

ΧΕΕΕ

Χαρτογραφική
Επιστημονική
Εταιρεία
Ελλάδας

Τμήμα
Αγρονόμων
και Τοπογράφων
Μηχανικών Α.Π.Θ.



10 Χρόνια

**ΠΡΑΚΤΙΚΑ
ΣΥΝΕΔΡΙΟΥ**

**Η Χαρτογραφία
του ευ ζην**

8^ο Εθνικό Συνέδριο Χαρτογραφίας

**ΘΕΣΣΑΛΟΝΙΚΗ
24-26
ΝΟΕΜΒΡΙΟΥ
2004**

ΤΕΛΕΤΗ ΕΝΑΡΞΗΣ: ΑΜΦΙΘΕΑΤΡΟ ΠΟΛΥΤΕΧΝΙΚΗΣ ΣΧΟΛΗΣ
"ΠΑΝΑΓΗΣ ΠΑΝΑΓΙΩΤΟΠΟΥΛΟΣ"
ΤΕΤΑΡΤΗ 24 ΝΟΕΜΒΡΙΟΥ 2004 ώρα 18:00

ΕΡΓΑΣΙΕΣ ΣΥΝΕΔΡΙΟΥ: ΑΜΦΙΘΕΑΤΡΟ ΠΟΛΥΤΕΧΝΙΚΗΣ ΣΧΟΛΗΣ
"ΠΑΝΑΓΗΣ ΠΑΝΑΓΙΩΤΟΠΟΥΛΟΣ"

Ομαδοποίηση Δεδομένων για Χωροπληθείς Χάρτες με τη Μέθοδο των Χαρτών Αυτο-Οργάνωσης

Ανδρέας Παπαχριστοδούλου¹, Βύρωνας Νάκος²

1. Μεταπτυχιακός Φοιτητής ΕΜΠ
2. Αναπληρωτής Καθηγητής ΕΜΠ

Περίληψη

Ένα από τα σημαντικότερα προβλήματα στη δημιουργία χωροπληθών χαρτών είναι η διαδικασία ομαδοποίησης των πολυπληθών δεδομένων σε ένα πεπερασμένο αριθμό διαδοχικών ομάδων (4-7), ώστε να διαφοροποιηθούν οπτικά με ευκρίνεια από το μέσο απόδοσης, εφαρμόζοντας την οπτική μεταβλητή της έντασης μιας απόχρωσης. Κατά την ομαδοποίηση τα δεδομένα πρέπει να ταξινομούνται με τέτοιο τρόπο, ώστε οι τιμές τους να παρουσιάζουν αφενός ομοιογένεια μέσα στις ομάδες, αφετέρου σημαντικές διαφορές μεταξύ των ομάδων. Στην εργασία εφαρμόζεται μια νέα μέθοδος ομαδοποίησης δεδομένων που βασίζεται στην τεχνολογία των νευρωνικών δικτύων και ειδικότερα στους «χάρτες αυτο-οργάνωσης» (Self-Organized Maps – SOM). Πιο συγκεκριμένα, αναπτύσσεται μια μεθοδολογία ομαδοποίησης με εφαρμογή του αλγορίθμου *SOM Toolbox* του λογισμικού *Matlab* στα δεδομένα της πυκνότητας πληθυσμού -που προέρχονται από επεξεργασία των στοιχείων της γενικής απογραφής πληθυσμού της ΕΣΥΕ για το έτος 1991- ως προς τους νομούς της χώρας σε 3, 5 & 7 ομάδες ταξινόμησης. Η προτεινόμενη μέθοδος συγκρίνεται με πέντε υπάρχουσες χαρτογραφικές μεθόδους ομαδοποίησης δια μέσου της οπτικοποίησης των παραγόμενων ομάδων. Ως συμπέρασμα προκύπτει ότι η μέθοδος ομαδοποίησης με τη βοήθεια των χαρτών αυτο-οργάνωσης δίνει πολύ καλά αποτελέσματα χωροπληθών χαρτών και υπερέχει ειδικότερα στο κριτήριο της ομοιογένειας μεταξύ των δεδομένων που κατατάσσονται σε κάθε ομάδα.

Data Classification of Choropleth Maps Using the Self-Organizing Maps Method

Papachristodoulou Andreas¹, Nakos Vyronas²

1. Postgraduate student NTUA
2. Associate Professor NTUA

Abstract

One of the most important problems in the creation of choropleth maps is the process of data classification into finite numbers of successive groups, so that their visual differentiation by applying the visual variable of value to be effective. The data classification

should be applied in such a way, that the homogeneity is preserved inside each group as well as the differentiation among the groups. In the present paper a method of data classification is proposed and discussed, based on the technology of neural networks and more specifically on the Self-Organizing Maps (SOM) method. More specifically, the introduced method applies the *SOM* algorithm in order to group data of population density. The proposed method is compared with five existing cartographic methods of data classification by analyzing visually the results. It is concluded that the proposed method gives very good results when applied on choropleth maps by preserving the homogeneity between the data that are classified in each group.



1. ΕΙΣΑΓΩΓΗ

Οι διάφορες μέθοδοι ομαδοποίησης δεδομένων αποτελούν τρόπους ταξινόμησης ενός συνόλου αριθμητικών δεδομένων σε ένα πεπερασμένο αριθμό ομάδων, ώστε σε κάθε ομάδα να περιέχεται «γεωγραφικά» ισοδύναμος αριθμός παρατηρήσεων. Η οργάνωση των δεδομένων σε ομάδες ανάλογα με τις ιδιότητές τους είναι ιδιαίτερα χρήσιμη, κυρίως όταν έχουμε να κάνουμε με πολυάριθμες οντότητες [1]. Γενικότερα, το πρόβλημα της ομαδοποίησης εστιάζεται στην ταξινόμηση ενός συνόλου δεδομένων σε διακριτές ομάδες, ανάλογα με την κατανομή που εμφανίζουν οι τιμές των δεδομένων [2]. Οι τιμές ομαδοποιούνται στις ομάδες προκειμένου να απλοποιηθούν τα αρχικά δεδομένα και να μπορούν να γίνουν πιο εύκολα αντιληπτά από τους χρήστες.

Η ομαδοποίηση των αριθμητικών δεδομένων για την κατασκευή ενός χωροπληθή χάρτη αρχίζει με τον προσδιορισμό του αριθμού των ομάδων. Ο αριθμός των ομάδων εξαρτάται άμεσα από το όριο διαφοροποίησης της οπτικής μεταβλητής που πρόκειται να εφαρμοστεί. Στους χωροπληθείς χάρτες η οπτική μεταβλητή που χρησιμοποιείται είναι η ένταση μιας απόχρωσης. Ανεξάρτητα από τη μέθοδο ομαδοποίησης ο αριθμός των ομάδων, έχει τυποποιηθεί αφού μετά από μελέτες διαπιστώθηκε ότι οι αναγνώστες των χαρτών δεν μπορούν να διακρίνουν εύκολα περισσότερες από πέντε έως επτά ή το πολύ οκτώ εντάσεις μιας απόχρωσης [3]. Στην πράξη, χρησιμοποιούνται συνήθως τέσσερις, πέντε ή το πολύ έξι ομάδες, για την καλύτερη διαχείριση των συμβόλων που επιλέγονται, αλλά και για τη βελτίωση της αναγνωσιμότητας των παραγόμενων χαρτών [4, 5, 6, 7].

Παρά το γεγονός ότι έχει αναπτυχθεί μεγάλος αριθμός μεθόδων ομαδοποίησης δεδομένων, δεν είναι εύκολο να χαρακτηριστεί μια από αυτές ως η «καλύτερη». Δεν υπάρχει μια μέθοδος που να μπορεί να εφαρμοστεί σε όλα τα πιθανά σύνολα δεδομένων, ανεξάρτητα του τρόπου που κατανέμονται οι τιμές τους. Αντίθετα, κάθε φορά χρειάζεται να γίνεται ανάλυση των δεδομένων με στόχο να προσδιοριστεί η «βέλτιστη» μέθοδος για την ομαδοποίηση του συγκεκριμένου συνόλου δεδομένων. Για την επιλογή της καλύτερης μεθόδου ομαδοποίησης λαμβάνεται υπόψη η οπτική απλότητα και η καλή οργάνωση του χάρτη αλλά και το πως κατανέμονται τα δεδομένα στις ομάδες [8]. Με βάση την παραπάνω επισήμανση στην εργασία αυτή διερευνάται μια νέα μέθοδος ομαδοποίησης που βασίζεται στον αλγόριθμο των χαρτών αυτο-οργάνωσης (Self-Organizing Maps) [9], ο οποίος εντάσσεται στο ευρύτερο γνωστικό πεδίο των νευρωνικών δικτύων [10, 11, 12, 13, 14, 15]. Η προτεινόμενη μέθοδος λαμβάνει υπόψη την κατανομή των δεδομένων αλλά και την ομοιογένεια των τιμών των δεδομένων που τοποθετούνται σε κάθε ομάδα.

2. ΟΙ ΧΑΡΤΕΣ ΑΥΤΟ-ΟΡΓΑΝΩΣΗΣ (SOM)

Ο αλγόριθμος των χαρτών αυτο-οργάνωσης (SOM), βασίζεται στην αρχή της εκπαίδευσης χωρίς εποπτεία και η εφαρμογή του απεικονίζει ένα πολυδιάστατο σύνολο δεδομένων, σε ένα μονοδιάστατο ή δισδιάστατο «χάρτη». Με τη μέθοδο των χαρτών αυτο-οργάνωσης μπορούν να επιτευχθούν δύο στόχοι [9]:

- (α) η ομαδοποίηση των δεδομένων να γίνεται με τρόπο που να παρουσιάζει ομοιογένεια εντός των ομάδων (τα χαρακτηριστικά διανύσματα που βρίσκονται εγγύτερα στο πεδίο εισόδου να κατανέμονται σε γειτονικές θέσεις στο πεδίο εξόδου) και
- (β) μείωση των διαστάσεων των δεδομένων με την παραγωγή μιάς μονοδιάστατης ή δισδιάστατης απεικόνισης.

2.1 Περιγραφή του αλγόριθμου SOM

Ο τρόπος λειτουργίας ενός νευρωνικού δικτύου βασίζεται στη διαδικασία που περιγράφεται στην ενότητα αυτή [9]. Θεωρείται ένα διάνυσμα εισόδου $x = [x_1, x_2, \dots, x_n]^T \in R^n$ και ένα διάνυσμα βάρους (synaptic weight vector) του νευρώνα i : $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$, με $i = 1, 2, \dots, m$, όπου m είναι ο συνολικός αριθμός των νευρώνων εξόδου. Κατά την έναρξη της επεξεργασίας όλα τα διανύσματα βαρών στους νευρώνες έχουν τυχαίες αρχικές τιμές. Όλα τα διανύσματα εισόδου συνδέονται με όλους τους νευρώνες μέσω των συνάψεων των νευρώνων, οι δε τιμές των συνάψεων είναι διαφορετικές για κάθε νευρώνα. Από το σύνολο των διανυσμάτων εισόδου επιλέγεται ένα τυχαίο διάνυσμα με το οποίο τροφοδοτείται το νευρωνικό δίκτυο. Στη συνέχεια, για κάθε νευρώνα υπολογίζεται η Ευκλείδεια απόσταση (d_i) μεταξύ του διανύσματος εισόδου και του διανύσματος βαρών του:

$$d_i = \|x - w_i\| = \sqrt{(x_1 - w_{i1})^2 + (x_2 - w_{i2})^2 + \dots + (x_n - w_{in})^2} \quad (2.1)$$

όπου x το διάνυσμα εισόδου και w_i το διάνυσμα βαρών του νευρώνα i .

Εναλλακτικά, θα μπορούσε να υπολογιστεί το εσωτερικό γινόμενο των διανυσμάτων x και w_i , που ως μέγεθος είναι αντίστροφα συσχετισμένο με την Ευκλείδεια απόσταση, δηλαδή, όσο μικρότερη γίνεται η Ευκλείδεια απόσταση τόσο μεγαλύτερο γίνεται το εσωτερικό γινόμενο. Ο πλησιέστερος νευρώνας (μικρότερη απόσταση) στο διάνυσμα εισόδου θεωρείται ο «νευρώνας-νικητής». Σε περίπτωση που υπάρχουν περισσότεροι του ενός νευρώνες με την ίδια μικρότερη απόσταση, τότε ο νευρώνας-νικητής επιλέγεται τυχαία μεταξύ τους.

Δεδομένου ότι κάθε νευρώνας έχει μια θέση, αυτό σημαίνει ότι γειτονεύει και με ορισμένους άλλους νευρώνες. Αφού βρεθεί ο νευρώνας-νικητής, ενημερώνεται με τα στοιχεία του διανύσματος εισόδου και στη συνέχεια ενημερώνονται οι γείτονες του νευρώνα-νικητή, που βρίσκονται μέσα σε μια προκαθορισμένη γειτονιά. Ακόμη, θα πρέπει να αναφερθεί ότι δεν μεταβάλλονται τα διανύσματα των βαρών των ηττημένων νευρώνων ή όσων βρίσκονται εκτός της περιοχής ανταγωνισμού. Τα διανύσματα των βαρών των νευρώνων προσαρμόζονται σύμφωνα με την εξίσωση:

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)] \quad (2.2)$$

όπου t υποδηλώνει το χρόνο, $w_i(t)$ είναι το διάνυσμα βάρους του i -οστού νευρώνα την χρονική στιγμή t , $x(t)$ είναι το διάνυσμα εισόδου την χρονική στιγμή t και $h_{ci}(t)$ η συνάρτηση γειτνίασης.

Επίσης θα πρέπει να ισχύει $h_{ci}(t) \rightarrow 0$ όταν $t \rightarrow \infty$ [9]. Η συνάρτηση γειτνίασης ορίζει την επιρροή που ασκεί ο νευρώνας-νικητής στους γειτονικούς του νευρώνες και συνήθως έχει τη μορφή ενός πυρήνα εξομάλυνσης.

Σε περίπτωση που υπάρχουν περισσότερα από ένα διανύσματα εισόδου, η διαδικασία επαναλαμβάνεται κυκλικά, μέχρι τα διανύσματα αναφοράς να τείνουν σε μια ασυμπτωτική κατάσταση. Μετά από αρκετές επαναλήψεις της διαδικασίας ορισμένες από τις τοπολογικές σχέσεις του χώρου εισόδου διατηρούνται και τα διανύσματα των βαρών του δικτύου μετατρέπονται σε μια οργανωμένη απεικόνισή τους.

Για την αντιμετώπιση του προβλήματος της ομαδοποίησης των δεδομένων σε χωροπληθείς χάρτες με την χρήση του αλγορίθμου SOM, χρησιμοποιήθηκε το λογισμικό *SOM Toolbox*, που αποτελεί ένα εργαλείο που έχει δημιουργηθεί από την ερευνητική ομάδα του Τεχνολογικού Πανεπιστημίου του Ελσίνκι και λειτουργεί στο περιβάλλον του λογισμικού πακέτου *Matlab* [16, 17].

2.2 Περιγραφή του τρόπου λειτουργίας του αλγόριθμου *SOM Toolbox*

Τα δεδομένα εισόδου μέσω του νευρώνα καταλήγουν σε ένα κομβικό σημείο, στο οποίο γίνεται η επεξεργασία τους και τελικά παράγονται τα δεδομένα εξόδου. Σε κάθε δεδομένο εισόδου αντιστοιχεί μια τιμή βάρους (w) η οποία εκφράζει τη σημασία που έχουν τα δεδομένα (x) στην τελική διαμόρφωση του αποτελέσματος.

Το πρώτο βήμα επεξεργασίας είναι η αρχικοποίηση των βαρών των νευρώνων του δικτύου, συνήθως με επιλογή μικρών τυχαίων τιμών. Η αρχικοποίηση των βαρών στη συγκεκριμένη εφαρμογή γίνεται αυτόματα από το πρόγραμμα. Κάθε νευρώνας του δικτύου έχει δύο είδη «συντεταγμένων», τη θέση του στο πλέγμα και τα βάρη του. Η θέση του στο πλέγμα είναι σταθερή και ορίζει τους γειτονικούς κόμβους του νευρώνα ενώ τα βάρη είναι αυτά που μεταβάλλονται κατά τη διάρκεια της εκπαίδευσης. Μετά την αρχικοποίηση του δικτύου ακολουθούν τρία στάδια:

1. *Το στάδιο ανταγωνισμού*: Για κάθε πρότυπο εισόδου προσδιορίζεται ο νευρώνας-νικητής.
2. *Το στάδιο συνεργασίας*: Εντοπίζεται η τοπολογική γειτονία του νευρώνα-νικητή.
3. *Το στάδιο προσαρμογής βαρών*: Τόσο ο νευρώνας-νικητής όσο και οι γειτονικοί του νευρώνες προσαρμόζουν τα βάρη τους με βάση το πρότυπο εισόδου.

Αρχικά, λοιπόν επιλέγεται τυχαία ένα πρότυπο εισόδου και στη συνέχεια προσδιορίζεται ο νευρώνας-νικητής. Για να προσδιοριστεί ο νευρώνας-νικητής υπολογίζεται η Ευκλείδεια απόσταση, για να βρεθεί ο νευρώνας του οποίου τα βάρη είναι πιο «κοντά» στο πρότυπο εισόδου. Ο νευρώνας με τη μικρότερη απόσταση θεωρείται και ο νευρώνας νικητής για το πρότυπο εισόδου.

Στη συνέχεια, ακολουθεί η διαδικασία της συνεργασίας, όπου ο νευρώνας-νικητής αποτελεί το κέντρο μιας γειτονιάς νευρώνων του πλέγματος οι οποίοι θα συνεργαστούν. Είναι λογικό, η γειτονία αυτή να αποτελείται κυρίως από νευρώνες που βρίσκονται τοπολογικά «κοντά» στο νικητή και να φθίνει όσο απομακρυνόμαστε από αυτόν. Η συνάρτη-

ση γειτνίασης που επιλέχθηκε για την εφαρμογή αυτή είναι η συνάρτηση της καμπύλης Gauss. Κατά τη διάρκεια της εκπαίδευσης το μέγεθος της γειτονιάς μεταβάλλεται. Στην αρχή στη γειτονιά του νευρώνα-νικητή περιλαμβάνεται ένας μεγάλος αριθμός νευρώνων, έτσι ώστε με τις πρώτες προσαρμογές να επιτυγχάνεται τοπολογική διάταξη των διανυσμάτων των βαρών. Προς το τέλος της εκπαίδευσης, είναι επιθυμητή η τελειοποίηση του χάρτη οπότε μειώνεται η τιμή της συνάρτησης γειτνίασης και περιλαμβάνονται ένας ή δύο το πολύ γειτονικοί νευρώνες γύρω από το νικητή.

Τέλος, ακολουθεί η διαδικασία της προσαρμογής των βαρών, η οποία χωρίζεται σε δύο φάσεις, τη φάση διάταξης και τη φάση σύγκλισης. Στη φάση της διάταξης γίνεται η αρχική οργάνωση του χάρτη, κατά την οποία αλλάζουν σημαντικά οι τιμές των βαρών, ώστε να επιτευχθεί η τοπολογική τους διάταξη ενώ η δεύτερη φάση -της σύγκλισης- χρησιμεύει για την ακριβή οργάνωση του χάρτη, αφού θεωρείται ότι ήδη από την προηγούμενη διαδικασία, τα βάρη των νευρώνων έχουν προσαρμοστεί έτσι ώστε να ακολουθούν σε γενικές γραμμές τα πρότυπα εισόδου [18].

Η ποιότητα και η ακρίβεια του αποτελέσματος μπορεί να προσδιοριστεί μέσα από δύο δείκτες αβεβαιότητας, την *αβεβαιότητα κβαντικοποίησης* (quantization error) και την *τοπογραφική αβεβαιότητα* (topographic error):

- *Αβεβαιότητα κβαντικοποίησης*: Είναι η μέση απόσταση μεταξύ του διανύσματος των δεδομένων και της καλύτερης ταυτιζόμενης μονάδας (BMU). Η αβεβαιότητα κβαντικοποίησης δείχνει πόσο ακριβής είναι η αναπαράσταση των πρότυπων εισόδου. Όσο μικρότερος είναι ο δείκτης αυτός τόσο τα διανύσματα αναφοράς προσεγγίζουν καλύτερα τα δεδομένα εισόδου.
- *Τοπογραφική αβεβαιότητα*: Είναι η αναλογία όλων των διανυσμάτων των στοιχείων για τις οποίες η πρώτη και η δεύτερη καλύτερα ταυτιζόμενη μονάδα (BMUs) δεν αποτελούν γειτονικές μονάδες. Η τοπογραφική αβεβαιότητα σχετίζεται με την ποιότητα της τοπολογίας του χάρτη, δηλαδή, κατά πόσο γειτονικοί νευρώνες αντιστοιχούν σε παρόμοια πρότυπα εισόδου.

Ο χάρτης, ο οποίος θα έχει τη μικρότερη μέση αβεβαιότητα κβαντικοποίησης και τοπογραφικής αβεβαιότητας μπορεί να θεωρηθεί ως η βέλτιστη λύση.

3. ΕΦΑΡΜΟΓΗΣ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΤΩΝ ΧΑΡΤΩΝ ΑΥΤΟ-ΟΡΓΑΝΩΣΗΣ

3.1 Διαδικασία επεξεργασίας πρωτογενών δεδομένων

Αρχικά συγκεντρώθηκαν και επεξεργάστηκαν τα δεδομένα της πιλοτικής εφαρμογής. Τα πρωτογενή δεδομένα, αφορούν στο σύνολο του πληθυσμού ανά νομό, της Ελλάδας και πάρθηκαν από πίνακες της ΕΣΥΕ της γενικής απογραφής πληθυσμού του έτους 1991. Στη συνέχεια, υπολογίζεται η πυκνότητα πληθυσμού (D), για κάθε νομό, με τη βοήθεια της σχέσης [19]:

$$D = \frac{N}{A} \quad (3.1)$$

όπου N ο συνολικός πληθυσμός του νομού και A το εμβαδόν του.

Πριν την κατασκευή των νευρωνικών δικτύων, κανονικοποιήθηκαν τα διανύσματα εισόδου, δηλαδή τα δεδομένα της πυκνότητας πληθυσμού. Με την κανονικοποίηση όλες

οι συνιστώσες αποκτούν τιμές στο διάστημα [0,1]. Η κανονικοποιημένη πυκνότητα πληθυσμού (P) υπολογίζεται από τη σχέση:

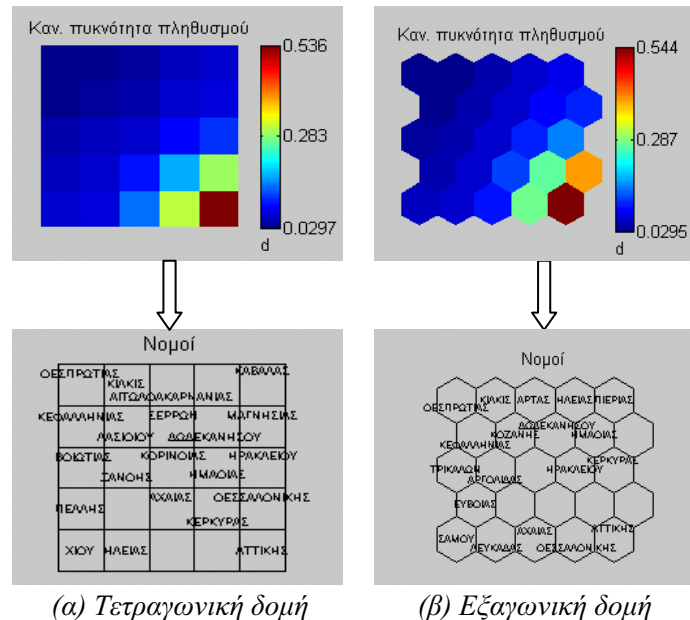
$$P = \frac{D}{D_{\max}} \quad (3.2)$$

όπου D η πυκνότητα πληθυσμού για κάθε νομό και D_{\max} η μεγαλύτερη τιμή της πυκνότητας πληθυσμού.

Το αρχείο με τις κανονικοποιημένες τιμές για κάθε νομό αποτελούν τα δεδομένα εισόδου για τη διαδικασία εκπαίδευσης του αλγόριθμου SOM.

3.2 Εκπαίδευση του δικτύου

Μετά την επεξεργασία των πρωτογενών δεδομένων εισάγονται στο περιβάλλον *SOM Toolbox* τα δεδομένα της κανονικοποιημένης πυκνότητας πληθυσμού καθώς και το όνομα του κάθε νομού. Συνολικά δημιουργήθηκαν 35 δοκιμαστικά νευρωνικά δίκτυα. Μέσα από τη διαδικασία των δοκιμών επιλέχθηκε εκείνο το δίκτυο που μοντελοποιούσε καλύτερα το πρόβλημα της ομαδοποίησης των δεδομένων για χωροπληθείς χάρτες, ανάλογα με τον αριθμό των ομάδων που είχε προκαθοριστεί. Η διαδικασία των δοκιμών είναι εξαιρετικά χρονοβόρα λόγω των δυνατών επιλογών των κατάλληλων παραμέτρων του δικτύου, έτσι ώστε να αναγνωριστεί η αρχιτεκτονική δομή που οδηγεί σε σύγκλιση των νευρωνικών δικτύων αλλά και η επιλογή του βέλτιστου δικτύου για κάθε περίπτωση ομαδοποίησης με κριτήριο την ελαχιστοποίηση των δύο δεικτών αβεβαιότητας. Οι δοκιμές έχουν ως στόχο να προσδιοριστεί σε σχέση με τα συγκεκριμένα δεδομένα ο καταλληλότερος τρόπος εκκίνησης (γραμμική ή τυχαία αρχικοποίηση) καθώς και η επιλογή της καταλληλότερης τοπολογίας του πλέγματος (τετραγωνική ή εξαγωνική). Για παράδειγμα, στο Σχήμα 1 παρουσιάζεται η δομή ενός νευρωνικού δικτύου 25 νευρώνων παραγόμενων



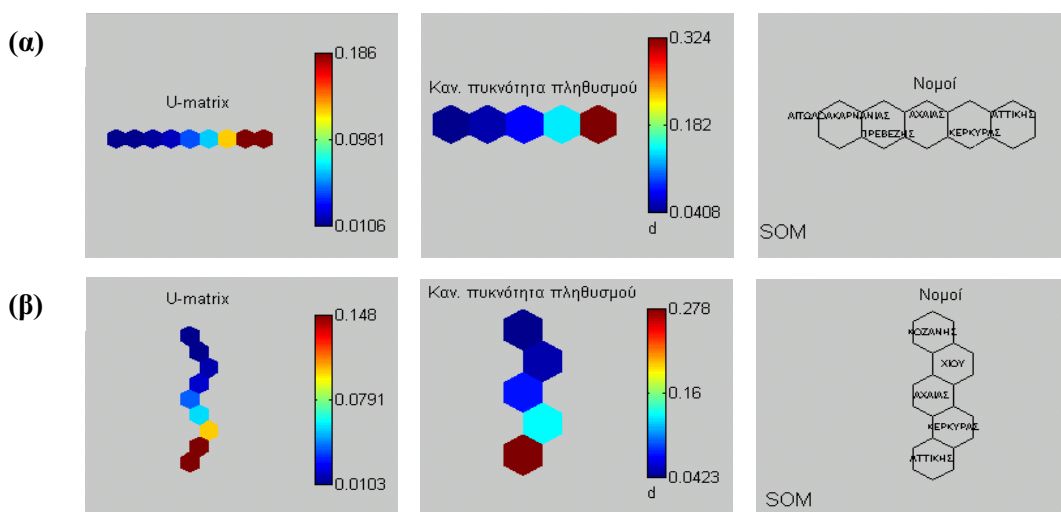
Σχήμα 1: Οπτικοποίηση του αποτελέσματος με τις δύο δομές

από τα αρχικά δεδομένα, με διάταξη (5 5), με τετραγωνική τοπολογία του πλέγματος (Σχήμα 1α) και με εξαγωνική τοπολογία (Σχήμα 1β).

Από τις δοκιμές που έγιναν, παρατηρήθηκε ότι σε σχέση με τη μέθοδο αρχικοποίησης, επιλέγοντας τη γραμμική αρχικοποίηση, οι δύο αβεβαιότητες (κβαντικοποίησης και τοπογραφική) δεν εξαρτώνται από την εκκίνηση του αλγορίθμου αλλά από τις υπόλοιπες παραμέτρους του δικτύου. Επιπλέον, παρατηρήθηκε ότι η αβεβαιότητα κβαντικοποίησης παραμένει ανεξάρτητη από το είδος της τοπολογικής δομής του δικτύου ενώ αντίθετα υπάρχει μεγάλη διαφοροποίηση στην τοπογραφική αβεβαιότητα. Στα δίκτυα με διάταξη τετραγωνικού πλέγματος η τοπογραφική αβεβαιότητα είναι κατά πολύ μεγαλύτερη. Το γεγονός αυτό οφείλεται στις οριζόντιες και κάθετες διευθύνσεις της διάταξης, που ευνοεί η τετραγωνική δομή σε αντίθεση με την εξαγωνική δομή. Για την επιλογή των τελικών παραμέτρων προτιμήθηκε η γραμμική διαδικασία αρχικοποίησης και η εξαγωνική δομή, επειδή δίνουν καλύτερο αποτέλεσμα στον παραγόμενο χάρτη αυτο-οργάνωσης.

Στη συνέχεια, για τις ανάγκες της πιλοτικής εφαρμογής δημιουργήθηκαν νευρωνικά δίκτυα που αποτελούνται από τρία, πέντε και επτά νευρώνες. Η επιλογή αυτή έγινε σύμφωνα με τον χαρτογραφικά επιτρεπόμενο αριθμό ομάδων για χωροπληθείς χάρτες. Πέρα από την τοπολογία που μπορούν να έχουν οι νευρώνες μεταξύ τους (π.χ. τετραγωνική ή εξαγωνική δομή) σημαντικό ρόλο παίζει και ο καθορισμός των διαστάσεων του χάρτη αυτο-οργάνωσης. Έγιναν δοκιμές που αφορούσαν διάφορους συνδυασμούς διαστάσεων ανάλογα με τον αριθμό των ομάδων. Αν για παράδειγμα θέλουμε να δημιουργήσουμε έξι νευρώνες, τότε η διάταξη τους μπορεί να καθοριστεί με τέσσερις διαφορετικούς τρόπους. Μπορούμε να καθορίσουμε τους συνδυασμούς (2 3), (3 2), (6 1) και (1 6), οπότε κάθε ένας από αυτούς θα δίνει και διαφορετικό αποτέλεσμα. Πράγματι, λόγω των διαφορετικών συντεταγμένων που παρουσιάζουν οι νευρώνες του κάθε δικτύου, το αποτέλεσμα της ομαδοποίησης είναι διαφορετικό, χωρίς όμως να παρουσιάζονται μεγάλες διαφοροποιήσεις. Στο Σχήμα 2 παρουσιάζονται τα αποτελέσματα των δύο νευρωνικών δικτύων που έχουν διάσταση πέντε νευρώνων με διάταξη (5 1) (Σχήμα 2α) και (1 5) (Σχήμα 2β).

Ως τελική λύση για την ομαδοποίηση των δεδομένων επιλέγεται εκείνο το νευρωνικό δίκτυο. Με βάση τα αποτελέσματα της ομαδοποίησης με τον αλγόριθμο SOM σε τρεις, πέντε και επτά ομάδες συντάχτηκαν οι αντίστοιχοι χωροπληθείς χάρτες. Στη συνέχεια,



Σχήμα 2: Οπτικοποίηση αποτελέσματος χάρτη αυτο-οργάνωσης (5 1) (α) και (1 5) (β)

εφαρμόστηκαν οι μέθοδοι των ίσων διαστημάτων εμβαδού, των ίσων διαστημάτων, των φυσικών διακοπών, της κανονικής τμηματοποίησης, καθώς και η μέθοδος της βέλτιστης προσαρμογής της απόλυτης αβεβαιότητας.

4. ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΩΝ ΟΜΑΔΟΠΟΙΗΣΗΣ

Με βάση τα αποτελέσματα της ομαδοποίησης με τον αλγόριθμο SOM σε τρεις, πέντε και επτά ομάδες συντάχθηκαν οι αντίστοιχοι χωροπληθείς χάρτες. Στη συνέχεια, εφαρμόστηκαν οι μέθοδοι των ίσων διαστημάτων εμβαδού, των ίσων διαστημάτων, των φυσικών διακοπών, της κανονικής τμηματοποίησης, καθώς και η μέθοδος της βέλτιστης προσαρμογής της απόλυτης απόκλισης (GADF), για τον ίδιο αριθμό ομάδων. Με βάση τα αποτελέσματα ομαδοποίησης των πέντε αυτών μεθόδων συντάχθηκαν οι αντίστοιχοι χωροπληθείς χάρτες. Στην ενότητα αυτή γίνεται σύγκριση του οπτικού αποτελέσματος που δίνουν οι προαναφερθέντες μέθοδοι σε σχέση με το οπτικό αποτέλεσμα που προέκυψε από την προτεινόμενη μέθοδο. Τα αποτελέσματα της ομαδοποίησης οπτικοποιημένα σε πέντε ομάδες με τις πέντε συγκρινόμενες μεθόδους παρουσιάζονται τους Χάρτες 1, 2, 3, 4 & 5. Στον Χάρτη 6 παρουσιάζεται οπτικοποιημένο το αποτέλεσμα της ομαδοποίησης των αρχικών δεδομένων σύμφωνα με τη μέθοδο SOM.

Τέλος, στον πίνακα 1 παρουσιάζεται ο αριθμός των νομών που αντιστοιχούν σε κάθε μια από τις πέντε ομάδες για κάθε μέθοδο ξεχωριστά.

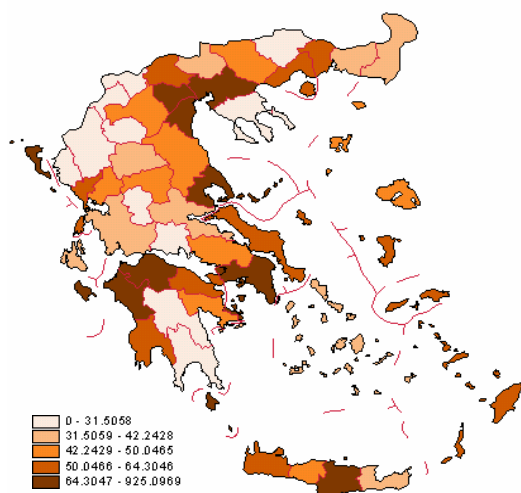
Πίνακας 1: Στατιστικά στοιχεία των μεθόδων ομαδοποίησης σε 5 ομάδες

Ομάδες	Ίσα διαστήματα εμβαδού	Ίσα διαστήματα	GADF	Φυσικές διακοπές	Κανονική τμηματοποίηση	SOM
1 ^η	12	50	22	4	15	21
2 ^η	9	1	27	28	1	16
3 ^η	9	0	1	17	19	11
4 ^η	12	0	1	2	7	2
5 ^η	10	1	1	1	10	2

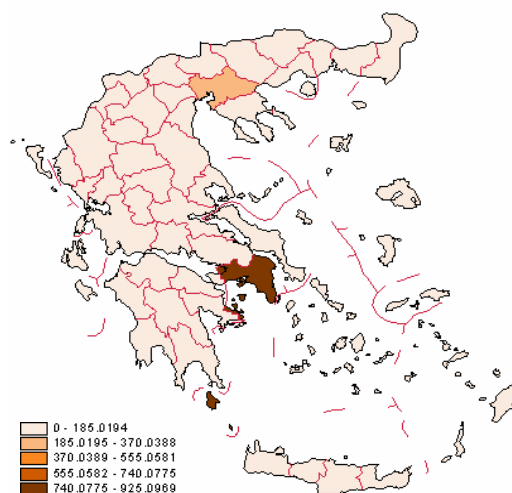
4.1 Ομαδοποίηση των δεδομένων σε 5 ομάδες

Με μια πρώτη παρατήρηση των χαρτών που προέκυψαν με την εφαρμογή της μεθόδου των ίσων διαστημάτων εμβαδού (Χάρτης 1) και της μεθόδου ομαδοποίησης δεδομένων με τον αλγόριθμο SOM (Χάρτης 6) θα μπορούσε να ειπωθεί ότι η πρώτη μέθοδος δίνει καλύτερο οπτικό αποτέλεσμα. Όμως παρατηρώντας και αναλύοντας καλύτερα τους δύο χάρτες, η μέθοδος των ίσων διαστημάτων εμβαδού δεν λαμβάνει υπόψη της την ομοιογένεια των δεδομένων που κατατάσσονται στην ίδια κατηγορία, με αποτέλεσμα ο αναγνώστης του χάρτη να εξάγει λανθασμένα συμπεράσματα. Η μέθοδος των ίσων διαστημάτων εμβαδού ομαδοποιεί 10 νομούς στην 5^η κατηγορία (πίνακας 1), παρά το γεγονός ότι παρουσιάζουν έντονη διαφοροποίηση ως προς τις τιμές της πυκνότητας πληθυσμού, ενώ αντίθετα η μέθοδος που βασίζεται στον αλγόριθμο SOM στην 5^η ομάδα ομαδοποιεί τους

δύο νομούς με τη μεγαλύτερη πυκνότητα πληθυσμού, αυτόν της Αττικής και Θεσσαλονίκης. Με βάση το παραπάνω κριτήριο και παρατηρώντας τους δύο χάρτες διαπιστώνεται ότι με τη μέθοδο των ίσων διαστημάτων εμβαδού ταξινομούνται οι νομοί της Ηλίας και Μαγνησίας στην ίδια κατηγορία με το νομό της Αττικής παρά την έντονη διαφοροποίηση της πυκνότητας πληθυσμού, ενώ αντίθετα με τη μέθοδο που βασίζεται στον αλγόριθμο SOM, τοποθετούνται οι νομοί Ηλίας και Μαγνησίας σε διαφορετική κατηγορία σε σχέση με το νομό της Αττικής. Η διάκριση αυτή προσφέρει στον αναγνώστη του χάρτη μια πιο ρεαλιστική αναπαράσταση των δεδομένων και μπορεί να διαπιστώσει κανείς εύκολα ποιός ή ποιοί νομοί έχουν τη μεγαλύτερη πυκνότητα πληθυσμού. Σε αντίθετη περίπτωση μπορεί να συμπεράνει κανείς π.χ. ότι ο νομός της Αττικής και Μαγνησίας παρουσιάζουν παραπλήσιες πυκνότητες πληθυσμού, γεγονός που δεν είναι αληθινό.



Χάρτης 1: Μέθοδος των ίσων διαστημάτων εμβαδού

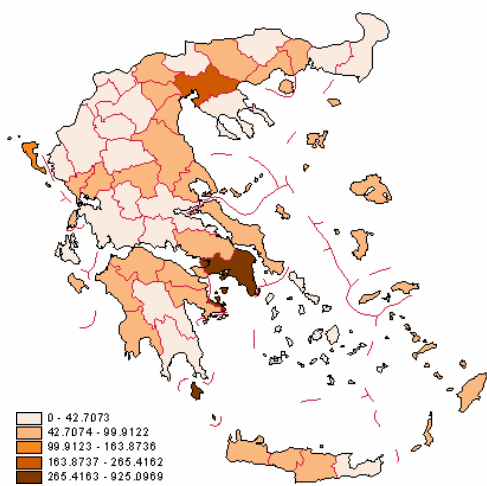


Χάρτης 2: Μέθοδος των ίσων διαστημάτων

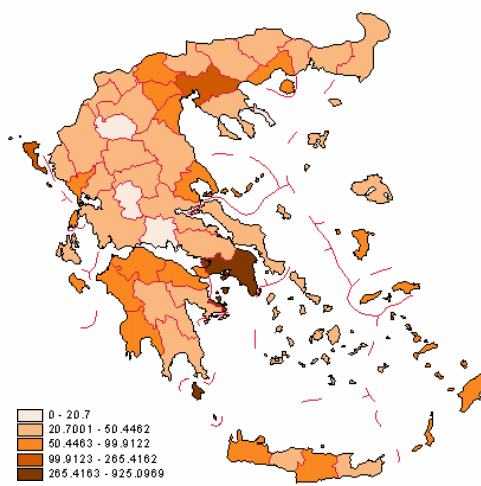
Παρατηρώντας το χάρτη που προέκυψε με την εφαρμογή της μεθόδου των ίσων διαστημάτων (Χάρτης 2), βλέπουμε ότι δεν αποδίδει καλά τα δεδομένα αφού όλοι σχεδόν οι νομοί τοποθετούνται στην 1^η ομάδα και δεν μπορεί να γίνει διάκριση της διαφοροποίησης της πυκνότητας πληθυσμού για κάθε νομό. Επίσης παρατηρούνται κενές ομάδες, αφού στην 3^η και 4^η ομάδα δεν τοποθετείται κανένας νομός. Συγκρίνοντας τη μέθοδο των ίσων διαστημάτων με τη μέθοδο που βασίζεται στον αλγόριθμο SOM, η δεύτερη αποδίδει πολύ καλύτερα τα δεδομένα και μπορεί ο αναγνώστης του χάρτη να αντιληφθεί πως κατανέμεται η πυκνότητα πληθυσμού.

Ακολουθώντας, συγκρίνοντας τα αποτελέσματα που προέκυψαν από τη μέθοδο της βέλτιστης προσαρμογής της απόλυτης απόκλισης -GADF- (Χάρτης 3) με τη μέθοδο που βασίζεται στον αλγόριθμο SOM (Χάρτης 6), με μια πρώτη ματιά μπορούμε να πούμε ότι οι δύο χάρτες είναι πανομοιότυποι. Με μια καλύτερη παρατήρηση όμως φαίνεται ότι με την εφαρμογή του κριτηρίου της βέλτιστης προσαρμογής της απόλυτης απόκλισης υπάρχει μεγάλη συγκέντρωση στοιχείων στην 1^η και τη 2^η ομάδα με αποτέλεσμα να κυριαρχούν στο οπτικό ερέθισμα. Συγκρίνοντας τους δύο αυτούς χάρτες μπορούμε να επισημάνουμε ορισμένες επιπλέον διαφορές. Πρώτα-πρώτα μπορούμε να πούμε ότι στην περιοχή της

Κρήτης η μέθοδος GADF τοποθετεί στην ίδια κατηγορία τους νομούς των Χανίων, Λασηθίου και Ηρακλείου. Το γεγονός αυτό δίνει λανθασμένη πληροφόρηση στον αναγνώστη του χάρτη, δεδομένου ότι μπορεί να συμπεράνει ότι οι τρεις αυτοί νομοί παρουσιάζουν παρόμοιες πυκνότητες πληθυσμού, γεγονός που δεν συμβαίνει στην πραγματικότητα. Αντίθετα, η μέθοδος που βασίζεται στον αλγόριθμο SOM, τοποθετεί το νομό του Ηρακλείου σε ξεχωριστή κατηγορία σε σχέση με τους νομούς των Χανίων και Λασηθίου. Με τον τρόπο αυτό απεικονίζεται η έντονη διαφοροποίηση της πυκνότητας πληθυσμού του νομού Ηρακλείου σε σχέση με τους δύο άλλους νομούς. Παρόμοια διαφορά παρουσιάζεται στην περιοχή της Πελοποννήσου, όπου η μέθοδος GADF ομαδοποιεί στην ίδια κατηγορία τους νομούς Αργολίδας, Μεσσηνίας, Κορινθίας, Ηλείας και Αχαΐας ενώ αντίθετα η μέθοδος SOM ομαδοποιεί σε ξεχωριστή κατηγορία τους νομούς Αργολίδας και Μεσσηνίας. Με τη μέθοδο SOM ο αναγνώστης του χάρτη μπορεί να διακρίνει τη διαφοροποίηση της πυκνότητας πληθυσμού που παρουσιάζουν οι πέντε αυτοί νομοί, γεγονός που δεν συμβαίνει με τη μέθοδο GADF. Τέλος, η παραπάνω διαφορά παρουσιάζεται και στην περιοχή των Δωδεκανήσων αλλά και σε ορισμένους νομούς της Θεσσαλίας. Με βάση τις παραπάνω επισημάνσεις η μέθοδος SOM παρέχει ορθότερα αποτελέσματα.



Χάρτης 3: Μέθοδος βέλτιστης προσαρμογής της απόλυτης απόκλισης (GADF)



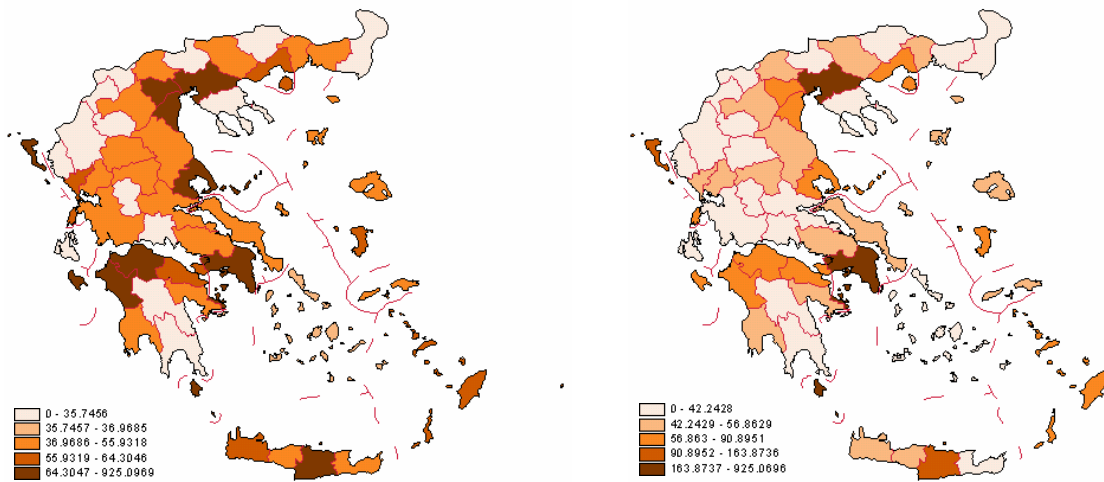
Χάρτης 4: Μέθοδος των φυσικών διακοπών

Ως μειονέκτημα της μεθόδου των φυσικών διακοπών (Χάρτης 4) θεωρείται ότι στην 1^η ομάδα τοποθετεί πολύ μικρό αριθμό δεδομένων, με αποτέλεσμα να κυριαρχούν οπτικά η 2^η και η 3^η ομάδα. Σύμφωνα με τη μέθοδο αυτή ο αναγνώστης του χάρτη δεν μπορεί να διακρίνει τις διαφοροποιήσεις της πυκνότητας πληθυσμού από νομό σε νομό. Για παράδειγμα, στη περιοχή της Βόρειας Ελλάδας, όλοι οι νομοί τοποθετούνται στην ίδια κατηγορία. Οι νομοί Σερρών, Δράμας, Καβάλας, Ξάνθης, Ροδόπης και Έβρου παρουσιάζουν έντονες διαφοροποιήσεις στην πυκνότητα πληθυσμού γεγονός που δεν απεικονίζεται στον χάρτη. Αντίθετα, με τη μέθοδο που βασίζεται στον αλγόριθμο SOM ομαδοποιούνται στην ίδια κατηγορία (1^η ομάδα) οι νομοί Δράμας, Ροδόπης και Έβρου, που παρουσιάζουν παρόμοιες τιμές, οι νομοί Σερρών και Ξάνθης στη 2^η ομάδα και ο νομός Καβάλας στην 3^η ομάδα. Με τον τρόπο αυτό ο αναγνώστης του χάρτη μπορεί να παρατηρήσει τη διαφοροποίηση της πυκνότητας πληθυσμού που παρουσιάζουν οι παραπάνω νομοί. Το ίδιο πρό-

βλημα παρουσιάζεται στην περιοχή της Ηπειρωτικής και Στερεάς Ελλάδας, της Πελοποννήσου και της Κρήτης.

Με την εφαρμογή της μεθόδου της κανονικής τμηματοποίησης (Χάρτης 5) το οπτικό αποτέλεσμα που προκύπτει θα μπορούσε να χαρακτηριστεί ικανοποιητικό, αφού ομαδοποιείται περίπου ίσος αριθμός δεδομένων σε κάθε ομάδα. Αυτό όμως έχει ως αποτέλεσμα να υπάρχει ανομοιογένεια των δεδομένων που τοποθετούνται σε κάθε ομάδα. Για παράδειγμα, οι νομοί της Αττικής και Μαγνησίας τοποθετούνται στην ίδια κατηγορία παρά την έντονη διαφοροποίηση που παρουσιάζει η τιμή της πυκνότητας πληθυσμού τους. Με αυτή τη μέθοδο ομαδοποίησης υπάρχει ενδεχόμενο λανθασμένης πληροφόρησης του αναγνώστη, γιατί αν γνωρίζει ότι η Αττική έχει τη μεγαλύτερη πυκνότητα πληθυσμού, τότε μπορεί να συμπεράνει ότι και οι υπόλοιποι νομοί που τοποθετούνται στην ίδια κατηγορία έχουν παραπλήσιες τιμές. Αντίθετα το πρόβλημα αυτό δεν παρατηρείται στη μέθοδο που βασίζεται στον αλγόριθμο SOM, αφού η μέθοδος αυτή λαμβάνει υπόψη την ομοιογένεια των τιμών του φαινομένου που τοποθετούνται σε κάθε κατηγορία.

Η μέθοδος ομαδοποίησης δεδομένων που βασίζεται στον αλγόριθμο SOM (Χάρτης 6) θα μπορούσαμε να πούμε ότι ταξινομεί καλύτερα τα δεδομένα σε σύγκριση με τις πέντε προϋπάρχουσες μεθόδους που εφαρμόστηκαν, κυρίως γιατί λαμβάνει υπόψη την ομοιογένεια των στοιχείων που τοποθετούνται σε κάθε κατηγορία. Επίσης, το οπτικό αποτέλεσμα που προκύπτει είναι πολύ καλό και μπορεί να μεταδώσει τη χαρτογραφική πληροφορία με σαφήνεια, χωρίς να προκαλεί σύγχυση ή ερωτηματικά στον αναγνώστη του χάρτη.



Χάρτης 5: Μέθοδος κανονικής τμηματοποίησης *Χάρτης 6: Μέθοδος του αλγόριθμου SOM*

5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Με την χρήση του αλγόριθμου SOM ύστερα από επιλογή κατάλληλων παραμέτρων ένα σύνολο δεδομένων μπορεί να ομαδοποιηθεί σε κατηγορίες ώστε να απεικονιστεί υπό τη μορφή χωροπληθών χαρτών. Όμως θα πρέπει να δίνεται ιδιαίτερη σημασία στο τρόπο «χωροθέτησης» των νευρώνων του δικτύου. Αν υποθέσουμε ότι θέλουμε να ομαδοποιήσουμε τα δεδομένα σε έξι ομάδες, τότε το νευρωνικό δίκτυο θα έχει έξι νευρώνες. Όμως,

οι νευρώνες αυτοί μπορούν να καθοριστούν με τέσσερις διαφορετικούς τρόπους διάταξης (2 3), (3 2), (6 1) και (1 6), οπότε κάθε ένας από αυτούς τους τρόπους θα δίνει και διαφορετικό αποτέλεσμα και επομένως επιλέγεται τελικά αυτός που δίνει τις μικρότερες αβειβαιότητες (κβαντικοποίησης και τοπογραφική). Ένα άλλο σημείο το οποίο θα πρέπει να προσεχθεί κατά τη δημιουργία του νευρωνικού δικτύου είναι η επιλογή του κατάλληλου αριθμού βημάτων για την εκτέλεση της διάταξης και της σύγκλισης του αλγορίθμου. Ακόμη, προσοχή θα πρέπει να δοθεί στον τρόπο εισαγωγής των δεδομένων στο νευρωνικό δίκτυο. Θα πρέπει να δημιουργηθεί ένα αρχείο που θα περιλαμβάνει τα δεδομένα κατά αύξουσα ή φθίνουσα σειρά έτσι ώστε το όνομα που θα απεικονίζεται σε κάθε κελί να αντιπροσωπεύει τη μεγαλύτερη πυκνότητα πληθυσμού της συγκεκριμένης ομάδας.

Συγκρίνοντας τις υπάρχουσες μεθόδους με την προτεινόμενη μέθοδο -που βασίζεται στον αλγόριθμο SOM- ως προς το οπτικό αποτέλεσμα, παρατηρούμε ότι με τη μέθοδο των ίσων διαστημάτων εμβαδού τοποθετείται σε κάθε ομάδα ίσος περίπου αριθμός δεδομένων με αποτέλεσμα να μη λαμβάνεται υπόψη η κατανομή των δεδομένων. Αντίθετα, με τη μέθοδο των ίσων διαστημάτων, επειδή διαχωρίζονται τα όρια των κατηγοριών σε ίσα διαστήματα, δημιουργούνται κενές ομάδες. Με τη μέθοδο της βέλτιστης προσαρμογής της απόλυτης απόκλισης παρατηρείται μεγάλη συγκέντρωση στοιχείων σε ορισμένες κατηγορίες με αποτέλεσμα να κυριαρχούν στο οπτικό αποτέλεσμα που προκύπτει. Αντίθετα, η μέθοδος των φυσικών διακοπών εξαρτάται άμεσα από τη φύση των δεδομένων και σε ορισμένες περιπτώσεις παρατηρείται μεγάλη συγκέντρωση στοιχείων σε δύο ή τρεις ομάδες με αποτέλεσμα στις υπόλοιπες να τοποθετείται πολύ μικρός αριθμός δεδομένων. Τέλος, με τη μέθοδο της κανονικής τμηματοποίησης τοποθετείται περίπου ίσος αριθμός δεδομένων σε κάθε ομάδα, χωρίς όμως, να λαμβάνεται υπόψη η ομοιογένεια των τιμών των δεδομένων που τοποθετούνται σε κάθε κατηγορία. Σε αντίθεση με όλες τις υπάρχουσες μεθόδους η προτεινόμενη μέθοδος, που βασίζεται στον αλγόριθμο SOM, λαμβάνει υπόψη της την ομοιογένεια που υπάρχει μεταξύ των δεδομένων που τοποθετούνται σε κάθε κατηγορία με αποτέλεσμα ο τρόπος προσδιορισμού των ορίων των ομάδων να έχει άμεση σχέση με τις τιμές των δεδομένων και με τον τρόπο κατανομής τους.

6. ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Chou, Y.H., 1996. **Exploring spatial analysis in Geographic Information Systems**. On-Word Press: 140-149.
- [2] Νάκος, Β., 2002. **Ομαδοποίηση αριθμητικών δεδομένων**. Διδακτικές Σημειώσεις, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα.
- [3] Robinson, A.H., Sale, R.D., Morrison, J.L., and Muehrcke, Ph.C., 1984. **Elements of Cartography** (5th edition). John Wiley and Sons, New York: 347-366.
- [4] Monmonier, M.S., 1982. Flat laxity, optimization and rounding in the selection of class intervals. **Cartographica**, **19** (1): 16-27.
- [5] Coulson, R.M., 1987. **In the matter of class intervals for choropleth maps: with particular reference to the work of George F Jenks**. University of Calgary, Calgary, Alberta: 16-39.
- [6] Dent, D.B., 1990. **Cartography. Thematic Map Design** (2nd edition). Wm C. Brown Pub., Dubuque: 153-167.

- [7] Slocum, A.T., 1998. **Thematic Cartography and Visualization**, Prentice Hall, New Jersey: 60-74.
- [8] Chang, D., 1978. Visual Aspects of Class intervals in Choroplethic Mapping. **The Cartographic Journal**, **15** (1): 42-48.
- [9] Kohonen, T., 1997. **Self-Organizing Maps** (2nd edition). Springer-Verlag, Berlin.
- [10] Hopfield, J.J., 1982. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. **Proceedings of the National Academy of Sciences of the USA**, Vol. 79: 2554-2558.
- [11] Fausett, L., 1994. **Fundamentals of neural networks**. Prentice Hall, New Jersey.
- [12] Haykin, S., 1994. **Neural Networks: A Comprehensive Foundation**. Macmillan, New York.
- [13] Ρίζος, Γ., 1996. **Τεχνητά Νευρωνικά Δίκτυα, Θεωρία και Εφαρμογές**. Εκδόσεις Νέων Τεχνολογιών, Αθήνα: 26-31.
- [14] Garson, G.D., 1998. **Neural Networks: An Introductory Guide for Social Scientists**. Sage Publications, London.
- [15] Ham, F.M., and Kostanic, I., 2001. **Principles of Neurocomputing for Science & Engineering**. McGraw-Hill, New York.
- [16] Vesanto, J., 2000. **Using SOM in Data Mining**. Helsinki University of Technology, Department of Computer Science and Engineering.
- [17] Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J., 2000. **SOM Toolbox for Matlab 5**, Helsinki University of Technology.
- [18] Matlab Help (Version 6.5), 2002. **The Language of Technical Computing**. The Math Works Inc.
- [19] Νάκος, Β. και Φιλιππακοπούλου, Β., 1992. **Θεματική Χαρτογραφία**. Διδακτικές Σημειώσεις, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα: 105-107.