



Πανεπιστήμιο Αιγαίου

Χωρική Ανάλυση

Ενότητα 10: Στοιχεία χωρικής παρεμβολής

Κυριακίδης Φαίδων

Τμήμα Γεωγραφίας

Άδειες Χρήσης

Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.

Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.

Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.

Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Στοιχεία Χωρικής Παρεμβολής

Φαίδων Κυριακίδης

Καθηγητής

phkyriakidis@geo.aegean.gr



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
Λόφος Πανεπιστημίου, 81100 Μυτιλήνη

Χωρική Ανάλυση

ΤΜΗΜΑ ΓΕΩΓΡΑΦΙΑΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΓΕΩΓΡΑΦΙΑ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΗ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ

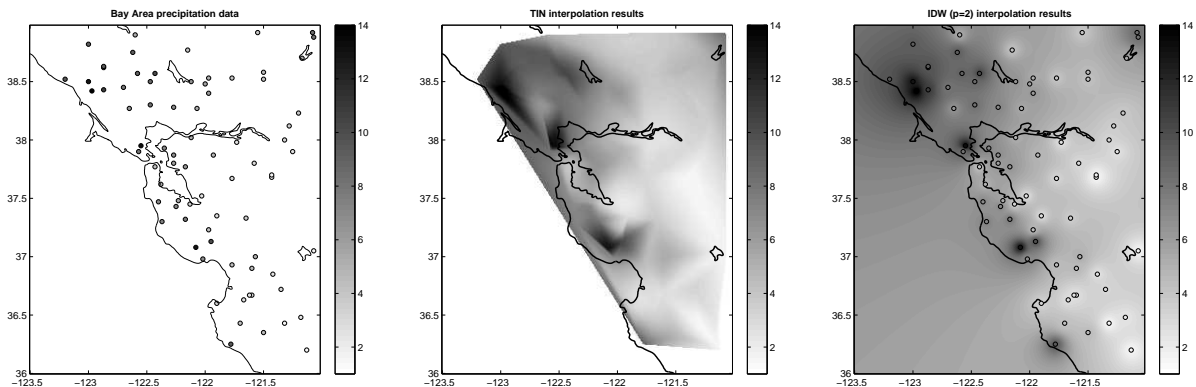
Εισαγωγικές Έννοιες

Εισαγωγή



Δεδομένα

Μετρήσεις μιας χωρικής μεταβλητής (spatial attribute) που έχουν καταμετρηθεί σε διάφορες μονάδες παρατήρησης, π.χ., σημεία δειγματοληψίας. Συνήθως, τα σημεία παρατήρησης είναι λίγα και χωρικά διεσπαρμένα (σε μή κανονική διάταξη)



Σκοπός του μαθήματος

Επισκόπηση βασικών μεθόδων **χωρικής παρεμβολής**, δηλ. της πρόβλεψης (εκτίμησης), με βάση τις γνωστές μετρήσεις, άγνωστων τιμών της εν λόγω μεταβλητής σε διάφορα σημεία (ή πολύγωνα). Συχνά, χωρική παρεμβολή μετρήσεων γίνεται στους κόμβους ενός κανονικού καννάβου για την κατασκευή εκτιμώμενης "επιφάνειας" τιμών μιας μεταβλητής



Συμβολισμός και Ορολογία (1)

Δεδομένα

Σύνολο N μετρήσεων μιας μεταβλητής $\{z(s_n), n = 1, \dots, N\}$ σε N μονάδες παρατήρησης $\{s_n, n = 1, \dots, N\}$, π.χ., s_n είναι το σημείο παρατήρησης n

Στόχος χωρικής παρεμβολής

Πρόβλεψη (εκτίμηση), με βάση τις N γνωστές μετρήσεις, M άγνωστων τιμών $\{z(t_m), m = 1, \dots, M\}$ σε ένα σύνολο M σημείων $\{t_m, m = 1, \dots, M\}$, για παράδειγμα, στους κόμβους ενός καννάβου

Ορολογία

- ▶ συχνά, οι μετρήσεις καλούνται δεδομένα πηγής (source data), ενώ οι άγνωστες τιμές καλούνται τιμές στόχου (target values), και ανάλογα οι αντίστοιχες μονάδες παρατήρησης ή πρόβλεψης – source and target zones or supports
- ▶ θεωρείται, προς το παρόν, ότι τόσο τα δεδομένα πηγής όσο και οι τιμές στόχου αφορούν σε μία και μόνο μεταβλητή Z , και οι αντίστοιχες μονάδες παρατήρησης είναι σημειακές

Συμβολισμός και Ορολογία (2)



Δεδομένα

Σύνολο N μετρήσεων μιας μεταβλητής $\{z(s_n), n = 1, \dots, N\}$ σε N μονάδες παρατήρησης $\{s_n, n = 1, \dots, N\}$, π.χ., s_n είναι το σημείο παρατήρησης n

Στόχος χωρικής παρεμβολής

Υπολογισμός, με βάση τις N γνωστές μετρήσεις, M παρεμβαλλόμενων (προβλεπόμενων) τιμών $\{\hat{z}(t_m), m = 1, \dots, M\}$ σε ένα σύνολο M σημείων $\{t_m, m = 1, \dots, M\}$.

Ισοδύναμα, παρεμβολή σε ένα σημείο t_m και κατόπιν επανάληψη παρεμβολής σε άλλα $M - 1$ σημεία, χωρίς να λαμβάνεται υπ' όψη η παρεμβαλλόμενη τιμή $\hat{z}(t_m)$ στο σημείο t_m

Ορολογία

- ▶ όταν όλες οι N μετρήσεις πηγής χρησιμοποιούνται για πρόβλεψη σε κάθε σημείο t_m , η μέθοδος χωρικής παρεμβολής ονομάζεται καθολική (global). Όταν χρησιμοποιείται μόνο ένα υποσύνολο $N(t_m) \ll N$ μετρήσεων, π.χ., των πλησιέστερων 5 για κάθε σημείο t_m , η μέθοδος ονομάζεται τοπική (local)
- ▶ ακριβής (exact) μέθοδος: όταν η θέση πρόβλεψης t_m ταυτίζεται με τη θέση παρατήρησης s_n , η προβλεπόμενη τιμή αναπαράγει την αντίστοιχη μέτρηση: $\hat{z}(s_n) = z(s_n)$
- ▶ κυρτή (convex) μέθοδος: οι εκτιμώμενες τιμές δεν είναι μικρότερες (μεγαλύτερες) από τη μικρότερη (μεγαλύτερη) μέτρηση στο δείγμα



Γραμμική Χωρική Παρεμβολή

Γραμμική συνδιασμός μετρήσεων

Η προβλεπόμενη τιμή $\hat{z}(t_m)$ στο σημείο t_m εκφράζεται ως ένα σταθμισμένο άθροισμα των N μετρήσεων $\{z(s_n), n = 1, \dots, N\}$:

$$\hat{z}(t_m) = \sum_{n=1}^N w_{mn} z(s_n)$$

w_{mn} = βάρος που αποδίδεται στη μέτρηση $z(s_n)$ κατά την πρόβλεψη της άγνωστης τιμής $z(t_m)$ στο σημείο t_m . Όταν τα σημεία στόχου αντιστοιχούν στους κόμβους ενός κανονικού καννάβου, μιλάμε για μια "επιφάνεια" προβλεπόμενων τιμών

Προσδιορισμός βαρών παρεμβολής

Τα N βάρη $\{w_{mn}, n = 1, \dots, N\}$ είναι συνήθως συνάρτηση της χωρικής διάταξης των σημείων πηγής και στόχου, και ειδικότερα, συνάρτηση των N αποστάσεων στόχου-πηγών $\{d_{mn}, n = 1, \dots, N\}$ μεταξύ του σημείου πρόβλεψης t_m και των N σημείων παρατήρησης. Με άλλα λόγια, πρώτα τα βάρη υπολογίζονται με βάση τη χωρική διάταξη των σημείων πρόβλεψης και παρατήρησης, και κατόπιν αντιστοιχίζονται με τις παρατηρούμενες μετρήσεις. . .

Μέθοδος Πλησιέστερου Γείτονα



Βάρη παρεμβολής

Συνάρτηση των αποστάσεων μεταξύ των σημείων πρόβλεψης και παρατήρησης:
1 για το σημείο παρατήρησης που είναι πλησιέστερος γείτονας του σημείου t_m ,
0 για όλα τα υπόλοιπα σημεία παρατήρησης:

$$w_{mn} = \begin{cases} 1 & \text{αν } d_{mn} = \min\{d_{mn}, n = 1, \dots, N\} \\ 0 & \text{διαφορετικά} \end{cases}$$

Τα βάρη παίρνουν τιμές **0** ή **1** και έχουν άθροισμα **1**

Παρεμβαλλόμενη τιμή

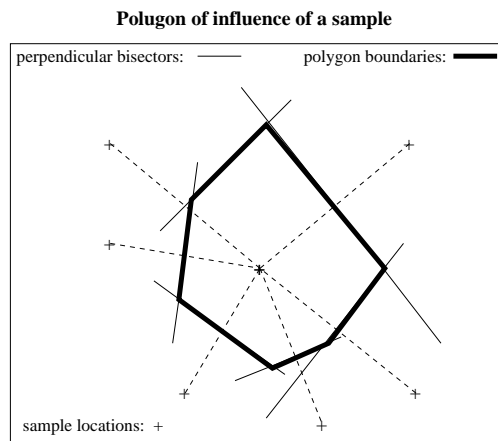
Η προβλεπόμενη τιμή $\hat{z}(t_m)$ ταυτίζεται με τη μέτρηση $z(s_n)$ του σημείου παρατήρησης s_n , το οποίο βρίσκεται πλησιέστερα στο σημείο πρόβλεψης t_m :

$$\hat{z}(t_m) = \sum_{n=1}^N w_{mn} z(s_n) = 1 \cdot z(s_n) + \sum_{\substack{n'=1 \\ n' \neq n}}^N 0 \cdot z(s_{n'})$$



Πολύγωνο Επιρροής Σημείου

- ▶ η περιοχή γύρω από ένα σημείο παρατήρησης s_n , μέσα στην οποία ένα οποιοδήποτε σημείο (πρόβλεψης) t_m βρίσκεται πλησιέστερα στο σημείο s_n απ' ότι σε οποιοδήποτε άλλο σημείο παρατήρησης $s_{n'}$. Γνωστό και ως πολύγωνο Voronoi ή Thiessen. Το σύνολο των N πολυγώνων επιρροής ονομάζεται ψηφίδωση Dirichlet
- ▶ τα όρια του πολυγώνου υπολογίζονται από την τομή ευθειών, οι οποίες ορίζονται κάθετα στα μεσοδιαστήματα των ευθύγραμμων τμημάτων που ενώνουν το σημείο παρατήρησης s_n με τα πλησιέστερα σημεία παρατήρησης
- ▶ τα πολύγωνα επιρροής δεν ορίζονται για σημεία παρατήρησης κοντά στα όρια της περιοχής μελέτης (λόγω απουσίας γειτόνων έξω από την περιοχή)



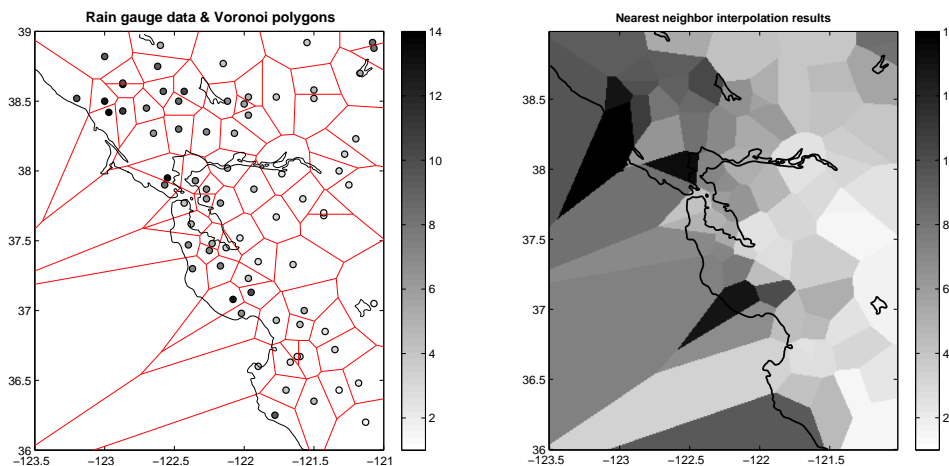
Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής 7 / 34

Παρεμβολή Πλησιέστερου Γείτονα

Παράδειγμα Παρεμβολής Πλησιέστερου Γείτονα



Χαρακτηριστικά

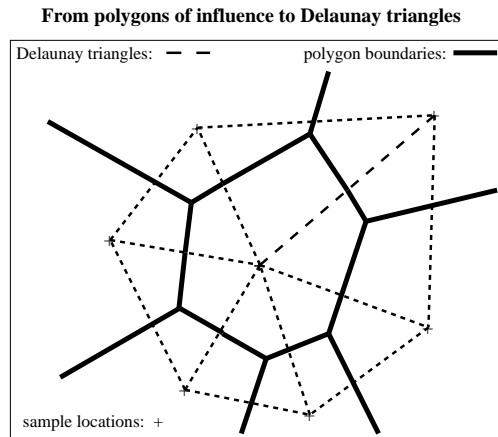
- (i) οι εκτιμώμενες τιμές απαρτίζουν μια ασυνεχή "επιφάνεια" με "μπαλλώματα" (ratches), (ii) κυρτή (convex) μέθοδος, (iii) ακριβής (exact) μέθοδος,
- (i) η τοπική και η καθολική εφαρμογή της μεθόδου ταυτίζονται



Τριγωνισμός Delaunay

Ορισμός και χαρακτηριστικά

- ▶ τρία σημεία παρατήρησης αποτελούν ένα τρίγωνο Delaunay, εάν τα αντίστοιχα πολύγωνα επιρροής τους έχουν μία κοινή κορυφή
- ▶ ένα τρίγωνο Delaunay προσεγγίζει κατά το δυνατόν το ισόπλευρο τρίγωνο
- ▶ ένα σύστημα τριγώνων Delaunay ορίζει ένα Triangulated Irregular Network (TIN), μια δομή δεδομένων που χρησιμοποιείται ευρώς στην ψηφιακή αναπαράσταση και μοντελοποίηση αναγλύφου



Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

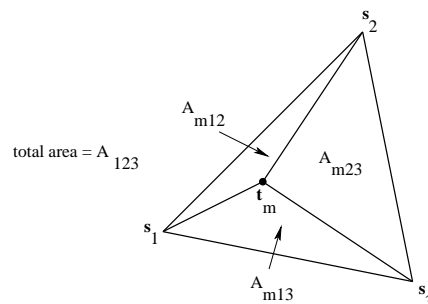
9 / 34

Χωρική Παρεμβολή Μέσω Τριγωνισμού Delaunay

Παρεμβολή Delaunay



Triangles defined between 3 source points and 1 target point



A_{123} είναι το εμβαδόν του τριγώνου με κορυφές τα σημεία παρατήρησης s_1 , s_2 , s_3

Υπολογισμός βαρών παρεμβολής

(i) εύρεση των τριών σημείων παρατήρησης, π.χ., s_1 , s_2 , και s_3 , που αποτελούν το τρίγωνο Delaunay μέσα στο οποίο εμπεριέχεται το σημείο πρόβλεψης t_m , και (ii) απόδοση βάρους στο σημείο s_1 ανάλογα με το εμβαδόν A_{m23} του απέναντι τριγώνου:

$$w_{m1} = \frac{A_{m23}}{A_{123}}, \quad w_{m2} = \frac{A_{m13}}{A_{123}}, \quad w_{m3} = \frac{A_{m12}}{A_{123}}$$

Τα βάρη έχουν άθροισμα 1 και κυμαίνονται μέσα στο διάστημα $[0, 1]$

Όσο μικραίνει η απόσταση μεταξύ του σημείου πρόβλεψης t_m και ενός σημείου παρατήρησης s_1 , τόσο αυξάνει το εμβαδόν A_{m23} του απέναντι τριγώνου, και τόσο αυξάνει το αντίστοιχο βάρος w_{m1}

Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

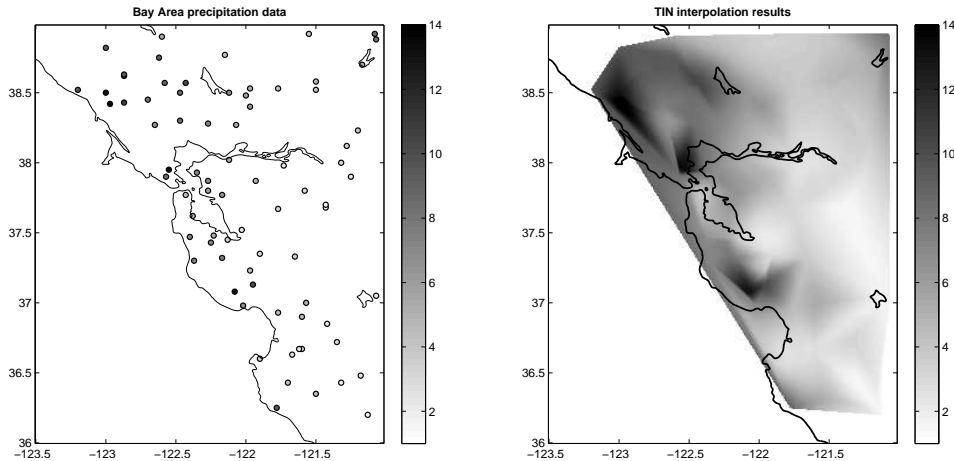
10 / 34



Παράδειγμα Παρεμβολής Delaunay

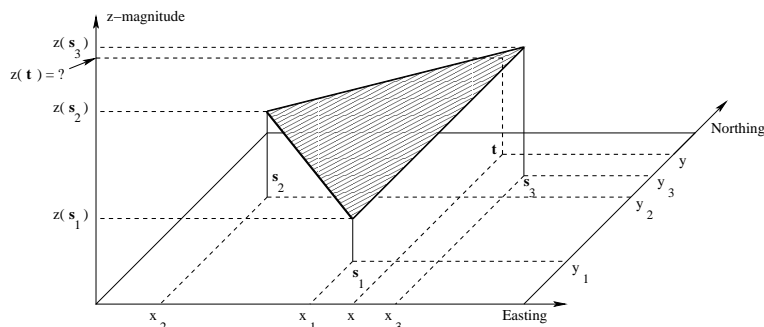
Η προβλεπόμενη τιμή $\hat{z}(t_m)$ είναι το σταθμισμένο άθροισμα των παρατηρούμενων μετρήσεων, π.χ., $z(s_1)$, $z(s_2)$ και $z(s_3)$, οι οποίες βρίσκονται στις κορυφές του τριγώνου Delaunay που εμπεριέχει το σημείο πρόβλεψης t_m :

$$\hat{z}(t_m) = \sum_{n=1}^N w_{mn} z(s_n) = \frac{A_{m23}}{A_{123}} \cdot z(s_1) + \frac{A_{m13}}{A_{123}} \cdot z(s_2) + \frac{A_{m12}}{A_{123}} \cdot z(s_3)$$



Η "επιφάνεια" των εκτιμώμενων τιμών εμφανίζει ασυνέχεις στις ακμές των τριγώνων Delaunay και απαρτίζεται από τριγωνικές "οφεις". Ακριβής και κυρτή, τοπική μέθοδος χωρικής παρεμβολής

Εναλλακτικός Ορισμός Παρεμβολής Delaunay



- ▶ με βάση μόνο τις συντεταγμένες $[x_1 \ y_1]$, $[x_2 \ y_2]$, και $[x_3 \ y_3]$, των κορυφών s_1 , s_2 και s_3 του τριγώνου Delaunay που εμπεριέχει το σημείο πρόβλεψης t_m με συντεταγμένες $[x_m \ y_m]$
- ▶ προσαρμογή επιπέδου με εξίσωση: $z = \beta_0 1 + \beta_1 x + \beta_2 y$ στις μετρήσεις $z_1 = z(s_1)$, $z_2 = z(s_2)$, και $z_3 = z(s_3)$:

$$\beta_0 1 + \beta_1 x_1 + \beta_2 y_1 = z_1$$

$$\beta_0 1 + \beta_1 x_2 + \beta_2 y_2 = z_2$$

$$\beta_0 1 + \beta_1 x_3 + \beta_2 y_3 = z_3$$

- ▶ επίλυση συστήματος 3 εξισώσεων με 3 αγνώστους για την εκτίμηση των παραμέτρων $\beta_0, \beta_1, \beta_2$ της εξίσωσης του επιπέδου που "περνά" από τις μετρήσεις $z(s_1), z(s_2), z(s_3)$
- ▶ με βάση τις συντεταγμένες του σημείου πρόβλεψης $t_m = [x_m \ y_m]$ και την εξίσωση του επιπέδου, υπολογισμός της προβλεπόμενης τιμής: $\hat{z}(t_m) = \hat{\beta}_0 1 + \hat{\beta}_1 x_m + \hat{\beta}_2 y_m$



Τοπικός Μέσος Όρος Μετρήσεων

Τα βάρη παρεμβολής είναι συνάρτηση του πλήθους $N(t_m)$ των σημείων παρατήρησης που εμπεριέχονται σε ένα κύκλο ακτίνας r , ή $r(t_m)$ στην προσαρμοζόμενη (adaptive) περίπτωση, με κέντρο το σημείο πρόβλεψης t_m :

$$w_{mn} = \begin{cases} 1/N(t_m) & \text{αν } d_{mn} \leq r \\ 0 & \text{διαφορετικά} \end{cases}$$

Το άθροισμα των βαρών είναι 1 , και οι τιμές τους κυμαίνονται στο διάστημα $[0, 1]$

Διαδικασία:

1. ορισμός ενός κύκλου ακτίνας r ή $r(t_m)$, με κέντρο το σημείο πρόβλεψης t_m
2. εύρεση του πλήθους $N(t_m)$ των σημείων παρατήρησης που εμπίπτουν στον κύκλο
3. καθένα από τα $N(t_m)$ σημεία παρατήρησης λαμβάνει βάρος $1/N(t_m)$, ενώ σε καθένα από τα υπόλοιπα σημεία παρατήρησης (εκτός του κύκλου) αντιστοιχεί βάρος 0

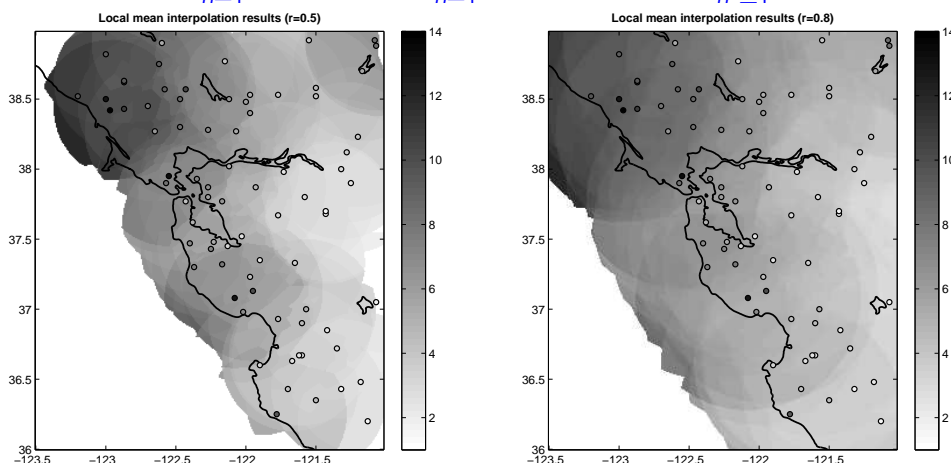
Εναλλακτικά, θα μπορούσε κανείς να ορίσει ένα μέγιστο αριθμό $\tilde{N}(t_m) < N(t_m)$ πλησιέστερων σημείων παρατήρησης μέσα σε κάθε γειτονία του σημείου πρόβλεψης t_m

Παράδειγμα Χωρικής Παρεμβολής Τοπικού Μέσου Όρου



Η προβλεπόμενη τιμή $\hat{z}(t_m)$ είναι ο μέσος όρος των $N(t_m)$ μετρήσεων που εμπίπτουν μέσα στον κύκλο με ακτίνα r και κέντρο το σημείο πρόβλεψης t_m :

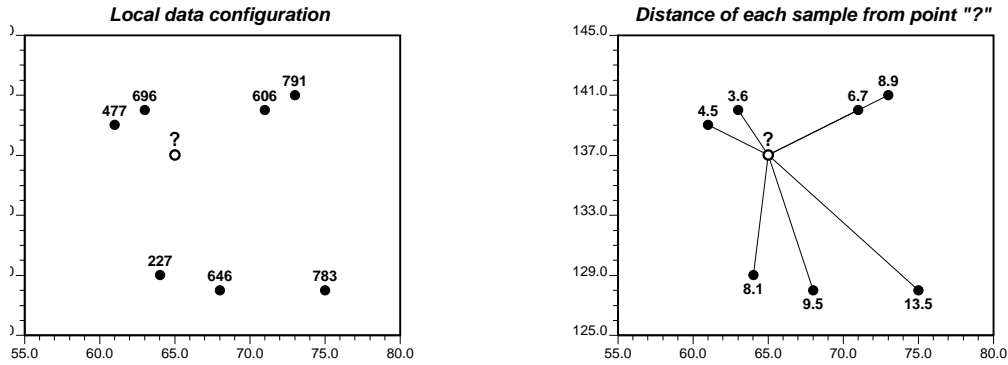
$$\hat{z}(t_m) = \sum_{n=1}^N w_{mn} z(s_n) = \sum_{n=1}^{N(t_m)} \frac{1}{N(t_m)} z(s_n) + \sum_{n'=1}^{N-N(t_m)} 0 z(s_{n'})$$



- (i) η ομαλότητα (smoothness) της εκτιμώμενης "επιφάνειας" εξαρτάται από την ακτίνα r ,
(ii) κυρή αλλά όχι ακριβής μέθοδος παρεμβολής, (iii) μόνο η τοπική εφαρμογή της μεθόδου έχει νόημα, (iv) προσαρμοζόμενες (adaptive) παραλλαγές της μεθόδου έχουν αναπτυχθεί (με ακτίνα που μεταβάλλεται χωρικά)



Μέθοδος Αντιστρόφων Αποστάσεων (1)

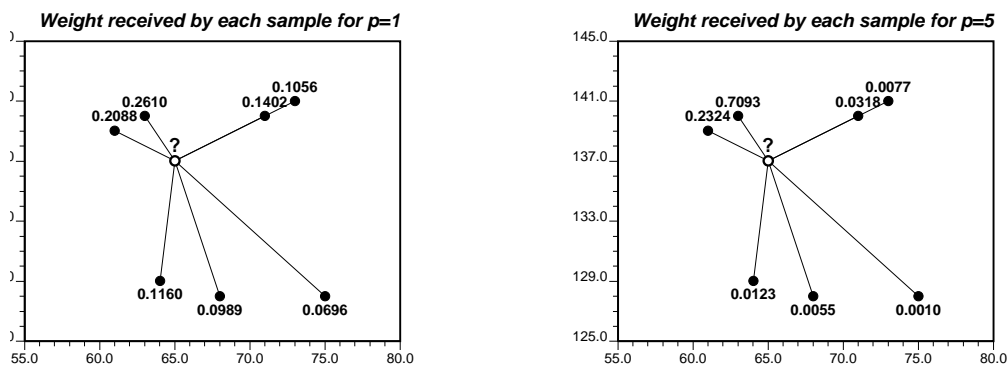


Στη μέθοδο αυτή, τα βάρη παρεμβολής ορίζονται αντιστρόφως ανάλογα (σε δύναμη p) της απόστασης d_{mn} μεταξύ του σημείου πρόβλεψης t_m και του σημείου παρατήρησης s_n :

$$w_{mn} = \frac{d_{mn}^{-p}}{\sum_{n=1}^N d_{mn}^{-p}} = \frac{1/d_{mn}^p}{\sum_{n=1}^N 1/d_{mn}^p}$$

Τα βάρη κυμαίνονται στο διάστημα $[0, 1]$ με άθροισμα 1 (λόγω του παρονομαστή). Το βάρος w_{mn} τείνει στο άπειρο όταν το σημείο πρόβλεψης ταυτίζεται με κάποιο σημείο παρατήρησης $t_m \equiv s_n$

Μέθοδος Αντιστρόφων Αποστάσεων (2)

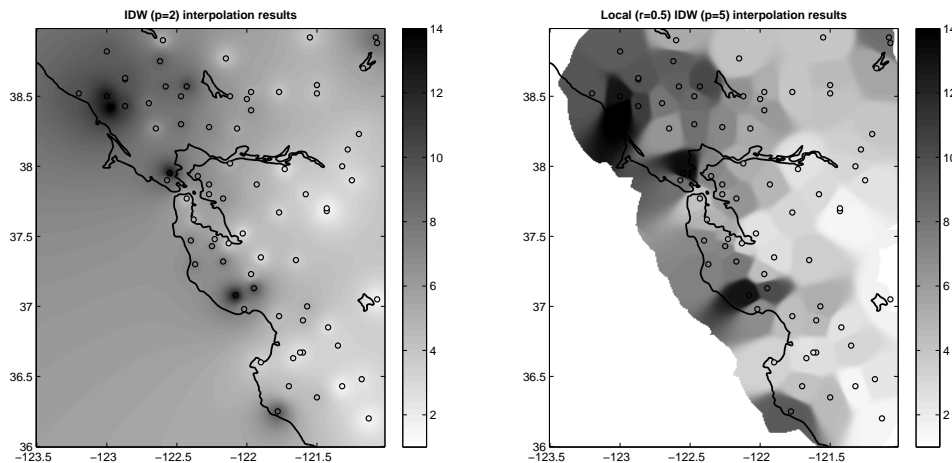


Η επίδραση της τιμής του εκθέτη

- ▶ $p \rightarrow 0$: τα βάρη ομογενοποιούνται: $w_{mn} \simeq 1/N, \forall n$.
Η προβλεπόμενη τιμή $\hat{z}(t_m)$ τείνει στο μέσο όρο των N μετρήσεων
- ▶ $p \rightarrow +\infty$: το πλησιέστερο σημείο παρατήρησης, π.χ., s_n , λαμβάνει το μεγαλύτερο βάρος: $w_{mn} \simeq 1$, ενώ $w_{mn'} \simeq 0, \forall n' \neq n$.
Η προβλεπόμενη τιμή $\hat{z}(t_m)$ τείνει στη μέτρηση $z(s_n)$ του πλησιέστερου σημείου παρατήρησης s_n (παρεμβολή πλησιέστερου γείτονα)



Παραδείγματα Παρεμβολής Αντιστρόφων Αποστάσεων



Η ομαλότητα (smoothness) της εκτιμώμενης "επιφάνειας" τιμών εξαρτάται από τον εκθέτη p , (ii) ακριβής και κυρτή μέθοδος παρεμβολής, (iii) καθολικές και τοπικές παραλλαγές της μεθόδου είναι διαθέσιμες, καθώς και προσαρμοζόμενες παραλλαγές (με τιμή εκθέτη p που μεταβάλλεται στο χώρο)

Εναλλακτική Ερμηνεία του Βάρους Παρεμβολής



Η εξίσωση της παρεμβολής: $\hat{z}(t_m) = \sum_{n=1}^N w_{mn} z(s_n)$ για το σημείο πρόβλεψης t_m , μπορεί να ερμηνευθεί και ως η σχέση που ορίζει τον τρόπο (w_{mn}) με τον οποίο η κάθε μέτρηση $z(s_n)$ συνεισφέρει στην προβλεπόμενη τιμή $\hat{z}(t_m)$

Επίδραση μίας μέτρησης στο χώρο

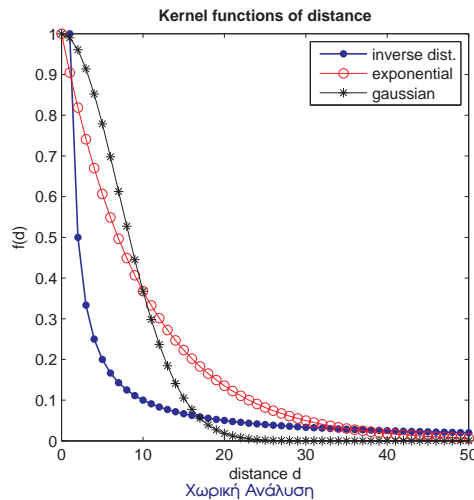
- ▶ μια μέτρηση $z(s_n)$ στο σημείο παρατήρησης s_n "διαχέεται" σε οποιοδήποτε σημείο t_m με βάση την τιμή του βάρους w_{mn} , που συνήθως είναι συνάρτηση της απόστασης d_{mn} μεταξύ του σημείου πρόβλεψης t_m και του σημείου παρατήρησης s_n
- ▶ αν οι τιμές του βάρους κυμαίνονται στο διάστημα $[0, 1]$, τότε η τιμή w_{mn} αντιστοιχεί στο ποσοστό της μέτρησης $z(s_n)$ που "φτάνει" μέχρι το σημείο t_m
- ▶ για ένα συγκεκριμένο σημείο παρατήρησης s_n και για M σημεία πρόβλεψης $\{t_m, m = 1, \dots, M\}$, αντιστοιχούν M βάρη $\{w_{mn}, m = 1, \dots, M\}$, τα οποία ποσοτικοποιούν την "επίδραση" του συγκεκριμένου σημείου παρατήρησης s_n στα M σημεία πρόβλεψης (δηλ. στο χώρο), ανεξάρτητα από την οποιαδήποτε τιμή $z(s_n)$ που έχει μετρηθεί στο σημείο s_n



Συναρτήσεις Πυρήνα (Kernel)

Φθίνουσες συναρτήσεις απόστασης

Η χωρική επίδραση ή επιρροή μιας μέτρησης ορίζεται συνήθως ως μία φθίνουσα συνάρτηση $k(d; \theta)$ της απόστασης d μεταξύ του σημείου μέτρησης και οποιουδήποτε άλλου σημείου στο χώρο (θ είναι μια παράμετρος). Τέτοιες συναρτήσεις μπορεί να είναι: η αντίστροφη συνάρτηση $k(d) = 1/(d^p)$, η εκθετική συνάρτηση $k(d) = \exp(-d/r)$, και η Γκαουσιανή (Gaussian) συνάρτηση $k(d) = \exp(-(d/r)^2)$, όπου r είναι μία παράμετρος που καθορίζει το ρυθμό ελάττωσης. Στις παραπάνω περιπτώσεις $\theta = p, r$, και r , αντίστοιχα



Φ. Κυριακίδης (Παν. Αιγαίου)

Στοιχεία Χωρικής Παρεμβολής

19 / 34

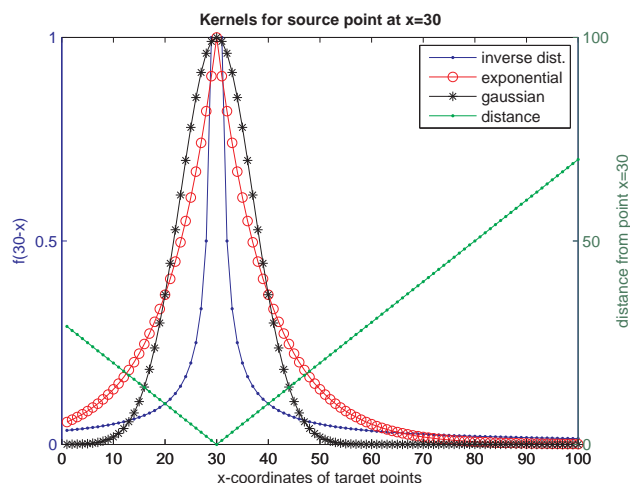
Χωρική Παρεμβολή με τη Μέθοδο Kernel

Συνάρτηση Πυρήνα για Ένα Σημείο Μέτρησης



Χωρική επιρροή μέτρησης

- ▶ για ένα σημείο μέτρησης s_n ορίζεται ένα $(M \times 1)$ διάνυσμα $\mathbf{d}_n = [\|t_m - s_n\|, m = 1, \dots, M]^T$ αποστάσεων μεταξύ όλων των M σημείων πρόβλεψης $\{t_m, m = 1, \dots, M\}$ και του σημείου s_n
- ▶ στις παραπάνω αποστάσεις αντιστοιχεί ένα $(M \times 1)$ διάνυσμα $\mathbf{k}_n = [k(\|t_m - s_n\|; \theta), m = 1, \dots, M]^T$ τιμών συνάρτησης πυρήνα για το σημείο s_n



Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

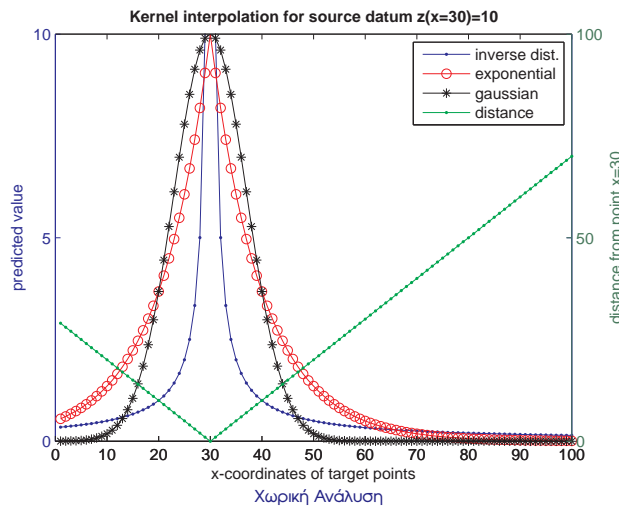
20 / 34



Χωρική Παρεμβολή με Μία Μέτρηση

Βάρη παρεμβολής και συνάρτηση πυρήνα

- ▶ σε ένα σημείο μέτρησης s_n αντιστοιχεί ένα $(M \times 1)$ διάνυσμα $\mathbf{k}_n = [k(\|t_m - s_n\|; \theta), m = 1, \dots, M]^T$ τιμών συνάρτησης πυρήνα για τα M σημεία πρόβλεψης $\{t_m, m = 1, \dots, M\}$
- ▶ για ένα σημείο μέτρησης s_n το σύνολο των τιμών συνάρτησης πυρήνα \mathbf{k}_n αποτελεί και τις τιμές του βάρους με το οποίο η μέτρηση $z(s_n)$ θεωρείται ότι συνεισφέρει στην πρόβλεψη σε κάθε σημείο t_m , ως: $\hat{z}(t_m) = w_{mn}z(s_n) = k(\|t_m - s_n\|; \theta)z(s_n)$



Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

21 / 34

Χωρική Παρεμβολή με τη Μέθοδο Kernel

Χωρική Παρεμβολή με Πολλές Μετρήσεις



$$\hat{z}(t_m) = \sum_{n=1}^N w_{mn}z(s_n) = \frac{\sum_{n=1}^N k(\|t_m - s_n\|; \theta)z(s_n)}{\sum_{n=1}^N k(\|t_m - s_n\|; \theta)} = \sum_{n=1}^N \frac{k(\|t_m - s_n\|; \theta)}{\sum_{n=1}^N k(\|t_m - s_n\|; \theta)} z(s_n)$$

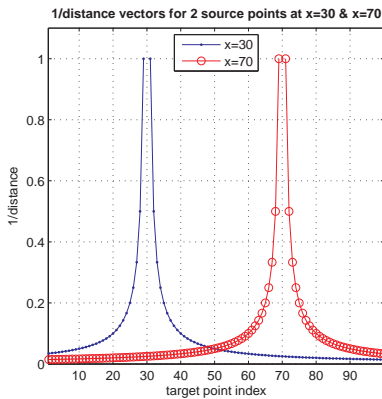
Παρατηρήσεις

- ▶ η συνάρτηση πυρήνα $k(\|t - s\|; \theta)$ ποσοτικοποιεί τη "χωρική διάχυση" στο σημείο στόχο t μίας και μόνο μέτρησης πηγής στο σημείο s . Ο πυρήνας $k(\|t_m - s_n\|; \theta)$, δηλαδή, αναφέρεται σε ένα και μόνο σημείο παρατήρησης s_n
- ▶ στην πράξη, όμως, υπάρχουν N τέτοια σημεία παρατήρησης, και συνεπώς πρέπει να ληφθεί υπ' όψη η "χωρική διάχυση" και των N αντίστοιχων μετρήσεων. Αυτό επιτυγχάνεται διαιρώντας τον αρχικό πυρήνα $k(\|t_m - s_n\|; \theta)$ του σημείου s_n με το άθροισμα όλων των N πυρήνων
- ▶ συνεπώς, το τελικό βάρος παρεμβολής w_{mn} του σημείου παρατήρησης s_n στο σημείο πρόβλεψης t_m είναι το ποσοστό συνεισφοράς του πυρήνα $k(\|t_m - s_n\|; \theta)$ του σημείου παρατήρησης s_n σε σχέση με τη συνολική συνεισφορά των πυρήνων όλων των N σημείων παρατήρησης

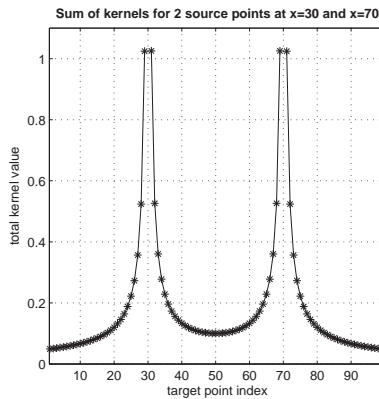


Υπολογισμός Βαρών Παρεμβολής

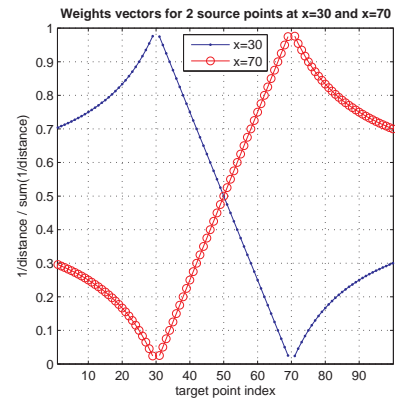
- ▶ άθροισμα N τιμών πυρήνα στο σημείο πρόβλεψης t_m : $\tilde{k}_m = \sum_{n=1}^N k(\|t_m - s_n\|; \theta)$ και ορισμός $(M \times 1)$ διανύσματος $\tilde{\mathbf{k}} = [\tilde{k}_m, m = 1, \dots, M]^T$
- ▶ διαίρεση της αρχικής τιμής του πυρήνα k_{mn} με την αντίστοιχη τιμή του αθροίσματος στο σημείο t_m για τον υπολογισμό του βάρους παρεμβολής: $w_{mn} = k_{mn} / \tilde{k}_m$. Ορισμός του $(M \times 1)$ διανύσματος βαρών $\mathbf{w}_n = [w_{mn}, m = 1, \dots, M]^T$ ως $\mathbf{w}_n = \mathbf{k}_n \oslash \tilde{\mathbf{k}}$. Τα στοιχεία w_{mn} του διανύσματος \mathbf{w}_n παίρνουν τιμές στο διάστημα $[0, 1]$ και έχουν άθροισμα 1 για κάθε σημείο t_m (θεωρούνται ποσοστά)



Φ. Κυριακίδης (Παν. Αιγαίου)



Χωρική Ανάλυση



Στοιχεία Χωρικής Παρεμβολής

Χωρική Παρεμβολή με τη Μέθοδο Kernel

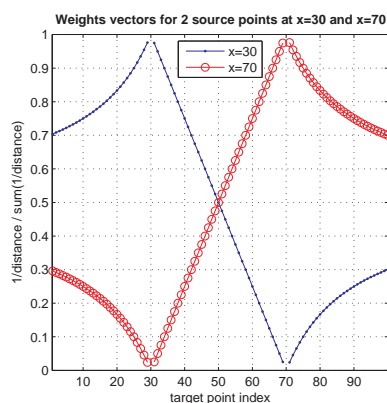
Χωρική Παρεμβολή και Επαλληλία Πυρήνων



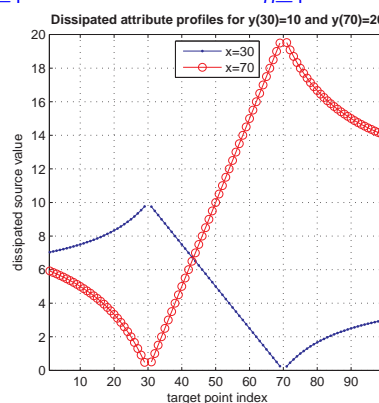
- ▶ διάχυση μέτρησης $z(s_n)$ στο σημείο t_m = κλιμάκωση – πολλαπλασιασμός με την τιμή της μέτρησης $z(s_n)$ – του βάρους w_{mn} . Για το σημείο μέτρησης s_n , υπάρχουν M τέτοιες τιμές, δηλαδή ένα $(M \times 1)$ διάνυσμα $\mathbf{w}_n z(s_n)$ που εμπεριέχει το πρότυπο χωρικής διάχυσης της μέτρησης $z(s_n)$
- ▶ οι τελικές παρεμβαλλόμενες τιμές προέρχονται από την επαλληλία (υπέρθωση ή άθροισμα) των N τέτοιων προτύπων διάχυσης:

$$\hat{z}(t_m) = \sum_{n=1}^N [k(\|t_m - s_n\|; \theta) z(s_n)] / \sum_{n=1}^N k(\|t_m - s_n\|; \theta), \quad m = 1, \dots, M$$

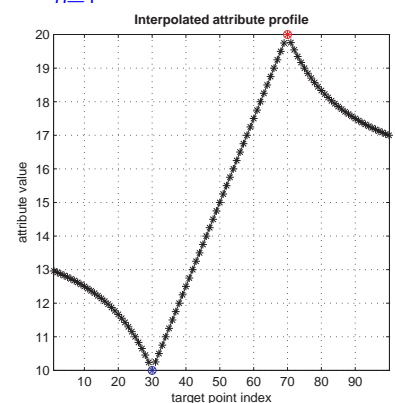
$$\hat{\mathbf{z}} = \sum_{n=1}^N [\mathbf{k}_n z(s_n)] \oslash \tilde{\mathbf{k}} = \sum_{n=1}^N [\mathbf{k}_n z(s_n)] \oslash \sum_{n=1}^N \mathbf{k}_n$$



Φ. Κυριακίδης (Παν. Αιγαίου)



Χωρική Ανάλυση



Στοιχεία Χωρικής Παρεμβολής



Επιλογή Πυρήνα και Εκτίμηση Παραμέτρων

$$\hat{z}(t_m) = \sum_{n=1}^N w_{mn} z(s_n) = \sum_{n=1}^N \frac{k(\|t_m - s_n\|; \theta)}{\sum_{n=1}^N k(\|t_m - s_n\|; \theta)} z(s_n)$$

Επιλογή συνάρτησης πυρήνα;

- ▶ η μέθοδος παρεμβολής με πυρήνα εφαρμόζεται συχνά χωρίς να γίνεται αυτό αντιληπτό: μέθοδος πλησιέστερου γείτονα, μέθοδος αντίστροφων αποστάσεων
- ▶ συνήθως επιλέγονται πυρήνες με το σχήμα της κανονικής κατανομής του Gauss
- ▶ η σωστή επιλογή της συνάρτησης πυρήνα, όμως, προϋποθετεί ένα μοντέλο της χωρικής διάχυσης μιας μέτρησης, πράγμα το οποίο είναι εφικτό μόνο σε λίγες σχετικά περιπτώσεις (συνήθως μέσω νόμων φυσικής, π.χ., διάδοση θερμότητας)

Εκτίμηση παραμέτρων πυρήνα;

- ▶ όταν επιλεγεί κάποια παραμετρική συνάρτηση πυρήνα, το επόμενο βήμα είναι η εκτίμηση της (ή των) παραμέτρου της θ (κάτι τέτοιο δεν είναι απαραίτητο για τη μέθοδο πλησιέστερου γείτονα και την ψηφίδωση Delaunay)
- ▶ σε γενικές γραμμές, ο υπολογισμός παραμέτρων βασίζεται στην ελαχιστοποίηση της διακύμανσης των σφαλμάτων πρόβλεψης/παρεμβολής γνωστών τιμών που αποκρύφθηκαν από το δείγμα των μετρήσεων

Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

25 / 34

Άλλες Μέθοδοι Χωρικής Παρεμβολής

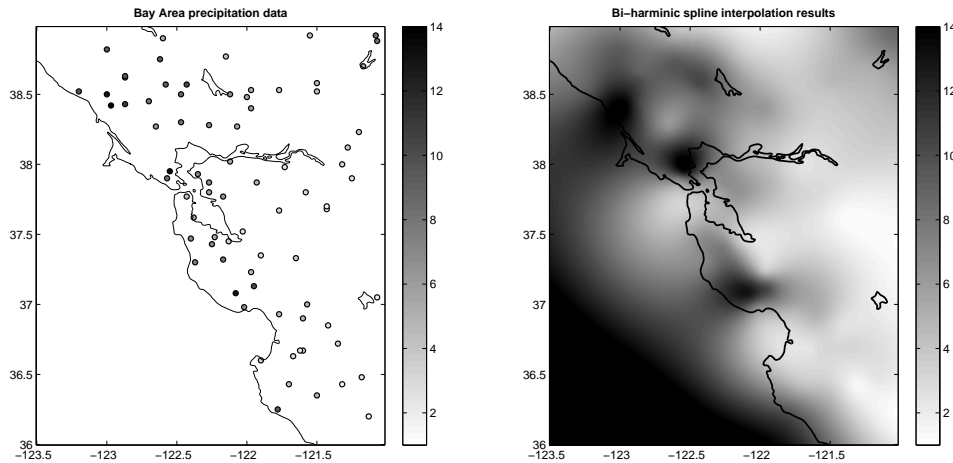
Splines και Radial Basis Functions



- ▶ έχουν αναπτυχθεί δύο συναφείς μέθοδοι χωρικής παρεμβολής, η μέθοδος Spline και η μέθοδος Radial Basis Functions (RBF), στις οποίες η επιφάνεια των παρεμβαλλόμενων τιμών κατασκευάζεται κάτω από τους εξής περιορισμούς (constraints): (i) πιστότητα στις μετρήσεις στα σημεία παρατήρησης, και (ii) πιστότητα σε ένα μοντέλο ομαλότητας (smoothness) της επιφάνειας, το οποίο εκφράζεται με τη μορφή μερικών παραγώγων (partial derivatives)
- ▶ επιπρόσθετα, οι παραπάνω μέθοδοι χωρικής παρεμβολής μοντελοποιούν με σαφήνεια το γεγονός ότι η παρουσία πολλών μετρήσεων θα πρέπει να οδηγεί σε αλλίωση/παραμόρφωση του πυρήνα που αντιστοιχεί σε κάθε σημείο μέτρησης
- ▶ οι παραπάνω μέθοδοι επιλύουν παρόμοια συστήματα γραμμικών εξισώσεων για τον υπολογισμό των βαρών παρεμβολής, αλλά διαφέρουν ως προς τις συναρτήσεις πυρήνα που επιλέγονται. Η μέθοδος RBF χρησιμοποιεί Gaussian πυρήνες, ενώ η μέθοδος spline χρησιμοποιεί πυρήνες που αντιστοιχούν στο όνομα της μεθόδου, π.χ., cubic spline ή thin-plate spline



Παράδειγμα Χωρικής Παρεμβολής Spline



Χαρακτηριστικά

(1) η ομαλότητα (smoothness) της "επιφάνειας" των παρεμβαλλόμενων τιμών εξαρτάται από το όνομα της συνάρτησης spline, εδώ, δι-αρμονικό (bi-harmonic) spline, (2) είναι ακριβής και μη κυρτή μέθοδος παρεμβολής, (3) υπάρχουν καθολικές και τοπικές παραλλαγές της μεθόδου, (4) υπάρχουν και εξομαλυντικά (smoothing) splines, στις οποίες οι μετρήσεις δεν αναπαραγονται λόγω της ύπαρξης σφάλματος μέτρησης – measurement error, (5) οδηγεί σε μη λογικές τιμές σε σημεία πρόβλεψης μακριά από το κυρτό κύτος (convex hull) των σημείων παρατήρησης

Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

27 / 34

Άλλες Μέθοδοι Χωρικής Παρεμβολής

Γεωστατιστικές Μέθοδοι Χωρικής Παρεμβολής



- ▶ έχει αναπτυχθεί μια γεωστατιστική μέθοδος χωρικής παρεμβολής, η μέθοδος Kriging, με πολλαπλές παραλλαγές, η οποία βασίζεται στην έννοια του ημιβαριογράμματος ή του συσχετογράμματος για τον υπολογισμό των βαρών της χωρικής παρεμβολής
- ▶ η μέθοδος Kriging επιλύει ένα διαφορετικό σύστημα γραμμικών εξισώσεων από ότι οι μέθοδοι RBF και splines, αλλά έχει αποδειχθεί ότι τα βάρη που προκύπτουν είναι όμοια σε συγκεκριμένες περιπτώσεις
- ▶ επιπρόσθετα, το ημιβαριόγραμμα και το συσχετογράμμα έχουν στατιστική ερμηνεία, και συνδέονται στενά με την έννοια ενός τυχαίου πεδίου. Το ίδιο όμως ισχύει και για τις συναρτήσεις πυρήνα, οι οποίες μπορούν να συνδεθούν με τις γεωστατιστικές έννοιες του ημιβαριογράμματος και του συσχετογράμματος
- ▶ για παράδειγμα, αν κανείς εφαρμόσει χωρική παρεμβολή με τη μέθοδο RBF χρησιμοποιώντας Gaussian πυρήνες, θα καταλήξει στα ίδια αποτελέσματα με την εφαρμογή της μεθόδου Kriging χρησιμοποιώντας ένα Gaussian μοντέλο συσχετογράμματος. Παρομοίως, τα αποτελέσματα της χωρικής παρεμβολής με τη μέθοδο cubic spline είναι ίδια με αυτά που προκύπτουν από την εφαρμογή της μεθόδου Kriging με ένα κυβικό μοντέλο ημιβαριογράμματος. . .

Αυτό που διαφοροποιεί τη γεωστατιστική μέθοδο παρεμβολής Kriging από τις παραπάνω μεθόδους είναι: (i) η αναζήτηση της συνάρτησης πυρήνα από τα δεδομένα μέσω του εμπειρικού ημιβαριογράμματος, και (ii) η ποσοτικοποίηση της αβεβαιότητας κάθε παρεμβαλλόμενης τιμής

Φ. Κυριακίδης (Παν. Αιγαίου)

Χωρική Ανάλυση

Στοιχεία Χωρικής Παρεμβολής

28 / 34



Σφάλμα Παρεμβολής

Δεδομένα επικύρωσης

Σύνολο μετρήσεων, των οποίων οι θέσεις (σημεία παρατήρησης) δεν συμπεριλήφθηκαν στον υπολογισμό των βαρών παρεμβολής.

Εναλλακτικά: Σύνολο M' σημείων πρόβλεψης με αντίστοιχες γνωστές τιμές

Σφάλμα παρεμβολής

Διαφορά μεταξύ προβλεπόμενης $\hat{z}(t_m)$ και γνωστής μέτρησης $z(t_m)$ στο σημείο επικύρωσης t_m (όπου η αντίστοιχη μέτρηση είναι γνωστή):

$$e(t_m) = \hat{z}(t_m) - z(t_m)$$

Διασταυρωμένη επικύρωση leave-one-out cross-validation

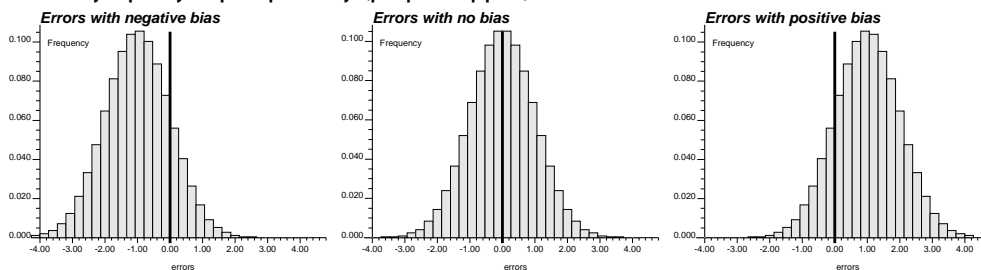
Κατασκευή ενός συνόλου δεδομένων επικύρωσης με τη μικρότερη απώλεια μετρήσεων:

- (i) απόκρυψη μιας γνωστής μέτρησης, (ii) πρόβλεψή της από τις $N - 1$ λοιπές μετρήσεις χρησιμοποιώντας κάποια επιλεγμένη μέθοδο παρεμβολής ή κάποιες τιμές παραμέτρων, (iii) επανάληψη των βημάτων (i) και (ii) για τις $N - 1$ λοιπές μετρήσεις για τον υπολογισμό N cross-validation προβλέψεων και αντίστοιχων σφαλμάτων

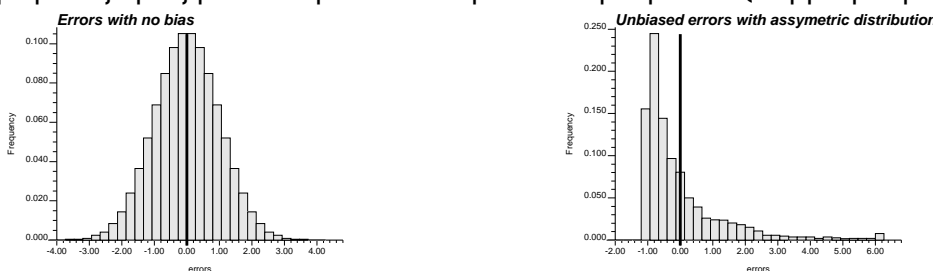
Κατανομές Σφαλμάτων



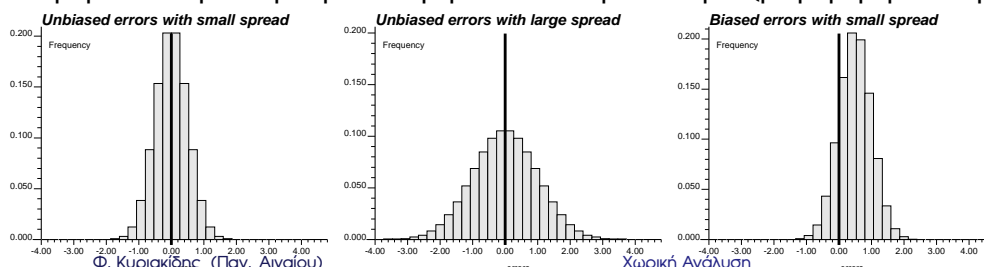
Μέσος όρος σφάλματος (μεροληψία) κοντά στο 0;



Παρόμοιες τιμές μέσου όρου και διάμεσου σφαλμάτων (συμμετρική κατανομή);



Μικρή διασπορά σφαλμάτων γύρω από το μέσο όρο (μικρή τιμή διακύμανσης);





Στατιστικά Σφαλμάτων

Μέτρα κεντρικής τάσης

- ▶ μέσος όρος σφαλμάτων ή μεροληψία (bias): $ME = \frac{1}{M} \sum_{m=1}^M e(t_m) = \bar{e}$
- ▶ μέσο απόλυτο σφάλμα (mean absolute error): $MAE = \frac{1}{M} \sum_{m=1}^M |e(t_m)|$

Μέτρα διασποράς

- ▶ άθροισμα τετραγωνικών σφαλμάτων (sum of squared error): $SSE = \sum_{m=1}^M [e(t_m) - \bar{e}]^2$
- ▶ διακύμανση σφαλμάτων (error variance): $VE = \frac{1}{M} \sum_{m=1}^M [e(t_m) - \bar{e}]^2$

Μέσο τετραγωνικό σφάλμα (mean squared error)

$$MSE = \frac{1}{M} \sum_{m=1}^M e^2(t_m)$$

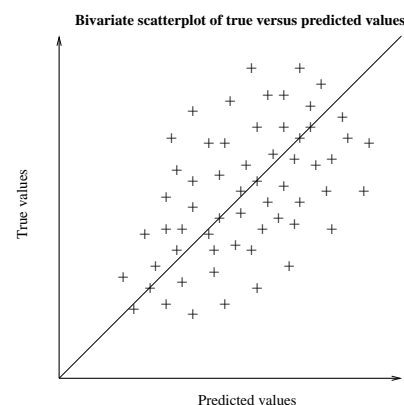
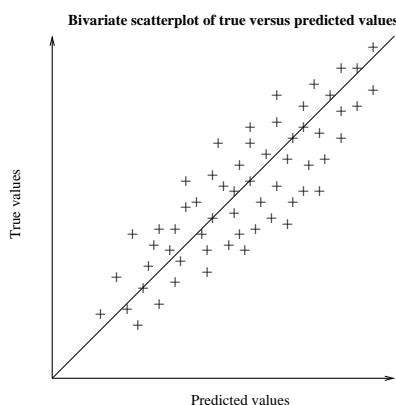
Σημείωση: $MSE = \text{μεροληψία}^2 + \text{διακύμανση}$, σφάλματος

Συσχέτιση Προβλέψεων και Παρατηρήσεων



Διασπορόγραμμα προβλεπόμενων και παρατηρούμενων τιμών

- ▶ βέλτιστη περίπτωση όταν τα σημεία (ζευγάρια τιμών $\{z(t_m), \hat{z}(t_m)\}$) βρίσκονται πολύ κοντά στη διχοτόμο
- ▶ ο συντελεστής συσχέτισης $\rho_{z\hat{z}}$, ή ο συντελεστής προσδιορισμού R^2 ενός γραμμικού μοντέλου, αποτελούν περιγραφικά στατιστικά του διασπορογράμματος



Μεγαλύτερες τιμές συντελεστή συσχέτισης (κοντά στο $+1$), υποδηλώνουν καλύτερη συμφωνία προβλεπόμενων και παρατηρούμενων τιμών, και συνεπώς καλύτερα αποτελέσματα χωρικής παρεμβολής



Βασικά Σημεία Διάλεξης (1)

Χωρική παρεμβολή

- ▶ πρόβλεψη (εκτίμηση), με βάση γνωστές μετρήσεις μιας χωρικής μεταβλητής, άγνωστων τιμών αυτής σε ένα σύνολο σημείων, είτε διεσπαρμένων είτε σε κανονικό κάρναβο (επιφάνεια τιμών)
- ▶ γραμμική χωρική παρεμβολή = σταθμισμένο άθροισμα μετρήσεων
- ▶ ο προσδιορισμός των βαρών της χωρικής παρεμβολής βασίζεται στη χωρική διάταξη των σημείων παρατήρησης σε σχέση με το εκάστοτε σημείο πρόβλεψης

Μέθοδοι χωρικής παρεμβολής

- ▶ πλησιέστερου γείτονα – Nearest Neighbor ή Voronoi
- ▶ τριγωνισμού Delaunay – Triangulated Irregular Network
- ▶ τοπικού μέσου όρου – Local Mean
- ▶ αντίστροφων αποστάσεων – Inverse Distance Weighting
- ▶ συναρτήσεων πυρήνα – Kernel functions
- ▶ συναρτήσεων Radial Basis Functions και Splines
- ▶ γεωστατιστικές μέθοδοι – βασιζόμενες στο ημιβαριόγραμμα

Βασικά Σημεία Διάλεξης (2)



Εκτίμηση ποιότητας χωρικής παρεμβολής

- ▶ απόκρυψη κάποιων παρατηρήσεων (δεδομένα επικύρωσης) και πρόβλεψη αυτών από τις υπόλοιπες μετρήσεις (cross-validation)
- ▶ σφάλματα παρεμβολής (διαφορά προβλεπόμενων και παρατηρούμενων τιμών) και κατανομή αυτών
- ▶ μέσος όρος σφαλμάτων (μεροληψία) και διασπορά σφαλμάτων
- ▶ μέσο τετραγωνικό σφάλμα (MSE) = μεροληψία² + διακύμανση (σφαλμάτων)
- ▶ διαγνωστικά στατιστικά γραφήματα, όπως το διασπορογράμμα προβλεπόμενων και παρατηρούμενων τιμών – συντελεστής συσχέτισης διασπορογράμματος: όσο μεγαλύτερος (κοντά στο +1) τόσο πιο ακριβής η παρεμβολή

Άλλες χρήσεις της διασταυρωμένης επικύρωσης (cross-validation)

- ▶ επιλογή βέλτιστης μεθόδου παρεμβολής από ένα σύνολο μεθόδων, με βάση τα χαρακτηριστικά των αντίστοιχων κατανομών σφαλμάτων
- ▶ υπολογισμός βέλτιστης τιμής παραμέτρου παρεμβολής, όπως ο εκθέτης στη μέθοδο των αντίστροφων αποστάσεων