



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ & ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
Δ.Π.Μ.Σ. ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ
Τομέας Τοπογραφίας
Εργαστήριο Τηλεπισκόπησης

ΑΝΑΠΤΥΞΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΕΘΟΔΟΛΟΓΙΑΣ
ΓΙΑ ΤΗΝ ΑΥΤΟΜΑΤΗ ΠΑΡΑΓΩΓΗ ΧΑΡΤΩΝ
ΚΑΛΥΨΗΣ ΓΗΣ ΓΙΑ ΤΗΝ ΕΛΛΑΔΑ

Μεταπτυχιακή Εργασία

Δημήτριος Μπληζιώτης

Αθήνα, Ιούνιος 2016



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF RURAL AND SURVEYING ENGINEERING
GEOINFORMATICS, post-graduate programme
Department of Topography
Remote Sensing Laboratory

**DEVELOPMENT AND EVALUATION OF AN
AUTOMATED LAND COVER MAPPING APPROACH
IN GREECE**

Master's Thesis

Dimitrios Bliziotis

Athens, June 2016



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ & ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
Δ.Π.Μ.Σ. ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ
Τομέας Τοπογραφίας
Εργαστήριο Τηλεπισκόπησης

**Ανάπτυξη και αξιολόγηση μεθοδολογίας για την
αυτόματη παραγωγή χαρτών κάλυψης γης για
την Ελλάδα**

Μεταπτυχιακή Εργασία

Δημήτριος Μπληζιώτης

Τριμελής εξεταστική επιτροπή:

Κ. Καράντζαλος

Δ. Αργιαλάς

Β. Καραθανάση

.....

.....

.....

Επ. Καθηγητής Ε.Μ.Π.
Επιβλέπων

Καθηγητής Ε.Μ.Π.

Αν. Καθηγήτρια Ε.Μ.Π.

Αθήνα, Ιούνιος 2016

Δημήτριος Μπληζιώτης

Διπλωματούχος Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π.

Copyright © All rights reserved. Δημήτριος Μπληζιώτης, 2016

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Αφιερώνεται

*Στους γονείς μου
Λεωνίδα και Ελευθερία*

*Στον αδερφό μου
Γιώργο*

Στην Κλεοπάτρα

ΠΡΟΛΟΓΟΣ

Η παρούσα Μεταπτυχιακή Διπλωματική Εργασία εκπονήθηκε στα πλαίσια ολοκλήρωσης των μεταπτυχιακών μου σπουδών στο Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών (ΔΠΜΣ) στο γνωστικό πεδίο της Γεωπληροφορικής στη Σχολή Αγρονόμων και Τοπογράφων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Το θέμα της διπλωματικής μου ανατέθηκε από το εργαστήριο Τηλεπισκόπησης του Τομέα Τοπογραφίας της σχολής.

Το ΔΠΜΣ έχει σαν στόχο την διεπιστημονική εμβάθυνση των διπλωματούχων Μηχανικών του ΕΜΠ ή και άλλων ΑΕΙ της χώρας, ή ισοτίμων της αλλοδαπής και άλλων επιστημόνων, αφενός στη συλλογή, γεωαναφορά, περιγραφή, ερμηνεία και απεικόνιση χωρικών δεδομένων του φυσικού, τεχνικού και κοινωνικοοικονομικού περιβάλλοντος και αφετέρου στην ανάλυση και στο σχεδιασμό του χώρου, με τη χρήση σύγχρονων μεθόδων και τεχνολογιών αιχμής και ιδιαίτερα της πληροφορικής.

Το παραπάνω μεταπτυχιακό πρόγραμμα σπουδών καλύπτει ένα μεγάλο εύρος επιστημονικών πεδίων όπως είναι για παράδειγμα η τηλεπισκόπηση, η φωτογραμμετρία, η γεωγραφία, η γεωδαισία, το κτηματολόγιο, τα γεωπληροφορικά συστήματα και άλλα. Τα τελευταία χρόνια με την είσοδο των σύγχρονων τεχνολογιών, τα αντικείμενα ενασχόλησης του Τοπογράφου Μηχανικού έχουν διευρυνθεί και έχουν αποκτήσει νέες διαστάσεις. Πιο συγκεκριμένα, με την ενσωμάτωση της δορυφορικής τεχνολογίας καθώς και των Γεωγραφικών Συστημάτων Πληροφοριών (GIS) έχουν αναπτυχθεί ειδικές μέθοδοι για την επεξεργασία ψηφιακών εικόνων και διανυσματικών στοιχείων παρέχοντας ιδιαίτερα χρήσιμες πληροφορίες, διευκολύνοντας και βελτιώνοντας τον εντοπισμό, τη χαρτογράφηση και την πλοήγηση σε πολλές εφαρμογές με πολύ μεγάλη χωρική ακρίβεια.

Στο πλαίσιο ενσωμάτωσης των νέων τεχνολογιών στο επιστημονικό φάσμα του Αγρονόμου και Τοπογράφου Μηχανικού ανέλαβα τη συγκεκριμένη Μεταπτυχιακή Διπλωματική Εργασία με σκοπό την αξιοποίηση μεθόδων παρατήρησης γης και των GIS για την ανάπτυξη αλγορίθμων τηλεπισκόπησης αξιοποιώντας μεγάλα και διαχρονικά γεωχωρικά δεδομένα. Το κίνητρο μου για να αναλάβω το συγκεκριμένο θέμα ήταν το ενδιαφέρον μου αφενός για τον τομέα της τηλεπισκόπησης και αφετέρου για την επιστήμη της πληροφορικής, των ανοιχτών λογισμικών και του προγραμματισμού.

Στο σημείο αυτό, θα ήθελα να ευχαριστήσω όσους βοήθησαν στην υλοποίηση της παρούσας εργασίας. Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Λέκτορα κύριο Κωνσταντίνο Καράντζαλο, για την απεριόριστη βοήθεια και συμπαράστασή του σε όλη τη διάρκεια εκπόνησης της εργασίας. Επιπλέον, θα ήθελα να ευχαριστήσω τους γονείς και τους φίλους μου για την αμέριστη συμπαράσταση και βοήθειά τους σε όλη τη διάρκεια του μεταπτυχιακού προγράμματος.

Περίληψη

Στη συγκεκριμένη μεταπτυχιακή διπλωματική εργασία αναπτύχθηκε και αξιολογήθηκε μεθοδολογική προσέγγιση για την αυτόματη παραγωγή χαρτών κάλυψης και χρήσεων γης, αξιοποιώντας ελεύθερα δορυφορικά τηλεπισκοπικά δεδομένα καθώς και εξειδικευμένους αλγόριθμους ταξινόμησης που συνδέονται με την επιστήμη της τηλεπισκόπησης και παρατήρησης της γης, για την περιοχή της Ελλάδας. Για την υλοποίηση των αλγορίθμων και την περάτωση των απαιτούμενων διαδικασιών με σκοπό την επίτευξη του παραπάνω στόχου χρησιμοποιήθηκαν λογισμικά και δεδομένα που διέπονται από τις αρχές του ανοιχτού λογισμικού και των ανοιχτών προτύπων αντίστοιχα.

Πιο συγκεκριμένα, η μεθοδολογική προσέγγιση που αναπτύχθηκε ακολουθεί μορφή ιεραρχικού δέντρου απόφασης, η οποία αξιοποιεί πολλαπλούς ταξινομητές ανάλογα με τη θεματική κατηγορία προς ταξινόμηση. Η συγκεκριμένη μεθοδολογία βασίζεται στη λογική του “per class classification”, δηλαδή κάθε κατηγορία ταξινομείται χωριστά και στη συνέχεια «μασκάρεται» έτσι ώστε ο αλγόριθμος ταξινόμησης να εστιάζει κάθε φορά στην ενδιαφερόμενη κατηγορία, αξιοποιώντας παράλληλα φαινολογικά χαρακτηριστικά, τα οποία εξάγονται ύστερα από κατάλληλη ανάλυση χρονοσειρών για την αύξηση της ακρίβειας στις θεματικές κατηγορίες της βλάστησης και των καλλιεργειών. Για το σύνολο των διαδικασιών δημιουργήθηκε πρόγραμμα γραμμένο σε γλώσσα προγραμματισμού Python αξιοποιώντας ελεύθερες βιβλιοθήκες όπως είναι η GDAL, NumPy και Orfeo Toolbox.

Η μεθοδολογική προσέγγιση που αναπτύχθηκε ταξινομεί την κάθε σκηνή σε 2 στάδια, ανάλογα με το επίπεδο λεπτομέρειας κάθε κατηγορίας. Πιο συγκεκριμένα, κατά το 1ο στάδιο εφαρμόζεται μία ιεραρχική ταξινόμηση ανά κλάση, λαμβάνοντας υπόψη λογικούς και ποσοτικούς κανόνες, ενώ στο 2ο στάδιο ενσωματώνονται διαχρονικά φαινολογικά χαρακτηριστικά σε συνδυασμό με το διαχρονικό στατιστικό μοντέλο εκπαίδευσης/πρόβλεψης που δημιουργήθηκε από δείγματα σε διάφορες περιοχές και χρονικές στιγμές στην Ελλάδα. Η ενσωμάτωση των χαρακτηριστικών που σχετίζονται με το χρόνο ήταν ιδιαίτερα σημαντική για τη βελτίωση της ακρίβειας των αποτελεσμάτων της ταξινόμησης αξιοποιώντας δορυφορικά δεδομένα από τον Landsat 8 καθώς και αναλύοντας χρονοσειρές NDVI του αντίστοιχου δορυφόρου.

Συνοψίζοντας, η μεθοδολογική προσέγγιση που ακολουθήθηκε, καθώς και τα αντίστοιχα προγράμματα που δημιουργήθηκαν, αντιμετώπισαν ικανοποιητικά την παραγωγή διαχρονικών χαρτών κάλυψης γης και χρήσεων γης για την περιοχή της Ελλάδας με ακρίβειες καλύτερες του 85 % στις περιοχές μελέτης που επιλέχθηκαν. Λαμβάνοντας τα παραπάνω υπόψη, μπορεί να θεωρηθεί πως η συγκεκριμένη μεθοδολογική προσέγγιση, ύστερα από κατάλληλες παραμετροποιήσεις και προσθήκες, μπορεί να αξιοποιηθεί ως μία εφικτή επιχειρησιακή λύση για την παραγωγή χαρτών κάλυψης γης σε χωρική ανάλυση 30 m για την περιοχή της Ελλάδας.

Λέξεις κλειδιά: Χρήσεις/Κάλυψη Γης, Μέθοδοι Παρατήρησης Γης, Μηχανική μάθηση, Εξόρυξη Δεδομένων, Ταξινομήσεις

Abstract

In this MSc thesis a methodological approach for the automatic production of land cover and land use maps for the region of Greece was developed and evaluated. The methodology takes advantage of free satellite remote sensing data (such as Landsat 8 imagery) and specialized classification algorithms, utilizing machine learning processes, associated with the science of remote sensing and earth observation. For the implementation of the algorithms and the completion of the necessary procedures open source software/programming languages as well as data governed by the principles of open standards were extensively used.

More specifically, the developed methodological approach for the production of automated LC/LU maps follows the form of a hierarchical decision tree, which utilizes multiple classifiers depending on the land use to be classified. The methodology is based on the logic of "per class classification", i.e. each category is classified separately and then is masked so that the classification algorithm focuses each time on the concerned category. Moreover, in order to increase the accuracy of the thematic categories of vegetation and crops the methodology incorporates multitemporal phenological characteristics, exported from the appropriate time series analysis. For all procedures a software written in Python programming language, utilizing free libraries such as GDAL, Numpy and Orfeo Toolbox, was developed.

The developed methodological approach classifies each scene in two stages, depending on the level of detail in each category. More specifically, during the first stage a hierarchical per class classification approach is applied, taking into account logical and quantitative rules, while during the second stage the multitemporal phenological features are incorporated combined with the multitemporal training/predictive statistical model created from samples at various locations and times in Greece. The integration of the characteristics associated with the time was particularly important in order to improve the accuracy of the classification results utilizing satellite data from Landsat 8 and analyzing time series of NDVI of the corresponding satellite.

In summary, the methodological approach followed, and the respective programs created, sufficiently experienced the goal of producing temporal land cover and land use maps in the region of Greece with accuracies better than 85% in the target areas selected. Taking the abovementioned into account, it can be assumed that the specific methodological approach, after appropriate configurations and additions, can be used as a viable operational solution for the production of land cover maps at 30 m spatial resolution in the region of Greece.

Keywords: Land Cover/Land Use, Earth Observation Methods, Machine Learning, Data Mining, Classification Algorithms

ΕΙΣΑΓΩΓΗ

Η δημιουργία χαρτών κάλυψης γης είναι μία πολύ σημαντική μεταβλητή για διάφορες μελέτες και εφαρμογές που σχετίζονται με την παρακολούθηση της γήινης επιφάνειας, όπως είναι για παράδειγμα η κλιματολογία, η υδρολογία, η ποιότητα της ατμόσφαιρας, η παρακολούθηση της βλάστησης και των συνεπειών των φυσικών καταστροφών κ.τ.λ. Λαμβάνοντας υπόψη πως η κατάσταση της γης είναι δυναμική, οι χρήσεις γης μεταβάλλονται στο χρόνο από διάφορους παράγοντες, οι οποίοι οφείλονται τόσο σε ανθρωπογενείς παρεμβάσεις όσο και σε φυσικά φαινόμενα. Για παράδειγμα, τα πλημμυρικά φαινόμενα διαρκούν συνήθως μεγάλα χρονικά διαστήματα επηρεάζοντας σε σημαντικό βαθμό τροπικές περιοχές καθώς και περιοχές σε χαμηλά υψόμετρα. Αντίστοιχα, μέσα σε διάστημα μηνών το γυμνό έδαφος μπορεί να έχει καλυφθεί από βλάστηση ή από καλλιέργειες, αλλάζοντας τη χρήση γης της περιοχής. Λαμβάνοντας υπόψη τα παραπάνω, η μελέτη και η παρακολούθηση των δυναμικών χρήσεων γης δημιουργούν προκλήσεις στη συνεχή και ακριβή χαρτογράφησή τους, ιδιαίτερα τη σύγχρονη εποχή όπου οι χρήσεις γης μεταβάλλονται σε σύντομα χρονικά διαστήματα λόγω της περιβαλλοντικής ρύπανσης, της υποβάθμισης του εδάφους κ.τ.λ. σε παγκόσμιο επίπεδο.

Με την ανάπτυξη της τεχνολογίας και τη συνεχή μείωση του κόστους απόκτησης δορυφορικών τηλεπισκοπικών δεδομένων, διάφορα παγκόσμια και τοπικά προϊόντα κάλυψης γης έχουν αναπτυχθεί τόσο σε πειραματικό όσο και σε επιχειρησιακό στάδιο. Ωστόσο, η πληθώρα των υπαρχόντων προϊόντων χαρακτηρίζονται από διαφορετικές προδιαγραφές και ακρίβειες στα αποτελέσματα και ως εκ τούτου αποτελεί σημαντικό παράγοντα η επιλογή του κατάλληλου συστήματος ανάλογα με τις ανάγκες των διάφορων χρηστών. Για αυτό το λόγο έχουν πραγματοποιηθεί διάφορες αναλύσεις σύγκρισης των υφιστάμενων προϊόντων ως προς την αβεβαιότητα και ακρίβειά τους.

Στο πλαίσιο αυτό, στόχος της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι η ανάπτυξη και αξιολόγηση μεθοδολογικής προσέγγισης για την αυτόματη παραγωγή χαρτών κάλυψης γης και χρήσεων γης, αξιοποιώντας ελεύθερα δορυφορικά τηλεπισκοπικά δεδομένα καθώς και εξειδικευμένους αλγόριθμους ταξινόμησης που συνδέονται με την επιστήμη της τηλεπισκόπησης και παρατήρησης της γης, για την περιοχή της Ελλάδας. Για την υλοποίηση των αλγορίθμων και την περάτωση των απαιτούμενων διαδικασιών με σκοπό την επίτευξη του παραπάνω στόχου χρησιμοποιήθηκαν λογισμικά και δεδομένα που διέπονται από τις αρχές του ανοιχτού λογισμικού και των ανοιχτών προτύπων αντίστοιχα.

Ακολουθεί η διάρθρωση της εργασίας:

- **Κεφάλαιο 1:** Στο συγκεκριμένο κεφάλαιο πραγματοποιείται εισαγωγή στα υφιστάμενα παγκόσμια συστήματα παραγωγής χαρτών κάλυψης και χρήσεων γης, τα οποία αξιοποιούν δορυφορικά δεδομένα και αλγορίθμους παρατήρησης γης σε συνδυασμό με γεωγραφικά υπολογιστικά συστήματα. Επίσης, παρουσιάζονται και αξιολογούνται προϊόντα που αξιοποιούν τόσο δεδομένα χαμηλής χωρικής ανάλυσης όσο και αντίστοιχα που χρησιμοποιούν δορυφορικά τηλεπισκοπικά δεδομένα πολύ υψηλής χωρικής ανάλυσης. Τέλος, πραγματοποιείται αναφορά και στα αντίστοιχα ευρωπαϊκά συστήματα καθώς και στα εθνικά που έχουν αναπτυχθεί σε κάθε ήπειρο.

- **Κεφάλαιο 2:** Στο δεύτερο κεφάλαιο παρουσιάζονται οι κυριότεροι αλγόριθμοι αιχμής στον τομέα της ταξινόμησης που χρησιμοποιούν την τεχνολογία της μηχανικής μάθησης (machine learning) και χρησιμοποιούνται ευρέως σε εφαρμογές ταξινόμησης σε μεγάλα και διαχρονικά γεωχωρικά δεδομένα. Οι συγκεκριμένοι αλγόριθμοι εντάσσονται στην ευρύτερη κατηγορία των διαδικασιών εξόρυξης δεδομένων (data mining).
- **Κεφάλαιο 3:** Το τρίτο κεφάλαιο αποτελεί το βασικό κορμό της συγκεκριμένης μεταπτυχιακής διπλωματικής εργασίας, όπου παρουσιάζεται, αναλύεται και αξιολογείται η μεθοδολογική προσέγγιση που αναπτύχθηκε με σκοπό την επίτευξη του στόχου της εργασίας. Πιο συγκεκριμένα, παρουσιάζεται η ανάπτυξη μεθοδολογικής προσέγγισης για την αυτόματη παραγωγή διαχρονικών χαρτών κάλυψης και χρήσεων γης αξιοποιώντας πολυφασματικά δορυφορικά δεδομένα. Για το σκοπό αυτό αξιοποιήθηκαν διαχρονικά δεδομένα του δορυφόρου Landsat 8 από διάφορες περιοχές της Ελλάδας. Λαμβάνοντας υπόψη τη διεθνή βιβλιογραφία καθώς και ύστερα από δοκιμές διάφορων μεθοδολογικών προσεγγίσεων και αλγορίθμων αναπτύχθηκε μεθοδολογία μορφής δέντρου απόφασης, η οποία αξιοποιεί πολλαπλούς ταξινομητές στα διάφορα στάδια της μεθοδολογίας.
- **Κεφάλαιο 4:** Στο συγκεκριμένο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προέκυψαν από την εφαρμογή της μεθοδολογίας που αναπτύχθηκε, επισημαίνονται τα σημαντικότερα προβλήματα που παρουσιάστηκαν στα διάφορα στάδια της διαδικασίας ανάλυσης, ενώ τέλος δίνονται κατευθύνσεις για περαιτέρω έρευνα σε θέματα που αφορούν τη βελτίωση της μεθοδολογικής προσέγγισης που ακολουθήθηκε.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΡΟΛΟΓΟΣ.....	i
Περίληψη.....	ii
Abstract.....	iii
ΕΙΣΑΓΩΓΗ.....	iv
1. ΠΑΓΚΟΣΜΙΑ ΣΥΣΤΗΜΑΤΑ ΠΑΡΑΓΩΓΗΣ ΧΑΡΤΩΝ ΚΑΛΥΨΗΣ ΓΗΣ.....	1
1.1 Παραδείγματα Παγκόσμιων Συστημάτων Κάλυψης Γης	1
2. ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΑΙΧΜΗΣ ΓΙΑ ΤΗΝ ΤΑΞΙΝΟΜΗΣΗ ΔΟΥΡΥΦΟΡΙΚΩΝ ΕΙΚΟΝΩΝ.....	8
2.1 Data Mining.....	8
2.2 Μέθοδοι Ταξινόμησης και Μηχανική Μάθηση	10
2.2.1 SVM.....	10
2.2.2 Decision tree	15
2.2.3 K-nearest neighbors.....	18
2.2.4 Αξιολόγηση στατιστικών μοντέλων	20
3. ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΔΙΑΧΡΟΝΙΚΩΝ ΧΑΡΤΩΝ ΚΑΛΥΨΗΣ ΓΗΣ	23
3.1 Περιοχές Μελέτης και Επιλογή Δεδομένων	23
3.2 Μεθοδολογική Προσέγγιση.....	31
3.2.1 Ενσωμάτωση φαινολογικών χαρακτηριστικών	33
3.2.2 Κατηγορίες χρήσεων γης	36
3.3 1 ^ο Στάδιο Ταξινόμησης: Περιγραφή Κανόνων για τις Κατηγορίες Χρήσεων Γης... ..	39
3.3.1 Σύννεφα/Σκιές	43
3.3.2 Χιόνι	46
3.3.3 Υδάτινοι όγκοι.....	48
3.3.4 Βλάστηση	49
3.3.5 Τεχνητές επιφάνειες.....	50
3.3.6 Αγροτικές εκτάσεις	51
3.3.7 Θαμνολίβαδα, χορτολιβαδικές εκτάσεις, διάφοροι τύποι βλάστησης	53
3.4 2 ^ο Στάδιο Ταξινόμησης: Δημιουργία και Αξιολόγηση Διαχρονικού Στατιστικού Μοντέλου Εκπαίδευσης	54
3.5 Αποτελέσματα / Αξιολόγηση	57
3.5.1 Περιοχή 1: Αττική.....	58
3.5.2 Περιοχή 2: Αιτωλοακαρνανία.....	61
3.5.3 Περιοχή 3: Βορειοδυτική Πελοπόννησος.....	64
3.5.4 Περιοχή 4: Θεσσαλονίκη	67
3.5.5 Περιοχή 5: Χανιά.....	70
3.5.6 Σύγκριση / Αξιολόγηση Μεθοδολογίας με Άλλους Αλγορίθμους Ταξινόμησης	73
4. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ.....	75
4.1 Συμπεράσματα.....	75

4.2 Προοπτικές Εξέλιξης	77
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	79

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1-1: Παγκόσμια συστήματα κάλυψης και χρήσεων γης (GLC) [19]	2
Εικόνα 1-2: Παραδείγματα των προϊόντων: MCD12Q1 – Western United States (πάνω αριστερά), UMD GLC – New Zealand (πάνω δεξιά), GLC 2000 – global (κάτω αριστερά), GlobCover – global (κάτω δεξιά)	3
Εικόνα 1-3: Παραδείγματα των προϊόντων: GLCNMO – global (πάνω αριστερά), Geo-Wiki (πάνω δεξιά), CCI-LC - global (κάτω αριστερά), GLCC 2.0 - global (κάτω δεξιά).....	4
Εικόνα 1-4: Παραδείγματα των προϊόντων: FROM-GLC-agr – global (αριστερά), GlobeLand30 - global (δεξιά)	5
Εικόνα 1-5: Παραδείγματα των προϊόντων: CORINE 2000 (πάνω αριστερά), CORINE 2006 (πάνω δεξιά), LUCAS 2009 TOPSOIL (κάτω αριστερά), GlobCorine (κάτω δεξιά).....	6
Εικόνα 1-6: Εθνικά συστήματα, σε παγκόσμια κλίμακα, κάλυψης και χρήσεων γης [19]	7
Εικόνα 2-1: Παράδειγμα εύρεσης υπερεπιπέδου μεταξύ δύο κλάσεων μέσω του αλγορίθμου SVM.....	11
Εικόνα 2-2: Πρόβλημα βελτιστοποίησης περιθωρίου σε γραμμικό ταξινομητή SVM	12
Εικόνα 2-3: Δέντρο απόφασης για την εύρεση συγκεκριμένων επιθέτων [42]	16
Εικόνα 2-4: Δέντρο απόφασης για την ταξινόμηση εικονοστοιχείων [43]	16
Εικόνα 2-5: Ψευδοκώδικας αλγορίθμου τερματισμού επέκτασης δέντρου απόφασης [42]	17
Εικόνα 2-6: Παράδειγμα εφαρμογής αλγορίθμου KNN.....	19
Εικόνα 2-7: Ψευδοκώδικας αλγορίθμου KNN [47]	19
Εικόνα 2-8: Δεδομένα εκπαίδευσης χωρίς την εφαρμογή του CNN	20
Εικόνα 2-9: 1NN ταξινόμηση χωρίς εφαρμογή του CNN	20
Εικόνα 2-10: Δεδομένα εκπαίδευσης με εφαρμογή του CNN	20
Εικόνα 2-11: 1NN ταξινόμηση με εφαρμογή του CNN.....	20
Εικόνα 3-1: Παραδείγματα από “artifacts” σε περιοχές με έντονη τοπογραφική ανομοιομορφία από τα ατμοσφαιρικά διορθωμένα προϊόντα του Landsat 8	25
Εικόνα 3-2: Συνοπτική παρουσίαση της μεθοδολογικής προσέγγισης που αναπτύχθηκε ...	33
Εικόνα 3-3: Αρχική χρονοσειρά NDVI (λεπτή γραμμή), τελική χρονοσειρά NDVI (παχιά γραμμή) με τη χρήση του φίλτρου Savitzky–Golay	35
Εικόνα 3-4: Ομαλοποιημένη συνάρτηση χρονοσειρών NDVI	36
Εικόνα 3-5: Παράμετροι εποχικότητας (a) αρχή περιόδου, (b) τέλος περιόδου, (c) εύρος, (d) τιμή βάσης, (e) χρόνος μέσου περιόδου,	36
Εικόνα 3-6: Αναλυτικό διάγραμμα ροής του 1 ^{ου} σταδίου της μεθοδολογικής προσέγγισης	41
Εικόνα 3-7: Περιγραφή κανόνων ιεραρχικού δέντρου ταξινόμησης για το 1 ^ο στάδιο ταξινόμησης.....	42

Εικόνα 3-8: Επεξήγηση τιμών για το κανάλι CFmask (http://landsat.usgs.gov/documents/provisional_l8sr_product_guide.pdf)	43
Εικόνα 3-9: Επεξήγηση τιμών για το κανάλι CFmask Cloud Confidence (http://landsat.usgs.gov/documents/provisional_l8sr_product_guide.pdf)	44
Εικόνα 3-10: Κανάλι για ανίχνευση σύννεφων	44
Εικόνα 3-11: Κανάλι CFmask.....	44
Εικόνα 3-12: Κανάλι CFmask Confidence	44
Εικόνα 3-13: Παραδείγματα από την ταξινόμηση των σύννεφων σε διάφορες περιοχές και χρονικές στιγμές	46
Εικόνα 3-14: Παραδείγματα από την ταξινόμηση του χιονιού σε διάφορες περιοχές και χρονικές στιγμές	47
Εικόνα 3-15: Παραδείγματα από την ταξινόμηση των υδάτινων όγκων σε διάφορες περιοχές και χρονικές στιγμές	48
Εικόνα 3-16: Παραδείγματα από την ταξινόμηση της βλάστησης σε διάφορες περιοχές και χρονικές στιγμές	50
Εικόνα 3-17: Παραδείγματα από την ταξινόμηση των τεχνητών επιφανειών σε διάφορες περιοχές και χρονικές στιγμές	51
Εικόνα 3-18: Διάγραμμα ροής για την ανίχνευση των καλλιεργειών.....	52
Εικόνα 3-19: Παραδείγματα από την ταξινόμηση των καλλιεργειών σε διάφορες περιοχές και χρονικές στιγμές	53
Εικόνα 3-19: Αρχική και ταξινομημένη εικόνα για την περιοχή της Αττικής (Πάνω 13/4/2015, Κάτω 4/9/2015)	58
Εικόνα 3-20: Αρχική και ταξινομημένη εικόνα για την περιοχή της Αιτωλοακαρνανίας (Πάνω 1/8/2015, Κάτω 5/11/2015)	61
Εικόνα 3-21: Αρχική και ταξινομημένη εικόνα για την περιοχή της Βορειοδυτικής Πελοποννήσου (Πάνω 14/6/2015, Κάτω 18/9/2015)	64
Εικόνα 3-22: Αρχική και ταξινομημένη εικόνα για την περιοχή της Θεσσαλονίκης (Πάνω 1/8/2015, Κάτω 18/10/2015)	67
Εικόνα 3-23: Αρχική και ταξινομημένη εικόνα για την περιοχή των Χανίων (Πάνω 4/9/2015, Κάτω 29/4/2015)	70

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 2-1: Ελεύθερα λογισμικά και βιβλιοθήκες που χρησιμοποιούνται για διαδικασίες υπολογιστικής μάθησης	10
Πίνακας 2-2: Πλεονεκτήματα-Μειονεκτήματα SVM.....	15
Πίνακας 2-3: Πλεονεκτήματα-Μειονεκτήματα decision tree	18
Πίνακας 2-4: Πλεονεκτήματα-Μειονεκτήματα των κ-πλησιέστερων γειτόνων	20
Πίνακας 3-1: Χαρακτηριστικά των ατμοσφαιρικά διορθωμένων προϊόντων του Landsat 8 μέσω του αλγορίθμου L8SR.....	24

Πίνακας 3-2: Επιλογή διαχρονικών περιοχών μελέτης από διαφορετικές γεωγραφικές ενότητες στην Ελλάδα	31
Πίνακας 3-3: Χρήσεις γης που ανιχνεύει η μεθοδολογική προσέγγιση που αναπτύχθηκε ..	36
Πίνακας 3-4: Παραδείγματα χρήσεων γης που ανιχνεύονται	39
Πίνακας 3-5: Συνθήκες και κατώφλια για την ανίχνευση των σύννεφων και των σκιών.....	45
Πίνακας 3-6: Συνθήκες και κατώφλια για την ανίχνευση του χιονιού	46
Πίνακας 3-7: Συνθήκες και κατώφλια για την ανίχνευση των υδάτινων όγκων.....	48
Πίνακας 3-8: Συνθήκες και κατώφλια για την ανίχνευση της βλάστησης	49
Πίνακας 3-9: Συνθήκες και κατώφλια για την ανίχνευση των τεχνητών επιφανειών	50
Πίνακας 3-10: Στατιστικά για κάθε κλάση του στατιστικού μοντέλου εκπαίδευσης/πρόβλεψης.....	55
Πίνακας 3-11: Πίνακας σύγχυσης του στατιστικού μοντέλου εκπαίδευσης	56
Πίνακας 3-12: Πίνακας σύγχυσης για την περιοχή της Αττικής 13/4/2015	59
Πίνακας 3-13: Πίνακας σύγχυσης για την περιοχή της Αττικής 4/9/2015.....	60
Πίνακας 3-14: Πίνακας σύγχυσης για την περιοχή της Αιτωλοακαρνανίας 1/8/2015	62
Πίνακας 3-15: Πίνακας σύγχυσης για την περιοχή της Αιτωλοακαρνανίας 5/11/2015	63
Πίνακας 3-16: Πίνακας σύγχυσης για την περιοχή της βορειοδυτικής Πελοποννήσου 14/6/2015	65
Πίνακας 3-17: Πίνακας σύγχυσης για την περιοχή της βορειοδυτικής Πελοποννήσου 18/9/2015	66
Πίνακας 3-18: Πίνακας σύγχυσης για την περιοχή της Θεσσαλονίκης 1/8/2015	68
Πίνακας 3-19: Πίνακας σύγχυσης για την περιοχή της Θεσσαλονίκης 18/10/2015	69
Πίνακας 3-20: Πίνακας σύγχυσης για την περιοχή των Χανίων 4/9/2015.....	71
Πίνακας 3-21: Πίνακας σύγχυσης για την περιοχή των Χανίων 29/4/2015.....	72
Πίνακας 3-22: Σύγκριση / αξιολόγηση μεθοδολογίας με άλλους αλγορίθμους ταξινόμησης (SVM multi class, decision tree, k – nearest neighbors, random forest) για επιλεγμένες περιοχές και χρονικές στιγμές	74

1. ΠΑΓΚΟΣΜΙΑ ΣΥΣΤΗΜΑΤΑ ΠΑΡΑΓΩΓΗΣ ΧΑΡΤΩΝ ΚΑΛΥΨΗΣ ΓΗΣ

Η δημιουργία χαρτών κάλυψης γης είναι μία πολύ σημαντική μεταβλητή για διάφορες μελέτες και εφαρμογές που σχετίζονται με την παρακολούθηση της γήινης επιφάνειας, όπως είναι για παράδειγμα η κλιματολογία, η υδρολογία, η ποιότητα της ατμόσφαιρας, η παρακολούθηση της βλάστησης και των συνεπειών των φυσικών καταστροφών κ.τ.λ. Λαμβάνοντας υπόψη πως η κατάσταση της γης είναι δυναμική, οι χρήσεις γης μεταβάλλονται στο χρόνο από διάφορους παράγοντες, οι οποίοι οφείλονται τόσο σε ανθρωπογενείς παρεμβάσεις όσο και σε φυσικά φαινόμενα. Για παράδειγμα, τα πλημμυρικά φαινόμενα διαρκούν συνήθως μεγάλα χρονικά διαστήματα επηρεάζοντας σε σημαντικό βαθμό τροπικές περιοχές καθώς και περιοχές σε χαμηλά υψόμετρα. Αντίστοιχα, μέσα σε διάστημα μηνών το γυμνό έδαφος μπορεί να έχει καλυφθεί από βλάστηση ή από καλλιέργειες, αλλάζοντας τη χρήση γης της περιοχής. Λαμβάνοντας υπόψη τα παραπάνω, η μελέτη και η παρακολούθηση των δυναμικών χρήσεων γης δημιουργούν προκλήσεις στη συνεχή και ακριβή χαρτογράφησή τους, ιδιαίτερα τη σύγχρονη εποχή όπου οι χρήσεις γης μεταβάλλονται σε σύντομα χρονικά διαστήματα λόγω της περιβαλλοντικής ρύπανσης, της υποβάθμισης του εδάφους κ.τ.λ. σε παγκόσμιο επίπεδο.

Η τηλεπισκόπηση αποτελεί μία από τις πιο ακριβείς, οικονομικές και ευρέως χρησιμοποιούμενη προσέγγιση για τη μελέτη και παρακολούθηση των χρήσεων γης σε μεγάλες περιοχές [1] [2]. Με την παράλληλη χρήση δορυφορικών τηλεπισκοπικών δεδομένων, αλγορίθμων επεξεργασίας εικόνας καθώς και των γεωγραφικών συστημάτων πληροφοριών (GIS) καθίσταται δυνατή η διαχρονική μελέτη των αλλαγών κάλυψης γης σε λιγότερο χρόνο και με μεγαλύτερη ακρίβεια, συμβάλλοντας στην εξαγωγή χρήσιμων συμπερασμάτων [1]. Παράλληλα, η έλευση των υψηλής και πολύ υψηλής χωρικής ανάλυσης εικόνων (VHR) σε συνδυασμό με τις προηγμένες τεχνολογίες επεξεργασίας εικόνων και των γεωχωρικών εργαλείων, έχει οδηγήσει σε μία πιο αποτελεσματική και ακριβή παρακολούθηση και μοντελοποίηση των μοτίβων και χαρακτηριστικών των χρήσεων/καλύψεων γης.

Στο κεφάλαιο που ακολουθεί παρουσιάζονται ενδεικτικά διαθέσιμα παγκόσμια συστήματα κάλυψης γης.

1.1 Παραδείγματα Παγκόσμιων Συστημάτων Κάλυψης Γης

Με την ανάπτυξη της τεχνολογίας και τη συνεχή μείωση του κόστους απόκτησης δορυφορικών τηλεπισκοπικών δεδομένων, διάφορα παγκόσμια και τοπικά προϊόντα κάλυψης γης έχουν αναπτυχθεί τόσο σε πειραματικό όσο και σε επιχειρησιακό στάδιο. Ωστόσο, η πληθώρα των υπαρχόντων προϊόντων χαρακτηρίζονται από διαφορετικές προδιαγραφές και ακρίβειες στα αποτελέσματα και ως εκ τούτου αποτελεί σημαντικό παράγοντα η επιλογή του κατάλληλου συστήματος ανάλογα με τις ανάγκες των διάφορων χρηστών. Στο πλαίσιο αυτό έχουν πραγματοποιηθεί διάφορες αναλύσεις σύγκρισης των υφιστάμενων προϊόντων ως προς την αβεβαιότητα και ακρίβειά τους [3]. Τα πιο ευρέως διαθέσιμα και επιχειρησιακά παγκόσμια προϊόντα κάλυψης γης είναι τα: GLCC 2.0 [4], International Satellite Land Surface Climatology Project (ISLSCP II) [5], MODIS [6], UMD [7], GLC 2000 [8], GlobCover [9] [10], GLCNMO [11] [12], GLC-SHARE (Global Land Cover-SHARE) [13], GeoWiki [14], Climate Change Initiative-Land Cover (CCI-LC) [15], Global Land Cover 250

m China (GLC250 m_CN) [16], Finer Resolution Observation and Monitoring-Global Land Cover (FROM-GLC) [17], και GlobeLand30 [18]. Ακολουθεί πίνακας που παρουσιάζει συνοπτικά τα τεχνικά χαρακτηριστικά του κάθε συστήματος.

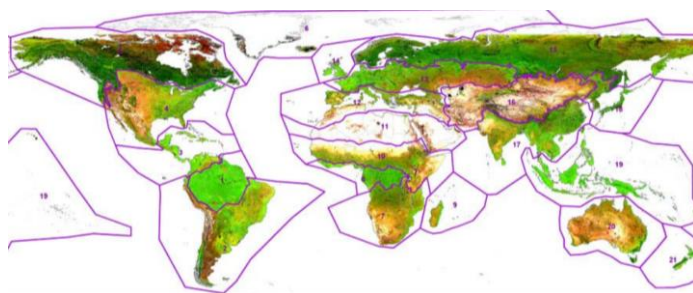
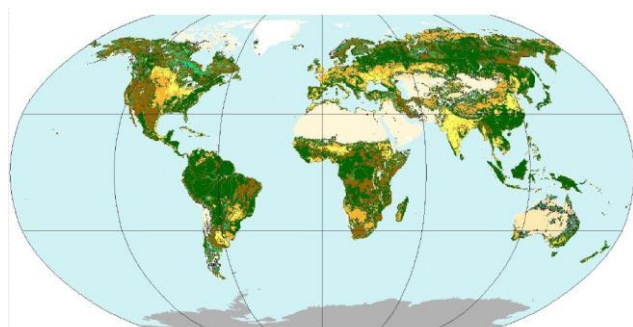
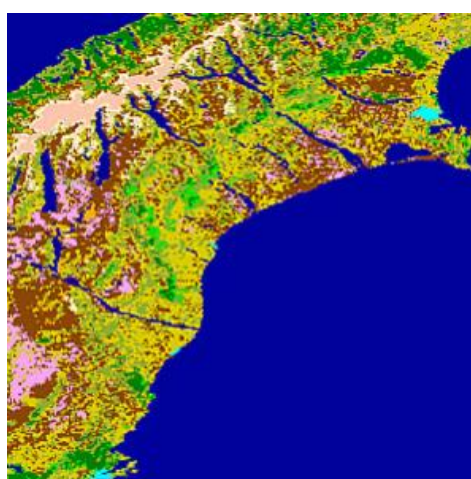
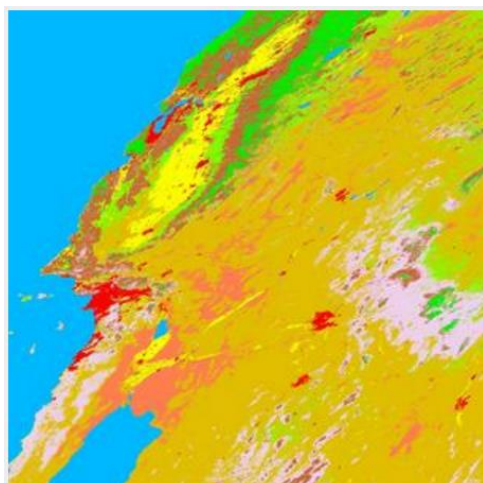
Data set/ product name	Spatial resolution	Overall accuracy (%)	Period of data acquisition	Sensor	Classification scheme-level I	Classification method	Change detection	Creator	Download data URL	Reference paper
GLCC 2.0 IGBP DISCover	1 km	66.9	Apr1992–Mar 1993	AVHRR	IGBP 17 classes	Unsupervised classification	Not supported	USGS, UNL	http://edc2.usgs.gov/glc/tabgoode_globe.php	Loveland et al. (2000)
ISLSCP II IGBP DISCover and SiB	0.25°, 0.5°, 1.0°	66.9	Apr1992–Mar 1993	AVHRR	IGBP 17 classes and SiB 15 classes	Unsupervised classification	Not supported	ORNL-DAAC/UMD	http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=969	Loveland et al. (2009)
MODIS Land Cover Type/ MCD12Q1 collection 5.1	500 m	71.6	Yearly since 2001	MODIS	IGBP 17 classes	Supervised classification	Not recommended	NASA	http://e4fil01.cr.usgs.gov/MOTA/MCD12Q1_051/	Friedl et al. (2010)
UMD	1 km	65.0	1992–1993	AVHRR	Simplified IGBP 12 classes	Supervised classification	Not supported	UMD	http://www.landcover.org/data/landcover/index.shtml	Hansen et al. (2000)
GLC 2000	1 km	68.6	Nov 1999–Dec 2000	SPOT4 VEGETATION	FAO LCCS 22 classes	Unsupervised classification	Not supported	JRC	http://bioval.jrc.ec.europa.eu/products/glc2000/products.php	Bartholomé and Belward (2005)
GlobCover				MERIS FR	FAO LCCS 22 classes	Supervised and unsupervised classification	Not recommended	ESA	http://due.esrin.esa.int/globcover/	Bicheron et al. (2008)
GlobCover2005 v2.2	300 m	73.1	2004–2006							Bontemps et al. (2011)
GlobCover2009	300 m	67.5	2009							
GLCNMO				MODIS	FAO LCCS 20 classes	Supervised classification	Not recommended	ISCGM	https://www.iscgm.org/gmd/	Tateishi et al. (2011)
GLCNMO V1	1 km	76.5	2003							Tateishi et al. (2014)
GLCNMO V2	500 m	77.9	2008							
GLC SHARE	1 km	80.2	Various, depending on database source	Various, depending on database source	FAO LCCS 11 classes	Data fusion	Not supported	FAO UN	http://www.glcn.org/databases/lc_glcshare_downloads_en.jsp	Latham et al. (2014)
GeoWiki			Data from: GLC2000	MODIS, SPOT4, MERIS FR	FAO LCCS 10 classes	GWR	Not supported	GeoWiki	http://geo-wiki.org/Application/index.php	See et al. (2015)
Hybrid 1	300 m	87.9	MODIS 2005							
Hybrid 2	300 m	82.8	GlobCover2005							
CCI-LC			Validation in 1998–2002 process	MERIS Full and Reduced Resolution/ SPOT VGT	FAO LCCS 22 classes	Unsupervised classification	Yes	ESA	http://maps.elie.ucl.ac.be/CCI/viewer/index.php	ESA (2014)
2000	300 m						Yes			
2005	300 m									
2010	300 m									
GLC250 m_CN				MODIS	11 classes	Random forest	Yes	CAS	Not available yet	Wang et al. (2015)
2001	250 m	74.93	2000–2001							
2010	250 m	75.17	2009–2011							
FROM-GLC					9 classes			CESSC	http://data.ess.tsinghua.edu.cn/	
FROM-GLC	30 m	63.69	Circa 2010	Landsat TM, ETM+		Supervised classification	No			Gong et al. (2013)
FROM-GLC-seg	30 m	64.42	Circa 2010	Landsat TM, ETM+		Supervised classification	No			Yu, Wang, and Gong (2013b)
FROM-GLC-agg	30 m	65.51	Circa 2010	Landsat TM, ETM+		Aggregation procedure	No			Yu, Wang, et al. (2014)
FROM-GLC-Hierarchy	30 m up to 100 km	69.50 for the 30 m	Circa 2010	Landsat TM, ETM+		Decision tree	No			Yu, Wang, et al. (2014)
GlobeLand30				Landsat TM, ETM7, HJ-1A/b/	10 classes	Supervised classification	Yes	UN/ NASG	http://www.globallandcover.com	Chen et al. (2015)
2000	30 m	78.6	2000							
2010	30 m	80.3	2010							

Εικόνα 1-1: Παγκόσμια συστήματα κάλυψης και χρήσεων γης (GLC) [19]

Το προϊόν **GLCC 2.0** διαθέτει χωρική ανάλυση 1km, αποτελείται από 17 γενικές κατηγορίες χρήσεων γης, καλύπτοντας την περίοδο 1992 – 1993, και παρουσιάζει μία λεπτομερή ερμηνεία της έκτασης των ανθρωπογενών παρεμβάσεων. Το συγκεκριμένο προϊόν χαρακτηρίζεται από συνολική ακρίβεια της τάξης του 70%. Το 2009 αξιοποιήθηκε στα πλαίσια της πρωτοβουλίας **ISLSCP II** [5], όπου συγχωνεύτηκε σε χωρικές αναλύσεις των 0.25°, 0.5°, και 1°.

Το προϊόν MODIS Land Cover Type (**MCD12Q1**) έχει σχεδιαστεί για να υποστηρίξει την έρευνα της τρέχουσας και εποχιακής κατάστασης καλύπτοντας 5 διαφορετικά συστήματα

ταξινόμησης. Το πρωτεύον είναι το IGBP, το οποίο προσδιορίζει 17 διαφορετικές κλάσεις με χωρική ανάλυση 500 m και ακρίβεια 71.6%. Το προϊόν **UMD GLC** δημιουργήθηκε από το πανεπιστήμιο του Maryland, από δορυφορικές εικόνες που λήφθηκαν το διάστημα 1992 και 1993. Το συγκεκριμένο σύστημα μπορεί να διακρίνει 12 κλάσεις σε χωρική ανάλυση 1 km και με ακρίβεια 65%. Το έργο **GLC 2000**, βασίζεται στη συνεργασία του Οργανισμού Τροφίμων και Γεωργίας των Ηνωμένων Εθνών με το Κοινό Κέντρο Ερευνών (JRC), χρησιμοποιώντας δεδομένα που αποκτήθηκαν από το δορυφόρο SPOT 4. Παράλληλα, έχουν αναπτυχθεί συστήματα, όπως είναι το Land Cover Classification System (**FAO LCCS**) [20], το οποίο εστιάζει στα κύρια οικολογικά συστήματα όπως είναι τα δάση, τα λιβάδια και οι καλλιεργήσιμες εκτάσεις. Το συγκεκριμένο σύστημα, μπορεί να ανιχνεύσει 23 κλάσεις με ακρίβεια 68.6 % στο σύνολο των δεδομένων. Επιπλέον, έχουν αναπτυχθεί παγκόσμια συστήματα, τα οποία αξιοποιούν τις προδιαγραφές και τα δεδομένα από προηγούμενα συστήματα, προκείμενου να παράγουν μεγαλύτερης ακρίβειας προϊόντα. Χαρακτηριστικό παράδειγμα είναι το προϊόν **GlobCover**, που αποτελεί το πρώτο σύστημα παραγωγής χαρτών χωρικής ανάλυσης 300 m και 22 διαφορετικών κλάσεων που προέρχονται από το FAO LCCS σύστημα. Καλύπτει δύο διαφορετικές περιόδους (Δεκέμβριος 2004 – Ιούνιος 2006, Ιανουάριος 2009 – Δεκέμβριος 2009) με συνολική ακρίβεια 73.1 % [21] [22] [23].



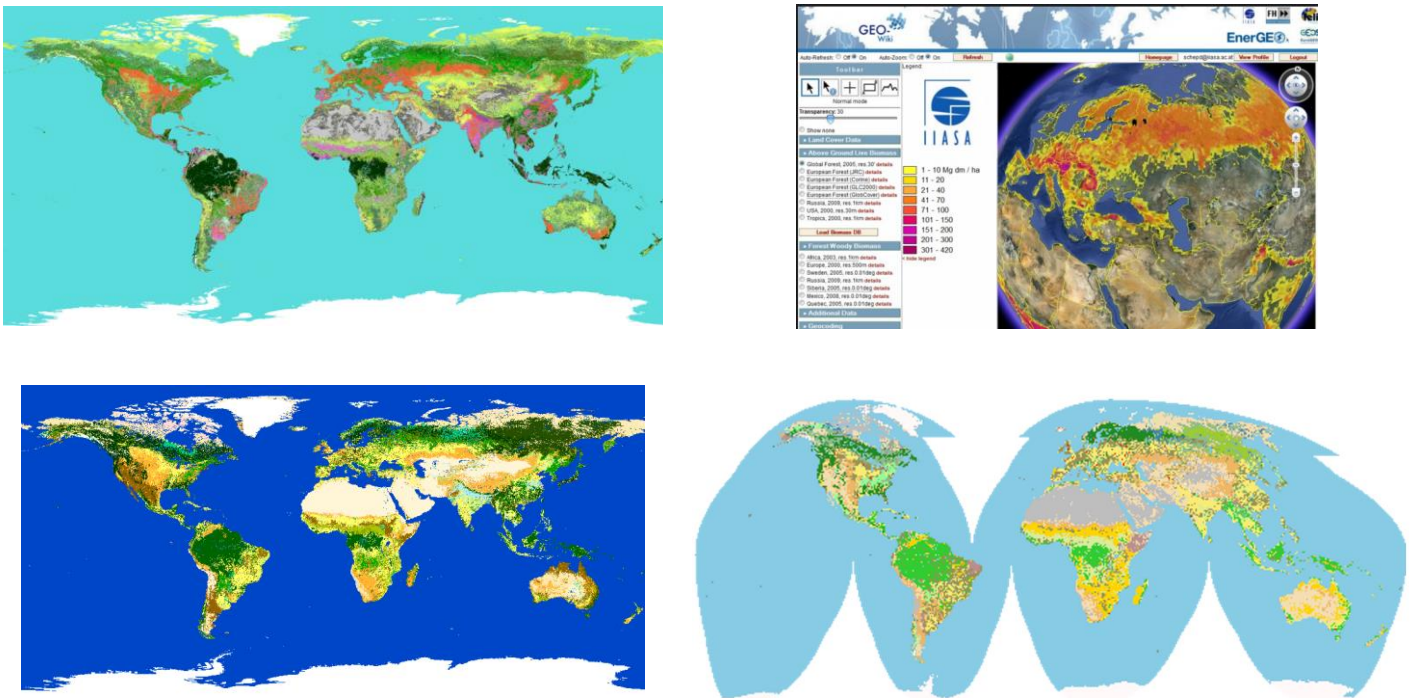
Εικόνα 1-2: Παραδείγματα των προϊόντων: MCD12Q1 – Western United States (πάνω αριστερά), UMD GLC – New Zealand (πάνω δεξιά), GLC 2000 – global (κάτω αριστερά), GlobCover – global (κάτω δεξιά)

Επιπλέον, το σύστημα **GLCNMO** έχει δημοσιευθεί σε δύο εκδόσεις, όπου η πρώτη καλύπτει 20 διαφορετικές κλάσεις, με χωρική ανάλυση 1 km (ακρίβειας 76.5 %) και βασίζεται σε δεδομένα MODIS του 2003. Η δεύτερη έκδοση, γνωστή ως GLCMNO2008, έχει 20 κλάσεις με

χωρική ανάλυση 500 m και βασίζεται σε δεδομένα MODIS που λήφθηκαν το 2008, με ακρίβεια 77.9%.

Μία διαφορετική προσέγγιση στη χαρτογράφηση των καλύψεων γης είναι το δίκτυο **Geo-Wiki**. Βασίζεται στην ιδέα των volunteered geographic information, όπου οι χρήστες σε όλον τον κόσμο αξιολογούν τα αποτελέσματα από τρία παγκόσμια συστήματα (GLC 2000, MODIS 2005, GlobCover 2005) βελτιώνοντας κάθε φορά τα αποτελέσματα. Παραδείγματα τέτοιων προϊόντων ενσωματώνουν αλγορίθμους γεωγραφικής σταθμισμένης παλινδρόμησης έτσι ώστε να παραχθούν υβριδικοί χάρτες καλύψεων γης που συνδυάζουν τα τηλεπισκοπικά δεδομένα με τη γνώση από τους χρήστες.

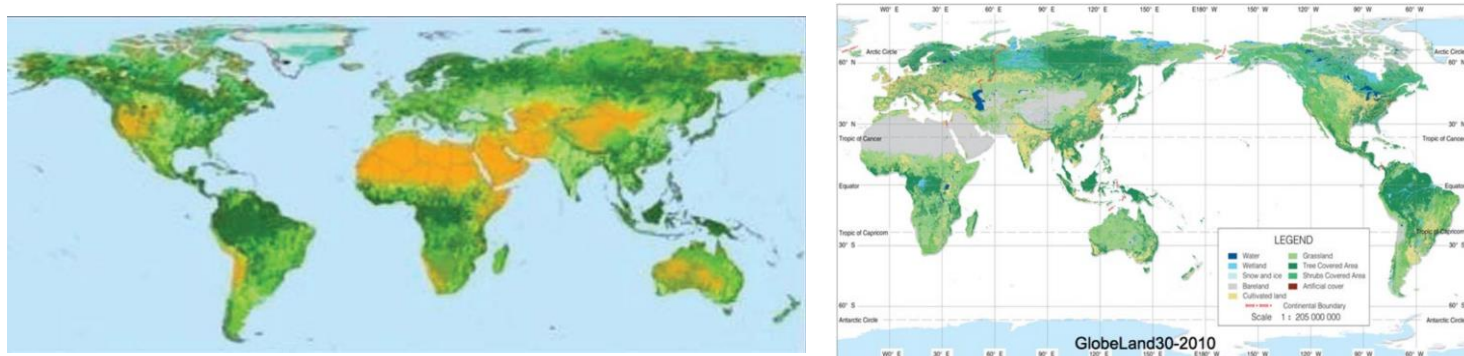
Παράλληλα, ο Ευρωπαϊκός Οργανισμός Διαστήματος (ESA), δημοσίευσε το 2014 το προϊόν **CCI-LC**, χωρικής ανάλυσης 300 m, για τρεις διαφορετικές εποχές έτσι ώστε να αξιοποιηθεί σε εφαρμογές κλιματικής μοντελοποίησης [15]. Επιπλέον, το συγκεκριμένο προϊόν παράγει χάρτες αλλαγών για τις περιόδους 2000 – 2005 και 2005 – 2010, το οποίο είναι και το μοναδικό στη συγκεκριμένη χωρική ανάλυση. Όμοια το **GLC250 m_CN** είναι το μοναδικό με χωρική ανάλυση 250 m, που προσφέρεται για τα έτη 2001 και 2010 (συμπεριλαμβανομένων και των αντίστοιχων μεταβολών) και λαμβάνει υπόψη ετήσια δεδομένα MODIS. Η ακρίβεια για το 2001 είναι 74.93 %, ενώ για το 2010 75.17% [17].



Εικόνα 1-3: Παραδείγματα των προϊόντων: GLCNMO – global (πάνω αριστερά), Geo-Wiki (πάνω δεξιά), CCI-LC - global (κάτω αριστερά), GLCC 2.0 - global (κάτω δεξιά)

Τα πιο σύγχρονα παγκόσμια συστήματα κάλυψης γης παράγονται σε χωρική ανάλυση 30 m, προσθέτοντας σημαντική πληροφορία σε σχέση με τα προγενέστερα συστήματα. Το πρώτο προϊόν που δημιουργήθηκε και χρησιμοποιεί δεδομένα από τον Landsat για το έτος 2010 είναι το **FROM-GLC** [24]. Το συγκεκριμένο σύστημα αξιοποιεί τέσσερις διαφορετικούς ταξινομητές για την παραγωγή τεσσάρων χαρτών κάλυψης γης, με 9 κλάσεις, στην πρώτη ιεραρχία, και 25 στο δεύτερο επίπεδο. Η μέγιστη ακρίβεια των τεσσάρων προϊόντων είναι

63.69%, κάτι που δημιουργεί περιορισμούς στην επιχειρησιακή χρήση του, ιδιαίτερα σε εφαρμογές που σχετίζονται με τη μελέτη φαινολογικών χαρακτηριστικών. Λαμβάνοντας υπόψη το προαναφερθέν, δημιουργήθηκε μία νεότερη έκδοσή του, με ονομασία **FROMGLC-Seg** [24], η οποία ενσωματώνει δεδομένα πολλαπλών χωρικών αναλύσεων από διαφορετικούς δέκτες, όπως είναι για παράδειγμα οι χρονοσειρές MODIS enhanced vegetation index (EVI), βιοκλιματικές και εδάφους μεταβλητές για τη βελτίωση της ακρίβειας. Στη συνέχεια το FROM-GLC και το FROMGLC-Seg συγχωνεύθηκαν για τη δημιουργία του προϊόντος **FROM-GLC-agr** [17] χωρικής ανάλυσης 30 m και ακρίβειας 65.51 %. Τέλος, η πιο πρόσφατη έκδοση παγκόσμιου συστήματος κάλυψης γης είναι το **GlobeLand30**, το οποίο κυκλοφόρησε το 2014. Το συγκεκριμένο δόθηκε στα Ηνωμένα Έθνη από την Εθνική Κινεζική Διοίκηση Τοπογραφίας, Χαρτογράφησης και Γεωπληροφορίας (NASG) με χωρική ανάλυση 30 m και 10 διαφορετικών κατηγοριών για τα έτη 2000 και 2010. Το συγκεκριμένο προϊόν επιτυγχάνει ακρίβειες της τάξης του 80%.



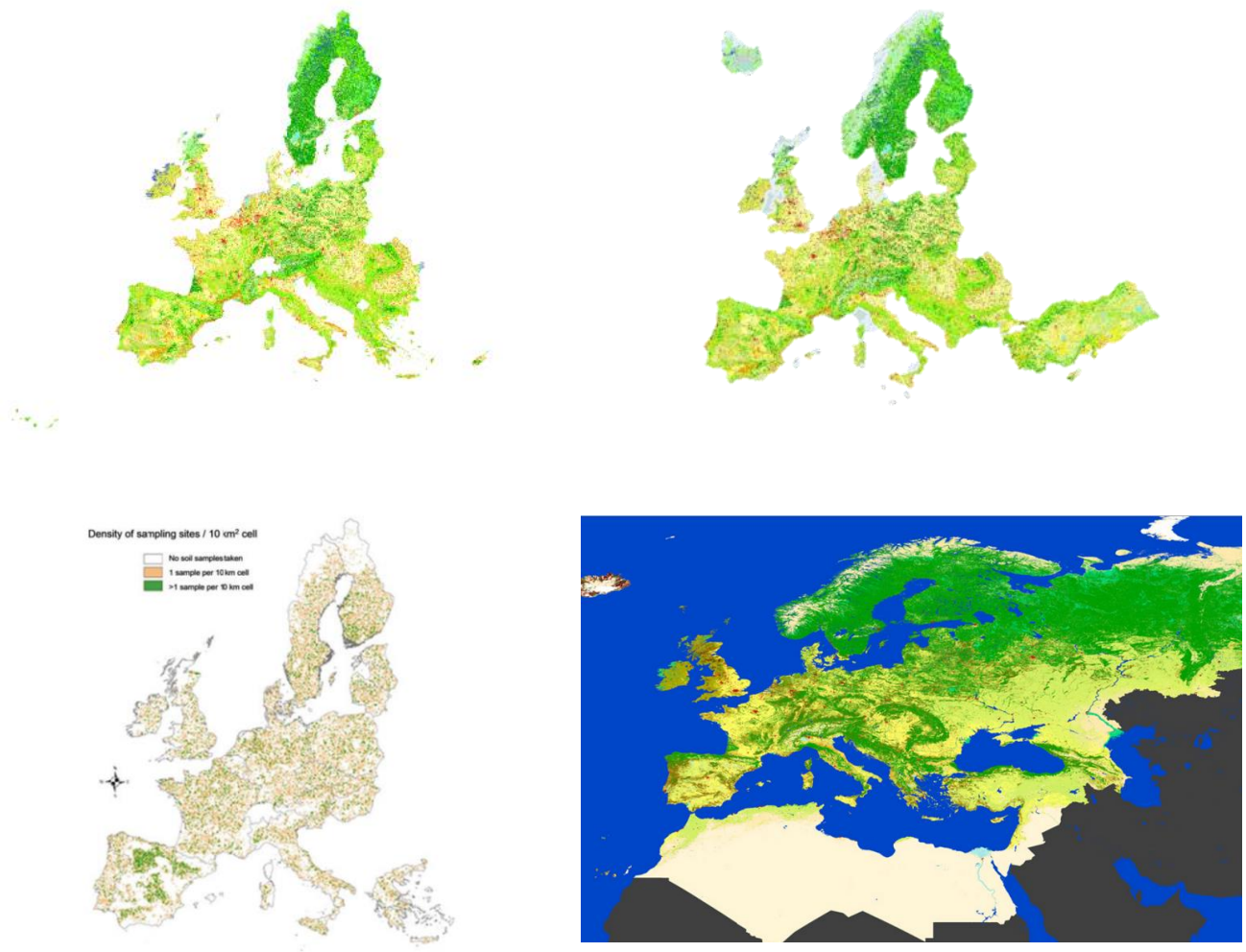
Εικόνα 1-4: Παραδείγματα των προϊόντων: FROM-GLC-agr – global (αριστερά), GlobeLand30 - global (δεξιά)

Όσον αφορά τα Ευρωπαϊκά συστήματα, έχουν αναπτυχθεί αρκετά προϊόντα κάλυψης των χρήσεων γης. Πιο συγκεκριμένα, το πρόγραμμα **CORINE** είναι το πρώτο που είχε ως στόχο την παραγωγή χαρτών χρήσεων γης και βασίστηκε σε φωτοερμηνεία δορυφορικών εικόνων, διαφορετικών αναλύσεων και δεκτών, τα έτη αναφορά 1990, 2000 και 2006. Στα πλαίσια του συγκεκριμένου προγράμματος έχουν αναπτυχθεί δύο διαφορετικά προϊόντα με χωρική ανάλυση 100 m και 250 m με τους αντίστοιχους χάρτες μεταβολών. Μία νέα έκδοση του CORINE, με ονομασία CLC2012, έχει ήδη δημοσιευθεί από την Ευρωπαϊκή Υπηρεσία Περιβάλλοντος, αλλά δεν έχει ακόμα επιτευχθεί πλήρη κάλυψη.

Ένα πιο πρόσφατο σύστημα κάλυψης γης για την Ευρώπη είναι το προϊόν Land Use and Cover Area frame Survey (**LUCAS**). Σε αντίθεση με τα προηγούμενα συστήματα το LUCAS βασίζεται σε διαφορετική μεθοδολογική προσέγγιση αξιοποιώντας δεδομένα πεδίου και αεροφωτογράφισης, έτσι ώστε να υπάρχει πιο ολοκληρωμένη και ακριβή χαρτογράφηση. Το συγκεκριμένο σύστημα ανανεώνεται κάθε 3 χρόνια, προσφέροντας χάρτες αλλαγών, και ξεκίνησε το 2006 από τους [25] [26] [27].

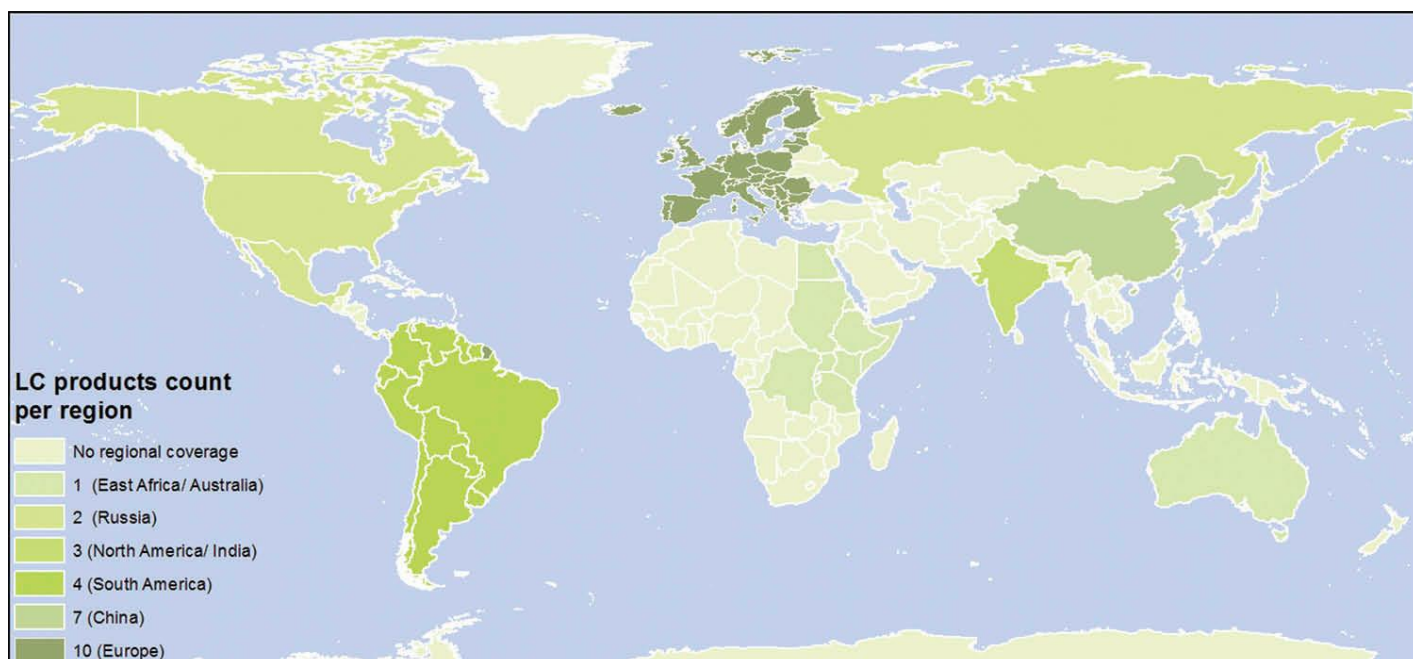
Παράλληλα έχουν αναπτυχθεί ευρωπαϊκά συστήματα, τα οποία βασίζονται σε αντίστοιχα παγκόσμια. Για παράδειγμα το **GlobCorine** βασίζεται σε σημαντικό βαθμό στο GlobCover, το οποίο παρουσιάστηκε νωρίτερα, για τα έτη 2005 και 2009 με χωρική ανάλυση 300 m. Το

GlobCorine ακολουθεί την ίδια διάταξη ταξινόμησης με το Corine Land Cover (CLC) σε συνδυασμό με το UN LCSS [28] [29].



Εικόνα 1-5: Παραδείγματα των προϊόντων: CORINE 2000 (πάνω αριστερά), CORINE 2006 (πάνω δεξιά), LUCAS 2009 TOPSOIL (κάτω αριστερά), GlobCorine (κάτω δεξιά)

Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα εθνικά προϊόντα κάλυψης χρήσεων γης. Όπως περιεγράφηκε και παραπάνω οι χώρες της Ευρώπης, Κίνας και Αμερικής έχουν πληθώρα προϊόντων κάλυψης γης και ιδιαίτερα για τη δεκαετία 2000-2010. Αντίθετα, η Ασία και η Αφρική δεν έχουν ολοκληρωμένα εθνικά προϊόντα, τα οποία να εστιάζουν αποκλειστικά για τις συγκεκριμένες περιοχές. Πιο συγκεκριμένα, η Ανατολική Αφρική καλύπτεται αποκλειστικά από το **Africover** και η Ασία καλύπτεται αραιά από εθνικές προσπάθειες χωρίς να υπάρχει πλήρη κάλυψη. Στην εικόνα που ακολουθεί παρουσιάζεται συνοπτικά ο αριθμός των προϊόντων κάλυψης γης ανά περιοχή.



Εικόνα 1-6: Εθνικά συστήματα, σε παγκόσμια κλίμακα, κάλυψης και χρήσεων γης [19]

Συνοψίζοντας, γίνεται αντιληπτό πως υπάρχει ένα ευρύ φάσμα υφιστάμενων προϊόντων κάλυψης χρήσεων γης χωρίς ωστόσο να έχουν λειτουργικό και επιχειρησιακό χαρακτήρα σε παγκόσμια κλίμακα. Παρόλο που η χρησιμότητα των συγκεκριμένων προϊόντων είναι ευρέως γνωστή υπάρχουν αρκετοί περιορισμοί, όπως είναι παράδειγμα η χρήση δεδομένων χαμηλής χωρικής ανάλυσης, διαφορετικών δεδομένων εισόδου, έλλειψη αξιόπιστων και αυτόματων μεθοδολογιών παραγωγής χαρτών κάλυψης γης καθώς και το ότι τα αποτελέσματα διαφέρουν σε σημαντικό βαθμό ανάλογα με την περιοχή.

Η χρήση των δορυφορικών εικόνων με χωρική ανάλυση 30 μέτρων έχει πρόσφατα οδηγήσει στη δημιουργία δύο παγκόσμιων συστημάτων κάλυψης χρήσεων γης (GlobeLand30 και FROM-GLC), ενώ τα υπόλοιπα ξεκινούν από τα 250 μέτρα. Σε περιφερειακό επίπεδο μόνο οι ΗΠΑ, ο Καναδάς, η Κίνα, η Νότια Αμερική και η Ανατολική Αφρική έχουν αντίστοιχα προϊόντα στα 30 μέτρα.

Παρά το γεγονός πως για συγκεκριμένες εφαρμογές οι χωρικές αναλύσεις κάτω των 250 m μπορούν να προσφέρουν ικανοποιητικά αποτελέσματα, το μεγαλύτερο μέρος των εφαρμογών απαιτούν υψηλότερη ανάλυση. Για παράδειγμα, δεδομένα με χαμηλή και μέτρια χωρική ανάλυση είναι συνήθως ακατάλληλα για την ταξινόμηση και μελέτη αστικών εκτάσεων, διαφοροποίηση τύπων καλλιεργειών κ.τ.λ. και ως εκ τούτου η χρήση δορυφορικών δεδομένων πολύ υψηλής ανάλυσης αποτελεί βασική προϋπόθεση. Τέλος, ένας σημαντικός περιορισμός των υφιστάμενων προϊόντων έγκειται στην έλλειψη διαχρονικών προϊόντων, όπου περιορίζονται συνήθως σε συγκεκριμένα έτη ή δεκαετίες ανάλογα με το προϊόν, δημιουργώντας ασυνέχειες σε εφαρμογές που απαιτούν ανάλυση χρονοσειρών.

2. ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΑΙΧΜΗΣ ΓΙΑ ΤΗΝ ΤΑΞΙΝΟΜΗΣΗ ΔΟΡΥΦΟΡΙΚΩΝ ΕΙΚΟΝΩΝ

Για την ταξινόμηση δορυφορικών εικόνων με σκοπό τη δημιουργία συστημάτων παραγωγής δεδομένων χρήσεων γης συμβατικοί αλγόριθμοι σε πολλές περιπτώσεις δεν αντιμετωπίζουν αποτελεσματικά το πρόβλημα, καθώς απαιτείται η ταυτόχρονη επεξεργασία διαχρονικών και από διαφορετικές πηγές εικόνων. Λαμβάνοντας υπόψη τα παραπάνω, είναι απαραίτητη η χρήση εξειδικευμένων αλγορίθμων ταξινόμησης που στηρίζονται σε αλγορίθμους υπολογιστικής νοημοσύνης. Οι συγκεκριμένοι αλγόριθμοι εκμεταλλεύονται διαφόρου τύπου δεδομένων και μέσω διαδικασιών εξόρυξης δεδομένων (data mining) δημιουργούν στατιστικά μοντέλα που αντιμετωπίζουν αποτελεσματικότερα τα προβλήματα ταξινόμησης. Παραδείγματα τέτοιου είδους αλγορίθμων είναι ο SVM, τα στατιστικά μοντέλα decision trees, random forest και άλλα [30].

Στα κεφάλαια που ακολουθούν παρουσιάζονται οι κυριότεροι αλγόριθμοι ταξινόμησης που χρησιμοποιούν την τεχνολογία της μηχανικής μάθησης (machine learning) και χρησιμοποιούνται σε εφαρμογές ταξινόμησης σε μεγάλα διαχρονικά γεωχωρικά δεδομένα. Οι συγκεκριμένοι αλγόριθμοι εντάσσονται στην ευρύτερη κατηγορία των διαδικασιών εξόρυξης δεδομένων (data mining) [31].

2.1 Data Mining

Η εξόρυξη δεδομένων (data mining) είναι μία αναλυτική διαδικασία που σχεδιάστηκε για την ανάλυση μεγάλων δεδομένων με σκοπό την αναζήτηση προτύπων ή συστηματικών σχέσεων μεταξύ των μεταβλητών. Στη συνέχεια, επικυρώνονται τα παραπάνω πορίσματα για την ανίχνευση μοτίβων σε νέα υποσύνολα δεδομένων σε εφαρμογές που απαιτείται πρόβλεψη χαρακτηριστικών. Η διαδικασία της εξόρυξης δεδομένων αποτελείται από τρία στάδια: 1) αρχική έρευνα, 2) μοντελοποίηση και επικύρωση, 3) ανάπτυξη και εφαρμογή του μοντέλου σε νέα δεδομένα ώστε να παραχθούν προβλέψεις. [32] [33] [34]

Στις εφαρμογές εξόρυξης δεδομένων χρησιμοποιούνται σε ευρεία κλίμακα διαδικασίες μηχανικής μάθησης μέσω αλγορίθμων ταξινόμησης. Παραδείγματα τέτοιου είδους αλγορίθμων ταξινόμησης είναι τα παρακάτω:

- Linear classifiers
 - Fisher's linear discriminant
 - Logistic regression
 - Naive Bayes classifier
 - Perceptron
- Support vector machines
 - Least squares support vector machines
- Quadratic classifiers
- Kernel estimation
 - k-nearest neighbor
- Boosting (meta-algorithm)

- Decision trees
 - Random forests
- Neural networks
- Gene Expression Programming
- Bayesian networks
- Hidden Markov models

Ακολουθεί πίνακας που παρουσιάζει τα πιο γνωστά και αποτελεσματικά ελεύθερα λογισμικά καθώς και τις βιβλιοθήκες που χρησιμοποιούνται για διαδικασίες υπολογιστικής μάθησης.

Ελεύθερα λογισμικά			
Όνομασία	Κύρια χαρακτηριστικά	Αλγόριθμοι Ταξινόμησης	Ιστότοπος
Apache Mahout	java cross platform Apache 2.0 licence	Logistic Regression Naive Bayes Random Forest Hidden Markov Models Multilayer Perceptron	http://mahout.apache.org
mlpy	Python, C, C++ Cross platform GPL licence	Linear Discriminant Analysis Logistic Regression SVM KNN Classification Tree	http://mlpy.sourceforge.net
OpenCV	C, C++ Cross platform BSD licence	Boosting Decision tree KNN Naive Bayes classifier ANN SVM etc	http://opencv.org
R	C, Fortran, R language Cross platform GNU licence	Neural Networks Random Forests Recursive Partitioning Boosting SVM Bayesian methods etc	http://www.r-project.org
Scikit-learn	Python, C, C++ Cross platform BSD Licence	Svm Generalized Linear Models Stochastic Gradient Descent KNN Decision Tree Naive Bayes etc	http://scikit-learn.org/stable/
Weka	Java	AdaBoost	http://www.cs.waikato.ac.nz/ml/weka/

	Cross platform GNU licence	Bayesian methods SVM Naive Bayes Random Forest ANN etc	
--	-------------------------------	--	--

Πίνακας 2-1: Ελεύθερα λογισμικά και βιβλιοθήκες που χρησιμοποιούνται για διαδικασίες υπολογιστικής μάθησης

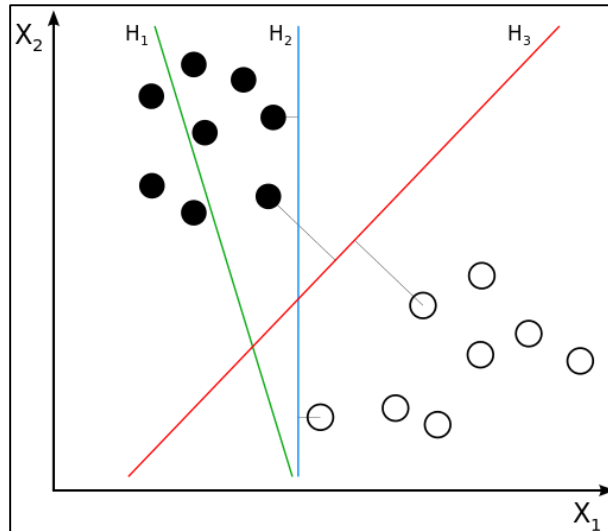
2.2 Μέθοδοι Ταξινόμησης και Μηχανική Μάθηση

Στο συγκεκριμένο κεφάλαιο παρουσιάζονται συνοπτικά διάφοροι αλγόριθμοι ταξινόμησης που βασίζονται σε διαδικασίες μηχανικής μάθησης και χρησιμοποιούνται ευρέως σε εφαρμογές που απαιτούν χρήση μεγάλων και ανομοιογενών δεδομένων. Πιο συγκεκριμένα, περιγράφονται οι εξής μέθοδοι: SVM, Decision tree, K-nearest neighbors, οι οποίοι χρησιμοποιούνται σε ευρεία κλίμακα σε εφαρμογές ταξινόμησης δορυφορικών εικόνων.

2.2.1 SVM

Οι μηχανές υποστήριξης διανυσμάτων (Support Vector Machines) είναι επιβλεπόμενα μοντέλα μάθησης που συνδυάζονται με αλγόριθμους εκμάθησης, με στόχο την ανάλυση δεδομένων και την αναγνώριση προτύπων για διαδικασίες ταξινόμησης και κατάτμησης (regression). Παρέχοντας ένα σύνολο δεδομένων εκπαίδευσης, το καθένα από αυτά διαχωρίζεται σε μία από τις δύο κατηγορίες και κάθε φορά ο αλγόριθμος εκπαίδευσης, SVM, χτίζει ένα μοντέλο που διαχωρίζει τα νέα δεδομένα στη μία ή στην άλλη κατηγορία δημιουργώντας ένα μη πιθανοτικό, γραμμικό και δυαδικό μοντέλο. Το μοντέλο SVM, πρακτικά, είναι μια αναπαράσταση των δεδομένων στο χώρο, τα οποία απεικονίζονται ως κατηγορίες διαχωρισμένες από ένα σαφές όριο, το οποίο χαρακτηρίζεται από συγκεκριμένο εύρος κάθε φορά. Τα νέα δεδομένα αντιστοιχίζονται στο συγκεκριμένο διάστημα-σύνολο που ανήκουν με βάση την πλευρά του ορίου που βρίσκονται. Εκτός από τη γραμμική μορφή ταξινόμησης, τα μοντέλα SVM μπορούν να εκτελέσουν αποτελεσματικά μη γραμμικές ταξινομήσεις χρησιμοποιώντας εξειδικευμένα τεχνάσματα του πυρήνα τους (kernel) και αντιμετωπίζοντας πολύπλοκα προβλήματα.

Οι μηχανές υποστήριξης διανυσμάτων κατασκευάζουν ένα υπερεπίπεδο (hyperplane) ή ένα σύνολο υπερεπιπέδων σε έναν υψηλό ή άπειρων διαστάσεων χώρο, το οποίο μπορεί να χρησιμοποιηθεί για ταξινόμηση, κατάτμηση ή άλλες εφαρμογές. Ένας καλός διαχωρισμός επιτυγχάνεται από το υπερεπίπεδο που έχει τη μεγαλύτερη δυνατή απόσταση από το πλησιέστερο σημείο των δεδομένων εκπαίδευσης, καθώς όσο μεγαλύτερο είναι το περιθώριο τόσο μικρότερη είναι η πιθανότητα λάθους από τον ταξινομητή [35].



Εικόνα 2-1: Παράδειγμα εύρεσης υπερεπιπέδου μεταξύ δύο κλάσεων μέσω του αλγορίθμου SVM

Στο παραπάνω παράδειγμα γίνεται αντιληπτό πως το όριο H1 δεν διαχωρίζει τις κλάσεις. Αντίθετα το όριο H2 διαχωρίζει τις κλάσεις παρέχοντας ωστόσο μικρό περιθώριο, ενώ το όριο H3 διαχωρίζει τις κλάσεις με το μέγιστο δυνατό περιθώριο.

Linear SVM

Έστω κάποια δεδομένα εκπαίδευσης D και μία σειρά n σημείων της μορφής:

$$D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^p, \quad y_i \in \{-1, 1\}\}_{i=1}^n$$

όπου τα y_i είναι 1 ή -1, υποδεικνύοντας την κλάση στην οποία το σημείο x_i ανήκει.

Κάθε σημείο x_i είναι ένα διάνυσμα p -διαστάσεων. Στόχος είναι να βρεθεί το μέγιστο περιθώριο του υπερεπιπέδου που διαχωρίζει τα σημεία με τιμή $y_i = 1$ από αυτά που έχουν τιμή $y_i = -1$. Κάθε υπερεπίπεδο μπορεί να γραφτεί ως το σύνολο των σημείων x που ικανοποιούν την παρακάτω μαθηματική σχέση:

$$w \cdot x - b = 0$$

όπου w το «κανονικό» διάνυσμα (normal vector) του υπερεπιπέδου.

Η παράμετρος $\frac{b}{\|w\|}$ δηλώνει την απόκλιση (offset) του υπερεπιπέδου από την αρχή και κατά μήκος του «κανονικού» διανύσματος w .

Εάν τα δεδομένα εκπαίδευσης είναι γραμμικώς διακριτά, είναι δυνατή η επιλογή δύο υπερεπιπέδων με τέτοιο τρόπο ώστε τα δεδομένα να διαχωρίζονται και να μην υπάρχουν σημεία μεταξύ τους. Στη συνέχεια, γίνεται προσπάθεια μεγιστοποίησης της απόστασής τους. Η περιοχή που οριοθετείται από αυτά ονομάζεται περιθώριο (margin). Τα παραπάνω υπερεπίπεδα μπορούν να περιγραφούν από τις εξής εξισώσεις:

$$w \cdot x - b = 1$$

και

$$w \cdot x - b = -1$$

Με εφαρμογή γεωμετρίας βρίσκεται η απόσταση μεταξύ των δύο υπερεπιπέδων (hyperplanes) που είναι $\frac{2}{\|w\|}$ και στόχος είναι η ελαχιστοποίηση του όρου $\|w\|$. Ωστόσο, επειδή πρέπει να τηρείται και ο περιορισμός των δεδομένων, δηλαδή να μην βρεθούν αυτά στο περιθώριο (margin), προστίθεται ένας παραπάνω περιορισμός για κάθε σημείο i :

$$w \cdot x_i - b \geq 1, \quad \text{για κάθε } x_i \text{ της πρώτης κλάσης}$$

ή

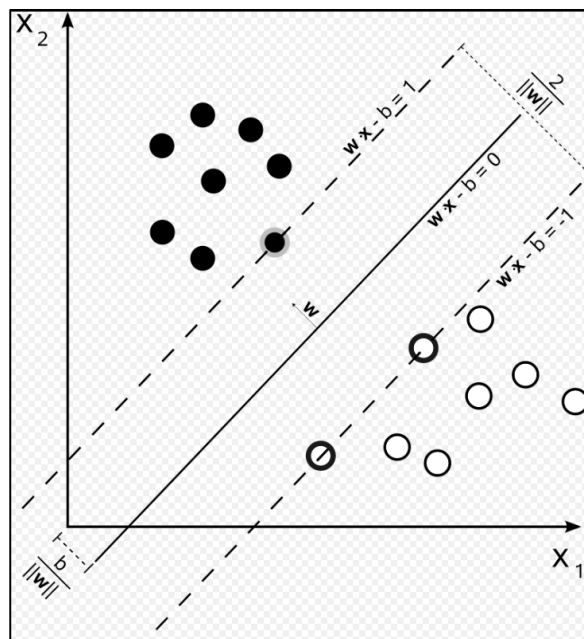
$$w \cdot x_i - b \leq -1, \quad \text{για κάθε } x_i \text{ της δεύτερης κλάσης}$$

Οι παραπάνω δύο περιορισμοί μπορούν να γραφτούν και ως εξής:

$$y_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n$$

Χρησιμοποιώντας τις παραπάνω σχέσεις καταστρώνεται το πρόβλημα βελτιστοποίησης:

Ελαχιστοποίηση w, b υπό τους περιορισμούς $y_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n$



Εικόνα 2-2: Πρόβλημα βελτιστοποίησης περιθωρίου σε γραμμικό ταξινομητή SVM

Για την επίλυση του προβλήματος της βελτιστοποίησης που περιεγράφηκε παραπάνω έχουν προταθεί διάφορες λύσεις, ωστόσο στα πλαίσια της συγκεκριμένης διπλωματικής εργασίας περιγράφονται δύο κλασσικές μέθοδοι βελτιστοποίησης που έχουν χρησιμοποιηθεί για το σκοπό αυτό.

- **Primal form**

Η δυσκολία του συγκεκριμένου προβλήματος βελτιστοποίησης έγκειται στο γεγονός πως η επίλυση του στηρίζεται στον όρο $\|w\|$, ο οποίος απαιτεί τη χρήση τετραγωνικής ρίζας. Για το συγκεκριμένο λόγο πραγματοποιείται αντικατάσταση της σχέσης $\|w\|$ με τη σχέση $\frac{1}{2}\|w\|^2$ για τη διευκόλυνση των μαθηματικών πράξεων, χωρίς όμως να αλλάζει το αποτέλεσμα της επίλυσης. Η επίλυση του προβλήματος της βελτιστοποίησης ανάγεται σε ένα πρόβλημα βελτιστοποίησης τετραγωνικού προγραμματισμού (quadratic programming optimization) [36].

Πιο συγκεκριμένα ισχύουν οι εξής σχέσεις:

$$\operatorname{argmin}_{(w,b)} \frac{1}{2} \|w\|^2, \text{ υπό τους περιορισμούς } y_i(w \cdot x_i - b) \geq 1, 1 \leq i \leq n$$

Εντάσσοντας τους πολλαπλασιαστές του Lagrange α , το προηγούμενο πρόβλημα περιορισμών μπορεί να εκφραστεί ως εξής:

$$\operatorname{argmin}_{(w,b)} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(w \cdot x_i - b) - 1] \right\}$$

Ο μετασχηματισμός του αρχικού προβλήματος βελτιστοποίησης στην παραπάνω σχέση οδηγεί στην επίλυσή του με τεχνικές τετραγωνικού προγραμματισμού, χρησιμοποιώντας γραμμικούς συνδυασμούς μεταξύ των διανυσμάτων εκπαίδευσης.

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

Μόνο ορισμένα α_i θα είναι μεγαλύτερα του μηδενός. Τα αντίστοιχα x_i είναι διανύσματα υποστήριξης, τα οποία βρίσκονται στο περιθώριο και καλύπτουν τη σχέση $y_i(w \cdot x_i - b) = 1$. Από το προηγούμενο εξάγεται το συμπέρασμα πως τα συγκεκριμένα διανύσματα θα καλύπτουν επίσης τη σχέση:

$$w \cdot x_i - b = \frac{1}{y_i} = y_i \leftrightarrow b = w \cdot x_i - y_i$$

μέσω της οποίας είναι δυνατός ο προσδιορισμός της απόκλισης (offset) b , χρησιμοποιώντας το μέσο όρο όλων των διανυσμάτων υποστήριξης από τη σχέση:

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (w \cdot x_i - y_i)$$

- **Dual form**

Γράφοντας τον κανόνα ταξινόμησης στη διπλή μορφή του γίνεται αντιληπτό πως το μέγιστο περιθώριο του υπερεπιπέδου και συνεπώς της διαδικασίας της ταξινόμησης είναι μία συνάρτηση των διανυσμάτων υποστήριξης και του υποσυνόλου των δεδομένων εκπαίδευσης που βρίσκονται στο περιθώριο.

Χρησιμοποιώντας τη σχέση $\|w\|^2 = w \cdot w$ και αντικαθιστώντας την τιμή του w με $\sum_{i=1}^n (w \cdot x_i - y_i)$ γίνεται αντιληπτό πως οδηγούμαστε στο παρακάτω πρόβλημα βελτιστοποίησης [35]:

$$\text{Ελαχιστοποίηση του } a_i: \tilde{L}(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j k(x_i, x_j)$$

υπό τους περιορισμούς: $0 \leq a_i \leq C$, $1 \leq i \leq n$ και $\sum_{i=1}^n a_i y_i = 0$

Nonlinear SVM

Η δημιουργία μη γραμμικών ταξινομητών πραγματοποιείται με την εφαρμογή τεχνάσματος στον πυρήνα (kernel) που στοχεύει στην μεγιστοποίηση του περιθωρίου στα υπερεπίπεδα. Ο αλγόριθμος που προκύπτει είναι τυπικά ο ίδιος, εκτός από το ότι κάθε εσωτερικό γινόμενο αντικαθίσταται από μία μη γραμμική συνάρτηση του πυρήνα. Αυτό επιτρέπει στον αλγόριθμο να «χωρέσει» το μέγιστο περιθώριο του υπερεπιπέδου σε ένα μετασχηματισμένο χώρο χαρακτηριστικών [37].

Μερικοί από τους πιο συνηθισμένους πυρήνες είναι οι εξής:

- Polynomial (ομογενείς): $k(x_i, x_j) = (x_i \cdot x_j)^d$
- Polynomial (ανομοιογενείς): $k(x_i, x_j) = (x_i \cdot x_j + 1)^d$
- Gaussian radial basis function: $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$
- Hyperbolic tangent: $k(x_i, x_j) = \tanh(kx_i \cdot x_j + c)$, $k > 0$ $c < 0$

Multiclass SVM

Η μηχανή υποστήριξης διανυσμάτων σε πολλαπλές κλάσεις έχει ως στόχο την εκχώρηση ετικετών (labels) σε περιπτώσεις (instances), όπου οι ετικέτες έχουν προέλθει από ένα πεπερασμένο σύνολο στοιχείων. Η κυρίαρχη προσέγγιση είναι ο μετασχηματισμός του εννιαίου προβλήματος πολλαπλών κλάσεων σε πολλαπλά δυαδικά προβλήματα ταξινόμησης [38].

Ακολουθούν τρεις από τις πιο συνηθισμένες μεθόδους για την αντιμετώπιση του παραπάνω προβλήματος:

- Δημιουργία δυαδικών ταξινομητών που διακρίνονται μεταξύ τους σε δύο κατηγορίες: 1) κάθε ετικέτα συγκρίνεται με τις άλλες (one-versus-one) ή 2) ζεύγη ετικετών συγκρίνονται μεταξύ τους (one-versus-one). Η ταξινόμηση στην πρώτη κατηγορία στηρίζεται στη λογική winer-takes-all, στην οποία ο ταξινομητής με την υψηλότερη συνάρτηση εξόδου καταχωρεί την κατηγορία, ενώ στη δεύτερη η ταξινόμηση στηρίζεται στη στρατηγική max-win voting, όπου η κατηγορία με τους περισσότερους «ψήφους» καταχωρεί την τιμή της σε κάθε περίπτωση [39].

- Directed acyclic graph [40]
- Error-correcting outpout coded [41]

Ακολουθεί πίνακας που παρουσιάζει τα κύρια πλεονεκτήματα και μειονεκτήματα των μηχανών υποστήριξης διανυσμάτων:

Πλεονεκτήματα	Μειονεκτήματα
Αποτελεσματικές σε πολυδιάστατους χώρους	Εάν ο αριθμός των χαρακτηριστικών είναι πολύ μεγαλύτερος από τον αριθμό των δειγμάτων δεν δίνουν καλά αποτελέσματα
Δεν απαιτούν μεγάλη υπολογιστική ισχύ	
Ευέλικτες ως προς τον προσδιορισμό των πυρήνων	Δεν παρέχουν άμεσα εκτιμήσεις πιθανοτήτων, απαιτούν αντεπικύρωση (cross-validation)
Αποτελεσματικές και με σχετικά μικρό αριθμό δειγμάτων	

Πίνακας 2-2: Πλεονεκτήματα-Μειονεκτήματα SVM

2.2.2 Decision tree

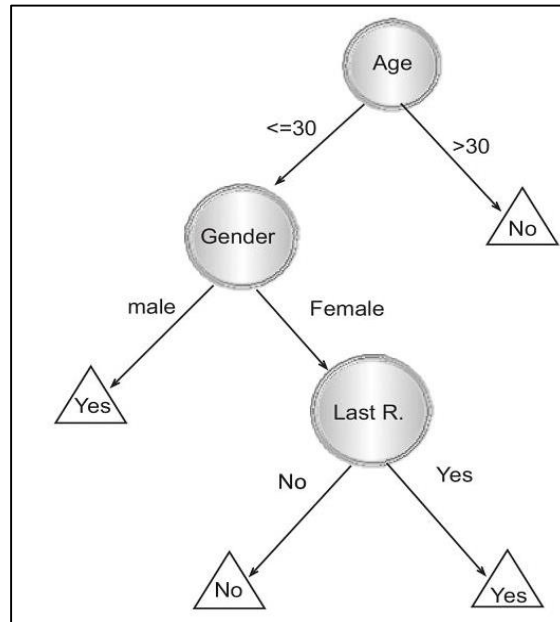
Στις τεχνικές εξόρυξης δεδομένων (data mining) το δέντρο αποφάσεων (decision tree) είναι ένα μοντέλο πρόβλεψης τόσο για ταξινόμηση όσο και για κατάτμηση. Πρόκειται για ένα ιεραρχικό μοντέλο αποφάσεων και συνεπειών που καλείται να βρει τη βέλτιστη στρατηγική για την επίτευξη του στόχου. Όταν ένα δέντρο αποφάσεων χρησιμοποιείται για διαδικασίες ταξινόμησης καλείται δέντρο ταξινόμησης (classification tree), ενώ όταν χρησιμοποιείται για διαδικασίες κατάτμησης καλείται δέντρο κατάτμησης (regression tree). Στη συνέχεια του κεφαλαίου, παρουσιάζονται οι ιδιότητες και τα χαρακτηριστικά των δέντρων ταξινόμησης, τα οποία χρησιμοποιούνται σε σημαντικό βαθμό σε τομείς που ασχολούνται με τα οικονομικά, το marketing, τις επιστήμες και την ιατρική [42].

Το δέντρο απόφασης είναι ένας ταξινομητής που εκφράζεται ως μια αναδρομική διχοτόμηση του χώρου. Αποτελείται από κόμβους (nodes) που σχηματίζουν ένα ριζωμένο δέντρο (rooted tree), δηλαδή ένα δέντρο με κατεύθυνση και με έναν κόμβο που δεν έχει εισερχόμενες ακμές. Ο κόμβος με εξερχόμενες ακμές ονομάζεται «εσωτερικός» ή «δοκιμαστικός» κόμβος, ενώ όλοι οι άλλοι κόμβοι ονομάζονται «φύλλα» (leaves). Κάθε φύλλο αντιπροσωπεύεται είτε από μια κλάση με το πιο αντιπροσωπευτικό χαρακτηριστικό, είτε από ένα διάνυσμα πιθανότητας που μεταφράζεται ως η πιθανότητα να έχει το φύλλο μια συγκεκριμένη κλάση. Σε κάθε εσωτερικό κόμβο υπάρχει η δυνατότητα να δημιουργηθούν περισσότερα από δύο κλαδιά (branches). Κάθε κομβός αντιπροσωπεύεται από ένα αποκλειστικό χαρακτηριστικό, ενώ κάθε κλαδί αντιπροσωπεύεται από ένα εύρος τιμών [42].

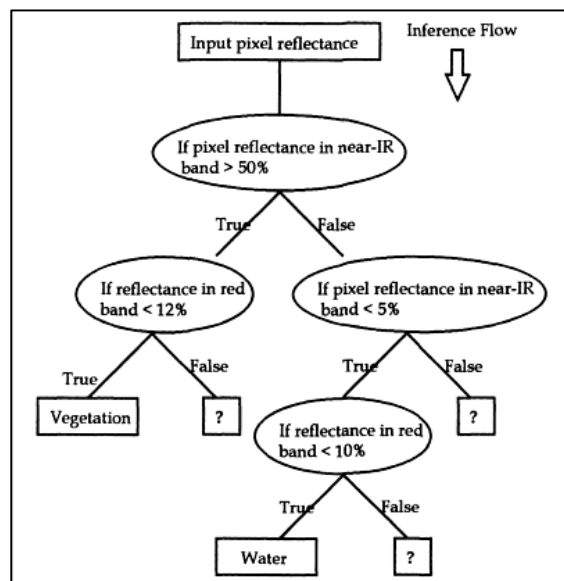
Κάθε αντικείμενο ταξινομείται ξεκινώντας από την ρίζα του δέντρου μέχρι κάποιο φύλλο του, ανάλογα με τα αποτελέσματα των ελέγχων στη διαδρομή που ακολουθεί και σύμφωνα με τα χαρακτηριστικά του. Πιο συγκεκριμένα, η διαδικασία ξεκινά από την ρίζα του δέντρου, ελέγχεται στη συνέχεια ποιο κλαδί είναι πιο αντιπροσωπευτικό ως προς τα χαρακτηριστικά του αντικειμένου και τέλος ελέγχεται ο κόμβος στον οποίο ανήκει το συγκεκριμένο κλαδί. Η διαδικασία επαναλαμβάνεται μέχρι να υπάρξει αντιστοίχιση – ταξινόμηση του αντικειμένου σε κάποιο φύλλο. [Rokach, 2008] Στις περισσότερες περιπτώσεις επιλέγεται ένα όχι

ιδιαίτερα πολύπλοκο δέντρο ώστε να είναι κατανοητό από τους μελετητές, ενώ λαμβάνεται υπόψη το γεγονός ότι η πολυπλοκότητα στα δέντρα απόφασης έχει επίδραση στην ακρίβεια της ταξινόμησης.

Ακολουθούν δύο διαγραμματικά παραδείγματα εφαρμογής του δέντρου απόφασης:



Εικόνα 2-3: Δέντρο απόφασης για την εύρεση συγκεκριμένων επιθέτων [42]



Εικόνα 2-4: Δέντρο απόφασης για την ταξινόμηση εικονοστοιχείων [43]

Η φάση εξάπλωσης του δέντρου συνεχίζεται μέχρις ότου καλυφθούν κάποια κριτήρια τερματισμού. Οι συνθηκές που οδηγούν σε τερματισμό είναι οι εξής:

- Εάν όλα τα στοιχεία των δεδομένων εκπαίδευσης έχουν μια μοναδική τιμή γ
- Εάν έχει εκπληρωθεί η συνθήκη του μέγιστου βάθους του δέντρου
- Εάν ο αριθμός των περιπτώσεων στον τερματικό κόμβο είναι μικρότερος από τον ελάχιστο αριθμό περιπτώσεων που έχει τεθεί για τους κόμβους

- Εάν ο κόμβος έχει διαχωριστεί και ο αριθμός των περιπτώσεων εμφάνισης είναι μικρότερος από αυτόν που έχει οριστεί
- Εάν ο αριθμός βέλτιστων διαχωρισμών είναι μεγαλύτερος από το συγκεκριμένο κατώφλι που έχει τεθεί

Ακολουθεί παράδειγμα ψευδοκώδικα του αλγόριθμου για την επέκταση του δέντρου απόφασης:

```

TreeGrowing (S, A, y, SplitCriterion, StoppingCriterion)
Where:
S - Training Set
A - Input Feature Set
y - Target Feature
SplitCriterion - the method for evaluating a certain split
StoppingCriterion - the criteria to stop the growing process
Create a new tree T with a single root node.
IF StoppingCriterion(S) THEN
    Mark T as a leaf with the most
    common value of y in S as a label.
ELSE
    ∀ai ∈ A find a that obtain the best SplitCriterion(ai, S).
    Label t with a
    FOR each outcome vi of a:
        Set Subtreei = TreeGrowing (σa=vi, S, A, y).
        Connect the root node of tT to Subtreei with
        an edge that is labelled as vi
    END FOR
END IF
RETURN TreePruning (S, T, y)
TreePruning (S, T, y)
Where:
S - Training Set
y - Target Feature
T - The tree to be pruned
DO
    Select a node t in T such that pruning it
    maximally improve some evaluation criteria
    IF t ≠ ∅ THEN T = pruned(T, t)
UNTIL t = ∅
RETURN T
    
```

Εικόνα 2-5: Ψευδοκώδικας αλγορίθμου τερματισμού επέκτασης δέντρου απόφασης [42]

Ακολουθεί πίνακας που παρουσιάζει τα κύρια πλεονεκτήματα και μειονεκτήματα των δέντρων απόφασης:

Πλεονεκτήματα	Μειονεκτήματα
Εύκολα στην κατανόηση και δυνατότητα αναπαράστασής τους	Προβλήματα ακρίβειας σε σύνθετες μορφές δέντρων
Δεν απαιτούν μεγάλη προεργασία δεδομένων	Απαιτείται εμπειρία από το μελετητή για να δημιουργηθούν σωστές συνθήκες
Δυνατότητα ενσωμάτωσης ποσοτικών και ποιοτικών κριτηρίων	Τα αποτελέσματα επηρεάζονται σε σημαντικό βαθμό από τις συνθήκες τερματισμού
Δυνατότητα αξιολόγησής τους μέσα από στατιστικούς ελέγχους	Απαιτούν αρκετά μεγάλο όγκο δεδομένων εκπαίδευσης

Δεν απαιτούν μεγάλη υπολογιστική ισχύ

Πίνακας 2-3: Πλεονεκτήματα-Μειονεκτήματα *decision tree*

2.2.3 K-nearest neighbors

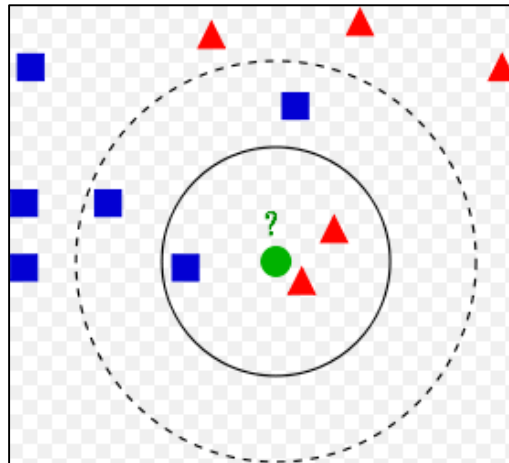
Στην αναγνώριση προτύπων, ο αλγόριθμος k-πλησιέστερων γειτόνων (k-nearest neighbors ή KNN) είναι μία μη παραμετρική μέθοδος που χρησιμοποιείται για ταξινόμηση και κατάτμηση. Και στις δύο περιπτώσεις ο αλγόριθμος στηρίζεται στα k-κοντινά δεδομένα εκπαίδευσης στο χώρο των αντικειμένων. Όσον αφορά την εφαρμογή του αλγορίθμου σε διαδικασίες ταξινόμησης, η έξοδος είναι η κατηγοριοποίηση των δεδομένων σε κλάσεις με βάση την πλειοψηφία των γειτόνων (majority voting). Ο αριθμός των γειτόνων k είναι ένας θετικός ακέραιος αριθμός, όπου αν $k=1$ τότε το αντικείμενο απλά αποδίδεται στην κατηγορία που ανήκει ο πλησιέστερος γείτονας. Ο αλγόριθμος KNN ανήκει στην κατηγορία των instance-based learning algorithms ή αλλιώς lazy learning algorithms καθώς η συνάρτηση απόδοσης τιμής και οι υπολογισμοί είναι κατά προσέγγιση σε τοπικό επίπεδο. Σε σύγκριση με άλλους αλγορίθμους υπολογιστικής μάθησης ο KNN είναι ο πιο απλός [44].

Επιπλέον, ο αλγόριθμος KNN δεν χρησιμοποιεί τα δεδομένα εκπαίδευσης για να κάνει οποιαδήποτε γενίκευση, οπότε δεν υπάρχει ρητή φάση εκπαίδευσης ή αυτή είναι ελάχιστη. Με τη μη γενίκευση των δεδομένων εκπαίδευσης ο αλγόριθμος κρατά όλα τα δεδομένα για εκπαίδευση σε όλες τις φάσεις δοκιμών σε αντίθεση με άλλους αλγορίθμους εκπαίδευσης, όπως είναι ο SVM, όπου υπάρχει η δυνατότητα απόρριψης δεδομένων. Κατά συνέπεια, ο αλγόριθμος KNN είναι πιο δαπανηρός από άποψη υπολογιστικής ισχύος καθώς αποθηκεύει και χρησιμοποιεί το σύνολο των δεδομένων για τη λήψη απόφασης [45].

Όπως αναφέρθηκε, ο αλγόριθμος KNN είναι μη παραμετρικός και το μόνο που απαιτείται είναι η επιλογή του αριθμού των γειτόνων k. Η καλύτερη επιλογή του k εξαρτάται από τα δεδομένα και σε γενικές γραμμές έχει παρατηρηθεί πως όσο μεγαλύτερη είναι η τιμή του k τόσο περισσότερο μειώνεται ο θόρυβος στην ταξινόμηση κάνοντας παράλληλα τα όρια των τάξεων λιγότερο διακριτά [46].

Για τη βέλτιστη επιλογή του αριθμού k πραγματοποιούνται διάφοροι αλγόριθμοι βελτιστοποίησης (π.χ. υπερπαραμετρική βελτιστοποίηση) για κάθε διαφορετική εφαρμογή. Στην ειδική περίπτωση που ο αριθμός k επιλέγεται να ισούται με τη μονάδα τότε ο αλγόριθμος ονομάζεται πλησιέστερου γείτονα.

Ακολουθεί ένα παράδειγμα του αλγορίθμου KNN:



Εικόνα 2-6: Παράδειγμα εφαρμογής αλγορίθμου KNN

Στο παραπάνω παράδειγμα γίνεται αντιληπτό πως ο πράσινος κύκλος πρέπει να ταξινομηθεί είτε στην κατηγορία μπλε τετράγωνο είτε στην κατηγορία κόκκινο τρίγωνο. Αν επιλεγεί ο αριθμός γειτόνων να είναι τρεις, $k=3$ (με τη συνεχόμενη γραμμή), τότε θα επιλεγεί η κατηγορία κόκκινο τρίγωνο αφού μέσα στον κύκλο υπάρχουν 2 κόκκινα τρίγωνα και 1 μπλε τετράγωνο. Αντίθετα, αν επιλεγεί ο αριθμός γειτόνων να είναι πέντε, $k=5$ (με τη διακεκομμένη γραμμή) τότε θα επιλεγεί η κατηγορία μπλε τετράγωνο αφού στο όριο υπάρχουν 3 μπλε τετράγωνα και 2 κόκκινα τρίγωνα.

Ακολουθεί ο ψευδοκώδικας του αλγορίθμου KNN:

```

TRAIN-KNN(C, D)
1  D' ← PREPROCESS(D)
2  k ← SELECT-K(C, D')
3  return D', k

APPLY-KNN(C, D', k, d)
1  Sk ← COMPUTE_NEAREST_NEIGHBORS(D', k, d)
2  for each cj ∈ C
3  do pj ← |Sk ∩ cj| / k
4  return arg maxj pj
    
```

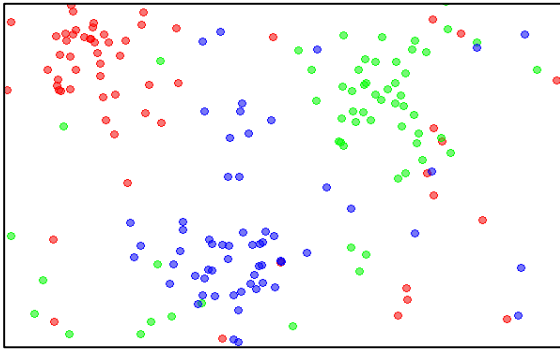
Εικόνα 2-7: Ψευδοκώδικας αλγορίθμου KNN [47]

Όσον αφορά στον αλγόριθμο KNN, έχει προταθεί μία βελτίωσή του με ονομασία CNN (condensed nearest neighbor), η οποία σχεδιάστηκε για να μειώσει το σύνολο των δεδομένων για τον αλγόριθμο ταξινόμησης KNN. Ο αλγόριθμος επιλέγει το σύνολο των προτύπων U από τα δεδομένα εκπαίδευσης, έτσι ώστε ο αλγόριθμος πλησιέστερου γείτονα με $k=1$ να μπορεί να ταξινομήσει το σύνολο των δεδομένων με την ακρίβεια που θα έκανε ο 1NN σε μικρότερα παραδείγματα. Πιο συγκεκριμένα, ακολουθείται η εξής διαδικασία:

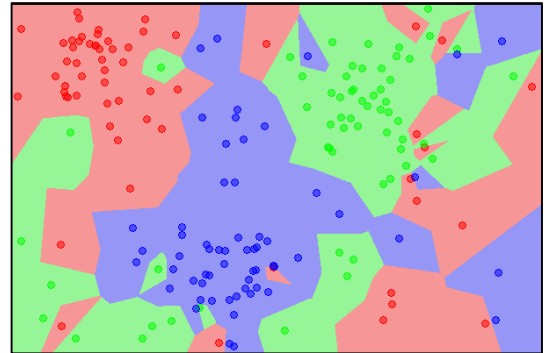
Έστω τα δεδομένα εκπαίδευσης X :

- Σάρωση όλων των στοιχείων X , ψάχνοντας το στοιχείο x του οποίου το κοντινότερο πρότυπο από το U έχει διαφορετική ετικέτα από το x
- Αφαίρεση του x από το X και ένταξη του στο U
- Επανάληψη της διαδικασίας μέχρις ότου κανένα πρότυπο να μην εντάσσεται στο U
- Χρήση του U αντί του X για ταξινόμηση

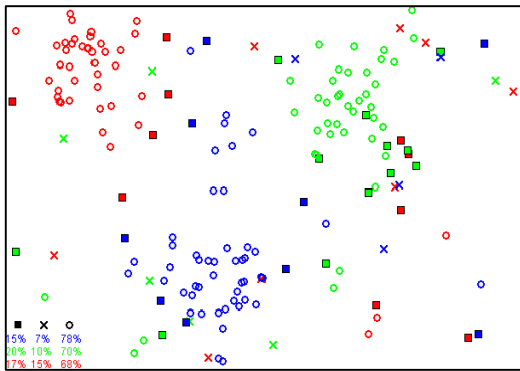
Ακολουθεί παράδειγμα που παρουσιάζει τα δεδομένα εκπαίδευσης και το αποτέλεσμα της ταξινόμησης χωρίς τη χρήση του αλγορίθμου CNN καθώς και αυτά με τη χρήση του αλγορίθμου CNN.



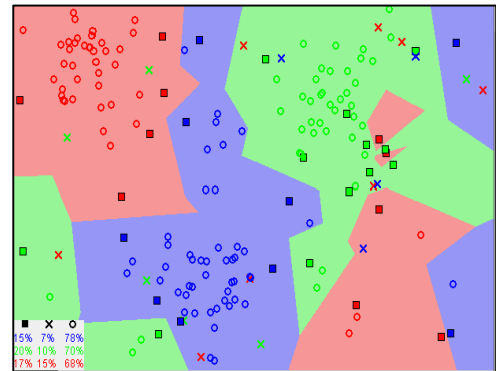
Εικόνα 2-8: Δεδομένα εκπαίδευσης χωρίς την εφαρμογή του CNN



Εικόνα 2-9: 1NN ταξινόμηση χωρίς εφαρμογή του CNN



Εικόνα 2-10: Δεδομένα εκπαίδευσης με εφαρμογή του CNN



Εικόνα 2-11: 1NN ταξινόμηση με εφαρμογή του CNN

Ακολουθεί πίνακας που παρουσιάζει τα κύρια πλεονεκτήματα και μειονεκτήματα των κ-πλησιέστερων γειτόνων:

Πλεονεκτήματα	Μειονεκτήματα
Εύκολα στην κατανόηση και εκσφαλμάτωσή τους	Απαιτούν μεγάλη υπολογιστική ισχύ και είναι αρκετά αργά
Αντιμέτωπη θορύβου στις ταξινομήσεις	Ευαίσθητα στην επιλογή γειτόνων k
Μη παραμετρικά	Δεν αντιμετωπίζουν αποτελεσματικά σύνθετα πρότυπα
Ευκολία αντιμετώπισης προβλημάτων με πολλά δεδομένα εκπαίδευσης	

Πίνακας 2-4: Πλεονεκτήματα-Μειονεκτήματα των κ-πλησιέστερων γειτόνων

2.2.4 Αξιολόγηση στατιστικών μοντέλων

Στο συγκεκριμένο κεφάλαιο παρουσιάζονται κάποιοι δείκτες που χρησιμοποιούνται προκειμένου να αξιολογηθούν τα στατιστικά μοντέλα εκμάθησης που εφαρμόζονται στα πλαίσια των διαδικασιών εξόρυξης δεδομένων. Πιο συγκεκριμένα, παρουσιάζεται ο δείκτης της ακρίβειας (precision), ο αριθμός εύστοχων αποκρίσεων (recall), ο δείκτης f1 score, ο δείκτης accuracy και ο δείκτης kappa.

Πριν δοθούν οι ορισμοί των παραπάνω δεικτών είναι αναγκαίο να αναφερθούν τέσσερις έννοιες που είναι απαραίτητες για την πλήρη κατανόηση των δεικτών. Συγκεκριμένα, εξηγούνται οι έννοιες true positives, true negatives, false positives, false negatives.

Αρχικά, οι έννοιες positive (θετικό) και negative (αρνητικό) αποδίδουν την πρόβλεξη του ταξινομητή (ή προσδοκία) ενώ οι έννοιες αληθές (true) και ψευδές (false) αναφέρονται στο αν η παραπάνω πρόβλεψη ανταποκρίνεται σε εξωτερική κρίση (ή παρατήρηση). Συνδυάζοντας τις παραπάνω έννοιες προκύπτει πως με την έννοια true positives (TP) ορίζονται τα αντικείμενα που ταξινομήθηκαν σωστά στην τάξη, με την έννοια true negative (TN) ορίζονται τα αντικείμενα που δεν συμπεριλήφθηκαν σωστά στην τάξη, με την έννοια false positive (FP) ορίζονται τα αντικείμενα που ταξινομήθηκαν στην τάξη αλλά δεν έπρεπε, ενώ με την έννοια false negative (FN) ορίζονται τα αντικείμενα που δεν συμπεριλήφθηκαν στην τάξη ενώ έπρεπε [48].

Precision

Σε εργασίες ταξινόμησης ο δείκτης precision (ακρίβεια) για μια συγκεκριμένη τάξη είναι ο αριθμός των αληθώς θετικών (true positives) αντικειμένων, δηλαδή ο αριθμός των αντικειμένων που επισημάνθηκαν σωστά ότι ανήκουν σε μία τάξη, διαιρούμενος με το συνολικό αριθμό των στοιχείων που επισημάνθηκαν ότι ανήκουν στη θετική τάξη, δηλαδή το άθροισμα των αληθών και ψευδών θετικών αντικειμένων [48].

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

Ο αριθμός εύστοχων αποκρίσεων (recall) ορίζεται ως ο αριθμός των αληθώς θετικών αντικειμένων διαιρούμενος με το συνολικό αριθμό των στοιχείων που πραγματικά ανήκουν στη θετική τάξη, δηλαδή το άθροισμα των πραγματικά αληθώς και ψευδώς αρνητικών αντικειμένων που έπρεπε να ανήκουν στη συγκεκριμένη τάξη αλλά δεν επισημάνθηκαν [48].

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score

Πρόκειται για ένα μέτρο που συνδυάζει την ακρίβεια με την ευστοχία μέσω ενός αρμονικού μέσου σταθμίζοντας τις δύο έννοιες ομοιόμορφα.

Πιο συγκεκριμένα:

$$F - \text{score} = 2 * \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Accuracy

Η έννοια της ακρίβειας (accuracy) εκφράζει το πηλίκο του αθροίσματος των αληθών αντικειμένων (θετικών και αρνητικών) με το σύνολο όλων των θετικών και αρνητικών αντικειμένων. Πιο συγκεκριμένα:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Καρρα

Η στατιστική τιμή καρρα είναι ένα μέτρο διαφοράς μεταξύ της πραγματικής συμφωνίας μεταξύ των επίγειων δεδομένων αναφοράς και ενός αυτόματου αλγορίθμου ταξινόμησης, και της τυχαίας συμφωνίας μεταξύ των επίγειων δεδομένων αναφοράς και ενός τυχαίου αλγορίθμου ταξινόμησης. Είναι μια ένδειξη του βαθμού, κατά τον οποίο οι ποσοστιαίες ορθές τιμές ενός πίνακα σφαλμάτων οφείλονται σε αληθή έναντι τυχαίας συμφωνίας.

$$k = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

3. ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΔΙΑΧΡΟΝΙΚΩΝ ΧΑΡΤΩΝ ΚΑΛΥΨΗΣ ΓΗΣ

Στόχος της συγκεκριμένης διπλωματικής εργασίας ήταν η ανάπτυξη μεθοδολογικής προσέγγισης για την αυτόματη παραγωγή διαχρονικών χαρτών κάλυψης και χρήσεων γης αξιοποιώντας πολυφασματικά δορυφορικά δεδομένα. Για το σκοπό αυτό αξιοποιήθηκαν διαχρονικά δεδομένα του δορυφόρου Landsat 8 από διάφορες περιοχές της Ελλάδας. Λαμβάνοντας υπόψη τη σύγχρονη διεθνή βιβλιογραφία [18] [49] [50] [51] [52] καθώς και ύστερα από δοκιμές διάφορων μεθοδολογικών προσεγγίσεων και αλγορίθμων αναπτύχθηκε μεθοδολογία μορφής δέντρου απόφασης, η οποία αξιοποιεί πολλαπλούς ταξινομητές. Η συγκεκριμένη μεθοδολογία βασίζεται στην λογική του “per class classification” και αξιοποιεί παράλληλα φαινολογικά χαρακτηριστικά, τα οποία εξάγονται ύστερα από κατάλληλη ανάλυση χρονοσειρών.

3.1 Περιοχές Μελέτης και Επιλογή Δεδομένων

Για την εφαρμογή και την αξιολόγηση της μεθοδολογικής προσέγγισης επιλέχθηκαν πέντε διαφορετικές περιοχές της Ελλάδας σε τέσσερις διαφορετικές χρονικές στιγμές. Ιδιαίτερη προσοχή δόθηκε στο να επιλεγούν δεδομένα που δεν καλύπτονταν σε μεγάλο ποσοστό από σύννεφα έτσι ώστε να είναι δυνατή η μετέπειτα επεξεργασία τους. Για τον σκοπό της διπλωματικής αξιοποιήθηκαν τα ατμοσφαιρικά διορθωμένα δεδομένα του δορυφόρου Landsat 8 από τον επίσημο ιστοχώρο της USGS.

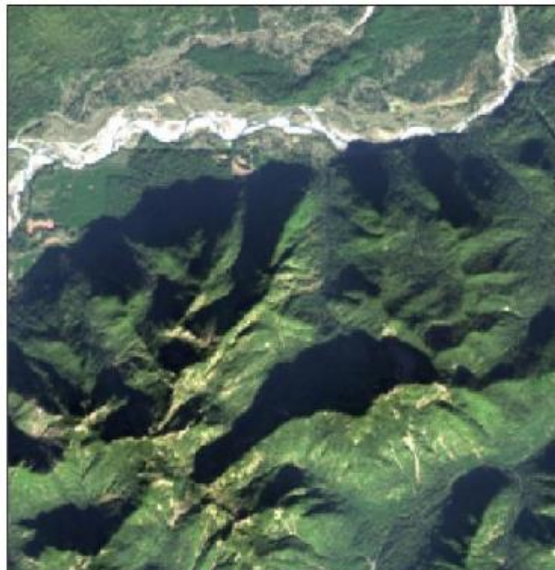
Τα ατμοσφαιρικά διορθωμένα δεδομένα του Landsat 8 είναι διαθέσιμα από τις αρχές του 2015 και παράγονται από ένα εξειδικευμένο λογισμικό της NASA, με ονομασία L8SR. Ο συγκεκριμένος αλγόριθμος είναι διαφορετικός από τον αντίστοιχο που χρησιμοποιείται για την επεξεργασία των δεδομένων Landsat 4-5 και Landsat 7 και βρίσκεται ακόμα σε στάδιο εξέλιξης, ανανεώνοντας σε τακτά διαστήματα τα υφιστάμενα επεξεργασμένα δεδομένα. Ακολουθεί πίνακας που παρουσιάζει τα χαρακτηριστικά των δεδομένων που προσφέρονται μετά την επεξεργασία από τον L8SR.

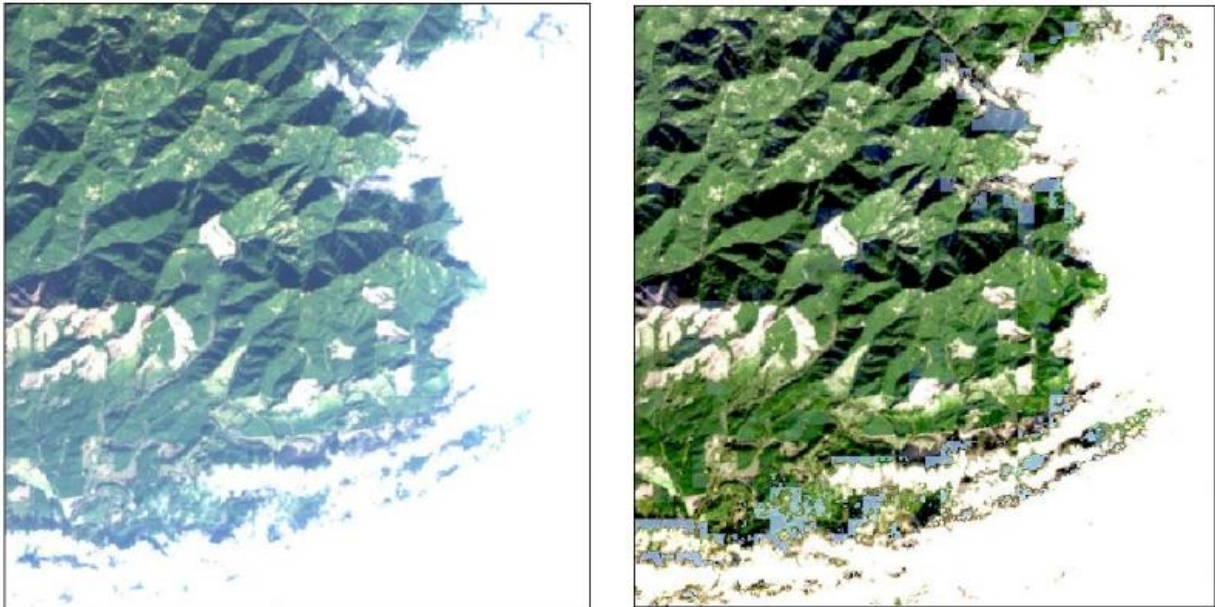
Παράμετρος	Landsat 8 OLI (L8SR)
Υπεύθυνο ερευνητικό κέντρο	NASA GSFC
Παγκόσμια κάλυψη	Ναι
TOA	Οπτικά (1-7,9) + Θερμικά (10-11)
SR	Οπτικά (1-7)
Μοντέλο μεταφοράς ακτινοβολίας	«Εσωτερικός» - πειραματικός αλγόριθμος
Διόρθωση θερμικών καναλιών	Μόνο TOA
Μονάδες θερμικών καναλιών	Kelvin
Πίεση	Η πίεση εδάφους υπολογίζεται με βάση το υψόμετρο
Υδρατμοί	MODIS CMA
Θερμοκρασία αέρα	MODIS CMA
DEM	Παγκόσμιο ψηφιακό μοντέλο εδάφους
Ozone	MODIS CMG
AOT	MODIS CMA

Γωνία ήλιου	Κέντρο σκηνής από τα μεταδεδομένα
Ζενίθια γωνία	Μετασχηματισμό (hard-coded) σε 0
Διόρθωση ανεπιθύμητων ζενίθιων γωνιών	Τα SR δεδομένα δεν επεξεργάζονται όταν η ζενίθια γωνία είναι μεγαλύτερη των 76 μοιρών
Επεξεργασία παγχρωματικού καναλιού	Όχι
Επεξεργασία μεταδεδομένων	Ναι
Μάσκα σύννεφων	CFmask
Μορφότυπος δεδομένων	INT16
Γέμισμα τιμών	-9999
Κανάλι QA	Σύννεφα Γειτονικά σύννεφα Σκιά σύννεφων Αεροζόλ Νεφέλη (cirrus)

Πίνακας 3-1: Χαρακτηριστικά των ατμοσφαιρικά διορθωμένων προϊόντων του Landsat 8 μέσω του αλγορίθμου L8SR

Πρέπει να τονιστεί πως τα ατμοσφαιρικά δεδομένα που παράγονται, μέχρι αυτή τη στιγμή, δεν βρίσκονται σε επιχειρησιακό στάδιο και πολύ συχνά, κυρίως σε περιοχές με έντονο ανάγλυφο ή έντονη συννεφοκάλυψη/χιονοκάλυψη, συναντώνται artifacts που επηρεάζουν την ποιότητα των δεδομένων. Ωστόσο, λαμβάνοντας υπόψη την αναγκαιότητα των ατμοσφαιρικά διορθωμένων δεδομένων σε εφαρμογές παραγωγής χαρτών κάλυψης γης και σε αλγορίθμους ταξινόμησης επιλέχθηκαν με προσοχή τα δεδομένα έτσι ώστε να μην έχουν artifacts σε μεγάλη έκταση της εικόνας. Ακολουθούν παραδείγματα από “artifacts” σε περιοχές με έντονη τοπογραφική ανομοιομορφία.






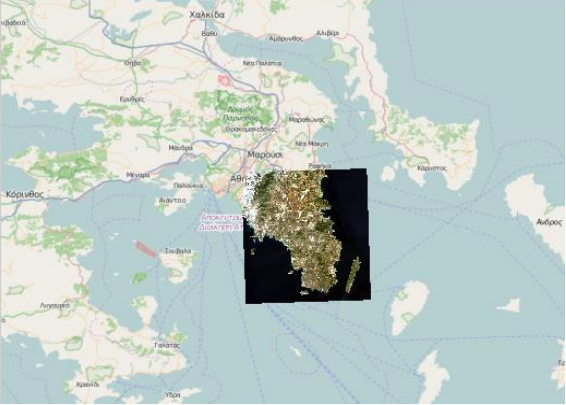

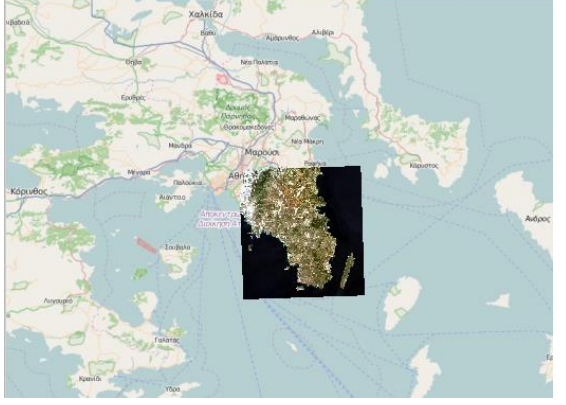

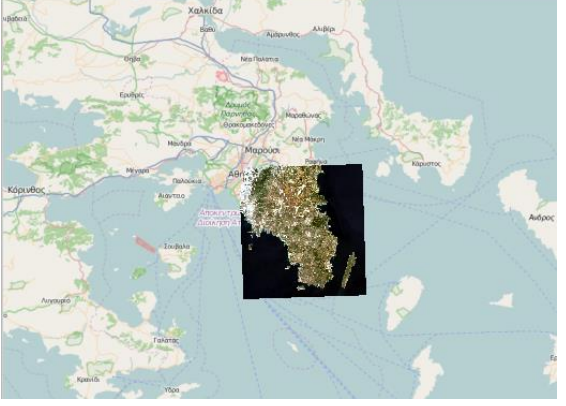
Εικόνα 3-1: Παραδείγματα από “artifacts” σε περιοχές με έντονη τοπογραφική ανομοιομορφία από τα ατμοσφαιρικά διορθωμένα προϊόντα του Landsat 8

Η επιλογή των δεδομένων πραγματοποιήθηκε λαμβάνοντας υπόψη τα παρακάτω κριτήρια:


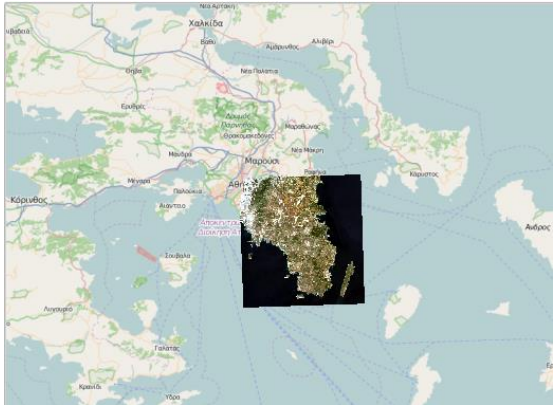

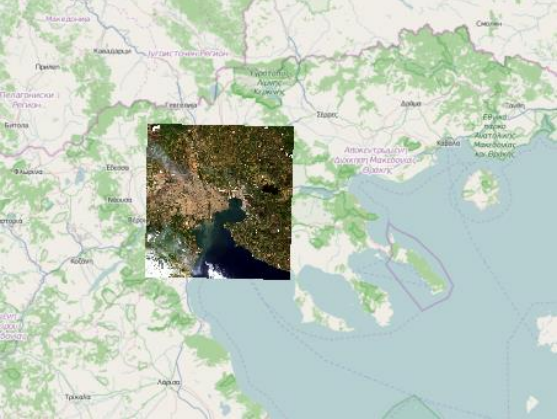

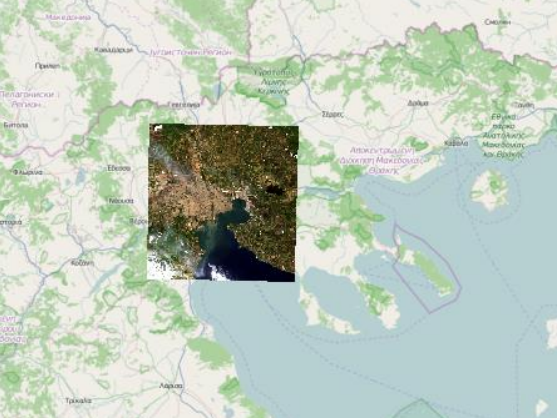
- Επιλογή δεδομένων από ημίξηρες περιοχές, όπου η εποχιακή συμπεριφορά της βλάστησης αντικατοπτρίζεται σε σημαντικό βαθμό, έτσι ώστε να είναι δυνατή η ανάλυση χρονοσειρών.
- Οι περιοχές δεν επηρεάζονταν από απότομες αλλαγές, σε σύντομο χρονικό διάστημα, στις χρήσεις γης, όπως είναι για παράδειγμα οι ανθρωπογενείς παρεμβάσεις.
- Επιλογή δεδομένων με όσο το δυνατό μικρότερη κάλυψη από σύννεφα (τέθηκε ως ανώτερο όριο το 10%).
- Οι περιοχές μελέτης αποτελούσαν αντιπροσωπευτικές πιλοτικές περιοχές για το σύνολο της χώρας.

Παρακάτω ακολουθεί πίνακας που παρουσιάζει τα δεδομένα που επιλέχθηκαν για τις ανάγκες της εργασίας σε διάφορες χρονικές στιγμές:


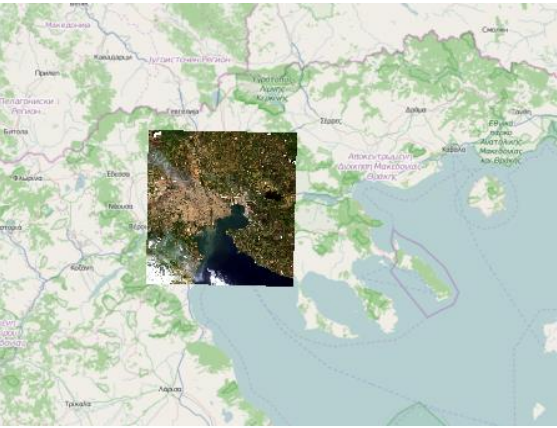

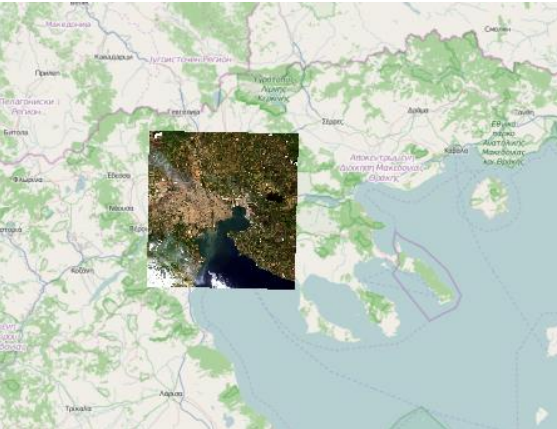
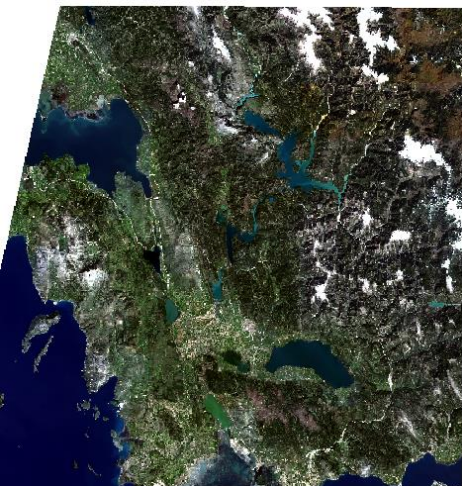

Κεφάλαιο 3. Ανάπτυξη Μεθοδολογίας για την Παραγωγή Διαχρονικών Χαρτών Κάλυψης Γης



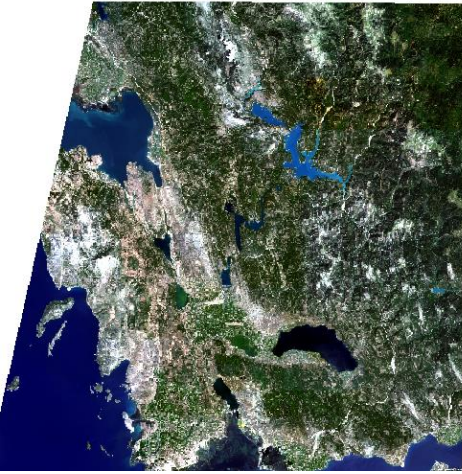



Περιοχή	Ημ/νία	Απόσπασμα Οθόνης	Path/ Row	Τοποθεσία
1	13/4/2015		182/34	
1	31/5/2015		182/34	
1	18/7/2015		182/34	

Κεφάλαιο 3. Ανάπτυξη Μεθοδολογίας για την Παραγωγή Διαχρονικών Χαρτών Κάλυψης Γης

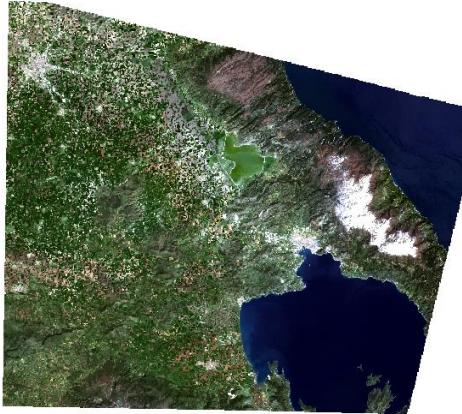
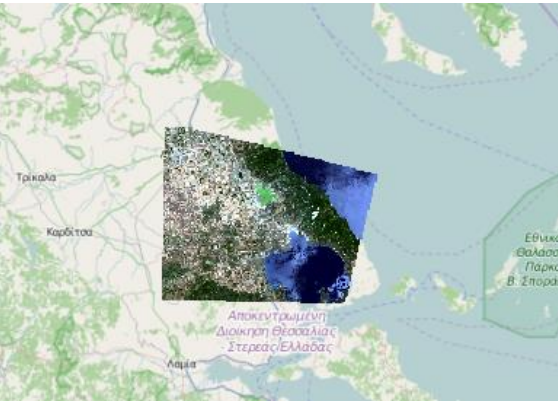

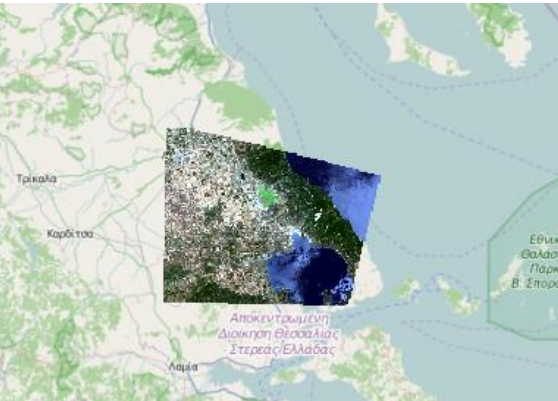
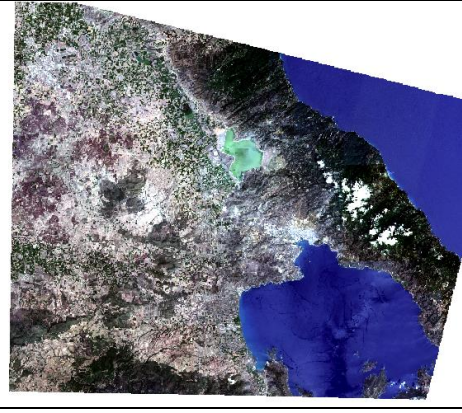
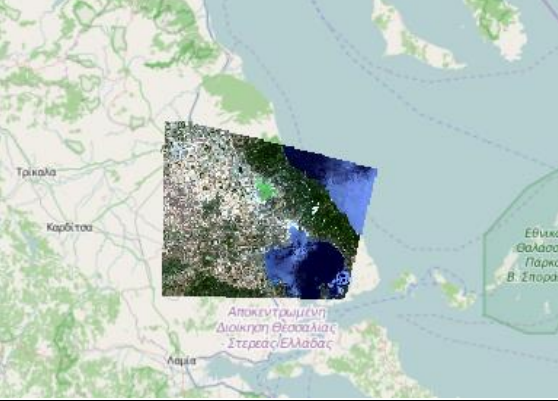
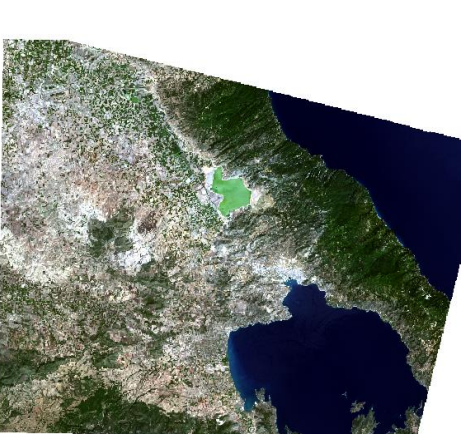
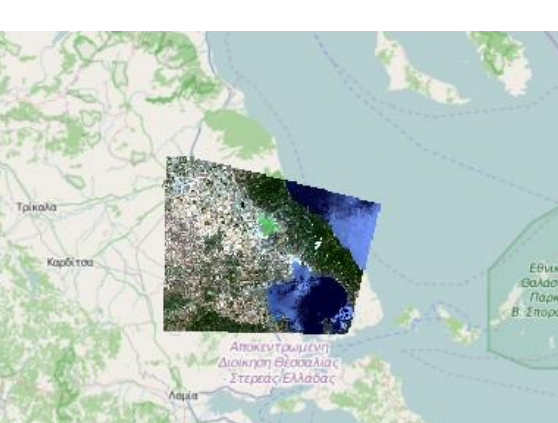
<p>1</p> <p>4/9/2015</p>		<p>182/34</p>	
<p>2</p> <p>5/1/2015</p>		<p>184/32</p>	
<p>2</p> <p>27/4/2015</p>		<p>184/32</p>	





Κεφάλαιο 3. Ανάπτυξη Μεθοδολογίας για την Παραγωγή Διαχρονικών Χαρτών Κάλυψης Γης

2	1/8/2015		184/32	
2	2/9/2015		184/32	
3	11/4/2015		184/33	

3	14/6/2015		184/33	
3	1/8/2015		184/33	
3	2/9/2015		184/33	

Κεφάλαιο 3. Ανάπτυξη Μεθοδολογίας για την Παραγωγή Διαχρονικών Χαρτών Κάλυψης Γης

4	11/4/2015		183/33	
4	14/6/2015		183/33	
4	1/8/2015		183/33	
4	2/9/2015		183/33	

5	13/6/2014		182/35	
5	18/7/2015		182/35	

Πίνακας 3-2: Επιλογή διαχρονικών περιοχών μελέτης από διαφορετικές γεωγραφικές ενότητες στην Ελλάδα

3.2 Μεθοδολογική Προσέγγιση

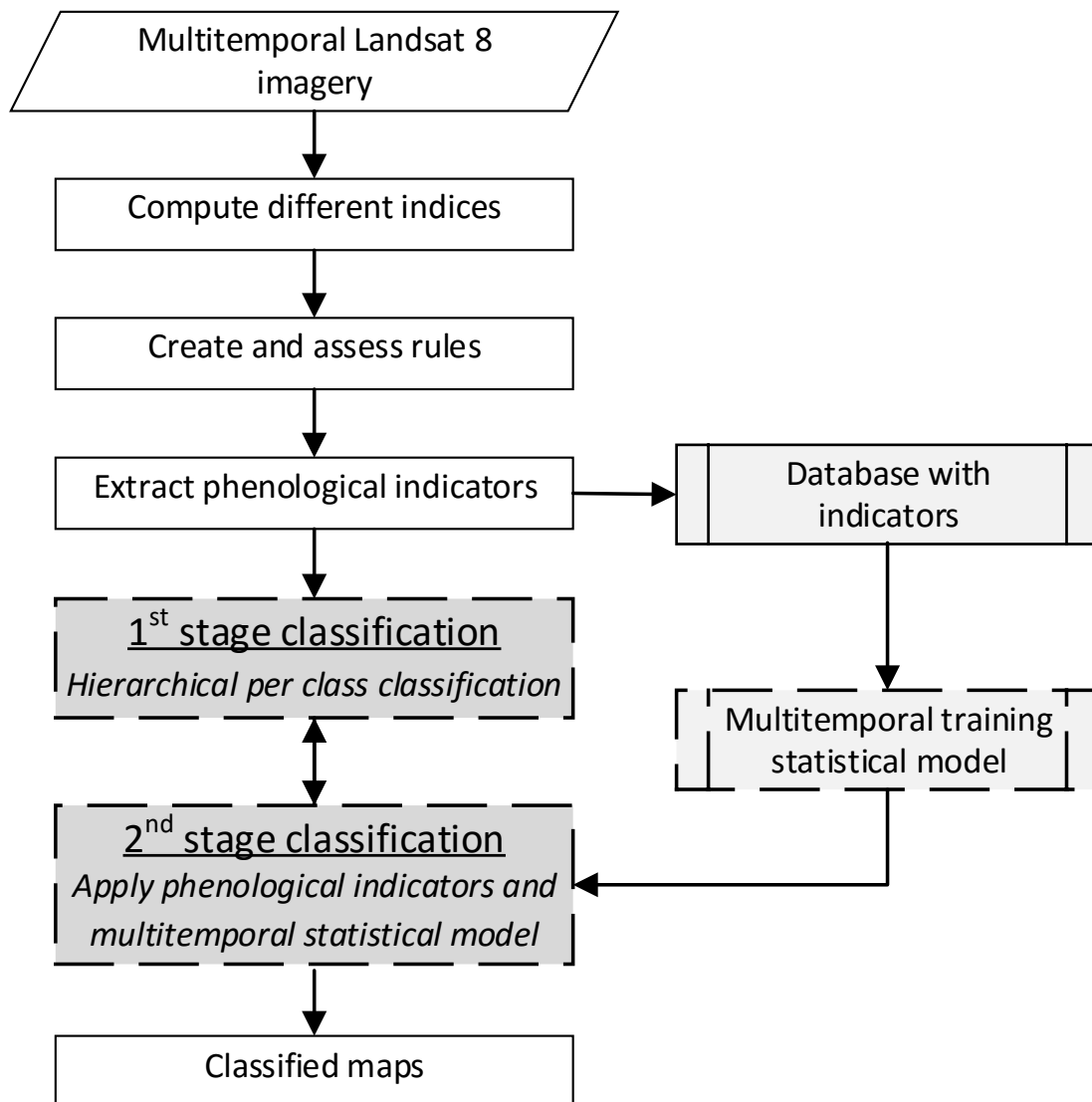
Όπως αναφέρθηκε νωρίτερα η μεθοδολογική προσέγγιση που αναπτύχθηκε ακολουθεί μορφή ιεραρχικού δέντρου απόφασης, η οποία αξιοποιεί πολλαπλούς ταξινομητές ανάλογα με τη θεματική κατηγορία προς ταξινόμηση. Η συγκεκριμένη μεθοδολογία βασίζεται στη λογική του “per class classification”, δηλαδή κάθε κατηγορία ταξινομείται χωριστά και στη συνέχεια «μασκάρεται» έτσι ώστε ο αλγόριθμος ταξινόμησης να εστιάζει κάθε φορά στην ενδιαφερόμενη κατηγορία, αξιοποιώντας παράλληλα φαινολογικά χαρακτηριστικά, τα οποία εξάγονται ύστερα από κατάλληλη ανάλυση χρονοσειρών για την αύξηση της ακρίβειας στις θεματικές κατηγορίες της βλάστησης και των καλλιεργειών. Για το σύνολο των διαδικασιών δημιουργήθηκε πρόγραμμα γραμμένο σε γλώσσα προγραμματισμού Python αξιοποιώντας ελεύθερες βιβλιοθήκες όπως είναι η GDAL, NumPy και Orfeo Toolbox.

Λαμβάνοντας υπόψη πως οι δορυφόροι Landsat προσφέρουν δεδομένα σε μεσαία χωρική ανάλυση (30 m), πολλές περιοχές ανήκουν σε διαφορετικούς τύπους κάλυψης γης δημιουργώντας προβλήματα στην ταξινόμηση των εικόνων. Δεδομένου, λοιπόν, πως οι συγκεκριμένοι τύποι κάλυψης γης σχηματίζουν μία φυσική ιεραρχική δομή, γίνεται

αντιληπτό πως η ιεραρχική ταξινόμηση σε δομή δέντρου ανά κλάση αποτελεί κατάλληλη στρατηγική για το διαχωρισμό των κατηγοριών κάλυψης γης. Με τον ορισμό της ιεραρχικής δομής της κάλυψης γης, είναι δυνατό να δημιουργηθούν κατάλληλοι κανόνες γνώσης (knowledge rules) για κάθε κατηγορία. Πιο συγκεκριμένα οι παραπάνω κανόνες μπορούν να ομαδοποιηθούν στις παρακάτω τρεις κατηγορίες:

- **Τομεακή φασματική γνώση:** Η φασματική γνώση μπορεί να χρησιμοποιηθεί για την κατασκευή της ιεραρχικής δομής των κλάσεων, όπως είναι για παράδειγμα ο χωρισμός των περιοχών μεταξύ βλάστησης και μη βλάστησης, αξιοποιώντας κατάλληλα κατώφλια από κανονικοποιημένους δείκτες ανάλογα με την κατηγορία.
- **Κανόνες φασματικής ταξινόμησης:** Οι συγκεκριμένοι ποσοτικοί κανόνες ταξινόμησης δημιουργούνται από αντιπροσωπευτικά δείγματα εκπαίδευσης, οι οποίοι μπορούν στη συνέχεια να χρησιμοποιηθούν για την περαιτέρω διάκριση των κατηγοριών γης με μεγαλύτερη ακρίβεια.
- **Χωροχρονική γνώση:** Δεδομένου πως η φασματική γνώση από μόνη της δεν είναι επαρκής για την ταξινόμηση όλων των τύπων κάλυψης γης, είναι ιδιαίτερα σημαντικό να ληφθούν υπόψη χωρικά κριτήρια, όπως είναι για παράδειγμα οι σχέσεις γειτνίασης, καθώς και χρονικά κριτήρια, όπως είναι τα φαινολογικά χαρακτηριστικά για τους διάφορους τύπους βλάστησης και καλλιεργειών.

Η μεθοδολογική προσέγγιση που αναπτύχθηκε ταξινομεί την κάθε σκηνή σε 2 στάδια, ανάλογα με το επίπεδο λεπτομέρειας κάθε κατηγορίας. Πιο συγκεκριμένα, κατά το 1^ο στάδιο εφαρμόζεται μία ιεραρχική ταξινόμηση ανά κλάση, λαμβάνοντας υπόψη λογικούς και ποσοτικούς κανόνες, ενώ στο 2^ο στάδιο ενσωματώνονται διαχρονικά φαινολογικά χαρακτηριστικά σε συνδυασμό με το διαχρονικό στατιστικό μοντέλο εκπαίδευσης/πρόβλεψης που δημιουργήθηκε από δείγματα σε διάφορες περιοχές και χρονικές στιγμές. Ακολουθεί η διαγραμματική παρουσίαση του συνόλου της μεθοδολογικής προσέγγισης.



Εικόνα 3-2: Συνοπτική παρουσίαση της μεθοδολογικής προσέγγισης που αναπτύχθηκε

3.2.1 Ενσωμάτωση φαινολογικών χαρακτηριστικών

Η ενσωμάτωση χαρακτηριστικών που σχετίζονται με το χρόνο είναι ιδιαίτερα σημαντική για τη βελτίωση των αποτελεσμάτων της ταξινόμησης αξιοποιώντας δορυφορικά δεδομένα. Για το σκοπό αυτό, αξιοποιήθηκαν φαινολογικά χαρακτηριστικά (συμπεριλαμβανομένων της έναρξης και λήξης της καλλιεργήσιμης περιόδου, της διάρκειας της καλλιεργήσιμης περιόδου, το εποχιακό πλάτος καθώς και της μέγιστης τιμής του κανονικοποιημένου δείκτη NDVI), τα οποία προήλθαν από την ανάλυση χρονοσειρών NDVI του Landsat 8.

Τα φαινολογικά χαρακτηριστικά εξήχθησαν μέσω του εργαλείου TIMESAT και ενσωματώθηκαν στα φασματικά δεδομένα για τη βελτίωση της ακρίβειας της ταξινόμησης μέσω του ταξινομητή SVM. Μέσω του εργαλείου TIMESAT, οι χρονοσειρές του NDVI, για κάθε περιοχή μελέτης, προσαρμόστηκαν αρχικά μέσω μεθόδου ελαχίστων τετραγώνων και στη συνέχεια εξάχθηκαν τα φαινολογικά χαρακτηριστικά. Για τη βελτίωση της ταξινόμησης των χρήσεων γης και για την καλύτερη προσαρμογή του μοντέλου στα φαινολογικά χαρακτηριστικά της κάθε περιοχής χρησιμοποιήθηκαν δεδομένα NDVI με βήμα 3 μηνών (εφόσον αυτό ήταν δυνατό) για κάθε έτος αναφοράς.

Οι διαδικασίες που καλύφθηκαν μέσω του εργαλείου TIMESAT ήταν οι εξής:

1. Καθορισμός των ρυθμίσεων εισόδου για την επεξεργασία των διαχρονικών NDVI δεδομένων.
2. Εισαγωγή των δεδομένων σε μορφή ASCII και καθορισμός του χρονικού βήματος δειγματοληψίας για κάθε περιοχή ενδιαφέροντος.
3. Εισαγωγή δεικτών ποιότητας των δεδομένων, οι οποίοι προέκυψαν από το μασκάρισμα των σύννεφων και των περιοχών με χιόνι που προέκυψαν από τα αρχικά στάδια της μεθοδολογικής προσέγγισης.
4. Προεπεξεργασία των χρονοσειρών και προσαρμογή συνάρτησης ομαλοποίησης, ανάλογα με την περιοχή μελέτης.
5. Εφαρμογή της ομαλοποιημένης συνάρτησης για την εξαγωγή των παραμέτρων εποχικότητας.
6. Αποθήκευση και συσχετισμός των τιμών των παραμέτρων και της συνάρτησης στα πολυφασματικά δεδομένα, λαμβάνοντας υπόψη την περιοχή και τη χρονική περίοδο.

Συμπληρωματικά δεδομένα για τον υπολογισμό της αβεβαιότητας

Σε πρώτο στάδιο αξιοποιήθηκε η μάσκα σύννεφων που δημιουργήθηκε προκειμένου να παραχθούν οι δείκτες αβεβαιότητας των τιμών NDVI καθώς και τα αντίστοιχα βάρη. Για τον καλύτερο καθορισμό των βαρών λήφθηκαν υπόψη και τα κανάλια CFmask, CFmask Cloud Confidence και δημιουργήθηκαν τρεις κλάσεις με βάρη 0, 0.5 και 1 για τις τιμές που ήταν σύννεφα, mixed και καθαρές αντίστοιχα. Στη συνέχεια, τα παραπάνω βάρη μετατράπηκαν σε αβεβαιότητες στο λογισμικό TIMESAT μέσω της σχέσης:

$$\sigma = \frac{1}{w + 0.0001}$$

Προσαρμογή και ομαλοποίηση

Το λογισμικό TIMESAT υποστηρίζει τρεις διαφορετικές μεθόδους επεξεργασίας, που βασίζονται στη μέθοδο ελαχίστων τετραγώνων για την προσαρμογή στο upper envelope των NDVI δεδομένων. Η πρώτη μέθοδος χρησιμοποιεί τοπικές πολυωνυμικές συναρτήσεις και μπορεί να χαρακτηριστεί ως ένα προσαρμοστικό φίλτρο Savitzky–Golay. Οι άλλες δύο αποτελούν κλασικές μεθόδους ελαχίστων τετραγώνων, όπου τα δεδομένα προσαρμόζονται για τη μοντελοποίηση συναρτήσεων διαφορετικής πολυπλοκότητας. Στις περιοχές μελέτης που επιλέχθηκαν αξιοποιήθηκε το προσαρμοστικό φίλτρο Savitzky–Golay, καθώς παρατηρήθηκε πως για το συγκεκριμένο χρονικό εύρος έδωσε τα πιο ικανοποιητικά αποτελέσματα σε σχέση με τις άλλες δύο μεθόδους.

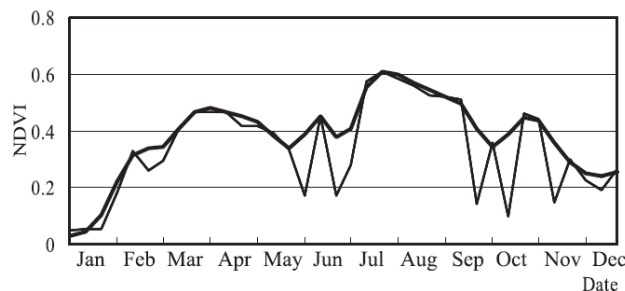
Οι Savitzky και Golay (1964) πρότειναν μία απλοποιημένη μέθοδο βασισμένη στη μέθοδο ελαχίστων τετραγώνων για την εξομάλυνση και τον υπολογισμό των παραγώγων από ένα σύνολο διαδοχικών τιμών, όπως είναι για παράδειγμα το φάσμα. Η συνέλιξη μπορεί να θεωρηθεί ως ένα σταθμισμένο κινούμενο μέσο φίλτρο, όπου η στάθμιση δίνεται ως πολυώνυμο ορισμένου βαθμού. Όταν οι συντελεστές βάρους εφαρμοστούν σε ένα σήμα, πραγματοποιείται προσαρμογή μέσω της μεθόδου ελαχίστων τετραγώνων εντός του παραθύρου του φίλτρου. Το συγκεκριμένο πολυώνυμο έχει σχεδιαστεί για να διατηρεί τις υψηλές τιμές και να μειώνει τα σφάλματα που εισάγονται από το φίλτρο. Η γενική εξίσωση

της απλοποιημένης μεθόδου ελαχίστων τετραγώνων για τη χρονοσειρά των NDVI μπορεί να δοθεί ως εξής:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+i}}{N}$$

όπου Y η αρχική NDVI τιμή, Y^* η τιμή NDVI που προκύπτει, C_i ο συντελεστής του φίλτρου (παράθυρο εξομάλυνσης) για την τιμή του $NDVI_i$ και N ο αριθμός των ακεραίων συνέλιξης για το συγκεκριμένο παράθυρο ομαλοποίησης ($2m + 1$). Ο δείκτης j είναι ο τρέχον δείκτης του αρχικού πίνακα δεδομένων. Ο ομαλοποιημένος πίνακας αποτελείται από $2m + 1$ σημεία, όπου το m ισούται με το μισό πλάτος του παραθύρου εξομάλυνσης.

Όταν το φίλτρο εφαρμοστεί στη χρονοσειρά NDVI πρέπει να καθοριστούν δύο παράμετροι. Η πρώτη παράμετρος είναι η m , που περιεγράφηκε παραπάνω, όπου μεγαλύτερες τιμές της έχουν ως αποτέλεσμα την παραγωγή ομαλότερου αποτελέσματος “συμπιέζοντας” ωστόσο τις κορυφές. Η δεύτερη παράμετρος είναι ο ακέραιος d , που καθορίζει το βαθμό του πολυωνύμου εξομάλυνσης, ο οποίος συνήθως παίρνει τιμές από 2 έως 4. Μικρότερη τιμή του d παράγει ομαλότερο αποτέλεσμα, αλλά μπορεί να εισάγει bias, ενώ υψηλότερη τιμή μειώνει το bias αλλά μπορεί να δημιουργήσει το πρόβλημα του over fit εισάγοντας θόρυβο στα τελικά προϊόντα.



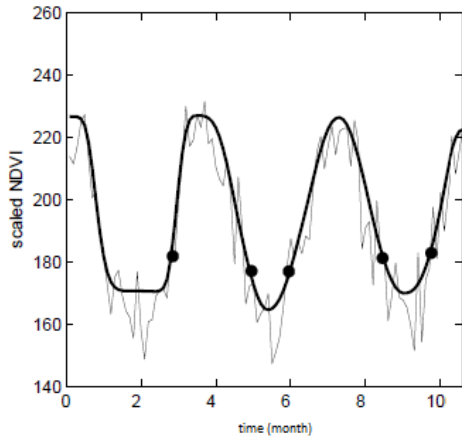
Εικόνα 3-3: Αρχική χρονοσειρά NDVI (λεπτή γραμμή), τελική χρονοσειρά NDVI (παχιά γραμμή) με τη χρήση του φίλτρου Savitzky–Golay

Εξαγωγή εποχιακών παραμέτρων

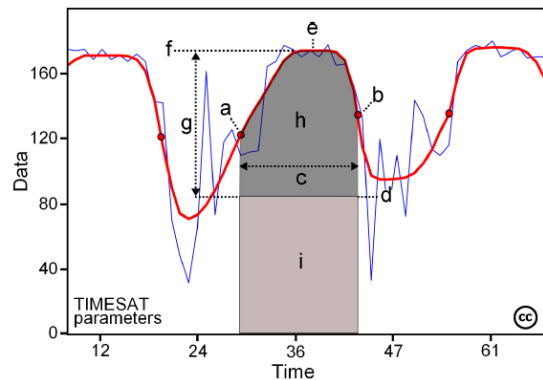
Με την ολοκλήρωση της ανάλυσης των χρονοσειρών NDVI, οι παράμετροι αποθηκευτήκαν σε κατάλληλη βάση δεδομένων και αξιοποιήθηκαν για την περαιτέρω ταξινόμηση των διάφορων τύπων βλάστησης καθώς και των καλλιεργειών. Οι παράμετροι που αξιοποιήθηκαν ήταν οι εξής:

- Ο χρόνος έναρξης της καλλιεργήσιμης περιόδου.
- Ο χρόνος ολοκλήρωσης της καλλιεργήσιμης περιόδου.
- Το χρονικό εύρος της περιόδου (με βάση την έναρξη και την ολοκλήρωση).
- Το επίπεδο βάσης (ως το μέσο όρο των ελαχίστων τιμών του αριστερού και του δεξιού άκρου της συνάρτησης).
- Ο χρόνος για το μέσο της εποχής, ο οποίος υπολογίζεται ως ο μέσος όρος των χρόνων όπου το αριστερό άκρο της συνάρτησης αυξήθηκε κατά 80% και το δεξί άκρο μειώθηκε κατά 80%.
- Η μέγιστη τιμή για κάθε ομαλοποιημένη συνάρτηση σε κάθε περίοδο.

Τα εποχιακά δεδομένα εξάγονται για κάθε μία από τις καλλιεργήσιμες περιόδους για κάθε έτος αναφοράς. Από την Εικόνα 3-5 γίνεται αντιληπτό πως η αρχή της περιόδου (a) ορίζεται από την ομαλοποιημένη συνάρτηση ως το σημείο για το οποίο η τιμή αυξήθηκε κατά έναν ορισμένο αριθμό ή ποσοστό της απόστασης μεταξύ του αριστερού ελάχιστου και μέγιστου επιπέδου. Το τέλος της περιόδου (b) ορίζεται με παρόμοιο τρόπο εξετάζοντας το δεξί τμήμα της ομαλοποιημένης συνάρτησης. Δυσκολία παρουσιάζει ο ορισμός του μέσου της περιόδου (e), του οποίου η θέση προκύπτει ανάμεσα από τις θέσεις (c,d), για τις οποίες οι τιμές στην προσαρμοσμένη συνάρτηση έχουν αυξηθεί κατά 90% μεταξύ των αριστερών και δεξιών ελάχιστων και μέγιστων επιπέδων. Τέλος, το πλάτος της περιόδου (f) λαμβάνεται ως η διαφορά μεταξύ της κορυφής και του μέσου όρου των ελάχιστων τιμών αριστερά και δεξιά.



Εικόνα 3-4: Ομαλοποιημένη συνάρτηση χρονοσειρών NDVI



Εικόνα 3-5: Παράμετροι εποχικότητας (a) αρχή περιόδου, (b) τέλος περιόδου, (c) εύρος, (d) τιμή βάσης, (e) χρόνος μέσου περιόδου, (f) μέγιστη τιμή, (g) πλάτος, (h) small integrated value, (h+i) large integrated value

3.2.2 Κατηγορίες χρήσεων γης

Η μεθοδολογική προσέγγιση που αναπτύχθηκε είναι ικανή να παράγει χάρτες καλύψεων γης, για τις εξής κατηγορίες:



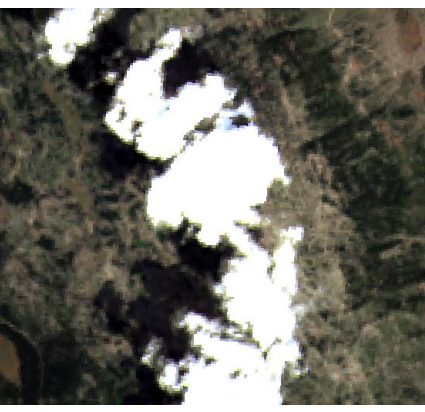

Κατηγορία	Υπό-κατηγορία	Χρώμα
Υδάτινος όγκος	Θαλάσσια ύδατα	Blue
	Ρηχά θαλάσσια ύδατα	Cyan
	Εσωτερικά ύδατα	Light Blue
Βλάστηση	Κωνοφόρα δάση	Dark Green
	Πλατύφυλλα δάση	Medium Green
	Μεικτά δάση	Bright Green
Αγροτικές εκτάσεις	Διάφορες υποκατηγορίες (φασματικές)	Yellow
Θαμνολίβαδα	Σκληρόφυλλη βλάστηση	Olive Green
Χορτολιβαδική έκταση	-	Orange
Τεχνητή επιφάνεια	-	Magenta
Σύννεφα/Σκιές	-	Black
Χιόνι	-	White

Πίνακας 3-3: Χρήσεις γης που ανιχνεύει η μεθοδολογική προσέγγιση που αναπτύχθηκε

Ακολουθούν αποσπάσματα οθόνης από παραδείγματα των αντίστοιχων κατηγοριών στα δεδομένα του Landsat 8.

Περιγραφή κατηγορίας	Landsat 8 / RGB - 432
Υδάτινος όγκος / Ρηχά θαλάσσια ύδατα	
Υδάτινος όγκος / Θαλάσσια ύδατα	
Υδάτινος όγκος / Εσωτερικά ύδατα	
Βλάστηση / Κωνοφόρα δάση	

<p>Βλάστηση / Πλατύφυλλα δάση</p>	 A satellite image showing a dense forest with a dark green color palette, characteristic of broadleaf vegetation.
<p>Βλάστηση / Μεικτά δάση</p>	 A satellite image showing a forest with a dark green color palette, characteristic of mixed forest.
<p>Αγροτικές εκτάσεις</p>	 A satellite image showing agricultural fields with a dark green color palette, characteristic of agricultural land.
<p>Θαμνολίβαδα / Σκληρόφυλλη βλάστηση</p>	 A satellite image showing scrubland with a dark green color palette, characteristic of scrubland.

Χορτολιβαδική έκταση	
Τεχνητή επιφάνεια	
Σύννεφα/Σκιές	
Χιόνι	

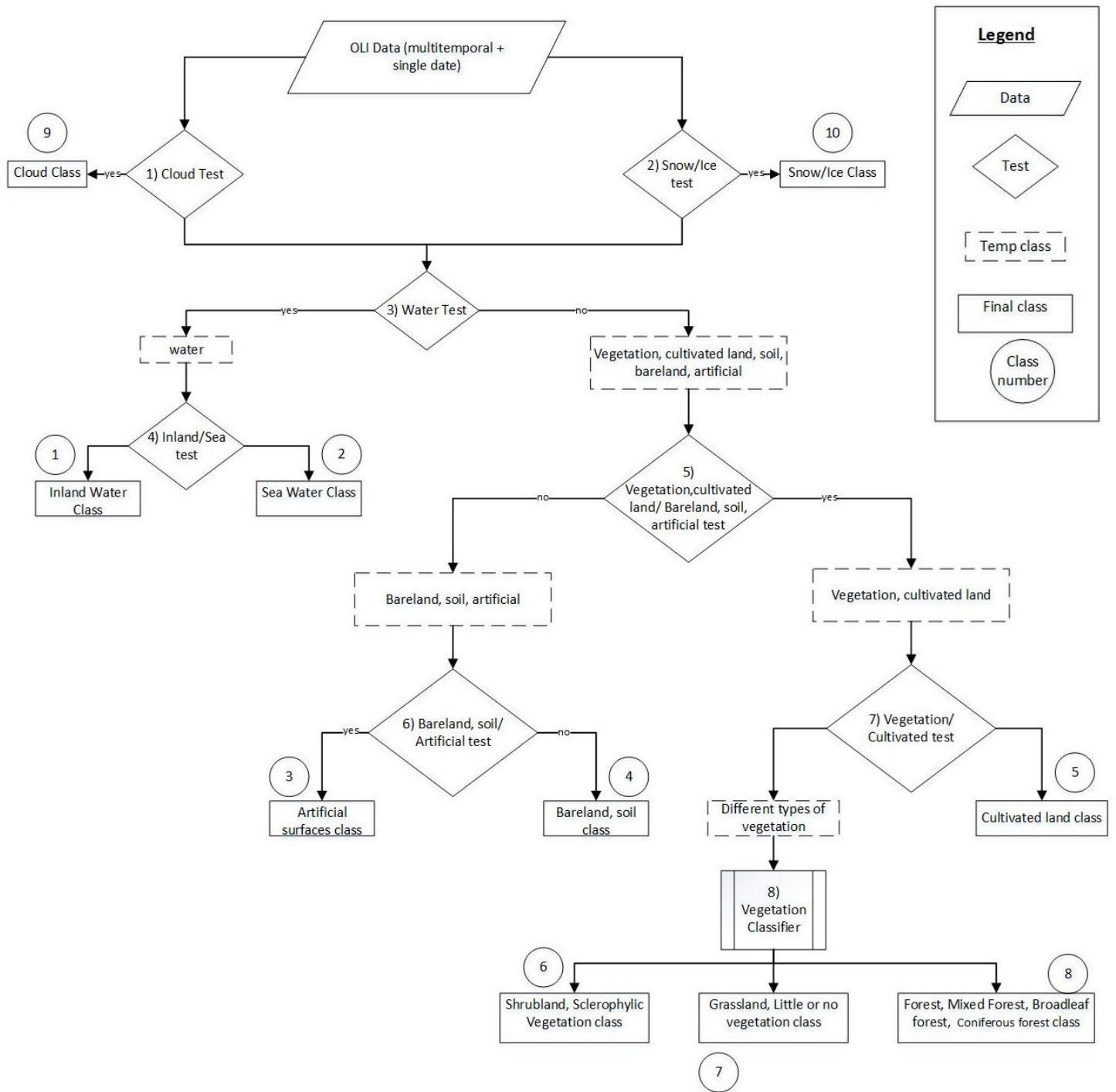
Πίνακας 3-4: Παραδείγματα χρήσεων γης που ανιχνεύονται

3.3 1^ο Στάδιο Ταξινόμησης: Περιγραφή Κανόνων για τις Κατηγορίες Χρήσεων Γης

Όπως περιεγράφηκε παραπάνω κάθε κατηγορία ταξινομείται αξιοποιώντας κάθε φορά διαφορετικά δεδομένα, διαφορετική μέθοδο καθώς και διαφορετικό αλγόριθμο ταξινόμησης. Συνοπτικά, η κατηγοριοποίηση και η ταξινόμηση των αρχικών γενικευμένων κατηγοριών πραγματοποιείται με αλγορίθμους επιλογής και κατωφλίωσης σε επιλεγμένα

κανάλια ή δείκτες (μίας χρονικής στιγμής ή διαχρονικά) ανάλογα με την κατηγορία. Αντίστοιχα, οι πιο ειδικές κατηγορίες ταξινομούνται με αλγορίθμους μηχανικής μάθησης, και πιο συγκεκριμένα μέσω των μηχανών υποστήριξης διανυσμάτων (SVM), συνδυάζοντας αποτελέσματα από την ανάλυση χρονοσειρών, την ενσωμάτωση φαινολογικών χαρακτηριστικών καθώς και από τα συλλεγμένα δείγματα. Τέλος, πρέπει να τονιστεί πως κάθε κατηγορία που ταξινομείται στη συνέχεια «μασκάρεται» και εξετάζονται τα υπόλοιπα εικονοστοιχεία αγνοώντας τα ήδη ταξινομημένα σε κάθε προηγούμενη κατηγορία.

Στη συνέχεια ακολουθεί διαγραμματική παρουσίαση του ιεραρχικού δέντρου απόφασης που παρουσιάζει την ροή εργασιών ανά κατηγορία καθώς και τα αντίστοιχα δεδομένα και τους αλγορίθμους που αξιοποιήθηκαν σε κάθε επιμέρους στάδιο για την τελική ταξινόμηση των σκηνών κατά το 1^ο στάδιο της ταξινόμησης. Παράλληλα, ακολουθεί πίνακας που παρουσιάζει συνοπτικά τα δεδομένα και τις διαδικασίες που πραγματοποιούνται για κάθε κανόνα και κατηγορία χρήσης γης:



Εικόνα 3-6: Αναλυτικό διάγραμμα ροής του 1^{ου} σταδίου της μεθοδολογικής προσέγγισης

Test Number	Description	Data	Process
1	Mask and classify clouds	Landsat 8 cfmask, cloud band, cloud conf band, Band 6	Selection, thresholding
2	Classify snow/ice	NDSI, Band 2, Band 3, Band 4	Thresholding
3	Classify water areas	NDWI, band 5	Thresholding
4	Classify inland water and sea water	Multispectral, NDWI	SVM classifier
5	Classify Vegetation, cultivated land/ Bareland, soil, artificial	Multitemporal NDVI, Multitemporal Band 4	Time series analysis, phenology, thresholding
6	Classify Bareland, soil/Artificial	Multitemporal NDBI, Multispectral	SVM, thresholding
7	Classify Vegetation/cultivated	Multitemporal NDVI, Multispectral	SVM classifier
8	Classify different types of vegetation	Multitemporal NDVI, Multispectral	SVM classifier

Εικόνα 3-7: Περιγραφή κανόνων ιεραρχικού δέντρου ταξινόμησης για το 1^ο στάδιο ταξινόμησης

1^ο Στάδιο ταξινόμησης: Περιγραφή ιεραρχικού δέντρου ταξινόμησης

- Ο αλγόριθμος δέχεται ως είσοδο διαχρονικές σκηνές της κάθε περιοχής μελέτης καθώς και την αντίστοιχη χρονική στιγμή προς ταξινόμηση σε τιμές ανάκλασης στο έδαφος.
- Στο πρώτο στάδιο πραγματοποιείται η ανίχνευση και η ταξινόμηση των σύννεφων με των σκιών τους και δημιουργείται η πρώτη μάσκα.
- Ακολουθεί η ταξινόμηση και η δημιουργία μάσκας των περιοχών που έχουν χιόνι ή πάγο.
- Στη συνέχεια πραγματοποιείται ο έλεγχος για την ταξινόμηση των υδάτινων περιοχών. Ο συγκεκριμένος κανόνας δημιουργεί δύο προσωρινές και γενικευμένες κατηγορίες χρήσεων γης, όπου η μία περιλαμβάνει το σύνολο των υδάτων (χερσαία και θαλάσσια) και η άλλη τις κατηγορίες της βλάστησης, των καλλιεργήσιμων εκτάσεων, του γυμνού εδάφους και των τεχνητών/ανθρωπογενών επιφανειών.
- Η κατηγορία των υδάτων διαχωρίζεται στη συνέχεια στην κατηγορία των χερσαίων υδάτων και των θαλάσσιων υδάτων.

- Αντίστοιχα η κατηγορία που περιέχει τη βλάστηση, τις καλλιεργήσιμες εκτάσεις, το γυμνό έδαφος και τις τεχνητές/ανθρωπογενείς επιφάνειες διαχωρίζεται σε δύο διαφορετικές κατηγορίες. Η μία περιέχει την κατηγορία του γυμνού εδάφους και των τεχνητών/ανθρωπογενών επιφανειών ενώ η άλλη τους διάφορους τύπους βλάστησης και καλλιεργειών.
- Στη συνέχεια με τους κατάλληλους ελέγχους και κανόνες η κατηγορία με το γυμνό έδαφος και τις τεχνητές/ανθρωπογενείς επιφάνειες διαχωρίζεται στις τελικές κλάσεις γυμνό έδαφος και τεχνητές/ανθρωπογενείς επιφάνειες.
- Αντίστοιχα η προσωρινή γενικευμένη κατηγορία με τους διάφορους τύπους βλάστησης και καλλιεργειών διαχωρίζεται σε δύο επιμέρους, όπου η κατηγορία των καλλιεργειών θεωρείται τελική ενώ η αντίστοιχη των διάφορων τύπων βλάστησης υφίστανται περαιτέρω ανάλυση.
- Τέλος, διαχωρίζεται η ευρύτερη κατηγορία των διάφορων τύπων βλάστησης στις εξής κατηγορίες: σκληρόφυλλη βλάστηση, λιβάδια, κωνοφόρα δάση, πλατύφυλλα δάση και μεικτά δάση.
- Ο αλγόριθμος δίνει ως έξοδο το χάρτη κάλυψης γης για τη χρονική στιγμή που επιλέχθηκε, αξιοποιώντας παράλληλα το διαχρονικό στατιστικό μοντέλο εκπαίδευσης που περιγράφεται στο κεφάλαιο 3.4.

Ακολουθούν οι αναλυτικοί κανόνες για κάθε επιμέρους κατηγορία χρήσης γης.

3.3.1 Σύννεφα/Σκιές

Σε πρώτο στάδιο πραγματοποιείται η ταξινόμηση των σύννεφων και των σκιών τους. Για την ανίχνευση τους αξιοποιήθηκαν τα προϊόντα CFmask, CFmask Cloud Confidence, το κανάλι SWIR 1 και παράλληλα δημιουργήθηκε το κανάλι Coastal Aerosol + $\frac{1}{4} \left(\frac{\text{Coastal Aerosol-Green}}{\text{Coastal Aerosol}} \right)$, έτσι ώστε να αυξηθεί η ακρίβεια της ταξινόμησης.

Σύμφωνα με την περιγραφή των ατμοσφαιρικά διορθωμένων προϊόντων του Landsat 8 τα κανάλια CFmask, CFmask Cloud Confidence έχουν τις παρακάτω ιδιότητες:

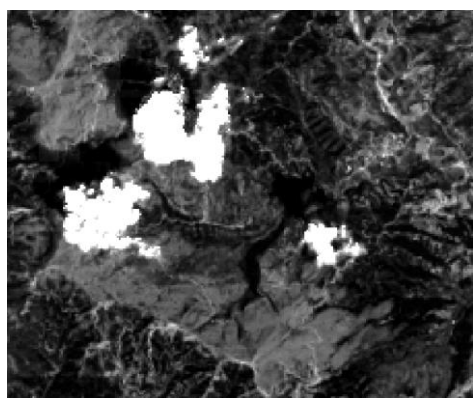
Pixel Value	Interpretation
255	Fill
0	Clear
1	Water
2	Cloud shadow
3	Snow
4	Cloud

Εικόνα 3-8: Επεξήγηση τιμών για το κανάλι CFmask
(http://landsat.usgs.gov/documents/provisional_l8sr_product_guide.pdf)

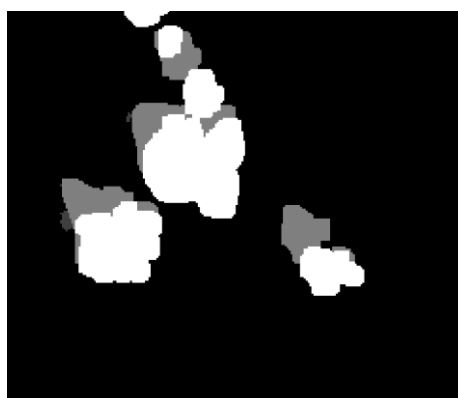
Pixel Value	Interpretation
255	Fill
0	None
1	<= 12.5% cloud confidence
2	> 12.5% and <= 22.5% cloud confidence
3	> 22.5% cloud confidence

Εικόνα 3-9: Επεξήγηση τιμών για το κανάλι CFmask Cloud Confidence (http://landsat.usgs.gov/documents/provisional_l8sr_product_guide.pdf)

Λαμβάνοντας υπόψη τους παραπάνω πίνακες ως σύννεφα και σκιές ταξινομούνται τα εικονοστοιχεία εκείνα που στο κανάλι CFmask έχουν τιμή 2 ή 4 και στο κανάλι CFmask Cloud Confidence έχουν τιμή 3. Ωστόσο, δεδομένου πως η τιμή 3 του τελευταίου καναλιού αναφέρεται σε ποσοστό εμπιστοσύνης μεγαλύτερο του 22.5 % είναι λογικό πως οι συγκεκριμένες δύο συνθήκες δεν αρκούν για την ακριβή ταξινόμηση των σύννεφων σκιών. Για το σκοπό αυτό αξιοποιείται παράλληλα το κανάλι Coastal Aerosol+ $(1)/4((\text{Coastal Aerosol-Green})/(\text{Coastal Aerosol}))$). Όπως φαίνεται από το παρακάτω απόσπασμα οθόνης τα σύννεφα και οι σκιές γίνονται πολύ εύκολα αντιληπτά ως προς τις ραδιομετρικές τους τιμές.



Εικόνα 3-10: Κανάλι για ανίχνευση σύννεφων



Εικόνα 3-11: Κανάλι CFmask



Εικόνα 3-12: Κανάλι CFmask Confidence

Παρατηρήθηκε πως οι ραδιομετρικές τιμές στις περιοχές με σύννεφα της εικόνας αναφοράς ήταν πολύ μεγαλύτερες (μεγαλύτερες των 0.1 τιμών ανακλαστικότητας) από τις αντίστοιχες ραδιομετρικές τιμές στις περιοχές της εικόνας χωρίς σύννεφα. Για τον παραπάνω λόγο, τέθηκε ένα κατώφλι έτσι ώστε να ξεχωρίσουν οι ραδιομετρικές τιμές που υποδηλώνουν την ύπαρξη σύννεφου στην εικόνα αναφοράς. Η τιμή του κατωφλιού αυτού ισούται με 0.1, αφού ύστερα από παρατηρήσεις, η τιμή αυτή θεωρήθηκε το ελάχιστο δυνατό όριο διαφοράς μεταξύ της ραδιομετρικής τιμής της εικόνας με σύννεφα και της αντίστοιχης χωρίς σύννεφα. Επιπλέον, ανιχνεύθηκε η τιμή εκείνη του καναλιού για την οποία υπήρχαν σίγουρα σύννεφα. Ύστερα από παρατηρήσεις έγινε αντιληπτό πως τα σύννεφα στο κανάλι που δημιουργήθηκε είχαν τιμή ανακλαστικότητας μεγαλύτερη ή ίση του 0.2.

Για την ανίχνευση των σκιών των σύννεφων παρατηρήθηκε πως οι ραδιομετρικές τιμές στο κανάλι 6 (SWIR 1) στις δορυφορικές εικόνες με σύννεφα ήταν πολύ μικρότερες (μικρότερες των 0.1 τιμών ανακλαστικότητας) από τις αντίστοιχες ραδιομετρικές τιμές της εικόνας χωρίς σύννεφα. Για τον παραπάνω λόγο τέθηκε ένα κατώφλι έτσι ώστε να ξεχωρίσουν οι ραδιομετρικές τιμές που υποδηλώνουν την ύπαρξη σκιάς σύννεφου στην εικόνα αναφοράς. Η τιμή του κατωφλιού αυτού ισούται με 0.1, αφού ύστερα από παρατηρήσεις, αυτή θεωρήθηκε το ελάχιστο δυνατό όριο διαφοράς μεταξύ της ραδιομετρικής τιμής της εικόνας

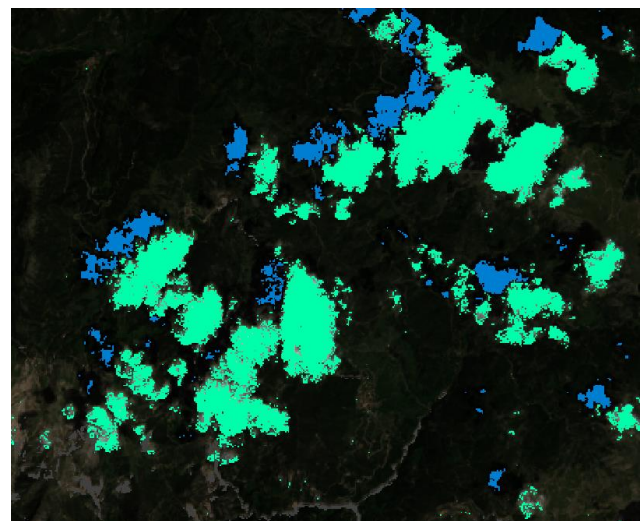
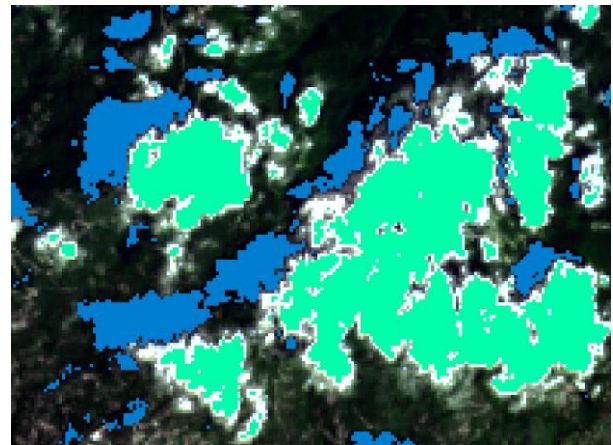
χωρίς σύννεφα και της αντίστοιχης με σύννεφα. Επιπλέον, ανιχνεύθηκε η τιμή εκείνη του καναλιού για την οποία υπήρχαν σκιές σύννεφου στην εικόνα. Ύστερα από παρατηρήσεις έγινε αντιληπτό πως οι σκιές στο κανάλι που 6 είχαν τιμή ανακλαστικότητας μικρότερη του 0.1.

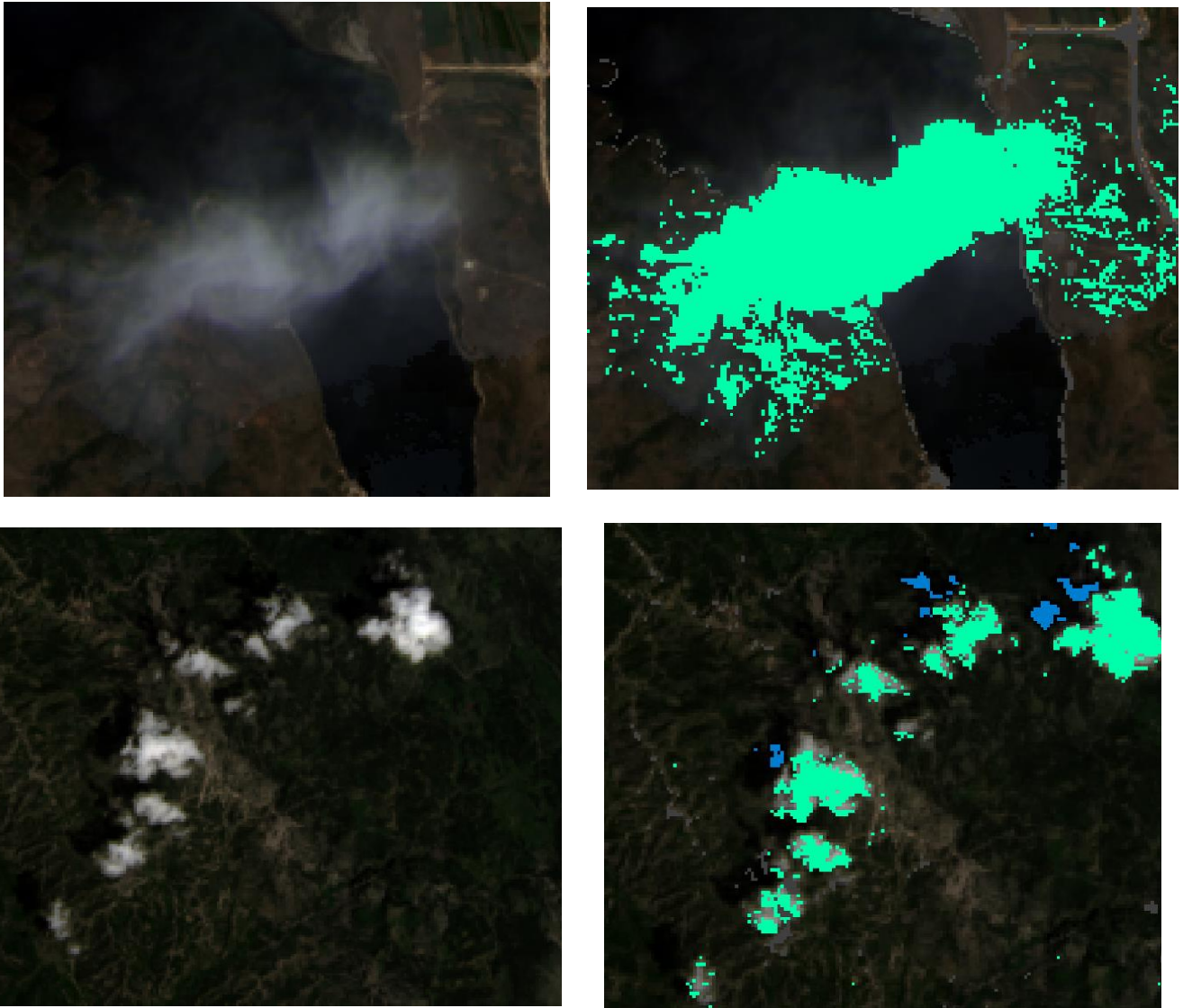
Ακολουθεί πίνακας που παρουσιάζει συνοπτικά τις συνθήκες και τα κατώφλια που χρησιμοποιήθηκαν για την ανίχνευση των σύννεφων και των σκιών.

Κανάλι	Συνθήκη	Παρατηρήσεις
CFmask	= 2 ή = 4	Εφαρμογή σε single image
CFmask Confidence	= 3	
Band 1+(1)/4((Band 1-Band 3)/(Band 1))	$t_{cloud} - t_{nocloud} > 0.1$ και $t_{cloud} > 0.2$	Λαμβάνονται υπόψη διαχρονικές σκηνές
Band 6	$t_{cloud} - t_{nocloud} < 0.1$ και $t_{cloud} > 0.1$	

Πίνακας 3-5: Συνθήκες και κατώφλια για την ανίχνευση των σύννεφων και των σκιών

Ακολουθούν αποσπάσματα οθόνης από την ταξινόμηση των σύννεφων σε διάφορες περιοχές και χρονικές στιγμές.





Εικόνα 3-13: Παραδείγματα από την ταξινόμηση των σύννεφων σε διάφορες περιοχές και χρονικές στιγμές

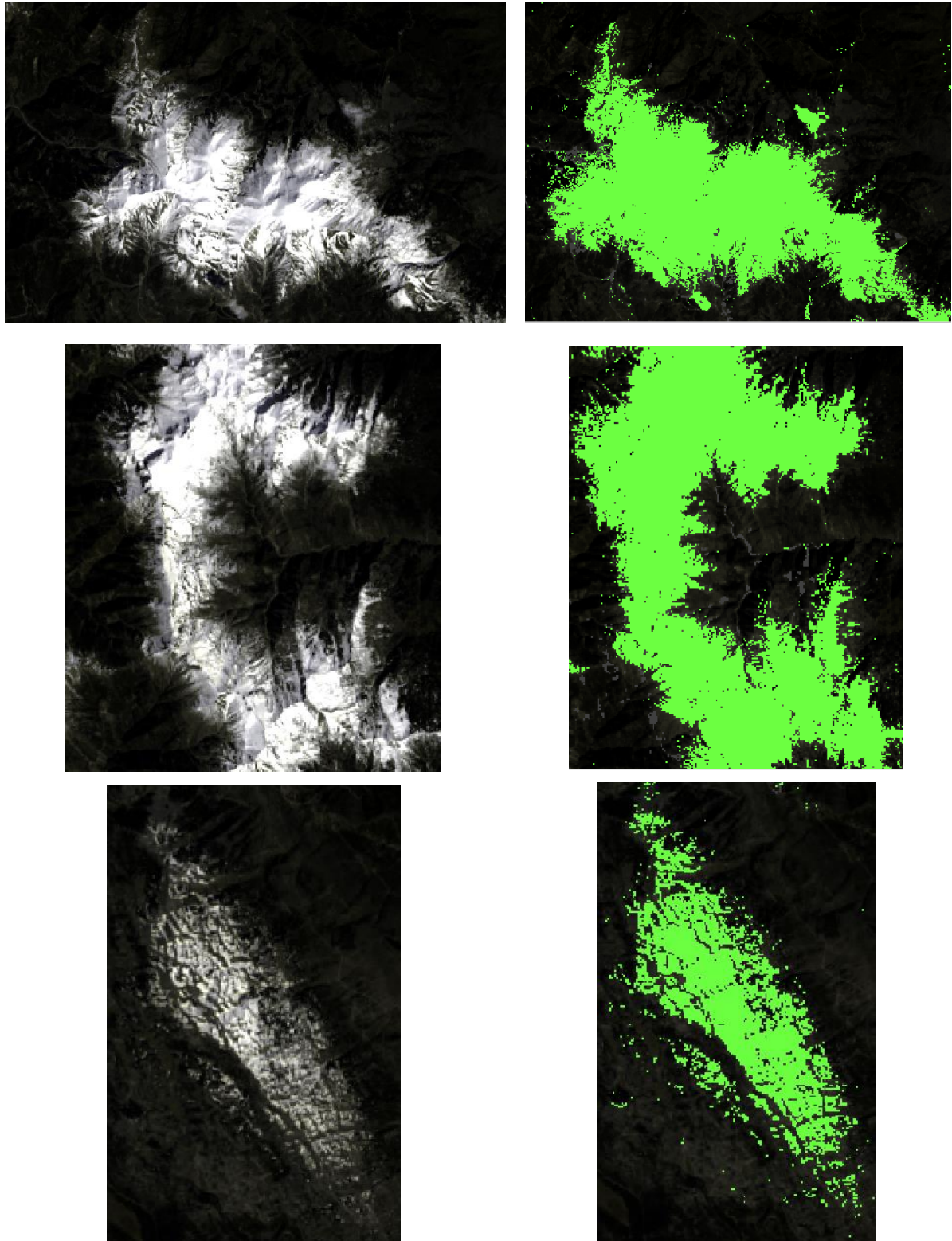
3.3.2 Χιόνι

Για την ταξινόμηση των περιοχών με χιόνι αξιοποιείται ο κανονικοποιημένος δείκτης NDSI καθώς και τα κανάλια 2, 3, 4. Πιο συγκεκριμένα παρατηρήθηκε πως ο NDSI στις περιοχές με χιόνι έχει τιμή μεγαλύτερη ή ίση του 0.6 και παράλληλα η ελάχιστη τιμή μεταξύ των καναλιών 2, 3, 4 είναι μεγαλύτερη του 0.3.

Κανάλι	Συνθήκη	Παρατηρήσεις
NDSI: $(\text{Band 3} - \text{Band 6}) / (\text{Band 3} + \text{Band 6})$	≥ 0.6	Εφαρμογή σε single image
Band 2, Band 3, Band 4	$\text{Min}(\text{Band 2}, \text{Band 3}, \text{Band 4}) > 0.3$	

Πίνακας 3-6: Συνθήκες και κατώφλια για την ανίχνευση του χιονιού

Ακολουθούν αποσπάσματα οθόνης από την ταξινόμηση του χιονιού σε διάφορες περιοχές και χρονικές στιγμές.



Εικόνα 3-14: Παραδείγματα από την ταξινόμηση του χιονιού σε διάφορες περιοχές και χρονικές στιγμές

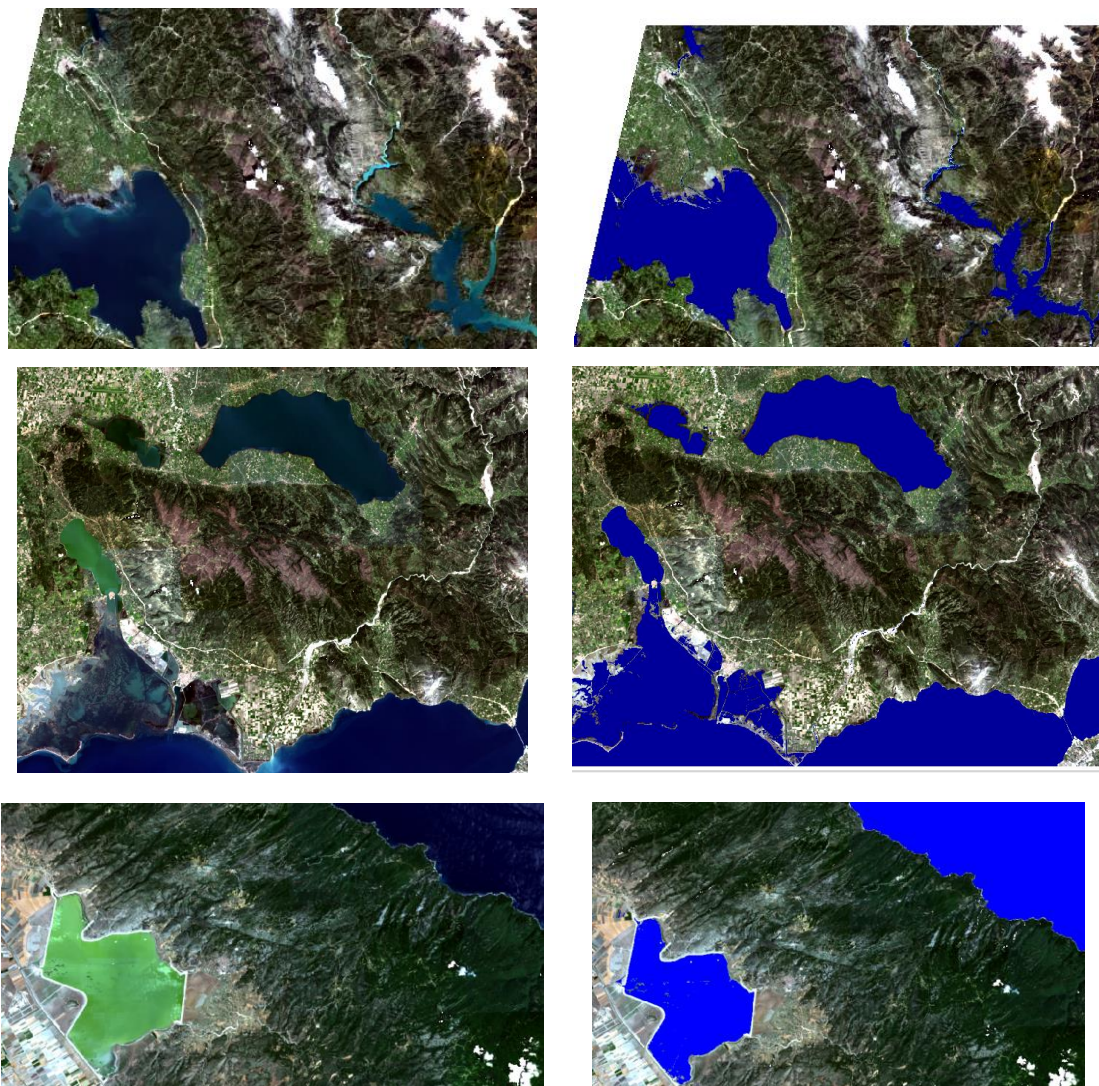
3.3.3 Υδάτινοι όγκοι

Για την ταξινόμηση των υδάτινων όγκων αξιοποιείται ο κανονικοποιημένος δείκτης NDWI καθώς και το κανάλι 5. Πιο συγκεκριμένα παρατηρήθηκε πως ο NDWI στις περιοχές με ύδατα έχει τιμή μεγαλύτερη 0 και παράλληλα το κανάλι 5 τιμή μικρότερη του 0.2. Στο συγκεκριμένο στάδιο γίνεται ανίχνευση της ευρύτερης κατηγορίας των υδάτινων όγκων και δεν πραγματοποιείται η διαχώριση των εδάφινων υδάτων, κάτι που αναλύεται στο κεφάλαιο 3.4.

Κανάλι	Συνθήκη	Παρατηρήσεις
NDWI: (Band 3 – Band 5) / (Band 3 + Band 5)	> 0	Εφαρμογή σε single image
Band 5	< 0.2	

Πίνακας 3-7: Συνθήκες και κατώφλια για την ανίχνευση των υδάτινων όγκων

Ακολουθούν αποσπάσματα οθόνης από την ταξινόμηση των υδάτινων όγκων σε διάφορες περιοχές και χρονικές στιγμές.



Εικόνα 3-15: Παραδείγματα από την ταξινόμηση των υδάτινων όγκων σε διάφορες περιοχές και χρονικές στιγμές

3.3.4 Βλάστηση

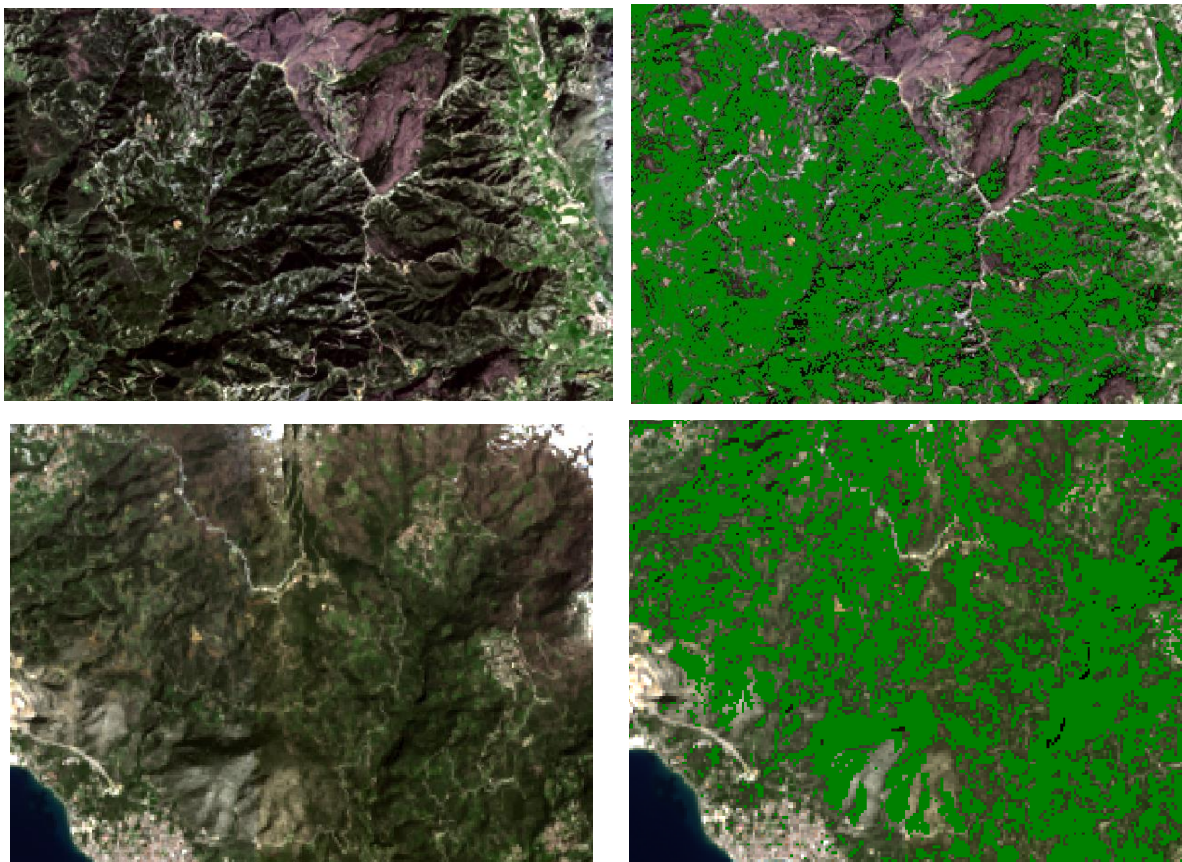
Όσον αφορά την κατηγορία της βλάστησης λαμβάνονται υπόψη διαχρονικές τιμές του κανονικοποιημένου λόγου NDVI καθώς και του καναλιού 4. Πιο συγκεκριμένα, η μεθοδολογία λαμβάνει υπόψη διαχρονικά δεδομένα από διάφορες περιόδους, χρονικού βήματος μεγαλύτερου ή ίσου των τριών μηνών, έτσι ώστε να είναι δυνατή η μελέτη των φαινολογικών σταδίων της βλάστησης. Ως βλάστηση θεωρούνται τα εικονοστοιχεία εκείνα όπου η μέγιστη τιμή του NDVI σε κάποια από τις χρονικές στιγμές είναι μεγαλύτερη 0.4 και παράλληλα η ελάχιστη τιμή του καναλιού 4 σε κάποια από τις χρονικές στιγμές είναι μικρότερη του 0.2.

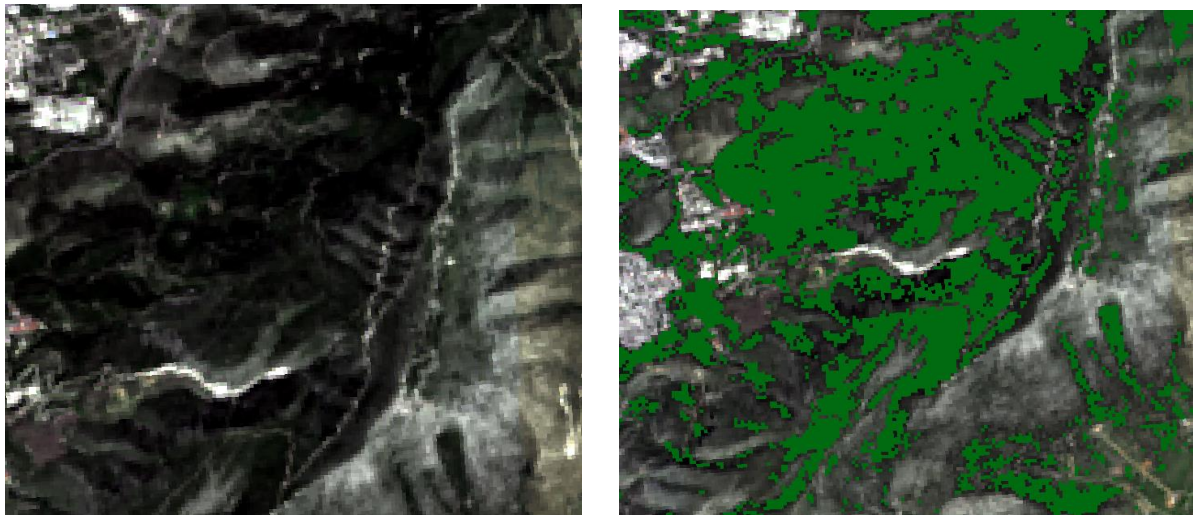
Στο συγκεκριμένο στάδιο γίνεται ανίχνευση της ευρύτερης κατηγορίας της βλάστησης και δεν πραγματοποιείται η διαχώριση μεταξύ των διαφορετικών ειδών της, κάτι που αναλύεται στο κεφάλαιο 3.4

Κανάλι	Συνθήκη	Παρατηρήσεις
NDVI: $(\text{Band } 5 - \text{Band } 4) / (\text{Band } 5 + \text{Band } 4)$	$\max(\text{NDVI}_t) > 0.4$	Λαμβάνονται υπόψη διαχρονικές σκηνές
Band 4	$\min(\text{Band } 4_t) < 0.2$	

Πίνακας 3-8: Συνθήκες και κατώφλια για την ανίχνευση της βλάστησης

Ακολουθούν αποσπάσματα οθόνης από την ταξινόμησης της βλάστησης σε διάφορες περιοχές και χρονικές στιγμές.





Εικόνα 3-16: Παραδείγματα από την ταξινόμηση της βλάστησης σε διάφορες περιοχές και χρονικές στιγμές

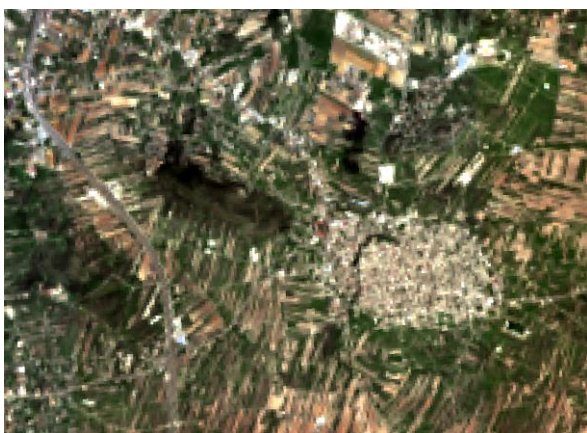
3.3.5 Τεχνητές επιφάνειες

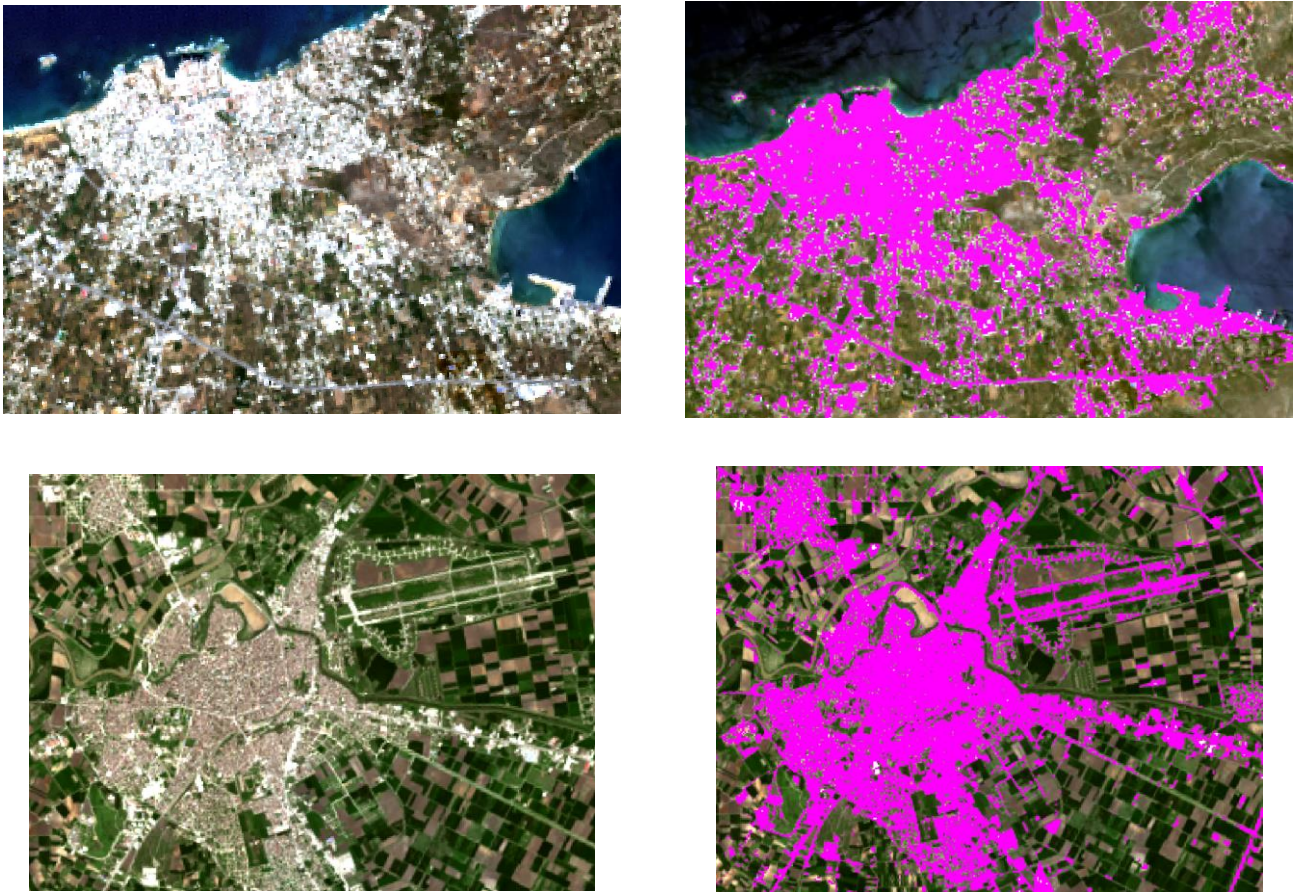
Η ταξινόμηση των τεχνητών επιφανειών πραγματοποιείται μέσα από την ανάλυση διαχρονικών τιμών του κανονικοποιημένου λόγου NDBI σε συνδυασμό με το στατιστικό μοντέλο εκμάθησης, το οποίο δημιουργήθηκε από δείγματα σε πολυφασματικά δεδομένα. Περαιτέρω ανάλυση του μοντέλου που δημιουργήθηκε πραγματοποιείται στο κεφάλαιο 3.4.

Κανάλι	Συνθήκη	Παρατηρήσεις
NDBI: (Band 6 – Band 5) / (Band 6 + Band 5)	$0 < NDBI_t < 0.3$, για κάθε χρονική στιγμή	Λαμβάνονται υπόψη διαχρονικές σκηνές
SVM σε multispectral δεδομένα	-	Εφαρμογή σε single image

Πίνακας 3-9: Συνθήκες και κατώφλια για την ανίχνευση των τεχνητών επιφανειών

Ακολουθούν αποσπάσματα οθόνης από την ταξινόμησης των τεχνητών επιφανειών σε διάφορες περιοχές και χρονικές στιγμές.



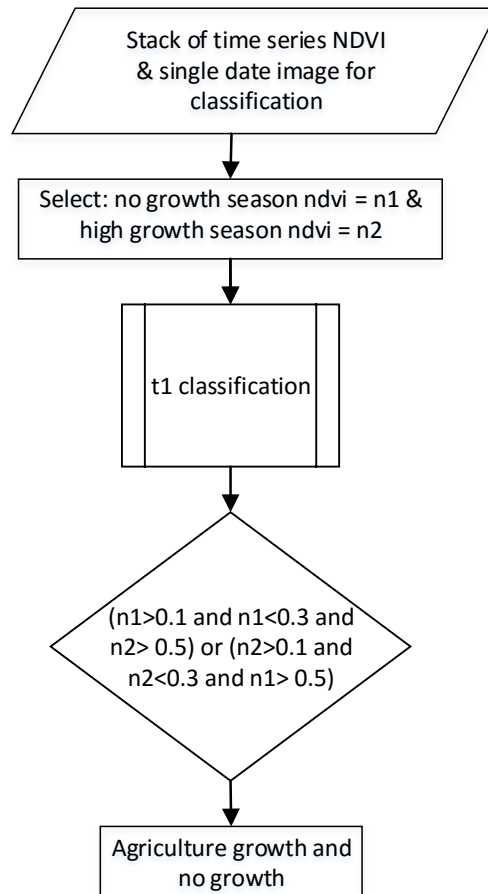


Εικόνα 3-17: Παραδείγματα από την ταξινόμηση των τεχνητών επιφανειών σε διάφορες περιοχές και χρονικές στιγμές

3.3.6 Αγροτικές εκτάσεις

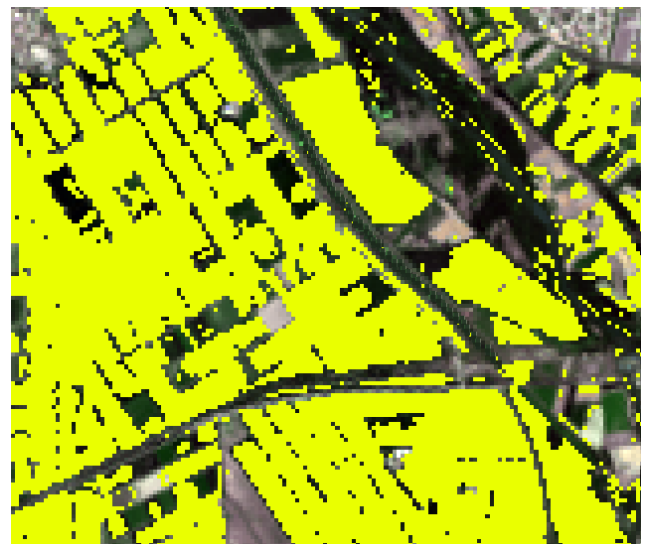
Οι αγροτικές εκτάσεις, λόγω της πολυπλοκότητάς τους, καθώς και της χαμηλής χωρικής ανάλυσης των δεδομένων του Landsat 8 ταξινομούνται σε δύο διαφορετικά στάδια. Στο πρώτο στάδιο πραγματοποιείται ανάλυση και κατωφλίωση των διαχρονικών κανονικοποιημένων λόγων NDVI, ενώ κατά το δεύτερο στάδιο ταξινομούνται μέσω του στατιστικού μοντέλου εκπαίδευσης και του αλγορίθμου SVM.

Στο πρώτο στάδιο αξιοποιούνται τα διαχρονικά δεδομένα του NDVI και επιλέγονται εκείνα που αντιστοιχούν στην περίοδο έναρξης της καλλιεργήσιμης περιόδου (έστω t_1 και $NDVI_1$) καθώς και αντίστοιχα κοντά στην περίοδο λήψης της καλλιεργήσιμης περιόδου (έστω t_2 και $NDVI_2$). Στη συνέχεια ελέγχεται αν η τιμή του NDVI για τη χρονική στιγμή t_1 ανήκει στο διάστημα 0.1 – 0.3 και παράλληλα η τιμή του NDVI για τη χρονική στιγμή t_2 είναι μεγαλύτερη του 0.5. Παράλληλα, πραγματοποιείται και ο αντίστροφος έλεγχος, δηλαδή αν η τιμή του NDVI για τη χρονική στιγμή t_2 ανήκει στο διάστημα 0.1 – 0.3 και παράλληλα η τιμή του NDVI για τη χρονική στιγμή t_1 είναι μεγαλύτερη του 0.5, έτσι ώστε να ανιχνευθούν όλα τα είδη καλλιεργήσιμων εκτάσεων. Οι παραπάνω συνθήκες ενισχύονται με το στατιστικό μοντέλο εκπαίδευσης που περιγράφεται στο κεφάλαιο 3.4. Ακολουθεί το διάγραμμα ροής των συνθηκών που αναλύθηκαν:



Εικόνα 3-18: Διάγραμμα ροής για την ανίχνευση των καλλιεργειών

Ακολουθούν αποσπάσματα οθόνης από την ταξινόμηση των αγροτικών εκτάσεων σε διάφορες περιοχές και χρονικές στιγμές.





Εικόνα 3-19: Παραδείγματα από την ταξινόμηση των καλλιεργειών σε διάφορες περιοχές και χρονικές στιγμές

Παρατήρηση: Η συγκεκριμένη προσέγγιση αντιμετωπίζει αποτελεσματικά την ταξινόμηση των καλλιεργειών που παρουσιάζουν διακύμανση στις διάφορες χρονικές περιόδους, ωστόσο δεν είναι ικανή να ανιχνεύσει τις καλλιέργειες εκείνες που δεν εμφανίζουν διακύμανση και χαρακτηρίζονται από υψηλές τιμές του NDVI σε όλη τη διάρκεια του έτους. Οι παραπάνω περιπτώσεις αντιμετωπίζονται στη συνέχεια με την εφαρμογή του αλγορίθμου SVM σε συνδυασμό με τη βιβλιοθήκη δειγμάτων που δημιουργήθηκε.

3.3.7 Θαμνολίβαδα, χορτολιβαδικές εκτάσεις, διάφοροι τύποι βλάστησης

Οι συγκεκριμένες χρήσεις γης δεν ταξινομούνται στο πρώτο στάδιο της μεθοδολογικής προσέγγισης, αλλά ανιχνεύονται μέσω του στατιστικού μοντέλου εκπαίδευσης που περιγράφεται στο κεφάλαιο 3.4. Οι παραπάνω κατηγορίες προκύπτουν από την περαιτέρω ταξινόμηση των περιοχών εκείνων που χαρακτηρίστηκαν ως βλάστηση ή γυμνό έδαφος και όλες οι άλλες κατηγορίες «μασκάρονται» κατάλληλα.

3.4 2^ο Στάδιο Ταξινόμησης: Δημιουργία και Αξιολόγηση Διαχρονικού Στατιστικού Μοντέλου Εκπαίδευσης

Για την ενίσχυση της ακρίβειας της ταξινόμησης καθώς και για να είναι δυνατή η ταξινόμηση των πιο εξειδικευμένων χρήσεων γης αναπτύχθηκε ένα διαχρονικό στατιστικό μοντέλο εκπαίδευσης/πρόβλεψης, με βάση τις περιοχές μελέτης που επιλέχθηκαν, το οποίο εφαρμόστηκε σε συνάρτηση με τους παραπάνω κανόνες που περιεγράφηκαν. Πιο συγκεκριμένα, για κάθε περιοχή και για κάθε χρονική στιγμή ψηφιοποιήθηκαν δείγματα για τις διάφορες κατηγορίες χρήσεων γης και χρησιμοποιήθηκαν για την εκπαίδευση του ταξινομητή SVM. Το μοντέλο δημιουργήθηκε αξιοποιώντας βιβλιοθήκες ανοιχτού κώδικα που εστιάζουν στην ανάλυση δεδομένων τηλεπισκόπησης, υπολογιστικής όρασης καθώς αντίστοιχων που ενσωματώνουν εξειδικευμένους αλγορίθμους ταξινόμησης. Πιο συγκεκριμένα, αξιοποιήθηκαν οι βιβλιοθήκες Orfeo Toolbox, OpenCV και LibSVM, οι οποίες προσφέρουν ένα αποτελεσματικό πλαίσιο για την ενσωμάτωση υπαρχόντων αλγορίθμων μηχανικής μάθησης σε τηλεπισκοπικά δεδομένα, μέσω προγράμματος γραμμένο σε γλώσσα προγραμματισμού Python.

Το μοντέλο εκπαίδευσης/πρόβλεψης δημιουργήθηκε με βάση τα στατιστικά στοιχεία των διαχρονικών εικόνων συμπεριλαμβανομένων των πολυφασματικών καναλιών (κανάλια 2-7), των κανονικοποιημένων λόγων NDVI, NDWI, NDSI, NDBI καθώς και των φαινολογικών χαρακτηριστικών που εξάχθηκαν. Το στατιστικό μοντέλο που δημιουργήθηκε είχε ακρίβεια μεγαλύτερη του 85% για το σύνολο των περιοχών και δείκτη Kappa μεγαλύτερο του 0.8. Με βάση το συγκεκριμένο μοντέλο ενισχύθηκε η ακρίβεια και η αποτελεσματικότητα της μεθοδολογικής προσέγγισης δίνοντας τη δυνατότητα να ταξινομηθούν οι διάφοροι τύποι βλάστησης και καλλιεργειών.

Ακολουθεί η ποσοτική παρουσίαση και αξιολόγηση του μοντέλου που δημιουργήθηκε:

ID	Class	Precision	Recall	F-score
1	Sea Water	0.814499	1	0.897767
2	Swallow Sea Water	0.989295	0.981416	0.985345
3	Inland Water	0.933333	0.859487	0.894889
4	Broadleaf Forest	0.829499	0.948916	0.885199
5	Coniferous Forest	0.815632	0.785412	0.800237
6	Mixed Forest	0.834933	0.742321	0.785908
7	Grasslands	0.60883	0.684073	0.644262
8	Sclerophyll forests	0.795432	0.811892	0.803578
9	Crops	0.703281	0.682463	0.692716
10	Baresoil	0.81791	0.895425	0.854914

11	Artificial Surfaces	0.936077	0.954464	0.945181
12	Snow	0.905137	0.884235	0.894564
13	Cloud	1	1	1
14	Cloud Shadow	0.989189	0.924242	0.955614

Πίνακας 3-10: Στατιστικά για κάθε κλάση του στατιστικού μοντέλου εκπαίδευσης/πρόβλεψης

# of pixels	Sea Water	Swallow Sea Water	Inland Water	Broadleaf Forest	Coniferous Forest	Mixed Forest	Grasslands	Sclerophyll forests	Crops	Baresoil	Artificial Surfaces	Snow	Cloud	Cloud Shadow	Total	UA (%)
Sea Water	1246	0	10	0	0	0	0	0	0	0	0	0	0	0	1256	99.20382
Swallow Sea Water	17	1247	23	0	0	0	0	0	0	0	0	0	0	0	1287	96.892
Inland Water	198	12	625	0	0	0	0	0	0	0	0	0	0	0	835	74.8503
Broadleaf Forest	0	0	0	821	40	29	0	1	3	0	0	0	0	0	894	91.83445
Coniferous Forest	0	0	0	0	742	8	94	40	26	0	0	0	0	0	910	81.53846
Mixed Forest	0	0	0	107	89	698	40	2	2	0	0	0	0	0	938	74.41365
Grasslands	0	0	0	6	0	27	1123	187	86	25	8	0	0	10	1472	76.29076
Sclerophyll forests	0	0	0	11	0	10	98	1289	52	17	9	0	0	0	1486	86.74293
Crops	0	0	0	0	0	12	198	153	823	0	40	0	0	12	1238	66.47819
Baresoil	0	0	0	0	0	0	9	7	0	799	16	0	0	0	831	96.14922
Artificial Surfaces	0	0	0	0	0	0	10	6	16	19	1234	30	0	0	1315	93.8403
Snow	2	0	0	0	0	0	0	0	0	0	10	623	5	0	640	97.34375
Cloud	0	0	0	0	0	0	0	0	0	0	0	30	523	0	553	94.57505
Cloud Shadow	0	0	0	2	0	0	0	13	0	0	0	0	0	289	304	95.06579
Total	1463	1259	658	947	871	784	1572	1698	1008	860	1317	683	528	311	13959	
Pa (%)	85.16746	99.04686	94.98	86.69483	85.189437	89.03	71.43766	75.912839	81.65	92.907	93.6978	91.215	99.053	92.926		

Overall Accuracy = 86.55 %, Kappa Index = 0.8253

Πίνακας 3-11: Πίνακας σύγκρισης του στατιστικού μοντέλου εκπαίδευσης

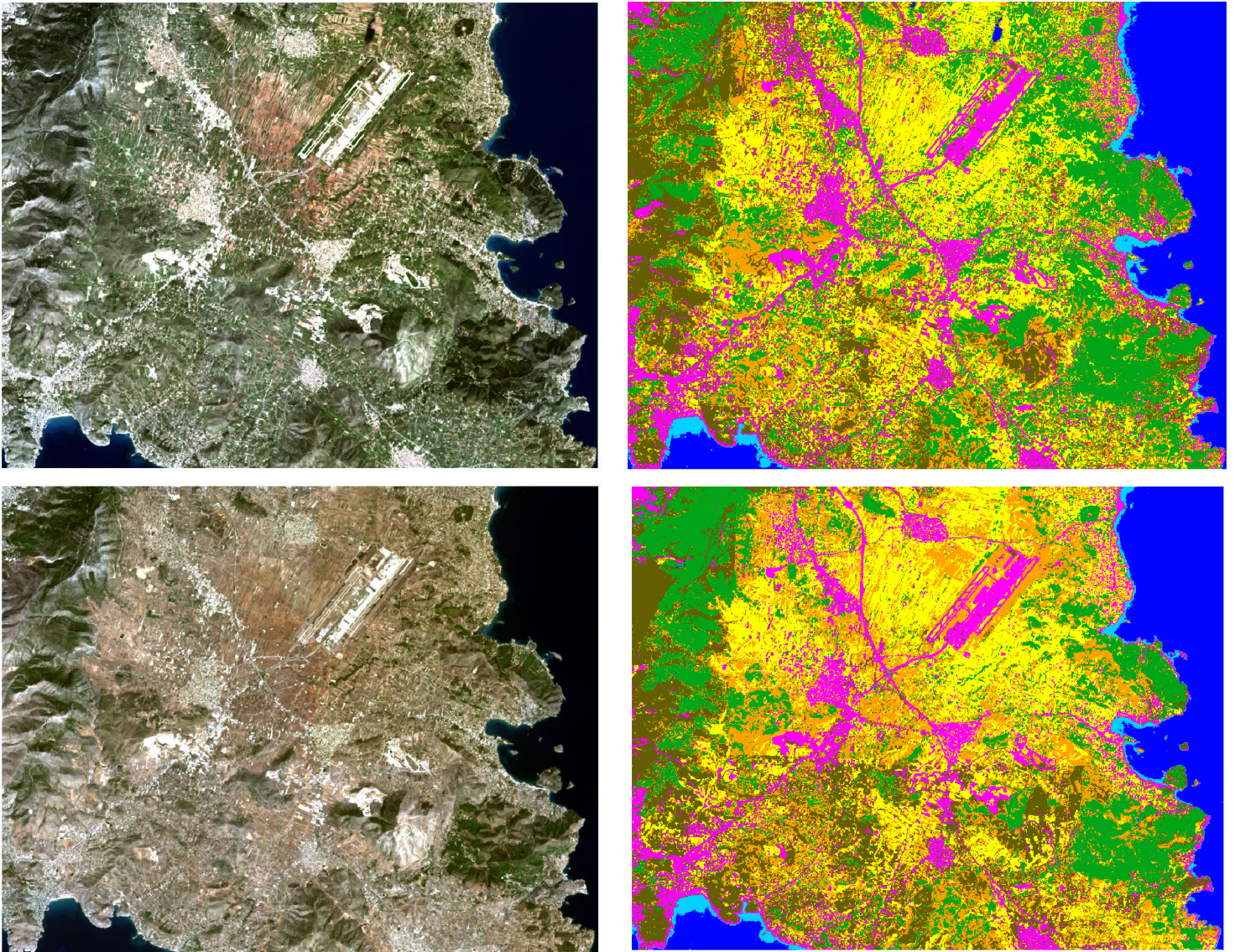
3.5 Αποτελέσματα / Αξιολόγηση

Η μεθοδολογική προσέγγιση εφαρμόστηκε σε διαφορετικές περιοχές, με πολυπλοκότητα και ποικιλία χρήσεων γης και σε διαφορετικές χρονικές στιγμές. Παράλληλα, δόθηκε προσοχή στη χωρική κατανομή των περιοχών προς ταξινόμηση εστιάζοντας στο ότι έπρεπε να καλύπτουν διαφορετικές περιοχές από το σύνολο του ελλαδικού χώρου. Παράλληλα, λαμβάνοντας υπόψη τα προβλήματα των ατμοσφαιρικά διορθωμένων προϊόντων του Landsat 8 (βλ. 3.1), ταξινομήθηκαν οι περιοχές εκείνες που δεν είχαν artifacts έτσι ώστε να μην επηρεαστεί η ακρίβεια της μεθοδολογικής προσέγγισης.

Παρατηρώντας τους πίνακες 3-12 και 3-13 γίνεται αντιληπτό πως στην περιοχή της Αττικής οι κατηγορίες του αστικού ιστού, των υδάτων και της βλάστησης ταξινομήθηκαν με μεγάλη ακρίβεια (μεγαλύτερη του 75 %), ενώ παρατηρήθηκαν προβλήματα στην κατηγορία των καλλιεργήσιμων εκτάσεων με ακρίβειες μικρότερες του 70 % ανάλογα με την εκάστοτε χρονική στιγμή. Στην περιοχή της Αιτωλοακαρνανίας (πίνακες 3-14 και 3-15) ανιχνεύθηκαν προβλήματα μεταξύ της κατηγορίας των θαμνολίβαδων και της αντίστοιχης των καλλιεργήσιμων εκτάσεων σε όλες τις χρονικές στιγμές. Το παραπάνω ήταν αναμενόμενο λαμβάνοντας υπόψη τη φασματική ομοιότητα των δύο κατηγοριών στις επιλεγμένες χρονικές στιγμές καθώς και τη χωρική ανάλυση του Landsat που δεν επιτρέπει το διαχωρισμό των δύο κατηγοριών με μεγάλη ακρίβεια. Αντίστοιχα, από τους πίνακες 3-16, 3-17 παρατηρείται πως και για την περιοχή της Βορειοδυτικής Πελοποννήσου παρουσιάστηκαν προβλήματα μεταξύ των δύο προαναφερθέντων κατηγοριών, ενώ καταγράφηκε υψηλή ακρίβεια (μεγαλύτερη του 75 %) στην κατηγορία του αστικού ιστού και των υδάτων. Αντίθετα, για την περιοχή της Θεσσαλονίκης (πίνακες 3-18, 3-19) η κατηγορία των καλλιεργήσιμων εκτάσεων χαρακτηρίστηκε από ακρίβεια μεγαλύτερη του 85 %, ενώ μεμονωμένα προβλήματα ανιχνεύθηκαν μεταξύ της σκληρόφυλλης βλάστησης και συγκεκριμένων τύπων καλλιεργειών σε τμήματα της περιοχής μελέτης. Τέλος, για την περιοχή των Χανίων (πίνακες 3-20 και 3-21) οι ακρίβειες που σημειώθηκαν ήταν ιδιαίτερα υψηλές (μεγαλύτερες του 90 % για κάθε κατηγορία) και δεν παρατηρήθηκαν χονδροειδή σφάλματα στη διαφοροποίηση των χρήσεων γης.

Ακολουθεί η αναλυτική περιγραφή και αξιολόγηση της κάθε περιοχής μελέτης για τις διάφορες χρονικές στιγμές που επιλέχθηκαν. Λαμβάνοντας υπόψη πως η μεθοδολογική προσέγγιση εκτελείται σε δύο διαφορετικά στάδια, ανάλογα με την εκάστοτε χρήση γης, κάθε κατηγορία αξιολογείται κατά το στάδιο ταξινόμησής της καθώς και στο σύνολο των κατηγοριών της εικόνας στο τέλος της διαδικασίας.

3.5.1 Περιοχή 1: Αττική



Εικόνα 3-20: Αρχική και ταξινομημένη εικόνα για την περιοχή της Αττικής (Πάνω 13/4/2015, Κάτω 4/9/2015)

# of pixels	Artificial Surfaces	Sea water	Swallow Sea Water	Coniferous Forest	Grasslands	Crops	Crops	Sclerophyll forests	Total	UA (%)
Artificial Surfaces	43	0	0	0	0	0	0	0	43	100
Sea water	0	52	0	0	0	0	0	0	52	100
Swallow Sea Water	0	0	43	1	0	12	0	0	56	76.78571
Coniferous Forest	0	0	0	37	0	0	12	2	51	72.54902
Grasslands	0	0	0	0	47	0	0	0	47	100
Crops	0	0	0	0	0	38	0	0	38	100
Crops	0	0	0	21	0	0	37	0	58	63.7931
Sclerophyll forests	0	0	0	0	0	0	2	52	54	96.2963
Total	43	52	43	59	47	50	51	54	399	
Pa (%)	100	100	100	62.71186	100	76	72.54902	100		

Overall Accuracy = 87.47 %, Kappa Index = 0.8567

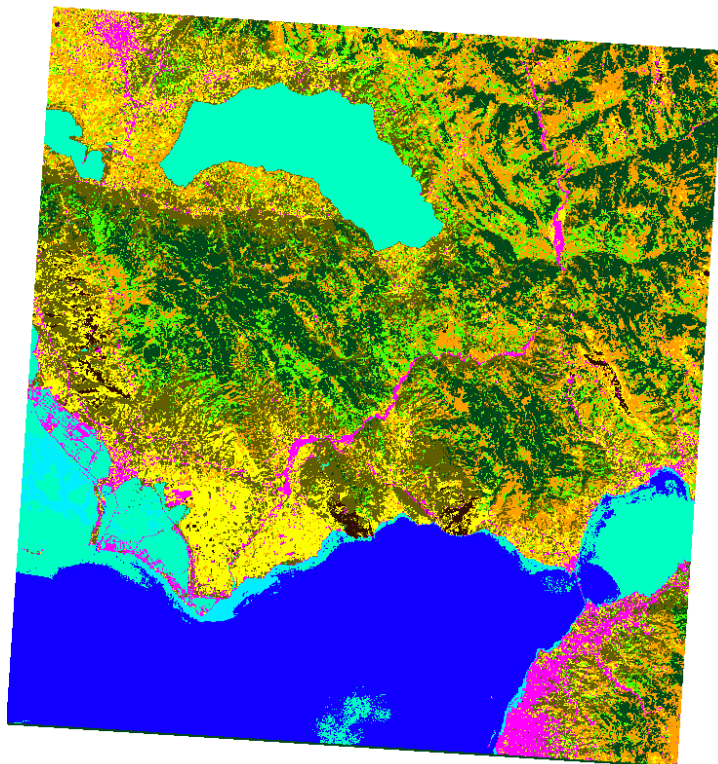
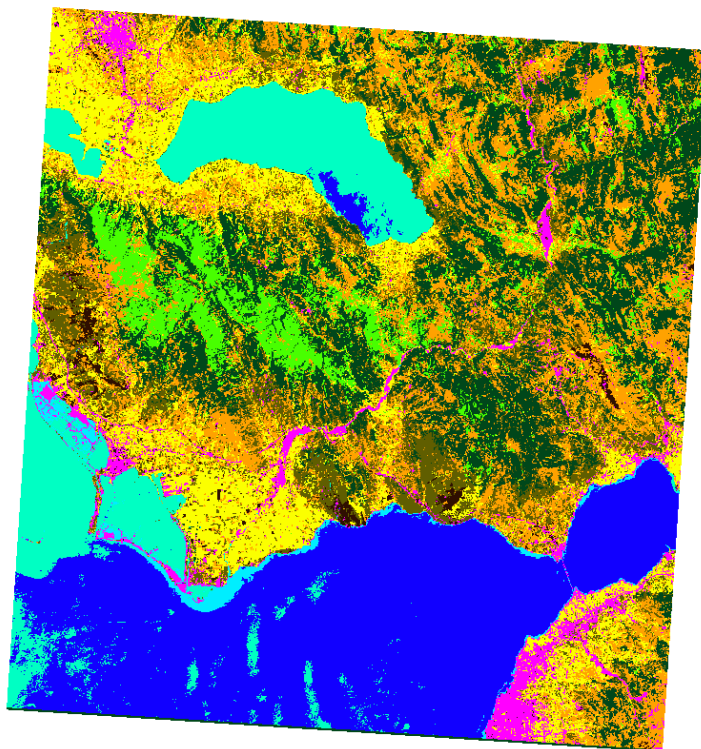
Πίνακας 3-12: Πίνακας σύγχυσης για την περιοχή της Αττικής 13/4/2015

# of pixels	Artificial Surfaces	Sea water	Swallow Sea Water	Coniferous Forest	Grasslands	Crops	Crops	Sclerophyll forests	Total	UA (%)
Artificial Surfaces	25	0	0	0	0	0	0	0	25	100
Sea water	0	30	0	0	0	0	0	0	30	100
Swallow Sea Water	0	0	29	0	0	0	0	0	29	100
Coniferous Forest	0	0	0	22	0	0	1	1	24	91.66667
Grasslands	0	0	0	0	26	1	0	0	27	96.2963
Crops	0	0	1	1	0	23	0	0	25	92
Crops	0	0	2	12	0	0	17	0	31	54.83871
Sclerophyll forests	0	0	0	0	0	0	2	26	28	92.85714
Total	25	30	32	35	26	24	20	27	219	
Pa (%)	100	100	90.625	62.85714	100	95.83333	85	96.2963		

Overall Accuracy = 90.41 %, Kappa Index = 0.8905

Πίνακας 3-13: Πίνακας σύγχυσης για την περιοχή της Αττικής 4/9/2015

3.5.2 Περιοχή 2: Αιτωλοακαρνανία



Εικόνα 3-21: Αρχική και ταξινομημένη εικόνα για την περιοχή της Αιτωλοακαρνανίας (Πάνω 1/8/2015, Κάτω 5/11/2015)

# of pixels	Broadleaf Forest	Mixed Forest	Grasslands	Baresoil	Sea Water	Swallow Sea Water	Inland Water	Inland Water	Artificial Surfaces	Sclerophyll forests	Crops	Total	UA (%)
Broadleaf Forest	80	0	0	0	0	0	0	0	0	0	0	80	100
Mixed Forest	0	79	0	0	0	0	0	0	0	0	0	79	100
Grasslands	0	1	80	0	0	0	0	0	0	6	0	87	91.95402
Baresoil	0	0	0	83	0	0	0	0	0	0	0	83	100
Sea Water	0	0	0	0	87	0	0	0	0	0	0	87	100
Swallow Sea Water	0	0	0	0	0	82	0	0	0	0	0	82	100
Inland Water	0	0	0	0	0	0	71	0	0	0	0	71	100
Inland Water	0	0	0	0	0	0	0	66	0	0	0	66	100
Artificial Surfaces	0	0	0	0	0	0	0	0	71	0	0	71	100
Sclerophyll forests	1	0	1	1	0	0	0	0	0	75	0	78	96.15385
Crops	0	0	4	1	0	0	0	0	0	2	71	78	91.02564
Total	81	80	85	85	87	82	71	66	71	83	71	862	
Pa (%)	98.76543	98.75	94.11765	97.64706	100	100	100	100	100	90.36145	100		

Overall Accuracy = 98.03 %, Kappa Index = 0.9782

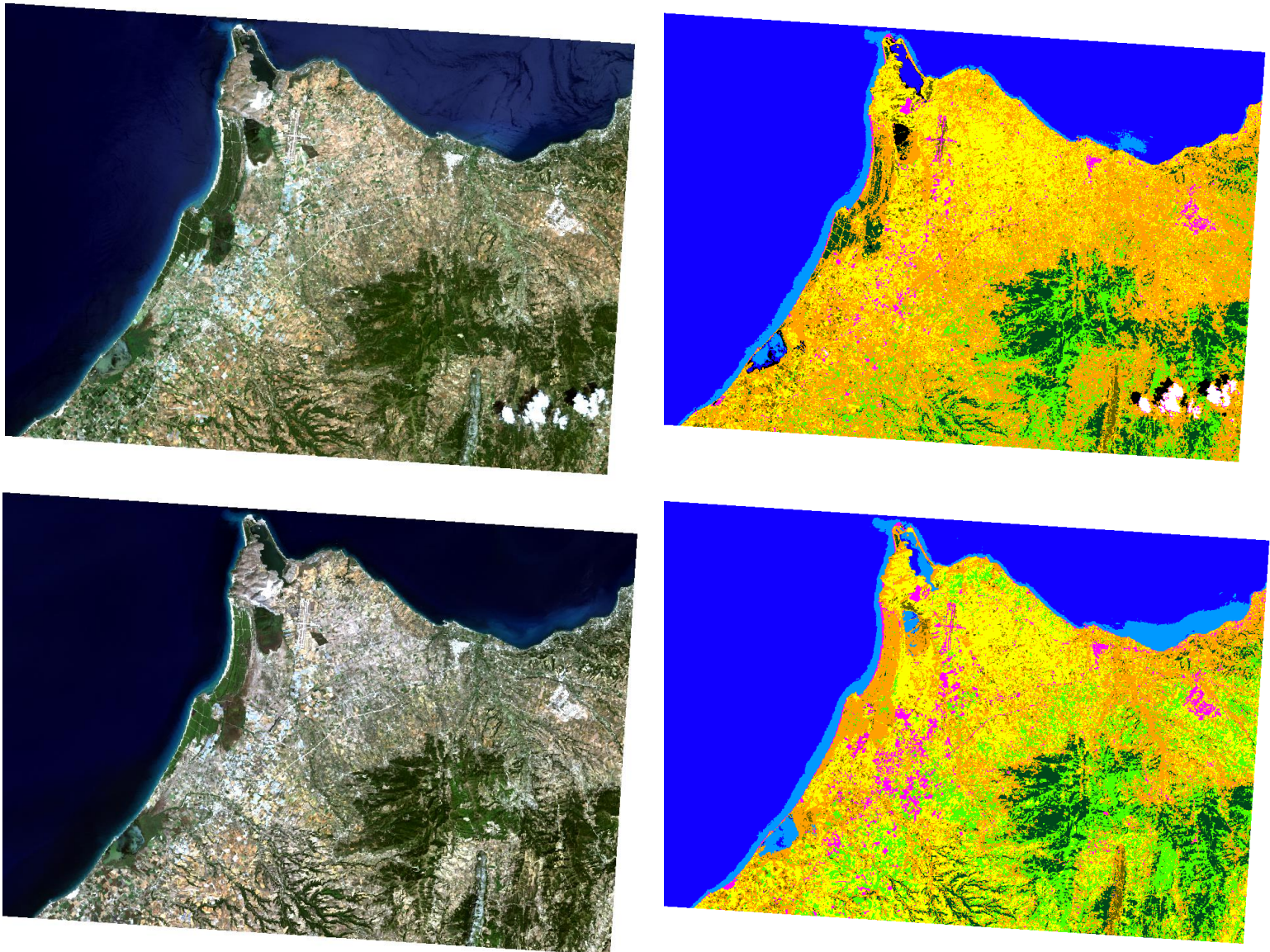
Πίνακας 3-14: Πίνακας σύγχυσης για την περιοχή της Αιτωλοακαρνανίας 1/8/2015

# of pixels	Broadleaf Forest	Mixed Forest	Grasslands	Baresoil	Sea Water	Swallow Sea Water	Inland Water	Inland Water	Artificial Surfaces	Sclerophyll forests	Crops	Total	UA (%)
Broadleaf Forest	80	0	0	0	0	0	0	0	0	0	0	80	100
Mixed Forest	0	78	0	0	0	0	0	0	0	1	0	79	98.73418
Grasslands	4	1	65	0	0	0	0	0	0	4	13	87	74.71264
Baresoil	0	0	0	82	0	0	0	0	0	0	1	83	98.79518
Sea Water	0	0	0	0	87	0	0	0	0	0	0	87	100
Swallow Sea Water	0	0	0	0	0	82	0	0	0	0	0	82	100
Inland Water	0	0	0	0	0	0	71	0	0	0	0	71	100
Inland Water	0	0	0	0	0	0	0	66	0	0	0	66	100
Artificial Surfaces	0	0	0	0	0	0	0	0	65	0	6	71	91.5493
Sclerophyll forests	1	0	0	0	0	0	0	0	0	73	4	78	93.58974
Crops	0	2	4	1	0	0	0	0	1	7	63	78	80.76923
Total	85	81	69	83	87	82	71	66	66	85	87	862	
Pa (%)	94.11765	96.2963	94.2029	98.79518	100	100	100	100	98.48485	85.88235	72.41379		

Overall Accuracy = 94.20 %, Kappa Index = 0.9362

Πίνακας 3-15: Πίνακας σύγχυσης για την περιοχή της Αιτωλοακαρνανίας 5/11/2015

3.5.3 Περιοχή 3: Βορειοδυτική Πελοπόννησος



Εικόνα 3-22: Αρχική και ταξινομημένη εικόνα για την περιοχή της Βορειοδυτικής Πελοποννήσου (Πάνω 14/6/2015, Κάτω 18/9/2015)

# of pixels	Broadleaf Forest	Mixed Forest	Grasslands	Cloud	Cloud Shadow	Sclerophyll forests	Sea Water	Swallow Sea Water	Artificial Surfaces	Crops	Total	UA (%)
Broadleaf Forest	378	1	0	0	0	1	0	0	0	34	414	91.30434
Mixed Forest	2	325	2	0	0	4	0	0	0	58	391	83.12020
Grasslands	0	5	273	0	0	26	0	0	0	1277	1581	17.26755
Cloud	0	0	0	313	0	0	0	0	0	6	319	98.11912
Cloud Shadow	0	0	0	0	135	0	0	0	0	0	135	100
Sclerophyll forests	0	0	0	0	0	292	0	0	0	273	565	51.68141
Sea Water	0	0	0	0	0	0	10536	0	0	0	10536	100
Swallow Sea Water	0	0	0	0	0	0	0	800	0	0	800	100
Artificial Surfaces	0	0	0	0	0	0	0	0	177	60	237	74.68354
Crops	0	0	19	0	0	7	0	0	2	1764	1792	98.4375
Total	380	331	294	313	135	330	10536	800	179	3472	16770	
Pa (%)	99.47368	98.18731	92.85714	100	100	88.48484	100	100	98.88268	50.80645		

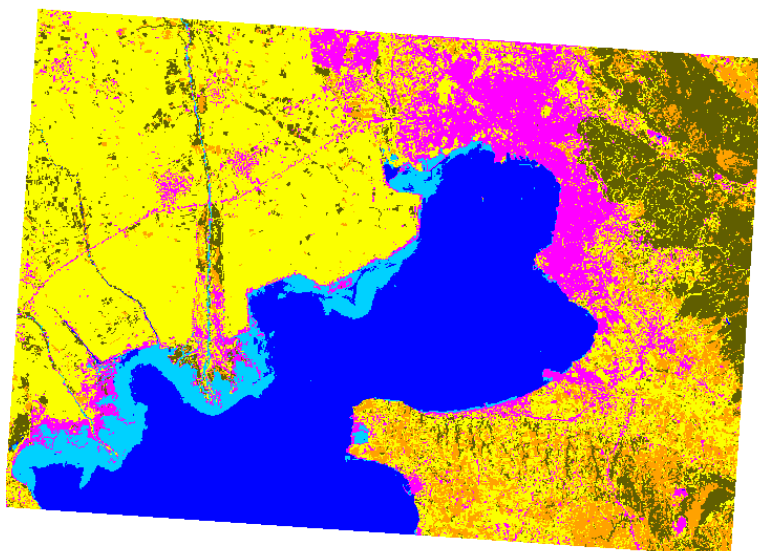
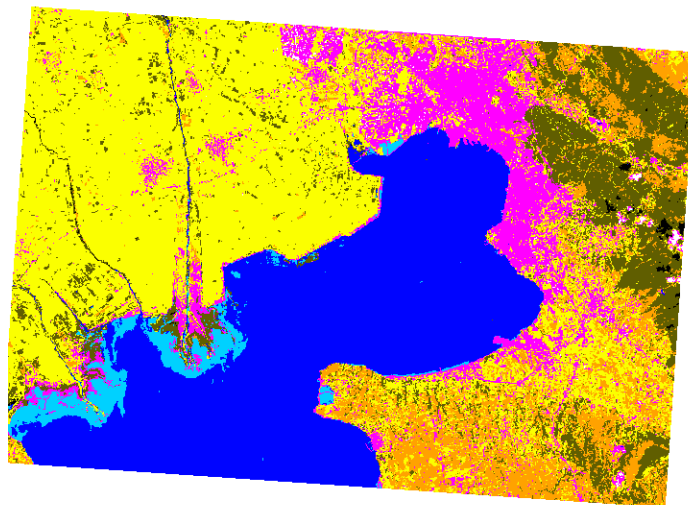
Overall Accuracy = 89.40 %, Kappa Index = 0.8163

Πίνακας 3-16: Πίνακας σύγκρισης για την περιοχή της βορειοδυτικής Πελοποννήσου 14/6/2015

# of pixels	Broadleaf Forest	Mixed Forest	Grasslands	Sclerophyll forests	Sea Water	Swallow Sea Water	Artificial Surfaces	Crops	Total	UA (%)
Broadleaf Forest	377	0	0	4	0	0	0	4	385	97.92207
Mixed Forest	3	311	4	3	0	0	0	395	716	43.43575
Grasslands	0	3	251	32	0	0	1	784	1071	23.43604
Sclerophyll forests	0	0	3	277	0	0	0	125	405	68.39506
Sea Water	0	0	0	1	10536	0	0	0	10537	99.99050
Swallow Sea Water	0	0	0	0	0	800	0	0	800	100
Artificial Surfaces	0	0	0	0	0	0	176	161	337	52.22551
Crops	0	17	36	13	0	0	2	2003	2071	96.71656
Total	380	331	294	330	10536	800	179	3472	16322	
Pa (%)	99.21053	93.9577	85.37415	83.93939	100	100	98.32402	57.69009		
Overall Accuracy = 90.25 %, Kappa Index = 0.8229										

Πίνακας 3-17: Πίνακας σύγχυσης για την περιοχή της βορειοδυτικής Πελοποννήσου 18/9/2015

3.5.4 Περιοχή 4: Θεσσαλονίκη



Εικόνα 3-23: Αρχική και ταξινομημένη εικόνα για την περιοχή της Θεσσαλονίκης (Πάνω 1/8/2015, Κάτω 18/10/2015)

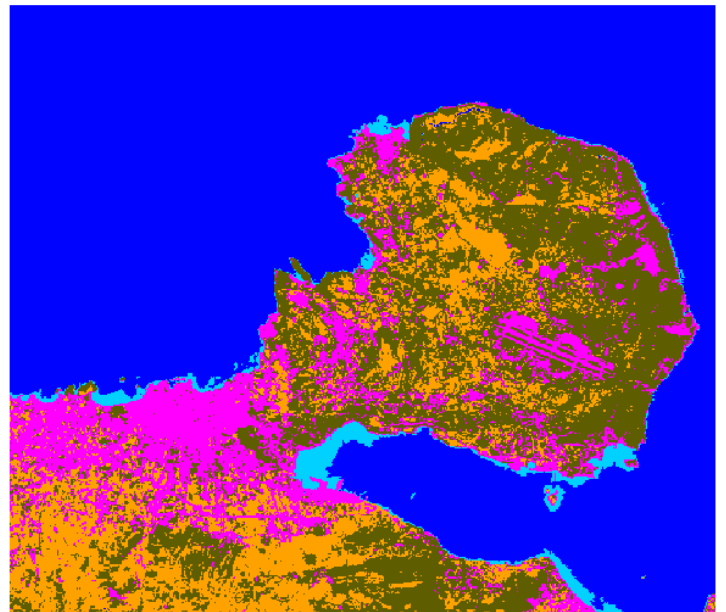
# of pixels	Sea Water	Swallow Sea Water	Artificial Surfaces	Sclerophyll forests	Grasslands	Crops	Cloud	Cloud Shadow	Total	UA (%)
Sea Water	9512	0	0	0	0	0	0	0	9512	100
Swallow Sea Water	0	264	0	0	0	0	0	0	264	100
Artificial Surfaces	0	0	905	0	0	31	0	0	936	96.68803
Sclerophyll forests	0	0	0	407	1	558	0	8	974	41.78645
Grasslands	0	0	3	0	397	71	0	0	471	84.28875
Crops	0	0	17	10	5	5224	0	0	5256	99.39117
Cloud	0	0	0	0	0	0	172	0	172	100
Cloud Shadow	0	0	0	0	0	0	0	56	56	100
Total	9512	264	925	417	403	5884	172	64	17641	
Pa (%)	100	100	96.68803	41.78645	84.28874	99.39117	100	100		
Overall Accuracy = 96.01 %, Kappa Index = 0.9340										

Πίνακας 3-18: Πίνακας σύγκρισης για την περιοχή της Θεσσαλονίκης 1/8/2015

# of pixels	Sea Water	Swallow Sea Water	Artificial Surfaces	Sclerophyll forests	Grasslands	Crops	Total	UA (%)
Sea Water	9512	0	0	0	0	0	9512	100
Swallow Sea Water	0	264	0	0	0	0	264	100
Artificial Surfaces	0	0	910	0	0	10	920	98.91304
Sclerophyll forests	0	0	0	404	1	374	779	51.86136
Grasslands	0	0	1	0	394	114	509	77.40668
Crops	0	0	14	13	8	6899	6934	99.49524
Total	9512	264	925	417	403	7397	18918	
Pa (%)	100	100	98.37838	96.88249	97.76674	93.26754		
Overall Accuracy = 97.17 %, Kappa Index = 0.9528								

Πίνακας 3-19: Πίνακας σύγχυσης για την περιοχή της Θεσσαλονίκης 18/10/2015

3.5.5 Περιοχή 5: Χανιά



Εικόνα 3-24: Αρχική και ταξινομημένη εικόνα για την περιοχή των Χανίων (Πάνω 4/9/2015, Κάτω 29/4/2015)

# of pixels	Sea Water	Swallow Sea Water	Artificial Surfaces	Sclerophyll forests	Grasslands	Total	UA (%)
Sea Water	1992	0	0	0	0	1992	100
Swallow Sea Water	0	39	0	0	0	39	100
Artificial Surfaces	0	0	241	4	0	245	98.36735
Sclerophyll forests	0	0	3	1235	7	1245	99.19678
Grasslands	0	0	0	12	443	455	97.36264
Total	1992	39	244	1251	450	3976	
Pa (%)	100	100	98.77049	98.721023	98.44444		

Overall Accuracy = 99.35 %, Kappa Index = 0.9896

Πίνακας 3-20: Πίνακας σύγχυσης για την περιοχή των Χανίων 4/9/2015

# of pixels	Sea Water	Swallow Sea Water	Artificial Surfaces	Sclerophyll forests	Grasslands	Coniferous Forest	Total	UA (%)
Sea Water	1986	0	0	0	0	0	1986	100
Swallow Sea Water	0	39	0	0	0	0	39	100
Artificial Surfaces	0	0	230	2	0	0	232	99.13793
Sclerophyll forests	0	0	14	818	2	48	882	92.74376
Grasslands	0	0	0	2	163	0	165	98.78788
Coniferous Forest	0	0	0	16	1	764	781	97.82330
Total	1986	39	244	838	166	812	4085	
Pa (%)	100	100	94.26229	97.61336	98.19277	94.08867		

Overall Accuracy = 97.92 %, Kappa Index = 0.9692

Πίνακας 3-21: Πίνακας σύγκρισης για την περιοχή των Χανίων 29/4/2015

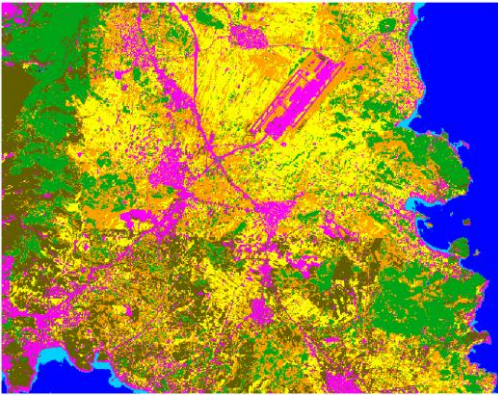
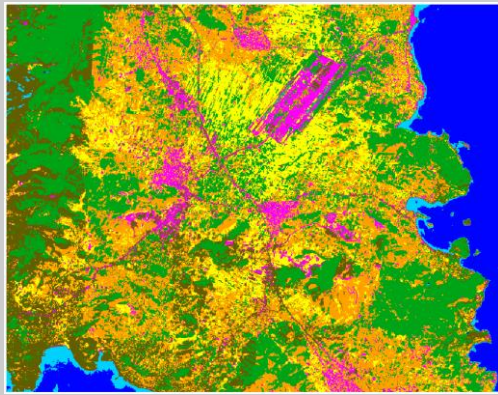
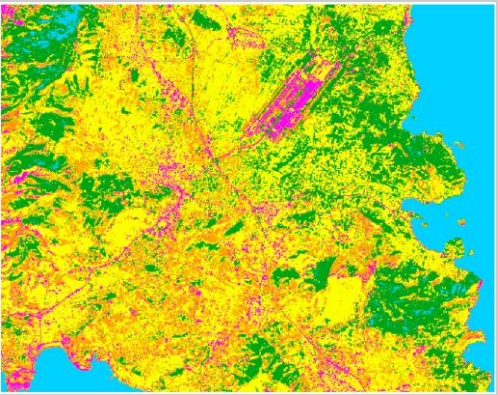
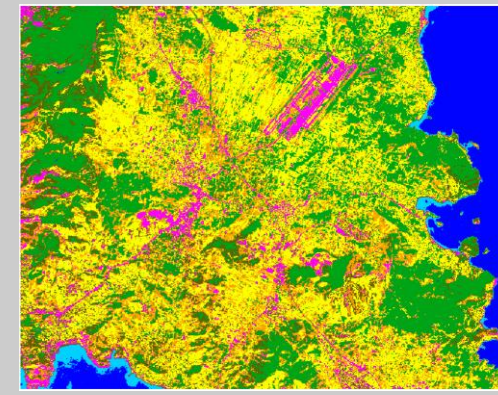
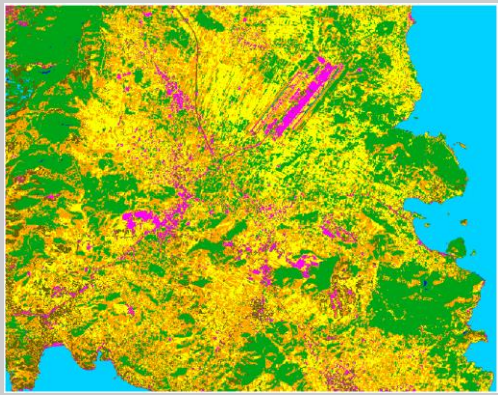
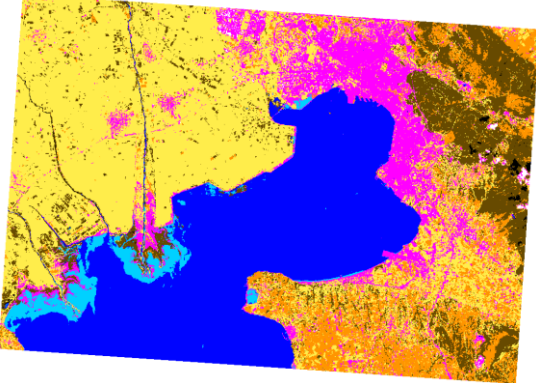
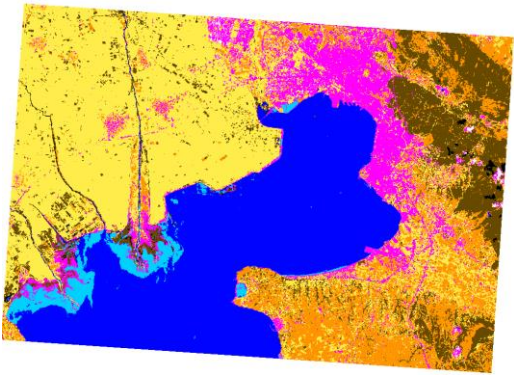
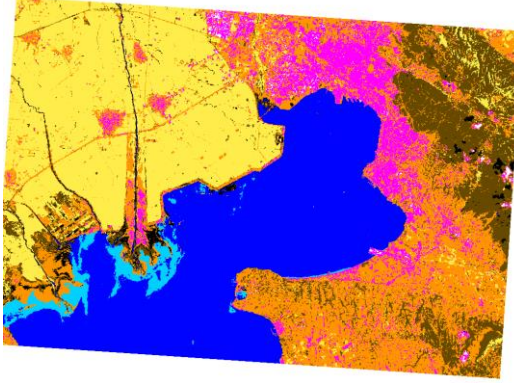
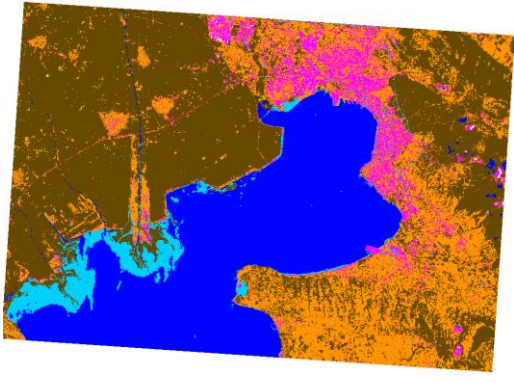
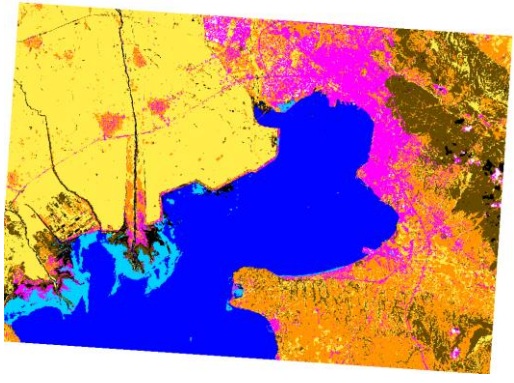

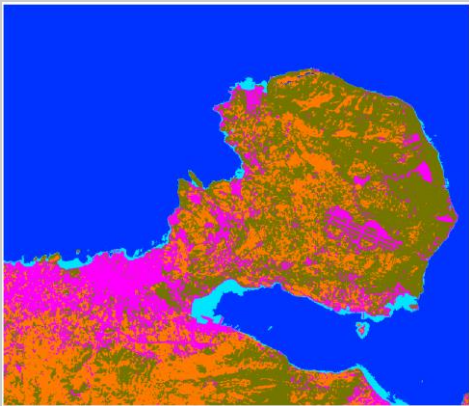
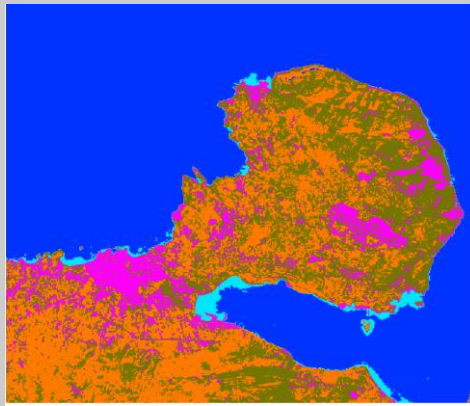


3.5.6 Σύγκριση / Αξιολόγηση Μεθοδολογίας με Άλλους Αλγορίθμους Ταξινόμησης

Στο συγκεκριμένο κεφάλαιο πραγματοποιήθηκε η σύγκριση και η αξιολόγηση των αποτελεσμάτων που προέκυψαν από τη μεθοδολογική προσέγγιση που αναπτύχθηκε με συμβατικές μεθοδολογίες ταξινόμησης, αξιοποιώντας διαφορετικούς ταξινομητές σε επιλεγμένες χρονικές στιγμές και περιοχές μελέτης. Πιο συγκεκριμένα, τα αποτελέσματα για την περιοχή της Αττικής, της Χαλκιδικής και των Χανίων συγκρίθηκαν με τους εξής αλγορίθμους ταξινόμησης: Support Vector Machine (SVM), Decision Tree (dt), K-nearest neighbors (KNN) και Random forest (rf).

Προκειμένου να είναι δυνατή η σύγκριση των αποτελεσμάτων χρησιμοποιήθηκαν ακριβώς τα ίδια δεδομένα εκπαίδευσης και αξιολόγησης για κάθε αλγόριθμο και περιοχή, έτσι ώστε να εξεταστεί η αποτελεσματικότητα του κάθε αλγορίθμου. Επιπλέον, σε αντίθεση με τη μεθοδολογική προσέγγιση που αναπτύχθηκε όλες οι κατηγορίες ταξινομούνται σε ένα στάδιο και όλες μαζί έτσι ώστε να αξιολογηθεί κατά πόσο η προσέγγιση που ακολουθήθηκε προσφέρει πιο ακριβή αποτελέσματα. Όσον αφορά την παραμετροποίηση των μεταβλητών των αλγορίθμων ταξινόμησης δεν πραγματοποιήθηκε κάποια αλλαγή στις default τιμές εκτός από τις απαραίτητες. Πιο συγκεκριμένα, για τον SVM επιλέχτηκε γραμμικός πυρήνας, για τον Decision Tree και για τον Random forest δεν περιορίστηκε το μέγιστο εύρος του δέντρου και τέλος για τον KNN δεν τέθηκε κάποιο συγκεκριμένο όριο στους γείτονες προς διερεύνηση. Ωστόσο, στο σημείο αυτό πρέπει να τονιστεί πως για να είναι αποδοτικοί οι αλγόριθμοι Decision Tree και Random forest απαιτείται ο πειραματισμός στην παραμετροποίησή τους, κάτι που δεν αποτέλεσε αντικείμενο ενασχόλησης στο συγκεκριμένο στάδιο, καθώς ο στόχος ήταν να διερευνηθεί η αποτελεσματικότητά τους με αυτοματοποιημένες διαδικασίες, λαμβάνοντας υπόψη πως η μεθοδολογική προσέγγιση που αναπτύχθηκε δεν απαιτεί παραμετροποίηση τις μεταβλητές των αλγορίθμων ταξινόμησης.

Παρατηρώντας τον πίνακα που ακολουθεί γίνεται αντιληπτό πως η μεθοδολογική προσέγγιση που αναπτύχθηκε αντιμετωπίζει στο σύνολο πιο αποτελεσματικά την ταξινόμηση των εικόνων με υψηλότερη ακρίβεια και μεγαλύτερο Kappa index. Το παραπάνω είναι λογικό καθώς η μεθοδολογική προσέγγιση λαμβάνει υπόψη τόσο διαχρονικές σκηνές όσο και τα φαινολογικά χαρακτηριστικά της κάθε εικόνας. Παράλληλα, κάθε κλάση ταξινομείται με τους δικούς της κανόνες και συνθήκες και σαν αποτέλεσμα δεν παρατηρούνται σημαντικές επικαλύψεις μεταξύ των κλάσεων. Επιπλέον, λόγω του ότι κάθε κλάση ταξινομείται ξεχωριστά και δεν συμμετέχουν όλες κατά τη διαδικασία κατασκευής του στατιστικού μοντέλου ο χρόνος εκτέλεσης του αλγορίθμου είναι αισθητά μικρότερος.

Από τον παρακάτω πίνακα προκύπτει επίσης πως για τις περιοχές μελέτης που επιλέχθηκαν ο αλγόριθμος SVM προσφέρει πιο ακριβή αποτελέσματα, έπεται ο Random forest, ακολουθεί ο Decision tree, ενώ ο KNN παρατηρήθηκε πως δεν αντιμετώπισε αποτελεσματικά την εφαρμογή εισάγοντας χονδροειδή λάθη στην ανίχνευση των κατηγοριών σε όλες τις περιοχές. Τέλος, στις περιοχές που χρησιμοποιήθηκαν μεγαλύτερες χρονοσειρές (π.χ. Χανιά) παρατηρείται πως οι ακρίβειες των αποτελεσμάτων είναι αρκετά υψηλές και όμοιες για τους διάφορους αλγορίθμους ταξινόμησης κάτι που αποδίδεται στο πιθανό "Overfitting" του μοντέλου εκπαίδευσης στις συγκεκριμένες περιοχές.

	Classifier/ Accuracy	Thesis methodology	SVM (all classes)	Decision tree	K – nearest neighbors	Random forest
Περιοχή 1: Αττική (4/9/2015)	Classified map					
	Overall Accuracy	87.47 %	81.80 %	72.51 %	65.95 %	80.54 %
	Kappa Index	0.8567	0.8006	0.7586	0.6857	0.7903
Περιοχή 4: Θεσσαλονίκη (1/8/2015)	Classified map					
	Overall Accuracy	96.01 %	90.05 %	91.13 %	63.28 %	93.21 %
	Kappa Index	0.9340	0.8925	0.8994	0.6288	0.9277
Περιοχή 5: Χανιά (4/9/2015)	Classified map					
	Overall Accuracy	99.35 %	91.35 %	92.21 %	89.52 %	95.26 %
	Kappa Index	0.9896	0.92961	0.9288	0.9076	0.9644

Πίνακας 3-22: Σύγκριση / αξιολόγηση μεθοδολογίας με άλλους αλγορίθμους ταξινόμησης (SVM multi class, decision tree, k – nearest neighbors, random forest) για επιλεγμένες περιοχές και χρονικές στιγμές

4. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Στη συγκεκριμένη μεταπτυχιακή διπλωματική εργασία αναπτύχθηκε και αξιολογήθηκε μεθοδολογική προσέγγιση για την αυτόματη παραγωγή διαχρονικών χαρτών κάλυψης και χρήσεων γης για τον ελλαδικό χώρο, αξιοποιώντας τα πολυφασματικά και ατμοσφαιρικά διορθωμένα δορυφορικά δεδομένα του Landsat 8, με χωρική ανάλυση 30 m. Από τα αποτελέσματα που προέκυψαν γίνεται αντιληπτό πως η ιεραρχική ταξινόμηση σε συνδυασμό με διαφορετικές τεχνικές ταξινόμησης (SVM, thresholding) και με την ενσωμάτωση διαχρονικών φαινολογικών χαρακτηριστικών μπορεί να οδηγήσει στην παραγωγή υψηλής ακρίβειας θεματικών χαρτών χρήσεων γης, χωρίς να απαιτείται μεγάλη υπολογιστική ισχύ, σε αντίθεση με συμβατικές μεθόδους που ταξινομούν το σύνολο της εικόνας σε ένα στάδιο.

Στη συνέχεια του συγκεκριμένου κεφαλαίου παρουσιάζονται τα συμπεράσματα που προέκυψαν από την εφαρμογή της μεθοδολογίας που αναπτύχθηκε, επισημαίνονται τα σημαντικότερα προβλήματα που παρουσιάστηκαν στα διάφορα στάδια της διαδικασίας ανάλυσης, ενώ τέλος δίνονται κατευθύνσεις για περαιτέρω έρευνα σε θέματα που αφορούν τη βελτίωση της μεθοδολογικής προσέγγισης που ακολουθήθηκε.

4.1 Συμπεράσματα

Με την εφαρμογή και την αξιολόγηση της μεθοδολογικής προσέγγισης που ακολουθήθηκε, στα διάφορα στάδια της διαδικασίας ανάλυσης, εξάχθηκαν τα παρακάτω συμπεράσματα:

- Έγινε αντιληπτή η αναγκαιότητα για την αξιοποίηση ατμοσφαιρικά διορθώσεων δορυφορικών δεδομένων στις περιπτώσεις που απαιτείται παράλληλη χρήση διαχρονικών δεδομένων για την εξαγωγή και αξιοποίηση φαινολογικών χαρακτηριστικών. Παρόλο που μέχρι στιγμής τα ατμοσφαιρικά προϊόντα που προσφέρει η USGS έχουν αρκετά προβλήματα, σε ορισμένες σκηνές, ως προς την ραδιομετρία παρατηρήθηκε πως όταν αυτά αποκοπούν οι ταξινομητές επιτυγχάνουν υψηλότερες ακρίβειες σε σύγκριση με τις ψηφιακές τιμές (DNs) ή τις TOA Reflectance. Σύμφωνα με τον επίσημο ιστότοπο της USGS τα παραπάνω προβλήματα έχουν ανιχνευθεί και βρίσκονται σε φάση επανεπεξεργασίας για την επίλυσή τους. Ως εκ τούτου αναμένεται η βελτίωση και η αύξηση της ακρίβειας των ταξινομητών και των ταξινομημένων προϊόντων.
- Η μεθοδολογική προσέγγιση που αναπτύχθηκε ακολουθεί μορφή ιεραρχικού δέντρου απόφασης, η οποία αξιοποιεί πολλαπλούς ταξινομητές ανάλογα με τη θεματική κατηγορία προς ταξινόμηση. Η συγκεκριμένη μεθοδολογία βασίζεται στη λογική του “per class classification”, δηλαδή κάθε κατηγορία ταξινομείται χωριστά και στη συνέχεια «μασκάρεται» έτσι ώστε ο αλγόριθμος ταξινόμησης να εστιάζει κάθε φορά στην ενδιαφερόμενη κατηγορία, και αξιοποιεί παράλληλα φαινολογικά χαρακτηριστικά, τα οποία εξάγονται ύστερα από κατάλληλη ανάλυση χρονοσειρών για την αύξηση της ακρίβειας στις θεματικές κατηγορίες της βλάστησης και των καλλιεργειών.

Ύστερα από την αξιολόγηση των αποτελεσμάτων και τη σύγκρισή τους με αντίστοιχες συμβατικές μεθόδους ταξινόμησης (όπου όλες οι κατηγορίες ανιχνεύονται και ταξινομούνται σε ένα στάδιο) παρατηρήθηκε πως η ιεραρχική προσέγγιση ταξινόμησης προσφέρει σημαντικά πλεονεκτήματα. Πιο συγκεκριμένα, τα φασματικά χαρακτηριστικά και η φασματική γνώση (spectral knowledge) που λαμβάνεται από τα δεδομένα εκπαίδευσης παρέχεται με έναν πιο εύκολα διαχειρίσιμο και κατανοητό τρόπο επιτρέποντας τη δημιουργία στοχευμένων κανόνων για κάθε κατηγορία. Επιπλέον, αποφεύγονται οι υπολογιστικά δαπανηρές λειτουργίες περιορίζοντας κάθε φορά τα κανάλια, τα δεδομένα εισόδου και τις κατηγορίες προς ταξινόμηση επιτρέποντας την παραλληλοποίηση των διαδικασιών. Τέλος, η συγκεκριμένη μεθοδολογική προσέγγιση είναι ιδιαίτερα ευέλικτη και μπορεί να εφαρμοστεί με μεγαλύτερη ευκολία σε διαφορετικές εφαρμογές διατηρώντας κάθε φορά τις γενικές κατηγορίες και μεταβάλλοντας τις αντίστοιχες υποκατηγορίες με τα όρια τους.

- Επιπλέον, παρατηρήθηκε πως η αξιοποίηση αλγορίθμων ταξινόμησης, που χρησιμοποιούν διαδικασίες μηχανικής μάθησης, σε εφαρμογές που απαιτούν ανάλυση χρονοσειρών δίνει ικανοποιητικά αποτελέσματα. Συγκεκριμένα, μέσω διαδικασιών εξόρυξης δεδομένων επιτυγχάνεται η αναλυτική δημιουργία μοντέλου εκπαίδευσης, το οποίο μπορεί να χρησιμοποιηθεί κατά τη διαδικασία ταξινόμησης διαφορετικών περιοχών και χρονικών με υψηλότερα ποσοστά ακρίβειας απ' ό,τι επιτυγχάνουν οι συμβατικές μέθοδοι ταξινόμησης.
- Η ενσωμάτωση χαρακτηριστικών που σχετίζονται με το χρόνο είναι ιδιαίτερα σημαντική για τη βελτίωση των αποτελεσμάτων της ταξινόμησης αξιοποιώντας δορυφορικά δεδομένα. Για το σκοπό αυτό, εξάχθηκαν και αξιοποιήθηκαν φαινολογικά χαρακτηριστικά (συμπεριλαμβανομένων της έναρξης και λήξης της καλλιεργήσιμης περιόδου, τη διάρκεια της καλλιεργήσιμης περιόδου, το εποχιακό πλάτος καθώς και της μέγιστης τιμής του κανονικοποιημένου δείκτη NDVI), τα οποία προήλθαν από την ανάλυση χρονοσειρών NDVI του Landsat 8. Η ενσωμάτωση των φαινολογικών χαρακτηριστικών ήταν απαραίτητη για την ακριβή ταξινόμηση των κατηγοριών της βλάστησης και των καλλιεργειών στις διάφορες χρονικές στιγμές, κάτι που δεν θα ήταν δυνατό αν δεν είχε προηγηθεί η παραπάνω διαδικασία. Ωστόσο, πρέπει να τονιστεί πως η ακριβή εξαγωγή των φαινολογικών χαρακτηριστικών αποτελεί μία χρονοβόρα και μη αυτόματη διαδικασία, η οποία εξαρτάται σε σημαντικό βαθμό από το μέγεθος της χρονοσειράς που θα χρησιμοποιηθεί καθώς και από την ποιότητα των πρωτογενών δεδομένων. Παρόλα αυτά προκειμένου να είναι δυνατή η ακριβή διαχρονική μελέτη των χρήσεων γης είναι απαραίτητη η ενσωμάτωσή του συγκεκριμένου σταδίου για διάφορες χρονικές περιόδους και σε δεδομένα που έχουν εξαρχής διορθωθεί ατμοσφαιρικά.
- Σε ορισμένες περιπτώσεις, κυρίως στην περιοχή των Χανίων, παρατηρήθηκε το φαινόμενο του “overfitting” στο στατιστικό μοντέλο εκπαίδευσης και σαν συνέπεια υπήρξαν χρονικές στιγμές που ταξινομήθηκαν με πολύ μεγάλη ακρίβεια ενώ αντίθετα άλλες που είχαν αρκετά χαμηλότερη ακρίβεια. Το παραπάνω πρόβλημα προέκυψε λόγω του γεγονότος πως στις συγκεκριμένες περιοχές μελέτης για ορισμένες χρονικές στιγμές αξιοποιήθηκαν περισσότερα δεδομένα εκπαίδευσης κάτι που μπορεί να

επιλυθεί με την εφαρμογή πιο σχολαστικής ανάλυσης ευαισθησίας στον αριθμό των δειγμάτων.

- Επιπλέον, με τη μελέτη των πινάκων σύγχυσης από την αξιολόγηση των αποτελεσμάτων προκύπτει πως οι κατηγορίες που ταξινομήθηκαν στο 1^ο στάδιο της μεθοδολογικής προσέγγισης έχουν υψηλή ακρίβεια στο σύνολο τους (> 85 %) κάτι που ήταν απαιτούμενο της μεθοδολογίας που αναπτύχθηκε. Το παραπάνω είναι ιδιαίτερα σημαντικό καθώς η ιεραρχική προσέγγιση ταξινόμησης απαιτεί κάθε κατηγορία να έχει ταξινομηθεί με την καλύτερη δυνατή ακρίβεια καθώς δεν επανεξετάζεται στη συνέχεια του αλγορίθμου και υπάρχει πιθανότητα να μεταφερθούν τα σφάλματα της κάθε κατηγορίας στις αντίστοιχες υποκατηγορίες.

Επίσης, παρόλη την ενσωμάτωση των φαινολογικών χαρακτηριστικών και της ανάλυσης των χρονοσειρών παρατηρείται πως τα δορυφορικά δεδομένα μέτριας ανάλυσης, όπως είναι του Landsat 8, δεν αποτελούν τα ιδανικά, τουλάχιστον σε επιχειρησιακό επίπεδο, για την ταξινόμηση και χαρτογράφηση των υποκατηγοριών της βλάστησης και των καλλιεργειών. Ωστόσο με την ενσωμάτωση της φαινολογικής πληροφορίας παρατηρείται βελτίωση της ακρίβειας της ταξινόμησης των συγκεκριμένων κατηγοριών τουλάχιστον κατά 10-20 % σε σχέση με μεθοδολογικές προσεγγίσεις που δεν την λαμβάνουν υπόψη. Αντίστοιχα, όσον αφορά την κατηγορία του αστικού ιστού παρατηρήθηκε πως οι αλγόριθμοι μηχανικής μάθησης, και πιο συγκεκριμένα ο SVM, αντιμετωπίζουν ικανοποιητικά την ανίχνευση και ταξινόμησή του με ακρίβειες καλύτερες του 80 % ιδιαίτερα όταν μελετώνται περιοχές με πυκνή δόμηση. Ωστόσο, η ακρίβεια του αστικού ιστού μειώνεται σε περιπτώσεις αραιής δόμησης, όπου η χρήση δορυφορικών δεδομένων μεσαίας χωρικής ανάλυσης δεν ενδείκνυται.

Συνοψίζοντας, η μεθοδολογική προσέγγιση που ακολουθήθηκε, καθώς και τα αντίστοιχα προγράμματα που δημιουργήθηκαν, αντιμετώπισαν ικανοποιητικά την παραγωγή διαχρονικών χαρτών κάλυψης γης και χρήσεων γης για την περιοχή της Ελλάδας με ακρίβειες καλύτερες του 85 % στις περιοχές μελέτης που επιλέχθηκαν. Λαμβάνοντας τα παραπάνω υπόψη, μπορεί να θεωρηθεί πως η συγκεκριμένη μεθοδολογική προσέγγιση, ύστερα από κατάλληλες παραμετροποιήσεις και προσθήκες, μπορεί να αξιοποιηθεί ως μία εφικτή επιχειρησιακή λύση για την παραγωγή χαρτών κάλυψης γης σε χωρική ανάλυση 30 m για την περιοχή της Ελλάδας.

Στο επόμενο κεφάλαιο παρουσιάζονται προοπτικές εξέλιξης όσον αφορά αφενός στη μεθοδολογική προσέγγιση και αφετέρου στους αντίστοιχους αλγορίθμους που αναπτύχθηκαν

4.2 Προοπτικές Εξέλιξης

Με την αξιοποίηση της συγκεκριμένης μεθοδολογίας καθίσταται δυνατή η δημιουργία υψηλής ανάλυσης και απαλλαγμένου από σύννεφα μωσαϊκού για το σύνολο της Ελλάδας, το οποίο μπορεί να χρησιμοποιηθεί τόσο ως υπόβαθρο για εφαρμογές που απαιτούν τη μελέτη χρήσεων γης όσο και για περαιτέρω ανάλυση. Επιπλέον, με την παράλληλη χρήση δορυφορικών δεδομένων Modis και Landsat είναι δυνατή η κατασκευή ημερήσιων και

απαλλαγμένων από σύννεφα μωσαικών εικόνων αξιοποιώντας χρονοσειρές δεδομένων. Με το συγκεκριμένο τρόπο, θα υπάρξει αύξηση των διαθέσιμων προς εκμετάλλευση δεδομένων, τα οποία θα συμβάλλουν στην ενίσχυση του στατιστικού μοντέλου εκπαίδευσης. Επιπλέον, προτείνεται η δημιουργία «τράπεζας» διανυσματικών δεδομένων για το σύνολο της χώρας και μετέπειτα η δημιουργία ενός στατιστικού μοντέλου εκμάθησης, το οποίο θα είναι ικανό να ταξινομήσει εικόνες που καλύπτουν το σύνολο της χώρας, αναπτύσσοντας παράλληλα αλγορίθμους που εκμεταλλεύονται τη μονάδα γραφικών (GPU) του υπολογιστή. Με τον τρόπο αυτό θα είναι δυνατή η αυτόματη παραγωγή ταξινομημένων εικόνων, χωρίς να απαιτείται η παρέμβαση του μελετητή, ανά 16 μέρες (χρονικό βήμα λήψης δεδομένων του Landsat) και η αξιοποίησή τους σε διάφορες εφαρμογές.

Παράλληλα, είναι απαραίτητη η ενσωμάτωση μεγαλύτερης χρονοσειράς δεδομένων εκπαίδευσης (τουλάχιστον 5ετίας) έτσι ώστε να εξεταστεί κατά πόσο θα αυξηθεί η ακρίβεια των αποτελεσμάτων στις κατηγορίες της βλάστησης και των καλλιεργειών. Απαραίτητο στοιχείο είναι επίσης η παράλληλη χρήση δορυφορικών δεδομένων πολύ υψηλής χωρικής ανάλυσης (π.χ. WorldView – 3) έτσι ώστε να είναι δυνατή η πιο ακριβή ταξινόμηση των παραπάνω κατηγοριών καθώς και του αστικού ιστού. Σε συνάρτηση με το προηγούμενο, προκειμένου η συγκεκριμένη μεθοδολογική προσέγγιση να καταστεί επιχειρησιακή πρέπει να αναπτυχθεί ένα ολοκληρωμένο σύστημα αξιολόγησης, το οποίο θα αξιοποιεί σημεία ελέγχου από επίγειες μετρήσεις ή/και δεδομένα από αεροφωτογραφήσεις.

Επίσης, μέσω της αξιοποίησης των αλγορίθμων καθώς και της μεθοδολογικής προσέγγισης που παρουσιάστηκε είναι δυνατή η δημιουργία ενός ολοκληρωμένου WebGIS σε τεχνολογία cloud, το οποίο θα έχει ως στόχο τη διαρκή παροχή ταξινομημένων εικόνων για την περιοχή ενδιαφέροντος του χρήστη. Ακόμη, είναι δυνατή η παροχή, σε πραγματικό χρόνο (on the fly), ταξινομημένων εικόνων καθώς και εικόνων μεταβολών για την αντίστοιχη περιοχή μελέτης που ορίζει ο χρήστης (on demand) χρησιμοποιώντας το στατιστικό μοντέλο εκμάθησης που θα έχει προκύψει από το σύνολο της Ελλάδας. Με αυτόν τον τρόπο, θα καθίσταται εφικτή η αυτόματη ταξινόμηση οποιασδήποτε περιοχής που ορίζει ο χρήστης, παρέχοντας άμεσα προϊόντα υψηλής ακρίβειας, χωρίς να απαιτείται η αλληλεπίδραση του.

Τέλος, ενδιαφέρον παρουσιάζει η εφαρμογή και η αξιολόγηση της συγκεκριμένης μεθοδολογικής προσέγγισης σε διαφορετικές γεωγραφικές περιοχές, πλέον της Ελλάδας, έτσι ώστε να εξεταστεί ο απαιτούμενος βαθμός παραμετροποίησης των μεταβλητών σε δεδομένα που δεν έχουν συμμετάσχει καθόλου κατά τη φάση κατασκευής του στατιστικού μοντέλου εκπαίδευσης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] J. Cihlar, "Land cover mapping of large areas from satellites: status and research priorities," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1093-1114, 2000.
- [2] P. Gong, "Remote sensing of environmental change over China: A review," *Chinese Science Bulletin*, vol. 57, no. 22, pp. 2793-2801, 2012.
- [3] R. Congalton, J. Gu, K. Yadav, P. Thenkabail and M. Ozdogan, "Global Land Cover Mapping: A Review and Uncertainty Analysis," *Remote Sensing*, vol. 6, no. 12, pp. 12070-12093, 2014.
- [4] T. Loveland, B. Reed, J. Brown, D. Ohlen, J. Zhu, L. Yang and J. Merchant, "Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data," *International Journal of Remote Sensing*, vol. 21, no. 6-7, p. 1303–1330, 2000.
- [5] T. Loveland, D. J. Brown, B. R. Ohlen, Z. Zhu, L. Yang and S. Howard, "ISLSCP II IGBP DISCover and SiB Land Cover, 1992-1993," Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A., 2009.
- [6] M. A. Friedla, D. Sulla-Menashea, B. Tanb, A. Schneiderc, N. Ramankuttyd, A. Sibleya and X. Huang, "MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets," *Remote Sensing of Environment*, vol. 114, no. 1, pp. 168-182, 2010.
- [7] M. C. Hansen, R. S. Defries, J. R. G. Townshend and R. Sohlberg, "Global land cover classification at 1 km spatial resolution using a classification tree approach," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1331-1364, 2000.
- [8] E. Bartholomé and A. S. Belward, "GLC2000: a new approach to global land cover mapping from Earth observation data," *International Journal of Remote Sensing*, vol. 26, no. 9, pp. 1959-1977, 2005.
- [9] A. O., D. Gross, F. Ranera, M. Leroy, P. Bicheron, C. Brockmann, P. Defourny, C. Vancutsem, L. Schouten, F. Achard, L. Bourg, J. Latham, A. D. Gregorio, R. Witt, M. Herold, J. Sambale, S. Plummer and J.-L. Weber, "GlobCover: ESA service for Global Land Cover from MERIS," in *International GeoScience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, 2007.
- [10] P. Bicheron, V. Amberg, L. Bourg, D. Petit, M. Huc, B. Miras, C. Brockmann, O. Hagolle, S. Delwart, F. Ranera, M. Leroy and O. Arino, "Geolocation Assessment of MERIS GlobCover Orthorectified Products," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, pp. 2972 - 2982, 2011.
- [11] R. Tateishia, B. Uriyangqaib, H. Al-Bilbisc, M. A. Ghard, J. Tsend-Ayushe, T. Kobayashia, A. Kasimuf, N. T. Hoana, A. Shalabyd, B. Alsaaidaha, T. Enkhzayaa, Gegentanaa and H. P. Satog, "Production of global land cover data – GLCNMO," *International Journal of Digital Earth*, vol. 4, no. 1, pp. 22-49, 2011.
- [12] R. Tateishi, N. T. Hoan, T. Kobayash, B. A. G. Tana and D. X. Phong, "Production of Global Land Cover Data – GLCNMO2008," *Journal of Geography and Geology*, vol. 6, no. 3, 2014.

- [13] J. Latham, R. Cumani, I. Rosati and M. Bloise, "Global Land Cover SHARE (GLC-SHARE)," 2014.[Online].Available:
http://www.glcn.org/downs/prj/glcshare/GLC_SHARE_beta_v1.0_2014.pdf. [Accessed 4 2016].
- [14] L. Seea, D. Schepaschenkoa and e. al, "Building a hybrid land cover map with crowdsourcing and geographically weighted regression," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 48-56, 2015.
- [15] ESA, "Land Cover CCI PRODUCT USER GUIDE V2.5," 01 26 2016. [Online]. Available:
<http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-PUG-v2.5.pdf>. [Accessed 04 2016].
- [16] J. Wanga, Y. Zhaob, C. Lic, L. Yub, D. Liud and P. Gong, "Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 38-47, 2015.
- [17] L. Yu, J. Wang, X. Li, C. Li, Y. Zhao and P. Gong, "A multi-resolution global land cover dataset through multisource data aggregation," *Science China Earth Sciences*, vol. 57, pp. 2317-2329, 2014.
- [18] J. Chena, J. Chenb, A. Liaoa, X. Caob, L. Chena, X. Chenb, C. Hea, G. Hana, S. Penga, M. Lua, W. Zhanga, X. Tongc and J. Millsd, "Global land cover mapping at 30 m resolution: A POK-based operational approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, p. 7–27, 2015.
- [19] G. Grekousis, G. Mountrakis and M. Kavouras, "An overview of 21 global and 43 regional landcover," *International Journal of Remote Sensing*, vol. 36, no. 21, pp. 5309-5335, 2015.
- [20] E. Bartholomé και A. S. Belward, «GLC2000: a new approach to global land cover mapping from Earth observation data,» *International Journal of Remote Sensing*, τόμ. 26, αρ. 9, pp. 1959-1977, 2005.
- [21] P. Bicheron, P. Defourny, C. Brockmann, L. Schouten, C. Vancutsem, M. Huc and M. S. Bontemps, "Globcover. Products Description v2.2," 2014. [Online]. Available:
<http://due.esrin.esa.int/globcover/>. [Accessed 2015].
- [22] O. Arino, P. Bicheron, F. Achard, J. Latham, R. Witt and J. L. Weber, "GLOBCOVER the Most Detailed Portrait of Earth," ESA Bulletin-European Space Agency, 2008.
- [23] S. Bontemps, P. Defourny, E. Bogaert, O. Arino, V. Kalogirou and J. Perez, "Globcover 2009 Products Description and Validation Reports v2.2," 2011. [Online]. Available:
http://due.esrin.esa.int/globcover/LandCover2009/GLOBCOVER2009_Validation_Report_2.2.pdf. [Accessed 2015].
- [24] P. Gong, J. W. L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu, S. Liu, C. Li, X. Li and e. al, "Finer Resolution Observation and Monitoring of Global Land Cover: First Mapping Results with Landsat TM and ETM+ Data," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2607-2654, 2013.

- [25] L. Martino and M. Fritz, "New Insights into Land Cover and Land Use in Europe," *Statistics in Focus* 33/2008, 2014.
- [26] LUCAS, "Technical Reference Document C-3: Landuse and Land Cover: Nomenclature," 2009.[Online].Available:
http://epp.eurostat.ec.europa.eu/portal/page/portal/lucas/documents/LUCAS2009_C3-Classification_20121004.pdf.
- [27] LUCAS, "Technical Reference Documentc-3 Land Use and Land Cover Classification (Revised),"2013.[Online].Available:
http://epp.eurostat.ec.europa.eu/portal/page/portal/lucas/documents/LUCAS2012_C3-Classification_20131004_0.pdf.
- [28] P. Defourny, S. Bontemps, E. v. Bogaert, J.-L. Weber and T. Soukup, "GLOBCORINE Validation Report,"2010.[Online].Available:
http://due.esrin.esa.int/files/p114/GLOBCORINE_VR_2.1_20100406.pdf.
- [29] P. Defourny, S. Bontemps, E. v. Bogaert, J.-L. Weber, C. Steenmans and L. Brodsky, "GLOBCORINE 2009 Description and Validation Report," 2010. [Online]. Available:
http://due.esrin.esa.int/files/p114/GLOBCORINE2009_DVR_2.1.pdf.
- [30] Z. Zhu and C. E. Woodcock, "Continuous change detection and classification of land cover using all available Landsat data," *Remote Sensing of Environment*, vol. 144, pp. 152-171, 2014.
- [31] S. Lhermitte, J. Verbesselt, W. Verstraeten and P. Coppin, "A comparison of time series similarity measures for classification and change detection of ecosystem dynamics," *Remote Sensing of Environment*, vol. 115, no. 12, pp. 3129-3152, 2011.
- [32] M. Berry, A. J. and G. S. Linoff, "Mastering data mining," *INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS*, 2000.
- [33] H. A. Edelstein, "Introduction to data mining and knowledge discovery (3rd ed)," *INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS*, 1999.
- [34] DELL, "What is Data Mining (Predictive Analytics, Big Data)," 2014. [Online]. Available:
<https://www.statsoft.com/Textbook/Data-Mining-Techniques>. [Accessed 2016].
- [35] V. Vladimir and C. Corinna, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [36] J. Nocedal and S. Wright, *Numerical Optimization*, New York: Springer-Verlag New York, 2006.
- [37] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *COLT '92 Proceedings of the fifth annual workshop on Computational learning theory*, New York, NY, USA, 1992.
- [38] K.-B. Duan and S. S. Keerthi, "Which Is the Best Multiclass SVM Method? An Empirical Study," *Multiple Classifier Systems*, vol. 3541, pp. 278-285, 2005.

- [39] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415 - 425, 2002.
- [40] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," Microsoft Research, 1998.
- [41] T. G. Dietterich, G. Bakiri and Bakiri, "Solving Multiclass Learning Problems via Error-Correcting Output Codes," *Journal of Artificial Intelligence Research*, vol. 22, p. 263–286, 1995.
- [42] L. Rokach and O. Maimon, "Data Mining with Decision Trees," *Series in Machine Perception and Artificial Intelligence*, vol. 69, 2008.
- [43] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data*, Second Edition, CRC Press Taylor & Francis Group, 2009.
- [44] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, pp. 175-185, 1992.
- [45] S. Thirumuruganathan, "A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm," 2010.[Online].Available: <https://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/>. [Accessed 2016].
- [46] B. S. Everitt, S. Landau, M. Leese and D. Stahl, *Cluster Analysis*, 5th Edition, Wiley, 2010.
- [47] C. D. Manning, P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, GAMBRIDGE Catalog, 2008.
- [48] D. M. W. Powers, "EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.
- [49] E. Li, P. Du, A. Samat, J. Xia and M. Che, "An automatic approach for urban land-cover," *International Journal of Remote Sensing*, vol. 36, no. 24, pp. 5983-6007, 2015.
- [50] B. Zhong, A. Yang, A. Nie, Y. Yao, H. Zhang, S. Wu and Q. Liu, "Finer Resolution Land-Cover Mapping Using Multiple Classifiers and Multisource Remotely Sensed Data in the Heihe River Basin," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4973 - 4992, 2015.
- [51] K. Karantzalos, D. Bliziotis and A. Karmas, "A Scalable Geospatial Web Service for Near Real-Time, High-Resolution Land Cover Mapping," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4665 - 4674, 2015.
- [52] K. Jia, S. Liang, X. Wei, Y. Yao, Y. Su, B. Jiang and X. Wang, "Land Cover Classification of Landsat Data with Phenological Features Extracted from Time Series MODIS NDVI Data," *Remote Sensing*, vol. 6, no. 11, pp. 11518-11532, 2014.