
Μεθοδολογίες Παρεμβολής για την Πλήρωση Κενών σε Δορυφορικές Εικόνες

Διπλωματική Εργασία

Αγγελική Καραβόλια



Πολυτεχνείο Κρήτης

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών

Υπολογιστών

Χανιά, Κρήτη, Ελλάδα

Επιβλέπων: Διονύσιος Χριστόπουλος, Καθηγητής, Σχολή Ηλεκτρολόγων
Μηχανικών και Μηχανικών Υπολογιστών, Πολυτεχνείου Κρήτης

Τριμελής Επιτροπή:

Διονύσιος Χριστόπουλος, Καθηγητής ΗΜΜΥ

Μιχάλης Ζερβάκης, Καθηγητής ΗΜΜΥ

Γεώργιος Πετρόπουλος, Καθηγητής Τμήμα Γεωγραφίας, Χαροκόπειο Πα-
νεπιστήμιο

Ευχαριστίες

Με την παρούσα διπλωματική εργασία, μου δόθηκε η ευκαιρία να ασχοληθώ με ένα αντικείμενο υψηλού ενδιαφέροντος, αυτού της σύγκρισης μεθοδολογιών παρεμβολής για την πλήρωση κενών σε δορυφορικές εικόνες. Θα ήθελα λοιπόν να ευχαριστήσω τον κύριο Χριστόπουλο Διονύσιο, Καθηγητή του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών και επιβλέποντα της διπλωματικής μου εργασίας, τον οποίο είχα την τιμή να συνεργαστώ τόσο στα πλαίσια της διπλωματικής μου, όσο και στα πλαίσια του μαθήματος Στοχαστικές Διαδικασίες και Ανάλυση Χρονοσειρών. Επίσης, θα ήθελα να τον ευχαριστήσω για την πολύτιμη καθοδήγηση που μου προσέφερε σε θέματα σύνταξης και δομής της διπλωματικής μου εργασίας καθώς επίσης και για τη γνώση που μου παρείχε.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω την οικογένεια μου, τη μητέρα μου Νούλα, τον πατέρα μου Γιώργο και την αδελφή μου Αμάντα. Τους ευχαριστώ για τη στήριξη που μου παρείχαν κατά τη διάρκεια των σπουδών μου, δίνοντας μου και τη δυνατότητα να σπουδάσω αυτό που ήθελα.

Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου, για τη συνεχή ενθάρρυνση, κατανόηση και στήριξη σε όλη τη διαδρομή των σπουδών μου.

Περίληψη

Στις μέρες μας οι δορυφορικές εικόνες είναι καθοριστικής σημασίας για ποικίλες εφαρμογές όπως η ανίχνευση αλλαγών στη Γη μετά από ανθρώπινες παρεμβάσεις και φυσικές καταστροφές. Οι δορυφορικές εικόνες συχνά περιέχουν κενά και αυτό δημιουργεί προβλήματα σε περαιτέρω επεξεργασία ιδιαίτερα αν τα βήματα επεξεργασίας είναι αυτοματοποιημένα. Στους στόχους της διπλωματικής εργασίας αυτής είναι η διερεύνηση μεθοδολογιών για την πλήρωση κενών στις δορυφορικές εικόνες. Πιο ειδικά, η σύγκριση των αιτιοκρατικών μεθοδολογιών με τη στοχαστική μέθοδο INNC που βασίζεται στο μοντέλο Ising. Η ύπαρξη κενών οφείλεται σε διάφορους λόγους οπτικά εμπόδια (π.χ. νέφωση), οι τροχιές που ακολουθούν οι δορυφόροι και δυσλειτουργίες των αισθητήρων. Τα δεδομένα που χρησιμοποιήθηκαν για τον έλεγχο των μεθοδολογιών πλήρωσης κενών χωρίζονται σε δύο κατηγορίες: συνθετικά και δορυφορικά δεδομένα. Τα συνθετικά δεδομένα δημιουργήθηκαν χρησιμοποιώντας τη συνάρτηση Ackley. Τα δορυφορικά δεδομένα κατηγοριοποιούνται σε δύο σύνολα, ένα πραγματικό και ένα προσομοιωμένο. Τα πραγματικά δεδομένα αναφέρονται σε θερμοκρασίες της επιφάνειας της Γης κατά τη διάρκεια της ημέρας, όπως μετρήθηκαν από το όργανο Terra του δορυφόρου MODIS, για το έτος 2016. Το σύνολο των προσομοιωμένων δεδομένων υλοποιήθηκε με την προσαρμογή ενός χωρικού μοντέλου Gauss σε ένα τυχαίο δείγμα 2 500 παρατηρήσεων από τα προαναφερόμενα δεδομένα MODIS.

Στα πλαίσια της εργασίας πραγματοποιήθηκε διερευνητική στατιστική ανάλυση της χωρικής συνέχειας με τη βοήθεια της συνάρτησης αυτοσυσχέτισης και του βαριογράμματος στις δύο ορθογώνιες διευθύνσεις του πλέγματος. Οι αιτιοκρατικές μεθοδολογίες οι οποίες χρησιμοποιήθηκαν για την πλήρωση κενών είναι οι εξής: η μέθοδος των πλησιέστερων γειτόνων (nearest neighbor), η μέθοδος παρεμβολής ελάχιστης καμπυλότητας (v4), η μέθοδος παρεμβολής φυσικών γειτόνων (natural neighbor), η μέθοδος κυβικής παρεμβολής (cubic) και η μέθοδος γραμμικής παρεμβολής (linear). Οι μεθοδολογίες αυτές υλοποιήθηκαν μέ-

σω της εντολής `griddata` και ορισμένες από αυτές μέσω της εντολής `scatteredInterpolant` στο προγραμματιστικό περιβάλλον Matlab. Αυτές οι δύο εντολές πραγματοποιούν παρεμβολή της επιφάνειας στα σημεία διερεύνησης με βάση γνωστές μετρήσεις στα σημεία του δείγματος. Οι αιτιοκρατικές μεθοδολογίες εφαρμόστηκαν και με τις δύο εντολές διότι ο τρόπος υλοποίησης των δύο εντολών διαφέρει. Επιπλέον, εφαρμόσαμε τη στοχαστική μεθοδολογία με αλληλεπιδράσεις κοντινότερων γειτόνων, η οποία βασίζεται στο μοντέλο Ising (INNC). Η μέθοδος INNC προσεγγίζει συνεχείς μεταβλητές ανάλογα με τον αριθμό των διακριτών τάξεων που έχουμε ορίσει. Επίσης, λαμβάνει υπόψιν τοπικά τις τιμές του δείγματος προκειμένου να σχηματίσει την πρόβλεψη. Εφαρμόζοντας μια προσέγγιση βασισμένη στη μεθοδολογία Monte Carlo, προβλέπει την τάξη της μεταβλητής σε σημεία όπου δεν υπάρχουν μετρήσεις.

Για την επιλογή της βέλτιστης μεθοδολογίας χωρικής παρεμβολής χρησιμοποιήθηκε η στατιστική μέθοδος της διασταυρωτικής επιβεβαίωσης (cross validation). Στα πλαίσια της συγκριτικής ανάλυσης των μεθοδολογιών χρησιμοποιήθηκαν η οπτική επισκόπηση των ανακατασκευασμένων εικόνων, τα διαγράμματα διασποράς των εκτιμώμενων και μετρημένων τιμών (τιμών επιβεβαίωσης), καθώς επίσης και στατιστικά μέτρα διασταυρωτικής επιβεβαίωσης. Επιπλέον, στην ανάλυση των συνθετικών δεδομένων διερευνήθηκαν διαφορετικές χωρικές διατάξεις και θέσεις κενών σημείων/περιοχών. Δηλαδή, για έλεγχο χρησιμοποιήσαμε τις εξής διατάξεις κενών: 1) τυχαία κατανομή 40 κενών εικονοστοιχείων, 2) συνεχή περιοχή 303 κενών εικονοστοιχείων και 3) τυχαία κατανομή 303 κενών εικονοστοιχείων.

Η μέθοδος παρεμβολής ελάχιστης καμπυλότητας (υλοποιείται με την εντολή `griddata`, επιλογή `v4`) είναι βέλτιστη για τα συνθετικά δεδομένα με τυχαία κατανομή 40 κενών εικονοστοιχείων. Η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 5×5 είναι καλύτερη, σε σύγκριση με τη μέθοδο των πλησιέστερων γειτόνων (υλοποιείται με την εντολή `griddata` και την `scatteredInterpolant`, επιλογή `nearest neighbor`). Η μέθοδος παρεμβολής φυσικών γειτόνων (υλοποιείται με την εντολή `griddata` και την `scatteredInterpolant`, επιλογή `natural neighbor`)

είναι υπολογιστικά γρηγορότερη σε σύγκριση με τις προηγούμενες μεθοδολογίες. Η μέθοδος γραμμικής παρεμβολής (υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `linear`) είναι βέλτιστη για τη συνεχή περιοχή 303 κενών εικονοστοιχείων στην περίπτωση που τα κενά είναι τοποθετημένα στο δεξιό άκρο της εικόνας. Επίσης, η μέθοδος παρεμβολής φυσικών γειτόνων (υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `natural neighbor`) υπερτερεί όσον αφορά την υπολογιστική ταχύτητα. Η μέθοδος INNC δεν εμφάνισε ικανοποιητικά αποτελέσματα, σε σχέση με τις προηγούμενες μεθοδολογίες. Στη συνέχεια, διερευνήθηκε η συνεχής περιοχή 303 κενών εικονοστοιχείων, με τα κενά τοποθετημένα στο κέντρο. Η μέθοδος γραμμικής παρεμβολής (υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `linear`) είναι βέλτιστη όπως συμπεράναμε και στην περίπτωση που τα κενά είναι τοποθετημένα στο δεξιό σύνορο της εικόνας. Η μέθοδος των πλησιέστερων γειτόνων (υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `nearest neighbor`) είναι πιο αποδοτική σε χρόνο. Η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 30×30 είναι καλύτερη, σε σύγκριση με τη μέθοδο των πλησιέστερων γειτόνων (υλοποιείται με την εντολή `griddata`, επιλογή `nearest neighbor`) και τη μέθοδο της παρεμβολής ελάχιστης καμπυλότητας (υλοποιείται με την εντολή `griddata`, επιλογή `v4`). Έπειτα, η τυχαία κατανομή 303 κενών εικονοστοιχείων ερευνήθηκε χρησιμοποιώντας τις ίδιες μεθοδολογίες. Η μέθοδος παρεμβολής ελάχιστης καμπυλότητας (υλοποιείται με την εντολή `griddata`, επιλογή `v4`) είναι βέλτιστη, όπως και στην περίπτωση της τυχαίας κατανομή 40 κενών εικονοστοιχείων. Η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 5×5 είναι πιο αποδοτική σε χρόνο, σε σύγκριση με τις αιτιοκρατικές μεθοδολογίες.

Η μέθοδος παρεμβολής φυσικών γειτόνων (υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `natural neighbor`) είναι βέλτιστη για τα δορυφορικά δεδομένα. Σύμφωνα με τα αποτελέσματα, η μέθοδος γραμμικής παρεμβολής (υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `linear`) υπερτερεί όσον αφορά την υπολογιστική ταχύτητα, στην περίπτωση των πραγματικών δεδομένων. Αντίθετα, στην περίπτωση των προσομοιωμένων δεδομένων, η μέθοδος παρεμβολής φυσικών γειτόνων

(υλοποιείται με την εντολή `scatteredInterpolant`, επιλογή `natural neighbor`) είναι η πιο αποδοτική σε χρόνο. Τέλος, η μέθοδος INNC χρειάστηκε σημαντικά περισσότερο χρόνο για την εφαρμογή της στα δορυφορικά δεδομένα.

Λέξεις κλειδιά Δορυφορικές εικόνες, πλήρωση κενών, αιτιοκρατικές μεθοδολογίες, Ising model, χωρική παρεμβολή.

Nowadays, satellite images are crucial for a variety of applications such as detecting changes on Earth following human interventions and natural disasters. Satellite images often contain gaps thus creating problems in further processing, especially if the processing steps are automatized. The objectives of this thesis are to investigate methodologies for gap filling in satellite images; more specifically, the comparison of deterministic methodologies with the stochastic INNC method based on the Ising model. The existence of gaps is due to optically thick aerosol layers, satellite trajectories and sensor malfunction. The data used to test the gap-filling methodologies are divided into two categories: synthetic and satellite data. The synthetic data were generated using the Ackley function. The satellite data involve two datasets, one real and one simulated. The real data comprises daytime surface temperatures of the Earth measured by the Terra instrument of the MODIS satellite for the year 2016. The simulated data set was implemented by fitting a Gaussian spatial model to a random sample of 2 500 observations from the aforementioned MODIS data.

We first performed an exploratory statistical analysis of spatial continuity in the data using the autocorrelation function and the variogram in the two orthogonal directions of the grid. The deterministic methodologies that were used for interpolation are: the nearest neighbor method, the minimum curvature interpolation method, the natural neighbor interpolation method, the cubic interpolation method and the linear interpolation method. These methodologies were implemented via the `griddata` command and some of them via the `scatteredInterpolant` command in the Matlab programming environment. Both commands perform interpolation of the surface at the points of investigation based on known measurements at the sample points. The deterministic interpolation methodologies were implemented with both functions because the implementation of the algorithms differs between the two functions. In addition, we have applied a stochastic methodology with nearest neighbor interactions, which is based on the Ising model (INNC). The INNC method approximates continuous variables according to a number of discrete classes we have defined. It also takes into account local values of the sample in order

to estimate the prediction. Applying an approach based on Monte Carlo methodology, it predicts the class of the target variable at points where no measurements are provided.

The statistical method of cross validation was used to select the optimal spatial interpolation methodology. A comparative analysis of the methodologies included a visual overview of the reconstructed images, scatterplots of the estimated and measured values (confirmation values), as well as statistical measures of cross validation. In addition, different spatial configurations and locations of missing points/areas were examined in the analysis of the synthetic data. Specifically, we used the following configurations with gaps for testing: 1) the random distribution of 40 missing pixels, 2) the continuous region of 303 missing pixels, and 3) the random distribution of 303 missing pixels.

The minimum curvature interpolation method (implemented by `griddata`, option `v4`) is the best method for the synthetic data with a random distribution of 40 missing pixels. The INNC method with eight classes and a 5×5 stencil is optimal, compared to the nearest neighbor method (implemented by `griddata` and `scatteredInterpolant`, option `nearest neighbor`). The natural neighbor interpolation method (implemented by `griddata` and `scatteredInterpolant`, option `natural neighbor`) is computationally faster compared to the previous methodologies. The linear interpolation method (implemented by `scatteredInterpolant`, option `linear`) is the best method for the continuous region of 303 missing pixels in the case where the missing pixels are located in the right part of the image. Also, the natural neighbor interpolation method (implemented by `scatteredInterpolant`, option `natural neighbor`) is preferable in terms of computational time. The INNC method did not have satisfying results, compared to the previous methodologies. Then the continuous region of 303 missing pixels was explored, with the gaps being located in the center. The linear interpolation method (implemented by `scatteredInterpolant`, option `linear`) is the best method as we concluded in the case when the gaps are located in the right part of the image. The nearest neighbor method (implemented by `scatteredInterpolant`,

option `nearest neighbor`) is more efficient method in terms of computational time. The INNC method with eight classes and a 30×30 stencil is a better method compared to the nearest neighbor method (implemented by `griddata`, option `nearest neighbor`) and the least curvature interpolation method (implemented by `griddata`, option `v4`). Furthermore, the random distribution of 303 missing pixels was examined using the same methodologies. The minimum curvature interpolation method (implemented by `griddata`, option `v4`) is considered to be the best method, like in the case of the random distribution of 40 missing pixels. The INNC method with 16 classes and a 5×5 stencil is more efficient method in terms of computational time, compared to the deterministic methodologies.

The natural neighbor interpolation method (implemented by `scatteredInterpolant`, option `natural neighbor`) is the best method for satellite data. According to the results, the linear interpolation method (implemented by `scatteredInterpolant`, option `linear`) is better in terms of computational time, for the real satellite data. In contrast, for simulated data, the method of natural neighbor interpolation (implemented by `scatteredInterpolant`, option `natural neighbor`) is more efficient in terms of computational time. Finally, the INNC method needed significantly more time for implementation on the satellite data.

Keywords Satellite images, gap filling, deterministic methodologies, Ising model, spatial interpolation.

Περιεχόμενα

Περίληψη	iii
1 Εισαγωγή	1
1.1 Σκοπός Εργασίας	2
2 Θεωρητικό Υπόβαθρο για Ανάλυση Χωρικών Δεδομένων	5
2.1 Βασικές Έννοιες	5
2.1.1 Τυχάίο Πεδίο	5
2.1.2 Κοινή Συνάρτηση Αθροιστικής Κατανομής	6
2.1.3 Συνάρτηση Συνδιακύμανσης	6
2.1.4 Συνάρτηση Αυτοσυσχέτισης	7
2.1.5 Βαριόγραμμα	8
2.1.6 Στασιμότητα	10
2.1.7 Εργοδικότητα	12
2.2 Ανακατασκευή Κενών σε Χωρικά Δεδομένα	15
2.2.1 Γενικές Έννοιες	15
2.2.2 Χωρική Παρεμβολή	16
2.2.3 Μέτρα Διασταυρωτικής Επιβεβαίωσης	17
3 Διερευνητική Ανάλυση Δεδομένων	21
3.1 Δορυφορικά Δεδομένα	21

3.2	Συνθετικά Δεδομένα με Χρήση της Συνάρτησης Ackley . .	23
3.3	Διερευνητική Στατιστική Ανάλυση	25
4	Μεθοδολογίες Χωρικής Παρεμβολής	35
4.1	Αιτιοκρατικές Μέθοδοι	35
4.2	Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων . .	39
4.2.1	Ο Αλγόριθμος Πολυεπίπεδης Παρεμβολής INNC . .	42
5	Αποτελέσματα	49
5.1	Αποτελέσματα για Συνθετικά Δεδομένα	49
5.1.1	Τυχαία Κατανομή 40 Κενών Εικονοστοιχείων . . .	51
5.1.2	Συνεχής Περιοχή 303 Κενών Εικονοστοιχείων . . .	56
5.1.3	Τυχαία Κατανομή 303 Κενών Εικονοστοιχείων . . .	71
5.2	Αποτελέσματα για Δορυφορικά Δεδομένα	78
5.2.1	Αιτιοκρατικές Μέθοδοι	78
5.2.2	Μέθοδος INNC	85
5.3	Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής	89
6	Συμπεράσματα	99
	Βιβλιογραφία	104

Κατάλογος Σχημάτων

2.1	Παράδειγμα βαριογράμματος [9]	9
3.1	(i): Συνθετικά δεδομένα από συνάρτηση Ackley. Τυχαία κατανομή 40 κενών εικονοστοιχείων (ii): μπλε και (iii): άσπρο χρώμα.	24
3.2	Πρώτη εικόνα (i): πλήρες σύνολο δεδομένων, (ii): πραγματικό σύνολο δεδομένων εκπαίδευσης. Δεύτερη εικόνα (iii): πλήρες σύνολο δεδομένων, (iv): προσομοιωμένο σύνολο δεδομένων εκπαίδευσης.	27
3.3	Αποτελέσματα συνάρτησης αυτοσυσχέτισης. Διακεκομμένη κόκκινη γραμμή αντιπροσωπεύει το μέσο όρο αυτοσυσχέτισης. Πρώτη εικόνα ανά (i): στήλες, (ii): γραμμές. Δεύτερη εικόνα ανά (iii): στήλες, (iv): γραμμές.	29
3.4	Βαριογράμματα. Πρώτη εικόνα ανά (i): στήλες, (ii): γραμμές. Δεύτερη εικόνα ανά (iii): στήλες, (iv): γραμμές. . . .	32
3.5	Μέσο βαριόγραμμα (i): Πρώτη εικόνα. (ii): Δεύτερη εικόνα. . . .	33

- 5.1 (i): Τυχαία κατανομή 40 κενών εικονοστοιχείων. Αποτελέσματα από την εντολή `griddata`. Μέθοδος παρεμβολής: (ii): επιλογή `nearest neighbor`, (iii): επιλογή `v4`, (iv): επιλογή `natural neighbor`, (v): επιλογή `cubic`, (vi): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nmb` (`nearest neighbor interpolation`). 52
- 5.2 Αποτελέσματα από την εντολή `scatteredInterpolant` για τυχαία κατανομή 40 κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (i): επιλογή `nearest neighbor`, (ii): επιλογή `natural neighbor`, (iii): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nmb` (`nearest neighbor interpolation`). 53
- 5.3 (i): Μέθοδος `INNC 8` τάξεις. (ii): Διάγραμμα διασποράς πρόβλεψης 40 κενών εικονοστοιχείων. 55
- 5.4 (i): Συνεχής περιοχή 303 κενών εικονοστοιχείων. Αποτελέσματα από την εντολή `griddata` για συνεχή περιοχή 303 κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (ii): επιλογή `nearest neighbor`, (iii): επιλογή `v4`, (iv): επιλογή `natural neighbor`, (v): επιλογή `cubic`, (vi): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nmb` (`nearest neighbor interpolation`). 58
- 5.5 Αποτελέσματα από την εντολή `scatteredInterpolant` για συνεχή περιοχή 303 κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (i): επιλογή `nearest neighbor`, (ii): επιλογή `natural neighbor`, (iii): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nmb` (`nearest neighbor interpolation`). 60
- 5.6 (i): Μέθοδος `INNC 8` τάξεις. (ii): Μέθοδος `INNC 16` τάξεις. 61

5.7	Διαγράμματα διασποράς πρόβλεψης 303 κενών εικονοστοιχείων σε συνεχή περιοχή. (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.	63
5.8	(i): Συνεχής περιοχή 303 κενών εικονοστοιχείων στο κέντρο. Αποτελέσματα από την εντολή <code>griddata</code> για συνεχή περιοχή 303 κενών εικονοστοιχείων στο κέντρο. Μέθοδος παρεμβολής: (ii): επιλογή <code>nearest neighbor</code> , (iii): επιλογή <code>v4</code> , (iv): επιλογή <code>natural neighbor</code> , (v): επιλογή <code>cubic</code> , (vi): επιλογή <code>linear</code> . Για λόγους συντομίας αναφερόμαστε στην επιλογή <code>nearest neighbor</code> με την ονομασία <code>nnb</code> (<code>nearest neighbor interpolation</code>).	66
5.9	Αποτελέσματα από την εντολή <code>scatteredInterpolant</code> για συνεχή περιοχή 303 κενών εικονοστοιχείων στο κέντρο. Μέθοδος παρεμβολής: (i): επιλογή <code>nearest neighbor</code> , (ii): επιλογή <code>natural neighbor</code> , (iii): επιλογή <code>linear</code> . Για λόγους συντομίας αναφερόμαστε στην επιλογή <code>nearest neighbor</code> με την ονομασία <code>nnb</code> (<code>nearest neighbor interpolation</code>).	67
5.10	(i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.	69
5.11	Διαγράμματα διασποράς πρόβλεψης 303 κενών εικονοστοιχείων σε συνεχή περιοχή στο κέντρο. (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.	70
5.12	(i): Τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Αποτελέσματα από την εντολή <code>griddata</code> για τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (ii): επιλογή <code>nearest neighbor</code> , (iii): επιλογή <code>v4</code> , (iv): επιλογή <code>natural neighbor</code> , (v): επιλογή <code>cubic</code> , (vi): επιλογή <code>linear</code> . Για λόγους συντομίας αναφερόμαστε στην επιλογή <code>nearest neighbor</code> με την ονομασία <code>nnb</code> (<code>nearest neighbor interpolation</code>).	73

5.13 Αποτελέσματα από την εντολή <code>scatteredInterpolant</code> για τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (i): επιλογή <code>nearest neighbor</code> , (ii): επιλογή <code>natural neighbor</code> , (iii): επιλογή <code>linear</code> . Για λόγους συντομίας αναφερόμαστε στην επιλογή <code>nearest neighbor</code> με την ονομασία <code>nrb</code> (<code>nearest neighbor interpolation</code>).	74
5.14 (i): Μέθοδος INNC 8 τάξεις, (ii): Μέθοδος INNC 16 τάξεις.	76
5.15 Διαγράμματα διασποράς πρόβλεψης ισάριθμων κενών εικονοστοιχείων σε τυχαία κατανομή. (i): Μέθοδος INNC 8 τάξεις, (ii): Μέθοδος INNC 16 τάξεις.	77
5.16 Αποτελέσματα από την εντολή <code>griddata</code> με μέθοδο παρεμβολής επιλογή <code>nearest neighbor</code> (i): πραγματικά δεδομένα, (iii): προσομοιωμένα δεδομένα. Από την εντολή <code>scatteredInterpolant</code> με μέθοδο παρεμβολής επιλογή <code>nearest neighbor</code> (ii): πραγματικά δεδομένα, (iv): προσομοιωμένα δεδομένα. Για λόγους συντομίας αναφερόμαστε στην επιλογή <code>nearest neighbor</code> με την ονομασία <code>nrb</code> (<code>nearest neighbor interpolation</code>).	79
5.17 Αποτελέσματα από την εντολή <code>griddata</code> . Πραγματικά δεδομένα με μέθοδο παρεμβολής (i): επιλογή <code>natural neighbor</code> , (ii): επιλογή <code>cubic</code> , (iii): επιλογή <code>linear</code>	81
5.18 Αποτελέσματα από την εντολή <code>griddata</code> . Προσομοιωμένα δεδομένα με μέθοδο παρεμβολής (i): επιλογή <code>natural neighbor</code> , (ii): επιλογή <code>cubic</code> , (iii): επιλογή <code>linear</code>	82
5.19 Αποτελέσματα από την εντολή <code>scatteredInterpolant</code> με μέθοδο παρεμβολής επιλογή <code>natural neighbor</code> (i): πραγματικά δεδομένα, (iii): προσομοιωμένα δεδομένα. Μέθοδος παρεμβολής επιλογή <code>linear</code> (ii): πραγματικά δεδομένα, (iv): προσομοιωμένα δεδομένα.	84

- 5.20 Αποτελέσματα από τη μέθοδο INNC. Πραγματικά δεδομένα (i): 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iii): 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Προσομοιωμένα δεδομένα (ii) 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iv) 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 86
- 5.21 Διαγράμματα διασποράς πρόβλεψης πραγματικών δεδομένων μέθοδος INNC (i): 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iii): 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Πρόβλεψης προσομοιωμένων δεδομένων μέθοδος INNC (ii): 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iv): 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . 88

Κατάλογος Πινάκων

5.1	Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της τυχαίας κατανομής 40 κενών εικονοστοιχείων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το <code>sI</code> την εντολή <code>scatteredInterpolant</code> και το <code>nnb</code> τη μεθοδολογία <code>nearest neighbor</code> . Το $N_c = 8$ και $m_{\max} = 5$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.	90
5.2	Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της συνεχόμενης περιοχής 303 κενών εικονοστοιχείων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το <code>sI</code> την εντολή <code>scatteredInterpolant</code> και το <code>nnb</code> τη μεθοδολογία <code>nearest neighbor</code> . Στην εντολή <code>griddata</code> σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 8$ και το $m_{\max} = 30$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 30$, τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.	91

- 5.3 Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της συνεχόμενης περιοχής 303 κενών εικονοστοιχείων, τοποθετημένων στο κέντρο. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το `sI` την εντολή `scatteredInterpolant` και το `nnb` τη μεθοδολογία `nearest neighbor`. Στην εντολή `griddata` σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 8$ και το $m_{\max} = 30$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 30$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα. 93
- 5.4 Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της τυχαίας κατανομής 303 κενών εικονοστοιχείων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το `sI` την εντολή `scatteredInterpolant` και το `nnb` τη μεθοδολογία `nearest neighbor`. Στην εντολή `griddata` σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 8$ και το $m_{\max} = 5$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 5$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα. 95

- 5.5 Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση των πραγματικών δεδομένων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του ορθογώνιου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το `sI` την εντολή `scatteredInterpolant` και το `nnb` τη μεθοδολογία `nearest neighbor`. Στην εντολή `griddata` σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 16$ και το $m_{\max} = 20$ και στην άλλη περίπτωση το $N_c = 16$ και $m_{\max} = 30$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα. 97
- 5.6 Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση των προσομοιωμένων δεδομένων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του ορθογώνιου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το `sI` την εντολή `scatteredInterpolant` και το `nnb` τη μεθοδολογία `nearest neighbor`. Στην εντολή `griddata` σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 16$ και το $m_{\max} = 20$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 30$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα. 98

Κεφάλαιο 1

Εισαγωγή

Οι δορυφορικές εικόνες, εδώ και πολλά χρόνια, προσφέρουν τεράστια ποσότητα δεδομένων για την επιφάνεια της γης και αυτό συμβάλλει στην λεπτομερή ανάλυση και στην εξαγωγή χαρακτηριστικών. Επιπλέον, η αύξηση της διακριτικής ικανότητας των δορυφόρων και η μείωση του κόστους απόκτησης των δορυφορικών εικόνων έχει βοηθήσει σε ένα εκτενές εύρος εφαρμογών. Συνεισφέρουν δηλαδή, στη γεωργία, στις τηλεπικοινωνίες καθώς και σε καταστάσεις έκτακτης ανάγκης, όπως σε πυρκαγιές και σε φυσικές καταστροφές.

Η ύπαρξη κενών στις δορυφορικές εικόνες αποτελεί ένα σημαντικό πρόβλημα που επηρεάζει τη χωρική διαθεσιμότητα των επιφανειακών παρατηρήσεων. Αυτό έχει ως αποτέλεσμα να εμποδίζει την ανάλυση της επιφάνειας της γης διότι με την παρουσία των κενών χάνονται πληροφορίες. Τα δεδομένα που λείπουν, λοιπόν είναι ένα σύνηθες ζήτημα και αυτό οφείλεται λόγω εσφαλμένων μετρήσεων και δυσλειτουργίας αισθητήρων [8, 13].

Διάφορες μελέτες έχουν αναφερθεί στο πως έχουν αντιμετωπίσει τα κενά που υπάρχουν στα χωροχρονικά δεδομένα, τα οποία προκαλούνται από διάφορους παράγοντες για παράδειγμα διαφορετικοί χρόνοι επαναφοράς δορυφόρων, παρουσία πάγου ή χιονιού και μόλυνση από παρεμβολές ραδιοσυχνοτήτων. Για παράδειγμα, η βάση δεδομένων για την υγρασία εδάφους (Soil Moisture Climate Change Initiative αλλιώς,

1.1. Σκοπός Εργασίας

CCI SM), η οποία είναι η μακροβιότερη διαθέσιμη καταγραφή δορυφορικών δεδομένων, έχει χρησιμοποιήσει μεθοδολογίες για την αντιμετώπιση του προβλήματος. Ορισμένες από τις μεθοδολογίες που έχουν προταθεί για τη συμπλήρωση των κενών είναι η γραμμική παλινδρόμηση (linear regression), η μέθοδος ordinary Kriging (OK), η μέθοδος παλινδρόμησης kriging (regression kriging) και τα γενικά γραμμικά μοντέλα (general linear models αλλιώς, GLM). Επίσης, η μέθοδος SVM (Support Vector Machines) λειτουργεί καλά για την κάλυψη χωρικών κενών στα δορυφορικά δεδομένα. Όμως, η αποτελεσματική συμπλήρωση των κενών αποτελεί δύσκολη πρόκληση επειδή το υπολογιστικό κόστος είναι υψηλό, ιδίως για εξαιρετικά μεγάλα σύνολα δορυφορικών δεδομένων [1].

Οι συμβατικές τεχνικές γεωστατιστικής παρεμβολής, όπως η μέθοδος Kriging, ενδέχεται να μην είναι εφαρμόσιμες λόγω υψηλού κόστους επεξεργασίας των δεδομένων. Για τη συμπλήρωση των κενών σε πολύ μεγάλα σύνολα δεδομένων απαιτούνται ιδιαίτερα αποδοτικές υπολογιστικές λύσεις. Συνεπώς, για την αντιμετώπιση του προβλήματος έχουν αναπτυχθεί μέθοδοι παρεμβολής που ασχολούνται με την πλήρωση κενών με πολύ θετικά αποτελέσματα. Αυτό θα παρουσιαστεί στην παρούσα διπλωματική εργασία η οποία έχει ως αντικείμενο την μελέτη μεθόδων παρεμβολής. Καθώς και την εφαρμογή τους στο πρόβλημα πλήρωσης κενών, με σκοπό την ανακατασκευή της εικόνας.

1.1 Σκοπός Εργασίας

Σκοπός της διπλωματικής εργασίας είναι να συγκριθούν διαφορετικές μεθοδολογίες χωρικής παρεμβολής. Η σύγκριση πραγματοποιείται χρησιμοποιώντας μέτρα διασταυρωτικής επιβεβαίωσης (cross validation measures). Τα μέτρα διασταυρωτικής επιβεβαίωσης που χρησιμοποιήσαμε είναι το μέσο σφάλμα (ME), η ρίζα τετραγωνικού σφάλματος (RMSE), το μέσο απόλυτο σφάλμα (MAE) και ο συντελεστής συσχέτισης Pearson (ρ). Αυτά τα μέτρα θα τα αναλύσουμε παρακάτω στο Κεφάλαιο 2. Στην παρούσα εργασία διερευνήθηκαν μεθοδολογίες για την πλήρωση κενών σε δορυφορικές εικόνες. Χρησιμοποιήθηκαν τόσο συνθετικά δεδομένα

1.1. Σκοπός Εργασίας

όσο και δορυφορικά δεδομένα θερμοκρασίας. Τα συνθετικά δεδομένα δημιουργήθηκαν με χρήση της συνάρτησης Ackley, η οποία είναι μία μη κυρτή συνάρτηση σε χώρο δύο διαστάσεων. Στην ανάλυση των συνθετικών δεδομένων διερευνήθηκαν διαφορετικές χωρικές διατάξεις και θέσεις κενών σημείων/περιοχών. Οι περιπτώσεις που χρησιμοποιήθηκαν είναι: 1) τυχαία κατανομή 40 κενών εικονοστοιχείων, 2) συνεχής περιοχής 303 κενών εικονοστοιχείων και 3) τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Έπειτα, τα δορυφορικά δεδομένα χρησιμοποιήθηκαν από την εργασία [5]. Πιο συγκεκριμένα, τα δορυφορικά δεδομένα αποτελούνται από ένα προσομοιωμένο και ένα πραγματικό σύνολο δεδομένων. Τα δεδομένα που αποτέλεσαν το πραγματικό σύνολο δεδομένων προέρχονται από θερμοκρασίες της επιφάνειας της Γης κατά τη διάρκεια της ημέρας, όπως μετρήθηκαν από το όργανο Terra του δορυφόρου MODIS στις 4 Αυγούστου 2016. Όσον αφορά το σύνολο των προσομοιωμένων δεδομένων, υλοποιήθηκε με την προσαρμογή ενός χωρικού μοντέλου Gauss σε ένα τυχαίο δείγμα 2500 παρατηρήσεων από τα προαναφερόμενα δεδομένα του δορυφόρου MODIS. Για τη διερεύνηση της χωρικής συνέχειας των δεδομένων πραγματοποιήθηκε διερευνητική χωρική στατιστική ανάλυση. Πιο ειδικά, έγινε εκτίμηση της συνάρτησης αυτοσυσχέτισης και του βαριογράμματος στις δύο ορθογώνιες διευθύνσεις του πλέγματος. Οι μεθοδολογίες που αναφέραμε προηγουμένως είναι οι παρακάτω: η μέθοδος των πλησιέστερων γειτόνων (υλοποιημένη με την εντολή `griddata` και `scatteredInterpolant`, επιλογή `nearest neighbor`), η μέθοδος παρεμβολής ελάχιστης καμπυλότητας (υλοποιημένη με την εντολή `griddata`, επιλογή `v4`), η μέθοδος παρεμβολής φυσικών γειτόνων (υλοποιημένη με την εντολή `griddata` και `scatteredInterpolant`, επιλογή `natural neighbor`), η μέθοδος κυβικής παρεμβολής (υλοποιημένη με την εντολή `griddata`, επιλογή `cubic`) και η μέθοδος γραμμικής παρεμβολής (υλοποιημένη με την εντολή `griddata` και `scatteredInterpolant`, επιλογή `linear`). Οι μεθοδολογίες αυτές καλούνται ως αιτιοκρατικές και θα τις αναλύσουμε στο Κεφάλαιο 4. Επίσης, η στοχαστική μέθοδος INNC, που βασίζεται στο μοντέλο Ising εφαρμόστηκε στα δεδομένα. Η ανάλυση της οποίας θα παρουσιαστεί στο Κεφάλαιο 4. Τέλος, στο Κεφάλαιο 5 αξιολογούνται τα αποτελέσματα με βάση τις μεθοδολογίες που χρησιμο-

1.1. Σκοπός Εργασίας

ποιήθηκαν και στο Κεφάλαιο 6 παρουσιάζονται τα συμπεράσματα και οι προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο για Ανάλυση Χωρικών Δεδομένων

Αυτό το κεφάλαιο επικεντρώνεται σε βασικές έννοιες οι οποίες είναι απαραίτητες για την κατανόηση των χωρικών δεδομένων. Ορισμένες από τις έννοιες που θα αναλύσουμε είναι το τυχαίο πεδίο, η κοινή συνάρτηση αθροιστικής κατανομής, η συνάρτηση συνδιακύμανσης, η συνάρτηση αυτοσυσχέτισης, το βαριόγραμμα, η στασιμότητα και η εργοδικότητα. Στη συνέχεια, όπως θα δούμε παρακάτω, παρουσιάζονται οι έννοιες παρεμβολή και χωρική παρεμβολή. Με τη βοήθεια της παρεμβολής θα πραγματοποιηθεί η ανακατασκευή των κενών στα δεδομένα. Τέλος, τα μέτρα διασταυρωτικής επιβεβαίωσης θα παρουσιαστούν, τα οποία είναι χρήσιμα για την επιλογή της βέλτιστης μεθοδολογίας.

2.1 Βασικές Έννοιες

2.1.1 Τυχαίο Πεδίο

Ένα χωρικό τυχαίο πεδίο συμβολίζεται $Z(\mathbf{s})$ σε ένα πολυδιάστατο χώρο με δείκτη $\mathbf{s} \in \mathbb{R}^d$ και περιλαμβάνει μια μέση συνάρτηση $m(\mathbf{s})$ και

2.1. Βασικές Έννοιες

μια διακύμανση μηδενικού μέσου $Z'(\mathbf{s})$. Το \mathbf{s} χρησιμοποιείται για να δηλώσει τη χωρική θέση. Όσον αφορά τη μέση τιμή και τη συνάρτηση συνδιακύμανσης, ορίζονται ως εξής, $m(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})]$ και $C_{zz}(\mathbf{s}_1, \mathbf{s}_2)$ [7].

Ομογενές ονομάζεται ένα τυχαίο πεδίο εφόσον έχει σταθερή μέση τιμή $m(\mathbf{s}) = m_z = m$ και η συνάρτηση συνδιακύμανσης $C_{zz}(\mathbf{s}_1, \mathbf{s}_2)$ εξαρτάται μόνο από το διάνυσμα $\mathbf{s}_1 - \mathbf{s}_2$ μεταξύ των δύο σημείων παρατήρησης, πιο συγκεκριμένα, υποθέτοντας $m = 0$,

$$C_Z(\mathbf{s}_1 + \mathbf{s}_2, \mathbf{s}_1) = \mathbb{E}[Z(\mathbf{s}_1 + \mathbf{s}_2) \cdot Z(\mathbf{s}_1)]. \quad (2.1)$$

Η συνάρτηση συνδιακύμανσης της στοχαστικής διαδικασίας σε δύο σημεία παραμέτρων εξαρτάται από την απόσταση καθώς και από την κατεύθυνση του διανύσματος μεταξύ των δύο αυτών σημείων.

2.1.2 Κοινή Συνάρτηση Αθροιστικής Κατανομής

Σε αυτή την ενότητα, δίνουμε βάση στις κοινές πιθανότητες που περιγράφουν την εξάρτηση των τιμών του τυχαίου πεδίου ταυτόχρονα σε δύο σημεία. Η κοινή αθροιστική συνάρτηση κατανομής στα σημεία \mathbf{s}_1 και \mathbf{s}_2 είναι η πιθανότητα να μην ξεπερνούν οι τιμές του τυχαίου πεδίου, το z_1 στο \mathbf{s}_1 και αντίστοιχα το z_2 στο \mathbf{s}_2 . Αυτό διατυπώνεται ως εξής:

$$\mathbb{F}_z(z_1, z_2; \mathbf{s}_1, \mathbf{s}_2) = P(Z(\mathbf{s}_1) \leq z_1 \wedge Z(\mathbf{s}_2) \leq z_2), \quad (2.2)$$

όπου το σύμβολο \wedge αναπαριστά το λογικό τελεστή "and".

2.1.3 Συνάρτηση Συνδιακύμανσης

Η συνάρτηση συνδιακύμανσης (covariance function) είναι το γινόμενο των τιμών του πεδίου σε δύο διαφορετικά σημεία μείον το γινόμενο των μέσων συναρτήσεων στα ίδια σημεία, πιο συγκεκριμένα,

$$C_{zz}(\mathbf{s}_1, \mathbf{s}_2) = \mathbb{E}[Z(\mathbf{s}_1)Z(\mathbf{s}_2)] - \mathbb{E}[Z(\mathbf{s}_1)]\mathbb{E}[Z(\mathbf{s}_2)]. \quad (2.3)$$

2.1. Βασικές Έννοιες

Επίσης, η παραπάνω εξίσωση είναι ισοδύναμη με την αναμενόμενη τιμή του γινομένου των διακυμάνσεων σε αυτές τις θέσεις, δηλαδή,

$$C_{zz}(\mathbf{s}_1, \mathbf{s}_2) = \mathbb{E}[Z'(\mathbf{s}_1)Z'(\mathbf{s}_2)]. \quad (2.4)$$

Εάν τα σημεία \mathbf{s}_1 και \mathbf{s}_2 συμπίπτουν, τότε το αποτέλεσμα της συνδιακύμανσης είναι ίδιο με αυτό της διασποράς του τυχαίου πεδίου στο σημείο αυτό, άρα,

$$C_{zz}(\mathbf{s}_1, \mathbf{s}_1) = \sigma_z^2(\mathbf{s}_1). \quad (2.5)$$

Η συνάρτηση συνδιακύμανσης ελαττώνεται καθώς αυξάνεται η απόσταση μεταξύ των δύο σημείων. Όταν η απόσταση τείνει στο άπειρο τότε η συνάρτηση τείνει στο μηδέν. Παρατηρείται ότι οι υψηλές τιμές, όσον αφορά τη συνάρτηση συνδιακύμανσης, υποδεικνύουν υψηλό βαθμό συσχέτισης μεταξύ των διακυμάνσεων του πεδίου στα δύο σημεία απ' ότι οι χαμηλές τιμές της συνάρτησης συνδιακύμανσης [6].

2.1.4 Συνάρτηση Αυτοσυσχέτισης

Η συνάρτηση αυτοσυσχέτισης (autocorrelation function) ορίζεται ως το πηλίκο της συνάρτησης συνδιακύμανσης προς τη διασπορά μεταξύ των τιμών του πεδίου σε συγκεκριμένα σημεία δηλαδή,

$$\rho_{zz}(\mathbf{s}_1, \mathbf{s}_2) = \frac{C_{zz}(\mathbf{s}_1, \mathbf{s}_2)}{\sigma_z(\mathbf{s}_1)\sigma_z(\mathbf{s}_2)}. \quad (2.6)$$

Να σημειωθεί ότι όλες οι δυνατές τιμές της συνάρτησης αυτοσυσχέτισης βρίσκονται εντός του διαστήματος $[-1, 1]$, ικανοποιώντας τη συνθήκη

$$-1 \leq \rho_{zz} \leq 1. \quad (2.7)$$

Η συνάρτηση αυτοσυσχέτισης προσδιορίζει τον τρόπο με τον οποίο η γραμμική εξάρτηση των τιμών του πεδίου τροποποιείται ως προς την απόσταση μεταξύ των δύο σημείων. Έτσι, μπορεί να χρησιμοποιηθεί

2.1. Βασικές Έννοιες

για τη σύγκριση της χωρικής διάστασης των συσχετίσεων σε χωρικές διαδικασίες με διαφορετική διασπορά.

Εάν $\rho_{zz}(\mathbf{s}_1, \mathbf{s}_2)$ είναι περίπου ίσο με το 0, τότε οι τυχαίες μεταβλητές $Z(\mathbf{s}_1)$ και $Z(\mathbf{s}_2)$ είναι ασυσχέτιστες. Από την άλλη πλευρά, οι αρνητικές τιμές του $\rho_{zz}(\mathbf{s}_1, \mathbf{s}_2)$ υποδηλώνουν ότι οι υψηλότερες τιμές του πεδίου στο \mathbf{s}_1 , συσχετίζονται με τις χαμηλότερες τιμές στο \mathbf{s}_2 και αντίστροφα. Αντίθετα, οι θετικές τιμές του $\rho_{zz}(\mathbf{s}_1, \mathbf{s}_2)$ υποδηλώνουν ότι οι τιμές του πεδίου στο \mathbf{s}_1 και στο \mathbf{s}_2 συσχετίζονται με τον ίδιο τρόπο για όλες τις υλοποιήσεις [6].

2.1.5 Βαριόγραμμα

Το βαριόγραμμα (variogram) περιγράφει τη διαφορά των τιμών του πεδίου συναρτήσει του μέτρου και της διεύθυνσης του διανύσματος απόστασης. Αξιοποιώντας το βαριόγραμμα, ουσιαστικά παίρνουμε ενδείξεις για το πόσο παρόμοιες ή ανόμοιες είναι οι μετρήσεις των γειτονικών σημείων ως συνάρτηση της απόστασης τους το ένα από το άλλο. Αν ισχύουν οι συνθήκες της εσωτερικής στασιμότητας, το βαριόγραμμα ορίζεται ως εξής:

Το βαριόγραμμα είναι ανάλογο της διασποράς της τιμής της προσαύξησης του πεδίου, $Z(\mathbf{s}_1) - Z(\mathbf{s}_2)$, μεταξύ δύο σημείων \mathbf{s}_1 και \mathbf{s}_2 σύμφωνα με την εξίσωση

$$\gamma_{zz}(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2} \text{Var}\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\}. \quad (2.8)$$

Εάν το τυχαίο πεδίο έχει σταθερή μέση τιμή, τότε η συνάρτηση του βαριογράμματος δίνεται ως εξής

$$\gamma_{zz}(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2} \mathbb{E}\{[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]^2\}. \quad (2.9)$$

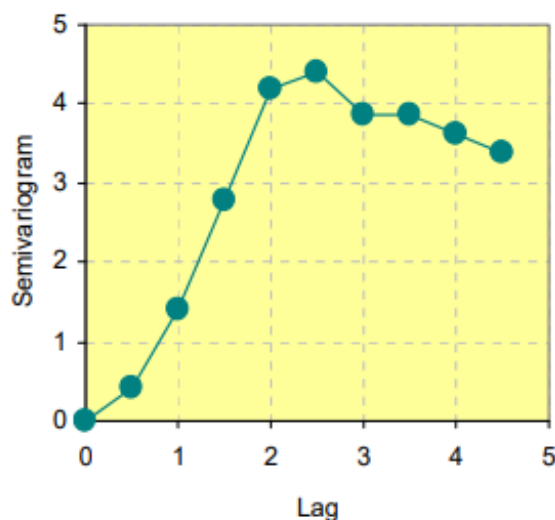
Η Εξ. (2.9) ισχύει για τυχαία πεδία που είναι στάσιμα.

Να σημειωθεί σε αυτό το σημείο ότι η ποσότητα που ορίζεται στην Εξ. (2.8) για λόγους απλότητας θα αναφέρεται με τον όρο βαριόγραμμα,

2.1. Βασικές Έννοιες

αν και στη βιβλιογραφία αναφέρεται ευρέως ως ημιβαριόγραμμα [3].

Το Σχήμα 2.1 αποτελεί ένα παράδειγμα για την καλύτερη κατανόηση του βαριογράμματος. Γενικά, για τα βαριογράμματα ισχύει ότι όσο μεγαλώνει η απόσταση, τόσο λιγότερα ζεύγη έχουμε. Το βαριόγραμμα λοιπόν είναι το άθροισμα των τετραγώνων των διαφορών διαιρεμένο με το πλήθος των ζευγών μαζί με το $1/2$. Στον οριζόντιο άξονα μετράμε τις αποστάσεις, σε γενική απόσταση βλέπουμε ότι το βαριόγραμμα έχει την τιμή 0. Επειδή το βαριόγραμμα εστιάζει στο τετράγωνο των διαφορών όταν η απόσταση είναι μηδενική, δεν υπάρχουν διαφορές. Άρα γι' αυτό ξεκινάει από το μηδέν το βαριόγραμμα. Επομένως, έχει μια συμπεριφορά αντίθετη από την αυτοσυσχέτιση. Σε μηδενική απόσταση είναι μηδέν και όσο αυξάνει η απόσταση, αυτό σημαίνει ότι τα σημεία απομακρύνονται το ένα από το άλλο. Αν δούμε τη διαφορά τους αυτή έχει μεγαλύτερη διασπορά, επομένως το βαριόγραμμα τείνει να έχει μεγαλύτερες τιμές. Αν υπάρχει στασιμότητα ή αλλιώς μια στάσιμη διαδικασία το βαριόγραμμα τείνει να πάει σε μια σταθερή τιμή που την ονομάζουμε οροφή.



Σχήμα 2.1: Παράδειγμα βαριογράμματος [9]

2.1.6 Στασιμότητα

Ένα τυχαίο πεδίο $Z(\mathbf{s})$ ονομάζεται ασθενώς στάσιμο (weakly stationary) εφόσον ισχύουν τα εξής:

$$\mathbb{E}[Z(\mathbf{s})] = m_z \text{ για κάθε } \mathbf{s} \in D \quad (2.10)$$

$$C_{zz}(\mathbf{s}_1, \mathbf{s}_2) = C_{zz}(\mathbf{s}_1 - \mathbf{s}_2) \text{ για κάθε } \mathbf{s}_1, \mathbf{s}_2 \in D. \quad (2.11)$$

Εάν το τυχαίο πεδίο είναι ασθενώς στάσιμο τότε η μέση τιμή είναι σταθερή και η συνάρτηση συνδιακύμανσης είναι συναρτήσεως του \mathbf{r} , όπου $\mathbf{r} := \mathbf{s}_1 - \mathbf{s}_2$ και αντιπροσωπεύει τη χωρική υστέρηση.

Οι συνέπειες όταν ένα τυχαίο πεδίο είναι ασθενώς στάσιμο είναι οι παρακάτω:

- Ομοιόμορφη διασπορά

$$\mathbb{E}[\{Z'(\mathbf{s})\}^2] = \sigma_z^2, \text{ όπου } Z'(\mathbf{s}) \text{ είναι η μέση μηδενική διακύμανση.} \quad (2.12)$$

- Αμετάβλητη συνάρτηση συσχέτισης

$$\rho_{zz}(\mathbf{s}_1, \mathbf{s}_2) = \rho_{zz}(\mathbf{s}_1 - \mathbf{s}_2). \quad (2.13)$$

- Βαριόγραμμα - Συνάρτηση συνδιακύμανσης

$$\gamma_{zz}(\mathbf{s}_1 - \mathbf{s}_2) = \sigma_z^2 - C_{zz}(\mathbf{s}_1 - \mathbf{s}_2). \quad (2.14)$$

Όταν το τυχαίο πεδίο είναι στάσιμο, το βαριόγραμμα έχει ορισμένες χαρακτηριστικές ιδιότητες οι οποίες αναλύονται παρακάτω.

- Όταν η χωρική υστέρηση είναι ίση με μηδέν δηλαδή, $\|\mathbf{r}\| = 0$ τότε η τιμή του βαριογράμματος είναι μηδέν.
- Το βαριόγραμμα εμφανίζει ασυνέχεια στη διασπορά του τυχαίου χωρικού θορύβου με μήκος υστέρησης το οποίο είναι ελάχιστα μεγαλύτερο από το μηδέν δηλαδή, $\|\mathbf{r}\| \rightarrow 0_+$.

2.1. Βασικές Έννοιες

- Το βαριόγραμμα παρατηρείται ότι αυξάνεται, καθώς το μήκος υστέρησης $\|\mathbf{r}\|$ αυξάνεται.
- Ο τρόπος με τον οποίο αυξάνεται το βαριόγραμμα προσδιορίζεται από ένα χαρακτηριστικό μήκος.
- Καθώς το $\|\mathbf{r}\|$ τείνει στο άπειρο, το βαριόγραμμα παρατηρείται ότι σταθεροποιείται.
- Εάν υπάρχει τμηματική ανισοτροπία, η τιμή του κατωφλιού εξαρτάται από τη διεύθυνση του μήκους υστέρησης \mathbf{r} .
- Εάν υπάρχει γεωμετρική ανισοτροπία, η τιμή του κατωφλιού, όπως και στην περίπτωση της τμηματικής ανισοτροπίας, εξαρτάται από τη διεύθυνση του μήκους υστέρησης \mathbf{r} .

Το βαριόγραμμα και η συνάρτηση συνδιασποράς αναφέρονται σε συσχετίσεις δύο σημείων. Στην περίπτωση που τα τυχαία πεδία χαρακτηρίζονται από στασιμότητα, τόσο το βαριόγραμμα όσο και η συνάρτηση συνδιασποράς περιέχουν ίδια πληροφορία. Αντίθετα, όταν τα τυχαία πεδία ορίζονται ως μη στάσιμα, το βαριόγραμμα και η συνάρτηση συνδιασποράς παρουσιάζουν διαφορές μεταξύ τους. Παρακάτω, παραθέτονται οι διαφορές που εντοπίζονται μεταξύ τους.

- Εάν η συνάρτηση συνδιασποράς του τυχαίου πεδίου $Z(\mathbf{s})$ εξαρτάται μόνο από τη χωρική υστέρηση, τότε το βαριόγραμμα θεωρείται αμετάβλητο ως προς τη μετάθεση σύμφωνα με την εξ. (2.14). Αυτό συμβαίνει διότι έχουμε υποθέσει ότι η συνάρτηση συνδιασποράς είναι αμετάβλητη ως προς τη μετάθεση.
- Προηγουμένως, έγινε αναφορά ότι εάν η συνάρτηση συνδιασποράς είναι αμετάβλητη ως προς τη μετάθεση, το ίδιο ισχύει και για το βαριόγραμμα. Το αντίστροφο όμως δεν ισχύει απαραίτητα. Δηλαδή, εάν το βαριόγραμμα του τυχαίου πεδίου $Z(\mathbf{s})$ είναι αμετάβλητο ως προς τη μετάθεση, δεν σημαίνει ότι το ίδιο ισχύει απαραίτητα και για τη συνάρτηση συνδιασποράς.

2.1. Βασικές Έννοιες

- Το βαριόγραμμα εξαρτάται μόνο από το διάνυσμα της χωρικής υστέρησης, ενώ η συνάρτηση συνδιασποράς εξαρτάται από τις θέσεις και των δύο σημείων.
- Το βαριόγραμμα μπορεί να καθοριστεί για τυχαία πεδία με διασπορά που απειρίζεται, σε αντίθεση με τη συνδιασπορά.
- Όσον αφορά την εκτίμηση του βαριογράμματος, δεν απαιτείται η γνώση της μέσης τιμής και αυτό αποτελεί ένα πλεονέκτημα για το βαριόγραμμα σε σύγκριση με τη συνδιασπορά. Στην περίπτωση που η μέση τιμή δεν είναι γνωστή εκ των προτέρων, ο υπολογισμός της γίνεται με βάση το δειγματικό μέσο όρο, ο οποίος χρησιμοποιείται για τον καθορισμό της συνάρτησης συνδιασποράς. Όμως, ο δειγματικός μέσος όρος ενδέχεται να δώσει μια μεροληπτική εκτίμηση της μέσης τιμής για τυχαία πεδία που είναι συσχετισμένα, με βάση τη διαμόρφωση της χωρικής δειγματοληψίας.

2.1.7 Εργοδικότητα

Η εργοδικότητα (ergodicity) είναι μια χαρακτηριστική ιδιότητα που συνδέει τους χωρικούς μέσους όρους και τους μέσους όρους συνόλου. Αυτή η ιδιότητα είναι εύχρηστη για την εκτίμηση τυχαίων πεδίων, επειδή από ένα μόνο δείγμα, οι χωρικοί μέσοι όροι μπορούν να αξιολογηθούν, ενώ οι μέσοι όροι συνόλου δεν μπορούν. Επίσης, η εργοδικότητα δηλώνει ότι οι ροπές μέχρι μια συγκεκριμένη τάξη μπορούν να εκτιμηθούν από ένα μόνο δείγμα. Για να προσδιοριστούν οι στατιστικές ιδιότητες σε όλα τα σημεία από ένα μόνο στιγμιότυπο του πεδίου, χρειάζεται να γίνει η υπόθεση ότι είναι αμετάβλητο ως προς τη μετάθεση επειδή χωρίς αυτή την υπόθεση είναι αδύνατο να συμβεί εκτίμηση.

Η έννοια της εργοδικότητας μας ενδιαφέρει διότι εφαρμόζεται σε τυχαία πεδία που χαρακτηρίζονται μόνο από τη στασιμότητά τους. Συνεπώς, ένα τυχαίο πεδίο θεωρείται εργοδικό όταν ισχύει ότι είναι και στάσιμο, ωστόσο το αντίστροφο δεν ισχύει δηλαδή ένα στάσιμο τυχαίο πεδίο δεν χρειάζεται να είναι και απαραίτητως εργοδικό. Θα ακο-

2.1. Βασικές Έννοιες

λουθήσει ανάλυση της εργοδικότητας στη μέση τιμή και στη συνάρτηση συνδιακύμανσης.

Ο δειγματικός μέσος όρος του τυχαίου πεδίου $Z(\mathbf{s})$ σε ένα χωρικό πεδίο \mathcal{D} δίνεται από την ακόλουθη σχέση του τυχαίου πεδίου.

$$Z(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{s}) d\mathbf{s}. \quad (2.15)$$

Ένα τυχαίο πεδίο το οποίο είναι στάσιμο, $Z(\mathbf{s})$ με σταθερή μέση τιμή m_z και με συνάρτηση συνδιακύμανσης $C_{zz}(\mathbf{r})$ που εξαρτάται μόνο από τη χωρική υστέρηση, είναι εργοδικό ως προς τη μέση τιμή εάν η Εξ. (2.15) τείνει ασυμπτωτικά προς τη μέση τιμή, άρα,

$$\lim_{\mathcal{D} \rightarrow \infty} Z(\mathcal{D}) = m_z. \quad (2.16)$$

Επομένως, με βάση την εργοδικότητα έπεται ότι ο χωρικός μέσος όρος $Z(\mathcal{D})$ δίνει μια ακριβή εκτίμηση της m_z καθώς το χωρικό πεδίο \mathcal{D} επεκτείνεται στο άπειρο σε κάθε ορθογώνια κατεύθυνση.

Στη συνέχεια, παρουσιάζεται το θεώρημα του Slutsky που αφορά την εργοδικότητα. Αυτό το θεώρημα δηλώνει ότι ένα στάσιμο τυχαίο πεδίο $Z(\mathbf{s})$ με σταθερή μέση τιμή m_z και συνάρτηση συνδιακύμανσης $C_{zz}(\mathbf{r})$, όπως αναφέραμε και προηγουμένως, είναι εργοδικό ως προς τη μέση τιμή, εφόσον ισχύει η εξής συνθήκη,

$$\lim_{V \rightarrow \infty} \frac{1}{|V|} \int_V C_{zz}(\mathbf{r}) d\mathbf{r} = 0, \quad (2.17)$$

όπου V ορίζεται ως ο διανυσματικός χώρος που δημιουργείται από τη διαφορά όλων των διανυσμάτων $\mathbf{s} \in \mathcal{D}$.

Μια συνθήκη η οποία κρίνεται ικανή αλλά όχι αναγκαία, για το θεώρημα του Slutsky είναι ότι για όλες τις κατευθύνσεις στο χώρο, η συνάρτηση συνδιακύμανσης τείνει στο μηδέν δεδομένου ότι το μέγεθος της χωρικής υστέρησης τείνει στο άπειρο.

Όσον αφορά για την εργοδικότητα στη συνάρτηση συνδιακύμανσης, σε περίπτωση που ένα τυχαίο πεδίο $Z(\mathbf{s})$ έχει μη μηδενική μέση τιμή

2.1. Βασικές Έννοιες

$m_z(\mathbf{s})$, επικεντρωνόμαστε στις διακυμάνσεις της μηδενικής μέσης τιμής. Για να εξετάσουμε την εργοδικότητα ενός στάσιμου τυχαίου πεδίου ως προς τη συνάρτηση συνδιακύμανσης, χρειάζεται να υλοποιήσουμε το ακόλουθο τυχαίο πεδίο,

$$Y(\mathcal{D}; \mathbf{r}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{s}) Z(\mathbf{s} + \mathbf{r}) d\mathbf{s}. \quad (2.18)$$

Η εργοδικότητα καθιστά αναγκαίο το όριο ενός άπειρου χωρικού πεδίου $Y(\mathcal{D}; \mathbf{r})$ να είναι μια ακριβής εκτίμηση της συνάρτησης συνδιακύμανσης, επομένως,

$$\hat{Y}(\mathbf{r}) = \lim_{\mathcal{D} \rightarrow \infty} Y(\mathcal{D}; \mathbf{r}) = C_{zz}(\mathbf{r}). \quad (2.19)$$

Είναι εύκολο να αποδειχθεί ότι η $\hat{Y}(\mathbf{r})$ είναι μια αντικειμενική εκτίμηση της συνάρτησης συνδιακύμανσης, εφόσον ισχύει ότι $\mathbb{E}[Y(\mathcal{D}; \mathbf{r})] = C_{zz}(\mathbf{r})$ για κάθε $\mathcal{D} \subset \mathbb{R}^d$. Αν και για να διαπιστωθεί ότι η διασπορά της $\hat{Y}(\mathbf{r})$ μηδενίζεται,

$$\lim_{\mathcal{D} \rightarrow \infty} \text{Var}\{Y(\mathcal{D}; \mathbf{r})\} = 0, \quad (2.20)$$

προϋποθέτει την εφαρμογή όρων στις ροπές τέταρτης τάξης του τυχαίου πεδίου $Z(\mathbf{s})$. Για τυχαία πεδία Gauss η κατάσταση είναι απλή, διότι οι ροπές τέταρτης τάξης, με τη χρήση του θεωρήματος Wick-Isserlis, μπορούν να αναλυθούν σε ροπές δεύτερης τάξης. Με βάση την ανάλυση αυτή, η συνθήκη για την εργοδικότητα στη συνάρτηση συνδιακύμανσης μετατρέπεται σε,

$$\lim_{\mathcal{D} \rightarrow \infty} \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} [C_{zz}(\mathbf{r})]^2 d\mathbf{r} = 0. \quad (2.21)$$

Η συνθήκη (2.21) ικανοποιείται αν η απόλυτη τιμή της συνάρτησης συνδιακύμανσης ελαττώνεται στο μηδέν ιδιαίτερα γρήγορα καθώς η χωρική υστέρηση $\|\mathbf{r}\|$ τείνει στο άπειρο.

2.2 Ανακατασκευή Κενών σε Χωρικά Δεδομένα

2.2.1 Γενικές Έννοιες

Η χωρική πρόβλεψη αναφέρεται στην εκτίμηση των τιμών που λείπουν σε μια χωρική διαδικασία. Παρεμβολή καλείται ένα πρόβλημα πρόβλεψης όταν τα σημεία της πρόβλεψης είναι εντός της περιοχής του δείγματος. Η διαδικασία χωρικής πρόβλεψης που πραγματοποιείται σε σημεία εκτός της περιοχής δειγματοληψίας ονομάζεται προέκταση (extrapolation).

Το πρόβλημα της παρεμβολής είναι η εκτίμηση των βέλτιστων τιμών του τυχαίου πεδίου σε έναν αριθμό σημείων πρόβλεψης $\mathbf{z}_{p=1}^P$ που συνήθως τοποθετούνται σε ένα κανονικό πλέγμα G . Το πεδίο που παρεμβάλλεται υπολογίζεται σε κάθε σημείο του πλέγματος ανεξάρτητα από άλλα σημεία του πλέγματος. Συνεπώς, μπορούμε να εστιάσουμε σε ένα μόνο σημείο για το πρόβλημα της παρεμβολής. Αυτό σημαίνει επίσης ότι οι υλοποιήσεις της παρεμβολής μπορούν να γίνονται ταυτόχρονα.

Η παρεμβολή διαθέτει μεθόδους και η παρακάτω θεωρείται η απλούστερη μέθοδος της, η οποία αποκαλείται μέθοδος παρεμβολής πλησιέστερου γείτονα (nearest neighbor interpolation). Η μέθοδος αυτή αναθέτει σε κάθε σημείο του πλέγματος παρεμβολής μια τιμή που είναι ίση με την τιμή του πλησιέστερου γείτονά του στο δείγμα. Οι τιμές παρεμβολής του πλησιέστερου γείτονα προέρχονται αποκλειστικά από το σύνολο του δείγματος. Η μέθοδος παρεμβολής φυσικών γειτόνων (natural neighbor interpolation) είναι μια επέκταση της προηγούμενης μεθόδου. Η μέθοδος αυτή βασίζεται στη διάταξη Voronoi ενός διακριτού συνόλου χωρικών σημείων. Εκφράζει την πρόβλεψη στο ζητούμενο σημείο ως άθροισμα των τιμών του δείγματος στα φυσικά γειτονικά σημεία, δηλαδή, τα κέντρα των γειτονικών πολυγώνων Voronoi.

Η μέθοδος γραμμικής παρεμβολής (linear interpolation) χρησιμοποιείται για την πρόβλεψη ενός τυχαίου πεδίου $Z(\mathbf{s})$ σε κάποιο μη μετρημένο σημείο \mathbf{s}_0 . Πιο συγκεκριμένα, εφαρμόζεται γραμμική επικάλυψη στα δεδομένα. Δηλαδή, η πρόβλεψη $\hat{z}(\mathbf{s}_0)$ στο σημείο \mathbf{s}_0 δίνεται παρακάτω:

$$\hat{z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i z(\mathbf{s}_i), \quad (2.22)$$

όπου,

- $\hat{z}(\mathbf{s}_0)$, η εκτιμώμενη τιμή της μεταβλητής στο σημείο \mathbf{s}_0 .
- $z(\mathbf{s}_i)$, η πραγματική τιμή της μεταβλητής στο σημείο \mathbf{s}_i .
- n , το σύνολο των σημείων του δείγματος.
- λ_i , ο συντελεστής βαρύτητας που έχει καθοριστεί με βάση το σημείο της δειγματοληψίας.

Η μέθοδος αυτή σε σύγκριση με την μέθοδο παρεμβολής του πλησιέστερου γείτονα (nearest neighbor interpolation), δίνει μια καλύτερη προσέγγιση. Παρ'όλα αυτά δεν θεωρείται ιδιαίτερα ακριβής.

2.2.2 Χωρική Παρεμβολή

Η χωρική παρεμβολή (spatial interpolation) αποτελεί μια διαδικασία εκτίμησης της τιμής ενός φαινομένου σε σημεία που δεν ανήκουν στο δείγμα, με βάση τις τιμές που είναι γνωστές από τις μετρήσεις στα σημεία του δείγματος. Τα σημεία του δείγματος αναφέρονται ως σημεία στο χώρο και στο χρόνο. Η πιο διαδεδομένη εφαρμογή της παρεμβολής είναι η προσομοίωση πεδίου δύο διαστάσεων, μπορεί όμως να προσδιοριστεί και σε μία διάσταση.

Η χωρική παρεμβολή στηρίζεται στον Πρώτο Νόμο της Γεωγραφίας, ο οποίος διατυπώθηκε από τον Waldo Tobler (1970) και ο οποίος αναφέρει ότι “όλα σχετίζονται με όλα τα άλλα, αλλά τα πράγματα που είναι πιο κοντά σχετίζονται περισσότερο από αυτά που είναι απομακρυσμένα”.

Σκοπός της χωρικής παρεμβολής είναι ο προσδιορισμός του προβλήματος και η εύρεσης μιας μαθηματικής συνάρτησης, η οποία θα ελέγχει τα σημεία στα οποία έχουμε δεδομένα και την εκτίμηση των τιμών σε εκείνα τα σημεία στα οποία τα αρχικά δεδομένα ήταν ελλιπή. Η ακρίβεια

2.2. Ανακατασκευή Κενών σε Χωρικά Δεδομένα

σε σημεία που έχουν γίνει μετρήσεις είναι μεγαλύτερη σε σχέση με τα σημεία όπου υπάρχει έλλειψη στοιχείων. Αυτό συμβαίνει επειδή αρκετά στοιχεία διαθέτουν συγκεκριμένα επίπεδα θορύβου που για να επιτευχθούν καλύτερα αποτελέσματα πρέπει να ελαχιστοποιηθούν. Επομένως, η αξιοπιστία μιας μεθόδου παρεμβολής θα πρέπει να καθορίζεται από τις παρακάτω ιδιότητες:

- Ικανοποιητική αξιοπιστία των υπό εξέταση φαινομένων.
- Συνδυασμός πολλαπλών διαστάσεων.
- Δυνατότητα υλοποίησης σε δεδομένα από π.χ τηλεανίχνευση, γεωστατιστικές μεθόδους κλπ.
- Δυνατότητα εφαρμογής σε μεγάλου μεγέθους βάσεις δεδομένων.
- Υπολογιστική ικανότητα.

Σύμφωνα με τις παραπάνω ιδιότητες, γίνεται αντιληπτό ότι η εύρεση μιας ικανοποιητικής μεθόδου που να πληρεί όλες τις προαναφερθείσες απαιτήσεις είναι δύσκολο. Έτσι, λοιπόν, για να είναι επιτυχημένη η επιλογή της μεθόδου για μία συγκεκριμένη εφαρμογή είναι ιδιαίτερα σημαντικό να είναι επαρκής η γνώση και να γίνει σύγκριση μεταξύ των διαθέσιμων μεθόδων.

2.2.3 Μέτρα Διασταυρωτικής Επιβεβαίωσης

Μέσο Σφάλμα (Mean Error ή αλλιώς ME) υπολογίζεται ως ο μέσος όρος των σφαλμάτων. Αν ο δείκτης που αντιπροσωπεύει το μέσο σφάλμα είναι θετικός τότε το μοντέλο πρόβλεψης δίνει αρνητικές προβλέψεις διότι η πραγματική τιμή είναι μικρότερη της προβλεπόμενης τιμής. Αν όμως ο δείκτης είναι αρνητικός τότε το μοντέλο πρόβλεψης δίνει θετικές προβλέψεις αφού κατά μέσο όρο οι προβλεπόμενες τιμές είναι μεγαλύτερες από τις πραγματικές τιμές.

$$ME = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i), \quad (2.23)$$

2.2. Ανακατασκευή Κενών σε Χωρικά Δεδομένα

όπου το x_i είναι οι πραγματικές τιμές και το \hat{x}_i είναι οι εκτιμώμενες τιμές.

Ρίζα Τετραγωνικού Σφάλματος (Root Mean Square Error ή αλλιώς RMSE) είναι ένας τετραγωνικός κανόνας βαθμολόγησης που υπολογίζει το μέσο εύρος του σφάλματος. Δηλαδή, η διαφορά μεταξύ των προβλέψεων και των αντίστοιχων παρατηρούμενων τιμών τετραγωνίζονται κι έπειτα βρίσκεται ο μέσος όρος του δείγματος. Λαμβάνοντας υπόψη ότι τα σφάλματα τετραγωνίστηκαν για να βρεθεί ο μέσος όρος τους, το RMSE δίνει ένα σχετικά σημαντικό βάρος σε μεγάλα σφάλματα. Αυτό δηλώνει ότι το RMSE είναι πιο χρήσιμο όταν μεγάλα σφάλματα είναι ανεπιθύμητα.

$$RMSE = \sqrt{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}, \quad (2.24)$$

όπου το x_i είναι οι πραγματικές τιμές και το \hat{x}_i είναι οι εκτιμώμενες τιμές.

Μέσο Απόλυτο Σφάλμα (Mean Absolute Error ή αλλιώς MAE) υπολογίζει το μέσο εύρος των σφαλμάτων σε μια σειρά από προβλέψεις, χωρίς να λαμβάνεται υπόψη η διεύθυνση τους. Μετρά την ακρίβεια για τις συνεχείς μεταβλητές. Δηλαδή, το MAE είναι ο μέσος όρος του δείγματος επαλήθευσης των απόλυτων τιμών, των διαφορών μεταξύ των προβλέψεων και της αντίστοιχης παρατήρησης. Είναι γραμμικό αποτέλεσμα, το οποίο σημαίνει ότι όλες οι ατομικές διαφορές έχουν την ίδια βαρύτητα στη μέση τιμή.

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i|, \quad (2.25)$$

όπου το x_i είναι οι πραγματικές τιμές και το \hat{x}_i είναι οι εκτιμώμενες τιμές.

2.2. Ανακατασκευή Κενών σε Χωρικά Δεδομένα

Το μέσο τετραγωνικό σφάλμα υπολογίζει ουσιαστικά τη μέση απόλυτη απόκλιση των προβλεπόμενων τιμών από τις πραγματικές για όλο το σύνολο της πρόβλεψης. Με βάση την απόλυτη τιμή του οι θετικές τιμές δεν αναιρούνται από τις αρνητικές.

Συνεπώς, το MAE και το RMSE μπορούν να χρησιμοποιηθούν μαζί για να ελεγχθεί η διακύμανση σφαλμάτων σε μια σειρά προβλέψεων. Το MAE θα είναι πάντα μικρότερο ή ίσο με το RMSE και παρατηρείται ότι όσο μεγαλύτερη η διαφορά μεταξύ τους, τόσο μεγαλύτερη η διακύμανση σφαλμάτων στο δείγμα. Τα σφάλματα είναι του ίδιου εύρους εάν το $MAE = RMSE$.

Συντελεστής Συσχέτισης Pearson (Pearson Correlation Coefficient ή αλλιώς ρ) μετρά τη στατιστική αλληλεπίδραση μεταξύ των πραγματικών και των εκτιμώμενων τιμών του δείγματος. Το ρ του Pearson είναι ευαίσθητο στις γραμμικές συσχετίσεις μεταξύ των πραγματικών και των εκτιμώμενων τιμών του δείγματος, αλλά δεν είναι αξιόπιστο στην περίπτωση μη γραμμικών σχέσεων.

$$\rho = \frac{\sum_{i=1}^N (x_i - \bar{x}_i)(\hat{x}_i - \bar{\hat{x}}_i)}{\sqrt{\sum_{i=1}^N (x_i - \bar{x}_i)^2} \sqrt{(\hat{x}_i - \bar{\hat{x}}_i)^2}}, \quad (2.26)$$

όπου το x_i είναι οι πραγματικές τιμές και το \hat{x}_i είναι οι εκτιμώμενες τιμές. Για την Εξ. (2.26) ισχύει $-1 \leq \rho \leq 1$.

2.2. Ανακατασκευή Κενών σε Χωρικά Δεδομένα

Κεφάλαιο 3

Διερευνητική Ανάλυση Δεδομένων

3.1 Δορυφορικά Δεδομένα

Για τη υλοποίηση της παρούσας διπλωματικής εργασίας χρησιμοποιήθηκαν δορυφορικά δεδομένα, δηλαδή δύο σύνολα δεδομένων εκπαίδευσης (ένα προσομοιωμένο και ένα πραγματικό) μαζί με ένα σύνολο θέσεων πρόβλεψης, τα οποία απεικονίζουν μία συγκεκριμένη περιοχή και παρουσιάζουν ποικιλία ως προς το ποσοστό εμφάνισης κενών. Όλα τα σύνολα δεδομένων χρησιμοποιήθηκαν από το την εργασία [5]. Τα δεδομένα που αποτέλεσαν το πραγματικό σύνολο δεδομένων προέρχονταν από θερμοκρασίες ($^{\circ}\text{C}$) της επιφάνειας της Γης κατά τη διάρκεια της ημέρας, όπως μετρήθηκαν από το όργανο Terra του δορυφόρου MODIS στις 4 Αυγούστου 2016 (δεδομένα Επιπέδου 3).

Τα δεδομένα αυτά ανακτήθηκαν αρχικά από τη διαδικτυακή διεπαφή του εργαλείου αναπαράστασης (MRTweb) για τα δορυφορικά δεδομένα MODIS. Λόγω διακοπής στην λειτουργία αυτού του εργαλείου, τα δεδομένα λήφθηκαν από το αποθετήριο [github](#). Και τα δύο σύνολα δεδομένων αποτελούνται από παρατηρήσεις στο ίδιο πλέγμα 500 επί 300 με τιμές γεωγραφικού μήκους που κυμαίνονται από -95.91153 έως -91.28381 και τιμές γεωγραφικού πλάτους από 34.29519 έως 37.06811.

3.1. Δορυφορικά Δεδομένα

Επίσης να σημειωθεί ότι το γεωγραφικό πλάτος και μήκος καθώς και η ημερομηνία, επιλέχθηκαν λόγω της αραιής νεφοκάλυψης της περιοχής στη συγκεκριμένη ημερομηνία.

Το δύο σύνολα δεδομένων χωρίζονται σε προσομοιωμένο και σε πραγματικό με βάση τη συνάρτηση συνδιακύμανσης. Δηλαδή, το προσομοιωμένο σύνολο δεδομένων αντιπροσώπευε μια περίπτωση όπου η συνάρτηση συνδιακύμανσης έχει καθοριστεί σωστά. Αντίθετα, το πραγματικό σύνολο δεδομένων αντιπροσώπευε μια περίπτωση όπου έχει καθοριστεί λανθασμένα η συνάρτηση συνδιακύμανσης.

Το σύνολο των προσομοιωμένων δεδομένων υλοποιήθηκε με την προσρμογή ενός χωρικού μοντέλου Gauss με σταθερή μέση τιμή, εκθετική συνάρτηση συνδιακύμανσης, και τυχαίο χωρικό θόρυβο (φαινόμενο κόκκου - nugget effect) σε ένα τυχαίο δείγμα 2 500 παρατηρήσεων από τα προαναφερόμενα δεδομένα MODIS. Οι τιμές των παραμέτρων που προέκυψαν ήταν 1.33 για τη χωρική ακτίνα της συνάρτησης συνδιακύμανσης, 16.4 για τη χωρική διασπορά, 0.05 για τη διασπορά του θορύβου (φαινόμενο κόκκου - nugget effect) και 44.49 για τη μέση τιμή. Για την υλοποίηση 150 000 παρατηρήσεων στο ίδιο πλέγμα με τα δεδομένα MODIS κρίθηκε απαραίτητο η χρήση των παραμέτρων αυτών. Όσον αφορά την παράμετρο του χωρικού εύρους, η οποία έχει τιμή ίση με 1.33, ισοδυναμεί με μια κατά προσέγγιση αποτελεσματικού χωρικού εύρους των περίπου 210 μιλίων (338 χιλιομέτρων).

Για τη δημιουργία των συνόλων δοκιμής και εκπαίδευσης αναφέρεται στην εργασία ότι χρησιμοποιήθηκε το σύνολο δεδομένων που έλειπε στις 6 Αυγούστου 2016 από το ίδιο προϊόν δορυφορικών δεδομένων MODIS. Αφού έγινε αυτός ο διαχωρισμός, το σύνολο εκπαίδευσης για τα δεδομένα MODIS αποτελούνταν από 105 569 παρατηρήσεις και οι 42 740 παρατηρήσεις που έμειναν προορίστηκαν για το σύνολο δοκιμής. Για τα προσομοιωμένα δεδομένα, το σύνολο εκπαίδευσης αποτελούνταν κι αυτά από 105 569 παρατηρήσεις αλλά το σύνολο δοκιμής σε αυτή την περίπτωση ήταν 44 431. Αυτή η διαφορά στο μέγεθος του συνόλου δοκιμής συμβαίνει εξαιτίας της νεφοκάλυψης στα αρχικά δεδομένα και αυτό έχει ως αποτέλεσμα να έχουν αλλοιωθεί κάποια δεδομένα.

3.2 Συνθετικά Δεδομένα με Χρήση της Συνάρτησης Ackley

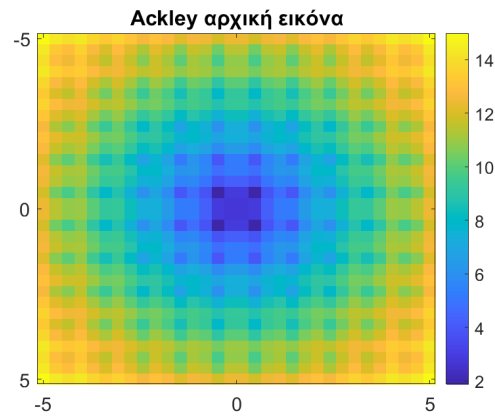
Εκτός από τα δορυφορικά δεδομένα, σε αυτή τη διπλωματική εργασία ασχοληθήκαμε και με τα συνθετικά δεδομένα. Τα δεδομένα αυτά δημιουργήθηκαν με χρήση της συνάρτησης Ackley. Η μη κυρτή δισδιάστατη συνάρτηση Ackley εφαρμόζεται στη δοκιμή αλγορίθμων βελτιστοποίησης. Η συνάρτηση αυτή βασίζεται στην προσθήκη της συνάρτησης συνημιτόνου της εκθετικής συνάρτησης η οποία χαρακτηρίζεται από μια σχεδόν επίπεδη περιοχή με διαμόρφωση του συνημιτόνου κύματος που δημιουργεί μια τρύπα ή μια κορυφή, προκαλώντας τον κυματισμό της επιφάνειας. Για τη συνάρτηση Ackley χρησιμοποιείται η εξίσωση:

$$f(x) = -20 \times \exp \left(-0.2 \sqrt{1/n \sum_{i=1}^n x_i^2} \right) - \exp \left(1/n \sum_{i=1}^n (\cos 2\pi x_i) \right) + 20 + \exp, \quad (3.1)$$

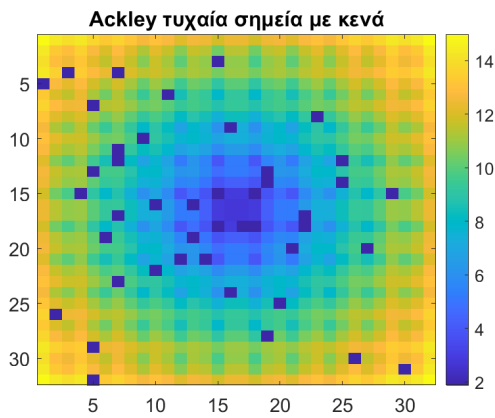
όπου $x_i \in [-32.768, 32.768]$ και το D αντιπροσωπεύει τη διάσταση της ανεξάρτητης μεταβλητής x . Στην Εξ. (3.1) θέτουμε το $D = 2$. Η συνάρτηση Ackley είναι πολυδιάστατη και είναι δυνατόν να ληφθεί η βέλτιστη τιμή 0 όταν το $(x_1, \dots, x_D) = (0, \dots, 0)$ [2, 11].

Με τη βοήθεια της συνάρτησης Ackley δημιουργήθηκαν τα συνθετικά δεδομένα και παρουσιάζονται στο παρακάτω Σχήμα 3.1. Στη συνέχεια τοποθετήθηκαν ενδεικτικά 40 τυχαία σημεία, τα οποία εμφανίζονται με μπλε χρώμα Σχήμα 3.1ii και με άσπρο χρώμα Σχήμα 3.1iii. Το άσπρο χρώμα χρησιμοποιήθηκε για μεγαλύτερη διευκόλυνση. Επίσης, να επισημανθεί ότι ο αριθμός των κόμβων ανά πλευρά του τετραγώνου είναι 32.

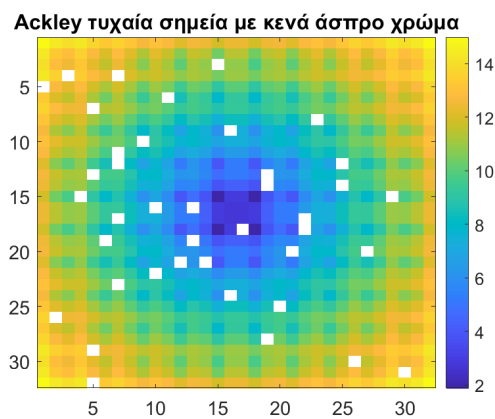
3.2. Συνθετικά Δεδομένα με Χρήση της Συνάρτησης Ackley



(i)



(ii)



(iii)

Σχήμα 3.1: (i): Συνθετικά δεδομένα από συνάρτηση Ackley. Τυχαία κατανομή 40 κενών εικονοστοιχείων (ii): μπλε και (iii): άσπρο χρώμα.

3.3. Διερευνητική Στατιστική Ανάλυση

Όπως θα δούμε στο Κεφάλαιο 5, τα συνθετικά δεδομένα διακρίνονται σε διάφορες περιπτώσεις όσον αφορά τα κενά. Πιο συγκεκριμένα, αναλύονται στις εξής διατάξεις με κενά: 1) τυχαία κατανομή 40 κενών εικονοστοιχείων, 2) συνεχής περιοχή 303 κενών εικονοστοιχείων και 3) τυχαία κατανομή 303 κενών εικονοστοιχείων. Αυτές οι διατάξεις πραγματοποιήθηκαν για να παρατηρήσουμε πως οι μεθοδολογίες συμπεριφέρονται και πόσο αποδίδουν ώστε να επιτευχθεί η πλήρωση κενών. Τα κενά στην πρώτη περίπτωση είναι 40 για να αξιολογήσουμε τις μεθοδολογίες με μικρό αριθμό κενών. Όσον αφορά τον αριθμό των κενών για τις άλλες δύο περιπτώσεις, υπολογίστηκε χρησιμοποιώντας το ποσοστό επί των συνολικών σημείων του πλέγματος που υπάρχουν κενά στην εικόνα με τα προσομοιωμένα δεδομένα (δορυφορικά δεδομένα). Αυτό εφαρμόστηκε για να συγκρίνουμε την απόδοση των μεθοδολογιών με το ίδιο ποσοστό ύπαρξης κενών τόσο στα συνθετικά δεδομένα όσο και στα προσομοιωμένα δεδομένα. Το ποσοστό υπολογίστηκε ως εξής,

$$\frac{44\,431}{150\,000} \times 100\% = 29,62\%. \quad (3.2)$$

Στη συνέχεια, ο αριθμός των κενών για τα συνθετικά δεδομένα προέκυψε ως εξής,

$$\frac{29,62\%}{100\%} \times 1024 \approx 303. \quad (3.3)$$

3.3 Διερευνητική Στατιστική Ανάλυση

Πριν την εφαρμογή των μεθοδολογιών που θα παρουσιαστούν στο Κεφάλαιο 4 έπρεπε να γίνει επεξεργασία του αρχικού συνόλου δεδομένων. Οι συγκεκριμένες επεξεργασίες περιγράφονται παρακάτω.

Αρχικά τα δεδομένα που αποκτήθηκαν από την εργασία ήταν σε περιβάλλον R και έγινε η μεταφορά των δεδομένων σε περιβάλλον matlab με σκοπό να γίνει η επεξεργασία αυτών και στη συνέχεια η υλοποίηση και η εφαρμογή των μεθόδων. Για να εμφανιστούν οι εικόνες πρώτα απ' όλα ανακατασκευάστηκαν οι εικόνες με την εντολή reshape. Η εντολή

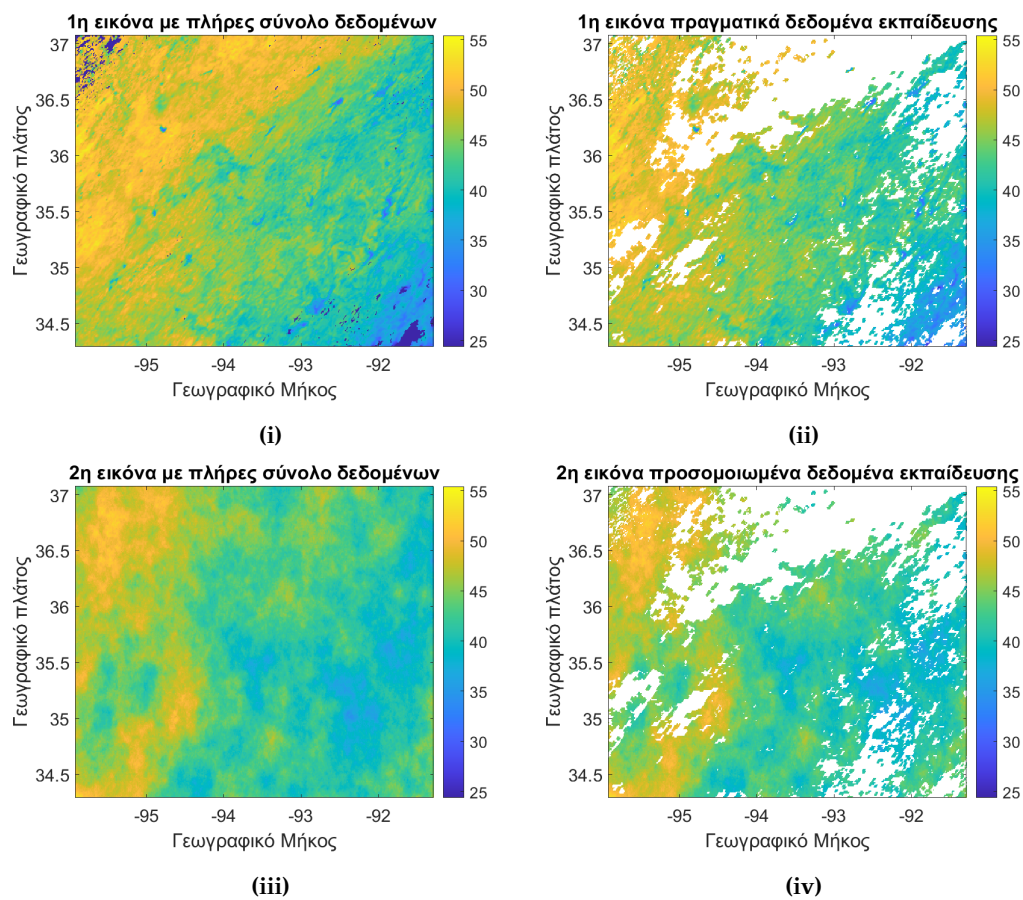
3.3. Διερευνητική Στατιστική Ανάλυση

`reshape` ουσιαστικά ανακατασκευάζει πίνακα και τον επιστρέφει με μέγεθος 500 επί 300 του οποίου τα στοιχεία λαμβάνονται κατά στήλες από τον πίνακα. Να σημειωθεί ότι ο πίνακας που αναφέρεται είναι διάνυσμα στήλης μεγέθους 150 000 επί 1. Με τη χρήση της εντολής αυτής γίνεται η ανακατασκευή και το τελικό αποτέλεσμα είναι ένας πίνακας με 300 γραμμές και 500 στήλες.

Το επόμενο βήμα ήταν η εμφάνιση των εικόνων έτσι όπως εμφανίζονται στην εργασία. Η εντολή `imagesc` χρησιμοποιήθηκε, η οποία είναι παρόμοια με την εντολή `image` με τη διαφορά ότι τα δεδομένα κλιμακώνονται για να αξιοποιηθεί ο πλήρης χάρτης χρωμάτων. Πριν γίνει αυτό θεωρήθηκε σημαντικό να καθοριστεί κοινός χρηματικός κώδικας για όλες τις εικόνες. Επομένως, για κάθε εικόνα με επαναληπτικό τρόπο βρέθηκε με την εντολή `min(min(x))`, όπου x είναι ο πίνακας, η ελάχιστη τιμή και αντίστοιχα η μέγιστη τιμή με την εντολή `max(max(x))`. Αφού βρέθηκε για κάθε μία από τις τέσσερις εικόνες η ελάχιστη και η μέγιστη τιμή στη συνέχεια, διαπιστώθηκε ανάμεσα στις τέσσερις τιμές ποιο είναι το γενικό ελάχιστο. Η ίδια διαδικασία έγινε και για το γενικό μέγιστο. Έτσι, ολοκληρώθηκε η διαδικασία εμφάνισης των εικόνων με κοινό χρωματικό κώδικα με τιμές από 25 μέχρι 55.

Στο Σχήμα 3.2i και στο Σχήμα 3.2iii, όπως φαίνεται παρακάτω, παρουσιάζονται τα πλήρη σύνολα δεδομένων. Στο Σχήμα 3.2ii απεικονίζεται το πραγματικό σύνολο εκπαίδευσης και στο Σχήμα 3.2iv το προσομοιωμένο σύνολο εκπαίδευσης. Ο οριζόντιος άξονας αντιπροσωπεύει το γεωγραφικό μήκος και ο κατακόρυφος άξονας το γεωγραφικό πλάτος.

3.3. Διερευνητική Στατιστική Ανάλυση



Σχήμα 3.2: Πρώτη εικόνα (i): πλήρες σύνολο δεδομένων, (ii): πραγματικό σύνολο δεδομένων εκπαίδευσης. Δεύτερη εικόνα (iii): πλήρες σύνολο δεδομένων, (iv): προσομοιωμένο σύνολο δεδομένων εκπαίδευσης.

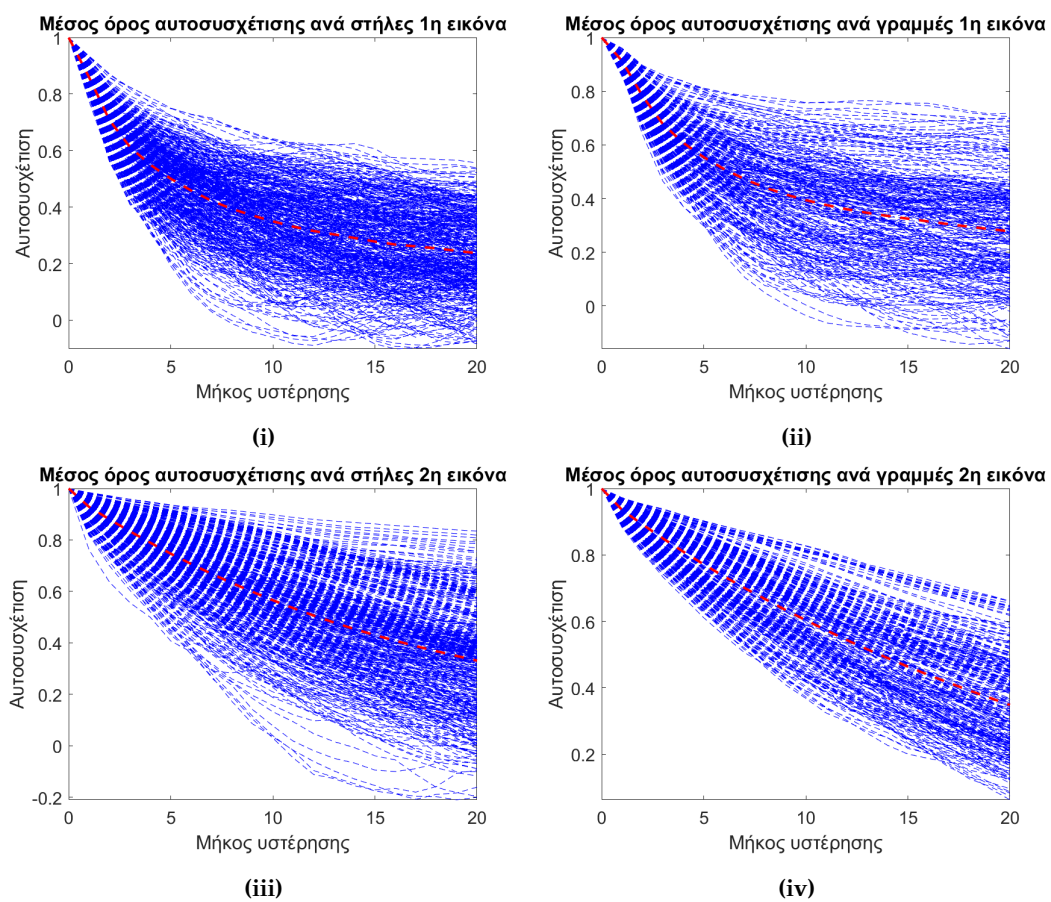
3.3. Διερευνητική Στατιστική Ανάλυση

Στη συνέχεια, υπολογίζεται η συνάρτηση αυτοσυσχέτισης πάνω στην πρώτη εικόνα (πραγματικά δεδομένα) με τα πλήρη δεδομένα, Σχήμα 3.2i. Αρχικά, για τον υπολογισμό της συνάρτησης αυτοσυσχέτισης, κάθε σειρά στην εικόνα θεωρήθηκε ως μια χρονοσειρά, το ίδιο και για κάθε στήλη. Από κάθε σειρά λοιπόν αφαιρέθηκε η αντίστοιχη μέση τιμή, η οποία υπολογίστηκε με την εντολή `mean`. Έπειτα, χρησιμοποιήθηκε η εντολή `autocorr` για να υπολογιστεί η συνάρτηση αυτοσυσχέτισης. Τα αποτελέσματα που προέκυψαν ήταν ένας πίνακας που είχε τις αυτοσυσχετίσεις ανά σειρά, ένα διάνυσμα αριθμών που αντιπροσωπεύει το μήκος υστέρησης (`lags`) και τέλος ένα διάνυσμα δύο στοιχείων που αναπαριστά προσεγγιστικά το άνω και το κάτω όριο εμπιστοσύνης (`bounds`). Στη συνέχεια, παρουσιάστηκαν σε διάγραμμα οι αυτοσυσχετίσεις για καθεμία από τις γραμμές καθώς και για τις στήλες.

Ακολούθως, υπολογίστηκε ο μέσος όρος ως προς τις αυτοσυσχετίσεις για τις γραμμές και αντίστοιχα για τις στήλες. Είναι απαραίτητο να διευκρινιστεί ότι έγινε εφαρμογή της εντολής `fillmissing` για να συμπληρωθούν κάποιες μη αριθμητικές τιμές (`NaN`) που υπήρχαν στα δεδομένα. Αυτό είχε αναφερθεί στην εργασία ότι λόγω νεφοκάλυψης αλλοιώθηκαν κάποιες τιμές και συγκεκριμένα μόνο το 1.1%. Τέλος, όλα αυτά τα βήματα για τον υπολογισμό της συνάρτησης αυτοσυσχέτισης εφαρμόστηκαν και στη δεύτερη εικόνα (προσομοιωμένα δεδομένα) με το πλήρες σύνολο δεδομένων.

Στο Σχήμα 3.3i και στο Σχήμα 3.3ii παρουσιάζονται οι αυτοσυσχετίσεις της πρώτης εικόνας ανά στήλες και ανά γραμμές αντίστοιχα. Πρώτη εικόνα θεωρούμε την εικόνα με τα πραγματικά δεδομένα. Στο Σχήμα 3.3iii και στο Σχήμα 3.3iv απεικονίζονται οι αυτοσυσχετίσεις της δεύτερης εικόνας ανά στήλες και ανά γραμμές αντίστοιχα. Δεύτερη εικόνα θεωρούμε την εικόνα με τα προσομοιωμένα δεδομένα. Επίσης, απεικονίζεται ο μέσος όρος της αυτοσυσχέτισης. Η διακεκομμένη κόκκινη γραμμή αναπαριστά τον μέσο όρο της αυτοσυσχέτισης.

3.3. Διερευνητική Στατιστική Ανάλυση



Σχήμα 3.3: Αποτελέσματα συνάρτησης αυτοσυσχέτισης. Διακεκομμένη κόκκινη γραμμή αντιπροσωπεύει το μέσο όρο αυτοσυσχέτισης. Πρώτη εικόνα ανά (i): στήλες, (ii): γραμμές. Δεύτερη εικόνα ανά (iii): στήλες, (iv): γραμμές.

3.3. Διερευνητική Στατιστική Ανάλυση

Για να ολοκληρωθεί η διερευνητική στατιστική ανάλυση, εφαρμόσαμε τα βαριόγραμματα. Στο Σχήμα 3.4 παρατηρούμε ότι τα βαριόγραμματα ξεκινάνε από το μηδέν οπότε συμπεραίνουμε ότι δεν υπάρχουν ασυνέχειες. Επειδή το βαριόγραμμα εστιάζει στο τετράγωνο των διαφορών, όταν η απόσταση είναι μηδενική, δεν υπάρχουν διαφορές άρα γιαυτό ξεκινάει από το μηδέν. Όσο αυξάνεται η απόσταση αυτό σημαίνει ότι τα σημεία απομακρύνονται το ένα από το άλλο. Στο Σχήμα 3.4ii παρατηρούμε ότι όσο μεγαλώνει η απόσταση, το βαριόγραμμα δεν τείνει να πάει σε μια σταθερή τιμή που την ονομάζουμε οροφή. Λόγω αυτού του γεγονότος συμπεραίνουμε ότι δεν έχει στασιμότητα. Αυτό μπορούμε να το επιβεβαιώσουμε, παρατηρώντας την αρχική εικόνα (πραγματικά δεδομένα). Το κίτρινο χρώμα το οποίο αντιπροσωπεύει τις υψηλές τιμές διακρίνουμε ότι καλύπτει περίπου το μισό της εικόνας και αντίστοιχα το μπλε/πράσινο χρώμα που είναι για τις χαμηλές τιμές καλύπτει το άλλο μισό της εικόνας. Το γεγονός αυτό μας υποδηλώνει ότι η διαδικασία αυτή δεν μπορεί να χαρακτηριστεί στάσιμη. Επομένως, κι αυτός είναι ο λόγος της μη στασιμότητας που καθρεφτίζεται στο βαριόγραμμα.

Στο Σχήμα 3.4i κάποιες στήλες φαίνεται ότι έχουν μια σταθερή τιμή η οποία είναι περίπου κάπου στο 10. Στις περισσότερες όμως στήλες φαίνεται να υπάρχει μεγάλη μεταβλητότητα και αυτό μπορούμε να το δούμε και στην αρχική εικόνα (πραγματικά δεδομένα). Όπου κατά μήκος της καθέτου παρατηρούμε ότι σε κάποια σημεία πηγαίνουμε από τα κίτρινα στα μπλε πράσινα περίπου και αυτό μας επιβεβαιώνει ότι έχουμε μια μεγάλη μεταβλητότητα. Για να ήταν στάσιμη θα έπρεπε να είχαμε πολλές επαναλήψεις των ίδιων στατιστικά μορφών δηλαδή των κορυφών και των κοιλάδων.

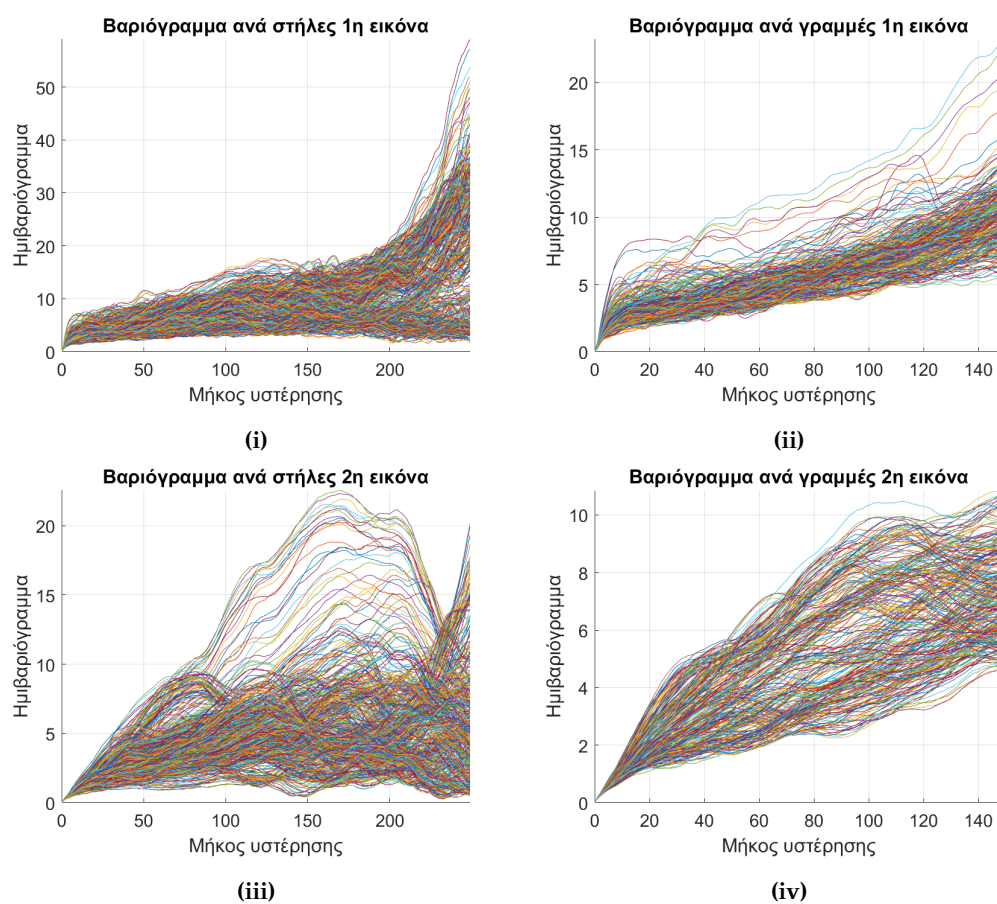
Στη συνέχεια, παρουσιάζονται τα βαριόγραμματα ανά στήλες και ανά γραμμές για τα προσομοιωμένα δεδομένα. Στο Σχήμα 3.4iv, κατά μήκος των γραμμών έχουμε μία μεταβλητότητα που δεν φαίνεται να φτάνει σε μια οροφή το βαριόγραμμα αλλά φαίνεται να συνεχίζει να αυξάνει αυτό. Παρατηρώντας την αρχική εικόνα (προσομοιωμένα δεδομένα) αν πάμε κατά μήκος μιας γραμμής πηγαίνουμε από τα κίτρινα στα μπλε άρα φαίνεται ότι έχουμε μια μεγαλύτερη μεταβλητότητα σε σχέση με αν

3.3. Διερευνητική Στατιστική Ανάλυση

πάμε κατά μήκος μιας στήλης. Αυτό επιβεβαιώνεται και στο βαριόγραμμα ανά γραμμές όπου το βαριόγραμμα αυξάνεται και αυτό συμβαίνει επειδή έχουμε αυτές τις εναλλαγές από υψηλές σε χαμηλές τιμές. Όσον αφορά το βαριόγραμμα ανά στήλες, δηλαδή στο Σχήμα 3.4iii αν εξαιρέσουμε κάποιες λίγες στήλες οι οποίες έχουν πολύ μεγάλη μεταβλητότητα, στις περισσότερες φαίνεται ότι πιάνει μια σταθερή τιμή η οποία είναι της τάξεως μεταξύ 5 και 10 περίπου. Κάπου στο 5 λοιπόν είναι η τιμή οροφής για πολλά από αυτά τα βαριογράμματα. Άρα αυτό δείχνει ότι έχουμε μια πιο στάσιμη κατάσταση κατά μήκος των στηλών. Αυτό επιβεβαιώνεται και στην αρχική εικόνα (προσομοιωμένα δεδομένα) όπου παρατηρούμε ότι κατά μήκος της καθέτου δεν έχουμε μεγάλη μεταβολή διότι αν ξεκινήσουμε από αριστερά που είναι κίτρινο, θα παραμείνουμε πάλι στο κίτρινο που αντιπροσωπεύει τις υψηλές τιμές. Αντίστοιχα αν ξεκινήσουμε από τα δεξιά που είναι μπλε/πράσινο πάλι στο ίδιο χρώμα θα παραμείνουμε.

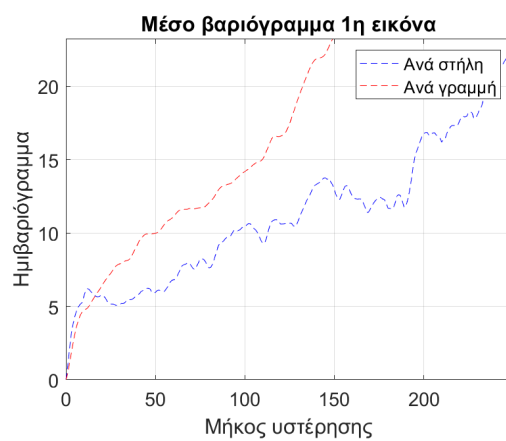
Στο Σχήμα 3.5 απεικονίζεται το μέσο βαριόγραμμα για τα πραγματικά και τα προσομοιωμένα δεδομένα.

3.3. Διερευνητική Στατιστική Ανάλυση

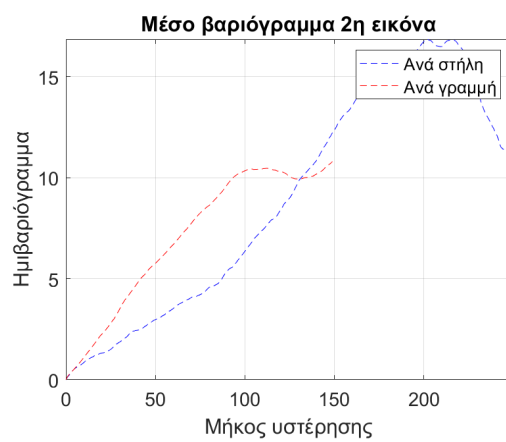


Σχήμα 3.4: Βαριογράμματα. Πρώτη εικόνα ανά (i): στήλες, (ii): γραμμές. Δεύτερη εικόνα ανά (iii): στήλες, (iv): γραμμές.

3.3. Διερευνητική Στατιστική Ανάλυση



(i)



(ii)

Σχήμα 3.5: Μέσο βαριόγραμμα (i): Πρώτη εικόνα. (ii): Δεύτερη εικόνα.

3.3. Διερευνητική Στατιστική Ανάλυση

Κεφάλαιο 4

Μεθοδολογίες Χωρικής Παρεμβολής

4.1 Αιτιοκρατικές Μέθοδοι

Σε αυτό το κεφάλαιο παρουσιάζονται οι μεθοδολογίες που χρησιμοποιήθηκαν για την πλήρωση κενών στις δορυφορικές εικόνες. Οι αιτιοκρατικές μεθοδολογίες υλοποιήθηκαν με τις έτοιμες συναρτήσεις της matlab, οι οποίες ονομάζονται `griddata` και `scatteredInterpolant`. Ως αιτιοκρατικές ορίζονται οι μέθοδοι παρεμβολής των οποίων οι εκτιμήσεις στηρίζονται αποκλειστικά σε συναρτήσεις της απόστασης και οδηγούν στην εκτίμηση μιας τιμής της μελετούμενης μεταβλητής για κάθε σημείο του χώρου.

Αρχικά, η συνάρτηση

```
vq=griddata(x,y,v,xq,yq)
```

διαμορφώνει μια επιφάνεια της μορφής $v = f(x, y)$ σε διάσπαρτα δεδομένα πάνω στα διανύσματα (x, y, v) . Αυτή η συνάρτηση παρεμβάλλει διάσπαρτα δεδομένα σε προκαθορισμένες θέσεις σημείων πλέγματος έτσι ώστε να παραχθούν δεδομένα πλέγματος, όπως δηλώνει και το όνομα της. Προκειμένου άρα, να συμβεί η παρεμβολή στις μετρήσεις που πα-
ρέχονται από τυχαία τοποθετημένους αισθητήρες, χρησιμοποιείται μια

4.1. Αιτιοκρατικές Μέθοδοι

ενσωματωμένη συνάρτηση της matlab που εκτελεί δισδιάστατη παρεμβολή από διάσπαρτα δεδομένα. Πιο αναλυτικά, η συνάρτηση `griddata` προσφέρει μια επιφάνεια παρεμβολής με βάση γνωστές μετρήσεις v που παρέχονται από αισθητήρες που έχουν αναπτυχθεί στα σημεία που καθορίζονται από τις συντεταγμένες x και y . Να σημειωθεί ότι η επιφάνεια διέρχεται πάντα από τα σημεία δεδομένων. Τα σημεία x_q και y_q δημιουργούν συνήθως ένα ομοιόμορφο πλέγμα, όπως παράγεται από τη συνάρτηση `meshgrid` της matlab. Η τελευταία παράμετρος της συνάρτησης `griddata` είναι η "μέθοδος" ("method"), η οποία ορίζει τον τύπο της προσαρμογής της επιφάνειας στα δεδομένα και έχει τις ακόλουθες τιμές:

- `linear`: γραμμική παρεμβολή με βάση τον τριγωνισμό (προεπιλεγμένη μέθοδος),
- `cubic`: κυβική παρεμβολή με βάση τον τριγωνισμό,
- `nearest neighbor`: παρεμβολή του πλησιέστερου γείτονα με βάση τον τριγωνισμό,
- `natural neighbor`: παρεμβολή φυσικών γειτόνων με βάση τον τριγωνισμό. Αυτή η μέθοδος είναι μια αποτελεσματική εξισορρόπηση μεταξύ γραμμικής και κυβικής,
- `v4`: παρεμβολή ελάχιστης καμπυλότητας. Αυτή η μέθοδος σε αντίθεση με τις άλλες μεθόδους δε βασίζεται σε τριγωνισμό.

Η μέθοδος κυβικής παρεμβολής και παρεμβολής ελάχιστης καμπυλότητας παρατηρείται ότι δημιουργούν ομαλές επιφάνειες, ενώ η μέθοδος γραμμικής παρεμβολής και πλησιέστερου γείτονα έχουν ασυνέχειες στην πρώτη και μηδενική παράγωγο αντίστοιχα. Όλες οι μέθοδοι εκτός από τη μέθοδο παρεμβολής ελάχιστης καμπυλότητας βασίζονται στον τριγωνισμό Delaunay των δεδομένων.

Στη συνέχεια, γίνεται αναφορά στη συνάρτηση `scatteredInterpolant`, η οποία χρησιμοποιείται κι αυτή για πλήρωση κενών. Η συνάρτηση

`F=scatteredInterpolant(x,y,v)`

δημιουργεί μια παρεμβολή που προσαρμόζεται σε μια επιφάνεια της μορφής $v = F(x, y)$. Τα ορίσματα x, y και v είναι διανύσματα και τα δύο πρώτα προσδιορίζουν τις συντεταγμένες (x, y) των σημείων του δείγματος και το τελευταίο περιέχει τις τιμές του δείγματος που σχετίζονται με τα σημεία (x, y) .

Η εντολή `scatteredInterpolant` επίσης διαθέτει τις ακόλουθες μεθόδους παρεμβολής:

- `nearest neighbor`: παρεμβολή πλησιέστερου γείτονα, όπου η επιφάνεια παρεμβολής είναι ασυνεχής. Αυτή είναι η απλούστερη μέθοδος, όπου ένα σημείο παρεμβολής λαμβάνει την ίδια τιμή με το πλησιέστερο δείγμα δεδομένων. Είναι η ταχύτερη σε σχέση με τις υπόλοιπες μεθόδους, αλλά θεωρείται επίσης και η λιγότερο ακριβής.
- `linear`: γραμμική παρεμβολή (προεπιλεγμένη μέθοδος), όπου η επιφάνεια παρεμβολής είναι συνεχής στη μηδενική παράγωγο. Πιο αναλυτικά, η μέθοδος αυτή αξιοποιεί μια αποσύνθεση του δειγματικού χώρου σε τρίγωνα. Όλα τα δείγματα στη συνέχεια ενώνονται σε ομάδες των τριών, φτιάχνοντας μια επιφάνεια που σχηματίζεται από τριγωνικά σχήματα, τα οποία ενώνονται με τις άκρες τους. Κάθε ένα από τα μεμονωμένα τρίγωνα καθορίζει μια επιφάνεια και οι παράμετροι αυτής της επιφάνειας μπορούν να αποκτηθούν, δεδομένων των τριών σημείων που τη διαμορφώνουν. Για να πραγματοποιηθεί η παρεμβολή ενός σημείου, πρώτα εντοπίζεται το σημείο. Το σημείο βρίσκεται σε ένα από τα τρίγωνα που σχηματίζουν το δειγματικό χώρο και η τιμή που παρεμβάλλεται για το σημείο είναι αυτή που ρυθμίζεται στην εξίσωση της επιφάνειας η οποία ορίζεται από τα σημεία του τριγώνου.
- `natural neighbor`: παρεμβολή φυσικών γειτόνων, όπου η επιφάνεια παρεμβολής είναι συνεχής στην πρώτη παράγωγο, με εξαίρεση τα σημεία δειγματοληψίας. Πιο συγκεκριμένα η μέθοδος αυτή βασίζεται στην αποσύνθεση Voronoi ενός συνόλου σημείων. Αυτή η αποσύνθεση παράγει ένα διαχωρισμό του επιπέδου όπου το

σύνολο σημείων βρίσκεται, αναθέτοντας σε κάθε σημείο ένα κελί Voronoi. Το κελί Voronoi ενός συγκεκριμένου σημείου είναι η περιοχή του επιπέδου που εντοπίζεται πιο κοντά σε αυτό σε σχέση με οποιοδήποτε άλλο σημείο του συνόλου. Στην περίπτωση που ένα νέο σημείο εισάγεται στο σύνολο ή παρεμβάλλεται, το διάγραμμα Voronoi τροποποιείται, καθώς ένα κελί πρέπει να ανατεθεί σε αυτό το νέο σημείο. Αυτό το νέο κελί θα έχει τομή με τα υπόλοιπα κελιά που σχημάτιζαν το διάγραμμα πριν από την εισαγωγή του, δηλαδή τα γειτονικά του κελιά. Τέλος, η τιμή του νέου σημείου που έχει εισαχθεί στο σύνολο υπολογίζεται ως ο μέσος όρος των τιμών των γειτονικών κελιών του. Τα βάρη που εφαρμόζονται είναι η αναλογία της περιοχής κάθε γειτονικού κελιού που έχει περάσει στο νέο κελί που περιλαμβάνει το σημείο που έχει γίνει η παρεμβολή.

Η εντολή `griddata` είναι χρήσιμη όταν πρέπει να γίνει παρεμβολή προκειμένου να βρεθούν οι τιμές σε ένα σύνολο προκαθορισμένων θέσεων των σημείων του πλέγματος. Όσον αφορά τα προβλήματα της παρεμβολής, η εντολή `scatteredInterpolant` παρέχει μεγαλύτερη προσαρμοστικότητα. Η συνάρτηση αυτή διαθέτει τα ακόλουθα πλεονεκτήματα:

- Η μέθοδος παρεμβολής μπορεί να αλλάξει ανεξάρτητα από τον τριγωνισμό.
- Οι τιμές στα σημεία δεδομένων μπορούν να τροποποιηθούν ανεξάρτητα από τον τριγωνισμό.
- Τα σημεία δεδομένων μπορούν να προστεθούν σταδιακά στην υπάρχουσα παρεμβολή χωρίς να δημιουργηθεί όμως πρόβλημα. Τα σημεία δεδομένων μπορούν επίσης να τροποποιηθούν και να μετακινηθούν επιτυχώς, υπό την προϋπόθεση ότι ο αριθμός των σημείων που έχουν επεξεργασθεί είναι μικρός σε σύγκριση με τον συνολικό αριθμό των σημείων του δείγματος.
- Παρέχει λειτουργικότητα της παρέκτασης για την προσέγγιση τιμών σε σημεία που εντοπίζονται εκτός του κυρτού περιβλήματος.

4.2 Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

Για την επίλυση του προβλήματος παρεμβολής χωρικών δεδομένων γίνεται εφαρμογή του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC). Ο αλγόριθμος αυτός μπορεί να εφαρμοστεί σε διακριτά και συνεχή δεδομένα πραγματικών τιμών (παλινδρόμηση). Σχετικά με το πρόβλημα παλινδρόμησης, ο INNC προσεγγίζει τις συνεχείς μεταβλητές μέσω ενός συγκεκριμένου αριθμού τάξεων. Το INNC καθορίζει την ταυτότητα της τάξης σε μη μετρούμενα σημεία αξιοποιώντας τη μεθοδολογία Monte Carlo. Επίσης, είναι μη παραμετρικός και αυτό είναι σημαντικό διότι αποδεικνύεται ότι είναι κατάλληλος για μη Γκαουσιανά δεδομένα. Παρατηρείται σύμφωνα με έρευνες ότι η μέθοδος είναι πολύ ανταγωνιστική αναφορικά με την ακρίβεια της παρεμβολής και την υπολογιστική αποδοτικότητα συγκριτικά με ορισμένες μεθόδους.

Στη συνέχεια, παρουσιάζεται ο ορισμός του προβλήματος παρεμβολής. Αρχικά, θεωρείται ένα σύνολο σημείων δειγματοληψίας $G_s = \vec{r}_i$ όπου $\vec{r}_i = (x_i, y_i) \in \mathbb{R}^2$ και $i = 1, \dots, N$. Τα σημεία αυτά αντιστοιχίζονται σε ένα ορθογώνιο πλέγμα \tilde{G} μεγέθους $N_G = L_x \times L_y$, όπου L_x και L_y είναι η οριζόντια και η κατακόρυφη διάσταση του ορθογωνίου αντίστοιχα και $N < N_G$. Στο σημείο r_i αποδίδεται μια τιμή z_i και έτσι το σύνολο $Z_s = \{z_i \in \mathbb{R}\}_{i=1}^N$ αντικατοπτρίζει το δείγμα της διαδικασίας. Το σύνολο των σημείων πρόβλεψης ορίστηκε ως $G_p = \{\vec{r}_p\}_{p=1}^P$ όπου $p = 1, \dots, P$ έτσι ώστε $\tilde{G} = G_s \cup G_p$. Το Z_s για διακριτά δεδομένα παίρνει τιμές σε ένα σύνολο διακριτών τάξεων (ετικετών) $\{C_q\}_{q=1}^{N_c}$, ενώ για εφαρμογές παλινδρόμησης, το Z_s μπορεί να αναπαρασταθεί ως υλοποίηση ενός υποκείμενου τυχαίου πεδίου συνεχών τιμών $Z(\vec{r}_i)$. Έπειτα, χρησιμοποιώντας έναν αριθμό τάξεων C_q με $q = 1, \dots, N_c$ διακριτοποιείται η συνεχής κατανομή. Οι τάξεις καθορίζονται σε σχέση με ένα σύνολο επιπέδων κατωφλίου t_k με $k = 1, \dots, N_c + 1$, όπου $t_1 = \min(z_1, \dots, z_N)$ και $t_{N_c+1} = \max(z_1, \dots, z_N)$. Κάθε τάξη C_q αντιστοιχεί σε ένα διάστημα ως εξής: $C_q = (t_q, t_{q+1}]$ με $q = 2, \dots, N_c - 1$, $C_1 = (t_1, t_2]$ και $C_{N_c} = (t_{N_c-1}, t_{N_c})$. Δηλώνεται το πεδίο δεικτών $I_Z(\vec{r}_p)$ να λαμβάνει α-

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

κέραιες τιμές $q \in 1, \dots, N_c$ ίσες με τον κατάλληλο δείκτη τάξης για μια συγκεκριμένη τιμή την $z(\vec{r}_p)$. Δηλαδή, από το $I_Z(\vec{r}_p) = q$ προκύπτει ότι $z(\vec{r}_p) \in C_q$, όπου C_q ορίζεται είτε μια συγκεκριμένη ετικέτα για το πρόβλημα ταξινόμησης είτε ένα συγκεκριμένο διάστημα για το πρόβλημα παλινδρόμησης. Επομένως, το πρόβλημα παρεμβολής μπορεί να καθοριστεί ως πρόβλημα ταξινόμησης τόσο για τις εφαρμογές ταξινόμησης όσο και για τις εφαρμογές παλινδρόμησης. Το μοντέλο Ising αξιολογεί την ταυτότητα τάξης του Z στα σημεία πρόβλεψης και αφού όλα τα σημεία πρόβλεψης έχουν αντιστοιχιστεί σε μια τάξη, μπορεί να δημιουργηθεί ένας χάρτης της διαδικασίας Z που περιλαμβάνει περιγράμματα ισοδύναμων τάξεων.

Για κάθε τάξη q , ας θεωρήσουμε ένα σύνολο μεταβλητών $\{s_i^q\}_{i=1}^N$, το οποίο αντιπροσωπεύει τιμές σπιν (“μαγνητικών ροπών”). Το σύνολο αυτό μπορεί να πάρει τις εξής τιμές $s_i^q = 1$, που αντιστοιχεί σε θετικό σπιν (spin-up) ή $s_i^q = -1$, που αντιστοιχεί σε αρνητικό σπιν (spin-down). Το μοντέλο Ising μελετά τις αλληλεπιδράσεις που γίνονται μεταξύ των σπιν (δηλαδή, από πάνω σε κάτω και αντίστροφα) κατά ζεύγη και εκφράζονται από την ακόλουθη σχέση:

$$H[\{s\}] = - \sum_{i,j} J_{i,j} s_i s_j - \sum_i h_i s_i, \quad (4.1)$$

όπου το σύμβολο $H[\{s\}]$ δηλώνει ότι η ενέργεια είναι μια συνάρτηση του συνόλου των τιμών των σπιν δηλαδή αναφερόμαστε στη χωρική διάταξη των σπιν.

Γενικά, οι διατάξεις σπιν που καταλήγουν σε χαμηλότερη ενέργεια, ενδέχεται να πραγματοποιηθούν. Ο πρώτος όρος στην ενέργεια αντιστοιχεί στην αλληλεπίδραση ενέργειας ανταλλαγής σπιν (spin-spin exchange). Όσον αφορά το $J_{i,j}$ είναι η ισχύς σύζευξης, η οποία συμβάλλει στον έλεγχο της ισχύς καθώς και το είδος της αλληλεπίδρασης: αν $J_{i,j} > 0$ είναι σιδηρομαγνητική και αυτό ευνοεί τις περιστροφές με το ίδιο πρόσημο, ενώ αν $J_{i,j} < 0$ είναι αντισιδηρομαγνητική και αυτό ευνοεί τις περιστροφές με αντίθετο πρόσημο. Ο δεύτερος όρος αντιστοιχεί σε ένα σφάλμα που δεν καταφέρνει τη συμμετρία και αυτό συμβαίνει λόγω

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

της παρουσίας ενός εξαρτώμενου από τη θέση εξωτερικού πεδίου h_i . Να σημειωθεί ότι οι θετικές και αντίστοιχα οι αρνητικές τιμές του εξωτερικού πεδίου ευνοούν τις περιστροφές με το ίδιο πρόσημο. Συνεπώς, το h_i ελέγχει τη συνολική κατανομή των τιμών του σπιν μεταξύ 1 και -1 , το οποίο προσδιορίζει την συνολική “μαγνήτιση”. Η ισχύς σύζευξης $J_{i,j}$ παρουσιάζεται συνήθως ομοιόμορφη και το εύρος της οριοθετείται στους πλησιέστερους γείτονες. Αν και το μοντέλο μπορεί να γενικευτεί έτσι ώστε να περιλαμβάνει και μη ομοιόμορφες ζεύξεις οι οποίες κατηγοριοποιούνται σε μεγαλύτερη εμβέλεια.

Στη συνέχεια, παρουσιάζεται η συνάρτηση πυκνότητας πιθανότητας για μια χωρική διάταξη των σπιν $\{s\}$, η οποία δίνεται από την ακόλουθη εκθετική σχέση Boltzmann-Gibbs:

$$f[\{s\}] = \frac{e^{\frac{-H[\{s\}]}{k_B T}}}{Z} \quad (4.2)$$

όπου k_B είναι η σταθερά του Boltzmann και T είναι η θερμοκρασία. Επίσης, η συνάρτηση κατανομής Z είναι ένας παράγοντας κανονικοποίησης που απορρέει από το άθροισμα του εκθετικού $\exp[-H[\{s\}]/k_B T]$ σε όλες τις πιθανές διατάξεις των σπιν. Επομένως, είναι συνάρτηση μόνο των παραμέτρων του μοντέλου $J_{i,j}$ και h_i αλλά όχι μιας συγκεκριμένης χωρικής διάταξης των σπιν.

Ας θεωρήσουμε ότι η ισχύς της σύζευξης και το πεδίο πόλωσης είναι γνωστά, και υπάρχει ενδιαφέρον για τις πιο πιθανές διατάξεις σπιν ή για τον υπολογισμό της συνάρτησης συσχέτισης των σπιν. Αν οι περιστροφές σε ορισμένες θέσεις είναι γνωστές, μπορούν δηλαδή να προσδιοριστούν από τις δειγματοληπτικές τιμές του πεδίου τότε η διαδικασία εκτίμησης εστιάζει στην εξαγωγή των παραμέτρων του μοντέλου (για παράδειγμα μέσω της μεθόδου μέγιστης πιθανοφάνειας) που αντικατοπτρίζουν καλύτερα τις παρατηρήσεις. Όμως, η σταθερά κανονικοποίησης Z είναι στις περισσότερες περιπτώσεις ανέφικτο να επιλυθεί με αναλυτικά μέσα και συγχρόνως ο αριθμητικός υπολογισμός της είναι ιδιαίτερα κοστοβόρος για μεγάλα συστήματα. Πιθανές λύσεις για την παράκαμψη αυτού του προβλήματος, όπως η προσέγγιση της μέγιστης ψευδο-πιθανοφάνειας ή

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

διάφορες τεχνικές εκτίμησης βασισμένες σε Markov Chain Monte Carlo, μπορεί να αποδειχθούν είτε ανακριβείς είτε αργές, αντίστοιχα.

Μόλις καθοριστούν οι παράμετροι του μοντέλου, οι βέλτιστες τιμές των σπιν στις μη δειγματοληπτημένες θέσεις δηλαδή, όπου υπάρχουν τα κενά δεδομένων, μπορούν να προσδιοριστούν μεγιστοποιώντας την υπό συνθήκη, ως προς τα δεδομένα, πιθανότητα f (ισοδύναμα ελαχιστοποιώντας την H) ως προς τις άγνωστες τιμές.

Για να αποφευχθεί το δύσκολο πρόβλημα της εκτίμησης των παραμέτρων, θα αξιοποιήσει μια μη παραμετρική μέθοδο, που θα αναφερθεί παρακάτω, η οποία δεν χρειάζεται να γνωρίζει τις παραμέτρους του μοντέλου Ising.

4.2.1 Ο Αλγόριθμος Πολυεπίπεδης Παρεμβολής INNC

Το μοντέλο Ising περιορίζεται στην απλούστερη ενεργειακή συνάρτηση. Θέτουμε ομοιόμορφα το πεδίο πόλωσης στο μηδέν, δηλαδή $h_i = 0, i = 1, \dots, N$ και περιορίζουμε τις αλληλεπιδράσεις ανταλλαγής σε ομοιόμορφη θετική τιμή μόνο για τους πλησιέστερους γείτονες (nnb), δηλαδή, $J_{i,j} = J > 0$ εάν $i \in \text{nnb}(j)$ και $J_{i,j} = 0$ διαφορετικά. Η χρήση μηδενικού πολωτικού πεδίου εμποδίζει τον προσεκτικό έλεγχο του λόγου των “θετικών” προς τα “αρνητικά” σπιν. Αυτό επιτυγχάνεται στις προσομοιώσεις με την επιλογή αρχικών τιμών “σπιν”. Αυτές αντιπροσωπεύουν την κατανομή των σπιν του δείγματος ανάμεσα σε θετικές και αρνητικές τιμές, όπως αναλύεται παρακάτω.

Όταν σε ένα πρόβλημα παρεμβολής εμπλέκονται μόνο δύο τάξεις, η παρεμβολή γίνεται σε ένα πέρασμα. Τα $Z_s = \{z_i\}_{i=1}^N$ μετατρέπονται σε διακριτές μεταβλητές (δηλαδή σε “σπιν”). Μια ιεραρχική στρατηγική παρεμβολής χρησιμοποιείται όταν το πρόβλημά μας έχει πολλές τάξεις (λόγω της διακριτοποίησης μιας συνεχούς μεταβλητής). Καθώς ο δείκτης τάξης q μετακινείται από το 1 στο N_c σε αυτή τη μέθοδο, τα υποπλέγματα δειγματοληψίας, G_s^p , και πρόβλεψης, G_p^q , ανανεώνονται σταδιακά. Για τη χαμηλότερη τάξη ισχύει ότι $G_s^1 = G_s$ και $G_p^1 = G_p$ όπου G_s και G_p είναι τα αρχικά πλέγματα δειγματοληψίας και παρεμβολής, αντίστοιχα.

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

Για όλες τις τάξεις $q = 1, \dots, N_c$, ισχύει ότι $G_p^q \cup G_s^q = \tilde{G}$.

Οι περιοχές με αρνητικά σπιν εντάσσονται στο ενημερωμένο υπόπλεγμα δείγματος όταν το q αυξάνεται και ταυτόχρονα αφαιρούνται από το υπόπλεγμα πρόβλεψης. Η διακριτοποίηση σε κάθε επίπεδο q είναι δυαδική σε σχέση με την τιμή κατωφλίου, δηλαδή, $s_i^q = -1$ εάν $z_i \leq t_{q+1}$ και $s_i^q = 1$ εάν $z_i > t_{q+1}$ για $i = 1, \dots, N_q$, όπου N_q είναι ο αριθμός των περιοχών με γνωστές τιμές στο επίπεδο q . Οι κόμβοι πλέγματος $\vec{r}_l \in G_p$, για τους οποίους $s_l^{q-1} = -1$, αυξάνουν (μειώνουν) το υπόπλεγμα δείγματος (πρόβλεψης) για $q > 1$. Ως αποτέλεσμα, έχουμε $N_1 = N$ και $N_{q>1} \geq N$. Όλες οι τιμές σπιν για τον δείκτη τάξης q περιλαμβάνονται στο σύνολο $S_s^q = \{s_i^q\}_{i=1}^{N_q}$, όπου $q = 1, \dots, N_c$. Οι τιμές “σπιν” για το συγκεκριμένο επίπεδο περιέχονται στην ένωση των δύο συνόλων που περιλαμβάνουν τις τιμές δείγματος και πρόβλεψης στο επίπεδο q , δηλαδή, $S^q = S_s^q \cup S_p^q$. Το μοντέλο Ising μπορεί στη συνέχεια να χρησιμοποιηθεί για την απεικόνιση των χωρικών αλληλεπιδράσεων μεταξύ των σπιν \tilde{S}^q για το επίπεδο q , υπονοώντας ότι τα σπιν ορίζονται ως προς το κατώφλι δυαδικοποίησης.

Η προτεινόμενη ιεραρχική δομή παρέχει έναν μη παραμετρικό τρόπο για την παράκαμψη του προβλήματος εξαγωγής συμπερασμάτων για τις παραμέτρους. Η μέθοδος αυτή χρησιμοποιεί μια συνάρτηση κόστους, $\mathcal{U}(S_p^q | S_s^q)$, για τον υπολογισμό της απόκλισης (τετραγωνική διαφορά) μεταξύ μιας κατάλληλα κανονικοποιημένης ενέργειας, C_s^q , της χωρικής διάταξης του δείγματος στο επίπεδο q και της αντίστοιχης ενέργειας της διάταξης των σπιν \tilde{C}^q σε ολόκληρο το πλέγμα. Αυτό δίνεται από την ακόλουθη εξίσωση:

$$\mathcal{U}(S_p^q | S_s^q) = (\tilde{C}^q - C_s^q)^2, \quad (4.3)$$

όπου η συσχέτιση ζεύγους σπιν της δειγματικής διάταξης των σπιν στο επίπεδο q δίνεται από τη σχέση $C_s^q = \langle s_i^q s_j^q \rangle_{G_s^q}$, ενώ η συσχέτιση ζεύγους σπιν σε όλο το πλέγμα δίνεται από τη σχέση $\tilde{C}^q = \langle s_i^q s_j^q \rangle_{\tilde{G}}$ η τελευταία περιέχει τόσο το S_p^q όσο και το S_s^q .

Ως αποτέλεσμα, η απόδοση της σωστής κατηγορίας στα σπιν S_p^q πε-

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

ριορίζεται στον προσδιορισμό της καλύτερης διάταξης των \hat{S}_p^q , που ελαχιστοποιεί τη συνάρτηση κόστους Εξ. (4.3) σε δεδομένη θερμοκρασία T .

$$\hat{S}_p^q = \arg \min_{S_p^q} \mathcal{U}(S_p^q | S_s^q). \quad (4.4)$$

Ο ιεραρχικός αλγόριθμος ξεκινά από το χαμηλότερο κατώφλι δυαδικοποίησης και ανεβαίνει μέχρι τον υψηλότερο δείκτη τάξης. Η δυαδική διακριτοποίηση και κατηγοριοποίηση των μη μετρημένων περιοχών γίνεται πρώτα σε σχέση με την πρώτη τάξη και στη συνέχεια για τις επόμενες τάξεις με διαδοχική σειρά. Καθώς ο αλγόριθμος προχωρά σε διαδοχικά επίπεδα, τα “κενά” στο υπόπλεγμα πρόβλεψης, G_p , συμπληρώνονται σταδιακά. Όλες οι θέσεις που ανιχνεύονται ως έχουσες τιμή spin -1 σε χαμηλότερα επίπεδα χρησιμοποιούνται ως είσοδος (δεδομένα δείγματος) στο τρέχον στάδιο σε κάθε επίπεδο. Για τον δείκτη τάξης q , το μειωμένο υπόπλεγμα πρόβλεψης, G_p^q , περιλαμβάνει P^q σημεία, που σημαίνει ότι $P^q \leq P^{q'}$ και $P^1 = P$ για $q > q'$. Όταν πρόκειται για συνεχείς μεταβλητές, οι τάξεις C_q μπορούν να δημιουργηθούν όπως θέλετε και δεν χρειάζεται να αντιπροσωπεύουν διαστήματα ομοιόμορφου μεγέθους.

Ο λόγος απόρριψης χρησιμοποιείται από τον αλγόριθμο του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC) και ορίζεται ως εξής:

$$\rho = \frac{\text{ο αριθμός των καταστάσεων που απορρίφθηκαν}}{\text{ο αριθμός των προσομοιωμένων καταστάσεων}}.$$

Ο λόγος απόρριψης ενημερώνεται σε τακτική βάση και χρησιμοποιείται για να καθορίσει πότε ο αλγόριθμος πρέπει να σταματήσει, να προτείνει νέες καταστάσεις και να μεταβεί στο επόμενο επίπεδο τάξης q .

Τα βασικά βήματα της μεθόδου του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC) απεικονίζονται στον ψευδοκώδικα του αλγορίθμου που βρίσκεται στην εργασία [14]. Η προσέγγιση αυτή παράγει ένα πεδίο δεικτών $\hat{I}_Z = I_Z(G_s) \cup \hat{I}_Z(G_p)$, το οποίο περιέχει τις αρχικές τάξεις του δείγματος καθώς και τις εκτιμήσεις των τάξεων G_p .

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

Επειδή η αρχική κατάσταση λαμβάνει υπόψιν αυτές τις τιμές και οι επαναληπτικές διαδικασίες παρακάμπτουν τις περιοχές στο ενημερωμένο σύνολο δειγμάτων S_s^q , οι τιμές των δεικτών στις περιοχές δειγματοληψίας αναδημιουργούνται με απόλυτη ακρίβεια. Το σύνολο εκπαίδευσης αναφέρεται ως $I_Z(G_s)$.

Αξίζει να σημειωθεί ότι ενώ ο αλγόριθμος προορίζεται για μη διανυσματική υλοποίηση, η δημιουργία νέων καταστάσεων επιτυγχάνεται στην πραγματικότητα με ενημέρωση του διανυσματικού σπιν Metropolis. Επειδή το τετραγωνικό πλέγμα μπορεί να καταταμηθεί σε δύο αλληλοδιεισδυτικά υποπλέγματα με τρόπο σκακιέρας, η διανυσματοποίηση είναι δυνατή. Ως αποτέλεσμα, τα σπιν στο πρώτο υπόπλεγμα αλληλεπιδρούν μόνο με τα σπιν στο δεύτερο υπόπλεγμα και αντίστροφα, λόγω της μικρής εμβέλειας της αλληλεπίδρασης που περιορίζεται στους πλησιέστερους γείτονες. Αυτή η αποσύνθεση επιτρέπει την παράλληλη εφαρμογή της ενημέρωσης της μεθόδου σε σπιν που ανήκουν στο ίδιο υπόπλεγμα. Ο αλγόριθμος επαναλαμβάνει τη σάρωση σε όλο το πλέγμα έως ότου ο λόγος απόρριψης ξεπεράσει το κατώφλι (εδώ έχει οριστεί σε ένα).

Η ενημέρωση σπιν πραγματοποιείται σε θερμοκρασία 0°C . Δεν υπάρχει στοχαστική επιλογή δυσμενών σπιν με τον περιορισμό $T = 0$. Ως αποτέλεσμα, τα υποψήφια “σπιν” που ενημερώνονται, αναστρέφονται μόνο εάν η αναστροφή μειώνει τη συνάρτηση κόστους. Αυτό αναφέρεται ως “άπληστος” αλγόριθμος Monte Carlo και εξασφαλίζει γρήγορη σύγκλιση. Η προσομοιωμένη ανόπτηση (simulated annealing), από την άλλη πλευρά, ξεκινά με μια κατάσταση υψηλής θερμοκρασίας και μειώνει σταδιακά το T . Αυτή η μέθοδος είναι πιο αργή από άποψη υπολογισμού, αλλά η προκύπτουσα διαμόρφωση είναι λιγότερο ευαίσθητη στην αρχική κατάσταση $\hat{S}_p^{q(0)}$. Η ευαισθησία του άπληστου αλγορίθμου είναι γνωστό ότι είναι ιδιαίτερα ισχυρή σε χώρους υψηλών διαστάσεων με μη κυρτές ενέργειες. Σε τέτοιες περιπτώσεις, αντί να συγκλίνει στο παγκόσμιο ελάχιστο, ο άπληστος αλγόριθμος είναι πιο επιρρεπής στο να παγιδευτεί σε τοπικά ελάχιστα. Το πρόβλημα της παρεμβολής, ωστόσο, δεν επηρεάζεται από αυτό. Στην πραγματικότητα, η εστίαση στο πα-

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

γκόσμιο ελάχιστο της συνάρτησης κόστους \mathcal{U} υπερεκτιμά την ενέργεια συσχέτισης του δείγματος ανά ζεύγος “σπιν” C_s^q , αγνοώντας το γεγονός ότι επηρεάζεται από τις διακυμάνσεις μεταξύ των δειγμάτων [10].

Υπάρχουν διάφορες μέθοδοι για την επιλογή της αρχικής διάταξης των σπιν. Η αρχική διαμόρφωση θα πρέπει να ελαχιστοποιεί τη διαδρομή χαλάρωσης (στο χώρο καταστάσεων) προς την κατάσταση ισορροπίας, επειδή το προτεινόμενο μοντέλο προσπαθεί να παρέχει μια γρήγορη και αυτόματη προσέγγιση παρεμβολής. Θα πρέπει επίσης να επιλέγεται με ελάχιστη ή καθόλου συμβολή του χρήστη. Το $\hat{S}_p^{q(0)}$ υπολογίζεται με βάση τις καταστάσεις S_s^q του “επαυξημένου δείγματος” στην άμεση γειτονιά κάθε μεμονωμένου σημείου πρόβλεψης, υποθέτοντας έναν ορισμένο βαθμό χωρικής συνέχειας, ο οποίος είναι συχνός στα σύνολα γεωχωρικών δεδομένων. Ένα προσαρμοστικό τετραγωνικό πλαίσιο INNC $m \times m$ (όπου $m = 2l + 1$) με κέντρο το \vec{r}_p χρησιμοποιείται για τον προσδιορισμό της γειτονιάς ενός τέτοιου κόμβου \vec{r}_p . Το μέγεθος του πλαισίου INNC $m \leq m_{\max}$ καθορίζεται προσαρμοστικά, λαμβάνοντας υπόψη την πυκνότητα δειγματοληψίας και τις κατανομές τιμών σπιν στη γύρω περιοχή.

Ξεκινώντας με μια αρχική τιμή $m = 3$, κοιτάξαμε να δούμε αν υπήρχε μια σαφής πλειοψηφία θετικών ή αρνητικών τιμών σπιν μέσα στο πλαίσιο INNC. Εάν δεν συμβαίνει αυτό, θα αυξήσουμε το m κατά ένα, θα δοκιμάσουμε ξανά και θα συνεχίσουμε τη διαδικασία όπως απαιτείται. Για να αποφευχθεί η υπερχείλιση και να μειωθεί η υπολογιστική προσπάθεια, τίθεται ένα αυθαίρετο άνω όριο m_{\max} στο μέγεθος του πλαισίου INNC. Στη συνέχεια, χρησιμοποιώντας τον κανόνα της πλειοψηφίας, το $\hat{s}_p^{q(0)}$ ορίζεται ανάλογα με την τρέχουσα τιμή των γειτόνων του στο S_s^q μέσα στο πλαίσιο INNC. Η αρχική τιμή αντιστοιχίζεται τυχαία εάν δεν υπάρχει επικρατέστερο πρόσημο (δηλαδή εάν υπάρχει ίσος αριθμός τιμών $+1$ και -1 ή εάν το \vec{r}_p δεν έχει γείτονες στο S_s^q μέσα στο πλαίσιο INNC).

Οι προτεινόμενες τροποποιήσεις του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC) εγκρίνονται ανεπιφύλακτα εφόσον μειώνουν τη συνάρτηση κόστους της Εξ. (4.3). Ολόκληρο το πλέγμα σαρώνεται σε δύο βήματα με τη διανυσματική μέθοδο σκακιέρας. Εάν μια

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

πλήρης σάρωση του υποπλέγματος παρεμβολής G_p^q δεν παρέχει ούτε μία επιτυχή ενημέρωση για δεδομένο δείκτη τάξης q , η προσομοίωση τερματίζεται. Η ιεραρχική δομή δείχνει ότι όσο αυξάνεται το q , μειώνεται ο υπολογιστικός φόρτος, γεγονός που αντιστοιχεί στη μείωση του μεγέθους του υποπλέγματος G_p^q .

Ο αλγόριθμος απαιτεί ως δεδομένα εισόδου τον ορισμό των τάξεων, καθώς και το μέγιστο μέγεθος πλαισίου INNC m_{\max} που χρησιμοποιείται για την κατασκευή της αρχικής κατάστασης. Ο αριθμός των τάξεων καθορίζεται από τη φύση του προβλήματος και τον στόχο της μελέτης. Όταν πρόκειται για διακριτές ετικέτες τάξεων, ο αριθμός των ετικετών είναι προκαθορισμένος και δεν απαιτείται διακριτοποίηση. Η διακριτοποίηση εξαρτάται από τον στόχο της μελέτης εάν το πρόβλημα παρεμβολής είναι μια διαδικασία συνεχούς αξίας. Η δυαδική κατηγοριοποίηση είναι επαρκής εάν ο σκοπός είναι ο υπολογισμός των ορίων υπέρβασης. Ένας μέτριος αριθμός τάξεων (π.χ. έξι ή οκτώ) είναι συχνά επαρκής για την περιβαλλοντική παρακολούθηση και τη λήψη αποφάσεων. Η χρήση όμως, μεγαλύτερου αριθμού τάξεων βοηθάει στον υπολογισμό των τιμών της μοντελοποιημένης διαδικασίας με πιο αποτελεσματικό τρόπο.

Οι εκτιμώμενες τιμές στους κόμβους εγκυρότητας συγκρίνονται με τις πραγματικές τιμές για την αξιολόγηση της απόδοσης του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC), οι οποίες αφαιρέθηκαν από το αντίστοιχο δείγμα. Οι τιμές του δείκτη $I_Z(G_p)$ στους κόμβους εγκυρότητας συγκρίνονται με τις εκτιμήσεις $\hat{I}_Z(G_p)$ που προκύπτουν μετά την εξαίρεση του συνόλου των κόμβων G_p από τα δεδομένα για την αξιολόγηση της ταξινόμησης. Χωρίς περαιτέρω επεξεργασία, χρησιμοποιούνται τα διακριτά δεδομένα.

4.2. Μοντέλο Ising με Συσχετίσεις Κοντινότερων Γειτόνων

Κεφάλαιο 5

Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα που προέκυψαν από το προγραμματιστικό περιβάλλον matlab για τα συνθετικά δεδομένα που δημιουργήθηκαν από τη συνάρτηση Ackley και για τα δορυφορικά δεδομένα. Για τα συνθετικά δεδομένα θα διακρίνουμε κι άλλες περιπτώσεις όσον αφορά τη θέση των κενών. Αρχικά, θα εφαρμόσουμε τις αιτιοκρατικές μεθοδολογίες που υλοποιήθηκαν με τις εντολές `griddata` και `scatteredInterpolant` οι οποίες συμβάλλουν στη χωρική παρεμβολή διάσπαρτων δεδομένων ή δεδομένων με συνεχείς τιμές, με μεθόδους που υποδείξαμε στο Κεφάλαιο 4 και στη συνέχεια τη μέθοδο του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC). Τα ίδια βήματα θα χρησιμοποιήσουμε και για τα δορυφορικά δεδομένα, τα οποία τα χωρίζουμε σε πραγματικά και προσομοιωμένα δεδομένα. Επιπλέον, θα αξιολογήσουμε τα αποτελέσματα των μέτρων διασταυρωτικής επιβεβαίωσης. Τα μέτρα που υπολογίστηκαν ήταν τα ρ , ME, RMSE και MAE. Τα συγκεκριμένα μέτρα παρουσιάστηκαν αναλυτικά στο Κεφάλαιο 2.

5.1 Αποτελέσματα για Συνθετικά Δεδομένα

Όπως αναφέραμε στο Κεφάλαιο 3, τα συνθετικά δεδομένα δημιουργήθηκαν με τη συνάρτηση Ackley. Ο λόγος που χρησιμοποιήθηκαν αυτά

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

τα δεδομένα είναι για να διερευνήσουμε πως συμπεριφέρεται η μέθοδος INNC σε αυτά και αργότερα στα δορυφορικά που θα δούμε στην επόμενη ενότητα. Θα δούμε στη συνέχεια ότι λάβαμε υπόψιν κι άλλες περιπτώσεις όσον αφορά τα κενά και εκτός από τη μέθοδο INNC, εφαρμόσαμε τις αιτιοκρατικές μεθόδους προκειμένου να έχουμε ολοκληρωμένη άποψη για την πλήρωση κενών. Οι περιπτώσεις κατηγοριοποιήθηκαν σε συνεχή περιοχή και σε τυχαία κατανομή κενών εικονοστοιχείων. Όπως θα δούμε και παρακάτω τα κενά χωρίστηκαν σε 40 και 303 κενά.

Στην αρχή, χρησιμοποιήσαμε την τυχαία κατανομή 40 κενών εικονοστοιχείων στα συνθετικά δεδομένα και θα αξιολογήσουμε παρακάτω τις μεθόδους ως προς την πρόβλεψη των κενών. Στη συνέχεια, τα κενά εικονοστοιχεία επιλέχθηκαν να είναι σε συνεχή περιοχή και ο αριθμός των κενών ρυθμίστηκε με βάση το ποσοστό επί των συνολικών σημείων του πλέγματος που υπάρχουν στα δορυφορικά δεδομένα συγκεκριμένα στα προσομοιωμένα. Στα προσομοιωμένα δεδομένα γνωρίζουμε ότι τα 44 431 από τα 150 000 είναι δεδομένα με κενά επομένως, το ποσοστό είναι 29.62%. Αντίστοιχα για τα συνθετικά δεδομένα, αφού το ποσοστό είναι ίδιο, ο αριθμός των κενών υπολογίζεται ότι είναι 303 περίπου.

Έπειτα, τοποθετήθηκαν τα κενά εικονοστοιχεία σε συνεχόμενη περιοχή στο κέντρο για να διαπιστώσουμε πως είναι η πρόβλεψη των κενών σε σύγκριση με την περίπτωση όπου η περιοχή των κενών είναι στη δεξιά πλευρά της εικόνας. Αυτή η περίπτωση συμπεριλαμβάνεται διότι τα κενά σε συνεχή περιοχή όταν τοποθετούνται μέχρι και τη τελευταία στήλη του πίνακα στο προγραμματιστικό περιβάλλον matlab, φαίνεται όπως θα δούμε και παρακάτω, ότι δεν λαμβάνεται υπόψιν. Έτσι, παραμένουν κάποιες τιμές της τελευταίας στήλης άθικτες. Αυτό θα το εξηγήσουμε παρακάτω στα αποτελέσματα. Τέλος, εφαρμόστηκε για τον ίδιο αριθμό κενών εικονοστοιχείων τυχαία κατανομή, για να παρατηρήσουμε πως συμπεριφέρονται σε αυτή την περίπτωση οι μέθοδοι.

5.1.1 Τυχαία Κατανομή 40 Κενών Εικονοστοιχείων

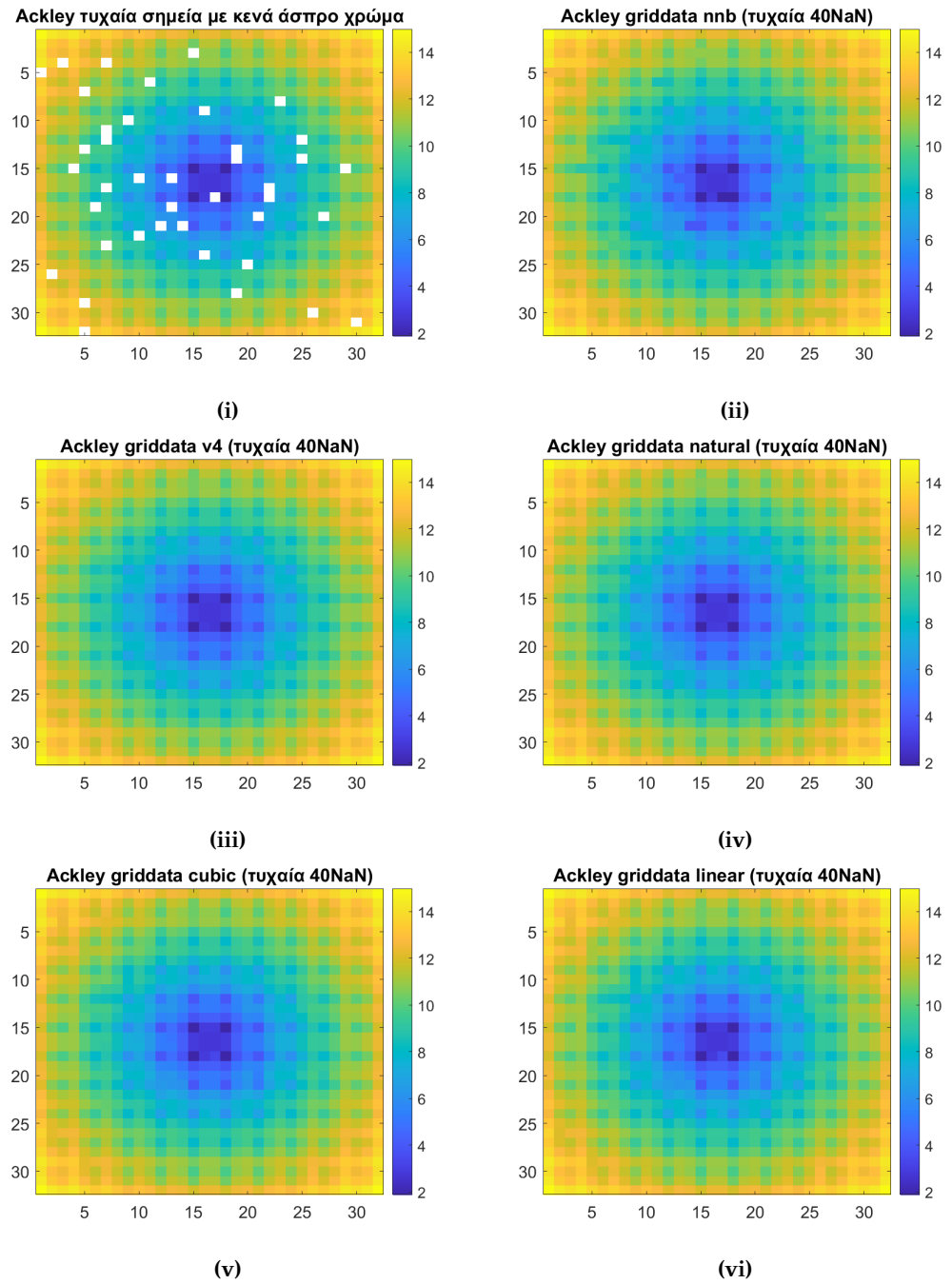
Στο Σχήμα 5.1i παρουσιάζονται τα κενά που έχουν τοποθετηθεί με τυχαίο τρόπο και στο σύνολο τους είναι 40. Αρχικά, εφαρμόσαμε τις αιτιοκρατικές μεθόδους στα συνθετικά δεδομένα. Στο Σχήμα 5.1ii χρησιμοποιήσαμε τη μέθοδο των πλησιέστερων γειτόνων (επιλογή `nearest neighbor`) στην εντολή `griddata`. Διακρίνουμε ότι η ανακατασκευή της εικόνας είναι επιτυχής. Το ίδιο παρατηρούμε με όλες τις μεθόδους χωρικής παρεμβολής που διαθέτει η εντολή `griddata` για τη συγκεκριμένη περίπτωση. Πιο αναλυτικά, στο Σχήμα 5.1iii με μέθοδο παρεμβολής ελάχιστης καμπυλότητας (επιλογή `n4`), στο 5.1iv με μέθοδο παρεμβολής φυσικών γειτόνων (επιλογή `natural neighbor`), στο 5.1v με μέθοδο κυβικής παρεμβολής (επιλογή `cubic`) και στο 5.1vi με μέθοδο γραμμικής παρεμβολής (επιλογή `linear`) έχει επιτευχθεί ο στόχος της πλήρωσης των κενών.

Είναι σημαντικό να επισημανθεί ότι με τη βοήθεια και των μέτρων διασταυρωτικής επιβεβαίωσης θα μπορέσουμε να έχουμε καλύτερη αξιολόγηση των μεθόδων.

Επιπλέον, διερευνήσαμε τη μέθοδο των πλησιέστερων γειτόνων (επιλογή `nearest neighbor`), μέθοδο παρεμβολής φυσικών γειτόνων (επιλογή `natural neighbor`) και τη μέθοδο γραμμικής παρεμβολής (επιλογή `linear`) υλοποιημένες με την εντολή `scatteredInterpolant` για τα Σχήματα 5.2i, 5.2ii και 5.2iii αντίστοιχα. Τα αποτελέσματα που προέκυψαν, παρέχουν την ίδια ανακατασκευή με αυτή από το προηγούμενο Σχήμα 5.1, εφόσον έχουν εφαρμοστεί οι ίδιες μέθοδοι χωρικής παρεμβολής, εξάίρεση αποτελεί η μέθοδος κυβικής παρεμβολής (επιλογή `cubic`). Η μέθοδος αυτή δεν υλοποιείται με την εντολή `scatteredInterpolant`.

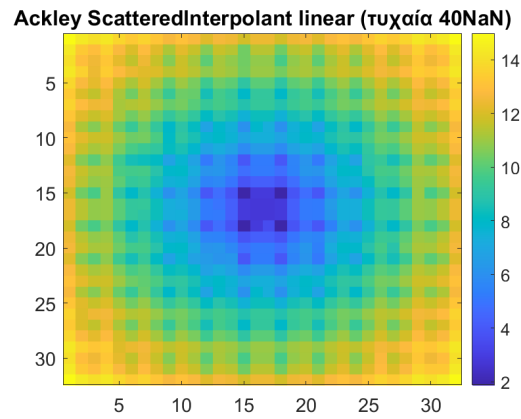
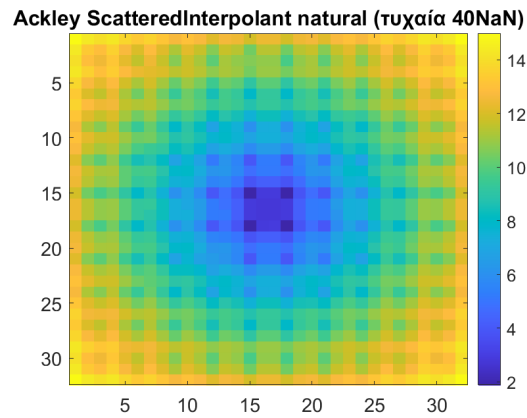
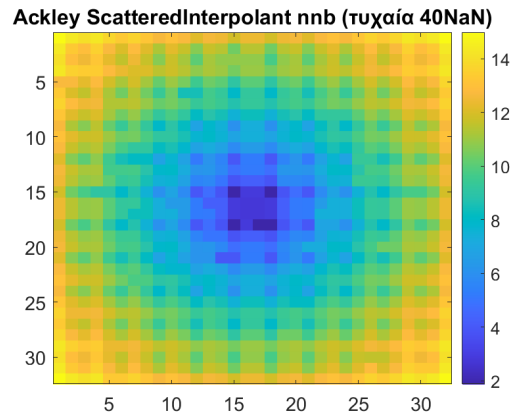
Το συμπέρασμα στο οποίο καταλήξαμε, μπορεί να εξακριβωθεί στη συνέχεια με τον υπολογισμό των μέτρων διασταυρωτικής επιβεβαίωσης.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



Σχήμα 5.1: (i): Τυχαία κατανομή 40 κενών εικονοστοιχείων. Αποτελέσματα από την εντολή griddata. Μέθοδος παρεμβολής: (ii): επιλογή nearest neighbor, (iii): επιλογή v4, (iv): επιλογή natural neighbor, (v): επιλογή cubic, (vi): επιλογή linear. Για λόγους συντομίας αναφερόμαστε στην επιλογή nearest neighbor με την ονομασία nnb (nearest neighbor interpolation).

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

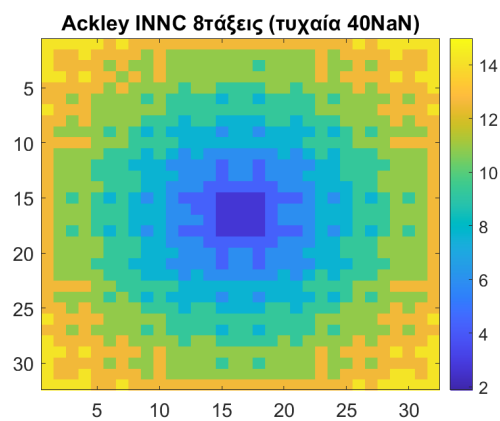


Σχήμα 5.2: Αποτελέσματα από την εντολή `scatteredInterpolant` για τυχαία κατανομή 40 κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (i): επιλογή `nearest neighbor`, (ii): επιλογή `natural neighbor`, (iii): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nnb` (`nearest neighbor interpolation`).

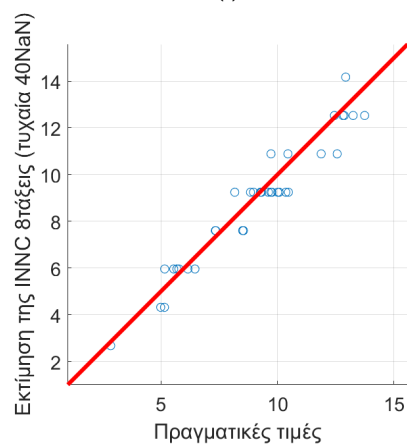
5.1. Αποτελέσματα για Συνθετικά Δεδομένα

Τέλος, στο Σχήμα 5.3i απεικονίζεται το αποτέλεσμα από την εφαρμογή της μεθόδου INNC, που έχει συμπληρώσει επιτυχώς τα κενά. Η μέθοδος έχει οριστεί με οχτώ τάξεις, με παραπάνω από οχτώ τάξεις δεν λειτουργεί σύμφωνα με το προγραμματιστικό περιβάλλον matlab. Επίσης, παρουσιάζονται τα διαγράμματα διασποράς όπου στον οριζόντιο άξονα εμφανίζονται οι αρχικές τιμές των συνθετικών δεδομένων και στον κατακόρυφο άξονα οι τιμές που έχουν εκτιμηθεί από τη μέθοδο INNC, για το Σχήμα 5.3ii. Τα κυκλάκια δηλώνουν τις τάξεις, οι οποίες έχουν οριστεί 8, όσα είναι κι αυτά. Όπως βλέπουμε στο διάγραμμα όλα τα σημεία είναι κοντά σε αυτή τη διαγώνια γραμμή και κάποια από αυτά είναι πάνω σε αυτή τη διαγώνιο που μας δείχνει ότι έχει γίνει καλή προσπάθεια όσον αφορά την πρόβλεψη.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



(i)



(ii)

Σχήμα 5.3: (i): Μέθοδος INNC 8 τάξεις. (ii): Διάγραμμα διασποράς πρόβλεψης 40 κενών εικονοστοιχείων.

5.1.2 Συνεχής Περιοχή 303 Κενών Εικονοστοιχείων

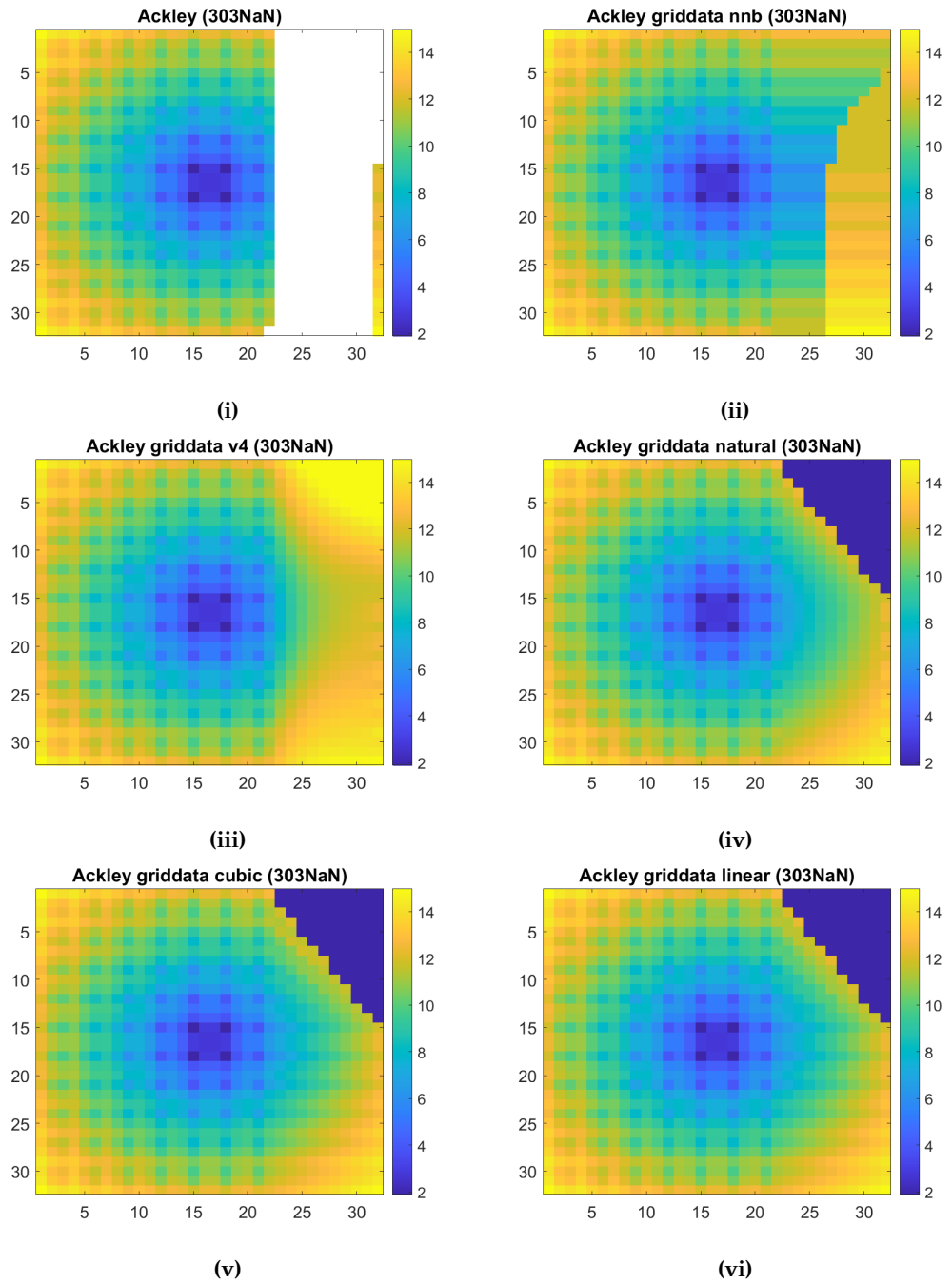
Όπως αναφέραμε και προηγουμένως, λάβαμε υπόψιν το ποσοστό επί των συνολικών σημείων του πλέγματος που υπάρχουν στα δορυφορικά δεδομένα, πιο συγκεκριμένα στα προσομοιωμένα, έτσι ώστε να παρατηρήσουμε τι αποτελέσματα θα έχουμε σε αυτά τα δεδομένα με τις ίδιες μεθόδους. Στο Σχήμα 5.4i έχουμε τοποθετήσει σε συνεχόμενη περιοχή 303 κενά εικονοστοιχεία. Τονίσαμε προηγουμένως ότι τα δεδομένα τα οποία έχουν κενά, οι θέσεις τους βρίσκονται στην τελευταία στήλη. Ωστόσο έχουμε σημειώσει ότι εκεί δεν υπάρχουν κενά, σύμφωνα με το προγραμματιστικό περιβάλλον matlab. Όπως φαίνεται λοιπόν στο Σχήμα 5.4i, στον κατακόρυφο άξονα 15 έως 32 και στον οριζόντιο άξονα 31 έως 32 υπάρχουν δεδομένα ενώ δεν θα έπρεπε και αυτό συμβαίνει επειδή αντισταθμίζονται τα δεδομένα κατά μισό εικονοστοιχείο (pixel) και απορρίπτουν μία σειρά και μία στήλη δεδομένων [4].

Αρχικά, αξιοποιήσαμε τις αιτιοκρατικές μεθόδους με όλες τις μεθόδους παρεμβολής που διαθέτουν για να μπορέσουμε να επιλέξουμε ποια είναι η πιο κατάλληλη για τη συμπλήρωση των κενών. Στο παρακάτω Σχήμα 5.4ii έχει εφαρμοστεί η μέθοδος των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `nearest neighbor`). Η ανακατασκευή που παράχθηκε, παρατηρούμε ότι έχει διαφορετική στρατηγική στον τρόπο με τον οποίο συμπληρώνονται τα κενά. Η μέθοδος δεν φαίνεται να έχει επιτευχθεί σωστά. Η εντολή `griddata` διαθέτει επίσης τη μέθοδο παρεμβολής ελάχιστης καμπυλότητας (επιλογή `v4`). Το αποτέλεσμα της εμφανίζεται στο Σχήμα 5.4iii, με το κίτρινο χρώμα ιδίως στις άκρες της δειγματοληπτημένης περιοχής να υπερισχύει. Στη μέση της επιφάνειας φαίνεται όμως να λαμβάνει υπόψιν την τιμή του γειτονικού σημείου.

Με βάση τα Σχήματα 5.4iv, 5.4v και 5.4vi συμπεραίνουμε ότι δεν είναι κατάλληλες αυτές οι μεθοδολογίες παρεμβολής που εφαρμόζονται. Χρησιμοποιώντας τη μέθοδο παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `natural neighbor`), τη μέθοδο κυβικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή

cubic) και τη μέθοδο γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `linear`) παρατηρούμε ότι στην πάνω άκρη της δειγματοληπτημένης περιοχής δεν έχει γίνει πλήρωση των κενών. Αυτό παρατηρείται διότι η εντολή `griddata` έχει μια ιδιομορφία. Πιο συγκεκριμένα, κατασκευάζει ένα κυρτό φλοιό ο οποίος ορίζεται από το σύνολο των σημείων δειγματοληψίας και γεμίζει μόνο μέσα στο κυρτό φλοιό. Στο Σχήμα 5.4iv, η μπλε συνεχής γραμμή (πολυγωνική) είναι ο κυρτός φλοιός. Αυτός σχηματίζει ένα κυρτό πολύγωνο στο οποίο περιέχει όλα τα σημεία του δείγματος. Αυτό σημαίνει ότι στα σημεία έξω από το κυρτό πολύγωνο, η εντολή `griddata` θα λαμβάνει NaN.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



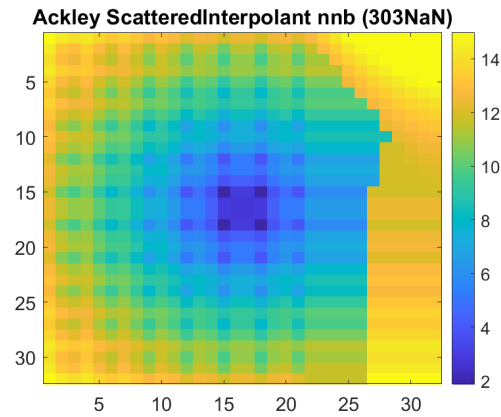
Σχήμα 5.4: (i): Συνεχής περιοχή 303 κενών εικονοστοιχείων. Αποτελέσματα από την εντολή griddata για συνεχή περιοχή 303 κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (ii): επιλογή nearest neighbor, (iii): επιλογή v4, (iv): επιλογή natural neighbor, (v): επιλογή cubic, (vi): επιλογή linear. Για λόγους συντομίας αναφερόμαστε στην επιλογή nearest neighbor με την ονομασία nnb (nearest neighbor interpolation).

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

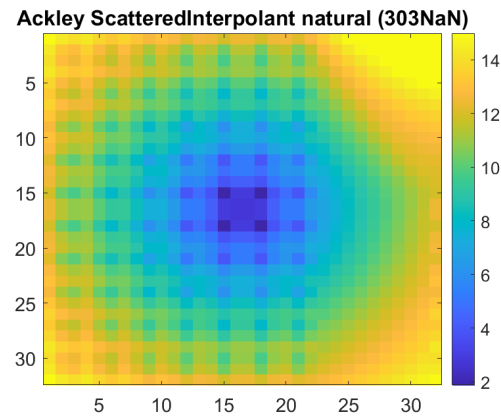
Έπειτα, ακολουθεί το Σχήμα 5.5 που παρουσιάζει τα αποτελέσματα της εντολής `scatteredInterpolant`, χρησιμοποιώντας τις μεθόδους παρεμβολής που διαθέτει. Στο Σχήμα 5.5i, η μέθοδος των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `nearest neighbor`), δεν φαίνεται να έχει εκτιμήσει σωστά τις τιμές στη δειγματοληπτημένη περιοχή. Επιπλέον, πραγματοποιήσαμε ανακατασκευή με τη βοήθεια της μεθόδου παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `natural neighbor`), όπως φαίνεται στο Σχήμα 5.5ii. Με τη μέθοδο αυτή παρατηρούμε ότι έχει γίνει καλύτερα ανακατασκευή, δηλαδή στον κατακόρυφο άξονα από 10 έως 25 και στον οριζόντιο από 22 έως 30 έχει συμβάλει στη συμπλήρωση κενών καλύτερα σε σχέση με τη μέθοδο των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `nearest neighbor`). Το ίδιο απεικονίζεται στο Σχήμα 5.5iii, με τη μέθοδο γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `linear`) με ελάχιστη διαφορά συγκριτικά με το προηγούμενο Σχήμα 5.5ii.

Στο Σχήμα 5.6i εφαρμόσαμε τη μέθοδο INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , η ανακατασκευή φαίνεται να έχει τροποποιήσει αρκετά την περιοχή που διέθετε πριν τα κενά. Επίσης, σύμφωνα με το προγραμματιστικό περιβάλλον `matlab`, η μέθοδος εστίασε και σε άλλα σημεία τα οποία δεν έχουν κενά. Αυτό το ελέγξαμε μέσω του τελικού πίνακα που διέθετε τα ανακατασκευασμένα δεδομένα. Πιο συγκεκριμένα, σημεία εκτός από την δειγματοληπτημένη περιοχή είχαν διαφοροποιηθεί οι τιμές τους. Επιπλέον, χρησιμοποιήσαμε τη μέθοδο για 16 τάξεις, με το ίδιο τετραγωνικό πλαίσιο INNC, όπως φαίνεται στο Σχήμα 5.6ii. Η ανακατασκευή δεν φαίνεται να είναι σωστή διότι έχει καλυφθεί ένα μεγάλο μέρος των κενών με γαλάζιο χρώμα. Να επισημανθεί ότι υπάρχει κι εδώ διαφοροποίηση στις τιμές ακόμα και σε σημεία όπου προηγουμένως δεν υπήρχαν κενά.

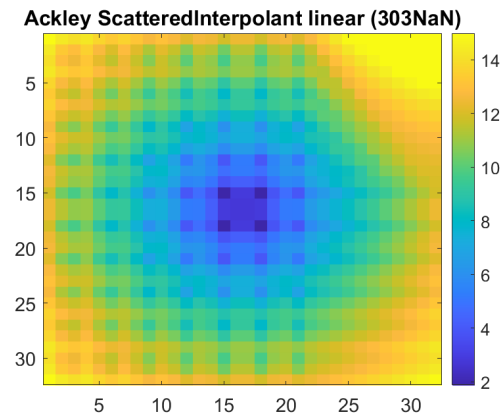
5.1. Αποτελέσματα για Συνθετικά Δεδομένα



(i)

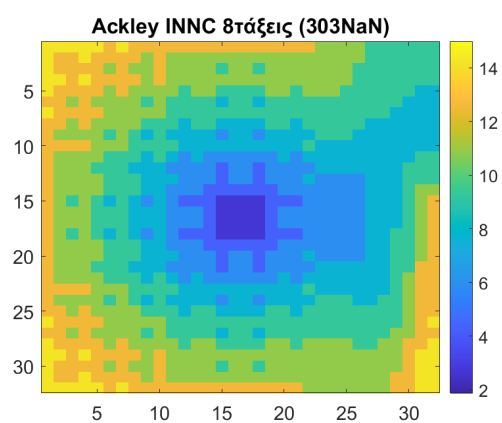


(ii)

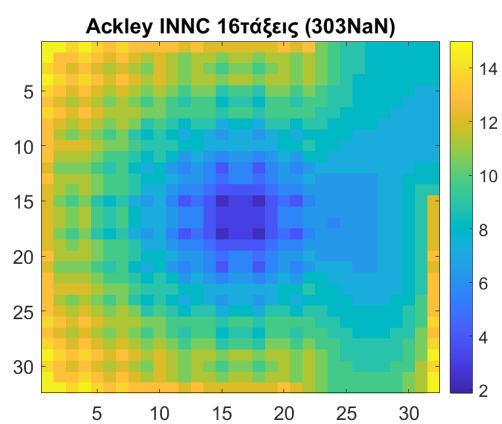


(iii)

Σχήμα 5.5: Αποτελέσματα από την εντολή `scatteredInterpolant` για συνεχή περιοχή 303 κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (i): επιλογή `nearest neighbor`, (ii): επιλογή `natural neighbor`, (iii): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nnb` (`nearest neighbor interpolation`). 60



(i)



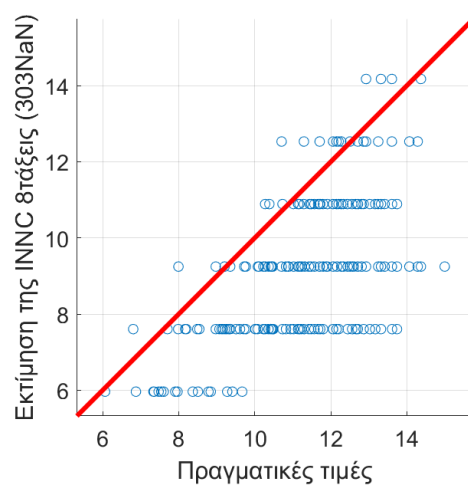
(ii)

Σχήμα 5.6: (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.

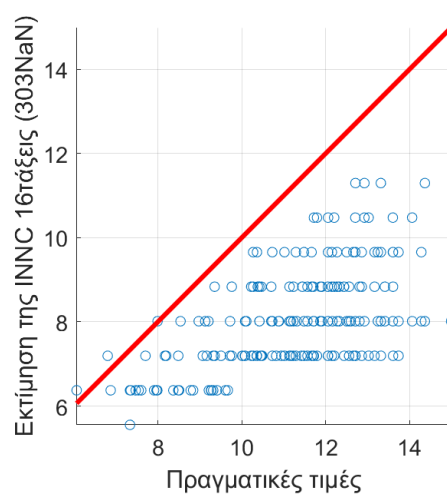
5.1. Αποτελέσματα για Συνθετικά Δεδομένα

Στη συνέχεια, παρουσιάζονται τα διαγράμματα διασποράς όπου στον οριζόντιο άξονα απεικονίζονται οι αρχικές τιμές των συνθετικών δεδομένων για τα Σχήματα 5.7i και 5.7ii και στον κατακόρυφο άξονα οι τιμές που έχουν εκτιμηθεί από τη μέθοδο INNC για οχτώ και 16 τάξεις αντίστοιχα. Αξιολογώντας τα δύο σχήματα παρατηρείται ότι με τη μέθοδο INNC για 16 τάξεις οι τιμές που έχουν εκτιμηθεί απέχουν πολύ σε σχέση με τις αρχικές τιμές των δεδομένων. Αυτό μπορούμε να το εξακριβώσουμε στη συνέχεια με τα μέτρα διασταυρωτικής επιβεβαίωσης. Σε σύγκριση με το Σχήμα 5.7i συμπεραίνουμε ότι είναι προτιμότερες οι προβλεπόμενες τιμές που έχουν υπολογιστεί από τη μέθοδο INNC για οχτώ τάξεις. Σε αυτό το διάγραμμα διασποράς, λοιπόν τα περισσότερα σημεία είναι μακριά από αυτή τη διαγώνια γραμμή. Φαίνεται ότι στον άξονα με τις πραγματικές τιμές για μεγάλες τιμές έχουν εκτιμηθεί πιο χαμηλά οι τιμές από τη μέθοδο.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



(i)



(ii)

Σχήμα 5.7: Διαγράμματα διασποράς πρόβλεψης 303 κενών εικονοστοιχείων σε συνεχή περιοχή. (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

Στη συνέχεια, λόγω της αντιστάθμισης δεδομένων κατά μισό εικονοστοιχείο και απόρριψης της τελευταίας στήλης των δεδομένων, εξετάσαμε τα ίδια δεδομένα τοποθετώντας τα 303 κενά στο κέντρο, όπως φαίνεται στο Σχήμα 5.8i. Αυτή την περίπτωση, τη λάβαμε υπόψιν προκειμένου να δούμε πως συμπεριφέρονται οι μεθοδολογίες όταν τα δεδομένα σε συνεχόμενη περιοχή βρίσκονται σε άλλο σημείο της εικόνας.

Αρχικά, εφαρμόσαμε τις αιτιοκρατικές μεθόδους με όλες τις μεθόδους παρεμβολής που διαθέτουν για να μπορέσουμε να επιλέξουμε ποια είναι η καταλληλότερη για τη συμπλήρωση των κενών. Στο Σχήμα 5.8ii έχει χρησιμοποιηθεί η μέθοδος των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `nearest neighbor`). Η ανακατασκευή που έχει δημιουργηθεί, συμπεραίνουμε ότι διαθέτει διαφορετική στρατηγική στον τρόπο με τον οποίο συμπληρώνονται τα κενά. Η μέθοδος δεν φαίνεται να έχει επιτευχθεί σωστά. Πιο συγκεκριμένα, στον οριζόντιο άξονα 15 με 22 και στον κατακόρυφο άξονα 15 με 20, έχει υπερισχύσει το χρώμα κίτρινο ενώ διακρίνουμε ότι το χρώμα που υπάρχει γύρω από αυτή την περιοχή είναι το γαλάζιο. Η μέθοδος παρεμβολής ελάχιστης καμπυλότητας (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `v4`). Το αποτέλεσμα της εμφανίζεται στο Σχήμα 5.8iii, με το κίτρινο χρώμα να υπερισχύει σε περιοχή που δεν έχει οριστεί ως δειγματοληπτημένη. Επομένως, η μέθοδος έχει συμβάλλει στην ανακατασκευή και της δεξιάς περιοχής της εικόνας ενώ δεν χρειαζόταν διότι δεν διέθετε κενά στη συγκεκριμένη περιοχή.

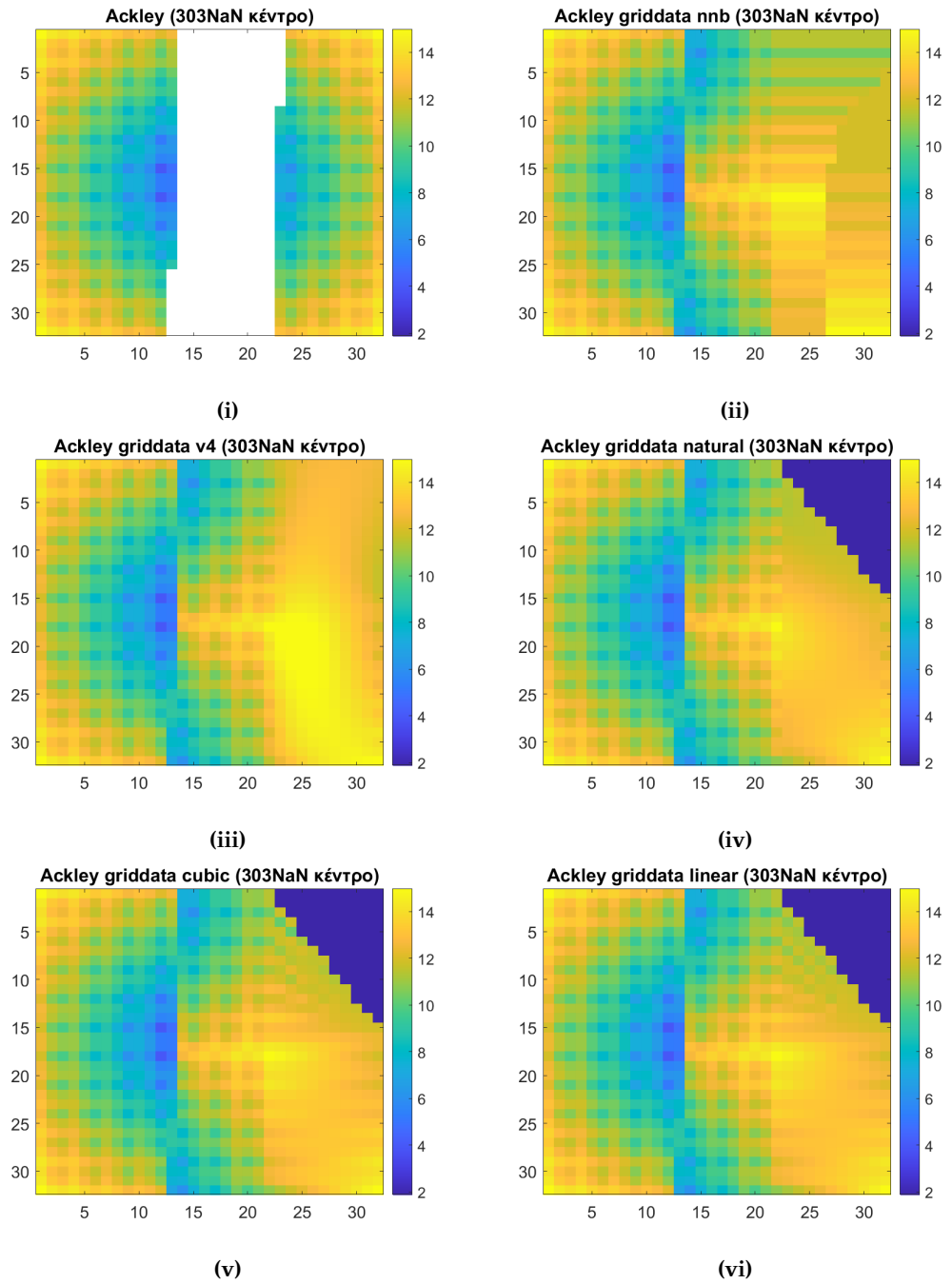
Επίσης, τα Σχήματα 5.8iv, 5.8v και 5.8vi καταλήγουμε στο συμπέρασμα ότι δεν είναι κατάλληλες αυτές οι μεθοδολογίες παρεμβολής που εφαρμόζονται. Αξιοποιώντας τη μέθοδο παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `natural neighbor`), τη μέθοδο κυβικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `cubic`) και τη μέθοδο γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `linear`) παρατηρούμε ότι στην πάνω άκρη της μη δειγματοληπτημένης περιοχής έχουν εμφανιστεί κενά ενώ πριν δεν υπήρχαν στο συγκεκριμένο σημείο.

Έπειτα, ακολουθεί το Σχήμα 5.9 που παρουσιάζει τα αποτελέσμα-

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

τα της εντολής `scatteredInterpolant`, εφαρμόζοντας τις μεθόδους παρεμβολής που διαθέτει. Στο Σχήμα 5.9i, η μέθοδος των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `nearest neighbor`), λαμβάνει την ίδια τιμή με το πλησιέστερο δείγμα των δεδομένων και με αυτό τον τρόπο συμπληρώνει τα κενά. Τα αποτελέσματα στο σχήμα αυτό είναι καλύτερα από αυτά του Σχήματος 5.8ii. Επίσης, πραγματοποιήσαμε ανακατασκευή με τη βοήθεια της μεθόδου παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `natural neighbor`), όπως φαίνεται στο Σχήμα 5.9ii. Με τη μέθοδο αυτή παρατηρούμε ότι έχει γίνει η συμπλήρωση των κενών με διαφορετικό τρόπο σε σύγκριση με τη προηγούμενη επιλογή `nearest neighbor`, επειδή το αποτέλεσμα που προκύπτει είναι αρκετά θολό. Στο Σχήμα 5.9iii, με τη μέθοδο γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `linear`) έχει επιτευχθεί ο στόχος της πλήρωσης των κενών. Αυτό μπορούμε να το επαληθεύσουμε στη συνέχεια με τα μέτρα διασταυρωτικής επιβεβαίωσης.

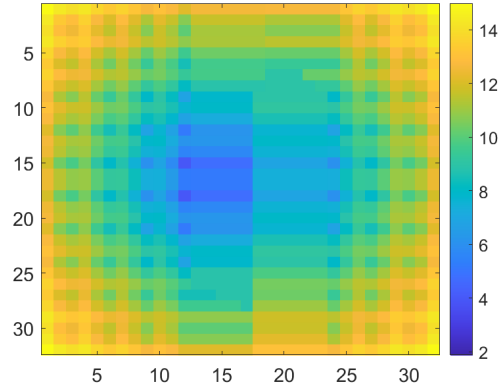
5.1. Αποτελέσματα για Συνθετικά Δεδομένα



Σχήμα 5.8: (i): Συνεχής περιοχή 303 κενών εικονοστοιχείων στο κέντρο. Αποτελέσματα από την εντολή griddata για συνεχή περιοχή 303 κενών εικονοστοιχείων στο κέντρο. Μέθοδος παρεμβολής: (ii): επιλογή nearest neighbor, (iii): επιλογή v4, (iv): επιλογή natural neighbor, (v): επιλογή cubic, (vi): επιλογή linear. Για λόγους συντομίας αναφερόμαστε στην επιλογή nearest neighbor με την ονομασία nnb (nearest neighbor interpolation).

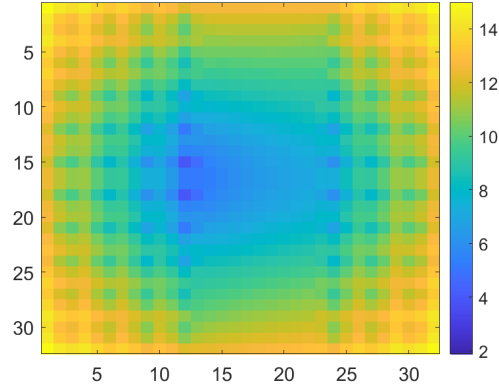
5.1. Αποτελέσματα για Συνθετικά Δεδομένα

Ackley ScatteredInterpolant nnb (303NaN κέντρο)



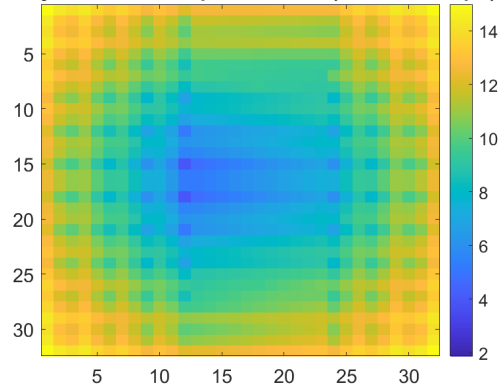
(i)

Ackley ScatteredInterpolant natural (303NaN κέντρο)



(ii)

Ackley ScatteredInterpolant linear (303NaN κέντρο)



(iii)

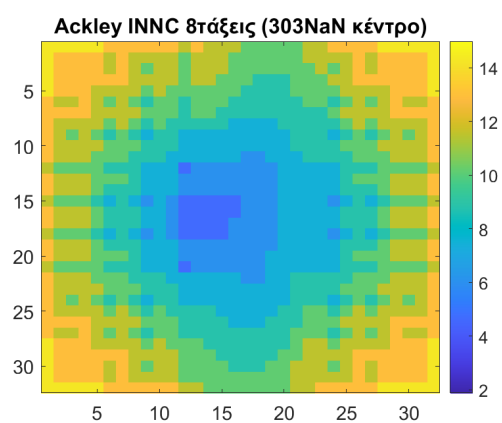
Σχήμα 5.9: Αποτελέσματα από την εντολή `scatteredInterpolant` για συνεχή περιοχή 303 κενών εικονοστοιχείων στο κέντρο. Μέθοδος παρεμβολής: (i): επιλογή `nearest neighbor`, (ii): επιλογή `natural neighbor`, (iii): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nnb` (`nearest neighbor interpolation`).

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

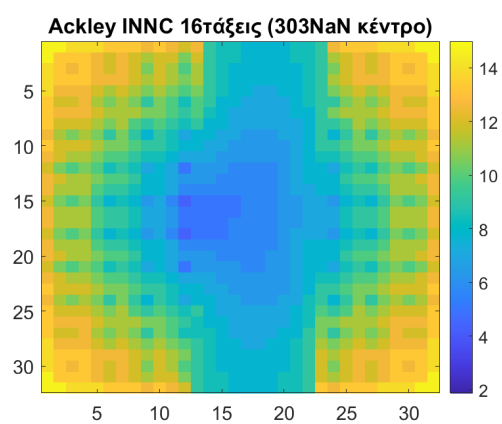
Στη συνέχεια, προβάλλουμε τα αποτελέσματα της μεθόδου INNC με δύο διαφορετικές τάξεις. Στο Σχήμα 5.10i έχουμε τη μέθοδο με παράμετρο οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 30×30 , παρατηρούμε ότι η συμπλήρωση των κενών δεν είναι ακριβής και δεν φαίνεται να λειτουργεί σωστά η μέθοδος σε αυτή την περίπτωση. Επιπλέον, αξιοποιούμε τη μέθοδο με παράμετρο 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 προκειμένου να δούμε πως συμπεριφέρεται με μεγαλύτερη τάξη. Τα αποτελέσματα, όπως φαίνονται στο Σχήμα 5.10ii, δεν ικανοποιούν το ζητούμενο και πάλι εντοπίζουμε ότι επικρατεί σε κάποιες περιοχές το μπλε χρώμα ενώ δεν φαίνεται αυτό να επικρατεί γύρω από την περιοχή. Για παράδειγμα, στον οριζόντιο άξονα από 15 έως 20 και στον κατακόρυφο από 5 έως 10, η πληροφορία που υπάρχει, όσον αφορά το χρώμα είναι το πράσινο αλλά η μέθοδος επιλέγει το μπλε χρώμα. Συμπεραίνουμε λοιπόν, ότι η μέθοδος σε αυτή την περίπτωση, με τα κενά τοποθετημένα σε συνεχόμενη περιοχή στο κέντρο, δεν λειτουργεί ικανοποιητικά για τη συμπλήρωση των κενών ούτε για οχτώ ούτε για 16 τάξεις.

Τέλος μετά την εφαρμογή της μεθόδου INNC για οχτώ και 16 τάξεις, μέσω του προγραμματιστικού περιβάλλοντος matlab κατασκευάσαμε τα διαγράμματα διασποράς. Σύμφωνα με τα Σχήματα 5.11i και 5.11ii, στον οριζόντιο άξονα απεικονίζονται οι αρχικές τιμές των συνθετικών δεδομένων και στον κατακόρυφο άξονα οι τιμές που έχουν εκτιμηθεί από τη μέθοδο INNC για οχτώ και 16 τάξεις αντίστοιχα. Σύμφωνα με το Σχήμα 5.11i στην αρχή παρατηρούμε ότι για μικρές τιμές η πρόβλεψη που έχει γίνει από τη μέθοδο INNC απέχει πολύ από τις πραγματικές τιμές και αυτό το συμπεραίνουμε από το γεγονός ότι τα σημεία είναι μακριά από τη διαγώνιο και όσο μεγαλώνουν οι τιμές φαίνεται να μην υπάρχει κάποια βελτίωση στην εκτίμηση των τιμών καθώς έχουν εκτιμηθεί πιο χαμηλά οι τιμές σε σύγκριση με τις πραγματικές τιμές. Παρατηρούμε ότι με τη μέθοδο INNC για 16 τάξεις οι τιμές που έχουν υπολογιστεί, δεν αντιστοιχούν σε μεγάλο βαθμό με τις αρχικές τιμές των δεδομένων. Αυτό μπορούμε να το επαληθεύσουμε στη συνέχεια με τα μέτρα διασταυρωτικής επιβεβαίωσης.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



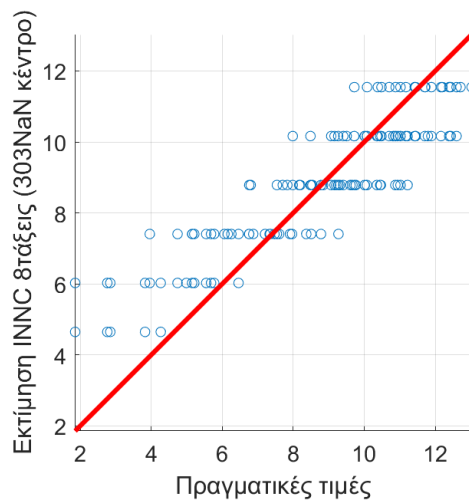
(i)



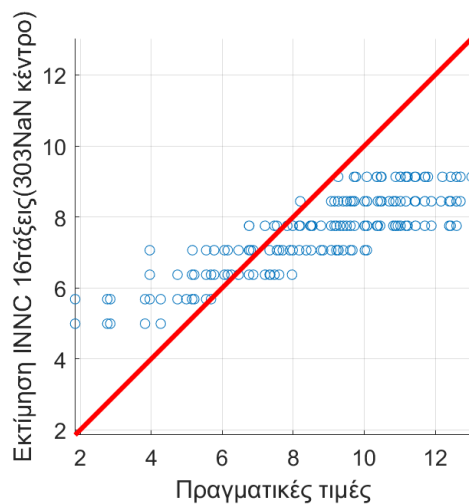
(ii)

Σχήμα 5.10: (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



(i)



(ii)

Σχήμα 5.11: Διαγράμματα διασποράς πρόβλεψης 303 κενών εικονοστοιχείων σε συνεχή περιοχή στο κέντρο. (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.

5.1.3 Τυχαία Κατανομή 303 Κενών Εικονοστοιχείων

Στο παρακάτω σχήμα, τα κενά είναι τόσα όσο στο προηγούμενο σχήμα, δηλαδή 303, με τη διαφορά ότι η κατανομή έχει γίνει τυχαία, όπως φαίνεται στο Σχήμα 5.12i. Ο λόγος που τοποθετήθηκαν τα κενά σε τυχαία σημεία, είναι για να παρατηρήσουμε τι αποτελέσματα προκύπτουν σε αυτή την περίπτωση με τις ίδιες μεθόδους που χρησιμοποιήσαμε πριν.

Αρχικά, εφαρμόσαμε τις αιτιοκρατικές μεθόδους στα συνθετικά δεδομένα. Στο Σχήμα 5.12ii χρησιμοποιήσαμε τη μέθοδο των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `nearest neighbor`) και διακρίναμε ότι η διαδικασία της ανακατασκευής της εικόνας είναι επιτυχής. Το ίδιο συμπεραίνουμε για τη μέθοδο παρεμβολής ελάχιστης καμπυλότητας (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `v4`), όπως φαίνεται στο Σχήμα 5.12iii. Οι υπόλοιπες μέθοδοι παρεμβολής που παρέχει η εντολή δεν είναι κατάλληλες. Με τη μέθοδο παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `natural neighbor`), Σχήμα 5.12iv, τη μέθοδο κυβικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `cubic`), Σχήμα 5.12v και τη μέθοδο γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `linear`), Σχήμα 5.12vi διακρίνουμε ότι στην άκρη πάνω δεξιά δεν έχει γίνει πλήρωση των κενών. Επομένως, δεν λειτουργούν σωστά για την παρεμβολή διάσκορων δεδομένων εκτός από τη μέθοδο πλησιέστερων γειτόνων και τη μέθοδο παρεμβολής ελάχιστης καμπυλότητας.

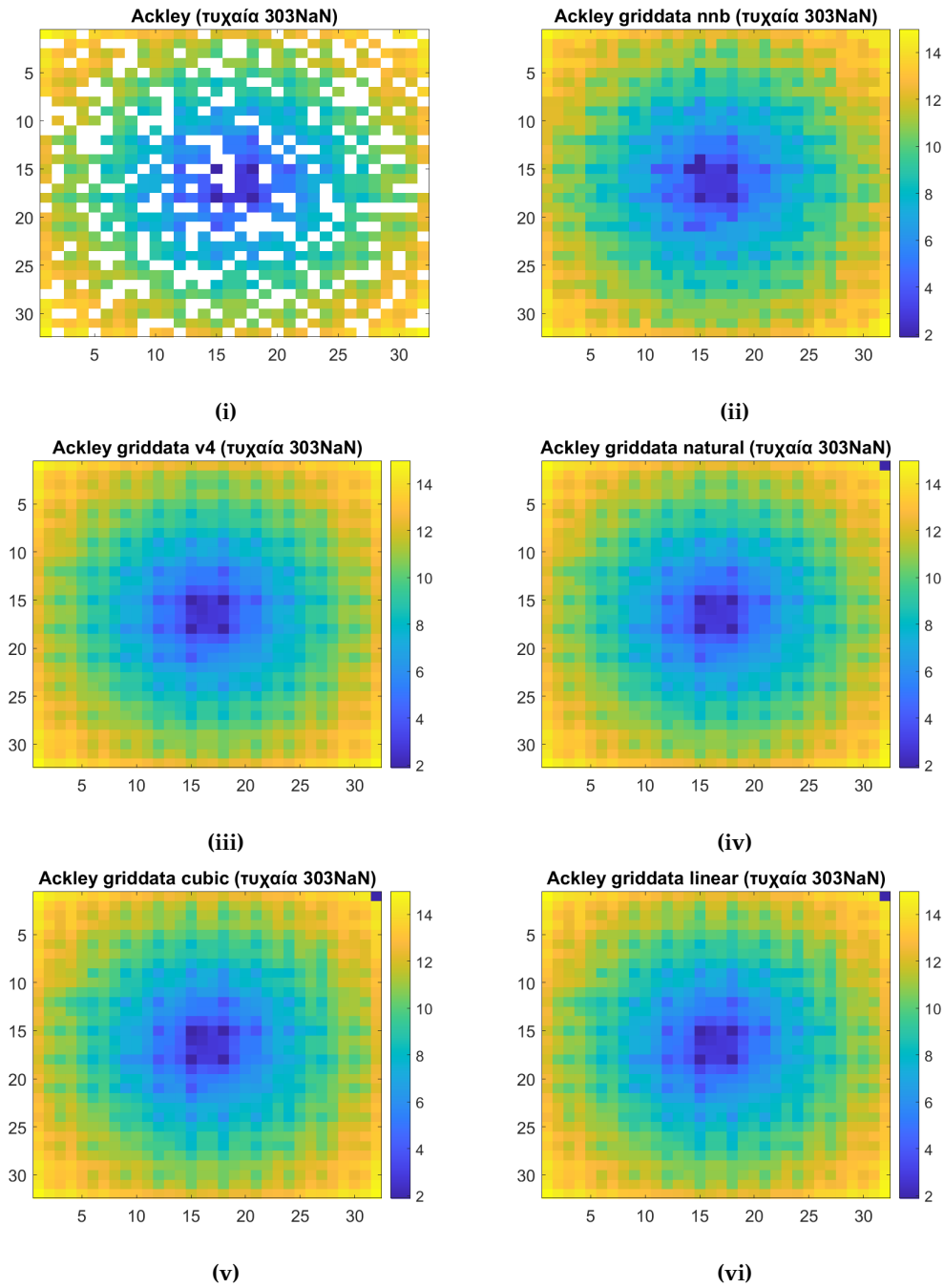
Επιπροσθέτως, διερευνήσαμε τη μέθοδο των πλησιέστερων γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `nearest neighbor`), μέθοδο παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `natural neighbor`) και τη μέθοδο γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `linear`) για τα Σχήματα 5.13i, 5.13ii και 5.13iii αντίστοιχα. Τα αποτελέσματα που προέκυψαν, έχουν ανακατασκευάσει την αρχική εικόνα και άρα, έχει επιτευχθεί ο στόχος της πλήρωσης των κενών. Η μέθοδος των πλησιέστερων γειτόνων (υλοποιή-

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

θηκε μέσω της εντολής `griddata` και μέσω της `scatteredInterpolant`) παρατηρείται ότι παρέχει παρόμοια ανακατασκευή και με τις δύο εντολές.

Είναι αξιοσημείωτο να αναφέρουμε ότι με τη βοήθεια και των μέτρων διασταυρωτικής επιβεβαίωσης θα μπορέσουμε να έχουμε καλύτερη εκτίμηση των μεθόδων.

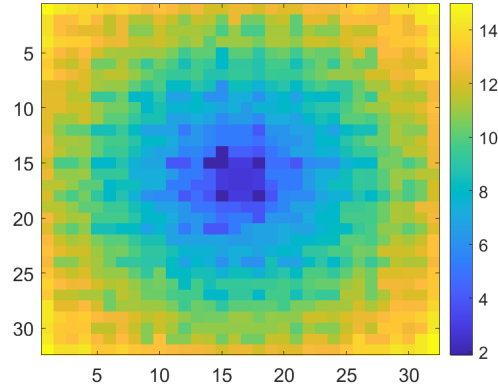
5.1. Αποτελέσματα για Συνθετικά Δεδομένα



Σχήμα 5.12: (i): Τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Αποτελέσματα από την εντολή `griddata` για τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (ii): επιλογή `nearest neighbor`, (iii): επιλογή `v4`, (iv): επιλογή `natural neighbor`, (v): επιλογή `cubic`, (vi): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nnb` (`nearest neighbor interpolation`).

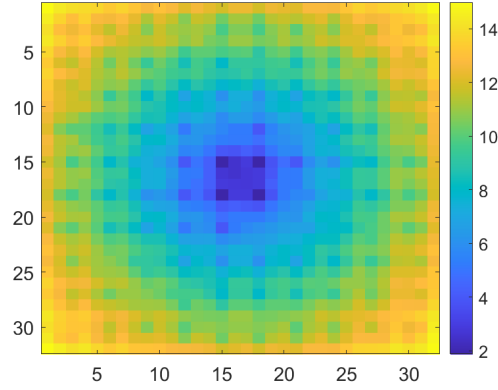
5.1. Αποτελέσματα για Συνθετικά Δεδομένα

Ackley ScatteredInterpolant nnb (τυχαία 303NaN)



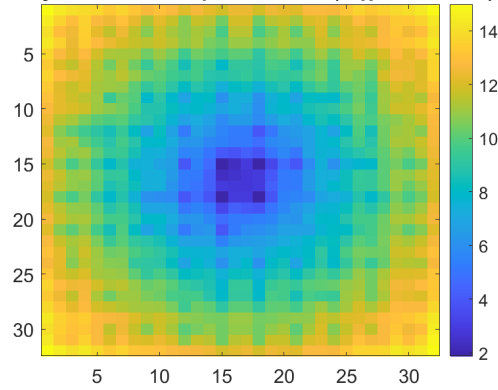
(i)

Ackley ScatteredInterpolant natural (τυχαία 303NaN)



(ii)

Ackley ScatteredInterpolant linear (τυχαία 303NaN)



(iii)

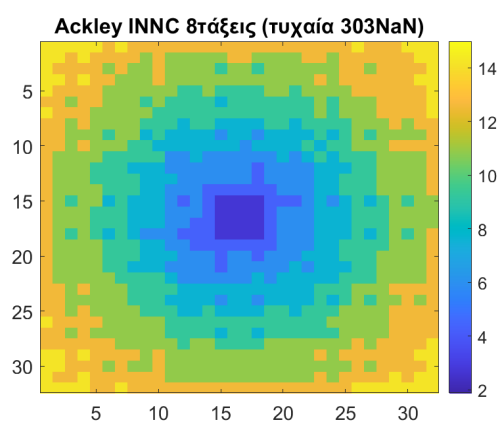
Σχήμα 5.13: Αποτελέσματα από την εντολή `scatteredInterpolant` για τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Μέθοδος παρεμβολής: (i): επιλογή `nearest neighbor`, (ii): επιλογή `natural neighbor`, (iii): επιλογή `linear`. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nnb` (`nearest neighbor interpolation`). 74

5.1. Αποτελέσματα για Συνθετικά Δεδομένα

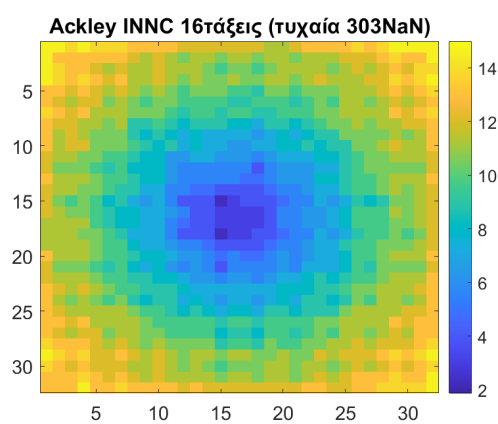
Στη συνέχεια, παρουσιάζουμε τα αποτελέσματα της μεθόδου INNC με δύο διαφορετικές τάξεις. Στο Σχήμα 5.14i έχει χρησιμοποιηθεί η μέθοδος με παράμετρο οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 5×5 . Όπως φαίνεται από το σχήμα, η συμπλήρωση των κενών που έχει συμβεί, έχει αλλοιώσει την εικόνα σε σύγκριση με την αρχική εικόνα 3.1i με αποτέλεσμα να μην λειτουργεί σωστά η μέθοδος σε αυτή την περίπτωση. Επίσης, εφαρμόζουμε την ίδια μέθοδο με παράμετρο 16 τάξεις και τετραγωνικό πλαίσιο INNC 5×5 προκειμένου να δούμε πως συμπεριφέρεται με μεγαλύτερη τάξη. Τα αποτελέσματα που απεικονίζονται στο Σχήμα 5.14ii, υποδεικνύουν ότι δεν έχει λειτουργήσει σωστά η μέθοδος. Άρα, συμπεραίνουμε ότι η μέθοδος σε αυτή την περίπτωση της τυχαίας κατανομής των κενών, δεν λειτουργεί ικανοποιητικά για τη συμπλήρωση των κενών ούτε για οχτώ ούτε για 16 τάξεις.

Τέλος μέσω του προγραμματιστικού περιβάλλοντος matlab κατασκευάσαμε τα διαγράμματα διασποράς. Σύμφωνα με τα Σχήματα 5.15i και 5.15ii, στον οριζόντιο άξονα απεικονίζονται οι αρχικές τιμές των συνθετικών δεδομένων και στον κατακόρυφο άξονα οι τιμές που έχουν αξιολογηθεί από τη μέθοδο INNC για οχτώ και 16 τάξεις αντίστοιχα. Παρατηρούμε ότι με τη μέθοδο INNC για οχτώ τάξεις οι τιμές που έχουν υπολογιστεί, δεν αντιστοιχούν σε μεγάλο βαθμό με τις αρχικές τιμές των δεδομένων. Για παράδειγμα, στον οριζόντιο άξονα η τιμή 10 που αντιπροσωπεύει την αρχική τιμή, αντιστοιχεί στις προβλεπόμενες τιμές από τη μέθοδο INNC 9 και 11 περίπου. Αυτό μπορούμε να το επαληθεύσουμε στη συνέχεια με τα μέτρα διασταυρωτικής επιβεβαίωσης.

Σε όλα τα σχήματα που αφορούν τα αποτελέσματα των συνθετικών δεδομένων που έχουν παραχθεί από τη συνάρτηση Ackley, είναι σημαντικό να σημειωθεί ότι το εύρος τιμών που παίρνουν τα δεδομένα αυτά είναι από το 2 έως το 14.



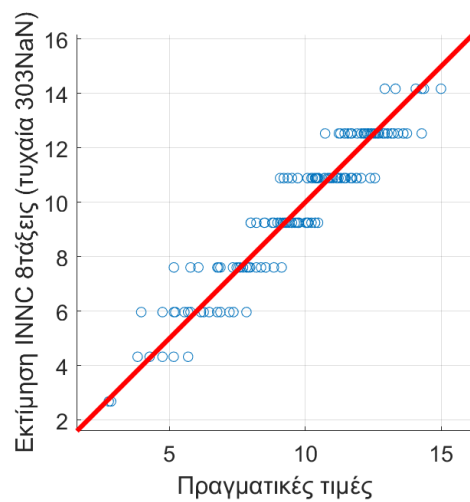
(i)



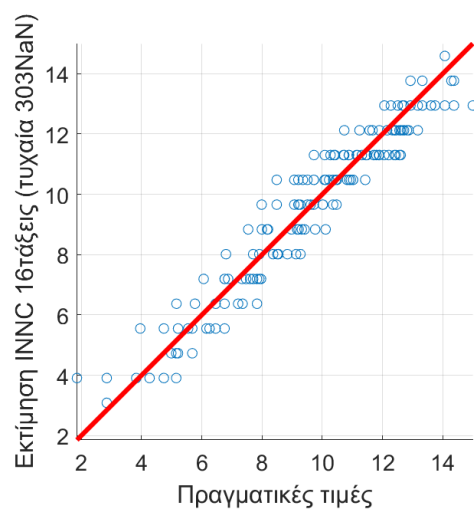
(ii)

Σχήμα 5.14: (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.

5.1. Αποτελέσματα για Συνθετικά Δεδομένα



(i)



(ii)

Σχήμα 5.15: Διαγράμματα διασποράς πρόβλεψης ισάριθμων κενών εικονοστοιχείων σε τυχαία κατανομή. (i): Μέθοδος INNC 8 τάξεις. (ii): Μέθοδος INNC 16 τάξεις.

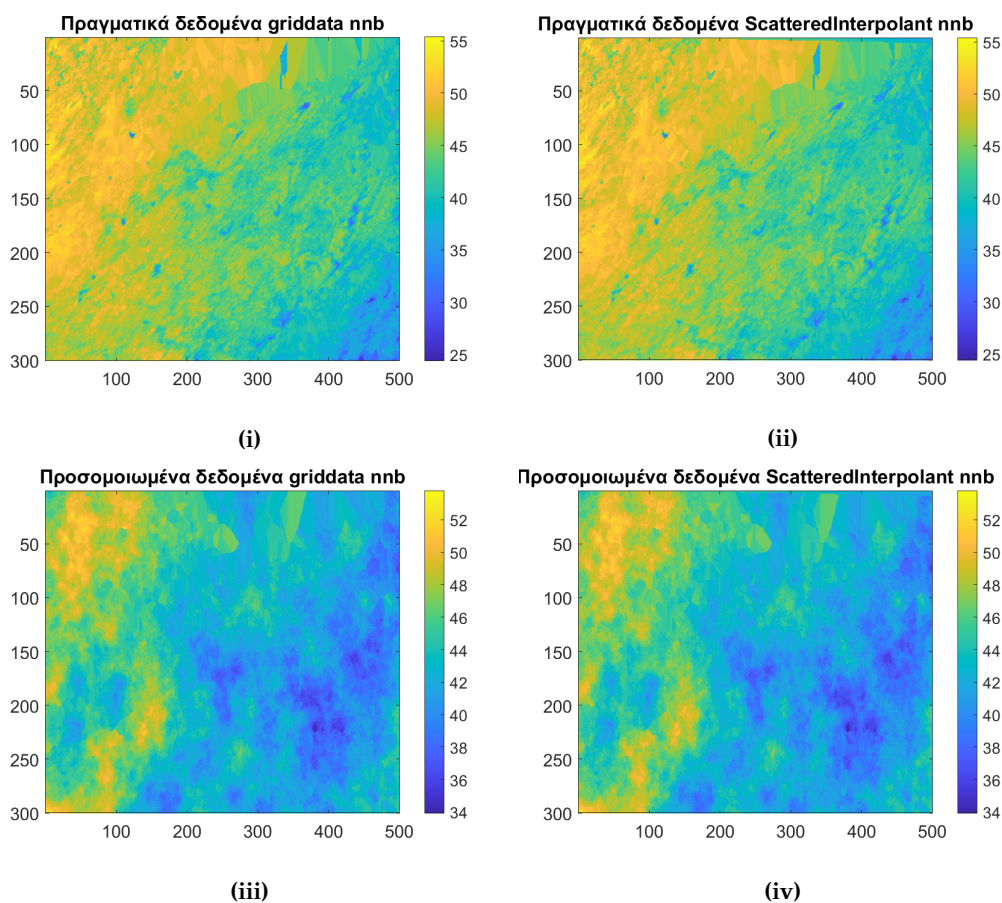
5.2 Αποτελέσματα για Δορυφορικά Δεδομένα

Στην ενότητα αυτή, αφού υλοποιήσαμε τις αιτιοκρατικές μεθόδους και τη μέθοδο του μοντέλου Ising με συσχετίσεις κοντινότερων γειτόνων (INNC) με σκοπό την παροχή προβλέψεων σε συγκεκριμένες θέσεις των δορυφορικών δεδομένων, προχωρήσαμε στην αξιολόγηση αυτών των μεθόδων. Επομένως, με βάση τις εικόνες που παράχθηκαν, χρησιμοποιώντας το προγραμματιστικό περιβάλλον matlab, θα αναλύσουμε τα αποτελέσματα σε σύγκριση με την αρχική εικόνα. Έτσι, για καθεμία μέθοδο θα διαπιστώσουμε εάν έγινε επιτυχώς η πλήρωση των κενών.

5.2.1 Αιτιοκρατικές Μέθοδοι

Στο επόμενο σχήμα εφαρμόζονται για τα ίδια δεδομένα οι αιτιοκρατικές μέθοδοι που παρουσιάσαμε στο Κεφάλαιο 4. Συγκεκριμένα στο Σχήμα 5.16i και στο Σχήμα 5.16ii αξιολογήσαμε για τα πραγματικά δεδομένα τη μεθοδολογία των πλησιέστερων γειτόνων (επιλογή nearest neighbor) υλοποιήθηκε μέσω της εντολής griddata και μέσω της εντολής scatteredInterpolant αντίστοιχα. Τα αποτελέσματα που προκύπτουν είναι ικανοποιητικά, η μέθοδος εφαρμόστηκε επιτυχώς με αποτέλεσμα οι ανακατασκευασμένες εικόνες να μην εμφανίζουν κενά. Το ίδιο ισχύει και για τα Σχήματα 5.16iii και 5.16iv που είναι για προσομοιωμένα δεδομένα. Η πλήρωση κενών με τη βοήθεια των δύο αιτιοκρατικών αυτών μεθόδων έχει διεξαχθεί σωστά με ικανοποιητικά αποτελέσματα όπως φαίνεται παρακάτω.

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα



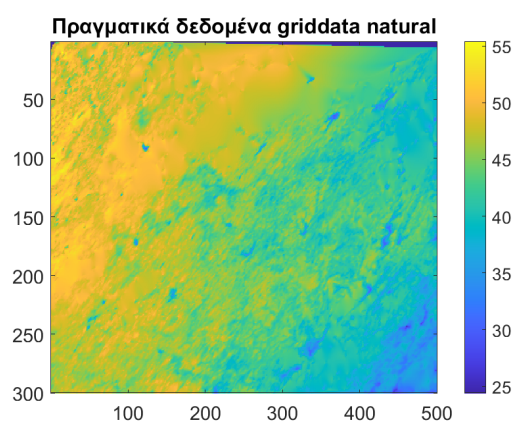
Σχήμα 5.16: Αποτελέσματα από την εντολή `griddata` με μέθοδο παρεμβολής επιλογή `nearest neighbor` (i): πραγματικά δεδομένα, (iii): προσομοιωμένα δεδομένα. Από την εντολή `scatteredInterpolant` με μέθοδο παρεμβολής επιλογή `nearest neighbor` (ii): πραγματικά δεδομένα, (iv): προσομοιωμένα δεδομένα. Για λόγους συντομίας αναφερόμαστε στην επιλογή `nearest neighbor` με την ονομασία `nnb` (`nearest neighbor interpolation`).

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα

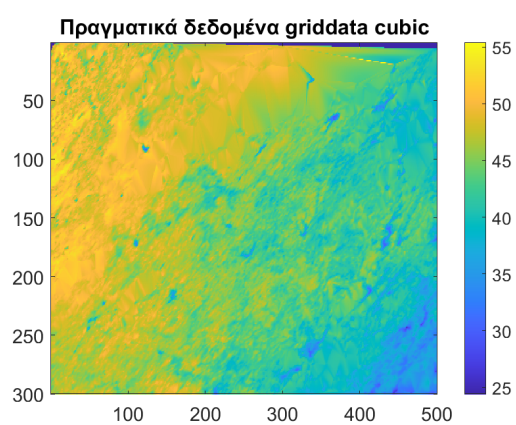
Έπειτα, αξιοποιήθηκε η εντολή `griddata` και με άλλες μεθόδους παρεμβολής, όπως φαίνεται στα Σχήματα 5.17 και 5.18. Με βάση το Σχήμα 5.17 απεικονίζονται οι μέθοδοι παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `natural neighbor`), κυβικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `cubic`) και γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `linear`) στα 5.17i, 5.17ii και 5.17iii αντίστοιχα, οι οποίες συμπεραίνουμε ότι δε συμβάλλουν στην πλήρωση των κενών. Σε όλες τις περιπτώσεις, διαπιστώνουμε κενά στο πάνω μέρος της εκάστοτε εικόνας. Όσον αφορά τα προσομοιωμένα δεδομένα στο Σχήμα 5.18, χρησιμοποιώντας τις ίδιες μεθοδολογίες παρεμβολής προκύπτει το ίδιο αποτέλεσμα με αυτό του προηγούμενου σχήματος.

Άρα, με την εντολή `griddata` εκτός από τη μέθοδο των πλησιέστερων γειτόνων (επιλογή `nearest neighbor`), οι υπόλοιπες μέθοδοι παρεμβολής δεν λειτουργούν ούτε για τα πραγματικά ούτε για τα προσομοιωμένα δεδομένα. Τέλος, να αναφέρουμε ότι διερευνήσαμε τη μέθοδο παρεμβολής ελάχιστης καμπυλότητας (υλοποιήθηκε μέσω της εντολής `griddata`, επιλογή `v4`), ωστόσο το προγραμματιστικό περιβάλλον `matlab` δεν έδωσε αποτέλεσμα διότι απαιτούσε την αντιστροφή ενός μεγάλου πίνακα.

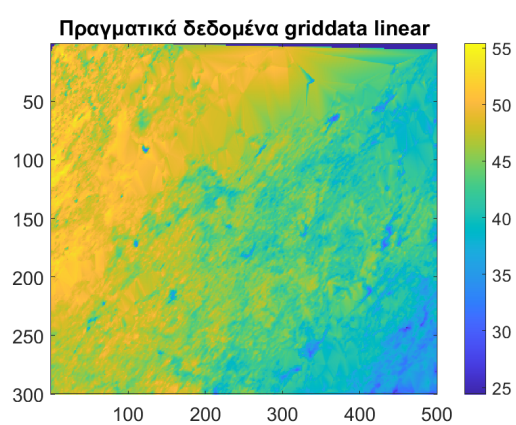
5.2. Αποτελέσματα για Δορυφορικά Δεδομένα



(i)



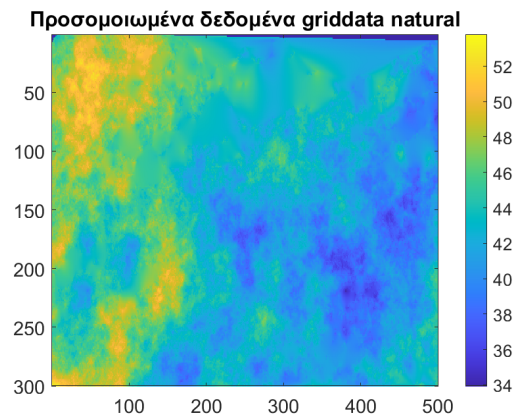
(ii)



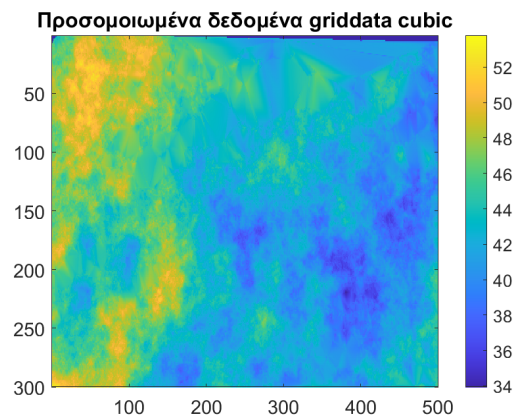
(iii)

Σχήμα 5.17: Αποτελέσματα από την εντολή `griddata`. Πραγματικά δεδομένα με μέθοδο παρεμβολής (i): επιλογή `natural neighbor`, (ii): επιλογή `cubic`, (iii): επιλογή `linear`.

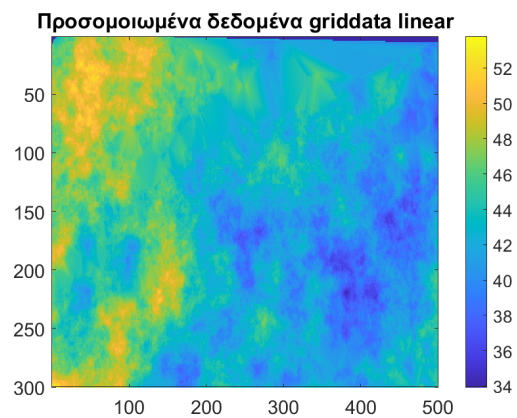
5.2. Αποτελέσματα για Δορυφορικά Δεδομένα



(i)



(ii)



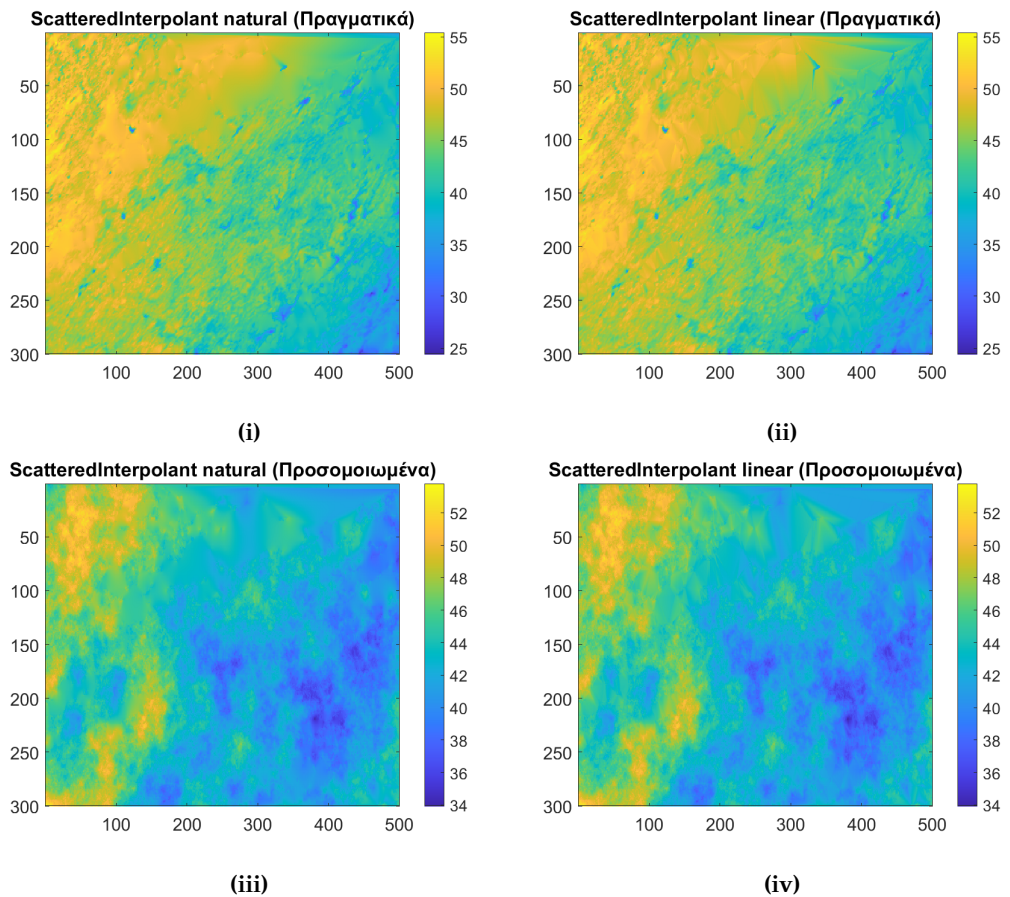
(iii)

Σχήμα 5.18: Αποτελέσματα από την εντολή griddata. Προσομοιωμένα δεδομένα με μέθοδο παρεμβολής (i): επιλογή natural neighbor, (ii): επιλογή cubic, (iii): επιλογή linear.

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα

Τέλος, αφού αναλύσαμε τα αποτελέσματα της εντολής `griddata` ανάλογα με τις μεθόδους παρεμβολής που εφαρμόσαμε, συνεχίσαμε με την εντολή `scatteredInterpolant`. Στα Σχήματα 5.19i και 5.19iii αξιολογήσαμε τη μεθοδολογία παρεμβολής φυσικών γειτόνων (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `natural neighbor`). Καταλήξαμε στο συμπέρασμα ότι τόσο για τα πραγματικά όσο και για τα προσομοιωμένα δεδομένα υλοποιείται η πλήρωση κενών. Επίσης, εφαρμόστηκε η μεθοδολογία γραμμικής παρεμβολής (υλοποιήθηκε μέσω της εντολής `scatteredInterpolant`, επιλογή `linear`) για τα δορυφορικά δεδομένα 5.19ii, 5.19iv. Όπως και προηγουμένως, εκτιμήθηκαν οι τιμές στα σημεία όπου υπήρχαν κενά με βέλτιστο τρόπο. Αυτό συνεπάγεται ότι έγινε επιτυχώς η πλήρωση των κενών.

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα



Σχήμα 5.19: Αποτελέσματα από την εντολή `scatteredInterpolant` με μέθοδο παρεμβολής επιλογή `natural neighbor` (i): πραγματικά δεδομένα, (iii): προσομοιωμένα δεδομένα. Μέθοδος παρεμβολής επιλογή `linear` (ii): πραγματικά δεδομένα, (iv): προσομοιωμένα δεδομένα.

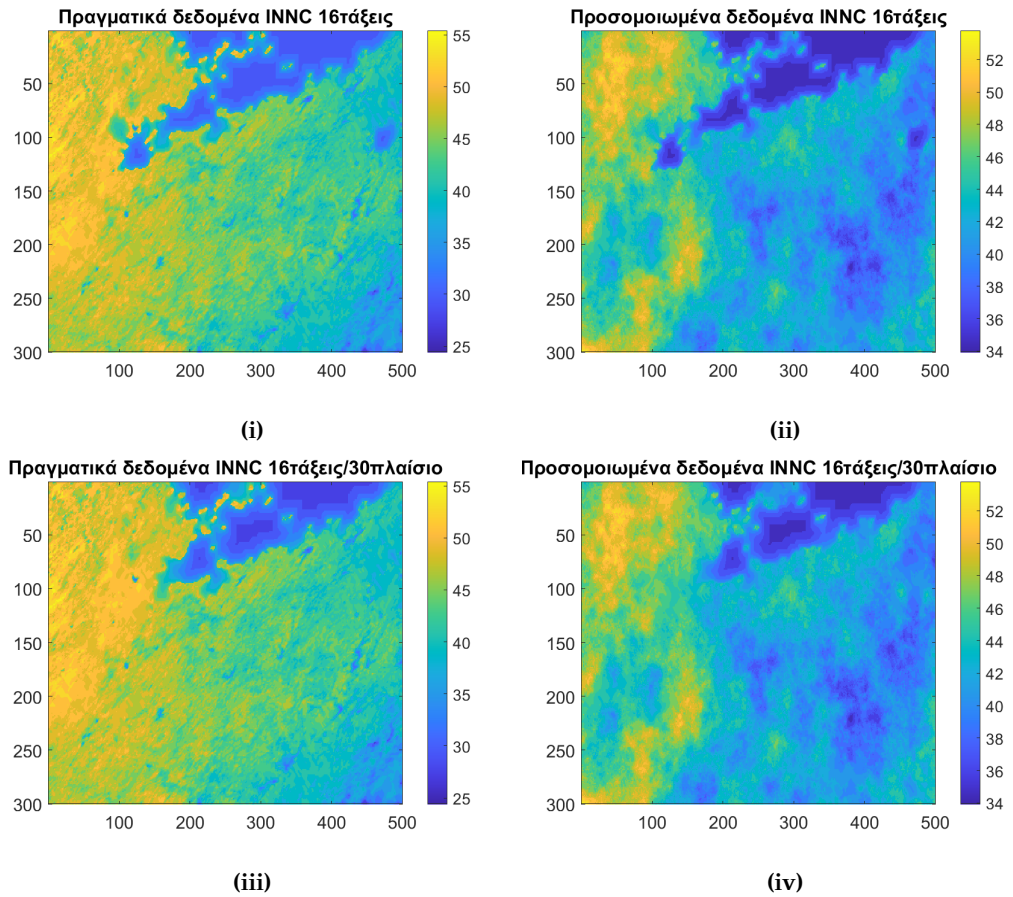
5.2.2 Μέθοδος INNC

Σε αυτό το σχήμα παρατηρούμε ότι τόσο για τα πραγματικά δεδομένα όσο και για τα προσομοιωμένα δεν έχει επιτευχθεί καλή ανακατασκευή. Πιο συγκεκριμένα στα Σχήματα 5.20i και 5.20ii, στα σημεία του κατακόρυφου άξονα μεταξύ των τιμών 200 με 300, έχει γίνει μια ικανοποιητική χωρική παρεμβολή σε σύγκριση και με τις αρχικές εικόνες των αντίστοιχων δεδομένων. Αντίθετα στα σημεία του κατακόρυφου άξονα μεταξύ των τιμών 0 με 150, φαίνεται ότι δεν έχει υλοποιηθεί σωστά η ανακατασκευή διότι τα δεδομένα που βρίσκονται γύρω από αυτό το σημείο έχουν κίτρινο και πράσινο χρώμα και αντί για το κίτρινο ή πράσινο χρώμα εντοπίζεται το μπλε να κυριαρχεί στα σημεία που ήταν τα κενά.

Σύμφωνα με τη μέθοδο που έχουμε εφαρμόσει, η συμπλήρωση των κενών γίνεται με βάση τη γειτονική πληροφορία. Αν κι η περιοχή με βάση τα σημεία που αναφέραμε προηγουμένως, περικλείεται από υψηλές τιμές (κίτρινο και πράσινο χρώμα), διακρίνουμε ότι η ανακατασκευή της με το μπλε χρώμα δεν διέθετε υψηλές τιμές. Τα δύο σχήματα που σχολιάσαμε, έχουν διεξαχθεί με 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 .

Για τα Σχήματα 5.20iii και 5.20iv ισχύουν τα ίδια που αναφέραμε παραπάνω με τη μόνη διαφορά ότι η μέθοδος INNC εφαρμόστηκε για τετραγωνικό πλαίσιο INNC 30×30 και όχι 20×20 . Αυτό υλοποιήθηκε για να δούμε εάν θα υπήρχε κάποια βελτίωση σε σχέση με τα άλλα δύο σχήματα. Υπάρχει μικρή βελτίωση για το Σχήμα 5.20iii σε σχέση με το Σχήμα 5.20i και το ίδιο παρατηρούμε και για το Σχήμα 5.20iv σε σύγκριση με το Σχήμα 5.20ii. Στα δύο σχήματα με τετραγωνικό πλαίσιο INNC 30×30 , εντοπίζουμε ότι στον κατακόρυφο άξονα 100 με 150 και αντίστοιχα στον οριζόντιο άξονα, έχει πραγματοποιηθεί και αποδοθεί σωστά η παρεμβολή στο συγκεκριμένο σημείο.

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα



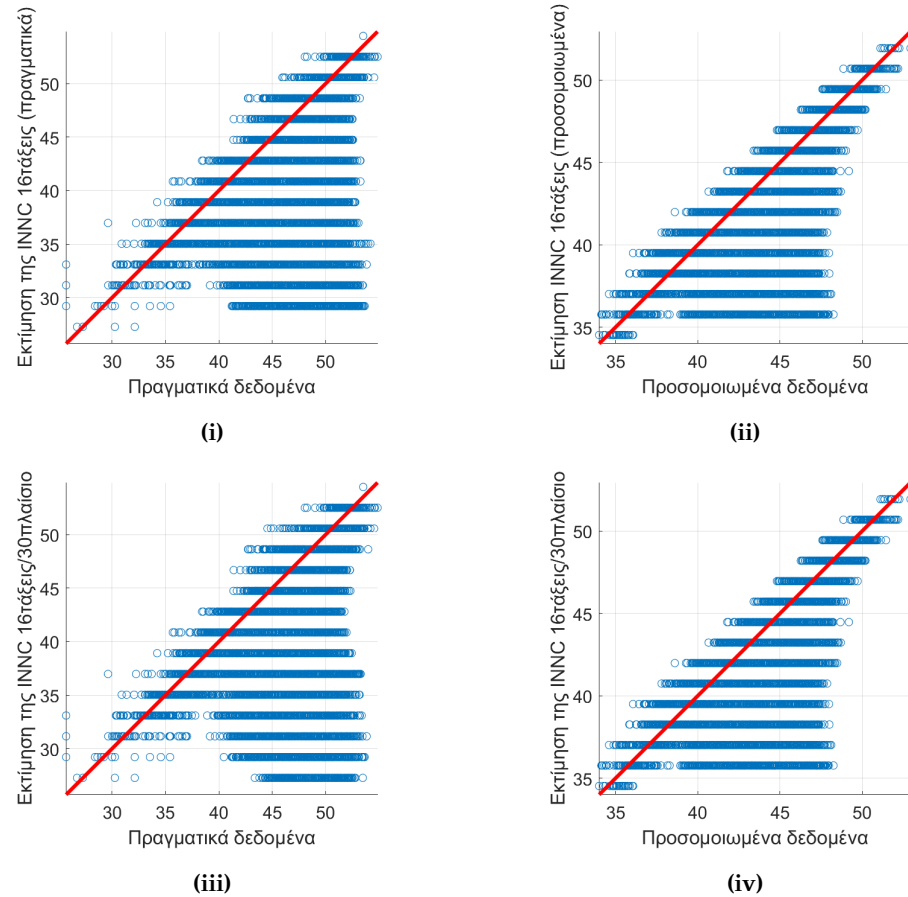
Σχήμα 5.20: Αποτελέσματα από τη μέθοδο INNC. Πραγματικά δεδομένα (i): 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iii): 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Προσομοιωμένα δεδομένα (ii) 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iv) 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 .

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα

Να σημειωθεί ότι στο προγραμματιστικό περιβάλλον matlab διερευνήσαμε κι άλλες περιπτώσεις με διαφορετικά πλαίσια για τα δορυφορικά δεδομένα. Αυτό είχε ως αποτέλεσμα να μην ανταποκρίνεται και να προκαλεί υπερφόρτωση του συστήματος, σε περιπτώσεις με τετραγωνικό πλαίσιο INNC μεγαλύτερο του 40×40 .

Στη συνέχεια, παρουσιάζονται τα διαγράμματα διασποράς όπου στον οριζόντιο άξονα απεικονίζονται οι αρχικές τιμές των πραγματικών δεδομένων για τα Σχήματα 5.21i και 5.21iii και στον κατακόρυφο άξονα οι τιμές που έχουν εκτιμηθεί από τη μέθοδο. Αξιολογώντας τα δύο σχήματα παρατηρείται μικρή διαφορά, την οποία θα αντιληφθούμε καλύτερα με τα μέτρα διασταυρωτικής επιβεβαίωσης που θα δούμε παρακάτω. Για τα Σχήματα 5.21ii και 5.21iv καταλήγουμε στο ίδιο συμπέρασμα, επομένως αναμένουμε τα αποτελέσματα που θα παρέχουν τα μέτρα που προαναφέραμε.

5.2. Αποτελέσματα για Δορυφορικά Δεδομένα



Σχήμα 5.21: Διαγράμματα διασποράς πρόβλεψης πραγματικών δεδομένων μέθοδος INNC (i): 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iii): 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Πρόβλεψης προσομοιωμένων δεδομένων μέθοδος INNC (ii): 16 τάξεις και τετραγωνικό πλαίσιο INNC 20×20 , (iv): 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 .

5.3 Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα των μέτρων διασταυρωτικής επιβεβαίωσης στο σύνολο συνθετικών δεδομένων καθώς και στο σύνολο των δορυφορικών δεδομένων. Επίσης, ακολουθεί ο σχολιασμός των τιμών που προέκυψαν από τις μεθόδους που εφαρμόσαμε και αναλύσαμε προηγουμένως.

Στον Πίνακα 5.1, παρουσιάζονται τα αποτελέσματα των μετρικών για την καθεμία μέθοδο που χρησιμοποιήθηκε στην περίπτωση της τυχαίας κατανομής των 40 κενών εικονοστοιχείων. Παρατηρούμε ότι οι τιμές των μετρικών συμφωνούν με τα οπτικά αποτελέσματα. Οι τιμές και των τεσσάρων μετρικών υποδεικνύουν ότι οι παραγόμενες εικόνες είναι ποιοτικές. Η τιμή του ME πρέπει να είναι 0 για να θεωρηθεί ότι μια μέθοδος παρεμβολής είναι ακριβής. Άρα, σύμφωνα με τον πίνακα, για το ME καλύτερη τιμή θεωρείται η 0.09. Όσον αφορά τα MAE και RMSE, καλύτερες τιμές παρουσιάζει η εντολή `griddata` με τη μέθοδο `v4`, με τιμές 0.22 και 0.28 αντίστοιχα. Για το συντελεστή συσχέτισης Pearson (ρ), η καλύτερη τιμή 0.99 εμφανίζεται στην εντολή `griddata` με τις μεθόδους `v4`, `natural neighbor` και στην εντολή `scatteredInterpolant` με τη μέθοδο `natural neighbor`. Επομένως, η εντολή `griddata` με τη μέθοδο `v4` έχει ικανοποιητικά αποτελέσματα κι αυτό φαίνεται από την παραγόμενη εικόνα. Όμως, η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 5×5 είναι καλύτερη επιλογή σε σύγκριση με τη μέθοδο `nearest neighbor` με εντολές `griddata` και `scatteredInterpolant` και έχει εκτελεστεί πιο γρήγορα σε σχέση με τη δεύτερη εντολή. Επίσης, με βάση τους χρόνους εκτέλεσης των μεθόδων, πιο γρήγορες είναι οι εντολές `griddata` και `scatteredInterpolant` με τη μέθοδο `natural neighbor`. Ο χρόνος εκτέλεσης είναι 0.67 και για τις δύο εντολές. Αντίθετα περισσότερο χρόνο διατέθηκε στην εντολή `scatteredInterpolant` με τη μέθοδο `nearest neighbor`.

Στον Πίνακα 5.2, εμφανίζονται τα αποτελέσματα των μετρικών για την περίπτωση της συνεχόμενης περιοχής 303 κενών εικονοστοιχείων.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Για το ME καλύτερη τιμή θεωρείται η 0.15. Όσον αφορά τα MAE και RMSE, καλύτερες τιμές εμφανίζει η εντολή `scatteredInterpolant` με τη μέθοδο `linear`, με τιμές 0.84 και 1.17 αντίστοιχα. Για το συντελεστή συσχέτισης Pearson (ρ), η καλύτερη τιμή 0.92 εμφανίζεται στην εντολή `scatteredInterpolant` με τη μέθοδο `linear`, η οποία καταλήγουμε ότι είναι καταλληλότερη σε σύγκριση με τις υπόλοιπες μεθόδους. Παρατηρώντας τους χρόνους εκτέλεσης, η πιο γρήγορη μέθοδος είναι η μέθοδος `natural neighbor` με εντολή `scatteredInterpolant`. Από όλες τις μεθόδους, χρειάζεται περισσότερο χρόνο η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Από τις αιτιοκρατικές μεθόδους, χρειάζεται η εντολή `griddata` με μέθοδο `v4`.

Μέτρα επιβεβαίωσης						
Μεθοδολογία	L	ρ	ME	MAE	RMSE	$T(\text{sec})$
<code>griddata(nnb)</code>	32	0.96	0.21	0.73	0.82	0.82
<code>griddata(v4)</code>	32	0.99	0.11	0.22	0.28	0.72
<code>griddata(natural neighbor)</code>	32	0.99	0.12	0.29	0.37	0.67
<code>griddata(cubic)</code>	32	0.98	0.09	0.41	0.49	0.68
<code>griddata(linear)</code>	32	0.98	0.09	0.47	0.55	0.68
<code>sI(nnb)</code>	32	0.96	0.21	0.73	0.82	1.07
<code>sI(natural neighbor)</code>	32	0.99	0.12	0.29	0.37	0.67
<code>sI(linear)</code>	32	0.98	0.09	0.47	0.55	0.90
INNC(8,5)	32	0.97	0.20	0.60	0.72	0.85

Πίνακας 5.1: Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της τυχαίας κατανομής 40 κενών εικονοστοιχείων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον `matlab`, το `sI` την εντολή `scatteredInterpolant` και το `nnb` τη μεθοδολογία `nearest neighbor`. Το $N_c = 8$ και $m_{\max} = 5$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Μέτρα επιβεβαίωσης						
Μεθοδολογία	L	ρ	ME	MAE	RMSE	$T(\text{sec})$
griddata(nnb)	32	0.74	0.15	1.34	1.57	0.78
griddata(v4)	32	0.89	-1.68	1.68	1.89	1.06
griddata(natural neighbor)	32	NaN	NaN	NaN	NaN	0.79
griddata(cubic)	32	NaN	NaN	NaN	NaN	0.97
griddata(linear)	32	NaN	NaN	NaN	NaN	0.89
sI(nnb)	32	0.86	0.60	1.48	1.72	0.76
sI(natural neighbor)	32	0.91	-0.76	0.87	1.21	0.67
sI(linear)	32	0.92	-0.71	0.84	1.17	0.68
INNC(8,30)	32	0.65	1.94	2.04	2.46	1.09
INNC(16,30)	32	0.59	2.92	2.93	3.27	1.02

Πίνακας 5.2: Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της συνεχόμενης περιοχής 303 κενών εικονοστοιχείων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το sI την εντολή scatteredInterpolant και το nnb τη μεθοδολογία nearest neighbor. Στην εντολή griddata σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 8$ και το $m_{\max} = 30$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 30$, τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.

Στον Πίνακα 5.3, προβάλλονται τα αποτελέσματα των μετρικών για τη συνεχή περιοχή 303 κενών εικονοστοιχείων, τοποθετημένα στο κέντρο. Για το ME καλύτερη τιμή θεωρείται η -0.24 . Επίσης, για τα MAE και RMSE, καλύτερες τιμές εμφανίζει η εντολή `scatteredInterpolant` με τη μέθοδο `linear`, με τιμές 1.06 και 1.34 αντίστοιχα. Για το συντελεστή συσχέτισης Pearson (ρ), η καλύτερη τιμή 0.97 εμφανίζεται στην εντολή `scatteredInterpolant` με τη μέθοδο `linear`. Μπορεί με βάση τα αποτελέσματα να υπερτερεί η μέθοδος αυτή, οπτικά όμως φαίνεται ότι έχει αλλοιωθεί σε σχέση με την αρχική εικόνα. Όμως, η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 30×30 είναι καλύτερη επιλογή σε σύγκριση με τις μεθόδους `nearest neighbor` και `v4` με εντολή `griddata` και έχει εκτελεστεί πιο γρήγορα σε σχέση με τη δεύτερη μέθοδο. Σύμφωνα με τους χρόνους εκτέλεσης, η πιο γρήγορη μέθοδος είναι η `nearest neighbor` με εντολή `scatteredInterpolant`. Σε σύγκριση με την εντολή `scatteredInterpolant` με μέθοδο `linear` λιγότερο χρόνο απαιτεί η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Από όλες τις μεθόδους, χρειάζεται περισσότερο χρόνο η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Από τις αιτιοκρατικές μεθόδους, χρειάζεται η εντολή `griddata` με μέθοδο `v4`.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Μέτρα επιβεβαίωσης						
Μεθοδολογία	L	ρ	ME	MAE	RMSE	$T(\text{sec})$
griddata(nnb)	32	-0.58	-2.88	4.01	4.99	0.77
griddata(v4)	32	-0.57	-2.89	4.01	4.99	0.96
griddata(natural neighbor)	32	NaN	NaN	NaN	NaN	0.69
griddata(cubic)	32	NaN	NaN	NaN	NaN	0.89
griddata(linear)	32	NaN	NaN	NaN	NaN	0.68
sI(nnb)	32	0.93	-0.98	1.16	1.48	0.44
sI(natural neighbor)	32	0.96	-1.07	1.14	1.48	0.69
sI(linear)	32	0.97	-0.95	1.06	1.34	0.85
INNC(8,30)	32	0.91	-0.24	1.11	1.39	0.81
INNC(16,30)	32	0.87	0.85	1.73	2.08	1.04

Πίνακας 5.3: Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της συνεχόμενης περιοχής 303 κενών εικονοστοιχείων, τοποθετημένων στο κέντρο. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το sI την εντολή scatteredInterpolant και το nnb τη μεθοδολογία nearest neighbor. Στην εντολή griddata σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 8$ και το $m_{\max} = 30$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 30$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Στον Πίνακα 5.4, παρουσιάζονται τα αποτελέσματα των μετρικών για την περίπτωση της τυχαίας κατανομής 303 κενών εικονοστοιχείων. Για το ME καλύτερη τιμή θεωρείται η -0.002 . Επίσης, για τα MAE και RMSE, καλύτερες τιμές εμφανίζει η εντολή `griddata` με τη μέθοδο `v4`, με τιμές 0.33 και 0.41 αντίστοιχα. Για το συντελεστή συσχέτισης Pearson (ρ), η καλύτερη τιμή 0.99 εμφανίζεται στην εντολή `griddata` με τη μέθοδο `v4`, η οποία συμπεραίνουμε ότι με βάση τις τιμές των μετρικών είναι πιο κοντά στην αρχική εικόνα. Όμως, η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 5×5 είναι καλύτερη επιλογή σε σύγκριση με τη μέθοδο `nearest neighbor` με εντολές `griddata` και `scatteredInterpolant`. Σύμφωνα με τους χρόνους εκτέλεσης, η πιο γρήγορη μέθοδος είναι η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 5×5 . Από τις αιτιοκρατικές μεθόδους, με τον ίδιο χρόνο εκτέλεσης γρήγορες είναι οι μέθοδοι `natural neighbor` και `linear` με εντολή `scatteredInterpolant`. Από όλες τις μεθόδους, χρειάζεται περισσότερο χρόνο η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 5×5 . Από τις αιτιοκρατικές μεθόδους, χρειάζεται η εντολή `griddata` με μέθοδο `nearest neighbor`.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Μέτρα επιβεβαίωσης						
Μεθοδολογία	L	ρ	ME	MAE	RMSE	$T(\text{sec})$
griddata(nnb)	32	0.95	0.04	0.70	0.80	0.76
griddata(v4)	32	0.99	0.08	0.33	0.41	0.71
griddata(natural neighbor)	32	NaN	NaN	NaN	NaN	0.69
griddata(cubic)	32	NaN	NaN	NaN	NaN	0.70
griddata(linear)	32	NaN	NaN	NaN	NaN	0.69
sI(nnb)	32	0.95	0.03	0.70	0.80	0.74
sI(natural neighbor)	32	0.98	0.02	0.38	0.48	0.67
sI(linear)	32	0.98	-0.002	0.44	0.54	0.67
INNC(8,5)	32	0.95	0.08	0.62	0.79	0.86
INNC(16,5)	32	0.96	0.11	0.58	0.69	0.45

Πίνακας 5.4: Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση της τυχαίας κατανομής 303 κενών εικονοστοιχείων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του τετραγώνου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το sI την εντολή scatteredInterpolant και το nnb τη μεθοδολογία nearest neighbor. Στην εντολή griddata σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 8$ και το $m_{\max} = 5$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 5$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Στη συνέχεια, παρουσιάζονται τα αποτελέσματα των μετρικών για τα δορυφορικά δεδομένα, χωρισμένα σε πραγματικά και προσομοιωμένα δεδομένα. Στον Πίνακα 5.5, που αφορά τα πραγματικά δεδομένα, το ME διαθέτει ως καλύτερη τιμή την 0.99. Επιπλέον, για τα MAE και RMSE, καλύτερες τιμές εμφανίζει η εντολή `scatteredInterpolant` με τη μέθοδο `natural neighbor`, με τιμές 1.36 και 1.89 αντίστοιχα. Για το συντελεστή συσχέτισης Pearson (ρ), η καλύτερη τιμή 0.93 εμφανίζεται στην ίδια μέθοδο. Άρα, η εντολή αυτή με τη συγκεκριμένη μέθοδο συμπεραίνουμε ότι έχει παράξει την εικόνα με τέτοιον τρόπο που πλησιάζει την αρχική εικόνα. Παρατηρώντας τους χρόνους εκτέλεσης, η πιο γρήγορη μέθοδος είναι η `linear` με εντολή `scatteredInterpolant`. Από όλες τις μεθόδους, χρειάζεται περισσότερο χρόνο η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Από τις αιτιοκρατικές μεθόδους, χρειάζεται η εντολή `scatteredInterpolant` με τη μέθοδο `nearest neighbor`.

Τέλος, στον Πίνακα 5.6, διατίθενται τα προσομοιωμένα δεδομένα, το ME παρέχει ως καλύτερη τιμή την -0.02 . Για τα MAE και RMSE, καλύτερες τιμές εμφανίζει η εντολή `scatteredInterpolant` με τη μέθοδο `natural neighbor`, με τιμές 0.69 και 1.02 αντίστοιχα. Για το συντελεστή συσχέτισης Pearson (ρ), η καλύτερη τιμή 0.93 εμφανίζεται στην εντολή `griddata` με τη μέθοδο `nearest neighbor` και στην εντολή `scatteredInterpolant` με τις μεθόδους `nearest neighbor` και `natural neighbor`. Επομένως, η εντολή `scatteredInterpolant` με τη μέθοδο `natural neighbor` έχει ικανοποιητικά αποτελέσματα κι αυτό επιβεβαιώνεται από την παραγόμενη εικόνα. Παρατηρώντας τους χρόνους εκτέλεσης, η πιο γρήγορη μέθοδος είναι η `natural neighbor` με εντολή `scatteredInterpolant`. Από όλες τις μεθόδους, χρειάζεται περισσότερο χρόνο η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 . Από τις αιτιοκρατικές μεθόδους, χρειάζεται η εντολή `scatteredInterpolant` με τη μέθοδο `nearest neighbor`.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Μέτρα επιβεβαίωσης						
Μεθοδολογία	L	ρ	ME	MAE	RMSE	$T(\text{sec})$
griddata(nnb)	300x500	0.92	0.99	1.41	1.99	1.81
griddata(natural neighbor)	300x500	NaN	NaN	NaN	NaN	1.63
griddata(cubic)	300x500	NaN	NaN	NaN	NaN	1.41
griddata(linear)	300x500	NaN	NaN	NaN	NaN	1.29
sI(nnb)	300x500	0.90	1.08	1.50	2.18	2.04
sI(natural neighbor)	300x500	0.93	1.05	1.36	1.89	1.56
sI(linear)	300x500	0.92	1.02	1.37	1.91	1.45
INNC(16,20)	300x500	-0.02	9.58	9.77	13.43	24.52
INNC(16,30)	300x500	0.15	6.89	7.10	10.43	95.41

Πίνακας 5.5: Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση των πραγματικών δεδομένων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του ορθογώνιου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το sI την εντολή scatteredInterpolant και το nnb τη μεθοδολογία nearest neighbor. Στην εντολή griddata σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 16$ και το $m_{\max} = 20$ και στην άλλη περίπτωση το $N_c = 16$ και $m_{\max} = 30$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.

5.3. Ανάλυση Απόδοσης Μεθόδων Χωρικής Παρεμβολής

Μέτρα επιβεβαίωσης						
Μεθοδολογία	L	ρ	ME	MAE	RMSE	$T(\text{sec})$
griddata(nnb)	300x500	0.93	-0.02	0.77	1.07	1.72
griddata(natural neighbor)	300x500	NaN	NaN	NaN	NaN	1.38
griddata(cubic)	300x500	NaN	NaN	NaN	NaN	1.13
griddata(linear)	300x500	NaN	NaN	NaN	NaN	1.20
sI(nnb)	300x500	0.93	-0.03	0.78	1.07	1.76
sI(natural neighbor)	300x500	0.93	0.10	0.69	1.02	1.34
sI(linear)	300x500	0.92	0.11	0.74	1.05	1.37
INNC(16,20)	300x500	0.44	3.15	3.43	4.95	19.47
INNC(16,30)	300x500	0.44	3.15	3.43	4.95	68.48

Πίνακας 5.6: Αποτελέσματα ρ , ME, MAE και RMSE από όλες τις μεθόδους πρόβλεψης για την περίπτωση των προσομοιωμένων δεδομένων. Το ρ αντιπροσωπεύει το συντελεστή συσχέτισης Pearson, το L τον αριθμό των κόμβων ανά πλευρά του ορθογώνιου, το T είναι ο χρόνος εκτέλεσης των μεθόδων σε προγραμματιστικό περιβάλλον matlab, το sI την εντολή scatteredInterpolant και το nnb τη μεθοδολογία nearest neighbor. Στην εντολή griddata σε ορισμένες μεθόδους έχουμε κενά, λόγω αποτυχίας ανακατασκευής της εικόνας. Το σύμβολο NaN δηλώνει μια μη αριθμητική τιμή. Στη μία περίπτωση, το $N_c = 16$ και το $m_{\max} = 20$ και στην άλλη περίπτωση το $N_c = 16$ και το $m_{\max} = 30$ τα οποία υποδηλώνουν τάξεις και τετραγωνικό πλαίσιο INNC αντίστοιχα.

Κεφάλαιο 6

Συμπεράσματα

Στη διπλωματική εργασία διερευνήσαμε αιτιοκρατικές μεθοδολογίες και τη μέθοδο παρεμβολής με αλληλεπιδράσεις κοντινότερων γειτόνων (INNC) για την πλήρωση κενών σε χωρικά δεδομένα (εικόνες). Χρησιμοποιήθηκαν συνθετικά και δορυφορικά δεδομένα. Ειδικότερα, τα συνθετικά δεδομένα δημιουργήθηκαν με χρήση της συνάρτησης Ackley, η οποία είναι μία μη κυρτή συνάρτηση σε χώρο δύο διαστάσεων. Σε αυτά τα δεδομένα εισήχθησαν κενά με διαφορετικές χωρικές διατάξεις και σε διαφορετικές θέσεις κενών σημείων/περιοχών. Τα δορυφορικά δεδομένα χρησιμοποιήθηκαν από την εργασία [5]. Πιο συγκεκριμένα, τα δορυφορικά δεδομένα αποτελούνται από ένα προσομοιωμένο και ένα πραγματικό σύνολο δεδομένων. Ο χωρισμός των δορυφορικών δεδομένων σε δύο σύνολα έγινε με βάση τη συνάρτηση συνδιακύμανσης. Δηλαδή, το προσομοιωμένο σύνολο δεδομένων αντιπροσώπευε μια περίπτωση όπου η συνάρτηση συνδιακύμανσης έχει καθοριστεί σωστά. Αντίθετα, το πραγματικό σύνολο δεδομένων μια περίπτωση όπου έχει καθοριστεί λανθασμένα η συνάρτηση συνδιακύμανσης. Αποδείξαμε ότι με τη χρήση των μεθοδολογιών αυτών επιτυγχάνεται η συμπλήρωση των κενών κι επομένως η ανακατασκευή της εικόνας. Η απόδοση της κάθε μεθόδου εξαρτάται από το μέγεθος του συνόλου δεδομένων και τα γεωμετρικά χαρακτηριστικά της περιοχής από την οποία λείπουν τα δεδομένα.

Όπως αναφέραμε στο Κεφάλαιο 4, χρησιμοποιήσαμε αιτιοκρατικές

μεθοδολογίες, οι οποίες υλοποιήθηκαν με τις εντολές `griddata` και `scatteredInterpolant` του προγραμματιστικού περιβάλλοντος `matlab`. Τόσο η εντολή `griddata` όσο και η εντολή `scatteredInterpolant` πραγματοποιούν παρεμβολή της επιφάνειας στα σημεία διερεύνησης με βάση γνωστές μετρήσεις στα σημεία του δείγματος. Επομένως, οι αιτιοκρατικές μεθοδολογίες έχουν εφαρμοστεί και με τις δύο εντολές επειδή ο τρόπος υλοποίησης των δύο εντολών είναι διαφορετικός.

Στη συνέχεια, εφαρμόστηκε η μέθοδος INNC, η οποία εμπνέεται από το μοντέλο Ising. Η μέθοδος αυτή βασίζεται στην ελαχιστοποίηση της συνάρτησης κόστους, η οποία μετρά την απόσταση μεταξύ της ενέργειας αλληλεπίδρασης μέσα σε δείγμα με την αντίστοιχη ενέργεια σε ολόκληρο το πλέγμα. Η μέθοδος INNC υλοποιείται με την εφαρμογή της μεθοδολογίας Monte Carlo. Επίσης, η μέθοδος INNC προσεγγίζει τις μεταβλητές ανάλογα με τον αριθμό τάξεων διακριτοποίησης που έχουμε ορίσει. Ο χρόνος εκτέλεσης της μεθόδου μεταβάλλεται σχεδόν γραμμικά σε σχέση με τον αριθμό των τάξεων και το μέγεθος του πλαισίου INNC [12, 14].

Σύμφωνα με τα αποτελέσματα που αναφέραμε διεξοδικά στο Κεφάλαιο 5, παρατηρήσαμε τα εξής. Όσον αφορά τα συνθετικά δεδομένα, η βέλτιστη μεθοδολογία διέφερε ανάλογα με τις θέσεις κενών σημείων/περιοχών. Για την επιλογή της βέλτιστης μεθοδολογίας, αναλύθηκαν οι εξής διατάξεις με κενά: 1) τυχαία κατανομή 40 κενών εικονοστοιχείων, 2) συνεχής περιοχή 303 κενών εικονοστοιχείων και 3) τυχαία κατανομή ισάριθμων κενών εικονοστοιχείων. Για τα δορυφορικά δεδομένα οι μεθοδολογίες εφαρμόστηκαν στα πραγματικά δεδομένα και στα προσομοιωμένα δεδομένα. Η σύγκριση των μεθοδολογιών για να βρεθεί η βέλτιστη μεθοδολογία πραγματοποιήθηκε με βάση τα μέτρα διασταυρωτικής επιβεβαίωσης.

Παρακάτω συνοψίζεται η απόδοση των διαφορετικών μεθοδολογιών για τις διαφορετικές γεωμετρικές κενών θέσεων. Στην περίπτωση των συνθετικών δεδομένων με 40 κενά εικονοστοιχεία βέλτιστη είναι η μέθοδος παρεμβολής ελάχιστης καμπυλότητας. Οι αμέσως καλύτερες μεθοδολογίες είναι η μέθοδος παρεμβολής φυσικών γειτόνων, η μέθοδος κυβικής παρεμβολής και η μέθοδος γραμμικής παρεμβολής. Η μέθοδος

INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 5×5 είναι καλύτερη, σε σύγκριση με τη μέθοδο των πλησιέστερων γειτόνων. Η μέθοδος παρεμβολής φυσικών γειτόνων υπερτερεί όσον αφορά την υπολογιστική ταχύτητα (χρόνος: 0.67sec), σε σύγκριση με τις υπόλοιπες μεθοδολογίες.

Η μέθοδος γραμμικής παρεμβολής είναι βέλτιστη για συνεχή περιοχή 303 κενών εικονοστοιχείων στην περίπτωση που τα κενά είναι τοποθετημένα στο δεξί τμήμα της εικόνας (ξεκινώντας από το σύνορο του πλέγματος). Οι αμέσως καλύτερες μεθοδολογίες είναι η μέθοδος παρεμβολής φυσικών γειτόνων, η μέθοδος των πλησιέστερων γειτόνων και η μέθοδος παρεμβολής ελάχιστης καμπυλότητας. Σε αυτή την περίπτωση όπως και στις παρακάτω που θα αναφερθούμε στη συνέχεια, η μέθοδος παρεμβολής φυσικών γειτόνων, η μέθοδος κυβικής παρεμβολής και η μέθοδος γραμμικής παρεμβολής (υλοποιημένες με την εντολή `griddata`) δεν καταλήγουν σε αριθμητικό αποτέλεσμα. Η μέθοδος INNC εμφάνισε χειρότερα αποτελέσματα σε σύγκριση με τις υπόλοιπες μεθοδολογίες. Επιπλέον, η μέθοδος παρεμβολής φυσικών γειτόνων είναι υπολογιστικά γρηγορότερη (χρόνος: 0.67sec), σε σύγκριση με τις υπόλοιπες μεθοδολογίες.

Στη συνέχεια, για συνεχή περιοχή 303 κενών εικονοστοιχείων εξετάσαμε την περίπτωση που η περιοχή των κενών είναι τοποθετημένη στο κέντρο. Η μέθοδος γραμμικής παρεμβολής είναι βέλτιστη, όπως και στην περίπτωση που τα κενά είναι τοποθετημένα στο δεξί τμήμα της εικόνας. Οι αμέσως καλύτερες μεθοδολογίες είναι η μέθοδος INNC με οχτώ τάξεις και τετραγωνικό πλαίσιο INNC 30×30 , η μέθοδος των πλησιέστερων γειτόνων και η μέθοδος παρεμβολής φυσικών γειτόνων. Στη συνέχεια, ακολουθούν η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 30×30 και η μέθοδος παρεμβολής ελάχιστης καμπυλότητας. Τέλος, η μέθοδος των πλησιέστερων γειτόνων είναι υπολογιστικά γρηγορότερη (χρόνος: 0.44sec), σε σύγκριση με τις υπόλοιπες μεθοδολογίες.

Η τυχαία κατανομή 303 κενών εικονοστοιχείων ήταν η τελευταία περίπτωση που διερευνήθηκε για τα συνθετικά δεδομένα. Η μέθοδος παρεμβολής ελάχιστης καμπυλότητας είναι βέλτιστη, όπως και στην περίπτωση της τυχαίας κατανομής 40 κενών εικονοστοιχείων. Οι αμέσως

καλύτερες μεθοδολογίες είναι η μέθοδος παρεμβολής φυσικών γειτόνων, η μέθοδος γραμμικής παρεμβολής, η μέθοδος INNC και η μέθοδος των πλησιέστερων γειτόνων. Επιπλέον, η μέθοδος INNC με 16 τάξεις και τετραγωνικό πλαίσιο INNC 5×5 υπερτερεί όσον αφορά την υπολογιστική ταχύτητα (χρόνος: 0.45sec), σε σύγκριση με τις υπόλοιπες μεθοδολογίες.

Όσον αφορά τα δορυφορικά δεδομένα, η μέθοδος παρεμβολής φυσικών γειτόνων είναι βέλτιστη για την πλήρωση κενών. Οι αμέσως καλύτερες μεθοδολογίες είναι η μέθοδος γραμμικής παρεμβολής και η μέθοδος των πλησιέστερων γειτόνων. Αντίθετα, η μέθοδος INNC δεν απέδωσε ικανοποιητικά τόσο για τα πραγματικά δεδομένα όσο και για τα προσομοιωμένα δεδομένα. Επίσης, η μέθοδος αυτή χρειάστηκε σημαντικά περισσότερο χρόνο για την εφαρμογή της στα δορυφορικά δεδομένα. Σύμφωνα με τα αποτελέσματα, η μέθοδος γραμμικής παρεμβολής είναι υπολογιστικά ταχύτερη (χρόνος: 1.45sec), σε σχέση με τις υπόλοιπες μεθοδολογίες στην περίπτωση των πραγματικών δεδομένων. Τέλος, η μέθοδος παρεμβολής φυσικών γειτόνων υπερτερεί σε χρόνο (χρόνος: 1.34sec) στην περίπτωση των προσομοιωμένων δεδομένων.

Συνοψίζοντας τα αποτελέσματα που προέκυψαν με βάση τα μέτρα διασταυρωτικής επιβεβαίωσης, καταλήγουμε στο συμπέρασμα ότι η μέθοδος παρεμβολής ελάχιστης καμπυλότητας είναι βέλτιστη για τα συνθετικά δεδομένα με τυχαία κατανομή κενών. Αντίθετα, η μέθοδος γραμμικής παρεμβολής είναι βέλτιστη για τα συνθετικά δεδομένα με συνεχή περιοχή κενών. Όσον αφορά τα δορυφορικά δεδομένα, τόσο για τα πραγματικά όσο και για τα προσομοιωμένα, η μέθοδος παρεμβολής φυσικών γειτόνων είναι καλύτερη σε σύγκριση με τις υπόλοιπες μεθοδολογίες.

Σύμφωνα με τα αποτελέσματα που αναλύσαμε προηγουμένως, οι αιτιοκρατικές μεθοδολογίες που παρουσιάσαμε έχουν μεγάλες δυνατότητες ως προς την κάλυψη κενών σε χωρικά δεδομένα (εικόνες). Η μέθοδος INNC, από την άλλη πλευρά, δεν εμφάνισε καλά αποτελέσματα ειδικά στα δορυφορικά δεδομένα. Επομένως, η επέκταση της μεθόδου INNC έτσι ώστε να μπορεί να αντιμετωπίσει προβλήματα στα οποία υπάρχουν συνεχή κενά παρουσιάζει ενδιαφέρον για μελλοντική έρευνα. Αυτό για να γίνει πρέπει ουσιαστικά να οριστεί η αρχικοποίηση της μεθόδου

με σκοπό να δουλέψει σε τέτοιες περιπτώσεις. Επιπροσθέτως, η τμηματοποίηση (Segmentation) θα ήταν χρήσιμη λόγω της ομαδοποίησης των εικονοστοιχείων ανάλογα με την τιμή και τη θέση του εικονοστοιχείου. Αν τα εικονοστοιχεία δηλαδή είναι χωρικά συνδεδεμένα και έχουν ίδιες τιμές τότε ομαδοποιούνται σε ένα μοναδικό τμήμα. Επίσης, για το πρόβλημα της πλήρωσης κενών μπορούν να εξεταστούν και να συγκριθούν διαφορετικές στοχαστικές μεθοδολογίες (όπως η μέθοδος Kriging ή η παλινδρόμηση βασισμένη σε γκαουσιανές στοχαστικές διαδικασίες).

Σε μια μελλοντική έρευνα, επίσης θα παρουσίαζε ιδιαίτερο ενδιαφέρον η διερεύνηση της αβεβαιότητας των προβλέψεων. Αυτό μπορεί να γίνει στα πλαίσια της χρήσης στοχαστικών μεθόδων. Επιπλέον, όσον αφορά την πλήρωση κενών, θα παρουσίαζε ενδιαφέρον η εφαρμογή των νευρωνικών δικτύων συνέλιξης (Convolutional neural networks). Τα νευρωνικά δίκτυα συνέλιξης εκμαιεύουν τα βασικά χαρακτηριστικά απευθείας από τις εικόνες εκπαίδευσης μέσω της αλληλεπικάλυψης των επιπέδων συνέλιξης και συγκέντρωσης. Τέλος, για μελλοντική έρευνα αξιοσημείωτο είναι να αναφερθούν τα γεννητικά ανταγωνιστικά δίκτυα (Generative adversarial networks). Αυτά τα δίκτυα διαθέτουν ένα ισχυρό τρόπο δημιουργίας τεχνητών εικόνων που μιμούνται σε μεγάλο βαθμό την επιθυμητή εικόνα.

Βιβλιογραφία

- [1] Laura Almendra-Martín, José Martínez-Fernández, María Piles, and Ángel González-Zamora. Comparison of gap-filling techniques applied to the cci soil moisture database in southern europe. *Remote Sensing of Environment*, 258:112377, 2021.
- [2] Thomas Back. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford university press, 1996.
- [3] Jean-Paul Chiles and Pierre Delfiner. *Geostatistics: Modeling Spatial Uncertainty*, volume 497. John Wiley & Sons, 2009.
- [4] Chad Greene. Offsets and missing data via pcolor and surf — matlab central file exchange, 2022.
- [5] Matthew J Heaton, Abhirup Datta, Andrew O Finley, Reinhard Furrer, Joseph Guinness, Rajarshi Guhaniyogi, Florian Gerber, Robert B Gramacy, Dorit Hammerling, Matthias Katzfuss, et al. A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):398–425, 2019.
- [6] Dionissios T Hristopulos. *Random Fields for Spatial Data Modeling*. Springer, Dordrecht, Netherlands, 2020.
- [7] Georg Lindgren. *Stationary Stochastic Processes: Theory and Applications*. CRC Press, 2012.

- [8] Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data*, volume 793. John Wiley & Sons, 2019.
- [9] Ricardo A Olea. A practical primer on geostatistics. Technical report, US Geological Survey, 2018.
- [10] Christos H Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Courier Corporation, 1998.
- [11] Xingmei Wang, Shu Liu, and Zhipeng Liu. Underwater sonar image detection: A combination of non-local spatial information and quantum-inspired shuffled frog leaping algorithm. *Plos One*, 12(5):e0177666, 2017.
- [12] Milan Žukovič and Dionissios T Hristopulos. Classification of missing values in spatial data using spin models. *Physical Review E*, 80(1):011116, 2009.
- [13] Milan Žukovič and Dionissios T Hristopulos. A directional gradient-curvature method for gap filling of gridded environmental spatial data with potentially anisotropic correlations. *Atmospheric Environment*, 77:901–909, 2013.
- [14] Milan Žukovič and Dionissios T Hristopulos. Ising model for interpolation of spatial data on regular grids. *Entropy*, 23(10):1270, 2021.