



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ΔΙΑΤΡΙΒΗ
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

**ΓΕΩΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΣΤΗΝ
ΠΡΟΣΟΜΟΙΩΣΗ ΡΟΗΣ ΥΠΟΓΕΙΩΝ ΥΔΑΤΩΝ:
ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ KRIGING**

Ευαγγελία Θ. Φλουρή

Εξεταστική επιτροπή:

Γεώργιος Καρατζάς, (Επιβλέπων)

Καθηγητής, Τμήμα Μηχ. Περιβάλλοντος, Πολυτεχνείο Κρήτης

Νικόλαος Νικολαΐδης,

Καθηγητής, Τμήμα Μηχ. Περιβάλλοντος, Πολυτεχνείο Κρήτης

Ιωάννης Νικολός,

Λέκτορας, Τμήμα Μηχ. Παρ. & Διοίκησης, Πολυτεχνείο Κρήτης

Χανιά, Σεπτέμβριος 2007

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της εργασίας Καθ. Γεώργιο Καρατζά για την πολύτιμη βοήθεια και υποστήριξη του καθ' όλη την διάρκεια των μεταπτυχιακών σπουδών μου, καθώς και τα μέλη της εξεταστικής επιτροπής: τον κ. Ιωάννη Νικολό και τον κ. Νικόλαο Νικολαΐδη για την καθοδήγηση και τις χρήσιμες συμβουλές τους για την ολοκλήρωση της παρούσας εργασίας.

Επίσης, ιδιαίτερες ευχαριστίες οφείλω στα μέλη της ομάδας Περιφερειακής Ανάλυσης του Ινστιτούτου Υπολογιστικών Μαθηματικών του Ιδρύματος Τεχνολογίας & Έρευνας: στον Δρ. Ι. Καμαριανάκη, για την πολύτιμη βοήθειά του σε θέματα Στατιστικής, και στον Δρ. Νεκτάριο Χρυσουλάκη, για την διάθεση των δεδομένων βροχόπτωσης που χρησιμοποιήθηκαν στην παρούσα μελέτη.

Τέλος, θα ήθελα να ευχαριστήσω τους προϊσταμένους μου στο Ινστιτούτου Υπολογιστικών Μαθηματικών του Ιδρύματος Τεχνολογίας & Έρευνας: τον Καθ. κ. Β. Δουγαλή, Διευθυντή του ΙΥΜ/ΙΤΕ, που μου έδωσε την δυνατότητα να παρακολουθήσω μεταπτυχιακές σπουδές, τον κ. Ι. Αικατερινάρη, Καθηγητή του Τμήματος Μηχανολόγων και Αεροναυπηγών Μηχανικών του Πανεπιστημίου Πατρών, και τον Δρ. Π. Πραστάκο, Αναπλ. Διευθυντή του ΙΥΜ/ΙΤΕ.

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1	ΕΙΣΑΓΩΓΗ	1
1.1	Μέθοδοι Χωρικής παρεμβολής.....	2
1.2	Γενικά για την Γεωστατιστική	3
1.3	Τυχαίες Συναρτήσεις.....	4
1.3.1	Η θεωρία των Αντιπροσωπευτικών Στοιχειωδών Όγκων	4
1.3.2	Η Θεωρία των Τυχαίων Συναρτήσεων	6
Κεφάλαιο 2	ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΕΚΤΙΜΗΣΗΣ: ΟΡΙΣΜΟΣ ΤΟΥ KRIGING	9
2.1	Το Kriging για την παραδοχή της στασιμότητας.....	11
2.1.1	Υπόθεση Ασθενούς Στασιμότητας ή Στασιμότητας 2 ^{ης} τάξης.....	11
2.1.2	Kriging με υπόθεση στασιμότητας 2 ^{ης} τάξης όταν η μέση τιμή και η συνδιασπορά είναι γνωστά.	11
2.1.3	Υπολογισμός της διασποράς της εκτίμησης	13
2.2	Το Kriging για την παραδοχή του ενδογενούς.....	14
2.2.1	Η υπόθεση του ενδογενούς (intrinsic hypothesis)	14
2.2.2	Το Kriging για την υπόθεση του ενδογενούς.....	16
2.2.3	Η Διασπορά του Σφάλματος Εκτίμησης.....	18
2.3	Ορισμένες παρατηρήσεις για το Kriging	19
2.3.1	Το Kriging ονομάζεται “BLUE”.....	19
2.3.2	Το Kriging είναι μια Ακριβής Παρεμβολή	19
2.3.3	Διάστημα Εμπιστοσύνης.....	19
2.3.4	Υπολογισμός του Πλήρους Πίνακα της Συνδιασποράς.....	20
2.3.5	Οι εξισώσεις στο σύστημα του Kriging δεν εξαρτώνται από τις τιμές μέτρησης Z_i	20
2.3.6	Σχεδιάζοντας ισοψείς χάρτες με το Kriging.....	20
2.3.7	Υπολογίζοντας τις Μέσες Τιμές πάνω σε ένα κάνναβο (mesh) αντί σημειακών τιμών.....	20
2.3.8	Το Kriging με «αβέβαια» δεδομένα.....	22
Κεφάλαιο 3	ΘΕΩΡΙΑ ΤΟΥ ΒΑΡΙΟΓΡΑΜΜΑΤΟΣ – ΒΑΡΙΟΓΡΑΦΙΑ	23
3.1	Ορισμός του ημιβαριογράμματος	23
3.2	Καθορισμός του βαριογράμματος.....	25
3.3	Είδη Ημιβαριογραμμάτων.....	27
3.4	Συμπεριφορά του ημιβαριογράμματος για μεγάλες αποστάσεις	30
3.5	Συμπεριφορά του ημιβαριογράμματος κοντά στο μηδέν.....	30
3.6	Ανισοτροπία στο ημιβαριογράμμα	31
Κεφάλαιο 4	ΕΙΔΗ ΜΕΘΟΔΩΝ KRIGING	33
4.1	Ordinary kriging.....	33

4.2 Simple kriging	34
4.3 Universal kriging	34
4.4 Indicator kriging	35
4.5 Co-kriging.....	36
4.6.1 Οι εξισώσεις του co-kriging.....	36
4.6.2 Cross-Συνδιασπορές και Cross-Βαριογράμματα	39
Κεφάλαιο 5 ΓΕΩΓΡΑΦΙΚΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΩΝ	43
5.1 Γενικά για τα ΓΣΠ	43
5.2 ΓΣΠ και Γεωστατιστική Ανάλυση.....	44
Κεφάλαιο 6 ΔΕΔΟΜΕΝΑ – ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ	47
6.1 Διαθέσιμα Δεδομένα.....	47
6.2 Προϋποθέσεις για την εφαρμογή της μεθόδου kriging	50
6.3 Ανάλυση των δεδομένων.....	50
6.3.1 Χάρτες Voronoi.....	50
6.3.2 Ανάλυση Τάσεων - Trend Analysis	55
Κεφάλαιο 7 ΕΦΑΡΜΟΓΗ ΤΗΣ ΓΕΩΣΤΑΤΙΣΤΙΚΗΣ ΜΕΘΟΔΟΥ	57
7.0 Γεωστατιστική Ανάλυση - Η μεθοδολογία για την εφαρμογή της μεθόδου kriging.....	58
7.1 Ordinary Kriging	59
7.2 Ημibaριογραφία και Ανάλυση Ανισοτροπίας	61
7.3 Η περιοχή επιρροής	65
Κεφάλαιο 8 ΑΠΟΤΕΛΕΣΜΑΤΑ	69
8.1 Χάρτες Εκτίμησης	69
8.2 Χάρτες Πιθανότητας.....	71
8.3 Επικύρωση των αποτελεσμάτων - Validation	72
8.3.1 Αυτό-επικύρωση (cross-validation)	72
8.3.2 Επικύρωση στο 20% των αποτελεσμάτων	78
8.3.3 Επικύρωση στο 40% των αποτελεσμάτων	83
8.3.4 Σχολιασμός της διαδικασίας επικύρωσης	87
8.4 Σύγκριση με άλλες μεθόδους παρεμβολής	89
8.5 Σχολιασμός αποτελεσμάτων.....	96
Κεφάλαιο 9 ΣΥΜΠΕΡΑΣΜΑΤΑ	98
ΒΙΒΛΙΟΓΡΑΦΙΑ	101

Κεφάλαιο 1

ΕΙΣΑΓΩΓΗ

Οι φυσικές παράμετροι που είναι σημαντικές στην υδρογεολογία (όπως π.χ. τα υδραυλικά ύψη, η μεταφερσιμότητα, η διαπερατότητα, ο συντελεστής αποθήκευσης, το πάχος των στρωμάτων, η βροχόπτωση, κ.ά.) αποτελούν όλες συναρτήσεις στο χώρο. Πολύ συχνά δε οι συναρτήσεις αυτές εμφανίζουν ισχυρές μεταβολές στο χώρο. Όμως αυτή η χωρική μεταβολή, γενικά, δεν είναι εντελώς τυχαία. Παραδείγματος χάρη αν πάρουμε μετρήσεις σε δύο διαφορετικές τοποθεσίες, παρατηρούμε ότι όσο πιο κοντινά μεταξύ τους είναι τα σημεία μέτρησης τόσο πλησιέστερες είναι οι τιμές που θα μετρηθούν. Με άλλα λόγια, υπάρχει κάποιου είδους συσχέτιση στην χωρική κατανομή των φυσικών αυτών παραμέτρων. Ο Georges Matheron, [1], ονόμασε «περιοχοποιημένες μεταβλητές» (regionalized variables) τέτοιου είδους ποσότητες. Αυτές είναι μεταβλητές που έχουν σημαντικό ρόλο στην ανάπτυξη ενός φαινομένου στο χώρο (ή/και στο χρόνο) και οι οποίες εμφανίζουν μια συγκεκριμένη δομή. Εδώ με τον όρο «δομή» αναφερόμαστε σε αυτή τη χωρική συσχέτιση η οποία φυσικά, είναι πολύ διαφορετική σε κάθε συγκεκριμένο υδρογεωολογικό σχηματισμό.

Οι regionalized μεταβλητές κατατάσσονται σε δύο βασικές κατηγορίες:

- στατικές (stationary), και
- μη στατικές (non-stationary).

Στις τελευταίες, η μεταβλητή έχει μια συγκεκριμένη «τάση» στο χώρο (π.χ. η μεταβλητή μειώνεται συστηματικά σε μια κατεύθυνση). Αυτό γενικά συμβαίνει όσον αφορά το υδραυλικό ύψος. Αντιθέτως, οι στατικές μεταβλητές δεν εμφανίζουν καμία συστηματική τάση στο χώρο. Κάτι τέτοιο γενικά συμβαίνει με την μεταφερσιμότητα (transmissivity). Στη συνέχεια οι δύο αυτές κατηγορίες θα ορισθούν με μεγαλύτερη ακρίβεια.

Ας δούμε όμως πρώτα το πρόβλημα του «πώς υπολογίζουμε μια περιοχοποιημένη μεταβλητή», το οποίο είναι το πιο κοινό πρόβλημα που αντιμετωπίζουν στην πράξη οι υδρογεωολόγοι και οι μηχανικοί περιβάλλοντος. Πώς υπολογίζουμε την τιμή μιας μεταβλητής στο χώρο, έχοντας λάβει μετρήσεις της μεταβλητής σε ένα σύνολο σημείων (π.χ. υδραυλικά ύψη σε διάφορα πιεζόμετρα, ή μεταφερσιμότητα σε διάφορες γεωτρήσεις). Πώς υπολογίζουμε την τιμή της μεταβλητής σε όλα τα υπόλοιπα σημεία της περιοχής, έτσι ώστε να μπορέσουμε να κατασκευάσουμε ένα χάρτη με ισουήψεις για την μεταβλητή αυτή; Η βέλτιστη μέθοδος υπολογισμού είναι η λεγόμενη μέθοδος **kriging** η οποία θα περιγραφεί στη συνέχεια, τόσο για στατικές, όσο και για μη στατικές μεταβλητές. [1]

1.1 Μέθοδοι Χωρικής παρεμβολής

Χωρική παρεμβολή ονομάζεται η διαδικασία υπολογισμού/εκτίμησης της τιμής μιας μεταβλητής (ή ενός φυσικού μεγέθους) σε σημεία στα οποία δεν είναι γνωστή (δεν υπάρχουν μετρήσεις). Για τον υπολογισμό χρησιμοποιούνται οι τιμές της μεταβλητής αυτής σε άλλα σημεία στο χώρο (ή σε σημεία της περιοχής όπου υπάρχουν μετρήσεις).

Οι μέθοδοι χωρικής παρεμβολής βασίζονται στον νόμο του Tobler, ο οποίος είναι γνωστός και ως «ο πρώτος νόμος της γεωγραφίας». Σύμφωνα με αυτόν «τα πάντα συσχετίζονται μεταξύ τους, αλλά αυτά που βρίσκονται σε κοντινή απόσταση συσχετίζονται μεταξύ τους περισσότερο από τα πιο μακρινά». Τούτο σημαίνει ότι η συσχέτιση των τιμών ενός φυσικού μεγέθους στο χώρο είναι αντιστρόφως ανάλογη (inverse distance) της απόστασης των σημείων στις οποίες αντιστοιχούν οι τιμές αυτές.

Έχουν αναπτυχθεί διάφοροι μέθοδοι χωρικής παρεμβολής και αντίστοιχοι αλγόριθμοι υλοποίησης. Δύο βασικές κατηγορίες είναι οι «σημειακοί» και οι «επιφανειακοί» ανάλογα με τις διαστάσεις των πληροφοριών εισόδου και εξόδου, που μπορεί να είναι σημεία, γραμμές ή επιφάνειες, [1]. Οι σημειακοί αλγόριθμοι διακρίνονται σε ακριβείς και σε προσεγγιστικούς, ανάλογα με το αν διατηρούν τις τιμές του δείγματος εισόδου ή όχι, αντίστοιχα. Οι επιφανειακοί αλγόριθμοι προϋποθέτουν ότι υπάρχει μια καθολική τάση μικρής μεταβλητότητας στην διακύμανση του υπό εξέταση φυσικού μεγέθους, στο οποίο υπερτίθενται τοπικές ανομοιομορφίες μεγάλης μεταβλητότητας. Οι ανομοιομορφίες αυτές προκαλούν σφάλματα δειγματοληψίας και αβεβαιότητες στις σημειακές μετρήσεις και έτσι οι αλγόριθμοι αυτοί χρησιμεύουν για την αναγνώριση τάσεων και την αφαίρεση των σφαλμάτων, μέσω της εξομάλυνσης της παραμετρικής επιφάνειας.

Για την χωρική παρεμβολή σημειακών δεδομένων υπάρχουν διάφορες μέθοδοι, οι βασικότερες των οποίων είναι οι εξής:

- Η μέθοδος Inverse Distance Weighted (IDW)
- Η πολυωνυμική παρεμβολή
- Η μέθοδος χωρικής παρεμβολής με Splines
- Η μέθοδος Kriging

Η κάθε μία από τις παραπάνω μεθόδους κάνει διαφορετικές υποθέσεις σχετικά με τον προσδιορισμό των τιμών που εκτιμώνται. Ανάλογα με το φυσικό μέγεθος ή την ιδιότητα που μελετάται κάθε φορά, αλλά και ανάλογα και με την κατανομή των δειγμάτων και των μετρήσεων, κάποιες μέθοδοι παρέχουν αποτελέσματα με εξαιρετική ακρίβεια. Παρόλα αυτά και ανεξάρτητα από την μέθοδο χωρικής παρεμβολής που χρησιμοποιείται, τα αποτελέσματα της εκτίμησης είναι τόσο περισσότερο ακριβή και αξιόπιστα, όσο περισσότερα είναι τα δεδομένα και όσο καλύτερη και ομαλότερη είναι η κατανομή τους στην περιοχή μελέτης.

Από τις παραπάνω μεθόδους, οι τρεις πρώτες είναι ντετερμινιστικές (προσδιοριστικές) μέθοδοι, καθώς βασίζονται άμεσα στις μετρήσεις και σε συγκεκριμένα μαθηματικά

μοντέλα για τον υπολογισμό της εκτίμησης. Οι ντετερμινιστικές τεχνικές χρησιμοποιούν την υπάρχουσα διάταξη των δειγμάτων για την δημιουργία μιας επιφάνειας εκτίμησης, ή παρεμβάλλουν μια μαθηματική συνάρτηση στα σημεία των μετρήσεων.

Η τελευταία μέθοδος, η μέθοδος Kriging ανήκει σε μια άλλη κατηγορία, την κατηγορία των λεγόμενων γεωστατιστικών μεθόδων, οι οποίες βασίζονται σε στατιστικά μοντέλα που λαμβάνουν υπόψη την τυχαιότητα και την στατιστικές ιδιότητες των δεδομένων. Καθώς η γεωστατιστική βασίζεται στην στατιστική οι τεχνικές αυτές παράγουν όχι μόνο επιφάνειες πρόγνωσης αλλά και επιφάνειες σφαλμάτων και αβεβαιότητας, δίνοντας έτσι μια ένδειξη του πόσο καλή είναι η πρόγνωση.

1.2 Γενικά για την Γεωστατιστική

Από την εφαρμογή της στατιστικής σε προβλήματα γεωλογίας και έρευνας μεταλλευτικών κοιτασμάτων που χρονολογείται από τα μέσα του 1960, προέκυψε ο επιστημονικός κλάδος της γεωστατιστικής. Ο κλάδος αυτός έχει παραδοσιακά ως αντικείμενο την εφαρμογή στατιστικών μεθόδων και στοχαστικής ανάλυσης σε θέματα που απασχολούν κυρίως τον υπολογισμό αποθεμάτων στην εκμετάλλευση μεταλλείων. Σήμερα πλέον, η γεωστατιστική χρησιμοποιείται ως πρακτικό εργαλείο στην υδρολογία, στην επιδημιολογία, στην στατιστική, στις επιστήμες της ατμόσφαιρας, στην γεωγραφία, στην γεωπονία κ.ά.

Η γεωστατιστική απαρτίζεται από ένα σύνολο τεχνικών μοντελοποίησης της χωρικής κατανομής φυσικών μεγεθών που θεωρούνται ως τυχαίες μεταβλητές. Η διαδικασία αυτή γίνεται σε δύο στάδια κυρίως, όπως:

1. Στατιστική ανάλυση των δειγματοληπτικών μετρήσεων κάποιου φυσικού μεγέθους που έχουν περισυλλεγεί από μια περιοχή
2. Παρεμβολή των τιμών των μεγεθών αυτών μεταξύ των σημείων ή περιοχών δειγματοληψίας, με χρήση διαφόρων τεχνικών όπως του πλησιέστερου γείτονα (Nearest neighbor), της δύναμης της αντίστροφης απόστασης (inverse distance), και του Kriging, οι οποίες είναι και οι πλέον διαδεδομένες.

Με τη χρήση μόνο κλασσικών στατιστικών μεθόδων, η μελέτη της χωρικής κατανομής μιας ποσότητας θα γινόταν με την κατασκευή στατιστικών κατανομών των διαφορών $Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})$ για δεδομένο μέτρο του \mathbf{h} και προσανατολισμό του διανύσματος που ενώνει τα δύο σημεία δειγματοληψίας, με τη χρήση στατιστικών παραμέτρων που περιγράφουν το ιστόγραμμα, και αυτή η διαδικασία να επαναλαμβανόταν για πολλές αποστάσεις δειγματοληψίας και προσανατολισμούς στο χώρο. Τα ιστογράμματα που προκύπτουν με τον τρόπο αυτό περιγράφονται από τις εξής παραμέτρους:

- τον μέσο όρο (mean)

$$m(\mathbf{h}) = \frac{1}{n} \sum [Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] ,$$

➤ τη **διασπορά** (variance):

$$\sigma^2 = \frac{1}{n} \sum [Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2.$$

Για να αποφευχθεί η παραπάνω χρονοβόρος διαδικασία, η γεωστατιστική χρησιμοποιεί δύο επιπλέον παραδοχές όσον αφορά την συμπεριφορά των χωρικών μεταβλητών. Οι παραδοχές αυτές είναι η παραδοχή της στασιμότητας 2^{ης} τάξης και η υπόθεση του ενδογενούς που θα παρουσιάστούν αναλυτικά στο Κεφ. 2.

Λόγω της πρώτης παραδοχής, κατά την γεωστατιστική ανάλυση, αν $Z(\mathbf{x}_1)$ και $Z(\mathbf{x}_2)$ είναι δύο διαφορετικές τιμές μιας ιδιότητας, σε δύο διαφορετικά σημεία $\mathbf{x}_1, \mathbf{x}_2$ στο χώρο, τότε οι τιμές αυτές εξαρτώνται μόνο από την απόσταση μεταξύ των σημείων και την διεύθυνσή τους. Η προϋπόθεση αυτή σημαίνει ότι δείγματα που βρίσκονται σε μικρότερη απόσταση $h=|\mathbf{h}|=|\mathbf{x}_1-\mathbf{x}_2|$ μεταξύ τους, συσχετίζονται καλύτερα από ότι όταν βρίσκονται σε μεγαλύτερες αποστάσεις.

1.3 Τυχαίες Συναρτήσεις

Για μια εισαγωγή στην θεωρία των τυχαίων συναρτήσεων θα χρησιμοποιήσουμε σαν παράδειγμα την έννοια του πορώδους, σημειώνοντας ότι τα παρακάτω ισχύουν και για άλλες φυσικές ιδιότητες των πορώδων μέσων όπως π.χ. για την διαπερατότητα (permeability).[1]

Για τον ορισμό των φυσικών ιδιοτήτων ενός πορώδους μέσου υπάρχουν δύο αποδεκτοί τρόποι: η θεωρία των αντιπροσωπευτικών στοιχειωδών όγκων (representative elementary volumes, REV), και η θεωρία των τυχαίων συναρτήσεων (random functions, RF). Θα δούμε ότι οι δύο αυτές θεωρήσεις επηρεάζουν έμμεσα κάθε περιγραφή της χωρικής μεταβλητότητας των υδρογεωλογικών παραμέτρων.

Το όλο πρόβλημα συνίσταται στο ότι οι έννοιες του πορώδους και της διαπερατότητας, οι οποίες είναι έννοιες που αφορούν σημεία σε μια εξίσωση με μερικές παραγώγους, δεν μπορούν να οριστούν ή να μετρηθούν σε μεμονωμένα σημεία καθώς ένα πορώδες μέσο είναι ένα σύνολο από στερεούς κόκκους και κενά. Έτσι, κάτω από μια συγκεκριμένη κλίμακα όγκου, το πορώδες και η διαπερατότητα παύουν να έχουν φυσική σημασία.

1.3.1 Η θεωρία των Αντιπροσωπευτικών Στοιχειωδών Όγκων

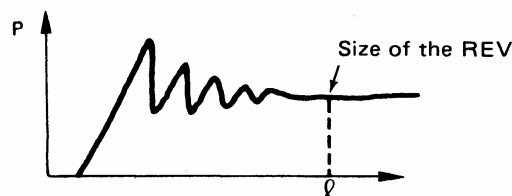
Η μέθοδος των αντιπροσωπευτικών στοιχειωδών όγκων (REV) συνίσταται στο ότι σε ένα μαθηματικό σημείο στο χώρο αποδίδεται η τιμή του πορώδους ή της διαπερατότητας ενός συγκεκριμένου όγκου του υλικού που περιβάλλει το σημείο αυτό, του REV, η οποία θα χρησιμοποιηθεί για τον ορισμό και πιθανά σαν μέτρο της μέσης τιμής της υπό εξέτασης ιδιότητας του όγκου. Συνεπώς, η θεώρηση αυτή εισάγει μια ολοκλήρωση στο χώρο. Είναι προφανές ότι μια τέτοια προσέγγιση είναι η πρώτη που μπορεί να σκεφτεί κανείς. Πίσω από αυτήν υπάρχει η ιδέα ενός δείγματος, το οποίο συλλέγεται και από το οποίο εκτιμάται

η τιμή της σχετικής ιδιότητας με μέτρηση. Για την ακρίβεια, το μέγεθος των REV ορίζεται ότι πρέπει να είναι:

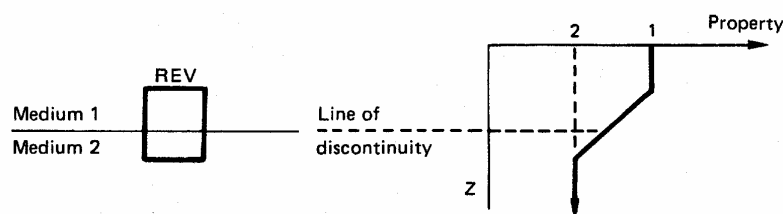
- (1) Αρκετά μεγάλο ώστε να περιλαμβάνει ένα μεγάλο αριθμό πόρων έτσι ώστε να μας επιτρέπει να ορίσουμε μια μέση τιμή της ιδιότητας, ενώ παράλληλα να διασφαλίζει ότι οι μεταβολές από τον ένα πόρο στον άλλο είναι αμελητέες. Κάποιος μπορεί να θεωρήσει ικανοποιητικά μεγάλο ένα μέγεθος της τάξης παραδείγματος χάριν του 1 cm^3 .
- (2) Αρκετά μικρό, έτσι ώστε οι μεταβολές της ιδιότητας από το ένα χωρίο στο επόμενο να προσεγγίζονται από συνεχείς συναρτήσεις, έτσι ώστε να μπορούμε να χρησιμοποιήσουμε τον απειροστικό λογισμό, χωρίς να εισάγεται με τον τρόπο αυτό κάποιο λάθος το οποίο θα μπορεί να εντοπιστεί με τα εργαλεία μέτρησης σε μακροσκοπική κλίμακα, όπου οι συνήθεις διαστάσεις είναι της τάξης μερικών μέτρων.

Πρέπει να σημειώσουμε εδώ ότι το μέγεθος των REV μπορεί να είναι εντυπωσιακά μεγάλο ώστε να ικανοποιείται η δεύτερη υπόθεση των «συνεχών συναρτήσεων» στην κλίμακα των οργάνων μέτρησης.

Το μέγεθος του REV (μετρημένου, παραδείγματος χάριν, με μια από τις χαρακτηριστικές του διαστάσεις ℓ , όπως είναι η ακτίνα μιας σφαίρας ή η πλευρά ενός κύβου) γενικά συνδέεται με την ύπαρξη μιας ευθείας στην καμπύλη που περιγράφει την υπό μελέτη ιδιότητα ως προς την διάσταση αυτή (βλ. Σχ.1). Όμως, κανείς δεν μας επιτρέπει να θεωρήσουμε ότι μια τέτοια ευθεία υπάρχει πάντα. Επομένως το μέγεθος του REV παραμένει αρκετά ακαθόριστο.



Σχήμα 1.



Σχήμα 2.

Άλλες σημαντικές αντιρρήσεις που μπορεί να έχει κανείς όσον αφορά την θεώρηση αυτή για τα πορώδη μέσα είναι δύο ειδών. Πρώτον, ότι είναι πολύ ακατάλληλη για τον χειρισμό ασυνεχειών στο μέσο. Όταν, υπό το πρίσμα ενός πειράματος, ο REV κινείται διασχίζοντας μια ασυνέχεια, η υπό μελέτη ιδιότητα υπόκειται μια συνεχής μεταβολή (βλ. Σχ. 2). Αυτό μερικές φορές εισάγει προβλήματα όσον αφορά την ακριβής προσομοίωση συνόρων και ορίων μεταξύ δύο μέσων. Η δεύτερη, και πιο σημαντική αντίρρηση είναι ότι δεν δίνει μια βάση για την μελέτη της δομής της ιδιότητας στο χώρο. Ως προς αυτό, οι χωρικές μεταβολές της υπό μελέτη ιδιότητας πρέπει να είναι ομαλές, σύμφωνα με την διαδικασία που αναφέρθηκε σχετικά με τις ασυνέχειες. [1]

1.3.2 Η Θεωρία των Τυχαίων Συναρτήσεων

Η θεωρία των τυχαίων συναρτήσεων είναι μια πολύ πιο ισχυρή θεώρηση και συνίσταται στο ότι το υπό μελέτη πορώδες μέσο είναι η πραγματοποίηση μιας τυχαίας διαδικασίας (random process). Θα προσπαθήσουμε να οπτικοποιήσουμε την θεώρηση αυτή. Ας υποθέσουμε ότι δημιουργούμε στο εργαστήριο μερικές στήλες άμμου, η καθεμία από τον ίδιο τύπο άμμου. Κάθε στήλη αντιπροσωπεύει το ίδιο πορώδες μέσο, είναι όμως κατά κάποιον τρόπο διαφορετική από τις υπόλοιπες. Κάθε στήλη είναι μια «πραγματοποίηση» του ίδιου πορώδες μέσου, το οποίο ορίζεται ως η συνισταμένη, το σύνολο, όλων των δυνατών πραγματοποιήσεων (άπειρων σε πλήθος) της ίδιας διαδικασίας.

Μια ιδιότητα όπως το πορώδες μπορεί να οριστεί τότε σε δεδομένο γεωμετρικό σημείο στον χώρο, σαν τον μέσο όρο, ως προς όλες τις πιθανές πραγματοποιήσεις, της τιμής του σημείου (ορισμένης σαν 0 σε κόκκο και 1 σε κενό). Κανείς μιλά για «συνισταμένη ή σύνολο μέσων όρων» αντί για «χωρικό μέσο όρο». Για τις στήλες άμμου που περιγράψαμε, είναι προφανές ότι η συνισταμένη του μέσου όρου (ή αναμενόμενη τιμή) αυτών των σημειακών πορώδων θα είναι ταυτόσημη με το χωρικό μέσο όρο, όπως θα οριζόταν αν θεωρούσαμε ότι κάθε στήλη αποτελεί έναν REV. Επιπλέον, αυτή η συνισταμένη μέσου όρου θα είναι η ίδια για κάθε σημείο της στήλης. Θα ορίσουμε στη συνέχεια ποιές είναι οι αναγκαίες συνθήκες για να είναι αυτό αληθές.

Με πιο γενικούς όρους, μια ιδιότητα Z θα καλείται **τυχαία συνάρτηση** (random function, RF) $Z(\mathbf{x}, \xi)$, αν μεταβάλλεται τόσο μέσα στο χωρικό σύστημα συντεταγμένων \mathbf{x} , όσο και με τη «μεταβλητή κατάστασης» ξ στο σύνολο των πραγματοποιήσεων. Τότε το $Z(\mathbf{x}, \xi_1)$ είναι μια πραγματοποίηση της Z , το $Z(\mathbf{x}_0, \xi)$ είναι μια τυχαία μεταβλητή, το σύνολο των πραγματοποιήσεων της τυχαίας συνάρτησης Z στο \mathbf{x}_0 , και $Z(\mathbf{x}_0, \xi_1)$ είναι η τιμή της Z στο \mathbf{x}_0 για την πραγματοποίηση ξ_1 . Για την απλοποίηση των συμβολισμών η μεταβλητή ξ , γενικά παραλείπεται.

Το βασικό πλεονέκτημα της στοχαστικής θεώρησης είναι ότι πέρα από την αναμενόμενη τιμή, κανείς μπορεί να μελετήσει και άλλες στατιστικές ιδιότητες του πορώδους μέσου στο σύνολο των πραγματοποιήσεων. Πολύ συχνά χρησιμοποιεί κανείς τη διασπορά ή την κατανομή της διασποράς της ιδιότητας, που χαρακτηρίζει το μέγεθος των αποκλίσεων ως προς τον μέσο όρο, και την αυτοσυνδιασπορά ή απλά συνδιασπορά, που χαρακτηρίζει την συσχέτιση μεταξύ των τιμών τιμών της ιδιότητας σε δύο γειτονικά σημεία στο χώρο.

Όμως, κατά την μελέτη ενός δεδομένου πορώδους μέσου θα υπάρχει μια μόνο πραγματοποίηση του θεωρούμενου τυχαίου μέσου. Επομένως, ορισμένες υποθέσεις είναι απαραίτητες για την χρήση της θεωρήσης. Οι πιο συνηθισμένες υποθέσεις είναι η στατικότητα και η εργοδικότητα.

Η **στατικότητα (stationarity)** υποθέτει ότι κάθε στατιστική ιδιότητα του μέσου (μέσος όρος, διασπορά, συνδιασπορά) είναι σταθερή στο χώρο, δηλ. δεν μεταβάλλεται με μια μεταφορά. Θα είναι η ίδια σε κάθε σημείο του μέσου. Η **ασθενής στατικότητα (weak stationarity)** αναφέρεται σε ένα μέσο όπου οι δύο πρώτες στιγμές (moments) είναι στατικές: αν $Z(\mathbf{x})$ είναι η υπό μελέτη ιδιότητα, \mathbf{x} είναι οι συντεταγμένες στις δύο ή τρεις διαστάσεις, τότε η τυχαία συνάρτηση $Z(\mathbf{x})$ ικανοποιεί τα εξής:

- (1) η αναμενόμενη τιμή $E[Z(\mathbf{x})] = m$ δεν είναι συνάρτηση του \mathbf{x} ,
- (2) η συνδιασπορά $E[(Z(\mathbf{x}) - m)(Z(\mathbf{x} + \mathbf{h}) - m)]$ δεν είναι συνάρτηση του \mathbf{x} , αλλά συνάρτηση μόνο της υστέρησης \mathbf{h} , που είναι ένα διάνυσμα στις 2 ή 3 διαστάσεις.

Αναπτύσσοντας την συνδιασπορά αυτή $C(\mathbf{h})$, έχουμε

$$C(\mathbf{h}) = E[Z(\mathbf{x}) \cdot Z(\mathbf{x} + \mathbf{h})] - m^2.$$

Εξ ορισμού,

$$C(0) = E[(Z(\mathbf{x}) - m)^2] = \sigma_z^2$$

είναι η **διασπορά** της Z .

Με πιο αυστηρούς όρους, ισχυρή στατικότητα σημαίνει ότι όλες οι συναρτήσεις κατανομής πιθανότητας της τυχαίας συνάρτησης $Z(\mathbf{x})$ είναι αναλλοίωτες κατά τη μεταφορά, αν θεωρήσουμε ένα σημείο $p(Z(\mathbf{x}))$ ή n σημεία $p(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))$.

Η **εργοδικότητα (ergodicity)** υπονοεί ότι η μοναδική διαθέσιμη πραγματοποίηση συμπεριφέρεται στο χώρο με την ίδια συνάρτηση κατανομής πιθανότητας όπως το σύνολο των πιθανών πραγματοποιήσεων. Με άλλα λόγια, είναι δυνατόν να καθορίσουμε την συνάρτηση κατανομής πιθανότητας μια τυχαίας συνάρτησης για όλες τις πραγματοποιήσεις, παρατηρώντας την μεταβολή στο χώρο της ιδιότητας. Αυτό ονομάζεται «στατιστική συμπερασματολογία» (statistical inference) της συνάρτησης κατανομής πιθανότητας της τυχαίας συνάρτησης $Z(\mathbf{x})$.

Στην ορολογία των στοχαστικών ανελίξεων, ένα φαινόμενο το οποίο είναι «στατικό» και «εργοδικό» ονομάζεται ομογενές. Συνεπώς θα χρησιμοποιήσουμε τον όρο «ενιαίο» (uniform) για να περιγράψουμε ένα μέσο στο οποίο κάποια ιδιότητα δεν μεταβάλλεται στον χώρο. Οι γεωλόγοι παραδοσιακά το ονομάζουν ομογενές.

Κεφάλαιο 2

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΕΚΤΙΜΗΣΗΣ:

ΟΡΙΣΜΟΣ ΤΟΥ KRIGING

Στο κεφάλαιο αυτό θα αναφερθούμε εκτενώς στην γεωστατιστική μέθοδο του kriging. Η μέθοδος πήρε το όνομά της από τον Danie G. Krige (Daniel Gerhardus Krige) έναν Νοτιοαφρικάνο Μηχανικό Μεταλλείων. Ο Krige υπήρξε πρωτοπόρος στον τομέα της γεωστατιστικής και ανέπτυξε μια σειρά από εμπειρικές τεχνικές για την εκτίμηση κοιτασμάτων χρυσού στην Νότιο Αφρική. Η δουλειά του Krige τυποποιήθηκε στις αρχές της δεκαετίας του 1960 από τον έναν Γάλλο μηχανικό, τον Georges Matheron, Καθηγητή στην Ecole Normale Superieure des Mines de Paris (School of Mines), ο οποίος και ονόμασε την μέθοδο Kriging προς τιμήν του Danie G. Krige.

Το Kriging είναι μια μέθοδος για βελτιστοποίηση της εκτίμησης ενός μεγέθους, το οποίο κατανέμεται στο χώρο και το οποίο μετράται σε ένα πλέγμα σημείων. Έστω $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ οι τοποθεσίες των n σημείων των μετρήσεων, όπου το \mathbf{x}_i υποδηλώνει τις συντεταγμένες στο χώρο του σημείου i , και έστω $Z_i = Z(\mathbf{x}_i)$ η τιμή η οποία μετρήθηκε στο σημείο i .

Το πρόβλημα της εκτίμησης των σημείων έγκειται στον καθορισμό της τιμής της ποσότητας Z_0 για κάθε σημείο \mathbf{x}_0 στο οποίο δεν έχει γίνει μέτρηση. Μεταβάλλοντας συνεχώς την θέση του σημείου \mathbf{x}_0 είναι δυνατόν να εκτιμηθεί όλο το πεδίο του Z .

Στην Υδρολογία, το kriging έχει ένα ευρύ πεδίο εφαρμογών, [12], [17], όπως είναι π.χ.:

- (1) Υπολογισμός βροχόπτωσης, [8], θερμοκρασίας, ηλιοφάνειας, [9], κ.ά. με βάση μετρήσεις από κλιματολογικούς σταθμούς.
- (2) Παρεμβολή του πάχους ή της υψομετρίας υπόγειων γεωλογικών σχηματισμών με βάση μετρήσεις από πηγάδια, [10], [11].
- (3) Εκτίμηση υδρογεωλογικών παραμέτρων όπως είναι η μεταφερσιμότητα (transmissivity) ενός ταμιευτήρα, [13], [14], το πιεζομετρικό φορτίο, [12], η υδραυλική αγωγιμότητα, [15], και οι συγκεντρώσεις διαλυτών, [16], με βάση μετρήσεις από πιεζόμετρα.
- (4) Απεικόνιση των συγκεντρώσεων ρυπαντών σε μια λίμνη, κ.ά.

Όμως, το kriging δεν περιορίζεται σε απλή εκτίμηση της τιμής ενός μεγέθους Z απλώς σε σημεία, αλλά μπορεί επίσης να χρησιμοποιηθεί για:

- (1) Υπολογισμό της διασποράς του μεγέθους Z , π.χ. ένδειξη του διαστήματος εμπιστοσύνης της εκτίμησης,
- (2) Εκτίμηση της μέσης τιμής του Z σε ένα δεδομένο μπλοκ, π.χ. στο πλέγμα ενός μοντέλου ή σε ένα υποχωρίο οποιουδήποτε σχήματος ενός ταμιευτήρα νερού.
- (3) Επιλογή της βέλτιστης τοποθεσίας για ένα νέο σημείο μέτρησης, [18], [19] π.χ. ελαχιστοποιώντας την συνολική αβεβαιότητα στο υπό εξέταση πεδίο.

Μια γενίκευση του kriging καθιστά επίσης εφικτή τη δημιουργία ενός άπειρου πλήθους προσομοιώσεων Monte Carlo υπό συνθήκη του πεδίου Z . Παραδείγματος χάριν διαφορετικές πραγματοποιήσεις του χάρτη του Z , που είναι συμβατές με τα δεδομένα των μετρήσεων. Αυτοί οι χάρτες μπορούν να χρησιμοποιηθούν για την οπτικοποίηση της αβεβαιότητας της εκτίμησης και σαν στοιχεία εισόδου σε στοχαστικά μοντέλα (Delhomme, [7]).

Για τον συγκεκριμένο σκοπό του υπολογισμού μιας εκτίμησης θα επιλέξουμε ένα πιθανοθεωρητικό πλαίσιο και θα θεωρήσουμε ότι το μέγεθος $Z(\mathbf{x})$ είναι μια τυχαία συνάρτηση, η $Z(\mathbf{x}, \xi)$, για την οποία μόνο έχουμε μια πραγματοποίηση. Εδώ το \mathbf{x} συμβολίζει το σημείο στον γεωμετρικό χώρο και το ξ την μεταβλητή κατάστασης στο χώρο των πραγματοποιήσεων, το $Z(\mathbf{x}, \xi_1)$ υποδηλώνει μια πραγματοποίηση και το $Z(\mathbf{x}_0, \xi)$ μια τυχαία συνάρτηση, για παράδειγμα, όλο το σύνολο των πραγματοποιήσεων της τυχαίας συνάρτησης Z στο σημείο \mathbf{x}_0 .

Για να χρησιμοποιήσουμε το kriging πρέπει να προσπαθήσουμε να ορίσουμε, στην βάση της μόνης πραγματοποίησης $Z(\mathbf{x}_0, \xi_1)$, τόσο (α) την «δομή» της τυχαίας συνάρτησης $Z(\mathbf{x}, \xi)$, π.χ. την συνάρτηση αυτοσυσχέτισής της (το πρόβλημα που ονομάζεται στοχαστική συμπερασματολογία), όσο και (β) την «βέλτιστη» εκτίμηση του $Z(\mathbf{x}_0, \xi_1)$ για κάθε σημείο \mathbf{x}_0 .

Η πιθανοθεωρητική προσέγγιση που θα χρησιμοποιήσουμε πρέπει να θεωρηθεί σαν μια γλώσσα, ένα εργαλείο, το οποίο απλά οδηγεί σε ένα σύστημα εξισώσεων, από την επίλυση του οποίου θα πάρουμε την επιθυμητή εκτίμηση. Στις περισσότερες περιπτώσεις υπάρχει μόνο μια πραγματοποίηση της Z , η οποία είναι πλήρως καθορισμένη στο χώρο. Η αβεβαιότητά μας για τις τιμές της Z προκύπτει μόνο λόγω της αδυναμίας διαθέσιμων δειγμάτων, και η πιθανοθεωρητική γλώσσα παρέχει ένα χρήσιμο εργαλείο για την έκφραση της αβεβαιότητας αυτής. Όταν θα αρχίσουμε να κάνουμε υποθέσεις για τη Z για να πάρουμε μια εκτίμηση, π.χ. όσον αφορά την στασιμότητά της, αυτές θα είναι υποθέσεις εργασίας και μόνο τοπικά απαιτείται να ικανοποιούνται από τα δεδομένα. Η μόνη αντικειμενική απόδειξη της καταλληλότητας της διαδικασίας θα έρθει από την επιβεβαίωση της πρόγνωσης από μετρήσεις πεδίου.

Στη συνέχεια παραθέτουμε τις εξισώσεις του kriging για τρεις περιπτώσεις: (1) για την υπόθεση στασιμότητας, (2) για την υπόθεση του ενδογενούς και (3) για την υπόθεση της μή στασιμότητας.

2.1 Το Kriging για την παραδοχή της στασιμότητας

2.1.1 Υπόθεση Ασθενούς Στασιμότητας ή Στασιμότητας 2^{ης} τάξης

Μια τυχαία συνάρτηση λέμε ότι ικανοποιεί την υπόθεση στασιμότητας 2^{ης} τάξης, αν

- (1) η μέση τιμή $E\{Z(\mathbf{x})\}$ υπάρχει και δεν εξαρτάται από το σημείο \mathbf{x} του χώρου, αλλά για κάθε \mathbf{x} είναι $E[Z(\mathbf{x})] = m$ σταθερή, και
- (2) η συνάρτηση της αυτοσυσχέτισης ή απλά της συνδιασποράς εξαρτάται μόνο από την απόσταση και όχι από τα σημεία αναφοράς

$$\text{cov}(\mathbf{x}_1, \mathbf{x}_2) = E\{[Z(\mathbf{x}_1, \xi) - m][Z(\mathbf{x}_2, \xi) - m]\} = C(h) ,$$

όπου $h = |\mathbf{x}_1 - \mathbf{x}_2|$.

Αναπτύσσοντας την παραπάνω έκφραση παίρνουμε

$$C(h) = E[Z(\mathbf{x}_1, \xi)Z(\mathbf{x}_2, \xi)] - m^2 . \quad (1)$$

Η συνάρτηση συνδιασποράς $C(h)$ καθορίζει την «δομή» του φαινομένου. Παρατηρούμε ότι $C(0) = \text{var}(Z) = \sigma_Z^2$ είναι η διασπορά (ή διακύμανση της διασποράς) της Z .

Παρατηρούμε ότι η αναμενόμενη τιμή θεωρείται εδώ ως προς όλες τις πραγματοποιήσεις της Z , π.χ. για όλες τις τιμές του ξ . Το ότι η μέση τιμή είναι σταθερή $E[Z(\mathbf{x})] = m$ σημαίνει ότι η αναμενόμενη αυτή τιμή θα είναι η ίδια για κάθε τοποθεσία \mathbf{x} . Αλλά αυτό δεν σημαίνει ότι για μια συγκεκριμένη πραγματοποίηση ξ_1 , το $Z(\mathbf{x}, \xi_1)$ θα είναι σταθερό ως προς το \mathbf{x} , καθώς μια τέτοια συνάρτηση δεν θα ήταν πλέον μεταβλητή στο χώρο!

2.1.2 Kriging με υπόθεση στασιμότητας 2^{ης} τάξης όταν η μέση τιμή και η συνδιασπορά είναι γνωστά.

Από εδώ και πέρα εγκαταλείπουμε την μεταβλητή κατάστασης ξ για απλοποίηση του συμβολισμού. Ορίζουμε μια διαδικασία με μέση τιμή μηδέν ως

$$Y(\mathbf{x}) = Z(\mathbf{x}) - m ,$$

η οποία δίνει $E(Y) = 0$. Θα υπολογίσουμε την τιμή της Y στο σημείο \mathbf{x}_0 . Αν $Y_i = Y(\mathbf{x}_i)$, $i=1, \dots, n$, τότε

$$Y_0^* = Y^*(\mathbf{x}_0) = \sum_{i=1}^n \lambda_0^i Y_i \quad (2)$$

Συμβολισμός: Y_0^* είναι ο εκτιμητής της άγνωστης τιμής Y_0 στο σημείο \mathbf{x}_0 . Ο αστερίσκος δείχνει εξ ορισμού ότι έχουμε να κάνουμε με μια εκτίμηση. Τα λ_0^i είναι τα βάρη του εκτιμητή kriging. Αυτοί είναι και οι άγνωστοι του προβλήματος. Τα i και 0 είναι και τα δύο δείκτες: ο δείκτης i υποδεικνύει ότι το βάρος λ^i σχετίζεται με την μέτρηση που έχει γίνει στο σημείο \mathbf{x}_i , ενώ ο δείκτης 0 δείχνει ότι σε κάθε σημείο \mathbf{x}_0 , όπου θα εκτιμηθεί το Y_0 , θα υπάρχει ένα διαφορετικό σύνολο από βάρη λ^i .

Ο εκτιμητής Y_0^* θα ονομάζεται «βέλτιστος» αν το σφάλμα της εκτίμησης ($Y_0^* - Y_0$) είναι ελάχιστο. Εφόσον οι πραγματική τιμή του Y_0 είναι άγνωστη θα ελαχιστοποιήσουμε μόνο την μαθηματική ελπίδα του τετραγωνικού μέσου του σφάλματος της εκτίμησης:

$$\min \left\{ E \left[\left(Y_0^* - Y_0 \right)^2 \right] \right\} \quad (3)$$

Συνεπώς, καθώς τα Y_0^* και Y_0 είναι τυχαίες μεταβλητές, ελαχιστοποιούμε τη διασπορά του σφάλματος της εκτίμησης ($Y_0^* - Y_0$). Εδώ πρέπει να σημειώσουμε ότι η μαθηματική ελπίδα θεωρείται για ένα συγκεκριμένο και καθορισμένο στο χώρο σημείο \mathbf{x}_0 για όλες τις πιθανές πραγματοποιήσεις της $Y_0^* - Y_0$, δηλ. για όλες τις πιθανές τιμές της μεταβλητής κατάστασης ξ στον συμβολισμό $Y(\mathbf{x}_0, \xi)$.

Με άλλα λόγια, αν είμασταν σε θέση να υπολογίσουμε το Y_0^* για άπειρο πλήθος πραγματοποιήσεων, χρησιμοποιώντας τα ίδια βάρη λ_0^i για κάθε πραγματοποίηση $Y(\mathbf{x}_0, \xi)$, θα μπορούσαμε να make on τον μέσο όρο το ελάχιστο σφάλμα. Βεβαίως, καθώς γενικά έχουμε μια μόνο πραγματοποίηση, θα κάνουμε για την τοποθεσία \mathbf{x}_0 ένα σφάλμα εκτίμησης το οποίο δεν μπορούμε να εκτιμήσουμε ποσοτικά. Αλλά εισάγοντας την εργοδική υπόθεση, μπορούμε να πούμε ότι στον μέσο όρο, ενός μεγάλου πλήθους τοποθεσιών \mathbf{x}_0 όπου εκτιμούμε την Y , το σφάλμα της εκτίμησης μας θα είναι ελάχιστο.

Μπορούμε να αναπτύξουμε τη σχέση (3) εισάγοντας την τιμή του Y_0^* από την εξίσωση (2) και να έχουμε

$$\begin{aligned} E \left[\left(Y_0^* - Y_0 \right)^2 \right] &= E \left[\left(\sum_i \lambda_0^i Y_i - Y_0 \right)^2 \right] \\ &= E \left[\left(\sum_i \lambda_0^i Y_i \right) \left(\sum_j \lambda_0^j Y_j \right) \right] - 2E \left[\left(\sum_i \lambda_0^i Y_i Y_0 \right) \right] + E \left[Y_0^2 \right] \\ &= \sum_i \sum_j \lambda_0^i \lambda_0^j E \left(Y_i Y_j \right) - 2 \sum_i \lambda_0^i E \left(Y_i Y_0 \right) + E \left(Y_0^2 \right) . \end{aligned}$$

Από τον ορισμό της συνάρτησης συνδιασποράς της σχέσης (1), μπορούμε να γράψουμε

$$E(Y_i Y_j) = C(\mathbf{x}_i - \mathbf{x}_j) + m^2 ,$$

αλλά ο μέσος όρος m της Y είναι μηδέν. Επομένως $E(Y_i Y_j) = C(\mathbf{x}_i - \mathbf{x}_j)$ και μάλιστα

$$E \left[\left(Y_0^* - Y_0 \right)^2 \right] = \sum_i \sum_j \lambda_0^i \lambda_0^j C(\mathbf{x}_i - \mathbf{x}_j) - 2 \sum_i \lambda_0^i C(\mathbf{x}_i - \mathbf{x}_0) + C(0) . \quad (4)$$

Η εξίσωση (4) είναι μια τετραγωνική συνάρτηση των βαρών λ_0^i . Για να ελαχιστοποιήσουμε την συνάρτηση αυτή, όλες οι μερικές παράγωγοι ως προς το λ_0^i εξισώνονται με το μηδέν

$$\frac{\partial}{\partial \lambda_0^i} E \left[\left(Y_0^* - Y_0 \right)^2 \right] = 2 \sum_j \lambda_0^j C(\mathbf{x}_i - \mathbf{x}_j) - 2C(\mathbf{x}_i - \mathbf{x}_0) = 0 , \quad i = 1, \dots, n.$$

Έτσι καταλήγουμε σε ένα γραμμικό σύστημα n εξισώσεων με n αγνώστους

$$\sum_j \lambda_0^j C(\mathbf{x}_i - \mathbf{x}_j) = C(\mathbf{x}_i - \mathbf{x}_0) , \quad i=1, \dots, n. \quad (5)$$

Το σύστημα αυτό έχει μόνο μία λύση αν το C είναι θετικά ορισμένη συνάρτηση και αν τα \mathbf{x}_i είναι διακριτά. Μπορούμε να υποθέσουμε ότι πράγματι αυτό ισχύει. Το γραμμικό σύστημα (5) μπορεί εύκολα να επιλυθεί με αντιστροφή του πίνακα των συντελεστών ή με απαλοιφή του Gauss και να υπολογισθούν έτσι τα βάρη $\lambda_0^j, j=1, \dots, n$.

Παρατηρούμε ότι το αριστερό σκέλος της εξίσωσης (5) δεν εξαρτάται από το \mathbf{x}_0 . Συνεπώς ο πίνακας του συστήματος της (5) χρειάζεται να αντιστραφεί μία φορά για όλα τα σημεία \mathbf{x}_0 . Μόνο το δεξιό σκέλος της εξίσωσης (5) είναι συνάρτηση του \mathbf{x}_0 .

2.1.3 Υπολογισμός της διασποράς της εκτίμησης

Τώρα είναι γνωστή η τιμή της εκτίμησης Y_0^* :

$$Y_0^* = \sum_{i=1}^n \lambda_0^i Y_i .$$

Δεν μπορούμε να υπολογίζουμε το σφάλμα της εκτίμησης αλλά μόνο της διασπορά του:

$$\text{var}(Y_0^* - Y_0) = E[(Y_0^* - Y_0)^2] - E[(Y_0^* - Y_0)]^2$$

Ο δεύτερος όρος του δεξιού μέρους της παραπάνω εξίσωσης είναι μηδέν, καθώς

$$E[(Y_0^* - Y_0)] = E(Y_0^*) - E(Y_0) = \sum_{i=1}^n \lambda_0^i E(Y_i) - E(Y_0) = 0 ,$$

εφόσον υποθέσαμε ότι $E(Y) = 0$. Συνεπώς

$$\text{var}(Y_0^* - Y_0) = E[(Y_0^* - Y_0)^2] .$$

Αλλά το δεξιό σκέλος της εξίσωσης έχει ήδη υπολογιστεί, από την εξίσωση (4) σαν συνάρτηση των λ_0^i . Αντικαθιστώντας στην εξίσωση (4) τις τιμές των λ_0^i που υπολογίστηκαν από την (5), έχουμε

$$\sum_j \lambda_0^j C(\mathbf{x}_i - \mathbf{x}_j) = C(\mathbf{x}_i - \mathbf{x}_0) , \quad i=1, \dots, n,$$

$$\sum_i \sum_j \lambda_0^i \lambda_0^j C(\mathbf{x}_i - \mathbf{x}_j) = \sum_i \lambda_0^i C(\mathbf{x}_i - \mathbf{x}_0) ,$$

$$\text{var}(Y_0^* - Y_0) = \text{var}(Y) - \sum_i \lambda_0^i C(\mathbf{x}_i - \mathbf{x}_0) .$$

Μπορούμε να δούμε ότι η *διακύμανση* της εκτίμησης της άγνωστης ποσότητας Y_0 είναι μικρότερη από την *διασπορά της διακύμανσης* της πραγματικής διασποράς της τυχαίας συνάρτησης Y . Καθώς η Y μετρήθηκε στα σημεία \mathbf{x}_i , η αβεβαιότητα για την Y μειώθηκε. Τώρα, μπορούμε να επιστρέψουμε στην αρχική μας μεταβλητή Z ($Z = Y + m$):

$$Z_0^* = m + \sum_i \lambda_0^i (Z_i - m) ,$$

$$\text{var}(Z_0^* - Z_0) = \text{var}(Z) - \sum_i \lambda_0^i C(\mathbf{x}_i - \mathbf{x}_0) .$$

Αργότερα θα εξετάσουμε πως θα υπολογίσουμε μια συνάρτηση συνδιασποράς C από τα δεδομένα. Είναι δυνατόν να παράγει κανείς εξισώσεις kriging και όταν ο μέσος όρος m είναι άγνωστος, είτε υπολογίζοντας κατευθείαν το Z , είτε υπολογίζοντας αυτόν τον μέσο όρο m (βλ. Kitanidis, [2], and Cressie, [3]). Στην συνέχεια προχωράμε κατευθείαν στην περίπτωση που ονομάζεται «παραδοχή του ενδογενούς», όπου το m είναι άγνωστο και δεν ικανοποιείται η υπόθεση στασιμότητας $2^{\text{ης}}$ τάξης.

2.2 Το Kriging για την παραδοχή του ενδογενούς

Στην βιομηχανία των μεταλλείων (εκτίμηση κοιτασμάτων χρυσού), έχει αποδειχθεί ότι η υπόθεση της στασιμότητας $2^{\text{ης}}$ τάξης με μια πεπερασμένη διασπορά $C(0)$ δεν ικανοποιείται από τα δεδομένα για ορισμένες περιπτώσεις. Ομοίως συχνά τέτοιες περιπτώσεις εμφανίζονται και στην υδρολογία. Η πειραματική διασπορά αυξάνει με το μέγεθος της υπό μελέτη περιοχής. Για τις περιπτώσεις αυτές έχει αναπτυχθεί μια λιγότερο αυστηρή υπόθεση, η οποία ονομάζεται «υπόθεση του ενδογενούς», που να καθιστά εφικτή την εκτίμηση.

2.2.1 Η υπόθεση του ενδογενούς (intrinsic hypothesis)

Η υπόθεση του ενδογενούς έγκειται στην θεώρηση ότι αν η διασπορά της Z δεν είναι πεπερασμένη, η διασπορά των $1^{\text{ης}}$ τάξης αυξήσεων της Z είναι πεπερασμένη και αυτές οι αυξήσεις είναι από μόνες τους στάσιμες $2^{\text{ης}}$ τάξης, δηλαδή ότι οι $[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$ ικανοποιούν

$$\left. \begin{aligned} E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] &= m(\mathbf{h}) \\ \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] &= 2\gamma(\mathbf{h}) \end{aligned} \right\} \text{ συναρτήσεις του } \mathbf{h} \text{ και όχι του } \mathbf{x} ,$$

όπου \mathbf{h} είναι ένα διάνυσμα στον διδιάστατο ή τριδιάστατο χώρο, και όπου $\gamma(\mathbf{h})$ είναι γενικά μόνο συνάρτηση της απόστασης \mathbf{h} .

Παρόλο που αυτό δεν είναι απολύτως απαραίτητο, συνήθως υποθέτουμε ότι $m = 0$. Αν αυτό δεν ισχύει, αλλά $m(\mathbf{x} + \mathbf{h}) - m(\mathbf{x}) = m(\mathbf{h})$ τότε η συνάρτηση $Z(\mathbf{x}) - m(\mathbf{x})$ πρέπει να ικανοποιεί αυτή την συνθήκη.

Τότε η διασπορά των αυξήσεων ορίζει μια νέα συνάρτηση που ονομάζεται **ημιβαριόγραμμα** $\gamma(\mathbf{h})$:

$$E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0 \quad (6)$$

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] \quad (7)$$

Οι εξισώσεις (6) και (7) μας επιτρέπουν να γράψουμε

$$\gamma(\mathbf{h}) = \frac{1}{2} E \left\{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \right\}$$

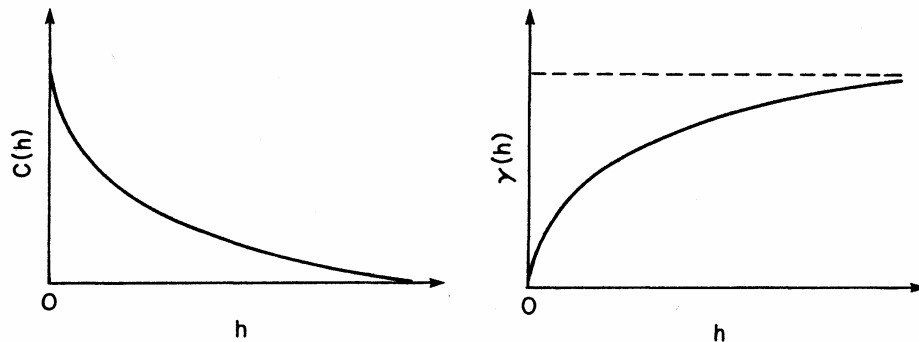
όπου $\gamma(\mathbf{h})$ είναι η μέση τετραγωνική αύξηση της Z μεταξύ δύο σημείων που τα χωρίζει μια απόσταση \mathbf{h} .

Αν συγκρίνουμε την υπόθεση του ενδογενούς με την υπόθεση στασιμότητας 2^{ης} τάξης, βλέπουμε ότι η μεν εξίσωση (6) είναι ισοδύναμη με την $E[Z(\mathbf{x})] = m$, αλλά ότι η εξίσωση (7) είναι λιγότερο αυστηρή από την συνθήκη για την συνδιασπορά $C(\mathbf{h}) = E[Z(\mathbf{x} + \mathbf{h})Z(\mathbf{x})] - m^2$.

Υπάρχει άραγε κάποια σχέση μεταξύ της συνδιασποράς και του ημιβαριόγραμματος; Στην περίπτωση που και οι δύο συναρτήσεις υπάρχουν, π.χ. στην παραδοχή στασιμότητας, μπορούμε να γράψουμε

$$\gamma(\mathbf{h}) = \frac{1}{2} E[Z(\mathbf{x} + \mathbf{h})^2] - E[Z(\mathbf{x} + \mathbf{h})Z(\mathbf{x})] + \frac{1}{2} E[Z(\mathbf{x})^2],$$

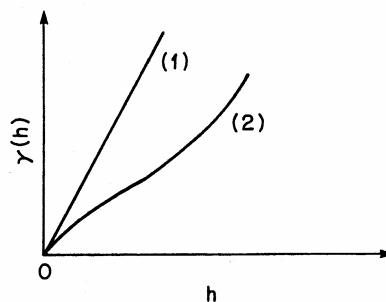
όπου μπορούμε να δούμε ότι $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$.



Σχήμα 3. Συνδιασπορά και ημιβαριόγραμμα.

Αν γνωρίζουμε την συνδιασπορά, το ημιβαριόγραμμα είναι απλά η αντανάκλασή της ως προς τον οριζόντιο άξονα με μια κάθετη μετατόπιση (βλ. Σχήμα 3).

Όταν η διασπορά της Z , $\text{var}(Z)$, είναι πεπερασμένη, το ημιβαριόγραμμα τείνει προς μια ασυμπτωτική τιμή ίση με αυτήν την διασπορά, η οποία καλείται οροφή (sill) του ημιβαριόγραμματος. Η απόσταση στην οποία το ημιβαριόγραμμα φτάνει την ασυμπτωτική του τιμή ονομάζεται εύρος (range) του ημιβαριόγραμματος. Όμως, αν το υπό εξέταση φαινόμενο δεν έχει πεπερασμένη διασπορά, το ημιβαριόγραμμα δεν θα έχει ποτέ μια οριζόντια ασυμπτωτική τιμή, όπως τα ημιβαριόγραμμα που φαίνονται στο ακόλουθο σχήμα.



Το ημιβαριόγραμμα όπως και η συνδιασπορά πρέπει να είναι μια θετικά ορισμένη συνάρτηση. Πράγματι, είναι δυνατόν ναδειχθεί ότι:

- (1) Η συνάρτηση $-\gamma$ πρέπει να είναι υπό συνθήκες θετικά ορισμένη, δηλαδή για όλα τα $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbf{R}^m$ ($m = 1, 2, \text{ ή } 3$), και για όλες τις n -άδες $\lambda_1, \dots, \lambda_n \in \mathbf{R}$ συντελεστών που ικανοποιούν τη σχέση $\sum_i \lambda_i = 0$, πρέπει να ισχύει

$$-\sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \geq 0 ,$$

- (2) Η $\gamma(\mathbf{h})$ για $|\mathbf{h}| \rightarrow \infty$ πρέπει απαραίτητα να αυξάνει λιγότερο γρήγορα από την $|\mathbf{h}|^2$, δηλαδή

$$\lim_{|\mathbf{h}| \rightarrow \infty} \frac{\gamma(\mathbf{h})}{|\mathbf{h}|^2} = 0 .$$

Στην πράξη, μόνο μια περιορισμένη κλάση συναρτήσεων χρησιμοποιούνται για την περιγραφή των ημιβαριογραμμάτων. Αργότερα, θα παραθέσουμε ορισμένες από αυτές μαζί με την στατιστική συμπερασματολογία (τον καθορισμό του βαριογράμματος από τα δεδομένα).

2.2.2 Το Kriging για την υπόθεση του ενδογενούς

Στην συνέχεια, θα προσπαθήσουμε να βρούμε τον εκτιμητή Z_0^* της άγνωστης ποσότητας Z_0 σαν ένα άθροισμα με βάρη, όλων των διαθέσιμων μετρήσεων:

$$Z_0^* = \sum_i \lambda_0^i Z_i . \quad (8)$$

Εφόσον δεν γνωρίζουμε την τιμή του σταθερού μέσου όρου m στην ενδογενή διαδικασία $Z(\mathbf{x})$, εισάγουμε μια επιπλέον συνθήκη για την εκτίμηση Z_0^* , ότι η μαθηματική της ελπίδα να είναι ίση με αυτής της τυχαίας συνάρτησης Z_0 :

$$E(Z_0^* - Z_0) = 0 , \quad \text{ή} \quad E(Z_0^*) = E(Z_0) \quad (9)$$

Έστω m η άγνωστη μαθηματική ελπίδα της διαδικασίας Z . Εισάγοντας την εξίσωση (8) στην σχέση (9), έχουμε

$$E\left[\sum_i \lambda_0^i Z_i\right] = E[Z_0] = m , \quad \text{ή} \quad \sum_i \lambda_0^i E(Z_i) = m , \quad \text{ή} \quad \sum_i \lambda_0^i m = m ,$$

π.χ.

$$\sum_i \lambda_0^i = 1 . \quad (10)$$

Αυτή η συνθήκη απαιτείται για να έχουμε έναν αμερόληπτο εκτιμητή. Τώρα επανα-ορίζουμε το σύνολο των βαρών λ_0^i στην σχέση (8), σύμφωνα με την συνθήκη της σχέσης (10), εισάγοντας την συνθήκη το σφάλμα της εκτίμησης να είναι ελάχιστο, δηλ.

$$E\left[\left(Z_0^* - Z_0\right)^2\right] = \text{ελάχιστο} , \quad (11)$$

ή $\text{var}(Z_0^* - Z_0) = \text{ελάχιστο}$, εφόσον $E(Z_0^* - Z_0) = 0$.

Αναπτύσσοντας τη σχέση (11) έχουμε:

$$\begin{aligned} E[(Z_0^* - Z_0)^2] &= E\left[\left(\sum_i \lambda_0^i Z_i - Z_0\right)^2\right] = E\left[\left(\sum_i \lambda_0^i Z_i - \sum_i \lambda_0^i Z_0\right)^2\right] \\ &= E\left[\left(\sum_i \lambda_0^i (Z_i - Z_0)\right)^2\right] = E\left[\left(\sum_i \lambda_0^i (Z_i - Z_0)\right)\left(\sum_j \lambda_0^j (Z_j - Z_0)\right)\right] \end{aligned}$$

και συνεπώς

$$E[(Z_0^* - Z_0)^2] = \sum_i \sum_j \lambda_0^i \lambda_0^j E[(Z_i - Z_0)(Z_j - Z_0)] . \quad (12)$$

Στη συνέχεια, χρησιμοποιούμε τον ορισμό του ημιβαριογράμματος

$$\begin{aligned} \gamma(\mathbf{x}_i - \mathbf{x}_j) &= \frac{1}{2} E[(Z_i - Z_j)^2] \\ &= \frac{1}{2} E[(Z_i - Z_0) - (Z_j - Z_0)]^2 \\ &= \frac{1}{2} E[(Z_i - Z_0)^2] + \frac{1}{2} E[(Z_j - Z_0)^2] - E[(Z_i - Z_0)(Z_j - Z_0)] \end{aligned}$$

άρα

$$\gamma(\mathbf{x}_i - \mathbf{x}_j) = \gamma(\mathbf{x}_i - \mathbf{x}_0) + \gamma(\mathbf{x}_j - \mathbf{x}_0) - E[(Z_i - Z_0)(Z_j - Z_0)] . \quad (13)$$

Από την εξίσωση (13) μπορούμε να υπολογίσουμε την μαθηματική ελπίδα, που χρειαζόμαστε στην εξίσωση (12). Με αφαίρεση βρίσκουμε

$$\begin{aligned} E[(Z_0^* - Z_0)^2] &= -\sum_i \sum_j \lambda_0^i \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \sum_i \sum_j \lambda_0^i \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_0) \\ &\quad + \sum_i \sum_j \lambda_0^i \lambda_0^j \gamma(\mathbf{x}_j - \mathbf{x}_0) . \end{aligned}$$

Μπορούμε να εξάγουμε το $\sum_i \lambda_0^i$ ή το $\sum_j \lambda_0^j$ ως κοινό παράγοντα στους δύο τελευταίους όρους του δεξιού μέρους της εξίσωσης, αλλά σύμφωνα με την εξίσωση (10) αυτά τα δύο αθροίσματα είναι ίσα με την μονάδα. Επιπλέον

$$\sum_i \lambda_0^i \gamma(\mathbf{x}_i - \mathbf{x}_0) = \sum_j \lambda_0^j \gamma(\mathbf{x}_j - \mathbf{x}_0) ,$$

συνεπώς

$$E[(Z_0^* - Z_0)^2] = -\sum_i \sum_j \lambda_0^i \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_j) + 2 \sum_i \lambda_0^i \gamma(\mathbf{x}_i - \mathbf{x}_0) . \quad (14)$$

Πάλι καταλήγουμε σε μία τετραγωνική μορφή των αγνώστων λ_0^i . Η ελαχιστοποίηση της εξίσωσης (14), με τον περιορισμό της εξίσωσης (10), γίνεται με τους πολλαπλασιαστές Lagrange. Απλά ελαχιστοποιούμε την έκφραση

$$\frac{1}{2} E[(Z_0^* - Z_0)^2] - \mu \left[\sum_i \lambda_0^i - 1 \right] , \quad (15)$$

όπου μ είναι ένας νέος άγνωστος, που ονομάζεται πολλαπλασιαστής Lagrange, ο οποίος έρχεται να προστεθεί στους n προηγούμενους αγνώστους λ_0^i . Μπορεί να αποδειχθεί ότι, όταν η παραπάνω έκφραση γίνεται ελάχιστη, η γραμμική συνθήκη $\sum_i \lambda_0^i = 1$ ικανοποιείται για την τιμή του $E[(Z_0^* - Z_0)^2]$, που είναι η μικρότερη συμβατή με τον περιορισμό. Στην εξίσωση (15) έχουμε διαιρέσει δια 2 και έχουμε εισάγει ένα μείον πριν το μ ώστε να απλοποιήσουμε τις εκφράσεις που ακολουθούν, αλλά καθώς ο μ είναι άγνωστος, αυτό είναι άνευ σημασίας.

Η ελάχιστη τετραγωνική μορφή των λ και μ λαμβάνεται εξισώνοντας με το μηδέν τις μερικές παραγώγους της ως προς τα λ_0^i και μ . Τότε έχουμε

$$\sum_j \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \mu = \gamma(\mathbf{x}_i - \mathbf{x}_0) \quad , \quad i = 1, \dots, n, \quad (16)$$

$$\sum_i \lambda_0^i = 1 \quad .$$

Ας γράψουμε το πλήρες γραμμικό σύστημα του kriging στην μορφή πινάκων:

$$\begin{pmatrix} 0 & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \dots & \gamma_{2n} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \dots & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_0^1 \\ \lambda_0^2 \\ \vdots \\ \lambda_0^n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix} \quad (17)$$

Σημείωση: Έχουμε συμβολίσει με γ_{ij} το $\gamma(\mathbf{x}_i - \mathbf{x}_j)$. Η διαγώνιος του πίνακα είναι μηδέν καθώς $\gamma_{ij} = \gamma(\mathbf{x}_i - \mathbf{x}_j) = \gamma(0) = 0$.

Ο πίνακας των εξισώσεων kriging είναι πάντα κανονικός αν το $-\gamma$ είναι θετικά ορισμένο υπό συνθήκη. Εδώ πάλι ο πίνακας χρειάζεται να αντιστραφεί μια φορά για όλα τα σημεία \mathbf{x}_0 .

2.2.3 Η Διασπορά του Σφάλματος Εκτίμησης

Η διασπορά του σφάλματος της εκτίμησης μπορεί επίσης να υπολογιστεί:

$$\text{var}(Z_0^* - Z_0) = E[(Z_0^* - Z_0)^2] \quad ,$$

Καθώς $E(Z_0^* - Z_0) = 0$. Μπορούμε να υπολογίσουμε την τιμή της αντικαθιστώντας την λύση των εξισώσεων (16) στην εξίσωση (14):

$$\text{var}(Z_0^* - Z_0) = \sum_i \lambda_0^i \gamma(\mathbf{x}_i - \mathbf{x}_0) + \mu \quad . \quad (18)$$

Τώρα έχουμε όλες τις εξισώσεις του kriging, που χρησιμοποιούνται για την ενδογενής υπόθεση και για την στασιμότητα δευτέρης τάξης όταν ο μέσος όρος είναι γνωστός.

2.3 Ορισμένες παρατηρήσεις για το Kriging

2.3.1 Το Kriging ονομάζεται “BLUE”

Το BLUE είναι ακρωνύμιο για το Βέλτιστος Γραμμικός Αμερόληπτος Εκτιμητής (Best Linear Unbiased Estimator).

Άλλες κλάσεις εκτιμητών χρησιμοποιούνται επίσης στην πράξη. Παραδείγματος χάριν, μη γραμμικοί εκτιμητές μπορούν να κατασκευαστούν με προηγούμενο μετασχηματισμό των δεδομένων. Το disjunctive kriging και το indicator kriging (βλ. de Marsily, [1]) είναι παραδείγματα μη γραμμικών εκτιμητών. Οι πολωτικοί (μεροληπτικοί) εκτιμητές ορισμένες φορές προτιμούνται από τους αμερόληπτους, π.χ. όταν επιπλέον περιορισμοί εισάγονται στην εκτίμηση ή απλά όταν η συνθήκη μη μεροληψίας έχει αφαιρεθεί. Θα επιστρέψουμε στο πρόβλημα αυτό αργότερα με ένα παράδειγμα.

2.3.2 Το Kriging είναι μια Ακριβής Παρεμβολή

Αν προσπαθήσουμε να υπολογίσουμε την τιμή της Z στο σημείο \mathbf{x}_k , η οποία έχει ήδη μετρηθεί, π.χ. $\mathbf{x}_k \in (\mathbf{x}_1, \dots, \mathbf{x}_n)$, το σύστημα του kriging δίνει

$$Z_k^* = Z_k ,$$

για μια μετρημένη τιμή, π.χ. $\lambda_k^k = 1$, $\lambda_k^i = 0$, $i \neq k$, και

$$\text{var}(Z_k^* - Z_k) = 0$$

Συνεπώς δεν υπάρχει καμμία αβεβαιότητα σε ένα σημείο μέτρησης.

Αυτό έρχεται σε αντίθεση παραδείγματος χάρι με την παρεμβολή ελαχίστων τετραγώνων ενός πολυωνύμου, η οποία δεν θα δώσει ποτέ την πραγματική τιμή στα σημεία μέτρησης.

2.3.3 Διάστημα Εμπιστοσύνης

Η γνώση της διασποράς του σφάλματος της εκτίμησης δεν είναι, γενικά, αρκετή για τον ορισμό του διαστήματος εμπιστοσύνης των εκτιμήσεων. Όμως, κανείς μπορεί πολύ συχνά να υποθέσει ότι η κατανομή του σφάλματος αυτού είναι κανονική: στην περίπτωση αυτή μπορούμε να πούμε, παραδείγματος χάριν, ότι το 95% του διαστήματος εμπιστοσύνης είναι $\pm 2\sigma$, όπου σ είναι η τυπική απόκλιση, δηλαδή η τετραγωνική ρίζα της διασποράς:

$$\sigma = \sqrt{\text{var}(Z_k^* - Z_k)} .$$

Τότε η εκτίμηση Z_0 με 95% επιτυχία είναι

$$Z_0^* = \sum_i \lambda_0^i Z_i \pm 2\sigma .$$

Γνωρίζουμε επίσης ότι πολλές άλλες συναρτήσεις κατανομής ικανοποιούν επίσης ένα $\pm 2\sigma$ διάστημα εμπιστοσύνης στο 95%, και συνεπώς η έκφραση αυτή χρησιμοποιείται πολύ συχνά.

2.3.4 Υπολογισμός του Πλήρους Πίνακα της Συνδιασποράς

Αντί να υπολογίζουμε μόνο τη διασπορά του σφάλματος της εκτίμησης, $\text{var}(Z_0^* - Z_0)$ είναι δυνατόν με το kriging να υπολογίσουμε τον πλήρη πίνακα της συνδιασποράς του σφάλματος της εκτίμησης. Μπορούμε να δείξουμε ότι

$$\begin{aligned} \text{cov}\left[(Z_1^* - Z_1)(Z_2^* - Z_2)\right] &= -\gamma(\mathbf{x}_1 - \mathbf{x}_2) - \sum_i \sum_j \lambda_1^i \lambda_2^j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \sum_i \sum_j \lambda_0^i \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_0) \\ &\quad + \sum_i \lambda_1^i \gamma(\mathbf{x}_i - \mathbf{x}_2) + \sum_j \lambda_2^j \gamma(\mathbf{x}_j - \mathbf{x}_1) . \end{aligned}$$

Αυτό ποσοτικοποιεί τη σχέση μεταξύ του σφάλματος της εκτίμησης στην τοποθεσία \mathbf{x}_1 και \mathbf{x}_2 .

2.3.5 Οι εξισώσεις στο σύστημα του Kriging δεν εξαρτώνται από τις τιμές μέτρησης Z_i

Πράγματι, το μόνο που χρειάζεται να γνωρίζουμε είναι οι συντεταγμένες \mathbf{x}_i των σημείων μέτρησης ώστε να υπολογίσουμε τα βάρη λ^i . Αν τα δεδομένα μεταβάλλονται με το χρόνο, παραδείγματος χάριν, αυτά τα βάρη λ^i μπορούν να χρησιμοποιηθούν για διάφορες καταστάσεις.

2.3.6 Σχεδιάζοντας ισοψείς χάρτες με το Kriging

Λύνοντας το σύστημα του kriging μπορούμε να υπολογίσουμε το Z_0 σε κάθε σημείο \mathbf{x}_0 . Για να σχεδιάσουμε ένα χάρτη με υσουψείς καμπύλες, γενικά επιλέγουμε ένα μεγάλο αριθμό σημείων \mathbf{x}_0 πάνω σε έναν κανονικό κάνναβο (mesh), ανεξάρτητα από την θέση των σημείων μέτρησης. Οι υσουψείς καμπύλες τότε σχεδιάζονται είτε με το χέρι ή με κάποιο κλασσικό πακέτο περιγράμματος, που συνήθως απαιτεί σαν είσοδο την τιμή του Z σε έναν τέτοιο κανονικό κάνναβο. Κανείς πρέπει να σχεδιάσει επίσης τον χάρτη της διασποράς της εκτίμησης για να κατανοήσει την αβεβαιότητα. Επίσης, μπορεί να σχεδιάσει το χάρτη του διπλάσιου της τυπικής απόκλισης, που αντιστοιχεί στο 95% του διαστήματος εμπιστοσύνης.

2.3.7 Υπολογίζοντας τις Μέσες Τιμές πάνω σε ένα κάνναβο (mesh) αντί σημειακών τιμών

Αντί να εκτιμήσουμε την τιμή Z_0 στο σημείο \mathbf{x}_0 , είναι επίσης δυνατόν να εκτιμήσουμε κάθε γραμμικό συνδυασμό της τιμής της μεταβλητής Z , και ειδικότερα της μέσης τιμής της πάνω σε μια δεδομένη περιοχή S_0 :

$$Z_{S_0} = \frac{1}{S_0} \int_{S_0} Z(\mathbf{x}) d\mathbf{x} .$$

Εδώ η S_0 μπορεί να είναι ένας δεδομένος κάνναβος ή και ολόκληρο το χωρίο (π.χ. για εκτιμήσεις της μέσης βροχόπτωσης σε μια λεκάνη απορροής κατά τη διάρκεια μιας καταιγίδας). Ο εκτιμητής της μέσης τιμής κατασκευάζεται κατευθείαν σαν ένας γραμμικός συνδυασμός των διαθέσιμων δεδομένων:

$$Z_{S_0}^* = \sum_i \lambda_0^i Z_i .$$

Χρησιμοποιώντας τις ίδιες συνθήκες αμερόληπτης ή βέλτιστης εκτίμησης, είναι επίσης δυνατόν να υπολογιστεί το παρακάτω νέο σύστημα kriging:

$$\sum_j \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_j) + \mu = \bar{\gamma}(\mathbf{x}_i, S_0) , \quad i = 1, \dots, n ,$$

$$\sum_i \lambda_0^i = 1 ,$$

με

$$\bar{\gamma}(\mathbf{x}_i, S_0) = \frac{1}{S_0} \int_{S_0} \gamma(\mathbf{x}_i - \mathbf{x}) d\mathbf{x}$$

(μέση τιμή του ημιβαριογράμματος μεταξύ των \mathbf{x}_i και της περιοχής S_0).

Η διασπορά της εκτίμησης δίνεται από

$$\text{var}(Z_{S_0}^* - Z_{S_0}) = \sum_i \lambda_0^i \bar{\gamma}(\mathbf{x}_i, S_0) + \mu - \bar{\bar{\gamma}}(S_0, S_0) ,$$

με

$$\bar{\bar{\gamma}}(S_0, S_0) = \frac{1}{S_0^2} \int_{S_0} \int_{S_0} \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

Παρόμοια, είναι δυνατόν να χρησιμοποιηθεί στη θέση των μετρήσεων η τιμή της Z η οποία έχει υπολογιστεί σαν μέση τιμή πάνω σε μία συγκεκριμένη περιοχή. Παραδείγματος χάρη, αν η μέτρηση Z_j είναι μια μέση τιμή της παραμέτρου στην περιοχή S_j , το $\gamma(\mathbf{x}_i - \mathbf{x}_j)$ θα αντικατασταθεί στις εξισώσεις του kriging από

$$\bar{\gamma}(\mathbf{x}_i, S_j) = \frac{1}{S_j} \int_{S_j} \gamma(\mathbf{x}_i - \mathbf{x}) d\mathbf{x} ,$$

και το $\bar{\gamma}(\mathbf{x}_j, S_0)$ στο δεξιό μέρος από το

$$\bar{\bar{\gamma}}(S_j, S_0) = \frac{1}{S_0 S_j} \int_{S_0} \int_{S_j} \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y} .$$

Με τον τρόπο αυτό καθίσταται δυνατή η ανά πάσα στιγμή χρήση δεδομένων που έχουν συλλεχθεί με διαφορετικές μεθόδους (μετρήσεις από δείγματα πυρήνων, slug tests, long pumping tests, κ.ά., για την περίπτωση της μεταφερσιμότητας, για παράδειγμα). Σημειώνουμε όμως ότι το γ εδώ είναι το ημιβαριόγραμμα των δεδομένων από σημειακές μετρήσεις. Δεδομένα από μη σημειακές μετρήσεις δεν θα πρέπει επομένως να χρησιμοποιηθούν για την εκτίμηση του γ (ή προηγουμένως θα πρέπει να γίνει μια deconvolution), εκτός εάν όλα τα δεδομένα έχουν μετρηθεί πάνω στην ίδια περιοχή. Σε τέτοιες περιπτώσεις, το $\bar{\gamma}$ καθορίζεται κατευθείαν. Αν μετρήσεις που έχουν γίνει σαν μέσες τιμές πάνω σε διαφορετικές περιοχές πρέπει να χρησιμοποιηθούν την ίδια στιγμή και για τον καθορισμό του γ και στο kriging, τότε θα πρέπει να χρησιμοποιηθεί το co-kriging.

2.3.8 Το Kriging με «αβέβαια» δεδομένα

Μέχρι τώρα έχουμε υποθέσει ότι οι μετρημένες τιμές Z_i είναι γνωστές χωρίς καμμία αβεβαιότητα. Στην πραγματικότητα αυτό δεν συμβαίνει πάντα, αλλά το kriging μπορεί επίσης να χειριστεί αβέβαια δεδομένα, [1]. Σε περιπτώσεις που τα σφάλματα ε_i που αντιστοιχούν σε κάθε μέτρηση Z_i είναι

- (1) μη συστηματικά, δηλ. $E[\varepsilon_i] = 0, i = 1, \dots, n$,
- (2) ασυσχέτιστα μεταξύ τους, δηλ. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$,
- (3) ασυσχέτιστα με το Z , δηλ. $\text{cov}[\varepsilon_i, Z(\mathbf{x})] = 0, \forall i, \forall \mathbf{x}$, και
- (4) έχουν μια γνωστή διασπορά σ_i^2 (διαφορετική για κάθε i),

εύκολα μπορεί ναδειχθεί ότι οι εξισώσεις του συστήματος του kriging, γίνονται για παράδειγμα, για την εκτίμηση σημειακών τιμών ως εξής:

- (1) Εκτιμητής

$$Z_0^* = \sum_i \lambda_0^i Z_i ,$$

- (2) Εξισώσεις kriging

$$\sum_j \lambda_0^j \gamma(\mathbf{x}_i - \mathbf{x}_j) - \lambda_0^i \sigma_i^2 + \mu = \gamma(\mathbf{x}_i, \mathbf{x}_0) , \quad i = 1, \dots, n ,$$
$$\sum_i \lambda_0^i = 1 ,$$

- (3) Διασπορά

$$\text{var}(Z_0^* - Z_0) = \sum_i \lambda_0^i \gamma(\mathbf{x}_i, \mathbf{x}_0) + \mu .$$

Η συνάρτηση γ θεωρείται ότι είναι το πραγματικό ημιβαριόγραμμα της ποσότητας Z (χωρίς σφάλματα). Αυτό μπορεί να γίνει εφικτό ορίζοντας το γ μόνο σε εκείνα τα σημεία όπου το σφάλμα των μετρήσεων είναι μηδέν. Τα Z_i είναι οι μετρημένες τιμές (δηλ. με τα σφάλματα της μέτρησης, αν υπάρχουν). Η ποσότητα Z_0 είναι η «πραγματική» (άγνωστη) τιμή (χωρίς σφάλματα μέτρησης).

Η μόνη αλλαγή από το σύστημα είναι ότι οι εξισώσεις τώρα έχουν $-\sigma_i^2$ στην διαγώνιο αντί για μηδενικά. Είναι επίσης δυνατόν να χρησιμοποιηθούν και «βέβαια» και «αβέβαια» δεδομένα την ίδια στιγμή: απλώς αντικαθιστούμε το $-\sigma_i^2$ με 0 για τα «βέβαια» δεδομένα. Αν τα σφάλματα ε_i συσχετίζονται μεταξύ τους, οι εξισώσεις στο σύστημα του kriging γίνονται λίγο πιο πολύπλοκες. Οι ίδιες εξισώσεις μπορούν να παραχθούν για την εκτίμηση πάνω σε μέσες τιμές αντί για σημειακές τιμές.

Κεφάλαιο 3

ΘΕΩΡΙΑ ΤΟΥ ΒΑΡΙΟΓΡΑΜΜΑΤΟΣ – ΒΑΡΙΟΓΡΑΦΙΑ

Στην στατιστική η διασπορά περιγράφει ποσοτικά την απόσταση ή διαφορά των τιμών από την μέση τιμή μιας κατανομής. Αντίστοιχα, στη γεωστατιστική, το ημιβαριόγραμμα περιγράφει την απόσταση ή τη διαφορά των τιμών συναρτήσει της απόστασης και της διεύθυνσης, συνεπώς έχει τρισδιάστατο χαρακτήρα.

3.1 Ορισμός του ημιβαριογράμματος

Αν θεωρήσουμε δύο αριθμητικές τιμές $Z(\mathbf{x})$ και $Z(\mathbf{x}+\mathbf{h})$ σε δύο σημεία \mathbf{x} και $\mathbf{x}+\mathbf{h}$ που ορίζονται από ένα διάνυσμα \mathbf{h} , τότε η διαφοροποίηση μεταξύ των τιμών περιγράφεται ποσοτικά από την συνάρτηση $2\gamma(\mathbf{x},\mathbf{h})$, που ορίζεται ως η μέση τιμή της τυχαίας μεταβλητής $[Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})]^2$, δηλαδή:

$$2\gamma(\mathbf{x},\mathbf{h}) = E \left\{ [Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2 \right\} .$$

Η συνάρτηση γ ονομάζεται *ημιβαριόγραμμα*.

Παράλληλα, ο βαθμός συσχέτισης δύο χωρικών μεταβλητών, δίνεται από την συνάρτηση της συνδιασποράς ή *συμμεταβλητότητας* (covariance). Αν δύο χωρικές μεταβλητές $Z(\mathbf{x})$ και $Z(\mathbf{x}+\mathbf{h})$ εμφανίζουν διασπορά στα σημεία \mathbf{x} και $\mathbf{x}+\mathbf{h}$, τότε εμφανίζουν και συμμεταβλητότητα στα \mathbf{x} και $\mathbf{x}+\mathbf{h}$ η οποία είναι συνάρτηση της μεταξύ τους απόστασης και ισούται με:

$$\begin{aligned} C(\mathbf{x}, \mathbf{x}+\mathbf{h}) &= E \left\{ [Z(\mathbf{x}) - m(\mathbf{x})][Z(\mathbf{x}+\mathbf{h}) - m(\mathbf{x}+\mathbf{h})] \right\} \Rightarrow \\ C(\mathbf{h}) &= E \left\{ Z(\mathbf{x}+\mathbf{h})Z(\mathbf{x}) \right\} - m^2 . \end{aligned}$$

Λόγω της χωρικής φύσης της συμμεταβλητότητας, της διασποράς και του ημιβαριογράμματος, μπορεί να αποδειχθεί ότι ισχύουν οι παρακάτω σχέσεις:

$$\text{Var} \{Z(\mathbf{x})\} = E \left\{ [Z(\mathbf{x}) - m]^2 \right\} = C(0) , \quad \forall \mathbf{x} .$$

και

$$\gamma(\mathbf{h}) = \frac{1}{2} E \left\{ [Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})]^2 \right\} = C(0) - C(\mathbf{h}) , \quad \forall \mathbf{x} .$$

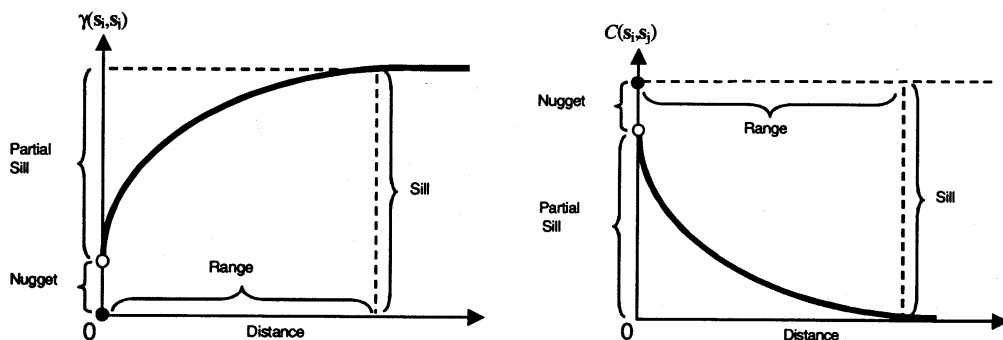
Στην φύση, η συσχέτιση μεταξύ δύο χωρικών μεταβλητών, $Z(\mathbf{x})$ και $Z(\mathbf{x}+\mathbf{h})$, μειώνεται μέχρι μηδενισμού, όταν το μέτρο του διανύσματος τείνει στο άπειρο ($|\mathbf{h}| \rightarrow \infty$). Στην

πράξη θεωρείται πως το $C(\mathbf{h})=0$, όταν η απόσταση $|\mathbf{h}| \geq a$. Η απόσταση αυτή, στην οποία το $C(\mathbf{h})=0$ καλείται **απόσταση συσχέτισης** ή εύρος ή μέγιστη απόσταση επιρροής (**correlation length** ή range) και περιγράφει την μετάβαση από μία κατάσταση στην οποία υπάρχει συσχέτιση μεταξύ των μεταβλητών ($|\mathbf{h}| < a$), σε μία κατάσταση όπου δεν είναι δυνατή καμία συσχέτιση. Από την εξίσωση (2.8) σε συνδυασμό με το γεγονός ότι το ημιβαριόγραμμα περιγράφει την διασπορά μεταξύ των θέσεων στο χώρο (\mathbf{x} και $\mathbf{x}+\mathbf{h}$), είναι δυνατό να εξαχθούν τα παρακάτω συμπεράσματα:

$$\gamma(0) = 0 \quad \text{και} \quad \gamma(\mathbf{h}) = \gamma(-\mathbf{h}) \geq 0.$$

Το ημιβαριόγραμμα αναπαριστά γραφικά την εξάρτηση της διασποράς από την απόσταση μεταξύ των δειγμάτων, ενώ ένα σύνολο από ημιβαριόγραμμα προς διάφορες κατευθύνσεις περιγράφει την ανισοτροπία του πληθυσμού. Ένα κλασικό ημιβαριόγραμμα και η συμμεταβλητότητα του απεικονίζονται στο Σχήμα 4.

Το ημιβαριόγραμμα, όπως φαίνεται στο Σχήμα 4, παύει να αυξάνει μετά από μία συγκεκριμένη απόσταση, που είναι η απόσταση συσχέτισης ή εύρος, και πλησιάζει το όριο $\gamma(\infty)$ που ονομάζεται **οροφή (sill)**, του ημιβαριογράμματος και αποτελεί την a priori διασπορά της χωρικής μεταβλητής: $\gamma(\infty) = \text{Var}\{Z(\mathbf{x})\} = C(0)$. Τέτοια ημιβαριόγραμμα που εμφανίζουν οροφή και απόσταση ονομάζονται μεταβατικά (transitional) μοντέλα, [1]. Σε ένα μεταβατικό φαινόμενο, οποιαδήποτε τιμή της μεταβλητής θα συσχετιστεί με οποιαδήποτε άλλη η οποία απέχει το πολύ απόσταση ίση με το εύρος. Όπως προαναφέρθηκε, η συσχέτιση αυτή και συνεπώς η ακτίνα επιρροής του ενός δείγματος επάνω σε άλλο μειώνεται καθώς η απόσταση μεταξύ των δειγμάτων αυξάνει, μέχρι η τελευταία να φτάσει την οριακή τιμή a , όπου οι χωρικές μεταβλητές $Z(\mathbf{x})$ και $Z(\mathbf{x}+\mathbf{h})$ είναι πλήρως ασυσχέτιστες μεταξύ τους.



Σχήμα 4. Απεικόνιση του ημιβαριογράμματος, της συμμεταβλητότητας και των βασικών παραμέτρων τους.

Ένα άλλο χαρακτηριστικό του ημιβαριογράμματος είναι το nugget. Όπως φαίνεται στο Σχήμα 4, η οροφή είναι το άθροισμα του nugget και του $C(0)$. Το nugget ορίζεται ως η διασπορά μεταξύ δύο δειγμάτων των οποίων η απόσταση είναι μηδενική. Κανονικά, η διασπορά αυτή πρέπει να είναι μηδέν, κάτι όμως που δεν συμβαίνει στη φύση και οφείλεται κυρίως σε σφάλματα κατά τη δειγματοληψία και στην ετερογένεια της

μεταβλητής που δεν είναι δυνατό να εντοπιστεί στην κλίμακα που γίνεται η δειγματοληψία. Οι μικροδιαφορές αυτές είναι ένας τύπος λευκού θορύβου (white noise) και ισχύει ότι $C(t_1, t_2) = 0$, ($t_1 \neq t_2$) [13].

3.2 Καθορισμός του βαριογράμματος

Έχουμε ορίσει το ημιβαριόγραμμα για την περίπτωση που η μέση τιμή είναι σταθερή, ως

$$\gamma(\mathbf{h}) = \frac{1}{2} E \left\{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \right\}.$$

Για να εκτιμήσουμε το ημιβαριόγραμμα χρησιμοποιούμε απλά τα μετρημένα σημεία Z_i και υποθέτουμε εργοδικότητα στις αυξήσεις (increments), δηλαδή ότι οι χωρικοί μέσοι όροι μπορούν να χρησιμοποιηθούν για την εκτίμηση των μέσων τιμών σε όλο το σύνολο των πραγματοποιήσεων.

Καταρχήν ορίζουμε ένα συγκεκριμένο αριθμό κλάσεων απόστασης μεταξύ των μετρημένων σημείων, π.χ.

$$0 < d_1 < 1 \text{ km}, \quad 1 < d_2 < 2 \text{ km}, \quad 2 < d_3 < 3 \text{ km}, \quad 3 < d_4 < 5 \text{ km}, \quad 5 \text{ km} < d_5.$$

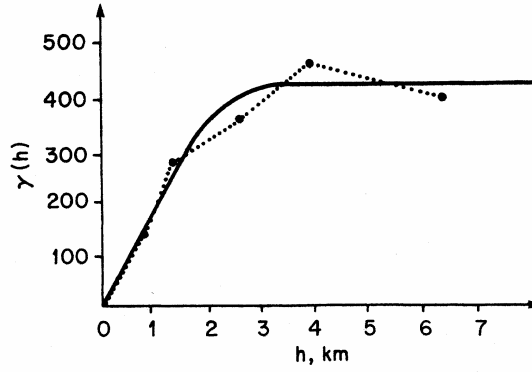
Στη συνέχεια, παίρνουμε όλα τα δυνατά ζεύγη σημείων i και j για κάθε κλάση απόστασης, υπολογίζουμε:

- (1) Το πλήθος των σημείων που υπάρχουν σε κάθε κλάση.
- (2) Την μέση απόσταση σε κάθε κλάση
- (3) Την μέση τετραγωνική αύξηση $\frac{1}{2}(Z_i - Z_j)^2$.

Με ένα σύνολο από 50 μετρήσεις, για παράδειγμα, θα έχουμε:

Κλάση	d_1	d_2	d_3	d_4	d_5
Πλήθος στοιχείων	500	350	250	100	25
Μέση Απόσταση	0.7	1.3	2.4	3.8	6.2
Μέση $\frac{1}{2}(Z_i - Z_j)^2$	130	275	350	570	400

Σημειώνουμε ότι το πλήθος των ζευγών που μπορεί να σχηματιστούν από ένα σύνολο n σημείων είναι $n(n-1)/2$, για 50 σημεία αυτό είναι 1225 ζεύγη. Όμως γενικά δεν είναι κανονικά κατανομημένα. Υπάρχουν περισσότερα ζεύγη στις κοντινές αποστάσεις από ότι στις μακρινές. Το ημιβαριόγραμμα γίνεται ολοένα και περισσότερο αβέβαιο καθώς το $|\mathbf{h}|$ αυξάνει. Στις μεγάλες αποστάσεις ορισμένα σημεία μπορεί να παίζουν κυρίαρχο ρόλο και εισάγουν σφάλματα στην εκτίμηση. Μπορεί να είναι απαραίτητο να αφαιρεθούν ορισμένες μετρήσεις κατά τον υπολογισμό του ημιβαριογράμματος.



Σχήμα 5. Απεικόνιση του πειραματικού ημιβαριογράμματος (σημεία) και το μοντελο ημιβαριογράμματος (συνεχής καμπύλη).

Στο Σχήμα 5 παρουσιάζεται ένα παράδειγμα πειραματικού ημιβαριογράμματος και η συνάρτηση ημιβαριογράμματος, δηλαδή το μαθηματικό μοντέλο που προσαρμόζεται στο πειραματικό ημιβαριογράμμα. Οι βασικές διαφορές μεταξύ της συνάρτησης ημιβαριογράμματος και του πειραματικού ημιβαριογράμματος είναι οι εξής:

- το πειραματικό ημιβαριογράμμα είναι *διακριτή* συνάρτηση, ενώ
- το μοντελοποιημένο ημιβαριογράμμα χρησιμοποιεί μαθηματικές εξισώσεις για να περιγράψει τη χωρική μεταβλητότητα του πειραματικού και χρησιμοποιείται από τον αλγόριθμο του Kriging.

Ο αριθμός των μαθηματικών συναρτήσεων που μπορούν να βαθμονομηθούν πάνω σε πειραματικά ημιβαριογράμματα, είναι περιορισμένος, [6]. Αυτό συμβαίνει επειδή οι εξισώσεις του Kriging για να έχουν μοναδική και σταθερή λύση πρέπει ο πίνακας \mathbf{A} των συμμεταβλητοτήτων

$$\mathbf{A} = \begin{bmatrix} \tilde{C}_{00} & \tilde{C}_{01} & \dots & \tilde{C}_{0n} \\ \tilde{C}_{10} & \tilde{C}_{11} & \dots & \tilde{C}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{C}_{n0} & \tilde{C}_{n1} & \dots & \tilde{C}_{nn} \end{bmatrix}$$

να είναι θετικά ορισμένος.

Ένας συμμετρικός $n \times n$ πίνακας \mathbf{A} είναι *θετικά ορισμένος*, ($|\mathbf{A}| > 0$), αν το γινόμενο $\mathbf{X}^t \mathbf{A} \mathbf{X}$ (όπου ο υπερυψωμένος δείκτης t συμβολίζει την πράξη της αναστροφής πινάκων) είναι μεγαλύτερο του μηδενός, δηλαδή: $\mathbf{X}^t \mathbf{A} \mathbf{X} > 0$, για κάθε $\mathbf{X} = [x_1, \dots, x_n]^T$ (όπου τα x_j δεν είναι όλα μηδενικά), ο \mathbf{A} είναι θετικά ορισμένος εάν όλες οι ορίζουσες των υποπινάκων του είναι μεγαλύτερες του μηδέν: $|A_j| > 0$, $i = (1, \dots, n)$. Μη θετικά ορισμένες συναρτήσεις οδηγούν σε πίνακες που δεν έχουν λύσεις ή είναι μαθηματικά ασταθείς.

Οι συναρτήσεις που χρησιμοποιούνται είναι περιορισμένες σε αριθμό αλλά αρκετές για να περιγράψουν ικανοποιητικά το σύνολο των περιπτώσεων που μπορούν να

συναντηθούν στην πράξη. Η επιλογή της κατάλληλης συνάρτησης εξαρτάται από το σχήμα του πειραματικού ημιβαριογράμματος και από τον τύπο των αρχικών δεδομένων.

3.3 Είδη Ημιβαριογραμμάτων

Όμως, έχουν ήδη δει ότι δεν μπορούν όλες οι συναρτήσεις να είναι ημιβαριογράμματα. Σε μια κλάση αποδεκτών αναλυτικών συναρτήσεων επιλέγουμε μια δεδομένη μορφή και προσαρμόζουμε τις παραμέτρους της συνάρτησης αυτής στα σημεία παρατήρησης.

Οι κύριες μορφές ημιβαριογραμμάτων που χρησιμοποιούνται είναι: το γραμμικό, στο $|\mathbf{h}|^\lambda$, $\lambda < 2$, το σφαιρικό, το εκθετικό το Gaussian και το κυβικό. Οι μορφές και οι εξισώσεις των ημιβαριογραμμάτων αυτών δίνονται στο Σχήμα 9 (όπου με h συμβολίζεται $h = |\mathbf{h}|$).

Ένα μεγάλο πρόβλημα που ανακύπτει με την χρήση των συναρτήσεων αυτών σχετίζεται κυρίως με τον ορισμό της απόστασης συσχέτισης α . Η απόσταση αυτή στο σφαιρικό ημιβαριόγραμμα είναι ίση με την μέγιστη απόσταση στην οποία τα δείγματα μπορούν να συσχετιστούν μεταξύ τους.

Στο εκθετικό μοντέλο

$$\gamma(\mathbf{h}) = 1 - \exp\left(\frac{-|\mathbf{h}|}{a}\right),$$

η πραγματική απόσταση συσχέτισης α είναι ίση με τρεις φορές το a , καθώς η συνάρτηση τείνει ασυμπτωτικά για να προσεγγίσει την μονάδα στο άπειρο, για αυτό το λόγο το a υπολογίζεται προσεγγιστικά ως η απόσταση εκείνη στην οποία το ημιβαριόγραμμα φτάνει το 95% της διασποράς [1], δηλ. $\alpha' = 3a$, όπου $\gamma(\alpha') = 1 - e^{-3} = 0.95 \approx 1$.

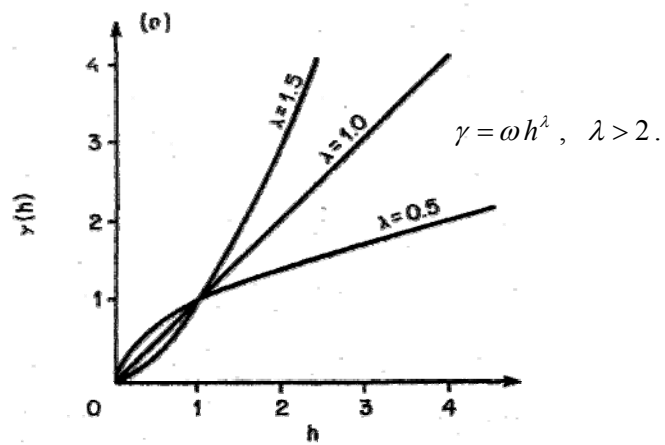
Η κυριότερη διαφορά μεταξύ του **σφαιρικού** και του εκθετικού μοντέλου είναι η απόσταση στην οποία οι εφαπτόμενες τους στο σημείο 0 τέμνονται με την οροφή (sill). Για το σφαιρικό $|\mathbf{h}| = 2a/3$ δηλ. δύο τρίτα της απόστασης α , ενώ για το εκθετικό $|\mathbf{h}| = a = \alpha'/3$, δηλ. το ένα τρίτο της πραγματικής απόστασης α' .

Ομοίως με το εκθετικό μοντέλο, στο **Gaussian** μοντέλο που περιγράφεται από τη σχέση

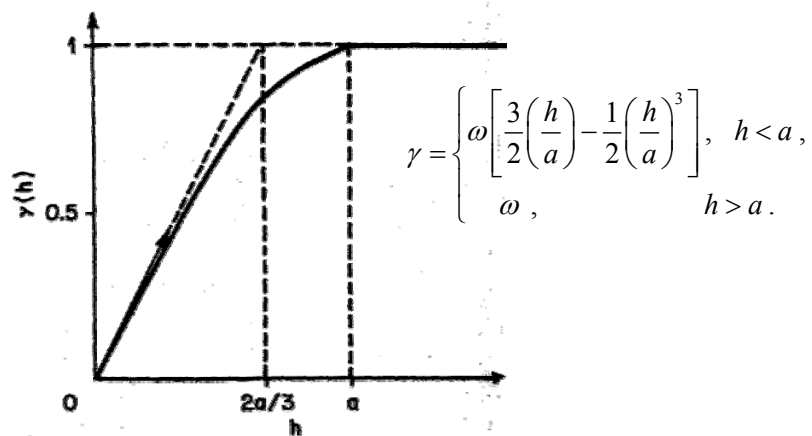
$$\gamma(\mathbf{h}) = 1 - \exp\left(\frac{-|\mathbf{h}|^2}{a^2}\right),$$

η μέγιστη απόσταση α' βρίσκεται από την προσεγγιστική σχέση $\alpha' = \sqrt{3}a$, όπου $\gamma(\alpha') = 0.95 \approx 1$.

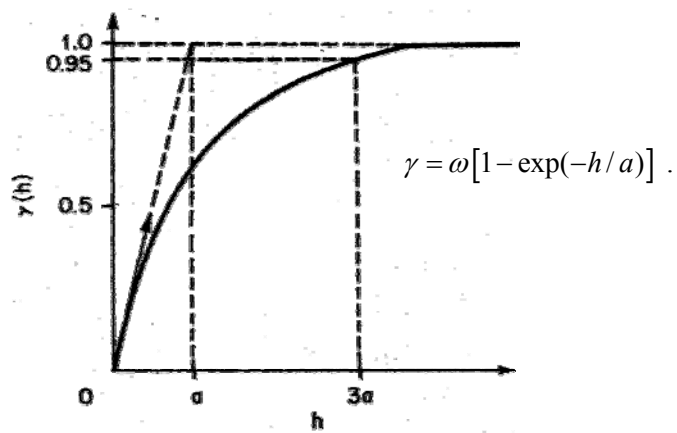
Για την σωστή επιλογή του μοντέλου του ημιβαριογράμματος που περιγράφει ακριβέστερα τις πειραματικές παρατηρήσεις, χρησιμοποιείται η διαδικασία της αυτοεπικύρωσης (cross-validation) που περιγράφεται σε επόμενο κεφάλαιο. Η πρόβλεψη άλλωστε της συμπεριφοράς μεταβλητών σε άγνωστες περιοχές δεν είναι απλή διαδικασία.



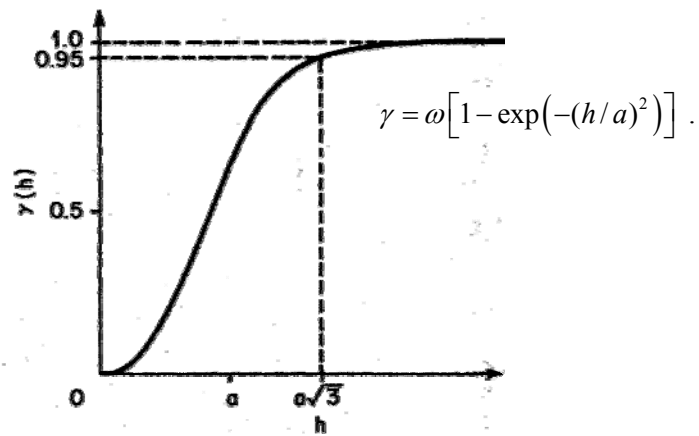
(Α) Γραμμικό ημιβαριόγραμμα ως προς $|h|^\lambda, \lambda < 2.$



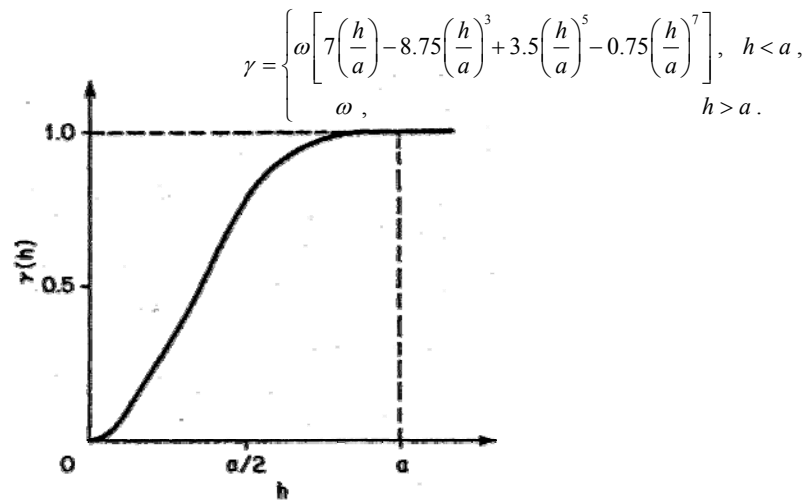
(Β) Σφαιρικό ημιβαριόγραμμα



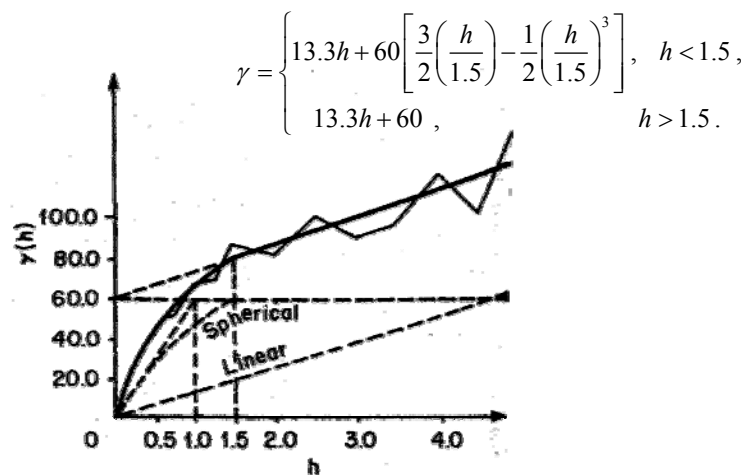
(Γ) Εκθετικό ημιβαριόγραμμα



(Δ) Gaussian ημιβαριόγραμμα



(Ε) Κυβικό ημιβαριόγραμμα



Σχήμα 9. Παράδειγμα προσαρμογής σε ένα «γραμμικό και σφαιρικό» μοντέλο.

3.4 Συμπεριφορά του ημιβαριογράμματος για μεγάλες αποστάσεις

Σημειώνουμε ότι ένα μη φραγμένο ημιβαριόγραμμα, π.χ. ένα γραμμικό (βλ. Σχήμα 9α) υποδηλώνει ότι το πεδίο έχει άπειρη διασπορά και ότι δεν υπάρχει η συνάρτηση συνδιασποράς: η υπόθεση του ενδογενούς είναι η μόνη αποδεκτή στην περίπτωση αυτή. Αλλά αν το ημιβαριόγραμμα φτάνει κάποια οροφή, κάποιο άνω φράγμα, όπως στο παράδειγμα του Σχήματος 9β, τότε η συνάρτηση συνδιασποράς υπάρχει για το υπό εξέταση φαινόμενο.

3.5 Συμπεριφορά του ημιβαριογράμματος κοντά στο μηδέν

Θεωρητικά, για $\mathbf{h} = 0$, $\gamma(\mathbf{h}) = 0$ ανεξάρτητα από το ημιβαριόγραμμα. Όμως, πολύ συχνά τα ημιβαριόγραμμα εμφανίζουν, όπως έχουμε δει και προηγουμένως, ένα «πήδημα» στην αρχή το λεγόμενο nugget. Αυτό το πήδημα στο μηδέν ονομάζεται «φαινόμενο ψήγματος» καθώς εντοπίστηκε στις βιομηχανίες ορυχείων. Πράγματι, αν ένας πυρήνας (core) περιέχει ένα ψήγμα μετάλλου, η συγκέντρωση θα είναι πολύ υψηλή, τέτοια ώστε γειτονικοί πυρήνες ακόμα και με υψηλή συγκέντρωση δεν θα μπορούν να τη φτάσουν. Υπάρχει δηλαδή κάποιου είδους αιρετική συμπεριφορά.

Τέτοια συμπεριφορά συναντάται πολύ συχνά όταν αναλύονται δεδομένα (π.χ. μεταφερσιμότητας). Για να τη λάβει κανείς υπόψιν του προσθέτει την ποσότητα C στο ημιβαριόγραμμα που προσαρμόζεται στα δεδομένα, έτσι ώστε η αρχική τιμή του να είναι το C :

$$\gamma(\mathbf{h}) = C[1 - \delta(\mathbf{h})] + \gamma'(\mathbf{h}) ,$$

όπου $\delta(\mathbf{h})$ είναι το δ του Kronecker ($\delta = 1$ αν $\mathbf{h} = 0$, $\delta = 0$ αν $\mathbf{h} \neq 0$) και $\gamma'(\mathbf{h})$ είναι το ημιβαριόγραμμα που προσαρμόζεται στα δεδομένα με αρχική τιμή C .

Αυτό το φαινόμενο nugget μπορεί επίσης να οφείλεται σε σφάλματα των μετρήσεων ή στο γεγονός ότι τα δεδομένα δεν έχουν συλλεχθεί με αρκετά μικρό διάστημα ώστε να δείχνουν την συνεχή συμπεριφορά του φαινομένου που υπάρχει.

Ένα οριζόντιο ημιβαριόγραμμα, π.χ. $\gamma(\mathbf{h}) = C, \forall \mathbf{h} > 0$, χαρακτηρίζεται ως ημιβαριόγραμμα με καθαρά ψηγματική (nugget) συμπεριφορά και εκφράζει ένα καθαρά τυχαίο φαινόμενο χωρίς χωρική δομή.

Όταν το ημιβαριόγραμμα έχει μια ψηγματική μορφή, το kriging εξακολουθεί να είναι μια ακριβής παρεμβολή, αλλά η εκτίμηση είναι ασυνεχής στα σημεία μέτρησης, π.χ. αν \mathbf{x}_k είναι ένα σημείο μέτρησης και $Z_k = Z(\mathbf{x}_k)$, τότε $Z^*(\mathbf{x}_k) = Z_k$, αλλά $Z^*(\mathbf{x}_k + d\mathbf{x}) \neq Z_k$ ακόμα και όταν $d\mathbf{x} \rightarrow 0$. Όμως η εκτίμηση είναι συνεχής παντού αλλού.

Έστω C η τιμή του ψήγματος (nugget). Ας γράψουμε το ημιβαριόγραμμα $\gamma(\mathbf{h}) = C(1 - \delta) + \gamma'$.

Μια εναλλακτική που δίνει ακριβώς το ίδιο αποτέλεσμα είναι να αφαιρέσουμε την τιμή του ψήγματος από το ημιβαριόγραμμα, θεωρώντας ότι αντιπροσωπεύει μόνο σφάλματα

των μετρήσεων, και να χρησιμοποιήσουμε μια αβεβαιότητα $\sigma_i^2 = C$ στην διαγώνιο του συστήματος του kriging, δεδομένου ότι και ο πίνακας του kriging και το δεξιό μέρος βασίζονται στο γ' αντί στο γ . Μπορεί εύκολα να δει κανείς ότι αυτό το νέο σύστημα kriging είναι ταυτόσιμο με το κλασσικό σύστημα του kriging, προσθέτοντας την τελευταία γραμμή του συστήματος πολλαπλασιασμένη επί C σε όλες τις υπόλοιπες.

Όμως, είναι επίσης δυνατόν να γράψει κανείς τις εξισώσεις του kriging χρησιμοποιώντας το γ στο αριστερό μέλος (με μηδέν στην διαγώνιο) και το γ' στο δεξιό μέλος. Αυτό αντιστοιχεί στην περίπτωση όπου το γ' θεωρείται το πραγματικό ημιβαριόγραμμα του φαινομένου (χωρίς σφάλματα μετρήσεων) και το γ το ημιβαριόγραμμα των δεδομένων με θόρυβο. Σε αυτήν την περίπτωση, η εκτίμηση Z_0^* είναι συνεχής παντού αλλά δεν είναι πλέον ακριβής στα σημεία μέτρησης. Η διασπορά του σφάλματος της εκτίμησης είναι επίσης μικρότερη σε σχέση με το κανονικό σύστημα και δίνεται από τη σχέση

$$\text{var}(Z_0^* - Z_0) = \sum_i \lambda_i^i \gamma'(\mathbf{x}_i - \mathbf{x}_0) + \mu + C.$$

Αν έχουμε ένα οριζόντιο ημιβαριόγραμμα (με καθαρά ψηγματική μορφή) τότε μπορεί ναδειχθεί ότι $\lambda^i = n^{-1}$, $\forall i$, όπου n είναι το πλήθος των σημείων μέτρησης. Η εκτίμηση Z_0^* είναι τότε σταθερή πάνω σε όλο το χωρίο και ίση με τον μέσο όρο όλων των μετρήσεων.

3.6 Ανισοτροπία στο ημιβαριόγραμμα

Μέχρι τώρα έχουμε θεωρήσει ότι η συνάρτηση της συνδιασποράς εξαρτάται μόνο από την απόσταση h . Αυτό ονομάζεται ισοτροπία, όταν δηλαδή

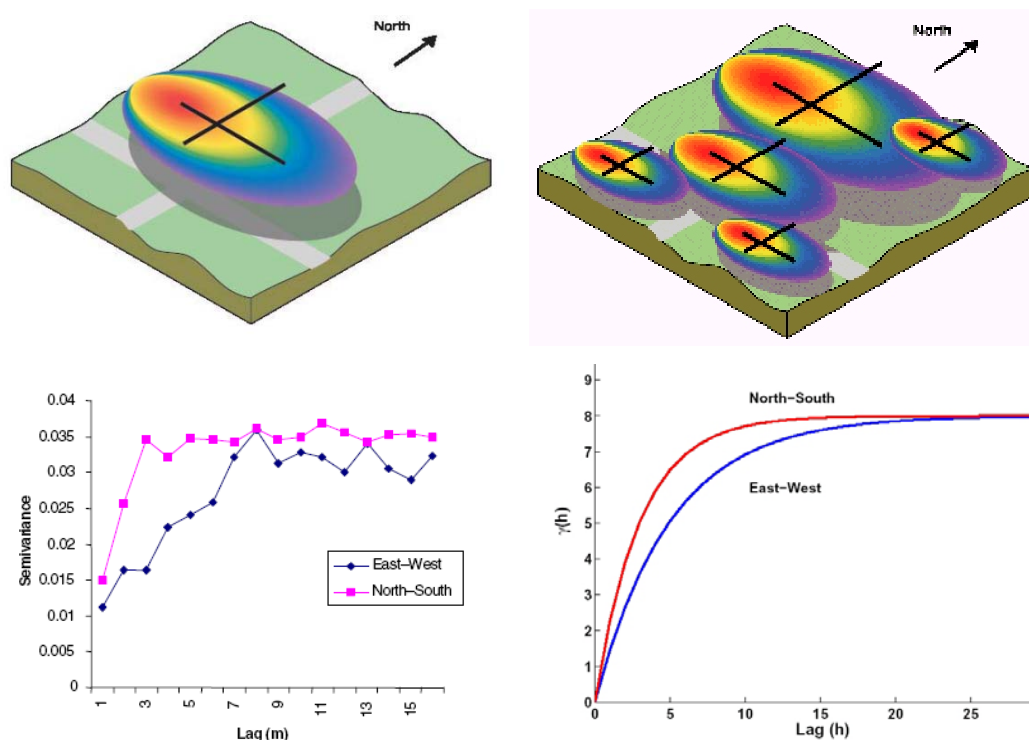
$$E\{[Z(x) - m][Z(y) - m]\} = C(|x - y|).$$

Κατά την γεωστατιστική ανάλυση κατασκευάζεται σε αρχικό στάδιο ένα ημιβαριόγραμμα του οποίου η ανοχή στην διεύθυνση είναι τόσο μεγάλη, ώστε η κατεύθυνση οποιουδήποτε διανύσματος που ενώνει δυο πειραματικά σημεία να μην είναι σημαντική. Η χρήση τέτοιων ημιβαριογραμμάτων συνεπάγεται άμεσα την ύπαρξη ισοτροπίας. Ακόμα και όταν αυτό δεν ισχύει η κατασκευή τέτοιων ημιβαριογραμμάτων βοηθά στον προσδιορισμό χρήσιμων παραμέτρων.

Συχνά όμως για παρατηρείται το φαινόμενο η συνάρτηση συνδιασποράς και συνεπώς και το ημιβαριόγραμμα να μην εξαρτώνται μόνο από την απόσταση, αλλά και από την διεύθυνση. Το φαινόμενο αυτό καλείται ανισοτροπία και σημαίνει ότι υπάρχει μια επίδραση της κατεύθυνσης στην χωρική αυτοσυσχέτιση των δεδομένων, δηλαδή ότι η ίδια τιμή συνδιασποράς μπορεί να προκύψει σε διαφορετικές αποστάσεις αν θεωρήσουμε διαφορετικές διευθύνσεις.

Σε τέτοιες περιπτώσεις η συνάρτηση του ημιβαριογράμματος θα μεταβάλλεται ανάλογα με την κατεύθυνση, βλ. Σχήμα 10. Η μελέτη λοιπόν του ημιβαριογράμματος σε διάφορες

κατευθύνσεις είναι δυνατόν να οδηγήσει στον εντοπισμό πιθανής ανισοτροπίας μιας χωρικής τυχαίας μεταβλητής. Είναι χρήσιμο επομένως να υπολογίζουμε το ημιβαριόγραμμα θεωρώντας ότι το $\gamma(\mathbf{h})$ είναι επίσης συνάρτηση της κατεύθυνσης του διανύσματος της απόστασης \mathbf{h} . Φυσικά, αυτό απαιτεί περισσότερα σημεία δεδομένων για να έχει νόημα. Θα μπορούσαμε, για παράδειγμα να χρησιμοποιήσουμε τέσσερις ή οκτώ κλάσεις κατεύθυνσης και να σχεδιάσουμε κάθε ημιβαριόγραμμα χωριστά όπως στο Σχήμα 10.



Σχήμα 10. Παραδείγματα φαινομένου ανισοτροπίας και ανισότροπα ημιβαριογράμματα.

Γενικά τα ημιβαριογράμματα δεν εμφανίζουν τέτοια ανισοτροπία όπως στο Σχήμα 10. Αν όμως η ανισοτροπία είναι σημαντική, τότε (1) αυτό μπορεί να είναι σημάδι ότι η παραδοχή της στασιμότητας (ή ακόμα και η παραδοχή του ενδογενούς) δεν ισχύουν, οπότε οι εξισώσεις είναι διαφορετικές, ή (2) αν η ενδογενής υπόθεση ισχύει, αυτή η ανισοτροπία μπορεί να ελαχιστοποιηθεί με κατάλληλο γραμμικό μετασχηματισμό του συστήματος συντεταγμένων. Αυτό μας επιτρέπει να χρησιμοποιήσουμε το kriging στο νέο σύστημα.

Κεφάλαιο 4

ΕΙΔΗ ΜΕΘΟΔΩΝ KRIGING

Με την πάροδο των ετών έχουν αναπτυχθεί διάφορες παραλλαγές της μεθόδου Kriging, [4], [20]. Στο κεφάλαιο αυτό θα αναφερθούμε στις σημαντικότερες από αυτές, οι οποίες είναι:

- Ordinary (κανονικό) Kriging
- Simple (απλό) Kriging
- Universal (καθολικό) Kriging
- Indicator (ενδεικτικό) Kriging

οι οποίες θα παρουσιαστούν συνοπτικά στην συνέχεια.

Επίσης, σε ορισμένες περιπτώσεις είναι δυνατόν να έχουμε δύο ή και περισσότερες μεταβλητές οι οποίες συσχετίζονται μεταξύ τους. Τότε οι μεταβλητές αυτές μπορούν να εκτιμηθούν μαζί με την μέθοδο του

- Co-Kriging

που είναι μια πιο σύνθετη παραλλαγή της μεθόδου Kriging ιδιαίτερα χρήσιμη στην υδρολογία.

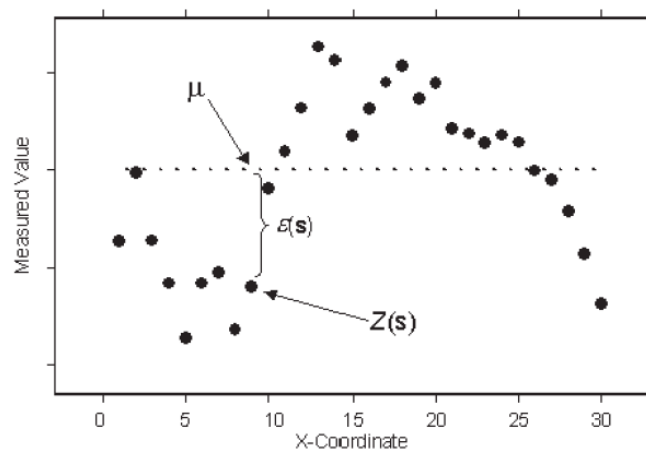
Οι παραπάνω παραλλαγές της μεθόδου Kriging παρουσιάζονται στο κεφάλαιο αυτό.

4.1 Ordinary kriging

Η μέθοδος του Ordinary kriging βασίζεται στο ημιβαριόγραμμα για πρόβλεψη της τιμής μιας μεταβλητής σε κάποιο σημείο, χωρίς να είναι γνωστή η μέση τιμή των δειγμάτων. Η μέση τιμή θεωρείται σταθερή μέσα στην περιοχή εκτίμησης.

Το Ordinary Kriging προϋποθέτει ότι το δείγμα ακολουθεί την κανονική κατανομή, ενώ λόγω της συνθήκης αμεροληψίας το άθροισμα των σταθμικών συντελεστών (βάρη) που υπεισέρχονται στην εξίσωση παρεμβολής είναι ίσο με την μονάδα

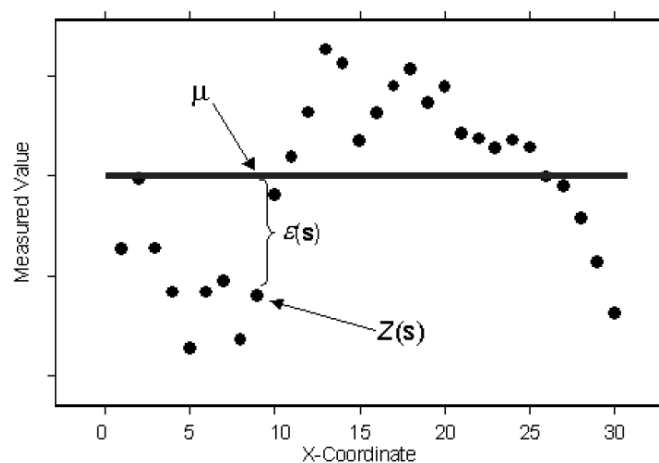
$$\sum_{i=1}^n \lambda_i = 1$$



4.2 Simple kriging

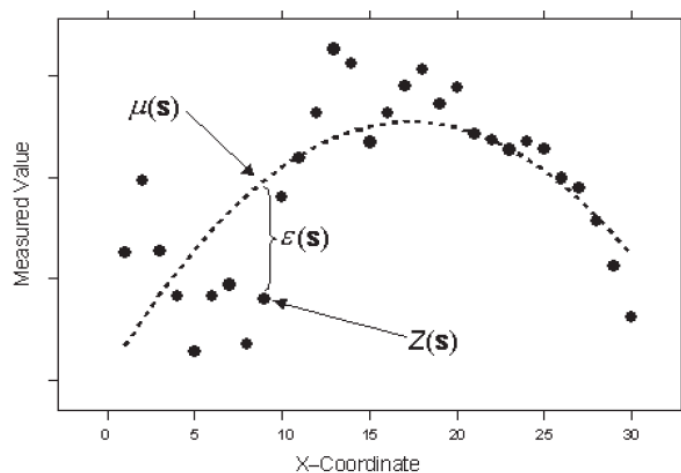
Το λεγόμενο Απλό (Simple) Kriging αποτελεί την πλέον κλασσική μέθοδο kriging. Όλες οι άλλες μέθοδοι kriging αποτελούν παραλλαγές αυτής.

Στο Απλό kriging υπολογίζεται η τιμή της άγνωστης μεταβλητής προϋποθέτοντας ότι είναι γνωστή η μέση τιμή των δειγμάτων.



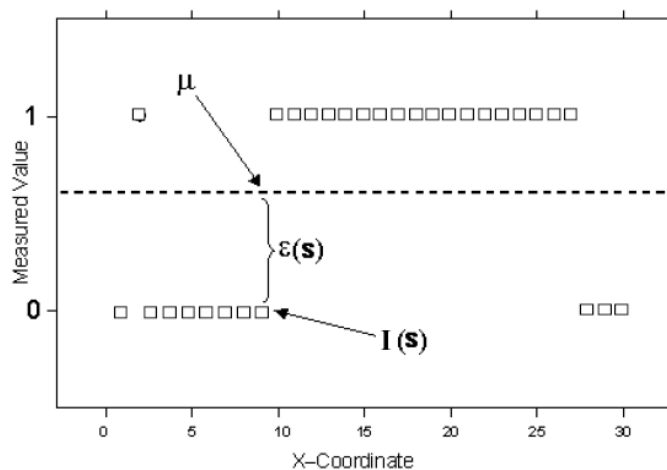
4.3 Universal kriging

Η μέθοδος του Γενικευμένου (Universal) Kriging χρησιμοποιείται όταν τα δεδομένα παρουσιάζουν κάποια «τάση» (trend), όπως φαίνεται στο παρακάτω σχήμα..



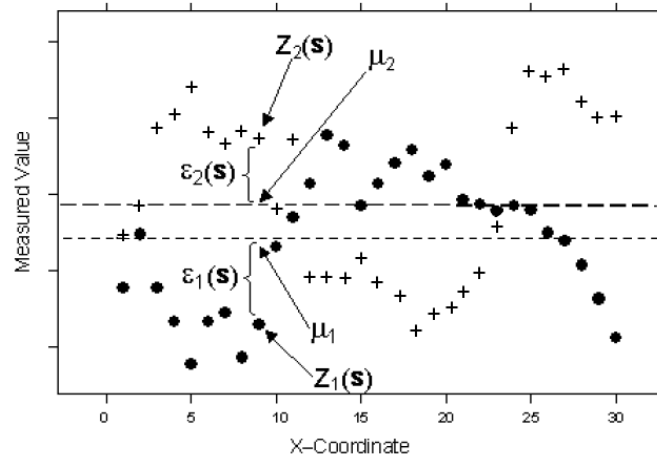
4.4 Indicator kriging

Το Ενδεικτικό (Indicator) Kriging μετασχηματίζει τα αρχικά δεδομένα χρησιμοποιώντας έναν ενδεικτικό (indicator) μετασχηματισμό σε δυαδικά. Αυτό σημαίνει ότι με βάση μια συγκεκριμένη τιμή (cut off), τα δεδομένα μεγαλύτερες τιμές εξισώνονται με μονάδα, ενώ αυτά με μικρότερες τιμές με μηδέν. Χρησιμοποιώντας δυαδικές μεταβλητές, το ενδεικτικό Kriging ακολουθεί την ίδια διαδικασία με το ordinary (κανονικό) kriging. Καθώς οι μεταβλητές είναι δυαδικές (0 ή 1), οι εκτιμήσεις από το ενδεικτικό kriging θα είναι μεταξύ 0 και 1 και μπορούν να θεωρηθούν σαν πιθανότητες της μεταβλητής να είναι 1 ή να ανοίκει στην κλάση του 1. Είναι φανερό ότι το ενδεικτικό kriging είναι κατάλληλο για δυαδικά δεδομένα (π.χ. αν έχουμε ένα δείγμα δεδομένων σχετικά με το αν ένα σημείο είναι δάσος ή κατοικημένη περιοχή).



4.5 Co-kriging

Το co-kriging είναι μια τεχνική εκτίμησης χρήσιμη όταν δύο (ή περισσότερες) μεταβλητές οι οποίες συσχετίζονται μεταξύ τους καταμετρούνται στο πεδίο και μπορούν να εκτιμηθούν μαζί. Για παράδειγμα, αν σε έναν υδροφορέα η συγκέντρωση στο νερό διαφόρων μετάλλων είναι συσχετισμένη, τότε είναι δυνατόν να εκτιμήσουμε σε μία τοποθεσία το ποσό του μετάλλου Z_1^* σε μετρήσεις όχι μόνο του Z_1 , αλλά και των μετάλλων Z_2 ή Z_3 . Ένα άλλο παράδειγμα είναι αυτό της εκτίμησης της μεταφερισιμότητας σε έναν υδροφορέα βασιζόμενοι όχι μόνο στις μετρήσεις της μεταφερισιμότητας καθεαυτής, που έχουν με test άντλησης, αλλά βασιζόμενοι την ίδια στιγμή και σε μετρήσεις της ειδικής χωρητικότητας, που συσχετίζεται με την μεταφερισιμότητα. Το co-kriging είναι τότε μια πιο εξεζητημένη και ακριβής μέθοδος για την χρησιμοποίηση και των δύο ειδών των δεδομένων από την απλή παλινδρόμηση και το kriging με αβέβαια δεδομένα.



4.6.1 Οι εξισώσεις του co-kriging

Έστω ότι $Z_1(\mathbf{x})$ και $Z_2(\mathbf{x})$ είναι δύο περιοχωποιημένες (regionalized) μεταβλητές που συσχετίζονται. Η εκτίμηση του Z_1 (και αν είναι απαραίτητο και του Z_2) με το co-kriging δίνονται, [1], με μία βέλτιστη γραμμική αμερόληπτη εκτίμηση της μορφής

$$Z_1^*(\mathbf{x}_0) = \sum_j \lambda_1^j Z_1(\mathbf{x}_j) + \sum_{\ell=1} \lambda_2^\ell Z_2(\mathbf{x}_\ell) . \quad (19)$$

Εδώ, το $Z_1^*(\mathbf{x}_0)$ είναι η εκτίμηση του Z_1 στην τοποθεσία \mathbf{x}_0 , τα $Z_1(\mathbf{x}_j)$ είναι οι μετρημένες τιμές του Z_1 , $j = 1, \dots, n$, και τα $Z_2(\mathbf{x}_\ell)$ είναι οι μετρημένες τιμές του Z_2 , $\ell = 1, \dots, m$. (Σημειώνουμε ότι το Z_1 και το Z_2 δεν είναι απαραίτητο να έχουν μετρηθεί στις ίδιες τοποθεσίες, και ότι τα n και m μπορεί να είναι διαφορετικά). Στην εξίσωση (19), τα λ είναι τα βάρη του co-kriging. Όμως, καλό είναι να αλλάξουμε λίγο τον συμβολισμό σε σχέση με τα προηγούμενα: το λ_j^i είναι το βάρος του kriging για την μεταβλητή Z_j μετρημένη στην τοποθεσία \mathbf{x}_i , και παραλείπουμε τον δείκτη 0 που χρησιμοποιούσαμε νωρίτερα (λ_0^i) ο οποίος καταδείκνυε ότι αυτά τα βάρη αφορούν την εκτίμηση του Z στην

τοποθεσία \mathbf{x}_0 . Παρόλα αυτά είναι ξεκάθαρο ότι αυτά τα λ θα αλλάζουν για κάθε νέα τοποθεσία \mathbf{x}_0 στην οποία θα γίνεται εκτίμηση. Θα ήταν επίσης απαραίτητο να έχουμε έναν άλλο δείκτη για να καταδεικνύει ότι αναφέρονται στην εκτίμηση του Z_1^* , καθώς κανείς θα μπορούσε να θέλει να εκτιμήσει το Z_2^* από τις ίδιες μετρήσεις. Για να διατηρήσουμε τον συμβολισμό απλό θα θεωρήσουμε εδώ μόνο την εκτίμηση της μίας μεταβλητής Z .

Όμως, θα γενικεύσουμε την εξίσωση (19) για N μεταβλητές γράφοντας

$$Z_i^*(\mathbf{x}_0) = \sum_{a=1}^N \sum_{j=1}^{n_a} \lambda_a^j Z_a(\mathbf{x}_j) . \quad (20)$$

Όπως και στο απλό kriging, τα λ θα προσδιοριστούν γράφοντας τις δύο συνήθειες συνθήκες: αμεροληψίας και βελτιστοποίησης.

(α) *Αμερόληπτη εκτίμηση.* Θα θεωρήσουμε ότι $E(Z_a) = m_a$, $k = 1, \dots, N$ (το οποίο ισχύει και για την παραδοχή στασιμότητας και για την παραδοχή του ενδογενούς). Τότε, η συνθήκη $E(Z_i^*) = m_i$ μπορεί να γραφτεί σαν

$$m_i = \sum_a m_a \sum_j \lambda_a^j . \quad (21)$$

Αν όλες οι μεταβλητές έχουν διαφορετική αναμενόμενη τιμή, τότε η εκτίμηση του Z_i^* απαιτεί:

$$\sum_{j=1}^{n_i} \lambda_i^j = 1 , \quad \sum_{j=1}^{n_a} \lambda_a^j = 0 , \quad a \neq i . \quad (22)$$

Αλλά αν η μεταβλητή Z_k έχει την ίδια αναμενόμενη τιμή με την Z_i , μια περίπτωση που είναι αρκετά συχνή στην πράξη, τότε η εξίσωση (21) δίνει:

$$\sum_{j=1}^{n_i} \lambda_i^j + \sum_{\ell=1}^{n_k} \lambda_k^\ell = 1 ,$$

$$\sum_{j=1}^{n_a} \lambda_a^j = 0 , \quad a \neq i \text{ και } k .$$

(β) *Βελτιστοποίηση.* Όταν είχαμε εισάγει την συνθήκη βελτιστοποίησης για το απλό kriging με την παραδοχή ότι η διασπορά του σφάλματος της εκτίμησης $E\{[Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0)]^2\}$ είναι ελάχιστη, τότε είχαμε επίσης εισάγει τη συνδιασπορά ή το ημιβαριόγραμμα της υπό μελέτη μεταβλητής λόγω της ανάγκης υπολογισμού όρων όπως το $E[Z(\mathbf{x}_i)Z(\mathbf{x}_j)]$. Στο co-kriging, πρόσθετοι όροι όπως το $E[Z_a(\mathbf{x}_i)Z_b(\mathbf{x}_j)]$ θα πρέπει επίσης να υπολογιστούν. Συνεπώς, τώρα θα εισάγουμε μια νέα συνάρτηση για την ποσοτικοποίηση της χωρικής συσχέτισης της μεταβλητής Z_a με την μεταβλητή Z_b , σαν συνάρτηση της απόστασης \mathbf{h} μεταξύ των σημείων στα οποία η Z_a και η Z_b έχουν μετρηθεί. Για την περίπτωση της στασιμότητας, αυτή θα ονομάζεται συνάρτηση cross-covariance $C_{a\beta}(\mathbf{h})$ και στην περίπτωση της παραδοχής του ενδογενούς θα ονομάζεται cross-ημιβαριόγραμμα $\gamma_{a\beta}(\mathbf{h})$. Αυτά θα οριστούν με ακρίβεια στην συνέχεια.

Επίσης, στις εξισώσεις του co-kriging θα χρειαστούμε την συνήθη συνδιασπορά ή ημιβαριόγραμμα κάθε μεταβλητής Z_a , που θα συμβολίζουμε σαν $C_{aa}(\mathbf{h})$ ή $\gamma_{aa}(\mathbf{h})$.

Οι εξισώσεις του co-kriging μπορούν να γραφούν πολύ εύκολα με τις παραπάνω νέες συναρτήσεις. Αρχικά θα τις γράψουμε για την εκτίμηση του Z_1 από n μετρήσεις του και m μετρήσεις του Z_2 (π.χ. με την συνθήκη αμεροληψίας (22)). Αν οι Z_1 και Z_2 είναι και οι δύο στάσιμες $2^{\text{ης}}$ τάξης, το σύστημα του co-kriging μπορεί να γραφεί σαν

$$\sum_{j=1}^n \lambda_1^j C_{11}(\mathbf{x}_i - \mathbf{x}_j) + \sum_{\ell=1}^m \lambda_2^\ell C_{12}(\mathbf{x}_i - \mathbf{x}_\ell) - \mu_1 = C_{11}(\mathbf{x}_0 - \mathbf{x}_i) , \quad i=1, \dots, n , \quad (23)$$

$$\sum_{j=1}^n \lambda_1^j C_{21}(\mathbf{x}_k - \mathbf{x}_j) + \sum_{\ell=1}^m \lambda_2^\ell C_{22}(\mathbf{x}_k - \mathbf{x}_\ell) - \mu_2 = C_{12}(\mathbf{x}_0 - \mathbf{x}_k) , \quad k=1, \dots, m , \quad (24)$$

με

$$\sum_{\ell=1}^n \lambda_1^\ell = 1 , \quad \text{και} \quad \sum_{\ell=1}^m \lambda_2^\ell = 0 . \quad (25)$$

Η διασπορά του σφάλματος της εκτίμησης είναι:

$$\begin{aligned} \text{var}[Z_1^*(\mathbf{x}_0) - Z_1(\mathbf{x}_0)] &= E\left\{[Z_1^*(\mathbf{x}_0) - Z_1(\mathbf{x}_0)]^2\right\} \\ &= -\sum_{j=1}^n \lambda_1^j C_{11}(\mathbf{x}_0 - \mathbf{x}_j) - \sum_{\ell=1}^m \lambda_2^\ell C_{12}(\mathbf{x}_0 - \mathbf{x}_\ell) + \mu_1 . \end{aligned} \quad (26)$$

Αν οι Z_1 και Z_2 ικανοποιούν και οι δύο την παραδοχή του ενδογενούς, τότε το σύστημα του kriging γράφεται αντικαθιστώντας τα C_{ij} με $-\lambda_{ij}$ στις εξισώσεις (23)–(26). Αυτό μπορεί εύκολα να γενικευθεί για περισσότερες από δύο μεταβλητές: για παράδειγμα, το σύστημα του kriging για την εκτίμηση της ενδογενούς υπόθεσης του $Z_i^*(\mathbf{x}_0)$ από την εξίσωση (20) δίνεται από τη σχέση

$$\sum_{a=1}^N \sum_{\ell=1}^{n_a} \lambda_a^\ell \gamma_{\beta a}(\mathbf{x}_j - \mathbf{x}_\ell) + \mu_\beta = \gamma_{i\beta}(\mathbf{x}_0 - \mathbf{x}_j) ,$$

για $j=1, \dots, n_\beta$ και $\beta=1, \dots, N$, και

$$\sum_{\ell=1}^{n_a} \lambda_a^\ell = \begin{cases} 1 & \alpha \nu \quad a=i \\ 0 & \alpha \nu \quad a \neq i \end{cases} \quad a=1, \dots, N .$$

Η διασπορά του σφάλματος της εκτίμησης είναι:

$$\text{var}[Z_i^*(\mathbf{x}_0) - Z_i(\mathbf{x}_0)] = \sum_{a=1}^N \sum_{\ell=1}^{n_a} \lambda_a^\ell \gamma_{ia}(\mathbf{x}_0 - \mathbf{x}_\ell) + \mu_i .$$

Οι παραπάνω σχέσεις μπορούν επίσης να γενικευθούν για kriging του μέσου όρου πάνω σε ένα κάνναβο (mesh) ή για αβέβαια δεδομένα. Μια σχέση με μορφή πινάκων για το co-kriging έχει προταθεί από τον Myers (1982, 1983, 1984).

4.6.2 Cross-Συνδιασπορές και Cross-Βαριογράμματα

Στην υπόθεση ότι οι μεταβλητές $Z_1(\mathbf{x})$ και $Z_2(\mathbf{x})$ είναι και οι δύο στάσιμες δεύτερης τάξης, η cross-συνδιασπορά $C_{12}(\mathbf{h})$ του Z_1 και Z_2 ορίζεται σαν

$$C_{12}(\mathbf{h}) = E\{[Z_1(\mathbf{x}) - m_1][Z_2(\mathbf{x} + \mathbf{h}) - m_2]\} ,$$

όπου $m_1 = E(Z_1)$ και $m_2 = E(Z_2)$.

Μπορεί να αποδειχθεί ότι γενικά, $C_{12}(\mathbf{h}) \neq C_{21}(\mathbf{h})$, αλλά ότι $C_{12}(\mathbf{h}) = C_{21}(-\mathbf{h})$ (de Marsily, [1]).

Στην υπόθεση του ενδογενούς, το cross-ημιβαριόγραμμα ορίζεται σαν

$$\gamma_{12}(\mathbf{h}) = \frac{1}{2} E\{[Z_1(\mathbf{x} + \mathbf{h}) - Z_1(\mathbf{x})][Z_2(\mathbf{x} + \mathbf{h}) - Z_2(\mathbf{x})]\} . \quad (27)$$

Στην περίπτωση της στασιμότητας, όπου και τα δύο $C_{12}(\mathbf{h})$ και $\gamma_{12}(\mathbf{h})$ υπάρχουν,

$$\gamma_{12}(\mathbf{h}) = C_{12}(0) - \frac{1}{2}[C_{12}(\mathbf{h}) + C_{21}(\mathbf{h})] .$$

Συνεπώς, είναι φανερό ότι το ημιβαριόγραμμα είναι πάντα συμμετρικό:

$$\gamma_{12}(\mathbf{h}) = \gamma_{21}(\mathbf{h}) = \gamma_{12}(-\mathbf{h}) = \gamma_{21}(-\mathbf{h}) .$$

Επομένως, η χρήση της cross-συνδιασποράς δίνει/έχει περισσότερες δυνατότητες/πιθανότητες από την χρήση του cross-βαριογράμματος στην περίπτωση της στασιμότητας. Η πειραματική cross-συνδιασπορά ή το περιβαματικό cross-ημιβαριόγραμμα ορίζονται με ανάλογο τρόπο με αυτόν που χρησιμοποιήθηκε για τον καθορισμό των απλών Ημιβαριογραμμάτων νωρίτερα.

Η κύρια δυσκολία στην χρήση του co-kriging είναι ότι αυτές οι συναρτήσεις πρέπει να ικανοποιούν περιορισμούς και κριτήρια για να είναι αποδεκτές, όπως ακριβώς δείξαμε ότι οι συναρτήσεις συνδιασποράς πρέπει να είναι θετικά ορισμένες και ότι το μείον του βαριογράμματος πρέπει να είναι θετικά ορισμένο υπό συνθήκη.

Έστω G ένα άθροισμα με βάρη όλων των μετρήσεων όπως αυτό που χρησιμοποιείται στο co-kriging

$$\sum_a \sum_j \lambda_a^j Z_a(\mathbf{x}_j) ,$$

τότε

$$\text{var}(G) = \sum_a \sum_\beta \sum_j \sum_\ell \lambda_a^j \lambda_\beta^\ell C_{a\beta}(\mathbf{x}_j - \mathbf{x}_\ell) ,$$

ή

$$\text{var}(G) = - \sum_a \sum_\beta \sum_j \sum_\ell \lambda_a^j \lambda_\beta^\ell \gamma_{a\beta}(\mathbf{x}_j - \mathbf{x}_\ell) ,$$

με $\sum_j \lambda_a^j = 0$ για το τελευταίο, όπως στην παράγραφο

Η θεώρηση ότι το $\text{var}(G)$ και στις δύο περιπτώσεις είναι θετικό για όλα τα λ και για κάθε σημείο \mathbf{x} δίνει τους περιορισμούς που πρέπει να ικανοποιούνται την ίδια στιγμή από τις συνδιασπορές, και τις cross-συνδιασπορές, και από τα βαριογράμματα και τα cross-

βαριογράμματα που χρησιμοποιούνται σε ένα σύστημα co-kriging. Δεν είναι καθόλου εύκολο να επιλέξει κανείς συναρτήσεις που να ικανοποιούν αυτούς τους περιορισμούς.

Το Γραμμικό Μοντέλο: Μια μέθοδος είναι να υποθέσουμε ότι οι N μεταβλητές Z_a μπορούν να θεωρηθούν σαν ένας γραμμικός συνδυασμός των M υποθετικών μεταβλητών Z'_i , οι οποίες θα είναι ασυσχέτιστες:

$$Z_a(\mathbf{x}) = \sum_{i=1}^M a_{ai} Z'_i(\mathbf{x}) \ , \quad \text{για } a=1, \dots, N \ .$$

Ακόμα και αν οι μεταβλητές Z'_i είναι ανεξάρτητες (π.χ., οι cross-συνδιασπορά ή το cross-ημιβαριόγραμμα είναι μηδέν) οι μεταβλητές Z_a θα είναι συσχετισμένες. Έστω $C'_i(\mathbf{h})$ ή $\gamma'_i(\mathbf{h})$ οι συνδιασπορές ή τα βαριογράμματα των μεταβλητών Z'_i . Τότε, είναι εύκολο να δείχθει ότι οι συνδιασπορές και cross-συνδιασπορές (ή τα βαριογράμματα και τα cross-βαριογράμματα) των μεταβλητών Z_a είναι

$$C_{a\beta}(\mathbf{h}) = \sum_{i=1}^M b_{a\beta}^i C'_i(\mathbf{h}) \ ,$$

ή

$$\gamma_{a\beta}(\mathbf{h}) = \sum_{i=1}^M b_{a\beta}^i \gamma'_i(\mathbf{h}) \ ,$$

όπου $b_{a\beta}^i = a_{ai} a_{\beta i}$ (επομένως $b_{a\beta}^i = b_{\beta a}^i$).

Τότε ο περιορισμός για τις συνδιασπορές και τα βαριογράμματα γίνεται απλά ότι οι πίνακες $[b_{a\beta}^i]$ για $i=1, \dots, M$ πρέπει να είναι θετικά ορισμένοι. Αυτό μπορεί να επιτευχθεί θεωρώντας ότι οι καθρέπτες (mirrors) δεύτερης τάξης των πινάκων αυτών είναι θετικά ορισμένοι:

$$\begin{vmatrix} b_{aa}^i & b_{a\beta}^i \\ b_{\beta a}^i & b_{\beta\beta}^i \end{vmatrix} > 0$$

για κάθε $i=1, \dots, M$ και για $a=1, \dots, N$ και $\beta=1, \dots, N$.

Αυτό συνεπάγεται τη συνθήκη

$$b_{aa}^i > 0, \quad b_{\beta\beta}^i > 0, \quad |b_{a\beta}^i| = |b_{\beta a}^i| < \sqrt{b_{aa}^i b_{\beta\beta}^i} \ . \quad (28)$$

Για παράδειγμα, έστω ότι έχουμε δύο μεταβλητές Z_1 και Z_2 , και τα βαριογράμματα των Z_1 , Z_2 και $Z_1 Z_2$, έχουν όλα την ίδια μορφή, την οποία θα ονομάζουμε γ . Αυτό σημαίνει ότι $M=1$ και $N=2$. Επομένως

$$\begin{aligned} \gamma_{11}(\mathbf{h}) &= b_1 \gamma(\mathbf{h}) \ , & \text{με } b_1 &> 0 \ , \\ \gamma_{22}(\mathbf{h}) &= b_2 \gamma(\mathbf{h}) \ , & \text{με } b_2 &> 0 \ , \\ \gamma_{12}(\mathbf{h}) &= \gamma_{21}(\mathbf{h}) = b_{12} \gamma(\mathbf{h}) \ . \end{aligned} \quad (29)$$

Τότε η μόνη συνθήκη είναι $b_{12} < \sqrt{b_1 b_2}$.

Αν τα βαριογράμματα γ_{11} , γ_{22} και γ_{12} είναι το καθένα το άθροισμα δύο συνιστωσών (π.χ., ένα nugget effect συν ένα ημιβαριόγραμμα γ , $C[1-\delta(\mathbf{h})]+b\gamma(\mathbf{h})$, τότε ένας παρόμοιος περιορισμός πρέπει να μπει για τα C.

Στην πράξη, τα πειραματικά βαριογράμματα και τα cross-βαριογράμματα καθορίζονται χρησιμοποιώντας όλα τα ζεύγη των διαθέσιμων σημείων μέτρησης χρησιμοποιώντας την εξίσωση (27). Τότε αυτά τα πειραματικά βαριογράμματα προσαρμόζονται με έναν γραμμικό συνδυασμό των ίδιων βασικών Ημιβαριογραμμάτων, και διασφαλίζεται ότι οι περιορισμοί (28) ή (29) ικανοποιούνται.

Κεφάλαιο 5

ΓΕΩΓΡΑΦΙΚΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΩΝ

5.1 Γενικά για τα ΓΣΠ

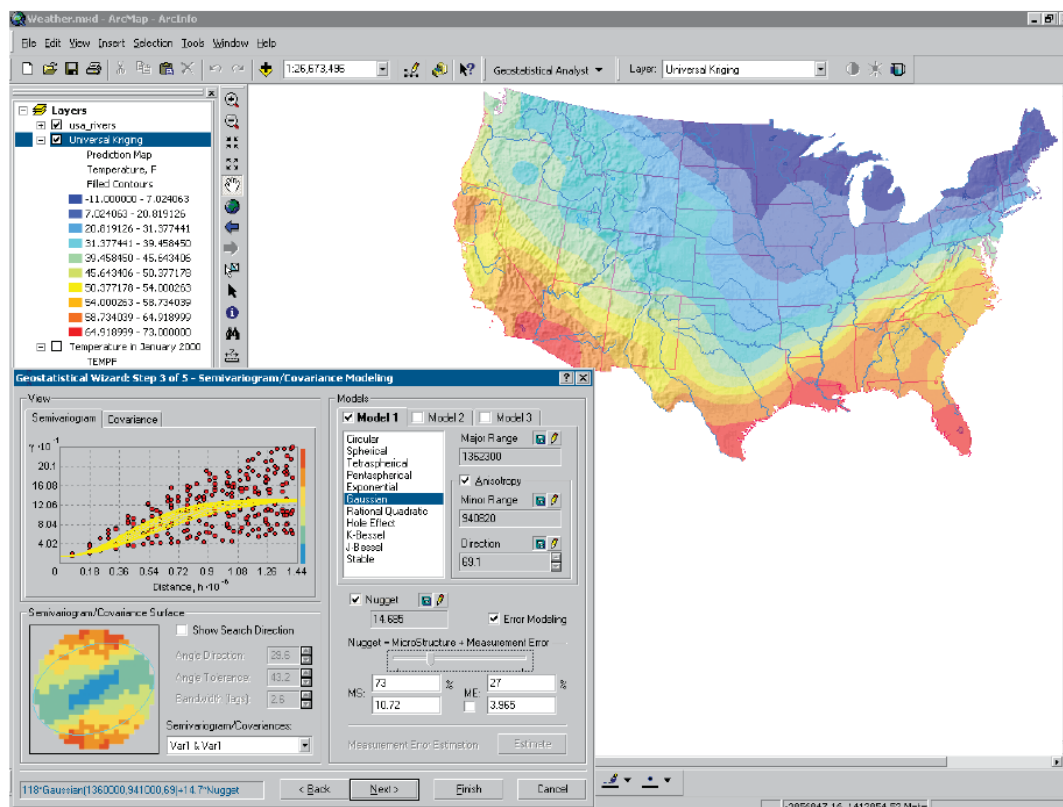
Ο όρος Γεωγραφικό Σύστημα Πληροφοριών (GIS) αναφέρεται σε Σύστημα Πληροφορικής που συνδυάζει ψηφιοποιημένα γραφικά υπόβαθρα (χάρτες) και περιγραφικές πληροφορίες, τα οποία με κατάλληλη επεξεργασία μπορούν να βοηθήσουν στην παραγωγή θεματικών χαρτών, περιγραφικών καταλόγων και διαγραμμάτων με σκοπό την ανάλυση και μελέτη του χώρου για τη λήψη αποφάσεων που αφορούν τον άνθρωπο και τη γη.

Τα γεωγραφικά συστήματα πληροφοριών είναι μια συλλογή υπολογιστικών συστημάτων, λογισμικών συστημάτων, γεωγραφικών δεδομένων και ανθρωπίνου δυναμικού με σκοπό τη συλλογή, καταχώρηση, διαχείριση, ανάλυση, επεξεργασία και απόδοση, κάθε μορφής πληροφορίας που αφορά το γεωγραφικό περιβάλλον. Τα γεωγραφικά συστήματα πληροφοριών αρχικά αποτελούν μια ηλεκτρονική βάση γεωγραφικών δεδομένων ή πιο απλά είναι ένας ψηφιακός χάρτης.

Ένα Γεωγραφικό Σύστημα Πληροφοριών (ΓΣΠ), γνωστό ευρέως και ως G.I.S. (Geographic Information System), είναι ένα σύστημα διεύθυνσης χωρικών δεδομένων (spatial data) και συσχετισμένων ιδιοτήτων. Στην πιο αυστηρή μορφή του, είναι ένα ηλεκτρονικό σύστημα, ικανό να ενσωματώσει, αποθηκεύσει, προσαρμόσει, αναλύσει και παρουσιάσει γεωγραφικά-συσχετισμένες (geographically-referenced) πληροφορίες. Σε πιο γενική μορφή, ένα ΓΣΠ είναι ένα εργαλείο "έξυπνου χάρτου" το οποίο επιτρέπει τους χρήστες του να δημιουργήσουν διαδραστικά ερωτήσεις (αναζητήσεις δημιουργούμενες από τον χρήστη), να αναλύσουν τα χωρικά δεδομένα (spatial data) και να προσαρμόσουν δεδομένα.

Η εφαρμογή Γεωγραφικών Συστημάτων Πληροφοριών είναι ιδιαίτερα χρήσιμη στην μελέτη περιβαλλοντικών προβλημάτων. Περιβαλλοντικά προβλήματα όπως η ρύπανση των ακτών από πετρέλαια, οι δασικές πυρκαγιές, η ατμοσφαιρική ρύπανση κ.ά. μελετούνται με την χρήση ΓΣΠ. Ιδιαίτερα χρήσιμη είναι η δυνατότητα των ΓΣΠ να οργανώνουν μεγάλες ποσότητες πληροφορίας που προέρχονται από συσκευές παρακολούθησης περιβαλλοντικών παραμέτρων (monitoring). Έτσι έχουν αναπτυχθεί εφαρμογές ΓΣΠ που παρακολουθούν, χαρτογραφούν και ελέγχουν υποθέσεις για διάφορα φαινόμενα όπως παραδείγματος χάρι οι ρύποι. Τα τελευταία χρόνια τα ΓΣΠ επεκτάθηκαν

και εφαρμόστηκαν και σε πιο εξειδικευμένα περιβαλλοντικά προβλήματα, όπως για παράδειγμα στην καταστροφή που προκλήθηκε στις ακτές της Αλάσκας από τη βύθιση του τάνκερ Exxon Valdez το 1989, η οποία μελετήθηκε συστηματικά με τη βοήθεια ενός ΓΣΠ. Έτσι διερευνήθηκε η χωροχρονική εξέλιξη της πετρελαιοκηκίδας και οι επιπτώσεις της στο οικοσύστημα των ακτών. Αντίστοιχες μελέτες έχουν γίνει με τη χρήση ΓΣΠ για τις ευρύτερες περιβαλλοντικές επιπτώσεις από τον πόλεμο του Κουβέιτ στον Περσικό κόλπο.



Στην παρούσα εργασία χρησιμοποιήθηκε το λογισμικό ΓΣΠ ArcGIS Desktop της εταιρείας ESRI που είναι ίσως το πιο γνωστό πακέτο ανάπτυξης εφαρμογών ΓΣΠ. Ο κύριος όμως λόγος που επιλέχθηκε το συγκεκριμένο πακέτο ήταν οι εξαιρετικές δυνατότητες που παρέχει στον χρήστη όσον αφορά την γεωστατιστική ανάλυση δεδομένων.

5.2 ΓΣΠ και Γεωστατιστική Ανάλυση

Το λογισμικό ArcGIS είναι εφοδιασμένο με ένα ισχυρό εργαλείο για την γεωστατιστική ανάλυση των δεδομένων, το οποίο παρέχει την δυνατότητα εφαρμογής πολλών, γεωστατιστικών και μη, μεθόδων χωρικής παρεμβολής. Έτσι για την Γεωστατιστική Ανάλυση των δεδομένων και την εφαρμογή της μεθόδου kriging στην παρούσα μελέτη

χρησιμοποιήθηκε το λογισμικό ArcGIS Geostatistical Analyst που αποτελεί μια επέκταση του λογισμικού ΓΣΠ ArcGIS Desktop και παρέχει μια ποικιλία εργαλείων για ανάλυση χωρικών δεδομένων, εντοπισμό ανωμαλιών των δεδομένων, βέλτιστη πρόβλεψη, εκτίμηση της αβεβαιότητας της πρόβλεψης και δημιουργία επιφανειών.

Το λογισμικό ArcGIS Geostatistical Analyst, [20], [21], δίνει την δυνατότητα στον χρήστη

- να διερευνήσει την μεταβλητότητα των δεδομένων, να εξετάσει την ύπαρξη γενικών τάσεων (trends) και να ερευνήσει την χωρική αυτοσυσχέτιση και την συσχέτιση μεταξύ πολλαπλών συνόλων δεδομένων
- να δημιουργήσει προβλέψεις, να υπολογίσει τα τυπικά σφάλματα των προβλέψεων και να δημιουργήσει χάρτες χρησιμοποιώντας διάφορα γεωστατιστικά μοντέλα και εργαλεία.

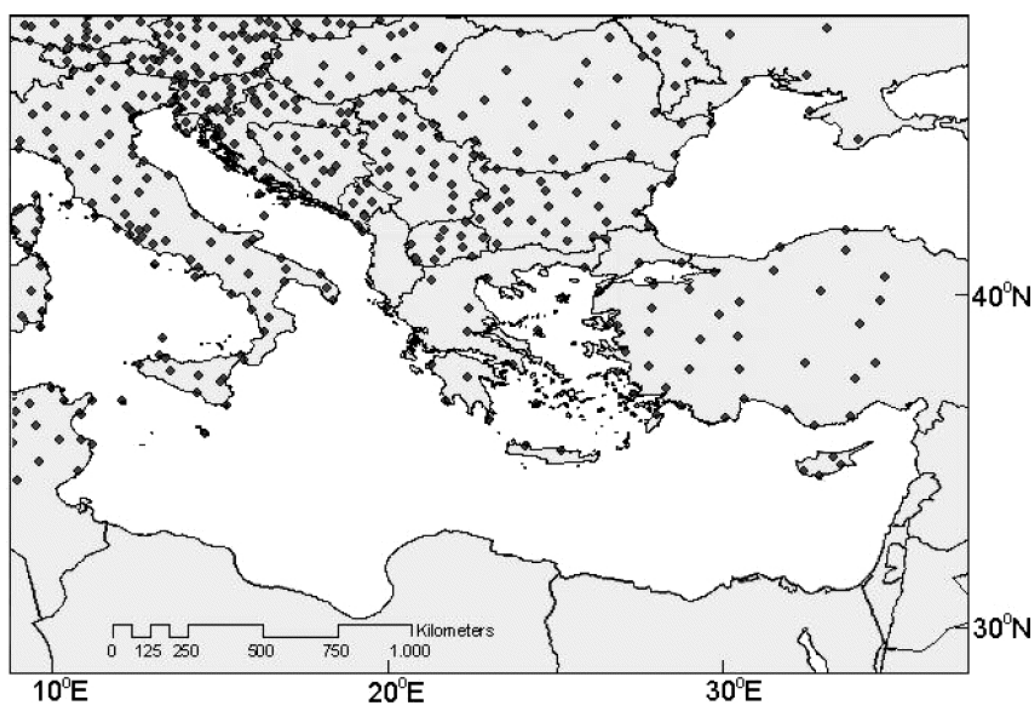
Το ArcGIS Geostatistical Analyst είναι ένα ολοκληρωμένο πακέτο λογισμικού για επεξεργασία χωρικών δεδομένων, γεωστατιστική ανάλυση των δεδομένων και την δημιουργία χαρτών πρόβλεψης και για το λόγο αυτό επιλέχθηκε να χρησιμοποιηθεί στην παρούσα μελέτη.

Κεφάλαιο 6

ΔΕΔΟΜΕΝΑ – ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

6.1 Διαθέσιμα Δεδομένα

Στην παρούσα μελέτη, τα διαθέσιμα δεδομένα αφορούσαν μετρήσεις βροχόπτωσης στην ευρύτερη περιοχή της Μεσογείου. Τα δεδομένα της μελέτης αποτελούνται από μετρήσεις αθροιστικής βροχόπτωσης 24 ωρών σε ένα δίκτυο από 233 έως 299 συνοπτικούς σταθμούς. Η γεωγραφική κατανομή των βροχομετρικών σταθμών καλύπτει την κεντρική – ανατολική Μεσόγειο και παρουσιάζεται στην Εικόνα 1.

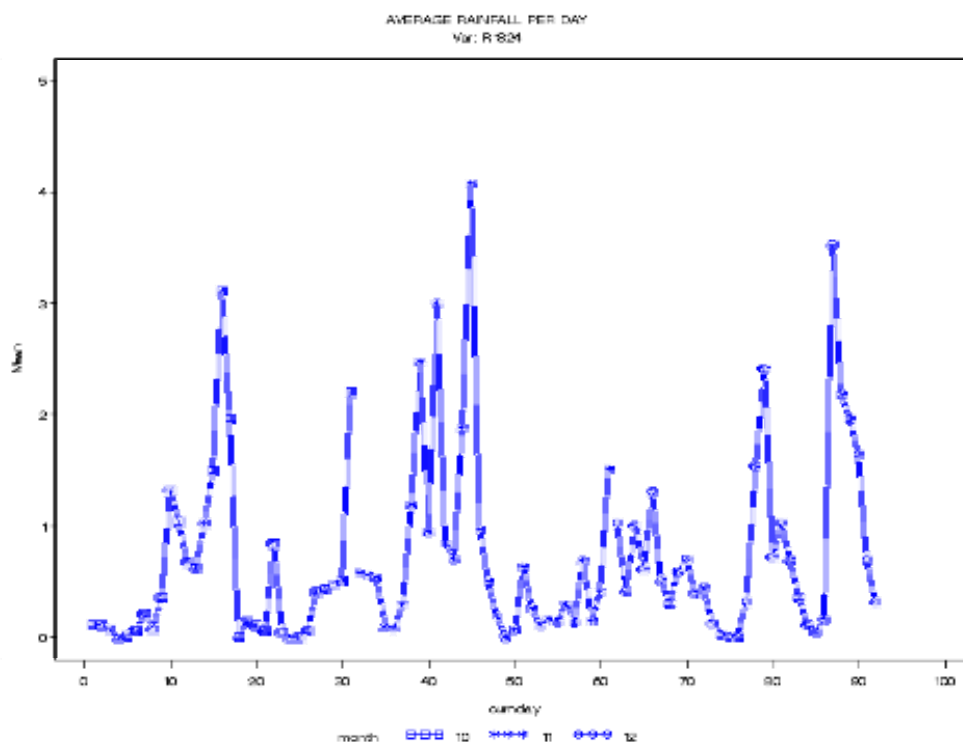


Εικόνα 1. Γεωγραφικές θέσεις των βροχομετρικών σταθμών στην υπό μελέτη περιοχή.

Τα διαθέσιμα δεδομένα για την παρούσα μελέτη συνίστανται σε δεδομένα 24-ωρης αθροιστικής βροχόπτωσης που έχουν καταγραφεί από ένα δίκτυο 233 έως 299 συνοπτικών σταθμών. Τα δεδομένα αυτά θα αποτελέσουν την βάση για την επιβεβαίωση των εκτιμήσεων της βροχόπτωσης στη συνέχεια. Η γεωγραφική κατανομή των βροχομετρικών σταθμών καλύπτει την κεντρική-ανατολική Μεσόγειο και απεικονίζεται στην Εικόνα 1. Το συγκεκριμένο πλήθος των σταθμών ήταν το μέγιστο διαθέσιμο από το

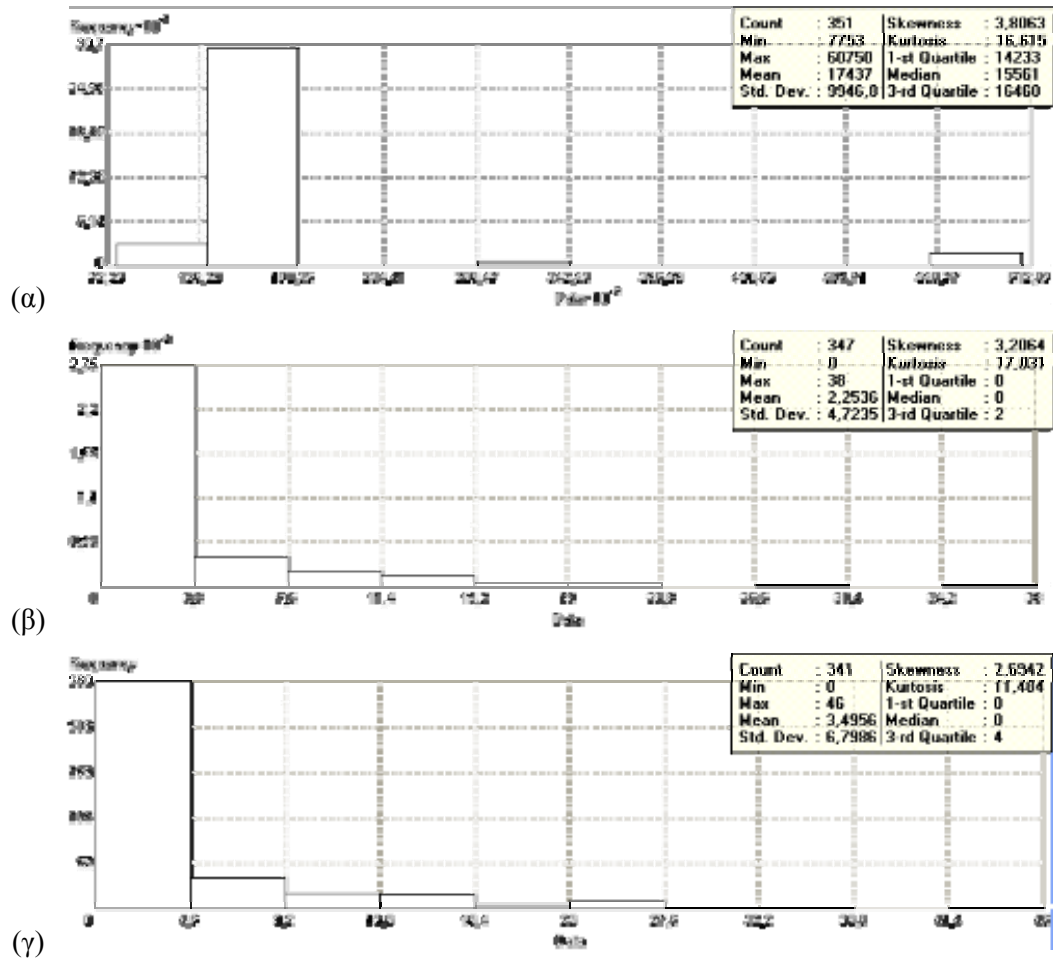
Ευρωπαϊκό Κέντρο Πρόγνωσης ECMWF (European Center for Medium Weather Forecast). Σημειώνουμε τέλος ότι τα δεδομένα αυτά έχουν ήδη περάσει από δοκιμασίες ελέγχου ποιότητας (quality tests) για τον εντοπισμό τυχόν ακραίων τιμών.

Πιο αναλυτικά, τα διαθέσιμα δεδομένα συνίστανται σε τιμές 24-ωρης αθροιστικής βροχόπτωσης στις 6:00 μ.μ., για 92 ημέρες κατά τους μήνες Οκτώβριο, Νοέμβριο και Δεκέμβριο του έτους 2004. Το γράφημα της Εικόνας 2 απεικονίζει τις τιμές της αθροιστικής βροχόπτωσης κατά τη διάρκεια των τριών αυτών μηνών.



Εικόνα 2. Αθροιστική 24-ωρη βροχόπτωση για τους 3 μήνες (Οκτώβριο-Δεκέμβριο).

Όσον αφορά τις ημέρες για τις οποίες θα εφαρμόσουμε χωρική παρεμβολή στην παρούσα μελέτη, έγινε η παρακάτω επεξεργασία. Αρχικά οι 92 ημέρες ταξινομήθηκαν όσον αφορά την τιμή της αθροιστικής βροχόπτωσης. Στην συνέχεια διαχωρίστηκαν σε 10 κατηγορίες ανάλογα με την τιμή αυτή και επιλέχθηκε μια ημέρα από κάθε κατηγορία. Από αυτές, στην παρούσα μελέτη επιλέξαμε να παρουσιάσουμε αποτελέσματα χωρικής παρεμβολής για τρεις ημέρες: τις 28/10, 8/11 και 26/12, που αντιστοιχούν σε χαμηλή, μεσαία και υψηλή τιμή μέσης ημερήσια τιμής βροχόπτωσης, αντίστοιχα. Στην Εικόνα 3 παρουσιάζονται τα ιστογράμματα για τις τρεις αυτές ημέρες (28/10, 8/11 και 26/12).



Εικόνα 3. Ιστογράμματα των τιμών αθροιστικής βροχόπτωσης για τις τρεις ημέρες (28/10 (α), 8/11 (β), και 26/12 (γ)).

Πρέπει να σημειώσουμε ότι πριν προχωρήσουμε στην ανάλυση των δεδομένων, τα δεδομένα ελέγχθηκαν για τυχόν ελλείψεις και για ακραίες τιμές. Εντοπίστηκαν ορισμένες ακραίες παρατηρήσεις με επίπεδα βροχόπτωσης 350, 400, κ.λ.π. Οι ελάχιστες αυτές παρατηρήσεις με τιμές πάνω από 100, καθώς και παρατηρήσεις με αρνητικές τιμές αφαιρέθηκαν από τα δεδομένα.

Ένα άλλο σημαντικό θέμα για την ανάλυση των δεδομένων είναι ότι δεν ακολουθούν την κανονική κατανομή, γεγονός που θα διευκόλυνε την εφαρμογή γεωστατιστικών μεθόδων και της μεθόδου kriging. Για το σκοπό αυτό δοκιμάστηκαν διάφοροι root-type και λογαριθμικοί μετασχηματισμοί (οι οποίοι εμφανίζονται αρκετά δημοφιλής στην βιβλιογραφία που σχετίζεται με την μέθοδο kriging), αλλά χωρίς ικανοποιητικά αποτελέσματα. Με δεδομένο ότι πολύ λίγοι ερευνητές μετασχηματίζουν τα δεδομένα τους, όπως φαίνεται για παράδειγμα στην εργασία [24] (SIC 97 Spatial Interpolation Comparison report), αποφασίσαμε να μην χρησιμοποιήσουμε μετασχηματισμό των δεδομένων στην παρούσα μελέτη.

6.2 Προϋποθέσεις για την εφαρμογή της μεθόδου kriging

Η κύρια προϋπόθεση για την εφαρμογή του Kriging για την χωρική παρεμβολή μιας μεταβλητής είναι το μοντέλο που προσαρμόζεται στο πειραματικό ημιβαριόγραμμα να χαρακτηρίζει την χωρική κατανομή της υπό μελέτη μεταβλητής. Υπάρχουν πολλοί τρόποι για να ελεγχθεί η ορθότητα της παραπάνω παραδοχής. Τις περισσότερες φορές, γίνονται έλεγχοι συγκρίνοντας τις πραγματικές τιμές των δειγμάτων με τις εκτιμημένες τιμές τους από τις παραμέτρους του μοντέλου του βαριογράμματος σε σχέση με τα δεδομένα των γύρω δειγμάτων, χρησιμοποιώντας την μέθοδο kriging. Οι παράμετροι του βαριογράμματος διορθώνονται μέχρι το μοντέλο να ανταποκρίνεται ικανοποιητικά στην πραγματικότητα, δηλαδή να δίνει αμερόληπτες εκτιμήσεις των τιμών των δειγμάτων και τα πραγματικά σφάλματα εκτίμησης να συμφωνούν με τα θεωρητικά προβλεπόμενα σφάλματα από το kriging. Η εκτίμηση σημειακών τιμών με τον τρόπο αυτό, είναι γενικά μια αρκετά δύσκολη και χρονοβόρος διαδικασία. Οι εκτιμήσεις των τιμών που προκύπτουν, έχουν πολύ χαμηλότερα σφάλματα εκτίμησης από τις σημειακές εκτιμήσεις που έγιναν κατά τον έλεγχο αξιοπιστίας του μοντέλου του βαριογράμματος. Αν η χωρική κατανομή αλλάζει μέσα στο χώρο ενδιαφέροντος, τότε υπάρχει μη σταθερότητα και ένα μόνο ημιβαριόγραμμα δεν είναι αρκετό για να γίνει σωστός χαρακτηρισμός σε όλα τα μέρη του πεδίου. Σε αυτές τις περιπτώσεις γίνεται χρήση μεθόδων kriging όπως είναι η μέθοδος του Γενικευμένου (universal) kriging, που απαιτούν πολυπλοκότερες εξισώσεις και συνεπώς λύσεις. Πολλές φορές για να αντιμετωπιστούν τέτοιες περιπτώσεις, το χωρίο χωρίζεται σε μικρότερες επιμέρους περιοχές στις οποίες η χωρική μεταβλητότητα παραμένει σταθερή και έτσι γίνεται χρήση του ordinary kriging. Το kriging στην μορφή αυτή μπορεί να χρησιμοποιηθεί στην πλειονοψηφία των περιπτώσεων.

6.3 Ανάλυση των δεδομένων

Για την κατασκευή ενός μοντέλου χωρικής παρεμβολής το πρώτο βήμα είναι η ανάλυση των διαθέσιμων δεδομένων (exploratory spatial data analysis), τα αποτελέσματα της οποίας θα χρησιμοποιηθούν στη συνέχεια για την επιλογή του κατάλληλου μοντέλου καθώς και για τον καθορισμό των παραμέτρων του.

Στα πλαίσια της ανάλυσης αυτής αρχικά παρουσιάζονται οι χάρτες Voronoi που απεικονίζουν τα διαφορετικά δυναμικά βροχόπτωσης για κάθε ημέρα, ενώ στην συνέχεια γίνεται μια ανάλυση των τάσεων που εμφανίζονται στα δεδομένα.

6.3.1 Χάρτες Voronoi

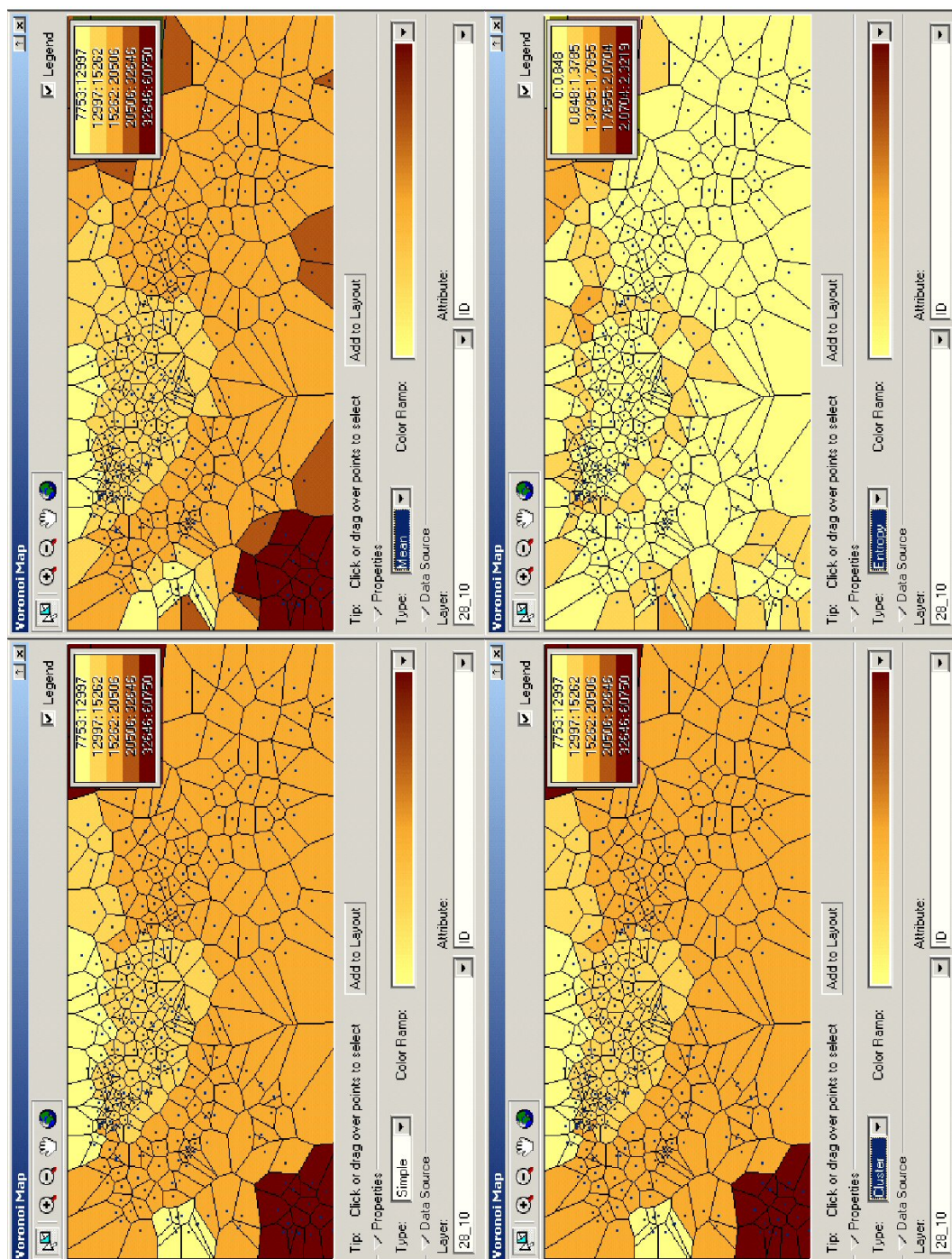
Οι χάρτες Voronoi κατασκευάζονται από μια σειρά από πολύγωνα γύρω από την τοποθεσία ενός σημείου μέτρησης. Κατασκευάζονται έτσι ώστε κάθε τοποθεσία μέσα σε ένα πολύγωνο να είναι πιο κοντά στο συγκεκριμένο σημείο μέτρησης από ότι σε κάθε άλλο σημείο μέτρησης.

Αφού κατασκευαστούν τα πολύγωνα, ως γείτονες ενός σημείου μέτρησης ορίζονται όλα τα άλλα σημεία μέτρησης των οποίων τα πολύγωνα έχουν κοινό σύνορο με το πολύγωνο του συγκεκριμένου σημείου μέτρησης. Χρησιμοποιώντας τον παραπάνω ορισμό «γειτονίας» μπορούν να υπολογιστούν μια σειρά από τοπικές στατιστικές. Για παράδειγμα, ένας τοπικός μέσος όρος (mean) υπολογίζεται παίρνοντας τον μέσο όρο των σημείων μέτρησης που ανοίκουν στην γειτονιά ενός πολύγωνου.

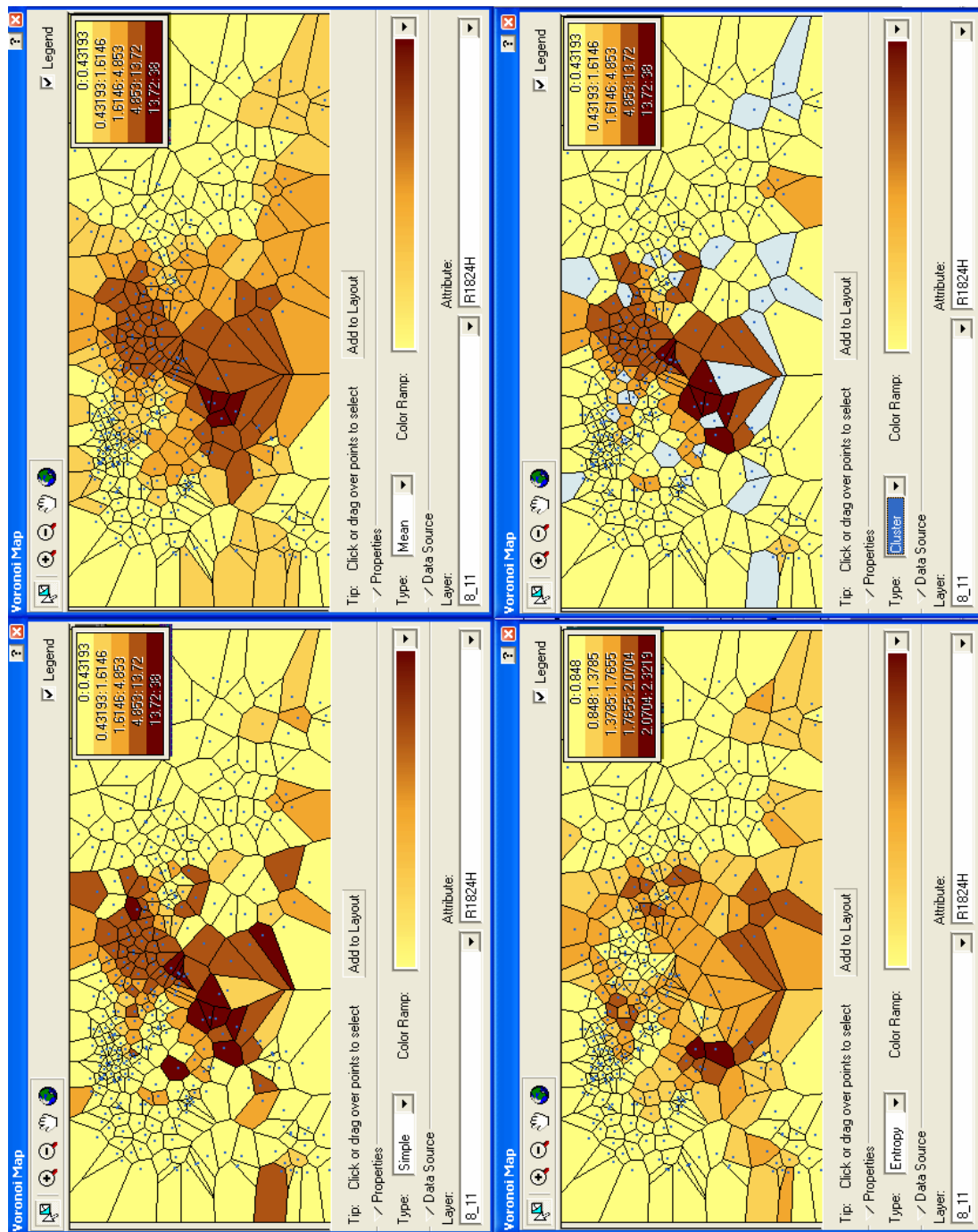
Διαφορετικές στατιστικές Voronoi χρησιμοποιούνται για διαφορετικούς σκοπούς. Στην παρούσα μελέτη παρουσιάζουμε τους χάρτες Voronoi που αντιστοιχούν στις εξής στατιστικές:

- Mean - Local smoothing
- Entropy - Local Variation
- Cluster- Local outliers
- simple - Local influence

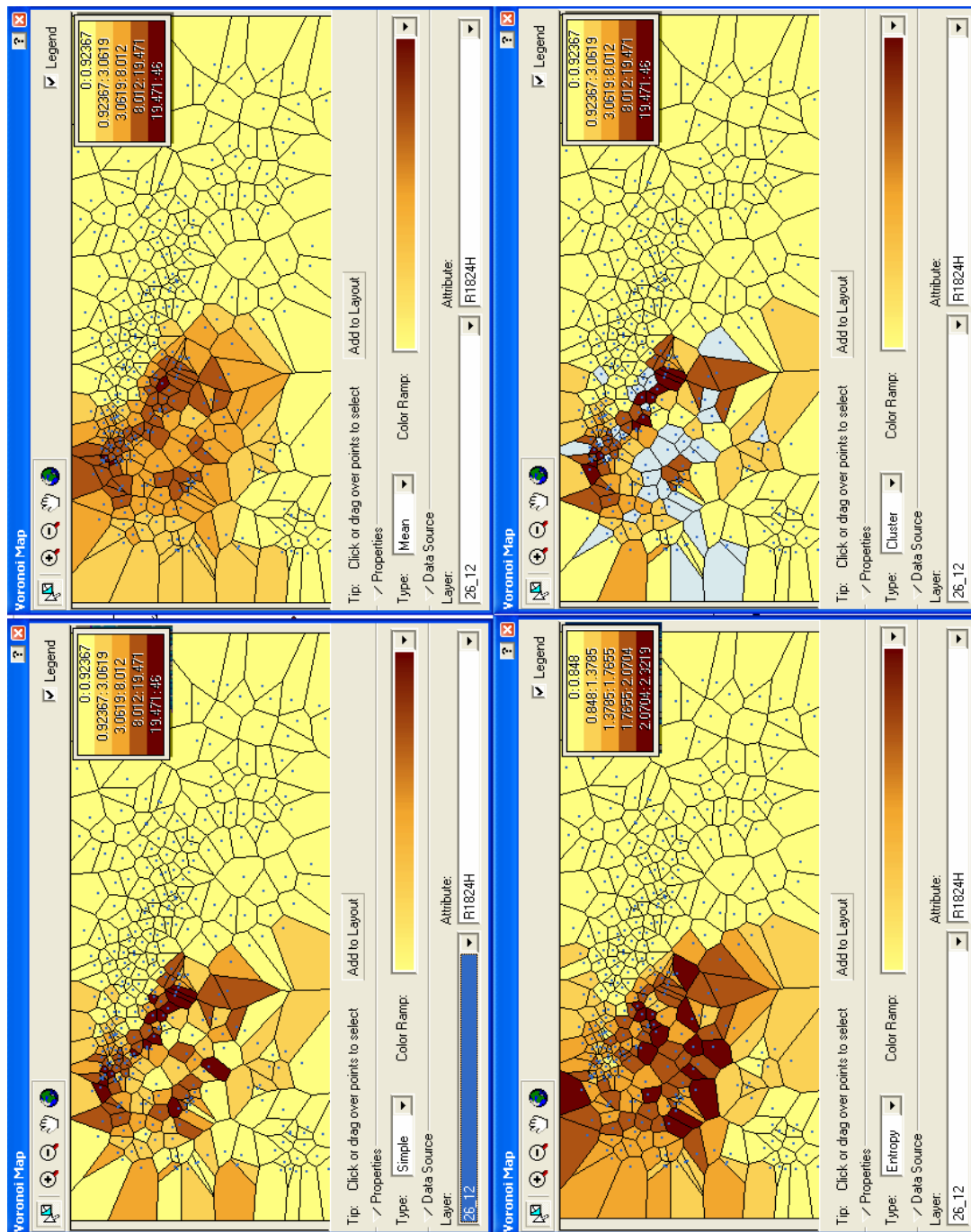
Στις Εκόνες 4, 5 και 6 που ακολουθούν, παρουσιάζονται οι χάρτες για simple, mean, entropy και cluster στατιστικές, για τα δεδομένα βροχόπτωσης κατά τις τρεις ημέρες 28/10, 8/11 και 26/12 στις οποίες επικεντρώνεται η παρούσα μελέτη, αντίστοιχα.



Εικόνα 4. Χάρτες Voronoi για Simple (πάνω αριστερά), Mean (πάνω δεξιά), Entropy (κάτω αριστερά) και Cluster (κάτω δεξιά) στατιστικές που αντιστοιχούν στην ημέρα με τα χαμηλά επίπεδα βροχόπτωσης (28/10).



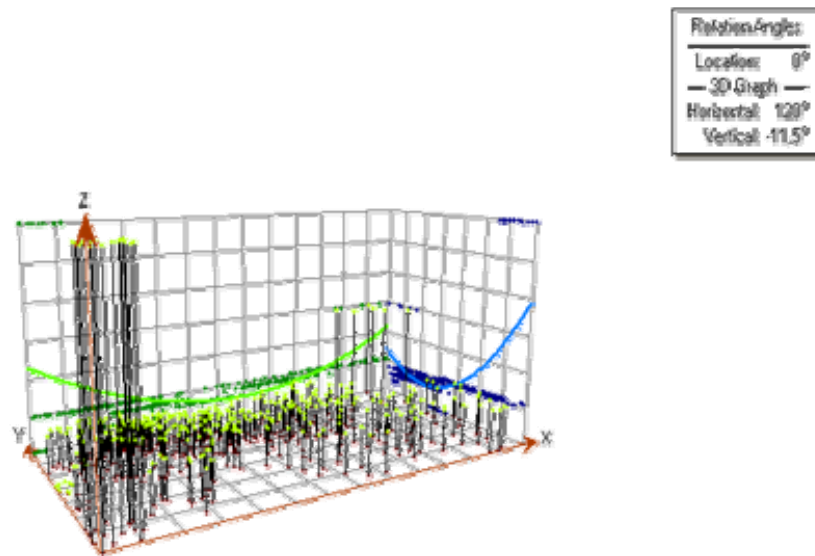
Εικόνα 5. Χάρτες Voronoi για Simple (πάνω αριστερά), Mean (πάνω δεξιά), Entropy (κάτω αριστερά) και Cluster (κάτω δεξιά) στατιστικές που αντιστοιχούν στην ημέρα με τα μεσαία επίπεδα βροχόπτωσης (8/11).



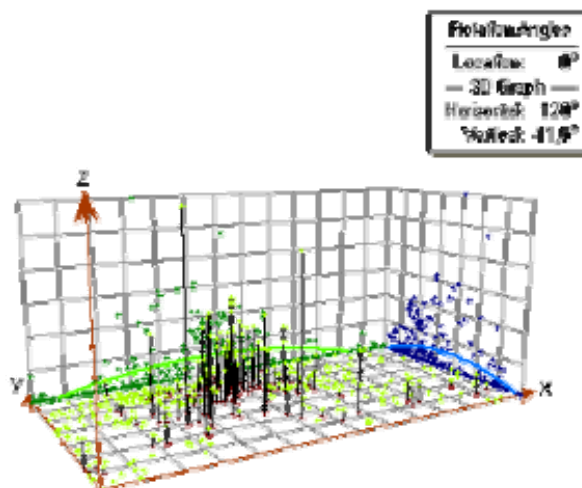
Εικόνα 6. Χάρτες Voronoi για Simple (πάνω αριστερά), Mean (πάνω δεξιά), Entropy (κάτω αριστερά) και Cluster (κάτω δεξιά) στατιστικές που αντιστοιχούν στην ημέρα με τα υψηλά επίπεδα βροχόπτωσης (26/12).

6.3.2 Ανάλυση Τάσεων - Trend Analysis

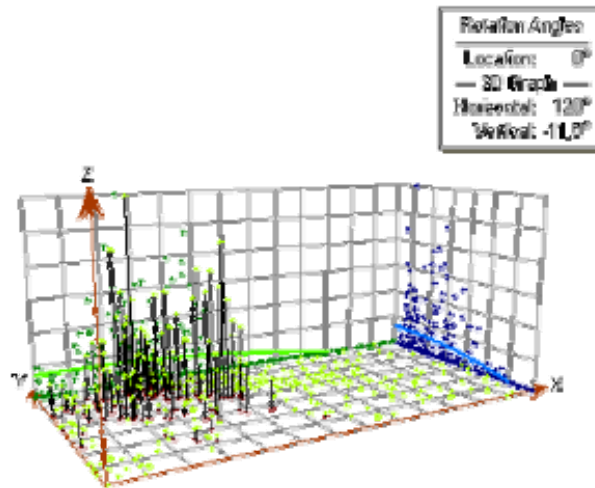
Η ύπαρξη καθολικών τάσεων στα δεδομένα θα πρέπει επίσης να διερευνηθεί. Στη συνέχεια, παρουσιάζουμε τρισδιάστατα διαγράμματα (scatterplots) των δεδομένων στα οποία διακρίνεται η παρουσία χωρικών τάσεων των δεδομένων. Στις Εικόνες 7α, 7β και 7γ παρουσιάζονται τα διαγράμματα αυτά για τις τρεις επιλεγμένες ημέρες.



Εικόνα 7α. Το τρισδιάστατο διάγραμμα για την ανίχνευση χωρικών τάσεων των δεδομένων για την ημέρα 28/10.



Εικόνα 7β. Το τρισδιάστατο διάγραμμα για την ανίχνευση χωρικών τάσεων των δεδομένων για την ημέρα 8/11.



Εικόνα 7γ. Το τρισδιάστατο διάγραμμα για την ανίχνευση χωρικών τάσεων των δεδομένων για την ημέρα 26/12.

Από τα παραπάνω γραφήματα ανίχνευσης χωρικών τάσεων παρατηρούμε ότι στα δεδομένων των δύο ημερών μεσαίας και υψηλής βροχόπτωσης (8/11 και 26/12) εμφανίζονται καμπυλόγραμμες τάσεις. Οι χωρικές αυτές τάσεις θα πρέπει να υπολογισθούν στη συνέχεια κατά την εφαρμογή της μεθόδου kriging.

Κεφάλαιο 7

ΕΦΑΡΜΟΦΗ ΤΗΣ ΓΕΩΣΤΑΤΙΣΤΙΚΗΣ ΜΕΘΟΔΟΥ

Στο κεφάλαιο αυτό παρουσιάζεται η τεχνική της γεωστατιστικής παρεμβολής. Όπως δηλώνεται και από το όνομα, οι γεωστατιστικές τεχνικές δημιουργούν επιφάνειες χρησιμοποιώντας τις στατιστικές ιδιότητες των δεδομένων. Εφόσον η γεωστατιστική βασίζεται στην στατιστική, οι μέθοδοι αυτές υπολογίζουν, πέρα από τις επιφάνειες πρόγνωσης, και επιφάνειες του σφάλματος ή της αβεβαιότητας, παρέχοντας μας με τον τρόπο αυτό ενδείξεις για το πόσο καλή είναι η πρόγνωση.

Πολλές μέθοδοι σχετίζονται με την γεωστατιστική, αλλά ένα μεγάλο κομμάτι της γεωστατιστικής αποτελεί η οικογένεια των μεθόδων kriging. Για περισσότερες πληροφορίες σχετικά με τις μεθόδους kriging, ο αναγνώστης παραπέμπεται στους Kitanidis, Cressie (1993), Stein (1999) και Isaaks and Srivastava (1989).

Διάφορες μέθοδοι μπορούν χρησιμοποιηθούν για την χωρική γραμμική παρεμβολή δεδομένων, όπως η μέθοδος inverse distance weighting, η μέθοδος των ελαχίστων τετραγώνων, η παρεμβολή με splines, ο τριγωνισμός, κ.ά. Μεταξύ των παραπάνω μεθόδων μόνο η μέθοδος kriging συνυπολογίζει την χωρική δομή του υπό μελέτη φαινομένου. Στην παρούσα εργασία, παρατηρούμε πολλές μεταβολές στην χωρική συμπεριφορά της βροχόπτωσης, και για το λόγο αυτό η τεχνική kriging αναμένεται να έχει καλύτερη συμπεριφορά από άλλες μεθόδους παρεμβολής. Συγκριμένα, η μέθοδος kriging λαμβάνει υπ' όψη τα εξής:

- την πυκνότητα και την κατανομή των δεδομένων στην υπό μελέτη περιοχή,
- την υψηλότερη ή την χαμηλότερη χωρική συνέχεια που μοντελοποιείται από την συνάρτηση του ημιβαριογράμματος, και
- τα διάφορα σφάλματα των μετρήσεων και τις αβεβαιότητες που σχετίζονται με τα δεδομένα.

Όπως και στην μέθοδο inverse distance weighting όπου η χωρική απόσταση των δεδομένων θεωρείται ότι σχετίζεται με την τιμή τους, η μέθοδος kriging αναγνωρίζει ότι η χωρική μεταβλητότητα είναι συνάρτηση της απόστασης. Το kriging θεωρείται ότι είναι ο βέλτιστος ανεξάρτητος (απροκατάληπτος) γραμμικός εκτιμητής (Best Linear Unbiased Estimator -BLUE). Είναι βέλτιστος γιατί μπορεί να αποδειχθεί ότι οι εκτιμήσεις του kriging έχουν την μικρότερη residual variance από κάθε άλλο εκτιμητή. Είναι γραμμικός γιατί η εκτίμηση είναι ένας γραμμικός συνδυασμός των δεδομένων με βάρη ανάλογα της

απόστασης και της κατεύθυνσής τους από το εκάστοτε σημείο για το οποίο γίνεται η εκτίμηση.

Η μέθοδος kriging είναι ουσιαστική η στατιστική προσέγγιση όσον αφορά την παρεμβολή δεδομένων και βασίζεται στην χωρική μεταβλητότητα. Οι μέθοδοι kriging αποτελούνται από τεχνικές βέλτιστης εκτίμησης που βασίζονται στην αρχή των ελαχίστων τετραγώνων και χρησιμοποιούν μια συνάρτηση χωρικής δομής, η οποία συνήθως είναι το ημιβαριόγραμμα.

Η μέθοδος βασίζεται στην διαπίστωση ότι η χωρική μεταβλητότητα π.χ. της βροχόπτωσης (στην περίπτωση μας) είναι υπερβολικά σύνθετη για να περιγραφεί από μια ομαλή μαθηματική συνάρτηση, αλλά μπορεί να περιγραφεί καλύτερα από μια στοχαστική επιφάνεια. Η παρεμβολή γίνεται διερευνώντας αρχικά και στην συνέχεια περιγράφοντας τις στοχαστικές ιδιότητες της βροχόπτωσης. Οι παρεμβαλλόμενες τιμές της φυσικής ιδιότητας (αθροιστικής βροχόπτωσης στην παρούσα μελέτη) σε κάθε σημείο υπολογίζεται σαν ένας γραμμικός συνδυασμός των μετρήσεων που έχουν ληφθεί σε μια περιοχή γύρω από το σημείο αυτό. Η περιοχή αυτή ονομάζεται «γειτονιά». Στην συνέχεια η πληροφορία αυτή χρησιμοποιείται για να εκτιμηθούν τα «βάρη» που θα χρησιμοποιηθούν για την παρεμβολή. Τα βάρη αυτά αντιστοιχίζονται στις μετρήσεις μέσω της χωρικής εξάρτησης της βροχόπτωσης όπως αυτή περιγράφεται από το ημιβαριόγραμμα.

Οι εφαρμογές των μεθόδων kriging έχουν σταδιακά εξαπλωθεί ευρέως σε διάφορους τομείς, με την μέθοδο του Κανονικού (Ordinary) kriging να εμφανίζεται ως η πλέον δημοφιλής.

7.0 Γεωστατιστική Ανάλυση - Η μεθοδολογία για την εφαρμογή της μεθόδου kriging

Η πρόβλεψη της συμπεριφοράς μεταβλητών σε περιοχές όπου δεν υπάρχουν μετρήσεις δεν είναι απλή διαδικασία. Η σωστή επιλογή του μοντέλου του ημιβαριογράμματος που περιγράφει ακριβέστερα τις πειραματικές παρατηρήσεις, καθώς και ο καθορισμός της ποιότητάς του είναι καθοριστική κατά την γεωστατιστική ανάλυση.

Ένας βασικός παράγοντας που εξετάζεται στην αρχή, είναι το ελάχιστο μέγεθος του διαθέσιμου δείγματος, καθώς από αυτό εξαρτάται και η ποιότητα του θεωρητικού μοντέλου. Αν το δείγμα αποτελείται από λιγότερες από 50 μετρήσεις, τότε δεν είναι δυνατόν να παραχθούν αξιόπιστα θεωρητικά αποτελέσματα. Αυτό δεν σημαίνει ότι μια μεγαλύτερη ποσότητα δειγμάτων μπορεί πάντα να δώσει αξιόπιστα θεωρητικά αποτελέσματα. Όμως, στην περίπτωση αυτή υπάρχει το πλεονέκτημα ότι τα αποτελέσματα της ανάλυσης θα είναι περισσότερα και ο βαθμός «υποστήριξης» του κάθε πειραματικού σημείου του ημιβαριογράμματος θα είναι μεγαλύτερος. Σε περιπτώσεις όπου το πλήθος των δεδομένων είναι ανεπαρκές, είτε γίνεται συλλογή επιπλέον

δεδομένων, είτε η ανάλυση γίνεται λαμβάνοντας υπόψη ότι τα αποτελέσματα δεν θα είναι αξιόπιστα.

Στην συνέχεια ακολουθεί η προσαρμογή του καταλληλότερου θεωρητικού μοντέλου που περιγράφει το πειραματικό ημιβαριόγραμμα. Το μοντέλο αυτό και η χρησιμότητά του ως εργαλείου πρόβλεψης εξαρτάται κυρίως από το πλήθος των δειγμάτων, την απόσταση μεταξύ των σημείων, τον βαθμό ανομοιογένειας κ.λ.π. Αν τα δεδομένα δεν είναι καλής ποιότητας, τότε τα αποτελέσματα του kriging δεν θα δώσουν αξιόπιστες προβλέψεις. Ένα χαρακτηριστικό παράδειγμα τέτοιας συμπεριφοράς είναι τα λεγόμενα μοντέλα καθαρού nugget. Στα μοντέλα αυτά, οι τιμές του ημιβαριογράμματος είναι ίσες ή μεγαλύτερες της διασποράς και είναι χρήσιμα μόνο για τον καθορισμό της μέσης τιμής της μεταβλητής στο χώρο, ενώ δεν ενδείκνυται η χρήση τους στην ανάλυση του kriging.

Κατά την γεωστατιστική ανάλυση κατασκευάζεται σε αρχικό στάδιο ένα ημιβαριόγραμμα του οποίου η ανοχή στην διεύθυνση είναι τόσο μεγάλη, ώστε η κατεύθυνση οποιουδήποτε διανύσματος που ενώνει δύο σημεία μετρήσεων να μην είναι σημαντική. Ο συνδυασμός όλων των δυνατών διευθύνσεων σε ένα ημιβαριόγραμμα καθιστά πλέον σημαντικό όμοιο το μέτρο του διανύσματος που ενώνει τα δύο σημεία, ενώ το ημιβαριόγραμμα μπορεί να θεωρηθεί ως ο μέσος όρος των ημιβαριογραμμάτων διαφόρων διευθύνσεων. Η χρήση τέτοιων ημιβαριογραμμάτων συνεπάγεται άμεσα την ύπαρξη ισοτροπίας στα δεδομένα, αλλά βοηθά στον προσδιορισμό παραμέτρων που χρειάζονται για των υπολογισμό των επόμενων ημιβαριογραμμάτων. Καθώς η διεύθυνση δεν παίζει σημαντικό ρόλο, γίνεται ευκολότερη η εύρεση παραμέτρων όπως ο αριθμός των διαστημάτων (lags) και το εύρος της απόστασης ανοχής.

7.1 Ordinary Kriging

Η μέθοδος του Κανονικού kriging έχει αναφερθεί στη βιβλιογραφία σαν ο βασικός αλγόριθμος της γεωστατιστικής (Deutsch και Journel, 1992, σελ. 64) λόγω της αξιόλογης αποτελεσματικότητάς της κάτω από διάφορες συνθήκες. Αυτός είναι και ο λόγος που η συγκεκριμένη μέθοδος επιλέχθηκε να εφαρμοσθεί στην παρούσα μελέτη.

Στο kriging η παρεμβολή υπολογίζεται μέσω ενός γραμμικού συνδυασμού των δεδομένων. Η επιφάνεια στην οποία πραγματοποιείται η παρεμβολή θεωρείται σαν μια πραγματοποίηση ενός τυχαίου πεδίου Z , του οποίου η τιμή είναι γνωστή σε ορισμένα σημεία $z(x_i)$ του \mathbb{R}^2 . Μέσω ορισμένων δεύτερης τάξης ιδιοτήτων του Z , κατασκευάζεται ένας εκτιμητής \hat{Z} του Z , ο οποίος ορίζεται σαν ένας γραμμικός συνδυασμός των γνωστών τιμών του Z . Επομένως, η χωρική εκτίμηση της μεταβλητής της βροχόπτωσης Z σε ένα σημείο στο x , για το οποίο δεν έχει γίνει μέτρηση, γίνεται από την σχέση

$$\hat{Z}(x) = \sum_{i=1}^n \lambda_i z(x_i)$$

όπου τα x_i είναι τα σημεία για τα οποία υπάρχουν διαθέσιμες μετρήσεις (δεδομένα), τα λ_i είναι τα βάρη που αντιστοιχούν στα σημεία μέτρησης x_i , και n είναι το πλήθος των σημείων μέτρησης.

Τα βάρη λ_i είναι επιλέγονται έτσι ώστε: (α) η εκτίμηση $\hat{Z}(x)$ της πραγματικής τιμής $z(x)$ να είναι αμερόληπτη, και (β) η μεταβλητότητα (variance) της εκτίμησης $\sigma^2(x)$ να ελαχιστοποιείται. Αυτό σημαίνει:

$$E[\hat{Z}(x) - Z(x)] = 0$$

και

$$\sigma^2(x) = \text{var}[\hat{Z}(x) - Z(x)] = \text{minimum}.$$

Για να διασφαλιστεί ότι η εκτίμηση θα είναι αμερόληπτη τα βάρη τα οποία αντιστοιχίζονται σε κάθε σημείο πρέπει να ικανοποιούν τη σχέση

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

Στην περίπτωση αυτή οι εξισώσεις του kriging που προκύπτουν είναι οι εξής:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma(h_{ij}) + \mu = \gamma(h_{ix}) , & \forall i \text{ (n equations)} \\ \sum_{i=1}^n \lambda_i = 1. \end{cases}$$

όπου το n είναι το πλήθος των μετρήσεων που χρησιμοποιούνται για την παρεμβολή στο σημείο x , τα $z(x_i)$ είναι οι τιμές της βροχόπτωσης που έχουν μετρηθεί στα σημεία x_i , το x είναι το σημείο για το οποίο εκτιμάται η τιμή της βροχόπτωσης, τα λ_i είναι οι συντελεστές βάρους που αντιστοιχούν στα σημεία x_i και τα οποία πρέπει να καθοριστούν, το μ είναι ένας πολλαπλασιαστής Lagrange, το $\gamma(h_{ij})$ είναι η τιμή του βαριογράμματος για την απόσταση h_{ij} μεταξύ των σημείων x_i και x_j , και $\gamma(h_{ix})$ είναι η τιμή του βαριογράμματος για την απόσταση h_{ix} μεταξύ των σημείων x_i και x .

Τα βάρη επιλέγονται μέσα από μια διαδικασία η οποία καλείται μοντελοποίηση βαριογράμματος (semivariogram modelling). Το Kriging είναι «βέλτιστο» θεωρώντας ένα μοντέλο βαριογράμματος της δομής της μεταβλητότητας των δειγμάτων βροχόπτωσης. Ο πίνακας των βαρών κατασκευάζεται με τέτοιο τρόπο που να εγγυάται ότι θα είναι θετικά ορισμένος, το οποίο με την σειρά του εξασφαλίζει ότι θα είναι αντιστρέψιμος. Η επιλογή του κατάλληλου βαριογράμματος είναι πάντα το πρώτο βήμα στην παρεμβολή με τη μέθοδο kriging καθώς περιγράφει την χωρική μεταβλητότητα μεταξύ των δεδομένων βροχόπτωσης, και θα παρουσιαστεί αναλυτικά στην επόμενη παράγραφο.

Σύμφωνα με την θεωρία του Kriging για την κατασκευή του πίνακα των βαρών πρέπει να χρησιμοποιηθεί το σύνολο των δεδομένων. Ο πίνακας των βαρών προκύπτει προκύπτει βρίσκοντας τον αντίστροφο ενός πίνακα που έχει διάσταση $(n+1) \times (n+1)$, όπου n είναι το πλήθος των δειγμάτων. Στην επιπλέον στήλη βρίσκονται οι πολλαπλασιαστές

Lagrange. Όταν το πλήθος των μετρήσεων είναι μεγάλο, η κοινή πρακτική είναι να επιλέγεται ένα σχετικά μικρό υποσύνολο μετρήσεων που αντιστοιχούν σε σημεία που περιβάλλουν το σημείο στο οποίο γίνεται η εκτίμηση. Επιλέγονται δηλαδή οι μετρήσεις σε μια γειτονιά και χρησιμοποιούνται μόνο αυτές για την εκτίμηση και όχι το σύνολο των δεδομένων.

Τα βάρη επιλέγονται με την λογική ότι μετρήσεις σε σημεία που βρίσκονται σε μεγάλη απόσταση από το σημείο στο οποίο γίνεται η εκτίμηση ασκούν μικρότερη επιρροή στην εκτιμώμενη τιμή από ότι τα κοντινά σημεία. Σκοπός της διαδικασίας είναι να ελαχιστοποιείται η μεταβλητότητα (variance) της εκτίμησης. Αυτό επιτυγχάνεται σε την κατάλληλη επιλογή των συντελεστών βάρους, για τους οποίους τίθεται ένας μόνον περιορισμός: ότι το άθροισμά τους πρέπει να είναι ίσο με τη μονάδα.

Γενικά κατά την γεωστατιστική ανάλυση απαιτείται τα δεδομένα να ακολουθούν την κανονική κατανομή. Το Κανονικό Kriging όμως είναι πολύ αποτελεσματικό και έτσι υπάρχει η δυνατότητα να χρησιμοποιηθεί ακόμα και όταν τα δεδομένα δεν ακολουθούν την κανονική κατανομή. Υπάρχουν όμως εναλλακτικές. Συνήθως αυτό που γίνεται είναι να μετασχηματίζονται τα δεδομένα έτσι ώστε να έχουν κανονική κατανομή, να χρησιμοποιείται η μέθοδος του kriging για τα μετασχηματισμένα δεδομένα, και στη συνέχεια να χρησιμοποιείται ο αντίστροφος μετασχηματισμός. Η προσέγγιση αυτή έχει διάφορα μειονεκτήματα, το μεγαλύτερο των οποίων είναι ότι ο αντίστροφος μετασχηματισμός μπορεί να εισάγει αβεβαιότητα στις εκτιμώμενες τιμές. Για τον λόγο αυτό, στην παρούσα μελέτη δεν υιοθετήθηκε η πρακτική του μετασχηματισμού των δεδομένων.

Όπως αναφέρθηκε παραπάνω η εφαρμογή της μεθόδου του kriging περιλαμβάνει δύο στάδια: την ποσοτικοποίηση της χωρικής δομής των δεδομένων και την κατασκευή της εκτίμησης. Το πρώτο στάδιο συνήθως αναφέρεται ως Ημιβαριογραφία (variography) και περιλαμβάνει την προσαρμογή ενός μοντέλου χωρικής συσχέτισης στα δεδομένα. Η διαδικασία αυτή παρουσιάζεται αναλυτικά για τα δεδομένα της βροχόπτωσης στην παράγραφο που ακολουθεί.

7.2 Ημιβαριογραφία και Ανάλυση Ανισοτροπίας

Για να κατασκευαστεί η χωρική παρεμβολή με την μέθοδο του ordinary kriging, πρέπει καταρχήν να μοντελοποιηθεί η χωρική μεταβλητότητα των δεδομένων. Η πρώτη ανάλυση που πρέπει να γίνει είναι η περιγραφή της χωρικής μεταβλητότητας όσον αφορά την κατεύθυνση και την απόσταση μεταξύ των σημείων των βροχομετρικών σταθμών. Κάθε ζεύγος σημείων ελέγχεται όσον αφορά αυτά τα δύο χαρακτηριστικά. Η χωρική μεταβολή της βροχόπτωσης στην υπό εξέταση περιοχή ποσοτικοποιήθηκε με την κατασκευή ημιβαριογραμμάτων (semivariograms). Στην παρούσα παράγραφο παρουσιάζονται τα κατευθυντικά (directional) ημιβαριογράμματα και η ανάλυση της ανισοτροπίας των δεδομένων.

Στη γεωστατιστική, η μεταβλητότητα στο χώρο ποσοτικοποιείται μέσω του ημιβαριογράμματος, το οποίο συμβολίζεται με $\gamma(h)$, και μετρά την μέση ανομοιομορφία μεταξύ των παρατηρήσεων σαν συνάρτηση της απόστασης και της κατεύθυνσης.

Οι συναρτήσεις του ημιβαριογράμματος ποσοτικοποιούν την θεώρηση ότι τα κοντινά πράγματα τείνουν να μοιάζουν περισσότερο από εκείνα που βρίσκονται σε μεγαλύτερη απόσταση μεταξύ τους. Το ημιβαριόγραμμα μετρά το μέγεθος της στατιστικής συχέτισης (statistical correlation) σαν συνάρτηση της απόστασης και ορίζεται από την σχέση

$$\gamma(h) = \frac{1}{2} \text{var}(Z(x_i + h) - Z(x_i)).$$

όπου $Z(x_i)$ και $Z(x_i + h)$ είναι δύο σημεία των οποίων η απόσταση είναι h .

Η συνάρτηση του ημιβαριογράμματος συνδέει την ημιδιασπορά (semivariance or dissimilarity) των σημείων των δεδομένων με την απόσταση που τα χωρίζει, και η γραφική της παράσταση μπορεί να χρησιμοποιηθεί για να παρέχει μια εικόνα της χωρικής συχέτισης των σημείων των δεδομένων με τα γειτονικά τους σημεία.

Η διαδικασία της μοντελοποίησης του βαριογράμματος περιλαμβάνει δύο ξεχωριστά στάδια, τα οποία είναι

- η κατασκευή του πειραματικού βαριογράμματος, και
- η προσαρμογή ενός μοντέλου (δηλ. μιας συνάρτησης) στο ημιβαριόγραμμα.

Επίσης, θα πρέπει να καθοριστεί μια τιμή επιβράδυνσης (lag size). Η τιμή επιβράδυνσης (lag) είναι ίση με το μέγεθος μιας κλάσης απόστασης στην οποία έχουν ομαδοποιηθεί ζεύγη σημείων (τοποθεσιών) έτσι ώστε να μειωθεί το πλήθος των πιθανών συνδυασμών. Αρχικά τα ζεύγη των σημείων (τοποθεσιών) ομαδοποιούνται σε κλάσεις απόστασης έτσι ώστε να μειωθεί το πλήθος των πιθανών συνδυασμών και να εμφανίζονται λιγότερα σημεία στο πειραματικό ημιβαριόγραμμα. Η διαδικασία αυτή ονομάζεται εγκιβωτισμός (binning). Η τιμή επιβράδυνσης είναι ίση με το μέγεθος μιας κλάσης απόστασης. Μια σωστή επιλογή στην τιμή της επιβράδυνσης βοηθά να αποκαλυφθούν οι χωρικές συσχετίσεις. Μείωση της τιμής επιβράδυνσης σημαίνει εστίαση στις λεπτομέρειες του μοντέλου και στις τοπικές μεταβολές μεταξύ γειτονικών σημείων.

Points separated by a lag distance greater than the range are spatially uncorrelated.

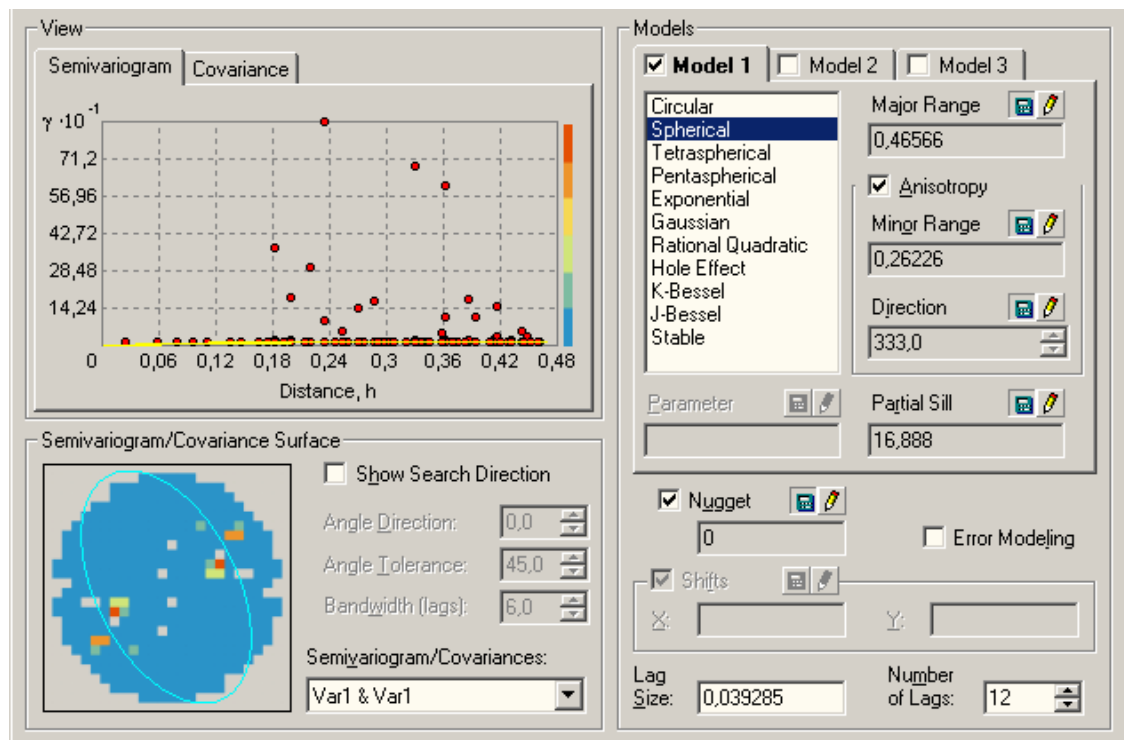
Ένας πολύ σημαντικός παράγοντας που πρέπει να διερευνηθεί είναι η ανισοτροπία. Η ανισοτροπία αναφέρεται σε περιπτώσεις που η χωρική μεταβλητότητα είναι συνάρτηση τόσο της απόστασης, αλλά και της κατεύθυνσης. Σε συγκεκριμένες κατευθύνσεις στα κοντινά σημεία τα πράγματα μπορεί να είναι πιο παρόμοια από ότι σε άλλες διευθύνσεις. Η επίδραση της κατεύθυνσης θα είναι τότε σημαντική και θα πρέπει να ληφθεί υπόψη, τόσο στα σημεία του πειραματικού ημιβαριογράμματος όσο και στο μοντέλο που θα προσαρμοσθεί σ' αυτά. Για να συμπεριληφθεί σωστά η επίδραση της

κατεύθυνσης στο μοντέλο του ημιβαριογράμματος θα πρέπει να υπολογισθεί το ανισοτροπικό ημιβαριόγραμμα ή η ανισοτροπική συνδυασπορά.

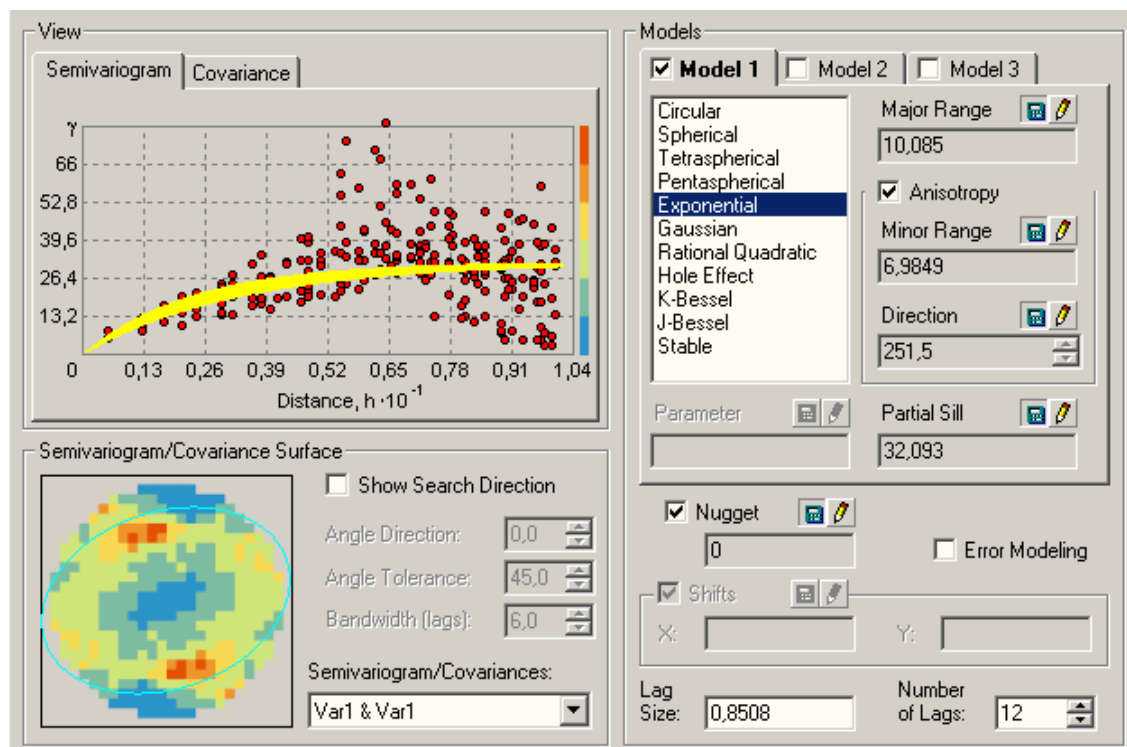
Στην παρούσα εργασία, εξετάστηκαν διάφορα μοντέλα ημιβαριογράμματος για την προσαρμογή στα πειραματικά ημιβαριογράμματα των δεδομένων βροχόπτωσης για τις τρεις υπό μελέτη ημέρες. Για την ανάλυση του ημιβαριογράμματος χρησιμοποιήθηκε το λογισμικό Geostatistical Analyst που είναι μια επέκταση του πακέτου ArcGIS και παρέχει την δυνατότητα διερεύνησης κατευθυντικών μεταβολών του ημιβαριογράμματος. Η ανισοτροπία της τυχαίας διαδικασίας εξετάστηκε χρησιμοποιώντας τους χάρτες ημιβαριογράμματος (ή επιφάνειες ημιβαριογράμματος) που είναι διαθέσιμοι μέσω του παραπάνω λογισμικού.

Για την προσαρμογή των μοντέλων (συναρτήσεων βαριογράμματος) στο πειραματικό ημιβαριόγραμμα. Διάφορα μοντέλα, δηλαδή διάφορες μαθηματικές συναρτήσεις, δοκιμάστηκαν να προσαρμοστούν στο πειραματικό ημιβαριόγραμμα. Η βέλτιστη προσαρμογή επιλέχθηκε με βάση την ελαχιστοποίηση της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Square Error) ή RMSE σφάλματος της πρόγνωσης που προκύπτει. Στην προσαρμογή του μοντέλου συνυπολογίστηκε και η ύπαρξη καθολικών τάσεων των δεδομένων.

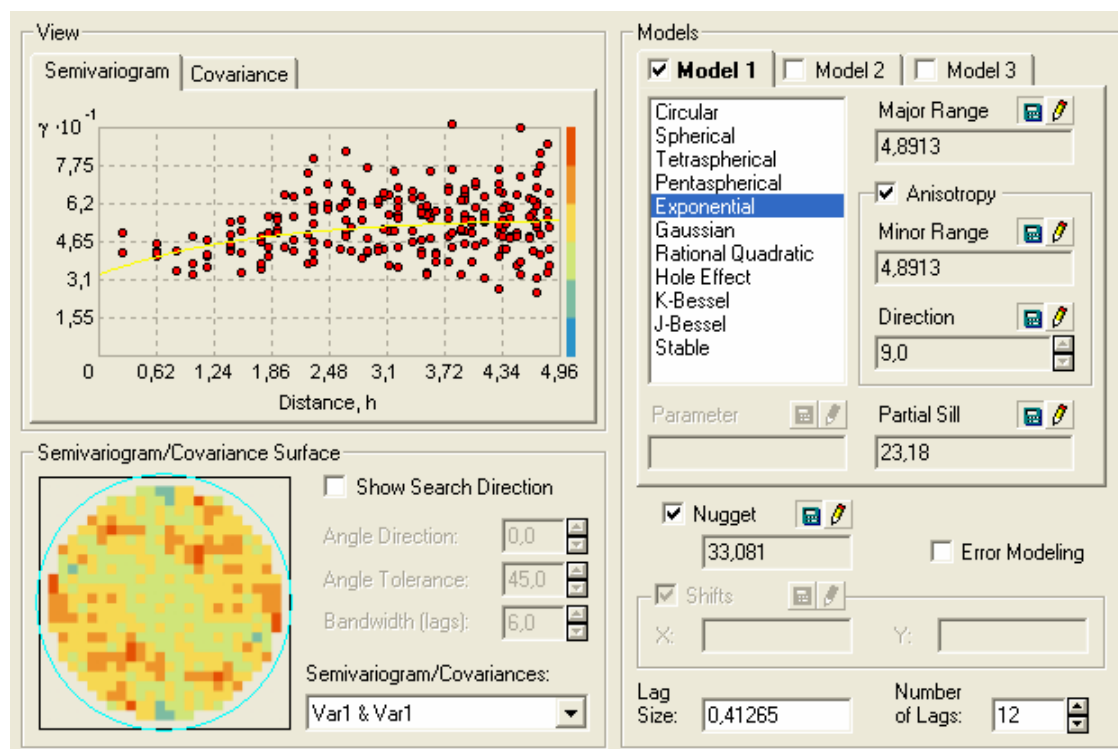
Στις Εικόνες 8α – 8γ παρουσιάζεται η ανάλυση του ημιβαριογράμματος για τις τρεις επιλεγμένες ημέρες που εξετάζονται στην παρούσα μελέτη.



Εικόνα 8α. Η ανάλυση του ημιβαριογράμματος για την ημέρα 28/10.



Εικόνα 8β. Η ανάλυση του ημιβαριογράμματος για την ημέρα 8/11.



Εικόνα 8γ. Η ανάλυση του ημιβαριογράμματος για την ημέρα 26/12.

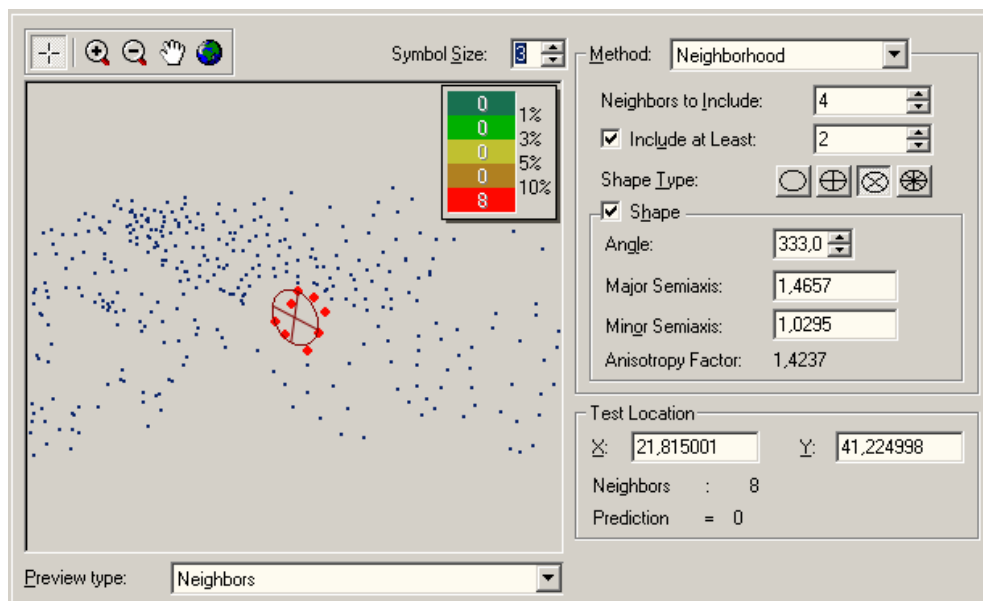
Στα αποτελέσματα που παρουσιάζονται στην παρούσα εργασία η ανισοτροπία των δεδομένων (όπου υπάρχει) έχει ενσωματωθεί στο μοντέλο έτσι ώστε να γίνεται διόρθωση ως προς την επιρροή της κατεύθυνσης της αυτοσυσχέτισης στην προκύπτουσα επιφάνεια, όπως φαίνεται στην επιφάνεια του ημιβαριογράμματος στις Εικόνες 8α-8γ.

Για τα δεδομένα που αφορούν την ημέρα με χαμηλή βροχόπτωση (28/10) η βέλτιστη προσαρμογή επιτεύχθηκε από ένα σφαιρικό μοντέλο ημιβαριογράμματος με τα χαρακτηριστικά που φαίνονται στην Εικόνα 8β. Για την ημέρα με μεσαία επίπεδα βροχόπτωσης (8/11) επιλέχθηκε ένα εκθετικό μοντέλο που ενσωματώνει την ανισοτροπία των δεδομένων όπως φαίνεται στην Εικόνα 8α. Τέλος για τα δεδομένα που αφορούν υψηλά επίπεδα βροχόπτωσης και αντιστοιχούν στην ημέρα 26/12 η χωρική τάση 2^{ης} τάξης που εμφανίζεται δεν έχει αφαιρεθεί από τα δεδομένα, αλλά έχει ενσωματωθεί σε ένα ανισοτροπικό εκθετικό μοντέλο ημιβαριογράμματος που φαίνεται στην Εικόνα 8γ.

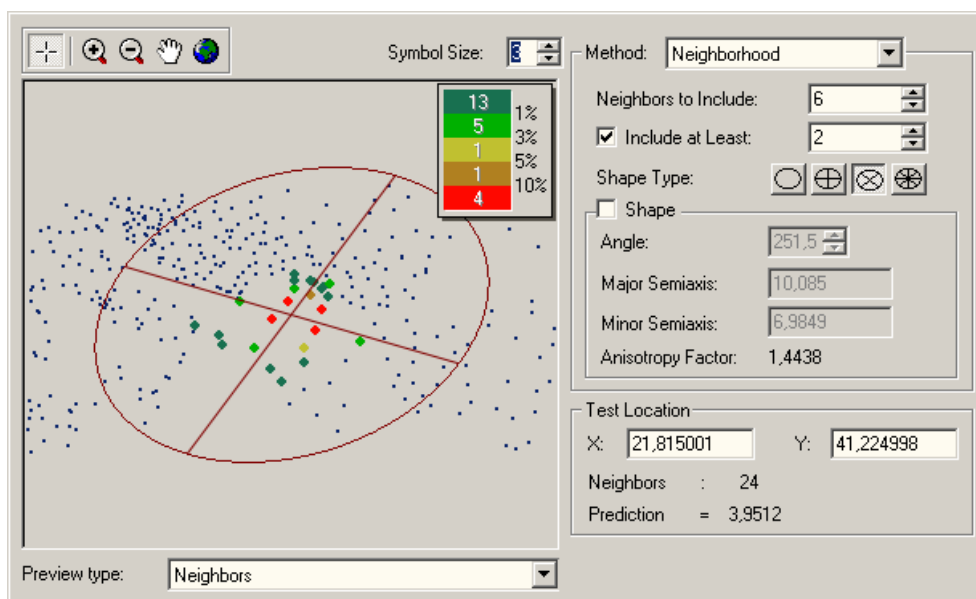
7.3 Η περιοχή επιρροής

Καθώς ο εκτιμητής kriging λειτουργεί ως ένας τελεστής κινητού παραθύρου (moving window operator) η επιλογή του παραθύρου αυτού είναι ένα σημαντικό βήμα της διαδικασίας εκτίμησης. Η επιλογή αυτή στην παρούσα μελέτη πραγματοποιήθηκε μέσω του πεδίου neighborhood search του λογισμικού Geostatistical Analyst.

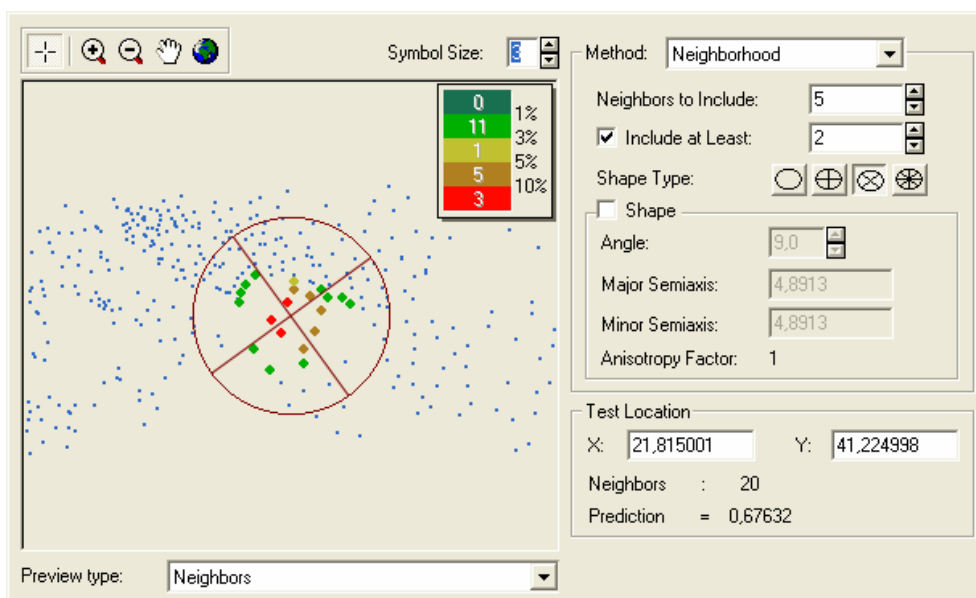
Είναι κοινή πρακτική να περιορίζονται οι μετρήσεις (δεδομένα) που θα χρησιμοποιηθούν για την πρόβλεψη ορίζοντας ένα κύκλο ή γενικά μια «περιοχή επιρροής» γύρω από ένα σημείο στο οποίο θα υπολογισθεί η τιμή της βροχόπτωσης με kriging. Μόνο τα δεδομένα που ανήκουν στην περιοχή αυτή λαμβάνονται υπόψη για τον υπολογισμό της τιμής στο σημείο όπου δεν υπάρχουν μετρήσεις. Η περιοχή αυτή ονομάζεται «περιοχή επιρροής» ή «γειτονιά».



Εικόνα 9β. Η περιοχή επιρροής για τα δεδομένα της ημέρας 28/10.



Εικόνα 9β. Η περιοχή επιρροής για τα δεδομένα της ημέρας 8/11.



Εικόνα 9γ. Η περιοχή επιρροής για τα δεδομένα της ημέρας 26/12.

Όταν υπάρχει ανισοτροπία στα δεδομένα, αντί για κύκλος, η περιοχή επιρροής μπορεί να έχει το σχήμα μιας έλλειψης έτσι ώστε να ακολουθεί την κατεύθυνση της ανισοτροπίας. Επίσης, για να αποφευχθεί η τυχόν μεροληψία προς μια συγκεκριμένη κατεύθυνση, ο κύκλος ή η έλλειψη μπορεί να χωριστεί σε τομείς, έτσι ώστε από κάθε τομέα να

επιλέγεται ίσος αριθμός σημείων. Η πρακτική αυτή ακολουθήθηκε και στην παρούσα μελέτη, όπως φαίνεται στις Εικόνες 9α -9γ, έτσι ώστε να διασφαλισθεί η αμεροληψία της πρόβλεψης.

Στις Εικόνες 9α, 9β και 9γ παρουσιάζεται η περιοχή επιρροής για τα δεδομένα που αφορούν την ημέρα με μεσαία, χαμηλά και υψηλά επίπεδα βροχόπτωσης, αντίστοιχα. Ο χρωματισμός των σημείων που απεικονίζονται μέσα στην περιοχή επιρροής δίνει μια ένδειξη του βάρους που θα έχουν τα δεδομένα αυτά κατά τον υπολογισμό της τιμής της πρόγνωσης στα σημεία όπου δεν υπάρχουν μετρήσεις. Στις παραπάνω εικόνες είναι φανερό η ενσωμάτωση της ανισοτροπίας από την ελλειπτική μορφή της περιοχής επιρροής. Επίσης, κάθε περιοχή χωρίστηκε, όπως φαίνεται, σε 4 τομείς ώστε να διασφαλισθεί κατά το δυνατόν η αμεροληψία.

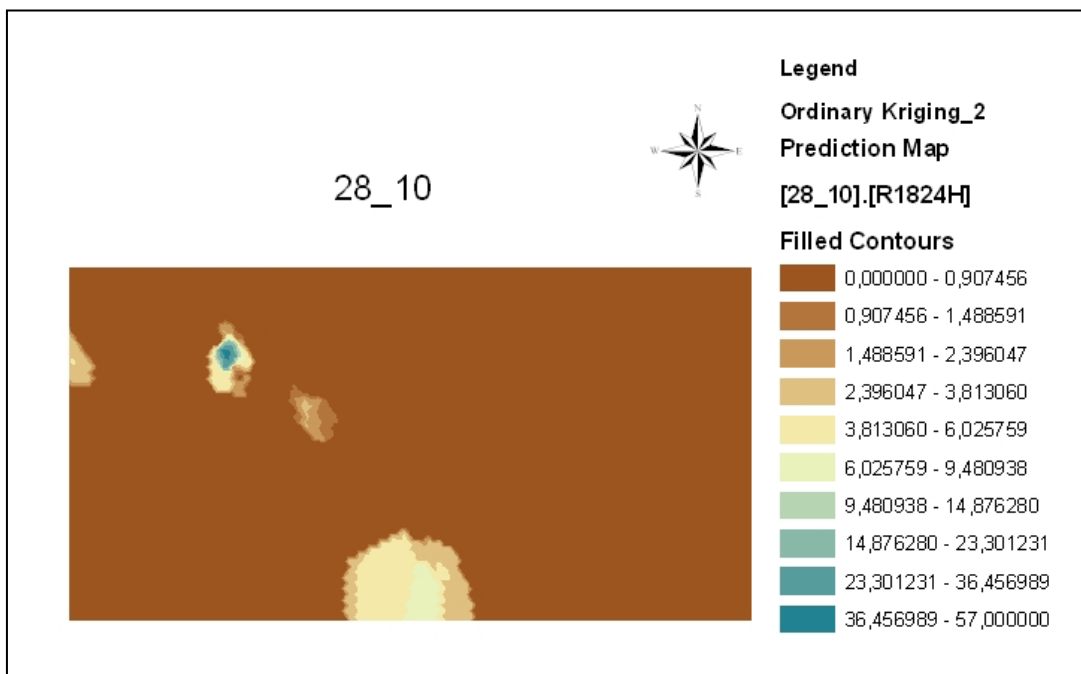
Κεφάλαιο 8

ΑΠΟΤΕΛΕΣΜΑΤΑ

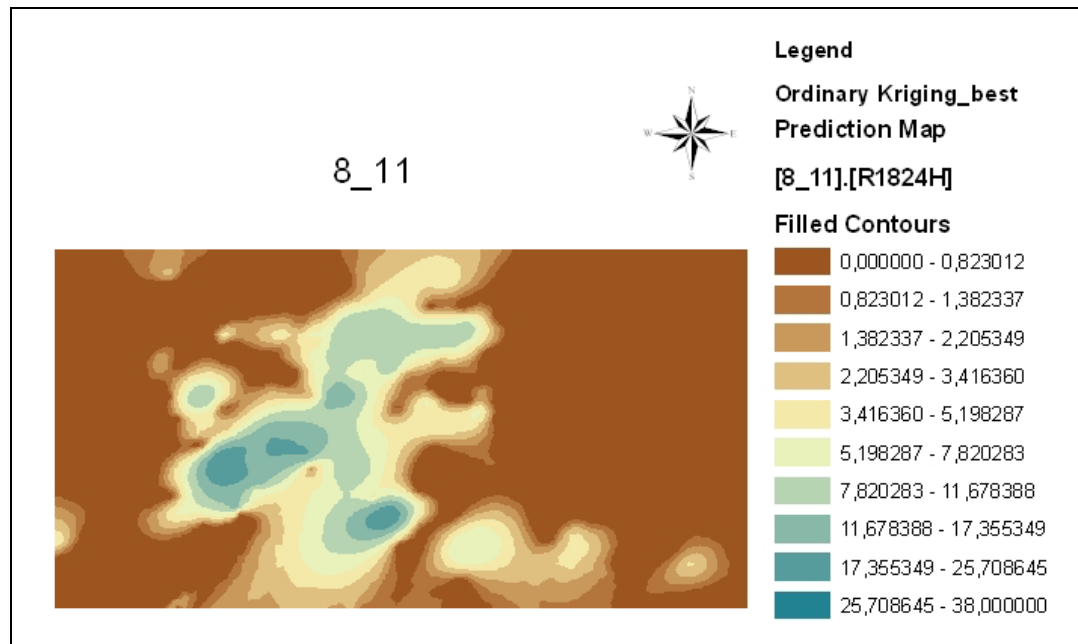
Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα της εφαρμογής της μεθόδου kriging στα δεδομένα βροχόπτωσης των τριών ημερών μελέτης. Στην συνέχεια τα αποτελέσματα αυτά επιβεβαιώνονται με cross-validation, validation σε ένα ποσοστό των δεδομένων, ενώ γίνεται και σύγκριση με τα αποτελέσματα που δίνουν άλλες μέθοδοι χωρικής παρεμβολής.

8.1 Χάρτες Εκτίμησης

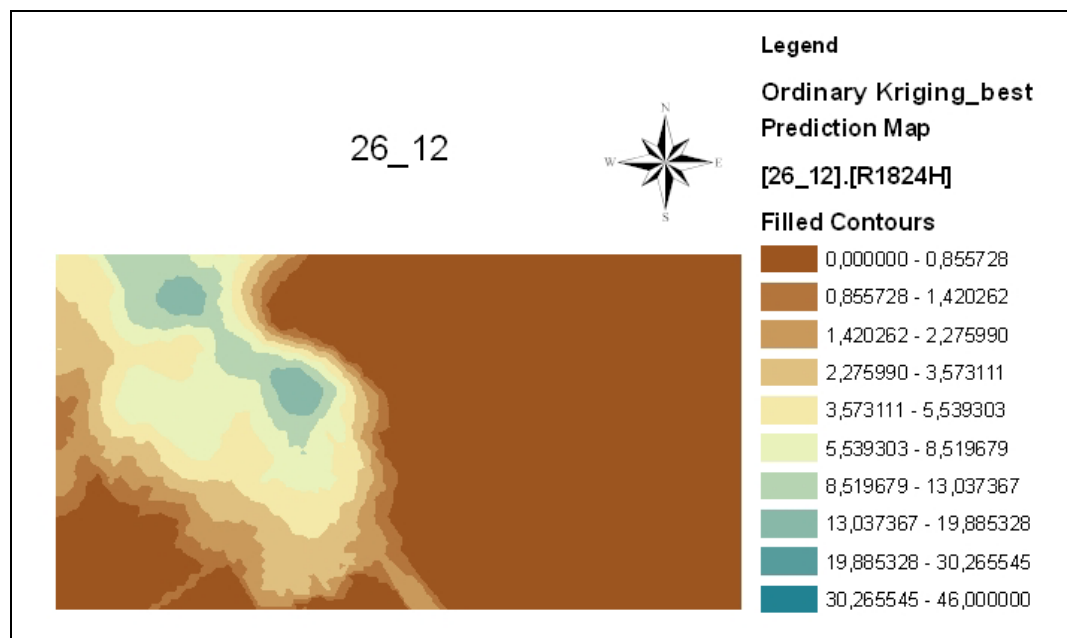
Μετά την ανάλυση που παρουσιάστηκε στο προηγούμενο κεφάλαιο, υπολογίστηκε η πρόγνωση με την μέθοδο Κανονικού Kriging και κατασκευάστηκαν οι χάρτες πρόγνωσης. Στις Εικόνες 10α – 10γ παρουσιάζονται οι χάρτες εκτίμησης που έδωσε η μέθοδος για τις τρεις ημέρες που εξετάζονται στην παρούσα μελέτη.



Εικόνα 10α. Ο χάρτης εκτίμησης για την ημέρα 28/10.



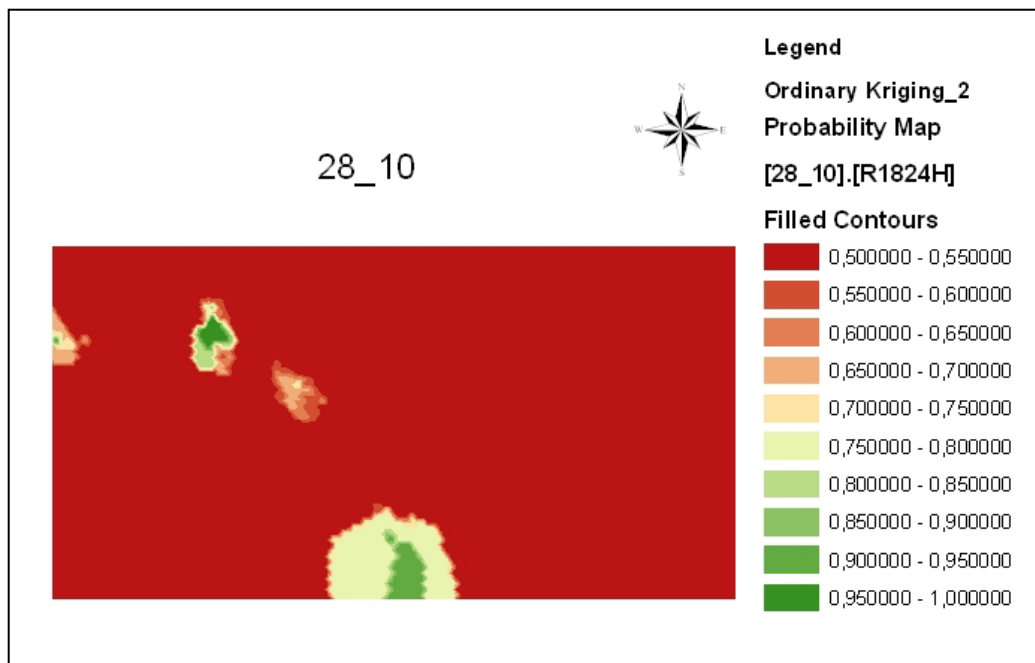
Εικόνα 10β. Ο χάρτης εκτίμησης για την ημέρα 8/11.



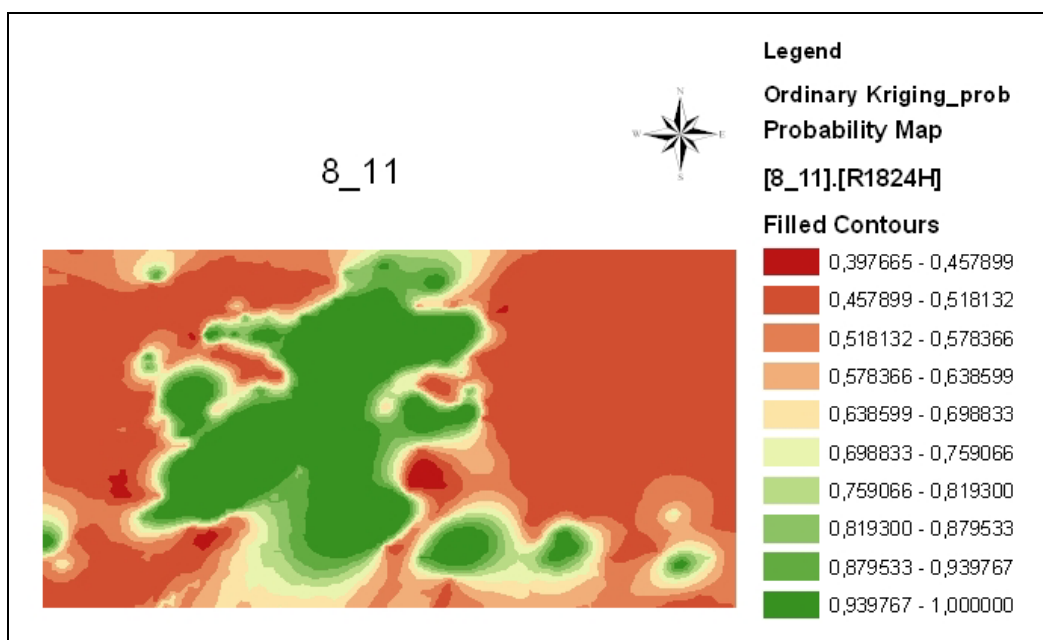
Εικόνα 10γ. Ο χάρτης εκτίμησης για την ημέρα 26/12.

8.2 Χάρτες Πιθανότητας

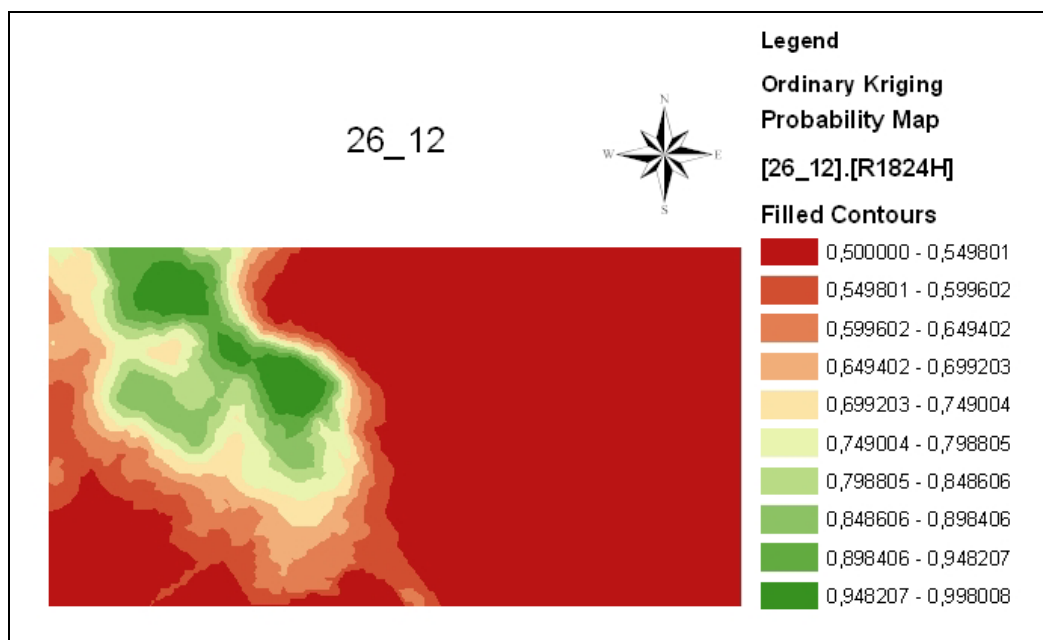
Στην συνέχεια κατασκευάστηκαν οι χάρτες πιθανότητας για κάθε μια από τις τρεις ημέρες που εξετάζονται στην παρούσα μελέτη. Στις Εικόνες 11α, 11β και 11γ παρουσιάζονται οι χάρτες πιθανότητας για τις ημέρες 8/11, 28/10 και 26/12, αντίστοιχα.



Εικόνα 11α. Ο χάρτης πιθανότητας για την ημέρα 28/10.



Εικόνα 11β. Ο χάρτης πιθανότητας για την ημέρα 8/11.



Εικόνα 11γ. Ο χάρτης πιθανότητας για την ημέρα 26/12.

8.3 Επικύρωση των αποτελεσμάτων - Validation

Για την επικύρωση των αποτελεσμάτων δύο τεχνικές χρησιμοποιήθηκαν. Αρχικά χρησιμοποιήθηκε μια μέθοδος αυτό-επικύρωσης (cross-validation) η οποία χρησιμοποιεί όλα τα δεδομένα για την κατασκευή του μοντέλου αυτοσυσχέτισης. Στην συνέχεια έγινε επικύρωση των αποτελεσμάτων σε ένα ποσοστό των δεδομένων χρησιμοποιώντας τα υπόλοιπα δεδομένα για τον υπολογισμό της εκτίμησης με την μέθοδο Kriging.

8.3.1 Αυτό-επικύρωση (cross-validation)

Όπως έχει γίνει φανερό, η εφαρμογή της μεθόδου kriging απαιτεί τον καθορισμό μιας πληθώρας παραμέτρων. Δυστυχώς, δεν υπάρχει κάποιος γενικός κανόνας για τον βέλτιστο καθορισμό των παραμέτρων αυτών. Παρόλα αυτά, σύμφωνα με τον Davis (1987), η λεγόμενη τεχνική της αυτό-επικύρωσης χρησιμοποιείται συχνά για την επιλογή του βέλτιστου εκτιμητή. Η μέθοδος αυτή βασίζεται στην ακόλουθη διαδικασία: Κάθε φορά αφαιρείται ένα σημείο μέτρησης από το σύνολο των δεδομένων και γίνεται παρεμβολή για τον καθορισμό της τιμής της βροχόπτωσης στο σημείο αυτό χρησιμοποιώντας τα υπόλοιπα δεδομένα (π.χ. σαν να μην υπήρχε μέτρηση στο σημείο που αφαιρέθηκε). Στην συνέχεια υπολογίζεται η διαφορά της πραγματικής τιμής μέτρησης από την τιμή που υπολογίστηκε για το σημείο αυτό από την παρεμβολή των υπολοίπων δεδομένων. Η

διαδικασία αυτή επαναλαμβάνεται για όλα τα σημεία μέτρησης, με διαδοχική αφαίρεση ενός σημείου κάθε φορά.

Η παραπάνω τεχνική παράγει μια ένδειξη του «πόσο καλά» το μοντέλο που έχει κατασκευαστεί με τη μέθοδο kriging προβλέπει την τιμή της βροχόπτωσης σε σημεία που δεν είναι γνωστή η τιμή της. Για όλα τα σημεία, η τεχνική της αυτό-επικύρωσης αφαιρεί διαδοχικά ένα σημείο, προβλέπει την τιμή του χρησιμοποιώντας τα υπόλοιπα δεδομένα και συγκρίνει την πραγματική τιμή μέτρησης με την τιμή της πρόβλεψης. Τα στατιστικά χαρακτηριστικά που προκύπτουν βοηθούν στην εκτίμηση του πόσο καλό είναι το μοντέλο.

Πίνακας 1. Αυτό-επικύρωση των αποτελεσμάτων για την ημέρα 28/10.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	0.01042
Σφάλμα RMSE	2.587
Μέσο Τυπικό Σφάλμα	4.871
Μέσο Τυποποιημένο Σφάλμα	0.006488
Σφάλμα RMSSE	0.5195

Πίνακας 2. Αυτό-επικύρωση των αποτελεσμάτων για την ημέρα 8/11.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.009617
Σφάλμα RMSE	3.543
Μέσο Τυπικό Σφάλμα	2.533
Μέσο Τυποποιημένο Σφάλμα	0.0003817
Σφάλμα RMSSE	1.259

Πίνακας 3. Αυτό-επικύρωση των αποτελεσμάτων για την ημέρα 26/12.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	0.0299
Σφάλμα RMSE	4.805
Μέσο Τυπικό Σφάλμα	6.739
Μέσο Τυποποιημένο Σφάλμα	0.002977
Σφάλμα RMSSE	0.7349

Στην παρούσα μελέτη εφαρμόστηκε η μέθοδος της αυτό-επικύρωσης για τα μοντέλα που κατασκευάστηκαν για τις τρεις επιλεγμένες ημέρες με την μέθοδο kriging. Στους Πίνακες 1 έως 3 παρουσιάζονται τα στατιστικά μεγέθη που υπολογίστηκαν κατά την διαδικασία αυτό-επικύρωσης για τις τρεις υπό εξέταση ημέρες.

Για κάθε περίπτωση δίνονται: το μέσο σφάλμα (mean error), η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error) ή αλλιώς σφάλμα RMSE, το μέσο τυπικό σφάλμα (average standard error), το μέσο τυποποιημένο σφάλμα (mean standardized error), και η τετραγωνική ρίζα του μέσου τετραγωνικού τυποποιημένου σφάλματος (Root Mean Square Standardised Error) ή εν συντομία σφάλμα RMSSE.

Για να θεωρείται ένα μοντέλο ότι παράγει ακριβείς προγνώσεις, το μέσο σφάλμα (Μέσο Σφάλμα) θα πρέπει να είναι κοντά στο μηδέν, το τετραγωνικό μέσο σφάλμα (Σφάλμα RMSE) και το μέσο κανονικό σφάλμα (Μέσο Τυπικό Σφάλμα) πρέπει να είναι όσο το δυνατόν πιο μικρά, και η τιμή του root-mean-square standardized error να είναι κοντά στην μονάδα.

Εδώ ο όρος σφάλμα πρόβλεψης (prediction error) χρησιμοποιείται για την διαφορά της πρόβλεψης από την πραγματική τιμή της μέτρησης. Για ένα μοντέλο που παράγει ακριβείς προβλέψεις, το μέσο σφάλμα πρόβλεψης (mean prediction error) πρέπει να είναι κοντά στο μηδέν αν οι προβλέψεις είναι αμερόληπτες. Επίσης, η τετραγωνική ρίζα του μέσου τετραγωνικού τυποποιημένου σφάλματος (Root Mean Square Standardized Error), ή εν συντομία σφάλμα RMSSE της πρόβλεψης, πρέπει να είναι κοντά στην μονάδα αν τα τυπικά σφάλματα (standard errors) είναι ακριβή. Τέλος, η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error) της πρόβλεψης, ή εν συντομία σφάλμα RMSE, πρέπει να είναι μικρό αν οι προβλέψεις είναι κοντά στις τιμές μέτρησης.

Στην παρούσα μελέτη, η συνολική απόδοση ενός μοντέλου κρίθηκε με βάση το

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z_{i(int)} - Z_i)^2}{n}}$$

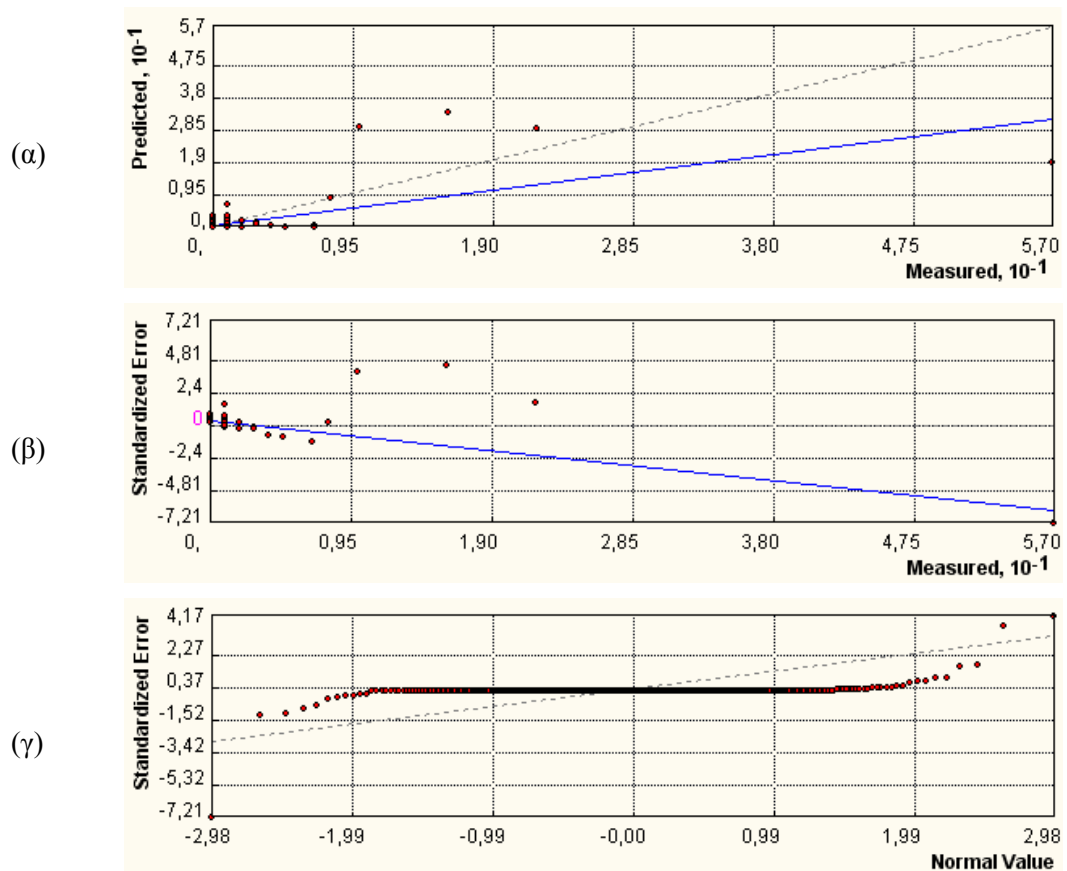
όπου με RMSE συμβολίζεται η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error), $Z_{i(int)}$ είναι η τιμή της παρεμβλλόμενης τιμής της βροχόπτωσης στο σημείο i υπολογισμένη με βάση τα υπόλοιπα $n-1$ σημεία, Z_i είναι η μετρημένη τιμή της βροχόπτωσης στο σημείο i , και n είναι το πλήθος των σημείων των δεδομένων.

Χαμηλές τιμές του σφάλματος RMSE αποτελούν ένδειξη ότι η παρεμβολή θα δώσει αξιόπιστες εκτιμήσεις για τις περιοχές όπου η ένταση της βροχόπτωσης δεν είναι γνωστή.

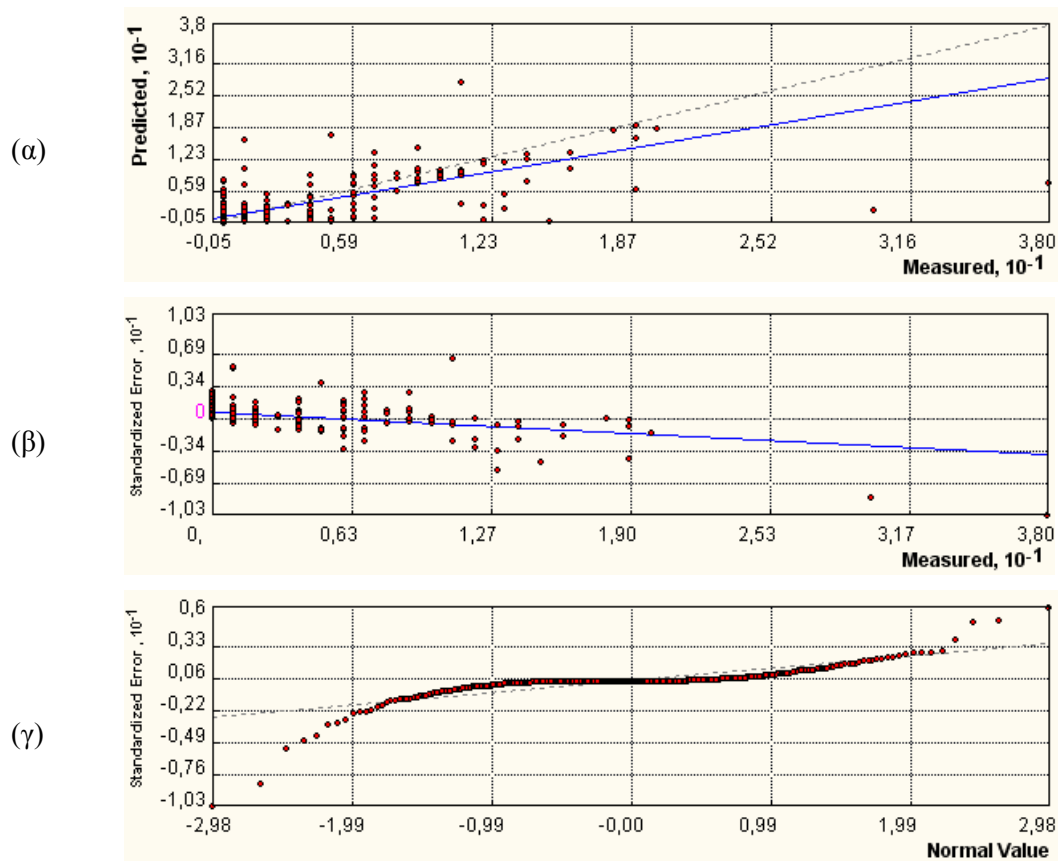
Στην παρούσα μελέτη η παραπάνω διαδικασία της αυτό-επικύρωσης των αποτελεσμάτων επαναλήφθηκε για διάφορες τιμές των παραμέτρων που απαιτούνται κατά την εφαρμογή της μεθόδου kriging. Στην συνέχεια, τα αποτελέσματα αξιολογήθηκαν με κριτήριο την τιμή του σφάλματος RMSE. Ως βέλτιστο επιλέχθηκε το σετ των παραμέτρων που

παρήγαγε το χαμηλότερο RMSE σφάλμα κατά της διαδικασίας της αυτό-επικύρωσης των αποτελεσμάτων.

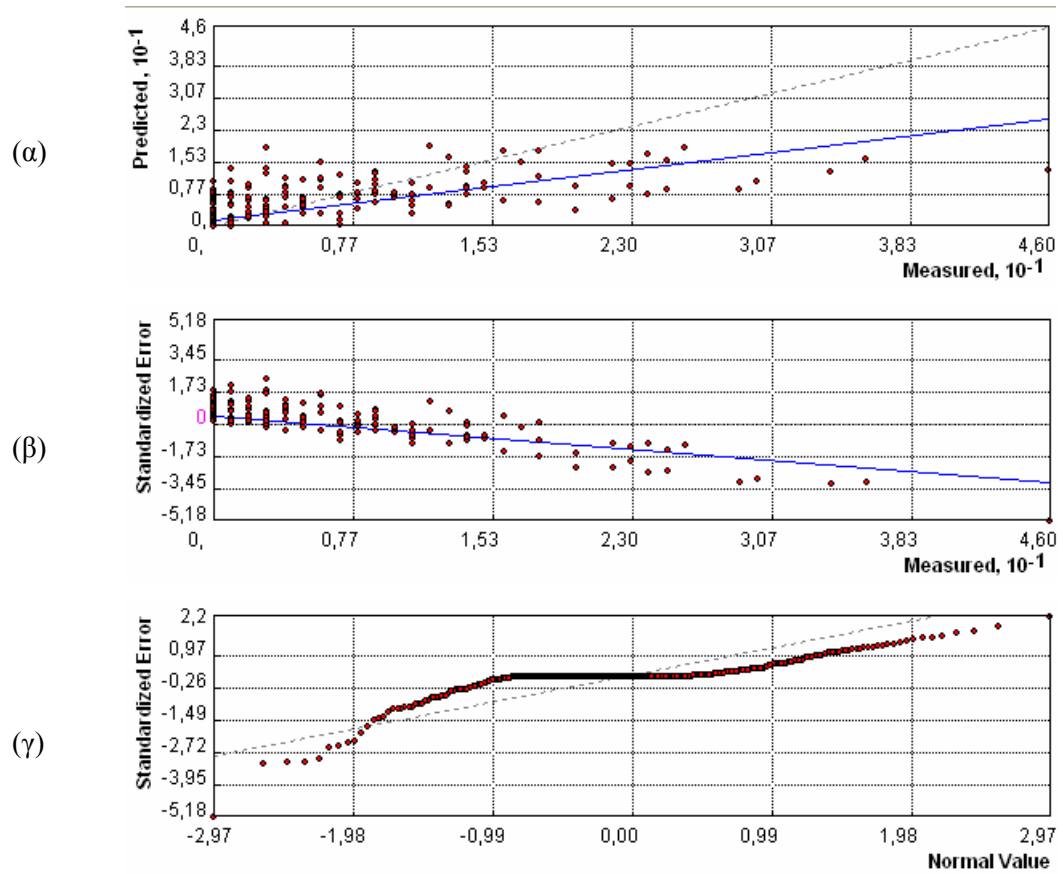
Στις Εικόνες 12α -12γ παρουσιάζονται οι γραφικές παραστάσεις που προκύπτουν από την ανάλυση σφάλματος της διαδικασίας, δηλ. το σφάλμα πρόγνωσης (α), το τυποποιημένο (standardized) σφάλμα (β) και το διάγραμμα QQplot (γ).



Εικόνα 12α. Ανάλυση σφάλματος από την αυτό-επικύρωση των αποτελεσμάτων για την ημέρα 28/10: σφάλμα πρόγνωσης (α), τυποποιημένο σφάλμα (β) και QQplot (γ).



Εικόνα 12β. Ανάλυση σφάλματος από την αυτό-επικύρωση των αποτελεσμάτων για την ημέρα 8/11: σφάλμα πρόγνωσης (α), τυποποιημένο σφάλμα (β) και QQplot (γ).

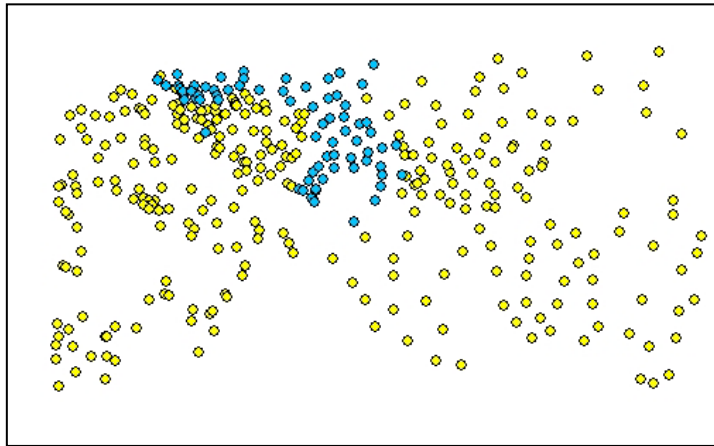


Εικόνα 12γ. Ανάλυση σφάλματος από την αυτό-επικύρωση των αποτελεσμάτων για την ημέρα 26/12: σφάλμα πρόγνωσης (α), τυποποιημένο σφάλμα (β) και QQplot (γ).

8.3.2 Επικύρωση στο 20% των αποτελεσμάτων

Στην συνέχεια ακολουθήθηκε μια άλλη διαδικασία επικύρωσης των αποτελεσμάτων.

Τα δεδομένα χωρίστηκαν σε δύο υποσύνολα. Ένα ποσοστό 20% των δεδομένων διαχωρίστηκε και διαμόρφωσε το λεγόμενο «σύνολο ελέγχου» (test dataset), ενώ το υπόλοιπο 80% των δεδομένων αποτέλεσε το λεγόμενο «σύνολο εκπαίδευσης» (training dataset. Στην συνέχεια, το σύνολο εκπαίδευσης χρησιμοποιήθηκε για την κατασκευή του μοντέλου πρόγνωσης με την μέθοδο kriging, ενώ το σύνολο ελέγχου χρησιμοποιήθηκε για την επικύρωση των αποτελεσμάτων του μοντέλου αυτού. Δηλαδή το μοντέλο που κατασκευάστηκε με βάση το 80% των δεδομένων, χρησιμοποιήθηκε για να προβλέψει την τιμή της βροχόπτωσης στο υπόλοιπο 20% των δεδομένων.



Εικόνα 13α. Σύνολο εκπαίδευσης 80% (κίτρινο χρώμα) και σύνολο ελέγχου 20% (μπλέ χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 28/10.

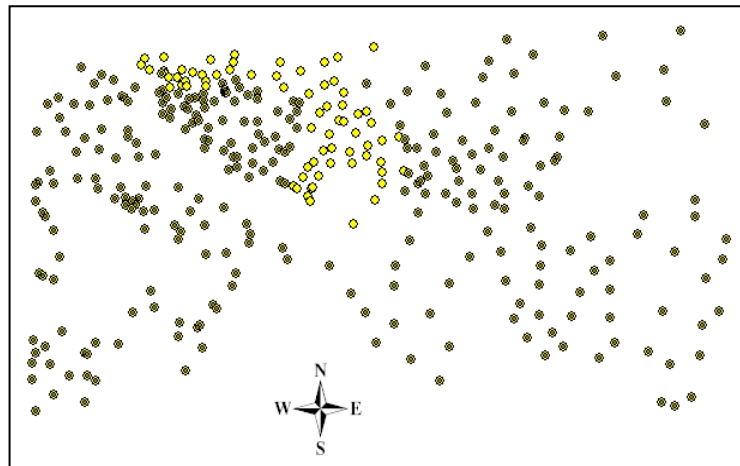
Πίνακας 4α. Επικύρωση στο 20% (με αυθαίρετη επιλογή) για την ημέρα 28/10.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.3882
Σφάλμα RMSE	5.071
Μέσο Τυπικό Σφάλμα	3.226
Μέσο Τυποποιημένο Σφάλμα	-0.2598
Σφάλμα RMSSE	2.685

Ένα ερώτημα που τίθεται στο σημείο αυτό είναι ο τρόπος επιλογής των σημείων που θα αποτελέσουν το σύνολο εκπαίδευσης. Στην παρούσα εργασία δύο διαφορετικοί τρόποι επιλογής ακολουθήθηκαν. Ο πρώτος έγκειται στην αυθαίρετη επιλογή των σημείων, ενώ

ο δεύτερος στην επιλογή των σημείων με έτσι ώστε να επιλέγονται σημεία από όλη την γεωγραφική περιοχή.

Στις Εικόνες 13α, 14α και 15α απεικονίζονται τα σύνολα εκπαίδευσης και τα σύνολα ελέγχου που προέκυψαν από αυθαίρετη επιλογή για τις ημέρες 28/10, 8/11 και 26/12, αντίστοιχα. Στους Πίνακες 4α έως 6α παρουσιάζονται τα σφάλματα πρόγνωσης που προέκυψαν από την διαδικασία αυτή με αυθαίρετη επιλογή των σημείων για τις τρεις υπό εξέταση ημέρες.

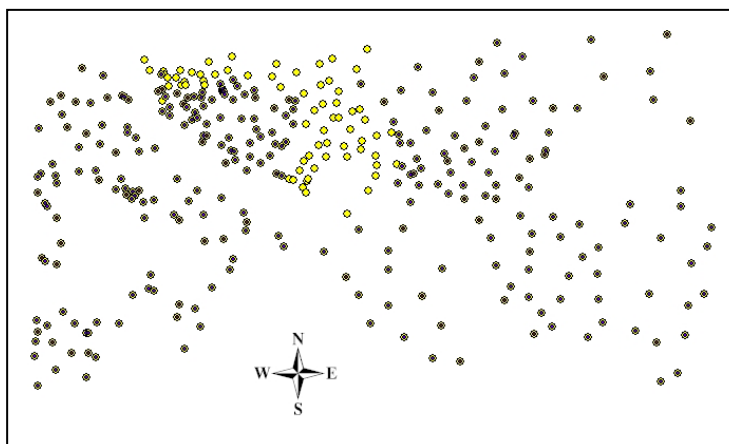


Εικόνα 14α. Σύνολο εκπαίδευσης 80% (καφέ χρώμα) και σύνολο ελέγχου 20% (κίτρινο χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 8/11.

Πίνακας 5α. Επικύρωση στο 20% (με αυθαίρετη επιλογή) για την ημέρα 8/11.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-1.308
Σφάλμα RMSE	3.855
Μέσο Τυπικό Σφάλμα	3.271
Μέσο Τυποποιημένο Σφάλμα	-0.3853
Σφάλμα RMSSE	1.114

Για κάθε περίπτωση δίνονται: το μέσο σφάλμα (mean error), η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error) ή αλλιώς σφάλμα RMSE, το μέσο τυπικό σφάλμα (average standard error), το μέσο τυποποιημένο σφάλμα (mean standardized error), και η τετραγωνική ρίζα του μέσου τετραγωνικού τυποποιημένου σφάλματος (Root Mean Square Standarised Error) ή εν συντομία σφάλμα RMSSE.



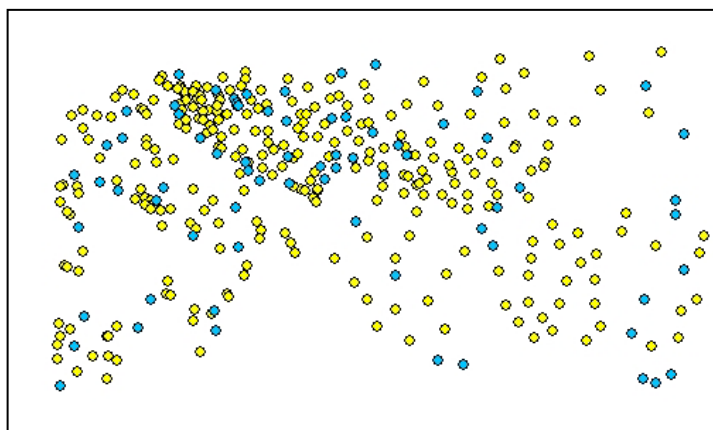
Εικόνα 15α. Σύνολο εκπαίδευσης 80% (καφέ χρώμα) και σύνολο ελέγχου 20% (κίτρινο χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 26/12.

Πίνακας 6α. Επικύρωση στο 20% (με αυθαίρετη επιλογή) για την ημέρα 26/12.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	1.904
Σφάλμα RMSE	6.485
Μέσο Τυπικό Σφάλμα	6.287
Μέσο Τυποποιημένο Σφάλμα	0.2905
Σφάλμα RMSSE	1.083

Από τους πίνακες 4α έως 6α παρατηρούμε ότι η διαδικασία επικύρωσης στο 20% (με αυθαίρετη επιλογή) δίνει υψηλότερα RMSE σφάλματα από τα αντίστοιχα σφάλματα της διαδικασίας αυτοεπικύρωσης.

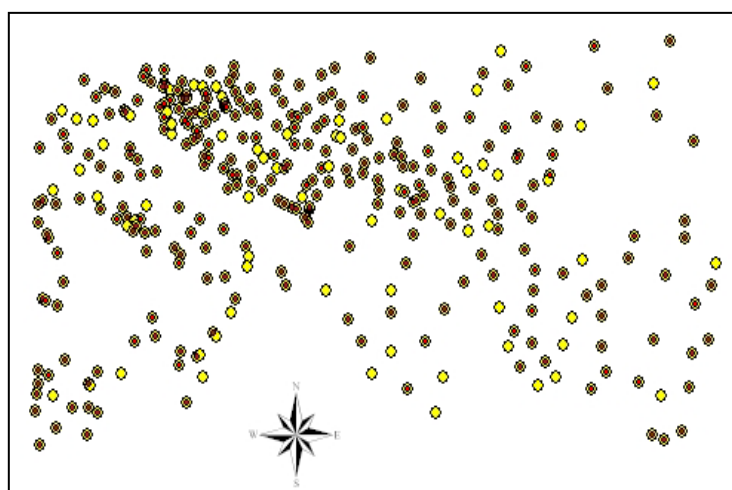
Στην συνέχεια η διαδικασία επαναλήφθηκε, αλλά η επιλογή των σημείων των δύο υποσυνόλων έγινε με τρόπο ώστε να είναι ομοιόμορφα γεωγραφικά κατανεμημένα στην υπό μελέτη περιοχή. Στις Εικόνες 13β, 14β και 15β απεικονίζονται τα σύνολα εκπαίδευσης και τα σύνολα ελέγχου που προέκυψαν από την ομοιόμορφη επιλογή για τις ημέρες 28/10, 8/11 και 26/12, αντίστοιχα. Στους Πίνακες 4β έως 6β παρουσιάζονται τα αντίστοιχα σφάλματα πρόγνωσης που προέκυψαν από την διαδικασία επικύρωσης στο 20% με ομοιόμορφη (γεωγραφικά) επιλογή των σημείων για τις τρεις υπό εξέταση ημέρες.



Εικόνα 13β. Σύνολο εκπαίδευσης 80% (κίτρινο χρώμα) και σύνολο ελέγχου 20% (μπλέ χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 28/10.

Πίνακας 4β. Επικύρωση στο 20% (με ομοιόμορφη επιλογή) για την ημέρα 28/10.

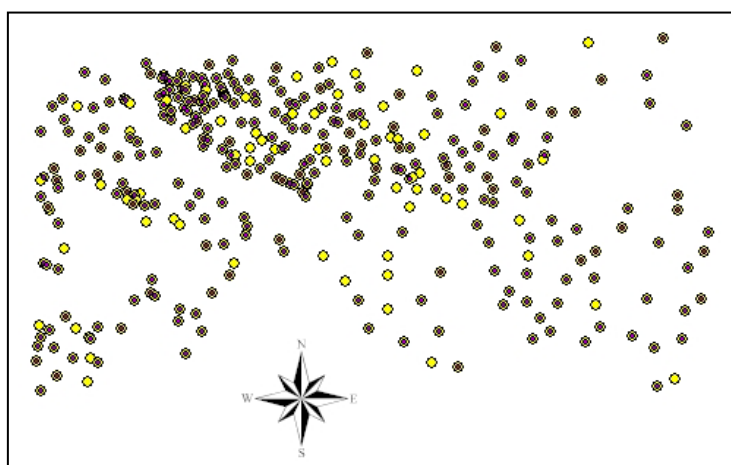
ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	0.3304
Σφάλμα RMSE	2.27
Μέσο Τυπικό Σφάλμα	4.751
Μέσο Τυποποιημένο Σφάλμα	0.06825
Σφάλμα RMSSE	0.4734



Εικόνα 14β. Σύνολο εκπαίδευσης 80% (καφέ χρώμα) και σύνολο ελέγχου 20% (κίτρινο χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 8/11.

Πίνακας 5β. Επικύρωση στο 20% (με ομοιόμορφη επιλογή) για την ημέρα 8/11.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	0.4665
Σφάλμα RMSE	3.385
Μέσο Τυπικό Σφάλμα	2.769
Μέσο Τυποποιημένο Σφάλμα	0.1926
Σφάλμα RMSSE	1.113



Εικόνα 15β. Σύνολο εκπαίδευσης 80% (καφέ χρώμα) και σύνολο ελέγχου 20% (κίτρινο χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 26/12.

Πίνακας 6β. Επικύρωση στο 20% για την ημέρα 26/12.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.235
Σφάλμα RMSE	5.32
Μέσο Τυπικό Σφάλμα	6.888
Μέσο Τυποποιημένο Σφάλμα	-0.03143
Σφάλμα RMSSE	0.7884

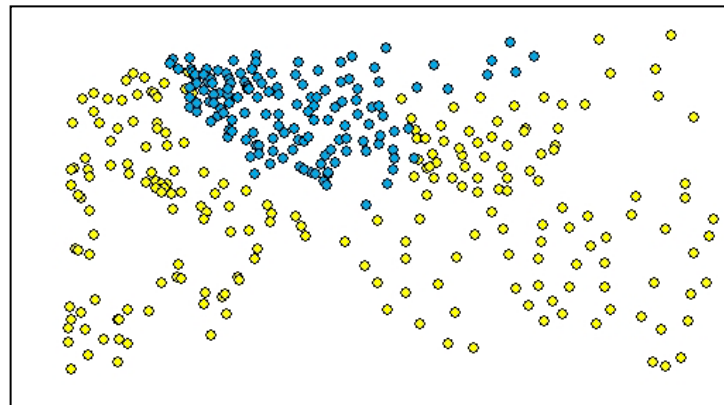
Παρατηρούμε ότι η ομοιόμορφη (γεωγραφικά) επιλογή των σημείων δίνει καλύτερα αποτελέσματα. Αυτό είναι αναμενόμενο καθώς το σύνολο του 80% των σημείων πρέπει να είναι αντιπροσωπευτικό του αρχικού συνόλου.

8.3.3 Επικύρωση στο 40% των αποτελεσμάτων

Στην συνέχεια, για κάθε μία από τις επιλεγμένες ημέρες, τα δεδομένα διαχωρίστηκαν σε υποσύνολα που περιλάμβαναν το 40% των δεδομένων σαν σύνολα ελέγχου, ενώ το υπόλοιπο 60% των δεδομένων χρησιμοποιήθηκαν σαν σύνολο εκπαίδευσης, δηλαδή βάση για την δημιουργία του μοντέλου πρόβλεψης με την μέθοδο kriging. Η επικύρωση των αποτελεσμάτων έγινε στην συνέχεια συγκρίνοντας τις τιμές πρόβλεψης με τα δεδομένα των συνόλων ελέγχου.

Στις Εικόνες 16α, 17α και 18α απεικονίζονται τα σύνολα εκπαίδευσης και τα σύνολα ελέγχου που προέκυψαν από αυθαίρετη επιλογή του 40% για τις ημέρες 28/10, 8/11 και 26/12, αντίστοιχα.

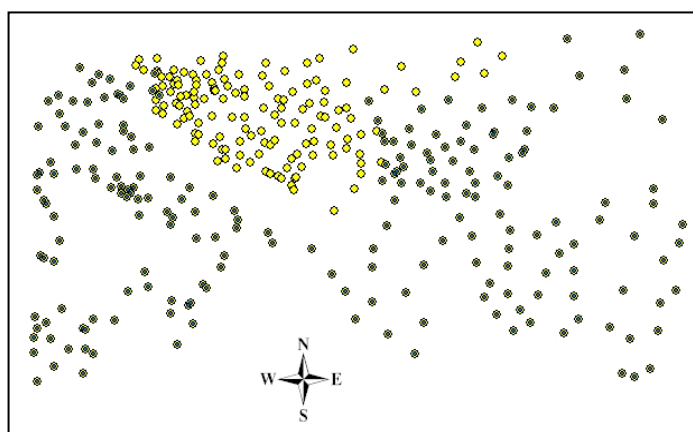
Τα σφάλματα πρόγνωσης από την διαδικασία επικύρωσης στο 40% των δεδομένων (με αυθαίρετη επιλογή) παρουσιάζονται στους Πίνακες 7α, 8α και 9α για τις ημέρες 8/11, 28/10 και 26/12, αντίστοιχα.



Εικόνα 16α. Σύνολο εκπαίδευσης 60% (κίτρινο χρώμα) και σύνολο ελέγχου 40% (μπλέ χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 28/10.

Πίνακας 7α. Επικύρωση στο 40% (με αυθαίρετη επιλογή) για την ημέρα 28/10.

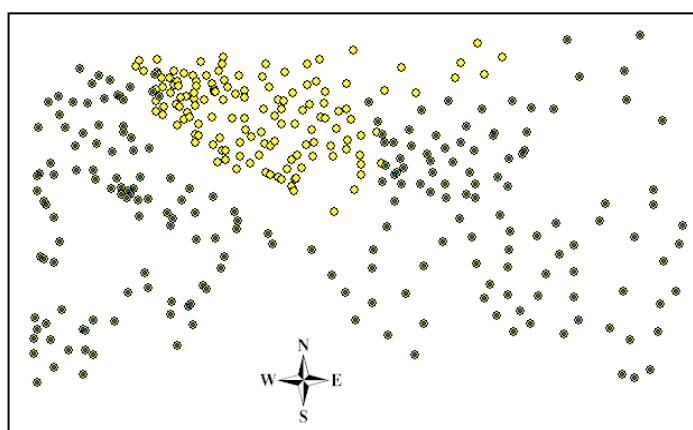
ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.9645
Σφάλμα RMSE	5.46
Μέσο Τυπικό Σφάλμα	0.6846
Μέσο Τυποποιημένο Σφάλμα	-1.613
Σφάλμα RMSSE	9.292



Εικόνα 17α. Σύνολο εκπαίδευσης 60% (καφέ χρώμα) και σύνολο ελέγχου 40% (κίτρινο χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 8/11.

Πίνακας 8α. Επικύρωση στο 40% (με αυθαίρετη επιλογή) για την ημέρα 8/11.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	3.757
Σφάλμα RMSE	5.442
Μέσο Τυπικό Σφάλμα	4.696
Μέσο Τυποποιημένο Σφάλμα	0.7904
Σφάλμα RMSSE	1.151

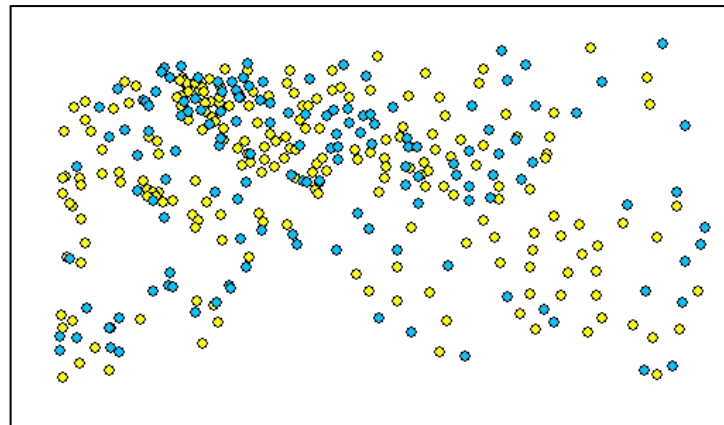


Εικόνα 18α. Σύνολο εκπαίδευσης 60% (καφέ χρώμα) και σύνολο ελέγχου 40% (κίτρινο χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 26/12.

Πίνακας 9α. Επικύρωση στο 40% (με αυθαίρετη επιλογή) για την ημέρα 26/12.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-1.147
Σφάλμα RMSE	8.34
Μέσο Τυπικό Σφάλμα	3.781
Μέσο Τυποποιημένο Σφάλμα	-0.3399
Σφάλμα RMSSE	2.26

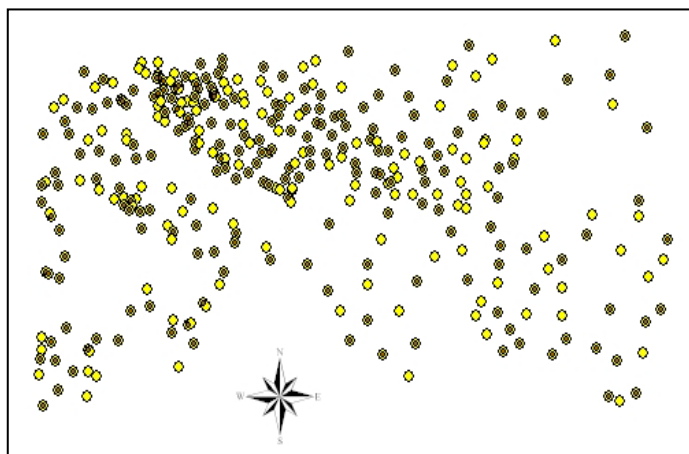
Στην συνέχεια η διαδικασία επαναλήφθηκε, αλλά η επιλογή των σημείων των δύο υποσυνόλων έγινε με τρόπο ώστε να είναι ομοιόμορφα γεωγραφικά κατανεμημένα στην υπό μελέτη περιοχή.



Εικόνα 16β. Σύνολο εκπαίδευσης 60% (κίτρινο χρώμα) και σύνολο ελέγχου 40% (μπλέ χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 28/10.

Πίνακας 7β. Επικύρωση στο 40% (ομοιόμορφη επιλογή για την ημέρα 28/10).

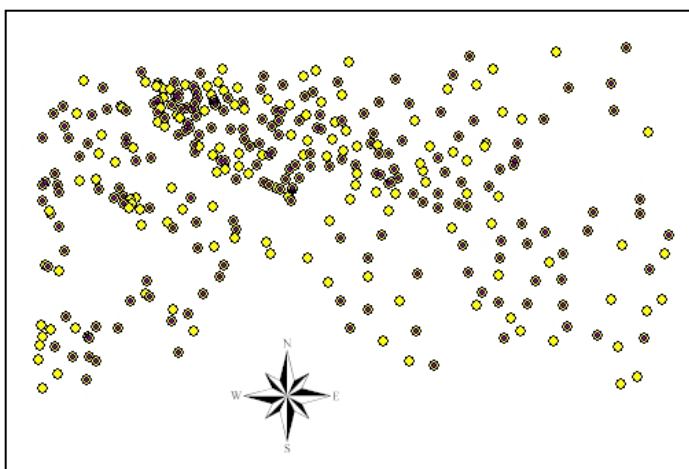
ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.2948
Σφάλμα RMSE	3.763
Μέσο Τυπικό Σφάλμα	2.875
Μέσο Τυποποιημένο Σφάλμα	-0.1757
Σφάλμα RMSSE	2.246



Εικόνα 17β. Σύνολο εκπαίδευσης 60% (καφέ χρώμα) και σύνολο ελέγχου 40% (κίτρινο χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 8/11.

Πίνακας 8β. Επικύρωση στο 40% (ομοιόμορφη επιλογή) για την ημέρα 8/11.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.03588
Σφάλμα RMSE	3.703
Μέσο Τυπικό Σφάλμα	2.821
Μέσο Τυποποιημένο Σφάλμα	-0.02818
Σφάλμα RMSSE	1.259



Εικόνα 18β. Σύνολο εκπαίδευσης 60% (κίτρινο χρώμα) και σύνολο ελέγχου 40% (μπλέ χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 26/12.

Πίνακας 9β. Επικύρωση στο 40% (ομοιόμορφη επιλογή για την ημέρα 26/12.

ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ	
Μέσο Σφάλμα	-0.936
Σφάλμα RMSE	6.259
Μέσο Τυπικό Σφάλμα	5.575
Μέσο Τυποποιημένο Σφάλμα	-0.1893
Σφάλμα RMSSE	1.214

Στις Εικόνες 16β, 17β και 18β απεικονίζονται τα σύνολα εκπαίδευσης (60%) και τα σύνολα ελέγχου (40%) που προέκυψαν από την ομοιόμορφη επιλογή για τις ημέρες 28/10, 8/11 και 26/12, αντίστοιχα. Στους Πίνακες 7β, 8β και 9β παρουσιάζονται τα αντίστοιχα σφάλματα πρόγνωσης από την διαδικασία επικύρωσης όταν η επιλογή των σημείων γίνεται ομοιόμορφα από όλη την γεωγραφική περιοχή.

Όπως φαίνεται από τους παραπάνω πίνακες τα RMSE σφάλματα είναι, όπως αναμένεται, μεγαλύτερα από τα αντίστοιχα της διαδικασίας επικύρωσης στο 20% των αποτελεσμάτων, όμως αποτελούν παρόλα αυτά μια επιπλέον ένδειξη ότι τα μοντέλα kriging που κατασκευάστηκαν είναι αξιόπιστα.

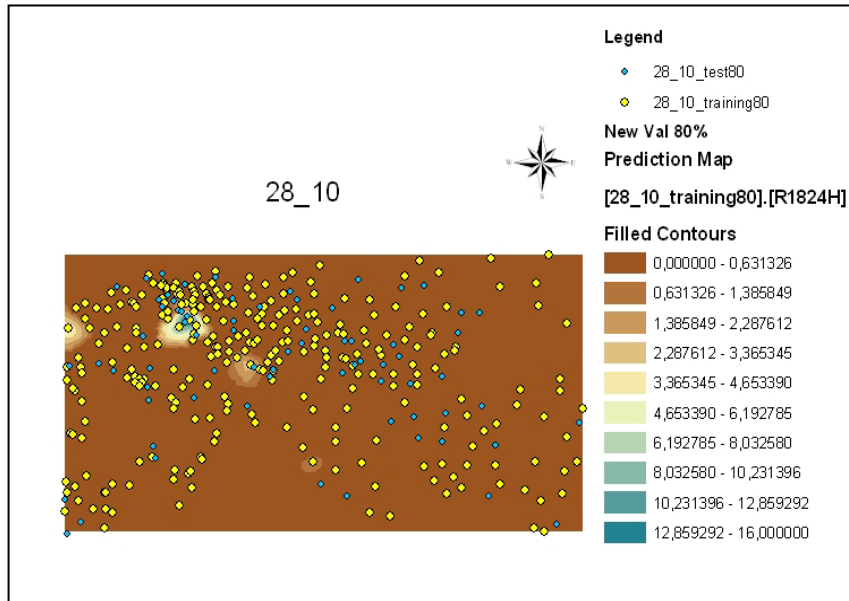
Παρατηρούμε ότι και στην περίπτωση αυτή η ομοιόμορφη, ως προς την γεωγραφική κατανομή, επιλογή των σημείων του συνόλου εκπαίδευσης (60%) δίνει πολύ χαμηλότερα σφάλματα και συνεπώς καλύτερες εκτιμήσεις.

8.3.4 Σχολιασμός της διαδικασίας επικύρωσης

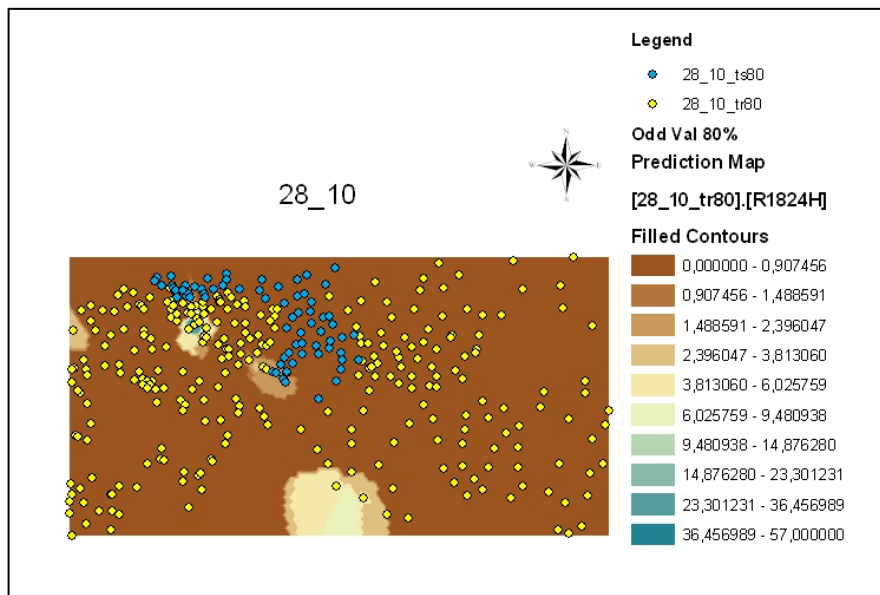
Όπως φαίνεται από τις προηγούμενες παραγράφους, η επιλογή των υποσυνόλων που θα χρησιμοποιηθούν για την διαδικασία επικύρωσης είναι πολύ σημαντική. Γενικά, θα πρέπει τα υποσύνολα που χρησιμοποιούνται για την επικύρωση της μεθόδου (π.χ. τα υποσύνολα του 80% και 60%, στην παρούσα μελέτη) να είναι κατά το δυνατόν αντιπροσωπευτικά του συνόλου των δεδομένων, έτσι ώστε και εφαρμογή της μεθόδου kriging να δώσει αξιόπιστες προβλέψεις για τα υπόλοιπα σημεία, δηλαδή για τα αντίστοιχα σύνολα ελέγχου (20% και 40%, αντίστοιχα, στην παρούσα μελέτη).

Συχνά όμως αυτό δεν είναι τόσο απλό, ειδικά όταν οι μετρήσεις παρουσιάζουν μεγάλη ανομοιομορφία. Παραδείγματος χάριν, στα δεδομένα της ημέρας χαμηλής βροχόπτωσης 28/10, γενικά οι μετρήσεις έχουν πολύ χαμηλές τιμές. Τα σημεία μέτρησης με υψηλές τιμές βροχόπτωσης είναι πολύ λίγα, αλλά παίζουν καθοριστικό ρόλο στην εφαρμογή της

μεθόδου kriging. Έτσι, μπορούμε να πάρουμε ένα 80% του συνόλου των σημείων ομοιόμορφα κατανεμημένων γεωγραφικά, που όμως δίνει πολύ κακή πρόγνωση, όπως αυτό που φαίνεται στην Εικόνα 19, καθώς και μεγάλο σφάλμα $RMSE = 5.434$.



Εικόνα 19. Σύνολο εκπαίδευσης 80% (κίτρινο χρώμα) και σύνολο ελέγχου 20% (μπλέ χρώμα) διαμορφωμένα με ομοιόμορφη επιλογή για την ημέρα 28/10, και χάρτης εκτίμησης με υψηλό σφάλμα $RMSE=5.434$.



Εικόνα 20. Σύνολο εκπαίδευσης 80% (κίτρινο χρώμα) και σύνολο ελέγχου 20% (μπλέ χρώμα) διαμορφωμένα με αυθαίρετη επιλογή για την ημέρα 28/10, και χάρτης εκτίμησης με πολύ χαμηλό σφάλμα $RMSE=0.6404$.

Αντίθετα, για την ίδια ημέρα μια αυθαίρετη επιλογή των σημείων, όπως αυτή που απεικονίζεται στην Εικόνα 20, μπορεί να δώσει πολύ καλύτερη πρόβλεψη και πολύ χαμηλό σφάλμα. Στην περίπτωση αυτή το σύνολο του 80% των σημείων περιέχει σχεδόν όλα τα σημεία στις περιοχές με έντονη μεταβλητότητα της βροχόπτωσης. Η εφαρμογή της μεθόδου kriging δίνει έτσι πολύ καλή πρόβλεψη για το υπόλοιπο 20% των σημείων και έχει πολύ μικρό σφάλμα $RMSE = 0.6404$.

Επομένως, είναι φανερό ότι αν και γενικά είναι καλό να γίνεται μια ομοιόμορφη γεωγραφικά επιλογή των σημείων που θα αποτελέσουν το υποσύνολο για την εφαρμογή της μεθόδου, χρειάζεται ιδιαίτερη προσοχή καθώς η επιλογή πρέπει να είναι ομοιόμορφη και όσον αναφορά τις τιμές των μετρήσεων. Όταν έχουμε έντονη τοπική μεταβλητότητα των τιμών της βροχόπτωσης η οποία εντοπίζεται σε λίγα σημεία χρειάζεται αντιπροσωπευτικός αριθμός των σημείων αυτών για να πάρουμε μια καλή εκτίμηση κατά την διαδικασία της επικύρωσης της μεθόδου kriging.

Σημειώνουμε επομένως ότι χρειάζεται ιδιαίτερη προσοχή στην επιλογή των σημείων του συνόλου εκπαίδευσης που αποτελεί τη βάση για την εκτίμηση της μεθόδου kriging κατά την διαδικασία επικύρωσης. Σε ορισμένα σύνολα δεδομένων η επιλογή ή όχι συγκεκριμένων σημείων εμφανίζεται να είναι ιδιαίτερα κρίσιμη στην διαδικασία επικύρωσης, άσχετα από τον τρόπο επιλογής (αυθαίρετος ή ομοιόμορφος γεωγραφικά).

8.4 Σύγκριση με άλλες μεθόδους παρεμβολής

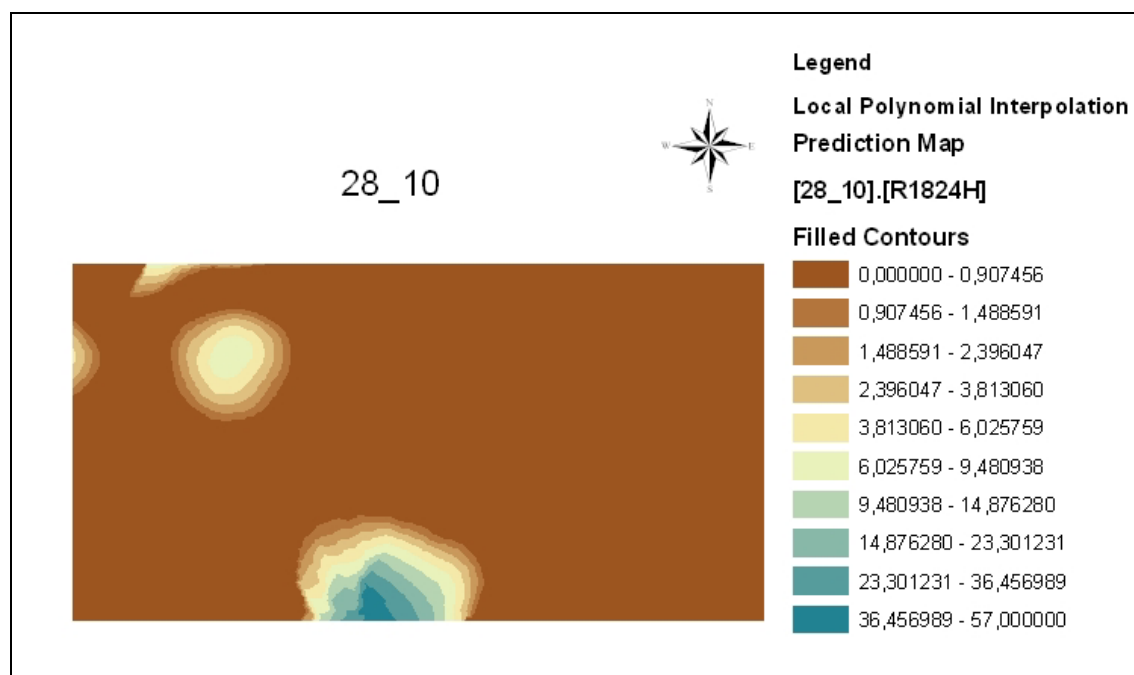
Στην συνέχεια, τα αποτελέσματα του kriging συγκρίθηκαν με τα αποτελέσματα άλλων μεθόδων παρεμβολής. Για το σκοπό αυτό εξετάστηκαν οι εξής μέθοδοι:

- καθολική πολυωνυμική παρεμβολή
- τοπική πολυωνυμική παρεμβολή
- μέθοδος Inverse Distance Weighting.

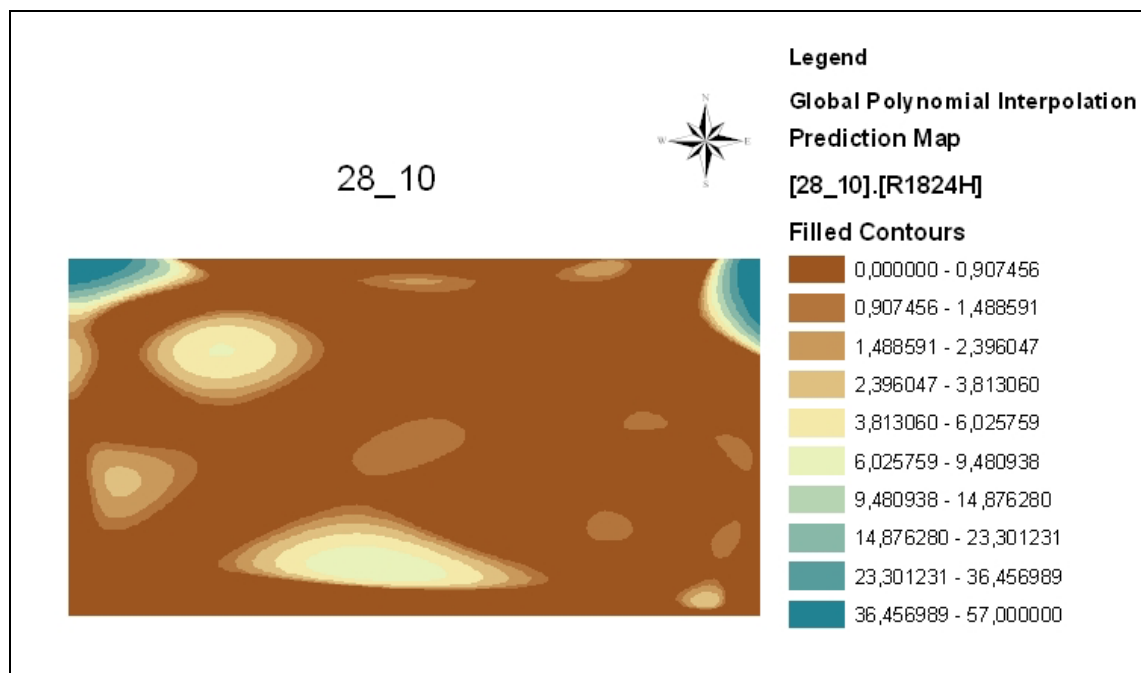
Στον Πίνακα 10 παρουσιάζονται τα συγκριτικά σφάλματα πρόγνωσης για την μέθοδο Κανονικού kriging, Καθολικής πολυωνυμικής παρεμβολής, τοπικής πολυωνυμικής παρεμβολής και για την μέθοδο inverse distance weighting, για την ημέρα 28/10. Στις Εικόνες 12, 13 και 14 παρουσιάζονται οι χάρτες πιθανότητας για τις μεθόδους Local Polynomial Interpolation, Global Polynomial Interpolation και Inverse Distance Weighting, αντίστοιχα, για την ημέρα 28/10.

Πίνακας 10. Σύγκριση με άλλες μεθόδους για την ημέρα 28/10.

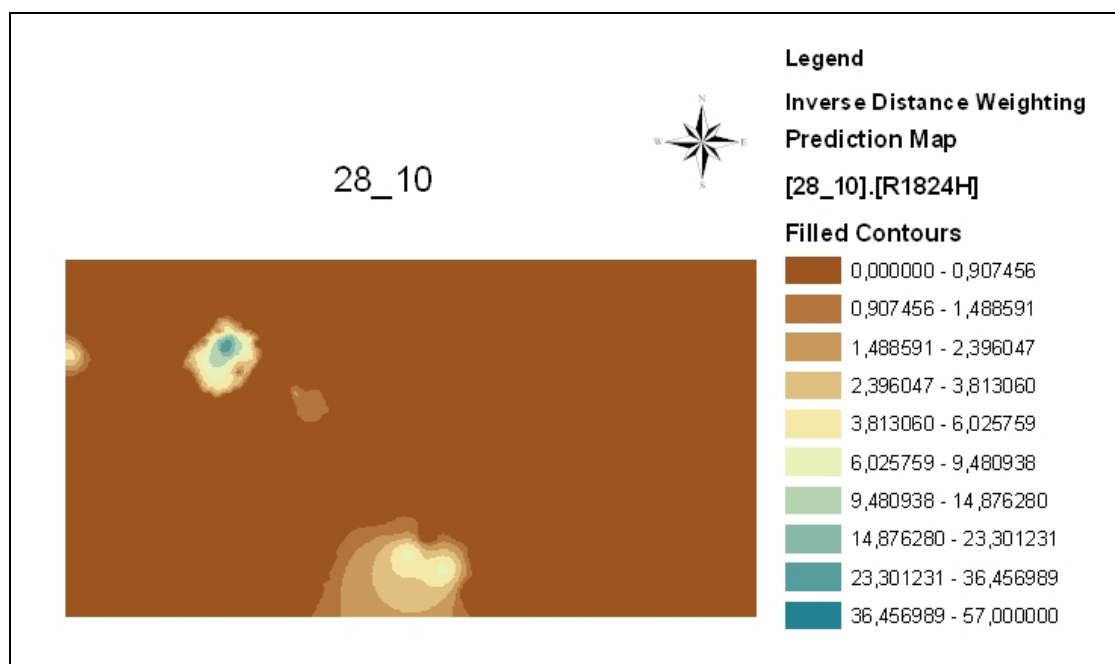
	Ordinary Kriging	Global Polynomial Interpolator	Local Polynomial Interpolator	Inverse Distance Weighting
Μέσο Σφάλμα	0.01042	-0.1018	0.07432	0.0195
Σφάλμα RMSE	2.587	4.608	3.14	2.69



Εικόνα 12. Ο χάρτης πρόγνωσης με την μέθοδο Local polynomial interpolation για την ημέρα 28/10.



Εικόνα 13. Ο χάρτης πρόγνωσης με την μέθοδο Global polynomial interpolation για την ημέρα 28/10.

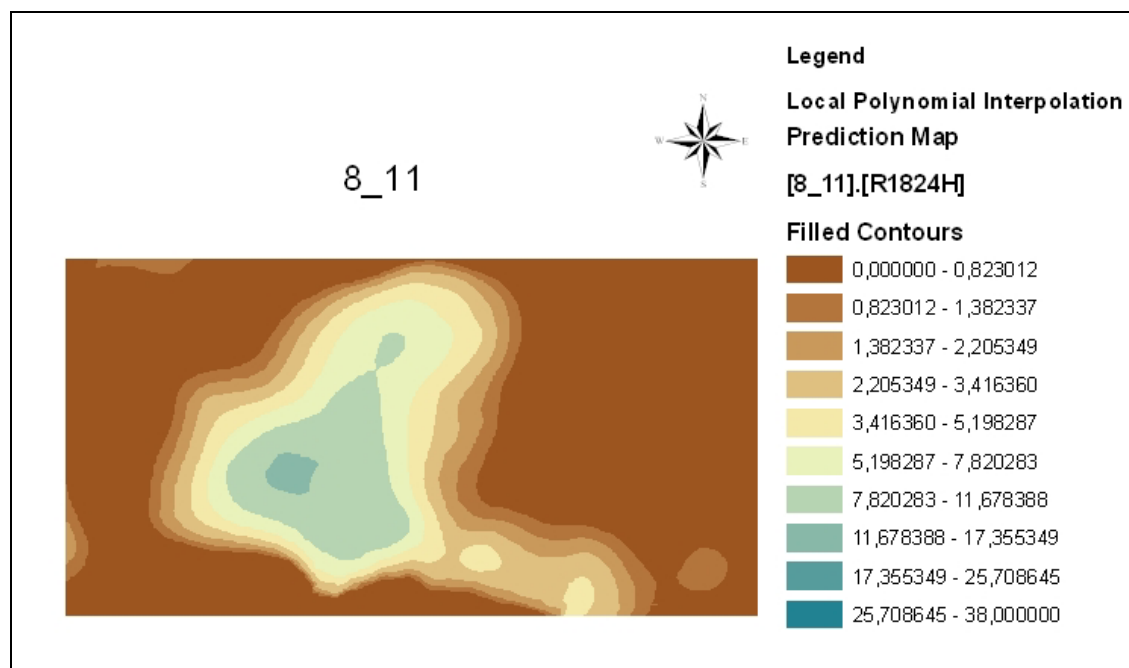


Εικόνα 14. Ο χάρτης πρόγνωσης με την μέθοδο Inverse Distance Weighting για την ημέρα 28/10.

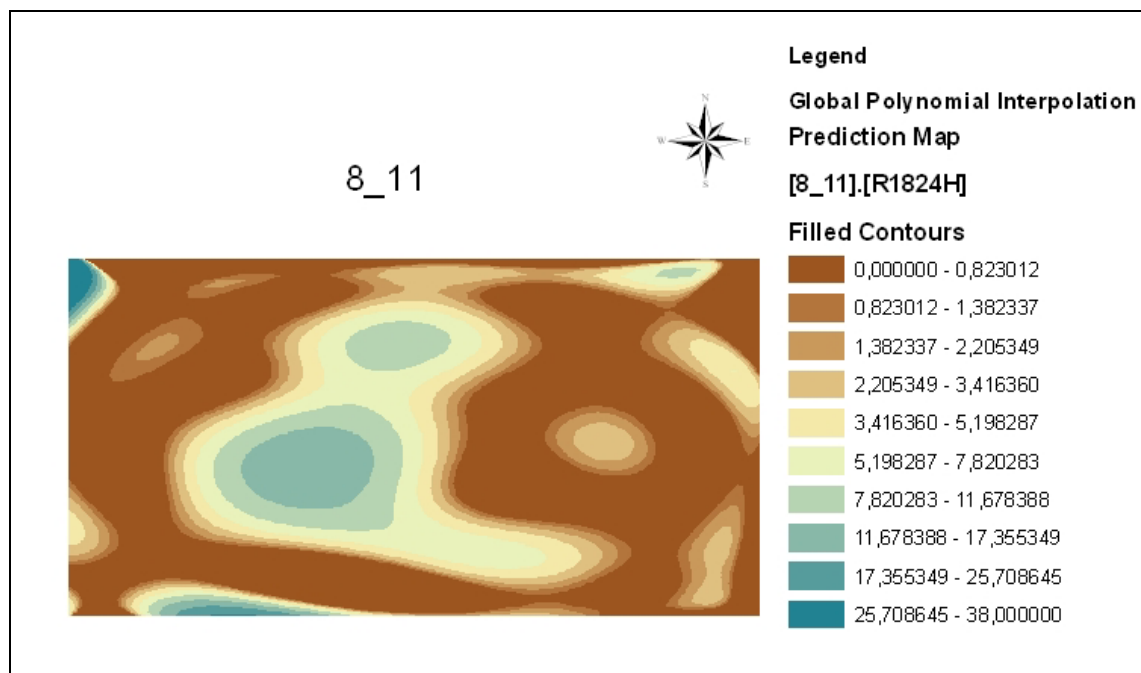
Στον Πίνακα 11 παρουσιάζονται τα συγκριτικά σφάλματα πρόγνωσης για την μέθοδο Κανονικού kriging, Καθολικής πολυωνυμικής παρεμβολής, τοπικής πολυωνυμικής παρεμβολής και για την μέθοδο inverse distance weighting, για την ημέρα 8/11. Στις Εικόνες 15, 16 και 17 παρουσιάζονται οι χάρτες πιθανότητας για τις μεθόδους Local Polynomial Interpolation, Global Polynomial Interpolation και Inverse Distance Weighting, αντίστοιχα, για την ημέρα 8/11.

Πίνακας 11. Σύγκριση με άλλες μεθόδους για την ημέρα 8/11.

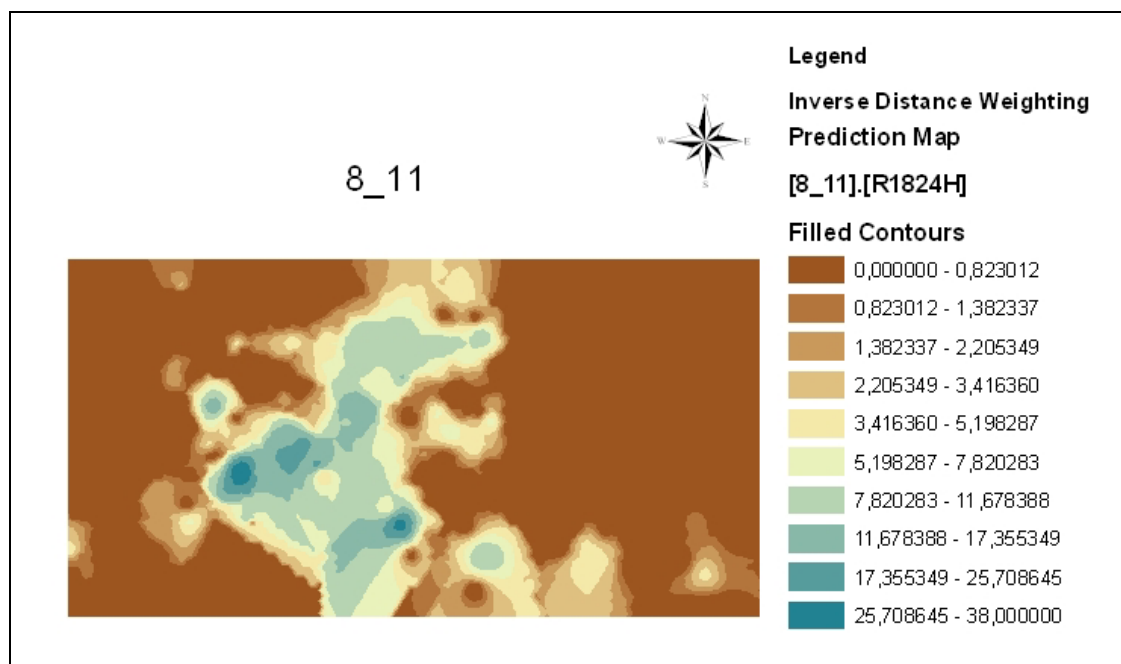
	Ordinary Kriging	Global Polynomial Interpolator	Local Polynomial Interpolator	Inverse Distance Weighting
Μέσο Σφάλμα	-0.009617	0.1531	-0.1216	0.01339
Σφάλμα RMSE	3.543	6.896	3.549	3.337



Εικόνα 15. Ο χάρτης πρόγνωσης με την μέθοδο Local polynomial interpolation για την ημέρα 8/11.



Εικόνα 16. Ο χάρτης πρόγνωσης με την μέθοδο Global polynomial interpolation για την ημέρα 8/11.



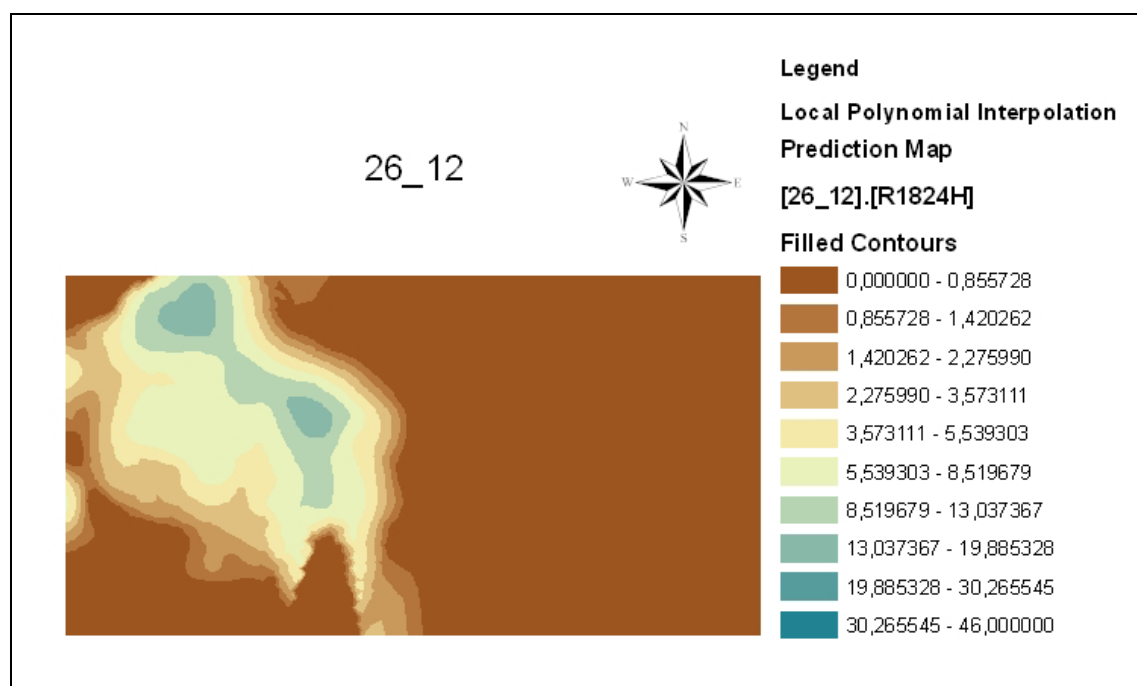
Εικόνα 17. Ο χάρτης πρόγνωσης με την μέθοδο Inverse Distance Weighting για την ημέρα 8/11.

Στον Πίνακα 12 παρουσιάζονται τα συγκριτικά σφάλματα πρόγνωσης για την μέθοδο Κανονικού kriging, Καθολικής πολυωνυμικής παρεμβολής, τοπικής πολυωνυμικής παρεμβολής και για την μέθοδο inverse distance weighting, για την ημέρα 26/12. Στις

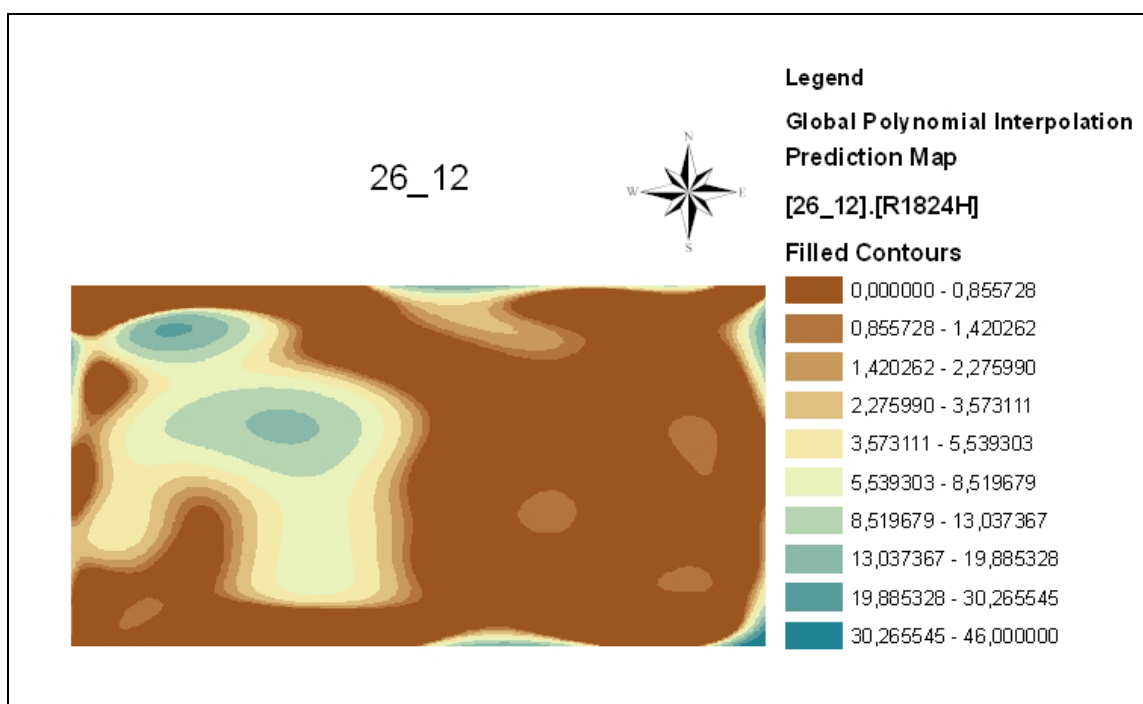
Εικόνες 18, 19 και 20 παρουσιάζονται οι χάρτες πιθανότητας για τις μεθόδους Local Polynomial Interpolation, Global Polynomial Interpolation και Inverse Distance Weighting, αντίστοιχα, για την ημέρα 26/12.

Πίνακας 12. Σύγκριση με άλλες μεθόδους για την ημέρα 26/12.

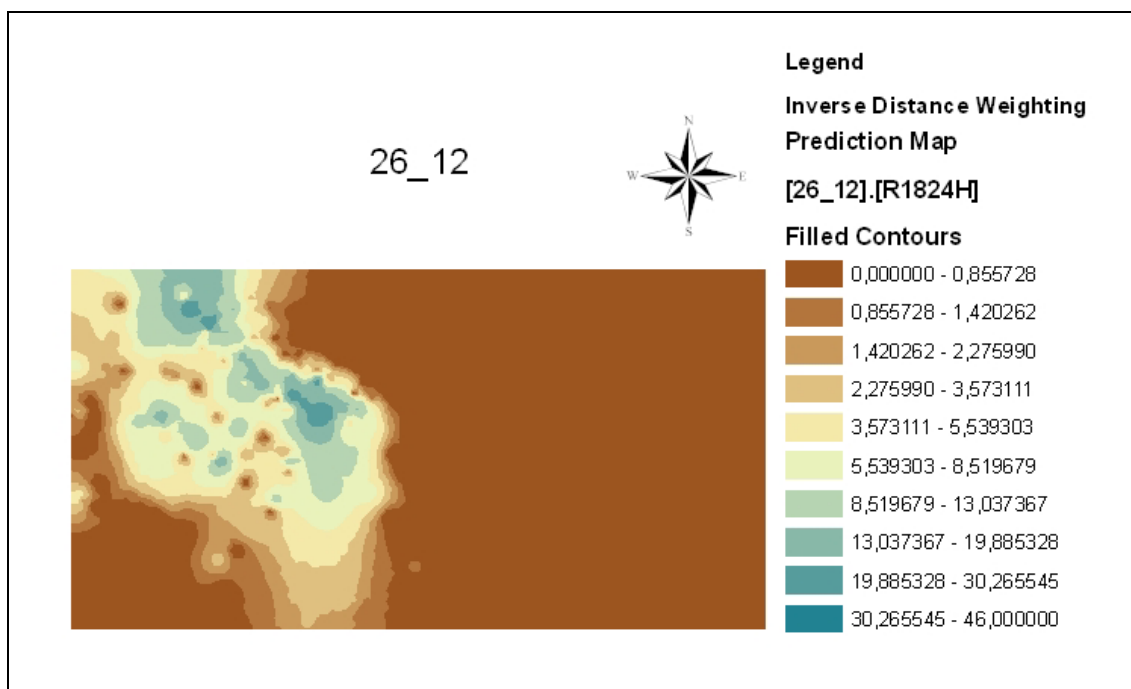
	Ordinary Kriging	Global Polynomial Interpolator	Local Polynomial Interpolator	Inverse Distance Weighting
Μέσο Σφάλμα	-0.0299	-0.1638	0.1211	0.1393
Σφάλμα RMSE	4.805	8.663	5.014	5.004



Εικόνα 18. Ο χάρτης πρόγνωσης με την μέθοδο Local polynomial interpolation για την ημέρα 26/12.



Εικόνα 19. Ο χάρτης πρόγνωσης με την μέθοδο Global polynomial interpolation για την ημέρα 26/12.



Εικόνα 20. Ο χάρτης πρόγνωσης με την μέθοδο Local polynomial interpolation για την ημέρα 26/12.

Η μέθοδος Inverse Distance Weighting είναι μια μέθοδος παρεμβολής που για την εκτίμηση της τιμής μιας μεταβλητής σε ένα σημείο χρησιμοποιεί τις τιμές της μεταβλητής στα γειτονικά σημεία. Η εξίσωση που δίνει την εκτιμώμενη τιμή είναι ίδια με την εξίσωση της μεθόδου Kriging,

$$Z_0 = \sum_i \lambda_i Z_i ,$$

μόνο που τα βάρη λ_i στην περίπτωση αυτή εξαρτώνται μόνο από την απόσταση.

Στην συγκεκριμένη εφαρμογή της μεθόδου Inverse Distance Weighting χρησιμοποιήθηκαν οι παράμετροι που φαίνονται στο πίνακα που ακολουθεί

power	2
Neighbors to include	15 (include at least 10)
Ellipsis major semiaxis	7.3966
Ellipsis minor semiaxis	7.3966

Στην μέθοδο Global polynomial interpolation χρησιμοποιήθηκε ένα πολυώνυμο με την μέγιστη δυνατή τάξη (10^η τάξη). Ένα τέτοιο σχήμα είναι κατάλληλο για την παρεμβολή μεταβολών ευρείας κλίμακας. Σοβαρό μειονέκτημα της μεθόδου είναι ότι μπορεί να δώσει αρνητικές προγνώσεις, το οποίο όμως δεν συνέβη στην συγκεκριμένη εφαρμογή.

Στην μέθοδο Local polynomial interpolation χρησιμοποιήθηκε ένα πολυώνυμο 2^{ης} τάξης για την προσαρμογή των δεδομένων της παρούσας μελέτης. Ένα τέτοιο σχήμα αναμένεται να δώσει καλύτερα αποτελέσματα από την προηγούμενη μέθοδο όσον αφορά τοπικές μεταβλητότητες. Και αυτή η μέθοδος έχει το μειονέκτημα ότι μπορεί να δώσει αρνητικές προβλέψεις, το οποίο δεν συνέβη όμως στην συγκεκριμένη εφαρμογή.

8.5 Σχολιασμός αποτελεσμάτων

Από τα αποτελέσματα που έδωσαν οι τέσσερις μέθοδοι χωρικής παρεμβολής που εξετάστηκαν μπορούμε να εξάγουμε τα παρακάτω συμπεράσματα.

Οι δύο μέθοδοι πολυωνυμικής παρεμβολής Global και Local έδωσαν υψηλότερα σφάλματα, ενώ και οι επιφάνειες πρόγνωσης που έδωσαν απέχουν σημαντικά από αυτές των δύο άλλων μεθόδων. Χειρότερη απόδοση είχε η μέθοδος Global polynomial interpolation και για τις τρεις ημέρες που εξετάστηκαν. Η μέθοδος έδωσε εκτιμήσεις που αποκλίνουν σημαντικά από αυτές των άλλων μεθόδων, με κύριο χαρακτηριστικό τις υψηλές τιμές εκτίμησης στα άκρα της γεωγραφικής περιοχής (π.χ. ΒΔ και ΒΑ άκρο για την ημέρα 28/10, ΒΔ και ΝΔ άκρο την ημέρα 8/11, και Ανατολικά άκρο για την ημέρα 26/12). Η μέθοδος Local polynomial interpolation έδωσε πολύ καλύτερα αποτελέσματα από την Global polynomial interpolation «πιάνοντας» τις τάσεις δίνουν και οι υπόλοιπες

μέθοδοι. Με εξαίρεση την ημέρα της 28/10 όπου γίνεται μια υπερεκτίμηση των τιμών της βροχόπτωσης στο Νότιο τμήμα, οι εκτιμήσεις της τοπικής πολυωνυμικής παρεμβολής είναι αρκετά κοντά σε αυτές των δύο άλλων μεθόδων, αλλά υστερούν σε ακρίβεια όπως φαίνεται και από τις τιμές του σφάλματος RMSE.

Οι μέθοδος Kriging και η μέθοδος Inverse Distance Weighting έδωσαν τα μικρότερα σφάλματα και τις πιο λεπτομερείς επιφάνειες πρόγνωσης. Όσον αφορά τα σφάλματα RMSE, η μέθοδος Kriging έδωσε χαμηλότερο σφάλμα για τις ημέρες χαμηλής και υψηλής βροχόπτωσης (28/10 και 26/12, αντίστοιχα), ενώ για την ημέρα μέτριας βροχόπτωσης, 8/11, η μέθοδος Inverse Distance Weighting έδωσε χαμηλότερο σφάλμα από αυτό του Kriging. Όσον αφορά τις εκτιμήσεις των δύο μεθόδων, παρατηρούμε ότι είναι αρκετά παρόμοιες και για τις τρεις ημέρες που εξετάστηκαν. Όμως, οι εκτιμήσεις της μεθόδου Inverse Distance Weighting παρουσιάζουν σε αρκετά σημεία ορισμένα πολύ τοπικά φαινόμενα που είναι γνωστά στην βιβλιογραφία σαν «bull's eyes» (βλ. [25] και [26]). Το φαινόμενο αυτό (bull's eye effect) είναι χαρακτηριστικά της μεθόδου και οφείλεται στο ότι η μέθοδος τείνει να θεωρεί μεμονωμένες ανώμαλες τιμές ως τοπικά ακρότατα, και αναφέρεται ως ένα σημαντικό μειονέκτημά της. Αντίθετα οι εκτιμήσεις της μεθόδου Kriging είναι αρκετά ομαλές επιφάνειες, και έχουν τα μικρότερα σφάλματα για τις δύο από τις τρεις ημέρες που εξετάστηκαν.

Συνεπώς, η μέθοδος kriging εμφανίζεται να έχει την καλύτερη απόδοση σε σχέση με τις άλλες μεθόδους που εφαρμόστηκαν, δίνοντας και τα μικρότερα RMSE σφάλματα και ομαλές εκτιμήσεις.

Κεφάλαιο 9

ΣΥΜΠΕΡΑΣΜΑΤΑ

Από την γεωστατιστική ανάλυση και μελέτη της βροχόπτωσης που πραγματοποιήθηκε στην παρούσα εργασία προέκυψαν κάποια γενικά συμπεράσματα τα οποία παρατίθενται στην συνέχεια.

Θα πρέπει να τονισθεί ότι πολλές από τις περιοχές της Μεσογείου, στις οποίες πραγματοποιήθηκε ανάλυση της χωρικής κατανομής της βροχόπτωσης δεν καλύπτονται επαρκώς από τα διαθέσιμα δεδομένα και συνεπώς υπάρχει ανάγκη πύκνωσης του δικτύου βροχομετρικών και μετεωρολογικών σταθμών από το οποίο θα συλλέγονται τα δεδομένα βροχόπτωσης. Στην παρούσα μελέτη, τα διαθέσιμα δεδομένα ήταν τέτοια που επιτρέπουν μεν την εφαρμογή της μεθόδου χωρικής παρεμβολής (ικανό πλήθος μετρήσεων), αλλά η γεωγραφική κατανομή των δεδομένων ήταν τέτοια που σε ορισμένες περιοχές υπήρχαν πολύ λίγες μετρήσεις. Αυτό έχει σαν αποτέλεσμα, η εφαρμογή της μεθόδου kriging να μην γίνεται στο βέλτιστο βαθμό, αλλά να πραγματοποιείται έτσι ώστε να αποκτηθεί μια κατά το δυνατόν κοντά στην πραγματικότητα εκτίμηση της τάξης μεγέθους της βροχόπτωσης στις περιοχές εκείνες που δεν καλύπτονται επαρκώς από σταθμούς.

Όσον αφορά την μελέτη και την εφαρμογή της μεθόδου Kriging για την εκτίμηση της χωρικής κατανομής της βροχόπτωσης προέκυψαν τα παρακάτω συμπεράσματα:

- Το Kriging είναι μια αρκετά γρήγορη τεχνική παρεμβολής που είναι ακριβής αν τα δεδομένα δεν έχουν σφάλματα μέτρησης και εξομαλύνει τυχόν υπάρχουν σφάλματα μέτρησης.
- Η μέθοδος kriging επιτρέπει την διερεύνηση της χωρικής αυτοσυσχέτισης των δεδομένων και λαμβάνει υπόψη την ανισοτροπία των δεδομένων. Αυτό την καθιστά κατάλληλη για την παρεμβολή υδρογεολογικών δεδομένων τα οποία μπορεί να εμφανίζουν έντονα φαινόμενα ανισοτροπίας.
- Η μέθοδος kriging αποδείχθηκε ότι είναι επίσης αρκετά ευέλικτη. Όμως, η ευελιξία της μεθόδου απαιτεί την λήψη αποφάσεων για τον καθορισμό μιας πληθώρας παραμέτρων. Αυτό απαιτεί πολύ σημαντικότερη γνώση από την πλευρά του χρήστη, σε σχέση με άλλες μεθόδους παρεμβολής. Πολύ χρήσιμο εργαλείο για την επιλογή του βέλτιστου μοντέλου ημιβαριογράμματος και των κατάλληλων παραμέτρων της μεθόδου kriging αποδείχθηκε ότι είναι η τεχνική της αυτό-επικύρωσης των αποτελεσμάτων.
- Η μέθοδος kriging έδωσε πολύ καλά αποτελέσματα. Σε σχέση με την μέθοδο Inverse Distance Weighting και τις πολωνυμικές παρεμβολές, το kriging αποδείχθηκε πιο αξιόπιστο όσον αφορά τις προβλέψεις που έδωσε. Οι εκτιμήσεις

της μεθόδου kriging είναι καλύτερες ή πολύ κοντά σε εκείνες που έδωσε η μέθοδος Inverse Distance Weighting, αλλά υπερτερούν δίνοντας πολύ πιο «ομαλές» επιφάνειες σε αντίθεση με το φαινόμενο «bull's eye» που εμφανίζουν οι εκτιμήσεις της Inverse Distance Weighting.

- Τέλος, η μέθοδος kriging επιτρέπει την εξαγωγή μιας πληθώρας αποτελεσμάτων, μεταξύ των οποίων είναι τα σφάλματα πρόγνωσης, οι χάρτες πρόγνωσης και πιθανότητας κ.ά.

Η προσαρμογή της κατάλληλης συνάρτησης βαριογράμματος στο πειραματικό ημιβαριόγραμμα είναι μια διαδικασία όπου η εμπειρία του χρήστη

Με βάση την παρούσα μελέτη, προτείνεται η εφαρμογή της μεθόδου kriging για την εκτίμηση της χωρικής κατανομής υδρογεωλογικών παραμέτρων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] de Marsily, G., *Quantitative Hydrogeology, Groundwater Hydrology for Engineers*, Academic Press, 1986.
- [2] Kitanidis, P. K., *Introduction to Geostatistics: Applications to Hydrogeology*, Cambridge University Press.
- [3] Cressie, N. A. C., *Statistics for Spatial Data*, John Wiley & Sons, 1993.
- [4] Stein, M., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, Berlin, 1999.
- [5] Isaaks, E. H., and R. M. Srivastava, *An introduction to applied geostatistics*. Oxford University Press, New York 1989.
- [6] Σαράτσης, Γεώργιος, *Συζευγμένη Γεωστατιστική – Μηχανική Ανάλυση Σηράγγων, Διατριβή Μεταπτυχιακού Διπλώματος Ειδίκευσης, Τμήμα Μηχανικών Ορυκτών Πόρων, Πολυτεχνείο Κρήτης, Δεκέμβριος 2004*.
- [7] Delhomme, J. P., “Kriging in Hydrosiences”, *Advances in Water Resources*, vol. 1, 1978.
- [8] Atkinson, P. M., and C. D. Lloyd, “Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging”, *Journal of Geographical Information & Decision Analysis*, vol. 2, 1998.
- [9] Ishida, T., and S. Kawashima, “Use of Cokriging to Estimate Surface Air Temperature from Elevation”, *Theoretical & Applied Climatology*, vol. 47, 1993.
- [10] Desbarats, A. J., and C. E. Logan, M. J. Hinton, and D.R. Sharpe, “On the kriging of water table elevations using collateral information from a digital elevation model”, *Journal of Hydrology*, vol. 255, 2002.
- [11] Marinoni, O., “Improving geological models using a combined ordinary-indicator kriging approach”, *Engineering Geology*, vol. 69, 2003.
- [12] Kumar, V., and Remadevi, “Kriging of Groundwater Levels – A case study”, *Journal of Spatial Hydrology*, vol. 6, 2006.
- [13] Bracq, P., and F. Delay, “Transmissivity and morphological features in a chalk aquifer: a geostatistical approach of their relationship”, *Journal of Hydrology*, vol. 191, 1997.
- [14] Hosseini, S. M., and M. Kholghi, “Estimation of Aquifer Transmissivity using Kriging, Artificial Neural Network, and Neuro-Fuzzy models”, *Geophysical Research Abstracts*, vol. 9, 2007.

- [15] Patriarche, D., M. C. Castro, and P. Goovaerts, "Estimating Regional Hydraulic Conductivity Fields – A Comparative Study of Geostatistical Methods", *Mathematical Geology*, vol. 37, 2005.
- [16] Salih, I., H.B.L. Petterson, A. Sivertum, and E. Lund, "Spatial correlation between radon in groundwater and bedrock uranium: GIS and geostatistical analyses", *Journal of Spatial Hydrology*, vol. 2.
- [17] Hu, K., Y. Huang, H. Li, B. Li, D. Chen, and R. E. White, "Spatial variability of shallow groundwater level, electrical conductivity and nitrate concentration, and risk assessment of nitrate contamination in North China Plain", *Environment International*, vol. 31, 2005.
- [18] Virdee, T. S., and N. T. Kottegoda, "A brief review of kriging and its application to optimal interpolation and observation well selection", *Hydrological Sciences*, vol. 29, 1984.
- [19] Theodossiou, N., and P. Latinopoulos, "Evaluation and optimisation of groundwater observation networks using the Kriging methodology", *Environmental Modelling & Software*, 2005
- [20] *Using ArcGIS Geostatistical Analyst*, ArcGIS 9, ESRI, 2003.
- [21] <http://www.esri.com/software/arcgis/extensions/geostatistical/index.html>
- [22] Van Beers, W. C. M. and J. P.C. Kleijnen, "Kriging Interpolation in Simulation: A Survey", *Proceedings of the 2004 Winter Simulation Conference*, R. G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, eds., pp. 113-121.
- [23] Atkinson, P.M., and C. D. Lloyd, Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging, *Journal of Geographical Information and Decision Analysis*, vol. 2, no 2, pp. 65-76, 1988.
- [24] Spatial Interpolation Comparison 97, Special Issue, *Journal of Geographic Information and Decision Analysis*, vol. 2, no. 1-2, 1998. http://www.geodec.org/gida_4.htm
- [25] Gonçalves, G., "Analysis of interpolation errors in urban digital surface models created from Lidar data", *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Edited by M. Caetano and M. Painho.
- [26] Arthur, J. D., and W. H. Pollock, "Use of ArcView GIS for Geologic Surface Modeling - Preliminary Results from Subsurface Mapping in Southwest Florida", *Digital Mapping Techniques '98 - Workshop Proceedings*, U.S. Geological Survey Open-File Report 98-487, 1998.