University of the Aegean – Department of Geography

Multi-scale modelling of land cover/land use (LCLU) change with Geoinformatics and scenario-based simulations

by

Dimitrios Gounaridis

A dissertation submitted to the
Department of Geography, University of the Aegean
in fulfillment of the requirements for the degree of
Doctor of Geography

Mytilene, 2018

Committee

- Sotirios Koukoulas, Professor (Associate), Department of Geography, University of the Aegean (Supervisor)
- Helen Briassoulis, Professor (Full), Department of Geography, University of the Aegean (member of advisory committee)
- Elias Symeonakis, Senior Lecturer, School of Science and the Environment, Manchester Metropolitan University (member of advisory committee)
- Nikolaos Soulakellis, Professor (Full), Department of Geography, University of the Aegean (examiner)
- Athanasios Kizos, Professor (Associate), Department of Geography, University of the Aegean (examiner)
- Dimitrios Kavroudakis, Professor (Assistant), Department of Geography, University of the Aegean (examiner)
- Ioannis Gitas, Professor (Full), Department of Forestry and Natural Environment, Aristotle University of Thessaloniki (examiner)

Ευχαριστίες

Αρχικά, θέλω να ευχαριστήσω τον επιβλέποντα μου, Αναπληρωτή Καθηγητή κ. Σωτήρη Κουκούλα για την αρχική εμπιστοσύνη, την απεριόριστη ελευθερία να προσαρμόσω το θέμα στα ενδιαφέροντα μου, την υλική υποστήριξη που μου παρείχε, τη σημαντική καθοδήγησή και τις εκατοντάδες ώρες συζητήσεων όλα αυτά τα χρόνια. Όλα αυτά συνέβαλαν σε μεγάλο βαθμό στην ολοκλήρωση της παρούσας διατριβής. Ευχαριστώ θερμά τον Αναπληρωτή Καθηγητή κ. Θανάση Κίζο για την πολύτιμη εμπειρία που αποκόμισα δουλεύοντας σε ευρωπαϊκά προγράμματα, για την άριστη συνεργασία και για τη βοήθεια του όποτε τη χρειάστηκα.

Ευχαριστώ επίσης τα δυο μέλη της τριμελούς συμβουλευτικής επιτροπής για την τιμή που μου έκαναν να συμμετάσχουν, την Καθηγήτρια κα. Ελένη Μπριασούλη για τα εύστοχα σχόλια και τις συμβουλές της καθώς και τον Επίκουρο Καθηγητή κ. Ηλία Συμεωνάκη που ήταν πάντα άμεσα προσβάσιμος και θετικός για οποιοδήποτε θέμα. Θα ήθελα να ευχαριστήσω επίσης τους Αναπληρωτή Καθηγητή κ. Γιάννη Χωριανόπουλο και Επίκουρο Καθηγητή Γιώργο Ζαίμη για την άριστη συνεργασία σε διάφορα στάδια του διδακτορικού.

Ευχαριστώ επίσης τους Καθηγητή κ. Γήτα, Επίκουρο Καθηγητή κ. Καβρουδάκη και Καθηγητή κ. Σουλακέλλη για την άμεση ανταπόκρισή τους και την τιμή που μου έκαναν να κρίνουν αυτή τη διατριβή.

Ευχαριστώ το εργαστήριο Χωρικής Ανάλυσης, Συστημάτων Γεωγραφικών Πληροφοριών και Τηλεποσκόπησης για τη φιλοξενία και την άριστη συνεργασία. Θεωρώ υποχρέωσή μου επίσης να ευχαριστήσω όλους του καθηγητές του Προγράμματος Μεταπτυχιακών Σπουδών του Τμήματος Γεωγραφίας για τις γνώσεις που μου μετέδωσαν και την έμμεση συμβολή τους στην εκπόνηση της παρούσας διατριβής.

Τις ευχαριστίες μου επίσης οφείλω προς το Ίδρυμα Κρατικών Υποτροφιών, του οποίου υπήρξα υπότροφος, για την οικονομική στήριξη που μου παρείχε.

Τέλος ευχαριστώ τη σύντροφο της ζωής μου Νίκη και την οικογένειά μου για την υπομονή, τη στωικότητα, τη συνεχή ενθάρρυνση και την αδιάλειπτη στήριξη όλων αυτών των ετών.

Content

List of figures

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of tables

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Εκτενής περίληψη

Οι αλλαγές στην κάλυψη/χρήση γης, είναι ένα πολυεπίπεδο φαινόμενο που λαμβάνει χώρα σε διάφορες μορφές. Ο όρος κάλυψη γης σχετίζεται με τον τύπο των χαρακτηριστικών που εμφανίζονται πάνω στην επιφάνεια της γης ενώ ο όρος χρήση γης σχετίζεται με την ανθρώπινη εκμετάλλευση και την οικονομική λειτουργία των διαφόρων τύπων κάλυψης. Το μεγαλύτερο ποσοστό των αλλαγών στην κάλυψη/χρήση γης οφείλεται σε ανθρωπογενή αίτια. Οι επιπτώσεις που συγκεντρώνουν το μεγαλύτερο κομμάτι της έρευνας, είναι εκείνες που αφορούν τις αλληλεπιδράσεις των αλλαγών στην κάλυψη/χρήσεις γης με θεμελιώδη χαρακτηριστικά και διεργασίες της γης όπως το κλίμα, ο βιοφυσικός και υδρολογικός κύκλος, η βιοποικιλότητα και εν τέλει οι υπηρεσίες του οικοσυστήματος. Ο μηχανισμός που υποκινεί αυτές τις αλλαγές είναι πολύπλοκος και πολυπαραγοντικός. Οι παράγοντες που συμβάλλουν στις αλλαγές της κάλυψης/χρήσης γης, είναι πολυδιάστατοι (περιβαλλοντικοί, κοινωνικοί, οικονομικοί, πολιτικοί κ.α.) και αλληλοεπιδρούν με πλήθος άλλων παραγόντων αλλά και μεταξύ τους.

Σκοπός της διατριβής είναι να διερευνήσει να αποτυπώσει και να κατανοήσει το σύστημα αλλαγών κάλυψης/χρήσης γης σε διάφορες κλίμακες αποκωδικοποιώντας τους παράγοντες που επιδρούν σε αυτό και τέλος να χρησιμοποιήσει την πληροφορία που προκύπτει για την προσομοίωση μελλοντικών σεναρίων εξέλιξης του συστήματος. Η βιβλιογραφία υπογραμμίζει ότι πολλές απόπειρες προς αυτή την κατεύθυνση έχουν ορισμένα μεθοδολογικά μειονεκτήματα που υπονομεύουν την εγκυρότητα και τη χρησιμότητα των αποτελεσμάτων. Για παράδειγμα πολλές μελέτες δεν λαμβάνουν υπόψη τις πολλαπλές κλίμακες που εμπλέκονται στις αλλαγές κάλυψης/χρήσης γης, συμπεριλαμβάνοντας στις αναλύσεις τους μόνο παράγοντες που λειτουργούν σε μία μόνο κλίμακα (π.χ. διοικητική ή λήψης αποφάσεων). Επιπλέον, σε πολλές μελέτες, οι παρελθοντικές αλλαγές κάλυψης/χρήσης γης αντιμετωπίζονται συχνά σε χαμηλή ανάλυση (θεματική, χωρική και χρονική), γεγονός το οποίο υπονομεύει τη χρησιμότητα των αποτελεσμάτων. Επιπλέον, πολλές μελέτες επικεντρώνονται σε μια συγκεκριμένη πλευρά του φαινομένου (π.χ. κοινωνικοοικονομική) χωρίς να λαμβάνουν υπόψη άλλους σημαντικούς παράγοντες ενώ η πλειονότητα των μελετών παρέχει αποτελέσματα μοντελοποίησης που δεν υπόκεινται σε ανάλυση ευαισθησίας. Για το λόγο αυτό, η διατριβή αυτή περιγράφει ένα ολοκληρωμένο μεθοδολογικό πλαίσιο το

οποίο με χρήση μεθόδων Γεωπληροφορικής θα συμβάλει στην κατανόηση του τρόπου λειτουργίας του συστήματος αλλαγών κάλυψης/χρήσης γης. Αυτή η διατριβή υποστηρίζει ότι ένα ολοκληρωμένο μεθοδολογικό πλαίσιο θα πρέπει να σχεδιαστεί με τέτοιο τρόπο ώστε i) να λαμβάνει υπόψη τις πολλαπλές κλίμακες που εμπλέκονται στο σύστημα αλλαγών κάλυψης/χρήσης γης, ii) να παρέχει πληροφορίες σε πολύ υψηλή θεματική ανάλυση λαμβάνοντας υπόψη όχι μόνο για τις αλλαγές μεταξύ κατηγοριών κάλυψης/χρήσης γης, αλλά και μεταβολές στην πυκνότητας τους, iii) να ανιχνεύει τις αλλαγές κάλυψης/χρήσης γης σε χρονική ανάλυση που επιτρέπει την ταυτοποίηση άνισων αλλαγών κατά την περίοδο μελέτης, κάτι που είναι απαραίτητο για την αντιπροσωπευτικότερη οριοθέτηση των σεναρίων, iv) να λαμβάνει υπόψη ένα ευρύ φάσμα παραγόντων αλλαγών κάλυψης/χρήσης γης (περιβαλλοντικοί, κοινωνικοί, οικονομικοί, πολιτικοί κ.α.), v) να παρέχει αποτελέσματα που υπόκεινται σε ανάλυση ευαισθησίας.

Αρχικά, διερευνάται ο βέλτιστος συνδυασμός μεθοδολογιών που απαιτούνται για επεξεργασία και ταξινόμηση δορυφορικών εικόνων, με σκοπό την ανίχνευση και ποσοτικοποίηση αλλαγών. Επιπρόσθετα διερευνάται ο ρόλος ειδικών μετρικών τοπίου στο χαρακτηρισμό της σύνθεσης και δομής του τοπίου. Η πρώτη περίπτωση μελέτης αφορά την προστατευόμενη περιοχή του Υμηττού, Αττικής. Οι περιοδικές αλλαγές που έλαβαν χώρα σε μια περίοδο 28 ετών χαρτογραφήθηκαν και ποσοτικοποιήθηκαν με χρήση μεθόδων τηλεπισκόπησης, Συστημάτων Γεωγραφικών Πληροφοριών (ΣΓΠ) και μετρικών τοπίου.

Με βάση αυτή την διερεύνηση, στη συνέχεια, το κομμάτι της ταξινόμησης δορυφορικών εικόνων βελτιώνεται με τη χρήση μια αποτελεσματικής ημιαυτόματης μεθοδολογίας κατά τη οποία η μοντελοποίηση λαμβάνει υπόψη πληροφορία που προέρχεται από ήδη διαθέσιμα δεδομένα κάλυψης/χρήσης γης. Η διαδικασία αυτή μειώνει σημαντικά το χρόνο που απαιτείται για την ταξινόμηση δορυφορικών εικόνων, με υψηλά ποσοστά ακρίβειας. Η μεθοδολογία εφαρμόζεται σε εθνική κλίμακα, σε ένα σύνολο 27 εικόνων Landsat. Ως ανεξάρτητες μεταβλητές το περιλαμβάνει συνολικά 20 επίπεδα που αφορούν εκτός από τα φασματικά κανάλια, και διάφορους πολυφασματικούς δείκτες, γεωμορφολογία εδάφους και απόσταση από το οδικό δίκτυο.

Στη συνέχεια, διερευνάται μια μεθοδολογία η οποία επιτρέπει την επίτευξη πολύ υψηλής θεματικής ανάλυσης, με ημιαυτόματη διαδικασία. Η μεθοδολογία περιλαμβάνει ενσωμάτωση πληροφορίας από ήδη υπάρχοντα δεδομένα, τα οποία

προέρχονται από πολλαπλές πηγές και είναι ετερογενή (πολλαπλές κλίμακες ανάλυσης, χωρική διακριτική ικανότητα). Η ομογενοποίηση τους έγινε με τη χρήση ενός αλγορίθμου μηχανικής εκμάθησης και το αποτέλεσμα διέκρινε την αστική κάλυψη/χρήση γης σε πέντε κατηγορίες ανάλογα με την πυκνότητα της. Η μεθοδολογία επίσης εφαρμόζεται σε εθνική κλίμακα.

Στο επόμενο κεφάλαιο εξετάζεται η αποτελεσματικότητα του συνδυασμού ενός μη παραμετρικού μοντέλου μηχανικής εκμάθησης με ένα χωρικό μοντέλο Cellular Automata σε μια προσέγγιση που συνδυάζει τα πλεονεκτήματα της κάθε μεθόδου. Περιγράφεται ένα παράδειγμα κατά το οποίο οι αλλαγές κάλυψης/χρήσης γης προσομοιώνονται με βάση διαφορετικά σενάρια τα οποία αναπαριστούν διαφορετικές οικονομικές συνθήκες και επίπεδα ανάπτυξης. Η προσέγγιση εφαρμόστηκε σε τοπική κλίμακα με περιοχή μελέτης τα Μεσόγεια Αττικής. Οι περιοδικές αλλαγές που έλαβαν χώρα την περίοδο 1985-2015 χαρτογραφήθηκαν, ποσοτικοποιήθηκαν και εξετάστηκε η συσχέτιση τους με 20 χωρικούς παράγοντες υποκίνησης (κοινωνικο-οικονομικοί, περιβαλλοντικοί, νομοθετικοί). Η χρονική κατανομή των αλλαγών οδήγησε στο σχεδιασμό τεσσάρων σεναρίων τα οποία αναπαριστούν διαφορετικές φάσεις οικονομικής ανάπτυξης. Η γεωγραφική κατανομή τους σε συνδυασμό με τους 20 παράγοντες υποκίνησης μοντελοποιήθηκαν για να προκύψουν οι επιφάνειες πιθανότητας αλλαγών. Με βάση αυτές τις επιφάνειες, οι αλλαγές προσομοιώθηκαν έως το έτος 2045 για κάθε ένα από τα τέσσερα σενάρια.

Τέλος, περιγράφεται η δημιουργία ενός ολοκληρωμένου μοντέλου προσομοίωσης των αλλαγών κάλυψης/χρήσης γης σε περιφερειακό επίπεδο (Αττική), συνδυάζοντας όλες τις μεθοδολογίες που συζητήθηκαν προηγουμένως. Ο στόχος είναι η διερεύνηση εναλλακτικών σεναρίων που αντικατοπτρίζουν διαφορετικές πραγματικότητες οικονομικής απόδοσης και εναλλακτικές επιλογές σχεδιασμού. Η ολοκληρωμένη προσέγγιση περιλαμβάνει, ημι-αυτόματη ταξινόμηση κάλυψη χρήσης γης σε πολύ υψηλή θεματική ανάλυση, διακρίνοντας συνολικά οκτώ κατηγορίες κάλυψης/χρήσης γης οι οποίες περιλαμβάνουν τη διάκριση της αστικής κάλυψης/χρήσης γης σε τέσσερις κατηγορίες ανάλογα με την πυκνότητα της. Ως ανεξάρτητες μεταβλητές το μοντέλο ταξινόμησης περιλαμβάνει συνολικά 20 επίπεδα που αφορούν εκτός από τα φασματικά κανάλια, και διάφορους πολυφασματικούς δείκτες καθώς και τη γεωμορφολογία εδάφους. Η ανίχνευση και ποσοτικοποίηση των διαχρονικών αλλαγών (1991-2016) οδήγησε στην γεωγραφική και χρονική αποτύπωση τους και εν συνεχεία στην συσχέτισή τους με 27 παράγοντες υποκίνησης (κοινωνικο-

οικονομικοί, περιβαλλοντικοί, νομοθετικοί), οι οποίοι προέρχονται από πολλαπλές πηγές και είναι ετερογενείς (πολλαπλές κλίμακες ανάλυσης, χωρική διακριτική ικανότητα). Η μοντελοποίηση της σχέσης αλλαγών και παραγόντων υποκίνησης οδήγησε στην κατασκευή 18 διαφορετικών επιφανειών πιθανότητας αλλαγών. Τα αποτελέσματα ενσωματώθηκαν σε ένα χωρικό μοντέλο Cellular Automata και οι αλλαγές προσομοιώθηκαν έως το έτος 2040 με βάση τρία σενάρια τα οποία αναπαριστούν διαφορετικές φάσεις οικονομικής ανάπτυξης. Τέλος εφαρμόστηκε ανάλυση ευαισθησίας των αποτελεσμάτων στη χωρική διακριτική ικανότητα των εισροών του μοντέλου. Σκοπός αυτού του εγχειρήματος ήταν να παραχθούν αποτελέσματα τα οποία είναι ακριβή ανεξάρτητα από τη χωρική διακριτική ικανότητα των εισροών του μοντέλου.

Chapter 1: Introduction

The intensification of research related to land cover and land use changes has emerged on the global environmental research agenda in the mid-1970s, with the discovery - realization that processes taking place at the surface of the Earth directly and indirectly affect the climate and the environment (Sagan et al., 1979). Land cover, according to Moser (1996), is a term referring to the physical attributes of the Earth's surface, including the soil, biomass, crops and human constructions. The term is often confused with that of land use which, by contrast, according to Mucher et al. (1993), is used to refer to the human activity that is directly related to the Earth. In other words, the term land use primarily refers to the purposes for which humans manage land cover by using its natural resources or impacting its ecological processes that in turn directly determine the land cover. An important distinction between the concepts of land use and land cover is the fact that the former focuses on economic activities occurring on a given surface of the land, while the second does not (Meyer and Turner II, 1996). Given that humankind alteration of Earth is substantial and ever growing, any significant changes in land use affect land cover and vice versa (Vitousek et al. 1997). Through a complex mechanism, pertaining to complex theory, changes in land cover react locally on land uses while also contributing to wider processes, such as climate change (Feddema et al. 2005), desertification (Gibbs and Salmon, 2015) and global environmental change (Turner II, 1994). Moreover, land cover changes hold wide-ranging significance for the structure and function of ecosystems, with equally far-reaching consequences for humans in every aspect.

Monitoring and understanding the land use/land cover (LULC) dynamics and the drivers behind it, in all aspects and scales, is therefore essential. The drivers involved to LULC change are multidimensional, inter-related, interact across different scales and can be broadly distinguished into bio-physical (eg physical characteristics of an area) and socio-economic (e.g. demographic, social, economic, political and institutional factors). For instance, the way the land resources are employed by land owners, generate immediate consequences on land cover which in turn dictate consequent adjustments in management strategies. Such adjustments usually have an influence on the socio-economic conditions that produced the original uses, and in turn lead to different uses. On the other hand, inversely, changes in the socio-economic

conditions alone may trigger land use and in turn land cover changes (Lambin and Meyfroidt, 2010).

The vast majority of land cover changes are human induced. Land cover changes are considered to be the most ancient influence of humans on the environment and these changes are substantial and growing (Sagan et al., 1979, Vitousek et al., 1997). As humanity grows, the technology expands along with the needs to resources and economic exploitation of the land. This human domination inevitably leads to Earth's transformation. This dual role of humanity, which contributes actively to land changes and at the same experiences the consequences of those changes, emphasizes the need for a better understanding of this human-environment interaction. Land cover changes have both desirable and undesirable impacts with a magnitude that varies from local to global scales.

The impacts that attracted the most attention by the scientific community are those adapted to the interactions of LULC changes with fundamental features and processes of the Earth such as climate, biochemical and hydrological cycles, biodiversity and ecosystem services (Foley et 2005, Vitousek et al., 1997). In particular, the composition and characteristics of the Earth's surface have an impact on the climate, determining the amount of evaporation and infiltration as well as the surface water runoff and thus the hydrological cycle (Becker and Bugmann, 2001) and the atmospheric composition (Falkowski et al., 2000). Changes in the hydrological cycle and the climate, in turn, have an impact on soil quality, leading to land degradation through soil erosion, which in the long-term triggers desertification (Le Houérou, 2002). The structure and composition of land cover is also directly related to biodiversity and rapid changes inevitably lead to fragmentation of habitats with multiple direct effects on species distribution (Tilman et al., 2001). All the above-mentioned, broad categories of impacts interact and in turn trigger further short-term and long-term effects. Equally important is the impact of changes in society. Changes in the climate, the water equilibrium, the quality and hence the productivity of soils, the biodiversity and the capacity of ecosystems to support human needs, make human society vulnerable to undesirable impacts which in turn directly threaten and may be disrupting to the economic and socio-political conditions of society (Foley et 2005, Schroter et al., 2005).

This dissertation is an effort to explore part of the complexity that characterizes the LULC changes system and provide insights into aspects that are commonly recognized as limitations still to be addressed. The literature stresses that many attempts toward this direction suffer from certain methodological drawbacks that undermine the validity of the outcomes and the transferability to other areas of study (see chapter 2;4;5;6;7;8). To name a few of these drawbacks, many studies do not properly address the multiple scales involved in LULC changes, concerning only factors operating at a single scale (e.g administration or planning scale). Additionally, in many studies, historical LULC changes are often addressed in low resolution (both thematic, spatial and temporal), which provides weak insights in every aspect of the phenomenon. Moreover, many studies are focused on a narrow perspective (e.g socioeconomic) disregarding other important factors while the majority of studies provide modeling results that are not subject to sensitivity analysis.

The overarching aim is to enhance the understanding of how the LULC system functions by building an integrated methodological framework devised for detecting historical changes, delineating and quantifying the factors of differing importance that drive these changes and sketching alternative future LULC trajectories. This dissertation argues that an integrated methodological framework should be designed in a way that sufficiently i) takes into account the multiple scales involved in LULC systems, ii) provides insights into hidden patterns, by taking into account not only the prominent changes between major LULC categories, but also changes in density, iii) detects LULC changes in a temporal resolution that enables the identification of uneven patterns throughout the study period which in turn enables the sound delineation of scenarios, iv) takes into account socioeconomic, biophysical, legislative and land use factors spanning a broad spectrum of LULC change driving forces, v) provides results that are subject to sensitivity analysis and unbiased to the technical details of inputs.

This dissertation will be able to provide answers to the following detailed research questions:

- How can heterogenous data, derived from various sources and expressed at multiple scales can be efficiently combined in a LULC modeling framework?

- How and on what extent do the patterns of outputs change over different scales and spatial resolutions?

- Which are the spatial determinants to the different types of LULC changes, in a region that experienced wide transformations?

- How and to what extent the socio-economic circumstances spatially influence the changes in LULC and how this information can be used to evaluate alternative pathways and policy options?

The originality of this dissertation lies in the focus to address key existing methodological shortcomings assembling in a unified framework the following elements: i) Development of a semi-automated process to achieve exhaustive training in conjunction with advanced processing while at the same time limiting the costs, both in terms of time and computational resources. ii) Efficient fusion of qualitative and quantitative data derived from multiple sources, expressed at various scales and resolution, allowing the full exploitation of available information. iii) Devise a robust modeling approach, designed to handle variables of different nature (continuous and categorical), insensitive to overfitting and collinearity and capable to handle large datasets without being computationally exhaustive. iv) Identification of important linkages and feedbacks between LULC patterns in a structural hierarchical manner delineating the contribution of various spatial determinants to each LULC change independently. v) High thematic resolution of analysis considering not only the actual type of land use but also the density, that is expected to reveal new patterns and different aspects. vi) Integrated modeling of LULC changes that takes into account the historical LULC spatial and temporal patterns and simulates alternative trajectories, by a multiple scales approach. vii) Accurate detection and projection of scattered unplanned development patches, which is often reported as a challenging task due to randomness and variation in structure and composition. viii) Coupling of non-parametric machine learning modeling with spatially explicit and application-oriented scenario-based simulation integrating the advantages of each approach to a unified ensemble. ix) Design of a methodological framework to generate results, derived from scenario-based simulations that are insensitive to bias and uncertainty related to spatial resolution of inputs and scaling issues.

The novelty of this dissertation can be summarized in four key aspects. i) It presents an integrated methodological framework that allows the efficient fusion of heterogeneous data expressed at multiple scales, forming a unified approach that allows the efficient detection and modeling of LULC changes, ii) it presents advances towards the detection and modelling of urban development in spatially continuous and discontinuous forms, iii) It introduces a sensitivity analysis for the identification of changes regardless of the spatial scale involved and iv) it demonstrates the utility of density based LULC change detection and modeling. The presented integrated framework is operational, cost effective and transferable.

The dissertation is structured in five core chapters. Chapter 4 explores the application of methodologies that employ multi-temporal satellite imagery and geo-informatics. This chapter describes the pre-processing steps of the satellite data, the classification and the change detection techniques adopted throughout the dissertation. Chapter 5 proceeds a step further the classification methodology and describes an efficient and robust semi-automated methodology for LULC classification using satellite imagery, and geo-informatics. Chapter 6 focuses on how a very high thematic resolution can be achieved. It describes a methodological framework for LULC thematic disaggregation, employing datasets from multiple sources, expressed in various scales and resolutions. Chapter 7 explores the effectiveness of coupling a non-parametric machine learning algorithm with a spatially explicit CA model, in an approach that combines the advantages of each method. It describes a scenario-based simulation modeling framework that sketches an appraisement of different alternative pathways related to economic circumstances and development. Finally, chapter 8 looks at the terrestrial part of Attica region. The aim was to explore, at the regional level, potential future LULC trajectories under three distinctive scenarios that reflect different economic performance realities and alternative planning options. To achieve this, an integrated approach that combined all previously discussed methodologies was designed. Chapter 9 provides answers to the detailed research questions and stresses the lessons learned and the concluding remarks of the dissertation.

References

Becker, A. and Bugmann, H. (2001) Global Change and Mountain Regions. IGBP Report 49.

Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., B. Moore, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V. and Steffen, W. (2000) Integrated understanding of the global carbon cycle- A test of our knowledge. *Science*, 290, 291–296.

Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., Washington, W.M. (2005). The Importance of Land-Cover Change in Simulating Future Climates. *Science*, 310, 1674-1678.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, C., Ramankutty, N. and Snyder, P.K. (2005) Global Consequences of Land Use. *Science*, 309, 570–574.

Gibbs, H.K. and Salmon, J.M. (2015). Mapping the world's degraded lands. *Applied Geography*, 57, 12-21.

Lambin, E.F. and Meyfroidt, P. (2010). Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use Policy*, 27, 108-118.

Le Houérou, H.N. (2002) Man-made deserts: Desertization processes and threats. *Arid Land Research and Management*, 16, 1–36.

Meyer, W.B. and Turner, B.L. II (1996) Land-use/land-cover change: challenges for geographers. *GeoJournal*, 39(3), 237–240.

Moser, S.C. (1996) A Partial Instructional Module on Global and Regional Land Use/Cover Change: Assessing the Data and Searching for General Relationships. *Geojournal*, 39(3), 241–283.

Mucher, C.A., Stomph, T.J. and Fresco, L.O. (1993) Proposal for a global land use classification. FAO/ITC/WAU, Rome/Wageningen.

Sagan, C., Toon, O.B. and Pollack, J. B. (1979) Anthropogenic albedo changes and the Earth's climate. *Science*, 206, 1363–1368.

Schroter D., Cramer W., Leemans R., Prentice C., Araujo M.B., Arnell N.W., Bondeau A., Bugmann H., Carter T.R., Gracia C.A., de la Vega-Leinert A.C., Erhard M., Ewert F., Glendining M., House J.I., Kankaanpaa S., Klein R.J., Lavorel S., Lindner M., Metzger M.J., Meyer J., Mitchell T.D., Reginster I., Rounsevell M., Sabate´ S., Sitch S., Smith B., Smith J., Smith P., Sykes M.T., Thonicke K., Thuiller W., Tuck G., Zaehle S. and Zierl B. (2005) Ecosystem Service Supply and Vulnerability to Global Change in Europe. *Science*, 310, 1333–1337.

Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W.H., Simberloff, D., and Swackhamer, D. (2001) Forecasting Agriculturally Driven Global Environmental Change. *Science*, 292, 282–284.

Turner, B.L. II. (1994). Local faces, global flows: the role of land use and land cover in global environmental change. *Land Degradation & Rehabilitation*, 5, 71-78.

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

Chapter 2: Land use/land cover monitoring and modeling

2.1 Land use/land cover monitoring

Given that a major proportion of the Earth's land cover is influenced and thus shaped by human activities and land use (Vitousek et al. 1997), long-term observation of LULC is essential (Lambin and Meyfroidt, 2011). Traditional field data approaches face several limitations as they are destined to a local extent. Moreover, they are source demanding in terms of personnel, equipment and time and also, they are limited by topographic and climatic conditions and low accessibility to remote areas. Remote sensing (RS) along with Geographic Information Systems (GIS) can be combined to successfully provide spatially consistent and detailed LULC information, a prerequisite in order to monitor the Earth effectively (Coppin et al., 2004; Rozenstein and Karnieli, 2011). An increased number of satellite data (Belward and Skøien, 2014) can facilitate the growing demand for multi-spectral and multi-temporal information of the Earth's surface over a wide range of scales and data types in order to monitor the Earth effectively (e.g Hansen et al., 2013; Schneider et al., 2009; Zhu and Woodcock, 2013). However, adopting RS techniques and relying on satellite data involves facing an imperative trade-off. Very high resolution (VHR) imagery comes with an amount of costs, a fact that acts as an obstacle in large scale and multi-temporal approaches. On the other hand, low resolution data are free of charge and, more recently, in abundance. However, this type of data may be unsuitable to monitor certain phenomena and to capture patterns that usually occur on a smaller scale, like LULC changes.

With the Landsat program running for almost five decades now, high spatial resolution satellite images have been widely used for monitoring LULC and its changes (Hansen and Loveland, 2012). The advantages of using Landsat data are the suitable spatial resolution for LULC monitoring, the very high temporal resolution due to the low revisit cycle of the satellite and the high spectral resolution offered. To add on that, the opening of the Landsat archive readily available for download with no costs, makes it the only feasible option for studies that span some decades of time and cover large extents (Wulder et al. 2016).

Recent technological and methodological advancements contributed to the wide spreading increase of digital spatial databases. Nowadays, various and at multiple

resolutions LULC datasets exist but with discussed limitations and challenges still to be addressed (Giri et al., 2013; Herold et al., 2008). On a regional scale, for Europe, two frequently used databases are the Coordinate Information on the Environment (CORINE), a pan-European LC map for the years 1990, 2000, 2006 and 2012 provided by the European Environmental Agency and the Pan-European Land Cover Monitoring database (Mucher et al., 2001). These two databases suffer from limitations, namely low spatial resolution and minimum mapping units (MMU), inconsistency from one country to another, lack of rigorous accuracy assessments and thus reliability (Neumann et al., 2007; Waser and Schwarz, 2006). The most important limitation, of the available, both global and regional datasets that is still to be addressed, is the discrepancy between LC classes, their overarching definitions, their nomenclature and thus the heterogeneity of information provided. Another serious issue about the CORINE LC database, which emanates from the fact that is produced by each country separately, is the outdated available information for some countries like for example Greece.

Additionally, the feasibility of using already available datasets in a range of research applications and management activities is limited by their low thematic resolution. Thematic resolution refers to the detail in the definition of LULC categories and thus it directly determines the amount of geospatial information of hard classified categorical data. The amount of detail in a LULC map, defines how meaningful and insightful the map is for a wide range of research questions. Several authors have explored the effects of thematic resolution in land use modeling (Conway, 2009; Pontius and Malizia, 2004), land-cover pattern analyses (Buyantuyev and Wu, 2007) and landscape indices behavior (Bailey et al. 2007), converging that the outcomes are significantly influenced. Whilst thematic resolution is important to a range of applications, available regional and global datasets in most cases represent important LULC categories lumped into one or two broad classes, a fact that is far from reality on the ground (Potere et al. 2009). Thus, the usage of these data to research efforts that are centered on areas that faced a multitude of transformations is limited. On the contrary, depending on the area of study and the dominant transformations occurred in that area, the discrimination of LULC categories according to their density and continuity is crucial. For example, in areas that faced various forms of urbanization, taking into account the changes occurred in not only the extent but also to the density of urban

areas, will reveal important insights. The same applies to research efforts focusing on areas with forested land or cropland.

Therefore, to avoid the limitations stressed above, studies tailored to assess the LULC changes on a specific area and period cannot rely on already existing hard classified datasets. Temporally consistent and accurate LULC maps need to be produced to satisfy the growing demand for spatially explicit data. To meet these requirements a number of researchers focused on introducing increasingly sophisticated approaches, which are at the same time less source demanding and labor intensive. Currently, there is a clear trend in the research agenda to develop and suggest automated (e.g. Chen et al., 2012; Comber et al., 2004; Huth et al., 2012; Radoux et al., 2014; Yuan et al., 2009) or semi-automated (Jiang et al., 2012; Xian et al., 2009) LULC classification approaches. A key element to accomplish minimum intervention by the user is the utilization of existing and readily available LULC data, to train the classifier (Chen et al., 2012; Jiang et al., 2012; Klein et al., 2012; Radoux et al., 2014; Xian et al., 2009). Under the assumption that changes usually occur only to a small fraction of the land, incorporating in the process accurate but outdated information is reasonable and promising to eliminate remaining gaps in LC mapping.

2.2 Land use/land cover modeling

In recent decades, a wide variety of LULC change models have been developed to meet the scientific community's need for understanding how LULC evolves and why (Agarwal et al. 2002, Briassoulis, 2000). Models are widely used to analyze the complex structure of linkages and feedbacks between drivers of change, determine their relevance to particular LULC 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, models can not exactly replicate complex interactions and nonlinear relations, but they are rather, at a fundamental level, a process that provides a platform that when formally expressed, consists of a tool that allows certain experiments (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 simple and straightforward task. But when dealing with inherently complex systems, as is the case with LULC changes, the models are capable to represent and exemplify only small fractions of the whole mechanism in order to highlight important processes.

The major technological breakthrough removed technical barriers and in conjunction with the rapid methodological advances, facilitated the proliferation of available LULC models, geographical datasets and software. Contemporary theoretical approaches and analytical tools are used to describe, interpret and predict, both qualitatively and quantitatively, the behavior of LULC changes. Nowadays models are frequently used to answer questions such as which factors contributed to the current state of land cover, or how much and what are their interdependencies (McBurney, 2012). Apart from the delineation of causes and consequences, the recent methodological and technological advancements have opened the way to more articulated models which are capable to answer more complex questions such as what the possible outcomes would be if alternative pathways were followed, which alternative outcome is the most desired out of many as well as a diverse range of similar 'what if' scenarios. Scenario-based analysis is now increasingly being adopted by a range of disciplines pertaining to LULC change, as fruitful experiments for exploring the possible future trajectories of the historical and current trends (Murray-Rust et al. 2013). Taking as granted that, in reality, the number of potential futures is infinite (Greeuw et al., 2000), scenarios are not used to exactly predict the future, but rather to explore a range of possible futures 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, by delineating plausible, presumably coherent and internally consistent storylines of different socio-economic development trajectories (Rounsevell and Metzger, 2010).

Modeling LULC have its origins in the family of spatial interaction models but the conceptualization of the models is built around economics, regional science, sociological and political economy, and nature–society interaction theories (Briassoulis, 2008). Precisely, given that land use, in other words the exploitation of land cover resources is directly related to economic motives and the principal drivers that drive LULC changes are socioeconomically oriented, the von Thunen's agricultural

land-rent and Alonso's urban land-rent theories still stand as a cornestone of modeling applications.

Models designed to analyze LULC dynamics can be divided into various 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, Parker et al. 2003). A simple, non-exhaustive, classification with regard to their methodological origins would include i) Empirical–statistical models using multivariate regression and geostatistical analysis (e.g. He and Lo, 2007, Millington et al., 2007, Poelmans and Van Rompaey, 2010). Models of this type often suffer from limitations such as sensitivity to outliers, collinearity issues and factors compatibility (Eastman et al. 2005). ii) Stochastic and optimization models (e.g. Brown et al. 2002), which are processes that consider one objective or simply convent multiple objectives into one, and the optimization takes place with the use of weighting methods (Ma and Zhao, 2017). iii) Dynamic process-based simulation (Veldkamp and Fresco 1996, Verburg et al., 2002) which often involve multiple models subdivided in modules that capture non-spatial (e.g demand) and spatially explicit (e.g allocation) processes. iv) Agent-based modeling (e.g. Manson, 2005, Robinson et al. 2012), which capture decision making processes at the individual, household or neighborhood levels. Agent-based models can be very complex as part of distributed artificial intelligence method and when it comes to agent behavior and are often parametrized with qualitative social survey data and other types of participatory approaches (Zagaria et al. 2017). v) Markov chains (e.g. Dongjie et al. 2008) which are frequently used to delineate LULC changes as a transition probability scheme that is statistically estimated based on past transition proportions between different types of LULC. vi) Cellular automata (CA) (literature review by Sante et al 2010) which are based on transition rules and neighborhood interactions between LULC categories. CA are spatially-explicit and application-oriented and are capable to represent stochastic, non-linear processes in a conceptually simple way (Batty et al. 1997). The basic principle of CA models is that LULC changes can be explained by the current state of a cell and its surroundings and the transition rules that dictate the possible change of a cell, can be expert-based or calculated from statistical analysis of historical LULC changes (White and Engelen, 2000).

When modeling LULC, the scale spatial resolution and 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 by the modeler 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). In turn, resolution refers not only to spatial resolution, but also to thematic, which is the level of precision in LULC categories and to temporal resolution which is used to refer to the time span and frequency of the analysis. Modelling LULC changes 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). 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 broader extent case studies and coarse 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. Taking as a fact that all models are driven by the data used as inputs, studies focusing on individual phenomena, considering only a single scale and using data that are only particularly suitable to their study area, are not always representative, transferable or reproducible to different scales and are characterized by uncertainty and various critical assumptions (Kok and Veldkamp 2001, Veldkamp et al., 2001, Verburg et al. 2006).

References

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.

Bailey, D., Herzog, F., Augenstein, I., Aviron, S., Billeter, R., Szerencsits, E. and Baudry, J. (2007). Thematic resolution matters: indicators of landscape pattern for European agro-ecosystems. *Ecological Indicators* 7, 692–709.

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

Belward, A.S. and Skøien, J.O. (2014). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation. *ISPRS Journal of Photogrammetry and Remote Sensing* 103, 115-128.

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.

Briassoulis, H. (2008). Land-use policy and planning, theorizing, and modeling: lost in translation, found in complexity? *Environment and Planning B: Planning and Design*, 35(1), 16–33.

Brown, D.G., Goovaerts, P., Bumlckl, A. and 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. and Lange, M.D. (2013). Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Environmental Sustainability*, 5, 452–457.

Buyantuyev, A. and Wu, J. (2007). Effects of thematic resolution on landscape pattern analysis. *Landscape Ecology* 22, 7–13.

Chen, X., Chen, J., Shi, Y. and Yamaguchi, Y. (2012). An automated approach for updating land cover maps based on integrated change detection and classification methods. *ISPRS Journal of Photogrammetry and Remote Sensing* 71, 86–95.

Comber, A.J., Law, A.N.R. and Lishman, J.R. (2004). Application of knowledge for automated land cover change monitoring. *International Journal of Remote Sensing* 25, 3177–3192.

Conway, T.M. (2009). The impact of class resolution in land use change models. *Computers Environment and Urban Systems* 33, 269–277.

Coppin, P., Jonckheere, I., Nackaerts, K. and Muys, B. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25, 1565 – 1596.

Dongjie, G., Weijun, G., Kazuyuki, W. and 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., and 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. and Ahn, T.K. (2000). The concept of scale and the human dimensions of global change: a survey. *Ecological Economics*, 32, 217–239.

Giri, C., Pengra, B., Long, J. and Loveland, T. R. (2013). Next generation of global land cover characterization, mapping, and monitoring. *International Journal of Applied Earth Observation and Geoinformation* 25, 30–37.

Greeuw, S., van Asselt, M., Grosskurth, J., Storms, C., Klomp, N., Rothman, D., Rotmans, J., Agency, E.E. (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.

Hansen, M.C. and Loveland, T.R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment* 122, 66–74.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. and Townshend, J.R.G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342, 850–853.

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

Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A. and Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and

accuracy in existing 1 km datasets. *Remote Sensing of Environment* 112, 2538–2556.

Huth, J., Kuenzer, C., Wehrmann, T., Gebhardt, S., Quoc Tuan, V. and Dech, S. (2012). Land Cover and Land Use Classification with TWOPAC: towards Automated Processing for Pixel- and Object-Based Image Classification. *Remote Sensing* 4, 2530-2553.

Jiang, D., Huang, Y., Zhuang, D., Zhu, Y., Xu, X. and Ren, H. (2012). A Simple Semi-Automatic Approach for Land Cover Classification from Multispectral Remote Sensing Imagery. *PLoS ONE* **7**(9), e45889.

Klein, I., Gessner, U. & Kuenzer, C. (2012). Regional land cover mapping and change detection in Central Asia using MODIS time-series. *Applied Geography* 35(1-2), 219‑234.

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

Lambin, E.F. and Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming scarcity. *Proceedings of the National Academy of Sciences of the United States of America* 108 (9), 3465–3472.

Ma, X. and 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., Gottschalk Druschke, C., Hychka, K., Kenney, M.A., Kok, K. and Borsuk, M.E. (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.

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.

Mucher, C.A., Champeaux, J.L., Steinnocher, K.T., Griguolo, S., Wester, K., Heunks, C., Winiwater, W., Kressler, F.P., Goutorbe, J.P., ten Brink, B., van Katwijk, V.F., Furberg, O., Perdigao, V. and Nieuwenhuis, G.J.A. (2001). Development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental monitoring; the PELCOM report. Alterra-rapport 18 178/CGI-report 6, Alterra, Wageningen, the Netherlands.

Murray-Rust, D., Rieser, V., Robinson, D.T., Milicic, V. and Rounsevell, M. (2013). Agent-based modelling of land use dynamics and residential quality of life for future scenarios. *Environmental Modelling & Software*, 46, 75-89.

Neumann, K., Herold, M., Hartley, A. and Schmullius, C. (2007). Comparative assessment of CORINE2000 and GLC2000: Spatial analysis of land cover data for Europe. *International Journal of Applied Earth Observation and Geoinformation* 9, 425–437.

Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J. and Deadman, P. (2003) Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review. *Annals of the Association of American Geographers*, 93(2), 314–337.

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

Pontius Jr., R.G. and Malizia, N.R. (2004). Effect of category aggregation on map comparison. In: Engenhofer, M.J., Freska, C., Miller, H.J. (Eds.), GIScience 2004. Springer, New York, pp. 251–268.

Potere, D., Schneider, A., Angel, S. and Civco, D.L. (2009). Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *International Journal of Remote Sensing* 30, 6531–6558.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C. and Defourny, P. (2014). Automated Training Sample Extraction for Global Land Cover Mapping. *Remote Sensing* 6, 3965-3987.

Robinson D.T., Murray-Rust D., Rieser V., Milicic V. and 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.

Rozenstein, O. and Karnieli, A. (2011). Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Applied Geography* 31, 533–544.

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. and 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.

Schneider, A., Friedl, M. A. and Potere, D. (2009). A new map of global urban extent from MODIS satellite data. *Environmental Research Letters* 4 article 044003.

Veldkamp A. and 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. and 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. and 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. and 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. and Veldkamp, A. (2006). Downscaling of land use change scenarios to assess the dynamics of European landscapes. *Agriculture, Ecosystems and Environment*, 116, 39–56.

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

Waser, L.T. and Schwarz, M. (2006). Comparison of large-area land cover products with national forest inventories and CORINE land cover in the European Alps. *International Journal of Applied Earth Observation and Geoinformation* 8, 196–207.

White, R. and 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.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R. and Woodcock, C.E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* 122, 2–10.

Xian, G., Collin, H. and Fry, J. (2009). Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. *Remote Sensing of Environment* 113 (6), 1133–1147.

Yuan, H., Van Der Wiele, C.F., & Khorram, S. (2009). An Automated Artificial Neural Network System for Land Use/Land Cover Classification from Landsat TM Imagery. *Remote Sensing* 1, 243-265.

Zagaria, C., Schulp, C.J.E., Kizos, T., Gounaridis, D. and 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.

Zhu, Z. and Woodcock, C.E. (2013). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment* 144, 152–171.

Chapter 3: Research methodology

This dissertation seeks to devise an integrated methodological framework for LULC changes modeling that will be able to sufficiently address all the aforementioned aims and objectives and provide answers to the research questions. The approach will demonstrate how to take into account the multiple scales involved in LULC systems, will detect, map and quantify the LULC historical changes in sufficient spatial, temporal and thematic resolution, will incorporate in the modeling a broad spectrum of LULC change driving forces socioeconomic, biophysical, legislative and land use factors and will present results that are subject to sensitivity analysis.

To meet the aims and objectives and to address the scientific challenges that emerge, the dissertation is structured into five methodological steps that are demonstrated in five applications (Table 1).

First, it describes the pre-processing steps of the satellite data, the classification and the change detection techniques adopted throughout the dissertation. The approach is demonstrated in Hymettus mountain, Athens. The LULC types are classified and quantified over a study period of 28 years. Post classification comparisons, in the form of cross-classification and cross-tabulation are applied to detect, map and quantify the LULC changes spatio-temporally. Additionally, a set of landscape metrics, suitable to delineate the structure and composition of the LULC are computed for each year.

Next, the LULC classification proceeds a step further by devising an efficient and robust semi-automated methodology for LULC classification using satellite imagery, and geo-informatics. The application is demonstrated at the national scale. Information extracted as training, from already available land cover datasets, reducing significantly the time consuming and labor-intensive process of training a classification model.

Next, the focus was on thematic disaggregation and efficient data fusion in order to achieve very high thematic resolution. The approach adopted a non-parametric machine learning modeling framework that allowed the fusion of existing, readily available and with acceptable accuracies datasets, in order to achieve a very high thematic resolution in which urban LULC is classified into five categories according to

continuity and density. This application is also demonstrated at the national scale with a high degree of automation.

Table 1. Methodological steps and applications of this dissertation.

Theme	Spatial Scale			Model	Techniques
	Local	Regional	National		
Forests	Chapter 4	-	-	Random Forest (RF) Classification	Change detection Cross tabulation-classification Landscape metrics
LULC	-	-	Chapter 5	RF Classification	Semi-automated LULC classification
Urban	-	-	Chapter 6	RF Classification	Semi-automated LULC classification Thematic disaggregation Data fusion
Urban/Artificial non-Urban	Chapter 7	-	-	RF Classification / RF Regression / Cellular Automata	Change detection Cross tabulation-classification Landscape metrics Data fusion Scenario -based simulation
LULC	-	Chapter 8	-	RF Classification / RF Regression / Cellular Automata	Semi-automated LULC classification Thematic disaggregation Data fusion Change detection Cross tabulation-classification Landscape metrics Data fusion Scenario -based simulation Multiple resolution Sensitivity analysis

Next, the effectiveness of coupling a non-parametric machine learning algorithm with a spatially explicit CA model is explored. A scenario-based simulation modeling framework is devised in order to sketch an appraisement of different alternative pathways related to economic circumstances and development. The focus is on a locality (Messoghia plain, Attica) that experienced vast transformations and looked at the urban and industrial LULC categories. After detecting and quantifying the

periodic LULC transitions occurred during 1980–2015, the observed changes were combined with 20 dynamic, biophysical, socio-economic and legislative factors, to produce transition potential surfaces. Four scenarios, that reflect four distinct chronological frames marked with uneven development, different economic performance realities and land-use planning, were projected until 2045.

Finally, the final application explores at the regional level, potential future LULC trajectories under three distinctive scenarios that reflect different economic performance realities and alternative planning options. To achieve this, an integrated approach that combined all previously discussed methodologies was designed. The focus is the terrestrial part of Attica region and the study period spanned 25 years (1991-2016) and LULC is classified into eight categories, achieving very high thematic resolution. Change detection techniques in the form of cross-classification and cross-tabulation are used in order to map and quantify the periodic LULC changes occurred. A total of 18 different possible LULC transitions are identified and combined with 27 different factors derived from multiple sources and represented in different scales, units and resolutions. Simulation models are calibrated and fine-tuned in order to project the LULC changes until 2040, under the three scenarios. Finally, the results are subject to a multi-resolution sensitivity analysis in a process that outputs of each model run are compared at several spatial resolutions in order to identify areas of future change disregarding the spatial resolution of inputs.

Chapter 4: Quantifying spatio-temporal patterns of forest fragmentation in Hymettus Mountain, Greece

Authors: Dimitrios Gounaridis, George N. Zaimes, Sotirios Koukoulas

Journal: Computers, Environment and Urban Systems 46, (2014), 35–44

Abstract

The rapid land use/cover change (LULCC) and landscape fragmentation occurring around the world is largely attributed to human induced factors. Landscape fragmentation has become a central issue in landscape ecology and conservation policies due to its direct influence on biodiversity which consequently endangers the sustainability of ecological goods and ecosystem services. Thus, fragmentation monitoring and assessment is a critical issue in land use planning and sustainable environmental management in order to avoid any irreversible negative consequences. This research explores the application of methodologies that employ multi-temporal satellite imagery, combined with geographical information systems and landscape metrics, to assess forest fragmentation. The objective is to determine spatio temporally the LULCCs focusing on the woody vegetation in Hymettus Mountain of Greece over the last decades. The study area, which has been designated as a Natura 2000 site, is situated near the city of Athens. It faces various perturbations triggered by socio-economic factors and the absence of an ongoing contextual appraisal for conservation. To quantify the LULCCs, nine Landsat images spanning 28 years are classified. Post classification comparison is applied to generate transition maps. Additionally, eight landscape metrics are calculated. The change detection results identify hot-spots of forest fragmentation where mitigation measures should be taken, so that further irreversible alteration of the ecosystem is prevented. The landscape metrics advocate that, during the last three decades, the woody vegetation has steadily been more fragmented. The primary direct causes are economic driven intense anthropogenic activities along with frequent wildland fires whereas the indirect cause is the absence of a sustainable environmental management and conservation strategy.

Keywords: Forest fragmentation, change detection, Landsat, post classification comparison, landscape metrics.

2.1 Introduction

Socioeconomic development in Greece has considerably been influenced by land-based economic activities. These are closely related to the structure and function of landscapes, as long as agriculture, grazing, forest harvesting and mining still constitute partially income source for its residents (Papanastasis et al., 2008). Landscape refers to a mosaic of heterogeneous territory composed of sets of interacting ecosystems (Forman, 1995). It is characterized by dynamics, composition and configuration that are governed by natural processes and human activities (Forman, 1995). The term composition describes the abundance and variety of different patch types, while configuration refers to the physical distribution and spatial character of patches within a landscape mosaic (McGarigal and Marks, 1995).

Over the last century, natural ecosystems in Europe have been substantially transformed because of socio-economic and political changes (Reger et al., 2007). These transformations are expected to continue. More specifically the structure of forested landscapes has changed as a result of natural and anthropogenic disturbances, ecological succession and degenerative trends (Lambin and Meyfroidt, 2010; Ji et al., 2006). Human activities have modified the environment to the extent that landscapes are increasingly becoming dominated by human settlements, artificial cultivation fields with only scattered fragments of natural ecosystems (Vitousek et al., 1997). Most natural conservation reserves are progressively being surrounded by intensively modified environments and in the long-term, are deemed to function as isolated natural ecosystems (Wolter and White, 2002).

Forest fragmentation is a dynamic process in which contiguous forest tracts are progressively being sub-divided into smaller, geometrically complex isolated patches (Gibson et al., 1988). Caused by either natural or anthropogenic agents, forest fragmentation seriously threatens key features and processes of the earth such as climate, biophysical and hydrological cycles, biodiversity and ultimately ecosystem services. More specifically, the composition and characteristics of the earth's forests, aggregated at a global scale, affect the climate. First the earth's forests can impact the hydrological cycle by determining the amount of evapotranspiration, infiltration and surface water runoff (Becker and Bugmann, 1999). Second forests affect the atmospheric composition, by determining the moisture content in lower atmospheric layers (Chase et al., 1999). Finally, they also determine the emission of greenhouse

gases functioning as a natural sink (Falkowski et al., 2000; Rockstrom et al., 2009). Moreover, alterations in the hydrological cycle and climate affect the soil quality, since the latter is gradually being degraded through erosion that progressively leads towards desertification (Le Houérou, 2002). The status of forest ecosystems is also directly related to biodiversity. Rapid changes in health, composition and structure of ecosystems inevitably lead to fragmentation of habitats with multiple effects such as species extinction (Gaston, 2005; Schroter et al., 2005; Tilman et al., 2001). Hence, forests alteration clearly endangers the sustainability of ecological goods and ecosystem services (Costanza et al., 1997).

To face these issues the European legislation via the Habitats and Birds Directives ((92/43/EEC and 79/409/EEC respectively) has identified the need to protect natural ecosystems providing the legal basis to establish the Natura 2000 network. The overall goal of this network is to implement management plans that will preserve high-value natural ecosystems, protect the biodiversity and the ecological functions of natural ecosystems and enhance sustainable management. This network comprises of approximately 26.000 protected areas (Special Areas of Protection and Special Areas of Conservation) and covers a total area of about 850.000 km^2, more than 20% of total EU territory (Apostolopoulou and Pantis, 2009; Dimopoulos et al., 2006; Papageorgiou and Vogiatzakis, 2006; Tsiafouli et al., 2013). The effectiveness of protection strategies (namely: conservation conventions, protocols and parks) has attracted the interest of scientists the last decades (e.g. Bruner et al., 2001; Mallinis et al., 2011; Seto and Fragias, 2007).

Understanding the landscape pattern and quantifying its spatial relationships and changes through time, is essential for the continuous monitoring and assessment of ecological processes. Remote sensing (RS) combined with geographic information systems (GIS) and landscape metrics (LM) can successfully provide spatially consistent and detailed information about landscape structure, a prerequisite to study ecosystem services, sustainable resources management and land use planning (Gustafson, 1998; Riitters et al., 1995; Shi et al., 2011).

Recent developments in the field of satellite RS have increased the use of spatially explicit landscape analyses. At the same time, numerous landscape indices have been developed to quantify landscape structure and spatial heterogeneity based on the composition and configuration of landscapes (Chen, 2002; Coppin et al., 2004; Cushman et al., 2004; Liu and Zhou, 2005; Seto and Fragias, 2007). Metrics are

calculated at three different hierarchical levels: landscape, class, and patch. The landscape level metrics includes all patches within a defined landscape. The class level metrics represent the spatial distribution and patterns of a land use/cover class, such as a woodland, and mainly involve differences between classes. Finally, the patch level metrics are calculated on the individual patches within each class.

The aim of this study is to identify general trends and subtle patterns of forest extent in Hymettus Mountain, Attica Prefecture, Greece that has been exposed to persistent anthropogenic activities over the last three decades and is divided into three main complementary axes: i) Evaluation of the potentials and limits of an integrated earth observation approach as a valuable tool for monitoring. Furthermore, the methodology presented in this paper is literally a cost-effective proposal that can be adopted by land use planners and ministry policy makers, management agencies and environmental researchers. ii) Exploration of the effectiveness of the protection status of the area, since it belongs to the Natura 2000 network. iii) Provision of important feedback and historical evidence associated with the implications of decision-making being monopolized by economic growth on the one hand and being deprived of effective conservation measures, on the other hand.

Athens, being the largest conurbation and the densest populated area of Greece, shows two major contradictions. On the one hand, it concentrates around one third of the total population of the country, a phenomenon triggered by socio-economic developmental needs and the comparative advantages of the city to attract investments and development opportunities (Chorianopoulos et al., 2010). On the other hand, it is located in a basin where mountains on the three sides and the sea on the other restrict its growth. As a consequence of those economic and demographic pressures, the region is facing urban sprawl problems due to the increasing population and the urgent socioeconomic development that has occurred during the last decades (Weber et al., 2005). The urban, industrial and construction grid is expanding considerably, along with the road network and the relative linear residential developments, in order to serve rapidly growing needs (Chrysoulakis et al., 2013; Nikolakopoulos, et al., 2005). All the above-mentioned changes have led to major environmental implications (Forman and Alexander, 1998). The urban expansion, which is in expense of the natural reserves, is expected to continue as a new road network expansion for the city is scheduled by the Ministry of the Environment, Spatial Planning and Public Works, (2006).

It is hypothesized that forest spatial extent, composition and distribution have been changing since the early 1980s, thus leading to fragmentation, shrinkage and attrition of forested areas. The focus is mainly on forest fragmentation induced by human activities, specifically agriculture, unplanned urbanization and industrialization, heavy exploitation and wildland fires. The spatial extent and distribution of LULCCs is assessed to acquire information about the dynamics of the area, by identifying increased fragmentation hot spots.

This paper is organized as follows: First the Hymettus Mountain and the Landsat imagery used are described. Second, the image preprocessing, the classification, employing a machine learning methodology and the accuracy assessment are presented. Following post classification comparisons to assess the periodic LULCC are presented along with the proportion of each class, revealed by the classification of the imagery. Next the time series of eight LM calculated to assess the landscape structure changes are presented. Finally, the results are discussed in detail and the main points are highlighted.

2.2 Methodology and Data

2.2.1 Study site

Hymettus Mountain is located in the south-central part of Attica Prefecture, between the Athens conurbation, Penteli Mountain, Messoghia Plain and Saronikos Gulf. Its north to south length is approximately 20 km and its width ranges from 4 to 6 km. It covers an area of 8820 ha with a maximum altitude of 1025 m. The dominant habitat classes of the region are coniferous (22%), evergreen broadleaved (33%) woodland and heaths, shrubs and garrigue (25%) (Georghiou et al., 1995). Hymettus is protected nationally and internationally. The ecosystem hosts a variety of rare flora (approximately 40 endemic species) and ecologically essential habitats for breeding, nesting and wintering grounds of rare bird species, regularly occurring migratory birds, mammals, and invertebrates. It is part of the Natura 2000 network (code GR 3000006 Hymettus mountain - Kaisariani Aesthetic Forest - Lake Vouliagmeni) as a Special Protection Area (SPA), Site of Community Importance (SCI) and Area of Conservation (SAC).

Although sustainable management has been promoted in the area because of its protected status, the ecosystem has been remarkably degraded during the last decades (Vlachogiannis et al., 2012; Weber et al., 2005). This is primarily due to rapid unplanned urban expansion, devastating wildland fires (e.g.1995, 1998 and 2007) that have frequently occurred in the last decades and intensive anthropogenic activities (industrial facilities, military installation, power production facilities and quarries). To fulfill the aims of this study, the wider area is chosen as the study site, in order to include the pressures on the boundaries (Figure 1).

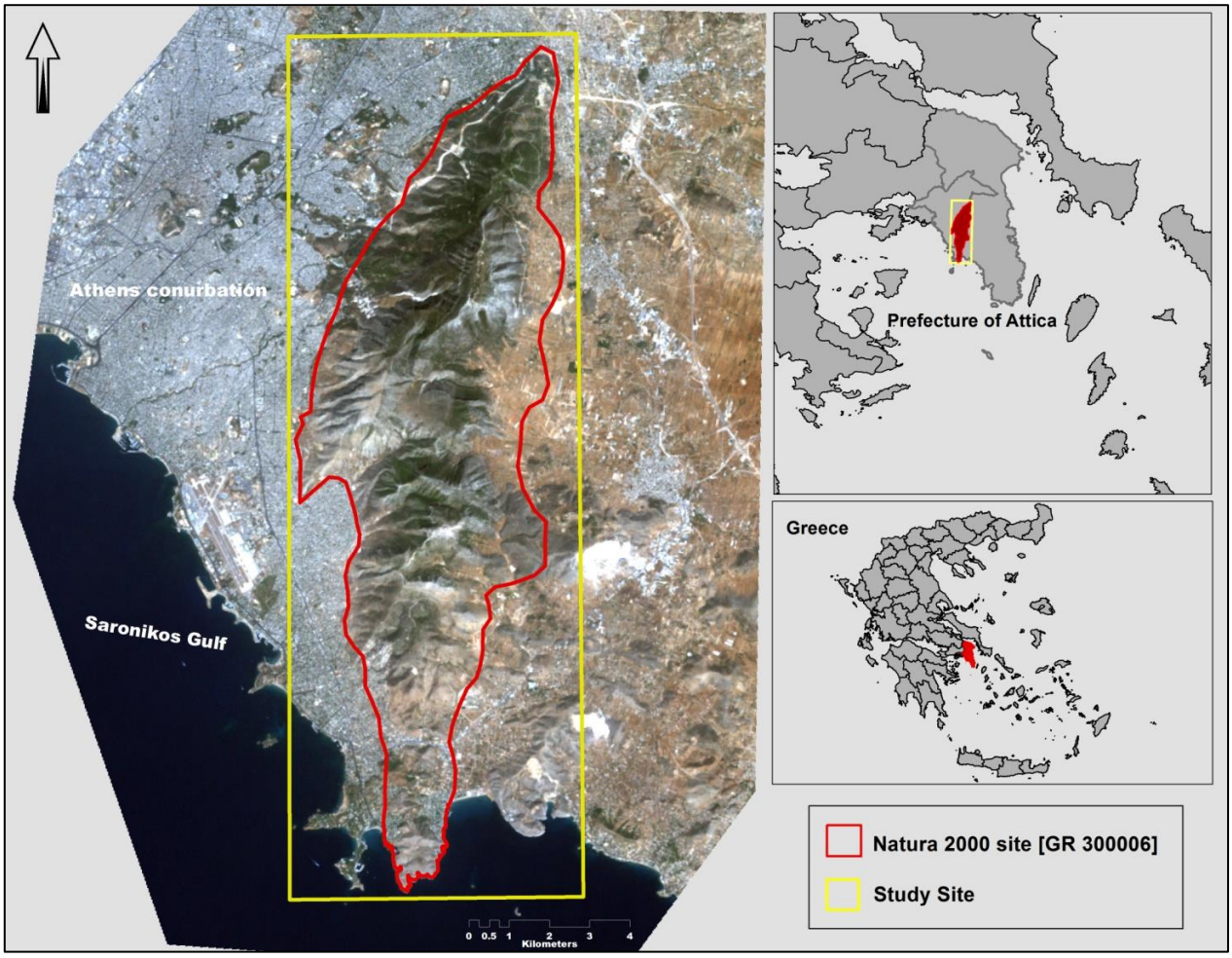


Figure 1. The location of the study site in Attica Prefecture, in Central Greece (Landsat TM 6 2011 – R: band3, G: band2, B: band1).

2.2.2 Landsat data

Landsat imagery is used to generate series of land cover data, integrated in a GIS framework to quantify changes in a spatial and temporal manner. Landsat imagery is a cost effective, with a high temporal scale, satellite record of the Earth's surface. In our research, nine mid-resolution Landsat satellite images (30m) spanning over 28 yrs (1985-2013) are employed to identify LULCC through time. Phenological variation can complicate classification consistency among images. During summer, late spring and early autumn the vegetation is vigorous and is considered the best period to distinguish among different land cover types (Mas, 1999; Shi et al., 2011). Therefore, ideally, for multi-temporal change detection of vegetated areas the images should be acquired during the summer months (June to August). However, additional criteria (e.g how to avoid the scan line corrector problem of Landsat 7 after 2003, budget limitations and availability of cloud free data) led to a set of nine Landsat images acquired from May – August (Table 1).

Table 1. The characteristics of the satellite images that were used as the primary data to corroborate the change detection analysis.

	Date	Sensor	Satelite type	Resolution (m)	Path/Row
1	19 May 1985	Thematic Mapper (TM)	Landsat 5	30	183/34
2	10 June 1987	Thematic Mapper (TM)	Landsat 4	30	183/34
3	21 June 1991	Thematic Mapper (TM)	Landsat 5	30	183/34
4	6 August 1999	Enhanced Thematic Mapper Plus (ETM+)	Landsat 7	30	183/34
5	26 July 2001	Enhanced Thematic Mapper Plus (ETM+)	Landsat 7	30	183/34
6	26 July 2003	Enhanced Thematic Mapper Plus (ETM+)	Landsat 7	30	183/34
7	24 July 2009	Thematic Mapper (TM)	Landsat 5	30	183/34
8	16 August 2011	Thematic Mapper (TM)	Landsat 5	30	183/34
9	19 July 2013	Operational Land Imager (OLI)	Landsat 8	30	183/34

2.2.3 Data pre-processing

The standard image processing techniques of extraction, layer stacking, re-projection, radiometric enhancement and topographic correction are performed on the nine Landsat images. First the nine images are referred to a common projection (Greek Geodetic Reference System, 1987). Second the images have to be atmospherically and radiometrically calibrated to avoid any discrepancies due to the multi-temporal and multi-sensor type of analysis (Hall et al., 1991; Lu et al., 2002). Three calibration steps are applied in the radiometric correction procedure for Landsat imagery. The top-of-atmosphere (TOA) reflectance is first calculated to correct the reflectance differences caused by the solar distance and angle (Vermote et al., 1997b). The equations used are:

a) to calculate the at-sensor radiance in W/(m² * sr * μm),

$$\text{Gain} = (L_{max} - L_{min}) / (DN_{max} - DN_{min}) \quad (1)$$

$$\text{Bias} = L_{min} - \text{gain} * DN_{min} \quad (2)$$

$$\text{Radiance} = \text{gain} * DN + \text{bias} \quad (3)$$

where, *Lmax* and *Lmin* are the calibration constants, *DN* are the initial digital numbers of the imagery, while *DNmax* and *DNmin* are the highest and the lowest points of the range of the rescaled radiance in *DN* (Chander et al., 2009),

b) to calculate the at-sensor reflectance:

$$\text{Sun radiance} = [Esun * \sin(e)] / (PI * d\text{^}2) \quad (4)$$

$$\text{Reflectance} = \text{radiance} / \text{sun\_radiance} \quad (5)$$

where, *d* is the earth-sun distance in astronomical units, *e* is the solar elevation angle, and *Esun* is the mean solar exoatmospheric irradiance in W/(m² * μm) and

c) to invert the TOA reflectance and obtain surface reflectance, the 6Smodel (Vermote et al., 1997a) is used. The model performs absolute atmospheric correction, taking into account the Bi-directional Reflectance Distribution Function (BRDF), and calculating Rayleigh and aerosol scattering, gaseous absorption and transmission. The minimum input variables in estimating the atmospheric conditions for every scene are related to meteorological visibility and aerosol conditions at the date and time of image acquisition.

After the radiometric calibration and atmospheric correction, the area has to be topographically corrected because it is mountainous. The first step in correcting the

topographic effects is to compute the illumination angle, based on the equation (Civco, 1989):

$$IL = \cos(s) * \cos(z) + \sin(s) * \sin(z) * \cos(a - o) \quad (6)$$

where, *s* is the terrain slope angle, *z* is the solar zenith angle, *a* the solar azimuth angle and *o* is the terrain aspect angle.

To compute the slope gradient and aspect for the illumination, a Digital Elevation Model (DEM) of similar resolution to the Landsat image obtained. The second step to achieve the removal of topographic shadows is to apply the C-correction method, a semi empirical approach developed by Teillet et al., (1982). The equation, as described by Meyer et al., (1993), is:

$$P_h = \rho_i (\cos(s) + C / IL + C) \quad (7)$$

where P_h is the reflectance of a horizontal surface, ρ_i is the reflectance of an inclined surface and *C* is the correction parameter b/m (m = the inclination of the regression line describing the correlation between the original band (b) and the illumination)

2.2.4 Classification

After the nine Landsat images have been calibrated, the classification into land use classes is performed implementing the Random Forests (RF) machine learning algorithm (Breiman, 2001). RF is a tree structured classifier generating a "forest" of randomized base regression trees. Each random tree in the "forest" is composed of nodes at different levels: a root node, a set of internal nodes (splits) and a set of terminal nodes. The root node is formed by all training samples and searches only across a randomly selected subset of the input variables, in order to determine the best split for each node. Each individual tree predicts the target response, while the forest predicts the target as the average of the individual tree predictions. This process is repeated until the desired number of trees has been built (1000 in this analysis). The various outputs are combined in a final result, using a majority vote. For a full detailed description of the RF algorithm, theory and applications, the reader is referred to Breiman (2001).

According to the approach by Symeonakis et al., (2007), Principal Components Analysis (PCA) and Tasseled Cap (TC) transformations are performed in order to produce an information portfolio for the training required by RF. PCA is computed separately for the three visible bands (1, 2, and 3) and for the middle infrared bands (5

and 7). The first Principal Component produced of each run, along with the near infra-red layer (band 4), which is the least correlated with the other bands, are used. The TC transformation products, Soil Brightness Index (SBI) and Green Vegetation Index (GVI) are also merged to the layer stack of the predictor variables, because of their capacity to separate vegetation from bare features during the classification process (Symeonakis et al., 2007). Considering that, the training data must represent all classes, additional criteria have been found after several tests, so as to improve the classification accuracy results in our case study. The Normalized Differential Vegetation Index (NDVI) (Tucker, 1979) is added because of its capacity to describe vegetation density and condition. It has been found that it improves the performance of RF in respect to the forested areas. Moreover, the slope layer has been found to significantly improve the accuracy of predictions related to the sparsely vegetation or bare land areas. Finally, a Euclidean distance to the road network layer of the area, treated under the assumption that intense human activity (thus artificial land use) involves dense road networks, has also been found to improve the artificial surfaces class results. A set of randomly distributed points (n=350) is used to train the algorithm. The values of the eight merged layers are collected on the location of every point and manipulated as predictor values.

For the accuracy assessment of the classified maps, a new set of points (n=200) is randomly distributed to the 1991, 2009 and 2013 reference images. The observed accuracy percentages reach 87% for the 1991, 84% for 2009 and 89% for 2013 scene (aerial photographs from 1995 and 2008, high resolution satellite imagery from 2013 and coarser resolution land cover maps from 1990 are used to define the reference data). Since matching reference data are not available for the rest of the produced land cover maps, a series of Boolean logic and "if-then-else" rules are devised based on possible transitions (e.g. an 'urban-forest-urban' transition is not possible). In this way, for each land cover change map, the land cover maps in between are used as control data (e.g. in 1985-1987-1991 series, the 1987 serves as a 'control' map so that the changes occurred between 1985-1991 would be accepted). This approach is inspired by the research of Symeonakis et al., (2012) who use series of intermediate satellite images and conditional probability networks to ensure the detection of valid changes. Since single pixels are often unreliably classified (probably as a result of spectral mixes), another commonly used step is to eliminate them from the classified images and to replace their values with the mode of the neighborhood pixels (Symeonakis et al., 2007;

Weber and Puissant, 2003). The neighborhood is defined with a 3x3 window centered to the single pixel that is eliminated. The minimum mapping unit is then defined as the patch area covered by at least 2 pixels.

The final step is the aggregation of the classes with a focus on forest fragmentation. Initially, seven land use classes were chosen to represent the land cover of the area. As the research is orientated towards forest fragmentation monitoring, change detection focuses on the transformation of woody vegetation land cover to sparsely vegetated area (sva)/bare or artificial land uses (indicating deforestation) and vice-versa (indicating reforestation). Therefore, the initial classification classes are aggregated into three in order to better be distinguished from the woody vegetation (Table 2).

Table 2. Initial land cover types derived from the Random Forests classification and the aggregated class covers made to improve the assessment of forest fragmentation.

Initial land cover classes	Forest change associated classes
Artificial surfaces Agricultural areas	Artificial land use
Forest-land Shrub maquis and garrigue	Woody vegetation
Open spaces with little or no vegetation / bare rocks	Sparsely vegetated areas (sva) / bare
Water bodies	N/A (masked out)

2.2.5 Change detection

The aim of the change detection analysis is to obtain spatial and quantitative information about periodic LULCC from forested (woody vegetation) to sparsely vegetated/bare areas or artificial land-uses and vice-versa. In this study, post classification (PC) comparison in the form of cross-classification is employed because of its advantage to provide quantitative information about changes and more specifically 'from-to' change class information. The pixel by pixel nature of this change

allows both the areal extent and spatial distribution of land-cover changes to be quantified (Coppin et al., 2004). PC comparison is conducted for three intervals, consisting of three consecutive images each (1st 1985-1987-1991, 2nd 1999-2001-2003, 3rd 2009-2011-2013). This is done to highlight the LULCCs that occurred during the study period, to temporarily allocate the LULCCs and to enable possible associations with significant events that occurred (e.g. establishment of the new International Airport, Athens 2004 Olympics, economic crisis). The generated transition maps quantitatively identify the periodic LULCC that have occurred during the last 28 yrs.

2.2.6 Landscape Metrics

After the remote sensing analysis, the time series LM for each of the maps are calculated. Selecting metrics for a given study involves a number of considerations. First many LM are highly correlated, providing redundant information (Riitters et al., 1995; Cushman et al., 2008). Second the objectives of the study, the spatial characteristics of the system and the ecological processes under investigation determine which metrics best describe the studied phenomenon (Gustafson, 1998; Herold et al., 2005). After the aforementioned have been taken into consideration, eight LM (Table 3) at the class level were chosen to examine the spatio-temporal forest changes in the Hymettus landscape composition and configuration during the last 28 years. Indices of size, density, shape, isolation, proximity, connectivity and aggregation are implemented to study forest fragmentation. Patch density (PD), is expected to increase as the forested area becomes more fragmented whereas edge density (ED) is also expected to increase depicting the shape complexity and spatial heterogeneity. The largest patch index (LPI) of woody vegetation is expected to decrease as a result of fragmentation, affecting directly the habitat quality (especially that of migratory birds). The mean shape index (MSI) depicts management measures and thus an increase would indicate mismanagement. The Euclidean area weighted mean nearest neighbor distance (ENN) and aggregation index (AI) are widely applied to characterize isolation, proximity and neighborhood of the landscape patches, illustrating effects on habitat quality, species distribution and landscape stability. The clumpiness index (CL) is a metric of distribution - isolation of classes and increases as they become more randomly distributed and fragmented. Finally, the spliting index (SPL) calculates the subdivision of classes and is expected to increase, while a class becomes more fragmented.

Complete descriptions and mathematical expressions of these metrics are provided in McGarigal and Marks, (1995).

Table 3. The Landscape Metrics (LM) computed for this study and their correlation to forest fragmentation.

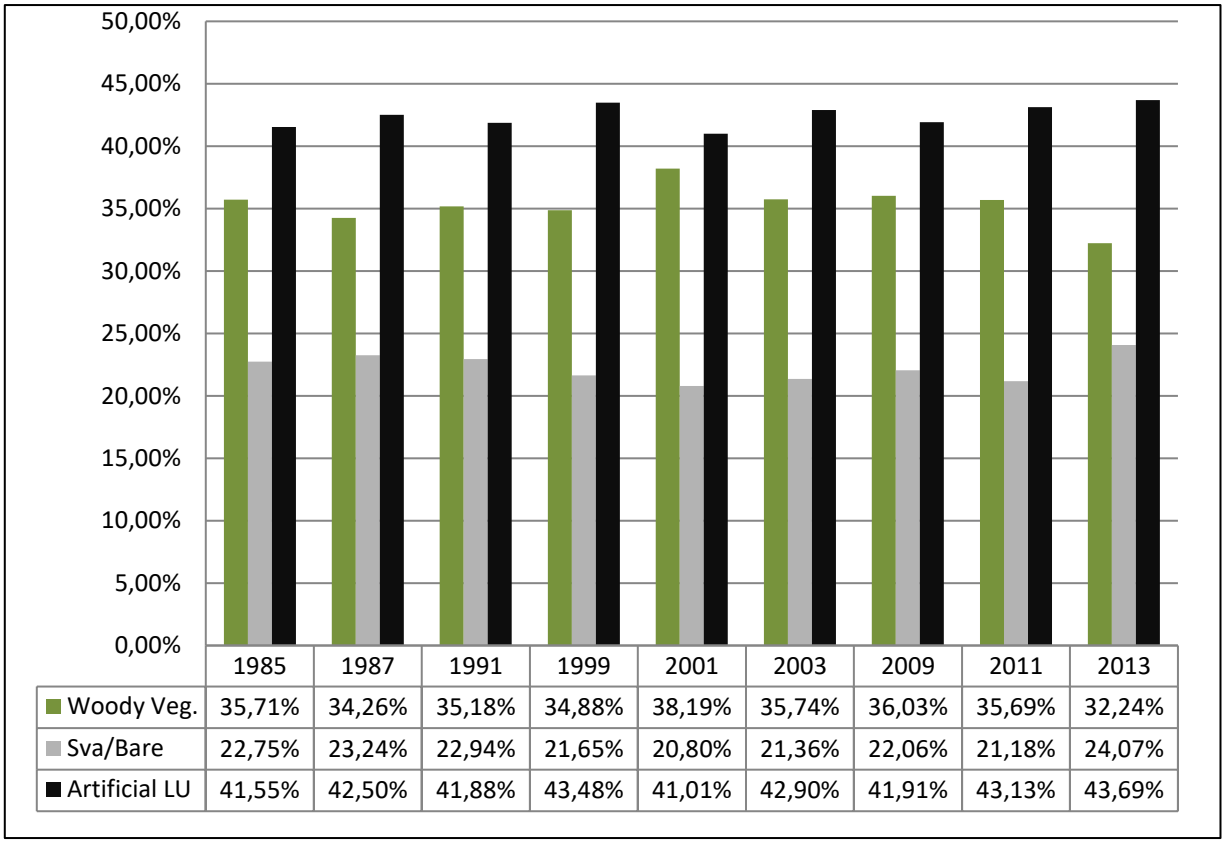
Landscape Metrics	Abbreviation	Description
Patch Density (# of patches/100 ha)	PD	Lower density of patches indicates less fragmented classes
Edge Density (m/ha)	ED	Indicates the shape complexity and spatial heterogeneity of a class
Largest Patch Index (%)	LPI	Percentage of total landscape area occupied by the largest-sized patch. Indicator of dominance
Mean Shape Index	MSI	Patches are less geometrically complex in managed forests
Mean Euclidean nearest-neighbor distance (m)	ENN_MN	Quantifies patch isolation and therefore fragmentation of classes
Aggregation Index (%)	AI	Measures isolation and compactness. Higher values indicate lower fragmentation
Clumpiness index	CL	Lower clumpiness indicates more fragmented classes
Splitting Index	SPL	Higher rates indicate higher subdivision of classes

2.3 Results and Discussion

2.3.1 Change detection

Figure 2 and the attached table depict the results generated by the classification of the nine images and the aggregation of the seven initial classes into three. The rates illustrate the proportion covered by each of the three classes, revealing some trends. From 1985 to 2001, the woody vegetation area seems to increase approximately by 2.3 % whereas the sva/bare land decreased by approximately 2 %. Following, the woody vegetation decreases with the total decrease reaching approximately the 6 % in 2013, which is also the lowest rate observed during the study period. At the same time, the sva/bare land proportion increases by approximately 4 %. Furthermore, the artificial land use class, which comprises of the urban areas along with the construction sites, the road network and the agriculture areas, remains consistently high with only slight

fLULCtuations throughout the study period. Assuming that urban areas, road networks and infrastructure sites rarely change, the slight decreases observed are attributed to changes in the agriculture sector (e.g. land abandonment).



	1985	1987	1991	1999	2001	2003	2009	2011	2013
Woody Veg.	35,71%	34,26%	35,18%	34,88%	38,19%	35,74%	36,03%	35,69%	32,24%
Sva/Bare	22,75%	23,24%	22,94%	21,65%	20,80%	21,36%	22,06%	21,18%	24,07%
Artificial LU	41,55%	42,50%	41,88%	43,48%	41,01%	42,90%	41,91%	43,13%	43,69%

Figure 2. The summary of the Landsat Area Classification statistics for 1985, 1987, 1991, 1999, 2001, 2003, 2009, 2011 and 2013. The values are relative percentage of the total area (no data pixels were masked out).

The generated maps from the PC comparison depict nine combinations ‘‘from – to’’ change information derived for each of the three time intervals 1985-1991, 1999-2003 and 2009-2012 (Figure 3). For the purpose of the analysis, the first interval can be linked to the early years before the announcement of the new International airport in Spata, the second to the pre-Olympic Games of 2004 period where infrastructures started to take place and the third is linked to the post-Olympic Games of 2004 period and the economic crisis.

The main focus is on the "woody vegetation to sva/bare or artificial land use" transition that describes the decline in forest extent, indicating forest fragmentation. During the years of the first interval, two core areas, distinguished with red color (Fig 3. 1a;1b), of "woody vegetation to sva/bare" transition are observed. Noteworthy, both are located next to urban core areas.

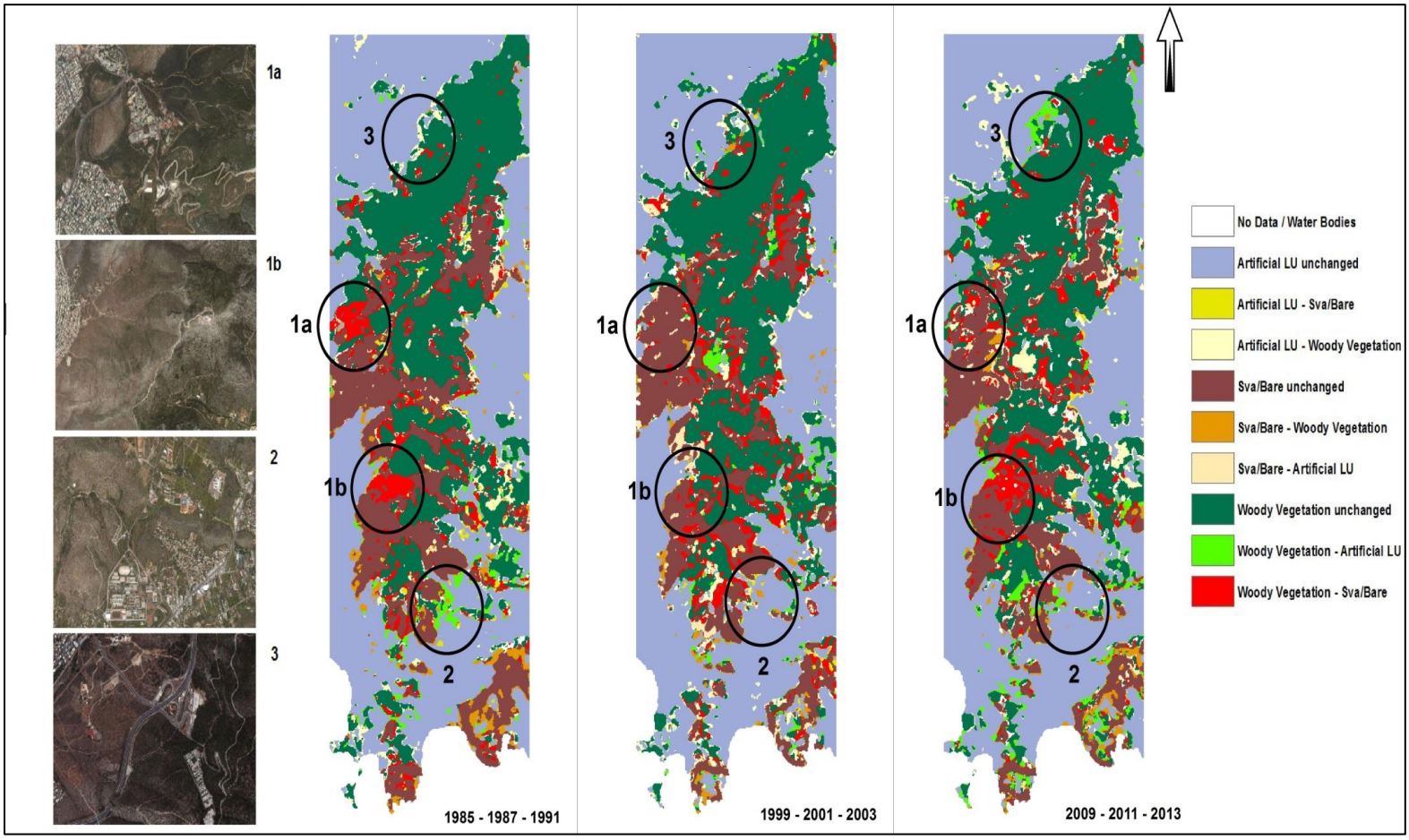


Figure 3. The PC comparison depicting in three groups (out of a total of nine Landsat images) the LULCCs that occurred in Hymettus Mountain Attica Prefecture, Greece during the last 28 years.

These two major areas of deforestation remain sva/bare land during the second interval. During the third interval, the north patch of deforestation reveals scattered artificial land use class patches while the south patch expands significantly. Another interesting part is illustrated in Figure 3 (2) that is related to the "woody vegetation to artificial land use", and colored in light green. A core area of this class, located to the south of the mountain, was converted to artificial land use (agricultural) during the first interval and remains as is until 2013. Another example of deforestation of this type is observed during the third interval in the north part (Fig 3. 3). This patch of woody

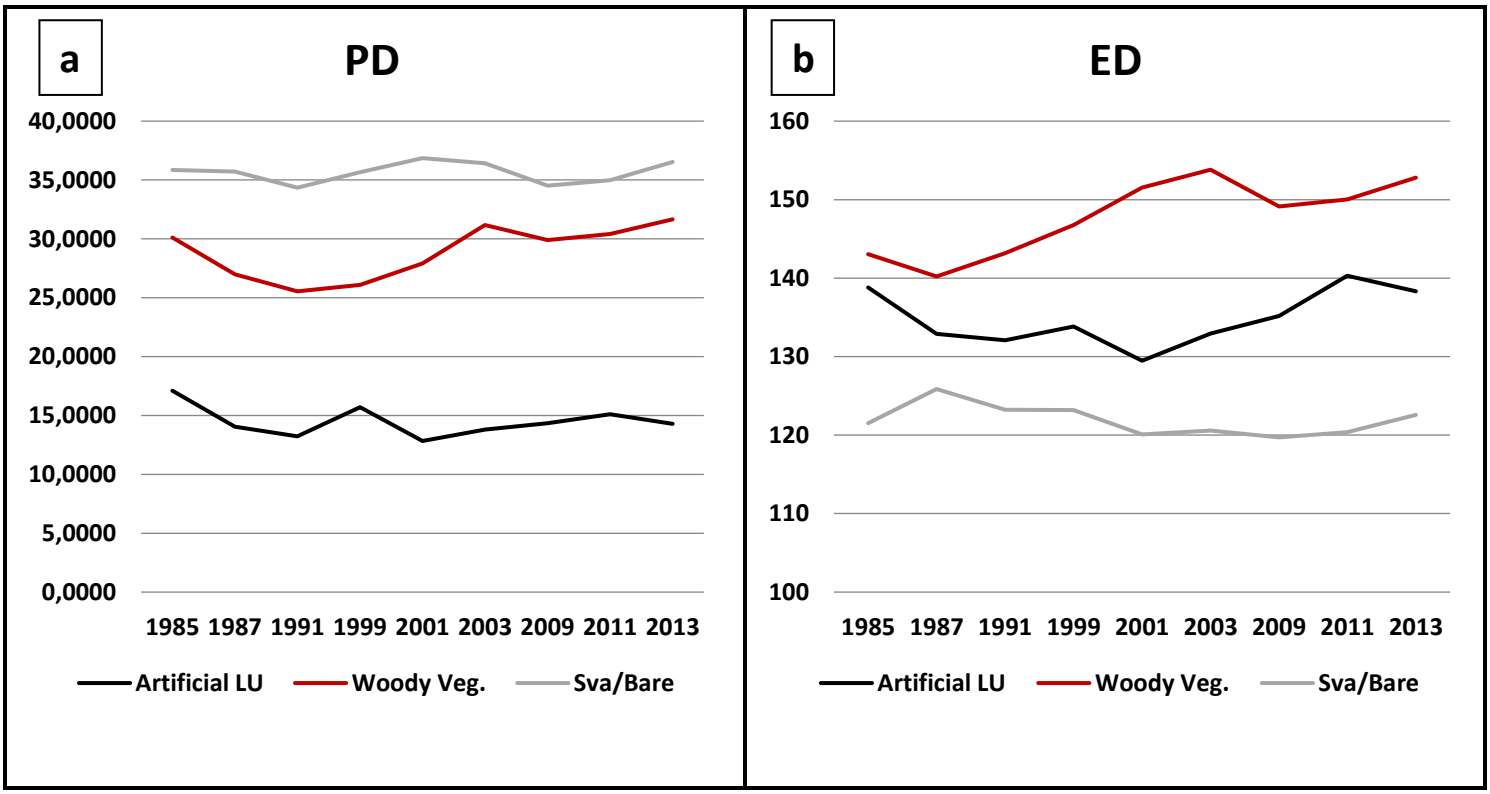
vegetation is the “victim” of the new road network expansion. Several other smaller patches of the same type behave the same way. Small patches of the “artificial land use to woody vegetation” class are mostly located in the agricultural area to the left of the mountain. Finally, what can be inferred from the transition maps is that the “woody vegetation to sva/bare” along with the “woody vegetation to artificial land use” patches are increasing during the second and third interval in comparison with the first. This observed increase along with their disorderly widespread distribution highlight the consistent and clear evidence of forest fragmentation.

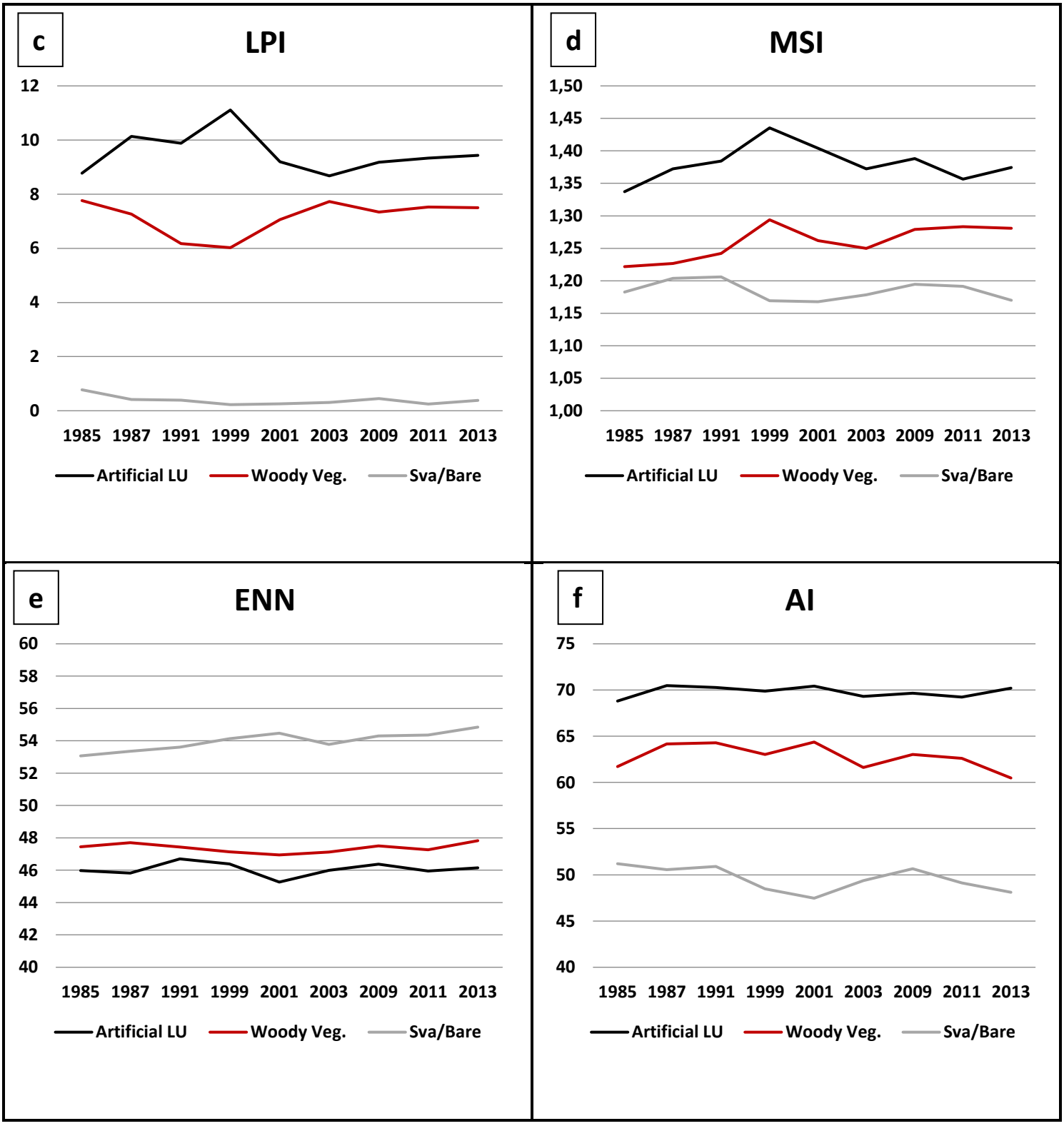
2.3.2 Landscape Metrics

The previous findings also coincide with the results of the eight LM that are computed for this study and provide an insight into the spatial configuration of the changing landscape. The PD and ED values (Figure 4a; 4b), of the woody vegetation class, depict an increasing trend during 1991-2003 that remains high until 2013. Another interesting trend derived both from the ED and the LPI (Figure 4c) values is the inversely proportional behavior between the woody vegetation and the artificial land use classes. This trend reflects the previous remarks that woody vegetation area progressively became surrounded by artificial land uses and shrunk. FLULCtuations in the woody vegetation ED, PD, and LPI, rapidly affect the habitat quality, especially for migratory species by altering their distribution and population density and thus their diversity. Therefore, if the landscape continues to follow the present trends of fragmentation, the habitat of certain species of the region will decrease and degrade.

Considering the spatial configuration of the boundaries and therefore the shape complexity, the MSI values of the woody vegetation class also indicate that the patches are progressively becoming more geometrically complicated probably because most changes occur at class boundaries (Figure 4d, also see ED in 4b). As expected the values behave inversely proportional with the sva/bare class. The area weighted ENN was preferred to be computed, instead of the mean ENN, so that the inter-patch connectivity in respect to the size of the patches is indicated. Thus, the ENN values illustrated in Figure 4e show that the connectivity core and larger areas of woody vegetation class remain relatively constant. This index provides additional information and highlights the findings of the PC results, showing that the majority of transitions, in expense of the woody vegetation coverage, are made at their boundaries. The AI values, (Figure

4f) suggest that the woody vegetation class successively tend to be more isolated and less compact, right after 2001, when the highest value is observed. Yet the values of artificial land use constantly remain high, with just slight fLULCtuations during the study period. Furthermore, the CL values (Figure 4g), indicate that the distribution of woody vegetation class increased during the 1985-1987 period and remained high until 1991. Right after 1991, which is the end of the first interval in the transition map, the values decrease steadily until they finally reach their lowest value in 2013. Another interesting part is the behavior of the sva/bare and the artificial land use class values that increase remarkably right after 2001 and 2003, respectively. Finally, the SPL values (Figure 4h), confirm that woody vegetation areas show a tendency to be subdivided after 2003, a crucial period for the area, due to pressures in the wider area that resulted from increased development during this period.




c
LPI
12
10
8
6
4
2
0
1985 1987 1991 1999 2001 2003 2009 2011 2013
Artificial LU
Woody Veg.
Sva/Bare
d
MSI
1,50
1,45
1,40
1,35
1,30
1,25
1,20
1,15
1,10
1,05
1,00
1985 1987 1991 1999 2001 2003 2009 2011 2013
Artificial LU
Woody Veg.
Sva/Bare
e
ENN
60
58
56
54
52
50
48
46
44
42
40
1985 1987 1991 1999 2001 2003 2009 2011 2013
Artificial LU
Woody Veg.
Sva/Bare
f
AI
75
70
65
60
55
50
45
40
1985 1987 1991 1999 2001 2003 2009 2011 2013
Artificial LU
Woody Veg.
Sva/Bare

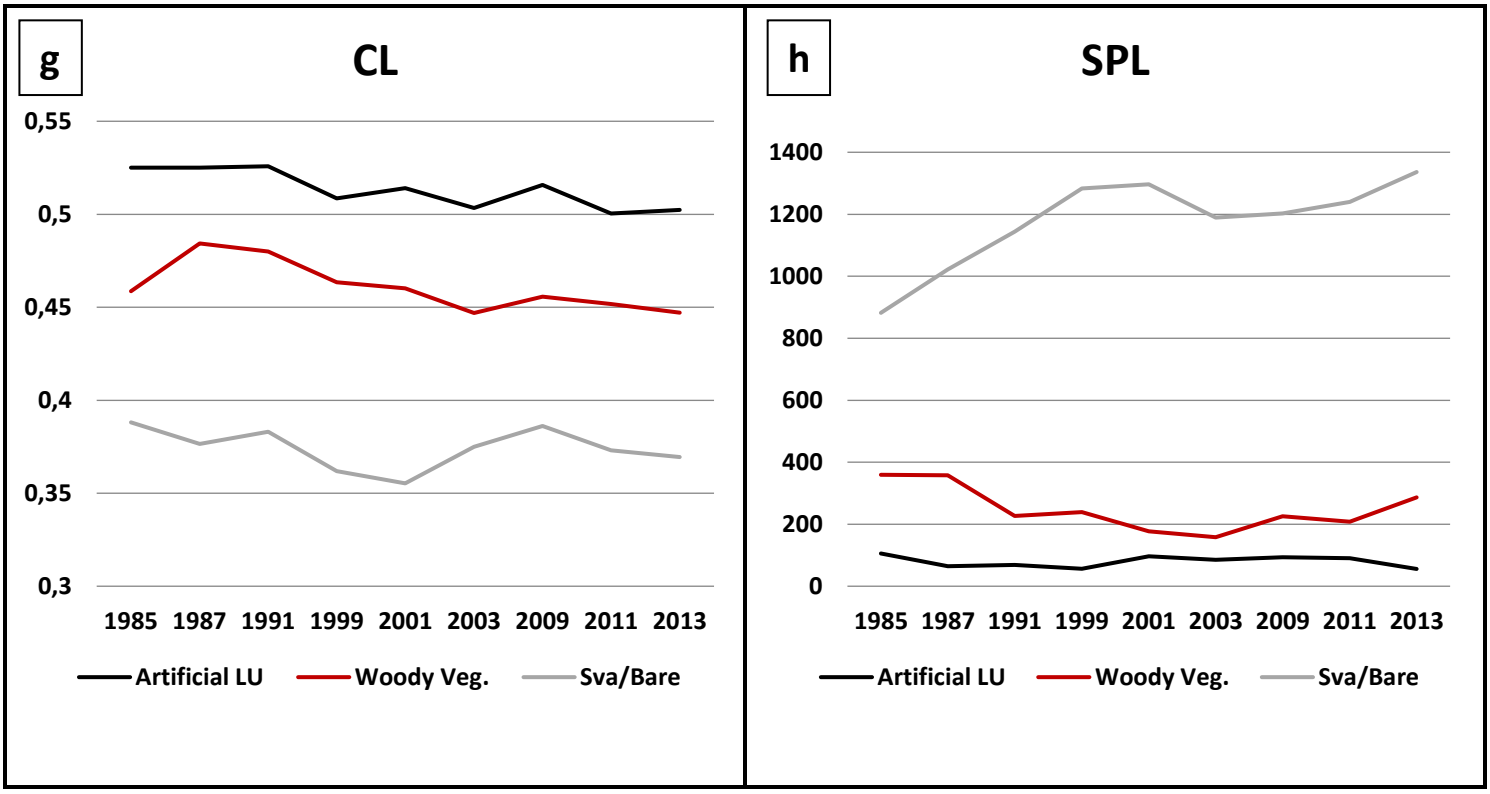


Figure 4. The time series (1985–2013) of the eight Landscape Metrics (LM) computed for each aggregated class level, for Hymettus Mountain, Attica Prefecture, Greece. Brief descriptions and full names of the LM are provided in Table 3.

2.3.3 Natura 2000 sites

In Greece, despite the establishment of many Natura 2000 sites, the implementation of conservation policy is problematic (Apostolopoulou and Pantis, 2009; Apostolopoulou, et al., 2012; Dimitrakopoulos, et al., 2004; Papageorgiou and Vogiatzakis, 2006; Tsiafouli et al., 2013). The lack of clear goals and the divergence between stated and actual goals leads to incorrect interpretations of conservation objectives and the decisions made to satisfy economic and/or developmental interests (Apostolopoulou and Pantis, 2009). There are internal contradictions of the institutional framework due to the fact that the levels of planning between the managing organizations overlap. There is also no clear differentiation of the responsibilities and functions among various organizations - such as Ministry of Environment, Energy and Climate Change, Organization on the Regulatory Planning and Protection of the Environment of Athens, Local Authorities, Forest Service of Penteli etc. - that manage in some capacity the area (Papageorgiou and Vogiatzakis, 2006). As a consequence,

some land-uses may not be compatible based on the current institutional framework (Tsiafouli et al., 2013) in some cases. As it is understood, Hymettus mountain is no exception to all aforementioned inconsistencies. To complicate even more the situation, a very complex ownership in the area, including the church, municipalities, military, cooperatives and individuals, usually functions as a stumbling block. A major concern is also the scheduled road network expansion (approximately 83 km) that will intensify the fragmentation of the forested ecosystem.

Thus, forest fragmentation can be largely attributed to the pursuit of economic development, such as self-driven land use policies and land tenure arrangements, demographic pressure and urban sprawl. This arbitrary land use policies are ambivalent to environmental sustainability and directly lead to mismanagement or passive management of the forested areas and lack of interest or awareness by the general public. Finally, the increasing density of the transportation network and high frequency of forest wildfires inevitably tend to further degrade the remaining natural reserves.

Given the importance of Greek biodiversity and the governmental failure to confront this neglected conservation policy, strong strategies should be designed, and decisions should be made. Effective management in the Natura 2000 sites in Greece has been lagging due to lack of science-based data that leads to incorrect interpretation of what the desirable state of an ecosystem is (Apostolopoulou and Pantis, 2009). Sustainable forest management, in any region or country has to rely on science-based information about the condition of the remaining forests in regard to their ownership, composition and structure. This type of information can act as a baseline to help establish efficient policy, by promoting the adaptation of sustainable forest management, public awareness and science-based decision making. Innovative use of technologies, such as satellite remote sensing, GIS analysis and landscape metrics, can facilitate in the spatio-temporal considerations of landscape patterns and processes. Medium resolution Landsat imagery provide a cost-effective primary data source, to facilitate the monitoring of such a phenomenon (Loveland and Dwyer, 2012), compared to high resolution satellite data that are often prohibitive because of their high cost. Additionally, the cost effectiveness and therefore, the wide applicability can be enhanced with the utilization of open source GIS software (Steiniger and Hunter, 2013). A potential limitation that can be encountered, if the presented methodology is adopted, is the availability of free satellite data, especially for past decades and their relative

efficient reference data. Nevertheless, additional criteria for the suitability of data selection (as discussed in 2.2) are also mandatory, a fact that complicates even more the situation.

2.4 Conclusions

The Hymettus Mountain is an area of great conservation significance, since it is a biodiversity hotspot near Athens. The PC results suggest that the fragmentation of woody vegetation of Hymettus Mountain had been increasing during the study period. The majority of the LULCCs in expense of forests extent occurred in areas where the forested areas bordering with artificial land uses. A general conclusion drown upon the transition maps is that the "woody vegetation to sva/bare" along with the "woody vegetation to artificial land use" patches, increased during the second and third interval in comparison with the first.

The LM results come also to support these conclusions for our study area. The woody vegetation class successively tends to be more isolated and less compact, whereas the artificial land uses are highly aggregated with slight fLULCtuations during the study period. In particular, woody vegetation areas tend to be subdivided after 2003, a crucial period for the area, due to pressures that resulted from increased development during this period.

Changes in the spatial configuration of the boundaries indicated that the woody vegetation patches are progressively becoming more geometrically complicated. The examinations of the inter-patch connectivity with respect to the size of the patches have shown that core and larger areas of woody vegetation class remain relatively constant. These together with the findings of the shape complexity show that the majority of the transitions are made at the boundaries of the woody vegetation areas.

The methodology presented here uses these technologies to determine the variability of forest extent through time in a nationally and internationally protected area via earth observation accurately and cost-effectively. It can also provide valuable information about the effectiveness of past activities. Such a cost-effective method could provide science-based information to the involved parties enhancing the development of effective and realistic management plans for Natura 2000 sites and thus promoting the collaboration among various managing organizations.

References

Apostolopoulou, E., & Pantis, D.J. (2009). Conceptual gaps in the national strategy for the implementation of the European Natura 2000 conservation policy in Greece. *Biol. Conserv*, 142, 221–237.

Apostolopoulou, E., Drakou, E.G., & Pediaditi, K. (2012). Participation in the management of Greek Natura 2000 sites: evidence from a cross-level analysis. *J. Environ. Manage*, 113, 308–318.

Becker, A., & Bugmann, H. (1999). Global Change and Mountain Regions. The Mountain Research Initiative. IGBP Report 49.

Breiman, L. (2001). Random Forests. *Mach. Learn*. 40, 5–32.

Bruner, A.G., Gullison, R.E., Rice, R.E., & da Fonseca, G.A.B. (2001). Effectiveness of Parks in Protecting Tropical Biodiversity. *Science*, 297, 125–128.

Chander, G, Markham, B.L, & Helder, D.L (2009). Summary of Current Radiometric Calibration Coeficients for Landsat MSS, TM, ETM+, and EO-1 ALI Sensors. *Remote Sens. Environ*, 113, 893–903.

Chase, T.N., Pielke, R.A., Kittel, T.G.F., Nemani, R.R., & Running, S.W. (1999). Simulated impacts of historical land cover changes on global climate in northern winter. *Climate Dynam*, 16, 93–105.

Chen, X. (2002). Using remote sensing and GIS to analyse land cover change and its impacts on regional sustainable development. *Int. J. Remote Sens*, 23, 107–124.

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

Chrysoulakis, N., Mitraka, Z., Stathopoulou, M., & Cartalis, C. (2013). A comparative analysis of the urban web of the greater Athens agglomeration for the last 20 years period on the basis of Landsat imagery. *Fresen. Environ. Bull*, 22, 2139-2144.

Civco, D.L. (1989). Topographic normalization of Landsat thematic mapper digital imagery. *Photogram. Eng. Remote Sensing*, 55, 1303–1309.

Coppin, P., Jonckheere, I., Nackaerts, K., & Muys, B. (2004). Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sensing*, 25, 1565 – 1596.

Costanza, R., dArge, R., deGroot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Suttonkk, P., & van den Belt,

M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387, 253–260.

Cushman, S.A., McGarigal, K., & Neel, M.C. (2008). Parsimony in landscape metrics: Strength, universality, and consistency. *Ecol. Indic*, 8, 691 -703.

Dimitrakopoulos, P.G., Memtsas, D., & Troumbis, A.Y. (2004). Questioning the effectiveness of the Natura 2000 special areas of conservation strategy: the case of Crete. *Global Ecol. Biogeogr*, 13, 199–207.

Dimopoulos, P., Bergmeier, E., & Fischer, P. (2006). Natura 2000 habitat types of Greece evaluated in the light of distribution, threat and responsibility. *Biol. Environ*, 106(3), 175-187.

Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., Moore, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., & Steffen W. (2000). Integrated understanding of the global carbon cycle- A test of our knowledge. *Science*, 290, 291–296.

Forman, R. (1995). Some general principles of landscape and regional ecology. *Landscape Ecol*, 10 (3), 133-142.

Forman, R.T.T., & Alexander, L.E (1998). Roads and their major ecological effects. *Annu. Rev. Ecol. Evol. Syst*, 29, 207–231.

Gaston, J.K. (2005). Biodiversity and extinction: Species and people. *Prog. Phys. Geo*, 29 (2), 239–247.

Georghiou, K., Arianoutsou, M., Kazanis, D., Economidou, E., Legakis, A., & Valakos, E. (1995). Natura 2000 Clean Standard Data Form - NATURA 2000 - GR3000006 (Ymittos – Aisthitiko Dasos Kaisarianis – Limni Vouliagmenis). University of Athens, The Ministry of the Environment, Physical Planning, and Public Works, and The Greek Biotope/Wetlands Center.

Gibson, D.J., Collins, S.L., & Good, R.E. (1988). Ecosystem fragmentation of oak-pine forest in New Jersey Pinelands. *Forest Ecol. Manag*, 25, 105–122.

Gustafson, E.J. (1998). Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems*, 1, 143-156.

Hall, F. G., Strebel, D. E., Nickeson, J. E., & Goetz, S.J. (1991). Radiometric rectification: toward a common radiometric response among multidate, multisensory images. *Remote Sens. Environ*, 35, 11–27.

Herold, M., Couclelis, H., & Clarke, K.C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change, *Comput. Environ. Urban*, 29(4), 369-399.

Ji, W., Ma, J., Twibell, R.W., & Underhill, K. (2006). Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. *Comput. Environ. Urban*, 30, 861–879.

Lambin, E.F., & Meyfroidt, P. (2010). Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use Policy*, 27, 108–118.

Le Houérou, H.N. (2002). Man-made deserts: Desertization processes and threats. *Arid Land Res. Manag*, 16, 1–36.

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

Liu, H., & Zhou, Q. (2005). Developing urban growth predictions from spatial indicators based on multi-temporal images, *Comput. Environ. Urban*, 29(5), 580-594.

Loveland, R.T., & Dwyer, L.J. (2012). Landsat: Building a strong future. *Remote Sens. Environ,* 122, 22 – 79.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2002). Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *Int. J. Remote Sensing*, 23(13), 2651-2671.

Mallinis, G., Emmanoloudis, D., Giannakopoulos, V., Maris, F., & Koutsias, N. (2011). Mapping and interpreting historical land cover/land use changes in a Natura 2000 site using earth observational data: The case of Nestos delta, Greece. *Appl. Geogr*, 31, 312–320.

Mas, J.F. (1999). Monitoring land-cover changes: a comparison of change detection techniques. *Int. J. Remote Sensing*, 20, 139-152.

McGarigal, K., & Marks, B.J. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Meyer, P., Itten, K.I., Kellenberger, T., Sandmeier, S., & Sandmeier, R. (1993). Radiometric corrections of topographically induced effects on Landsat TM data in alpine environment. *ISPRS J. Photogram*, 48, 17–28.

Ministry of the Environment, Planning and Public Works (2006). 107017/06 (ΦΕΚ 1225 B/5-9-2006 - Environmental assessment of certain plans and programs. (in Greek).

Neteler, M., & Mitasova, H. (2008). Open source GIS: A GRASS GIS approach (3rd ed.). Berlin: Springer.

Nikolakopoulos, K., Pavlopoulos, K., Chalkias, C., & Manou, D. (2005). Monitoring the urban expansion of Athens, using remote sensing and GIS techniques in the last 35 years. *Proc. SPIE-Int. Soc. Opt. Eng*, 5983 (598309).

Papageorgiou, K., & Vogiatzakis, I.N. (2006). Nature Protection in Greece: An appraisal of the factors shaping integrative conservation and policy effectiveness. *Environ. Sci. Policy*, 9, 476–486.

Papanastasis, P.V., Mantzanas, K., Dini-Papanastasi O., & Ispikoudis, I. (2008). Traditional Agroforestry Systems and Their Evolution in Greece. In A. Rigueiro-Rodróguez, J. McAdam, & M.R. Mosquera-Losada (Eds.), Agroforestry in Europe Current Status and Future Prospects (pp. 89-109). Springer.

Quantum GIS Development Team (2013). Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reger, B., Otte, A., & Waldhardt, R. (2007). Identifying patterns of land-cover change and their physical attributes in a marginal European landscape. *Landscape Urban Plann*, 81(1-2), 104-113.

Riitters, K.H., O'Neil, R.V., Hunsaker, C.T., Wickham, J.D., Yankee, D.H., Timmins, S.P., Jones, K.B., & Jackson, B.L. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecol*, 10, 23-39.

Rockstrom, J., Steffen, W., Noone, K., Persson, E., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sφrlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., & Foley, J.A. (2009). A safe operating space for humanity. *Nature*, 461, 472–275.

Seto, K.C., & Fragias, M. (2007). Mangrove conversion and aquaculture development in Vietnam: A remote sensing-based approach for evaluating the Ramsar Convention on Wetlands. *Global Environ. Chang*, 17, 486–500.

Shi, Z.H., Li, L., Yin, W., Ai, L., Fang, N.F., & Song, Y.T. (2011). Use of multi-temporal Landsat images for analyzing forest transition in relation to socioeconomic factors and the environment. *Int. J. Appl. Earth Obs*, 13, 468–476.

Schroter, D., Cramer, W., Leemans, R., Prentice, C., Araujo, M.B., Arnell, N.W., et al. (2005). Ecosystem Service Supply and Vulnerability to Global Change in Europe. *Science*, 310, 1333–1337.

Steiniger, S., & Hunter, J.S.A. (2013). The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption. *Comput. Environ. Urban*, 39, 136–150.

Symeonakis, E., Calvo-Cases, A., & Arnau-Rosalen, E. (2007). Land use change and land degradation in southeastern Mediterranean Spain. *Environ. Manage*, 40, 80–94.

Symeonakis, E., Caccetta, P., Koukoulas, S., Furby, S., & Karathanasis, N. (2012). Multi-temporal land cover classification and change analysis with Conditional Probability Networks: the case of Lesvos Island (Greece). *Int. J. Remote Sensing*, 33(13), 4075-4093.

Teillet, P.M., Guindon, B., & Goodenough, D.G. (1982). On the slope-aspect correction of multispectral scanner data. *Can. J. Remote Sens*, 8 (2), 84–106.

Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R. et al. (2001). Forecasting Agriculturally Driven Global Environmental Change. *Science*, 292, 282–284.

Tsiafouli, A.M., Apostolopoulou, E., Mazaris, D.A., Kallimanis, S.A., Drakou, G.E., & Pantis, D.J. (2013). Human Activities in Natura 2000 Sites: A Highly Diversified Conservation Network. *Environ. Manage*, 51(5), 1025-1033.

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ*, 8, 127–150.

Vermote, E., Tanré, D., Deuzé, J.L., Herman, M., & Morcrette, J.J. (1997a) Second Simulation of the Satellite Signal in the Solar Spectrum (6S), 6S User Guide Version 2.

Vermote, E Tanre, D., Deuze, J.L., Herman, M., & Morcrette, J.J., (1997b), Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Trans. Geosci. Remote Sens*, 35(3), 675–686.

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

Vlachogiannis, D., Sfetsos, A., Gounaris, N., & Papadopoulos, A. (2012). Computational study of the effects of induced land use changes on meteorological patterns during hot weather events in an urban environment. *Int. J. Environ. Pollut,* 50 (1-4), 460-468.

Weber, C., Puissant, A. (2003). Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. *Remote Sens. Envir,* 86, 341–352.

Weber, C., Petropoulou, C., & Hirsch, J. (2005). Urban development in the Athens metropolitan area using remote sensing data with supervised analysis and GIS. *Int. J. Remote Sensing*, 26 (4), 785–796.

Wolter, P.T., & White, M.A. (2002). Recent forest cover type transitions and landscape structural changes in northeast Minnesota, USA. *Landscape Ecol*, 17(2), 133–155.

Chapter 5: Land cover mapping of Greece: A semi-automated classification approach based on Landsat data and Random Forests

Authors: Dimitrios Gounaridis, Anastasios Apostolou, Sotirios Koukoulas
Journal: Journal of Maps 12(5), (2016), 1055-1062.

Abstract

Information about land cover (LC) and land use is fundamental in various areas of research regarding the Earth's surface. However, field campaigns are costly and time consuming while existing data sets have strong limitations. Classification of LC by remote sensing, although considered a technically and methodologically challenging task, can facilitate mapping initiatives at various scales. This study suggests an efficient and robust methodology of LC classification with minimal user requirements. The study site is Greece which faces a lack of up to date LC maps at national scale. In this context we employed Landsat imagery, open source software and the random forest classification algorithm to produce a high resolution national LC map for 2010. The algorithm was trained semi-automatically, extracting information from available data sets. The results are promising, achieving an overall accuracy of 83%. The methodology presented minimizes many obstacles that lead to data deficiencies and can act as a baseline for future LC mapping initiatives.

Keywords: Land cover mapping; Greece; Landsat; random forests; semi-automated classification; open source software

3.1 Introduction

A vast portion of the Earth's land cover (LC) is directly influenced and shaped by human activities through its land use (LU) (Turner, 1994). Hence, long term observation of LC, at various scales, is essential to understanding Earth surface processes and the anthropogenic influence on human and natural systems (Turner, Lambin, and Reenberg, 2007). Research concerning LC has intensified, especially after discovering the impacts on climate and the environment (Foley et al., 2005). To this end, accurate and up to date LU and LC spatial information serves as a principal component in a variety of research, management and planning activities, while for studies related to global, environmental and/or climate change it is considered a pre-requisite.

Traditional field campaigns are limited by their local extent and high resource demand (personnel, cost, time). Recent advances in Remote Sensing (RS) and Geographic Information Systems (GIS) can successfully provide spatially consistent multi-spectral and multi-temporal LC information. However, very high resolution (VHR) imagery is costly particularly for small scale mapping. In this respect, a number of recent studies conducted virtual "field visits" utilizing VHR imagery provided by Google Earth or similar engines either for data training or to validate outputs (e.g. Knorn et al., 2009; Schneider, 2012) with good results.

Currently, various global LC datasets exist, with different spatial resolutions and each with specific limitations (Giri et al. 2014; Herold et al. 2008). Regarding Europe, two significant databases exist, the Coordinate Information on the Environment (CORINE), a pan-European LC map for the years 1990, 2000 and 2006 provided by the European Environmental Agency (EEA) and the Pan-European Land Cover Monitoring (PELCOM) database (Mucher et al., 2001). The CORINE database is available at 100 m spatial resolution with a minimum mapping unit (MMU) of 25 ha (Bossard et al. 2000), while the PELCOM database was constructed at 1 km spatial resolution. These two frequently used databases suffer from limitations, to resolution and MMU, inconsistency from one country to another, lack of rigorous accuracy assessments and reliability (Neumann et al. 2007). One important limitation, of most available datasets, still unaddressed, is the discrepancy between LC classes, their overarching definitions, their nomenclature and thus the heterogeneity of information

provided. For instance, the CORINE database, although it is established and recognized as an LC database, strictly represents a mixture of LU and LC classes (Comber et al. 2005). Another serious issue of the CORINE LC database, emanating from the fact that it is produced by each country separately, is the out dated information available for some countries. For instance, the latest update of LC information for Greece was for the year 2000.

To meet the need for consistent and accurate LC maps, increasingly sophisticated approaches have been introduced that are less source data demanding and less labour intensive (Symeonakis et al. 2012), showing a trend in developing automated (e.g. Chen et al. 2012; Radoux et al., 2014) or semi-automated (e.g. Jiang et al., 2012) LC classification methodologies. An effective way to minimize user intervention is the utilization of existing LC data, to train the classifier (Chen et al., 2012; Jiang et al., 2012; Radoux et al., 2014). Additionally, cost effective methodologies utilizing open source software (OSS) (Steiniger and Hunter, 2013) and freely available satellite data (Wulder et al. 2012) are increasingly being adopted by the research community.

This study aims to cover the observed gap in LC data for Greece at a national scale, demonstrating a robust and cost-effective methodology. In this context we utilized the NASA–USGS Global Land Survey (GLS) Landsat data for 2010 with a nominal pixel size of 30 m (Gutman et al. 2013). Classification is performed implementing the Random Forests (RF) machine learning algorithm (Breiman, 2001) that has proven to perform well with heterogeneous classes (Rodriguez-Galiano et al. 2012). To train the algorithm we used the existing CORINE 2000 dataset of Greece. Visual inspection of the training samples was conducted utilizing VHR imagery, from the Google and Bing engines, which are accessible via the OpenLayers plugin for QGIS (Quantum GIS Development Team, 2013)

3.2 Materials and Methods

3.2.1 Study Area

Greece is a Mediterranean country, located in the Balkan peninsula (Figure 1), covering an estimated total area of 131,957 km^2, with its coastline stretching for 15,021 km while 20% of that territory is distributed on its approximately - both inhabited and uninhabited - 3000 islands (Minetos and Polyzos, 2010). According to the 2011 census, the resident population is approximately 11 million (Hellenic Statistical Authority, 2013). Dominant vegetation types are broadleaved and coniferous forests, sclerophyllous maquis and garrigue (Arianoutsou et al., 1997).

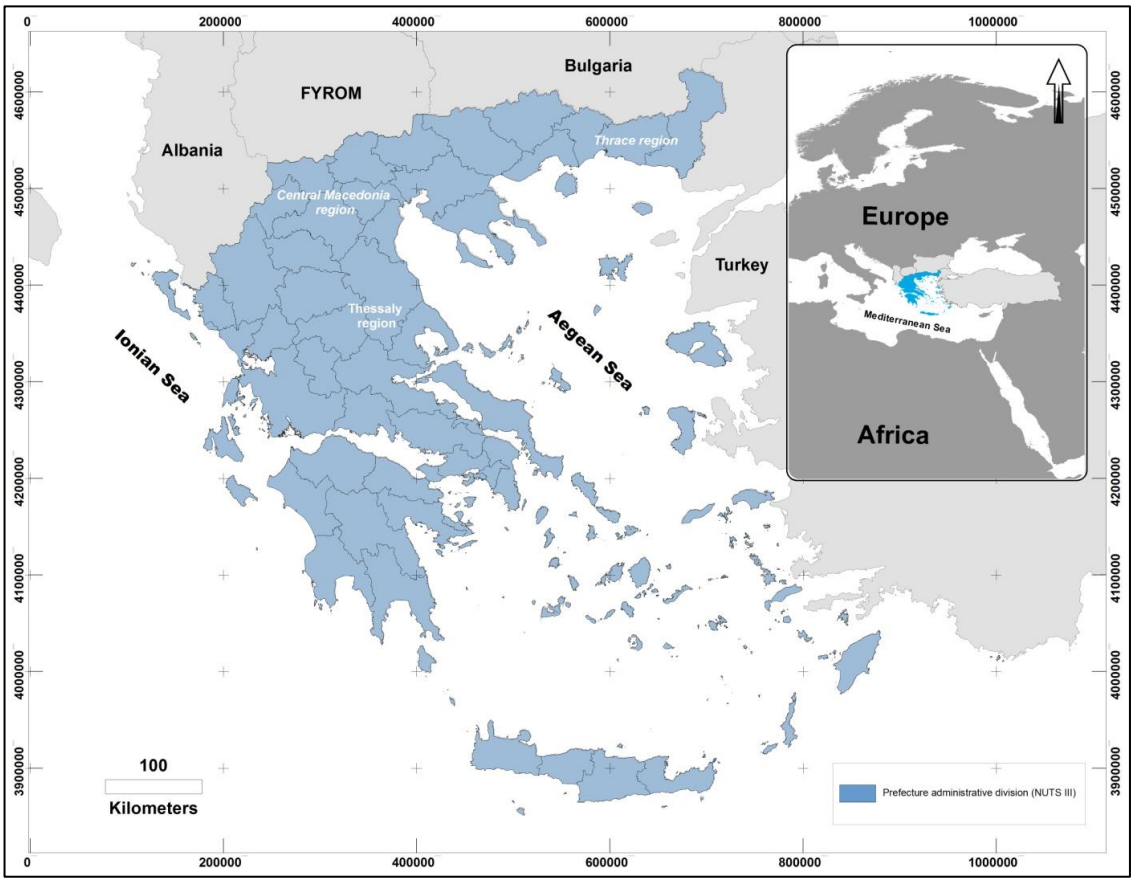


Figure 1. Location of the study site, Greece

The climate is typical Mediterranean, with hot and dry summers and relatively mild and wet winters. Geomorphologically, two thirds of the Greek territory are mountainous or hilly and more than 40% of the land exceeds 500 m in altitude (EEA,

2010) while extensive agricultural plains are primarily located in Thessaly, Central Macedonia and Thrace regions. Socioeconomic development in Greece has been depended largely on land-based economic activities, including agriculture, farming, forest cultivation for timber and mining. These are closely related to the structure and function of the land's surface constituting a source of income for its residents (Demoussis, 2003; Papanastasis et al. 2009).

3.2.2 Landsat data

We employed the GLS 2010 Landsat imagery dataset that uses a combination of Landsat 5 TM and Landsat 7 ETM+ images acquired between 2008 and 2011. The GLS datasets are created using the Landsat sensor operating at the time, meeting quality and cloud cover standards and undergoing a level of pre-processing (Gutman et al., 2013). A total of 27 images (spanning path: 180–186 and row: 31–36) were selected, achieving full coverage of the country. The successive steps of the procedure in order to achieve the goals of this research are summarized in Fig. 2.

3.2.2.1 Data pre-processing

Initially, all images were re-projected to the Greek Geodetic Reference System (GGRS 1987). Data analysis was conducted at the prefecture level using the "Kapodistrias" administrative division system (NUTS III), which divides Greece into 51 prefectures. To avoid any discrepancies due to the multi-temporal and double-sensor analysis, all images underwent a series of corrections and calibration (Gounaridis et al. 2014). To radiometrically and atmospherically correct the Landsat images the initial DN values were converted to top of atmosphere reflectance (TOA). To obtain surface reflectance and achieve data normalization, we used the 6S model (Vermote et al. 1997). Finally, to correct any topographic effects we applied the C-correction method (Teillet et al. 1982). The 27 corrected images were clipped into 51 parts according to every prefecture boundary (Figure 1). In cases where a prefecture was covered by more than one image, the clipped pieces were mosaiced into one.

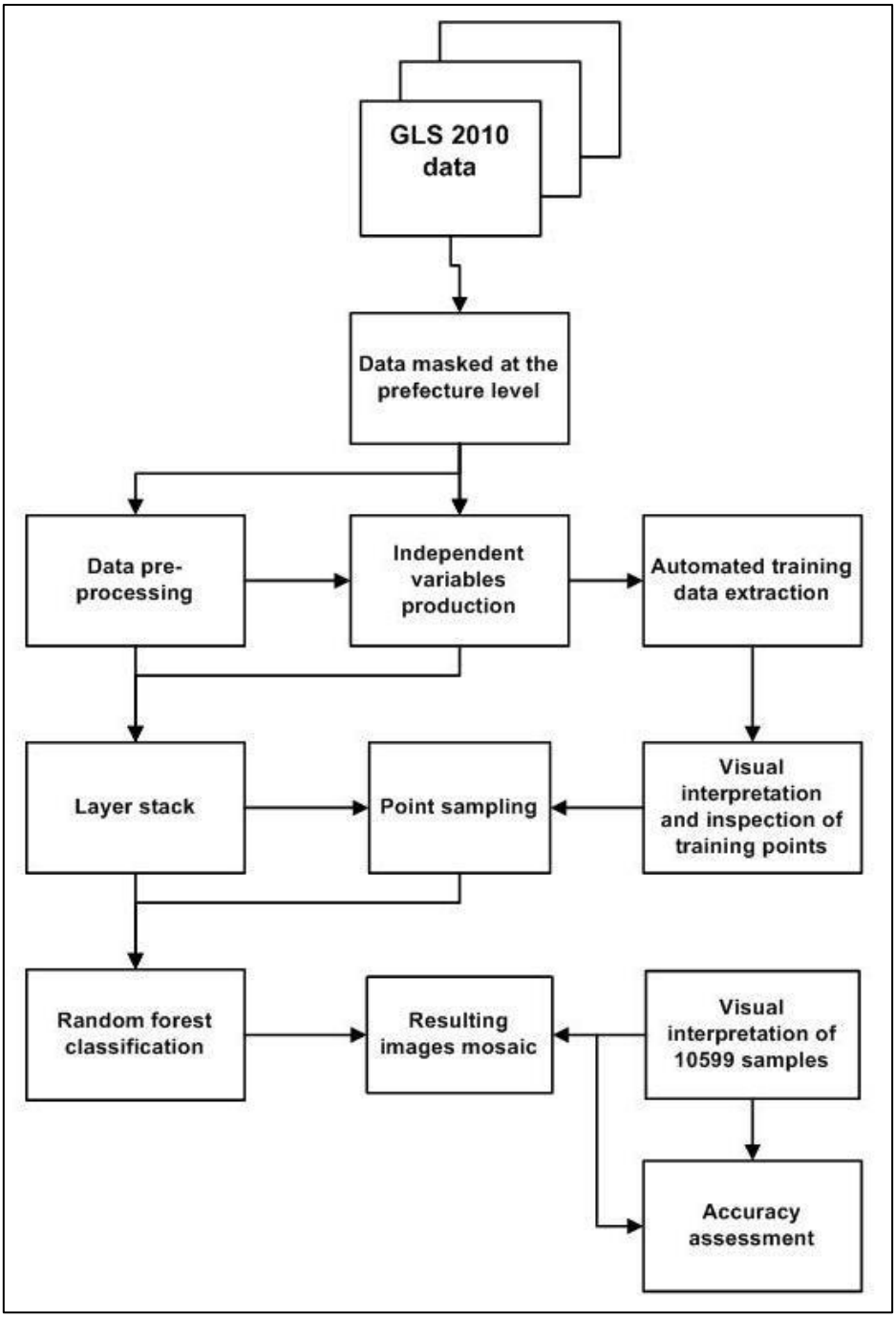


Figure 2. Flowchart of the successive steps followed in the methodology

3.2.2.2 Image classification with Random forests

We implemented the Random Forest (RF) classification algorithm. The algorithm starts with a random selection of the predictor variables resulting in a collection of, independent to each other and identically distributed, tree structured

classifiers. Each individual tree casts a unit vote for the most popular class while the outputs of classification are determined from the majority of votes for each class (Breiman, 2001). The nomenclature adopted for this study, is in line with the CORINE's classification scheme at Level 2 (Table 1). As mentioned in the introduction, the CORINE classes do not refer literally to pure LC but rather to a mixture of LC and LU and thus this applies to our approach as well (referred to as the LC map hereafter).

Table 1. Classification scheme and the portion occupied by each class. For complete description of the classes the reader is referred to Bossard et al. (2000).

	Class Name	% of coverage
1	Urban fabric	1.53%
2	Industrial, commercial and transport units	0.22%
3	Mine, dump and construction sites	0.09%
4	Arable land	18.15%
5	Permanent crops	14.09%
6	Pastures	0.10%
7	Heterogeneous agricultural areas	5.48%
8	Forests	17.03%
9	Scrub and/or herbaceous vegetation associations	41.26%
10	Open spaces with little or no vegetation	1.18%
11	Inland wetlands	0.15%
12	Inland waters	0.72%

3.2.2.3 Sample design for training

Training samples are a crucial part to any classification process and at the same time the main source of errors (Foody and Arora, 1997). We used the CORINE 2000 dataset, which is a universally adopted nomenclature and the only available dataset at national scale for Greece. To train the classifier, 500 - 1000 randomly distributed points (depending on the complexity of the landscape) were applied to each of the 51 scenes, achieving a balance in the representation of all classes. A 50m buffer was drawn around all points and these areas were assigned class values from the CORINE 2000 dataset. Subsequently, we verified the class membership visually by inspecting the points over VHR images of the area (via the OpenLayers plugin in QGIS). In several cases we had

to relocate, delete or re-assign values for problematic points. All points close to the boundaries of the classes were either relocated or eliminated so as to provide a clear sample. This was an important step in order to avoid misclassified training samples due to differences in spatial resolution and MMU, between CORINE and Landsat data. Special attention was given to areas whose classification results were not satisfying, enhancing the training data with extra points originating from visual interpretation of the VHR imagery.

3.2.3 Predictor variables

The RF algorithm can handle multiple types of auxiliary data, biophysical and spectral variables, both continuous and categorical, to improve classification performance and discrimination between LC classes. In addition to the 6 reflective Landsat bands (bands 1-5 & 7), we used Principal Components Analysis (PCA) separately for the three visible bands (1, 2 & 3) and the middle infrared bands (5 & 7). The first of the principal components produced for each run was used (Gounaridis et al., 2014; Symeonakis et al. 2007). We also included Landsat's thermal band (band 6) since different LC classes have specific temperature ranges (Southworth, 2004). We used Tasseled Cap (TC) transformations Soil Brightness Index (SBI) and Green Vegetation Index (GVI) since they are capable of discerning vegetation from bare features (Kolios and Stylios, 2013). In order to further enhance the representation of vegetation we extracted the normalized difference vegetation index (NDVI). Additionally we tested several band ratios that provide unique information and subtle spectral reflectance differences between surface materials (Southworth, 2004; Weiser et al. 1986); we used five of them in our model (band 4/band 3; band 7/band 2; band 3/band 2; (band 2-band 6)/(band 2+band 6); TC.GVI/band 6). We also derived auxiliary variables (elevation and slope) from the Global Land Survey Digital Elevation Model (GLSDEM) as they helped in discriminating between bare and sparsely vegetated areas. Assuming that human activity and the artificial environment are connected to the existence of a road network we derived a distance-from-the road-network layer based on the Euclidean distance function (Gounaridis et al., 2014). Finally, all predictor variables were merged into a single layer stack containing 20 layers in total.

3.2.4 Random forests model implementation

Training and predictor variable data were collated in a database. The values of the 20 merged layers were sampled on the location of every training point, already containing LC class values and manipulated as predictor values. The next step included the classification process, implemented through the use of the RandomForest package in R (Liaw and Wiener, 2002). RF requires two primary parameters being (i) the number of predictor variables randomly sampled at each decision tree split and (ii) the number of classification trees. We used five (5) predictor variables for each tree split (equal to the square root of the total number of predictor variables) and 1000-2000 trees for each run.

3.2.5 Classification post processing

All classification outputs, originally at the prefecture level were mosaiced into a single LC layer. In order to further improve the final product, we removed isolated patches (smaller than 0.1 ha) which are often unreliably classified, usually due to a mix up of spectral values. We replaced the values of isolated pixels with the mode of their neighborhood pixels, defined by a 3x3 window centered on the pixel to be eliminated (Gounaridis et al., 2014). Lastly, we calculated the percentage of coverage occupied by each of the 12 classes and incorporated our findings in Table 1.

3.2.6 Accuracy Assessment

To evaluate the results, we assessed the accuracy using a randomly distributed sample of 10599 -independent to the original training- points employing VHR imagery (via the OpenLayers plugin in QGIS) in order to visually interpret and label them according to the adopted nomenclature. Accuracy was assessed by generating a confusion matrix that compared the visually interpreted samples with the final map. Additionally, we computed both the overall accuracy of the map and the errors of commission and omission for each category.

3.3 Results and Discussion

3.3.1 RF performance

Results obtained from this study demonstrate that RF discriminated 12 classes relatively well (Figure 3). The performance of the RF classifier was influenced by the computational demands and the depiction of a landscape's composition heterogeneity at medium scales (e.g. as for Landsat's). Performing all the analysis on a single PC, lead us to divide the study area into smaller parts due to the computational demands of the algorithm. This in turn, increased the workload as we needed to calibrate the model separately for each part of the study area. With regard to landscape composition heterogeneity, even though Landsat's 30 m spatial resolution is acceptable for LC characterization, when working with more semantically complex and with less pronounced differences in spectral reflectance classes, a higher spatial resolution is likely required. Nevertheless, our study shows that Landsat's resolution achieved satisfactory results for the production of a medium scale classification of 12 classes for Greece.

RF performance, as expected for all machine learning family algorithms, significantly depends on the predictor variables and training data quality. Extracting information from a known dataset instead of independently classifying the remotely acquired images is good, especially when dealing with a large extent, with respect to workload and required sources. Under the assumption that under normal circumstances changes only occur to a small proportion of land and especially at the edges, utilizing unchanged areas as a training source is reasonable (Chen et al., 2012; Jiang et al., 2012; Radoux et al., 2014; Xian et al., 2009). However, susceptibility to errors exists due to the complexity of CORINE class definitions, with some of them unsuitable for algorithmic classification and certainly due to incompatibility concerning different scales between the image to be classified and the reference data (in our case CORINE data with a MMU of 25 ha were used to train Landsat data with spatial resolution of 30 m).

3.3.2 Accuracy assessment

Results were satisfactory, with an overall accuracy (percent correctly classified) of 83% (table 1). Of the 1773 (out of 10599) validation points assigned as misclassified,

940 (53%) could be considered minor misclassifications (confusion between certain classes due to visual or spectral similarities). Most "disagreements" were discovered in "Heterogeneous agricultural areas", "Permanent crops", "Scrub and/or herbaceous vegetation associations" and "Forests". These classes are spectrally heterogeneous due to different cropping systems and cycles in the region, resulting in confusion between the spectral signatures (or sometimes an interpreter's mislabeling). If classes were aggregated (e.g. to adopt the level 1 scheme of CORINE) the overall accuracy would significantly increase. Using the current validation data, an accuracy assessment at level 1 yields an approximate overall accuracy of 91% (table 2).

Table 2. Error matrix. The reference corresponds to validation points visually interpreted utilizing VHR imagery

		Reference												
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**	**Commission error**
Classified	**1**	**256**	28	1	8	1		6		4	1	1	1	307	16.6%
	2	6	**46**	8	1								2	63	27.0%
	3		2	**43**			1	2		1			1	50	10.0%
	4	22	31	30	**1711**	31	4	34	4	39	7	47	27	1987	13.9%
	5	1	2		20	**381**	1	26	1	22		1		455	16.3%
	6				3		**15**	1		2				21	14.3%
	7	45	19	8	147	183	4	**707**	18	134		5	4	1274	44.5%
	8			1	5	3		16	**1692**	103		4	8	1832	7.6%
	9	27	8	27	23	83	4	148	165	**3567**	8	3	6	4069	12.3%
	10	3	2	11	2					49	**96**	1	2	166	38.6%
	11			1	4							**50**	6	61	20.0%
	12											52	**262**	314	16.6%
	Total	360	138	130	1924	682	29	940	1880	3921	112	164	319	**10599**	
	Omission error	28.1%	66.7%	66.9%	11.1%	44.1%	48.3%	24.8%	10.0%	9.0%	14.3%	68.9%	17.6%		
		Overal Acuuracy : 83%													

Note: The reference corresponds to validation points visually interpreted utilizing VHR imagery.

Classes

Level 1: Artificial surfaces

1: Urban fabric

2: Industrial, commercial and transport units

3: Mine, dump and construction sites

Level 1: Agricultural areas

4: Arable land

5: Permanent crops

6: Pastures

7: Heterogeneous agricultural areas

Level 1: Forests and semi natural areas

8: Forests

9: Scrub and/or herbaceous vegetation associations

10: Open spaces with little or no vegetation

Level 1: Wetlands

11: Inland wetlands

Level 1: Water bodies

12: Inland waters

3.4 Conclusions

This study demonstrated an efficient and cost-effective approach to produce an LC map, in order to fill the observed gap in up to date LU and LC data at the national scale for Greece (Figure 3). The high cost of VHR imagery acquisition, the need for time consuming and labor-intensive field surveys, inconsistency and incomparability of the existing databases and the need for exhaustive and advanced methodologies in order to train classification algorithms, prevents many researchers from producing LC and LU maps. The presented sequences of methodologies minimize many of these obstacles that in turn have led to data deficiencies especially for large extents. The approach presented can act as a baseline to continuously monitor LC and to assist ongoing and upcoming LC mapping initiatives.

References

Arianoutsou, M., Delipetrou, P., Dimopoulos, P., Economidou, E., Karagiannakidou, V., Kostantinidis, P. et al. (1997). Habitat Types Present in Greece. In Dafis, et al. (Ed.), The Greek "Habitat" Project NATURA 2000: An overview (pp. 402-434). The Goulandris Natural History Museum – Greek Biotope/Wetland Centre.

Bossard, M., Feranec, J., & Otahel, J. (2000). CORINE land cover technical guide – Addendum 2000. Technical report, 40. European Environment Agency. Copenhagen.

Breiman, L. (2001). Random forests. *Mach Learn, 40*, 5–32.

Chen, X., Chen, J., Shi, Y., & Yamaguchi, Y. (2012). An automated approach for updating land cover maps based on integrated change detection and classification methods. *ISPRS J Photogramm, 71*, 86–95.

Comber, A.J., Fisher, P.F. & Wadsworth, R.A. (2005). What is land cover?. *Environ Plan B 32*, 199-209.

Demoussis, M. (2003) Transformations of the CAP and the Need for Reorganizing Agricultural Policy in Greece. In C. Kasimis and G. Stathakis (Eds.), The Reform of the CAP and Rural Development in Southern Europe (pp.173–186). Aldershot, Ashgate.

EEA, 2010. The European environment – state and outlook 2010: synthesis. European Environment Agency, Copenhagen.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, … Snyder, P.K. (2005). Global consequences of land use. *Science, 309*, 570–574.

Foody, G., & Arora, M. (1997). An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *Int J Remote Sens, 18*, 799–810.

Giri, C., Pengra, B., Long, J., & Loveland, T.R. (2013). Next generation of global land cover characterization, mapping, and monitoring. *Int J Appl Earth Obj, 25*, 30–37.

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

Gutman, G., Huang, C., Chander, G., Noojipady, P., & Masek, J.G. (2013). Assessment of the NASA–USGS Global Land Survey (GLS) datasets. *Remote Sens Environ, 134*, 249–265.

Hellenic Statistical Authority (2013). Announcement of the demographic and social characteristics of the Resident Population of Greece according to the 2011 Population - Housing Census.

Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens Environ, 112*, 2538–2556.

Jiang, D., Huang, Y., Zhuang, D., Zhu, Y., Xu, X., & Ren, H. (2012). A Simple Semi-Automatic Approach for Land Cover Classification from Multispectral Remote Sensing Imagery. *PLoS ONE, 7*, e45889.

Knorn, J., Rabe, A., Radeloff, V.C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sens Environ, 113*, 957–964.

Knudby, A., Mtwana Nordlund, L., Palmqvist, G., Wikstrom, K., Koliji, A., Lindborg, R., & Gullstrom, M. (2014). Using multiple Landsat scenes in an ensemble classifier reduces classification error in a stable nearshore environment. *Int J Appl Earth Obj, 28*, 90–101.

Kolios, S. & Stylios, C.D. (2013). Identification of land cover/land use changes in the greater area of the Preveza peninsula in Greece using Landsat satellite data. *Appl Geogr, 40*, 150–160.

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

Minetos, D., & Polyzos, S. (2010). Deforestation processes in Greece: A spatial analysis by using an ordinal regression model. *Forest Policy Econ, 12*, 457–472.

Mucher, C.A., Champeaux, J.L., Steinnocher, K.T., Griguolo, S., Wester, K., Heunks, C. et al. (2001). Development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental monitoring; the PELCOM report. Alterra-rapport 18 178/CGI-report 6, Alterra, Wageningen, the Netherlands.

Neumann, K., Herold, M., Hartley, A., & Schmullius, C. (2007). Comparative assessment of CORINE2000 and GLC2000: Spatial analysis of land cover data for Europe. *Int J Appl Earth Obj, 9*, 425–437.

Papanastasis, P.V., Mantzanas, K., Dini-Papanastasi O., & Ispikoudis, I. (2009). Traditional Agroforestry Systems and Their Evolution in Greece. In A. Rigueiro-Rodriguez, J. McAdam, & M.R. Mosquera-Losada (Eds.), Agroforestry in Europe Current Status and Future Prospects (pp. 89-109). Springer.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., & Defourny, P. (2014). Automated Training Sample Extraction for Global Land Cover Mapping. *Remote Sens, 6*, 3965-3987.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm, 67*, 93–104.

Schneider, A. (2012). Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sens Environ, 124*, 689–704.

Southworth, J. (2004). An assessment of Landsat TM band 6 thermal data for analyzing land cover in tropical dry forest regions. *Int. J. Remote Sens, 25*, 689–706.

Steiniger, S., & Hunter, J.S.A. (2013). The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption. *Comput Environ Urban, 39*, 136–150.

Symeonakis, E., Calvo-Cases, A., & Arnau-Rosalen, E. (2007). Land use change and land degradation in southeastern Mediterranean Spain. *Environ Manage, 40*, 80–94.

Symeonakis, E., Caccetta, P., Koukoulas, S., Furby, S. & Karathanasis, N. (2012). Multi-temporal land cover classification and change analysis with Conditional Probability Networks: the case of Lesvos Island (Greece). *Int. J. Remote Sens 33(13)*, 4075-4093.

Teillet, P.M., Guindon, B., & Goodenough, D.G. (1982). On the slope-aspect correction of multispectral scanner data. *Can. J. Remote Sens, 8(2)*, 84–106.

Turner II, B.L. (1994). Local faces, global flows: the role of land use and land cover in global environmental change. *Land Degrad Rehabil, 5*, 71–78.

Turner II, B.L., Lambin, E.F., & Reenberg, A. (2007). The emergence of land change science for global environmental change and sustainability. *Proc Natl Acad Sci, 104*, 20666–20671.

Vermote, E., Tanré, D., Deuzé, J.L., Herman, M., & Morcrette, J.J. (1997) Second Simulation of the Satellite Signal in the Solar Spectrum (6S), 6S User Guide Version 2.

Weiser, R.L., Asrar, G., Miller, G.P., & Kanemasu, E.T. (1986). Assessing grassland biophysical characteristics from spectral measurements. *Remote Sens Environ, 20*, 141-152.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., & Woodcock, C.E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens Environ, 122*, 2–10.

Chapter 6: Urban land cover thematic disaggregation, employing datasets from multiple sources and Random Forests modeling

Authors: Dimitrios Gounaridis, Sotirios Koukoulas
Journal: International Journal of Applied Earth Observation and Geoinformation 51, (2016), 1–10.

Abstract

Urban land cover mapping has lately attracted a vast amount of attention as it closely relates to a broad scope of scientific and management applications. Late methodological and technological advancements facilitate the development of datasets with improved accuracy. However, thematic resolution of urban land cover has received much less attention so far, a fact that hampers the produced datasets utility. This paper seeks to provide insights towards the improvement of thematic resolution of urban landcover classification. We integrate existing, readily available and with acceptable accuracies datasets from multiple sources, with remote sensing techniques. The study site is Greece and the urban land cover is classified nationwide into five classes, using the Random Forests algorithm. Results allowed us to quantify, for the first time with a good accuracy, the proportion that is occupied by each different urban landcover class. The total area covered by urban land cover is 2280 km2(1.76% of total terrestrial area), the dominant class is discontinuous dense urban fabric (50.71% of urban land cover) and the least occurring class is discontinuous very low density urban fabric (2.06% of urban land cover).

Keywords: Urban land cover, Thematic disaggregation, Urban atlas, Landsat, Random Forests.

4.1 Introduction

Urban areas determine, both positively and negatively, several functions of the Earth system, from local to global scales (DeFries et al. 2010; Folke et al. 1997). Accurate information about urban land cover (ULC) is critical to a wide range of social, economic, and environmental research questions not only for descriptive but also for analytical and predictive modeling purposes. Thus, reliable spatial information about ULC composition and configuration serves as a principal component in a variety of scientific activities, across several disciplines, while for studies related to global, environmental and/or climate change it is considered a pre-requisite (Grimm et al. 2008; Mills, 2007).

In pursuit of spatial information about land cover (LC), traditional field data approaches face certain drawbacks as they are limited to a local extent due to their prohibitively expensive nature in means of time, costs and personnel. Technological and methodological advances in remote sensing (RS) and geographic information systems (GIS) successfully provide spatially consistent LC information. Nowadays, an increased number of satellite sensors has been launched and facilitate the growing demand for multi-spectral and multi-temporal information of the Earth's surface over a wide range of scales and data types (Belward and Skøien, 2014).

A number of studies have generated several datasets regarding ULC, or LC in general. The majority of them consider studies at the scale of individual cities, analyzing changes and patterns over multiple years or exploiting spatial information and structure on a single date (Yu et al. 2014). On a global scale, more than ten datasets have been produced with spatial resolutions ranging from approximately 10km to 30m (Chen et al. 2014; Potere et al. 2009). The limitations and drawbacks of these global datasets have been discussed in detail by several researchers (Congalton et al. 2014; Giri et al. 2013; Potere et al. 2009). The predominant conclusion stressed by these studies is that the most prominent drawback is the variability in ULC definition.

On a regional scale, for Europe, the CORINE land cover (CLC) is the most frequently used dataset with a hierarchical classification scheme comprising of 44 classes (at level 3) and a minimum mapping unit of 25 ha. The urban category denoted as 'urban fabric' is divided in two classes (continuous and discontinuous). Recently,

the Copernicus land monitoring service has released the Urban Atlas (UA) database. It consists of LC maps for 305 European large urban zones with more than 100.000 inhabitants for the reference year 2006 (European Commission, 2011). It has been derived by very high-resolution satellite data (spatial resolution between 2.5 and 5 m) and has a significantly lower minimum mapping unit of 0.25 ha, compared to CLC. The thematic resolution of UA, regarding ULC is also much more detailed than CLC - although it has limited geographic coverage- dividing the urban class into five classes differentiated by their degree of imperviousness.

Despite the unquestionable value of the datasets produced so far, their application to a range of research applications and management activities is inefficient. The reason is their resolution, both spatial and thematic, a constraint in cases when these data are to be used in studies that finer scale of analysis is mandatory (e.g. urban planning) or in studies that require sufficient thematic ULC detail (e.g. population density mapping). As far as spatial resolution is concerned, the previous efforts mainly employed coarse resolution satellite data for feasibility reasons (data availability, technical innovation, human and financial resources). However, ULC delineation employing coarse resolution primary data is not a simple task. On the one hand, ULC class has a limited areal extent in comparison with other classes, while on the other hand, it is a class with extreme variability in terms of spectral and textural characteristics. Thus, data derived by coarse spatial resolution are due to the mixed pixel effect, especially for the ULC class, where small area urban areas are often completely omitted, spatial details are lacking and the edges are erroneously presented (Potere et al. 2009; Schneider et al. 2010). Thematic resolution refers to the number of classes and the detail in their definition that determines the amount of geospatial information of hard classified categorical data. The more detail in a land user/cover map, the more meaningful and insightful the map is for a wide range of research questions. Several authors have explored the effects of thematic resolution in land use modeling (Conway, 2009; Pontius and Malizia, 2004), land-cover pattern analyses (Buyantuyev and Wu, 2007) and landscape indices behavior (Bailey et al. 2007), converging that the outcomes are significantly influenced. Whilst thematic resolution is important to a range of applications, available regional and global datasets in most cases represent ULC lumped into one or two broad classes, a fact that is far from reality on the ground, given the heterogeneity of urban areas across space (Potere et al. 2009).

In this paper we successfully disaggregate ULC patterns into five categories achieving nationwide coverage (for Greece). Additionally, we demonstrate a sequence of steps towards the improvement of existing shortcomings and scarcity of high quality data related to ULC. Our main focus was to achieve the highest possible thematic resolution without compromising accuracy. To this end, we employ the Random Forests (RF) machine learning algorithm (Breiman, 2001) that is proven to perform well in the face of heterogeneous classes. Our model is trained intensively by the polygon centroids of the UA dataset -available for nine cities- to finally 'predict' ULC for the rest of the geographic coverage of Greece. Road density, population, LC and spectral indices derived by Landsat satellite, serve as predictor variables.

The rest of the paper is structured as follows: We first present the study site, Greece, along with information about morphology, recent population dynamics and some causal factors that contributed to the existing ULC scenery. Next, we present an overview and the data used as both response and predictor variables to train our models, along with the data pre-processing steps. Then, the RF classifier application and the accuracy assessment process are described in detail. In the next section, we present the obtained results and we discuss the model performance. Finally, in the last section we discuss the conclusions drown and we highlight some key points.

4.2 Material and methods

4.2.1 Study site

Greece is a Mediterranean country of Southeast Europe situated between latitudes 34° and 42° N, and longitudes 19° and 30° E (Figure 1) and is populated by approximately 11 million inhabitants. Two-thirds of the inhabitants live in urban areas, while the remaining one-third are rural inhabitants (Hellenic Statistical Authority, 2013). Almost two thirds of the Greek territory are mountainous, with Mount Olympus being the highest at 2.917m (European Environment Agency, 2010). Extensive agricultural plains are primarily located in Thessaly, Central Macedonia and Thrace regions, constituting key economic sources.

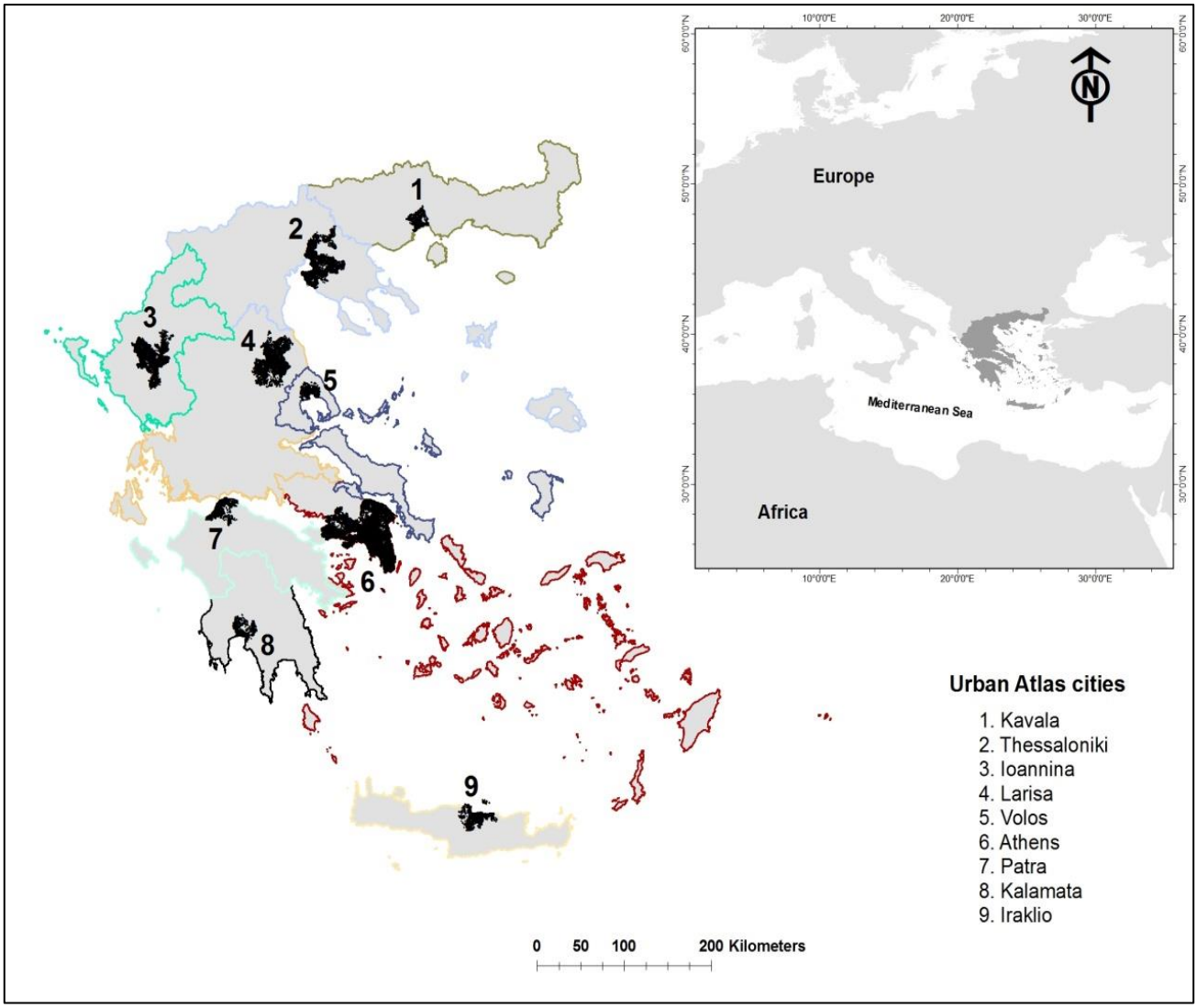


Figure 1. Study site Greece, divided into nine areas containing one UA city each

Greece has a long history of land use, ranging from prehistoric to present times constituting a country of people with strong dependency on the land. The last decades of the 20th century the economic potential of urban centers motivated a constant societal demand to capture new economic opportunities, a fact that consequently triggered a shift of rural population (Kasimis et al. 2003). In turn, rural land abandonment progressively led to marginalization of remote areas especially in the uplands (MacDonald et al., 2000) while leading to agricultural and farming intensification at the lowlands (Beopoulos and Skuras, 1997). Significant expansion of the tourism sector as well as a trend in second homes gave a boost in growth dynamics of the built environment, especially in the coastal zones. At the same time, the aforementioned

developments are perceived as both consequences and driving forces of public works and transport infrastructures expansion. Thus, the demographic dynamics and the major socio-economic changes that have taken place progressively brought radical changes in Greek landscapes (Zomeni et al. 2008).

Statistics clearly advocate all the aforementioned. According to the latest census of 2011, the Greek ULC scenery consists of 13220 settlements, 746 (5.6 %) of which are uninhabited, 6897 (52.2 %) have less than 100 residents and 8806 (66.6 %) have less than 200 inhabitants. At the same time, according to the 2011 census, Attica prefecture is inhabited by 3.827.624 residents (35% of total population) and the prefecture of Thessaloniki by 1.880.058 residents (17% of total population).

4.2.2 Overview

The RF algorithm is a robust non-parametric machine learning algorithm (Breiman, 2001) that has been widely used for LC classification. Initially, the algorithm uses a randomly selected part of training observations (response variable) as well as a sample of predictor variables, resulting in a number of independent to each other classification trees. This process is repeated several hundreds of times, thus forming a 'forest' of classifiers. Each tree contributes with a single vote to the assignment of the most frequent class. The final outputs of classification are determined from the majority of votes for each class (Breiman, 2001). The main advantages of adopting RF in our task are: i) the independency of each classification tree, on the one hand, and the randomness of variable selection, on the other, reduce the problem of overfitting and at the same time make the models insensitive to noise and outliers (Breiman, 2001; Chan and Paelinckx, 2008). ii) The algorithm can efficiently handle predictor variables, with different nature (both continuous and categorical) and from multiple sources (Gounaridis et al. 2014; Gounaridis et al. 2015) which is the case for our approach. iii) The first two advantages of RF contribute to good performance in the classification of heterogeneous landscapes (Rodriguez-Galiano et al. 2012a; Timm and McGarigal, 2012) such as ULC. iv) RF can handle large datasets and thousands of input variables being computationally faster than other classifiers (Rodriguez-Galiano et al. 2012b) and thus can efficiently handle broad scale tasks. v) The importance of each input variable

is quantified allowing for several tests to determine whether a variable will be included in the model or not.

Discrimination and classification of ULC with the use of remote sensing techniques is not a simple task due to the complex nature of this LC category. Built-up environment is commonly characterized by heterogeneous patterns and a variety of mixed land uses. Residential buildings, transportation networks, industrial and commercial infrastructures, open spaces and vegetated patches co-exist in a ULC patch and are often composed by a wide range of surface types, materials and thus, spectral responses. A number of authors attempted to delineate built-up areas and at the same time deal with the challenge of the mixed pixel problem. The majority of studies exploit full spectral information combined with texture information (Lu and Weng, 2005), spectral indices (Xu, 2007), spectral unmixing techniques (Wu, 2004) and Nighttime Lights (Elvidge et al. 2007). Lately, the use of auxiliary non-spectral variables has been found to improve the classification accuracy of certain LC classes (Rodriguez-Galiano and Chica-Olmo, 2012). Additionally, methodological advancements of algorithms that can usefully handle different type of data, such as RF, increase the inclusion of more and more morphological and socioeconomic variables to serve as proxies to ULC classification.

4.2.3 Data processing

4.2.3.1 Response variable

To train the RF algorithm we extracted the polygon centroids of the UA dataset. By using the centroids instead of another sampling strategy, we ensured taking clear samples of each category avoiding at the same time any sampling near boundaries of adjacent categories that would lead to discrepancies. The major advantages of using the UA dataset as training for our models was that i) it has the highest thematic resolution available, distinguishing ULC into five categories. ii) Despite its limited geographic coverage, it is derived from high resolution imagery and providing thousands of accurate samples and thus consisting a valuable source for extensive training for our task. iii) It is free and readily available. For Greece nine cities in total have been mapped so far (Figure 1). The UA data were re-projected to the Greek Geodetic Reference System (GGRS 1987) and reclassified into six final classes. Table 1 presents the training used for each of the nine regions of this study.

Table 1. Number of UA polygon centroids used to train the RF algorithm for each of the nine areas depicted in Figure 1.

	1*	2*	3*	4*	5*	6*	Total
Kavala	774	352	126	49	8	1151	2460
Thessaloniki	7313	3925	962	111	43	8963	21317
Ioannina	1731	1584	740	372	56	6308	10791
Larisa	2827	1909	535	130	78	5502	10981
Volos	2065	655	155	110	71	2001	5057
Athens	28562	19638	11260	7428	1329	30847	99064
Patra	1311	1430	855	458	21	2754	6829
Kalamata	685	425	398	461	240	1703	3912
Iraklio	1160	593	506	405	102	1858	4624

1*. Continuous urban fabric

2*. Discontinuous dense urban fabric

3*. Discontinuous medium density urban fabric

4*. Discontinuous low density urban fabric

5*. Discontinuous very low density urban fabric

6*. Other use

4.2.3.2 Predictor variables

Several datasets can be used as predictor variables for ULC classification. RF algorithm offers the flexibility that data of different nature and value scaling can be incorporated in the model. After several tests (not presented here) we concluded to a set of predictor variables that best facilitate the discrimination of ULC classes, while taking into account two criteria: 1) Simplicity of the model and feasibility in terms of computational cost and 2) availability of data to enhance reproducibility of our approach to other regions.

Assuming that human activity and artificial environment are connected with the existence of roads, one of the most prominent proxies of ULC is the road network (Gounaridis et al. 2014; Gounaridis et al. 2015; Hawbaker et al. 2004). Another useful

proxy with apparent interrelation with ULC is population density (Mesev, 1998). Additionally, the last few decades there have been efforts towards the impervious surface mapping with the use of remote sensing (Sutton et al. 2009). Inclusion of impervious surface information can serve as a base and enhances the accuracy of ULC mapping (Lu and Weng, 2006).

4.2.4 Data processing

Spectral indices: To retrieve spectral information, for our case study, we employed the Global Land Survey (GLS) 2010 Landsat imagery dataset that uses a combination of Landsat 5 TM and Landsat 7 ETM+ images acquired between 2008–2011 (Gutman et al., 2013). A total of 27 GLS images (spanning path: 180–186 and row: 31–36) were selected and processed, achieving full coverage of the country. To avoid any discrepancies due to the multi-temporal and double-sensor type of analysis all images underwent a series of corrections and calibration. To radiometrically and atmospherically correct the Landsat images, first we converted the DN numbers into top of atmosphere reflectance and second to obtain surface reflectance and achieve data normalization, we applied the 6S model (Vermote et al. 1997). Finally, to correct for topographic effects, we applied the C-correction method (Teillet et al. 1982). The corrected bands of the 27 Landsat images were processed in order to calculate two spectral indices, the normalized difference built-up index (NDBI) (Zha et al. 2003) and the enhanced built-up and bareness index (EBBI) (As-Syakur et al. 2012). NDBI has the capacity to help classification models in distinguishing built-up areas and barren land as it involves spectral information of near and middle infrared wavelengths which are sensitive to the spectral response of these classes. EBBI both involves near and middle infrared wavelengths spectral information and incorporates the thermal band of Landsat as well. This was reported that it enhances the distinction between built-up areas and barren land classes, thus achieving better results compared to other similar spectral indices (As-Syakur et al. 2012).

Population: The latest national population census data of 2011 (Hellenic Statistical Authority, 2013) were acquired and geo-coded in point vector format. Next, the point data were spatially aggregated in administrative polygons and finally converted into raster format. We did not use the raster of population density surface available by the European Environment Agency (EEA), as it is outdated and of coarser resolution (Gallego, 2010).

Road network: Spatial representation of the national road network was acquired from the OpenStreetMap (OSM) project. The OSM project is one of the most established sources of voluntarily derived, free access and up to date datasets (Ramm et al. 2011). The degree of completeness specifically of the road network is high especially for the cities, and the data coverage is expected to be improved in the coming years (Jokar Arsanjani et al. 2013). The data were re-projected onto the Greek Geodetic Reference System (GGRS 1987). Roads are constructed for a variety of purposes and thus different types of roads may not indicate ULC, a fact that might diminish the model's performance. To this end, the lines assigned as "residential road" and "living street" class were extracted. Finally, a raster density map was generated using a moving Gaussian kernel density function with equal split and a bandwidth of 200m (Okabe et al. 2009).

Land cover: We used a newly produced land cover dataset for Greece for the year 2010 (Gounaridis et al. 2015). This dataset was derived by the 2010 Landsat GLS imagery and the RF algorithm. The algorithm was trained semi-automatically, extracting information from the CORINE 2000 LC dataset. The overall accuracy is 83%. The nomenclature adopted was similar to CORINE at hierarchical level 2, thus ULC is represented in one single class.

Soil sealing: We acquired the sealed land dataset produced as part of the Global Monitoring for Environment and Security (GMES) Fast Track Service on Land Monitoring in 2006–2008 and disseminated by EEA in 2009. This dataset, available in 20m spatial resolution, refers to the degree of soil sealing (continuously ranging from 0% to 100%) for 2006, with classification accuracy higher than 85% per hectare (European Environment Agency 2011). From this dataset, our case study area, Greece, was extracted and the resulting raster map was re-sampled to 30m and re-projected on the GGRS 1987 system.

4.2.5 Variables exploratory analyses

We concluded in a set of predictor variables that best describe ULC. Figure 2 illustrates the interrelationships between the predictors and response variable for the five ULC classes, excluding "other use" (Table 1). The values of EBBI and NDBI are within a certain range for all road density and soil sealing values, where negative values correspond to non-built-up areas. The population values coincide with almost all values

of sealed land. Unsurprisingly, higher population values coincide with high values of the soil sealing layer, while smaller values (including classes of lower ULC densities) can be encountered in less sealed areas. The values of road density behave similarly. Areas with all values of road density coincide with higher amounts of soil sealing, while there is a small portion of areas with less dense road network that can be encountered in less sealed land. All predictor values, except the LC_2010 dataset, follow a reasonable route across different ULC densities. Soil sealing, road density, EBBI and NDBI seem to decrease accordingly as ULC density also decreases, while population has, as expected, higher values for continuous and discontinuous dense urban fabric. All five categories of ULC though coincide with all LC_2010 categories, where they should be expected to coincide with only urban values. The fact that ULC values coincide with non-urban categories indicates that LC_2010 dataset underestimates the ULC extent.

4.2.6 Model implementation

Figure 3 summarizes the sequence of procedures followed for our analyses. Initially, Greece was split into nine, relatively equal parts, each one containing an UA city (Figure 1). All data were masked into these nine parts, trained and run individually. This step allowed the process to be held, otherwise the computational demand at a national scale of analysis would be prohibitively high. The values of each predictor variable layer were collected on the location of the UA centroid, already containing LC class values (Table 1) and used as predictor values. Next, the classification was implemented through the use of randomForest package (Liaw and Wiener, 2002) in R open-source statistical software. For classification, RF requires two primary parameters: (i) the number of predictor variables randomly sampled at each decision tree split and (ii) the number of decision trees. We used the value of three (3) predictor variables for each tree split (equal to the square root of the total number of predictor variables). We did not apply any optimization method, as it increases the computational sources demand. The number of decision trees was set to 5000 for each run since it was found that the higher the number, the more the classification error converges (Rodriguez-Galiano et al., 2012a).

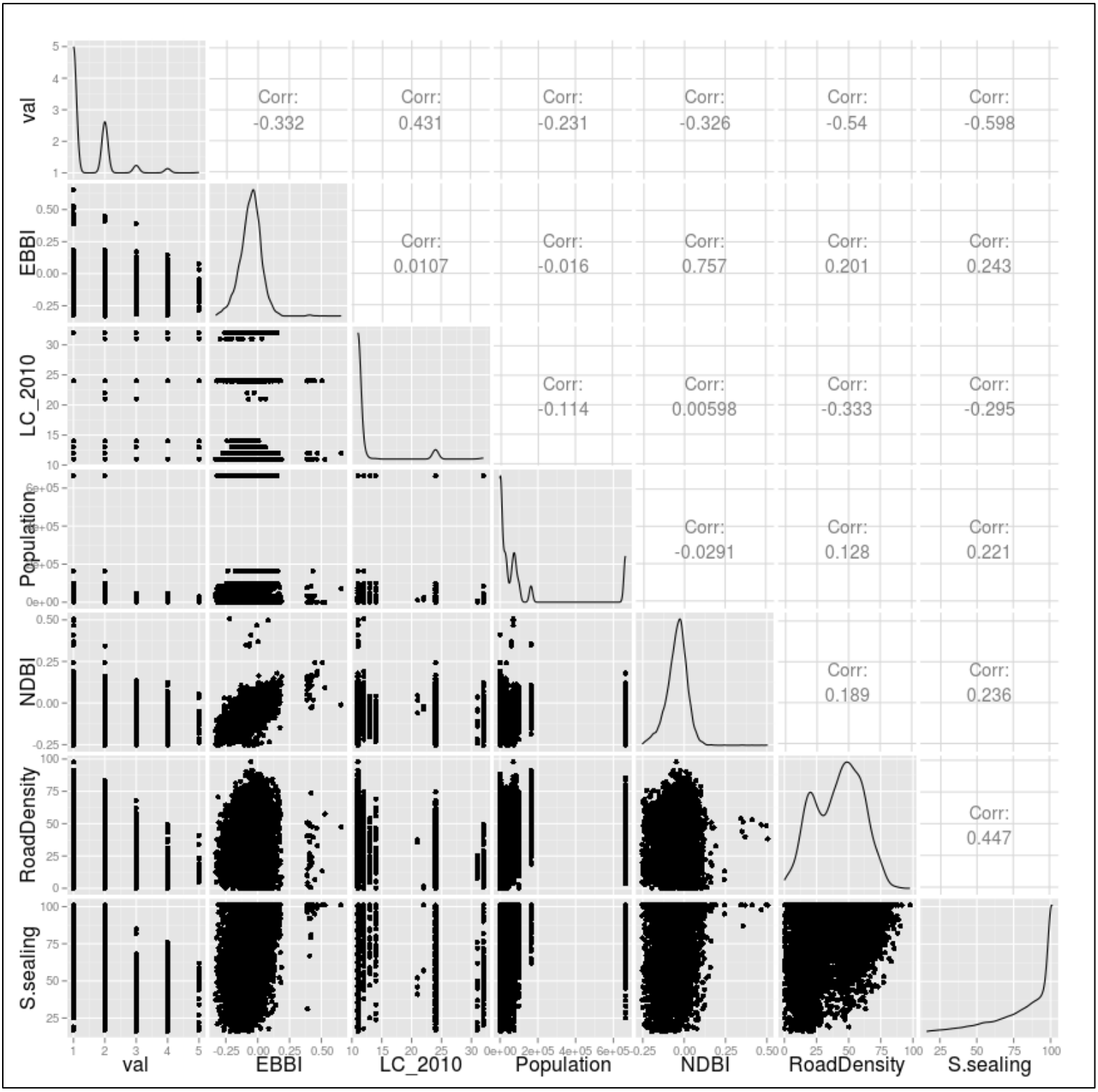


Figure 2. Inter-relationship between predictor and response variables to be used in RF models. The values consist a 10% random sample of the dataset used to train the model of Athens. The 6th category (other use - please see table 1) was excluded.

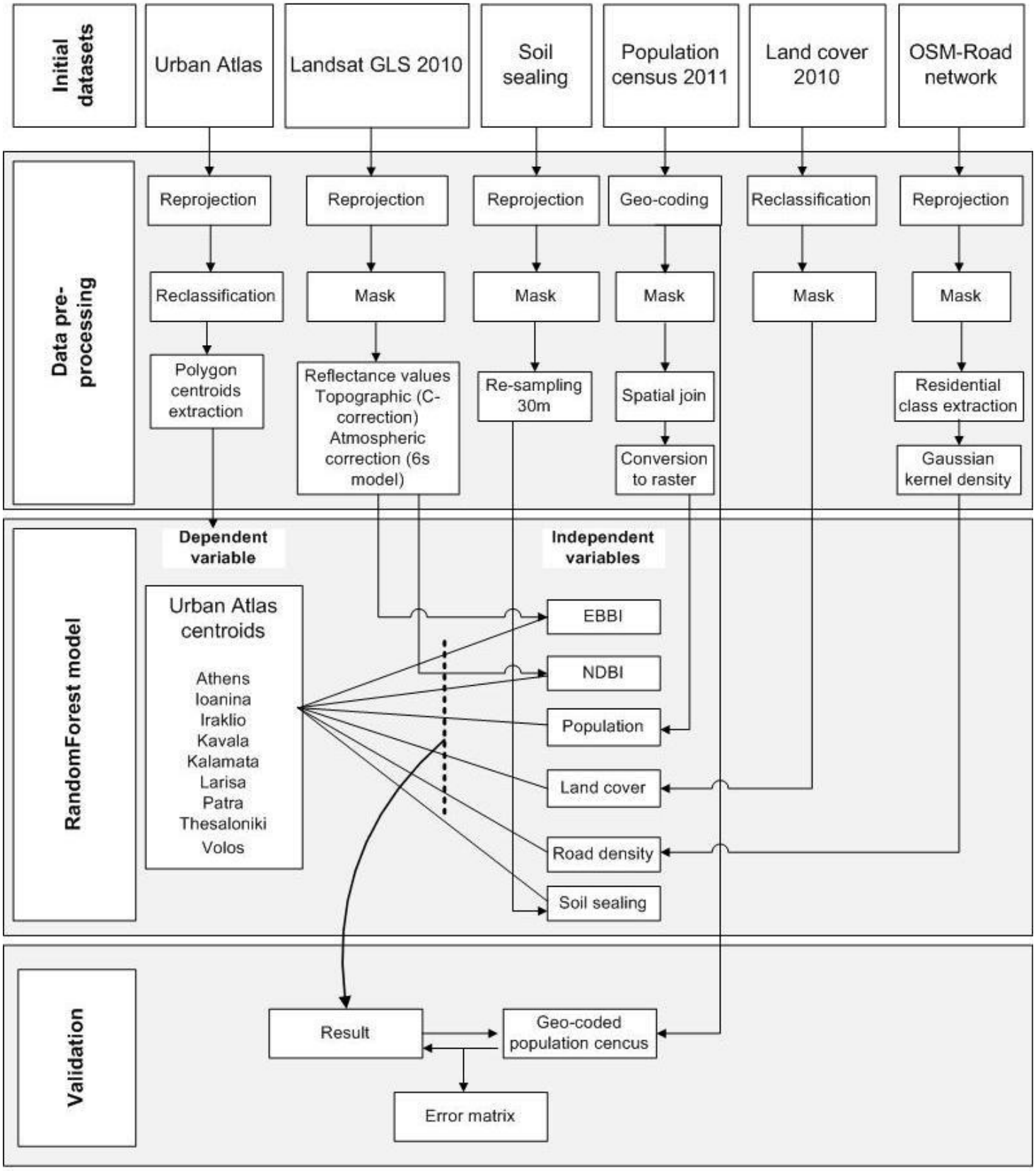


Figure 3. Workflow of the presented methodology

4.2.7 Accuracy assessment

We assessed the classification's accuracy at two levels, both for positional errors and thematically. Initially, we used the geo-coded data of the population census, in point vector format. Given the small portion of ULC compared to other LC classes, a random distribution of points to our map would result in weak assessment of the ULC classes accuracy. By using geo-coded points, this weakness was diminished as this dataset actually represents all settlements of Greece for the year 2011, recorded via *in situ* survey. Moreover, we included a validation dataset consisting of a set of points that were photo-interpreted using very high-resolution imagery (Gounaridis et al. 2015). To assess the accuracy for the thematic categorization, we re-classified the soil sealing layer according to Lu and Weng (2006). The re-classified soil sealing values along with the values of the resulting map were collected at the settlements location. Results were tabulated, generating an error matrix that allowed the computation of overall accuracy and the errors of commission and omission for each class.

4.3 Results and discussion

Table 2 shows that results were mainly satisfactory with an overall accuracy of 81.8 %. We also estimated omission and commission errors at the class level. Some errors were from the confusion between adjacent classes, which is a frequently encountered issue in classification. The major disagreement appears between the "Discontinuous very low density urban fabric" with "Discontinuous low density urban fabric" and "Other use". These two classes represent very sparse built-up urban fabric with less than 10 % and 10 - 30 % sealed land respectively. They consist of heterogeneous areas encompassing irregularly, incompact and scattered settlements with significant differences from adjacent LC classes (e.g. agricultural fields, forests). Especially in the "Discontinuous very low density urban fabric" the asphalt road network is often absent. Thus, in many cases the adjacent LC dominates the spectral response of the area leading to 'weak' evidences of their built-up patches existence. In other words, the mixed pixel effect is particularly noticeable in such heterogeneous landscapes.

Table 2. Error matrix

	Resulting map							
Reference	**1***	**2***	**3***	**4***	**5***	**6***	**Total**	**Omission Error (%)**
1*	**980**	314	71	34	19	83	**1501**	34,7
2*	224	**1765**	102	33	11	29	**2164**	18,4
3*	38	55	**707**	46	19	49	**914**	22,6
4*	42	23	69	**394**	22	90	**640**	38,4
5*	14	26	22	47	**65**	53	**227**	65,2
6*	51	95	42	41	54	**4249**	**4532**	6,7
Total	**1349**	**2278**	**1013**	**595**	**190**	**4553**	**9978**	
Comission Error (%)	27,3	22,5	30,2	33,7	65,7	6,6		
Overal accuracy	81,8%							

1*. Continuous urban fabric

2*. Discontinuous dense urban fabric

3*. Discontinuous medium density urban fabric

4*. Discontinuous low density urban fabric

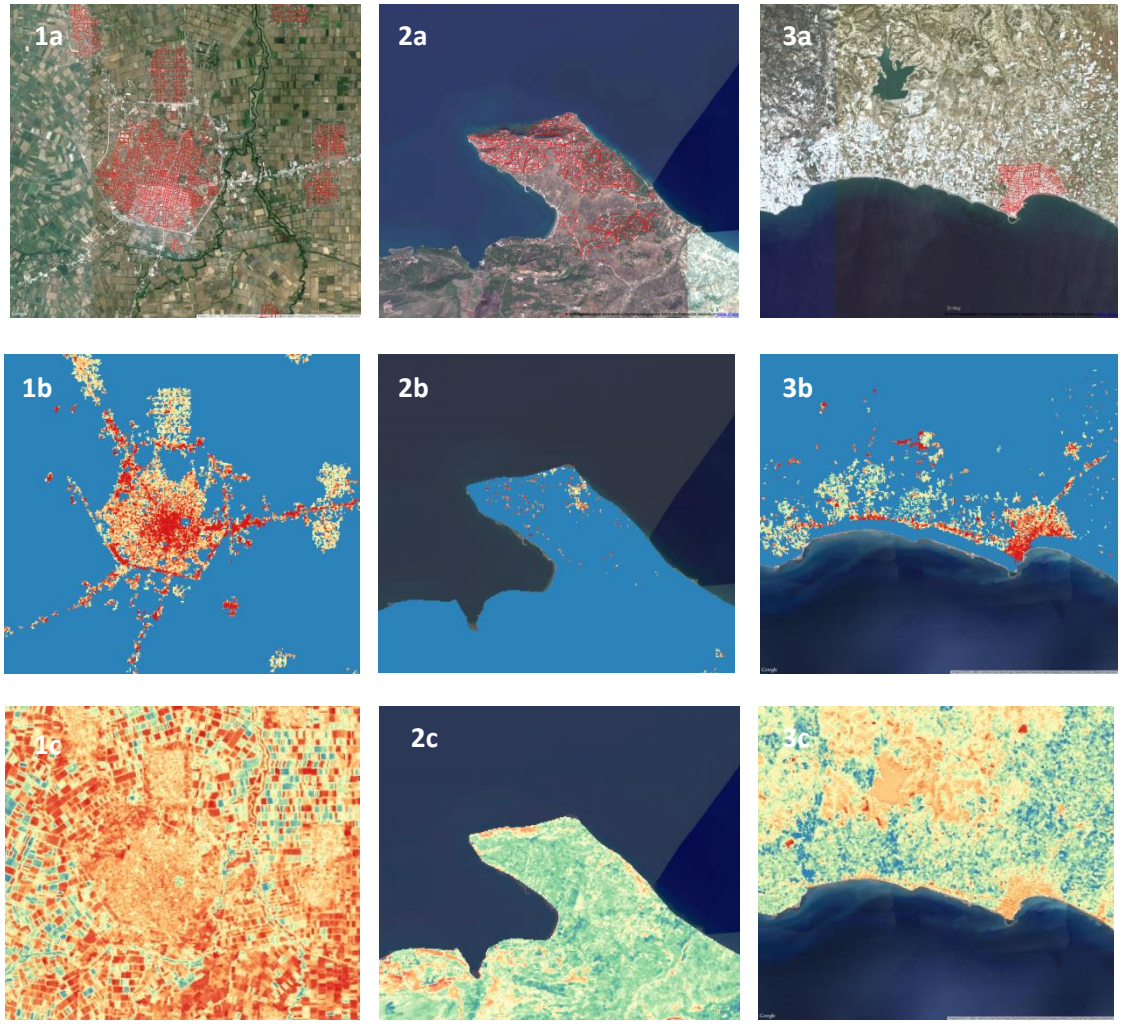
5*. Discontinuous very low density urban fabric

6*. Other use

A sound way to overcome these shortcomings would be the utilization of very high resolution (both spatial and spectral) imagery. The nature and the extent of our analyses ruled out this option. Medium to high spatial resolution data and Landsat imagery for spectral derivatives were the only viable way. RF model's performance, as expected for all machine learning algorithms, was found to be highly dependent on the predictor variables and training data quality. All included predictor variables were datasets of acceptable accuracies, as discussed in the "2.3.2 Predictor variables" section. The UA centroids, used as training, were also of acceptable accuracy and quantity (165,035 in total). One important part regarding quantity is the relatively equal representation of classes that in our case, was an unreached goal. Nevertheless, the nine cities that have already been mapped by UA had inequalities in size and composition resulting in unequal representation of classes (Table 1).

However, the distinction of ULC classes and our primary objective to achieve high thematic resolution has been met. Our models performed well in cases were previous efforts failed due to spectral and textural confusion of similar but different adjacent LC

classes. All six variables included in our models contributed more or less to obtain satisfactory results. Figure 5 illustrates the variables ordered top-to-bottom as most to least important. The mean decrease in accuracy is a score calculated during the out of bag error calculation phase and it informs about how much the accuracy decreases if a variable would be permuted or in other words excluded from the model. Therefore, the larger the value of mean decrease, the higher the importance of a variable is. The mean decrease in Gini coefficient is a score informing about each variable's contribution to the impurity of the resulting random forest. Variables with a high value in the decrease Gini score, have nodes with high purity and thus contributes to the model's homogeneity. In all cases, the road density had the most important role in the models while the population and the EBBI were the second most important variables. The order of importance for both scores varied slightly between the models of the nine sub-regions.



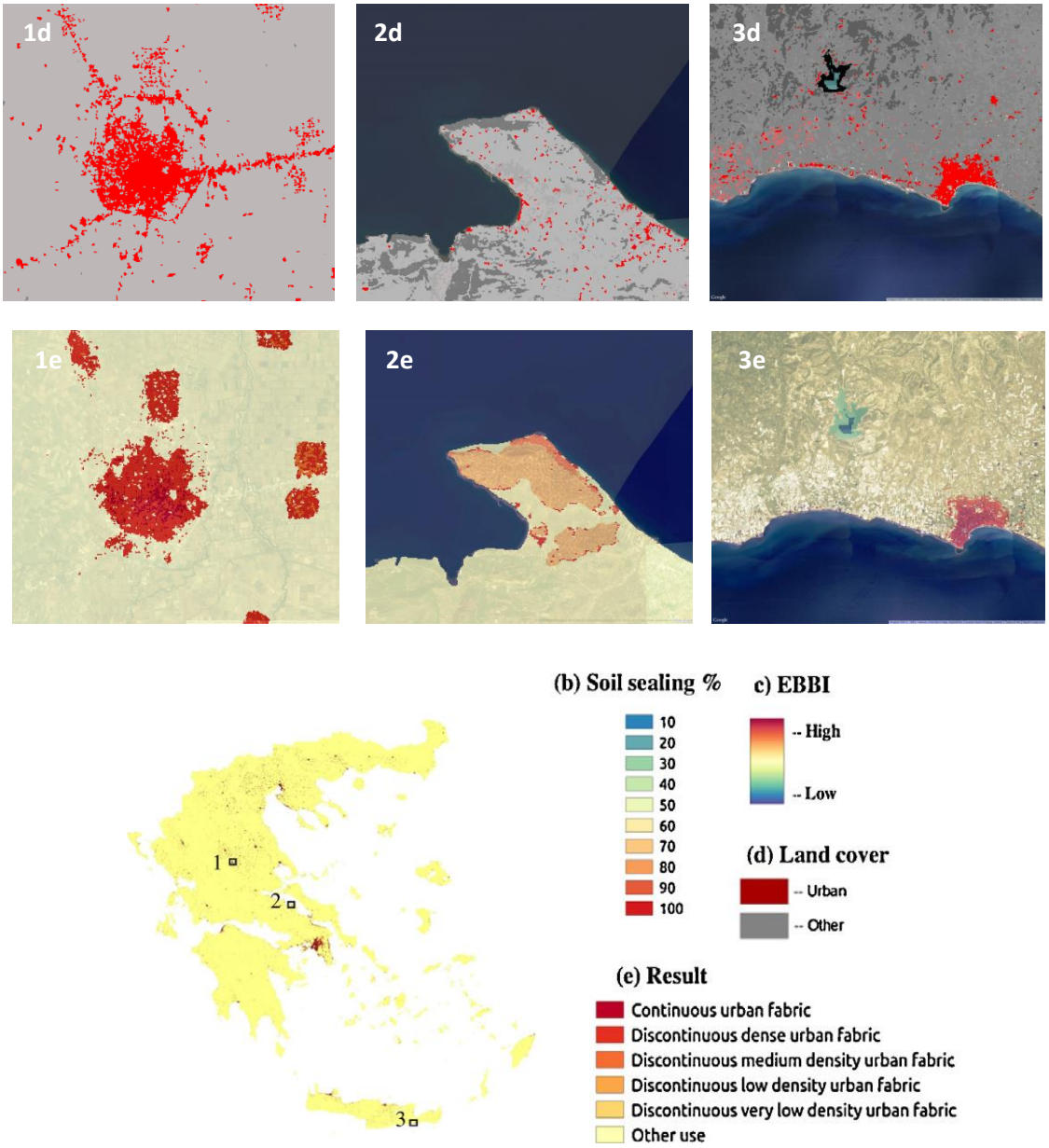


Figure 4. Representative examples of results and predictor variables. 1a;2a;3a: Google earth imagery with the red line representing the residential road network. 1b;2b;3b: Degree of soil sealing. 1c;2c;3c: EBBI. 1d;2d;3d: LC 2010 dataset. 1e;2e;3e: Resulting map.

Figure 4 illustrates three representative examples of RF good performance and highlights how the combination of predictor variables used allowed the adequate discrimination of ULC. The first example depicts the city of Karditsa, which lies in the plain of Thessaly. The city is surrounded by a highway, industrial facilities, agricultural fields and smaller ULC patches dispersed in the peri-urban zone (Figure 1a). The soil sealing layer (Figure 1b), as expected, has higher values of imperviousness in the highway than in many ULC areas, while the 2010 LC dataset considers all bult-up areas

as ULC (Figure 1d). EBBI index alone (Figure 1c) fails to delineate ULC as the area is surrounded by agricultural fields in the dry season, where soil dominates and spectral responses are confusing.

However, our model achieved an adequate discrimination of the classes (Figure 1e). The second example depicts a settlement of the prefecture of Fthiotida in central Greece. This landscape is a typical example of very low density built-up area, surrounded by forests, schrublands and tree crops (Figure 2a). The mixed pixel effect is particularly evident here, as it can be seen by the soil sealing values (Figure 2b). Following, the 2010 LC dataset partially underestimates the ULC extent (Figure 2d) while EBBI (Figure 2c) is prone to the mixed pixel problem, underestimating, in most parts, the actual ULC area and in other parts overestimating due to confusion with barren land (left part of Figure 2c). As seen in Figure 2e, our model classifies accurately this landscape. The third example depicts the city of Ierapetra, located in the southeast of Crete. The economy of Ierapetra relies mainly on farming and agriculture. Thus, the area next to the city is occupied by agricultural fields, pastures and farms (Figure 3a). The landscape is dominated by numerous greenhouses and tin roofed farm establishments. As a result, the impervious surfaces expand extensively outside of the city (Figure 3b). At the same time, EBBI successfully distinguishes the greenhouses from ULC materials but failes to distinguish ULC from bare rocks (Figure 3c). Finally, 2010 LC dataset (Figure 3d) delineates rather well the area, with minor overestimation of ULC. Results show (Figure 3e) that our model confuses neither greenhouse material nor bare rock with ULC.

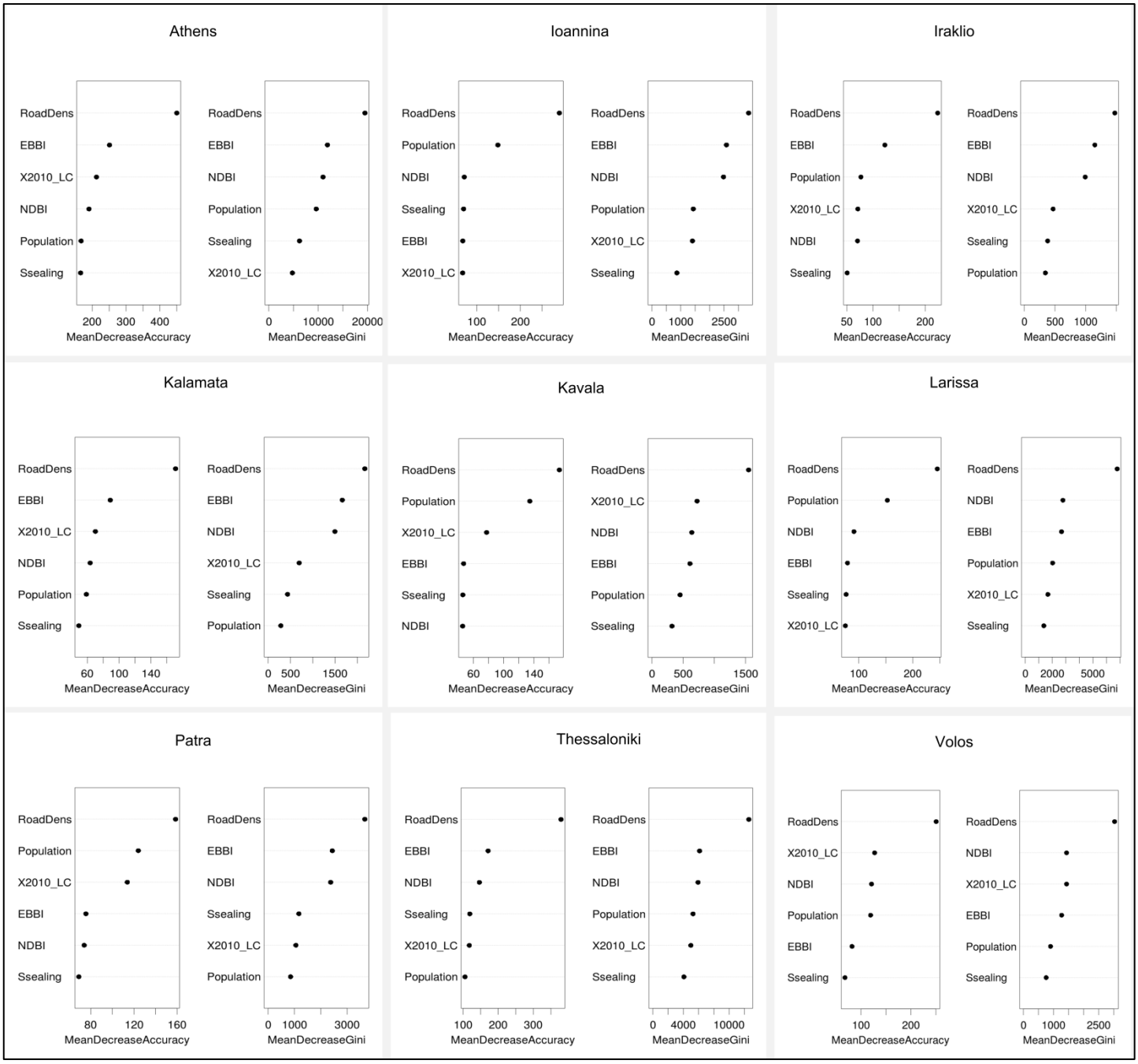


Figure 5. Importance of variables used for the nine RF models.

Results allowed us to quantify, with a certainty, the total area occupied by each ULC class (Table 3). ULC occupies an area of approximately 2280 km^2 which is approximately 1.8 % of the total terrestrial area (including the inland water). Dominant ULC type is discontinuous dense urban fabric, whereas the least occurring is discontinuous very low density urban fabric with 50.71 % and 20.06 % respectively. These numbers provide an insight into the ULC scenery of Greece and can be attributed

to the developmental and demographic shifts that took place during the last decades of the 20th century.

Table 3. Quantification of the area occupied by each ULC class.

Urban land cover class	Area (km^2)	Percentage of total terrestrial land cover (%)	Percentage of urban land cover (%)
Continuous urban fabric	456,66	0,35	20,03
Discontinuous dense urban fabric	1155,92	0,89	50,71
Discontinuous medium density urban fabric	391,53	0,30	17,18
Discontinuous low density urban fabric	228,39	0,18	10,02
Discontinuous very low density urban fabric	46,93	0,04	2,06
Other use	127,792,48	98,25	
Total	130,071,91	100,00	100

4.4 Conclusions

The objectives of this paper were to investigate if we can effectively fuse existing datasets and remote sensing techniques in order to classify ULC into five classes and finally to accurately quantify the area occupied by each class. To do so, we integrate data of soil sealing, population, road network density, LC and two spectral indices for built-up area delineation. We demonstrate that data from multiple sources, meeting certain criteria (e.g. costs, quality, computational feasibility), can be combined in a modeling framework and provide satisfactory results and high thematic resolution. Considering the general characteristics of the described approach, we expect that the present paper will contribute to the generation of improved ULC mapping, which was also one of our goals. The modeling framework discussed here is suitable for a wide range of applications and can act as a baseline in planning, steering and monitoring of LC and its associated changes.

References

As-Syakur, A.R., Adnyana, I., Arthana, I.W. & Nuarsa, I.W. (2012). Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area. *Remote Sens,* 4, 2957–2970.

Bailey, D., Herzog, F., Augenstein, I., Aviron, S., Billeter, R., Szerencsits, E. & Baudry, J. (2007). Thematic resolution matters: indicators of landscape pattern for European agro-ecosystems. *Ecol Indic,* 7, 692-709.

Belward, A.S., & Skøien, J.O. (2014). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites *ISPRS J Photogram* (in press).

Beopoulos, N. & Skuras, D. (1997). Agriculture and the Greek Rural Environment. *Sociol Ruralis*, 37(2), 255–269.

Breiman, L. (2001). Random forests. *Mach Learn*, 40, 5–32.

Buyantuyev, A. & Wu, J. (2007). Effects of thematic resolution on landscape pattern analysis. *Landscape Ecol*, 22, 7–13.

Chan, J.C.W. & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens Environ*, 112 (6), 2999–3011.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X. et al. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J Photogram,* 103, 7-27.

Congalton, R.G., Gu, J., Yadav, K., Thenkabail, P. & Ozdogan, M. (2014). Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sens,* 6, 12070-12093.

Conway, T.M. (2009). The impact of class resolution in land use change models. *Computers Environ Urban*, 33, 269–277.

DeFries, R.S., Rudel, T., Uriarte, M. & Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nat Geosci,* 3, 178–181.

Elvidge, C.D., Tuttle, B.T., Sutton, P.C., Baugh, K.E., Howard, A.T., Milesi, C., Bhaduri, B.L., Nemani, R. (2007). Global distribution and density of constructed impervious surfaces. *Sensors,* 7, 1962–1979.

European Commission. (2011). Urban atlas - Delivery of land use/cover maps of major European agglomerations. Final report V 2.0. Available at http://ec.europa.eu/regional_policy/sources/tender/pdf/2012066/urban_atlas_final_report_112011.pdf (assessed in 11/3/2015)

European Environment Agency, 2010. The European environment – state and outlook 2010: synthesis. European Environment Agency, Copenhagen.

European Environment Agency. (2011). Urban Soil Sealing in Europe. European Environment Agency, Copenhagen.

Foley J.A., DeFries R., Asner G.P., Barford C., Bonan G., Carpenter S.R. et al. (2005). Global Consequences of Land Use. *Science*, 309, 570–574.

Folke, C., Jansson, A., Larsson, J. & Costanza, R. (1997). Ecosystem appropriation by cities. *Ambio*, 26 (3), 167-172.

Gallego, F.J. (2010). A population density grid of the European Union. *Popul Environ*, 31(6), 460–473.

Giri, C., Pengra, B., Long, J. & Loveland, T.R. (2013). Next generation of global land cover characterization, mapping, and monitoring. *Int J Appl Earth Obs*, 25, 30–37.

Gounaridis, D., Zaimes, N.G. & Koukoulas, S. (2014). Quantifying spatio-temporal patterns of forest fragmentation in Hymettus Mountain, Greece. *Computers Environ Urban*, 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.

Gutman, G., Huang, C., Chander, G., Noojipady, P., & Masek, J.G. (2013). Assessment of the NASA–USGS Global Land Survey (GLS) datasets. *Remote Sens Environ*, 134, 249–265.

Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J.G., Bai, X. & Briggs, J.M. (2008). Global change and the ecology of cities. *Science*, 319, 756–760.

Hawbaker, T.J., Radeloff, V.C., Hammer, R.B. & Clayton, M.K. (2004). Road density and landscape pattern in relation to housing density, land ownership, land cover, and soils. *Landscape Ecol*, 20, 609–625.

Hellenic Statistical Authority (EL.STAT) (2013). 2011 Greek Census. http://www.statistics.gr/portal/page/portal/ESYE

Jokar Arsanjani, J., Helbich, M., Bakillah, M. & Loos, L. (2013). The emergence and evolution of OpenStreetMap: A cellular automata approach. *Int J Digital Earth*, 8, 1-30.

Kasimis, C., Papadopoulos, A.G. & Zacopoulou, E. (2003). Migrants in Rural Greece. *Sociol Ruralis*, 43, 167–184.

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

Lu, D. & Weng, Q. (2005). Urban Classification Using Full Spectral Information of Landsat ETM_ Imagery in Marion County, Indiana. *Photogramm Eng Rem Sens*, 71 (11), 1275–1284.

Lu, D. & Weng, Q. (2006). Use of impervious surface in urban land-use classification. *Remote Sens Environ*, 102, 146–160.

MacDonald, D., Crabtree, J.R., Wiesinger, G., Dax, T., Stamou, N., Fleury, P. et al. (2000). Agricultural abandonment in mountain areas of Europe: environmental consequences and policy response. *J Environ Manage*, 59, 47–69.

Mesev, V. (1998). The Use of Census Data in Urban Image Classification. *Photogramm Eng Rem Sens*, 64(5), 431-438.

Mills, G. (2007). Cities as agents of global change. *Int J Climatol*, 27, 1849–1857.

Okabe, A., Satoh, T. & Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int J Geogr Inf Sci*, 23(1), 7-32.

Pontius, R. G., Jr., & Malizia, N. R. (2004). Effect of category aggregation on map comparison. In M. J. Engenhofer, C. Freska, & H. J. Miller (Eds.), GIScience 2004 (pp. 251–268). New York: Springer.

Potere, D., Schneider, A., Angel, S., & Civco, D.L. (2009). Mapping urban areas on a global scale: Which of the eight maps now available is more accurate? *Int J Remote Sens*, 30, 6531–6558.

Ramm, F., Topf, J. & Chilton, S. (2011). OpenStreetMap: using and enhancing the free map of the world. 3rd ed. Cambridge: UIT Cambridge.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012a). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogram*, 67, 93–104.

Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C. (2012b). Random Forest classification of Mediterranean land

cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens Environ,* 121, 93–107.

Rodriguez-Galiano, V. & Chica-Olmo, M. (2012). Land cover change analysis of a Mediterranean area in Spain using different sources of data: Multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. *Appl Geogr*, 35, 208-218.

Schneider, A., Friedl, M.A. & Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sens Environ*, 114, 1733–1746.

Sutton, P.C., Anderson, S.J., Elvidge, C.D., Tuttle, B.T. & Ghosh, T. (2009). Paving the planet: impervious surface as proxy measure of the human ecological footprint. *Prog Phys Geog*, 33(4), 510-527.

Teillet, P.M., Guindon, B. & Goodenough, D.G. (1982). On the slope-aspect correction of multispectral scanner data. *Can J Remote Sens*, 8(2), 84–106.

Timm, B.C. & McGarigal, K. (2012). Fine-scale remotely-sensed cover mapping of coastal dune and salt marsh ecosystems at Cape Cod National Seashore using Random Forests. *Remote Sens Environ,* 127, 106–117.

Vermote, E., Tanré, D., Deuzé, J.L., Herman, M. & Morcrette, J.J. (1997a). Second Simulation of the Satellite Signal in the Solar Spectrum (6S), 6S User Guide Version 2.

Wu, C. (2004). Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sens Environ*, 93, 480-492.

Xu, H. (2007). Extraction of Urban Built-up Land Features from Landsat Imagery Using a Thematic oriented Index Combination Technique. *Photogramm Eng Rem Sens*, 73 (12), 1381–1391.

Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L. et al. (2014). Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int J Remote Sens,* 35(13), 4573-4588.

Zha, Y., Gao, J. & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int J Remote Sens,* 24, 583–594.

Zomeni, M., Tzanopoulos, J. & Pantis, J.D. (2008). Historical analysis of landscape change using remote sensing techniques: an explanatory tool for agricultural transformation in Greek rural areas. *Landscape Urban Plann*, 86, 38–46.

Chapter 7: Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: The case of Athens

Authors: Dimitrios Gounaridis, Ioannis Chorianopoulos, Sotirios Koukoulas

Journal: Applied Geography 90, (2018), 134–144.

Abstract

Simulation modeling along with scenario analysis is a useful tool for urban planning, by providing an appraisement of different alternatives and tradeoffs and thus contributing to improved decision making. The objective of this study is to explore potential future urban dynamics in the Messoghia plain, (peri-urban Athens, Greece) under four scenarios that reflect future growth traits in the area related to different economic performance realities and alternative policies. Messoghia, a predominantly rural area, experienced significant and unregulated urban growth, during the past decades, due to the construction of the international airport in the area, the significant allocation of funds triggered by 2004 Olympics and the absence of planning controls. However, the late economic circumstances significantly affected the growth trends in the area. First, the paper looks at the periodic changes occurred during the past three decades (1980–2015) employing remote sensing techniques and Landsat data. The observed changes are then combined with 20 dynamic, biophysical, socio-economic and legislative factors, to produce transition potential maps using the Random Forests algorithm. Scenarios are projected until 2045 by implementing a spatially explicit Cellular Automata model. Under an economically optimistic scenario which means high or medium development circumstances, and given the absence of an adequate controlling mechanism, the artificial surfaces are expected to nearly double in size, by 2045. In case of a continuation of economic scarcity which can be translated in low or very low development, the artificial surfaces are expected to increase by 9% or 6% respectively, by 2045.

Keywords: Urban growth, Random forests, Cellular automata, Simulation, Development scenarios, Land use planning

5.1 Introduction

Changes in land cover and land use are among the most important human made alterations on earth, reflecting a wide range of interactions between society and the environment (Turner II, Lambin & Reenberg, 2007). The rapid transformation of land into artificial surfaces, has rightfully attracted the attention of scholars, planners and policy makers, concerned with the negative environmental implications it entails (Johnson, 2001). Research on the environmental impact of unregulated urban expansion centres on a number of issues, ranging from soil sealing, ecosystem fragmentation and the increased consumption of rural and natural land (Hasse & Lathrop, 2003; Jongman, 2002; Milesi, 2003), to broader concerns regarding the demotion of "urban sustainability" goals (Wilson & Chakraborty, 2013).

In view of these consequences, emphasis in the literature is placed on the capacity of land-use planning to influence the form, degree and direction of urban growth tendencies. Key in the fruitfulness of such efforts - aptly termed "smart growth" policy initiatives (EEA, 2006) - is the availability of two distinct types of information. First, an appraisal of urban growth trends, encompassing the pivotal factors that influence urban expansion. Second, an estimation of the impact of particular spatial planning choices on future land cover patterns (Xiang & Clarke, 2003). In the absence of such insights, spatial planning is insufficiently informed to adequately intervene and regulate urban growth pressures, risking the emergence of sprawl type phenomena (Chorianopoulos, Pagonis, Koukoulas & Drymoniti, 2010).

Spatially explicit modeling, constitutes a useful tool for conducting computational experiments that quantify the importance of various driving forces of change, contributing to an enhanced understanding of such a complex phenomenon (Veldkamp & Lambin, 2001). Modeling of growth dynamics is meaningful when adopting a two-phase approach. To start with, the history of the place explored has to be comprehensively looked at, part of an attempt to identify the key socio-spatial variables influencing the traits and direction of urban growth. Subsequently, by quantifying the driving forces of local change, a model can be build, capable of predicting possible future growth trajectories in the area under certain scenarios.

Scenario-based analysis has emerged in order to explore variations for a limited, but consistent, set of model parameters, delineating feasible future development trends

under a set of pre-defined conditions (Feng & Liu, 2016; Murray-Rust, Rieser, Robinson, Milicic & Rounsevell, 2013). The key step in such an attempt is the creation of the so-called transition potential maps, an exercise that draws from an area's recorded trends and performances, to indicate the degree of potential change in the future. This process is based on the change detection outputs of the historical land cover and the quantification of the ways in which the respective driving forces contributed to such changes (Kolb, Mas & Galicia, 2013).

While, by definition, models cannot replicate complex interactions and nonlinear socio-economic relations, spatial simulation approaches are increasingly being adopted and used. Advancements in geo-informatics as well as in computer capacity triggered the proliferation of modeling techniques (Berling-Wolff & Wu, 2004), the availability of geographic datasets and the methodological achievements in data processing and change detection (Tewkesbury, Comber, Tate, Lamb, & Fisher, 2015). Various approaches have been adopted to model the dynamics of the built-up environment and to explore future scenarios, using regression modeling (Feng, Liu, Chen, & Liu, 2016, Poelmans & Van Rompaey, 2010), agent based modeling (Batty, Xie, & Zhao, 2007), markov chains (Ku, 2016), system dynamics (He, Okada, Zhang, Shia, & Zhang, 2006; Zheng et al. 2012), and cellular automata (CA) (Lagarias, 2012; Vliet, White, & Dragicevic, 2009).

Accounting for the increased interest in CA applications (Sante et al. 2010), the literature stresses the capacity of the respective approach to represent stochastic, non-linear processes in a conceptually simple way (Batty et al, 1997). Additionally, CA are spatially-explicit and application-oriented and therefore fully consistent with Geographic Information Systems (GIS) and remote sensing (Feng, 2017, Liu & Feng, 2016; Feng, Yang, Hong, & Cui, 2016). Another important advantage of the CA approach is their incorporation in a plethora of modeling frameworks and platforms (Aburas, Ho, Ramli & Ash'aari, 2016). Examples include, among others, SLEUTH (Clarke, Gaydos, & Hoppen ,1997), Environment Explorer (Engelen, White, & de Nijs, 2003), the MOLAND (Lavalle et al., 2004), IDRISI's CA_MARKOV (Paegelow & Camacho Olmedo, 2005), iCity (Stevens, Dragicevic, & Rothley, 2007) and Dinamica EGO (Soares-Filho, Pennachin, & Cerqueria, 2002). However, some of these frameworks and platforms are restricted to certain methodologies, steady schemes and fixed parameters. Instead, Dinamica EGO is a flexible open platform where modelers

are able to elaborate and to extend the methodological procedures according to their specific needs. Recently, Mas et al. (2014) in a comparative assessment of four modeling frameworks, outlined two key advantages of Dinamica EGO that are crucial for simulating growth dynamics (section 2.2.6).

In this paper we employ and elaborate this particular methodology to explore a challenging example of a dynamically growing peri-urban area. The case in point is the Messoghia plain in Athens, Greece. Since the early 1980s and for the following two decades, Athenian urban growth was channeled towards Messoghia, initiating a rural to urban transformation process with marked environmental implications. During this time, land use planning controls were not in place, as unregulated built-up expansion was approached as a shortcut to economic growth (Chorianopoulos et al., 2014). Environmental deterioration, however, triggered a belated planning response (2003), aiming to curb emergent sprawling tendencies. Built-up expansion has also been affected by the insolvency crisis the country is facing since 2010. The Messoghia plain, therefore, is an area that faces strong development pressures that have only been temporarily weakened as a result of extreme economic circumstances. In fact, the area's strong development potential is underscored and encouraged in the city's revised Master Plan (2014), shaping a pro-growth policy trajectory for the forthcoming decade (GGN, 2014).

In this paper we argue that estimating the future growth traits in the area in light of different economic performance realities and land-use planning contexts and choices, is a prerequisite in any attempt to address the undesirable consequences of unregulated urban expansion. From this perspective, we attempt to delineate the future growth dynamics in Messoghia under four different economic and spatial policy scenarios and to illustrate accordingly the respective the urban scenery in the medium (10 years) and the long (30 years) term. With regard to key concepts of reference, urban land is defined in respect to all human-constructed elements, such as continuous or discontinuous residential areas (hereafter called urban fabric) and industrial, commercial, infrastructure and transport units (hereafter called artificial non-urban areas).

The paper is organized in four parts. In the first part, we look at the case study area, contextualizing the research exercise. In the second part, we outline the research

methodology followed for imagery processing, classification and change detection along with the 20 factors description and processing, including the calculation of the Leap-frog development index. Accordingly, we present the sequence of methodologies for the transition potential modeling, the scenarios development, the model calibration and the projection of results to the future. In the third part of the paper, we illustrate and discuss in detail the results obtained. In the concluding section, we revisit the area's prospective futures, highlighting the relevance or our approach to the quest for effective planning responses and sustainable urban development trajectories.

5.2 Material and Methods

5.2.1 Study site

The Messoghia plain is located eastwards of the Athens conurbation (Figure 1). Until the early 1980s the area retained at large a rural character, escaping the rapid urbanization wave that transformed Athens in the postwar years. The main reason for this particularity is the Hymettus Mountain, a physical barrier separating the plain from the city that obstructed accessibility and delayed the development of an adequate transportation network. Since the mid-1980s, however, the area displayed notable urban expansion signs, associated in the literature with the sprawling tendencies of Athens; a congested city with rapidly deteriorating environmental conditions (Leontidou, et al., 2007). In the succeeding decades, change in Messoghia was swift and multifaceted, a turn of events bringing to the foreground the fundamental antagonism between economic growth preoccupations and regulated urban expansion goals. The key developments that altered the area's features are hereafter discussed and categorized for methodological purposes in four distinct chronological frames of reference.

- 1985-1995: In the mid-1980s, and in light of urban expansion tendencies noted in the region, the planning authorities decided to intervene. The introduction of the Master Plan of Athens (1985) was expected to guide urban growth in the region via the launch of detailed land-use plans. In the case of Messoghia, the Master Plan was geared towards the protection of the area's rural character. This, however, did not happen. In a parallel trajectory, the decision to relocate the Athens international airport in Messoghia was taken, cancelling de facto the

respective Master Plan directions. Consequently, the implementation of the spatial planning framework for Messoghia was indefinitely postponed, enabling the unobstructed continuation of the sprawling tendencies noted in the area.

- 1995-2006: In the mid-1990s, the national authorities put Athens forward as a candidate city for hosting the 2004 Olympic Games. The bid was successful (1997), and a number of large scale physical infrastructure projects were expeditiously initiated in Messoghia, including Olympics' related venues (Equestrian Centre, Shooting Centre) and transportation networks of metropolitan importance (ring road, suburban railway). Investment in transportation infrastructure, in particular, enhanced Messoghia's accessibility, triggering a population influx and a concurrent increase in urban land-uses. The inflexible deadline of the 2004 Olympics resulted in the prioritization of development planning goals over the spatial planning ones (Souliotis, 2013). The land-use zoning scheme that was supposed to guide growth in Messoghia, for instance, was put into force as late as in 2003. In the intervening period, rural to urban transformation in the region proceeded apace.

- 2006-2010: The post-Olympics era is characterized by the relative soundness of the economy, displaying annual growth rates that exceeded, on average, three per cent of the GDP (Bank of Greece, 2014). Alongside, the long-awaited land-use planning scheme for Messoghia was finally in place. This period of stability, however, was cut short by the impact of the global financial crisis (2008) on the state of public finances.

- 2010-2015: In the wake of the global financial crisis, the general government deficit and the public debt stretched respectively to 15.4 and 126.8 percent of Gross Domestic Product (GDP). As the state practically lost access to the international financial markets, and in order to avoid a solvency crisis, the government agreed a series of loans with the European Commission, the European Central Bank and the IMF. The loans were conditional upon Greece implementing an adjustment programme including, amongst others, the introduction of steep austerity measures and the privatization of state owned assets (Eurogroup, 2015). As a result, between 2008 and 2015 the economy lost a cumulative 27 percent of its GDP (Bank of Greece, 2016). The relevance of these developments for Messoghia are twofold. Fist, by reason of the economic

depression, transactions in the real estate market in the region fell by 78 percent, grinding the built-up expansion trend to a halt (Municipality of Athens, 2014). Second, the ownership and, hence, the fate of key state owned real estate properties in Messoghia (land and structures) was transferred to a privatization fund, the sales of which aim to reduce the government's debt burden. The future usage of these sites in not regulated by the area's planning framework, arresting in practice the effectiveness of the respective land use planning scheme (Pagonis & Chorianopoulos, 2015).

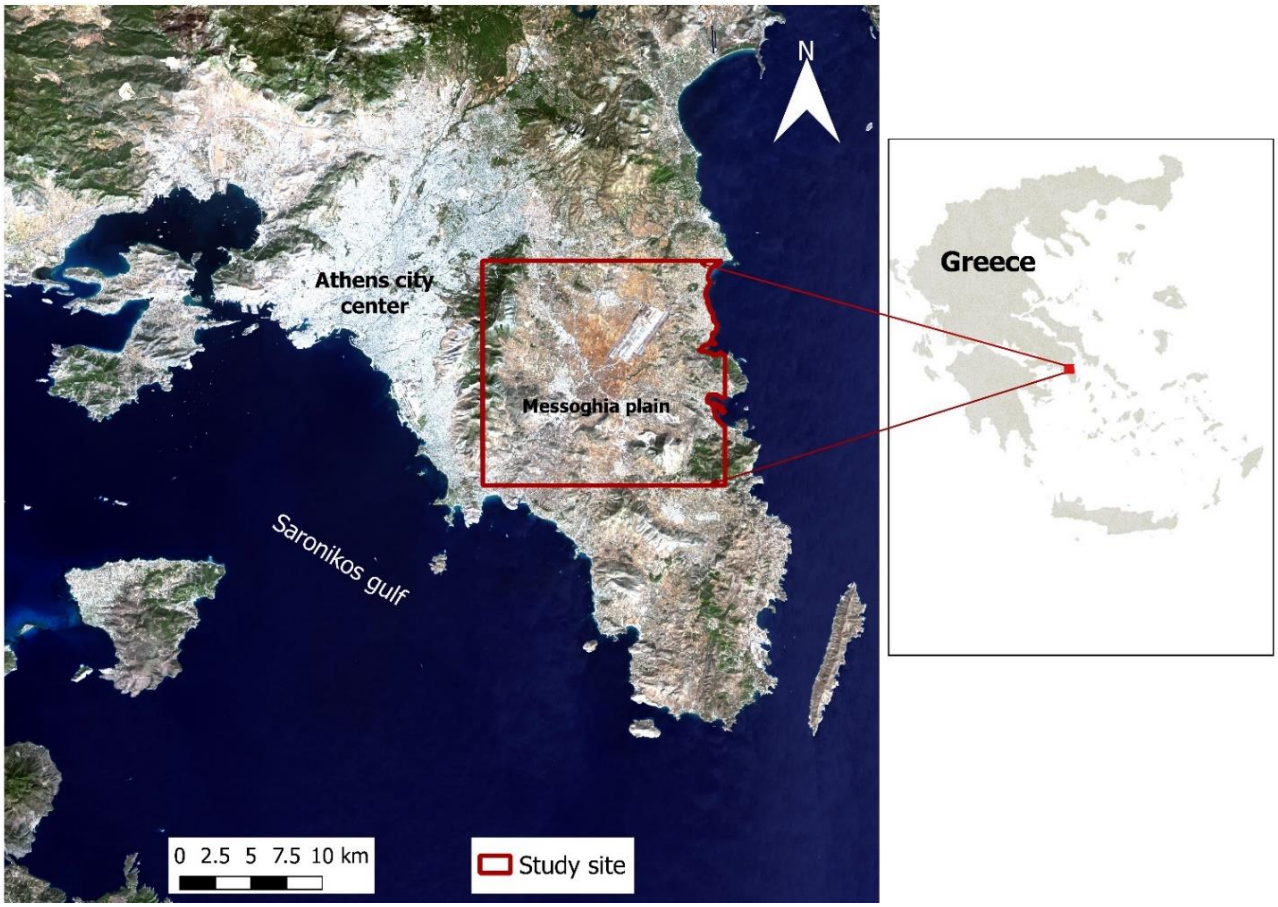


Figure 1. Location of the Messoghia plain (background from Landsat 8 OLI, 6 May 2015, path/row: 183/34, Bands: 4;3;2).

These four chronological frames marked with uneven development, different economic performance realities and land-use planning, will consist our basis for the scenario analysis in our attempt to explore potential future growth dynamics in Messoghia.

5.2.2 Approach

5.2.2.1 Images pre-processing and classification

We employed seven Landsat images (Table 1) spanning 30 years (1985-2015). Four images (1990; 2000; 2006 and 2010) were part of the Global Land Survey (GLS) datasets which are a collection of images that meet high quality and minimum cloud cover standards (Gutman, Huang, Chander, Noojipady, & Masek, 2013). The rest of the images (1985; 1995 and 2015) also meet the quality standards and had no cloudiness in the study area. Ideally all images should be acquired the same month and preferably during summer where phenological variations are less evident. However due to availability of images that meet certain standards, and the scope of the study (artificial surfaces) we use images acquired from May to August.

To avoid any discrepancies due to the multi-temporal and double-sensor type of analysis, and to compute spectral indices (Song, Woodcock, Seto, Pax Lenney & Macomber, 2001), all images underwent radiometric and atmospheric correction. We first converted the DN numbers into top of atmosphere reflectance using the dark-object subtraction method introduced by Chavez (1988). To obtain surface reflectance and achieve data normalization, we applied the 6S model introduced by Vermote, Tanré, Deuzé, Herman, & Morcrette (1997).

Table 1. Characteristics of the Landsat satellite images

Date	Satelite	Sensor	Path	Row	Resolution
19/5/1985	Landsat 5	Thematic Mapper (TM)	183	34	30
14/8/1990	Landsat 5	Thematic Mapper (TM)	182	34	30
11/7/1995	Landsat 5	Thematic Mapper (TM)	182	34	30
30/6/2000	Landsat 7	Enhanced Thematic Mapper + (ETM+)	182	34	30
30/5/2006	Landsat 7	Enhanced Thematic Mapper + (ETM+)	182	34	30
12/8/2010	Landsat 5	Thematic Mapper (TM)	183	34	30
6/5/2015	Landsat 8	Operational Land Imager (OLI)	183	34	30

We classified all images into three categories, implementing the RF classification algorithm through the RandomForest package available in R (Liaw & Wiener, 2002). The three categories were i) urban fabric, ii) artificial non-urban and iii) Other land

cover/uses (cultivated land, vegetation, bare land). To train the model, a set of randomly distributed points (n = 650) was plotted against the Landsat images and very high-resolution images available via Google Earth. Category values assigned by visual interpretation. All points close to the boundaries of adjacent categories were relocated, ensuring that clear samples of each category were taken and thus eliminating any source of confusion to the model. As predictor variables, besides the 6 reflective Landsat bands (bands 1–5 & 7 for Landsat 5 TM and Landsat 7 ETM+, bands 2-7 for Landsat 8 OLI), we used the first layer produced by principal components analysis (PCA) separately for the three visible bands (1, 2 and 3) and the infrared bands (5 and 7), as it appears to increase classification accuracy (Gounaridis, Zaimes & Koukoulas, 2014; Gounaridis, Apostolou & Koukoulas, 2015). In addition, we included the normalized difference built-up index (NDBI) (Zha, Gao, & Ni, 2003) and the enhanced built-up and bareness index (EBBI) (As-Syakur, Adnyana, Arthana, & Nuarsa, 2012).

Since RF requires two primary parameters to be specified by the user being (i) the number of predictor variables randomly sampled at each decision tree split and (ii) the number of classification trees to be built, we used three (3) predictor variables for each tree split, which is equal to the square root of the total number of predictor variables and 500 trees for each run. Last, to sidestep the so called 'salt n pepper effect' we removed the isolated patches (area less than 0.1 ha), by replacing their category value with the mode of their neighborhood pixels, defined by a 3x3 window (Gounaridis et al., 2014; 2015).

To assess the accuracy of each classified map, a group of 200 to 350 random points were distributed and values assigned via visual interpretation. Overall accuracy for all images ranged from 90% to 93%, while the two categories of focus had omission and commission errors that ranged from 88% to 93% (Table 2).

Table 2. Error matrix - Resulting map per year against reference samples (20% of initial samples).

Resulting maps																												
	1985				1990				1995				2000				2006				2010				2015			
Reference	1	2	3	O.E (%)	1	2	3	O.E (%)	1	2	3	O.E (%)	1	2	3	O.E (%)	1	2	3	O.E (%)	1	2	3	O.E (%)	1	2	3	O.E (%
1	**77**	2	5	91,7	**76**	2	7	89,4	**77**	2	3	93,9	**79**	2	6	90,8	**86**	1	5	93,5	**88**	4	4	91,7	**104**	3	6	92,0
2	3	**53**	3	89,8	1	**57**	7	87,7	2	**54**	5	88,5	2	**77**	8	88,5	7	**72**	2	88,9	5	**80**	3	90,91	5	**80**	3	90,9
3	5	4	**90**	90,9	1	6	**106**	93,8	2	4	**113**	95,0	2	3	**112**	95,7	3	6	**117**	92,9	3	3	**120**	95,2	2	5	**121**	94,5
C.E	90,6	89,8	91,8		97,4	87,7	88,3		95,1	90,0	93,4		95,2	93,9	88,9		89,6	91,1	94,4		91,7	92,0	94,5		93,7	90,9	93,1	
O.A	90,9				90,9				93,1				92,1				92,0				92,9				92,7			

O.E: Omission Error; C.E: Commission Error; O.A: Overall Accuracy

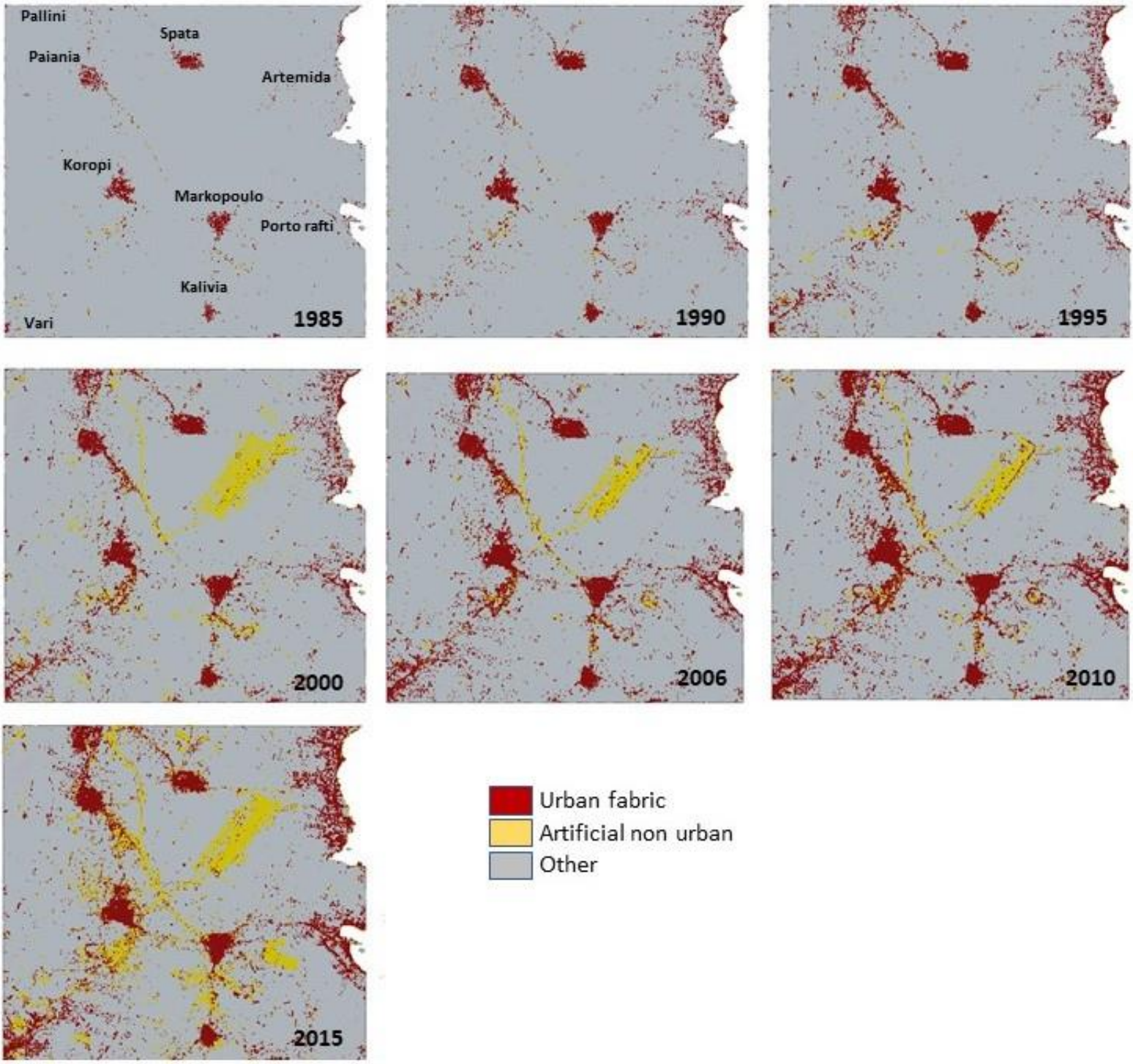


Figure 2. Resulting maps after classifying the nine Landsat images.

5.2.2.2 Predictor variables

Given that human decisions and actions drive the growth dynamics, we aimed at incorporating, in the model, factors and processes that explain peoples' choices about residential and infrastructure location. Social shifts, economic motives, inherent quality and attractiveness of a given place, the effects of neighboring areas and proximity to basic needs were assumed to play a key role. Taking into account previous efforts and data availability and accessibility we concluded to a suite of 20 predictor variables (Table 3). We hypothesized that these biophysical, socioeconomic, legislative and land use factors can spherically explain the changes occurred during the last decades in Messoghia.

Territorial variables such as elevation, slope and aspect influence the quality of a certain location. Proximity to the sea as well as to areas of high nature value are also an adding value in pursuit of a better quality of life and aesthetics for both primary or second-homes (Leontidou et al. 2007). City center of Athens as well as of nearest towns consist important poles for markets, services and jobs (Aburas et al. 2016). Distance to transportation is strongly related to urban dynamics, both acting as a cause and a consequence of urban expansion, while access to health provision centers is considered a prerequisite for many people. Demography and socio-economic variables such as changes in population density, employment and unemployment rates reflect the shifts in population dynamics and the labour market that shape the socio-economic profile of the area. Regarding land use, livestock, agriculture and building rates delineate the socio-economic restructuring occurred during the last decades. Additionally, business activity in Messoghia, in the last 25 years was recorded by the respective local champers of the area (Chorianopoulos, Tsilimigkas, Koukoulas & Balatsos, 2014). Details of all enterprises established throughout this period were obtained by the Athens Chamber of Commerce and Industry and geocoded using BatchGeo (https://www.batchgeo.com/). Finally, the Urban Development Control Areas zoning scheme is the only available planning mechanism for regulating the urban expansion (Chorianopoulos et al. 2010).

Variables available at the prefecture level were not included, since the municipalities in Greece are responsible for local land management and thus representing a meaningful spatial unit for this type of analysis (Salvati, Mavrakis, Serra & CarLULCci, 2015). All census data were then collated in a GIS environment at the

municipality level while all distances computed using the Euclidean distance function. All variables were converted to raster format and resampled at 30m spatial resolution.

Table 3. List of predictors used in the transition potential modeling process.

Acronym	Variable	Discription	Source	Time interval
Territorial variables				
DEM	Elevation	Elevation in m	GLSDEM[1]	(-)
SL	Slope	Slope in degrees	GLSDEM	(-)
AS	Aspect	Aspect in degrees	GLSDEM	(-)
DS	Distance from the sea	Euclidean distance from the shoreline in m		(-)
Socio-economic variables				
DATH	Distance from Athens	Euclidean distance from the centre of Athens in m	OSM[2]	(-)
DT	Distance from nearest town	Euclidean distance from the center of the nearest town (Markopoulo, Paiania, Koropi, Keratea, Artemida) in m	OSM	(-)
DAIR	Distance from airport	Euclidean distance from the airport in m	OSM	(-)
DPH	Distance from public health	Euclidean distance from public hospitals and other public health care units in m	OSM	(-)
DPT	Distance from public transport	Euclidean distance from public transport stops (bus, metro, suburban train) in m	OSM	(-)
DR	Distance from road network	Euclidean distance from road network in m	OSM	(-)
POP	Demographics	Changes in population density at the municipality level	ELSTAT[3]	1991-2011
EMP	Employment rate	Difference between: Total number of employed persons per total population at the municipality level	ELSTAT	1991-2001
UNEMP	Unemployment rate	Difference between: Total number of unemployed persons per total population at the municipality level	ELSTAT	1991-2001
Legislative				
UDCA	UDCA zones	Urban Development Control Areas (UDCAs) zoning scheme		2003
Land use				
DGU	Distance from green urban areas	Euclidean distance from green urban patches in m	UA[4]	2006
LIV	Livestock rate	Difference between: Total number of animals (cuttle, pigs, goats, poultry, rabbits) per municipality total area	ELSTAT	1991-2000
AGR	Agriculture rate	Difference between: Total area devoted for agriculture per municipality total area	ELSTAT	1991-2000
BU	Built-up rate	Cumulative total number of new houses built per municipality total area	ELSTAT	2000-2008
ENT	Enterprises rate	Cumulative total number of new enterprises registered to ACCI per municipality total area	ACCI[5]	1985-2010
DN	Distance from nature	Euclidean distance from forested patches, areas of high nature value and protected areas in m	OSM	(-)

[1] Global Land Survey Digital Elevation Model (GLSDEM) http://glcf.umd.edu/data/glsdem/
[2] OpenStreetMap https://www.openstreetmap.org/
[3] Hellenic Statistical Authority http://www.statistics.gr/
[4] Urban Atlas provided by European Environmental Agency http://www.eea.europa.eu/data-and-maps/data/urban-atlas
[5] Athens Chamber of Commerce and Industry http://www.acci.gr/acci/Home/tabid/28/language/el-GR/Default.aspx

4.2.3 Leap-frog development index

Based on the 1985 and 2010 land cover maps, we calculated the Leap-frog development index, originally proposed by Xu et al. (2007). The index is calculated by dividing the length of the common boundaries of newly developed artificial patches (in our case patches appeared in 2010) with already existing artificial patches (1985) with the perimeter of the newly developed patches. When the result is more than 0.5 the growth type is identified as infilling, when the result is less than 0.5 it is identified as edge growth while when the result is 0, which mean that there is not shared boundary, the growth is identified as Leap-frog development.

Initially, both land cover maps converted to vector format. Artificial areas of 1985 along with artificial areas of 2010 were moved in a new vector layer after being assigned a different value. Subsequently, common boundaries length and perimeter were calculated using GIS functions. The index was then calculated and the file converted to a raster format at 30m spatial resolution.

4.2.4 Transition potential modeling

Probability maps in Dinamica EGO, are usually computed using the weights of evidence method which is based on the Bayes theorem of conditional probability (Bonham-Carter, 1994). In this method, it is necessary to select independent explanatory variables in order to avoid collinearity issues, and this could be considered a limitation. Another possible limitation of this method is that continuous variables have to be transformed into categorical.

Recently, Kamusoko & Gamba, (2015) combined CA with the Random Forest (RF) algorithm for simulating urban growth. To test the effectiveness of their approach, they compared the performance of RF, support vector machines and logistic regression for producing transition potential maps. The RF model outperformed the two others. The only drawback reported is that in general all models failed to detect the so-called leap-frog development.

RF is a robust, non-parametric machine learning algorithm introduced by Breiman (2001) and has certain advantages when coping with complex systems such as

built-up growth dynamics: i) RF can efficiently handle both categorical and continuous variables facilitating the incorporation of any type of inputs in terms of data type and scaling of values. Thus, data from multiple sources like remote sensing or census can be incorporated in the model (Gounaridis et al. 2015; Gounaridis & Koukoulas, 2016). ii) RF is insensitive to overfitting, to collinearity issues as well as to noise and outliers (Chan and Paelinckx, 2008). iii) Normal distribution of inputs is not a prerequisite. iv) The algorithm performs well when coping with non-linear relationships between response and predictor variables (Kamusoko & Gamba, 2015). v) The importance of each predictor variable is computed using several metrics, allowing the user to determine whether a variable will be incorporated in the model or not (Gounaridis & Koukoulas, 2016).

All predictors were included in a single stack and served as independent variables. To overcome the limitation of failure in detecting leap-frog development, which is widely evident in Messoghia (Chorianopoulos et al. 2010), we incorporated in the model the Leap-frog development index. The initial maps resulted from classification had 3 categories, urban fabric, artificial non-urban and other (Figure 2). To train the model, we used 2500 randomly placed points. Values assigned as to indicate change (from Other to Artificial) and no change. The regression version of RF (Breiman, 2001) was then implemented in R using the RandomForest package (Liaw & Wiener, 2002). To fine tune the RF, we used five (5) predictor variables (equal to the square root of the total number of predictor variables) for each tree split and 700 trees for each run.

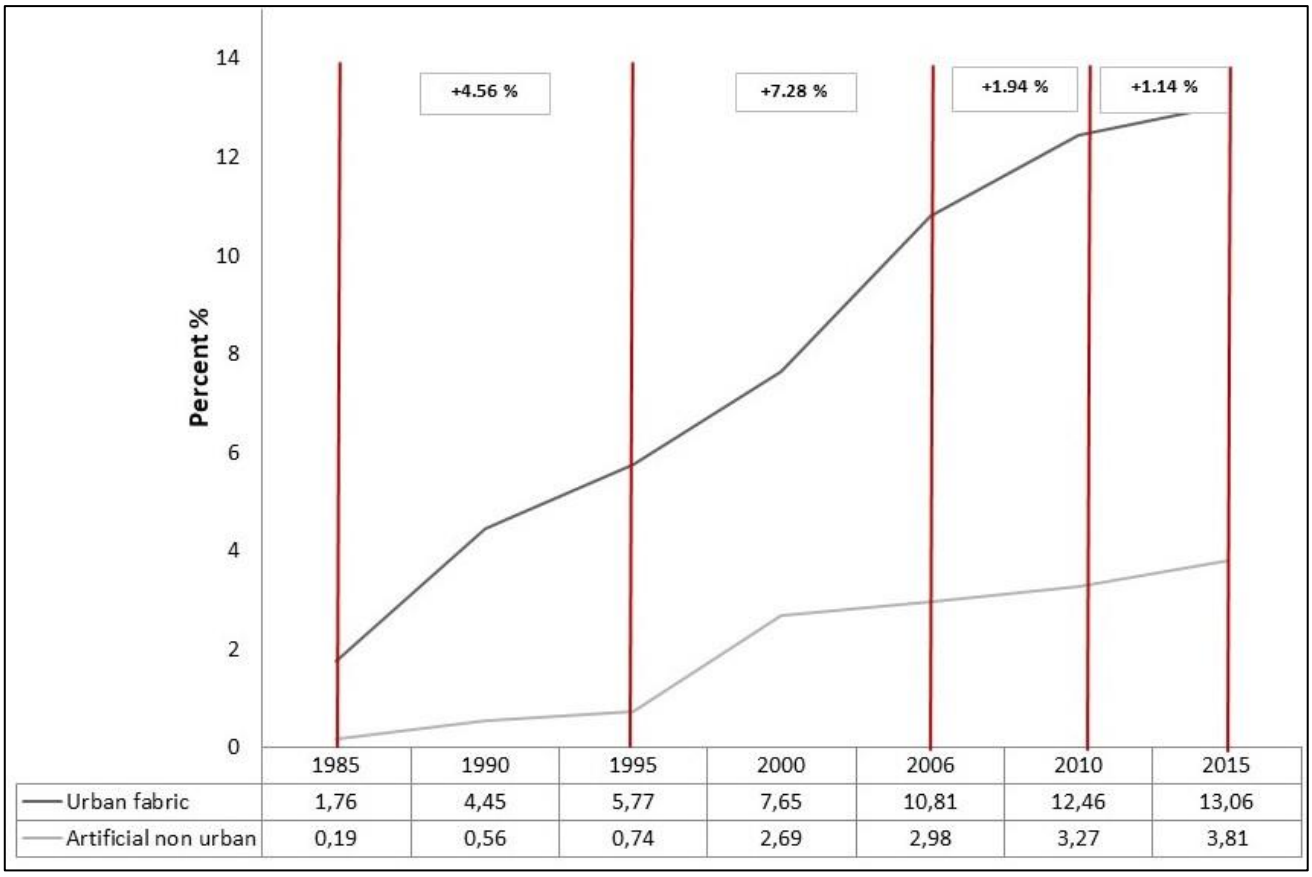
5.2.5 Scenarios

Our aim was to project future changes under four scenarios sketching out different economic development policies and options. Thus, our scenarios are based on the observed historical trends during the last three decades. Looking at the resulting land cover maps, it is evident that the built-up dynamics in (Messoghia were dramatically uneven, reflecting different phases of economic development (Figure 3). Thus, we delineated four scenarios of projected residential development, distinguished by different levels of development. The scenarios include:

Medium development reflects the period 1985-1995 and *High development* reflects the period 1995-2006. In these scenarios (Messoghia are re-brought to the fore and into the development path, experiencing an exponential rise in economic sectors and industries following the trends observed during the past decades. Planning and spatial policies contribute towards this direction, allocating funds into the area and promoting the socio-economic restructuring. Consequently, infrastructure, firm headquarters, enterprises and shopping centers colonize the area, leading to an economic polarization and economic functions re-concentration. Extensive regeneration of the waterfront also takes place, channeling changes in real estate dynamics towards tourism-specialized settlements and second homes. Last but not least, a steady population increase as a result of Athenian de-concentration, rural depopulation and external migration because of the job opportunities and better quality of life, increases the housing demand and in turn the housing construction.

The *Low development* scenario reflects the period 2006-2010 while the *Very low development* reflects the period 2010-2015 and keeps the development pace very low as a consequence of economic scarcity and lack of investments. The economic functions as well as the population flows remain stable following a low to very low pace. An amount of already built residences, intended to meet the needs for both second or primary housing, remain uninhabited (unsold or unfinished) while many already constructed industrial and commercial facilities remain unexploited. At the same time the demand, shaped by economic scarcity and population low rates, is lower than the already built and available buildings leading to a low to very low building rate in the area.

In these scenarios, we assume that profound social and political changes will not occur, accessibility in the area will remain stable and new roads and railway links will not be constructed. As far as land use regulations and legislative frameworks are concerned, we assume that especially in the first two scenarios, will continue to be loose in the face of development potential.



	1985	1990	1995	2000	2006	2010	2015
Urban fabric	1,76	4,45	5,77	7,65	10,81	12,46	13,06
Artificial non urban	0,19	0,56	0,74	2,69	2,98	3,27	3,81

Figure 3. Summary of statistics based on the nine classified maps and relative percentages of artificial areas. The observed trends reflect four different levels of development during the last 30 years in Messoghia.

5.2.6 Model calibration and simulation

We used the Dinamica EGO platform in order to develop a CA model able to simulate the changes occurred in Messoghia. We opted to use this platform because of certain advantages: Firstly, Dinamica EGO incorporates two complimentary, but distinct, allocation functions, i) the Patcher, which generates new patches based on a seeding mechanism and ii) the Expander, which expands previously formed patches. Additionally, the user is able to manually set parameters like patch size variance, isometry of transitions and mean patch size. These advantages of the platform allow the user to directly intervene and calibrate the model in a way that complies with the reality of a specific area. Secondly, Dinamica EGO allows multi-resolution accuracy assessment employing a fuzzy similarity index (Hagen, 2003). By employing this

methodology, accuracy of simulated versus observed patterns is not restricted to a strict cell by cell overlay but gradually considers the cells in the neighborhood.

To calibrate the model and evaluate the goodness of fit, a comparison of simulated maps with a reference/observed maps is the most efficient way. Therefore, we trained the model based on the 1985-2010 period, simulated these changes up to 2015 and compared the simulated result with the 2015 classified land cover map. To do so, we calculated the annual rates of change between 1985 and 2010 using the transition matrix function (Soares-Filho et al. 2002). Next, we adjusted the mean and the variance of new patch sizes and the patch isometry in order to replicate the actual conditions of the area in terms of structure and composition. 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 (equal) the newly developed patches. We calibrated the CA model by computing the mean patch size and mean patch variance of the input land cover map (2010) and adjusted the isometry through trial and error. We set the model to run at a 5-year time step from 1985 to 2015. To evaluate the model's performance, we compared the simulated land cover map of 2015 to the observed land cove map of 2015 using the fuzzy similarity index at multiple resolutions. Finally, taking 2015 as the initial year and 2045 as the final year, we simulated the land-use changes for Messoghia in a 5-year time step under the four scenarios

5.3 Results and discussion

5.3.1 Historical land cover change

Results generated by the classification of the nine Landsat images and the subsequent aggregation of classes are depicted in Figure 2. During the 80s the built-up scenery of Messoghia was characterized by small towns (Paiania, Spata, Markopoulo, Kalivia and Koropi) and a few sea-side resorts while during the 90s and before the airport, the Olympic venues and the transport infrastructure as well as the effects of a loose regulatory framework are evident. Especially in the north-western (Palini) and south-western (Vari) parts of the area, which are closer to Athens urban conurbation, changes in the form of sprawl can be observed. At the same time the small towns show

a tendency to infill and expand while the sea-side residences (Porto rafti and Artemida) start to increase. These trajectories during the 90s are attributed to the gains in population of the area, as a result of the decentralization trends of the Athenians seeking better quality of life and the waves of internal rural-urban immigrants. The figures after 2000, clearly portray the effects of the Olympics-related large-scale projects and the infrastructural investments in Messoghia. Especially the effects of the airport and the road expansion are more than evident. All types of sprawl (namely suburban growth, leap-frog development, strip development and scattered development), in both urban fabric and artificial non-urban classes, can be observed in the area after 2000. Quantification of the results (Figure 3) reveals that the period between 1985 and 1995 reflects the first boost in terms of development while the next decade 1995-2006 is the peak. After 2006 the development trends remain positive but obviously start to decline reflecting the post-Olympic era with the dramatic decrease of investments and the deterioration of economy.

5.3.2. Model calibration and performance

Using the fuzzy similarity index, we evaluated the model's performance over a range of resolutions. We found a spatial fit of 82.72 % within a 1x1 window size radius which improved to 91.72 % when widened to a 15x15 window size (Figure 4).

Figure 5 shows the result after comparing the observed and simulated land cover maps of 2015. The model was relatively accurate at predicting the allocation of urban fabric and artificial non-urban surfaces both in the form of suburban growth, strip development and scattered development. The good performance suggests that the suite of 20 predictor variables explains the observed historical changes efficiently while the RF algorithm performed well with an adequate fit. Most importantly, the model achieved to predict accurately the leap-frog development and this is mainly attributed to the incorporation of the Leap-frog development index and the extensive training of the RF model.

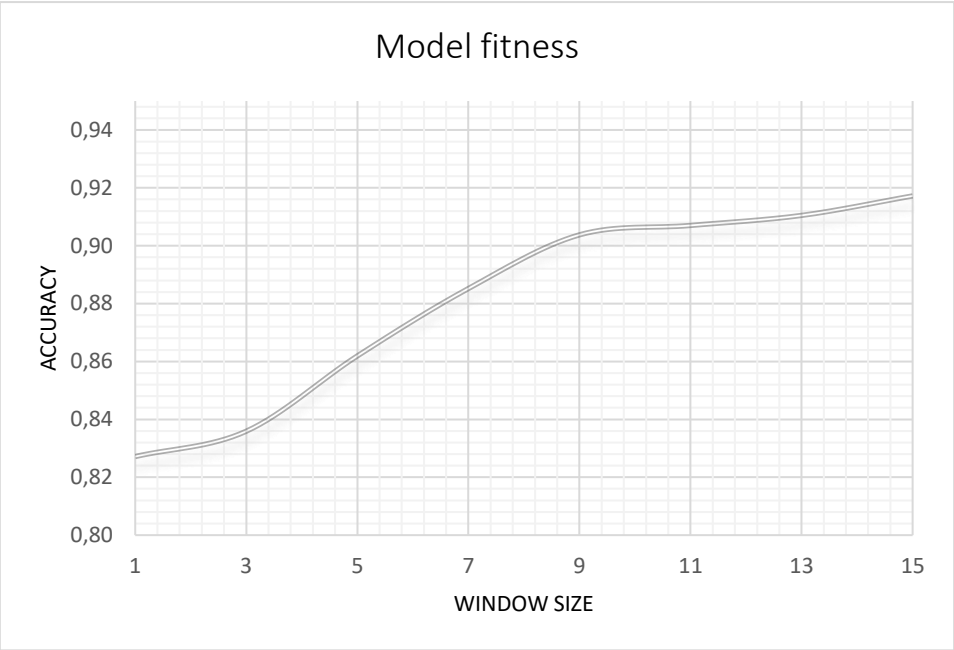


Figure 4. Results of multi-resolution spatial evaluation of model fitting using the fuzzy similarity index.

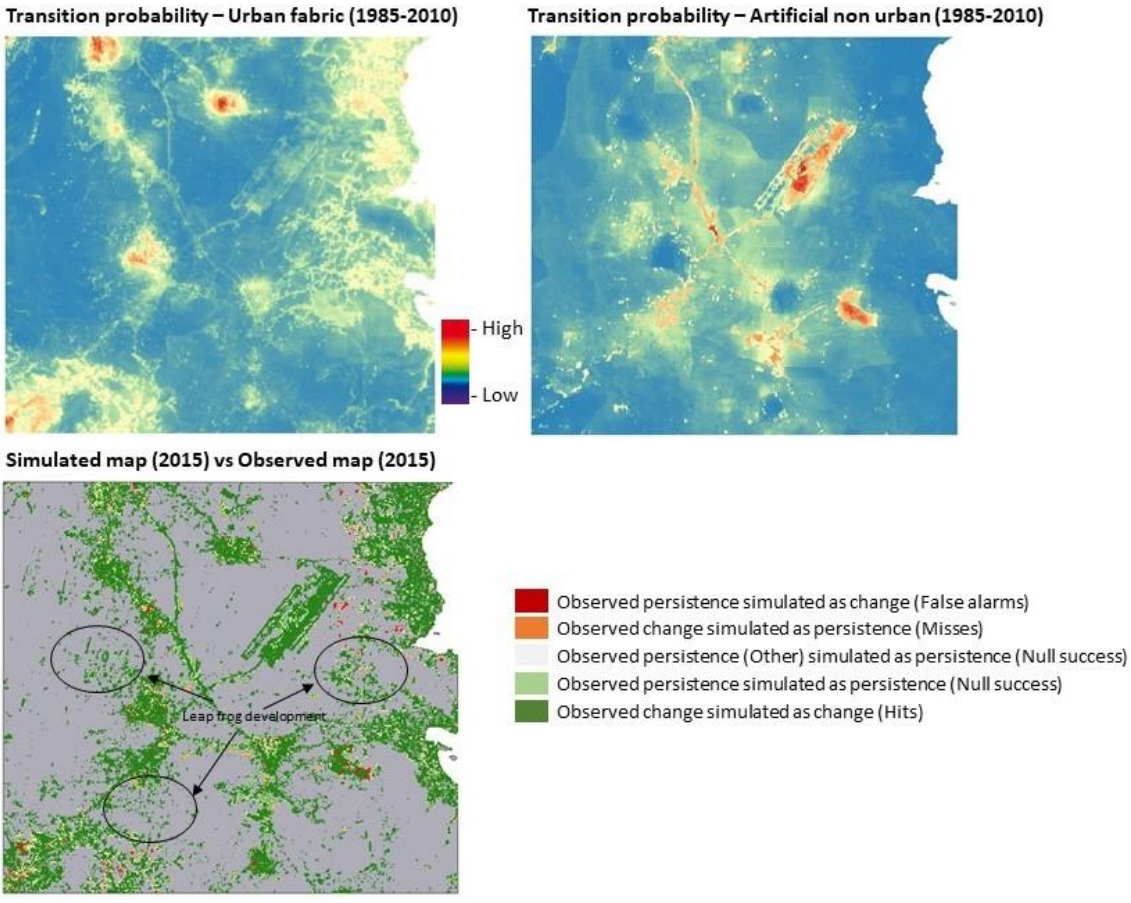


Figure 5. a) Transition potential surface obtained through the RF modeling with 20 predictors. b) Result of cross classification between the simulated vs the observed map of 2015.

5.3.3 Simulations

In this last substantive section, we demonstrate the simulated results on how Messoghia will look in the future under four scenarios related to different levels of economic development and policies. Figure 6 illustrates a prediction of the built-up expansion, over a 30-year period, generated by the CA model while Figure 7 illustrates the quantification of results in a 5-year timestep. The black line indicates a ten-year projection in respect of the stated life span of the Athens Master Plan, expected to be revised in 2025 (Hellenic Parliament, 2014: 5).

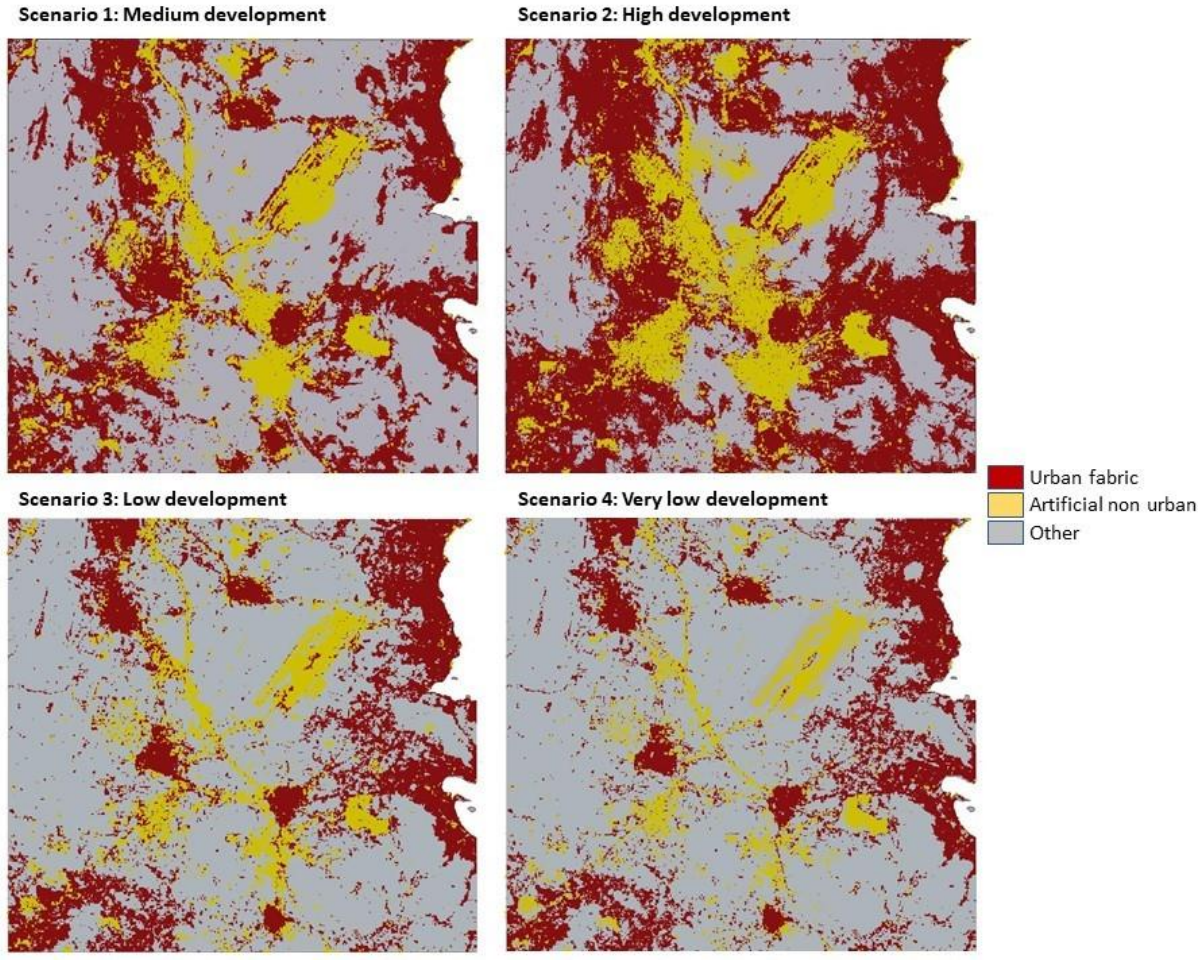


Figure 6. Artificial areas extent, simulated for 2045, under the four different development level scenarios.

Under the medium and high development scenarios the artificial surfaces expand predominantly along the transportation links. Pre-existing urban and industrial clusters

infill, expand and finally appear connected especially along the nodal points where the radial roads inter-connect the small towns (Paiania, Spata, Markopoulo, Kalivia and Koropi) and Athens. The waterfront also acts as a remarkable core of agglomeration. Seashore existing towns obviously expand and become densely infilled, servicing second or permanent housing demands and tourism related activities. Leap-frog development has also increased sharply around junctions of infilled areas and main roads, mostly in areas previously occupied by agriculture, indicating that the development pressures and the assumed permissive planning scheme allowed the high consumption of land at the expense of other less profitable land uses.

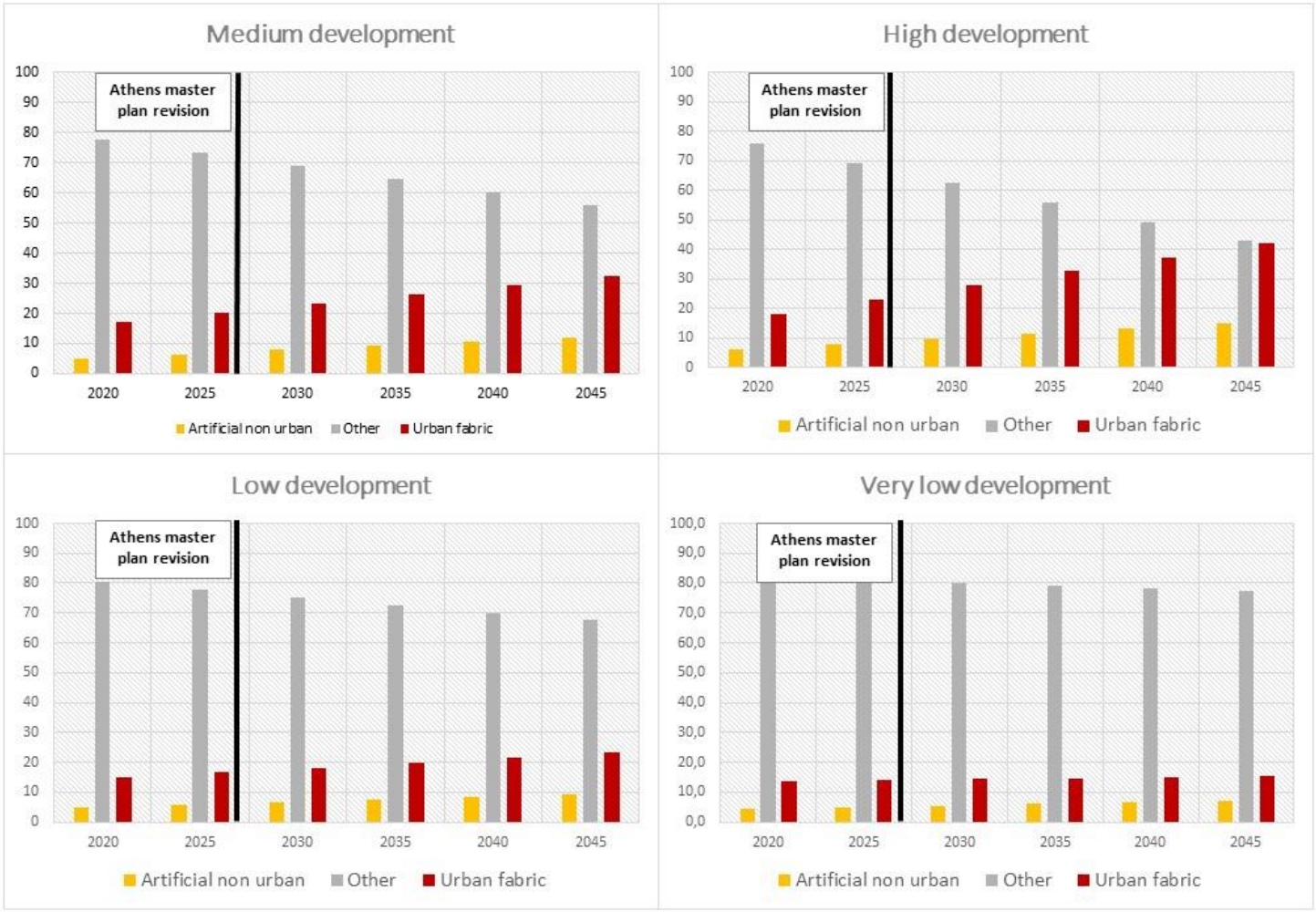


Figure 7. Rates of artificial areas simulated for 2045 in a 5 year step, under the four different development level scenarios.

The artificial surfaces are expected to increase considerably under the medium and high development scenarios (Figure 7) and with a pace of development equivalent to that of 1985-1995 and 1995-2006 periods. Under the high development scenario, the artificial surfaces are expected to occupy almost half of the total surface of the study area by 2045 while for both development scenarios the artificial surfaces are expected to nearly double in size (16.87% in 2015).

Under the low and very low scenarios, expansion of artificial surfaces is also observed but in a lower extent, less dispersed and with a considerably lower magnitude. Again, the changes observed are mostly around the road network and the waterfront. Already existing patches of artificial surfaces appear infilled rather than expanded while leap-frog development is also evident throughout the area. Notably, the leap-frog development can be observed mostly around areas characterized by favorable conditions such as proximity to the town centers (Paiania, Spata, Markopoulo, Kalivia and Koropi), proximity to Athens (around the motorway Vari-Koropi in the south-west part and around Pallini in the north west part) and proximity to the sea (seashore towns of Artemida and Markopoulo). Results indicate that these areas are the most likely to become urban in the future. The pace of development of the low and very low scenarios, follows equivalent trends of the 2006-2010 and 2010-2015 periods respectively and the total area occupied by artificial surfaces is expected to be 28.88 % for the low development scenario and 25.02 % for the very low scenario.

5.4 Conclusions

We explored potential future growth dynamics in the Messoghia plain under certain scenarios that reflect different economic development trajectories and policy options. Methodologically, coupling of CA and RF proved to be a sound way to overcome certain limitations reported in previous efforts. On the one hand, implementing the RF algorithm in order to generate transition potential surfaces allowed us to incorporate in the model spatial determinants of different nature in terms of scale and origin, sidestepping collinearity and distribution issues. The suite of 20 predictors proved to describe well the complicated issue of built-up expansion in Messoghia plain, while the incorporation of the Leap-frog index boosted the performance of the models in the face of sprawl detection and, in turn, prediction. On

the other hand, implementing CA modeling through the Dinamica EGO framework proved fully compatible with transition potential generated in a different environment and provided certain advantages. The two complimentary sub-models Patcher and Expander are fully operational with the built-up growth dynamics concept allowing the user to intervene efficiently and calibrate models according to case specific needs by taking into account actual parameters of the study area. Thus, coupling these two frameworks reduces several limitations that are commonly encountered. An important limitation in our case, one that is also commonly encountered by researchers, is data scarcity in terms of both spatial, thematic and temporal resolution.

Regarding Messoghia, the area maintained a predominantly rural character up until the early 80s, due to cumbersome accessibility and limited attraction of investments. The construction of the international airport in the area and the significant allocation of funds set off by 2004 Olympics, did transform the region. The pace of transformation, however, declined considerably during the last decade as a result of insolvency and the subsequent sovereign debt crisis. In light of this inopportune economic reality, Messoghia escaped the exhaustive consequences of unordered urban expansion, retaining noteworthy reserves of agricultural and natural land. The four simulation models illustrate how Messoghia would feature in the future under a set of different circumstances. The simulation of the pre-Olympics period, based on a sound economic growth record and the absence of an overarching spatial planning framework, is expected to yield an unprecedented development of artificial surfaces, at the expense of agricultural and natural land. Although this is a promising scenario with respect to economic growth in the short-term, it also carries significant negative environmental externalities (such as pollution, congestion, and sub-optimal land allocation), capable of undermining the very economic prospects of the area in the medium term (Cervero, 2001).

Of course, scenario modeling is meaningful when dealing with plausible and realistic projections, based on present data and trends. In that sense, the low and very low development scenarios are deemed as the ones that are closer to the current course of events. However, as economic development goals are taking precedence over virtually all other spatial planning priorities, the medium and high development scenarios cannot be dismissed as implausible. In this case, results generated from the model stress out the major impact of artificial areas expansion on natural resources,

agriculture and environmental quality indicators. From this spectrum, the presence of a spatial planning tool that is capable to capture and quantify the potential consequences of prospective spatial planning strategies is key in invigorating the opposite to sustainable development socio-political responses.

In short, simulation modeling is a tool that contributes to the endeavor of researchers and planners to approach and apprehend a broader spectrum of growth issues. By incorporating social, economic and political aspects into the analysis, it brings to surface conflicts and synergies in land use allocation and planning. In this light, it provides a sound framework for a richer spatial analysis, also facilitating well-informed decision-making processes.

References

Aburas, M.M., Ho, M.Y., Ramli, M.F. and 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.

As-Syakur, A.R., Adnyana, I., Arthana, I.W. & Nuarsa, I.W. (2012). Enhanced built-upand bareness index (EBBI) for mapping built-up and bare land in an urban area. *Remote Sensing*, 4, 2957–2970.

Bank of Greece. 2014. *The Chronicle of the Great Crisis. The Bank of Greece 2008-2013*. Athens: Bank of Greece.

Bank of Greece. 2016. *Summary of the Annual Report (2015)*. Athens: Bank of Greece.

Barredo, J., Kasanko, M., McCormick, M., & Lavalle, C. (2003). Modeling dynamic spatial processes: Simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, 64, 145–160.

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

Batty, M., Xie, Y. & Zhao, K. (2007). Simulating emergent urban form using agentbased modeling: Desakota in the Suzhou-Wuxian region in China. *Annals of the Association of American Geographers*, 97 (3), 477–495.

Bonham-Carter, G.F. (1994). Geographic Information Systems for Geoscientists: Modelling with GIS. In: Computer Methods in the Geosciences, vol. 13, p. 398. Pergamon.

Berling-Wolff, S., & Wu, J. (2004). Modeling urban landscape dynamics: A review. *Ecological Research*, 19, 119–129.

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

Cervero, R. (2001). Efficient urbanisation: Economic performance and the shape of the metropolis. *Urban Studies*, 38, 1651–1671.

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

Chavez, P.S., Jr. (1988), An improved dark-object subtraction technique for atmospheric scattering correction of multi-spectral data. Remote Sensing of Environment, 24, 459–479.

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

Chorianopoulos, I., Tsilimigkas, G., Koukoulas, S. & Balatsos, Th. (2014). The shift to competitiveness and a new phase of sprawl in the Mediterranean city: Enterprises guiding growth in Messoghia – Athens. *Cities*, 39, 133–143.

Clarke, K.C., Gaydos, L. & Hoppen, S. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, 24, 247-261.

EEA (2006) Urban Sprawl in Europe: The Ignored Challenge. Office for Official Publications of the European Communities, Luxemburg.

Engelen, G., White, R. & de Nijs, T. (2003). The Environment Explorer: spatial support system for integrated assessment of socio-economic and environmental policies in the Netherlands. *Integrated Assessment*, 4(2), 97–105.

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.

Feng, Y. & Liu, Y. (2016). Scenario prediction of emerging coastal city using CA modeling under different environmental conditions: a case study of Lingang New City, China. *Environmental Monitoring and Assessment,* 188, 540.

Feng, Y., Yang, Q., Hong, Z. & Cui, L. (2016). Modelling coastal land use change by incorporating spatial autocorrelation into cellular automata models. Geocarto International. DOI: 10.1080/10106049.2016.1265597.

Feng, Y., Liu, M., Chen, L. & Liu, Y. (2016). Simulation of Dynamic Urban Growth with Partial Least Squares Regression-Based Cellular Automata in a GIS Environment. *International Journal of Geo-Information*, 5, 243.

GGN (2014) New Master Plan for Athens-Attika. *Greek Government Newspaper*, 156(A): 4901-4996.

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.

Gutman, G., Huang, C., Chander, G., Noojipady, P. and Masek, J.G. (2013). Assessment ofthe NASA–USGS global land survey (GLS) datasets. *Remote Sensing of Environment*, 134,249–265.

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

Hasse, J.E. & Lathrop, R.G. (2003). Land resource impact indicators of urban sprawl. *Applied Geography*, 23, 159–175.

He, C., Okada, N., Zhang, Q., Shia, P. & Zhang, J. (2006). Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Applied Geography*, 26, 323–345.

Hellenic Parliament (2014) Explanatory Report. New Master Plan of Athens, New Master Plan of Thessaloniki and Other Provisions". Athens: Hellenic Parliament.

Johnson, M. (2001). Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda. *Environment and Planning A*, 33, 717-735.

Jongman, R.H.G. (2002). Homogenisation and fragmentation of the European landscape: ecological consequences and solutions. *Landscape and Urban Planning*, 58(2-4), 211-221.

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.

Kolb, M., Mas, J-F. & Galicia, L. (2013). Evaluating drivers of land-use change and transition potential models in a complex landscape in Southern Mexico. *International Journal of Geographical Information Science,* 27(9), 1804-1827.

Ku, C-A. (2016). Incorporating spatial regression model into cellular automata for simulating land use change. *Applied Geography* ,69, 1-9.

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.

Lavalle, C., Barredo, J., McCormick, N., Engelen, G., White, R., & Uljee, I. (2004). The MOLAND model for urban and regional growth forecast. A tool for the definition of sustainable development paths. European Communities Joint Research Centre.

Leontidou, L., Afouxenidis, A., Korliouros, E. & 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 and Policy Eds. C Couch, L Leontidou, G Petschel-Held (Blackwell, Oxford) pp 71-101.

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

Liu, Y. & Feng, Y. (2016). Simulating the Impact of Economic and Environmental Strategies on Future Urban Growth Scenarios in Ningbo, China. *Sustainability*, 8, 1045.

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., Kolb, M., Paegelow, M., Camacho Olmedo, M.T. & Houet, T. (2014). Inductive pattern-based land use/cover change models: A comparison of four software packages. *Environmental Modelling & Software*, 51, 94-111.

Municipality of Athens (2014) Study on market trends and development. Athens: Centre for entrepreneurial support.

Murray-Rust, D., Rieser, V., Robinson, D.T., Milicic, V. & Rounsevell, M. (2013). Agent-based modelling of land use dynamics and residential quality of life for future scenarios. *Environmental Modelling & Software***,** 46, 75–89.

Milesi, C. (2003). Assessing the impact of urban land development on net primary productivity in the southeastern United States. *Remote Sensing of Environment*, 86, 401– 410.

Paegelow, M. & Camacho Olmedo, M.T. (2005). Possibilities and limits of prospective GIS land cover modelling - a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). *International Journal of Geographical Information Science,* 19(6), 697-722.

Pagonis, A. & Chorianopoulos, I. (2015). Spatial planning and governance: Path dependent trajectories of rescaling in Metropolitan Athens. *Geographies*, 25: 77-91.

Poelmans, L. & Van Rompaey, A. (2010). Complexity and Performance of Urban Expansion Models. *Computer Environment and Urban Systems*, 34, 17-27.

Salvati, L., Mavrakis, A., Serra, P. & CarLULCci, M. (2015). Lost in translation, found in entropy: An exploratory data analysis of latent growth factors in a Mediterranean city (1960-2010). *Applied Geography*, 60, 107-119.

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.

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.

Song, C., Woodcock, C.E., Seto, K.C., Pax Lenney,M., & Macomber, S.A. (2001). Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? *Remote Sensing of Environment*, 75, 230–244.

Souliotis, N. (2013). Cultural economy, sovereign debt crisis and the importance of local contexts: The case of Athens. *Cities*, 33, 61–68.

Stevens, D., Dragicevic, S. & Rothley, K. (2007). iCity: A GIS- CA Modelling Tool for Urban Planning and Decision Making. *Environmental Modelling & Software*, 22(6), 761-773.

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A.& Fisher, P.F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, 1-14.

Turner II, B.L., Lambin, E.F. & Reenberg, A. (2007). The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, 104(52), 20666–20671.

Veldkamp, A. & Lambin, E.F. (2001). Predicting land-use change. *Agriculture, Ecosystems and Environment*, 85, 1–6.

Vermote, E., Tanré, D., Deuzé, J.L., Herman, M., & Morcrette, J.J. (1997). Second simulation of the satellite signal in the solar spectrum (6S). 6S User Guide Version 2.

Vliet, J., White, R. & Dragicevic, S. (2009). Modeling Urban Growth Using a Variable Grid Cellular Automaton. *Computers, Environment and Urban Systems*, 33, 35-43.

Wilson, B. & Chakraborty, A. (2013). The Environmental Impacts of Sprawl: Emergent Themes from the Past Decade of Planning Research. *Sustainability, 5*, 3302-3327.

Xiang, W.N., & Clarke, K.C. (2003). The use of scenarios in land-use planning. *Environment and Planning B: Planning and Design*, 30(6),885–909.

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.

Zha, Y., Gao, J. & Ni, S. (2003). Use of normalized difference built-up index inautomatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24,583–594.

Zheng, X-Q., Zhao, L., Xiang, W-N., Li, N., Lv, L-N. & Yang, X. (2012). A coupled model for simulating spatio-temporal dynamics of land-use change: A case study in Changqing, Jinan, China. *Landscape and Urban Planning*, 106, 51–61.

Chapter 8: Multi-scale modelling of land use/land cover (LULC) change with Geoinformatics and scenario-based simulations: The case of Attica region.

Abstract

The objective of this final chapter is to explore potential future land use/cover (LULC) dynamics in the terrestrial Attica region, under three scenarios that reflect future growth traits in the area related to different economic performance realities and alternative land use planning options. Attica experienced significant and, in places, unregulated urban growth, during the past decades, due to the absence of planning controls and the substantial increase of demand, as a consequence of migration and a boost in second homes, especially in the coastal zones. First, this chapter looks at the periodic LULC changes occurred during a period of 25 years (1991–2016) employing remote sensing techniques and Landsat satellite data. The observed changes are then related with 27 dynamic, biophysical, socio-economic and territorial factors, to generate transition potential maps implementing Random Forests (RF) regression modeling. Scenarios are projected until 2040 by implementing a spatially explicit Cellular Automata (CA) model. Finally, the resulting maps, are subject to a multiple resolution sensitivity analysis. Change detection reveal that the vast majority of the built-up land expansion took place at the expense of natural areas and croplands and during the last decades Attica region experienced remarkable land transformations. However, the late economic circumstances significantly affected the growth trends. Under an economically optimistic scenario which means high development, and assuming the absence of an adequate controlling mechanism, the built-up surfaces are expected to increase by almost 24%, by 2040, and consequently the natural areas and croplands are expected to decrease significantly. In case the economic scarcity persists, which can be translated in low development, the artificial surfaces are expected to slightly increase by approximately 7.5%, by 2040.

6.1 Introduction

In this last chapter the area of focus is the terrestrial part of Attica region which serves as a perfect example, that paradigmatically mirrors the rapid socio-economic transformations, the demographic dynamics and the population redistribution (mostly in the form of rural depopulation) occurred in Greece during the last decades. The region includes Athens, the capital of Greece, which has been the central pole of job opportunities and economic prosperity, and according to the late census (2010) is inhabited almost 4 million residents which accounts for the 35% of the total population. During the recent decades, the population influx triggered a persistent increase in housing demand and supply (Mantouvalou et al, 1995). In turn, concurrent socio-economic transformations triggered the redistribution of middle-class Athenian residents, seeking better quality of life, in areas outside the compact Athenian center, within a commuting distance from their jobs (Leontidou et al. 2007). At the same time, the socio-economic conditions favored a persistent amenity-driven second homes trend along the sea coast (Arapoglou and Sayas, 2009). As a consequence, to all the aforementioned, the landscape of the area, especially the peri-urban Athens, has changed substantially over the years. The whole structure of unplanned development was encouraged, or even forced, by the total absence of actual, on the ground, regulation mechanism and the weak land use planning (Pagonis, 2013). In fact, urban planning of the area was legally established since decades ago, but in the actual mechanisms that shape urban development on the ground, are not evident. On the contrary, development was permitted at any environmental, functional or operative cost. Moreover, after successfully attracting national and foreign funds and in the face of hosting 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). However, after the phase of progressive economic growth the area has been exposed to the negative consequences of the sovereign crisis and the economic recession, attributed to the global financial crisis, that in turn had a drastic influence on the development rates and significantly altered the housing and construction industries (Gounaridis et al.2018).

Topographically, Athens also constitutes an interesting case for study since it is characterized by undulated morphology (Figure 1). The geomorphological features of the region dictate the land availability and determine the accessibility and the optimal conditions for construction of built-up land. The plain of Athens is surrounded by large mountains (Aegaleo, Parnitha, Penteli and Hymettus), and this acts as a constraint, that separates Athens from the proximate flat districts (Thriasio, Messoghia, Marathonas), making them the only available areas to host residential and industrial settlements.

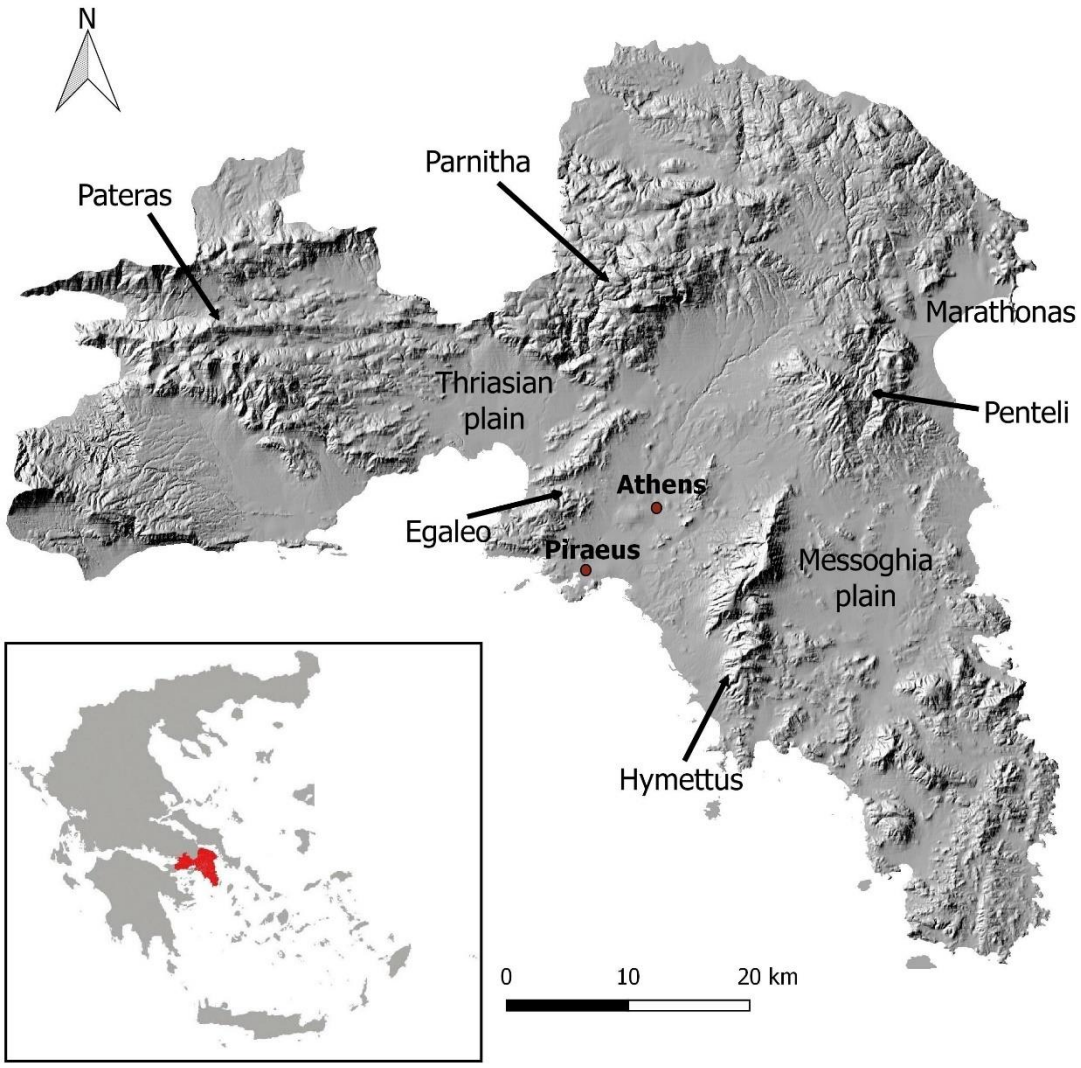


Figure 1. Topography of Attica region.

The aim of this chapter is to explore potential future LULC dynamics in the terrestrial Attica region, under three scenarios that reflect different economic performance realities and alternative planning options. To achieve this, it was deemed critical to implement an integrated methodological framework that combines the advantages of methodologies discussed is the previous chapters. The central premise was to simulate all categories of LULC changes (a total of 18 different possible LULC transitions identified) at the regional level, to evaluate the effects of different proximate and underlying causes (chapter 4 - Gounaridis et al. 2018). Since the area experienced vast land transformations during the recent decades, which are mostly related to built-up categories, a scientific objective of this approach was to achieve a very high thematic resolution in these categories, following the methodological framework described in chapter 3 (Gounaridis and Koukoulas, 2016). The non-artificial land categories were efficiently discriminated following the approach presented in chapter 2 (Gounaridis et al. 2016). Change detection techniques in the form of cross-classification and cross-tabulation were applied in order to map and quantify the periodic LULC changes occurred during the study period, following the methodology presented in chapter 1 (Gounaridis et al. 2014). Attempting to geographically associate the driving forces with the observed historical LULC changes, a suite of 27 different factors derived from multiple different sources and expressed in different scales, units and resolutions was incorporated in the modeling framework. The effective fusion of these data was achieved implementing the Random Forests (RF) algorithm following a similar approach to chapter 4 (Gounaridis et al. 2018). The models generated the transition probability surfaces that served as a basis to simulate the observed LULC changes and project future changes under different scenarios (chapter 4 - Gounaridis et al. 2018). Three different scenarios were composed that fully reflect the phases of uneven development observed in the area. To calibrate and fine tune the simulation models, landscape metrics were computed and introduced to the models (chapter 1 - Gounaridis et al. 2014). Finally, the results were subject to a multi-resolution sensitivity analysis in order to be more robust and unaffected by the technical details of inputs and the bias they entail. In this process the outputs generated by each model run were compared at several spatial resolutions in order to identify areas of future LULC change regardless the spatial resolution of the inputs. The results are expected to shed light in different economic performance realities and land-use planning contexts and choices while the

whole attempt will quantify the importance of various driving forces of change, contributing to an enhanced understanding of such a complex phenomenon.

6.2 Material and methods

6.2.1 Images pre-processing and classification

Since the Attica region is fully covered by two consecutive images (path: 183, row: 033 - 034), 10 Landsat images (Table 1) spanning 25 years (1991-2016) were chosen to achieve full geographical coverage. The acquired images meet certain quality standards, namely no cloudiness in the study area, acquisition during summer months to avoid phenological variations and absence of the scan line corrector problem of Landsat 7 after 2003.

Table 1. The characteristics of the satellite images that were used as the primary data to corroborate the change detection analysis.

Date	Sensor	Satelite type	Resolution (m)	Path/Row
17/9/1991	Thematic Mapper (TM)	Landsat 4	30	183/034
29/6/1991	Thematic Mapper (TM)	Landsat 4	30	183/033
22/8/1999	Enhanced Thematic Mapper Plus (ETM+)	Landsat 7	30	183/034
22/8/1999	Enhanced Thematic Mapper Plus (ETM+)	Landsat 7	30	183/033
12/10/2003	Thematic Mapper (TM)	Landsat 5	30	183/034
12/10/2003	Thematic Mapper (TM)	Landsat 5	30	183/033
12/8/2010	Thematic Mapper (TM)	Landsat 5	30	183/034
12/8/2010	Thematic Mapper (TM)	Landsat 5	30	183/033
29/9/2016	Operational Land Imager (OLI)	Landsat 8	30	183/034
29/9/2016	Operational Land Imager (OLI)	Landsat 8	30	183/033

To avoid any discrepancies due to the multi-temporal and multi-sensor type of analysis and to efficiently compute spectral indices, all images underwent radiometric as well as atmospheric correction following the methods described in Chapters 1 and 2. Topographic correction was also important in order to minimize the topographical effects due to the mountainous nature of the study area. The methodology followed for this step is also described in Chapters 1 and 2. Next, the two consecutive calibrated images per year were mosaiced ending up with five images spanning 25 years (1991, 1999, 2003, 2010, 2016).

All images were classified into eight LULC categories, implementing the RF classification algorithm through the RandomForest package available in R (Liaw & Wiener, 2002). Since the objective was to obtain LULC information with the best possible thematic resolution, especially for the urban land types, the LULC categories were 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).

The LULC categories distinguished by devising a semi-automated sampling extraction based on a context that combined the no-change areas, prior knowledge and spectral controlling. More specifically, starting with 2010 and 2016, an extensive sampling was designed based on visual interpretation of very high spatial resolution data from Google Earth and on existing available reference LULC data. Particularly, the datasets created by Gounaridis and Koukoulas (2016) as described in chapter 3 and by Gounaridis et al. (2016) as described in chapter 2, were used as reference data for the semi-automated sampling extraction. For the non-artificial LULC types of croplands, natural areas and other, additional samples from the Urban Atlas and Corine datasets were also assembled to strengthen the training. Last but not least, a quality control mechanism based on the spectral signatures of the samples to remove outliers from the analysis was applied (Radoux et al. 2014) (Figure 2). For the 1991, 1999 and 2003 images, a backwards automated training strategy was adopted. Given that changes usually occur on a fraction of the total area, the use of the unchanged areas as training samples for the desired past date is reasonable (Chen et al. 2012; Kim et al. 2014). No-change areas were identified via visual interpretation of very high spatial resolution data from Google Earth. These no-change areas were then used to semi-automatically generate training samples as input for the subsequent classification of each year. Special attention was payed to avoid taking points close to the boundaries of adjacent LULC categories, ensuring that clear samples of each category were taken and thus eliminating any source of confusion to the model. 70% of the samples were used to train the RF algorithm while the other 30% were kept independently for the accuracy assessment of the results (Table 2).

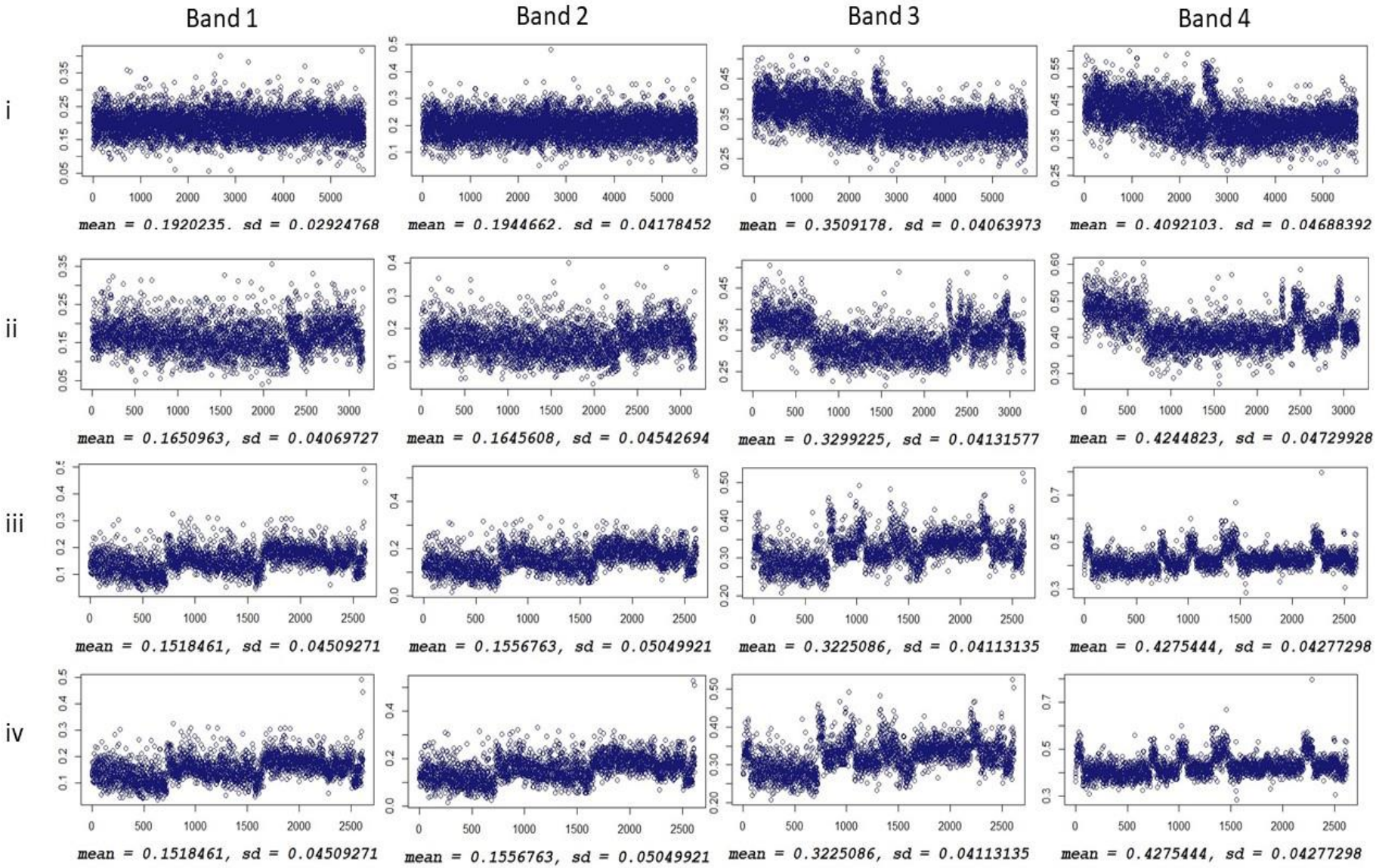


Figure 2. Example of spectral controlling (TM image 2010 – Red; Green; Blue; IR bands. i) Continuous urban fabric, ii) Discontinuous dense urban fabric, iii) Discontinuous medium density urban fabric, iv) Discontinuous low density urban fabric).

Table 2. Training and validation samples used for classification

	Training				
	1991	**1999**	**2003**	**2010**	**2016**
Continuous urban fabric	1798	2321	2622	4095	5707
Discontinuous dense urban fabric	969	1401	1808	2329	3159
Discontinuous medium density urban fabric	1021	1286	1446	1888	2617
Discontinuous low density urban fabric	502	895	1175	2331	3221
Industrial, commercial and transport units	473	685	991	1245	1717
Arable land and permanent crops	2009	2119	2409	2009	2776
Forests, Scrubs and other natural areas	1449	1463	1559	1568	1974
Other (open spaces, bare land, mine, inland water)	453	460	475	525	574
Total	8674	10630	12485	15990	21745
	Validation				
	3637	4319	5419	6919	9399

Since RF has proven efficient with large data handling, provides reduced likelihood of over-fitting and is suitable for multi-source inputs (Gounaridis et al. 2014; Gounaridis and Koukoulas 2016, Gounaridis et al 2016, Gounaridis et al. 2018), the classification models involved 20 variables in total. Besides the 6 reflective Landsat bands (bands 1–5 & 7 for Landsat 5 TM and Landsat 7 ETM+, bands 2-7 for Landsat 8 OLI), the thermal band was also used as it provenly helps in the classification process (Rodríguez-Galiano and Chica-Olmo, 2012). In addition, the first layer produced by principal components analysis (PCA) separately for the three visible bands (1, 2 and 3) and the infrared bands (5 and 7), was also incorporated as it appears to increase classification accuracy (Gounaridis et al. 2014; Gounaridis et al. 2016). In addition, following the approach by Gounaridis and Koukoulas, (2016), the normalized difference built-up index (NDBI) and the enhanced built-up and bareness index (EBBI) were also computed and incorporated in order to enhance the urban LULC types discrimination. The Enhanced Vegetation Index (EVI), the Normalized Difference Moisture Index (NDMI), the Normalized Difference Bareness Index (NDBaI) and the Normalized Differential Vegetation Index (NDVI) were also included because of their capacity to separate vegetation from bare features during the classification process. The three widely used Tasseled Cap (TC) transformations namely, Soil Brightness Index (SBI), Green Vegetation Index (GVI) and Moisture Content of Soil/Vegetation

(Wetness) were also computed and incorporated (Gounaridis et al. 2014; Gounaridis et al. 2016). Finally, auxiliary variables (elevation and slope) acquired from the Global Land Survey Digital Elevation Model (GLSDEM) were also included.

To set up the models, RF requires two primary parameters to be specified by the user: the number of predictor variables randomly sampled at each decision tree split and the number of classification trees to be built. Four (4) predictor variables were chosen for each tree split, which is equal to the square root of the total number of predictor variables and 500 trees for each run. The variables' importance for each classification run is illustrated in Figure 3. The elevation and slope always rank among the top, showing the major influence these attributes have on different types of LULC. The inclusion of the thermal band had also an influence in most of the models.

To sidestep the so called 'salt n pepper effect' of the resulting maps, all isolated patches (defined as area less than 0.1 ha), were removed by replacing their category value with the mode of their neighborhood pixels, defined by a 3x3 window (Gounaridis et al. 2014; Gounaridis and Koukoulas 2016, Gounaridis et al 2016, Gounaridis et al. 2018). Results were plotted against the 30% of the initial samples and validated using the cross-tabulation approach.

6.2.2 Leap-frog development index

To enhance the accuracy of the model, and to ensure the accurate detection and representation of scattered development, the Leap-frog development index, originally proposed by Xu et al. (2007) was calculated and included in the modeling scheme. 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. 2018). Based on the 1991 and 2016 maps, the index was calculated following the methodology described in chapter.

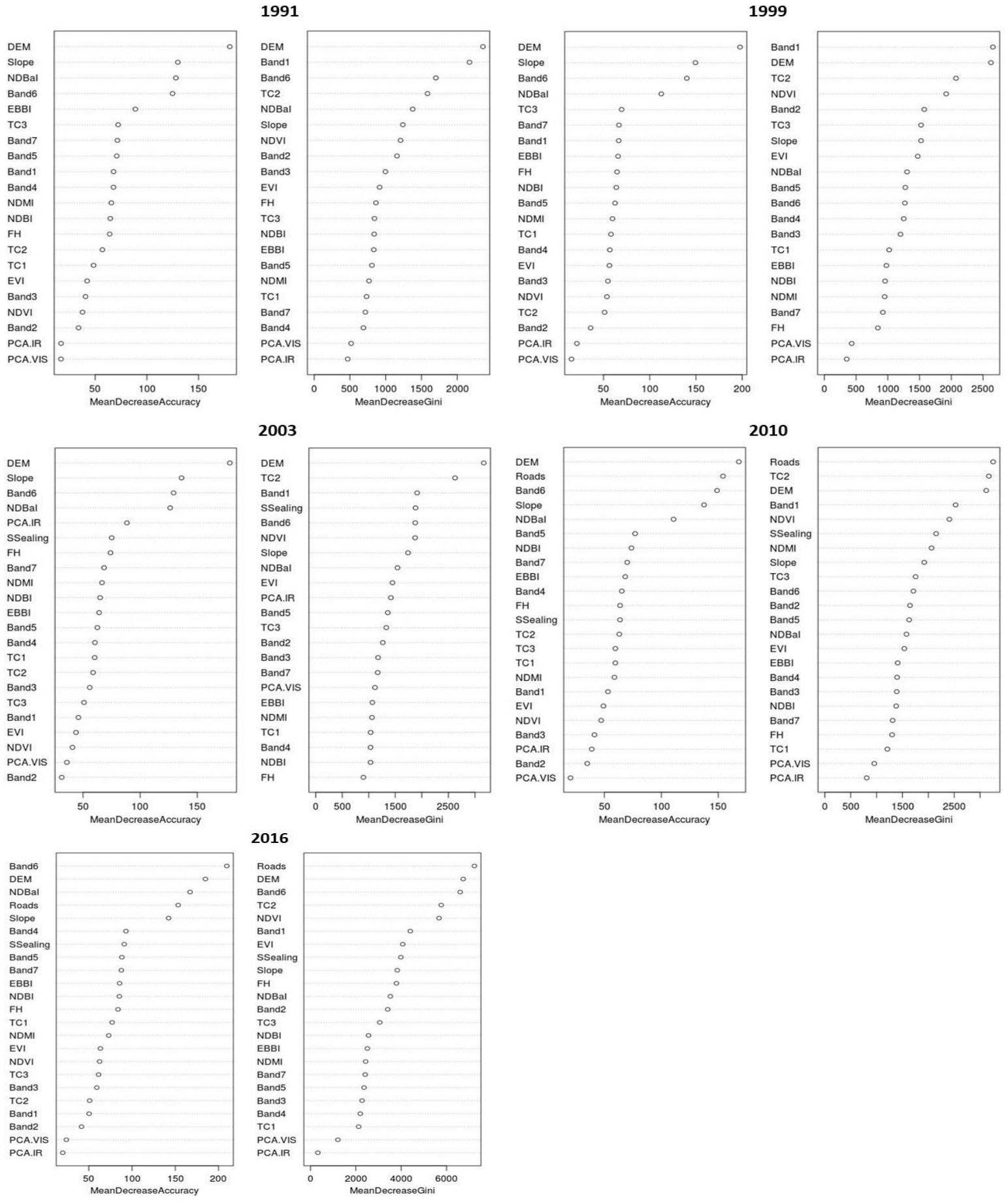


Figure 3. Predictor variables' importance for each classification model

6.2.3 Transition potential modeling

6.2.3.1 Predictor variables

This section introduces the suite of predictors used to model the transition probabilities. Taking into account previous efforts, data availability and accessibility, a suite of 27 variables was concluded to best represent the LULC change phenomenon occurred throughout Attica region during the period of focus (1991-2016). Since the changes related to artificial surfaces were dominant in Attica, the majority of the chosen variables represent factors and processes that explain peoples' choices about residential and infrastructure location. Social shifts, economic motives, inherent quality and attractiveness of a given place, the effects of neighboring areas and proximity to basic needs and amenities were assumed to play a key role (Table 3). It was hypothesized that these biophysical, socioeconomic, territorial and land use factors can spherically explain the changes occurred during the last decades in Attica region (Gounaridis et al. 2018). Similarly to chapter 4, the variables used are of different nature (both categorical and continuous), derived from multiple sources, with different scales and resolutions. Many of these variables act as proxies to other non-measurable factors. It is worth noting that, variables available at the prefecture or regional level were not included, since the municipalities in Greece are responsible for local land management and thus representing a meaningful spatial unit for this type of analysis (Salvati, et al. 2015; Panori et al. 2016; Gounaridis et al. 2018). After processing, all no-spatial data were then collated in a GIS environment at the municipality level while all distances were computed using the Euclidean distance function. The last step was to convert all variables to raster format after resampling them at 30 m spatial resolution to match the resolution of the classification products.

Table 3. List of predictors used in the transition potential modeling process.

Acronym	Variable	Discription	Source	Time interval
Territorial variables				
DEM	Elevation	Elevation in m	GLSDEM[a]	(-)
SL	Slope	Slope in degrees	GLSDEM	(-)
AS	Aspect	Aspect in degrees	GLSDEM	(-)
CQI	Climate Quality	Climate quality index	EEA[b]	1961-1990
VIEW	Viewshed	Visibility from residential areas at the parcel level (centroids from UA).	GLSDEM and Urban Atlas[c]	(-)
DB	Distance from beaches	Euclidean distance from beaches signed with a blue flag in m	Ministry of Environment & Energy[d]	2010
DS	Distance from the sea	Euclidean distance from the shoreline in m		(-)
Socio-economic variables				
DEDU	Distance from Education centers	Euclidean distance from public education centers (all levels)	Ministry of Education & OSM[e]	2010
DPH	Distance from public health centers	Euclidean distance from public health centers	Society of Information[f] & OSM	(-)
DT	Distance from nearest town	Euclidean distance from the center of the nearest town in m	OSM	(-)
DPB	Distance from public buildings	Euclidean distance from public buildings	Society of Information & OSM	
DPH	Distance from public health	Euclidean distance from public hospitals and other public health care units in m	OSM	(-)
DPT	Distance from public transport	Euclidean distance from public transport stops (bus, metro, tram, suburban train) in m	OSM & opendata	(-)
DRN	Distance from road network	Euclidean distance from road network in m	OSM	(-)
POP	Demographics	Changes in population density at the municipality level	ELSTAT[g]	1991-2011
EMP	Employment rate	Total number of employed persons per total population at the municipality level	ELSTAT	1991-2011
UNEMP	Unemployment rate	Total number of unemployed persons per total population at the municipality level	ELSTAT	1991-2011
LVI	Landscape values Instagram	Landscape values quantified using Instagram data	van Zanten et al. (2016)[h]	2004-2015
LVF	Landscape values Flickr	Landscape values quantified using Flickr data	van Zanten et al. (2016)	2004-2015
LVP	Landscape values Panoramio	Landscape values quantified using Panoramio data	van Zanten et al. (2016)	2004-2015
Land use				
DGU	Distance from green urban areas	Euclidean distance from green urban patches (centroids from UA) in m	Urban Atlas	2006
SSM	Soil Sealing rate	Average soil sealing per municipality	EEA	2006-2012
MTC	Tree cover	Average tree cover canopy percentage per municipality	USGS[i]	2010
BU	Built-up rate	Cumulative total number of new houses built per municipality	ELSTAT	1997-2016
ENTH	HeatMap of Enterprizes	HeatMap of new enterprises registered to ACCI	ACCI[j]	1991-2016
ENT	Enterprises count	Cumulative total number of new enterprises registered to ACCI per municipality	ACCI	1991-2016
DN	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	(-)

[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/figures/climate-quality-index-map
[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

6.2.3.2 Random Forests regression models

Following the approach adopted by Gounaridis et al. (2018), the transition probability surfaces were generated by employing the regression type of the RF algorithm. All predictor variables, including the calculated Leap-frog index, were combined in a single stack and served as independent variables for the model. The initial maps resulted from classification had eight categories, as depicted in Figure 8. The possible transitions identified were 18 (Table 4), under three assumptions: it is impossible a) the urban fabric to convert to any other land type as well as to decrease in density, b) the industrial, commercial and transport units to convert to any other land type and c) the "other" category that includes inland waters, bare land and mines to interact with other classes. To train each of the 18 models, 5000 randomly placed points were dispersed throughout the extent of the study area. Values assigned in binary scale to indicate change (from category A to category B) and no change. The regression version of RF was then implemented in R using the RandomForest package (Liaw & Wiener, 2002). To fine tune the RF, five (5) predictor variables (equal to the square root of the total number of predictor variables) were used for each tree split and 700 trees for each run. The modeling process generated 18 transition probability surfaces, each one of them indicating the degree of potential change in the future per transition. The performance of the models was assessed using the Areas Under Curve (AUC) metric, derived from the Receiver Operating Characteristic (ROC) curve (Figure 13). The importance of each predictor variable was computed using the Mean Decrease Accuracy (%IncMSE) and the Mean Decrease Gini (IncNodePurity) metrics (Figure 17 and 18).

Table 4. Transition probabilities of the eight LULC categories.

	From	To
1	Discontinuous dense urban fabric	Continuous urban fabric
2	Discontinuous medium density urban fabric	Continuous urban fabric
3	Discontinuous medium density urban fabric	Discontinuous dense urban fabric
4	Discontinuous low density urban fabric	Continuous urban fabric
5	Discontinuous low density urban fabric	Discontinuous dense urban fabric
6	Discontinuous low density urban fabric	Discontinuous medium density urban fabric
7	Arable land and permanent crops	Continuous urban fabric
8	Arable land and permanent crops	Discontinuous dense urban fabric
9	Arable land and permanent crops	Discontinuous medium density urban fabric
10	Arable land and permanent crops	Discontinuous low density urban fabric
11	Arable land and permanent crops	Industrial, commercial and transport units
12	Arable land and permanent crops	Forests, Scrubs and other natural areas
13	Forests, Scrubs and other natural areas	Continuous urban fabric
14	Forests, Scrubs and other natural areas	Discontinuous dense urban fabric
15	Forests, Scrubs and other natural areas	Discontinuous medium density urban fabric
16	Forests, Scrubs and other natural areas	Discontinuous low density urban fabric
17	Forests, Scrubs and other natural areas	Industrial, commercial and transport units
18	Forests, Scrubs and other natural areas	Arable land and permanent crops

6.2.4 Scenarios

The overarching aim was to design scenarios that reflect different economic performance realities and planning options. The scenarios were based on the observed historical trends during the 25 years period of focus. After mapping (Figure 8) and quantifying (Figure 4) the historical trends, it is evident that the LULC change dynamics in Attica region were notably uneven, a fact that clearly reflects different phases of economic development and performance.

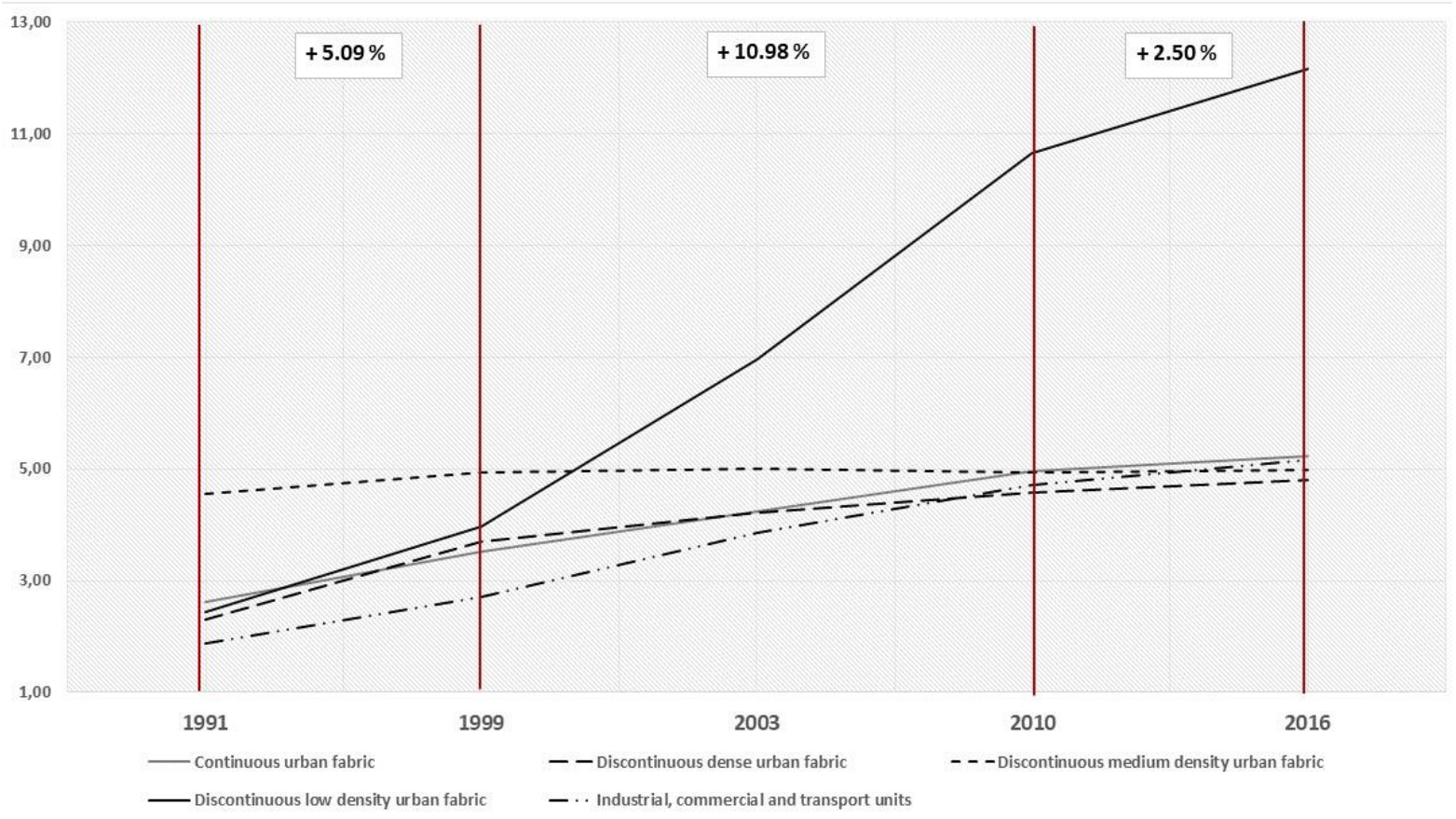


Figure 4. Summary of statistics based on the classified maps and relative percentages of artificial areas. Three different trends reflect three different levels of development during the last 25 years.

Thus, similarly to the idea behind the approach presented in chapter 4, three scenarios, namely "medium development", "high development" and "low development", were composed in order to reflect the uneven historical trends occurred in Attica region:

Medium development reflects the period 1991–1999 where the outskirts of Athens conurbation, especially the uplands and Messoghia plain are re-brought to the fore and into the development path. The Athens conurbation experiences a steady population increase as a result of rural depopulation and external migration because of the economic potential, the job opportunities and the social amenities that Athens has to offer. This constant societal demand to capture new economic opportunities and the centralization tendency triggered another tendency where Athenians, seeking a better quality of life, moved to the outskirts of Athens to areas with less urban density. Consequently, an increase in the housing demand and in turn the housing construction

led to a boost in urban growth at the expense of other less profitable land uses and progressively brought radical changes in the peri-urban landscape.

High development reflects the period 1999–2010 where the Attica region experiences exponential rise and permissive urban policies following the trends observed during the pre-olympic and meta-olympic period when the fiscal crisis was still less evident. During this time, the steady population increase along with the housing demand culminates while extensive regeneration of the waterfront also takes place, channeling changes in real estate dynamics towards tourism-specialized settlements and second homes. Planning and spatial policies contribute towards this direction, allocating funds and promoting the socio-economic restructuring. Specifically, investment in transportation infrastructure, enhances the accessibility to both three sides of Athenian growth (Northern outskirts- Maranthon, Oropos, Messoghia and Thriasio plains), triggering a population influx and a concurrent increase in housing construction. Consequently, infrastructure, firm headquarters, enterprises and shopping centers colonize these areas, leading to an economic polarization and economic functions re-concentration. Last but not least, during this time, land use planning controls are not in place, as unregulated built-up expansion is approached as a shortcut to economic growth.

Low development scenario reflects the period 2010–2016 and keeps the development pace very low as a consequence of economic scarcity and lack of investments. The economic functions as well as the population flows remain stable following a low pace. An amount of already built residences, intended to meet the needs for both second or primary housing, remain uninhabited (unsold or unfinished) while many already constructed industrial and commercial facilities remain unexploited. At the same time the demand, shaped by economic scarcity and population low rates, is lower than the already built and available buildings leading to a low to very low building rate in the area. All three scenarios were composed under the assumptions that profound social and political changes will not occur, as well as any other extreme events and accessibility in the area will remain stable. As far as land use regulations and legislative frameworks are concerned, it is assumed that especially in the medium and high development scenarios, will continue to be loose in the face of development potential.

6.2.5 Model calibration

The CA model designed and implemented in Dinamica EGO platform following a similar methodology to chapter 4. A crucial step, prior to the prediction phase is the model's calibration. To calibrate the model and evaluate the goodness of fit, a comparison of simulated maps with a reference/observed maps is the most efficient way (Gounaridis et al. 2018). Any CA modeling 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.

In this case, initially the model was set to train 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 employing the transition matrix function. Next, the parameterization set up required the computation of two landscape metrics, the mean and the variance of patch size and the patch isometry in order to replicate the actual conditions of the area in terms of structure and composition. 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 (equal) the newly developed patches. The first metric was computed for the input LULC map (2010) while the latter was adjusted through the trial and error process. Last, the 18 transition probabilities were stacked and consisted the cell allocation target area where cells with the highest likelihood values are supposed to change first. The model was then set to run and predict 2016. To evaluate the model's performance, the simulated LULC map of 2016 was compared with the observed LULC map of 2016 (resulted from classification) using the fuzzy similarity index at multiple resolutions (Gounaridis et al. 2018, Hagen, 2003).

6.2.6 Scenarios simulation

After calibration, the projection of LULC changes under the three scenarios 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 (1991-1999; 1999-2010; 2010-2016), to reveal the quantity

of each possible transition per scenario (Table 5). Ideally, the predictor variables (Table 4) and in turn the transition probabilities surfaces would also change per scenario, to better reflect the socio-economic conditions of each epoch, but in this case this option was impossible due to data unavailability and temporal mismatch.

Table 5. Transition probabilities allocated per scenario. The numbers indicate transition rates per year in hectares.

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

6.2.7 Multi-resolution sensitivity analysis

After completing the simulation process per scenario at 30m spatial resolution, a sensitivity analysis conducted following an approach with multiple resolutions. The central premise behind this step was that the spatial resolution of the models' inputs can have important and substantial effects on the output, and thus potentially 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. It was hypothesized that when all other parameters of the model are held constant and only spatial resolution of inputs changes, the quantities, the spatial allocation and thus the spatial patterns of outputs can differ.

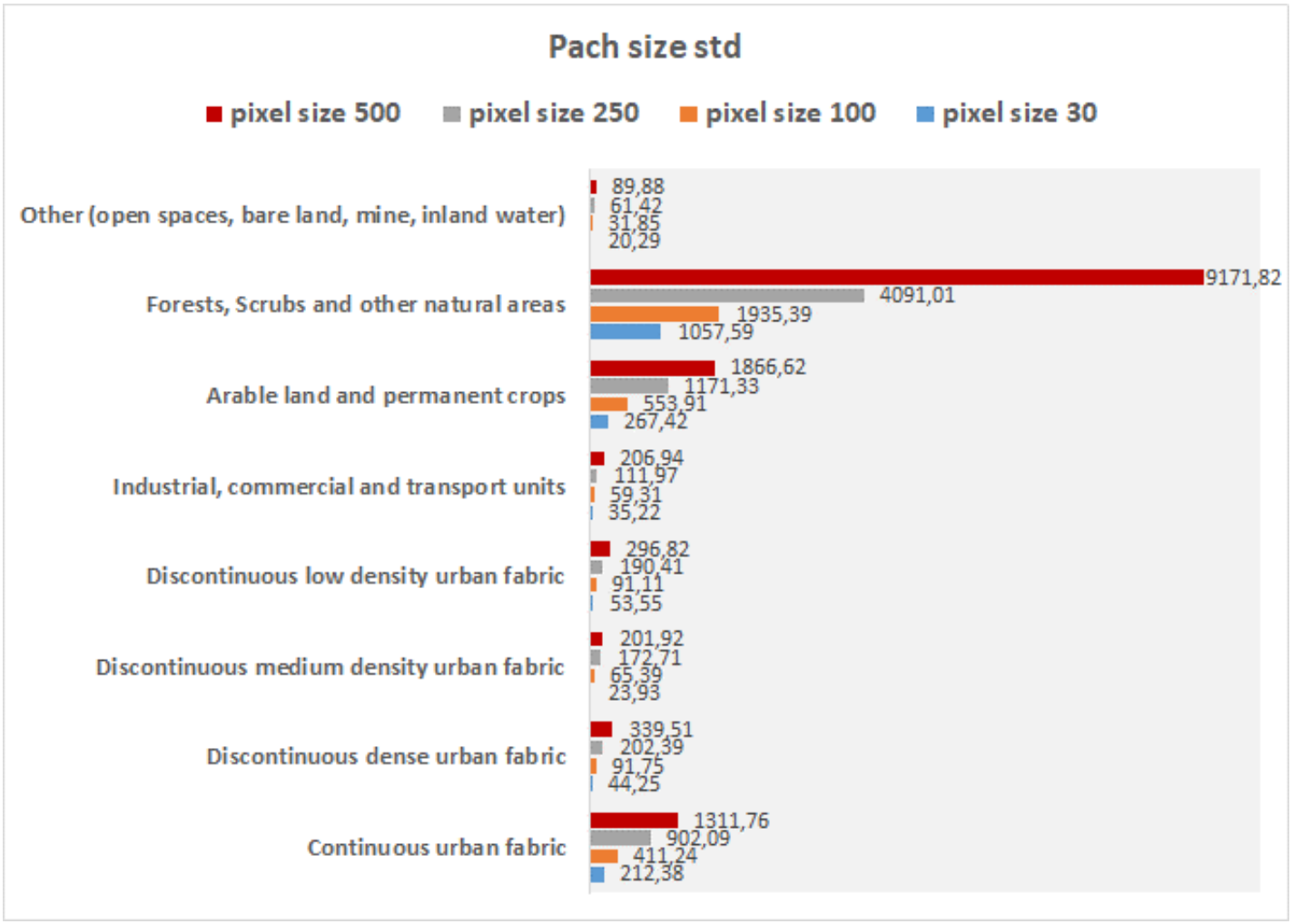


Figure 5. Variation of Patch Size Std landscape metric used to feed each calibration model, according to spatial resolution

Thus, the overarching aim of this step was to compare the outputs of models at the native resolution (30m) and at several coarser resolutions (100m, 250m, 500m) and identify areas of change that are common regardless the spatial resolution of the inputs. To do so, the outputs of the initial LULC classifications were resampled to 100, 250 and 500 meters respectively and change detection was performed for each case. Next, the transition probabilities were re-constructed through RF regression after resampling 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 (Figures 5 & 6). 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 a cross classification technique, in order to produce the final map per scenario.

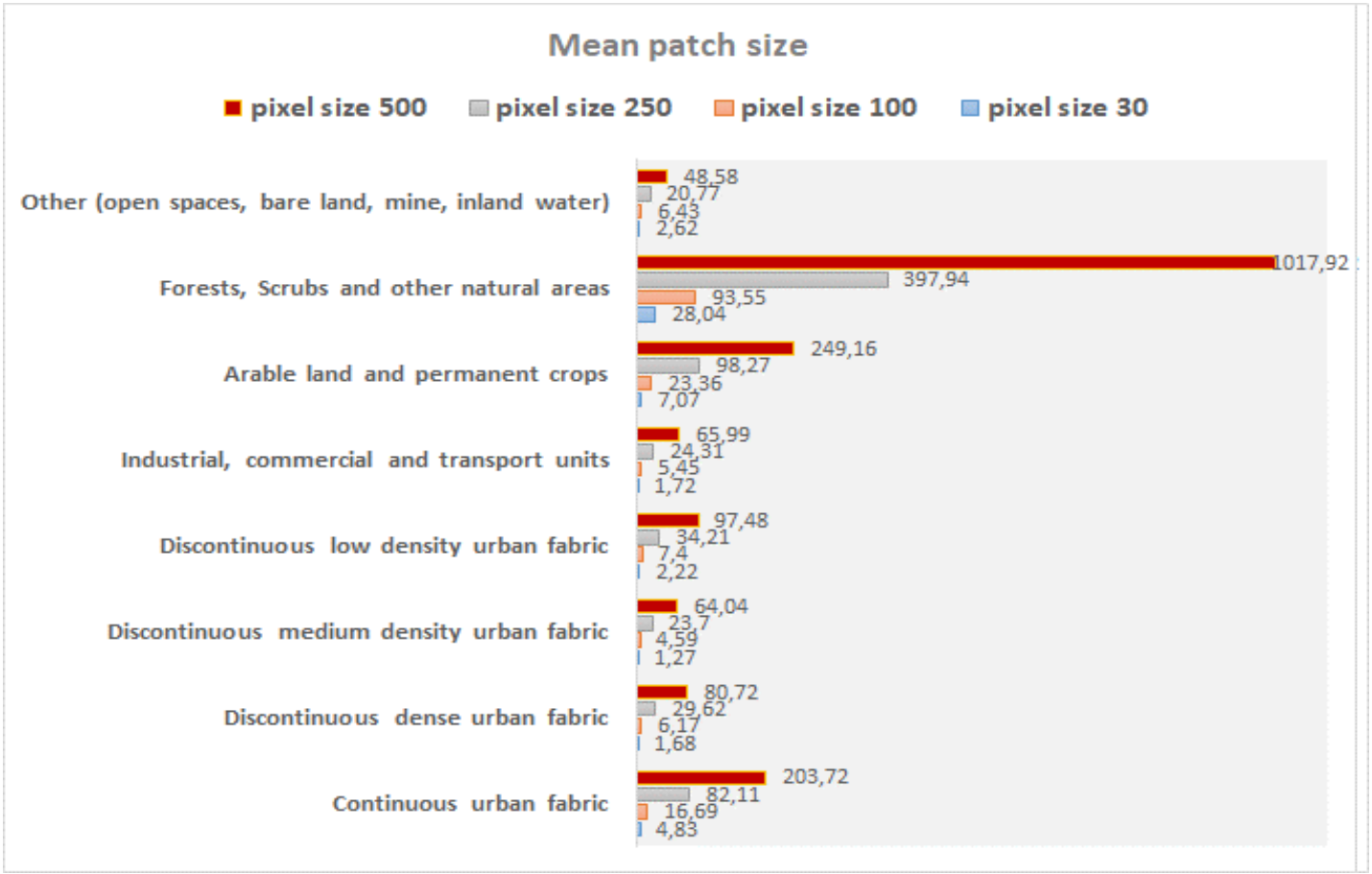


Figure 6. Variation of Mean Patch Size landscape metric used to feed each calibration model, according to spatial resolution

6.3 Results and discussion

6.3.1 Historical land use/cover change

The classification of the five mosaiced Landsat images reveal the LULC changes occurred in Attica region during a 25 years period. Overall accuracy for all images ranged from 90.5% to 93.5%. (Table 6). Results are depicted in Figure 8. Apart from mapping, the resulting maps were also quantified in order to have a quantitative insight on the transitions occurred. In conjunction with Figure 4 which depicts the increase in the built-up land categories, Figure 7 provides a quantified picture of the LULC changes observed in Attica region.

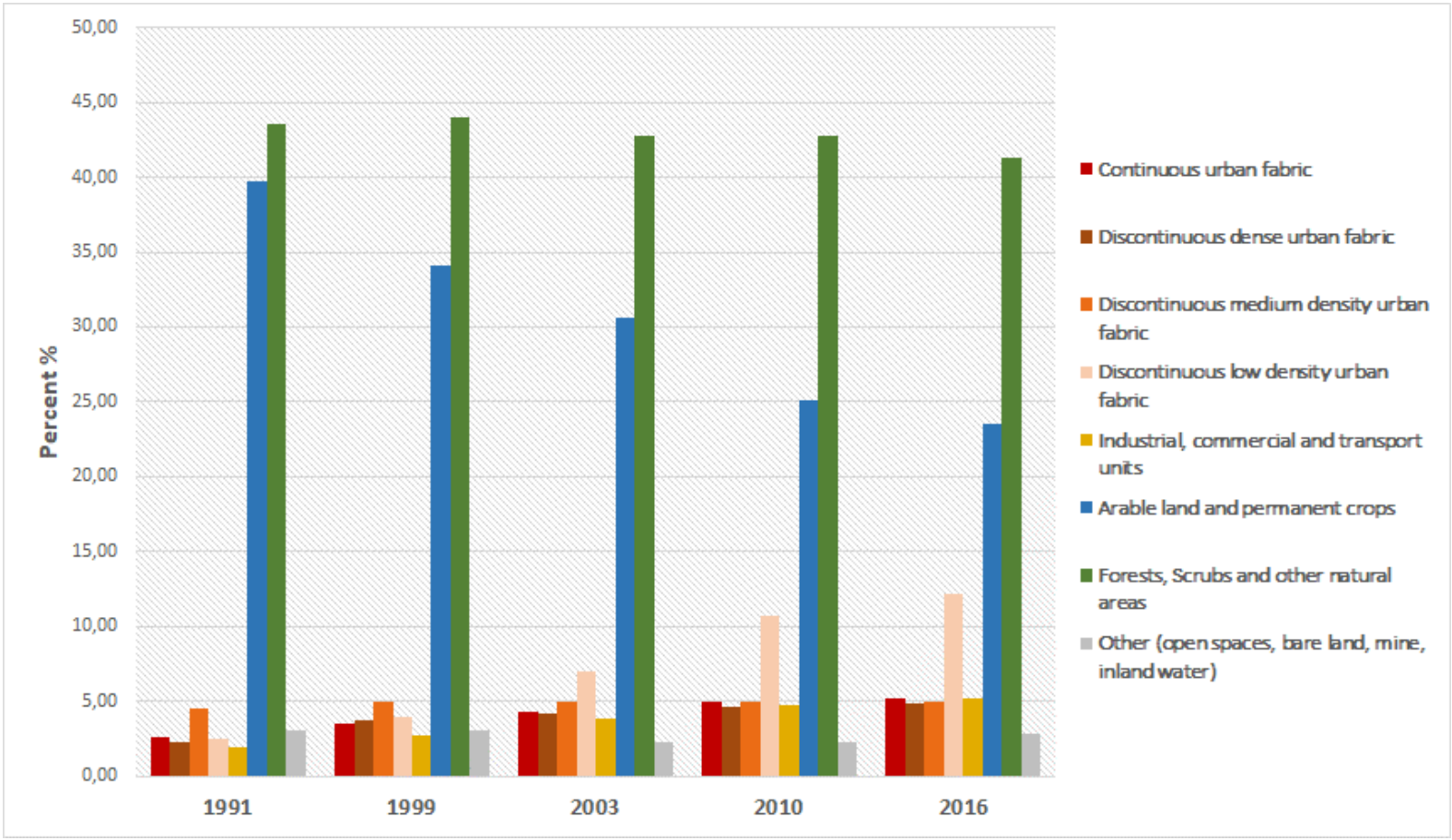


Figure 7. Summary of statistics based on the five classified maps and relative percentages of LULC categories.

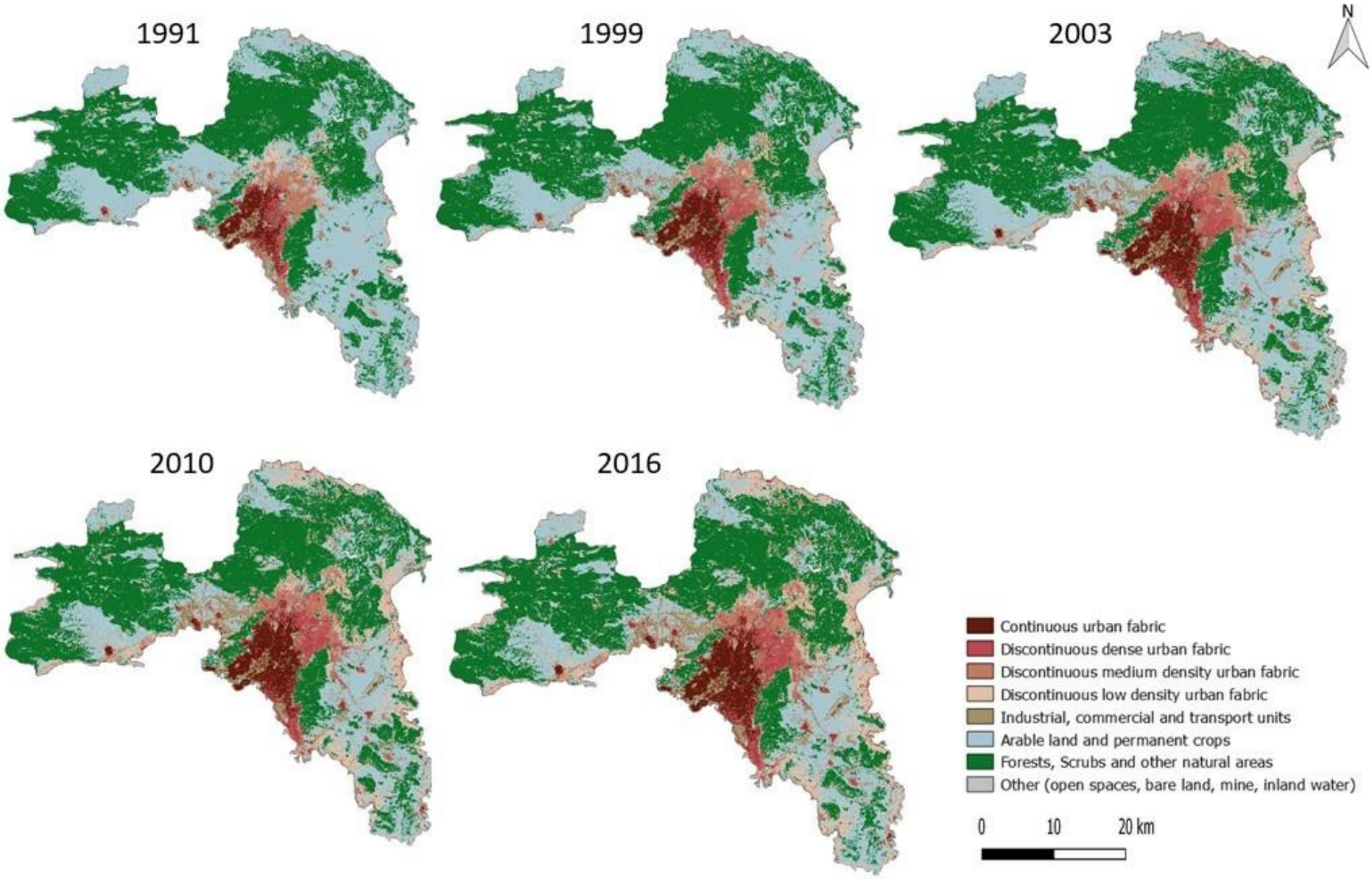


Figure 8: Results of the LULC classificatio**n**

Table 6. Error matrix - Resulting map per year against reference samples. O.E: Omission Error; C.E: Commission Error; 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, mine, inland water)

		1991									**1999**								
		Result									***Result***								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**O.E**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**O.E**
Reference	**1**	**681**	11	3		9				97%	**1012**	28	6		8				96%
	2	14	**631**	28		12	2	2	2	91%	58	**1174**	22		18	2	6	14	91%
	3	31	51	**1710**	58	4	6	6	4	91%	32	67	**1833**	15	15	6	3	11	92%
	4	8	20	104	**1288**	12	20	10	8	88%		28	90	**1426**	36	24	12	4	88%
	5		10	15	10	**515**	15	5	10	89%	25	20	15	5	**859**	15		10	91%
	6	12	13	40	148	24	**6576**	168	28	94%	6	46	48	196	26	**4946**	120	26	91%
	7		8	6	42	11	109	**2337**	11	93%		21	63	22	16	188	**2427**	19	88%
	8			1	7	10	16	11	**258**	85%			10	24	14	28	14	**456**	84%
	C.E	91%	85%	90%	83%	86%	98%	92%	80%		89%	85%	88%	84%	87%	95%	94%	84%	
	O.A	92,2%									90,5%								

2003

Reference \ *Result*	1	2	3	4	5	6	7	8	O.E
1	**1323**	42	10		19			1	95%
2	56	**1658**	52	4	20	8	6	4	92%
3	30	35	**2276**	30	38	3	12	16	93%
4		48	86	**2682**	56	64	24	12	90%
5	25	20	15	20	**1165**	30	10	10	90%
6		32	46	228	38	**4128**	60	16	91%
7		7	21	61	21	224	**2860**	24	89%
8			8	6	14	20	19	**443**	87%
C.E	92%	90%	91%	88%	85%	92%	96%	84%	
O.A	90,7%								

2010

Reference \ *Result*	1	2	3	4	5	6	7	8	O.E
1	**1838**	51	21		14		1		95%
2	72	**2241**	42	10	36	4	8	12	92%
3	21	46	**2299**	63	31	15	21	9	92%
4		48	91	**3246**	76	42	40	10	91%
5	33	20	40	21	**1798**	25		7	92%
6		6	18	162	54	**4091**	64	12	93%
7		14		48	28	207	**3528**	14	92%
8		8		16	11	14	20	**490**	88%
C.E	94%	92%	92%	91%	88%	93%	96%	88%	
O.A	92,3%								

2016

Reference \ *Result*	1	2	3	4	5	6	7	8	O.E
1	**2683**	59	19		24		2	2	96%
2	80	**3252**	76	18	42	22	10	8	93%
3	24	61	**2964**	101	35	17	36	20	91%
4	4	32	92	**5446**	78	112	44	1	94%
5	55	30	30	30	**2325**	105		17	90%
6		6	12	36	30	**4866**	66	26	97%
7		14		35	28	182	**4823**	22	94%
8				16	16	34	24	**616**	87%
C.E	94%	94%	93%	96%	90%	91%	96%	87%	
O.A	93,5%								

The most significant changes are the urban and industrial expansion which started to be evident since 1999 and culminated in 2010. Especially the discontinuous low density urban fabric started to increase rapidly by 2003, reaching 7% (from 2.5% in 1991) and this trend continued until 2016, reaching 12%. The continuous as well as the discontinuous dense urban fabric, almost doubled throughout the study period, reaching 5.5% and 4.8% respectively in 2016, while in 1991 they were about 2.6% and 2.3% respectively. These trends clearly reflect the previously discussed decentralization of Athenians to areas with less density. It is worth noting that after 2010 the development trends remain positive but start to decline as a consequence of the dramatic decrease of investments and the overall economic deterioration of the demand and supply equilibrium. All the aforementioned development trajectories took place at the expense of agricultural areas which were about 40% in 1991 and declined to 23.5% in 2016.

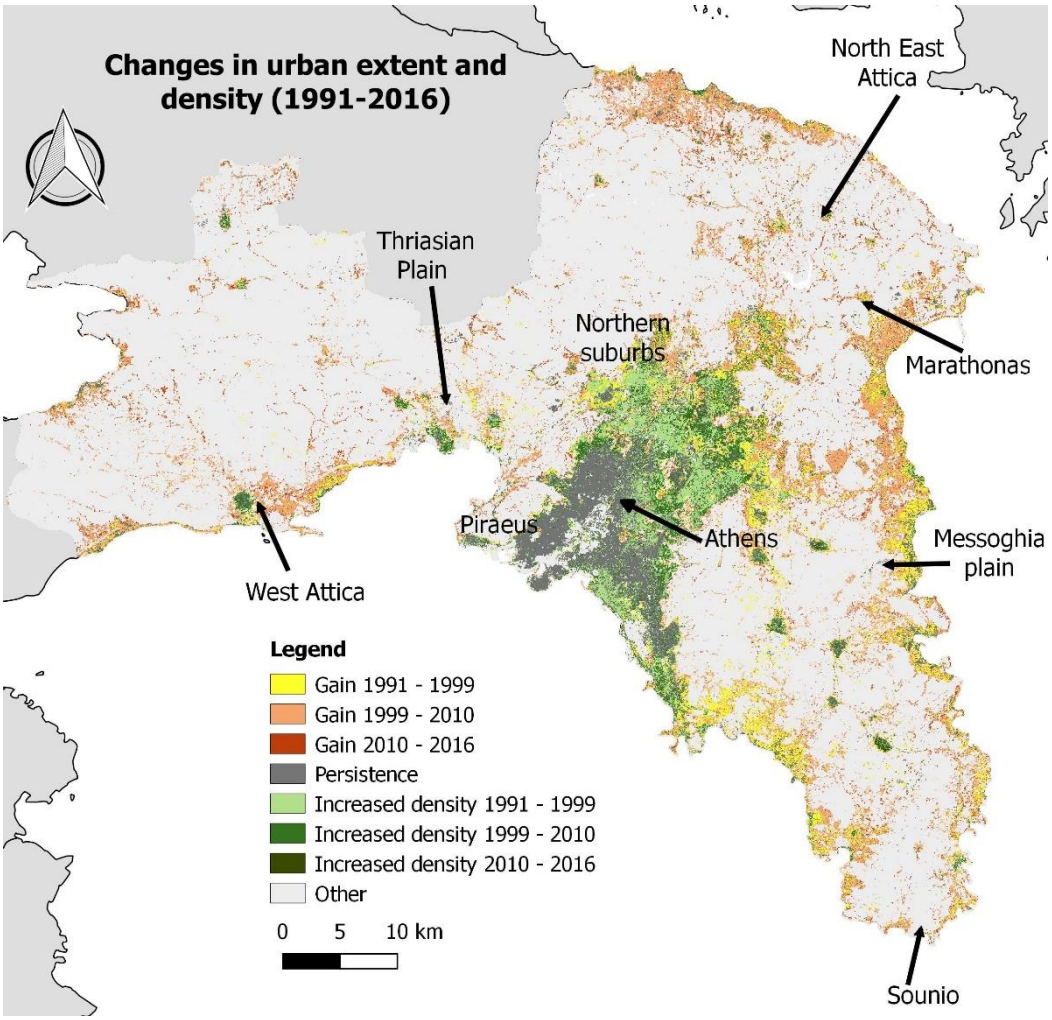


Figure 9. Urban trajectories observed between 1991-2016 (includes all urban fabric categories).

The forests and natural areas category remained relatively stable, decreasing only by 3% in 25 years. To even better highlight the observed trajectories of the most prominent changes, a suite of maps portraying each major LULC change was created. Figure 9 depicts the urban expansion as well as the increase in density occurred during the last 25 years in Attica region. As can be seen, urban sprawl is obvious mostly in the northern and eastern parts of Attica, while in the west, it is less but still evident. The majority of sprawl occurred in the waterfront, especially in the Messoghia, in Marathonas, in Oropos and in the south east of Athens. The northern suburbs of Athens also experienced an amount of expansion, but the most dominant type of change in this area was the infill that consequently led to significant increase in density. Also notable is the pattern of uneven development between the three different periods. The majority of LULC changes took place during the 1999-2010 period in all types of urban categories while the period 2010-2016 demonstrates the less changes. Regarding the types of sprawl, all types (namely suburban growth, leap-frog development, strip development and scattered development), can be observed in the area, a fact that can be attributed to a loose regulatory framework and to the absence of a planning scheme.

Figure 10 depicts the expansion of industrial, commercial and transport units, occurred the last 25 years in Attica region. Thriassian plain, located in the west of Athens as well as Messoghia plain located to the east, experienced the largest amount of this type of LULC change. During the period 1991-1999, Thriassian plain faced a notable industrial expansion while during 1999-2010 period the constraction of the new international airport in Messoghia, dominates. These two areas facilitated the expansion and were targeted due to two main advantages. First, they were both interlinked with Athens both in terms of easy access from and to Athens and as a pole of working hands. Additionally, they had morphological features suitable for construction activities, with available land and very low land costs. However, these two areas have different attributes. Thriassian plain, is mainly occupied by industrial facilities such as oil refineries, steel mills, military bases and centers of transshipment. Messoghia plain, on the other hand, is occupied by commercial clusters surrounding the newly development transport units and by large physical infrastructure facilities such as the Olympics related venues and the new international airport.

Urban and industrial expansion occurred the last 25 years in Attika region, mostly affected the agricultural land leading to a considerable amount to be consumed

over built-up land categories. Figure 11 depicts the aggregated loss of agricultural land over the past 25 years. As can be seen, almost all the surrounding areas of the greater Athens, experienced severe agricultural land loss. In an analogous fashion with the urban and industrial expansion (Figures 9 & 10), the agricultural land loss is more than evident in the Thriassian plain, west Attica, Oropos, Marathonas, Messoghia plain and southeast of Athens. The vast majority of this loss occurred during the period 1991-2010.

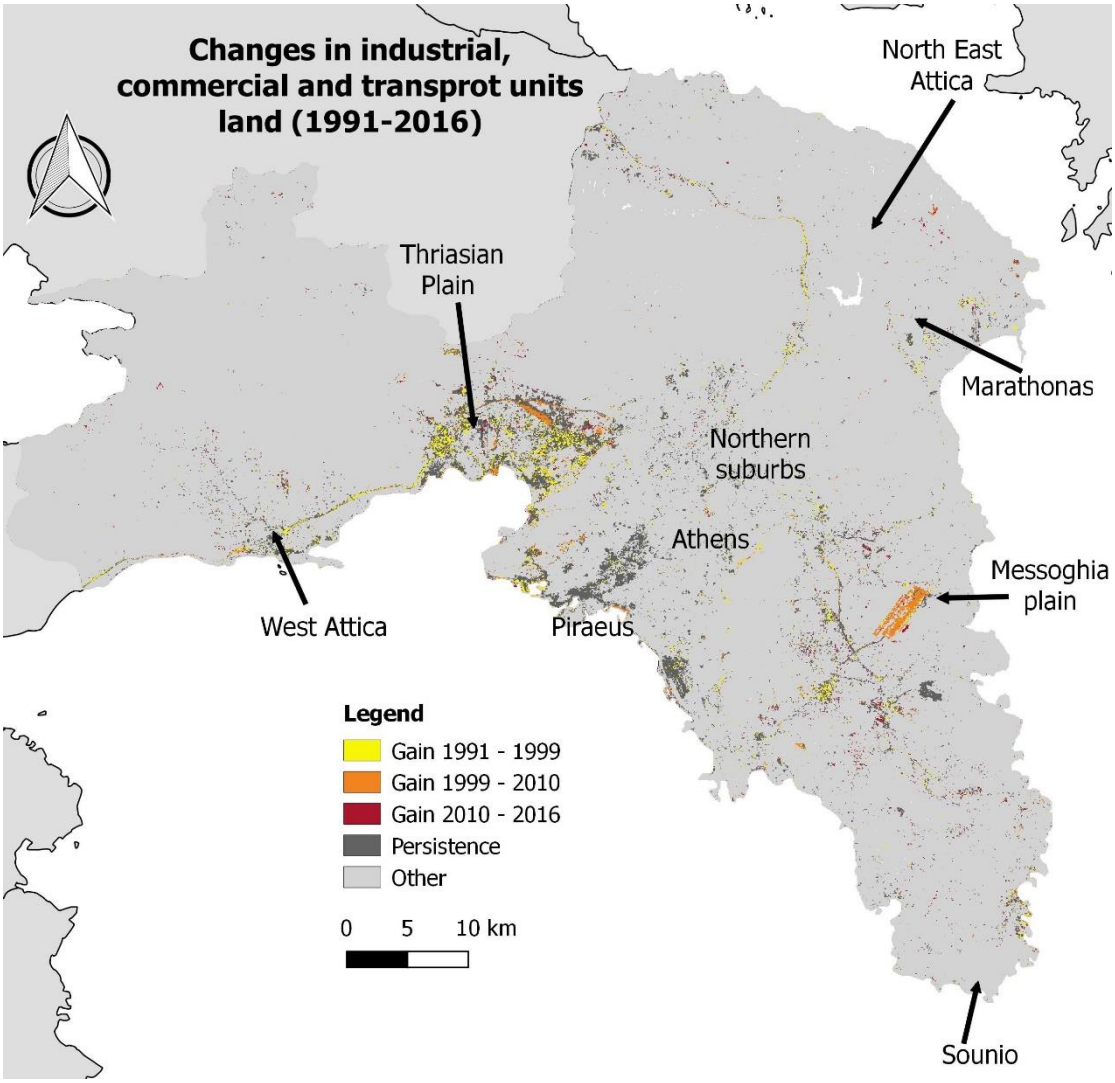


Figure 10. Industrial expansion observed between 1991-2016.

This is primarily due to the socio-economic and political changes occurred in each of these localities, encouraging population shifts that in turn affected the housing (both residential and second home) demand as well as the demand for construction sites and thus leading to different land uses competing for the available land. Urban development commonly associates with an increase in the market value of nearby lands uses and this why usually, residential, industry and commercial uses tend to dominate over less profitable lands in the bid for space. Consequently, owners of proximate agricultural lands may welcome nearby urban development, considering it as a way to expand the value of their properties. Moreover, land owners of Attica, were encouraged by the weak regulation mechanism. The only available land for development at low costs, in the case of Attica, was agricultural land and in conjunction with the loose regulatory mechanisms, land owners switched their land use into more profitable pathways.

Finally, Figure 12 depicts the changes related to the forests and natural areas land. As revealed also from the quantitative information (Figure 7), the forests and natural areas had a relatively slight decrease. Changes are located mostly in the urban periphery especially in the boundaries of already established urban agglomerations, in the northern part of Athens. Also, notable losses can be seen in the northern suburbs of Athens, in the boundaries of Hymettus mountain and in cape of Sounio. Most of these changes are associated with urban sprawl tendencies as well as forest fires which are frequent in the region and often accused of being human-intended due to land speculation.

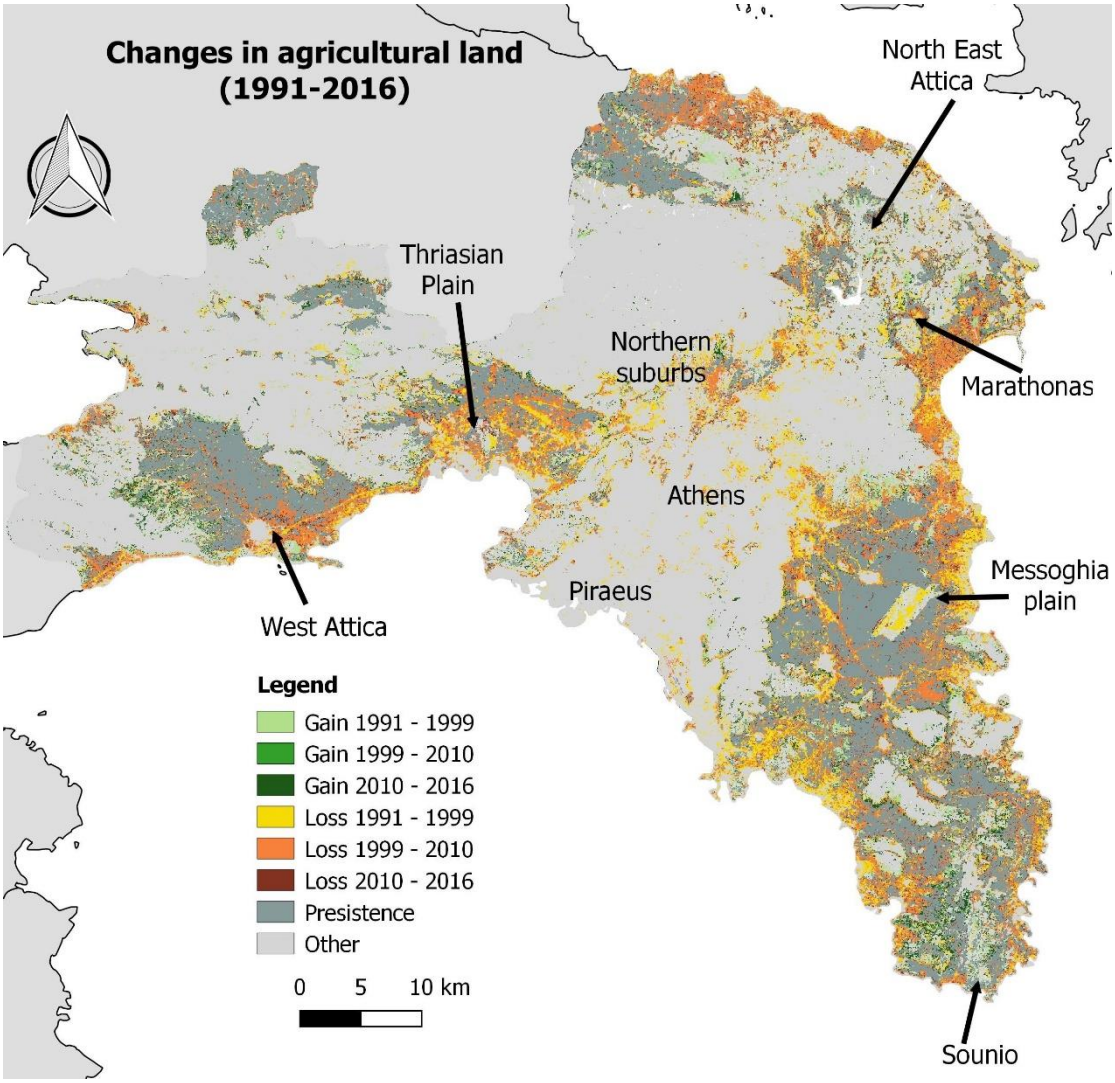


Figure 11. Agriculture loss observed between 1991-2016.

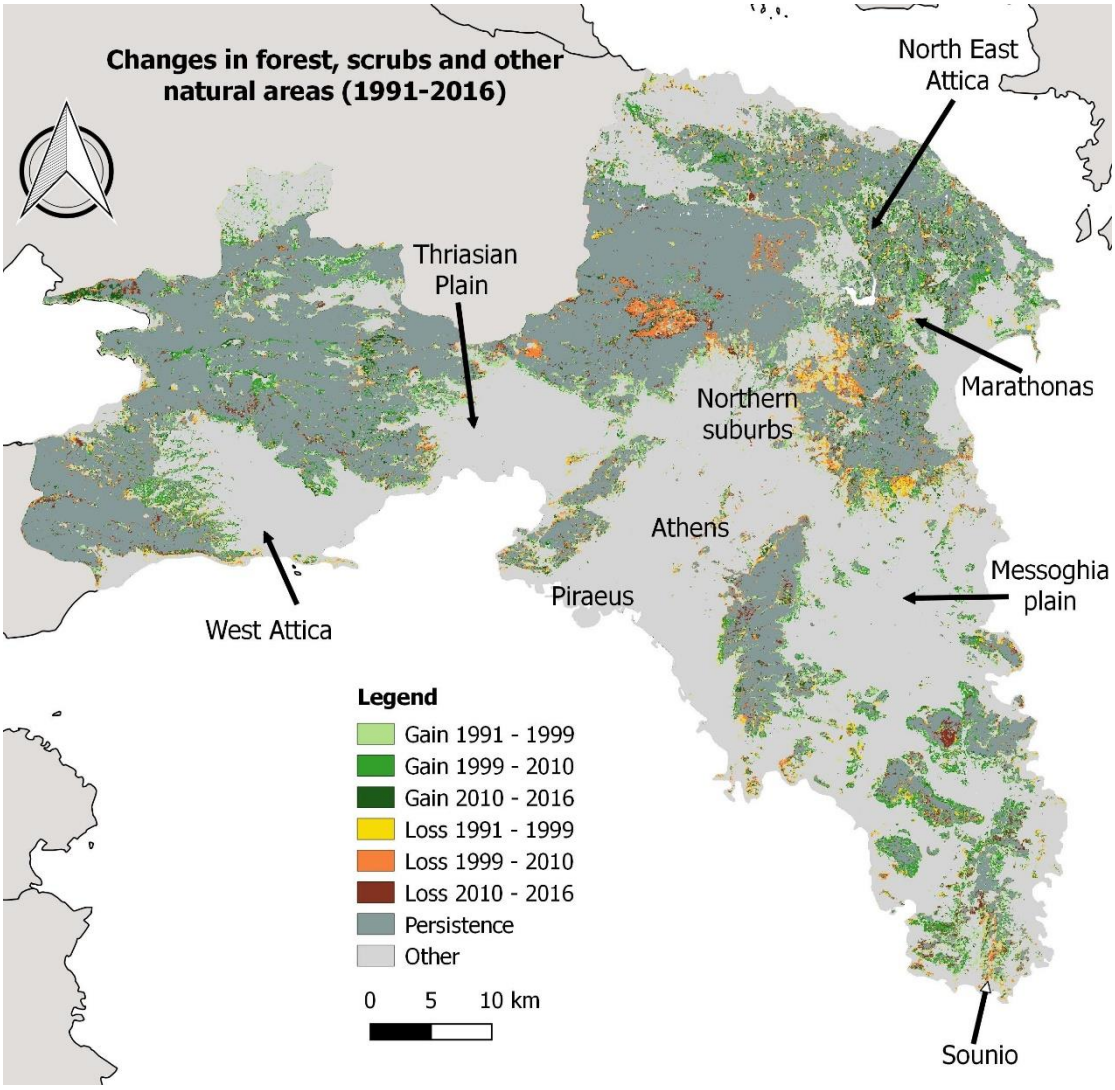


Figure 12. Forests and natural areas loss observed between 1991-2016.

6.3.2 Transition probability modeling performance

The transition probability surfaces were constructed using the RF modeling framework. The performance of the models was assessed using the Areas Under Curve (Figure 13). As can be seen, the algorithm efficiently handled 27 heterogenous factors derived from multiple sources and expressed in different scales, units and resolutions.

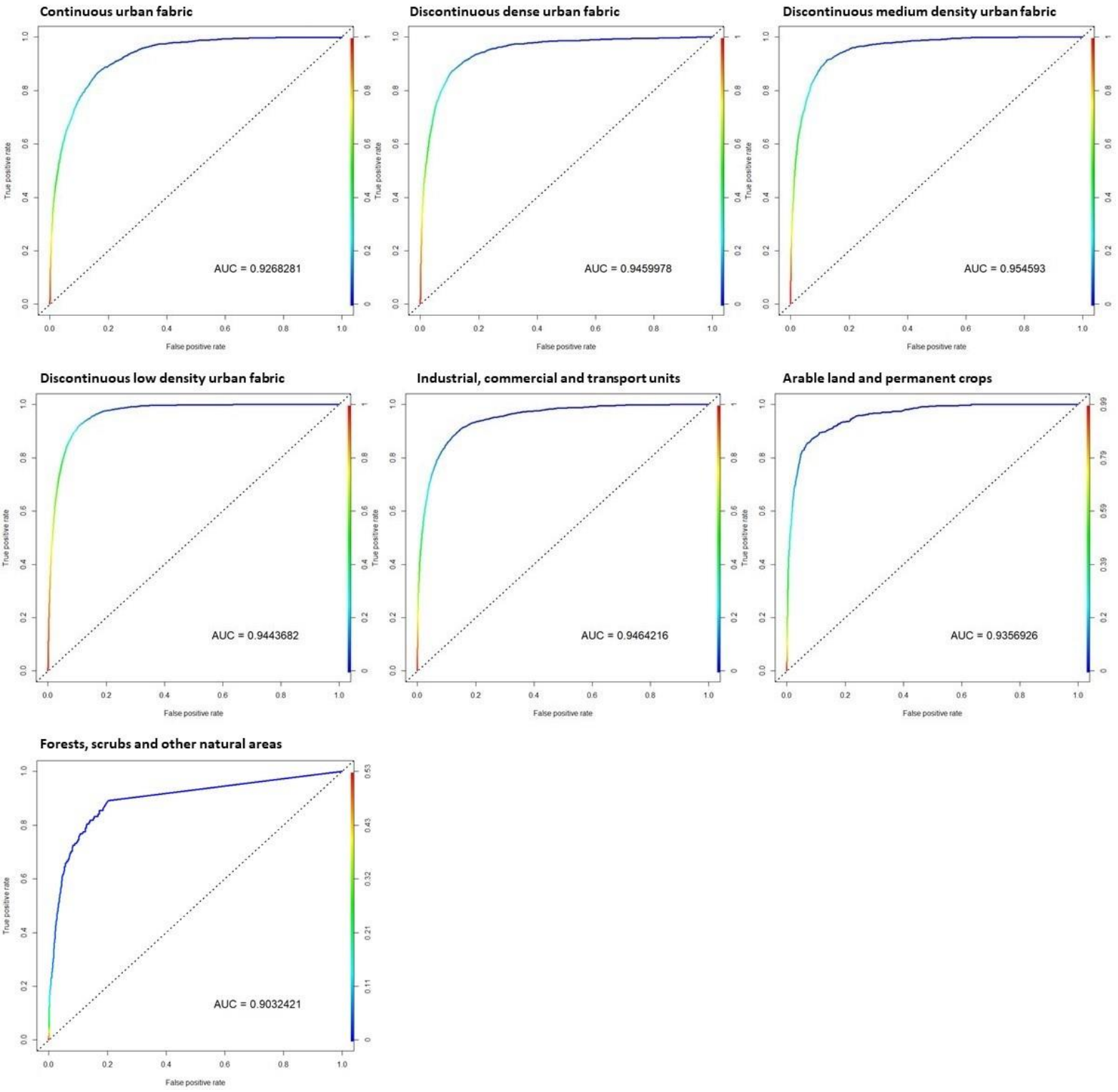


Figure 13. Performance of the transition probability modeling based on Relative operating characteristic curves (ROC) and area under curves (AUC) (LULC types are aggregated for demonstration purpose.

6.3.3 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 classification of satellite data. Figure 14 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.

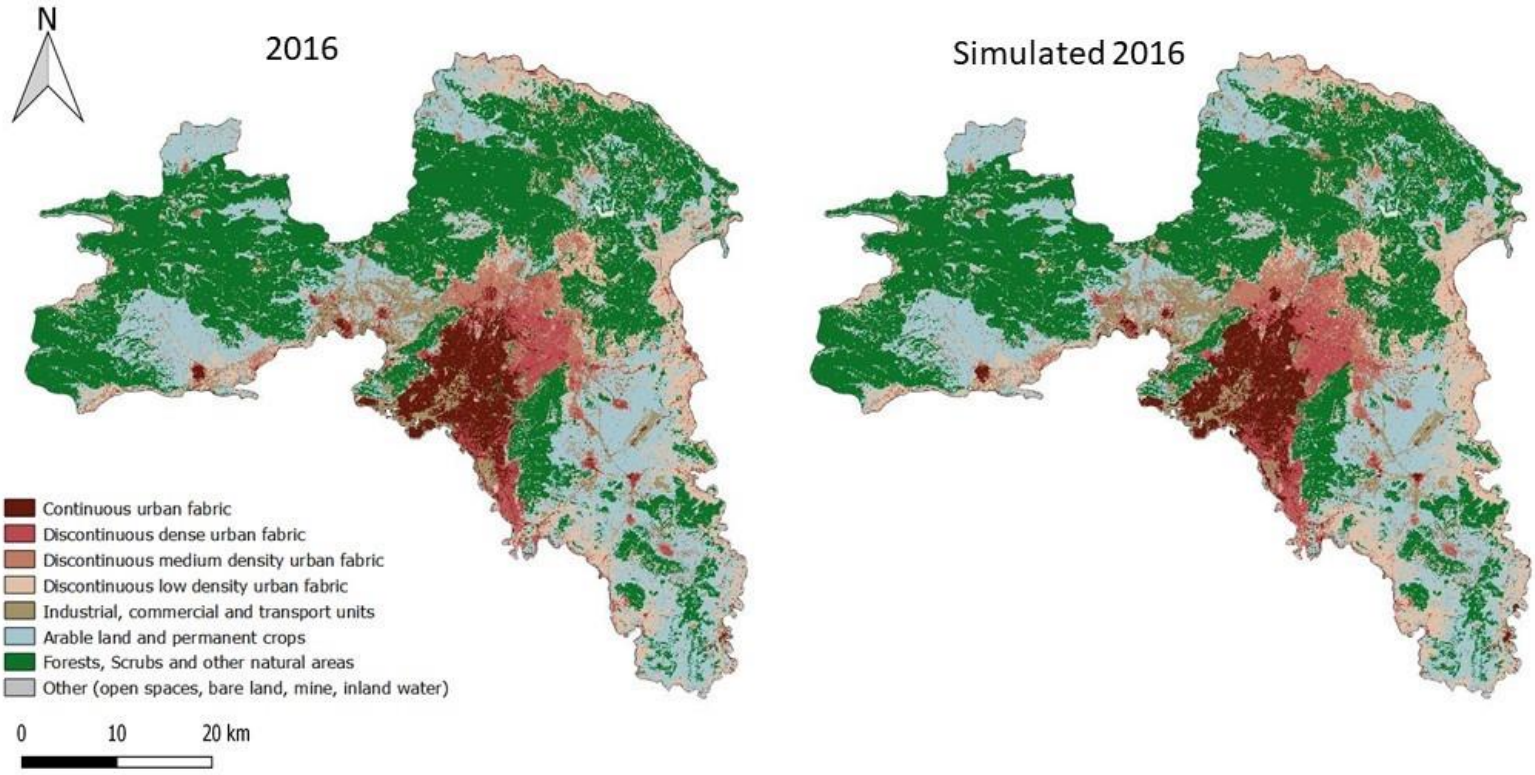


Figure 14. The simulated map of 2016 versus the observed of 2016 (reality).

Figure 15 illustrates the fuzzy similarity index computed based on the overlay of the two maps. The fuzzy similarity index evaluates the model's performance over a range of resolutions. 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. Again, the high scores in performance suggests that the suite of 27 predictor variables were used efficiently and the RF algorithm performed well with an adequate fit.

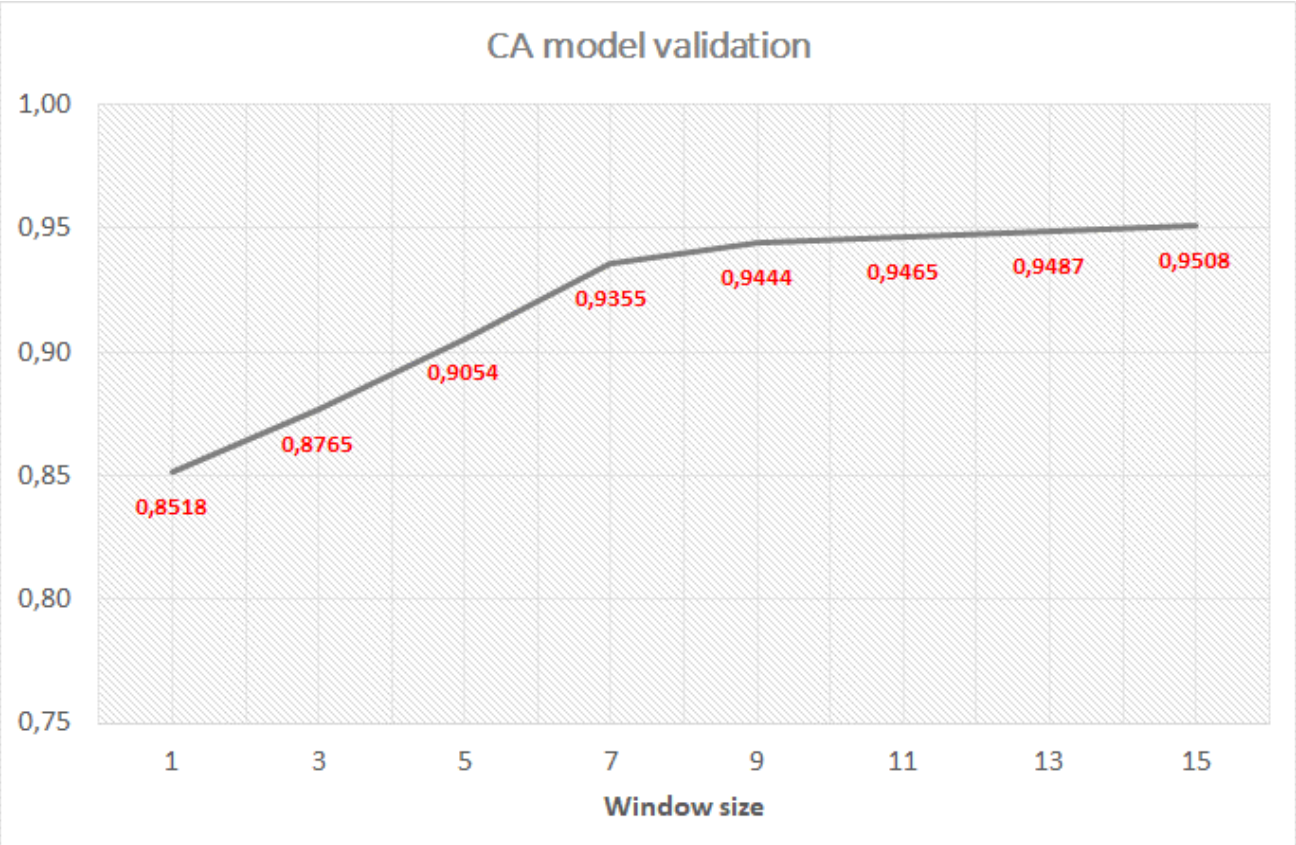


Figure 15. Multi-resolution evaluation of model fitting using the fuzzy similarity index.

Figure 16 depicts the components of agreement and disagreement between the simulated versus the observed maps. This type of accuracy assessment reveals information about the (i) observed change simulated correctly as change (hits), (ii) observed persistence (that is, LULC remained unchanged) simulated correctly as persistence (null successes), (iii) observed change simulated incorrectly as persistence (misses), and (iv) observed persistence simulated incorrectly as change (false alarms). Most importantly, the model predicted accurately the leap-frog development and this proves the added value of the Leap-frog development index and the extensive training of the RF model.

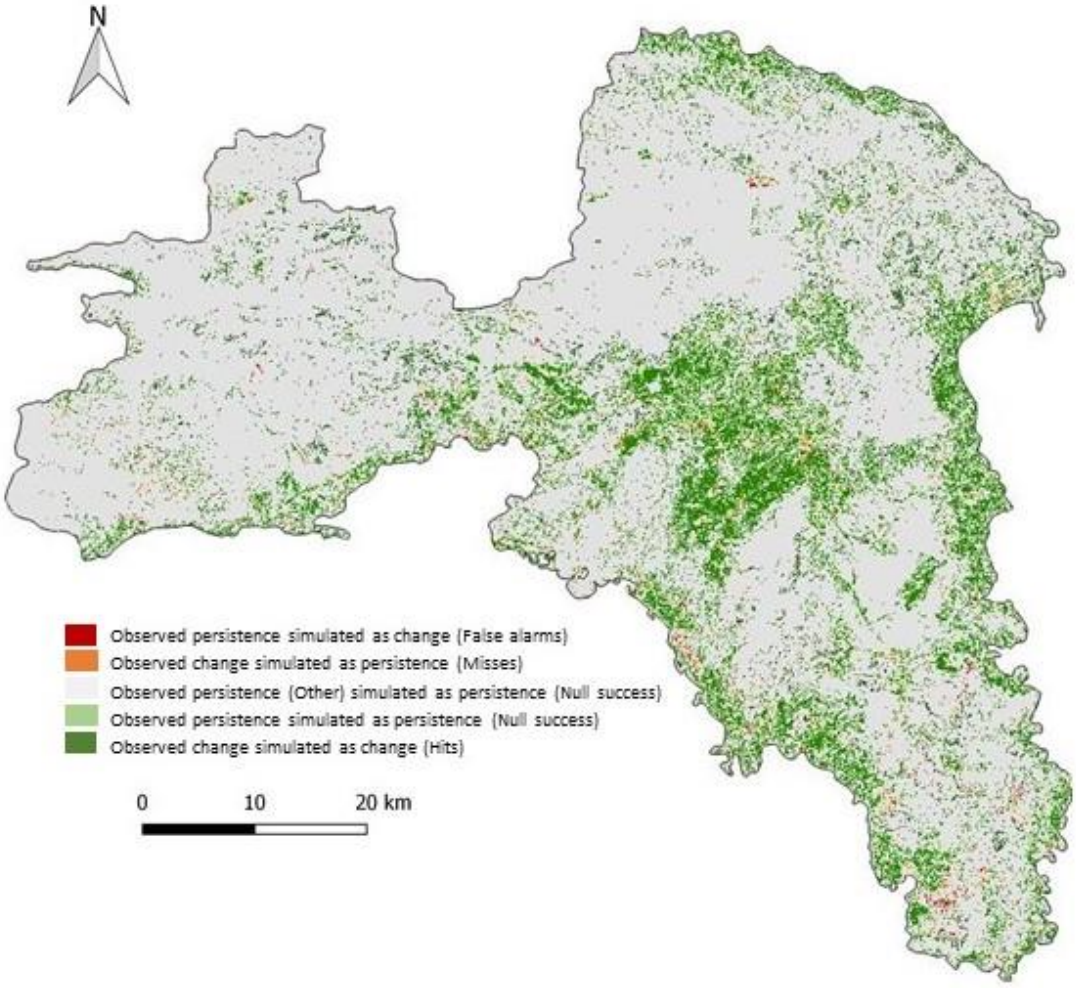


Figure 16. Result of cross classification between the simulated vs the observed map of 2016.

6.3.4 Factors contribution to land use/cover changes

The process of model calibration involved a critical step for the transition probabilities surfaces construction. In the form of spatial variables, 27 factors (Table 4) that were able to serve as proxies that describe the historical LULC patters were incorporated into a RF regression model. The variables spanned a range of different aspects, assumed to geographically explain the phenomenon and can be broadly categorized into territorial, socio-economic and land use factors. Given that the majority of transformations experienced in the case study, were related to urban and industrial categories, the list of variables was formed with an aim to incorporate in the model, spatial determinants that reflect that peoples' choices about residential and infrastructure location. For instance, territorial variables 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 flagged beaches as well as to areas of high nature value or urban green are also an adding value in pursuit of a better quality of life and aesthetics for both primary or second-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 while distance to education, health, public buildings and enterprises density serving as proxies to amenities. Demographic and socio-economic proxies such as changes in population density, employment and unemployment rates provide insights on the shifts in the socio-economic profile of the area (per municipality).

This suite of geographical data was derived from multiple different sources and represented in different scales, units and resolutions. The modeling scheme employing the RF regression, efficiently (Figure 13) handled the fusion of these data in order to produce the transition probabilities surfaces and to quantify each variable's contribution to the outcomes (Figures 17 & 18). Certain advantages for adopting the RF algorithm for data fusion (also demonstrated in chapters 1,2,3 & 4) were notable in this case. First, RF efficiently handled different types of data expressed in both categorical and continuous variables facilitating the incorporation of any type of inputs. Second, RF proved insensitive to overfitting and collinearity of inputs was not an issue. Third, normal distribution of inputs was not a prerequisite and the algorithm performed well when coping with non-linear relationships between response and predictor variables. Fourth, RF offers meaningful metrics about the importance of each predictor variable. To quantify the actual importance and contribution for each of the 27 predictor variables, two metrics, the Mean Decrease Accuracy (%IncMSE) and the Mean Decrease Gini (IncNodePurity) were computed.

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 score, tend to have nodes with high purity which is a measure of model's homogeneity. As can be seen in Figure 17, road density, enterprises density and elevation contributed the most for changes related to dense urban fabric, while 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.

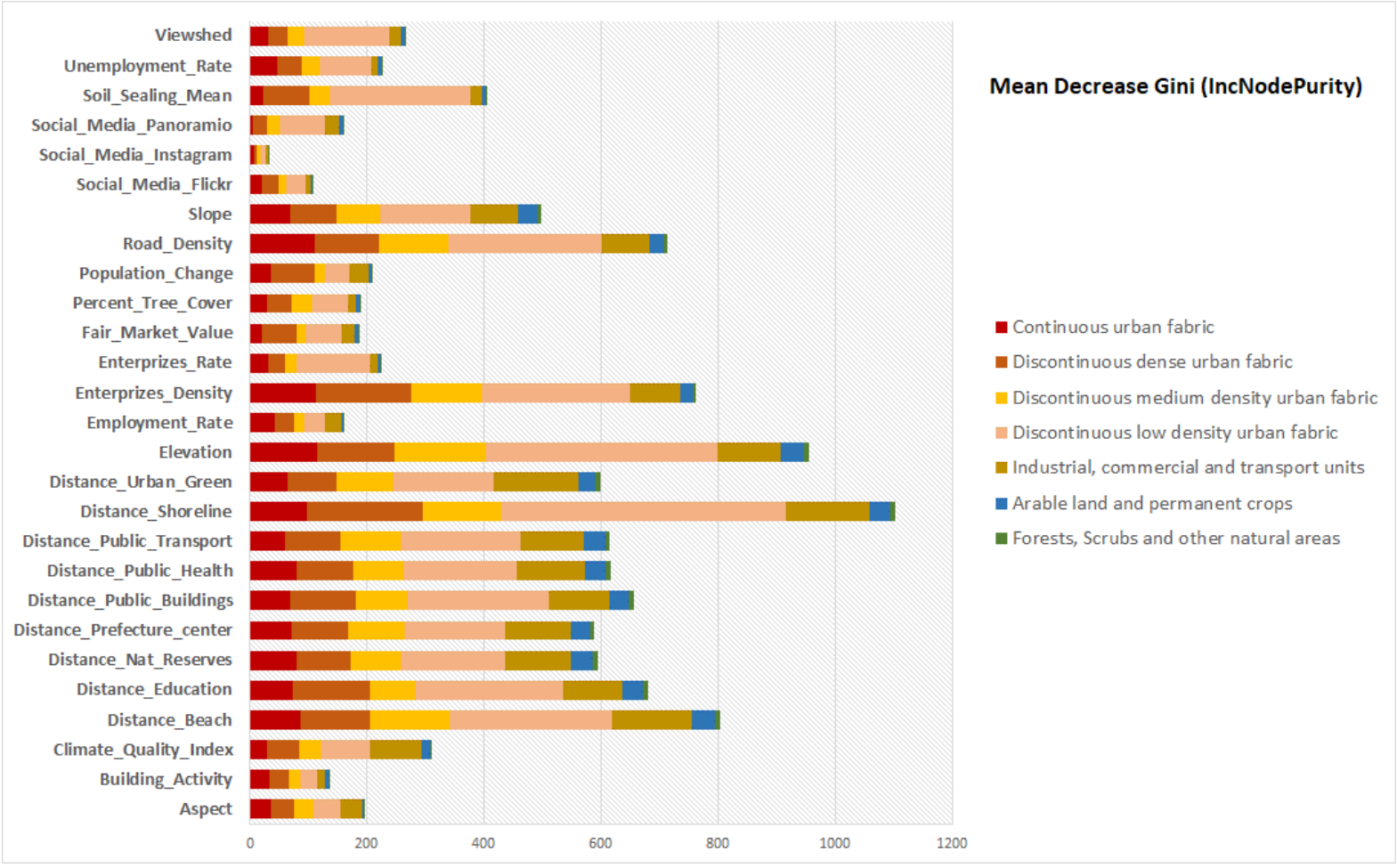


Figure 17. Mean Decrease Gini (IncNodePurity) as assigned by the RF regression algorithm

The mean decrease in accuracy is a score that informs about how much the accuracy decreases if a variable would be excluded from the model. Therefore, the larger the value of mean decrease, the higher the importance of a variable is. As shown in Figure 18, 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.

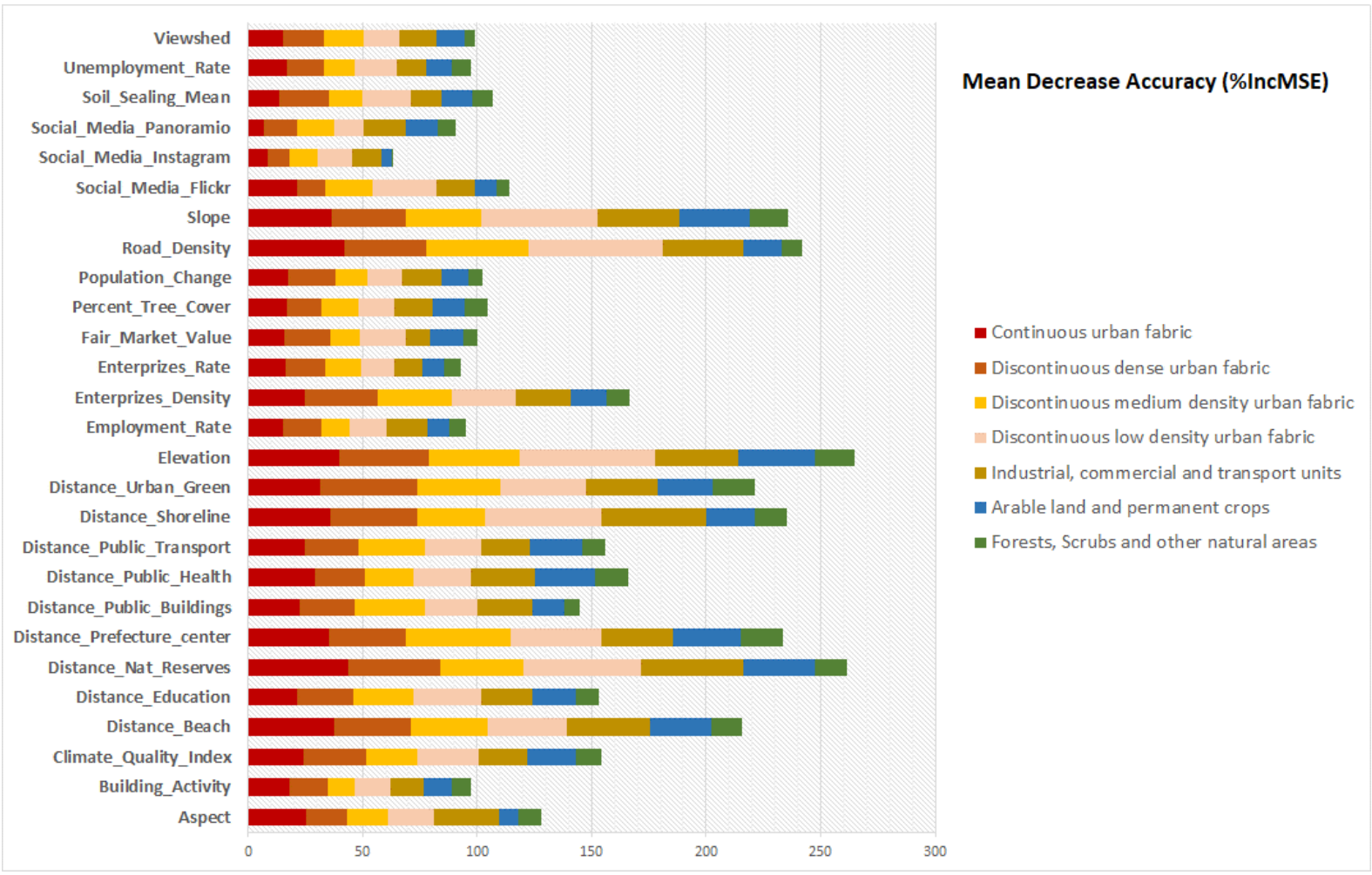


Figure 18. Mean Decrease Accuracy (%IncMSE) as assigned by the RF regression algorithm

6.3.5 Multi-resolution sensitivity analysis

The simulation models implemented in order to explore alternative trajectories of LULC changes that will occur in Attica region by 2040, under the three scenarios. To gain in robustness of predictions and to obtain unbiased results in regard to the spatial resolution of inputs, the modeling process was subject to a multiple resolution sensitivity analysis. To do so, besides the nominal resolution of 30m, the whole process was repeated resampling all inputs, required for the CA modeling, at 100m, 250m and 500m. It is worth noting that all models were identical in terms of inputs and parameters. Transition probabilities were re-modeled and the landscape metrics were re-calculated and re-introduced to the model for each case. The outputs of each model, implemented in different resolutions, were overlapped and cross classified in order to produce the final map for each of the three scenarios. Figures 19-21 depict the results for the medium, high and low development scenarios respectively. As can be observed, the models yield similar patterns for each scenario but as the resolution increases, the patterns tend to become more aggregated and smaller patches of change tend to be lost. This provides evidence that the technical characteristics and quality of inputs 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 analogously to the spatial resolution, and the patterns revealed in the results will become less informative. Figure 22 illustrates the correlation between transition probabilities for continuous urban fabric per different spatial resolution. It is another evident of the influence the spatial resolution has on various consecutive steps of the modeling process. The values were collected at the location of 1000 random samples, dispersed across the transition probability surface generated at 30m, 100m, 250m and 500m. Gradually as the pixel size increases, the correlation between the transition probability surfaces tend to decrease. The higher correlation value can be observed between the 30m and 100m pixel size while the 500m is the least correlated with the rest.

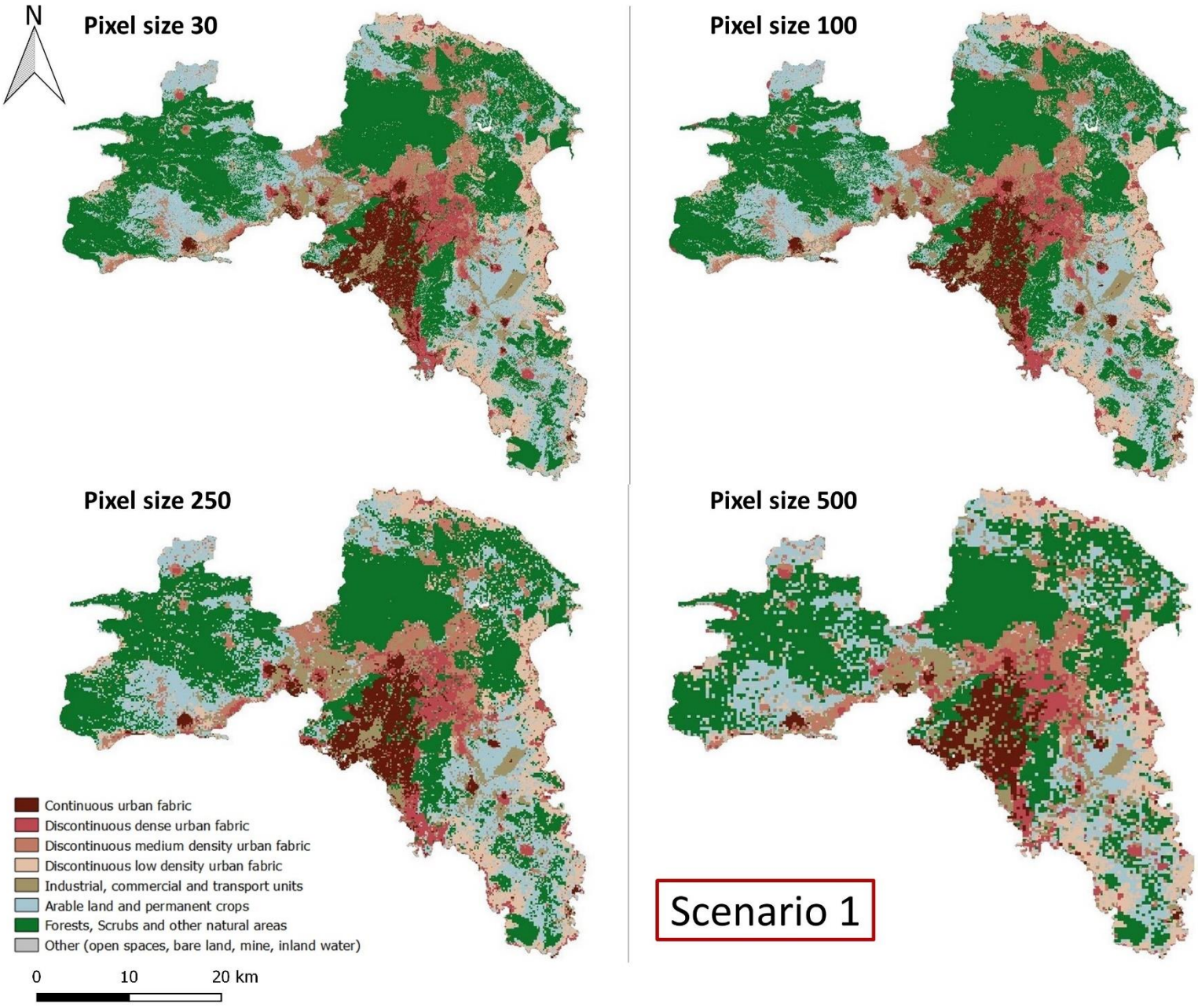


Figure 19. Resulting map of "medium development" scenario at various spatial resolutions

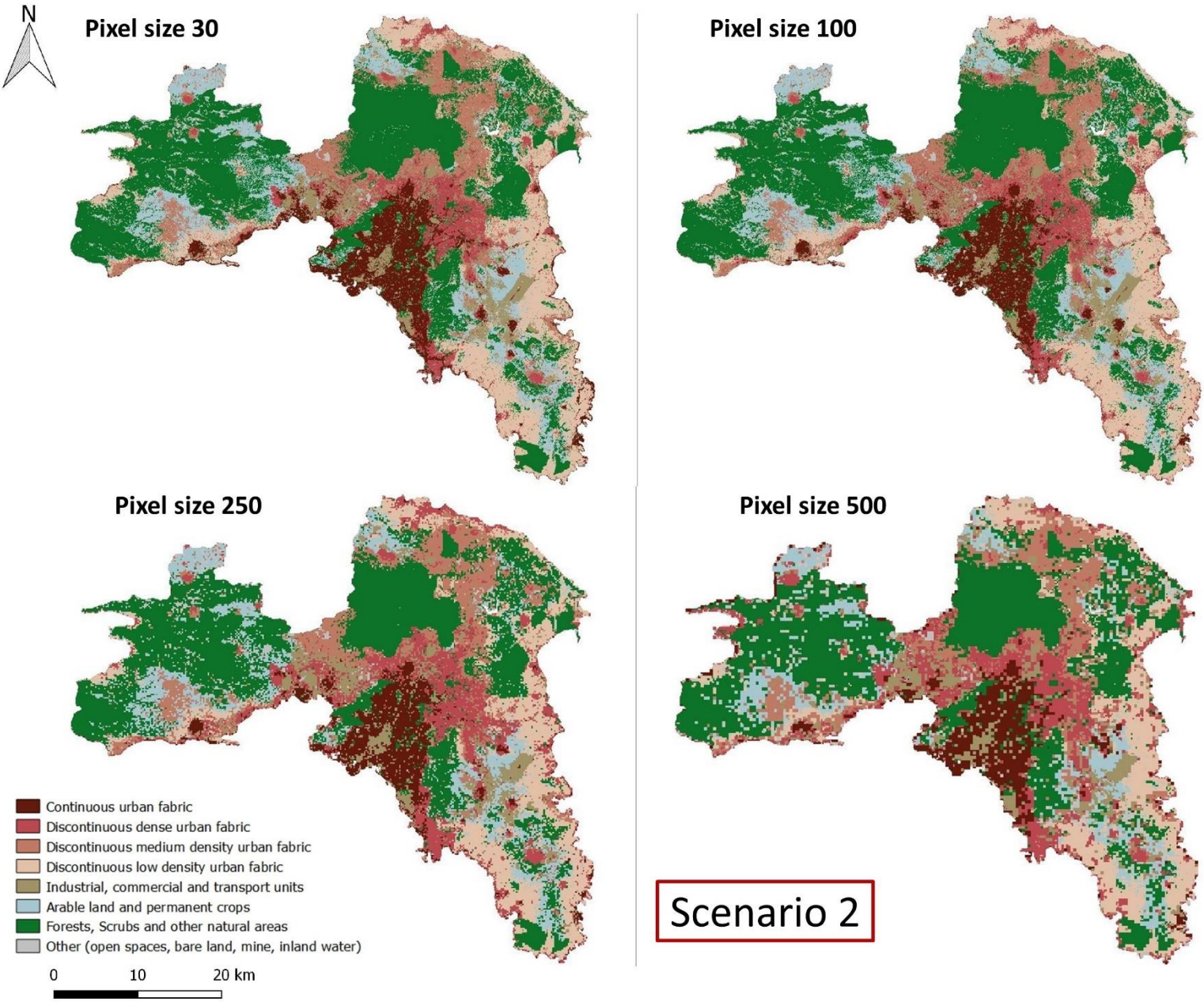


Figure 20. Resulting map of "high development" scenario at various spatial resolutions

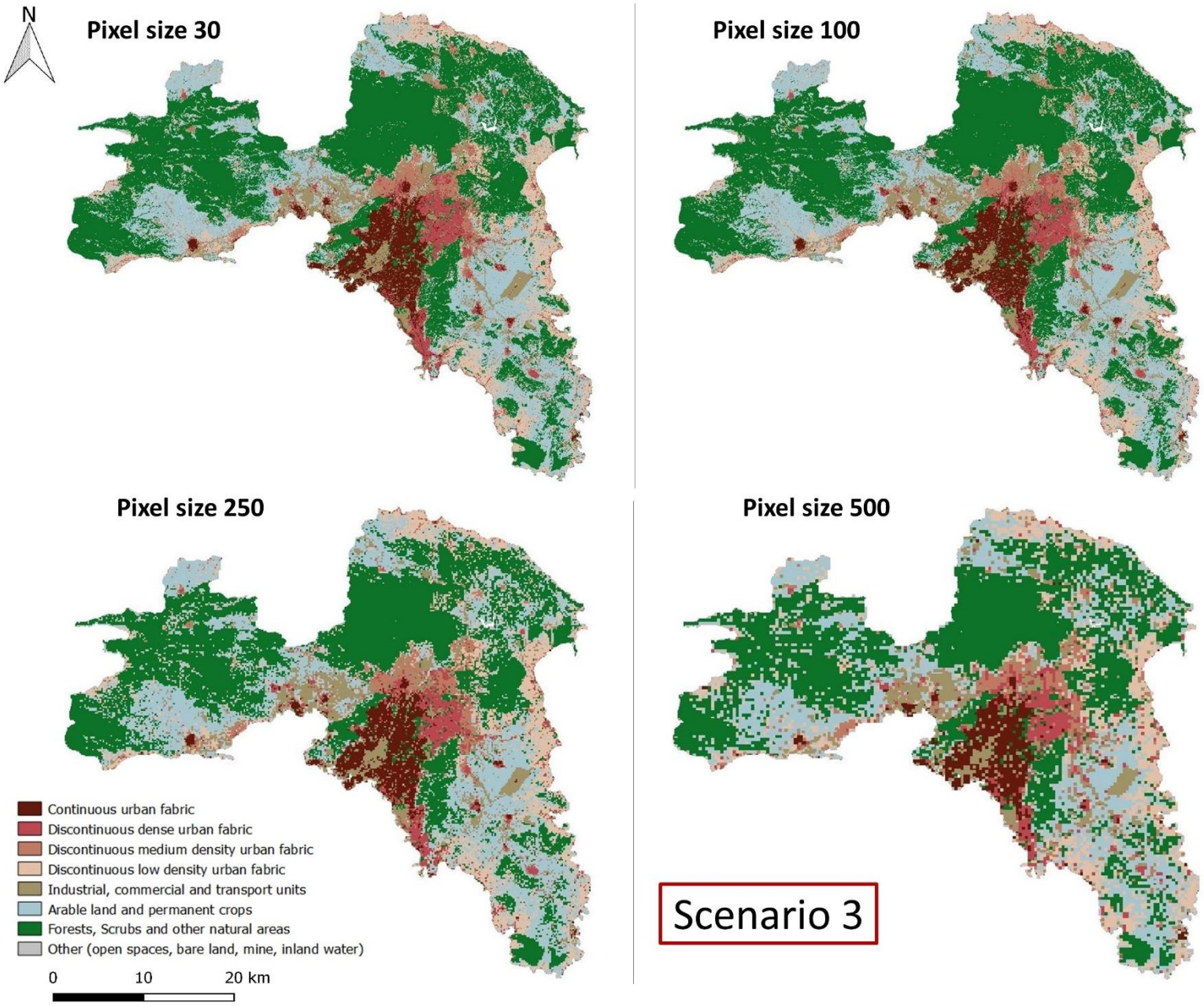


Figure 21. Resulting map of "low development" scenario at various spatial resolutions

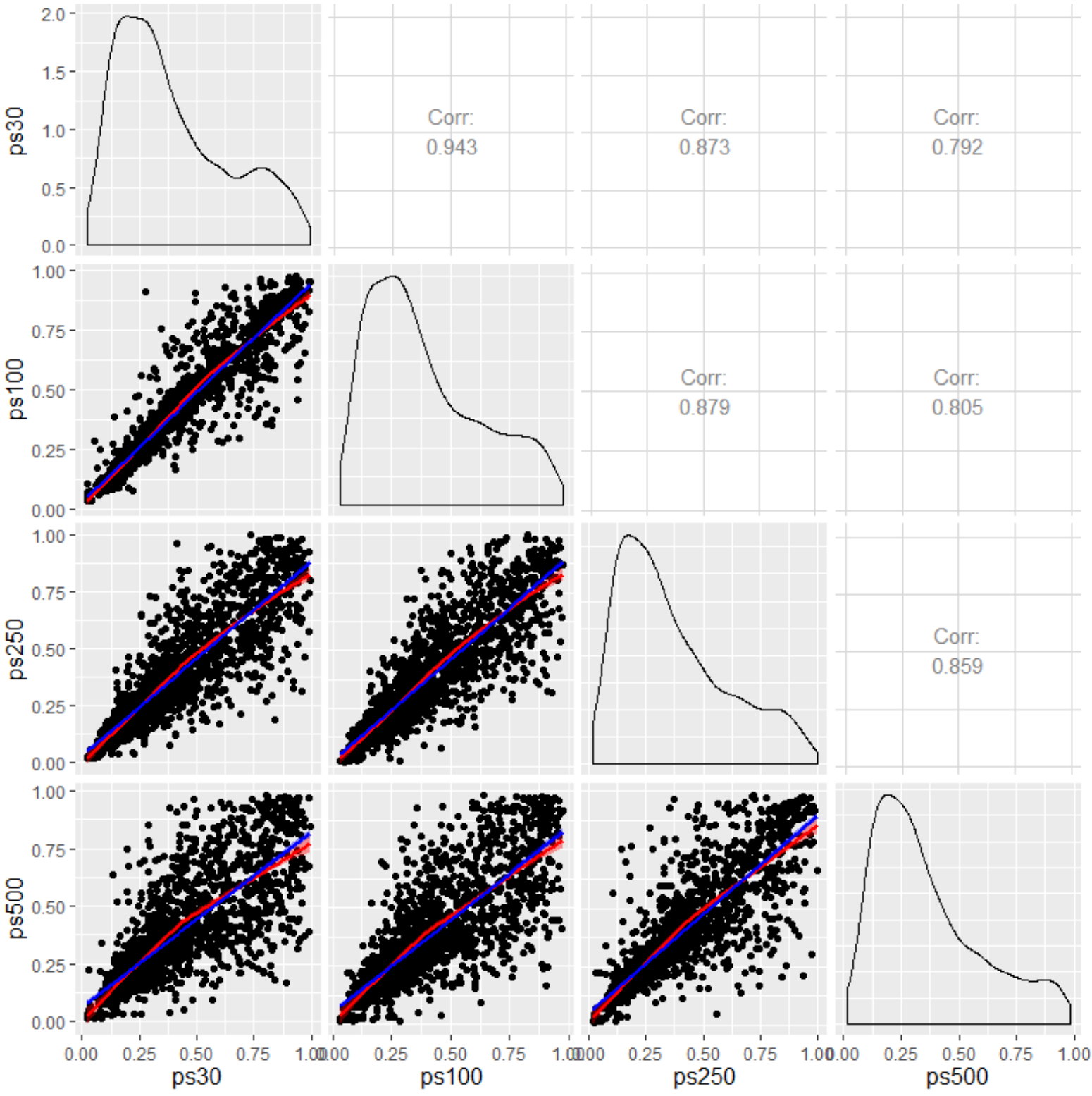


Figure 22. Pearson correlation between transition probabilities for continuous urban fabric category. The values derived from 1000 random samples, dispersed across the study area

6.3.5 Final results

After the multiple resolution sensitivity analysis, the final maps illustrate the LULC synthesis and configuration of Attica region in 2040. Figures 23-25 depict the land use/cover changes projection under the three scenarios and Figure 26 provides a quantified insight to the final results.

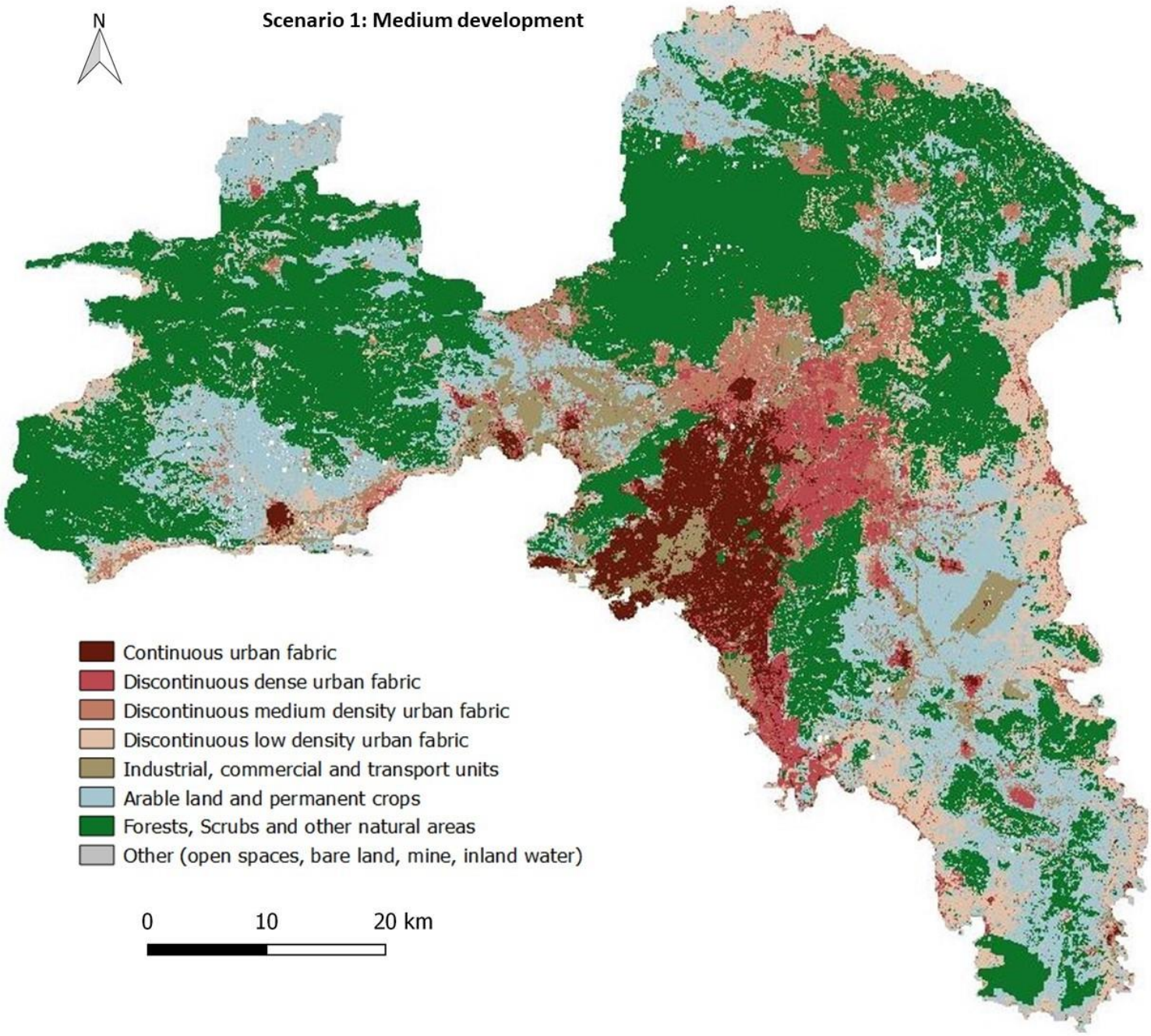


Figure 23. LULC spatial configuration simulated for 2045, under the Medium development scenario.

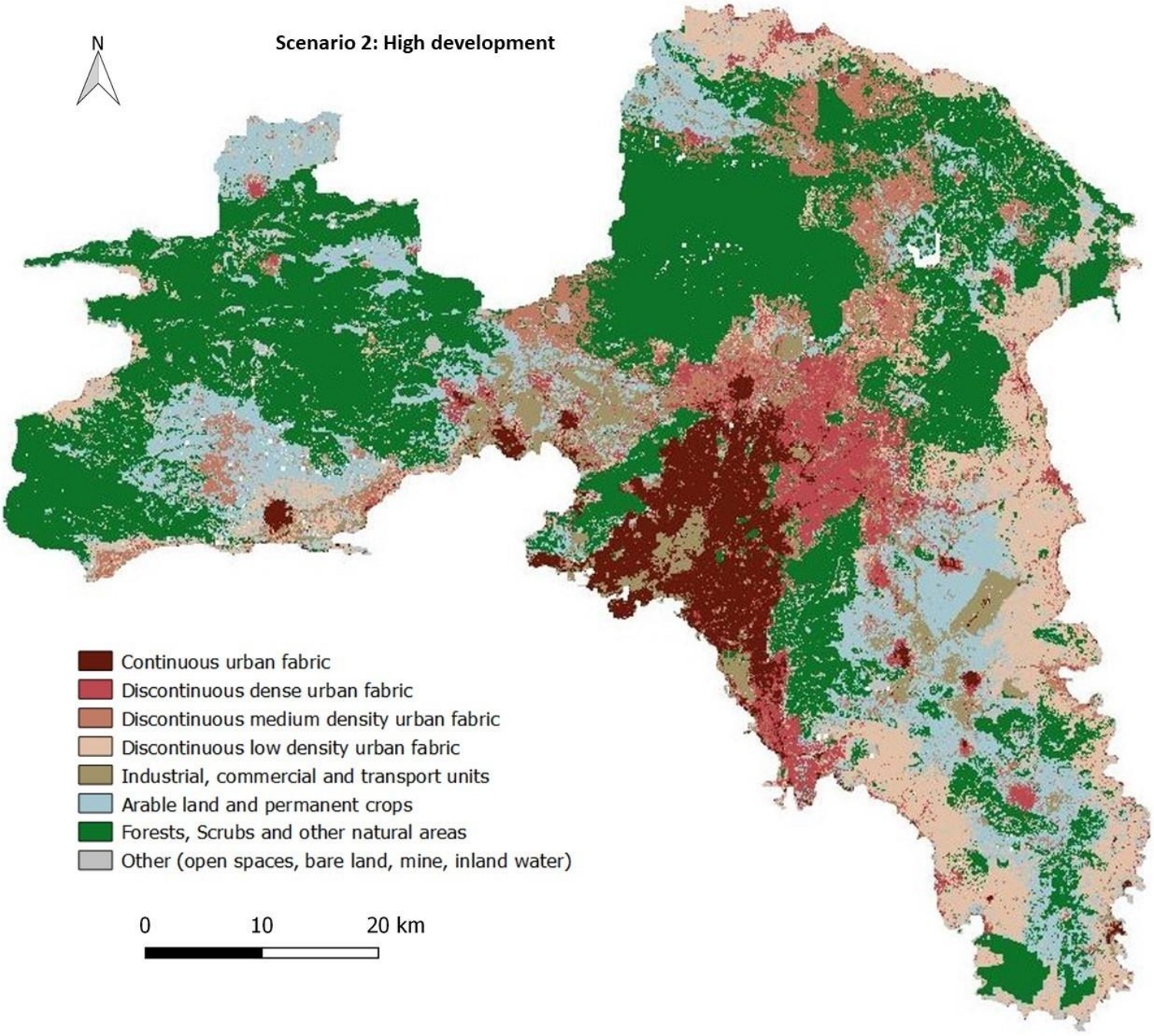


Figure 24. LULC spatial configuration simulated for 2045, under the High development scenario.

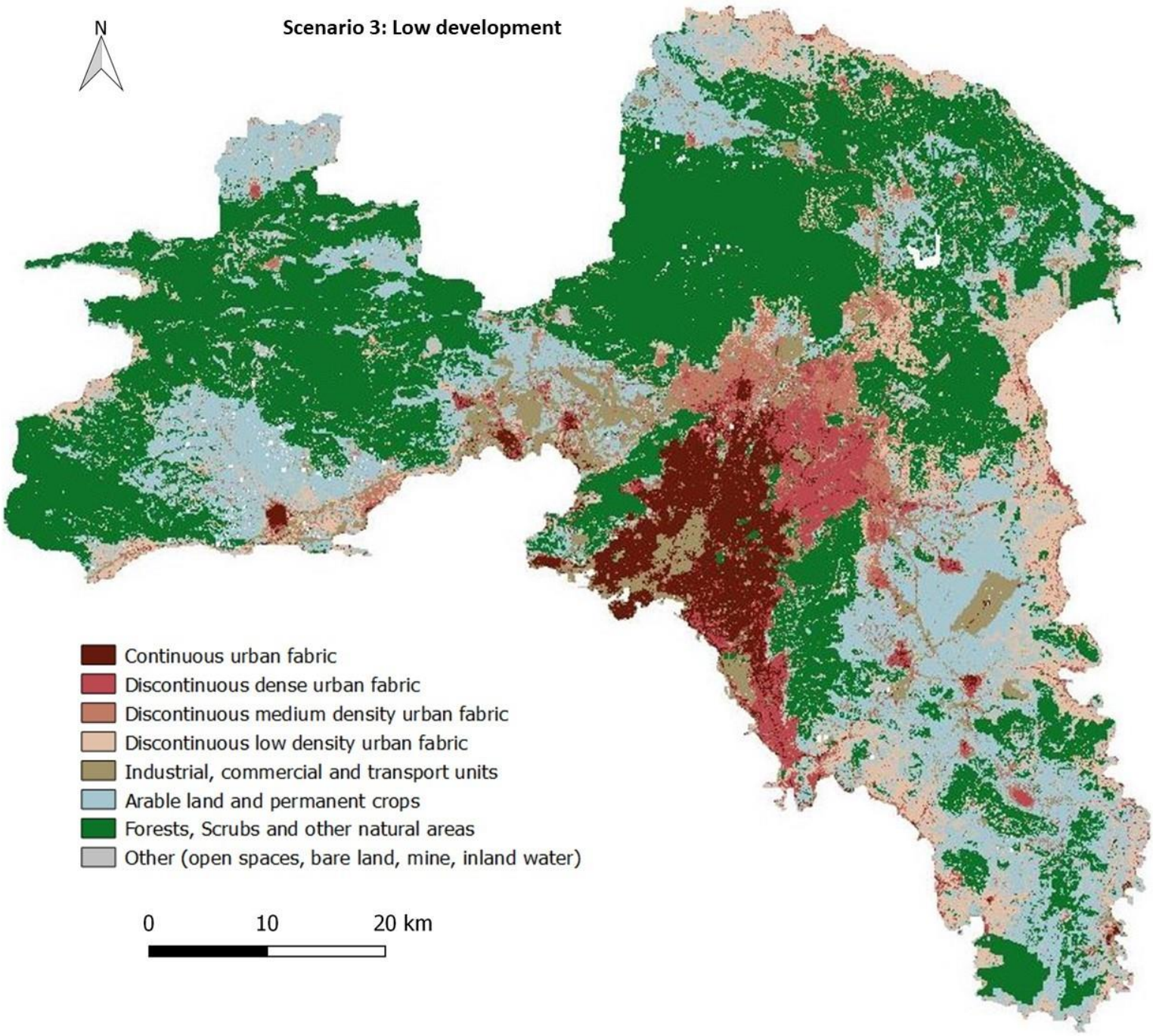


Figure 25. LULC spatial configuration simulated for 2045, under the Low development scenario.

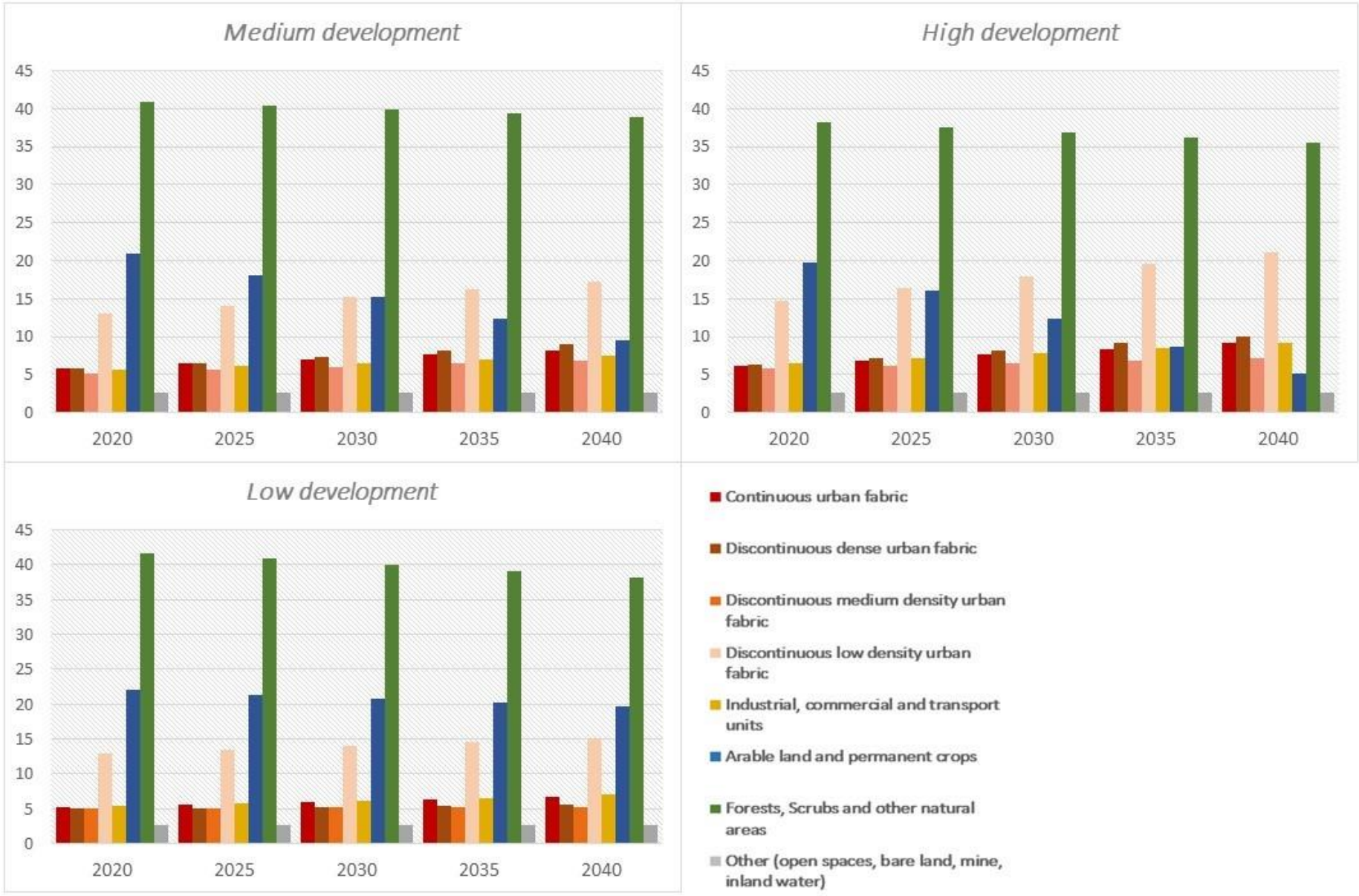


Figure 26. Rates of LULC simulated for 2040 in a 5-year step, under the three different development level scenarios

Under the *medium development* scenario and with a pace of development equivalent to that of 1991–1999, the artificial surfaces are expected to expand predominantly at the expense of other less profitable land uses. Urban areas are expected to reach 41%, 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 areas are expected to decline from 23.5% in 2016 to 10% in 2040. Most changes will occur along the waterfront and to the urban periphery of Athens, particularly at the Messoghia, Thriassian, Marathonas, Oropos and Sounio areas. Additionally, in these areas, pre-existing urban and industrial clusters portray a tendency to infill, to become denser and to expand considerably, ending up almost connected with the Athens urban fringe, especially the northern suburbs. The waterfront especially in Messoghia, Marathonas, Oropos and Sounio, is also expected to exhibit remarkable changes as the existing towns obviously tend to expand and become densely infilled while the shoreline almost will almost convert to a large and solid low density urban patch. Leap-frog development is also expected to increase sharply around junctions of infilled areas, main roads and already established urban patches.

Under the *high development* scenario, where the pace of development reflects the period between 1999 and 2010, the artificial surfaces are expected to increase remarkably and occupy more than half of the total surface of Attica region (56.7%). The urban use, is expected to occupy an area of almost 48% in 2040 which can be translated to a 21% increase while at the same time the agricultural areas are expected to decrease by 18%, occupying only 5.2% of the total area (Figure 24). Discontinuous low density urban fabric will reach a high peak of almost 21% of the total area, while continuous dense and discontinuous high density urban fabric are expected to reach 9% and 10% respectively. All these accelerated landscape transformations are expected to occur throughout Attica region leading to a mosaic of mixed land uses of which the settlements dominate. Again, most changes are observed along the waterfront and to the surroundings of Athens, particularly at the Northern suburbs of Athens, Messoghia, Thriassian, Marathonas, Oropos and Sounio areas. Pre-existing urban and industrial clusters will infill, become denser and expand considerably. The urban fringe of Athens will be channeled to all directions and is expected to completely connect with the waterfront existing patches (Marathonas, Messoghia, Sounio, Thriassian, west Attica) and the northern suburbs up to Oropos area forming a large and solid urban patch with

low, and at places, medium density. The effect of proximity to the shoreline as well as the amenity driven residential tendency is especially pronounced in the outer city due to the residential preferences. 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.

Under the *low development* scenario, an increase of approximately 6% in the area occupied by artificial surfaces is also expected but with considerably lower magnitude, in terms of landscape structure and composition, compared to the previous scenarios. For instance, the discontinuous low density urban fabric is expected to occupy 15% of the total area, which is only a 3% increase from 2016. Continuous dense and discontinuous high density urban fabric are expected to reach 6.5% and 5% respectively. Following the pace of the "recession period" between 2010 and 2016, expansion is observed throughout the study area but in lower extent and in a more compact form. The expected changes will mostly occur around the road network and the waterfront (particularly to areas of the eastern and northern parts of Attica). Already existing patches of urban areas appear infilled rather than expanded while leapf-rog development is also expected mostly around areas characterized by favorable conditions such as proximity to Athens as well as to the town centers and around the motorways connecting Athens to the periphery. Regarding the urban density, slight changes are expected in the northern suburbs of Athens.

6.4 Conclusions

Methodologically, the semi-automated sampling used for classifying the LULC categories proved efficient and reduced significantly the limitations regarding the resources consumption. The spectral controlling approach also played an important role in building robust and accurate models for each year as well as to the uncompromised backwards automated training strategy. For the modeling part, in accordance with the approach described in chapter 4, 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 modeling, allows the efficient fusion of qualitative and quantitative data derived from multiple sources, with different nature in terms of scale and origin,

overcoming the collinearity and distribution issues. The predictors incorporated in the models proved capable to spatially determine the phenomenon while the incorporation of the Leap-frog index, at the regional level this time, assisted the models in the face of sprawl detection and, in turn, prediction. In this approach, a total of 18 distinct transitions were identified and equal transition probability surfaces were generated. Their combination in a CA modeling environment seemed challenging and required intense training and calibration through trial and error. Currently, most models can only simulate limited possible transitions due to complexity in definitions, attributes and transition rules. But in reality, even to a locality, different LULC dynamics occur simultaneously and affect each other. Thus, a comprehensive outlook of these processes is much more effective in order to determine realistically the future trajectories. The interactions and competition among different type of LULC was explored by using a simple yet effective competition mechanism, in which the combined probabilities are manipulated as a single layer stack containing all the probability surfaces. Each layer represents one single possible transition while each cell contains 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 landscape metrics, that fed the two complimentary sub-models Patcher and Expander, by taking into account actual parameters of the study area. The adoption of the fuzzy similarity index at multiple resolutions for assessing the models' 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. Finally, the approach presented in this chapter provided results that are insensitive to spatial resolution bias, after implementing the multiple resolution sensitivity analysis. Since the modeling approaches generate outputs that are more or less driven by the parameters and characteristics of input data, the results obtained by this approach and the patterns demonstrated, are consistent to all pixel sizes and thus insensitive to the effect of pixel size.

This chapter demonstrated an integrated approach to explore potential future LULC dynamics in the Attica region under three scenarios that reflect different economic performances and policy options. The third scenario which is linked to low development can be translated to the so-called business as usual scenario where the

current economic functions continue relatively unchanged and the financial crisis persists and keeps the building demand and supply at low levels. Results obtained from the medium and high development scenarios, can be translated in various ways. Apart from the straightforward research question of how LULC of Attica will be structured and composed under potential economic development rates, it can also be translated to how Attica would look like in case the economic crisis would be sidestepped. Another important aspect could be in regard to the absence of a regulation mechanism and the permissive and weak overarching spatial planning framework. The results obtained can be valuable in gaining insights and visualizing the outcomes of economic development goals that take precedence over virtually all other spatial planning priorities. They can be informative about the significant negative environmental and cultural externalities, carried by each alternative economic performance reality and land-use planning context and choice. The results stress that the negative consequences can undermine the very economic prospects and the sustainability of the area. The various ways the results can be interpreted, makes the scope of the findings to become wider and forms a broader foundation for debate. In the absence of intervention and adequate regulation, the first two scenarios demonstrate continued expansion of urbanization-driven development that is expected to compete with, and likely consume, a large amount of areas occupied by agricultural land uses and natural areas, throughout the region. The magnitude and distribution of development, demonstrated in both scenarios, can diminish the ecological and cultural equilibrium of the region, and this finding underscores a significant tradeoff.

References

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

Chen, X., Chen, J., Shi, Y. & Yamaguchi, Y. (2012). An automated approach for updating land cover maps based on integrated change detection and classification methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 86–95.

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.

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. (2018). Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: The case of Athens. *Applied Geography* 90, 134-144.

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

Kim, D-H., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S. et al. (2014). Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sensing of Environment*, 155, 178-193.

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.

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

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.

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. and 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.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brock- Mann, C., & Defourny, P. (2014). Automated training sample extraction for global land cover mapping. *Remote Sensing*, 6(5), 3965–3987.

Rodríguez-Galiano, V.F., Ghimire, B., Pardo-Igúzquiza, E., Chica-Olmo, M. and Congalton, R.G. (2012). Incorporating the Downscaled Landsat TM Thermal Band in Land-cover Classification using Random Forest. *Photogrammetric Engineering & Remote Sensing*, 78(2), 129–137.

Rodriguez-Galiano, V.F. and Chica-Olmo, M. (2012). Land cover change analysis of a Mediterranean area in Spain using different sources of data: multi-seasonal landsat images, land surface temperature, digital terrain models and texture. *Applied Geography*, 35, 208–218.

van Zanten, B.T., Van Berkel, D.B., Meentemeyer, R.K., Smith, J.W., Tieskens, K.F. & Verburg, P.H. (2016). Continental-scale quantification of landscape values using social media data. *PNAS*, 113(46), 12974–12979.

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.

Chapter 9: Conclusions

This dissertation aimed to explore part of the complexity that characterizes the LULC changes system and accomplished to meet the primary objectives and the scientific challenges that emerged during the process. A range of state of the art methodologies that lie in the Geoinformatics, satellite remote sensing and spatial modeling disciplines were assembled in order to build an integrated methodological framework for detecting historical changes, delineating and quantifying the factors and sub-factors that drive these changes and sketching alternative future LULC trajectories.

The integrated methodological framework was devised in order to sufficiently i) take into account the multiple scales involved in LULC systems, 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) detect LULC changes in a temporal resolution that enables the identification of uneven patterns throughout the study period which in turn enables the sound delineation of scenarios, iv) take into account socioeconomic, biophysical, legislative and land use factors spanning a broad spectrum of LULC change driving forces and finally to v) provide results that are subject to sensitivity analysis and unbiased to the technical details of inputs.

To answer the research question of how can heterogenous data, be efficently combined in a LULC modeling framework, it is demonstrated in two case studies by incorporating in the modeling 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, the validity, 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 and their transferability is limited. Data expressed at coarse scales might hold information and patterns that are invincible at more detailed scales and vice versa. 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. This dissertation exploited all possible resources and efficiently fused and integrated the available multi-scale and multi-resolution data. ii) The models' calibration was based on a fuzzy similarity index that considers the similarities of neighborhood in a growing

moving window and not a strict cell by cell comparison. To overcome the fact that the spatial resolution of inputs plays a major role to the outcomes of models, a multiple resolution sensitivity analysis was also applied. iii), The simulation results were subject to a multiple resolution sensitivity analysis. Under the assumption that the spatial resolution of the models' inputs can have important effects on the output, this parameter is central to the ability of a model to project future scenarios of LULC change. Thus, it was hypothesized that when the spatial resolution of inputs changes while all other parameters of the model are held constant the quantities, the spatial allocation and the spatial patterns of outputs can differ. The results after this step identify areas of future LULC change disregarding the spatial resolution of inputs and are unaffected by the bias they entail. iv) Finally, this dissertation included a modeling case study at the local scale (Messoghia plain) and one at the regional scale (Attica region). After detecting and quantifying the historical LULC changes, it was noted that the results were uneven. For instance, Messoghia (which are located inside the wider Attica region) experienced four distinct periods of development, while for Attica region, the periods of development were three. This fact underscores the aforementioned that different scales of analysis reveal different patterns and that the results obtained from a scale specific case study are not representative for the wider context.

Regarding the spatial determinants to the different types of LULC changes, the dissertation incorporated a total of 27 variables into the modeling process. By implementing 18 different models representing every possible LULC transition, the contribution of each factor was quantified using two meaningful metrics. After applying all these models, four clear messages emerge: First, the results demonstrate that depending on the LULC type, different factors are dominant in spatial determination of changes. Especially the interrelations of urban related categories, can be clearly distinguished according to density which translates in different residential use (eg second homes). In densely built urban areas, spatial determinants such as road density, enterprises density, amenities (health, education) and accessibility to the municipality center were the most dominant. In urban areas with less density, distance to shoreline, to blue flagged beaches were among the most important. The results converge with the literature about those factors, especially with studies that looked at the coastal zone of Mediterranean countries. Second, many of these factors that deemed as "drivers of change", can be actually "driven by" the changes. In other words, it is hard to interpret

whether these factors are causes or consequences of urban growth. As an example, the road or the enterprises densities that rank among the top correlations, might actually just follow the patterns of settlements. Third, 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 and this is mostly due to the geomorphology of Attica region, where a large amount of settlements is concentrated into plains. Finally, it is important to be noted that all these patterns and numbers are case specific and the conclusions drawn from the quantitative insights are not necessarily fully transferable to other regions. Even if Attica region can be aptly categorized into groups like coastal areas, Mediterranean cities or regions containing a big metropolitan area, and apparently share some common attributes or patterns. This is mostly due to specificities apparent only in the region, for example the physical constraints related to geomorphology, might be less pronounced in other areas. Another aspect would be the peoples' choices for residency, or the presence or absence of a regulation mechanism which, more or less, shapes the urban patterns in other countries.

Regarding how the socio-economic circumstances relate to the changes in LULC, this dissertation included two approaches that clearly advocate their importance. The economy dictates the development trends and in turn forms the demand and supply chain. In particular the built-up expansion rates are highly correlated with the level of economic development as demonstrated by the LULC change detection performed in two cases spanning a three decades period. The spatio-temporal dynamics that Attica region experienced, revealed uneven development trends that fully reflect the conditions of each epoch. Higher development rates were evident in conjunction with significant fund allocation, competitiveness and economic soundness and at the same time the built-up land expanded remarkably. During economic depression times, transactions in the real estate market followed the opposite way. It should be stressed though that all the historical LULC changes were driven by the development as well as by the absence of a regulation mechanism. Observing these uneven historical trends, served as a basis to sketch distinct and alternative scenarios and project the observed trends to future decades exploring plausible alternative pathways.

Several other conclusions can be drawn from this dissertation and should be noted:

Earth observation coupled with Geoinformatics is a sound way to provide a wide range of spatio-temporal information accurately and cost-effectively. Landsat imagery are particularly suitable for applications related to detecting historical LULC changes, since the satellite was launched in the early 1970s and constitutes the longest record of the Earth's surface. The long archives are readily available for download with no costs, and this makes it the only feasible option for studies that span some decades of time. It is obvious that for studies that require large extents. The only compromise a researcher has to make, is the spatial resolution since Landsat data come with a nominal pixel size of 30m. Satellite sensors record the emitted energy of objects and each satellite image is therefore a file of spectral signatures, translated by users as information about the objects and each pixel represents the spectral characteristics of all objects found in a 900 m^2 area. Apparently, this translates to much loss of information and might be crucial to the results. However, the ratio price/spatial resolution/size is almost inversely proportional and for this reason, it is necessary to be taken into account that the level of detail, the available budget and the purpose of study are complementary. For example, for detecting historical LULC changes over Attica region, instead of using more than 10 Landsat images at 30 m spatial resolution, would engage considerably more images if very high spatial resolution was deemed important. Apart from the high costs, the sizes of files and thus the time and computational costs for all processes would proliferate. Thus, given that the analysis was operated at the regional level and not on the block, such amount of information was not imperative.

The semi-automated techniques demonstrated, are proven a sound way to overcome the need of exhaustive methodologies in order to train classification algorithms, which, in fact, prevent many researchers from producing LULC data in high temporal resolution. Given that changes usually occur in a small fraction of land and especially at the edges, extracting information as training from already available datasets, utilizing unchanged areas as a training source, can be reasonable. For doing so, the potential error that will likely propagate undermining the whole process would be due to incompatibility and unsuitability between datasets and scales. To overcome this obstacle, relocation and elimination of points close to the boundaries between adjacent categories is the only option. The semi-automated techniques demonstrated, are fully transferable and can act as a baseline for continuous monitoring of LULC. Also, additional conclusions, regarding the classification process, emerged throughout

this dissertation. The amount of training samples, was crucial for accurate classifications and also the whole process benefits from samples that are proportional to the occurrence of each LULC category. Generally, it was found that classification accuracy will increase with larger training data but sample distribution with a good range of intra-class variability to be represented, is of equal importance.

The RF algorithm for classification and later for regression analysis was proven to be robust and advantages of using it are stressed in all chapters. Especially the variable importance functions were efficient as they report on not only the influence of each predictor separately, but also their multivariate interactions. The algorithm successfully handled the fusion of multiple and heterogenous data allowing the accomplishment of very high thematic resolution disaggregating the urban-related LULC categories. The discrimination of LULC categories according to their density and continuity was an important step, because the detection and quantification of such changes and their projection provided unique insights into the whole process of LULC system. In case this step was avoided, and a more conventional and straightforward nomenclature of categories, was adopted, the majority of changes would have been ignored (e.g the changes in density observed in the northern suburbs of Athens where the extent remained relatively constant while the density increased dramatically).

The good performance of the models designed for Messoghia and for Attica suggests that the predictors incorporated were capable to spatially determine historical LULC changes for each case. But an important challenge was the accurate identification of scattered unplanned development patches, the so-called leap-frog development. In both models, the incorporation of the Leap-frog index boosted the performance of the RF algorithm facilitating an adequate fit in the face of sprawl detection. This task was previously reported by several researchers as difficult to detect and predict since this type of sprawl forms patches that vary in shape, structure, composition and place of occurrence.

The attempt to couple CA and RF was a sound way to overcome certain limitations in an approach that combines the advantages of each method. On the one hand, the RF algorithm provided a robust modeling option to generate accurate transition potential surfaces, by fusing heterogenous data without overfitting and collinearity issues, while on the other hand CA modeling proved fully compatible with

transition probability surfaces produced by the RF models. Additionally, the two separate sub-models Patcher and Expander allowed the efficient calibration of the models according to case specific needs by taking into account actual parameters of the study area. Thus, coupling these two frameworks is fully operational and reduces several limitations that are commonly reported in the literature. In fact, recent comparative approaches stress that the combination of RF and CA, outperforms other methods.

As for Attica region and the wide transformations evident throughout the previous decades, it should be noted that scenarios reveal plausible outcomes that reflect the envisioned and encouraged economic re-growth. The economic goals set by the policy makers, appear capable of taking precedence over virtually all other spatial planning priorities undermining the enforcement of actual regulation, once more. Therefore, the scenarios presented in this dissertation can not be dismissed as implausible, not even the most optimistic in terms of economic growth. Results generated from the models stress out the major impact that would arise from unplanned artificial areas expansion. From this spectrum, the presence of a spatial planning scheme that is visually and quantitatively informed about the potential consequences is a key step towards finding the optimal balance between development and sustainability. A regulatory mechanism should be operational and unobtrusive, reducing the negative consequences of development without hindering growth, by shifting the distribution of new development to locations that are more ecologically suitable.

Projecting LULC patterns is a useful experiment for evaluating the causes and identifying the impact of these changes. The scenario-based simulations are a useful way to sketch out how the LULC patterns evolve under different pathways with a level of plausibility. Embedded in every analysis engaging scenarios, there will always be a level of potential uncertainties originated from the general nature relating to the socioeconomic predictions that drive the scenarios, the inability to foresee any unexpected circumstances and integrate any emerging discontinuities or the data used for the models. Especially when dealing with complex systems such as LULC changes, those assumptions are unavoidable. Combining an empirical analysis and sketching different scenarios attributes and conditions that deviate from historic trends in LULC changes, is a way to minimize the uncertainty. Another important aspect is the careful observation of these trends, identifying any uneven patterns throughout the study

period, that should be taken into account in the design process. Finally, a central point of the LULC system science that should be noted is that the LULC system as a whole requires scientific advances by bringing together diverse disciplines to co-design integrated approaches and jointly work towards such a multi-disciplinary scientific problem. This is why this dissertation avoided to explicitly deal with subjects pertaining to different disciplines, like policy making, spatial planning, environmental management or socio-economics and human geography and is rather centered to the perspective of how geo-informatics can advance the methodological framework in various ways and how the wide array of methodologies it entails, can be assembled in an integrated multidisciplinary approach.