ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΓΕΝΙΚΟ ΤΜΗΜΑ

ΠΜΣ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ ΓΕΩΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΧΩΡΙΚΗΣ ΠΑΡΕΜΒΟΛΗΣ ΣΕ ΠΕΡΙΒΑΛΛΟΝ G.I.S.

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΚΑΤΣΕΝΟΣ Δ. ΠΕΤΡΟΣ

Επιβλέπων

Καλύβας Διονύσιος Αναπληρωτής Καθηγητής

Αθήνα 2014

ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ ΓΕΩΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΧΩΡΙΚΗΣ ΠΑΡΕΜΒΟΛΗΣ ΣΕ ΠΕΡΙΒΑΛΛΟΝ G.I.S.

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΚΑΤΣΕΝΟΣ Δ. ΠΕΤΡΟΣ

Επιβλέπων

Καλύβας Διονύσιος Αναπληρωτής Καθηγητής

Τριμελής Συμβουλευτική και Εξεταστική Επιτροπή :

Καλύβας Διονύσιος, Αναπληρωτής Καθηγητής

Τσιλιγκιρίδης Θεόδωρος, Καθηγητής

Παπαδόπουλος Γεώργιος, Επίκουρος Καθηγητής

ΑΘΗΝΑ 2014

ΠΡΟΛΟΓΟΣ

Η παρούσα πτυχιακή μελέτη σηματοδοτεί το πέρας της φοίτησης μου, στο μεταπτυχιακό της Γεωπληροφορικής του Γενικού τμήματος του Γεωπονικού Πανεπιστημίου Αθηνών. Αντικείμενο της ήταν η ανάπτυξη ενός εργαλείου επιλογής του κατάλληλου θεωρητικού μοντέλου βαριογράμματος.

Εκφράζω τις θερμότατες ευχαριστίες μου στον αναπληρωτή καθηγητή κ. Δ. Καλύβα για την εμπιστοσύνη που έδειξε στο πρόσωπό μου. Επίσης θα ήθελα να ευχαριστήσω για τον χρόνο τους τα μέλη της τριμελούς επιτροπής Καθηγητή κ. Θ. Τσιλιγκιρίδη και τον Επίκουρο Καθηγητή κ. Γ. Παπαδόπουλο.

Δε θα μπορούσα να μην ευχαριστήσω όλο το προσωπικό του Εργαστηρίου Πληροφορικής, του Εργαστηρίου των Μαθηματικών, και της ερευνητικής ομάδας ΓΠΣ, για την έμμεση, αλλά ουσιαστική συμβολή τους, παρέχοντας γνώσεις και εποικοδομητικές κρίσεις.

Τέλος θα ήθελα να ευχαριστήσω την καλή φίλη Έλενα για την παρέα και επειδή αντιλήφθηκε όλους μου τους προβληματισμούς,την Σοφία για τις πρακτικές συμβουλές της, την οικογένεια μου για την στήριξη και την εμπιστοσύνη τους καθ’όλη τη διάρκεια των σπουδών μου στο μεταπτυχιακό και περισσότερο απ όλους την Νατάσσα για την ευεργετική παρουσία της στη ζωή μου, τη βοήθεια της αλλά και για το γεγονός ότι ήταν κοντά μου σε κάθε δύσκολη στιγμή.

Κάτσενος Δ. Πέτρος

Η Σατραπεία

Τι συμφορά, ενώ είσαι καμωμένος
για τα ωραία και μεγάλα έργα
η άδικη αυτή σου η τύχη πάντα
ενθάρρυνσι κ’ επιτυχία να σε αρνείται·
να σ’ εμποδίζουν ευτελείς συνήθειες,
και μικροπρέπειες, κι αδιαφορίες.
Και τι φρικτή η μέρα που ενδίδεις,
(η μέρα που αφέθηκες κ’ ενδίδεις),
και φεύγεις οδοιπόρος για τα Σούσα,
και πηγαίνεις στον μονάρχην Αρταξέρξη
που ευνοϊκά σε βάζει στην αυλή του,
και σε προσφέρει σατραπείες και τέτοια.
Και συ τα δέχεσαι με απελπισία
αυτά τα πράγματα που δεν τα θέλεις.
Άλλα ζητεί η ψυχή σου, γι’ άλλα κλαίει·
τον έπαινο του Δήμου και των Σοφιστών,
τα δύσκολα και τ’ ανεκτίμητα Εύγε·
την Αγορά, το Θέατρο, και τους Στεφάνους.
Αυτά πού θα σ’ τα δώσει ο Αρταξέρξης,
αυτά πού θα τα βρεις στη σατραπεία·
και τι ζωή χωρίς αυτά θα κάμεις.

Κ.Π.Καβάφης

Περίληψη

Τα γεωγραφικά πληροφοριακά συστήματα (ΓΠΣ) είναι ένα σύγχρονο κομμάτι της επιστήμης που ασχολείται με την συλλογή, την ανάλυση και την επεξεργασία πληροφοριών με καθορισμένη θέση στο χώρο. Σημαντική διαδικασία επεξεργασίας των χωρικών πληροφοριών είναι η χωρική παρεμβολή με την βοήθεια της οποίας δημιουργούνται συνεχείς επιφάνειες, έχοντας ως αρχική πληροφορία σημειακά δεδομένα.

Το Kriging αποτελεί μια γεωστατιστική μέθοδο χωρικής παρεμβολής που στηρίζεται στην πληροφορία που λαμβάνεται από το ημιβαριόγραμμα και συγκεκριμένα από την μοντελοποίηση του πειραματικού ημιβαριογράμματος. Ενώ έχει δημιουργηθεί ένα πλήθος λογισμικών που εκτελούν την διαδικασία του Kriging, η επιλογή του κατάλληλου μοντέλου ημιβαριογράμματος, που περιγράφει με την μικρότερη απώλεια πληροφορίας το πειραματικό ημιβαριόγραμμα, γίνεται με εμπειρικό τρόπο, κυρίως οπτικά, ή χρησιμοποιώντας ένα σύνολο εμπειρικών κανόνων.

Στην παρούσα εργασία αναπτύχθηκε ένα λογισμικό, το οποίο ενσωματώθηκε στο περιβάλλον επεξεργασίας χωρικών δεδομένων ArcGIS 10.1, που προσφέρει στο χρήστη την δυνατότητα δημιουργίας και πλήρους ανάλυσης πειραματικών ημιβαριογραμμάτων, με απλά βήματα και με ελεγχόμενες διαδικασίες, καθώς και προσαρμογής θεωρητικών μοντέλων τόσο με τη χρήση κριτηρίων-στατιστικών δεικτών, όσο και με την οπτικοποίηση των διαδικασιών και των αποτελεσμάτων. Αποτέλεσμα είναι η επιλογή του κατάλληλου μοντέλου ημιβαριογράμματος, με τις βέλτιστες τιμές των παραμέτρων του.

Συγκεκριμένα αρχικά γίνεται ανάλυση στα χαρακτηριστικά των δεδομένων για την επιλογή των τιμών της παραμέτρου του εύρους και της παραμέτρου του χωρικού βήματος, ώστε να γίνει υπολογισμός του πειραματικού ημιβαριογράμματος.

Εν συνεχεία παρουσιάζονται σε διαγράμματα και συγκρίνονται τα θεωρητικά ημιβαριογράμματα, αφού πρώτα έχουν υπολογισθεί οι τιμές των παραμέτρων για την βέλτιστη προσαρμογή στο πειραματικό ημιβαριόγραμμα και τελικά παρουσιάζονται οι τιμές των κριτηρίων ανά μοντέλο.

Η επιλογή του καταλληλότερου μοντέλου στηρίζεται, στα κριτήρια, Akaike, Bayesian και Hannan-Quinn, που λαμβάνουν ως παραμέτρους τον αριθμό των ελεύθερων ή μη παραμέτρων, του εκάστοτε μοντέλου. Όμως επειδή τα κριτήρια αυτά δεν αποτελούν δείκτες καταλληλότητας ενός μοντέλου αλλά είναι δείκτες σύγκρισης μεταξύ των υποψηφίων μοντέλων, τα θεωρητικά ημιβαριογράμματα οπτικοποιούνται ώστε και ο χρήστης να έχει άποψη επί των αποτελεσμάτων.

Το πρόγραμμα έχει γραφτεί σε γλώσσα Python, που επιλέχτηκε γιατί συνδυάζει απλότητα ενώ διεθνώς χαρακτηρίζεται ως πολύ ισχυρή. Μερικά από τα πλεονεκτήματά της είναι :α) υπάρχουν βιβλιοθήκες με μακροεντολές ειδικά δημιουργημένες για γεωγραφικές εφαρμογές, β) τα προγράμματα που έχουν γραφτεί σε γλώσσα Pythonμπορεί να υλοποιηθούν σε πολλές πλατφόρμες λειτουργικών συστημάτων γ) δίνει μεγάλα περιθώρια βελτίωσης των προγραμμάτων που είναι γραμμένα σε γλώσσα Pythonμε τις βιβλιοθήκες που ενσωματώνουν τη (low level) γλώσσα προγραμματισμού C ή και με βιβλιοθήκες που επιτρέπουν την παράλληλη επεξεργασία.

Το λογισμικό που αναπτύχθηκε στα πλαίσια της εργασίας, είναι ένα απλό στη χρήση, με φιλικά GUI, χωρίς να προϋποθέτει κάποια ιδιαίτερη γνώση κάποιου εξειδικευμένου περιβάλλοντος-γλώσσας, όπως γίνεται μέχρι σήμερα με εργαλεία που έχουν ανατηχθεί σε γλώσσα R ή Matlab.

Τα λογισμικό δοκιμάστηκε σε πραγματικά δεδομένα από αναλύσεις εδαφικών δειγμάτων και παρουσιάζονται τα αποτελέσματα της βαριογραφικής ανάλυσης.

Το λογισμικό που αναπτύχθηκε πολύ εύκολα μπορεί να τροποποιηθεί και να ενσωματωθεί και σε άλλα εμπορικά GIS λογισμικά ή ελεύθερου κώδικα ή ακόμη και να χρησιμοποιηθεί ανεξάρτητα από περιβάλλοντα γεωγραφικών πληροφοριακών συστημάτων.

Λέξεισκλειδιά : ημιβαριόγραμμα, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), επιλογή μοντέλου

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝΣΕΛΙΔΕΣ

1. ΕΙΣΑΓΩΓΗ

1.1.Εισαγωγή – Βιβλιογραφική Ανασκόπηση

Είναι γνωστό σε όλους όσους ασχολούνται με την επεξεργασία και ανάλυση χωρικών δεδομένων, ότι ένα από τα σημαντικότερα κομμάτια της χωρικής επεξεργασίας είναι η πρόβλεψη της διακύμανσης της τιμής που εμφανίζει μια μελετούμενη ιδιότητα, στο χώρο.

Πριν τη δεκαετία του 70' σύμφωνα με του Beckett και Webster (1971) όλες οι προσπάθειες για την ανάπτυξη μεθόδων εξήγησης της διακύμανσης, σχετίζονταν με τις τεχνικές δειγματοληψίας. Για να υπάρξουν αποτελεσματικές εκτιμήσεις έπρεπε πρώτα να γίνει κάποια ταξινόμηση των περιοχών δειγματοληψίας και στη συνέχεια να ακολουθηθεί η σωστή μέθοδος δειγματοληψίας (Webster, 1985). Η προηγούμενη μέθοδος όμως είχε περιορισμούς στη χρήση της και ήταν ανεπαρκής να εξηγήσει την μεταβολή μιας ιδιότητας στο χώρο.

Πολλές προσπάθειες υπήρξαν για να συνδέσουν με μαθηματικά την διακύμανση με την απόσταση (Καλύβας, 1989), τέτοιες ήταν των Webster και Cuanalo το 1975 όπου χρησιμοποίησαν συναρτήσεις αυτοσυσχέτισης, και αποτύπωσαν τα αποτελέσματα στο αυτοσυσχετόγραμμα, για να μελετήσουν τις διαφοροποιήσεις των τιμών μιας ιδιότητας. Την ίδια περίπου εποχή ο Matheron ανέπτυξε την θεωρία των περιφερειακών μεταβλητών (Regionalist Variable Theory) που αποτέλεσε την βάση για την ποσοτική περιγραφή της μεταβλητότητας μια ιδιότητας στο χώρο (Καλυβας, 1989). Κάπως έτσι φτάσαμε στις μεθόδους χωρικής παρεμβολής, σήμερα υπάρχουν παρά πολλές. Χωρίζονται σε Γεωστατιστικές και μη γεωστατιστικές. Οι Franke (1982) και Meyers (1994) στις εργασίες τους συγκρίνουν πάνω από 30 μεθόδους χωρικής παρεμβολής συνολικά και οι πιο δημοφιλείς από αυτές συμπεριλήφθησαν σε προγράμματα χωρικής ανάλυσης και γεωγραφικά πληροφοριακά συστήματα (ΓΠΣ) (CooperandJarvis, 2004).

Η πραγματική όμως άνθισης της γεωστατιστικής και αναπόφευκτα των μεθόδων χωρικής παρεμβολής, ήρθε με την εξάπλωση της χρήσης του ηλεκτρονικού υπολογιστή. Παρόλο που τα γεωγραφικά πληροφοριακά συστήματα αναπτύσσονταν από την δεκαετία του '60, μέχρι τα τέλη της δεκαετίας του '80, υπήρχε έλλειψη σε εξειδικευμένα λογισμικά χωρικής στατιστικής, εμποδίζοντας την εξάπλωση της χρήσης αυτών των μεθόδων

(Haining, 1989). Ο λόγος ήταν ότι στην αρχή οι χωρικές επεξεργασίες και στατιστικές επεξεργασίες των χωρικών δεδομένων, ήταν κάτι ξεχωριστό (Openshaw 1990, Ding & Fotheringham, 1992). Ώθηση στην ανάπτυξη εξειδικευμένων λογισμικών χωρικής ανάλυσης, δόθηκε από εργασίες και ακαδημαϊκές έρευνες (Goodchild 1987), υπήρξαν αρκετές προσπάθειες ανάπτυξης κάποιων εργαλείων χωρικής στατιστικής επεξεργασίας, όπως το Geo-EAS (Geostatistical Environmental Assessment Software) των Englund και Sparks(1988) που ήταν γραμμένο σε γλώσσα FORTRAN, αλλά και στην Ελλάδα από τον Καλύβα το 1989 όπου επίσης σε γλώσσα FORTRAN ανέπτυξε ένα εργαλείο βαριογραφίας.

Από τις πρώτες προσπάθειες ανάπτυξης εξειδικευμένου λογισμικού ήταν το Spiderand Regardsoftware, ανεπτυγμένο να δουλεύει σε περιβάλλον machintosh (Haslettet. al. 1990, 1991, Unwin 1994), όπου περιείχε οπτικοποίηση δεδομένων και διαγραμμάτων αλλά είχε περιορισμένες δυνατότητες στην περεταίρω δυνατότητα επεξεργασίας των δεδομένων, με εξαίρεση την βαριογραφία στην έκδοση του 1992.

Το λογισμικό Spacestat αποτέλεσε το πρώτο ανεξάρτητο λογισμικό χωρικής στατιστικής και παρουσιάστηκε απλ την εταιρία NCGIA το 1991 (Anselin, 1992). Για πολύ καιρό ήταν το μοναδικό λογισμικό που υπήρχε με ποικιλία στατιστικών λειτουργιών όπως χωρικής παρεμβολή, και της μεθόδου μέγιστης πιθανοφάνειας.

Ακολούθησε το 1996 το λογισμικό S+SpatialStats το 1996 από την Mathsoft, ως επέκταση του στατιστικού πακέτου S-Plus και περιείχε γεωστατιστικές αναλύσεις όπως βαριογραφία και Kriging. Ενώ από το 1997 μπορούσε να χρησιμοποιηθεί από το περιβάλλον του λογισμικού της ESRIArcView αλλά και από το περιβάλλον του ελεύθερου λογισμικού GRASSGIS (Bao & Martin 1997, Bao et. al. 2000).

Αρκετά χρόνια αργότερα αναπτύχθηκαν βιβλιοθήκες για γλώσσα R αλλά και σε περιβάλλον Matlab όπως η BMElib του Χρηστάκου το 1992 (Christakos et. al. 2002).

Ενώ το 2010 παρουσιάστηκε η βιβλιοθήκη PySAL της γλώσσας προγραμματισμού Python.

Όπως γίνεται αντιληπτό στις μέρες μας υπάρχει πλήθος προγραμμάτων στατιστικής ανάλυσης χωρικών δεδομένων και τα περισσότερα από αυτά προσφέρουν γεωστατιστικές μεθόδους χωρικής παρεμβολής Kriging, όμως ακόμα η ανάλυση βαριογράμματος γίνεται είτε εμπειρικά από το χρήστη, είτε με αδιαφανείς διαδικασίες που δεν ελέγχονται ούτε υπάρχει δυνατότητα επέμβασης και διόρθωσης της διαδικασίας.

Υπάρχει φυσικά και η δυνατότητα χρήσης γλωσσών προγραμματισμού 4ης γενιάς όπως της Matlabκαι της R όπου υπάρχουν εξειδικευμένες βιβλιοθήκες για χωρική στατιστική με εργαλεία για βαριογραφική ανάλυση και μεθόδους επιλογής μοντέλου βαριογράμματος. Λόγο του γεγονότος ότι είναι γλώσσες προγραμματισμού, υπάρχει δυσκολία στο χειρισμό και η χρήση τους απαιτεί εξειδικευμένη γνώση, τόσο της γλώσσας όσο και πολύ καλή γνώση μαθηματικών.

Αυτό το κενό έρχεται να καλύψει η παρούσα εργασία καθώς προτείνει έναν εύκολο και απλό τρόπο ανάλυσης του ημίβαριογραμματος, ο οποίος είναι πλήρως ελεγχόμενος. Η ανάπτυξη των εργαλείων έγινε έτσι ώστε να δουλεύουν σε περιβάλλον ArcGIS10.1 με τη μορφή toolbox. Δίνοντας την ευκολία στο χρήστη να μπορεί να χρησιμοποιήσει τα δεδομένα που υπάρχουν σε αρχεία του ArcGIS (shapefiles), χωρίς να είναι αναγκαία καμιά μετατροπή αρχείου για την επεξεργασία του, πράγμα που είναι σύνηθες σε πολλά προγράμματα στατιστικής επεξεργασίας. Επίσης αμέσως μετά την βαριογραφική ανάλυση ο χρήστης μπορεί κατευθείαν να επιλέξει τις εξαγόμενες τιμές των παραμέτρων ώστε να εκτέλεση τη διαδικασία του Krigingπου προσφέρει το λογισμικό ArcGIS10.1. Τα εργαλεία που αναπτύχθηκαν να δίνουν πάντα εικόνα σε όλα τα στάδια της ανάλυσης της βαριογραφίας ώστε να μπορεί ο χρήστης να κάνει τις απαραίτητες αλλαγές ώστε να εξάγει αποτελέσματα ακριβή και ποιοτικά.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Γεωγραφικά Πληροφοριακά Συστήματα

Τα Γεωγραφικά Πληροφοριακά Συστήματα, ή για συντομία ΓΠΣ, είναι συστήματα αυτόματης χωρικής επεξεργασίας που περιγράφουν οντότητες του πραγματικού κόσμου σε σχέση με: α) τη θέση τους ως προς ένα σύστημα συντεταγμένων, β) παραμέτρους που δεν έχουν σχέση με τη χωρική τους θέση και γ) τους τοπολογικούς συνδέσμους που έχουν με άλλες οντότητες. Τα ΓΠΣ είναι μία τεχνολογία που έχει θεωρητικό και πρακτικό υπόβαθρο. Τι είναι όμως αυτή η τεχνολογία που έχει προκαλέσει τόσο μεγάλο ενδιαφέρον ώστε να αποτελεί αντικείμενο πολλών επιστημονικών περιοδικών και συνεδρίων (Κόλλια, 2012); Τα ΓΠΣ δεν είναι απλά συστήματα σχεδίασης με ηλεκτρονικό υπολογιστή (Η/Υ), αλλά και ένα σημαντικό εργαλείο που προσφέρει στους χρήστες και υπεύθυνους για τη λήψη αποφάσεων ισχυρά εργαλεία για την επίλυση των σύνθετων και όχι πάντα πλήρως δομημένων χωρικών προβλημάτων. Τα συστήματα αυτά οφείλουν να έχουν μια αποτελεσματική απόδοση (π.χ., ταχεία απόκριση), ώστε να υποστηρίξουν τις ανάγκες των χρηστών σε ένα περιβάλλον παραγωγής (Στεφανάκης, 2010).

Τεχνικές αυτόματης σχεδίασης οπωσδήποτε χρειάζονται σε ένα καλό ΓΠΣ, αλλά στα ΓΠΣ μεγάλη σπουδαιότητα έχουν παράλληλα με τις χωρικές και οι μη χωρικές πληροφορίες. Τα ΓΠΣ έχουν πολλά κοινά σημεία με αυτό που ονομάζεται«σχεδίαση με Η/Υ» (ComputerAidedDesign) και που χρησιμοποιείται ευρέως για τον σχεδιασμό αεροπλάνων, υπολογιστών κ.λπ. Η σχεδίαση με χρήση Η/Υ χρησιμοποιεί μη χωρικές πληροφορίες, αλλά τα ΓΠΣ περιέχουν πολύ μεγαλύτερο εύρος και όγκο πληροφοριών και η φύση των επεξεργασιών που παρέχουν είναι πολύ διαφορετική (Dangermond, 1986, Cowen, 1988).

2.1.1 Ορισμός

Ενώ έχουν προταθεί πολλοί ορισμοί για τα ΓΠΣ κανένας δεν έχει θεωρηθεί απολύτως ικανοποιητικός ώστε να επικρατήσει. Στην πιο απλή του μορφή ένα ΓΠΣ μπορεί να θεωρηθεί σαν ένα σύστημα Βάσης δεδομένων (ΒΔ), στο οποίο είναι αποθηκευμένα χωρικά δεδομένα (spatialdata) που πάνω τους εφαρμόζεται ένα σύνολο από επεξεργασίες, με σκοπό να απαντώνται ερωτήσεις για τις χωρικές οντότητες της ΒΔ. Ο απλός αυτός ορισμός των ΓΠΣ είναι φυσικά απέχει πολύ από τις δυσκολίες που συναντούν οι επιστήμονες και οι τεχνικοί που υλοποιούν μεγάλα ΓΠΣ. Τα προβλήματα των ΓΠΣ όμως ξεκινούν όταν ο όγκος των χωρικών δεδομένων, που επεξεργάζονται,αρχίζει να γίνεται μεγάλος.

Επίσης, ένα ΓΠΣ μπορεί να θεωρηθεί σαν μία αποθήκη ψηφιακών χαρτών. Όλοι εκτιμούμε την αξία των χαρτών και η αίσθηση ότι μπορούμε να επεξεργασθούμε χάρτες με έναν Η/Υ είναι ανάλογη με αυτή της χρήσης του επεξεργαστή κειμένου σε άλλου τύπου πληροφορίες. Άλλος ορισμός είναι, ότι τα ΓΠΣ είναι εργαλεία Η/Υ για τη λύση γεωγραφικών προβλημάτων και την εφαρμογή επεξεργασιών σε χωρικά δεδομένα ή είναι συστήματα λογισμικού που συνδέουν πληροφορίες για τη θέση των αντικειμένων με πληροφορίες για το είδος τους. Ένας πιο πλήρης ορισμός είναι ότι τα ΓΠΣ είναι συστήματα για τη συλλογή, την αποθήκευση, την ανάκτηση με όποιο τρόπο χρειάζεται, την επεξεργασία και την αποτύπωση των γεωγραφικών δεδομένων, για κάποιο σκοπό (Dangermond, 1988). Ο Tomlin (1990) αναφέρει ότι ένα ΓΠΣ είναι ένα σύστημα για την προετοιμασία, παρουσίαση και κατανόηση φαινομένων που συμβαίνουν στην επιφάνεια της γης. Ένας πιο ακριβής ορισμός από τον ίδιο είναι ότι ένα ΓΠΣ είναι ένας συνδυασμός υλικού και λογισμικού σχεδιασμένου ειδικά για την αναζήτηση, συντήρηση και χρήση των χαρτογραφικών δεδομένων.

Τα ΓΠΣ είναι κάτι παραπάνω από εργαλεία για τις παραπάνω διαδικασίες. Χρησιμοποιούνται νια την οργάνωση - παρουσίαση χωρικών πληροφοριών και την ανάλυση επιδράσεων εναλλακτικών αποφάσεων στο χώρο. Είναι εργαλεία για τη μελέτη του περιβάλλοντος και των δυνατών αποτελεσμάτων που θα προκύψουν από κάποιο συγκεκριμένο σχεδιασμό. Δηλαδή χρησιμοποιούνται όπως οι προσομοιωτές (simulators) των αεροπορικών πτήσεων για να αποφευχθούν ορισμένες λανθασμένες ενέργειες πριν εφαρμοσθούν στη γη (π.χ. σχεδιασμός χρήσεων γης, αξιολόγηση καλλιεργειών, εγγειοβελτιωτικά έργα κ.λπ.) (Star and Estes, 1990, Jones, 1997, Burrough and McDonnell, 1998, Longley et al., 2001).

2.1.2 Ιστορική Αναδρομή

Η πρώτη καταγεγραμμένη οργανωμένη χρήση γεωγραφική μεθόδου τοποθετείται στο 1854 και στον Άγγλο επιδημιολόγο JohnSnowόπου κατέγραψε και αναπαρέστησε χωρικά την εξάπλωση χολέρας, στην ευρύτερη περιοχή του Λονδίνου, χρησιμοποιώντας σημεία για την αναπαράσταση της θέσης μεμονωμένων περιστατικών. Το αποτέλεσμα της μελέτης του γύρω από την κατανομή των κρουσμάτων χολέρας οδήγησε στον εντοπισμό της πηγής της ασθένειας, η οποία ήταν μια μολυσμένη αντλία νερού στο κέντρο της περιοχής εμφάνισης της ασθένειας. Ο χάρτης που είχε δημιουργήσει το 1854 ο JohnSnow αναλύθηκε από τον γεωγράφο και καθηγητή του Πανεπιστημίου της Οξφόρδης. E.W. Gilbert τέσσερα χρόνια αργότερα. O Gilbert ουσιαστικά απεικόνισε την ομαδοποιημένη κατανομή των κρουσμάτων χολέρας που είχε καταγράψει ο Snow. Ενώ από παλιότερα οι βασικές αρχές της τοπογραφίας έβρισκαν εφαρμογή στην χαρτογραφία. ο χάρτης του Snow ήταν μοναδικός καθώς χρησιμοποιούσε για πρώτη φορά χαρτογραφικές μεθόδους όχι μόνο για να απεικονίσει αλλά και να αναλύσει ομάδες ανεξάρτητων γεωγραφικών φαινομένων.

Η πρώτη μεγάλη αλλαγή στη γεωγραφική μέθοδο ήρθε στις αρχές του εικοστού αιώνα όπου σηματοδοτήθηκε η απαρχή της ανάπτυξης της φωτολιθογραφίας και είχε ως αποτέλεσμα τον διαχωρισμό των χαρτών σε πολλά επίπεδα. Στο δεύτερο μισό του 20ου αιώνα υπήρξε η επανάσταση στη γεωγραφία με την χρήση σε αυτή των υπολογιστικών μηχανών, έτσι στις αρχές του 1960 πραγματοποιήθηκε η απαρχή των χαρτογραφικών εφαρμογών με χρήση μηχανικών υπολογιστικών συστημάτων .

Το 1962 αναπτύχθηκε, το πρώτο σε παγκόσμιο επίπεδο, Γεωγραφικό Σύστημα Πληροφοριών την Οτάβα του Οντάριο, Καναδά από το Ομοσπονδιακό Τμήμα Δασών και Αγροτικής Ανάπτυξης. Πρωτεργάτης της προσπάθειας αυτής ήταν ο Άγγλος γεωγράφος Δρ. Roger Tomlison γνωστός και ως «Πατέρας των G.I.S.», και το ονόμασε "Canada Geographic Information System (C.G.I.S.)". Βασικός στόχος αυτού του πρωτοπόρου συστήματος ήταν η αποθήκευση, ανάλυση και διαχείριση των συλλεγόμενων δεδομένων από την υπηρεσία που είχε αναλάβει την καταγραφή γαιών σε ολόκληρο τον Καναδά.

Αποτέλεσμα αυτού του εγχειρήματος ήταν να καταγραφούν οι πληροφορίες των αγροτικών περιοχών του Καναδά σχετικά με τις διάφορες χρήσεις των εδαφών, σε χάρτες κλίμακας 1:50.000. Για την επίτευξη της ανάλυσης προστέθηκε στα δεδομένα μια κλίμακα βαθμολόγησης.

Το εν λόγω σύστημα αποτέλεσε μια σημαντική βελτίωση στις χαρτογραφικές εφαρμογές με τη χρήση ηλεκτρονικού υπολογιστή καθώς διέθετε δυνατότητα επίθεσης, μέτρησης και ψηφιοποίησης. Διέθετε ένα εθνικό σύστημα συντεταγμένων, κωδικό ποιημένες γραμμές σαν καμπύλες οι οποίες υποστήριζαν πραγματική τοπολογία και αποθήκευε περιγραφικά χαρακτηριστικά καθώς και χωρικές πληροφορίες σε ξεχωριστά αρχεία. Το σύστημα που οραματίστηκε και εφάρμοσε η ομάδα του Tomlison χρησιμοποίησε την αλληλεπίθεση (overlay) επιπέδων για την προώθηση της χωρικής ανάλυσης συγκλινόντων γεωγραφικών δεδομένων. Το συγκεκριμένο σύστημα συνέχισε τη λειτουργία του μέχρι τη δεκαετία του 1990 και δημιούργησε μια μεγάλη βάση δεδομένων για τους εδαφικούς πόρους του Καναδά. Παρόλο την συνεχή χρήση του υπήρξε ποτέ ως εμπορικό λογισμικό.

Στις αρχές του 70 οι Η.Π.Α. ανέπτυξαν το δικό τους σύστημα ΓΠΣ με σκοπό την καταγραφή και επεξεργασία των αποτελεσμάτων της τότε προγραμματισμένης απογραφής. Το DIME (Dual Independent Map Encoding) είχε με ακρίβεια καταγεγραμμένα σε ψηφιακά αρχεία όλα τα δεδομένα της απογραφής κατά οδούς (Peuquet&Marble, 1990). Μέχρι τα τέλη της δεκαετίας του ’70 όλες οι υπηρεσίες είχαν αυτοματοποιήσει όλες τις εργασίες που σχετίζονταν με χωρικές πληροφορίες.

Αργότερα τη δεκαετία του ’80 έκαναν την πρώτη τους εμφάνισηγνωστές ακόμη και σήμερα εταιρείες όπως Intergraph, EnvironmentalSystemResearchInstitute (E.S.R.I), ComputerAidedResourceInformationSystem (C.A.R.D.I.S.) καιERDASόπου παρείχαν στην αγορά πακέτα λογισμικών Γεωγραφικών Συστημάτων Πληροφοριών.

Παράλληλα με την ανάπτυξη των ΓΠΣ υπήρξε και η ανάπτυξη της τηλεπισκόπησης που βοήθησε στην ανάπτυξη των ΓΠΣ καθώς παρείχε σε μεγάλη γκάμα αλλά και όγκο πληροφορίες προς επεξεργασία. Αυτό συνέβη όταν την δεκαετία του πενήντα κατασκευάστηκαν οι πρώτοι στρατιωτικοί δορυφόροι με σκοπό την κατασκοπία χρησιμοποιώντας ψηφιακά φιλμ, που αργότερα αντικαταστάθηκαν από ψηφιακούς ανιχνευτές και κάπως έτσι φτάσαμε στους πρώτους πολιτικούς δορυφόρους που παρείχαν έναν τεράστιο όγκο πληροφοριών.

Όσο η ανάπτυξη των ηλεκτρονικών υπολογιστών ανθούσε και η τεχνολογία γινόταν ολοένα και πιο εύκολα προσβάσιμη τόσο πλήθαιναν οι εφαρμογές των ΓΠΣ.

Με την πάροδο του χρόνου τα εμπορικά πακέτα ΓΠΣ προσαρμόστηκαν στα νέα λειτουργικά συστήματα στους οικιακούς ηλεκτρονικούς υπολογιστές ενώ με την διάδοση του διαδικτύου δημιουργήθηκε η έννοια του WebGIS δηλαδή των συστημάτων ΓΠΣ όπου η χρήση και η επεξεργασία των χωρικών δεδομένων γινόταν εξ ολοκλήρου στο διαδικτύων

Τέλος αναπτυχθεί λογισμικά πακέτα ΓΠΣ ανοικτού κώδικα όπου ο χρήστης έχει τη δυνατότητα να ελέγξει και να προσαρμόσει τις διαδικασίες των ΓΠΣ ή ακόμη και να συμβάλει στην ανάπτυξη νέων μεθόδων και εργαλείων για τις διάφορες χωρικές επεξεργασίες.

2.1.3 Απαιτήσεις χρηστών και Εφαρμογών

Υπάρχει μια μεγάλη ποικιλία σε εφαρμογές, που σχετίζονται με τα χωρικά δεδομένα και τις χρηστικές απαιτήσεις. Γίνεται αντιληπτό ότι ο αποτελεσματικός σχεδιασμός ενός Γεωγραφικού Πληροφοριακού Συστήματος (ΓΠΣ) αποτελεί μια σημαντική διαδικασία προσαρμογής ενός ΓΠΣ στην εκάστοτε εφαρμογή. Στηρίζεται στο είδος της εφαρμογής, των προσβάσιμων μέσων και το είδος των δεδομένων αλλά και στις βασικές κατηγορίες λειτουργιώνγια την επεξεργασία των δεδομένων αυτών.

Οι κατηγορίες λειτουργιών επεξεργασίας των γεωγραφικών δεδομένων ομαδοποιούνται σε τέσσερεις αν και υπάρχει παραλλαγών και επεκτάσεων σε αυτές ανάλογα με το είδος των εφαρμογών και την εξέλιξη των ΓΠΣ.

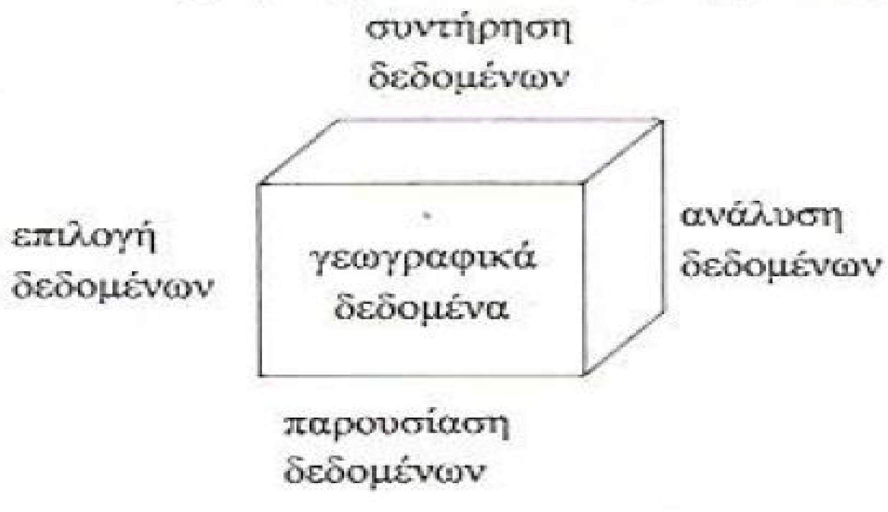
Οι τέσσερεις κατηγορίες αναλύονται παρακάτω (παράθεση από Εμμανουήλ Στεφανάκη, 2010)

1. ***Συντήρηση δεδομένων:*** Η κατηγορία αυτή περιστοιχίζει μια ευρεία ποικιλία μεθόδων για εισαγωγή, διαγραφή, και ενημέρωση των δεδομένων του συστήματος.
2. ***Επιλογή δεδομένων***. Οι λειτουργίες αυτής της κατηγορίας υποστηρίζουν την επιλογή των γεωγραφικών οντοτήτων με βάση τις τιμές των γνωρισμάτων τους. Ανάλογα με τις διαστάσεις που εμπλέκονται, οι λειτουργίες επιλογής δεδομένων διαιρούνται στις ακόλουθες τέσσερις υποκατηγορίες: (α) *χωρικές επιλογές* (βασίζονται στη χωρική διάσταση των γεωγραφικών οντοτήτων), (β) *θεματικές επιλογές* (βασίζονται στην ταυτότητα ή τη θεματική διάσταση των γεωγραφικών οντοτήτων), (γ)*χρονικές επιλογές* (βασίζονται στη χρονική συμπεριφορά των γεωγραφικών οντοτήτων), και (δ) *μικτές επιλογές* (εμπλέκουν περισσότερες από μία διαστάσεις των γεωγραφικών οντοτήτων, καθώς και τη χρονική τους συμπεριφορά, π.χ., χωρική και/ή θεματική και/ή χρονική επιλογή). Να σημειωθεί ότι, τα μέτρα ποιότητας των δεδομένων

πρέπει να λαμβάνονται υπόψη, ώστε να υπάρχει μια ένδειξη της ακρίβειας και αξιοπιστίας των αποτελεσμάτων. Οι λειτουργίες επιλογής μαζί με αυτές της συντήρησης συνθέτουν το σύνολο των *λειτουργιών διαχείρισης* των γεωγραφικών δεδομένων, τις οποίες παρέχει ο διαχειριστής δεδομένων του πληροφοριακού συστήματος .

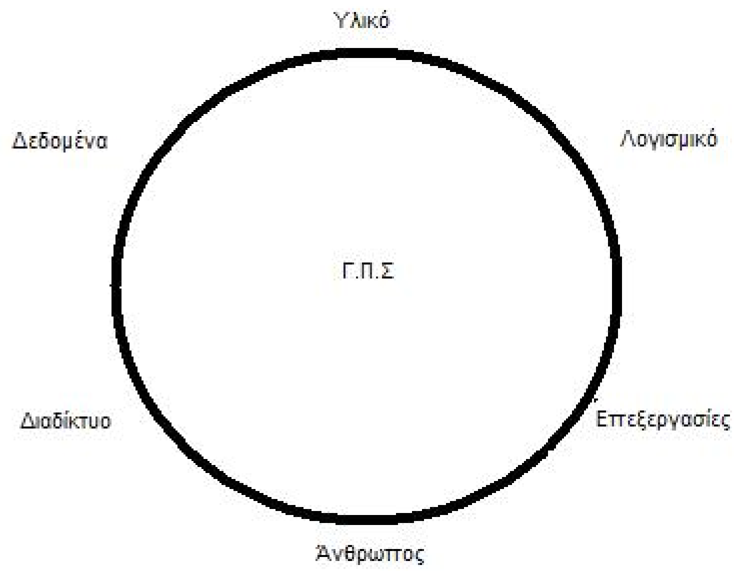
3. ***Ανάλυση δεδομένων:*** Οι λειτουργίες ανάλυσης (ή ερμηνείας) των γεωγραφικών δεδομένων είναι πολυάριθμες και εξαρτώνται στενά από το πεδίο εφαρμογών. Στο παρελθόν έχουν προταθεί αρκετές κατηγοριοποιήσεις των λειτουργιών αυτών, παρόλα αυτά συνεχίζει να απουσιάζει μια κοινά αποδεκτή άλγεβρα για τα γεωγραφικά δεδομένα. Οι λειτουργίες ανάλυσης επεξεργάζονται τα γνωρίσματα (χωρικά, θεματικά, χρονικά, ποιότητας) των γεωγραφικών οντοτήτων και καταλήγουν είτε στη δημιουργία νέων οντοτήτων, στην τροποποίηση των τιμών των γνωρισμάτων αυτών ή στην εξαγωγή γενικότερων συμπερασμάτων, μέτρων και πληροφοριών για τα γεωγραφικά δεδομένα. Συνοπτικά, οι λειτουργίες ανάλυσης γεωγραφικών δεδομένων περιλαμβάνουν τις: (α) λειτουργίες ταξινόμησης και γενίκευσης, (β) λειτουργίες μέτρησης, (γ) λειτουργίες επίθεσης, (δ) λειτουργίες αναγνώρισης χωρικών σχέσεων (τοπολογικές, κατεύθυνσης, μετρητικές), (ε) λειτουργίες παρεμβολής, (ζ) λειτουργίες σύνδεσης, κ.ά.

4. ***Παρουσίαση δεδομένων:*** Η κατηγορία αυτή συγκεντρώνει μια ποικιλία μεθόδων για την παρουσίαση των αποτελεσμάτων των λειτουργιών επιλογής ή ανάλυσης των δεδομένων. Π παρουσίαση μπορεί να είναι σε μορφή χάρτη, γραφήματος, κειμένου, πίνακα, αντικειμένων πολυμέσων, κ.λπ., και δίνεται στο χρήστη μέσω κατάλληλα σχεδιασμένων διεπαφών.



Εικόνα 2.1: βασικές κατηγορίες λειτουργιών ενός ΓΠΣ (Ε. Στεφανάκης 2010)

2.1.4 Συστατικά Μέρη ενός Γεωγραφικού Συστήματος πληροφοριών



Εικόνα 2.2: Τα συστατικά μέρη ενός ΓΠΣ

Κάθε ΓΠΣ αποτελεί ένα υπολογιστικό σύστημα με έξι βασικά συστατικά μέρη (εικόνα πάνω): (α) ταγεωγραφικά δεδομένα, (β) τολογισμικό, (γ) τουλικό, *(δ) οι* επεξεργασία, (ε) ο άνθρωπος και (στ) το διαδίκτυο. Το υλικό και το λογισμικό είναι εξελίξιμα συστατικά έτσι λέμε ότι έχουν καθορισμένο κύκλο ζωής. Τα δεδομένα είναι το πιο σημαντικό μέρος ενός συστήματος γι αυτό και διακινούνται συχνά με αντίτιμο μεγάλα ποσά. Η επεξεργασία σπάνια αλλάζει καθώς και οι δεξιότητες των ανθρώπων .

2.1.4.1 Τα Γεωγραφικά Δεδομένα

Αποτελεί την καρδία των ΓΠΣη δημιουργία της ΒΔ, είναιη πιο χρονοβόρα και δαπανηρή διαδικασία κατά την υλοποίηση ενός ΓΠΣ είναι. Τα δεδομένα που αποθηκεύονται στις ΒΔ ενός ΓΠΣ είναι τριών ειδών α) χωρικά, β) παραμετρικά/περιγραφικά (μη χωρικά) και γ) εικόνες και προέρχονται κυρίως από υπάρχοντες χάρτες, παρατηρήσεις αγρού και την τηλεπισκόπηση. Η εισαγωγή τους στο ΓΠΣ γίνεται με τη χρήση συμβατικών Η/Υ, ψηφιοποιητές (digitizers), μαγνητικά αρχεία, σαρωτές (scanners), το διαδίκτυο και άλλα μαγνητικά και οπτικά μέσα. Τα χωρικά δεδομένα αποτελούνται από σημεία, γραμμές και περιοχές και καθορίζουν τις θέσεις και τα σχήματα των χαρακτηριστικών του χάρτη π.χ. κτίρια, δρόμοι, πόλεις κ.λπ. Τα παραμετρικά δεδομένα αποτελούν τις πληροφορίες που συνοδεύουν τα χωρικά χαρακτηριστικά. Για παράδειγμα ένας χάρτης εδαφοτομών μπορεί να συνοδεύεται από περιγραφική πληροφορία που αφορά τα μορφολογικά χαρακτηριστικά τους. Οι εικόνες μπορεί να περιλαμβάνουν δορυφορικές εικόνες, αεροφωτογραφίες ή σαρωμένους χάρτες (χάρτες που έχουν μετατραπεί από χαρτί σε ψηφιακή μορφή).

2.1.4.2 Το Λογισμικό

Υπάρχουν πολλά πακέτα λογισμικού ΓΠΣ διαθέσιμα σήμερα. Όλα ία πακέτα πρέπει να περιλαμβάνουν προγράμματα για:

1. την είσοδο και επαλήθευση των δεδομένων,
2. την αποθήκευση των δεδομένων και τον χειρισμό της Βάσης Δεδομένων,
3. την έξοδο των δεδομένων και παρουσίαση τους,
4. το μετασχηματισμό των δεδομένων,
5. την ανάλυση / επεξεργασία των δεδομένων,
6. την επικοινωνία με τον χρήστη.

Τα σημερινά ΓΠΣ επιτρέπουν την αποθήκευση των χωρικών και μη χωρικών δεδομένων στην ίδια ΒΔ (object-relationalmodel). Το μοντέλο που χρησιμοποιούνταν προηγουμένως (geo-relationalmodel) χειριζόταν τα χωρικά και μη χωρικά δεδομένα χωριστά. Επίσης, τα σύγχρονα πακέτα περιλαμβάνουν ένα σύνολο εργαλείων λογισμικού τα οποία μπορούν να προσαρμοσθούν στις απαιτήσεις του χρήστη.

2.1.4.3 Το Υλικό

Το υλικό είναι οι συσκευές εκείνες που είναι απαραίτητες για την υλοποίηση ενός ΓΠΣ, φυσικά πάντα συμπεριλαμβάνεται ένας ισχυρός υπολογιστή που να μπορεί να τρέχει το απαιτητικό λογισμικό των ΓΠΣ, να μπορεί να αποθηκεύει και να επεξεργάζεται το μεγάλο όγκο των δεδομένων και μονάδες εισόδου των δεδομένων και εξόδου των αποτελεσμάτων (Maguire,1991).

Το βασικό υλικό από το οποίο αποτελείται ένα ΓΠΣ περιλαμβάνει:

- Κεντρικό επεξεργαστή (Central processing unit = CPU).
- Μονάδες εισόδου των χωρικών δεδομένων: οθόνη γραφικών, ψηφιοποιητές (digitizers), σαρωτές (scanners).
- Μονάδες εισόδου μη χωρικών δεδομένων: Πληκτρολόγιο.
- Μονάδες παρουσίασης των δεδομένων: Εκτυπωτές (printers), σχεδιαστές (plotters).
- Μονάδες αποθήκευσης των δεδομένων: Μαγνητικοί και Οπτικοί Δίσκοι
- Συσκευή εντοπισμού της θέσης των χωρικών οντοτήτων (GPS).

2.1.4.4 Επεξεργασία

Οι επεξεργασίες περιλαμβάνουν το τρόπο με τον οποίο θα ανακτηθούν τα δεδομένα, θα εισαχθούν στο σύστημα, θα αποθηκευθούν, θα αναλυθούν, θα μετασχηματισθούν και τελικά θα παρουσιασθούν σε τελική μορφή. Οι επεξεργασίες αφορούν τα βήματα τα πρέπει να ακολουθηθούν για να απαντηθούν οι ερωτήσεις των χρηστών. Η δυνατότητα των ΓΠΣ για χωρική ανάλυση και στο να απαντούν χωρικές ερωτήσεις είναι αυτό που τα διαφοροποιεί από τα άλλα πληροφοριακά συστήματα. Ο μετασχηματισμός των δεδομένων περιλαμβάνει διαδικασίες όπως η διόρθωση του συστήματος νεοαναφοράς, διόρθωση των λαθών της ψηφιοποίησης και τη μετατροπή από διανυσματική σε πλεγματική μορφή και αντίστροφα κ.α.

2.1.4.5 Ο Άνθρωπος

Οι άνθρωποι είναι το μέρος του ΓΠΣ που κάνει τη δουλειά. Περιλαμβάνει ένα πλήθος από ειδικότητες: managers, διαχειριστές των ΒΔ, ειδικούς των εφαρμογών, προγραμματιστές κ.λπ.

Οι Κόλια και συν. (2012) αναφέρουν ότι : "Ένα ΓΠΣ είναι ένα μέσο επεξεργασίας της γεωγραφικής πληροφορίας, αλλά από μόνο του δεν μπορεί να εγγυηθεί ότι αυτή η πληροφορία θα χρησιμοποιηθεί ικανοποιητικά. Δεν αρκεί δηλαδή ένας οργανισμός να πάρει ένα ΓΠΣ, να εκπαιδεύσει μερικούς ενθουσιώδεις υπαλλήλους και να περιμένει αυτόματη επιτυχία. Όπως και κάθε καινούργιο εργαλείο, ένα ΓΠΣ μπορεί να είναι επιτυχημένο αν ενσωματωθεί στις διαδικασίες του οργανισμού (ChanandWilliamson, 1996). Δηλαδή δεν χρειάζονται μόνο επενδύσεις σε λογισμικό και υλικό αλλά και εκπαίδευση του προσωπικού, διευθυντικού και υπαλληλικού, για να χρησιμοποιήσει την καινούργια τεχνολογία. Επιπλέον ένας οργανισμός πρέπει επίσης να ορίσει τους ελέγχους και τους μηχανισμούς ώστε οι διαδικασίες του ΓΠΣ αφενός να εξυπηρετούν τις ανάγκες του οργανισμού, αφετέρου να μην ξεφύγουν από το οικονομικό του πλαίσιο (PintoandAzad, 1994)".

Επίσης είναι απαραίτητοι οι άνθρωποι που σχεδιάζουν και προγραμματίζουν το ΓΠΣ, το εφοδιάζουν με δεδομένα και να ερμηνεύουν τα αποτελέσματα. Οι άνθρωποι των ΓΠΣ πρέπει να έχουν διάφορες ικανότητες, ανάλογα με τον τομέα που δουλεύουν. Μπορεί να είναι ειδικοί στη συλλογή των δεδομένων, στην ανάπτυξη προγραμμάτων λογισμικού κ.λπ., πωλητές λογισμικού, εκπαιδευτές ΓΠΣ, χρήστες ΓΠΣ κ.λπ.

2.1.4.6. Το Διαδίκτυο

Το 1972, το υπουργείο Εθνικής Άμυνας των ΗΠΑ, ξεκίνησε ένα πρόγραμμα το οποίο είχε το όνομα ARPANET και αποτέλεσε τον αρχική μορφή αυτού που σήμερα γνωρίζουμε ως διαδίκτυο. Η άνθηση ήρθε τη δεκαετία του 1980 όταν και άρχισαν να αναπτύσσονται οι πρώτες εφαρμογές που έφεραν το WorldWideWeb στην καθημερινή ζωή. Οι γεωγράφοι και όσοι ασχολιόντουσαν εκείνη την εποχή με τα ΓΠΣ, αντελήφθησαν την αξία του διαδικτύου και ο πρώτος Web χάρτης εμφανίσθηκε το 1993. Από τότε το διαδίκτυο και τα ΓΠΣ έχουν επηρεάσει σημαντικά το ένα την ανάπτυξη του άλλου. απεδείχθησαν σημαντική εφαρμογή που ώθησε πολύ κόσμο να χρησιμοποιήσει το διαδίκτυο και αντίστροφα τα ΓΠΣ ωφελήθηκαν από τη δημοσιότητα του διαδικτύου. Σήμερα υπάρχουν πάρα πολλές εφαρμογές ΓΠΣ στο διαδίκτυο, πολλές από τις οποίες τρέχουν σε προσωπικούς υπολογιστές και έχουν την δυνατότητα να παρέχουν online χωρική πληροφορία.Το πιο βασικό στοιχείο ενός ΓΠΣ είναι το διαδίκτυο (Internet), χωρίς το οποίο δεν μπορούμε να έχουμε ταχεία επικοινωνία και ανταλλαγή πληροφοριών. Σήμερα τα ΓΠΣ εξαρτώνται σημαντικά από το διαδίκτυο. Το διαδίκτυο αρχικά σχεδιάσθηκε σαν ένα δίκτυο Η/Υ, αλλά γρήγορα έγινε το κοινωνικό μέσο ανταλλαγής πληροφοριών.

2.2 Η γλώσσα προγραμματισμού Python

2.2.1 Γενικά

Η γλώσσα προγραμματισμού Python είναι μια γλώσσα που θα μπορούσε κανείς να ισχυριστεί ότι είναι πολύ απλή, αλλά παράλληλακαι εξαιρετικά ισχυρή. Η γλώσσα προγραμματισμού Python δίνει το προνόμιο στο χρήστη να μπορεί να συγκεντρωθεί στη λύση του προβλήματος και όχι σε κάποιο αυστηρό συντακτικό ή την σωστή δομή της γλώσσας στην οποία προγραμματίζετε.

Έχει αποδοτικές δομές δεδομένων υψηλού επιπέδου και μια απλή αλλά αποτελεσματική προσέγγιση στον αντικειμενοστρεφή προγραμματισμό. Η απλή σύνταξη η δυναμική χρήση της μνήμης, μαζί με τη λειτουργία της ως διερμηνευόμενης (αντί μεταγλωττιζόμενης) γλώσσας, την καθιστούν την ιδανική γλώσσα μικροεφαρμογών (scripts), γι αυτό και πολλές φορές αναφέρεται ως ‘σκριπτογλώσσα’.

Χαρακτηριστικάτης Python:

- **Απλή :** Η Python είναι μια απλή και ‘μινιμαλιστική’ γλώσσα. Το διάβασμα ενός καλού προγράμματος γραμμένο σε Python είναι σαν το διάβασμα οδηγιών σε φυσική γλώσσα. Η σύνταξή της μοιάζει αρκετά με την σύνταξη του ψευδοκώδικα, χαρακτηριστικό που αποτελεί ένα από τα πιο ισχυρά σημεία της.
- **Εύκολη εκμάθηση :** Λόγο της απλότητας της, και του χαλαρού τρόπου σύνταξης, επιτρέπει την εύκολη και γρήγορη εκμάθησή της
- **Ελεύθερη και Ανοικτού Κώδικα :** Η Python είναι ένα τυπικό παράδειγμαΜεαπλά λόγια, μπορείτε να διανείμετε αντίγραφα αυτού του λογισμικού, να διαβάσετε τον πηγαίο κώδικάτου, να κάνετε αλλαγές σ’ αυτό και να χρησιμοποιήσετε κομμάτια του σε νέα ελεύθερα προγράμματα. ΤοΕΛΛΑΚ βασίζεται στην ιδέα μιας κοινότητας που μοιράζεται τη γνώση. Η Pythonσαν γλώσσα προγραμματισμού είναι μια δυναμική γλώσσα, δημιουργήθηκε και βελτιώνεται συνεχώς από μια κοινότητα που το μόνο που θέλει να χρησιμοποιεί μια βελτιούμενη, μέρα με τη μέρα Python.
- **Γλώσσα υψηλού επιπέδου :** Όταν γράφετε προγράμματα στην Python, δε χρειάζεται ποτέ να νοιάζεστε για τις χαμηλού επιπέδουλεπτομέρειες όπως η διαχείριση της μνήμης που χρησιμοποιείται από τα προγράμματά σας. Αν και

αυτό πολλοί το θέτουν και ως μειονέκτημα, στην πραγματικότητα έχουν αναπτυχτεί βιβλιοθήκες διασύνδεσης με γλώσσες χαμηλού επιπέδου. Έτσι δίνεται η δυνατότητα στον χρήστη ανάλογα με τις απαιτήσεις να διευθετήσει λεπτομέρειες όπως η διαχείριση της μνήμης, σαν να γράφει με γλώσσα χαμηλού επιπέδου.

- **Φορητή :** Λόγω του ανοικτού της κώδικα, η Python έχει υλοποιηθεί (αλλάχθηκε για να λειτουργεί) σε πολλές πλατφόρμες λειτουργικών συστημάτων. Όλα τα προγράμματά που είναι γραμμένα σε Python, μπορούν να δουλέψουν σε οποιαδήποτε από αυτές τις πλατφόρμες λειτουργικών συστημάτων, χωρίς να χρειάζονται καθόλου αλλαγές αν είστε αρκετά προσεκτικοί ώστε να αποφύγετε να χρησιμοποιήσετε χαρακτηριστικά που εξαιρούνται κάθ'έκαστο λειτουργικό σύστημα. Μπορείτε να χρησιμοποιήσετε την Python στο Linux, στα Windows, στο FreeBSD, σε Macintosh, στο Solaris, στο OS/2, στην Amiga, στο AROS, στο AS/400, στο BeOS, στο OS/390, στο z/OS, στο Palm OS, στο QNX, στο VMS, στο Psion, στο Acorn RISC OS, στο VxWorks, σε PlayStation, στο SharpZaurus, στα Windows CE.
- **Διερμηνευόμενη :** Ένα πρόγραμμα που γράφεται σε μια μεταγλωττιζόμενη γλώσσα όπως η C ή η C++ μετατρέπεται σε πηγαία γλώσσα μηχανής, (δυαδικός κώδικας δηλαδή 0 και 1), χρησιμοποιώντας ένα μεταγλωττιστή με διάφορες σημαίες καιεπιλογές. Όταν τρέχετε το πρόγραμμα, ο συνδέτης αντιγράφει το πρόγραμμα στη μνήμη και αρχίζει νατο τρέχει. Η γλώσσα Python, από την άλλη, δε χρειάζεται μεταγλώττιση σε δυαδικό αρχείο. Απλά τρέχετετο πρόγραμμα απ'ευθείας από τον πηγαίο κώδικα. Εσωτερικά, η Python μετατρέπει τον πηγαίο κώδικα σε μια ενδιάμεση μορφή που ονομάζεται bytecode και μετά το μεταφράζει στη γλώσσα του υπολογιστή και μετά το τρέχει. Όλο αυτό, στην πραγματικότητα κάνει τη χρήση της Python πολύ πιο εύκολη αφού δε χρειάζεται να ανησυχείτε για τη μεταγλώττιση του προγράμματος, τη σύνδεση με τις κατάλληλες βιβλιοθήκες, κ.α. Αυτό επίσης κάνει τα προγράμματα της Python εξαιρετικά φορητά, αφού μπορείτε απλά να αντιγράψετε το πρόγραμμα Python που φτιάξατε σε έναν άλλο υπολογιστή.
- **Αντικειμενοστρεφής :** Η Python υποστηρίζει τόσο το διαδικασιοστρεφή προγραμματισμό (procedure-oriented) όσο και τον αντικειμενοστρεφή προγραμματισμό (object-oriented).

Στο διαδικασιοστρεφή προγραμματισμό, το πρόγραμμα δομείται πάνω σε διαδικασίες ή συναρτήσεις οι οποίες δεν είναι τίποτε άλλο από επαναχρησιμοποιήσιμα κομμάτια από προγράμματα. Στις αντικειμενοστρεφείς γλώσσες, το προγράμματα δομούνται πάνω σε αντικείμενα τα οποία συνδυάζουν τα δεδομένα με τη λειτουργικότητα. Η Python έχει έναν πολύ ισχυρό αλλά πολύ απλό τρόπο για αντικειμενοστρεφή προγραμματισμό, ειδικά όταν συγκρίνεται με μεγάλες γλώσσες όπως η C++ ή η Java. Καθώς στην Python, τα πάντα είναι αντικείμενα, ακόμα και οι συναρτήσεις, οι γεννήτορες (generators), οι εξαιρέσεις και ότι άλλο μπορεί να χρησιμοποιηθεί. Έτσι, προσφέρεται ένας διαισθητικός τρόπος συγγραφής των προγραμμάτων μας με συνέπεια στην αντικειμενοστρέφεια, παρόλο που υποστηρίζονται και άλλοι τρόποι προγραμματισμού (όπως συναρτησιακός, διαδικαστικός κ.α.).Στα βασικά χαρακτηριστικά της γλώσσας θα συναντήσουμε των χειρισμό εξαιρέσεων, πακέτα, κλάσεις, συναρτήσεις, γεννήτορες (ειδική περίπτωση συναρτήσεων) με ένα τρόπο όπου κάθε ένα θα συμπληρώνει αρμονικά το άλλο για την διευκόλυνση του προγραμματιστή. Αυτό το δέσιμο καθοδηγείται πάντα από την συνέπεια στην αρχή της Python ότι θα έπρεπε να υπάρχει ένας και μόνο ένας προφανής τρόπος για να γίνει κάτι, οδηγώντας έτσι στην απλοποίηση των προβλημάτων που μπορεί να αντιμετωπίσει ένα προγραμματιστής.

- **Επεκτάσιμη :** Αν χρειάζεστε ένα κρίσιμο κομμάτι κώδικα να τρέχει πολύ γρήγορα ή αν πρέπει να έχετε ένα κομμάτι ενός αλγόριθμου που να μην είναι ανοικτό, τότε μπορείτε να προγραμματίσετε εκείνο το κομμάτι σε C ήC++ ή java ή ακόμη και Cuda και μετά να το χρησιμοποιείτε από το Python πρόγραμμά σας διατηρώντας τα περισσότερα από τα πλεονεκτήματα των άλλων γλωσσών.
- **Ενσωματώσιμη :** Μπορείτε να ενσωματώσετε την Python μέσα σε προγράμματα γραμμένα σε άλλες γλώσσες όπως C ή C++ ή ακόμα και να ενσωματωθεί σε πολλά εμπορικά προγράμματα όπως στην παρούσα εργασία στο ArcGIS10.1 έχοντας τις δυνατότητες του 'scripting'.
- **Εκτεταμένες βιβλιοθήκες :** Η Πρότυπη βιβλιοθήκη της Python είναι πραγματικά τεράστια. Μπορεί να σας βοηθήσει να κάνετε διάφορα πράγματα σχετικά με κανονικές εκφράσεις, δημιουργία τεκμηρίωσης, δοκιμές μονάδων,

νημάτωση, βάσεις δεδομένων, περιηγητές ιστού, CGI, FTP, email, XML, XML-RPC, HTML, αρχεία WAV, κρυπτογράφηση, γραφικές διεπαφές χρήστη (GUI : Graphical User Interfaces), και άλλα πράγματα που εξαρτούνται από το σύστημα. Θυμηθείτε ότι όλα αυτά είναι διαθέσιμα όποτε είναι εγκατεστημένη η Python. Επιπλέον από την πρότυπη βιβλιοθήκη, υπάρχουν διάφορες άλλες βιβλιοθήκες υψηλής ποιότητας όπως η wxPython, η Twisted, η PythonImagingLibrary και πολλές άλλες.

(παράθεση από το βιβλίο ABiteofPython)

2.2.2 Μεταβλητές και Βασικοί τελεστές

Μεταβλητές είναι ο χώρος στην μνήμη που αποθηκεύονται προσωρινά δεδομένα. Κάθε τιμή κατά την διάρκεια της λειτουργίας ενός προγράμματος είναι αποθηκευμένη σε κάποια μεταβλητή.

Η λειτουργία των μεταβλητών μοιάζει με αυτή των συρταριών ενός γραφείου. Εκεί, μπορούν να αποθηκεύονται διάφορα αντικείμενα, όπως κάθε μεταβλητή περιέχει κάποια τιμή. Όταν γίνεται προσπέλαση μια τιμής μια μεταβλητής είναι σαν να ανοίγει ένα συρτάρι. Στον υπολογιστή με τον ορισμό μιας μεταβλητή και το είδος της, στη μνήμη δημιουργείται ένας χώρος ώστε να δεχτεί τα δεδομένα κατά την διάρκεια της επεξεργασίας.

Στην Python χωρίζουμε τα είδη των μεταβλητών ανάλογα με το εμβέλεια δράσης τους σε τοπικές και Γενικές (Global) μεταβλητές.

- Τοπικές μεταβλητές

Όταν μία μεταβλητή δηλώνεται μέσα σε μία ορισμένη συνάρτηση, αυτή δεν έχει καμία σχέση με οποιαδήποτε άλλη μεταβλητή εκτός της συνάρτησης ακόμη κι αν έχει την ίδια ονομασία και χρησιμοποιείται έξω από αυτή τη συνάρτηση, δηλαδή η χρησιμότητα των μεταβλητώνπεριορίζεται τοπικάεντός της συνάρτησης. Αυτό ονομάζεται εμβέλεια(scope) των μεταβλητών. Όλες οι μεταβλητές έχουν την εμβέλεια του τμήματος όπου έχουν δηλωθεί, αρχίζοντας από το σημείο στο οποίο ορίζεται το όνομα.

- Γενικές (Global) μεταβλητές

Είναι κάθε εντολή εκτός συνάρτησης. Εκτός αν μια μεταβλητή ορισθεί μέσα σε κάποια συνάρτηση αλλά οριστεί ως γενική, αυτό γίνεται με τη χρήση της εντολής global.

Χρησιμοποιώντας την εντολή global γίνεται ξεκάθαρο ότι η μεταβλητή μπορεί να χρησιμοποιηθεί από κάθε εντολή βρίσκεται εντός ή εκτός της προκείμενης συνάρτησης. Εδώ πρέπει να προστεθεί ότι αυτού του είδους οι μέθοδοι πρέπει να αποφεύγονται και να χρησιμοποιούνται μόνο σε εξαιρετικές περιπτώσεις, καθώς αυξάνει την πολυπλοκότητα του προγράμματος και το καθιστά δύσκολο στο να διαβαστεί.

- Τελεστές

Τελεστής	Όνομα	Εξήγηση
+	Συν	Προσθέτει δύο αντικείμενα.
-	Μείον	Είτε δίνει έναν αρνητικό αριθμό, ή αφαιρεί έναν αριθμό από έναν άλλο.
*	Επί	Δίνει το γινόμενο δύο αριθμών ή μία συμβολοσειρά (string) επαναλαμβανόμενη τόσες φορές.
**	Δύναμη	Επιστρέφει το x υψωμένο στη δύναμη y.
/	Διά	Διαιρεί το x με το y.
//	Διαίρεση στρογγυλοποιημένη προς τα κάτω (Floor Division)	Επιστρέφει τον κοντινότερο (προς τα κάτω) ακέραιο στο πηλίκο.
%	Υπόλοιπο	Επιστρέφει το υπόλοιπο της διαίρεσης.
<<	Αριστερή μετάθεση	Μεταθέτει τα δυαδικά ψηφία (bits) του αριθμού προς τα αριστερά κατά το πλήθος των θέσεων που καθορίστηκε. (Κάθε αριθμός αναπαρίσταται στη μνήμη με δυαδικά ψηφία (bits, binary digits) -δηλαδή με 0 και 1).
>>	Δεξιά μετάθεση	Μεταθέτει τα bits του αριθμού προς τα δεξιά κατά το πλήθος των θέσεων που καθορίστηκε.

&	Δυαδικό ΚΑΙ	Δυαδικό ΚΑΙ των αριθμών.
|	Δυαδικό Ή	Δυαδικό Ή των αριθμών.
^	Δυαδικό αποκλειστικό Ή	Δυαδικό αποκλειστικό Ή των αριθμών.
~	Δυαδική αντιστροφή	Το δυαδικό αντίστροφο του x είναι -(x+1).
<	Μικρότερο από	Επιστρέφει το αν το x είναι μικρότερο από το y. Όλοι οι τελεστές σύγκρισης επιστρέφουν True (Αληθής) ή False (Ψευδής). Σημειώστε ότι τα ονόματα αυτά ξεκινούν με κεφαλαίο.
>	Μεγαλύτερο από	Επιστρέφει το αν το x είναι μεγαλύτερο Από το y.
<=	Μικρότερο ή ίσο	Επιστρέφει το αν το x είναι μικρότερο από ή ίσο με το y.
>=	Μεγαλύτερο ή ίσο	Επιστρέφει το αν το x είναι μεγαλύτερο από ή ίσο με το y.
==	Ίσο	Συγκρίνει αν τα αντικείμενα είναι ίσα. Το x = 2; y = 2; x == y επιστέφει True.
!=	Διαφορετικό	Συγκρίνει αν τα αντικείμενα ΔΕΝ είναι ίσα.
not	Λογικό ΌΧΙ	Αν το x είναι True, επιστρέφει False. Αν το x είναι False, επιστρέφει True.

and	Λογικό ΚΑΙ	Το xandy επιστρέφει False αν το x είναι False, αλλιώς επιστρέφει υπολογίζει και επιστρέφει την τιμή του y.
or	Λογικό Ή	Αν το x είναι True, επιστρέφει True, αλλιώς υπολογίζει και επιστρέφει την τιμή του y.

(παράθεση από το βιβλίο ABiteofPython)

2.2.3 Έλεγχος Ροής εκτέλεσης

- Η εντολή if

Η εντολή if χρησιμοποιείται για να ελεγχθεί μια συνθήκη. Αν η συνθήκη που ακολουθεί είναι αληθής, τότε εκτελείται ένα σύνολο εντολών που ονομάζεται if-block, διαφορετικά πάμε στην εντολή elif όπου ελέγχετε η εκεί συνθήκη και αν είναι αληθής, αν ισχύει τότε εκτελείται το σύνολο των εντολών που ονομάζεται elif-block, τέλος στην εντολή else εκτελείτε ένα άλλο σύνολο εντολών που ονομάζεται else-block. Η χρήση των όρων elif και else είναι προαιρετική.

```
x = -6
if x > 0:
  print "Positive"
elif x == 0:
  print "Zero"
else:
  print "Negative"
```

Εικόνα 2.3: Παράδειγμα της δομής επιλογής if

- Η δομή επανάληψης while

Αποτελεί δομή επανάληψης καθώς η εντολή whileεπιτρέπει να εκτελεστεί επανειλημμένα ένα σύνολο εντολών, όσο η λογική συνθήκη δίπλα στην εντολή while παραμένει αληθές. Όταν η λογική συνθήκη δίπλα στην εντολή while δεν είναι αληθής το σύνολο των εντολών δεν θα εκτελεστεί.

```
x = 5
while x > 0:
    print x
    x = x - 1
```

Εικόνα 2.4: Παράδειγμα της δομής while

- Η δομή επανάληψης do…while

Είναι και αυτή μια δομή επανάληψης παραλλαγή της while στην οποία εκτελούνται επανειλημμένα οι εντολές ανάμεσα στο do και το while, όσο η λογική συνθήκη παραμένει αληθές.

Μοναδική διαφορά είναι ότι το σύνολο των εντολών θα εκτελεστεί τουλάχιστον μία φορά πριν γίνει έλεγχος της συνθήκης.

- Η δομή επανάληψης for

Η δομή επανάληψηςfor..in είναι άλλη μία δομή επιλογής, στην οποία επίσης υπάρχει επανειλημμένη εκτέλεση μιας ομάδας εντολών αλλά αυτή τη φορά σε προκαθορισμένο αριθμό επαναλήψεων. Σε αυτή την δομή επιλογής ορίζεται η αρχική τιμή και η τελική τιμή ενός δείκτη καθώς και το βήμα που θα αυξάνεται ή θα μειώνεται η τιμή του δείκτη αυτού.

```
for i in range(10, 0, -2):
    print i
```

Εικόνα 2.5: Παράδειγμα της δομής επανάληψης for

- Η εντολή break

Η εντολή break χρησιμοποιείται για τη διακοπή μιας επαναλαμβανόμενης διαδικασίας εκτέλεσης εντολών, σε κάθε περίπτωση και ανεξάρτητα από το αν είναι αληθής ή ψευδής η συνθήκη στο βρόγχο επιλογής ή επανάληψης.

- Η εντολή continue

Η εντολή continue χρησιμοποιείται για την παράληψη των υπολοίπων εντολών στο τμήμα των εντολών που εκτελούνται σε κάθε επανάληψη και να συνεχίσει με την επόμενη επανάληψη του βρόχου.

```
while True:
    s = raw_input('Εισάγετε κάτι : ')
    if s == 'quit':
        break
    if len(s) < 3:
        print('Πολύ μικρό')
        continue
    print('Το μήκος των εισαχθέντων είναι επαρκές')
```

Εικόνα 2.6: Παράδειγμα εντολών brake και continue

2.2.4 Συναρτήσεις

Οι συναρτήσεις αποτελούν ξεχωριστά μέρη ενός προγράμματος.

Έχουν δικό τους όνομα παίρνουν σαν είσοδο συγκεκριμένο αριθμό τιμών και επιστρέφουν συγκεκριμένες τιμές. Ορίζουν μια διαδικασία επεξεργασίας που μπορεί να χρησιμοποιηθεί πολλές φορές σε διάφορα σημεία το κώδικα.

Οι συναρτήσεις ορίζονται χρησιμοποιώντας τη δεσμευμένη λέξη def, μετά την οποία ακολουθεί ένα όνομα το χαρακτηριστικό όνομα της συνάρτησης και κατόπιν ακολουθεί εντός παρενθέσεων τα ονόματα μεταβλητών εκείνων, που θα δεχτούν τις τιμές τις εισόδου όταν θα καλεστεί η συγκεκριμένη συνάρτηση.

Αυτές οι παράμετροι μοιάζουν με τις μεταβλητές, διαφέροντας ως προς το ότι οι τιμές αυτών των μεταβλητών ορίζονται όταν καλούμε τη συνάρτηση και τους έχουν ήδη εκχωρηθεί τιμές όταν τρέχει η συνάρτηση. Οι παράμετροι που καθορίζονται μέσα στις παρενθέσεις στον ορισμό της συνάρτησης και διαχωρίζονται με κόμμα. Όταν καλούμε τη συνάρτηση δίνουμε και τις τιμές με τον ίδιο τρόπο.

```
def printMax(a, b):
    if a > b:
        print(a, 'είναι το μέγιστο')
    elif a == b:
        print(a, 'είναι ίσο με το', b)
    else:
        print(b, 'είναι το μέγιστο')


printMax(3, 4)
```

Εικόνα 2.7: Παράδειγμα συνάρτησης

2.2.5 Δομές δεδομένων

Οι δομές δεδομένων όπως λέει και το όνομά τους, είναι δομέςπου μπορούν να κρατήσουν μαζικά μερικά δεδομένα. Με άλλα λόγια χρησιμοποιούνται για να αποθηκεύουν δεδομένα που έχουν σχέση μεταξύ τους.

Στη βασική βιβλιοθήκη της Python, υπάρχουν ενσωματωμένες τέσσερις δομές δεδομένων, οι λίστες, οι πλειάδες, τα λεξικά και τα σύνολα. Η χρήση τους έχει σχέση με την αιτία που χρησιμοποιούνται.

- Λίστα

Οι λίστεςείναι μια δομή δεδομένων που μπορεί να χρησιμοποιηθεί για αποθηκεύσει έναν μεγάλο αριθμό όμοιων στοιχείων, αριθμών συμβολοσειρών ή άλλων λιστών.

Η λίστα δηλώνεται με το όνομα της λίστας ένα = και μετά αγκύλες ([ ]) αν η λίστα είναι κενή. Αν η λίστα δεν είναι κενή τότε εντός των αγκυλών ([ ]) δίνονται τα στοιχεία της λίστας που θα υπάρχουν αρχικά μέσα σε αυτή. Ύστερα από τη δημιουργία μιας λίστας μπορούν να προστεθούν, να μετακινηθούν ή να προσπελαστούν τα στοιχεία μέσα σε αυτή τη λίστα

```
# Αυτή είναι η λίστα αγορών μου
shoplist = ['μήλο', 'μάνγκο', 'καρότο', 'μπανάνα']

# -*- coding: utf-8 -*-
# Filename: using_list.py

for item in shoplist:
    print(item, end=' ')
```

Εικόνα 2.8: Παράδειγμα λίστας

- Πλειάδα

Οι πλειάδες είναι μια δομή δεδομένων που χρησιμοποιούνται για να συγκρατήσουν μαζί πολλαπλά αντικείμενα. Ένα κύριο χαρακτηριστικό των πλειάδων είναι ότι είναι αμετάβλητες όπως οι συμβολοσειρές, δηλαδή δε μπορούν να τροποποιηθούν.

```
zoo = ('πύθωνας', 'ελέφαντας', 'πιγκουίνος')
 είναι προαιρετικές
print('Ο αριθμός των ζώων στο ζωολογικό κήπο είναι', len(zoo))
```

Εικόνα 2.9: Παράδειγμα πλειάδας

- Λεξικό

Ένα λεξικό είναι σαν ένας κατάλογος για ψώνια όπου υπάρχουν οι ονομασίες των προϊόντων που θέλουμε να αγοράσουμε και δίπλα ο αριθμός των τεμαχίων. Υπάρχει άμεση συσχέτιση των κλειδιών (ονομασίες προϊόντων) με τιμές (αριθμό τεμαχίων). Το κλειδί σε κάθε περίπτωση είναι μοναδικό. Ενώ οι τιμές (αριθμός των τεμαχίων) μπορεί να είναι λίστα, ή άλλα λεξικά .

Θυμηθείτε ότι τα ζευγάρια κλειδί-τιμή σε ένα λεξικό δεν ταξινομούνται με κανένα τρόπο. Αν χρειάζεται να γίνει κάποιο είδος ταξινόμησης, τότε αυτή η διαδικασία πρέπει γίνει πριν καταχωρηθούν τα στοιχεία σε κάποιο λεξικό.

```
ab = {  'Swaroop'   : 'swaroop@swaroopch.com',
        'Larry'     : 'larry@wall.org',
        'Matsumoto' : 'matz@ruby-lang.org',
        'Spammer'   : 'spammer@hotmail.com'
    }

print("Swaroop's address is", ab['Swaroop'])
```

Εικόνα 2.10: Παράδειγμα λεξικού

2.2.6 Κλάσεις και Αντικείμενα

Ο αντικειμενοστραφής προγραμματισμός αποτελεί μια προγραμματιστική τεχνική όπου βασιζόμαστε στα αντικείμενα και στις μεταξύ τους αλληλεπιδράσεις. Έτσι, στον αντικειμενοστραφή προγραμματισμό, μπορούμε να δούμε ένα πρόγραμμα επικεντρώνοντας στην οπτική μιας συλλογής αντικειμένων και των αλληλεπιδράσεων μεταξύ τους. Κάθε αντικείμενο μπορεί να αλληλεπιδράσει με τα υπόλοιπα αντικείμενα δίνοντας στοιχεία και περνώντας στοιχεία από άλλα αντικείμενα, επεξεργαζόμενο δεδομένα και γενικά σαν μια ανεξάρτητη οντότητα που όμως βρίσκεται σε ένα περιβάλλον αλληλεπίδρασης και συνεργασίας με τα υπόλοιπα στοιχεία που απαρτίζουν το πρόγραμμα. Για αυτό τον λόγο και οι συναρτήσεις που χρησιμοποιούνται στα αντικείμενα, είναι στενά συνδεδεμένες με τα ίδια τα αντικείμενα. Κλάσεις και αντικείμενα είναι οι δυο κύριες πτυχές του αντικειμενοστραφούς προγραμματισμού. Πιο συγκεκριμένα η κλάση δημιουργεί ένα νέο τύπο, όπου τα αντικείμενα είναι στιγμιότυπα της κλάση.

Τα βασικά χαρακτηριστικά που παρατηρούμε κατά τον σχεδιασμό και την υλοποίηση ενός τυπικού προγράμματος μέσω αντικειμενοστρέφειας είναι:

- **Κλάση (Class):** Καθορίζει τα βασικά χαρακτηριστικά (attributes) και συμπεριφορές (method) των αντικειμένων που περιγράφονται από αυτή την κλάση. Ένα παράδειγμα θα μπορούσε να είναι η κλάση σκύλος. Γενικότερα μπορούμε να σκεφτόμαστε ότι αντιπροσωπεύουν τα αδρά χαρακτηριστικά που ενυπάρχουν σε όλα τα στιγμιότυπα μιας κλάσης
- **Αντικείμενο (object):** Μια συγκεκριμένη μορφή μιας κλάσης, δηλαδή ένα από τα αντικείμενα που περιγράφει η κλάση που το χαρακτηρίζει.
- **Στιγμιότυπο (Instance):** Η κατάσταση ενός αντικείμενο κατά την εκτέλεση του προγράμματος μέσα από την συγκεκριμενοποίηση του.
- **Μέθοδος (Method):** Η συμπεριφορά και οι δυνατότητες ενός αντικειμένου.
- **Κληρονομικότητα (Inheritance):** Όταν μια κλάση αποτελεί εξειδίκευσημιας άλλης, και έτσι έχει όλα τα χαρακτηριστικά που περιγράφουν την κλάση από την οποία κληρονομεί, καθώς και κάποια ιδιαίτερα που περιγράφουν ειδικά την συγκεκριμένη ομάδα αντικειμένων. Για παράδειγμα η κλάση

(παράθεση από το βιβλίο ABiteofPython)

2.2.7 Εξαιρέσεις

Ο χειρισμός εξαιρέσεων είναι μια κατασκευή η οποία επιτρέπει στον προγραμματιστή να χειριστεί ειδικές συνθήκες που αλλάζουν την φυσιολογική ροή του προγράμματος.

Οι εξαιρέσεις χρησιμοποιούνται μόνο σε ειδικές περιπτώσεις που η ροή του προγράμματος αλλάζει εξαιτίας ενός συγκεκριμένου συμβάντος, που συνήθως δεν μπορεί να πραγματοποιηθεί. Ένα σύνηθες παράδειγμα είναι η περίπτωση όπου δεν μπορεί να ανοίξει ένα αρχείο ώστε να προσπελαστούν τα περιεχόμενα του. Γενικά οι εξαιρέσεις δημιουργούνται όταν κάτι δεν πάει σύμφωνα με τον αρχικό προγραμματισμό εκτέλεσης του προγράμματος

Ο γενικότερος μηχανισμός που λειτουργούν οι εξαιρέσεις είναι:

1. Αν κάτι πάει στραβά η Python εγείρει μια εξαίρεση.
2. Αναγνωρίζεται η κλάση της εξαίρεσης (πχ name Error, type Error, index Error).
3. Εκτελείται το ειδικό κώδικα που έχει γραφτεί για αυτή την περίπτωση.
4. Αν δεν είναι δυνατό να χειριστεί μια εξαίρεση που εγείρεται, τότε το πρόγραμμα μας σταματάει την εκτέλεση του.
5. Εξαιρέσεις που δεν έχουν χειριστεί, εμφανίζονται στο traceback, όπου δηλώνεται από το πρόγραμμα η αιτία της παύσης της εκτέλεσης του κώδικα.

(παράθεση από το βιβλίο ABiteofPython)

2.3 Βαριογραφία

2.3.1 Μη γεωστατιστικές μέθοδοι χωρικής παρεμβολής

Σύμφωνα με τους Burrough και McDonnell οι μέθοδοι χωρικής παρεμβολής διαιρούνται σε δύο ομάδες, τις γενικές και τις τοπικές. Οι γενικές μέθοδοι παρεμβολής χρησιμοποιούν όλα τα διαθέσιμα δεδομένα για να κάνουν εκτιμήσεις σε ολόκληρη την εξεταζόμενη περιοχή, ενώ οι τοπικές μέθοδοι βασίζονται στον πρώτο νόμο της γεωγραφίας (Tobler, 1970) και 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 δυνατών πιο ακριβείς. Τέτοιες είναι οι μέθοδοι αντίστροφης απόστασης (Inverse Distance Weighted) και η γεωγραφικά σταθμισμένη παλινδρόμηση (Geographically Weighted Regression), η οποία συνέβαλε στη μοντελοποίηση χωρικών ετερογενών διεργασιών **(**Fotheringhametal, 2002).

2.3.2 Γεωστατιστική

Στις προαναφερόμενες μεθόδους χωρικής παρεμβολής, η ποιότητα των προβλέψεων δεν μπορεί να εκτιμηθεί καθώς, δεν υπάρχουν οι ανάλογες τεχνικές εκτίμησης. Ο μόνος τρόπος αξιολόγησης της καταλληλότητας της πρόβλεψης είναι ο υπολογισμός των εκτιμώμενων τιμών για μια ομάδα σημείων που δεν χρησιμοποιήθηκαν κατά την εφαρμογή της μεθόδου χωρικής παρεμβολής (Burrough & McDonnell, 1998).

Επίσης, δεν υπάρχει για αυτές τις μεθόδους κάποια τεχνική που να υποδεικνύει ποιες είναι οι καλύτερες τιμές παραμέτρων που μπορεί να επιλεχθούν, με στόχο την τελειοποίηση του αποτελέσματος. Τέτοιες παράμετροι είναι ο αριθμός των σημείων που χρειάζεται να συνεισφέρουν στην πρόβλεψη, το μέγεθος, ο προσανατολισμός και το σχήμα της γειτονίας αναζήτησης. Επιπλέον, τα βάρη υπολογίζονται με απλουστευμένο τρόπο ως απλή συνάρτηση της απόστασης, ενώ δεν εξετάζονται τα σφάλματα που σχετίζονται με τις τιμές που «παρεμβάλλονται».

Στα μέσα της δεκαετίας του ’60 οι G.Matheron και D.G. Krige ανέπτυξαν, ξεχωριστά, γεωστατιστικές μεθόδους που επιλύουν τα προβλήματα των μεθόδων παρεμβολής που προαναφέρθηκαν και εξασφαλίζουν την καλύτερη δυνατή πρόβλεψη των τιμών βάσει των υποθέσεων που υιοθετούνται. (Burrough & McDonnell, 1998).

Αυτό έγινε συνδέοντας τη χωρική στατιστική με τη θεωρία των χωρικών μεταβλητών, δημιουργώντας έναν απλό τρόπο μέτρησης της χωρικής εξάρτησης οποιουδήποτε συνεχούς μεταβαλλόμενου φαινομένου

Οι γεωστατιστικές μέθοδοι βασίζονται στη διαπίστωση ότι η χωρική μεταβολή των ιδιοτήτων που παίρνουν τιμές σε συνεχές πεδίο είναι τόσο ακανόνιστη, ώστε δεν μπορεί να αντιπροσωπευθεί ικανοποιητικά από μια απλή μαθηματική συνάρτηση, αλλά από μια τυχαία κατανομή. Σε αυτή την περίπτωση η ιδιότητα ονομάζεται "Regionalized" μεταβλητή (Burrough & McDonnell, 1998).

2.3.3 Θεωρία της περιφερειακής μεταβλητής

Υπάρχουν διάφορα επιστημονικά μοντέλα πρόβλεψης, τα οποία λόγο των ομοιοτήτων τους εντάσσονται σε δύο μεγάλες κατηγορίες:

1 τα προσδιοριστικά ή αιτιοκρατικά (*deterministic*) μοντέλα στα οποία οι γνωστές μεταβλητές (π.χ. αρχικές συνθήκες) αρκούν για την ακριβή πρόβλεψη των αποτελεσμάτων τους,

2 τα στοχαστικά (*stochastic, probabilistic*) μοντέλα στα οποία οι γνωστές μεταβλητές δεν είναι αρκετές για την ακριβή πρόβλεψη των αποτελεσμάτων τους καθώς υπάρχουν επιδράσεις άγνωστες στον παρατηρητή (επηρεάζονται από τον παράγοντα «τύχη»)

Στα στοχαστικά μοντέλα υπάρχει η έννοια της πιθανότητας, μιας τιμής επηρεάζεται από τον παράγοντα τύχη. Κατά τον κλασσικό ορισμό της πιθανότητας του Laplace,αν Ω το πλήθος των δυνατών αποτελεσμάτων και Α το πλήθος των ευνοϊκών αποτελεσμάτων για το ενδεχόμενο (Α) τότε η πιθανότητα του ενδεχομένου είναι :

$$P(A) = \frac{N(A)}{N(\Omega)}.$$

Τυχαία μεταβλητή (randomvariable) ορίζεται ως η μεταβλητή που μπορεί να πάρει μία τιμή από ένα σύνολο μερικών πιθανών τιμών, έχει επομένως συνδέσει κάθε τιμή της με καθορισμένη πιθανότητα σύμφωνα με το προηγούμενο ορισμό. Εάν λαμβάνει πεπερασμένο ή αριθμήσιμο πλήθος τιμών καλείται απαριθμητή ή διακριτή ενώ εάν παίρνει κάθε τιμή ενός διαστήματος ονομάζεται συνεχής. Τόσο για τις διακριτές όσο και για τις συνεχείς μεταβλητές ορίζονται οι συναρτήσεις κατανομών τους που περιγράφουν μέσω μιας σχέσης τις πιθανότητες που αντιστοιχούν σε κάθε τιμή που μπορεί να λάβει η μεταβλητή. Οι κατανομές μπορεί να είναι γνωστές ή συνηθέστερα θα πρέπει να εκτιμηθούν.

Η διακύμανση μιας τυχαίας μεταβλητής, όπως εδαφικής ιδιότητας Z σε οποιαδήποτε θέση i, δίνεται από το μοντέλο :

$$Z_i = m + e_i$$

Όπου m ο μέσος όρος και e_i ένας καθαρά τυχαίος όρος που προέρχεται από πληθυσμό με μέση τιμή 0, διακύμανση σ^2 και κανονική κατανομή (Καλύβας, 1989).

Από τα παραπάνω γίνεται φανερό ότι η θεωρία της τυχαίας μεταβλητής δεν εξετάζει τη μεταβλητότητα μιας ιδιότητας σε συνάρτηση με τη γεωγραφική της θέση. Με το θέμα αυτό ασχολείται η επιστήμη της γεωστατιστικής. Η γεωστατιστική ασχολείται με χωρικά ή χωροχρονικά δεδομένα (Krige, 1951)και οι τυχαίες μεταβλητές αντικαθίστανται από άλλες τις λεγόμενες περιφερειακές (τοπικές- γεωγραφικές) μεταβλητές.

Για να πραγματοποιηθεί οποιοσδήποτε υπολογισμός σε κάποια μεταβλητή, στη γεωστατιστική υποθέτουμε την ύπαρξη μια περιφερειακή μεταβλητή

$$\{Z(x), x \in \Omega\}$$

Όπου Ζ(χ) είναι μια περιφερειακή μεταβλητή που καθορίζει την τιμή του εξεταζόμενου χαρακτηριστικού σε μία χωρική περιοχή Ω όπου η τιμή χ μέσα σε αυτή ποικίλει συνεχώς παίρνοντας συντεταγμένες x, y όταν αναφερόμαστε σε δύο διαστάσεις και x, y, z, όταν αναφερόμαστε σε τρεις διαστάσεις (Cressie, 1991). Πολλές φορές οι ιδιότητες που πρέπει να περιγραφούν μέσω τέτοιων συναρτήσεων είναι δυνατόν να μεταβάλλονται τόσο ποικιλότροπα, ώστε να είναι αδύνατον να περιγραφούν απευθείας με μαθηματικό τρόπο.

Αν θεωρήσουμε όλες τις τιμές Ζ(χ) για όλα τα χ μίας έκτασης τότε η περιφερειακή μεταβλητή ζ(χ) είναι μέλος ενός συνόλου τυχαίων μεταβλητών Ζ(χ) για όλες τις θέσεις χ. αυτό το σύνολο ονομάζεται τυχαία κατανομή Ζ(χ) γιατί συνδέει τη τυχαία μεταβλητή Ζ με κάθε θέση χ (Καλύβας, 1989).

Η έκφραση της περιφερειακής μεταβλητής Ζ(χ) μέσω της τυχαίας κατανομής Ζ(χ) έχει λειτουργική σημασία μόνο όταν αναφέρεται σε όλο ή σε μέρος του νόμου των πιθανοτήτων που αποκλειστικά καθορίζει την τυχαία συνάρτηση (Journeland Huizbregts, 1978).

Προφανώς υπάρχει αδυναμία καθορισμού του νόμου των πιθανοτήτων από μια μόνο έκφραση της τυχαίας κατανομής Ζ(χ) με αποτέλεσμα να απαιτούνται πολλές εκφράσεις $Z_1(\chi)$, $Z_2(\chi)$,..., $Z_3(\chi)$, της τυχαίας κατανομής Ζ(χ). Επειδή όμως έχουμε δυνατότητα μόνο ενός συνόλου τιμών ζ(χ,) στις θέσεις χ, είναι απαραίτητο να γίνουν ορισμένες υποθέσεις που εισάγουν διάφορους βαθμούς ομοιογένειας στο χώρο και συνθέτουν την υπόθεση της στασιμότητας (Stationarity) (Καλύβας, 1989). Αναλυτικά έχουμε:

Η Θεωρία της Περιφερειακής Μεταβλητής υποθέτει ότι η χωρική διακύμανση κάθε μεταβλητής μπορεί να εκφραστεί ως άθροισμα τριών συνιστωσών (Burrough & McDonnell 1998):

• Μίας δομικής συνιστώσας που έχει ένα σταθερό μέσο ή τάση m(χ).

• Μίας τυχαίας, αλλά χωρικά συσχετισμένης συνιστώσας, γνωστή ως η διακύμανση της περιφερειακής μεταβλητής [ε'(χ)].

• Μιας τυχαίας αλλά μη χωρικά συσχετισμένης συνιστώσας, που αντιπροσωπεύει τα σφάλματα μετρήσεων και τη χωρική διακύμανση πολύ μεγάλης κλίμακας και ακολουθεί κανονική κατανομή και αποκαλείται «θόρυβος», ε" .

Έτσι λοιπόν, έστω χ είναι η θέση σε μονοδιάστατο, 2-διάστατο, ή 3-διάστατο χώρο. Τότε η τιμή της τυχαίας μεταβλητής Ζ στο χ δίνεται από την εξίσωση:

$$Z(\chi) = m(\chi) + \varepsilon'(\chi) + \varepsilon''$$

Η αναμενόμενη τιμή της regionalized μεταβλητής είναι ίδια για κάθε θέση της υπό εξέταση περιοχής, οπότε λέμε ότι χαρακτηρίζεται από σταθερότητα πρώτης τάξης και παράλληλα χρησιμοποιείτε η ενδογενής υπόθεση (intrinsic chypothesis) στασιμότητας (Matheron, 1971; Kitanidis, 1997). Με βάση την υπόθεση αυτή οι διαφορές Z(x) - Z(x+h) είναι ασθενώς στάσιμες (secondorderstationary – weakstationary).

Άρα η h(χ) ισούται με τη μέση τιμή στην περιοχή του δείγματος, οπότε ο μέσος όρος ή αναμενόμενη διαφορά μεταξύ δύο θέσεων χ και χ + h που απέχουν μεταξύ τους απόσταση h, θα ισούται με το μηδέν.

$$E\ [Z(\chi)\ - Z(\chi + h)]\ =\ 0$$

Όπου: Z(χ), Z(χ+h) είναι οι τιμές της τυχαίας μεταβλητής Z στις θέσεις χ και χ+h.
Επίσης η σταθερότητα δεύτερης τάξης ισχύει όταν η διακύμανση υπάρχει για κάθε ζεύγος σημείων και εξαρτάται από την απόσταση h που τα χωρίζει και όχι από τη θέση τους στο χώρο, οπότε :

$$\mathrm{E}[\{Z(\chi) - Z(\chi + h)\}2]\ =\ E[\{\varepsilon'(\chi) - \varepsilon'(\chi + h)\}2]\ =\ 2\gamma(h)$$

όπου γ(h) είναι γνωστή ως ημιβαριόγραμμα (variogram).
Οι δύο παραπάνω συνθήκες :
Α. για τη σταθερή διαφορά και
Β. για τη διακύμανση

αποτελούν τις αναγκαίες προϋποθέσεις για την ισχύ της θεωρίας της περιφερειακής μεταβλητής (γενική υπόθεση στασιμότητας) (Burrough & McDonnell,1998).
Όταν ικανοποιούνται οι παραπάνω συνθήκες η ημιδιακύμανση μπορεί να προσδιοριστεί από τα δεδομένα μας με την εξίσωση:

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}\{Z(x_i) - Z(x_i + h)\}^2$$

(Burrough & McDonnell,1998).

2.3.4 Ημιβαριόγραμμα

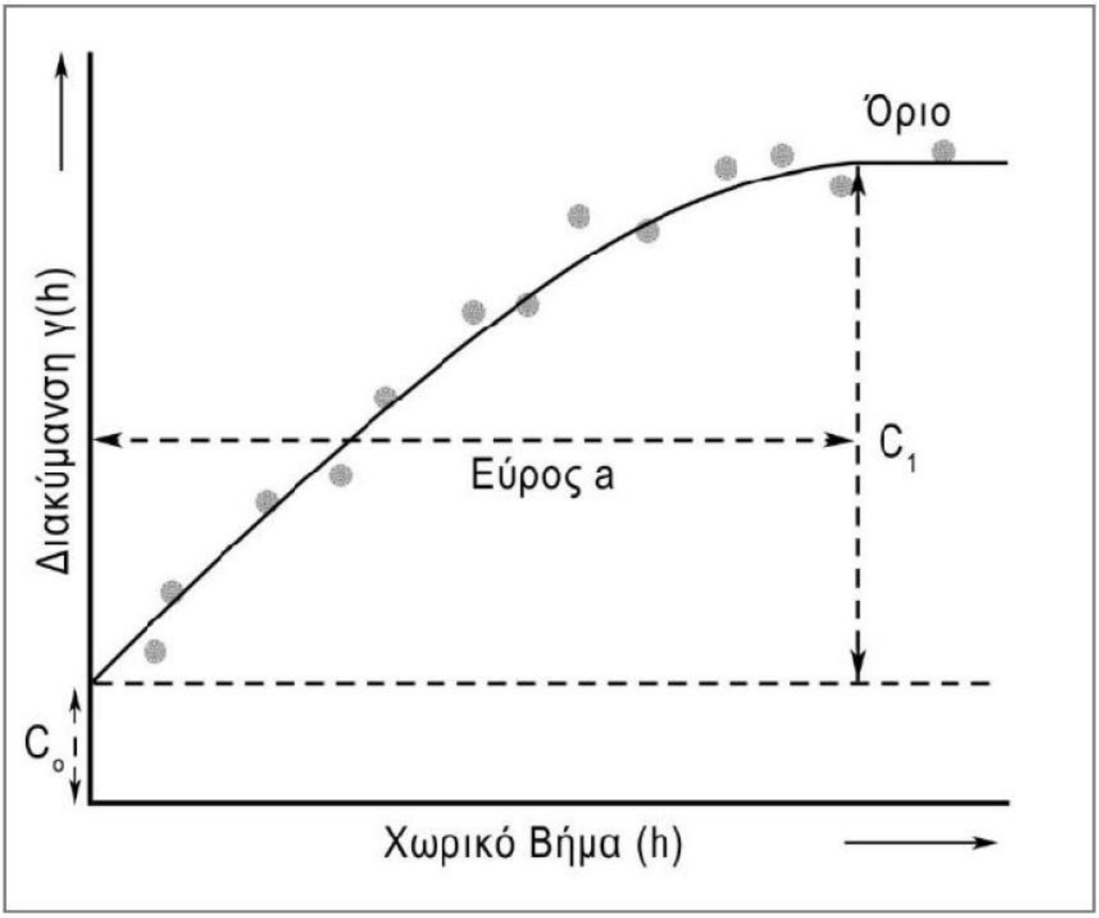
Το ημιβαριόγραμμα ή ημιμεταβλητόγραμμα, αναπαριστά τη χωρική εξάρτηση σε γειτονικά πεδία, δηλαδή την πιθανότητα ότι οι παρατηρήσεις που βρίσκονται κοντά είναι πιο όμοιες από αυτές που βρίσκονται μακρύτερα, ως συνάρτηση της ημιδιακύμανσης σε δεδομένες αποστάσεις (Atkinson, 1993, Cohenet.al., 1990, Legendre και Fortin 1989). Αυτό συμβαίνει με βάση τον τρόπο που ορίζονται δύο σημεία στο χώρο, την μετρική τους απόσταση και την μεταξύ τους κατεύθυνση (OliverandWebster, 1990).

Αποτελεί βασικό διαγνωστικό εργαλείο για τον χωρικό χαρακτηρισμό μιας χωρομεταβλητής αλλά και βασικό εργαλείο για την εφαρμογή της γεωστατιστικής μεθόδου χωρικής παρεμβολής Kriging και ορίζεται από την συνάρτηση

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}\{Z(x_i) - Z(x_i + h)\}^2$$

Προσεγγίζοντας την πρακτική πλευρά της γεωστατιστικής η βαρύτητα επικεντρώνεται στη μελέτη διαφορών μεταξύ τιμών μιας ιδιότητας σε θέσεις που απέχουν κάποια απόσταση η μία από την άλλη.

Στη πράξη δεν ενδιαφερόμαστε για το πρόσημο της διαφοράς αλλά για την απόλυτη τιμή της. Στο χρησιμοποιούμενο τύπο και για την αποφυγή των δυσκολιών που εισάγει ο υπολογισμός απολύτων τιμών χρησιμοποιούμε τετράγωνα διαφορών: $(Z(x)-Z(x+h))^2$. Η τιμή της σχέσεως αυτής εκφράζει την ανομοιότητα μεταξύ δύο σημείων αλλά εκείνο που έχει σημασία για τον προσδιορισμό της συναρτήσεως που περιγράφει τη σχέση μεταβολής μιας ιδιότητας σε μια έκταση είναι ο μέσος όρος των διαφορών (αποστάσεως h) όλων των ζευγαριών. Ο μέσος αυτός όρος του αθροίσματος των διαφορών εκφράζει την ημιδιακύμανση (Καλύβας, 1989).



Εικόνα 2.11: Πειραματικό βαριόγραμμα

2.3.5 Τα χαρακτηριστικά του ημιβαριογράμματος

- Εύρος (range)

  Το εύρος είναι εκείνη η απόσταση, εντός της οποίας η ημιδιακύμανση των τιμών της εξεταζόμενης μεταβλητής αυξάνεται με τρόπο που να μπορεί να αποτυπωθεί και χωρικά. Δείχνει το κατά πόσο επηρεάζει η γνωστή τιμή ενός σημείου την τιμή των υπολοίπων αγνώστων σημείων ανάλογα με την απόσταση που απέχουν μεταξύ τους, έτσι όσο πιο κοντά βρίσκονται δύο σημεία τόσο πιο όμοιες είναι. Το εύρος επίσης προσδιορίζει τη ζώνη επίδρασης ενός δείγματος (Καλύβας, 1989) καθώς σε αποστάσεις μεγαλύτερες του εύρους δεν υπάρχει συσχέτιση εκτός και αν οι τιμές της εξεταζόμενης μεταβλητής παρουσιάζουν περιοδικότητα

(Burrough&McDonnell,1998). Στα σχήματα της προκείμενης εργασίας συμβολίζεται με a.

- Όριο (Sill)

  Πρωτοαναφέρθηκε από τους Journel & Huijbregts το 1978.Είναι το σημείο που σταθεροποιείτε η διακύμανση ή σταθεροποιείτε τοπικά αν οι τιμές της εξεταζόμενης μεταβλητής παρουσιάζουν περιοδικότητα και καλείται όριο. Η τιμή αυτή εκφράζει τη διακύμανση των δειγμάτων. Πρακτικά εισάγει το συλλογισμό ότι σε μεγάλες αποστάσεις η συμπεριφορά της εξεταζόμενης μεταβλητής είναι τυχαία και δεν υπάρχει εξάρτηση μεταξύ της τιμής σε μια θέση με την τιμή σε κάποια άλλη θέση (Καλύβας, 1989). Το ποσοστό αύξησης των εξεταζόμενων τιμών αντανακλά τον βαθμό ομοιογένειας των δειγμάτων (Chilès & Delfiner, 1999). Στα σχήματα της προκείμενης εργασίας συμβολίζεται με C_1.

- Διακύμανση επί της αρχής (Nugget)

  Η διακύμανση (nugget) μας δείχνει τις χωρικές μεταβολές που παρατηρούνται σε αποστάσεις πολύ μικρότερες από αυτές που μπορούν να μελετηθούν ή τις χωρικές μεταβολές που οφείλονται σε σφάλματα κατά τη λήψη των τιμών. Κατά τον Cressie αν η διακύμανση έχει τιμή ίδια ή σχεδόν ίδια με το όριο τότε δεν υπάρχει χωρική συσχέτιση στις εξεταζόμενες τιμές.

  Ο Cambardella το 1994 εισήγαγε τον ακόλουθο τρόπο εκτίμησης της χωρικής εξάρτησης, αυτό υπολογίζεται με τον επί της εκατό λόγο της διακύμανσης με το όριο

$$\frac{nugget}{sill} * 100 = \%$$

  Αν η τιμή εκτίμησης είναι μικρότερη ή ίση του 25%, υπάρχει ισχυρή χωρική εξάρτηση, αν η τιμή βρίσκεται μεταξύ 26-75% υπάρχει σχετική χωρική συσχέτιση, ενώ για τιμές άνω του 75% υπάρχει ασθενής χωρική συσχέτιση. Στα σχήματα της προκείμενης εργασίας συμβολίζεται με C_0.

2.3.6 Τύποι ημιβαριογράμματος

1. **Μεταβατικά (transitive)**: Είναι το "τυπικό" ημιβαριόγραμμα όπου η ημιδιακύμανση σε μικρές αποστάσεις αυξάνεται και φθάνει ένα όριο (sill). Χαρακτηρίζεται την ύπαρξη συγκεκριμένου εύρους (range).
2. **Καθαρού nuggeteffect ή τυχαία:** Αυτά φανερώνουν την έλλειψη συσχέτισης μεταξύ των δειγμάτων για το συγκεκριμένο lag πρόβλημα που αντιμετωπίζεται με μείωση της απόστασης δειγματοληψίας.
3. **Συνεχή**: Εδώ το ημίβαριογραμμα έχει παραβολική μορφή και δείχνει συνεχή αύξηση της διακύμανσης. Εκφράζει την ύπαρξη του φαινομένου drift ή trend, το οποίο περιγράφει συνεχή συστηματική μεταβολή (αύξηση ή μείωση) της τιμής μιάς ιδιότητας κατά μήκος της μελετούμενης διεύθυνσης.
4. **Γραμμικά**: Τα ημίβαριογράμματα αυτού του τύπου απεικονίζουν γραμμική μεταβολή της συσχέτισης μεταξύ των δειγμάτων, τη γραμμική δηλαδή σχέση μεταξύ ημιδιακυμάνσεως και αποστάσεως.

2.3.7 Ανισοτροπία

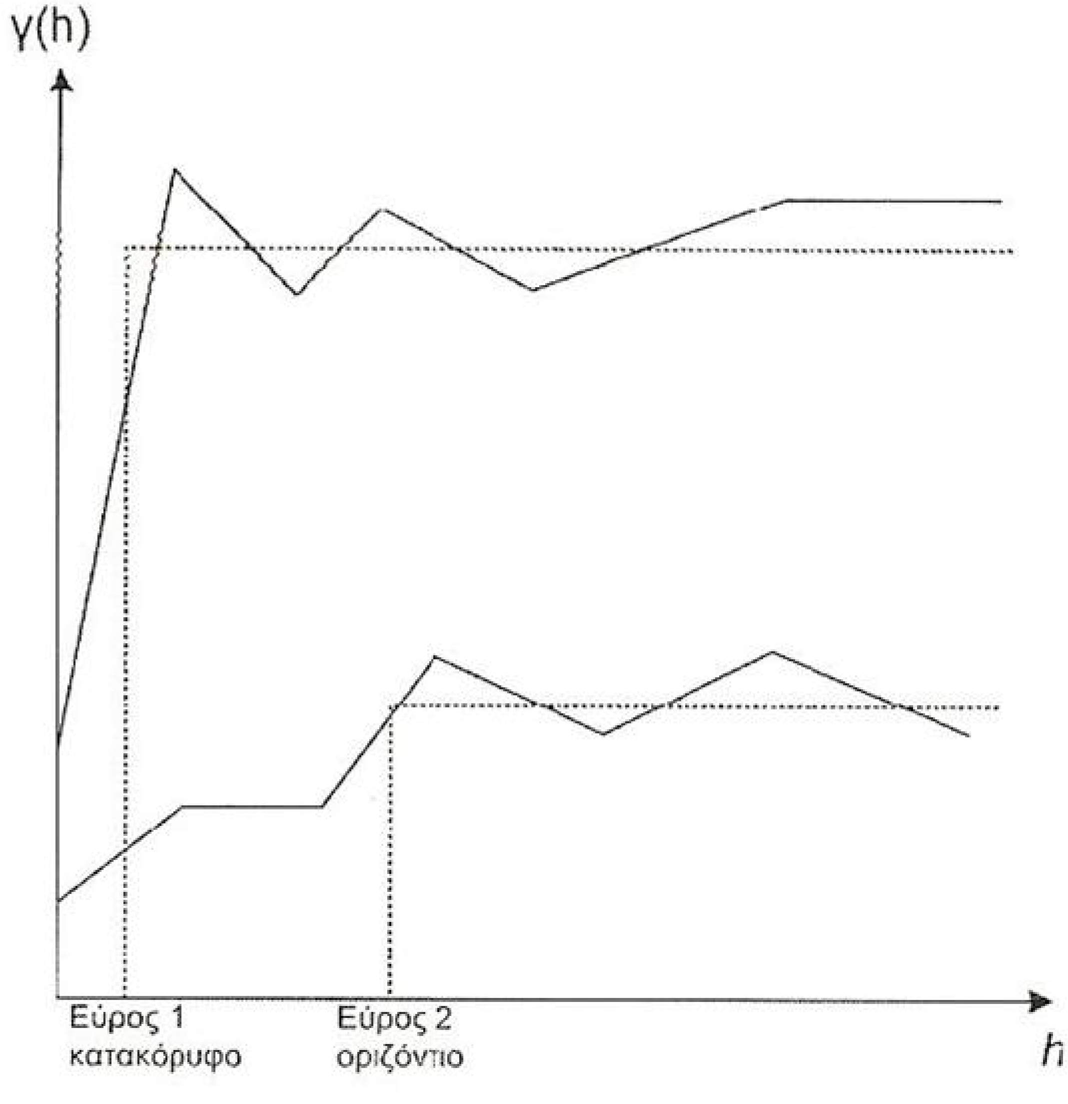
Η διακύμανση των τιμών του πειραματικού βαριογράμματος υπολογίζεται με μοναδικό παράγοντα την απόσταση των θέσεων μέτρησης ανεξάρτητα από τη διεύθυνση που θα έχει το διάνυσμα της απόστασης. Η λογική αυτή είναι βάσιμη μόνο αν οι τιμές της υπό εξέταση ιδιότητας μεταβάλλονται με τον ίδιο τρόπο προς όλες της κατευθύνσεις και αυτή η περίπτωση λέγεται ισοτροπία.

Ένα παράδειγμα ανισοτροπίας είναι όταν θέλουμε να προβλέψουμε τις μεταβολές στη μέση θερμοκρασία μιας περιοχής, ο τρόπος μεταβολής ανά κατεύθυνση ήταν διαφορετικός ανάλογα με τον τοπογραφικό σχηματισμό του εδάφους.

Στη βιβλιογραφία υπάρχουν πολλά είδη ανισοτροπίας αλλά δύο είναι τα πιο κοινά η γεωμετρική, που είναι και η πιο σημαντική (Καλύβας, 1989) και η ζωνική.

- Ζωνική Ανισοτροπία

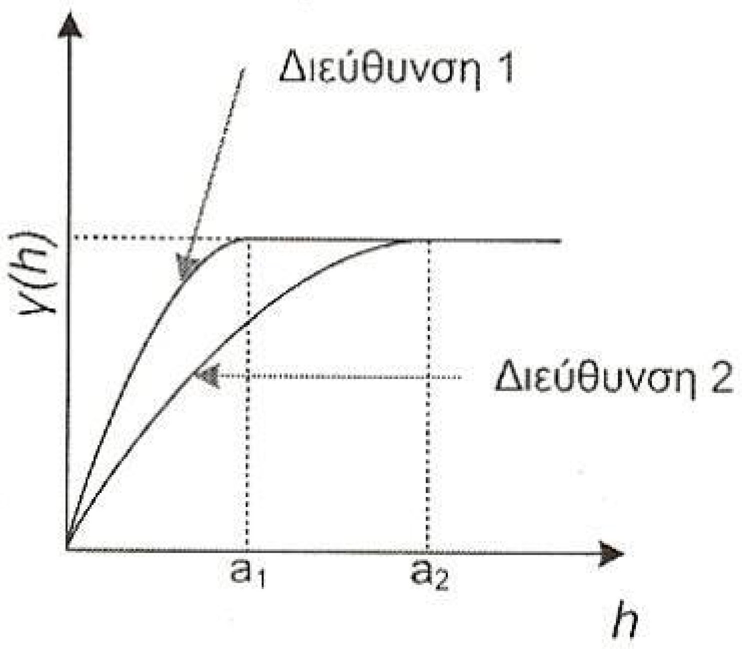
Ή αλλιώς nugget anisotropy (LeMay, 1998) ορίζεται ως η περίπτωση κατά την οποία το εύρος παραμένει σταθερό ανά διεύθυνση αλλά αλλάζει το όριο(Isaaksand Srivastava, 1989). Γενικά δεν μπορεί να μειωθεί με ένα απλό γραμμικό μετασχηματισμό της απόστασης δειγματοληψίας (Journel & Huijbrects, 1978). Η ζωνική ανισοτροπία είναι ένα κοινό χαρακτηριστικό των ιδιοτήτων που εμφανίζουν γεωχημικές ή γεωφυσικές μεταβολές (κλίσεις) προκαλούμενες από προσανατολισμένες αποθέσεις ή ορυκτοποιήσεις. Αυτή η ανισοτροπία μπορεί να συμβεί ως προς τρεις διαστάσεις και αν και παρατηρείται συχνά σε ορυκτά κοιτάσματα, δεν είχε περιγράφει στην εδαφολογία κατά τον Trangram (1985), αλλά σε εργασία του ο Καλύβας(1989) περιέγραψε ζωνική ανισοτροπία σε εδαφικά δεδομένα.



Εικόνα 2.12: Ζωνική και ΓεωμετρικήΑνισοτροπία

- Γεωμετρική Ανισοτροπία

Χαρακτηρίζεται από βαριογράμματα ανά κατεύθυνση τα οποία παρουσιάζουν σχεδόν το ίδιο όριο αλλά διαφορετικό εύρος (Journel & Huijbregts, 1978, Isaaksand Srivastava, 1989). Αυτό το είδος ανισοτροπίας μπορεί να αρθεί με έναν απλό γραμμικό μετασχηματισμό. Στην προκείμενη περίπτωση η μεταβολή για μια δεδομένη απόσταση h σε μια διεύθυνση είναι ισοδύναμη με τη μεταβολή δύο σημείων σε απόσταση k φορές την h, σε μια άλλη διεύθυνση. Ο συντελεστής k καλείται λόγος ανισοτροπίας και εκφράζει το σχετικό μέγεθος των διαφοροποιήσεων των μεταβολών στις εξεταζόμενες διευθύνσεις (Καλύβας, 1989).



Εικόνα 2.13: Γεωμετρική Ανισοτροπία

Διαδικασία άρσης της γεωμετρικής ανισοτροπίας: (Krūminiene,2010)

1. Τον εντοπισμό των διευθύνσεων μέγιστου και ελάχιστου εύρους (range).
2. Την στροφή των αξόνων ώστε να ταυτίζονται με τις διευθύνσεις μέγιστου – ελάχιστου εύρους.
3. Τα κατευθυντικά βαριογράμματα προσαρμόζονται σε ένα με τυποποιημένο εύρος 1.
4. Την κατασκευή ενός πίνακα μετασχηματισμών

Στη γενική έκφραση του μετασχηματισμού καθόσον οι διευθύνσεις μέγιστου και ελάχιστου εύρους είναι δυνατόν να μην είναι ορθογώνιες (Krūminiene,2010) και ορίζοντας *x*ατο ελάχιστο εύρος και *y*ατο μέγιστο τότε ο μετασχηματισμός γίνεται:

$$\begin{pmatrix} h'_x \\ h'_y \end{pmatrix} = \begin{pmatrix} \frac{1}{a_x} & 0 \\ 0 & \frac{1}{a_y} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} h_x \\ h_y \end{pmatrix} = TRh$$

Επίσης ο λόγος a_x/a_y ορίζεται και ως λόγος ανισοτροπίας (Krūminiene,2010) και κυμαίνεται μεταξύ 0 και 1, και το βαριόγραμμα μπορεί να γραφεί $\gamma_{(h)=}\,\gamma_{(Ah)}$(Chiles & Delfiner, 1999).

2.3.8 Μοντέλα Βαριογράμματος

Για να είναι το πειραματικό διάγραμμα διακύμανσης χρήσιμο σε μία μέθοδο παρεμβολής (Kriging), θα πρέπει να προσεγγιστεί από ένα θεωρητικό μοντέλο από όπου θα προκύψει ένα προσαρμοσμένο γράφημα διακύμανσης. Έτσι το δεύτερο στάδιο μετά την κατασκευή του ημιβαριογράμματος είναι η εκλογή ενός μαθηματικού μοντέλου που να απεικονίζει τον τρόπο μεταβολής της ημιδιακυμάνσεως με την απόσταση lag, που να περιγράφει δηλαδή, μέσω μιας συναρτήσεως την καμπύλη, η οποία ενώνει τα ζευγάρια των σημείων h, γ(h). Στο σημείο αυτό, θα πρέπει να τονιστεί ότι η επιλογή των διαφόρων αυτών μοντέλων δεν μπορεί να είναι μια απλή, υποκειμενική διαδικασία, αφού η χρήση τους πρέπει να ακολουθεί συγκεκριμένες μαθηματικές προϋποθέσεις (Isaaks & Srivastava, 1989). Γενικά, η προσαρμογή των στοιχείων ενός δείγματος σε ένα συγκεκριμένο μοντέλο είναι μία διαδικασία, όπου η γνώση, η εμπειρία και τα στατιστικά στοιχεία αρμονικά την καθορίζουν (Κουτσόπουλος, 2005), αλλά καθορίζεται επίσης από τους σκοπούς για τους οποίους εκτελείται αυτή η εκλογή (Webster, 1985).

Κάθε μοντέλο περιέχει τις εξής παραμέτρους:

- Lag: h
- Range: a
- Nugget variance: C_0
- Sill: $C_0 + C$

Τα θεωρητικά μοντέλα που εφαρμόζονται συνήθως στις εδαφικές μεταβλητές είναι το εκθετικό (exponential), το σφαιρικό (spherical), και το κανονικό μοντέλο ή γκαουσιανό (Gaussian)

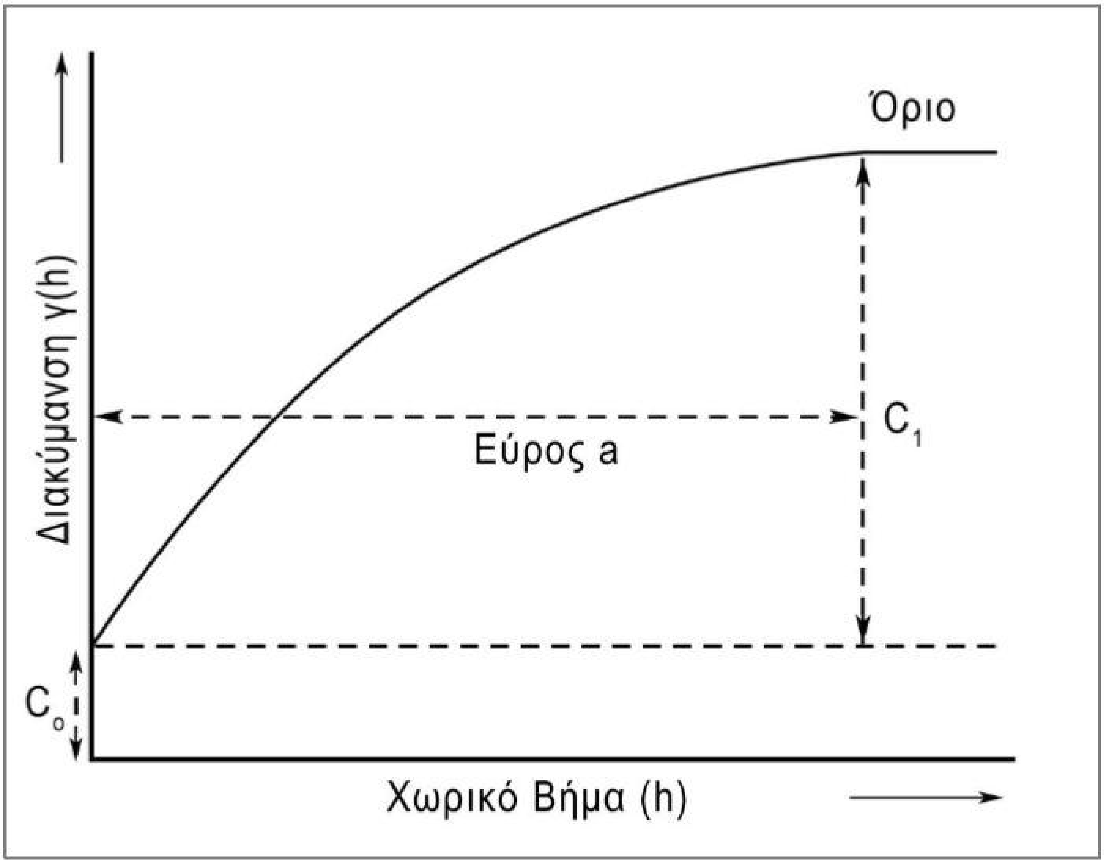
2.3.8.1 Εκθετικό μοντέλο

Το εκθετικό μοντέλο μπορεί να παριστά επαρκώς την έννοια της χωρικής τυχαιότητας και μπορεί να περιγράψει τα βαριογράμματα για αρκετές εδαφολογικές ιδιότητες (Olover, 1984) Χρησιμοποιείται όταν παρατηρείται βαθμιαία μετάβαση προς τη μέγιστη τιμή της ζώνης επιρροής (Κουτσόπουλος, 2005)

Περιγράφεται από την κάτωθι σχέση:

$$\gamma(h) = C_0 + C_1\left[1 - e^{\frac{-|h|}{r}}\right] \quad \text{για} \quad |h| > 0$$

$$\gamma(h) = 0 \text{για} |h| = 0$$



Εικόνα 2.14: Εκθετικό Μοντέλο

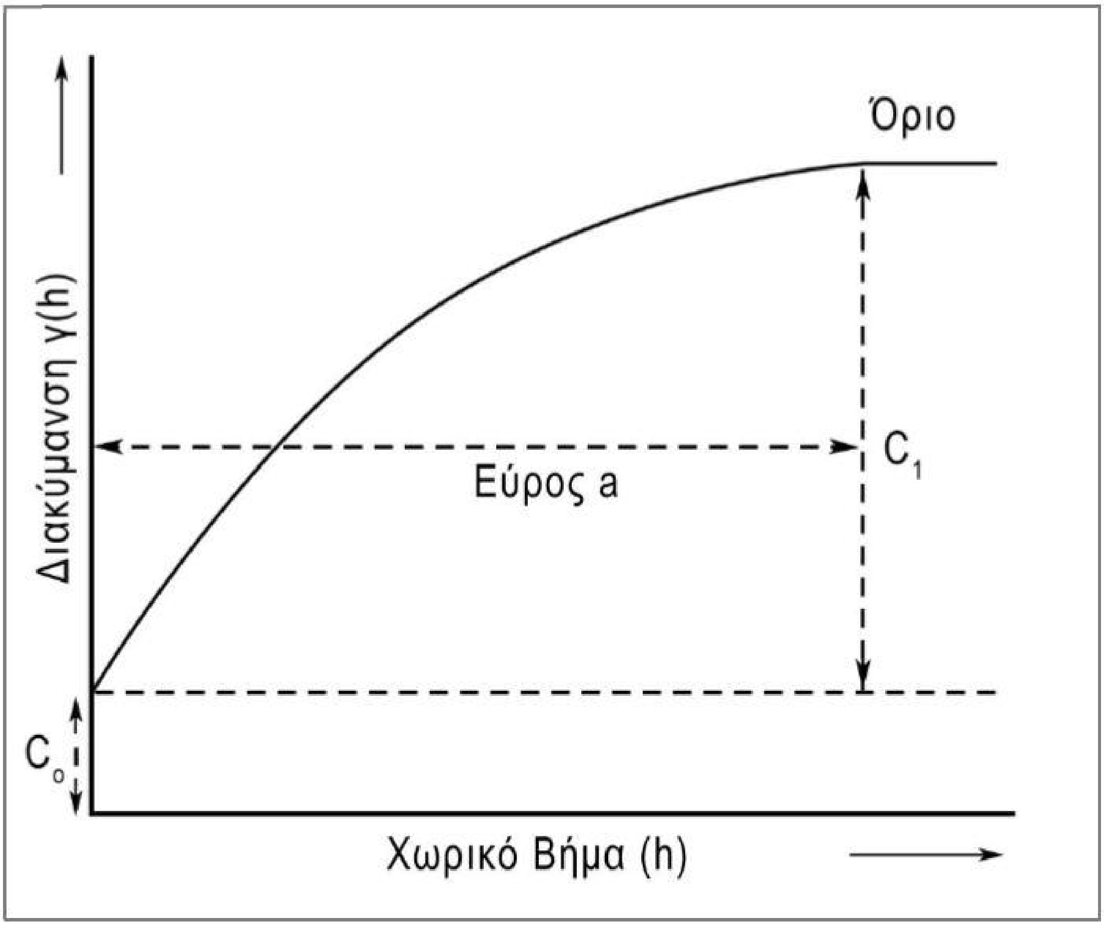
Toreίναι η παράμετρος της απόστασης που ελέγχει την ακτίνα ή το εύρος της χωρικής διακύμανσης. Ο Armstrong (1998) όρισε το δραστικό εύρος (effectiverange) ως a=3r και είναι η απόσταση όπου το όριο προσεγγίζει την τιμή C_0+0.95C

2.3.8.2 Σφαιρικό μοντέλο

Η χρήση του σφαιρικού μοντέλου θεωρείτε ιδανική για βαριόγραμμα διότι ορίζεται πολύ καθαρά το όριο και η ερμηνεία του εύρους είναι εύκολη. (Woodcocket.al., 1988). Κατά τον Armstrong (1998) είναι το πιο ευρέως χρησιμοποιούμενο μοντέλο στη γεωστατιστική.

$\gamma(h) = C_0 + C_1\left[1{,}5\,\frac{|h|}{a} - 0{,}5\left(\frac{|h|}{a}\right)^3\right]$ για $0 < |h| < a$

$\gamma(h) = 0$ για $h = 0$



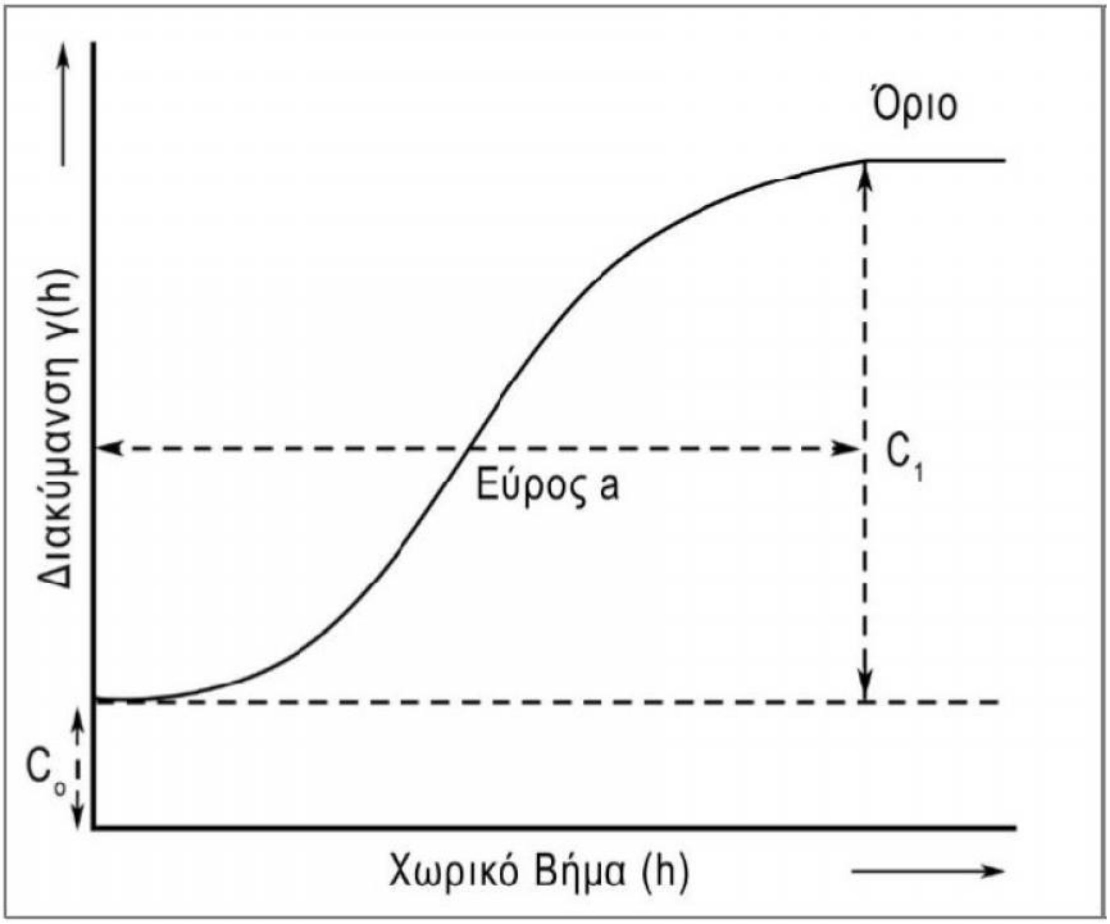
Εικόνα 2.15: Σφαιρικό Μοντέλο

2.3.8.3 Κανονικό μοντέλο (Gaussian)

Το Γκαουσσιανό μοντέλο χρησιμοποιείται ώστε να αντιπροσωπεύει έντονα συνεχή φαινόμενα, αφού το μοντέλο εμφανίζει μια παραβολική συμπεριφορά κοντά στην αρχή (Morgan, 2005) ή όπως αναφέρει ο Bohling (2005) ιδιότητες με μεγάλη μεταβλητότητα και μικρό εύρος.

$$\gamma(h) = C_0 + C_1\left[1 - e^{\left(\frac{-|h|}{r}\right)^2}\right]$$

$$\gamma(h) = 0$$



Εικόνα 2.16: Κανονικό (Gaussian) Μοντέλο

2.4 Μοντέλα Επιλογής

2.4.1 Γενικά

Η επιλογή μοντέλου, είναι η σωστή κατά περίπτωση, επιλογή ενός στατιστικού μοντέλου από ένα σετ πιθανών υποψήφιων μοντέλων, με σκοπό να αντιπροσωπεύσει τα πειραματικά δεδομένα. Σε αρκετές περιπτώσεις, ένα σετ δεδομένων που προϋπάρχει μπορεί να παίξει τον ρόλο του μοντέλου. Παρόλα αυτά η κοινή διαδικασία περιλαμβάνει την επεξεργασία πειραματικών δεδομένων με τρόπο ώστε να μπορεί να γίνει προσαρμογή σε κάποιο στατιστικό μοντέλο. Δεδομένου ότι υπάρχουν υποψήφια μοντέλα με παρόμοια αποτελέσματα ως προς την προσαρμογή, προτεραιότητα στην επιλογή έχει το απλούστερο μοντέλο.

Στις πιο βασικές μορφές, η επιλογή μοντέλου αποτελεί ένα από τις θεμελιώδεις ικανότητες της επιστημονικής έρευνας. Καθορίζοντας την αρχή που επεξηγεί τη σειρά παρατηρήσεων συχνά συνδέεται άμεσα με ένα μαθηματικό μοντέλο που προβλέπει τις τιμές αυτές. Ένα γνωστό ιστορικό παράδειγμα είναι αυτό του Γαλιλαίου όταν παρουσίασε το πειράματα του κεκλιμένου επιπέδου και κατέδειξε πως η κίνηση που έκαναν οι μπάλες ταίριαζε στην παραβολή που είχε προβλεφθεί από το μοντέλο του.

Πολλές φορές η επιλογή του μοντέλου από το σετ των υποψήφιων μοντέλων, γίνεται εμπειρικά “με το μάτι” κάτι που είναι αμφιλεγόμενο, όμως η μαθηματική ανάλυση επιτρέπει την επιλογή του πραγματικά καλύτερου από αυτά τα μοντέλα. Βασικοί γνώμονες στην επιλογή του κατάλληλου μοντέλου είναι το πόσο αντιπροσωπευτικό είναι, πόσο απλό είναι το μοντέλο, καθώς και το αν οι παράγοντες του αντιπροσωπεύουν κάτι σημαντικό για την εξαγωγή συμπερασμάτων. Τα πιο πολύπλοκα μοντέλα προσαρμόζουν εύκολα το σχήμα τους ώστε να ταιριάζει με τα δεδομένα αλλά οι επιπρόσθετες παράμετροι μπορεί να μην αντιπροσωπεύουν κάτι χρήσιμο. Η πολυπλοκότητα γενικά μετράται με το να υπολογίζεται ο αριθμός των παραμέτρων στο μοντέλο.

Οι τεχνικές επιλογής μοντέλου μπορεί να θεωρηθούν σαν εκτιμήσεις κάποιων φυσικών ποσοτήτων, όπως η πιθανότητα του μοντέλου να παράγει τα δεδομένα. Η προκατάληψη και μεταβλητότητα είναι εξίσου σημαντικές μετρήσεις της ποιότητας του εκτιμητή.

2.4.2 Μέθοδος Ελαχίστων Τετραγώνων

Η απλούστερη και πιο συνηθισμένη μέθοδος προσαρμογής ενός μοντέλου στα πειραματικά δεδομένα είναι η μέθοδος εκτίμησης ελαχίστων τετραγώνων (Cressie , 1993). Η κανονική μέθοδος ελαχίστων τετραγώνων έχει σαν γενικό τύπο

$$RSS = \sum_{i=1}^{n} (\hat{\chi}_i - \chi_i)^2$$

Όπου $\hat{\chi}_i$ οι τιμές από το μοντέλο

και χ_i οι τιμές του πειραματικού ημιβαριογράμματος

Βασικό μειονέκτημα της μεθόδου, είναι ότι το χρησιμοποιούμε με την παραδοχή ότι οι διαφορές της εξίσωσης κατανέμονται κανονικά (normal distribution) (Cressie, 1993).

2.4.3 Bayesian Information criterion (BIC)

Στη στατιστική το Bayesianinformationcriterion (BIC) ή αλλιώς Schwarzcriterion (SBC ή SBIC) είναι ένα κριτήριο για την επιλογή του μοντέλου από μια ομάδα γνωστών μοντέλων, με την καλύτερη προσαρμογή στα πειραματικά δεδομένα. Βασίζεται εν μέρει στη συνάρτηση πιθανόφάνειας (likelihood function), ενώ έχει πολλές ομοιότητες αλλά και σημαντικές διαφορές με το Akaike information criterion (AIC). Όταν ψάχνουμε το καταλληλότερο μοντέλο η προσαρμογή βελτιώνεται με την προσθήκη στο μοντέλο νέων παραμέτρων που το κάνει σύνθετο, το BIC όπως και το AIC λύουν το παραπάνω πρόβλημα εισάγοντας την έννοιας στα κριτήρια του ορίου των ελευθέρων παραμέτρων.

Η γενική μορφή είναι η

$$BIC = -2 \ln \hat{L} + k \left(\ln(n) - \ln(2\pi)\right)$$

Όπου $\hat{L}$ η τιμή της συνάρτησης πιθανοφάνειας

k ο αριθμός των ελεύθερων παραμέτρων

n ο αριθμός των τιμών των πειραματικών δεδομένων.

Για μεγάλο αριθμό δειγμάτων μπορεί να υπολογιστεί και ως κάτωθει

$$BIC = n \ln(\hat{L}) + k \ln(n)$$

Με την προϋπόθεση ότι τα σφάλματα του είναι ισοκατανεμημένα σύμφωνα με την κανονική κατανομή και η διαφορά του λογαρίθμου της πιθανοφάνειας (likelihood) σε σχέση με τις πραγματικές τιμές τείνει στο μηδέν τότε η σχέση γίνεται (M.B. Priestley, 1981)

$$BIC = n \ln\left(\hat{\sigma}_e^2\right) + k \ln(n)$$

Όπου $\hat{\sigma}_e^2$ η διακύμανση του σφάλματος

$$\hat{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^{n}(\chi_i - \hat{\chi}_i)^2$$

Δεδομένων των εκτιμήσεων, το μοντέλο με τη μικρότερη τιμή BIC είναι αυτό που θα επιλεγεί τελικά. Η τιμή του BIC αυξάνεται με την αύξηση της τιμής του σ^2_e και με την αύξηση του αριθμού των ελεύθερων παραμέτρων (k) του κάθε μοντέλου. Ως εκ τούτου η μικρότερη τιμή του BIC συνδέεται είτε με τις λιγότερες ελεύθερες παράμετροι, είτε με την καλύτερη προσαρμογή του μοντέλου στα δεδομένα ή ακόμη μπορούν να συμβαίνουν και τα δύο.

2.4.4 Akaike Information Criterion

Το Akaike Information Criterion (AIC) είναι ένα κριτήριο για την επιλογή του μοντέλου από μια ομάδα γνωστών μοντέλων, με την καλύτερη προσαρμογή στα πειραματικά δεδομένα σε συνδυασμό με την απλότητα του μοντέλου. Αποτελεί μια από τις πιο αξιόπιστες δημοφιλείς και εύκολα εφαρμόσιμες στρατηγικές επιλογής μοντέλου. Το κριτήριο αυτό το οποίο βασίζεται στην εντροπία πληροφορίας, προσφέρει μια σχετική εκτίμηση των πληροφοριών που χάνονται όταν ένα συγκεκριμένο μοντέλο χρησιμοποιείται για να αναπαραστήσει τη διαδικασία που δημιουργεί τα δεδομένα.

Το μειονέκτημα του κριτηρίου αυτού είναι ότι δεν μπορεί να πει τίποτα για την ποιότητα του μοντέλου με την απόλυτη έννοια. Δηλαδή εάν κανένα από τα υποψηφία μοντέλα δεν ταιριάζει επαρκώς τότε το AIC δεν θα δώσει καμιά προειδοποίηση γι' αυτό.

Σαν σύντομο ιστορικό να πούμε ότι το Akaike information criterion (AIC) αναπτύχθηκε από τον Hirotugu Akaike, με αρχικό όνομα "ένα κριτήριο πληροφόρησης". Πρωτο δημοσιεύτηκε από τον Akaike το 1974. Ο Akaike αρχικά ονόμασε την προσέγγιση σαν μεγιστοποίηση της αρχής της εντροπίας, μιας και στηρίχτηκε στην ιδέα της εντροπίας στην θεωρία πληροφορίας.

Η αρχική προέλευση του AIC βασίστηκε σε κάποιες ισχυρές υποθέσεις του Takeuchi (1976) που έδειξε πως οι υποθέσεις "τρωτά" σημεία που μπορούν να καταδείξουν αδυναμίες. Η δουλειά του Takeuchi παρόλα αυτά ήταν στα ιαπωνικά και δεν ήταν ευρέως γνωστή για πολλά χρόνια.

Το AICc είχε αρχικά προταθεί για γραμμική παλινδρόμηση από τον Sugiura (1978). Αυτό υποκίνησε τη δουλειά των Hurvich & Tsai (1989), και πολλές εργασίες των ίδιων όπου διεύρυναν τις εφαρμογές στις οποίες μπορεί να χρησιμοποιηθεί το AICc. Η δουλειά των Hurvich & Tsai αποτέλεσε μία συνεισφορά στο να δημοσιευθεί μία δεύτερη έκδοση από τους Brockwell & Davis (1991), όπου αποτελεί τη σταθερή αναφορά για γραμμικές χρονοσειρές.

Το 2002 οι Burnham & Anderson το επέκτειναν και πήγαν την προσέγγιση στους νόμου της θερμοδυναμικής του Ludwig Boltzmann. Συνοπτικά αναφερόμενοι το να ελαχιστοποιηθεί το AIC σε ένα στατιστικό μοντέλο είναι εξίσου σημαντικό με το να μεγιστοποιηθεί η εντροπία σε ένα θερμοδυναμικό σύστημα. Με άλλα λόγια η θεωρητική προσέγγιση στην πληροφορία πάνω στη στατιστική εφαρμόζεται στον $2^{ο}$ νόμο της θερμοδυναμικής.

Η γενική μορφή του κριτηρίου είναι η

$$AIC = 2k - \ln(\hat{L})$$

Όπου k είναι ο αριθμός των παραμέτρων του μοντέλου

Και όπου L είναι η τιμή της πιθανοφάνειας

Δεδομένων των εκτιμήσεων, το μοντέλο με τη μικρότερη τιμή BIC είναι αυτό που θα επιλεγεί τελικά. Ως εκ τούτου το AIC δείχνει το πόσο καλά προσαρμόστηκε ένα μοντέλο στα πειραματικά δεδομένα σε σχέση με τα άλλα υποψήφια μοντέλα αλλά η τιμή αυξάνεται όσο αυξάνονται οι εκτιμώμενοι παράμετροι του μοντέλου, διότι αυξάνουν την πολυπλοκότητα του εκάστοτε μοντέλου. Αυτή είναι μια λειτουργιά του κριτηρίου με στόχο να αποτρέψει την αύξηση του αριθμού των παραμέτρων στο μοντέλο (overfitting) ή τη μόνιμη επιλογή μοντέλου με πολλές παραμέτρους, καθώς ο αυξανόμενος αριθμός παραμέτρων είναι συνώνυμος της καλύτερης προσαρμογής στα πειραματικά δεδομένα.

Μια ακόμη επεξεργασία σε περίπτωση που δεν μπορούμε να βρούμε με σιγουριά το υποψήφιο μοντέλο ώστε να αντιπροσωπεύσει τα δεδομένα λόγο μικρής διαφοράς των τιμών, μια προτεινόμενη διαδικασία είναι η παρακάτω, αφού βρούμε τις τιμές των κριτηρίων AIC για τα υποψήφια μοντέλα.

Έστω AIC1, AIC2, AIC3,…AICn. Ορίζουμε AICmin είναι η ελάχιστη τιμή των AIC και AICi είναι οι διάφορες τιμές των AIC. Στη συνέχεια για κάθε τιμή υπολογίζουμε το $e^{((AICmin-AICi)/2)}$ που δίνει την πιθανότητα να το εξεταζόμενο μοντέλο να έχει την ίδια απώλεια πληροφορίας με το μοντέλο που είχε την μικρότερη τιμή AIC.

Σαν παράδειγμα ας υποθέσουμε πως υπάρχουν 3 υποψήφια μοντέλα, με τιμές AICτιμές 100,102 και 110. Στη συνέχεια το δεύτερο μοντέλο $e^{((100-102)/2)} = 0{,}368$ φορές εξίσου πιθανό να ελαχιστοποιηθεί η απώλεια πληροφορίας όπως και στο $1^{ο}$

μοντέλο, με παρόμοιο τρόπο το 3ο μοντέλο $e^{((100-110)/2)} = 0{,}007$ φορές εξίσου πιθανό να ελαχιστοποιηθεί η απώλεια πληροφορίας όπως και στο 1ο μοντέλο.

Διορθωμένο AkaikeInformationCriterion (AICc)

Το γεγονός ότι η πιθανοφάνεια ενός μοντέλου αυξάνει με την αύξηση του αριθμού των δειγμάτων, αλλά και ότι το AIC προσμετρά όλες τις παραμέτρους του μοντέλου ελεύθερες και μη, έτσι το AIC έχει την τάση να δίνει μικρότερες τιμές στα πιο πολύπλοκα μοντέλα. Έτσι υπήρξε η ανάγκη διόρθωσης του κριτηρίου και η γένεση του AICc.

Το AICc αρχικά είχε προταθεί από τους Hurvich&Tsai (1989). Οι Brockwell&Davis το 1991 και αργότερα το 1997 οι Burnham&Anderson, andCavanaugh έδωσαν την δική τους εκδοχή στο διορθωμένο κριτήριο Akaike. Όλα υποθέτουν ένα μονομεταβλητό γραμμικό μοντέλο με κανονικά κατανεμημένα σφάλματα. Αν δεν μπορεί να τηρηθεί αυτό τότε αλλάζουν τη μορφή του AICc.

Αν όλα τα υποψήφια μοντέλα έχουν τον ίδιο αριθμό παραμέτρων k, τότε το AIC και το AICc ακριβώς τις ίδιες εκτιμήσεις και αρκεί να χρησιμοποιηθεί το AIC γενικά.

Ο γενικός τύπος του AICc είναι:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

Όπου k ο αριθμός των παραμέτρων του μοντέλου

και n το μέγεθος των δεδομένων

Η σύνδεση του AIC με το X^2

Συχνά η επιλογή μοντέλων γίνεται με μεγάλο ανταγωνισμό ανάμεσα στα μοντέλα όπου η πιθανοφάνεια και οι διαφορές υποθέτουμε πως κατανέμονται ομαλά και είναι ανεξάρτητα. Η παραπάνω υπόθεση οδηγεί στο προσαρμοσμένο με το X^2.

Με το X^2 η συνάρτηση πιθανοφάνειας γίνεται:

$$L = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2} \exp\left(-\sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{2\sigma_i^2}\right)$$
$$\therefore \ln(L) = \ln\left(\prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2}\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{\sigma_i^2}$$
$$\therefore \ln(L) = C - \chi^2/2$$

όπου το C είναι μία σταθερά ανεξάρτητη από το μοντέλο και εξαρτώμενη μόνο από τη χρήση συγκεκριμένων σημειακών δεδομένων. Δεν αλλάξει αν δεν αλλάξουν και τα δεδομένα.

Ακολούθως το AIC γίνεται:

$$AIC = 2k - 2\ln(L) = 2k - 2\left(C - \frac{\chi^2}{2}\right) = 2k - 2C - \chi^2$$

Αφού στο AIC έχουν σημασία μόνο οι διαφορές, η σταθερά C μπορεί να παραβλεφθεί:

$$AIC = \chi^2 + 2k$$

Το οποίο μπορεί να γραφεί και

$$AIC = n \ \ln(RSS/n) + 2k + C$$

και όπου

$$RSS = \sum_{i=1}^{n} \left(y_i - f(\chi_i)\right)^2$$

είναι το άθροισμα του τετραγώνου της διαφοράς.

Εδώ και πάλι το C μπορεί να αγνοηθεί στη σύγκριση των μοντέλων οπότε ο τύπος παίρνει την τελική του μορφή

$$AIC = n \ \ln(RSS/n) + 2k$$

2.4.5 Hannan–Quinn Information criterion (HQC)

Στη στατιστική το Hannan–Quinninformationcriterion (HQC) είναι ένα κριτήριο για την επιλογή μοντέλου. Αποτελεί μία εναλλακτική για το Akaikei nformation criterion (AIC) ή το Bayesian information criterion (BIC).

Δίνεται από τη σχέση:

$$HQC = n\,log\left(\frac{RSS}{n}\right) + 2k\,log\big(log(n)\big)$$

όπου κ είναι ο αριθμός των παραμέτρων του μοντέλου

n ο αριθμός των συμμετεχόντων παρατηρήσεων

και το RSS είναι το άθροισμα του τετραγώνου της διαφοράς το οποίο είναι αποτέλεσμα από τη γραμμική παλινδρόμηση και άλλα στατιστικά μοντέλα.

Οι Burnham&Anderson (2002) αναφέρουν πως το HQC ενώ συχνά αναφέρεται, φαίνεται να έχει μικρή συμβολή στην πράξη. Επίσης αναφέρουν πως το HQC όπως και το BIC δεν αποτελεί εκτίμηση της Kullback–Leibler απόκλισης. Οι Claeskens & Hjort (2008) αναφέρουν πως το HQC όπως BIC δεν είναι ασυμπτωματικά αποτελεσματικό και επιπρόσθετα αναφέρουν πως όποια μέθοδος χρησιμοποιείται για εξομάλυνση των βραχυχρόνιων διακυμάνσεων το κριτήριο θα είναι σημαντικότερο στην πράξη από τον όρο log(log(n)) μιας και ο τελευταίος αριθμός είναι μικρός για μεγάλο αριθμό παρατηρήσεων.

3. Η ΕΦΑΡΜΟΓΗ

Στο τελευταίο κεφάλαιο θα γίνει ανάλυση του εργαλείου επιλογής θεωρητικού μοντέλου βαριογράμματος αλλά και των βοηθητικών εργαλείων που αναπτύχθηκαν καθώς και η εφαρμογή του σε πραγματικά δεδομένα. Πιο αναλυτικά στο πρώτο μέρος (3.1. Κώδικας) θα γίνει λεπτομερής ανάλυση των εργαλείων καθώς και η μέθοδος εργασίας που υιοθετήθηκε για τη δημιουργία τους. Στο δεύτερο μέρος θα γίνει παρουσίαση των εφαρμογών των παραπάνω εργαλείων σε δεδομένα καθώς και του πως πρέπει να χρησιμοποιηθούν.

3.1. Ο Κώδικας

Κατά τη διάρκεια της πτυχιακής αναπτύχθηκαν τρία εργαλεία που μας καταδεικνύουν, η χρήση ποιου μοντέλου προσαρμόζεται καλύτερα στο πειραματικό ημιβαριόγραμμα, που προέρχεται από τα προς επεξεργασία δεδομένα ή με άλλα λόγια το μοντέλο που αν χρησιμοποιηθεί θα αποφέρει τη μικρότερη απώλεια πληροφορίας από το πειραματικό ημιβαριόγραμμα.

- Single Semivariogram εργαλείο: είναι απλό και γρήγορο, δημιουργεί και οπτικοποιεί το πειραματικό ημιβαριόγραμμα με παραμέτρους (εύρος-range, χωρικό βήμα - lagsize) τις οποίες επιλέγει ο χρήστης. Βασικός στόχος η διερεύνηση της ύπαρξης της χωρικής εξάρτησης και ο προσδιορισμός του εύρους (range) αυτής.
- Semivariogram Comparison εργαλείο: μοιάζει με το πρώτο με μοναδική διαφορά ότι σαν αποτέλεσμα μας δίνει τρεις παραλλαγές του πειραματικού ημιβαριογράμματος, μεταβάλλοντας το χωρικό βήμα, ώστε να προσανατολιστούμε καλύτερα, ως προς το ποιες τιμές πρέπει να δοθούν στο εργαλείοSingle Semivariogram ώστε να βελτιωθεί το τελικό πειραματικό ημιβαριόγραμμα.
- Model Selection Criteria εργαλείο: μας παρέχει εικόνα για το πώς τα μοντέλα προσαρμόζονται στο πειραματικό ημιβαριόγραμμα αλλά και τις τιμές των κριτηρίων επιλογής Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), και του Hannan-Quinn Information Criterion (HQC).

3.1.1. Γενικά

Και στα τρία εργαλεία χρησιμοποιήθηκαν οι ίδιες βιβλιοθήκες της Python και υιοθετήθηκε η ίδια μέθοδος για την εισαγωγή των δεδομένων αλλά και για την αρχική βασική επεξεργασία τους.

Οι βιβλιοθήκες είναι οι εξής:

- Βιβλιοθήκη arcPy:είναι μια βιβλιοθήκη που αναπτύχτηκε από την εταιρία ESRIμε σκοπό ο εκάστοτε χρήστης του ArcGISνα μπορεί να δημιουργήσει τα δικά του εργαλεία χρησιμοποιώντας τις χωρικές, τις στατιστικές αλλά λειτουργίες βάσεων δεδομένων που υπάρχουν στο ArcGIS. Ακόμη με το arcpyμπορούν να ενσωματωθούν αυτά τα εργαλεία στα εργαλεία που υπάρχουν στο ArcGIS.
- Βιβλιοθήκη scipy: η scipy είναι μια υπερβιβλιοθήκη που περιέχει άλλες σημαντικές αλλά μικρότερες σε μέγεθος βιβλιοθήκες. Εμείς στα εργαλεία που αναπτύξαμε χρησιμοποιήσαμε συγκεκριμένα τα εργαλεία από τον προορισμό της βιβλιοθήκης spatial.distance, όπου παρέχει εργαλεία εξειδικευμένα στη χρήση τους πάνω σε χωρικά δεδομένα, τα pdist και squareform. Το πρώτο υπολογίζει την απόσταση δύο σημείων από τις συντεταγμένες x,y που υπάρχουν σε έναν πίνακα και το δεύτερο δημιουργεί ισορροπημένους πίνακες με συγκεκριμένες θέσεις που έχουν σχέση με τα δεδομένα από τα οποία προέκυψαν.
- Βιβλιοθήκη Numpy: Είναι από τις πιο σημαντικές βιβλιοθήκες στην Python και επιτρέπει στον κώδικα να κάνει γρήγορους υπολογισμούς, υπολογισμούς πινάκων και παρέχει την πιο γρήγορη δομή δεδομένων, τον πίνακα Numpy
- Βιβλιοθήκη matplotlib: είναι η βιβλιοθήκη η οποία εισάγει τα περισσότερα από τα εργαλεία του matlab, στα εργαλεία χρησιμοποιούνται τα απαραίτητα εργαλεία για την εμφάνιση των ημίβαριογραμματων.
- Βιβλιοθήκη Tkinter: είναι μια βιβλιοθήκη για graphicaluserinterface (GUI), που δημιουργεί το περιβάλλον που αλληλεπίδρασης του χρήστη με την εφαρμογή.

```
import arcpy
import matplotlib.pyplot as plt
import numpy as np
from scipy.spatial.distance import pdist, squareform
from Tkinter import *
from arcpy.sa import *
```

Εικόνα 3.1: Η εισαγωγή των βιβλιοθηκών

Για την εισαγωγή των δεδομένων από το ArcGIS στο εργαλείο χρησιμοποιήθηκε η βιβλιοθήκη arcpy .

```
8   input = arcpy.GetParameter(0)
9
10  x = arcpy.da.TableToNumPyArray(input,"SHAPE@X")
11  x = np.array( x, dtype=np.float )
12
13  y =arcpy.da.TableToNumPyArray(input,"SHAPE@Y")
14  y = np.array( y, dtype=np.float )
15
16  zsthlh = arcpy.GetParameter(1)
17  zsthlh = str(zsthlh)
18
19  z = arcpy.da.TableToNumPyArray(input,zsthlh)
20  z = np.array( z, dtype=np.float )
21
22  bw = arcpy.GetParameter(2)
23  bw = int(bw)
24
25  lag = arcpy.GetParameter(3)
26  lag = int(lag)
```

Εικόνα 3.2: Η εισαγωγή των δεδομένων από το ArcGIS10.1

Στην γραμμή 8 του κώδικα, δηλώνουμε ότι η πρώτη παράμετρος που θα εισάγουμε από τη φόρμα στο ArcGIS,θα αποθηκεύσει στην μεταβλητή inputτην ονομασία του shapefile, με τα δεδομένα του οποίου θα εργαστούμε.

Στις γραμμές 10 και 13 του κώδικα, δίνεται εντολή να παρθούν τα x,y αντίστοιχα, για τα σημεία του shapefile που δηλώθηκε πριν, και να τα εισάγει σε έναν Numpy πίνακα με την ονομασία x, y.

Στις γραμμές 11 και 14 του κώδικα, μετατρέπουμε τα στοιχεία των πινάκων σε δεκαδικούς αριθμούς. Αυτή είναι μια διαδικασία μορφοποίησης των εισαγόμενων στοιχείων σε μορφή χειρίσημη από το πρόγραμμα.

Στη γραμμή 17 του κώδικα, δηλώνουμε ότι η δεύτερη παράμετρος που θα εισάγουμε από τη φόρμα στο ArcGIS,θα αποθηκεύσει στην μεταβλητή zsthlhτην ονομασία της στήλης των δεδομένων, που βρίσκονται στο επιλεγμένο shapefile,και θα είναι οι τιμές τις εξεταζόμενης παραμέτρου.

Στη γραμμή 18 του κώδικα, μορφοποιείται η ονομασία της στήλης σε μορφή κειμένου (str)

Στις γραμμές 19,20 του κώδικα, όπως στις γραμμές10,11,12,13 τα στοιχεία της επιλεγμένης στήλης εισάγονται σε πίνακα Numpy και μορφοποιούνται σε δεκαδική μορφή.

Στις γραμμές 22,23,25,26 του κώδικα γίνεται η εισαγωγή και μορφοποίηση των παραμέτρων bandwith και lag από την Τρίτη και τέταρτη θέση της φόρμας εισαγωγής στο ArcGIS.

```
28    slag = np.arange(0,lag,bw)
29
30    pdis = squareform( pdist( Pin[:,:2] ))
```

Εικόνα 3.3: αρχική προεπεξεργασία

Στις γραμμές 28 έως 30 γίνεται η πρώτη προεπεξεργασία των δεδομένων. Στη γραμμή 28 δημιουργείτε πίνακας με τις αποστάσεις ανά χωρικό βήμα και στη γραμμή 30 δημιουργείτε μια μήτρα με τις αποστάσεις μεταξύ των σημείων με το ακόλουθο τύπο

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

3.1.2. Το εργαλείο Single Semivariogram

Το εργαλείο αυτό αναπτύχθηκε με σκοπό να δώσει στο χρήστη μία εικόνα για το πώς είναι το πειραματικό βαριόγραμμα των εκάστοτε δεδομένων, καθώς και για να μπορέσει να αναγνωρίσει ο χρήστης την ύπαρξη και την απόσταση της χωρικής εξάρτησης. Επίσης μπορεί να οπτικοποιήσει τη συμπεριφορά του βαριογράμματος στις διαφορετικές τιμές του χωρικού βήματος (h).

Αποτελείται από δύο συναρτήσεις που υπολογίζουν τις τιμές του πειραματικού ημιβαριογράμματος και αναλύονται παρακάτω:

- Η συνάρτηση distance:

Η συνάρτηση distance έχει ως σκοπό τη δημιουργία ενός δισδιάστατου πίνακα δεδομένων που θα περιέχει τις αποστάσεις όλων των σημείων και τις αντίστοιχες διαφορές τιμών μεταξύ των σημείων αυτών.

Αναλυτικότερα όπως φαίνεται στην εικόνα 3.3, στη γραμμή 32, παίρνει σαν είσοδο τον πίνακα με τις τιμές των σημείων, καθώς και τον πίνακα που έχουμε ήδη δημιουργήσει, ο οποίος περιέχει τη μήτρα των αποστάσεων των σημείων. Στη γραμμή 35 αναγνωρίζεται το πλήθος των ζευγών των σημείων. Στις σειρές 36, 37 υπάρχει μία εμφωλευμένη δομή επανάληψης όπου ώς σκοπό έχει την ταυτόχρονη προσπέλαση των δεδομένων της μήτρας αποστάσεων και κατά αντιστοιχία την εύρεση του τετραγώνου των διαφορών στις τιμές που έχουν τα σημεία αυτά.

$$\left(Z(x_i) - Z(x_j)\right)^2$$

Στις σειρές του κώδικα 40 έως 43 τοποθετούμε τα νέα δεδομένα σε ένα νέο πίνακα. Τον πίνακα αυτόν κατανέμουμε κατά αύξουσα σειρά και τον επιστρέφουμε.

```python
def distance(pin, pdis):
    pdis2=[]
    pdif = []
    N = len(Pin[:,0])
    for i in range(N):
        for j in range((i+1),N):
            pdis2.append(pdis[i,j])
            pdif.append((Pin[i,2] - Pin[j,2] )**2.0)
    ola = np.array([pdis2,pdif]).T
    ind = np.lexsort((ola[:,1],ola[:,0]))
    ola= ola[ind]
    return ola
```

Εικόνα 3.4: Η Συνάρτηση distance

- Η συνάρτηση semi:

Η συνάρτηση semi έχει ως σκοπό τη δημιουργία ενός δισδιάστατου πίνακα που θα περιέχει τις τιμές του βαριογράμματος, καθώς και τον αριθμό των ζευγών που συμμετέχουν ανά χωρικό βήμα (h).

Αναλυτικότερα, όπως φαίνεται στην εικόνα 3.5, στη σειρά 46 φαίνεται ότι λαμβάνει σαν είσοδο τον πίνακα των τιμών, τη μήτρα των αποστάσεων, τον πίνακα των καθορισμένων αποστάσεων του βαριογράμματος, καθώς και το σταθερό χωρικό βήμα μεταξύ των αποστάσεων του βαριογράμματος. Στη σειρά 43, καλούμε τη συνάρτηση distance που περιγράφτηκε παραπάνω, με σκοπό να μας δώσει τον πίνακα τιμών και αποστάσεων. Στις σειρές 44 εώς 46, δηλώνουμε τις απαραίτητες λίστες που θα αποθηκευτούν προσωρινά τα αποτελέσματα. Στις σειρές 47 εώς 53, υπολογίζουμε τις τιμές του βαριογράμματος για κάθε απόσταση βάση του τύπου

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}\{Z(x_i) - Z(x_i + h)\}^2$$

Στη σειρά 54, βάζουμε όλες τις τιμές που υπολογίσαμε σε έναν πίνακα numpy αφού πρώτα ελέγξουμε πως όλες οι τιμές είναι έγκυρες. Στη σειρά 55, η συνάρτηση επιστρέφει τον πίνακα σε μορφή όπου όλες οι τιμές είναι τοποθετημένες σε στήλες.

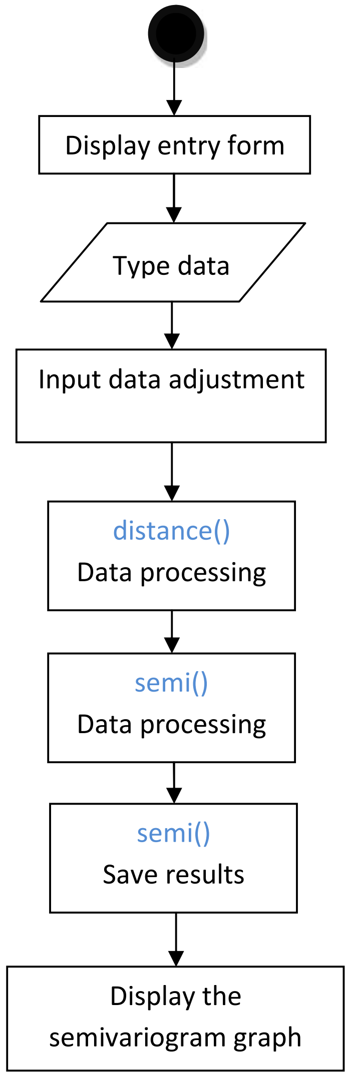
```
46 def semi(Pin,slag,bw,pdis):
47     Pin=distance(Pin, pdis)
48     se=list()
49     pingh=list()
50     piN=list()
51     for h in slag:
52         idx=Pin[:,0]<=h+bw
53         N=len(Pin[idx])
54         piN.append(int(N))
55         s = (np.sum(Pin[idx], axis=0))[1]
56         x=s/(N*2.0)
57         se.append(x)
58     pingh = [ [ slag[i], se[i], piN[i]]for i in range(len(slag)) if se[i]>0]
59     return np.array(pingh).T
```

Εικόνα 3.5: Η Συνάρτηση semi

- Τέλος στις σειρές του κώδικα 57 έως 65, οπτικοποιούμε με τη βοήθεια της βιβλιοθήκης matplotlib, το τελικό διάγραμμα των διακυμάνσεων.

```
57 sv = semi(Pin,slag,bw)
58 N =len(slag)
59 plt.plot( sv[0], sv[1], '.-' );
60 plt.ylabel('Semivariance')
61 plt.xlabel("Lag (m)")
62 plt.savefig('semivariogram.png',fmt='png',dpi=200)
63 plt.show()
```

Έικόνα 3.6: Η Οπτικοποίηση του Ημιβαριογράμματος



Εικόνα 3.7: Διάγραμμα Ροής του εργαλείου SingleVariogram

3.1.3. Το εργαλείο Semivariogram Comparison.

Το εργαλείο αυτό αποτελεί παραλλαγή του εργαλείου single semivariogram με σκοπό να δώσει στο χρήστη μία εικόνα, για την επιλογή του κατάλληλου δείγματος αποστάσεων. Οπτικοποιεί παράλληλα τρία ημιβαριογράμματα με απόσταση βήματος h, καθώς και με 2h και h/2. Πέραν αυτού ο κώδικας είναι ίδιος.

```
import arcpy
import matplotlib.pyplot as plt #για να εισάγουμε το ματλαμπ (διαγράμματα)
import numpy as np
from scipy.spatial.distance import pdist, squareform #μας λύνει τα χέρια σε ότι αφορά τον υπολογισμό των αποστάσεων με συντεταγμένες
from Tkinter import *
from arcpy.sa import *
input = arcpy.GetParameter(0)
x = arcpy.da.TableToNumPyArray(input,"SHAPE@X")
x = np.array( x, dtype=np.float )
y =arcpy.da.TableToNumPyArray(input,"SHAPE@Y")
y = np.array( y, dtype=np.float )
zsthlh = arcpy.GetParameter(1)
zsthlh = str(zsthlh)
z = arcpy.da.TableToNumPyArray(input,zsthlh)
z = np.array( z, dtype=np.float )
Pin= np.array([x,y,z]).T

bw = arcpy.GetParameter(2)
bw = int(bw)
lag = arcpy.GetParameter(3)
lag = int(lag)
lag += bw #έχει σχέση με την προσπέλαση των δεδομένων στον πίνακα
slag = np.arange(0,lag,bw)
pdis = squareform( pdist( Pin[:,:2] ))

slag1 = np.arange(0,lag,bw)
slag2 = np.arange(0,lag,bw*2)
slag3 = np.arange(0,lag,bw/2)

def distance(pin):
    pdis2=[]
    pdif = []
    N = len(Pin[:,0])
    for i in range(N):
        for j in range((i+1),N):
            pdis2.append(pdis[i,j])
            pdif.append((Pin[i,2] - Pin[j,2] )**2.0)
    ola = np.array([pdis2,pdif]).T
    ind = np.lexsort((ola[:,1],ola[:,0]))
    ola= ola[ind]
    return ola
ola=distance(Pin)

def semi(Pin,slag,bw):
    Pin=distance(Pin)
    se=list()
    pingh=list()
    piN=list()
    for h in slag:
        idx=Pin[:,0]<=h+bw
        N=len(Pin[idx])
        piN.append(int(N))
        s = (np.sum(Pin[idx], axis=0))[1]
        x=s/(N*2.0)
        se.append(x)
    pingh = [ [ slag[i], se[i], piN[i]]for i in range(len(slag)) if se[i]>0]
    return np.array(pingh).T

sv1 = semi(Pin,slag,bw)
sv2 = semi(Pin,slag2,bw*2)
sv3 = semi(Pin,slag3,bw/2)
N =len(slag)
plt.plot( sv1[0], sv1[1], '.-' );
plt.plot( sv2[0], sv2[1],'.-' ) ;
plt.plot( sv3[0], sv3[1],'.-' ) ;
plt.ylabel('Semivariance')
plt.xlabel("Lag (m)")
plt.savefig('semicomparison.png',fmt='png',dpi=200)
plt.show()
```

Εικόνα 3.8: Ο κώδικας του εργαλείου semivariogram comprison

3.1.4. Το εργαλείο Model Selection Criteria

Το εργαλείο αυτό αναπτύχθηκε με σκοπό όχι μόνο τη δημιουργία του κατάλληλου πειραματικού ημιβαριογράμματος, αλλά και την εύρεση του καλύτερα προσαρμοζόμενου μοντέλου ημιβαριογράμματος. Υπολογίζει για κάθε μοντέλο τις απαραίτητες τιμές των παραμέτρων ώστε να προσαρμόζονται καλύτερα στο πειραματικό ημιβαριόγραμμα και το οπτικοποιεί παρουσιάζοντας και τις τιμές των παραμέτρων του. Στη συνέχεια βάσει των κριτηρίων AIC, AICc, BIC καιHQC, βοηθά στην επιλογή του κατάλληλου μοντέλου ημιβαριογράμματος σύμφωνα με τις τιμές που προκύπτουν από αυτά. Οι μικρότερες τιμές ανά κριτήριο σημαίνουν μικρότερη απώλεια πληροφορίας στην μετάβαση από το πειραματικό ημιβαριόγραμμα στο μοντέλο ημιβαριογράμματος. Ο κώδικας του εργαλείου αποτελείται από 7 συναρτήσεις που περιγράφονται αναλυτικά παρακάτω.

- Η συνάρτηση distance:
  Είναι όμοια με αυτήν που έχει περιγραφεί παραπάνω.

- Η συνάρτηση semi:
  Είναι όμοια με αυτήν που έχει περιγραφεί παραπάνω.

- Η συνάρτηση evrosnugg:

Σκοπός αυτής της συνάρτησης είναι η εύρεση των κατάλληλων τιμών των παραμέτρων του κάθε μοντέλου ημιβαριογράμματος για τη βέλτιστη προσαρμογή αυτού στα πειραματικά δεδομένα.
Όπως φαίνεται από την Εικόνα 3.9, στη σειρά 63, σαν είσοδο λαμβάνει το είδος του μοντέλου (σφαιρικό εκθετικό ή κανονικό), τις τιμές του πειραματικού ημιβαριογράμματος, και το μέρος του χωρικού βήματος(h). Στη σειρά 64, υπολογίζεται το όριο (sill, C1), που είναι η διαφορά της μέγιστης και ελάχιστης τιμής του πειραματικού ημιβαριογράμματος. Στη σειρά 65, δημιουργούμε ένα δισδιάστατο πίνακα μεγέθους 500X50 όπου θα τοποθετηθούν οι τιμές του αθροίσματος των

τετραγώνων των διαφορών. Στη σειρά 61, χωρίζουμε το εύρος του πειραματικού ημιβαριογράμματος σε 500 ίσα διαστήματα. Στη σειρά 67, χωρίζουμε τη διακύμανση επι της αρχής (nugget) σε 50 ίσα διαστήματα. Στις σειρές 68 έως 70, υπολογίζουμε για όλου τους συνδυασμούς εύρους και nugget, τις τιμές του αθροίσματος του τετραγώνου των διαφορών και το διαιρούμαι με το πλήθος τους, με τον παρακάτω τύπο.

$$r = \frac{rss}{n} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\chi}_i - \chi_i)^2$$

Στη σειρά 72, επιλέγουμε τη μικρότερη τιμή αθροίσματος του τετραγώνου των διαφορών και τέλος στη σειρά 73, η συνάρτηση επιστρέφει την τιμή αυτή μαζί με το αντίστοιχο εύρος και την αντίστοιχη διακύμανση επί της αρχής (nugget) και το αντίστοιχο όριο (sill,C1).

```
63 def evrosnugg( fct, x, y, meshSize=500,mesSize = 50 ):
64     C1=np.abs(y[-1]-y[0])
65     rs = np.zeros( (meshSize,mesSize) )
66     a = np.linspace( x[1], x[-1], meshSize )
67     nugget =np.linspace(0,y[0],mesSize)
68     for i in range( meshSize ):
69         for j in range(mesSize):
70             rs[i,j] = np.mean( ( y - fct( x, a[i], C1, nugget[j] ) )**2.0 )
71 
72     z= np.unravel_index(rss.argmin(), rss.shape)
73     return rs[z[0],z[1]],a[z[0]],nugget[z[1]],C1
```

Εικόνα 3.9: Η Συνάρτησηevrosnugg

- Η Συνάρτηση cvmodel:

Η συγκεκριμένη συνάρτηση είναι μια κομβική συνάρτηση που έχει ως σκοπό να μαζέψει τα απαραίτητα δεδομένα και να επιστρέψει τις τιμές του πειραματικού και θεωρητικού ημιβαριογράμματος.

Όπως φαίνεται στην Εικόνα 3.10, στη σειρά 75, σαν είσοδο εισάγωνται τα δεδομένα του πίνακα με τις τιμές των σημείων, η μήτρα ε τις αποστάσεις μεταξύ των σημείων ο πίνακας με τις αποστάσεις ανά χωρικό βήμα (h), η σταθερή του διαφορά (bw) καθώς και το είδος του μοντέλου βαριογράμματος σε κάθε περίπτωση.

```
75  def cvmodel( Pin, model, slag,bw,pdis):
76
77      sv = semi( Pin, slag, bw, pdis)
78      #υπολογιζουμε τις πραγματικες τιμες του βαριογράμματος
79
80      rs,evros,nugget,C1 = evrosnugg( model, sv[0], sv[1])
81
82      covfct = lambda h, a=evros: model( h, a, C1, nugget )
83
84      return covfct,sv,rs,evros,nugget
```

Εικόνα 3.10: Η Συνάρτηση cvmodel

Στη σειρά 77, καλείτε η συνάρτηση semi να υπολογίσει τις τιμές του πειραματικού ημιβαριογράμματος
Στη σειρά 80 υπολογίζονται καλώντας την συνάρτηση evrosnugg, οι απαραίτητες παράμετροι για την δημιουργία του πειραματικού ημιβαριογράμματος (όριο, εύρος, διακύμανση επι της αρχής), καθώς και ο μέσος όρος των αθροισμάτων των τετραγώνων των διαφορών που θα χρησιμοποιηθεί αργότερα για τον υπολογισμό των τιμών των κριτηρίων.
Στη σειρά 82 υπολογίζονται οι τιμές του θεωρητικού ημιβαριογράμματος με τις τιμές των παραμέτρων που δόθηκαν στην γραμμή 80.
Τέλος στη γραμμή 84, η συνάρτηση επιστρέφει τις τιμές του πειραματικού ημιβαριογράμματος, τις τιμές του θεωρητικού ημιβαριογράμματος καθώς και τον μέσο όρο των αθροισμάτων των τετραγώνων των διαφορών.

- Η Συνάρτηση spherical:

Σκοπός αυτής της συνάρτησης είναι ο υπολογισμός των τιμών του θεωρητικού ημιβαριογράμματος με τη συνάρτηση του σφαιρικού μοντέλου σε κάθε καθορισμένη απόσταση.

Όπως φαίνεται στην Εικόνα 3.11, στη σειρά 86, σαν είσοδο παίρνει την τιμή του χωρικού βήματος, το εύρος του βαριογράμματος, την διακύμανση επί της αρχής και το όριο του ημιβαριογράμματος.

Η συνάρτηση μετατρέπει τα δεδομένα σε πίνακες δεδομένων, αν δεν είναι ήδη, και στη συνέχεια κάνει πράξεις πινάκων και υπολογίζει την τιμή για την κάθε απόσταση βάση του τύπου:

$$\gamma(h) = C_0 + C_1\left[1{,}5\,\frac{|h|}{a} - 0{,}5\left(\frac{|h|}{a}\right)^3\right] \quad \text{για} \quad 0 < |h| < a$$

$$\gamma(h) = 0 \text{για} h = 0$$

```
86  def spherical( h, a, C1, nugget):
87      '''
88      σφαιρικό μοντέλο
89      '''
90      # αν είναι δεκαδικά
91      if type(h) == np.float64:
92          # για το εύρος που δόθηκε υπολογίζει τις τιμές του μοντέλου
93          if h <= a:
94              return (C1*( 1.5*(h/a) - 0.5*(h/a)**3.0 ))+nugget
95          else:
96              return C1
97      # αλλιώς δημιουργεί τους ανάλογους πίνεκς ώστε να είναι δυνατοί οι υπολογίσμοί
98      else:
99          a = np.ones( h.size ) * a
100         C1 = np.ones( h.size ) * C1
101         nugget = np.ones(h.size)* nugget
102         return map( spherical, h, a, C1, nugget)
```

Εικόνα 3.11: Η Συνάρτηση spherical

- Η Συνάρτηση exponential:

Σκοπός αυτής της συνάρτησης είναι ο υπολογισμός των τιμών του θεωρητικού ημιβαριογράμματος με τη συνάρτηση του εκθετικού μοντέλου σε κάθε καθορισμένη απόσταση.

Όπως φαίνεται στην Εικόνα 3.12, στη σειρά 104, σαν είσοδο παίρνει την τιμή του χωρικού βήματος, το εύρος του βαριογράμματος, την διακύμανση επί της αρχής και το όριο του ημιβαριογράμματος.

Η συνάρτηση μετατρέπει τα δεδομένα σε πίνακες δεδομένων, αν δεν είναι ήδη, και στη συνέχεια κάνει πράξεις πινάκων και υπολογίζει την τιμή για την κάθε απόσταση βάση του τύπου:

$$\gamma(h) = C_0 + C_1\left[1 - e^{\frac{-|h|}{r}}\right] \quad \text{για} \quad |h| > 0$$

$$\gamma(h) = 0 \text{για} |h| = 0$$

```
104 def exponential( h, a, C1, nugget):
105     '''
106     εκθετικό μοντέλο
107     '''
108     # αν είναι δεκαδικά
109     if type(h) == np.float64:
110         # για το εύρος που δόθηκε υπολογίζει τις τιμές του μοντέλου
111 
112         return (C1*( 1-np.exp(-(np.absolute(h))/a )))+nugget
113 
114     # αλλιώς δημιουργεί τους ανάλογους πίνεκς ώστε να είναι δυνατοί οι υπολογίσμοί
115     else:
116         a = np.ones( h.size ) * a
117         C1 = np.ones( h.size ) * C1
118         nugget = np.ones(h.size)* nugget
119         return map( exponential, h, a, C1, nugget )
```

Εικόνα 3.12: Η Συνάρτηση exponential

- Η Συνάρτηση gaussian:

Σκοπός αυτής της συνάρτησης είναι ο υπολογισμός των τιμών του θεωρητικού ημιβαριογράμματος με τη συνάρτηση του κανονικού μοντέλου (γκαουσιανού ή Gaussian) σε κάθε καθορισμένη απόσταση.

Όπως φαίνεται στην Εικόνα 3.13, στη σειρά 121, σαν είσοδο παίρνει την τιμή του χωρικού βήματος, το εύρος του βαριογράμματος, την διακύμανση επί της αρχής και το όριο του ημιβαριογράμματος.

Η συνάρτηση μετατρέπει τα δεδομένα σε πίνακες δεδομένων, αν δεν είναι ήδη, και στη συνέχεια κάνει πράξεις πινάκων και υπολογίζει την τιμή για την κάθε απόσταση βάση του τύπου:

$$\gamma(h) = C_0 + C_1\left[1 - e^{\left(\frac{-|h|}{r}\right)^2}\right]$$

$$\gamma(h) = 0$$

```
121 def gaussian( h, a, C1, nugget):
122     '''
123     γκαουσιανό μοντέλο
124     '''
125     # αν είναι δεκαδικά
126     if type(h) == np.float64:
127         # για το εύρος που δόθηκε υπολογίζει τις τιμές του μοντέλου
128         return (C1*( 1-np.exp(-(np.absolute(h))**2/a**2 )))+nugget
129 
130     # αλλιώς δημιουργεί τους ανάλογους πίνακες ώστε να είναι δυνατοί οι υπολογίσμοί
131     else:
132         a = np.ones( h.size ) * a
133         C1 = np.ones( h.size ) * C1
134         nugget = np.ones(h.size)* nugget
135         return map( gaussian, h, a, C1, nugget)
```

Εικόνα 3.13: Η Συνάρτηση gaussian

- Η Συνάρτηση criteria

Η συνάρτηση criteria αποτελεί μια κομβική συνάρτηση που έχει ως σκοπό να δώσει τα απαραίτητα δεδομένα και να επιστρέψει τις τιμές των κριτηρίων AIC, AICc, BIC, HQC καθώς και να οπτικοποιήσει τα βέλτιστα υπολογισμένα μοντέλα ημιβαριογραμμάτων μαζί με τις τιμές των παραμέτρων τους.

Όπως φαίνεται στην εικόνα 3.14, στην σειρά 137, η συνάρτηση παίρνει σαν είσοδο τον πίνακα των τιμών των σημείων, την μήτρα των αποστάσεων μεταξύ των σημείων, των πίνακα με τις αποστάσεις των χωρικών βημάτων και την τιμή της ισαποχής τους. Εν συνεχεία στην σειρά 138 καλεί την συνάρτηση cvmodel με είσοδο το είδος του σφαιρικού μοντέλου ώστε να λάβει τις τιμές του πειραματικού και του θεωρητικού ημιβαριογράμματος καθώς και τον μέσο όρο του αθροίσματος του τετραγώνου των διαφορών των τιμών, πειραματικού με θεωρητικού ημιβαριογράμματος.

Στη σειρά 140, υπολογίζεται η τιμή του κριτηρίου AIC, για το σφαιρικό μοντέλο, χρησιμοποιήθηκε ο τύπος

$$AIC = n \ln(RSS/n) + 2k$$

Στη σειρά 141 βάση της υπολογισθείσας τιμής του AIC στη σειρά 140 υπολογίζεται η τιμή του κριτηρίου AICc, χρησιμοποιήθηκε ο τύπος

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

Στη σειρά 142, υπολογίζεται η τιμή του κριτηρίου BIC, για το σφαιρικό μοντέλο, χρησιμοποιήθηκε ο τύπος

$$BIC = n \ln(RSS/n) + k \ln(n)$$

Στη σειρά 143, υπολογίζεται η τιμή του κριτηρίου HQC, για το σφαιρικό μοντέλο, χρησιμοποιήθηκε ο τύπος

$$HQC = n\ log\left(\frac{RSS}{n}\right) + \ 2k\ log\big(log(n)\big)$$

Στις σειρές 145 με 151 η συνάρτηση με την βοήθεια της βιβλιοθήκης matplotlib εμφανίζει τα διαγράμματα του πειραματικού ημιβαριογράμματος με το αντίστοιχα προσαρμοσμένο σφαιρικό μοντέλο.

Όμοια διαδικασία ακολουθείτε στις σειρές 155 έως 167 για το εκθετικό μοντέλο και στις σειρές 171 έως 183 για κανονικό ή γκαουσιανό (Gaussian) μοντέλο.

Τέλος σαν επιστροφή υπάρχει ένα μορφοποιημένο κείμενο όπου περιέχει τις τιμές των κριτηρίων ανά μοντέλο με σκοπό να δοθούν στην συνάρτηση gui που θα αναλυθεί παρακάτω.

```
def criteria(pin, slag, bw, pdis)
    sp,sv,rs,evros,nugget=cvmodel(Pin,spherical,slag,bw,pdis)
    N=len(sv[1])
    A=np.abs((N*np.log(rs))+4)
    AC=np.abs(A+(12/(N-3)))
    B=np.abs((N*np.log(rs))+(2*np.log(N)))
    H=np.abs((N*np.log10(rs))+(4*np.log10(np.log10(N))))

    plt.plot( sv[0], sv[1], '.-' );
    plt.plot( sv[0], sp( sv[0] ) ) ;
    plt.title('Spherical Model')
    plt.ylabel('Semivariance')
    plt.xlabel("Lag (m)\n range : %d, nugget : %d"%(evros,nugget))
    plt.savefig('semivariogram_model_spherical.png',fmt='png',dpi=200)
    plt.show()
    sp,sv,rs,evros,nugget=cvmodel(Pin,exponential,slag,bw,pdis)
    A2=np.abs((N*np.log(rs))+4)
    AC2=np.abs(A2+(12/(N-3)))
    B2=np.abs((N*np.log(rs))+(np.log(N)))
    H2=np.abs((N*np.log10(rs))+(4*np.log10(np.log10(N))))

    plt.plot( sv[0], sv[1], '.-' );
    plt.plot( sv[0], sp( sv[0] ) ) ;
    plt.title('Exponential Model')
    plt.ylabel('Semivariance')
    plt.xlabel("Lag (m)\n range : %d, nugget : %d"%(evros,nugget))
    plt.savefig('semivariogram_model_exponential.png',fmt='png',dpi=200)
    plt.show()
    sp,sv,rs,evros,nugget=cvmodel(Pin,gaussian,slag,bw,pdis)
    A3=np.abs((N*np.log(rs))+4)
    AC3=np.abs(A3+(12/(N-3)))
    B3=np.abs((N*np.log(rs))+(np.log(N)))
    H3=np.abs((N*np.log10(rs))+(4*np.log10(np.log10(N))))

    plt.plot( sv[0], sv[1], '.-' );
    plt.plot( sv[0], sp( sv[0] ) ) ;
    plt.title('Gaussian Model')
    plt.ylabel('Semivariance')
    plt.xlabel("Lag (m)\n range : %d, nugget : %d"%(evros,nugget))
    plt.savefig('semivariogram_model_Gaussian.png',fmt='png',dpi=200)
    plt.show()
    apant="spherical model \n AIC : %d \n AICc : %d \n BIC : %d \n HQC : %

    return apant
```

Εικόνα 3.14: Η Συνάρτηση Criteria

- Η Συνάρτηση gui

Είναι η συνάρτηση όπου ξεκινά το πρόγραμμά και μας δίνει τις τελικές απαντήσεις από εδώ καλείτε η συνάρτηση criteria ώστε να υπολογισθούν το πειραματικό και τα θεωρητικά ημιβαριογράμματα. Ο τελικός σκοπός που επιτελεί η συγκεκριμένη συνάρτηση είναι να επιστέφει ένα αναδυόμενο παράθυρο με τις τιμές των κριτηρίων ανά μοντέλο.

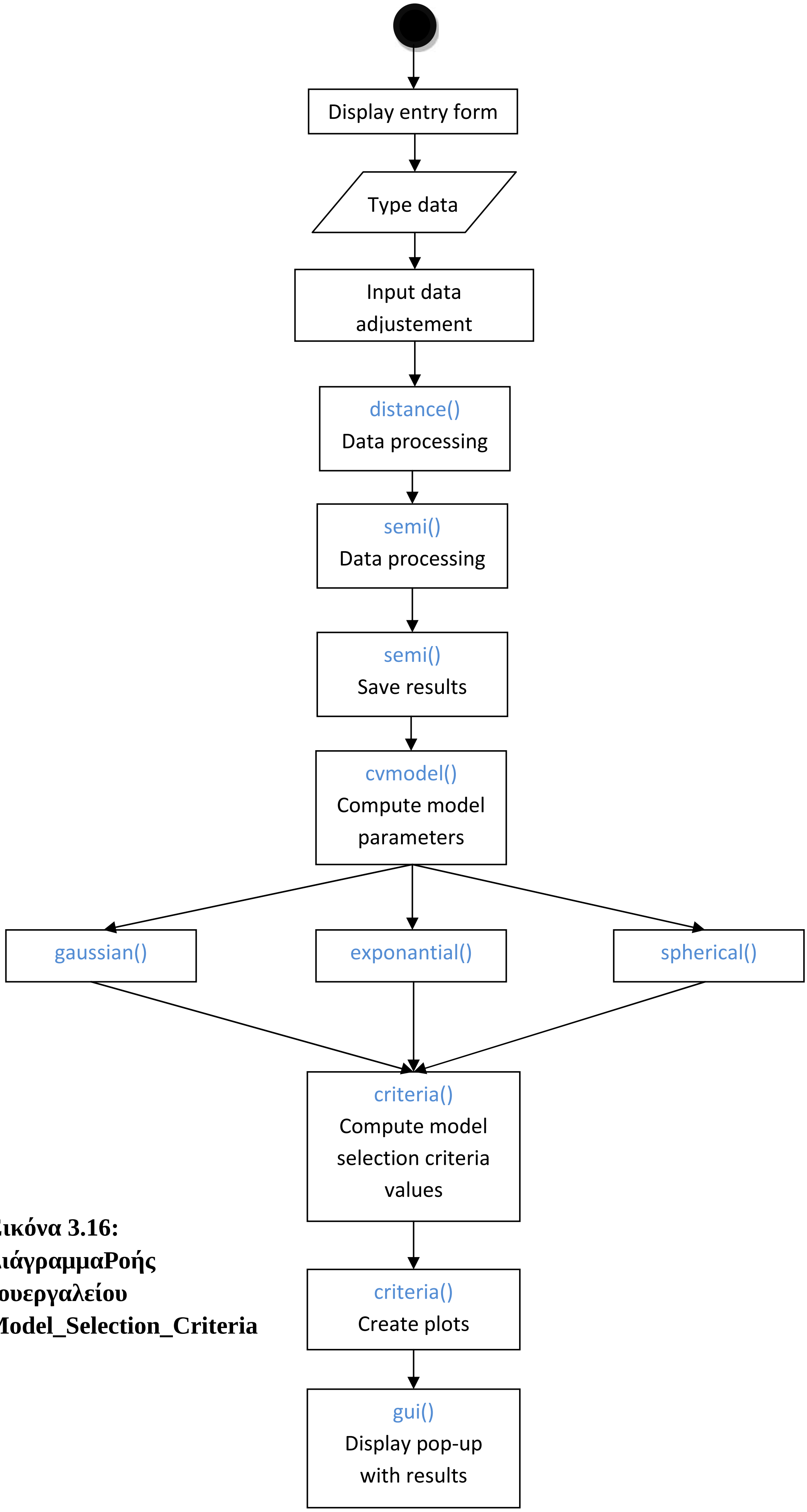
Αναλυτικότερα όπως φαίνεται στην εικόνα 3.15, στη σειρά 188, σαν είσοδο παίρνει τα δεδομένα τις προεπεξεργασίας, δηλαδή τον πίνακα με τις τιμές των σημείων, τον πίνακα με τις τιμές του χωρικού βήματος, την ισαποχή τους, και την μήτρα των αποστάσεων των σημείων.

Στη σειρά 189, καλεί την συνάρτηση criteria, ώστε άμεσα να πάρει το κείμενο, με τις τιμές των κριτηρίων ανά μοντέλο, που θα προβάλει αλλά και έμεσα για να ενεργοποιηθούν οι προβολές των ημιβαριογραμμάτων.

Τέλος στις σειρές 190 έως 195 με τη βοήθεια της βιβλιοθήκης Tkinter δημιουργείτε το αναδυόμενο παράθυρο που θα περιέχει τις τελικές τιμές των κριτηρίων.

```
188 def gui(Pin, slag, bw, pdis, input):
189     apant = criteria(Pin, slag, bw)
190     master = Tk()
191     lbl = Label(master,text = apant)
192     lbl.pack()
193     Q = Button(master, text='quit', command= master.quit)
194     Q.pack()
195     master.mainloop()
```

Εικόνα 3.15: Η Συνάρτηση gui



Εικόνα 3.16: ΔιάγραμμαΡοής τουεργαλείου Model_Selection_Criteria

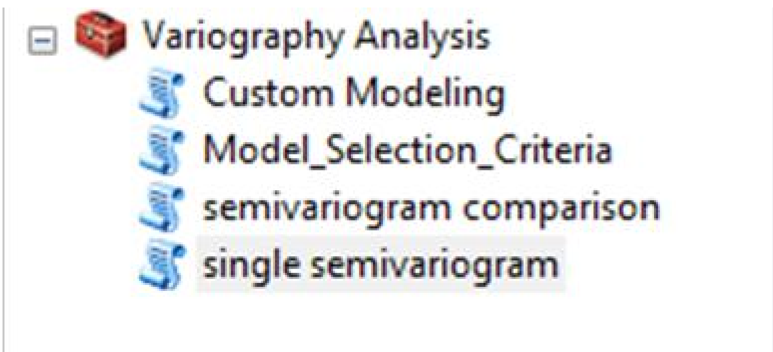
3.2 Εφαρμογή σε Πειραματικά Δεδομένα

Στο δεύτερο μέρος που ακολουθεί γίνεται εφαρμογή του των εργαλείων που αναπτύχθηκαν κατά τη διάρκεια της μεταπτυχιακής εργασίας σε πραγματικά πειραματικά δεδομένα που προέρχονται από εδαφολογικές μελέτες που εκπονήθηκαν από τον Τομέα Εδαφολογίας του ΓΠΑ.

Η πρώτη εφαρμογή έγινε σε δεδομένα κοκκομετρικής και ειδικότερα για την τιμή της περιεχόμενης στο έδαφος άργιλο σύστασης σε δείγματα που ελήφθησαν στα πλαίσια του ερευνητικού προγράμματος αξιολόγησης εδαφικών και κλιματικών συνθηκών εκτάσεων της κοινότητας Μουζάκι του νομού Μεσσηνίας για την καλλιέργεια ποικιλιών αμπέλου για οινοποίηση το έτος 2008 για λογαριασμό της Κυνηγός Α.Ε.. Χρησιμοποιήθηκαν και τα τρία εργαλεία με σειρά και τρόπο που πρέπει να γίνεται η χρήση τους.

Η δεύτερη εφαρμογή έγινε σε δεδομένα εδαφικών ιδιοτήτων που ελήφθησαν στα πλαίσια της δράσης για την προστασία του Αμβρακικού κόλπου από την γεωργική δραστηριότητα, με τίτλο «Δημιουργία εδαφικών χαρτών και έλεγχος της ποιότητας των εδαφών των πεδινών εκτάσεων του βορείου Αμβρακικού», το έτος 2009 για λογαριασμό της νομαρχίας Άρτας. Πιο συγκεκριμένα έγινε εφαρμογή στις τιμές του pH των εδαφικών δειγμάτων.

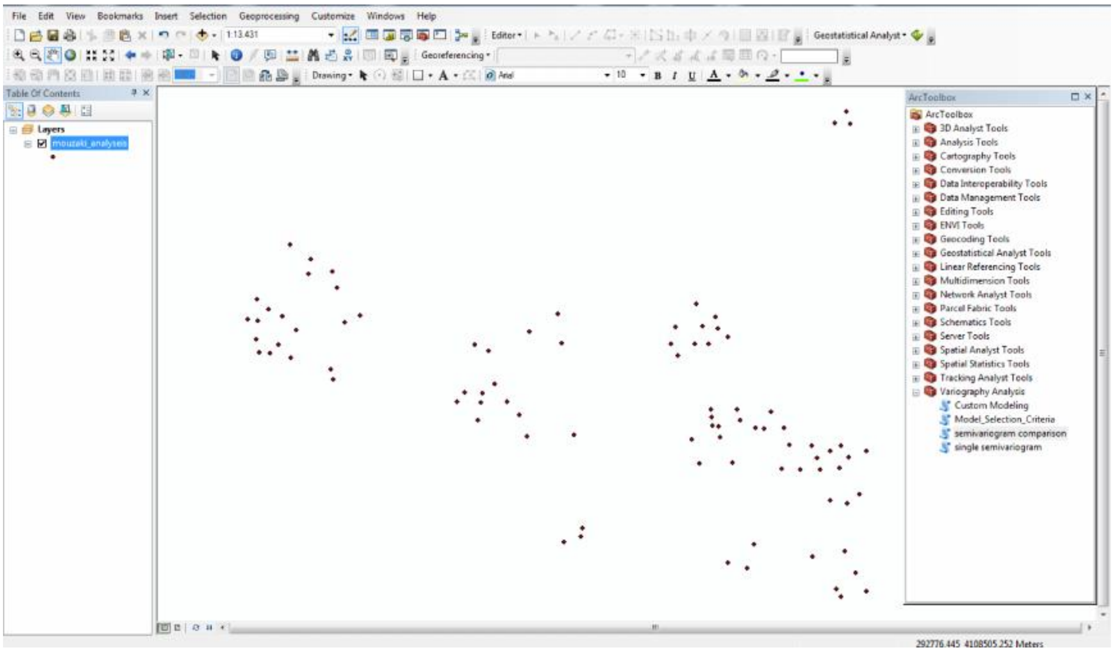
Η τρίτη εφαρμογή έγινε επίσης σε δεδομένα που πάρθηκαν από την περιοχή της Άρτας αλλά αυτή τη φορά έγινε εφαρμογή για τις τιμές της περιεχόμενης στο έδαφος άργιλο.



Εικόνα 3.17: Το toolbox που αναπτύχθηκε όπως φαίνεται στο περιβάλλον ArcGIS10.1

3.2.1 Εφαρμογή Πρώτη

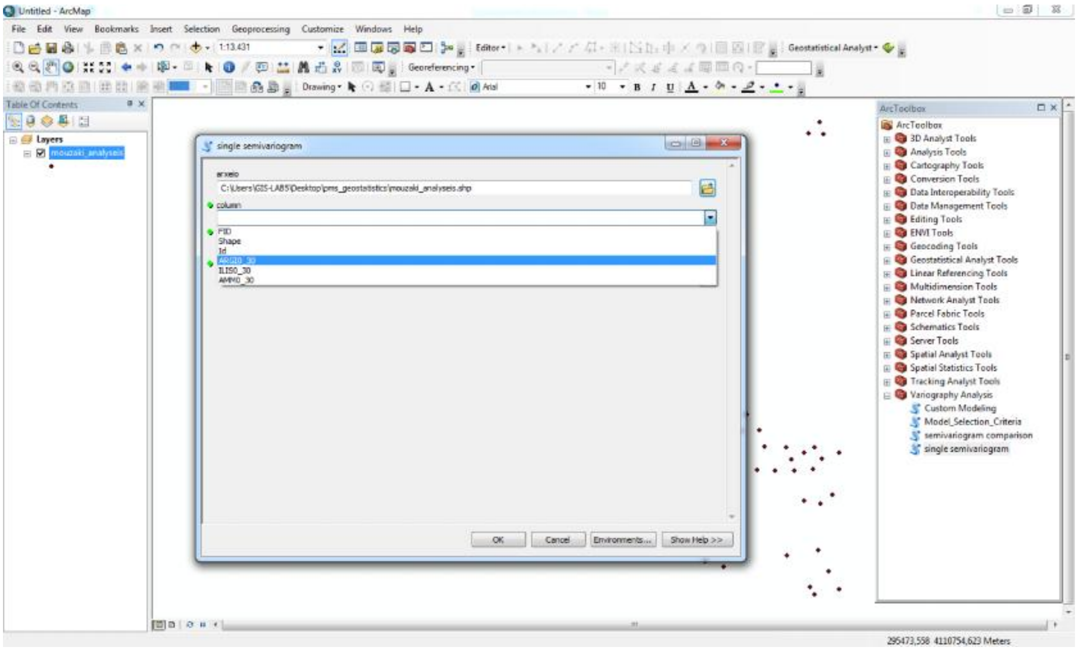
Στην πρώτη εφαρμογή έγινε χρήση όλων των εργαλείων που αναπτύχθηκαν στην παρούσα εργασία και ο τρόπος που επεξεργάστηκαν τα δεδομένα αποτελεί τον ενδεδειγμένο ώστε να πάρουμε το καλύτερα προσαρμοσμένο στα δεδομένα ημιβαριόγραμμα. Τα δεδομένα πάρθηκαν την περιοχή Μουζάκι που βρίσκεται στον νομό Μεσσηνίας. Είναι δεδομένα που περιγράφουν την κοκκομετρκή σύσταση του εδάφους. Από αυτά τα δεδομένα επιλέχθηκαν για επεξεργασία οι τιμές της περιεχόμενης στο έδαφος άργιλο.



Εικόνα 3.18: Τα σημεία του αρχείου mouzaki σε περιβάλλον ArcGIS10.1

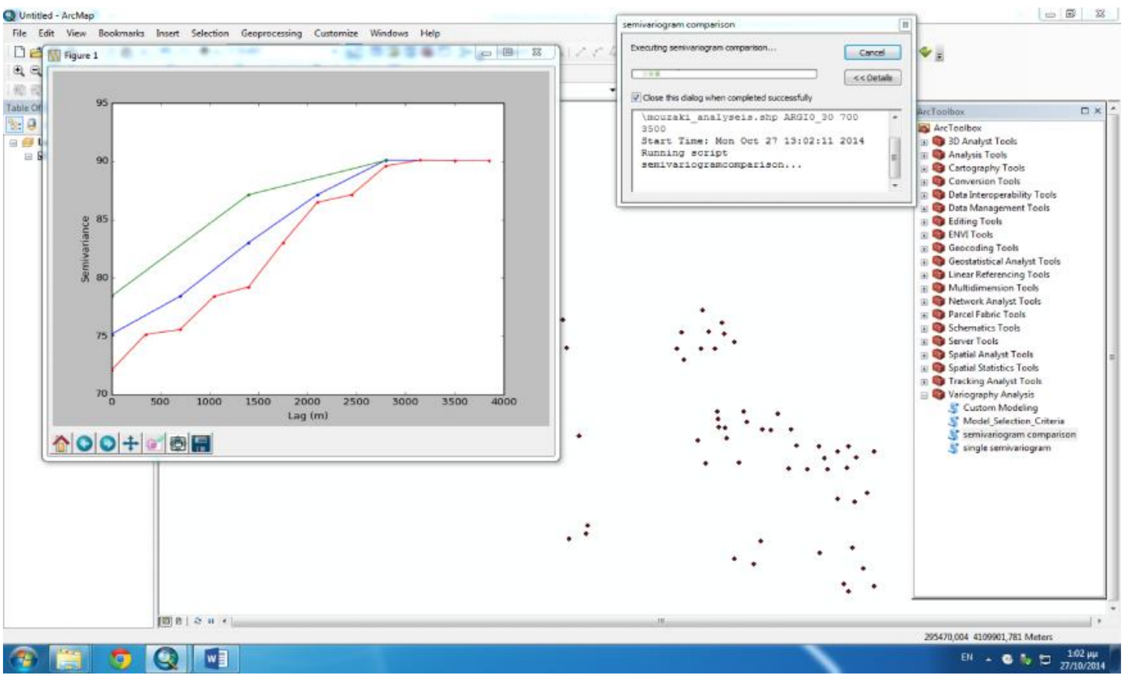
Αφού τα δεδομένα που βρίσκονται σε αρχείο μορφής shapefile (.shp) εισήχθησαν στο ArcGIS10.1 (εικόνα 3.18), πατάμε πάνω στο όνομα του εργαλείου semivariogram comparison και στην φόρμα που θα εμφανιστεί (εικόνα 3.19) εισάγουμε το shapefile που περιέχει τους πίνακες με τις τιμές των δεδομένων, επιλέγουμε την στήλη με τις τιμές της αργίλου και δίνουμε την τιμή 700 μέτρα στην τιμή του χωρικού βήματος και 4000 μέτρα το εύρος της αύξησης της διακύμανσης.

*(υποσημείωση στη σελίδα) Η επιλογή των τιμών των αρχικών τιμών του εύρους και του χωρικού βήματος στηρίζεται στην μέγιστη απόσταση των σημείων και τη μέση απόσταση των δειγματοληψιών που προσδιορίζονται εύκολα από διαδικασίες του ArcGIS10.1.



Εικόνα 3.19: Η φόρμα εισαγωγής δεδομένων

Το αποτέλεσμα των παραπάνω ενεργειών θα είναι μια εικόνα τριών πειραματικών ημίβαριογραμματων, με χωρικό βήμα στο κάθε ημιβαριόγραμμά, 700 , 350 και 1400 μέτρα όπως φαίνεται στην Εικόνα 3.20.

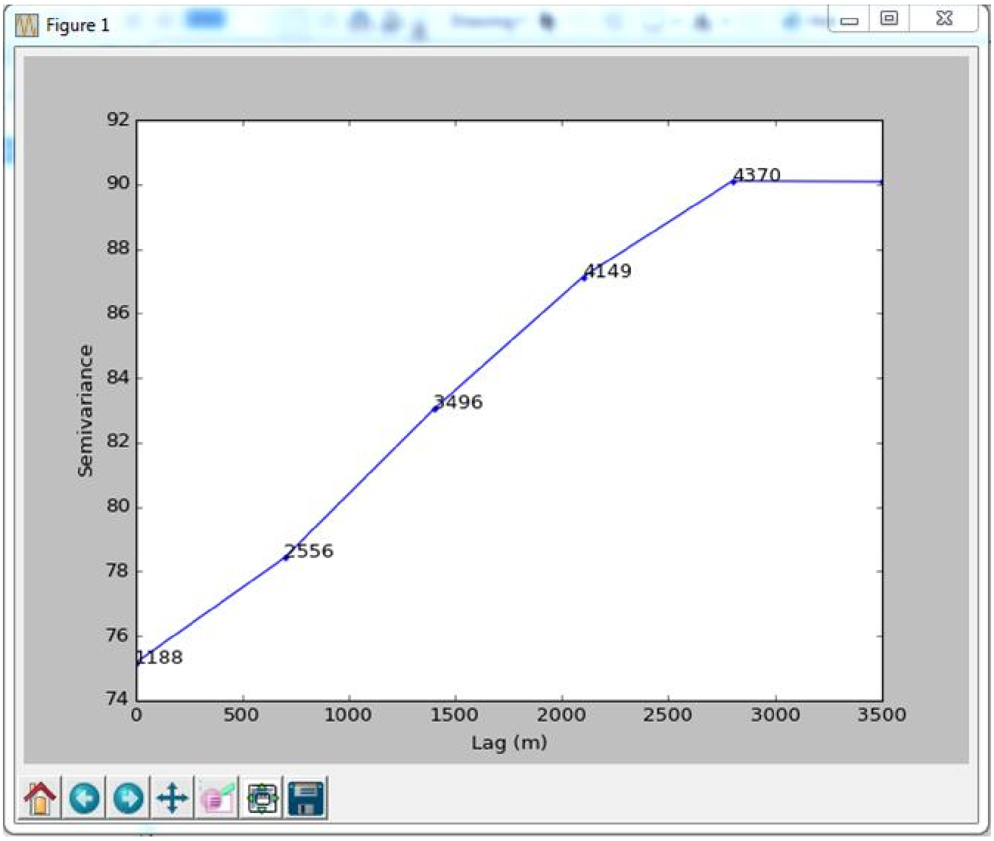


Εικόνα 3.20: Τα τρία πειραματικά ημιβαριογράμματα

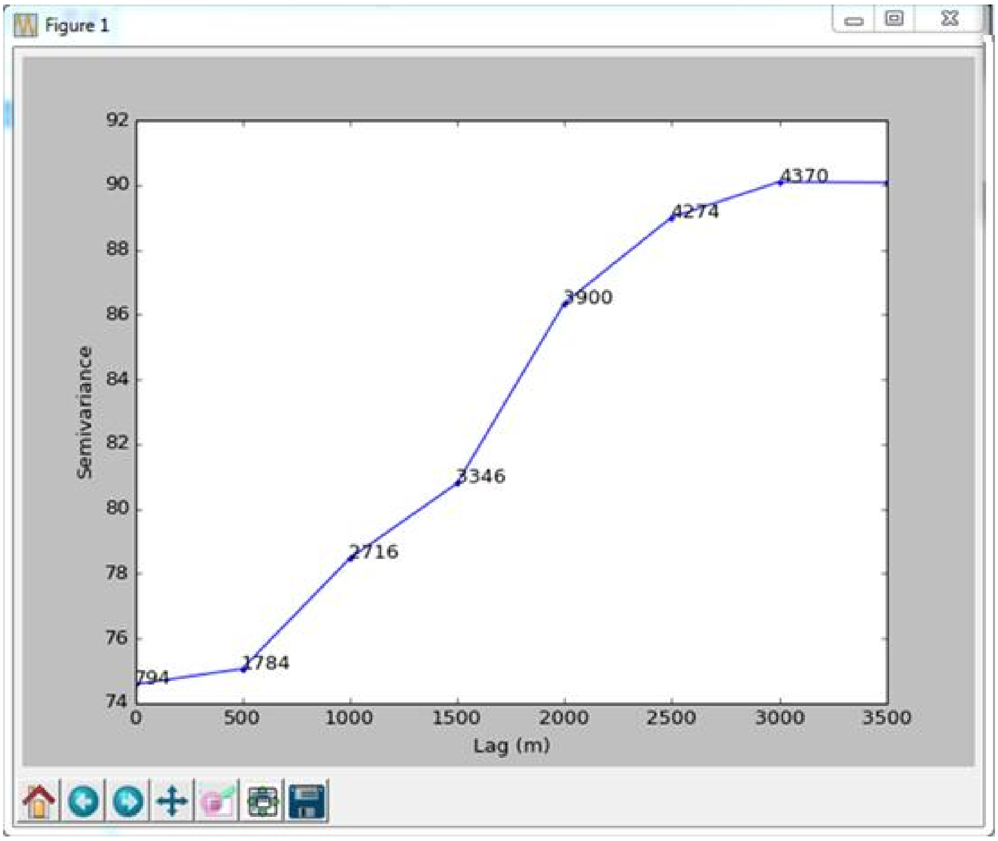
Από το διάγραμμα της Εικόνα 3.20 φαίνεται ότι η αύξηση της ημιδιακύμανσης έχει εύρος περίπου τα 3500 μέτρα και όσον αφορά την τιμή του χωρικού βήματος αυτή πρέπει να είναι σίγουρα μικρότερη από 700 μέτρα. Οπότε χρησιμοποιούμε το εργαλείο Single Semivariogram για να αποκτήσουμε μια καλύτερη εικόνα γύρω από το επιθυμητό πειραματικό ημιβαριόγραμμα.

Εισάγουμε με την ιδία διαδικασία τα δεδομένα αλλά αυτή την φορά επιλέγουμε ως εύρος αύξησης της διακύμανσης τα 3500 μέτρα και για την τιμή του χωρικού βήματος, διαδοχικά τις τιμές 700, 500 και 250 μέτρα.

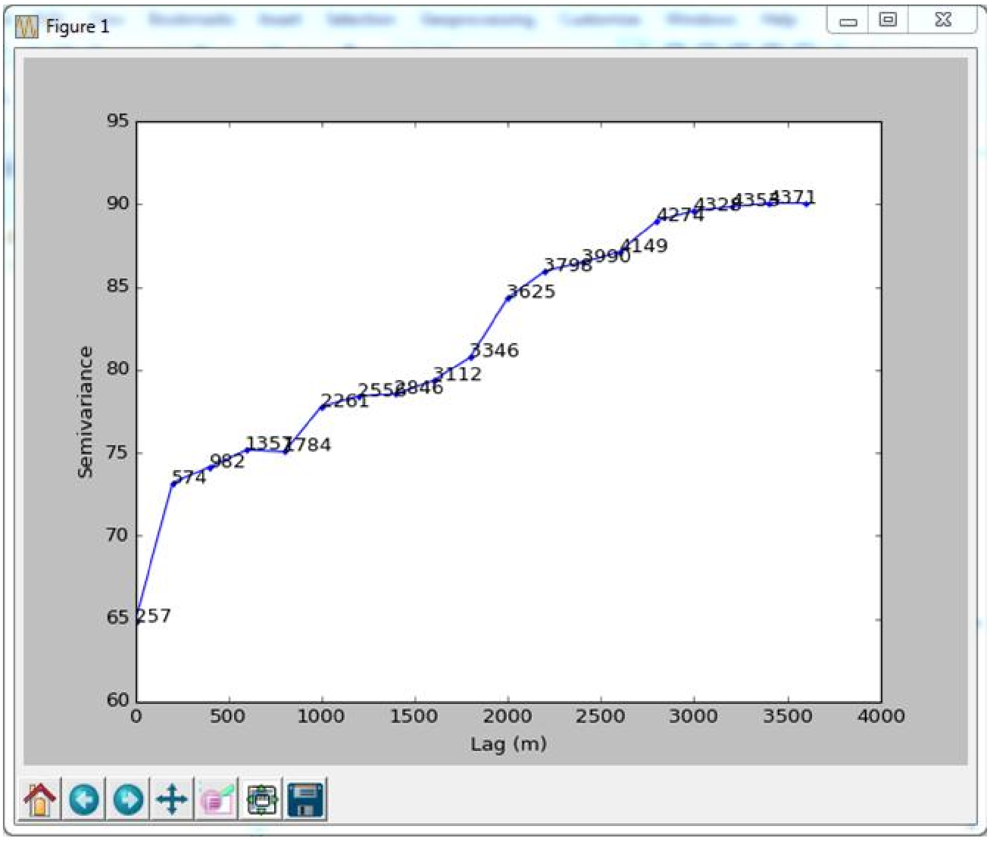
Το αποτέλεσμα φαίνεται στις παρακάτω εικόνες.



Εικόνα 3.21: ημιβαριόγραμμα με τιμή χωρικού βήματος 700 μέτρα



Εικόνα 3.22: Ημιβαριόγραμμα με τιμή χωρικού βήματος 500 μέτρα

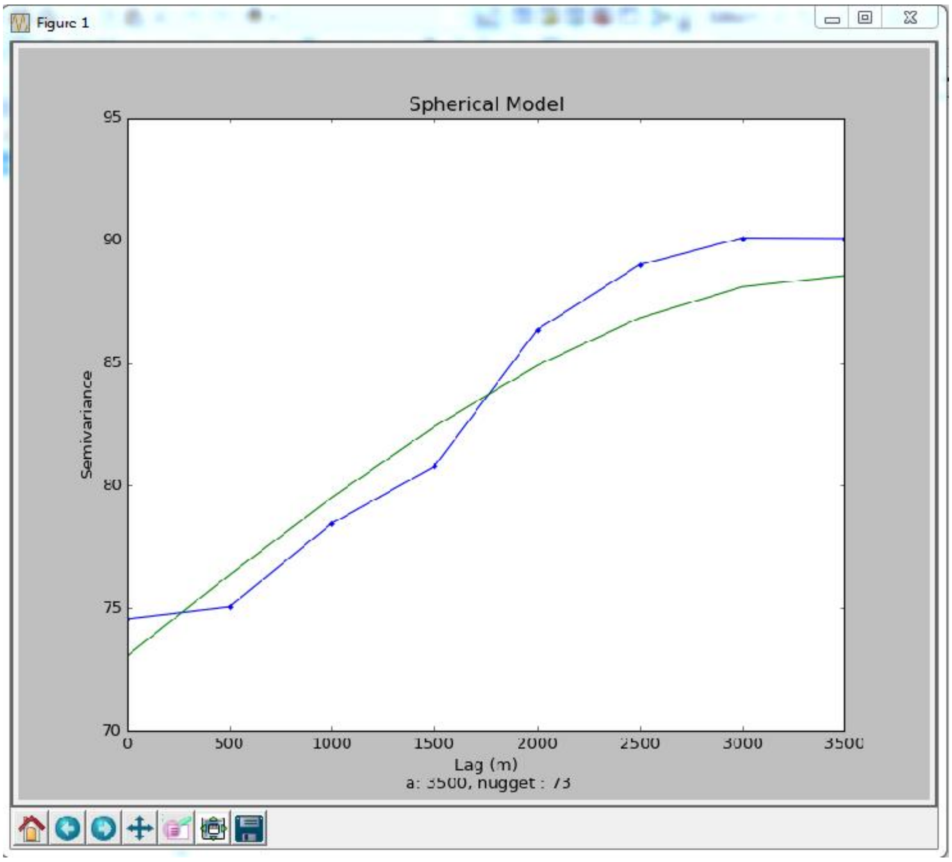


Εικόνα 3.23: Ημιβαριόγραμμα με τιμή χωρικού βήματος 250 μέτρα

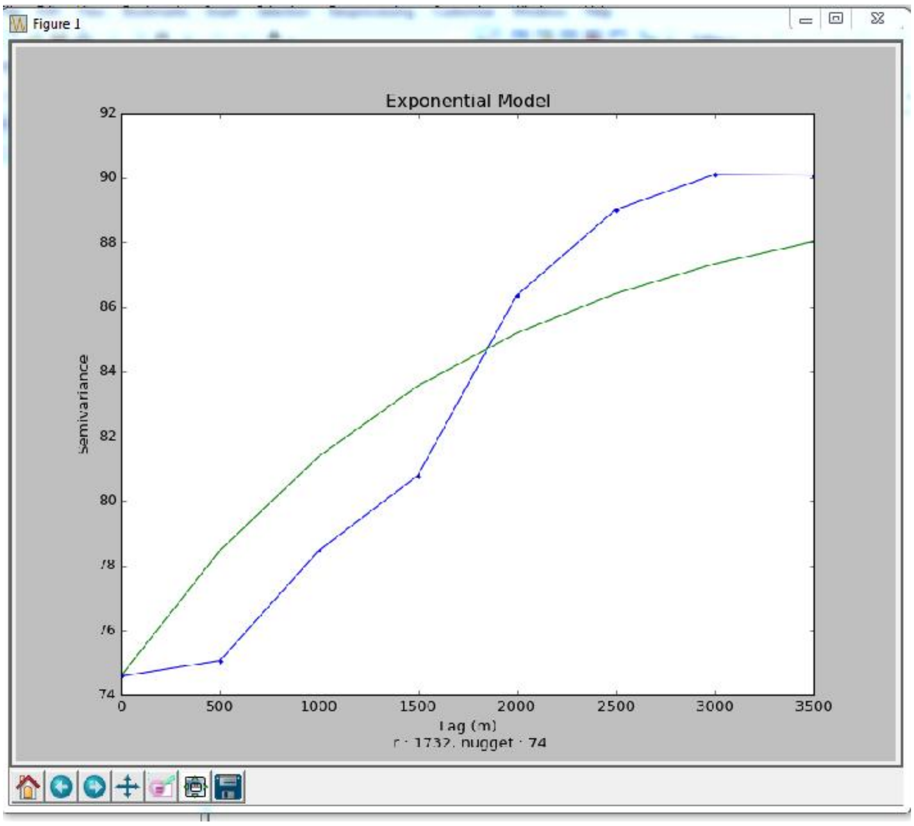
Οι τιμές που παρουσιάζονται εντός του βαριογράμματος είναι τα ζεύγη των σημείων που λαμβάνουν μέρος στον υπολογισμό της κάθε τιμής, του πειραματικού ημιβαριογράμματος.

Επιλέχτηκε το ημιβαριόγραμμα με τιμές παραμέτρων 3500 μέτρα για το εύρος και 500 μέτρα για την τιμή του χωρικού βήματος (εικόνα 3.22).

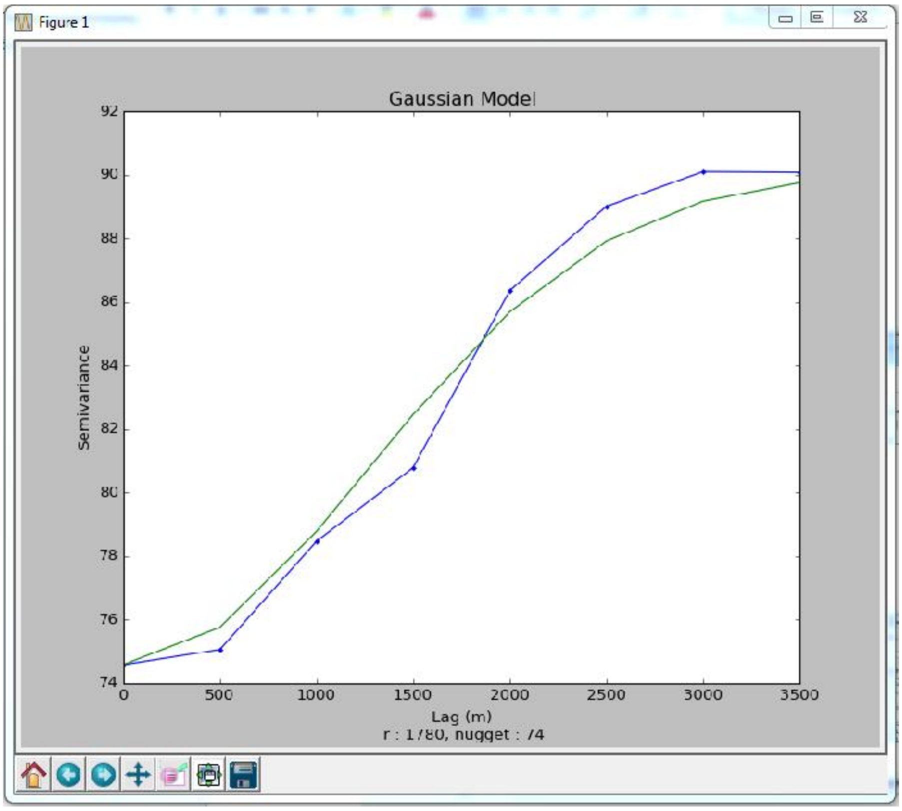
Για την τελική φάση της επεξεργασίας των δεδομένων χρησιμοποιήθηκε το εργαλείο ModelSelectionCriteria για να υπολογιστούν τα θεωρητικά ημιβαριογράμματα και να μας δοθεί μια έγκυρη εικόνα για το πιο έχει την καλύτερη προσαρμογή στο πειραματικό ημιβαριόγραμμα, αλλά και με ποιες παραμέτρους. Αυτό γίνεται τόσο οπτικά με την εμφάνιση των μοντελοποιημένων ημιβαριογραμμάτων όσο και με τη χρήση των δεικτών AIC, AICc, BIC, HQC που αναλύθηκαν σε προηγούμενο κεφάλαιο.



Εικόνα 3.24: Το ημιβαριόγραμμα του σφαιρικού μοντέλου



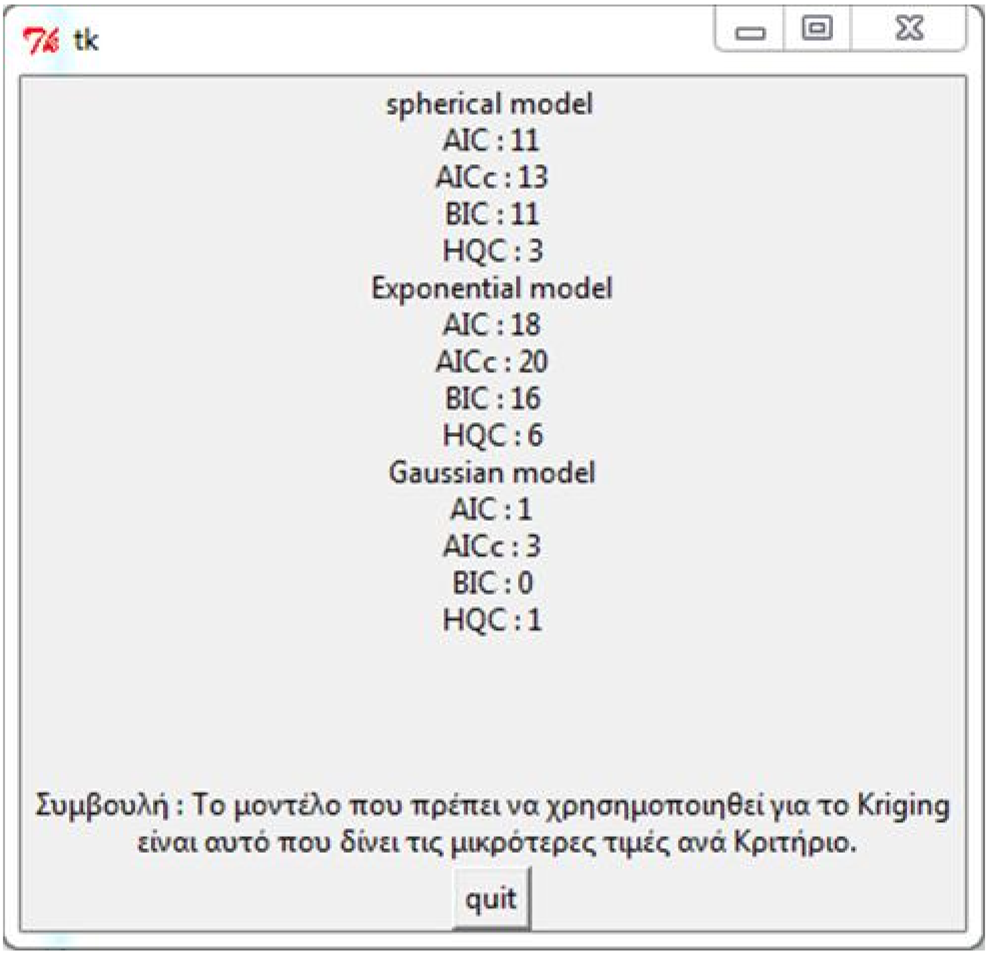
Εικόνα 3.25: Το ημιβαριόγραμμα του εκθετικού μοντέλου



Εικόνα 3.26: Το ημιβαριόγραμμα του κανονικού (Gaussian) μοντέλου

Από τα διαγράμματα (Εικόνες 3.24, 3.25, 3.26) έχουμε μια πρώτη εικόνα αφού φαίνεται καθαρά ότι το κανονικό (gaussian) μοντέλο έχει την καλύτερη προσαρμογή στο πειραματικό ημιβαριόγραμμα με τιμές στις παραμέτρους, του εύρους (range) 1780 μέτρα και διακύμανση επι της αρχής (nugget) 74 μέτρα.

Αποτέλεσμα το οποίο συμφωνεί κι με τις τιμές των δεικτών των κριτηρίων επιλογής μοντέλου, αφού όλες οι τιμές των δεικτών για το κανονικό (gaussian) μοντέλο είναι οι μικρότερες όπως φαίνεται στην παρακάτω εικόνα (Εικόνα 3.27) που θα εμφανισθεί μετα το πέρας τις επεξεργασίας των δεδομένων από το εργαλείο.

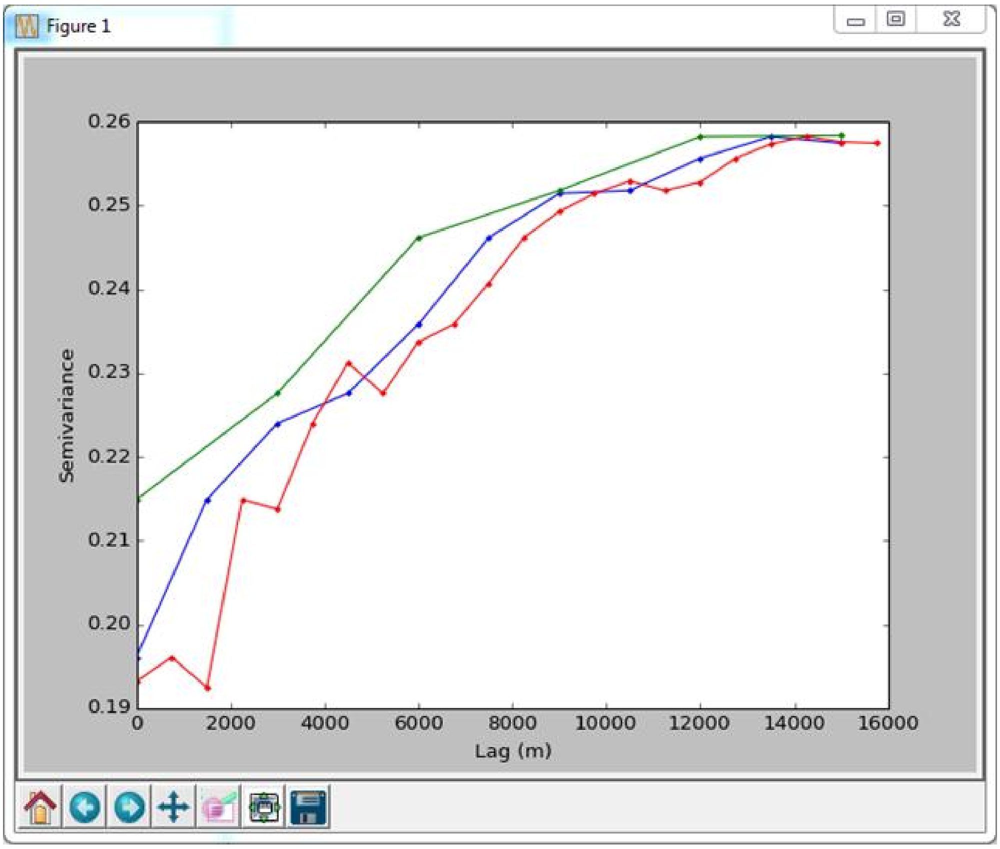


Εικόνα 3.27: Αναδυόμενο παράθυρο με τις τιμές των δεικτών ανά μοντέλο

Έτσι αν προχωρήσουμε σε περεταίρω επεξεργασία γεωστατιστικής χωρικής παρεμβολής (Kriging) είμαστε σίγουροι ότι θα πρέπει να χρησιμοποιήσουμε το κανονικό (Gaussian) μοντέλο, και τις τιμές 1780 μέτρα για την τιμή του εύρους και 74 μέτρα για την τιμή της ημιδιακύμανσης επί της αρχής.

3.2.2 Δεύτερη Εφαρμογή

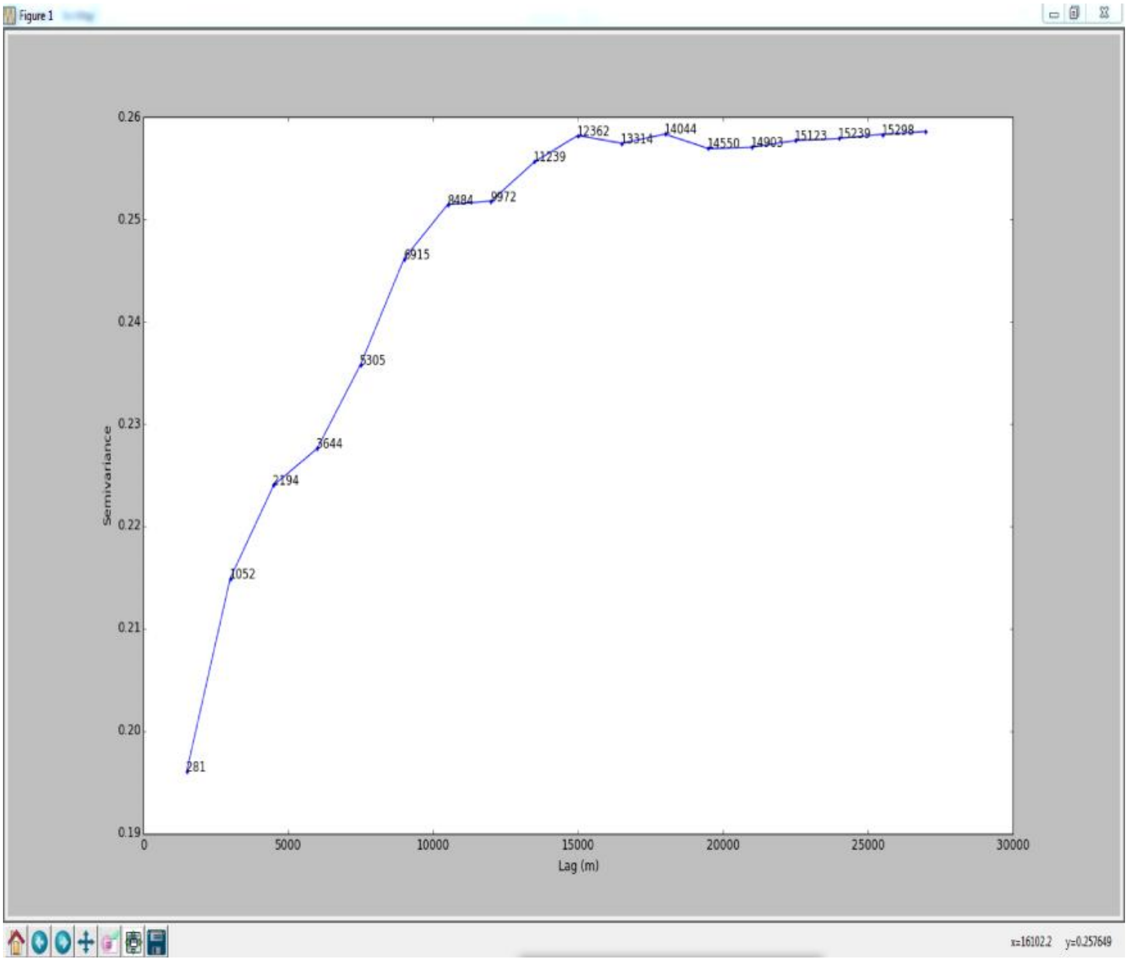
Στη δεύτερη εφαρμογή έγινε χρήση όλων των εργαλείων που αναπτύχθηκαν στην παρούσα εργασία και ο τρόπος που επεξεργάστηκαν τα δεδομένα είναι όμοιος με αυτόν που εργαστήκαμε στην πρώτη εφαρμογή. Τα δεδομένα πάρθηκαν την περιοχή της Άρτας που βρίσκεται στον νομό Άρτας. Είναι δεδομένα που περιγράφουν την κοκκομετρκή σύσταση του εδάφους. Από αυτά τα δεδομένα επιλέχθηκαν για επεξεργασία οι τιμές του εδαφικού pH.



Εικόνα 3.28: Τα τρία πειραματικά ημιβαριογράμματα

Στην αρχή έγινε χρήση του εργαλείου Semivariogram Comparison με τιμή εύρους 15000 μέτρα και τιμές του χωρικού βήματος τα 700, 1500 και 3000 μέτρα από το διάγραμμα (εικόνα 3.28) έγινε αντιληπτό ότι η τιμή του χωρικού βήματος πρέπει να είναι περίπου στα 1500 μέτρα.

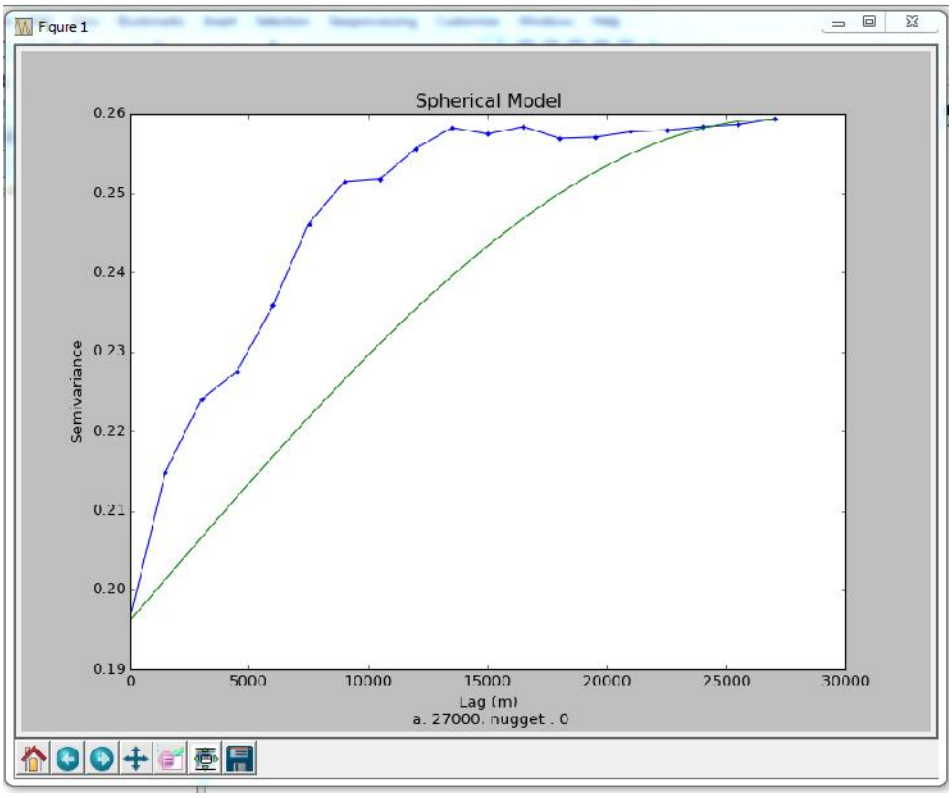
Στη συνέχεια αφού έγιναν δοκιμές στις τιμές εύρους και ισαποχής, με τη βοήθεια του εργαλείου Single Semivariogram. Κατάληξη των δοκιμών ήταν το πειραματικό ημιβαριόγραμμα της Εικόνας 3.29, με τιμή του χωρικού βήματος τα 1500 μέτρα και εύρος αύξησης της ημιδιακύμανσης ορίσαμε τα 27000 μέτρα.



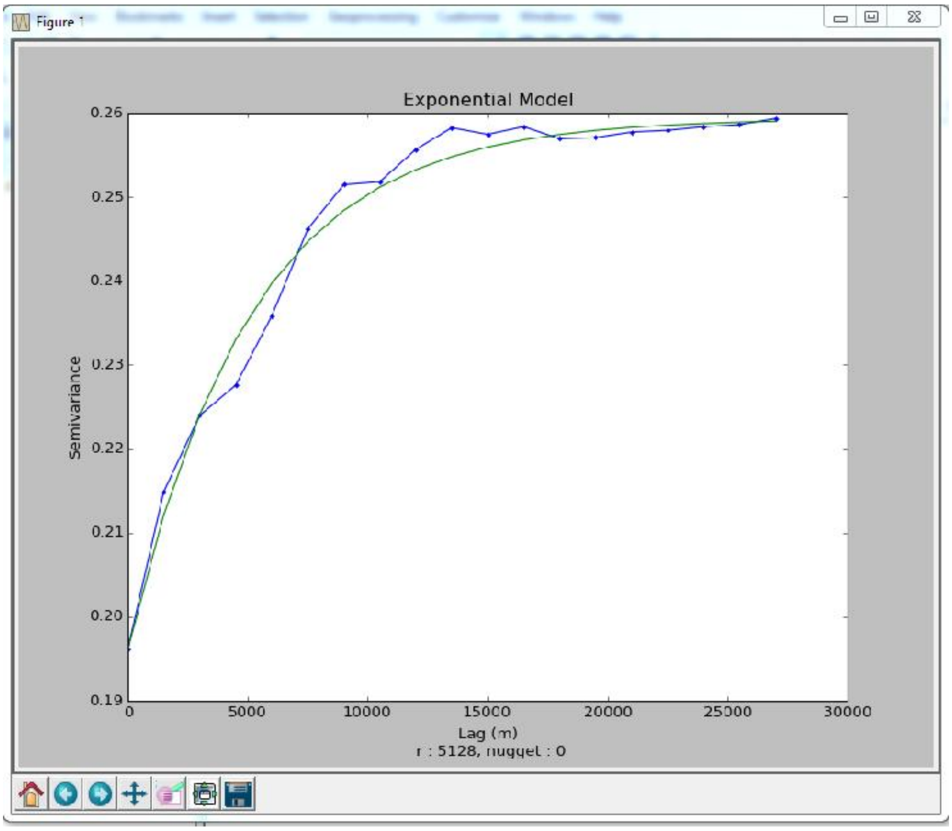
Εικόνα 3.29: Το πειραματικό ημιβαριόγραμμα για το pH

Τέλος χρησιμοποιήσαμε το εργαλείο Model Selection Criteria όπου δώσαμε σαν τιμές παραμέτρων τις τιμές που ορίσαμε στο πειραματικό ημιβαριόγραμμα με την χρήση του εργαλείου Single Semivariogram, δηλαδή 1500 μέτρα για την τιμή του χωρικού βήματος και 27000 το εύρος αύξησης της διακύμανσης.

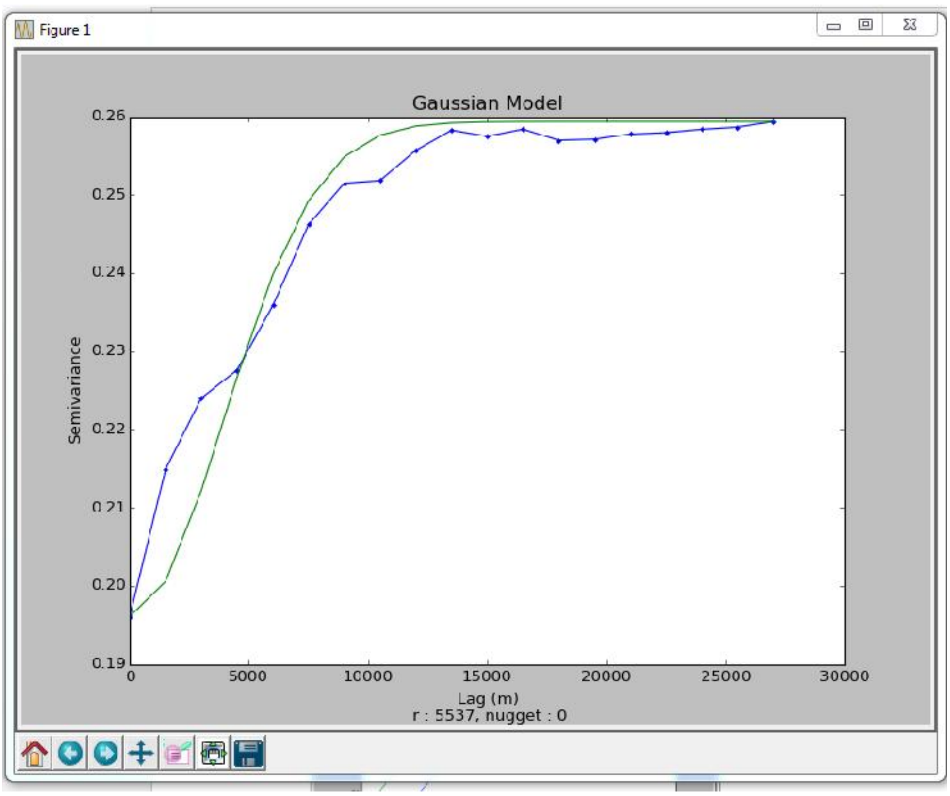
Το αποτέλεσμα ήταν η εμφάνιση των παρακάτω διαγραμμάτων που περιέχουν το πειραματικό ημιβαριόγραμμα και τα θεωρητικά ημιβαριογράμματα (Εικόνες 3.30, 3.31, 3.32)



Εικόνα 3.30: Το ημιβαριόγραμμα του σφαιρικού μοντέλου



Εικόνα 3.31: Το ημιβαριόγραμμα του εκθετικού μοντέλου



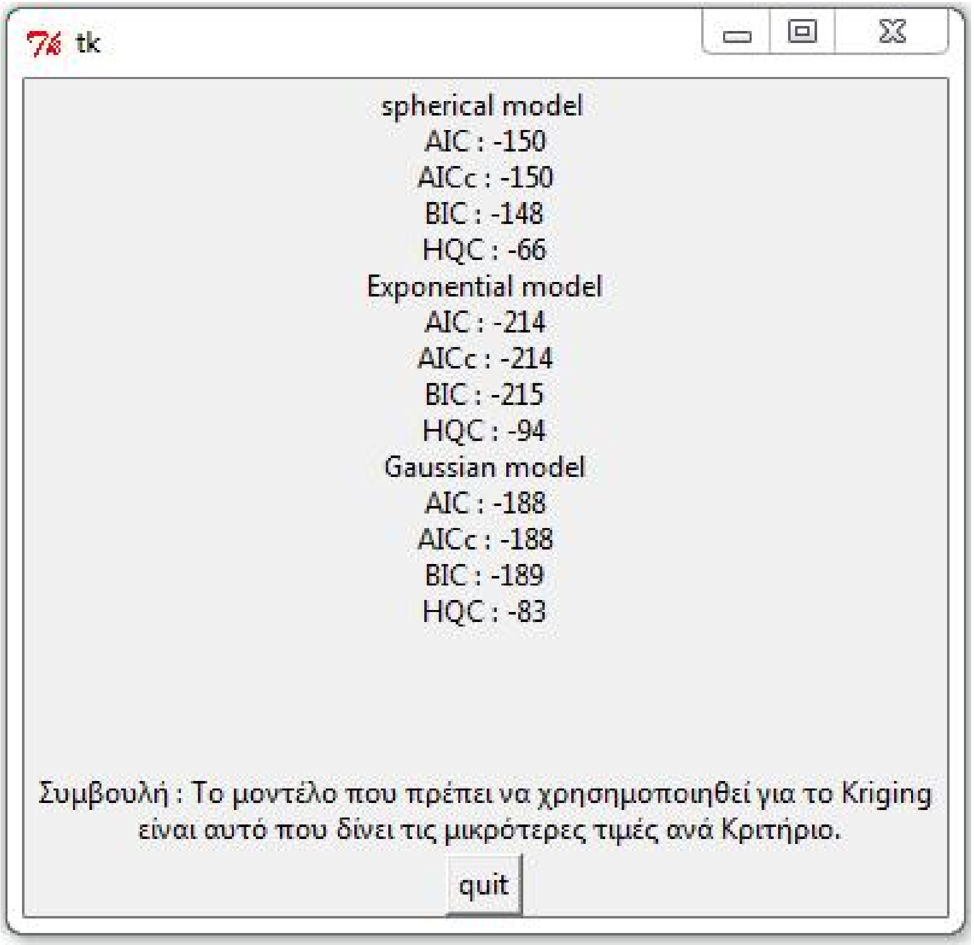
Εικόνα 3.32: Το ημιβαριόγραμμα του κανονικού (Gaussian) μοντέλου

Από τα διαγράμματα των ημιβαριογραμμάτων παρατηρούμαι ότι δεν μπορούμε να είμαστε σίγουροι αν η επιλογή του εκθετικού μοντέλου είναι η βέλτιστη επιλογή και αυτό γιατί οπτικά θα μπορούσε να ήταν μια πιθανή επιλογή και το κανονικό μοντέλο.

Όμως με την εμφάνιση των δεικτών μπορούμε να είμαστε σίγουροι ότι το μοντέλο που προσαρμόζεται καλύτερα στις τιμές του πειραματικού ημιβαριογράμματος είναι το εκθετικό αφού όλα τα κριτήρια επιλογής του κατάλληλου μοντέλου συμφωνούν. Καθώς οι τιμές τους είναι οι μικρότερες με αρκετή διαφορά.

Οι τιμές των παραμέτρων του εκθετικού μοντέλου είναι 5128 μέτρα και 0.19 η τιμή της ημιδιακύμανσης επί της αρχής

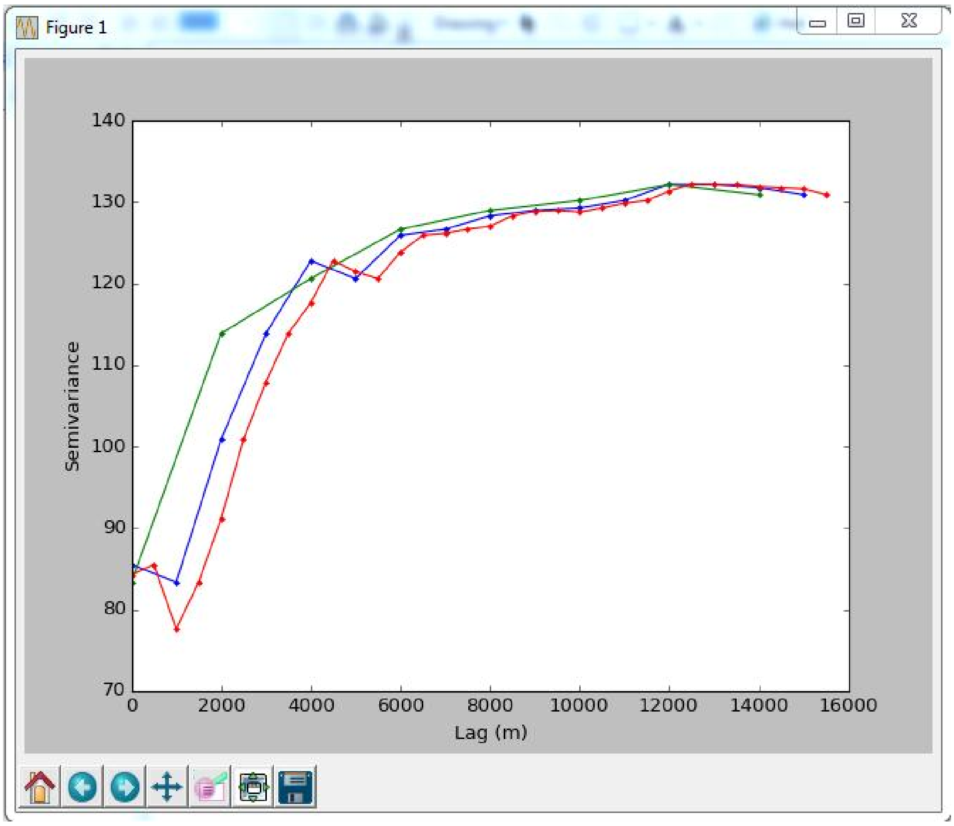
Οι αρνητικές τιμές στους δείκτες των κριτηρίων, έχουν σχέση με την φύση των τιμών του pH καθώς είναι πολύ μικρές.



Εικόνα 3.33: Αναδυόμενο παράθυρο με τις τιμές των δεικτών ανά μοντέλο

3.2.3 Τρίτη Εφαρμογή

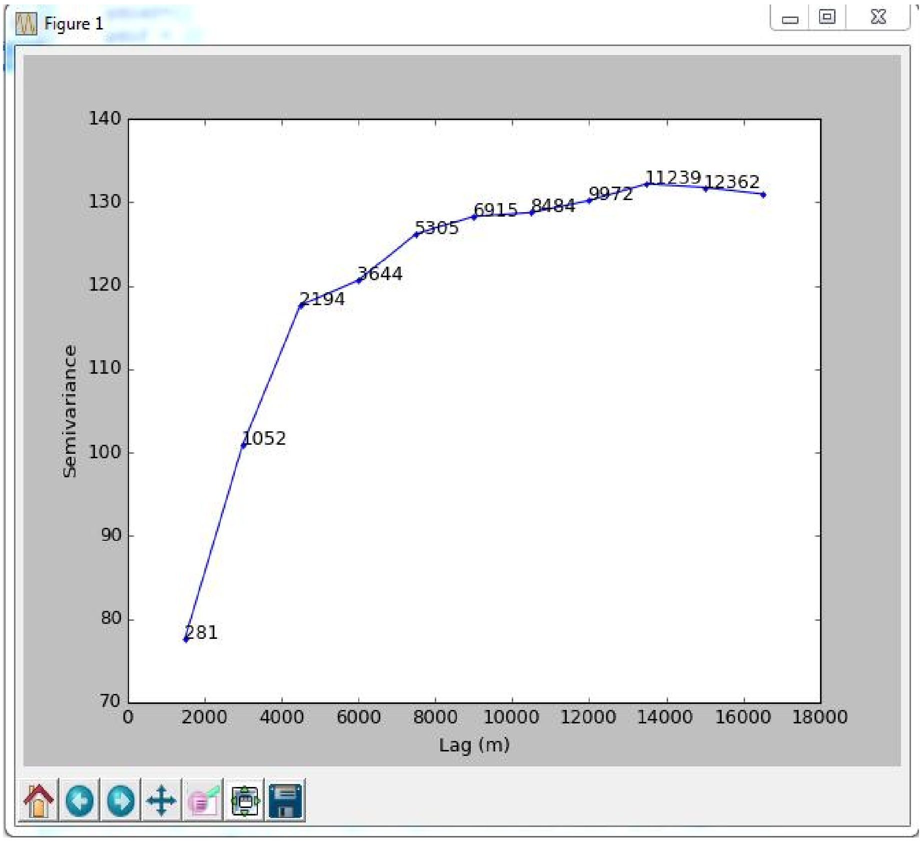
Η τρίτη και τελευταία εφαρμογή για την εργασία αύτη επιλέχτηκε να είναι άλλη μια επανάληψη των δύο προηγουμένων εφαρμογών, με δεδομένα από την περιοχής της Άρτας. Έγινε επεξεργασία των τιμών της περιεχόμενης στο έδαφος αργίλου.



Εικόνα 3.34: Τα τρία πειραματικά ημιβαριογράμματα

Στην αρχή έγινε χρήση του εργαλείου SemivariogramComparison με τιμή εύρους 15000 μέτρα και τιμές του χωρικού βήματος τα 500, 1000 και 2000 μέτρα από το διάγραμμα (εικόνα 3.34) έγινε αντιληπτό ότι η τιμή του χωρικού βήματος πρέπει να είναι περίπου στα 1500 μέτρα.

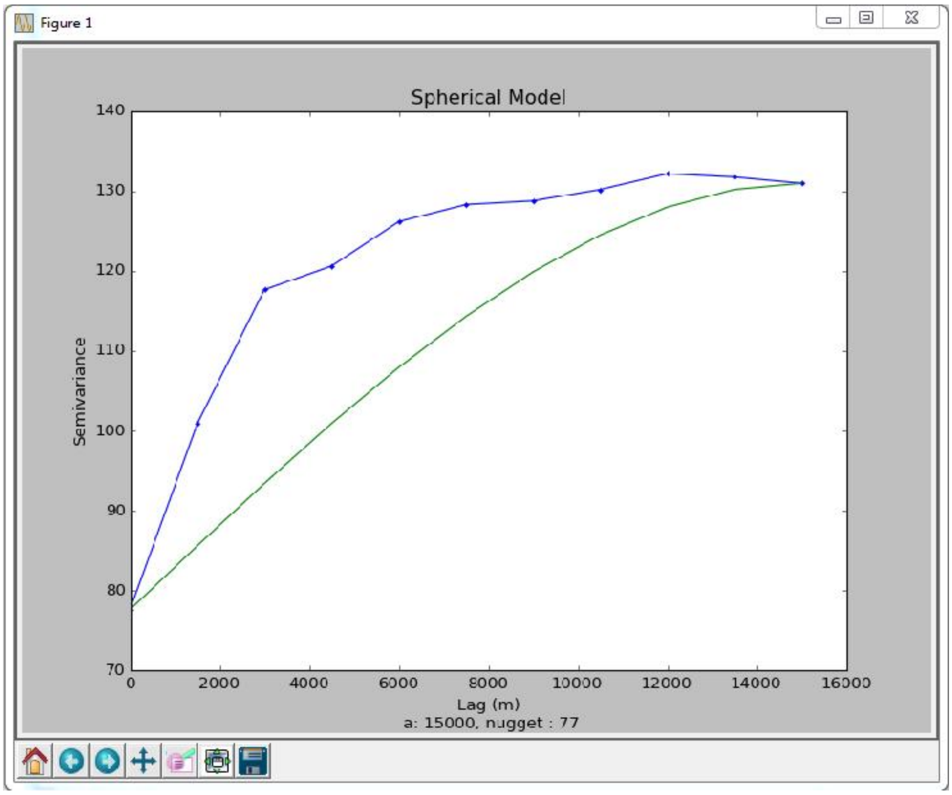
Στη συνέχεια αφού έγιναν δοκιμές στις τιμές εύρους και ισαποχής, με τη βοήθεια του εργαλείου Single Semivariogram. Κατάληξη των δοκιμών ήταν το πειραματικό ημιβαριόγραμμα της Εικόνας 3.35, με τιμή του χωρικού βήματος τα 1500 μέτρα και εύρος αύξησης της ημιδιακύμανσης ορίσαμε τα 15000 μέτρα.



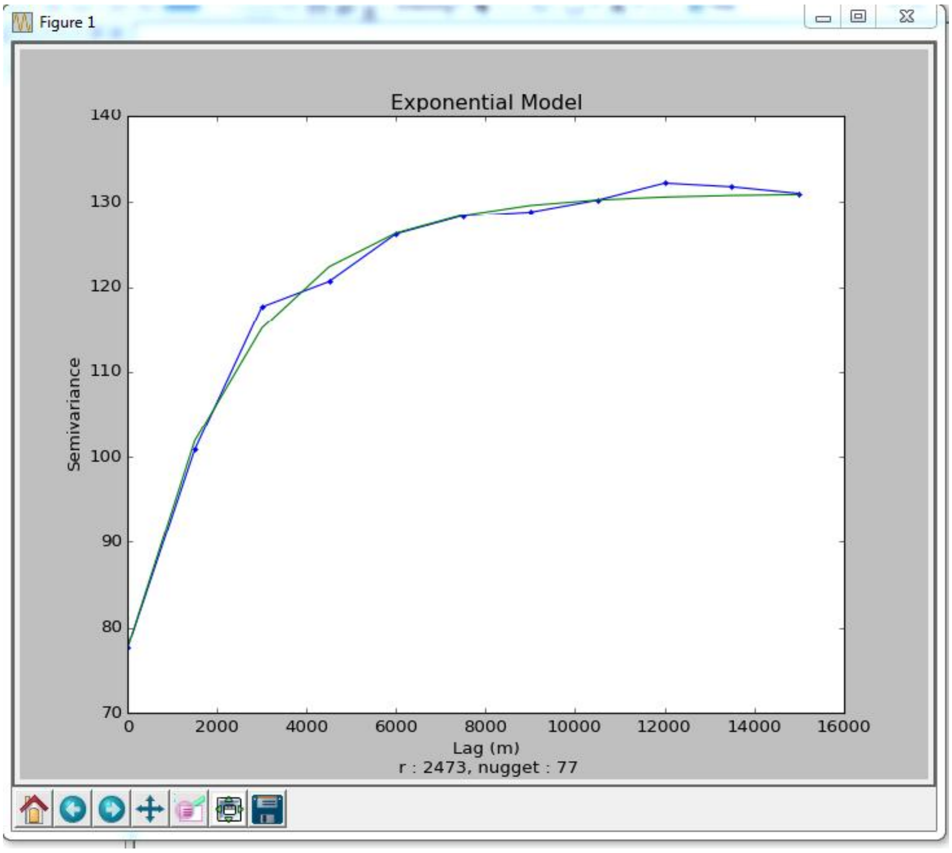
Εικόνα 3.35: Το πειραματικό ημιβαριόγραμμα για τις τιμές της περιεχόμενης στο έδαφος αργίλου

Τέλος χρησιμοποιήσαμε το εργαλείο Model Selection Criteria όπου δώσαμε σαν τιμές παραμέτρων τις τιμές που ορίσαμε στο πειραματικό ημιβαριόγραμμα με την χρήση του εργαλείου Single Semivariogram, δηλαδή 1500 μέτρα για την τιμή του χωρικού βήματος και 15000 το εύρος αύξησης της διακύμανσης.

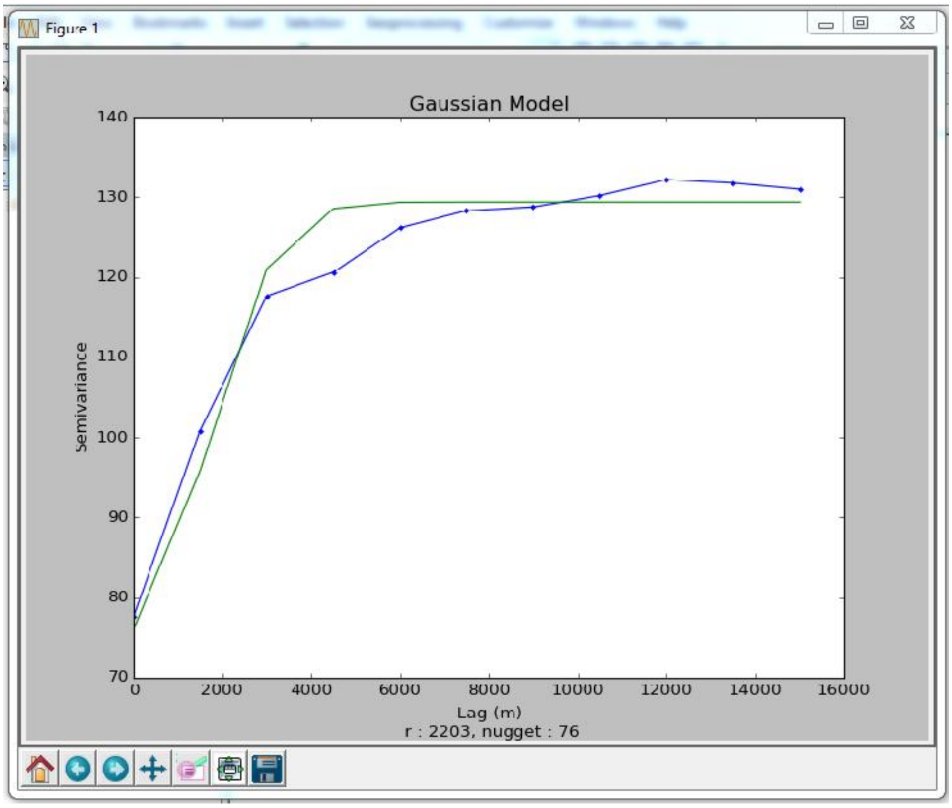
Το αποτέλεσμα ήταν η εμφάνιση των παρακάτω διαγραμμάτων που περιέχουν το πειραματικό ημιβαριόγραμμα και τα θεωρητικά ημιβαριογράμματα (Εικόνες 3.36, 3. 37, 3.38)



Εικόνα 3.36: Το ημιβαριόγραμμα του σφαιρικού μοντέλου



Εικόνα 3.37: Το ημιβαριόγραμμα του εκθετικού μοντέλου

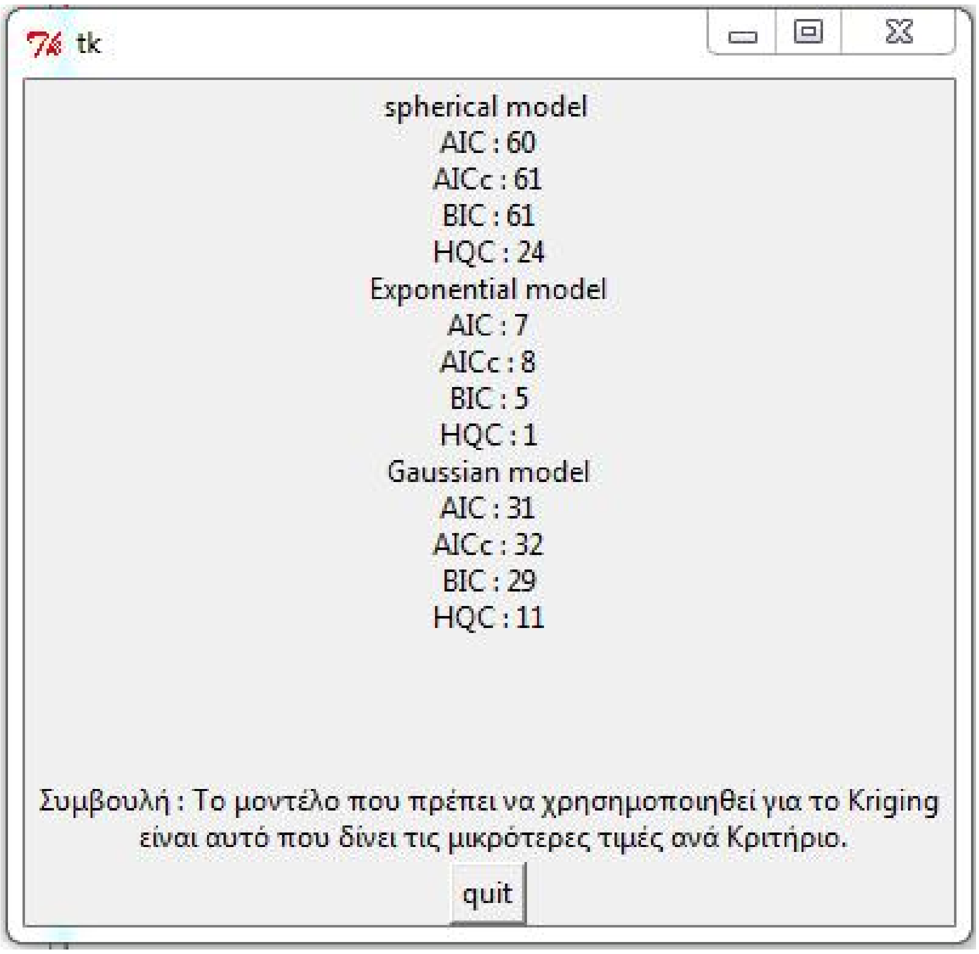


Εικόνα 3.38: Το ημιβαριόγραμμα του κανονικού (Gaussian) μοντέλο

Από τα διαγράμματα των ημιβαριογραμμάτων παρατηρούμαι ότι δεν μπορούμε να είμαστε σίγουροι αν η επιλογή του εκθετικού μοντέλου είναι η βέλτιστη επιλογή και αυτό γιατί οπτικά θα μπορούσε να ήταν μια πιθανή επιλογή και το κανονικό μοντέλο.

Όμως με την εμφάνιση των δεικτών μπορούμε να είμαστε σίγουροι ότι το μοντέλο που προσαρμόζεται καλύτερα στις τιμές του πειραματικού ημιβαριογράμματος είναι το εκθετικό αφού όλα τα κριτήρια επιλογής του κατάλληλου μοντέλου συμφωνούν. Καθώς οι τιμές τους είναι οι μικρότερες με αρκετή διαφορά.

Οι τιμές των παραμέτρων του εκθετικού μοντέλου είναι 2473 μέτρα και 77 η τιμή της ημιδιακύμανσης επί της αρχής



Εικόνα 3.39: Αναδυόμενο παράθυρο με τις τιμές των δεικτών ανά μοντέλο

4. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΣΥΖΗΤΗΣΗ

Βασικός σκοπός της εργασίας ήταν η ανάπτυξη των κατάλληλων εργαλείων για να γίνει μια όσο το δυνατόν καλύτερη ανάλυση της ημιδιακύμανσης. Πιο αναλυτικά αναπτύχθηκαν εργαλεία αρχικά για την διερεύνηση του πειραματικού ημιβαριογράμματος, την εύρεση των κατάλληλων τιμών των παραμέτρων αυτού και την εύρεση εκείνου του μοντέλου ημιβαριογράμματος που η χρήση του θα φέρει την μικρότερη απώλεια πληροφορίας από το πειραματικό ημιβαριόγραμμα.

Η επιλογή που έγινε στην αρχή της εργασίας, το λογισμικό να αναπτυχθεί σε περιβάλλον ArcGISήταν απόλυτα συνειδητή καθώς είναι ένα από τα πιο δημοφιλή λογισμικά στο χώρο των ΓΠΣ με χιλιάδες χρήστες ανά τον κόσμο και είναι ένα πρόγραμμα όπου διδάσκεται σε πολλές σχολές που ασχολούνται με χωρικά δεδομένα όπως και στο Γεωπονικό πανεπιστήμιο Αθηνών.

Πολλά δεδομένα υπάρχουν σε μορφή αρχείων που λειτουργούν σε περιβάλλον ArcGIS,επίσης η δυνατότητα, να χρησιμοποιηθεί το εργαλείο σαν συμπλήρωμα σε κάποια ανάλυση χωρικών δεδομένων χωρίς να χρειαστεί να γίνει αλλαγή στο περιβάλλον εργασίας προσφέρει μια ακόμη ευκολία.

Παρόλο που το ArcGISαποτελεί ένα εξαιρετικά εξελιγμένο και ενημερωμένο επιστημονικά λογισμικό, στον τομέα της ανάλυσης του βαριογράμματος εμφανίζει μειονεκτήματα, καθώς δεν παρέχει κάποια τέτοια δυνατότητα. Συγκεκριμένα η επιλογή του μοντέλου και των παραμέτρων του βαριογράμματος που υπάρχει στο εργαλείο Geostatistical Analyst (μέσω του οποίου γίνονται οι επεξεργασίες χωρικής παρεμβολής) του ArcGIS λογισμικού, δεν δίνει την δυνατότητα ελέγχου ή περαιτέρω ανάλυσης για την σωστή επιλογή τόσο του μοντέλου του ημιβαριογράμματος όσο και των τιμών στις παραμέτρους αυτού.

Η επιλογή της γλώσσας προγραμματισμού Python που γράφτηκε το συγκεκριμένο λογισμικό επιλέχτηκε σκόπιμα γιατί αποτελεί μια γλώσσα απλή και ταυτόχρονα απίστευτα ισχυρή προσφέροντας τη δυνατότητα επιλογής μέσα από μια πληθώρα εργαλείων και διαφορετικών τεχνοτροπιών προγραμματισμού προκειμένου να φτάσουμε στο επιθυμητό αποτέλεσμα. Επίσης μας δίνει τις δυνατότητες ανάπτυξης και εξέλιξης του συγκεκριμένου προγράμματος σε όποιο τομέα χρειαστεί να εξελιχθεί.

Ερευνώντας διαπιστώσαμε ότι δεν υπήρχε κάποια προσπάθεια ανάπτυξης παρόμοιου λογισμικού, που να δίνει την δυνατότητα να γίνει η ανάλυση του βαριογράμματος με τρόπο πλήρως ελεγχόμενο.

Εργαλεία βαριογραφίας υπήρχαν όπως το εργαλείο για Simple Kriging του Connor Johnson που είχε κάποια στοιχειώδη ανάλυση βαριογράμματος, αλλά δεν έδινε την δυνατότητα σύγκρισης διαφορετικών μοντέλων, είτε εμπειρικά είτε με την βοήθεια κάποιου κρατητηρίου.

Άλλη μια πολύ καλή προσπάθεια ήταν αυτή των L.Pesquer κ.α.(2011) που εφάρμοσαν Krigingμε τη χρήση μάλιστα παράλληλου προγραμματισμού σε γλώσσα C. Στη συγκεκριμένη προσπάθεια όμως δεν μπορούσε να γίνει κάποιου είδους βαριογραφίας καθώς όλα ήταν αυτοματοποιημένα.

Τέλος να αναφέρουμε ότι τα εργαλεία για βαριογραφική ανάλυση που έχουν αναπτυχθεί σε γλώσσα προγραμματισμού Rκαι που προσφέρουν σύγκριση μεταξύ μοντέλων μέσω κριτηρίων είναι αρκετά δύσκολα στη χρήση που απαιτούν γνώσεις προγραμματισμού από τον εκάστοτε χρήστη. Το ίδιο φυσικά ισχύει και για βιβλιοθήκες που υπάρχουν διαθέσιμες στο περιβάλλον-γλώσσα προγραμματισμού Matlab.

Κατά τη διάρκεια της εργασίας αναπτύχθηκαν 3 εργαλεία.

Το πρώτο εργαλείο ως σκοπό είχε να μας δώσει μια εικόνα των πιθανών τιμών του πειραματικού ημιβαριογράμματος.

Το δεύτερο εργαλείο φτιάχτηκε με σκοπό να προσφέρει τη δυνατότητα μιας λεπτομερούς ανάλυσης του πειραματικού ημιβαριογράμματος με δοκιμές στις τιμές των παραμέτρων αυτού.

Το τρίτο εργαλείο δημιουργήθηκε με σκοπό να δίνει εικόνα του κατά πόσο ένα μοντέλο ημιβαριογράμματος προσεγγίζει το πειραματικό ημιβαριόγραμμα, αυτό γίνεται με την βοήθεια διαγραμμάτων αλλά και με τη χρήση δεικτών, των κυριότερων κριτηρίων επιλογής μοντέλου.

Κατά τη διάρκεια των δοκιμών ζητήσαμε να χρησιμοποιηθεί το λογισμικό από χρήστες ArcGIS, ώστε να διαπιστωθεί και να αμβλυνθεί η ευχρηστία του λογισμικού, σε κάθε περίπτωση διαπιστώθηκε ότι η βαριογραφική ανάλυση μπορεί να γίνει εύκολα από κάποιο χρήστη ΓΠΣ.

Το λογισμικό χρησιμοποιεί τα τρία κυριότερα μοντέλα προσέγγισης των πειραματικών ημιβαριογραμμάτων που από την βιβλιογραφία είναι γνωστό ότι καλύπτουν το μεγαλύτερο μέρος των περιπτώσεων της ημιδιακύμανσης των εδαφικών ιδιοτήτων .

Οι δείκτες για την επιλογή μοντέλου Akaike (AIC)καιBayesian(BIC)είναι πολύ δημοφιλείς σε πολλούς κλάδους επιστημών που εφαρμόζεται στατιστική ανάλυση και προσαρμογή μοντέλων (π.χ. οικονομετρία, χωρική οικονομετρία), και έχουν μελετηθεί και χρησιμοποιηθεί επιτυχώς για πολλά χρόνια. Το Κριτήριο Akaikeχρησιμοποιείται και από λογισμικό τουR περιβάλλοντος,που αναπτύχθηκε για εφαρμογές σε χωρικά δεδομένα και συγκεκριμένα στην επιλογή του μοντέλου του βαριογράμματος. Το κριτήριο Hannan_Quinn (HQC) αν και δεν είναι τόσο δημοφιλές χρησιμοποιήθηκε για να δοκιμαστεί καθώς συνδυάζει τις αρχές τόσο του κριτηρίου Bayesianόσο και του κριτηρίου Akaike.

Τέλος το λογισμικό δοκιμάστηκε επιτυχώς σε πραγματικά δεδομένα ώστε να διαπιστωθεί τόσο η εγκυρότητα των αποτελεσμάτων όσο και αποτελεσματικότητα του λογισμικού σε πραγματικές συνθήκες ανάλυσης δεδομένων. Κάποιες από αυτές τις περιπτώσεις τις παρουσιάσαμε σε αυτή την εργασία.

Η ανάπτυξη του λογισμικού αυτού πρέπει να εξελιχτεί τόσο στον τομέα της ανάπτυξης λογισμικού υιοθετώντας τεχνικές παράλληλης επεξεργασίας και προσέγγισης του προγράμματος περισσότερο κοντά στη γλώσσα μηχανής (lowlevel).

Τέλος όσον αφορά την στατιστική ανάλυση πρέπει να διευρυνθεί το πεδίο εφαρμογών με τη χρήση περισσότερων βαριογραφικών μοντέλων. Καθώς και με ανάλυση υπολογισμού ημίβαριογραμματων σε διαφορετικές διευθύνσεις (μελέτη ανισοτροπίας).

5. ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξένη Βιβλιογραφία

1. Agterberg, F P, *Geomathematics, Mathematical Background and Geo-Science Applications*, Elsevier Scientific Publishing Company, Amsterdam, 1974
2. Akaike, H., 1969. Fitting autoregressive models for prediction. Annals of the Institute of Statistical Mathematics. 21: 243-247
3. Anselin, L. and A. Getis (1992). Spatial statistical analysis and geographic informationsystems. The Annals of Regional Science 26, 19-33.
4. Anselin, L. and S. Hudak (1992). Spatial econometrics in practice: a review of softwareoptions. Regional Science and Urban Economics 22, 509-536.
5. Atkinson, D. E., Gajewski K., 2002: High-resolution estimation of summer surface airtemperature in the Canadian Arctic Arcipelago. *J. Clim. 15*, 3601-3614.
6. Aznar Grasa, A. (1989). *Econometric Model Selection: A New Approach*, Springer.
7. Bao, S., L. Anselin, D. Martin and D. Stralberg (2000). Seamless integration of spatialstatistics and GIS: the S-Plus for ArcView and the S+Grassland links. Journal ofGeographical Systems 2, 287-306.
8. Bao, S. and D. Martin (1997). User’s Reference for the S+ArcView Link. Seattle, WA,MathsoftInc.
9. Becket P.H.T. and R. Webster, 1971. Soil variability. Solis and Feritilaizer 34:1-15
10. Bhat, H. S. and Kumar, N., "On the derivation of the Bayesian Information Criterion".
11. Bohling, G.C., 2005, Chronos Age-Depth Plot: A Java application for stratigraphic data analysis, Geosphere, vol. 1, no. 2,
12. Burnham, K.P. and Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer-Verlag..
13. Burrough, P.A.,McDonnell,R.A.,1998.PrinciplesofGeographicalInformation Systems. OxfordUniversityPress,333pp.

14. Cambardella A.C., T. B. Moorman, T. B. Parkin, D. L. Karlen, J. M. Novak, R. F. Turco and A. E. Konopka, 1994. Field-Scale Variability of Soil Properties in CentralIowa Soils. Soil Science Soc. ASm. J. Vol. 58 No. 5, 1501-1511
15. Chan, T. 0., and Williamson, I. P., 1996. A model of the decision process for GIS adoption and diffusion in a government environment. In The URISA Proceedings, {Salt Lake City, Utah, 26th July -1st August: URISA) pp. 247-260.
16. Chiles, J.-P.,Delfiner,P.,1999.Geostatistics:ModellingSpatialUncertainty.Wiley, New York,687pp.
17. Chrisman, N.R. 1997. Exploring Geographic Information Systems. John Wiley and 5ons.
18. Christakos G, Bogaert P, Serre ML (2002) Temporal GIS: advanced functions for fieldbasedapplications. Springer, Berlin, Heidelberg and New York
19. Claeskens, G. and Hjort, N.L.(2008). *Model Selection and Model Averaging*, Cambridge.
20. Cooper, W., Jarvis,C., 2004.AJava-baseintelligentadvisorfor selecting a context-appropriate spatialinterpolationalgorithm.Computers&GeoSciences30, 1003–1018.
21. Cowen, O. J., 1988. GIS versus CAD versus DB999MS: what are the differences? Photogram- metric Engineering and Remote Sensing 54, 1551 4.
22. Cressie, N.A.C.,1993.StatisticsforSpatialData,WileySeriesinProbabilityand Mathematical Statistics.JohmWiley&Sons,NewYork,900pp.
23. Cressie, N. A. C., *The Origins of Kriging, Mathematical Geology*, v. 22, pp 239–252, 1990
24. Christopoulos Dionissios T., (2003), Spartan Gibbs random fields models for geostatistical applications, Society for industrial and applied Mathematics
25. Dangermond, J., 1988. Introduction and overview of GIS. In Geographic Information Systems Seminar, Data Sharing - Myth or Reality, (Ontario, Canada. 3rd-5th October: Ministry of Natural Resources)
26. Dangermond, J.,1986**.** CAD vs GIS. Computer Graphics World, Vol.**9,** No10**,**p.73-74**.**

27. Ding, Y. and S. Fotheringham (1992). The integration of spatial analysis and GIS.Computers, Environment and Urban Systems 16, 3-19.
28. EnglundE, Sparks A (1988) Geo-EAS 1.2.1 user's guide. EPA Report 60018-91/008. EPAEMSL,Las Vegas [NV]
29. Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.
30. Franke, R., 1982. Scattered data interpolation: tests of some methods. Mathematics of Computation 38, 181–199.
31. Goodchild, M. (1987). A spatial analytical perspective on geographical informationsystems. International Journal of Geographical Information Systems 1, 31-45.
32. *GoovaertsPierre*Geostatistical Software 2009
33. Haining, R. (1989). Geography and spatial statistics: current positions, futuredevelopments. In B. Macmillan (ed.), Remodeling Geography, pp. 191-203.Oxford, Basil Blackwell.
34. Hannan, E. J., and B. G. Quinn (1979) The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society*, B, 41, 190–195.
35. Haslett, J., R. Bradley, P. Craig, A. Unwin and G. Wills (1991). Dynamic graphics forexploring spatial data with applications to locating global and local anomalies.The American Statistician 45, 234-242.
36. Haslett, J., G. Wills and A. Unwin (1990). SPIDER – An interactive statistical tool forthe analysis of spatially distributed data. International Journal of GeographicalInformation Systems 4, 285-296.
37. Isaaks, E.H., and Srivastava, R.M., 1989: Applied Geostatistics. New York: OxfordUniv. Press.
38. Journel, A.G., and Huijbregts, C.J., 1978: Mining Geostatistics. London: AcademicPress.
39. Kitanidis, K.P., 1997: Introduction to Geostatistics: Applications in Hydrogeology,Cambridge University Press, 249 pages.
40. Krige, D.G, *A statistical approach to some mine valuations and allied problems at the Witwatersrand*, Master's thesis of the University of Witwatersrand, 1951

41. Krūminiene I., 2006: Analysis of anisotropic variogram models for prediction of theCuronian lagoon data, Mathematical Modelling and Analysis, 11:1, 73-86.
42. LeMay, E.N., 1998: Variogram Modeling and Estimation. Master Thesis. University of Colorado, Denver.
43. Legendre, P., Fortin, M.-J., 1989. Spatial pattern and ecological analysis.Vegetatio 80, 107–138.
44. Liddle, A. R., "Information criteria for astrophysical model selection",
45. Link, R F and Koch, G S, *Experimental Designs and Trend-Surface Analsysis, Geostatistics,* A colloquium, Plenum Press, New York, 1970
46. Longley, P.A., Goodchild, M.F., Maguire, D.J. and D. W. Rhind, 2001. Geographic Information Systems and Science. Wiley, 454 p.
47. Luc Anselin2011 From SpaceStat to CyberGIS:Twenty Years of Spatial Data AnalysisSoftware Working Paper Number 06
48. Matheron, G., "The intrinsic random functions, and their applications", *Adv. Appl. Prob*., 5, pp 439–468, 1973
49. Matheron, G., "Principles of geostatistics", *Economic Geology*, 58, pp 1246–1266, 1963
50. Matheron, G., 1962. Traite´ de Ge´ ostatistique Applique´ e. Editions Technip, Paris, 334 pp.
51. McQuarrie, A. D. R., and Tsai, C.-L., 1998. *Regression and Time Series Model Selection*. World Scientific.
52. Merriam, D F, Editor, *Geostatistics*, a colloquium, Plenum Press, New York, 1970
53. Meyers, D.E. (Ed.), 1994. Spatial Interpolation: An Overview. Geoderma (62), pp. 17–231 (special issue).
54. Oliver, M.A. and Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. Int. J. Geographical Information Systems, 4(3), pp.313-332.
55. Oliver, M.A., Webster, R., Gerrard, J., 1989. Geostatistics in physical geography. PartI: theory. The Royal Geographical Society, Vol.14, No 3, pp. 259-269.
56. Oliver, M.A., 1984. Soil Variation in the Wyre Forest: its Elucidation and Measurement. PhD Thesis, University of Birmingham.

57. Openshaw, S., 1987. Some applications of supercomputers in urban and regional analysis and modelling. Environment & Planning A 19, 853–860.
58. Unwin, A. (1994). REGARDing geographic data. In P. Dirschedl and R. Osterman (eds.)Computational Statistics, pp. 345-354. Heidelberg, PhysicaVerlag.
59. PesquerLluı, AnaCorte, XavierPons Parallel ordinary kriging interpolation in corporating automatic variogram fitting Lluı´sPesquer, AnaCorte´ s b, XavierPons 2011
60. Peuquet, D.J. and Marble, D.F. (eds.) 1990. Introductory Readings in Geographic Information Systems. London, Taylor and Francis.
61. Schwarz, Gideon E. (1978). "Estimating the dimension of a model". Annals of Statistics 6 (2): 461–464. doi:10.1214/aos/1176344136. MR 468014.
62. Trangram B.B. R.S. Yost, ang G. Uehara, 1985. Application of geostatistics to spatial studies of soil properties. Advances in Agronomy.
63. Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography,* 46(2): 234-240.
64. Tomlin, D. C., 1990. Geographic Information Systems and Cartographic Modelling. Englewood Cliffs, NJ, Prenticc-Hall.
65. Webster R., 1985. Quantitave spatial analysis of soil in the field. Advances in soil Scienxe, Vol.3:1-70.
66. Webster R.andCuanalo de la C., H.E. 1975. Soil transectcorrelograms of North Oxfordshire and their interpretation. Journal of Soil Science 35:127-140.
67. Webster R., 1973 Automatic soil-boundary location from transect data Math. Geol. 5:27-37
68. Wiley, Chichester. Lloyd, C. D. (2007). Local Models for Spatial Analysis . CRC, Boca Raton.
69. Woodcock C E, Strahler A H, Jupp D L B, et al. The use ofvariograms in remote sensing: I. Scene models and simulatedimages. Remote Sensing of Environment, 1988.

Ελληνική Βιβλιογραφία

1. Δήμου, Π., 2010. Εφαρμογή γεωστατιστικών μεθόδων στην εκτίμηση εδαφικών ιδιοτήτων γεωργικών εδαφών με τη χρήση γεωγραφικών πληροφοριακών συστημάτων. Μεταπτυχιακή διατριβή. Γεωπονικό Πανεπιστήμιο Αθήνας.
2. Καλυβας Δ. 1989., Μελετη Μεταβλητότητας Εδαφικών ιδιοτήτων με τη βοήθεια της γεωστατιστικής στην περιοχή Καρδίτσας. Μεταπτυχιακή διατριβή, Αθήνα.
3. Κολλια Β., Καλυβας Δ., Τραιντακωνσταντης Δ. 2012, Γεωγραφικά Πληροφοριακά συστήματα Αθηνα
4. Κουτσόπουλος Κ. 2005 Γεωγραφικά συστήματα πληροφοριών και ανάλυσης χωρου εκδόσεις ΠαπασωτηρίουΑθηνα
5. ΣτεφανακηςΕ.Βασεις Γεωγραφικών Δεδομένων & Συστήματα ΦεωγραφικώνΠληροφοριων,2010,2ηεκδοση, ΠαπασωτηρίουΑθηνα

Ιστοσελίδες

1. http://connor-johnson.com/2014/03/20/simple-kriging-in-python/
2. http://en.wikipedia.org/wiki/Model_selection
3. http://en.wikipedia.org/wiki/Bayesian_information_criterion
4. http://en.wikipedia.org/wiki/Hannan%E2%80%93Quinn_information_criterion
5. http://en.wikipedia.org/wiki/Residual_sum_of_squares
6. http://en.wikipedia.org/wiki/Akaike_information_criterion