



**ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΑΞΙΟΠΟΙΗΣΗΣ ΦΥΣΙΚΩΝ ΠΟΡΩΝ
& ΓΕΩΡΓΙΚΗΣ ΜΗΧΑΝΙΚΗΣ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΦΥΣΙΚΟΙ ΠΟΡΟΙ, ΓΕΩΠΕΡΙΒΑΛΛΟΝ, ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ
& ΓΕΩΡΓΙΚΗ ΜΗΧΑΝΙΚΗ**

Μεταπτυχιακή Διατριβή

Μελέτη περιβαλλοντικών επιπτώσεων στην περιοχή του Θριάσιου πεδίου
με μεθόδους χωρικής παρεμβολής
και χωρικής ανάλυσης

Κωνσταντίνος Γ. Σταμάτης

Επιβλέπων Καθηγητής:

Διονύσιος Καλύβας, Καθηγητής ΓΠΑ

**ΑΘΗΝΑ
2021**

**ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΑΞΙΟΠΟΙΗΣΗΣ ΦΥΣΙΚΩΝ ΠΟΡΩΝ
& ΓΕΩΡΓΙΚΗΣ ΜΗΧΑΝΙΚΗΣ**

Μεταπτυχιακή Διατριβή

Μελέτη περιβαλλοντικών επιπτώσεων στην περιοχή του Θριάσιου πεδίου
με μεθόδους χωρικής παρεμβολής
και χωρικής ανάλυσης

“Environmental impact study in the area of Thriassion
with spatial interpolation methods
and spatial analysis”

Κωνσταντίνος Γ. Σταμάτης

Εξεταστική Επιτροπή:

Διονύσιος Καλύβας, Καθηγητής ΓΠΑ (επιβλέπων)

Γεώργιος Παπαδόπουλος, Αναπληρωτής Καθηγητής ΓΠΑ

Μάσσας Ιωάννης, Καθηγητής ΓΠΑ

Μελέτη περιβαλλοντικών επιπτώσεων στην περιοχή του Θριάσιου πεδίου με μεθόδους χωρικής παρεμβολής και χωρικής ανάλυσης

Τμήμα Αξιοποίησης Φυσικών Πόρων και Γεωργικής Μηχανικής

ΠΕΡΙΛΗΨΗ

Το Θριάσιο πεδίο βρίσκεται Δυτικά της πόλης της Αθήνας, πρωτεύουσας της Αττικής. Υπάρχουν δύο μεγάλες πόλεις, η Ελευσίνα και ο Ασπρόπυργος, με έντονη βιομηχανοποίηση. Η περιοχή μελέτης πραγματοποιήθηκε τόσο τον αστικό όσο και τον περιαστικό ιστό των δυο πόλεων. Στόχος της είναι να μελετήσουμε τις επιπτώσεις των βαρέων μετάλλων με τις μεθόδους της χωρικής παρεμβολής και της χωρικής ανάλυσης. Χρησιμοποιώντας τη γεωστατιστική μέθοδο: Ordinary Kriging και τη μη γεωστατιστική μέθοδο εκτίμηση αντίστροφης απόστασης Inverse Distance Weighting (IDW) εφαρμόστηκαν σε δεδομένα, των επιφανειακών και των υποεπιφανειακών σε βάθος 5 εκατοστών, των βαρέων μετάλλων όπως Cr, Zn, Ni, Pb, Co, Mn, Ba, Cu, και Fe. Η μέθοδος cross-validation χρησιμοποιήθηκε για να ελέγξουμε την ακρίβεια των αποτελεσμάτων των μεθόδων πρόβλεψης. Η αξιολόγηση των μεθόδων στηρίχθηκε στις τιμές του δείκτη της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) και του μέσου σφάλματος (M.E.).

Επιστημονική περιοχή: Γεωστατιστικές μέθοδοι

Λέξεις- κλειδιά: Γεωστατιστική, kriging, ordinary kriging, IDW, Moran's I δείκτης, Εκθετικό και Σφαιρικό μοντέλο.

Environmental impact study in the area of Thriassion with spatial interpolation methods and spatial analysis

Department of Natural Resources Development and Agricultural Engineering

ABSTRACT

The Thriasio field is located west of the city of Athens, the capital of Attica. There are two major cities, Eleusis and Aspropyrgos, with strong industrialization. The study area took place in both the urban and the periurban fabric of the two cities. Its aim is to study the effects of heavy metals with the methods of spatial interpolation and spatial analysis. Using the Geostatistical method: Ordinary Kriging and Non-geostatistical method Inverse Distance Weighting (IDW) were applied to data, surface and subsurface at a depth of 5 cm, heavy metals such as Cr, Zn, Ni, Pb, Co, Mn, Ba, Cu, and Fe. The cross-validation method used to control the accuracy of the results of forecasting methods. The evaluation of the methods was based on the values of the root mean square error (R.M.S.E) and the mean error (M.E.).

Scientific area: Geostatistical methods

Keywords : Geostatistical methods, kriging, ordinary kriging , IDW, Moran's I index, Exponential and Spherical type.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	3
ABSTRACT	4
1.ΕΙΣΑΓΩΓΗ – ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ	6
2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ.....	9
2.1. ΜΕΘΟΔΟΙ	9
2.1.1. ΣΤΑΤΙΣΤΙΚΕΣ ΕΠΕΞΕΡΓΑΣΙΕΣ- ΜΗ ΧΩΡΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	9
2.1.2. ΧΩΡΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	14
2.1.2.Α. ΧΩΡΙΚΗ ΑΝΑΛΥΣΗ	14
2.1.2.Β. ΧΩΡΙΚΗ ΠΑΡΕΜΒΟΛΗ	18
2.2. ΠΕΡΙΟΧΗ ΜΕΛΕΤΗΣ	28
2.2.1. ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΕΔΑΦΩΝ	30
2.2.2. ΔΕΙΚΤΕΣ ΜΟΛΥΝΣΗΣ ΕΔΑΦΩΝ	31
2.2.3. ΓΕΩΓΡΑΦΙΚΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	32
3.ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	32
3.1.ΜΗ ΧΩΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	32
3.1.1. ΥΠΟΛΟΓΙΣΜΟΣ OUTLIERS.....	36
3.1.2. ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΜΕΤΑΒΛΗΤΩΝ.....	39
3.1.3. ΕΛΕΓΧΟΣ ΤΗΣ ΔΙΑΦΟΡΑΣ ΤΩΝ ΜΕΣΩΝ ΟΡΩΝ	44
3.1.4. ΟΡΙΣΜΟΣ ΤΕΧΝΙΚΗΣ ΑΝΑΛΥΣΗΣ (TREND ANALYSIS)	45
3.1.5. ΔΙΕΡΕΥΝΗΣΗ ΧΩΡΙΚΩΝ ΑΚΡΑΙΩΝ ΤΙΜΩΝ	48
3.1.6.ΔΙΕΡΕΥΝΗΣΗ ΓΙΑ ΤΗΝ ΥΠΑΡΞΗ ΧΩΡΙΚΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ (SPATIAL AUTOCORRELATION) ΚΑΙ ΕΝΤΟΠΙΣΜΟΣ ΑΠΟΣΤΑΣΗΣ ΣΤΗΝ ΟΠΟΙΑ ΕΜΦΑΝΙΖΕΤΑΙ	49
3.1.7.ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ KRIGING	57
4.ΑΠΟΤΕΛΕΣΜΑΤΑ-ΣΥΜΠΕΡΑΣΜΑΤΑ	68
4.1. ΑΠΟΤΕΛΕΣΜΑΤΑ	68
4.2.ΣΥΜΠΕΡΑΣΜΑΤΑ.....	71
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	73
ΞΕΝΗ.....	73
ΕΛΛΗΝΙΚΗ	74

1.ΕΙΣΑΓΩΓΗ – ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

Το έδαφος αποτελεί τον εξωτερικό φλοιό της Γης και θεωρείται το θεμέλιο της ζωής μας αφού είναι η βάση όλων των ανθρωπίνων γεωργικών δραστηριοτήτων.

Είναι ένας πολύ σημαντικός φυσικός πόρος από τον οποίο εξαρτάται όχι μόνο η οικονομία, σε τοπική και παγκόσμια κλίμακα, αλλά και η διατήρηση τουλάχιστον της χερσαίας ζωής στον πλανήτη μας. Για το έδαφος ως φυσικός σχηματισμός της επιφάνειας της γης έχουν δοθεί διάφοροι ορισμοί, επειδή χρησιμοποιείται από τον άνθρωπο με ποικίλους τρόπους. Δηλαδή το έδαφος ορίζεται ανάλογα με τις εκάστοτε χρήσεις του από τον άνθρωπο. Για παράδειγμα ένας μεταλλειολόγος αντιλαμβάνεται το έδαφος ως αδρανές υλικό που πρέπει να απομακρύνει προκειμένου να εξορύξει τα επιθυμητά μεταλλεύματα. Ένας πολιτικός μηχανικός αντιλαμβάνεται το έδαφος ως υπόστρωμα των δομικών δραστηριοτήτων του και τον ενδιαφέρουν οι ιδιότητες του εδάφους που σχετίζονται με τη συνοχή, συνάφεια και σταθερότητα αυτού του υποστρώματος στα διάφορα δομικά φορτία. Ο γεωπόνος αντιλαμβάνεται το έδαφος ως υπόστρωμα ανάπτυξης φυτών και τον ενδιαφέρουν κυρίως οι ιδιότητες που σχετίζονται με τη γονιμότητα του εδάφους, την άρδευση κ.λπ. Ο διαχειριστής φυσικών πόρων αντιλαμβάνεται το έδαφος ως "ζωντανό" οργανισμό που επιδρά στο περιβάλλον και δέχεται επιδράσεις από αυτό. Τον ενδιαφέρει το σύνολο των εδαφικών ιδιοτήτων και η αλληλεξάρτηση αυτών με τις διάφορες χρήσεις της γης, όπως γονιμότητα, εδαφική υποβάθμιση και διάβρωση εδαφική προστασία, εδαφογένεση και εξέλιξη. Επίσης συσχετίζει το έδαφος με τους υπόλοιπους φυσικούς πόρους και μελετά την αλληλεξάρτηση μεταξύ τους.

Σήμερα, γίνεται δεκτό ότι το έδαφος είναι ένας φυσικός σχηματισμός που δημιουργείται στην επιφάνεια της γης από τα προϊόντα αποσάθρωσης των πετρωμάτων με την μακρόχρονη επίδραση του κλίματος και των ζώντων οργανισμών. Το έδαφος είναι ένα φυσικό, ανοικτό σύστημα που δέχεται επιδράσεις από το περιβάλλον και επιδρά στο περιβάλλον. Υπάρχει με άλλα λόγια μια δυναμική ενεργειακή ισορροπία μεταξύ του εδάφους και του περιβάλλοντος. Λόγω αυτής της δυναμικότητάς του, το έδαφος συνεχώς μεταβάλλει τις ιδιότητές του και κατά συνέπεια τα δομικά του χαρακτηριστικά. Σύμφωνα με αυτόν τον ορισμό (Jenny's 1941) το έδαφος μπορεί να "περιγραφεί" με την ακόλουθη μαθηματική συνάρτηση:

$$s = f (p, c, o, r, t)$$

όπου: s = εδαφική ιδιότητα

p = μητρικό υλικό

c = κλίμα

o = οργανισμοί

r = τοπογραφία

t = χρόνος

όπου p,c,o,r,t : Παράγοντες εδαφογένεσης

Όταν ένας από τους παράγοντες εδαφογένεσης διαφοροποιηθεί, προκύπτει διαφορετικός τύπος εδάφους.

Η ταχεία ανάπτυξη, η εκβιομηχάνιση και η αστικοποίηση έχουν οδηγήσει σε σοβαρή μόλυνση του εδάφους από ανθρωπογενείς πηγές σε πολλές περιοχές του κόσμου (Yaylali-Avanza 2011). Σε αντίθεση με τα περισσότερα οργανικά υλικά, δεν διασπώνται με το χρόνο. Αν και ορισμένα βαρέα μέταλλα (π.χ., Zn, Cu και Mn) είναι απαραίτητα για τους ζωντανούς οργανισμούς μέχρι ορισμένες συγκεντρώσεις, μπορούν να καταστούν επιβλαβείς όταν υπερβαίνουν τις τιμές και οι τοξικές επιδράσεις είναι πιθανό να συμβούν όταν τα βαρέα μέταλλα, εισάγονται στον άνθρωπο σε υψηλότερες δόσεις από αυτά που προτείνει ο Παγκόσμιος Οργανισμός Υγείας (Nordberg et al. 2007).

Η ύπαρξη των διάφορων χημικών ουσιών στο έδαφος δε συνιστά από μόνη της ρύπανση. Οι χημικές αυτές οργανικές και ανόργανες ουσίες για να χαρακτηρισθούν ως ρύποι και να προκαλέσουν ρύπανση στο εδαφικό οικοσύστημα, πρέπει να παρεμποδίζουν μία ή περισσότερες εδαφικές λειτουργίες.

Το έδαφος, μαζί με τα υδάτινα συστήματα, είναι ο μεγαλύτερος αποδέκτης της ανθρωπογενούς ρύπανσης. Οι βιομηχανικές, βιοτεχνικές και εμπορικές δραστηριότητες ρυπαίνουν σε μεγάλο βαθμό το έδαφος. Τα υγρά και στερεά απόβλητα αρχικά εναποτίθενται και επεξεργάζονται σε χερσαίες εγκαταστάσεις. Γεωργικά και κτηνοτροφικά απόβλητα, λιπάσματα και φυτοφάρμακα ρυπαίνουν τις καλλιεργημένες εκτάσεις. Ατυχήματα και διαρροές πετρελαίου, απόβλητα ορυχείων και λατομείων είναι μερικές αιτίες ρύπανσης εδαφών. Το έδαφος γίνεται αποδέκτης και των ατμοσφαιρικών ρύπων που κατακρημνίζονται με αργούς ρυθμούς ανάλογα με τις συνθήκες και τη γεωμορφολογία των περιοχών, οργανικών και ανόργανων ρύπων στο νερό, απορροφούνται από το έδαφος και μεταφέρονται στα διάφορα εδαφικά περιβαλλοντικά διαμερίσματα με τη βροχή και στα υπόγεια νερά. Οι συντελεστές κατανομής ρύπων μεταξύ του εδάφους και νερού, εδάφους και ατμόσφαιρας, εδάφους και οργανικών συστατικών, προδιαγράφουν και την τοξικότητα ρύπων στα φυτά και τα χερσαία ζώα, τη διάσπαση από φυσικές διεργασίες μέσα στο έδαφος και την αποικοδόμηση τους μέσω των εδαφικών μικροοργανισμών. Αυτά είναι τα βασικά προβλήματα της περιβαλλοντικής τοξικολογίας σε σχέση με τους τοξικούς και επικίνδυνους χημικούς ρύπους στα εδάφη.

Σε αρκετές περιοχές, η κύρια πηγή των μετάλλων στο έδαφος είναι η αποσάθρωση του γεωλογικού υποστρώματος μέσω των διαδικασιών εδαφογένεσης. Ωστόσο, ακόμη και σε αυτές τις περιοχές τα εδάφη μπορεί να επηρεαστούν από τη σκόνη και τη μεγάλη βεληνεκούς ατμοσφαιρική μεταφορά των μετάλλων. Δεύτερον, ο εμπλουτισμός των εδαφών από μέταλλα σε περιοχές της ανατολικής λεκάνης της Μεσογείου λόγω ατμοσφαιρικών εναποθέσεων που προέρχονται από την Βόρεια Αφρική, την Κεντρική και Δυτική Ευρώπη, και στη Σαχάρα έχουν αναφερθεί (Chester et al 1996, Κοζακ et al 2007, Koulouvaris et al. 2009).

Αναγνωρίζοντας ότι το έδαφος είναι ένας από τους βασικούς πόρους που στηρίζουν τη ζωή στη γη, την παρακολούθηση της κατάστασης των βαρέων μετάλλων στα εδάφη που επηρεάζονται από τις ανθρώπινες δραστηριότητες είναι πρωταρχικής σημασίας. Οι αυξημένες συγκεντρώσεις βαρέων μετάλλων έχουν παρατηρηθεί σε βιομηχανικές, αστικές (Bretzel, Calderisi 2006, De Miguel et al. 2007, Massas et al 2009) (Srinivasa Gowd et al 2010), ή ακόμα και αγροτικές περιοχές, καθώς και σε περιοχές μικτών χρήσεων γης (Koulouvaris et al 2009, Yaylali-Abanuz 2011).

Ο συμβατικός εδαφολογικός χάρτης περιορίζεται στην εμφάνιση μόνο των πληροφοριών που είναι απαραίτητες για τη διάκριση των χαρτογραφικών μονάδων και δεν έχει τη δυνατότητα αποτύπωσης της διακύμανσης

των τιμών μια εδαφικής ιδιότητας. Η ακριβής χαρτογράφηση μιας περιοχής είναι σημαντική για τη διαχείριση των φυσικών πόρων (Lin et al, 2005).

Οι εδαφικές ιδιότητες διαφέρουν από σημείο σε σημείο (Webster and Burgess, 1983), η δειγματοληψία με έναν πεπερασμένο αριθμό θέσεων δεν είναι επαρκής για την πλήρη απεικόνιση της κατάστασης των εδαφικών συνθηκών. Η συλλογή δειγμάτων σε πολλές θέσεις είναι απαγορευτική λόγω κόστους, χρόνου και ανάλυσης των εδαφικών ιδιοτήτων. Εξαιτίας των παραπάνω παραγόντων οδηγούμαστε στην πρόβλεψη των εδαφικών ιδιοτήτων σε άγνωστες θέσεις. Πρόβλεψη είναι η οποιαδήποτε υποθετική εκτίμηση σχετικά με τις τιμές των εδαφικών ιδιοτήτων σε άλλες θέσεις πέρα των γνωστών.

Για αυτό τον λόγο χρησιμοποιούμε μεθόδους χωρικής παρεμβολής όπως η γεωστατιστική προσέγγιση κατά την οποία το έδαφος θεωρείται συνεχής επιφάνεια και οι εδαφικές ιδιότητες αντιμετωπίζονται ως χωρικές μεταβλητές (regionalized variables). Οι στατιστικές μέθοδοι μπορούν να παρέχουν αξιόπιστες εκτιμήσεις σε μη δειγματοληπτικά σημεία υπό την προϋπόθεση ότι το διάστημα δειγματοληψίας επιλύει τη μεταβολή στο επίπεδο ενδιαφέροντος (Kerry and Oliver, 2004). Σύμφωνα με τη θεωρία των χωρικών δεδομένων τα διαθέσιμα δειγματολογικά δεδομένα είναι αποτέλεσμα μιας τυχαίας διεργασίας ή εκφράσεις μιας τυχαίας συνάρτησης εισάγοντας έτσι ένα βαθμό αβεβαιότητας για την εκτίμηση μιας εδαφικής ιδιότητας (Gressie N.A.C. , 1993).

Οι πιο κοινές τεχνικές παρεμβολής υπολογίζουν τις εκτιμήσεις σε οποιαδήποτε δεδομένη θέση με ένα σταθμισμένο μέσο όρο των κοντινών στοιχείων. Η στάθμιση αντιστοιχίζεται είτε σύμφωνα με ντετερμινιστικό ή στατιστικά κριτήρια. Ένας αριθμός παραγόντων επηρεάζουν την ποιότητα του χάρτη, όπως η φύση της διακύμανσης εδάφους (Sadler et al., 1998), η ένταση της δειγματοληψίας και η μέθοδος παρεμβολής. Η ποικιλία των διαθέσιμων μεθόδων παρεμβολής έχει οδηγήσει σε ερωτήσεις σχετικά με το ποια είναι η πλέον κατάλληλη σε διαφορετικά περιβάλλοντα και έχει προωθήσει αρκετές συγκριτικές μελέτες με σχετική ακρίβεια. Μεταξύ των στατιστικών μεθόδων, η γεωστατιστική τεχνική Kriging και τα είδη αυτής, όπως simple και ordinary kriging, universal kriging και simple cokriging έχουν συχνά χρησιμοποιηθεί για την χωρική ανάλυση (Deutsch, 2002). Μεταξύ των μη γεωστατιστικών μεθόδων παρεμβολής, η μέθοδος εκτίμησης αντίστροφης απόστασης (IDW) και οι τροποποιήσεις της (Nalder και Wein, 1998) εφαρμόζονται πιο συχνά. Οι Kriging και IDW είναι οι πιο συχνά χρησιμοποιούμενες μέθοδοι στη γεωργία πρακτικών (Franzen και Peck, 1995. Weisz et al, 1995). Η Kriging απαιτεί το προκαταρκτικό στάδιο διαμόρφωσης μιας σχέσης διακύμανσης αποστάσεων, ενώ η IDW δεν απαιτεί τέτοιο βήμα και είναι πολύ απλή και γρήγορη. Και οι δύο μέθοδοι υπολογίζουν τις τιμές σε μη δειγματοληπτικές θέσεις με βάση τις μετρήσεις σε γύρω περιοχές.

Η γεωστατιστική εισήχθη στην επιστήμη της εδαφολογίας για την αποτύπωση της εδαφικής διακύμανσης. Οι Burgess και Webster (1980) ήταν οι πρώτοι που χρησιμοποίησαν πρακτικά την μέθοδο Kriging στην εδαφολογία. Πολλές μελέτες έχουν συγκρίνει την IDW και Kriging. Σε ορισμένες περιπτώσεις, η απόδοση του kriging ήταν γενικά καλύτερα από ό,τι η IDW (Hosseini et al, 1994.. Dalthorp κ.ά., 1999 Kravchenko και Bullock, 1999 Kravchenko, 2003 Reinstorf et al., 2005). Ο Warrick et al. (1988) ανέφερε επίσης πως η μέθοδος kriging είναι καλύτερη από ό,τι η IDW για χαρτογράφηση πατάτας και ιδιότητες του εδάφους, όπως τοις εκατό της άμμου, το περιεχόμενο Ca και ο ρυθμός διεύθυνσης του εδάφους. Σε άλλες μελέτες, η IDW γενικά ξεπέρασε την Kriging (Weisz κ.ά., 1995. Nalder και Wein, 1998). Ο Gotway et al. (1996) παρατήρησε καλύτερα αποτελέσματα στην οργανική ύλη του εδάφους και τα επίπεδα NO₃ εδάφους για διάφορους τομείς, όταν η IDW χρησιμοποιήθηκε ως τεχνική παρεμβολή. Συχνά, τα αποτελέσματα

αναμιγνύονται (Schloeder κ.ά., 2001. Mueller κ.ά., 200. Lapen και Hayhoe, 2003). Η μέθοδος kriging μπορεί να επηρεαστεί σημαντικά από τη μεταβλητότητα και τη χωρική δομή των δεδομένων (Leenaers et al., 1990) , από την επιλογή ενός μοντέλου βαριογράμματος, την ακτίνα αναζήτησης και τον αριθμό των πλησιέστερων γειτονικών σημείων που χρησιμοποιούνται για την εκτίμηση. Όπως ήταν αναμενόμενο, η kriging είναι βελτιωμένη σε σχέση με την IDW όταν η χωρική δομή είναι γνωστή.

Στόχος της εργασίας είναι η σύγκριση της μεθόδου Kriging και της Inverse Distance Weighted για τις επιπτώσεις στην περιοχή του Θριάσιου πεδίου. Με την χρήση αυτών των μεθόδων προσδιορίζονται οι συνολικά διαθέσιμες συγκεντρώσεις των βαρέων μετάλλων στην περιοχή, περιγράφεται η κατανομή τόσο των συνολικών όσο και των διαθέσιμων μορφών των μετάλλων στα εδάφη με την ιδιότητα του επί τοις εκατό ποσοστού των βαρέων μετάλλων στην περιοχή του Θριάσιου πεδίου.

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1. ΜΕΘΟΔΟΙ

2.1.1. ΣΤΑΤΙΣΤΙΚΕΣ ΕΠΕΞΕΡΓΑΣΙΕΣ- ΜΗ ΧΩΡΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Για τη διεξαγωγή μίας στατιστικής μελέτης απαιτείται αρχικά η διατύπωση και ο σχεδιασμός του προς επίλυση προβλήματος. Έπειτα συλλέγονται τα δεδομένα, δηλαδή πληροφορίες που έχουν σχέση με την υπό εξέταση μεταβλητή, που αφορούν το προς επίλυση πρόβλημα. Το επόμενο καθοριστικό βήμα, για την ανάλυση των δεδομένων, την ερμηνεία των αποτελεσμάτων και την εξαγωγή συμπερασμάτων, είναι η καταχώρηση των δεδομένων που θα συλλέξουμε σε κάποιο λογισμικό πακέτο ανάλυσης δεδομένων. Γίνεται σαφές ότι η καταχώρηση των προς ανάλυση δεδομένων στο λογισμικό επιδρά στη μετέπειτα ανάλυση και στην εξαγωγή των συμπερασμάτων. Από τη θεωρία της Στατιστικής γνωρίζουμε ότι οι μεταβλητές διακρίνονται σε ποσοτικές και ποιοτικές, οι οποίες με τη σειρά τους διακρίνονται σε διατάξιμες και ονοματικές (Ζωγράφος, 2003).

Ποσοτική μεταβλητή είναι εκείνη που μπορεί να μετρηθεί (έχει δηλαδή αριθμητικές τιμές). Ποιοτική μεταβλητή είναι εκείνη που περιγράφει χαρακτηριστικά του πληθυσμού που μεταβάλλονται κατά ποιότητα ή είδος, αλλά όχι κατά μέγεθος.

Η γεωστατιστική μεθοδολογία έχει καθιερωθεί τις τελευταίες δεκαετίες ως η μόνη επίσημα αποδεκτή τεχνική διεθνώς για την εκτίμηση μεταλλευτικών αποθεμάτων καθώς και για την αντιμετώπιση ενός μεγάλου εύρους προβλημάτων παρεμβολής και εκτίμησης. Γνώρισε και γνωρίζει συνεχή ανάπτυξη και αποτελεί αντικείμενο έρευνας παγκοσμίως. Είναι η εφαρμογή της θεωρίας των πιθανοτήτων σε δεδομένα που απαντώνται στον χώρο μας και επηρεάζονται μόνο από την θέση τους στο χώρο αυτό. Η διαφορά από την συμβατική στατιστική είναι το ότι προσπαθεί να αναδείξει την δομή των στοιχείων και την σχέσης τους στο χώρο ανεξαρτήτως της δεδομένης πολυπλοκότητας που έχουν τα φυσικά φαινόμενα. Το πεδίο εφαρμογής της είναι ουσιαστικά απεριόριστο και περιλαμβάνει μεταξύ των άλλων εκτίμηση αποθεμάτων των κοιτασμάτων, εκτίμηση κατανομής ιχνοστοιχείων, στερεών υγρών και αερίων ρυπαντών, του τοπογραφικού ανάγλυφου και γενικά σε όλα τα δεδομένα που μετρούνται σε μια θέση με ορισμένες γεωγραφικές συντεταγμένες. Οι τεχνικές αυτές βασίζονται στην υπόθεση ότι η χωρική

διακύμανση της μεταβλητής είναι τυχαία, οπότε χρησιμοποιούν στατιστικές μεθοδολογίες για οποιαδήποτε εκτίμηση προέρχεται από τις σημειακές μετρήσεις της μεταβλητής. Σημαντικό πλεονέκτημα των γεωστατιστικών μεθόδων είναι το γεγονός ότι ποσοτικοποιούν και τελικά ελαχιστοποιούν το σφάλμα εκτίμησης. Ωστόσο, οι μέθοδοι είναι αρκετά πολύπλοκες στην εφαρμογή τους, η οποία προϋποθέτει τη χρήση κατάλληλων υπολογιστικών προγραμμάτων.

Σύμφωνα με τον Ζωγράφο (2003) «η συνοπτική παρουσίαση των δεδομένων ποσοτικών μεταβλητών περιλαμβάνει τον υπολογισμό των τιμών διάφορων στατιστικών μέτρων, όπως η μέση τιμή (Mean), η τυπική απόκλιση (Std Deviation), οι συντελεστές κύρτωσης και λοξότητας (Kurtosis, Skewness, αντίστοιχα), η διάμεσος (median), η επικρατούσα τιμή (mode), το εύρος (range), τα ποσοστιαία σημεία (Percentile values) κ.ά. Το δεύτερο στάδιο περιλαμβάνει την πιθανή κατασκευή του ιστογράμματος (histogram) και θηκογράμματος (boxplot) της υπό εξέταση ποσοτικής μεταβλητής, τον έλεγχο ύπαρξης ακραίων τιμών στις δειγματικές τιμές της υπό εξέταση μεταβλητής, καθώς και τον έλεγχο αν οι διαθέσιμες δειγματικές τιμές μπορούν να θεωρηθούν ότι προέρχονται από έναν πληθυσμό που περιγράφεται ικανοποιητικά από την κανονική κατανομή».

Ο υπολογισμός μέτρων κεντρικής θέσης και διασποράς :

- Ο αριθμητικός μέσος (mean) είναι ένα μέτρο κεντρικής τάσης. Υπολογίζεται από το άθροισμα όλων των τιμών διαιρούμενο με το πλήθος των παρατηρήσεων. Βασικό πλεονέκτημά του είναι ότι βασίζεται σε όλες τις τιμές του δείγματος ενώ μειονέκτημά του είναι ότι παρασύρεται από τις ακραίες τιμές και μπορεί να οδηγήσει αν δεν ερμηνευθεί σωστά σε λανθασμένα συμπεράσματα.
- Η διάμεσος (median) είναι η μεσαία τιμή μιας ομάδας τιμών ιεραρχημένων σε αύξουσα τάξη μεγέθους.
- Επικρατούσα Τιμή (mode): Η τιμή με τη μεγαλύτερη συχνότητα εμφάνισης. Χρησιμοποιείται συνήθως σε ποιοτικές μεταβλητές.
- Η τυπική απόκλιση (Std. deviation) είναι η τετραγωνική ρίζα της διακύμανσης. Είναι προφανές πως όσο μικρότερη είναι η τιμή της διακύμανσης και ως εκ τούτου και της τυπικής απόκλισης τόσο μικρότερη είναι η διασπορά των τιμών γύρω από το μέσο όρο.
- Η ασυμμετρία (skewness) μετρά τον βαθμό στον οποίο ο μεγάλος όγκος των μετρήσεων σε μία κατανομή βρίσκεται προς το ένα ή το άλλο μέρος του μέσου όρου. Μεγάλος βαθμός ασυμμετρίας σημαίνει ότι τα δεδομένα δεν ακολουθούν την κανονική κατανομή και επομένως δεν προέρχονται από κανονικά κατανεμημένο πληθυσμό.

$$s = E (x - \mu)^3 / \sigma^3 \quad (1)$$

όπου μ είναι η μέση τιμή του x , σ είναι η τυπική απόκλιση του x , και $E (t)$ αντιπροσωπεύει την αναμενόμενη τιμή της ποσότητας t .

- Η κύρτωση (kurtosis) αποτελεί την αιχμηρότητα της κορυφής μιας κατανομής. Κατηγορίες που αναγνωρίζονται: Λεπτόκυρτη (kurtosis > 3), Πλατύκυρτη (kurtosis < 3) Μεσόκυρτη (kurtosis = 3).

$$k = E (x - \mu)^4 / \sigma^4 \quad (2)$$

όπου μ είναι η μέση τιμή του x , σ είναι η τυπική απόκλιση του x , και $E (t)$ αντιπροσωπεύει την αναμενόμενη τιμή της ποσότητας t .

- Εύρος (range) είναι η διαφορά μεταξύ της μεγαλύτερης (max) και της μικρότερης τιμής (min) των δεδομένων.

Ο έλεγχος κανονικότητας για να ελέγξουμε αν η κατανομή μιας μεταβλητής είναι συμβατή με την κανονική εφαρμόζουμε το test Kolmogorov-Smirnov και Shapiro-Wilk. Οι υποθέσεις που κάνουμε είναι οι εξής:

H_0 : Το δείγμα προέρχεται από κανονικό πληθυσμό.

H_1 : Το δείγμα δεν προέρχεται από κανονικό πληθυσμό.

Αν p -value (Sig) < α , $\alpha = 0,05$ τότε απορρίπτεται η H_0 .

Για τη διεξαγωγή των ελέγχων υποθέσεων χρησιμοποιούνται κάποιοι μαθηματικοί τύποι, που καλούνται ελεγχουσυναρτήσεις. Με βάση το αποτέλεσμα τους οδηγούμαστε στο συμπέρασμα ότι η μηδενική υπόθεση απορρίπτεται ή όχι. Το επίπεδο στατιστικής σημαντικότητας ορίζεται συνήθως ίσο με 0.05 ή 5%. Το παρατηρηθέν επίπεδο στατιστικής σημαντικότητας ορίζεται ως η πιθανότητα η τιμή του ελέγχου (ελεγχουσυνάρτησης) να πάρει μία τιμή τόσο ακραία ή περισσότερο ακραία από αυτή που πήρε στο συγκεκριμένο δείγμα κάτω από τη μηδενική υπόθεση. Αν η p -value είναι μικρότερη του 0.05, τότε λέμε ότι η μηδενική υπόθεση απορρίπτεται. Αν η p -value είναι μεγαλύτερη ή ίση του 0.05, τότε λέμε ότι η μηδενική υπόθεση δεν απορρίπτεται. Ο λόγος που χρειαζόμαστε την κανονικότητα των δεδομένων, είναι για να έχουν ισχύ κάποιες στατιστικές τεχνικές που θα χρησιμοποιήσουμε όπως οι έλεγχοι υποθέσεων για τους μέσους, η γραμμική παλινδρόμηση, η ανάλυση διακύμανσης κ.ά

Στη θεωρία πιθανοτήτων και της στατιστικής, ο συντελεστής διακύμανσης (CV), είναι γνωστός ως σχετική τυπική απόκλιση (RSD), δηλαδή είναι ένα τυποποιημένο μέτρο της διασποράς μιας κατανομής πιθανότητας ή κατανομή συχνότητας. Συχνά εκφράζεται ως ποσοστό, και ορίζεται ως ο λόγος της τυπικής απόκλισης σ με τον αριθμητικό μέσο μ .

$$CV = \sigma/\mu \quad (3)$$

Δείχνει την έκταση της μεταβλητότητας σε σχέση με το μέσο όρο του πληθυσμού. Ο συντελεστής μεταβλητότητας πρέπει να υπολογίζεται μόνο για τα δεδομένα που μετρούνται σε μια κλίμακα αναλογίας, καθώς αυτά είναι μετρήσεις που μπορεί να πάρει μόνο μη αρνητικές τιμές. Ο συντελεστής διακύμανσης μπορεί να μην έχει κανένα νόημα για τα δεδομένα σε μια υποδιαίρεση της κλίμακας. Ο συντελεστής διακύμανσης είναι χρήσιμος επειδή η τυπική απόκλιση των δεδομένων πρέπει πάντοτε να εννοείται στο πλαίσιο του μέσου όρου των δεδομένων. Σε αντίθεση, η πραγματική τιμή της CV είναι ανεξάρτητη από τη μονάδα στην οποία έχει ληφθεί η μέτρηση, έτσι είναι ένας αδιάστατος αριθμός. Για λόγους σύγκρισης μεταξύ συνόλων δεδομένων με διαφορετικές μονάδες ή πολύ διαφορετικά μέσα, πρέπει κανείς να χρησιμοποιήσει τον συντελεστή μεταβλητότητας, αντί της τυπικής απόκλισης.

Όταν η μέση τιμή είναι κοντά στο μηδέν, ο συντελεστής διακύμανσης θα προσεγγίσει το άπειρο και επομένως είναι ευαίσθητα σε μικρές μεταβολές στη μέση. Αυτή είναι συχνά η περίπτωση που οι τιμές δεν προέρχονται από μια κλίμακα αναλογίας. Σε αντίθεση με την τυπική απόκλιση, αυτό δεν μπορεί να χρησιμοποιηθεί άμεσα για την δημιουργία διαστημάτων εμπιστοσύνης για τον αριθμητικό μέσο. Η εκτίμηση παραλλακτικότητας δεν είναι ένας ιδανικός δείκτης της βεβαιότητας της μέτρησης όταν ο αριθμός των επαναλήψεων ποικίλλει ανάλογα με τα δείγματα, επειδή είναι αναλλοίωτη με τον αριθμό των επαναλήψεων, ενώ η βεβαιότητα του μέσου βελτιώνεται με την αύξηση επαναλήψεων. Σε αυτή την περίπτωση το τυπικό σφάλμα σε ποσοστό τείνει να είναι μεγαλύτερο.

Προϋπόθεση για τον υπολογισμό των ακραίων τιμών (outliers) είναι η κανονικότητα των πληθυσμών από τους οποίους προέρχονται τα δείγματά μας. Ο υπολογισμός του ανώτερου και κατώτερου ορίου γίνεται με βάση τους παρακάτω τύπους :

$$U = Q_3 + (2.2 * (Q_3 - Q_1))^{(4)}$$

$$L = Q_1 - (2.2 * (Q_3 - Q_1))^{(5)}$$

Όπου Q_1 είναι το κατώτατο τεταρτημόριο, Q_2 είναι η μεσαία τιμή και Q_3 είναι το ανώτατο τεταρτημόριο.

Η ανίχνευση είναι τόσο εύκολη όσο και δύσκολη. Είναι εύκολη, δεδομένου ότι υπάρχουν πολλές σχετικά απλές δοκιμές για την παρουσία των ακραίων τιμών. Είναι δύσκολη γιατί δεν υπάρχουν αυστηροί κανόνες ως προς το πότε η αφαίρεση ακραίων τιμών είναι κατάλληλη. Οι ακραίες τιμές μπορεί να οφείλονται σε :

- Σφάλμα μέτρησης
- Πειραματικό σφάλμα.

Οι ακραίες τιμές μπορούν ή δεν μπορούν να είναι πρόβλημα, ανάλογα με πολλούς παράγοντες:

- Μερικά στατιστικά τεστ είναι ισχυρά και μπορεί να φιλοξενήσουν ακραίες τιμές, άλλα μπορεί να επηρεαστούν σοβαρά από ακραίες τιμές.
- Το παραμετρικό τεστ μπορεί αδικαιολόγητα να επηρεάζεται.
- Ορισμένοι τύποι δεδομένων θα περιέχουν φυσικά ακραίες τιμές.
- Τα επίπεδα ακτινοβολίας έχουν συχνά ακραίες τιμές (αιχμές).

Η παρουσία των ακραίων τιμών μπορεί, στην πραγματικότητα, να είναι ενδιαφέρον. Η ακραία(ες) τιμή(ες), μπορεί να πέσει σε μια περιοχή επικάλυψης πληθυσμού. Αυτό το είδος της ακραίας πρέπει να αφαιρεθεί από το σύνολο δεδομένων.

Ο συντελεστής γραμμικής συσχέτισης μεταξύ δύο μεταβλητών συμβολίζεται με το r ή Pearson product-moment. Μια θετική τιμή r εκφράζει μια θετική σχέση μεταξύ των δύο μεταβλητών (όσο μεγαλύτερο είναι το A , τόσο μεγαλύτερο το B), ενώ μια αρνητική τιμή r υποδεικνύει μια αρνητική σχέση (όσο μεγαλύτερο το A , τόσο μικρότερο το B). Ένας συντελεστής συσχέτισης μηδέν δείχνει καμία σχέση μεταξύ των μεταβλητών. Ακόμη και αν ο συντελεστής συσχέτισης είναι μηδέν, θα μπορούσε να υπάρχει μια μη γραμμική σχέση.

Σύγκριση Μέσων Όρων t-test μία ανεξάρτητη μεταβλητή με δύο επίπεδα, και μία εξαρτημένη μεταβλητή και οι δύο μορφές του τεστ έχουν το ίδιο σκεπτικό:

Συλλέγουμε δείγματα δεδομένων και υπολογίζουμε τους μέσους όρους τους στη μεταβλητή που μας ενδιαφέρει. Αν τα δείγματα προέρχονται από τον ίδιο πληθυσμό, αναμένουμε ότι οι μέσοι όροι τους δεν θα παρουσιάζουν στατιστικά σημαντική διαφορά (θα είναι σχεδόν ίδιοι), και η διαφορά που υπάρχει θα οφείλεται σε τυχαίο σφάλμα μέτρησης (η πιθανότητα η διαφορά να οφείλεται σε τυχαίο σφάλμα μέτρησης θα είναι πολύ μικρή)

Εξετάζουμε την υπόθεση ότι οι μέσοι όροι των δειγμάτων δεν θα διαφέρουν μεταξύ τους [Μηδενική υπόθεση (H_0)].

Ουσιαστικά, συγκρίνουμε τη διαφορά μεταξύ των μέσων όρων των δειγμάτων μας με τη διαφορά των μέσων όρων που θα περιμέναμε να λάβουμε τυχαία. Χρησιμοποιούμε το τυπικό σφάλμα (standard error) για τον υπολογισμό αυτό. Αν το τυπικό σφάλμα είναι μικρό, περιμένουμε ότι οι μέσοι όροι των δειγμάτων δεν θα διαφέρουν. Όταν το τυπικό σφάλμα είναι μεγάλο, περιμένουμε ότι οι μέσοι όροι των δειγμάτων θα έχουν μεγάλες τυχαίες

διαφορές. Αν οι μέσοι όροι των δειγμάτων διαφέρουν περισσότερο από ότι θα αναμέναμε με βάση το τυπικό σφάλμα, τότε μπορούμε να κάνουμε δύο υποθέσεις:

$$t = \frac{\text{Παρατηρούμενη διαφορά μεταξύ των μέσων όρων} - \text{αναμενόμενη διαφορά μεταξύ των μέσων όρων των πληθυσμών (αν η } H_0 \text{ ισχύει)}}{\text{Εκτίμηση του τυπικού σφάλματος της διαφοράς μεταξύ των δειγμάτων}}$$

A. Οι μέσοι όροι στον πληθυσμό μας κυμαίνονται σε μεγάλο βαθμό λόγω τύχης, και έχουμε συλλέξει δύο δείγματα που δεν είναι αντιπροσωπευτικά του πληθυσμού από τον οποίο προέρχονται.

B. Τα δύο δείγματα προέρχονται από διαφορετικούς πληθυσμούς, και είναι αντιπροσωπευτικά των πληθυσμών από τους οποίους προέρχονται (με άλλα λόγια η διαφορά των δύο δειγμάτων είναι πραγματική και η μηδενική υπόθεση πρέπει να απορριφθεί).

Καθώς μεγαλώνει η παρατηρούμενη διαφορά των μέσων όρων των δύο δειγμάτων αυξάνεται η βεβαιότητα μας ότι η δεύτερη εξήγησή μας είναι σωστή (ότι δηλαδή η μηδενική υπόθεση θα πρέπει να απορριφθεί).

Η ακριβής μορφή της εξίσωσης θα εξαρτηθεί από το αν τα δείγματα είναι ανεξάρτητα ή εξαρτημένα. Από τις προϋποθέσεις (assumptions), τα δεδομένα αν ακολουθούν την κανονική κατανομή, αν η μέτρηση έχει γίνει τουλάχιστον σε κλίμακα ίσων διαστημάτων. Για το τεστ με ανεξάρτητα δείγματα:

- Η διακύμανση των δύο πληθυσμών είναι περίπου ίση (homogeneity of variance).
- Είναι ανεξάρτητα (προέρχονται από διαφορετικά άτομα).

Τα Independent Samples t-test (Independent measures, or independent means) η συχνότερα απαντημένη μορφή ελέγχου t χρησιμοποιείται για τη σύγκριση των μέσων όρων δύο ανεξάρτητων δειγμάτων. Έχουμε δύο επίπεδα της ανεξάρτητης μεταβλητής και κάθε επίπεδο περιλαμβάνει διαφορετικούς συμμετέχοντες. Όταν αυτή η διαφορά είναι στατιστικά σημαντική θεωρούμε ότι τα δύο δείγματα ανήκουν σε διαφορετικούς πληθυσμούς, και δεν οφείλεται σε τυχαίο σφάλμα δειγματοληψίας. Επίσης χρησιμοποιούμε το t-test για ζευγαρωτές παρατηρήσεις (σύγκριση κατά ζεύγη, εξαρτημένα δείγματα).

Στη στατιστική, η ANOVA είναι μια τεχνική που χρησιμοποιείται για να συγκρίνει μέσα από τρία ή περισσότερα δείγματα (χρησιμοποιώντας την κατανομή F). Αυτή η τεχνική μπορεί να χρησιμοποιηθεί μόνο για αριθμητικά δεδομένα. Η ANOVA ελέγχει την μηδενική υπόθεση ότι τα δείγματα σε δύο ή περισσότερες ομάδες που προέρχονται από τους πληθυσμούς με τις ίδιες μέσες τιμές. Για να γίνει αυτό, οι δύο εκτιμήσεις έγιναν από την διακύμανση του πληθυσμού. Οι εκτιμήσεις αυτές βασίζονται σε διάφορες υποθέσεις. Τα αποτελέσματα της ANOVA μπορεί να θεωρηθούν αξιόπιστα όσο διάστημα πληρούνται οι ακόλουθες υποθέσεις: αν τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διακυμάνσεις και οι παρατηρήσεις είναι ανεξάρτητες. Η ANOVA είναι μια σχετικά

αξιόπιστη διαδικασία σε σχέση με τις παραβιάσεις του στην υπόθεση ομαλότητας (Kirk, RE 1995). Η ANOVA παράγει μια F-στατιστική, ο λόγος της διακύμανσης που υπολογίζεται μεταξύ των μέσων για τη διακύμανση εντός των δειγμάτων. Εάν οι μέσοι της ομάδας που προέρχονται από τους πληθυσμούς με τις ίδιες μέσες τιμές, η διακύμανση μεταξύ των μέσων της ομάδας πρέπει να είναι χαμηλότερη από την διακύμανση των δειγμάτων. Συνεπώς, η μεγαλύτερη αναλογία υποδηλώνει ότι τα δείγματα προέρχονται από πληθυσμούς με διαφορετικά μέσες τιμές. Συνήθως, όμως, η ANOVA με ένα παράγοντα χρησιμοποιείται για τις διαφορές μεταξύ των τουλάχιστον τριών ομάδων, δεδομένου ότι η υπόθεση δύο ομάδων μπορεί να καλυφθεί από ένα t-test (Gosset, 1908). Όταν υπάρχουν μόνο δύο μέσα για να συγκρίνουν, το t-test και το F-test είναι ισοδύναμες, η σχέση μεταξύ ANOVA και t δίνεται από $F = t^2$. Η προέκταση της one-way ANOVA είναι η two-way ANOVA που εξετάζει την επίδραση των δύο διαφορετικών κατηγορηματικών ανεξάρτητων μεταβλητών σε μία εξαρτημένη μεταβλητή.

2.1.2. ΧΩΡΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

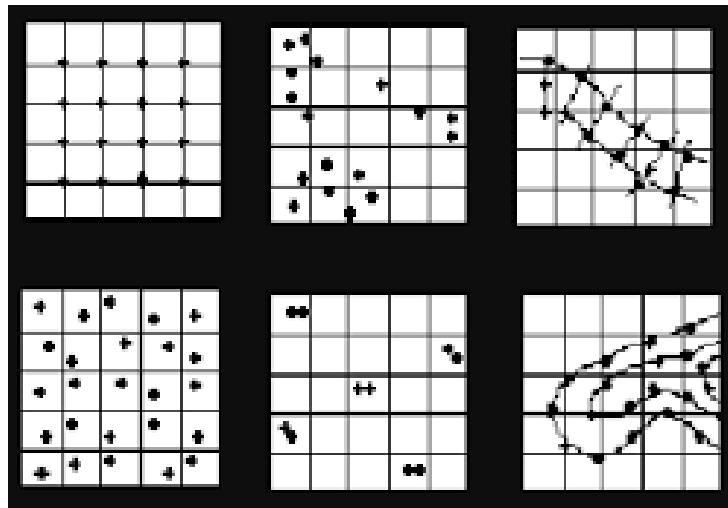
2.1.2.α. ΧΩΡΙΚΗ ΑΝΑΛΥΣΗ

Η Χωρική Ανάλυση (Spatial Analysis) αποτελεί μια από τις βασικές περιοχές ενδιαφέροντος της επιστήμης της Γεωγραφίας. Σκοπός της Γεωγραφικής Ανάλυσης είναι η μελέτη των γεωγραφικών κατανομών με έμφαση στον προσδιορισμό χωρικών προτύπων, την ερμηνεία των διαδικασιών που τα προκαλούν και τη γενίκευσή τους. Πεδία εφαρμογής είναι οι χωρικές κατανομές τόσο του ανθρωπογενούς όσο και του φυσικού περιβάλλοντος. Η σύγχρονη Χωρική Ανάλυση έχει τις ρίζες της στην Ποσοτική Επανάσταση της δεκαετίας του 1960 στις ΗΠΑ, όταν μαθηματικές μέθοδοι από τις φυσικές επιστήμες εισήχθησαν στη Γεωγραφία με ανάπτυξη νέων μαθηματικών προσαρμοσμένων στην ανάλυση των γεωγραφικών δεδομένων. Από την εποχή αυτή η έμφαση είναι στη στατιστική ανάλυση των χωρικών δεδομένων. Στο πλαίσιο αυτό η σύγχρονη Γεωγραφική Ανάλυση συνδέεται με τον έλεγχο υποθέσεων, τη δόμηση μοντέλων και την ανάπτυξη θεωρίας με αυξανόμενη χρήση μαθηματικών και στατιστικών στοιχείων.

Η κλασική Γεωγραφική Ανάλυση χρησιμοποιεί λογισμικό στατιστικής ανάλυσης για τη μελέτη γεωγραφικών φαινομένων τα οποία είναι οργανωμένα σε μορφή πινάκων στους οποίους οι παρατηρήσεις (γραμμές) έχουν κάποια γεωγραφική διάσταση, όπως γεωγραφικές περιφέρειες (πολύγωνα) ή σημεία (π.χ. πόλεις). Μία σύγχρονη εξέλιξη της Γεωγραφικής Ανάλυσης αποτελεί η ενσωμάτωση της γεωγραφικής διάστασης στις στατιστικές τεχνικές, με την έννοια του υπολογισμού της γειτνίασης των γεωγραφικών οντοτήτων, δεδομένου ότι βασική ιδιότητα των γεωγραφικών δεδομένων είναι η συσχέτιση των χαρακτηριστικών γειτονικών περιοχών. Η εξέλιξη αυτή συχνά ονομάζεται τοπική χωρική ανάλυση (local spatial analysis), κυρίως σε σχέση με μοντέλα παλινδρόμησης και έχει διευκολυνθεί από την ανάπτυξη του λογισμικού των Συστημάτων Γεωγραφικών Πληροφοριών. Η χωρική – γεωγραφική ανάλυση προϋπήρχε των GIS. Ανάλογα με τον τρόπο χωρικής κατανομής των θέσεων διακρίνουμε έξι σχήματα δειγματοληψίας:

- Κανονική, στην οποία οι θέσεις δειγματοληψίας απέχουν μεταξύ τους σταθερή απόσταση που πρακτικά επιτυγχάνεται με δειγματοληψία στις κορυφές ενός τετραγωνικού κάναβου.
- Τυχαία, όπου δεν ακολουθείται κανένας κανόνας για τις αποστάσεις δειγματοληψίας με αποτέλεσμα να εμφανιστούν τμήματα της περιοχής μελέτης στα οποία να μην έχει γίνει ούτε μια παρατήρηση.

- Εγκάρσια , που χαρακτηρίζει δειγματοληψίες κατά μήκος μιας διεύθυνσης ή πιο απλά κατά μήκος μιας γραμμής.
- Τυχαία επιβλεπόμενη, όταν εφαρμόζεται ίδια πυκνότητα δειγματοληψίας ανά μονάδα επιφάνειας π.χ. ένα δείγμα ανά στρέμμα, ένα δείγμα ανά τετραγωνικό χιλιόμετρο. Εδώ δεν απαιτείται να ισαπέχουν τα δείγματα, αλλά να εμφανίζουν μια ομοιόμορφη χωρική κατανομή.
- Σε ομάδες, όταν τα δείγματα σχηματίζουν στο χώρο ομάδες, με αποτέλεσμα σε μεγάλες εκτάσεις να μην εμφανίζεται θέση δειγματοληψίας , ενώ αντίθετα από ορισμένες άλλες να έχουν ληφθεί πολλά δείγματα.
- Περιμετρική, όταν η δειγματοληψία έχει γίνει περιμετρικά κατά μήκος μιας γραμμής π.χ. ισούψους.



Εικόνα 1 Χωρική κατανομή θέσεων

ΧΩΡΙΚΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗ

Μπορούμε να εξερευνήσουμε τη χωρική αυτόσυσχετιση των δεδομένων μας, εξετάζοντας τα διάφορα ζεύγη στις τοποθεσίες δείγματος. Μετρώντας την απόσταση ανάμεσα σε δύο θέσεις και την διαφορά των μεταξύ τους τιμών στις διάφορες τοποθεσίες, και υποτυπώνοντας τη διαφορά τετραγώνου μεταξύ των τιμών στις τοποθεσίες δημιουργείται ένα σύννεφο ημιβαριογράμματος (Semivariogram/Covariance Cloud). Στον άξονα x βρίσκεται η απόσταση μεταξύ των θέσεων και στον άξονα y είναι η διαφορά των τετραγώνων των τιμών τους. Κάθε κουκκίδα στο ημιβαριογράφημα αντιπροσωπεύει ένα ζεύγος τοποθεσιών και όχι τις μεμονωμένες τοποθεσίες στον χάρτη.

Εάν υπάρχει χωρική συσχέτιση, τα ζεύγη σημείων που βρίσκονται κοντά (στην αριστερή άκρη του άξονα x) θα πρέπει να έχουν λιγότερη διαφορά (να είναι χαμηλή στον άξονα y). Καθώς τα σημεία γίνονται μακρύτερα το ένα από το άλλο (κινούνται δεξιά στον άξονα x), γενικά, η διαφορά τετραγώνου θα πρέπει να είναι μεγαλύτερη (κινείται πάνω στον άξονα y). Συχνά υπάρχει μια ορισμένη απόσταση πέρα από την οποία τα τετράγωνα διαφέρουν επίπεδα. Ζεύγη τοποθεσιών πέρα από αυτήν την απόσταση θεωρείται ότι δεν σχετίζονται.

Μια θεμελιώδης υπόθεση για τις γεωστατιστικές μεθόδους είναι ότι οποιοσδήποτε δύο τοποθεσίες που έχουν παρόμοια απόσταση και κατεύθυνση μεταξύ τους θα πρέπει να έχουν μια παρόμοια διαφορά τετραγωνική. Αυτή η σχέση ονομάζεται στασιμότητα.

Η χωρική αυτοσυσχέτιση μπορεί να εξαρτάται μόνο από την απόσταση ανάμεσα σε δύο σημεία, που ονομάζεται **ισοτροπία**. Ωστόσο, είναι πιθανό η ίδια τιμή αυτόματου συσχετισμού μπορεί να εμφανιστεί σε διαφορετικές αποστάσεις όταν εξετάζουμε διαφορετικές κατευθύνσεις. Ένας άλλος τρόπος για να σκεφτούμε ότι τα

πράγματα είναι περισσότερο όμοια για μεγάλες αποστάσεις σε ορισμένες κατευθύνσεις από ότι σε άλλες κατευθύνσεις. Η κατευθυντική επιρροή φαίνεται στα βαριογράμματα όπου ονομάζεται **ανισοτροπία**.

Η χωρική αυτοσυσχέτιση έχει εφαρμογές σε μια ποικιλία προβλημάτων όπως οι κοινωνικές επιστήμες. Κοινό χαρακτηριστικό πρόβλημα της χωρικής στατιστικής είναι ότι τα υπό μελέτη δεδομένα απαρτίζονται από μεταβλητές, οι παρατηρήσεις των οποίων αναπαριστώνται στο χώρο. Αυτές λοιπόν οι παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους, όπως επιτάσσει η κλασική στατιστική (Κανάρογλου, 2001). Η χωρική αυτοσυσχέτιση, εστιάζει το ενδιαφέρον της στον βαθμό που η τιμή μιας μεταβλητής σε κάποια θέση είναι παρόμοια με τις τιμές της ίδιας μεταβλητής, που βρίσκονται γεωγραφικά κοντά της (Goodchild, 1987). Δηλαδή, η συσχέτιση μας δείχνει αν υπάρχει εξάρτηση μεταξύ των τιμών μιας συγκεκριμένης μεταβλητής σε εγγύς περιοχές. Όπως αναφέρει άλλωστε και ο καθηγητής Waldo Tobler (1970) στον πρώτο νόμο της γεωγραφίας, τα πάντα σχετίζονται μεταξύ τους από ότι τα αποκρουσμένα (Tobler, 1970). Οι δείκτες που χρησιμοποιεί η χωρική στατιστική αφορούν σε δείκτες που εξετάζουν είτε τη χωρική διασπορά της θέσης των σημείων είτε τη χωρική διασπορά των τιμών χαρακτηριστικών της κατανομής. Οι συνήθεις τεχνικές που χρησιμοποιούνται στην πρώτη περίπτωση είναι η μέθοδος ανάλυσης καννάβου και της απόστασης μεταξύ σημείων, στη δεύτερη οι δείκτες χωρικής αυτοσυσχέτισης Moran's I και G - Statistic.

Παράλληλα, γίνεται διάκριση:

- i. Σε γενικευμένους δείκτες (Global Statistics)
- ii. Σε τοπικούς δείκτες (Local Statistics).

Οι μεν πρώτοι εστιάζουν στην αναζήτηση σχηματισμού κάποιου προτύπου σε όλη την περιοχή μελέτης, οι δε δεύτεροι ασχολούνται με τις τοπικές διακυμάνσεις των τιμών. Όσον αφορά στη χωρική στατιστική ανάλυση δημογραφικών στοιχείων, αυτά επιπίπτουν στο πεδίο των ασυνεχών κατανομών επιφανειών, καθώς αντιστοιχεί μία τιμή σε κάθε χωρική ενότητα, και αυτό που ενδιαφέρει είναι η ανάλυση της χωρικής διασποράς των τιμών των πληθυσμιακών μεταβλητών και η εξέταση πιθανού προτύπου τους. Κατάλληλος δείκτης χωρικής στατιστικής ανάλυσης είναι ο Moran's I τόσο στη γενικευμένη όσο και στην τοπική του εκδοχή.

Ο Global Moran's I αποτελεί ένα συνηθισμένο δείκτη χωρικής αυτοσυσχέτισης για ποσοτικά δεδομένα, ο οποίος χρησιμοποιεί το μέγεθος των τιμών των χαρακτηριστικών των χωρικών στοιχείων για να ταυτοποιήσει την ύπαρξη χωρικού προτύπου. Η χωρική αυτοσυσχέτιση προκύπτει ως έννοια από τη χωρική επίδραση της χωρικής εξάρτησης και βασίζεται στο χαρακτηριστικό των χωρικών στοιχείων να παρουσιάζουν όμοιες τιμές τα κοντινά στοιχεία λόγω πιθανών όμοιων συνθηκών. Έτσι, υποδεικνύεται κατά πόσο η χωρική κατανομή των τιμών των χαρακτηριστικών εξαρτάται από τη χωρική κατανομή των θέσεων στις οποίες αναφέρονται οι τιμές, δηλαδή κατά πόσο οι συγκεκριμένες τιμές του υπό εξέταση φαινομένου εξαρτώνται από τη θέση που κατέχει η περιοχή στην οποία εμφανίζονται ή είναι ανεξάρτητες της θέσης αυτής. Η ύπαρξη χωρικού προτύπου είτε ομαδοποιημένου είτε διασκορπισμένου αντιστοιχεί σε θετική ή αρνητική χωρική αυτοσυσχέτιση αντίστοιχα. Ομαδοποιημένο χωρικό πρότυπο και συνεπώς θετική χωρική αυτοσυσχέτιση παρουσιάζεται στην περίπτωση που τα κοντινά στοιχεία είναι περισσότερο όμοια από τα πιο απομακρυσμένα, ενώ διασκορπισμένο πρότυπο και αρνητική χωρική αυτοσυσχέτιση παρουσιάζεται αν τα γειτονικά χωρικά στοιχεία τείνουν να είναι ανομοιογενή.

Ο δείκτης Global Moran's I λειτουργεί με βάση την ομοιότητα ή όχι στις τιμές γειτονικών στοιχείων. Στην ουσία υπολογίζεται η διαφορά του μέσου όρου όλων των τιμών από την τιμή ενός συγκεκριμένου χωρικού στοιχείου καθώς

και η διαφορά του μέσου όρου από την τιμή κάθε γείτονα του εξεταζόμενου χωρικού στοιχείου και συγκρίνονται οι διαφορές μεταξύ τους, μία διαδικασία που γίνεται για όλα τα χωρικά στοιχεία.

Συγκεκριμένα, ο τύπος με βάση τον οποίο υπολογίζεται είναι ο ακόλουθος :

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^{n-1} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^{n-1} w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

Από τη σχέση φαίνεται ότι ο δείκτης Global Moran's I υπολογίζει τη συνδιασπορά των τιμών των στοιχείων ανά ζευγάρια αποτελούμενα από τιμές ενός στοιχείου κάθε φορά (x_i) και τιμές των γειτονικών του στοιχείων (x_j), μετά πολλαπλασιάζει το κάθε αποτέλεσμα με το αντίστοιχο βάρος (w_{ij}), συναθροίζει όλες τις συνδιασπορές και διαιρεί το τελικό αποτέλεσμα με τη διασπορά όλων των χωρικών στοιχείων n της υπό εξέταση περιοχής, η οποία έχει πολλαπλασιαστεί με το άθροισμα όλων των βαρών.

Ο δείκτης Moran's I μεταβάλλεται από -1 έως +1. Παρ' όλα αυτά μπορεί να παρατηρηθούν τιμές μεγαλύτερες από το +1 και μικρότερες από το -1. Η τιμή 0 δείχνει απουσία αυτοσυσχέτισης. Θετική αυτοσυσχέτιση σημαίνει ότι χωρικά γειτονικές τιμές μιας μεταβλητής είναι όμοιες ενώ αρνητική αυτοσυσχέτιση σημαίνει ότι γειτονικές τιμές είναι ανόμοιες. Επίσης, χρησιμοποιείται σε διανυσματικά και σε πλεγματικά δεδομένα. Η γειτνίαση είναι μια παράξενη ιδιότητα του πλέγματος αφού κάθε κυψελίδα έχει κοινό όριο με τέσσερις γειτονικές. Κατάλληλες τεχνικές εφαρμόζονται για τον υπολογισμό του δείκτη σε πλεγματικά δεδομένα (Shortridge, 2007).

Η χρήση του τοπικού δείκτη αυτοσυσχέτισης (Local Moran I) βοηθάει στον εντοπισμό οντοτήτων που φέρουν τιμές διαφορετικές από τον περίγυρο τους και μπορούν να αποτελέσουν περιοχές δυναμικές ή αδύναμες για την εκάστοτε μεταβλητή (Anselin, 2010). Η εφαρμογή αυτή για τις επιλεγμένες μεταβλητές εντοπίζει κάθε φορά ένα περιορισμένο αριθμό οντοτήτων που είναι στατιστικά σημαντικές. Το άθροισμα όλων των I_i για όλες τις παρατηρήσεις είναι ανάλογο του ολικού Moran's I. Ο τοπικός δείκτης Moran's δίνεται από την παρακάτω εξίσωση :

$$L'Moran'I = (X_i - \bar{X}) \frac{\sum_j (X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2 / n} \quad (7)$$

όπου x_i και x_j σημαίνουν παρατηρούμενη τιμή από την περιοχή i και j , $\sum_i (x_i - \bar{x})$ και $\sum_j (x_j - \bar{x})$ είναι η απόκλιση της μεταβλητής X σε διάφορες περιοχές σε σχέση με το μέσο όρο. Ο \bar{x} είναι η μέση τιμή του x και n είναι ο αριθμός των παρατηρήσεων.

Τα αποτελέσματα του δείκτη τοπικού Moran μπορούν να τυποποιηθούν, έτσι ώστε το επίπεδο σημαντικότητας του να μπορεί να ελεγχθεί με βάση την παραδοχή της κανονικής κατανομής (Levine 2004). Όταν χρησιμοποιούμε τον τοπικό δείκτη Moran για να αναλύσουμε το χωρικό σχέδιο, τα αποτελέσματα επηρεάζονται από τον ορισμό της συνάρτησης βάρους, τα δεδομένα μετασχηματισμού, και την ύπαρξη ακραίων τιμών (Zhang et al. 2008). Αυτοί οι παράγοντες λαμβάνονται υπόψη προκειμένου να αποκτήθούν αξιόπιστα και σταθερά αποτελέσματα.

Μια υψηλή θετική τιμή του τοπικού Moran συνεπάγεται η τιμή να είναι παρόμοια με τη γειτονιά της, και στη συνέχεια οι θέσεις να είναι χωρικές ομάδες, που περιλαμβάνουν μεγάλες-μεγάλες ομάδες (υψηλές τιμές σε μια γειτονιά μεγάλων τιμών) hot spots και μικρές-μικρές ομάδες (χαμηλές τιμές σε μια γειτονιά χαμηλών τιμών) cold spots (Lalor and Zhang, 2001. Wulder et al., 2007). Εν τω μεταξύ, μια υψηλά αρνητική τιμή του τοπικού Moran συνεπάγεται μια πιθανή χωρική ακραία τιμή, η οποία είναι προφανώς διαφορετική από τις τιμές γύρω των περιοχών. Χωρικά ακραίες τιμές περιλαμβάνουν high-low (μεγάλες - μικρές) και low-high (μικρές-μεγάλες) ακραίες τιμές (Lalor and Zhang, 2001).

2.1.2.β. ΧΩΡΙΚΗ ΠΑΡΕΜΒΟΛΗ

Χωρική παρεμβολή (interpolation) είναι η διαδικασία του υπολογισμού της τιμής μιας παραμέτρου σε θέσεις που δεν υπάρχουν μετρήσεις χρησιμοποιώντας μετρήσεις που έχουν γίνει σε σημεία μέσα στην ίδια την περιοχή. Η πρόβλεψη των τιμών μιας ιδιότητας σε θέσεις έξω από την περιοχή των παρατηρήσεων καλείται extrapolation. Οι μέθοδοι χωρικής παρεμβολής διαιρούνται σε δύο ομάδες, τις γενικές και τις τοπικές (Burrough and McDonnell, 1999). Οι γενικές μέθοδοι παρεμβολής χρησιμοποιούν όλα τα διαθέσιμα δεδομένα για να κάνουν εκτιμήσεις σε ολόκληρη την εξεταζόμενη περιοχή, ενώ οι τοπικές μέθοδοι εκτιμούν την τιμή μιας μεταβλητής σε ένα σημείο με βάση μια μικρή περιοχή γύρω από αυτό, ώστε οι υπολογισμοί να γίνονται, με δεδομένα από αυτή τη γειτονική περιοχή του σημείου και να είναι όσο το δυνατόν ακριβείς. Δηλαδή, οι τοπικές μέθοδοι χωρικής παρεμβολής βασίζονται στον πρώτο νόμο της γεωγραφίας (Tobler, 1970) και ο υπολογισμός των τιμών σε άγνωστα σημεία γίνεται από έναν αριθμό γειτονικών τους σημείων.

Οι γενικές μέθοδοι χρησιμοποιούνται κυρίως όχι για άμεση εκτίμηση, αλλά για την εξέταση και πιθανώς την απομάκρυνση των επιδράσεων των γενικών μεταβολών που προέρχονται από κάποιες τάσεις. Όταν οι γενικές επιδράσεις προσδιοριστούν και απομακρυνθούν, τότε οι υπόλοιπες μεταβολές μπορεί να εκτιμηθούν τοπικά.

Μια άλλη μέθοδος διάκρισης των μεθόδων χωρικής παρεμβολής είναι ο τρόπος χειρισμού των γνωστών μετρημένων τιμών στις αρχικές θέσεις. Η μέθοδος χωρικής παρεμβολής που εκτιμά, δηλαδή διατηρεί, σε ένα γνωστό σημείο ίδια τιμή με την μετρημένη τιμή της μεταβλητής στο σημείο αυτό καλείται ακριβής παρεμβολέας (exact interpolator). Όλες οι άλλες μέθοδοι είναι μη ακριβείς παρεμβολείς (inexact interpolator).

Τα αποτελέσματα των μεθόδων χωρικής παρεμβολής εκτιμώνται βάσει των τιμών ορισμένων στατιστικών μεγεθών – δεικτών, που υπολογίζονται από τις διαφορές μεταξύ των μετρημένων τιμών στα γνωστά σημεία από τις εκτιμηθείσες τιμές από τη χωρική παρεμβολή (Burrough and McDonnell, 1999). Ο υπολογισμός των διαφορών μπορεί να γίνει εάν αντί για τα αρχικά σημεία που χρησιμοποιήθηκαν για την χωρική παρεμβολή χρησιμοποιηθεί μια άλλη ομάδα σημείων στα οποία έχουν γίνει μετρήσεις.

ΜΕΘΟΔΟΙ ΤΟΠΙΚΩΝ ΕΚΤΙΜΗΣΕΩΝ

Σύμφωνα με τον νόμο του Tobler (1970), η τιμή σε ένα σημείο που δεν υπάρχει μέτρηση είναι παρόμοια με αυτές των κοντινών του σημείων. Επομένως υπάρχουν οι τοπικές μέθοδοι χωρικής εκτίμησης που χρησιμοποιούν πληροφορίες από τα κοντινά σημεία. Για την προσέγγιση οι τοπικές μέθοδοι χωρικής παρεμβολής περιλαμβάνουν (Burrough and McDonnell, 1998) :

- i. Τον ορισμό μίας περιοχής ή γειτονίας γύρω από το σημείο που πρέπει να εκτιμηθεί
- ii. Την εύρεση σημείων μέσα στη γειτονία
- iii. Την επιλογή μίας μαθηματικής συνάρτησης για την παράσταση της μεταβολής μεταξύ αυτού του μικρού αριθμού των σημείων
- iv. Την επαναληπτική εφαρμογή των παραπάνω βημάτων για να υπολογισθεί η τιμή της συνάρτησης σε κάθε σημείο ενός κανονικού πλέγματος.

Παραδείγματα μεθόδων τοπικών εκτιμήσεων είναι τα πολύγωνα Thiessen (Nearest neighbors), η Inverse Distance Weighting (IDW) και οι Splines.

Παρεμβολή Inverse Distance Weighting

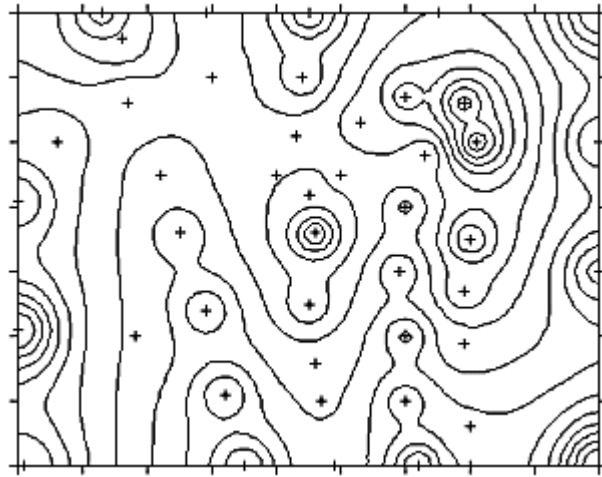
Στην απλούστερη περίπτωση, μπορούμε να προχωρήσουμε με τις μεθόδους βάση της απόστασης με τον ίδιο τρόπο όπως και με τη μέθοδο της «moving windows»: ορίζουμε μια ορισμένη γειτονία των γνωστών σημειακών δεδομένων γύρω από μια άγνωστη θέση για να εκτιμηθεί κάθε φορά ο αριθμητικός μέσος. Οι αριθμητικοί μέσοι των γνωστών τιμών είναι κινητός μέσος όρος. Η γειτονία μπορεί να οριστεί με διάφορους τρόπους: ένα χωρικό διορθωμένο σχήμα (ορθογώνιο, κύκλος, κλπ.) και ένας ορισμένος αριθμός των πλησιέστερων γειτονικών σημείων. Ωστόσο, αυτή η μέθοδος είναι αρκετά ασαφής λόγω των διαφορετικών αποστάσεων μεταξύ της θέσης για να εκτιμηθεί και η κακή ενσωμάτωση των γνωστών σημείων στην παρεμβολή. Οι πραγματικοί μέθοδοι με βάση την απόσταση χρησιμοποιούν ακριβώς αυτές τις αποστάσεις μεταξύ των σημείων εκτίμησης και των γνωστών σημείων μέτρησης για να υπολογίσουν την επιρροή τους στον υπολογισμό της εκτιμώμενης τιμής. Με την ευκαιρία, απαιτούν μία γραμμική χωρική συσχέτιση μεταξύ των φαινομένων.

Χρησιμοποιώντας την αποκαλούμενη μέθοδο "Inverse Distance Weighting" ή IDW, η τιμή οποιουδήποτε γνωστού σημείο έχει οριστεί αντιστρόφως ανάλογα με την απόστασή του από το εκτιμώμενο σημείο. Υπολογίζεται ως εξής:

$$\hat{v}_1 = \frac{\sum_{i=1}^n \frac{1}{d^p_i} v_i}{\sum_{i=1}^n \frac{1}{d^p_i}} \quad (8)$$

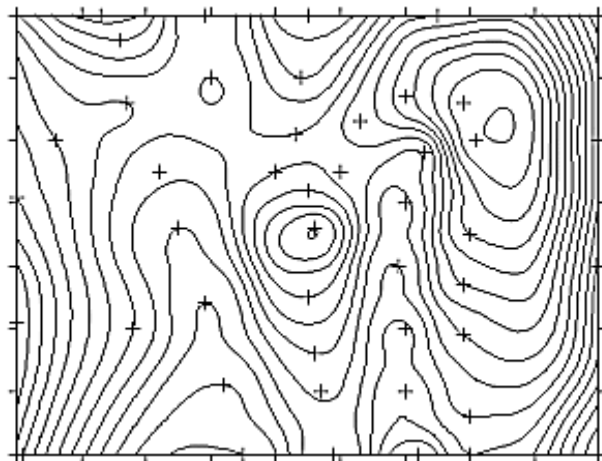
Όπου v η εκτιμώμενη τιμή, v_i γνωστή τιμή και $d^p_1 \dots d^p_n$ = αποστάσεις από τα σημεία n δεδομένα εις την p στην εκτίμηση σημείου n .

Πλεονεκτήματα της παρεμβολής IDW: 1) Παρέχει τη δυνατότητα για πολύ γρήγορους υπολογισμούς. 2) Διαφορετικές αποστάσεις ενσωματώνονται στην εκτίμηση. 3) Ο εκθέτης σταθμισμένη απόσταση είναι σε θέση να ελέγχει με ακρίβεια την επίδραση των αποστάσεων. Μειονεκτήματα της παρεμβολής IDW: 1) Δεν είναι δυνατόν να κάνουμε μια κατεύθυνση που εξαρτάται από την στάθμιση. Αυτό σημαίνει ότι οι σχέσεις των χωρικών προσανατολισμών αγνοούνται (π.χ. υψομετρικά σημεία κατά μήκος μιας κορυφογραμμής). 2) Κρυμμένα αντικείμενα τα λεγόμενα "Bulls-eyes" - αυτά είναι κυκλικές περιοχές ίσων τιμών γύρω από τα γνωστά σημεία δεδομένων. Ωστόσο, εφαρμόζοντας μια παραλλαγή της IDW-Interpolation που αναπτύχθηκε από τον Shepard (1968) μπορεί να μειώσει τα Bulls-eyes:



Εικόνα 2 IDW Bulls-eyes

IDW "Bulls-eye" αποτέλεσμα: ομόκεντρων περιοχών της ίδιας αξίας γύρω από τα γνωστά σημεία - ένα ανεπιθύμητο αντικείμενο της παρεμβολής IDW.



Εικόνα 3 IDW τροποποιημένη μετά τον Shepard: οι Bulls-eyes έχουν σίγουρα μειωθεί.

ΓΕΩΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΑΡΕΜΒΟΛΗΣ

Ένα από τα μειονεκτήματα της παρεμβολής IDW είναι η έλλειψη συγκεκριμένων κατευθύνσεων (ανισότροπα) πληροφοριών. Ως εκ τούτου, οι χωρικές συσχετίσεις αγνοούνται και δεν έχουν ενσωματωθεί στο αποτέλεσμα της εκτίμησης. Αυτό το μειονέκτημα εξαλείφεται από γεωστατιστικές παρεμβολές. Το όνομα "γεω" επισημαίνει ήδη το πιο σημαντικό χαρακτηριστικό αυτών των μεθόδων: χωρικών-στατιστικών παραμέτρων που αποτελούν την κύρια βάση για αυτές τις μεθόδους παρεμβολής.

Το βαριόγραμμα ή βαριογραφία (η μέθοδος για να αντλήσει από τα δεδομένα χωρικά σημεία) είναι η βάση για μια επιτυχημένη γεωστατιστική παρεμβολή. Οι γεωστατιστικές παρεμβολές έχουν προχωρήσει και σε κάποιο βαθμό έχουν περιπλέξει τις μεθόδους. Η λογική εφαρμογή τους απαιτεί ένα μεγάλο ποσό γνώσης και εμπειρίας.

Οι κύριες διαδικασίες είναι οι μέθοδοι Kriging. Ονομάστηκαν έτσι μετά από ένα Αφρικανό μηχανικό, D. Krige. Στην διπλωματική του εργασία το 1951, έθεσε τα θεμέλια για kriging. Ωστόσο, οι κυριότερες εξελίξεις προέρχονται από το έργο του G. Matheron στη δεκαετία του 1960.

Χρησιμοποιώντας τη βαριογραφία, παίρνουμε ενδείξεις πόσο παρόμοιες ή ανόμοιες είναι οι μετρήσεις των γειτονικών σημείων ως συνάρτηση της απόστασής τους το ένα από το άλλο.

Για μια χωρική στοχαστική διαδικασία $\{Z(s), s \in A\}$, όπου $E(Z(s)) = \mu(s)$ και $\text{var}(Z(s)) = \sigma^2(s)$, η συνδιασπορά αυτής της διαδικασίας σε δύο σημεία s_i και s_j ορίζεται ως:

$$C(s_i, s_j) = E[(Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j))] \quad (9) \quad \text{ενώ η αντίστοιχη συσχέτιση ορίζεται ως:}$$

$$\rho(s_i, s_j) = \frac{C(s_i, s_j)}{\sigma(s_i)\sigma(s_j)} \quad (10) \quad \text{βέβαια η } C(s_i, s_j) = \sigma^2(s) \quad (11)$$

Η παραπάνω διαδικασία θεωρείται στάσιμη αν $\mu(s) = \mu$ και $\sigma^2(s) = \sigma^2$ και επιπλέον:

$$C(s_i, s_j) = C(s_i - s_j) = C(h) \quad (12) \quad \text{όπου βέβαια: } C(h=0) = \sigma^2$$

Η $C(h)$ αναφέρεται ως συνάρτηση συνδιασποράς ή συνβαριόγραμμα της διαδικασίας.

Η $\rho(h)$ η αντίστοιχη συσχέτιση αναφέρεται ως συσχετόγραμμα. Η παραπάνω εξίσωση δείχνει ότι η $C(s_i, s_j)$ εξαρτάται μόνο από τη διανυσματική διαφορά h μεταξύ s_i και s_j και όχι από την απόλυτη θέση τους, δηλαδή η διαδικασία είναι ισοτροπική. Στην περίπτωση αυτή: $C(s_i, s_j) = C(h)$ και $\rho(s_i, s_j) = \rho(h)$

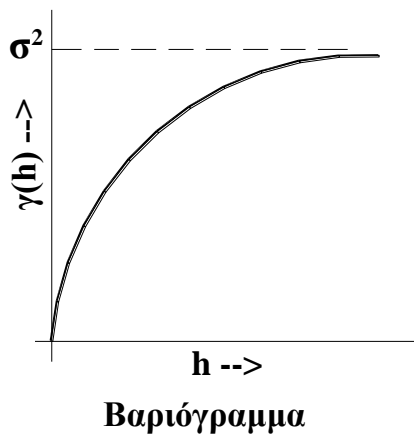
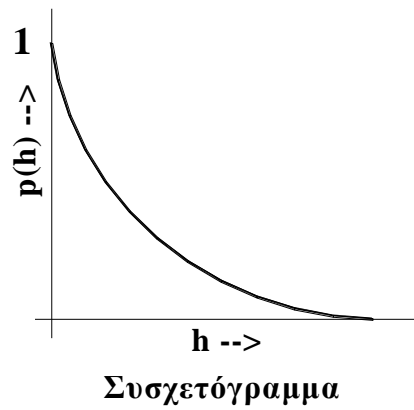
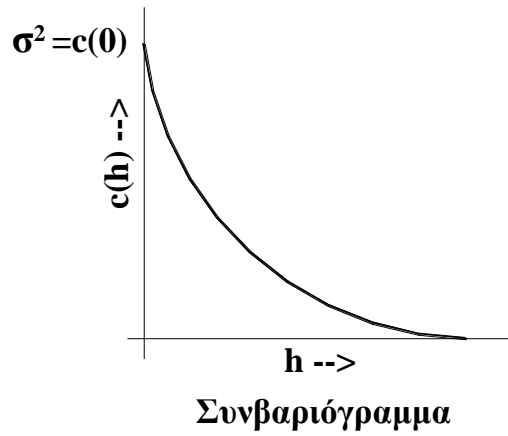
Αν υποθεθεί ότι η μέση τιμή και η διασπορά των διαφορών μεταξύ των τιμών σε δύο σημεία, που η θέση τους απέχει μια συγκεκριμένη απόσταση και διεύθυνση, είναι σταθερές (εσωτερική στασιμότητα), τότε:

$$E(Z(s+h) - Z(s)) = 0 \quad (13)$$

$$\text{VAR}(Z(s+h) - Z(s)) = 2\gamma h \quad (14)$$

Η $\gamma(h)$ αναφέρεται ως βαριόγραμμα. Το βαριόγραμμα είναι το πρώτο βήμα για την ποσοτική περιγραφή της χωρικής μεταβολής. Πρακτικά υπολογίζεται η ημιδιακύμανση για αποστάσεις $h, 2h, 3h, 4h, \dots$, όπου h είναι το χωρικό βήμα απόστασης που στη διεθνή βιβλιογραφία ονομάζεται lag. Η απόσταση h προσδιορίζεται από τη μέση απόσταση δειγματοληψίας, εάν πρόκειται για τυχαία, ή από την απόσταση δειγματοληψίας στην περίπτωση της κανονικής δειγματοληψίας. Σε μια στάσιμη χωρική διαδικασία το συνβαριόγραμμα, το συσχετόγραμμα και το βαριόγραμμα σχετίζονται άμεσα ως εξής:

$$\rho(h) = \frac{C(h)}{\sigma^2} \quad \gamma(h) = \sigma^2 - C(h)$$



Η θεωρία των περιφερειοποιημένων μεταβλητών υποθέτει ότι η χωρική διαφοροποίηση μιας οποιασδήποτε μεταβλητής μπορεί να θεωρηθεί ως το άθροισμα των εξής τριών βασικών συνισταμένων:

- Της δομικής συνισταμένης, η οποία έχει μια σταθερή μέση τιμή ή «τάση».
- Μιας τυχαίας αλλά χωρικά συσχετισμένης συνισταμένης, γνωστής και ως περιφερειοποιημένη μεταβλητή, και
- Ενός μη χωρικά συσχετισμένου τυχαίου θορύβου ή ενός υπόλοιπου τυχαίου σφάλματος.

Η τιμή της μεταβλητής Z σε ένα σημείο s_i δίνεται από τη σχέση: $Z(s_i) = m(s_i) + U(s_i) + \varepsilon$ ⁽¹⁵⁾

Όπου :

$m(s_i)$ = μια αιτιοκρατική συνάρτηση που περιγράφει τη δομική συνισταμένη της Z στο σημείο s_i .

$U(s_i)$ = ένας στοχαστικός όρος που εκφράζει τα υπόλοιπα από την $m(s_i)$ που μεταβάλλονται τοπικά αλλά είναι χωρικά συσχετισμένα, δηλαδή την περιφερειοποιημένη μεταβλητή, και

ε = τα χωρικά ανεξάρτητα σφάλματα που ακολουθούν μια κανονική κατανομή με μέση τιμή μηδέν και διασπορά σ^2 .

Αν υποθεθεί ότι η προηγούμενη εξίσωση εκφράζει μια εσωτερική στάσιμη χωρική διαδικασία. Αν οι δομικές επιπτώσεις (π.χ. η μέση τιμή στην περιοχή μελέτης) μπορούν να εξηγηθούν η εναπομείνουσα διασπορά είναι ομοιογενής ως προς τη διαφοροποίησή της. Οι διαφορές μεταξύ σημείων του δείγματος είναι απλώς μια συνάρτηση της απόστασης μεταξύ τους. Η εξίσωση μπορεί να πάρει τη μορφή: $Z(s_i) = m(s_i) + \gamma(h) + \varepsilon$ για την κατανομή της οποίας αρκεί η εκτίμηση της $\gamma(h)$. Το βαριόγραμμα μπορεί να εκτιμηθεί από τα στοιχεία ενός δείγματος ως εξής:

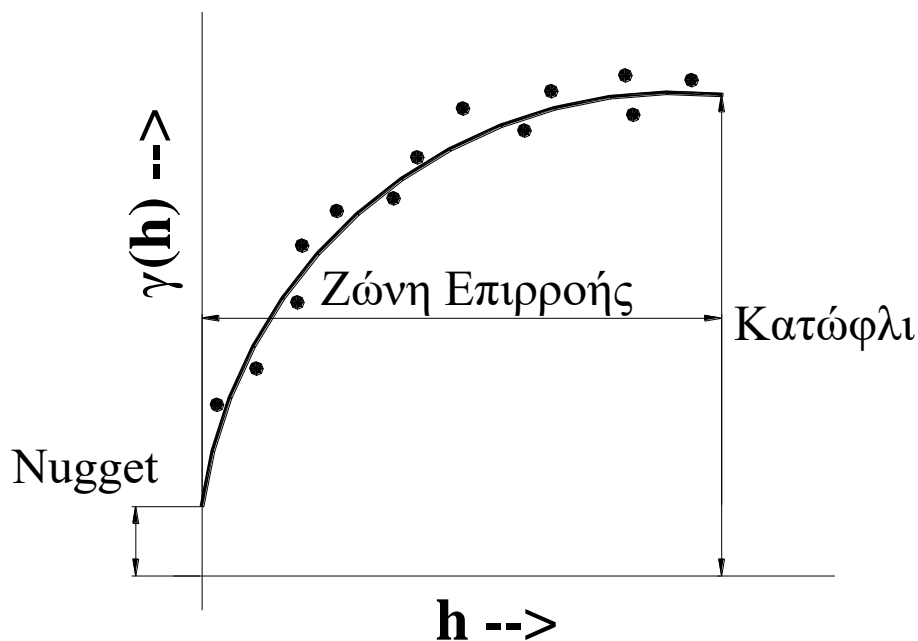
$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{s_i - s_j = h}^n (Z(s_i) - Z(s_j))^2 \quad (16)$$

$\sum_{s_i - s_j = h}^n$ = το άθροισμα για όλα τα ζευγάρια των παρατηρούμενων σημείων που απέχουν μεταξύ τους κατά το

διάστημα h .

$n(h)$ = ο αριθμός αυτών των ζευγαριών.

Καθώς το h μεταβάλλεται, μια σειρά από τιμές εκτιμώνται, δημιουργώντας το δειγματικό ή πειραματικό βαριόγραμμα.



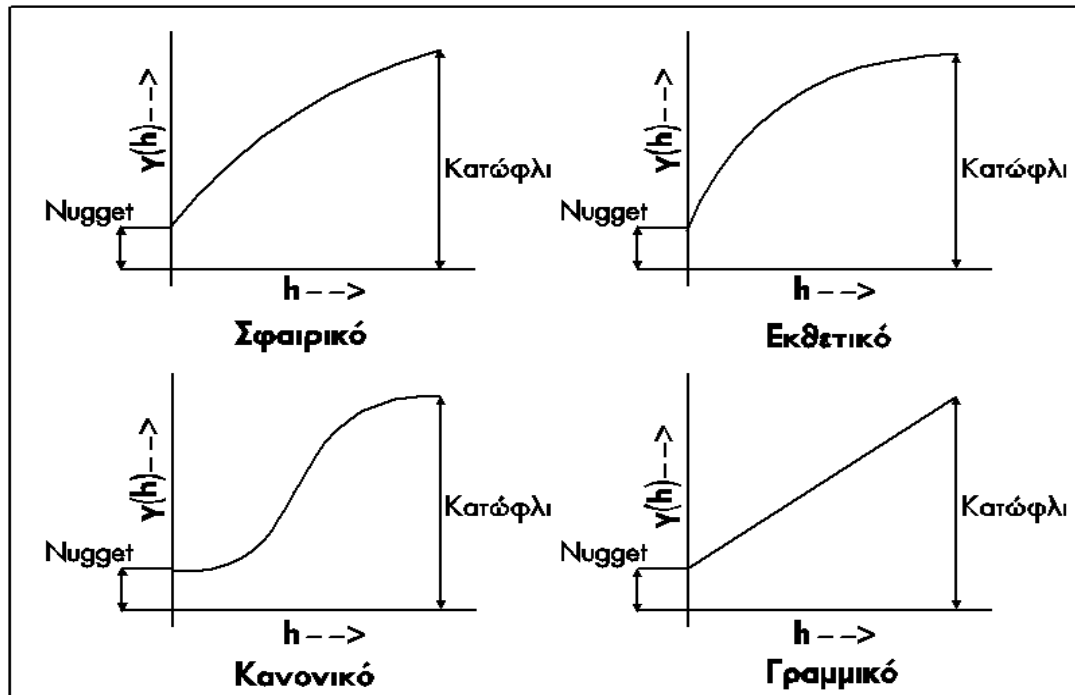
Σχήμα 1

Εξετάζοντας το σχήμα 1 παρατηρούνται τα εξής: Καθώς οι τιμές του h αυξάνουν, η τιμή της $\gamma(h)$ αυξάνεται ασύμπτωτα προς ένα ανώτερο όριο που ονομάζεται κατώφλι (sill). Η καμπύλη της $\gamma(h)$ αυξάνεται μέχρι να φτάσει στο κατώφλι, οπότε λαμβάνει τη μέγιστη τιμή της σε μια απόσταση που ορίζει τη ζώνη επιρροής (range), η οποία με

τη σειρά της καθορίζει το χώρο εντός του οποίου οι μεταξύ των σημείων διαφοροποιήσεις είναι χωρικά εξαρτημένες. Η εκτιμηθείσα καμπύλη δεν περνά από την αρχή των αξόνων αλλά τέμνει τον κάθετο άξονα σε ένα σημείο με θετική τιμή. Επομένως, θετική τιμή $h \rightarrow 0$ είναι μια εκτίμηση του ϵ , δηλαδή του μη χωρικού θορύβου ή υπόλοιπου, που είναι γνωστό ως nugget (C_0). Στην περίπτωση που παρατηρείται ένα βαριόγραμμα όπου οι τιμές των διασπορών είναι ευρέως διασκορπισμένες, αυτό δηλώνει καθαρά ότι η εκτίμηση της έγινε με τη χρήση ενός μικρού δείγματος.

Η μορφή του βαριογράμματος είναι ιδιαίτερα κατατοπιστική για το είδος της χωρικής διαφοροποίησης που υπάρχει στην περιοχή μελέτης και για να βοηθήσει σημαντικά στη διαδικασία χωρικής παρεμβολής.

Τέσσερα είναι τα βασικά μοντέλα (μορφές) του βαριογράμματος



Σχήμα 2

Σφαιρικό Μοντέλο

$$\gamma(h) = \alpha + (\sigma^2 - \alpha) \left(\frac{3h}{2r} - \frac{h^3}{2r^3} \right) \text{ για } 0 < h < r$$

$$\gamma(h) = 0 \text{ για } h = 0$$

$$\gamma(h) = \sigma^2 \text{ για } h \geq r$$

Όπου: r = η ζώνη επιρροής (range), σ^2 = το κατώφλι-διασπορά, h = η απόσταση μεταξύ των σημείων και α = το τμήμα της διασποράς που αφορά το nugget

Εκθετικό Μοντέλο

$$\gamma(h) = \alpha + (\sigma^2 - \alpha) \left(1 - e^{-\frac{3h}{r}} \right) \text{ για } h > 0$$

$$\gamma(h) = 0 \text{ για } h = 0$$

Κανονικό ή Μοντέλο Gaussian

$$\gamma(h) = \alpha + (\sigma^2 - \alpha) \left(1 - e^{-3h^2/r^2} \right) \text{ για } h > 0$$

$$\gamma(h) = 0 \text{ για } h = 0$$

Γραμμικό Μοντέλο

$$\gamma(h) = \sigma^2 + \beta h$$

όπου β η κλίση της γραμμής του βαριογράμματος.

ΜΕΘΟΔΟΣ ΓΕΩΣΤΑΤΙΣΤΙΚΗΣ ΠΑΡΕΜΒΟΛΗΣ-KRIGING

Ordinary Kriging

Η μέθοδος Kriging είναι η πιο διαδεδομένη γεωστατιστική μέθοδος που εφαρμόζεται στα περισσότερα λογισμικά πακέτα GIS. Με τον όρο γεωστατιστική ορίζεται ένα σύνολο στατιστικών τεχνικών που σχετίζονται με μεταβλητές που μεταβάλλονται στο χώρο. Οι τεχνικές αυτές βασίζονται στην υπόθεση ότι η χωρική διακύμανση της μεταβλητής είναι τυχαία. Γενικά οι γεωστατιστικές μέθοδοι είναι αρκετά πολύπλοκες στην εφαρμογή τους, η οποία προϋποθέτει τη χρήση κατάλληλων υπολογιστικών προγραμμάτων. Οι γεωστατιστικές μέθοδοι σε σχέση με τις υπόλοιπες μεθόδους που περιγράφηκαν παραπάνω έχουν τα εξής πλεονεκτήματα :

- Έχουν τη δυνατότητα να δώσουν άμεσες εκτιμήσεις για την ποιότητα των προβλέψεων, δηλαδή μία εκτίμηση της διασποράς των προβλεπόμενων τιμών στα σημεία που δεν ανήκουν στο δείγμα.
- Εκτιμούν εκ των προτέρων τις βέλτιστες τιμές για τις διάφορες παραμέτρους που χρησιμοποιούν οι μέθοδοι.
- Παρέχουν τη δυνατότητα επιλογής του αριθμού των σημείων των αναγκαίων για την εφαρμογή των μεθόδων πλησιέστερου γείτονα.
- Γνωστοποιούν τα λάθη – αβεβαιότητες που σχετίζονται με τις εκτιμηθείσες τιμές της χωρικής παρεμβολής.

Η γεωστατιστική ανάλυση περιλαμβάνει δύο κύριες φάσεις:

- την χωρική ανάλυση που περιλαμβάνει την επιλογή και προσαρμογή ενός μοντέλου που περιγράφει τη χωρική μεταβλητότητα των σημειακών μετρήσεων.
- τη βέλτιστη γραμμική αμερόληπτη εκτίμηση (best linear unbiased estimation – BLUE) που σχετίζεται με τον υπολογισμό των εκτιμήσεων των αγνώστων ως γραμμικών συναρτήσεων των μετρήσεων. Οι εκτιμήσεις είναι αμερόληπτες, έχουν την ελάχιστη μεταβλητότητα, ενώ για τον υπολογισμό τους χρησιμοποιείται η μοντελοποίηση της χωρικής μεταβλητότητας.

Η τιμή μιας συνεχούς χωρικής μεταβλητής δίνεται από το άθροισμα δύο συνισταμένων, μιας πρώτου και μιας δεύτερου βαθμού. Η συνισταμένη πρώτου βαθμού είναι ο γενικευμένος μέσος όρος $\mu(s)$, ο οποίος είναι σταθερός και δεν χρειάζεται να εκτιμηθεί από τις παρατηρηθείσες τιμές. Αν η τιμή αυτή αφαιρεθεί από τις παρατηρούμενες τιμές σε κάθε σημείο του δείγματος, τα υπόλοιπα δίνουν τις τιμές $U(s)$ με γνωστή διακύμανση σ^2 και συνδιασπορά $C(\cdot)$, που εκφράζονται από το βαριόγραμμα.

Το πρόβλημα που επιλύεται με τη μέθοδο της συνήθους Kriging είναι η εκτίμηση:

$\hat{V}(s)$ = η τιμή της τυχαίας μεταβλητής $V(s)$ στη θέση s (με βάση τις παρατηρηθείσες τιμές $V(s_i)$ σε n θέσεις s_i του δείγματος).

$Z(s)$ = που εκτιμάται αυτόματα (αρκεί στην τιμή $V(s)$ για το σημείο s να προστεθεί η γνωστή μέση τιμή ή τάση $\mu(s)$).

Θεωρείται ότι οι εκτιμήσεις αποτελούν ένα γραμμικό, με βάρη συνδυασμό των παρατηρηθεισών τιμών των περιφερειοποιημένων μεταβλητών, που παίρνει τη μορφή:

$$\hat{V}(s_0) = \sum_{i=1}^n \lambda_i V(s_i) \quad (17)$$

Όπου: $\hat{V}(s_0)$ = Η εκτιμηθείσα τιμή στη θέση s_0 .

$V(s_i)$ = Η παρατηρηθείσα τιμή στο σημείο s_i .

λ_i = Τα βάρη που αντιστοιχούν σε κάθε σημείο του δείγματος s_i , δηλαδή εξαρτώνται από τη θέση τους σε σχέση με την υπό εκτίμηση θέση s_0 .

Τα βάρη λ_i επιλέγονται έτσι ώστε η εκτίμηση $\hat{V}(s_0)$ να τηρεί τον όρο της μη προκατάληψης ότι η εκτιμηθείσα διακύμανση είναι μικρότερη από κάθε άλλο γραμμικό συνδυασμό των παρατηρούμενων τιμών. Λόγω της απαίτησης για μη προκατάληψη και επειδή η αναμενόμενη τιμή είναι γραμμικός τελεστής, ισχύουν:

$$E[V(s_0) - \hat{V}(s_0)] = 0 \Rightarrow \mu_{s_0} - \sum_{i=1}^n \lambda_i \cdot \mu_{s_0} = 0$$

Επειδή η μέση τιμή είναι σταθερή, η απαίτηση για μη προκατάληψη σημαίνει ότι:

$$\sum_{i=1}^n \lambda_i = 1.0$$

Η απαίτηση για βέλτιστη πρόβλεψη οδηγεί στην ελαχιστοποίηση της ποσότητας Q ως εξής:

$$Q = E\left[\left(V(s_0) - \hat{V}(s_0)\right)^2\right] = E\left[\left(V(s_0) - \sum_{i=1}^n \lambda_i V(s_i)\right)^2\right]$$

Μετά από μετασχηματισμούς η εξίσωση γίνεται:

$$Q = E\left[\left(V(s_0) - \mu\right)^2\right] + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E[(V(s_i) - \mu)(V(s_j) - \mu)] - 2 \sum_{i=1}^n \lambda_i \cdot E[(V(s_i) - \mu)(V(s_0) - \mu)] \quad (18)$$

Από τους τρεις όρους:

Ο πρώτος είναι η διασπορά της τυχαίας συνάρτησης, δηλαδή C_0 . Ο μεσαίος όρος είναι η συνδιασπορά των μεταβλητών στις θέσεις s_i, s_j . Ο τρίτος όρος είναι η συνδιασπορά της μεταβλητής που είναι επιθυμητό να προβλεφθεί και της μεταβλητής στη θέση s_i . Επομένως:

$$Q = C_0 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_i \quad (19)$$

Με δεδομένο ότι $\sum_{i=1}^n \lambda_i = 1.0$ και με τη χρήση πολλαπλασιαστών Lagrange, δημιουργείται ένα σύστημα

εξισώσεων ως εξής:

$$\frac{\partial}{\partial \lambda_k} \left[Q - 2\phi \left(1 - \sum_{i=1}^n \lambda_i \right) \right] = 0 \quad \text{για } k=1,2,3,4,\dots,n$$

$$\sum_{i=1}^n \lambda_i = 1.0$$

Αν υλοποιηθεί η μερική παράγωγος για κάθε k , τότε:

$$2 \sum_{i=1}^n \lambda_i C_{ki} - 2C_k + 2\phi = 0 \quad \text{για } k=1,2,3,\dots,n$$

$$\sum_{i=1}^n \lambda_i = 1$$

Το σύστημα των n+1 εξισώσεων σε αναλυτική μορφή είναι:

$$\begin{bmatrix} C & 1 \\ 1^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda \\ \phi \end{bmatrix} = \begin{bmatrix} C \\ 1 \end{bmatrix}$$

όπου: 1 = διάνυσμα μήκους n που περιλαμβάνει μονάδες, C = πίνακας με συνδιασπορές ανάμεσα στα σημεία με γνωστές τιμές, και λ = το διάνυσμα των αγνώστων γραμμικών συντελεστών.

Η λύση του συστήματος είναι:

$$\begin{bmatrix} \lambda \\ \phi \end{bmatrix} = \begin{bmatrix} C & 1 \\ 1^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} C \\ 1 \end{bmatrix}$$

Για απλοποίηση η εξίσωση μπορεί να γραφεί και ως εξής:

$$\begin{bmatrix} \lambda \\ \phi \end{bmatrix} = A^{-1} \cdot \beta \quad (20)$$

Η εξίσωση αυτή δίνει και τη λύση για τη συνήθη Kriging.

Η μέθοδος Kriging μαζί με την πρόβλεψη δίνει και την εκτίμηση του σφάλματος. Το σφάλμα σ_k , γνωστό και ως σφάλμα Kriging, είναι:

$$\sigma_k = \sqrt{C_0 - \sum_{k=1}^n \lambda_k C_k - \phi} \quad (21)$$

Ένας άλλος προσδιορισμός του σφάλματος είναι η μέθοδος Cross Validation (η επαλήθευση του μοντέλου του βαριογράμματος με την χρησιμοποίηση του kriging που περιλαμβάνει τον υπολογισμό του αθροίσματος των διαφορών για όλα τα σημεία, όταν κάθε σημείο δεδομένων παραλείπεται διαδοχικά και εκτιμάται από τα υπόλοιπα). Σε αυτήν μια τιμή αφαιρείται από το δείγμα και η αναμενόμενη τιμή αφαιρείται από την πραγματική. Η διαδικασία αυτή επαναλαμβάνεται για όλες τις τιμές και έτσι λαμβάνονται n υπόλοιπα τα οποία θεωρητικά έχουν κανονική κατανομή. Χρησιμοποιώντας τις εκτιμώμενες και τις πραγματικές τιμές για όλα τα σημεία υπολογίζονται ορισμένοι δείκτες, όπως το μέσο kriging λάθος και το τετράγωνο του μέσου kriging λάθους.

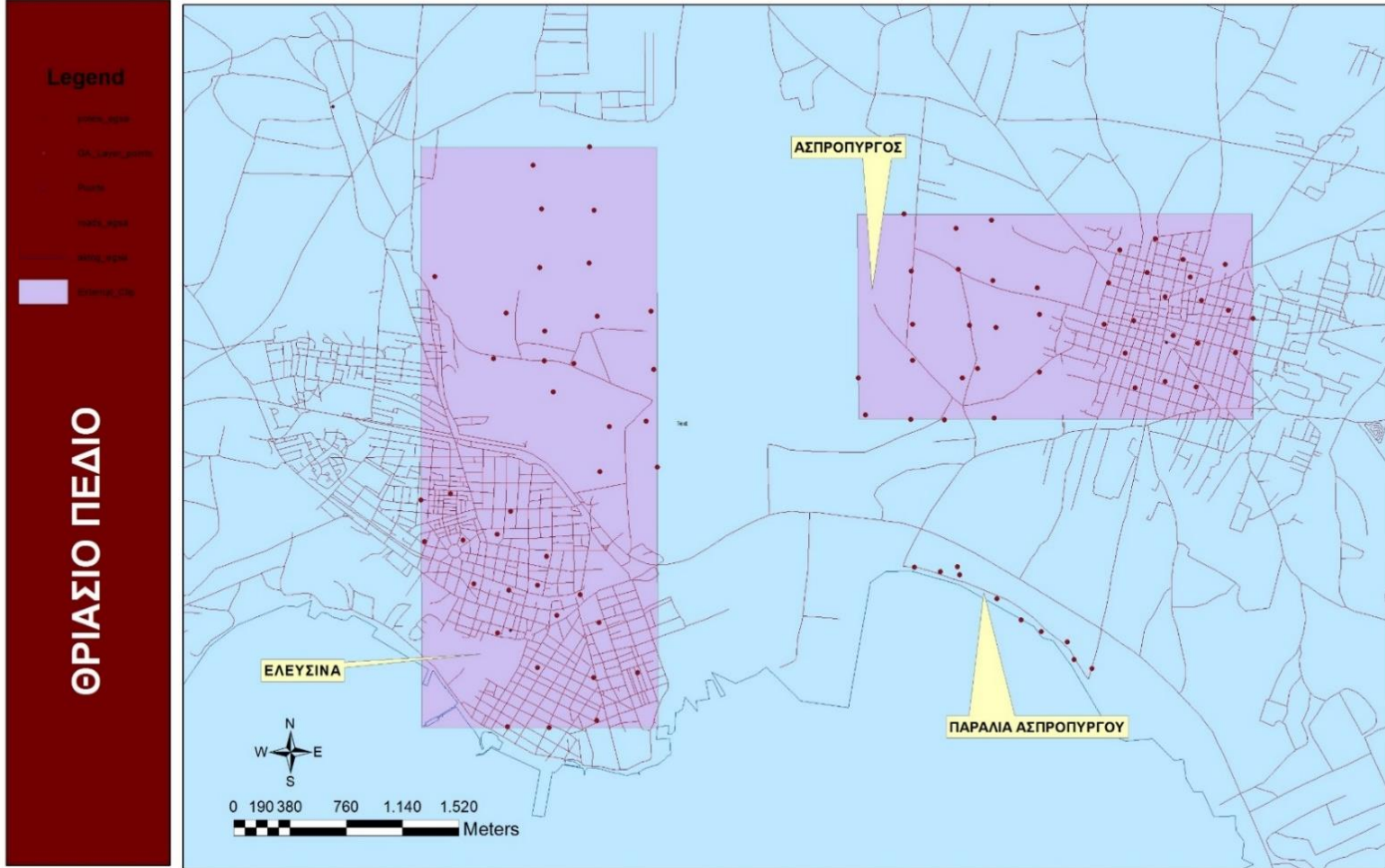
Το σφάλμα εκτίμησης προσδιορίζεται με τη μέθοδο Cross Validation, στην οποία κάθε μετρημένη τιμή συγκρίνεται με την αντίστοιχη εκτιμώμενη τιμή. Συγκεκριμένα από όλο το δείγμα μία μετρημένη τιμή κάθε φορά αφαιρείται και η εκτίμηση της τιμής της προσδιορίζεται από τις εναπομένουσες μετρημένες τιμές. Η προσδιορισμένη τιμή αφαιρείται από την πραγματική και έτσι προκύπτει το υπόλοιπο το οποίο λέγεται και θεωρητικό σφάλμα εκτίμησης. Η διαδικασία επαναλαμβάνεται για όλες τις τιμές και έτσι λαμβάνονται τόσα υπόλοιπα όσες είναι και οι μετρήσεις. Η μέθοδος αυτή χρησιμοποιεί όλο το δείγμα των μετρήσεων και προσδιορίζει την αυτοσυσχέτιση που εμφανίζεται μεταξύ των μετρημένων τιμών του δείγματος. Το RMSE είναι ένα στατιστικό μέγεθος που ποσοτικοποιεί το θεωρητικό σφάλμα που παράγεται από τη διαδικασία παρεμβολής, ουσιαστικά πρόκειται για την τυπική απόκλιση των υπολοίπων που προκύπτουν. Η διαδικασία προσδιορισμού του σφάλματος εκτίμησης προηγείται της παραγωγής

της τελικής επιφάνειας παρεμβολής γι' αυτό και μπορεί να θεωρηθεί ότι είναι ένα διαγνωστικό στάδιο όπου μελετάται αν οι σχετικές τιμές των παραμέτρων είναι λογικές οπότε και η παραγόμενη επιφάνεια παρεμβολής είναι αντιπροσωπευτική του φαινομένου που ο χρήστης θέλει να απεικονίσει.

Εκτός από το μέσο τετραγωνικό σφάλμα RMSE η μέθοδος Kriging σε αντίθεση με τις άλλες μεθόδους, προσδιορίζει και άλλα τυπικά σφάλματα, όπως το μέσο σφάλμα εκτίμησης (Mean Prediction Error), το μέσο τυπικό σφάλμα (Average Standard Error), το μέσο σφάλμα που προκύπτει από την προσαρμογή των εκτιμώμενων σημείων στην κανονική κατανομή (Average Standardized Error), και το μέσο τετραγωνικό σφάλμα από την προσαρμογή της κανονικής κατανομής (Root Mean Square Standardized Error). Για ένα μοντέλο που απεικονίζει ακριβείς μετρήσεις, θεωρητικά το Mean Prediction Error πρέπει να είναι κοντά στο μηδέν. Το μέσο τετραγωνικό σφάλμα RMSE πρέπει να είναι σχετικά μικρό αν οι εκτιμώμενες τιμές πλησιάζουν τις μετρημένες τιμές. Επιπλέον αν οι μετρήσεις είναι αμερόληπτες πρέπει το Root Mean Square Standardized Error να είναι κοντά στη μονάδα.

2.2. ΠΕΡΙΟΧΗ ΜΕΛΕΤΗΣ

Η πεδιάδα του Θριάσιου βρίσκεται 25 χιλιόμετρα βορειοδυτικά από το κέντρο της Αθήνας και χωρίζεται από τη γειτονική λεκάνη της Αθήνας και το βουνό Αιγάλεω. Αποτελείται από προσχλωσιγενή και κατακλυσμαία σχηματισμούς των καταθέσεων Τεταρτογενούς (Alexoudis-Λειβαδίτη et al. 1997). Η περιοχή μελέτης είναι γενικά επίπεδη. Στις τελευταίες πέντε δεκαετίες, το Θριάσιο Πεδίο έχει σε μεγάλο βαθμό βιομηχανίες και εξελίχθηκε από μια κατοικημένη-γεωργική έκταση σε μια περιοχή με μικτές χρήσεις γης (δηλαδή, οικιστικών, γεωργικών και βιομηχανικών). Διάφοροι τύποι των ρυπογόνων βιομηχανιών (δύο διυλιστήρια πετρελαίου, δύο βιομηχανίες χάλυβα, δύο εργοστάσια τσιμέντου, μονάδες ανακύκλωσης πετρελαίου, μια βιομηχανία πυρομαχικών, μεγάλες αποθήκες, εγκαταστάσεις διανομής πετρελαίου, και πολλές χημικές βιομηχανίες) βρίσκονται στην περιοχή μελέτης μεταξύ Ελευσίνας και Ασπροπύργου, δύο από τις σημαντικότερες πόλεις της πεδιάδας. Ο συνολικός πληθυσμός αυτών των πόλεων είναι περίπου 50.000 κατοίκους (απογραφή 2001), και η περιοχή μελέτης είναι περίπου 18,8 km². Η κοντινή παρουσία του ΧΥΤΑ της Αθήνας επιβάλλει πρόσθετη περιβαλλοντική απειλή για την περιοχή. Η πεδιάδα επίσης διασχίζεται από αυτοκινητόδρομους, κύριους και δευτερεύοντες οδούς. Το κλίμα στην πεδιάδα του Θριασίου είναι τυπικά μεσογειακό. Η μέση ετήσια βροχόπτωση είναι 390 χιλιοστά, η μέση θερμοκρασία είναι 18 ° C, και βόρειοι άνεμοι επικρατούν.



Χάρτης 1 Θρίασιο Πεδίο



Χάρτης 2

2.2.1. ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΕΔΑΦΩΝ

Η περιοχή μελέτης χωρίστηκε σε τέσσερα μέρη δειγματοληψίας που προσανατολίζονται από την παρουσία του στρατιωτικού αεροδρομίου και δύο κύριες οδικές αρτηρίες. Για να αποκτήσουμε ένα σχέδιο κατανομής της ρύπανσης των βαρέων μετάλλων στην περιοχή ακολουθήθηκε ένα σύστημα δειγματοληψίας πλέγμα τύπου (250 × 250 m) σε κάθε μέρος της δειγματοληψίας, και συνολικά 90 δείγματα εδάφους (0-5 cm βάθος) που συλλέχθηκαν προέρχονται από αστικά και αγροτικά εδάφη σε ίσους αριθμούς. Δύο δείγματα από σαράντα σημεία και άλλο ένα δείγμα από δέκα σημεία λήφθηκαν και αναμίχθηκαν για να κάνουν κάθε ένα από τα 90 σημεία τρία επιμέρους δείγματα από μια επιφάνειας 40 x 40 cm.

Τα δείγματα εδάφους ξηραίνονται στον αέρα και κοσκινίζεται από 2-mm. Η περιεκτικότητα σε οργανική ύλη και μηχανική σύσταση των δειγμάτων προσδιορίστηκαν από την Walkley-Black διαδικασία (Nelson και Sommers 1982) και με τη υδρομετρική μέθοδο (Βουγουκος 1951), αντίστοιχα. Οι τιμές pH μετρήθηκαν με τη χρήση τυποποιημένων ηλεκτροδίων glass/calomel σε 1:1 w/v λόγος εδάφους-νερού, και το αντίστοιχο ποσοστό CaCO₃ εκτιμήθηκε με μέτρηση του CO₂ ακολουθώντας HCl διάλυση. Για να αποκτήσουμε τις συγκεντρώσεις μετάλλων "ψευδο-σύνολο", τα δείγματα χώματος μετρήθηκαν με aqua regia (Gasparatos and Haidouti 2001). Ο όρος "ψευδο-σύνολο" χρησιμοποιείται επειδή η διάβρωση του aqua regia δεν καταστρέφει εντελώς τα πυριτικά (Facchinelli et al. 2001). Το διαθέσιμο κλάσμα των υπό μελέτη μετάλλων: Pb, Ni, Zn, Fe, Co, Cr, Cu, Mn, Ba, Cd, Pb₂, Ni₂, Zn₂, Fe₂,

Co₂, Cr₂, Cu₂, Mn₂, Ba₂ και Cd₂ εκχυλίζεται με DTPA (Lindsay και Norvell 1978). Στο κείμενο, οι όροι συνολικός και ο διαθέσιμος χρησιμοποιούνται για την aqua regia και DTPA. Όλες οι συγκεντρώσεις μετάλλων προσδιορίστηκαν από το ατομικό φασματοφωτομετρητή απορρόφησης, χρησιμοποιώντας ένα Varian spectra A300. Τα πρότυπα βαθμονόμησης παρασκευάστηκαν στην ίδια μήτρα που χρησιμοποιείται για τα δείγματα εδάφους. Σε όλα τα στάδια της προετοιμασίας και της ανάλυσης του δείγματος ελήφθησαν αυστηρές προφυλάξεις για την ελαχιστοποίηση της μόλυνσης μέσω του αέρα, των γυάλινων σκευών και των αντιδραστηρίων. Ο έλεγχος των δειγμάτων αναλύθηκε για κάθε δέκα δείγματα και η αναπαραγωγιμότητα δοκιμάστηκε επανεξετάζοντας το 30% των δειγμάτων. Η αναλυτική ακρίβεια, εκτιμήθηκε ως σχετική τυπική απόκλιση, ότι ήταν μικρότερη από 3%.

2.2.2. ΔΕΙΚΤΕΣ ΜΟΛΥΝΣΗΣ ΕΔΑΦΩΝ

Η περιβαλλοντική επίδραση των μετάλλων και το επίπεδο της ρύπανσης στα ιζήματα μπορεί να προσδιοριστεί με τη βοήθεια δύο παραμέτρων την αναλογία εμπλουτισμού Enrichment Factor (EF) και του δείκτη γεωσυσσώρευση Index of geoaccumulation (Igeo). Ο παράγοντας εμπλουτισμού (EF), λόγω της γενικής φόρμουλα, είναι ένα σχετικά απλό και εύκολο εργαλείο για την αξιολόγηση του βαθμού εμπλουτισμού και σύγκριση της μόλυνσης των διαφόρων περιβαλλοντικών μέσων (Benhaddya και Hadjel, 2013). Το EF είναι μια μέθοδος κανονικοποίησης που προτείνει ο Simex και Helz (1981) για την εκτίμηση της συγκέντρωσης των μετάλλων. Ομαλοποιεί την συγκέντρωση μετάλλου ως αναλογία σε άλλο συστατικό των ιζημάτων. Ο Rubio et al. (2000) αναφέρει ότι δεν υπάρχει συναίνεση σχετικά με το πιο είναι κατάλληλο συστατικό του ιζήματος για να χρησιμοποιηθεί για την εξομάλυνση. Μεταξύ εκείνων που χρησιμοποιούνται έχουν Al, Fe, ολικού οργανικού άνθρακα και το μέγεθος των κόκκων. Η σύσταση που επιλέγεται για το σκοπό αυτό θα πρέπει να συνδέονται με λεπτότερα σωματίδια (σε σχέση με το μέγεθος κόκκου) και η συγκέντρωσή του δεν θα πρέπει να μεταβληθεί ανθρωπογενός (Ackerman, 1980). Ο εμπλουτισμός παραγόντων των μετάλλων (EF) υπολογίστηκε όπως περιγράφεται από τον Massas et al. (2010), και η ακόλουθη εξίσωση που χρησιμοποιήθηκε:

$$EF = C_{it} / A_{it}^{(22)}$$

όπου C_{it} είναι η συνολική συγκέντρωση των μετάλλων στο σημείο δειγματοληψίας i και A_{it} είναι η αξία του μετάλλου A του ολλανδικού συστήματος, που λαμβάνεται με την προσαρμογή της σταθερής αξίας A_{tv} του συστήματος για την περιοχή δειγματοληψίας i σύμφωνα με την εξίσωση:

$$A_i = A_{tv} \times \{ \{ a + (b \times \%clay) + (c \times \%organic\ matter) \} / \{ (a + (b \times 25) + (c \times 10)) \} \}$$

Όπου A_{tv} σταθερή A τιμή του μετάλλου στο ολλανδικό σύστημα, $\%clay$ ποσοστό αργίλλου (μέγεθος κόκκου <2 mm) στο έδαφος που πρέπει να αξιολογηθούν, $\%organic\ matter$ ποσοστό οργανικής ύλης (w / w) στο έδαφος που πρέπει να αξιολογηθούν και a, b, c εξαρτώμενες σταθερές για κάθε μέταλλο. $EF > 1$ μπορεί να θεωρηθεί πιθανή για τον εμπλουτισμό του εδάφους.

Ο δείκτης γεωσυσσώρευσης (Igeo) εισήχθη από τους Muller (1969) και έχει εφαρμοστεί σε πρόσφατες μελέτες ρύπανσης για την αξιολόγηση της ρύπανσης του εδάφους από βαρέα μέταλλα (Srinivasa Gowd et al 2010;. Yaylali-Abanoz 2011). Για τον υπολογισμό, η ακόλουθη εξίσωση χρησιμοποιήθηκε είναι :

$$I_{geo} = \log_2 [C_n / 1,5B_n] \quad (23)$$

όπου C_n είναι η μετρούμενη συγκέντρωση στο ίζημα για το μέταλλο n , B_n είναι η τιμή υποβάθρου για το μέταλλο n (Turekian και Wedepohl, 1961), και ο συντελεστής 1,5 χρησιμοποιείται λόγω των πιθανών διακυμάνσεων των δεδομένων λόγω λιθολογικών παραλλαγών. Η ποσότητα I_{geo} υπολογίζεται χρησιμοποιώντας τον παγκόσμιο μέσο όρο των δεδομένων σχιστόλιθου από Turekian και Wedepohl (1961). Οι επτά τάξεις του δείκτη I_{geo} όπως προτείνεται από Muller (1969) έχουν ως εξής:

$I_{geo} \leq 0$: αμόλυντα

$0 < I_{geo} < 1$: αμόλυντα ή μέτρια μολυσμένα

$1 < I_{geo} < 2$: μέτρια μολυσμένα

$2 < I_{geo} < 3$: μέτρια προς μολυσμένα

$3 < I_{geo} < 4$: βαριά μολυσμένα

$4 < I_{geo} < 5$: βαριά προς εξαιρετικά μολυσμένα

$5 \leq I_{geo}$: εξαιρετικά μολυσμένο

Ο ακόλουθος τύπος εφαρμόστηκε για τον υπολογισμό του δείκτη διαθεσιμότητας Availability ratio (AR) (Massas et al. 2010):

$$AR = (C_{ia} / C_{it}) \times 10^2 \quad (24)$$

Όπου C_{ia} η διαθέσιμη συγκέντρωση μετάλλου στο σημείο δειγματοληψίας i και C_{it} η συνολική συγκέντρωση μετάλλου στο σημείο δειγματοληψίας i

2.2.3. ΓΕΩΓΡΑΦΙΚΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

Για την μελέτη των βαρέων μετάλλων αναπτύχθηκε μετά από ειδική ανάλυση και σχεδιασμό ένα Γεωγραφικό Σύστημα Πληροφοριών ArcGIS 10.2 στο οποίο μετά από διαδικασία γεωαναφοράς και ψηφιοποίησης αποθηκεύτηκαν πληροφορίες από χάρτες. Καταχωρήθηκαν πληροφορίες από σημεία δειγματοληψίας καθώς και αναλύσεις από τις εδαφικές ιδιότητες. Τα ψηφιακά δεδομένα παραχωρήθηκαν από την Ερευνητική Μονάδα Γεωγραφικών Πληροφοριακών Συστημάτων του Γεωπονικού Πανεπιστημίου Αθηνών.

3. ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

3.1. ΜΗ ΧΩΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Στην παρούσα μελέτη υπολογίζουμε τα μέτρα κεντρικής θέσης και διασποράς για τα 90 σημεία. Στην συνέχεια για να ελέγξουμε αν η κατανομή μιας μεταβλητής είναι συμβατή με την κανονική εφαρμόζουμε το test Kolmogorov – Smirnov και Shapiro-Wilk. Ανάλογα με το p -value (Sig) < α , $\alpha=0,05$, τότε απορρίπτεται η H_0 .

	N		Mean	Median	Mode	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	CV
	Valid	Missing										
sand	90,00	0,00	52,70	52,53	46,60	8,67	75,12	0,01	0,25	-0,02	0,50	16,45
clay	90,00	0,00	19,96	20,49	20,800000a	5,34	28,49	-0,16	0,25	-0,94	0,50	26,74
silt	90,00	0,00	27,34	26,86	24,60	6,04	36,46	0,13	0,25	0,76	0,50	22,09
pH	90,00	0,00	7,91	7,96	7,790000a	0,35	0,12	-0,54	0,25	-0,17	0,50	4,38
CaCO3	90,00	0,00	25,75	24,91	14,145000a	11,72	137,33	0,73	0,25	0,79	0,50	45,50
Org_Matter	90,00	0,00	3,41	3,22	3,48	2,05	4,19	1,19	0,25	2,64	0,50	60,01
Na	90,00	0,00	1,53	0,67	0,35	2,86	8,16	4,42	0,25	23,10	0,50	186,57
k	90,00	0,00	1,18	1,16	0,81	0,60	0,36	0,73	0,25	1,27	0,50	50,40
Pb	90,00	0,00	137,59	113,25	78,000000a	119,18	14203,70	5,25	0,25	30,80	0,50	86,62
Ni	90,00	0,00	76,67	74,20	43,650000a	27,14	736,34	0,61	0,25	-0,36	0,50	35,39
Zn	90,00	0,00	182,26	153,50	131,50	121,30	14714,89	3,69	0,25	17,69	0,50	66,56
Fe	90,00	0,00	15220,05	15272,50	6390,000000a	4813,63	23171043,54	0,15	0,25	-0,79	0,50	31,63
Co	90,00	0,00	23,00	24,00	24,000000a	4,18	17,48	-0,11	0,25	-0,32	0,50	18,17
Cr	90,00	0,00	78,36	79,49	46,550000a	28,95	837,82	0,69	0,25	-0,09	0,50	36,94
Cu	90,00	0,00	45,17	38,18	24,350000a	41,00	1681,24	4,44	0,25	21,68	0,50	90,78
Mn	90,00	0,00	326,19	309,25	233,500000a	107,09	11468,80	0,22	0,25	0,19	0,50	32,83
Ba	90,00	0,00	840,13	841,00	752,500000a	244,01	59540,37	0,59	0,25	2,28	0,50	29,04
Cd	90,00	0,00	6,80	5,76	0,00	4,47	19,99	0,18	0,25	-0,92	0,50	65,74
clay_2	90,00	0,00	50,24	49,20	49,20	7,56	57,22	-0,15	0,25	0,03	0,50	15,06
sand_2	90,00	0,00	22,47	22,67	20,20	4,88	23,77	-0,06	0,25	-0,75	0,50	21,70
silt_2	90,00	0,00	27,30	27,48	22,60	5,21	27,20	0,04	0,25	1,14	0,50	19,10
Ph_2	90,00	0,00	8,08	8,14	7,97	0,31	0,09	-1,13	0,25	1,57	0,50	3,78
CaCO3_2	90,00	0,00	25,27	23,78	13,530000a	12,94	167,32	1,03	0,25	2,22	0,50	51,19
Na_2	90,00	0,00	1,20	0,58	0,35	1,67	2,79	3,43	0,25	14,28	0,50	139,51
k_2	90,00	0,00	1,00	0,97	,592195a	0,50	0,25	0,73	0,25	1,17	0,50	50,49
Pb_2	90,00	0,00	120,61	111,18	97,000000a	53,39	2850,41	3,17	0,25	13,91	0,50	44,26
Ni_2	90,00	0,00	86,48	74,85	86,500000a	56,31	3170,37	6,21	0,25	49,49	0,50	65,11
Zn_2	90,00	0,00	172,17	154,00	120,50	107,54	11565,83	3,16	0,25	12,45	0,50	62,47
Fe_2	90,00	0,00	16675,72	16567,50	19465,00	4541,50	20625200,88	0,03	0,25	-0,51	0,50	27,23
Co_2	90,00	0,00	23,76	24,43	21,100000a	3,97	15,78	-0,45	0,25	0,10	0,50	16,72
Cr_2	90,00	0,00	81,61	78,08	52,750000a	32,48	1054,67	1,03	0,25	0,84	0,50	39,79
Cu_2	90,00	0,00	41,20	36,03	26,250000a	27,06	732,16	4,98	0,25	35,94	0,50	65,68
Mn_2	90,00	0,00	350,74	336,75	313,000000a	97,54	9514,59	0,38	0,25	-0,62	0,50	27,81
Ba_2	90,00	0,00	827,10	825,75	751,500000a	234,89	55172,77	-0,34	0,25	0,32	0,50	28,40
Cd_2	90,00	0,00	7,35	6,44	0,00	4,54	20,61	0,12	0,25	-0,84	0,50	61,78
Org_Matt_2	90,00	0,00	2,69	2,61	1,4070a	1,57	2,46	1,09	0,25	2,65	0,50	58,24
a. Multiple modes exist. The smallest value is shown												

Πίνακας 1

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
sand	,075	90	,200*	,987	90	,523
clay	,077	90	,200*	,972	90	,052
silt	,085	90	,139	,987	90	,495
pH	,102	90	,022	,964	90	,014
CaCO3	,082	90	,185	,965	90	,016
Org_Matter	,110	90	,009	,928	90	,000
Na	,343	90	,000	,443	90	,000
k	,076	90	,200*	,964	90	,013
Pb	,340	90	,000	,399	90	,000
Ni	,129	90	,001	,939	90	,000
Zn	,245	90	,000	,630	90	,000
Fe	,070	90	,200*	,979	90	,142
Co	,133	90	,000	,970	90	,037
Cr	,146	90	,000	,921	90	,000
Cu	,270	90	,000	,493	90	,000
Mn	,082	90	,177	,978	90	,124
Ba	,085	90	,120	,964	90	,014
Cd	,102	90	,022	,961	90	,009
clay_2	,089	90	,078	,987	90	,517
sand_2	,065	90	,200*	,984	90	,326
silt_2	,062	90	,200*	,980	90	,184
Ph_2	,133	90	,000	,914	90	,000
CaCO3_2	,085	90	,135	,941	90	,000
Na_2	,310	90	,000	,569	90	,000
k_2	,069	90	,200*	,966	90	,020
Pb_2	,229	90	,000	,690	90	,000
Ni_2	,213	90	,000	,498	90	,000
Zn_2	,230	90	,000	,669	90	,000
Fe_2	,064	90	,200*	,990	90	,746
Co_2	,109	90	,010	,974	90	,064
Cr_2	,129	90	,001	,903	90	,000
Cu_2	,168	90	,000	,615	90	,000
Mn_2	,093	90	,052	,969	90	,030
Ba_2	,079	90	,200*	,983	90	,272
Cd_2	,129	90	,001	,955	90	,004
Org_Matt_2	,092	90	,056	,939	90	,000

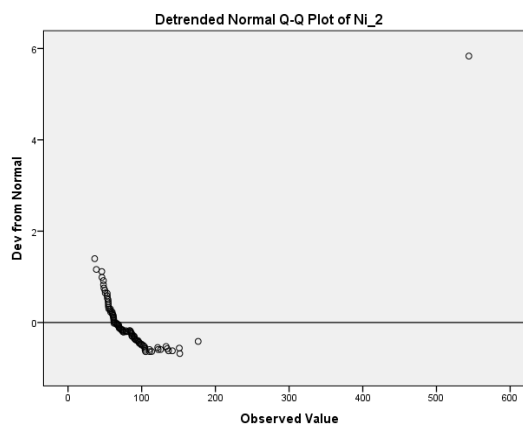
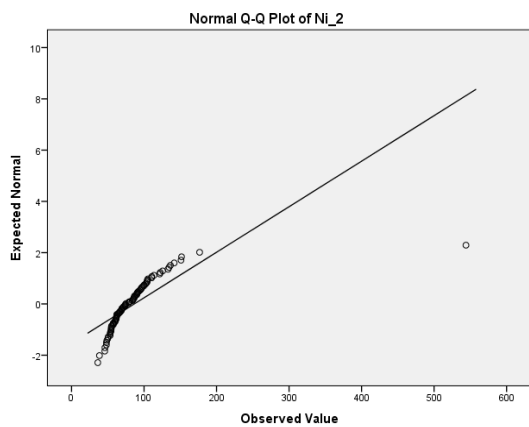
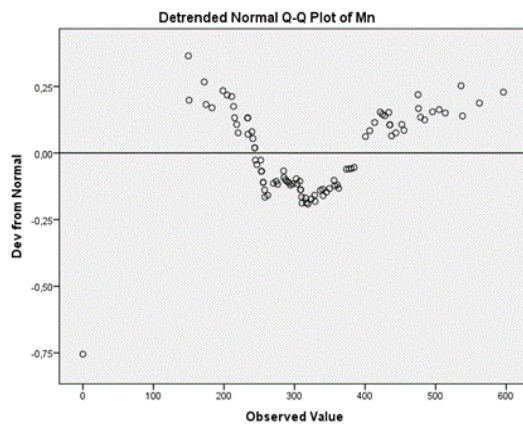
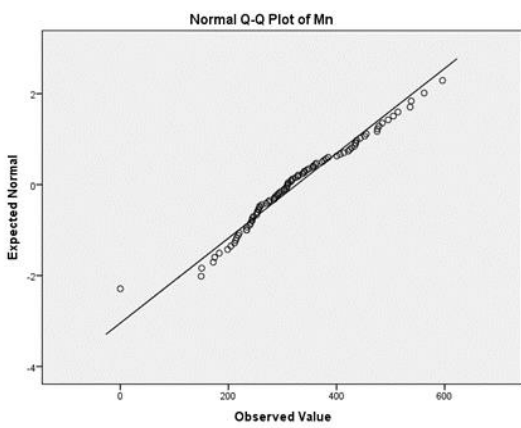
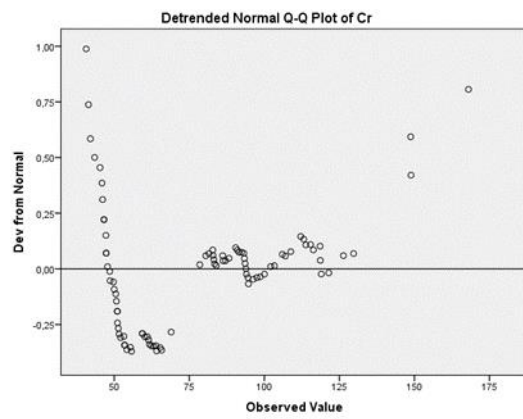
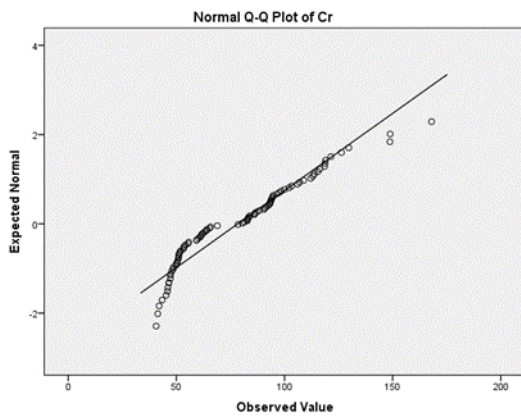
*. This is a lower bound of the true significance.

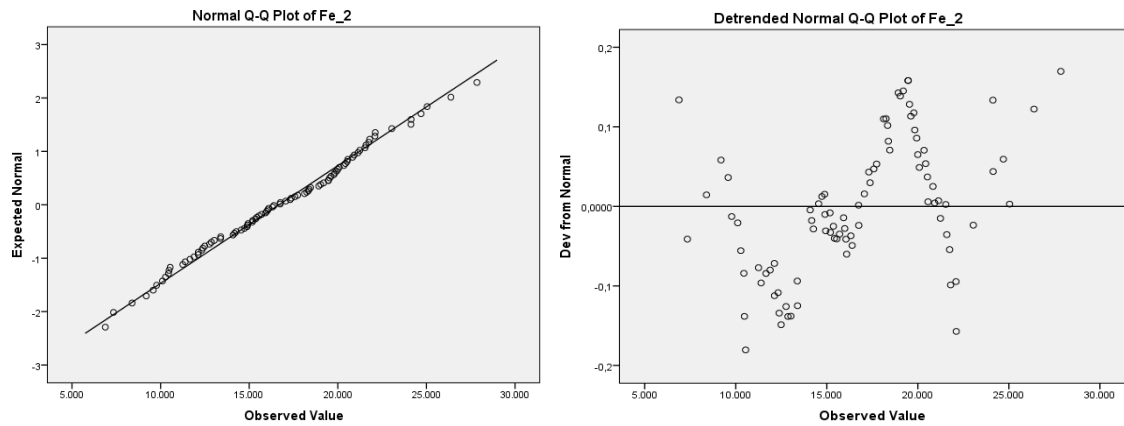
a. Lilliefors Significance Correction

Πίνακας 2

Στις περιπτώσεις που το Sig. < 0,05 η μηδενική υπόθεση απορρίπτεται και η κατανομή του αντίστοιχου πληθυσμού δεν θεωρείται κανονική. Ενώ όταν το Sig. > 0,05 η μηδενική υπόθεση δεν απορρίπτεται και η κατανομή του αντίστοιχου πληθυσμού θεωρείται κανονική. Για πολλούς από τους πληθυσμούς η στατιστική επεξεργασία

προκύπτει ότι και με τα δύο τεστ αυτοί ακολουθούν κανονική κατανομή. Παρατηρώντας και τα αντίστοιχα διαγράμματα καταλήγουμε στα ίδια.





Διαγράμματα Normal Q-Q Plot και Detrended Normal Q-Q Plot

Στα διαγράμματα Normal Q-Q Plot, στον άξονα των x βρίσκονται οι παρατηρούμενες τιμές της εκάστοτε μεταβλητής και στον άξονα των y βρίσκονται οι αναμενόμενες τιμές αυτής. Σε μια ιδανική κατάσταση για την κανονικότητα, όλα τα σημεία θα βρίσκονται πάνω στην διχοτόμο των αξόνων. Αυτό όμως είναι σχεδόν απίθανο, οπότε μας ενδιαφέρει τα σημεία να βρίσκονται πολύ κοντά στην διχοτόμο των αξόνων. Σε μια τέτοια περίπτωση μπορούμε να πούμε ότι ο πληθυσμός ακολουθεί προσεγγιστικά την κανονική κατανομή.

Τα διαγράμματα Detrended Normal Q-Q Plot δείχνουν τις διαφορές μεταξύ των παρατηρούμενων και των αναμενόμενων τιμών σε μια κανονική κατανομή. Εάν η κατανομή είναι κανονική (ιδανική περίπτωση), όλα τα σημεία πρέπει να είναι ομοιόμορφα κατανεμημένα σε μια οριζόντια ζώνη γύρω από το μηδέν.

Οι τιμές της μεταβλητής με τον μεγαλύτερο συντελεστή μεταβλητότητας (Coefficient of Variation) είναι πιο διασκορπισμένες γύρω από τη μέση τιμή της μεταβλητής. Επίσης, και τα δείγματα των παραπάνω παραδειγμάτων που εξετάζουμε δεν έχουν ομοιογένεια, αφού οι συντελεστές μεταβλητότητάς τους ξεπερνούν το 10%.

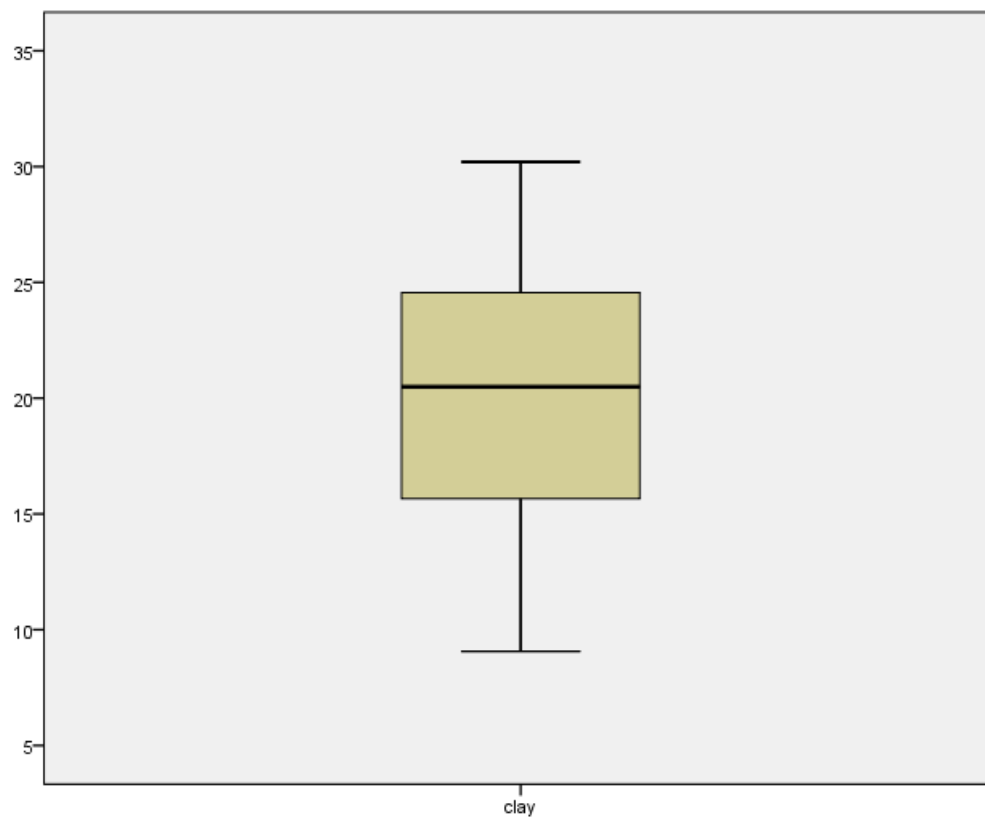
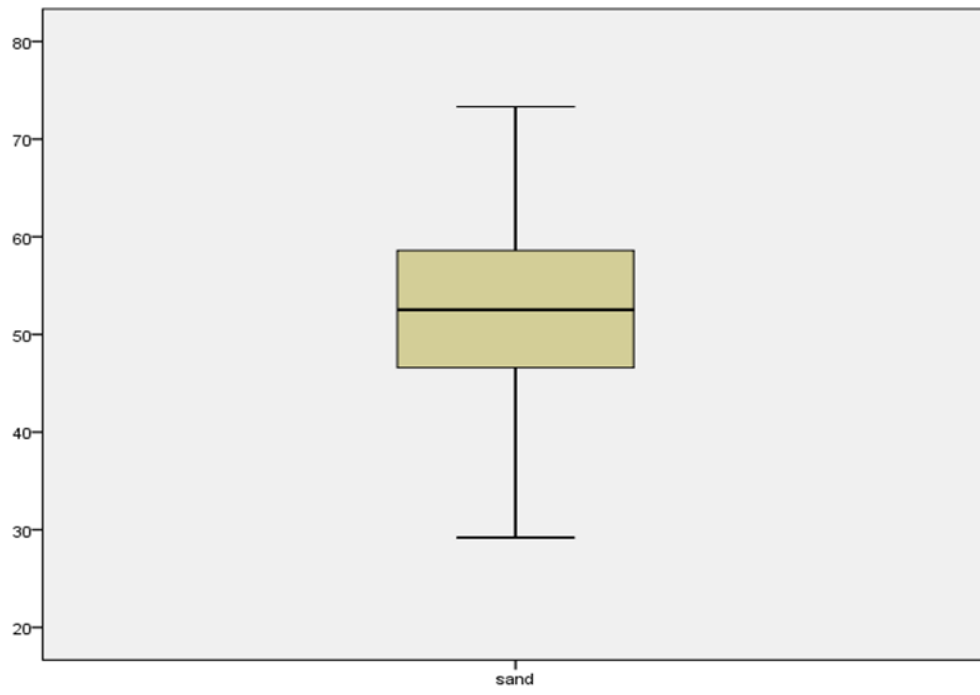
3.1.1. ΥΠΟΛΟΓΙΣΜΟΣ OUTLIERS

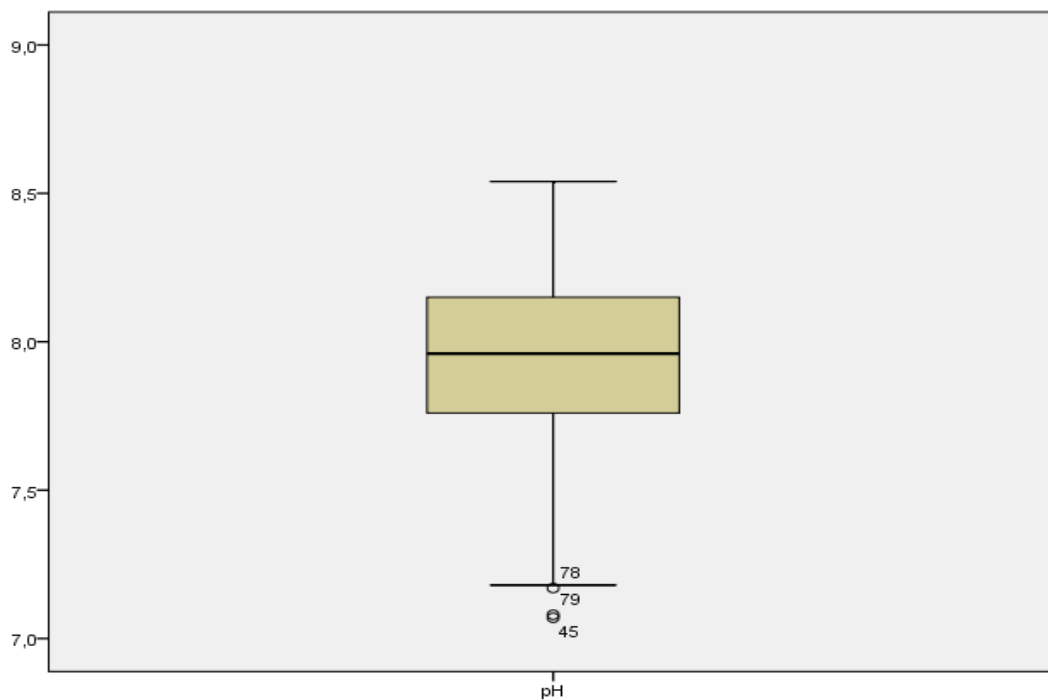
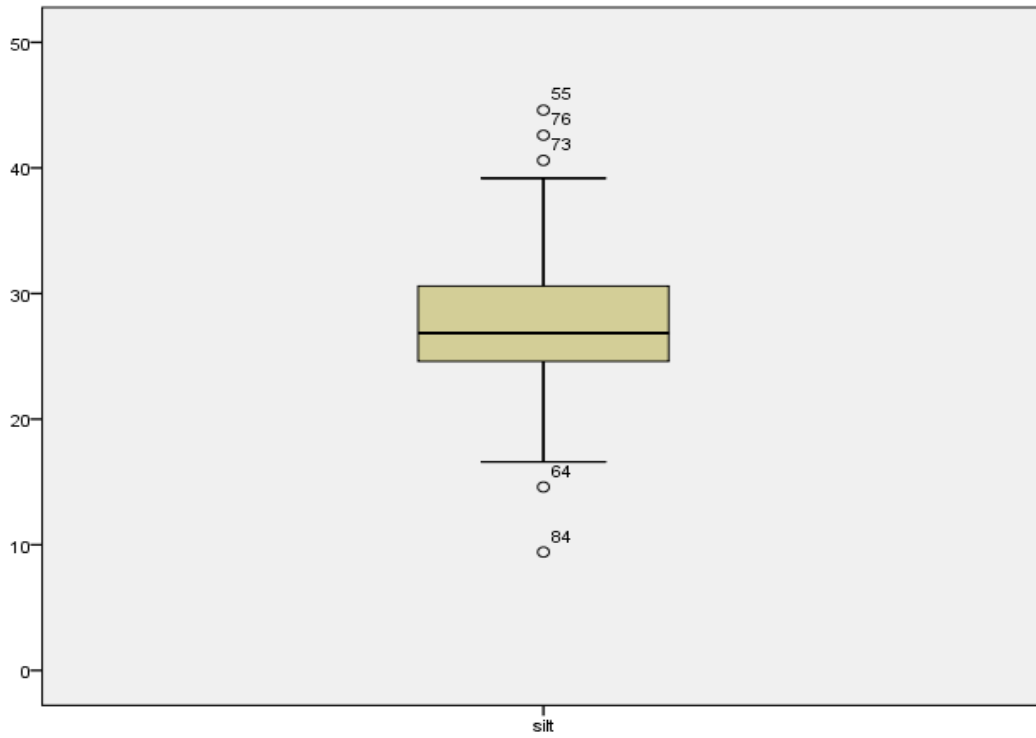
Προϋπόθεση για τον υπολογισμό των outliers είναι η κανονικότητα των πληθυσμών από τους οποίους προέρχονται τα δείγματά μας.

Ο υπολογισμός του ανώτερου και κατώτερου ορίου γίνεται με βάση τους παρακάτω τύπους:

$$U=Q_3+[2.2*(Q_3-Q_1)]$$

$$L=Q_1-[2.2*(Q_3-Q_1)]$$





Από τους προηγούμενους τύπους και με βάση τον πίνακα Percentiles και από τα boxplot μπορούμε να υπολογίσουμε τα U και L για τις μεταβλητές. Από τον πίνακα Extreme Values, μπορούμε να δούμε αν υπάρχουν τιμές (παρατηρήσεις) που να βρίσκονται εκτός των ορίων outliers. Μόνο για τις μεταβλητές clay, sand, Ni, Fe, Co, Mn, Cd, clay_2, sand_2, silt_2, Fe_2, Co_2, Mn_2 και Cd_2 παρατηρούμε ότι όλες οι τιμές βρίσκονται εντός των διαστημάτων ($\mu-2\sigma$, $\mu+2\sigma$) ή ($\mu-3\sigma$, $\mu+3\sigma$).

3.1.2. ΣΥΣΧΕΤΙΣΕΙΣ ΜΕΤΑΞΥ ΜΕΤΑΒΛΗΤΩΝ

Ο συντελεστής γραμμικής συσχέτισης μιας μεταβλητής μπορεί να πάρει τιμές στο διάστημα $[-1,1]$. Η συσχέτιση συνήθως καλείται διμεταβλητή συσχέτιση για να υποδηλώσει μια απλή συσχέτιση μεταξύ δύο μεταβλητών. Παρότι ο συντελεστής γραμμικής συσχέτισης r στηρίζεται στην υπόθεση ότι οι δύο εμπλεκόμενες μεταβλητές ακολουθούν περίπου την κανονική κατανομή, ο τύπος συχνά αποδίδει καλά ακόμα κι όταν η υπόθεση της κανονικότητας δεν ικανοποιείται. Όταν οι μεταβλητές δεν ακολουθούν κανονική κατανομή, ο συντελεστής συσχέτισης Spearman (μια τιμή που βασίζεται στη διάταξη των τιμών) είναι πιο κατάλληλος. Η τιμή $+1$ δείχνει μια σημαντική θετική συσχέτιση. Μία τέλεια έλλειψη συγγένειας αποδεικνύεται με ένα συντελεστή συσχέτισης μηδέν ενώ όταν η τιμή είναι -1 δείχνει αρνητική συσχέτιση. Από τα αποτελέσματα προκύπτει ότι στις περισσότερες περιπτώσεις υπάρχει στατιστικά σημαντική γραμμική συσχέτιση.

3.1.3. ΕΛΕΓΧΟΣ ΤΗΣ ΔΙΑΦΟΡΑΣ ΤΩΝ ΜΕΣΩΝ ΟΡΩΝ

Για τον έλεγχο της διαφοράς μέσων όρων χρησιμοποιούμε τη μέθοδο paired T-test

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	sand & clay_2	90	,783	,000
Pair 2	clay & sand_2	90	,744	,000
Pair 3	silt & silt_2	90	,666	,000
Pair 4	pH & Ph_2	90	,596	,000
Pair 5	CaCO3 & CaCO3_2	90	,667	,000
Pair 6	Org_Matter & Org_Matt_2	90	,644	,000
Pair 7	Na & Na_2	90	,783	,000
Pair 8	k & k_2	90	,796	,000
Pair 9	Pb & Pb_2	90	,662	,000
Pair 10	Ni & Ni_2	90	,245	,020
Pair 11	Zn & Zn_2	90	,689	,000
Pair 12	Fe & Fe_2	90	,715	,000
Pair 13	Co & Co_2	90	,603	,000
Pair 14	Cr & Cr_2	90	,699	,000
Pair 15	Cu & Cu_2	90	,704	,000
Pair 16	Mn & Mn_2	90	,865	,000
Pair 17	Ba & Ba_2	90	,626	,000
Pair 18	Cd & Cd_2	90	,889	,000

Πίνακας 5

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	sand - clay_2	2,47	5,45	0,57	1,33	3,61	4,30	89	0,00
Pair 2	clay - sand_2	-2,51	3,68	0,39	-3,28	-1,74	-6,47	89	0,00
Pair 3	silt - silt_2	0,04	4,66	0,49	-0,93	1,02	0,09	89	0,93
Pair 4	pH - Ph_2	-0,18	0,30	0,03	-0,24	-0,11	-5,66	89	0,00
Pair 5	CaCO3 - CaCO3_2	0,48	10,12	1,07	-1,64	2,60	0,45	89	0,65
Pair 6	Org_Matter - Org_Matt_2	0,72	1,58	0,17	0,38	1,05	4,29	89	0,00
Pair 7	Na - Na_2	0,33	1,87	0,20	-0,06	0,72	1,70	89	0,09
Pair 8	k - k_2	0,18	0,36	0,04	0,11	0,26	4,84	89	0,00
Pair 9	Pb - Pb_2	16,97	92,88	9,79	-2,48	36,43	1,73	89	0,09
Pair 10	Ni - Ni_2	-9,80	56,18	5,92	-21,57	1,96	-1,66	89	0,10
Pair 11	Zn - Zn_2	10,09	91,09	9,60	-8,99	29,17	1,05	89	0,30
Pair 12	Fe - Fe_2	-1455,67	3540,54	373,21	-2197,22	-714,11	-3,90	89	0,00
Pair 13	Co - Co_2	-0,76	3,64	0,38	-1,52	0,00	-1,98	89	0,05
Pair 14	Cr - Cr_2	-3,25	24,06	2,54	-8,29	1,79	-1,28	89	0,20
Pair 15	Cu - Cu_2	3,97	29,18	3,08	-2,14	10,08	1,29	89	0,20
Pair 16	Mn - Mn_2	-24,55	53,87	5,68	-35,83	-13,27	-4,32	89	0,00
Pair 17	Ba - Ba_2	13,03	207,20	21,84	-30,36	56,43	0,60	89	0,55
Pair 18	Cd - Cd_2	-0,55	2,12	0,22	-0,99	-0,10	-2,45	89	0,02

Πίνακας 6

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sand	52,7	90,0	8,7	0,9
	clay_2	50,2	90,0	7,6	0,8
Pair 2	clay	20,0	90,0	5,3	0,6
	sand_2	22,5	90,0	4,9	0,5
Pair 3	silt	27,3	90,0	6,0	0,6
	silt_2	27,3	90,0	5,2	0,6
Pair 4	pH	7,9	90,0	0,3	0,0
	Ph_2	8,1	90,0	0,3	0,0
Pair 5	CaCO3	25,8	90,0	11,7	1,2
	CaCO3_2	25,3	90,0	12,9	1,4
Pair 6	Org_Matter	3,4	90,0	2,0	0,2
	Org_Matt_2	2,7	90,0	1,6	0,2
Pair 7	Na	1,5	90,0	2,9	0,3
	Na_2	1,2	90,0	1,7	0,2
Pair 8	k	1,2	90,0	0,6	0,1
	k_2	1,0	90,0	0,5	0,1
Pair 9	Pb	137,6	90,0	119,2	12,5
	Pb_2	120,6	90,0	53,4	5,6
Pair 10	Ni	76,7	90,0	27,1	2,9
	Ni_2	86,5	90,0	56,3	5,9
Pair 11	Zn	182,3	90,0	121,3	12,8
	Zn_2	172,2	90,0	107,5	11,3
Pair 12	Fe	15.220	90,0	4.813,6	507,4
	Fe_2	16.675,7	90,0	4.541,5	478,7
Pair 13	Co	23,0	90,0	4,2	0,4
	Co_2	23,8	90,0	4,0	0,4
Pair 14	Cr	78,4	90,0	28,9	3,1
	Cr_2	81,6	90,0	32,5	3,4
Pair 15	Cu	45,2	90,0	41,0	4,3
	Cu_2	41,2	90,0	27,1	2,9
Pair 16	Mn	326,2	90,0	107,1	11,3
	Mn_2	350,7	90,0	97,5	10,3
Pair 17	Ba	840,1	90,0	244,0	25,7
	Ba_2	827,1	90,0	234,9	24,8
Pair 18	Cd	6,8	90,0	4,5	0,5
	Cd_2	7,3	90,0	4,5	0,5

Πίνακας 7

3.1.4. ΟΡΙΣΜΟΣ ΤΕΧΝΙΚΗΣ ΑΝΑΛΥΣΗΣ (Trend Analysis)

Τεχνική ανάλυση μιας τιμής ή ενός Δείκτη, ή μιας ολόκληρης οργανωμένης μελέτης, είναι η μελέτη της κίνησης των τιμών, όπως αυτή απεικονίζεται σε ένα διάγραμμα τιμής-χρόνου, με στόχο τον χαρακτηρισμό της συμπεριφοράς της και την πρόβλεψη της μελλοντικής της κίνησης. Στηρίζεται στα γραφήματα της Αναλυτικής Γεωμετρίας και της Στατιστικής, τα οποία όμως έχουν εκλαϊκευτεί, για να μπορεί να τα καταλαβαίνει ο απλός παρατηρητής.

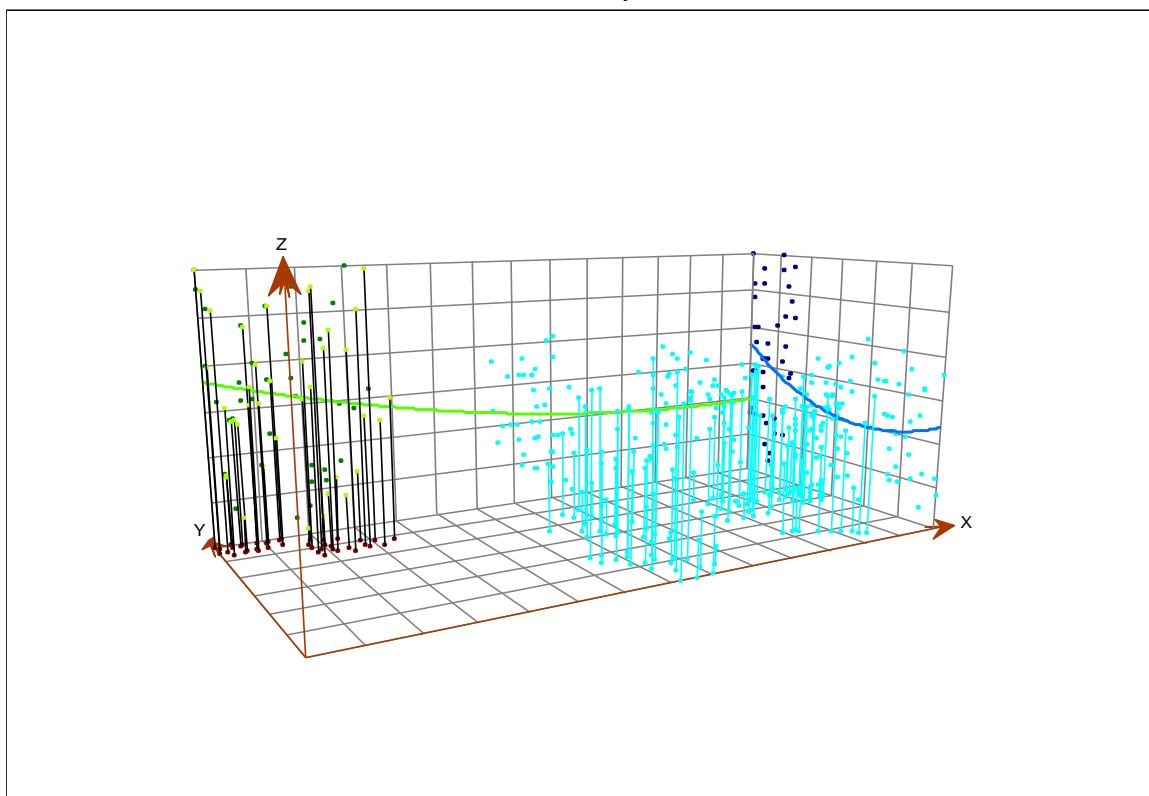
Βασική υπόθεση της τεχνικής ανάλυσης είναι πως ότι είναι γνωστό για ένα συντελεστή ή παράγοντα είναι ήδη ενσωματωμένο στην τιμή της. Για την τεχνική ανάλυση δεν έχει σημασία τι προκάλεσε την άνοδο ή κάθοδο της τιμής ενός παράγοντα, αλλά πόσο θα διαρκέσει αυτή η μεταβολή, είτε προς τα πάνω είτε προς τα κάτω.

Επίσης βασική παραδοχή της τεχνικής ανάλυσης αποτελεί το γεγονός ότι η τιμή ενός παράγοντα απεικονίζει τα δεδομένα, τις πληροφορίες και επηρεάζεται σημαντικά από τις μεταβολές εδάφους και τα φυσικά φαινόμενα που επικρατούν κάθε περίοδο.

Η Τεχνική Ανάλυση μελετάει τις μεταβολές τιμών από το παρελθόν όπου προσπαθεί να εντοπίσει έγκαιρα τις μεταβολές που θα γίνουν στο μέλλον και κυρίως το σημείο αντιστροφής μιας κίνησης τιμών.

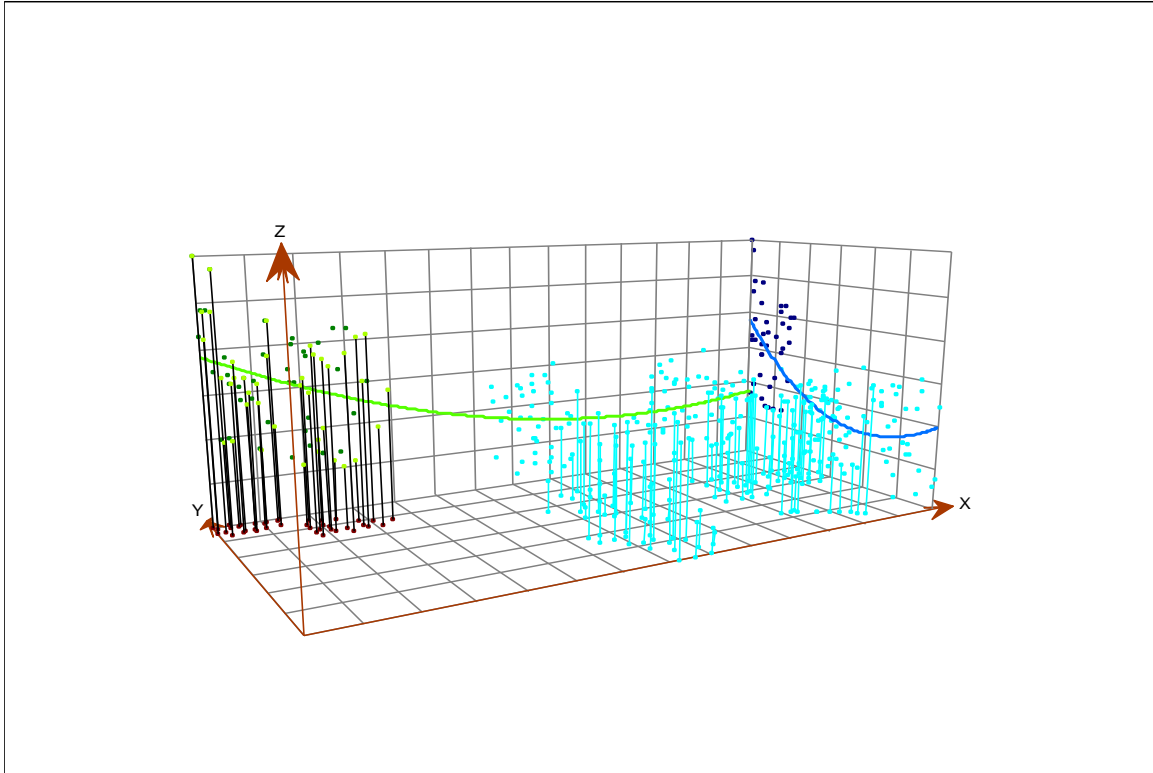
Θεωρώντας δεδομένο ότι δεν γνωρίζουμε τι θα συμβεί στο μέλλον, με την Τεχνική Ανάλυση αναζητούμε πιθανές συμπεριφορές στο παρελθόν που μπορεί να μας βοηθήσουν ώστε να εφαρμόσουμε μια στρατηγική, βασιζόμενοι στο σκεπτικό ότι οι συμπεριφορές τείνουν να επαναλαμβάνονται με παρόμοιο τρόπο.

Trend Analysis



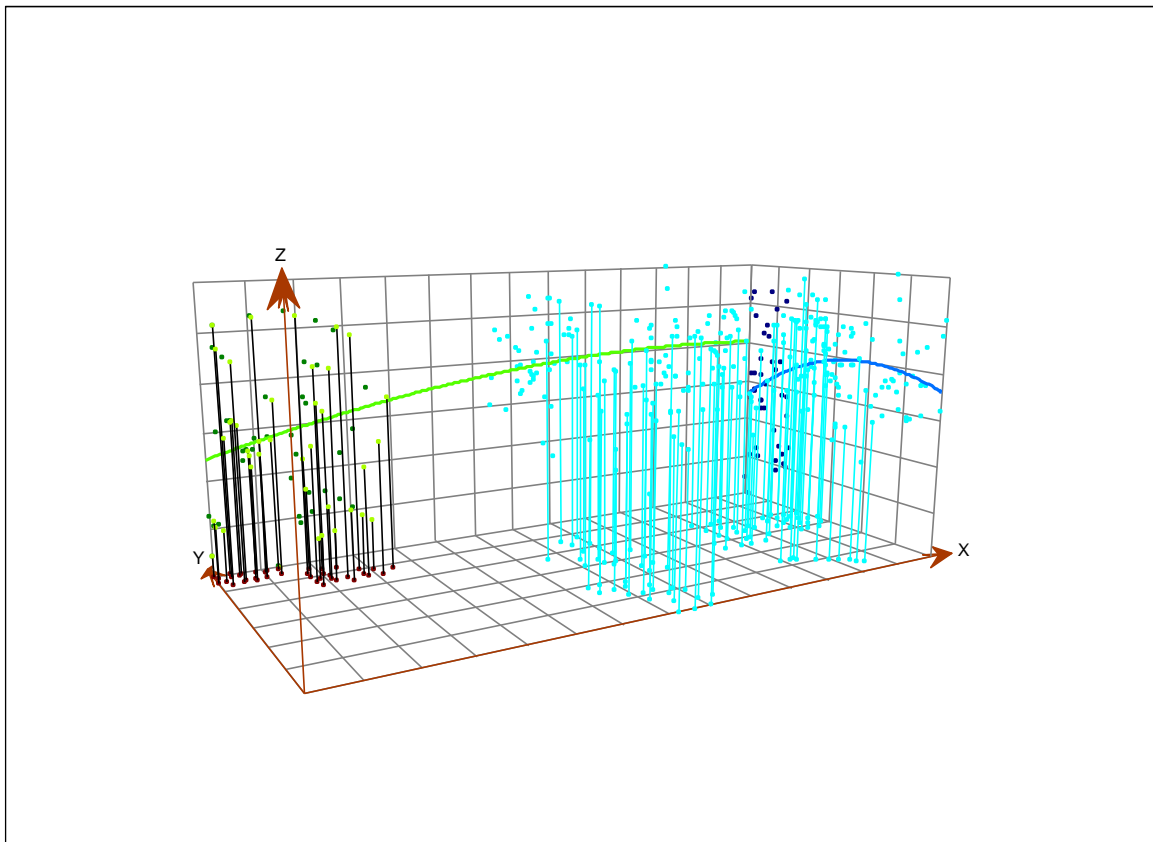
Dataset : Points Attribute: SAND_2

Trend Analysis



Dataset : Points Attribute: CLAY

Trend Analysis



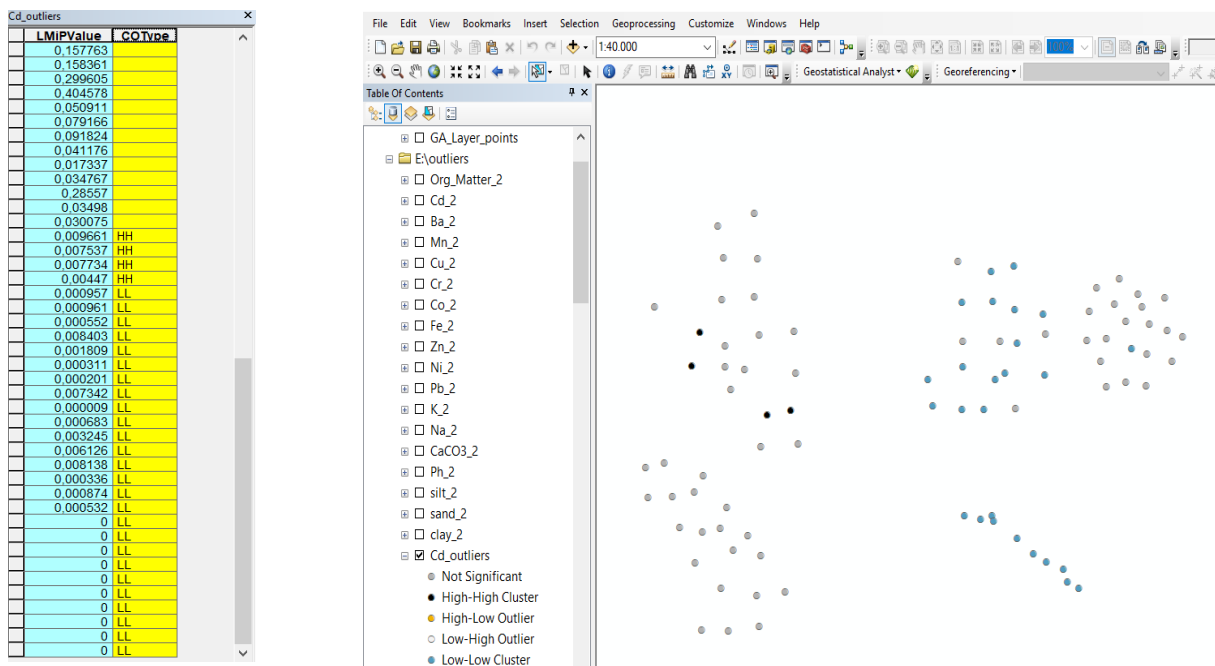
Dataset : Points Attribute: CLAY_2

Στο σημείο αυτό θα παρατηρήσουμε τις μεταβολές των τιμών της αργίλου (clay,sand_2) και άμμου (sand,clay_2) σε περιβάλλον Arc Map. Αυτό που παρατηρούμε για τη μεταβλητή της αργίλου είναι ότι η πράσινη

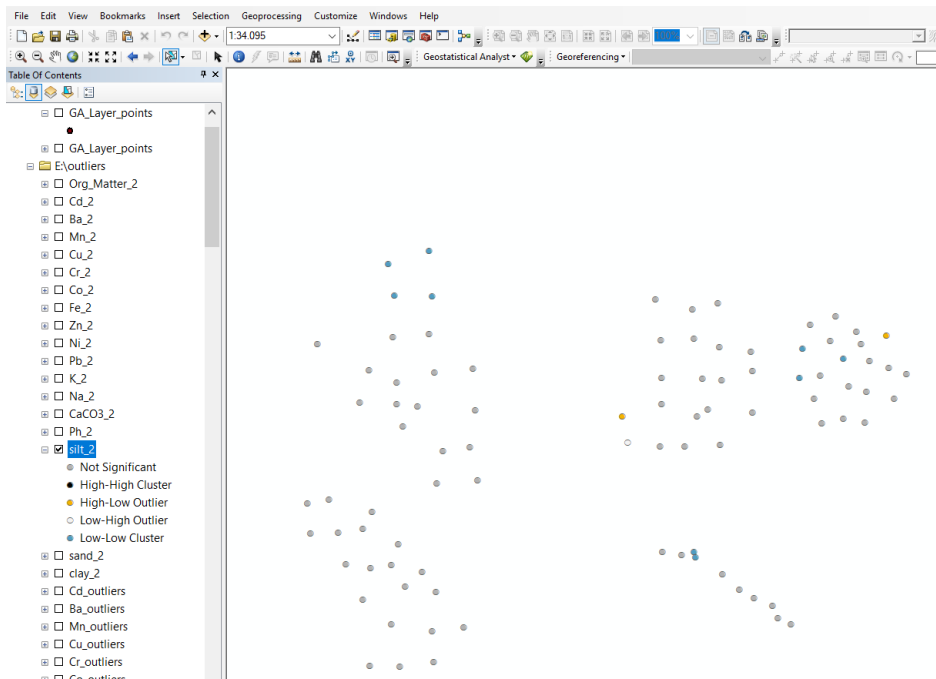
γραμμή είναι η καμπύλη που μας δείχνει την τάση μεταβολής των τιμών της αργίλου όπου παρατηρούμε πτωτική τάση. Η μπλε γραμμή είναι η καμπύλη που δείχνει την πτωτική τάση μεταβολής που έχουν οι τιμές της αργίλου από τον βορρά προς τον νότο. Αντίστοιχα αυτό που παρατηρούμε για την μεταβλητή της άμμου είναι πράσινη γραμμή είναι η καμπύλη που μας δείχνει την τάση μεταβολής των τιμών της άμμου όπου παρατηρούμε ανοδική τάση. Η μπλε γραμμή είναι η καμπύλη που δείχνει την ανοδική τάση μεταβολής που έχουν οι τιμές της άμμου από τον βορρά προς τον νότο.

3.1.5. ΔΙΕΡΕΥΝΗΣΗ ΧΩΡΙΚΩΝ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

Η ασυνήθιστα μικρές ή μεγάλες τιμές, απομακρυσμένες από το κύριο σώμα των δεδομένων ίσως να οφείλονται σε λάθος καταγραφή, ή να κρύβουν χρήσιμες πληροφορίες. Το ενδοτεταρτημοριακό εύρος (IQR) δεν επηρεάζεται από πιθανές ακραίες τιμές που μπορεί να υπάρχουν στα δεδομένα. Σε περιβάλλον Arc Map εάν μία από τις ακραίες τιμές (outlier), είναι ένα σφάλμα, θα πρέπει να διορθωθούν ή να αφαιρεθούν από την ανάλυση. Εάν μια ακραία τιμή αντανακλά έγκυρα δεδομένα, μπορεί να εκτελέσουμε την ανάλυση με και χωρίς την ακραία τιμή για να δούμε πόσο είναι και πόσο επηρεάζουν τα αποτελέσματα. Προκειμένου να διαπιστώσουμε αν υπάρχουν χωρικά ακραίες τιμές στην περιοχή του Ασπροπύργου (και στην παραλία) και της Ελευσίνας για τις μεταβλητές χρησιμοποιούμε το ArcToolbox Window.



Για τη μεταβλητή Cd δεν υπάρχουν τέτοιου είδους σημεία (outliers), πράγμα που σημαίνει ότι γειτονικά σημεία έχουν παρόμοιες τιμές για τη μεταβλητή Cd.



LMPValue	COType
0.971656	
0.969449	
0.645924	
0.965743	
0.872892	
0.948335	
0.936088	
0.367309	
0.979738	
0.113074	
0.571905	
0.557291	
0.78666	
0.844599	
0.994018	
0.835082	
0.800786	
0.799139	
0.597921	
0.543118	
0.822104	
0.420261	
0.886227	
0.336267	
0.902487	
0.850502	
0.755979	
0.628921	
0.996527	
0.893532	
0.969707	
0.009356	HL
0.003002	HL
0.038767	LH
0.035294	LL
0.01153	LL
0.029782	LL
0.000089	LL
0.001203	LL
0	LL
0	LL
0.002552	LL
0.002275	LL

Για τη μεταβλητή silt_2 υπάρχουν τέτοιου είδους σημεία (outliers), πράγμα που σημαίνει ότι γειτονικά σημεία δεν έχουν παρόμοιες τιμές για τη μεταβλητή silt_2.

3.1.6.ΔΙΕΡΕΥΝΗΣΗ ΓΙΑ ΤΗΝ ΥΠΑΡΞΗ ΧΩΡΙΚΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ (spatial autocorrelation) ΚΑΙ ΕΝΤΟΠΙΣΜΟΣ ΑΠΟΣΤΑΣΗΣ ΣΤΗΝ ΟΠΟΙΑ ΕΜΦΑΝΙΖΕΤΑΙ

Η χωρική αυτοσυσχέτιση έχει εφαρμογές σε μια ποικιλία προβλημάτων τόσο από τις κοινωνικές επιστήμες. Κοινό χαρακτηριστικό πρόβλημα της χωρικής στατιστικής είναι ότι τα υπό μελέτη δεδομένα απαρτίζονται από μεταβλητές, οι παρατηρήσεις των οποίων αναπαρίστανται στο χώρο. Αυτές λοιπόν οι παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους, όπως επιτάσσει η κλασική στατιστική (Κάνοργλου, 2001).

Η χωρική αυτοσυσχέτιση, εστιάζει το ενδιαφέρον της στον βαθμό που η τιμή μιας μεταβλητής σε κάποια θέση είναι παρόμοια με τις τιμές της ίδιας μεταβλητής, που βρίσκονται γεωγραφικά κοντά της (Goodchild, 1987).

Δηλαδή, η συσχέτιση μας δείχνει αν υπάρχει εξάρτηση μεταξύ των τιμών μιας συγκεκριμένης μεταβλητής σε εγγύς περιοχές. Όπως αναφέρει άλλωστε και ο καθηγητής Waldo Tobler (1970) στον πρώτο νόμο της γεωγραφίας, τα πάντα σχετίζονται μεταξύ τους, αλλά τα κοντινά πράγματα σχετίζονται περισσότερο από τι τα απομακρυσμένα (Tobler, 1970).

Οι δείκτες που χρησιμοποιεί η χωρική στατιστική αφορούν σε δείκτες που εξετάζουν είτε τη χωρική διασπορά της θέσης των σημείων είτε τη χωρική διασπορά των τιμών χαρακτηριστικών της κατανομής. Οι συνήθεις τεχνικές που χρησιμοποιούνται στην πρώτη περίπτωση είναι:

- i. Η μέθοδος ανάλυσης καννάβου και της απόστασης μεταξύ σημείων,
- ii. Η δεύτερη οι δείκτες χωρικής αυτοσυσχέτισης Moran's I και G - Statistic.

Παράλληλα, γίνεται διάκριση:

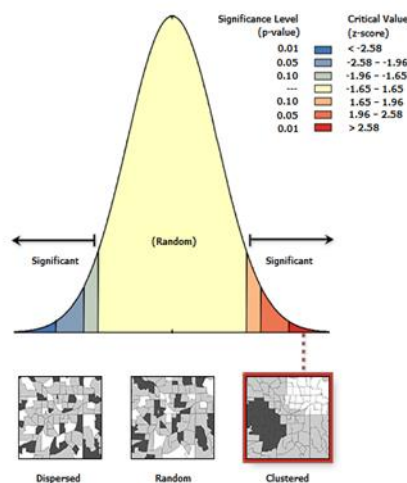
- i. Σε γενικευμένους δείκτες (Global Statistics)
- ii. Σε τοπικούς δείκτες (Local Statistics).

Οι μεν πρώτοι εστιάζουν στην αναζήτηση σχηματισμού κάποιου προτύπου σε όλη την περιοχή μελέτης, οι δε δεύτεροι ασχολούνται με τις τοπικές διακυμάνσεις των τιμών.

Όσον αφορά στη χωρική στατιστική ανάλυση δημογραφικών στοιχείων, αυτά επιπίπτουν στο πεδίο των ασυνεχών κατανομών επιφανειών, καθώς αντιστοιχεί μία τιμή σε κάθε χωρική ενότητα, και αυτό που ενδιαφέρει είναι η ανάλυση της χωρικής διασποράς των τιμών των πληθυσμιακών μεταβλητών και η εξέταση πιθανού προτύπου τους. Κατάλληλος δείκτης χωρικής στατιστικής ανάλυσης είναι ο Moran's I τόσο στη γενικευμένη όσο και στην τοπική του εκδοχή.

Εφαρμογή Spatial Autocorrelation (Global Moran's I)

Προκειμένου, να διαπιστώσουμε εάν υπάρχει χωρική αυτοσυσχέτιση μιας μεταβλητής, εξετάζουμε το Γενικευμένο δείκτη Moran's I (Global), που ορίζεται ως το μέτρο της συσχέτισης μεταξύ των γειτονικών παρατηρήσεων και παίρνει τιμές μεταξύ [-1,1]. Οι θετικές τιμές αντιπροσωπεύουν τη θετική χωρική αυτοσυσχέτιση, ενώ το αντίστροφο ισχύει για τις αρνητικές τιμές. Η μηδενική τιμή του δείκτη δεν αντιπροσωπεύει καμία χωρική αυτοσυσχέτιση. Οι υπολογισμοί του δείκτη χωρικής αυτοσυσχέτισης αρχικά αναφέρονται στη default επιλογή της απόστασης (distance band =387,0439 από τα σημεία δειγματοληψίας για την περιοχή Ασπροπύργου και distance band=538,0584 από τα σημεία δειγματοληψίας για την περιοχή Ελευσίνας) και κατόπιν για διάφορες τιμές της απόστασης (300 m, 400m, 500m, 600m, 700m, 800 m και 900m).

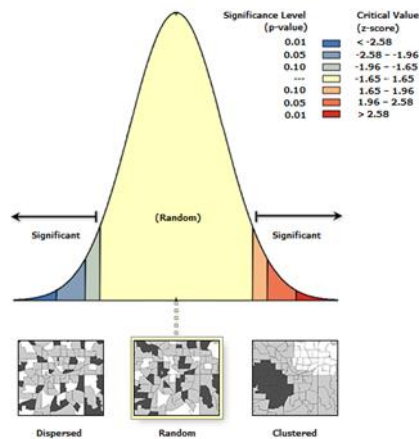


Μη Ασπροπύργου

Moran's Index:	0,445045
Expected Index:	-0,020408
Variance:	0,011054
z-score:	4,426976
p-value:	0,000010
Distance Threshold:	387,0439
Input Field:	MN

Παρατηρούμε ότι ο δείκτης Moran's I είναι θετικός με υψηλή τιμή, πράγμα που σημαίνει ότι υπάρχει υψηλή θετική χωρική αυτοσυσχέτιση, δηλαδή τα γειτονικά σημεία έχουν παρόμοιες τιμές για τη μεταβλητή Mn (συγκεντρωμένες υψηλές τιμές και συγκεντρωμένες χαμηλές τιμές → υπάρχει κατάταξη –clustering τιμών). Επίσης το p-value είναι σχεδόν μηδενικό, και το z_score πολύ μεγάλο, πράγμα που σημαίνει ότι υπάρχουν στατιστικά σημαντικές αποδείξεις εναντίον της μηδενικής υπόθεσης (H_0 : δεν υπάρχει χωρική αυτοσυσχέτιση για τη μεταβλητή Mn, δηλαδή ο δείκτης Moran's I είναι μηδενικός ή ότι το πρότυπο της χωρικής κατανομής είναι τυχαίο → αυτή η υπόθεση απορρίπτεται). Συνεπώς δεχόμαστε την εναλλακτική υπόθεση H_1 : Υπάρχει χωρική αυτοσυσχέτιση και η κατανομή που ακολουθείται σίγουρα δεν είναι το τυχαίο πρότυπο. Οι διαπιστώσεις αυτές οπτικοποιούνται και στο παραπάνω γράφημα.

Επαναλαμβάνουμε την ίδια διαδικασία και για τη μεταβλητή clay.



clay Ελευσίνας

Moran's Index:	0,040953
Expected Index:	-0,025641
Variance:	0,008765
z-score:	0,711297
p-value:	0,476900
Distance Threshold:	550,0000 Meters
Input Field:	CLAY

Παρατηρούμε ότι ο δείκτης Moran's I είναι θετικός, πράγμα που σημαίνει ότι υπάρχει θετική χωρική αυτοσυσχέτιση, δηλαδή τα γειτονικά σημεία έχουν παρόμοιες τιμές για τη μεταβλητή clay, ομοιότητα (συγκεντρωμένες υψηλές τιμές και συγκεντρωμένες χαμηλές τιμές → υπάρχει κατάταξη –clustering τιμών). Επίσης το p-value είναι υψηλό, και το z_score μικρό, πράγμα που σημαίνει ότι υπάρχουν στατιστικά σημαντικές αποδείξεις εναντίον της μηδενικής υπόθεσης (H_0 : δεν υπάρχει χωρική αυτοσυσχέτιση για τη μεταβλητή clay, δηλαδή ο δείκτης Moran's I είναι μηδενικός ή ότι το πρότυπο της χωρικής κατανομής είναι τυχαίο). Συνεπώς δεχόμαστε την εναλλακτική υπόθεση H_1 : Υπάρχει χωρική αυτοσυσχέτιση και η κατανομή που ακολουθείται σίγουρα δεν είναι το τυχαίο πρότυπο. Οι διαπιστώσεις αυτές οπτικοποιούνται και στο παραπάνω γράφημα.

Για να βρούμε την απόσταση στην οποία εμφανίζεται η μέγιστη χωρική αυτοσυσχέτιση για τις ομαδοποιημένες μεταβλητές, θα υπολογίσουμε που υπάρχει το μέγιστο Moran's I. Χρησιμοποιώντας το ArcToolbox και εφαρμόζοντας το Incremental Spatial Autocorrelation (ISA) για τις δυο περιοχές βρίσκουμε που έχουμε τη μέγιστη χωρική αυτοσυσχέτιση για διάφορες τιμές απόστασης για τις μεταβλητές και ελέγχοντας κάθε φορά τα σημεία τα οποία περιλαμβάνονται μέσα σε κάθε γειτονιά, παίρνουμε τα παρακάτω αποτελέσματα:

Για την περιοχή του Ασπροπύργου:

Cr

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,184459	-0,020408	0,009769	2,072733	0,038197
500,00	0,200765	-0,020408	0,006289	2,788970	0,005288
600,00	0,153387	-0,020408	0,004440	2,608254	0,009101
700,00	0,145739	-0,020408	0,003439	2,833389	0,004606
800,00	0,134107	-0,020408	0,002808	2,915651	0,003549
900,00	0,157709	-0,020408	0,002362	3,665139	0,000247
1000,00	0,125395	-0,020408	0,001961	3,292527	0,000993
1100,00	0,113387	-0,020408	0,001559	3,388785	0,000702
1200,00	0,081942	-0,020408	0,001371	2,763753	0,005714
1300,00	0,071894	-0,020408	0,001215	2,648396	0,008087

Co

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,312610	-0,020408	0,009759	3,371073	0,000749
500,00	0,302009	-0,020408	0,006282	4,067801	0,000047
600,00	0,239337	-0,020408	0,004435	3,900207	0,000096
700,00	0,235089	-0,020408	0,003435	4,359419	0,000013
800,00	0,219625	-0,020408	0,002806	4,531739	0,000006
900,00	0,233366	-0,020408	0,002359	5,224701	0,000000
1000,00	0,229375	-0,020408	0,001959	5,643599	0,000000
1100,00	0,194462	-0,020408	0,001557	5,445140	0,000000
1200,00	0,175986	-0,020408	0,001370	5,306027	0,000000
1300,00	0,202655	-0,020408	0,001213	6,403671	0,000000

Zn

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,276146	-0,020408	0,005997	3,829334	0,000128
500,00	0,294844	-0,020408	0,003865	5,070862	0,000000
600,00	0,280147	-0,020408	0,002728	5,754597	0,000000
700,00	0,295387	-0,020408	0,002115	6,866042	0,000000
800,00	0,273884	-0,020408	0,001729	7,077479	0,000000
900,00	0,233714	-0,020408	0,001454	6,665276	0,000000
1000,00	0,191356	-0,020408	0,001207	6,095376	0,000000
1100,00	0,151407	-0,020408	0,000960	5,544863	0,000000
1200,00	0,157582	-0,020408	0,000845	6,122954	0,000000
1300,00	0,153842	-0,020408	0,000750	6,363002	0,000000

Mn**Global Moran's I Summary by Distance**

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,322299	-0,020408	0,009886	3,446718	0,000567
500,00	0,299877	-0,020408	0,006364	4,014811	0,000059
600,00	0,258759	-0,020408	0,004493	4,164767	0,000031
700,00	0,265628	-0,020408	0,003480	4,849022	0,000001
800,00	0,248039	-0,020408	0,002842	5,035563	0,000000
900,00	0,250386	-0,020408	0,002390	5,539210	0,000000
1000,00	0,230571	-0,020408	0,001984	5,634101	0,000000
1100,00	0,205808	-0,020408	0,001577	5,695781	0,000000
1200,00	0,155006	-0,020408	0,001388	4,708746	0,000002
1300,00	0,150194	-0,020408	0,001229	4,866241	0,000001

Cr 2**Global Moran's I Summary by Distance**

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,090503	-0,020408	0,009784	1,121307	0,262157
500,00	0,101530	-0,020408	0,006298	1,536489	0,124419
600,00	0,105136	-0,020408	0,004446	1,882728	0,059737
700,00	0,092913	-0,020408	0,003444	1,931101	0,053471
800,00	0,103491	-0,020408	0,002813	2,336214	0,019480
900,00	0,083976	-0,020408	0,002365	2,146342	0,031846
1000,00	0,065478	-0,020408	0,001964	1,938067	0,052615
1100,00	0,064718	-0,020408	0,001561	2,154516	0,031200
1200,00	0,060985	-0,020408	0,001373	2,196242	0,028075
1300,00	0,033344	-0,020408	0,001216	1,541171	0,123275

Co 2

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,315821	-0,020408	0,010044	3,354945	0,000794
500,00	0,306287	-0,020408	0,006465	4,062971	0,000048
600,00	0,270548	-0,020408	0,004565	4,306510	0,000017
700,00	0,270837	-0,020408	0,003535	4,898599	0,000001
800,00	0,240326	-0,020408	0,002887	4,852538	0,000001
900,00	0,232430	-0,020408	0,002428	5,131344	0,000000
1000,00	0,246649	-0,020408	0,002016	5,948022	0,000000
1100,00	0,256505	-0,020408	0,001602	6,917650	0,000000
1200,00	0,227702	-0,020408	0,001410	6,608022	0,000000
1300,00	0,228525	-0,020408	0,001248	7,045142	0,000000

Zn 2

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,021538	-0,020408	0,007177	0,495144	0,620498
500,00	0,026374	-0,020408	0,004623	0,688049	0,491422
600,00	0,117098	-0,020408	0,003263	2,407145	0,016078
700,00	0,114064	-0,020408	0,002529	2,673900	0,007497
800,00	0,091876	-0,020408	0,002067	2,470000	0,013511
900,00	0,024587	-0,020408	0,001738	1,079436	0,280393
1000,00	-0,013003	-0,020408	0,001443	0,194962	0,845423
1100,00	-0,009881	-0,020408	0,001147	0,310795	0,755957
1200,00	-0,004011	-0,020408	0,001010	0,516049	0,605820
1300,00	0,003253	-0,020408	0,000895	0,790783	0,429071

Mn_2

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
400,00	0,408897	-0,020408	0,010240	4,242462	0,000022
500,00	0,419162	-0,020408	0,006591	5,414248	0,000000
600,00	0,379147	-0,020408	0,004654	5,857096	0,000000
700,00	0,382493	-0,020408	0,003604	6,711609	0,000000
800,00	0,364509	-0,020408	0,002943	7,095102	0,000000
900,00	0,367159	-0,020408	0,002475	7,790310	0,000000
1000,00	0,338381	-0,020408	0,002055	7,914531	0,000000
1100,00	0,298885	-0,020408	0,001634	7,900016	0,000000
1200,00	0,246400	-0,020408	0,001437	7,038036	0,000000
1300,00	0,230929	-0,020408	0,001273	7,045329	0,000000

Από τους παραπάνω πίνακες, παρατηρώ ότι όσο αυξάνεται η τιμή της απόστασης δηλαδή όσο διευρύνονται τα όρια της γειτονιάς, τόσο ο δείκτης Moran's I μειώνεται δηλαδή η χωρική αυτοσυσχέτιση εξασθενεί οπότε, η μέγιστη τιμή και για τις μεταβλητές, εμφανίζεται στην μικρότερη τιμή της απόστασης.

Μεταβλητές	Απόσταση	Moran's I
Cr	500	0,2
Co	400	0,31
Zn	700	0,295
Mn	400	0,32
Cr_2	600	0,105
Co_2	400	0,315
Zn_2	500	0,26
Mn_2	500	0,42

Πίνακας 8

Το ίδιο ισχύει και για την περιοχή της Ελευσίνας μέγιστη τιμή της χωρικής αυτοσυσχέτισης εμφανίζεται στην μικρότερη τιμή της απόστασης.

Μεταβλητές	Απόσταση	Moran's I
Ph	600	0,42
Cr	600	0,23
Ni	600	0,227
Co	600	0,29
Zn	600	0,11
Mn	600	0,41
Ph_2	600	0,208
Cr_2	600	0,07
Ni_2	600	0,227
Co_2	600	0,18
Zn_2	600	0,107
Mn_2	600	0,067

Πίνακας 9

3.1.7.ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ Kriging

Σε περιβάλλον ArcMap και με την βοήθεια του Geostatistical tool εφαρμόζουμε την μέθοδο Ordinary Kriging για τις μεταβλητές χρησιμοποιώντας το θεωρητικό σφαιρικό και εκθετικό μοντέλο με ισοτροπία και ανισοτροπία για το του εύρος τιμών των σημείων. Στο τέλος κάθε εφαρμογής εμφανίζεται ένα πίνακας που μας δείχνει τα αποτελέσματα από την χωρική παρεμβολή με τα δεδομένα και τις αλλαγές που έχουν γίνει με τις παραμέτρους που έχουμε επιλέξει. Ο πίνακας περιλαμβάνει τα εξής αποτελέσματα:

Mean Error (M.E): Μέσο Σφάλμα.

Root Mean Square Error (R.M.S.E): Τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος.

Standardized Mean Error (S.M.E): Μέσο τυποποιημένο σφάλμα.

Average Standard Error (A.S.E): Μέγιστο τυπικό σφάλμα.

Standardized Root Mean Square Error (S.R.M.S.E): Τυποποιημένη τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος.

Με βάση τον παραπάνω πίνακα που θα εμφανίζεται στο τέλος της εφαρμογής αυτομάτως τα συμπεράσματα μας θα αναλύονται με βάση κάποιων κριτηρίων αξιολόγησης των αποτελεσμάτων. Για την αξιολόγηση των αποτελεσμάτων της χωρικής παρεμβολής, και τον προσδιορισμό της ακρίβειας για την πρόβλεψη των τιμών σε άγνωστες θέσεις από τα μοντέλα που εφαρμόστηκαν, χρησιμοποιήθηκε η μέθοδος cross-validation.

Στην μέθοδο αυτή αφαιρείται κάθε φορά ένα σημείο από το σύνολο των δεδομένων και χρησιμοποιούνται τα υπόλοιπα για την πρόβλεψη της τιμής στο σημείο αυτό. Η επανάληψη της διαδικασίας για το σύνολο των τιμών μας παρέχει την δυνατότητα της σύγκρισης των εκτιμώμενων με τις πραγματικές τιμές για κάθε χωρική θέση, και την εκτίμηση της ακρίβειας των προβλέψεων με την μορφή των διαφορών πραγματικών-εκτιμώμενων και τετραγωνισμένων διαφορών.

Για την αξιολόγηση των αποτελεσμάτων θα πρέπει:

Τα **M.E** και **S.M.E** να είναι όσο το δυνατό πιο κοντά στο μηδέν.

Το **R.M.S.E** όσο πιο μικρό γίνεται γεγονός το οποίο σημαίνει ότι οι προβλέψεις είναι αμερόληπτες και κοντά στις πραγματικές τιμές αντίστοιχα.

Οι τιμές του **A.S.E** χρησιμοποιούνται για να αξιολογήσουν την μεταβλητότητα των προβλέψεων από τις μετρημένες τιμές. Επομένως το **A.S.E** πρέπει να είναι ίδιο με το **R.M.S.E** προκειμένου να αξιολογηθεί σωστά η μεταβλητότητα των προβλέψεων.

Αν η τιμή του **A.S.E** είναι μεγαλύτερη από το **R.M.S.E** σημαίνει ότι οι τιμές των προβλέψεων είναι υπερεκτιμημένες.

Αν αντιθέτως το **A.S.E** είναι μικρότερο από το **R.M.S.E** τότε οι προβλέψεις είναι υποεκτιμημένες.

Οι τιμές του **S.R.M.S.E** θα πρέπει να βρίσκονται κοντά στην μονάδα, ενώ αυτές του **S.M.E** να είναι όσο το δυνατό μικρότερες.

Σε περιβάλλον ArcMap και με την βοήθεια του Geostatistical tool εφαρμόζουμε την μέθοδο Ordinary Kriging για τις μεταβλητές των βαρέων μετάλλων. Για εδαφικές ιδιότητες κυρίως εξετάζουμε σφαιρικό και εκθετικό μοντέλο, ενώ προσπαθούμε και η διακύμανση επί την αρχή (nugget) να είναι αρκετά μικρή σε σχέση με το όριο (sill), εμπειρικά πρέπει να είναι από 1/3 έως 1/4 του sill. Στο τέλος κάθε εφαρμογής εμφανίζεται ένα πίνακας που μας δείχνει τα αποτελέσματα από την χωρική παρεμβολή με τα δεδομένα και τις αλλαγές που έχουν γίνει με τις παραμέτρους που έχουμε επιλέξει.

Για την αξιολόγηση των αποτελεσμάτων της χωρικής παρεμβολής, και τον προσδιορισμό της ακρίβειας για την πρόβλεψη των τιμών σε άγνωστες θέσεις από τα μοντέλα που εφαρμόστηκαν, χρησιμοποιήθηκε η μέθοδος cross-validation.

Στην μέθοδο αυτή αφαιρείται κάθε φορά ένα σημείο από το σύνολο των δεδομένων και χρησιμοποιούνται τα υπόλοιπα για την πρόβλεψη της τιμής στο σημείο αυτό. Η επανάληψη της διαδικασίας για το σύνολο των τιμών μας παρέχει την δυνατότητα της σύγκρισης των εκτιμώμενων με τις πραγματικές τιμές για κάθε χωρική θέση, και την εκτίμηση της ακρίβειας των προβλέψεων με την μορφή των διαφορών πραγματικών-εκτιμώμενων και τετραγωνισμένων διαφορών.

Οι στατιστικές παράμετροι που υπολογίστηκαν από την μέθοδο cross-validation και αποτέλεσαν κριτήρια αξιολόγησης της χωρικής παρεμβολής με χρήση της γεωστατιστικής μεθόδου ordinary kriging ήταν το μέσο σφάλμα (Mean Error) και η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error). Επίσης το μέσο τυπικό σφάλμα (Average Standard Error) και η τυποποιημένη τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Standardized Root Mean Square Error).

Για την αξιολόγηση των αποτελεσμάτων της χωρικής παρεμβολής, και τον προσδιορισμό της ακρίβειας για την πρόβλεψη των τιμών σε άγνωστες θέσεις από τα μοντέλα που εφαρμόστηκαν, χρησιμοποιήθηκε η μέθοδος cross-validation.

Στην μέθοδο αυτή αφαιρείται κάθε φορά ένα σημείο από το σύνολο των δεδομένων και χρησιμοποιούνται τα υπόλοιπα για την πρόβλεψη της τιμής στο σημείο αυτό. Η επανάληψη της διαδικασίας για το σύνολο των τιμών μας παρέχει την δυνατότητα της σύγκρισης των εκτιμώμενων με τις πραγματικές τιμές για κάθε χωρική θέση, και την εκτίμηση της ακρίβειας των προβλέψεων με την μορφή των διαφορών πραγματικών-εκτιμώμενων και τετραγωνισμένων διαφορών.

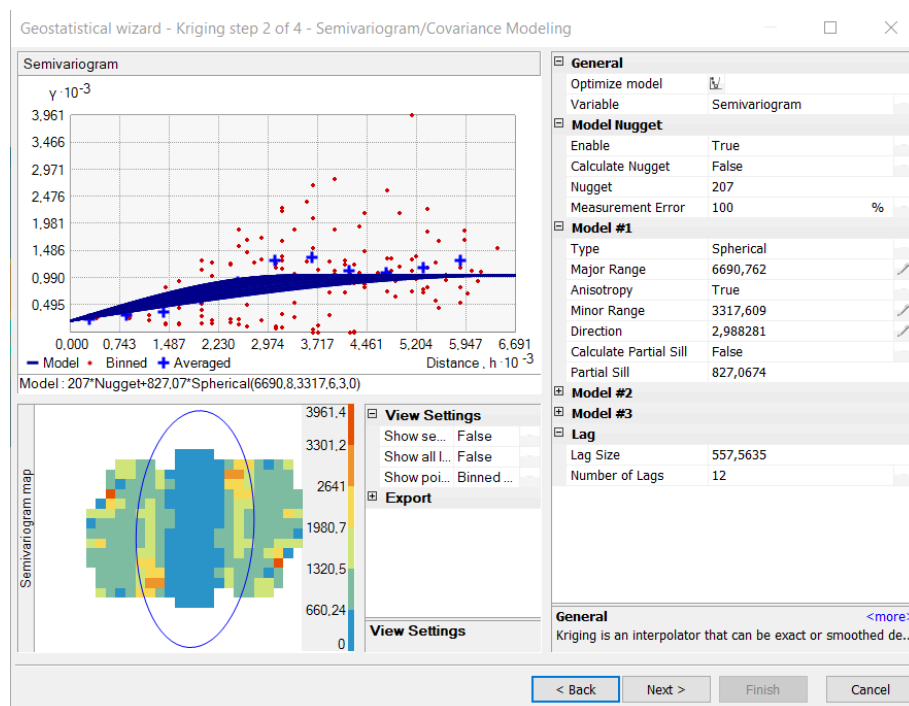
Οι στατιστικές παράμετροι που υπολογίστηκαν από την μέθοδο cross-validation και αποτέλεσαν κριτήρια αξιολόγησης της χωρικής παρεμβολής με χρήση της γεωστατιστικής μεθόδου ordinary kriging ήταν το μέσο σφάλμα (Mean Error) και η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error). Επίσης το μέσο τυπικό σφάλμα (Average Standard Error) και η τυποποιημένη τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Standardized Root Mean Square Error).

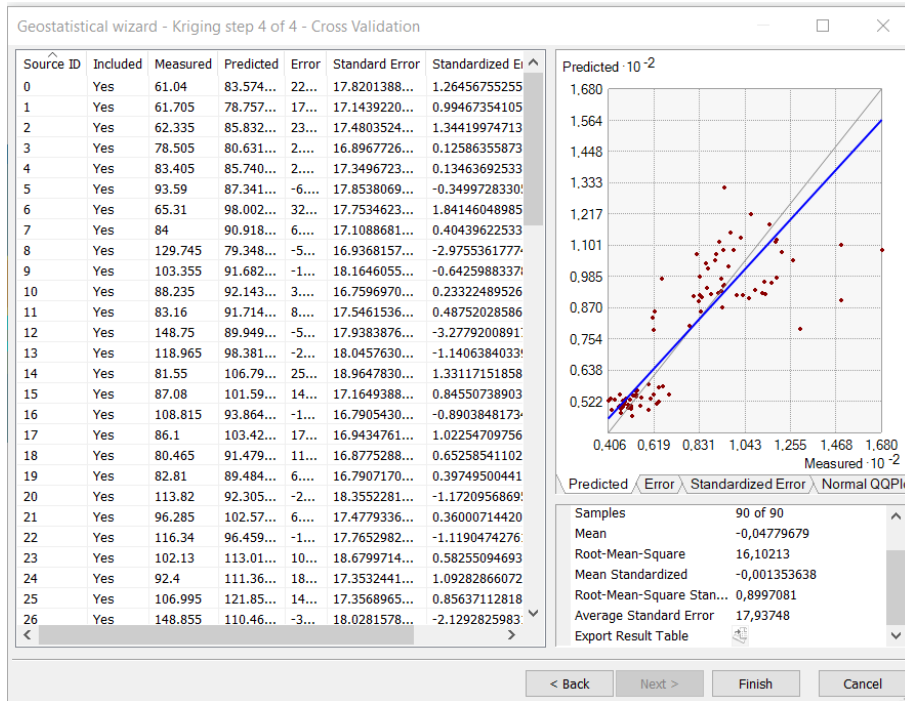
Για την αξιολόγηση των αποτελεσμάτων πρέπει τα ME και SME να είναι όσο το δυνατό πιο κοντά στο μηδέν και το RMSE όσο πιο μικρό γίνεται γεγονός το οποίο σημαίνει ότι οι προβλέψεις είναι αμερόληπτες και κοντά στις πραγματικές τιμές αντίστοιχα. Οι τιμές του ASE χρησιμοποιούνται για να αξιολογήσουν την μεταβλητότητα των προβλέψεων από τις μετρημένες τιμές. Επομένως το ASE πρέπει να είναι ίδιο με το RMSE προκειμένου να αξιολογηθεί σωστά η μεταβλητότητα των προβλέψεων. Αν η τιμή του ASE είναι μεγαλύτερη από το RMSE σημαίνει ότι οι τιμές των προβλέψεων είναι υπερεκτιμημένες. Αν αντιθέτως το ASE είναι μικρότερο από το RMSE τότε οι προβλέψεις είναι υποεκτιμημένες. Οι τιμές του SRMSE θα πρέπει να βρίσκονται κοντά στην μονάδα, ενώ αυτές του SME να είναι όσο το δυνατό μικρότερες.

Το ME μετρά την μεροληψία από την πρόβλεψη και θα πρέπει να είναι κοντά στο μηδέν για αμερόληπτες μεθόδους. Εάν το MSE μετρά την μέση ακρίβεια της πρόβλεψης θα πρέπει να είναι όσο το δυνατόν μικρότερο. Το μοντέλο που εκτελεί καλύτερα θα είναι αυτό με το **μικρότερο MSE**.

Εφαρμογή 1^η

Εφαρμόζουμε την γεωστατιστική μέθοδο Ordinary Kriging για την μεταβλητή Cr επιλέγοντας σφαιρικό μοντέλο με εύρος τιμών (η διακύμανση επί την αρχή (nugget) να είναι αρκετά μικρή σε σχέση με το όριο (sill), εμπειρικά πρέπει να είναι από 1/3 έως 1/4 του sill) 207 και με ανισοτροπία όπου τα αποτελέσματα φαίνονται στα παρακάτω σχήματα.





Με βάση τα αποτελέσματα του πίνακα Cross Validation παρατηρούμε ότι:

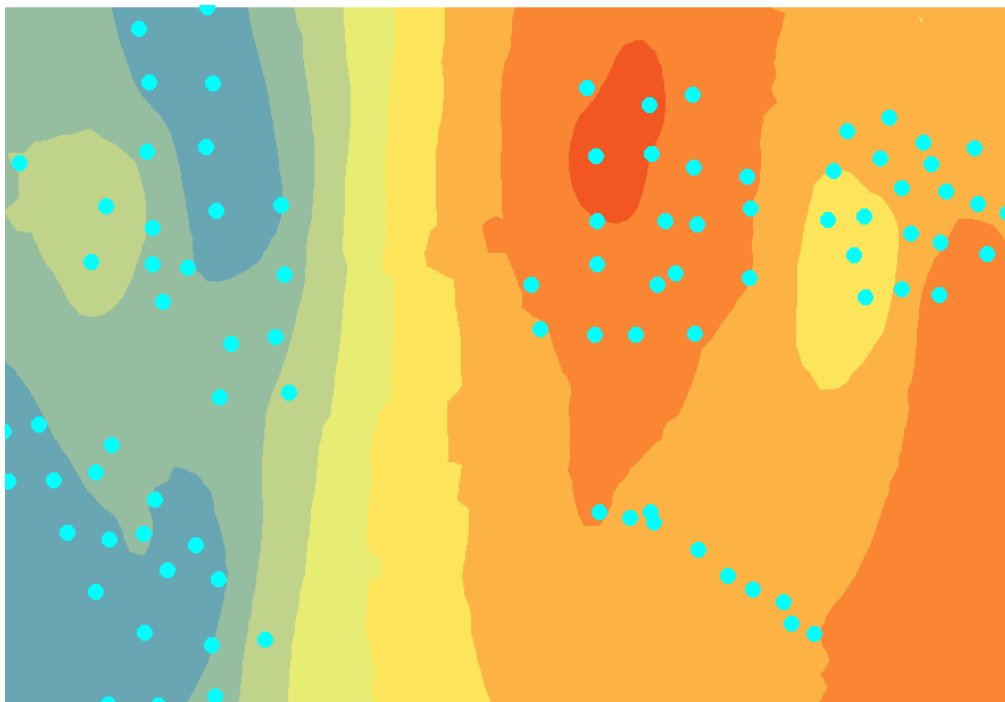
Το μέσο σφάλμα (M.E) και το Μέσο τυποποιημένο σφάλμα (S.M.E) τείνουν στο μηδέν.

Ο αριθμός της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) είναι μεγάλος άρα υπάρχει απόκλιση τιμών.

Ο αριθμός του Μέσου τυπικού σφάλματος (A.S.E) είναι μεγαλύτερος από τον αριθμό της Τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) που σημαίνει ότι οι τιμές των προβλέψεων μας είναι υπερεκτιμημένες.

Ο αριθμός της Τυποποιημένης τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (S.R.M.S.E) είναι κοντά στην μονάδα.

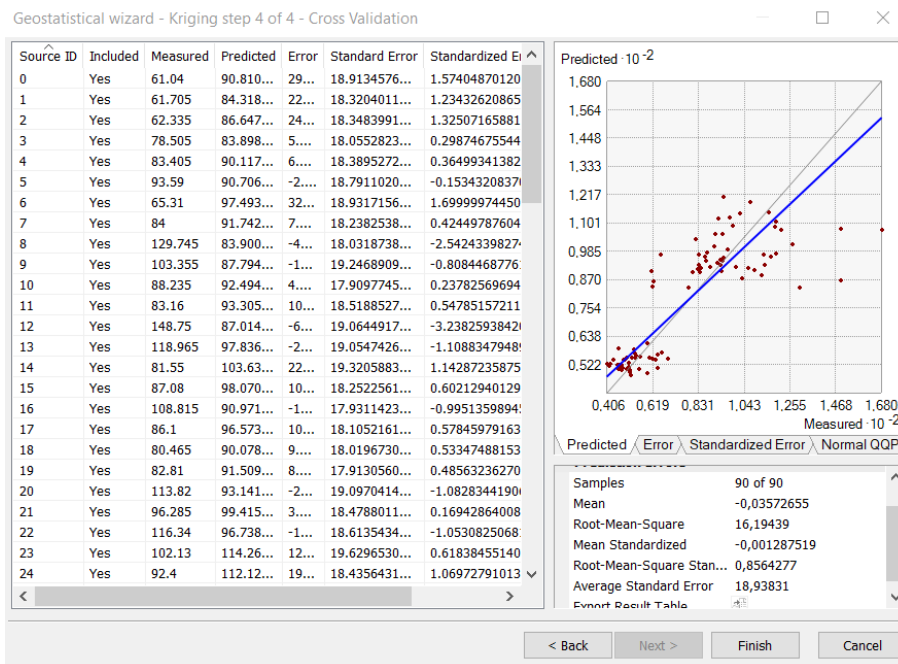
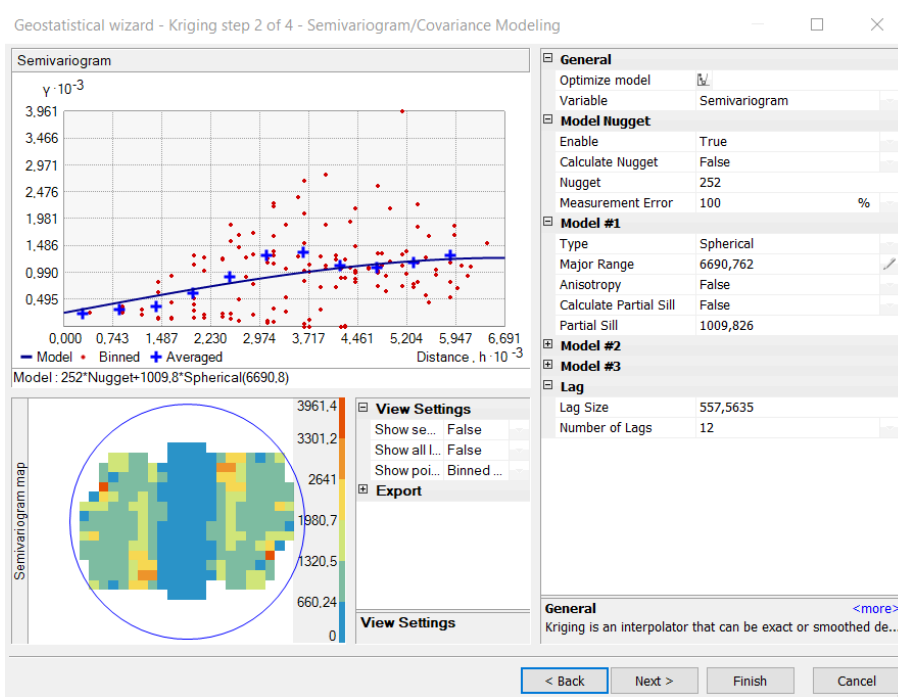
Ο αριθμός του Μέσου τυποποιημένου σφάλματος (S.M.E) είναι πολύ μικρός.



Εικόνα 4: Σφαιρικό μοντέλο με ανισοτροπία

Εφαρμογή 2^η

Εφαρμόζουμε Ordinary Kriging για την μεταβλητή Cr επιλέγοντας σφαιρικό μοντέλο με εύρος τιμών 252 και με ιστροπία όπου τα αποτελέσματα φαίνονται στα παρακάτω σχήματα.



Με βάση τα αποτελέσματα του πίνακα Cross Validation παρατηρούμε ότι:

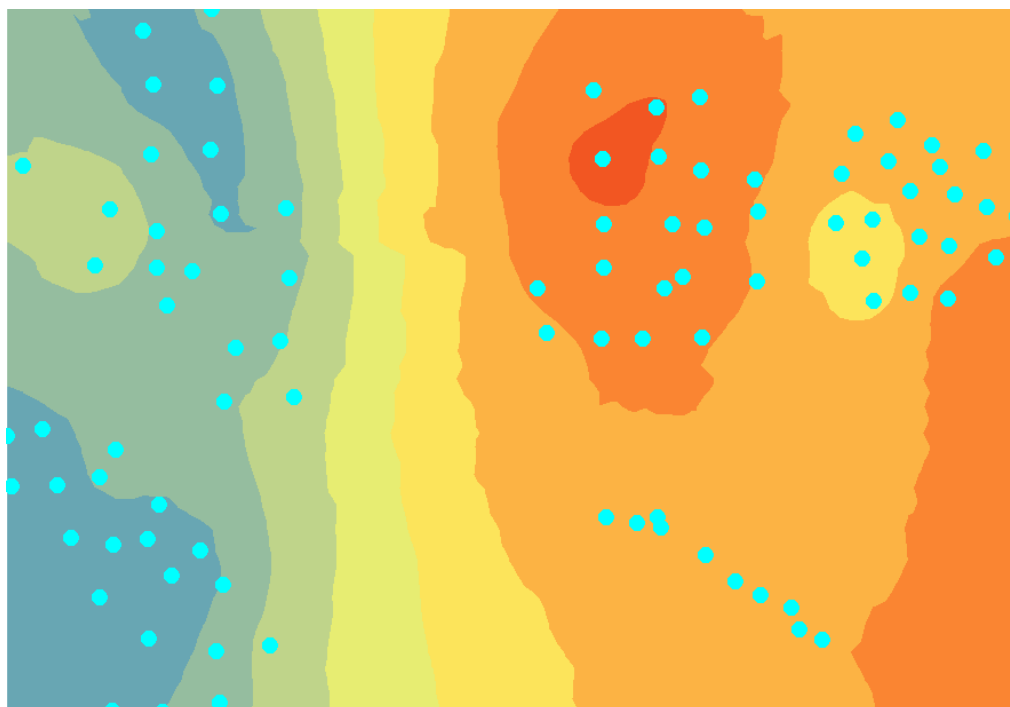
Το μέσο σφάλμα (Μ.Ε) και το Μέσο τυποποιημένο σφάλμα (S.Μ.Ε) τείνουν στο μηδέν.

Ο αριθμός της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.Μ.Σ.Ε) είναι μεγάλος άρα υπάρχει απόκλιση τιμών.

Ο αριθμός του Μέσου τυπικού σφάλματος (A.Σ.Ε) είναι μεγαλύτερος από τον αριθμό της Τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.Μ.Σ.Ε) που σημαίνει ότι οι τιμές των προβλέψεων μας είναι υπερεκτιμημένες.

Ο αριθμός της Τυποποιημένης τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (S.R.M.S.E) είναι κοντά στην μονάδα.

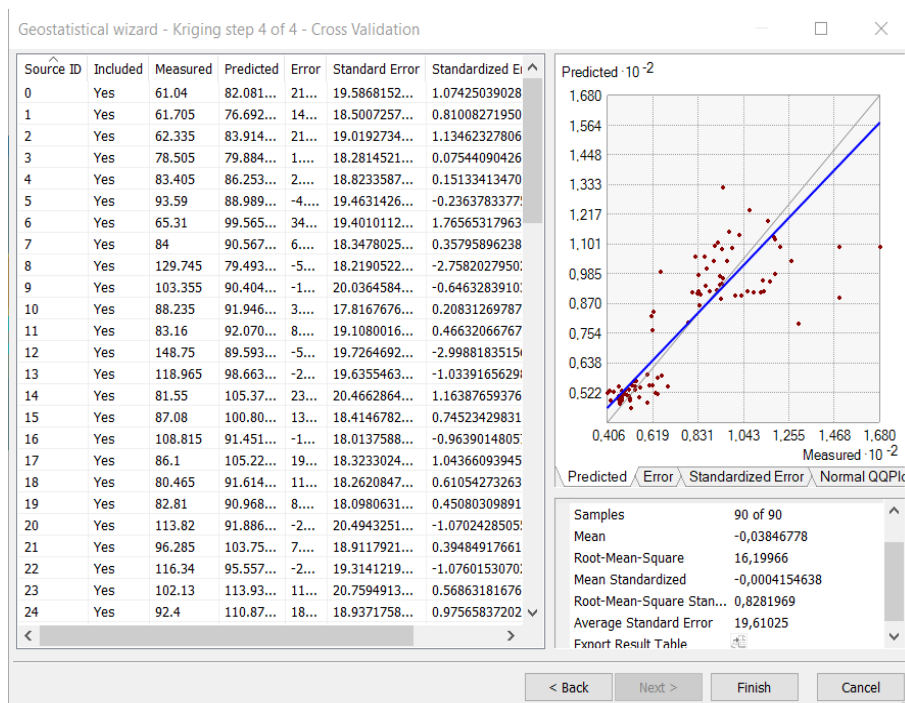
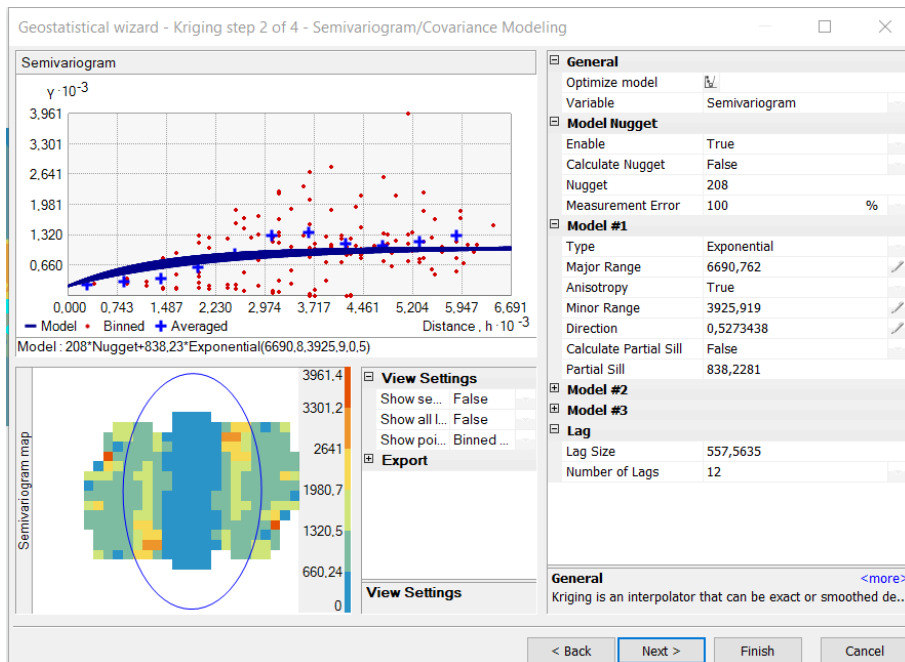
Ο αριθμός του Μέσου τυποποιημένου σφάλματος (S.M.E) είναι πολύ μικρός.



Εικόνα 5: Σφαιρικό μοντέλο με ισοτροπία

Εφαρμογή 3^η

Εφαρμόζουμε Ordinary Kriging για την μεταβλητή Cr επιλέγοντας εκθετικό μοντέλο με εύρος τιμών 208 και με ανισοτροπία όπου τα αποτελέσματα φαίνονται στα παρακάτω σχήματα.



Με βάση τα αποτελέσματα του πίνακα Cross Validation παρατηρούμε ότι:

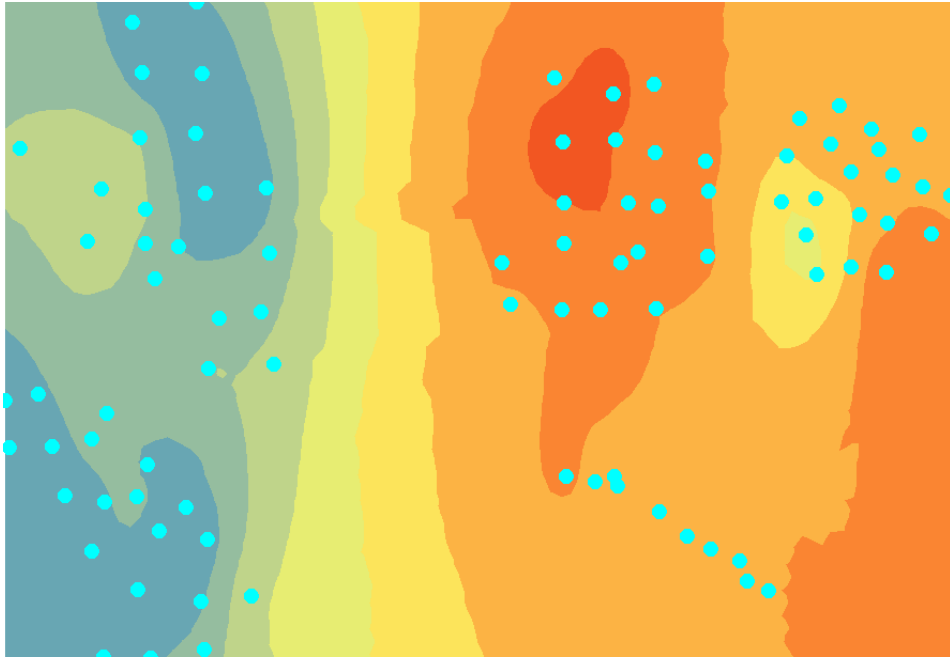
Το μέσο σφάλμα (M.E) και το Μέσο τυποποιημένο σφάλμα (S.M.E) τείνουν στο μηδέν.

Ο αριθμός της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) είναι μεγάλος άρα υπάρχει απόκλιση τιμών.

Ο αριθμός του Μέσου τυπικού σφάλματος (A.S.E) είναι μεγαλύτερος από τον αριθμό της Τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) που σημαίνει ότι οι τιμές των προβλέψεων μας είναι υπερεκτιμημένες.

Ο αριθμός της Τυποποιημένης τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (S.R.M.S.E) είναι κοντά στην μονάδα.

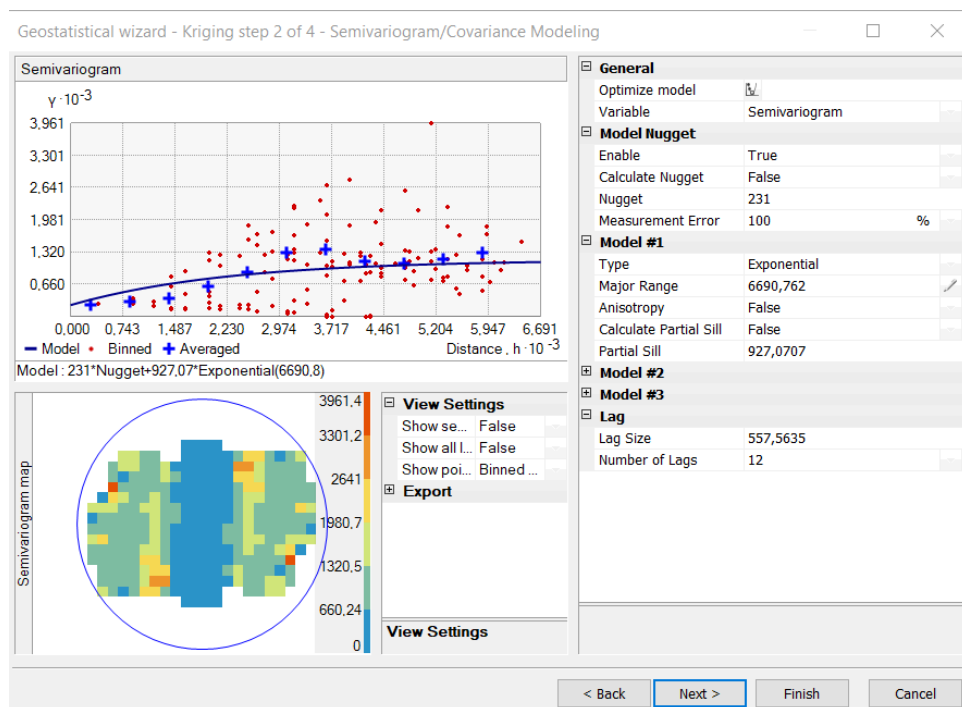
Ο αριθμός του Μέσου τυποποιημένου σφάλματος (S.M.E) είναι πολύ μικρός.

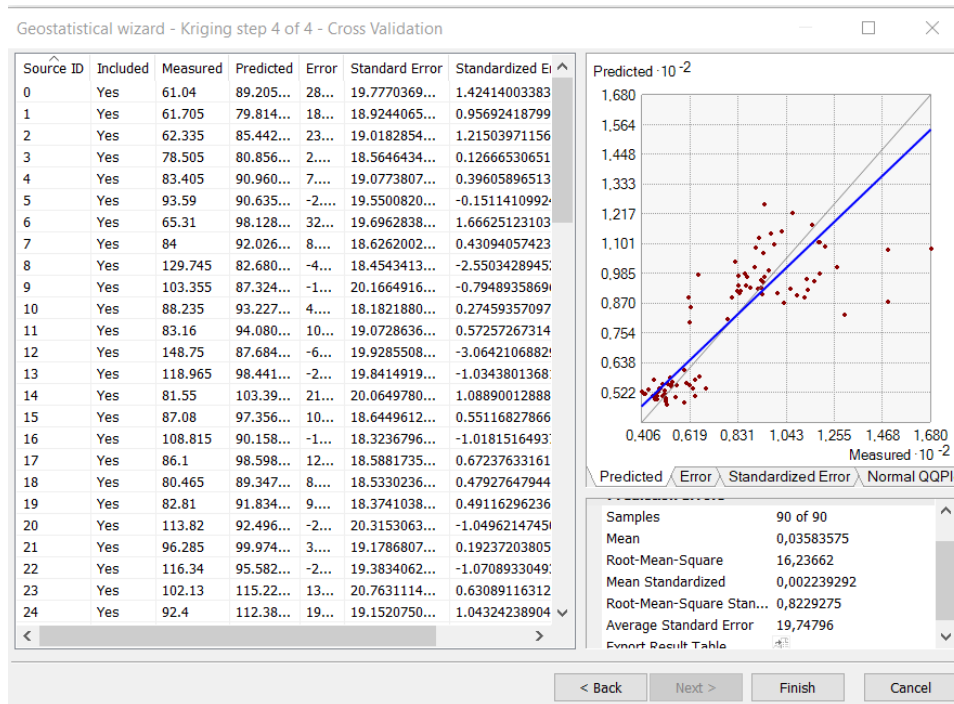


Εικόνα 6: Εκθετικό μοντέλο με ανισοτροπία

Εφαρμογή 4^η

Εφαρμόζουμε Ordinary Kriging για την μεταβλητή Cr επιλέγοντας εκθετικό μοντέλο με εύρος τιμών 231 και με ισοτροπία όπου τα αποτελέσματα φαίνονται στα παρακάτω σχήματα.





Με βάση τα αποτελέσματα του πίνακα Cross Validation παρατηρούμε ότι:

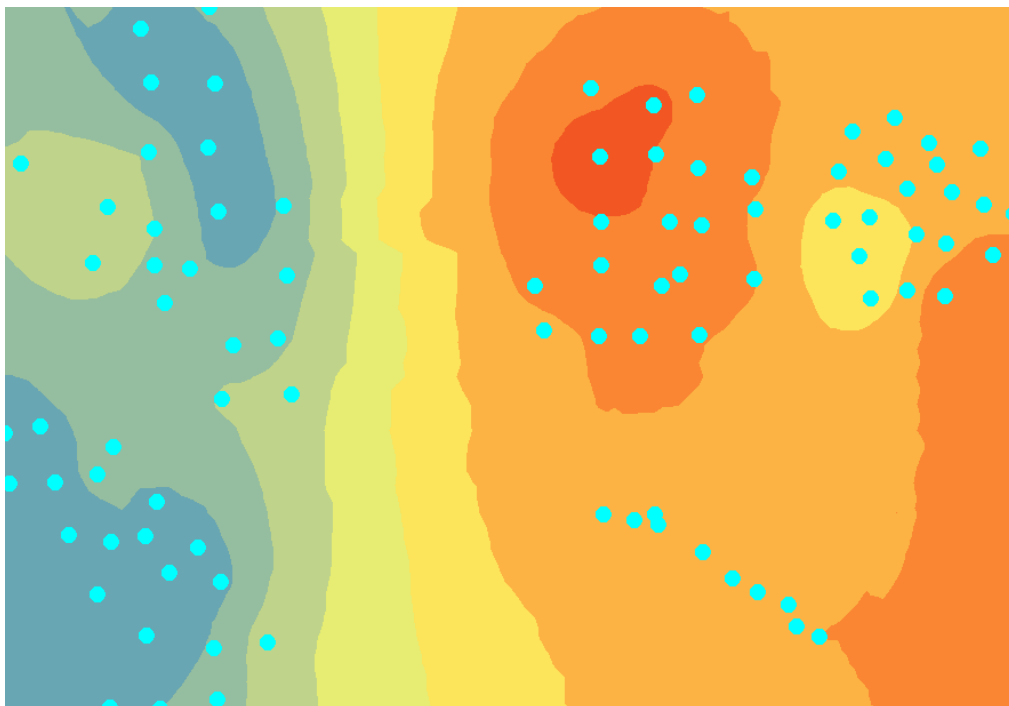
Το μέσο σφάλμα (M.E) και το Μέσο τυποποιημένο σφάλμα (S.M.E) τείνουν στο μηδέν.

Ο αριθμός της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) είναι μεγάλος άρα υπάρχει απόκλιση τιμών.

Ο αριθμός του Μέσου τυπικού σφάλματος (A.S.E) είναι μεγαλύτερος από τον αριθμό της Τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (R.M.S.E) που σημαίνει ότι οι τιμές των προβλέψεων μας είναι υπερεκτιμημένες.

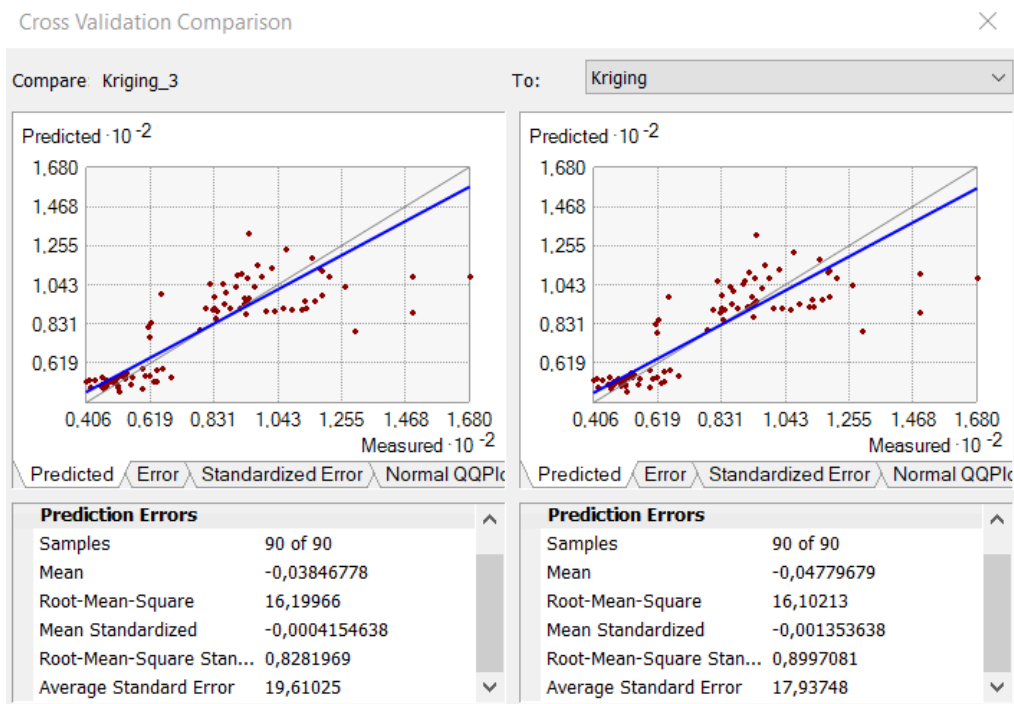
Ο αριθμός της Τυποποιημένης τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (S.R.M.S.E) είναι κοντά στην μονάδα.

Ο αριθμός του Μέσου τυποποιημένου σφάλματος (S.M.E) είναι πολύ μικρός.



Εικόνα 7: Εκθετικό μοντέλο με ισοτροπία

Μία κλασική μέθοδος για την αξιολόγηση χωρικής παρεμβολής σε μια περιοχή είναι τα αποτελέσματα από την χωρική παρεμβολή που εφαρμόζουμε σε μια περιοχή μελέτης και στην συγκεκριμένη περιοχή εξετάζουμε το Cr με σφαιρικό και εκθετικό μοντέλο με ανισοτροπία. Στο σχήμα που ακολουθεί παρατηρούμε τα αποτελέσματα από τους πίνακες Cross Validation με σφαιρικό και εκθετικό μοντέλο.



Εικόνα 8: εκθετικό και σφαιρικό μοντέλο

Συγκεντρωτικός Πίνακας Αποτελεσμάτων Χωρικής Παρεμβολής για τα βαρέα μέταλλα (επιφανειακά και υποεπιφανειακά) σφαιρικού και εκθετικού μοντέλου που εκτελούν καλύτερα με το μικρότερο MSE.

Μεταβλητές	Μοντέλο	Mean (M.E)	Root Mean Square Error (R.M.S.E.)	Mean Standardized Error (M.S.E.)	Root Mean Square Standardized Error (S.R.M.S.E.)	Average Standard Error (A.S.E.)
Cr	Σφαιρικό	-0,0357	16,19	-0,00128	0,856	18,93
	Εκθετικό	-0,038	16,19	-0,00041	0,828	19,6
Pb	Σφαιρικό	0,18	93,47	0,01	0,78	119,7
	Εκθετικό	1,1	93,79	0,015	0,79	118,1
Ba	Σφαιρικό	-4,5	189,7	-0,018	1,18	154,5
	Εκθετικό	-3,14	184,7	-0,01	1,08	166,8
Cu	Σφαιρικό	-0,31	46,6	0,0012	1,45	30,7
	Εκθετικό	0,07	45,3	0,0095	1,27	34,6
Ni	Σφαιρικό	-0,08	18,05	0,0001	2,26	7,88
	Εκθετικό	-0,35	16,6	-0,035	1,93	8,52
Co	Σφαιρικό	-0,031	2,66	-0,008	1,49	1,81
	Εκθετικό	-0,012	2,69	-0,0009	0,92	2,94
Zn	Σφαιρικό	0,175	114,6	-0,0002	0,89	133,86
	Εκθετικό	-0,18	117,9	-0,0022	1,1	109,2
Fe	Σφαιρικό	-5,8	3508,2	0,00052	1,13	3116,3
	Εκθετικό	-6,3	3433,4	0,002	1,11	3204,7
Mn	Σφαιρικό	0,084	80,71	0,007	1,3	62,7
	Εκθετικό	-0,04	79,7	0,0035	1,17	70,5

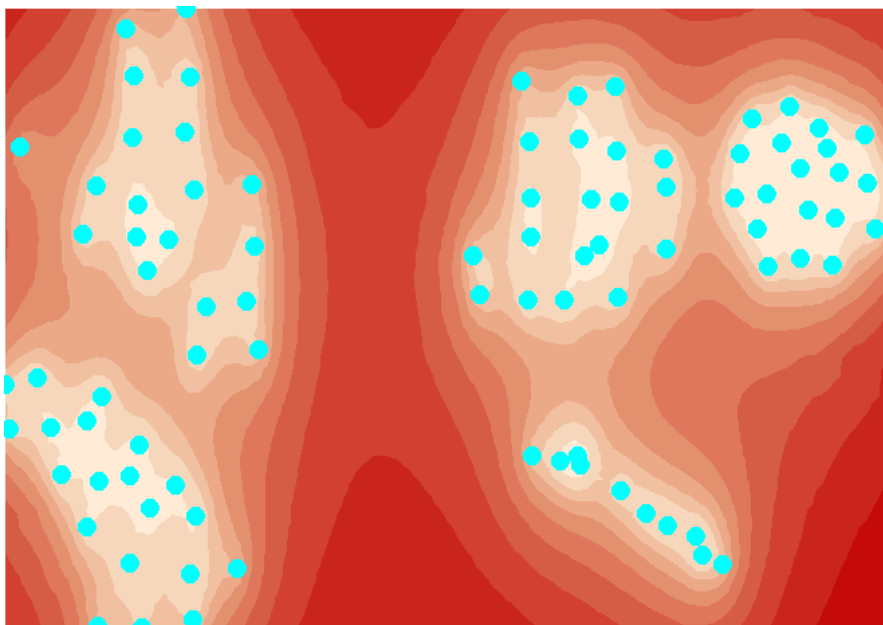
Cr_2	Σφαιρικό	-0,017	20,3	-0,0012	1,068	19,06
	Εκθετικό	0,0017	22,15	0,0052	1,86	12,13
Pb_2	Σφαιρικό	-1,2	48,02	-0,0088	1,02	47,6
	Εκθετικό	-0,18	51,04	0,006	0,89	57,4
Ba_2	Σφαιρικό	2,85	161,1	0,015	1,14	141,1
	Εκθετικό	1,89	159,2	0,008	0,9	174,9
Cu_2	Σφαιρικό	-0,04	27,1	-0,0013	1,6	15,5
	Εκθετικό	-0,004	26,7	0,0065	1,5	17,02
Ni_2	Σφαιρικό	0,25	58,9	0,002	3,59	17,2
	Εκθετικό	-0,49	60,7	-0,006	1,05	57,9
Co_2	Σφαιρικό	4,24	3,2	-0,00099	1,39	2,34
	Εκθετικό	0,04	3,38	0,03	2,61	1,3
Zn_2	Σφαιρικό	0,2	108,05	0,0085	1,28	83,1
	Εκθετικό	-0,9	107,7	-0,006	1,72	61,9
Fe_2	Σφαιρικό	-5,6	3532,5	-0,005	2,01	1778,1
	Εκθετικό	3,5	3582,5	0,0028	1,12	3197,2
Mn_2	Σφαιρικό	-0,28	72,7	-0,00056	1,13	66,04
	Εκθετικό	-0,25	70,81	0,00086	0,99	73,3

Πίνακας 10

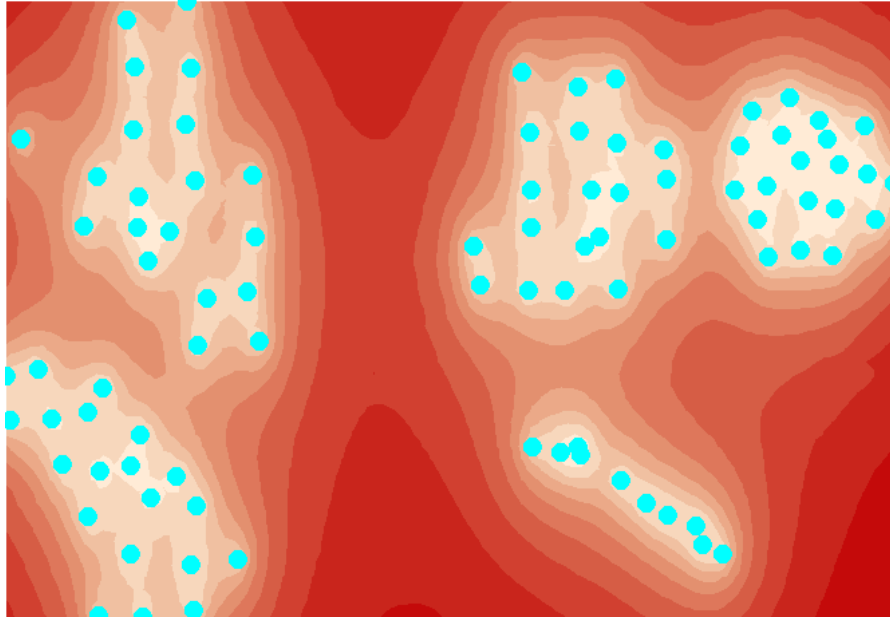
Η τεχνική Kriging της γεωστατιστικής μεθόδου εξάγει και χωρικά εντοπισμένα σφάλματα (πλεονέκτημα έναντι των μη γεωστατιστικών μεθόδων χωρικής επεξεργασίας δεδομένων).

Τα μεγέθη Standardize Error, αναφέρονται σε κανονικοποιημένα αποτελέσματα της μεθόδου kriging και είναι αυτά που χρησιμοποιούμε προκειμένου να έχουμε συγκρίσιμα αποτελέσματα. Επίσης, μπορούμε να πάρουμε εκτίμηση των λαθών πρόγνωσης (error of prediction).

Η διαδικασία για την επίτευξη της συγκεκριμένης τεχνικής λαθών πρόγνωσης προϋποθέτει πρώτα την ολοκλήρωση της διαδικασίας χωρικής επεξεργασίας δεδομένων (Kriging) όπου τα αποτελέσματα φαίνονται στις παρακάτω εικόνες:



Εικόνα 9: Σφαιρικό μοντέλο



Εικόνα 10: Εκθετικό μοντέλο

4. ΑΠΟΤΕΛΕΣΜΑΤΑ-ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία παρουσιάζονται τα αποτελέσματα για την ιδιότητα του επί τοις εκατό ποσοστού των βαρέων μετάλλων στην περιοχή του Θριάσιου πεδίου.

4.1. ΑΠΟΤΕΛΕΣΜΑΤΑ

Η εφαρμογή της γεωστατιστικής ανάλυσης για την πρόβλεψη των επιλεγμένων συγκεντρώσεων βαρέων μετάλλων στην περιοχή μελέτης, μας αποκάλυψε σημαντικές πληροφορίες τόσο για την χωρική διακύμανση των ιδιοτήτων του εδάφους, όσο και για την μέθοδο kriging.

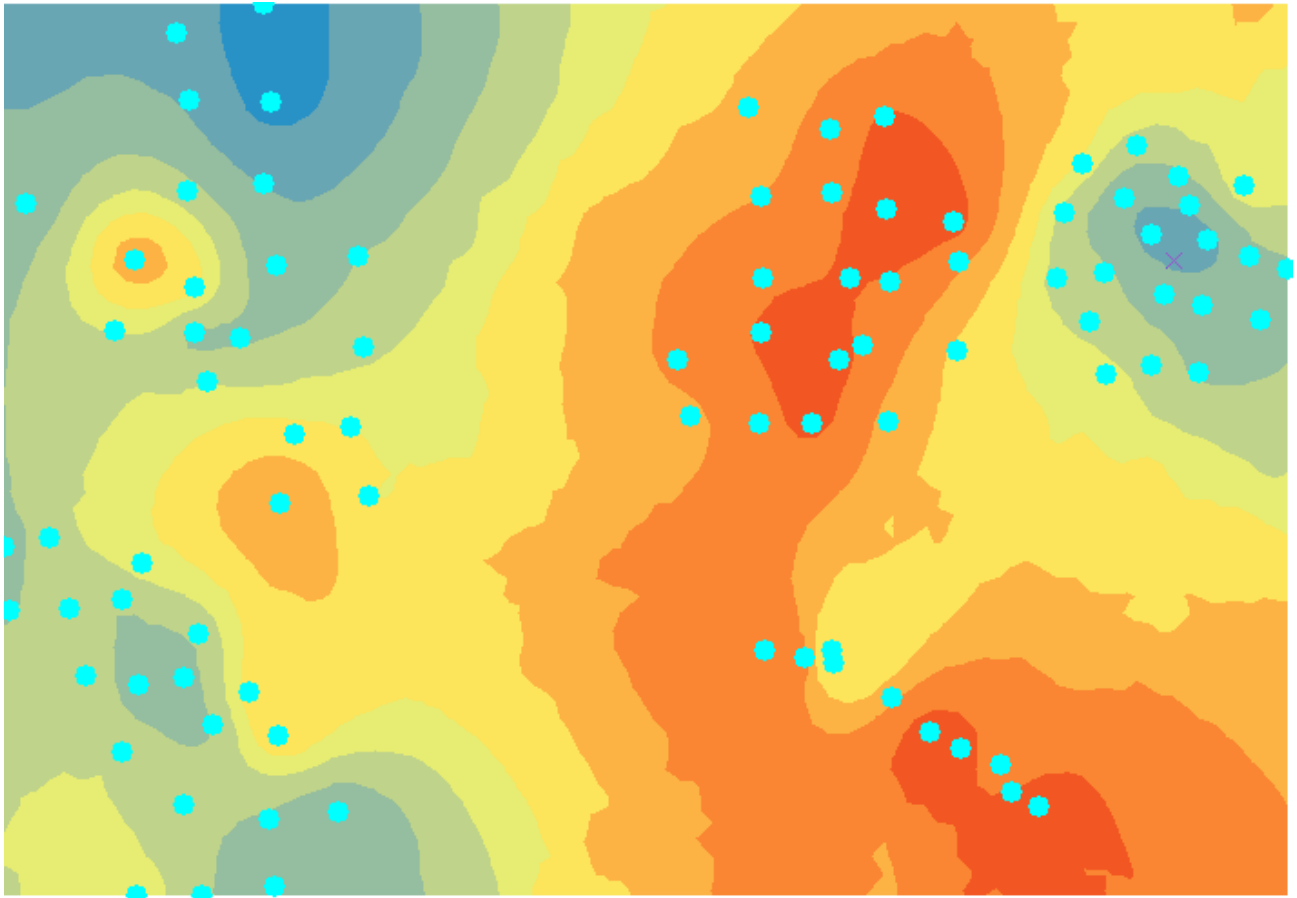
Όπως είδαμε και στην προηγούμενη ενότητα η στατιστική μεθοδολογία του cross validation χρησιμοποιήθηκε για να συγκρίνει την ακρίβεια στις προβλέψεις μεταξύ του εκθετικού και σφαιρικού μοντέλου της γεωστατιστικής μεθόδου Ordinary Kriging και της μη γεωστατιστικής μεθόδου IDW.

Root Mean Square Error (R.M.S.E.)	IDW	Ordinary Kriging	
		Σφαιρικό	Εκθετικό
Cr	175,6	16,19	16,19
Pb	16	93,47	93,79
Ba	96,1	189,7	184,7
Cu	44,1	46,6	45,3
Ni	16,55	18,05	16,6
Co	2,74	2,66	2,69
Zn	120,67	114,6	117,9
Fe	3540,45	3508,2	3433,4
Mn	72,29	80,71	79,7
Cr_2	21,74	20,3	22,15
Pb_2	51,03	48,02	51,04
Ba_2	156,98	161,1	159,2
Cu_2	25,7	27,1	26,7
Ni_2	60,9	58,9	60,7
Co_2	3,41	3,2	3,38
Zn_2	106,25	108,05	107,7
Fe_2	3556,3	3532,5	3582,5
Mn_2	71,6	72,7	70,81
Mean (M.E.)	IDW	Ordinary Kriging	
		Σφαιρικό	Εκθετικό
Cr	-5,92	-0,0357	-0,038
Pb	-0,23	0,18	1,1
Ba	-5,55	-4,5	-3,14
Cu	-0,83	-0,31	0,07
Ni	-0,54	-0,08	-0,35
Co	-0,066	-0,031	-0,012
Zn	-0,3	0,175	-0,18
Fe	-73,21	-5,8	-6,3
Mn	-3,9	0,084	-0,04
Cr_2	-0,45	-0,017	0,0017
Pb_2	-2,65	-1,2	-0,18
Ba_2	1,43	2,85	1,89
Cu_2	-0,53	-0,04	-0,004
Ni_2	2,84	0,25	-0,49
Co_2	-0,013	4,24	0,04
Zn_2	-2,78	0,2	-0,9
Fe_2	-91,5	-5,6	3,5
Mn_2	-3,89	-0,28	-0,25

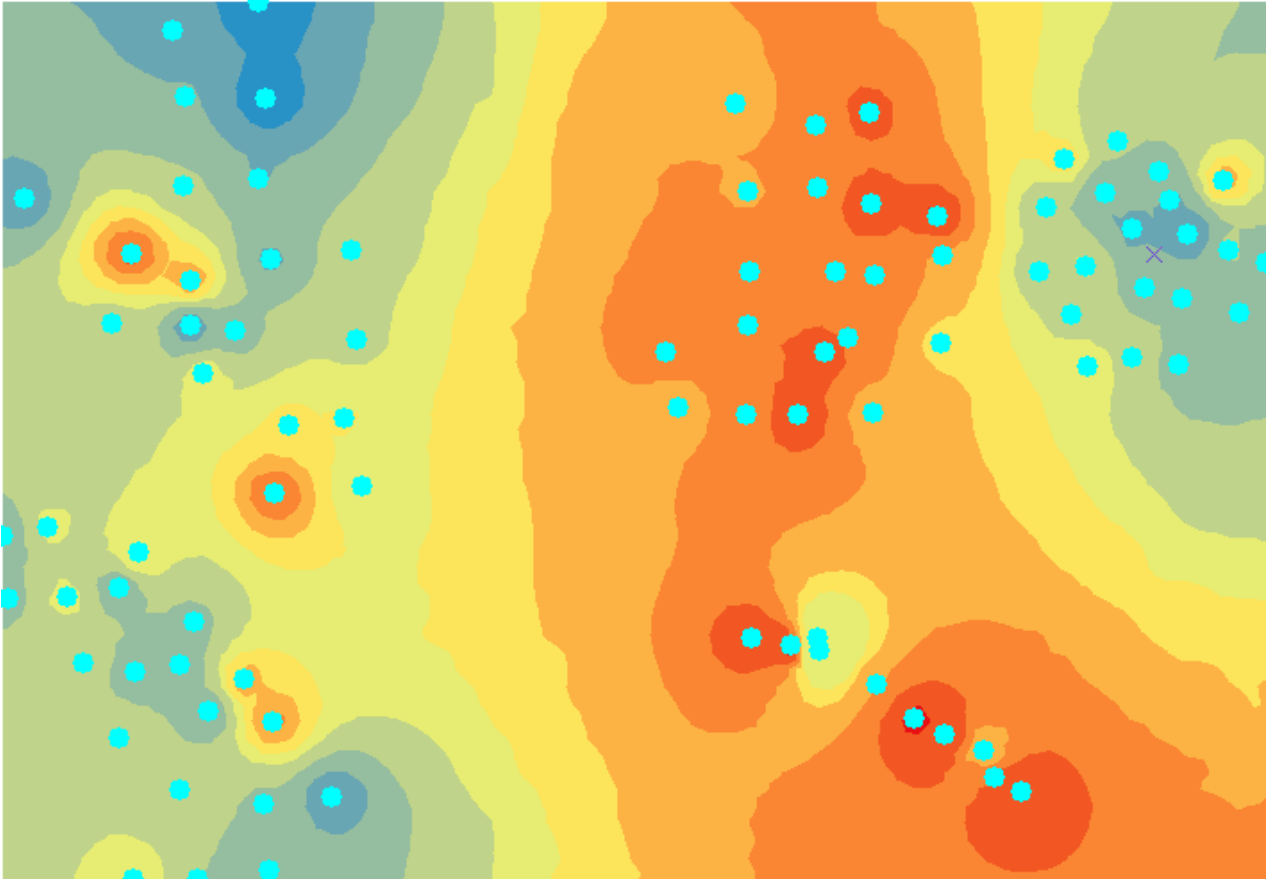
Πίνακας 11 Τιμές δεικτών R.M.S.E. και M.E.

Τα αποτελέσματα που βρέθηκαν από τη γεωστατιστική ανάλυση των δεδομένων δικαιολογούνται από την κοκκομετρική σύσταση. Η σχετικά παρόμοια εικόνα που δίνουν οι μέθοδοι, όπως το σχετικά μεγάλο RMSE στο Cross-Validation Set για τη μέθοδο Ordinary Kriging και IDW οφείλεται στο γεγονός ότι όπως είδαμε εξαρχής δεν υπάρχει ξεκάθαρα ισχυρή συσχέτιση μεταξύ της περιεκτικότητας των βαρέων μετάλλων και των τοπογραφικών ιδιοτήτων. Ωστόσο μπορούμε να θεωρήσουμε πως υπήρξε βελτίωση. Παρόμοια εικόνα δίνουν οι μέθοδοι στις τιμές του δείκτη

Mean Error όπου στο Cross-Validation Set είναι όλες κοντά στο μηδέν, με μεγαλύτερες τιμές αυτή της IDW. Επομένως έχουμε μια αρκετά καλή προσαρμογή των εκτιμωμένων τιμών στην πραγματική εικόνα.



Εικόνα 11 Ordinary Kriging Mn_2



Εικόνα 12 IDW Mn_2

Στην επιφάνεια του Inverse Distance Weighting παρουσιάζονται πολλές περιοχές που ταυτίζονται με σημεία δειγματοληψίας. Στην επιφάνεια του Ordinary Kriging παρουσιάζονται κάποιες μικρές περιοχές που ταυτίζονται με σημεία δειγματοληψίας, ενώ υπάρχει μια σαφής εικόνα σχετικά με την διακύμανση της περιεκτικότητας των μετάλλων σε όλη τη περιοχή. Το γεγονός αυτό είναι αναμενόμενο και επιβεβαιώνει το χαρακτηρισμό του Ordinary Kriging ως έναν ακριβή εκτιμητή με την έννοια ότι οι εκτιμημένες τιμές συμπίπτουν με τις τιμές των σημείων δεδομένων.

4.2.ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα εδάφη της περιοχής μελέτης αναπτύχθηκαν κυρίως σε προσχωσιγενή κοιτάσματα, και οι βασικές τους φυσικοχημικές ιδιότητες φαίνονται στον Πίνακα 1. Τα εδάφη δείχνουν μια ελαφριά υφή, είναι ελαφρώς βασικά και το ισοδύναμο ανθρακικού ασβεστίου που ορίζεται ως η ικανότητα εξουδετέρωσης της οξύτητας από δοθέν υλικό ασβέστωσης είναι μέτριο.

	N	Mean	Median	Mode	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	CV	
	Valid	Missing										
sand	90,00	0,00	52,70	52,53	46,60	8,67	75,12	0,01	0,25	-0,02	0,50	16,45
clay	90,00	0,00	19,96	20,49	20,800000a	5,34	28,49	-0,16	0,25	-0,94	0,50	26,74
silt	90,00	0,00	27,34	26,86	24,60	6,04	36,46	0,13	0,25	0,76	0,50	22,09
pH	90,00	0,00	7,91	7,96	7,790000a	0,35	0,12	-0,54	0,25	-0,17	0,50	4,38
CaCO3	90,00	0,00	25,75	24,91	14,145000a	11,72	137,33	0,73	0,25	0,79	0,50	45,50
Org_Matter	90,00	0,00	3,41	3,22	3,48	2,05	4,19	1,19	0,25	2,64	0,50	60,01
Na	90,00	0,00	1,53	0,67	0,35	2,86	8,16	4,42	0,25	23,10	0,50	186,57
k	90,00	0,00	1,18	1,16	0,81	0,60	0,36	0,73	0,25	1,27	0,50	50,40
Pb	90,00	0,00	137,59	113,25	78,000000a	119,18	14203,70	5,25	0,25	30,80	0,50	86,62
Ni	90,00	0,00	76,67	74,20	43,650000a	27,14	736,34	0,61	0,25	-0,36	0,50	35,39
Zn	90,00	0,00	182,26	153,50	131,50	121,30	14714,89	3,69	0,25	17,69	0,50	66,56
Fe	90,00	0,00	15220,05	15272,50	6390,000000a	4813,63	23171043,54	0,15	0,25	-0,79	0,50	31,63
Co	90,00	0,00	23,00	24,00	24,000000a	4,18	17,48	-0,11	0,25	-0,32	0,50	18,17
Cr	90,00	0,00	78,36	79,49	46,550000a	28,95	837,82	0,69	0,25	-0,09	0,50	36,94
Cu	90,00	0,00	45,17	38,18	24,350000a	41,00	1681,24	4,44	0,25	21,68	0,50	90,78
Mn	90,00	0,00	326,19	309,25	233,500000a	107,09	11468,80	0,22	0,25	0,19	0,50	32,83
Ba	90,00	0,00	840,13	841,00	752,500000a	244,01	59540,37	0,59	0,25	2,28	0,50	29,04
Cd	90,00	0,00	6,80	5,76	0,00	4,47	19,99	0,18	0,25	-0,92	0,50	65,74
clay_2	90,00	0,00	50,24	49,20	49,20	7,56	57,22	-0,15	0,25	0,03	0,50	15,06
sand_2	90,00	0,00	22,47	22,67	20,20	4,88	23,77	-0,06	0,25	-0,75	0,50	21,70
silt_2	90,00	0,00	27,30	27,48	22,60	5,21	27,20	0,04	0,25	1,14	0,50	19,10
Ph_2	90,00	0,00	8,08	8,14	7,97	0,31	0,09	-1,13	0,25	1,57	0,50	3,78
CaCO3_2	90,00	0,00	25,27	23,78	13,530000a	12,94	167,32	1,03	0,25	2,22	0,50	51,19
Na_2	90,00	0,00	1,20	0,58	0,35	1,67	2,79	3,43	0,25	14,28	0,50	139,51
k_2	90,00	0,00	1,00	0,97	,592195a	0,50	0,25	0,73	0,25	1,17	0,50	50,49
Pb_2	90,00	0,00	120,61	111,18	97,000000a	53,39	2850,41	3,17	0,25	13,91	0,50	44,26
Ni_2	90,00	0,00	86,48	74,85	86,500000a	56,31	3170,37	6,21	0,25	49,49	0,50	65,11
Zn_2	90,00	0,00	172,17	154,00	120,50	107,54	11565,83	3,16	0,25	12,45	0,50	62,47
Fe_2	90,00	0,00	16675,72	16567,50	19465,00	4541,50	20625200,88	0,03	0,25	-0,51	0,50	27,23
Co_2	90,00	0,00	23,76	24,43	21,100000a	3,97	15,78	-0,45	0,25	0,10	0,50	16,72
Cr_2	90,00	0,00	81,61	78,08	52,750000a	32,48	1054,67	1,03	0,25	0,84	0,50	39,79
Cu_2	90,00	0,00	41,20	36,03	26,250000a	27,06	732,16	4,98	0,25	35,94	0,50	65,68
Mn_2	90,00	0,00	350,74	336,75	313,000000a	97,54	9514,59	0,38	0,25	-0,62	0,50	27,81
Ba_2	90,00	0,00	827,10	825,75	751,500000a	234,89	55172,77	-0,34	0,25	0,32	0,50	28,40
Cd_2	90,00	0,00	7,35	6,44	0,00	4,54	20,61	0,12	0,25	-0,84	0,50	61,78
Org_Matt_2	90,00	0,00	2,69	2,61	1,4070a	1,57	2,46	1,09	0,25	2,65	0,50	58,24

a. Multiple modes exist. The smallest value is shown

Πίνακας 1

Το περιεχόμενο της οργανικής ύλης είναι πολύ διαφορετικό, όπως αναμενόταν από τη δειγματοληψία που εφαρμόζεται, τόσο στις αστικές όσο και στις περιαστικές περιοχές. Αυτά τα χαρακτηριστικά δείχνουν τα εδάφη των διαφόρων κατιόντων την ικανότητα ανταλλαγής και την επαρκή αποστράγγιση που οδηγεί σε επαρκή αερισμό του εδάφους και συνθήκες οξειδωσης. Λαμβάνοντας υπόψη την τυπική κατανομή των βροχοπτώσεων στη Μεσόγειο, το περιβάλλον του εδάφους ευνοεί μια ενεργή κοινότητα βιοτόπων του εδάφους για το μεγαλύτερο μέρος του έτους, με εξαίρεση την ξηρή θερινή περίοδο για την κορυφή του εδάφους. Τα εδάφη συλλέχθηκαν από τις θέσεις δειγματοληψίας που βρίσκεται σε μια περιοχή μικτών χρήσεων γης. Επομένως, το ανώτερο στρώμα του εδάφους μπορεί να επηρεαστεί από την συμπίκνωση ή ακόμη και την προσθήκη υλικών εδάφους από μακρινές περιοχές.

Παρ' όλα αυτά, οι τιμές του μεγαλύτερου μέρους του εδάφους, συστατικά όπως η άργιλος, η άμμος, το ανθρακικό ασβέστιο κ.α. ακολουθούν μια κανονική κατανομή, πράγμα που συνεπάγεται το ελάχιστο ανθρώπινη παρέμβαση στα εδάφη της περιοχής που μελετήθηκε. Από τον δείκτη Moran's I συμπεραίνουμε αν τα δείγματά μας ακολουθούν κανονική κατανομή και σε ποια απόσταση. Με το συντελεστή παραλλακτικότητας μπορούμε να δούμε αν τα δείγματά μας έχουν ομοιογένεια. Η μέση συγκέντρωση των μετάλλων είναι πάνω από την επικρατούσα τιμή και όπως φαίνεται σε κάποιες περιοχές η διαθεσιμότητα των μετάλλων στο έδαφος είναι πολύ υψηλή.

Από την περιοχή μελέτης τα μέταλλα που έχουν εισέλθει στο έδαφος μπορούν να εισχωρήσουν προς το εσωτερικό του εδάφους και να μολύνουν τα υπόγεια ύδατα, επιβάλλοντας έτσι μια σημαντική απειλή για το τοπικό περιβάλλον. Μελετάμε τις μεταβολές των τιμών από το παρελθόν για τις μεταβολές που θα γίνουν στο μέλλον. Έτσι για την περιοχή του Θριασίου έχουμε μια σταθερή αυξητική τάση των περισσότερων βαρέων μετάλλων στις τιμές τους με διεύθυνση από τον Βορρά προς τον Νότο. Για μερικά μέταλλα υπάρχουν ξεχωριστές διαφορές μεταξύ των

επιφανειακών και υποεπιφανειακών τιμών που παρατηρήθηκε στο μοτίβο διανομής στην περιοχή που μελετήθηκε, υποδηλώνοντας έτσι μια διαφορετική προέλευση των δύο μορφών.

Με την χωρική παρεμβολή βλέπουμε πως οι προβλέψεις μας με την μέθοδο Ordinary Kriging, είναι αξιόπιστες δηλαδή αντικειμενικές και ακριβείς, σε σχέση με την μέθοδο IDW. Τόσο με το σφαιρικό μοντέλο όσο και με το εκθετικό μοντέλο της μεθόδου OK, οι τιμές του μέσου σφάλματος είναι μικρότερες από τις τιμές της μεθόδου IDW. Αυτό δείχνει ότι τιμές της μεθόδου OK είναι αντικειμενικές και το μέσο τυποποιημένο σφάλμα μας δείχνει την ακρίβεια της πρόβλεψης, τείνει στο μηδέν, άρα έχουμε ακριβείς μετρήσεις.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΞΕΝΗ

1. Burrough and McDonnell, 1998, Principles of Geographical Information Systems. Oxford University Press Inc., New York.
2. Anselin, L., 1995. Local indicators of spatial association – LISA. *Geographical Analysis*, 27, 93-115.
3. Bailey, T., and Gatrell, A., 1995. *Interactive Spatial Analysis*, Prentice Hall, Essex, England.
4. Getis A., and Ord, J.K., 1996. Local spatial statistics an overview. In Longley P., and Batty M., (eds). *Spatial analysis: Modeling in a GIS environment*. Geoinformation International, Gambridge, UK.
5. Shortridge, A., 2007. Practical limits of Moran's autocorrelation index for raster class maps, *Computers Environment and Urban Systems*, 31, pp. 362-371s
6. D. P. Kalivas, D. P. Triantakonstantis and V. J. Kollias, 2002. "Spatial prediction of two soil properties using topographic information". *International Journal of Global Nest*, Vol. 4, pp. 41-49
7. Alexoudi-Livaditi, A., Livaditis, G., & Sachpazis, C. (1997). Geomorphological investigation of the drainage network and calculation of the peak storm runoff (Qp) and sediment yield of Sarantapotamos and Katsimidi streams, Attica, Greece. In Marinos, Koukis, Tsiambaos, & Stournaras (Eds.), *Engineering geology and the environment*. Rotterdam: Balkema.
8. Bouyoucos, G. H. (1951). A recalibration of the hydrometer method for making mechanical analysis of soils. *Agronomy Journal*, 43, 434-438.
9. Bretzel, F., & Calderisi, M. (2006). Metal contamination in urban soils of coastal Tuscany (Italy). *Environmental Monitoring and Assessment*, 118, 319-335.
10. Danellis, G. (1993). Copper, chromium, iron and nickel pollution in soils of Thriassio plain. N.T.U.A., pp. 127 (in Greek).
11. Dousis, P., Anastopoulos, I., Gasparatos, D., Ehaliotis C., & Massas I. (2013). Effect of time and glucose-C on the fractionation of Zn and Cu in a slightly acid soil. *Communications in Soil Science and Plant Analysis*, 44.
12. Gasparatos, D., & Haidouti, C. (2001). A comparison of wet oxidation methods for determination of total phosphorus in soils. *Journal of Plant Nutrition and Soil Science*, 164, 435-439.

13. Giannakopoulou, F., Gasparatos, D., Haidouti, C., & Massas, I. (2012). Sorption behavior of cesium in two Greek soils: effects of Cs initial concentration, clay mineralogy and particle size fraction. *Soil and Sediment Contamination*, 21(8), 937–950.
14. Lindsay, W. L., & Norvell, W. A. (1978). Development of a DTPA soil test for zinc, iron, manganese and copper. *Soil Science Society of America Journal*, 42, 421–428.
15. Loska, K., Wiechula, D., & Korus, I. (2004). Metal contamination of farming soils affected by industry. *Environment International*, 30, 159–165.
16. Lu, A. X., Zhang, S., & Shan, X. Q. (2005). Time effect of the fractionation of heavy metals in soils. *Geoderma*, 125, 225–234.
17. Massas, I., Ehaliotis, C., Gerontidis, S., & Sarris, E. (2009). Elevated heavy metal concentrations in top soils of an Aegean island town (Greece): total and available forms, origin and distribution. *Environmental Monitoring and Assessment*, 151, 105–116.
18. Nakos, G. (1982). Pollution of soil and vegetation in the Thriassian Plain, Greece. *Plant and Soil*, 66, 271–277.
19. Nelson, D. W., & Sommers, L. E. (1982). Total carbon, organic carbon and organic matter. In A. L. Page, R. H. Miller, & D. R. Keeney (Eds.), *Methods of soil analysis*
20. Wang, X. S., Qin, Y., & Sang, S. X. (2005). Accumulation and sources of heavy metals in urban topsoils: a case study from the city of Xuzhou, China. *Environmental Geology*, 48, 101–107.
21. ArcGis desktop help, v.10.3. “What is a z-score? What is a p-value?”
22. <https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/examining-spatial-autocorrelation-and-directional-variation.htm>

ΕΛΛΗΝΙΚΗ

1. Μήτσιος Ι., 1999. Εδαφολογία. Εκδόσεις Ζυμη. Αθήνα.
2. Κόλλια Β., Καλύβας Δ., Τριαντακωνσταντής Δ., 2012. Γεωγραφικά Πληροφοριακά Συστήματα. Εκδόσεις Έμβρυο.