Πολυτεχνείο Κρήτης

Τμήμα Ηλεκτρονικών Μηχανικών και
Μηχανικών Υπολογιστών

“Βελτίωση της δυνατότητας εκτίμησης κινδύνου έναρξης πυρκαγιάς στα περιαστικά δάση Ηλιούπολης Αττικής και Σερρών με χρήση προηγμένων τεχνικών ομαδοποίησης στα μετεωρολογικά δεδομένα των περιοχών αυτών.”

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Λύκας Χρήστος

Επιβλέπων Καθηγητής: *Καθ. Σταυρακάκης Γεώργιος*

Εξεταστική Επιτροπή: *Καθ. Σταυρακάκης Γεώργιος*
Καθ. Ζερβάκης Μηχάλης
Επ. Καθ. Κολοκοτσά Διονυσία

ΧΑΝΙΑ 2012

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τους καθηγητές μου κ. Σταυρακάκη Γιώργο, κα. Κολοκοτσά Διονυσία και κ. Ζερβάκη Μιχάλη για την ουσιαστική καθοδήγηση και τις πολύτιμες συμβουλές τους.

Ένα μεγάλο ευχαριστώ στον Κώστα Γομπάκη για τις συμβουλές, την άψογη συνεργασία και την βοήθεια που μου προσέφερε από την αρχή ως το τέλος της εργασίας.

Περίληψη

Κάθε χρόνο, οι πυρκαγιές καίνε εκατομμύρια στρέμματα έκτασης των παγκόσμιων δασών πράγμα το οποίο οδηγεί στην υποβάθμιση των δασικών εκτάσεων με ευρύτατες συνέπειες για τη βιοποικιλότητα, την υγεία και την οικονομία. Οι άμεσες επιπτώσεις των δασικών πυρκαγιών μπορεί να είναι καταστροφικές τόσο για τις ανθρώπινες κοινότητες όσο και για τα δασικά οικοσυστήματα. Τι πρέπει λοιπόν να κάνουμε ώστε να τις αντιμετωπίσουμε;

Για την αντιμετώπισή τους προτείνεται ένα ολοκληρωμένο, επιχειρησιακό σύστημα διαχείρισης των δασικών πυρκαγιών με το οποίο θα είναι δυνατή: η εκτίμηση της επικινδυνότητας για πρόκληση πυρκαγιάς, η άμεση ανίχνευση και μεθοδολογία διαχείρισης της κρίσης καθώς και η χαρτογράφηση της καμένης έκτασης και αποτίμηση του μεγέθους της καταστροφής.

Σκοπός της διατριβής αυτής είναι η εφαρμογή και συγκριτική μελέτη αλγορίθμων ομαδοποίησης για την επιλογή του κατάλληλου αλγορίθμου και του κατάλληλου αριθμού clusters ώστε να εκτιμηθεί ο μετεωρολογικός δείκτης επικινδυνότητας ο οποίος σε συνδυασμό με τον τοπογραφικό δείκτη επικινδυνότητας εξάγουν τον τελικό δείκτη επικινδυνότητας έναρξης πυρκαγιάς μιας περιοχής την δεδομένη χρονική στιγμή. Οι παραπάνω δείκτες παρέχουν μια ένδειξη του κινδύνου πυρκαγιάς σε πέντε κατηγορίες οι οποίες κυμαίνονται από 1 (δεν υπάρχει κίνδυνος πυρκαγιάς) έως 5 (αυξημένος κίνδυνος πυρκαγιάς). Για το λόγο αυτό θα δοθεί μεγαλύτερη έμφαση για αριθμό clusters ίσο με 5. Αυτές οι πέντε κατηγορίες ορίζονται συνήθως από εμπειρογνώμονες με ένα ευρηματικό τρόπο. Στην παρούσα εργασία αναπτύχθηκαν και δοκιμάστηκαν οι προηγμένες τεχνικές ομαδοποίησης. Τα δεδομένα που χρησιμοποιήθηκαν αφορούν δύο περιοχές της Ελλάδος, την περιοχή των Σερρών καθώς και την περιοχή της Ηλιούπολης.

Οι αλγόριθμοι ομαδοποίησης που εφαρμόστηκαν ήταν οι k-Means, Fuzzy c-Means και Gaussian Mixture Model καθώς υπάρχει και μια εκτενή αναφορά στον αλγόριθμο Hierarchical. Στη συνέχεια για να βρεθεί ο βέλτιστος αλγόριθμος και η βέλτιστη κατάταξη χρησιμοποιήθηκαν οι δείκτες επικύρωσης Silhouette και Davies – Bouldin.

Abstract

Every year, fires incinerate millions of acres of land of global forest which leads to degradation of forest land with sweeping consequences for biodiversity, health and economy. The impact of forest fires can be devastating both for human communities as well as forest ecosystems. What should we do to address them?

In order to cope, a complete forest fire management system is proposed. It is able to estimate the risk of fire, detect, devise a plan of action, map the burnt area and assess the magnitude of the disaster.

The point of this thesis is to implement and comparative study clustering algorithms for selecting the appropriate algorithm and an appropriate number of clusters to estimate the meteorological risk index, which combined with the topographic risk index will export the final index of fire hazard of a given region at a given time.

These indexes provide an indication of fire risk into five categories ranging from 1 (no risk of fire) to 5 (increased risk of fire). For this reason, greater emphasis on number of clusters equal to 5 is given. These five categories are usually defined by experts empirically. In the present study advanced clustering techniques are developed and tested. The data used relate to two areas of Greece, the region of Serres and the district of Heliopolis.

The clustering algorithms applied were k-Means, Fuzzy c-Means and Gaussian Mixture Model along with an extensive report on the algorithm Hierarchical. Afterwards in order to find the optimal algorithm and the optimal classification, validation indicators Silhouette and Davies - Bouldin were used.

Περιεχόμενα

Κεφάλαιο 1-Εισαγωγή.

1.1 Το πρόβλημα των δασικών πυρκαγιών

Οι δασικές πυρκαγιές είναι πολύπλοκα φαινόμενα τα οποία λαμβάνουν χώρα ως αποτέλεσμα φυσικών και ανθρώπινων δραστηριοτήτων. Η αδυναμία ελέγχου μιας πυρκαγιάς οδηγεί συχνά σε εκτεταμένες ζημιές και για το λόγο αυτό τα φαινόμενα αυτά αποκαλούνται “φυσικές καταστροφές”. Ο όρος “φυσική καταστροφή” έχει οριστεί με τους εξής τρόπους (1):

• Τα στοιχεία εκείνα του φυσικού περιβάλλοντος που είναι βλαβερά για τον άνθρωπο και προκαλούνται από δυνάμεις ξένες και άγνωστες σε αυτόν.

• Η πιθανότητα εμφάνισης ενός δυνητικά καταστροφικού γεγονότος μέσα σε μια συγκεκριμένη χρονική περίοδο και σε μια συγκεκριμένη γεωγραφική περιοχή.

• Μια φυσική ή ανθρωπογενής γεωλογική κατάσταση ή φαινόμενο κατά την οποία παρουσιάζεται πραγματικός ή δυνητικός κίνδυνος για την ανθρώπινη ζωή ή τις περιουσίες.

Η διαχείριση κάθε φυσικής καταστροφής θέτει τη λήψη προληπτικών μέτρων ως ένα βασικό στοιχείο, ώστε, όταν προκύψει το φαινόμενο, να εκτονωθεί μέσα στα όρια των μέτρων και των υποδομών του προκατασταλτικού σχεδιασμού και η συμπεριφορά του να μην ξεπεράσει τις δυνατότητες του μηχανισμού καταστολής.

Υπάρχουν πολλοί λόγοι που τα ελληνικά δάση είναι ευάλωτα στις πυρκαγιές -τα παρατεταμένα θερμά και ξηρά καλοκαίρια, οι ήπιοι χειμώνες (χαρακτηριστικοί του Μεσογειακού κλίματος), οι δυνατοί άνεμοι, το έντονο ανάγλυφο των δασικών εδαφών και η εύφλεκτη ξηροφυτική βλάστηση. Όταν σε αυτούς τους παράγοντες προστεθεί και η έντονη ανθρώπινη δραστηριότητα, η ελλιπής διαχείριση των εύφλεκτων αυτών δασών και η επικράτηση της αντίληψης ότι η προστασία από τις δασικές πυρκαγιές ταυτίζεται με την δασοπυρόσβεση, αυξάνεται ο αριθμός των πυρκαγιών όπως και οι δασικές εκτάσεις που αυτές καταστρέφουν. Ιδιαίτερα κατατοπιστικά είναι τα στατιστικά στοιχεία που παρουσιάζονται στην αμέσως επόμενη ενότητα.

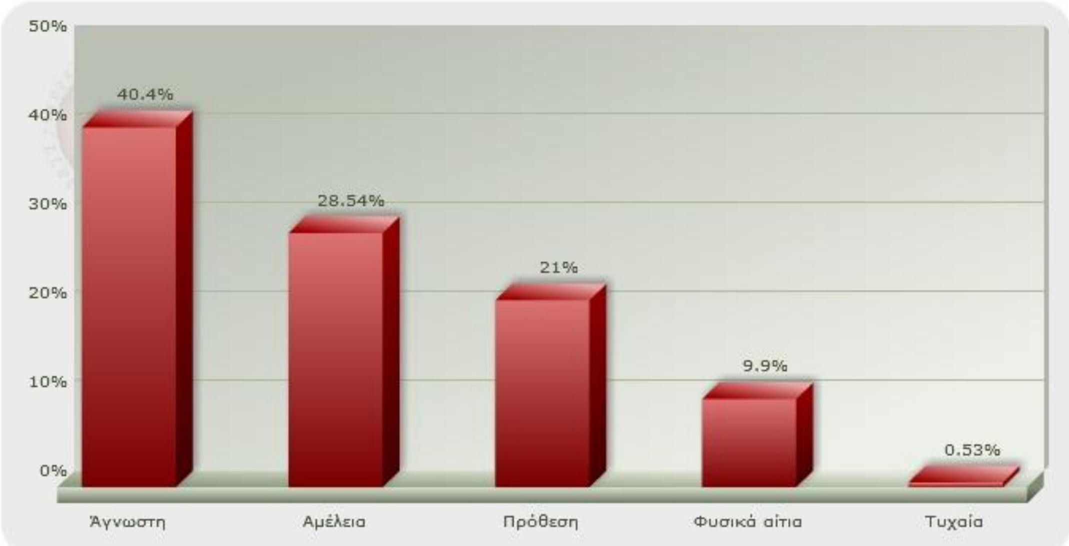
1.2 Οι πυρκαγιές στην Ελλάδα – Στατιστικά στοιχεία

Οι δασικές πυρκαγιές αποτελούν την κύρια απειλή και ένα από τα σημαντικότερα προβλήματα για τα δάση της Ελλάδας. Κάθε καλοκαίρι, οι πυρκαγιές προκαλούν τεράστιες καταστροφές στα δασικά οικοσυστήματα αφήνοντας στο πέρασμά τους στάχτες και αποκαΐδια. Υπάρχει ένα ευρύ φάσμα παραγόντων που συμβάλει στην αύξηση του κινδύνου δασικών πυρκαγιών στην Ελλάδα όπως είναι οι ψηλές θερμοκρασίες και η παρατεταμένη ξηρασία, οι δυνατοί άνεμοι, οι μεγάλες κλίσεις των δασικών εδαφών και η εύφλεκτη ξηροφυτική βλάστηση. Ένας άλλος επίσης σημαντικός παράγοντας είναι και η συσσώρευση βιομάζας, λόγω της εγκατάλειψης πολλών αγροτικών περιοχών. Ο αυξανόμενος τουρισμός και η μαζική έξοδος των κατοίκων των πόλεων για αναψυχή στα δάση, συμβάλουν επίσης σημαντικά στην αύξηση του κινδύνου πρόκλησης πυρκαγιών.

Σύμφωνα με τα στατιστικά στοιχεία, 9 στις 10 πυρκαγιές προκαλούνται από διάφορες ανθρώπινες δραστηριότητες. Οι κυριότερες αιτίες δασικών πυρκαγιών είναι οι εξής :

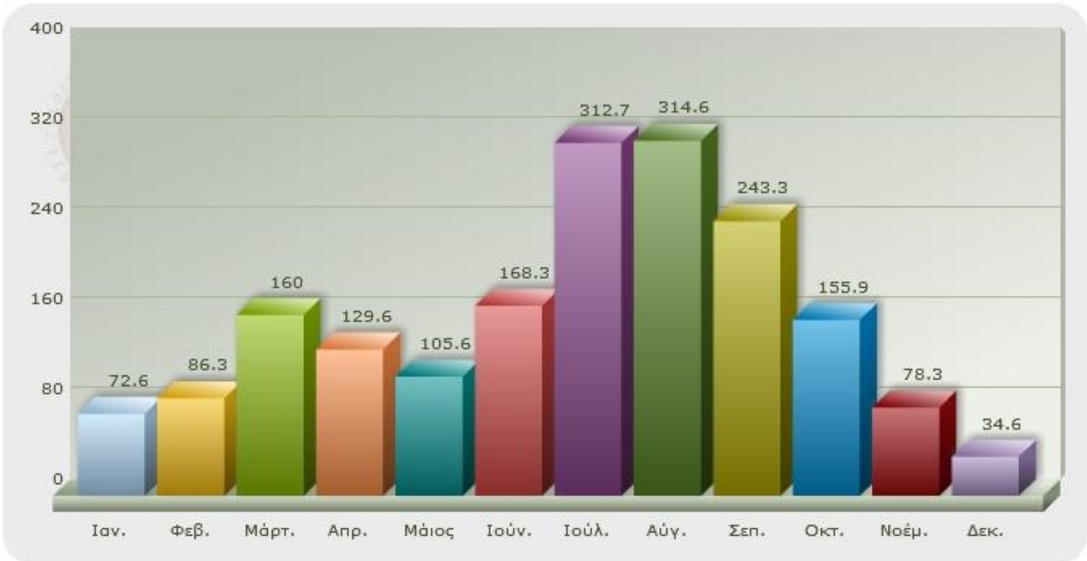
- Οι διάφορες γεωργικές δραστηριότητες και κυρίως το καψάλισμα της αποκαλάμης και το κάψιμο ξηρών χόρτων.
- Η απόρριψη αναμμένων τσιγάρων και σπίρτων.
- Το κάψιμο σκουπιδιών σε μη οργανωμένους σκυβαλοτόπους.
- Οι στρατιωτικές ασκήσεις.
- Οι δραστηριότητες κυνηγών κυρίως κατά τους θερινούς μήνες.
- Οι διάφορες δραστηριότητες σε εξοχικές κατοικίες όπως η χρήση ηλεκτρικών εργαλείων και το κάψιμο σκυβάλων.
- Οι κακόβουλες ενέργειες (εμπρησμοί).
- Βραχυκύκλωμα ηλεκτροφόρων καλωδίων.
- Οι κεραυνοί.

Όπως βλέπουμε στην εικόνα (Εικόνα 1), το 40.4 % των πυρκαγιών προέρχονται από άγνωστα αίτια, το 28.54% από αμέλεια, το 21% από πρόθεση και μόλις το 9.9% και 0.53% από φυσικά και τυχαία αίτια αντίστοιχα.



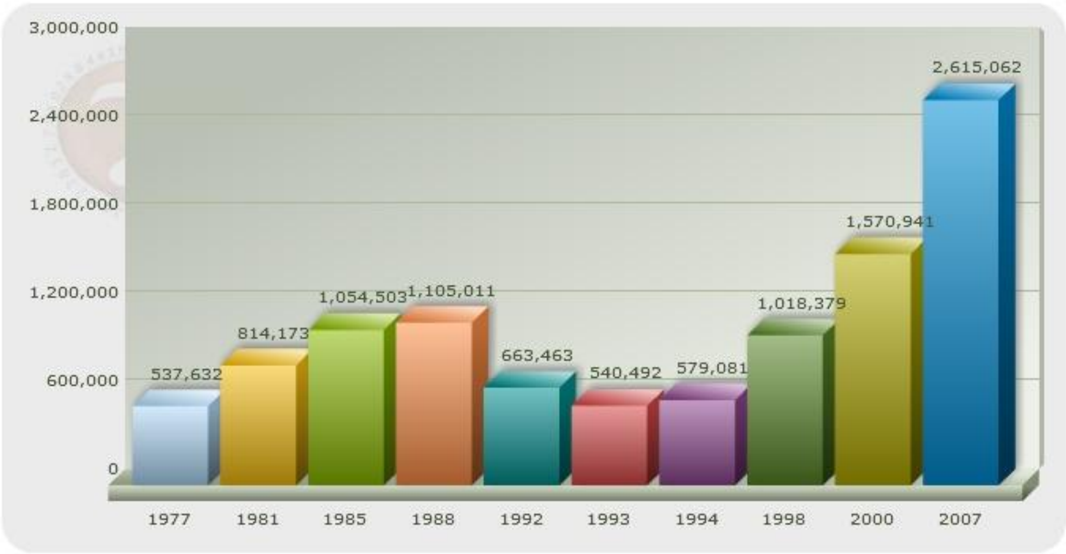
Εικόνα 1: Ανάλυση αιτιών δασικών πυρκαγιών.

Όσον αφορά τη χρονική περίοδο εμφάνισης των πυρκαγιών, σύμφωνα με την εικόνα (Εικόνα 2), ο Ιούλιος και ο Αύγουστος είναι οι δυο μήνες με την υψηλότερη συχνότητα εκδήλωσής τους και αυτό οφείλεται κυρίως στις κλιματικές συνθήκες που επικρατούν την περίοδο αυτη.



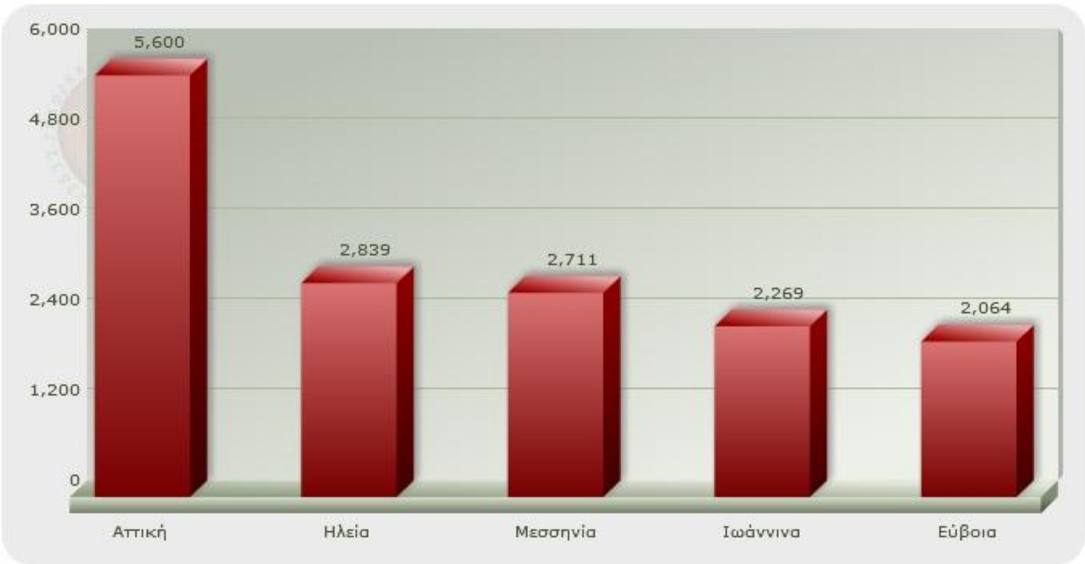
Εικόνα 2: Συχνότητα εκδήλωσης δασικών πυρκαγιών κατά μήνα τα έτη 2000-2008.

Με την πάροδο των χρόνων βλέπουμε από την εικόνα (Εικόνα 3) ότι από το 2000 και μετά οι πυρκαγιές έχουν αυξηθεί σε πολύ μεγάλο βαθμό με αποκορύφωμα το έτος 2007 όπου ο αριθμός των πυρκαγιών ανέρχεται στις 12.016 και οι καμένες δασικές εκτάσεις ανέρχονται στα 2.614.062 στρέμματα.



Εικόνα 3: Δέκα χρόνια με τις περισσότερες καμένες δασικές εκτάσεις σε στρέμματα.

Η περιοχή με τις περισσότερες εμφανίσεις πυρκαγιών κατά την περίοδο 2001-2007 όπως φαίνεται στην εικόνα (Εικόνα 4), είναι η Αττική με σχεδόν διπλάσιο αριθμό από τις υπόλοιπες περιοχές.



Εικόνα 4: Οι περισσότερες πυρκαγιές (2001-2007).

Αξιοπρόσεκτα στατιστικά στοιχεία είναι τα εξής (2) :

- 32.278 πυρκαγιές εκδηλώνονται κατά μέσο όρο ετησίως.
- 11.346 από αυτές σε δασικές εκτάσεις.
- 615.000 στρέμματα γης κατά μέσο όρο κάθε χρόνο παραδίδονται στις φλόγες.

- Ο Αύγουστος είναι ο κατ' εξοχήν μήνας των πυρκαγιών με τις περισσότερες δασικές πυρκαγιές: 314 κατά μέσο όρο από το 2000-2008 στη χώρα μας.
- 21 δασικές πυρκαγιές εκδηλώνονται κάθε μέρα στη χώρα μας, πάνω από μια για κάθε ώρα που περνά.
- 37% αύξηση σημειώθηκε το 2008 στον αριθμό των δασικών πυρκαγιών σε σχέση με το 2002.

1.3 Επιπτώσεις των πυρκαγιών στο φυσικό και ανθρωπογενές περιβάλλον

Σύμφωνα με τον Smith (3), η αντίδραση των διαφόρων ειδών πανίδας σε μια πυρκαγιά επηρεάζεται από την εποχή, τη δριμύτητα, την ένταση, το ρυθμό εξάπλωσης και το μέγεθος της πυρκαγιάς. Οι επιπτώσεις στην πανίδα συμπεριλαμβάνουν τραυματισμούς, θανατώσεις και μετανάστευση, ενώ τα άτομα που έχουν περιορισμένη κινητικότητα, όπως τα νεαρά, είναι πιο ευπαθή από τα αναπτυγμένα.

Επίσης, αρκετά σημαντικές επιδράσεις προκαλούνται στον πληθυσμό αλλά και στις κοινότητες της πανίδας (4). Έπειτα από μια πυρκαγιά παρατηρείται συνήθως αύξηση της παραγωγικότητας και της διαθεσιμότητας των νιτρικών στους βοσκότοπους. Αυτές οι αλλαγές έχουν ως αποτέλεσμα να αυξάνονται μεν σημαντικά οι πληθυσμοί των φυτοφάγων ζώων, να μετριάζεται δε η ικανότητα προσαρμογής και ανάπτυξής τους στο καινούργιο μεταπυρικό περιβάλλον.

Τα είδη που ευνοούνται κυρίως από τη φωτιά είναι τα αρπακτικά πουλιά, επειδή μειώνονται οι τρόποι κάλυψης των θηραμάτων τους (3). Επίσης, αν και ο πληθυσμός των μεγάλων σαρκοφάγων και παμφάγων ειδών μεταβάλλεται λίγο μετά από μια φωτιά, εντούτοις αναπτύσσεται εκ νέου στις πρόσφατα καμένες περιοχές λόγω της μεγαλύτερης ποσότητας θηραμάτων. Γενικά, έπειτα από τη λήξη μιας πυρκαγιάς παρατηρείται αναδιοργάνωση στις κοινότητες των ζώων -αύξηση μερικών ειδών και μείωση άλλων- η οποία σχετίζεται άμεσα με το μέγεθος των αλλαγών που έχει υποστεί η βλάστηση της περιοχής. Από την άλλη πλευρά, οι επιπτώσεις της φωτιάς στη χλωρίδα εξαρτώνται τόσο από τα χαρακτηριστικά της φωτιάς όσο και από το είδος της χλωρίδας (5). Η συμπεριφορά και η διάρκεια της πυρκαγιάς, η ποσότητα της καύσιμης ύλης που καταναλώθηκε και η θερμοκρασία που δημιουργήθηκε στο υπέδαφος επηρεάζουν σε μεγάλο βαθμό τον τραυματισμό και τη θνησιμότητα των φυτών καθώς και τη μετέπειτα ανάκαμψή τους. Η μεταπυρική αντίδραση των φυτών διαφοροποιείται ανάλογα με το είδος τους και εξαρτάται από τη φυσιολογία των χαρακτηριστικών τους αλλά και από τα προσαρμοστικά μέσα τα οποία διαθέτουν.

Η βλάστηση σε μεσογειακού τύπου οικοσυστήματα έχει αναπτύξει μηχανισμούς προσαρμογής που της επιτρέπει να επιβιώσει μετά από μια πυρκαγιά. Το κάθε είδος έχει αναπτύξει τέτοιους μηχανισμούς ώστε, είτε να μπορέσει να επιβιώσει στις υψηλές θερμοκρασίες που αναπτύσσονται, είτε να ενεργοποιήσει τους μηχανισμούς της αναπαραγωγικής διαδικασίας του (6). Για παράδειγμα, τα αείφυλλα σκληρόφυλλα ενεργοποιούν τους οφθαλμούς που βρίσκονται κάτω από το έδαφος λόγω της υψηλής θερμοκρασίας ενώ τα πεύκα αναπαράγονται λόγω των σπόρων που διασπείρονται από τους κώνους μετά από μια πυρκαγιά.

Σχετικά με τις επιπτώσεις των πυρκαγιών στον άνθρωπο, μπορούν να είναι άμεσες ή έμμεσες. Κύρια άμεση επίπτωση είναι η απειλή της ζωής τους καθώς αρκετοί είναι αυτοί που χάνουν τη ζωή τους ή τραυματίζονται προσπαθώντας να διαφύγουν από μια καιόμενη περιοχή, ενώ απώλειες υπάρχουν και στο προσωπικό που ασχολείται με την κατάσβεση. Στις έμμεσες επιπτώσεις, περιλαμβάνονται οι απώλειες στις ανθρώπινες περιουσίες ενώ η δημιουργία τοπίων καταστροφής επηρεάζει αρνητικά την ανθρώπινη ψυχολογία.

Επιπλέον, οι πυρκαγιές έχουν δυσμενείς επιπτώσεις στις ανθρώπινες δραστηριότητες γιατί συμβάλλουν στη δημιουργία καταστροφικών πλημμυρών, οι οποίες παρασύρουν ταυτόχρονα τα ορεινά εδάφη προκαλώντας διάβρωση και σταδιακή ερημοποίηση των πληγέντων περιοχών (7). Το σημαντικότερο φυσικό χαρακτηριστικό του εδάφους που επηρεάζεται από την πυρκαγιά είναι η εδαφολογική δομή του. Το αποτέλεσμα της μεταφοράς θερμότητας στο έδαφος επιφέρει αύξηση της θερμοκρασίας, η οποία επηρεάζει τις φυσικές, χημικές και βιολογικές ιδιότητές του (8).

1.4 Στόχος της παρούσας διπλωματικής

Κεντρική ιδέα της συγκεκριμένης διατριβής είναι η εκτίμηση του μετεωρολογικού δείκτη επικινδυνότητας, ο οποίος συμβάλει κατά πολύ μεγάλο ποσοστό στην έναρξη και εξάπλωση πυρκαγιάς από φυσικά κυρίως αίτια, για δυο περιοχές της Ελλάδος, την περιοχή των Σερρών και την περιοχή της Ηλιούπολης. Η εκτίμηση του δείκτη θα γίνει με την εφαρμογή αλγορίθμων ομαδοποίησης (clustering algorithms), βρίσκοντας έτσι τα όρια των clusters, τα οποία αντιστοιχούν στα όρια των πέντε κατηγοριών επικινδυνότητας έναρξης πυρκαγιάς. Στη συνέχεια, τα αποτελέσματα των αλγορίθμων ελέγχονται με κανόνες επικύρωσης ώστε να βρεθεί ο βέλτιστος αλγόριθμος καθώς γίνεται και έλεγχος των αλγορίθμων για περίπτωση clusters μικρότερη και μεγαλύτερη του πέντε ώστε να βρεθούν οι βέλτιστες ομαδοποιήσεις. Τέλος, αφού έχουν υπολογιστεί τα όρια των clusters γίνετε σύγκριση των αποτελεσμάτων με τα αποτελέσματα που υπολογίζουμε με ευρηματικό τρόπο τα οποία είναι όσο πιο κοντά γίνεται στους δείκτες επικινδυνότητας που εκδίδει με τη μορφή χάρτη η Γενική Γραμματεία Πολιτικής Προστασίας με το ήδη υπάρχον σύστημα πρόβλεψης έναρξης πυρκαγιάς το οποίο θα αναφερθεί λεπτομερώς στην επόμενη ενότητα.

Κεφάλαιο 2-Υπολογισμός μετεωρολογικού δείκτη (FDI)

2.1 Εισαγωγή

Ο όρος Κίνδυνος Δασικής Πυρκαγιάς εκφράζει το αποτέλεσμα της επίδρασης σταθερών και μεταβλητών παραγόντων, οι οποίοι καθορίζουν την έναρξη, εξάπλωση, δυσκολία ελέγχου της φωτιάς και τις απώλειες που προξενεί (9). Ως σχετικά σταθεροί στο χρόνο φυσικοί παράγοντες θεωρούνται το κλίμα, το ανάγλυφο και τα δασικά καύσιμα και ως μεταβλητοί στο χώρο και στο χρόνο, οι μετεωρολογικές παράμετροι. Για να είναι λοιπόν δυνατή η δημιουργία χαρτών στους οποίους θα παρουσιάζονται οι περιοχές με τον δείκτη επικινδυνότητας έναρξης πυρκαγιάς που επικρατεί τη δεδομένη χρονική στιγμή, πρέπει να έχει προηγηθεί η συλλογή των κατάλληλων δεδομένων για περεταίρω μελέτη τους. Η διαδικασία αυτή είναι συνήθως χρονοβόρα αλλά όταν γίνει σωστά και υπεύθυνα μπορεί να εγγυηθεί για την ποιότητα του τελικού αποτελέσματος. Έτσι, θα χρειαστεί να έχουμε στη διάθεσή μας δεδομένα όπως είναι η ταχύτητα του ανέμου, η θερμοκρασία αέρα καθώς και η σχετική υγρασία για τον υπολογισμό του μετεωρολογικού δείκτη επικινδυνότητας και δεδομένα όπως είναι η τοπογραφία της περιοχής, η καύσιμη ύλη καθώς και διάφοροι εξωτερικοί παράγοντες, όπως για παράδειγμα ύπαρξη κατοικημένων περιοχών, για τον υπολογισμό του τοπογραφικού δείκτη επικινδυνότητας. Στην συγκεκριμένη διατριβή θα ασχοληθούμε με τον υπολογισμό του μετεωρολογικού δείκτη και την χρήση τεχνικών ομαδοποίησης σ’ αυτόν για τη δημιουργία 5 κατηγοριών κινδύνου τις οποίες θα αναφέρουμε παρακάτω.

2.2 Συστήματα και Δείκτες εκτίμησης του Κινδύνου Πυρκαγιάς

Η αξιολόγηση του κινδύνου δασικής πυρκαγιάς γίνεται με τη δημιουργία και χρήση μεμονωμένων δεικτών ή με την ενοποίηση επιλεγμένων παραγόντων του κινδύνου πυρκαγιάς σε ένα ή περισσότερους ποιοτικούς ή ποσοτικούς δείκτες (9). Τα συστήματα και οι δείκτες κινδύνου χρησιμοποιούνται ως οδηγοί και βοηθήματα για τον καθορισμό μεθόδων και μέτρων διαχείρισης των δασικών πυρκαγιών σε μια περιοχή. Οι περισσότεροι από τους υπάρχοντες δείκτες κινδύνου χρησιμοποιούν ως εισροές μετεωρολογικές παραμέτρους και κάποιοι από αυτούς εκτιμούν την επίδραση που αυτές έχουν στην υγρασία των δασικών καυσίμων. Ακολουθεί μια σύντομη παρουσίαση των δυο κύριων συστημάτων που έχουν αναπτυχθεί σε Καναδά και ΗΠΑ καθώς και σημαντικών μετεωρολογικών δεικτών από την Ευρώπη (10) και την Αυστραλία.

Το Καναδικό Σύστημα Μετεωρολογικών Δεικτών Δασικής Πυρκαγιάς (Canadian Forest Fire Weather Index System, FWI) περιλαμβάνει τρεις αριθμητικούς δείκτες και αποτελεί το ένα από τα τέσσερα άλλα συστατικά-συστήματα, που χρησιμοποιεί το Καναδικό Σύστημα Εκτίμησης του Κινδύνου Δασικής Πυρκαγιάς (Canadian Forest Fire Danger Rating System, CFFDRS). Το Σύστημα FWI χρησιμοποιεί δεδομένα θερμοκρασίας αέρος, σχετικής υγρασίας, ταχύτητας ανέμου και βροχόπτωσης και με αυτά εκτιμάται η υγρασία των λεπτών καυσίμων, της φυλλάδας, του χούμου και σε πιο βαθιά στρώματα του εδάφους. Με δυο ενδιάμεσους δείκτες υπολογίζονται παράμετροι ανάφλεξης των λεπτών καυσίμων και συμπεριφοράς της φωτιάς. Εκτός από τον Καναδά, τα συστήματα FWI και CFFDRS έχουν εφαρμοστεί, σε διάφορες χώρες της νότιας Ευρώπης, σε περιφερειακό επίπεδο, στο πλαίσιο συγκριτικών μελετών με άλλους δείκτες και συστήματα. Στις ΗΠΑ, χρησιμοποιείται το Εθνικό Σύστημα Εκτίμησης Κινδύνου Δασικής Πυρκαγιάς (The National Fire Danger Rating System, NFDRS). Το σύστημα αυτό χρησιμοποιεί τρεις δείκτες για την εκτίμηση της πιθανότητας έναρξης πυρκαγιών με βάση μετεωρολογικές παραμέτρους, μοντέλα καύσιμης ύλης και ανθρωπογενείς παράγοντες. Με τους υπόλοιπους τρεις δείκτες του συστήματος εκτιμούνται παράμετροι του δυναμικού της συμπεριφοράς των πυρκαγιών. Ο συνολικός δείκτης έναρξης και αντιμετώπισης δασικής πυρκαγιάς είναι αθροιστικός δείκτης και παρέχει τον κίνδυνο έναρξης και τη δυσκολία ελέγχου μιας πυρκαγιάς. Το σύστημα NFDRS έχει χρησιμοποιηθεί μόνο του και σε συνδυασμό με το σύστημα BEHAVE, για την ανάπτυξη χαρτών, σε εθνικό επίπεδο και σε ημερήσια βάση (σε συγκεκριμένη μεσημβρινή ώρα), οι οποίοι χρησιμοποιούνται για προληπτικούς και προκατασταλτικούς σκοπούς.

Στη Γαλλία, ο δείκτης ξηρασίας του Orieux (1977), χρησιμοποιεί τη διαθέσιμη χωρητικότητα νερού στο έδαφος και τη δυναμική εξατμισοδιαπνοή. Με βάση το δείκτη αυτό και με την ταχύτητα του ανέμου καθορίζεται το επίπεδο κινδύνου πυρκαγιάς. Το 1985 ο Carrega προτείνει δείκτη κινδύνου, ο οποίος εκτιμάται με βάση το ημερήσιο απόθεμα νερού στο έδαφος, την ελάχιστη τιμή σχετικής υγρασίας, την ταχύτητα ανέμου και ένα συντελεστή, που αναφέρεται στο φαινολογικό στάδιο της βλάστησης. Στη Γαλλία, ο Wybo (1997) προτείνει νέες μεθόδους παρακολούθησης του κινδύνου πυρκαγιάς, στις οποίες η απόκτηση δεδομένων, όπως μετεωρολογικών, ανάγλυφου και των θέσεων των μέσων καταπολέμησης, γίνονται με τη χρήση GIS, ενώ για την ανάλυση και τελική παρουσίαση του κινδύνου χρησιμοποιείται ένα έμπειρο σύστημα. Στην Ιταλία αναπτύχθηκε από την ομάδα του Bovio (1984) μέθοδος εκτίμησης του κινδύνου πυρκαγιάς, με την οποία εκτιμάται η απώλεια του νερού στο έδαφος, λόγω της πραγματικής εξατμισοδιαπνοής και σε συνδυασμό με την δυναμική εξατμισοδιαπνοή. Οι μετεωρολογικές παράμετροι, που χρησιμοποιούνται, είναι οι μέσες ημερήσιες τιμές της θερμοκρασίας αέρος, σχετικής υγρασίας, ανέμου, ηλιακής ακτινοβολίας και βροχόπτωσης. Στη Σουηδία, αναπτύχθηκε ο δείκτης του Angstrom, με τον οποίο εκτιμάται ο κίνδυνος έναρξης πυρκαγιάς, με βάση τη θερμοκρασία αέρος και τη σχετική υγρασία. Η Πορτογαλική μέθοδος βασίζεται στο ρωσικό δείκτη του Nesterov (1949) και χρησιμοποιείται από το 1988 από το Μετεωρολογικό Ινστιτούτο της Πορτογαλίας. Χρησιμοποιεί δυο δείκτες, ένα ημερήσιο και ένα αθροιστικό. Κατά την εφαρμογή του Πορτογαλικού δείκτη, για τα έτη 1989-90, σε οκτώ περιφέρειες της Πορτογαλίας, επισημάνθηκε ότι οι αδυναμίες που παρουσιάστηκαν οφείλονται, κυρίως, στην έλλειψη μετεωρολογικών δεδομένων σε μερικές από τις περιοχές, λόγω του ανεπαρκούς αριθμού σταθμών σ’ αυτές. Η Ισπανική μέθοδος ICONA (1998) υπολογίζει από τη θερμοκρασία αέρος και τη σχετική υγρασία, ένα δείκτη, σχετικά, με την υγρασία των λεπτών νεκρών καυσίμων, ο οποίος συνδυάζεται με δεδομένα ταχύτητας του ανέμου για την εκτίμηση του κινδύνου. Οι Chuvieco and Salas (1996), χρησιμοποίησαν τη μέθοδο ICONA και χαρτογράφησαν τον κίνδυνο δασικής πυρκαγιάς με τη χρήση GIS, σε μια ορεινή περιοχή της κεντρικής Ισπανίας περιοχή έκτασης 60.000ha περίπου. Τα μετεωρολογικά δεδομένα που χρησιμοποίησαν προέρχονταν από 27 περίπου μετεωρολογικούς σταθμούς, εκ των οποίων οι τέσσερις μόνο βρίσκονταν μέσα στην περιοχή μελέτης, ενώ οι υπόλοιποι ήταν μακριά από αυτήν. Στην νοτιοανατολική Αυστραλία, ο κίνδυνος πυρκαγιάς καθορίζεται χρησιμοποιώντας τον McArthur δείκτη επικινδυνότητας για δάση και βοσκότοπους (11), (12) ο οποίος λαμβάνει τη μορφή κυκλικών κανόνων διαφάνειας, σύμφωνα με τον τύπο του καυσίμου υπό εξέταση. Το περιεχόμενο των μέτρων του δείκτη μπορεί να εκφραστεί με εξισώσεις (13) οι οποίες εκφράζουν τον κίνδυνο πυρκαγιάς ως εκθετική λειτουργία της θερμοκρασίας, της σχετικής υγρασίας, της βροχόπτωσης, της ταχύτητας του ανέμου, της ανάπτυξης της χλόης, της περιεκτικότητας της υγρασίας στα καύσιμα καθώς και του παράγοντα της ξηρασίας. Οι εξισώσεις αυτές παρέχουν την δυνατότητα εισαγωγής των μέτρων κινδύνου πυρκαγιάς σε σύγχρονα συστήματα ηλεκτρονικών υπολογιστών, επιτρέποντας τη σύνθεση ενός μοντέλου της συμπεριφοράς της φωτιάς και της εξάπλωσής της, μεταξύ άλλων εφαρμογών. Αυτός θα είναι και ο δείκτης που θα αναλυθεί περεταίρω στις επόμενες ενότητες και θα χρησιμοποιηθεί στην παρούσα διατριβή.

2.3 Υπολογισμός Μετεωρολογικού δείκτη (FDI) .

Η ταχύτητα του ανέμου είναι ο πιο κρίσιμος παράγοντας που επηρεάζει την έναρξη πυρκαγιάς (14) καθώς και ένα απ’ τα κύρια στοιχεία που καθορίζουν τον ρυθμό εξάπλωσης και την κατεύθυνσή της. Επιπλέον, βοηθά στην ανάφλεξη καθώς οι φλόγες φτάνουν πιο κοντά στην άκαυστη καύσιμη ύλη, εφοδιάζοντας την φλόγα με οξυγόνο διώχνοντας τον αέρα με υγρασία που σε άλλη περίπτωση θα περιόριζε τη ποσότητα θερμότητας που αναφλέγει την άκαυστη καύσιμη ύλη.

Η περιεκτικότητα σε υγρασία στην καύσιμη ύλη επηρεάζει επίσης την έναρξη πυρκαγιάς. Είναι ένας συνδυασμός από μια μακροπρόθεσμη συνιστώσα ξηρότητας και μιας συνιστώσας ξηρότητας που διέπεται από τον ατμοσφαιρικό αέρα. Η μακροχρόνια ξηρασία της καύσιμης ύλης επηρεάζει γενικά την ευφλεκτότητά της. Πρόκειται ουσιαστικά για ένα μέτρο του ποσοστού της καύσιμης ύλης που χαρακτηρίζεται ως εύφλεκτη. Η υγρασία του αέρα του περιβάλλοντος επηρεάζει επίσης την ευφλεκτότητα της καύσιμης ύλης, δεδομένου ότι είναι ένα μέτρο της αποτελεσματικότητας της καύσης της.

Άλλοι παράγοντες, όπως η θερμοκρασία και η σχετική υγρασία του περιβάλλοντος έχουν μόνο έμμεσες επιπτώσεις, μέσω της επιρροής τους στους παράγοντες που περιγράψαμε παραπάνω. Είναι επίσης σημαντικό να σημειωθεί ότι στη συγκεκριμένη προσπάθεια υπολογισμού του μετεωρολογικού δείκτη επικινδυνότητας, δεν λαμβάνονται υπόψη παράγοντες όπως το έδαφος και το είδος της καύσιμης ύλης.

Διαισθητικά λοιπόν, ο κίνδυνος έναρξης πυρκαγιάς θα πρέπει να αυξάνεται καθώς αυξάνεται η ταχύτητα του ανέμου και θα πρέπει να μειώνεται καθώς αυξάνεται η περιεκτικότητα σε υγρασία στην καύσιμη ύλη. Έτσι, σε γενικές γραμμές, ο μετεωρολογικός δείκτης επικινδυνότητας μπορεί να περιγραφεί από τον ακόλουθο τύπο:

$$FDI \approx \frac{U_f}{FMC} \quad (1)$$

όπου U_f είναι η ταχύτητα του ανέμου και *FMC* είναι η περιεκτικότητα σε υγρασία της καύσιμης ύλης.

2.3.1 *Υπολογισμός της ταχύτητας του ανέμου*

Μια πυρκαγιά επηρεάζεται άμεσα από τον άνεμο. Συνήθως το πιο σχετικό μέρος του ανέμου είναι σε ύψος 2 μέτρα σε αντίθεση με το μετεωρολογικό άνεμο ο οποίος μετράται στα 10 μέτρα με ανεμπόδιστη ροή. Ο άνεμος στο μέτωπο της φωτιάς (U_f) μπορεί να εκτιμηθεί από τον μετεωρολογικό άνεμο με τη βοήθεια ενός συντελεστή μείωσης (γ), ο οποίος κυμαίνεται από 0 ως 1,

$$U_f = \gamma U_{10} \quad (2)$$

όπου U_{10} είναι ο μετεωρολογικός άνεμος. Όλες οι μετρήσεις του ανέμου έχουν μονάδες χιλιόμετρα ανά ώρα. Ο συντελεστής μείωσης (γ) είναι ένας όρος που περιγράφει την εξασθένηση των ανέμων από τη βλάστηση. Τυπικά, για ένα δάσος παίρνει τιμή ίση με $\gamma \approx 0.3$, δηλαδή το 70% της ταχύτητας του ανέμου είναι εξασθενημένο από τα διάφορα φυσικά εμπόδια που συναντά και δεν συμβάλει στην έναρξη της πυρκαγιάς. Τέλος, για να καθορίσουμε τον κίνδυνο πυρκαγιάς, η ταχύτητα του ανέμου στο μέτωπο της φωτιάς (U_f) δεν πρέπει να πέφτει κάτω από μια ελάχιστη τιμή U_b η οποία αντιστοιχεί στην ταχύτητα διάδοσης της φωτιάς υποστήριξης.

2.3.2 *Υπολογισμός της περιεκτικότητας σε υγρασία της καύσιμης ύλης*

Η περιεκτικότητα σε υγρασία της καύσιμης ύλης εκφράζεται σε εκατοστημόρια. Το βάρος της υγρασίας της καύσιμης ύλης εκφράζεται ως ποσοστό επί του βάρους μετά από ξήρανση. Για παράδειγμα, αν ένα δείγμα ενός κιλού καύσιμης ύλης ξηραίνεται και ζυγίζει 900g, τότε η περιεκτικότητά του σε υγρασία είναι 11%, και επηρεάζεται από τρείς παράγοντες, τον αέρα, το έδαφος και τη βροχόπτωση. Η επίδραση του ατμοσφαιρικού αέρα για το ποσοστό υγρασίας στην καύσιμη ύλη μπορεί να προσεγγιστεί με την ακόλουθη απλή σχέση :

$$FMC = 10 - \frac{(T-H)}{4} \quad (3)$$

όπου T είναι η θερμοκρασία του αέρα (C^o) και H είναι η σχετική υγρασία (%). Η παραπάνω εξίσωση υποδηλώνει ότι ο θερμότερος και ξηρότερος αέρας μειώνουν τη περιεκτικότητα σε υγρασία (*FCM*). Σημειώνουμε ότι η παραπάνω εξίσωση είναι μια προσέγγιση με βάση τα αποτελέσματα παλινδρόμησης που βρέθηκαν από τον Pook(1993) (15) και Pook and Gill(1993) (16). Οι συντελεστές παλινδρόμησης έχουν στρογγυλοποιηθεί για ευκολία.

2.4 Συλλογή Δεδομένων

Όπως αναφέραμε στην προηγούμενη ενότητα, οι περιοχές μελέτης μας είναι δυο, η περιοχή των Σερρών και η περιοχή της Ηλιούπολης. Για την συλλογή των απαραίτητων δεδομένων ακολουθήθηκε διαφορετική μέθοδος για την κάθε περιοχή. Πιο συγκεκριμένα, για την περιοχή των Σερρών τα δεδομένα συλλέχθηκαν από την Εθνική Μετεωρολογική Υπηρεσία (Ε.Μ.Υ) (17). Έτσι έχουμε στη διάθεσή μας την θερμοκρασία του αέρα, την ταχύτητα του ανέμου καθώς και τη σχετική υγρασία της συγκεκριμένης περιοχής από τη 1η Ιουνίου του 2005 ως τη 31η Οκτωβρίου του 2011 με μετρήσεις οι οποίες πραγματοποιήθηκαν κάθε μια ώρα. Μια χαρακτηριστική αναπαράσταση των δεδομένων φαίνεται στον παρακάτω πίνακα (Πίνακας 1) όπου παρουσιάζονται οι μέγιστες και οι ελάχιστες τιμές καθώς και η μέση τιμή τους. Αντίστοιχα για την περιοχή τις Ηλιούπολης, οι παρατηρήσεις μας είναι από τη 2^η Ιουνίου του 2005 ως τη 4^η Οκτωβρίου του 2011 ανά μια ώρα και αυτές, και συλλέχθηκαν από ένα συγκεκριμένο σταθμό ο οποίος βρίσκεται στο δήμο Ηλιούπολης. Ο σταθμός είναι εγκατεστημένος στους πρόποδες του βουνού Υμηττού, σε υψόμετρο εδάφους 206 μέτρων και είναι ενεργός από τις 20/05/2005 υπό την εποπτεία του Εθνικού Μετσόβιου Πολυτεχνείου. Τα αποτελέσματα των μετρήσεων καθώς και αναλυτικότερες πληροφορίες σχετικά με το σταθμό βρίσκονται στο χώρο του διαδικτύου και πιο συγκεκριμένα στη διεύθυνση http://hoa.ntua.gr/stations/d/8/ . Στον πίνακα (Πίνακας 2) βλέπουμε μια αναπαράσταση των δεδομένων παρόμοια με αυτή της περιοχής των Σερρών.

Πίνακας 1: Μετεωρολογικά δεδομένα περιοχής Σερρών.

Μετεωρολογικά δεδομένα περιοχή Σερρών			
Τιμές δεδομένων	*Θερμοκρασία (C^o)*	*Ταχύτητα Ανέμου km/h*	*Υγρασία (%)*
Ελάχιστη	6.4	0	14
Μέγιστη	40.3	38.6	91
Μέση	23.94	5.19	55.14

Πίνακας 2: Μετεωρολογικά δεδομένα περιοχής Ηλιούπολης.

Μετεωρολογικά δεδομένα περιοχή Ηλιούπολης			
Τιμές δεδομένων	*Θερμοκρασία (C^o)*	*Ταχύτητα Ανέμου km/h*	*Υγρασία (%)*
Ελάχιστη	7.71	0	9.43
Μέγιστη	41.91	36.18	97.90
Μέση	24.94	8.12	48.99

Με μια πρώτη ματιά στους παραπάνω πίνακες μπορούμε να διαπιστώσουμε ότι η περιοχή των Σερρών αναμένεται να έχει μικρότερο κίνδυνο έναρξης πυρκαγιάς καθώς έχει χαμηλότερη μέση θερμοκρασία (ένα βαθμό κελσίου διαφορά) από την περιοχή της Ηλιούπολης και η μέση ταχύτητα ανέμου είναι αρκετά μικρότερη από αυτή της Ηλιούπολης (περίπου 3 km/h διαφορά). Επίσης η μέση τιμή του ποσοστού υγρασίας της περιοχής των Σερρών είναι αρκετά υψηλότερη σε σχέση με αυτή της Ηλιούπολης (περίπου 6 ποσοστιαίες μονάδες).

2.5 Δείκτης επικινδυνότητας Πολιτικής Προστασίας

Στα πλαίσια των προσπαθειών για την αποτελεσματικότερη αντιμετώπιση των δασικών πυρκαγιών και σε εφαρμογή του Γενικού Σχεδίου Πολιτικής Προστασίας με τη συνθηματική λέξη "Ξενοκράτης" καθώς και του εγκεκριμένου "Γενικού Σχεδίου Αντιμετώπισης Εκτάκτων Αναγκών εξαιτίας δασικών πυρκαγιών", η Γενική Γραμματεία Πολιτικής Προστασίας του Υπουργείου Προστασίας του Πολίτη (ΓΓΠΠ) εκδίδει ημερήσιο δελτίο πρόβλεψης κινδύνου πυρκαγιών υπό μορφή θεματικού χάρτη ο οποίος φαίνεται στην εικόνα (Εικόνα 5). Η πρόγνωση από τη ΓΓΠΠ βασίζεται σε μετεωρολογικές προβλέψεις διαφόρων ελληνικών μοντέλων, ενώ η κατάσταση της βλάστησης εκτιμάται μέσω του κανονικοποιημένου δείκτη βλάστησης NDVI (18).

Εικόνα 5: Χάρτης πρόβλεψης κινδύνου πυρκαγιάς

Πιο αναλυτικά, ο χάρτης παρουσιάζεται με 5 διαφορετικές κατηγορίες κινδύνου όπου για κάθε κατηγορία λαμβάνονται και διαφορετικές αποφάσεις προκατασταλτικών ενεργειών, οι οποίες αναφέρονται αναλυτικά παρακάτω σύμφωνα με την ιστοσελίδα της Γενικής Γραμματείας Πολιτικής Προστασίας (ΓΓΠΠ 2012) :

- **Κατηγορία Κινδύνου 1 (Χαμηλή):** Ο κίνδυνος είναι χαμηλός. Η πιθανότητα για εκδήλωση πυρκαγιάς δεν είναι ιδιαίτερα υψηλή. Εάν εκδηλωθεί πυρκαγιά, οι συνθήκες (κατάσταση καύσιμης ύλης, μετεωρολογικές συνθήκες) δεν θα ευνοήσουν τη γρήγορη εξέλιξή της.

- **Κατηγορία Κινδύνου 2 (Μέση):** Ο κίνδυνος είναι συνήθης για τη θερινή περίοδο. Πυρκαγιές που ενδέχεται να εκδηλωθούν, αναμένεται να είναι μέσης δυσκολίας στην αντιμετώπισή τους και εφόσον υπάρξει αποτελεσματική αντίδραση σε κάθε εκδηλωμένη πυρκαγιά δεν πρέπει να υπάρχουν προβλήματα .Ο δασοπυροσβεστικός μηχανισμός πρέπει να είναι στην κανονική του, για την αντιπυρική περίοδο, καθημερινή στελέχωση και ετοιμότητα .
- **Κατηγορία Κινδύνου 3 (Υψηλή):** Ο κίνδυνος είναι υψηλός. Είναι πιθανό να εκδηλωθεί αυξημένος αριθμός πυρκαγιών μέσης δυσκολίας αρκετές από τις οποίες θα είναι δύσκολο να αντιμετωπισθούν .Και στις δύο περιπτώσεις είναι απαραίτητο να καταβληθεί κάθε προσπάθεια για την άμεση κινητοποίηση του μηχανισμού σε κάθε επεισόδιο, την αποφυγή οποιασδήποτε χρονοτριβής και την αποστολή επαρκών δυνάμεων για να ολοκληρωθεί γρήγορα το έργο της κατάσβεσης δεδομένης της υπάρχουσας δυσκολίας. Με καλή οργάνωση και ετοιμότητα οι δυνάμεις του Πυροσβεστικού Σώματος αναμένεται να αντιμετωπίσουν τις δυσκολίες με επιτυχία .
- **Κατηγορία Κινδύνου 4 (Πολύ Υψηλή):** Ο κίνδυνος είναι ιδιαίτερα υψηλός. Ο αριθμός των πυρκαγιών που αναμένεται να εκδηλωθούν, πιθανόν να είναι μεγάλος αλλά, το κυριότερο, κάθε πυρκαγιά μπορεί να λάβει μεγάλες διαστάσεις εφόσον ξεφύγει από την αρχική προσβολή. Απαιτείται απόλυτη ετοιμότητα και πλήρης στελέχωση των δυνάμεων καταστολής και παραμονή του προσωπικού σε επαφή με τις υπηρεσίες. Όλες οι εμπλεκόμενες υπηρεσίες (νομαρχίες, ΟΤΑ, Δασικές Υπηρεσίες ,κτλ) τίθενται σε επιφυλακή και δρουν σε αρωγή του έργου του Πυροσβεστικού Σώματος σύμφωνα με τα προβλεπόμενα στα αντιπυρικά σχέδια για την κατάσταση κινδύνου επιπέδου 4 (π.χ. αυξάνονται οι περιπολίες πρόληψης και η αστυνόμευση των κρίσιμων δασικών εκτάσεων κλπ.)
- **Κατηγορία Κινδύνου 5 (Κατάσταση ΣΥΝΑΓΕΡΜΟΥ):** Ο κίνδυνος είναι ακραίος. Οι συνθήκες (ισχυρός άνεμος, χαμηλή σχετική υγρασία, υψηλή σχετικά θερμοκρασία, κλπ.) είναι πιθανό να οδηγήσουν σε ανεξέλεγκτη κατάσταση με μεγάλο αριθμό πυρκαγιών ακραίας συμπεριφοράς. Η δυσκολία ελέγχου αναμένεται να είναι πολύ μεγάλη μέχρι να μεταβληθούν οι συνθήκες κάτω από τις οποίες αναπτύσσονται οι πυρκαγιές. Επειδή αυτές οι συνθήκες είναι δυνατό να ξεπεράσουν τις συνολικές δυνατότητες του δασοπυροσβεστικού μηχανισμού, απαιτείται να μεγιστοποιηθούν οι προσπάθειες πρόληψης και η ετοιμότητα του μηχανισμού για άμεση επέμβαση με όλες τις δυνάμεις. Για την περίπτωση αυτή πρέπει να υπάρχει στο αντιπυρικό σχέδιο αλλά και σε κάθε εμπλεκόμενο στις πυρκαγιές φορέα,

κατάλληλος σχεδιασμός ετοιμότητας (αύξηση επιφυλακής προσωπικού, διατήρηση επαφής με όλο το προσωπικό και έλεγχοι ετοιμότητας, επίγειες και εναέριες περιπολίες, μηνύματα ενημέρωσης των πολιτών, πλήρης κινητοποίηση Νομαρχιών, στενή παρακολούθηση της κατάστασης με τη Ε.Μ.Υ. κλπ.) και συντονισμού με τους άλλους φορείς. Επίσης πρέπει να δίδονται ιδιαίτερες οδηγίες στο προσωπικό.

Ο χάρτης (Εικόνα 5) παρουσιάζει τις παραπάνω κατηγορίες με διαφορετικό χρώμα τη κάθε μια. Έτσι για κατηγορία κινδύνου 1 (Χαμηλή) η περιοχή παρουσιάζεται με πράσινο χρώμα, για κατηγορία κινδύνου 2 (Μέση) η περιοχή παρουσιάζεται με μωβ χρώμα, για κατηγορία κινδύνου 3 (Υψηλή) η περιοχή παρουσιάζεται με κίτρινο χρώμα, για κατηγορία κινδύνου 4 (Πολύ Υψηλή) η περιοχή παρουσιάζεται με πορτοκαλί χρώμα και τέλος για κατηγορία κινδύνου 5 (Κατάσταση ΣΥΝΑΓΕΡΜΟΥ) η περιοχή παρουσιάζεται με κόκκινο χρώμα.

Έχοντας στην διάθεσή μας όλους τους χάρτες πρόβλεψης κινδύνου πυρκαγιάς για τις χρονικές στιγμές όπου έχουμε και τις μετρήσεις των καιρικών συνθηκών όπως αναφέραμε στην προηγούμενη ενότητα, εξάγαμε τις τιμές του δείκτη επικινδυνότητας που έχει εκδώσει η Γενική Γραμματεία Πολιτικής Προστασίας με την ακόλουθη διαδικασία:

- Με τη χρήση του ***imtool*** εργαλείου του Matlab βρίσκουμε τις συντεταγμένες (x,y) των περιοχών του χάρτη που μας ενδιαφέρουν (περιοχή Σερρών και περιοχή Ηλιούπολης).
- Εισάγοντας τις περιοχές της εικόνας που θέλουμε σε ενα script σε περιβάλλον Matlab, ελέγχουμε το κάθε pixel της.
- Εξάγουμε μια τιμή από 1 έως 5, ανάλογα με το τι χρώμα έχει το κάθε pixel ελέγχοντας το RGB του.

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στους πίνακες (Πίνακας 3)(Πίνακας 4).

Πίνακας 3: Δείκτης επικινδυνότητας Πολιτικής Προστασίας για την περιοχή των Σερρών.

Δείκτης Επικινδυνότητας Πολιτικής Προστασίας περιοχής Σερρών		
Κλίμακα	*Αριθμός Παρατηρήσεων*	*Ποσοστό(%)*
1	15695	61.06
2	9601	37.35
3	384	1.49
4	24	0.09
5	0	0.00

Πίνακας 4: Δείκτης επικινδυνότητας Πολιτικής Προστασίας για την περιοχή της Ηλιούπολης.

Δείκτης Επικινδυνότητας Πολιτικής Προστασίας περιοχής Ηλιούπολης		
Κλίμακα	*Αριθμός Παρατηρήσεων*	*Ποσοστό(%)*
1	6072	23.56
2	9432	36.59
3	6648	25.79
4	3624	14.06
5	0	0.00

2.6 Συμπεράσματα κεφαλαίου

Από την περιγραφή των συστημάτων και των δεικτών κινδύνου φαίνεται πόσο σημαντική είναι η ύπαρξη αξιόπιστων και σε συνεχή βάση μετεωρολογικών και κλιματικών δεδομένων που να μπορούν να χρησιμοποιηθούν για την προστασία των δασικών και γενικά των ορεινών περιοχών από τις πυρκαγιές, αλλά και για άλλες μελέτες και εφαρμογές.

Για τις ανάγκες της συγκεκριμένης διατριβής επιλέχθηκε ο μετεωρολογικός δείκτης επικινδυνότητας FDI ο οποίος χρησιμοποιείτε στην νοτιοανατολική Αυστραλία. Τα κριτήρια που αξιολογίθηκαν για την επιλογή του συγκεκριμένου δείκτη είναι τα εξής:

- Ο FDI είναι δυναμικά υπολογίσιμος κάθε ώρα καθώς εξάγεται από μετεωρολογικά δεδομένα για τα οποία μπορούμε να έχουμε μετρήσεις κάθε μια ώρα, γεγονός που τον κάνει ιδιαίτερα ευέλικτο και παραμετροποιήσιμο σε περιοχές με έντονες κλιματολογικές αλλαγές.
- Δεν απαιτεί ιδιαίτερα μετεωρολογικά δεδομένα για να λειτουργίσει.
- Είναι εξίσου ακριβής σε σχέσει με άλλους δείκτες.

Μια πρώτη εκτίμηση των αποτελεσμάτων του μετεωρολογικού δείκτη αναμένεται να συμβαδίζει με την ΓΓΠΠ, σύμφωνα με την οποία ο κίνδυνος εκδήλωσης πυρκαγιάς για την περιοχή των Σερρών δεν είναι μεγάλος καθώς μόλις το 1.49% του δείκτη επικινδυνότητας που εκδόθηκε τη συγκεκριμένη χρονική περίοδο ανήκει στην 3η κατηγορία (Υψηλή) και το 0.09% ανήκει στην 4η κατηγορία (Πολύ Υψηλή). Σε αντίθετη περίπτωση, στην περιοχή της Ηλιούπολης ο κίνδυνος πυρκαγιάς αποτελεί μεγαλύτερο πρόβλημα καθώς το 25.79% ανήκει στην 3η κατηγορία και το 14.06% στην 4η κατηγορία, ποσοστά αρκετά υψηλότερα σε σχέση με την περιοχή των Σερρών πράγμα το οποίο είναι αναμενόμενο καθώς η περιοχή της Αττικής είναι με διαφορά η περιοχή με τις περισσότερες εμφανίσεις πυρκαγιών σύμφωνα με τα στατιστικά στοιχεία της εικόνας (Εικόνα 4) καθώς και έχει και ευνοϊκότερες συνθήκες λόγο μετεωρολογικών δεδομένων σύμφωνα με τους πινάκες (Πίνακας 1) και (Πίνακας 2). Με την χρήση προηγμένων τεχνικών ομαδοποίησης για τις οποίες θα αναφερθούμε στο αμέσως επόμενο κεφάλαιο, θα κατηγοριοποιήσουμε τον μετεωρολογικό δείκτη FDI ώστε να έχουμε μια εμπεριστατομένη εκτίμηση του κινδύνου έναρξης πυρκαγιάς στις δυο αυτές περιοχές.

Κεφάλαιο 3-Ανάλυση αλγορίθμων ομαδοποίησης

3.1 Εισαγωγή:

Η «Μέθοδος Ομαδοποίησης Δεδομένων» (Data Clustering) είναι μια διαδικασία κατά την οποία παρόμοια δεδομένα τοποθετούνται σε ομάδες. Ένας αλγόριθμος ομαδοποίησης χωρίζει ένα σύνολο δεδομένων σε διάφορες ομάδες δεδομένου ότι η ομοιότητα των στοιχείων του σε μια ομάδα είναι μεγαλύτερη από τις υπόλοιπες ομάδες. Η ιδέα της μεθόδου ομαδοποίησης (clustering) είναι απλή και είναι κοντά στον ανθρώπινο τρόπο σκέψης. Όταν για παράδειγμα έχουμε να διαχειριστούμε μεγάλο όγκο δεδομένων, συνοψίζουμε αυτό το τεράστιο αριθμό δεδομένων σε ένα μικρό αριθμό ομάδων ή κατηγοριών προκειμένου να διευκολυνθεί η περαιτέρω ανάλυσή τους. Επιπλέον, τα περισσότερα στοιχεία που συλλέγονται σε πολλά προβλήματα, έχουν ορισμένες έμφυτες ιδιότητες που προσφέρονται για φυσικές ενώσεις. Παρ' όλα αυτά, η εξεύρεση αυτών των ομάδων ή η προσπάθεια να κατηγοριοποιήσουμε τα δεδομένα δεν είναι μια απλή διαδικασία για τον άνθρωπο, εκτός εάν τα στοιχεία είναι χαμηλής διαστατικότητας (δύο η τριών διαστάσεων το περισσότερο). Έτσι για την επίλυση του συγκεκριμένου προβλήματος προτείνονται ορισμένοι μέθοδοι εν ονόματι «Μέθοδοι Ομαδοποίησης Δεδομένων» (Data Clustering Methods). Στην συγκεκριμένη διατριβή οι μέθοδοι ομαδοποίησης δεδομένων καλούνται να ομαδοποιήσουν τον μετεωρολογικό δείκτη επικινδυνότητας (FDI), ο οποίος αν και έχει χαμηλή διαστατικότητα (μονοδιάστατος πίνακας) αποτελείτε από ένα μεγάλο αριθμό δεδομένων.

3.2 Ορισμός Ομάδας (Cluster).

Ομάδα είναι μια συλλογή από όμοια αντικείμενα που έχουν μαζευτεί ή συγκεντρωθεί μαζί. Ο Everitt (19) δίνει μεταξύ άλλων τους εξής ορισμούς της Ομάδας :

- Ομάδα είναι το σύνολο των οντοτήτων που είναι όμοιες, και οντότητες από διαφορετικά σύνολα δεν είναι όμοιες.
- Ομάδα είναι μια συσσωμάτωση σημείων του πεδίου αναφοράς τέτοια ώστε η απόσταση κάθε δύο σημείων της Ομάδας να είναι μικρότερη από την απόσταση μεταξύ κάθε σημείου της Ομάδας και οποιουδήποτε άλλου σημείου.
- Οι ομάδες μπορούν να περιγραφούν ως συνδεδεμένες περιοχές ενός πολυδιάστατου χώρου, που περιέχουν σχετικά υψηλή πυκνότητα σημείων και

που είναι χωρισμένες μεταξύ τους με περιοχές, η πυκνότητα των οποίων είναι σχετικά μικρότερη.

Ο ορισμός της ομάδας ή αλλιώς cluster που δίνετε από τους Θεοδωρίδη και Κουτρουμπά (20) είναι ο εξής :

Ας θεωρηθεί ότι N είναι το πλήθος των αντικειμένων x του συνόλου δεδομένων X ,δηλαδή,

$$X=\{x_i,i=1,\ldots,N\}$$

Επιπλέον, αν συμβολιστεί με R η διαμέριση του X σε m ομάδες , C_j , j=1,…,m, τότε αυτές οι ομάδες ονομάζονται clusters και πρέπει να ικανοποιούν τρεις συνθήκες :

- $C_i \neq \emptyset, i = 1, \ldots, m$
- $\bigcup_{i=1}^{m} C_i = X$
- $C_i \cap C_j = \emptyset, i \neq j, i, j = 1, \ldots, m$

3.3 Διαδικασία Ομαδοποίησης.

Η διαδικασία για την εύρεση ομάδων που μπορεί να υπάρχουν στο σύνολο των δεδομένων περιλαμβάνει τέσσερα διακριτά βήματα, τα οποία περιγράφονται συνοπτικά παρακάτω (21):

i. ***Επιλογή και Εξαγωγή Χαρακτηριστικών*****:** Σε αυτό το βήμα γίνεται επιλογή ενός υποσυνόλου διακριτών χαρακτηριστικών, από το σύνολο των χαρακτηριστικών που περιγράφουν τα αντικείμενα, ή δημιουργούνται νέα χαρακτηριστικά από τα αρχικά. Γενικά, τα χαρακτηριστικά που θα χρησιμοποιηθούν θα πρέπει να συντελούν στην διάκριση των αντικειμένων που ανήκουν σε διαφορετικές ομάδες, να έχουν ανοχή στον θόρυβο, να κωδικοποιούν με τον καλύτερο δυνατό τρόπο την πληροφορία και να είναι εύκολο να εξαχθούν και να αναπαρασταθούν.
ii. ***Επιλογή του αλγορίθμου ομαδοποίησης*****:** Σε αυτό το βήμα, επιλέγεται ποιο θα είναι το μέτρο που θα δηλώνει το πόσο όμοια είναι δύο αντικείμενα μεταξύ τους ώστε να μπορούν να ομαδοποιηθούν στο ίδιο cluster. Στη συνέχεια, επιλέγεται ένας αλγόριθμος ομαδοποίησης. Όπως αναφέρθηκε και προηγουμένως, υπάρχουν πολλοί αλγόριθμοι ομαδοποίησης αλλά δεν υπάρχει κάποιος που να μπορεί να χρησιμοποιηθεί για να δώσει λύση σε όλα τα προβλήματα. Γι’ αυτό είναι σημαντικό να μελετηθεί σε βάθος το είδος των

χαρακτηριστικών που περιγράφουν τα αντικείμενα καθώς και η φύση του προβλήματος ώστε να επιλεχθεί ο κατάλληλος αλγόριθμος ομαδοποίησης.

iii. ***Επικύρωση των αποτελεσμάτων*:** Κάθε αλγόριθμος ομαδοποίησης μπορεί να παράγει μία διαμέριση σε ένα σύνολο δεδομένων είτε αυτή υφίσταται είτε όχι. Επιπλέον, διαφορετικοί αλγόριθμοι δίνουν διαφορετικές διαμερίσεις για το ίδιο σύνολο δεδομένων. Για τον λόγο αυτό, θα πρέπει τα αποτελέσματα των αλγορίθμων να αξιολογηθούν με βάση κάποια κριτήρια που θα είναι αντικειμενικά για όλους τους αλγορίθμους.

iv. ***Απεικόνιση των αποτελεσμάτων*:** Τέλος, αφού οι αλγόριθμοι ομαδοποίησης εφαρμόστηκαν και αξιολογήθηκαν μπορούν τα αποτελέσματα να απεικονιστούν με μία μορφή που να είναι κατάλληλη και να έχει νόημα για το συγκεκριμένο πρόβλημα.

Θα πρέπει να τονιστεί ότι η διαδικασία που περιγράφηκε μπορεί να είναι αναδρομική με την έννοια ότι αν σε οποιοδήποτε βήμα δεν προκύπτουν ικανοποιητικά αποτελέσματα μπορεί να γίνει μετάβαση σε κάποιο προηγούμενο και είτε για παράδειγμα να επιλεχθούν άλλα χαρακτηριστικά είτε να εφαρμοστεί κάποιος άλλος αλγόριθμος ομαδοποίησης είτε διαφορετικές τεχνικές επικύρωσης των αποτελεσμάτων. Στη συνέχεια, αναλύονται οι σημαντικότερες έννοιες που αφορούν την διαδικασία της ομαδοποίησης, αναλύεται ο τρόπος με τον οποίο λειτουργούν τέσσερις σημαντικοί αλγόριθμοι ομαδοποίησης καθώς και οι τεχνικές με τις οποίες μπορούν να αξιολογηθούν τα αποτελέσματά τους.

3.4 Αλγόριθμοι Ομαδοποίησης

Υπάρχουν τρεις γενικές κατηγορίες αλγορίθμων ομαδοποίησης (22):

- Οι αλγόριθμοι βασισμένοι σε διαχωρισμούς (partition based), που προσπαθούν να βρουν τον καλύτερο διαχωρισμό ενός συνόλου δεδομένων σε ένα συγκεκριμένο αριθμό ομάδων.
- Οι ιεραρχικοί (hierarchical) αλγόριθμοι, που προσπαθούν με ιεραρχικό τρόπο να ανακαλύψουν τον αριθμό και τη δομή των ομάδων.
- Οι πιθανοκρατικοί (probabilistic) αλγόριθμοι, που βασίζονται σε μοντέλα πιθανοτήτων.

Η ομαδοποίηση απαιτεί κάποιο μέτρο της ομοιότητας ή διαφοράς μεταξύ των δεδομένων. Συνήθως υπολογίζεται η "απόσταση" μεταξύ των δεδομένων. Έστω ένα σύνολο δεδομένων D, και δύο δεδομένα του, x, y που περιγράφονται από m χαρακτηριστικά $(x_1,x_2,\ldots,x_m)$, $(y_1,y_2,\ldots,y_m)$. Τυπικά μέτρα απόστασης αυτών των δύο δεδομένων είναι η απόσταση *Μανχάταν* και η *Ευκλείδεια* απόσταση:

$$d(x,y)=\sum_i |x_i - y_i| \qquad d(x,y) = \sqrt{\sum_i (x_i - y_i)^{\,2}}$$

Απόσταση Μανχάταν *Ευκλείδεια Απόσταση*

Αν κάποια χαρακτηριστικά είναι διακριτά, τότε η απόσταση των τιμών τους θεωρείται 0 αν πρόκειται για την ίδια τιμή και 1 αν πρόκειται για διαφορετικές τιμές. Τα αριθμητικά χαρακτηριστικά θα πρέπει να ομογενοποιούνται ώστε η απόστασή τους να πέφτει μέσα στο διάστημα [0,1].

3.4.1 *Ο αλγόριθμος k-Means*

Ένας από τους πιο γνωστούς αλγόριθμους ομαδοποίησης ο οποίος βασίζεται σε διαχωρισμούς είναι ο αλγόριθμος των K-μέσων (k-Means) (23), (21) η αλλιώς Hard c-Means. Είναι ένας από τους βασικούς επαναληπτικούς αλγόριθμους ο οποίος αποτελεί την βάση για πολλούς άλλους αλγορίθμους αυτού του τύπου. Εφαρμόζεται για την ομαδοποίηση μεγάλων συνόλων δεδομένων σε ομοιογενείς ομάδες. Όπως σε όλες τις επαναληπτικές μεθόδους ομαδοποίησης, έτσι και ο k–Means βασίζεται στην ιδέα της βελτιστοποίησης κάποιας συνάρτησης την οποία θα δούμε παρακάτω.

Τυπολόγιο

Η Συνάρτηση Κόστους που προσπαθεί να ελαχιστοποιήσει ο k-Means, δηλαδή η ελαχιστοποίηση της απόστασης ενός σημείο από το κέντρο ενός cluster, είναι η εξής :

$$J = \sum_{i=1}^{C} J_i = \sum_{i=1}^{C} \sum_{k, X_k, \in G_i} d(X_k, C_i) \quad (4)$$

Όπου C_i τα κέντρα των cluster i, $d(X_k, C_i)$ η απόσταση Μανχάταν ανάμεσα στο i_{th} κέντρο (C_i) και στο k_{th} σημείο και G_i i=1,…c το κάθε cluster. Για απλότητα χρησιμοποιείτε η Ευκλείδεια απόσταση ως μέτρο ανομοιότητας και η αντίστοιχη συνάρτηση είναι η εξής :

$$J = \sum_{i=1}^{C} J_i = \sum_{i=1}^{C} \Big(\sum_{k, X_k, \in G_i} \|X_k - C_i\|^{\,2} \Big) \quad (5)$$

Ο διαχωρισμός σε ομάδες μπορεί να οριστεί από έναν πίνακα U διάστασης c*n δυαδικού τύπου, όπου τα u_{ij} στοιχεία θα είναι 1 αν το j_{th} στοιχείο (x_j) ανήκει στη i ομάδα και 0 στην αντίθετη περίπτωση .Η εξίσωση που περιγράφει τον U είναι η εξής:

$$u_{ij} = \begin{cases} 1 \; \varepsilon\alpha\nu \; \left\|x_j - c_i\right\|^2 \leq \left\|x_j - c_k\right\|^2 \; \gamma\iota\alpha \; \kappa\acute{\alpha}\theta\varepsilon \; k \neq i \\ 0 \qquad\qquad\qquad\qquad \alpha\lambda\lambda o\acute{\upsilon} \end{cases} \quad (6)$$

Όσο ένα στοιχείο βρίσκεται σε μια ομάδα, ο πίνακας U έχει δυο προτεραιότητες οι οποίες δίνονται από τις παρακάτω εξισώσεις :

$$\sum_{i=1}^{C} u_{ij} = 1, \forall\, j = 1, \dots, n \quad (7)$$

$$\sum_{i=1}^{C} \sum_{j=1}^{n} u_{ij} = n \qquad\qquad (8)$$

Τα κέντρα υπολογίζονται ως τις μέσες τιμές από όλα τα διανύσματα της κάθε ομάδας i και δίνονται απ την παρακάτω εξίσωση:

$$C_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \quad (9)$$

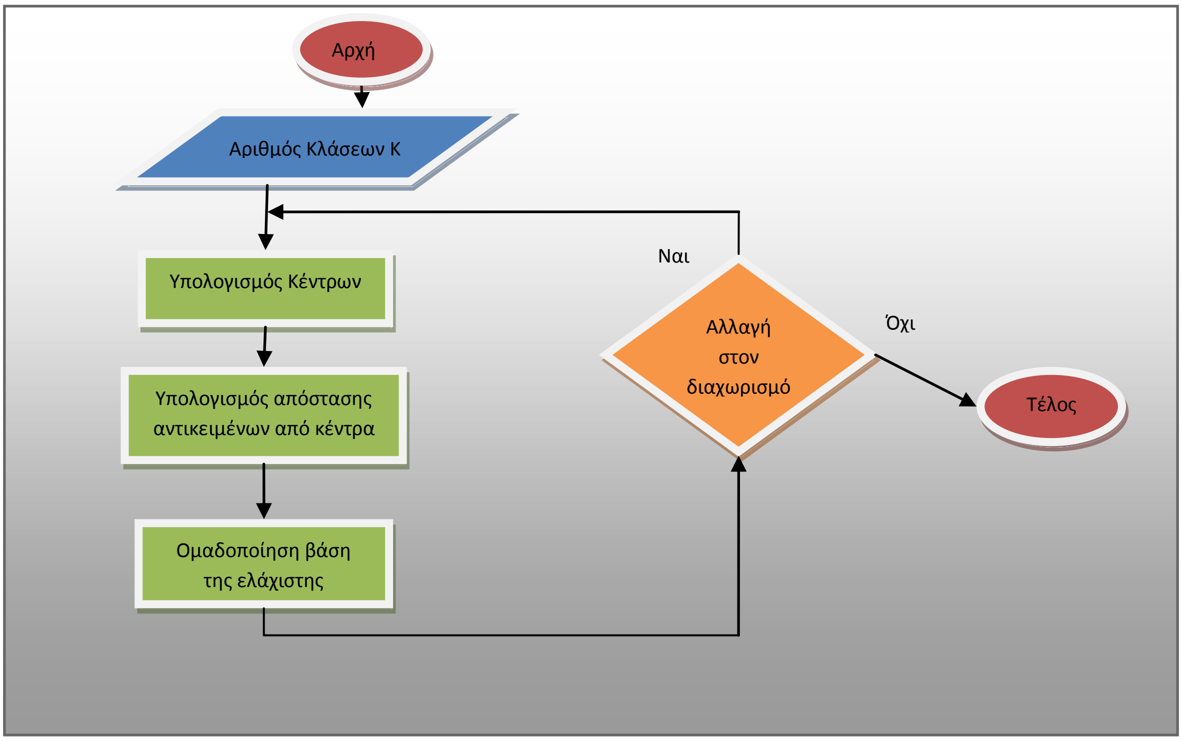
Όπου $|G_i|$ είναι το μέγεθος του G_i .

Η επίδοση του αλγορίθμου εξαρτάται από τα την αρχικοποίηση των κέντρων έτσι ο αλγόριθμος δεν δίνει εγγυημένα αποτελέσματα.

Μια σύντομη περιγραφή των βημάτων που ακολουθεί ο αλγόριθμος είναι η εξής :

- Ο αριθμός των k clusters καθορίζεται πριν (a priori) την εκτέλεση του αλγορίθμου.
- Ο αλγόριθμος ξεκινά διαλέγοντας k τυχαία σημεία από τα δεδομένα ως τα κέντρα των clusters.
- Αναθέτει κάθε σημείο στο cluster της οποίας το κέντρο είναι πιο κοντά (μικρότερη απόσταση) σε αυτό το σημείο.
- Υπολογίζει για κάθε cluster το μέσο όρο όλων των σημείων του (μέσο διάνυσμα) και ορίζει αυτό ως νέο κέντρο του.

- Τα δύο τελευταία βήματα επαναλαμβάνονται για ένα προκαθορισμένο αριθμό βημάτων ή μέχρι να μην υπάρχει αλλαγή στο διαχωρισμό των σημείων σε ομάδες.



Εικόνα 6: Περιγραφή του αλγόριθμου k-Means.

3.4.2 *Ο αλγόριθμος των Fuzzy c-Means (FCM)*

Ο αλγόριθμος Fuzzy c-Means (23), (21), γνωστός και ως Fuzzy ISODATA, είναι μια τεχνική ομαδοποίησης η οποία διαφέρει από τον αυστηρό διαχωρισμό της k-Means μεθόδου. Ο αλγόριθμος δηλαδή δεν αντιστοιχίζει το κάθε αντικείμενο σε μια μόνο ομάδα αλλά υλοποιεί ασαφής διαμερίσεις έτσι ώστε κάθε αντικείμενο να ανήκει σε όλες τις ομάδες με διαφορετικό βαθμό συμμετοχής (degree membership) για τη καθεμία ο οποίος κυμαίνετε ανάμεσα στο 0 και 1. Αυτό είναι ιδιαίτερα χρήσιμο σε περιπτώσεις που τα όρια μεταξύ των ομάδων δεν είναι ξεκάθαρα και καλά διαχωρισμένα. Ο Fuzzy c-Means (FCM) είναι ένας επαναληπτικός αλγόριθμός καθώς και ένας από τους πιο δημοφιλείς fuzzy αλγόριθμους και η βασική του ιδέα είναι να βρει ένα partition (c fuzzy cluster) για το σύνολο δεδομένων $x_j \in R^d$, j=1,…,N ελαχιστοποιώντας τη συνάρτηση που μετρά το βαθμό «μη ομοιότητας» ενός σημείου σε ένα cluster.

Τυπολόγιο:

Ο πίνακας (U) περιέχει το βαθμό συμμετοχής κάθε αντικειμένου σε κάθε cluster και αρχικοποιείται τυχαία τηρώντας την παρακάτω εξίσωση :

$$\sum_{i=1}^{C} u_{ij} = 1, \forall\, j = 1, \dots, n \quad (10)$$

Η συνάρτηση κόστους (ανομοιότητας) που χρησιμοποιεί ο FCM και καλείται να την ελαχιστοποιήσει είναι η εξής :

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^{C} J_i = \sum_{i=1}^{C} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \quad (11)$$

Όπου u_{ij} παίρνει τιμές ανάμεσα στο 0 και 1, c_i είναι το κέντρο του cluster i, d_{ij} είναι η Ευκλείδεια απόσταση ανάμεσα στο i_{th} κέντρο (c_i) και στο j_{th} σημείο των δεδομένων και m ϵ [1,∞] είναι η παράμετρος ασάφειας που συνήθως ορίζεται στο 2.

Ο υπολογισμός του κέντρου ενός cluster γίνετε λαμβάνοντας υπόψη τις τιμές που έχουν τα σημεία που ανήκουν στο cluster καθώς και το βαθμός συμμετοχής τους. Η συνάρτηση είναι η εξής :

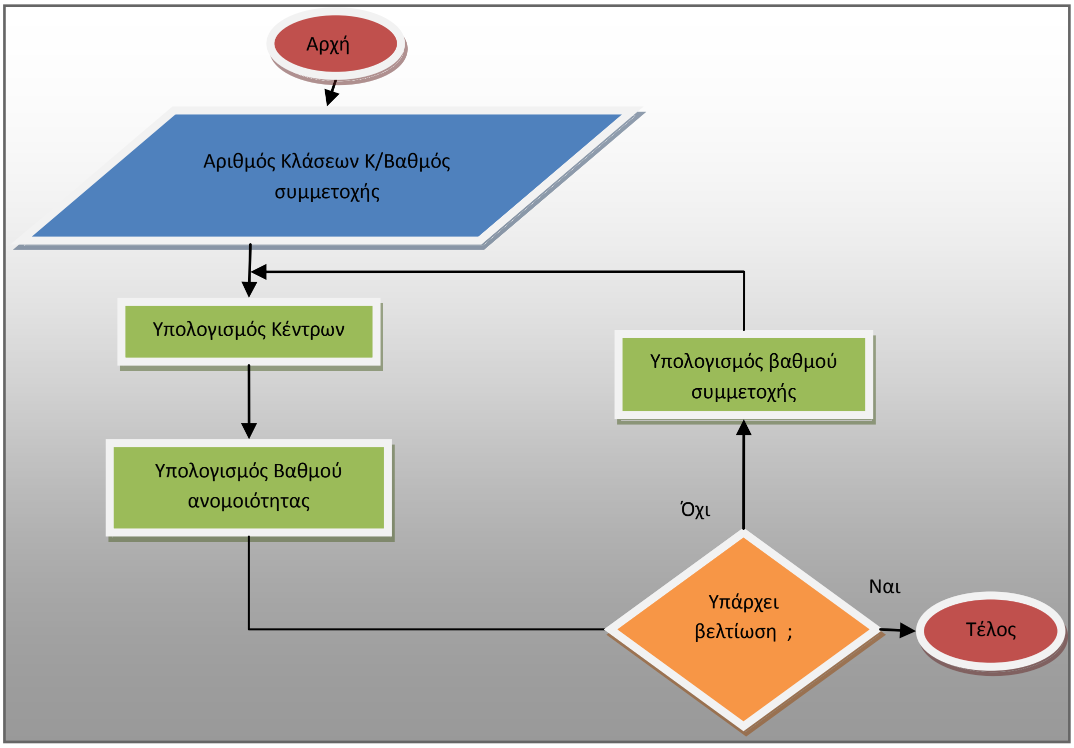
$$C_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \quad (12)$$

Ο υπολογισμός του βαθμού συμμετοχής ενός σημείου σε ένα cluster δίνεται από τον ακόλουθο τύπο:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{kj}})^{\,2/(m-1)}} \quad (13)$$

Μια σύντομη περιγραφή των βημάτων που ακολουθεί ο αλγόριθμος FCM είναι η ακόλουθη:

- Καθορισμός των k clusters και τυχαία αρχικοποίηση του πίνακα (U) που περιέχει το βαθμό συμμετοχής από τον τύπο (10).
- Υπολογισμός των κέντρων των clusters με τον τύπο (12) .
- Υπολογισμός ανομοιότητας ανάμεσα στα κέντρα και στα αντικείμενα των δεδομένων με τον τύπο (11).
- Σταματάει εάν η βελτίωση σε σχέση με τη προηγούμενη επανάληψη είναι επαρκής (πλήρη το κριτήριο για το τερματισμό του αλγορίθμου).
- Σε διαφορετική περίπτωση υπολογίζει νέες τιμές για τον πίνακα U χρησιμοποιώντας τον τύπο (13), και συνεχίζει με τον υπολογισμό νέων κέντρων στο δεύτερο βήμα.



Εικόνα 7: Περιγραφή του αλγόριθμου Fuzzy c-means

Η απόδοση του FCM εξαρτάται από την αρχικοποίηση του πίνακα U. Ενδείκνυται ο αλγόριθμος να τρέχει περισσότερες από μια φορές με διαφορετικές αρχικές τιμές για το U κάθε φορά. Τα αποτελέσματα που θα δώσει, δηλαδή τις τιμές των κέντρων των clusters, επηρεάζονται αρνητικά από σημεία που δεν ανήκουν 100% σε κάποιο cluster, δηλαδή ο βαθμός συμμετοχής τους είναι πολύ κοντά στο 0 (noise points).

Παρότι ο FCM έχει γραμμική πολυπλοκότητα, είναι πιο αργός σε σχέση με τον k-Means .Ο λόγος είναι ότι έχει να εκτελέσει πολλές πράξεις για να υπολογίζει βαθμούς συμμετοχής, κάτι το οποίο κοστίζει.

3.4.3 *Ιεραρχικοί Αλγόριθμοι (Hierarchical Clustering)*

Οι ιεραρχικοί αλγόριθμοι (21) είναι απ τις πιο διαδεδομένες clustering μεθόδους εξαιτίας της απλότητας τους. Οργανώνουν τα δεδομένα σε μια ιεραρχική δομή σύμφωνα με τον Πίνακα Γειτνίασης (proximity matrix) και παρουσιάζουν συνήθως τα αποτελέσματα σε μια δενδρική δομή που ονομάζεται δενδρόγραμμα (dendrogram). Η ρίζα του δενδρογράμματος αναπαριστά το σύνολο των δεδομένων (ενιαίο cluster που περιέχει όλα τα δεδομένα) και κάθε φύλλο αντιστοιχεί σε ένα μεμονωμένο αντικείμενο του συνόλου. Οι ενδιάμεσοι κόμβοι περιγράφουν την εγγύτητα μεταξύ των αντικειμένων και το ύψος του δενδρογράμματος εκφράζει την απόσταση μεταξύ ζευγαριών αντικειμένων ή clusters ή αντικειμένων με clusters. Τα clusters μπορούν να προκύψουν χωρίζοντας το δενδρόγραμμα σε διαφορετικά επίπεδα. Η αναπαράσταση αυτή είναι ιδιαίτερα χρήσιμη καθώς παρέχει οπτική αναπαράσταση και έτσι υπάρχει καλύτερη αντίληψη για ιεραρχικές σχέσεις που υπάρχουν σε δεδομένα. Ένα χαρακτηριστικό των ιεραρχικών μεθόδων ομαδοποίησης είναι ότι η ανάθεση ενός αντικειμένου σε μια ομάδα είναι αμετάκλητη. Όταν δηλαδή ένα αντικείμενο ενταχθεί σε μια ομάδα ποτέ δεν απομακρύνεται από αυτή και ποτέ δεν θα ενωθεί με αντικείμενα που ανήκουν σε άλλη ομάδα.

Οι ιεραρχικοί αλγόριθμοι ομαδοποίησης χωρίζονται κυρίως σε δύο κατηγορίες, τους συσσωρευτικούς (agglomerative) και τους διαιρετικούς (divisive). Οι συσσωρευτικοί ξεκινούν με N clusters που ο καθένας περιέχει ένα μόνο αντικείμενο. Έπειτα από μία σειρά από συνενώσεις τελικά όλα τα αντικείμενα ομαδοποιούνται στο ίδιο cluster. Οι διαιρετικοί αλγόριθμοι λειτουργούν με τον αντίθετο ακριβώς τρόπο. Στην αρχή όλα τα αντικείμενα ανήκουν στο ίδιο cluster και με μία διαδικασία συνεχώς χωρίζονται μέχρι όλα τα αντικείμενα να αποτελούν ένα cluster από μόνα τους. Οι διαιρετικοί αλγόριθμοι έχουν μεγάλο υπολογιστικό κόστος και γι' αυτό δεν χρησιμοποιούνται στην πράξη. Για το λόγο αυτό δίνουμε μια περιγραφή του συσσωρευτικού αλγόριθμου παρακάτω.

Τυπολόγιο:

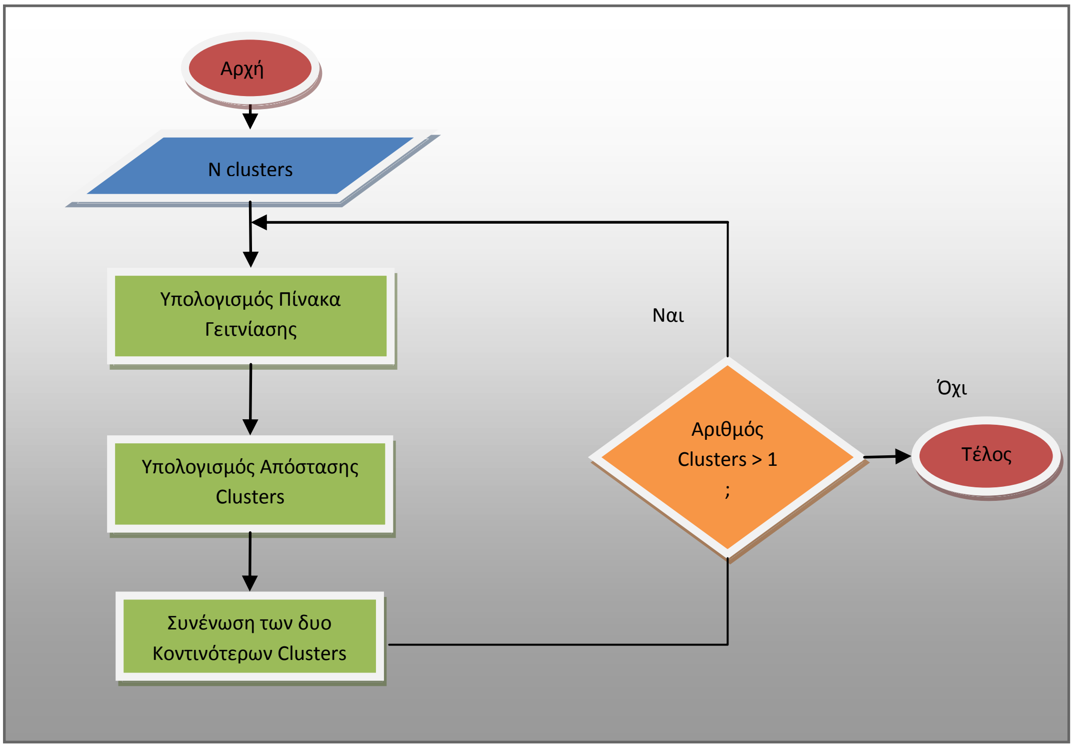
Η συνάρτηση που ψάχνει την ελάχιστη απόσταση των σημείων είναι η εξής :

$$D(C_i, C_j) = \min_{1 \le \substack{m,l \\ m \ne l} \le N} D(C_m, C_l) \quad (14)$$

Όπου *D(*,*)* είναι η συνάρτηση απόστασης στον πίνακα γειτνίασης, η οποία συνδυάζει το cluster C_i και το κοντινότερό του C_j σε ένα νέο cluster.

Μια σύντομη περιγραφή των βημάτων που ακολουθεί ο αλγόριθμος hierarcical agglomerative είναι η ακόλουθη:

- Εκκίνηση με N clusters. Υπολογισμός του πίνακα γειτνίασης για τα N clusters.
- Εύρεση της ελάχιστης απόστασης των clusters σύμφωνα με τον τύπο (14) και συνένωσή τους.
- Ενημέρωση του πίνακα γειτνίασης με τον υπολογισμό των αποστάσεων μεταξύ των νέων clusters με τους υπόλοιπους.
- Επανάληψη των δυο προηγούμενων βημάτων μέχρι όλα τα αντικείμενα να ανήκουν στο ίδιο cluster.



Εικόνα 8: Περιγραφή του ιεραρχικού αλγόριθμου Συστάδων (agglomerative).

Υπάρχουν πολλές παραλλαγές για τους HC αλγόριθμους που βασίζονται σε διαφορετικούς ορισμούς για την απόσταση μεταξύ δύο clusters. Το μεγαλύτερο μειονέκτημά τους είναι ότι δεν έχουν ανοχή σε θόρυβο. Όπως αναφέραμε και προηγούμενος, όταν ένα στοιχείο ενταχθεί σε μια ομάδα, παραμένει σ' αύτη και δεν γίνετε να αλλάξει αποκλίνοντας έτσι το ενδεχόμενο διόρθωσης τυχόν σφαλμάτων.

3.4.4 *Μέθοδος* ***Gaussian Mixture Model (GMM)***

Στους Πιθανοκρατικούς αλγόριθμους (21), τα στοιχεία των αντικειμένων παράγονται σύμφωνα με διάφορες κατανομές πιθανοτήτων. Τα στοιχεία διαφορετικών clusters, παράχθηκαν από διαφορετικές κατανομές πιθανοτήτων οι οποίες προέρχονται από διαφορετικούς τύπους συναρτήσεων πυκνότητας πιθανότητας (π.χ. Gaussian ή *t*-distribution). Αν οι κατανομές είναι γνωστές, ο υπολογισμός των clusters των δεδομένων ισοδυναμεί με την εκτίμηση των παραμέτρων διάφορων υποκείμενων υποδειγμάτων. Η ομαδοποίηση Gaussian Mixture περιλαμβάνει την επιλογή μοντέλου, δηλαδή τον προσδιορισμό του αριθμού των ομάδων σε ένα μείγμα και την εκτίμηση των παραμέτρων κάθε ομάδας, μέσω των αντικειμένων που παρατηρούνται, τα οποία βρίσκονται σε μορφή διανυσμάτων. Οι ταξινομητές τύπου Gaussian, εξαρτώνται σε μεγάλο βαθμό από την μέθοδο που βασίζεται στην αποστεριόρι πιθανότητα (posterior probability) και από την μέθοδο υποκείμενης απόστασης, η οποία ονομάζεται και απόσταση Mahalanobis.

Τυπολόγιο:

Η εκτίμηση της Μέγιστης Πιθανότητας (Maximum likehood (ML)) είναι μια σημαντική στατιστική προσέγγιση για την εκτίμηση των παραμέτρων και θεωρεί ότι η καλύτερη εκτίμηση είναι εκείνη που μεγιστοποιεί την πιθανότητα που παράγονται όλες οι παρατηρήσεις, η οποία δίνετε από την εξής συνάρτηση πυκνότητας :

$$p(\{x_1, \dots, x_N\}|\theta) = \prod_{j=1}^{N} p(x_j|\theta) \quad (15)$$

ή σε λογαριθμική μορφή:

$$I(\theta) = \sum_{j=1}^{N} ln\, p(x_j|\theta) \quad (16)$$

Ένας απ' τους δημοφιλέστερους αλγόριθμους που υλοποιεί τον παραπάνω συλλογισμό είναι ο Expectation Maximization (EM) αλγόριθμος, ο ποίος δέχεται τα δεδομένα ως ελλιπής και διαιρεί κάθε στοιχείο x_j σε δυο μέρη $x_j = \{x_j^g, x_j^m\}$ όπου x_j^g είναι τα χαρακτηριστικά των στοιχείων και $x_j^m = (x_{j1}^m, \dots x_{jK}^m)$ είναι τα ελλιπής στοιχεία και τα x_{ji}^m παίρνουν τις τιμές 1 ή 0 αν το x_j ανήκει στην ομάδα *i* ή όχι. Έτσι η πιθανότητα των πλήρη στοιχείων υπολογίζεται σε λογαριθμική μορφή απ' τη παρακάτω συνάρτηση :

$$I(\theta) = \sum_{j=1}^{N} \sum_{i=1}^{K} x_{ji}^m\, log[P(C_j)p(x_j^\theta \theta_i)] \quad (17)$$

3.5 Δείκτες Επικύρωσης

Με δεδομένο ότι τα αποτελέσματα των αλγορίθμων ομαδοποίησης δεν είναι εντελώς αξιόπιστα, είναι απαραίτητο, να αξιολογηθούν περαιτέρω. Η Επικύρωση της Ομαδοποίησης (Cluster Validity) είναι η διαδικασία της ποσοτικής αξιολόγησης των αποτελεσμάτων ενός αλγορίθμου ομαδοποίησης. Στην πράξη, σπάνια ελέγχεται ξεχωριστά αν το *X* διέπεται από μία δομή, αφού με την επικύρωση της ομαδοποίησης μπορεί να ελεγχθεί η ορθότητα των αποτελεσμάτων των αλγορίθμων ομαδοποίησης και έτσι να καταλήξει κάποιος στα αντίστοιχα συμπεράσματα. Επιπλέον, τις περισσότερες φορές οι αλγόριθμοι ομαδοποίησης εφαρμόζονται γνωρίζοντας εκ των προτέρων ότι υπάρχει μία δομή που διέπει το σύνολο των δεδομένων (21).

Έστω ότι με C συμβολίζεται η δομή της ομαδοποίησης που προέκυψε μετά από την εφαρμογή του αλγορίθμου ομαδοποίησης στο *X*. Για να μελετηθεί η εγκυρότητα της ομαδοποίησης υπάρχουν γενικά δύο προσεγγίσεις. Η πρώτη βασίζεται σε σύγκριση της διαμέρισης του αλγορίθμου ομαδοποίησης με μία εξωτερική ανεξάρτητη δομή (εξωτερικά κριτήρια) και η δεύτερη βασίζεται στην σύγκριση μιας διαμέρισης που έχει προκύψει από έναν αλγόριθμο ομαδοποίησης με άλλες διαμερίσεις του ίδιου αλγορίθμου αλλά με διαφορετικές παραμέτρους (σχετικά κριτήρια). Παρακάτω αναλύονται δυο δείκτες επικύρωσης (Silhouette και Davies-bouldin) οι οποίοι ανήκουν στην δεύτερη κατηγορία (σχετικά κριτήρια) και θα χρησιμοποιηθούν για την αξιολόγιση των αποτελεσμάτων.

3.5.1 *Δείκτης Silhouette*

Ο δείκτης Silhouette ((20), (24), (25)) είναι ιδιαίτερα χρήσιμος όταν το ζητούμενο είναι η εύρεση συμπαγών και καλά διαχωρισμένων clusters. Για την κατασκευή των silhouettes χρειάζεται μία διαμέριση του συνόλου δεδομένων που έχει προκύψει από την εφαρμογή ενός αλγορίθμου ομαδοποίησης και ο πίνακας εγγύτητας που περιέχει τις αποστάσεις μεταξύ των αντικειμένων. Δοθέντος ενός cluster, αυτή η μέθοδος αντιστοιχίζει σε κάθε αντικείμενο του cluster μία τιμή $s(i)$ που ονομάζεται εύρος του silhouette και δείχνει τον βαθμό συμμετοχής του αντικειμένου i στο cluster στο οποίο αντιστοιχήθηκε. Ας συμβολιστεί με i ένα αντικείμενο του συνόλου δεδομένων, με C_j το cluster στο οποίο έχει αντιστοιχηθεί το i και με a_i ο μέσος όρος της ανομοιότητας μεταξύ του αντικειμένου i με όλα τα άλλα αντικείμενα στο cluster C_j. Επιπλέον για οποιοδήποτε cluster C_k διαφορετικό από το cluster C_j :

$$b(i) = \min_{C_k \neq C_j} d(i, C_k) \quad (18)$$

όπου $k = 1,2,\dots,c$ και $k \neq j$.
Τότε το εύρος του silhouette καθορίζεται από τη σχέση:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (19)$$

Ένας γείτονας του αντικειμένου i είναι το cluster C_k όταν η τιμή $d(i, C_k) = b(i)$ είναι η ελάχιστη. Το cluster C_k αντιπροσωπεύει την δεύτερη καλύτερη επιλογή για το αντικείμενο i. Από τον ορισμό είναι προφανές ότι $-1 \leq s(i) \leq 1$. Το $s(i)$ πλησιάζει το 1 όταν η μέση (εντός του cluster) ανομοιότητα $a(i)$ είναι πολύ μικρότερη από την ελάχιστη (εκτός του cluster) ανομοιότητα $b(i)$ και έτσι το αντικείμενο i έχει ταξινομηθεί επιτυχώς. Από την άλλη μεριά, εάν το $s(i)$ παίρνει τιμές κοντά στο -1 αυτό σημαίνει ότι το $a(i)$ είναι πολύ μεγαλύτερο από το $b(i)$. Σε αυτήν την περίπτωση το αντικείμενο i δεν έχει ταξινομηθεί επιτυχώς και έτσι θα πρέπει να αντιστοιχηθεί ξανά. Εάν τα $a(i)$ και $b(i)$ έχουν παρόμοιες τιμές τότε το $s(i)$ είναι περίπου 0 και σε αυτήν την περίπτωση το αντικείμενο i βρίσκεται εξίσου μακριά από τα cluster C_j και C_k.

Εάν για τα δεδομένα χρησιμοποιούνται μέτρα ομοιότητας και τα $a^{'}(i)$ και $d^{'}(i, C)$ αναπαριστούν τις αντίστοιχες μέσες τιμές ομοιοτήτων, τότε $b^{'}(i) = \max_{C \neq A} d^{'}(i, C)$. Η αναπαράσταση, σε αυτήν την περίπτωση, για το εύρος του silhouette θα δίνεται από τη σχέση:

$$s(i) = \frac{a^{'}(i) - b^{'}(i)}{\max\{a^{'}(i), b^{'}(i)\}} \quad (20)$$

Είναι δυνατό να υπολογιστεί το silhouette του cluster, το οποίο ονομάζεται μέσο εύρος silhouette και αναπαριστά την ετερογένεια του cluster C_j. Το silhouette του cluster δίνεται από τη σχέση:

$$S_i = \frac{1}{m} \sum_{i=1}^{m} s(i) \quad (21)$$

Επιπλέον, το καθολικό εύρος του silhouette δίνεται από τη σχέση:

$$GS_u = \frac{1}{C} | \sum_{j=1}^{C} S_j \quad (22)$$

όπου το U είναι οποιαδήποτε διαμέριση, $U \leftrightarrow C: C_i \cup C_2 \cup \ldots \cup C_c$. Το καθολικό εύρος του silhouette χρησιμοποιείται σαν ένα μέτρο επικύρωσης για το *U*. Για να επιλεχθεί ο βέλτιστος αριθμός των clusters για ένα σύνολο δεδομένων με τη χρήση αυτού του δείκτη, θα πρέπει να επιλεχθεί η διαμέριση U με τη μέγιστη τιμή για το GS_u.

3.5.2 *Δείκτης Davies-Bouldin*

Έστω ότι s_i είναι το μέτρο της διασποράς ενός cluster C_i και $d(C_i, C_j) \equiv d_{ij}$ η ανομοιότητα μεταξύ των δύο clusters. Ένας δείκτης ομοιότητας R_{ij} μεταξύ των C_i και C_j ικανοποιεί τα παρακάτω:

- $R_{ij} \geq 0$
- $R_{ij} = R_{ij}$
- Εάν $s_i = 0$ και s_j=0 τότε $R_{ij} > R_{ik}$
- Εάν $s_j = s_k$ και $d_{ij} < d_{ik}$ τότε $R_{ij} > R_{ik}$

Αυτές οι συνθήκες δηλώνουν ότι το R_{ij} είναι θετικό και συμμετρικό. Μία επιλογή για το R_{ij} που να ικανοποιεί αυτές τις συνθήκες είναι:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \quad (23)$$

Σε αυτήν την περίπτωση ο δείκτης των Davies-Bouldin ορίζεται ως :

$$DB = \frac{1}{m}\sum_{i=1}^{m} R_i \quad (24)$$

όπου $R_i == \max_{j=1,\dots j\neq i\dots,m} R_{ij}, i = 1, \dots, m$.

Η ανομοιότητα μεταξύ των clusters C_i και C_j σε ένα 1-διάστατο χώρο ορίζεται ως (24), (26):

$$d_{ij} = \|\overline{x_\iota} - \overline{x_j}\| = \sqrt{\sum_{k=1}^{l} |\overline{x_{\iota k}} - \overline{x_{jk}}|^{\,2}} \quad (25)$$

Η διασπορά ενός cluster C_i δηλώνεται ως :

$$s_i = \sqrt{\frac{1}{n_i}\sum_{x\in C_i}\|x - \overline{x_\iota}\|^{\,2}} \quad (26)$$

Ο DB_m είναι ο μέσος όρος της ομοιότητας μεταξύ κάθε cluster και ενός άλλου cluster που είναι ο πιο όμοιός του. Μικρές τιμές για τον δείκτη DB εκφράζουν την ύπαρξη συμπαγών και καλά διαχωριζόμενων clusters.

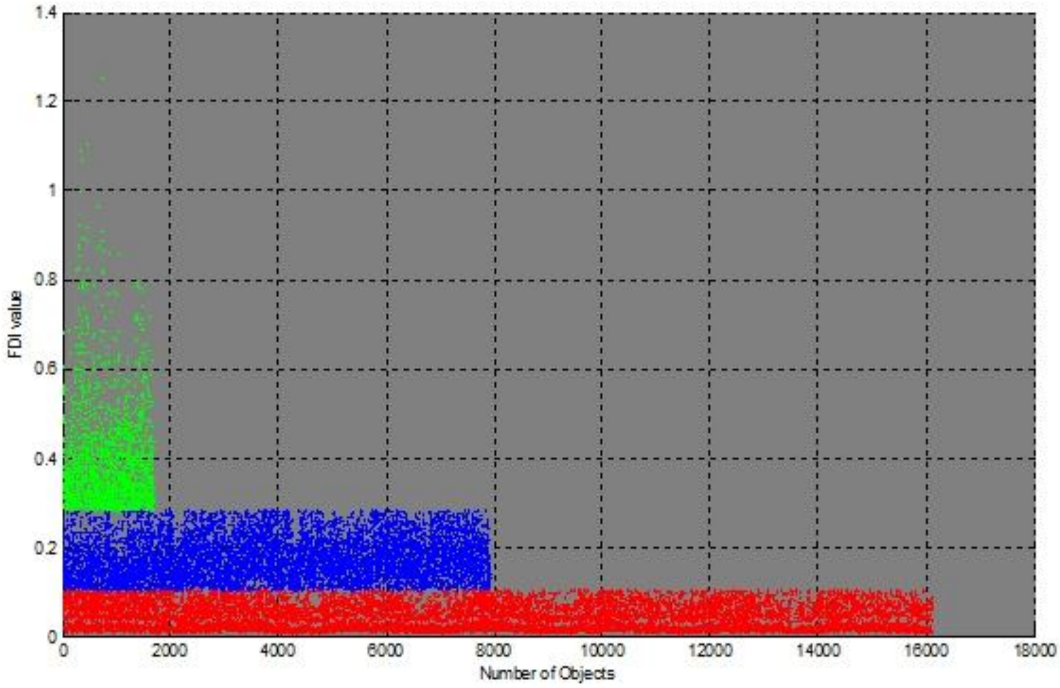
Κεφάλαιο 4- Εφαρμογή αλγορίθμων και Παρουσίαση Αποτελεσμάτων.

4.1 Εισαγωγή.

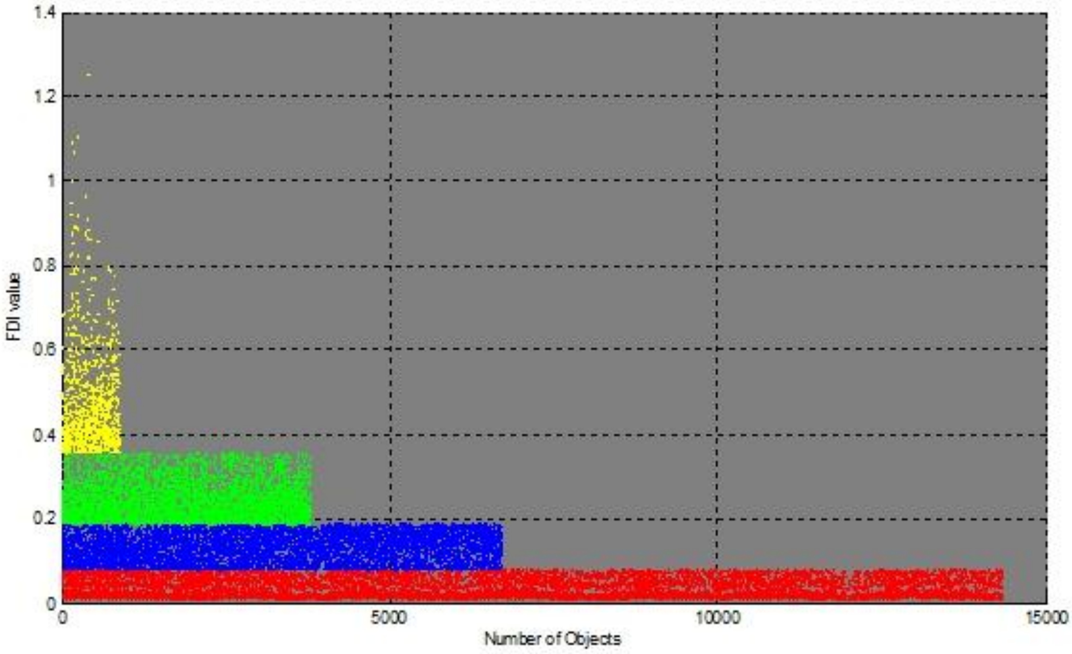
Σ' αυτό το κεφάλαιο αναφέρονται τα αποτελέσματα που παρατηρήθηκαν από την υλοποίηση των διάφορων τεχνικών ομαδοποίησης. Πιο συγκεκριμένα, σ' αυτή την ενότητα αναφέρονται τα σχετικές αποτελέσματα για τους τρείς αλγόριθμους (k-Means, Fuzzy c-Means, GMM) που εφαρμόστηκαν στα δεδομένα που αντιπροσωπεύουν τον μετεωρολογικό δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών καθώς και της περιοχής της Ηλιούπολης. Έτσι, οι πίνακες που παρουσιάζονται παρακάτω, περιέχουν τα όρια του κάθε cluster, τον αριθμό των δεδομένων που εντάσσονται σε κάθε cluster καθώς και το ποσοστό που καταλαμβάνει to κάθε cluster ως προς το συνολικό αριθμό των δεδομένων που χρησιμοποιήθηκαν. Ο αριθμός των clusters (k) που δοκιμάστηκαν οι αλγόριθμοι κυμαίνετε από 3 ως 7 και οι παράμετροι που χρησιμοποιήθηκαν αναφέρονται παρακάτω.

4.2 Αποτελέσματα Fuzzy c-Means

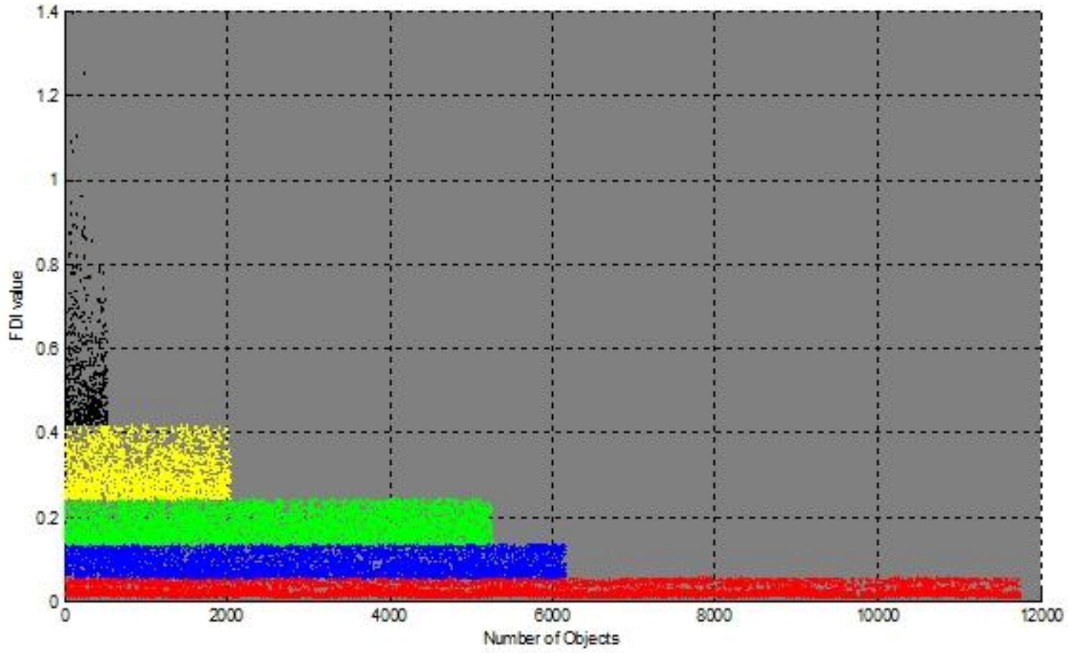
Κατά την εφαρμογή του αλγόριθμου FCM (27), καθορίσαμε τον αριθμό των clusters (3-7). Έπειτα με την συνάρτηση *initfcm* αρχικοποιήσαμε τον πίνακα *U*, ο οποίος περιέχει τον βαθμό συμμετοχής του κάθε στοιχείου, με τυχαίες τιμές. Τέλος, τρέξαμε τη συνάρτηση *stepfcm*, η οποία υλοποιεί τον αλγόριθμο FCM, με παράμετρο ασάφειας '*expo*' ίση με 2 για 100 φορές έτσι ώστε να εξάγουμε ικανοποιητικά αποτελέσματα τα οποία φαίνονται στη συνέχεια :



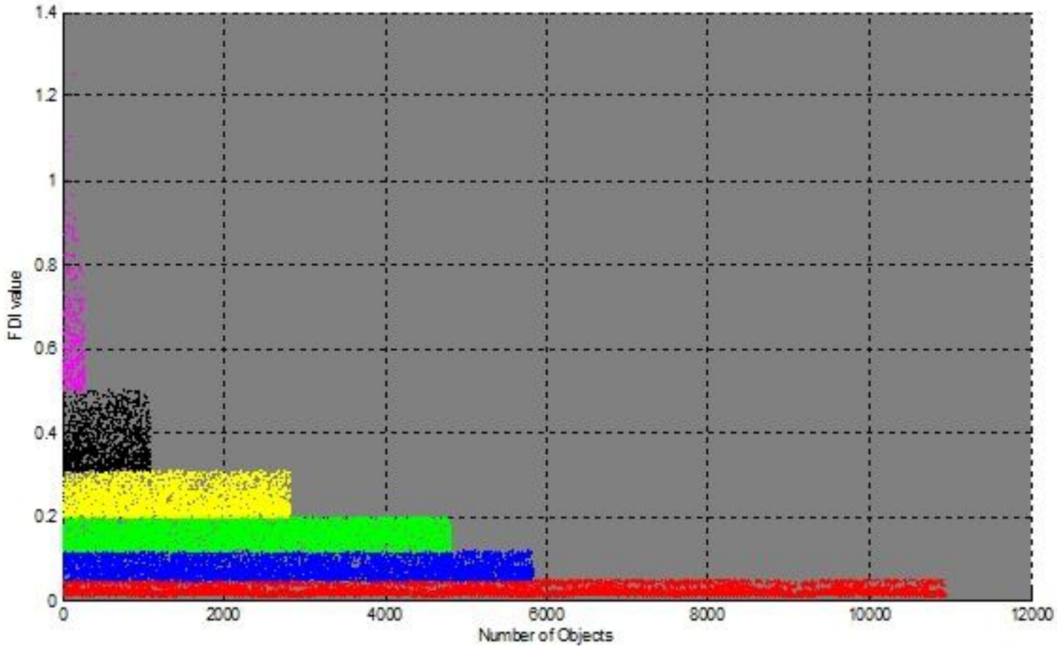
Εικόνα 9: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 3 κλάσεις με τη μέθοδο Fuzzy c-Means.



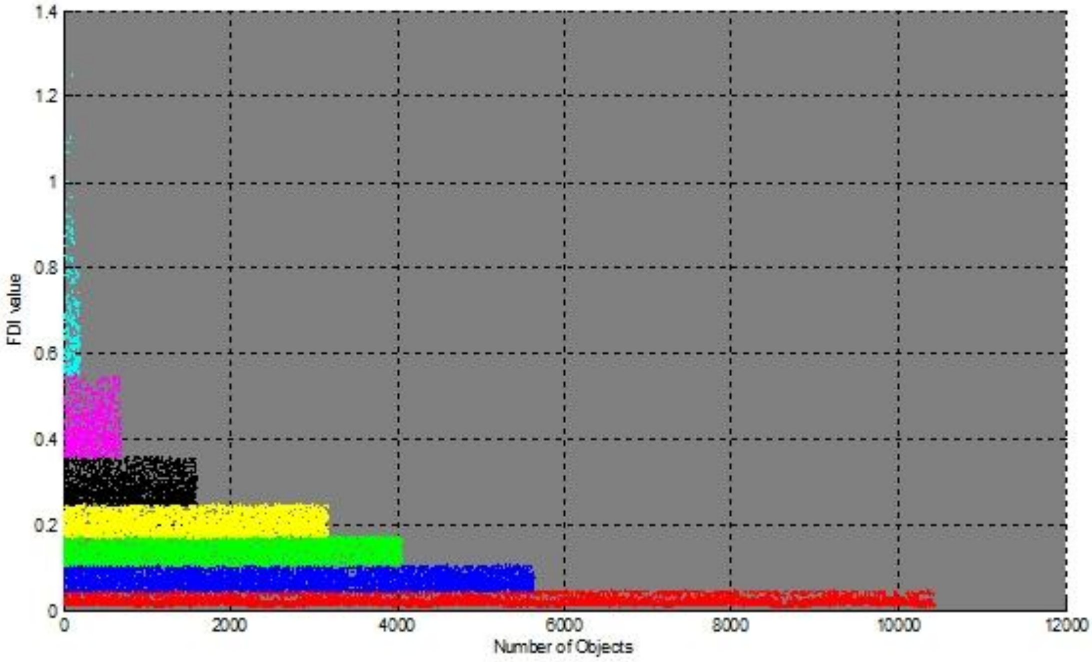
Εικόνα 10: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 4 κλάσεις με τη μέθοδο Fuzzy c-Means.



Εικόνα 11: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 5 κλάσεις με τη μέθοδο Fuzzy c-Means.



Εικόνα 12: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 6 κλάσεις με τη μέθοδο Fuzzy c-Means.



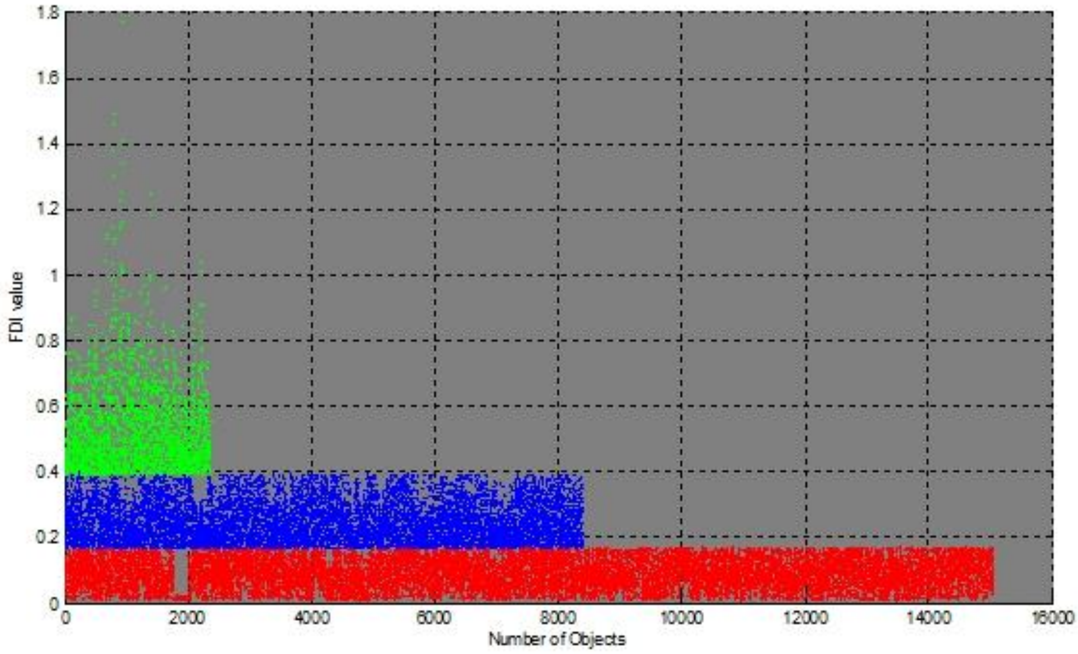
Εικόνα 13: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 7 κλάσεις με τη μέθοδο Fuzzy c-Means.

Πίνακας 5: Όρια κλάσεων (από 3 έως 7 κλάσεις) για τα δεδομένα της περιοχής των Σερρών με τη μέθοδο Fuzzy c-Means.

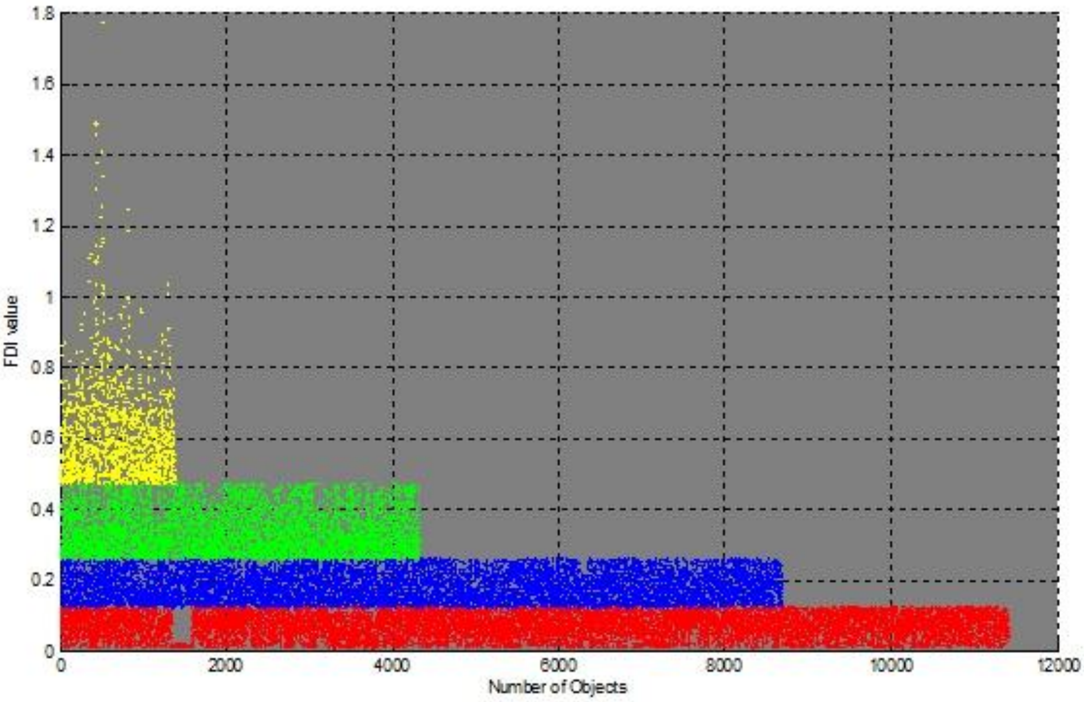
ΟΡΙΑ ΤΩΝ CLUSTERS ΓΙΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΡΙΟΧΗΣ ΤΩΝ ΣΕΡΡΩΝ					
Classes	*k=3*	*k=4*	*k=5*	*k=6*	*k=7*
A Class	0.0103-0.1065	0.0103-0.0810	0.0103-0.0565	0.0103-0.0495	0.0103-0.0458
B Class	0.1065-0.2849	0.0810-0.1890	0.0565-0.1366	0.0495-0.1177	0.0458-0.1053
C Class	0.2849-1.2545	0.1890-0.3580	0.1366-0.2434	0.1177-0.1992	0.1053-0.1716
D Class		0.3580-1.2545	0.2434-0.4198	0.1992-0.3091	0.1716-0.2475
E Class			0.4198-1.2545	0.3091-0.5004	0.2475-0.3592
F Class				0.5004-1.2545	0.3592-0.5489
G Class					0.5489-1.2545

Πίνακας 6: Αριθμός δεδομένων και ποσοστό συμμετοχής σε κάθε κλάση (από 3 έως 7 κλάσεις) για την περιοχή των Σερρών με τη μέθοδο Fuzzy c-Means.

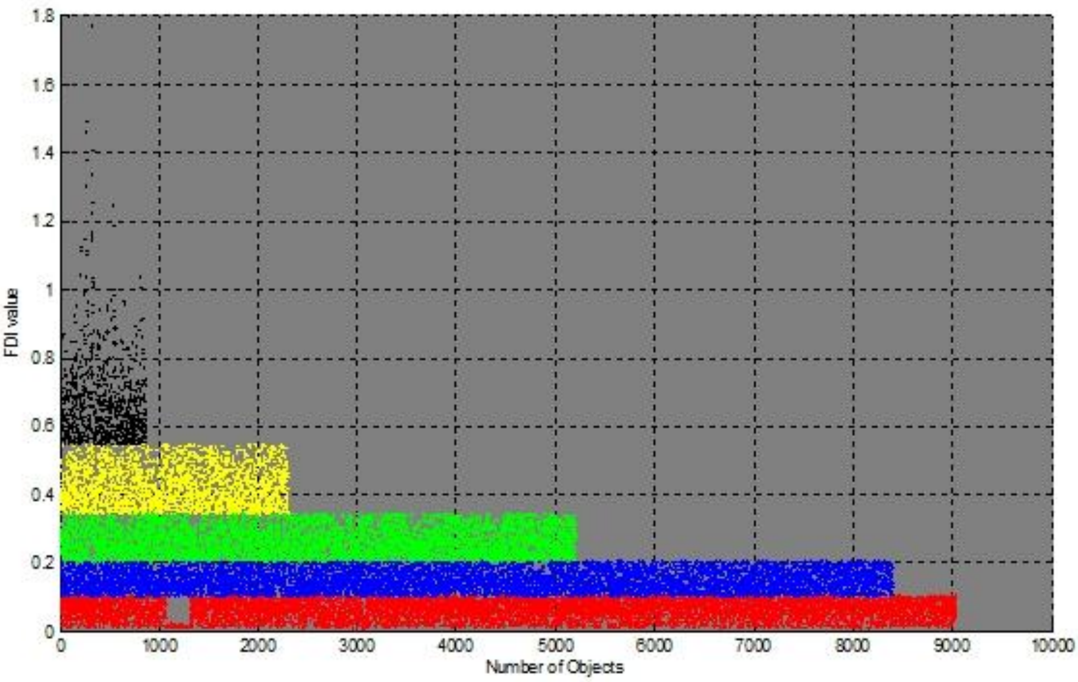
Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) για την περιοχή των Σερρών										
Classes	*k=3*		*k=4*		*k=5*		*k=6*		*k=7*	
A Class	16112	62.68	14339	55.79	11748	45.74	10931	42.53	10427	40.57
B Class	7911	30.78	6701	26.07	6149	23.92	5828	22.67	5625	21.88
C Class	1681	6.54	3790	14.74	5251	20.42	4798	18.67	4059	15.79
D Class			874	3.40	2031	7.90	2807	10.92	3151	12.26
E Class					515	2.01	1076	4.19	1584	6.16
F Class							264	1.03	673	2.62
G Class									185	0.72



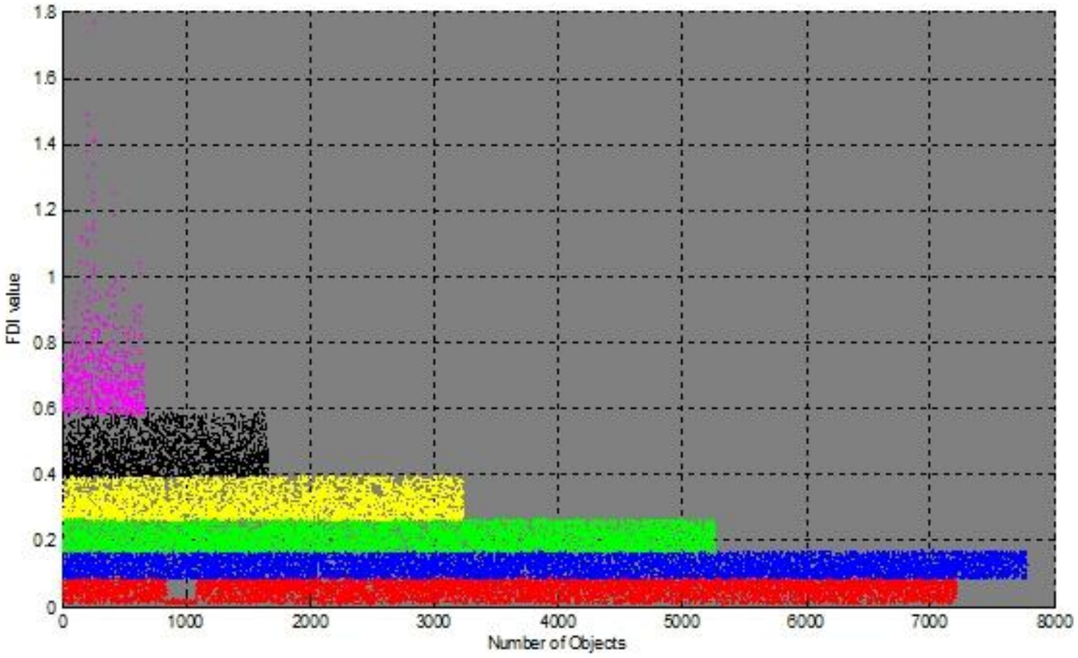
Εικόνα 14: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 3 κλάσεις με τη μέθοδο Fuzzy c-Means.



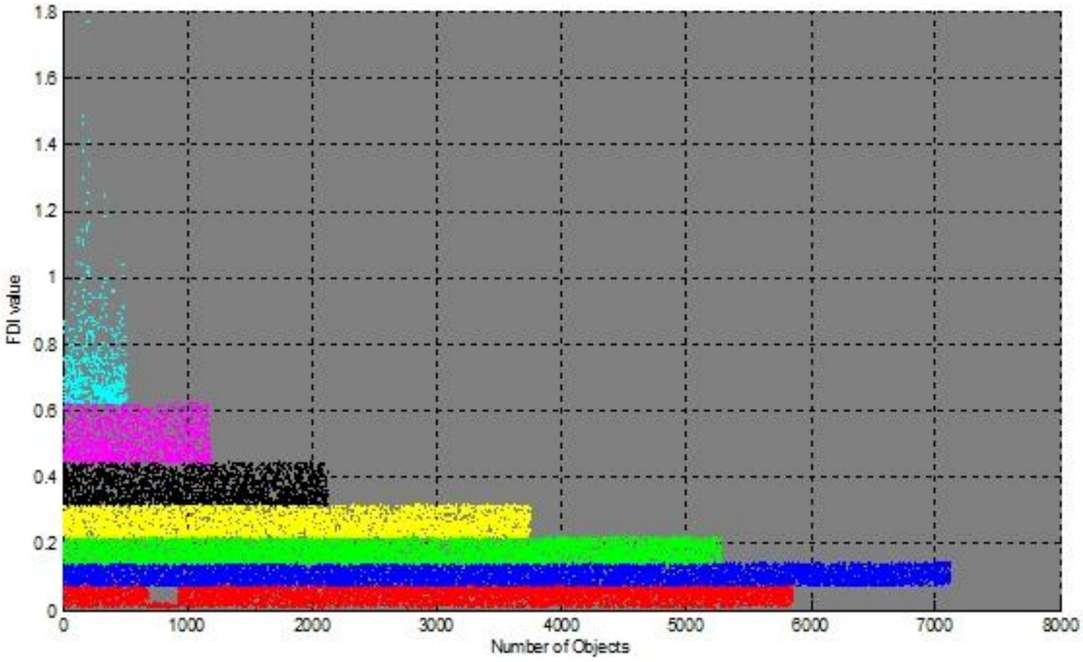
Εικόνα 15: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 4 κλάσεις με τη μέθοδο Fuzzy c-Means.



Εικόνα 16: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 5 κλάσεις με τη μέθοδο Fuzzy c-Means.



Εικόνα 17: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 6 κλάσεις με τη μέθοδο Fuzzy c-Means.



Εικόνα 18: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 7 κλάσεις με τη μέθοδο Fuzzy c-Means.

Πίνακας 7: Όρια κλάσεων (από 3 έως 7 κλάσεις) για τα δεδομένα της περιοχής της Ηλιούπολης με τη μέθοδο Fuzzy c-Means.

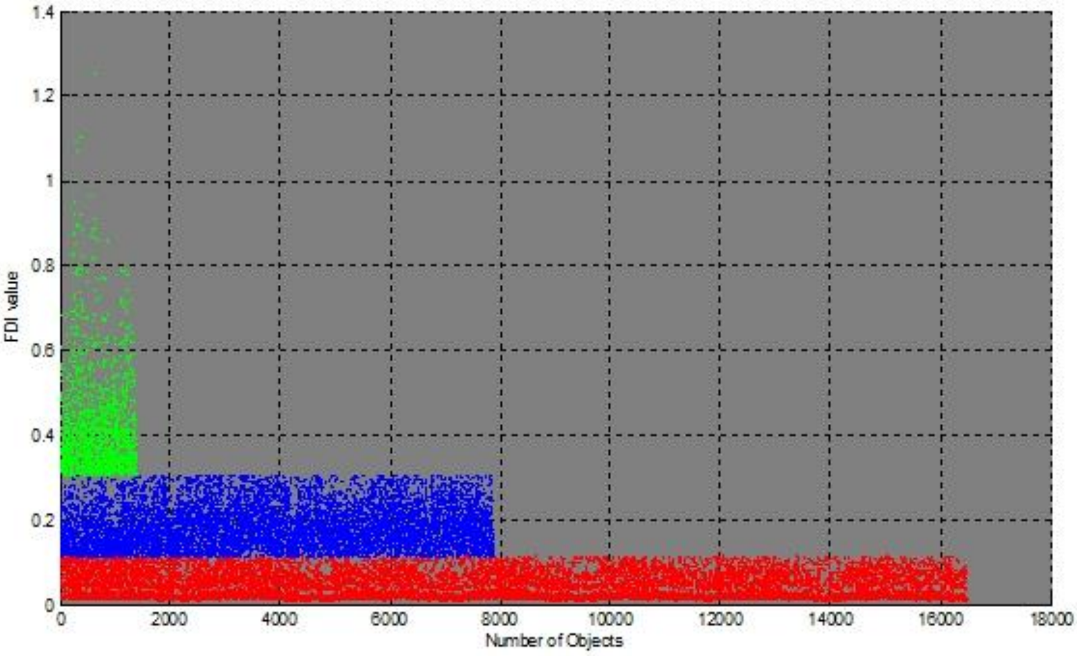
ΟΡΙΑ ΤΩΝ CLUSTERS ΓΙΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΡΙΟΧΗΣ ΤΗΣ ΗΛΙΟΥΠΟΛΗΣ					
Classes	*k=3*	*k=4*	*k=5*	*k=6*	*k=7*
A Class	0.0102-0. 1689	0. 0102-0.1264	0. 0102-0.1037	0. 0102-0.0874	0. 0102-0.0758
B Class	0. 1689-0.3939	0.1264-0.2619	0.1037-0.2062	0. 0874-0.1679	0.0758-0.1428
C Class	0.3939-1.7747	0.2619-0.4748	0.2062-0.3455	0.1679-0.2653	0.1428-0.2212
D Class		0.4748-1.7747	0. 3455-0.5479	0.2653-0.3962	0.2212-0.3186
E Class			0.5479-1.7747	0.3962-0.5863	0.3186-0.4463
F Class				0.5863-1.7747	0.4463-0.6243
G Class					0.6243-1.7747

Πίνακας 8: Αριθμός δεδομένων και ποσοστό συμμετοχής σε κάθε κλάση (από 3 έως 7 κλάσεις) για την περιοχή της Ηλιούπολης με τη μέθοδο Fuzzy c-Means.

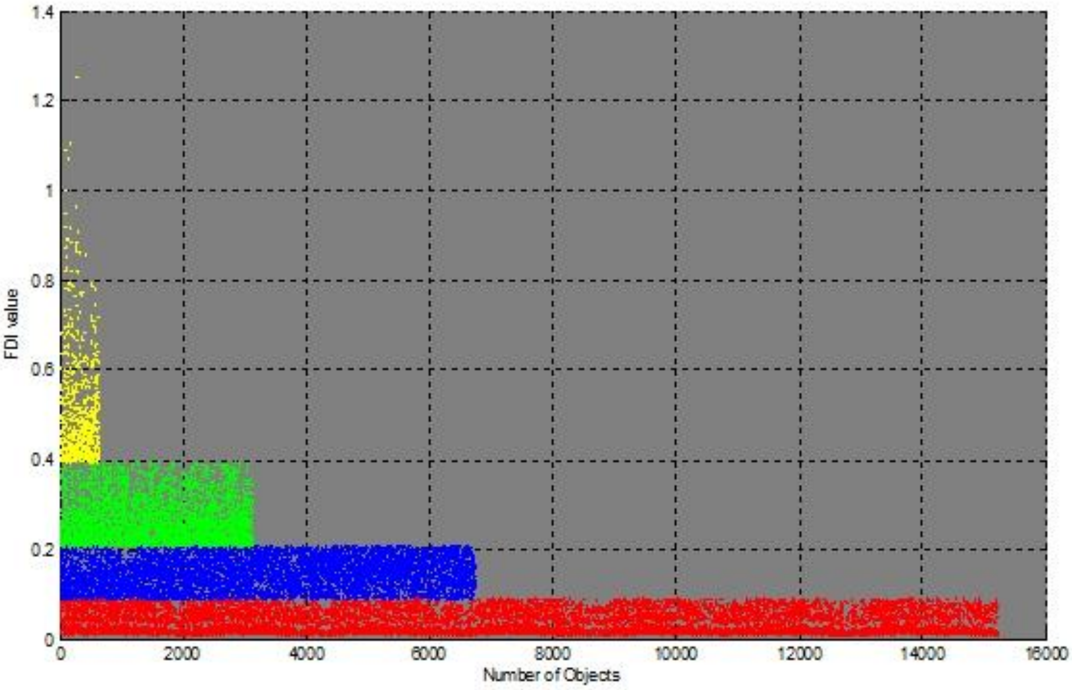
Αριθμός δεδομένων/Ποσοστό συμμετοχής (%)για την περιοχή της Ηλιούπολης										
Classes	*k=3*		*k=4*		*k=5*		*k=6*		*k=7*	
A Class	15038	58.34	11399	44.22	9028	35.02	7204	27.95	5847	22.68
B Class	8389	32.55	8685	33.69	8390	32.55	7773	30.16	7118	27.61
C Class	2349	9.11	4317	16.75	5197	20.16	5263	20.42	5271	20.45
D Class			1375	5.34	2296	8.91	3227	12.52	3742	14.52
E Class					865	3.36	1654	6.42	2119	8.22
F Class							655	2.54	1178	4.58
G Class									501	1.94

4.3 Αποτελέσματα k-Means

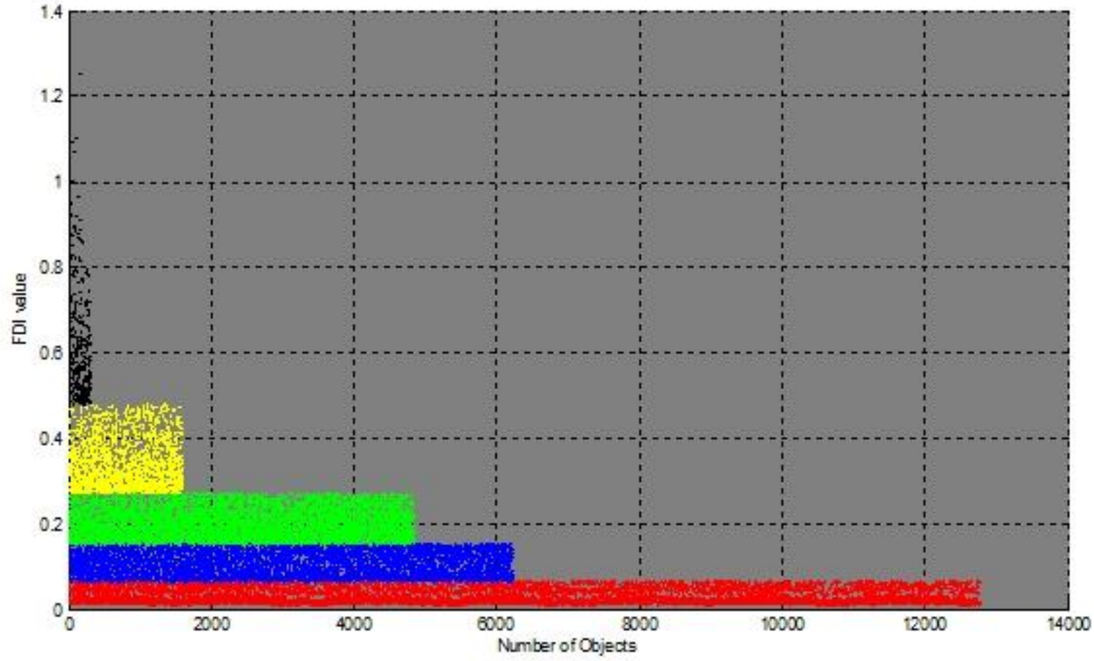
Όπως αναφέραμε και στην προηγούμενη ενότητα, ο αριθμός των clusters για τον συγκεκριμένο αλγόριθμο καθορίζεται από την αρχή (3-7) (27). Η παράμετρος *'distance'* της συνάρτησης *kmeans* που χρησιμοποιήθηκε στο matlab για την υλοποίηση του αλγόριθμου αναφέρετε στον τρόπο μέτρησης της απόστασης των δεδομένων. Στη συγκεκριμένη περίπτωση υπολογίζετε η ευκλείδεια απόσταση χρησιμοποιώντας την τιμή *'sqEuclidean'* για την παραπάνω παράμετρο. Η παράμετρος *'start'* αναφέρετε στη μέθοδο αρχικοποίησης των κέντρων. Η προεπιλεγμένη τιμή *'sample'*, επιλέγει *k* τυχαία στοιχεία από το σύνολο των δεδομένων μας και τα θέτει ως αρχικά κέντρα. Για να περιορίσουμε αυτή τη τυχαία αρχικοποίησης, χρησιμοποιήσαμε την τιμή *'cluster'* η οποία επιλέγει τυχαία στοιχεία από το 10 % του συνόλου των δεδομένων μας. Τα αποτελέσματα της συγκεκριμένης μεθόδου φαίνονται παρακάτω :



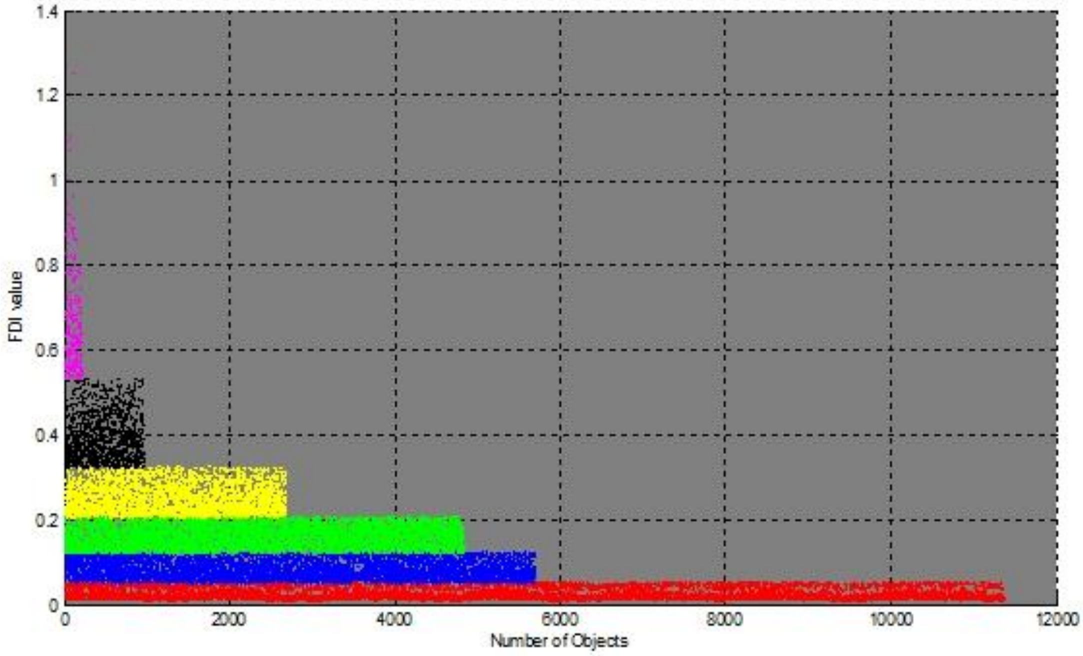
Εικόνα 19: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 3 κλάσεις με τη μέθοδο k-Means.



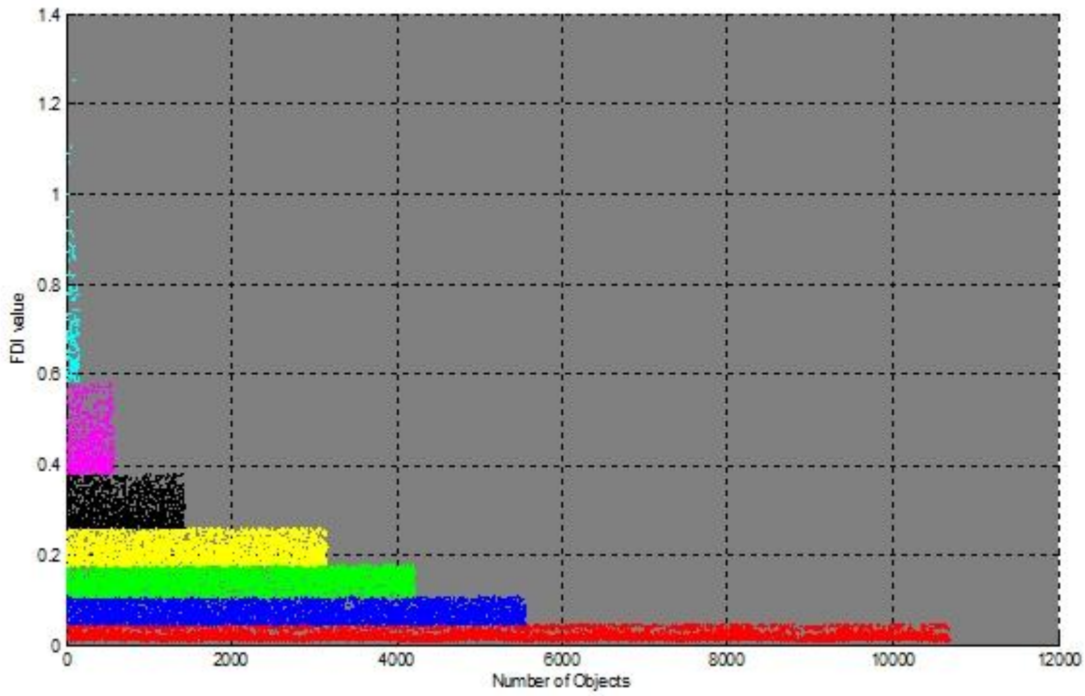
Εικόνα 20: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 4 κλάσεις με τη μέθοδο k-Means.



Εικόνα 21: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 5 κλάσεις με τη μέθοδο k-Means.



Εικόνα 22: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 6 κλάσεις με τη μέθοδο k-Means.



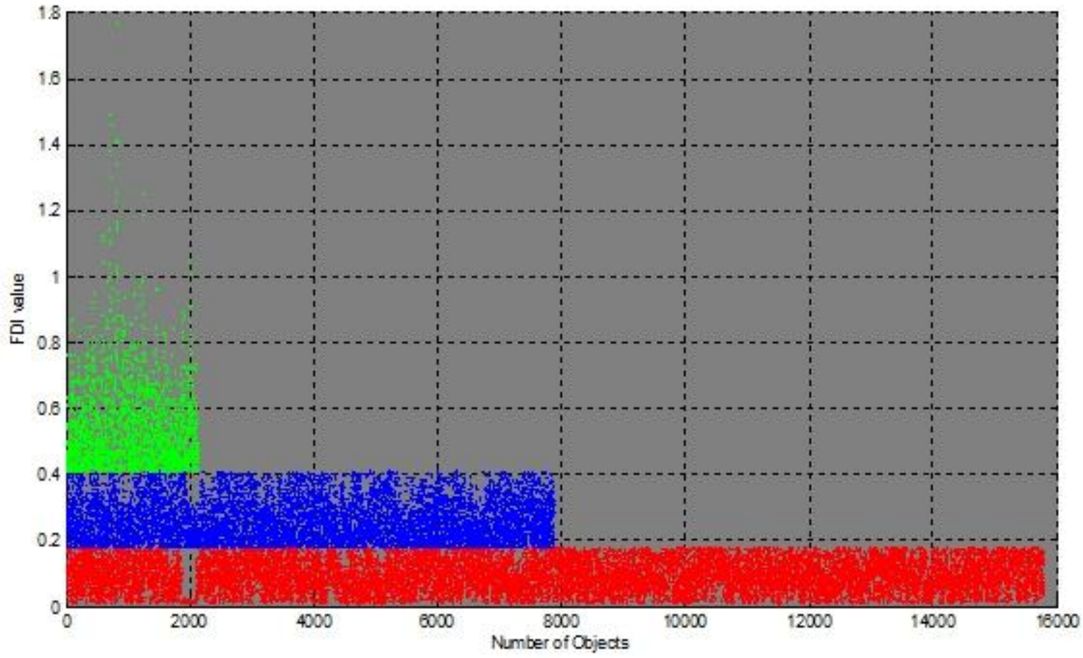
Εικόνα 23: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 7 κλάσεις με τη μέθοδο k-Means.

Πίνακας 9: Όρια κλάσεων (από 3 έως 7 κλάσεις) για τα δεδομένα της περιοχής των Σερρών με τη μέθοδο k-Means.

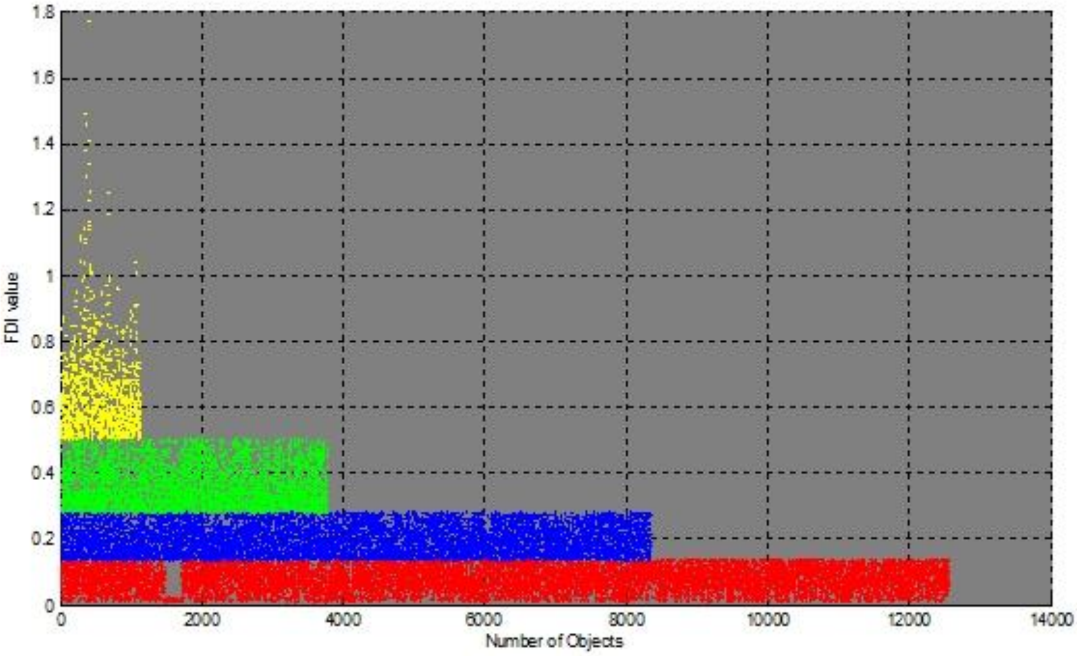
ΟΡΙΑ ΤΩΝ CLUSTERS ΓΙΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΡΙΟΧΗΣ ΤΩΝ ΣΕΡΡΩΝ					
Classes	*k=3*	*k=4*	*k=5*	*k=6*	*k=7*
A Class	0.0103-0.1129	0.0103-0.0916	0.0103-0.0662	0.0103-0. 0530	0.0103-0. 0477
B Class	0.1129-0.3059	0.0916-0.2094	0.0662-0.1531	0. 0530-0.1224	0. 0477-0.1085
C Class	0.3059-1.2545	0.2094-0.3959	0.1531-0.2727	0.1224-0.2070	0.1085-0.1781
D Class		0.3959-1.2545	0.2727-0.4795	0.2070-0.3248	0.1781-0.2609
E Class			0.4795-1.2545	0.3248-0.5324	0.2609-0.3813
F Class				0.5324-1.2545	0.3813-0.5819
G Class					0.5819-1.2545

Πίνακας 10: Αριθμός δεδομένων και ποσοστό συμμετοχής σε κάθε κλάση (από 3 έως 7 κλάσεις) για την περιοχή των Σερρών με τη μέθοδο k-Means.

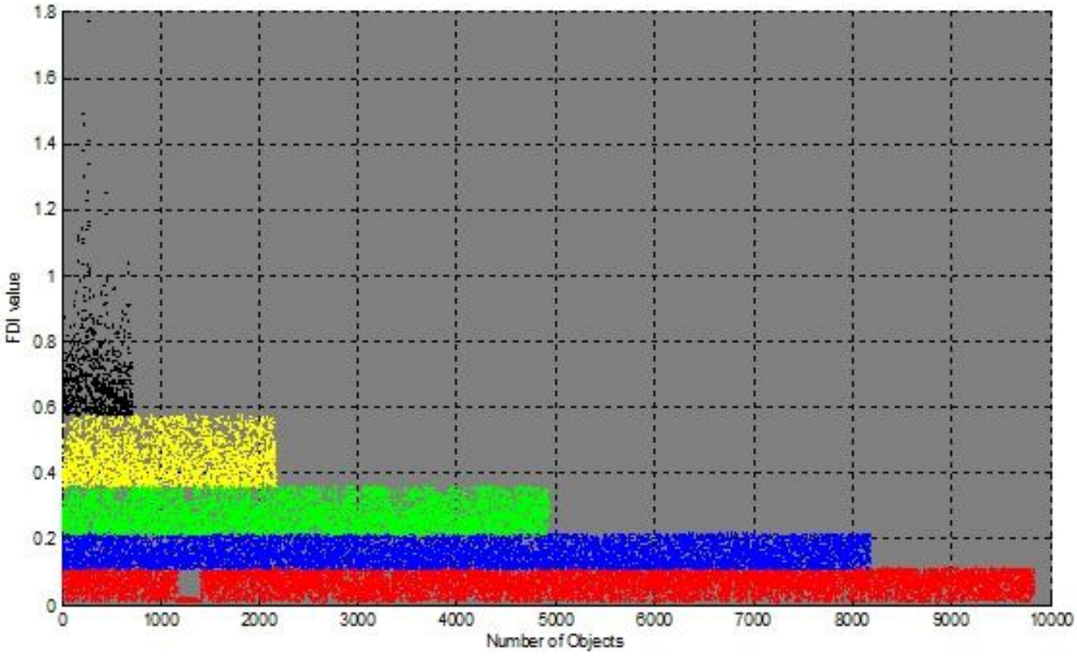
Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) για την περιοχή των Σερρών										
Classes	*k=3*		*k=4*		*k=5*		*k=6*		*k=7*	
A Class	16466	64.06	15219	59.21	12757	49.63	11382	44.28	10680	41.55
B Class	7869	30.61	6731	26.19	6214	24.18	5686	22.12	5545	21.57
C Class	1369	5.33	3123	12.15	4836	18.81	4826	18.78	4211	16.38
D Class			631	2.45	1590	6.19	2660	10.35	3141	12.22
E Class					307	1.19	944	3.67	1417	5.51
F Class							206	0.80	566	2.20
G Class									144	0.57



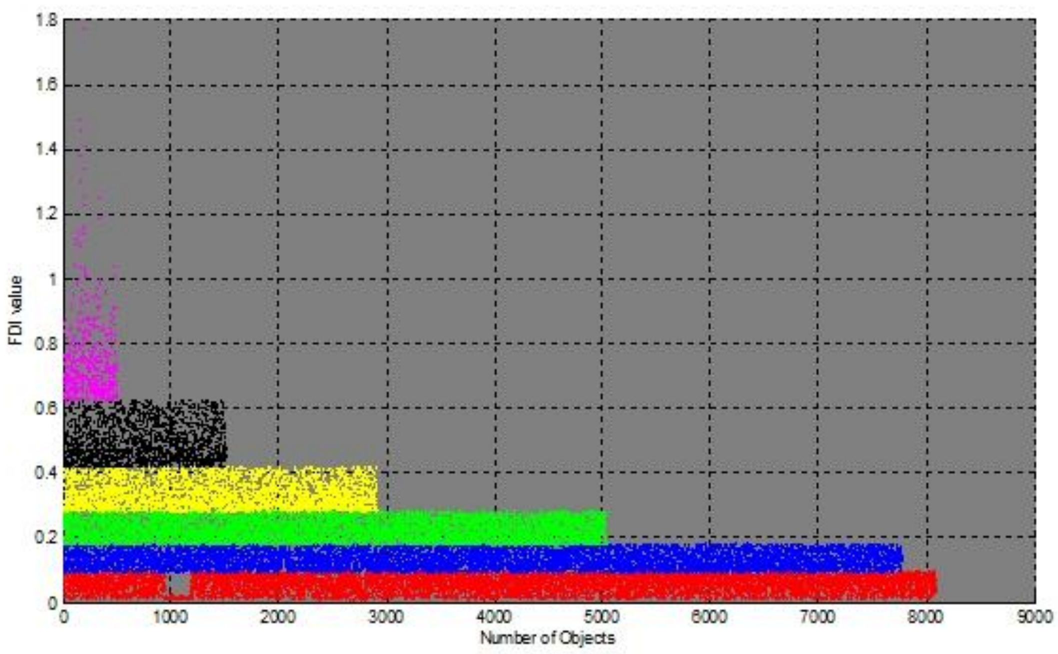
Εικόνα 24: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 3 κλάσεις με τη μέθοδο k-Means.



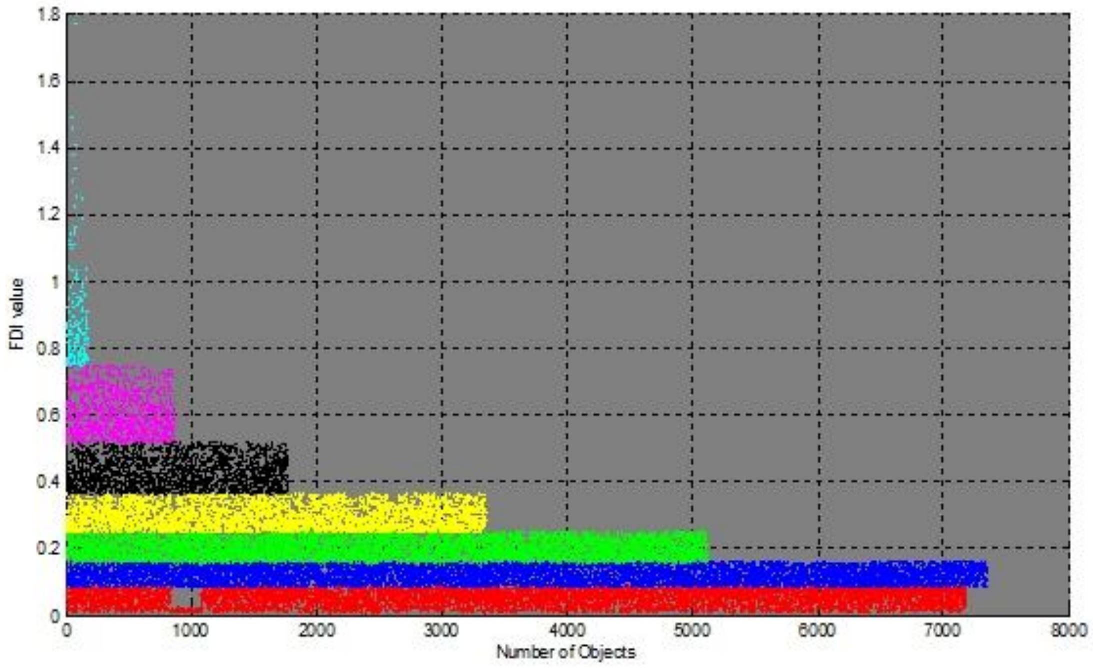
Εικόνα 25: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 4 κλάσεις με τη μέθοδο k-Means.



Εικόνα 26: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 5 κλάσεις με τη μέθοδο k-Means.



Εικόνα 27: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 6 κλάσεις με τη μέθοδο k-Means.



Εικόνα 28: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 7 κλάσεις με τη μέθοδο k-Means.

Πίνακας 11: Όρια κλάσεων (από 3 έως 7 κλάσεις) για τα δεδομένα της περιοχής της Ηλιούπολης με τη μέθοδο k-Means.

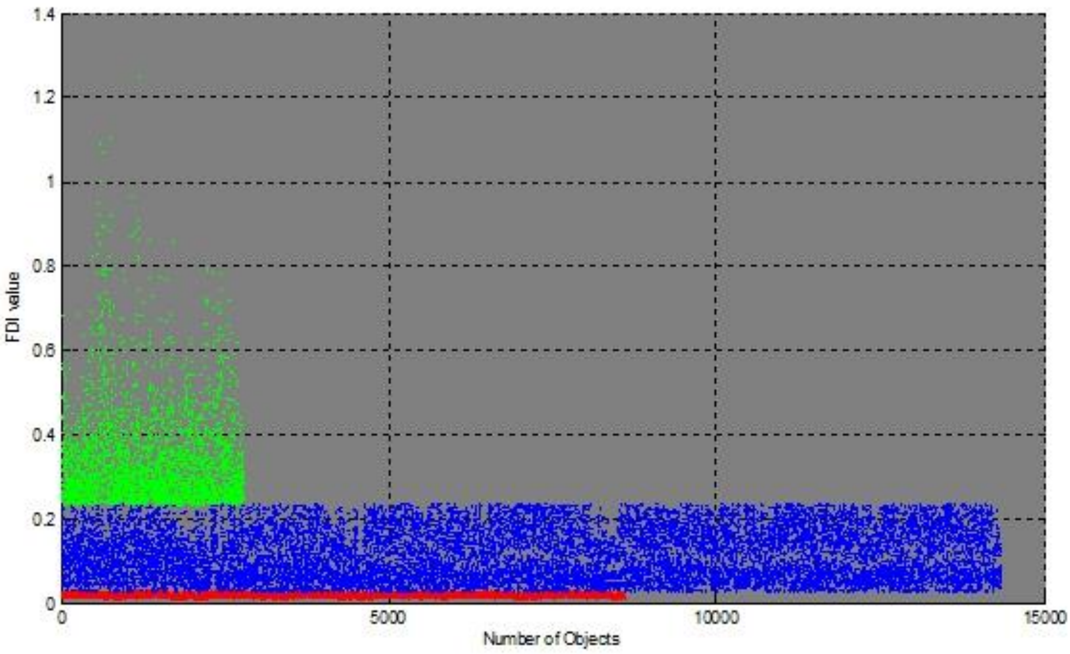
ΟΡΙΑ ΤΩΝ CLUSTERS ΓΙΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΡΙΟΧΗΣ ΤΗΣ ΗΛΙΟΥΠΟΛΗΣ					
Classes	*k=3*	*k=4*	*k=5*	*k=6*	*k=7*
A Class	0.0102-0. 1792	0.0102-0.1386	0.0102-0.1112	0.0102-0.0953	0.0102-0.0871
B Class	0.1792-0.4095	0.1386-0.2833	0.1112-0.2170	0.0953-0.1801	0.0871-0.1621
C Class	0.4095-1.7747	0.2833-0.5059	0.2170-0.3616	0.1801-0.2824	0.1621-0.2508
D Class		0.5059-1.7747	0.3616-0.5769	0.2824-0.4203	0.2508-0.3652
E Class			0.5769-1.7747	0.4203-0.6266	0.3652-0.5206
F Class				0.6266-1.7747	0.5206-0.7511
G Class					0.7511-1.7747

Πίνακας 12: Αριθμός δεδομένων και ποσοστό συμμετοχής σε κάθε κλάση (από 3 έως 7 κλάσεις) για την περιοχή της Ηλιούπολης με τη μέθοδο k-Means.

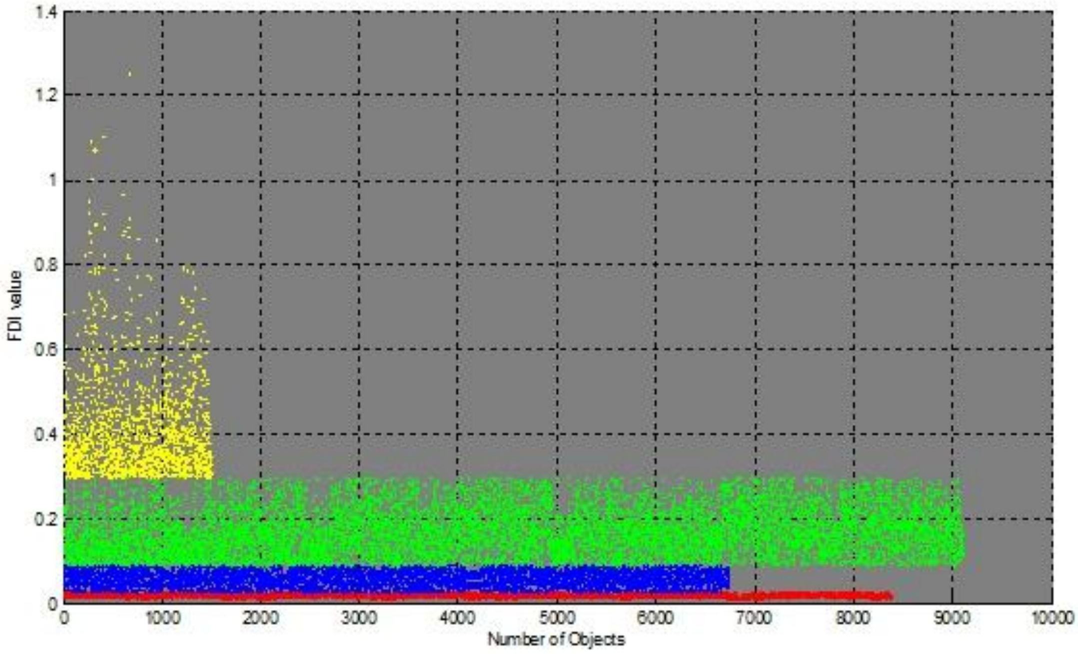
Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) της περιοχής της Ηλιούπολης										
Classes	*k=3*		*k=4*		*k=5*		*k=6*		*k=7*	
A Class	15806	61.32	12765	49.52	9898	38.40	8083	31.36	7177	27.84
B Class	7848	30.45	8264	32.06	8152	31.63	7766	30.13	7345	28.50
C Class	2122	8.23	3663	14.21	4886	18.96	5026	19.50	5113	19.84
D Class			1084	4.21	2137	8.29	2903	11.26	3342	12.97
E Class					703	2.72	1505	5.84	1770	6.87
F Class							493	1.91	859	3.32
G Class									170	0.66

4.4 Αποτελέσματα GMM

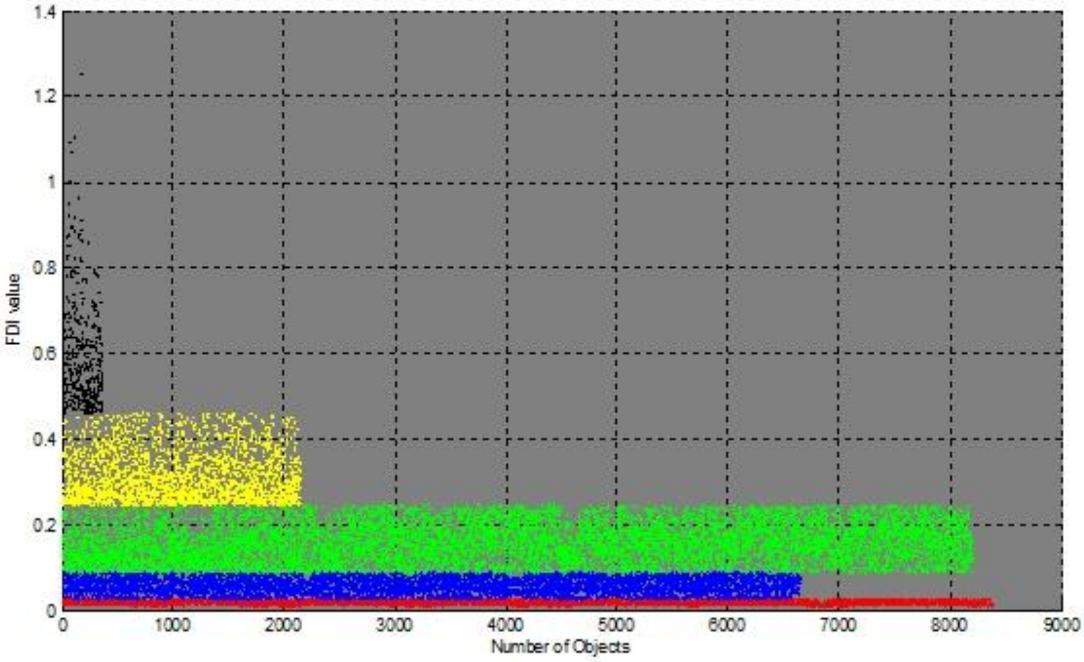
Όπως αναφέραμε σε προηγούμενη ενότητα, η μέθοδος GMM προϋποθέτει τον προσδιορισμό του αριθμού των clusters (27). Το Gaussian μοντέλο μπορεί να δημιουργηθεί με την εγκατάσταση ενός παραμετρικού μοντέλου με καθορισμένο αριθμό ομάδων. Στο Statistics Toolbox στο Matlab, η συνάρτηση ***gmdistribution*** τοποθετεί τα δεδομένα χρησιμοποιώντας τον Expectation Maximization (EM) αλγόριθμο. Πιο συγκεκριμένα, χρησιμοποιούμε την μέθοδο ***fit*** δίνοντας ως είσοδο τα δεδομένα (τον δείκτη επικινδυνότητας κάθε περιοχής) και τον αριθμό των ομάδων που θέλουμε να εξάγουμε. Η παράμετρος *'start'* καθορίζει την μέθοδο αρχικοποίησης των ομάδων και ως τιμή δίνουμε ένα διάνυσμα το οποίο περιέχει μια αρχική εκτίμηση στο πώς θα κατανεμηθούν τα δεδομένα μας. Πιο συγκεκριμένα, το διάνυσμα αυτό περιέχει τα αποτελέσματα που εξάγαμε από τη μέθοδο k-Means. Στη συνέχεια, για να δημιουργήσουμε τις ομάδες από το GMM, χρησιμοποιούμε την μέθοδο ***cluster*** η οποία χωρίζει τα δεδομένα σε ένα πίνακα διάστασης $n_x b$ όπου n είναι ο αριθμός των δεδομένων και b είναι η διάσταση των ήδη ταξινομημένων δεδομένων. Το κάθε στοιχείο εντάσσεται στην ομάδα με τη μεγαλύτερη αποστεριόρι πιθανότητα. Τα αποτελέσματα της παραπάνω μεθόδου φαίνονται παρακάτω :



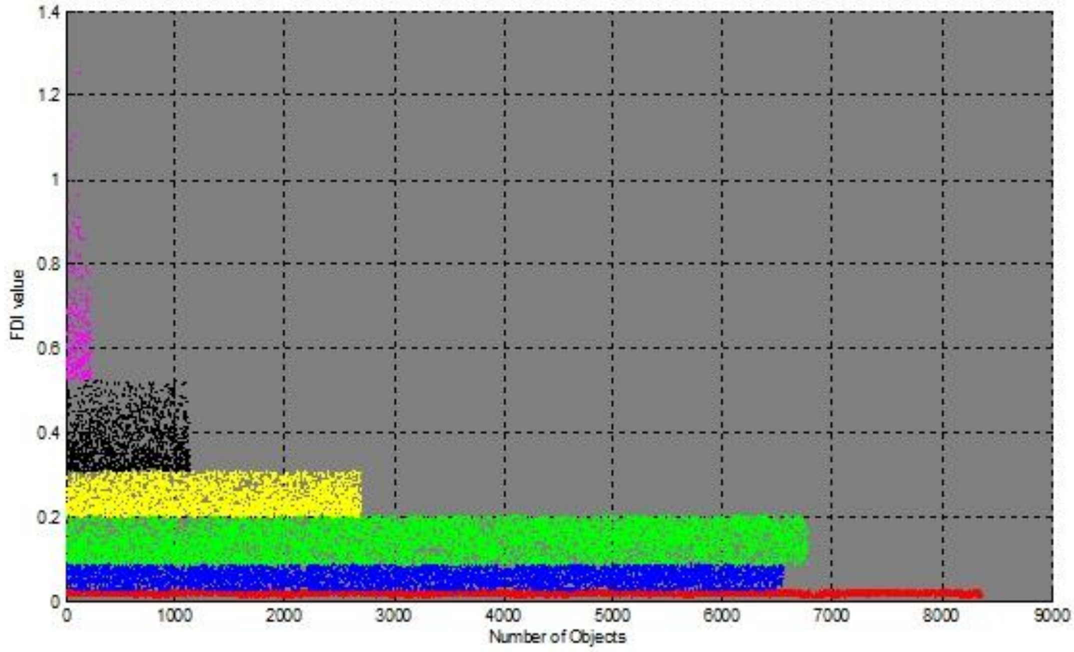
Εικόνα 29: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 3 κλάσεις με τη μέθοδο GMM.



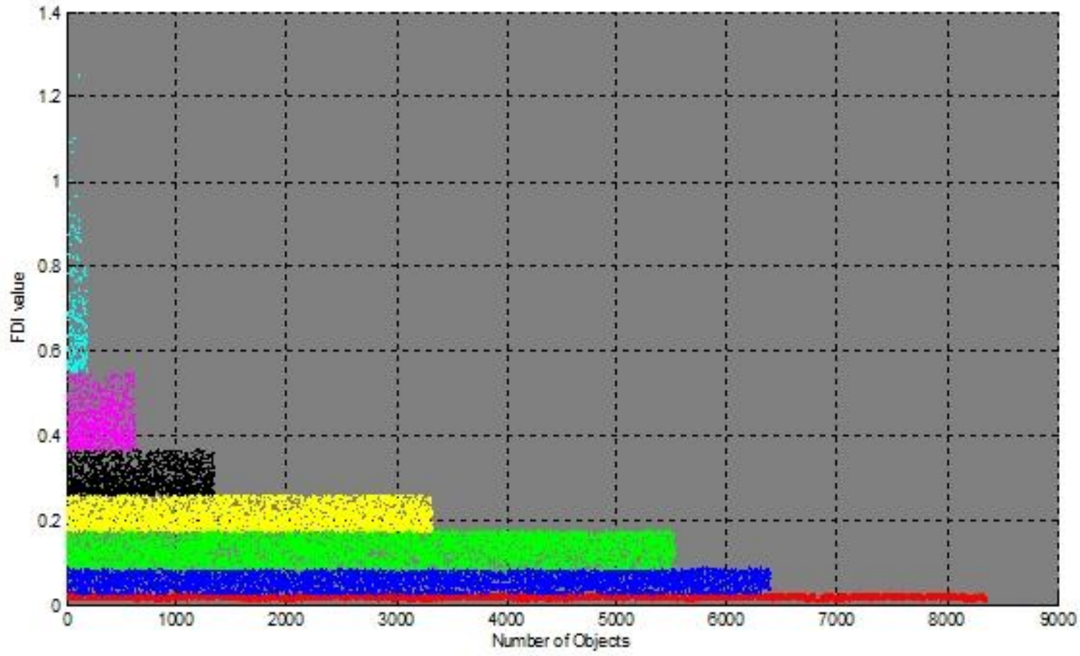
Εικόνα 30: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 4 κλάσεις με τη μέθοδο GMM.



Εικόνα 31: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 5 κλάσεις με τη μέθοδο GMM.



Εικόνα 32: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 6 κλάσεις με τη μέθοδο GMM.



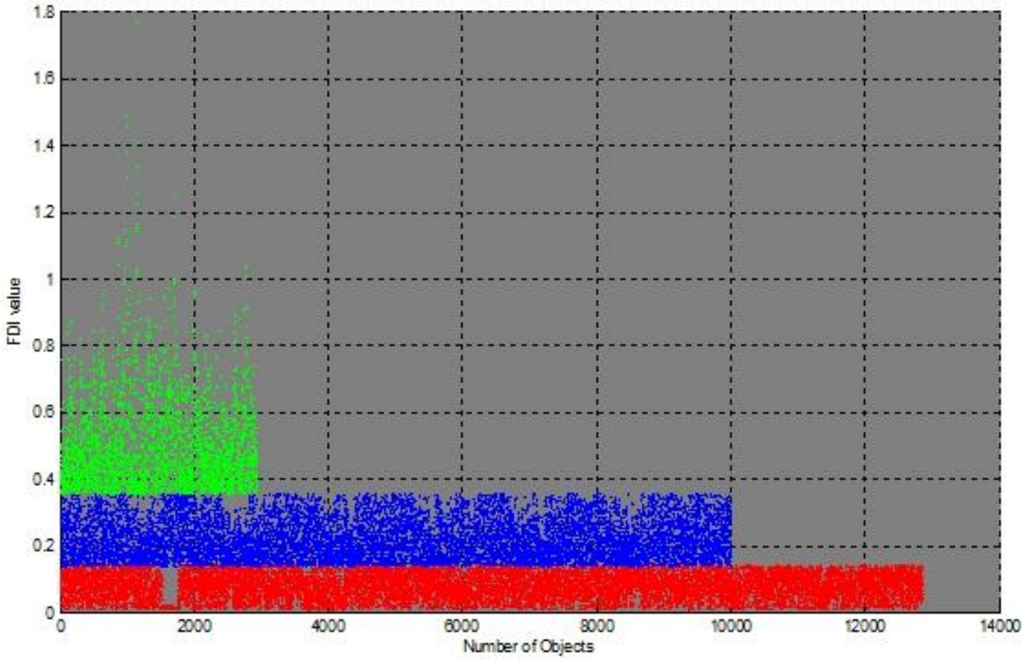
Εικόνα 33: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής των Σερρών για 7 κλάσεις με τη μέθοδο GMM.

Πίνακας 13: Όρια κλάσεων (από 3 έως 7 κλάσεις) για τα δεδομένα της περιοχής των Σερρών με τη μέθοδο GMM.

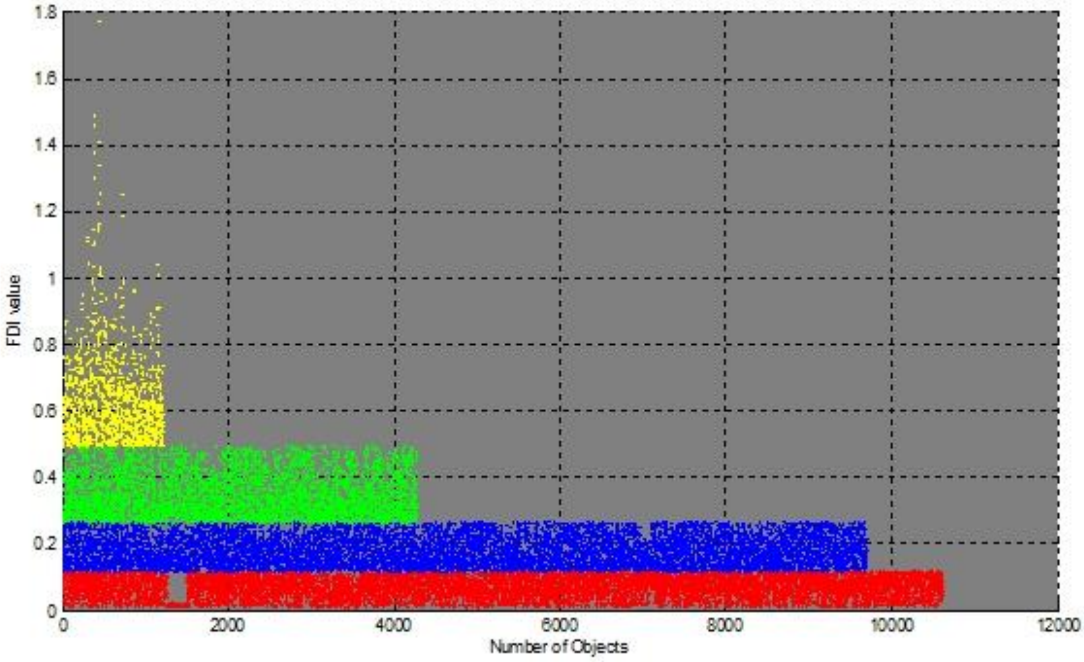
ΟΡΙΑ ΤΩΝ CLUSTERS ΓΙΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΡΙΟΧΗΣ ΤΩΝ ΣΕΡΡΩΝ					
Classes	*k=3*	*k=4*	*k=5*	*k=6*	*k=7*
A Class	0.0103-0.0283	0.0103-0.0269	0.0103-0.0269	0.0103-0.0269	0.0103-0.0268
B Class	0.0283-0.2354	0.0269-0.0903	0.0269-0.0893	0.0269-0.0876	0.0268-0.0855
C Class	0.2354-1.2545	0.0903-0.2971	0.0893-0.2457	0.0876-0.2017	0.0855-0.1743
D Class		0.2971-1.2545	0.2457-0.4610	0.2017-0.3081	0.1743-0.2604
E Class			0.4610-1.2545	0.3081-0.5225	0.2604-0.3668
F Class				0.5225-1.2545	0.3668-0.5472
G Class					0.5472-1.2545

Πίνακας 14: Αριθμός δεδομένων και ποσοστό συμμετοχής σε κάθε κλάση (από 3 έως 7 κλάσεις) για την περιοχή των Σερρών με τη μέθοδο GMM.

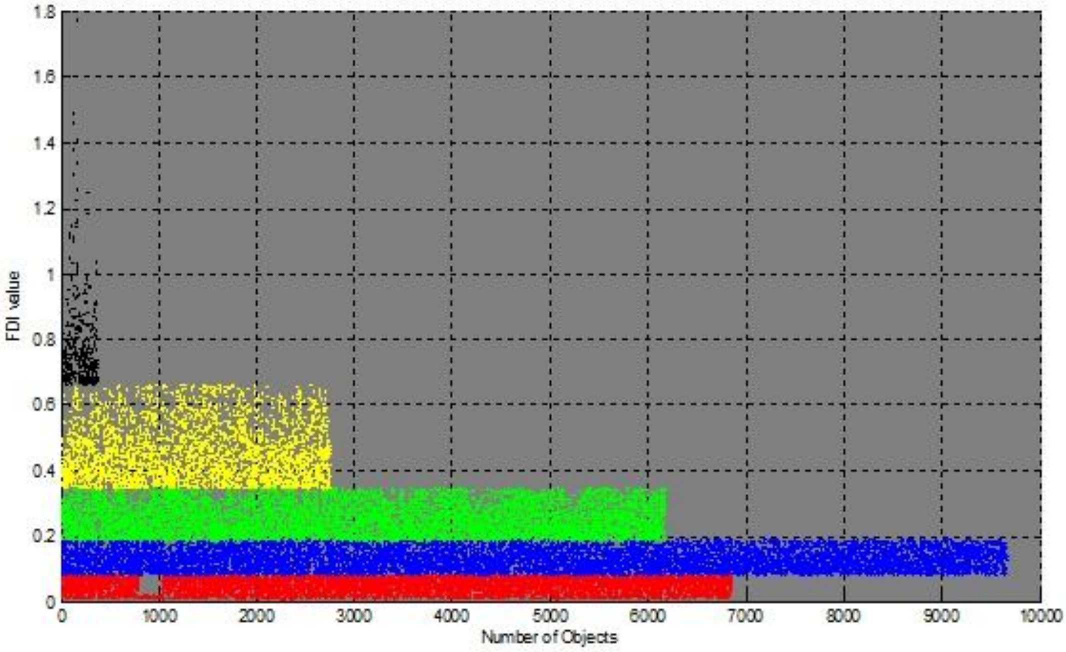
Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) της περιοχής των Σερρών										
Classes	*k=3*		*k=4*		*k=5*		*k=6*		*k=7*	
A Class	8599	33.45	8377	32.59	8372	32.57	8361	32.53	8355	32.50
B Class	14292	55.60	6717	26.13	6637	25.82	6538	25.44	6380	24.82
C Class	2813	10.95	9117	35.47	8257	31.73	6758	26.29	5519	21.47
D Class			1493	5.81	2176	8.47	2701	10.51	3313	12.89
E Class					362	1.41	1122	4.36	1334	5.19
F Class							224	0.87	617	2.40
G Class									186	0.73



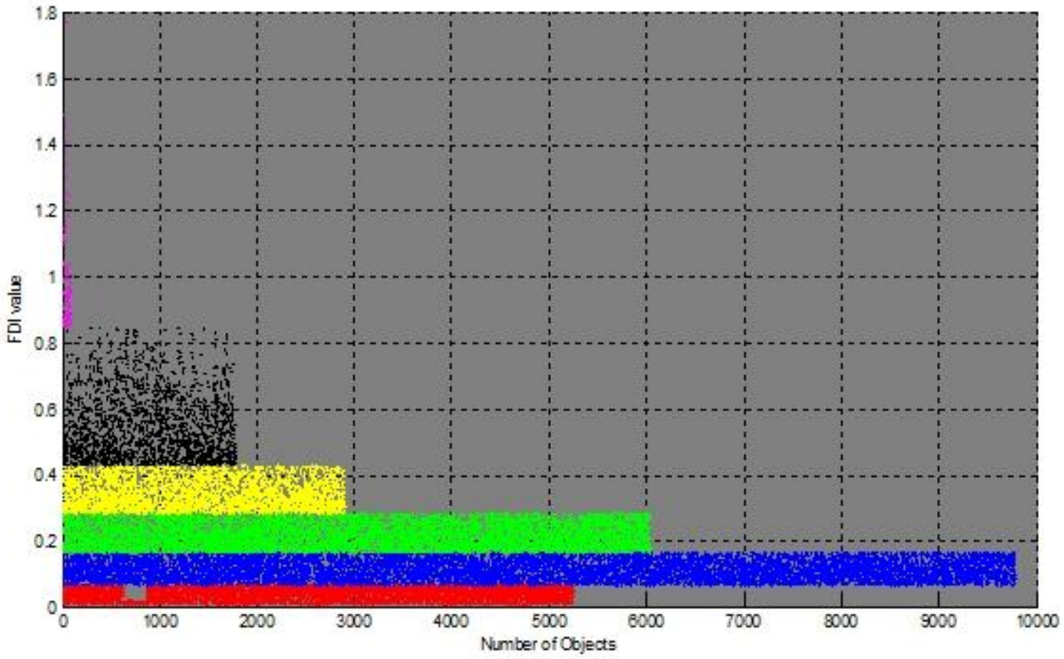
Εικόνα 34: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 3 κλάσεις με τη μέθοδο GMM.



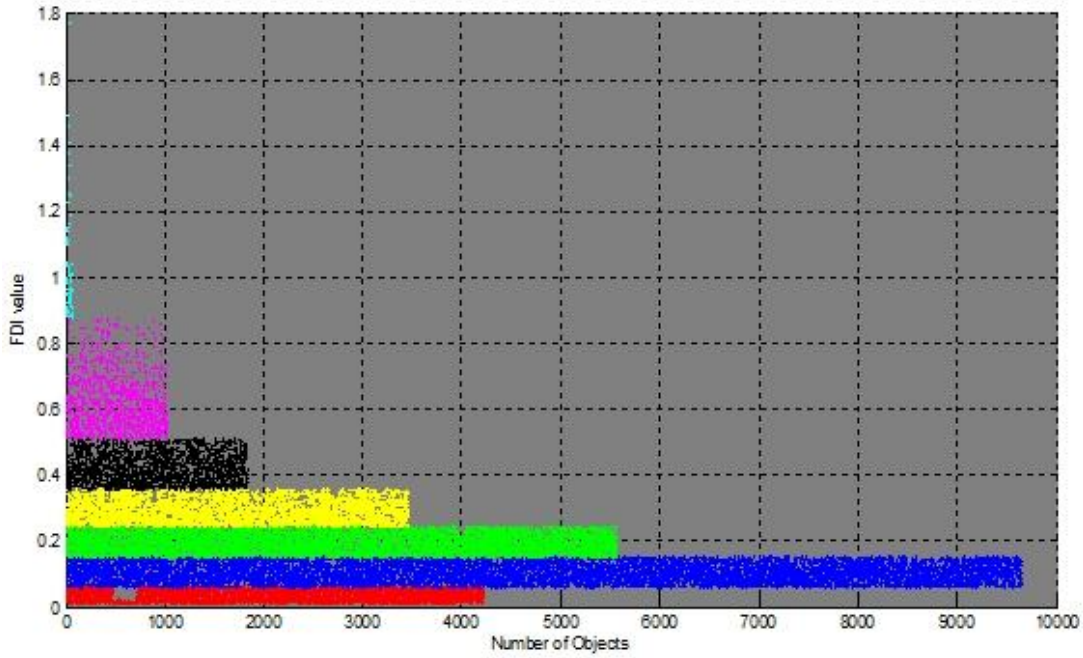
Εικόνα 35: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 4 κλάσεις με τη μέθοδο GMM.



Εικόνα 36: : Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 5 κλάσεις με τη μέθοδο GMM.



Εικόνα 37: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 6 κλάσεις με τη μέθοδο GMM.



Εικόνα 38: Διάγραμμα δείκτη επικινδυνότητας (FDI) της περιοχής της Ηλιούπολης για 7 κλάσεις με τη μέθοδο GMM.

Πίνακας 15: Όρια κλάσεων (από 3 έως 7 κλάσεις) για τα δεδομένα της περιοχής της Ηλιούπολης με τη μέθοδο GMM.

ΟΡΙΑ ΤΩΝ CLUSTERS ΓΙΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΡΙΟΧΗΣ ΤΗΣ ΗΛΙΟΥΠΟΛΗΣ					
Classes	*k=3*	*k=4*	*k=5*	*k=6*	*k=7*
A Class	0.0102-0.1416	0.0102-0.1186	0.0102-0.0844	0.0102-0.0705	0.0102-0.0612
B Class	0.1416-0.3573	0.1186-0.2669	0.0844-0.1908	0.0705-0.1685	0.0612-0.1532
C Class	0.3573-1.7747	0.2669-0.4959	0.1908-0.3477	0.1685-0.2873	0.1532-0.2456
D Class		0.4959-1.7747	0.3477-0.6629	0.2873-0.4327	0.2456-0.3590
E Class			0.6629-1.7747	0.4327-0.8506	0.3590-0.5119
F Class				0.8506-1.7747	0.5119-0.8787
G Class					0.8787-1.7747

Πίνακας 16: Αριθμός δεδομένων και ποσοστό συμμετοχής σε κάθε κλάση (από 3 έως 7 κλάσεις) για την περιοχή της Ηλιούπολης με τη μέθοδο GMM.

Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) της περιοχής της Ηλιούπολης										
Classes	*k=3*		*k=4*		*k=5*		*k=6*		*k=7*	
A Class	12815	49.72	8377	32.59	6847	26.56	5239	20.33	4109	15.94
B Class	10020	38.87	6717	26.13	9652	37.45	9774	37.92	9664	37.49
C Class	2941	11.41	9117	35.47	6165	23.92	6024	23.37	5561	21.57
D Class			1493	5.81	2744	10.65	2891	11.22	3515	13.64
E Class					368	1.42	1771	6.86	1851	7.18
F Class							77	0.30	1016	3.94
G Class									60	0.24

4.5 Συμπεράσματα κεφαλαίου

Με μια πρώτη ματιά στα διαγράμματα του δείκτη επικινδυνότητας καθώς και στους παραπάνω πίνακες, παρατηρούμε ότι τα αποτελέσματα της Fuzzy c-Means μεθόδου είναι αρκετά κοντά στα αποτελέσματα της k-Means μεθόδου για τις δυο περιοχές, ενώ τα αποτελέσματα της GMM μεθόδου έχουν μεγάλη απόκλιση για την περιοχή των Σερρών ενώ είναι την περιοχή της Ηλιούπολης έχουν αρκετά μικρότερη απόκλιση απο τις άλλες δυο μεθόδους. Επισής αξιοσημείοτο είναι το γεγονός ότι ενώ μελετήθηκε ο αλγόριθμος Hierarchical στο προηγούμενο κεφάλαιο, δεν παραθέτονται αποτελέσματα σε αυτό το κεφάλαιο. Αυτό σημβαίνει καθώς τα δεδομένα του δείκτη επικινδυνότητας FDI είναι πάρα πολλά ώστε να μπορεί να τρέξει ο αλγόριθμος σε έναν κοινό υπολογιστή χωρίς να παρουσιάσει σφάλμα ανεπαρκούς μνήμης. Στο επόμενο κεφάλαιο τα αποτελέσματα αναλύονται περισσότερο με σκοπό την επιλογή του καλύτερου αλγορίθμου και του βέλτιστου αριθμου clusters και τέλος την εξαγωγή των 5 κατηγοριών επικινδυνότητας που χρειάζονται για την δημιουργία χάρτη.

Κεφάλαιο5-Συγκριτική ανάλυση αλγορίθμων

5.1 Εισαγωγή

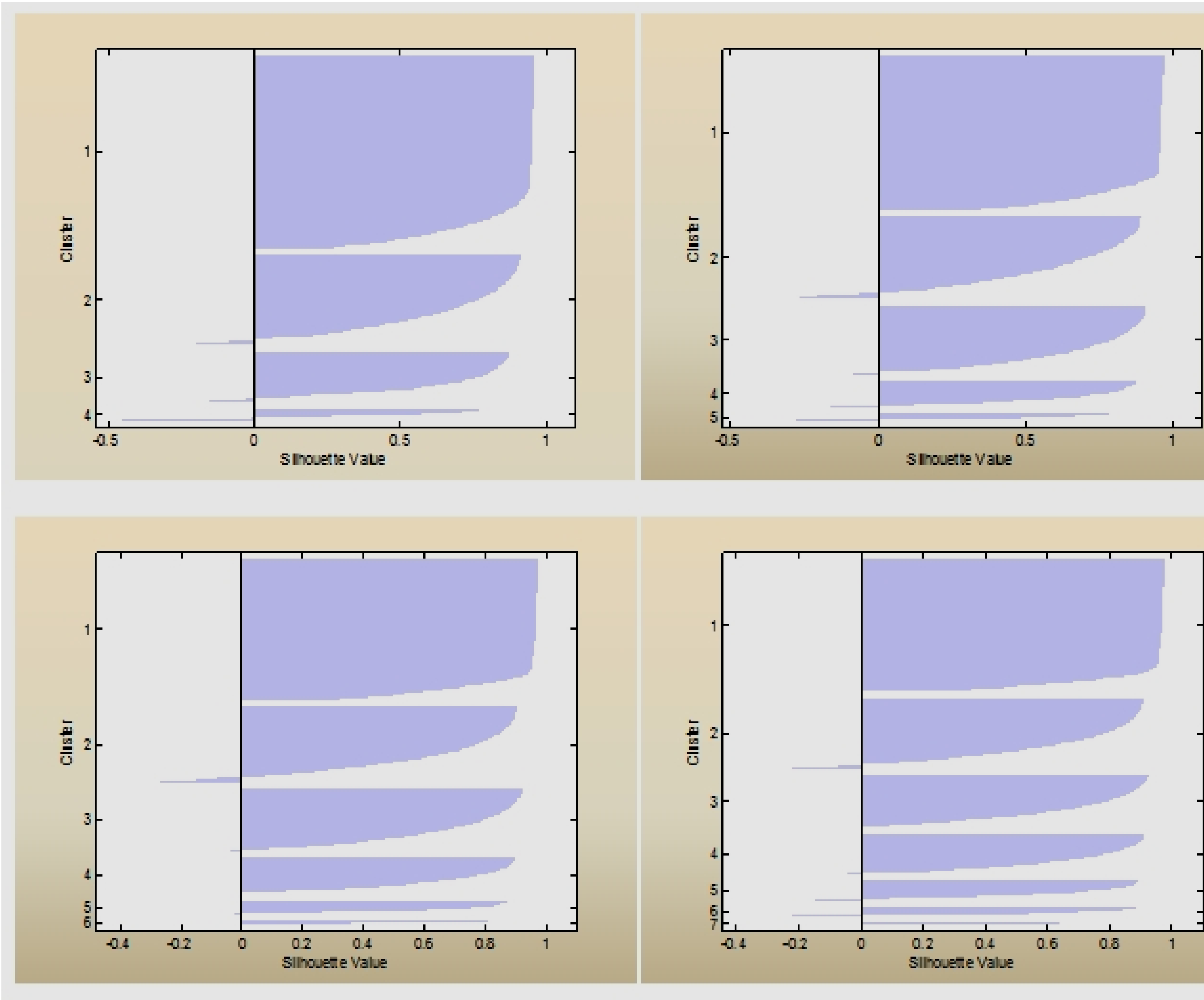
Όλες οι περιπτώσεις στις οποίες εφαρμόστηκαν οι αλγόριθμοι ομαδοποίησης και παρουσιάστηκαν στο Κεφάλαιο 4, ελέγχθηκαν με κανόνες επικύρωσης με σκοπό να βρεθεί αρχικά ποιος είναι ο αριθμός των clusters ο οποίος ταιριάζει καλύτερα στα δεδομένα και στη συνέχεια να επιλεχθεί, στις περιπτώσεις που αυτό είναι δυνατό, ο αλγόριθμος με την πιο αποδοτική ομαδοποίηση στα δεδομένα. Η αξιολόγηση των διαμερίσεων που προέκυψαν από την εφαρμογή των αλγορίθμων ομαδοποίησης έγινε με τη χρήση των σχετικών κριτηρίων:

- Silhouette
- Davies – Bouldin

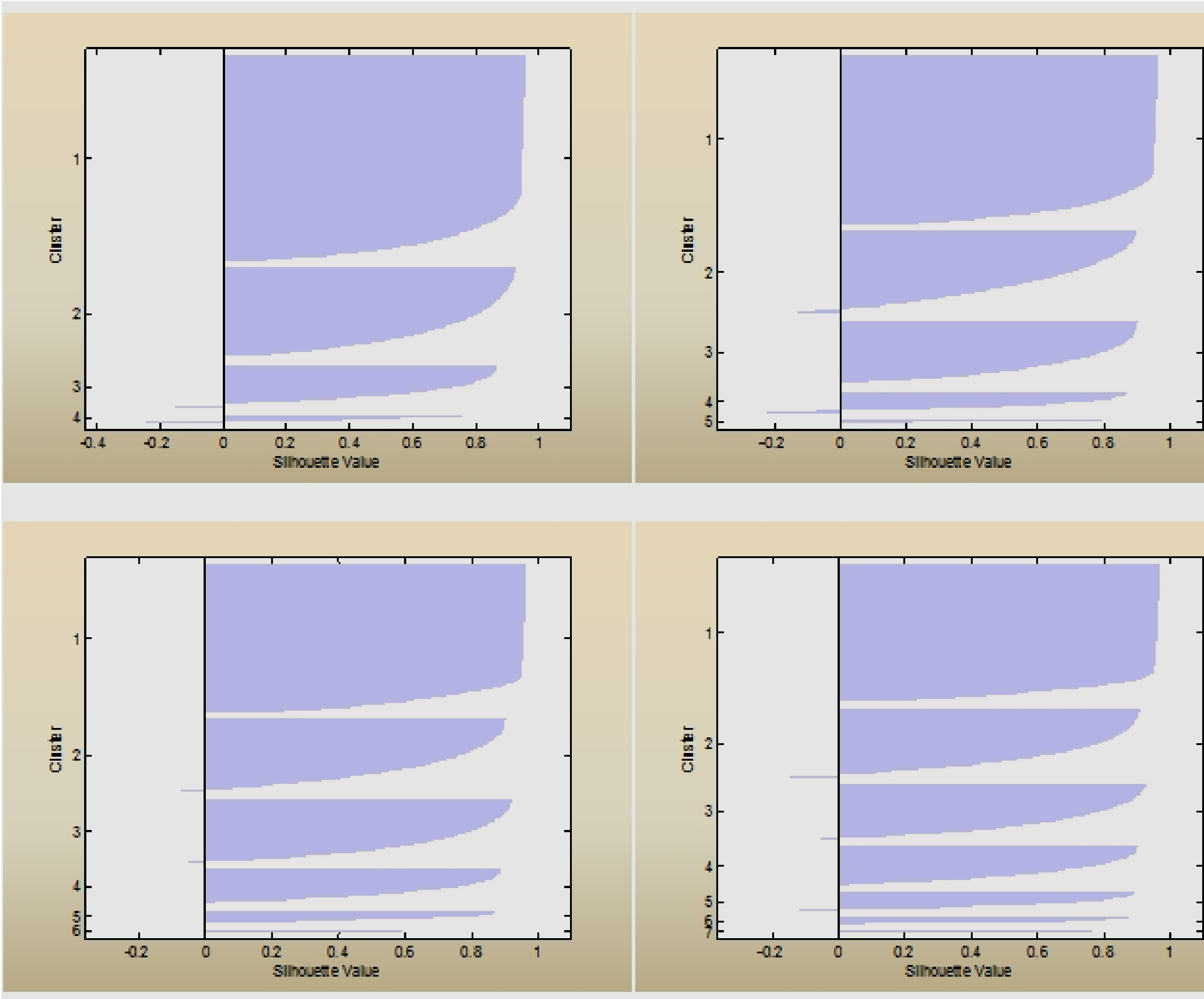
Στο σημείο αυτό θα πρέπει να τονιστεί ότι σε κάποιες περιπτώσεις οι κανόνες επικύρωσης θα πρέπει να συνδυάζονται και με την φύση του εκάστοτε προβλήματος ώστε το τελικό αποτέλεσμα να ανταποκρίνεται όσο το δυνατόν καλύτερα στις ανάγκες του προβλήματος. Στην περίπτωση μας, η προτεινόμενη ταξινόμηση θα πρέπει να συμβαδίζει με την Γενική Γραμματεία Πολιτικής Προστασίας ώστε να μπορεί να εφαρμοστεί στην έκδοση δείκτη επικινδυνότητας έναρξης πυρκαγιάς και στην σύγκριση με τον ήδη υπάρχον δείκτη. Για τον λόγο αυτό, αν και όλοι οι αλγόριθμοι ομαδοποίησης εφαρμόστηκαν για αριθμό clusters από 3 έως και 7, ο αριθμός clusters που θα χρησιμοποιηθεί για περαιτέρω μελέτη και προσπάθεια επίλυσης του προβλήματος θα είναι 5 όσες δηλαδή και οι κατηγορίες επικινδυνότητας που εκδίδει η ΓΓΠΠ.

Στη συνέχεια ακολουθούν τα γραφήματα silhouette και οι πίνακες με τις τιμές των δεικτών επικύρωσης για κάθε αλγόριθμο ομαδοποίησης για την περιοχή των Σερρών και για την περιοχή της Ηλιούπολης. Κάθε πίνακας περιλαμβάνει την τιμή των δεικτών επικύρωσης για τους τρείς αλγορίθμους ομαδοποίησης και για όλες τις περιπτώσεις αριθμών των clusters που δοκιμάστηκαν (από 3 έως και 7). Για κάθε δείκτη σε κάθε αλγόριθμο, η βέλτιστη τιμή του, που δηλώνει και τον βέλτιστο αριθμό των clusters, εμφανίζεται με πιο έντονο χρώμα. Υπενθυμίζεται ότι για τους δείκτες Silhouette μεγάλες τιμές δηλώνουν καλύτερες διαμερίσεις ενώ για τον δείκτη Davies-Bouldin οι τιμές αυτές θα πρέπει να είναι όσο το δυνατόν μικρότερες.

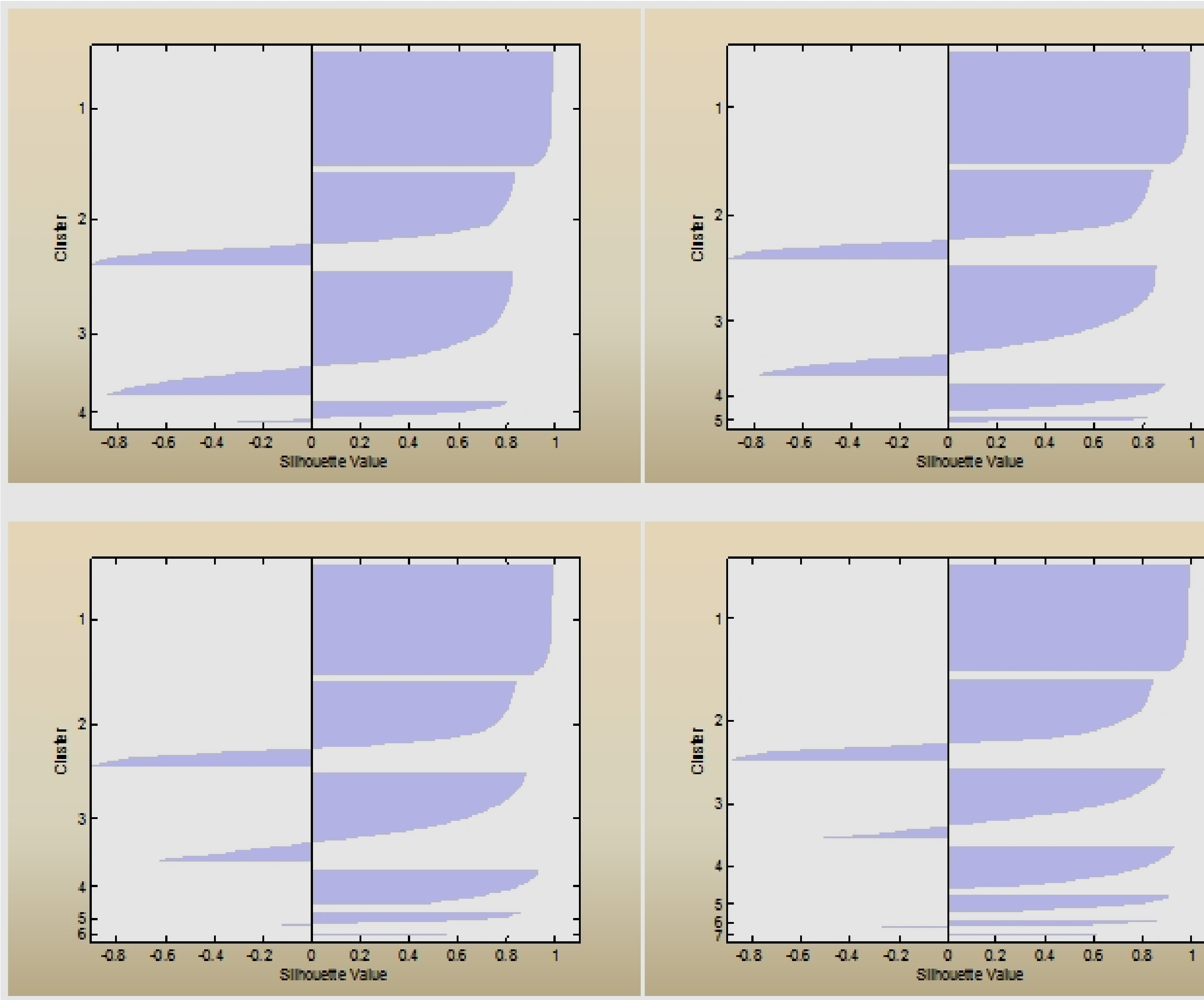
5.2 Σύγκριση διαμερίσεων δείκτη επικινδυνότητας - περιοχή Σερρών.



Εικόνα 39: Γραφικές παραστάσεις Silhouette για αριθμό clysters από 4 έως 7 για την περιοχή των Σερρών με τη μέθοδο Fuzzy c-Means.



Εικόνα 40: : Γραφικές παραστάσεις Silhouette για αριθμό clysters από 4 έως 7 για την περιοχή των Σερρών με τη μέθοδο k-Means .



Εικόνα 41: Γραφικές παραστάσεις Silhouette για αριθμό clysters από 4 έως 7 για την περιοχή των Σερρών με τη μέθοδο GMM .

Πίνακας 17: Δείκτες επικύρωσης για την περιοχή των Σερρών.

Fuzzy c-Means Clustering

	Αριθμός των clusters				
	3	**4**	**5**	**6**	**7**
Silhouette Index	**0.8042**	0.7690	0.7582	0.7652	0.7669
Davies – Bouldin Index	0.5106	0.5256	0.5128	0.4950	**0.4926**

k-Means Clustering

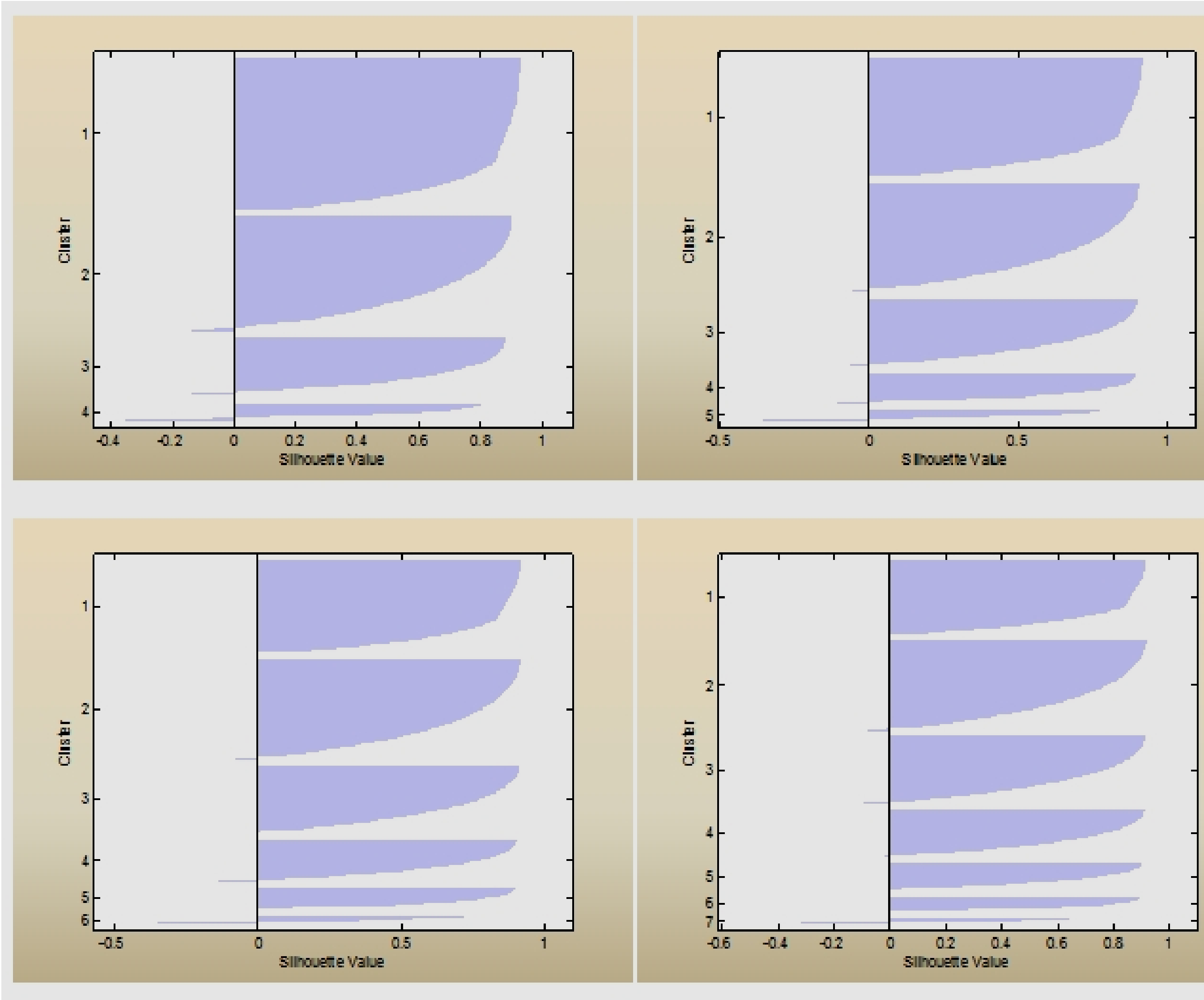
	Αριθμός των clusters				
	3	**4**	**5**	**6**	**7**
Silhouette Index	**0.8117**	0.7898	0.7607	0.7682	0.7693
Davies – Bouldin Index	0.5007	0.5137	0.5128	0.4946	**0.4893**

GMM Clustering

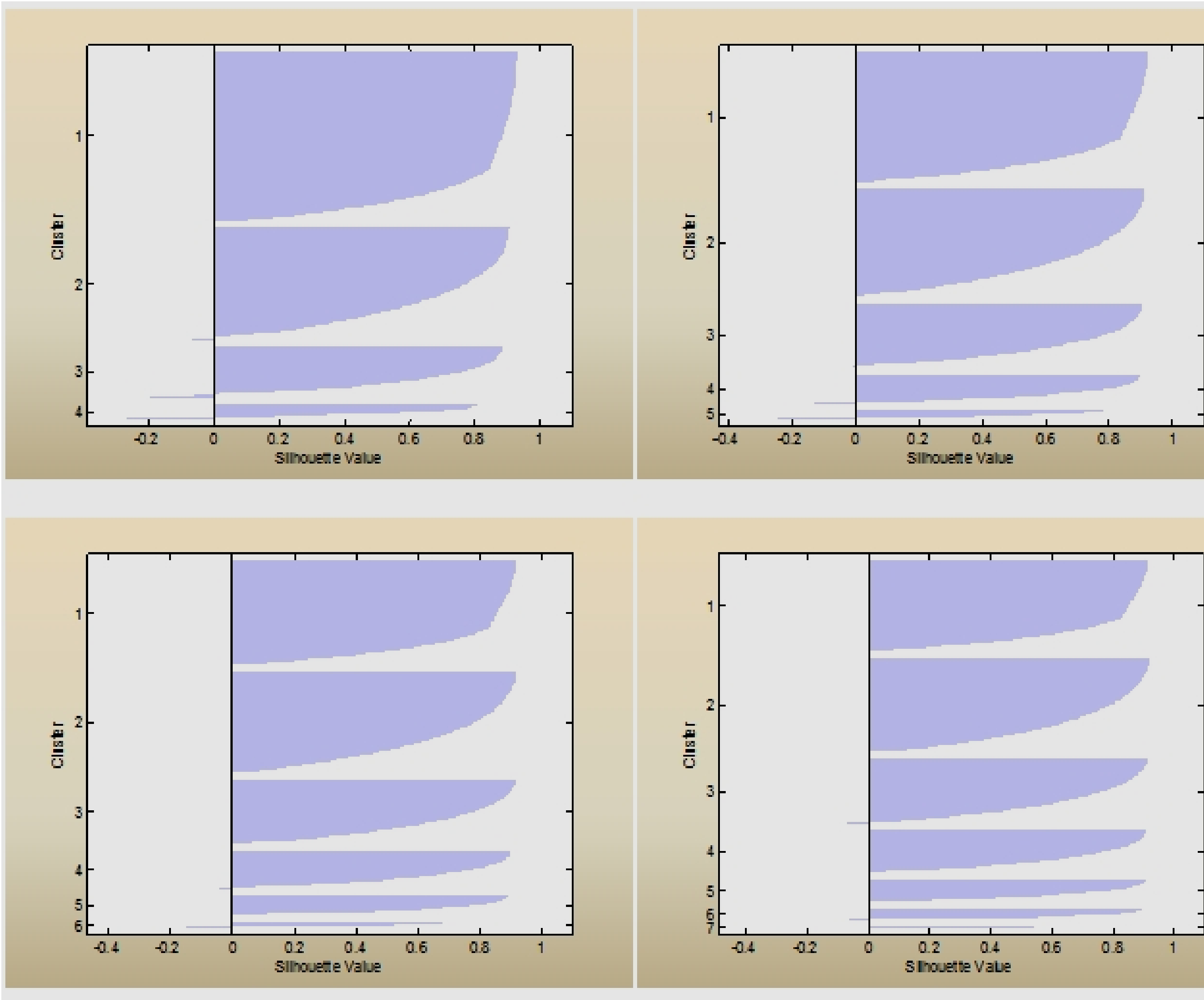
	Αριθμός των clusters				
	3	**4**	**5**	**6**	**7**
Silhouette Index	0.4203	0.5965	0.6319	0.6509	**0.6745**
Davies – Bouldin Index	0.5816	0.5128	0.5070	**0.5061**	0.5088

Όπως παρατηρούμε από τις παραπάνω εικόνες, τα γραφήματα του δείκτη silhouette με τη μέθοδο Fuzzy c-Means (Εικόνα 39) παρουσιάζονται αρκετά ικανοποιητικά για τους διάφορους αριθμούς clusters καθώς οι αρνητικές τους τιμές είναι ελάχιστες και έτσι συνεπάγεται ότι τα δεδομένα είναι καλά διαχωρισμένα. Αυτό φαίνετε και από τη μέση τιμή του δείκτη silhouette (Πίνακας 17) η οποία είναι αρκετά υψηλή για τις διάφορες κατηγορίες με μεγαλύτερη τιμή να έχουμε για αριθμό clusters ίσο με 3 (0.8042). Όσον αφορά τον δείκτη Davies – Bouldin , η μικρότερη τιμή παρουσιάζεται για αριθμό clusters ίσο με 7 (0.4926). Ακριβώς τα ίδια αποτελέσματα έχουμε και με τη μέθοδο k-Means (Εικόνα 40) με λίγο καλύτερες τιμές για τον δείκτη silhouette (0.8117) και για τον δείκτη dabies-Bouldin (0.4893) σε σχέση με την FCM μέθοδο. Για τη μέθοδο GMM (Εικόνα 41) οι αρνητικές τιμές είναι αρκετά μεγάλες. Έτσι η μέση τιμή του δείκτη silhouette είναι αρκετά πιο χαμηλή (0.6745 για 7 clusters) και η τιμή του δείκτη Davies – Bouldin αρκετά υψηλή (0.5061 για 6 clusters) συμπεραίνοντας ότι τα δεδομένα δεν διαχωρίστηκαν επιτυχώς.

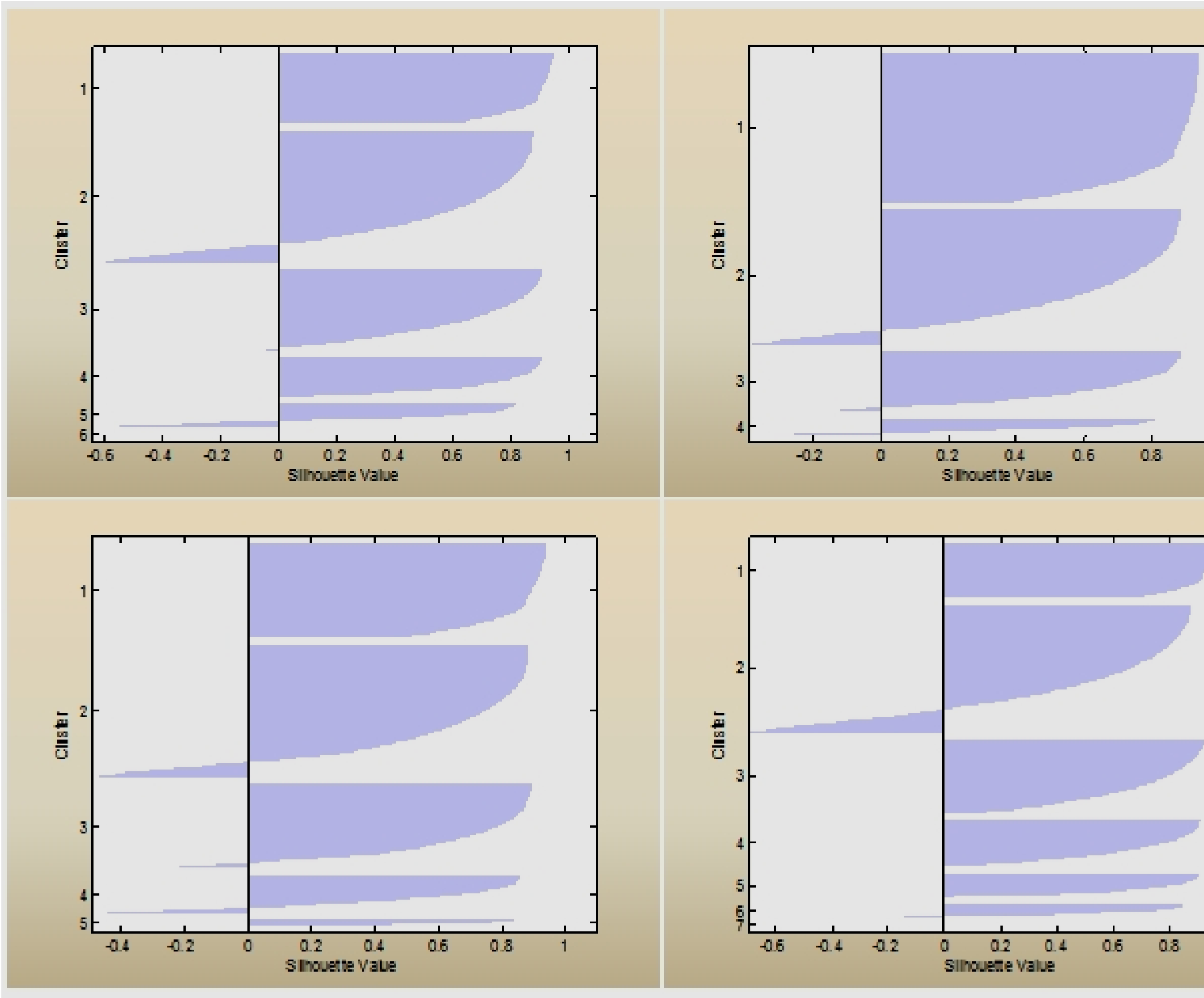
5.3 Σύγκριση διαμερίσεων δείκτη επικινδυνότητας - περιοχή Ηλιούπολης .



Εικόνα 42: Γραφικές παραστάσεις Silhouette για αριθμό clysters από 4 έως 7 για την περιοχή της Ηλιούπολης με τη μέθοδο Fuzzy c-Means .



Εικόνα 43: Γραφικές παραστάσεις Silhouette για αριθμό clysters από 4 έως 7 για την περιοχή της Ηλιούπολης με τη μέθοδο k-Means .



Εικόνα 44: Γραφικές παραστάσεις Silhouette για αριθμό clysters από 4 έως 7 για την περιοχή της Ηλιούπολης με τη μέθοδο GMM .

Πίνακας 18: Δείκτες επικύρωσης για την περιοχή της Ηλιούπολης.

Fuzzy c-Means Clustering

	Αριθμός των clusters				
	3	**4**	**5**	**6**	**7**
Silhouette Index	**0.7455**	0.7087	0.7018	0.6929	0.6837
Davies – Bouldin Index	0.5366	0.5380	0.5269	**0.5261**	0.5312

K-means Clustering

	Αριθμός των clusters				
	3	**4**	**5**	**6**	**7**
Silhouette Index	**0.7566**	0.7244	0.7081	0.6985	0.7001
Davies – Bouldin Index	0.5300	0.5340	0.5255	0.5280	**0.5237**

GMM Clustering

	Αριθμός των clusters				
	3	**4**	**5**	**6**	**7**
Silhouette Index	0.6977	**0.6969**	0.6655	0.6428	0.6383
Davies – Bouldin Index	0.5529	0.5367	0.5259	0.5165	**0.5101**

Αναλύοντας τις παραπάνω εικόνες με τα silhouette γραφήματα, διαπιστώνουμε ότι για τους αλγόριθμους Fuzzy c-Means (Εικόνα 42) και k-Means (Εικόνα 43) ισχύει σε γενικές γραμμές ότι και με τα δεδομένα της περιοχής των Σερρών. Οι αρνητικές τιμές είναι ελάχιστες για τους διάφορους αριθμούς των clusters με μια τάση να μειώνονται όλο και περισσότερο καθώς ο αριθμός τους αυξάνετε (εμφανή μείωση των αρνητικών τιμών για αριθμό clusters ίσο με 7). Μεγαλύτερη μέση τιμή για τον δείκτη silhouette παρατηρείτε για αριθμό clusters ίσο με 3 και στους δυο αλγόριθμους (0.7455 για τον FCM και 0.7566 για τον k-Means) και μικρότερη τιμή για τον δείκτη Davies – Bouldin παρατηρείτε για αριθμό clusters ίσο με 6 και 7 για τη μέθοδο Fuzzy c-Means και k-Means αντίστοιχα (0.5261 και 0.5237). Όσον αφορά τη μέθοδο GMM (Εικόνα 44), τα δεδομένα της περιοχής της Ηλιούπολης φαίνονται να έχουν διαχωριστεί αρκετά καλύτερα σε σχέση με τα δεδομένα της περιοχής των Σερρών, αλλά οι αρνητικές τιμές στα silhouette γραφήματα δεν παύει να είναι πολύ περισσότερες σε σχέση με τους άλλους δυο αλγόριθμους.

5.4 Βέλτιστες διαμερίσεις – διαμερίσεις για 5 clusters.

Συνοπτικά, σύμφωνα με τους δείκτες επικύρωσης, οι καλύτερες διαμερίσεις για κάθε περίπτωση παρουσιάζονται στους παρακάτω πίνακες:

Πίνακας 19: Βέλτιστες διαμερίσεις για την περιοχή των Σερρών.

Βέλτιστες διαμερίσεις για την περιοχή των Σερρών		
Δείκτης	*Αλγόριθμος Ομαδοποίησης*	*Αριθμός των clusters*
Silhouette	k-Means	3
Davies-Bouldin	k-Means	7

Πίνακας 20: Βέλτιστες διαμερίσεις για την περιοχή της Ηλιούπολης.

Βέλτιστες διαμερίσεις για την περιοχή της Ηλιούπολης		
Δείκτης	*Αλγόριθμος Ομαδοποίησης*	*Αριθμός των clusters*
Silhouette	k-Means	3
Davies-Bouldin	GMM	7

Πίνακας 21: Βέλτιστος αλγόριθμος για αριθμό clusters ίσο με 5.

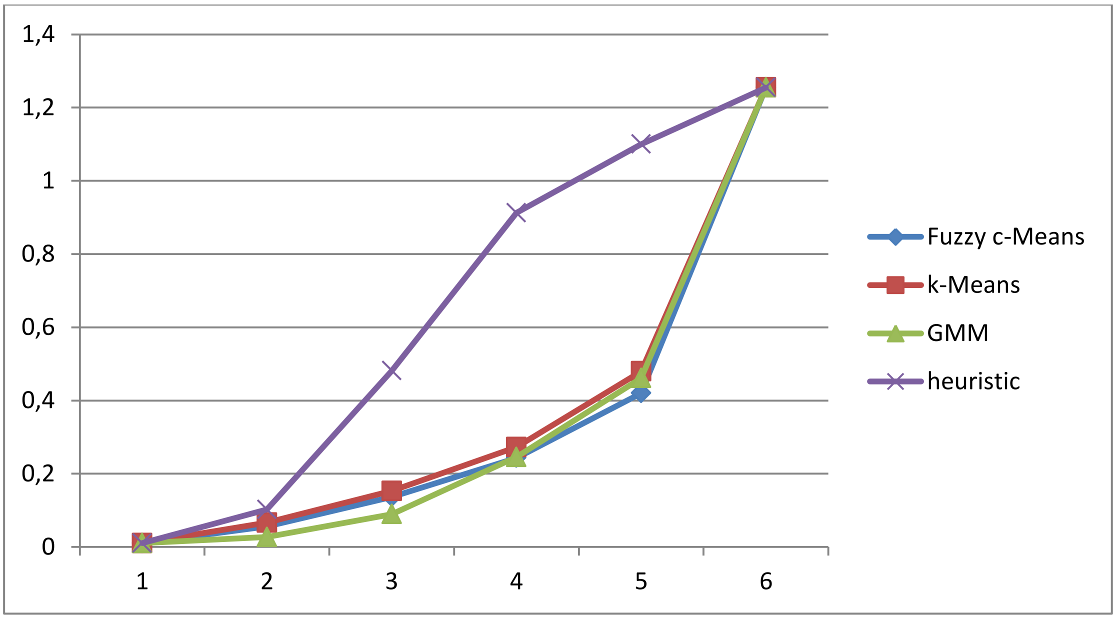
Βέλτιστος αλγόριθμος για αριθμό clusters ίσο με 5		
Δείκτης	*Δεδομένα Σερρών*	*Δεδομένα Ηλιούπολης*
Silhouette	k-Means	k-Means
Davies-Bouldin	GMM	k-Means

Ο αποδοτικότερος αλγόριθμος ομαδοποίησης σύμφωνα με τους δείκτες silhouette και Davies-Bouldin του πίνακα (Πίνακας 19) για την περιοχή των Σερρών είναι ο k-Means για 3 και για 7 clusters αντίστοιχα ενώ για την περιοχή της Ηλιούπολης (Πίνακας 20) είναι ο k-Means για 3 clusters με δείκτη επικύρωσης silhouette και ο GMM για 7 clusters με δείκτη επικύρωσης Davies-Bouldin.

Για αριθμό clusters ίσο με 5, ο οποίος αντιστοιχεί και στις 5 κατηγορίες επικινδυνότητας έναρξης πυρκαγιάς, αποδοτικότερος αλγόριθμος είναι ο k-Means με μικρή διαφορά από τον αλγόριθμο Fuzzy c-Means τόσο στις τιμές των δεικτών επικύρωσης όσο και στις τιμές των ορίων των clusters. Αν και από το πίνακα (Πίνακας 20), καλύτερη τιμή για τον δείκτη επικύρωσης davied-bouldin παρατηρείτε για τον αλγόριθμο GMM, οι αρνητικές τιμές του δείκτη επικύρωσης silhouette (Εικόνα 41) είναι αρκετές, άρα δεν ενδείκνυται η επιλογή του συγκεκριμένου αλγορίθμου. Επίσης όπως προαναφέραμε στο κεφάλαιο 4, για την εκτέλεση του αλγορίθμου χρησιμοποιήθηκε ένα διάνυσμα το οποίο περιέχει μια εκτίμηση των αποτελεσμάτων από την εφαρμογή του αλγορίθμου k-Means γεγονός το οποίο αυξάνει την πολυπλοκότητά του. Παρακάτω παρουσιάζονται τα όρια των clusters για τους τρεις αλγόριθμους, καθώς και για την ευριστική (heuristic) μέθοδο με την οποία υπολογίζονται οι 5 κατηγορίες επικινδυνότητας που εκδίδει η Γενική Γραμματεία Πολιτικής Προστασίας. Η ευριστική μέθοδος υλοποιήθηκε με σκοπό τα ποσοστά της κάθε κατηγορίας να είναι όσο πιο κοντά γίνεται με τα ποσοστά που εξάγαμε από τους χάρτες της ΓΓΠΠ (Πίνακας 3 , Πίνακας 4) στο 2ο κεφάλαιο.

Πίνακας 22: Όρια κατηγοριών δείκτη επικινδυνότητας για την περιοχή των Σερρών.

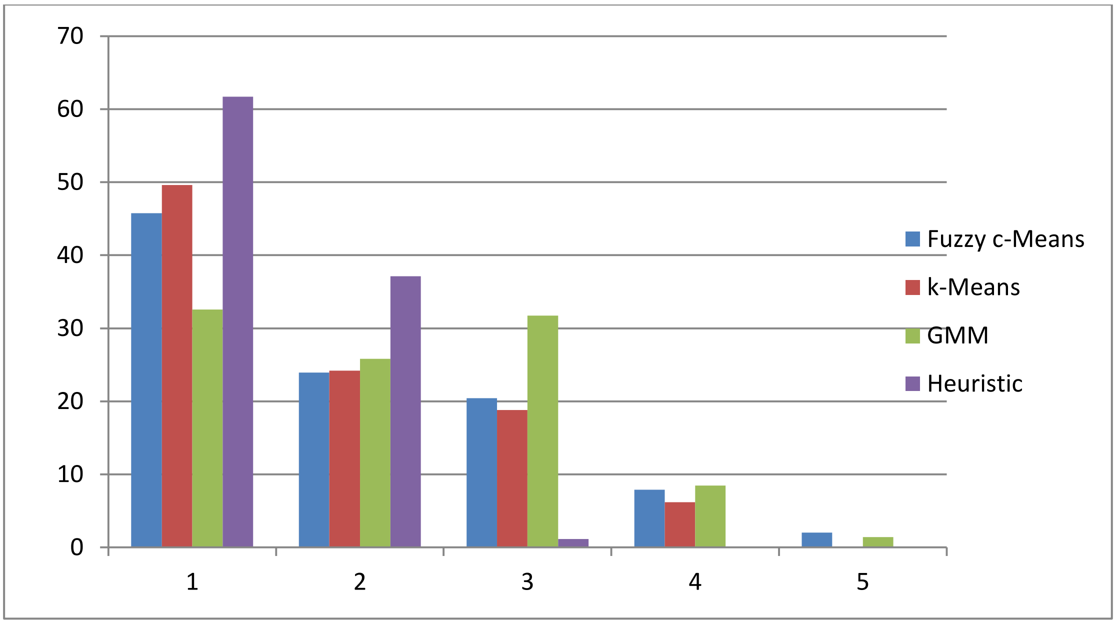
Όρια κατηγοριών δείκτη επικινδυνότητας για την περιοχή των Σερρών				
Κατηγορίες Κινδύνου	*Fuzzy c-Means*	*k-Means*	*Gaussian Mixture*	*Heuristic*
1	0.0103-0.0565	0.0103-0.0662	0.0103-0.0269	0.0103-0.1020
2	0.0565-0.1366	0.0662-0.1531	0.0269-0.0893	0.1020-0.4809
3	0.1366-0.2434	0.1531-0.2727	0.0893-0.2457	0.4809-0.9124
4	0.2434-0.4198	0.2727-0.4795	0.2457-0.4610	0.9124-1.1000
5	0.4198-1.2545	0.4795-1.2545	0.4610-1.2545	1.1000-1.2545



Εικόνα 45: Όρια δείκη επικινδυνότητας Σερρών.

Πίνακας 23: Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) για την περιοχή των Σερρών

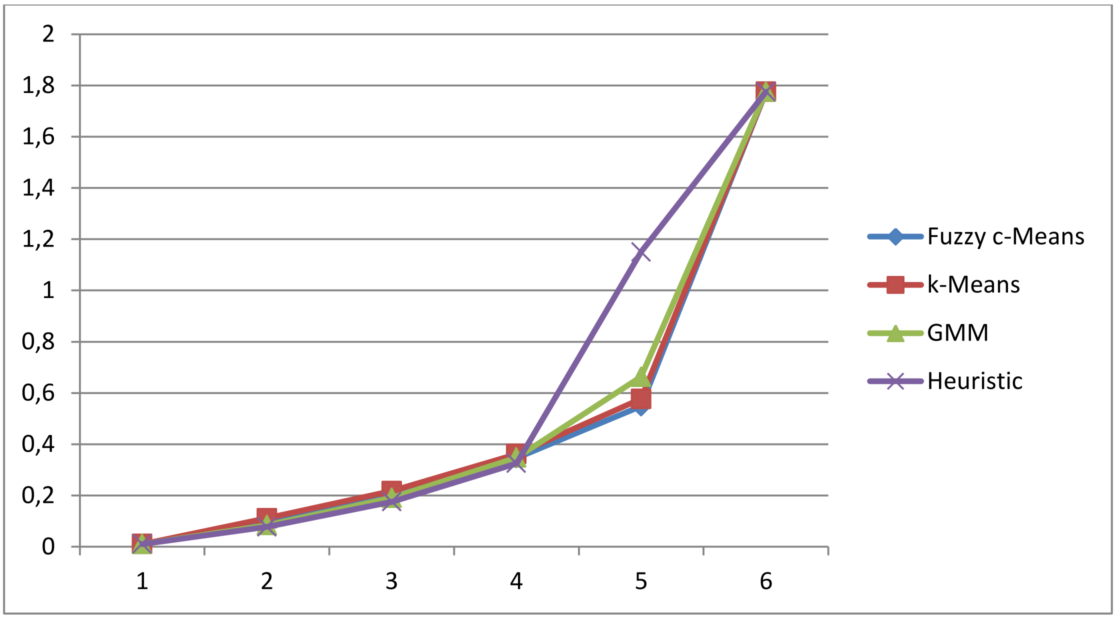
Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) για την περιοχή των Σερρών								
Κατηγορίες Κινδύνου	*Fuzzy c-Means*		*k-Means*		*Gaussian Mixture*		*Heuristic*	
1	11748	45.74	12757	49.63	8372	32.57	15861	61.71
2	6149	23.92	6214	24.18	6637	25.82	9538	37.11
3	5251	20.42	4836	18.81	8257	31.73	295	1.15
4	2031	7.90	1590	6.19	2176	8.47	8	0.03
5	515	2.01	307	1.19	362	1.41	1	0.00



Εικόνα 46: Ποσοστά κάθε κατηγορίας για την περιοχή των Σερρών.

Πίνακας 24: Όρια κατηγοριών δείκτη επικινδυνότητας για την περιοχή της Ηλιούπολης.

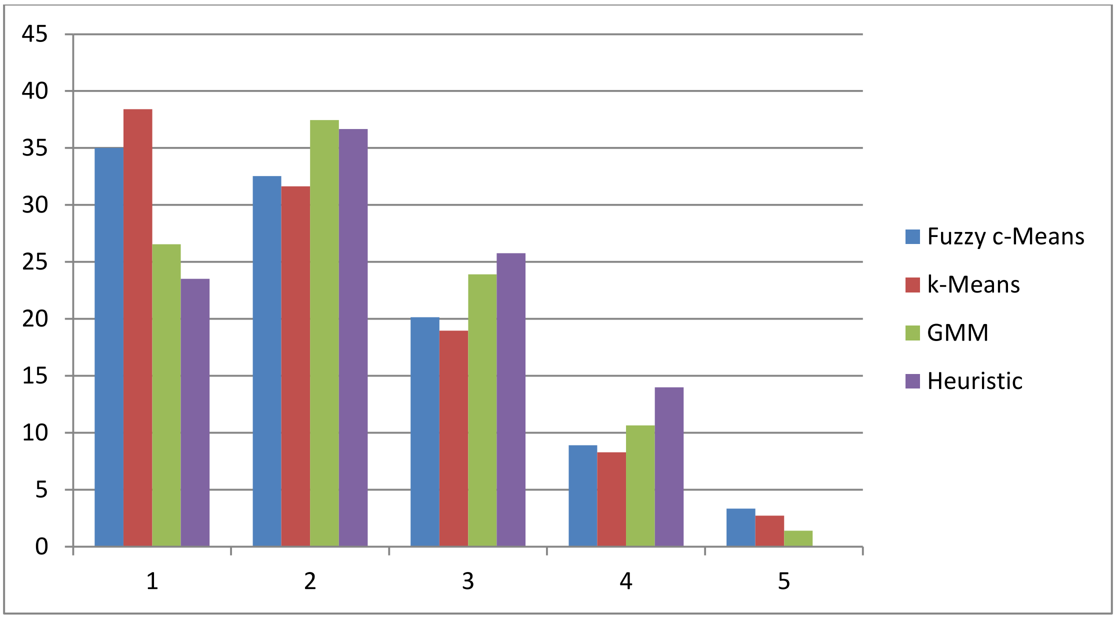
Όρια κατηγοριών δείκτη επικινδυνότητας για την περιοχή της Ηλιούπολης				
Κατηγορίες Κινδύνου	*Fuzzy c-Means*	*k-Means*	*Gaussian Mixture*	*Heuristic*
1	0.0102-0.1037	0.0102-0.1112	0.0102-0.0844	0.0102-0.0776
2	0.1037-0.2062	0.1112-0.2170	0.0844-0.1908	0.0776-0.1756
3	0.2062-0.3455	0.2170-0.3616	0.1908-0.3477	0.1756-0.3254
4	0.3455-0.5479	0.3616-0.5769	0.3477-0.6629	0.3254-1.1500
5	0.5479-1.7747	0.5769-1.7747	0.6629-1.7747	1.1500-1.7747



Εικόνα 47: Όρια δείκτη επικινδυνότητας Ηλιούπολης.

Πίνακας 25: Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) για την περιοχή της Ηλιουπολης

Αριθμός δεδομένων/Ποσοστό συμμετοχής (%) για την περιοχή της Ηλιουπολης								
Κατηγορίες Κινδύνου	*Fuzzy c-Means*		*k-Means*		*Gaussian Mixture*		*Heuristic*	
1	9028	35.02	9898	38.40	6847	26.56	6046	23.52
2	8390	32.55	8152	31.63	9652	37.45	9425	36.67
3	5197	20.16	4886	18.96	6165	23.92	6623	25.77
4	2296	8.91	2137	8.29	2744	10.65	3596	13.99
5	865	3.36	703	2.72	368	1.42	13	0.05



Εικόνα 48: Ποσοστά κάθε κατηγορίας για την περιοχή της Ηλιούπολης.

Αναλύοντας τους πίνακες (Πίνακας 22),(Πίνακας 24) και τις εικόνες (Εικόνα 45),(Εικόνα 47), διαπιστώνουμε ότι τα όρια κατηγοριών του δείκτη επικινδυνότητας για την περιοχή των Σερρών όπως και για την περιοχή της Ηλιούπολης, είναι αρκετά κοντά κατά την εφαρμογή των αλγορίθμων Fuzzy c-Means και k-Means. Τα αποτελέσματα του GMM αλγορίθμου έχουν σημαντική απόκλιση από τα αποτελέσματα των άλλων δυο αλγορίθμων για την περιοχή των Σερρών, ιδιαίτερα για τις πρώτες κατηγορίες κινδύνου (από 1 έως 3) ενώ είναι αρκετά πιο κοντά για τα δεδομένα της περιοχής της Ηλιούπολης. Αυτό φαίνεται και από τους πίνακες (Πίνακας 21),(Πίνακας 23) και τις εικόνες (Εικόνα 46),(Εικόνα 48), από τα ποσοστά που καταλαμβάνει η κάθε κατηγορία ως προς το συνολικό αριθμό των παρατηρήσεων. Σε σχέση με τα heuristic αποτελέσματα, η απόκλιση είναι αρκετά μεγάλη για τα δεδομένα των Σερρών ενώ για τα δεδομένα της Ηλιούπολης η απόκλιση μειώνεται αισθητά. Το συγκεκριμένο γεγονός ίσως να οφείλεται στα διάφορα τοπογραφικά κριτήρια που λαμβάνονται υπόψη από την ΓΓΠΠ για την έκδοση του χάρτη επικινδυνότητας, τα οποία δεν μελετούνται στην συγκεκριμένη διατριβή.

Κεφάλαιο 6-Συμπεράσματα και μελλοντική έρευνα

6.1 Συμπεράσματα

Τα τελευταία χρόνια οι προσπάθειες πρόβλεψης και διαχείρισης του κινδύνου πυρκαγιάς έχουν εντατικοποιηθεί τόσο στην Ευρώπη όσο και παγκοσμίως. Η πλειοψηφία των χωρών στην Ευρώπη και ιδιαίτερα στην περιοχή της Μεσογείου έχουν υιοθετήσει διάφορους δείκτες για την αξιολόγηση του κινδύνου πυρκαγιάς. Στην συγκεκριμένη διατριβή μελετήθηκε ο μετεωρολογικός δείκτης κινδύνου πυρκαγιάς (FDI) ο οποίος σε συνδυασμό με τον τοπογραφικό δείκτη κινδύνου (FHI) εξάγουν τον δείκτη επικινδυνότητας έναρξης πυρκαγιάς μιας περιοχής την δεδομένη χρονική στιγμή (FRI) ο οποίος παρουσιάζεται στον πίνακα (Πίνακας 26).

Πίνακας 26: Δείκτης επικινδυνότητας έναρξης πυρκαγιάς (FRI) σε συνδειασμό του μετεωρολογικού δείκτη (FDI) με τον τοπογραφικό δείκτη (FRI).

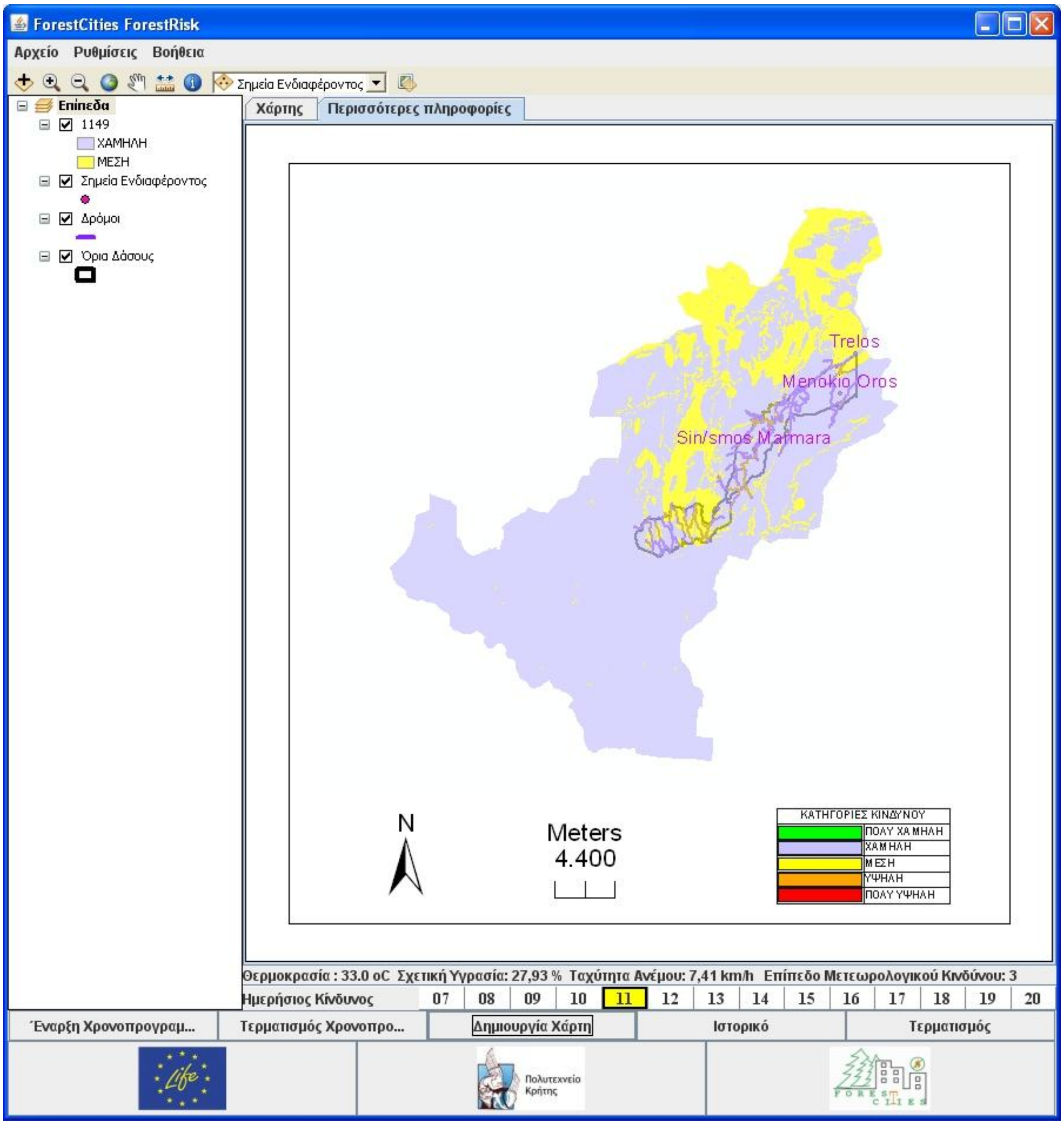
FRI	very low FDI	low FDI	medium FDI	high FDI	very high FDI
low FHI	1	2	3	4	5
medium FHI	2	2	3	4	5
high FHI	3	3	3	4	5

Πιο συγκεκριμένα, με τη χρήση προηγμένων τεχνικών ομαδοποίησης στα μετεωρολογικά δεδομένα των περιαστικών δασών Ηλιούπολης και Σερρών, μελετήθηκε η βελτίωση της δυνατότητας εκτίμησης του κινδύνου έναρξης πυρκαγιάς, ο οποίος κατηγοριοποιείται σε πέντε κατηγορίες με τελικό σκοπό τη δημιουργία χάρτη.

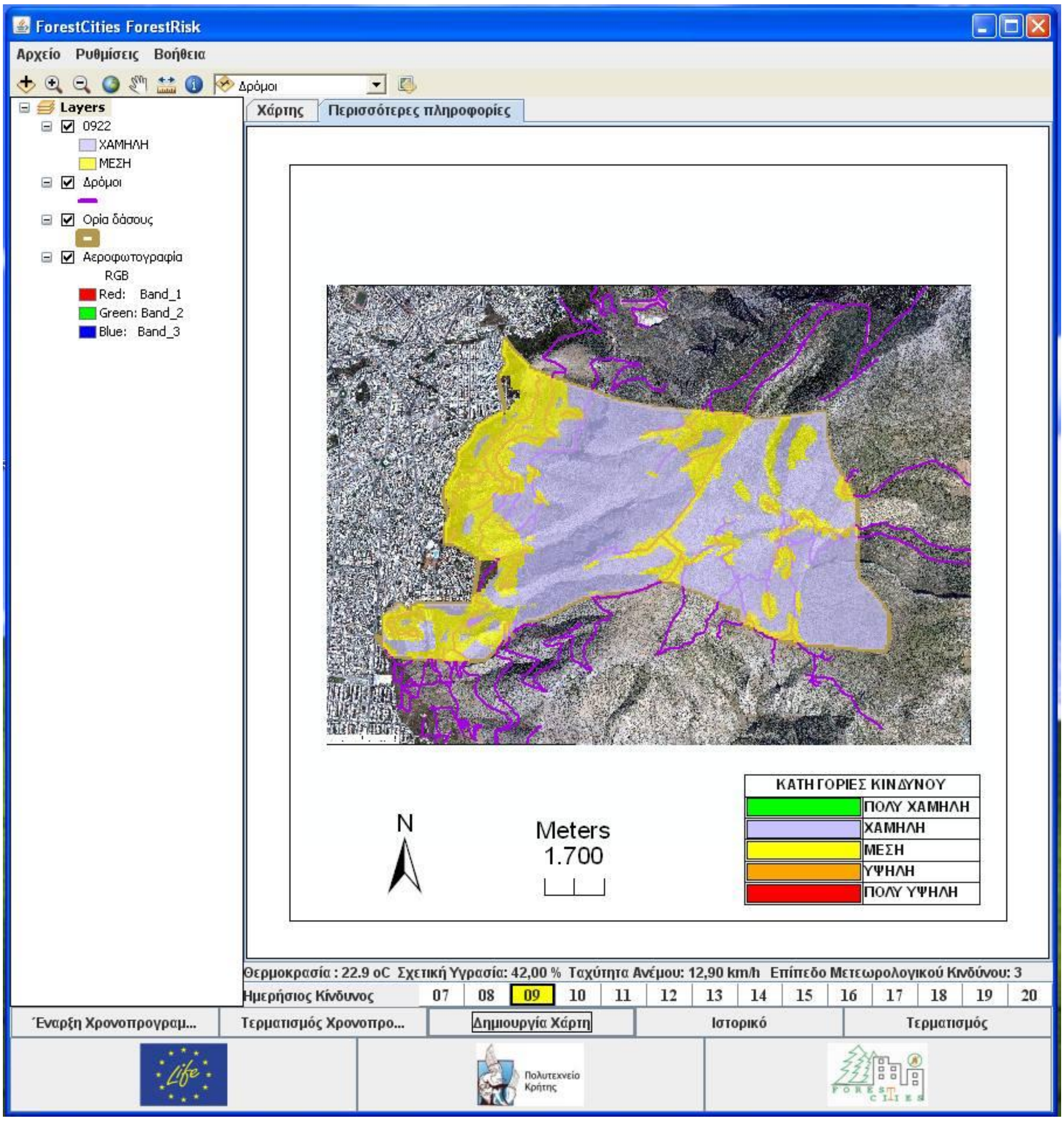
Οι αλγόριθμοι ομαδοποίησης που χρησιμοποιήθηκαν ήταν οι Fuzzy c-Means, k-Means και Gaussian Mixture Model καθώς έγινε και ανεπιτυχής προσπάθεια εφαρμογής του Hierarchical αλγόριθμου λόγο μεγάλου όγκου δεδομένων και ανεπαρκούς μνήμης από έναν κοινό ηλεκτρονικό υπολογιστή. Σε γενικές γραμμές, τα αποτελέσματα των αλγορίθμων Fuzzy c-Means και k-Means είναι αρκετά κοντά μεταξύ τους ενώ σημαντική διαφοροποίηση παρατηρείται για τον αλγόριθμο Gaussian Mixture Model. Εκτός από την επιλογή του βέλτιστου αλγορίθμου για τις ανάγκες της εκτίμησης του δείκτη επικινδυνότητας σε πέντε κατηγορίες, γίνεται και εκτενή μελέτη για την εύρεση του βέλτιστου αριθμού διαμερίσεων (clusters).

Για την επιλογή του βέλτιστου αλγόριθμου και την εύρεση των καλύτερων διαμερίσεων χρησιμοποιήθηκαν δυο τεχνικές επικύρωσης: ο δείκτης Silhouette και ο δείκτης Davies – Bouldin. Τα αποτελέσματα των δεικτών επικύρωσης δεν συμφωνούσαν σε όλες τις περιπτώσεις ως προς τον βέλτιστο αριθμό των clusters αν και θα πρέπει να τονιστεί ότι στην πλειοψηφία τους οι καλύτερες διαμερίσεις προέρχονταν από τον k-Means αλγόριθμο.

Τα αποτελέσματα της παρούσας διατριβής χρησιμοποιήθηκαν για τις ανάγκες της εφαρμογής DFF-Risk με σκοπό τη δημιουργία χάρτη εκτίμησης του κινδύνου πυρκαγιάς ο οποίος παρουσιάζεται στις εικόνες (Εικόνα 49),(Εικόνα 50). Περισσότερες πληροφορίες για τη συγκεκριμένη εφαρμογή βρίσκονται στο internet στην διεύθυνση http://www.forestcities.gr/.



Εικόνα 49: Τελικός Χάρτης επικινδυνότητας έναρξης πυρκαγιάς για την περιοχή των Σερρών.



Εικόνα 50: Τελικός Χάρτης επικινδυνότητας έναρξης πυρκαγιάς για την περιοχή των Σερρών.

6.2 Μελλοντική έρευνα.

Στη συγκεκριμένη διατριβή, οι αλγόριθμοι ομαδοποίησης εφαρμόστηκαν για δυο μόνο περιοχές, την περιοχή των Σερρών και την περιοχή της Ηλιούπολης. Θα πρέπει λοιπόν να γίνει η συλλογή μετεωρολογικών δεδομένων και από άλλες περιοχές της Ελλάδος, ώστε να έχουμε μια πιο ολοκληρωμένη εικόνα για την συμπεριφορά των αλγορίθμων. Επίσης, η εφαρμογή και άλλων αλγορίθμων στα συγκεκριμένα δεδομένα, πέρα των τριών για τους οποίους έχουμε αποτελέσματα, θα μπορούσε να μελετηθεί για την επίτευξη καλύτερων αποτελεσμάτων καθώς και η προσπάθεια εκτέλεσης του hierarchical αλγόριθμου σε έναν υπολογιστή με αρκετή μνήμη ώστε να μπορεί να υποστηρίζει μεγάλο όγκο δεδομένων.

Η σύγκριση των αποτελεσμάτων του δείκτη επικινδυνότητας FDI κατά το πέρασμα του χρόνου θα ήταν μια ενδιαφέρον μελλοντική έρευνα. Με την πάροδο του χρόνου όλο και περισσότερα μετεωρολογικά δεδομένα συλλέγονται σε βάσεις δεδομένων διάφορων σταθμών. Η μελέτη της αλλαγής των ορίων του δείκτη επικινδυνότητας είναι βασική προτεραιότητα που πρέπει να ληφθεί σοβαρά υπόψη για την καλύτερη πρόβλεψη και προστασία των δασών από ενδεχόμενη πυρκαγιά.

Η εφαρμογή DFF-risk όπως προαναφέραμε χρησιμοποιεί τα δεδομένα της παρούσας διατριβής, αφού πρώτα τρέξει τους αλγόριθμους στο πρόγραμμα του Matlab. Επομένως, ενδιαφέρον θα ήταν η ενσωμάτωση των αλγορίθμων στο περιβάλλον της συγκεκριμένης εφαρμογής ώστε να εξάγονται τα όρια επικινδυνότητας απευθείας χωρίς την χρήση εξωτερικών προγραμμάτων.

7 Βιβλιογραφία

1. **Λέκκας, Ε.Λ.** *Φυσικές και Τεχνολογικές Καταστροφές. .* Αθήνα : Access Pre-Press., 2000.

2. **Δ.Α.Ε.Ε. / Πυροσβεστικού Σώματος, Γ.Γ.Δ.Φ.Π. / Υπουργείο Γεωργίας.**

3. **Smith, J.K.** *Wildland fire in ecosystems: effects of fire on fauna. General Technical Report RMRS-GTR-42-vol.1, 83 pp. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.* 2000.

4. **Singer, F.J., & Schullery, P.** *Yellowstone wildlife: populations in process. Western Wildlands, 15(2), 18-22.* 1989.

5. **Brown, J.K., & Smith, J.K.** *Wildland fire in ecosystems: effects of fire on flora. General Technical Report RMRS-GTR-42 vol.2, 257 pp. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.* 2000.

6. **Pausas, J.G., & Vallejo, V.R.** *The role of fire European Mediterranean ecosystems. . In E. Basin. & E. Chuvieco (Eds.), Remote Sensing of Large Wildfires in the European Mediterannean (pp.3-16). Berlin: Springer-Verlag.* 1999.

7. **Miller, J.D., Nyham, J.W., & Yool, S.R.** *Modeling potential erosion due to the Cerro Grande Fire wih a GIS-based implementation of the Revised Universal Soil Loss Equation. International Journal of Wildland Fire, 12(2003), 85-100.* 2003.

8. **Neary, D.G., Ryan, K.C., & DeBano, L.F.** *Wildland fire in ecosystems: effects of fire on soils and water. General Technical Report RMRS-GTR-42-vol.4, 250 pp. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research.* 2005.

9. **NWCG, National Wildfire Coordinating Group.** *Glossary of Wildland Fire Terminology, PMS205, 186p.* 2008.

10. **Γκούμα, Β.** *Μεθοδολογία χωροχρονικής εκτίμησης του μετεωρολογικού κινδύνου δασικής πυρκαγιάς σε ορεινές περιοχές: Εφαρμογή στον ορεινό όγκο της Πάρνηθας. Διδακτορική διατριβήΓΠΑ, 221 σελ. (+Παράρτημα).* 2001.

11. **McArthur, A.G.** *Weather and grassland fire behaviour. Department of National Development, Forestry and Timber Bureau Leaflet No. 100. .* Canberra, Australia. : s.n., 1966.

12. —. *Fire behaviour in eucalypt forests. Department of National Development, Forestry and Timber Bureau Leaflet No. 107.* Canberra, Australia. : s.n., 1967.

13. **Noble, I.R., Bary, G.A.V, Gill, A.M.** *McArthur's fire-danger meters expressed as equations. Australian Journal of Ecology, 5, 201-203.* 1980.

14. **Gorski, C.J., Farnsworth, A.** *Fire weather and smoke management. In: Whiteman, C.D., Mountain meteorology fundamentals and applications. Oxford University Press.* . New York. : s.n., 2000.

15. **Pook, E.W.** *Empirical models evaluated for the prediction of fine fuel moisture in Australian Pinus radiata plantations. New Zealand Journal of Forestry, 23(3), 278-297.* New Zealand : s.n., 1993.

16. **Pook, E.W., Gill, A.M.** *Variation of live and dead fine fuel moisture in Pinus radiata plantations of the Australian Capital Territory. International Journal of Wildland Fire, 3(3), 155-168.* 1993.

17. http://www.hnms.gr/hnms/greek/index_html. [Ηλεκτρονικό]

18. **Rouse' J.W., Hass R.H., Shell J.A. and Deering D.W.** *Monitoring Vegetation.* 1974.

19. **B.S, Everitt.** *"Cluster Analysis".* New York : John Wiley & Sons, 1974.

20. **S. Theodoridis, K. Koutroubas.** *"Pattern Recognition".* s.l. : Academic Press, 1999.

21. **R. Xu, D. Wunsch.** *"Survey on Clustering Algorithms", IEEE Transactions on Neural Networks, volume 16,pp. 645-678.* 2005.

22. **Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου.** *Τεχνητή Νοημοσύνη - B' ΈκδοσηΙ.*

23. **Songül Albayrak, Fatih Amasyalı.** *Fuzzy c-means clustering on medical diagnostic systems .*

24. **M. Halkidi, Y. Batistakis, M. Vazirgiannis.** *"On Clustering Validation Techniques", Journal of Intelligent Information Systems, 17(2-3), pp. 107-145.* 2001.

25. **Rousseeuw, P.** *"Silhouettes: A graphical aid to the interpretation and validation of cluster analysis", Journal of Computational and Applied Mathematics, volume 20, pp. 53-65.* 1987.

26. **D. Davies, D. Bouldin.** *"A cluster separation measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 1, pp. 224-227.* 2009.

27. http://www.mathworks.com/. [Ηλεκτρονικό]