicevic, S. (2006). Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers Environment and Urban Systems,* 30, 921-953.

Kok, K. & Veldkamp, A. (2001). Evaluating impact of spatial scales on land use pattern analysis in Central America. *Agriculture. Ecosystems and Environment*, 85, 205–221.

Lagarias, A. (2012). Urban sprawl simulation linking macro-scale processes to micro-dynamics through cellular automata, an application in Thessaloniki, Greece. *Applied Geography*, 34, 146-160.

Larondelle, N. & Lauf, S. (2016). Balancing demand and supply of multiple urban ecosystem services on different spatial scales. *Ecosystem Services*, 22, 18–31.

Leontidou, L., Afouxenidis, A., Kourliouros, E. and Marmaras, E. (2007). Infrastructure-Related Urban Sprawl: Mega-Events and Hybrid Peri-Urban Landscapes in Southern Europe, in Urban Sprawl in Europe: Landscapes, Land-Use Change & Policy (eds C. Couch, L. Leontidou and G. Petschel-Held), Blackwell Publishing Ltd, Oxford, UK.

Li, X. & Yeh, A. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, *16*, 323–343.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.

Lin, L. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45, 255 - 268.

Lin, L. (2000). A note on the concordance correlation coefficient. *Biometrics* 56, 324 - 325.

Liu, X., Li, X., Shi, X., Wu, S. & Liu, T. (2008). Simulating complex urban development using kernel-based non-linear cellular automata. *Ecological Modelling*, 211, 169- 181.

Liu, X., Liang, X., Li, X., Xu, X., Oua, J., Chen, Y. et al. (2017). A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning,* 168, 94-116.

Ma, X. & Zhao, X. (2017). Land Use Allocation Based on a Multi-Objective Artificial Immune Optimization Model: An Application in Anlu County, China. *Sustainability,* 7, 15632-15651.

Mallampalli, V.R., Mavrommati, G., Thompson, J., Duveneck, M., Meyer, S., Ligmann-Zielinska, A., et al. (2016). Methods for translating narrative scenarios into quantitative assessments of land use change. *Environmental Modelling & Software,* 82, 7-20.

Manson, S.M. (2005) Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture, Ecosystems and Environment*, 111, 47–62.

Mantouvalou M., Mavridou M. and Vaiou, D. (1995). Processes of social integration and urban development in Greece: Southern challenges to European unification. *European Planning Studies* 3(2), 189-204.

Marraccini, E., Debolini, M., Moulery, M., Abrantes, P., Bouchier, A., Chery, J-P. et al. (2015). Common features and different trajectories of land cover changes in six Western Mediterranean urban regions. *Applied Geography*, 62, 347-356.

Mas, J-F., Perez-Vega, A. and Clarke, K.C. (2012). Assessing simulated land use/cover maps using similarity and fragmentation indices. *Ecological Complexity* 11, 38–45.

McBurney, P. (2012). What Are Models for? 9th European Workshop on Multi-Agent Systems. In: Cossentino M. et al. (Eds) LNAI 7541. Springer, Maastricht, The Netherlands, 175–188.

Meentemeyer, V. (1989). Geographical perspectives of space, time, and scale. *Landscape Ecology,* 3, 163–173.

Millington J.D. (2005) Wildfire risk mapping: considering environmental change in space and time. *Journal of Mediterranean Ecology*, 6, 33–42.

Pacione, M. (2009) Urban Geography: A Global Perspective. Oxon, Routledge.

Pagonis, A. (2013). The evolution of metropolitan planning policy in Athens over the last three decades: Linking shifts in the planning discourse with institutional changes and spatial transformation. Changing Cities: Spatial, morphological, formal and socioeconomic dimensions. University of Thessaly. Skiathos, 18-21 June.

Panori, A., Ballas, D. & Psycharis, Y. (2016). SimAthens: A spatial microsimulation approach to the estimation and analysis of small area income distributions and poverty rates in the city of Athens, Greece. *Computers, Environment and Urban Systems*, 63, 15-25.

Petrakos G. and Mardakis P. (1999) 'Recent changes in the Greek system of urban centers' In D. Economou and G. Petrakos (eds) *The development of Greek cities: Interdisciplinary approaches to urban analysis and policy*. Volos: University of Thessaly (pp. 45-64).

Petrov, L.O., Lavalle, C. & Kasanko, M. (2009). Urban land use scenarios for a tourist region in Europe: Applying the MOLAND model to Algarve, Portugal. *Landscape and Urban Planning*, 92(1), 10-23.

Poelmans, L. & Van Rompaey, A. (2010). Complexity and performance of urban expansion models. *Computer Environment and Urban Systems*, 34, 17–27.

Robinson D.T., Murray-Rust D., Rieser V., Milicic V. & Rounsevell M. (2012) Modelling the impacts of land system dynamics on human well-being: Using an agent-based approach to cope with data limitations in Koper, Slovenia. *Computers, Environment and Urban Systems*, 36, 164–176.

Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment* 121, 93–107.

Rounsevell, M. & Metzger, M. (2010). Developing qualitative scenarios and storylines. Wiley Interdisciplinary Reviews: *Climate Change* 1(4), 606-619.

Sante, I., Garcia, A. M., Miranda, D. & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96(2), 108–122.

Schrojenstein Lantman, J., Verburg, P.H. Bregt, A. & Geertman. S. (2011). Core principles and concepts in land-use modelling: A literature review. Chapter 3 in Land-Use Modelling in Planning Practice, ed. E. Koomen and J. Borsboom-van Beurden. New York: Springer.

Serraos, K. Greve T. Asprogerakas, E. Balampanidis D. & Chani. A. (2016). Athens, a capital in crisis. Tracing the spatial impacts. In Knieling J. & Othengrafen, F. (Eds.) *Cities in Crisis, Socio-spatial impacts of the economic crisis in Southern European cities*. Oxon, Routledge (pp. 116-138).

Soares-Filho, B., Pennachin, C. L. & Cerqueria, G. (2002). DINAMICA- a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154(3), 217–235.

Van Delden, H., Van Vliet, J., Rutledge, D.T. & Kirkby, M.J. (2011). Comparison of scale and scaling issues in integrated land-use models for policy support. *Agriculture, Ecosystems and Environment*, 142, 18– 28.

Veldkamp A. & Fresco L.O. (1996) CLUE: a conceptual model to study the Conversion of Land Use and its Effects. *Ecological Modelling*, 85, 253–270.

Veldkamp, A., Verburg, P.H., Kok, K., de Koning, G.H.J., Priess, J. & Bergsma, A.R. (2001). The need for scale sensitive approaches in spatially explicit land use change modeling. *Environmental Modeling and Assessment*, 6, 111–121.

Verburg P.H., Soepboer W., Limpiada R., Espaldon M.O., Sharifa M. & Veldkamp A. (2002) Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. *Environmental Management*, 30, 391–405.

Verburg, P.H., Schot, P.P., Dijst, M.J. & Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. *GeoJournal,* 61, 309–324.

Verburg, P.H., Schulp, C.J.E., Witte, N. & Veldkamp, A. (2006). Downscaling of land use change scenarios to assess the dynamics of European landscapes. *Agriculture, Ecosystems and Environment*, 116, 39–56.

Verburg, P.H., Dearing, J.A., Dyke, J.G., van der Leeuw, S., Seitzinger, S., Steffen, W. & Syvitski, J. (2016). Methods and approaches to modelling the Anthropocene. *Global Environmental Change*, 39, 328–340.

Vitousek, P.M., Mooney, H.A., Lubchenco, J. & Melillo, J.M. (1997) Human domination of Earth's ecosystems. *Science,* 277, 494–499.

White, R. & Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24(5), 383–400.

Xu, C., Liu, M., Zhang, C., An, S., Yu, W., & Chen, J. (2007). The spatiotemporal dynamics of rapid urban growth in the Nanjing metropolitan region of China. *Landscape Ecology*, 22(6), 925–937.

Yang, Q., Li, X. & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, 34, 592–602.

Zagaria, C., Schulp, C.J.E., Kizos, T., Gounaridis, D. & Verburg, P.H. (2017). Cultural landscapes and behavioral transformations: An agent-based model for the simulation and discussion of alternative landscape futures in East Lesvos, Greece. *Land Use Policy*, 65, 26–44.

List of Figures

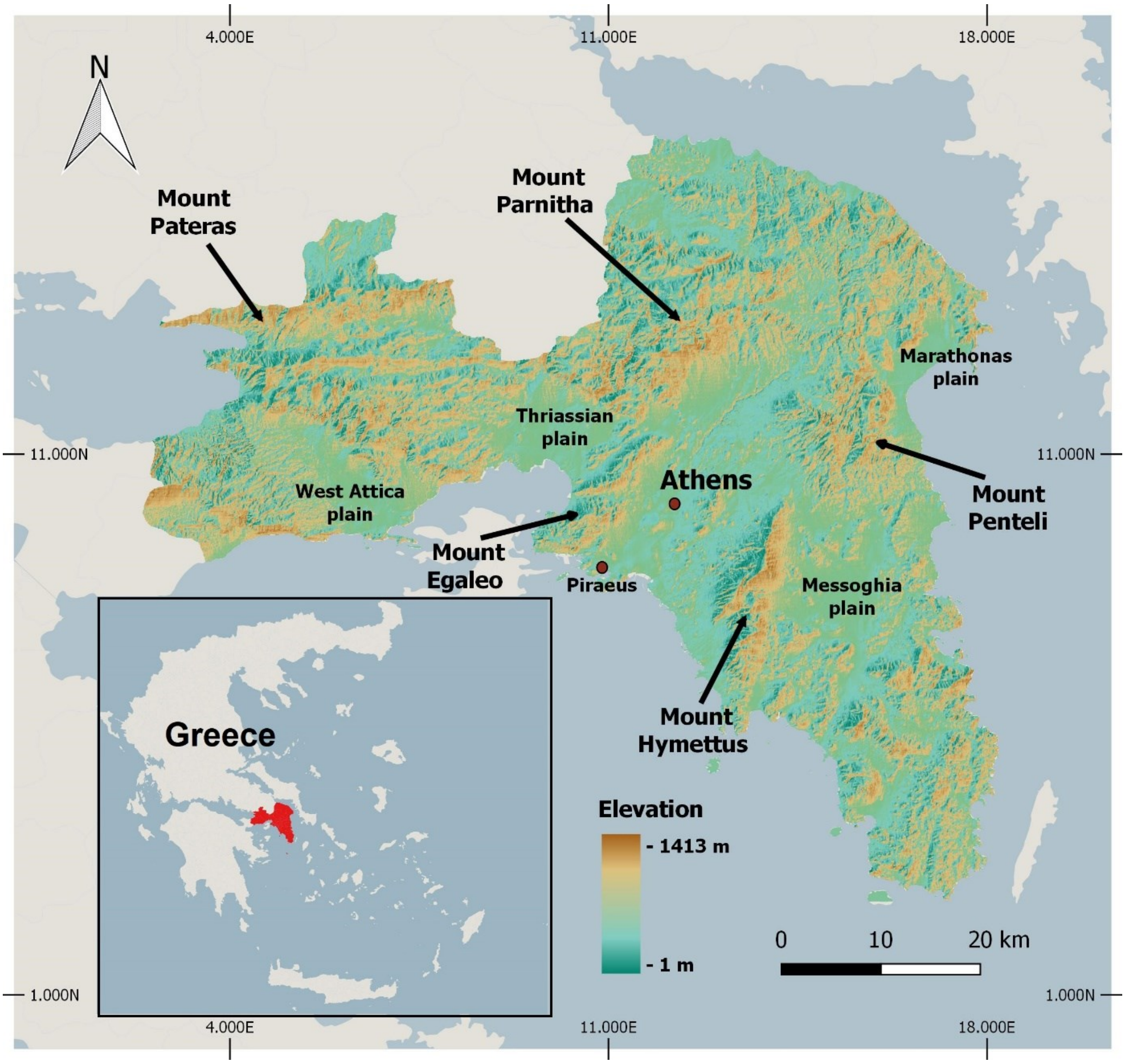
Figure 1. Location and topography of Attica region
Figure 2. Flowchart of the presented methodology
Figure 3. Summary of statistics based on the classified maps and relative percentages of artificial areas. Three different trends reflect three different levels of development during the last 25 years
Figure 4. The simulated map of 2016 versus the observed of 2016.
Figure 5. Multi-resolution evaluation of model fitting using the fuzzy similarity index.
Figure 6. Result of cross classification between the simulated vs the observed map of 2016
Figure 7. Mean Decrease Gini (IncNodePurity) as assigned by the RF regression algorithm
Figure 8. Mean Decrease Accuracy (%IncMSE) as assigned by the RF regression algorithm
Figure 9. Concorance correlation coeficient between transition probabilities for continuous urban fabric category. The values derived from 1000 random samples, dispersed across the study area
Figure 10. LULC spatial configuration simulated for 2040, under the Medium development scenario
Figure 11. LULC spatial configuration simulated for 2040, under the High development scenario
Figure 12. LULC spatial configuration simulated for 2040, under the Low development scenario
Figure 13. Rates of LULC simulated for 2040 in a 5-year step, under the three different development level scenarios

Tables

Table 1. List of predictors used in the transition potential modelling process.

Table 2. Transition probabilities of the eight LULC categories, allocated per scenario. The numbers indicate transition rates per year in hectares.

Table 3. Error matrix - Simulated map (2016) versus the observed (2016). U.A: User's Accuracy; P.A: Producer's Accuracy; O.A: Overall Accuracy. 1: Continuous urban fabric. 2: Discontinuous dense urban fabric. 3: Discontinuous medium density urban fabric. 4: Discontinuous low density urban fabric. 5: Industrial, commercial, and transport units. 6: Arable land and permanent crops. 7: Forests, scrubs, and other natural areas. 8: Other (open spaces, bare land, mines, inland water).



4.000E
11.000E
18.000E
N
Mount Pateras
Mount Parnitha
Marathonas plain
Thriassian plain
11.000N
Athens
Mount Penteli
West Attica plain
Mount Egaleo
Piraeus
Messoghia plain
Greece
Mount Hymettus
Elevation
- 1413 m
- 1 m
0
10
20 km
1.000N

Figure 1

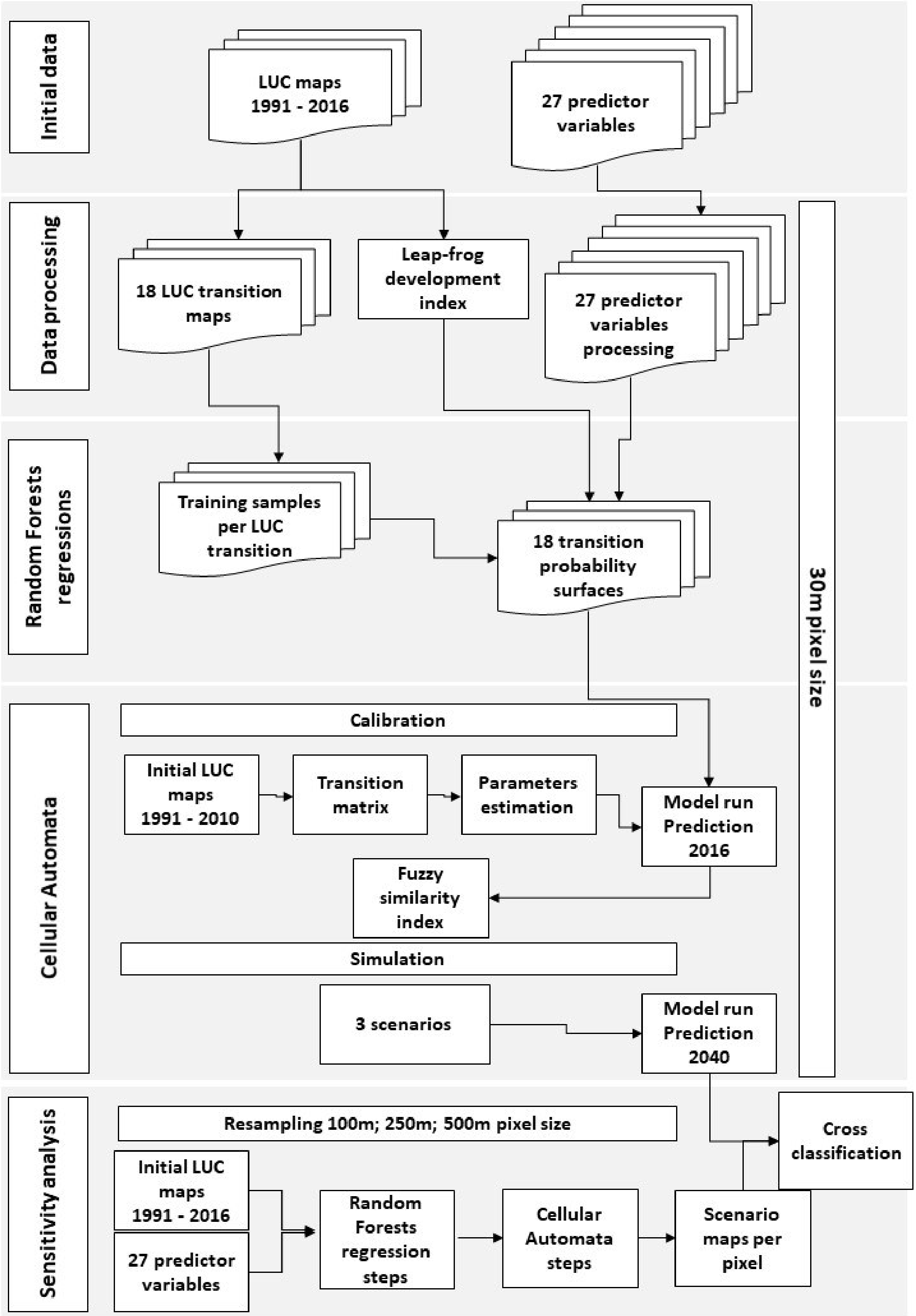

Initial data
LUC maps
1991 - 2016
27 predictor variables
Data processing
18 LUC transition maps
Leap-frog development index
27 predictor variables processing
Random Forests regressions
Training samples per LUC transition
18 transition probability surfaces
30m pixel size
Cellular Automata
Calibration
Initial LUC maps 1991 - 2010
Transition matrix
Parameters estimation
Model run Prediction 2016
Fuzzy similarity index
Simulation
3 scenarios
Model run Prediction 2040
Sensitivity analysis
Resampling 100m; 250m; 500m pixel size
Cross classification
Initial LUC maps 1991 - 2016
27 predictor variables
Random Forests regression steps
Cellular Automata steps
Scenario maps per pixel

Figure 2

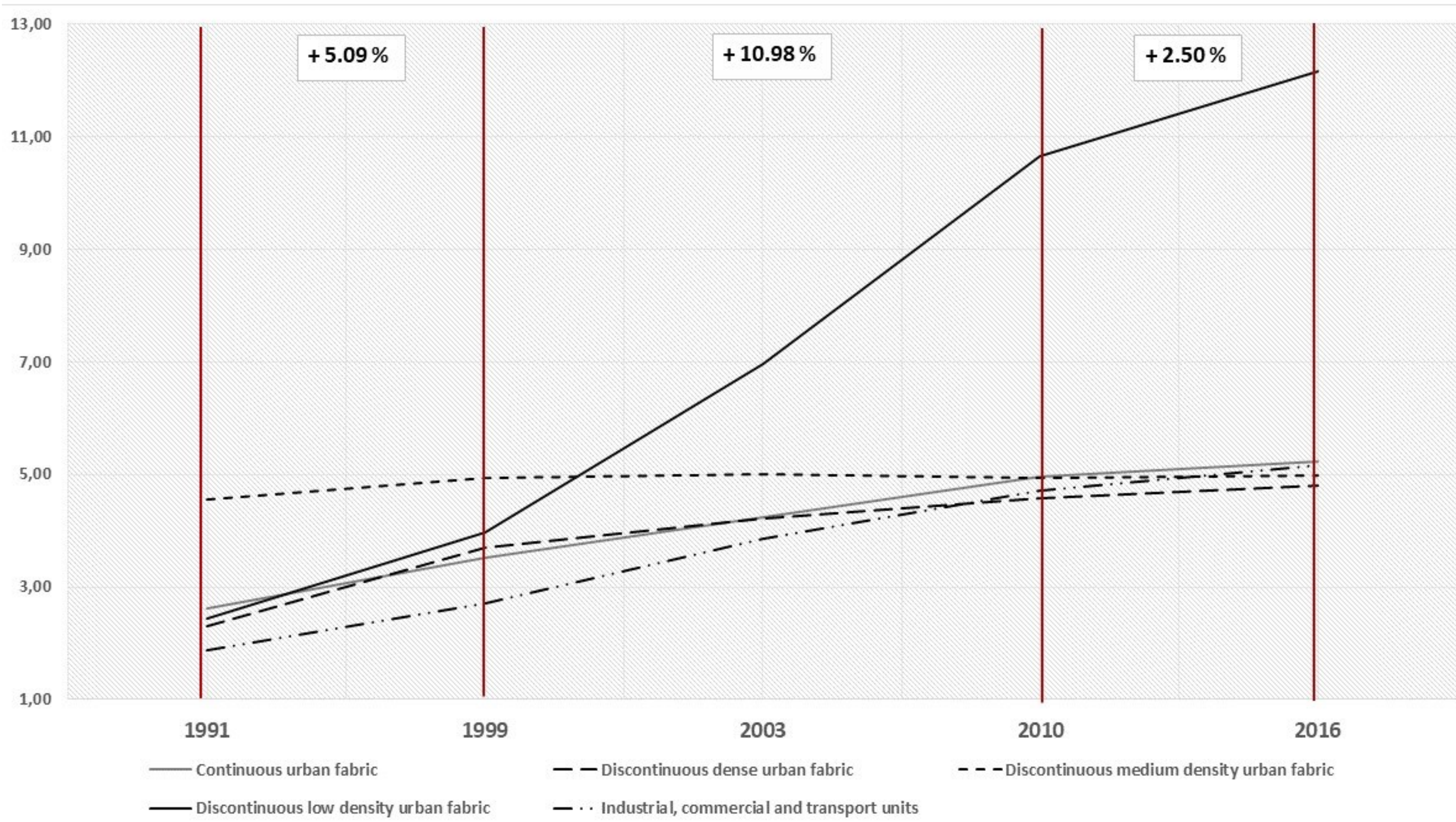

13,00
11,00
9,00
7,00
5,00
3,00
1,00
+ 5.09 %
+ 10.98 %
+ 2.50 %
1991
1999
2003
2010
2016
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units

Figure 3

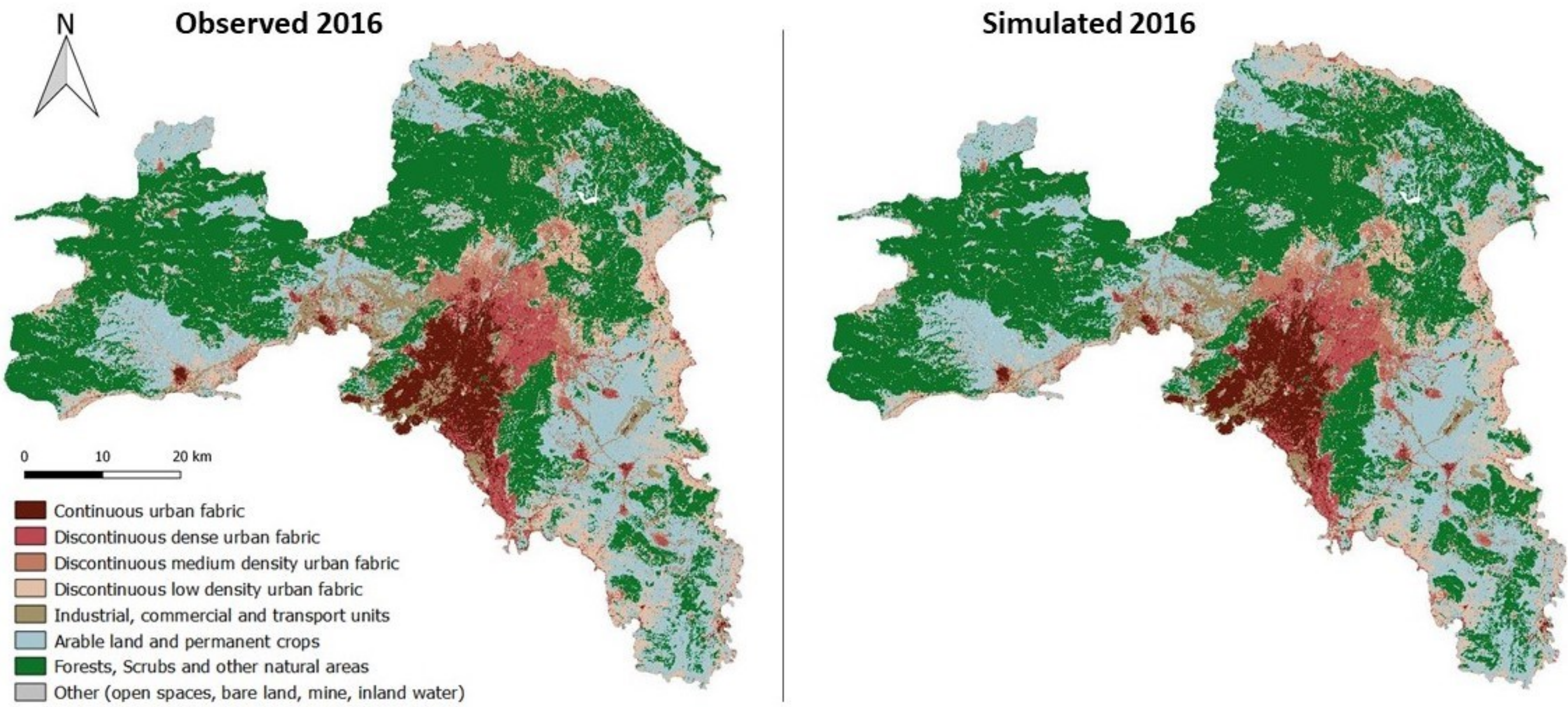

N
Observed 2016
Simulated 2016
0
10
20 km
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas
Other (open spaces, bare land, mine, inland water)

Figure 4

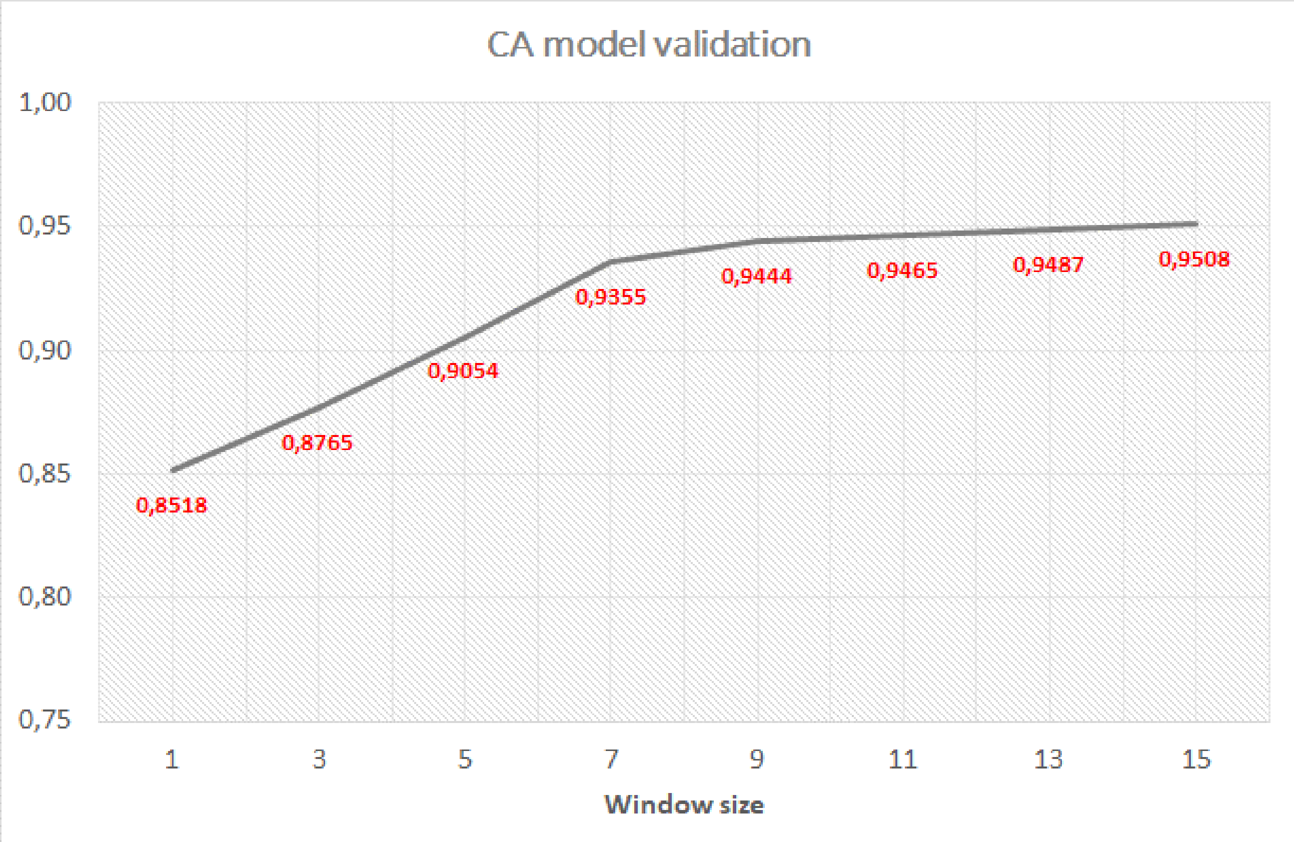

CA model validation
1,00
0,95
0,90
0,85
0,80
0,75
0,8518
0,8765
0,9054
0,9355
0,9444
0,9465
0,9487
0,9508
1
3
5
7
9
11
13
15
Window size

Figure 5

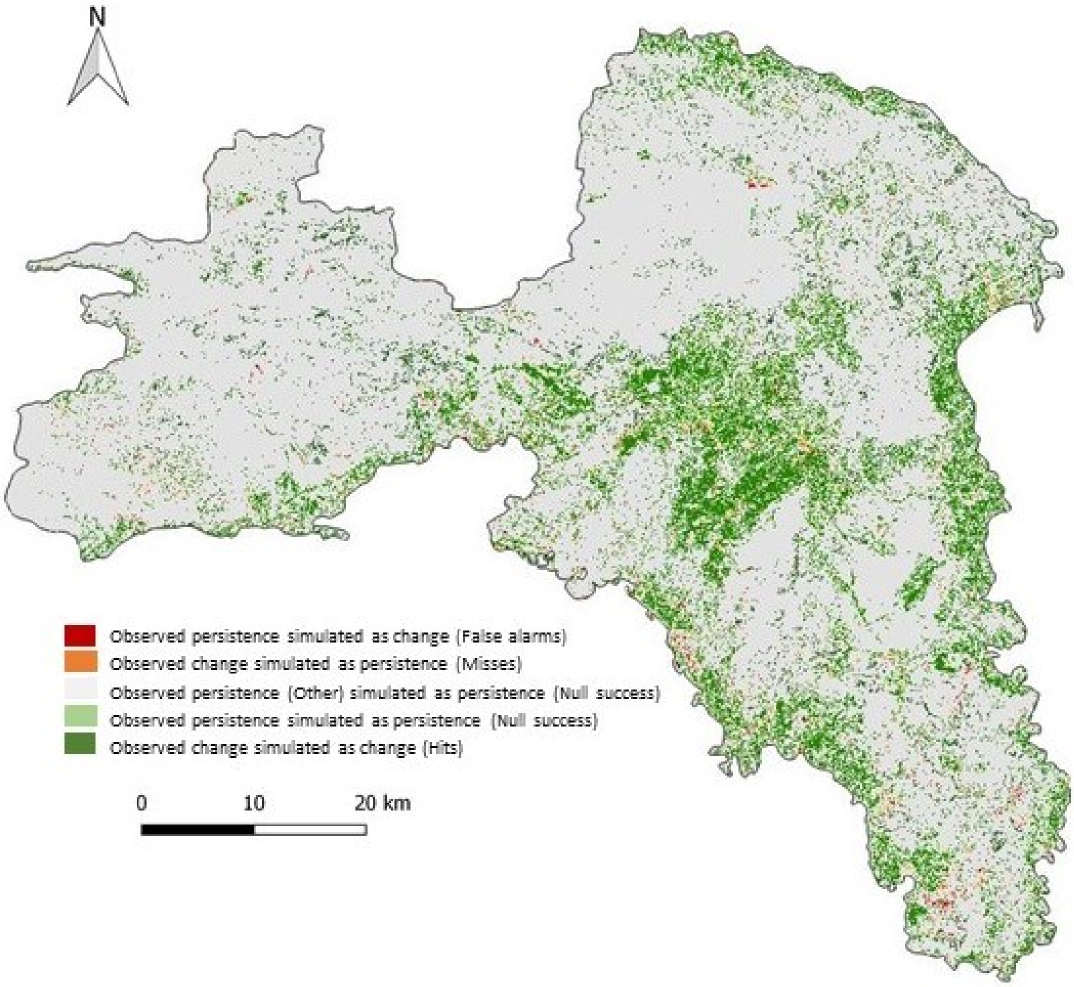

N
Observed persistence simulated as change (False alarms)
Observed change simulated as persistence (Misses)
Observed persistence (Other) simulated as persistence (Null success)
Observed persistence simulated as persistence (Null success)
Observed change simulated as change (Hits)
0
10
20 km

Figure 6

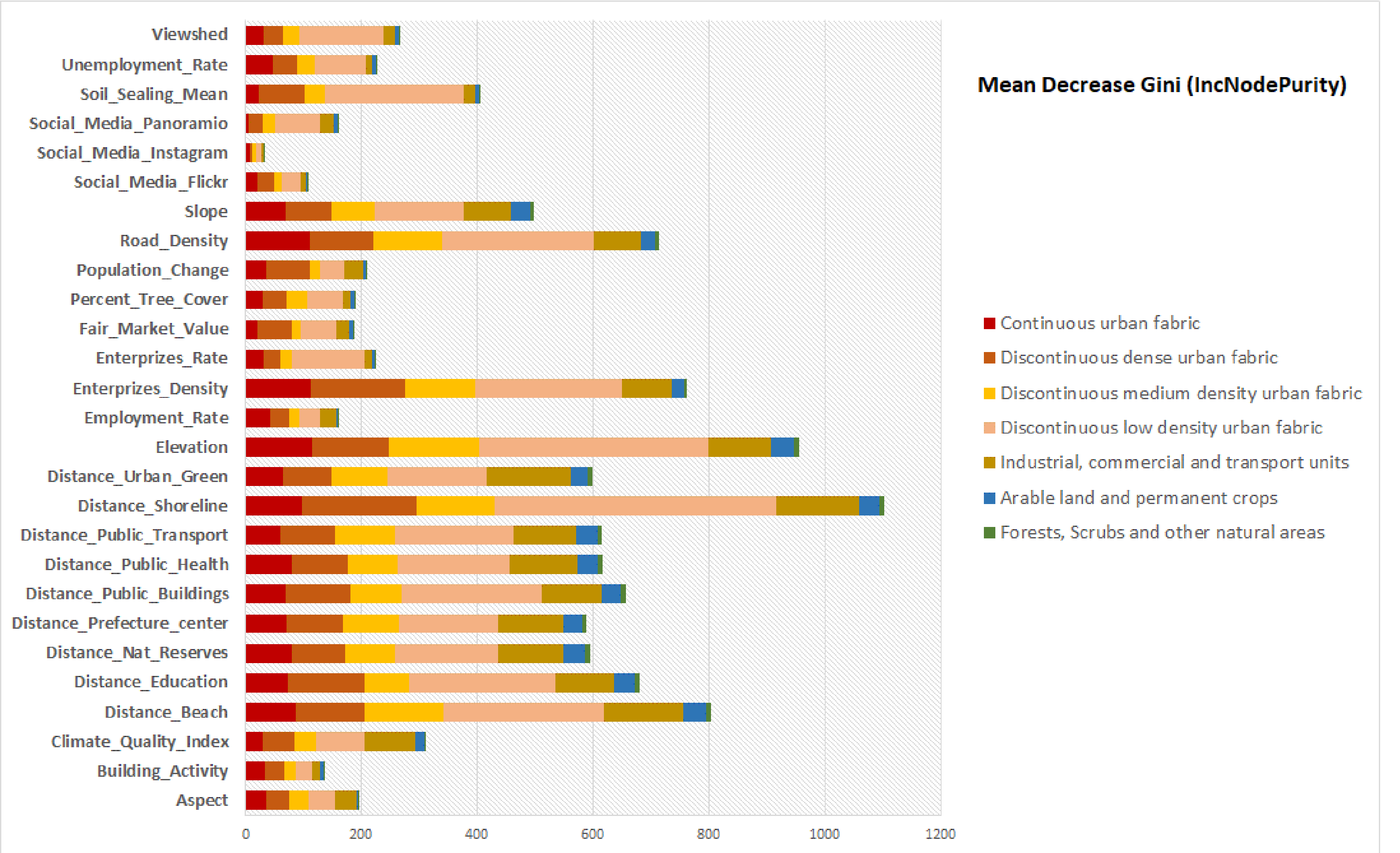

Mean Decrease Gini (IncNodePurity)
Viewshed
Unemployment_Rate
Soil_Sealing_Mean
Social_Media_Panoramio
Social_Media_Instagram
Social_Media_Flickr
Slope
Road_Density
Population_Change
Percent_Tree_Cover
Fair_Market_Value
Enterprizes_Rate
Enterprizes_Density
Employment_Rate
Elevation
Distance_Urban_Green
Distance_Shoreline
Distance_Public_Transport
Distance_Public_Health
Distance_Public_Buildings
Distance_Prefecture_center
Distance_Nat_Reserves
Distance_Education
Distance_Beach
Climate_Quality_Index
Building_Activity
Aspect
0
200
400
600
800
1000
1200
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas

Figure 7

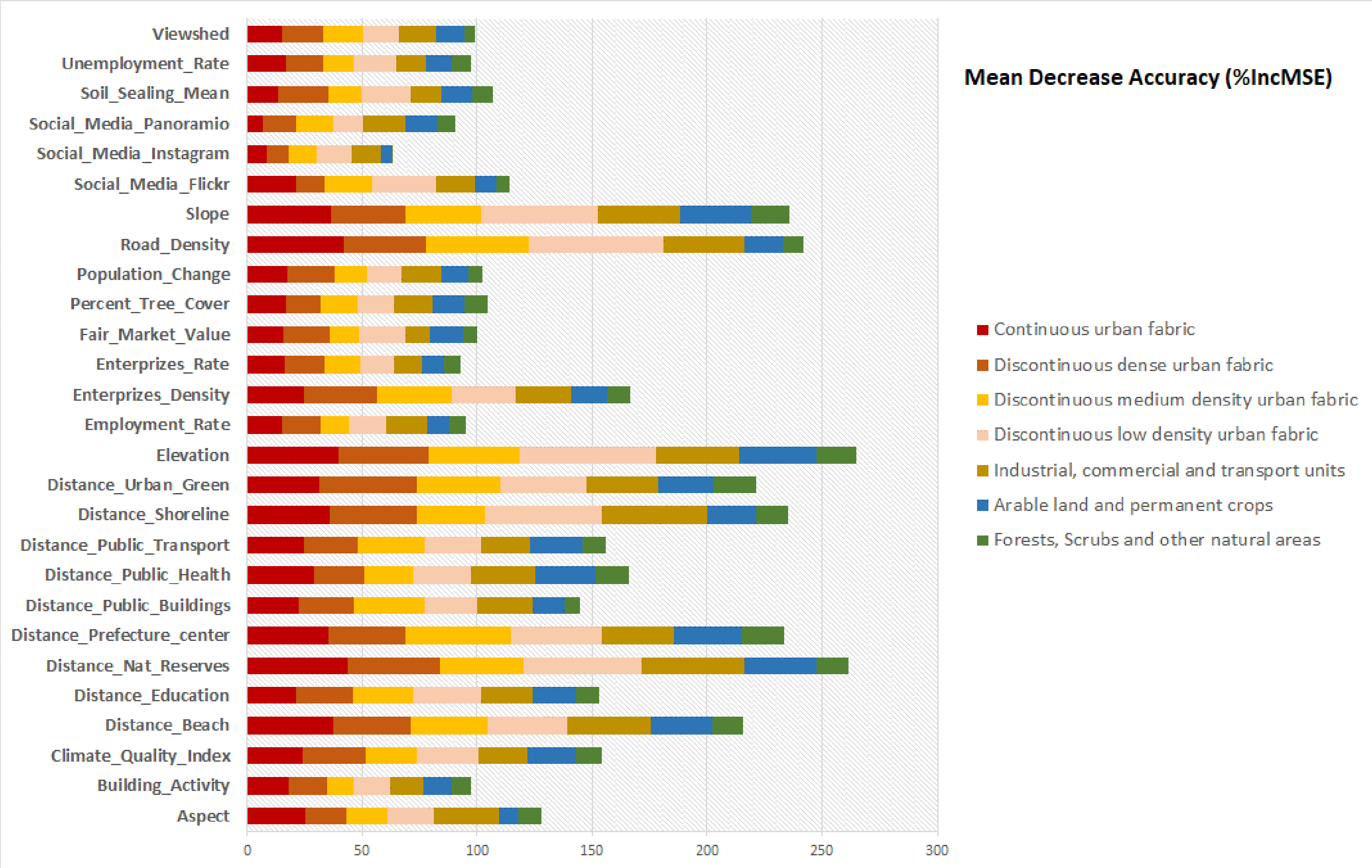

Mean Decrease Accuracy (%IncMSE)
Viewshed
Unemployment_Rate
Soil_Sealing_Mean
Social_Media_Panoramio
Social_Media_Instagram
Social_Media_Flickr
Slope
Road_Density
Population_Change
Percent_Tree_Cover
Fair_Market_Value
Enterprizes_Rate
Enterprizes_Density
Employment_Rate
Elevation
Distance_Urban_Green
Distance_Shoreline
Distance_Public_Transport
Distance_Public_Health
Distance_Public_Buildings
Distance_Prefecture_center
Distance_Nat_Reserves
Distance_Education
Distance_Beach
Climate_Quality_Index
Building_Activity
Aspect
0
50
100
150
200
250
300
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas

Figure 8

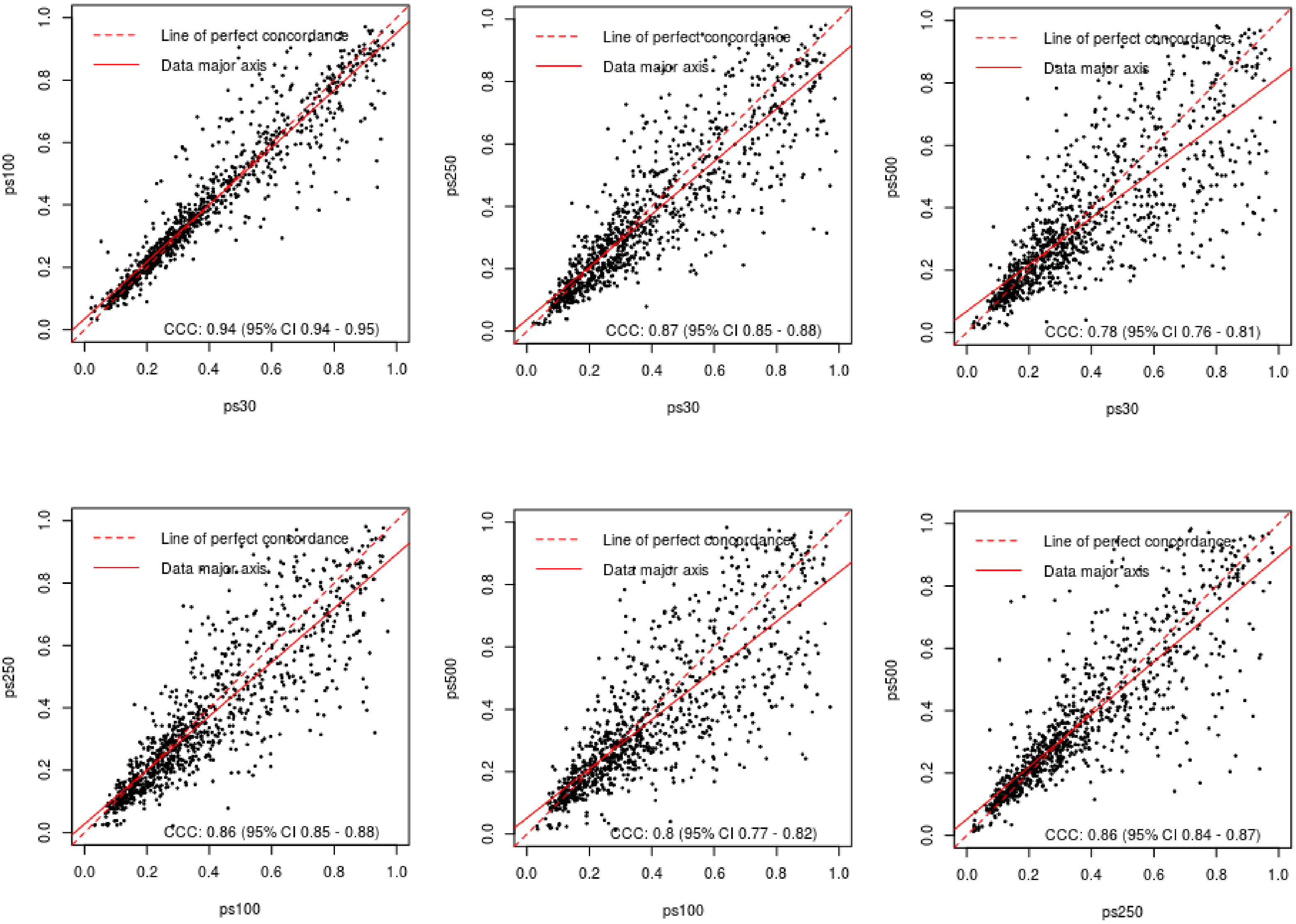

Line of perfect concordance
Data major axis
CCC: 0.94 (95% CI 0.94 - 0.95)
ps30
ps100
CCC: 0.87 (95% CI 0.85 - 0.88)
ps30
ps250
CCC: 0.78 (95% CI 0.76 - 0.81)
ps30
ps500
CCC: 0.86 (95% CI 0.85 - 0.88)
ps100
ps250
CCC: 0.8 (95% CI 0.77 - 0.82)
ps100
ps500
CCC: 0.86 (95% CI 0.84 - 0.87)
ps250
ps500

Figure 9

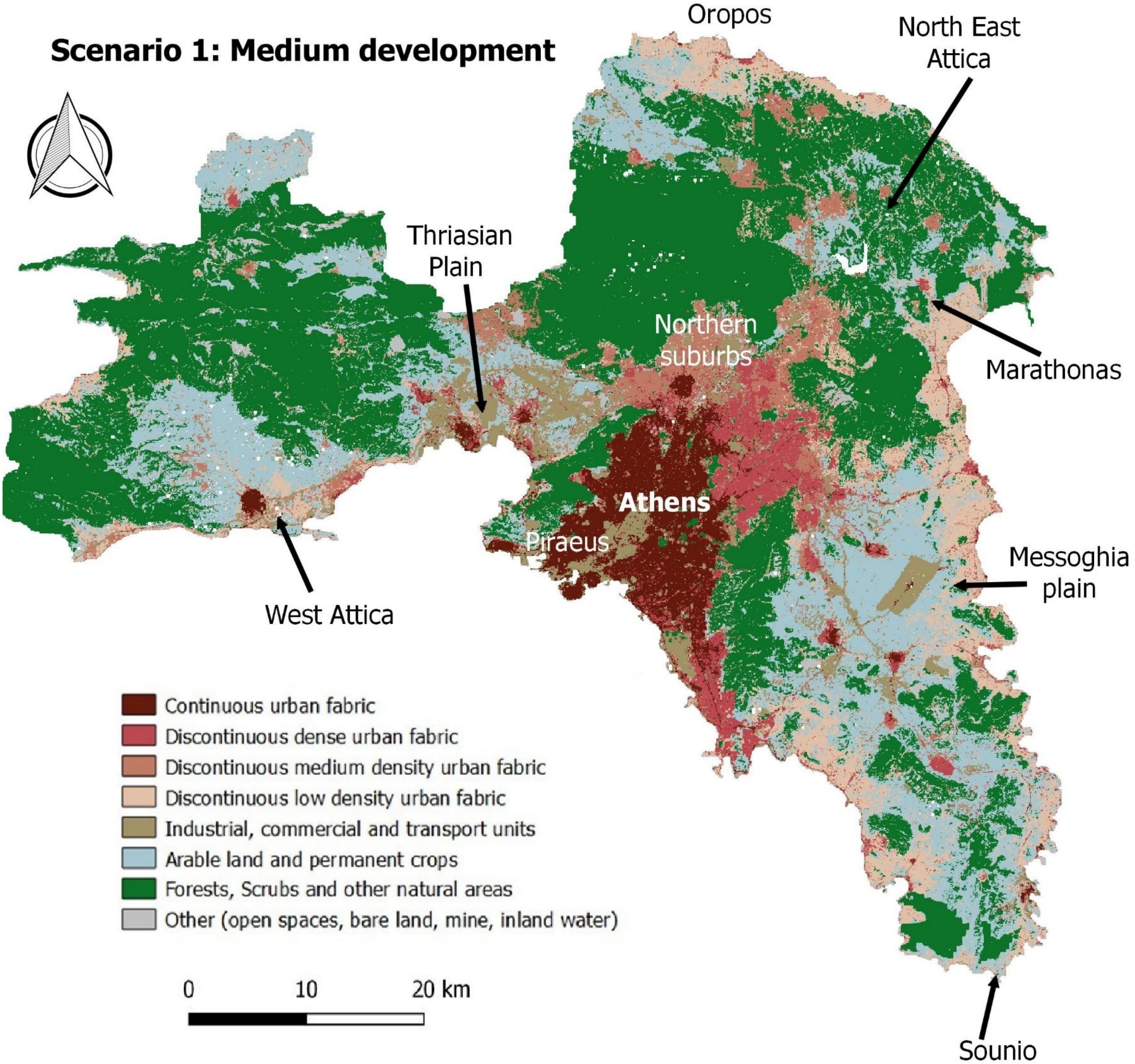

Scenario 1: Medium development
Oropos
North East Attica
Thriasian Plain
Northern suburbs
Marathonas
Athens
Piraeus
Messoghia plain
West Attica
Sounio
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas
Other (open spaces, bare land, mine, inland water)
0
10
20 km

Figure 10

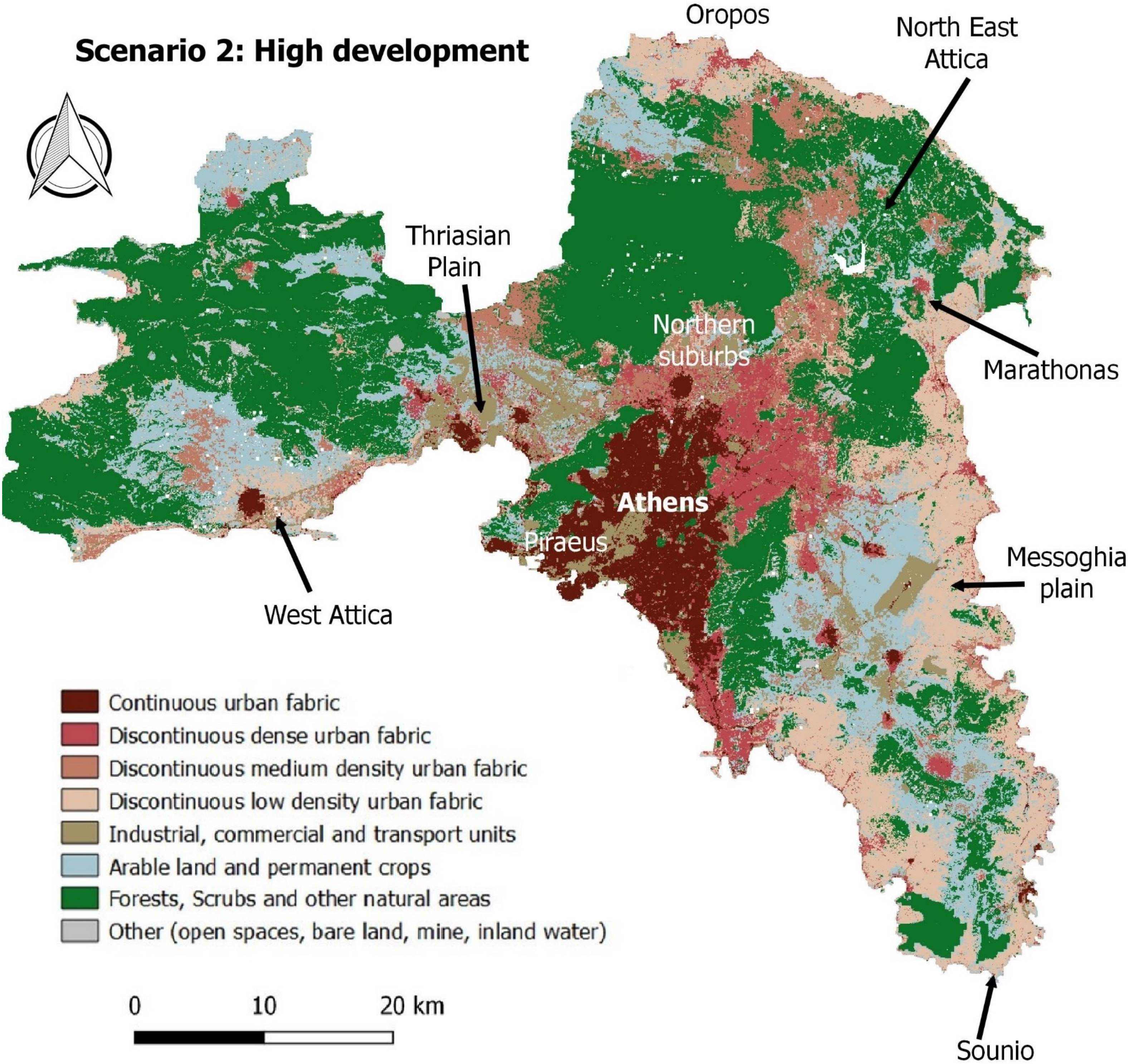

Scenario 2: High development
Oropos
North East
Attica
Thriasian
Plain
Northern
suburbs
Marathonas
Athens
Piraeus
Messoghia
plain
West Attica
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas
Other (open spaces, bare land, mine, inland water)
0
10
20 km
Sounio

Figure 11

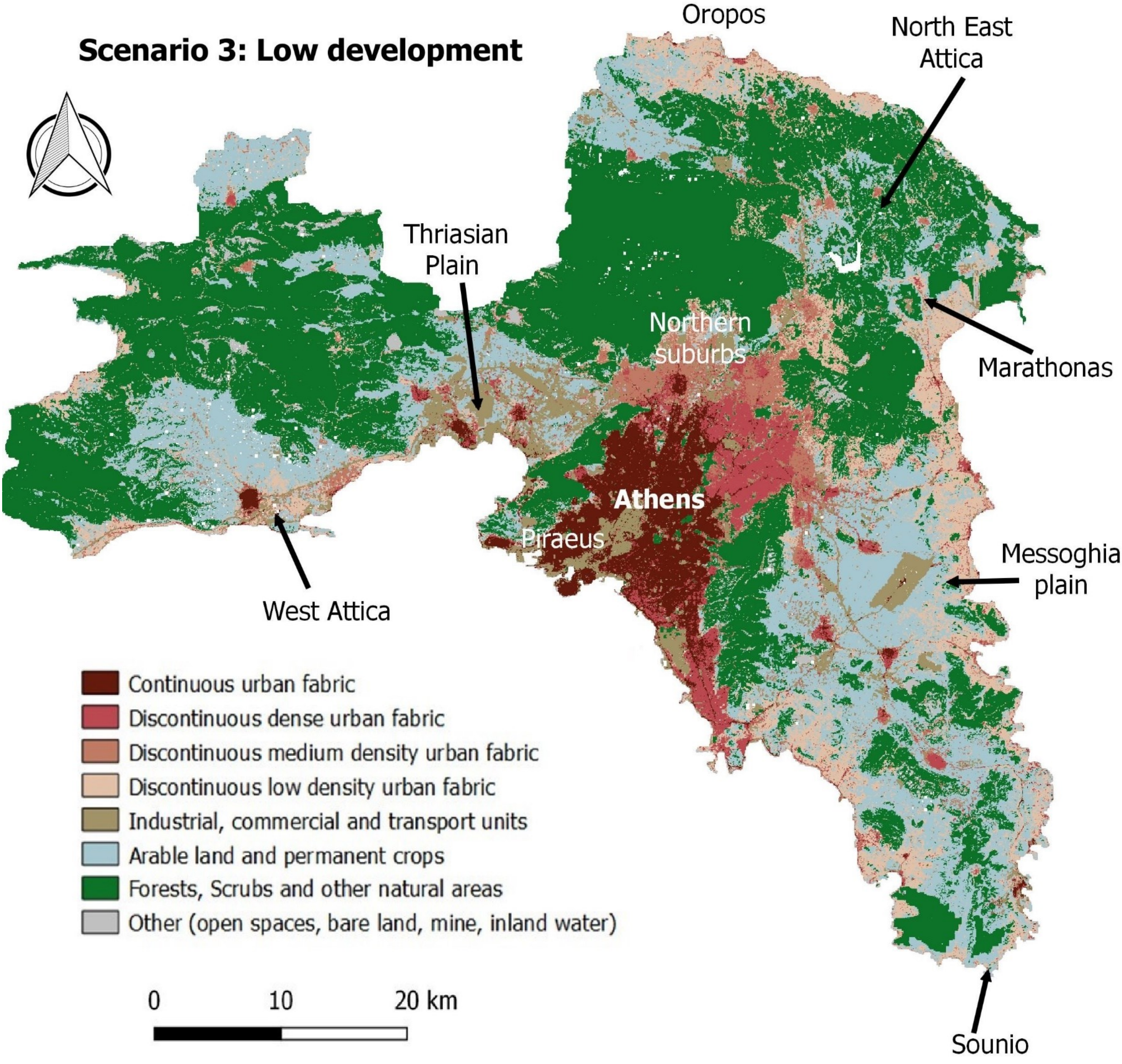
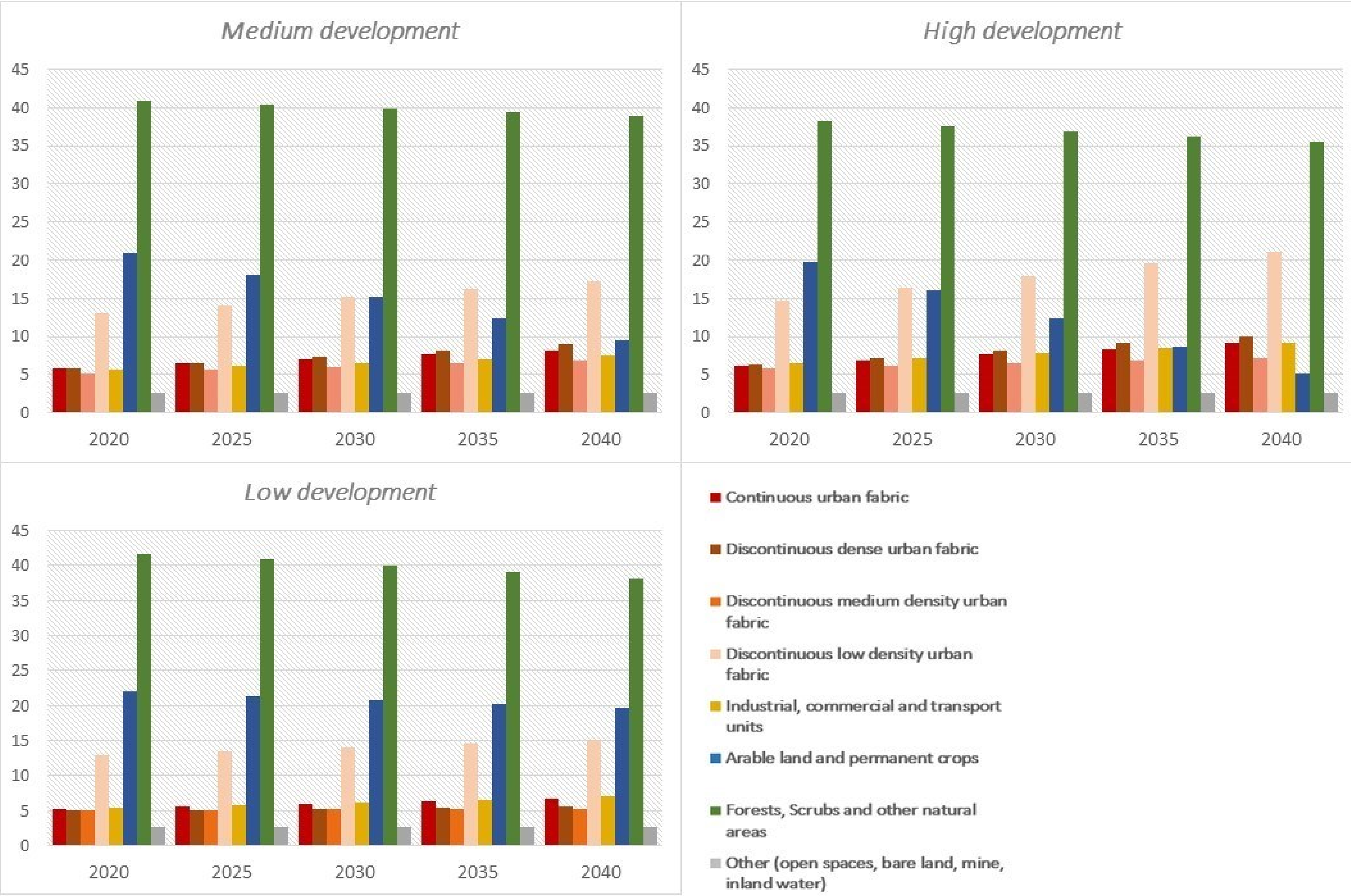

Scenario 3: Low development
Oropos
North East Attica
Thriasian Plain
Northern suburbs
Marathonas
Athens
Piraeus
Messoghia plain
West Attica
Sounio
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas
Other (open spaces, bare land, mine, inland water)
0
10
20 km

Figure 12



Medium development
High development
Low development
45
40
35
30
25
20
15
10
5
0
2020
2025
2030
2035
2040
Continuous urban fabric
Discontinuous dense urban fabric
Discontinuous medium density urban fabric
Discontinuous low density urban fabric
Industrial, commercial and transport units
Arable land and permanent crops
Forests, Scrubs and other natural areas
Other (open spaces, bare land, mine, inland water)

Figure 13

Table 1

Variable	Discription	Source	Time interval	Spatial resolution
Territorial variables				
Elevation	Elevation in m	GLSDEM*	(-)	30m
Slope	Slope in degrees	GLSDEM	(-)	30m
Aspect	Aspect in degrees	GLSDEM	(-)	30m
Climate Quality	Climate quality index	EEA*	1961-1990	1km
Viewshed	Visibility from residential areas at the parcel level (centroids from UA)	GLSDEM and Urban Atlas*	(-)	30m
Distance from beaches	Euclidean distance from beaches signed with a blue flag in m	Ministry of Environment & Energy*	2010	30m
Distance from the sea	Euclidean distance from the shoreline in m		(-)	30m
Socio-economic variables				
Distance from Education centers	Euclidean distance from public education centers (all levels)	Ministry of Education & OSM*	2010	30m
Distance from public health centers	Euclidean distance from public health centers	Society of Information* & OSM	(-)	30m
Distance from nearest town	Euclidean distance from the center of the nearest town (Markopoulo, Paiania, Koropi, Keratea, Artemida) in m	OSM	(-)	30m
Distance from public buildings	Euclidean distance from public buildings	Society of Information & OSM		30m
Distance from public health	Euclidean distance from public hospitals and other public health care units in m	OSM	(-)	30m
Distance from public transport	Euclidean distance from public transport stops (bus, metro, tram, suburban train) in m	OSM & opendata	(-)	30m
Distance from road network	Euclidean distance from road network in m	OSM	(-)	30m
Demographics	Changes in population density at the municipality level	ELSTAT*	1991-2011	30m
Employment rate	Total number of employed persons per total population at the municipality level	ELSTAT	1991-2011	30m
Unemployment rate	Total number of unemployed persons per total population at the municipality level	ELSTAT	1991-2011	30m
Landscape values Instagram	Landscape values quantifyed using Instagram data	van Zanten et al. (2016)	2004-2015	1km
Landscape values Flickr	Landscape values quantifyed using Flickr data	van Zanten et al. (2016)	2004-2015	1km
Landscape values Panoramio	Landscape values quantifyed using Panoramio data	van Zanten et al. (2016)	2004-2015	1km
Land use				
Distance from green urban areas	Euclidean distance from green urban patches in m	Urban Atlas	2006	30m
Soil Sealing rate	Average soil sealing per municipality	EEA	2006-2012	30m
Tree cover	Average tree cover canopy percentage per municipality	USGS*	2010	30m
Built-up rate	Cumulative total number of new houses built per municipality	ELSTAT	1997-2016	30m
HeatMap of Enterprizes	HeatMap of new enterprises registered to ACCI	ACCI*	1991-2016	30m
Enterprises count	Cumulative total number of new enterprises registered to ACCI per municipality	ACCI	1991-2016	30m
Distance from natural reserves	Euclidean distance from forested patches, areas of high nature value and protected areas in m	Ministry of Environment & Energy & OSM & Natura 2000	(-)	30m

[a] Global Land Survey Digital Elevation Model (GLSDEM) http://glcf.umd.edu/data/glsdem/

[b] European Environmental Agency. https://www.eea.europa.eu/data-and-maps/data/indices-of-climate-soil-and-vegetation-quality-1#tab-metadata

[c] European Environmental Agency. Urban Atlas. GMES/Copernicus land monitoring services. https://www.eea.europa.eu/data-and-maps/data/urban-atlas

[d] Ministry of Environment & Energy. http://geodata.gov.gr/dataset/poioteta-udaton-akton-kolumbeses-2013

[e] Open Street Map. https://www.openstreetmap.org

[f] Society of Information. http://geodata.gov.gr/dataset/demosia-kteria

[g] Hellenic statistical authority. http://www.statistics.gr/

[h] van Zanten et al. (2016). PNAS. http://geoplaza.vu.nl/data/dataset/continental-scale-quantification-of-landscape-values-using-social-media-data

[i] USGS. Global Tree Canopy Cover. https://landcover.usgs.gov/glc/TreeCoverDescriptionAndDownloads.php

[j] Athens chamber of commerce and industry http://www.acci.gr/acci/catalogue/search.jsp?context=201

Table 2

From	*To*	*Medium development*	*High development*	*Low development*
Discontinuous dense urban fabric	Continuous urban fabric	0,319	0,392	0,051
Discontinuous medium density urban fabric	Continuous urban fabric	0,029	0,040	0,005
Discontinuous medium density urban fabric	Discontinuous dense urban fabric	0,356	0,384	0,070
Discontinuous low density urban fabric	Continuous urban fabric	0,001	0,004	0,001
Discontinuous low density urban fabric	Discontinuous dense urban fabric	0,044	0,049	0,008
Discontinuous low density urban fabric	Discontinuous medium density urban fabric	0,383	0,436	0,022
Arable land and permanent crops	Continuous urban fabric	0,001	0,002	0,000
Arable land and permanent crops	Discontinuous dense urban fabric	0,010	0,019	0,001
Arable land and permanent crops	Discontinuous medium density urban fabric	0,026	0,043	0,005
Arable land and permanent crops	Discontinuous low density urban fabric	0,049	0,174	0,055
Arable land and permanent crops	Industrial commercial and transport units	0,018	0,045	0,014
Arable land and permanent crops	Forests Scrubs and other natural areas	0,090	0,099	0,083
Forests Scrubs and other natural areas	Continuous urban fabric	0,000	0,000	0,000
Forests Scrubs and other natural areas	Discontinuous dense urban fabric	0,001	0,002	0,000
Forests Scrubs and other natural areas	Discontinuous medium density urban fabric	0,002	0,004	0,001
Forests Scrubs and other natural areas	Discontinuous low density urban fabric	0,007	0,029	0,002
Forests Scrubs and other natural areas	Industrial commercial and transport units	0,001	0,002	0,001
Forests Scrubs and other natural areas	Arable land and permanent crops	0,060	0,064	0,056

Table 3

Observed 2016	Simulated 2016								
	1	2	3	4	5	6	7	Totals	P.A
1	1731	148	18	25	11			1933	89,55
2	112	1371	78	33	10			1604	85,47
3	36	77	1293	61	19		1	1487	86,95
4	3	23	99	1420	7	29	3	1584	89,65
5	17	21	11	7	529	12		597	88,61
6		3	9	121	14	957	32	1136	84,24
7		1	2	14	1	36	1004	1058	94,90
Totals	1899	1644	1510	1681	591	1034	1040	9399	
U.A	91,2	83,4	85,6	84,5	89,5	92,6	96,5		
O.A	88,36								