



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΝΑΥΠΗΓΩΝ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
«ΝΑΥΤΙΚΗ ΚΑΙ ΘΑΛΑΣΣΙΑ ΤΕΧΝΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΗ»

**Συγκριτική Μελέτη Εναλλακτικών Μεθόδων Παλινδρόμησης και  
Εφαρμογή σε Ανεμολογικά Δεδομένα**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΦΩΡΑΣ Ε. ΚΑΡΑΘΑΝΑΣΗ**

Διπλωματούχου Αγρονόμου και Τοπογράφου Μηχανικού Ε.Μ.Π. (2009)

**Επιβλέπων:** Τακβόρ Χ. Σουκισιάν  
Δρ. Ερευνητής Β', Ελληνικό Κέντρο Θαλάσσιων  
Ερευνών (ΕΛΚΕΘΕ)

Αθήνα, Ιούλιος 2013





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΝΑΥΠΗΓΩΝ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ  
ΣΠΟΥΔΩΝ ΣΕ  
«ΝΑΥΤΙΚΗ ΚΑΙ ΘΑΛΑΣΣΙΑ ΤΕΧΝΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΗ»

**Συγκριτική Μελέτη Εναλλακτικών Μεθόδων Παλινδρόμησης και  
Εφαρμογή σε Ανεμολογικά Δεδομένα**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΦΛΩΡΑΣ Ε. ΚΑΡΑΘΑΝΑΣΗ**

Διπλωματούχου Αγρονόμου και Τοπογράφου Μηχανικού Ε.Μ.Π. (2009)

**Επιβλέπων:** Τακβόρ Χ. Σουκισιάν  
Δρ. Ερευνητής Β', Ελληνικό Κέντρο Θαλάσσιων  
Ερευνών (ΕΛΚΕΘΕ)

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή τη 16<sup>η</sup> Ιουλίου 2013.

.....

Τακβόρ Χ. Σουκισιάν  
Δρ. Ερευνητής Β', ΕΛ.ΚΕ.Θ.Ε.

.....

Σπυρίδων Α. Μαυράκος  
Καθηγητής Ε.Μ.Π.

.....

Κων/νος Μπελιμπασάκης  
Αναπλ. Καθηγητής Ε.Μ.Π.

.....  
Φλώρα Ε. Καραθανάση

Κάτοχος Διατμηματικού Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Επιστημονική Περιοχή «Ναυτική και Θαλάσσια Τεχνολογία και Επιστήμη»

Copyright © Φλώρα Καραθανάση, 2013

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	1
ABSTRACT.....	2
ΚΕΦΑΛΑΙΟ 1 <sup>ο</sup> – ΕΙΣΑΓΩΓΗ.....	3
1.1. ΓΕΝΙΚΑ.....	3
1.2. ΣΚΟΠΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ.....	4
1.3. ΔΟΜΗ ΔΙΠΛΩΜΑΤΙΚΗΣ.....	4
ΚΕΦΑΛΑΙΟ 2 <sup>ο</sup> – ΠΗΓΕΣ ΚΑΙ ΕΙΔΗ ΑΝΕΜΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ.....	6
2.1. ΕΠΙΤΟΠΙΕΣ ΜΕΤΡΗΣΕΙΣ.....	6
2.2. ΜΕΤΡΗΣΕΙΣ ΑΠΟ ΠΛΟΙΑ ΕΘΕΛΟΝΤΙΚΗΣ ΠΑΡΑΤΗΡΗΣΕΙΣ.....	9
2.3. ΠΑΡΑΤΗΡΗΣΕΙΣ ΤΗΛΕΑΝΙΧΝΕΥΣΗΣ.....	9
2.3.1. Επίγειες Τεχνικές Τηλεανίχνευσης.....	11
2.3.2. Αερομεταφερόμενες συσκευές τηλεανίχνευσης.....	14
2.3.3. Δορυφορικές μετρήσεις.....	15
2.4. ΔΕΔΟΜΕΝΑ ΑΡΙΘΜΗΤΙΚΗΣ ΠΡΟΣΟΜΟΙΩΣΗΣ.....	19
ΚΕΦΑΛΑΙΟ 3 – ΜΕΘΟΔΟΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....	24
3.1. ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....	24
3.1.1. Γενικά.....	24
3.1.2. Μέθοδος Συνηθισμένων Ελαχίστων Τετραγώνων.....	24
3.1.3. Ρύθμιση/διόρθωση μέσω γραμμικής παλινδρόμησης.....	30
3.1.4. Σημεία Επιρροής (Influential Points).....	31
3.1.5. Έκτροπες παρατηρήσεις (Outliers).....	32
3.2. ΠΑΛΙΝΔΡΟΜΗΣΗ ΜΕ ΣΦΑΛΜΑΤΑ ΜΕΤΡΗΣΕΩΣ ΣΤΙΣ ΜΕΤΑΒΛΗΤΕΣ.....	36
3.3. ΑΝΘΕΚΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....	41
3.3.1. Γενικά.....	41
3.3.2. Μέτρα Ανθεκτικότητας.....	42
3.3.3. Ανθεκτικοί εκτιμητές.....	45
ΚΕΦΑΛΑΙΟ 4 – ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ.....	57
4.1. ΑΝΑΛΥΣΗ ΑΝΕΜΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ.....	57
4.1.1. Διαδικασία Ανάλυσης Ανεμολογικών Δεδομένων.....	57
4.1.2. Μέθοδοι Παλινδρόμησης.....	60
4.2. ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΑΠΟ ΤΙΣ ΔΙΑΦΟΡΕΤΙΚΕΣ ΜΕΘΟΔΟΥΣ.....	67
ΚΕΦΑΛΑΙΟ 5 – ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ.....	79
5.1. ΣΥΜΠΕΡΑΣΜΑΤΑ.....	79
5.2. ΠΡΟΤΑΣΕΙΣ.....	80
6. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	81

## ΕΥΡΕΤΗΡΙΟ ΓΡΑΦΗΜΑΤΩΝ

<b>Γράφημα 4. 1.</b> Τιμές της παραμέτρου $\hat{b}_0$ της ευθείας παλινδρόμησης όλων των τοποθεσιών για όλες τις μεθόδους πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.....	71
<b>Γράφημα 4. 2.</b> Τιμές της παραμέτρου $\hat{b}_1$ της ευθείας παλινδρόμησης όλων των τοποθεσιών για όλες τις μεθόδους πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.....	71
<b>Γράφημα 4. 3.</b> Τιμές του συντελεστή $R_\alpha^2$ όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους. ....	72
<b>Γράφημα 4. 4.</b> Τιμές του δείκτη <i>RMSE</i> όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.....	73
<b>Γράφημα 4. 5.</b> Τιμές του δείκτη <i>BIAS</i> όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.....	74
<b>Γράφημα 4. 6.</b> Τιμές του δείκτη <i>Absolute BIAS</i> όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.....	75
<b>Γράφημα 4. 7.</b> Τιμές του δείκτη <i>SI</i> όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους. ....	76

## ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

<b>Εικόνα 2. 1.</b> Μετρητικός σταθμός τύπου Sea-Watch του συστήματος <i>POSEIDON</i> .....	7
<b>Εικόνα 2. 2.</b> Παρασυρόμενος πλωτήρας. ....	7
<b>Εικόνα 2. 3.</b> Υπεράκτιος μετεωρολογικός ιστός. ....	8
<b>Εικόνα 2. 4.</b> Διαφορετικές διατάξεις σχετικά με τις παρατηρήσεις στη βιομηχανία του ανέμου που βασίζεται σε επίγειες τεχνικές τηλεανίχνευσης. Στα αριστερά εμφανίζονται οι προσεγγιστικές διαστάσεις μιας ανεμογεννήτριας, στη μέση οι κώνοι παρατήρησης του <i>LiDAR</i> τοποθετημένο πάνω στο δρομέα ή την άτρακτο και στο έδαφος, στα δεξιά φαίνεται το κατακόρυφο προφίλ του ανέμου επηρεασμένο από το φαινόμενο <i>wake</i> και άκρως δεξιά το αδιατάρακτο κατακόρυφο προφίλ του ανέμου [29].....	10
<b>Εικόνα 2. 5.</b> Τοποθέτηση συσκευής <i>SoDAR</i> σε πλατφόρμα ανοικτά των ακτών <i>Georgetown</i> , Βόρεια Καρολίνα. ....	11
<b>Εικόνα 2. 6.</b> Εγκατεστημένο <i>LiDAR</i> (μοντέλο <i>ZephIR</i> ) σε υποσταθμό πλατφόρμας στο πρώτο παράκτιο αιολικό πάρκο της Σκωτίας, το <i>Robin Rigg</i> (180MW) στο <i>Solway Firth</i> . ...	12
<b>Εικόνα 2. 7.</b> Η πλωτή συσκευή <i>FLiDAR</i> σε λειτουργία.....	13
<b>Εικόνα 2. 8.</b> Μετεωρολογικό μπαλόνι έτοιμο να τεθεί σε λειτουργία για τη μέτρηση του ανέμου. ....	14
<b>Εικόνα 2. 9.</b> Αρχή λειτουργίας του σκεδασίμετρου.....	17
<b>Εικόνα 2. 10.</b> Η ταχύτητα του ανέμου μετρημένη από το <i>SSM/I</i> την 01/08/2012 [65]. ....	19
<b>Εικόνα 2. 11.</b> Φυσικές διεργασίες των αριθμητικών μοντέλων καιρού [64]. ....	20
<b>Εικόνα 2. 12.</b> Το τρισδιάστατο πλέγμα σημείων ενός παγκόσμιου μοντέλου [64]. ....	20
<b>Εικόνα 4. 1.</b> Θέσεις πλωτήρων του συστήματος <i>POSEIDON</i> . ....	57

## ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

<b>Πίνακας 3. 1.</b> Συνήθεις συναρτήσεις για τους M-εκτιμητές [60].	51
<b>Πίνακας 3. 2.</b> Βασικές ιδιότητες ανθεκτικότητας για διάφορους εκτιμητές ανθεκτικότητας.	56
<b>Πίνακας 4. 1.</b> Θέσεις και περίοδος καταγραφής των πλωτήρων του συστήματος <i>POSEIDON</i> .	57
<b>Πίνακας 4. 2.</b> Βασικά στατιστικά στοιχεία των συνταυτισμένων ανεμολογικών δεδομένων για τη χρονική περίοδο 2000 – 2004.	59
<b>Πίνακας 4. 3.</b> Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση <i>Athos</i> πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.	68
<b>Πίνακας 4. 4.</b> Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση <i>Avgo</i> πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.	68
<b>Πίνακας 4. 5.</b> Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση <i>Lesvos</i> πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.	68
<b>Πίνακας 4. 6.</b> Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση <i>Mykonos</i> πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.	69
<b>Πίνακας 4. 7.</b> Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση <i>Santorini</i> πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.	69
<b>Πίνακας 4. 8.</b> Τιμές του <i>RMSE</i> για όλες τις μεθόδους σε όλες τις θέσεις για το 100% και το 50% (σε παρένθεση) του δειγματικού πλήθους των ανεμολογικών δεδομένων.	77
<b>Πίνακας 4. 9.</b> Τιμές του <i>Absolute BIAS</i> για όλες τις μεθόδους σε όλες τις θέσεις για το 100% και το 50% (σε παρένθεση) του δειγματικού πλήθους των ανεμολογικών δεδομένων.	78
<b>Πίνακας 4. 10.</b> Τιμές του <i>SI</i> για όλες τις μεθόδους σε όλες τις θέσεις για το 100% και το 50% (σε παρένθεση) του δειγματικού πλήθους των ανεμολογικών δεδομένων.	78

## ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

<b>Σχήμα 3. 1.</b> Γραφική απεικόνιση της κατακόρυφης απόστασης ενός καταλοίπου από τη γραμμή παλινδρόμησης στη μέθοδο ελαχίστων τετραγώνων.	27
<b>Σχήμα 3. 2.</b> Δεδομένα με έκτροπες παρατηρήσεις. Η συνεχής γραμμή αντιστοιχεί στον εκτιμητή ελαχίστων τετραγώνων ενώ η διακεκομμένη σε ανθεκτικό εκτιμητή. (a) $y$ -ακραία παρατήρηση, (b) κακό σημείο μόχλευσης, (c) καλό σημείο μόχλευσης [1].	34
<b>Σχήμα 3. 3.</b> Χώρος δύο διαστάσεων με μία έκτροπη παρατήρηση.	34
<b>Σχήμα 3. 4.</b> Ελαχιστοποίηση κάθετης απόστασης από τη γραμμή παλινδρόμησης για την παλινδρόμηση ορθογώνιας απόστασης.	40
<b>Σχήμα 3. 5.</b> Ελαχιστοποίηση κατακόρυφης και οριζόντιας απόστασης για τη συναρτησιακή σχέση γεωμετρικού μέσου.	40
<b>Σχήμα 3. 6.</b> Το διάγραμμά αυτό περιέχει 79 σημεία εκ των οποίων τα 40 προκύπτουν από το μοντέλο $y = 3x + 7 + \varepsilon_i$ . Παρόλο που τα υπόλοιπα 39 σημεία θεωρούνται ως έκτροπες παρατηρήσεις, ο MM-εκτιμητής έχει υπολογίσει σχεδόν απόλυτα τις πραγματικές τιμές των συντελεστών $\hat{b}$ , δίνοντας το αποτέλεσμα $\hat{b} = (7.151, 2.998)$ [60].	55
<b>Σχήμα 4. 1.</b> Σχηματική αναπαράσταση για τη χωρική συνταύτιση.	59
<b>Σχήμα 4. 2.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Athos</i> .	62
<b>Σχήμα 4. 3.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Avgo</i> .	62

<b>Σχήμα 4. 4.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Lesvos</i> .....	63
<b>Σχήμα 4. 5.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Mykonos</i> .....	63
<b>Σχήμα 4. 6.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Santorini</i> .....	64
<b>Σχήμα 4. 7.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Athos</i> .....	64
<b>Σχήμα 4. 8.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Avgo</i> .....	65
<b>Σχήμα 4. 9.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Lesvos</i> .....	65
<b>Σχήμα 4. 10.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Mykonos</i> .....	66
<b>Σχήμα 4. 11.</b> Γραμμές παλινδρόμησης των $u_B$ και $u_M$ μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση <i>Santorini</i> .....	66



---

## ΠΕΡΙΛΗΨΗ

Από τη μελέτη της σχετικής επιστημονικής και τεχνικής βιβλιογραφίας προκύπτει με σαφήνεια η αναγκαιότητα προσδιορισμού των πεδίων της ταχύτητας του ανέμου με όσο το δυνατόν μεγαλύτερη ακρίβεια, προκειμένου να επιτευχθεί ο ορθότερος σχεδιασμός ενός αιολικού πάρκου. Παρόλο που το αιολικό δυναμικό της Ελλάδας έχει μελετηθεί και έχει κριθεί κατάλληλο για τη δημιουργία θαλασσίων αιολικών πάρκων εν τούτοις ο σχεδιασμός τους βρίσκεται σε πολύ πρώιμο στάδιο. Σήμερα, οι πιο συχνά χρησιμοποιούμενες πηγές ανεμολογικών δεδομένων, αναφορικά με μετρήσεις στη θάλασσα, είναι οι πλωτοί μετρητικοί σταθμοί, οι δορυφορικές μετρήσεις και τα αριθμητικά μοντέλα προσομοίωσης. Όμως, λόγω των σημαντικών αδυναμιών που εντοπίζεται σε καθεμία μέθοδο, για διαφορετικούς λόγους, καθίσταται απαραίτητη η συνδυαστική τους μεταχείριση. Προς τούτο, το πλέον κλασσικό μαθηματικό εργαλείο για την ανάδειξη των γραμμικών σχέσεων μεταξύ δεδομένων που προέρχονται από διαφορετικές πηγές είναι η κλασσική παλινδρόμηση. Ωστόσο, η μέθοδος αυτή θεωρείται υποπερίπτωση μιας γενικότερης κατηγορίας, της παλινδρόμησης με σφάλματα μετρήσεως στις μεταβλητές, η οποία θεωρεί ότι υπεισέρχονται σφάλματα τόσο στην ανεξάρτητη όσο και την εξαρτημένη μεταβλητή. Αυτή η υπόθεση δεν ισχύει για την απλή γραμμική παλινδρόμηση.

Στην παρούσα εργασία, δίνεται έμφαση στην περιγραφή και εφαρμογή στατιστικών μεθόδων που λαμβάνουν με σαφήνεια υπόψη τις παρατηρήσεις που διαφέρουν από την πλειοψηφία των δεδομένων και ονομάζονται έκτροπες παρατηρήσεις. Οι παρατηρήσεις μπορεί να χαρακτηρίζονται ως έκτροπες για διαφορετικούς λόγους, ο συνηθέστερος των οποίων είναι τα σφάλματα κατά τη διαδικασία μέτρησης ή κατά τη διαδικασία μοντελοποίησης ενός φυσικού φαινομένου. Οι μέθοδοι αυτές είναι γνωστές στη βιβλιογραφία ως ανθεκτικές μέθοδοι παλινδρόμησης και χαρακτηρίζονται από προσαρμοστικότητα στην πληθώρα των δεδομένων με ταυτόχρονη αντιμετώπιση πλήθους έκτροπων παρατηρήσεων. Επιπρόσθετα, γίνεται σύγκριση μεταξύ των μεθόδων με τη βοήθεια στατιστικών δεικτών τόσο σε επίπεδο παλινδρόμησης όσο και σε επίπεδο ρύθμισης/διόρθωσης (*calibration*). Από όσο είμαστε σε θέση να γνωρίζουμε, είναι η πρώτη φορά στη βιβλιογραφία που γίνεται μια τέτοιου είδους συστηματική ανάλυση και διερεύνηση των ανθεκτικών μεθόδων σε ένα τόσο βασικό φυσικό χαρακτηριστικό όπως είναι η ταχύτητα ανέμου. Αξιοσημείωτο, τέλος, είναι το γεγονός ότι άλλες μέθοδοι δίνουν τα "καλύτερα" αποτελέσματα στην παλινδρόμηση και άλλες στην διόρθωση.

**Λέξεις κλειδιά:** ταχύτητα ανέμου, ανθεκτικές μέθοδοι, έκτροπες παρατηρήσεις, παλινδρόμηση, διόρθωση

**ABSTRACT**

From the study of relevant scientific and technical literature, it clearly emerges the necessity of quantifying the wind speed fields with the highest possible accuracy in order to efficiently apply a sitting methodology of a wind park. Although the wind potential in Greece has been studied and has been deemed suitable for the construction of offshore wind parks, at the same time little progress has been made in the planning phase. Nowadays, the most frequently used wind data sources, as regards wind measurements at sea, are floating measurement stations (buoys), satellite measurements and numerical simulation models. But due to significant weaknesses detected in each method, for different reasons, it is required to combine them. To this effect, the most classic mathematical tool for highlighting linear relationships between data from different sources is the classical regression. However, this method is a subcase of a more general category, the so-called Error-In-Variables, which considers that there are measurement errors both in the independent and the dependent variable. This assumption does not apply to the simple linear regression.

In this work, the emphasis is on the description and application of statistical methods that take clearly into consideration the observations that diverge from the bulk of the data and are called outliers. The observations may be characterized as outliers for various reasons, the most usual of which is the errors during the measurement process or during the process of modeling a physical phenomenon. These methods are known in the literature as robust regression methods and are characterized by the adaptability to the majority of the data while going through a large number of outliers. Furthermore, all the aforementioned methods are compared to each other concerning two levels, the regression as well as the calibration/correction relation. As far as we are aware, this is the first time in the literature that such a systematic analysis and examination of robust methods take place for a basic physical feature such as wind speed. Finally, noteworthy is the fact that other methods give the “best” results in regression and others in calibration.

**Keywords:** wind speed; robust methods; outliers; regression; calibration

## ΚΕΦΑΛΑΙΟ 1<sup>ο</sup> – ΕΙΣΑΓΩΓΗ

### 1.1. ΓΕΝΙΚΑ

Από τα τέλη της δεκαετίας του '70 το ενδιαφέρον του ανθρώπου έχει στραφεί προς την συστηματική αξιοποίηση των Ανανεώσιμων Πηγών Ενέργειας (ΑΠΕ), τόσο σε τοπικό όσο και σε εθνικό επίπεδο, και την ανάπτυξη πιο αξιόπιστων και οικονομικά αποδοτικότερων τεχνολογιών, καθόσον, κυρίως την τελευταία εικοσαετία, έχουν ενταθεί τα παγκόσμια περιβαλλοντικά προβλήματα. Πιο συγκεκριμένα, τα τελευταία χρόνια γίνονται σοβαρές επενδύσεις στον τομέα της αιολικής ενέργειας, κυρίως στις πιο ανεπτυγμένες χώρες, καθώς η αξιοποίησή της θεωρείται από τεχνικοοικονομικής απόψεως πιο συμφέρουσα και με τις λιγότερες κοινωνικές και περιβαλλοντικές επιπτώσεις από άλλες κλασσικές μορφές ενέργειας.

Σήμερα, είναι σαφές ότι ο παράκτιος χώρος αποτελεί μία ξεχωριστή χωρική ενότητα και έχει πολλές ιδιαιτερότητες που τον διαφοροποιούν από όλους τους υπολοίπους. Οι μεγάλες διαθέσιμες εκτάσεις και η αντίστοιχη έλλειψη θέσεων στην ξηρά, το σημαντικά πλουσιότερο αιολικό δυναμικό στη θάλασσα σε συνδυασμό με την αύξηση του μεγέθους και της αποδοτικότητας των ανεμογεννητριών καθώς και της μεγαλύτερης εμπειρίας στον τομέα αυτό, έχουν κάνει εφικτή τη δημιουργία θαλάσσιων αιολικών πάρκων.

Με βάση τα σημερινά δεδομένα, το αιολικό δυναμικό μιας περιοχής θεωρείται εκμεταλλεύσιμο ενεργειακά, όταν η μέση ετήσια ταχύτητα του ανέμου ξεπερνά την τιμή των 5.1 m/s. Στην Ελλάδα εντοπίζονται αρκετές περιοχές με ανάλογη ή και μεγαλύτερη διαθέσιμη ενέργεια, παρόλο που πρόκειται για μία σχετικά κλειστή και μικρή θαλάσσια περιοχή λόγω της παρουσίας νησιωτικών συμπλεγμάτων. Επομένως, είναι σημαντικό να γνωρίζουμε με όσο το δυνατόν μεγαλύτερη ακρίβεια και αξιοπιστία το αιολικό δυναμικό μιας περιοχής, έτσι ώστε με τον κατάλληλο και ορθό σχεδιασμό του αιολικού πάρκου να πραγματοποιείται η βέλτιστη εκμετάλλευση της αιολικής ενέργειας.

Ο πιο συνηθισμένος τρόπος για την εκτίμηση της αιολικής ενέργειας σε μια θαλάσσια περιοχή είναι η εγκατάσταση ενός μετεωρολογικού ιστού ή ενός πλωτήρα για τουλάχιστον ένα χρόνο λειτουργίας. Παρόλο που θεωρείται η πιο ποιοτική πηγή ανεμολογικών δεδομένων εν τούτοις οι μετρήσεις αφορούν περιορισμένη έκταση, έχουν γενικά περιορισμένη χωρική κάλυψη και συνοδεύεται από υψηλό κόστος εγκατάστασης, λειτουργίας και συντήρησης. Εξαιτίας αυτών των δυσκολιών, χρησιμοποιούνται επιπλέον δεδομένα που προέρχονται είτε από δορυφόρο με χρήση ραντάρ αλτιμετρίας, *sar* και σκεδασίμετρου είτε από ατμοσφαιρικά αριθμητικά μοντέλα πρόγνωσης/ανασύνθεσης με κατάλληλη επεξεργασία για να αποκτήσουν την επιθυμητή χωρική ανάλυση ανάλογα με την έκταση της περιοχής μελέτης. Παρόλα αυτά, και οι τελευταίες δύο πηγές ανεμολογικών δεδομένων χαρακτηρίζονται από σημαντικές αδυναμίες και προβλήματα. Πιο συγκεκριμένα, η χρονική συχνότητα εξαρτάται αυστηρά από την τροχιά κάθε δορυφόρου και υπάρχει πιθανότητα να μην είναι βολική λόγω των μεγάλων χρονικών διαστημάτων μεταξύ των διαβάσεων του δορυφόρου πάνω από μια συγκεκριμένη περιοχή. Αδυναμία επίσης παρουσιάζεται και αναφορικά με τη χωρική κάλυψη καθώς τα δορυφορικά δεδομένα θεωρούνται συχνά ακατάλληλα για μικρές και κλειστές περιοχές, όπως είναι το Αιγαίο Πέλαγος, λόγω της ευρείας παρουσίας νησιωτικών συμπλεγμάτων (οι δορυφορικές μετρήσεις θεωρούνται γενικά αναξιόπιστες στο σύνορο ξηράς – θάλασσας). Τα

ατμοσφαιρικά μοντέλα, από την πλευρά τους, παρουσιάζουν ανακρίβειες και αβεβαιότητες που μπορεί να οφείλονται σε σφάλματα κατά την παραμετροποίηση του μοντέλου, λάθη στις αρχικές συνθήκες (κυρίως κατά την εισαγωγή των πεδίων ανέμου) και στη διακριτοποίηση του πλέγματος [58].

Γενικά, για τη βέλτιστη αξιοποίηση των ανεμολογικών δεδομένων είναι απαραίτητη η συνδυασμένη χρήση τους. Σε αρκετές μελέτες μέχρι σήμερα, οι πλωτήρες χρησιμοποιούνται ως πηγή αναφοράς τόσο για την αξιολόγηση όσο και για τη ρύθμιση/διόρθωση των μετρήσεων που προέρχονται από λιγότερο ακριβή όργανα μέτρησης (δορυφόρος, αριθμητικό μοντέλο) συνήθως με την καθιέρωση μιας γραμμικής σχέσης διόρθωσης.

## 1.2. ΣΚΟΠΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ

Οι πιο συνηθισμένες μέθοδοι που προτείνονται και χρησιμοποιούνται για τη διόρθωση και ρύθμιση των δεδομένων από δορυφόρο και μοντέλο βασίζονται στην αρχή της γραμμικής παλινδρόμησης. Η πιο συνηθισμένη είναι η απλή γραμμική παλινδρόμηση, όπου όμως συχνά γίνεται μια βασική παράλειψη. Η ανεξάρτητη μεταβλητή θεωρείται ότι είναι μετρημένη χωρίς σφάλματα, κάτι που έρχεται σε αντίθεση με αυτό που πραγματικά συμβαίνει στην πλειοψηφία των περιπτώσεων στην πράξη. Η μέθοδος που λαμβάνει υπόψη της αυτό το γεγονός ονομάζεται Errors-In-Variables (*EIV*). Η μέθοδος *EIV* υποθέτει ότι υπεισέρχονται σφάλματα και στις δύο μεταβλητές και αποτελεί μια γενικότερη μεθοδολογία για την παραγωγή της δομικής σχέσης (*structural relation*) μεταξύ των μεταβλητών. Από την άλλη πλευρά, η εμπειρία δείχνει ότι τα ανεμολογικά δεδομένα συχνά περιλαμβάνουν (παρόλο τον ποιοτικό έλεγχο που ενδεχομένως γίνεται) λάθος μετρήσεις, οι οποίες μπορεί να επηρεάσουν την ευθεία παλινδρόμησης. Έτσι, εκτός από τις προαναφερθείσες μεθόδους παλινδρόμησης, αναλύονται στη συνέχεια και οι ανθεκτικές μέθοδοι, οι οποίες είναι σχεδιασμένες να ελέγχουν την επίδραση των εσφαλμένων μετρήσεων (μεγάλων καταλοίπων) στις μεταβλητές και παράλληλα να ικανοποιούν τις βασικές υποθέσεις της κλασσικής γραμμικής παλινδρόμησης.

Ο κύριος σκοπός αυτής της εργασίας είναι η εισαγωγή και χρήση των ανθεκτικών μεθόδων στην επεξεργασία και ανάλυση των ανεμολογικών δεδομένων από διαφορετικές πηγές και η σύγκρισή τους, σε ότι αφορά τη συμπεριφορά τους, με πιο παραδοσιακές και ευρέως χρησιμοποιούμενες μεθόδους παλινδρόμησης. Εξίσου σημαντικός στόχος είναι να βρεθεί, μέσα από δείκτες που χρησιμοποιούνται στη στατιστική ανάλυση, η μέθοδος που διορθώνει ορθότερα καινούριες μετρήσεις ανεμολογικών δεδομένων και προσαρμόζεται καλύτερα σε αυτές.

## 1.3. ΔΟΜΗ ΔΙΠΛΩΜΑΤΙΚΗΣ

Η διπλωματική αυτή εργασία αποτελείται από πέντε κεφάλαια, προκειμένου να αναλυθεί διεξοδικά η συγκριτική μελέτη των εναλλακτικών μεθόδων παλινδρόμησης και να εφαρμοστεί στη συνέχεια σε ανεμολογικά δεδομένα.

Στο **Δεύτερο Κεφάλαιο**, περιγράφονται οι εναλλακτικές πηγές και τα είδη των ανεμολογικών δεδομένων που χρησιμοποιούνται σήμερα στον τομέα της αιολικής ενέργειας για τη μέτρηση της ταχύτητας του ανέμου.

Στο **Τρίτο Κεφάλαιο**, αναλύεται το θεωρητικό υπόβαθρο των μεθόδων παλινδρόμησης που εφαρμόζονται στην ανάλυση των ανεμολογικών δεδομένων. Ειδικότερα, περιγράφονται εκτενώς η απλή μέθοδος παλινδρόμησης, η παλινδρόμηση που θεωρεί σφάλματα και στις δύο μεταβλητές και οι ανθεκτικές μέθοδοι παλινδρόμησης. Επίσης, παρουσιάζονται οι βασικές ιδιότητές τους μαζί με τα πλεονεκτήματα και τα μειονεκτήματά τους καθώς επίσης και η συμπεριφορά και η ερμηνεία των ασυνήθιστων παρατηρήσεων που υπεισέρχονται στις μετρήσεις.

Στο **Τέταρτο Κεφάλαιο**, περιγράφεται εκτενώς η διαδικασία της ανάλυσης που ακολουθήθηκε για τις επιλεγμένες περιοχές μελέτης του ελληνικού θαλάσσιου χώρου, παρατίθενται τα αριθμητικά αποτελέσματα της παρούσας εργασίας και τελικά, συγκρίνονται οι μέθοδοι παλινδρόμησης βάσει συγκεκριμένων στατιστικών δεικτών.

Το **Πέμπτο Κεφάλαιο** περιλαμβάνει τα τελικά συμπεράσματα και τις προτάσεις όπως προέκυψαν στην παρούσα εργασία από τη συγκριτική μελέτη των εναλλακτικών μεθόδων παλινδρόμησης με εφαρμογή σε ανεμολογικά δεδομένα του ελληνικού θαλάσσιου χώρου.

## ΚΕΦΑΛΑΙΟ 2<sup>ο</sup> – ΠΗΓΕΣ ΚΑΙ ΕΙΔΗ ΑΝΕΜΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Ο πρώτος που πραγματοποίησε συστηματικά τη συλλογή και χαρτογράφηση του ανέμου και της ωκεάνιας κυκλοφορίας ήταν ο *Matthew F. Maury* με τη δημοσίευση του πρώτου εκτεταμένου και λεπτομερούς βιβλίου σχετικά με την ωκεανογραφία “*The Physical Geography of the Sea*” (1855). Η πιο δημοφιλής μέθοδος εκείνης της περιόδου για τη μέτρηση της ταχύτητας και της διεύθυνσης του ανέμου ήταν με τη χρήση απλών ανεμόμετρων. Σήμερα, η συγκέντρωση ανεμολογικών στοιχείων επιτυγχάνεται με χρήση πολλών και διαφορετικών, ως προς την αρχή τους, μεθόδων και η ανάλυση τους γίνεται με τη βοήθεια ψηφιακών συστημάτων. Τα ανεμολογικά δεδομένα μπορεί να προέρχονται από τις παρακάτω πηγές:

- επιτόπιες μετρήσεις (*in situ measurements*) από πλωτούς ωκεανογραφικούς μετρητικούς σταθμούς (*buoys*), παρασυρόμενους πλωτήρες (*drifters*), πλατφόρμες εξόρυξης πετρελαίου, μετεωρολογικούς ιστούς (*meteorological masts*),
- μετρήσεις από πλοία εθελοντικής παρατήρησης (*Voluntary Observation Ships*),
- παρατηρήσεις τηλεανίχνευσης με χρήση *radar, lidar, sodar, sar* κ.α.,
- δεδομένα αριθμητικής προσομοίωσης (μετεωρολογικά μοντέλα).

### 2.1. ΕΠΙΤΟΠΙΕΣ ΜΕΤΡΗΣΕΙΣ

Στην Ελλάδα, την τελευταία δεκαετία, έχει ολοκληρωθεί η προσπάθεια δημιουργίας, εγκατάστασης και παρακολούθησης ενός δικτύου **πλωτών ωκεανογραφικών μετρητικών σταθμών (*buoys*)**, το οποίο αποτελεί ένα από τα βασικά μέρη του συστήματος *POSEIDON*. Φορέας υλοποίησης του έργου, το οποίο ολοκληρώθηκε το 2001, ήταν το Ελληνικό Κέντρο Θαλάσσιων Ερευνών.

Κάθε σταθμός *Sea-Watch* (**Εικόνα 2. 1**) είναι εφοδιασμένος με μετεωρολογικούς αισθητήρες (θερμοκρασία, ατμοσφαιρική πίεση, διεύθυνση ανέμου, ταχύτητα και ριπή ανέμου), αισθητήρα μέτρησης κυματικών παραμέτρων (*Seatex MRU – Motion Reference Unit*), ακουστικό ρευματογράφο σε βάθος 3 m με επιπλέον αισθητήρες θερμοκρασίας και αλατότητας. Η ενέργειά του εξασφαλίζεται από ηλιακές κυψέλες και επαναφορτιζόμενες μπαταρίες. Η συλλογή των δεδομένων γίνεται κάθε τρεις ώρες και ακολουθεί η αποστολή τους με το δορυφορικό σύστημα *Inmarsat-C* και με κινητή τηλεφωνία *GSM* [5].

Πιο συγκεκριμένα, σε ό,τι αφορά τις ανεμολογικούς παραμέτρους που μετρώνται από τους μετρητικούς σταθμούς του δικτύου *POSEIDON*, αυτές είναι οι εξής:

- Μέση ταχύτητα ανέμου  $U_w$  (m/sec). Είναι η μέση τιμή της ταχύτητας του ανέμου που καταγράφεται στη διάρκεια 600 sec.
- Ριπή ανέμου  $u_{Gust}$  (m/sec). Είναι η μέγιστη τιμή της ταχύτητας ανέμου στη διάρκεια των 600 sec της καταγραφής.
- Κατεύθυνση ανέμου  $\theta_{WIND}$  (deg). Είναι η κύρια κατεύθυνση του ανέμου κατά τη διάρκεια των 600 sec. Υποδηλώνει την κατεύθυνση από την οποία πνέει ο άνεμος.



**Εικόνα 2. 1.** Μετρητικός σταθμός τύπου Sea-Watch του συστήματος *POSEIDON*.

Το ανεμόμετρο που είναι τοποθετημένο στους πλωτούς ωκεανογραφικούς μετρητικούς σταθμούς βρίσκεται σε ύψος περίπου 3 m από το επίπεδο ηρεμίας της επιφάνειας της θάλασσας και λειτουργεί με περίοδο δειγματοληψίας 1 Hz [8]. Εν τούτοις, το κόστος αυτών των μετρήσεων, ιδιαίτερα σε μακροχρόνια βάση, παραμένει υψηλό, κυρίως λόγω της απαιτούμενης συντήρησης των ενεργειακών συστημάτων των μετρητικών οργάνων, δεδομένου ότι αυτά λειτουργούν σε άμεση επαφή με ένα δυσμενές περιβάλλον [8].

Μια πιο οικονομική λύση αποτελούν οι **παρασυρόμενοι πλωτήρες (*drifters*)** (Εικόνα 2. 2), καθώς προσφέρουν χαμηλή τιμή απόκτησης και λειτουργίας και μεγάλη ευελιξία στο σχεδιασμό και τον τρόπο τοποθέτησης στη θάλασσα. Χρησιμοποιούνται κυρίως για την καταγραφή των επιφανειακών ρευμάτων, αλλά διαθέτουν επιπρόσθετα και ατμοσφαιρικούς αισθητήρες για τη μέτρηση του ανέμου. Οι πλωτήρες αυτοί λειτουργούν αξιόπιστα 450 ημέρες κατά μέσο όρο.



**Εικόνα 2. 2.** Παρασυρόμενος πλωτήρας.

Οι πιο χρήσιμες και υψηλής ποιότητας επιτόπιες μετρήσεις για την πρόβλεψη της παράκτιας αιολικής ενέργειας είναι περίπου από ύψος 50 m και μεγαλύτερο χρησιμοποιώντας ένα **μετεωρολογικό ιστό (*meteorological mast*)** [23]. Ο ιστός αυτός αποτελείται από τα θεμέλια, μία πλατφόρμα με φόρτωση σκάφους, μετεωρολογικά και άλλα όργανα, φώτα πλοήγησης

και σήμανσης, καθώς και συναφή εξοπλισμό (**Εικόνα 2. 3**). Οι αισθητήρες που διαθέτει συλλέγουν δεδομένα σε κατακόρυφα προφίλ της ταχύτητας και διεύθυνσης του ανέμου, θερμοκρασίας αέρα και βαρομετρικής πίεσης, ταχύτητας θαλάσσιων ρευμάτων και θερμοκρασίας θαλάσσιου νερού [36].



**Εικόνα 2. 3.** Υπεράκτιος μετεωρολογικός ιστός.

Ο εν λόγω ιστός πρέπει να τοποθετηθεί κατάλληλα στην υποψήφια θέση που προτείνεται για εγκατάσταση υπεράκτιων ανεμογεννητριών για τουλάχιστον ένα χρόνο. Τα ανεμολογικά δεδομένα πρέπει να αναφέρονται κοντά ή στο ύψος του δρομέα (*hub*) των ανεμογεννητριών, το οποίο μπορεί να είναι 80 – 100 m πάνω από την επιφάνεια της θάλασσας, και να συγκριθούν με δεδομένα που έχουν συλλεχθεί από έναν κοντινό μετεωρολογικό σταθμό για την ίδια περίοδο. Ο συσχετισμός των δύο ομάδων δεδομένων χρησιμοποιείται για τη μεταφορά των μακροπρόθεσμων δεδομένων του σταθμού στην ενδεχόμενη τοποθεσία [15]. Παρόλα αυτά, τα δεδομένα είναι ανεπαρκή, γιατί τέτοιου είδους μετρητικά συστήματα θεωρούνται ακριβά για να εγκατασταθούν στη θάλασσα, και σημειακά καθώς δεν μπορούν να αναπαραστήσουν τη χωρική μεταβλητότητα του ανέμου.

Τέλος, υπάρχει η δυνατότητα τοποθέτησης κατάλληλων οργάνων πάνω σε αγκυρωμένες πλατφόρμες εξόρυξης πετρελαίου κ.λπ., προκειμένου να πραγματοποιηθούν ανεμολογικές μετρήσεις.

Αξίζει να σημειωθεί ότι, στην Ελλάδα, σποραδικές μετρήσεις ανεμολογικών χαρακτηριστικών γίνονται και από άλλους δημόσιους και ερευνητικούς φορείς (Εθνικό Μετσόβιο Πολυτεχνείο, Κέντρο Ανανεώσιμων Πηγών Ενέργειας, κ.λπ.), κυρίως στα πλαίσια ερευνητικών προγραμμάτων. Οι μετρήσεις αυτές είναι τοπικού χαρακτήρα και, ως εκ τούτου, είναι επαρκείς μόνο για τη μελέτη των τοπικών χαρακτηριστικών του ανεμολογικού καθεστώτος [8].



## 2.2. ΜΕΤΡΗΣΕΙΣ ΑΠΟ ΠΛΟΙΑ ΕΘΕΛΟΝΤΙΚΗΣ ΠΑΡΑΤΗΡΗΣΕΙΣ

Τα πλοία εθελοντικής παρακολούθησης (*Voluntary Observation Ships – VOS*) ανήκουν σε ένα διεθνές πρόγραμμα, όπου συμμετέχουν 49 χώρες, υπό την αιγίδα του Παγκόσμιου Μετεωρολογικού Οργανισμού (*World Meteorological Organization – WMO*). Τα πλοία αυτά διεξάγουν μετεωρολογικές και ωκεανογραφικές παρατηρήσεις με τη βοήθεια είτε οργάνων (ανεμόμετρων) είτε μέσω οπτικών παρατηρήσεων από εξειδικευμένο προσωπικό (κλίμακα *Beaufort*) και στη συνέχεια τις μεταδίδουν προκειμένου να χρησιμοποιηθούν από μετεωρολόγους για την πρόγνωση του καιρού, από ωκεανογράφους, αλιείς και άλλες υπηρεσίες.

Θα πρέπει να σημειωθεί ότι συστηματικά σφάλματα υπάρχουν και στις δύο μεθόδους καταγραφής ενώ έχει γίνει μεγάλη προσπάθεια για περαιτέρω διορθώσεις των δεδομένων. Ενδεικτικά αναφέρουμε ότι για τα δεδομένα *VOS*, βρέθηκε ότι ένα τυπικό σφάλμα για μία παρατήρηση του ανέμου είναι περίπου 2.2 m/s (μετά από διόρθωση για το ύψος του ανεμόμετρου πάνω από τη θάλασσα) [39]. Για τις παρατηρήσεις όπως καταγράφονται, τα σφάλματα είναι 15% μεγαλύτερα, 2.5 m/s. Αυτό δείχνει ότι παρ' όλες τις ποικίλες επιδράσεις της παραμόρφωσης της ροής του ανέμου γύρω από το πλοίο, η διόρθωση των δεδομένων για το ύψος του ανεμόμετρου πράγματι ελαττώνει τα σφάλματα.

Τα πλοία εθελοντικής παρακολούθησης δίνουν ακόμα μεγάλο όγκο δεδομένων από οπτικές εκτιμήσεις του ανέμου από την κλίμακα *Beaufort*. Την τελευταία δεκαετία χρησιμοποιείται μια καινούργια μέθοδος μετατροπής κλίμακας *Beaufort* σε ταχύτητα ανέμου, την οποία εξήγαγε ο *Lindau* (1995) από τη σύγκριση μετρήσεων της ταχύτητας του ανέμου από ωκεάνιους σταθμούς καιρού (*Offshore Weather Stations – OWS*) με οπτικές μετρήσεις από τα εμπορικά πλοία που έπλεαν στην περιοχή.

Παρόλο που η νέα γενιά των μετεωρολογικών δορυφόρων συμβάλλει στην αντιμετώπιση σημαντικών προβλημάτων, όπως είναι η έλλειψη δεδομένων σε μεγάλα τμήματα ωκεανών, δεδομένα που προέρχονται από συμβατικές μεθόδους, και προπαντός από πλοία εθελοντικής παρακολούθησης, παραμένουν ουσιώδη. Αυτές οι παρατηρήσεις παρέχουν σημαντικές επίγειες πληροφορίες, που οι δορυφόροι δεν μπορούν να παρατηρήσουν, και συνεισφέρουν ουσιαστικά τόσο στα αριθμητικά μοντέλα μετεωρολογικής πρόγνωσης όσο και σε αναφορές σε πραγματικό χρόνο.

## 2.3. ΠΑΡΑΤΗΡΗΣΕΙΣ ΤΗΛΕΑΝΙΧΝΕΥΣΗΣ

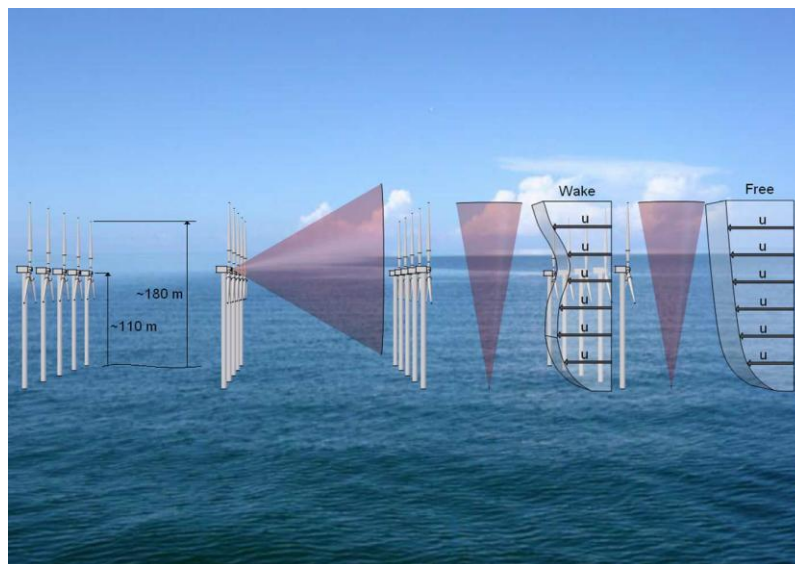
Με τον όρο τηλεπισκόπηση ή τηλεανίχνευση (*remote sensing*) χαρακτηρίζεται η επιστήμη και τεχνική που ασχολείται με τις αρχές, τις μεθόδους και τα όργανα με τα οποία επιτυγχάνεται από μακριά η συλλογή, επεξεργασία και ανάλυση πλήθους ποιοτικών και μετρητικών πληροφοριών για τη γη, τους ωκεανούς, την ατμόσφαιρα και το περιβάλλον γενικότερα, αλλά και για οποιοδήποτε αντικείμενο, γεγονός και συμβάν και στηρίζονται στην αλληλοεπίδραση της ηλεκτρομαγνητικής ακτινοβολίας με την ύλη. (Ρόκος Δ., Αθήνα 1996).

Η ηλεκτρομαγνητική ακτινοβολία είναι η ενέργεια η οποία εκπέμπεται από πηγή και μεταδίδεται στο διάστημα με την ταχύτητα του φωτός. Η πηγή αυτή είναι η περιοχή του ηλεκτρομαγνητικού φάσματος. Κάθε παρατηρούμενο αντικείμενο, μοιράζει τις προσπίπτουσες ακτινοβολίες σε ανακλώμενες, διερχόμενες και απορροφούμενες. Επιπλέον,

εκπέμπει μια δική του ακτινοβολία που καθορίζεται από την απόλυτη θερμοκρασία του. Όλες οι ιδιότητες αυτές, αποτελούν τα φασματικά χαρακτηριστικά του ή τη φασματική υπογραφή, ο καθορισμός της οποίας βοηθά στη διαδικασία διαχωρισμού του αντικειμένου από το περιβάλλον του. Γνωρίζοντας λοιπόν την φασματική υπογραφή ενός αντικειμένου και χρησιμοποιώντας δέκτες στην περιοχή της ακτινοβολίας που εκπέμπει ή ανακλά το αντικείμενο αυτό, καταγράφουμε ένα σημαντικό μέρος από την καμπύλη ανακλαστικότητας του, με αποτέλεσμα να αναγνωρισθεί η καμπύλη και επομένως να προσδιορισθεί το αντικείμενο[6].

Η ιστορία των τεχνικών τηλεανίχνευσης για τη μέτρηση του ανέμου στη θάλασσα ξεκίνησε παράλληλα με την ανάπτυξη της παράκτιας αιολικής ενέργειας. Για παράδειγμα, ο δορυφόρος *ERS-1* της *ESA* (*European Space Agency*), που ήταν εξοπλισμένος με ραντάρ συνθετικού ανοίγματος (*Synthetic Aperture Radar – SAR*) και σκεδασίμετρο (*scatterometer*), τέθηκε σε λειτουργία το 1991 για ερευνητικούς σκοπούς. Όσο η τεχνολογία των ανεμογεννητριών εξελίσσεται, το ύψος των ανεμογεννητριών αυξάνεται περίπου 5 m το χρόνο από το 1990 ενώ στις μεγάλες ανεμογεννήτριες ισχύος από 1 έως 2.5 MW, που έχουν εγκατασταθεί από το 2006, ο δρομέας βρίσκεται περίπου στα 110 m και το άκρο των πτερυγίων μπορεί να φθάσει τα 180 m, όπως φαίνεται και στην **Εικόνα 2. 4**. Για το λόγο αυτό κρίνεται απαραίτητη η γνώση του πεδίου ροής του ανέμου σε υψηλά επίπεδα της ατμόσφαιρας. Καθώς όμως το κόστος λειτουργίας ενός σταθερού μετεωρολογικού ιστού είναι ιδιαίτερα υψηλό και δεν είναι εφικτό να τοποθετηθεί και να λειτουργήσει σε τέτοια ύψη, πραγματοποιούνται παρατηρήσεις μέσω τηλεανίχνευσης για την εκτίμηση του ανέμου. Ανάλογα με το χωρικό μέσο που λειτουργούν οι δέκτες, οι εν λόγω παρατηρήσεις διακρίνονται σε τρεις κύριες κατηγορίες:

- Τις επίγειες τεχνικές τηλεανίχνευσης
- Τις αερομεταφερόμενες συσκευές τηλεανίχνευσης
- Τις δορυφορικές μετρήσεις



**Εικόνα 2. 4.** Διαφορετικές διατάξεις σχετικά με τις παρατηρήσεις στη βιομηχανία του ανέμου που βασίζεται σε επίγειες τεχνικές τηλεανίχνευσης. Στα αριστερά εμφανίζονται οι προσεγγιστικές διαστάσεις μιας ανεμογεννήτριας, στη μέση οι κώνοι παρατήρησης του *LiDAR* τοποθετημένο πάνω στο δρομέα ή την άτρακτο και στο έδαφος, στα δεξιά φαίνεται το κατακόρυφο προφίλ του ανέμου επηρεασμένο από το φαινόμενο *wake* και άκρως δεξιά το αδιατάρακτο κατακόρυφο προφίλ του ανέμου [29].

### 2.3.1. Επίγειες Τεχνικές Τηλεανίχνευσης

Οι επίγειες τεχνικές τηλεανίχνευσης χρησιμοποιούνται για τη μέτρηση των ανεμολογικών χαρακτηριστικών. Τα πιο συνηθισμένα όργανα είναι τα *SoDAR* (*Sound Detection and Ranging*) και τα *LiDAR* (*Light Detection and Ranging*), τα οποία βασίζονται στην εκμετάλλευση του φαινομένου *Doppler* (*Doppler effect*) με ηχητική ή οπτική διέγερση αντίστοιχα.

Το *SoDAR* (Εικόνα 2. 5) λαμβάνει ανακλάσεις από την ατμοσφαιρική σκέδαση του ήχου. Μέρος των μεταδιδόμενων παλμών του ήχου ανακλώνται πίσω προς τους ανιχνευτές ήχου του οργάνου λόγω των διακυμάνσεων της θερμοκρασίας στα διάφορα στρώματα της ατμόσφαιρας. Αυτές οι διακυμάνσεις στη θερμοκρασία προκαλούνται από θερμικά επαγόμενο στροβιλισμό. Στην περίπτωση των οργάνων *LiDAR* (Εικόνα 2. 6), εκπέμπεται μία παλμική ακτινοβολία *laser* σε ένα σημείο εστίασης στην ατμόσφαιρα και, σε όλη τη διαδρομή, αυτή διασκορπίζεται από αερολύματα, σκόνη, σταγόνες νερού και εξασθενεί. Συγκρίνοντας λοιπόν την εισερχόμενη και την εξερχόμενη ακτίνα *laser*, υπολογίζεται το μέγεθος της επίδρασης του φαινομένου *Doppler* στη συχνότητα της ακτίνας, που προκαλείται από την κίνηση των αερολυμάτων, και συνεπώς, είναι δυνατή η εκτίμηση της ταχύτητας του αέρα ακτινικά στο οπτικό πεδίο της συσκευής.



Εικόνα 2. 5. Τοποθέτηση συσκευής *SoDAR* σε πλατφόρμα ανοικτά των ακτών *Georgetown*, Βόρεια Καρολίνα.



**Εικόνα 2. 6.** Εγκατεστημένο *LiDAR* (μοντέλο *ZephIR*) σε υποσταθμό πλατφόρμας στο πρώτο παράκτιο αιολικό πάρκο της Σκωτίας, το *Robin Rigg* (180MW) στο *Solway Firth*.

Οι συσκευές αυτές μπορούν να εγκατασταθούν σε χαμηλές πλατφόρμες ή να μεταφερθούν από πλοία για την παρατήρηση του ανέμου. Το *LiDAR* μπορεί επίσης να τοποθετηθεί στην άτρακτο της τουρμπίνας προκειμένου να λειτουργεί με την πνοή του ανέμου προς τα πάνω (*upwind*) ή προς τα κάτω (*downwind*). Με αυτό τον τρόπο, δεν μελετάται μόνο ο ελεύθερος ατμοσφαιρικός αέρας αλλά και το φαινόμενο *wake* που προκαλείται από την τουρμπίνα ή το αιολικό πάρκο, πληροφορία που είναι ιδιαίτερα σημαντική για τους κατασκευαστές των ανεμογεννητριών λόγω των διαφόρων τύπων φορτίων που εντείνονται από τις επιπτώσεις του φαινομένου *wake*. Αυτές οι τεχνικές παρέχουν το πλεονέκτημα της κινητικότητας. Βελτιώνονται συνεχώς για να είναι εύκολα στη μεταφορά τους και για το λόγο αυτό χρησιμοποιούνται για να πραγματοποιήσουν μετρήσεις του προφίλ του ανέμου σε ένα αιολικό πάρκο μεταφέροντας το όργανο σε διάφορες θέσεις εντός ή πλησίον της περιοχής του πάρκου. Η σύγχρονη τεχνολογία παρέχει περαιτέρω προοπτικές για το μελλοντικό σχεδιασμό των αιολικών πάρκων και, ενδεχομένως, τη δυνατότητα να αποτελέσουν βασικό μέρος μιας ανεμογεννήτριας [29].

Σε σχέση με τους ιστούς, τα *SoDAR* και τα *LiDAR* φέρουν το πλεονέκτημα της πιο εύκολης αλλά και πιο φθηνής τοποθέτησης και εγκατάστασης, της πιο απλής συντήρησης και της απόλυτης εγγύησης ότι δεν είναι ευάλωτα σε δυσμενείς καιρικές συνθήκες (κρύο, βροχή, χιόνι, αέρας). Επίσης, πρόκειται για συσκευές μικρής μάζας και όγκου, αθόρυβες και διακριτικές κατά τη λειτουργία τους, παρέχουν τη δυνατότητα μέτρησης ζητούμενων μεγεθών σε ποικίλα ύψη, σταθερή υψηλή ανάλυση πλήρους φάσματος και αυτόματο



φιλτράρισμα δεδομένων. Αντίθετα, οι ιστοί συχνά χρειάζονται άμεση συντήρηση λόγω του σχήματός τους, είναι ευάλωτοι σε καιρικές συνθήκες στις οποίες κυριαρχούν δυνατοί άνεμοι κάτι που καθιστά αναγκαία τη μελέτη για αντοχή σε ταλαντώσεις από μηχανικά και δυναμικά φορτία και τη χρήση πιο ακριβών υλικών και αδυνατούν να μετρήσουν ταχύτητες σε ύψη που σχετίζονται με την αιολική ενέργεια, όπως ήδη αναφέρθηκε. Οι συσκευές *SoDAR* σε σύγκριση με τις *LiDAR* έχουν το εξής μειονέκτημα: οποιαδήποτε ηχητική παρεμβολή, είτε λόγω εξωτερικών πηγών, είτε λόγω της τοπογραφίας (αντίλαλος) μπορεί να επηρεάσει τη λειτουργία της συσκευής οδηγώντας σε μετρήσεις με σημαντική απόκλιση από τις πραγματικές [9]. Εξαιτίας της εξάρτησης των συσκευών *SoDAR* με τις εξωτερικές συνθήκες, συστήνεται πάντοτε η επαναβαθμονόμηση του συστήματος με έναν μετεωρολογικό πύργο πριν την εκτίμηση του αιολικού δυναμικού [29].

Μία πρότυπη πλωτή συσκευή πραγματικών διαστάσεων, γνωστή ως *Floating LiDAR* (*FLiDAR*) (Εικόνα 2. 7), αναπτύχθηκε και κατασκευάστηκε από έμπειρους ερευνητές της εταιρείας *OWA* (*Offshore & Wind Assistance*), θυγατρικής της βελγικής *GeoSea*, σε συνεργασία με τον παγκόσμιο σύμβουλο ανανεώσιμων πηγών ενέργειας, την 3E. Τον Οκτώβριο του 2011, δοκιμάστηκε με επιτυχία στις συνθήκες της Βόρειας Θάλασσας σε απόσταση 15 km από την ξηρά, αφού εγκαταστάθηκε δίπλα σε ένα σταθερό *LiDAR* (*WINDCUBEv2 LiDAR* της *Leosphere*) προκειμένου να ελεγχθεί η ικανότητα επιβίωσης σε ακραίες συνθήκες περιβάλλοντος και να εγκριθεί η ακρίβεια των μετρήσεων της ταχύτητας του ανέμου. Το *FLiDAR* έχει ως στόχο να μειώσει σημαντικά το απαγορευτικό κόστος ενός μετεωρολογικού ιστού στη θάλασσα και να παρέχει μεγαλύτερη ευελιξία σε προγραμματιστές και διαχειριστές κεφαλαίων που ασχολούνται με την υπεράκτια αιολική ενέργεια. Επιπλέον, το *FLiDAR* μπορεί να παρέχει καλύτερες και πληρέστερες εκτιμήσεις της ταχύτητας του ανέμου (κατακόρυφο προφίλ, οριζόντιο προφίλ, μετρήσεις στροβιλισμού) και να αναδιατάσσεται γρήγορα και εύκολα για πιο αποδοτικές και ευέλικτες εκστρατείες μετρήσεων. Από τα τέλη του 2012 έχει ξεκινήσει και η εμπορική του χρήση [63].

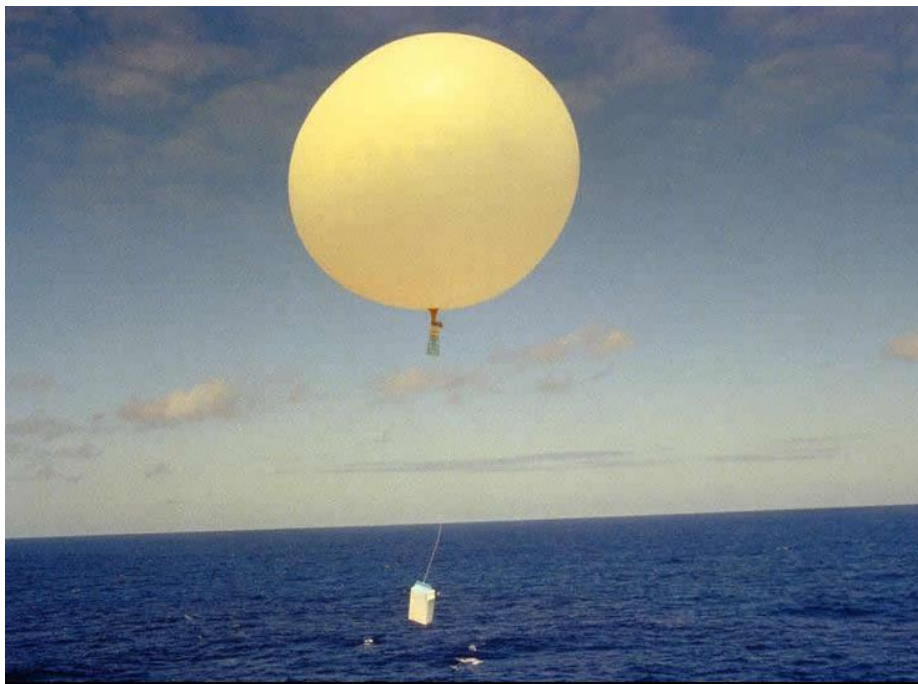


Εικόνα 2. 7. Η πλωτή συσκευή *FLiDAR* σε λειτουργία.

Αξίζει να αναφερθεί ότι, εκτός από τη μέτρηση των επιφανειακών ρευμάτων, έχει αποδειχθεί ότι τα **ραντάρ υψηλών συχνοτήτων (*high frequency*)** είναι ιδιαίτερα αποτελεσματικά στη μέτρηση της διεύθυνσης του ανέμου σε κλίμακες της τάξης του ενός μέχρι μερικών χιλιομέτρων και πάνω από εκτάσεις εκατοντάδων τετραγωνικών χιλιομέτρων. Η δυνατότητα εκτίμησης της διεύθυνσης του ανέμου σε υψηλή ανάλυση και για μεγάλες εκτάσεις καθιστούν χρήσιμα τα ραντάρ υψηλών συχνοτήτων, επίσης είναι επίγεια και συλλέγουν ωκεανογραφικές μετρήσεις από απόσταση που μπορούν να αξιοποιηθούν από διάφορες ειδικότητες όπως είναι οι μετεωρολόγοι, οι ωκεανογράφοι, οι παράκτιοι διαχειριστές.

### 2.3.2. Αερομεταφερόμενες συσκευές τηλεανίχνευσης

Σε χιλιάδες χερσαία και θαλάσσια σημεία της γης, πραγματοποιούνται καθημερινές και συνεχείς καταγραφές του καιρού και των μετεωρολογικών παραμέτρων από ειδικευμένους παρατηρητές, αυτόματους μετεωρολογικούς σταθμούς και σημαδούρες. Αυτές οι καταγραφές μεταδίδονται σε κωδικοποιημένη μορφή. Σε κάποια σημεία καταγραφής εκτελούνται και οι λεγόμενες **ραδιοβολήσεις**, συνήθως κάθε 12 ώρες. Απελευθερώνονται δηλαδή μετεωρολογικά μπαλόνια (**Εικόνα 2. 8**) εξοπλισμένα με όργανα καταγραφής και πομπό, τα οποία ανέρχονται μέσα στην ατμόσφαιρα και καταγράφουν και μεταδίδουν προς το έδαφος σε συνεχή ροή τις τιμές διαφόρων μετεωρολογικών παραμέτρων καθ' ύψος, μία εκ των οποίων είναι και ταχύτητα και η διεύθυνση του ανέμου. Στη συνέχεια όλα τα δεδομένα συλλέγονται και αποκωδικοποιούνται και με αυτόματες στατιστικές μεθόδους απαλείφονται οι εσφαλμένες μετρήσεις [64].



**Εικόνα 2. 8.** Μετεωρολογικό μπαλόνι έτοιμο να τεθεί σε λειτουργία για τη μέτρηση του ανέμου.

Τα συστήματα **ραντάρ συνθετικού ανοίγματος (*Synthetic Aperture Radar – SAR*)** προτάθηκαν από τον *Carl Wiley* το 1951. Τα συστήματα *SAR* επιτυγχάνουν υψηλή διακριτικότητα στην παρατήρηση της επιφάνειας της γης με την επεξεργασία πολλών διαδοχικών παλμών, που εκπέμπονται από τον κινούμενο φορέα που μεταφέρει το ραντάρ.

Συνήθως η παρατήρηση της γης γίνεται από αεροσκάφος (*airborne*) ή δορυφόρο (*spaceborne*). Το σημαντικότερο πλεονέκτημα σε ένα ραντάρ συνθετικής απεικόνισης είναι ότι μπορεί να παρέχει εικόνες εκτενών περιοχών της γήινης επιφάνειας ανεξάρτητα από τις καιρικές συνθήκες και της νεφοκάλυψης που επικρατούν στην ατμόσφαιρα ή εάν είναι ημέρα ή νύχτα. Αυτή η ικανότητα παροχής εικόνων, ανεξάρτητα από τις καιρικές συνθήκες, οφείλεται στη χρησιμοποίηση μήκων κύματος που προέρχονται από την *X-band* (3 cm), *C-band* (6 cm) και την *L-band* (24 cm) των μικροκυματικών συχνοτήτων. Για ένα αερομεταφερόμενο ραντάρ, που κινείται μέσα στην ζώνη της τροπόσφαιρας, η διαδρομή διάδοσης βρίσκεται μέσα σε έντονες βροχοπτώσεις, ένα φαινόμενο που δεν είναι τόσο έντονο, παρά μόνο σε ακραίες καιρικές συνθήκες (π.χ. τυφώνες), σε περίπτωση που το ραντάρ είναι εγκατεστημένο σε δορυφόρο[4].

### 2.3.3. Δορυφορικές μετρήσεις

#### 2.3.3.1. Γενικά

Σε έντονη ανάπτυξη βρίσκονται οι μέθοδοι παρακολούθησης του θαλάσσιου περιβάλλοντος που βασίζονται στη δορυφορική τηλεπισκόπηση (*satellite remote sensing*). Η χρήση δορυφόρων για τη συλλογή ανεμολογικών χαρακτηριστικών έχει αρχίσει πριν αρκετά χρόνια, κυρίως στις ΗΠΑ, ενώ ο πρώτος δορυφόρος που χρησιμοποιήθηκε συστηματικά για τη συλλογή κυματικών και ανεμολογικών δεδομένων είναι ο *GEOSAT* του Αμερικανικού Ναυτικού, ο οποίος τέθηκε σε τροχιά τον Μάρτιο του 1985. Οι δορυφόροι που καλύπτουν ευρείες γεωγραφικές εκτάσεις και χρησιμοποιούνται για σκοπούς μελέτης και πρόβλεψης ανεμολογικού και κυματικού κλίματος είναι ο Γαλλοαμερικανικός *TOPEX/Poseidon*, ο διάδοχός του *JASON*, και οι Ευρωπαϊκοί *ERS-1* και *ERS-2*. Κύριο χαρακτηριστικό των δεδομένων που συλλέγονται από δορυφόρους αποτελεί η ευρεία χωρική διασπορά και η μεγάλη περίοδος δειγματοληψίας [8].

Στη βιβλιογραφία, έχουν προταθεί ορισμένα όργανα για την εκτίμηση του αιολικού δυναμικού, τα οποία διακρίνονται σε ενεργητικούς δέκτες όπως είναι τα ραντάρ συνθετικού ανοίγματος (*SAR*), τα σκεδασίμετρα (*scatterometers*), τα ραντάρ αλτιμετρίας (*radar altimeters*) και παθητικούς δέκτες όπως είναι οι παθητικοί μικροκυματικοί δέκτες. Οι δέκτες αυτοί μπορούν να πραγματοποιήσουν μετρήσεις ανεξαρτήτως χρόνου και καιρικών συνθηκών, αλλά παρέχουν διαφορετικές χωρικές αναλύσεις και χρονικές καλύψεις [15].

#### 2.3.3.2. Ραντάρ Συνθετικού Ανοίγματος (*SAR*)

Τα συστήματα *SAR* συνθέτουν τεχνητά κεραίες ιδιαίτερα μεγάλου μήκους εκμεταλλεζόμενα την κίνηση του δορυφόρου και την αρχή του φαινομένου *Doppler*, με αποτέλεσμα τη σημαντική αύξηση της χωρικής διακριτικής ικανότητας στη διεύθυνση των αζιμούθιων.

Το πρώτο διαστημικό σύστημα *SAR* που τέθηκε σε τροχιά εγκαταστάθηκε από την *NASA* στον γεωστατικό δορυφόρο *SEASAT* το 1978 και λειτούργησε για περίπου έξι εβδομάδες [40]. Παρά την σύντομη διάρκεια της αποστολής, τέθηκαν οι βάσεις για την κατασκευή εξελιγμένων εικονοληπτικών συστημάτων *SAR* στη δεκαετία του 1980 και ιδιαίτερα στο πλαίσιο του διαστημικού προγράμματος του διαστημικού λεωφορείου της *NASA* (*SIR-A* και *SIR-B*) [22]. Κατά τη διάρκεια της δεκαετίας του 1990, με τη διαθεσιμότητα στο ευρύ επιστημονικό κοινό δεδομένων από τους δορυφόρους *ERS-1* (ESA, 1991), *ERS-2* (ESA,

1995), *JERS-1* [48] και *RADARSAT-1* [50] παρατηρείται ιδιαίτερο ενδιαφέρον για στην περιοχή των μικροκυματικών αισθητήρων.

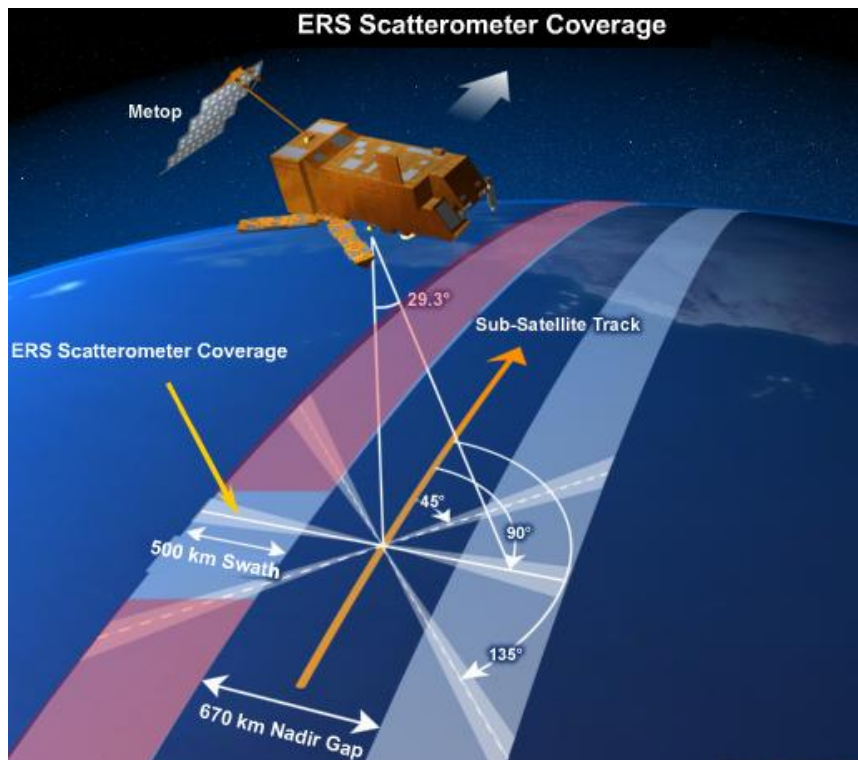
Την επιτυχία των προηγούμενων αποστολών ακολούθησε η κατασκευή και εκτόξευση εξελιγμένων δορυφορικών συστημάτων εικονοληπτικών *SAR*, αρχικά με τους *ENVISAT ASAR* (Ευρώπη) και *ALOS PALSAR* (Ιαπωνία) και στη συνέχεια με τους *RADARSAT-2* (Καναδάς) [61], *TerraSAR-X* (Γερμανία) [21] και *Cosmo-SkyMed* (Ιταλία) [17]. Αυτοί οι τύποι ραντάρ συνδυάζουν τη μέτρηση από το συντελεστή οπισθοσκέδασης (*backscattering coefficient*) με την ανάλυση συχνότητας των σημάτων και έτσι πετυχαίνουν υψηλή διακριτική ικανότητα. Θα πρέπει να αναφερθεί ότι δεν είναι δυνατή η εκτίμηση της ταχύτητας και της διεύθυνσης του ανέμου από κοινού εξαιτίας της ύπαρξης μίας μόνο κεραίας σε αυτούς τους αισθητήρες και επομένως μία από τις δύο παραμέτρους θα πρέπει να είναι γνωστή *a priori*. Μία προτεινόμενη μέθοδος για την ανάκτηση της διεύθυνσης του ανέμου βασίζεται στη *multi – resolution* ανάλυση των εικόνων *SAR* χρησιμοποιώντας ένα μετασχηματισμό *Fourier* [15].

### 2.3.3.3. Σκεδασίμετρο (*scatterometer*)

Το σκεδασίμετρο (**Εικόνα 2. 9**) είναι ένας ενεργητικός μικροκυματικός δέκτης, ο οποίος αποστέλλει ένα σήμα και μετράει το ποσοστό του σήματος που επιστρέφει μετά από την αλληλεπίδρασή του με το προς παρατήρηση αντικείμενο για διάφορες γωνίες πρόσπτωσης και παρέχει διανύσματα ανέμου μέσω εμπειρικών σχέσεων (συναρτήσεων μοντέλου) που σχετίζουν το ποσοστό αυτό της ενέργειας που επιστρέφει στο δορυφόρο (οπισθοσκέδαση κατά *Bragg*) με την ταχύτητα και την κατεύθυνση του ανέμου στα 10 m πάνω από την επιφάνεια της θάλασσας, σε συνθήκες ουδέτερης ευστάθειας. Ως αρχικό παράγωγο, το όργανο αυτό, αποδίδει πολλαπλές πιθανές κατευθύνσεις ανέμου λόγω ασαφειών και προσδιορίζεται η πιο πιθανή τιμή του.

Μεταξύ των διάφορων δορυφορικών αισθητήρων, το σκεδασίμετρο είναι το μοναδικό που έχει την ικανότητα να καθορίζει και να παρέχει συνεχείς, ακριβείς και υψηλής ανάλυσης μετρήσεις της ταχύτητας και της κατεύθυνσης του ανέμου πάνω από ηπειρωτικές και θαλάσσιες επιφάνειες ανεξαρτήτως καιρικών συνθηκών. Αυτές οι παρατηρήσεις του ανέμου έχουν μια ευρεία ποικιλία εφαρμογών, συμπεριλαμβανομένων των μετεωρολογικών προβλέψεων, τη θαλάσσια ασφάλεια, την εμπορική αλιεία, την πρόβλεψη και παρακολούθηση του φαινομένου *El Nino*, καθώς και μακροπρόθεσμες μελέτες για το κλίμα. Το 1978 ήταν η πρώτη φορά που σκεδασίμετρο τοποθετημένο σε δορυφόρο (*SEASAT*) χρησιμοποιήθηκε για τον προσδιορισμό της ταχύτητας και διεύθυνσης του ανέμου πάνω από την επιφάνεια ωκεανού. Από τότε έχουν σημειωθεί αρκετές αποστολές, όπως το *Scat* στο δορυφόρο *ERS-1* και τον *ERS-2*, το *NSCAT* στον *ADEOS-I*, *SeaWinds* στον *QuikSCAT* και *SeaWinds* στον *ADEOS-II*, η εξαιρετική ακρίβεια των οποίων οδηγεί στην ανάπτυξη νέων εφαρμογών.





Εικόνα 2. 9. Αρχή λειτουργίας του σκεδασίμετρου.

#### 2.3.3.4. Ραντάρ αλτίμετρο (*radar altimeter*)

Η διαθεσιμότητα και η μεγάλη κάλυψη που προσφέρει το ραντάρ αλτίμετρο έδωσαν νέα ώθηση και ενδιαφέρον για την ανάκτηση ωκεανογραφικών πληροφοριών από τις δορυφορικές μετρήσεις [8]. Το ραντάρ αλτίμετρο χρησιμοποιείται ευρέως για τη μέτρηση της στάθμης της θάλασσας και τις μεταβολές της σε παγκόσμια κλίμακα. Το 1975 εκτοξεύθηκε ο δορυφόρος *GEOS-3* της *NASA* με το πρώτο λειτουργικό ραντάρ αλτιμετρίας και από τότε υπάρχει μία μεγάλη χρονοσειρά μετρήσεων για τη στάθμη της θάλασσας. Εκτός όμως από αυτές τις μετρήσεις, χρησιμοποιώντας την πληροφορία από το σήμα ανασκέδασης, μπορεί επίσης να εκτιμηθεί το ύψος των θαλάσσιων κυματισμών και να εξαχθεί η ταχύτητα του ανέμου [56].

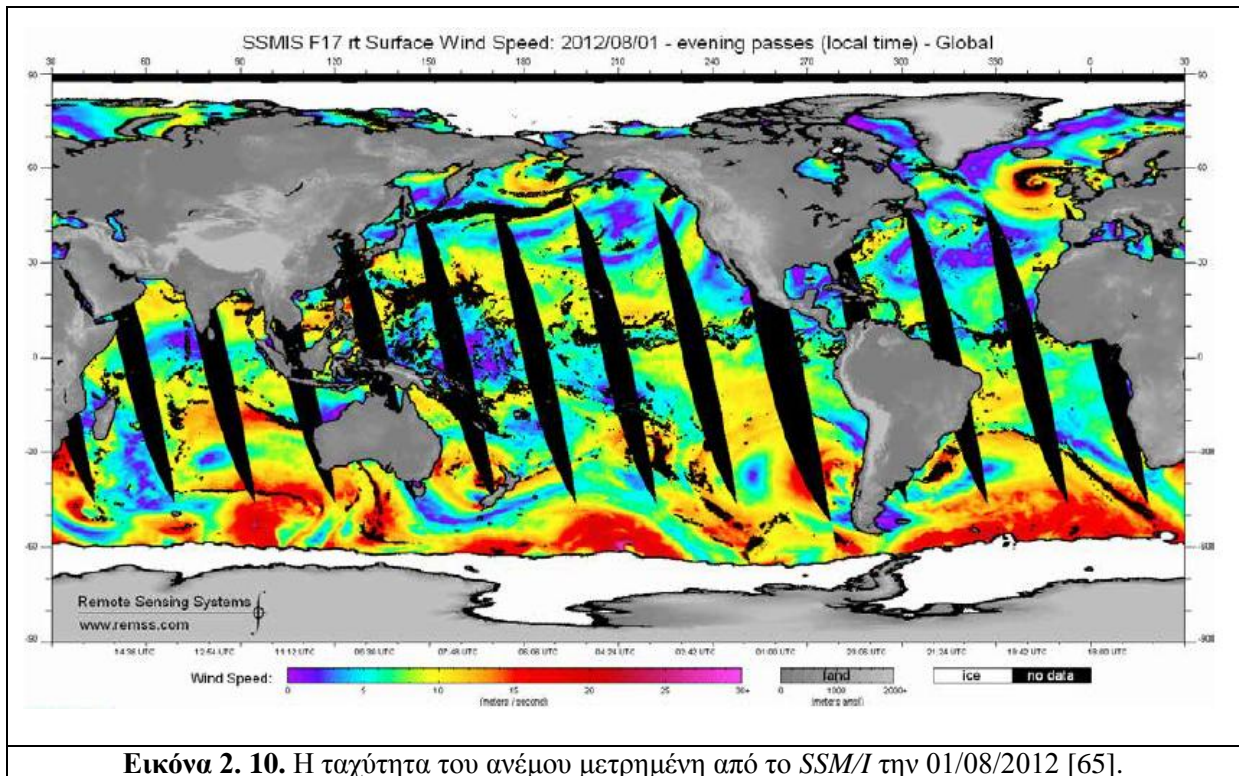
Οι σχετικοί αλγόριθμοι και η αξιοπιστία των μετρήσεων σε σχέση με τις επιτόπιες μετρήσεις έχουν βελτιωθεί σημαντικά. Οι περισσότεροι αλγόριθμοι που χρησιμοποιούνται σήμερα προέρχονται εμπειρικά συσχετίζοντας δεδομένα επιτόπιων μετρήσεων, αναφερόμενα σε ύψος 10 m από την επιφάνεια της θάλασσας, με τις δορυφορικές μετρήσεις [56]. Η μέτρηση της ταχύτητας του ανέμου γίνεται πλέον αρκετά αξιόπιστα εν τούτοις, η μέση περίοδος του κύματος υπολογίζεται ακόμη χονδρικά [8].

Δεδομένου ότι το βασικότερο μειονέκτημα του δορυφορικού αλτίμετρου για επιχειρησιακές εφαρμογές αποτελεί η ασυνέχεια στη δειγματοληψία από μια συγκεκριμένη περιοχή, η αγγλική εταιρεία *Satellite Observing System (SOS)* σχεδίασε έναν συνδυασμό φθηνών αλτίμετρων ραντάρ και μικρών δορυφόρων που κάνουν χαρτογράφηση. Με αυτόν τον τρόπο δημιουργήθηκε ένα δίκτυο δορυφόρων (*Global Altimeter Network to Eliminate Risk – GANDER*) που συνίσταται σε ένα απλό και οικονομικό σύστημα με 12 – 16 δορυφόρους σαν αυτόν του *ESA Envisat mission* [8].

### 2.3.3.5. Παθητικοί μικροκυματικοί δέκτες (*passive microwave radiometers*)

Οι παθητικοί μικροκυματικοί δέκτες είναι ραδιόμετρα τα οποία ευαισθητοποιούνται στην περιοχή των μικροκυμάτων της ηλεκτρομαγνητικής ακτινοβολίας και καταγράφουν την εκπεμπόμενη μικροκυματική ακτινοβολία από τα ίδια σώματα ή/και την ανακλώμενη ηλιακή μικροκυματική ακτινοβολία από τα σώματα. Επειδή όμως η ποσότητα της ανακλώμενης ή/και εκπεμπόμενης ακτινοβολίας στην περιοχή των μικροκυμάτων είναι πολύ μικρή, απαιτείται μεγάλη χρονική διάρκεια καταγραφής της ακτινοβολούμενης ενέργειας ανά μοναδιαία επιφάνεια για την καταγραφή ικανού σήματος. Αυτό επιτυγχάνεται με μικρή ταχύτητα σάρωσης και χαμηλή χωρική διακριτική ικανότητα (από 10 έως μερικές εκατοντάδες χιλιόμετρα). Οι εν λόγω δέκτες λειτουργούν στις συχνότητες 4.995, 10.69, 15.375 και 19.35 GHz ενώ ραδιομετρικές παρατηρήσεις της ατμόσφαιρας πραγματοποιούνται κοντά στη συχνότητα των 22 GHz όπου παρουσιάζεται ισχυρή απορρόφηση από υδρατμούς [3].

Με τους παθητικούς μικροκυματικούς δέκτες εκτιμάται η ταχύτητα του ανέμου με βάση το γεγονός ότι η θερμική ακτινοβολία, που εκπέμπεται από την επιφάνεια της θάλασσας σε μήκη κύματος της τάξης του χιλιοστού, καθορίζεται από την τραχύτητα της επιφάνειας της θάλασσας, που προκαλείται από τον άνεμο επιφανείας, και εν μέρει από την κάλυψη από αφρισμένη επιφάνεια. Στην περίπτωση αυτή, δεν προσδιορίζεται άμεσα η αντίστοιχη διεύθυνση ανέμου και γίνεται προσπάθεια για την ανάπτυξη μεθόδων εκτίμησής της (*polarization methods*). Αν δεν υπάρχουν ατμοσφαιρικές κατακρημνίσεις, η ακρίβεια στην εκτίμηση της ταχύτητας του ανέμου είναι 2 m/sec ή καλύτερη ενώ για ταχύτητες μεγαλύτερες των 20 m/sec οι μετρήσεις δεν θεωρούνται αξιόπιστες και επομένως, δεν μπορούν να χρησιμοποιηθούν για έντονες καταιγίδες ή τροπικούς κυκλώνες. Θα πρέπει να αναφερθεί ότι οι εν λόγω μετρήσεις είναι έγκυρες για αποστάσεις μεγαλύτερες των 80 km από τις ακτές δεδομένου ότι ο θόρυβος των μικροκυματικών πλευρικών λοβών καθιστά τις ταχύτητες άχρηστες. Αυτό το φαινόμενο μερικές φορές εντοπίζεται σε εικόνες ταχύτητας ανέμου με τη μορφή «δακτύλιου» των υψηλότερων τιμών περιβάλλοντας ακτογραμμές και νησιά και επομένως πρόκειται για άκυρες εκτιμήσεις της ταχύτητας του ανέμου. Το *Special Sensor Microwave Imager (SSM/I)* αποτελεί παράδειγμα τέτοιας εφαρμογής στα πλαίσια του προγράμματος *DMSP (Defense Meteorological Satellite Program)* [65].

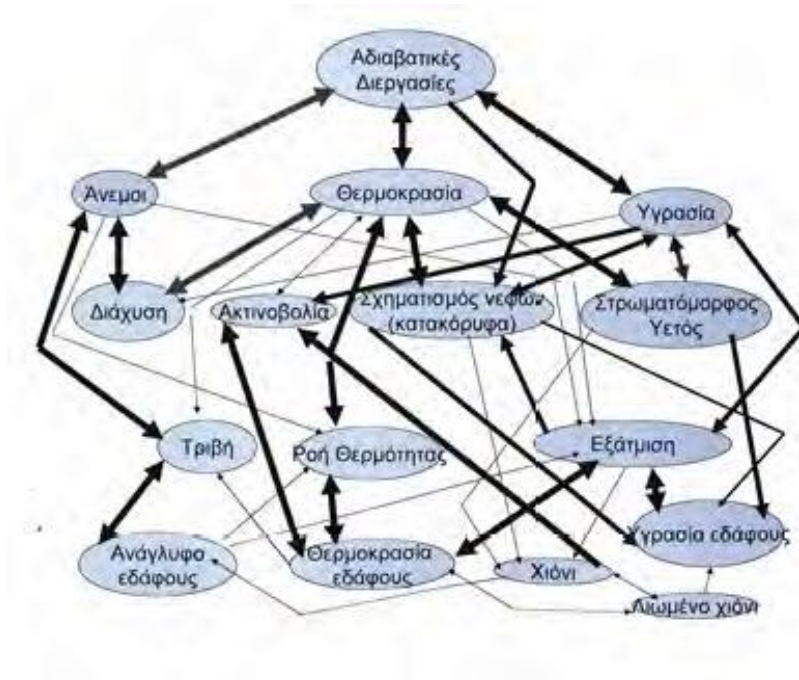


**Εικόνα 2. 10.** Η ταχύτητα του ανέμου μετρημένη από το SSM/I την 01/08/2012 [65].

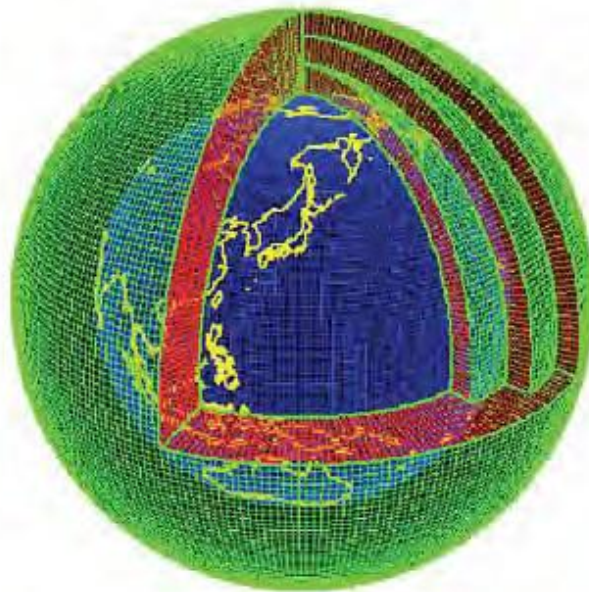
#### 2.4. ΔΕΔΟΜΕΝΑ ΑΡΙΘΜΗΤΙΚΗΣ ΠΡΟΣΟΜΟΙΩΣΗΣ

Η ικανότητα της πρόγνωσης (*forecasting*) ή *hindcast* των μετεωρολογικών συνθηκών σχετικά με την αιολική ενέργεια με χρήση των αριθμητικών μοντέλων καιρού (*numerical weather models*) καθίσταται αναπόσπαστη στα πρότυπα σχεδιασμού και λειτουργίας ενός θαλάσσιου αιολικού πάρκου.

Με τον όρο αριθμητικό μοντέλο καιρού νοείται ένα «κλειστό» σύστημα εξισώσεων που μοντελοποιούν την εξέλιξη των ατμοσφαιρικών – φυσικών φαινομένων. Οι εξισώσεις αυτές είναι μια εφαρμογή των εξισώσεων της υδροδυναμικής στην ατμόσφαιρα με τις διεργασίες που συμβαίνουν σε αυτή (**Εικόνα 2. 11**). Η ατμόσφαιρα προσομοιώνεται με διακριτά σημεία (*grid points*) σε συγκεκριμένα επίπεδα (*levels*) που αρχίζουν από το έδαφος και τελειώνουν στα ανώτερα σημεία της ατμόσφαιρας. Με τον τρόπο αυτό δημιουργείται ένα τρισδιάστατο πλέγμα σημείων πάνω στο οποίο πραγματοποιούνται όλοι οι απαιτούμενοι υπολογισμοί (**Εικόνα 2. 12**). Ορισμένες από τις εξισώσεις έχουν ως μεταβλητή το χρόνο και μάλιστα η μεταβολή των παραμέτρων δίνεται συναρτήσει της μεταβολής  $\Delta t$  του χρόνου. Έτσι, κατά την επίλυση, γίνεται η παραδοχή ότι «τα πάντα συμβαίνουν» σε διακριτές χρονικές στιγμές, που απέχουν μεταξύ τους κατά  $\Delta t$  [7].



Εικόνα 2. 11. Φυσικές διεργασίες των αριθμητικών μοντέλων καιρού [64].



Εικόνα 2. 12. Το τρισδιάστατο πλέγμα σημείων ενός παγκόσμιου μοντέλου [64].

Τα πιο συχνά χρησιμοποιούμενα μοντέλα, τα οποία παράγουν μέσω προσομοίωσης τα ανεμολογικά πεδία (ταχύτητα και διεύθυνση του ανέμου σε διάφορα ύψη) σε επιθυμητές περιοχές, είναι τα ακόλουθα [52]:

- **Μοντέλα μεσαίας κλίμακας (*mesoscale models*):** Τα μοντέλα μεσαίας κλίμακας περιγράφουν μετεωρολογικά φαινόμενα χωρικής ανάλυσης 20 – 2000 km και χρονικής ανάλυσης από ώρες μέχρι μέρες. Αυτά τα μοντέλα έχουν ως εισαγόμενες ποσότητες τα



δεδομένα ανασύνθεσης, το υψόμετρο και δεδομένα τραχύτητας. Τα πιο ευρέως διαδεδομένα μοντέλα αυτής της κατηγορίας είναι τα MM5, KAMM και MC2.

- **Υπολογιστική Ρευστοδυναμική (Computational Fluid Dynamics – CFD):** Τα μοντέλα CFD χρησιμοποιούνται για να μοντελοποιήσουν α) την ροή του ανέμου σε πολυσύνθετο ανάγλυφο και β) τις θερμικές επιδράσεις. Βασίζονται στην επίλυση των εξισώσεων *Reynolds-Averaged Navier-Stokes* σε συνδυασμό με μοντέλα τύρβης (*turbulence models*), παρέχοντας έτσι μια καλύτερη αντίληψη των τυρβωδών ροών όταν γίνεται χρήση επαρκούς χωρικής ανάλυσης. Τα δεδομένα εισόδου είναι το ψηφιακό μοντέλο αναγλύφου, χάρτης τραχύτητας και ανεμολογικά δεδομένα σε διάφορες τοποθεσίες για τουλάχιστον ένα χρόνο ή περισσότερο.
- **Μοντέλα μικρής κλίμακας (microscale models):** Ένα μοντέλο μικρής κλίμακας (ίσως το πιο γνωστό) που αναπτύχθηκε, στα πλαίσια της έρευνας του *European Wind Atlas*, από τους *Troen* και *Petersen* στα τέλη της δεκαετίας του '80 είναι το *WAsP (Wind Atlas Analysis and Application Program)*. Το *WAsP* έχει σκοπό την εξομοίωση των στατιστικών των ανέμων κοντά στην επιφάνεια σε μια περιοχή, με βάση τις παρατηρήσεις του καιρού, αλλά και τις ορογραφικές και τοπογραφικές πληροφορίες. Είναι ένα γρήγορο και φιλικό προς το χρήστη εργαλείο και από την πρώτη στιγμή που δημιουργήθηκε, επικράτησε για την εκτίμηση των τοπικών ανέμων, την εξακρίβωση των επιφανειακών ανέμων, σε περιοχές όπου επικρατούν ίδιες συνθήκες ανέμου.

Σημειώνεται ότι υπάρχει έλλειψη παρατηρήσεων για την εισαγωγή αρχικών συνθηκών προκειμένου να εφαρμοστούν αυτά τα μοντέλα πάνω από την επιφάνεια του ωκεανού. Αυτό συμβαίνει διότι τα μετεωρολογικά φαινόμενα που συναντώνται στην παράκτια ζώνη είναι αρκετά πολύπλοκα, εμφανίζουν σημαντικές ιδιαιτερότητες και έτσι καθίσταται δύσκολη η πρόβλεψη του ανέμου και των διαταραχών για τις υπεράκτιες ανεμογεννήτριες [23]. Παρόλα αυτά, έχουν αναπτυχθεί συγκεκριμένες μεθοδολογίες που βασίζονται σε μακροχρόνιες μετρήσεις σε κοντινούς επίγειους σταθμούς σε σύγκριση με βραχυχρόνιες υπεράκτιες καταγραφές. Πιο συγκεκριμένα, έχουν χρησιμοποιηθεί τρεις διαφορετικές μέθοδοι για την εκτίμηση της κλιματολογίας του ανέμου υπεράκτια, με εφαρμογή σε υπεράκτιες θέσεις στη Δανία, δίνοντας υποσχόμενα αποτελέσματα [13]. Συνοπτικά πρόκειται για τις εξής μεθόδους [57]:

- **Η μέθοδος MCP (Measure-Correlate-Predict).** Υποθέτει ότι υπάρχει μια γραμμική σχέση μεταξύ δύο σημείων, όπου το ένα σημείο έχει τις μακροχρόνιες μετρήσεις και λειτουργεί ως εκτιμητής (*predictor*) ενώ η ταχύτητα του ανέμου σε θέσεις με σύντομες χρονικά μετρήσεις θεωρείται ως εκτιμώμενη ποσότητα (*predictand*). Από την εξίσωση παλινδρόμησης που ορίζεται με βάση την κοινή περίοδο μετρήσεων, χρησιμοποιούνται οι παράμετροι της παλινδρόμησης για να προκύψει μία καταγραφή εκτεταμένων δεδομένων για το σημείο ενδιαφέροντος.
- **Risø's WAsP.** Υπολογίζει την κλιματολογία του ανέμου σε ένα σημείο από την μακροχρόνια κλιματολογία του ανέμου αντιπροσωπευτικών σταθμών. Το *WAsP* βασίζεται σε νόμους της φυσικής και χρησιμοποιεί μια τυποποιημένη θερμική ροή υπεράκτια και παράκτια για να υπολογίσει μια σταθερή διόρθωση και την αλλαγή στην τραχύτητα προκειμένου να διορθώσει τη ροή της ορμής [46].
- **Η μέθοδος διόρθωσης Weibull** για την επέκταση χρονοσειρών ανεμολογικών δεδομένων. Στηρίζεται στην ιδέα της τροποποίησης των παραμέτρων της συνάρτησης *Weibull* των βραχυχρόνιων χρονοσειρών δεδομένων για τον χαρακτηρισμό μιας μεγαλύτερης χρονικά περιόδου δειγματοληψίας και συγκρίνει τις κατανομές της ταχύτητας του ανέμου σε παράκτιες και υπεράκτιες θέσεις λαμβάνοντας υπόψη τις παράκτιες μεγάλες χρονοσειρές ως αντιπροσωπευτικές για την περιοχή μελέτης [57].

Επειδή η μεγάλη διακριτική ικανότητα στα μοντέλα παγκόσμιας κλίμακας, καθιστά τον αριθμό των σημείων του πλέγματος τεράστιο και επομένως, η διαδικασία επίλυσης τους γίνεται χρονοβόρα, τα τελευταία χρόνια έχουν αναπτυχθεί και χρησιμοποιούνται ολοένα και περισσότερο τα Περιοχικά Μοντέλα, γνωστά και ως *LAM (Limited Area Models)*. Η λειτουργία των *LAM* έγκειται στην επιλογή μιας σχετικά μικρής περιοχής ενδιαφέροντος στην οποία χρησιμοποιείται πυκνότερο πλέγμα σημείων με μικρή διακριτική ικανότητα. Το σύστημα των εξισώσεων λύνεται στα σημεία του νέου πλέγματος, με την παραδοχή ότι στα άκρα της περιοχής του τοπικού μοντέλου, η λύση ταυτίζεται με αυτήν του παγκόσμιου μοντέλου, δίνοντας με τον τρόπο αυτό τις απαιτούμενες συνοριακές συνθήκες (*boundary conditions*) [7].

Καθοριστικό ρόλο στην ποιότητα της εκτίμησης του αιολικού δυναμικού παίζει και η διακριτική ικανότητα των μοντέλων. Όσο μεγαλύτερη είναι η διακριτική τους ικανότητα (που σημαίνει όσο μικρότερες είναι οι αποστάσεις μεταξύ των σημείων του πλέγματος), τόσο καλύτερα και ακριβέστερα μπορεί να προσομοιώσουν διαταραχές και φαινόμενα μικρής κλίμακας, γεγονός που συνεπάγεται αναλυτικότερα και ποιοτικότερα αποτελέσματα. Ένα χαρακτηριστικό παράδειγμα της επίδρασης της διακριτικής ικανότητας αποτελεί η αναπαράσταση της ορογραφίας στο μοντέλο. Η διακριτική ικανότητα των 7.5 km και των 2.5 km μπορεί να δώσει σημαντικές διαφορές στην εκτίμηση του αιολικού δυναμικού, αφού στην πρώτη περίπτωση δε λαμβάνονται υπόψη τα φαινόμενα εμποδισμού ή καναλισμού της ροής του ανέμου με τη σωστή διάστασή τους. Προφανώς, σε μικρότερη κλίμακα επίλυσης των εξισώσεων χρειάζεται καλύτερη ανάλυση και εισαγωγή αναλυτικότερης φυσικής με έμφαση στις μη υδροστατικές διεργασίες που ισχύουν σε μικρότερες κλίμακες ατμοσφαιρικής ροής, καθώς προσομοιώνουν τις ατμοσφαιρικές διαδικασίες μικρότερου μήκους κύματος. Κατά την ολοκλήρωση των μοντέλων πρόγνωσης με την εισαγωγή της αναλυτικότερης τοπογραφίας και γεωμορφολογίας αναγκάζεται η ατμοσφαιρική ροή να ακολουθήσει τη δυναμική των εξισώσεων ολοκλήρωσης με αποτέλεσμα την καλύτερη προσομοίωση του πεδίου του ανέμου [7].

Στην **Ευρώπη**, υπάρχει μία συστηματική προσπάθεια στην εκτίμηση των δεδομένων από διάφορες ανεμολογικές πηγές με τη δημοσίευση ενός Ευρωπαϊκού Ανεμολογικού Άτλαντα (*European Wind Atlas*) το 1989. Περιλαμβάνει μία γενική άποψη των ανεμολογικών πηγών, παρέχει πληροφορίες για τον προσδιορισμό του αιολικού δυναμικού και θέσεις εγκατάστασης ανεμογεννητριών και εξηγεί το μετεωρολογικό υπόβαθρο και την ανάλυση του συγκεκριμένου άτλαντα περιλαμβάνοντας τη φυσική και στατιστική βάση των μοντέλων [47]. Επιπρόσθετα, τα ευρωπαϊκά μετεωρολογικά ινστιτούτα λαμβάνουν τις προκαταρκτικές προβλέψεις για όλη τη γη από το Ευρωπαϊκό Κέντρο Μεσοπρόθεσμων Μετεωρολογικών Προγνώσεων (*European Centre for Medium-Range Weather Forecasts – ECMWF*), το οποίο έχει ολοκληρώσει τη δημιουργία μιας μεγάλης βάσης μετεωρολογικών δεδομένων (*ERA-40*), χρησιμοποιώντας σύγχρονες τεχνικές αφομοίωσης διαθέσιμων παρατηρήσεων στα αποτελέσματα του πλανητικού μοντέλου *IFS CY23r4*. Σήμερα, η βάση *ERA-40* και κυρίως η βάση *ERA-Interim* θεωρούνται οι καλύτερες διαθέσιμες πληροφορίες για την αναλυτική περιγραφή της παγκόσμιας ατμοσφαιρικής και κυματικής κατάστασης για μια περίοδο περίπου 35 ετών (1976-2012).

Στην **Ελλάδα**, το μετεωρολογικό σύστημα *SKIIRON* αναπτύχθηκε από στο Καποδιστριακό Πανεπιστήμιο της Αθήνας με σκοπό την εφαρμογή του στην Ελληνική Εθνική Μετεωρολογική Υπηρεσία (ΕΕΜΥ) για την τοπική πρόγνωση καιρού στην περιοχή της Μεσογείου. Είναι ένα αξιόπιστο υπολογιστικό σύστημα το οποίο παράγει προβλέψεις καιρού

και είναι ιδιαίτερα χρήσιμο για τοπικές ατμοσφαιρικές συνθήκες. Το *SKIRON* είναι η εξέλιξη του μοντέλου *Eta* το οποίο έχει αναπτυχθεί από το Πανεπιστήμιο του Βελιγραδίου με σκοπό την πρόγνωση του καιρού σε περιοχές με βουνά απότομης κλίσης.

Ο βασικός σκοπός του ερευνητικού προγράμματος *SKIRON* ήταν η δημιουργία ενός υπολογιστικά ευέλικτου μοντέλου που να παράγει προβλέψεις ιδιαίτερα για τοπικές συνθήκες καιρού. Τα μοντέλα πρόβλεψης της παραγωγής ενός αιολικού πάρκου χρησιμοποιούν τις αριθμητικές προβλέψεις καιρού με αυτή τη μορφή. Τις αρχικές και οριακές συνθήκες τις λαμβάνει από το *ECMWF* οι οποίες ανανεώνονται κάθε τρεις ώρες. Η συχνότητα ενημέρωσης των μετεωρολογικών προβλέψεων είναι μια φορά τη μέρα και η πρόγνωση καλύπτει τις επόμενες 72 ώρες [37].

## ΚΕΦΑΛΑΙΟ 3 – ΜΕΘΟΔΟΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

### 3.1. ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

#### 3.1.1. Γενικά

Η ανάλυση της γραμμικής παλινδρόμησης ανήκει στην οικογένεια των στατιστικών μεθόδων, που χρησιμοποιούνται για τη διενέργεια προβλέψεων και τον προσδιορισμό των αντίστοιχων περιοχών σφάλματος. Αυτή η διαδικασία γίνεται μέσω της εκτίμησης και της μοντελοποίησης της γραμμικής σχέσης μεταξύ του μέσου μιας εξαρτημένης μεταβλητής (*response*) και μιας ή πολλών ανεξάρτητων μεταβλητών (*regressor*).

Στην πραγματικότητα, η ανάλυση παλινδρόμησης είναι μία από τις πιο ευρέως χρησιμοποιούμενες στατιστικές τεχνικές. Μέθοδοι παλινδρόμησης χρησιμοποιούνται εδώ και πολλά χρόνια σε πεδία εφαρμογών συμπεριλαμβανομένων της μηχανικής, των φυσικών, χημικών, οικονομικών, περιβαλλοντικών, βιολογικών και κοινωνικών επιστημών. Η παλαιότερη μορφή της γραμμικής παλινδρόμησης είναι η μέθοδος των ελαχίστων τετραγώνων που αναπτύχθηκε ολοκληρωμένα από τον *Legendre* (1805). Η προτίμηση της συγκεκριμένης μεθόδου οφείλεται στην εννοιολογικά απλή και γρήγορη διαδικασία εκτίμησης των άγνωστων παραμέτρων και στην άρτια ανεπτυγμένα στατιστική θεωρία μέσω κομψών μαθηματικών σχέσεων.

#### 3.1.2. Μέθοδος Συνηθισμένων Ελαχίστων Τετραγώνων

Η απλή γραμμική παλινδρόμηση δηλώνει ότι ο πραγματικός μέσος (*true mean*) της εξαρτημένης μεταβλητής  $Y$  μεταβάλλεται συνεχώς όσο η τιμή της ανεξάρτητης μεταβλητής  $x$  αυξάνεται ή μειώνεται. Κατά συνέπεια, η συναρτησιακή σχέση μεταξύ του πραγματικού μέσου  $E[Y | x]$  των  $Y$  και  $x$ , εκφράζεται ως εξής:

$$E[Y | x] = b_0 + b_1 x \quad (1)$$

όπου  $b_0$  είναι ένας σταθερός όρος (*intercept*), δηλαδή η τιμή όπου  $E[Y | x] = 0$  και  $b_1$  είναι η κλίση της ευθείας (*slope*), δηλαδή ο ρυθμός μεταβολής της  $E[Y]$  ανά μονάδα μεταβολής της  $x$ . Ο όρος «απλή» αναφέρεται στο γεγονός ότι υπάρχει μόνο μια ανεξάρτητη μεταβλητή και ο όρος «γραμμική» σημαίνει ότι η συνάρτηση παλινδρόμησης είναι μια γραμμική συνάρτηση των παραμέτρων  $b_0$  και  $b_1$  και όχι ότι η σχέση μεταξύ των μεταβλητών μπορεί να απεικονιστεί γραφικά ως ευθεία γραμμή [55].

Εφόσον η μεταβλητή  $Y$  είναι τυχαία, για κάθε πραγματοποίηση της  $x$ , έστω  $x_i$ , υπάρχει μια πραγματοποίηση της  $Y$ , έστω  $y_i$ , η οποία θεωρείται ότι βρίσκεται στη γειτονιά της  $E[Y]$ . Στην πραγματικότητα, οι παρατηρήσεις  $y_i$  είναι τυχαίες από πληθυσμούς με τυχαίες μεταβλητές με μέσο κάθε πληθυσμού  $E[y_i]$ . Έτσι, σε αυτό το σημείο εισάγουμε το τυχαίο σφάλμα ή αλλιώς διαταρακτικό όρο (*disturbance term*)  $\varepsilon_i$ , οπότε το βασικό μοντέλο της απλής γραμμικής παλινδρόμησης παίρνει τη μορφή:



$$y_i = b_0 + b_1 x_i + \varepsilon_i \quad (2)$$

που είναι ισοδύναμο με τη σχέση  $y_i = E[Y | x_i] + \varepsilon_i$ , όπου ο δείκτης  $i$  υποδηλώνει την εκάστοτε συγκεκριμένη παρατήρηση  $i = 1, 2, \dots, n$ . Ο τελευταίος όρος δηλώνει ουσιαστικά την έλλειψη ακριβούς σχέσης μεταξύ των  $x$  και  $y$  λόγω της παρουσίας σφαλμάτων.

Για το τυχαίο σφάλμα  $\varepsilon_i$  πρέπει να γίνουν οι ακόλουθες υποθέσεις:

1. Η σχέση (2) υποδηλώνει ότι υπάρχει μια γραμμική σχέση που συνδέει τα  $y$  και  $x$  και πιο συγκεκριμένα κάθε τιμή της  $y_i$  είναι γραμμική συνάρτηση της τιμής  $x_i$  συν το διαταρακτικό όρο  $\varepsilon_i$ .
2. Η τυχαία μεταβλητή  $\varepsilon_i$  ακολουθεί την κανονική κατανομή με μέση τιμή μηδέν και σταθερή διακύμανση  $\sigma^2$ , δηλαδή  $\varepsilon_i \sim N(0, \sigma^2)$ . Η πρώτη υπόθεση πρακτικά σημαίνει ότι τα θετικά σφάλματα αντισταθμίζονται κατά μέσο όρο από τα αρνητικά. Η υπόθεση της σταθερής διακύμανσης ονομάζεται και υπόθεση ομοσκεδαστικότητας (*homoscedasticity*).
3. Οι διαταρακτικοί όροι είναι ασυσχέτιστοι μεταξύ τους και επομένως, η συνδιακύμανση του διαταρακτικού όρου της παρατήρησης  $i$  με το διαταρακτικό όρο οποιασδήποτε άλλης παρατήρησης  $j$  είναι μηδέν. Δηλαδή  $Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i - E\varepsilon_i)(\varepsilon_j - E\varepsilon_j) = E\varepsilon_i \varepsilon_j = 0$  για κάθε  $i \neq j$ .
4. Η ανεξάρτητη μεταβλητή  $x$  δεν είναι στοχαστική, οι τιμές της παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους σε μια υποθετική διαδικασία επαναλαμβανόμενης δειγματοληψίας. Δηλαδή  $Cov(x_i, \varepsilon_i) = 0$ .

Θα πρέπει να αναφερθεί ότι, γενικά, στις εφαρμογές της απλής γραμμικής παλινδρόμησης οι παρατηρήσεις  $x_i$  της ανεξάρτητης μεταβλητής θεωρούνται ότι έχουν μετρηθεί χωρίς σφάλμα και δεν είναι τυχαίες. Επομένως, η εξαρτημένη μεταβλητή  $y$  είναι συνάρτηση του τυχαίου όρου  $\varepsilon$  με αποτέλεσμα η  $y$  να είναι επίσης μια τυχαία μεταβλητή, η οποία μπορεί να προβλεφθεί μόνο εάν απαλλαγεί από αυτόν το διαταρακτικό όρο. Εφόσον εκτιμηθούν οι τιμές των παραμέτρων  $b_0$  και  $b_1$ , η γραμμή παλινδρόμησης θα είναι της μορφής:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i \quad (3)$$

ή πιο γενικά

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x \quad (4)$$

όπου  $\hat{b}_0$  και  $\hat{b}_1$  είναι οι εκτιμητές για τους συντελεστές  $b_0$  και  $b_1$  και  $\hat{y}$  είναι η εκτίμηση της τιμής της  $y$  που υπολογίζεται από τη γραμμή παλινδρόμησης.

Για τον υπολογισμό των τιμών των παραμέτρων  $\hat{b}_0$  και  $\hat{b}_1$  της εξίσωσης (4) εφαρμόζεται η **Μέθοδος Συνηθισμένων Ελαχίστων Τετραγώνων (Ordinary Least Squares or OLS)**

*method*) έτσι ώστε το άθροισμα των τετραγώνων των όρων σφάλματος  $\varepsilon_i$  που ορίζεται από τη συνάρτηση  $SSE = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$  να ελαχιστοποιείται. Αυτό επιτυγχάνεται αν υπολογιστούν οι μερικές παράγωγοι της  $SSE$  ως προς τους συντελεστές  $b_0$  και  $b_1$  και εξισωθούν με μηδέν. Έτσι καταλήγουμε σε ένα γραμμικό σύστημα δύο εξισώσεων, που ονομάζονται κανονικές εξισώσεις, με αγνώστους τα  $\hat{b}_0$  και  $\hat{b}_1$ . Η λύση αυτού του συστήματος είναι η ακόλουθη:

$$\hat{b}_1 = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n-1} (x_i - \bar{x})^2} \quad (5)$$

και

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \quad (6)$$

όπου τα  $\bar{x}$  και  $\bar{y}$  είναι οι δειγματικές μέσες τιμές των μεταβλητών  $x$  και  $y$  και υπολογίζονται από τις σχέσεις  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  και  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  αντίστοιχα. Θέτοντας όπου

$$S_{XX} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7)$$

$$S_{YY} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (8)$$

και

$$S_{XY} = S_{YX} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (9)$$

ο συντελεστής  $\hat{b}_1$  μπορεί να γραφτεί ως εξής:

$$\hat{b}_1 = \frac{S_{XY}}{S_{XX}}.$$

Με άλλα λόγια, από όλες τις δυνατές γραμμικές συναρτήσεις των παρατηρήσεων, εκείνη που ορίζεται από τους εκτιμητές των ελαχίστων τετραγώνων  $\hat{b}_0$  και  $\hat{b}_1$  έχει τη μικρότερη μεταβλητότητα.

Για κάθε  $x_i$  υπάρχει ένα μέρος της παρατήρησης  $y_i$  το οποίο μένει ανερμηνευτο από το εκτιμηθέν μοντέλο και αποτελεί εκτίμηση του διαταρακτικού όρου  $\varepsilon_i$ . Αυτές οι διαφορές

των εκτιμήσεων  $\hat{y}_i$  από τις παρατηρήσεις  $y_i$  συμβολίζονται με  $\hat{\varepsilon}_i$  και ονομάζονται κατάλοιπα (*residuals*) (βλ. **Σχήμα 3. 1**). Η ποσότητα αυτή ισούται με:

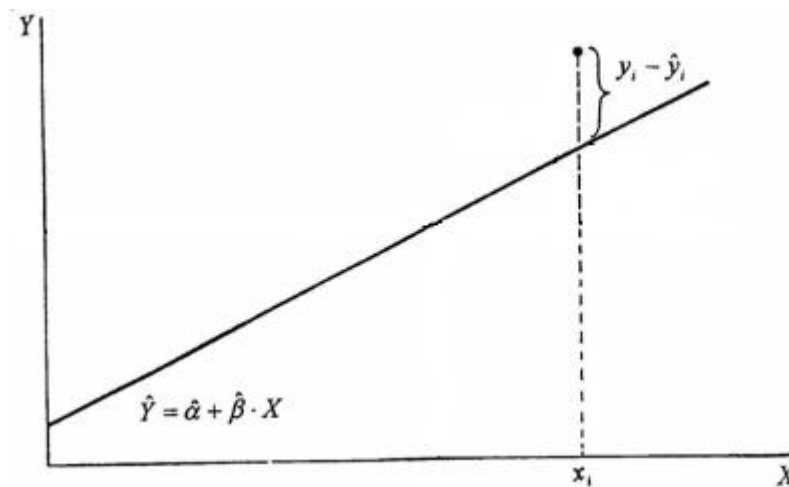
$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{b}_0 - \hat{b}_1 x_i, \quad i = 1, 2, \dots, n. \quad (10)$$

Επομένως, η συμπεριφορά των καταλοίπων  $\hat{\varepsilon}_i, i = 1, 2, \dots, n$  δίνει χρήσιμες πληροφορίες για τη συμπεριφορά των διαταρακτικών όρων.

Τα κατάλοιπα  $\hat{\varepsilon}_i, i = 1, 2, \dots, n$  της εκτίμησης ενός μοντέλου έχουν τις ακόλουθες ιδιότητες:

1. Το άθροισμα τους ισούται με μηδέν, δηλαδή  $\sum_{i=1}^n \hat{\varepsilon}_i = 0$ .
2. Το άθροισμα των γινομένων  $x_i \hat{\varepsilon}_i$  ισούται με μηδέν, δηλαδή  $\sum_{i=1}^n x_i \hat{\varepsilon}_i = 0$ .
3. Το άθροισμα των γινομένων  $y_i \hat{\varepsilon}_i$  ισούται με μηδέν, δηλαδή  $\sum_{i=1}^n y_i \hat{\varepsilon}_i = 0$ .

Η πρώτη και δεύτερη ιδιότητα χαρακτηρίζουν την ευθεία ελαχίστων τετραγώνων δηλαδή, αρκεί μια ευθεία να έχει τις δύο αυτές ιδιότητες για να είναι ευθεία ελαχίστων τετραγώνων [2].



**Σχήμα 3. 1.** Γραφική απεικόνιση της κατακόρυφης απόστασης ενός καταλοίπου από τη γραμμή παλινδρόμησης στη μέθοδο ελαχίστων τετραγώνων.

Κάθε παρατήρηση  $y_i$  μπορεί να γραφεί ως εξής:

$$y_i = \hat{y}_i + \hat{\varepsilon}_i, \quad i = 1, 2, \dots, n, \quad (11)$$

όπου  $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$  η εκτίμηση που ορίζεται από την ευθεία ελαχίστων τετραγώνων και  $\hat{\varepsilon}_i = y_i - \hat{y}_i$  το κατάλοιπο της εκτίμησης. Η σχέση (11) μπορεί ισοδύναμα να γραφεί ως εξής:

$$y_i = \hat{y}_i + (y_i - \hat{y}_i) \Leftrightarrow y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i), \quad i = 1, 2, \dots, n.$$

Υψώνοντας στο τετράγωνο και αθροίζοντας για όλες τις τιμές του  $i$  την παραπάνω σχέση προκύπτει ότι:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (12)$$

Θα αναφερόμαστε στον όρο  $\sum_{i=1}^n (y_i - \bar{y})^2$  ως ολικό άθροισμα τετραγώνων ή  $S_{YY}$  ή  $SST$  (*total sum of squares*), στον όρο  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$  ως άθροισμα τετραγώνων παλινδρόμησης ή  $SSR$  (*regression sum of squares*) και στον όρο  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$  ως άθροισμα τετραγώνων των καταλοίπων ή  $RSS$  (*residual sum of squares*). Άρα η σχέση (12) γράφεται ως έξης:

$$SST = SSR + RSS. \quad (13)$$

Ο όρος  $SST$  είναι ένα μέτρο της συνολικής μεταβλητότητας των παρατηρήσεων  $y_i$  γύρω από τη μέση τιμή τους, ενώ ο όρος  $RSS$  είναι το αντίστοιχο μέτρο των παρατηρήσεων  $y_i$  γύρω από την ευθεία των ελαχίστων τετραγώνων. Η διαφορά  $SST - RSS$  είναι η μείωση του σφάλματος εκτίμησης που επιτυγχάνεται με το μοντέλο παλινδρόμησης. Η αναλογία της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από την παλινδρόμηση ονομάζεται συντελεστής προσδιορισμού (*coefficient of determination*) και παριστάνεται με την παράμετρο  $R^2$ . Επίσης μπορεί να ερμηνευθεί ως η συσχέτιση μεταξύ της προβλεπόμενης μεταβλητής  $Y$  και της εκτιμώμενης της (μέσω της ανεξάρτητης μεταβλητής). Το  $R^2$  υπολογίζεται από τη σχέση:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{RSS}{SST}. \quad (14)$$

Ο συντελεστής προσδιορισμού δεν μπορεί να πάρει αρνητικές τιμές ή μεγαλύτερες από τη μονάδα, δηλαδή  $0 \leq R^2 \leq 1$ . Όσο πιο κοντά στη μονάδα είναι ο εν λόγω συντελεστής, τόσο μεγαλώνει και το ποσοστό της μεταβλητότητας στην  $y_i$  που εξηγείται από το μοντέλο παλινδρόμησης.

Θα πρέπει να σημειωθεί ότι υπάρχουν αρκετές παρανοήσεις σχετικά με το συντελεστή προσδιορισμού. Γενικά, το  $R^2$  δεν εκτιμάει το μέγεθος της κλίσης της γραμμής παλινδρόμησης. Μεγάλη τιμή του συντελεστή προσδιορισμού δε συνεπάγεται και απότομη κλίση. Επιπρόσθετα, το  $R^2$  δεν εκτιμάει την καταλληλότητα του γραμμικού μοντέλου, καθώς συχνά μπορεί να έχει μεγάλη τιμή παρόλο που τα  $y$  και  $x$  είναι μη γραμμικά συσχετισμένα. Υπενθυμίζουμε ότι παρόλο που το  $R^2$  μπορεί να είναι μεγάλο, δε σημαίνει απαραίτητα ότι το μοντέλο παλινδρόμησης θα αποτελεί μια ακριβή αναπαράσταση [45]. Ας σημειωθεί, επίσης, ότι χαμηλή ή ακόμα μηδενική τιμή του  $R^2$  δε σημαίνει αναγκαστικά έλλειψη εξαρτήσεως ανάμεσα στις μεταβλητές  $y$  και  $x$  [10].

Το μεγαλύτερο μειονέκτημα του συντελεστή προσδιορισμού έγκειται στο γεγονός ότι όσες περισσότερες μεταβλητές προστεθούν στο μοντέλο (για την περίπτωση της πολλαπλής

παλινδρόμησης), το  $R^2$  μπορεί μόνο να αυξηθεί, διότι το  $RSS$  με την προσθήκη μιας μεταβλητής μπορεί μόνο να μειωθεί (ακόμα και αν η καινούρια μεταβλητή δεν προσδίδει κάτι στην ερμηνεία του μοντέλου παλινδρόμησης και στην ουσία δεν ερμηνεύει κάποιο ποσοστό διακύμανσης στην πραγματικότητα). Το πρόβλημα αυτό αντιμετωπίζεται με τον προσαρμοσμένο συντελεστή προσδιορισμού (*adjusted coefficient of determination*) που συμβολίζεται με  $R_\alpha^2$  και, λαμβάνοντας υπόψη τους αντίστοιχους βαθμούς ελευθερίας, υπολογίζεται από τη σχέση:

$$R_\alpha^2 = 1 - (1 - R^2) \left( \frac{n-1}{n-2} \right). \quad (15)$$

Το πλεονέκτημα με το  $R_\alpha^2$  είναι ότι μπορεί να μειωθεί με την προσθήκη μιας νέας μεταβλητής και αυξάνεται αν όντως μια μεταβλητή προσφέρει σημαντική μείωση στην διακύμανση της εξαρτημένης μεταβλητής  $y$ .

Ένα μέτρο του βαθμού συσχέτισης δύο τυχαίων μεταβλητών είναι ο συντελεστής συσχέτισης (*correlation coefficient*)  $r_{XY}$ , που ορίζεται ως:

$$r_{XY} = \frac{S_{XY}}{S_{XX} S_{YY}}. \quad (16)$$

Ο συντελεστής  $r_{XY}$  είναι ένας «καθαρός» αριθμός, επιτρέποντας έτσι τις συγκρίσεις, και μπορεί να πάρει τιμές στο κλειστό διάστημα  $[-1, 1]$ . Επίσης, μετράει την ένταση γραμμικής συμμεταβολής. Έτσι, δύο μεταβλητές  $X$ ,  $Y$  μπορεί να έχουν συντελεστή συσχέτισης ίσο με μηδέν και οι μεταβλητές να μην είναι ανεξάρτητες, αλλά να συνδέονται με μία σχέση η οποία δεν είναι γραμμική. Επίσης, ο συντελεστής συσχέτισης έχει το ίδιο πρόσημο με το συντελεστή παλινδρόμησης  $\hat{b}_1$ . Όταν η συμμεταβολή των  $X$  και  $Y$  είναι προς την ίδια κατεύθυνση τότε και η κλίση  $\hat{b}_1$  της γραμμής παλινδρόμησης είναι θετική, διαφορετικά είναι αρνητική [2].

Είναι φανερό πως ο συντελεστής συσχέτισης ισούται με την τετραγωνική ρίζα του συντελεστή προσδιορισμού. Αν και υπάρχει τόσο στενή μαθηματική σχέση μεταξύ  $r_{XY}$  και  $R^2$ , εν τούτοις υπάρχει σημαντική διαφορά ως προς την ερμηνεία τους. Ο συντελεστής συσχέτισης είναι μέτρο μόνο της γραμμικής συσχέτισης ή εξαρτήσεως δύο μεταβλητών ενώ ο συντελεστής προσδιορισμού αναφέρεται στην αναλογία της μεταβλητότητας της  $Y$  που ερμηνεύει η μεταβλητή  $X$  [10].

Τέλος, η μεταβλητότητα των σφαλμάτων  $\sigma_e^2 = \text{Var}(e)$  δεν μπορεί να υπολογιστεί άμεσα καθώς τα σφάλματα δεν είναι παρατηρήσιμα. Πρακτικά, ως στενός συγγενικός δείκτης, χρησιμοποιείται το μέσο τετραγωνικό σφάλμα (*mean square error*)  $s_\varepsilon^2$  που ορίζεται από τη σχέση:

$$s_\varepsilon^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-2} = \frac{RSS}{n-2}. \quad (17)$$

Η θετική τετραγωνική ρίζα του  $s_{\varepsilon}^2$ ,  $s_{\varepsilon} = \sqrt{s_{\varepsilon}^2}$ , ονομάζεται τυπική απόκλιση των καταλοίπων (*residual standard deviation*) [58].

### 3.1.3. Ρύθμιση/διόρθωση μέσω γραμμικής παλινδρόμησης

Τα περισσότερα προβλήματα παλινδρόμησης που αφορούν την πρόβλεψη ή εκτίμηση προϋποθέτουν τον προσδιορισμό μιας τιμής  $y$  που αντιστοιχεί σε δοσμένη τιμή της  $x$ , όπως είναι η  $x_0$ . Μια ειδική χρήση της απλής γραμμικής παλινδρόμησης αποτελεί η αντίστροφη παλινδρόμηση (*inverse regression*) και χρησιμοποιείται αποτελεσματικά κυρίως σε εργασίες βαθμονόμησης, παρόλο που οι εφαρμογές της δεν περιορίζονται μόνο εκεί. Σε αυτή την περίπτωση, θεωρούμε ότι έχει παρατηρηθεί μια τιμή της  $y$ , έστω η  $y_0$ , και πρέπει να υπολογιστεί η τιμή της  $x$  που αντιστοιχεί σε αυτήν την τιμή.

Ας υποθέσουμε ότι έχουμε δύο μετρητικά όργανα, το ένα είναι αρκετά ακριβές αλλά εξίσου δαπανηρό στη χρήση του και το άλλο θεωρείται πιο οικονομικό, αλλά λιγότερο ακριβές. Επίσης, υποθέτουμε ότι οι μετρήσεις που έχουν ληφθεί από τα δύο όργανα έχουν υψηλή συσχέτιση επομένως, η μέτρηση που θα μπορούσε να είχε γίνει με την ακριβή μετρητική συσκευή μπορεί να προβλεφθεί ικανοποιητικά από τη μέτρηση που προκύπτει από την οικονομικά προσιτή συσκευή [55]. Ορίζουμε με  $x$  τη μέτρηση του ακριβούς οργάνου (όργανο αναφοράς) και με  $y$  τη μέτρηση από το λιγότερο αξιόπιστο όργανο. Το πρόβλημα βαθμονόμησης είναι να βρεθεί ένας τρόπος διόρθωσης του δεύτερου οργάνου έτσι ώστε να προκύψουν μετρήσεις πολύ κοντά με αυτές της συσκευής αναφοράς.

Πολύ συχνά εφαρμόζεται η γραμμική παλινδρόμηση κατά την οποία η εξαρτημένη μεταβλητή  $y$ , η οποία είναι μετρημένη με κάποιο σφάλμα που ακολουθεί την κανονική κατανομή, παλινδρομείται πάνω στην ανεξάρτητη μεταβλητή  $x$  και το αποτέλεσμα που προκύπτει χρησιμοποιείται για την εκτίμηση της τιμής μιας μέτρησης  $x_0$  ενός άγνωστου δείγματος από την αντίστοιχη μετρημένη τιμή της  $y_0$ . Αυτή η διαδικασία ονομάζεται κλασική προσέγγιση ρύθμισης/διόρθωσης (*classical calibration approach*). Εφαρμόζοντας την παλινδρόμηση της  $y$  επί της  $x$ , προκύπτουν οι εκτιμητές  $\hat{b}_0$  και  $\hat{b}_1$  των παραμέτρων  $b_0$  και  $b_1$  και το μοντέλο που προκύπτει με αναδιάταξη της σχέσης (4) θα είναι της μορφής:

$$\hat{x}_0 = \frac{y_0 - \hat{b}_0}{\hat{b}_1}, \quad (18)$$

με  $\hat{b}_1 \neq 0$  [58].

Μια εναλλακτική μέθοδος για την εκτίμηση της  $x$  είναι να εφαρμοστεί η παλινδρόμηση της ανεξάρτητης μεταβλητής  $x$  επί της εξαρτημένης μεταβλητής  $y$  παίρνοντας με αυτό τον τρόπο την αντίστροφη γραμμή παλινδρόμησης. Αυτό το μοντέλο περιγράφεται ως εξής:

$$\hat{x}_0 = \alpha_0 + \alpha_1 y_0. \quad (19)$$

Η παραπάνω προσέγγιση προτάθηκε από τον *Krutchkoff* (1967) και ονομάζεται αντίστροφη παλινδρόμηση. Όμως, η χρήση της τελευταίας εξίσωσης προϋποθέτει ότι η εξαρτημένη μεταβλητή  $x$  είναι τυχαία. Στις περισσότερες περιπτώσεις, η μεταβλητή  $x$  είναι ελεγχόμενη, γεγονός που καθιστά τη χρήση της ακατάλληλη, διότι παραβιάζονται υποθέσεις που υποκρύπτονται στη γραμμική παλινδρόμηση. Εν τούτοις, εάν η  $x$  είναι μια τυχαία παρατηρήσιμη μεταβλητή, τότε η προσέγγιση αυτή μπορεί να είναι αρκετά λογική [45].

Τέλος, θα πρέπει να σημειωθεί ότι τόσο η αντίστροφη παλινδρόμηση όσο και η κλασσική προσέγγιση παλινδρόμησης δίνουν αποτελέσματα πολύ κοντά μεταξύ τους, εάν τα δεδομένα των μετρήσεων είναι υψηλά συσχετισμένα [55].

### 3.1.4. Σημεία Επιρροής (Influential Points)

Κατά τον υπολογισμό διάφορων στατιστικών μεγεθών όπως είναι η μέση τιμή και η τυπική απόκλιση, κάθε μία από τις  $n$  παρατηρήσεις έχουν το ίδιο βάρος στον προσδιορισμό των τιμών αυτών των στατιστικών μεγεθών, εκτός από τις περιπτώσεις που υπάρχουν κάποιες «ασυνήθιστες» παρατηρήσεις στα δεδομένα και οι οποίες μπορεί να αλλοιώσουν τις εκτιμήσεις αυτών των μεγεθών.

Στην περίπτωση της ανάλυσης της παλινδρόμησης, ανάλογη κατάσταση μπορεί να οδηγήσει σε λάθος εκτίμηση της κλίσης της ευθείας των ελαχίστων τετραγώνων και γι' αυτό θα πρέπει να δοθεί ιδιαίτερη προσοχή και να εξεταστεί αναλυτικά κατά πόσο οι παρατηρήσεις που ασκούν μεγάλη «επιρροή» στο μοντέλο πρέπει να παραμείνουν ή να αποκλειστούν από τη διαδικασία εκτίμησης των παραμέτρων του. Τέτοιες παρατηρήσεις είναι αυτές που έχουν τιμές της  $x_i$  αρκετά πιο απομακρυσμένες από τα υπόλοιπα  $x_j$ ,  $i \neq j$  ή διαφορετικά, έχουν  $x_i$  αρκετά μακριά από τη μέση τιμή  $\bar{x}$ .

Ένας άμεσος τρόπος για τον εντοπισμό πιθανών σημείων επιρροής είναι το διάγραμμα διασποράς (*scatter plot*). Επειδή όμως υπάρχει το ενδεχόμενο μια παρατήρηση να απέχει από τη γραμμή παλινδρόμησης αλλά παράλληλα να έχει υψηλή επιρροή ή το αντίστροφο, δηλαδή να βρίσκεται πολύ κοντά στην ευθεία παλινδρόμησης και να ασκεί παρόλα αυτά επιρροή, υπάρχουν επιπρόσθετα διάφοροι στατιστικοί έλεγχοι για τον εντοπισμό αυτών των σημείων [55].

Μία άλλη κατηγορία στατιστικού ελέγχου του μεγέθους επιρροής στην εκτίμηση των παραμέτρων παλινδρόμησης είναι ο υπολογισμός κάθε παραμέτρου με ή χωρίς την κάθε παρατήρηση. Τα  $DFBETAS_{ij}$  προτάθηκαν από τους *Belsley et al.* (1980) και στηρίζονται στην επίδραση που έχει η  $i$ -οστή παρατήρηση στον  $j$ -οστό συντελεστή παλινδρόμησης. Στην απλή γραμμική παλινδρόμηση, μπορούμε να θέσουμε ότι για  $j=0$  αντιστοιχεί στο  $\hat{b}_0$  και για  $j=1$  αντιστοιχεί στο  $\hat{b}_1$ . Τότε προκύπτει ότι:

$$DFBETAS_{i0} = \frac{\hat{b}_0 - \hat{b}_0(i)}{s(i) \sqrt{\sum X^2 / nS_{XX}}}, \quad (20)$$

$$DFBETAS_{i1} = \frac{\hat{b}_1 - \hat{b}_1(i)}{s(i)\sqrt{1/S_{xx}}}, \quad (21)$$

όπου τα  $\hat{b}_j(i)$  και  $s(i)$  υποδηλώνουν ότι η  $i$ -οστή παρατήρηση αφαιρείται από τον υπολογισμό της εκτίμησης των τιμών των παραμέτρων.

Εάν είναι προτιμότερο να δοθεί προσοχή στην πρόβλεψη παρά στην εκτίμηση, είναι λογικό να χρησιμοποιηθεί ένα μέτρο σχετικό με τη διαφορά των προβλεπόμενων τιμών με και χωρίς την  $i$ -οστή παρατήρηση. Αυτό το διαγνωστικό μέτρο λέγεται  $DFFITS_i$  και ορίζεται με τον παρακάτω τρόπο:

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_i(i)}{s(i)\sqrt{h_i}}, \quad (22)$$

όπου το  $\hat{y}_i(i)$  υποδηλώνει ότι η  $i$ -οστή προβλεπόμενη τιμή προέκυψε χωρίς τη χρήση της  $i$ -οστής παρατήρησης.

Οι κρίσιμες τιμές για τα  $DFBETAS_{ij}$  και  $DFFITS_i$ , που έχουν προταθεί από τους *Belsley et al.*, είναι  $2/\sqrt{n}$  και  $2/\sqrt{(p+1)/n}$  αντίστοιχα.

Ένα επιπλέον μέτρο της επίδρασης ενός τέτοιου σημείου στο εκτιμηθέν μοντέλο που χρησιμοποιείται συχνά και σχετίζεται κάπως με το  $DFFITS_i$  είναι η απόσταση του Cook (*Cook's Distance*) που ορίζεται ως εξής:

$$D_i = \frac{1}{ps^2} \sum_{j=1}^n (\hat{Y}_{j(i)} - \hat{Y}_j)^2, \quad (23)$$

όπου  $\hat{Y}_{j(i)}$  η  $j$ -οστή προβλεπόμενη τιμή όταν η παρατήρηση  $i$  δεν συμπεριλαμβάνεται στους υπολογισμούς και  $p$  ο αριθμός των παραμέτρων. Κατά συνέπεια, στην απλή παλινδρόμηση, ισχύει ότι  $p=2$  [55].

### 3.1.5. Έκτροπες παρατηρήσεις (Outliers)

Στη στατιστική βιβλιογραφία, η ανάλυση των έκτροπων παρατηρήσεων παίζει πάντα σημαντικό ρόλο [12]. Οι έκτροπες παρατηρήσεις αποτελούν δεδομένα που δεν είναι συνηθισμένα σε σχέση με τα υπόλοιπα δεδομένα των μετρήσεων. Ανάλογα με τη θέση τους στο χώρο, οι παρατηρήσεις αυτές μπορεί να έχουν μέτριες έως σοβαρές επιπτώσεις στην εκτίμηση του μοντέλου παλινδρόμησης [45]. Ο εντοπισμός των έκτροπων σημείων είναι στενά συνδεδεμένος με τον εντοπισμό των σημείων επιρροής, καθώς τα έκτροπα σημεία είναι αρκετές φορές και σημεία επιρροής [55].

Πριν εφαρμοστεί η μέθοδος παλινδρόμησης είναι απαραίτητο να διερευνηθεί προσεκτικά κατά πόσον μπορεί να ερμηνευθεί η ασυνήθιστη συμπεριφορά των έκτροπων σημείων.

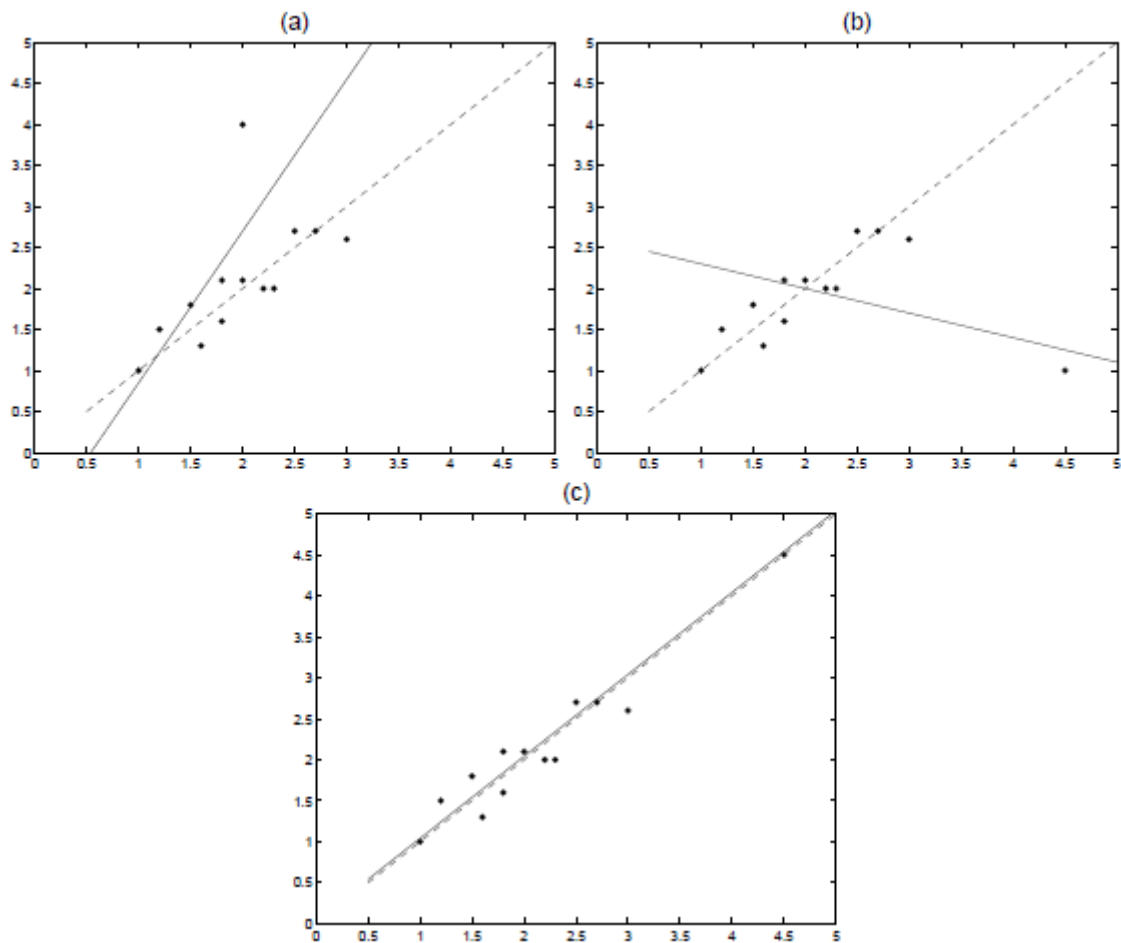


Μερικές φορές τα έκτροπα σημεία θεωρούνται ότι έχουν «κακές» τιμές, με αποτέλεσμα να σημειώνονται ασυνήθιστα αλλά δικαιολογήσιμα περιστατικά. Σχετικά παραδείγματα περιλαμβάνουν λανθασμένη μέτρηση ή ανάλυση, ανακριβή καταγραφή των δεδομένων και αποτυχία κατά τη λειτουργία του οργάνου μέτρησης. Εάν ισχύει κάτι από τα προηγούμενα, τότε το σημείο αυτό πρέπει να διορθωθεί (αν είναι εφικτό) ή να διαγραφεί από το σύνολο των δεδομένων. Απορριψη τέτοιων τιμών είναι αναγκαία, καθώς η μέθοδος των ελάχιστων τετραγώνων θα μετακινήσει την προσαρμοσμένη ευθεία προς το έκτροπο σημείο. Παρόλα αυτά πρέπει να ληφθεί σοβαρά υπόψη και το γεγονός ότι πρέπει να υπάρχει επιπρόσθετα μία ισχυρή (μη-στατιστική) ένδειξη ότι πρόκειται όντως για «κακή» τιμή προτού αυτή απορριφθεί [45].

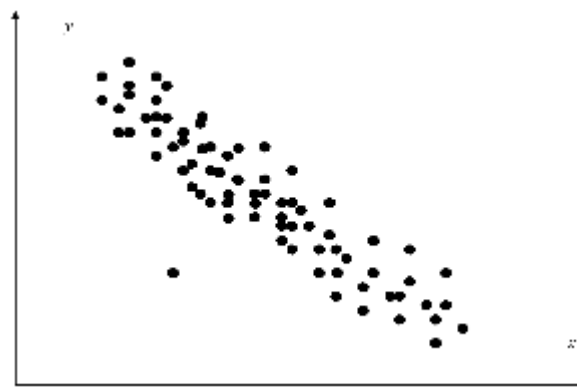
Υπάρχουν περιπτώσεις που μία έκτροπη παρατήρηση μπορεί να είναι ασυνήθιστη αλλά ταυτόχρονα και απόλυτα αποδεκτή. Διαγράφοντας αυτά τα σημεία για να «βελτιωθεί η προσαρμογή της εξίσωσης», μπορεί να είναι επικίνδυνο, καθώς μπορεί να οδηγήσει σε εσφαλμένη αίσθηση της ακρίβειας της εκτίμησης ή της πρόβλεψης. Περιστασιακά διαπιστώνουμε ότι το έκτροπο σημείο είναι περισσότερο σημαντικό από τα υπόλοιπα δεδομένα διότι μπορεί να ελέγχει αρκετές βασικές ιδιότητες του μοντέλου και να επισημαίνει ανεπάρκειες [45].

Οι έκτροπες παρατηρήσεις διακρίνονται στις παρακάτω κατηγορίες λαμβάνοντας υπόψη τη θέση τους στο διάγραμμα διασποράς:

- **$x$ -έκτροπη παρατήρηση.** Αποτελεί μια συνηθισμένη κατηγορία έκτροπων παρατηρήσεων και αφορά εκείνη που παρεκκλίνει αναφορικά με τις τιμές της ανεξάρτητης μεταβλητής  $x$  και ονομάζεται μοχλός ή σημείο μόχλευσης (*leverage point*). Μια παρατήρηση καλής μόχλευσης (*good leverage observation*) θεωρείται εκείνη που έχει ασυνήθιστα μεγάλη ή μικρή τιμή της μεταβλητής  $X$  συγκριτικά με τις υπόλοιπες μετρήσεις, αλλά βρίσκεται πολύ κοντά στη γραμμή παλινδρόμησης και δεν την επηρεάζει σημαντικά αν απομακρυνθεί από τις παρατηρήσεις, σε αντίθεση με εκείνη που είναι κακής μόχλευσης (*bad leverage observation*) (**Σχήμα 3. 2. b**). Θα πρέπει να σημειωθεί ότι οι μοχλοί δεν αντιστοιχούν απαραίτητα και σε έκτροπα σημεία.
- **$y$ -έκτροπη παρατήρηση.** Οι έκτροπες παρατηρήσεις στην  $y$ -κατεύθυνση συμπεριλαμβάνουν τα σημεία που παρουσιάζουν ασυνήθιστη τιμή της εξαρτημένης μεταβλητής  $y$ . Μπορούν να θεωρηθούν έκτροπα, γιατί παρουσιάζουν μεγάλη τιμή του τυποποιημένου κατάλοιπου [55] και επιδρούν στην εκτίμηση της παλινδρόμησης καθώς αυξάνουν το μέγεθος των καταλοίπων. Για παράδειγμα, η  $y$ -έκτροπη παρατήρηση στο **Σχήμα 3. 2 a** έχει σημαντική επίδραση στην κλίση της γραμμής παλινδρόμησης και την εκτίμηση της τυπικής απόκλισης των καταλοίπων.
- **$x$ - και  $y$ - έκτροπες παρατηρήσεις.** Υπάρχουν πολλές περιπτώσεις που οι παρατηρήσεις δεν μπορούν να εντοπιστούν, όταν κάθε μεταβλητή εξετάζεται ανεξάρτητα. Ενδεικτικό παράδειγμα αποτελεί η έκτροπη παρατήρηση, που εντοπίζεται κάτω αριστερά στο παρακάτω σχήμα (**Σχήμα 3. 3**) και χαρακτηρίζεται ως πολυμεταβλητή (*multivariate outlier*). Εάν μελετηθεί κάθε μέτρηση ξεχωριστά σε σχέση με το εύρος των τιμών στους άξονες  $x$  και  $y$ , θα διαπιστώσουμε ότι πέφτουν πολύ κοντά στο κέντρο των μονομεταβλητών κατανομών. Συνεπώς, κατά τον έλεγχο ανίχνευσης αυτών των σημείων, πρέπει να λαμβάνεται υπόψη η σχέση μεταξύ των δύο μεταβλητών.



**Σχήμα 3. 2.** Δεδομένα με έκτροπες παρατηρήσεις. Η συνεχής γραμμή αντιστοιχεί στον εκτιμητή ελαχίστων τετραγώνων ενώ η διακεκομμένη σε ανθεκτικό εκτιμητή. (a)  $y$  –ακραία παρατήρηση, (b) κακό σημείο μόγλευσης, (c) καλό σημείο μόγλευσης [1].



**Σχήμα 3. 3.** Χώρος δύο διαστάσεων με μία έκτροπη παρατήρηση.

Κατηγοριοποιώντας τις μεθόδους που υπάρχουν στη βιβλιογραφία για τον εντοπισμό των έκτροπων παρατηρήσεων στη γραμμική παλινδρόμηση προκύπτουν δύο ομάδες. Η πρώτη αφορά τις **άμεσες** (*direct*), οι οποίες βασίζονται στα αρχικά δεδομένα και τα κριτήριά τους είναι συναρτήσεις των καταλοίπων και της μοχλότητας, και η δεύτερη είναι οι **έμμεσες** (*indirect*), οι οποίες χρησιμοποιούν κυρίως τα κατάλοιπα που προκύπτουν από μια ανθεκτική παλινδρόμηση [1].

Στις άμεσες μεθόδους συναντάμε πολλά διαγνωστικά μέτρα τα οποία στηρίζονται στα κατάλοιπα που απορρέουν από τη μέθοδο των ελαχίστων τετραγώνων και έχουν σχεδιαστεί με σκοπό να ανιχνεύσουν μεμονωμένες περιπτώσεις ή σύνολο περιπτώσεων που διαφέρουν από τα υπόλοιπα δεδομένα ή ασκούν μεγαλύτερη επίδραση στους εκτιμητές. Θα πρέπει να σημειωθεί ότι ο εντοπισμός των έκτροπων παρατηρήσεων μέσω των διαγνωστικών μέτρων πολλές φορές αποτυγχάνει, καθώς ο εκτιμητής των ελαχίστων τετραγώνων, στην προσπάθειά του να αποφύγει τα μεγάλα κατάλοιπα, προσαρμόζει μία και μόνο έκτροπη παρατήρηση εις βάρος των υπόλοιπων παρατηρήσεων, η οποία μπορεί να προκαλέσει μια φτωχή προσαρμογή στην πλειοψηφία των δεδομένων. Επομένως, μια έκτροπη παρατήρηση μπορεί να έχει μικρό κατάλοιπο, όπως συμβαίνει με τα σημεία μόχλευσης, αλλά η ανίχνευσή της να μην είναι εφικτή.

Μία εμπειροτεχνική μέθοδος, που προτάθηκε από τους *Hoaglin* και *Welsch* (1978), είναι να θεωρηθεί ένα σημείο έκτροπο εάν ισχύει ότι  $h_i > 2p/n$ , όπου  $h_i$  είναι τα διαγώνια στοιχεία του πίνακα  $H$  (*Hat matrix*) που μετρούν τη μοχλότητα της κάθε παρατήρησης  $(x_i, y_i)$  στη γραμμή παλινδρόμησης,  $p$  είναι ο αριθμός των παραμέτρων του μοντέλου και  $n$  είναι το πλήθος των παρατηρήσεων. Στην περίπτωση της απλής γραμμικής παλινδρόμησης, η ανισότητα αυτή γίνεται  $h_i > 4/n$ . Αξίζει να τονιστεί ότι ο πίνακας  $H$ , που ορίζεται ως  $H = X(X^T X)^{-1} X^T$ , βασίζεται μόνο στις ανεξάρτητες μεταβλητές  $x_i$  και αγνοούνται τελείως τα έκτροπα σημεία ως προς την εξαρτημένη μεταβλητή  $y_i$ .

Ένα άλλο διαγνωστικό μέτρο που μετράει την απόσταση κάθε παρατήρησης από το κέντρο μάζας των ανεξάρτητων μεταβλητών είναι η απόσταση *Mahalanobis* (*Mahalanobis distance*) [18], η οποία δίνεται από τη σχέση:

$$MD_i^2 = (x_i - \bar{x}) \Sigma^{-1} (x_i - \bar{x})^T, \quad (24)$$

όπου  $\Sigma^{-1}$  είναι ο πίνακας συνδιακύμανσης των διανυσμάτων  $x_i$  που δίνεται από την εξίσωση

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x}). \quad (25)$$

Η σχέση που συνδέει τη *Mahalanobis* απόσταση και τα διαγώνια στοιχεία του πίνακα  $H$  είναι:

$$MD_i^2 = (n-1) \left[ h_i - \frac{1}{n} \right]. \quad (26)$$

Έχουν προταθεί και άλλα διαγνωστικά μέτρα που βασίζονται στα κατάλοιπα της γραμμικής παλινδρόμησης, καθώς ο πίνακας  $H$  είναι χρήσιμος μόνο στην περίπτωση που περιέχεται ένα και μόνο σημείο μόχλευσης στα δεδομένα. Η πιο κοινή μέθοδος είναι όταν τα κατάλοιπα είναι σημαντικά μεγαλύτερα κατά απόλυτη τιμή συγκριτικά με τα υπόλοιπα, με τυπική απόκλιση τρεις ή τέσσερις φορές μακριά από τη μέση τιμή [45]. Επίσης, μπορούν να χρησιμοποιηθούν και τα τυποποιημένα κατάλοιπα (*standardized residuals*), τα οποία

προκύπτουν όταν τα κατάλοιπα διαιρεθούν με την τυπική απόκλιση των καταλοίπων. Πιο συγκεκριμένα, το  $i$ -οστό τυποποιημένο κατάλοιπο ισούται με:

$$\hat{\varepsilon}_i' = \frac{\hat{\varepsilon}_i}{s_{\varepsilon} \sqrt{1-h_i}}, \quad (27)$$

όπου  $s_{\varepsilon}$  είναι η τυπική απόκλιση των καταλοίπων και  $h_i$  η τιμή μόχλευσης του  $i$ -οστού σημείου που ισούται με  $h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$ .

Αν σε κάποιο σύνολο δεδομένων περιέχεται μία μόνο έκτροπη παρατήρηση, τότε ο εντοπισμός και η αντιμετώπισή της είναι σχετικά μια εύκολη διαδικασία με τα παραπάνω διαγνωστικά μέτρα. Όμως η πιο συνηθισμένη περίπτωση είναι να εμφανίζονται πολλά έκτροπα σημεία ή ομάδες αυτών, καθιστώντας το πρόβλημα εντοπισμού τους πολύ πιο δύσκολο. Είναι γνωστό ότι φαινόμενα επικάλυψης (*masking*) και εμπύθισης (*swamping*) αποτελούν κοινά προβλήματα σε τέτοιες διαδικασίες εντοπισμού έκτροπων παρατηρήσεων από ένα σύνολο μετρήσεων [26, 30] και ειδικά σε πολυμεταβλητά δεδομένα. Το φαινόμενο επικάλυψης εμφανίζεται όταν μία ομάδα από έκτροπες παρατηρήσεις δεν μπορεί να ανιχνευθεί λόγω της παρουσίας μιας άλλης, συνήθως γειτονικής, ομάδας [1]. Το φαινόμενο της εμπύθισης εμφανίζεται όταν καλές παρατηρήσεις αναγνωρίζονται εσφαλμένα ως έκτροπες, εξαιτίας της παρουσίας μιας άλλης, συνήθως απόμακρης, ομάδας παρατηρήσεων.

Μία άλλη κατηγορία διαγνωστικών μέτρων των άμεσων μεθόδων στηρίζεται στη διαγραφή μίας παρατήρησης διαδοχικά ή ομάδας αυτών κάθε φορά σε επαναληπτική διαδικασία. Αυτά τα διαγνωστικά μέτρα μετρούν το μέγεθος της επίδρασης που έχει η διαγραφή μιας παρατήρησης ή ομάδας αυτών στην εκτίμηση των παραμέτρων σε σχέση με την αρχική εκτίμηση από το σύνολο των δεδομένων. Τέτοια είναι τα  $DFBETAS_{ij}$  και  $DFFITS_i$ , που έχουν αναπτυχθεί στο προηγούμενο εδάφιο.

Ένα συχνά χρησιμοποιούμενο διαγνωστικό μέτρο είναι η απόσταση του *Cook* (*Cook's Distance*), η οποία ορίστηκε σε προηγούμενη ενότητα, και μετράει το μέγεθος της αλλαγής που πραγματοποιείται στους συντελεστές παλινδρόμησης όταν αφαιρεθεί μία παρατήρηση από το δείγμα.

Όπως ήδη αναφέρθηκε, τα παραπάνω διαγνωστικά μέτρα βασίζονται σε κάποια συνάρτηση των καταλοίπων  $\hat{\varepsilon}_i$  της παλινδρόμησης καθώς και στα διαγώνια στοιχεία  $h_i$  του πίνακα  $H$ . Δυστυχώς όμως, καθίστανται αναξιόπιστα και ανίσχυρα όταν στα δεδομένα περιέχονται περισσότερες από μία έκτροπες παρατηρήσεις σε συνδυασμό με το φαινόμενο της επικάλυψης. Κρίνεται λοιπόν αναγκαία η εφαρμογή των ανθεκτικών μεθόδων για την επιτυχή αντιμετώπιση των προαναφερθέντων προβλημάτων κατά τον εντοπισμό των έκτροπων παρατηρήσεων.

### 3.2. ΠΑΛΙΝΔΡΟΜΗΣΗ ΜΕ ΣΦΑΛΜΑΤΑ ΜΕΤΡΗΣΕΩΣ ΣΤΙΣ ΜΕΤΑΒΛΗΤΕΣ

Στη γραμμική παλινδρόμηση (βλ. Υποκεφ. 3.1.) έχουμε υποθέσει είτε ότι δεν υπάρχουν σφάλματα στις μεταβλητές είτε ότι υπεισέρχονται σφάλματα μόνο στην εξαρτημένη

μεταβλητή. Παρόλα αυτά, η υπόθεση αυτή δεν ανταποκρίνεται στην πραγματικότητα, γιατί τόσο η ανεξάρτητη όσο και η εξαρτημένη μεταβλητή είναι πιθανό να περιέχουν σφάλματα μετρήσεως, τα οποία μπορεί να οφείλονται σε σφάλματα δειγματοληψίας, σε σφάλματα οργάνων καθώς και σε άλλα αίτια. Η περίπτωση αυτή αναφέρεται στη βιβλιογραφία ως **σφάλματα μετρήσεως στις μεταβλητές (*Error-In-Variables or EIV*)**.

Ξεκινάμε παίρνοντας την απλούστερη περίπτωση. Δηλαδή θεωρούμε δύο μεταβλητές  $X$  και  $Y$ , οι οποίες συνδέονται γραμμικά με την παρακάτω συναρτησιακή σχέση (*functional relation*) (Σχέση (28)) και θέλουμε να προσδιορίσουμε τα  $b_0$  και  $b_1$ :

$$Y = b_0 + b_1 X. \quad (28)$$

Επειδή δεν μπορούμε να παρατηρήσουμε τις ακριβείς τιμές των μεταβλητών  $X$  και  $Y$ , μπορούμε να παρατηρήσουμε μόνο τις τυχαίες μεταβλητές  $\xi$  και  $\eta$  που ορίζονται ως εξής:

$$\left. \begin{aligned} \xi_i &= X_i + \delta_i \\ \eta_i &= Y_i + \varepsilon_i, \end{aligned} \right\} \quad (29)$$

όπου  $\delta_i$  και  $\varepsilon_i$  είναι τα σφάλματα μετρήσεως των μεταβλητών  $X$  και  $Y$  για  $i=1,2,\dots,n$ . Αξίζει να σημειωθεί ότι γενικά, τα σφάλματα μπορεί να εξαρτώνται από το  $i$ . Για παράδειγμα, για μεγάλες τιμές της μεταβλητής  $X$  τα σφάλματα μπορεί να έχουν την τάση να μεγαλώνουν σε σύγκριση με μικρότερες τιμές της  $X$ . Όμως, στην απλή περίπτωση που εξετάζουμε, μπορούμε να υποθέσουμε ότι τα σφάλματα μετρήσεως  $\delta_i$  και  $\varepsilon_i$  έχουν μέσο ίσο με μηδέν και σταθερή διακύμανση για όλα τα  $i$ . Επίσης, υποθέτουμε ότι τα σφάλματα μετρήσεως είναι ασυσχέτιστα μεταξύ των μερών τους αλλά και μεταξύ τους. Δηλαδή υποθέτουμε ότι:

$$\left. \begin{aligned} E(\delta_i) &= E(\varepsilon_i) = 0, \quad i = 1, 2, \dots, n \\ V(\delta_i) &= \sigma_\delta^2, \quad V(\varepsilon_i) = \sigma_\varepsilon^2, \quad i = 1, 2, \dots, n \\ Cov(\delta_i, \delta_j) &= Cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j, \quad Cov(\delta_i, \varepsilon_j) = 0, \quad \forall i, j. \end{aligned} \right\} \quad (30)$$

Για λόγους απλοποίησης, η σχέση (29) μπορεί να γραφεί ως εξής:

$$\left. \begin{aligned} \xi &= X + \delta, \\ \eta &= Y + \varepsilon. \end{aligned} \right\} \quad (31)$$

Με μια πρώτη ματιά, η σχέση (28) μοιάζει με εξίσωση παλινδρόμησης, γεγονός που μπορεί να δημιουργήσει σύγχυση. Στην πραγματικότητα η (28) καταλήγει σε περίπτωση κλασσικής παλινδρόμησης όταν  $V(\delta_i) = \sigma_\delta^2 = 0$ .

Αντικαθιστώντας τη σχέση (31) στην (28) προκύπτει ότι:

$$\eta = b_0 + b_1\xi + (\varepsilon - b_1\delta). \quad (32)$$

Η παραπάνω εξίσωση ονομάζεται δομική συνάρτηση (*structural relation*) μεταξύ των παρατηρούμενων τυχαίων μεταβλητών  $\xi$  και  $\eta$  και είναι αποτέλεσμα της συναρτησιακής σχέσης μεταξύ των μαθηματικών μεταβλητών  $X$  και  $Y$  [38].

Όμως, σε πρακτικές εφαρμογές, υποθέτουμε ότι οι  $X$  και  $Y$  είναι τυχαίες μεταβλητές και κάνοντας την παρακάτω υπόθεση  $Cov(x, \delta) = Cov(x, \varepsilon) = Cov(y, \delta) = Cov(y, \varepsilon) = 0$  (33), το μοντέλο που θα ισχύει τώρα έχει την εξής μορφή:

$$\xi_i = x_i + \delta_i, \quad \eta_i = y_i + \varepsilon_i \quad (34)$$

$$y_i = b_0 + b_1x_i. \quad (35)$$

Το παρόν μοντέλο αποτελεί μία γενίκευση του προηγούμενου. Στην πραγματικότητα, η συναρτησιακή σχέση (Σχέση (28)) μεταξύ των μεταβλητών έχει αντικατασταθεί από μία δομική σχέση (Σχέση (35)), που εκφράζει μία απόλυτα γραμμική σχέση μεταξύ δύο μη παρατηρήσιμων τυχαίων μεταβλητών  $x$ ,  $y$ .

Η προσέγγιση της μεθόδου των ελαχίστων τετραγώνων στην ανάλυση παλινδρόμησης δεν είναι κατάλληλη όταν εφαρμόζεται για την εκτίμηση των  $b_0$  και  $b_1$ , γιατί τείνει στην υποτίμηση της τιμής της παραμέτρου  $b_1$ , που εκφράζει την κλίση της γραμμής [51]. Από την άλλη πλευρά, μπορούμε να εφαρμόσουμε τη μέθοδο μέγιστης πιθανοφάνειας για να εκτιμήσουμε τις παραμέτρους του μοντέλου δομικής σχέσης, όπως έχει οριστεί από τις σχέσεις (32), (33), (34) και (35). Επιπρόσθετα, υποθέτουμε ότι τα ζευγάρια των παρατηρήσεων  $\xi_i$  και  $\eta_i$  είναι ισόνομα (έχουν την ίδια κατανομή) και ακολουθούν την κανονική κατανομή. Τότε, έχουμε ότι:

$$\left. \begin{aligned} E(\xi) = E(x) = \mu, \quad E(\eta) = E(y) = b_0 + b_1\mu \\ V(\xi) = \sigma_x^2 + \sigma_\delta^2, \quad V(\eta) = b_1^2\sigma_x^2 + \sigma_\varepsilon^2 \\ Cov(\xi, \eta) = Cov(x, y) = b_1\sigma_x^2. \end{aligned} \right\} \quad (36)$$

Στο μοντέλο (36) υπάρχουν έξι παράμετροι: οι παράμετροι της δομικής συνάρτησης  $b_0$  και  $b_1$ , τα σφάλματα διακύμανσης  $\sigma_\delta^2$  και  $\sigma_\varepsilon^2$ , ο μέσος  $\mu$  και η διακύμανση  $\sigma_x^2$ . Εφαρμόζοντας τη μέθοδο μέγιστης πιθανοφάνειας, προκύπτουν πέντε στατιστικές για τις αντίστοιχες παραμέτρους της διμεταβλητής κανονικής κατανομής, που αποτελούν και τους εκτιμητές μέγιστης πιθανοφάνειας αυτών των παραμέτρων [38].

$$\left. \begin{aligned} \bar{\xi} &= \hat{\mu}, & \bar{\eta} &= \hat{b}_0 + \hat{b}_1 \hat{\mu}, \\ S_x^2 &= \hat{\sigma}_x^2 + \hat{\sigma}_\delta^2, & S_y^2 &= \hat{b}_1^2 \hat{\sigma}_x^2 + \hat{\sigma}_\varepsilon^2, & S_{xy} &= \hat{b}_1 \hat{\sigma}_x^2. \end{aligned} \right\} \quad (37)$$

Επειδή, όμως, στην περίπτωση αυτή το μοντέλο δεν είναι αναγνωρίσιμο, καθώς ο αριθμός των άγνωστων παραμέτρων είναι μεγαλύτερος από τον αριθμό των εξισώσεων, απαιτούνται περαιτέρω υποθέσεις. Έτσι, διακρίνουμε περιπτώσεις ανάλογα με τα σφάλματα διακύμανσης  $\sigma_\delta^2$  και  $\sigma_\varepsilon^2$ .

**1<sup>η</sup> περίπτωση:**  $\sigma_\delta^2$  γνωστό

$$\hat{b}_1 = \frac{S_{xy}}{S_x^2 - \sigma_\delta^2} \quad (38)$$

και

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}. \quad (39)$$

Αν  $\sigma_\delta^2 = 0$ , επιστρέφουμε στην περίπτωση παλινδρόμησης και πιο συγκεκριμένα, στον εκτιμητή ελαχίστων τετραγώνων.

**2<sup>η</sup> περίπτωση:**  $\sigma_\varepsilon^2$  γνωστό

$$\hat{b}_1 = \frac{S_y^2 - \sigma_\varepsilon^2}{S_{xy}} \quad (40)$$

και  $\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$ .

Αν  $\sigma_\varepsilon^2 = 0$ , που σημαίνει και εδώ ότι μόνο στη μία μεταβλητή υπεισέρχεται σφάλμα, η εκτίμηση του  $b_1$  με τη μέθοδο μέγιστης πιθανοφάνειας ισοδυναμεί με τη γραμμή των ελαχίστων τετραγώνων (χρησιμοποιώντας τη μεταβλητή που είναι απαλλαγμένη από σφάλμα ως παλινδρομητή).

**3<sup>η</sup> περίπτωση:**  $\sigma_\varepsilon^2$  και  $\sigma_\delta^2$  άγνωστα

Θεωρούμε την παράμετρο  $\lambda$  που ισούται με το λόγο των μεταβλητοτήτων των σφαλμάτων διακύμανσης, δηλαδή  $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ . Τότε οι εκτιμητές της μέγιστης πιθανοφάνειας υπολογίζονται ως εξής:

$$\hat{b}_1 = \frac{S_y^2 - \lambda S_x^2 + \sqrt{(S_y^2 - \lambda S_x^2)^2 + 4\lambda S_{xy}^2}}{2S_{xy}} \quad (41)$$

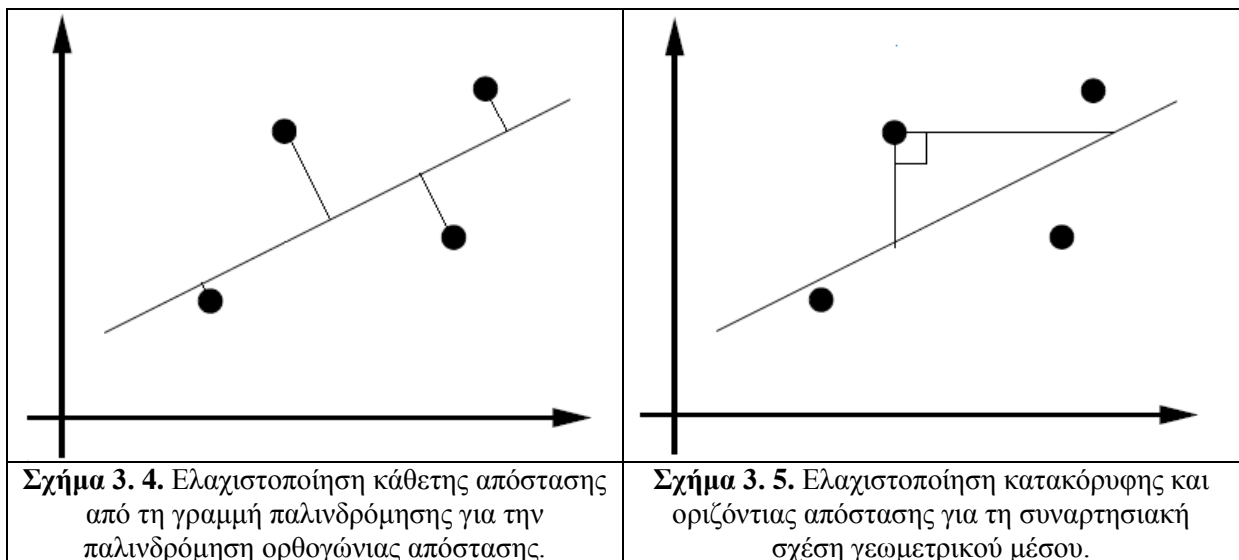
$$\text{και } \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}.$$

Διακρίνονται δύο περιπτώσεις, που χρησιμοποιούνται ευρέως στη θεωρία *EIV*, όταν τα  $\sigma_\varepsilon^2$  και  $\sigma_\delta^2$  είναι άγνωστα:

- Για  $\lambda = 1 \Rightarrow \sigma_\varepsilon^2 = \sigma_\delta^2$ , ακολουθείται η παραπάνω διαδικασία που ονομάζεται στη σχετική βιβλιογραφία **παλινδρόμηση ορθογώνιας απόστασης** (*orthogonal distance regression*), όπου ελαχιστοποιείται το άθροισμα των τετραγώνων της ορθογώνιας απόστασης των παρατηρούμενων σημείων από τη γραμμή παλινδρόμησης (**Σχήμα 3. 4**).
- Για  $\lambda = \frac{S_y^2}{S_x^2}$ , η σχέση (41) είναι γνωστή ως **συναρτησιακή σχέση γεωμετρικού μέσου** (*geometric mean functional relationship*), όπου ελαχιστοποιείται το άθροισμα των εμβαδών που περικλείονται από την οριζόντια και κάθετη προβολή των παρατηρούμενων σημείων με τη γραμμή παλινδρόμησης και την ίδια τη γραμμή παλινδρόμησης (**Σχήμα 3. 5**) [42].

$$\hat{b}_1 = \sqrt{\frac{S_y^2}{S_x^2}}, \quad (42)$$

$$\text{και } \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}.$$



Μια ενδιαφέρουσα ιδιότητα αυτής της περίπτωσης είναι ότι ο εκτιμητής  $b_1$  είναι ο γεωμετρικός μέσος των κλίσεων των εξισώσεων παλινδρόμησης της  $Y$  στην  $X$  και της  $X$  στην  $Y$  και επομένως, ο γεωμετρικός μέσος μπορεί να θεωρηθεί σαν μία μέση τιμή μεταξύ των δύο κλίσεων.



### 3.3. ΑΝΘΕΚΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

#### 3.3.1. Γενικά

Από τις κλασικές μεθόδους παλινδρόμησης, εκείνη που έχει επικρατήσει είναι η μέθοδος των ελαχίστων τετραγώνων εξαιτίας κυρίως της ευκολίας υπολογισμού της. Μικρές αποκλίσεις από τις εκ των προτέρων υποθέσεις, όπως είναι η κανονικότητα, η γραμμικότητα και η ανεξαρτησία, μπορεί να επιφέρουν μικρά σφάλματα στα τελικά αποτελέσματα, καθιστώντας τη μέθοδο λιγότερο βέλτιστη. Εύκολα όμως διαπιστώνεται η ευαισθησία αυτής της μεθόδου να αντιμετωπίσει ένα μεγάλο αριθμό έκτροπων παρατηρήσεων και γενικών σφαλμάτων (*gross errors*), τα οποία εμφανίζονται συχνότατα σε πραγματικά δεδομένα. Ο *Hampel* (1973) δηλώνει ότι ένα ποσοστό 5 με 10% από μικτά σφάλματα εμφανίζονται σε ένα σύνολο δεδομένων, γεγονός που αποτελεί κανόνα παρά εξαίρεση. Ο κίνδυνος από την ύπαρξη αυτών των σφαλμάτων έγκειται στο γεγονός ότι η μέθοδος των ελαχίστων τετραγώνων μπορεί να επηρεάσει σημαντικά τους εκτιμητές και να οδηγήσει σε λανθασμένη πρόβλεψη ή εκτίμηση, καθώς αδυνατεί να χειριστεί τέτοιου είδους παρατηρήσεις.

Η Ανθεκτική Στατιστική (*Robust Statistics*) είναι το σύνολο των γνώσεων και τεχνικών που χρησιμοποιούνται για την εκτίμηση των παραμέτρων ενός παραμετρικού μοντέλου ενώ ταυτόχρονα αντιμετωπίζει και τις αποκλίσεις από τις υποθέσεις που γίνονται στη στατιστική. Οι βασικότεροι στόχοι της ανθεκτικής στατιστικής είναι:

- η περιγραφή της δομής της καλύτερης προσαρμογής στην πληθώρα των δεδομένων,
- ο εντοπισμός των έκτροπων παρατηρήσεων και η μέθοδος αντιμετώπισής τους και
- η αναγνώριση και η προειδοποίηση σημείων μόχλευσης ή/και επιρροής.

Η ανθεκτική παλινδρόμηση έχει ως στόχο τη δημιουργία εκτιμητών που δεν επηρεάζονται από τις έκτροπες παρατηρήσεις. Ουσιαστικά ελέγχει τα κατάλοιπα, από όπου μπορούν να εντοπιστούν οι έκτροπες παρατηρήσεις, κάτι το οποίο δεν μπορεί να επιτευχθεί με τα μέσα των καταλοίπων των ελαχίστων τετραγώνων. Κατά συνέπεια, τα διαγνωστικά μέτρα και η ανθεκτική παλινδρόμηση ενώ έχουν τον ίδιο στόχο, εκτελούν την ακριβώς αντίθετη διαδικασία: με τη χρήση των διαγνωστικών εργαλείων, σε πρώτη φάση διαγράφονται τα έκτροπα σημεία και μετά γίνεται η προσπάθεια προσαρμογής των «καλών» δεδομένων με τα ελάχιστα τετράγωνα ενώ η ανθεκτική ανάλυση πρώτα προσαρμόζει μία γραμμή παλινδρόμησης στην πλειοψηφία των δεδομένων και σε δεύτερη φάση προσπαθεί να εντοπίσει τις έκτροπες παρατηρήσεις που έχουν μεγάλα κατάλοιπα μέσω της συγκεκριμένης ανθεκτικής λύσης [54].

Το επόμενο βήμα είναι να αναλυθεί η δομή που προκύπτει με τη βοήθεια αυτής τη μεθόδου. Για παράδειγμα, μπορεί κάποιος να ανατρέξει στα αρχικά δεδομένα, να μελετήσει τις έκτροπες παρατηρήσεις και να προσπαθήσει να εξηγήσει την προέλευσή τους. Επίσης, μπορεί κάποιος να εξετάσει αν οι αποκλίσεις οφείλονται σε αποτυχημένο μοντέλο, γεγονός που θα μπορούσε, για παράδειγμα, να αντιμετωπιστεί με την εισαγωγή ενός δευτεροβάθμιου όρου.

### 3.3.2. Μέτρα Ανθεκτικότητας

Η ανθεκτική παλινδρόμηση παρέχει έναν εναλλακτικό τρόπο από τη μέθοδο των ελαχίστων τετραγώνων, όταν οι βασικές υποθέσεις δεν πληρούνται λόγω της φύσης των δεδομένων, και έχει ως βασικό στόχο τον προσδιορισμό ενός παραμετρικού μοντέλου, το οποίο να περιγράψει τη γενική πληροφορία και συμπεριφορά της πλειοψηφίας των δεδομένων που εμπεριέχουν έκτροπες παρατηρήσεις. Σε θεωρητική βάση, έχουν οριστεί συγκεκριμένα μέτρα τα οποία χρησιμοποιούνται για να προσδιορίσουν το βαθμό ανθεκτικότητας μιας διαδικασίας και στα οποία γίνεται αναφορά στη συνέχεια. Καθένα από τα μέτρα αυτά περιγράφει διαφορετικά χαρακτηριστικά της διαδικασίας, συνεπώς λειτουργούν συμπληρωματικά.

#### 3.3.2.1. Σημείο κατάρρευσης

Οι εκτιμητές θα πρέπει να αντέχουν σε δεδομένα τα οποία περιέχουν ένα συγκεκριμένο ποσοστό έκτροπων παρατηρήσεων. Προκειμένου να διατυπωθεί το μέγεθος αυτού του ποσοστού με μαθηματικό τρόπο, έχει εισαχθεί ο όρος του **σημείου κατάρρευσης (breakdown point)**. Ας θεωρήσουμε ένα δείγμα από  $n$  παρατηρήσεις  $Z = (x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$  για  $i = 1, 2, \dots, n$  και  $T$  ένας εκτιμητής παλινδρόμησης. Εφαρμόζοντας τον εκτιμητή  $T$  στο δείγμα  $Z$  προκύπτει το διάνυσμα των άγνωστων παραμέτρων παλινδρόμησης, για το οποίο ισχύει  $T(Z) = \hat{b}$ . Τώρα ας θεωρήσουμε όλα τα πιθανά αλλοιωμένα δείγματα  $Z'$ , τα οποία προκύπτουν με αντικατάσταση οποιασδήποτε  $m$  παρατήρησης από τα αρχικά δεδομένα με αυθαίρετες τιμές, επιτρέποντας με αυτό τον τρόπο την είσοδο έκτροπων παρατηρήσεων. Η μέγιστη μεροληψία που μπορεί να προκληθεί από αυτή την αλλοίωση σε όλα τα πιθανά  $Z'$  εκφράζεται ως εξής:

$$bias(m; T, Z) = \sup_{Z'} \|T(Z') - T(Z)\|. \quad (43)$$

Όταν η μέγιστη μεροληψία τείνει στο άπειρο, αυτό σημαίνει ότι οι  $m$  ακραίες παρατηρήσεις έχουν αυθαίρετα μεγάλη επίδραση στον εκτιμητή  $T$  και τότε ο εκτιμητής καταρρέει. Από τους *Donoho* και *Huber* (1983) το σημείο κατάρρευσης ενός εκτιμητή  $T$  σε ένα δείγμα  $Z$  έχει οριστεί ως εξής:

$$\varepsilon_n^*(T, Z) = \min \left\{ \frac{m}{n}; bias(m; T, Z) = \infty \right\}. \quad (44)$$

Με άλλα λόγια, το σημείο κατάρρευσης είναι το μικρότερο κλάσμα της αλλοίωσης που μπορεί να προκαλέσει τον εκτιμητή  $T$  να πάρει αυθαίρετα τιμές μακριά από τον εκτιμητή  $T(Z)$ . Στη μέθοδο των ελαχίστων τετραγώνων, το σημείο κατάρρευσης ισοδυναμεί με:

$$\varepsilon_n^*(T, Z) = \frac{1}{n}, \quad (45)$$

το οποίο τείνει στο μηδέν όσο το μέγεθος  $n$  του δείγματος αυξάνεται και επομένως μπορεί να ειπωθεί ότι το σημείο κατάρρευσης είναι 0%. Αυτό το γεγονός, αντικατοπτρίζει πάλι την αδυναμία της μεθόδου των ελαχίστων τετραγώνων στην αντιμετώπιση των έκτροπων παρατηρήσεων. Πρακτικά, η μεγαλύτερη τιμή του σημείου κατάρρευσης ενός εκτιμητή που

μπορεί κάποιος να ελπίζει είναι το 50%, γιατί τότε, για μεγαλύτερες τιμές, θα είναι αδύνατος ο διαχωρισμός των «καλών» από τις έκτροπες παρατηρήσεις.

Παρόλο που το σημείο κατάρρευσης αποτελεί ένα κοινώς αποδεκτό μέτρο αξιοπιστίας, δεν είναι αρκετό για την εκτίμηση του βαθμού ανθεκτικότητας ενός εκτιμητή. Για αυτό το λόγο δίνεται έμφαση και σε άλλα μέτρα ανθεκτικότητας, που θα αναφερθούν στη συνέχεια.

### 3.3.2.2. Σχετική αποτελεσματικότητα – Ασυμπτωτική μεροληψία – Ασυμπτωτική διακύμανση

Ένας εκτιμητής με υψηλό σημείο κατάρρευσης θα έχει την τάση να παράγει λογικές εκτιμήσεις ακόμα και αν εφαρμόζεται σε δεδομένα με πολλαπλά σημεία υψηλής μόχλευσης, μπορεί εν τούτοις να γίνει λιγότερο ελκυστικός εάν οι εκτιμήσεις που έχουν προκύψει από «καλά» δεδομένα είναι υποκειμένες σε μεγαλύτερα ποσοστά αβεβαιότητας σε σύγκριση με τις εκτιμήσεις από τα ελάχιστα τετράγωνα. Για αυτό το λόγο θα πρέπει να λαμβάνεται υπόψη η **σχετική αποτελεσματικότητα** (*relative efficiency*). Σύμφωνα με τον Andersen (2008), η **αποτελεσματικότητα** (*efficiency*) ενός εκτιμητή για μια συγκεκριμένη παράμετρο ορίζεται ως ο λόγος της ελάχιστης δυνατής διακύμανσης προς την πραγματική. Πιο αυστηρά, ένας εκτιμητής θεωρείται «αποτελεσματικός» όταν αυτός ο λόγος ισούται με τη μονάδα. Ένας εκτιμητής που προσεγγίζει ένα αποδεκτό επίπεδο αποτελεσματικότητας σε μεγάλα δείγματα ονομάζεται **ασυμπτωτικά αποτελεσματικός** (*asymptotic efficient*). Βασικά κριτήρια αποτελεσματικότητας σε μεγάλα δείγματα, σύμφωνα με τον Huber (1981), θεωρούνται η **ασυμπτωτική μεροληψία** (*asymptotic bias*)  $T(F) - T(F_0)$  και η **ασυμπτωτική διακύμανση** (*asymptotic variance*)  $V(T, F)$  σε κάποια περιοχή  $P_\varepsilon(F_0)$  του μοντέλου κατανομής  $F_0$ . Έτσι, σε μια ποσοτική ανθεκτικότητα, η μέγιστη μεροληψία (*maximum bias*)  $b(\varepsilon) = \sup_{F \in P_\varepsilon} |T(F) - T(F_0)|$  και η μέγιστη διακύμανση (*maximum variance*)  $v(\varepsilon) = \max_{F \in P_\varepsilon} V(T, F)$

αποτελούν δύο από τα πιο σημαντικά χαρακτηριστικά. Η σχετική αποτελεσματικότητα συγκρίνει την αποτελεσματικότητα ενός εκτιμητή με κάποιον που έχει προκύψει από μία γνωστή μέθοδο. Στην παλινδρόμηση, οι εκτιμητές συγκρίνονται με αυτούς που προκύπτουν με τη μέθοδο των ελαχίστων τετραγώνων, καθώς παρέχει εκτίμηση με τη μικρότερη διακύμανση, όταν πληρούνται οι προϋποθέσεις, και έτσι θεωρείται ο πιο αποτελεσματικός γνωστός εκτιμητής.

### 3.3.2.3. Συνάρτηση επίδρασης

Η **συνάρτηση επίδρασης** (*influence function – IF*) χρησιμοποιείται στη μελέτη ιδιοτήτων ανθεκτικότητας και την εύρεση ασυμπτωτικών διασπορών και μετράει την επίδραση των απειροστών διαταραχών στον εκτιμητή. Η συνάρτηση επίδρασης ενός εκτιμητή  $T$  μιας κατανομής  $F$  ορίζεται ως

$$IF(x; T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T((1-\varepsilon)F + \varepsilon\Delta_x) - T(F)}{\varepsilon}, \quad (46)$$

όπου  $\Delta_x$  είναι η συνάρτηση πιθανότητας με όλη τη μάζα πιθανότητας συγκεντρωμένη στο σημείο  $x$ .

## 3.3.2.4. Ευαισθησία γενικού σφάλματος

Ένα σημαντικό μέτρο ανθεκτικότητας που προκύπτει από τη συνάρτηση επίδρασης είναι η **ευαισθησία του γενικού σφάλματος** (*gross error sensitivity*) και ορίζεται ως εξής:

$$\gamma^* = \sup_y |IF(x; T, F)|. \quad (47)$$

Το  $\gamma^*$  περιγράφει τη μέγιστη επίδραση που επιφέρει μία μικρή αλλοίωση της κατανομής στην τιμή του εκτιμητή. Για την εξασφάλιση της ανθεκτικότητας επιδιώκεται να είναι πεπερασμένες οι τιμές του  $\gamma^*$  καθώς έχουν σαν αποτέλεσμα να προκαλούν μικρές αλλαγές στον εκτιμητή σε περίπτωση που προστεθεί ένας μικρός αριθμός έκτροπων παρατηρήσεων.

## 3.3.2.5. Καμπύλη ευαισθησίας

Η **καμπύλη ευαισθησίας** (*sensitivity curve – SC*) μετράει την επίδραση που θα ασκηθεί στον εκτιμητή από μία μόνο επιπρόσθετη παρατήρηση  $x$ . Η καμπύλη ευαισθησίας ενός εκτιμητή  $T_n$  σε ένα σημείο  $x$  δοθέντος ενός τυχαίου δείγματος  $x_1, \dots, x_{n-1}$  ορίζεται από τη σχέση:

$$SC_n(x) = n[T(x_1, \dots, x_{n-1}, x) - T_n(x_1, \dots, x_{n-1})]. \quad (48)$$

## 3.3.2.6. Ίση μεταβολή ως προς την παλινδρόμηση και κλίμακα

Τέλος, υπάρχουν ορισμένες ιδιότητες ίσης μεταβολής για τους εκτιμητές θέσης  $T_n$  και κλίμακας  $S_n$ , οι οποίες, όταν ισχύουν, δείχνουν τον τρόπο με τον οποίο μεταβάλλεται ένας εκτιμητής στους διάφορους μετασχηματισμούς των δεδομένων. Ταξινομώντας αυτές τις ιδιότητες ξεκινώντας από εκείνη με την υψηλότερη προτεραιότητα και καταλήγοντας σε εκείνη με τη χαμηλότερη, για έναν εκτιμητή θέσης  $T_n$  που στηρίζεται στο τυχαίο δείγμα  $(x, y)$ , θα έχουμε ότι ο  $T_n$  είναι:

1. Ίσης μεταβολής ως προς την παλινδρόμηση (*regression equivariant*) όταν

$$T_n(x_i, y_i + x_i v) = T_n(x_i, y_i) + v, \quad (49)$$

για κάθε διάνυσμα  $v$ .

2. Ίσης μεταβολής ως προς την κλίμακα (*scale equivariant*) όταν

$$T_n(x_i, c y_i) = c T_n(x_i, y_i), \quad (50)$$

για κάθε σταθερά  $c$ .

3. Ίσης μεταβολής ως προς την παλινδρόμηση και κλίμακα (*affine equivariant*) όταν

$$T_n(x_i A, y_i) = A^{-1} T_n(x_i, y_i), \quad (51)$$

για κάθε αντιστρέψιμο πίνακα  $A$ , που σημαίνει ότι ένας γραμμικός μετασχηματισμός του  $x_i$  θα πρέπει να μετασχηματίσει ανάλογα και τον εκτιμητή  $T_n$ , διότι  $\hat{y}_i = x_i T_n = (x_i A)(A^{-1} T_n)$ .

Η συμπεριφορά ενός εκτιμητή κλίμακας  $S_n$  που υπόκειται σε αντίστοιχους μετασχηματισμούς θα είναι ίσης μεταβολής ως προς την κλίμακα (*scale equivariant*) όταν ισχύει

$$S_n(x_i, cy_i) = |c| S_n(x_i, y_i), \quad (52)$$

για κάθε σταθερά  $c$ . Να σημειωθεί ότι η απόλυτη τιμή είναι απαραίτητη καθώς οι εκτιμητές κλίμακας είναι πάντα θετικοί.

Οι *Huber* και *Ronchetti* (2009) επισημαίνουν τρεις ιδιότητες που θα πρέπει να έχει μία «καλή» ανθεκτική διαδικασία παλινδρόμησης:

- Όταν το υποτιθέμενο μοντέλο είναι σωστό, θα πρέπει τα αποτελέσματα της διαδικασίας που εφαρμόζεται να έχει μικρή δειγματική διασπορά, δηλαδή να έχει αποτελεσματικότητα.
- Μικρές αποκλίσεις από τις υποθέσεις του μοντέλου θα πρέπει να έχουν την ελάχιστη επίδραση στην απόδοση του εκτιμητή.
- Μεγάλες αποκλίσεις από τις υποθέσεις του μοντέλου δεν θα πρέπει να οδηγούν σε καταστροφή του εκτιμητή.

### 3.3.3. Ανθεκτικοί εκτιμητές

Σε αυτό το σημείο θα αναλυθούν οι βασικότεροι ανθεκτικοί εκτιμητές που έχουν αναπτυχθεί για τη θεωρία της ανθεκτικής ανάλυσης. Ενώ οι αρχικές τεχνικές ανθεκτικής παλινδρόμησης παρουσίαζαν αδυναμίες αναφορικά με ένα ή περισσότερα από τα προαναφερθέντα μέτρα, εν τούτοις, πολλές από τις πιο πρόσφατες μεθόδους συμπεριφέρονται καλύτερα όσον αφορά αυτές τις ιδιότητες, απεικονίζοντας με αυτό τον τρόπο την εξέλιξη αυτών των εκτιμητών. Η ανάλυση ξεκινάει με τους λιγότερο ανθεκτικούς  $L$ -εκτιμητές για να καταλήξει στους υψηλά ανθεκτικούς και δημοφιλή  $MM$ -εκτιμητές.

#### 3.3.3.1. $L$ -εκτιμητές

Κάθε εκτιμητής που μπορεί να υπολογιστεί από ένα γραμμικό συνδυασμό στατιστικών στοιχείων ανήκει στην κατηγορία των  **$L$ -εκτιμητών** (*L-estimators*). Ένα πρώτο βήμα προς την εκτίμηση ενός περισσότερου ανθεκτικού εκτιμητή έγινε από τον *Edgeworth* (1887), ο οποίος βελτίωσε μία πρόταση του *Boscovich* (1757) και παρουσίασε μία μέθοδο για την γραμμική παλινδρόμηση χρησιμοποιώντας τον εκτιμητή των ελαχίστων απόλυτων τιμών των καταλοίπων (*least absolute values*) αντί για τα ελάχιστα τετράγωνα αυτών, ο οποίος προσδιορίζεται από την παρακάτω μορφή:

$$\sum_{i=1}^n |\hat{\varepsilon}_i|. \quad (53)$$

(Αυτή η τεχνική αναφέρεται συνήθως ως  $L_1$  παλινδρόμηση ( $L_1$ -regression), ενώ τα ελάχιστα τετράγωνα ως  $L_2$ ). Ο  $L_1$ -εκτιμητής παλινδρόμησης μπορεί να γενικευθεί σαν  $\alpha$ -παλινδρόμηση ποσοστιαίου σημείου ( $\alpha$ -regression quantile) και η συνάρτηση που πρέπει να ελαχιστοποιηθεί έχει την εξής μορφή:

$$\sum_{i=1}^n \rho_{\alpha}(\hat{\varepsilon}_i), \quad (54)$$

όπου  $\alpha$  είναι ένα ποσοστιαίο σημείο, τότε

$$\rho_{\alpha}(\hat{\varepsilon}_i) = \begin{cases} \alpha \hat{\varepsilon}_i, & \hat{\varepsilon}_i \geq 0 \\ (\alpha - 1) \hat{\varepsilon}_i, & \hat{\varepsilon}_i < 0. \end{cases} \quad (55)$$

Αποδεικνύεται ότι ο  $L_1$ -εκτιμητής παλινδρόμησης μπορεί να αντιμετωπίσει τις  $y$ -έκτροπες παρατηρήσεις σε ένα δείγμα, αλλά εξακολουθεί να παρουσιάζει αδυναμίες με τα σημεία μόχλευσης, που έχουν μεγαλύτερη επίδραση στην προσαρμογή. Κατά συνέπεια, το σημείο κατάρρευσης της  $L_1$  παλινδρόμησης δεν είναι καλύτερο από 0%, αν και οι απόλυτες τιμές αυξάνονται λιγότερο γρήγορα από τα αντίστοιχα τετράγωνα των καταλοίπων. Επίσης, ο  $L_1$ -εκτιμητής έχει χαμηλή αποδοτικότητα. Ο συνδυασμός χαμηλού σημείου κατάρρευσης και χαμηλής αποδοτικότητας καθιστά την  $L_1$  παλινδρόμηση λιγότερο ελκυστική σε σχέση με άλλες μεθόδους ανθεκτικής στατιστικής.

Άλλοι γνωστοί  $L$ -εκτιμητές που προτάθηκαν από τον Rousseeuw (1984) είναι ο **εκτιμητής ελαχίστων διαμέσων τετραγώνων** (*least median of squares, LMS*) και ο **εκτιμητής ελαχίστων περικομμένων τετραγώνων** (*least trimmed of squares, LTS*).

Η *LMS* παλινδρόμηση αντικαθιστά το άθροισμα των τετραγώνων των καταλοίπων, που χαρακτηρίζει τη μέθοδο των ελαχίστων τετραγώνων, με την αποδεδειγμένα πιο ανθεκτική διάμεσο τους. Αυτή η εκτίμηση δίνεται από τη σχέση:

$$\min \text{med}(\hat{\varepsilon}_i^2). \quad (56)$$

Αυτή η πρόταση βασίζεται ουσιαστικά σε μια ιδέα του Hampel (1975) και διαπιστώνεται ότι ο εκτιμητής που προκύπτει είναι ανθεκτικός τόσο στις  $y$ - όσο και στις  $x$ -έκτροπες παρατηρήσεις με σημείο κατάρρευσης ίσο με 50%. Παρόλα αυτά η *LMS* έχει σημαντικές ελλείψεις που περιορίζουν τη χρήση της. Στην καλύτερη περίπτωση έχει μία σχετική αποτελεσματικότητα περίπου 37% [53] και αργό ρυθμό σύγκλισης του εκτιμητή. Παρόλα αυτά, οι *LMS* εκτιμητές παίζουν σημαντικό ρόλο στον υπολογισμό των πιο αποτελεσματικών *MM*-εκτιμητών, όπως θα δούμε παρακάτω, καθώς παρέχουν αρχικές εκτιμήσεις στα κατάλοιπα.

Ένας τρόπος βελτίωσης του αργού ρυθμού σύγκλισης του εκτιμητή *LMS* είναι η μέθοδος *LTS* και επιτυγχάνεται με τη χρήση διαφορετικής αντικειμενικής συνάρτησης (*objective function*). Αντί να ελαχιστοποιήσουμε το άθροισμα των τετραγώνων των καταλοίπων, ελαχιστοποιούμε το άθροισμα των περικομμένων τετραγώνων, αφού αρχικά έχουμε κατατάξει σε αύξουσα σειρά τα τετράγωνα των καταλοίπων και έχουμε περικόψει το άνω και κάτω  $\alpha$  ποσοστό αυτών. Η *LTS* δίνεται από τη σχέση:

$$\min \sum_{i=1}^h (\hat{\varepsilon}^2)_{i:n}, \quad (57)$$

όπου  $(\hat{\varepsilon}^2)_{1:n} \leq \dots \leq (\hat{\varepsilon}^2)_{n:n}$  τα διατεταγμένα τετράγωνα των καταλοίπων και  $h$  είναι ο αριθμός των παρατηρήσεων που απομένουν μετά την περικοπή του δείγματος κατά  $\alpha\%$  και συμπεριλαμβάνονται στον υπολογισμό του εκτιμητή. Το  $h$  δίνεται από τη σχέση:

$$h = \lceil n(1-\alpha) + 1 \rceil. \quad (58)$$

Οι καλύτερες ιδιότητες ανθεκτικότητας επιτυγχάνεται για  $h = \frac{n}{2} + 1$ , όπου σε αυτή την περίπτωση το σημείο κατάρρευσης φθάνει το 50%, όπως και με την *LMS*. Θα πρέπει να σημειωθεί ότι η μέθοδος *LTS* δεν κρίνεται αρκετά αποτελεσματική, πάρα την υψηλή της ανθεκτικότητα, όταν στο δείγμα υπάρχουν περισσότερες έκτροπες παρατηρήσεις σε σχέση με τις περικομμένες.

### 3.3.3.2. *R*-εκτιμητές

Οι ***R*-εκτιμητές** (*R-estimators*) προτάθηκαν στον τομέα της πολλαπλής παλινδρόμησης από τον *Jaeckel* (1972) και αφορούν μέτρα διασποράς που βασίζονται σε γραμμικούς συνδυασμούς των διατεταγμένων καταλοίπων ή αλλιώς σε τάξεις των καταλοίπων. Αν θεωρήσουμε ότι  $R_i$  είναι η τάξη των καταλοίπων  $\hat{\varepsilon}_i$  τότε ο σκοπός είναι να ελαχιστοποιηθεί η παρακάτω σχέση:

$$\min \sum_{i=1}^n \alpha_n(R_i) \hat{\varepsilon}_i, \quad (59)$$

όπου  $\alpha_n(i)$  είναι μία συνάρτηση επίδοσης (*score function*) που είναι μονότονη και ικανοποιεί τη σχέση:

$$\sum_{i=1}^n \alpha_n(i) = 0. \quad (60)$$

Έχουν προταθεί πολλές εκδοχές για τις συναρτήσεις επίδοσης, αλλά η πιο απλή, και ίσως η πιο συχνά χρησιμοποιημένη, είναι οι *Wilcoxon* συναρτήσεις επίδοσης, η οποία υπολογίζει απευθείας την τάξη των παρατηρήσεων από τη διάμεσο

$$\alpha_n(i) = i - \left( \frac{n+1}{2} \right). \quad (61)$$

Τα *median scores* είναι μία απλή προσαρμογή πάνω στις *Wilcoxon* συναρτήσεις επίδοσης,

$$\alpha_n(i) = \text{sgn} \left[ i - \left( \frac{n+1}{2} \right) \right]. \quad (62)$$

Τα *Van der Waerden scores* προσαρμόζουν τις τάξεις σύμφωνα με το αντίστροφο της κανονικής συνάρτησης πυκνότητας πιθανότητας  $\Phi^{-1}$ :

$$\alpha_n(i) = \Phi^{-1}\left(\frac{i}{n+1}\right). \quad (63)$$

Τέλος, τα *Bounded normal scores* προσαρμόζουν τα *Van der Waerden scores* περιορίζοντας τα σύμφωνα με μία σταθερά,  $c$ :

$$\alpha_n(i) = \min\left\{c, \max\left[\Phi^{-1}\left(\frac{i}{n+1}\right), -c\right]\right\}. \quad (64)$$

Ένα σημαντικό πλεονέκτημα των  $R$ -εκτιμητών συγκριτικά με τους  $M$ -εκτιμητές είναι ότι γίνονται αυτόματα ίσης μεταβολής ως προς την κλίμακα και έτσι δεν εξαρτώνται από την ταυτόχρονη ύπαρξη ενός εκτιμητή κλίμακας [54]. Παρόλα αυτά παρουσιάζουν αρκετά ανεπιθύμητα χαρακτηριστικά. Ένα παράδειγμα είναι ότι η βέλτιστη επιλογή της συνάρτησης επίδοσης είναι ασαφής. Ένα άλλο πρόβλημα είναι ότι η αντικειμενική συνάρτηση είναι αμετάβλητη σε σχέση με το σταθερό όρο. Επίσης, ακόμα πιο δυσεπίλυτο είναι το γεγονός ότι οι περισσότεροι  $R$ -εκτιμητές δεν μπορούν να αντιμετωπίσουν τις  $x$ -έκτροπες παρατηρήσεις με αποτέλεσμα το σημείο κατάρρευσης να είναι 0. Λόγω των προαναφερθέντων, η χρήση των  $R$ -εκτιμητών γενικά αποφεύγεται και προτιμώνται περισσότερο ανθεκτικοί εκτιμητές.

### 3.3.3.3. $M$ -εκτιμητές

Μια άλλη προσέγγιση της ανθεκτικής παλινδρόμησης είναι οι  **$M$ -εκτιμητές** ( *$M$ -estimators*). Ο *Huber* (1973) εισήγαγε τη χρήση των  $M$ -εκτιμητών, οι οποίοι αποτελούν γενίκευση των εκτιμητών μέγιστης πιθανοφάνειας (*maximum likelihood estimators*). Ένας  $M$ -εκτιμητής ορίζεται από την ελαχιστοποίηση της αντικειμενικής συνάρτησης:

$$\min \sum_{i=1}^n \rho(\hat{\epsilon}_i), \quad (65)$$

όπου η συνάρτηση  $\rho$  εκφράζει τη συνεισφορά του εκάστοτε καταλοίπου στην αντικειμενική συνάρτηση. Μία συνάρτηση  $\rho$  θα πρέπει να έχει τις ακόλουθες ιδιότητες:

- είναι πάντα θετική,  $\rho(\hat{\epsilon}_i) \geq 0$ ,
- ισούται με μηδέν όταν  $\hat{\epsilon}_i = 0$ ,  $\rho(0) = 0$ ,
- είναι συμμετρική, δηλαδή  $\rho(-u) = \rho(u)$  για κάθε  $u$ ,
- είναι μονότονη, δηλαδή  $\rho(u_i) \geq \rho(u'_i)$  για  $|u_i| \geq |u'_i|$ .

Ο εκτιμητής των ελαχίστων τετραγώνων είναι στην πραγματικότητα μία ειδική, λιγότερο ανθεκτική περίπτωση των  $M$ -εκτιμητών. Θα πρέπει όμως να γνωρίζουμε πως κατανομονται τα σφάλματα για να χρησιμοποιηθεί η κατάλληλη συνάρτηση  $\rho(u)$ , κάτι το οποίο είναι αρκετά ασαφές. Για αυτό το λόγο επιλέγονται συναρτήσεις ύστερα από προσεκτική εξέταση σχετικά με το πώς ο εκτιμητής που προκύπτει «ελαφραίνει» τα μεγάλα κατάλοιπα [60].



Για να βρεθεί ο  $M$ -εκτιμητής με την ελαχιστοποίηση της εξίσωσης (65) απαιτείται η μερική παραγωγή αυτής της εξίσωσης ως προς τις άγνωστες παραμέτρους  $\hat{b}$ , καταλήγοντας στο παρακάτω σύστημα εξισώσεων

$$\sum_{i=1}^n \psi(\hat{\varepsilon}_i) \vec{x}_i = \vec{0}, \quad (66)$$

όπου  $\psi$  είναι η παράγωγος συνάρτηση της  $\rho$  και ονομάζεται συνάρτηση επίδοσης και  $x_i$  είναι το διάνυσμα γραμμής των ανεξάρτητων μεταβλητών της  $i$ -οστής παρατήρησης:

$$\vec{x}_i = (x_{i1}, \dots, x_{ip}),$$

$$\vec{0} = (0, \dots, 0).$$

Για το λόγο αυτό η παραπάνω εξίσωση είναι στην πραγματικότητα ένα σύστημα  $p$  εξισώσεων, με λύση που δεν είναι πάντα εύκολο να βρεθεί. Στην πράξη, χρησιμοποιούνται επαναληπτικές μέθοδοι που βασίζονται στα επανασταθμισμένα ελάχιστα τετράγωνα (*iteratively reweighted least squares – IRLS*) ή στο λεγόμενο  $H$ -αλγόριθμο [33]. Εφόσον η λύση της εξίσωσης (66) δεν είναι ίσης μεταβολής σε μετασχηματισμούς ως προς τον άξονα  $y$ , θα πρέπει να τυποποιηθούν τα κατάλοιπα με κάποιο εκτιμητή της κλίμακας των καταλοίπων  $s$  δίνοντας την εξής σχέση:

$$\sum_{i=1}^n \psi\left(\frac{\hat{\varepsilon}_i}{s}\right) \vec{x}_i = \vec{0}, \quad (67)$$

όπου το  $s$  εκτιμάται ταυτόχρονα. [Αξίζει να σημειωθεί ότι, στην περίπτωση που υπάρχουν  $x$  – και  $y$  – έκτροπες παρατηρήσεις, θα πρέπει η συνάρτηση  $\psi$  να είναι επανακατερχόμενη (*redescending*) [44].

Ένας συχνά χρησιμοποιούμενος εκτιμητής της κλίμακας  $s$  είναι ο re-scaled εκτιμητής των απόλυτων αποκλίσεων (*median absolute deviation – MAD*), ο οποίος ορίζεται ως εξής:

$$s = \frac{MAD}{0.6745}, \quad (68)$$

όπου ο εκτιμητής  $MAD$ , γνωστός και ως εκτιμητής της διαμέσου, υπολογίζεται από την παρακάτω σχέση:

$$MAD = \text{median}|\hat{\varepsilon}_i|. \quad (69)$$

Αυτός ο εκτιμητής θεωρείται πιο ανθεκτικός στις έκτροπες παρατηρήσεις, με υψηλό σημείο κατάρρευσης, καθώς βασίζεται στη διάμεσο και όχι στον μέσο. Ο εκτιμητής  $s$  σμικρύνει τον εκτιμητή  $MAD$  κατά  $0.6745^{-1}$ , έτσι ώστε για μεγάλα δείγματα και για σφάλματα που ακολουθούν την κανονική κατανομή, ο  $s$  εκτιμά την τυπική απόκλιση.

Έπειτα ορίζεται η συνάρτηση βάρους ως εξής:

$$w(u) = \frac{\psi(u)}{u}. \quad (70)$$

Σύμφωνα με τους *Draper* και *Smith* (1998), ο αλγόριθμος των επανασταθμισμένων ελαχίστων τετραγώνων μπορεί να περιγραφεί συνοπτικά στα παρακάτω βήματα:

- Στην πρώτη επανάληψη ( $I=0$ ), εφαρμόζεται η μέθοδος των ελαχίστων τετραγώνων στα δεδομένα  $(x_i, y_i)$ , οπότε προκύπτουν οι εκτιμώμενες τιμές  $\hat{y}_i^{(0)}$ , τα κατάλοιπα  $\hat{\varepsilon}_i^{(0)}$  και οι συντελεστές παλινδρόμησης  $\hat{b}^{(0)}$ .
- Τα αρχικά  $\hat{\varepsilon}_i^{(0)}$  και  $\hat{b}^{(0)}$  χρησιμοποιούνται για τον υπολογισμό του εκτιμητή κλίμακας  $s^{(0)}$ .
- Επιλέγεται μία συνάρτηση βάρους  $w(u)$  και εφαρμόζεται στο λόγο  $\hat{\varepsilon}_i^{(0)}/s^{(0)}$  προκειμένου να υπολογιστούν τα αρχικά βάρη  $w_i^{(0)}$ . Έτσι προκύπτει η τιμή του  $W^{(0)}$  για  $\hat{b}^{(0)}$ .
- Για  $I=1$ , με χρήση του  $W^{(0)}$  μπορεί να υπολογιστεί η εκτίμηση  $\hat{b}^{(1)}$  από τη σχέση  $\hat{b}^{(1)} = (X^T W^{(0)} X)^{-1} X^T W^{(0)} Y$ .
- Με χρήση του  $\hat{b}^{(1)}$  μπορούν να βρεθούν νέα κατάλοιπα  $\hat{\varepsilon}_i^{(1)}$ , τα οποία με τη σειρά τους αποδίδουν το  $W^{(1)}$  μέσω του υπολογισμού του  $s^{(1)}$  και της εφαρμογής της συνάρτησης βάρους.
- Για  $I=2$ , βρίσκεται μια νέα εκτίμηση των  $\hat{b}$ , η  $\hat{b}^{(2)}$ , χρησιμοποιώντας το  $W^{(1)}$ . Ένας νέος πίνακας βαρών, ο  $W^{(2)}$ , μέσω των  $\hat{\varepsilon}_i^{(2)}$  και  $s^{(2)}$ .
- Ο τελικός  $M$ -εκτιμητής υπολογίζεται όταν οι εκτιμήσεις των  $\hat{b}$  συγκλίνουν.

Η σύγκλιση επιτυγχάνεται γρήγορα και η διαδικασία σταματάει όταν ο εκτιμητής αλλάζει κατά ένα ορισμένο ελάχιστο ποσοστό μεταξύ των επαναλήψεων ή όταν έχει εκτελεστεί ένας προκαθορισμένος αριθμός επαναλήψεων. Το κριτήριο που πρέπει να ισχύει στην περίπτωση της σύγκλισης είναι της μορφής

$$\frac{\|\hat{b}^{(q+1)} - \hat{b}^{(q)}\|}{\|\hat{b}^{(q+1)}\|} < \theta, \quad (71)$$

όπου  $\theta$  είναι μια μικρή θετική ποσότητα και συνήθως ισούται με 0.0001.

	Objective Function $\rho(u)$	Score Function $\psi(u)$	Weight Function $w(u) = \frac{\psi(u)}{u}$
(a) Least Squares	$\frac{1}{2}u^2 \quad -\infty \leq u \leq \infty$	$u$	1
(b) Huber (1973), $a > 0$	$\begin{cases} \frac{1}{2}u^2 & \text{if }  u  < a \\ a u  - \frac{1}{2}a^2 & \text{if }  u  \geq a \end{cases}$	$\begin{cases} u & \text{if }  u  < a \\ a \operatorname{sign} u & \text{if }  u  \geq a \end{cases}$	$\begin{cases} 1 & \text{if }  u  < a \\ \frac{a}{ u } & \text{if }  u  \geq a \end{cases}$
(c) Hampel, $a, b, c > 0$	$\begin{cases} \frac{1}{2}u^2 & \text{if }  u  < a \\ a u  - \frac{1}{2}a^2 & \text{if } a \leq  u  < b \\ a \frac{c u  - \frac{1}{2}u^2}{c-b} - \frac{7a^2}{6} & \text{if } b \leq  u  \leq c \\ a(b+c-a) & \text{otherwise} \end{cases}$	$\begin{cases} u & \text{if }  u  < a \\ a \operatorname{sign} u & \text{if } a \leq  u  < b \\ a \frac{c \operatorname{sign} u - u}{c-b} & \text{if } b \leq  u  \leq c \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 & \text{if }  u  < a \\ \frac{a}{ u } & \text{if } a \leq  u  < b \\ a \frac{c/ u  - 1}{c-b} & \text{if } b \leq  u  \leq c \\ 0 & \text{otherwise} \end{cases}$
(d) Tukey bisquare, $a > 0$	$\begin{cases} \frac{a^2}{6} \left(1 - \left(1 - \left(\frac{u}{a}\right)^2\right)^3\right) & \text{if }  u  \leq a \\ \frac{1}{6}a^2 & \text{if }  u  > a \end{cases}$	$\begin{cases} u \left(1 - \left(\frac{u}{a}\right)^2\right)^2 & \text{if }  u  \leq a \\ 0 & \text{if }  u  > a \end{cases}$	$\begin{cases} \left(1 - \left(\frac{u}{a}\right)^2\right)^2 & \text{if }  u  \leq a \\ 0 & \text{if }  u  > a \end{cases}$

**Πίνακας 3. 1.** Συνήθεις συναρτήσεις για τους M-εκτιμητές [60].

Υπό τις υποθέσεις *Gauss-Markov*, οι *M*-εκτιμητές είναι στατιστικά πιο αποτελεσματικοί από την  $L_1$  παλινδρόμηση, ενώ παράλληλα είναι ανθεκτικοί με τις  $y$ -έκτροπες παρατηρήσεις. Ωστόσο επηρεάζονται από τις  $x$ -έκτροπες παρατηρήσεις και δεν λαμβάνουν υπόψη τα σημεία μόχλευσης. Ένα σημείο κακής μόχλευσης μπορεί να προκαλέσει την κατάρρευση του εκτιμητή και για το λόγο αυτό οι *M*-εκτιμητές έχουν σημείο κατάρρευσης ίσο με 0.

Συμπερασματικά, οι *M*-εκτιμητές ήταν οι πρώτοι εκτιμητές που ήταν όσο το δυνατόν πιο ανθεκτικοί αναφορικά με τις  $y$ -έκτροπες παρατηρήσεις και ταυτόχρονα διατηρούσαν υψηλή σχετική αποτελεσματικότητα, αλλά το χαμηλό σημείο κατάρρευσης υποδηλώνει ότι σε μερικές περιπτώσεις δεν εκτελούν το ρόλο τους καλύτερα από τα ελάχιστα τετράγωνα [54]. Παρόλα αυτά όμως, συνεχίζουν να θεωρούνται σημαντικοί, όπως θα δούμε παρακάτω, στον υπολογισμό άλλων, πιο ανθεκτικών και αποτελεσματικών εκτιμητών.

### 3.3.3.4. Γενικευμένοι *M*-εκτιμητές

Εξαιτίας αυτής της ευπάθειας των σημείων μόχλευσης, εισήχθησαν οι **Γενικευμένοι *M*-εκτιμητές** (*generalized M-estimators or GM-estimators*) με βασικό στόχο τη βελτίωση της ανθεκτικότητας αναφορικά με τις  $x$ -έκτροπες παρατηρήσεις μέσω κάποιων συναρτήσεων βάρους. Η γενική σχέση των γενικευμένων *M*-εκτιμητών είναι η εξής:

$$\sum_{i=1}^n w(\bar{x}_i) \psi\left(\frac{\hat{\epsilon}_i}{v(\bar{x}_i)s}\right) \bar{x}_i = \bar{0}, \tag{72}$$

όπου  $\psi$  είναι μία συνάρτηση επίδοσης, όπως στην περίπτωση των *M*-εκτιμητών είναι τυπικά μία *Huber* ή *Tukey bisquare* (*biweight*) συνάρτηση, και  $w_i$  και  $v_i$  είναι συναρτήσεις βαρών που εξαρτώνται από τον πίνακα  $X$  των ανεξάρτητων μεταβλητών, οι οποίες παίρνουν νέες τιμές μέσω επαναληπτικής μεθόδου.

Ο πρώτος  $GM$ -εκτιμητής προτάθηκε από τον *Mallows* (1975) και περιλαμβάνει μόνο τα  $w_i$  βάρη, δηλαδή  $v_i(\bar{x}_i) = 1$ , επομένως η εξίσωση (72) γίνεται:

$$\sum_{i=1}^n w(\bar{x}_i) \psi\left(\frac{\hat{\varepsilon}_i}{s}\right) \bar{x}_i = \bar{0}. \quad (73)$$

Παρόλο που ο συγκεκριμένος εκτιμητής υπερέχει από τους άλλους όταν περιέχονται στις ανεξάρτητες μεταβλητές έκτροπες παρατηρήσεις, παρουσιάζει ένα σοβαρό μειονέκτημα. Συχνά, υπάρχουν έκτροπα σημεία τα οποία είναι καλά σημεία μόχλευσης και συμβάλλουν στην αποτελεσματικότητα του εκτιμητή. Επομένως, κάθε ελάφρυνση που εξαρτάται μόνο από τις τιμές των  $x_i$  και δεν εξετάζει κατά πόσο τα έκτροπα σημεία επηρεάζουν την προσαρμογή στην πληθώρα των δεδομένων, δεν είναι αρκετά αποτελεσματική [1].

Ο *Schwerpe* [28] επιχείρησε να πετύχει το στόχο ενός πιο ανθεκτικού εκτιμητή προτείνοντας μία λύση, η οποία ουσιαστικά προσαρμόζει τα βάρη των σημείων μόχλευσης ανάλογα με το μέγεθος των καταλοίπων  $\hat{\varepsilon}_i$ . Έτσι κατέληξε να διαιρέσει τα κατάλοιπα  $\hat{\varepsilon}_i$  με τα βάρη  $w_i$ , οπότε προέκυψε η παρακάτω σχέση:

$$\sum_{i=1}^n w(\bar{x}_i) \psi\left(\frac{\hat{\varepsilon}_i}{w(\bar{x}_i)s}\right) \bar{x}_i = \bar{0}. \quad (74)$$

Αποδεικνύεται όμως ότι το σημείο κατάρρευσης όλων των  $GM$ -εκτιμητών δεν μπορεί να είναι καλύτερο από την τιμή  $1/(p+1)$ , όπου  $p$  ο αριθμός των συντελεστών παλινδρόμησης που εκτιμώνται από το μοντέλο. Με άλλα λόγια, όσο αυξάνεται η διάσταση του μοντέλου, το σημείο κατάρρευσης πλησιάζει το 0. Αυτό το γεγονός είναι ιδιαίτερα προβληματικό, διότι συνεπάγεται ότι το σημείο κατάρρευσης μειώνεται με την αύξηση της διάστασης, όπου υπάρχουν περισσότερες πιθανότητες εμφάνισης έκτροπων παρατηρήσεων.

Σε μία προσπάθεια για να ξεπεραστούν τα προαναφερθέντα προβλήματα, προτάθηκαν διαδικασίες με τη χρήση  $GM$ -εκτιμητών, όπου αφαιρούνται τελείως αρκετά έκτροπα σημεία και έπειτα εφαρμόζεται η  $M$ -εκτίμηση στα εναπομείναντα «καλά» σημεία [16].

### 3.3.3.5. $S$ -εκτιμητές

Λόγω του χαμηλού σημείου κατάρρευσης των  $M$ -εκτιμητών, ο *Hampel* (1975) πρότεινε να ληφθεί υπόψη η κλίμακα των καταλοίπων. Ακολουθώντας αυτή την ιδέα, οι *Rousseeuw* και *Yohai* (1987) εισήγαγαν τους λεγόμενους  **$S$ -εκτιμητές** (*S-estimators*), καθώς βασίζονταν στους εκτιμητές κλίμακας. Αυτοί οι εκτιμητές έχουν ως λύση εκείνη που βρίσκει τη μικρότερη δυνατή διασπορά των καταλοίπων

$$\min s\left(\hat{\varepsilon}_1(\vec{b}), \dots, \hat{\varepsilon}_n(\vec{b})\right), \quad (75)$$

με τελική εκτίμηση για την παράμετρο κλίμακας

$$\hat{\sigma} = s\left(\hat{\varepsilon}_1\left(\hat{\vec{b}}\right), \dots, \hat{\varepsilon}_n\left(\hat{\vec{b}}\right)\right). \quad (76)$$

Η διασπορά  $s(\hat{\epsilon}_1(\vec{b}), \dots, \hat{\epsilon}_n(\vec{b}))$  ορίζεται ως η λύση της εξίσωσης:

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{\hat{\epsilon}_i}{s}\right) = K. \quad (77)$$

Το  $K$  είναι μία σταθερά που ισούται με  $E_\phi[\rho]$  και το  $\Phi$  αντιπροσωπεύει την τυπική κανονική κατανομή. Η συνάρτηση  $\rho$  πρέπει να ικανοποιεί τις ακόλουθες συνθήκες:

1. Η  $\rho$  είναι συμμετρική και συνεχώς παραγωγίσιμη, και  $\rho(0) = 0$ .
2. Υπάρχει  $c > 0$  τέτοιο ώστε η  $\rho$  αυστηρά να αυξάνεται στο διάστημα  $[0, c]$  και να παραμένει σταθερή στο  $[c, \infty]$ .
3.  $K/\rho(c) = 0.5$ .

Η δεύτερη συνθήκη σχετικά με την αντικειμενική συνάρτηση δηλώνει ότι η συσχετιζόμενη συνάρτηση επίδοσης θα είναι επανακατερχόμενη ενώ η τρίτη συνθήκη προϋποθέτει την εξασφάλιση του σημείου κατάρρευσης στο 50%, αλλά δεν είναι αυστηρά απαραίτητη.

Παραγωγίζοντας τη σχέση (77) προκύπτει:

$$\frac{1}{n} \sum_{i=1}^n \psi\left(\frac{\hat{\epsilon}_i}{s}\right) = 0, \quad (78)$$

όπου η συνάρτηση  $\psi$  αντικαθίσταται με μία κατάλληλη συνάρτηση βάρους. Όπως στις περισσότερες διαδικασίες  $M$ -εκτιμήσεων, έτσι και εδώ οι πιο συχνά χρησιμοποιούμενες συναρτήσεις βάρους είναι είτε η *Huber* είτε η *Tukey bisquare*.

Προκύπτει ότι οι  $S$ -εκτιμητές έχουν ουσιαστικά την ίδια ασυμπτωτική συμπεριφορά με τους  $M$ -εκτιμητές παλινδρόμησης και μπορούν να πετύχουν υψηλό σημείο κατάρρευσης έως και 50% με την κατάλληλη επιλογή των ρυθμιστικών σταθερών που υπεισέρχονται [54]. Όμως, το πρόβλημα που εντοπίζεται είναι ότι αδυνατούν να πετύχουν ταυτόχρονα και υψηλή αποτελεσματικότητα (περίπου 30%). Ενδεικτικά αναφέρεται ότι το πρόβλημα αυτό μπορεί να αντιμετωπιστεί αρκετά ικανοποιητικά με τους **GS-εκτιμητές** (*generalized S-estimates*) που προτάθηκαν από τους *Croux et al.* (1994) και αφορούν εκτιμητές που υπολογίζονται βρίσκοντας έναν  $GM$ -εκτιμητή της κλίμακας των καταλοίπων. Παρόλα αυτά, όπως θα αναφερθεί στη συνέχεια, τέτοιοι εκτιμητές με υψηλό σημείο κατάρρευσης αποφέρουν μία πολύ χρήσιμη πρώτη εκτίμηση μέσα σε πιο σύνθετες διαδικασίες ανθεκτικής παλινδρόμησης, καθώς οι εκτιμητές που προκύπτουν «υιοθετούν» το υψηλό σημείο κατάρρευσης τους [54].

### 3.3.3.6. $MM$ -εκτιμητές

Ο *Yohai* (1985) ήταν ο πρώτος που πρότεινε τους  **$MM$ -εκτιμητές** (*MM-estimators*), οι οποίοι γίνονται ολοένα και δημοφιλέστεροι δεδομένου ότι συνδυάζουν υψηλό σημείο κατάρρευσης (50%) και ταυτόχρονα υψηλή αποτελεσματικότητα. Η ονομασία τους οφείλεται στο γεγονός ότι χρησιμοποιείται παραπάνω από μία διαδικασία  $M$ -εκτίμησης για τον υπολογισμό των τελικών εκτιμητών.

Οι  $MM$ -εκτιμητές ορίζονται σε τρία στάδια:

1. Σε πρώτη φάση υπολογίζεται μία αρχική υψηλής κατάρρευσης εκτίμηση  $b^*$ , όπως με χρήση της  $LMS$ ,  $LTS$  ή της  $S$ -εκτίμησης, καθώς και τα αντίστοιχα κατάλοιπα  $\hat{\varepsilon}_i(b^*)$ . Αυτός ο αρχικός εκτιμητής παλινδρόμησης είναι ανθεκτικά συνεπής με ένα υψηλό σημείο κατάρρευσης, αλλά δεν είναι απαραίτητο να είναι αποτελεσματικός.
2. Τα κατάλοιπα  $\hat{\varepsilon}_i(b^*)$  της αρχικής εκτίμησης της πρώτης φάσης χρησιμοποιούνται για τον υπολογισμό της  $M$ -εκτίμησης της κλίμακας των καταλοίπων,  $s_n$ .
3. Στο τελευταίο βήμα, ο  $MM$ -εκτιμητής  $\hat{b}$  ορίζεται ως κάθε λύση της σχέσης

$$\sum_{i=1}^n \psi \left( \frac{\hat{\varepsilon}_i(\vec{b})}{s_n} \right) \vec{x}_i = \vec{0}, \quad (79)$$

η οποία ικανοποιεί την ανισότητα

$$s_n(\vec{b}) \leq s_n(\vec{b}^*), \quad (80)$$

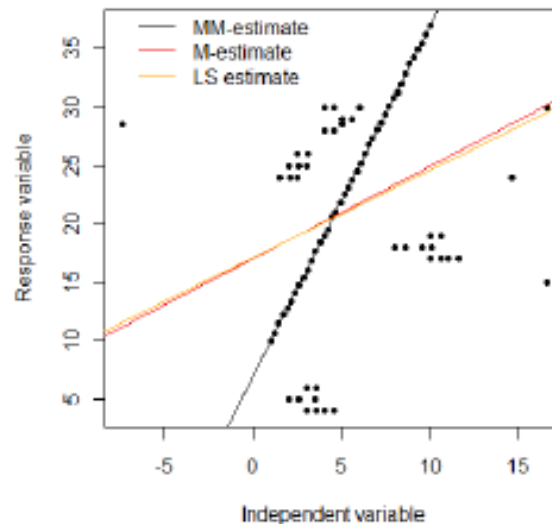
όπου

$$s_n(\vec{b}) = \sum_{i=1}^n \rho \left( \frac{\hat{\varepsilon}_i(\vec{b})}{s_n} \right). \quad (81)$$

Η συνάρτηση  $\rho$  θα πρέπει να πληροί τις συνθήκες που αναφέρθηκαν για τους  $S$ -εκτιμητές. Εφόσον το τρίτο στάδιο αποτελεί ουσιαστικά μία  $M$ -εκτίμηση με μια επιπλέον συνθήκη για τη λύση, η διαδικασία των  $IRLS$ , που περιγράφηκε σε προηγούμενη ενότητα, μπορεί να εφαρμοστεί για την πιθανή λύση της εξίσωσης (79), απλά κρατώντας σταθερό τον εκτιμητή κλίμακας σε μία τιμή  $s_n$  για κάθε επανάληψη [60].

Επιπλέον, οι  $MM$ -εκτιμητές είναι ίσης μεταβολής ως προς την παλινδρόμηση, εάν είναι και ο εκτιμητής που χρησιμοποιείται στην πρώτη φάση, αλλά και ως προς την κλίμακα [62]. Μία ακόμα σημαντική ιδιότητα που μπορεί να «υιοθετήσει» ο  $MM$ -εκτιμητής από τον εκτιμητή του πρώτου σταδίου είναι η ιδιότητα της ακριβούς προσαρμογής. Αυτή η ιδιότητα αποτελεί ένα διαφορετικό τρόπο προσέγγισης της ανθεκτικότητας ενός εκτιμητή στις έκτροπες παρατηρήσεις. Ένας εκτιμητής χαρακτηρίζεται από αυτή την ιδιότητα της ακριβούς προσαρμογής εάν σε ένα δείγμα  $n$  παρατηρήσεων  $(x_i, y_i)$ , οι  $n - \frac{n}{2} + 1$  από τις παρατηρήσεις

ικανοποιούν επακριβώς τη σχέση  $y = x_i b$ , θα πρέπει και η τελική εκτίμηση να είναι ακριβώς η ίδια με τα  $b$  ανεξαρτήτως των υπόλοιπων παρατηρήσεων. Τα ελάχιστα τετράγωνα και οι  $M$ -εκτιμητές δεν έχουν αυτή την ιδιότητα, σε αντίθεση με τους πιο ανθεκτικούς  $S$ -εκτιμητές και έτσι με τη σειρά τους και οι  $MM$ -εκτιμητές, καθώς τους χρησιμοποιούν στο πρώτο στάδιο υπολογισμού τους (**Σχήμα 3. 6**).



**Σχήμα 3. 6.** Το διάγραμμα αυτό περιέχει 79 σημεία εκ των οποίων τα 40 προκύπτουν από το μοντέλο  $y = 3x + 7 + \varepsilon_i$ . Παρόλο που τα υπόλοιπα 39 σημεία θεωρούνται ως έκτροπες παρατηρήσεις, ο MM-εκτιμητής έχει υπολογίσει σχεδόν απόλυτα τις πραγματικές τιμές των συντελεστών  $\hat{b}$ , δίνοντας το αποτέλεσμα  $\hat{b} = (7.151, 2.998)$  [60].

Τέλος, ο Yohai απέδειξε ότι οι MM-εκτιμητές:

- 1) έχουν σημείο κατάρρευσης 50% από το πρώτο στάδιο,
- 2) είναι υψηλά αποτελεσματικοί όταν τα σφάλματα κατανέμονται κανονικά και είναι ανεξάρτητες αυτές οι δύο ιδιότητες μεταξύ τους, καθώς το σημείο κατάρρευσης εξαρτάται από την επιλογή των ρυθμιστικών σταθερών στα δύο πρώτα στάδια ενώ
- 3) η ασυμπτωτική σχετική αποτελεσματικότητα καθορίζεται από τις ρυθμιστικές σταθερές στο τρίτο στάδιο.

Ο Πίνακας 3. 2 συνοψίζει δύο βασικές ιδιότητες της ανθεκτικότητας, το σημείο κατάρρευσης και την ασυμπτωτική αποτελεσματικότητα του εκτιμητή σε σχέση με τον εκτιμητή των ελαχίστων τετραγώνων, για τους εκτιμητές ανθεκτικότητας που παρουσιάστηκαν προηγουμένως. Διαπιστώνεται άμεσα ότι οι  $L_1$ - και  $M$ -εκτιμητές έχουν πολύ μικρό σημείο κατάρρευσης συγκριτικά με τους υπόλοιπους εκτιμητές και μερικές φορές μπορεί να μη συμπεριφέρονται καλύτερα από αυτούς των ελαχίστων τετραγώνων, ανάλογα με τη συσχέτιση των δεδομένων. Το ίδιο ισχύει και για τους  $R$ -εκτιμητές, που έχουν σημείο κατάρρευσης χαμηλότερο από 0.2. Αυτές οι μέθοδοι θα πρέπει γενικά να αποφεύγονται να χρησιμοποιούνται μόνες τους.

Επιπρόσθετα, θα πρέπει να δοθεί προσοχή στη χαμηλή ασυμπτωτική αποτελεσματικότητα των  $LMS$ ,  $LTS$  και  $S$ -εκτιμητών. Αν ο στόχος είναι μόνο η εξασφάλιση της ανθεκτικότητας και όχι η εξαγωγή συμπερασμάτων για τον πληθυσμό, τότε αυτοί οι εκτιμητές μπορεί να είναι κατάλληλοι. Από την άλλη πλευρά, δε θα πρέπει να χρησιμοποιούνται αν δεν είναι γνωστή η φύση των ασυνήθιστων παρατηρήσεων, επειδή η απερισκεπτη χρήση τους μπορεί να οδηγήσει σε λιγότερο αποτελεσματικούς εκτιμητές σε σύγκριση με αυτό που θα ήταν εφικτό σε διαφορετική περίπτωση.

Εκτιμητής	Σημείο Κατάρρευσης	Ασυμπτωτική Αποτελεσματικότητα
OLS	0	100
$L_1$ -εκτιμητής	0	64
LMS	0.5	37
LTS	0.5	8
R-εκτιμητής	<0.2	90
M-εκτιμητής (Huber, Biweight)	0	95
GM-εκτιμητής (Mallows, Schweppe)	$1/(p+1)$	95
S-εκτιμητής	0.5	33
MM-εκτιμητής	0.5	95

**Πίνακας 3. 2.** Βασικές ιδιότητες ανθεκτικότητας για διάφορους εκτιμητές ανθεκτικότητας.

Παρά το χαμηλό σημείο κατάρρευσης, η αποτελεσματικότητα του  $M$ -εκτιμητή είναι μία επιθυμητή ιδιότητα. Όταν εφαρμόζεται σε συνδυασμό με πιο ανθεκτικούς εκτιμητές, τότε οι νέοι εκτιμητές που προκύπτουν είναι ανθεκτικοί στις έκτροπες παρατηρήσεις και εξίσου υψηλά αποτελεσματικοί. Για παράδειγμα, μία  $M$ -εκτίμηση που χρησιμοποιεί τα κατάλοιπα που έχουν προκύψει από μία αρχική ανθεκτική προσαρμογή του  $LMS$  ή  $S$ -εκτιμητή, οδηγεί στον  $MM$ -εκτιμητή, που χαρακτηρίζεται από υψηλή ανθεκτικότητα στις  $x$ - και  $y$ -έκτροπες παρατηρήσεις και εξακολουθεί να διατηρεί την αποτελεσματικότητα στο 95% σε σχέση με τον εκτιμητή των ελαχίστων τετραγώνων.

Κατά συνέπεια, οι ανθεκτικοί εκτιμητές αποτελούν ένα ισχυρό εργαλείο στην ανίχνευση των έκτροπων παρατηρήσεων σε περίπλοκα δεδομένα, καθώς περιορίζουν την επιρροή τους και παράλληλα τα αποκαλύπτουν. Παρόλα αυτά, ο Anderson (2008) ενθαρρύνει μια προσεκτική προσέγγιση όταν ασχολείται κάποιος με εκτιμητές που προκύπτουν από μεθόδους ανθεκτικής παλινδρόμησης, γιατί μπορεί διάφοροι ανθεκτικοί εκτιμητές να δώσουν πολύ διαφορετικές εκτιμήσεις, εκτός και αν τα δεδομένα συμπεριφέρονται καλά. Επίσης, προτείνει ως βασικό στάδιο στην ανάλυση παλινδρόμησης τη σύγκριση των αποτελεσμάτων που έχουν προκύψει από ανθεκτικές μεθόδους και από αυτή των ελαχίστων τετραγώνων.



## ΚΕΦΑΛΑΙΟ 4 – ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

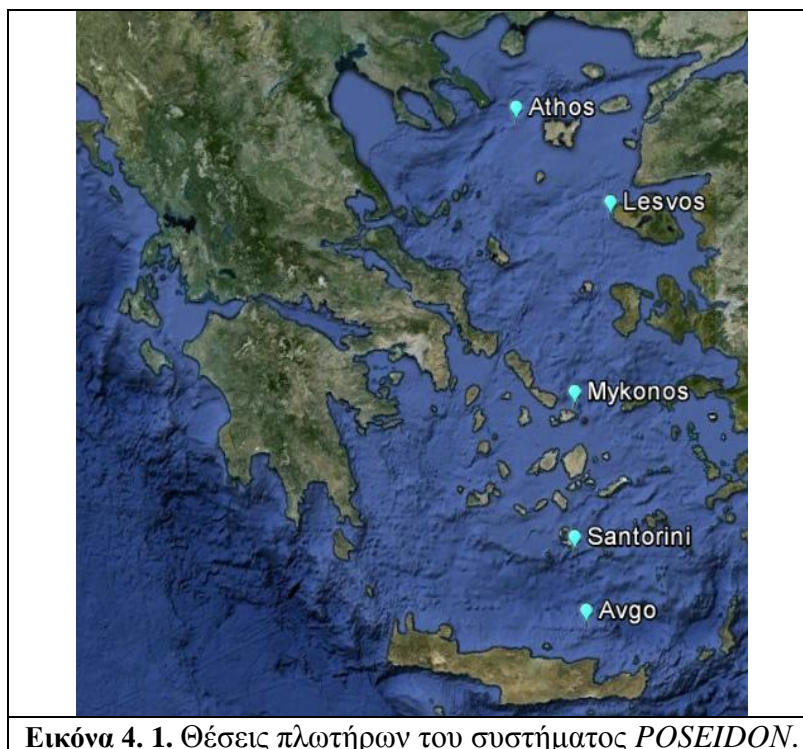
### 4.1. ΑΝΑΛΥΣΗ ΑΝΕΜΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

#### 4.1.1. Διαδικασία Ανάλυσης Ανεμολογικών Δεδομένων

Αναφορικά με τις επιτόπιες μετρήσεις, τα ανεμολογικά δεδομένα που χρησιμοποιήθηκαν για τη σύγκριση των μεθόδων παλινδρόμησης συνίστανται σε χρονοσειρές της ταχύτητας ανέμου καλύπτοντας μια περίοδο που κυμαίνεται από 6 έως 11 χρόνια (2000 έως 2011). Οι μετρήσεις ανέμου έχουν διάρκεια καταγραφής 600s με συχνότητα μέτρησης 1 Hz, διενεργούνται κάθε 3 ώρες ενώ το ανεμόμετρο της συσκευής είναι τοποθετημένο στα 3 m πάνω από την επιφάνεια της θάλασσας. Ειδικότερα, οι μετρήσεις έχουν αποκτηθεί από 5 πλωτούς μετρητικούς σταθμούς (*buoys*) του συστήματος *POSEIDON*, οι οποίοι είναι εγκατεστημένοι σε περιοχές βαθέων υδάτων (βλ. **Εικόνα 4. 1**) στο Αιγαίο Πέλαγος, και η ακριβή τους θέση μαζί με την περίοδο καταγραφής τους παρουσιάζονται σε μορφή πίνακα (**Πίνακας 4. 1**) στη συνέχεια.

Όνομασία	Γεωγραφικό πλάτος ( <i>degrees</i> )	Γεωγραφικό μήκος ( <i>degrees</i> )	Περίοδος καταγραφής
<i>Athos</i>	39.9688	24.7222	2000-2011
<i>Avgo</i>	35.6205	25.6392	2000-2006
<i>Lesvos</i>	39.1648	25.8115	2000-2011
<i>Mykonos</i>	37.5156	25.4608	2000-2011
<i>Santorini</i>	36.2591	25.4964	2000-2011

**Πίνακας 4. 1.** Θέσεις και περίοδος καταγραφής των πλωτήρων του συστήματος *POSEIDON*.



**Εικόνα 4. 1.** Θέσεις πλωτήρων του συστήματος *POSEIDON*.

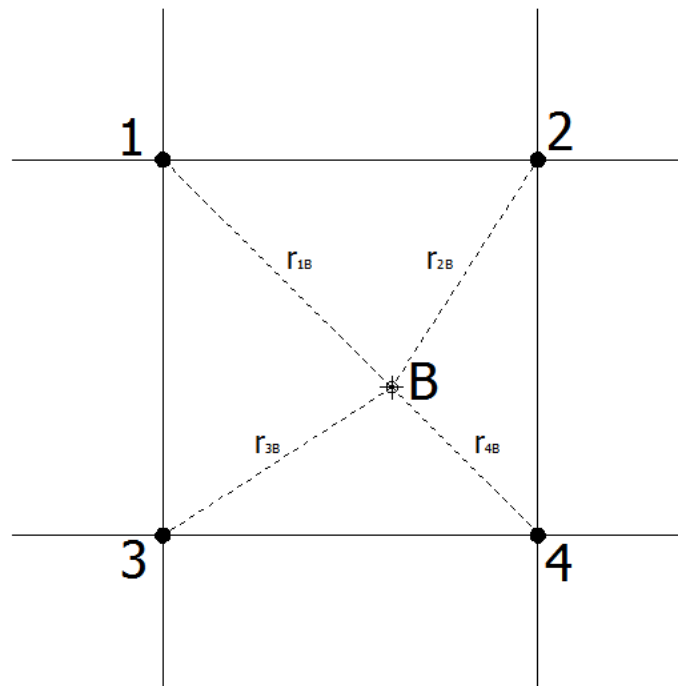
Το ατμοσφαιρικό μοντέλο που χρησιμοποιήθηκε για τη δημιουργία των ανεμολογικών δεδομένων βασίστηκε σε ένα βελτιωμένο μη-υδροστατικό μοντέλο *SKIRON-Eta* με χωρική ανάλυση  $0.1^\circ \times 0.1^\circ$  (περίπου  $9 \times 11$  km) και χρονικό βήμα αριθμητικής ολοκλήρωσης 36 sec [8]. Τα ανεμολογικά αποτελέσματα του μοντέλου που χρησιμοποιήθηκαν για την εκτίμηση του υπεράκτιου αιολικού δυναμικού συνίστανται σε χρονοσειρές 10 ετών (1994 έως 2004) της ταχύτητας ανέμου. Τα αποτελέσματα του μοντέλου παράγονται επίσης ανά 3 ώρες.

Πριν την ανάλυση παλινδρόμησης, είναι απαραίτητο να διορθωθούν τα ανεμολογικά δεδομένα από τους πλωτούς μετρητικούς σταθμούς και το ατμοσφαιρικό μοντέλο και να συνταυτιστούν μεταξύ τους χωρικά και χρονικά, προκειμένου να επιτευχθεί μία ορθή σύγκριση. Τα δεδομένα των επιτόπιων μετρήσεων που λήφθηκαν, ελέγχθηκαν ποιοτικά και απαλείφθηκαν αρνητικές, μηδενικές και κενές τιμές που μπορεί να υπήρχαν για διάφορους λόγους, όπως δυσλειτουργία του οργάνου και του συστήματος επικοινωνίας, βλάβη κατά την αποθήκευση ή επεξεργασία των δεδομένων, απώλεια ισχύος μπαταρίας, βανδαλισμός κ.α. Στις εγγραφές που τελικά απέμειναν, μετά από την εφαρμογή του ελέγχου ποιότητας, έγινε η μετατροπή της ταχύτητας του ανέμου από τα 3 στα 10 m, καθώς η ταχύτητα του ανέμου που λαμβάνεται από το μοντέλο αναφέρεται στα 10 m από την επιφάνεια της θάλασσας. Η σχέση που μετατρέπει την ταχύτητα ανέμου  $u_z$  από ένα επίπεδο  $z$ , στο επίπεδο των 10 m από την επιφάνεια της θάλασσας είναι η εξής:

$$\frac{u_z}{u_{10}} = \frac{\ln(z/0.0016)}{\ln(10/0.0016)}, \quad (82)$$

όπου  $u_{10}$  είναι η ταχύτητα του ανέμου στα 10 m από την επιφάνεια της θάλασσας. Για το αναλυτικό φυσικό και μαθηματικό υπόβαθρο της παραπάνω σχέσης βλ. [59]. Με αυτό τον τρόπο καθίσταται δυνατή η ομογενοποίηση των ταχυτήτων του ανέμου από δύο πηγές ανεμολογικών δεδομένων με διαφορετικά επίπεδα αναφοράς. Προκειμένου όμως να γίνει εφικτή η σύγκριση των ταχυτήτων του ανέμου από αυτές τις δύο πηγές θα πρέπει επίσης να συνταυτιστούν μεταξύ τους στο χώρο και το χρόνο, όπως αναφέρθηκε και προηγουμένως. Αναφορικά με τη χωρική συνταύτιση, θα πρέπει οι μετρήσεις των δεδομένων του μοντέλου, που είναι σε μορφή πλέγματος, να μεταφερθούν στις αντίστοιχες θέσεις των πλωτήρων. Εφόσον η θέση κάθε πλωτήρα περικλείεται από 4 κομβικά σημεία του πλέγματος του μοντέλου (βλ. **Σχήμα 4. 1**), μεταφέρονται οι χρονοσειρές και των τεσσάρων αυτών σημείων με βάρος αντιστρόφως ανάλογο του τετραγώνου της απόστασής τους από τον πλωτήρα. Η σχέση που χρησιμοποιείται για να προκύψει η τελική χρονοσειρά της ταχύτητας του ανέμου από το μοντέλο για κάθε μία θέση των πλωτήρων του συστήματος *POSEIDON* είναι η εξής:

$$u_{M_B} = \frac{\frac{1}{r_{1B}^2} u_1 + \frac{1}{r_{2B}^2} u_2 + \frac{1}{r_{3B}^2} u_3 + \frac{1}{r_{4B}^2} u_4}{\frac{1}{r_{1B}^2} + \frac{1}{r_{2B}^2} + \frac{1}{r_{3B}^2} + \frac{1}{r_{4B}^2}}. \quad (83)$$



Σχήμα 4. 1. Σχηματική αναπαράσταση για τη χωρική συνταύτιση.

Στη συνέχεια ακολουθεί η συνταύτιση των τελικών διορθωμένων δεδομένων των δύο ανεμολογικών πηγών στο χρόνο με κατάλληλο προγραμματισμό στο λογισμικό πακέτο *Matlab*, όπου ουσιαστικά συνταυτίζονται οι παρατηρήσεις τους έχοντας ως κοινό παράγοντα την ακριβή ημερομηνία και ώρα καταγραφής τους. Τελικά, προκύπτει ότι το κοινό διάστημα καταγραφής των επιτόπιων μετρήσεων και του ατμοσφαιρικού μοντέλου είναι μία περίοδος 5 ετών (2000 έως 2004).

Στον Πίνακα 4. 2 που ακολουθεί, παρουσιάζονται τα βασικά στατιστικά στοιχεία (μέγεθος δείγματος, μέση τιμή, τυπική απόκλιση, ελάχιστη τιμή, μέγιστη τιμή, συντελεστής μεταβλητότητας, συντελεστής προσδιορισμού, κύρτωση και ασυμμετρία) των δεδομένων της ταχύτητας του ανέμου, που αναλύθηκαν και συσχετίστηκαν, για τη χρονική περίοδο 2000 – 2004.

Ονομασία σταθμού	Πηγή δεδομένων	N	m	SD	min	max	CV	R	Ku	Sk
Athos	BUOY	9340	5.207	4.047	0.003	25.472	77.720	0.822	1.100	1.084
	MODEL		4.810	3.155	0.003	20.733	65.589		1.543	1.152
Avgo	BUOY	11277	6.400	3.847	0.000	22.908	60.113	0.706	0.175	0.533
	MODEL		5.814	2.882	0.075	20.531	49.569		-0.306	0.171
Lesvos	BUOY	11453	7.123	4.065	0.003	31.031	57.067	0.780	0.536	0.602
	MODEL		5.740	3.099	0.114	23.240	53.982		0.372	0.574
Mykonos	BUOY	8792	7.962	4.294	0.003	21.480	53.936	0.847	-0.751	0.071
	MODEL		6.517	3.244	0.166	17.945	49.778		-0.584	0.348
Santorini	BUOY	11442	6.647	3.734	0.000	21.342	56.170	0.759	0.000	0.483
	MODEL		4.840	2.450	0.098	18.959	50.627		1.033	0.866

Πίνακας 4. 2. Βασικά στατιστικά στοιχεία των συνταυτισμένων ανεμολογικών δεδομένων για τη χρονική περίοδο 2000 – 2004.

Από τον Πίνακα 4. 2 προκύπτει ότι η μεγαλύτερη μέση τιμή, τυπική απόκλιση και συντελεστής προσδιορισμού αντιστοιχούν στο μετρητικό σταθμό που βρίσκεται στη θέση *Mykonos*, ενώ η μεγαλύτερη καταγραφή ταχύτητας και για τις δύο πηγές ανεμολογικών

δεδομένων αντιστοιχούν στη θέση *Lesvos*. Οι τιμές του συντελεστή *CV* είναι μέγιστες στη θέση *Athos* τόσο για τα δεδομένα του μετρητικού σταθμού όσο και για αυτά του μοντέλου. Επίσης, τη μεγαλύτερη κύρτωση και ασυμμετρία εμφανίζουν και πάλι τα δεδομένα του σταθμού στη θέση *Athos*. Τέλος, παρατηρείται ότι οι τιμές της ταχύτητας του ανέμου του μοντέλου είναι συστηματικά μικρότερες από εκείνες των σταθμών. Επομένως, συμπεραίνεται ότι το μοντέλο έχει την τάση να υποεκτιμά την ταχύτητα του ανέμου.

#### 4.1.2. Μέθοδοι Παλινδρόμησης

Σε αυτό το εδάφιο, θα αναφερθούμε στις μεθόδους που χρησιμοποιήθηκαν στα πλαίσια της συγκριτικής μελέτης με σκοπό τη διόρθωση των δεδομένων από τα αριθμητικά μοντέλα καιρού με τη βοήθεια των ανεμολογικών μετρήσεων που συλλέχθηκαν και επεξεργάστηκαν κατάλληλα σε πέντε επιλεγμένες θαλάσσιες θέσεις στην περιοχή του Αιγαίου Πελάγους (βλ. Υποκεφ. 4.1.1). Θα πρέπει να σημειωθεί ότι, στη σχετική βιβλιογραφία, οι μέθοδοι που χρησιμοποιούνται για την πρόβλεψη τιμών χαρακτηριστικών μεγεθών από αριθμητικά μοντέλα βασίζονται επίσης στο ίδιο στατιστικό υπόβαθρο της γραμμικής παλινδρόμησης (απλής ή πολλαπλής). Οι πολύπλοκες ατμοσφαιρικές αβεβαιότητες και η περιορισμένη γνώση της μαθηματικής διατύπωσης της ατμοσφαιρικής φυσικής και δυναμικής καθιστούν απαραίτητες τις συγκεκριμένες στατιστικές τεχνικές για τη βελτίωση των αποτελεσμάτων των αριθμητικών μοντέλων βάσει των οποίων καθορίζεται μια στατιστική σχέση μεταξύ εκτιμώμενης ποσότητας και μεταβλητών (βλ. [24, 25, 35]).

Με τη βοήθεια του λογισμικού πακέτου *Matlab*, υλοποιήθηκαν οι κώδικες για την περιγραφή των διαφόρων μεθόδων παλινδρόμησης και για τη μετέπειτα σύγκρισή τους. Πιο συγκεκριμένα, οι μέθοδοι που χρησιμοποιήθηκαν είναι:

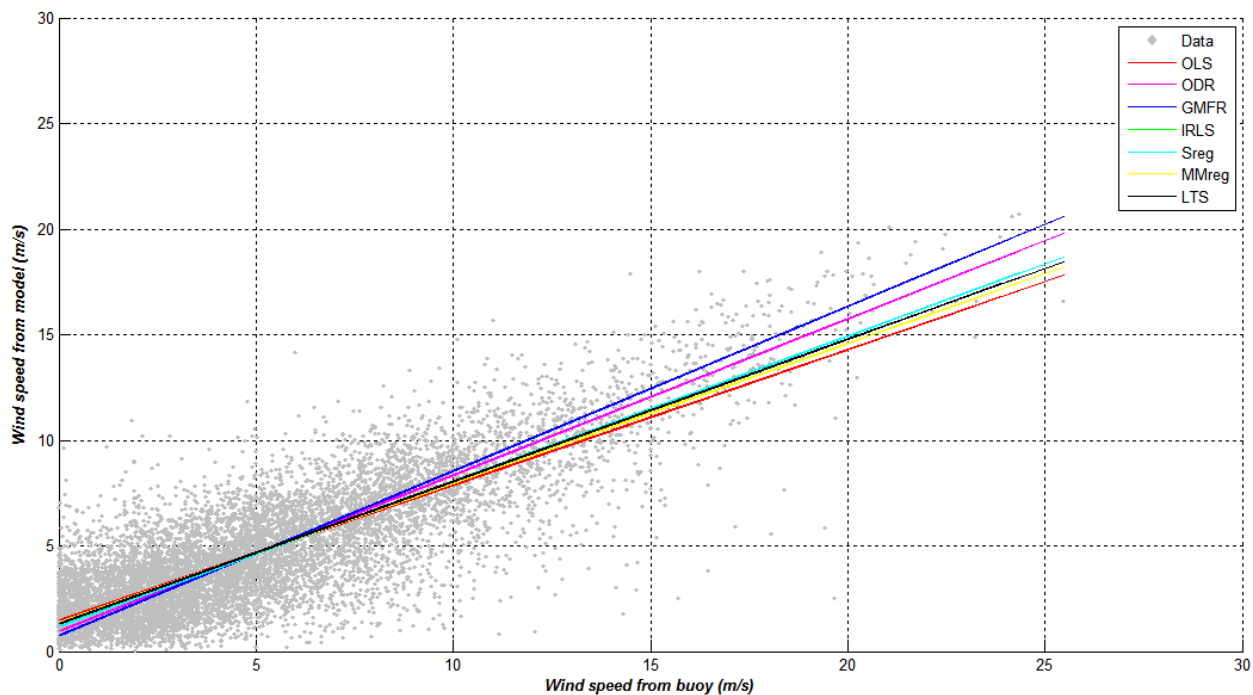
- **Κλασσικές μέθοδοι:**
  1. Μέθοδος Ελαχίστων τετραγώνων (*Ordinary Least Squares – OLS*),
  2. Παλινδρόμηση Ορθογώνιας Απόστασης (*Orthogonal Distance Regression – ODR*),
  3. Συναρτησιακή Σχέση Γεωμετρικού Μέσου (*Geometric Mean Functional Relationship – GMFR*),
- **Ανθεκτικές μέθοδοι:**
  1. Εκτιμητής Ελαχίστων Περικομμένων Τετραγώνων (*Least Trimmed of Squares – LTS*),
  2. S-εκτιμητές,
  3. MM-εκτιμητές,
  4. Επανασταθμισμένα Ελάχιστα Τετράγωνα (*Iteratively Reweighted Least Squares – IRLS*).

Οι μέθοδοι παλινδρόμησης ήταν είτε έτοιμες συναρτήσεις στη *Matlab* (*OLS*, *IRLS*) είτε υπήρχαν σε άλλα στατιστικά πακέτα (*LIBRA*, *FSDA*). Στη δεύτερη περίπτωση έγιναν ορισμένες τροποποιήσεις στον κώδικα για να προσαρμοστούν τα δεδομένα εισόδου κάθε μεθόδου στις ανάγκες της παρούσας συγκριτικής μελέτης. Εκτός από τις παραμέτρους της παλινδρόμησης  $\hat{b}_0$  και  $\hat{b}_1$  που υπολογίζονται για κάθε μέθοδο, για τις ανθεκτικές μεθόδους υπολογίζονται επιπρόσθετα τα βάρη, τα κατάλοιπα, ο εκτιμητής κλίμακας και εμφανίζονται σε μορφή πίνακα ο αύξων αριθμός των έκτροπων παρατηρήσεων (πλην της μεθόδου *IRLS*).

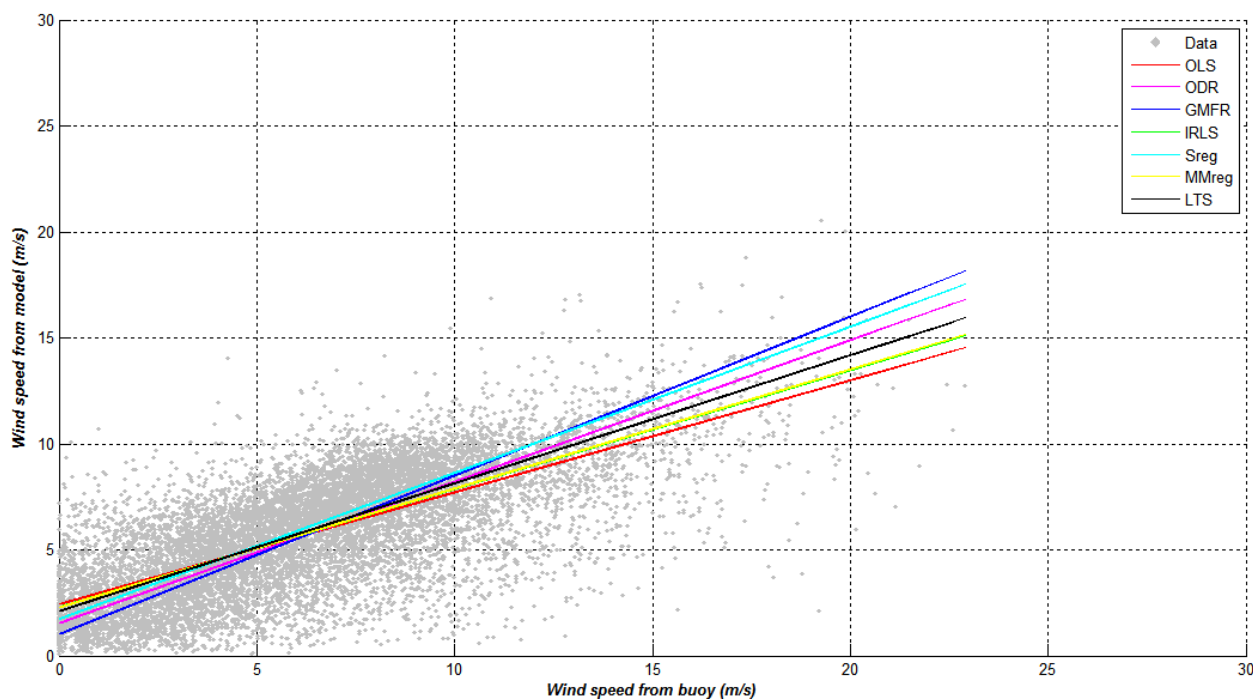
Η σύγκριση των αποτελεσμάτων που ακολουθεί στο επόμενο εδάφιο χωρίζεται σε δύο κατηγορίες.

Η πρώτη, αφορά τον έλεγχο της **αποδοτικότητας της εξίσωσης παλινδρόμησης**. Ο έλεγχος αυτός με τη σειρά του χωρίζεται σε δύο επιμέρους ομάδες. Η πρώτη περιλαμβάνει τη σύγκριση όλων των μεθόδων παλινδρόμησης για κάθε θέση μελέτης έχοντας τα τελικά συνταυτισμένα ανεμολογικά δεδομένα από τους μετρητικούς σταθμούς και το ατμοσφαιρικό μοντέλο. Στη δεύτερη κατηγορία, αφού έχουν προκύψει οι κοινές έκτροπες παρατηρήσεις των μεθόδων S, MM και LTS για κάθε θέση, αφαιρούνται οι αντίστοιχες παρατηρήσεις και από τις τρεις κλασσικές μεθόδους παλινδρόμησης και επανεκτελούνται για να υπολογιστούν οι νέες παράμετροι  $\hat{b}_0$  και  $\hat{b}_1$ . Στη δεύτερη κατηγορία αποτελεσμάτων ελέγχεται η **αποδοτικότητα της εξίσωσης διόρθωσης** με δύο τρόπους. Στον πρώτο τρόπο, η εξίσωση διόρθωσης περιλαμβάνει όλο το δειγματικό πλήθος των συνταυτισμένων ανεμολογικών δεδομένων κάθε θέσης. Στο δεύτερο τρόπο, εφαρμόζεται η εξίσωση διόρθωσης στα πρώτο μισό του δείγματος πλήθους και κατόπιν οι παράμετροι αυτής της σχέσης χρησιμοποιούνται στο υπόλοιπο μισό του θεωρώντας τις ως «νέες» τιμές ανεμολογικών δεδομένων. Ουσιαστικά, πρόκειται για έναν πιο αυστηρό και άμεσα εφαρμόσιμο έλεγχο μιας και ο επιθυμητός τρόπος για την εκτίμηση της σχέσης διόρθωσης είναι μέσω του ελέγχου αυτής της σχέσης σε νέα συνταυτισμένα ανεμολογικά δεδομένα που έχουν προκύψει από μετρητικούς σταθμούς και το ατμοσφαιρικό μοντέλο.

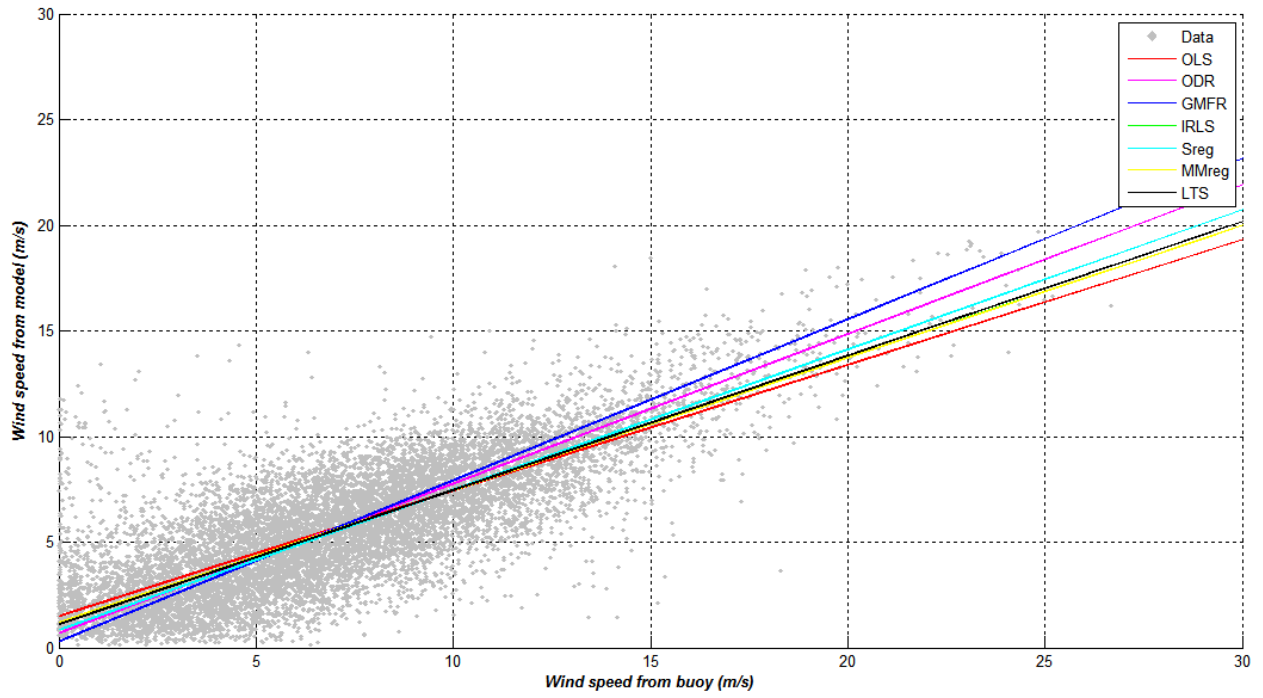
Στα **Σχήμα 4. 2 – Σχήμα 4. 6** παρουσιάζονται, σε μορφή διαγραμμάτων σκεδασμού, οι ευθείες παλινδρόμησης όλων των μεθόδων σε όλες τις θέσεις που εξετάζονται όπως αυτές υπολογίζονται με τη βοήθεια της *Matlab*. Συμπεραίνεται ότι οι γραμμικές σχέσεις μεταξύ της ταχύτητας του ανέμου από τους μετρητικούς σταθμούς ( $u_B$ ) και το μοντέλο ( $u_M$ ) όπως προκύπτουν από τις ανθεκτικές μεθόδους δίνουν παραπλήσιες γραμμές, ιδιαίτερα στις θέσεις *Mykonos* και *Santorini*. Επίσης, η μέθοδος *GMFR* δίνει πάντοτε τη μεγαλύτερη κλίση της ευθείας ( $\hat{b}_1$ ) και τη χαμηλότερη τιμή του σταθερού όρου ( $\hat{b}_0$ ) σε όλα τα σημεία μελέτης. Στα **Σχήμα 4. 7 – Σχήμα 4. 11** παρουσιάζονται οι ίδιες μέθοδοι στις ίδιες θέσεις μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους. Προκύπτει ότι η μέθοδος *OLS* παρουσιάζει τη μεγαλύτερη διαφορά συγκριτικά με τις μεθόδους *GMFR* και *ODR* καθώς προσεγγίζει συστηματικά τις ανθεκτικές μεθόδους. Τέλος, η μέθοδος *GMFR* εξακολουθεί να παρουσιάζει την ίδια συμπεριφορά όπως και προηγουμένως.



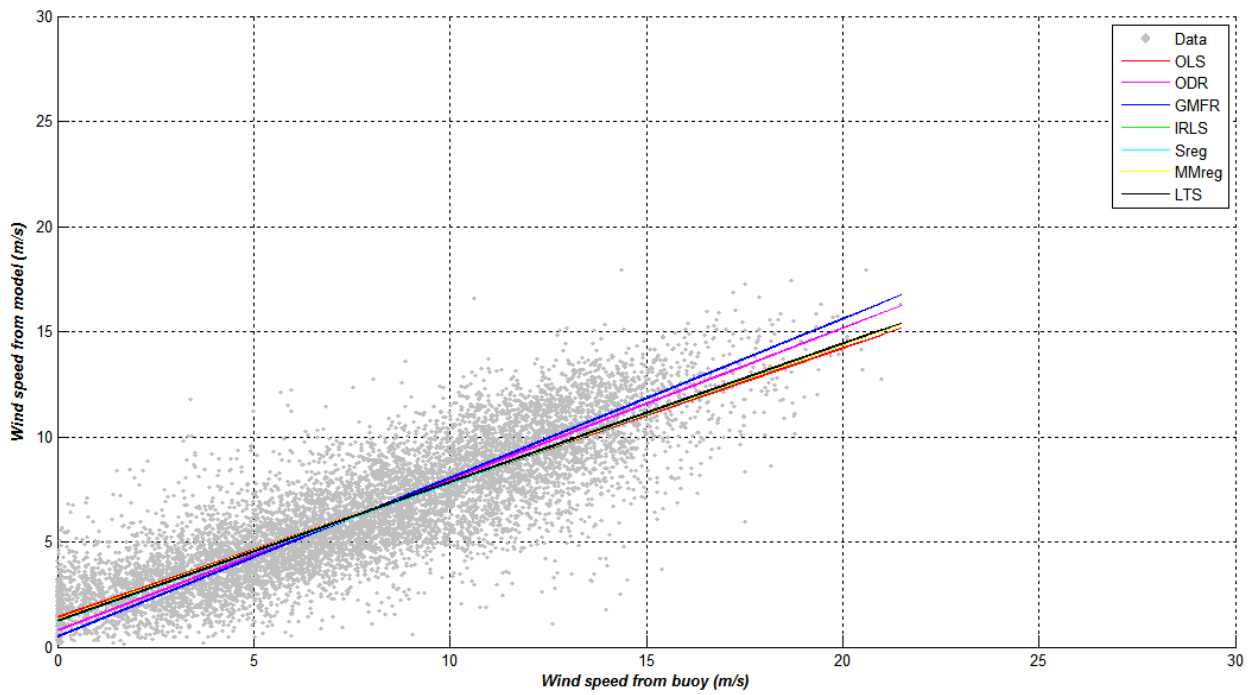
Σχήμα 4. 2. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  για διάφορες μεθόδους παλινδρόμησης στη θέση *Athos*.



Σχήμα 4. 3. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  για διάφορες μεθόδους παλινδρόμησης στη θέση *Avgo*.

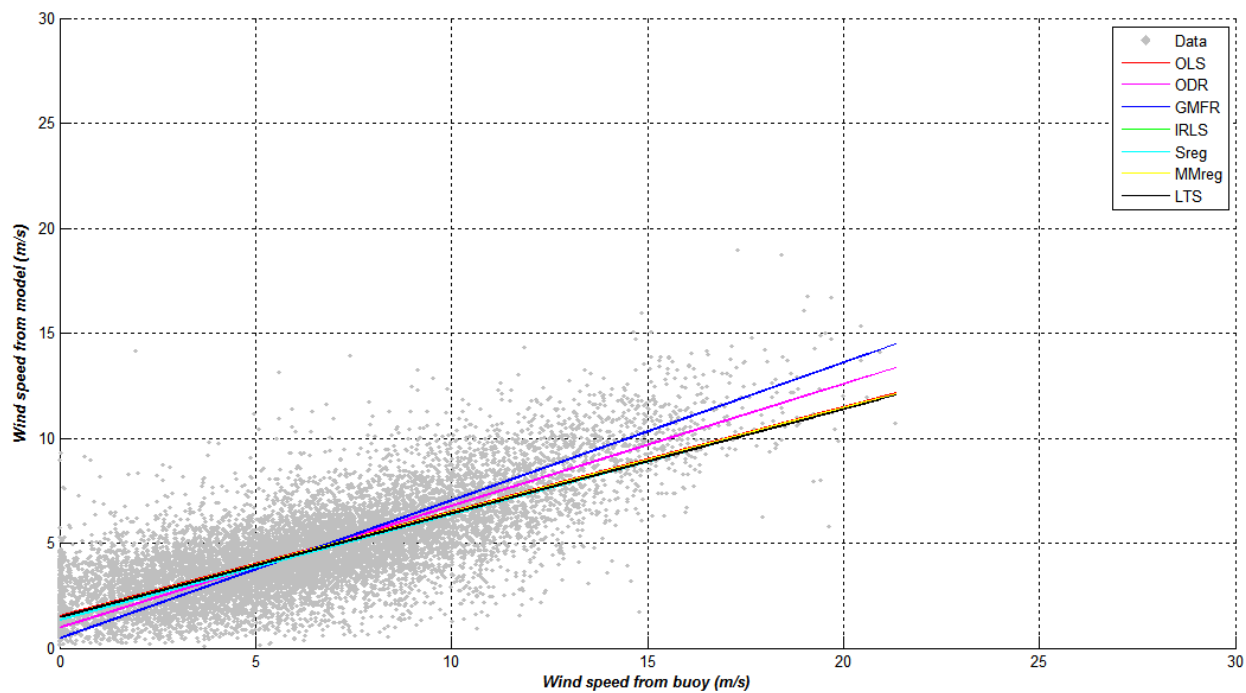


Σχήμα 4. 4. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  για διάφορες μεθόδους παλινδρόμησης στη θέση *Lesvos*.

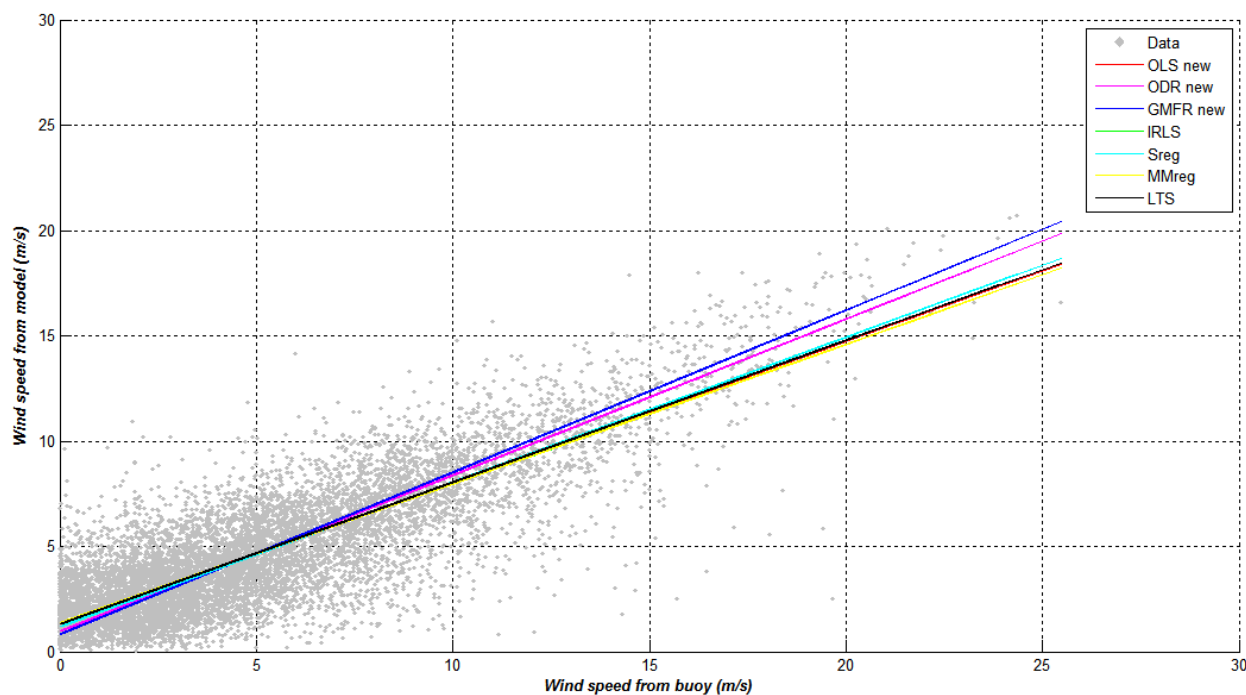


Σχήμα 4. 5. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  για διάφορες μεθόδους παλινδρόμησης στη θέση *Mykonos*.



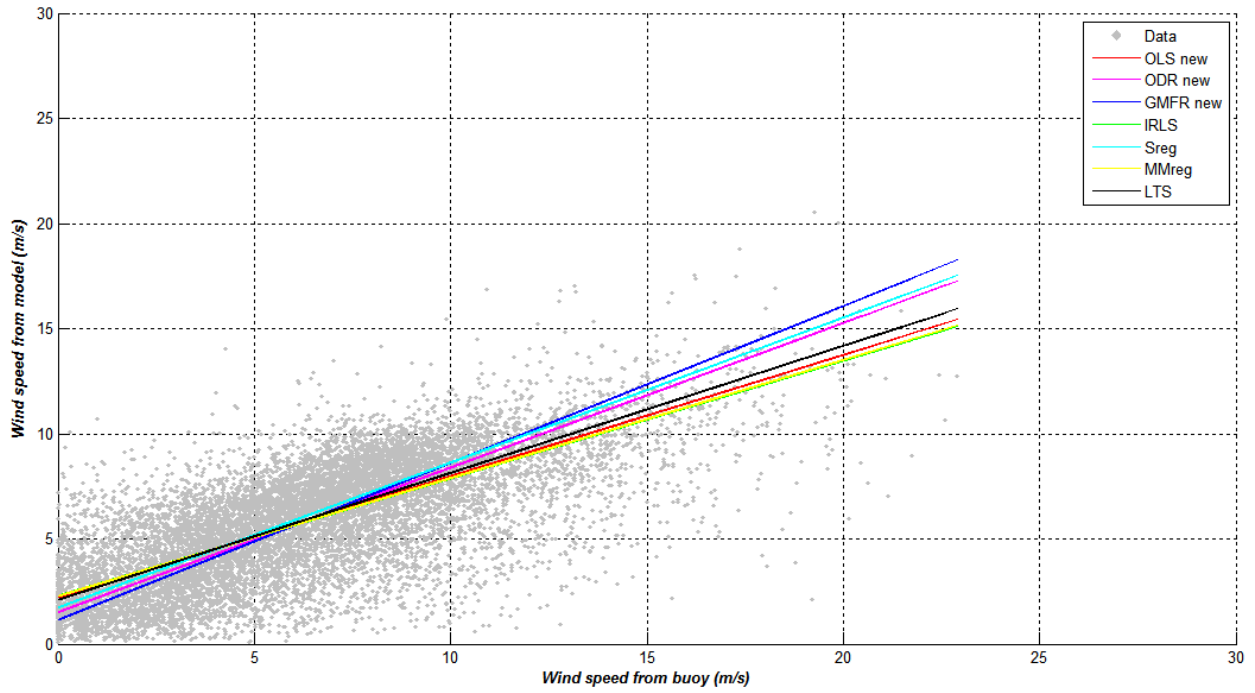


Σχήμα 4. 6. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  για διάφορες μεθόδους παλινδρόμησης στη θέση *Santorini*.

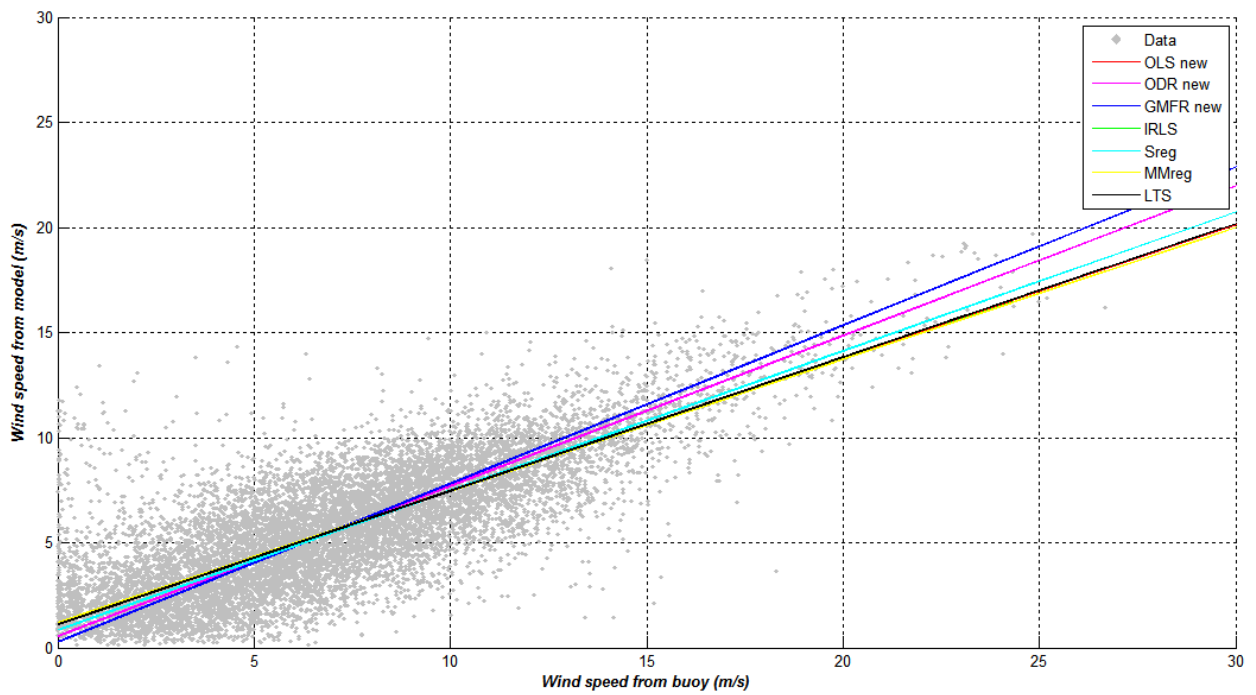


Σχήμα 4. 7. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση *Athos*.

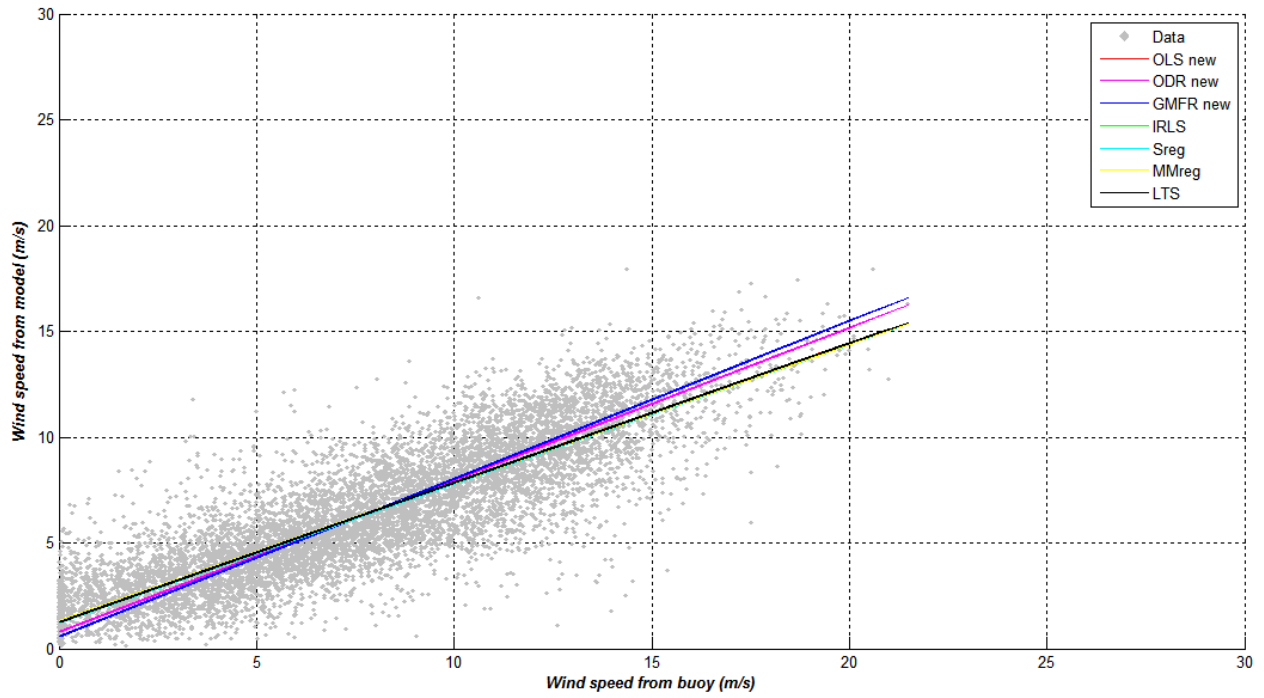




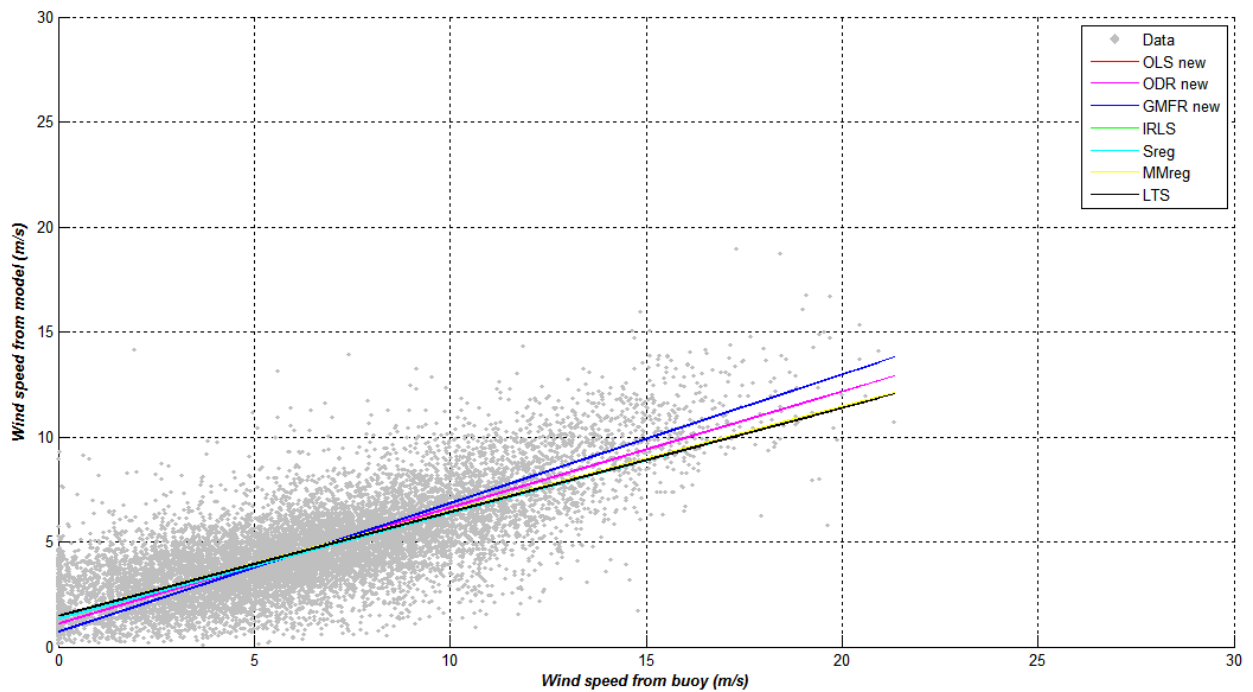
Σχήμα 4. 8. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση *Avgo*.



Σχήμα 4. 9. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση *Lesvos*.



Σχήμα 4. 10. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση *Mykonos*.



Σχήμα 4. 11. Γραμμές παλινδρόμησης των  $u_B$  και  $u_M$  μετά την αφαίρεση των έκτροπων παρατηρήσεων για διάφορες μεθόδους παλινδρόμησης στη θέση *Santorini*.

## 4.2. ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΑΠΟ ΤΙΣ ΔΙΑΦΟΡΕΤΙΚΕΣ ΜΕΘΟΔΟΥΣ

Σε αυτό το εδάφιο, θα χρησιμοποιηθούν συγκεκριμένοι στατιστικοί δείκτες ως κριτήρια για την αξιολόγηση και τον έλεγχο της γραμμικής προσαρμογής, όπως αυτή προέκυψε από την εφαρμογή των διαφόρων μεθόδων παλινδρόμησης. Υπενθυμίζεται ότι όταν συγκρίνονται δύο ομάδες  $n$  παρατηρήσεων,  $\hat{y}_i$  και  $y_i$ , η μεταβλητή  $\hat{y}_i$  είναι η προβλεπόμενη ή αναμενόμενη τιμή της μεταβλητής (*expected value*) και  $y_i$  η παρατηρηθείσα ή παρατηρούμενη τιμή της μεταβλητής (*observed value*). Στη συνέχεια δίνονται οι ορισμοί των δεικτών αυτών.

- Προσαρμοσμένος συντελεστής προσδιορισμού (*adjusted coefficient of determination*)

$$R_a^2 = 1 - (1 - R^2) \left( \frac{n-1}{n-2} \right), \text{ όπου } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

- Τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (*Root Mean Square Error – RMSE*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}}. \quad (84)$$

- Συστηματικό σφάλμα μεροληψίας (*BIAS*):

$$BIAS = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n}. \quad (85)$$

- Απόλυτο συστηματικό σφάλμα μεροληψίας (*Absolute BIAS*):

$$Absolute\_BIAS = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n}. \quad (86)$$

- Δείκτης διασποράς (*Scatter Index – SI*)

$$SI = \frac{RMSE}{(\bar{\hat{y}} + \bar{y})/2}. \quad (87)$$

Αρχικά, όλες οι τιμές των παραπάνω στατιστικών δεικτών καθώς και αυτές των παραμέτρων  $\hat{b}_0$  και  $\hat{b}_1$  του μοντέλου, όπως προκύπτουν από την εφαρμογή διαφόρων μεθόδων παλινδρόμησης, συνοψίζονται και παρουσιάζονται σε μορφή πίνακα (**Πίνακας 4. 3 – Πίνακας 4. 7**). Οι τιμές που είναι σημειωμένες με κόκκινο χρώμα αντιπροσωπεύουν την καλύτερη τιμή του εκάστοτε στατιστικού δείκτη για όλες τις μεθόδους πριν την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους παλινδρόμησης. Οι τιμές που

είναι με μπλε χρώμα αφορούν τις καλύτερες τιμές των στατιστικών δεικτών μετά την αφαίρεση των έκτροπων παρατηρήσεων. Η έντονη γραμματοσειρά υποδηλώνει ότι η συγκεκριμένη τιμή αποτελεί την καλύτερη τιμή και για τις δύο περιπτώσεις που διερευνούνται.

	$\hat{b}_0$	$\hat{b}_1$	$R_\alpha^2$	$RMSE$ (m/s)	$BIAS$ (m/s)	$Absolute$ $BIAS$ (m/s)	$SI$
<i>OLS</i>	1.4730 (1.329)	0.6409 (0.6706)	<b>0.6758</b> (0.6743)	<b>1.7964</b> (1.8004)	0 (0.0109)	1.3757 <b>(1.3700)</b>	<b>0.3735</b> (0.3739)
<i>ODR</i>	0.9582 (0.9708)	0.7398 (0.7411)	0.6597 (0.6592)	1.8404 (1.8417)	0 (0.0195)	1.3942 (1.3949)	0.3826 <b>(0.3736)</b>
<i>GMFR</i>	0.7508 (0.8266)	0.7796 (0.7695)	0.6441 (0.6485)	1.8820 (1.8703)	0 (0.0230)	1.4228 (1.4141)	0.3913 (0.3879)
<i>IRLS</i>	1.3682	0.6612	<b>0.6751</b>	<b>1.7982</b>	0	<b>1.3730</b>	0.3738
<i>S</i>	1.1873	0.6871	0.6721	1.8066	-0.0448	1.3749	0.3774
<i>MM</i>	1.3678	0.6613	<b>0.6751</b>	<b>1.7982</b>	0	<b>1.3730</b>	0.3738
<i>LTS</i>	1.3121	0.6737	0.6740	1.8013	0.0101	1.3732	0.3741

**Πίνακας 4. 3.** Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση *Athos* πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.

	$\hat{b}_0$	$\hat{b}_1$	$R_\alpha^2$	$RMSE$ (m/s)	$BIAS$ (m/s)	$Absolute$ $BIAS$ (m/s)	$SI$
<i>OLS</i>	2.4309 (2.2185)	0.5286 (0.5766)	<b>0.4979</b> (0.4927)	<b>2.0420</b> (2.0525)	0 (0.0947)	1.6318 (1.6181)	0.3512 (0.3502)
<i>ODR</i>	1.5411 (1.5197)	0.6676 (0.6876)	0.4634 (0.4515)	2.1109 (2.1343)	0 (0.1061)	1.6352 (1.6408)	0.3631 <b>(0.3499)</b>
<i>GMFR</i>	1.0197 (1.1461)	0.7491 (0.7469)	0.4112 (0.4114)	2.2112 (2.2108)	0 (0.1122)	1.6892 (1.6814)	0.3803 (0.3766)
<i>IRLS</i>	2.2933	0.5587	<b>0.4959</b>	<b>2.0460</b>	0.0548	1.6214	0.3503
<i>S</i>	1.7413	0.6893	0.4380	2.1604	0.3391	1.6486	0.3611
<i>MM</i>	2.2796	0.5618	0.4955	2.0469	0.0612	1.6206	<b>0.3502</b>
<i>LTS</i>	2.1055	0.6039	0.4848	2.0684	0.1566	<b>1.6176</b>	0.3511

**Πίνακας 4. 4.** Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση *Avgo* πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.

	$\hat{b}_0$	$\hat{b}_1$	$R_\alpha^2$	$RMSE$ (m/s)	$BIAS$ (m/s)	$Absolute$ $BIAS$ (m/s)	$SI$
<i>OLS</i>	1.5046 (1.1450)	0.5946 (0.6329)	<b>0.6084</b> (0.6051)	<b>1.9389</b> (1.9471)	0 (-0.0868)	1.4946 <b>(1.4862)</b>	<b>0.3378</b> (0.3418)
<i>ODR</i>	0.6972 (0.5579)	0.7080 (0.7149)	0.5863 (0.5827)	1.9929 (2.0017)	0 (-0.0897)	1.5095 (1.5121)	0.3472 (0.3419)
<i>GMFR</i>	0.3103 (0.2844)	0.7623 (0.7531)	0.5600 (0.5643)	2.0552 (2.0452)	0 (-0.0911)	1.5539 (1.5431)	0.3581 (0.3592)
<i>IRLS</i>	1.2036	0.6263	<b>0.6061</b>	<b>1.9447</b>	-0.0754	1.4866	<b>0.3410</b>
<i>S</i>	0.8473	0.6640	0.5974	1.9661	-0.1630	1.4908	0.3475
<i>MM</i>	1.2037	0.6263	<b>0.6061</b>	<b>1.9447</b>	-0.0754	1.4866	<b>0.3410</b>
<i>LTS</i>	1.1125	0.6361	0.6045	1.9486	-0.0966	<b>1.4862</b>	0.3424

**Πίνακας 4. 5.** Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση *Lesvos* πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων.

	$\hat{b}_0$	$\hat{b}_1$	$R_\alpha^2$	$RMSE$ (m/s)	$BIAS$ (m/s)	$Absolute$ $BIAS$ (m/s)	$SI$
<i>OLS</i>	1.4235 (1.2619)	0.6398 (0.6592)	<b>0.7172</b> (0.7165)	<b>1.7253</b> (1.7274)	0 (-0.0066)	1.3381 (1.3379)	<b>0.2647</b> (0.2652)
<i>ODR</i>	0.7905 (0.7893)	0.7193 (0.7189)	0.7061 (0.7062)	1.7588 (1.7585)	0 (-0.0037)	1.3592 (1.3592)	0.2699 (0.2651)
<i>GMFR</i>	0.5024 (0.5726)	0.7554 (0.7463)	0.6937 (0.6973)	1.7954 (1.7850)	0 (-0.0024)	1.3864 (1.3785)	0.2755 (0.2739)
<i>IRLS</i>	1.3110	0.6535	<b>0.7168</b>	1.7264	-0.0027	<b>1.3375</b>	<b>0.2649</b>
<i>S</i>	1.2354	0.6569	0.7164	1.7277	-0.0519	1.3383	0.2662
<i>MM</i>	1.3123	0.6534	<b>0.7168</b>	<b>1.7263</b>	-0.0026	<b>1.3375</b>	<b>0.2649</b>
<i>LTS</i>	1.2553	0.6595	0.7165	1.7275	-0.0109	1.3379	0.2653

**Πίνακας 4. 6.** Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση *Mykonos* πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων

	$\hat{b}_0$	$\hat{b}_1$	$R_\alpha^2$	$RMSE$ (m/s)	$BIAS$ (m/s)	$Absolute$ $BIAS$ (m/s)	$SI$
<i>OLS</i>	1.5302 (1.4775)	0.4980 (0.4967)	<b>0.5757</b> (0.5750)	<b>1.5963</b> (1.5975)	0 (-0.0609)	1.2160 ( <b>1.2145</b> )	<b>0.3298</b> (0.3321)
<i>ODR</i>	0.9833 (0.1121)	0.5803 (0.5526)	0.5599 (0.5682)	1.6256 (1.6102)	0 (-0.0548)	1.2369 (1.2236)	0.3358 (0.3319)
<i>GMFR</i>	0.4777 (0.7180)	0.6563 (0.6129)	0.5175 (0.5446)	1.7023 (1.6537)	0 (-0.0483)	1.2983 (1.2579)	0.3517 (0.3434)
<i>IRLS</i>	1.4779	0.4988	<b>0.5753</b>	<b>1.5970</b>	-0.0470	1.2146	<b>0.3315</b>
<i>S</i>	1.3524	0.5024	0.5719	1.6033	-0.1485	1.2160	0.3364
<i>MM</i>	1.4783	0.4988	<b>0.5753</b>	<b>1.5970</b>	-0.0465	1.2146	<b>0.3315</b>
<i>LTS</i>	1.4755	0.4949	0.5747	1.5981	-0.0749	<b>1.2145</b>	0.3327

**Πίνακας 4. 7.** Τιμές των στατιστικών δεικτών για όλες τις μεθόδους για τη θέση *Santorini* πριν και μετά (σε παρένθεση) την αφαίρεση των έκτροπων παρατηρήσεων

Στη συνέχεια, εξετάζεται αναλυτικά κάθε στατιστικός δείκτης σε κάθε θέση και προβάλλονται τα ανάλογα γραφήματα πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους παλινδρόμησης.

Συνοπτικά, μελετώντας τους **Πίνακας 4. 3 – Πίνακας 4. 7** προκύπτει ότι:

1. Πριν την αφαίρεση των έκτροπων παρατηρήσεων, η μέθοδος που παρουσιάζει τις καλύτερες τιμές αναφορικά με το σύνολο των στατιστικών δεικτών αξιολόγησης που εφαρμόστηκαν σε όλες τις θέσεις είναι η **OLS**. Η ίδια μέθοδος δίνει συστηματικά τις μεγαλύτερες τιμές για τον συντελεστή  $R_\alpha^2$  σε κάθε θέση. Σε σύγκριση με τις υπόλοιπες μεθόδους παλινδρόμησης οι **IRLS**, **MM** και **LTS** δίνουν τις ελάχιστες τιμές κυρίως ως προς το δείκτη **Absolute BIAS**, ενώ οι υπόλοιπες μέθοδοι (**ODR**, **GMFR**, **S**) δεν εμφανίζουν σε καμία μέθοδο και θέση κάποια τιμή στατιστικού δείκτη σημαντικά καλύτερη από τις άλλες διαδικασίες.
2. Μετά την αφαίρεση των έκτροπων παρατηρήσεων, οι μέθοδοι που παρουσιάζουν τις καλύτερες τιμές αναφορικά με το σύνολο των στατιστικών δεικτών αξιολόγησης που εφαρμόστηκαν σε όλες τις θέσεις είναι οι **IRLS** και **MM**. Αυτό σημαίνει ότι η απόδοση των κλασσικών μεθόδων μετά την αφαίρεση των έκτροπων παρατηρήσεων, γίνεται

μάλλον χειρότερη. Επίσης, αξίζει να αναφερθεί ότι οι τιμές του συντελεστή  $R_\alpha^2$  και του δείκτη  $RMSE$  είναι μέγιστες και ελάχιστες αντίστοιχα σε κάθε θέση μόνο με τις δύο εν λόγω μεθόδους. Σε σύγκριση με τις υπόλοιπες μεθόδους παλινδρόμησης, οι μέθοδοι  $OLS$  και  $LTS$  παρουσιάζουν τις ελάχιστες τιμές ως προς το δείκτη  $Absolute\ BIAS$  και η  $ODR$  ως προς το δείκτη  $SI$ , ενώ και πάλι οι μέθοδοι  $GMFR$  και  $S$  δεν εμφανίζουν σε καμία μέθοδο και θέση κάποια τιμή στατιστικού δείκτη σημαντικά καλύτερη από τις άλλες διαδικασίες.

### Α. Συντελεστές $\hat{b}_0$ και $\hat{b}_1$ της ευθείας παλινδρόμησης

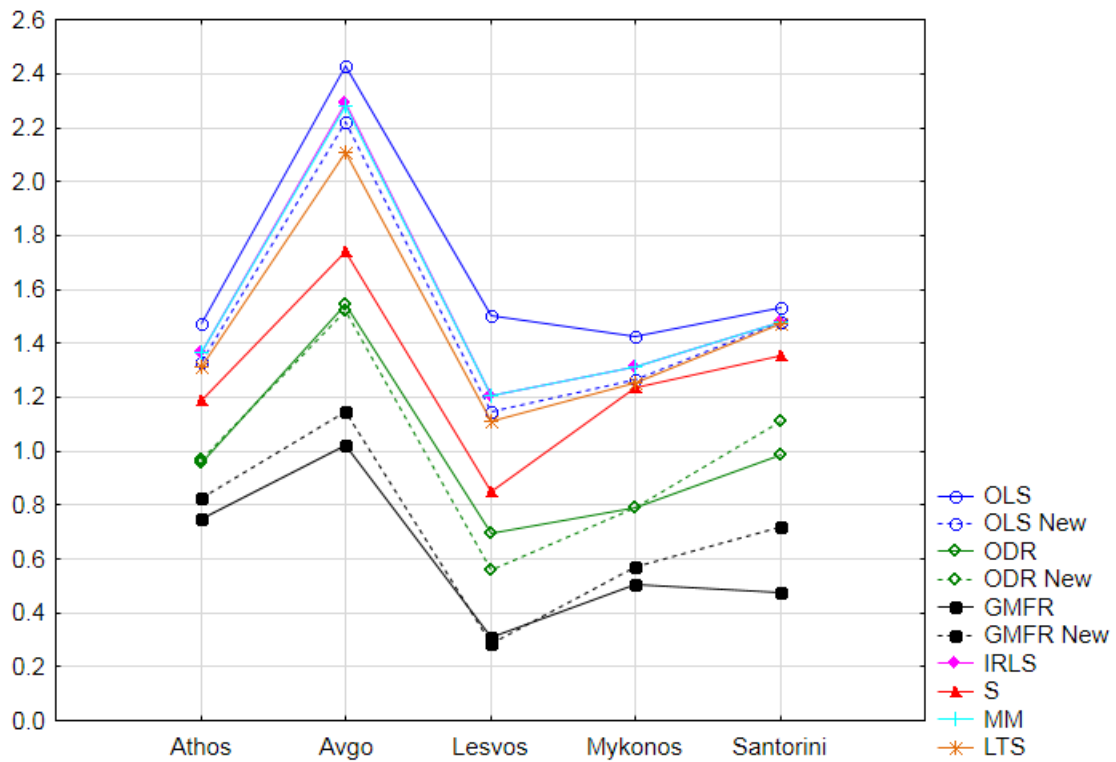
Εξετάζοντας τα **Γράφημα 4. 1** και **Γράφημα 4. 2** προκύπτουν τα ακόλουθα συμπεράσματα:

- Πριν την αφαίρεση των έκτροπων παρατηρήσεων

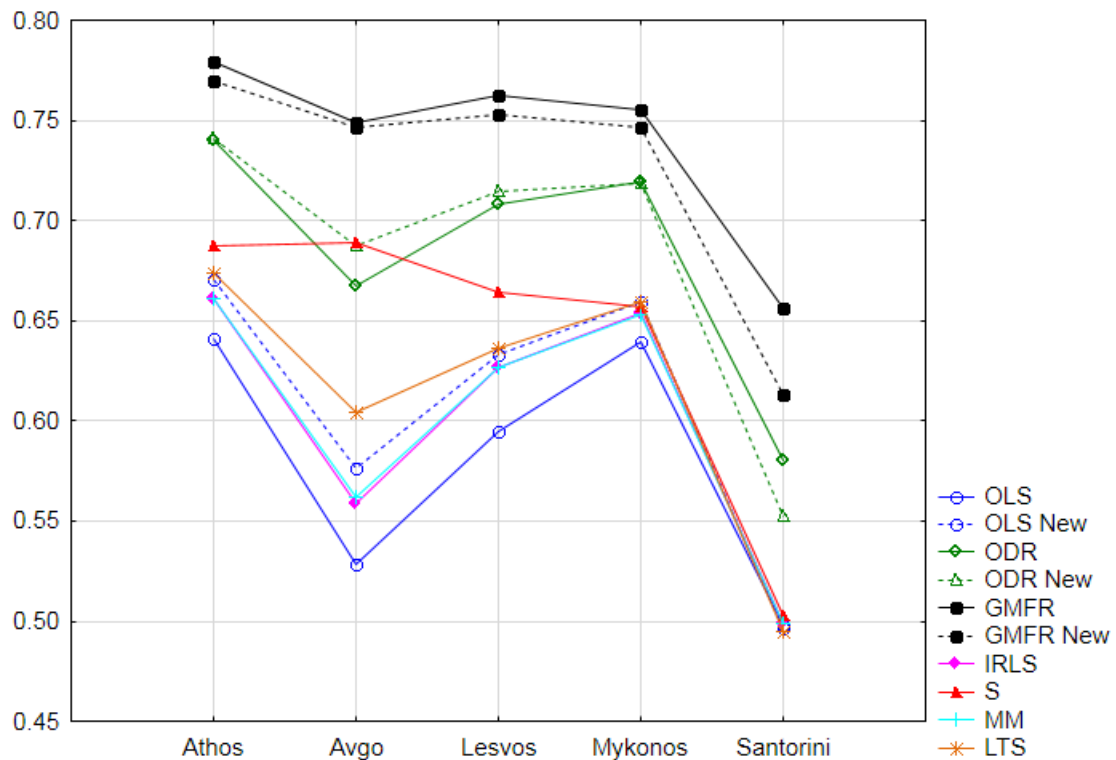
1. Οι μέθοδοι  $GMFR$  και  $OLS$  παρουσιάζουν τις ακραίες τιμές των παραμέτρων  $\hat{b}_0$  και  $\hat{b}_1$  της ευθείας παλινδρόμησης για όλες τις τοποθεσίες. Πιο συγκεκριμένα, στην περίπτωση της παραμέτρου  $\hat{b}_0$  οι μέγιστες τιμές αφορούν τη μέθοδο  $GMFR$  και οι ελάχιστες την  $OLS$  (βλ. Γράφημα 4.1.). Το αντίστροφο ισχύει για την παράμετρο  $\hat{b}_1$  (βλ. Γράφημα 4.2.). Η μέθοδος  $GMFR$  εμφανίζει το μικρότερο εύρος τιμών για όλες τις θέσεις. Έτσι, από την άποψη της επίδρασης της τοπικότητας στις τιμές των παραμέτρων  $\hat{b}_0$  και  $\hat{b}_1$ , φαίνεται ότι η μέθοδος  $GMFR$  έχει τη μικρότερη ευαισθησία. Επίσης, και στις δύο περιπτώσεις, το μεγαλύτερο εύρος τιμών δίνεται από ανθεκτικές μεθόδους. Για την παράμετρο  $\hat{b}_0$  της ευθείας παλινδρόμησης δίνεται από τη μέθοδο  $IRLS$  και για την παράμετρο  $\hat{b}_1$  από τη μέθοδο  $S$  και την  $ODR$  (μετά). Τέλος, οι μέθοδοι  $IRLS$  και  $MM$  δίνουν πολύ παραπλήσιες μεταξύ τους τιμές των παραμέτρων  $\hat{b}_0$  και  $\hat{b}_1$  σε όλες τις τοποθεσίες.
2. Οι τιμές των παραμέτρων  $\hat{b}_0$  και  $\hat{b}_1$  για όλες τις ανθεκτικές και για τη μέθοδο  $OLS$  είναι πολύ κοντά μεταξύ τους στις θέσεις *Mykonos* και *Santorini*.
3. Η θέση *Avgo* παρουσιάζει το μεγαλύτερο εύρος τιμών και για τους δύο συντελεστές για όλες τις μεθόδους, ενώ οι θέσεις *Athos* και *Mykonos* παρουσιάζουν το μικρότερο για το συντελεστή  $\hat{b}_0$  και  $\hat{b}_1$  αντίστοιχα.

- Μετά την αφαίρεση των έκτροπων παρατηρήσεων

1. Η μέθοδος  $GMFR$  συνεχίζει να παρουσιάζει την ίδια συμπεριφορά όπως και ανωτέρω, ενώ η μέθοδος  $OLS$  προσεγγίζει τώρα τις ανθεκτικές μεθόδους ( $IRLS$ ,  $MM$ ,  $LTS$ ) και για τις δύο παραμέτρους σε όλες τις τοποθεσίες. Για τη συμπεριφορά των μεθόδων  $GMFR$  και  $IRLS$  ισχύουν οι ίδιες διαπιστώσεις που έγιναν παραπάνω αναφορικά με την παράμετρο  $\hat{b}_0$ , ενώ η μέθοδος  $ODR$  δίνει τώρα το μεγαλύτερο εύρος τιμών αναφορικά με την παράμετρο  $\hat{b}_1$ .
2. Αναφορικά με τις τοποθεσίες, η θέση *Avgo* εξακολουθεί να παρουσιάζει και τώρα το μεγαλύτερο εύρος τιμών για όλες τις μεθόδους παλινδρόμησης. Οι παραπάνω διαπιστώσεις ισχύουν για τα σημεία *Mykonos* και *Athos*. Γενικά, παρατηρείται ότι, και για τις δύο παραμέτρους της ευθείας παλινδρόμησης, το εύρος τιμών έχει ελαττωθεί σε όλες τις θέσεις.



**Γράφημα 4. 1.** Τιμές της παραμέτρου  $\hat{b}_0$  της ευθείας παλινδρόμησης όλων των τοποθεσιών για όλες τις μεθόδους πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.

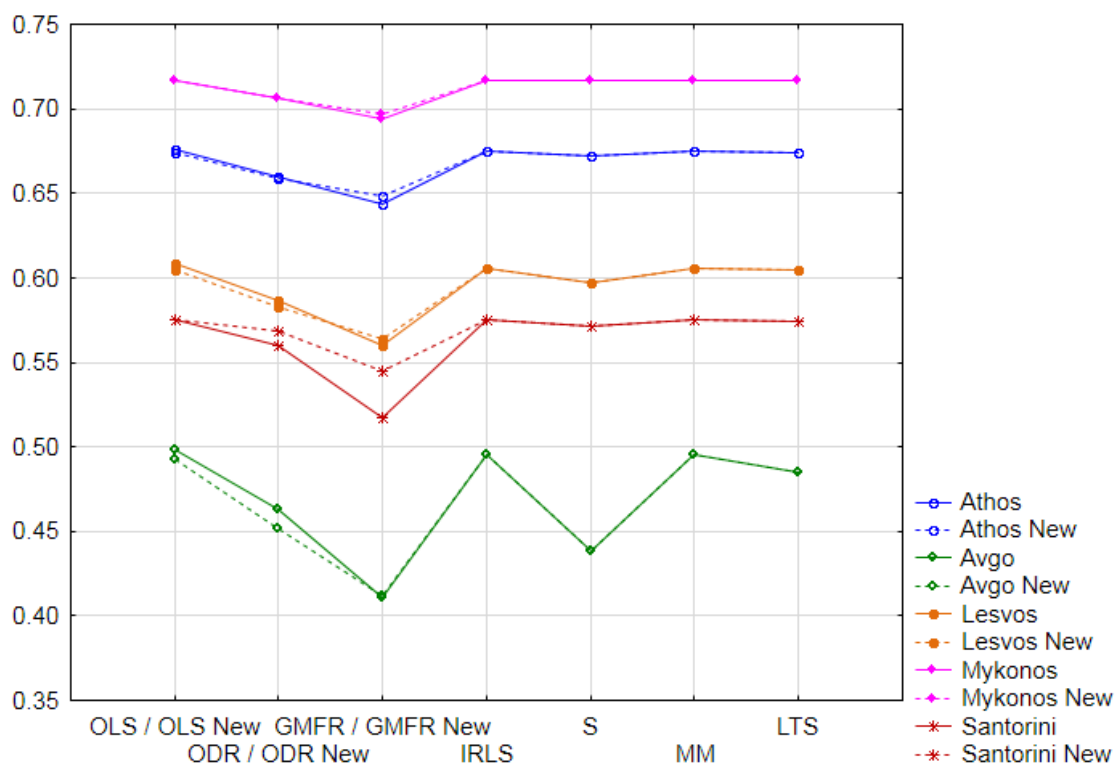


**Γράφημα 4. 2.** Τιμές της παραμέτρου  $\hat{b}_1$  της ευθείας παλινδρόμησης όλων των τοποθεσιών για όλες τις μεθόδους πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.

## Β. Προσαρμοσμένος συντελεστής προσδιορισμού ( $R_\alpha^2$ )

Εξετάζοντας το **Γράφημα 4.3** προκύπτουν τα ακόλουθα συμπεράσματα:

- Πριν την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η θέση *Mykonos* παρουσιάζει το μικρότερο εύρος τιμών του συντελεστή  $R_\alpha^2$  (0.0235) για όλες τις μεθόδους παλινδρόμησης, ενώ στη θέση *Avgo* το μεγαλύτερο εύρος (0.0867).
  2. Η μέθοδος *GMFR* δίνει το μεγαλύτερο εύρος τιμών του συντελεστή  $R_\alpha^2$  (0.2825), ενώ η μέθοδος *OLS* το μικρότερο (0.2193) με μικρή διαφορά από τις ανθεκτικές μεθόδους *IRLS* και *MM*.
  3. Με εξαίρεση το σημείο *Avgo*, στα υπόλοιπα σημεία επισημαίνεται ότι υπάρχει αξιοσημείωτη σταθερότητα αναφορικά με τις τιμές του συγκεκριμένου στατιστικού δείκτη σε όλες τις ανθεκτικές μεθόδους.
- Μετά την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η συμπεριφορά στις θέσεις *Mykonos* και *Avgo* παραμένει η ίδια.
  2. Η μέθοδος *GMFR* συνεχίζει να δίνει το μεγαλύτερο εύρος τιμών του συντελεστή  $R_\alpha^2$  (0.2859), ενώ η μέθοδος *IRLS* το μικρότερο (0.2209), με μικρή διαφορά από τις μεθόδους *MM* και *OLS*. Γενικά, παρατηρείται ότι, μετά την αφαίρεση των έκτροπων παρατηρήσεων, οι τιμές του συγκεκριμένου συντελεστή έχουν την τάση να αυξάνονται, κυρίως για τη μέθοδο *GMFR*.



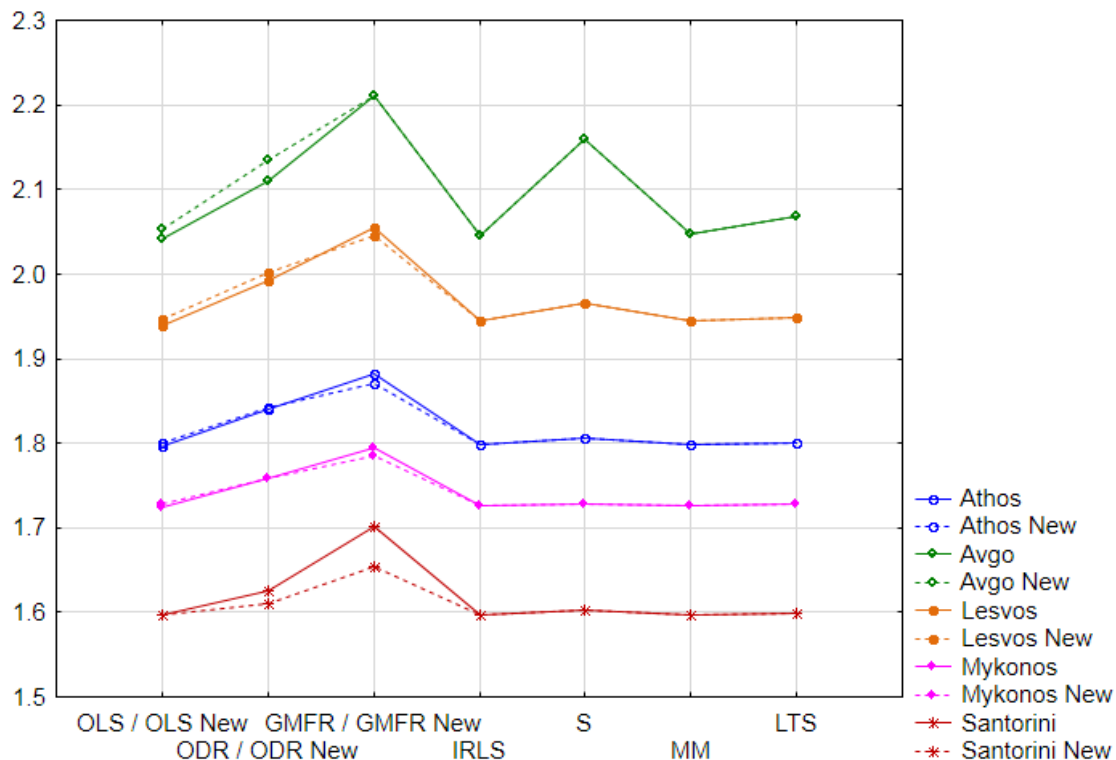
**Γράφημα 4.3.** Τιμές του συντελεστή  $R_\alpha^2$  όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.



### Γ. Τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (*RMSE*)

Εξετάζοντας το **Γράφημα 4. 4** προκύπτουν τα ακόλουθα συμπεράσματα:

- Πριν την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η θέση *Avgo* παρουσιάζει το μεγαλύτερο εύρος τιμών του *RMSE* (0.1692) για όλες τις μεθόδους παλινδρόμησης, ενώ στη θέση *Mykonos* συναντάται το μικρότερο (0.0701).
  2. Η μέθοδος *S* δίνει το μεγαλύτερο εύρος τιμών του *RMSE* (0.5571), ενώ η μέθοδος *OLS* το μικρότερο (0.4457) με μικρή διαφορά από τις μεθόδους *IRLS* και *MM*.
  3. Με εξαίρεση τα σημεία *Avgo* και *Lesvos*, στα υπόλοιπα σημεία επισημαίνεται ότι υπάρχει αξιοσημείωτη σταθερότητα αναφορικά με τις τιμές του συγκεκριμένου στατιστικού δείκτη σε όλες τις ανθεκτικές μεθόδους.
- Μετά την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η συμπεριφορά του *RMSE* στη θέση *Avgo* παραμένει ποιοτικά η ίδια, ενώ η θέση *Santorini* δίνει τώρα το μικρότερο εύρος τιμών του *RMSE* (0.0567).
  2. Η μέθοδος *GMFR* δίνει πάλι το μεγαλύτερο εύρος τιμών του *RMSE* (0.5571) μαζί με την *S*, ενώ η μέθοδος *IRLS* το μικρότερο (0.4490) με μικρή διαφορά από τις μεθόδους *MM* και *OLS*. Γενικά, παρατηρείται ότι, μετά την αφαίρεση των έκτροπων παρατηρήσεων, οι τιμές του συγκεκριμένου συντελεστή έχουν την τάση να μειώνονται, κυρίως για τη μέθοδο *GMFR* και *ODR*, σε όλες τις θέσεις εκτός από το *Avgo*.

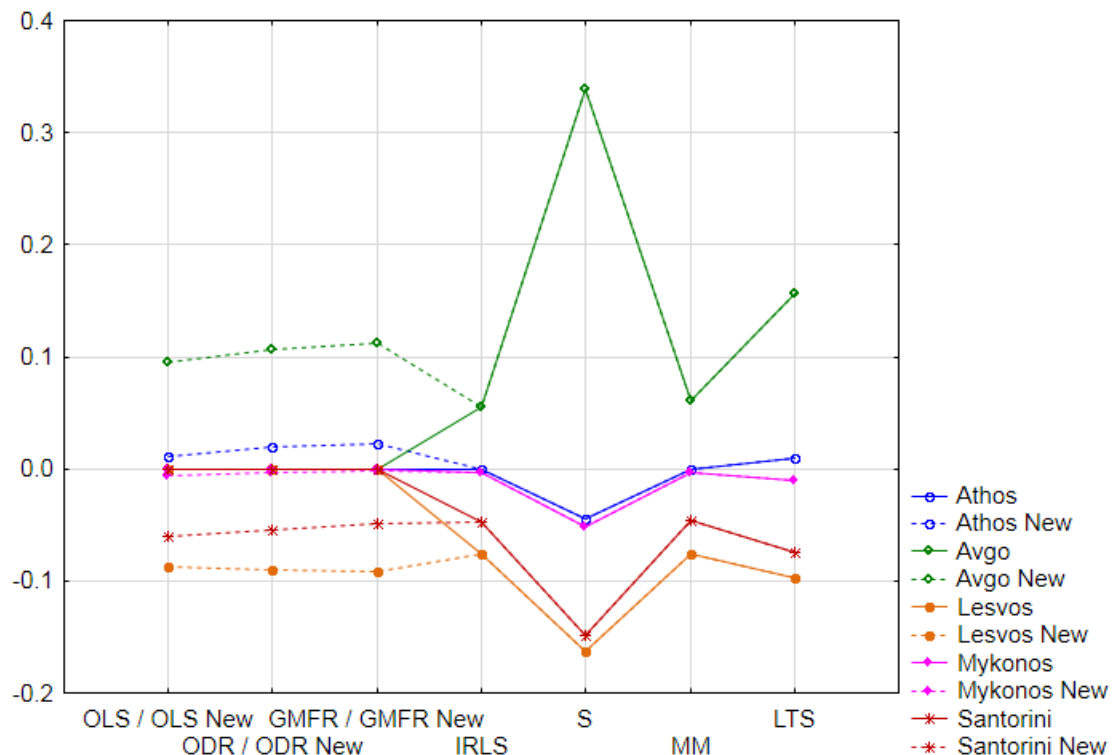


**Γράφημα 4. 4.** Τιμές του δείκτη *RMSE* όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.

#### Δ. Συστηματικό σφάλμα μεροληψίας (*BIAS*)

Εξετάζοντας το **Γράφημα 4.5** προκύπτουν τα ακόλουθα συμπεράσματα:

- Πριν την αφαίρεση των έκτροπων παρατηρήσεων
  1. Οι κλασσικές μέθοδοι παλινδρόμησης δίνουν μηδενικές τιμές του δείκτη *BIAS* και οι αντίστοιχες για τις ανθεκτικές τείνουν στο μηδέν, όπως ήταν αναμενόμενο, για όλες τις θέσεις. Αναφορικά με τις ανθεκτικές μεθόδους, εκείνη η μέθοδος που δίνει το μικρότερο εύρος τιμών του *BIAS* είναι η *IRLS* (0.1302) και εκείνη που δίνει το μεγαλύτερο είναι η *S* (0.5021).
  2. Οι ανθεκτικές μέθοδοι παρουσιάζουν σε όλες τις θέσεις αρνητικές τιμές (δηλαδή υποεκτιμούν το σύνολο των τιμών της εξαρτημένης μεταβλητής) με εξαίρεση το *Avgo*. Στην ίδια θέση παρουσιάζεται το μεγαλύτερο εύρος τιμών του *BIAS* (0.3391), ενώ το μικρότερο εμφανίζεται στη θέση *Mykonos* (0.0519) για όλες τις μεθόδους παλινδρόμησης.
- Μετά την αφαίρεση των έκτροπων παρατηρήσεων
  1. Για όλες τις θέσεις, με εξαίρεση το *Avgo*, όλες οι μέθοδοι παλινδρόμησης εκτός της ανθεκτικής μεθόδου *S* δίνουν αποτελέσματα πολύ κοντά μεταξύ τους.
  2. Η θέση *Avgo* εξακολουθεί να παρουσιάζει το μεγαλύτερο εύρος τιμών του *BIAS* (0.2843), ενώ το μικρότερο εύρος τιμών του εμφανίζεται και πάλι στη θέση *Mykonos* (0.0495) για όλες τις μεθόδους παλινδρόμησης.
  3. Η συμπεριφορά του *BIAS* για τις μεθόδους *IRLS* και *S* παραμένει ποιοτικά η ίδια.

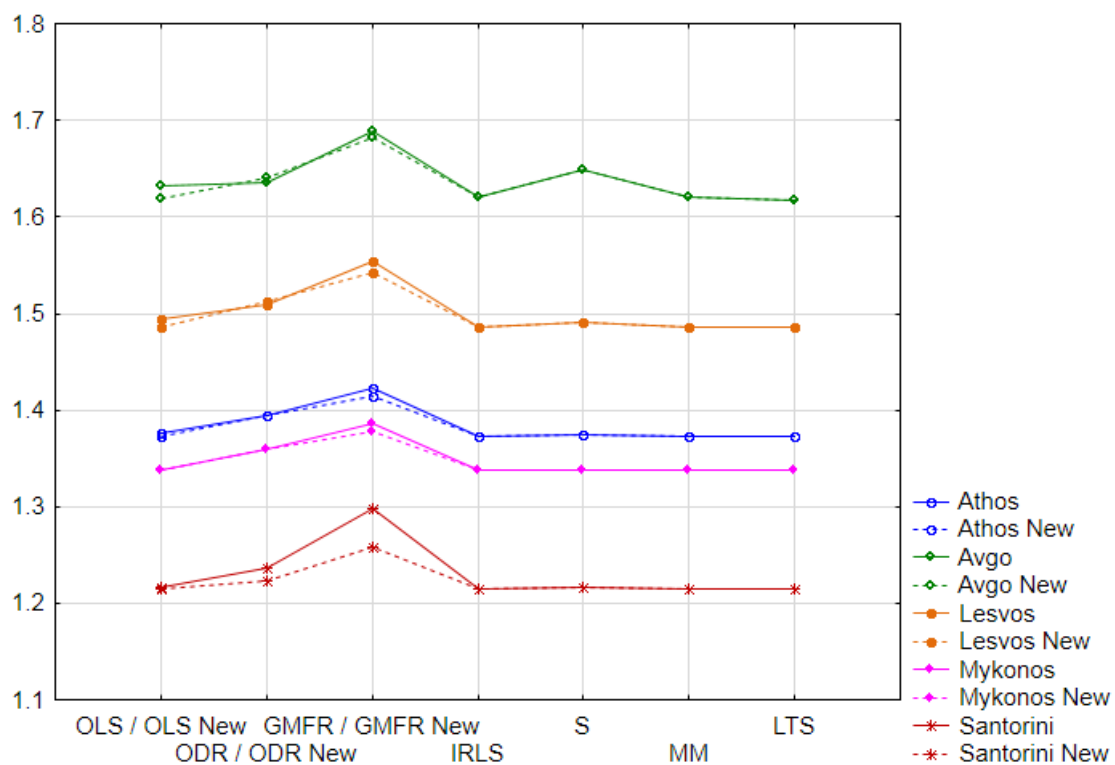


**Γράφημα 4.5.** Τιμές του δείκτη *BIAS* όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.

### Ε. Απόλυτο συστηματικό σφάλμα μεροληψίας (*Absolute BIAS*)

Εξετάζοντας το **Γράφημα 4. 6** προκύπτουν τα ακόλουθα συμπεράσματα:

- Πριν την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η θέση *Santorini* παρουσιάζει το μεγαλύτερο εύρος τιμών του *Absolute BIAS* (0.0838), ενώ το μικρότερο εύρος τιμών του εμφανίζεται στη θέση *Mykonos* (0.0489).
  2. Η μέθοδος *GMFR* δίνει το μικρότερο εύρος τιμών του *Absolute BIAS* (0.3909) σε σύγκριση με τις υπόλοιπες μεθόδους σε όλες τις θέσεις και, μαζί με τη μέθοδο *ODR*, δίνει τιμές αρκετά διαφορετικές σε αντίθεση με τις υπόλοιπες μεθόδους παλινδρόμησης, που φαίνεται από το Γράφημα 4.6. να προσεγγίζει πολύ η μία την άλλη, εκτός από τη θέση *Avgo*. Αντίθετα, το μεγαλύτερο εύρος τιμών δίνεται από τη μέθοδο *S* (0.4326). Τέλος, η σταθερότητα στις τιμές του δείκτη για τις ανθεκτικές μεθόδους παλινδρόμησης είναι εμφανής σε κάθε εξεταζόμενη περιοχή, εκτός από το *Avgo*.
- Μετά την αφαίρεση των έκτροπων παρατηρήσεων
  1. Στη θέση *Mykonos* η συμπεριφορά του δείκτη παραμένει ποιοτικά η ίδια όπως και ανωτέρω, ενώ το σημείο *Avgo* δίνει τώρα το μεγαλύτερο εύρος τιμών του *Absolute BIAS* (0.0638).
  2. Η μέθοδος *S* δίνει ποιοτικά τα ίδια αποτελέσματα όπως και ανωτέρω και η ανθεκτική μέθοδος *LTS* δίνει τώρα το μικρότερο εύρος τιμών.
  3. Η γενική τάση που παρατηρείται για τις κλασσικές μεθόδους παλινδρόμησης στις τιμές του συγκεκριμένου δείκτη είναι ότι υπάρχει μείωση, με το μεγαλύτερο ποσοστό να εμφανίζεται στη μέθοδο *GMFR* και τη θέση *Santorini*.

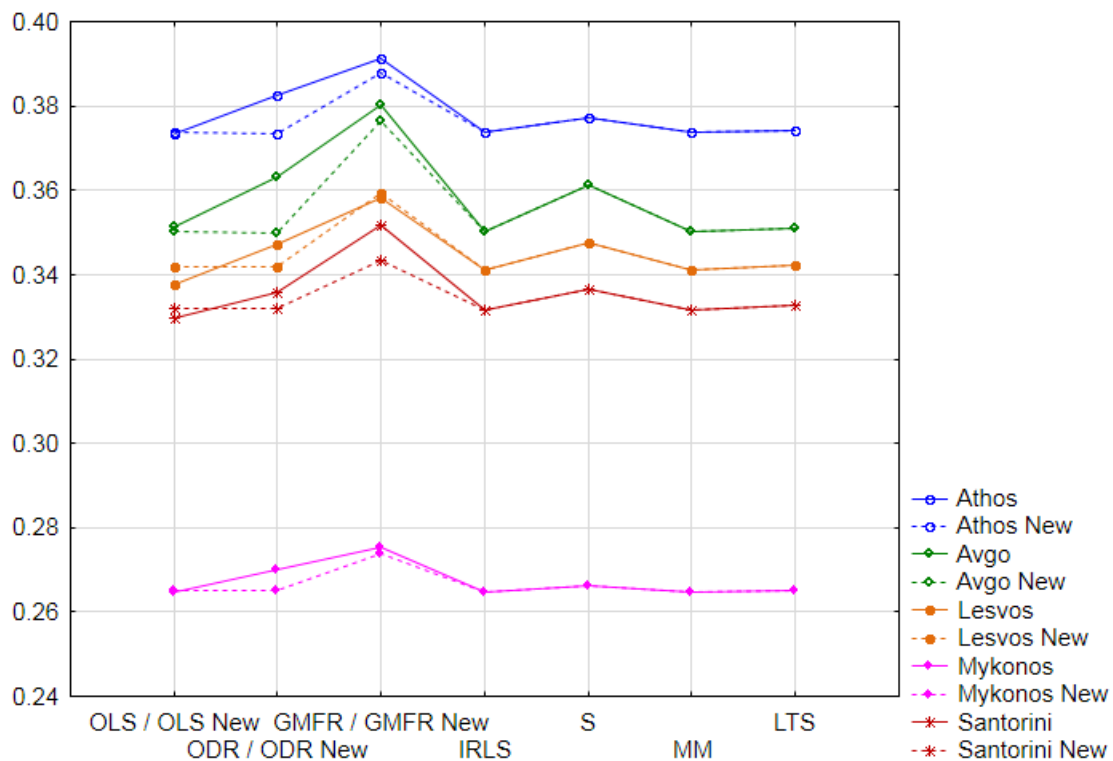


**Γράφημα 4. 6.** Τιμές του δείκτη *Absolute BIAS* όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.

## ΣΤ. Δείκτης διασποράς (SI)

Εξετάζοντας το **Γράφημα 4.7** προκύπτουν τα ακόλουθα συμπεράσματα:

- Πριν την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η θέση *Avgo* παρουσιάζει το μεγαλύτερο εύρος τιμών του δείκτη *SI* (0.0301), ενώ στη θέση *Μυκονος* το μικρότερο (0.0108) για όλες τις μεθόδους παλινδρόμησης.
  2. Η μέθοδος *GMFR* δίνει το μεγαλύτερο εύρος των τιμών του δείκτη *SI* (0.1158) σε σύγκριση με τις υπόλοιπες μεθόδους για όλες τις θέσεις. Αντίθετα, οι μέθοδοι *OLS*, *IRLS*, *MM* και *LTS* εμφανίζουν το μικρότερο εύρος τιμών (~0.1090).
- Μετά την αφαίρεση των έκτροπων παρατηρήσεων
  1. Η συμπεριφορά του δείκτη *SI* στις θέσεις *Avgo* και *Μυκονος* παραμένει ποιοτικά η ίδια όπως και ανωτέρω.
  2. Η μέθοδος *GMFR* εξακολουθεί να παρουσιάζει την ίδια συμπεριφορά όπως και ανωτέρω. Αντίθετα, η μέθοδος που εμφανίζει μετά την αφαίρεση των έκτροπων παρατηρήσεων το μικρότερο εύρος τιμών είναι η μέθοδος *ODR* (0.1085) με μικρή διαφορά από τις μεθόδους *OLS*, *IRLS*, *MM* και *LTS*, που έδιναν πριν τα μικρότερα εύρη τιμών.
  3. Παρατηρείται ότι όλες οι μέθοδοι παλινδρόμησης, με εξαίρεση τις μεθόδους *GMFR* και *S*, όπου παρατηρούνται λίγο μεγαλύτερες τιμές του εν λόγω δείκτη σε όλες τις θέσεις, έχουν πολύ κοντινές τιμές του δείκτη *SI* σε όλες τις θέσεις. Γενικά, παρατηρείται ότι, μετά την αφαίρεση των έκτροπων παρατηρήσεων, οι τιμές του συγκεκριμένου συντελεστή στις μεθόδους *GMFR* και *ODR* έχουν την τάση να μειώνονται.



**Γράφημα 4.7.** Τιμές του δείκτη *SI* όλων των μεθόδων για όλες τις τοποθεσίες πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων από τις κλασσικές μεθόδους.

Συνοψίζοντας τους παραπάνω σχολιασμούς των αποτελεσμάτων, προκύπτουν τα ακόλουθα κοινά συμπεράσματα:

- Για όλους τους στατιστικούς δείκτες που χρησιμοποιήθηκαν, η θέση *Avgo* παρουσιάζει το μεγαλύτερο εύρος τιμών συγκριτικά με τις υπόλοιπες θέσεις και πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων, με εξαίρεση το *Absolute BIAS* (πριν). Αντίθετα, η θέση *Mykonos* παρουσιάζει το μικρότερο, με εξαίρεση το *RMSE* (μετά).
- Για όλους τους δείκτες, το μεγαλύτερο εύρος τιμών δίνεται από τις μεθόδους *GMFR* και *S* συγκριτικά με τις υπόλοιπες μεθόδους παλινδρόμησης και πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων. Το μικρότερο εύρος των  $R_\alpha^2$ , *RMSE* και *SI* δίνεται από τις μεθόδους *OLS* (με ελάχιστες διαφορές από τις μεθόδους *IRLS* και *MM*) και *IRLS* (με ελάχιστες διαφορές από τις μεθόδους *OLS* και *MM*) πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων, αντίστοιχα.
- Εκτός από τη θέση *Avgo*, η σταθερότητα των τιμών των  $R_\alpha^2$ , *RMSE* και *Absolute BIAS* για όλες τις ανθεκτικές μεθόδους παλινδρόμησης είναι εμφανής.
- Μετά την αφαίρεση των έκτροπων παρατηρήσεων, οι τιμές όλων των στατιστικών δεικτών που προκύπτουν από την εφαρμογή της μεθόδου *GMFR* καλυτερεύουν αισθητά σε σύγκριση με τις άλλες δύο κλασσικές μεθόδους παλινδρόμησης.

Σε δεύτερη φάση, ελέγχεται η αποδοτικότητα της εξίσωσης διόρθωσης με δύο τρόπους, όπως αναφέρθηκε. Στους **Πίνακας 4. 8** – **Πίνακας 4. 10** παρουσιάζονται οι τιμές των δεικτών *RMSE*, *Absolute BIAS* και *SI* σε όλες τις θέσεις και για όλες τις μεθόδους παλινδρόμησης. Οι τιμές που είναι σημειωμένες με κόκκινο χρώμα αντιπροσωπεύουν την καλύτερη τιμή του εκάστοτε στατιστικού δείκτη σε κάθε θέση μελέτης για όλες τις μεθόδους και για όλα τα δειγματικό πλήθος των ανεμολογικών δεδομένων. Οι τιμές που είναι με μπλε χρώμα αφορούν τις καλύτερες τιμές των στατιστικών δεικτών για το μισό μέγεθος του δείγματος κάθε θέσης.

	<i>Athos</i>	<i>Avgo</i>	<i>Lesvos</i>	<i>Mykonos</i>	<i>Santorini</i>
<i>OLS</i>	2.8029 (2.9933)	3.8631 (3.8664)	3.2608 (3.2778)	2.6969 (2.8010)	3.2055 (3.2565)
<i>ODR</i>	2.4878 (2.5760)	3.1618 (3.3731)	2.8150 (2.8811)	2.4453 (2.4924)	2.8015 (2.8537)
<i>GMFR</i>	<b>2.4141</b> (2.4580)	<b>2.9518</b> (3.2391)	<b>2.6962</b> (2.7795)	<b>2.3767</b> (2.3807)	<b>2.5937</b> (2.6139)
<i>IRLS</i>	2.7196 (2.8414)	3.6624 (3.6996)	3.1051 (3.1640)	2.6416 (2.7301)	3.2018 (3.2746)
<i>S</i>	2.6292 (2.6704)	3.1340 (3.4265)	2.9610 (3.0635)	2.6302 (2.7208)	3.1912 (3.1530)
<i>MM</i>	2.7194 (2.8411)	3.6436 (3.6906)	3.1051 (3.1634)	2.6422 (2.7305)	3.2017 (3.2745)
<i>LTS</i>	2.6736 (2.7212)	3.4251 (3.5854)	3.0635 (3.1277)	2.6193 (2.7002)	3.2289 (3.3008)

**Πίνακας 4. 8.** Τιμές του *RMSE* για όλες τις μεθόδους σε όλες τις θέσεις για το 100% και το 50% (σε παρένθεση) του δειγματικού πλήθους των ανεμολογικών δεδομένων.

	<i>Athos</i>	<i>Avgo</i>	<i>Lesvos</i>	<i>Mykonos</i>	<i>Santorini</i>
<i>OLS</i>	2.1466 (2.3142)	3.0870 (3.0568)	2.5135 (2.4865)	2.0916 (2.1789)	2.4419 (2.4686)
<i>ODR</i>	1.8846 (1.9836)	2.4492 (2.6140)	2.1321 (2.1478)	1.8899 (1.9364)	2.1316 (2.1638)
<i>GMFR</i>	1.8250 (1.8931)	2.2550 (2.4969)	2.0385 (2.0690)	1.8352 (1.8494)	1.9782 (1.9830)
<i>IRLS</i>	2.0764 (2.1944)	2.9022 (2.9045)	2.3737 (2.3828)	2.0466 (2.1222)	2.3737 (2.4779)
<i>S</i>	2.0009 (2.0579)	2.3916 (2.6308)	2.2451 (2.2925)	2.0374 (2.1145)	2.4203 (2.3826)
<i>MM</i>	2.0762 (2.1941)	2.8847 (2.8961)	2.3737 (2.3823)	2.0471 (2.1226)	2.4351 (2.4779)
<i>LTS</i>	2.0381 (2.0986)	2.6786 (2.7969)	2.3364 (2.3499)	2.0287 (2.0986)	2.4539 (2.4955)

**Πίνακας 4. 9.** Τιμές του *Absolute BIAS* για όλες τις μεθόδους σε όλες τις θέσεις για το 100% και το 50% (σε παρένθεση) του δειγματικού πλήθους των ανεμολογικών δεδομένων.

	<i>Athos</i>	<i>Avgo</i>	<i>Lesvos</i>	<i>Mykonos</i>	<i>Santorini</i>
<i>OLS</i>	0.5596 (0.5723)	0.6326 (0.6641)	0.5070 (0.5082)	0.3725 (0.3932)	0.5581 (0.5747)
<i>ODR</i>	0.4967 (0.4940)	0.5178 (0.5723)	0.4377 (0.4456)	0.3378 (0.3486)	0.4877 (0.5017)
<i>GMFR</i>	0.4820 (0.4720)	0.4834 (0.5472)	0.4192 (0.4295)	0.3283 (0.3325)	0.4516 (0.4582)
<i>IRLS</i>	0.5431 (0.5453)	0.6046 (0.6409)	0.4783 (0.4869)	0.3648 (0.3829)	0.5529 (0.5734)
<i>S</i>	0.5216 (0.5145)	0.5348 (0.6135)	0.4518 (0.4665)	0.3613 (0.3790)	0.5417 (0.5475)
<i>MM</i>	0.5430 (0.5452)	0.6020 (0.6397)	0.4783 (0.4868)	0.3649 (0.3830)	0.5529 (0.5734)
<i>LTS</i>	0.5346 (0.5251)	0.5730 (0.6254)	0.4708 (0.4798)	0.3614 (0.3784)	0.5548 (0.5757)

**Πίνακας 4. 10.** Τιμές του *SI* για όλες τις μεθόδους σε όλες τις θέσεις για το 100% και το 50% (σε παρένθεση) του δειγματικού πλήθους των ανεμολογικών δεδομένων.

Συνοπτικά, μελετώντας τους **Πίνακας 4. 8 – Πίνακας 4. 10** προκύπτει ότι:

1. Η μέθοδος *GMFR* παρουσιάζει συστηματικά τις καλύτερες τιμές των δεικτών για όλες τις μεθόδους παλινδρόμησης, σε όλες τις θέσεις και για τους δύο τρόπους ελέγχου της αποδοτικότητας της εξίσωσης διόρθωσης που εξετάζονται.
2. Οι τιμές των δεικτών για τις ανθεκτικές μεθόδους είναι πολύ κοντά μεταξύ τους σε κάθε θέση. Επίσης, οι τιμές που προκύπτουν από όλα τα δεδομένα είναι σχετικά καλύτερες από εκείνες του μισού δείγματος.
3. Αναφορικά με τις θέσεις, η *Mykonos* παρουσιάζει τις καλύτερες τιμές των δεικτών για όλες τις μεθόδους που εφαρμόστηκαν.

## ΚΕΦΑΛΑΙΟ 5 – ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

### 5.1. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία γίνεται, για πρώτη φορά, η συγκριτική μελέτη διαφόρων μεθόδων παλινδρόμησης σε συνταυτισμένα ανεμολογικά δεδομένα από διαφορετικές πηγές σε πέντε περιοχές μελέτης στο Αιγαίο Πέλαγος. Τα συνταυτισμένα ανεμολογικά δεδομένα που χρησιμοποιήθηκαν βασίζονται σε επιτόπιες μετρήσεις ανέμου από το δίκτυο παρακολούθησης *POSEIDON* και σε δεδομένα αριθμητικής προσομοίωσης, είναι υπό τη μορφή μακροχρόνιων χρονοσειρών και καλύπτουν μια περίοδο 5 ετών (2000 – 2004).

Η απλή γραμμική παλινδρόμηση, κυρίως μέσω της μεθόδου των ελαχίστων τετραγώνων, θεωρείται η πιο κλασσική μέθοδος για τη γραμμική περιγραφή της εξαρτημένης μεταβλητής λόγω της ευκολίας στον υπολογισμό των άγνωστων παραμέτρων. Δεδομένου ότι τόσο ο εκτιμητής όσο και η εκτιμώμενη ποσότητα είναι τυχαίες μεταβλητές που περιέχουν σφάλματα στις μετρήσεις εφαρμόστηκε η μέθοδος EIV. Οι δύο πιο ευρέως χρησιμοποιούμενες περιπτώσεις αυτής της μεθόδου, που εφαρμόστηκαν και σε αυτή την εργασία, είναι η παλινδρόμηση ορθογώνιας απόστασης και η συναρτησιακή σχέση γεωμετρικού μέσου. Όμως αναφέρθηκε ότι εξαιτίας της φύσης των μετρήσεων υπεισέρχονται έκτροπες παρατηρήσεις και στις δύο μεταβλητές, οι οποίες θα πρέπει να αντιμετωπιστούν κατάλληλα, αφού πρώτα εντοπιστούν και ερμηνευθούν. Για το λόγο αυτό είναι αναγκαίο να χρησιμοποιηθεί μια πιο αποτελεσματική μέθοδος που να λαμβάνει υπόψη τις εν λόγω παρατηρήσεις, αλλά και τυχόν αποκλίσεις από τις υποθέσεις της γραμμικής παλινδρόμησης, όπως είναι οι ανθεκτικές μέθοδοι.

Από τη στατιστική ανάλυση των μετρήσεων και από τον τρόπο αξιολόγησης της παλινδρόμησης και της εξίσωσης διόρθωσης προκύπτουν τα ακόλουθα συμπεράσματα:

1. **Για την εξίσωση παλινδρόμησης**, οι μέθοδοι *OLS* και οι *IRLS* και *MM* παρουσιάζουν τη καλύτερη συμπεριφορά (απόδοση) αναφορικά με το σύνολο των στατιστικών δεικτών αξιολόγησης που εφαρμόστηκαν σε όλες τις θέσεις πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων αντίστοιχα. Επίσης, για όλους τους δείκτες, το μεγαλύτερο εύρος τιμών δίνεται από τις μεθόδους *GMFR* και *S* συγκριτικά με τις υπόλοιπες μεθόδους παλινδρόμησης και πριν και μετά την αφαίρεση των έκτροπων παρατηρήσεων. Μετά την αφαίρεση των έκτροπων παρατηρήσεων, οι τιμές όλων των στατιστικών δεικτών που προκύπτουν από την εφαρμογή της μεθόδου *GMFR* βελτιώνονται αισθητά σε σύγκριση με τις άλλες δύο κλασσικές μεθόδους, ενώ οι τιμές των νέων συντελεστών παλινδρόμησης της *OLS* προσεγγίζουν αυτές των ανθεκτικών μεθόδων.
2. **Για την εξίσωση διόρθωσης**, η μέθοδος που παρουσιάζει συστηματικά τις καλύτερες τιμές σε όλες τις θέσεις μελέτης, για όλους τους στατιστικούς δείκτες που εφαρμόστηκαν τόσο για το 100% όσο και για το 50% του δειγματικού πλήθους των συνταυτισμένων ανεμολογικών δεδομένων είναι η *GMFR*. Το γεγονός ότι η μέθοδος *GMFR* είναι αυτή που συστηματικά προσαρμόζεται καλύτερα σε νέα δεδομένα αποτελεί ένα βασικό και πρωτότυπο αποτέλεσμα της παρούσας εργασίας.
3. Στους δύο ελέγχους παρατηρείται ότι οι τιμές των στατιστικών μεγεθών και των συντελεστών παλινδρόμησης όπως υπολογίζονται από τις ανθεκτικές μεθόδους είναι πολύ κοντά μεταξύ τους σε κάθε θέση.

## 5.2. ΠΡΟΤΑΣΕΙΣ

Μέσα από αυτή την εργασία, προκύπτουν νέες μέθοδοι για τη διόρθωση ανεμολογικών δεδομένων που προέρχονται από αριθμητικά μοντέλα προσομοίωσης, καθώς και νέοι τρόποι προσέγγισης στην επεξεργασία των ανεμολογικών δεδομένων. Βάσει των αποτελεσμάτων της συγκριτικής μελέτης που παρουσιάστηκε στα προηγούμενα εδάφια, προκύπτει ότι ανάλογα με το σκοπό της εφαρμογής που επιδιώκεται, αντίστοιχη πρέπει να είναι και η επιλογή της καταλληλότερης μεθόδου παλινδρόμησης.

Τέλος, ορισμένα νέα ερωτήματα που εγείρονται προς περαιτέρω έρευνα και συζήτηση αφορούν: i) τη διερεύνηση των έκτροπων παρατηρήσεων, όπως, για παράδειγμα την επέμβαση του χρήστη στα αποτελέσματα των έκτροπων παρατηρήσεων και την ερμηνεία της προέλευσής τους, και ii) την εξέταση και εκτίμηση της ποιότητας και αρτιότητας των παραμέτρων του μοντέλου παλινδρόμησης στην προσαρμογή των δεδομένων μέσω γραφικών διαγνωστικών εργαλείων, (π.χ., γράφημα των καταλοίπων). Η συγκεκριμένη στατιστική ανάλυση μπορεί επίσης να πραγματοποιηθεί και για τα ανεμολογικά δεδομένα που προέρχονται από παρατηρήσεις τηλεανίχνευσης.



## 6. ΒΙΒΛΙΟΓΡΑΦΙΑ

### Ελληνική Βιβλιογραφία

1. Αβραμίδης, Α. (2010). *Ανθεκτική Παλινδρόμηση με Μαθηματικό Προγραμματισμό*, Διδακτορική Διατριβή, ΑΠΘ, Θεσσαλονίκη.
2. Ζαχαροπούλου, Χ. (1995). *Παλινδρόμηση – Συσχέτιση: Θεωρία και Πράξη*, Θεσσαλονίκη.
3. Καραθανάση, Β. (2006). *Μικροκυματική Τηλεπισκόπηση*, Εργαστήριο Τηλεπισκόπησης, ΕΜΠ, Αθήνα.
4. Κιοκός, Γ. (2004). *Μελέτη Ανάπτυξης Ραντάρ Συνθετικής Απεικόνισης Μικρών Διαστάσεων*, Διπλωματική εργασία, ΕΜΠ, Αθήνα.
5. Νίττης, Κ. Α, Χρόνης, Γ. Θ. (2003). *Σημειώσεις Επιχειρησιακή Ωκεανογραφία*. Τμήμα Επιστήμης της Θάλασσας, Πανεπιστήμιο Αιγαίου.
6. Παπανικολάου, Θ. (2010). *Μελέτη ανθρωπογενούς παρέμβασης σε περιοχή “NATURA” με χρήση φωτογρ/τρίας, τηλεπισκόπησης και Γεωγρ. Συστ. Πληροφοριών – ΔΕΛΤΑ ΑΧΕΛΩΟΥ*, Μεταπτυχιακή εργασία, ΕΜΠ, Αθήνα.
7. Σακελλαρίδης, Γ., Κωστόπουλος, Β., Χαραντώνης, Θ. (2010). *Η μεθοδολογία υπολογισμού του ανέμου για τη δημιουργία βιώσιμων αιολικών πάρκων*, Συνέδριο ΤΕΕ με θέμα «Ενέργεια: Σημερινή εικόνα – Σχεδιασμός – Προοπτικές», Αθήνα.
8. Σουκισιάν, Τ., Χατζηνάκη, Μ., Κορρές, Γ., Παπαδόπουλος, Α., Κάλλος, Γ., Αναδρανιστάκης, Ε. (2007). *Άτλας Άνεμου και Κύματος των Ελληνικών Θαλασσών*, Εκδόσεις Ελληνικό Κέντρο Θαλασσίων Ερευνών, Αθήνα, 300 σελ.
9. Φωτιάδης, Μ. Α. (2011). *Ανάπτυξη λογισμικού για την ανάλυση δεδομένων συσκευών μέτρησης ανέμου τύπου LIDAR*, Διπλωματική εργασία, ΕΜΠ, Αθήνα.
10. Χρήστου, Γ. Κ. (2006). *Εισαγωγή στην Οικονομετρία*, Τόμος Α', Β' Έκδοση, Gutenberg, Αθήνα, ISBN 960-01-0954-0.

### Ξενόγλωσση Βιβλιογραφία

11. Andersen, R. (2008). *Modern Methods for Robust Regression*, Thousand Oaks: SAGE Publications.
12. Barnett, V., Lewis, T. (1994). *Outliers in Statistical Data*, Third Edition, John Wiley & Sons Ltd, Baffins Lane, Chichester, West Sussex, England, ISBN 0-471-93094-6.
13. Barthelmie, R. J. (2001b). *Modelling and measurements of coastal wind speeds, in Renewable Energy: Trends and Prospects*, Academy of Science, Easton, PA.
14. Belsley, D.A., Kuh, E., Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: Wiley.
15. Benticha, M. B., Thierry, R., Lucien, W. (2005). *Using several data sources for offshore wind resource assessment*, Copenhagen Offshore Wind.
16. Coakley, C., Hettmansperger, T. (1993). *A bounded influence, high breakdown, efficient regression estimator*. Journal of the American Statistical Association, 88:872–880.
17. De Luca, G.F., Marano, G., Piemontese, M., Versini, B., Caltagirone, F., Casonato, G., Coletta, A., De Carlo, M. (2007). *Interoperability, expandability and multi mission-sensor COSMOSkyMed capabilities*. In: Proc. International Geoscience and Remote Sensing Symposium (IGARSS 2007), Barcelona, Spain.
18. De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L. (2000), *The Mahalanobis Distance*, Chemometrics and Intelligent Laboratory Systems, Volume 50, Elsevier, pp.1-18.

19. Donoho, D., Huber, P. (1983). *The notion of breakdown point*, In A Festschrift for Erich Lehmann, pages 157–184, Belmont, CA. Wadsworth.
20. Draper, N. R., Smith, K. (1998). *Applied Regression Analysis*, Third edition. New York: Wiley.
21. Eineder, M., Breit, H., Fritz, T., Schuettler, B., Roth, A. (2005). *TerraSAR-X SAR products and processing algorithms*, In: Proc. International Geoscience and Remote Sensing Symposium (IGARSS 2005), Seoul, Korea.
22. Elachi, C., Cimino, J. B., Settle, M. (1986). *Overview of the Shuttle Imaging Radar – B: Preliminary Scientific Results*. Science, Vol. 232, pp. 1511-1516.
23. Elliot, D, Frame, C., Gill, C., Hanson, H., Moriarty, P., Powell, M., Shaw, W. J., Wilczak, J., Wynne, J. (2012). *Offshore Resource Assessment and Design Conditions: A Data Requirements and Gaps Analysis for Offshore Renewable Energy Systems*, U.S. Department of Energy, Wind and Water Power Program.
24. Foley, A. M., Leahy, P. G., Marvuglia, A., and McKeogh, E. J. (2012). *Current methods and advances in forecasting of wind power generation*, Renewable Energy, Elsevier, 1-8.
25. Glahn, D. A., and Lowry, D. A. (1972). *The Use of Model Output Statistics (MOS) in Objective Weather Forecasting*, 1203-1211
26. Hadi, A. S., Simonoff, J. (1993). *Procedures for the identification of multiple outliers in linear models*. Journal of the American Statistical Association, 88:1264–1272.
27. Hampel, F.R. (1973). *Robust Estimation: a Condensed Partial Survey*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 27, 87-104.
28. Handschin, E., Schweppe, F., Kohlas, J., Fiechter, A. (1975). *Bad data analysis for power system state estimation*. IEEE Transactions of Power Apparatus and Systems, PAS-94:329-337.
29. Hasager, C. B., Peña, A., Christiansen, M. B., Astrup P., Nielsen, M., Monaldo F., Thompson, D., Nielsen, P. (2008). *Remote Sensing Observation Used in Offshore Wind Energy*, IEEE Journal of selected topics in applied earth observations and remote sensing, Vol. 1, No. 1.
30. Hekimoglu, S., Erenoglou, R.C., Kalina, J. (2006). *Outlier detection by means of robust regression estimators for use in engineering science*, Journal of Zhejiang University, 909-921.
31. Hoaglin, D., Welsch, R. (1978). *The hat matrix in regression and ANOVA*. The American Statistician, 32:17–22.
32. Huber, P. (1981). *Robust Statistics*. John Wiley, New York.
33. Huber, P J, Dutter, R. (1974). *Numerical solutions of robust regression problems*, Proceedings in Computational Statistics. Vienna: Physika Verlag, 1974, 165–172.
34. Huber, P. J., Ronchetti, E. M. (2009). *Robust Statistics*, Second edition. Hoboken: Wiley.
35. Idowu, O. S., and Rautenbach C. J. de W (2009). *Model Output Statistics to improve severe storms prediction over Western Sahel*, Atmospheric Research, Elsevier, 419-425.
36. Kaiser, M.J., Snyder, B.F. (2012). *Offshore Wind Energy Cost Modeling*, Green Energy Technology, Springer-Verlag, London.
37. Kallos, G. (1997). *The Regional weather forecasting system SKIRON*. Proceedings, Symposium on Regional Weather Prediction on Parallel Computer Environments, pp.9, 15-17, Athens, Greece.
38. Kendall, M. G., Stuart, A. (1961). *The Advanced Theory of Statistics*, Volume 2, Charles Griffin & Company Limited, London.
39. Kent, E.C., Taylor, P. K., Truscott, B. S., Hopkins, J. S. (1993). *The accuracy of voluntary observing ships' meteorological observations – Results of the VSOP-NA.*, J. Atmos. Oceanic Technol., 10, 591-608.

40. Lame, D. B., Born, G. H. (1982). *SEASAT Measurement System Evaluation: Achievements and Limitations*, J. Geophys. Res., 87(C5), 3175–3178.
41. Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., and Yan, Z. (2009). *A review on the forecasting of wind speed and generated power*, Renewable and Sustainable Energy Reviews, Elsevier, 915-920.
42. Leng, L., Zhang, T., Kleinman, L., Zhu, W. (2007). *Ordinary Least Square Regression, Orthogonal Regression, Geometric Mean Regression and their Applications in Aerosol Science*, Journal of Physics, Conference Series 78.
43. Mallows, C. (1975). *On some topics in robustness*, unpublished memorandum. Murray Hill, New Jersey.
44. Maronna, R. A., Martin, R. D., Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons, Ltd, ISBN: 0-470-01092-4.
45. Montgomery, D. C., Peck, E. A. (1992). *Introduction to Linear Regression Analysis*, Second Edition, John Wiley & Sons, Inc., USA, ISBN 0-471-53387-4.
46. Mortensen, N. G., Heathfield, D. N., Myllerup, L., Landberg, L., Rathmann, O. (2005). *Wind Atlas Analysis and Application Program: WASP 8 Help Facility*, Risø National Laboratory, Roskilde, Denmark. 335 topics. ISBN 87-550-3457-8, DOI, <http://www.risoe.dk/vea/projects/nimo/WAsPHelp/Wasp8.htm>.
47. Nelson, V. (2009). *Wind energy, renewable energy and the environment*, CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, ISBN 978-1-4200-7568-7.
48. Nemoto, Y., Nishino, H., Ono, M., Mizutamari, H., Nishikawa, K., Tanaka, K. (1991). *Japanese Earth Resources Satellite-1 Synthetic Aperture Radar*. Proc. IEEE, 79, 800-809.
49. Nishidai, T., Yoshie, T., Tsu, H. (1994). *Overall results of the Earth Resources Satellite 'Fuyo-T' data evaluation*. Proceedings of the Result Reporting Meeting of JERS-I System Verification Program. Tokyo. Japan. November 1994 (Tokyo: NASDA), pp,31-39.
50. Raney, R. K., Luscombe, A. P., Langham, E. J., Ahmed, S. (1991). *RADARSAT*. Proceedings of the IEEE. 79, pp. 839-849.
51. Riggs, D. S., Guarnieri, J. A., Addelman, S. (1978). *Fitting straight lines when both variables are subject to error*, Life Science, Vol. 22, pp.1305-1360.
52. Rodrigo, J. S. (2010), *State-of-the-Art Wind Resource Assessment*, WAUDIT.
53. Rousseeuw, P. J., Croux, C. (1993). *Alternatives to the Median Absolute Deviation*, Journal of the American Statistical Association (American Statistical Association) 88 (424): 1273–1283.
54. Rousseeuw, P. J., Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, John Wiley & Sons, Inc., USA, ISBN 0-471-85233-3.
55. Ryan, T. P. (1997). *Modern Regression Methods*, John Wiley & Sons, Inc., 605 Third Avenue, NY, USA, ISBN 0-471-52912-5.
56. Schöne, T., Eickschen, S. *Wind Speed and SWH Calibration for Radar Altimetry in the North Sea*, Potsdam, Germany.
57. Sempreviva, A. M., Barthelmie, R., Giebel, G., Lange, B., Sood, A. (2007). *Offshore wind resource assessment in European Seas, state-of- the –art. A survey within the FP6 "POW'WOW" Coordination Action Project*.
58. Soukissian, T., Kechris, C. (2006). *About applying linear structural method on ocean data: Adjustment of satellite wave data*, Ocean Engineering, Elsevier, pp. 371-389.
59. Soukissian, T., Papadopoulos, A, (2013). *Effects of different wind data sources in offshore wind power assessment*, submitted for publication in Renewable Energy.
60. Stuart, C. (2011). *Robust Regression*, Department of Mathematical Sciences, Durham University.

61. Thompson, A.A., Luscombe, A., James, K., Fox, P. (2008). *RADARSAT-2 mission status: capabilities demonstrated and image quality achieved*. In: Proc. 7th European Conference on SAR (EUSAR 2008), Friedrichshafen, Germany.
62. Yohai, V. J. (1987). *High Breakdown-Point and High Efficiency Robust Estimates for Regression*, The Annals of Statistics, Vol. 15, No. 2, pp. 642-656.

### Ιστοσελίδες

63. <http://www.windtech-international.com>
64. [www.meteo.gr](http://www.meteo.gr)
65. <http://earth.eo.esa.int/trainingcourses/EOSummerSchool2012>