# Εθνικο Μετσοβιο Πολυτεχνειο

## Σχολη Πολιτικων Μηχανικων

### Τομεας Υδατικων Πορων και Περιβαλλοντος


# Βέλτιστη συμπλήρωση ελλιπών υδρομετεωρολογικών δεδομένων με χρήση γειτονικών χρονικά παρατηρήσεων


Διπλωματική Εργασία


Παππάς Χριστόφορος

Επιβλέπων Καθηγητής

Κουτσογιάννης Δημήτρης



Αθήνα, Οκτώβριος 2010

***"Ακόμα και αν οι φυσικοί νόμοι δεν είχαν άλλα μυστικά από εμάς, θα μπορούσαμε να ξέρουμε την αρχική κατάσταση μόνο κατά προσέγγιση. Αν αυτό μας επιτρέπει να προβλέψουμε τη μεταγενέστερη κατάσταση με τον ίδιο βαθμό προσέγγισης, αυτό αρκεί να πούμε ότι το φαινόμενο είχε προβλεφθεί, ότι υπόκειται σε νόμους. Όμως, το ζήτημα δεν είναι πάντοτε έτσι: υπάρχει περίπτωση οι πολύ λεπτές διαφορές στις αρχικές συνθήκες να παράγουν πολύ μεγαλύτερες διαφορές στα τελικά φαινόμενα, ένα ελάχιστο σφάλμα στην αρχή να προκαλεί ένα τεράστιο σφάλμα στο τέλος. Η πρόβλεψη τότε γίνεται αδύνατη κι έτσι έχουμε το φαινόμενο της τύχης."*** *Henri Poincare*

Το σκίτσο του εξωφύλλου χρησιμοποιήθηκε μετά από έγγραφη άδεια του σκιτσογράφου Ηλία Μακρή

# ΕΥΧΑΡΙΣΤΙΕΣ

Φτάνοντας στο τέλος αυτής της εργασίας, που σηματοδοτεί παράλληλα και την ολοκλήρωση των σπουδών μου στη Σχολή Πολιτικών Μηχανικών, νιώθω την ανάγκη να ευχαριστήσω τους ανθρώπους που με βοήθησαν και με στήριξαν στην προσπάθεια αυτή.

Πρώτα απ' όλα πρέπει να ευχαριστήσω την οικογένεια μου, τον πατέρα μου και τον αδερφό μου, που στήριξαν και συνεχίζουν να στηρίζουν τις επιλογές μου, χωρίς ενδοιασμούς.

Θερμά επίσης θέλω να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Δημήτρη Κουτσογιάννη, για την πολύτιμη βοήθειά του στην εκπόνηση της παρούσας εργασίας. Θα ήθελα επίσης να τον ευχαριστήσω για τις γνώσεις και το διαφορετικό τρόπο σκέψης που μου μετέδωσε καθώς και για την εμπιστοσύνη που μου έδειξε, ανοίγοντας απλόχερα νέους ορίζοντες για τη συνέχιση των σπουδών μου.

Επίσης, ένα μεγάλο ευχαριστώ στον φίλο, υποψήφιο διδάκτορα, Σίμωνα Παπαλεξίου για τη στήριξή του σε κρίσιμες χρονικές στιγμές, όταν η ολοκλήρωση της εργασίας φάνταζε στα μάτια μου αδύνατη. Τον ευχαριστώ πολύ επίσης για τις γνώσεις που μου μετέδωσε και για την άψογη συνεργασία που είχαμε. Η συμβολή του ήταν πολύτιμη για την ολοκλήρωση αυτής της εργασίας.

Θα ήταν παράληψη να μην ευχαριστήσω τους φίλους και συμφοιτητές μου, Αρχοντία, Γιάννη, Κυριάκο και Παναγιώτη, αφού το θέμα μιας κοινής μας παρουσίασης στο συνέδριο του E.G.U στη Βιέννη αποτέλεσε την αφορμή για την παρούσα διπλωματική εργασία.

# ΠΕΡΙΕΧΟΜΕΝΑ

## ΠΕΡΙΛΗΨΗ

Η κατανόηση των υδρομετεωρολογικών φαινομένων βασίζεται σε μεγάλο βαθμό στη μελέτη και επεξεργασία των υπαρχουσών παρατηρήσεων (χρονοσειρών). Πολύ συχνά όμως, οι χρονοσειρές παρουσιάζουν κενά, δηλαδή για κάποιες χρονικές περιόδους δεν υπάρχουν μετρήσεις. Η συμπλήρωση των ελλιπών τιμών είναι απαραίτητη για την περαιτέρω διερεύνηση των φυσικών φαινομένων. Υπάρχει πληθώρα μεθοδολογιών για τη συμπλήρωση αυτών των κενών στη βιβλιογραφία, ωστόσο οι περισσότερες μπορούν να εκφραστούν με μια απλή γραμμική σχέση. Η συμπλήρωση των κενών μιας χρονοσειράς μπορεί να βασίζεται σε μετρήσεις γειτονικών σταθμών ή ακόμη και μόνο στις μετρήσεις της ελλιπούς χρονοσειράς.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η μελέτη αυτού του συνήθους προβλήματος, της συμπλήρωσης δηλαδή των ελλιπών υδρομετεωρολογικών δεδομένων. Πιο συγκεκριμένα, μελετάται το πρόβλημα της ύπαρξης μεμονωμένων ή ακόμα και σποραδικών κενών στις χρονοσειρές και εξετάζεται η συμπλήρωσή τους με χρήση των γειτονικών χρονικά παρατηρήσεων. Ανάλογα με τη δομή της αυτοσυσχέτισης της υπό εξέταση χρονοσειράς, διερευνάται αν η χρήση ενός σταθμισμένου τοπικού μέσου όρου είναι κατάλληλη και κάτω από ποιες προϋποθέσεις. Μελετώνται δύο τύποι δομών αυτοσυσχέτισης (εκθετικής μορφής και αυτοσυσχέτιση μορφής δύναμης). Τα αποτελέσματα που προκύπτουν είναι ιδιαίτερα ικανοποιητικά και η προτεινόμενη μεθοδολογία ενδείκνυται για τη γρήγορή και άμεση συμπλήρωση μικρού πλήθους ελλιπών μετρήσεων.

## ABSTRACT

The understanding of hydrometeorological phenomena is based on the study of existing observations (time series). Often, the time series have gaps, in other words, there are some periods with no measurements. Filling the missing values is necessary for further investigation of natural phenomena. In literature, there are many methodologies for infilling those gaps, but most of them, can expressed in a simple linear relationship. The interpolation can be based either on measurements of neighbouring gauges or on measurements of neighbouring time steps.

In this diploma thesis, we study the problem of interpolation in time-series with sporadic gaps and we examine the interpolation by using neighboring observation in time (local average). Depending on the autocorrelation structure of time-series, we investigated whether the use of a weighted local average is appropriate for the infilling. Assuming that the underlying hydrometeorological process behaves like either a Markovian or a Hurst-Kolmogorov process we estimate the missing values using different techniques based on a weighted local average. In each of the cases we determine the unknown quantities so as to minimize the estimation mean square error. The results are very satisfactory and the proposed methodology is appropriate for quick and immediate filling of a small number of missing measurements.

# ΚΕΦΑΛΑΙΟ 1$^{ο}$

## 1 ΕΙΣΑΓΩΓΗ

### 1.1 Η σημασία της γνώσης του παρελθόντος στην υδρολογία: «μια μέτρηση=1000 υπολογισμοί»

Οι θετικές επιστήμες έχοντας ως κύριο αντικείμενο την μελέτη φυσικών διεργασιών, βασίζονται στην παρατήρηση των φυσικών φαινομένων. Ο κλάδος της υδρολογίας επίσης, στηρίζεται στην παρατήρηση των φυσικών-υδρολογικών διεργασιών (βροχόπτωση, εξάτμιση, διαπνοή κτλ.). Ο λόγος για τον οποίο η υδρολογία και γενικότερα οι γεωφυσικές επιστήμες στηρίχτηκαν σε μεγάλο βαθμό στην παρατήρηση των φυσικών φαινομένων και στη μελέτη και επεξεργασία των υπαρχουσών μετρήσεων είναι το γεγονός ότι οι απλοί νόμοι της κλασικής μηχανικής (π.χ. διατήρηση μάζας, ορμής και ενέργειας) αδυνατούν να περιγράψουν ικανοποιητικά το χαοτικό χαρακτήρα των υδρολογικών συστημάτων (υδρολογικές λεκάνες ποταμοί, λίμνες, ατμόσφαιρα). Συνεπώς, η μελέτη της συμπεριφοράς ενός υδρολογικού συστήματος στο παρελθόν αποτελεί μονόδρομο για την κατανόηση της σημερινής και της μελλοντικής του συμπεριφοράς.

Ειδικότερα, όταν μελετάται ένα έργο που στοχεύει στην ανάπτυξη και αξιοποίηση των υδατικών πόρων ή στην προστασία από τις πλημμύρες, το βασικότερο στοιχείο που καθορίζει το σχεδιασμό του έργου είναι η ποσότητα του νερού που μπορεί να αξιοποιηθεί ή το μέγεθος της πλημμύρας που καλείται να αντιμετωπιστεί. Επίσης, στην περίπτωση μελέτης ενός συστήματος έργων, υδρευτικών, αρδευτικών, αντιπλημμυρικών ή υδροηλεκτρικών, πάλι θα πρέπει να γνωρίζουμε τη χρονική διακύμανση των ποσοτήτων νερού. Ο σωστός λοιπόν σχεδιασμός ενός υδραυλικού έργου που θα οδηγήσει στη βέλτιστη λύση, προϋποθέτει την άριστη γνώση των υδρολογικών φαινομένων που επικρατούν στη συγκεκριμένη περιοχή ώστε να είναι δυνατή η ακριβέστερη πρόβλεψη της εξέλιξης των φυσικών διεργασιών. Οι μετρήσεις όμως εκτός από την άμεση και προφανή χρησιμότητά τους στον προγραμματισμό, στο σχεδιασμό και στη λειτουργία των έργων έχουν και μια ευρύτερη επιστημονική χρησιμότητα: την κατανόηση των φυσικών φαινομένων και των μηχανισμών που τα διέπουν (Κουτσογιάννης, 2000).

Συνεπώς, βασικός σκοπός της επιστήμης της υδρολογίας είναι οι κατανόηση και η προσομοίωση των υδρολογικών διεργασιών. Η προσομοίωση είναι η τεχνική της μίμησης ενός πραγματικού συστήματος όπως αυτό εξελίσσεται στο χρόνο. Στόχος έτσι της υδρολογίας είναι η σύνθεση μοντέλων που να μπορούν να περιγράφουν-

προσομοιώνουν με μεγάλη ακρίβεια τις διάφορες φυσικές διεργασίες με απώτερο στόχο την εκτίμηση ή ορθότερα την πρόβλεψη[2] των μελλοντικών τιμών αυτών των διεργασιών. Η σωστή πρόβλεψη των φυσικών διεργασιών, ενισχύει την ασφάλεια των τεχνικών έργων (πρόβλεψη ακραίων γεγονότων) αλλά ταυτόχρονα καθιστά και οικονομικότερο το σχεδιασμό του έργου (σωστή διαστασιολόγηση των κατασκευών).

Η μελέτη λοιπόν των φυσικών αυτών διεργασιών και η κατανόηση τους βασίζεται σε μετρήσεις που γίνονται, σε διάφορες χρονικές κλίμακες, σε ειδικούς σταθμούς, με τις απαραίτητες υποδομές, εξοπλισμένους με κατάλληλα όργανα[3] για τη μέτρηση της βροχόπτωσης και άλλων μετεωρολογικών συνθηκών (μετεωρολογικοί και βροχομετρικοί σταθμοί). Για τη μελέτη και την ορθή καταγραφή της κάθε φυσικής διεργασίας χρησιμοποιούνται διάφορα όργανα[4]. Οι νέες τεχνολογίες μετρήσεων που έχουν αναπτυχθεί (αυτόματοι ψηφιακοί αισθητήρες μετρήσεων, τηλεμετρία, μετεωρολογικά ραντάρ κ.ά.) καθώς επίσης και το έμψυχο δυναμικό των αντίστοιχων σταθμών (καταρτισμένοι παρατηρητές, εκπαιδευμένα συνεργεία) βοηθούν ακόμη περισσότερο στην ακριβέστερη παρατήρηση των φυσικών διεργασιών. Η παρατήρηση αυτή των φυσικών φαινομένων και η γνώση της χρονικής τους εξέλιξης έχει ως αποτέλεσμα τη δημιουργία χρονοσειρών με μετρήσεις.

Με τον όρο χρονοσειρά, εννοούμε ένα σύνολο από παρατηρήσεις οι οποίες αναφέρονται σε κάποιο συγκεκριμένο φυσικό μέγεθος και λαμβάνονται σε ορισμένες χρονικές στιγμές ή περιόδους οι οποίες ισαπέχουν μεταξύ τους. Λεπτομερέστερος ορισμός της έννοιας της χρονοσειράς και οι διάφορες κατηγορίες χρονοσειρών που παρουσιάζονται στην υδρολογία δίνονται στο υποκεφάλαιο 2.3.

## 1.2 Το πρόβλημα των ελλιπών δεδομένων στις υδρομετεωρολογικές χρονοσειρές και η αναγκαιότητα συμπλήρωσής τους

Η περιγραφή των υδρολογικών διεργασιών, όπως ήδη αναφέραμε, βασίζεται στη μελέτη και επεξεργασία των υπαρχουσών χρονοσειρών. Η επεξεργασία αυτών των χρονοσειρών οδηγεί στη δημιουργία μοντέλων που αναπαράγουν τα στατιστικά

[2]Το σημαντικότερο αντικείμενο της στατιστικής είναι η εκτίμηση. Η εκτίμηση διακρίνεται σε εκτίμηση των παραμέτρων και την πρόγνωση. Τα δύο αυτά προβλήματα (εκτίμηση παραμέτρων και πρόγνωση) αντιμετωπίζονται με παρόμοιες μεθόδους στη στατιστική και για το λόγο αυτό αναφέρονται με τον όρο εκτίμηση.

[3] Όργανα σημειακής μέτρησης βροχοπτώσεων:

Βροχόμετρα : δίνουν την ολική σημειακή βροχόπτωση και το ισοδύναμο νερού μιας χιονόπτωσης ανά ορισμένα χρονικά διαστήματα, με ανάγνωση της ένδειξης από έναν παρατηρητή.

Βροχογράφοι : καταγράφουν με απλό ωρολογιακό μηχανισμό την μεταβολή του ύψους βροχής στο χρόνο, περιγράφοντας έτσι τη χρονική κατανομή της σημειακής βροχόπτωσης (Κουτσογιάννης & Ξανθόπουλος, 1999 σ. 95).

[4] Ακτινόμετρα, ανεμόμετρα, βαρόμετρα, βροχογράφοι, βροχόμετρα, εξατμησίμετρα, θερμόμετρα, υγρογράφοι κ.α.

χαρακτηριστικά των παρατηρήσεων με αποτέλεσμα την περιγραφή της εκάστοτε φυσικής διεργασίας. Με την κατάλληλη επεξεργασία των χρονοσειρών μπορούμε να εκτιμήσουμε τις αντίστοιχες στοχαστικές ανελίξεις[5] που αντιπροσωπεύουν τις συγκεκριμένες διεργασίες.

Συμπεραίνομε πως το πρόβλημα της περιγραφής μιας φυσικής διεργασίας σχετίζεται άμεσα με την επεξεργασία και τη μελέτη των χρονοσειρών. Είναι λοιπόν προφανές πως όσο πιο αξιόπιστες είναι οι τιμές[6] της χρονοσειράς που εξετάζουμε και όσο πιο μεγάλο είναι το μέγεθος του υδρολογικού δείγματος τόσο πιο ολοκληρωμένη εικόνα θα αποκτήσουμε για την αντίστοιχη φυσική διεργασία και τόσο πιο αξιόπιστες θα είναι και οι εκτιμήσεις και οι προβλέψεις μας.

Ένα από τα συνηθέστερα προβλήματα το οποίο καλείται να αντιμετωπίσει όποιος ασχολείται με την επιστήμη της υδρολογίας είναι η έλλειψη (συστηματική ή μη) κάποιων μετρήσεων. Η έλλειψη αυτή μπορεί να οφείλεται σε διάφορους παράγοντες, ανθρωπογενείς ή μη, τυχαίους ή προκαθορισμένους. Πιο συγκεκριμένα, η έλλειψη μετρήσεων για συνεχείς χρονικές περιόδους μπορεί να οφείλεται σε κάποια φθορά του οργάνου μέτρησης ή σε κάποια διακοπή λειτουργίας λόγω ακραίων συνθηκών (π.χ. ακραία καιρικά φαινόμενα ή ακραίοι κοινωνικοοικονομικοί παράγοντες). Πιθανή βλάβη των οργάνων μπορεί να οδηγήσει σε συστηματικά σφάλματα των μετρήσεων με αποτέλεσμα λάθος εκτίμηση των φυσικών διεργασιών. Έτσι οι εσφαλμένες αυτές τιμές, δεν λαμβάνονται υπόψη ή διορθώνονται κατάλληλα. Λάθη στο χειρισμό των οργάνων από το αρμόδιο προσωπικό, μπορούν να οδηγήσουν επίσης σε ελλιπείς μετρήσεις. Όλοι οι παραπάνω λόγοι έχουν ως αποτέλεσμα οι χρονοσειρές να έχουν πολλά κενά. Τα κενά αυτά μπορεί να είναι συγκεντρωμένα σε κάποια σημεία ή και να παρουσιάζονται σποραδικά σε όλο το μήκος της χρονοσειράς.

Η συμπλήρωση αυτών των κενών είναι αναγκαία για την περαιτέρω επεξεργασία των χρονοσειρών, όπως για παράδειγμα για την φασματική ανάλυση και για την εκτίμηση της αυτοσυσχέτισης. Επίσης η μετάβαση σε άλλες χρονικές κλίμακες (π.χ. από μία χρονοσειρά ωριαίων μετρήσεων να παράγουμε την αντίστοιχη ημερήσια) δυσχεραίνεται ιδιαίτερα από την παρουσία κενών. Επιπρόσθετα, τα κενά στις χρονοσειρές αλλοιώνουν τα στατιστικά χαρακτηριστικά της διεργασίας με αποτέλεσμα η ανέλιξη που θα παραχθεί να μην είναι αντιπροσωπευτική. Ο κίνδυνος αυτός είναι εντονότερος όταν τα κενά είναι πάρα πολλά συγκριτικά με το μέγεθος του δείγματος. Η παραγωγή συνθετικών χρονοσειρών δυσχεραίνεται επίσης από την παρουσία κενών.

Όλοι οι προαναφερθέντες λόγοι καταδεικνύουν τη σημασία της συμπλήρωσης των ελλιπών τιμών στις χρονοσειρές. Η θεωρία των πιθανοτήτων και η στατιστική (και κυρίως η γεωστατιστική) διαδραματίζουν πολύ σημαντικό ρόλο στην συμπλήρωση αυτών των ελλιπών τιμών. Οι στατιστικές έννοιες και μέθοδοι είναι εργαλεία όχι

---

[5] Πρέπει να υπογραμμιστεί η διαφορά ανάμεσα σε μία στοχαστική ανέλιξη και μία χρονοσειρά, συγκεκριμένα, η χρονοσειρά αποτελεί υλοποίηση μίας στοχαστικής ανέλιξης.

[6] Η αξιοπιστία των μετρήσεων εξαρτάται από πολλούς παράγοντες όπως : καταλληλότητα θέσης μέτρησης, καταλληλότητα και συντήρηση οργάνων, εμπειρία προσωπικού κ.α.

απλά χρήσιμα αλλά και απαραίτητα. Αν μία διεργασία δεν παρουσιάζει καθόλου αβεβαιότητα ή τυχαιότητα, τότε δε χρειάζεται η στατιστική προσέγγιση αφού τα ντετερμινιστικά μοντέλα μπορούν να περιγράψουν το φαινόμενο επακριβώς. Στις διεργασίες όμως που μελετά ο κλάδος της υδρολογίας σχεδόν πάντα υπάρχει ο παράγοντας της αβεβαιότητας ή τυχαιότητας και αυτό καθιστά σημαντική τη γνώση των στατιστικών μεθόδων και μοντέλων. Η αβεβαιότητα που παρουσιάζεται στις υδρολογικές διεργασίες οφείλεται σε τρείς βασικές αιτίες (Hirsch, et al. p. 17.1):

- στην εγγενή αβεβαιότητα και τυχαιότητα των φυσικών διεργασιών (βροχόπτωση, χιονόπτωση, θερμοκρασιακή μεταβολή) αλλά και των παραμέτρων που αποτελούν το υδρολογικό σύστημα (τοπογραφία, υδροφόρος ορίζοντας, χαρακτηριστικά του εδάφους κ.α.)
- στα σφάλματα των δειγμάτων καθώς το μέγεθος τους είναι πολύ μικρό σε σχέση με τον πληθυσμό της αντίστοιχης μεταβλητής που μελετάται που είναι πρακτικώς άπειρος. Ο μικρός όγκος των δειγμάτων έχει ως αποτέλεσμα την ανακριβή περιγραφή των αντίστοιχων διεργασιών
- στην ανεπαρκή κατανόηση των διεργασιών που διέπουν ένα φυσικό φαινόμενο.

Το πρόβλημα λοιπόν που καλείται να επιλύσει όποιος ασχολείται με τη μελέτη και επεξεργασία των χρονοσειρών είναι η εκτίμηση των τιμών που λείπουν. Η εκτίμηση αυτή βασίζεται στις υπάρχουσες μετρήσεις από το συγκεκριμένο σταθμό, αλλά και σε άλλες μετρήσεις που αναφέρονται στην ίδια φυσική διεργασία, από άλλους γειτονικούς σταθμούς.

Η βιβλιογραφία προσφέρει μια τεράστια ποικιλία από μεθόδους για την συμπλήρωση (είτε σε χρονική κλίμακα είτε σε χωρική) των ελλιπών δεδομένων των χρονοσειρών. Ειδικά για τις υδρομετεωρολογικές παρατηρήσεις έχουν αναπτυχθεί διάφορες μεθοδολογίες, μερικές από τις όποιες είναι πολύ απλές αλλά και άλλες πολύ πιο σύνθετες. Αναλυτική περιγραφή των βασικότερων μεθόδων συμπλήρωσης ελλιπών δεδομένων από υδρομετεωρολογικές χρονοσειρές παρουσιάζεται σε επόμενα κεφάλαια. Ενδεικτικά αναφέρονται : η μέθοδος της μέσης τιμής (με χρήση του ολικού ή τοπικού μέσου όρου), του σταθμισμένου μέσου όρου, των υπερετήσιων λόγων, η γραμμική παλινδρόμηση κ.α. Πολύ συχνά, χρησιμοποιούνται και στοχαστικά μοντέλα για την συμπλήρωση των ελλιπών δεδομένων. Ευρέως διαδεδομένη είναι επίσης στην υδρολογία, αλλά και σε άλλες επιστήμες, η μέθοδος kriging ή και παραλλαγές της, καθώς λαμβάνει υπόψη τη χωρική και χρονική συσχέτιση των μετρήσεων. Ωστόσο, η πληθώρα αυτή των μεθόδων, δεν συνεπάγεται δυστυχώς και ποικιλία στον τρόπο προσέγγισης του προβλήματος. Οι περισσότερες μέθοδοι βασίζονται σε ένα ζυγισμένο – σταθμισμένο μέσο όρο των υπαρχουσών παρατηρήσεων (Koutsoyiannis & Langousis, 2010).

Βασικό στοιχείο όμως στην συμπλήρωση των ελλιπών υδρομετεωρολογικών δεδομένων είναι η κατανόηση της αντίστοιχης φυσικής διεργασίας. Δεν είναι όλες οι μέθοδοι κατάλληλες για όλες τις διεργασίες, από την τεράστια αυτή ποικιλία μεθόδων θα πρέπει να επιλέγεται η καταλληλότερη. Επίσης, πρέπει να τονιστεί ότι σε καμία

περίπτωση η θεωρία των πιθανοτήτων και η στατιστική δεν επαρκούν για να αντιμετωπιστεί η έλλειψη της υδρολογικής πληροφορίας. Για το λόγο αυτό στην τεχνική υδρολογία, όπως και σε άλλες επιστήμες του μηχανικού, η επιτυχής αντιμετώπιση των προβλημάτων προϋποθέτει, πέρα από τη γνώση των μαθηματικών εργαλείων, ευρύτητα πνεύματος, εμπειρία, αλλά και φαντασία και διαίσθηση (Κουτσογιάννης, 1996).

## 1.3 Αντικείμενο της εργασίας

Αντικείμενο αυτής της εργασίας είναι η διερεύνηση των διαφόρων μεθόδων που προτείνονται από τη βιβλιογραφία για τη συμπλήρωση των ελλιπών δεδομένων των χρονοσειρών και η διατύπωση μιας εναλλακτικής μεθοδολογίας η οποία βασίζεται σε κατάλληλες παραδοχές λαμβάνοντας υπόψη τα ιδιαίτερα χαρακτηριστικά και το στοχαστικό χαρακτήρα των υδρολογικών διεργασιών.

Πιο συγκεκριμένα, η παρούσα εργασία, ασχολείται με τη συμπλήρωση ελλιπών τιμών στο χρόνο. Αρχικά παρουσιάζονται οι πιο διαδεδομένες μέθοδοι συμπλήρωσης ελλιπών δεδομένων που προτείνονται στη βιβλιογραφία, και στη συνέχεια διατυπώνεται μια εναλλακτική προσέγγιση του προβλήματος. Η προτεινόμενη μεθοδολογία δεν αποτελεί κάποια ιδιαίτερα πολύπλοκη και εξεζητημένη μέθοδο. Αντιθέτως πρόκειται για μια απλή και πολύ εύκολα εφαρμόσιμη μεθοδολογία η οποία χωρίς μεγάλο υπολογιστικό φόρτο κατορθώνει να ελαχιστοποιεί το σφάλμα της εκτίμησης. Η παρούσα μεθοδολογία είναι κατάλληλη για την συμπλήρωση μεμονωμένων κενών στις χρονοσειρές ή και λίγων συνεχόμενων ελλειπουσών τιμών (εξετάζεται η περίπτωση συμπλήρωσης μέχρι και τριών συνεχόμενων τιμών). Η επιτυχία της προτεινόμενης μεθόδου επαληθεύεται μειώνοντας το σφάλμα της εκτίμησης που υπολογίζεται με διάφορες στατιστικές μεθοδολογίες.

Η παρουσία μεμονωμένων κενών στις υδρομετεωρολογικές χρονοσειρές και η ανάγκη συμπλήρωσής τους μόνο από τις μετρήσεις του ίδιου σταθμού (δηλαδή χωρίς τη χρήση μετρήσεων από γειτονικούς σταθμούς) δεν είναι σπάνιο φαινόμενο, και δεν έχει μόνο ακαδημαϊκό ενδιαφέρον. Αντίθετα, παρουσιάζεται πολύ συχνά στη μελέτη παλαιοκλιματικών δεδομένων καθώς επίσης και στη μελέτη δεδομένων από τον 19$^{ο}$ αιώνα ή το πρώτο μισό του 20$^{ου}$ αιώνα, όπου είτε δεν υπήρχαν καθόλου υδρομετεωρολογικοί σταθμοί για να συλλέγουν δεδομένα, είτε αυτοί που υπήρχαν ήταν αρκετά αραιά τοποθετημένοι ώστε να μην είναι δυνατή η συσχέτιση των τιμών του σταθμού που παρουσιάζονται ελλείψεις με κάποιον γειτονικό του.

Στόχος λοιπόν αυτής της εργασίας δεν είναι η επανάληψη ή η σύγκριση των μεθόδων που χρησιμοποιούνται για την πρόβλεψη των ελλιπών τιμών, αλλά η παρουσίαση μιας νέας μεθοδολογίας η οποία βασίζεται σε λογικές παραδοχές και η οποία είναι ευκολότερη υπολογιστικά και ταυτόχρονα βέλτιστη στατιστικά (έχει το μικρότερο σφάλμα εκτίμησης). Η νέα αυτή μέθοδος αποτελεί επέκταση των

υπαρχουσών μεθόδων παλινδρόμησης, στηρίζεται στη μελέτη της δομής αυτοσυσχέτισης που παρουσιάζεται σε κάθε υδρομετεωρολογική ανέλιξη και περιορίζεται στη συμπλήρωση λίγων, μεμονωμένων τιμών.

## 1.4 Διάρθρωση της εργασίας

Η παρούσα εργασία αποτελείται από έξι κεφάλαια:

- Το πρώτο κεφάλαιο αποτελεί μια εισαγωγή στο θέμα της εργασίας. Περιγράφει τη σημασία των μετρήσεων στην υδρολογία καθώς επίσης και το πολύ συχνό πρόβλημα της έλλειψης μετρήσεων από ένα δείγμα και τις αιτίες που οδηγούν σε κενά στις χρονοσειρές.

- Στο δεύτερο κεφάλαιο συνοψίζονται οι βασικές έννοιες των πιθανοτήτων, της στατιστικής και των στοχαστικών ανελίξεων, που χρησιμοποιούνται στη συνέχεια της εργασίας καθώς και οι ιδιαιτερότητες των υδρολογικών χρονοσειρών που πρέπει να λαμβάνονται υπόψη στη μελέτη των υδρολογικών φαινομένων.

- Στο τρίτο κεφάλαιο παρουσιάζονται οι συνήθεις μέθοδοι που χρησιμοποιούνται για τη συμπλήρωση ελλιπών τιμών στις υδρομετεωρολογικές χρονοσειρές καθώς επίσης και οι βασικοί τρόποι αξιολόγησης των διάφορων μεθόδων με βάση διάφορα εργαλεία της στατιστικής.

- Το τέταρτο κεφάλαιο χωρίζεται σε δύο βασικά μέρη: το πρώτο μέρος ασχολείται με την συμπλήρωση μεμονωμένων κενών στις χρονοσειρές και στο δεύτερο διερευνάται η συμπλήρωση σποραδικών ελλιπών τιμών (μέχρι τρείς συνεχόμενες). Όσον αφορά τη συμπλήρωση μεμονωμένων κενών εξετάζεται η χρήση ενός τοπικού μέσου όρου στη θέση του ολικού που συχνά διαισθητικά αλλά ατεκμηρίωτα προτιμάται. Συγκεκριμένα, ανάλογα με τη δομή της αυτοσυσχέτισης της χρονοσειράς, εξετάζεται πότε και κάτω από ποιες προϋποθέσεις είναι καλύτερη η χρήση ενός τοπικού μέσου όρου στη θέση του ολικού. Επίσης, διερευνάται η βελτίωση της εκτίμησης με τη χρήση είτε ενός διευρυμένου τοπικού μέσου όρου, με χρήση δύο γειτονικών μηνιαίων τιμών και των αντίστοιχων γειτονικών ετήσιων είτε με τη χρήση ενός συνδυασμού του τοπικού μέσου όρου με τον αντίστοιχο ολικό με τη χρήση μιας παραμέτρου. Για λόγους πληρότητας εξετάζεται και η περίπτωση συμπλήρωσης με σταθμισμένα βάρη που θεωρητικά αντιστοιχεί στη εφαρμογή μιας μεθόδου BLUE (μέθοδος kriging). Στο

δεύτερο μέρος εξετάζεται η συμπλήρωση συνεχών ελλιπών τιμών (εξετάζεται η περίπτωση τριών συνεχόμενων ελλιπών τιμών) και η μεθοδολογία που προτείνεται είναι η χρήση σταθμισμένων βαρών (μέθοδος kriging) καθώς δεν είναι δυνατή η χρήση ενός τοπικού μέσου όρου γιατί οι αμέσως γειτονικές τιμές λείπουν.

- Στο πέμπτο κεφάλαιο γίνεται μια εφαρμογή των προτεινόμενων μεθοδολογιών σε πραγματικά δεδομένα ώστε να επιβεβαιωθούν τα θεωρητικά αποτελέσματα του τέταρτου κεφαλαίου.

- Στο έκτο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προκύπτουν για τη χρήση των διάφορων μεθόδων. Επιχειρείται επίσης και μια σύγκριση μεταξύ των διάφορων μεθοδολογιών που παρουσιαστήκαν ώστε να συμπεράνουμε ποια μέθοδος δίνει τα βέλτιστα αποτελέσματα.

- Η διπλωματική εργασία συνοδεύεται επίσης από εφτά παραρτήματα τα οποία περιέχουν τις αποδείξεις των μαθηματικών σχέσεων που χρησιμοποιούνται για τον υπολογισμό του μέσου τετραγωνικού σφάλματος κάθε μίας από τις προτεινόμενες μεθοδολογίες καθώς επίσης και πολλά διαγράμματα που παραθέτονται για λόγους πληρότητας και αναφέρονται στις προτεινόμενες μεθοδολογίες.

# ΚΕΦΑΛΑΙΟ 2ο

## 2 ΘΕΜΕΛΙΩΔΕΙΣ ΕΝΝΟΙΕΣ – ΟΡΙΣΜΟΙ

### 2.1 Βασικά στοιχεία πιθανοτήτων

Στα επόμενα κεφάλαια, για την μελέτη του προβλήματος της συμπλήρωσης των κενών που παρουσιάζονται στις μετρήσεις των υδρομετεωρολογικών φαινομένων, γίνεται χρήση διαφόρων εργαλείων της θεωρίας των πιθανοτήτων. Σκοπός μας εδώ δεν είναι να αναλύσουμε πλήρως τη θεωρία των πιθανοτήτων, αλλά μόνο να περιγράψουμε τα εργαλεία που χρησιμοποιούνται στα επόμενα κεφάλαια και βοηθούν στην κατανόηση των διαφόρων υδρολογικών διεργασιών. Πριν λοιπόν την ανασκόπηση των μεθόδων συμπλήρωσης των χρονοσειρών και την περεταίρω διερεύνηση του βέλτιστου τρόπου συμπλήρωσης των κενών, θα δοθούν οι ορισμοί των βασικών εννοιών και οι συμβολισμοί που θα χρησιμοποιηθούν στα επόμενα κεφάλαια ώστε να είναι ευκολότερη η ανάγνωση και η κατανόησή της εργασίας. Οι ορισμοί που δίνονται βασίζονται στην θεμελίωση της θεωρίας πιθανοτήτων κατά Kolmogorov.

#### *2.1.1 Τυχαίες μεταβλητές*

Οι διάφορες υδρολογικές μεταβλητές, παρουσιάζονται και αντιμετωπίζονται ως τυχαίες μεταβλητές (τ.μ.). Τυχαία μεταβλητή $X$ είναι μια συνάρτηση που σε κάθε απλό ενδεχόμενο $\omega$ ενός δειγματικού χώρου[7] $\Omega$ αντιστοιχεί ένα πραγματικό αριθμό. Δηλαδή, η τ.μ. αποτελεί απεικόνιση του δειγματικού χώρου στο σύνολο των πραγματικών αριθμών $\mathbb{R}$. Εάν η τ.μ. παίρνει πεπερασμένο ή άπειρο αριθμήσιμο πλήθος τιμών, ονομάζεται διακριτή, ενώ εάν παίρνει άπειρο μη αριθμήσιμο πλήθος τιμών, ονομάζεται συνεχής. Οι ανελίξεις[8] που περιγράφουν τις υδρομετεωρολογικές διεργασίες και που θα εξεταστούν στη συνέχεια αποτελούνται από διακριτές τ.μ.. Στη

[7] Ως δειγματικός χώρος $\Omega$ ορίζεται το σύνολο που τα στοιχεία του $\omega$ αντιστοιχούν στις δυνατές εκβάσεις ενός πειράματος ή μιας διεργασίας.

[8] Η μεταβολή της καταστάσεως ενός συστήματος μέσα στο χρόνο ονομάζεται ανέλιξη (process). Όπως διευκρινίζεται στο υδρογλωσσικό που είναι διαθέσιμο στον ιστότοπο http://www.itia.ntua.gr/dk-el/hydroglossica/orologia, «…η λέξη ανέλιξη (από το ρήμα *ανελίσσω* = *ξετυλίγω*, *εξελίσσομαι*) αποδίδει τον αγγλικό όρο process. Οι υδρολογικές διεργασίες περιγράφονται από μαθηματικά μοντέλα που λέγονται ανελίξεις (στοχαστικές ανελίξεις)».
Λεπτομερής ορισμός της έννοιας της ανέλιξης και ειδικότερα της στοχαστικής ανέλιξης δίνεται στο υποκεφάλαιο 2.3.

συνέχεια, με κεφαλαία και πλάγια στοιχεία του λατινικού αλφαβήτου (π.χ. $X$) θα συμβολίζομε τις τ.μ. και με πεζά (π.χ. $x_1, x_2, \ldots, x_n$) τις τιμές της, τα στοιχεία δηλαδή της αντίστοιχης χρονοσειράς.

Η συνάρτηση της πραγματικής μεταβλητή $x$ που δίνεται από την εξίσωση

$$F_x(x) := P(X \leq x) \tag{2.1}$$

ονομάζεται συνάρτηση κατανομής και ορίζεται $\forall\, x \in \mathbb{R}$.

Η συνάρτηση κατανομής της τ.μ. $X$, δίνει την πιθανότητα η τ.μ. $X$ να πάρει όλες της τιμές της μέχρι και την τιμή $x$.

Η παράγωγος της συνάρτησης κατανομής, δηλαδή η

$$f_x(x) := \frac{dF(x)}{dx} \tag{2.2}$$

λέγεται συνάρτηση πυκνότητας πιθανότητας.

Αξίζει να σημειωθεί πως ενώ για τις συνεχείς τ.μ. η συνάρτηση αυτή ορίζεται $\forall\, x \in \mathbb{R}$, δεν ισχύει το ίδιο και για την περίπτωση των διακριτών μεταβλητών.

### *2.1.2 Αναμενόμενες τιμές και παράμετροι κατανομών*

Έστω $X$ μια τ.μ. με συνάρτηση κατανομής $F_x(x)$, τότε αν η $X$ είναι συνεχής, ορίζεται ως αναμενόμενη τιμή ή μέση τιμή ή προσδοκία της $X$ το μέγεθος

$$m_X := E(X) := \int_R x\, f(x) dx \tag{2.3}$$

ενώ αν η $X$ είναι διακριτή με τιμές $x_1, x_2 \ldots, x_n$ με $n \in \mathbb{Z}$, η αναμενόμενη τιμή παίρνει τη μορφή

$$m_X := E(X) := \sum_{i=1}^{\infty} x\, p(x) \tag{2.4}$$

Η μέση τιμή μιας τ.μ. αποτελεί μια παράμετρο θέσης της αντίστοιχης κατανομής και το φυσικό της νόημα είναι η περιγραφή του κέντρου βάρους του σχήματος που ορίζει η συνάρτηση πυκνότητας πιθανότητας με τον οριζόντιο άξονα. Η μέση τιμή ισοδυναμεί επίσης και με την αντίστοιχη στατική ροπή ως προς τον κατακόρυφο άξονα του σχήματος που περιγράφεται πιο πάνω με δεδομένο πως το εν λόγω εμβαδό ισούται με 1 (Κουτσογιάννης, 1996 σ. 19).

Αν $g(X)$ είναι συνάρτηση της συνεχούς τ.μ. $X$ τότε ορίζεται ως αναμενόμενη τιμή ή προσδοκία της $g(X)$ το μέγεθος

$$m_{g(X)} := E\big[g(X)\big] := \int_R g(x)\, f(x)\, dx \tag{2.5}$$

ενώ στην περίπτωση που η τ.μ. $X$ είναι διακριτή, και παίρνει τις τιμές $x_1, x_2 \ldots, x_n$ με $n \in \mathbb{Z}$, η παραπάνω σχέση γίνεται

$$m_{g(X)} := E\left[g(X)\right] := \sum_{i=1}^{n} g(x_i) P(X = x_i) \tag{2.6}$$

Στη ειδική περίπτωση που η συνάρτηση $g(X)$ είναι της μορφής $g(X) = X^r$

με $r = 0, 1, 2, \ldots, n$ όπου $n \in \mathbb{Z}$, το μέγεθος

$$m_X^{(r)} := E\left[X^r\right] \tag{2.7}$$

ονομάζεται ροπή περί την αρχή ή ροπή τάξης $r$ της τ.μ. $X$, και για $r$=1 έχουμε την ροπή πρώτης τάξης η οποία προφανώς ταυτίζεται με την μέση τιμή.

Στη περίπτωση που $g(X) = (X - m_X)^r$ με $r = 0, 1, 2, \ldots, n$ όπου $n \in \mathbb{Z}$, τότε το μέγεθος

$$\mu_X^{(r)} := E\left[(X - m_X)^r\right] \tag{2.8}$$

ονομάζεται κεντρική ροπή τάξης $r$ της $X$.

Η κεντρική ροπή 2ης τάξης (δηλαδή για $r = 2$) ισούται με

$$Var\left[X\right] = s_X^2 := m_X^{(2)} := E\left[(X - m_X)^2\right] \tag{2.9}$$

και ονομάζεται διασπορά της $X$.

Η παράμετρος της διασποράς μιας τυχαίας μεταβλητής δηλώνει πόσο συγκεντρωμένες γύρω από τη μέση τιμή είναι οι τιμές της τυχαίας μεταβλητής. Το γεωμετρικό αντίστοιχο της διασποράς είναι η ροπή αδράνειας περί τον κατακόρυφο κεντροβαρικό άξονα του σχήματος που ορίζει η συνάρτηση πυκνότητας πιθανότητας με τον οριζόντιο άξονα (Κουτσογιάννης, 1996 σ. 20).

Η τετραγωνική ρίζα τις διασποράς:

$$\sigma_X = \sqrt{Var\left[X\right]} \tag{2.10}$$

έχει ίδιες διαστάσεις με τη τ.μ. και ονομάζεται τυπική απόκλιση.

Ο συντελεστής μεταβλητότητας ορίζεται ως

$$C_{v_X} := \frac{\sigma_X}{m_X} \tag{2.11}$$

Η 3η κεντρική ροπή περιγράφει την ασυμμετρία της κατανομής και αντίστοιχα ορίζεται ο συντελεστής ασυμμετρίας

$$C_{s_X} := \frac{\mu_X^{(3)}}{\sigma_X^3} \tag{2.12}$$

Τέλος, η 4η κεντρική ροπή χρησιμοποιείται ως παράμετρος κύρτωσης, περιγράφει δηλαδή πόσο 'αιχμηρή' είναι η συνάρτηση πυκνότητας πιθανότητας γύρω από την κορυφή της. Ο συντελεστής κύρτωσης είναι

$$C_{k_X} := \frac{\mu_X^{(4)}}{\sigma_X^4} \tag{2.13}$$

### *2.1.3 Από κοινού ιδιότητες δύο τυχαίων μεταβλητών*

Οι ορισμοί που δόθηκαν πιο πάνω αναφέρονται σε μια μεμονωμένη τυχαία μεταβλητή. Στην υδρολογία όμως συχνά μας ενδιαφέρει η ταυτόχρονη μελέτη περισσότερων μεταβλητών. Έστω λοιπόν το ζεύγος (*X, Y*) των τ.μ. και οι αντίστοιχοι δειγματικοί χώροι ($\Omega_X$, $\Omega_Y$).

Η από κοινού συνάρτηση κατανομής του ζεύγους των τ.μ. (*X, Y*) ορίζεται ως

$$F_{XY}(x, y) := P(X \le x, Y \le y) \tag{2.14}$$

Η παράγωγος της από κοινού συνάρτησης κατανομής (με την προϋπόθεση ότι αυτή είναι παραγωγίσιμη), όπως και στην περίπτωση της μεμονωμένης τ.μ., ορίζει τη συνάρτηση πυκνότητας πιθανότητας των μεταβλητών, η οποία ισούται με

$$f_{XY} := \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \tag{2.15}$$

Οι περιθώριες συναρτήσεις κατανομής και οι αντίστοιχες περιθώριες συναρτήσεις πυκνότητας πιθανότητας ορίζονται ως

$$F_X(x) := P(X \le x) = \lim_{y \to \infty} F_{XY}(x, y) \tag{2.16}$$

$$F_Y(y) := P(Y \le y) = \lim_{x \to \infty} F_{XY}(x, y) \tag{2.17}$$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy \tag{2.18}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx \tag{2.19}$$

για *X* και *Y* αντίστοιχα.

Σε αναλογία με όσα είπαμε στην περίπτωση μεμονωμένων μεταβλητών, και εδώ, μπορούμε αναλόγως να ορίσουμε αναμενόμενες τιμές και ροπές. Πιο συγκεκριμένα, στη περίπτωση δύο τ.μ., η αναμενόμενη τιμή ή προσδοκία της συνάρτησης *g(X,Y)* ορίζεται από τη σχέση

$$E\left[g(X,Y)\right] := \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y) dy dx \tag{2.20}$$

Η από κοινού ροπή *p+q* τάξης των *X* και *Y* ορίζεται ως $E\left[X^p Y^q\right]$ και ισούται με

$$m_{pq} \equiv E\left[X^p Y^q\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f_{XY}(x, y)\, dx\, dy \tag{2.21}$$

Ενώ από κοινού κεντρική ροπή τάξης $p+q$ των $X$ και $Y$ ορίζεται το μέγεθος

$$\mu_{pq} \equiv E\left[\left(X-m_X\right)^p\left(Y-m_Y\right)^q\right] := \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-m_X)^p(y-m_Y)^q f_{XY}(x,y)\,dx\,dy \quad (2.22)$$

Η πιο συχνά χρησιμοποιούμενη στην υδρολογία κεντρική ροπή δύο μεταβλητών είναι αυτή που προκύπτει για $p = q = 1$. Η ροπή αυτή συνήθως ονομάζεται συνδιασπορά των $X$ και $Y$ και ισούται με

$$\mathrm{Cov}\left[X,Y\right] \equiv \sigma_{XY} := E\left[\left(X-m_X\right)\left(Y-m_Y\right)\right] = E\left[XY\right] - m_X m_Y \quad (2.23)$$

Η συνδιασπορά διαιρούμενη με τις τυπικές αποκλίσεις των $X$ και $Y$ μας δίνει τον αδιάστατο συντελεστή που ονομάζεται συντελεστής συσχέτισης και συμβολίζεται με $\rho_{XY}$. Δηλαδή

$$\rho_{XY} = \frac{\mathrm{Cov}\left[X,Y\right]}{\sqrt{\mathrm{Var}\left[X\right]\mathrm{Var}\left[Y\right]}} \equiv \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (2.24)$$

Ο συντελεστής συσχέτισης μπορεί να λάβει τιμές από -1 έως και 1.

Η έννοια όμως τις συσχέτισης υφίσταται και όταν έχουμε μία μόνο τ.μ.. Ο συντελεστής συσχέτισης στην περίπτωση αυτή ονομάζεται συντελεστής αυτοσυσχέτισης και αναφέρεται στη συσχέτιση μεταξύ των τιμών της τ.μ. για διάφορες τιμές υστέρηση[9]. Δηλαδή, η συνάρτηση αυτοσυσχέτισης περιγράφει τις ιδιότητες μιας χρονοσειράς στο πεδίο του χρόνου.

Για παράδειγμα, έστω $x_1, x_2 \ldots, x_n$ με $n \in \mathbb{Z}$ οι τιμές μιας τ.μ. $X$, αν μετατοπίσουμε τη χρονοσειρά κατά 1 χρονικό βήμα εμπρός (δηλ. lag = 1) και εξετάσουμε τη συσχέτιση μεταξύ των διαφόρων τιμών της αρχικής χρονοσειράς και της μετατοπισμένης μπορούμε να εξάγουμε σημαντικά συμπεράσματα για το πόσο εξαρτημένες είναι οι τιμές μεταξύ τους. Ο αντίστοιχος συντελεστής αυτοσυσχέτισης συμβολίζεται $\rho_1$.

Ο συντελεστής αυτοσυσχέτισης $\rho(\tau)$ μιας χρονοσειράς δίνεται λοιπόν από τη σχέση

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} \quad (2.25)$$

όπου

$$\gamma_\tau = C(t,\tau) := \mathrm{Cov}\left[X(t), X(t+\tau)\right] = \mathrm{E}\left[\left(\mathrm{X}(t)-\mu(t)\right)\left(\mathrm{X}(t+\tau)-\mu(t+\tau)\right)\right]$$

και περιγράφει τη γραμμική συσχέτιση μεταξύ των τιμών $x(t)$ και $x(t+\tau)$ της χρονοσειράς.

---

[9] Ο όρος υστέρηση χρησιμοποιείτε ευρέως στη στατιστική, και αναφέρεται στη μετατόπιση της χρονοσειράς κάποια χρονικά βήματα εμπρός ή αντίστοιχα κάποια βήματα πίσω. Πολύ συχνά συμβολίζεται ως lag.

Η γραφική παράσταση του συντελεστή αυτοσυσχέτισης σε σχέση με την υστέρηση λέγεται συσχετόγραμμα και είναι πολύ χρήσιμη στην ανάλυση των χρονοσειρών.

## 2.2 Βασικά εργαλεία στατιστικής – γεωστατιστικής

Η στατιστική, είναι η επιστήμη που ασχολείται με την οργάνωση, την ανάλυση και την παρουσίαση δεδομένων. Η στατιστική δηλαδή συνίσταται στη συλλογή δεδομένων, την περιγραφή και κυρίως την ανάλυση τους με στόχο την εξαγωγή συμπερασμάτων. Ειδικότερα, τα στατιστικά μοντέλα προσπαθούν να προσαρμόσουν μαθηματικές εξισώσεις στις χρονοσειρές με στόχο την πρόβλεψη των άγνωστων ποσοτήτων. Αν για ένα φαινόμενο δεν υπάρχει καθόλου αβεβαιότητα ή τυχαιότητα τότε είναι προφανές πως δε χρειάζεται η στατιστική προσέγγιση αφού καθοριστικά μοντέλα (deterministic models) μπορούν να περιγράψουν το φαινόμενο επακριβώς. Στα πραγματικά φαινόμενα και ειδικότερα στις υδρολογικές διεργασίες, που είναι το βασικό αντικείμενο της επιστήμης της υδρολογίας, σχεδόν πάντα υπάρχει ο παράγοντας της αβεβαιότητας ή τυχαιότητας κι αυτό κάνει σημαντική τη γνώση των στατιστικών μεθόδων και μοντέλων.

Η γεωστατιστική έχει υιοθετήσει πολλά στοιχεία από τη θεωρία των πιθανοτήτων καθώς επίσης και διάφορες στατιστικές μεθόδους που αποβλέπουν στην ανάλυση δεδομένων στο χρόνο και στο χώρο. Βρίσκει έτσι τεράστια εφαρμογή στην επιστήμη της υδρολογίας. Πολλές υδρολογικές μεταβλητές δεν μπορούν να περιγραφούν ικανοποιητικά από ντετερμινιστικούς νομούς. Αυτό το κενό καλείτε να καλύψει η γεωστατιστική. Η γεωστατιστική λοιπόν, αποτελείται από ένα σύνολο στατιστικών μεθόδων εκτίμησης, που αφορούν ποσότητες που μεταβάλλονται στο χρόνο αλλά και στο χώρο (Kitanidis σσ. 20.1-20.3).

Θα πρέπει σε αυτό το σημείο, να διαχωρίσουμε την έννοια του πληθυσμού και του δείγματος. Πληθυσμός είναι μια ομάδα ή μια κατηγορία στην οποία αναφέρεται η τ.μ.. Ο πληθυσμός μπορεί να είναι είτε πεπερασμένος είτε άπειρος. Ενώ δείγμα είναι ένα υποσύνολο του πληθυσμού που εξετάζουμε και αναφέρεται σε ένα σύνολο μετρήσεων για το συγκεκριμένο πληθυσμό, για παράδειγμα οι τιμές της ημερήσιας βροχόπτωσης από το 1995 έως το 2010.

Για την επεξεργασία λοιπόν των δεδομένων (χρονοσειρές παρατηρήσεων) βασιζόμαστε στην κλασική στατιστική και υπολογίζουμε τις εκτιμήτριες των ροπών (συνήθως στην υδρολογία χρησιμοποιούμε μέχρι και ροπές τρίτης τάξης[10]).

Για κάθε παράμετρο $\eta$ πληθυσμού μπορούν να βρεθούν μία ή περισσότερες στατιστικές συναρτήσεις της μορφής $\Theta = g(X_1,...,X_n)$ κατάλληλες για την εκτίμηση

[10] Η εξαγωγή αμερόληπτων εκτιμητριών είναι αρκετά πολύπλοκη για την περίπτωση ροπών τάξης μεγαλύτερης του 3 και έχουν μεγάλη διασπορά για μικρά δείγματα (Κουτσογιάννης, 1996 σ. 53).

αυτής της παραμέτρου. Σε αυτή την περίπτωση λέμε ότι η $\Theta = g(X_1,...,X_n)$ είναι εκτιμήτρια της παραμέτρου $\eta$ και η αριθμητική της τιμή $\theta = g(x_1,...,x_n)$ αποτελεί εκτίμηση της $\eta$. Υπάρχουν διάφορες κατηγορίες εκτιμητριών :

- Μια συνάρτηση $\Theta$ είναι αμερόληπτη εκτιμήτρια της παραμέτρου $\eta$ αν $E(\Theta) = \eta$. Διαφορετικά είναι μεροληπτική εκτιμήτρια και η διαφορά $E(\Theta) - \eta$ λέγεται μεροληψία.
- Μια στατιστική συνάρτηση $\Theta$ είναι συνεπής εκτιμήτρια της παραμέτρου $\eta$ αν το σφάλμα εκτίμησης $\Theta$ - $\eta$ τείνει στο μηδέν με πιθανότητα 1 για $n \to \infty$. Διαφορετικά είναι ασυνεπής εκτιμήτρια.
- Μια στατιστική συνάρτηση $\Theta$είναι βέλτιστη εκτιμήτρια της παραμέτρου $\eta$ αν το μέσο τετραγωνικό σφάλμα εκτίμησης $(\Theta - \eta)^2$ είναι ελάχιστο.
- Μία στατιστική συνάρτηση $\Theta$ είναι η πιο αποτελεσματική εκτιμήτρια της παραμέτρου $\eta$ αν είναι αμερόληπτη και έχει την ελάχιστη διασπορά.

Οι τυπικές στατιστικές εκτιμήτριες που αναφέρονται σε στατιστικές ροπές του πληθυσμού και είναι ανεξάρτητες της συνάρτησης κατανομής πιθανότητας είναι :

- Δειγματική μέση τιμή
  Αποτελεί την πιο κοινή στατιστική συνάρτηση και είναι εκτιμήτρια της μέσης τιμής και ορίζεται από τη σχέση

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.26}$$

  Η αριθμητική τιμή της

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.27}$$

  ονομάζεται παρατηρημένη ή αριθμητική μέση τιμή ή απλώς μέσος όρος του δείγματος.
  Επειδή ισχύει

$$E\left[\bar{X}\right] = E\left[X\right] \quad \text{και} \quad \text{Var}\left[\bar{X}\right] = \frac{\text{Var}\left[X\right]}{n} \tag{2.28}$$

  η εκτιμήτρια είναι αμερόληπτη και συνεχής.
- Διασπορά και τυπική απόκλιση
  Η μεροληπτική εκτιμήτρια της διασποράς $\sigma_X^2$ του πληθυσμού δίνεται από τη σχέση

$$S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} \tag{2.29}$$

Στη περίπτωση που οι τ.μ. $X_i$ είναι ανεξάρτητες και όμοια κατανεμημένες, η αμερόληπτη εκτιμήτρια της διασποράς γίνεται

$$S_X^{*2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \qquad (2.30)^{11}$$

- Τρίτη κεντρική ροπή και συντελεστής ασυμμετρίας
  Η μεροληπτική εκτιμήτρια της τρίτης κεντρικής ροπής του πληθυσμού δίνεται από την σχέση

$$\widehat{m_X^{(3)}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{n} \qquad (2.31)$$

Στη περίπτωση που οι τ.μ. $X_i$ είναι ανεξάρτητες και όμοια κατανεμημένες, η αμερόληπτη εκτιμήτρια της τρίτης ροπής γίνεται

$$\widehat{m_X^{*(3)}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{(n-1)(n-2)} \qquad (2.32)$$

Η εκτιμήτρια του συντελεστή ασυμμετρίας δίνεται από τη σχέση

$$\widehat{C_{S_X}} = \frac{\widehat{M_X^{(3)}}}{S_X^3} \qquad (2.33)$$

- Συνδιασπορά και συσχέτιση
  Η μεροληπτική εκτιμήτρια της συνδιασποράς δυο ανεξάρτητων τ.μ. $X$ και $Y$ δίνεται από τη σχέση

$$S_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n} \qquad (2.34)$$

Στη περίπτωση που οι τ.μ. $X_i$ είναι ανεξάρτητες και όμοια κατανεμημένες, η αμερόληπτη εκτιμήτρια της συνδιασποράς γίνεται

[11] Παρατηρούμε πως γίνεται χρήση του $n$-1 αντί του $n$. Όπως η δειγματική μέση τιμή εκτιμά τη μέση τιμή του πληθυσμού $\mu$, έτσι και η δειγματική διασπορά $s^2$ εκτιμά τη διασπορά του πληθυσμού. Αν γνωρίζαμε τη $\mu$ τότε θα τη χρησιμοποιούσαμε στον τύπο για τον υπολογισμό του $s^2$, αλλά συνήθως η $\mu$ είναι άγνωστη. Οι παρατηρήσεις $x_i$, τείνουν να είναι πιο κοντά στη $\bar{x}$ παρά στη $\mu$ και άρα οι υπολογισμοί με βάση τις αποκλίσεις $xi - \bar{x}$ δίνουν μικρότερες τιμές από ότι αν χρησιμοποιούσαμε τις αποκλίσεις $xi - \mu$. Για να αντισταθμίσουμε αυτήν την τάση για υποεκτίμηση της διασποράς του πληθυσμού διαιρούμε με $n - 1$ αντί με $n$.

$$S_{XY}^{*} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{2.35}$$

Η εκτιμήτρια του συντελεστή συσχέτισης δίνεται από τη σχέση

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{S_{XY}^{*}}{S_X^{*} S_Y^{*}} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{2.36}$$

## 2.3 Στοχαστικές ανελίξεις

Πολλές υδρομετεωρολογικές διεργασίες είναι τόσο περίπλοκες, που μπορούν να ερμηνευθούν και να προσομοιωθούν μόνο με πιθανοτική θεώρηση. Τα υδρολογικά γεγονότα παρουσιάζουν τεράστια αβεβαιότητα και οι ανελίξεις που τα προσομοιώνουν περιέχουν τυχαίους παράγοντες με στοχαστικό χαρακτήρα. Οι στοχαστικές ανελίξεις[12] και γενικότερα τα διάφορα στοχαστικά εργαλεία, χρησιμοποιούνται ευρύτατα στην υδρολογία γιατί διευκολύνουν την προσομοίωση των φυσικών διεργασιών. Πιο συγκεκριμένα, η στοχαστική προσομοίωση υπερτερεί έναντι των απλών αιτιοκρατικών νόμων, γιατί δίνει τη δυνατότητα παραγωγής (με τη χρήση του κατάλληλου στοχαστικού μοντέλου) συνθετικών χρονοσειρών με μήκος πολλαπλάσιο του πραγματικού δείγματος. Το ιστορικό δείγμα, σπανίως είναι αρκετό για την πρόβλεψη με βεβαιότητα μελλοντικών γεγονότων και αυτό γιατί οι ιστορικές παρατηρήσεις δεν μπορούν να θεωρηθούν αντιπροσωπευτικές για όλες τις πιθανές μελλοντικές διακυμάνσεις ενός φαινομένου, που μπορούν να εμφανιστούν κατά τη διάρκεια ζωής ενός έργου (Warren, Viessman Jr.; Gary L., Lewis σ. 535). Για το λόγο αυτό, είναι συνήθως αναγκαία η επέκταση των ιστορικών μετρήσεων. Ενώ λοιπόν η χρονοσειρά των πραγματικών μετρήσεων έχει περιορισμένο μήκος, ένα στοχαστικό μοντέλο, βασισμένο στις υπάρχουσες μετρήσεις και λαμβάνοντας υπόψη τις ιδιαιτερότητες της εκάστοτε φυσικής διεργασίας που πρόκειται να προσομοιωθεί, μπορεί να παράγει συνθετικές χρονοσειρές με μήκος πολλαπλάσιο της ιστορικής, κατάλληλες για τη μελέτη και επεξεργασία του φαινομένου.

[12] Όπως επισημαίνεται στο υδρογλωσσικό που είναι διαθέσιμο στον ιστότοπο http://www.itia.ntua.gr/dk-el/hydroglossica/orologia, «…το επίθετο στοχαστικός εδώ δεν έχει τη σημασία που έχει στην καθομιλουμένη, αλλά αυτήν του τυχαίου. Με αυτή την έννοια έχει εισαχθεί ως επιστημονικός όρος από τον Ελβετό μαθηματικό Giacomo Bernoulli πριν 300 και πλέον χρόνια. Το επίθετο προέρχεται από το αρχαιοελληνικό ρήμα *στοχάζομαι* με την έννοια του *εικάζω* (η αρχική σημασία του *στοχάζομαι* είναι *σημαδεύω το στόχο*, κατόπιν έγινε *εικάζω*, *νομίζω*, και τέλος *συλλογίζομαι*).».

Ένα μοντέλο που περιγράφει την πιθανοτική δομή μιας ακολουθίας παρατηρήσεων ονομάζεται στοχαστική ανέλιξη (Box and Jenkins, 1976 σ. 21). Αναλυτικότερα, οι στοχαστικές ανελίξεις αποτελούν οικογένειες τ.μ.. Για παράδειγμα μια τ.μ. $X_t$, όπου $t$ είναι μια παράμετρος που παίρνει τιμές από ένα κατάλληλο σύνολο $T$. Το σύνολο $T$ συνήθως παριστάνει χρόνο. Η χρονοσειρά αποτελεί την υλοποίηση μιας ακολουθίας τυχαίων μεταβλητών η οποία είναι γνωστή ως στοχαστική ανέλιξη. Η θεωρία στοχαστικών ανελίξεων χρησιμοποιείται όταν η χρονοσειρά που μελετάμε αποτελείται από στοιχεία στοχαστικά εξαρτημένα.

Αν $T = \{0,1,2,...\}$ δηλαδή η παράμετρος $t$ αντιστοιχεί σε διακριτό χρόνο τότε έχουμε ανέλιξη σε διακριτό χρόνο. Η αντίστοιχη χρονοσειρά ονομάζεται διακριτή χρονοσειρά. Η διακριτή χρονική κλίμακα είναι διαδεδομένη για την περιγραφή υδρολογικών διεργασιών όπως η βροχόπτωση και η απορροή. Η διακριτή χρονοσειρά μπορεί να προκύψει μετά από ολοκλήρωση σε διαδοχικά διακριτά χρονικά διαστήματα. Αν οι τιμές της χρονοσειράς έχουν σταθερό βήμα μεταξύ τους, τότε η χρονοσειρά λέγεται σταθερού βήματος αλλιώς λέγεται μεταβλητού βήματος. Μια χρονοσειρά με ελλείπουσες τιμές την οποία πρόκειται να συμπληρώσουμε την ονομάζουμε ελλιπή χρονοσειρά. Η χρονοσειρά ή οι χρονοσειρές που χρησιμοποιούμε για να συμπληρώσουμε μια ελλιπή χρονοσειρά ονομάζεται χρονοσειρά αναφοράς.

Σε αυτή την εργασία, θα ασχοληθούμε με διακριτές χρονοσειρές σταθερού βήματος, τις οποίες για λόγους απλότητας θα τις καλούμε στο εξής χρονοσειρές.

Οι αναμενόμενες τιμές μίας στοχαστικής ανέλιξης ορίζονται όπως ακριβώς παρουσιαστήκαν και στην περίπτωση των τ.μ. στο υποκεφάλαιο 2.1.2. Ιδιαίτερο ενδιαφέρον όμως παρουσιάζουν:

Η μέση τιμή της ανέλιξης, δηλαδή η αναμενόμενη τιμή της μεταβλητής *X(t)* :

$$\mu_t = E\left[X(t)\right] \tag{2.37}$$

Για διάφορες τιμές $\tau$, με $\tau \in \mathbb{Z}$, της υστέρησης (lag) ορίζεται η

Αυτοσυνδιασπορά :

$$\gamma_\tau = C(t,\tau) := \mathrm{Cov}\left[X(t), X(t+\tau)\right] = \mathrm{E}\left[\left(\mathrm{X}(t) - \mu(t)\right)\left(\mathrm{X}(t+\tau) - \mu(t+\tau)\right)\right] \tag{2.38}$$

Διασπορά :

$$\gamma_0 = C(t,0) = \mathrm{Var}\left[X(t)\right] = \mathrm{Cov}\left[X(t), X(t)\right] \tag{2.39}$$

και ο συντελεστής αυτοσυσχέτισης :

$$\rho(t,\tau) := \frac{\mathrm{Cov}\left[X(t), X(t+\tau)\right]}{\sqrt{\mathrm{Var}\left[X(t)\right]\mathrm{Var}\left[X(t+\tau)\right]}} \tag{2.40}$$

Για δύο διαφορετικές ανελίξεις $X$ και $Y$ ορίζεται η ετεροσυνδιασπορά

$$C_{XY}(t,\tau) := \mathrm{Cov}\left[X(t), Y(t+\tau)\right] \tag{2.41}$$

και ο συντελεστής ετεροσυσχέτισης

$$\rho_{XY} := \frac{\mathrm{Cov}\left[X(t), Y(t+\tau)\right]}{\sqrt{\mathrm{Var}\left[X(t)\right]\mathrm{Var}\left[Y(t+\tau)\right]}} \tag{2.42}$$

των δύο μεταβλητών.

Αξίζει να σημειωθεί, όπως γίνεται φανερό και από τις πιο πάνω σχέσεις, πώς οι στατιστικές παράμετροι μιας στοχαστικής ανέλιξης, όπως για παράδειγμα η μέση τιμή ή η τυπική απόκλιση, μεταβάλλονται με το χρόνο. Οι χρονοσειρές αυτές ονομάζονται μη στάσιμες.

Η ειδική περίπτωση ανελίξεων, που τα στατιστικά της χαρακτηριστικά παραμένουν σταθερά με το χρόνο είναι οι στάσιμες ανελίξεις. Έτσι, μια στοχαστική ανέλιξη λέγεται στάσιμη με την αυστηρή έννοια όταν η από κοινού συνάρτηση κατανομής της ανέλιξης δεν επηρεάζεται από τη χρονική μεταβολή. Ειδική περίπτωση στάσιμης ανέλιξης έχουμε όταν η μέση τιμή $\mu$ της τ.μ. παραμένει σταθερή με το χρόνο και η αυτοσυνδιασπορά της εξαρτάται μόνο από την μεταβολή του χρόνου. Η περίπτωση αυτή στασιμότητας ονομάζεται στασιμότητα με την ευρεία (ελαστική) έννοια. Συνοψίζοντας λοιπόν τις δυο προηγούμενες περιπτώσεις έχουμε:

Στάσιμη με την αυστηρή έννοια στοχαστική ανέλιξη: όλα τα στατιστικά χαρακτηριστικά της ανέλιξης παραμένουν σταθερά με χρόνο.

Στάσιμη με την ευρεία έννοια στοχαστική ανέλιξη:

$$E\left[x(t)\right] = E\left[x(1)\right] = E\left[x(2)\right] = \ldots = E\left[x(n)\right] = \mu = \text{σταθερά}$$

$$\mathrm{Cov}\left[X(t), X(t+\tau)\right] = E\left[\left(X(t)-\mu\right)\left(X(t+\tau)-\mu\right)\right] = E\left[X(t)X(t+\tau)-\mu^2\right] = C(\tau)$$

όπου $\tau$ η χρονική υστέρηση.

Έτσι, μπορούμε να αναπαραστήσουμε μια στάσιμη (με την αυστηρή έννοια) στοχαστική ανέλιξη με την ακόλουθη σχέση

$$Z(t) = \mu + \varepsilon(t) \tag{2.43}$$

όπου

$Z(t)$ η τιμή της τ.μ $Z$ τη στιγμή $t$

$\mu$ η μέση τιμή της στοχαστικής ανέλιξης

$\varepsilon(t)$ ένας τυχαίος όρος που ακολουθεί κατανομή με μέση τιμή 0 και συνδιασπορά που δίνεται από τον τύπο $C(h) = E\left[\varepsilon(t)\varepsilon(t+h)\right]$

Στην περίπτωση που δεν έχουμε στάσιμη (είτε με την αυστηρή είτε με την ευρεία έννοια) στοχαστική ανέλιξη, δεν μπορούμε να ορίσουμε τη συνάρτηση της συνδιασποράς, γιατί η μέση τιμή μεταβάλλεται για κάθε χρονική μεταβολή. Στην περίπτωση ασθενούς στασιμότητας (weak stasionarity) της στοχαστικής ανέλιξης που μελετάμε, δηλαδή η μέση τιμή να μεταβάλλεται ελαφρώς στην περιοχή που

εξετάζουμε και η διασπορά να αυξάνει όσο η περιοχή που εξετάζουμε διευρύνεται, πάλι η συνάρτηση της συνδιασποράς δεν είναι δυνατόν να οριστεί.

Αυτό το κενό μελετήθηκε από το τον (Matheron, 1965), κάνοντας την παραδοχή ότι η μέση τιμή μπορεί μεν να μεταβάλλεται με το χρόνο, αλλά για ένα μικρό χρονικό διάστημα $\Delta t = |x-(x+h)| = |h|$ μπορεί να θεωρηθεί ως σταθερή. Έτσι έχουμε

$$E\left[Z(t)-Z(t+h)\right]=0 \tag{2.44}$$

Αν αντικαταστήσουμε τις συνδιασπορές με τις αντίστοιχες τιμές της διασποράς των διαφορών των $Z(x)$ και $Z(x+h)$, σαν μέτρο για τον προσδιορισμό της συσχέτισης μεταξύ των τιμών της τ.μ. έχουμε ότι

$$\mathrm{Var}\left[Z(t)-Z(t+h)\right]=E\left[\left(Z(t)-Z(t+h)\right)^2\right]=2\mathrm{Var}\left[Z(t)\right]-2\mathrm{Cov}\left[Z(t)\right]=2\gamma(h) \tag{2.45}$$

Η παραπάνω σχέση αποτελεί την εγγενή υπόθεση του Matheron (Matheron intrinsic hypothesis), και συνιστά ένα πολύ καλό εργαλείο για να αντιμετωπιστούν οι περιορισμοί των ασθενώς στάσιμων χρονοσειρών (weak stasionarity – second order stasionarity). Η ποσότητα $\gamma(h)$ είναι γνωστή ως ημιδιασπορά (semivariance) για υστέρηση $h$. Ο όρος ημιδιασπορά στηρίζεται στο γεγονός ότι είναι το μισό της διασποράς.

Στην περίπτωση ομοιόμορφου ισότροπου πεδίου, η συνάρτηση της ημιδιασποράς συνδέεται με την συνάρτηση αυτοσυνδιασποράς με την παρακάτω σχέση

$$\gamma(h)=C(0)-C(h) \tag{2.46}$$

για $h\to\infty, \gamma(h)\to C(0)=\sigma^2$

αν όμως η $Z(x)$ δεν είναι στάσιμη, τότε $\gamma(h)\to\infty$ για $h\to\infty$

Η απεικόνιση της ημιδιασποράς για διάφορες τιμές της υστέρησης είναι γνωστή ως ημι-μεταβλητόγραμμα (semivariogram) ή συνηθέστερα απλός μεταβαλητόγραμμα (variogam) (Webster & Oliver, 2001 σσ. 47-59).

Μια άλλη πολύ σημαντική ιδιότητα των στοχαστικών ανελίξεων είναι η εργοδικότητα. Η έννοια της εργοδικότητας σχετίζεται με τον προσδιορισμό της κατανομής μιας στοχαστικής ανέλιξης από μια σειρά παρατηρήσεών της. Πιο συγκεκριμένα, μια στάσιμη στοχαστική ανέλιξη ονομάζεται εργοδική όταν κάθε παράμετρος της κατανομής της μπορεί να προσδιοριστεί από μια απλή δειγματοσυνάρτηση της ανέλιξης. Ένας γενικότερος ορισμός της εργοδικής στοχαστικής ανέλιξης είναι ο εξής: μια ανέλιξη είναι εργοδική αν οι χρονικοί μέσοι είναι ίσοι με τους συνολικούς μέσους (δηλαδή τις αναμενόμενες τιμές) (Κουτσογιάννης, 1996 σ. 36).

Μια απλή στοχαστική ανέλιξη είναι ο λευκός θόρυβος. Για το λευκό θόρυβο έχουν δοθεί διάφοροι ορισμοί. Οι πιο σημαντικοί είναι ο ορισμός που δόθηκε από τον Brown (1983), και αυτός που δόθηκε από τον Papoulis (1991). Πιο συγκεκριμένα, ο

Brown όρισε ως λευκό θόρυβο μια στάσιμη τυχαία ανέλιξη, η οποία έχει σταθερή συνάρτηση φασματικής πυκνότητας[13]. Ενώ ο Papoulis θεωρεί ότι μια ανέλιξη $V(t)$ είναι λευκός θόρυβος αν οι τιμές της $V(t_i)$ και $V(t_j)$ είναι ασυσχέτιστες για οποιοδήποτε $t_i$ και $t_j$ με $t_i \neq t_j$. Δηλαδή $E\left[t_i, t_j\right] = 0$.

Γενικεύοντας τους δύο προηγούμενους ορισμούς μπορούμε να πούμε πως ο λευκός θόρυβος είναι μια ακολουθία από τυχαίες μεταβλητές $X(t)$ οι οποίες έχουν μέση τιμή 0, σταθερή τυπική απόκλιση και είναι ασυσχέτιστες δηλαδή $E\left[X(t)\right] = 0$, $E\left[X(t)^2\right] = \sigma^2$ και $E\left[x_t, x_{t+i}\right] = 0$.

Η συνάρτηση αυτοσυνδιασποράς της τ.μ. $X$ δίνεται από τη σχέση

$$\gamma_k = E\left[X_t X_{t+k}\right] = \begin{cases} \sigma_a^2, & k = 0 \\ \\ 0, & k \neq 0 \end{cases} \tag{2.47}$$

Και η συνάρτηση αυτοσυσχέτισης είναι

$$\rho_k = \begin{cases} 1, & k = 0 \\ \\ 0, & k \neq 0 \end{cases} \tag{2.48}$$

Η τ.μ. του λευκού θορύβου ακολουθεί δεδομένη κατανομή. Όταν η κατανομή που ακολουθεί η τ.μ. είναι η κανονική κατανομή, τότε ο η ανέλιξη του λευκού θορύβου λέγεται κανονική.

[13] Η συνάρτηση φασματική πυκνότητας (Spectral Density Function) είναι το φάσμα ισχύος εκφρασμένο ως προς τις αυτοσυσχετίσεις,(Box and Jenkins, 1976 σ. 41) δηλαδή $g(f) = \frac{p(f)}{\sigma_z^2} = 2\left\{1 + 2\sum_{k=1}^{\infty} \rho_\kappa \cos 2\pi f\, k\right\}, 0 \leq f \leq \frac{1}{2}$.

## 2.4 Ιδιαιτερότητες υδρολογικών διεργασιών

Οι υδρολογικές διεργασίες, όπως και οι περισσότερες φυσικές διεργασίες, παρουσιάζουν τεράστια ποικιλία και πολυπλοκότητα. Ο περίπλοκός αυτός χαρακτήρας καθιστά πολύ δύσκολή την μελέτη, την κατανόηση, την προσομοίωση και τελικά την πρόβλεψη των φαινομένων αυτών. Η στοχαστική προσομοίωση χρησιμοποιείται στην υδρολογία ως μια αριθμητική μέθοδος η οποία μπορεί να επιλύσει πολύπλοκα προβλήματα, λαμβάνοντας υπόψη την τυχαιότητα των φυσικών διεργασιών. Από την οπτική λοιπόν γωνία της θεωρίας των πιθανοτήτων, οι υδρολογικές διεργασίες αντιμετωπίζονται ως στοχαστικές ανελίξεις.

Το γεγονός όμως ότι μια φυσική διεργασία περιγράφεται από μια στοχαστική ανέλιξη δεν σημαίνει πως δεν υπακούει και σε κανένα αιτιοκρατικό νόμο. Αντιθέτως, τα υδρολογικά φαινόμενα εμφανίζουν κάποιες ιδιαιτερότητες, η γνώση των οποίων βοηθάει στην κατανόηση της χρονικής τους εξέλιξης και τελικά στην καλύτερη προσομοίωση αυτών των φαινομένων. Πιο συγκεκριμένα, τα υδρολογικά μεγέθη εμφανίζουν περιοδικές διακυμάνσεις κατά τη διάρκεια του έτους. Οι διακυμάνσεις αυτές οφείλονται στην ετήσια κίνηση της γης και στα μετεωρολογικά φαινόμενα που αυτή επηρεάζει. Οι περιοδικές αυτές διακυμάνσεις είναι γνωστές στη βιβλιογραφία ως προσδιοριστική συνιστώσα των διεργασιών.

Έστω λοιπόν $X(t)$ μια ανέλιξη σε συνεχή χρόνο και μια σειρά μετρήσεων σε τακτά χρονικά διαστήματα (χρονοσειρά). Όπως είδαμε πιο πάνω, η στοχαστική ανέλιξη αποτελείται από μια στοχαστική συνιστώσα $J(t)$ και μια προσδιοριστική $D(t)$. Δηλαδή

$$X(t) = D(t) + J(t) \tag{2.49}$$

Στη βιβλιογραφία συχνά συμπεριλαμβάνονται στη ντετερμινιστική συνιστώσα και μακροχρόνιες τάσεις ή πολυετείς μεταβολές των μέσων χαρακτηριστικών της $X(t)$. Όμως επειδή οι μεταβολές που προκαλούν αυτές τις τάσεις στις ανελίξεις δεν είναι πλήρως καθορισμένες, ουσιαστικά πρόκειται για τυχαίες διακυμάνσεις και ως τέτοιες αντιμετωπίζονται στη συνέχεια. Έτσι, το τυχαίο μέρος της ανέλιξης $J(t)$ δεν είναι εντελώς τυχαίο, αλλά παρουσιάζει στοχαστική δομή ή μνήμη. Δηλαδή, υπάρχει στοχαστική εξάρτηση μεταξύ γειτονικών τιμών. Η εξάρτηση αυτή, που περιγράφεται από τον συντελεστή συσχέτισης, μπορεί να είτε χρονική (για παράδειγμα η βροχόπτωση του Σεπτεμβρίου μιας συγκεκριμένης χρονιάς παρουσιάζει έντονη συσχέτιση με την αντίστοιχη του ίδιου μήνα της επόμενης χρονιάς) είτε χωρική (δηλαδή οι μετρήσεις ενός βροχομετρικού σταθμού αναμένεται να παρουσιάζουν έντονη συσχέτιση με τις αντίστοιχες μετρήσεις ενός γειτονικού σταθμού).

Οι υδρολογικές διεργασίες, για λόγους απλοποίησης των υπολογισμών, περιγράφονται από στοχαστικές ανελίξεις σε διακριτό χρόνο.

Επίσης, για την ευκολότερη μελέτη των φαινομένων, γίνονται οι παρακάτω παραδοχές:

- Οι ανελίξεις είναι στάσιμες (η κατανομή κάθε μεταβλητής παραμένει σταθερή, ανεξάρτητα της χρονικής μεταβολής)
- Οι ανελίξεις είναι εργοδικές (οι αναμενόμενες τιμές είναι ίσες με τους χρονικούς μέσους)
- Οι μεταβλητές που αναφέρονται σε διαφορετικές χρονικές τιμές είναι στοχαστικά ανεξάρτητες

Έχοντας κάνει τις πιο πάνω παραδοχές μπορούμε να θεωρήσουμε ως τυχαία μεταβλητή τις ανελίξεις των ετήσιων ή των μηνιαίων τιμών (ή ακόμη και μικρότερων χρονικών κλιμάκων) και τις τιμές που παίρνει η πιο πάνω τ.μ. στα διάφορα υδρολογικά έτη να τις θεωρήσουμε ως διαφορετικές εμφανίσεις της ίδιας τ.μ. (Κουτσογιάννης, 1996 σ. 94).

Όμως οι παραδοχές αυτές δεν έχουν γενική ισχύ. Ανάλογα με τη φυσική διεργασία που μελετάται και αναλόγως τη χρονική κλίμακα που εξετάζουμε, οι παραπάνω παραδοχές πρέπει να επανεξετάζονται και αν χρειάζεται να αναθεωρούνται. Απλά στάσιμα μοντέλα, βασισμένα στις παραπάνω παραδοχές δεν είναι συχνά η καλύτερη επιλογή για την προσομοίωση των υδρολογικών διεργασιών λόγω των ιδιαιτεροτήτων που αυτές παρουσιάζουν. Οι πιο σημαντικές ιδιαιτερότητες των υδρολογικών διεργασιών οι οποίες σχετίζονται με τη χρονική εξέλιξη των φαινομένων είναι (Koutsoyiannis, 2005) :

- Εποχικότητα ή περιοδικότητα (seasonality)

Όταν η χρονική κλίμακα που μελετάμε είναι μικρότερη από την ετήσια, τότε η υδρολογική διεργασία δεν είναι δυνατό να αντιμετωπιστεί ως στάσιμη λόγω της επιρροής της εκάστοτε εποχής του έτους στις ιδιότητες του φαινομένου. Για παράδειγμα, η παραδοχή της στασιμότητας δεν ισχύει στην περίπτωση που η χρονοσειρά που εξετάζουμε είναι αυτή της μηνιαίας βροχόπτωσης γιατί τότε ανάλογα με την εποχή του έτους (φθινόπωρο, χειμώνας, άνοιξη, καλοκαίρι) αναμένονται και αντίστοιχες διακύμανσης στις τιμές της αντίστοιχής χρονοσειράς. Στην περίπτωση αυτή, χρειάζεται να τυποποιήσουμε τις τιμές του δείγματος (συνήθως η τυποποίηση γίνεται αφαιρώντας τη μέση τιμή και διαιρώντας με την τυπική απόκλιση της χρονοσειράς) επιδιώκοντας η νέα χρονοσειρά (τυποποιημένη) να είναι στάσιμη. Πρέπει να σημειωθεί όμως πως δεν είναι εφικτό να μετατρέψουμε μια μη στάσιμη χρονοσειρά σε στάσιμη. Η νέα χρονοσειρά που προκύπτει μετά την τυποποίηση μπορεί να διατηρεί τη μέση τιμή και την τυπική απόκλιση σταθερή με το χρόνο, αλλά δεν είναι σίγουρο πως θα διατηρούνται και οι υπόλοιπες στατιστικές ιδιότητες της χρονοσειράς όπως για παράδειγμα η αυτοσυσχέτιση και η κύρτωση. Στη περίπτωση αυτή λοιπόν, που μετά την τυποποίηση διατηρείται σταθερή μόνο η μέση τιμή και η τυπική απόκλιση, δεν έχουμε στάσιμη ανέλιξη αλλά κυκλοστάσιμη (cyclostationary) ή αλλιώς περιοδική.

- Μακρά μνήμη (long-term persistence)

Η μελέτη μεγάλων ιστορικών χρονοσειρών υδρολογικών και άλλων γεωφυσικών φαινομένων έχει δείξει πως η αυτοσυσχέτιση των τιμών του δείγματος είναι πολύ

σημαντική για μεγάλες χρονικές υστερήσεις (για παράδειγμα για υστέρηση 50 ή και 100 χρόνια). Ο πρώτος που παρατήρησε και μελέτησε αυτήν την τάση ήταν ο βρετανός μηχανικός Harold Edwin Hurst το 1950, στα πλαίσια της μελέτης του φράγματος του Ασουάν. Η έρευνα του Hurst βασίστηκε στις καταγεγραμμένες χρονοσειρές της στάθμης του Νείλου από το Νειλόμετρο στο νησί Roda στο Κάιρο (τη μεγαλύτερη καταγεγραμμένη υδρολογική χρονοσειρά). Παρατήρησε λοιπόν ο Hurst ότι τα έτη με μεγάλη απορροή τείνουν να ομαδοποιούνται όπως επίσης και τα έτη μειωμένων παροχών. Περιέγραψε αυτή τη φυσική συμπεριφορά ως εξής: «Αν και σε τυχαία γεγονότα ομάδες υψηλών και χαμηλών τιμών συμβαίνουν, η τάση τους να εμφανίζονται σε φυσικά γεγονότα είναι μεγαλύτερη. Αυτή είναι η κύρια διαφορά μεταξύ φυσικών και τυχαίων γεγονότων» (Hurst, 1951).

Παρόλο που το φαινόμενο αυτό παρατηρήθηκε το 1950 σε γεωφυσικές πραγματικές χρονοσειρές, η μαθηματική του περιγραφή είχε γίνει δέκα χρόνια νωρίτερα από τον Ρώσο μαθηματικό Andrey Nikolaevich Kolmogorov. Έτσι είναι πιο δόκιμη η αναφορά στο φαινόμενο αυτό με τον όρο «δυναμική Hurst – Kolmogorov». Στη συνέχεια της εργασίας θα χρησιμοποιούμε τη συντομογραφία «δυναμική ΗΚ» για να αναφερθούμε στο παραπάνω φαινόμενο.

Το ίδιο φαινόμενο αναφέρεται στην βιβλιογραφία και ως «φαινόμενο Ιωσήφ»[14]. Εναλλακτικές ονομασίες του φαινομένου αυτού είναι επίσης «φαινόμενο μακράς μνήμης», «εμμονή μακράς διάρκειας», «κλασματική κίνηση Brown» (Koutsoyiannis, 2006). Ο όρος «φαινόμενο μακράς μνήμης» δεν θεωρείται ιδιαίτερα εύστοχος, καθώς όπως αναφέρει ο (Koutsoyiannis, 2002), μια απλή και εύκολα κατανοητή εξήγηση του φαινομένου Hurst βασίζεται στην τυχαία διακύμανση και μεταβλητότητα των υδρομετεωρολογικών διεργασιών που παρουσιάζεται σε διάφορες χρονικές κλίμακες. Συγκεκριμένα, ο (Koutsoyiannis, 2002) υποστηρίζει, πως το φαινόμενο Hurst ουσιαστικά βασίζεται στην έλλειψη μνήμης, που παρουσιάζεται σε μεγάλες χρονικές κλίμακες, παρά στη λογική της μακράς μνήμης και αυτό γιατί είναι λογικότερο η φύση να «ξεχνάει» τα στατιστικά χαρακτηριστικά (π.χ. τη μέση τιμή) κάποιας υδρολογικής διεργασίας για μεγάλες τιμές τις υστέρησης (π.χ. 100 χρόνια) παρά να προσπαθεί να τα διατηρεί. Η εξήγηση αυτή βασίζεται στη διαπίστωση ότι το κλίμα μεταβάλλεται ακανόνιστα, για άγνωστους λόγους σε όλες τις χρονικές κλίμακες.

Η συμπεριφορά αυτή δεν απαντάται αποκλειστικά σε υδρολογικές χρονοσειρές, αλλά και σε άλλες γεωφυσικές διεργασίες όπως για παράδειγμα στην ένταση πνοής ανέμων, στις μέσες σημειακές και παγκόσμιες θερμοκρασίες, στις απορροές σε διάφορους ποταμούς κ.α. Το φαινόμενο Hurst, θεωρείται από πολλούς ως ένα από τα σημαντικότερα άλυτα προβλήματα της υδρολογίας(Mesa, Oscar J.; Poveda, German, 1993) και η παρουσία του αυξάνει δραματικά την αβεβαιότητα στις υδρολογικές αλλά και γενικότερα στις κλιματικές διεργασίες.

[14] Η ονομασία αυτή δόθηκε από τον Benoit Mandelbrot, και βασίζεται στο μύθο της παλαιάς διαθήκης που αναφέρει τις επτά ισχνές και επτά παχιές αγελάδες που επισκέφθηκαν τον Φαραώ στον ύπνο του.

Γίνεται λοιπόν φανερό πως η παραδοχή για ανεξάρτητα γεγονότα και ανεξάρτητες μεταβλητές, σε πολλές περιπτώσεις δεν ευσταθεί και δεν είναι ρεαλιστική σε σχέση με την δομή της συσχέτισης που εμφανίζεται στη φύση. Τα μοντέλα που προτείνονταν από τους (Box and Jenkins, 1976) και που δεν είχαν ουσιαστικά μνήμη (short memory ή short-range dependence) αφού η δομή της αυτοσυσχέτισης τους ήταν τέτοια που μειώνονταν πολύ γρήγορα με τον αντίστοιχο χρόνο υστέρησης, αποδείχτηκαν ακατάλληλα για την περιγραφή των υδρολογικών διεργασιών. Ακόμη και αν προσπαθήσουμε να προσομοιώσουμε μια διεργασία που παρουσιάζει μακρά μνήμη με μοντέλα ασθενής μνήμης (όπως για παράδειγμα μοντέλα ARMA ή ανελίξεις Markov) τα αποτελέσματα δεν θα είναι ικανοποιητικά καθώς θα απαιτείται η χρήση πολλών παραμέτρων, γεγονός που τα καθιστά πρακτικώς ακατάλληλα (Beran, 1992).

- Έλλειψη συνέχειας (intermittency)

Στις μικρές χρονικά κλίμακες (ημερήσια, ωριαία κ.α.) μερικές υδρολογικές διεργασίες παρουσιάζουν ασυνέχειες. Για παράδειγμα η τυχαία μεταβλητή που παριστάνει τη βροχόπτωση παίρνει είτε θετικές τιμές, όταν υπάρχει βροχόπτωση είτε μηδενικές, όταν δεν υπάρχει καθόλου βροχόπτωση. Η ιδιαιτερότητα αυτή έχει ως αποτέλεσμα τη δημιουργία ασυνέχειας στην περιθώρια κατανομή πιθανότητας του ύψους βροχής στο σημείο μηδέν. Αυτή η ασυνέχεια πρέπει να συμπεριληφθεί στο μοντέλο της προσομοίωσης ώστε τα αποτελέσματα που θα προκύψουν να είναι κοντά στην πραγματικότητα.

- Ασυμμετρία (skewness)

Η ασυμμετρία που παρουσιάζεται στη συνάρτηση κατανομής των χρονοσειρών των υδρολογικών ανελίξεων, είναι μία ακόμη ιδιαιτερότητά τους. Η ασυμμετρία είναι ιδιαίτερα έντονη στις μικρότερες χρονικά κλίμακες. Έτσι, ενώ άλλα φυσικά φαινόμενα περιγράφονται από κανονικές κατανομές, συμμετρικές με κωδωνοειδές σχήμα, οι υδρολογικές διεργασίες και ειδικότερα όταν αναφερόμαστε σε μικρές κλίμακες, παρουσιάζουν έντονη ασυμμετρία και πρέπει να λαμβάνεται υπόψη στην προσομοίωση.

Η ασυμμετρία που παρουσιάζεται στις υδρολογικές μεταβλητές γίνεται εύκολα κατανοητή με το παράδειγμα της βροχόπτωσης. Η βροχόπτωση, σαν φυσικό μέγεθος, έχει έντονα τυχαίο χαρακτήρα, παρουσιάζει μεγάλη διασπορά, έχει έντονη θετική ασυμμετρία και σχήμα πυκνότητας πιθανότητας τύπου ανεστραμμένου J. Αυτό συμβαίνει γιατί το μεγαλύτερο μέρος των τιμών της είναι κοντά στο μηδέν (δεν υπάρχει βροχόπτωση), ενώ παράλληλα εμφανίζονται και ακραίες θετικές τιμές (δηλαδή έντονη βροχόπτωση) με μικρή βέβαια πιθανότητα. Λαμβάνοντας επίσης υπόψη πως δεν είναι δυνατό να υπάρξουν αρνητικές τιμές, οδηγούμαστε σε θετικά ασύμμετρες κατανομές με μεγάλη «ουρά» προς τα δεξιά (Κουτσογιάννης, 1996 σ. 24).

# ΚΕΦΑΛΑΙΟ 3ο

## 3 ΣΥΝΗΘΕΣΤΕΡΕΣ ΜΕΘΟΔΟΙ ΣΥΜΠΛΗΡΩΣΗΣ ΜΕΤΡΗΣΕΩΝ

### 3.1 Εισαγωγή

Όπως έχουμε ήδη αναφέρει, το πρόβλημα της ύπαρξης ελλιπών τιμών είναι σύνηθες στις υδρομετεωρολογικές χρονοσειρές. Τα κενά που υπάρχουν στις χρονοσειρές δυσχεραίνουν την επεξεργασία των δεδομένων, οπότε είναι επιθυμητή η συμπλήρωσή τους. Η συμπλήρωση, ακόμα κι αν είναι πρόχειρη εκτίμηση με μεγάλα περιθώρια σφάλματος, μπορεί να είναι πολύ χρήσιμη, κυρίως όταν θέλουμε να εξαγάγουμε χρονοσειρές μικρότερης διακριτότητας. Λόγω της σημασίας της συμπλήρωσης αυτών των ελλείψεων έχουν προταθεί πολλές διαφορετικές μέθοδοι. Κοινό χαρακτηριστικό όλων των μεθόδων είναι ότι στηρίζονται στις υπάρχουσες μετρήσεις, είτε από τον ίδιο σταθμό είτε από γειτονικούς, προκειμένου να εκτιμήσουν τις τιμές που λείπουν.

Οι μέθοδοι συμπλήρωσης υπάγονται σε δύο γενικές κατηγορίες, τις εμπειρικές και τις στατιστικές. Οι στατιστικές μέθοδοι μπορούν αν αναλυθούν σε μεθόδους βασισμένες στην γραμμική παλινδρόμηση και σε μεθόδους βασισμένες σε στοχαστικά μοντέλα.

Η έννοια της συμπλήρωσης είναι πολύ κοντινή με την έννοια της επέκτασης[15] μιας υδρομετεωρολογικής χρονοσειράς. Τον όρο συμπλήρωση τον χρησιμοποιούμε όταν το δείγμα μας παρουσιάζει λίγα κενά (π.χ. 1 έως 3), τα οποία πρέπει να συμπληρωθούν ενώ τον όρο επέκταση όταν λείπουν συστηματικά πολλές συνεχόμενες τιμές. Αυτό που επιδιώκουμε με την επέκταση και τη συμπλήρωση είναι η απόκτηση ενός δείγματος μελέτης με μεγαλύτερο μήκος από το αρχικό και κατά συνέπεια μεγαλύτερη αξιοπιστία εκτιμήσεων.

---

[15] Η συμπλήρωση και η επέκταση είναι ουσιαστικά το ίδιο πράγμα, πρόκειται δηλαδή για την εκτίμηση τιμών που δεν υπάρχουν, αλλά έχουν διαφορετικούς στόχους και γίνονται με διαφορετικές μεθόδους. Η συμπλήρωση γίνεται σε λίγες τιμές, με στόχο τη διευκόλυνση εξαγωγής χρονοσειρών μικρότερης διακριτότητας, και η ακρίβεια της εκτίμησης δεν έχει μεγάλη σημασία ενώ η επέκταση γίνεται γιατί η επεκτεταμένη χρονοσειρά μπορεί, σε σχέση με την αρχική, να οδηγεί σε καλύτερη εκτίμηση των στατιστικών παραμέτρων ή να έχει χαρακτηριστικά πιο κοντά στα χαρακτηριστικά της διεργασίας που εκπροσωπεί.

Παρά τη μεγάλη ποικιλία σε μεθόδους συμπλήρωσης ελλιπών δεδομένων, οι περισσότερες μέθοδοι θα μπορούσαν να περιγραφούν από μια απλή γραμμική σχέση

$$Y = w_1 x_1 + w_2 x_2 + ... + w_n x_n + e \tag{3.1}$$

όπου

$w_j$ είναι μια αριθμητική σταθερά (συντελεστής βαρύτητας)

$e$ είναι το σφάλμα της εκτίμησης

Όλες οι μέθοδοι συμπλήρωσης εκτιμούν τους διάφορους συντελεστές βαρύτητας $w_j$. Οι στατιστικές μέθοδοι υπερτερούν έναντι των απλών εμπειρικών μεθόδων διότι δίνουν περισσότερες πληροφορίες για το σφάλμα της εκτίμησης $e$ (όπως για παράδειγμα τη μέση τιμή και την τυπική απόκλιση του σφάλματος $\mu_e := E[e]$ και $\sigma_e := \sqrt{\mathrm{Var}[e]}$.

Στη συνέχεια παρουσιάζονται οι πιο διαδεδομένες μέθοδοι συμπλήρωσης υδρομετεωρολογικών χρονοσειρών και περιγράφονται τα θετικά και τα αρνητικά στοιχεία κάθε μεθόδου.

## 3.2 Απλές – εμπειρικές μέθοδοι

Οι μέθοδοι που περιγράφονται σε αυτό το σημείο, παρέχουν πρόχειρες εκτιμήσεις των ελλιπών τιμών και το βασικό τους πλεονέκτημα είναι η γρήγορη και απλή τους εφαρμογή χωρίς πολλές αριθμητικές πράξεις. Στην περίπτωση λοιπόν που οι ελλείψεις είναι σποραδικές και αφορούν μικρές χρονικές περιόδους (για παράδειγμα μερικές μέρες μέχρι λίγους μήνες) συχνά χρησιμοποιούνται για τη συμπλήρωση αυτών των κενών, κάποια από τις ακόλουθες απλές εμπειρικές μεθόδους:

- Η μέθοδος της μέσης τιμής

Πρόκειται για την απλούστερη μέθοδο συμπλήρωσης, κατά την οποία η τιμή που λείπει συμπληρώνεται με το μέσο όρο των υπαρχουσών τιμών της ελλιπούς χρονοσειράς. Η μέση τιμή μπορεί να χρησιμοποιηθεί για τη συμπλήρωση ελλειπουσών τιμών όταν δεν μπορεί να βρεθεί κάποια συσχέτιση (συσχέτιση είτε με τις τιμές της ίδιας της χρονοσειράς, είτε με άλλους γειτονικούς σταθμούς) που θα βοηθούσε στη συμπλήρωση των κενών.

Πιο συγκεκριμένα, η μέθοδος της μέσης τιμής εκφράζεται ως εξής

$$Y = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.2}$$

όπου

$Y$ η τιμή που λείπει από τη χρονοσειρά

$n$ το μέγεθος του δείγματος της χρονοσειράς

$x_i$ οι υπάρχουσες τιμές της χρονοσειράς

Όπως είναι φανερό, η σχέση (3.2) προκύπτει από την (3.1), αν θεωρήσουμε για όλες τις τιμές $x_i$ βάρη ίσα με $\frac{1}{n}$, δηλαδή $w_j=\frac{1}{n}$ για κάθε $j$. Έτσι στη περίπτωση αυτή συμπληρώνουμε την ελλείπουσα τιμή της χρονοσειράς με τον αριθμητικό μέσο των μετρήσεων.

Η μέθοδος της μέσης τιμής, αν και γρήγορη και απλή στην εφαρμογή, δεν προσφέρεται για τη συμπλήρωση χρονοσειρών με πολλά κενά, γιατί τότε αλλοιώνονται σημαντικά τα στατιστικά χαρακτηριστικά της χρονοσειράς που περιγράφουν τη φυσική διεργασία. Πιο συγκεκριμένα, η εφαρμογή της μέσης τιμής έχει ως αποτέλεσμα την υποεκτίμηση της τυπικής απόκλισης του συμπληρωμένου δείγματος. Επίσης, για μηνιαίες ή ημερήσιες μετρήσεις, η τιμή που συμπληρώνει το κενό πρέπει να είναι η μέση μηνιαία ή η μέση ημερήσια αντίστοιχα για το συγκεκριμένο μήνα ώστε να λαμβάνεται υπόψη η περιοδικότητα.

- Μέθοδος των κανονικών λόγων

Πρόκειται για μια γενίκευση της μεθόδου του αριθμητικού μέσου με την διαφορά ότι για τη συμπλήρωση της ελλείπουσας τιμής λαμβάνονται υπόψη οι μετρήσεις γειτονικών σταθμών, οι οποίες σταθμίζονται με βάση τις αναλογίες των ετήσιων τιμών τις τ.μ., σύμφωνα με τον τύπο :

$$h_Y = \frac{1}{n}\sum_{i=1}^{n}\frac{H_Y}{H_i}h_i \tag{3.3}$$

όπου

$H_Y$ και $H_i$ οι μέσες ετήσιες τιμές (συχνά αποκαλούμενες και κανονικές, εξ ου και η ονομασία της μεθόδου) του σταθμού $Y$ που θέλουμε να συμπληρώσουμε τις ελλείπουσες τιμές και του γειτονικού σταθμού $i$ αντίστοιχα.

Η μέθοδος των κανονικών λόγων είναι πολύ απλή και ευρέως διαδεδομένη. Δίνει συντελεστή βαρύτητας σε κάθε γειτονικό σταθμό. Συνήθως επιλέγονται τρείς σταθμοί, οι οποίοι απέχουν περίπου ίσες αποστάσεις και περιβάλλουν τον υπό συμπλήρωση σταθμό. Η μέθοδος αυτή έχει ευρεία εφαρμογή στη συμπλήρωση χρονοσειρών βροχοπτώσεων. Τα ύψη βροχής των γειτονικών βροχομετρικών σταθμών σταθμίζονται με βάση τις αναλογίες των μέσων ετήσιων βροχοπτώσεων σύμφωνα με την παραπάνω σχέση.

- Μέθοδος της αντίστροφης απόστασης

Και αυτή η μέθοδος είναι μια γενίκευση της μεθόδου του αριθμητικού μέσου με τη διαφορά ότι εδώ λαμβάνονται υπόψη για τη στάθμιση των επιμέρους υψών βροχής τα αντίστροφα των αποστάσεων των σταθμών, υψωμένα σε κατάλληλη δύναμη. Η τιμή λοιπόν που λείπει στο σταθμό $Y$ υπολογίζεται από τη σχέση

$$h_Y = \sum_{i=1}^{n} w_i \, h_i \tag{3.4}$$

όπου ο συντελεστής βάρους $w_i$ δίνεται από τη σχέση

$$w_i = \frac{d_i^{-b}}{\sum_{i=1}^{n} d_i^{-b}} \tag{3.5}$$

όπου $d_i$ είναι η απόσταση του σταθμού $i$ από το σταθμό $Y$ και $b$ είναι σταθερά που κατά κανόνα λαμβάνεται ίση με 2.

## 3.3 Μέθοδοι βασισμένες στη γραμμική παλινδρόμηση

Παλινδρόμηση ονομάζεται κάθε συσχέτιση η οποία στηρίζεται στη μέθοδο των ελαχίστων τετραγώνων και πρακτικά ταυτίζεται με τη συσχέτιση, αφού η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείτε σχεδόν σε όλες τις περιπτώσεις. Η συσχέτιση μεταξύ διαφορετικών χρονοσειρών μπορεί να μας δώσει χρήσιμα συμπεράσματα σχετικά με τον βαθμό της φυσικής συσχέτισης των δεδομένων (π.χ. λόγω γεωγραφικής θέσης).

Η παλινδρόμηση εξετάζει λοιπόν τη συσχέτιση που παρουσιάζεται στις παρατηρήσεις του προς συμπλήρωση σταθμού με κάποιους γειτονικούς σταθμούς αναφοράς (έναν ή και περισσότερους). Η τυπική χρήση της παλινδρόμησης στην τεχνική υδρολογία αφορά στη συμπλήρωση των ελλείψεων ενός υδρολογικού δείγματος με βάση κάποιο πληρέστερο δείγμα, το οποίο συσχετίζεται με το πρώτο, ή και την επέκταση του πρώτου δείγματος.

Στην τεχνική υδρολογία χρησιμοποιείται ευρύτατα η γραμμική παλινδρόμηση. Η επιλογή της γραμμικής παλινδρόμησης δεν γίνεται μόνο για λόγους απλότητας, αλλά και γιατί αποτελεί τη βέλτιστη παλινδρόμηση για μεταβλητές που ακολουθούν κανονική κατανομή. Η κανονική κατανομή είναι πολύ διαδεδομένη γιατί συχνά είναι κατάλληλη για μεταβλητές που αναφέρονται σε μεγάλες χρονικές κλίμακες ώστε σε κάθε διάστημα να αντιστοιχεί μεγάλος αριθμός υδρολογικών επεισοδίων. Η καταλληλότητα της κανονικής κατανομής εξηγείται από το κεντρικό οριακό θεώρημα[16].

Η μέθοδος της γραμμικής παλινδρόμησης είναι κατάλληλη για παράδειγμα, για τη συμπλήρωση ετήσιων υψών βροχής ή και μηνιαίων τιμών. Για χρονικές περιόδους

[16] Το κεντρικό οριακό θεώρημα συνοψίζεται ως εξής: το άθροισμα ανεξάρτητων και όμοια κατανεμημένων τυχαίων μεταβλητών τείνει στην κανονική κατανομή όσο το άθροισμα των προσθετέων τείνει στο άπειρο.

μικρότερες του μήνα, η μέθοδος δεν θεωρείται κατάλληλη, γιατί συνήθως ο συντελεστής συσχέτισης παίρνει αρκετά χαμηλές τιμές.

Υπάρχουν δυο βασικές κατηγορίες παλινδρόμησης στην υδρολογία: η απλή και η πολλαπλή. Απλή παλινδρόμηση είναι αυτή που χρησιμοποιεί ένα μοναδικό δείγμα ως δείγμα αναφοράς, ενώ πολλαπλή αυτή που ως δείγματα αναφορά έχει περισσότερα από ένα. Στη συνέχεια θα αναφερθούμε στην απλή γραμμική παλινδρόμηση και στις διάφορες κατηγορίες αυτής.

Πρέπει να τονίσουμε πως στις μεθόδους που περιγράφονται παρακάτω, θα πρέπει να λαμβάνεται υπόψη η εποχικότητα, που είναι βασικό χαρακτηριστικό των υδρολογικών διεργασιών. Αν για παράδειγμα η χρονοσειρά που εξετάζεται είναι ημερήσια ή μηνιαία, θα πρέπει ουσιαστικά να θεωρήσουμε δώδεκα διαφορετικές χρονοσειρές μία για όλους τους Ιανουαρίους, μία για τους Φεβρουαρίου, κτλ, και να γίνουν δώδεκα γραμμικές παλινδρομήσεις. Επίσης, για να πραγματοποιηθεί η παλινδρόμηση, πρέπει το σύνολο των χρονοσειρών που θα χρησιμοποιηθούν, ανεξάρτητες και εξαρτημένες, να έχουν κοινή περίοδο μετρήσεων, δηλαδή σύνολο χρονικών στιγμών για τις οποίες υπάρχουν τιμές σε όλες τις χρονοσειρές.

Στη συνέχεια παρουσιάζονται οι βασικοί τύποι απλής γραμμικής παλινδρόμησης που έχουν ευρεία εφαρμογή στην υδρολογία.

- Ομογενής γραμμική παλινδρόμηση (ή παλινδρόμηση με μηδενικό σταθερό όρο)

Αν $Y$ η τιμή που λείπει από τη χρονοσειρά και $X_i$ οι υπάρχουσες μετρήσεις τότε με βάση την ομογενή γραμμική παλινδρόμηση θα έχουμε

$$Y = wX \tag{3.6}$$

όπου $w = \dfrac{E[YX]}{E\left[X^2\right]} = \dfrac{\sigma_{XY} + \mu_X \mu_Y}{\sigma_X^2 + \mu_X^2}$ (αποτελεί μεροληπτική λύση)

Η χρήση της ομογενούς γραμμικής παλινδρόμησης (ομογενής ευθεία) για επέκταση του δείγματος, θα πρέπει γενικά να αποφεύγεται γιατί οδηγεί σε μεροληψία ως προς την εκτίμηση της μέσης τιμής αλλά και της διασποράς. Ωστόσο, στη περίπτωση συμπλήρωσης λίγων τιμών, οι οποίες δεν επηρεάζουν τα στατιστικά χαρακτηριστικά του δείγματος μπορεί η μέθοδος αυτή να είναι κατάλληλη.

- Απλή γραμμική παλινδρόμηση (χωρίς όρο σφάλματος)

Η απλή γραμμική παλινδρόμηση είναι η γνωστή μεθοδολογία βέλτιστης προσαρμογή μίας ευθείας στο καρτεσιανό επίπεδο όπου διατάσσονται διάφορα σημεία. Η απλή γραμμική παλινδρόμηση βασίζεται στην αρχή των ελάχιστων τετραγώνων, δηλαδή δίνεται μία λύση η οποία ελαχιστοποιεί το άθροισμα των τετραγώνων των αποστάσεων των σημείων από την ευθεία. Στην απλή γραμμική παλινδρόμηση οι τιμές της ελλιπούς χρονοσειράς $Y$ συσχετίζονται γραμμικά με τις τιμές μιας άλλης χρονοσειράς αναφοράς $X$ που μπορεί να έχει ίδια ή σχετικά

δεδομένα. Για παράδειγμα, μια χρονοσειρά από παροχές μπορεί να συσχετιστεί με μια χρονοσειρά βροχοπτώσεων.

Σύμφωνα λοιπόν με τη γραμμική παλινδρόμηση, η προς συμπλήρωση τιμή $Y$, εκτιμάται από την αντίστοιχη τιμή $X$ του γειτονικού σταθμού (για την περίοδο όπου σημειώνεται η έλλειψη στον υπό εξέταση σταθμό) με βάση τη σχέση

$$Y = a + bX \tag{3.7}$$

όπου $a$ και $b$ παράμετροι που εκτιμώνται ώστε να ελαχιστοποιείτε το τετραγωνικό σφάλμα της εκτίμησης. Αν $x_i$ και $y_i$ ταυτόχρονες μετρήσεις στους σταθμούς $X$ και $Y$ αντίστοιχα, τη χρονική στιγμή $i$ τότε

$$b = \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3.8}$$

και

$$a = \bar{y} - b\bar{x} \tag{3.9}$$

όπου $\bar{x}, \bar{y}$ οι μέσες τιμές των $x_i$ και $y_i$ αντίστοιχα, δηλαδή

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \end{aligned} \tag{3.10}$$

και $n$ το μήκος του δείγματος.

Ο βαθμός καταλληλότητας της μεθόδου για τα συγκεκριμένα δεδομένα αποδίδεται από το μέγεθος

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{3.11}$$

ή ισοδύναμα

$$r = \frac{n\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \tag{3.12}$$

Το μέγεθος αυτό ονομάζεται συντελεστής γραμμής συσχέτισης και οι τιμές του κυμαίνονται στο διάστημα [-1,1]. Όσο πιο κοντά στο -1 ή στο 1 βρίσκεται η τιμή του

*r*, τόσο ισχυρότερη είναι η συσχέτιση. Η μηδενική τιμή του συντελεστή συσχέτισης εκφράζει ανυπαρξία συσχέτισης.

Για να είναι στατιστικά σημαντική η συσχέτιση θα πρέπει ο συντελεστής *r* να είναι σε απόλυτη τιμή μεγαλύτερος από την κρίσιμη τιμή

$$r_c \approx \frac{2}{\sqrt{n}} \tag{3.13}$$

Θετική τιμή του συντελεστή συσχέτισης δείχνει ότι η αύξηση στην τιμή του *x* συνδέεται με την αύξηση στην τιμή του *y*. Αντίθετα, η αρνητική τιμή του συντελεστή συσχέτισης δείχνει ότι η αύξηση στην τιμή του *x* συνδέεται με τη μείωση στην τιμή του *y*. Στην περίπτωση βέβαια των υδρολογικών μεταβλητών, έχει κυρίως νόημα η θετική τιμή του συντελεστή συσχέτισης.

Ωστόσο, η μέθοδος της απλής γραμμικής παλινδρόμησης έχει το μειονέκτημα ότι το διευρυμένο δείγμα που παράγεται με την εφαρμογή της μεθόδου δίνει μεν αμερόληπτη εκτίμηση της μέσης τιμής αλλά η εκτίμηση της διασποράς είναι μεροληπτική.

- Οργανική συσχέτιση

Η οργανική συσχέτιση εφαρμόζεται κυρίως στην επέκταση δειγμάτων όπου είναι επιθυμητή η διατήρηση των στατιστικών χαρακτηριστικών του αρχικού (προ της συμπλήρωσης) δείγματος. Στη μέθοδο της οργανικής συσχέτισης, δεν υπάρχει η απαίτηση για ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος, όπως συμβαίνει στη περίπτωση της απλής γραμμικής παλινδρόμησης, αντίθετα, εκτιμώνται οι παράμετροι της συσχέτισης διατηρώντας την αρχική μέση τιμή και διασπορά.

Η χρήση της οργανικής συσχέτισης είναι ίδια όπως και στη περίπτωση της γραμμικής παλινδρόμησης χωρίς όρο σφάλματος. Η μόνη διαφορά τους είναι ο τρόπος εκτίμησης της παραμέτρου *b*.

Συγκεκριμένα, η παράμετρος *b* δίνεται από το τύπο

$$b = \mathrm{sgn}(\rho_{XY}) \frac{\sigma_X}{\sigma_Y} \tag{3.14}$$

όπου $\mathrm{sgn}(\rho_{XY})$ είναι το πρόσημο του συντελεστή συσχέτισης (+1 ή -1).

Ο όρος του προσήμου έχει τεθεί για να είναι συνεπής η εκτίμηση με την πραγματικότητα, δηλαδή για θετικά συσχετισμένες μεταβλητές να προκύπτει θετική τιμή του συντελεστή *b* και αντίστροφα.

Σε αναλογία με όσα αναφέρθηκαν στη περίπτωση της απλής γραμμικής παλινδρόμησης, για να έχει νόημα η εφαρμογή της οργανικής συσχέτισης θα πρέπει ο συντελεστής *r* να είναι σε απόλυτη τιμή μεγαλύτερος από την κρίσιμη τιμή:

$$|r_C| \geq \max\left\{0.5, \frac{2}{\sqrt{n}}\right\} \tag{3.15}$$

- Απλή γραμμική παλινδρόμηση με τυχαίο όρο (όρο σφάλματος)

Όταν στη γραμμική παλινδρόμηση, όπως στην περίπτωση της απλής γραμμικής παλινδρόμησης, δεν χρησιμοποιείται τυχαίος όρος, οι εκτιμημένες τιμές έχουν μεν τη μεγαλύτερη πιθανότητα να είναι πλησιέστερα στις πραγματικές, αλλά η διασπορά του συμπληρωμένου δείγματος θα έχει υποεκτιμηθεί. Για την αντιμετώπιση αυτού του προβλήματος, προσθέτουμε στη σχέση της απλής γραμμικής παλινδρόμησης χωρίς όρο σφάλματος, ένα τυχαίο όρο $\varepsilon$ και έχουμε

$$Y = a + bX + \varepsilon \tag{3.16}$$

με

$$\varepsilon = Z\,\sigma_Y\,a\sqrt{1-\rho^2} \tag{3.17}$$

όπου :

$Z$: ασυσχέτιστη κανονική τυχαία μεταβλητή με μέση τιμή 0 και διασπορά 1

$\rho$: συντελεστής συσχέτισης μεταξύ $y$ και $x$

$\alpha$: συντελεστής εξαρτώμενος από το πλήθος των τιμών των χρονοσειρών

Στην εφαρμογή της μεθόδου, όπως φαίνεται και από την προηγούμενη σχέση, ο όρος του σφάλματος δεν εκτιμάται, αλλά γεννάται ή προσομοιώνεται. Χρησιμοποιείται δηλαδή μια γεννήτρια τυχαίων αριθμών $Z$ η οποία παράγει τις τυχαίες τιμές του σφάλματος $\varepsilon$ (Κουτσογιάννης, 1996 σσ. 231-232). Πρέπει όμως να τονίσουμε πως η εισαγωγή τυχαίου όρου δεν έχει νόημα για μικρό αριθμό ελλείψεων.

Η τεχνική της εισαγωγής του τυχαίου όρου μπορεί να χρησιμοποιηθεί εναλλακτικά της οργανικής συσχέτισης, μάλιστα διατηρώντας καλύτερα τον αρχικό συντελεστή συσχέτισης. Η μεθοδολογία έχει το μειονέκτημα της μη επαναληψιμότητας των αποτελεσμάτων για διαφορετικές γεννήτριες τυχαίων αριθμών ή διαφορετικών συνθηκών αρχικοποίησης.

Στη βιβλιογραφία υπάρχουν και άλλες μέθοδοι γραμμικής παλινδρόμησης, όπως για παράδειγμα πολλαπλή γραμμική παλινδρόμηση, όμως δεν κρίνεται σκόπιμο να αναλυθούν όλες οι υπάρχουσες μέθοδοι σε αυτό το σημείο.

## 3.4 Μέθοδοι βασισμένες σε στοχαστικά μοντέλα

Μια άλλη μεθοδολογία, η οποία και αυτή ουσιαστικά αποτελεί γενίκευση της γραμμικής παλινδρόμησης, βασίζεται στη χρήση στοχαστικών μοντέλων για τη συμπλήρωση των κενών που παρουσιάζονται στις χρονοσειρές (Salas, 1993 σσ. 19.43-19.48).

- Χρήση μοντέλου AR(1) και PAR(1)

Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν για τη συμπλήρωση ελλιπών τιμών από χρονοσειρές όταν δεν υπάρχουν διαθέσιμα δεδομένα από γειτονικούς σταθμούς.

Έστω χρονοσειρά που περιγράφεται από ένα μοντέλο AR(1):

$$Y_t = \mu_Y + \varphi(Y_{t-1} - \mu_Y) + \varepsilon_t \tag{3.18}$$

όπου οι παράμετροι του μοντέλου μπορούν να εκτιμηθούν από τα διαθέσιμα δεδομένα.

Αν λείπει μια τιμή $Y_t$, αλλά είναι γνωστή η προηγούμενη τιμή, $Y_{t-1}$, τότε η $Y_t$ μπορεί να εκτιμηθεί από τη πιο πάνω σχέση. Το παραπάνω μοντέλο όμως, προϋποθέτει στάσιμη χρονοσειρά. Σε περίπτωση που έχουμε μη στάσιμη χρονοσειρά, τότε θα πρέπει να την κάνουμε στάσιμη πριν εφαρμόσουμε το πιο πάνω μοντέλο.

Στη περίπτωση που η χρονοσειρά που εξετάζουμε είναι μηνιαία ή ημερήσια, θα πρέπει να ληφθεί υπόψη η εποχικότητα. Δηλαδή, θα έχουμε 12 παραμέτρους $\varphi$ και 12 μέσες τιμές $\mu_Y$, μία για κάθε μήνα. Για κάθε ελλείπουσα τιμή θα χρησιμοποιείται και το κατάλληλο ζεύγος παραμέτρων. Το μοντέλο αυτό ονομάζεται PAR(1).

Είναι φανερό πως το μοντέλο AR(1) ή PAR(1) προκύπτει από γραμμική παλινδρόμηση της ελλιπούς χρονοσειράς με την ίδια χρονοσειρά αλλά υστερημένη κατά 1.

Η χρήση του πιο πάνω μοντέλου είναι κατάλληλη μόνο όταν δεν υπάρχουν γειτονικοί σταθμοί με διαθέσιμες χρονοσειρές, που να μπορεί να συσχετιστεί η ελλιπής χρονοσειρά. Επίσης, ο Salas (1993), αναφέρει πως το πιο πάνω μοντέλο δεν επαρκεί για τη συμπλήρωση των κενών στις χρονοσειρές, όταν λείπουν διαδοχικές τιμές διότι τότε όλες οι τιμές που συμπληρώνονται, εκτός της πρώτης, θα είναι συσχετιζόμενες με τις προηγούμενές τους με αποτέλεσμα να πολλαπλασιάζεται το σφάλμα.

- Πολύμεταβλητά μοντέλα

Όταν κάνουμε συσχέτιση της χρονοσειράς αναφοράς (δηλαδή της χρονοσειράς που θα χρησιμοποιήσουμε για τη συμπλήρωση) και παράλληλα χρησιμοποιήσουμε και την αυτοσυσχέτιση, τότε έχουμε ένα πολυμεταβλητό μοντέλο.

Έστω $Y_t$ οι τιμές της ελλιπούς χρονοσειράς και $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(n)}$ οι τιμές των χρονοσειρών αναφοράς, τότε η γενική μορφή του μοντέλου είναι

$$Y_t = a + \sum_{j=1}^{p} b_j Y_{t-j} + \sum_{j=0}^{p_1} b_j^{(1)} X_{t-j}^{(1)} + \sum_{j=0}^{p_2} b_j^{(2)} X_{t-j}^{(2)} + \ldots + \varepsilon_t \tag{3.19}$$

όπου

$a, b_j, j = 1, \ldots, p$ και $b_j^{(1)}, j = 0, \ldots, p_1$ και $b_j^{(2)}, j = 0, \ldots, p_2$ είναι παράμετροι που μπορούν να εκτιμηθούν από τα διαθέσιμα δεδομένα.

Συχνά, η συσχέτιση της ελλιπούς χρονοσειράς με τις χρονοσειρές αναφοράς με υστέρηση μεγαλύτερη από 0, είναι πολύ μικρή για χρονικές κλίμακες ίσες ή μεγαλύτερες της ημέρας όπως επίσης και η αυτοσυσχέτιση για υστέρηση μεγαλύτερη του 1. Έτσι χρησιμοποιούμε το πιο πάνω πολυμεταβλητό μοντέλο με τις εξής παραμέτρους $p = 1$ και $p_1 = p_2 = ... = 0$. Άρα η (3.19) γίνεται

$$Y_t = a + b_1 Y_{t-1} + b_0^{(1)} X_t^{(1)} + b_0^{(2)} X_t^{(2)} + ... + \varepsilon_t \tag{3.20}$$

Η προηγούμενη σχέση αποτελεί έκφραση της πολλαπλής γραμμικής παλινδρόμησης.

## 3.5 Μέτρα αξιολόγησης μεθόδων συμπλήρωσης χρονοσειρών

Προκειμένου να αξιολογήσουμε τις διάφορες μεθόδους συμπλήρωσης ελλιπών υδρομετεωρολογικών δεδομένων, ώστε να αποφανθούμε για το ποία είναι αυτή που δίνει τα καλύτερα αποτελέσματα, δηλαδή την καλύτερη εκτίμηση της ελλείπουσας τιμής, μπορούμε να χρησιμοποιήσουμε διάφορα μέτρα σφάλματος, που προτείνονται στη βιβλιογραφία.

Τα πιο ευρέως χρησιμοποιήσιμα είναι:

- το σφάλμα πρόγνωσης (forecast error)
  Ορίζεται ως η διαφορά της πραγματικής τιμής $x_t$ (της τιμής δηλαδή που έχει παρατηρηθεί) από την αντίστοιχη εκτιμημένη τιμή $\hat{x}_t$ (της αντίστοιχης δηλαδή τιμής που έχει εκτιμηθεί με την χρήση κάποιας μεθόδου συμπλήρωσης). Συγκεκριμένα,
  $$e_t = x_t - \hat{x}_t \tag{3.21}$$
- μέση απόκλιση σφάλματος (mean deviation, MD)
  Εκφράζει τη μεροληψία (bias) της μεθόδου εκτίμησης και δίνεται από το τύπο
  $$\mathrm{MD} = \frac{1}{n}\sum_{t=1}^{n} e_t \tag{3.22}$$
- μέση απόλυτη απόκλιση (mean absolute deviation, MAD)
  $$\mathrm{MAD} = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{3.23}$$
- μέσο τετραγωνικό σφάλμα (mean square error, MSE)
  $$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2 = \frac{1}{n}\sum_{t=1}^{n} (x_t - \hat{x}_t)^2 \tag{3.24}$$

- τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (root mean square error, RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (x_t - \hat{x}_t)^2} \tag{3.25}$$

- μέσο ποσοστιαίο σφάλμα (mean percent error, MPE)

$$\mathrm{MPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{e_t}{x_t} \tag{3.26}$$

- μέσο απόλυτο ποσοστιαίο σφάλμα (mean absolute percent error, MAPE)

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{\left|e_t\right|}{x_t} \tag{3.27}$$

Οι βασικές ιδιότητες μιας καλής πρόβλεψης είναι

- η αμεροληψία (unbiasedness)
  $\mathrm{E}\left[X_n(k)\right] = X_{n+k}$
- η αποτελεσματικότητα (efficiency)
  $\mathrm{Var}\left[\varepsilon_n(k)\right] = \mathrm{Var}\left[X_{n+k} - X_n(k)\right]$

Ο συνδυασμός της αμεροληψίας με την αποτελεσματικότητα γίνεται στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος πρόβλεψης, έτσι στη συνέχεια της εργασίας, θα χρησιμοποιήσουμε σαν μέτρο σύγκρισης των διαφόρων μεθόδων το μέσο τετραγωνικό σφάλμα της εκτίμησης (MSE). Βέλτιστη κάθε φορά θα θεωρούμε τη μέθοδο εκείνη που ελαχιστοποιεί το MSE.

# ΚΕΦΑΛΑΙΟ 4ο

## 4 ΠΡΟΤΕΙΝΟΜΕΝΗ ΜΕΘΟΔΟΛΟΓΙΑ ΣΥΜΠΛΗΡΩΣΗΣ ΜΕΤΡΗΣΕΩΝ

### 4.1 Εισαγωγή

Όπως διατυπώθηκε στο προηγούμενο κεφάλαιο, οι βασικές μέθοδοι που χρησιμοποιούνται στην υδρολογία για τη συμπλήρωση των ελλειπουσών τιμών που παρουσιάζονται στις χρονοσειρές μπορούν να περιγραφούν με μια απλή γραμμική σχέση. Οι περισσότερες μέθοδοι που προτείνονται στη βιβλιογραφία βασίζονται σε ένα σταθμισμένο μέσο όρο όλων των υπαρχουσών παρατηρήσεων της χρονοσειράς που μελετάται. Δηλαδή, για να εκτιμηθούν οι τιμές που λείπουν χρησιμοποιείται ως χρονοσειρά αναφοράς είτε η υπάρχουσα ελλιπής χρονοσειρά (δηλαδή η χρονοσειρά που μελετάται και που παρουσιάζει τα κενά στις μετρήσεις), είτε πιο συχνά συσχετίζεται η ελλιπής χρονοσειρά με κάποιους γειτονικούς σταθμούς με γραμμική παλινδρόμηση και χρησιμοποιούνται ως χρονοσειρές αναφοράς αυτές των γειτονικών σταθμών, με κάποιους συντελεστές βαρύτητας. Τέτοιες μέθοδοι, εκτός από αυτή της αντίστροφης απόστασης που παρουσιάστηκε προηγούμενα στην παράγραφο 3.2, είναι τα πολύγωνα Thiessen, η μέθοδος BLUE καθώς επίσης και η πολύ δημοφιλής γεωστατιστική μέθοδος kriging[17] και οι παραλλαγές της.

Παρατηρούμε λοιπόν πως υπάρχουν δύο διαφορετικές προσεγγίσεις για τη συμπλήρωση υδρομετεωρολογικών χρονοσειρών:

- συμπλήρωση των κενών που παρουσιάζονται στις χρονοσειρές με βάση μετρήσεις από γειτονικούς σταθμούς, ανάλογα με τη συσχέτιση που παρουσιάζεται μεταξύ των σταθμών, και
- συμπλήρωση της χρονοσειράς με χρήση μόνο της υδρολογικής πληροφορίας που προέρχεται από αυτή καθ' αυτή τη χρονοσειρά. Δηλαδή με βάση τις υπάρχουσες τιμές της χρονοσειράς που εξετάζεται.

Στα πλαίσια αυτής της διπλωματικής εργασίας, λαμβάνοντας υπόψη τις ιδιαιτερότητες που παρουσιάζουν οι υδρολογικές διεργασίες, και που αναλύθηκαν εκτενώς σε προηγούμενο κεφάλαιο, θα εξετάσουμε μια εναλλακτική μεθοδολογία για τη συμπλήρωση των ελλειπουσών τιμών των υδρομετεωρολογικών χρονοσειρών που βασίζεται στη συμπλήρωση των κενών με χρήση ως χρονοσειρά αναφοράς την

[17] Αναλυτικά η μέθοδος kriging καθώς και εφαρμογή της στις υδρομετεωρολογικές διεργασίες θα παρουσιαστεί στη συνέχεια στο υποκεφάλαιο 4.3.5.

ελλιπή χρονοσειρά, εντάσσεται δηλαδή στη δεύτερη προσέγγιση που αναφέρεται πιο πάνω.

Εξετάζουμε λοιπόν το πρόβλημα της συμπλήρωσης των κενών που παρουσιάζονται σε μια χρονοσειρά βασιζόμενοι μόνο στις τιμές της ίδιας χρονοσειράς, χωρίς δηλαδή χρήση υδρολογικών πληροφοριών από γειτονικούς σταθμούς. Το πρόβλημα αυτό είναι σύνηθες στην υδρολογία, και ιδιαίτερα στον Ελλαδικό χώρο, όπου οι υδρολογικοί σταθμοί δεν είναι πυκνά τοποθετημένοι, έτσι για την συμπλήρωση των μετρήσεων του σταθμού δεν μπορεί να χρησιμοποιηθεί κάποιος γειτονικός σταθμός, αφού δεν θα βρίσκεται σε κοντινή γεωγραφικά απόσταση. Επίσης, πολλές φορές είναι αδύνατη η συσχέτιση ενός σταθμού με κάποιον άλλο γειτονικό του, λόγω των διαφορετικών κλιματολογικών συνθηκών που επικρατούν σε κάθε περιοχή ή λόγω της παρεμβολής κάποιου ορεινού σχηματισμού. Στη περίπτωση επίσης μελέτης και ανάλυσης παλαιοκλιματικών δεδομένων ή δεδομένων που προέρχονται από τον $19^{ο}$ αιώνα ή ακόμη και το πρώτο μισό του περασμένου αιώνα η ανάγκη συμπλήρωσης των ελλιπών τιμών με χρήση ως χρονοσειρά αναφοράς την ελλιπή χρονοσειρά, είναι σύνηθες φαινόμενο.

### *4.1.1 Παράδοξο: ολικός μέσος όρος έναντι τοπικού;*

Κινητήριως δύναμη για την ανάπτυξη αυτής της νέα μεθοδολογίας συμπλήρωσης, ήταν η παρατήρηση του εξής παράδοξου στις υπάρχουσες μεθόδους:

Ενώ στην χωρική συμπλήρωση των χρονοσειρών δεν χρησιμοποιούνται όλοι οι υπάρχοντες-διαθέσιμοι σταθμοί αλλά μόνο οι γειτονικοί, που παρουσιάζουν και μεγαλύτερη συσχέτιση με την ελλιπή χρονοσειρά, στη χρονική συμπλήρωση, δηλαδή στη συμπλήρωση με χρήση των υπαρχουσών μετρήσεων της ίδιας της χρονοσειράς, προτιμάται η χρήση του ολικού μέσου όρου (σταθμισμένου ή μη) έναντι ενός τοπικού μέσου όρου που θα περιελάμβανε μόνο τις γειτονικές τιμές που παρουσιάζουν και τη μεγαλύτερη συσχέτιση.

Εξετάζουμε λοιπόν αν αυτή η πρακτική είναι βάσιμη, δηλαδή αν η εκτίμηση που προκύπτει από χρήση του ολικού μέσου όρου είναι καλύτερη από την χρήση ενός τοπικού μέσου όρου και κάτω από ποιες προϋποθέσεις.

## 4.2 Βασικές παραδοχές

### *4.2.1 Βασικές παραδοχές προσέγγισης*

Στην ανάλυση που ακολουθεί, έγιναν κάποιες απλουστευτικές μεν αλλά πολύ ρεαλιστικές παραδοχές με σκοπό την ευκολότερη προσέγγιση αυτού του πολύπλοκου προβλήματος, της μελέτης των φυσικών διεργασιών.

Πιο συγκεκριμένα, θεωρούμε πως οι ανελίξεις που προσομοιώνουν τις διάφορες υδρομετεωρολογικές διεργασίες είναι στάσιμες, δηλαδή τα στατιστικά τους χαρακτηριστικά μένουν αμετάβλητα ανεξαρτήτως της χρονικής περιόδου. Επίσης, για λόγους σύγκρισης και πληρότητας, εξετάζουμε δυο διαφορετικούς τύπους στοχαστικών ανελίξεων, τις ανελίξεις Markov, που έχουν βραχυπρόθεσμη μνήμη, και τις ανελίξεις με δυναμική ΗΚ που αναπαράγουν το φαινόμενο Hurst.

### *4.2.2 Ανελίξεις τύπου Markov και ανελίξεις με δυναμική ΗΚ*

Μελετάμε δυο τύπους στοχαστικών ανελίξεων οι οποίοι παρουσιάζουν τελείως διαφορετική συμπεριφορά όσον αφορά τη χρονική συσχέτιση μεταξύ των διαδοχικών τιμών της κάθε ανέλιξης. Εξετάζουμε λοιπόν στοχαστικές ανελίξεις τύπου Markov οι οποίες παρουσιάζουν ασθενή χρονική εξάρτηση (μνήμη) και ανελίξεις οι οποίες παρουσιάζουν μακρά μνήμη, δηλαδή αναπαράγουν το φαινόμενο Hurst και έχουν έντονη δομή αυτοσυσχέτισης. Στη συνέχεια παρουσιάζονται τα βασικά χαρακτηριστικά των δυο αυτών τύπων χρονικής εξάρτησης.

#### *4.2.2.1 Στοχαστικές ανελίξεις τύπου Markov*

Οι ανελίξεις Markov[18] είναι μια στοχαστική διαδικασία κατά την οποία η μετάβαση από τη μία κατάσταση στην επόμενη εξαρτάται μόνο από την τελευταία και όχι από τις υπόλοιπες προηγούμενες. Δηλαδή, όταν κάποια φυσική διεργασία περιγράφεται από μια ανέλιξη τύπου Markov, τότε μια μελλοντική τιμή της τυχαίας μεταβλητής της ανέλιξης, δεν εξαρτάται από τις προηγούμενες τιμές, όταν είναι γνωστή η τιμή της στο παρόν. Για το λόγο αυτό, λέμε ότι οι ανελίξεις Markov περιγράφουν διεργασίες με ασθενή μνήμη. Οι ανελίξεις Markov λοιπόν αδυνατούν να περιγράψουν το φαινόμενο της εμμονής που όπως είδαμε στο υποκεφάλαιο 2.4 αποτελεί μια βασική ιδιαιτερότητα το υδρομετεωρολογικών διεργασιών. Τα μοντέλα Markov χρησιμοποιούνται συνήθως για μεγάλες χρονικές κλίμακες (ετήσιες ή και μεγαλύτερες) όπου η συσχέτιση μεταξύ των διαφόρων τιμών είναι μικρότερη. Ένα παράδειγμα διεργασίας Markov είναι η κίνηση Brown[19].

Πιο συγκεκριμένα, μια ανέλιξη $X(t)$, στην οποία αν είναι γνωστό το παρόν, το μέλλον δεν εξαρτάται από το παρελθόν αλλά μόνο από το παρόν, λέγεται ανέλιξη Markov. Συμβολικά, για $t_1 \le t_2 \le ... \le t_n \le t$, και $\tau > 0$ έχουμε

$$P\left\{X(t+\tau) \le x \middle| X(t), X(t_n), ..., X(t_1)\right\} = P\left\{X(t+\tau) \le x \middle| X(t)\right\} \qquad (4.1)$$

[18] Οι ανελίξεις αυτού του τύπου πήραν το όνομά τους από τον Ρώσο μαθηματικό Andrey Markov (1856-1922), και αποτελούν ένα μαθηματικό εργαλείο που συμβάλει στην προσομοίωση διεργασιών που δεν έχουν ισχυρή μνήμη. Οι ιδιότητες των ανελίξεων Markov συνοψίζονται στην εξής φράση : «…the future is independent of the past given the present», σε ελεύθερη μετάφραση «…το μέλλον είναι ανεξάρτητο από το παρελθόν όταν το παρόν είναι δεδομένο».

[19] Κίνηση Brown καλείται η τυχαία κίνηση στερεών σωματιδίων μέσα σε ένα υγρό ή αέριο. Το φαινόμενο πήρε το όνομά του από το Βρετανό βοτανολόγο Robert Brown που πρώτος το παρατήρησε το 1827 όταν μελετούσε κόκκους γύρης στο νερό.

Το φαινόμενο της ασθενής μνήμης μπορούμε λοιπόν να το περιγράψουμε με μια ανέλιξη Markov και αφού οι μετρήσεις των διεργασιών που εξετάζουμε είναι σε διακριτό χρόνο, θα χρησιμοποιήσουμε ένα μοντέλο Markov σε διακριτό χρόνο, ή αλλιώς ένα μοντέλο αυτοπαλινδρόμησης (autoregression) τάξης 1. Το μοντέλο αυτό το συμβολίζουμε AR(1) και περιγράφεται από τη σχέση:

$$X_i = a\,X_{i-1} + V_i \tag{4.2}$$

όπου

$X_i$: είναι μια στάσιμη στοχαστική ανέλιξη σε διακριτό χρόνο

$V_i$: είναι μια ακολουθία λευκού θορύβου επίσης σε διακριτό χρόνο

$\alpha$ : παράμετρος που υπολογίζεται αναλυτικά από τα χαρακτηριστικά των $X$ και $V$

Αν $\mu_x$ και $\mu_v$ οι μέσες τιμές των $X_i$ και $V_i$ αντίστοιχα, $\gamma_m$ η αυτοσυνδιασπορά της $X_i$ για υστέρηση $m$, $\sigma_V^2$ η διασπορά της $V_i$ και $\mu_{3X}$ και $\mu_{3V}$ οι τρίτες κεντρικές ροπές των $X_i$ και $V_i$ αντίστοιχα, τότε προκύπτουν οι ακόλουθες εξισώσεις:

$$\begin{aligned} &\mathrm{Cov}\left[X_i, V_i\right] = \mathrm{Cov}\left[X_{i-1}, V_i\right] = 0 \\ &\mathrm{Cov}\left[X_i, V_i\right] = \sigma_V^2 \\ &\mathrm{Cov}\left[X_{i+m}, V_i\right] = a^m \sigma_V^2, (m > 0) \end{aligned} \tag{4.3}$$

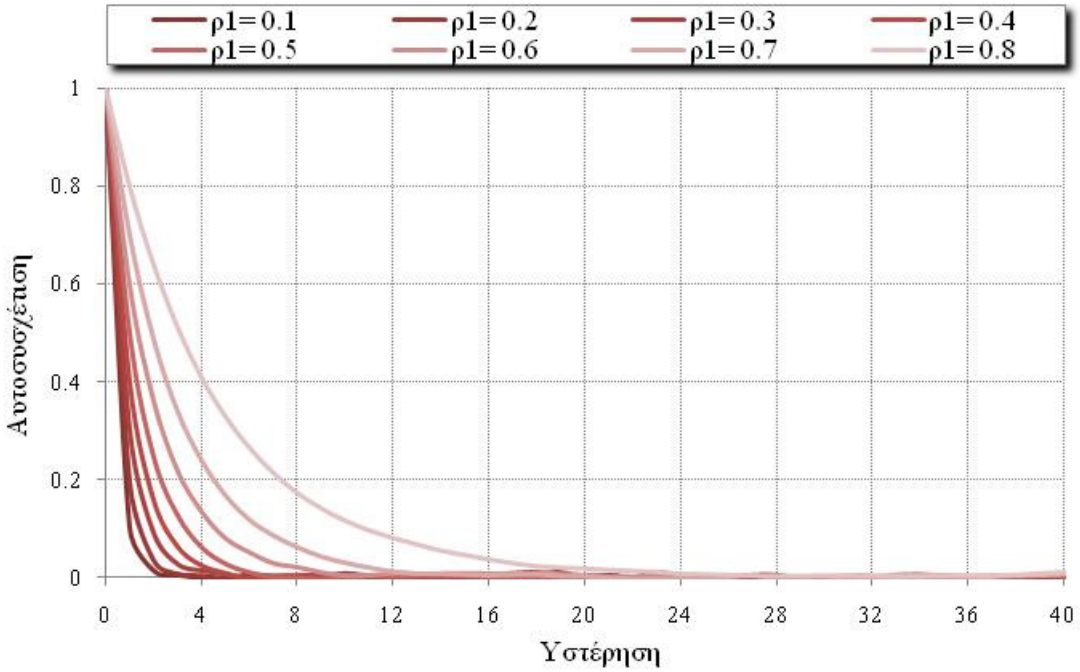
$$\gamma_m = a^m \gamma_0 \rightarrow \gamma_1 = a\gamma_0 \Leftrightarrow \alpha = \frac{\gamma_1}{\gamma_0} \tag{4.4}$$

Βασικό χαρακτηριστικό των ανελίξεων Markov είναι η εκθετική μείωση της αυτοσυνδιασποράς (ή της αυτοσυσχέτισης) με τη χρονική υστέρηση $m$ όπως φαίνεται και στη πιο πάνω σχέση.

Επίσης για τις ροπές των ανελίξεων AR(1) ισχύει :

$$\begin{aligned} \mu_V &= \mu_X (1-a) \\ \sigma_V^2 &= \gamma_0 (1-a^2) \\ \mu_{3V} &= \mu_{3X} (1-a^3) \end{aligned} \tag{4.5}$$

Το επόμενο διάγραμμα αποτελεί ένα αυτοσυσχετόγραμμα για μια στοχαστική ανέλιξη Markov που προσομοιώνεται με ένα μοντέλο AR(1) με συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) που μεταβάλλεται από 0.1 έως και 0.8. Παρατηρούμε πως η τιμή της αυτοσυσχέτισης μειώνεται ταχύτατα (εκθετικά) σε σχέση με την υστέρηση. Πιο συγκεκριμένα, η αυτοσυσχέτιση γίνεται πρακτικά μηδέν για υστέρηση μικρότερη του 4, όταν το μοντέλο AR(1) που χρησιμοποιείτε έχει μικρό συντελεστή $\rho_1$ ενώ για μεγαλύτερους συντελεστές $\rho_1$ μηδενίζεται για υστέρηση ίση με 10 περίπου.




Σχήμα 4.1 Αυτοσυσχετόγραμμα στοχαστικών ανελίξεων τύπου Markov με χρήση μοντέλου AR(1) για διάφορες τιμές του συντελεστή $\rho_1$.

#### *4.2.2.2 Στοχαστικές ανελίξεις με δυναμική Hurst-Kolmogorov*

Όπως είδαμε στην παράγραφο 2.4, μια από τις ιδιαιτερότητες των υδρομετεωρολογικών χρονοσειρών είναι η μακρά μνήμη, η έντονη δηλαδή αυτοσυσχέτιση που παρατηρείται μεταξύ των τιμών ακόμη και για πολύ μεγάλες τιμές της υστέρησης. Το φαινόμενο αυτό όπως έχουμε ήδη αναφέρει, ονομάζεται φαινόμενο Hurst, προς τιμή του βρετανού μηχανικού Harold Edwin Hurst, που πρώτος το ανακάλυψε.

Η μακρά μνήμη (long range dependence ή long-term persistence) ή η δυναμική Hurst-Kolmogorov (HK dynamic) όπως συχνά ονομάζεται, είναι μια ιδιαιτερότητα των υδρολογικών, και όχι μόνο, διεργασιών η οποία πρέπει να ληφθεί υπόψη στη μελέτη και προσομοίωση αυτών των φαινομένων. Η προσομοίωση της μακράς μνήμης που εμφανίζεται στις υδρολογικές διεργασίες όμως, δεν μπορεί να γίνει με τα κλασσικά μοντέλα που προτείνονται από τους (Box and Jenkins, 1976).

Ο Mandelbort προσπαθώντας να αναπαράγει μαθηματικά το φαινόμενο Hurst εισάγαγε μια νέα στοχαστική ανέλιξη (Mandelbrot, 1977) γνωστή ως κλασματικός Γκαουσιανός θόρυβος (Fractional Gaussian Noise FGN). Ο κλασματικός Γκαουσιανός θόρυβος (FGN) ή αλλιώς η ανέλιξη απλής ομοιοθεσίας (Simple Scaling Process) όπως συχνά αναφέρεται, είναι μια στάσιμη στοχαστική ανέλιξη που ορίζεται από τη σχέση

$$Z_i^{(k)} - k\mu =_d \left(\frac{k}{l}\right)^H \left(Z_j^{(l)} - l\mu\right) \tag{4.6}$$

όπου

$=_d$ συμβολίζει την ισότητα στην πεπερασμένης διάστασης από κοινού κατανομή

*i, j, k* και *l* είναι οποιοιδήποτε ακέραιοι

*H* είναι μια σταθερά ($0 < H < 1$) γνωστή ως συντελεστής (ή εκθέτης) Hurst

$Z_i^{(k)}$ είναι η συναθροισμένη ανέλιξη σε κλίμακα *k* για την όποια ισχύει

$$Z_i^{(k)} = \frac{1}{k} \sum_{i=(j-1)k+1}^{jk} Z_i^{(1)} \tag{4.7}$$

$\mu$ είναι η μέση τιμή της συναθροισμένης ανέλιξης $Z_i^{(k)}$

Εφόσον η ανέλιξη είναι στάσιμη για κάθε κλίμακα *k* , (δηλαδή $i = j$), για $l = 1$ και $\mu = 0$, έχουμε

$$\begin{aligned} & Z_i^{(k)} = k^H X_i \\ & \Rightarrow \mathrm{Var}\left[ Z_i^{(k)} \right] = k^{2H} \mathrm{Var}\left[ X_i \right] \\ & \Rightarrow \gamma_0^{(k)} := \mathrm{Var}\left[ Z_i^{(k)} \right] = k^{2H} \gamma_0 \end{aligned} \tag{4.8}$$

Κατά συνέπεια η τυπική απόκλιση κλίμακας *k* είναι ανάλογη της ποσότητας $k^{2H}$ γεγονός που επιβεβαιώνεται από τις παρατηρημένες στη φύση διεργασίες.

Η σταθερά *H*, που είναι ο εκθέτης Hurst, μαθηματικά, μπορεί να πάρει τιμές από 0 έως και 1. Για $H = 0.5$, η ανέλιξη γίνεται λευκός θόρυβος, δηλαδή δεν υπάρχει συσχέτιση μεταξύ των διαφόρων τιμών της ανέλιξης, ενώ οι τιμές $0 < H < 0.5$, αν και μαθηματικώς μπορεί να υπάρξουν, δεν έχουν κανένα απολύτως νόημα στην υδρολογία (Koutsoyiannis, 2005).

Ο συντελεστής Hurst λοιπόν, υπολογίζεται από την κλίση της ευθείας που προκύπτει από τη γραφική παράσταση της τυπικής απόκλισης για κλίμακα *k* συναρτήσει της κλίμακας *k* σε διπλό λογαριθμικό διάγραμμα (Κουτσογιάννης, 2008). Ωστόσο, υπάρχει ποικιλία προτάσεων στη διεθνή βιβλιογραφία, όσον αφορά τον τρόπο και την ακρίβεια υπολογισμού του συντελεστή Hurst σε μια υδρομετεωρολογική χρονοσειρά (Mesa, Oscar J.; Poveda, German, 1993).

Σε κάθε κλίμακα συνάθροισης *k*, η συνάρτηση αυτοσυσχέτισης είναι ανεξάρτητη της κλίμακας *k* (Κουτσογιάννης, 2008) και δίνεται από τη σχέση

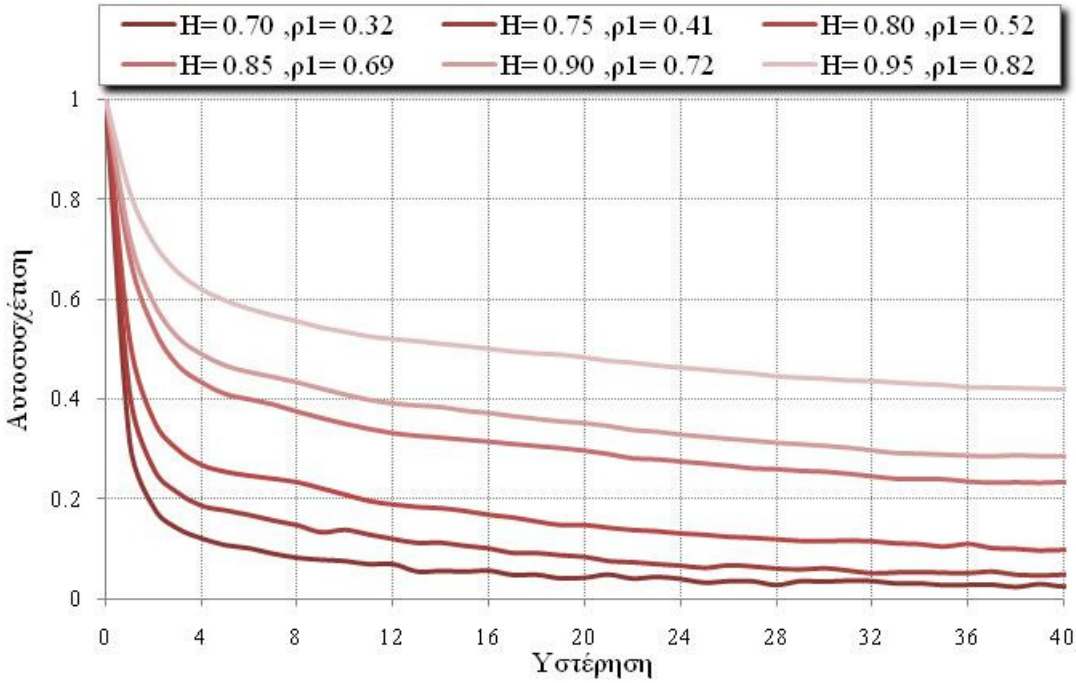
$$\rho_j^{(k)} = \rho_j = \left( \frac{1}{2} \right) \left[ \left( j+1 \right)^{2H} + \left( j-1 \right)^{2H} \right] - j^{2H}, \quad \text{με} \quad j > 0 \tag{4.9}$$

και προσεγγιστικά η πιο πάνω σχέση απλοποιείται σε

$$\rho_j^{(k)} = \rho_j = H \left( 2H - 1 \right) j^{2H-2} \tag{4.10}$$

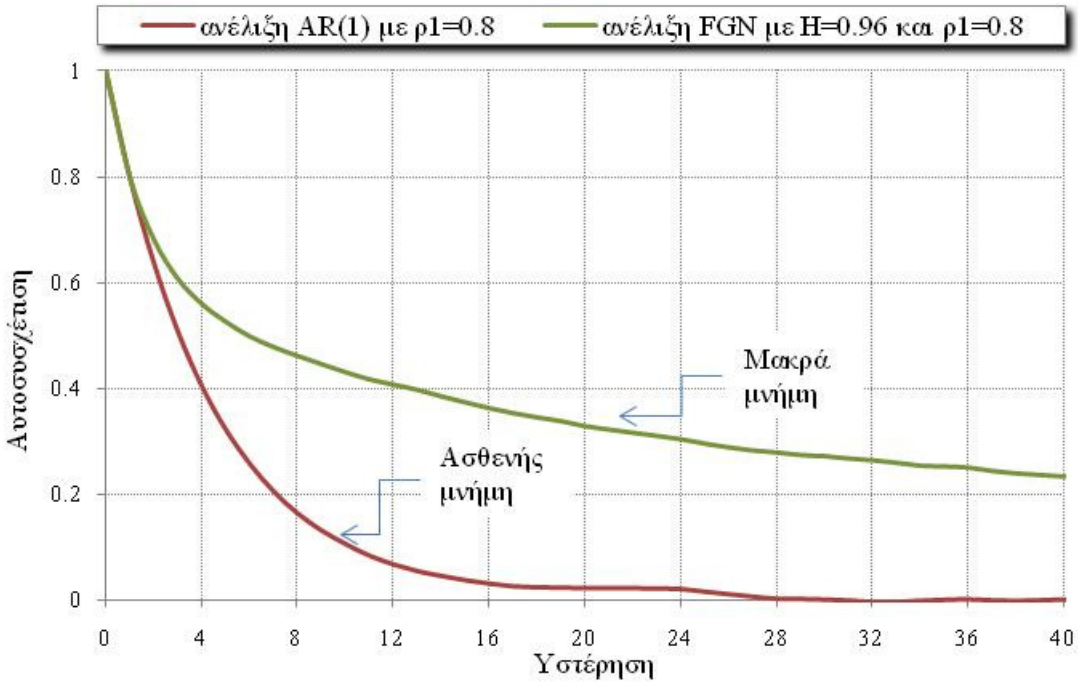
που δείχνει πως η αυτοσυσχέτιση είναι συνάρτηση δύναμης της υστέρησης.

Στο πιο κάτω σχήμα φαίνεται η σχέση της αυτοσυσχέτισης μιας στοχαστικής ανέλιξης που παρουσιάζει μακροπρόθεσμη εμμονή για διάφορες τιμές της υστέρησης (lag). Παρατηρούμε πως για μεγάλες τιμές του συντελεστή Hurst και συνεπώς για μεγάλες τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$), παρουσιάζεται σημαντική συσχέτιση μεταξύ των τιμών της χρονοσειράς, ακόμη και για μεγάλες τιμές της υστέρησης.




Σχήμα 4.2 Αυτοσυσχετόγραμμα στοχαστικών ανελίξεων που παρουσιάζουν δυναμική ΗΚ, με χρήση μοντέλου διαδοχικών επιμερισμών 3AR(1) για διάφορες τιμές του συντελεστή Hurst.

Σε αντίθεση λοιπόν με τις ανελίξεις τύπου Markov που η αυτοσυσχέτιση φθίνει με εκθετικό ρυθμό καθώς αυξάνει η υστέρηση, στις ανελίξεις με δυναμική ΗΚ, η αυτοσυσχέτιση φθίνει πιο ομαλά, και μάλιστα με μορφή δύναμης. Αυτή η διαπίστωση επιβεβαιώνεται στο Σχήμα 4.3. Συγκρίνοντας λοιπόν το αυτοσυσχετόγραμμα που προκύπτει για τις ανελίξεις Markov σε σχέση με το αντίστοιχο που προκύπτει για ανελίξεις με δυναμική ΗΚ για ίδιο συντελεστή συσχέτισης για υστέρηση 1 έχουμε:




Σχήμα 4.3 Αυτοσυσχετόγραμμα στοχαστικής ανέλιξης τύπου Markov με χρήση μοντέλου AR(1) για συντελεστή $\rho_1 = 0.8$ και στοχαστικής ανέλιξης με δυναμική ΗΚ παραγμένη με χρήση μοντέλου FGN για συντελεστή $\rho_1 = 0.8$ και συντελεστή $H = 0.96$ .

Σαν μέτρο λοιπόν της εμμονής (μακρά μνήμη) χρησιμοποιείται ο συντελεστής Hurst, ενώ σαν μέτρο της βραχυπρόθεσμης μνήμης (μοντέλα Markov) χρησιμοποιείται ο συντελεστής αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$).

Στη βιβλιογραφία υπάρχει πλήθος αλγορίθμων που προσπαθούν να αναπαράγουν απεικονίσεις ανελίξεων με δυναμική ΗΚ. Όμως, πολλοί από αυτούς δεν είναι εύκολοι ούτε στην κατανόηση αλλά ούτε και στην εφαρμογή. Βάση τους συχνά είναι οι ιδιότητες της ανέλιξης του κλασματικού Γκαουσιανού θορύβου (FGN). Οι πιο δημοφιλείς λοιπόν προσεγγίσεις συνοψίζονται σε:

- προσέγγιση τυχαίων διακυμάνσεων πολλαπλής κλίμακας
- προσέγγιση βασισμένη σε διαδοχικούς επιμερισμούς
- προσέγγιση συμμετρικού κυλιόμενου μέσου όρου (SMA)

Στη συνέχεια της εργασίας, ο αλγόριθμος που θα χρησιμοποιηθεί για την δημιουργία ανελίξεων που αναπαράγουν το φαινόμενο Hurst, βασίζεται στην προσέγγιση τυχαίων διακυμάνσεων πολλαπλής κλίμακας (Koutsoyiannis, 2002), και συνοπτικά παρουσιάζεται πιο κάτω.

Η ανέλιξη $X_t$ που θέλουμε να δημιουργήσουμε, και που θα αναπαράγει το φαινόμενο Hurst, παράγεται ως ένα άθροισμα τριών ανελίξεων AR(1), δηλαδή

$$X_i = A_i + B_i + C_i \tag{4.11}$$

με συντελεστές αυτοσυσχέτισης για υστέρηση 1, αντίστοιχα

$$\rho = 1.52(H - 0.5)^{1.32},$$

$$\varphi = 0.953 - 7.69(1 - H)^{3.85}, \tag{4.12}$$

$$\xi = \begin{cases} 0.932 + 0.087H, & H \le 0.76 \\ 0.993 + 0.007H, & H \ge 0.76 \end{cases}$$

και διασπορές

$$\begin{aligned} \mathrm{Var}\left[A_i\right] &= (1 - c_1 - c_2)\,\gamma_0 \\ \mathrm{Var}\left[B_i\right] &= c_1\,\gamma_0 \\ \mathrm{Var}\left[C_i\right] &= c_2\,\gamma_0 \end{aligned} \tag{4.13}$$

όπου $c_1$ και $c_2$ εκτιμώνται με τρόπο ώστε η αυτοσυσχέτιση του αθροίσματος των τριών ανελίξεων

$$\rho_j = (1 - c_1 - c_2)\rho^j + c_1\varphi^j + c_2\xi^j \tag{4.14}$$

να ταυτίζεται με τη θεωρητική αυτοσυσχέτιση του κλασματικού Γκαουσιανού θορύβου για υστέρηση 1 και 100.

Δηλαδή, για τον υπολογισμό των $c_1$ και $c_2$ λύνουμε το σύστημα:

$$\begin{aligned} \rho_1 &= (1 - c_1 - c_2)\rho^1 + c_1\varphi^1 + c_2\xi^1 \\ \rho_{100} &= (1 - c_1 - c_2)\rho^{100} + c_1\varphi^{100} + c_2\xi^{100} \end{aligned} \tag{4.15}$$

Πρέπει να διευκρινιστεί πως οι μέθοδοι που παρουσιάζονται πιο κάτω, στηρίζονται στη συμπλήρωση ελλιπών τιμών από μια χρονοσειρά βασιζόμενες στις διάφορες μετρήσεις του ίδιου σταθμού. Δηλαδή αξιοποιείται όλη η διαθέσιμη πληροφορία που προέρχεται από τον ίδιο σταθμό, και δεν εξετάζεται η χωρική του συσχέτιση με γειτονικούς σταθμούς.

## 4.3 Μεμονωμένα κενά στις χρονοσειρές

### *4.3.1 Εισαγωγή*

Εξετάζουμε το πρόβλημα της ύπαρξης μεμονωμένων κενών στις χρονοσειρές. Δηλαδή, οι ελλείπουσες τιμές παρουσιάζονται σποραδικά ώστε να μη δημιουργούνται

κενά για μεγάλα χρονικά διαστήματα στις μετρήσεις. Με τον τρόπο αυτό, εξασφαλίζουμε πως για κάθε τιμή που λείπει και θέλουμε να συμπληρώσουμε, θα υπάρχει πάντα ικανοποιητικός αριθμός παρατηρήσεων πριν και μετά από το κενό. Έχοντας λοιπόν εξασφαλίσει πως θα υπάρχουν οι γειτονικές τιμές, εξετάζουμε τη χρήση ενός τοπικού μέσου όρου στη θέση του ολικού, που συνήθως αβίαστα προτιμάται.

Πιο συγκεκριμένα, υπολογίζουμε το μέσο τετραγωνικό σφάλμα της εκτίμησης για χρήση ενός τοπικού μέσου όρου με μεταβλητό πλήθος γειτονικών τιμών ($n$). Συγκρίνοντας λοιπόν τα μέσα τετραγωνικά σφάλματα που υπολογίστηκαν, καταλήγουμε στο ποιος είναι κάθε φορά ο βέλτιστος αριθμός των γειτονικών τιμών ($n$), δηλαδή, ποιο είναι το πλήθος των τιμών πριν και μετά το κενό, ώστε να έχουμε το ελάχιστο τετραγωνικό σφάλμα της εκτίμησης.

Όπως έχουμε ήδη τονίσει, το πρόβλημα της συμπλήρωσης ελλιπών τιμών αναφέρεται στην εκτίμηση της τιμής $y$ μιας τυχαίας μεταβλητής, από ένα πλήθος γνωστών παρατηρήσεων $x_i$ με $i = 1,...,n$, της ίδιας τυχαίας μεταβλητής είτε σε άλλες χρονικές περιόδους στο ίδιο σημείο, είτε σε γειτονικά σημεία για την ίδια χρονική περίοδο. Το πρόβλημα λοιπόν της συμπλήρωσης των ελλιπών τιμών μπορεί να εκφραστεί μαθηματικά από την γραμμική σχέση

$$y = w_1 x_1 + w_2 x_2 + ... + w_n x_n + e \tag{4.16}$$

όπου

$y$ συμβολίζεται η τιμή που λείπει

$w_i$ είναι οι συντελεστές βαρύτητας

$e$ συμβολίζεται το σφάλμα της εκτίμησης

Σε μορφή πινάκων η παραπάνω σχέση γράφεται

$$Y = \mathbf{w}^T \mathbf{X} + e \quad \Leftrightarrow \quad e = Y - \mathbf{w}^T \mathbf{X} \tag{4.17}$$

όπου

$$\mathbf{w} := \left[ w_1, ..., w_n \right]^T$$

$$\mathbf{X} := \left[ x_1, ..., x_n \right]$$

### *4.3.2 Τοπικός μέσος όρος έναντι ολικού*[20]

Για λόγους απλότητας, αλλά και για ευκολία στους υπολογισμούς, και συνεπώς ευκολότερη εφαρμογή της μεθόδου, θεωρούμε πως οι συντελεστές βαρύτητας είναι σταθεροί και ίσοι μεταξύ τους. Δηλαδή

$$w_1 = w_2 = \ldots = w_n = \frac{1}{n} \tag{4.18}$$

όπου $n$ το πλήθος τον γειτονικών τιμών που θα χρησιμοποιηθούν για την εκτίμηση της ελλείπουσας τιμής.

Το μέσο τετραγωνικό σφάλμα (MSE) δίνεται από το τύπο

$$\mathrm{MSE} := E\left[e^2\right] = \sigma_e^2 + \mu_e^2 \tag{4.19}$$

Και με βάση τις παραδοχές που έχουμε προαναφέρει (στάσιμη στοχαστική ανέλιξη και ίσοι συντελεστές βαρύτητας) το MSE μπορεί να υπολογιστεί ως εξής

$$\mathrm{MSE} := E\left[e^2\right] = E\left[\left(x_t - \frac{\sum_{i=1}^{-n} x_{t+i} + \sum_{i=1}^{n} x_{t+i}}{2n}\right)^2\right] \tag{4.20}$$

Και μετά από αλγεβρικές πράξεις και απλοποιήσεις[21] καταλήγουμε όπως επισημαίνεται και στο (Dialynas, Y., P. Kossieris, K. Kyriakidis, A. Lykou, Y. Markonis, C. Pappas, S.M. Papalexiou, and D. Koutsoyiannis, 2010) στη πιο κάτω σχέση

$$\mathrm{MSE} := E\left[e^2\right] = \frac{1}{2}\left(\frac{\sigma}{n}\right)^2\left[(2n+1)\left(n - 2\sum_{i=1}^{n}\rho_i\right) + \sum_{i=1}^{2n}(2n+1-i)\rho_i\right] \tag{4.21}$$

Παρατηρούμε πως η παραπάνω σχέση, που μας δίνει το μέσο τετραγωνικό σφάλμα της εκτίμησης συναρτήσει του πλήθους $n$ των γειτονικών τιμών που χρησιμοποιούνται, είναι ανεξάρτητη της μέσης τιμής του δείγματος. Εξαρτάται μόνο από την τυπική απόκλιση $\sigma$, το πλήθος $n$ και τους συντελεστές αυτοσυσχέτισης για διάφορες τιμές της υστέρησης. Για το λόγο αυτό δεν κρίνεται σκόπιμη και αναγκαία η κανονικοποίηση και η τυποποίηση της χρονοσειράς πριν την εφαρμογή της μεθόδου. Την τυπική απόκλιση τη θεωρούμε απλουστευτικά ίση με 1, και εξετάζουμε όπως έχουμε ήδη αναφέρει δύο τύπους στοχαστικών ανελίξεων, Markov και ανελίξεις με δυναμική ΗΚ, οι οποίες παρουσιάζουν και διαφορετική δομή αυτοσυσχέτισης, δηλαδή διαφορετικές συναρτήσεις $\rho_i$.

---

[20] Το κεφάλαιο αυτό παρουσιάστηκε αρχικά στα Αγγλικά στο *European Geosciences Union General Assembly 2010* με τίτλο «Optimal infilling of missing values in hydrometeorological time series». Για περισσότερες λεπτομέρειες βλέπε (Dialynas, Y., P. Kossieris, K. Kyriakidis, A. Lykou, Y. Markonis, C. Pappas, S.M. Papalexiou, and D. Koutsoyiannis, 2010).

[21] Για την απόδειξη της σχέσης (4.21) βλέπε ΠΑΡΑΡΤΗΜΑ Α

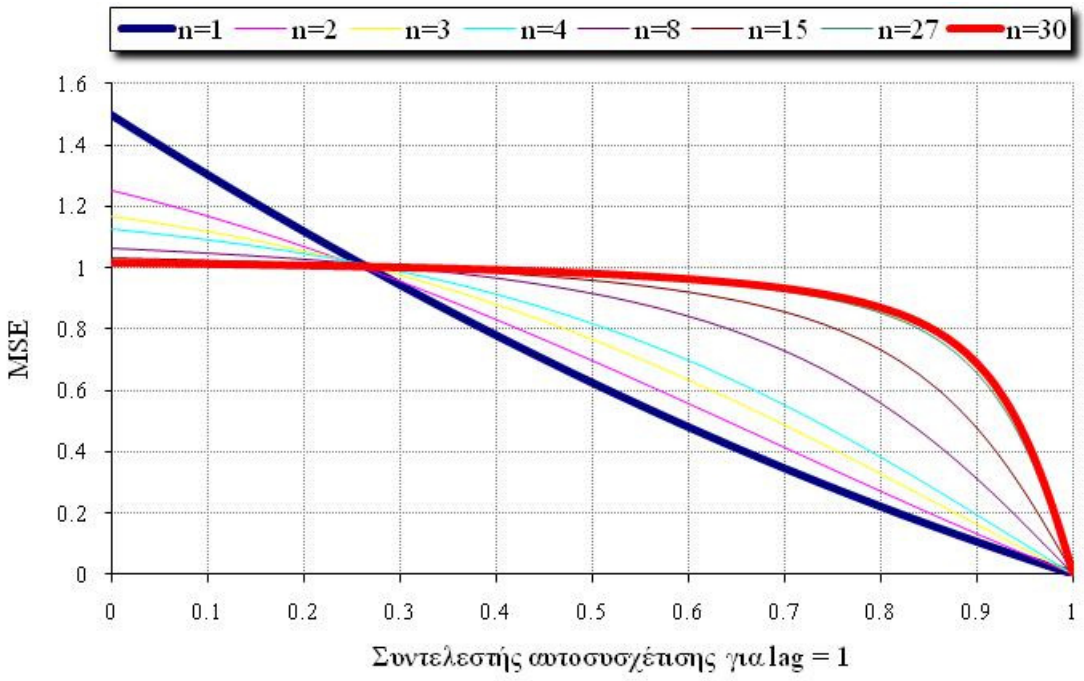
Βέλτιστη συμπλήρωση θεωρείτε εκείνη που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα. Έτσι ο αριθμός των γειτονικών τιμών που χρησιμοποιούνται για τη συμπλήρωση θα θεωρείτε βέλτιστος αν ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα. Εφαρμόζουμε λοιπόν την παραπάνω σχέση για διάφορες τιμές της παραμέτρου *n,* και υπολογίζουμε το αντίστοιχο MSE. Η τιμή του *n* που θα μας δώσει και το ελάχιστο MSE θα είναι η βέλτιστη και ο τοπικός μέσος όρος που θα χρησιμοποιήσουμε για την εκτίμηση της ελλείπουσας τιμής θα περιλαμβάνει *n* τιμές πριν και *n* μετά την ελλείπουσα τιμή.

#### *4.3.2.1 Χρονοσειρές που προσομοιώνονται με ανελίξεις Markov*

Οι στοχαστικές ανελίξεις Markov, όπως αναλυτικά περιγράψαμε και στο υποκεφάλαιο 4.2.2.1 έχουν ασθενή μνήμη και περιγράφονται ικανοποιητικά από το μοντέλο AR(1). Η συνάρτηση αυτοσυσχέτισης δίνεται από τη σχέση

$$\rho_j = \left(\rho_1\right)^j \tag{4.22}$$

Έτσι, από τη σχέση (4.21) με βάση τη συνάρτηση αυτοσυσχέτισης που δίνεται από τη σχέση (4.22) προκύπτει το πιο κάτω γράφημα, για διάφορες τιμές της παραμέτρου *n*.




Σχήμα 4.4 MSE συναρτήσει του συντελεστή αυτοσυσχέτισης με υστέρηση 1, για διάφορες τιμές *n* γειτονικών παρατηρήσεων για τη συμπλήρωση της ελλείπουσας τιμής.

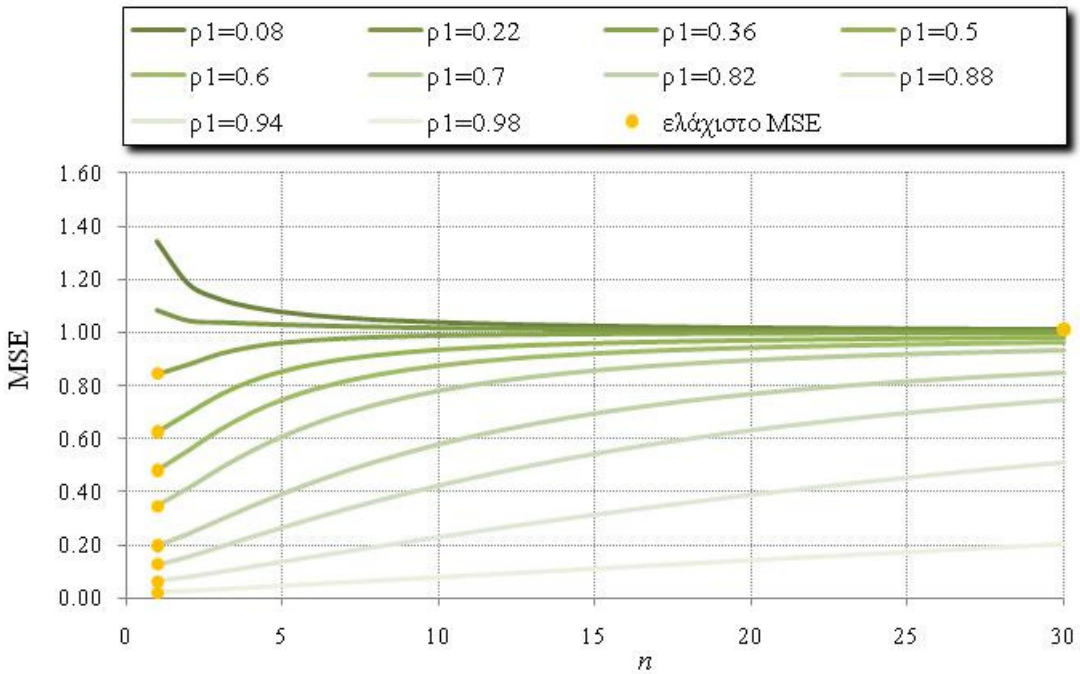
Παρατηρούμε πως για μικρές τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) το MSE γίνεται ελάχιστο για $n = 30$. Το $n = 30$ πρακτικά υποδηλώνει τον ολικό

μέσο όρο. Άρα για μικρές τιμές του $\rho_1$ ο βέλτιστος αριθμός των γειτονικών τιμών, που ελαχιστοποιεί δηλαδή το MSE αντιστοιχεί στον ολικό μέσο όρο.

Όμως, για μεγαλύτερες τιμές του $\rho_1$ παρατηρούμε πως αυτή η συμπεριφορά αλλάζει και μάλιστα δραματικά. Συγκεκριμένα, όπως φαίνεται και στο πιο πάνω σχήμα, υπάρχει μια κρίσιμη τιμή του $\rho_1$ πέρα από την οποία η χρήση του ολικού μέσου όρου κρίνεται ακατάλληλη, καθώς το MSE που προκύπτει δεν είναι το ελάχιστο. Η κρίσιμη αυτή τιμή του $\rho_1$ είναι $\rho_{cr} \approx 0.24$ και για τιμές του $\rho_1$ μεγαλύτερες του 0.24 η βέλτιστη συμπλήρωση προκύπτει με χρήση τοπικού μέσου όρου και μάλιστα με χρήση μιας τιμής πριν και μιας μετά την ελλείπουσα τιμή.

Συνοψίζοντας λοιπόν τις παρατηρήσεις για το πιο πάνω διάγραμμα, έχουμε πως στις ανελίξεις Markov ο βέλτιστος αριθμός γειτονικών βημάτων που πρέπει να χρησιμοποιηθούν για την συμπλήρωση μιας ελλείπουσας τιμής καθορίζεται από το συντελεστή $\rho_1$,και μάλιστα για

- $\rho_1 < \rho_{cr} \approx 0.24$ προτείνεται η χρήση του ολικού μέσου όρου
- $\rho_1 \geq \rho_{cr} \approx 0.24$ η χρήση ενός τοπικού μέσου όρου χρησιμοποιώντας μια τιμή πριν και μία μετά ενδείκνυται.

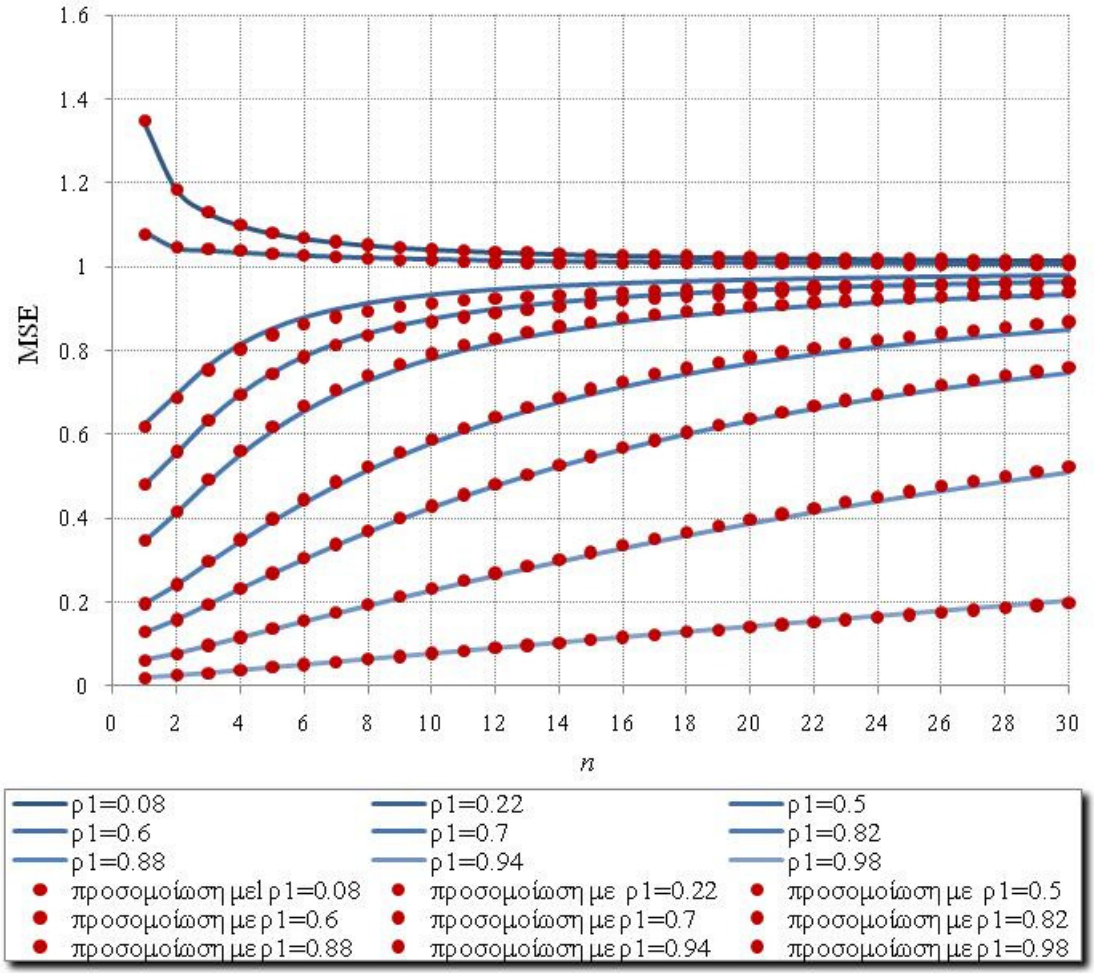



Σχήμα 4.5 MSE συναρτήσει του πλήθους των γειτονικών τιμών $n$ για διάφορες τιμές του συντελεστή $\rho_1$

Στο παραπάνω σχήμα γίνεται ακόμη πιο αισθητή η απότομη μεταβολή του βέλτιστου πλήθους των γειτονικών τιμών που απαιτούνται ώστε να ελαχιστοποιηθεί το MSE. Συγκεκριμένα, παρατηρούμε πως για χαμηλές τιμές της αυτοσυσχέτισης το

ελάχιστο MSE αντιστοιχεί στο $n$ = 30 και καθώς ο συντελεστής $\rho_1$ αυξάνει, το ελάχιστο MSE προκύπτει για $n$ = 1, χωρίς να πάρει καμία ενδιάμεση τιμή στο διάστημα [1, 30].

Για λόγους πληρότητας, προκειμένου να επαληθευτούν τα πιο πάνω θεωρητικά αποτελέσματα, παραγάγαμε με ένα μοντέλο AR(1) 600000 τιμές και κάθε φορά αφαιρούσαμε μια τιμή και την συμπληρώναμε με την πιο πάνω μεθοδολογία και στη συνέχεια υπολογίζαμε το MSE που προέκυπτε με χρήση κάθε φορά διαφορετικού πλήθους γειτονικές τιμές $n$. Τα αποτελέσματα ήταν ιδιαίτερα ενθαρρυντικά, και παρουσιάζονται στο πιο κάτω διάγραμμα.




Σχήμα 4.6 Θεωρητικές και προσομοιωμένες καμπύλες MSE - $n$ για διάφορες τιμές του συντελεστή $\rho_1$ .

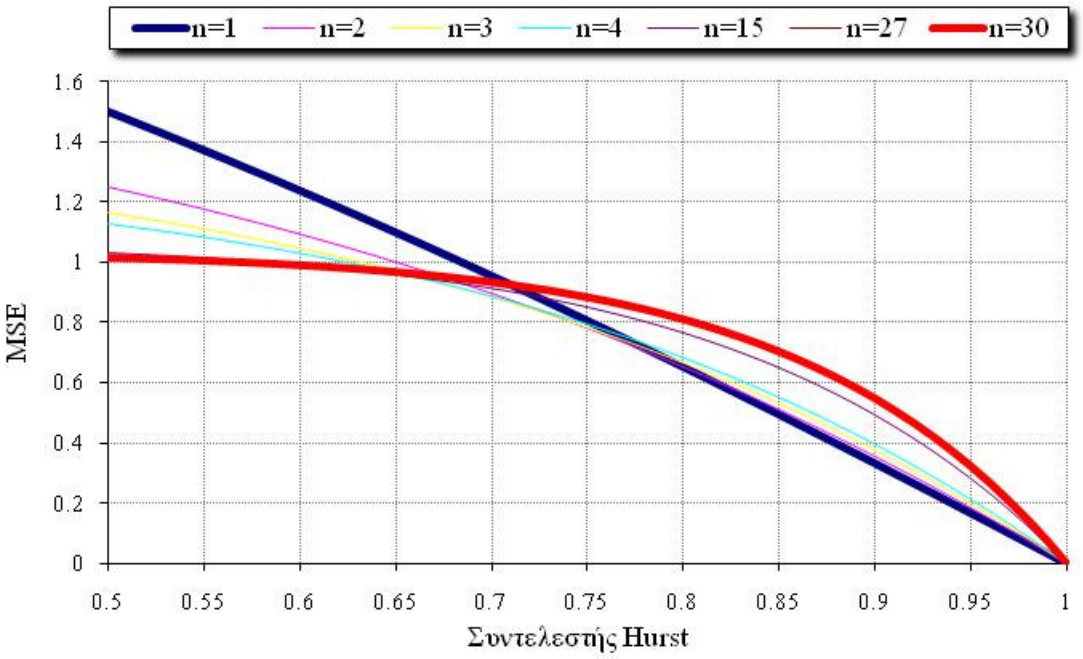
Παρατηρούμε πως οι προσομοιωμένες τιμές του MSE σχεδόν συμπίπτουν με τις αντίστοιχες θεωρητικές καμπύλες που προέκυψαν από την εφαρμογή της σχέσης (4.21).

*4.3.2.2 Χρονοσειρές που παρουσιάζουν δυναμική Hurst – Kolmogorov*

Σε αντίθεση με τις ανελίξεις Markov, οι ανελίξεις με δυναμική ΗΚ, παρουσιάζουν μακρά μνήμη. Όπως είδαμε στο υποκεφάλαιο 4.2.2.2 οι ανελίξεις με δυναμική ΗΚ έχουν έντονη δομή αυτοσυσχέτισης και μάλιστα η αυτοσυσχέτιση τους δεν μεταβάλλεται εκθετικά με την υστέρηση, όπως συμβαίνει στις ανελίξεις Markov, αλλά μεταβάλλεται με μορφή δύναμης. Πιο συγκεκριμένα, η αυτοσυσχέτιση, για διάφορες τιμές της υστέρησης *j* δίνεται από τη σχέση (4.9), δηλαδή

$$\rho_j = \left(\frac{1}{2}\right)\left[(j+1)^{2H} + (j-1)^{2H}\right] - j^{2H}$$

Με βάση λοιπόν τη σχέση (4.21), που υπολογίζει το MSE συναρτήσει του αριθμού των γειτονικών τιμών *n* που λαμβάνονται υπόψη στον υπολογισμό της ελλείπουσας τιμής και τη συνάρτηση αυτοσυσχέτισης για διάφορες τιμές της υστέρηση που δίνεται από τη σχέση (4.9), προκύπτει το πιο κάτω γράφημα, για διάφορες τιμές της παραμέτρου *n*.




4.7 MSE συναρτήσει του συντελεστή Hurst για διάφορες τιμές *n* γειτονικών παρατηρήσεων για τη συμπλήρωση της ελλείπουσας τιμής.

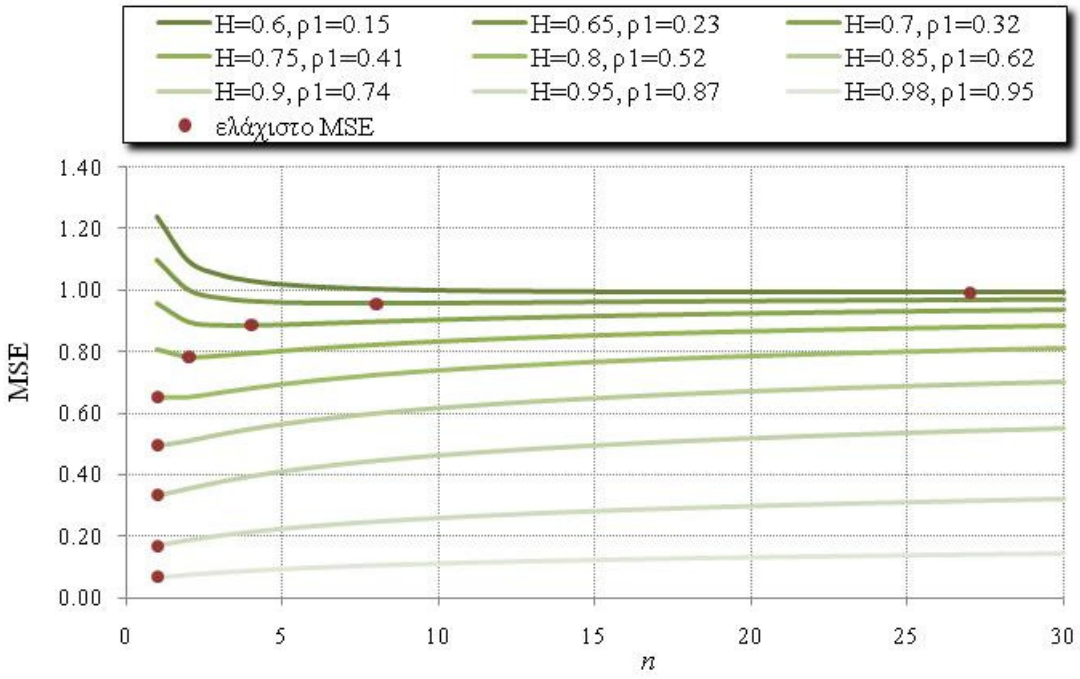
Παρατηρούμε πως όσο αυξάνει ο συντελεστής Hurst, δηλαδή όσο η δομή της αυτοσυσχέτισης γίνεται πιο έντονη, τόσο μικραίνει ο αριθμός των γειτονικών τιμών που πρέπει να χρησιμοποιήσουμε ώστε να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα. Πιο συγκεκριμένα, για μικρές τιμές του συντελεστή Hurst προτιμάται η χρήση του ολικού μέσου όρου, αφού το μέσο τετραγωνικό σφάλμα, όπως φαίνεται και στο πιο πάνω γράφημα, γίνεται ελάχιστο για n = 30, δηλαδή 30 τιμές πριν και 30

μετά την ελλείπουσα τιμή που ουσιαστικά είναι ο ολικός μέσος όρος. Αντίθετα, για μεγάλες τιμές του συντελεστή Hurst ο βέλτιστος αριθμός των γειτονικών τιμών που απαιτούνται για την συμπλήρωση μειώνεται δραματικά.

Η τελευταία αυτή παρατήρηση διευκολύνει πάρα πολύ τη συμπλήρωση ελλειπουσών τιμών σε χρονοσειρές με έντονη αυτοσυσχέτιση, γιατί ουσιαστικά το μόνο που απαιτείται είναι η γνώση της αμέσως προηγούμενης και της αμέσως επόμενης τιμής ώστε η πρόβλεψη που θα κάνουμε να έχει το ελάχιστο MSE.

Ειδικότερα, συνοψίζοντας τα αποτελέσματα του παραπάνω γραφήματος, διαπιστώνουμε ότι για τη συμπλήρωση ελλείπουσας τιμής σε μια χρονοσειρά με συντελεστή Hurst

- $H$ = 0.50-0.60, προτιμάται η χρήση του ολικού μέσου όρου
- $H$ = 0.70, απαιτούνται 4 τιμές πριν και 4 μετά την ελλείπουσα τιμή
- $H$ = 0.72, απαιτούνται 3 τιμές πριν και 3 μετά την ελλείπουσα τιμή
- $H$ = 0.74, απαιτούνται 2 τιμές πριν και 2 μετά την ελλείπουσα τιμή
- $H \geq$ 0.80, χρειαζόμαστε μόνο μία τιμή πριν και μία μετά

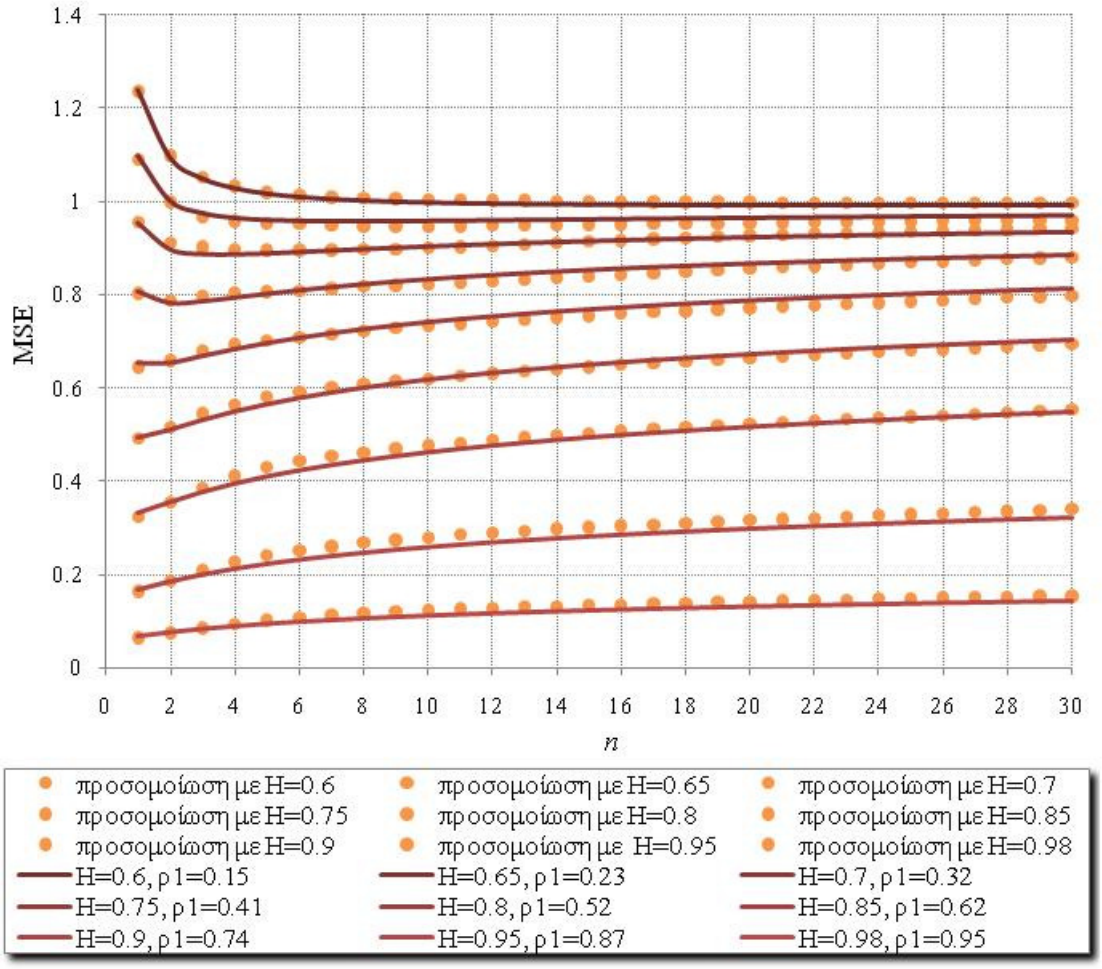



4.8 MSE συναρτήσει του πλήθους των γειτονικών τιμών $n$ για διάφορες τιμές του συντελεστή Hurst ($H$).

Και στο πιο πάνω γράφημα γίνεται φανερό πώς στις ανελίξεις με μακρά μνήμη, για μεγάλες τιμές του συντελεστή Hurst, προτιμάται ένας τοπικός μέσος όρος και μάλιστα περιορισμένος σε πολύ γειτονικά χρονικά βήματα (από 3 έως και 1) έναντι του ολικού.

Σε σύγκριση με το αντίστοιχο διάγραμμα για τις ανελίξεις Markov, παρατηρούμε πως εδώ η μετάβαση από τον ολικό μέσο όρο στον τοπικό και μάλιστα στον τοπικό μέσο όρο μόνο μιας προηγούμενης και μιας επόμενης τιμής, δεν γίνεται απότομα, αλλά ομαλά.

Όπως στην περίπτωση των ανελίξεων Markov, έτσι και εδώ, προκειμένου να επαληθευτούν τα θεωρητικά αποτελέσματα και για λόγους πληρότητας, παραγάγαμε με ένα μοντέλο FGN, και συγκεκριμένα με έναν αλγόριθμο τυχαίων διακυμάνσεων πολλαπλής κλίμακας με χρήση τριών μοντέλων AR(1), 600000 τιμές και κάθε φορά αφαιρούσαμε μια τιμή και τη συμπληρώναμε με τη πιο πάνω μεθοδολογία. Στη συνέχεια υπολογίζαμε το MSE που προέκυπτε με χρήση κάθε φορά διαφορετικού πλήθους γειτονικών τιμών *n*. Τα αποτελέσματα και πάλι ήταν ιδιαίτερα ενθαρρυντικά, και παρουσιάζονται στο πιο κάτω διάγραμμα.




4.9 Θεωρητικές και προσομοιωμένες καμπύλες MSE - *n* για διάφορες τιμές του συντελεστή Hurst (*H*).

Παρατηρούμε πως οι προσομοιωμένες τιμές του MSE σχεδόν συμπίπτουν με τις αντίστοιχες θεωρητικές καμπύλες που προέκυψαν από εφαρμογή της σχέσης (4.21).

*4.3.2.3 Σχόλια – παρατηρήσεις*

Με τη πιο πάνω μεθοδολογία, παρατηρούμε πως η δομή της αυτοσυσχέτισης των χρονοσειρών, διαδραματίζει καθοριστικό ρόλο στο πρόβλημα της συμπλήρωσης των ελλειπουσών τιμών. Πιο συγκεκριμένα, εξετάσαμε δυο περιπτώσεις ανελίξεων, με διαφορετική δομή αυτοσυσχέτισης, τις ανελίξεις Markov και τις ανελίξεις με δυναμική ΗΚ, και τα αποτελέσματα δείχνουν πως όσο η δομή της αυτοσυσχέτισης μεταξύ των τιμών της χρονοσειράς γίνεται εντονότερη, τόσο περιορίζεται και ο αριθμός των γειτονικών τιμών που χρειάζεται να ληφθούν υπόψη για την συμπλήρωση της κενής τιμής.

Στις χρονοσειρές που προσομοιώνονται με ανελίξεις Markov, ο βέλτιστος αριθμός γειτονικών τιμών που πρέπει να λάβουμε υπόψη για τη συμπλήρωση μιας ελλείπουσας τιμής καθορίζεται από μια κρίσιμη τιμή του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_{cr}$). Ειδικότερα, για

- $\rho_1 < \rho_{cr} \approx 0.24$

  προτιμάται η χρήση του ολικού μέσου όρου

- $\rho_1 \geq \rho_{cr} \approx 0.24$

  η χρήση ενός τοπικού μέσου χρησιμοποιώντας μια τιμή πριν και μία μετά την ελλείπουσα τιμή ενδείκνυται.

Αντίθετα, στις χρονοσειρές που παρουσιάζουν δυναμική ΗΚ, η χρήση ενός τοπικού μέσου όρου με βάση το βέλτιστο πλήθος γειτονικών τιμών $n$ που παρουσιάστηκε στην πιο πάνω μεθοδολογία, προτιμάται από τη χρήση του ολικού μέσου όρου. Ανάλογα λοιπόν το συντελεστή Hurst της χρονοσειράς, ο αριθμός των γειτονικών τιμών που απαιτούντα για τη συμπλήρωση περιορίζεται αρκετά, με αποτέλεσμα το πρόβλημα της συμπλήρωσης ελλειπουσών τιμών να απλοποιείται αρκετά και η υπολογιστική διαδικασία να γίνεται ταχύτερη. Πιο συγκεκριμένα, για τη συμπλήρωση ελλείπουσας τιμής σε μια χρονοσειρά με συντελεστή Hurst

- $H = 0.50 - 0.60$,
  προτιμάται η χρήση του ολικού μέσου όρου
- $H = 0.70$
  απαιτούνται 4 τιμές πριν και 4 μετά την ελλείπουσα τιμή
- $H = 0.72$
  απαιτούνται 3 τιμές πριν και 3 μετά την ελλείπουσα τιμή
- $H = 0.74$
  απαιτούνται 2 τιμές πριν και 2 μετά την ελλείπουσα τιμή
- $H \geq 0.80$
  χρειαζόμαστε μόνο μία τιμή πριν και μία μετά

Συνοψίζοντας λοιπόν τη πιο πάνω μεθοδολογία, πρέπει να υπογραμμίσουμε, πως παρά την απλότητα που παρουσιάζει υπολογιστικά, τα αποτελέσματα της εφαρμογής της είναι ιδιαιτέρως ικανοποιητικά και διευκολύνον αισθητά την αντιμετώπιση του

δύσκολου αυτού προβλήματος της συμπλήρωσης μεμονωμένων ελλειπουσών τιμών στις υδρομετεωρολογικές χρονοσειρές.

Η εφαρμογή της μεθόδου είναι πολύ απλή και γρήγορη υπολογιστικά, καθώς το μόνο που απαιτείται είναι είτε η εκτίμηση του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$), αν η χρονοσειρά μας προσομοιώνεται από ανελίξεις Markov, είτε η εκτίμηση του συντελεστή Hurst, αν η χρονοσειρά που εξετάζουμε παρουσιάζει δυναμική ΗΚ. Στη συνέχεια, ανατρέχουμε στο αντίστοιχό διάγραμμα και για το συντελεστή $\rho_1$ ή το συντελεστή Hurst της χρονοσειράς που μελετάμε, διαβάζουμε ποια καμπύλη δίνει το ελάχιστο MSE και έχουμε έτσι υπολογίσει τον απαιτούμενο για τη συμπλήρωση αριθμό των γειτονικών τιμών.

### *4.3.3 Χρήση των δυο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών*

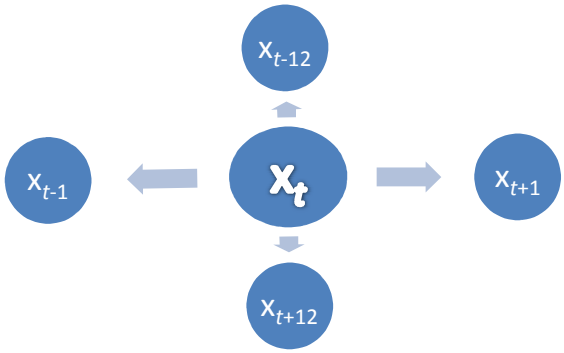
Όπως διαπιστώθηκε στην πιο πάνω προσέγγιση, η αυτοσυσχέτιση διαδραματίζει καθοριστικό ρόλο στην επίλυση του προβλήματος της συμπλήρωσης ελλείπων υδρομετεωρολογικών δεδομένων. Συγκεκριμένα διαπιστώθηκε πως όσο πιο ισχυρή είναι η αυτοσυσχέτιση μεταξύ των τιμών της χρονοσειράς, τόσο πιο εύστοχη είναι η χρήση ενός τοπικού μέσου όρου έναντι του ολικού. Μάλιστα, για πολύ υψηλές τιμές του συντελεστή αυτοσυσχέτισης, αποδείξαμε πως ο τοπικός μέσος που απαιτείται για τη συμπλήρωση, περιορίζεται σε δύο μόλις γειτονικά βήματα, ένα πριν και ένα μετά την ελλείπουσα τιμή. Λαμβάνοντας υπόψη αυτή τη διαπίστωση, επεκτείναμε την πιο πάνω μεθοδολογία και την απλουστεύσαμε αισθητά, προκειμένου να γίνει ακόμη πιο εύκολη η εφαρμογή της.

Συγκεκριμένα, εξετάσαμε λεπτομερέστερα, το πρόβλημα της συμπλήρωσης μηνιαίων υδρομετεωρολογικών χρονοσειρών και κάναμε τις εξής πρόσθετες παραδοχές:

- Για λόγους απλότητας αλλά και άμεσης εφαρμογής, θεωρήσαμε ίσα βάρη μεταξύ των τιμών του δείγματος που θα χρησιμοποιηθούν για τη συμπλήρωση
- Θεωρήσαμε επίσης πως θα λάβουμε υπόψη μόνο δύο γειτονικές τιμές, μια πριν και μια μετά την ελλείπουσα τιμή
- Προκειμένου να βελτιώσουμε την εκτίμηση, θα λάβουμε υπόψη και δύο επιπλέον γειτονικές τιμές αλλά στην ετήσια κλίμακα, δηλαδή την τιμή του αντίστοιχου μήνα που θέλουμε να συμπληρώσουμε, ένα χρόνο πριν και ένα χρόνο μετά την ελλείπουσα τιμή
- Επειδή όμως στη μηνιαία χρονοσειρά παρουσιάζεται το πρόβλημα της περιοδικότητας, πρέπει να τυποποιήσουμε τις τιμές της χρονοσειράς με κατάλληλους μετασχηματισμούς ώστε να μην παρουσιάζονται φαινόμενα περιοδικότητας και όλες οι τιμές της χρονοσειράς να βρίσκονται στο διάστημα [0,1].

Συνοψίζοντας λοιπόν τις πιο πάνω παραδοχές, θα χρησιμοποιήσουμε συνολικά για την εκτίμηση της ελλείπουσας τιμής τέσσερεις μόνο «γειτονικές» τιμές, δύο γειτονικές στη μηνιαία κλίμακα και δύο γειτονικές στην ετήσια κλίμακα.

Αναλυτικά, οι τιμές που θα χρησιμοποιηθούν για την πρόγνωση φαίνονται στο πιο κάτω σχήμα:




Εικόνα 4.1 Τέσσερεις «γειτονικές» χρονικά τιμές που χρησιμοποιούνται στην πρόγνωση της ελλείπουσας τιμής με ίσους συντελεστές βαρύτητας. Με $x_t$ συμβολίζεται η τιμή που λείπει τον μήνα $t$, με $x_{t-1}$ η αντίστοιχη τιμή ένα μήνα πριν και με $x_{t+1}$ ένα μήνα μετά ενώ με $x_{t-12}$ η αντίστοιχη τιμή ένα χρόνο πριν και με $x_{t+12}$ ένα χρόνο μετά.

Για να γίνει πιο κατανοητή η εν λόγω μεθοδολογία, παρουσιάζεται το εξής παράδειγμα. Έστω ότι επεξεργαζόμαστε τα βροχομετρικά δεδομένα που προέρχονται από τον σταθμό Αλιάρτου, απόσπασμα των οποίων φαίνεται στο παρακάτω πίνακα.

Πίνακας 1 Απόσπασμα χρονοσειρών βροχόπτωσης (σε mm) στο σταθμό Αλιάρτου.[22]

| Υδρ. Έτος | Οκτ | Νοε | Δεκ | Ιαν | Φεβ | Μαρ | Απρ | Μαϊ | Ιουν | Ιουλ | Αυγ | Σεπ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1924-25 | 103.30 | 104.00 | 40.30 | 30.50 | 90.60 | 176.90 | 43.20 | 80.80 | 29.20 | 46.90 | 0.00 | 0.00 |
| 1925-26 | 45.50 | 112.00 | 31.90 | 130.40 | 67.40 | 39.30 | 11.40 | 27.90 | 6.60 | 2.30 | 6.10 | 3.80 |
| 1926-27 | 1.30 | 46.70 | 221.50 | 70.20 | 81.30 | 51.60 | 61.90 | 19.80 | 0.00 | 0.00 | 5.80 | 18.00 |
| 1927-28 | 237.40 | 9.40 | 138.80 | 331.80 | 67.30 | 141.70 | 34.00 | 5.80 | 0.00 | 0.00 | 0.00 | 7.90 |
| 1928-29 | 19.70 | 219.80 | 99.10 | 79.20 | 127.20 | 37.50 | 17.10 | 8.90 | 4.60 | 0.00 | 0.00 | 124.20 |
| 1929-30 | 73.10 | 112.00 | 65.50 | 90.10 | 233.40 | 49.40 | 111.00 | 87.30 | 83.50 | 50.50 | 0.00 | 55.60 |
| 1930-31 | 70.10 | 82.90 | 131.40 | 92.00 | 135.60 | $x_t$ | 77.20 | 75.00 | 30.20 | 0.00 | 1.30 | 13.70 |
| 1931-32 | 63.50 | 44.00 | 210.10 | 52.80 | 118.20 | 159.10 | 11.20 | 24.40 | 26.70 | 0.00 | 41.10 | 0.50 |
| 1932-33 | 14.70 | 60.40 | 16.60 | 139.20 | 78.40 | 13.50 | 35.60 | 37.00 | 41.80 | 6.60 | 6.90 | 12.70 |
| 1933-34 | 44.70 | 42.60 | 169.70 | 138.20 | 153.40 | 114.90 | 18.30 | 22.40 | 15.50 | 11.40 | 0.00 | 1.00 |
| 1934-35 | 34.40 | 45.20 | 114.60 | 184.90 | 71.20 | 99.80 | 8.40 | 4.30 | 20.60 | 2.50 | 0.00 | 0.00 |

[22] Ο πλήρης πίνακας είναι διαθέσιμος στον ιστότοπο www.itia.ntua.gr/courses/stochwatres/more/

Έστω λοιπόν ότι θέλουμε να συμπληρώσουμε την τιμή $x_t$ που αντιστοιχεί στη βροχόπτωση το μήνα Μάρτιο, το υδρολογικό έτος 1930-31. Συμφώνα με την προτεινόμενη μεθοδολογία, θα πρέπει λάβουμε υπόψη μας στην πρόβλεψη της ελλείπουσας τιμής, την τιμή της βροχόπτωσης τον προηγούμενο μήνα, δηλαδή το Φεβρουάριο, αλλά και τον επόμενο, δηλαδή τον Απρίλιο καθώς επίσης και τις αντίστοιχες τιμές ένα χρόνο πριν, δηλαδή το υδρολογικό έτος 1929-30 και ένα χρόνο μετά, δηλαδή το υδρολογικό έτος 1931-32.

Προκειμένου όμως να λάβουμε περισσότερο υπόψη τις τιμές $x_{t-1}$ και $x_{t+1}$ που έχουν ισχυρότερη συσχέτιση με την $x_t$ θα χρησιμοποιήσουμε ένα συντελεστή βαρύτητας $\theta$ και έτσι η εκτίμηση της ελλείπουσας τιμής $x_t$ που θα τη συμβολίζουμε $\hat{x}_t$ θα ισούται με

$$\hat{x}_t = \theta \frac{x_{t-1} + x_{t+1}}{2} + (1-\theta) \frac{x_{t-12} + x_{t+12}}{2} \tag{4.23}$$

Και το μέσο τετραγωνικό σφάλμα της εκτίμησης[23] (MSE) θα είναι

$$\text{MSE} = E\left[e^2\right] = \sigma^2 \left[ \begin{array}{l} 1 - 2\theta(\rho_1 - \rho_{12}) - 2\rho_{12} + \frac{\theta^2}{2}(\rho_2 + 1) + \\ \frac{(1-\theta)^2}{2}(\rho_{24} + 1) + \theta(1-\theta)(\rho_{11} + \rho_{13}) \end{array} \right] \tag{4.24}$$

Εξετάζουμε πάλι, δύο τύπους στοχαστικών ανελίξεων, τις ανελίξεις Markov και τις ανελίξεις με δυναμική ΗΚ, που παρουσιάζουν διαφορετική δομή αυτοσυσχέτισης.

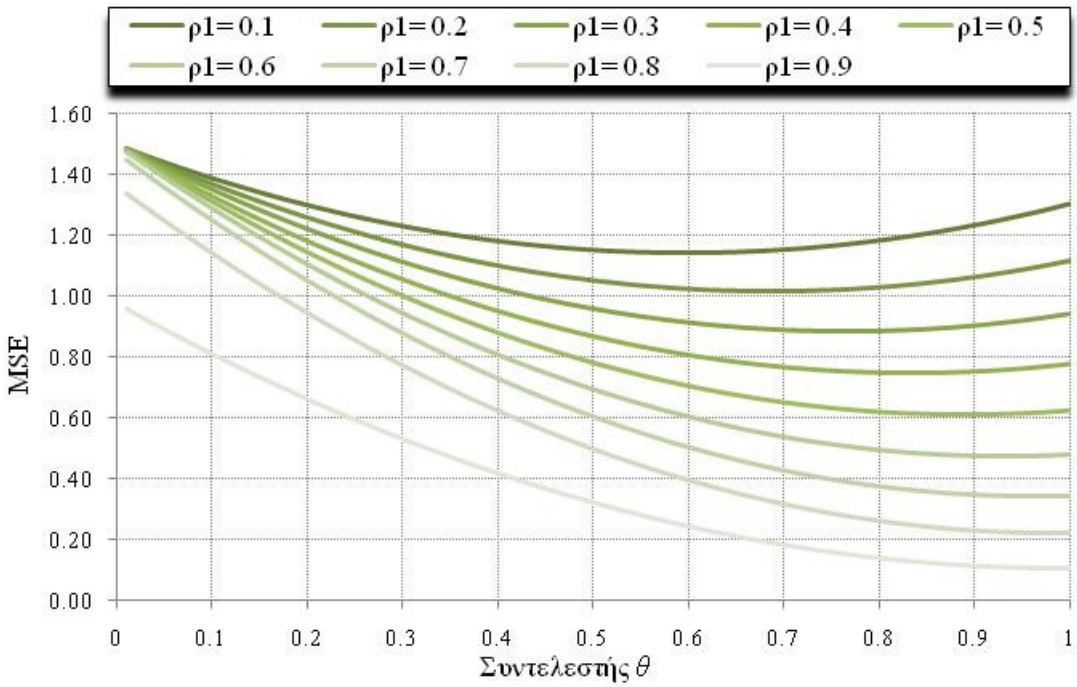
#### *4.3.3.1 Χρονοσειρές που προσομοιώνονται με ανελίξεις Markov*

Αν στη σχέση (4.24) υπολογίσουμε τους συντελεστές αυτοσυσχέτισης που παρουσιάζονται, για διάφορες τιμές της υστέρησης $j$, από τη σχέση (4.22), δηλαδή

$$\rho_j = \left(\rho_1\right)^j$$

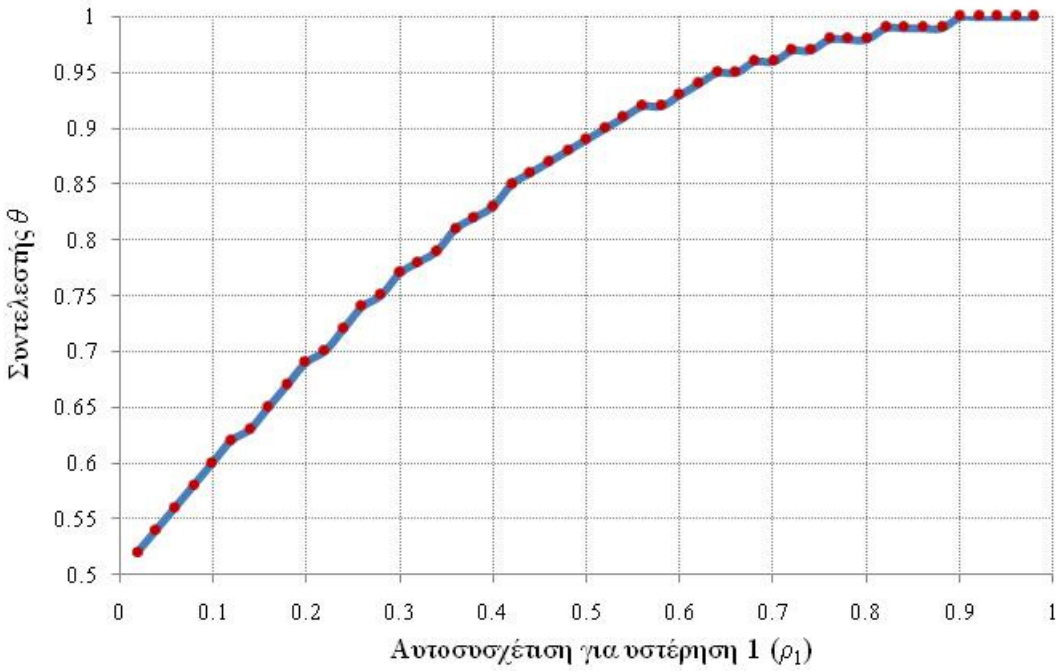
τότε μπορούμε να εκτιμήσουμε το MSE συναρτήσει του συντελεστή βαρύτητας $\theta$. Στο επόμενο διάγραμμα παρουσιάζονται καμπύλες που αντιστοιχούν σε διαφορετικούς συντελεστές αυτοσυσχέτισης για υστέρηση 1 και φαίνεται η σχέση μεταξύ του μέσου τετραγωνικού σφάλματος της εκτίμησης (MSE) και της αντίστοιχης τιμής της παραμέτρου $\theta$.

[23] Για τον αναλυτικό υπολογισμό του MSE βλέπε ΠΑΡΑΡΤΗΜΑ Β.

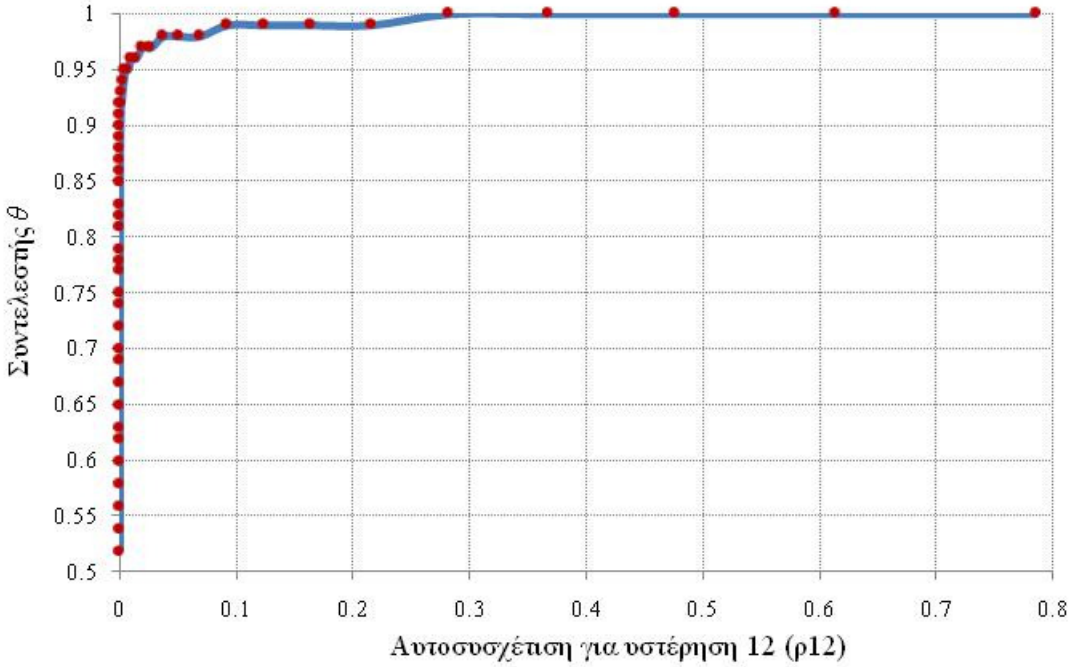



Σχήμα 4.10 MSE - $\theta$ για διάφορες τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$)

Στο πιο πάνω διάγραμμα παρατηρούμε πως για μικρές τιμές του συντελεστή $\rho_1$, η αντίστοιχη παράμετρος $\theta$ για την οποία ελαχιστοποιείτε το MSE παίρνει τιμές κοντά στο 0.5 και συνεπώς ο συντελεστής $(1-\theta)$ θα κυμαίνεται και αυτός κοντά στο 0.5. Συμπεραίνουμε λοιπόν ίσους περίπου συντελεστές βαρύτητας για τις τιμές $\left(x_{t-1}+x_{t+1}\right)$ και $\left(x_{t-12}+x_{t+12}\right)$. Αντίθετα για μεγαλύτερες τιμές του συντελεστή $\rho_1$, η παράμετρος $\theta$ είναι σημαντικά μεγαλύτερη, δηλαδή οι γειτονικές τιμές $\left(x_{t-1}+x_{t+1}\right)$ έχουν μεγαλύτερη βαρύτητα από τις $\left(x_{t-12}+x_{t+12}\right)$. Οι πιο πάνω διαπιστώσεις γίνονται ευκολότερα κατανοητές από τα παρακάτω διαγράμματα:




Σχήμα 4.11 Παράμετρος $\theta$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή $\rho_1$.




Σχήμα 4.12 Παράμετρος $\theta$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή $\rho_{12}$.

### *4.3.3.2 Χρονοσειρές που παρουσιάζουν δυναμική Hurst – Kolmogorov*

Στην περίπτωση που οι χρονοσειρές που εξετάζουμε αναπαράγουν το φαινόμενο Hurst, τότε η δομή της αυτοσυσχέτισης μεταξύ των τιμών είναι αρκετά ισχυρή και η

συνάρτηση που συνδέει την αυτοσυσχέτιση με τις διάφορες τιμές της υστέρησης $j$ δίνεται από την σχέση (4.9), δηλαδή

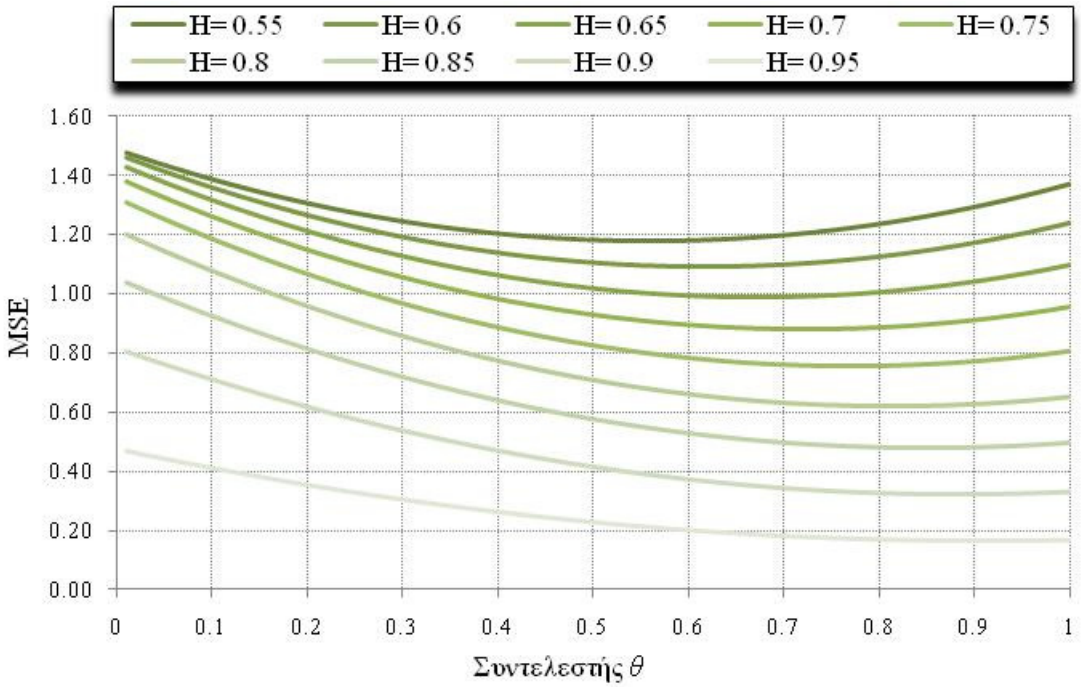
$$\rho_j = \left(\frac{1}{2}\right)\left[(j+1)^{2H} + (j-1)^{2H}\right] - j^{2H}$$

Αν λοιπόν υπολογίσουμε τους συντελεστές αυτοσυσχέτισης από την πιο πάνω σχέση, και εφαρμόσουμε την εξίσωση (4.24)

$$\mathrm{MSE} = E\left[e^2\right] = \sigma^2 \left[\begin{array}{l} 1 - 2\theta(\rho_1 - \rho_{12}) - 2\rho_{12} + \\ \dfrac{\theta^2}{2}(\rho_2 + 1) + \dfrac{(1-\theta)^2}{2}(\rho_{24} + 1) + \theta(1-\theta)(\rho_{11} + \rho_{13}) \end{array}\right]$$

μπορούμε να υπολογίσουμε το MSE συναρτήσει της παραμέτρου $\theta$.

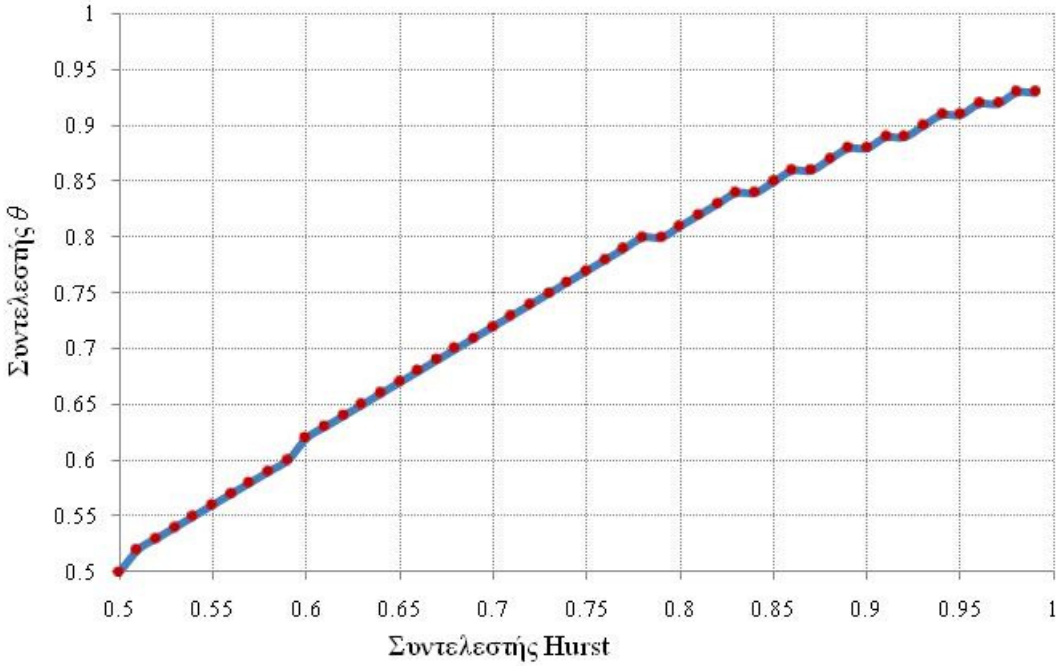
Έτσι, προκύπτουν οι πιο κάτω καμπύλες



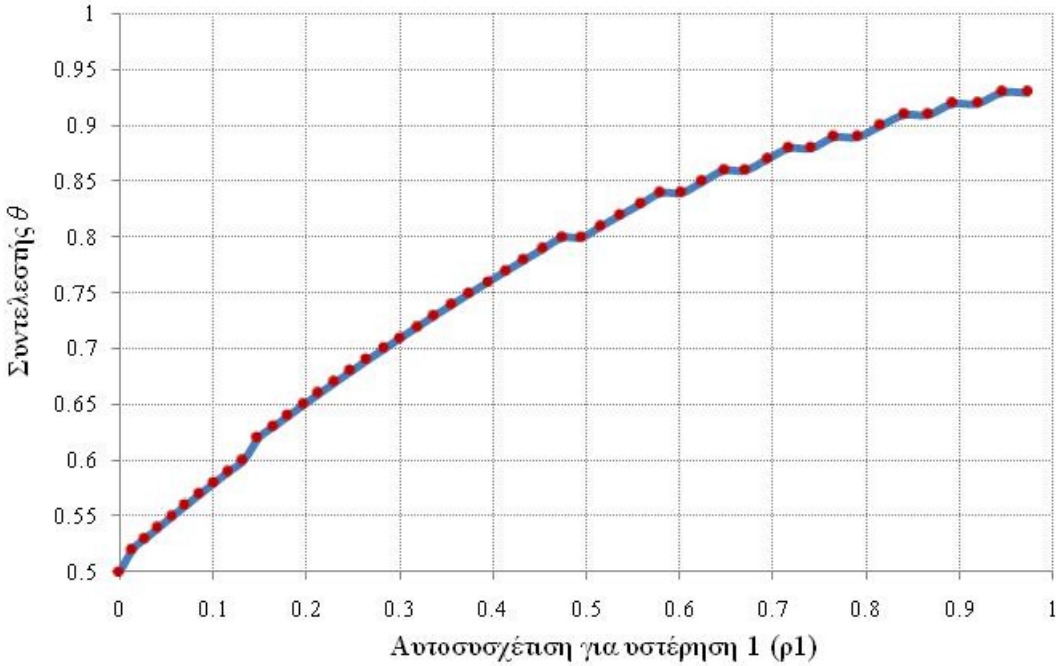

Σχήμα 4.13 MSE - $\theta$ για διάφορες τιμές του συντελεστή Hurst ($H$)

Παρατηρούμε, πως για ισχυρή δομή αυτοσυσχέτισης, δηλαδή υψηλές τιμές του συντελεστή Hurst, η παράμετρος $\theta$ για την οποία ελαχιστοποιείτε το MSE κυμαίνεται πάνω από 0.8 γεγονός που σηματοδοτεί μεγάλη βαρύτητα στις τιμές $\left(x_{t-1} + x_{t+1}\right)$ και αντίστοιχα μικρότερη στις $\left(x_{t-12} + x_{t+12}\right)$.

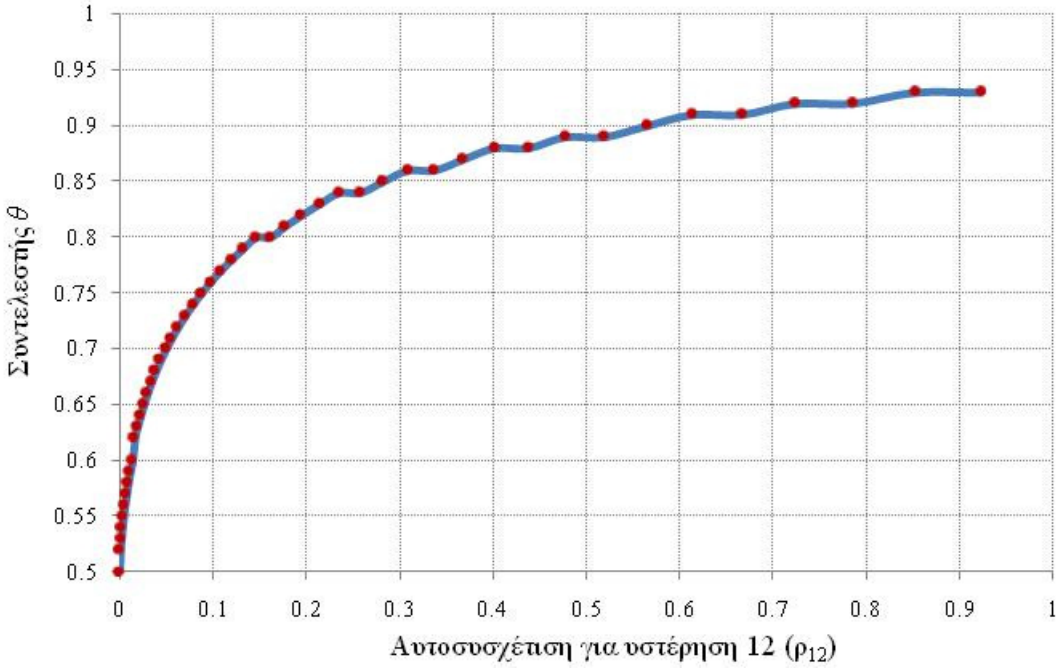
Τα πιο πάνω συμπεράσματα, φαίνονται ξεκάθαρα στα επόμενα γραφήματα.




Σχήμα 4.14 Παράμετρος $\theta$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή Hurst.




Σχήμα 4.15 Παράμετρος $\theta$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή $\rho_1$.




Σχήμα 4.16 Παράμετρος $\theta$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή $\rho_{12}$.

Στα πιο πάνω γραφήματα, γίνεται φανερό πως όσο αυξάνει ο συντελεστής Hurst και συνεπώς ο συντελεστής $\rho_1$ τόσο αυξάνει και η τιμή της παραμέτρου $\theta$, δηλαδή οι γειτονικές τιμές $\left(x_{t-1}+x_{t+1}\right)$ έχουν μεγαλύτερη βαρύτητα στην εκτίμηση της ελλείπουσας τιμής.

Επίσης παρατηρούμε πως ο συντελεστής Hurst έχει την ίδια συμπεριφορά με το συντελεστή $\rho_1$ και περιγράφει παρόμοια τη σχέση του MSE με την παράμετρο $\theta$.

### *4.3.3.3 Σχόλια – παρατηρήσεις*

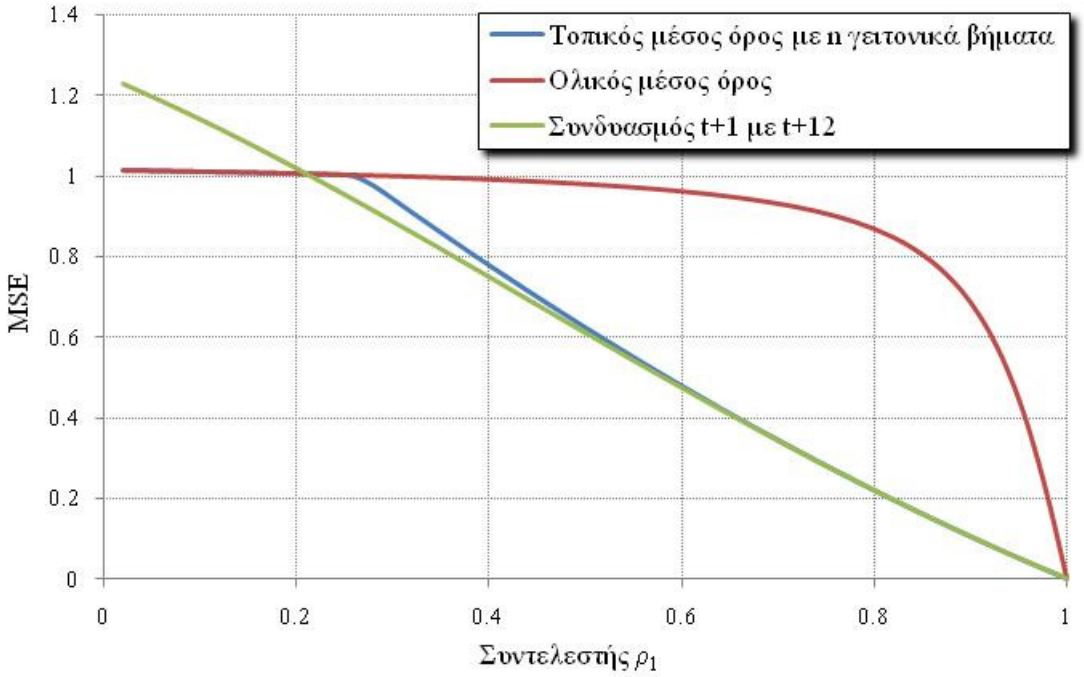
Η εφαρμογή της πιο πάνω μεθοδολογίας είναι ιδιαίτερα εύκολη και αρκετά γρήγορη. Τα αποτελέσματα επίσης που προκύπτουν από τη χρήση της είναι πολύ ικανοποιητικά. Πιο συγκεκριμένα, παρατηρούμε, πως αν η χρονοσειρά που εξετάζουμε έχει πολύ ισχυρή δομή αυτοσυσχέτισης, τότε το μόνο που χρειαζόμαστε προκειμένου να συμπληρώσουμε μια τιμή που λείπει είναι οι γειτονικές τιμές, όπως πολύ εποπτικά φαίνεται στον πίνακα με τις τιμές της βροχόπτωσης στην Αλίαρτο.

Συγκρίνοντας τώρα τα αποτελέσματα αυτής της μεθοδολογίας και της προσέγγισης που παρουσιάστηκε στο υποκεφάλαιο 4.3.2, καταλήγουμε στα πιο κάτω διαγράμματα.

Για την περίπτωση χρονοσειρών τύπου Markov, η πρώτη μέθοδος, δηλαδή αυτή που παρουσιάστηκε στο υποκεφάλαιο 4.3.2.1 καθώς επίσης και η χρήση του ολικού μέσου όρου, φαίνεται να έχει καλύτερα αποτελέσματα για μικρές τιμές του συντελεστή αυτοσυσχέτισης (για $\rho_1 \leq 0.2$), ενώ για τιμές στο διάστημα $0.2 \leq \rho_1 \leq 0.5$

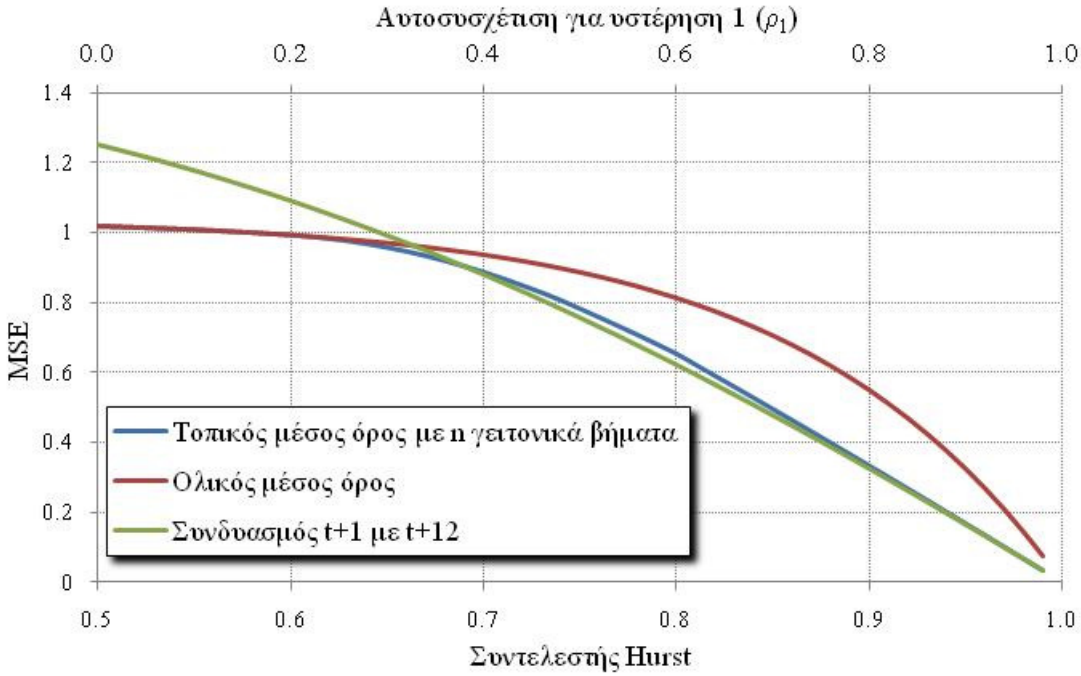
υπερτερεί η μέθοδος που βασίζεται στη χρήση τεσσάρων γειτονικών τιμών. Για τιμές του συντελεστή αυτοσυσχέτισης $\rho_1$ μεγαλύτερες του 0.5, η μέθοδος με το βέλτιστο αριθμό γειτονικών βημάτων που παρουσιάστηκε στο υποκεφάλαιο 4.3.2 και η μέθοδος που βασίζεται στη χρήση τεσσάρων γειτονικών τιμών που παρουσιάζεται εδώ, έχουν παρόμοια αποτελέσματα.

Αναλυτικά, τα πιο πάνω συμπεράσματα φαίνονται στο παρακάτω διάγραμμα.




Σχήμα 4.17 Ελάχιστη τιμή του MSE – συντελεστής αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) με εφαρμογή των μεθόδων που έχουν παρουσιαστεί μέχρι στιγμής.

Στη περίπτωση χρονοσειρών που παρουσιάζουν το φαινόμενο Hurst, η μέθοδος του κεφαλαίου 4.3.2 καθώς και ο ολικός μέσος όρος έχουν καλύτερα αποτελέσματα για $H \leq 0.65$, όπως φαίνεται και στο πιο κάτω διάγραμμα, ενώ για $H \geq 0.65$ η μέθοδος με το βέλτιστο αριθμό γειτονικών βημάτων και η αντίστοιχη που στηρίζεται στη χρήση τεσσάρων γειτονικών τιμών που παρουσιάζεται εδώ, δίνουν παρόμοια αποτελέσματα. Τα πιο πάνω επιβεβαιώνονται στο ακόλουθο διάγραμμα.




Σχήμα 4.18 Ελάχιστη τιμή του MSE – συντελεστής $H$ (κύριος οριζόντιος άξονας) και συντελεστής $\rho_1$ (δευτερεύον οριζόντιος άξονας) με εφαρμογή των μεθόδων που έχουν παρουσιαστεί μέχρι στιγμής.

### *4.3.4 Συνδυασμός ολικού και τοπικού μέσου όρου*

Στα πλαίσια της διερεύνησης της χρήσης ενός τοπικού μέσου όρου στη θέση του ολικού, εξετάσαμε την πιο κάτω παραλλαγή. Συγκεκριμένα, προσπαθήσαμε να συνδυάσουμε τον ολικό μέσο όρο, με ένα τοπικό που θα αποτελείται από $2n$ γειτονικά βήματα, δηλαδή $n$ παρατηρήσεις πριν και $n$ μετά την ελλείπουσα τιμή. Έτσι, η εκτίμηση της ελλείπουσας τιμής προκύπτει από το άθροισμα του ολικού μέσου όρου, πολλαπλασιασμένο με μια παράμετρο $\lambda$, και του τοπικού μέσου που αποτελείται από $n$ χρονικά βήμα πριν και $n$ μετά την ελλείπουσα τιμή, πολλαπλασιασμένο με το συντελεστή (1-$\lambda$), δηλαδή, η εκτίμηση της ελλείπουσας τιμής δίνεται από τον τύπο:

$$\hat{x}_t = \lambda \frac{\sum_{i=-N}^{N} x_i}{2N} + (1-\lambda) \frac{\sum_{i=-n}^{-1} x_i + \sum_{i=1}^{n} x_i}{2n} \qquad (4.25)$$

Ο τοπικός μέσος όρος περιορίστηκε εξαρχής σε μόλις δύο γειτονικές τιμές, δηλαδή ο πιο πάνω τύπος εφαρμόστηκε για $n$=1. Η επιλογή του συγκεκριμένου τοπικού μέσου όρου έγινε αφενός για την ευκολότερη και πιο άμεση εφαρμογή της μεθόδου και αφετέρου γιατί η χρήση περισσότερων γειτονικών τιμών δεν θα είχε ουσιαστικό νόημα στη σύγκριση που επιχειρούμε να κάνουμε μεταξύ του τοπικού και του ολικού μέσου όρου. Έτσι, η εκτίμηση που θα χρησιμοποιήσουμε για την ελλείπουσα τιμή δίνεται από τη σχέση

$$\hat{x}_t = \lambda \frac{\sum_{i=-N}^{N} x_i}{2N} + (1-\lambda)\frac{x_{-1}+x_1}{2} \tag{4.26}$$

Διερευνώνται λοιπόν οι διάφορες τιμές της παραμέτρου λ στο διάστημα [0,1] για τις οποίες ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα. Το μέσο τετραγωνικό σφάλμα της εκτίμησης δίνεται από τη σχέση:

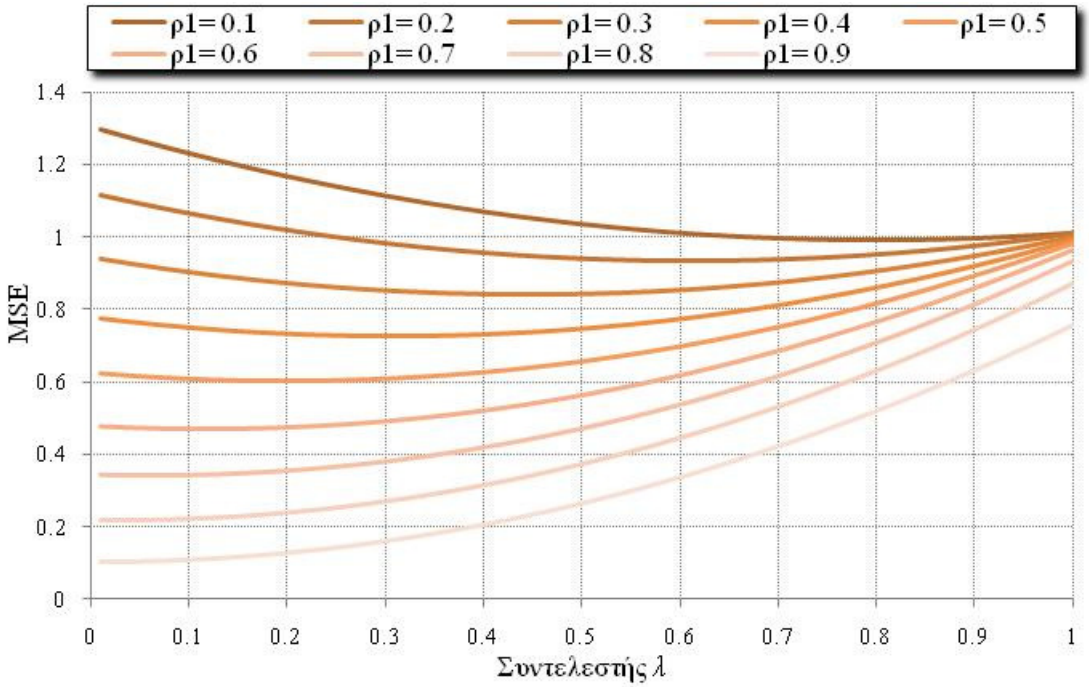
$$\begin{aligned} MSE &:= E\left[e^2\right] \\ &= \frac{1}{2}\sigma^2(3-4\rho_1+\rho_2) \\ &-2\lambda\sigma^2\left[\frac{1}{N}\sum_{i=1}^{N}\rho_t - \frac{1}{2N}\left(\sum_{i=1}^{N-1}\rho_i - \sum_{i=2}^{N+1}\rho_i + 1\right) - \rho_1 + \frac{\rho_2}{2} + 0.5\right] \\ &+\lambda^2\sigma^2\left[\begin{array}{l}\frac{1}{2N^2}\left(2\sum_{i=1}^{N-1}(N-i)\rho_i + \sum_{i=2}^{N+1}(i-1)\rho_i + \sum_{i=N+2}^{2N}(2N+1-i)\rho_i + N\right) \\ +\frac{\rho_2}{2} + \frac{1}{2} - \frac{1}{N}\left(\sum_{i=1}^{N-1}\rho_i + \sum_{i=2}^{N+1}\rho_i + 1\right)\end{array}\right] \end{aligned} \tag{4.27}$$

Ο αναλυτικός υπολογισμός του μέσου τετραγωνικού σφάλματος (MSE) παρουσιάζεται λεπτομερώς στο ΠΑΡΑΡΤΗΜΑ C.

Και σε αυτή τη μεθοδολογία, όπως και στις προηγούμενες περιπτώσεις που εξετάσαμε, μελετώνται δύο τύπου στοχαστικών ανελίξεων, οι ανελίξεις με ασθενή μνήμη, ανελίξεις Markov και οι ανελίξεις με μακρά μνήμη, που παρουσιάζουν δυναμική Hurst-Kolmogorov. Ανάλογα λοιπόν με τη δομή της αυτοσυσχέτισης της χρονοσειράς που εξετάζουμε, υπολογίζουμε το μέσο τετραγωνικό σφάλμα της εκτίμησης που προκύπτει για διάφορες τιμές της παραμέτρου $\lambda$ και διερευνάμε πότε και κάτω από ποιες προϋποθέσεις αυτό ελαχιστοποιείτε.

#### *4.3.4.1 Χρονοσειρές που προσομοιώνονται με ανελίξεις Markov*

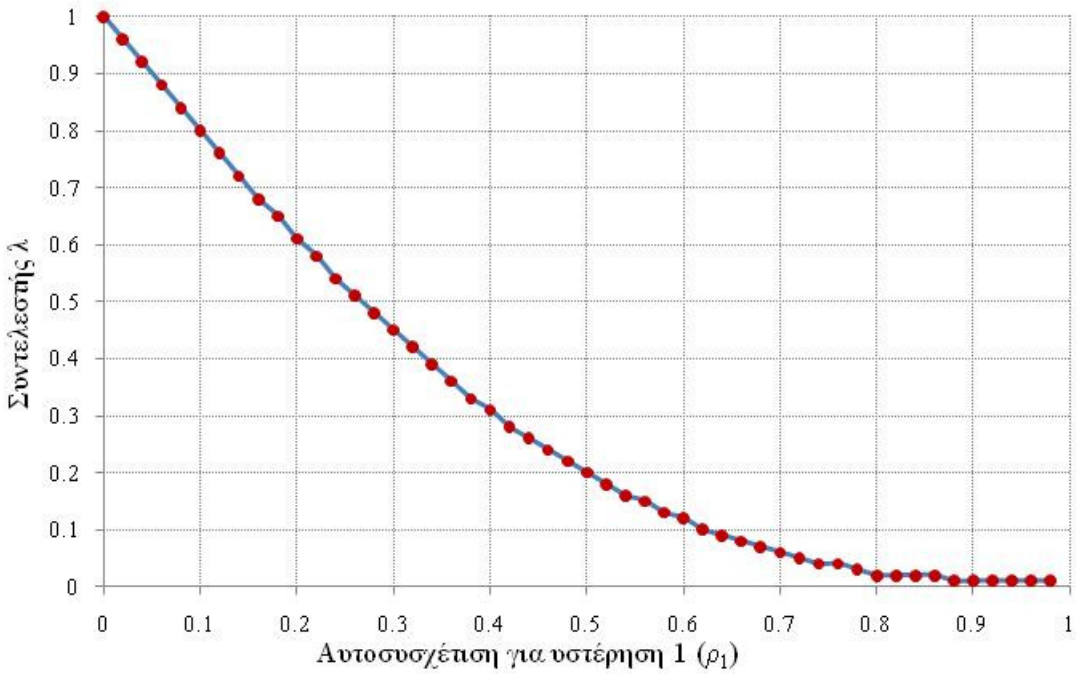
Στις ανελίξεις Markov έχουμε τονίσει πως η δομή της αυτοσυσχέτισης μειώνεται εκθετικά με την υστέρηση. Συγκεκριμένα, η συσχέτιση $\rho$, για διάφορες τιμές της υστέρησης $j$ (lag) δίνεται από την σχέση $\rho_j = \left(\rho_1\right)^j$. Εφαρμόζοντας λοιπόν τη σχέση του MSE με βάση τη δομή της αυτοσυσχέτισης των ανελίξεων Markov για διάφορες τιμές της παραμέτρου $\lambda$, υπολογίζουμε την τιμή του $\lambda$ για την οποία ελαχιστοποιείται το MSE.




Σχήμα 4.19 MSE - $\lambda$ για διάφορες τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$).

Στο πιο πάνω διάγραμμα παρουσιάζονται καμπύλες που αντιστοιχούν σε διαφορετικούς συντελεστές αυτοσυσχέτισης για υστέρηση 1 και φαίνεται η σχέση μεταξύ του μέσου τετραγωνικού σφάλματος της εκτίμησης (MSE) και της αντίστοιχης τιμής της παραμέτρου $\lambda$.

Παρατηρούμε πως όταν ο συντελεστής αυτοσυσχέτισης για υστέρηση 1 παίρνει μικρές τιμές, ο αντίστοιχος συντελεστής $\lambda$ για τον οποίο προκύπτει το μικρότερο MSE είναι κοντά στο 1 ενώ αντίθετα όταν ο συντελεστής αυτοσυσχέτισης για υστέρηση 1 είναι υψηλός, η παράμετρος $\lambda$ παίρνει τιμές κοντά στο 0. Η διαπίστωση αυτή φαίνεται πιο ξεκάθαρα στο επόμενο διάγραμμα.




Σχήμα 4.20 Παράμετρος $\lambda$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή $\rho_1$.

Συγκεκριμένα, παρατηρούμε πώς όταν έχουμε υψηλές τιμές του $\rho_1$ το $\lambda$ τείνει στο 0, δηλαδή στη σχέση

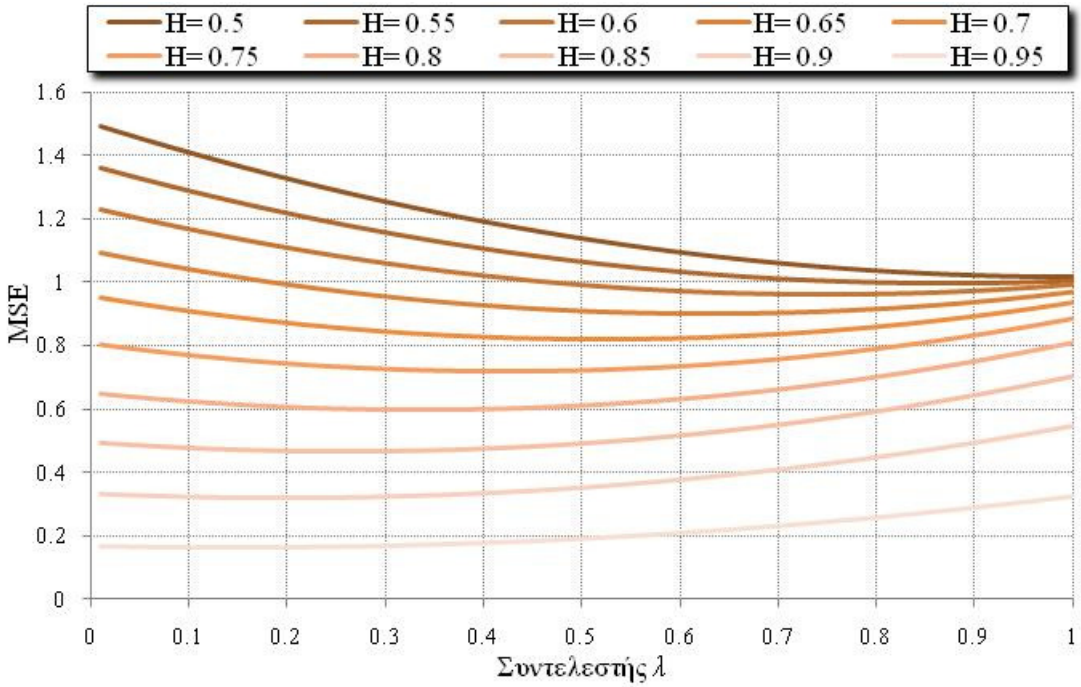
$$\hat{x}_t = \lambda \frac{\sum_{i=-N}^{N} x_i}{2N} + (1-\lambda)\frac{x_{-1}+x_1}{2}$$

ο πρώτος όρος που αντιστοιχεί στον ολικό μέσο όρο τείνει να μηδενιστεί και ο δεύτερος όρος (τοπικός μέσος) έχει αυξημένη βαρύτητα. Το αντίθετο ακριβώς συμβαίνει στην περίπτωση χαμηλών τιμών του $\rho_1$, ο ολικός μέσος έχει μεγαλύτερη βαρύτητα καθώς το $\lambda$ τείνει στο 1 και ο όρος που αντιστοιχεί στον τοπικό μέσο όρο πρακτικώς μηδενίζεται.

### *4.3.4.2 Χρονοσειρές που παρουσιάζουν δυναμική Hurst – Kolmogorov*

Οι χρονοσειρές που αναπαράγουν το φαινόμενο Hurst, έχουν ισχυρή δομή αυτοσυσχέτισης και η συνάρτηση που συνδέει την αυτοσυσχέτιση με τις διάφορες τιμές της υστέρησης $j$ δίνεται από την σχέση $\rho_j = \left(\frac{1}{2}\right)\left[(j+1)^{2H} + (j-1)^{2H}\right] - j^{2H}$.

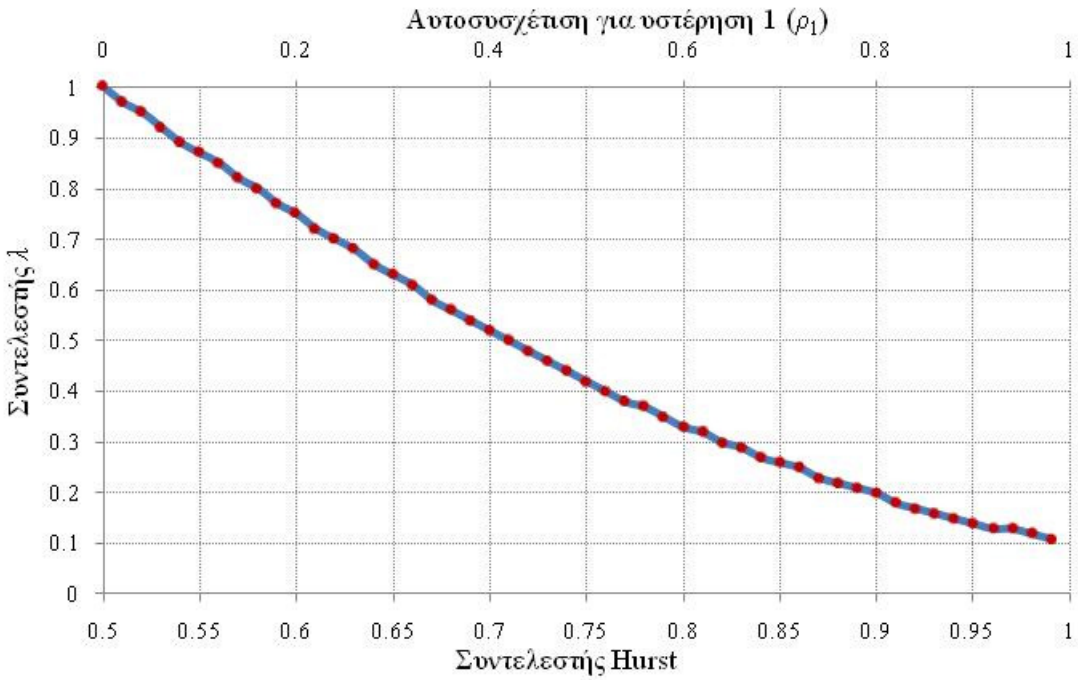
Οι τιμές του MSE για διάφορες τιμές της παραμέτρου $\lambda$, που προκύπτουν για ανελίξεις με συντελεστές Hurst από 0.5 έως 0.95, φαίνονται στο πιο κάτω διάγραμμα.




Σχήμα 4.21 MSE - $\lambda$ για διάφορες τιμές του συντελεστή Hurst ($H$)

Παρατηρούμε, πως για ισχυρή δομή αυτοσυσχέτισης, δηλαδή υψηλές τιμές του συντελεστή Hurst ο συντελεστής $\lambda$ κυμαίνεται κοντά στο 0, γεγονός που σηματοδοτεί μεγάλη βαρύτητα στον τοπικό μέσο όρο και μηδενισμός πρακτικά του ολικού μέσου όρου. Αντίθετα, για ασθενή δομή αυτοσυσχέτισης, δηλαδή χαμηλές τιμές του συντελεστή Hurst ο ολικός μέσος όρος υπερτερεί έναντι του ολικού και έχει μεγαλύτερη βαρύτητα στην εκτίμηση της ελλείπουσας τιμής.

Στο επόμενο διάγραμμα φαίνεται ξεκάθαρα, πώς ανάλογα με το πόσο ισχυρή ή ασθενής είναι η αυτοσυσχέτιση τόσο αυξάνεται ή μειώνεται αντίστοιχα η συμβολή του τοπικού μέσου όρου.




Σχήμα 4.22 Παράμετρος $\lambda$ για την οποία προκύπτει το μικρότερο MSE συναρτήσει του συντελεστή Hurst (κύριος οριζόντιος άξονας) και του συντελεστή αυτοσυσχέτισης για υστέρηση 1 (δευτερεύον οριζόντιος άξονας).

### *4.3.4.3 Σχόλια – παρατηρήσεις*

Σκοπός αυτής της παραλλαγής, που στηρίζεται στη χρήση ενός τοπικού μέσου όρου σε συνδυασμό με τον αντίστοιχο ολικό, είναι να βελτιωθεί η εκτίμηση της ελλείπουσας τιμής, να μειωθεί δηλαδή κατά το δυνατό το μέσο τετραγωνικό σφάλμα της εκτίμηση.

Υπολογιστικά, ο φόρτος που απαιτεί η μέθοδος είναι ελάχιστος, καθώς σε κάθε περίπτωση (ανελίξεις Markov αλλά και ανελίξεις με δυναμική ΗΚ) το μόνο που απαιτείται είναι ο υπολογισμός του ολικού μέσου όρου και του τοπικού μέσου όρου με μια τιμή πριν και μια μετά την ελλείπουσα τιμή και στη συνέχεια, ανάλογα με την τιμή του συντελεστή $\rho_1$ για τις ανελίξεις Markov και του συντελεστή Hurst για τις ανελίξεις με δυναμική ΗΚ, ο υπολογισμός της παραμέτρου $\theta$ από τα αντίστοιχα διαγράμματα που παρατίθενται.

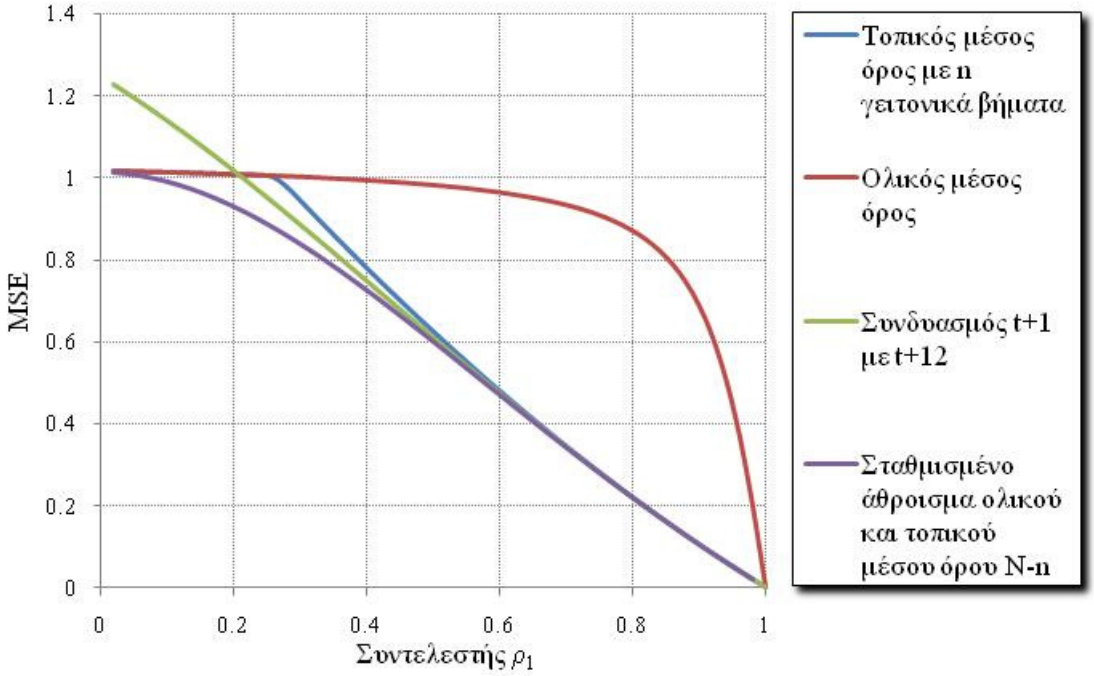
Όσον αφορά τη βελτίωση που προσφέρει αυτή η μέθοδος σε σχέση με τις προηγούμενες, αυτή είναι πολύ σημαντική. Αν δε, λάβουμε υπόψη και τον ελάχιστα επιπλέον υπολογιστικό φόρτο που αυτή απαιτεί, οδηγούμαστε στο συμπέρασμα ότι έχουμε μια βέλτιστη αριθμητικά λύση (ελάχιστο MSE), με το μικρότερο υπολογιστικό φόρτο.

Τα πιο πάνω συμπεράσματα επιβεβαιώνονται στα επόμενα συγκριτικά διαγράμματα καθώς επίσης και στο κεφάλαιο 5 όπου εφαρμόζεται η προτεινόμενη μεθοδολογία σε πραγματικά δεδομένα και δίνει το μικρότερο MSE αλλά και στο

κεφάλαιο 6 όπου συγκρίνονται και σχολιάζονται λεπτομερώς όλες οι προτεινόμενες μεθοδολογίες.

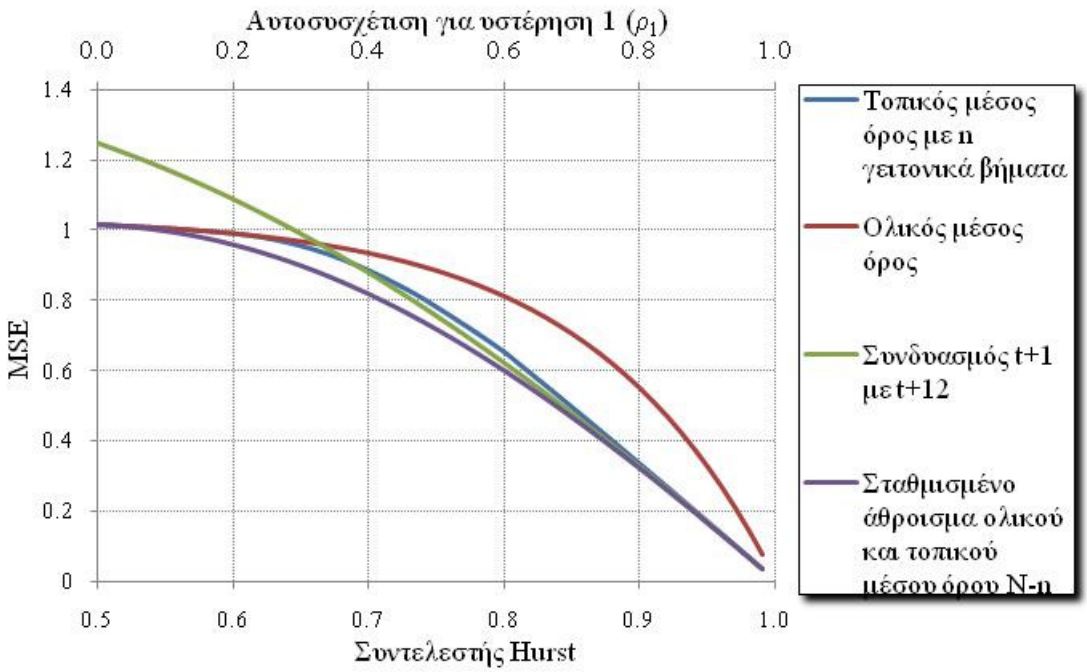
Συγκεκριμένα, για την περίπτωση χρονοσειρών που προσομοιώνονται από ανελίξεις Markov τα αποτελέσματα της σύγκρισης των μεθοδολογιών φαίνονται στο ακόλουθο διάγραμμα.

Παρατηρούμε πως η καμπύλη που αντιστοιχεί στη μεθοδολογία που περιγράψαμε πιο πάνω, δηλαδή στο συνδυασμό του τοπικού και του ολικού μέσου όρου ($N$-$n$), για οποιαδήποτε τιμή του συντελεστή $\rho_1$ δίνει το μικρότερο μέσο τετραγωνικό σφάλμα σε σχέση με τις υπόλοιπες μεθόδους που έχουν παρουσιαστεί μέχρι στιγμής.




Σχήμα 4.23 Ελάχιστη τιμή του MSE – συντελεστής αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) με εφαρμογή των μεθόδων που έχουν παρουσιαστεί μέχρι στιγμής.

Το ίδιο ενθαρρυντικά αποτελέσματα προκύπτουν και στην περίπτωση χρονοσειρών που προσομοιώνονται από ανελίξεις που αναπαράγουν το φαινόμενο Hurst. Πιο συγκεκριμένα, όπως φαίνεται στο επόμενο διάγραμμα, η καμπύλη που αντιστοιχεί στη μεθοδολογία που περιγράψαμε πιο πάνω, δηλαδή στο συνδυασμό του τοπικού και του ολικού μέσου όρου ($N$-$n$) για οποιαδήποτε τιμή του συντελεστή Hurst, δίνει το μικρότερο μέσο τετραγωνικό σφάλμα σε σχέση με τις υπόλοιπες μεθόδους που έχουν παρουσιαστεί μέχρι στιγμής.




Σχήμα 4.24 Ελάχιστη τιμή του MSE – συντελεστής $H$ (κύριος οριζόντιος άξονας) και συντελεστής $\rho_1$ (δευτερεύον οριζόντιος άξονας) με εφαρμογή των μεθόδων που έχουν παρουσιαστεί μέχρι στιγμής.

### *4.3.5 Σταθμισμένα βάρη*

Όπως έχουμε ήδη αναφέρει στο υποκεφάλαιο 4.3.1, το πρόβλημα της συμπλήρωσης των ελλιπών τιμών μπορεί να εκφραστεί μαθηματικά από τη σχέση (4.16) ή σε μορφή πινάκων από τη σχέση (4.17) που είναι:

$$Y = \mathbf{w}^T \mathbf{X} + e$$

όπου:

$$\mathbf{w} := \left[ w_1, ..., w_n \right]^T$$

$$\mathbf{X} := \left[ x_1, ..., x_n \right]$$

Στις προηγούμενες προσεγγίσεις, για λόγους απλότητας, αλλά και ευκολίας στους υπολογισμούς και συνεπώς γρήγορη εφαρμογή και εκτίμηση της ελλείπουσας τιμής, θεωρήσαμε ίσα βάρη μεταξύ των τιμών της χρονοσειράς. Τώρα, θα επιλύσουμε την πιο πάνω εξίσωση, προκειμένου να υπολογίσουμε αναλυτικά τους συντελεστές βαρύτητας $w_i$ που αντιστοιχούν σε κάθε τιμή της χρονοσειράς $X_i$.

Η γεωστατιστική παρέχει μια συστηματική διαδικασία, η οποία είναι κατάλληλη για τον υπολογισμό των διαφόρων συντελεστών βαρύτητας που θα χρησιμοποιηθούν σε ένα πρόβλημα γραμμικής συμπλήρωσης κάποιας ή και κάποιων ελλιπών τιμών. Ο υπολογισμός των βαρών $w_i$, βασίζεται είτε στη συνάρτηση αυτοσυσχέτισης, είτε στο μεταβαλλόγραμμα.

Συγκεκριμένα, οι διάφορες τιμές των $w_i$ επιλέγονται ώστε το σφάλμα της εκτίμησης, δηλαδή η ποσότητα: $e = \left(x_t - \hat{x}_t\right)$, όπου $x_t$ η πραγματική τιμή και $\hat{x}_t$ η εκτίμηση της ελλείπουσας τιμής, να έχει μέση τιμή μηδέν, δηλαδή να αποτελεί αμερόληπτη εκτίμηση, και να έχει το ελάχιστο μέσο τετραγωνικό σφάλμα, δηλαδή η ποσότητα $e^2 = \left(x_t - \hat{x}_t\right)^2$ να είναι ελάχιστη, ώστε να έχουμε τη μικρότερη διασπορά (Kitanidis σσ. 20.14-20.15).

Με βάση τους δύο παραπάνω περιορισμούς, (δηλαδή την απαίτηση για αμεροληψία και ελάχιστη διασπορά) δημιουργείται ένα σύστημα γραμμικών εξισώσεων, που η επίλυσή του μας δίνει της τιμές $w_i$.

Έτσι, προκειμένου να υπολογίσουμε τους συντελεστές βαρύτητας $w_i$, εφαρμόζουμε τη μέθοδο BLUE (Best Linear Unbiased Estimator), δηλαδή προσπαθούμε να βρούμε εκείνη τη λύση, που θα είναι βέλτιστη (Best), δηλαδή θα έχει το μικρότερο τετραγωνικό σφάλμα, ενώ παράλληλα θα είναι και αμερόληπτη (Unbiased), δηλαδή θα έχει μηδενική μέση τιμή. Η μεθοδολογία BLUE, περιλαμβάνει ένα σύνολο μεθόδων γραμμικής βέλτιστης παρεμβολής στις οποίες η εκτίμηση της κάθε τιμής είναι αμερόληπτη και το σφάλμα της εκτίμησης είναι το ελάχιστο. Η πιο διαδεδομένη από τις μεθόδους BLUE είναι η μέθοδος kriging[24].

Ενώ όμως στη μέθοδο kriging χρησιμοποιούνται τα ημι-μεταβλητογράμματα, εμείς στη συνέχεια θα χρησιμοποιήσουμε τη συνάρτηση συνδιασποράς, η οποία εκφράζει άμεσα τη δομή αυτοσυσχέτισης μεταξύ των τιμών της χρονοσειράς που μελετάμε και είναι κατάλληλη για ανελίξεις που παρουσιάζουν το φαινόμενο Hurst.

Έστω λοιπόν $X$ η τυχαία μεταβλητή η οποία ακολουθεί κανονική κατανομή[25] με τιμές $x_1, x_2, \ldots x_n$ η οποία θεωρούμε πως έχει μέση τιμή $\mu$ και διασπορά $\sigma^2$. Οι $\mu$, $\sigma^2$ είναι άγνωστοι. Για να υπολογίσουμε λοιπόν τις τιμές των διαφόρων βαρών $w_i$, πρέπει να ικανοποιούνται οι εξής περιορισμοί (σύμφωνα με την μέθοδο BLUE):

- το μέσο τετραγωνικό σφάλμα της εκτίμησης να παίρνει την ελάχιστη τιμή
- οι λύσεις να είναι αμερόληπτες.

Μια μεροληπτική λοιπόν λύση της εξίσωσης $Y = \mathbf{w}^T\mathbf{X} + e$ , η οποία όμως δίνει ελάχιστο τετραγωνικό σφάλμα, είναι (Koutsoyiannis & Langousis, 2010 σσ. 32-33):

[24] Η μέθοδος kriging αναπτύχθηκε στις αρχές της δεκαετίας του 50 από το μηχανικό ορυχείων Krige (1951) με σκοπό την πρόγνωση της περιεκτικότητας σε μετάλλευμα μιας περιοχής εξόρυξης αξιοποιώντας μεμονωμένες μετρήσεις περιεκτικότητας σε συγκριμένα σημεία. Η περιεκτικότητα αυτή μοντελοποιείται ως μια στοχαστική συνάρτηση στις τρεις διαστάσεις, δηλαδή ως ένα τυχαίο πεδίο. Η μέθοδος έχει εφαρμογή και σε άλλα προβλήματα πρόγνωσης όπως αυτά της υδρολογίας.

[25] Η κατανομή Gauss ή κανονική κατανομή είναι η πιο γνωστή και ευρέως χρησιμοποιούμενη κατανομή στην στατιστική ανάλυση δεδομένων. Οι λόγοι που τεκμηριώνουν τη χρήση της κατανομής αυτής είναι: το κεντρικό οριακό θεώρημα το οποίο λέει ότι το άθροισμα ανεξάρτητων και όμοια κατανεμημένων τυχαίων μεταβλητών τείνει στη κανονική κατανομή όσο το άθροισμα των προσθετέων τείνει στο άπειρο και η αρχή της μεγίστης εντροπίας η οποία δηλώνει ότι απ' όλες τις γνωστές κατανομές με γνωστή μέση τιμή και διασπορά, η κανονική κατανομή είναι αυτή που μεγιστοποιεί την εντροπία.

$$\begin{aligned} \mathbf{w} &= \mathbf{C}^{-1}\boldsymbol{\eta} \\ \mu_e &= \mu_y - \mathbf{w}^T\boldsymbol{\mu}_X \\ \sigma_e{}^2 &= \sigma_y{}^2 - \boldsymbol{\eta}^T\mathbf{C}^{-1}\boldsymbol{\eta} = \sigma_y{}^2 - \mathbf{w}^T\boldsymbol{\eta} \end{aligned} \tag{4.28}$$

όπου

$\boldsymbol{\eta} := \mathrm{Cov}[X, Y]$, είναι ένας πίνακας τα στοιχεία του οποίου είναι οι διάφορες τιμές της συνδιασποράς της τ.μ. $X$ με τη τ.μ. $Y$

$\mathbf{C} := \mathrm{Cov}[X, X]$, είναι ο θετικά ορισμένος, συμμετρικός πίνακας, που τα στοιχεία του είναι οι τιμές της αυτοσυνδιασποράς της τ.μ. $X$.

Προκειμένου να κάνουμε την παραπάνω λύση αμερόληπτη, μπορούμε, είτε να προσθέσουμε μια βοηθητική μεταβλητή $X_{n+1}$, της οποίας η τιμή θα είναι συνεχώς ίση με 1, είτε να προσθέσουμε ένα περιορισμό στην επίλυσης του συστήματος.

Ο περιορισμός που πρέπει να προσθέσουμε ώστε να εξασφαλίσουμε την αμεροληψία είναι

$$\mu_Y = \mathbf{w}^T\boldsymbol{\mu}_X \tag{4.29}$$

Με τον παραπάνω περιορισμό, το μέσο τετραγωνικό σφάλμα της εκτίμησης υπολογίζεται από τη σχέση

$$\mathrm{MSE} = \sigma_e^2 = \sigma_Y^2 + \mu_Y^2 + \mathbf{w}^T\left(\mathbf{C} + \boldsymbol{\mu}_X\boldsymbol{\mu}_X^T\right)\mathbf{w} - 2\mathbf{w}^T(\boldsymbol{\eta} + \mu_Y\boldsymbol{\mu}_X) \tag{4.30}$$

Η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος γίνεται με τη χρήση πολλαπλασιαστών Lagrange και οδηγεί στο παρακάτω σύστημα εξισώσεων:

$$\begin{cases} \mathbf{C}\mathbf{w} + \boldsymbol{\mu}_X\lambda = \boldsymbol{\eta} \\ \\ \boldsymbol{\mu}_X^T\mathbf{w} = \mu_Y \end{cases} \tag{4.31}$$

Η λύση του παραπάνω συστήματος, για τους $n$+1 αγνώστους, δηλαδή, τα βάρη, $w_1,\ldots,w_n$, και τον πολλαπλασιαστή Lagrange $\lambda$ είναι

$$\mathbf{w}' = \mathbf{C}'^{-1}\boldsymbol{\eta}' \tag{4.32}$$

όπου

$$\mathbf{w}' := \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix}, \quad \mathbf{C}' := \begin{bmatrix} \mathbf{C} & \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_X^T & 0 \end{bmatrix}, \quad \boldsymbol{\eta}' := \begin{bmatrix} \boldsymbol{\eta} \\ \mu_Y \end{bmatrix} \tag{4.33}$$

Και το ελάχιστο μέσο τετραγωνικό σφάλμα (mMSE) υπολογίζεται ως:

$$\mathrm{mMSE} = \sigma_e^2 = \sigma_Y^2 + \mathbf{w}^T\mathbf{C}\mathbf{w} - 2\mathbf{w}^T\boldsymbol{\eta} \tag{4.34}$$

Με βάση όμως τις παραδοχές που κάναμε στην αρχή του κεφαλαίου (βλέπε υποκεφάλαιο 4.2.1), οι στοχαστικές ανελίξεις που εξετάζουμε είναι στάσιμες και εργοδικές, έτσι ισχύουν οι πιο κάτω απλοποιήσεις:

$$\begin{aligned} \mu_{X_i} &= \mu_Y = \mu \\ \sigma_{X_i} &= \sigma_Y = \sigma \\ \sigma_{ij} &:= \mathrm{Cov}\left[X_i, X_j\right] \end{aligned} \tag{4.35}$$

Σε αυτή λοιπόν τη περίπτωση οι σχέσεις (4.33), απλοποιούνται και γίνονται

$$\mathbf{w}' := \begin{bmatrix} \mathbf{w} \\ \lambda' \end{bmatrix}, \quad \mathbf{C}' := \begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}, \quad \boldsymbol{\eta}' := \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{1} \end{bmatrix} \tag{4.36}$$

όπου

$\lambda' = \lambda\ \mu$ και

$\mathbf{1}$ είναι ένας πίνακας όπου όλα του τα στοιχεία ισούνται με 1.

Οι πιο πάνω λύσεις είναι γνωστές ως kriging, αν και συνήθως στη μέθοδο kriging οι λύσεις αυτές είναι εκφρασμένες σε όρους ημι-μεταβλητογράμματος και όχι σε όρους συνδιασπορών όπως γίνεται εδώ. Αυτό συμβαίνει γιατί τα ημι-μεταβλητογράμματα δεν είναι κατάλληλα για διεργασίες που παρουσιάζουν δυναμική HK (Koutsoyiannis & Langousis, 2010 σ. 33).

Συνοψίζοντας λοιπόν τα παραπάνω αποτελέσματα για την επίλυση της σχέσης $Y = \mathbf{w}^T\mathbf{X} + e$ και τον υπολογισμό των βαρών $w_i$ έχουμε:

- μεροληπτικές λύσεις
  $\mathbf{w} = \mathbf{C}^{-1}\boldsymbol{\eta}$
  με
  $\boldsymbol{\eta} := \mathrm{Cov}\left[X, Y\right]$
  και $\mathbf{C} := \mathrm{Cov}\left[X, X\right]$
- αμερόληπτες λύσεις
  $\mathbf{w}' = \mathbf{C}'^{-1}\,\boldsymbol{\eta}'$
  με $\mathbf{w}' := \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix}, \quad \mathbf{C}' := \begin{bmatrix} \mathbf{C} & \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_X^T & 0 \end{bmatrix}, \quad \boldsymbol{\eta}' := \begin{bmatrix} \boldsymbol{\eta} \\ \mu_Y \end{bmatrix}$
- αμερόληπτες λύσεις με την παραδοχή στασιμότητας και εργοδικότητας
  $\mathbf{w}' = \mathbf{C}'^{-1}\,\boldsymbol{\eta}'$
  με $\mathbf{w}' := \begin{bmatrix} \mathbf{w} \\ \lambda' \end{bmatrix}, \quad \mathbf{C}' := \begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}, \quad \boldsymbol{\eta}' := \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{1} \end{bmatrix}$

Στη συνέχεια, θα εφαρμόσουμε τις παραπάνω λύσεις, σε δύο τύπους στοχαστικών ανελίξεων, τις ανελίξεις τύπου Markov και τις ανελίξεις που παρουσιάζουν δυναμική ΗΚ.

Ανάλογα λοιπόν με τη συνάρτηση συνδιασποράς, τη δομή δηλαδή αυτοσυσχέτισης της υπό μελέτη χρονοσειράς, αλλά και αναλόγως το μέγεθος της χρονοσειράς και τη θέση της ελλείπουσας τιμής εφαρμόζουμε τις πιο πάνω σχέσεις και υπολογίζουμε τους συντελεστές βαρύτητας $w_i$ με τους οποίους θα εκτιμήσουμε την ελλείπουσα κάθε φορά τιμή.

### *4.3.5.1 Χρονοσειρές που προσομοιώνονται με ανελίξεις Markov*

Όπως έχουμε ήδη αναφέρει, οι ανελίξεις τύπου Markov, παρουσιάζουν ασθενή μνήμη και ο συντελεστής αυτοσυσχέτισης μειώνεται ταχύτατα σε σχέση με την υστέρηση. Συγκεκριμένα, ισχύει η σχέση:

$$\rho_j = \left(\rho_1\right)^j$$

Με βάση λοιπόν την πιο πάνω εξίσωση και έχοντας υπόψη πως για υστέρηση $\tau$, η συνδιασπορά συνδέεται με το συντελεστή αυτοσυσχέτισης με τη σχέση

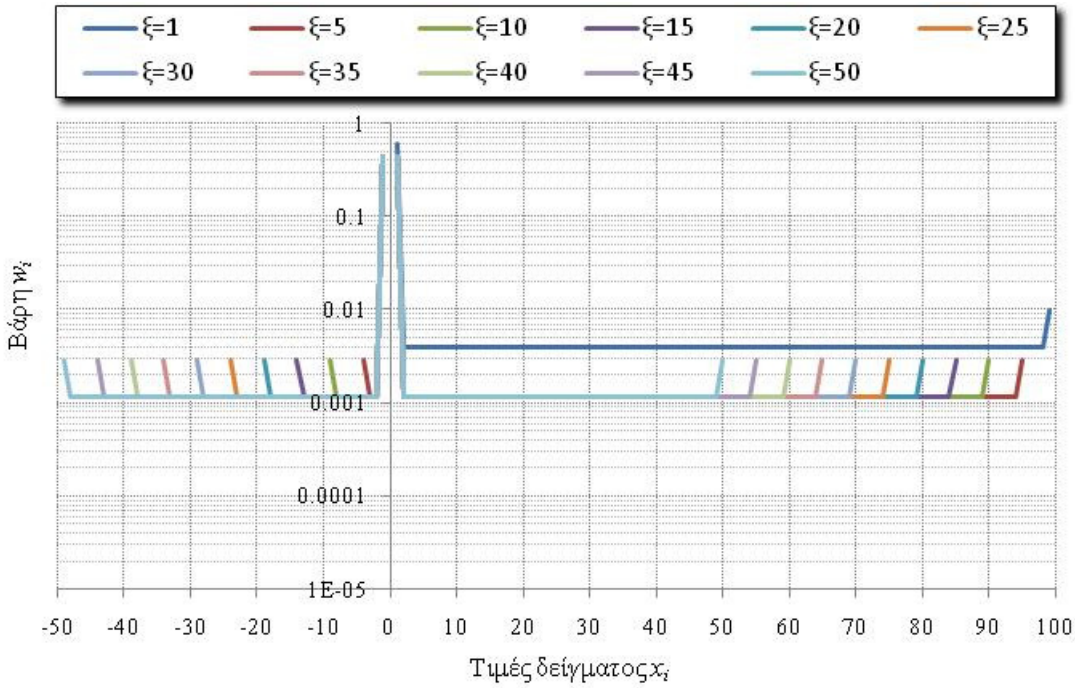
$$\mathrm{Cov}\left[X_i X_{i+\tau}\right] = \sigma^2 \rho_\tau + \mu^2$$

εφαρμόζουμε τις σχέσεις (4.28) και (4.36) για μεροληπτικές και αμερόληπτες λύσεις αντίστοιχα, και υπολογίζουμε τις διάφορες τιμές των βαρών $w_i$.

Στο ΠΑΡΑΡΤΗΜΑ D, παρουσιάζονται αναλυτικά διαγράμματα των βαρών $w_i$ που προκύπτουν για διάφορα πλήθη δείγματος (20, 40, 80 και 100 τιμές) καθώς επίσης και για διάφορες τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1, συγκεκριμένα για $\rho_1$ ίσο με [0.1, 0.3, 0.6, 0.7, 0.9].

Για οικονομία χώρου και αποφυγή επανάληψης, παρουσιάζεται σε αυτό το σημείο, το διάγραμμα των συντελεστών βαρύτητας $w_i$ για δείγμα μόνο 100 τιμών σε μια χρονοσειρά με αρκετά ισχυρή δομή αυτοσυσχέτισης ($\rho_1 = 0.6$), για διάφορες θέσεις της ελλείπουσας τιμής.

Το δείγμα λοιπόν που μελετάμε αποτελείται από 100 τιμές $x_1$, $x_2$,…, $x_{100,}$ και θεωρούμε πως λείπει μια από αυτές τις 100 τιμές. Ανάλογα με την τιμή θέλουμε να συμπληρώσουμε, θέτουμε την ελλείπουσα τιμή για εποπτικούς λόγους, ίση με $X_0$ και αντίστοιχα τις προηγούμενες $x_{-\kappa}$…, $x_{-2}$, $x_{-1}$ και τις επόμενες $x_1$, $x_2$,… $x_\lambda$. Έτσι, στο παρακάτω διάγραμμα, και σε όλα τα διαγράμματα που θα ακολουθήσουν, στη θέση 0 θα βρίσκεται πάντα η ελλείπουσα τιμή (και για αυτό το λόγο οι διάφορες καμπύλες δεν θα τέμνουν ποτέ τον άξονα των βαρών), και εκατέρωθεν οι συντελεστές βαρύτητας των γειτονικών τιμών.

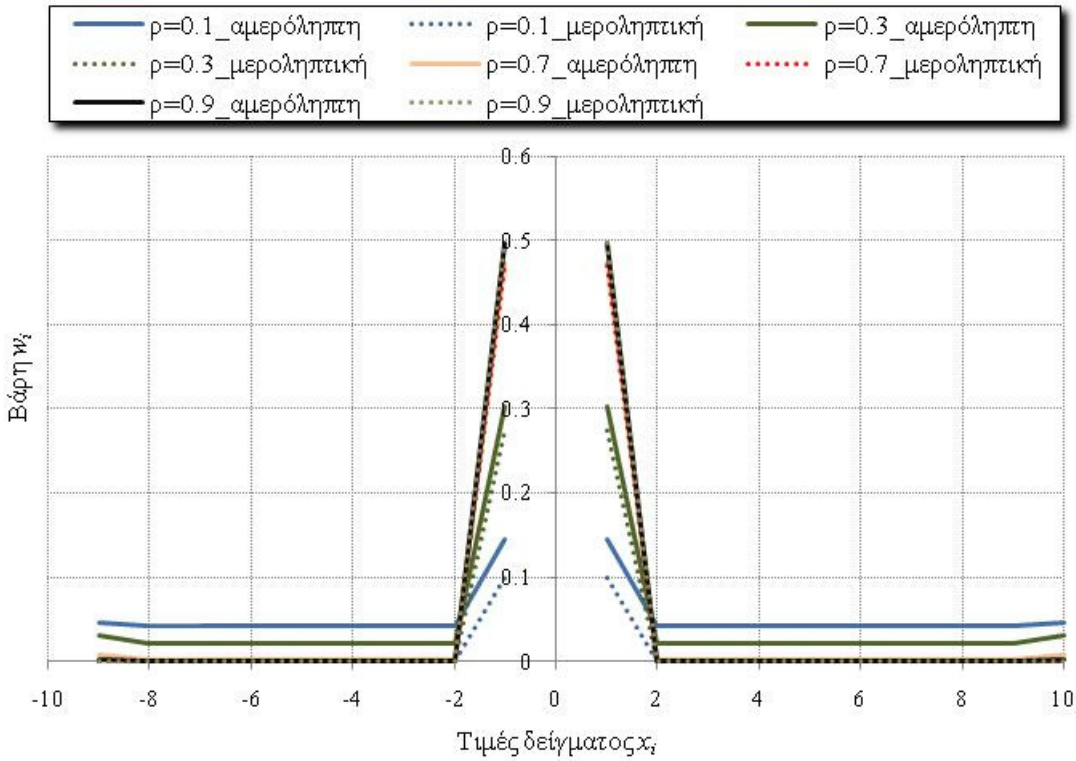



Σχήμα 4.25 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, μεγέθους 100 τιμών, ανάλογα με την θέση $\xi$ της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1 = 0.6$.

Στο πιο πάνω διάγραμμα, γίνεται φανερό πως οι γειτονικές τιμές έχουν αυξημένο συντελεστή βαρύτητας σε σχέση με τις υπόλοιπες, και αυτό οφείλετε στην αρκετά υψηλή τιμή του συντελεστή $\rho_1$, ενώ καθώς απομακρυνόμαστε από την ελλείπουσα τιμή, παρατηρούμε πως οι συντελεστές βαρύτητας σταθεροποιούνται σε μια τιμή λίγο πιο πάνω από το 0.001, ανεξαρτήτως της θέσης της ελλείπουσας τιμής.

Επίσης δεν πρέπει να μας ξαφνιάζει το γεγονός ότι στις τελευταίες τιμές, παρατηρείται αύξηση στους συντελεστές βαρύτητας (η αύξηση αυτή βέβαια δεν είναι ιδιαίτερα σημαντική αν αναλογιστεί κανείς ότι στο πιο πάνω διάγραμμα ο άξονας των βαρών $w_i$ παρουσιάζεται σε λογαριθμική κλίμακα). Η αύξησης αυτή οφείλεται στο γεγονός ότι οι τιμές $w_i$ απαιτήσαμε να είναι αμερόληπτες, έτσι θα πρέπει να έχουν άθροισμα ίσο με 1. Προκειμένου λοιπόν να ισχύει η προηγούμενη απαίτηση, οι τελευταίες τιμές των βαρών, ανάλογα με τη θέση της ελλείπουσας τιμής, είναι ελαφρώς αυξημένες.

Προκειμένου να επαληθευτεί ο πιο πάνω ισχυρισμός, παραστήσαμε γραφικά στο πιο κάτω διάγραμμα, τις μεροληπτικές (biased) τιμές των $w_i$ με διακεκομμένη γραμμή και τις αντίστοιχες αμερόληπτες (unbiased) με συνεχή γραμμή. Γίνεται φανερό, πως στις μεροληπτικές λύσεις, όπου δεν υπάρχει απαίτηση για άθροισμα των βαρών ίσο με 1 δεν αυξάνονται οι συντελεστές βαρύτητας των τελευταίων τιμών του δείγματος, σε αντίθεση με τις αμερόληπτες λύσεις που προκειμένου το άθροισμα των βαρών να είναι 1 αυξάνεται ελαφρώς ο συντελεστής βαρύτητας των τελευταίων τιμών.




Σχήμα 4.26 Σταθμισμένα βάρη για διάφορες τιμές του συντελεστή $\rho_1$, σε ένα δείγμα 20 τιμών, αν λείπει η $10^{η}$ τιμή. Παρουσιάζονται οι μεροληπτικές αλλά και οι αμερόληπτες τιμές των $w_i$.

Στο πιο πάνω διάγραμμα πρέπει να τονίσουμε επίσης τον σημαντικό ρόλο του συντελεστή αυτοσυσχέτισης στην επιλογή των βαρών. Παρατηρούμε πως όσο αυξάνει ο συντελεστής αυτοσυσχέτισης για υστέρηση 1, τόσο αυξάνει και ο συντελεστής βαρύτητας των γειτονικών στην ελλείπουσα τιμή παρατηρήσεων. Συγκεκριμένα, για $\rho_1 = 0.9$ η προηγούμενη από την ελλείπουσα τιμή παρατήρηση καθώς επίσης και η επόμενη, έχουν βαρύτητα 49% η κάθε μια, και οι υπόλοιπες 17 τιμές της χρονοσειράς ισομοιράζονται το υπόλοιπο 2%.

### *4.3.5.2 Χρονοσειρές που παρουσιάζουν δυναμική Hurst – Kolmogorov*

Στις χρονοσειρές που αναπαράγουν το φαινόμενο Hurst, εφαρμόζουμε την ίδια ακριβώς μεθοδολογία, όπως και στην περίπτωση χρονοσειρών που προσομοιώνονται από ανελίξεις Markov, με τη διαφορά ότι τώρα η δομή της αυτοσυσχέτισης είναι ισχυρότερη. Συγκεκριμένα έχουμε ήδη αναφέρει πως η σχέση που συνδέει το συντελεστή αυτοσυσχέτισης $\rho$ με την υστέρηση $\tau$, για διάφορες τιμές του συντελεστή Hurst $H$ είναι

$$\rho_j = \left(\frac{1}{2}\right)\left[(j+1)^{2H} + (j-1)^{2H}\right] - j^{2H}$$

Εφαρμόζοντας τις αμερόληπτες και τις μεροληπτικές λύσεις που δίνονται από τις σχέσεις (4.36) και (4.28) αντίστοιχα, ανάλογα με το συντελεστή Hurst και το μέγεθος

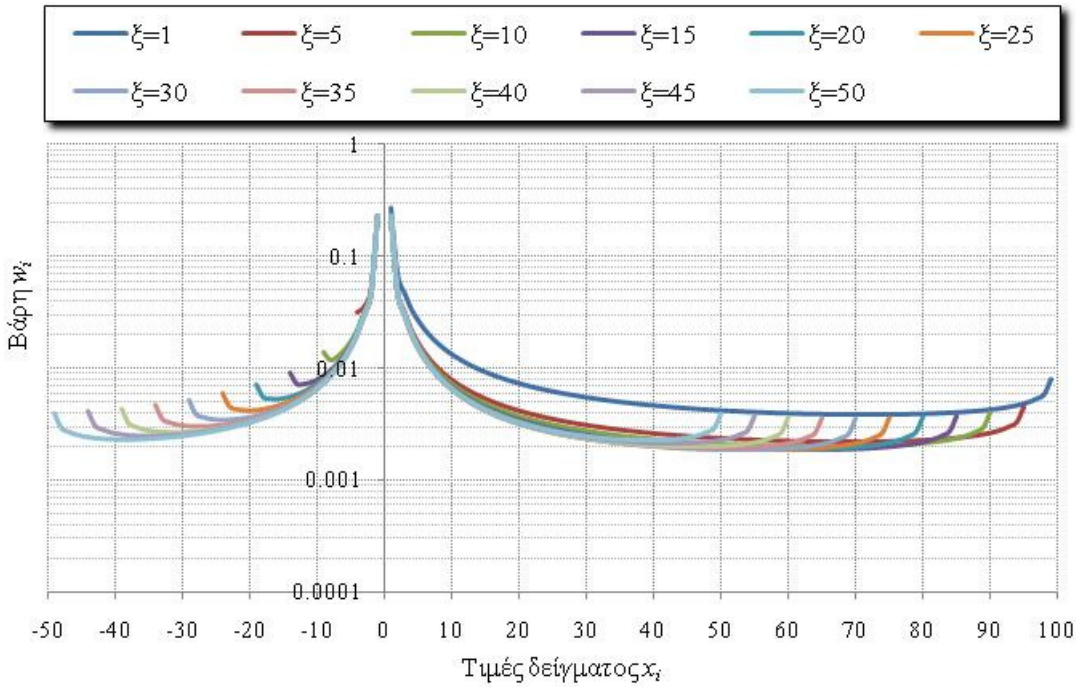
της χρονοσειράς που εξετάσουμε, προκύπτουν οι αντίστοιχες τιμές των συντελεστών βαρύτητας $w_i$ για διάφορες θέσεις της ελλείπουσας τιμής.

Και εδώ, όπως στη περίπτωση των ανελίξεων Markov δεν παρουσιάζονται όλα τα διαγράμματα για διάφορους συντελεστές Hurst και για διάφορα μεγέθη δείγματος, αλλά παρουσιάζεται μόνο το διάγραμμα των συντελεστών βαρύτητας $w_i$ για δείγμα 100 τιμών σε μια χρονοσειρά με αρκετά ισχυρή δομή αυτοσυσχέτισης ($H$ = 0.7), για διάφορες θέσεις της ελλείπουσας τιμής.

Στο ΠΑΡΑΡΤΗΜΑ Ε όμως παρουσιάζονται τα διαγράμματα των συντελεστών βαρύτητας $w_i$ για χρονοσειρές μεγέθους 20, 40, 80 και 100 τιμών και για συντελεστές Hurst $H$ ίσο με [0.7, 0.8, 0.9].

Από την πρώτη κιόλας στιγμή γίνεται φανερό πως οι χρονοσειρές που παρουσιάζουν δυναμική ΗΚ θα έχουν τελείως διαφορετική κατανομή βαρών $w_i$ σε σχέση με τις χρονοσειρές τύπου Markov.

Συγκεκριμένα, όπως φαίνεται και στο πιο κάτω διάγραμμα, οι καμπύλες των βαρών $w_i$ για διάφορες θέσεις της ελλείπουσας τιμής είναι πιο ομαλές σε σχέση με τις αντίστοιχες των ανελίξεων Markov. Αυξημένους συντελεστές βαρύτητας δεν έχουν μόνο η αμέσως προηγούμενη και η αμέσως επόμενη από την ελλείπουσα τιμή παρατηρήσεις, αλλά και οι υπόλοιπες.

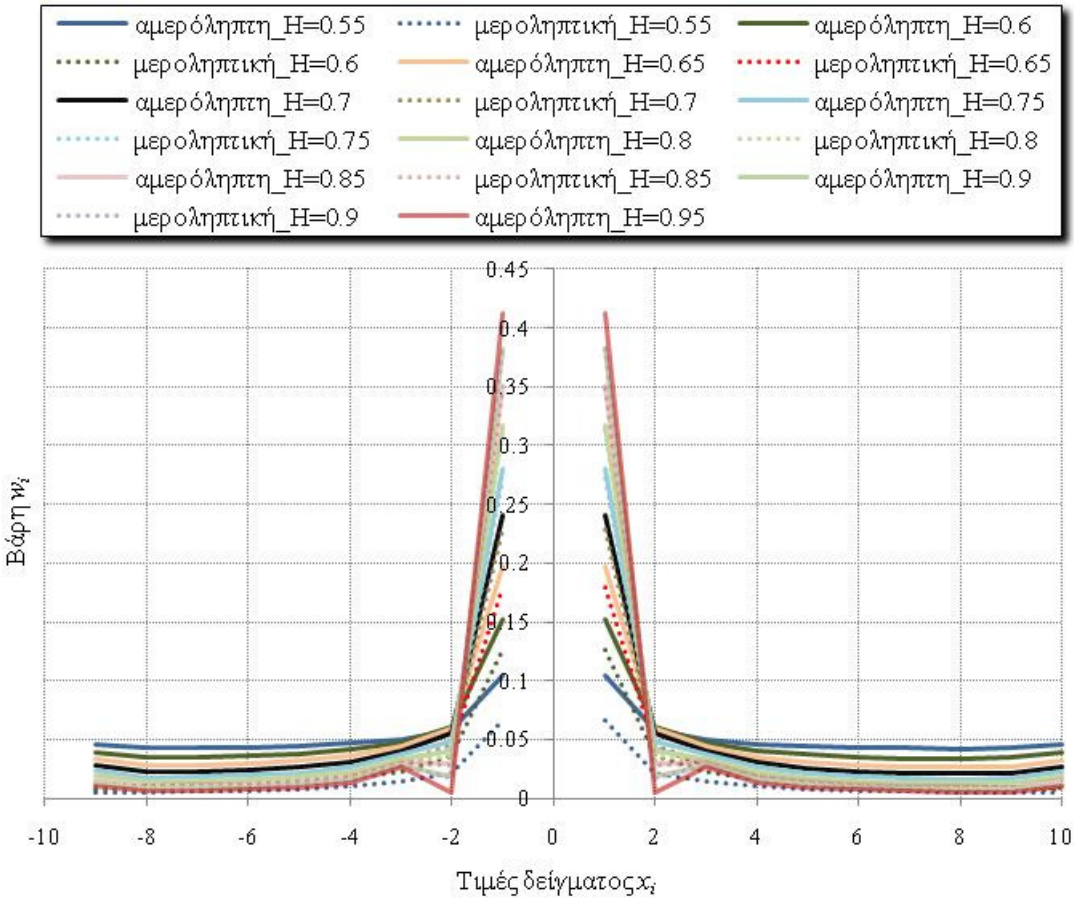



Σχήμα 4.27 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, μεγέθους 100 τιμών, ανάλογα με την θέση $\xi$ της ελλείπουσας τιμής, για συντελεστή Hurst $H$ = 0.7.

Ωστόσο, για πολύ υψηλές τιμές του συντελεστή Hurst, τα βάρη των αμέσως γειτονικών στην ελλείπουσα τιμή παρατηρήσεων αποκτούν χαρακτηριστικό προβάδισμα έναντι των υπολοίπων βαρών.

Στο επόμενο διάγραμμα γίνεται φανερό πως όσο ο συντελεστής Hurst αυξάνει, τόσο αυξάνει και ο συντελεστής βαρύτητας των αμέσως γειτονικών τιμών. Χαρακτηριστικά αναφέρεται η ακραία περίπτωση όπου ο συντελεστής Hurst είναι 0.95, οι αμέσως γειτονικές τιμές (πριν και μετά την ελλείπουσα τιμή) έχουν συντελεστή βαρύτητας 41% η κάθε μία, και οι υπόλοιπες 97 παρατηρήσεις της χρονοσειράς μοιράζονται το υπόλοιπο 18% περίπου.

Αναλυτικά, οι αμερόληπτες και οι μεροληπτικές τιμές των συντελεστών $w_i$ για διάφορες τιμές του συντελεστή Hurst, για δείγμα μεγέθους 20 τιμών όπου λείπει η $10^{η}$ τιμή, φαίνονται στο πιο κάτω σχήμα. Με συνεχή γραμμή παριστάνονται οι αμερόληπτες (unbiased) τιμές των $w_i$ ενώ με διακεκομμένη οι μεροληπτικές (biased) τιμές των $w_i$.




Σχήμα 4.28 Σταθμισμένα βάρη για διάφορες τιμές του συντελεστή Hurst ($H$), σε ένα δείγμα 20 τιμών, αν λείπει η $10^{η}$ τιμή, παρουσιάζονται οι μεροληπτικές αλλά και οι αμερόληπτες τιμές των $w_i$.

Επίσης και εδώ παρατηρείται αύξηση των συντελεστών βαρύτητας $w_i$ των τελευταίων τιμών της χρονοσειράς που λαμβάνονται υπόψη στη συμπλήρωση. Το γεγονός αυτό έχει διττή αιτία: αφενός μεν οφείλεται στην απαίτηση για άθροισμα όλων των συντελεστών βαρύτητας $w_i$ ίσο με 1, αφετέρου οφείλεται στο γεγονός ότι επειδή ο πληθυσμός δεν μπορεί να είναι ποτέ πεπερασμένος, οι τελευταίες τιμές λαμβάνονται με αυξημένους συντελεστές βαρύτητας προκειμένου να συμπεριληφθεί η πληροφορία από τις προηγούμενες τιμές που δεν υπάρχουν μετρήσεις.

#### *4.3.5.3 Σχόλια – παρατηρήσεις*

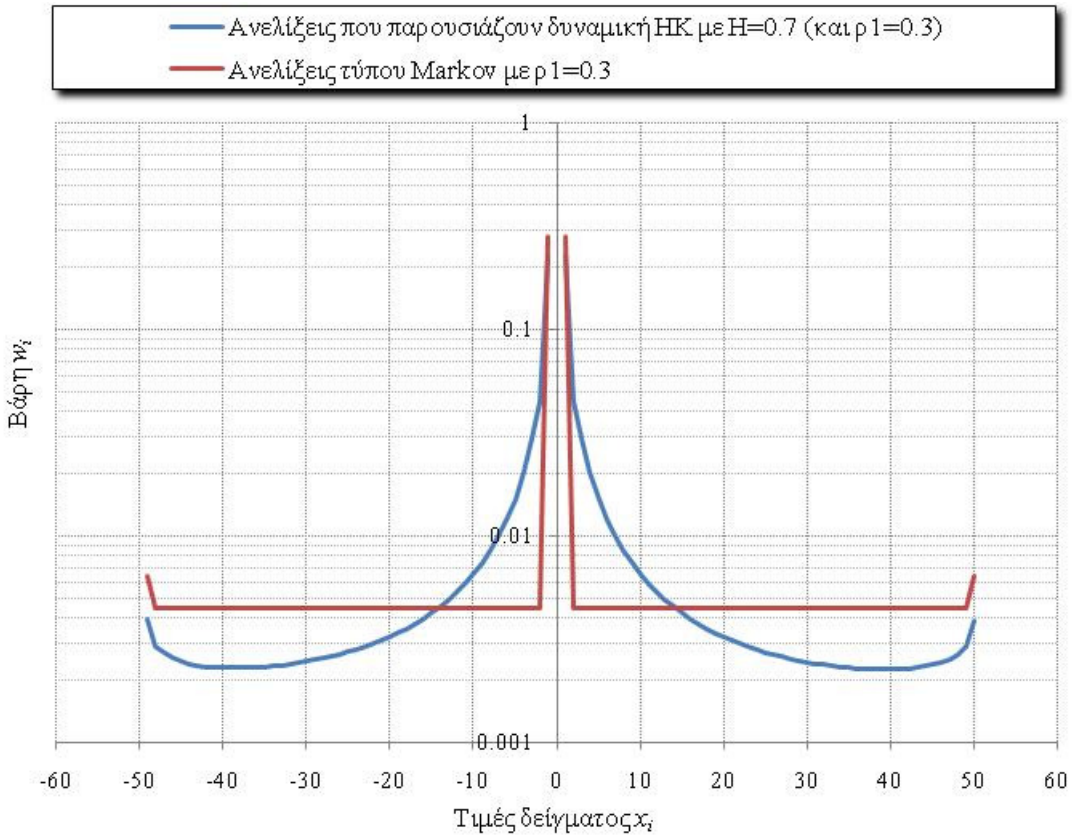
Η πιο πάνω μεθοδολογία, γίνεται φανερό πως έχει μεγάλο υπολογιστικό φόρτο, ειδικά αν δεν γίνουν οι παραδοχές στασιμότητας και εργοδικότητας των στοχαστικών ανελίξεων. Ο υπολογισμός των $w_i$ καθίσταται αδύνατος, αφού απαιτείται ο υπολογισμός[26] $\frac{(n^2+3n)}{2}$ διαφορετικών τιμών συνδιασποράς, αν θεωρήσουμε συμμετρικό τον πίνακα των συνδιασπορών (Koutsoyiannis & Langousis, 2010 σ. 33).

Ωστόσο με τις παραδοχές στασιμότητας και εργοδικότητας, οι υπολογισμοί διευκολύνονται αρκετά, και καθιστούν εφικτή υπολογιστικά τη διαδικασία υπολογισμού των συντελεστών $w_i$.

Παρατηρούμε και εδώ, που χρησιμοποιούμε σταθμισμένα βάρη, αυξημένη βαρύτητα, ανάλογα με τη δομή αυτοσυσχέτισης, έχουν οι γειτονικές τιμές. Αυτή η διαπίστωση επιβεβαιώνει τη μεθοδολογία που παρουσιάστηκε στο υποκεφάλαιο 4.3.2 και εξετάζει τη χρήση ενός τοπικού μέσου όρου στη θέση του ολικού, στη περίπτωση που η χρονοσειρά που εξετάζουμε έχει έντονη δομή αυτοσυσχέτισης.

Συγκρίνοντας τώρα τους συντελεστές $w_i$ που προκύπτουν για ανελίξεις Markov και τους αντίστοιχους που προκύπτουν για ανελίξεις που παρουσιάζουν δυναμική ΗΚ, παρατηρούμε μια εντελώς διαφορετική συμπεριφορά. Όπως φαίνεται και στο πιο κάτω διάγραμμα, οι συντελεστές $w_i$ κατανέμονται ομαλότερα στις ανελίξεις με δυναμική ΗΚ σε αντίθεση με τις ανελίξεις τύπου Markov όπου οι γειτονικές μόνο τιμές έχουν ουσιαστικά τον κύριο ρόλο στη συμπλήρωση.

[26] Όπου $n$ το πλήθος των τιμών του δείγματος.




Σχήμα 4.29 Συντελεστές βαρύτητας $w_i$ για ανελίξεις με δυναμική ΗΚ και ανελίξεις τύπου Markov με $\rho_1 = 0.3$,για δείγμα 100 τιμών, που λείπει η $50^{η}$ τιμή.

## 4.4 Σποραδικά κενά στις χρονοσειρές

### *4.4.1 Εισαγωγή*

Σύνηθες είναι το πρόβλημα της ύπαρξης συνεχόμενων κενών στις υδρομετεωρολογικές χρονοσειρές. Στο πρόβλημα αυτό όμως δεν είναι δυνατό να χρησιμοποιηθεί ένας τοπικός μέσος όρος, καθώς οι γειτονικές στην ελλείπουσα τιμή παρατηρήσεις θα λείπουν. Έτσι, δε θα μπορούμε να συσχετίσουμε την ελλείπουσα τιμή με τις γειτονικές τιμές. Η μέθοδος των σταθμισμένων βαρών είναι λοιπόν μονόδρομος στην περίπτωση πολλών συνεχόμενων ελλειπουσών τιμών.

Στη παρούσα εργασία, αναλύεται μόνο η περίπτωση της ύπαρξης τριών συνεχόμενων ελλειπουσών τιμών. Παρόμοια μεθοδολογία ακολουθείται και για περιπτώσεις με περισσότερες κενές τιμές. Πρέπει να τονιστεί όμως πως για να εφαρμοστεί η μεθοδολογία των σταθμισμένων βαρών, το πλήθος των ελλειπουσών τιμών δεν μπορεί να είναι πολύ μεγάλο σε σχέση με το μέγεθος του δείγματος.

### *4.4.2 Σταθμισμένα βάρη*

Εφαρμόζουμε λοιπόν τις λύσεις της εξίσωσης $Y = \mathbf{w}^T\mathbf{X} + e$ που είδαμε και στο υποκεφάλαιο 4.4.2, για διάφορες θέσεις ελλείπουσας τιμής. Συγκεκριμένα, έχουμε

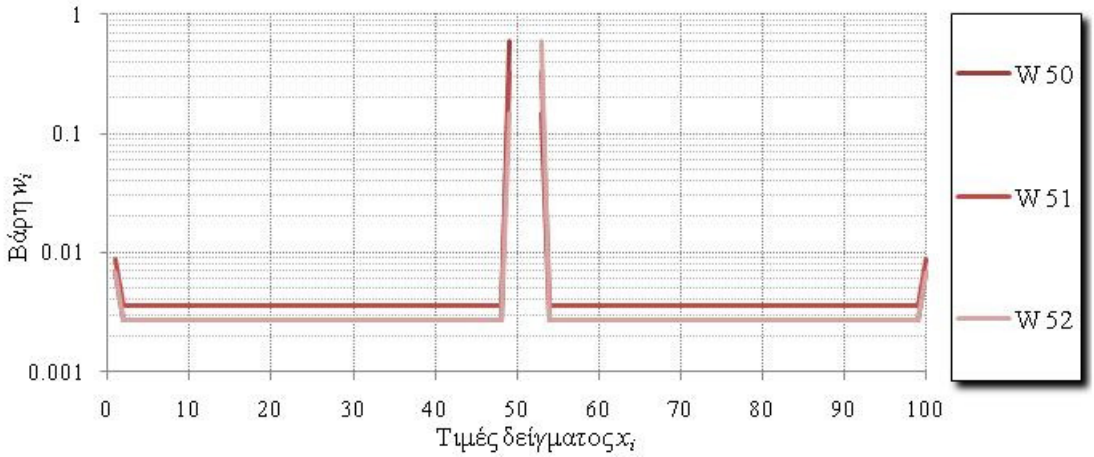
- μεροληπτικές τιμές των συντελεστών βαρύτητας $w_i$:
  $\mathbf{w} = \mathbf{C}^{-1}\boldsymbol{\eta}$
  με
  $\boldsymbol{\eta} := \mathrm{Cov}[X, Y]$
  και $\mathbf{C} := \mathrm{Cov}[X, X]$
- αμερόληπτες τιμές των συντελεστών βαρύτητας $w_i$ με τη παραδοχή στάσιμων και εργοδικών ανελίξεων:
  $\mathbf{w}' = \mathbf{C}'^{-1}\boldsymbol{\eta}'$
  με $\mathbf{w}' := \begin{bmatrix} \mathbf{w} \\ \lambda' \end{bmatrix}$, $\mathbf{C}' := \begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}$, $\boldsymbol{\eta}' := \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{1} \end{bmatrix}$

Στη συνέχεια εξετάζονται τα αποτελέσματα για δύο τύπους στοχαστικών ανελίξεων, τις ανελίξεις τύπου Markov και τις ανελίξεις που παρουσιάζουν δυναμική ΗΚ.

#### *4.4.2.1 Χρονοσειρές που προσομοιώνονται με ανελίξεις Markov*

Στο ΠΑΡΑΡΤΗΜΑ F παρουσιάζονται αναλυτικά διαγράμματα των βαρών $w_i$ που προκύπτουν για διάφορα πλήθη δείγματος (20, 40, 80 και 100 τιμές), για διάφορες θέσεις των ελλειπουσών τιμών καθώς επίσης και για διάφορες τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1, συγκεκριμένα για $\rho_1$ ίσο με [0.1, 0.3, 0.6, 0.7, 0.9].

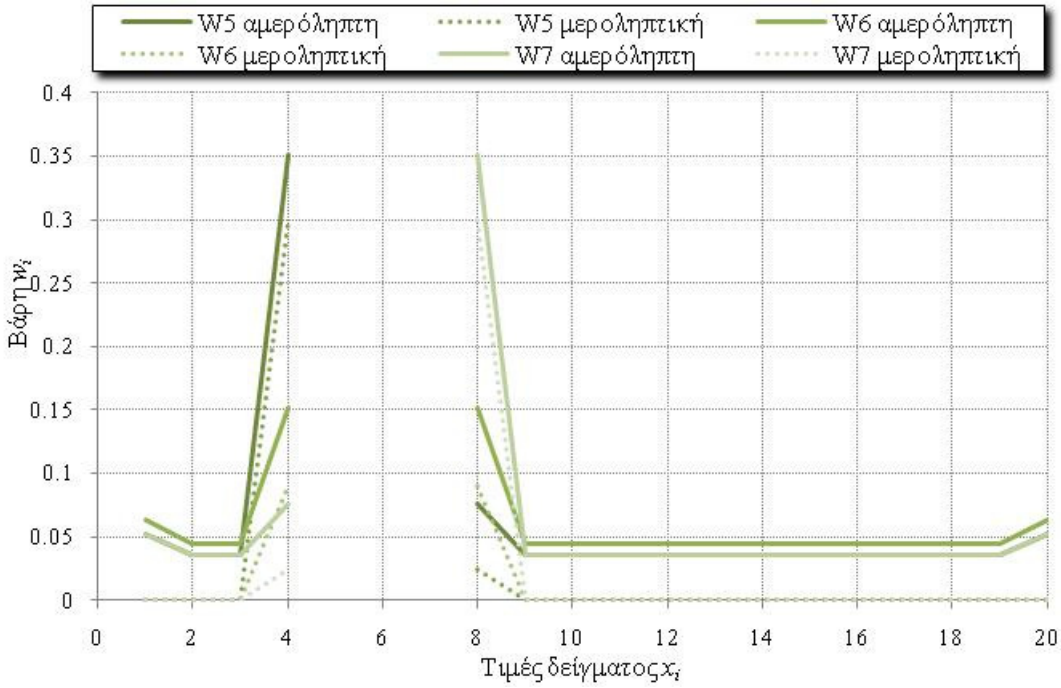
Στη συνέχεια, παραθέτουμε μόνο το διάγραμμα των συντελεστών βαρύτητας $w_i$ για δείγμα 100 τιμών σε μια χρονοσειρά με αρκετά ισχυρή δομή αυτοσυσχέτισης $(\rho_1 = 0.6)$ όπου λείπουν 3 συνεχόμενες τιμές στις θέσεις 50, 51 και 52.



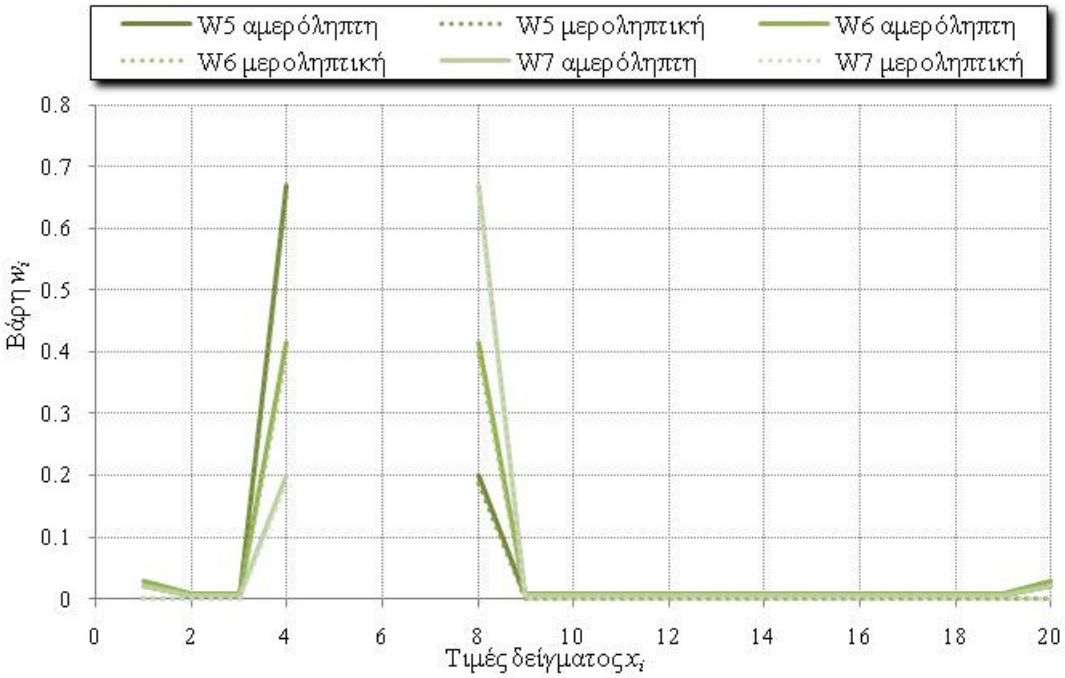

Σχήμα 4.30 Σταθμισμένα βάρη $w_i$ για δείγμα 100 τιμών με $\rho_1 = 0.6$ και κενή την $50^η$, $51^η$ και $52^η$ τιμή της χρονοσειράς.

Παρατηρούμε πως, όπως και στην περίπτωση των μεμονωμένων κενών, αυξημένη βαρύτητα έχουν οι αμέσως γειτονικές τιμές (λόγω της υψηλής τιμής του συντελεστή $\rho_1$) ενώ πάλι, στις τελευταίες τιμές της χρονοσειράς υπάρχει μία άνοδος των συντελεστών βαρύτητας που οφείλεται, όπως έχουμε ήδη τονίσει, στην απαίτηση για αμεροληψία των τιμών $w_i$.

Στη συνέχεια παρουσιάζονται οι κατανομές των συντελεστών βαρύτητας $w_i$ (αμερόληπτες και μεροληπτικές τιμές) για δείγμα μεγέθους 20 τιμών από το οποίο λείπουν 3 συνεχόμενες παρατηρήσεις στις θέσεις 5, 6 και 7 για διάφορες τιμές του συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$).




Σχήμα 4.31 Σταθμισμένα βάρη $w_i$ για δείγμα 20 τιμών με $\rho_1 = 0.3$ και κενές την $5^η$ ,$6^η$ και $7^η$ τιμή της χρονοσειράς.




Σχήμα 4.32 Σταθμισμένα βάρη $w_i$ για δείγμα 20 τιμών με $\rho_1 = 0.7$ και κενές την $5^{η}$ ,$6^{η}$ και $7^{η}$ τιμή της χρονοσειράς.

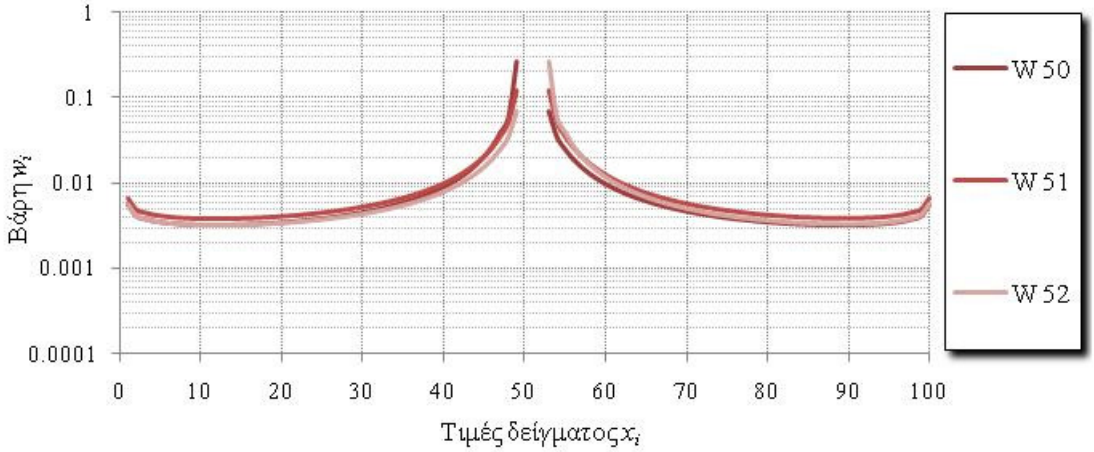
Παρατηρούμε πως καθώς αυξάνει ο συντελεστής $\rho_1$, αυξάνει και ο συντελεστής βαρύτητας των τιμών αμέσως πριν και αμέσως μετά το κενό. Συγκεκριμένα, για πολύ υψηλές τιμές του $\rho_1$ παρατηρούμε πως ουσιαστικά λαμβάνονται υπόψη μόνο οι αμέσως προηγούμενες και επόμενες τιμές και ελάχιστα (πρακτικώς καθόλου) οι υπόλοιπες τιμές της χρονοσειράς. Την ίδια διαπίστωση είχαμε κάνει και στη περίπτωση μεμονωμένων κενών στο υποκεφάλαιο 4.3.5.1.

Επίσης, και εδώ όπως και στη περίπτωση των μεμονωμένων κενών παρατηρούμε πως οι τιμές των $w_i$ αυξάνουν ελαφρώς στις τελευταίες τιμές για την περίπτωση των αμερόληπτων λύσεων. Αυτό όπως έχουμε ήδη εξηγήσει, οφείλεται στην απαίτηση για αμεροληψία. Δηλαδή αυξάνουν τα $w_i$ των τελευταίων τιμών προκειμένου το συνολικό τους άθροισμα να είναι ίσο με 1. Αυτή η αύξηση των $w_i$ για τις τελευταίες τιμές δεν συμβαίνει στη περίπτωση μεροληπτικών λύσεων, όπως γίνεται φανερό και στα πιο πάνω διαγράμματα.

#### *4.4.2.2 Χρονοσειρές που παρουσιάζουν δυναμική Hurst – Kolmogorov*

Και εδώ, όπως στη περίπτωση των ανελίξεων Markov δεν παρουσιάζονται όλα τα διαγράμματα για διάφορους συντελεστές Hurst και για διάφορα μεγέθη δείγματος, αλλά παρουσιάζεται μόνο το διάγραμμα των συντελεστών βαρύτητας $w_i$ για δείγμα 100 τιμών σε μια χρονοσειρά με αρκετά ισχυρή δομή αυτοσυσχέτισης ($H = 0.7$) όπου λείπουν 3 συνεχόμενες τιμές στις θέσεις 50, 51 και 52.

Στο ΠΑΡΑΡΤΗΜΑ Ε όμως παρουσιάζονται διαγράμματα των συντελεστών βαρύτητας $w_i$ για χρονοσειρές μεγέθους 20, 40, 80 και 100 τιμών και για συντελεστές Hurst $H$ ίσο με [0.7, 0.8, 0.9] για διάφορες θέσεις των ελλειπουσών τιμών.



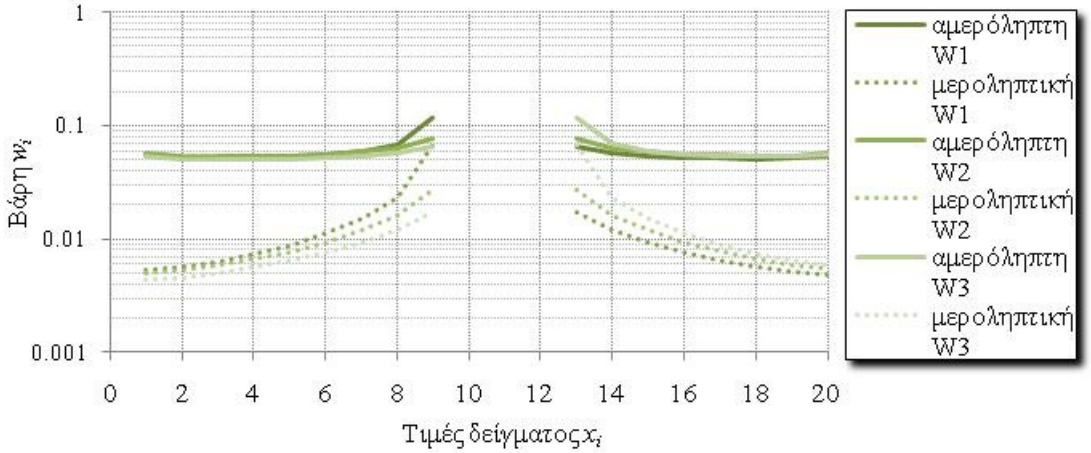

Σχήμα 4.33 Σταθμισμένα βάρη $w_i$ για δείγμα 100 τιμών με $H$ = 0.7 και κενή την 50$^η$, 51$^η$ και 52$^η$ τιμή της χρονοσειράς.

Όπως έχουμε τονίσει επανειλημμένως, οι χρονοσειρές που παρουσιάζουν δυναμική ΗΚ έχουν τελείως διαφορετική κατανομή βαρών $w_i$ σε σχέση με τις χρονοσειρές τύπου Markov και αυτό οφείλεται στο διαφορετικό χαρακτήρα αυτών των ανελίξεων, στη διαφορετική δηλαδή δομή αυτοσυσχέτισης που παρουσιάζεται μεταξύ των τιμών των ανελίξεων.

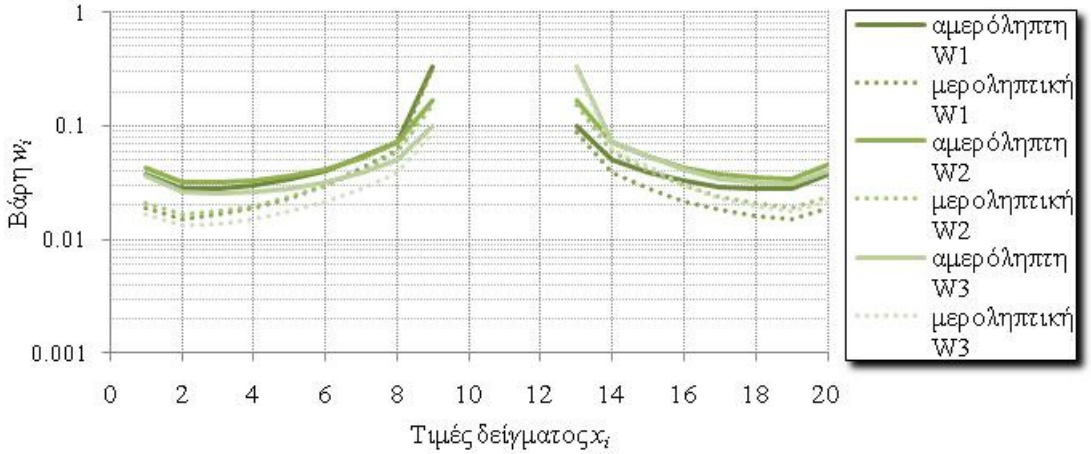
Συγκεκριμένα, όπως φαίνεται και στο πιο πάνω διάγραμμα, οι καμπύλες των βαρών $w_i$ είναι πιο ομαλές σε σχέση με τις αντίστοιχες των ανελίξεων Markov. Αυξημένους συντελεστές βαρύτητας δεν έχουν μόνο οι προηγούμενες και οι επόμενες από τις κενές τιμές παρατηρήσεις, αλλά και οι υπόλοιπες.

Στη συνέχεια παρουσιάζονται οι κατανομές των συντελεστών βαρύτητας $w_i$ (αμερόληπτες και μεροληπτικές τιμές) για δείγμα μεγέθους 20 τιμών από το οποίο λείπουν 3 συνεχόμενες παρατηρήσεις στις θέσεις 5, 6 και 7 για διάφορες τιμές του συντελεστή Hurst ($H$). Διαπιστώνουμε πως όσο ο συντελεστής Hurst αυξάνει, τόσο αυξάνει και ο συντελεστής βαρύτητας των αμέσως γειτονικών τιμών.

Αναλυτικά, οι αμερόληπτες και οι μεροληπτικές τιμές των συντελεστών $w_i$ για διάφορες τιμές του συντελεστή Hurst, για δείγμα μεγέθους 20 τιμών όπου λείπει η 10$^η$, 11$^η$ και 12$^η$ τιμή, φαίνονται στο πιο κάτω σχήμα. Με συνεχή γραμμή παριστάνονται οι αμερόληπτες (unbiased) τιμές των $w_i$ ενώ με διακεκομμένη οι μεροληπτικές (biased) τιμές των $w_i$.




Σχήμα 4.34 Κατανομή των συντελεστών $w_i$, αμερόληπτες (συνεχείς γραμμές) και μεροληπτικές (διακεκομμένες γραμμές), για δείγμα μεγέθους 20 τιμών, με κενή τη $10^{η}$, $11^{η}$ και $12^{η}$ τιμή, για $H = 0.55$.




Σχήμα 4.315 Κατανομή των συντελεστών $w_i$, αμερόληπτες (συνεχείς γραμμές) και μεροληπτικές (διακεκομμένες γραμμές), για δείγμα μεγέθους 20 τιμών, με κενή τη $10^{η}$, $11^{η}$ και $12^{η}$ τιμή, για $H = 0.75$.

Τέλος, και πάλι παρατηρείται αύξηση των συντελεστών βαρύτητας $w_i$ των τελευταίων τιμών της χρονοσειράς που λαμβάνονται υπόψη στη συμπλήρωση. Η συμπεριφορά αυτή οφείλεται αφενός στην απαίτηση για άθροισμα όλων των συντελεστών βαρύτητας $w_i$ ίσο με 1, αφετέρου στο γεγονός ότι επειδή ο πληθυσμός δεν μπορεί να είναι ποτέ πεπερασμένος, οι τελευταίες τιμές λαμβάνονται με αυξημένους συντελεστές βαρύτητας προκειμένου να συμπεριληφθεί η πληροφορία από τις προηγούμενες τιμές που δεν υπάρχουν διαθέσιμες μετρήσεις.

### *4.4.2.3 Σχόλια – παρατηρήσεις*

Η μέθοδος των σταθμισμένων βαρών στην περίπτωση διαδοχικών ελλειπουσών τιμών, παρατηρούμε πως έχει την ίδια ακριβώς συμπεριφορά όπως και στην περίπτωση μεμονωμένων κενών, όπως αυτή παρουσιάστηκε στο υποκεφάλαιο 4.3.5.

Συνεπώς τα συμπεράσματα και οι παρατηρήσεις που προκύπτουν έχουν ήδη αναλυθεί στο υποκεφάλαιο 4.3.5.3.

Το μόνο που πρέπει να προσθέσουμε είναι το γεγονός ότι η μέθοδος δεν έχει πρακτικό ενδιαφέρων στη περίπτωση που λείπουν πολλές διαδοχικές τιμές και αποτελούν σημαντικό ποσοστό σε σχέση με το μέγεθος του δείγματος. Γίνεται εύκολα αντιληπτό πως αν λείπουν πολλές τιμές, η συμπλήρωσή τους θα αλλάξει εντελώς τα στατιστικά χαρακτηριστικά της χρονοσειράς με αποτέλεσμα να αλλοιωθεί η πληροφορία που περιέχουν οι παρατηρήσεις και υπάρχει κίνδυνος για χονδροειδή σφάλματα στις περεταίρω διαδικασίες επεξεργασίας των δεδομένων.

Η συμπλήρωση μεγάλου όγκου ελλειπουσών τιμών αποτελεί ένα εντελώς διαφορετικό πρόβλημα και για το λόγω αυτό είναι επιβεβλημένη μια διαφορετική μεθοδολογία προσέγγισης που δεν εμπίπτει στο αντικείμενο της παρούσας εργασίας.

# ΚΕΦΑΛΑΙΟ 5ο

## 5 ΕΦΑΡΜΟΓΗ ΤΩΝ ΜΕΘΟΔΩΝ ΣΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

### 5.1 Εισαγωγή

Προκειμένου να επιβεβαιώσουμε τα θεωρητικά αποτελέσματα που παρουσιάστηκαν στο προηγούμενο κεφάλαιο θα εφαρμόσουμε τις μεθοδολογίες που εξετάστηκαν σε κάποιες πραγματικές υδρομετεωρολογικές παρατηρήσεις.

Συγκεκριμένα θα υπολογίσουμε το μέσο τετραγωνικό σφάλμα της εκτίμησης που προκύπτει από τη συμπλήρωση της ελλείπουσας τιμής με χρήση:

- του ολικού μέσου όρου
- της μεθοδολογίας του τοπικού μέσου όρου με βέλτιστο αριθμό γειτονικών βημάτων (βλέπε υποκεφάλαιο 4.3.2)
- του σταθμισμένου αθροίσματος δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών στη περίπτωση που η χρονοσειρά που μελετάμε αποτελείται από μηνιαίες τιμές (βλέπε υποκεφάλαιο 4.3.3)
- του σταθμισμένου αθροίσματος του τοπικού και του ολικού μέσου όρου (βλέπε υποκεφάλαιο 4.3.3)
- των σταθμισμένων βαρών – μέθοδος BLUE (βλέπε υποκεφάλαιο 4.3.3)

Πρέπει να τονίσουμε πώς για τον προσδιορισμό του σφάλματος της εκτίμησης, εφαρμόσαμε τις πιο πάνω μεθοδολογίες σε χρονοσειρές που ενώ γνωρίζαμε εκ τον προτέρων τις μετρήσεις, αφαιρούσαμε διαδοχικά μια-μια τις παρατηρήσεις και τις συμπληρώσαμε με κάθε μία από τις προηγούμενες μεθόδους. Στη συνέχεια υπολογίσαμε το τετραγωνικό σφάλμα κάθε μιας συμπλήρωσης. Τέλος υπολογίσαμε το μέσο τετραγωνικό σφάλμα που προέκυπτε μετά τη συμπλήρωση με κάθε μια από τις προηγούμενες μεθοδολογίες. Η μέθοδος που έδωσε το ελάχιστο μέσο τετραγωνικό σφάλμα είναι η ιδανική για τη συμπλήρωση της συγκεκριμένης χρονοσειράς.

Εξετάζονται στη συνέχεια τρείς διαφορετικοί τύποι υδρομετεωρολογικών χρονοσειρών :

- βροχόπτωση
- θερμοκρασία (παλαιοκλιματικά δεδομένα)
- ένταση πνοής ανέμου

## 5.2 Βροχόπτωση

Θα εφαρμόσουμε τις μεθόδους που παρουσιάστηκαν στα προηγούμενα κεφάλαια σε δεδομένα από δύο διαφορετικούς σταθμούς. Η πρώτη χρονοσειρά που πρόκειται να εξετάσουμε προέρχεται από το έναν μετεωρολογικό σταθμό στην Αυστραλία, συγκεκριμένα στην περιοχή Maatsuyker Island Lighthouse, με συντεταγμένες:-43.65B, 146.27A. Ο σταθμός βρίσκεται σε υψόμετρο 147m. Αναλυτικότερα, η γεωγραφική θέση του σταθμού φαίνεται στη πιο κάτω εικόνα.


Εικόνα 5.1 Γεωγραφική θέση μετεωρολογικού σταθμού Maatsuyker Island Lighthouse[27]

Η χρονοσειρά[28] αποτελείται από μηνιαίες μετρήσεις βροχόπτωσης και περιέχει δεδομένα 113 χρόνων, από το 1892 έως και το 2005. Έτσι θα μπορέσουμε να εφαρμόσουμε τις μεθόδους συμπλήρωσης που παρουσιάστηκαν στη μηνιαία και στην αντίστοιχη ετήσια χρονοσειρά.

Επειδή δεν υπάρχουν ημερήσιες μετρήσεις στο συγκεκριμένο σταθμό, θα εξετάσουμε ένα ακόμη μετεωρολογικό σταθμό ο οποίος αποτελείται από ημερήσιες

[27] Πηγή: www.bom.gov.au/jsp/ncc/cdio/cvg/av?p_stn_num=094041&p_prim_element_index=0&p_comp_element_index=0&period_of_avg=&normals_years=&redraw=null&p_display_type=enlarged_map

[28] Τα δεδομένα προέρχονται από τον ιστότοπο: http://climexp.knmi.nl/getprcpall.cgi?someone@somewhere+94962+MAATSUYKER_ISLAND_LIGHTHOUSE+

μετρήσεις ώστε να εφαρμόσουμε τις προτεινόμενες μεθοδολογίες συμπλήρωσης και σε ημερήσια δεδομένα.

Συγκεκριμένα μελετήθηκε η χρονοσειρά ημερήσιας βροχόπτωσης ενός σταθμού στην Ολλανδία, στη περιοχή De Bilt με συντεταγμένες 52.10B, 5.18A. Ο σταθμός βρίσκεται σε υψόμετρο 2m. Η ακριβής θέση του σταθμού φαίνεται στην ακόλουθη εικόνα.


Εικόνα 5.2 Γεωγραφική θέση μετεωρολογικού σταθμού De Bilt (Πηγή:http://www.knmi.nl/klimatologie/metadata/debilt.html)

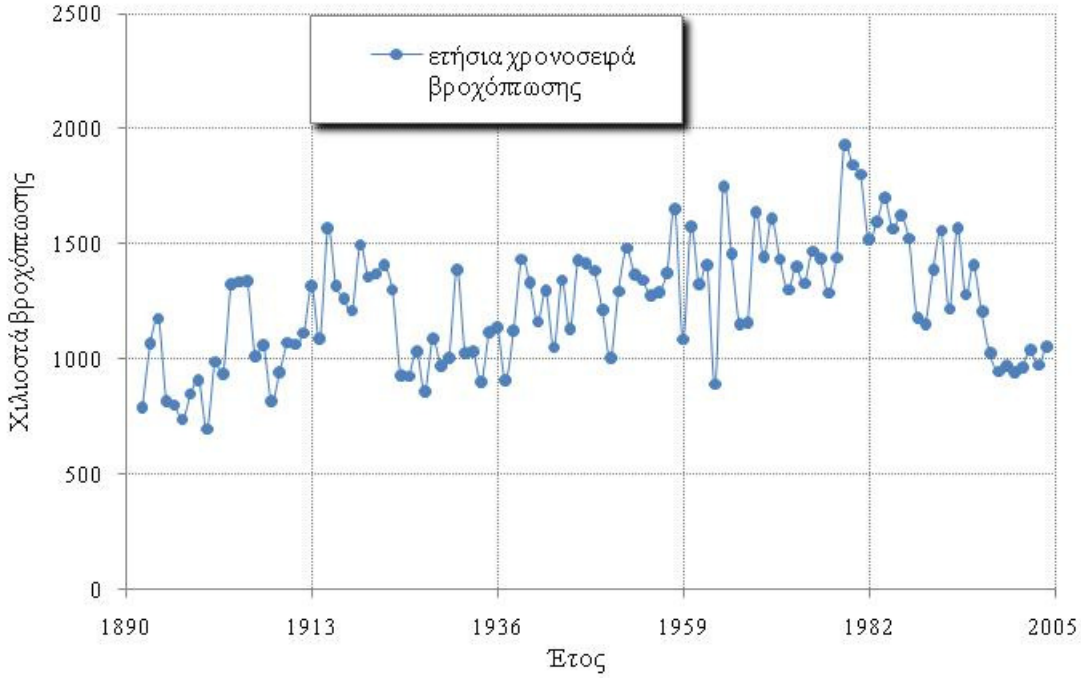
Η χρονοσειρά[29] που μελετήσαμε αποτελείται από ημερήσιες τιμές βροχόπτωσης 104 χρόνων. Συγκεκριμένα περιέχει μετρήσεις από το 1906 έως και το Σεπτέμβριο του 2010.

Στη συνέχεια παρουσιάζονται διαδοχικά η ετήσια, η μηνιαία και η ημερήσια χρονοσειρά στις οποίες έγινε συμπλήρωση, καθώς επίσης και η ρίζα του μέσου τετραγωνικού σφάλματος που προκύπτει από την εφαρμογή κάθε μιας από τις προηγούμενες μεθοδολογίες.

[29] Τα δεδομένα προέρχονται από τον ιστότοπο: http://climexp.knmi.nl/getdutchrh.cgi?someone@somewhere+260+De_Bilt+

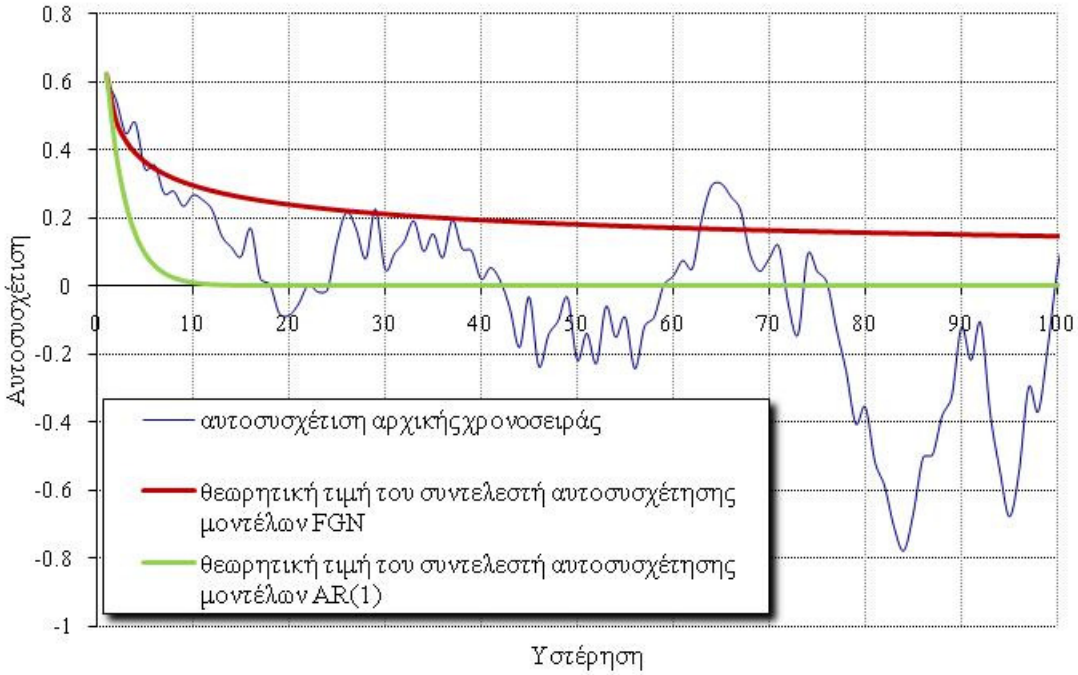
### *5.2.1 Ετήσια βροχόπτωση*

Η χρονοσειρά τις ετήσιας βροχόπτωσης προέρχεται από το σταθμό στη περιοχή Maatsuyker Island Lighthouse της Αυστραλίας και παρουσιάζεται στο ακόλουθο γράφημα.




Χρονοσειρά 5.1 Χιλιοστά ετήσιας βροχόπτωσης του σταθμού Maatsuyker Island Lighthouse από το 1892 έως το 2005.

Προκειμένου να εκτιμήσουμε αν η συγκεκριμένη χρονοσειρά προσομοιώνεται από ανελίξεις Markov ή ανελίξεις που παρουσιάζουν δυναμική Hurst-Kolmogorov, σχεδιάζουμε το διάγραμμα της αυτοσυσχέτισης σε συνάρτηση με την υστέρηση.




Σχήμα 5.1 Αυτοσυσχέτιση - υστέρηση για την πραγματική χρονοσειρά καθώς επίσης και οι θεωρητικές τιμές που προκύπτουν από την εφαρμογή μοντέλου FGN με $H = 0.85$ και μοντέλου AR(1) $\rho_1 = 0.62$

Όπως φαίνεται και στο προηγούμενο διάγραμμα το μοντέλο FGN με ένα συντελεστή $H$ ίσο με 0.85 φαίνεται να προσεγγίζει καλύτερα την πραγματική χρονοσειρά σε σχέση με το αντίστοιχο AR(1), οπου η αυτοσυσχέτιση φθίνει ταχύτατα με την υστέρηση.

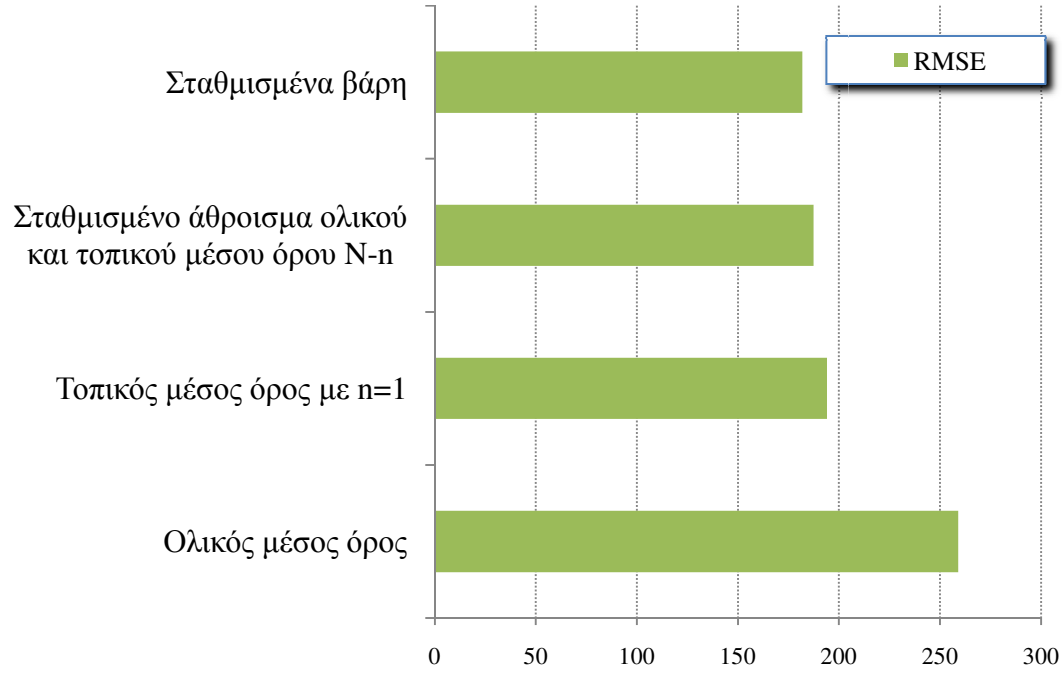
Αφού εκτιμήσαμε λοιπόν το συντελεστή Hurst της χρονοσειράς, μπορούμε να ανατρέξουμε στο κεφάλαιο 4 που παρουσιάζονται οι προτεινόμενες μεθοδολογίες και να εκτιμήσουμε τις διάφορες παραμέτρους που χρειάζονται για τη συμπλήρωση, αναλόγως τη μέθοδο που θα χρησιμοποιήσουμε.

Συγκεκριμένα, για τη περίπτωση της χρήσης του τοπικού μέσου όρου, από το υποκεφάλαιο 4.3.2 προκύπτει πως ο βέλτιστος αριθμός γειτονικών βημάτων $n$ είναι ίσος με 1. Δηλαδή χρειαζόμαστε μια τιμή πριν και μια μετά την ελλείπουσα παρατήρηση. Η μέθοδος του σταθμισμένου αθροίσματος δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών (βλέπε υποκεφάλαιο 4.3.3), δεν μπορεί να χρησιμοποιηθεί στη περίπτωση της ετήσιας χρονοσειράς, καθώς έχει εφαρμογή μόνο στη συμπλήρωση μηνιαίων τιμών. Όσον αφορά το σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου, από το υποκεφάλαιο 4.3.3, υπολογίζουμε την τιμή της παραμέτρου $\lambda$ που αντιστοιχεί στο συντελεστή $H$ της χρονοσειράς μας. Για $H = 0.85$ προκύπτει $\lambda = 0.26$.

Τέλος, για τη μέθοδο των σταθμισμένων βαρών, για συντελεστή $H = 0.85$, με βάση τη μέθοδο BLUE υπολογίζουμε τα αντίστοιχα σταθμισμένα βάρη. Συγκεκριμένα, για τις 113 τιμές της ετήσιας χρονοσειράς υπολογίσαμε 112 βάρη, τα οποία μεταβάλλονταν καθώς άλλαζε η θέση της ελλείπουσας τιμής. Επαναλάβαμε

αυτή τη διαδικασία, 113 φορές, όσες δηλαδή και οι πιθανές θέσεις τις ελλείπουσας τιμής.

Τα αποτελέσματα της συμπλήρωσης με κάθε μια από τις προηγούμενες μεθοδολογίες φαίνονται στο πιο κάτω γράφημα. Σαν μέτρο σύγκρισης των μεθοδολογιών χρησιμοποιήθηκε η ρίζα του μέσου τετραγωνικού σφάλματος.



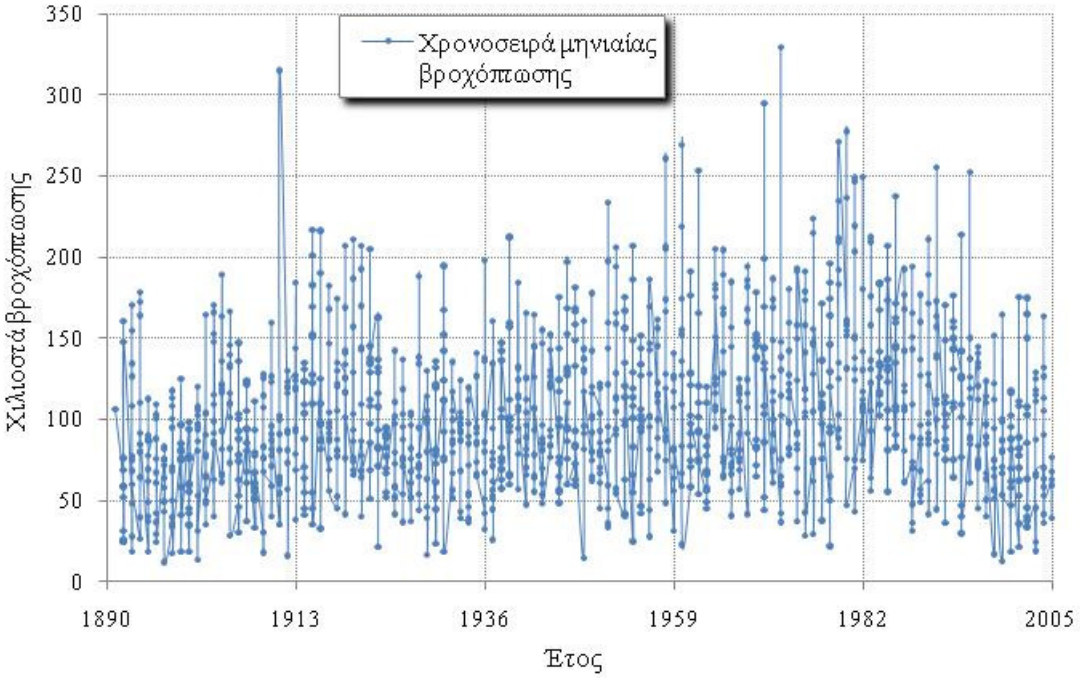

Διάγραμμα 5.1 Ρίζα του μέσου τετραγωνικού σφάλματος που προκύπτει μετά από τη συμπλήρωση με τις προτεινόμενες μεθόδους

Όπως φαίνεται και στο πιο πάνω διάγραμμα, η χρήση του ολικού μέσου όρου για τη συμπλήρωση των ελλιπών τιμών δίνει το μεγαλύτερο τετραγωνικό σφάλμα, ενώ οι υπόλοιπες μέθοδοι δίνουν παρόμοια αποτελέσματα με βέλτιστα βέβαια αυτά που προκύπτουν από τη χρήση των σταθμισμένων βαρών. Ωστόσο η συμπλήρωση με το σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου δίνει αποτελέσματα πολύ κοντά σε αυτά που προκύπτουν από τη χρήση των σταθμισμένων βαρών, ενώ παράλληλα απαιτείται ελάχιστος υπολογιστικός φόρτος.

Κατά την εφαρμογή της μεθόδου του σταθμισμένου αθροίσματος του ολικού και του τοπικού μέσου όρου, το μόνο που χρειάζεται να υπολογίσουμε είναι τον ολικό μέσο όρο, το τοπικός μέσος αποτελούμενος από μια τιμή πριν και μια μετά την κενή παρατήρηση και την παράμετρο $\lambda$. Αντίθετα, για την εφαρμογή της μεθόδου των σταθμισμένων βαρών απαιτείται να υπολογίσουμε 112 βάρη 113 φορές. Το γεγονός αυτό καθιστά τη μέθοδο του σταθμισμένου αθροίσματος τοπικού και ολικού μέσου όρου ιδανική για τη συμπλήρωση της συγκεκριμένης χρονοσειράς.

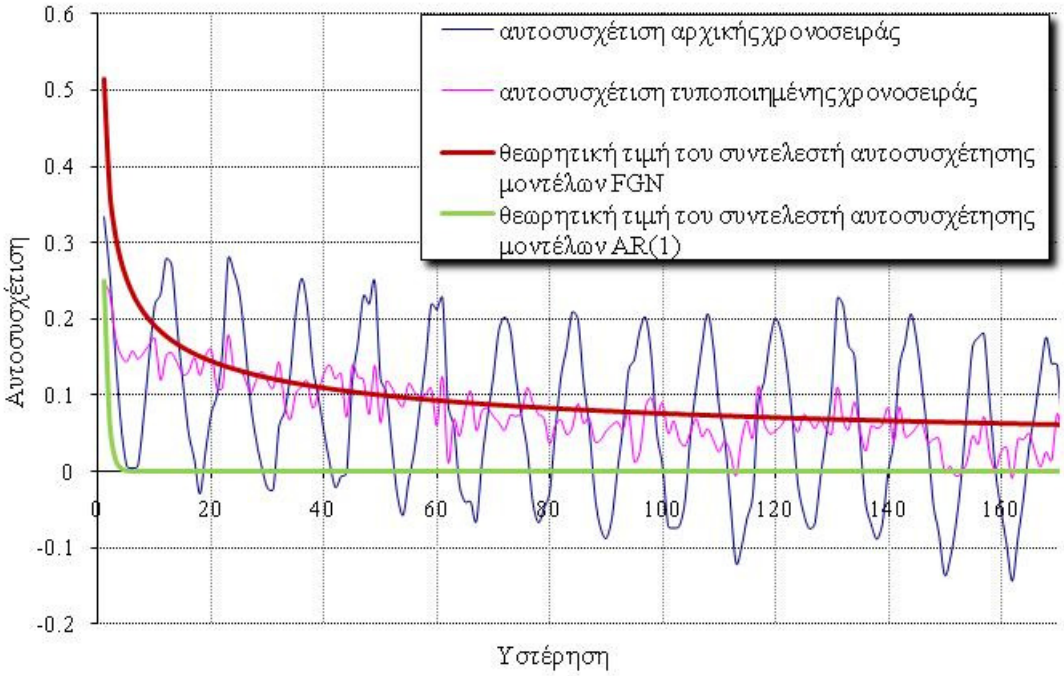
### *5.2.2 Μηνιαία βροχόπτωση*

Η χρονοσειρά τις μηνιαίας βροχόπτωσης προέρχεται από το σταθμό στην περιοχή Maatsuyker Island Lighthouse της Αυστραλίας και παρουσιάζεται στο ακόλουθο γράφημα.




Χρονοσειρά 5.2 Χιλιοστά μηνιαίας βροχόπτωσης του σταθμού Maatsuyker Island Lighthouse από το 1892 έως το 2005.

Όπως και στη περίπτωση της ετήσιας χρονοσειράς, θα εκτιμήσουμε αν η χρονοσειρά προσομοιώνεται καλύτερα από τις ανελίξεις Markov ή από τις ανελίξεις τύπου Hurst-Kolmogorov. Στο παρακάτω διάγραμμα φαίνεται η αυτοσυσχέτιση της χρονοσειράς σε σχέση με την υστέρηση, καθώς επίσης και η θεωρητική τιμή της αυτοσυσχέτισης με την υστέρηση για ανελίξεις FGN με $H$ = 0.85, και ανελίξεις AR(1) με $\rho_1$ = 0.25.




Σχήμα 5.2 Αυτοσυσχέτιση - υστέρηση για την πραγματική χρονοσειρά καθώς επίσης και οι θεωρητικές τιμές που προκύπτουν από την εφαρμογή μοντέλου FGN με $H = 0.80$ και μοντέλου AR(1) $\rho_1 = 0.25$

Παρατηρούμε πως το μοντέλο FGN με συντελεστή $H$ = 0.8 προσεγγίζει ικανοποιητικά την αυτοσυσχέτιση της πραγματικής χρονοσειράς, σε αντίθεση με το αντίστοιχο μοντέλο Markov.

Έτσι, με δεδομένο το συντελεστή Hurst της χρονοσειράς, μπορούμε να υπολογίσουμε τις τιμές των παραμέτρων που χρειαζόμαστε για την εφαρμογή των προτεινόμενων μεθοδολογιών.

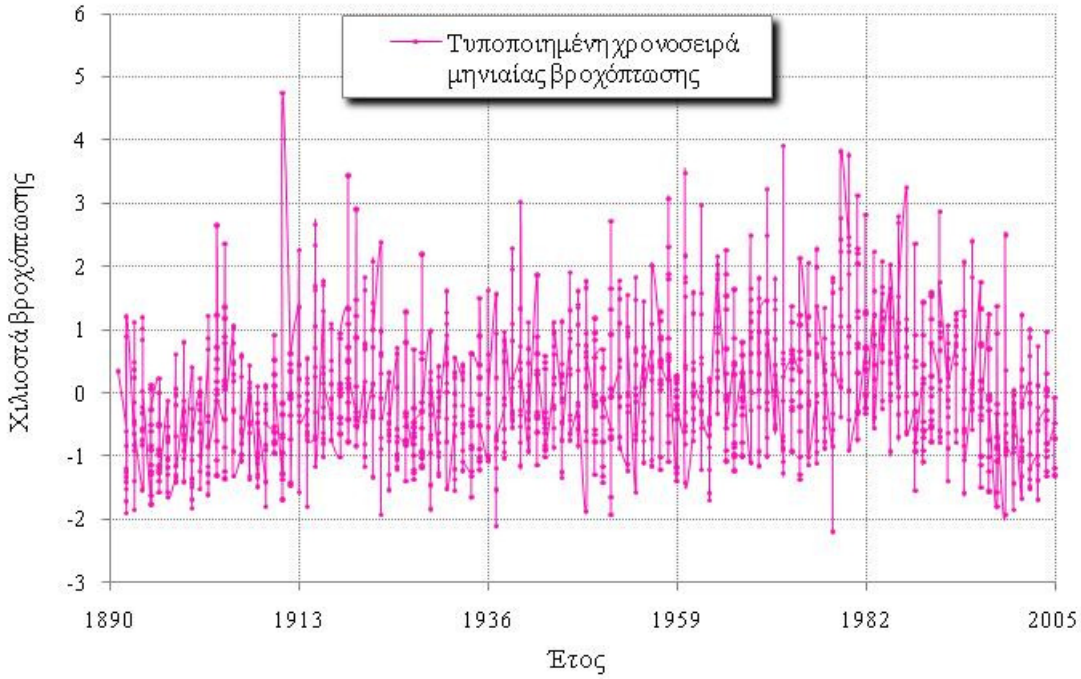
Συγκεκριμένα, για την περίπτωση της χρήσης του τοπικού μέσου όρου, από το υποκεφάλαιο 4.3.2 προκύπτει πως ο βέλτιστος αριθμός γειτονικών βημάτων $n$ είναι ίσος με 1. Για τη μέθοδο του σταθμισμένου αθροίσματος δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών (βλέπε υποκεφάλαιο 4.3.3) προκύπτει πως η τιμή της παραμέτρου $\theta$ είναι 0.81. Όσον αφορά το σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου, από το υποκεφάλαιο 4.3.3, υπολογίζουμε την τιμή της παραμέτρου $\lambda$ που αντιστοιχεί στο συντελεστή $H$ της χρονοσειράς μας. Για $H$ = 0.80 προκύπτει $\lambda$ = 0.33.

Τέλος, για τη μέθοδο των σταθμισμένων βαρών, για συντελεστή $H$ = 0.80, με βάση τη μέθοδο BLUE πρέπει να υπολογίζουμε 1355 σταθμισμένα βάρη (113 χρόνια*12μήνες-1μήνας[30]). Επιπλέον, αυτά τα 1355 βάρη πρέπει να τα υπολογίσουμε 1356 φορές, όσες δηλαδή και οι πιθανές θέσεις της ελλείπουσας τιμής. Επειδή όμως όλοι αυτοί οι υπολογισμοί των σταθμισμένων βαρών δεν συμβαδίζουν με το στόχο

[30] Αφαιρούμε ένα μήνα γιατί υποθέτουμε κάθε φορά πως αφαιρούμε μια τιμή (δηλαδή μια μηνιαία μέτρηση) και τη συμπληρώνουμε με τις προτεινόμενες μεθοδολογίες.

της παρούσας εργασίας, τη γρήγορή δηλαδή συμπλήρωση των υδρομετεωρολογικών χρονοσειρών, στη συνέχεια θα υπολογίσουμε μόνο 200 σταθμισμένα βάρη, για 100 προηγούμενες και τις 100 επόμενες στην ελλείπουσα τιμή παρατηρήσεις. Επίσης θα θεωρήσουμε πως τα βάρη αυτά είναι σταθερά και δεν επηρεάζονται από τη θέση της ελλείπουσας τιμής.

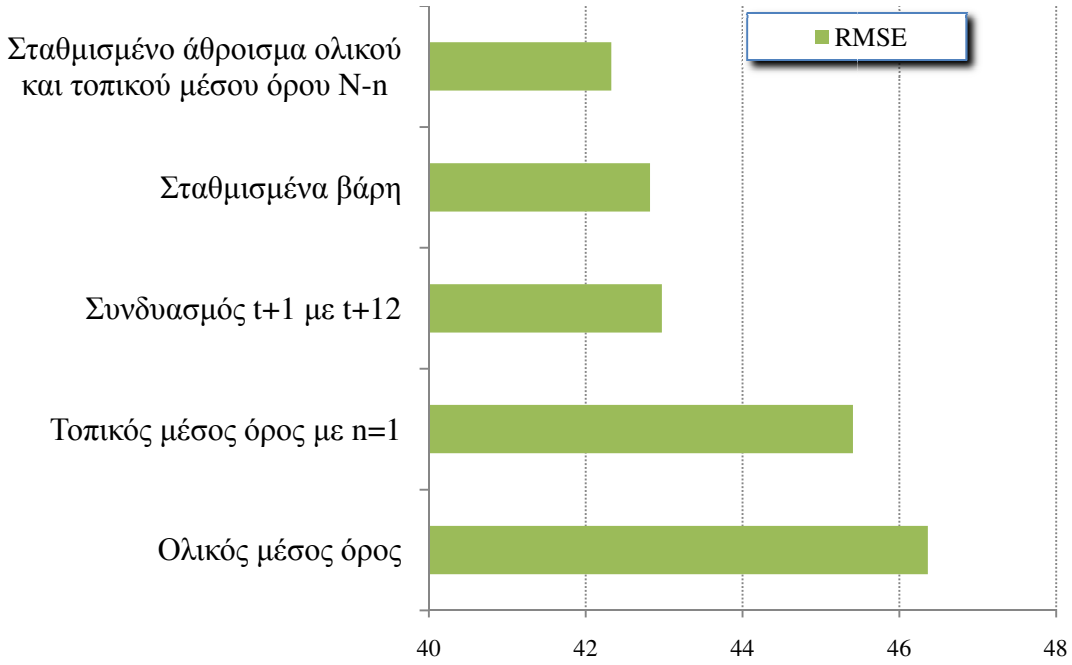
Όσον αφορά τη μεθόδο του σταθμισμένου αθροίσματος δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών, πρέπει πρώτα να τυποποιήσουμε τη χρονοσειρά των παρατηρήσεων, ώστε να απαλείψουμε την περιοδικότητα που παρουσιάζεται στις μηνιαίες παρατηρήσεις. Η τυποποιημένη χρονοσειρά παρουσιάζεται στο ακόλουθο διάγραμμα.




Χρονοσειρά 5.3 Τυποποιημένες τιμές της μηνιαίας βροχόπτωσης (σε χιλιοστά) του σταθμού Maatsuyker Island Lighthouse από το 1892 έως το 2005.

Τα αποτελέσματα της συμπλήρωσης με κάθε μια από τις προηγούμενες μεθοδολογίες και η αντίστοιχη τιμή της ρίζας του μέσου τετραγωνικού σφάλματος που προκύπτει, φαίνονται στο ακόλουθο διάγραμμα.

Στο ακόλουθο συγκριτικό γράφημα, η μέθοδος των σταθμισμένων βαρών δεν δίνει το ελάχιστο σφάλμα εκτίμησης, γιατί δεν την εφαρμόσαμε επακριβώς για λόγους ευκολίας. Όπως αναφέρθηκε προηγουμένως, υπολογίσαμε μόνο 200 βάρη, τα οποία τα θεωρήσαμε και ανεξάρτητα της θέσης της ελλείπουσας τιμής.

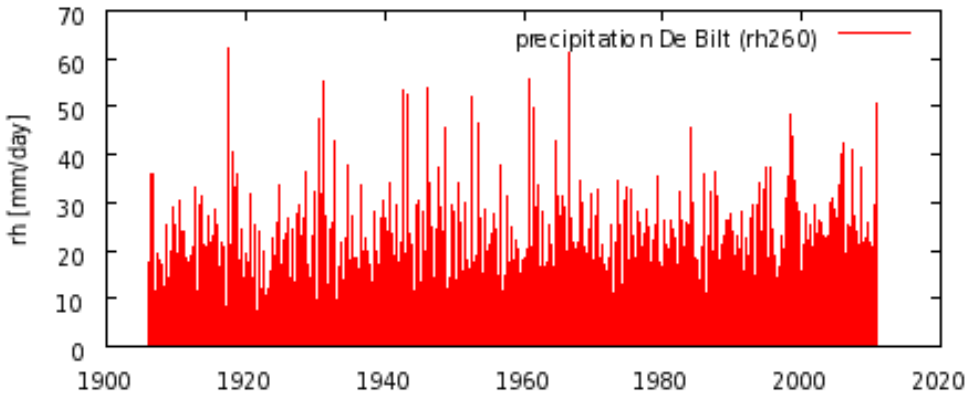



Διάγραμμα 5.2 Ρίζα του μέσου τετραγωνικού σφάλματος που προκύπτει μετά από τη συμπλήρωση με τις προτεινόμενες μεθόδους

Και στη περίπτωση της μηνιαίας χρονοσειράς, παρατηρούμε πως η χειρότερη εκτίμηση προκύπτει από τη χρήση του ολικού μέσου όρου, καθώς τότε έχουμε τη μεγαλύτερη τιμή του μέσου τετραγωνικού σφάλματος. Αντίθετα, η καλύτερη εκτίμηση δίνεται από τη χρήση του σταθμισμένου αθροίσματος του τοπικού και του ολικού μέσου όρου. Η εκτίμηση που προκύπτει από την εφαρμογή αυτής της μεθοδολογίας, για την συγκεκριμένη χρονοσειρά, είναι καλύτερη από αυτή των σταθμισμένων βαρών, γιατί όπως είπαμε υπολογίσαμε μόνο 200 βάρη. Άρα, και στη περίπτωση της μηνιαίας χρονοσειράς, η εκτίμηση της ελλείπουσας τιμής με χρήση του σταθμισμένου αθροίσματος του τοπικού και του ολικού μέσου όρου δίνει τα καλύτερα αποτελέσματα.

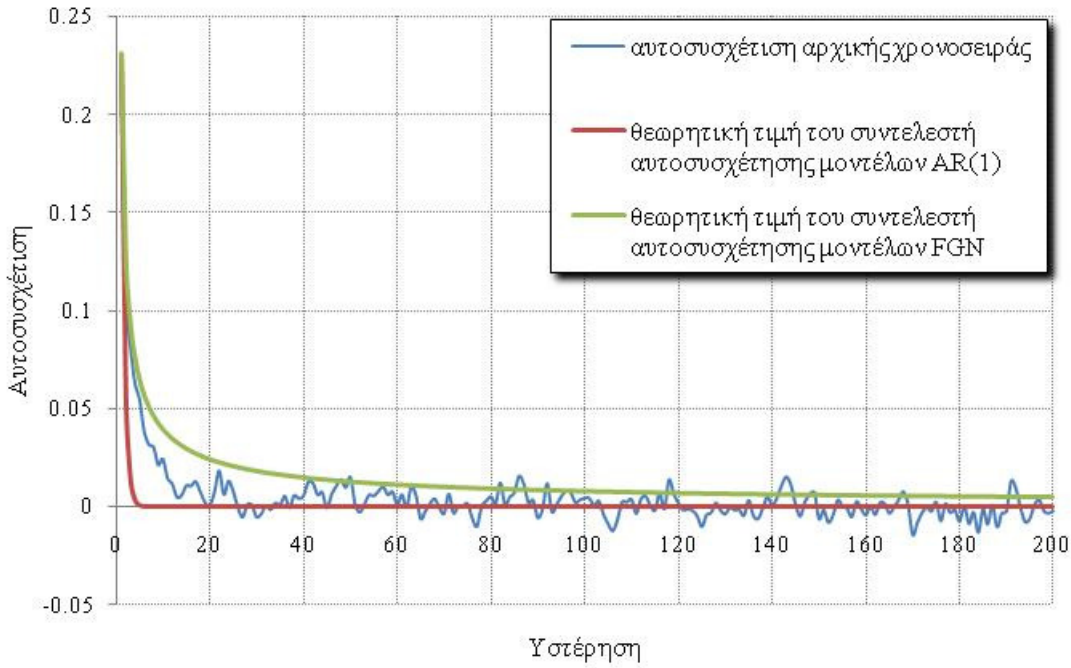
### *5.2.3 Ημερήσια βροχόπτωση*

Η χρονοσειρά τις ημερήσιας βροχόπτωσης προέρχεται από τον σταθμό στη περιοχή De Bilt της Ολλανδίας και παρουσιάζεται στο ακόλουθο γράφημα.




Χρονοσειρά 5.4 Χιλιοστά ημερήσιας βροχόπτωσης[31] του σταθμού De Bilt από το 1906 έως το 2010.

Από το πιο κάτω αυτοσυσχετόγραμμα συμπεραίνουμε πως η αυτοσυσχέτιση της ημερήσιας χρονοσειράς μειώνεται με την υστέρηση, και η μείωση αυτή προσομοιώνεται καλύτερα από τα μοντέλα μακράς μνήμης (δηλαδή μείωση τύπου δύναμης, και όχι εκθετικής μορφής, όπως στη περίπτωση των ανελίξεων Markov). Συμπεραίνουμε λοιπόν πως η ημερήσια χρονοσειρά που εξετάζουμε μπορεί να προσομοιωθεί από μια ανέλιξη FGN με συντελεστή Hurst $H = 0.65$.




Σχήμα 5.3 Αυτοσυσχέτιση - υστέρηση για την πραγματική χρονοσειρά καθώς επίσης και οι θεωρητικές τιμές που προκύπτουν από την εφαρμογή μοντέλου FGN με $H = 0.65$ και μοντέλου AR(1) $\rho_1 = 0.23$.

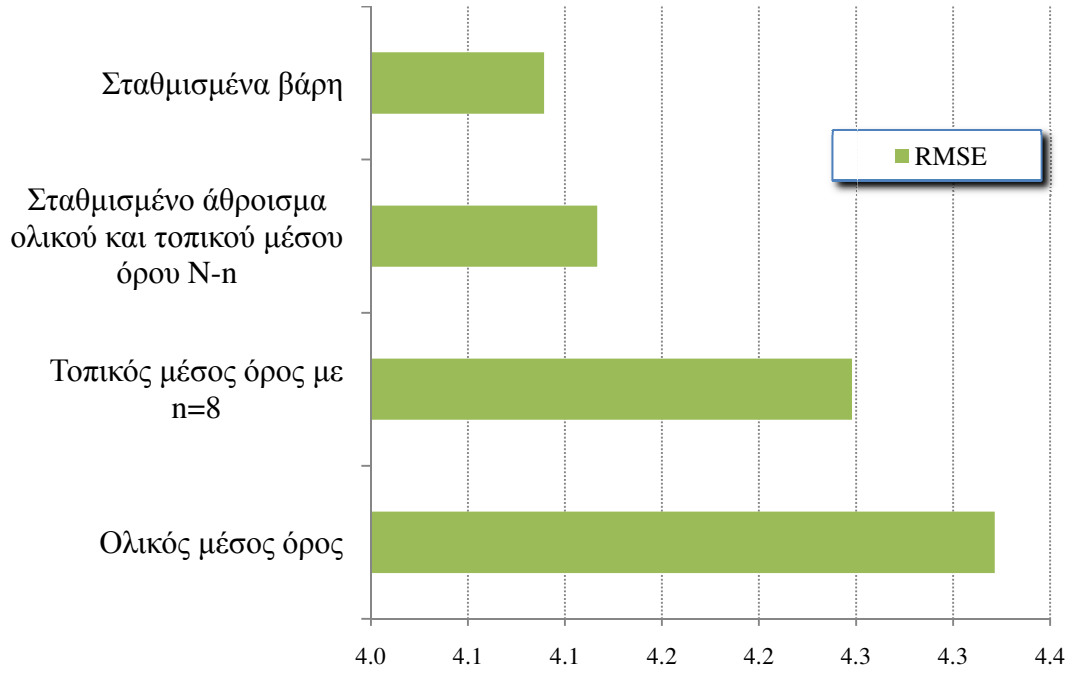
[31] Πηγή: http://climexp.knmi.nl/getdutchrh.cgi?someone@somewhere+260+De_Bilt+

Γνωρίζοντας λοιπόν το συντελεστή Hurst για την χρονοσειρά, υπολογίζουμε τις παραμέτρους για την εφαρμογή των μεθοδολογιών. Συγκεκριμένα, για την περίπτωση της χρήσης του τοπικού μέσου όρου, από το υποκεφάλαιο 4.3.2 προκύπτει πως ο βέλτιστος αριθμός γειτονικών βημάτων $n$ είναι ίσος με 8 ενώ για το σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου, από το υποκεφάλαιο 4.3.3, υπολογίζουμε την τιμή της παραμέτρου $\lambda$ που αντιστοιχεί στο συντελεστή $H$ της χρονοσειράς μας. Για $H = 0.65$ προκύπτει $\lambda = 0.63$.

Τέλος, για την μέθοδο των σταθμισμένων βαρών, για συντελεστή $H = 0.65$ υπολογίζουμε 1000 μόνο σταθμισμένα βάρη, 500 πριν και 500 μετά την ελλείπουσα τιμή, τα οποία τα θεωρούμε και ανεξάρτητα από τη θέση της ελλείπουσας τιμής.

Και στην περίπτωση της ημερήσιας χρονοσειράς, όπως και στην περίπτωση της ετήσιας που εξετάσαμε προηγουμένως, η μέθοδος του σταθμισμένου αθροίσματος δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών (βλέπε υποκεφάλαιο 4.3.3) δεν μπορεί να εφαρμοστεί, καθώς αφορά μόνο τη συμπλήρωση μηνιαίων χρονοσειρών.

Τα αποτελέσματα της συμπλήρωσης με κάθε μια από τις προηγούμενες μεθοδολογίες και η αντίστοιχη τιμή της ρίζας του μέσου τετραγωνικού σφάλματος που προκύπτει, φαίνονται στο ακόλουθο διάγραμμα.




Διάγραμμα 5.3 Ρίζα του μέσου τετραγωνικού σφάλματος που προκύπτει μετά από τη συμπλήρωση με τις προτεινόμενες μεθόδους.

Στο πιο πάνω γράφημα παρατηρούμε πως δεν υπάρχουν μεγάλες διακυμάνσεις στο μέσο τετραγωνικό σφάλμα της εκτίμησης που προκύπτει από τη συμπλήρωση με τον

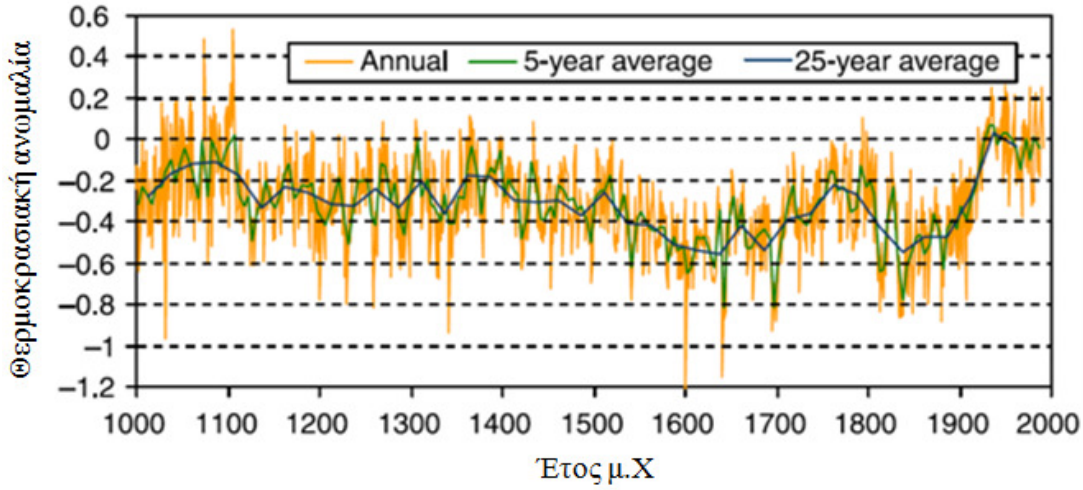
ολικό μέσο όρο και στο αντίστοιχο που προκύπτει από τον τοπικό μέσο όρο με βέλτιστο $n = 8$. Το γεγονός αυτό οφείλεται στο ότι η ημερήσια χρονοσειρά που εξετάσαμε δεν είχε υψηλό συντελεστή Hurst.

Αντίθετα το σφάλμα της μεθόδου του σταθμισμένου αθροίσματος του τοπικού και του ολικού μέσου όρου προσεγγίζει ικανοποιητικά το αντίστοιχο των σταθμισμένων βαρών.

## 5.3 Θερμοκρασία

Το δείγμα που πρόκειται να μελετήσουμε είναι τα παλαιοκλιματικά δεδομένα της χρονοσειράς Jones[32] (P.D. Jones, K.R. Briffa, T.P. Barnett and S.F.B. Tett, 1998). Η συγκεκριμένη χρονοσειρά περιέχει τις θερμοκρασιακές ανωμαλίες (σε $^{o}$C) που παρουσιάζονται στο βόρειο ημισφαίριο για 992 χρόνια, με αναφορά στη μέση τιμή των χρόνων 1961-1990. Η χρονοσειρά αυτή έχει ανακατασκευαστεί (P.D. Jones, K.R. Briffa, T.P. Barnett and S.F.B. Tett, 1998) χρησιμοποιώντας θερμοκρασιακά ευαίσθητα παλαιοκλιματικά υποκατάστατα (proxy) δεδομένα από 10 σημεία παγκοσμίως.

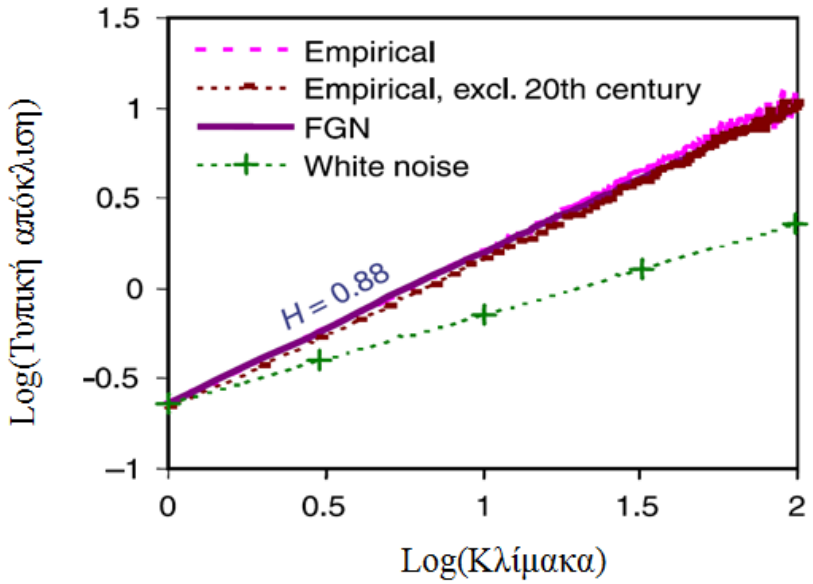
Τα υποκατάστατα (proxy) δεδομένα περιέχουν δακτυλίους δένδρων, πυρήνες πάγου, κοράλλια και ιστορικά δεδομένα. Μόνο τέσσερις από τις υποκατάστατες χρονοσειρές περιέχουν τιμές πριν το 1400 μ.Χ. και για το λόγο αυτό δεδομένα προγενέστερα των 600 χρόνων έχουν μεγαλύτερη αβεβαιότητα. Στο επόμενο διάγραμμα παρουσιάζεται η εν λόγω χρονοσειρά.




Χρονοσειρά 5.5 Παλαιοκλιματικά δεδομένα θερμοκρασιακών ανωμαλιών 992 ετών της χρονοσειράς Jones. (Πηγή: Koutsoyiannis, 2006).

[32] Η χρονοσειρά είναι διαθέσιμη στον ιστότοπο: ftp.ngdc.noaa.gov/paleo/contributions_by_author/jones1998/

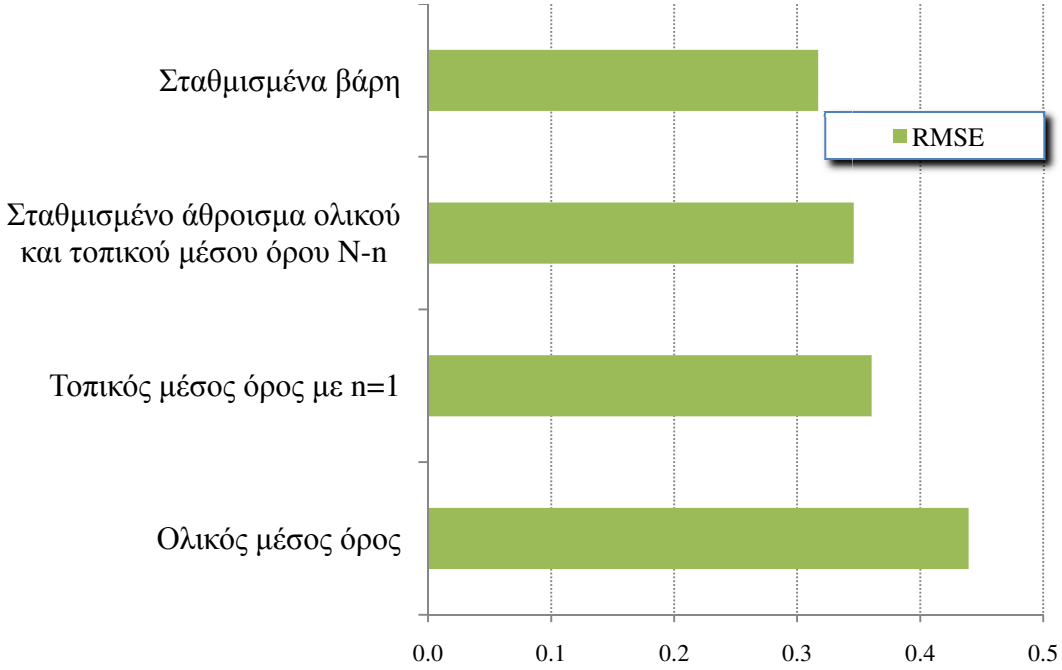
Η χρονοσειρά αυτή παρουσιάζει έντονη δομή αυτοσυσχέτισης και ο συντελεστής Hurst εκτιμάται ίσος με 0.88 (Koutsoyiannis, 2006). Αναλυτικότερα η συγκεκριμένη εκτίμηση φαίνεται στο ακόλουθο διάγραμμα.




Διάγραμμα 5.4 Διάγραμμα συναθροισμένης τυπικής απόκλισης. (Πηγή: Koutsoyiannis, 2006.)

Για συντελεστή $H = 0.88$, υπολογίζουμε τις παραμέτρους που χρειαζόμαστε για την εφαρμογή των μεθοδολογιών.. Συγκεκριμένα, για την περίπτωση της χρήσης του τοπικού μέσου όρου, ο βέλτιστος αριθμός γειτονικών βημάτων $n$ είναι ίσος με 1. Για το σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου, για $H = 0.88$ προκύπτει $\lambda = 0.22$. Για την εφαρμογή της μεθόδου των σταθμισμένων βαρών, υπολογίσαμε μόνο 200 βάρη, 100 πριν και 100 μετά την ελλείπουσα τιμή.

Τα αποτελέσματα της συμπλήρωσης με κάθε μια από τις προηγούμενες μεθοδολογίες και η αντίστοιχη τιμή της ρίζας του μέσου τετραγωνικού σφάλματος που προκύπτει, φαίνονται στο ακόλουθο διάγραμμα.




Διάγραμμα 5.5 Ρίζα του μέσου τετραγωνικού σφάλματος που προκύπτει μετά από τη συμπλήρωση με τις προτεινόμενες μεθόδους.

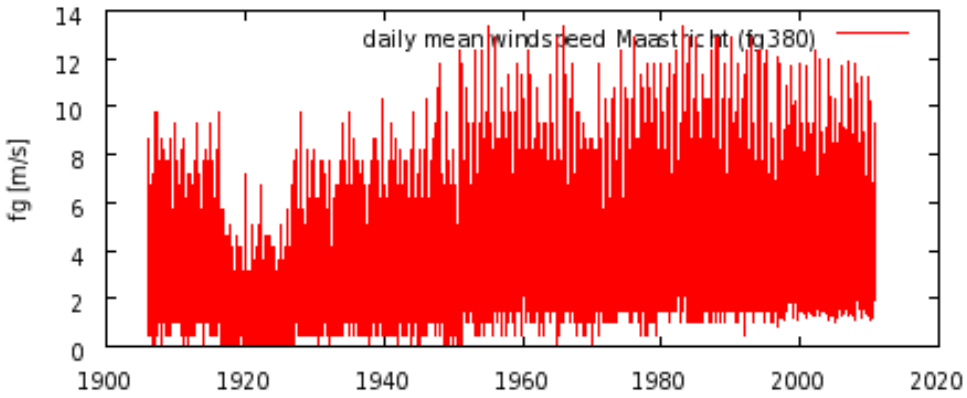
Παρατηρούμε πως πάλι ο ολικός μέσος όρος δεν δίνει ικανοποιητικά αποτελέσματα. Η βέλτιστη συμπλήρωση, όπως φαίνεται στο πιο πάνω διάγραμμα προέρχεται από τη χρήση των σταθμισμένων βαρών, ωστόσο, το σταθμισμένο άθροισμα του ολικού και του τοπικού μέσου όρου δίνει αποτελέσματα πολύ ικανοποιητικά.

## 5.4 Ένταση πνοής ανέμου

Μελετάται η χρονοσειρά ημερήσιων παρατηρήσεων της μέσης έντασης πνοής του ανέμου (μονάδα μέτρησης: m/s) στη περιοχή του Maastricht της Ολλανδίας. Ο σταθμός βρίσκεται σε υψόμετρο 114m και οι συντεταγμένες του είναι 50.92B, 5.78A. Η ακριβής γεωγραφική θέση του σταθμού φαίνεται στην επόμενη εικόνα.


Εικόνα 5.3 Γεωγραφική θέση μετεωρολογικού σταθμού στη περιοχή Maastricht (Πηγή: http://www.knmi.nl/klimatologie/metadata/maastricht.html)

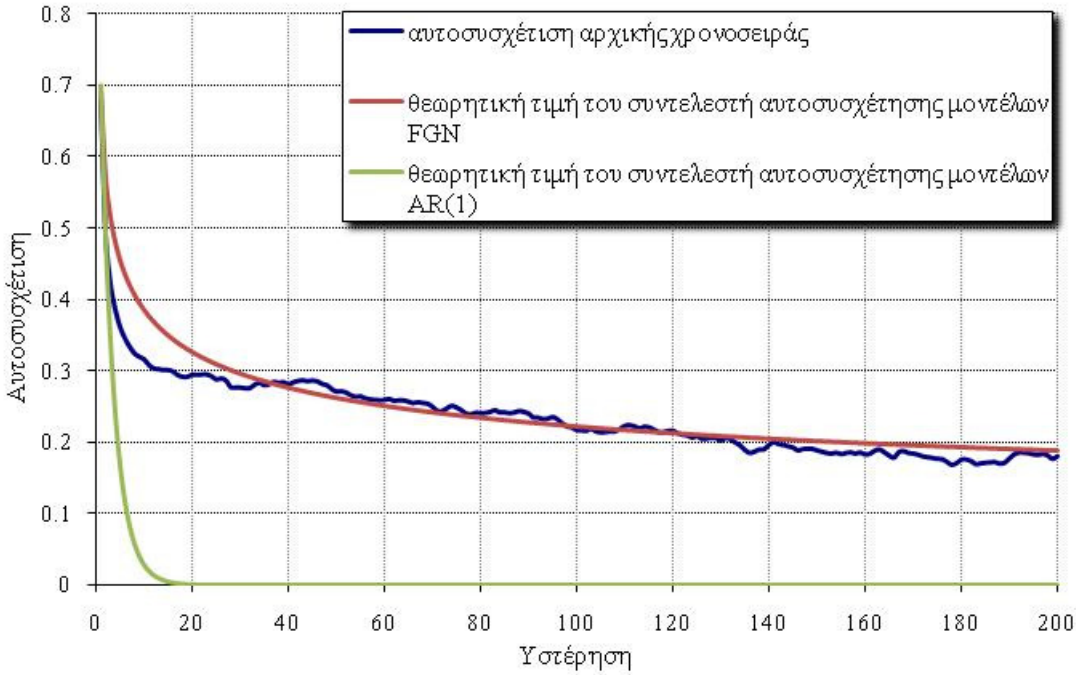
Η χρονοσειρά στην οποία εφαρμόσαμε τις διάφορες μεθοδολογίες συμπλήρωσης αποτελείται από ημερήσιες παρατηρήσεις 104 ετών. Στο επόμενο διάγραμμα φαίνεται η τιμή της έντασης της πνοής του ανέμου, για διάφορες ημέρες.




Χρονοσειρά 5.6 Ένταση πνοής ανέμου[33] (m/s), του σταθμού στο Maastricht, από το 1906 έως και το 2010.

Από το πιο κάτω αυτοσυσχετόγραμμα είναι προφανές πως η χρονοσειρά που μελετάμε παρουσιάζει έντονη αυτοσυσχέτιση.

[33] Πηγή: http://climexp.knmi.nl/getdutchfg.cgi?someone@somewhere+380+Maastricht+

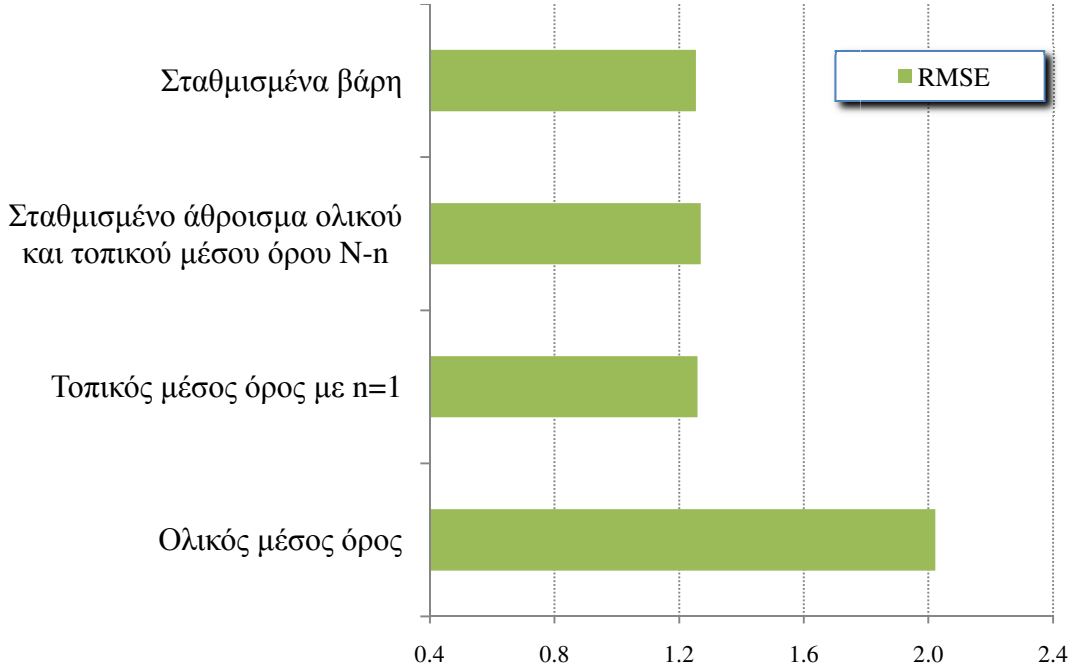



Σχήμα 5.4 Αυτοσυσχέτιση - υστέρηση για την πραγματική χρονοσειρά καθώς επίσης και οι θεωρητικές τιμές που προκύπτουν από την εφαρμογή μοντέλου FGN με $H = 0.88$ και μοντέλου AR(1) $\rho_1 = 0.70$.

Ακόμη και για μεγάλες τιμές της υστέρησης, η αυτοσυσχέτισης του δείγματος είναι αρκετά υψηλή. Έτσι η χρονοσειρά που εξετάζουμε μπορεί να προσομοιωθεί από μια ανέλιξη FGN με συντελεστή Hurst $H = 0.88$.

Υπολογίζουμε τις παραμέτρους για την εφαρμογή των μεθοδολογιών με βάση το συντελεστή Hurst που εκτιμήσαμε. Συγκεκριμένα, για την περίπτωση χρήσης του τοπικού μέσου όρου, ο βέλτιστος αριθμός γειτονικών βημάτων $n$ είναι ίσος με 1. Για το σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου, για $H = 0.88$ προκύπτει $\lambda = 0.22$. Για την εφαρμογή της μεθόδου BLUE υπολογίζουμε μόνο 1000 σταθμισμένα βάρη (500 πριν και 500 μετά την ελλείπουσα τιμή) και τα θεωρήσαμε ανεξάρτητα από τη θέση της ελλείπουσας τιμής.

Τα αποτελέσματα της συμπλήρωσης με κάθε μια από τις προηγούμενες μεθοδολογίες και η αντίστοιχη τιμή της ρίζας του μέσου τετραγωνικού σφάλματος που προκύπτει, φαίνονται στο ακόλουθο διάγραμμα.




Διάγραμμα 5.6 Ρίζα του μέσου τετραγωνικού σφάλματος που προκύπτει μετά από τη συμπλήρωση με τις προτεινόμενες μεθόδους.

Λόγω της υψηλής τιμής του συντελεστή Hurst, παρατηρούμε πως το σφάλμα της εκτίμησης που προκύπτει από τη συμπλήρωση με τον ολικό μέσο όρο, είναι αισθητά μεγαλύτερο από αυτό που δίνουν οι υπόλοιπες μέθοδοι. Επίσης, παρατηρούμε πως τα αποτελέσματα των τριών άλλων μεθοδολογιών (τοπικός μέσος όρος με βέλτιστο αριθμό γειτονικών βημάτων, σταθμισμένο άθροισμα του τοπικού και του ολικού μέσου όρου και τα σταθμισμένα βάρη -μέθοδος BLUE-) συγκλίνουν. Έχουν δηλαδή το ίδιο περίπου μέσο τετραγωνικό σφάλμα. Άρα, η χρήση του τοπικού μέσου όρου ενδείκνυνται.

# ΚΕΦΑΛΑΙΟ 6ο

## 6 ΓΕΝΙΚΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ-ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΟΛΟΓΙΩΝ

Κατά τη διερεύνηση του προβλήματος της συμπλήρωσης ελλιπών υδρομετεωρολογικών μετρήσεων εξετάζεται λεπτομερώς η συμπλήρωση μεμονωμένων κενών στις χρονοσειρές. Αναλυτικότερα, παρουσιάζονται οι υπάρχουσες μεθοδολογίες συμπλήρωσης ελλιπών υδρομετεωρολογικών τιμών και προτείνονται εναλλακτικές μέθοδοι βασισμένες στη δομή αυτοσυσχέτισης της χρονοσειράς.

Κινητήριος δύναμη και βασική αφορμή για τη διερεύνηση των μεθόδων συμπλήρωσης ελλιπών υδρομετεωρολογικών δεδομένων που παρουσιαστήκαν στα προηγούμενα κεφάλαια, ήταν η διαπίστωση ενός παραδόξου όσον αφορά τη συμπλήρωση ελλιπών υδρομετεωρολογικών τιμών. Συγκεκριμένα, διαπιστώνεται ότι ενώ όταν καλούμαστε να συμπληρώσουμε κάποιες μετρήσεις από μια χρονοσειρά, χρησιμοποιούμε τιμές από γειτονικούς μόνο σταθμούς και όχι από όλους τους διαθέσιμους, στην περίπτωση που καλούμαστε να συμπληρώσουμε τη χρονοσειρά με χρήση μόνο των υπαρχουσών τιμών της, τότε πολύ συχνά, διαισθητικά αλλά ατεκμηρίωτα, προτιμάται η χρήση του ολικού μέσου όρου.

Ανάλογα με τη δομή της αυτοσυσχέτισης της χρονοσειράς που μελετάται, εξετάζονται διάφορες παραλλαγές μεθοδολογιών συμπλήρωσης με στόχο την μείωση του σφάλματος της εκτίμησης. Τα αποτελέσματα που προκύπτουν είναι ιδιαίτερα ικανοποιητικά καθώς αποδεικνύεται θεωρητικά αλλά και αριθμητικά πως η επιλογή ενός τοπικού μέσου όρου με χρήση μόνο των γειτονικών μετρήσεων δίνει καλύτερα αποτελέσματα σε σχέση με αυτά του ολικού μέσου όρου.

Έχοντας πραγματοποιήσει την πιο πάνω ανάλυση, για δύο τύπους στοχαστικών ανελίξεων, που αποτελούν και τους πιο βασικούς στη μελέτη και προσομοίωση των υδρομετεωρολογικών φαινομένων, διερευνήσαμε πότε και κάτω από ποιες προϋποθέσεις κρίνεται σκόπιμη η χρήση ενός τοπικού μέσου όρου στη θέση του ολικού. Μελετήθηκαν και παρουσιάστηκαν επίσης διεξοδικά και κάποιες παραλλαγές, ή καλύτερα κάποιοι συνδυασμοί του τοπικού μέσου όρου ώστε να βελτιωθεί η εκτίμηση της ελλείπουσας τιμής.

Πιο συγκεκριμένα, οι μέθοδοι που παρουσιάστηκαν συνοψίζονται ως εξής:

- Ολικός μέσος όρος.
  Δηλαδή η ελλείπουσα τιμή εκτιμάται ως ο ολικός μέσος όρος των υπαρχουσών παρατηρήσεων με ίσα μεταξύ τους βάρη.
- Τοπικός μέσος όρος με $n$ τιμές πριν και $n$ μετά την ελλείπουσα τιμή.

Διερευνήθηκε ο βέλτιστος αριθμός *n* των γειτονικών τιμών πριν και μετά το κενό που πρέπει να χρησιμοποιηθούν, ώστε να προκύπτει το ελάχιστο σφάλμα εκτίμησης.

- Χρήση δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών.
  Περιορίσαμε τον τοπικό μέσο όρο σε ένα βήμα πριν και ένα μετά την ελλείπουσα τιμή και προκειμένου να βελτιώσουμε την εκτίμηση προσθέσαμε άλλον έναν όρο που περιελάμβανε δύο γειτονικές τιμές αλλά στην ετήσια κλίμακα, δηλαδή την τιμή του μήνα που θέλουμε να συμπληρώσουμε, ένα χρόνο πριν και ένα μετά.
- Συνδυασμός ολικού και τοπικού μέσου όρου.
  Προκειμένου πάλι να βελτιώσουμε την εκτίμηση της ελλείπουσας τιμής προσπαθήσαμε συνδυάζοντας με μια παράμετρο το τοπικό μέσο όρο (με μία τιμή πριν και μια μετά την ελλείπουσα μέτρηση) με τον ολικό, να ελαχιστοποιήσουμε το MSE.
- Σταθμισμένα βάρη.
  Εφαρμόζοντας μια παραλλαγή της μεθόδου kriging υπολογίσαμε τις τιμές των σταθμισμένων βαρών για τις οποίες ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα της εκτίμησης.

Σαν μέτρο αξιολόγησης όλων το μεθόδων χρησιμοποιήσαμε το μέσο τετραγωνικό σφάλμα (MSE). Δηλαδή, η μέθοδος που μας δίνει το ελάχιστο MSE είναι και η βέλτιστη αριθμητικά. Ωστόσο, στην αξιολόγηση των μεθόδων λάβαμε υπόψη και τον υπολογιστικό φόρτο γιατί εξαρχής τονίσαμε πως δεν αναζητούμε μία εξεζητημένη και πολύπλοκη μέθοδο συμπλήρωσης ελλιπών τιμών αλλά μια απλή και εύκολα εφαρμόσιμη μεθοδολογία η οποία θα επιτρέπει τη γρήγορη συμπλήρωση μεμονωμένων κενών στη χρονοσειρά, ενώ παράλληλα θα δίνει και το μικρότερο σφάλμα εκτίμησης. Έτσι η τελευταία μεθοδολογία που αναφέραμε, που βασίζεται στη χρήση σταθμισμένων βαρών, ουσιαστικά παρουσιάζεται για λόγους πληρότητας αλλά και σαν μέτρο σύγκρισης με τις άλλες μεθόδους. Αυτό γιατί ο υπολογισμός των σταθμισμένων βαρών είναι ιδιαίτερα επίπονος αν και τα αποτελέσματα που δίνει είναι τα βέλτιστα συγκριτικά με τις υπόλοιπες μεθόδους. Ωστόσο η χρήση των σταθμισμένων βαρών είναι αναγκαία στην περίπτωση πολλών συνεχών ελλιπών παρατηρήσεων.

Στη παρούσα εργασία λοιπόν εξετάζεται και αποδεικνύεται πως το πολύπλοκο και συχνό πρόβλημα της συμπλήρωσης ελλιπών υδρομετεωρολογικών μετρήσεων που παρουσιάζονται στις χρονοσειρές, μπορεί να αντιμετωπιστεί πολύ ικανοποιητικά και με ελάχιστο υπολογιστικό φόρτο στη περίπτωση μεμονωμένων κενών με τη μελέτη της δομής αυτοσυσχέτισης της υπό εξέταση χρονοσειράς. Στην περίπτωση λοιπόν ισχυρής δομής αυτοσυσχέτισης, ο τοπικός μέσος, και οι διάφορες παραλλαγές του, που αναλύθηκαν στην παρούσα εργασία, οδηγούν σε καλύτερη εκτίμηση της ελλείπουσας τιμής.

Πιο συγκεκριμένα, σε περιπτώσεις χρονοσειρών με ισχυρή δομή αυτοσυσχέτισης η χρήση ενός τοπικού μέσου όρου που αποτελείται από δύο μόνο τιμές, ήτοι μόνο την

προηγούμενη και την επόμενη στην ελλείπουσα τιμή, έχει μικρότερο σφάλμα εκτίμησης σε σχέση με άλλες πολυπλοκότερες μεθόδους. Επίσης αποδεικνύεται πως το σταθμισμένο άθροισμα δύο μεγεθών, ήτοι ενός τοπικού μέσου ορού και του αντίστοιχου ολικού έχει πολύ ικανοποιητικά αποτελέσματα που σχεδόν ταυτίζονται με αυτά της μεθόδου kriging ενώ ταυτόχρονα απαιτείται η χρήση μιας μόνο παραμέτρου και αισθητά λιγότερους υπολογισμούς, συνεπώς πολύ λιγότερο χρόνο.

Ένα πολύ σημαντικό επίσης πλεονέκτημα των μεθοδολογιών που προτείνονται είναι το γεγονός ότι δεν απαιτείται κανονικοποίηση και τυποποίηση των παρατηρήσεων. Εξαίρεση αποτελεί η μέθοδος που συνδυάζει δυο γειτονικές μηνιαίες με δυο γειτονικές ετήσιες τιμές (υποκεφάλαιο 4.3.3) όπου απαιτείται η τυποποίηση της χρονοσειράς, ώστε να λάβουμε υπόψη την περιοδικότητα.

Επιπρόσθετα, σε αντίθεση με τη μέθοδο της παλινδρόμησης που πρέπει να λαμβάνεται υπόψη η θετική ασυμμετρία που παρουσιάζεται στις υδρομετεωρολογικές χρονοσειρές (π.χ. η βροχόπτωση και η θερμοκρασία δεν παίρνουν αρνητικές τιμές και δημιουργούνται έτσι θετικά ασύμμετρες κατανομές) η προτεινόμενη μεθοδολογία δεν επηρεάζεται καθόλου από αυτή την ιδιαιτερότητα των φυσικών διεργασιών. Η ύπαρξη επίσης μηδενικών τιμών για συνεχείς περιόδους (π.χ. κατά τους θερινούς μήνες μεγάλα διαστήματα με μηδενική βροχόπτωση), σε αντίθεση με τη μέθοδο της γραμμικής παλινδρόμησης, δεν επηρεάζει ούτε περιορίζει την εφαρμογή της προτεινόμενης μεθοδολογίας.

Οι πιο πάνω παρατηρήσεις καθιστούν τη προτεινόμενη μεθοδολογία ιδανική για τη γρήγορη και άμεση συμπλήρωση μεμονωμένων κενών στις υδρομετεωρολογικές χρονοσειρές Τα προηγούμενα συμπεράσματα, επιβεβαιώνονται και για τους δύο τύπους στοχαστικών ανελίξεων που εξετάσθηκαν. Συγκεκριμένα, στη περίπτωση ανελίξεων Markov, με ασθενή δηλαδή μνήμη, αλλά και στη περίπτωση ανελίξεων που παρουσιάζουν δυναμική Hurst-Kolmogorov, δηλαδή ανελίξεων με μακρά μνήμη, τα αποτελέσματα που προκύπτουν από την χρήση ενός τοπικού μέσου όρου με διάφορες παραλλαγές, είναι πολύ καλύτερα από εκείνα που δίνει ο ολικός μέσος όρος. Πιο αναλυτικά, τα συμπεράσματα που προκύπτουν για τους δύο τύπους στοχαστικών ανελίξεων παρουσιάζονται στη συνέχεια.

## 6.1 Χρονοσειρές που προσομοιώνονται με ανελίξεις Markov

Το πιο κάτω διάγραμμα παρουσιάζει συνολικά τις μεθόδους που αναλύσαμε για την συμπλήρωση μεμονωμένων κενών στις υδρομετεωρολογικές χρονοσειρές που προσομοιώνονται από ανελίξεις τύπου Markov, και δείχνει πως μεταβάλλεται το μέσο τετραγωνικό σφάλμα της εκτίμησης (MSE) για διάφορες τιμές της αυτοσυσχέτισης της χρονοσειράς. Κάθε καμπύλη αντιστοιχεί σε μια από τις προηγούμενες μεθοδολογίες. Είναι προφανές πως η μέθοδος που αναπαριστάται με τη

καμπύλη με το μικρότερο MSE δίνει και την καλύτερη εκτίμηση της ελλείπουσας τιμής.

Μελετώντας λοιπόν το πιο κάτω διάγραμμα, μπορούμε γρήγορα να συμπεράνουμε πως η χρήση του ολικού μέσου όρου (global average) δεν είναι ιδιαίτερα ικανοποιητική ειδικά για υψηλές τιμές του συντελεστή αυτοσυσχέτισης. Συγκεκριμένα, παρατηρούμε πως η καμπύλη που αντιστοιχεί στη μέθοδο του ολικού μέσου όρου βρίσκεται πάνω από όλες τις υπόλοιπες, δηλαδή δίνει το μεγαλύτερο MSE.

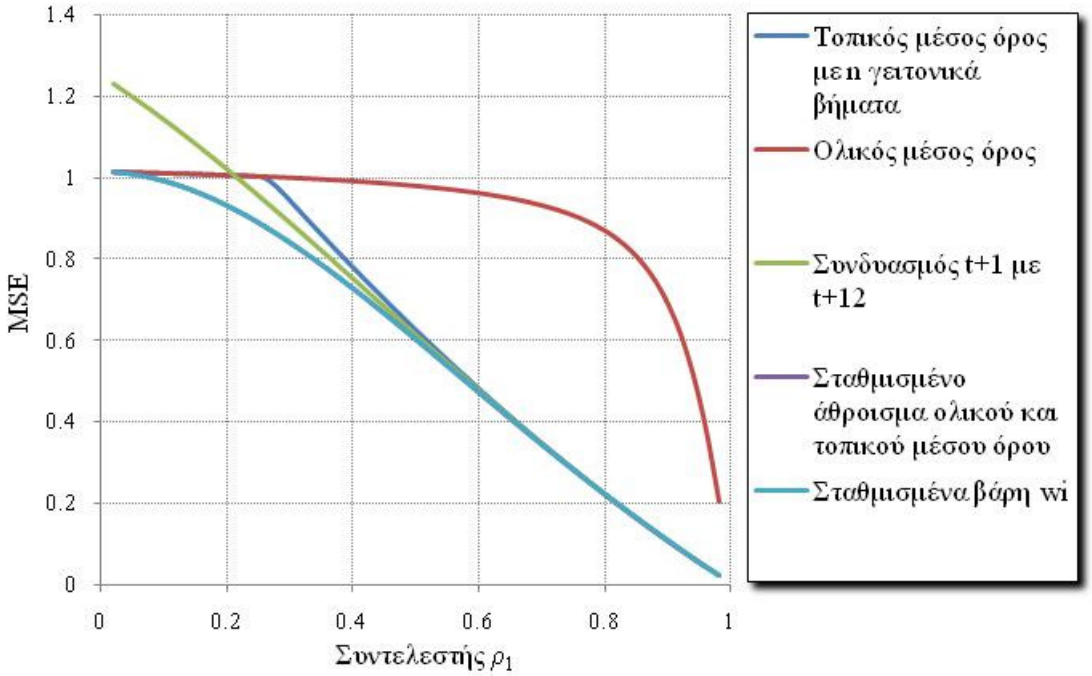
Αντίθετα η χρήση του τοπικού μέσου όρου που προκύπτει από το βέλτιστο αριθμό γειτονικών βημάτων (optimal steps *n*), δίνει πολύ καλά αποτελέσματα που πρακτικώς ταυτίζονται και με τις υπόλοιπες τρείς μεθοδολογίες για υψηλές τιμές του συντελεστή αυτοσυσχέτισης.

Πιο συγκεκριμένα, μπορούμε να συνοψίσουμε τα αποτελέσματα που προκύπτουν από την εφαρμογή της μεθόδου του τοπικού μέσου όρου, ανάλογα με τη τιμή του συντελεστή αυτοσυσχέτισης ($\rho_1$), ως εξής:

- για $\rho_1 < \rho_{cr} \approx 0.24$, προτείνεται η χρήση του ολικού μέσου όρου
- για $\rho_1 \geq \rho_{cr} \approx 0.24$, η χρήση ενός τοπικού μέσου όρου χρησιμοποιώντας μια τιμή πριν και μία μετά ενδείκνυται.

Η χρήση δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών ($t$+1, $t$+12) βελτιώνει ελαφρώς την εκτίμηση συγκριτικά με την μέθοδο που βασίζεται στο τοπικό μέσο όρο για τιμές του $\rho_1$ πάνω από 0.2, ενώ για μικρότερες τιμές έχει χειρότερα αποτελέσματα.

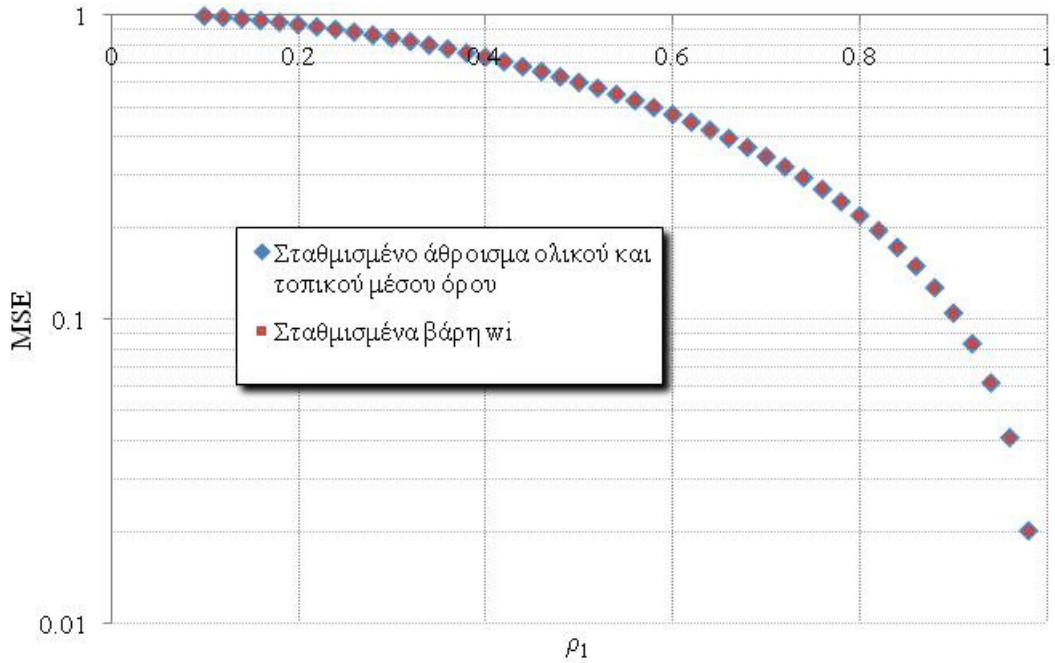
Πιο συγκεκριμένα, η χρήση του ολικού μέσου όρου (global average) καθώς επίσης και η μέθοδος του τοπικού μέσου όρου (optimal steps *n*), φαίνεται να έχει καλύτερα αποτελέσματα για μικρές τιμές του συντελεστή αυτοσυσχέτισης (για $\rho_1 \leq 0.2$), ενώ για τιμές στο διάστημα $0.2 \leq \rho_1 \leq 0.5$ υπερτερεί η μέθοδος που βασίζεται στη χρήση τεσσάρων γειτονικών τιμών ($t$+1, $t$+12). Για τιμές του συντελεστή αυτοσυσχέτισης $\rho_1$ μεγαλύτερες του 0.5, η μέθοδος με το βέλτιστο αριθμό γειτονικών βημάτων (optimal steps *n*) και η μέθοδος που βασίζεται στη χρήση τεσσάρων γειτονικών τιμών ($t$+1,$t$+12), έχουν παρόμοια αποτελέσματα. Αναλυτικότερα, τα πιο πάνω συμπεράσματα επιβεβαιώνονται από το ακόλουθο διάγραμμα.




Σχήμα 6.1 Ελάχιστη τιμή του MSE – συντελεστής αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) με εφαρμογή των μεθόδων που παρουσιαστήκαν.

Παρατηρούμε επίσης πως τα βέλτιστα αποτελέσματα προκύπτουν από την εφαρμογή της μεθόδου των σταθμισμένων βαρών (weights $w_i$). Όπως όμως έχουμε ήδη τονίσει, η μέθοδος των σταθμισμένων βαρών (παραλλαγή της μεθόδου kriging) μπορεί να έχει τα καλύτερα αποτελέσματα, να δίνει δηλαδή το μικρότερο MSE, αλλά απαιτεί μεγάλο υπολογιστικό φόρτο, γεγονός που την καθιστά δύσχρηστη για την γρήγορη συμπλήρωση μεμονωμένων ελλιπών τιμών.

Όμως και ο συνδυασμός του τοπικού με τον ολικό μέσο όρο ($N$-$n$) δίνει εξίσου καλά αποτελέσματα που πρακτικώς ταυτίζονται με τα αντίστοιχα που προκύπτουν για σταθμισμένα βάρη. Η εκτίμηση δηλαδή που προκύπτει από το σταθμισμένο άθροισμα του ολικού μέσου όρου και ενός τοπικού που αποτελείται μόνο από δύο τιμές, μια πριν και μια μετά, με τη χρήση μόνο μιας παραμέτρου δίνει πολύ ικανοποιητικά αποτελέσματα. Αναλυτικότερα η διαπίστωση αυτή φαίνεται στο παρακάτω διάγραμμα.




Σχήμα 6.2 Ελάχιστη τιμή του MSE – συντελεστής αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$).

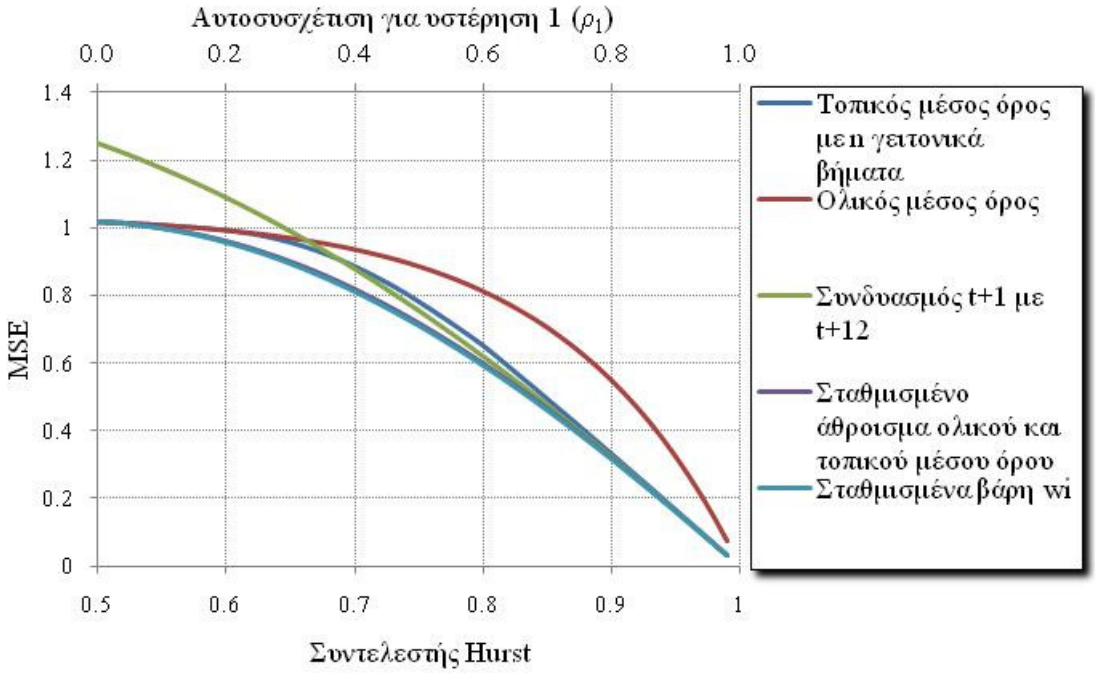
Η τόσο καλή συμπεριφορά της μεθόδου *N-n*, δηλαδή του συνδυασμού του ολικού και του τοπικού μέσου όρου, καθιστά τη μέθοδο αυτή βέλτιστη καθώς δίνει τα ίδια αποτελέσματα με αυτά της μεθόδου των σταθμισμένων βαρών ενώ παράλληλα απαιτεί λιγότερες πράξεις και συνεπώς είναι πιο εύκολα εφαρμόσιμη. Δηλαδή, ανεξάρτητα από την τιμή της αυτοσυσχέτισης της χρονοσειράς, η μέθοδος που βασίζεται στο συνδυασμό του ολικού μέσου όρου και του τοπικού με χρήση μιας γειτονικής τιμής πριν και μιας μετά με την κατάλληλη παράμετρο $\lambda$, δίνει ταυτόσημα σχεδόν αποτελέσματα με αυτά που προκύπτουν από τη χρήση της πολύπλοκης μεθόδου των σταθμισμένων βαρών.

## 6.2 Χρονοσειρές που παρουσιάζουν δυναμική Hurst - Kolmogorov

Όσον αφορά τις χρονοσειρές που προσομοιώνονται με ανελίξεις με δυναμική Hurst–Kolmogorov, δηλαδή χρονοσειρές με έντονη δομή αυτοσυσχέτισης, τα αποτελέσματα που προκύπτουν για κάθε μια από τις μεθόδους που παρουσιάσαμε συνοψίζονται στο πιο κάτω διάγραμμα.

Και εδώ, όπως και στην περίπτωση των ανελίξεων Markov, διαπιστώνουμε πως η χρήση του ολικού μέσου όρου (global average) δεν είναι δικαιολογημένη. Ειδικά όσο αυξάνεται ο συντελεστής Hurst τόσο πιο δυσμενείς είναι τα αποτελέσματα που προκύπτουν από τη χρήση του ολικού μέσου όρου. Αντίθετα, ικανοποιητικά αποτελέσματα δίνει η χρήση ενός τοπικού μέσου όρου με επιλογή κατάλληλου

αριθμού γειτονικών βημάτων (optimal steps *n*), όπως αυτή παρουσιάστηκε στο υποκεφάλαιο 4.3.2.




Σχήμα 6.3 Ελάχιστη τιμή του MSE – συντελεστής *H* (κύριος οριζόντιος άξονας) και συντελεστής $\rho_1$ (δευτερεύον οριζόντιος άξονας) με εφαρμογή των μεθόδων που παρουσιαστήκαν.

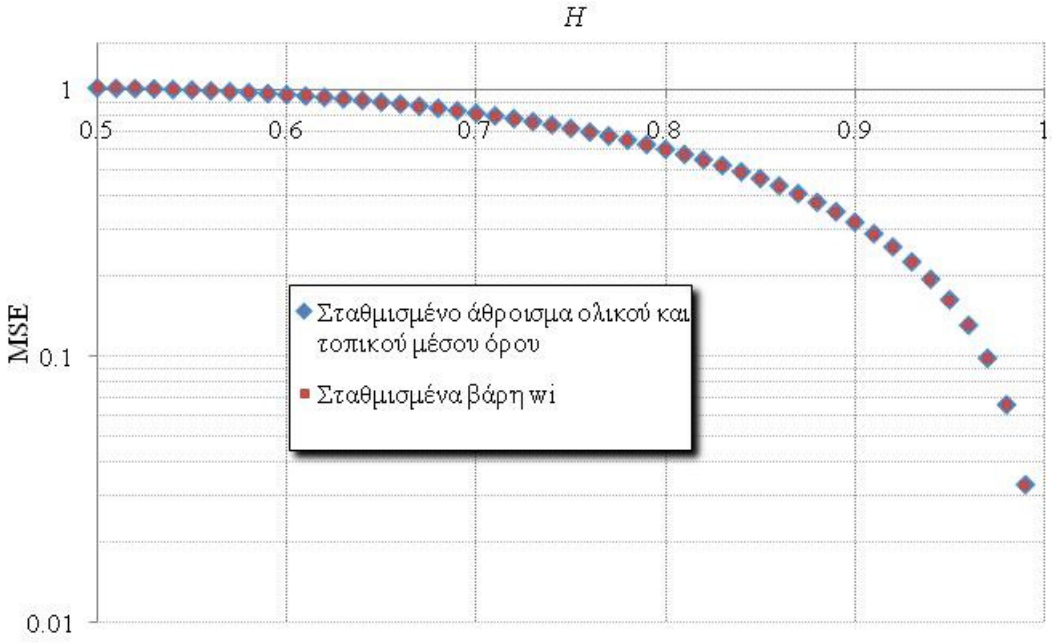
Πιο συγκεκριμένα, για τη μέθοδο του τοπικού μέσου όρου (optimal steps *n*), ανάλογα με τη τιμή του συντελεστή Hurst (*H*), μπορούμε να βγάλουμε τα ακόλουθα συμπεράσματα:

- Για $H = 0.50\text{-}0.60$
  προτιμάται η χρήση του ολικού μέσου όρου
- Για $H = 0.70$
  απαιτούνται 4 τιμές πριν και 4 μετά την ελλείπουσα τιμή
- Για $H = 0.72$
  απαιτούνται 3 τιμές πριν και 3 μετά την ελλείπουσα τιμή
- Για $H = 0.74$
  απαιτούνται 2 τιμές πριν και 2 μετά την ελλείπουσα τιμή
- Για $H \geq 0.80$
  χρειαζόμαστε μόνο μία τιμή πριν και μία μετά

Η χρήση δύο γειτονικών μηνιαίων και δύο γειτονικών ετήσιων τιμών (*t*+1, *t*+12), δεν έχει ικανοποιητικά αποτελέσματα για χρονοσειρές με ασθενή δομή αυτοσυσχέτισης (για τιμές του συντελεστή *H* μικρότερες του 0.65) ενώ για υψηλότερες τιμές του συντελεστή *H*, η μέθοδος αυτή δίνει ικανοποιητικά αποτελέσματα που συγκλίνουν με αυτά της μεθόδου του τοπικού μέσου όρου.

Και εδώ, όπως και στη περίπτωση των ανελίξεων Markov, τα καλύτερα αποτελέσματα, δηλαδή τη μικρότερη τιμή του MSE, προκύπτουν από τη χρήση των σταθμισμένων βαρών. Το εντυπωσιακό όμως είναι πως τα αποτελέσματα αυτά προσεγγίζονται πολύ ικανοποιητικά, πρακτικώς ταυτίζονται, με τα αντίστοιχα που προκύπτουν από το συνδυασμό του ολικού μέσου όρου και ενός τοπικού με μια τιμή πριν και μια μετά (*N-n*), χρησιμοποιώντας την κατάλληλη παράμετρο *λ*.

Στο πιο κάτω διάγραμμα φαίνονται οι τιμές του MSE που προκύπτουν για διάφορους συντελεστές Hurst με εφαρμογή της μεθόδου των σταθμισμένων βαρών και της μεθόδου του σταθμισμένου αθροίσματος του τοπικού και του ολικού μέσου όρου με μια παράμετρο *λ*.




Σχήμα 6.4 Ελάχιστη τιμή του MSE – συντελεστής *H*.

Παρατηρούμε πως το MSE που προκύπτει για τις διάφορες τιμές του συντελεστή Hurst με εφαρμογή της μεθόδου *N-n*, σχεδόν συμπίπτει με τις αντίστοιχες τιμές που δίνει η μέθοδος των σταθμισμένων βαρών. Παράλληλα όμως η μέθοδος *N-n* έχει λιγότερο υπολογιστικό φόρτο σε σχέση με τον υπολογισμό των σταθμισμένων βαρών.

Οι πιο πάνω διαπιστώσεις καθιστούν τη μέθοδο του συνδυασμού του ολικού και ενός τοπικού μέσου όρου που αποτελείται από δύο τιμές, ήτοι την αμέσως προηγούμενη και την αμέσως επόμενη στην ελλείπουσα τιμή, με την κατάλληλη παράμετρο *λ*, τη βέλτιστη για την συμπλήρωση μεμονωμένων κενών που παρουσιάζονται στις υδρομετεωρολογικές χρονοσειρές.

## ΒΙΒΛΙΟΓΡΑΦΙΑ


Beran, J.A., Statistical methods for data with long-range dependence. *Statistical Science 7* (4), 404–427, 1992.

Box and Jenkins. *Time Series Analysis forecasting and control.* Holden-Day, 1976.

Brown. *Introduction to Random Signal Analysis and Kalman Filtering*. John Wiley and Sons, 1983.

Dialynas, Y., P. Kossieris, K. Kyriakidis, A. Lykou, Y. Markonis, C. Pappas, S.M. Papalexiou, and D. Koutsoyiannis, Optimal infilling of missing values in hydrometeorological time series, *European Geosciences Union General Assembly 2010, Geophysical Research Abstracts, Vol. 12*, Vienna, EGU2010-9702, European Geosciences Union, 2010.

Haward R. A. *Dynamic Probabilistic Systems* New York : John Wiley & Sons, 1971.

Hirsch R. M., Helsel, D. R., Cohn, T. A., & Gilroy, E. J. Statistical Analysis of Hydrologic Data. Maidment, *Handbook of Hydrology*. McGraw-Hill.

Hurst, H. E., Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng*., *116*, 776–808, 1951.

Jones, P.D., Briffa, K.R., Barnett, T.P., Tett, S.F.B., High-resolution paleoclimatic records for the last millennium: interpretation, integration and comparison with general circulation model control-run temperatures. Holocene *Vol. 8* (4), 455–471, 1998.

Kitanidis. Geostatistics. Maidment, *Handbook of Hydrology*. McGraw-Hill.

Koutsoyiannis, D., and A. Langousis, Precipitation, *Treatise on Water Science*, edited by S. Uhlenbrook, Elsevier, 2011, (in press).

Koutsoyiannis, D., A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, 2010.

Koutsoyiannis, D., Hurst, Joseph, colours and noises: The importance of names in an important natural behaviour, *Niche Modeling*, 10 pages, 2006.

Koutsoyiannis, D., A toy model of climatic variability with scaling behaviour, *Journal of Hydrology*, 322, 25–48, 2006.

Koutsoyiannis, D., Hydrologic persistence and the Hurst phenomenon, *Water Encyclopedia, Vol. 4, Surface and Agricultural Water*, edited by J. H. Lehr and J. Keeley, 210–221, Wiley, New York, 2005.

Koutsoyiannis, D., Stochastic simulation of hydrosystems, *Water Encyclopedia, Vol. 4, Surface and Agricultural Water*, edited by J. H. Lehr and J. Keeley, 421–430, Wiley, New York, 2005.

Koutsoyiannis, D., The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47 (4), 573–595, 2002.

Mandelbrot, B. B., *The Fractal Geometry of Nature*, Freeman, New York, 1977.

Matheron *Les variables regionalisees et leur estimation*. Paris. Masson, 1965.

Mesa, O. J. and Poveda, G. The Hurst effect: the scale of fluctuation approach. *Water Resources Research*, *29* (12) 3995-4002, 1993.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1991.

Salas, J. D., Analysis and modeling of hydrologic time series, *Handbook of Hydrology*, Maidment, 19, McGraw-Hill, New York, 1993.

Warren, Viessman Jr.; Gary L., Lewis *Introduction to Hydrology*. 4th Edition.

Webster & Oliver *Geostatistics for Environmental Scientists*, 2001.

Κουτσογιάννης, Δ., *Σημειώσεις Στοχαστικών Μεθόδων στους Υδατικούς Πόρους*, Έκδοση 3, 100 σελίδες, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 2007.

Κουτσογιάννης, Δ., και Θ. Ξανθόπουλος, *Τεχνική Υδρολογία*, Έκδοση 3, 418 σελίδες, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1999.

Κουτσογιάννης, Δ., *Στατιστική Υδρολογία*, Έκδοση 4, 312 σελίδες, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1997.

Κουτσογιάννης, Δ., 1 μέτρηση = 1000 υπολογισμοί, *Εφημερίδα "Το Βήμα της Κυριακής", Ειδικό ένθετο για το νερό*, 18–20, 12 Νοεμβρίου 2000.

# ΠΑΡΑΡΤΗΜΑΤΑ

## ΠΑΡΑΡΤΗΜΑ Α

Υπολογισμός του μέσου τετραγωνικού σφάλματος εκτίμησης $\hat{x}_t$ όπου:

$$\hat{x}_t = \frac{\sum_{i=1}^{-n} x_{t+i} + \sum_{i=1}^{n} x_{t+i}}{2n}$$

Απόδειξη της σχέσης:

$$MSE := E\left[e^2\right] = \frac{1}{2}\left(\frac{\sigma}{n}\right)^2\left[(2n+1)\left(n - 2\sum_{i=1}^{n}\rho_i\right) + \sum_{i=1}^{2n}(2n+1-i)\,\rho_i\right]$$

Είδαμε ότι το μέσο τετραγωνικό σφάλμα υπολογίζεται, εξ' ορισμού από τη σχέση

$$MSE := E\left[e^2\right] = E\left[\left(x_t - \frac{\sum_{i=1}^{-n} x_{t+i} + \sum_{i=1}^{n} x_{t+i}}{2n}\right)^2\right],$$

όπου

$$\left(x_t - \frac{\sum_{i=1}^{-n} x_{t+i} + \sum_{i=1}^{n} x_{t+i}}{2n}\right)^2 = {x_t}^2 - 2x_t\,\frac{\sum_{i=1}^{-n} x_{t+i} + \sum_{i=1}^{n} x_{t+i}}{2n} + \left(\frac{\sum_{i=1}^{-n} x_{t+i} + \sum_{i=1}^{n} x_{t+i}}{2n}\right)^2$$

$$= {x_t}^2 - \frac{1}{n}x_t\sum_{i=1}^{-n} x_{t+i} - \frac{1}{n}x_t\sum_{i=1}^{n} x_{t+i} + \left(\frac{\left(\sum_{i=1}^{-n} x_{t+i}\right)^2 + \left(\sum_{i=1}^{n} x_{t+i}\right)^2 + 2\sum_{i=1}^{-n} x_{t+i}\sum_{i=1}^{n} x_{t+i}}{4n^2}\right)$$

$$= {x_t}^2 - \frac{1}{n}x_t\sum_{i=1}^{-n} x_{t+i} - \frac{1}{n}x_t\sum_{i=1}^{n} x_{t+i} + \frac{1}{4n^2}\left(\sum_{i=1}^{-n} x_{t+i}\right)^2 + \frac{1}{4n^2}\left(\sum_{i=1}^{n} x_{t+i}\right)^2 + \frac{1}{2n^2}\sum_{i=1}^{-n} x_{t+i}\sum_{i=1}^{n} x_{t+i}$$

με αναμενόμενες τιμές η παραπάνω σχέση γίνεται

$$MSE := E\left[e^2\right]$$
$$= E\left[x_t^{\,2} - \frac{1}{n}x_t\sum_{i=1}^{-n}x_{t+i} - \frac{1}{n}x_t\sum_{i=1}^{n}x_{t+i} + \frac{1}{4n^2}\left(\sum_{i=1}^{-n}x_{t+i}\right)^2 + \frac{1}{4n^2}\left(\sum_{i=1}^{n}x_{t+i}\right)^2 + \frac{1}{2n^2}\sum_{i=1}^{-n}x_{t+i}\sum_{i=1}^{n}x_{t+i}\right]$$
$$= E\left[x_t^{\,2}\right] - \frac{1}{n}E\left[x_t\sum_{i=1}^{-n}x_{t+i}\right] - \frac{1}{n}E\left[x_t\sum_{i=1}^{n}x_{t+i}\right] + \frac{1}{4n^2}E\left[\left(\sum_{i=1}^{-n}x_{t+i}\right)^2\right]$$
$$+ \frac{1}{4n^2}E\left[\left(\sum_{i=1}^{n}x_{t+i}\right)^2\right] + \frac{1}{2n^2}E\left[\sum_{i=1}^{-n}x_{t+i}\sum_{i=1}^{n}x_{t+i}\right]$$

όπου

$$E\left[x_t^{\,2}\right] = \sigma^2 + \mu^2$$
$$E\left[\frac{1}{n}x_t\sum_{i=1}^{-n}x_{t+i}\right] = E\left[\frac{1}{n}x_t\sum_{i=1}^{n}x_{t+i}\right] = \frac{1}{n}\sigma^2\sum_{i=1}^{n}\rho_i + n\mu^2$$
$$E\left[\frac{1}{4n^2}\left(\sum_{i=1}^{-n}x_{t+i}\right)^2\right] = E\left[\frac{1}{4n^2}\left(\sum_{i=1}^{n}x_{t+i}\right)^2\right] = \frac{1}{4n^2}\left[\sigma^2\left(n + 2\sum_{i=1}^{n-1}(n-i)\rho_i\right) + n^2\mu^2\right]$$
$$E\left[\frac{1}{2n^2}\sum_{i=1}^{-n}x_{t+i}\sum_{i=1}^{n}x_{t+i}\right] = \frac{1}{2n^2}\left[\sigma^2\left(\sum_{i=2}^{n+1}(i-1)\rho_i + \sum_{i=n+2}^{2n}(2n+1-i)\rho_i\right) + n^2\mu^2\right]$$

Άρα,

$$MSE := E\left[e^2\right] = \sigma^2 + \mu^2 - \frac{2}{n}\sigma^2\sum_{i=1}^{n}\rho_i + n\mu^2 + \frac{2}{4n^2}\left[\sigma^2\left(n + 2\sum_{i=1}^{n-1}(n-i)\rho_i\right) + n^2\mu^2\right]$$
$$+ \frac{1}{2n^2}\left[\sigma^2\left(\sum_{i=2}^{n+1}(i-1)\rho_i + \sum_{i=n+2}^{2n}(2n+1-i)\rho_i\right) + n^2\mu^2\right]$$

Και συνεπώς, μετά από κάποιες αλγεβρικές πράξεις και απλοποιήσεις έχουμε την τελική σχέση:

$$MSE := E\left[e^2\right] = \frac{1}{2}\left(\frac{\sigma}{n}\right)^2\left[(2n+1)\left(n - 2\sum_{i=1}^{n}\rho_i\right) + \sum_{i=1}^{2n}(2n+1-i)\rho_i\right]$$

## ΠΑΡΑΡΤΗΜΑ Β

Υπολογισμός του μέσου τετραγωνικού σφάλματος εκτίμησης $\hat{x}_t$ όπου:

$$\hat{x}_t = \theta \frac{x_{t-1} + x_{t+1}}{2} + (1-\theta) \frac{x_{t-12} + x_{t+12}}{2}$$

$$\begin{aligned}
e^2 = \left(x_t - \hat{x}_t\right)^2 &= \left[ x_t - \left( \theta \frac{x_{t-1} + x_{t+1}}{2} + (1-\theta) \frac{x_{t-12} + x_{t+12}}{2} \right) \right]^2 \\
&= x_t^2 + \frac{\theta^2}{2^2} (x_{t-1} + x_{t+1})^2 + \frac{(1-\theta)^2}{2^2} (x_{t-12} + x_{t+12})^2 + \frac{\theta(1-\theta)}{2} (x_{t-1} + x_{t+1})(x_{t-12} + x_{t+12}) \\
&- 2x_t \left( \theta \frac{x_{t-1} + x_{t+1}}{2} + (1-\theta) \frac{x_{t-12} + x_{t+12}}{2} \right) \\
&= x_t^2 + \frac{\theta^2}{4} \left( x_{t-1}^2 + x_{t+1}^2 + 2x_{t-1}x_{t+1} \right) + \frac{(1-\theta)^2}{4} \left( x_{t-12}^2 + x_{t+12}^2 + 2x_{t-12}x_{t+12} \right) \\
&+ \frac{\theta(1-\theta)}{2} \left( x_{t-1}x_{t-12} + x_{t-1}x_{t+12} + x_{t+1}x_{t-12} + x_{t+1}x_{t+12} \right) - \theta \left( x_t x_{t-1} + x_t x_{t+1} \right) - (1-\theta) \left( x_t x_{t-12} + x_t x_{t+12} \right)
\end{aligned}$$

Η παραπάνω σχέση με αναμενόμενες τιμές μας δίνει το μέσο τετραγωνικό σφάλμα της εκτίμησης, που είναι

$$\begin{aligned}
MSE &= E\left[ e^2 \right] \\
&= E\left[ \left( x_t - \hat{x}_t \right)^2 \right] \\
&= \sigma^2 \left[ 1 - 2\theta(\rho_1 - \rho_{12}) - 2\rho_{12} + \frac{\theta^2}{2} (\rho_2 + 1) + \frac{(1-\theta)^2}{2} (\rho_{24} + 1) + \theta(1-\theta)(\rho_{11} + \rho_{13}) \right]
\end{aligned}$$

## ΠΑΡΑΡΤΗΜΑ C

Υπολογισμός του μέσου τετραγωνικού σφάλματος εκτίμησης $\hat{x}_t$ όπου

$$\hat{x}_t = \lambda \frac{\sum_{i=-N}^{N} x_i}{2N} + (1-\lambda)\frac{\sum_{i=-n}^{-1} x_i + \sum_{i=1}^{n} x_i}{2n}$$

Προκειμένου να απλοποιήσουμε την πιο πάνω σχέση, μελετάμε την περίπτωση όπου ο αριθμός των γειτονικών βημάτων που χρησιμοποιούνται για τον υπολογισμό του τοπικού μέσου όρου είναι 2, δηλαδή n=1, μια μέτρηση πριν και μια μετά την ελλείπουσα τιμή. Με την πιο πάνω απλοποίηση η σχέση γίνεται:

$$\hat{x}_t = \lambda \frac{\sum_{i=-N}^{N} x_i}{2N} + (1-\lambda)\frac{x_{-1} + x_1}{2}$$

Όπως έχουμε ήδη αναφέρει, το τετραγωνικό σφάλμα $e^2$ δίνεται από τη σχέση:

$$e^2 = \left(x_t - \hat{x}_t\right)^2$$

$$= \left[ x_t - \left( \lambda \frac{\sum_{i=-N}^{N} x_i}{2N} + (1-\lambda)\frac{x_{-1} + x_1}{2} \right) \right]^2$$

$$= \left[ \left( x_t - \frac{x_{-1} + x_1}{2} \right) - \lambda \left( \frac{\sum_{i=-N}^{N} x_i}{2N} - \frac{x_{-1} + x_1}{2} \right) \right]^2$$

$$= \left( x_t - \frac{x_{-1} + x_1}{2} \right)^2 - 2\lambda \left( x_t - \frac{x_{-1} + x_1}{2} \right) \left( \frac{\sum_{i=-N}^{N} x_i}{2N} - \frac{x_{-1} + x_1}{2} \right) + \lambda^2 \left( \frac{\sum_{i=-N}^{N} x_i}{2N} - \frac{x_{-1} + x_1}{2} \right)^2$$

Θέτουμε

$$A == \left( x_t - \frac{x_{-1} + x_1}{2} \right)^2,$$

$$B = \left( x_t - \frac{x_{-1} + x_1}{2} \right) \left( \frac{\sum_{i=-N}^{N} x_i}{2N} - \frac{x_{-1} + x_1}{2} \right)$$

και

$$C = \left( \frac{\sum_{i=-N}^{N} x_i}{2N} - \frac{x_{-1} + x_1}{2} \right)^2$$

Και έχουμε, $e^2 = A - 2\lambda B + \lambda^2 C$ .

Έτσι, το μέσο τετραγωνικό σφάλμα $E\left[e^2\right]$ υπολογίζεται από τη σχέση:

$$MSE := E\left[e^2\right] = E\left[A\right] - 2\lambda E\left[B\right] + \lambda^2 E\left[C\right]$$

Αρκεί λοιπόν να υπολογίζουμε τα $E[A]$, $E[B]$ και $E[C]$.

Όμως, έχουμε αποδείξει (βλέπε ΠΑΡΑΡΤΗΜΑ Α) πώς:

$$E\left[ \left( x_t - \frac{\sum_{i=-n}^{-1} x_i + \sum_{i=1}^{n} x_i}{2n} \right)^2 \right] = \frac{1}{2}\left(\frac{\sigma}{n}\right)^2 \left[ (2n+1)\left( n - 2\sum_{i=1}^{n} \rho_i \right) + \sum_{i=1}^{2n} (2n+1-i)\,\rho_i \right]$$

άρα

$$E\left[A\right] = \frac{1}{2}\sigma^2 (3 - 4\rho_1 + \rho_2)$$

Για την ποσότητα $B$ έχουμε:

$$B = \left( x_t - \frac{x_{-1} + x_1}{2} \right) \left( \frac{\sum_{i=-N}^{N} x_i}{2N} - \frac{x_{-1} + x_1}{2} \right)$$

$$= \frac{1}{2N} x_t \sum_{i=-N}^{N} x_i + \frac{1}{2} x_t \frac{x_{-1} + x_1}{2} - \frac{1}{4N}\left(x_{-1} + x_1\right) \sum_{i=-N}^{N} x_i - \frac{1}{4}\left(x_{-1} + x_1\right)^2$$

Εξετάζουμε τώρα κάθε ποσότητα χωριστά, και έχουμε:

$$x_t \sum_{i=-N}^{N} x_i = x_t \sum_{i=-N}^{-1} x_i + x_t \sum_{i=1}^{N} x_i$$

Όμως,

$$E\left[ x_t \sum_{i=-N}^{-1} x_i \right] = E\left[ x_t \sum_{i=1}^{N} x_i \right] = \sigma^2 \sum_{i=1}^{N} \rho_i + N\mu^2$$

$$\Rightarrow E\left[ x_t \sum_{i=-N}^{N} x_i \right] = 2\sigma^2 \sum_{i=1}^{N} \rho_i + 2N\mu^2$$

Αντίστοιχα,

$$E\left[x_t\left(x_{-1}+x_1\right)\right]=2\sigma^2\rho_1+2\mu^2$$

Επίσης,

$$E\left[\left(x_{-1}+x_1\right)\sum_{i=-N}^{N}x_i\right]=2\sigma^2\left(\sum_{i=1}^{N-1}\rho_i+\sum_{i=2}^{N+1}\rho_i+1\right)+4N\mu^2$$

Τέλος,

$$E\left[\left(x_{-1}+x_1\right)^2\right]=2\sigma^2\left(\rho_2+1\right)+4\mu^2$$

Συνεπώς, η ποσότητα $E[B]$ ισούται με

$$E\left[B\right]=\sigma^2\left[\frac{1}{N}\sum_{i=1}^{N}\rho_t-\frac{1}{2N}\left(\sum_{i=1}^{N-1}\rho_i-\sum_{i=2}^{N+1}\rho_i+1\right)-\rho_1+\frac{\rho_2}{2}+0.5\right]$$

Για την ποσότητα $C$ έχουμε

$$C=\left(\frac{\sum_{i=-N}^{N}x_i}{2N}-\frac{x_{-1}+x_1}{2}\right)^2=\frac{\left(\sum_{i=-N}^{N}x_i\right)^2}{4N^2}+\frac{x_{-1}^2+x_1^2+2x_{-1}x_1}{4}-\frac{1}{2N}\sum_{i=-N}^{N}x_i\left(x_{-1}+x_1\right)$$

όπου

$$\left(\sum_{i=-N}^{N}x_i\right)^2=\left(\sum_{i=-N}^{-1}x_i+\sum_{i=1}^{N}x_i\right)^2=\left(\sum_{i=-N}^{-1}x_i\right)^2+\left(\sum_{i=1}^{N}x_i\right)^2+2\sum_{i=-N}^{-1}x_i\sum_{i=1}^{N}x_i$$

και

$$\begin{aligned}E\left[\left(\sum_{i=-N}^{N}x_i\right)^2\right]&=2\left[\sigma^2\left(N+2\sum_{i=1}^{N-1}\left(N-i\right)\rho_i\right)+N^2\mu^2\right]\\&\quad+2\left[\sigma^2\left(\sum_{i=2}^{N+1}\left(i-1\right)\rho_i+\sum_{i=N+2}^{2N}\left(2N+1-i\right)\rho_i\right)+N^2\mu^2\right]\\&=2\sigma^2\left[N+2\sum_{i=1}^{N-1}\left(N-i\right)\rho_i+\sum_{i=2}^{N+1}\left(i-1\right)\rho_i+\sum_{i=N+2}^{2N}\left(2N+1-i\right)\rho_i\right]+4N^2\mu^2\end{aligned}$$

Και

$$E\left[\frac{x_{-1}^2+x_1^2+2x_{-1}x_1}{4}\right]=\frac{\sigma^2}{2}\left(\rho_2+1\right)+\mu^2$$

Και

$$E\left[\sum_{i=-N}^{N}x_i\left(x_{-1}+x_1\right)\right]=2\sigma^2\left(\sum_{i=1}^{N-1}\rho_i+\sum_{i=2}^{N+1}\rho_i+1\right)+4N\mu^2$$

Άρα

$$E[C]=\frac{1}{2}\left(\frac{\sigma}{N}\right)^2\left(2\sum_{i=1}^{N-1}(N-i)\rho_i+\sum_{i=2}^{N+1}(i-1)\rho_i+\sum_{i=N+2}^{2N}(2N+1-i)\rho_i+N\right)$$
$$+\frac{\sigma^2}{2}(\rho_2+1)-\frac{\sigma^2}{N}\left(\sum_{i=1}^{N-1}\rho_i+\sum_{i=2}^{N+1}\rho_i+1\right)$$

$$=\sigma^2\left[\begin{array}{l}\frac{1}{2N^2}\left(2\sum_{i=1}^{N-1}(N-i)\rho_i+\sum_{i=2}^{N+1}(i-1)\rho_i+\sum_{i=N+2}^{2N}(2N+1-i)\rho_i+N\right)\\+\frac{\rho_2}{2}+\frac{1}{2}-\frac{1}{N}\left(\sum_{i=1}^{N-1}\rho_i+\sum_{i=2}^{N+1}\rho_i+1\right)\end{array}\right]$$

Συνεπώς προκύπτει,

$$MSE:=E\left[e^2\right]=E[A]-2\lambda E[B]+\lambda^2E[C]$$

$$=\frac{1}{2}\sigma^2(3-4\rho_1+\rho_2)$$

$$-2\lambda\sigma^2\left[\frac{1}{N}\sum_{i=1}^{N}\rho_\iota-\frac{1}{2N}\left(\sum_{i=1}^{N-1}\rho_i-\sum_{i=2}^{N+1}\rho_i+1\right)-\rho_1+\frac{\rho_2}{2}+0.5\right]$$

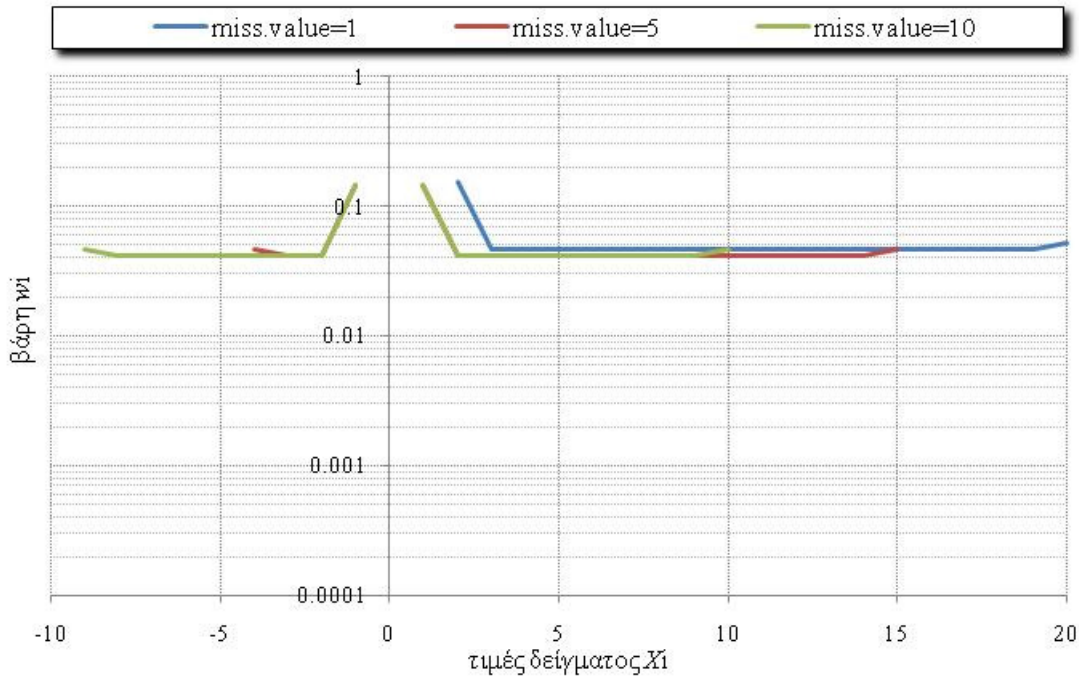
$$+\lambda^2\sigma^2\left[\begin{array}{l}\frac{1}{2N^2}\left(2\sum_{i=1}^{N-1}(N-i)\rho_i+\sum_{i=2}^{N+1}(i-1)\rho_i+\sum_{i=N+2}^{2N}(2N+1-i)\rho_i+N\right)\\+\frac{\rho_2}{2}+\frac{1}{2}-\frac{1}{N}\left(\sum_{i=1}^{N-1}\rho_i+\sum_{i=2}^{N+1}\rho_i+1\right)\end{array}\right]$$

## ΠΑΡΑΡΤΗΜΑ D

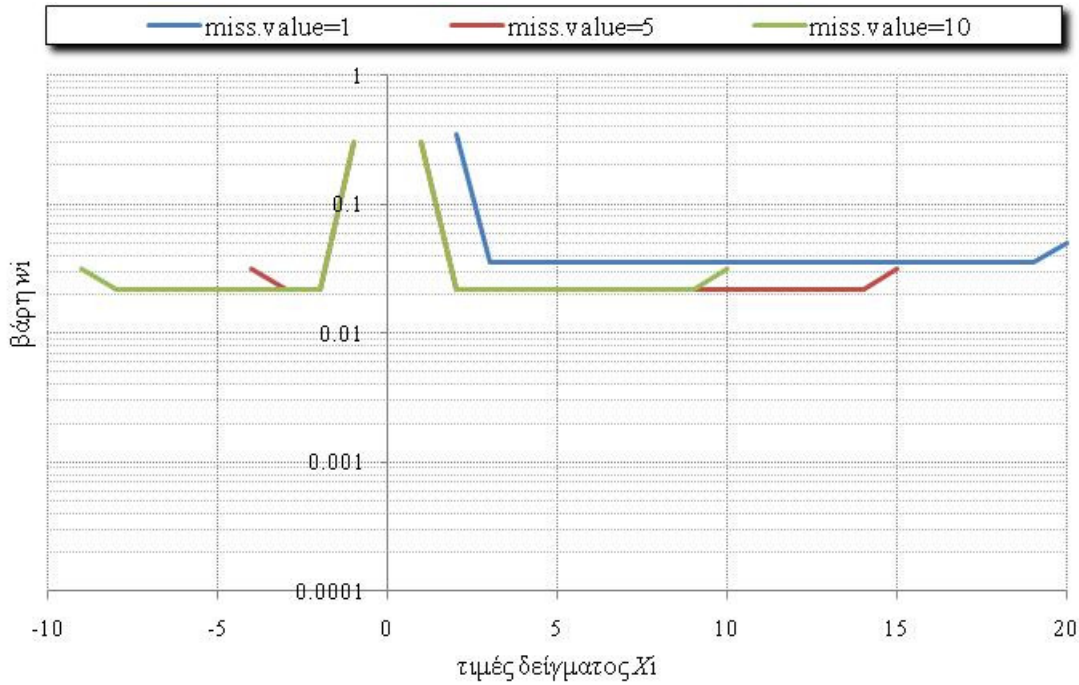
Συμπλήρωση μεμονωμένων ελλειπουσών τιμών με σταθμισμένα βάρη $w_i$ σε υδρομετεωρολογικές χρονοσειρές που προσομοιώνονται από ανελίξεις τύπου Markov.

Ανάλογα με το συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) κατασκευάστηκαν τα πιο κάτω διαγράμματα που παρουσιάζουν τα σταθμισμένα βάρη των τιμών της χρονοσειράς, ανάλογα με την θέση της ελλείπουσας τιμής.

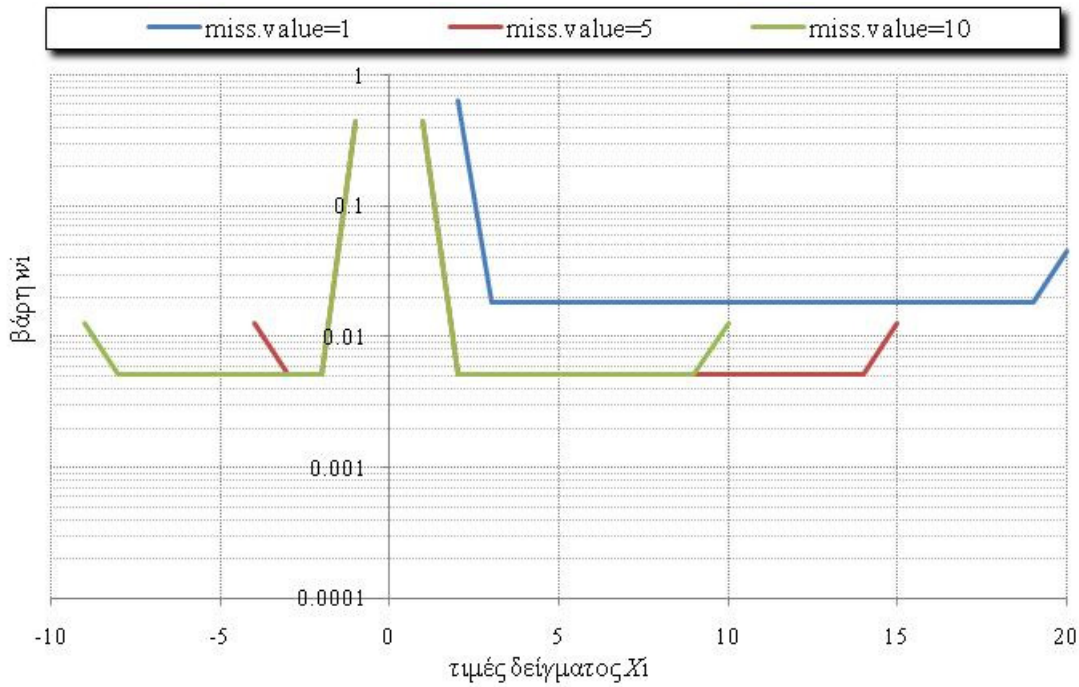
- Για δείγμα μεγέθους 20 τιμών έχουμε:



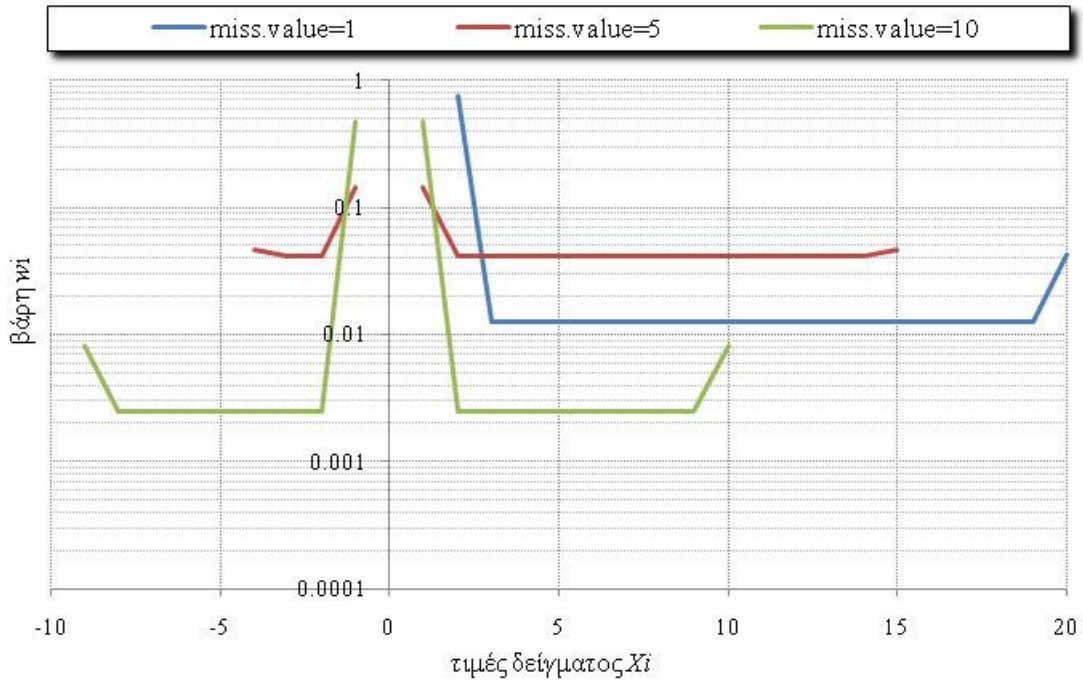

Διάγραμμα 0.1 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.1.



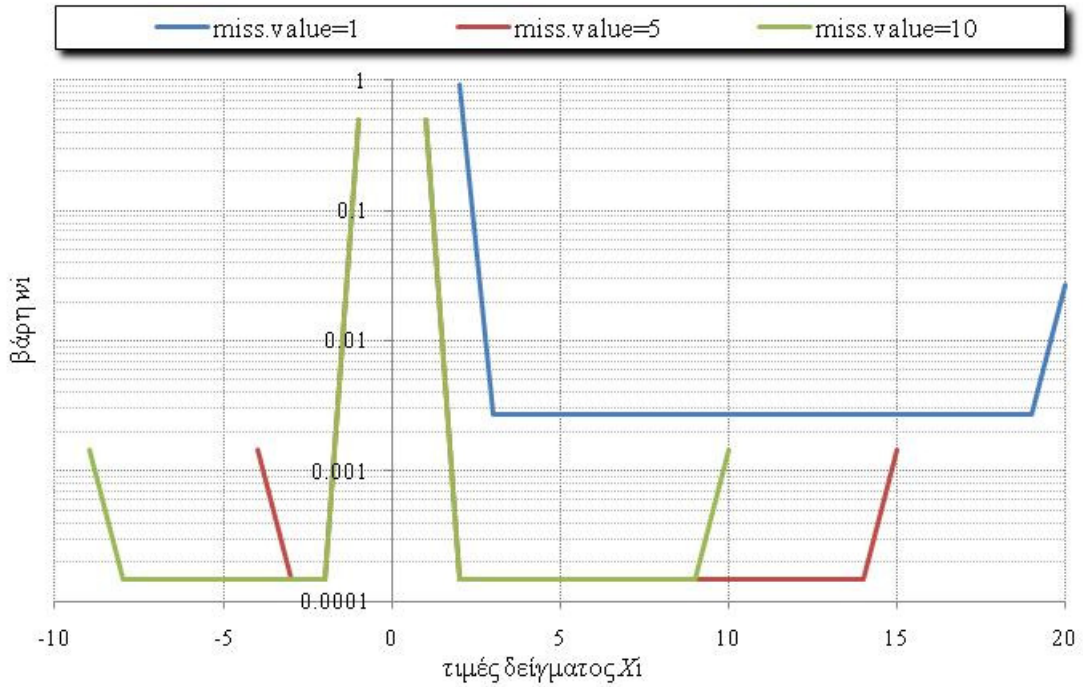

Διάγραμμα 0.2 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.3.




Διάγραμμα 0.3 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.6.

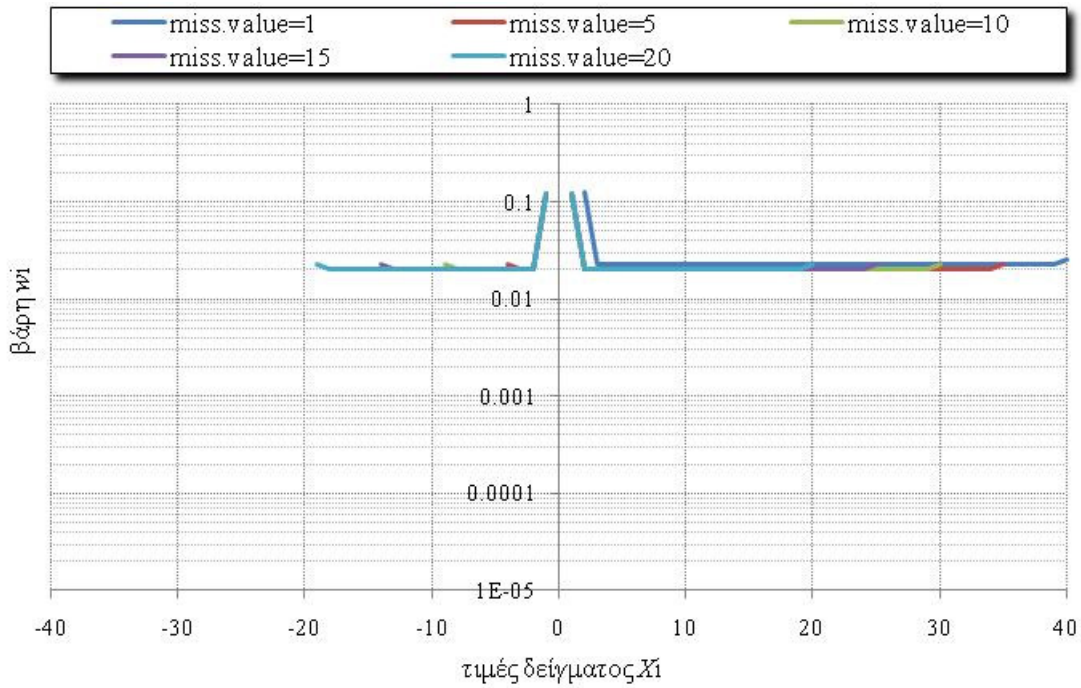



Διάγραμμα 0.4 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.7.

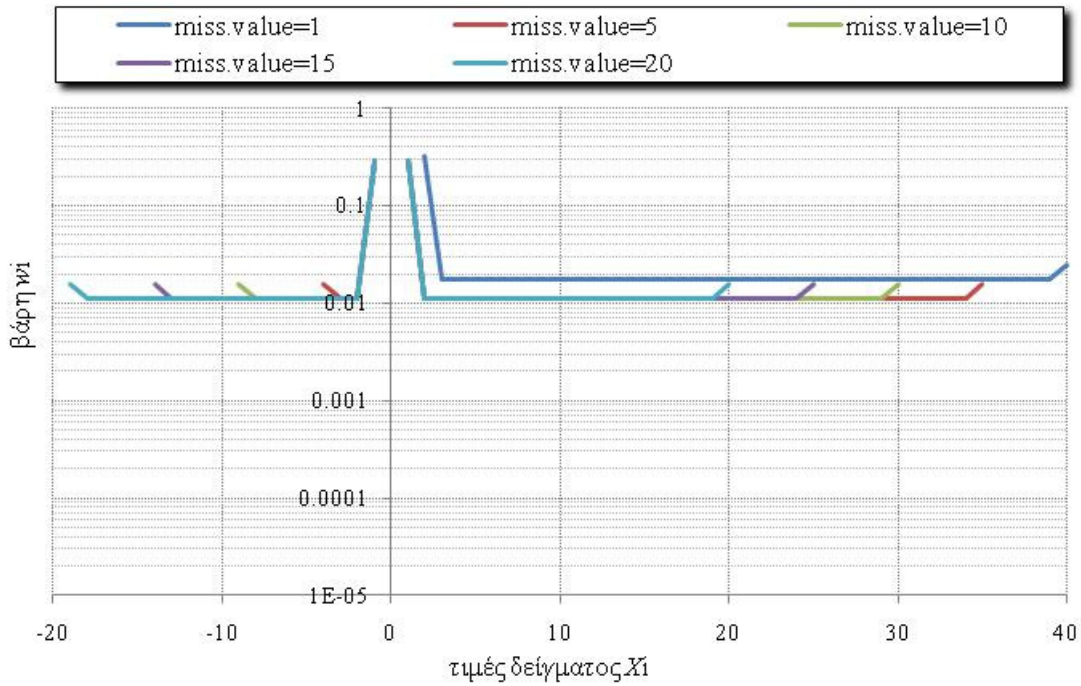



Διάγραμμα 0.5 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.9.

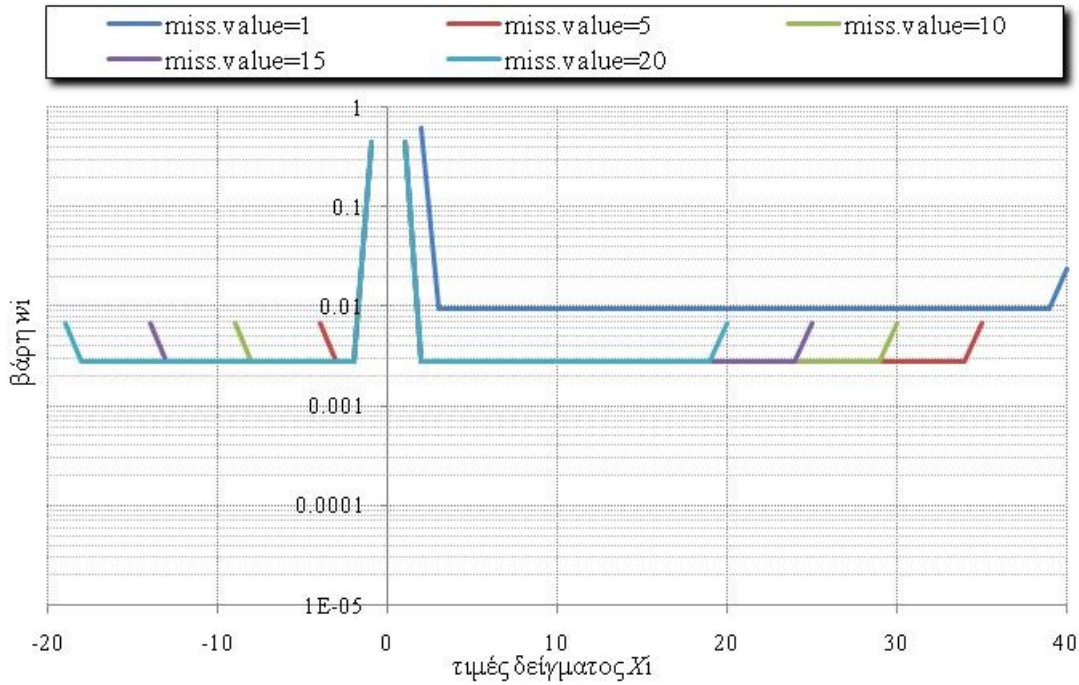
- Για δείγμα μεγέθους 40 τιμών έχουμε:



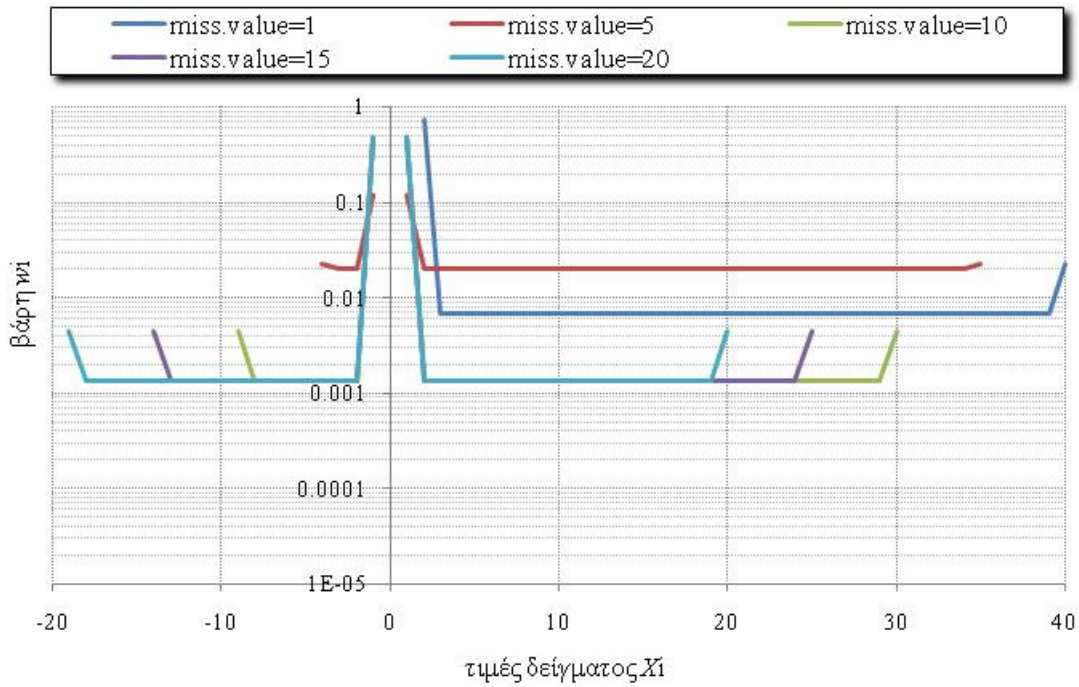

Διάγραμμα 0.6 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.1.



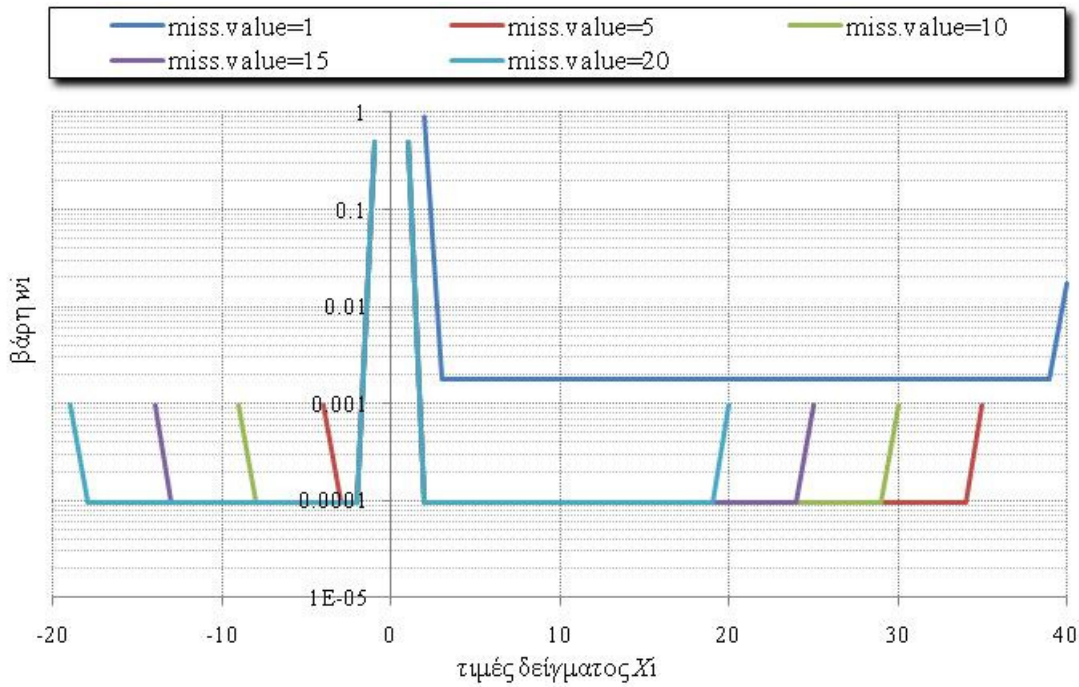

Διάγραμμα 0.7 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.3.




Διάγραμμα 0.8 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.6.

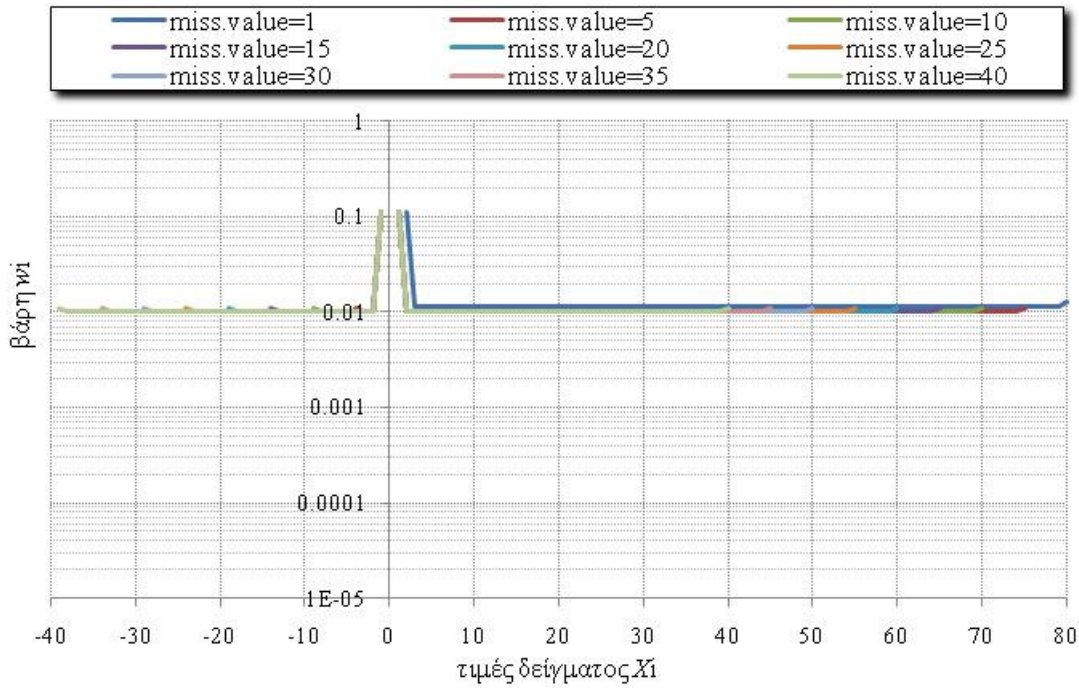



Διάγραμμα 0.9 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.7.

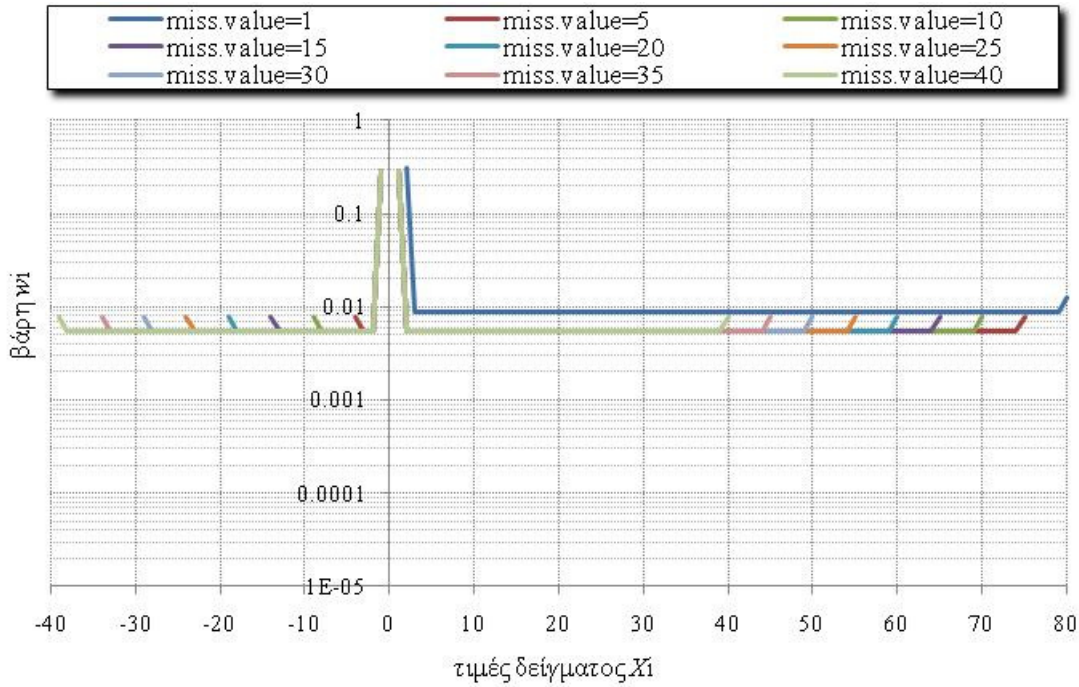



Διάγραμμα 0.10 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.9.

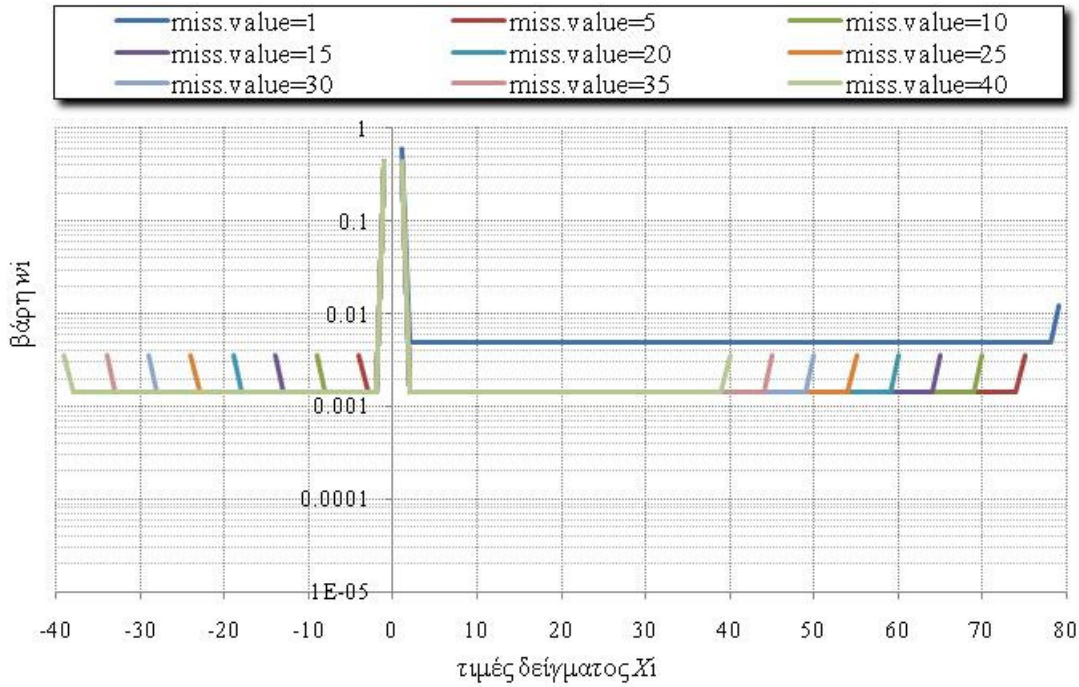
- Για δείγμα μεγέθους 80 τιμών έχουμε:



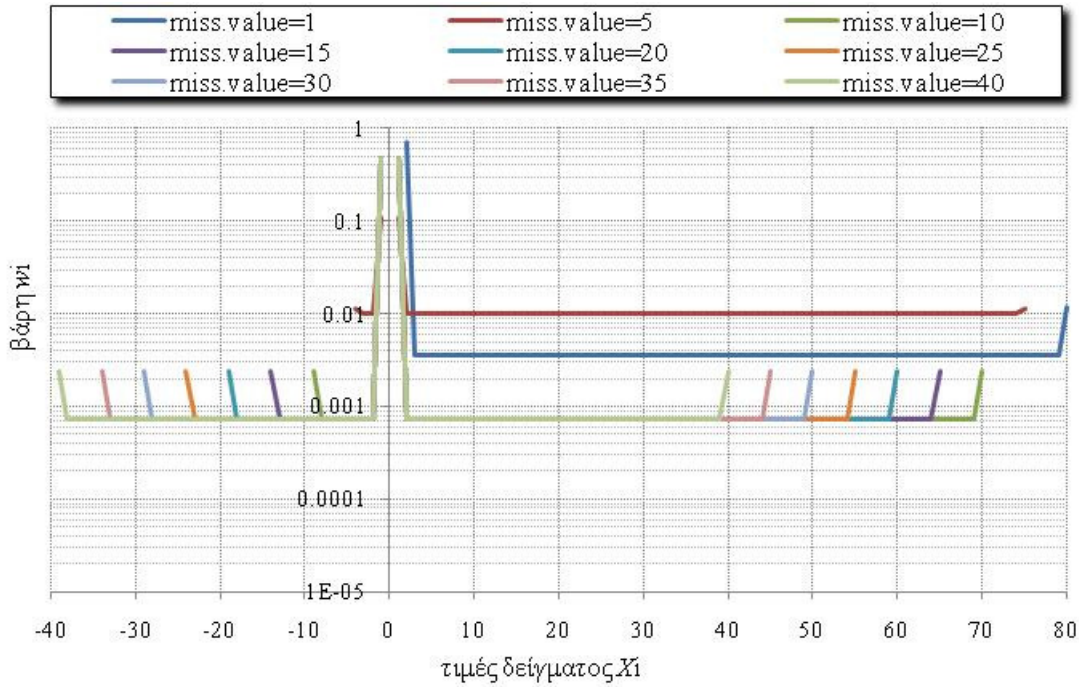

Διάγραμμα 0.11 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.1.



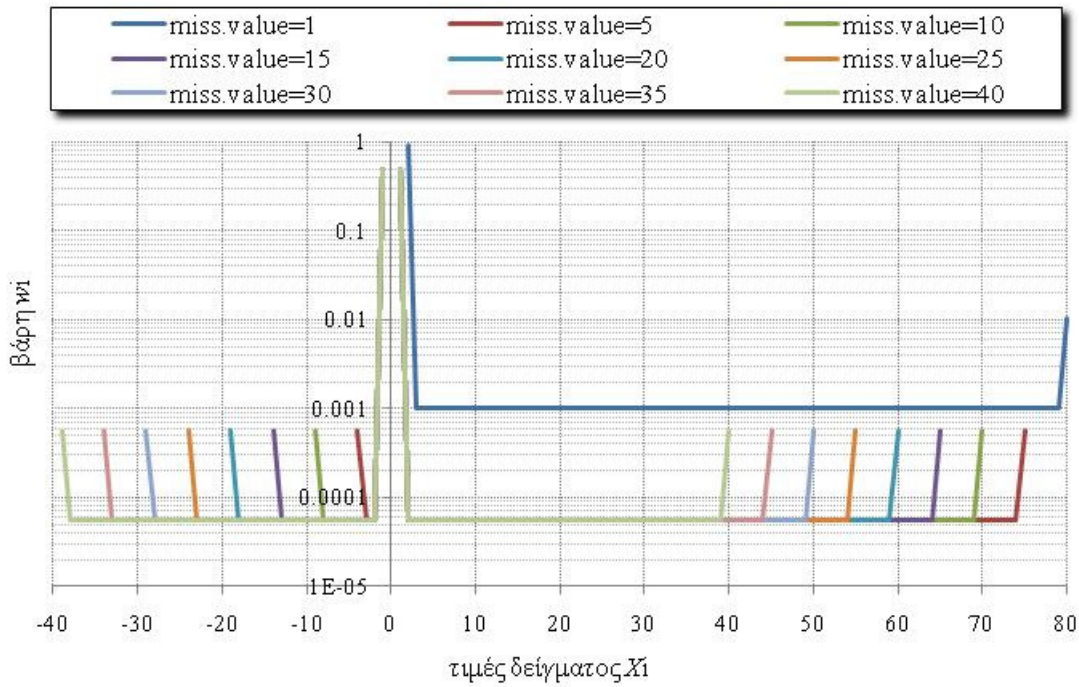

Διάγραμμα 0.12 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.3.




Διάγραμμα 0.13 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.6.

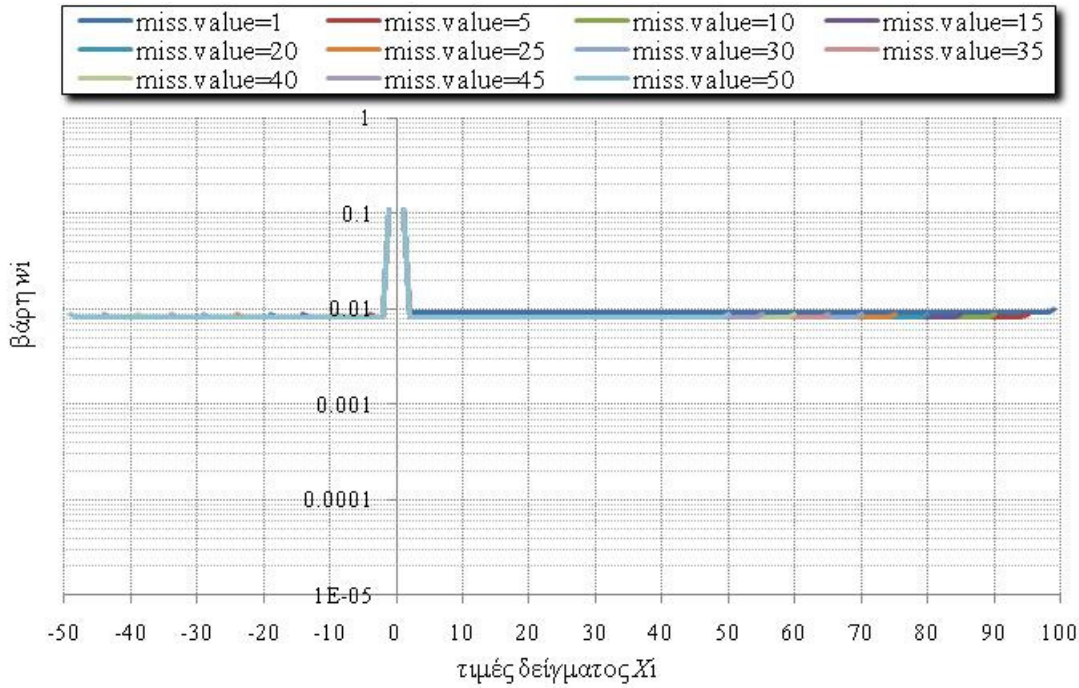



Διάγραμμα 0.14 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.7.

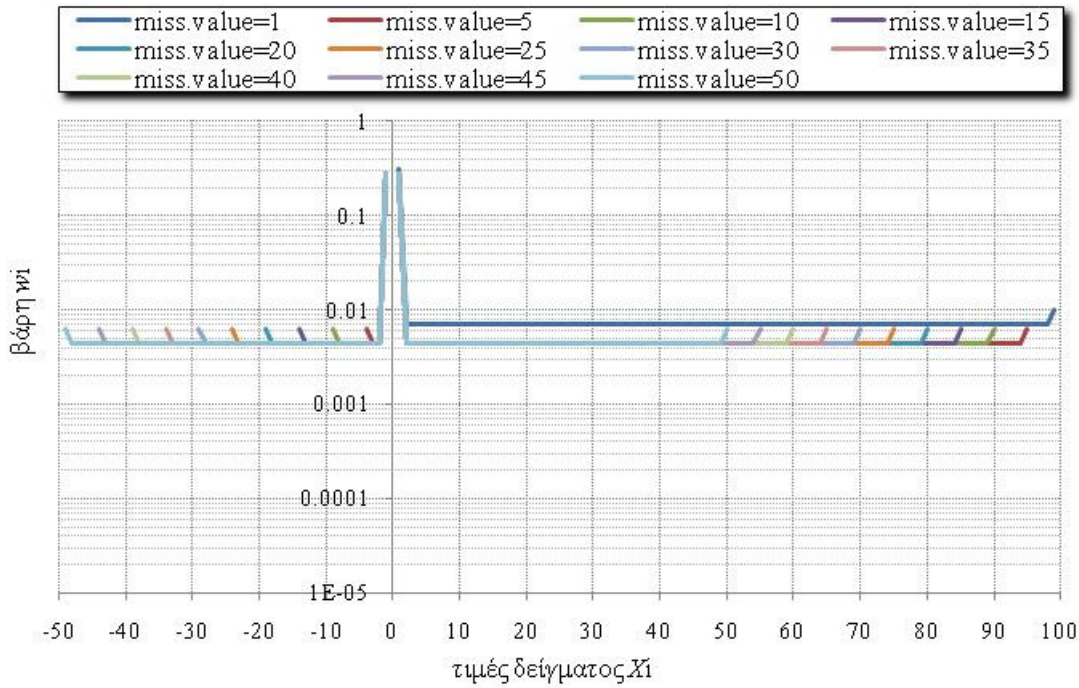



Διάγραμμα 0.15 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.9.

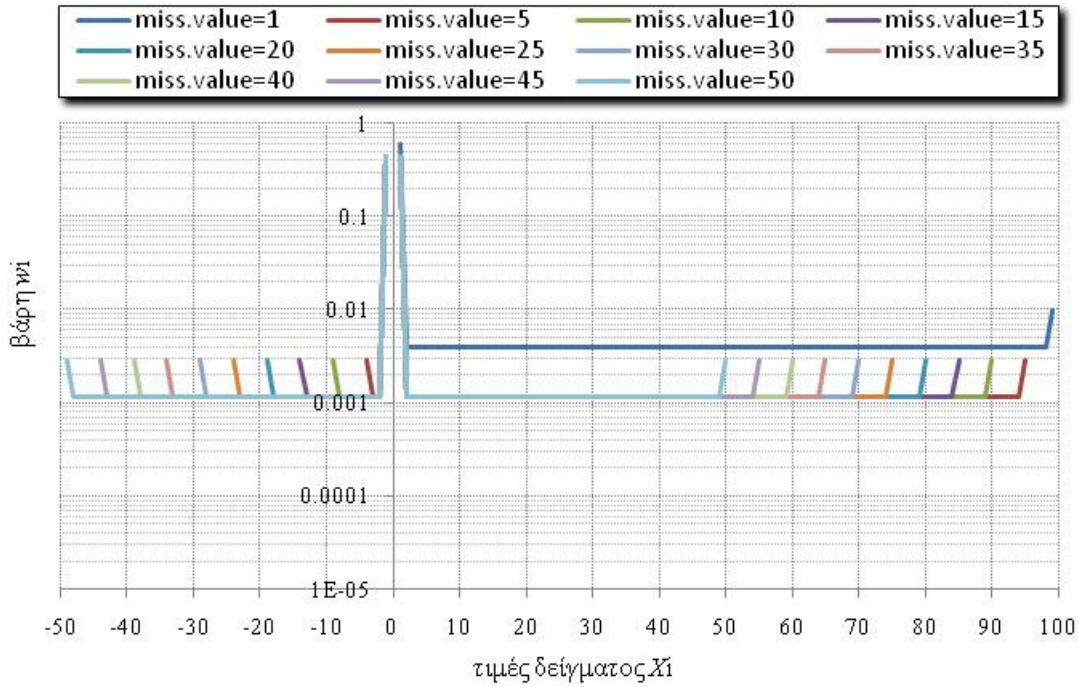
- Για δείγμα μεγέθους 100 τιμών έχουμε:



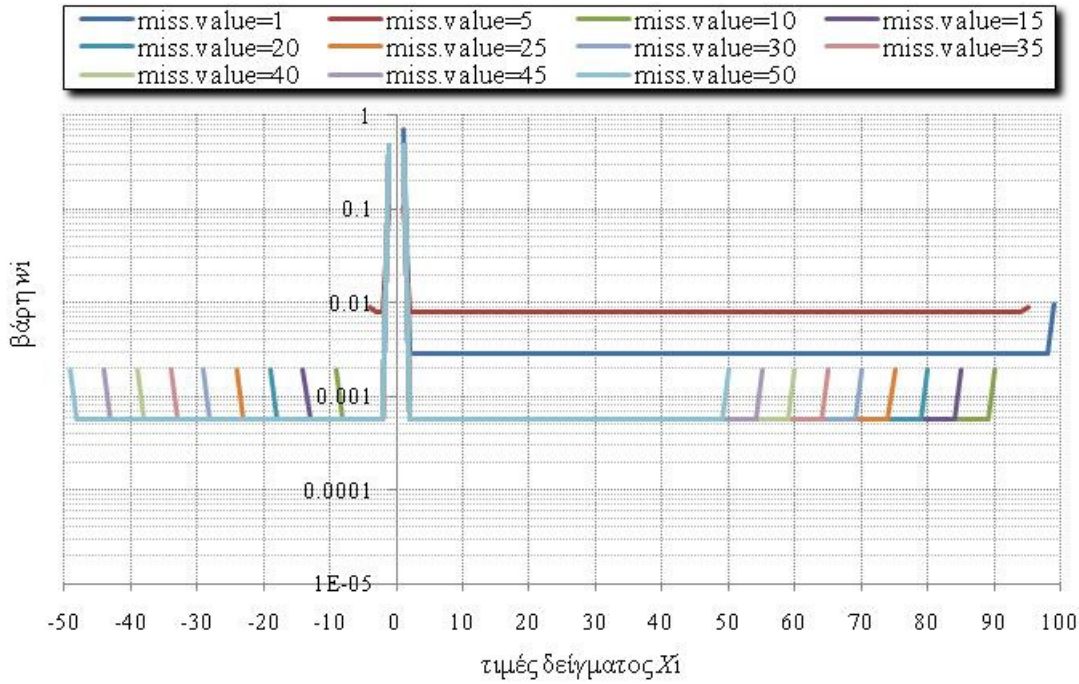

Διάγραμμα 0.16 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.1.



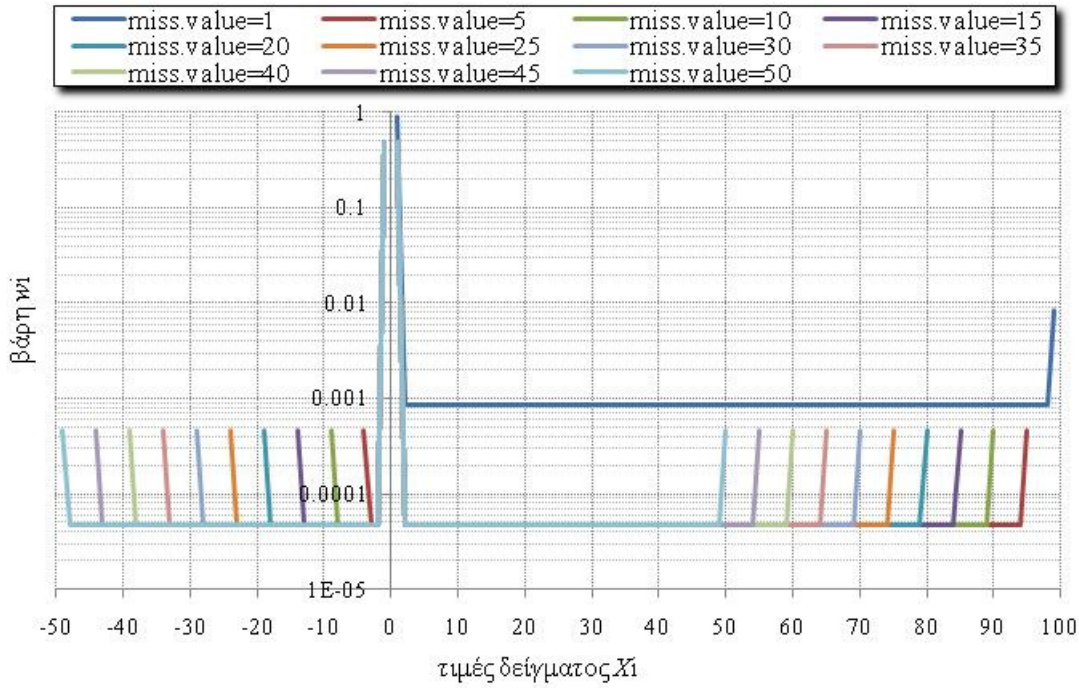

Διάγραμμα 0.17 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.3.




Διάγραμμα 0.18 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.6.




Διάγραμμα 0.19 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.7.



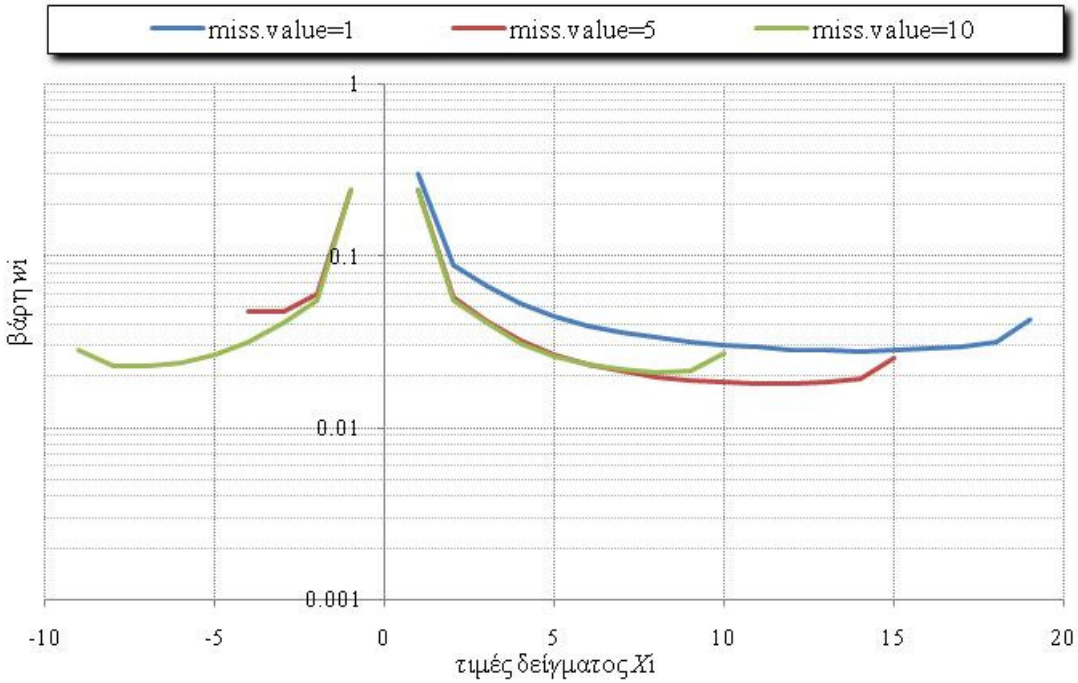

Διάγραμμα 0.20 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή αυτοσυσχέτισης $\rho_1$=0.9.

## ΠΑΡΑΡΤΗΜΑ Ε

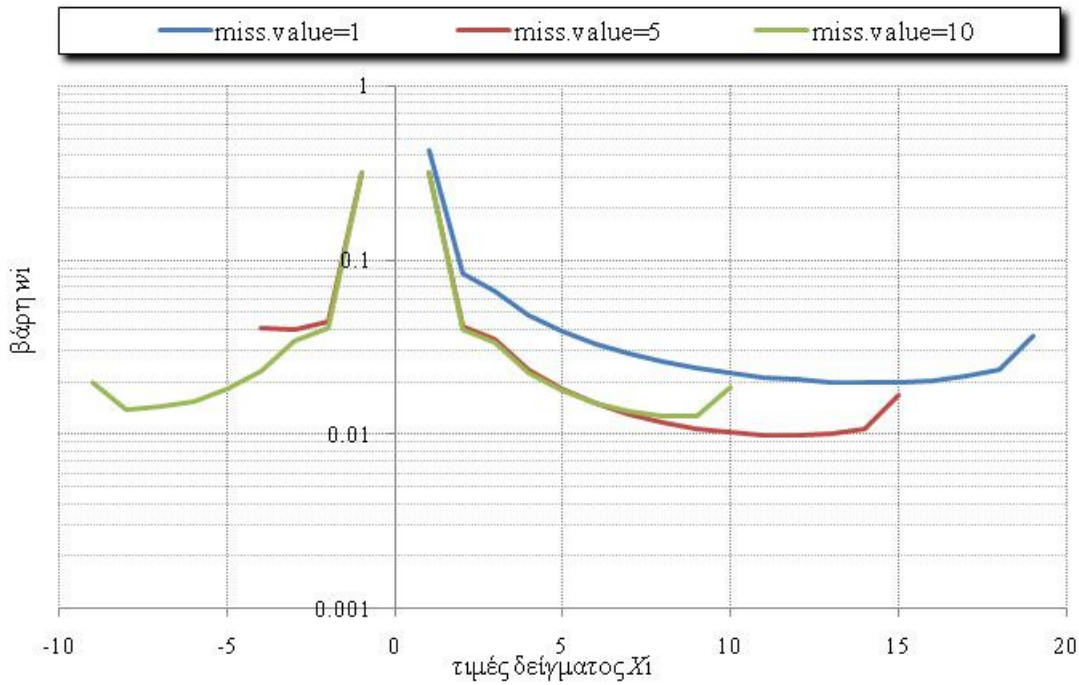
Συμπλήρωση μεμονωμένων ελλειπουσών τιμών με σταθμισμένα βάρη $w_i$ σε υδρομετεωρολογικές χρονοσειρές που παρουσιάζουν το φαινόμενο Hurst.

Ανάλογα με το συντελεστή Hurst ($H$) κατασκευάστηκαν τα πιο κάτω διαγράμματα που παρουσιάζουν τα σταθμισμένα βάρη των τιμών της χρονοσειράς, ανάλογα με την θέση της ελλείπουσας τιμής.

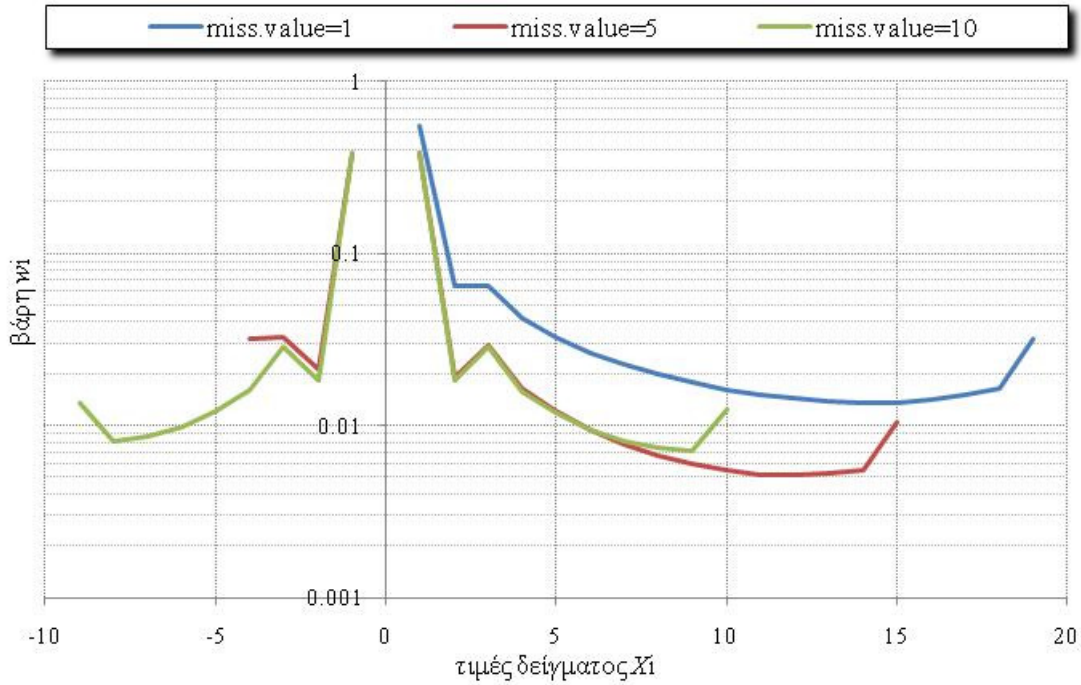
- Για δείγμα μεγέθους 20 τιμών έχουμε:




Διάγραμμα 0.21 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.7.

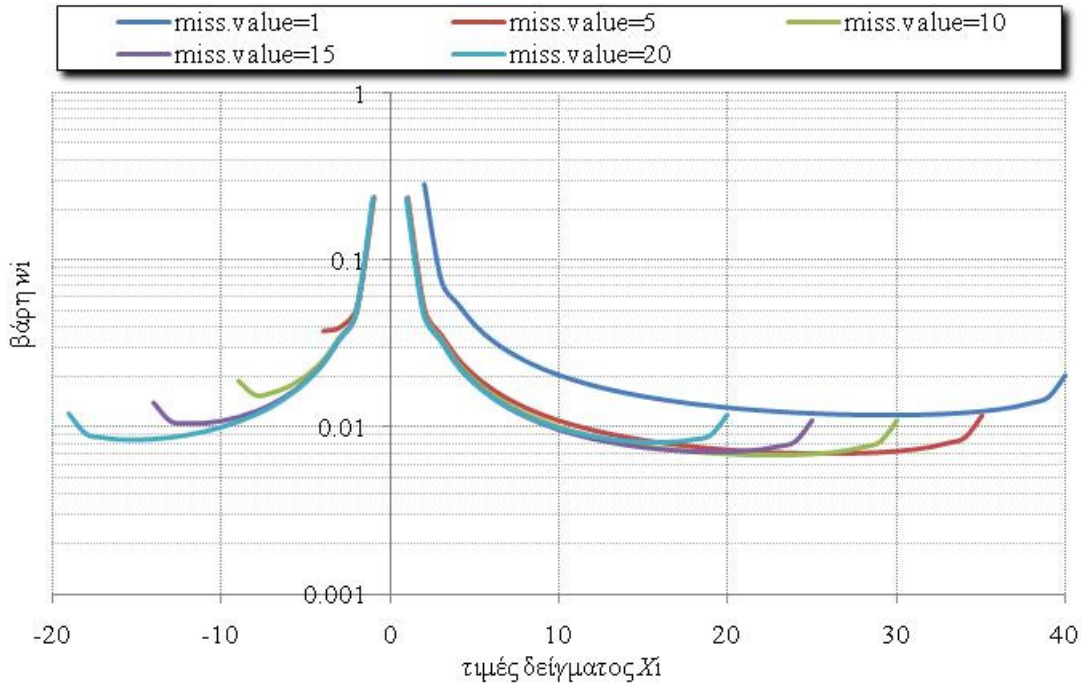



Διάγραμμα 0.22 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.8.

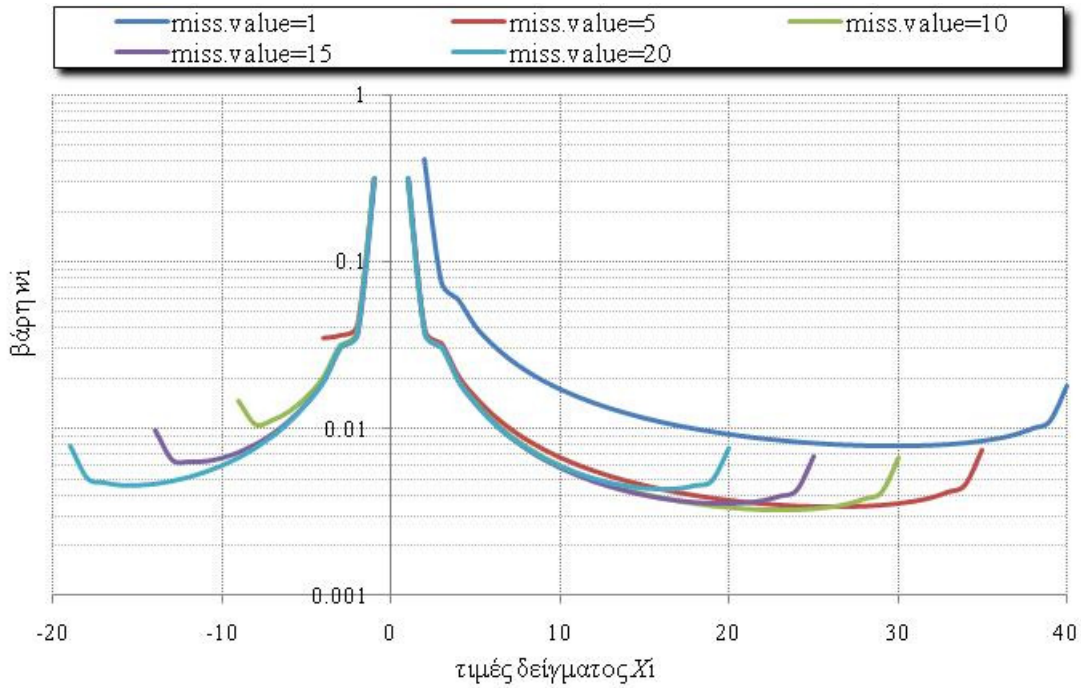



Διάγραμμα 0.23 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.9.

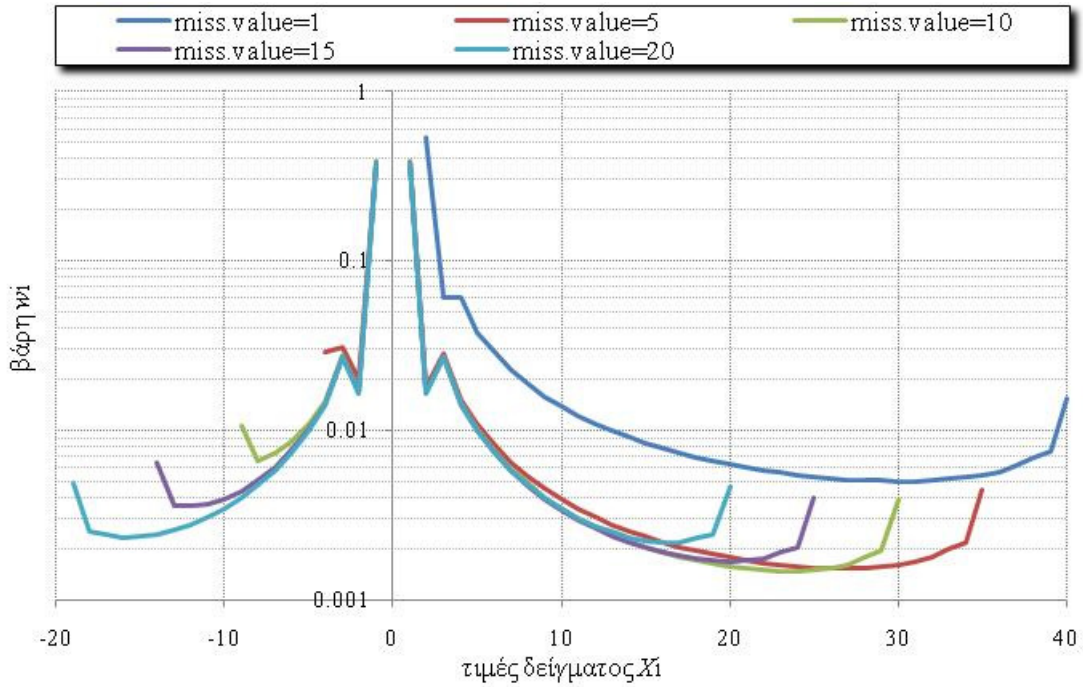
- Για δείγμα μεγέθους 40 τιμών έχουμε:




Διάγραμμα 0.24 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.7.

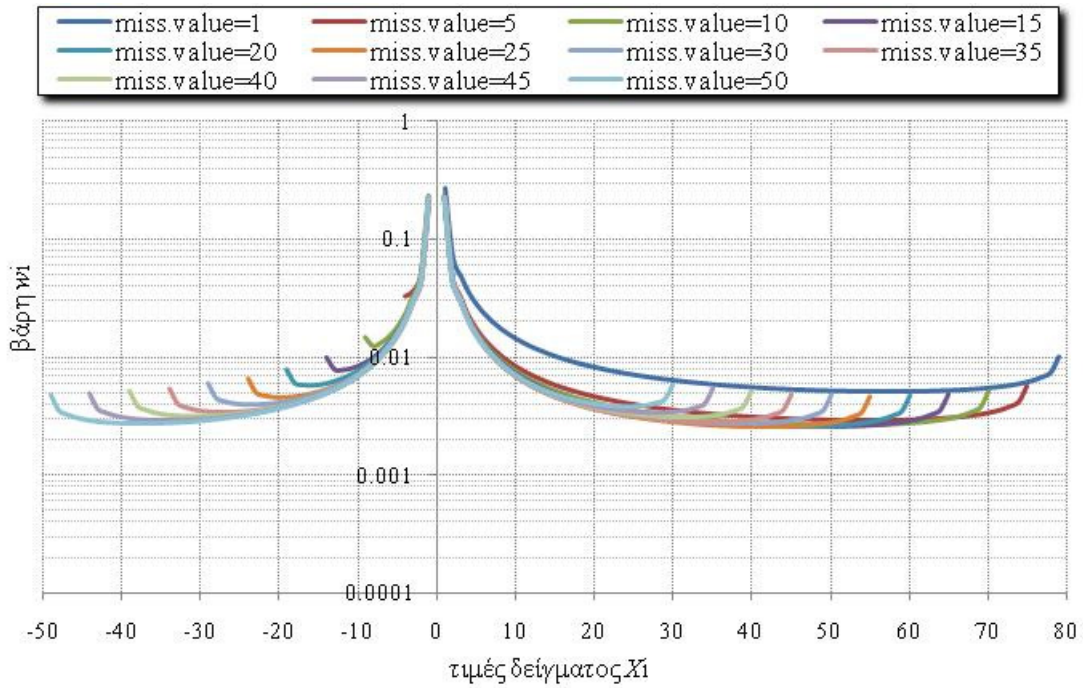



Διάγραμμα 0.25 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.8.

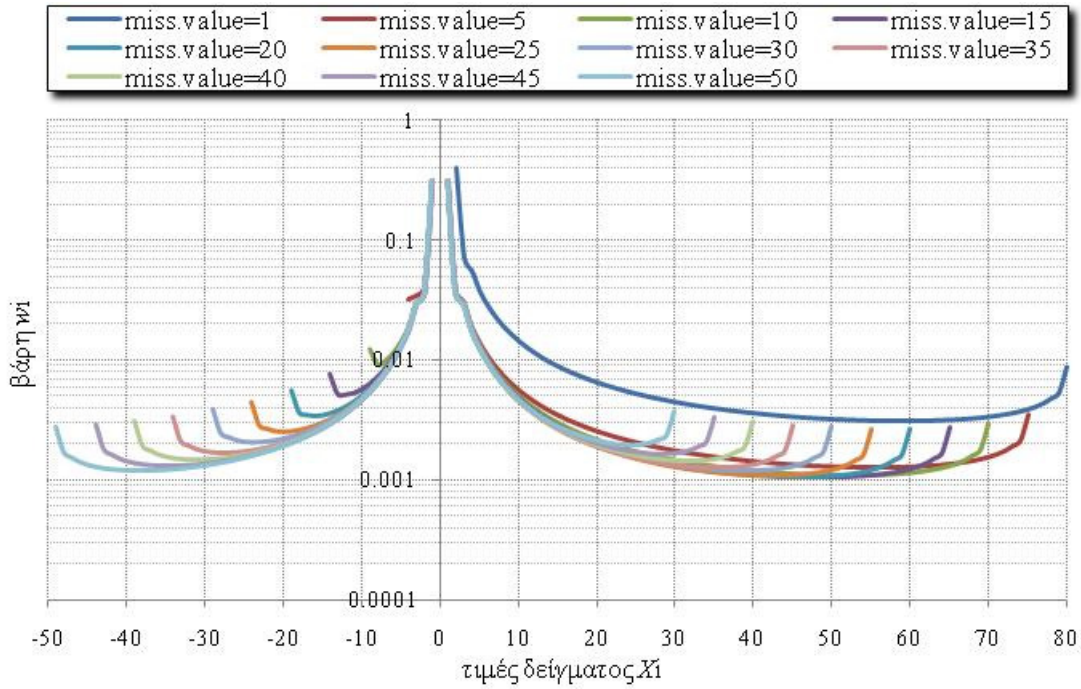



Διάγραμμα 0.26 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.9.

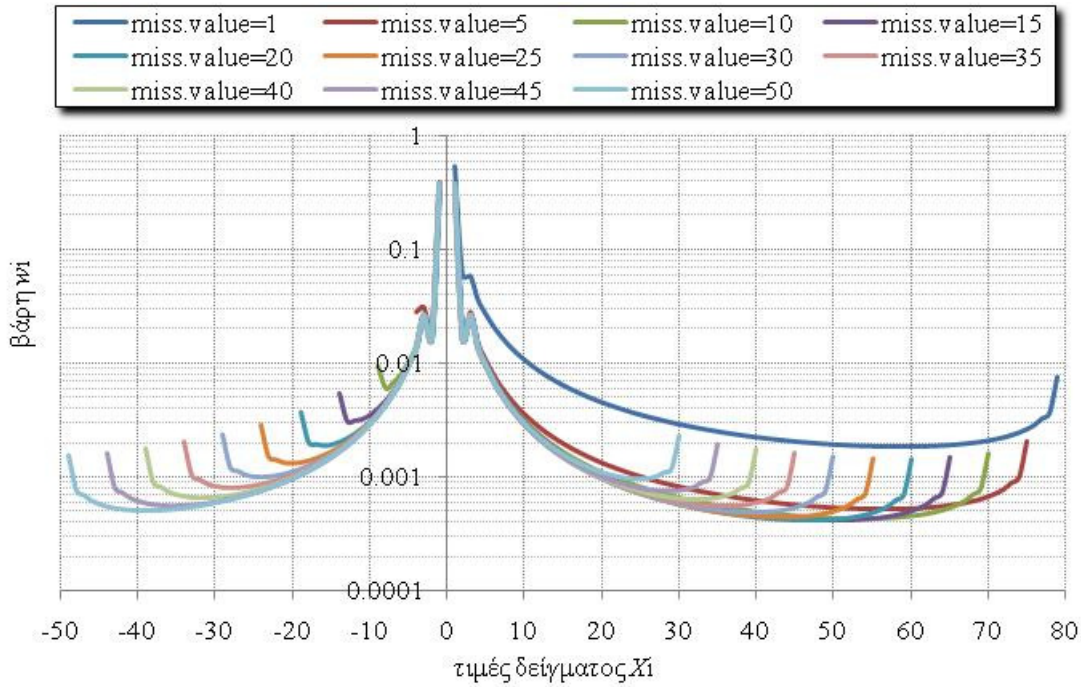
- Για δείγμα μεγέθους 80 τιμών έχουμε:




Διάγραμμα 0.27 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.7.

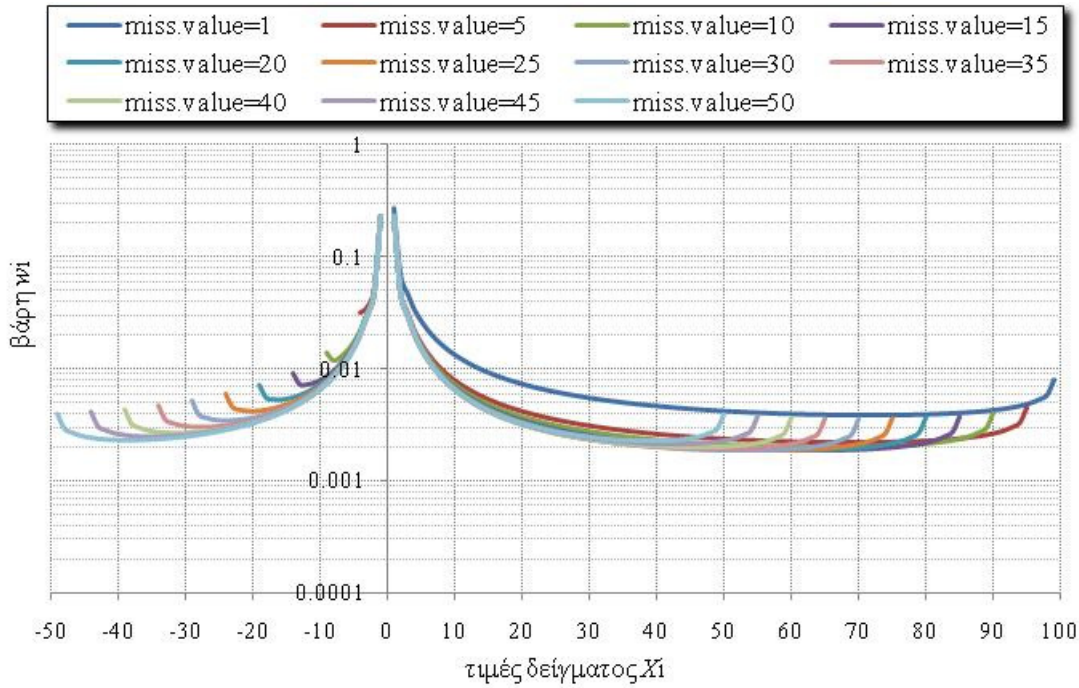



Διάγραμμα 0.28 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.8.

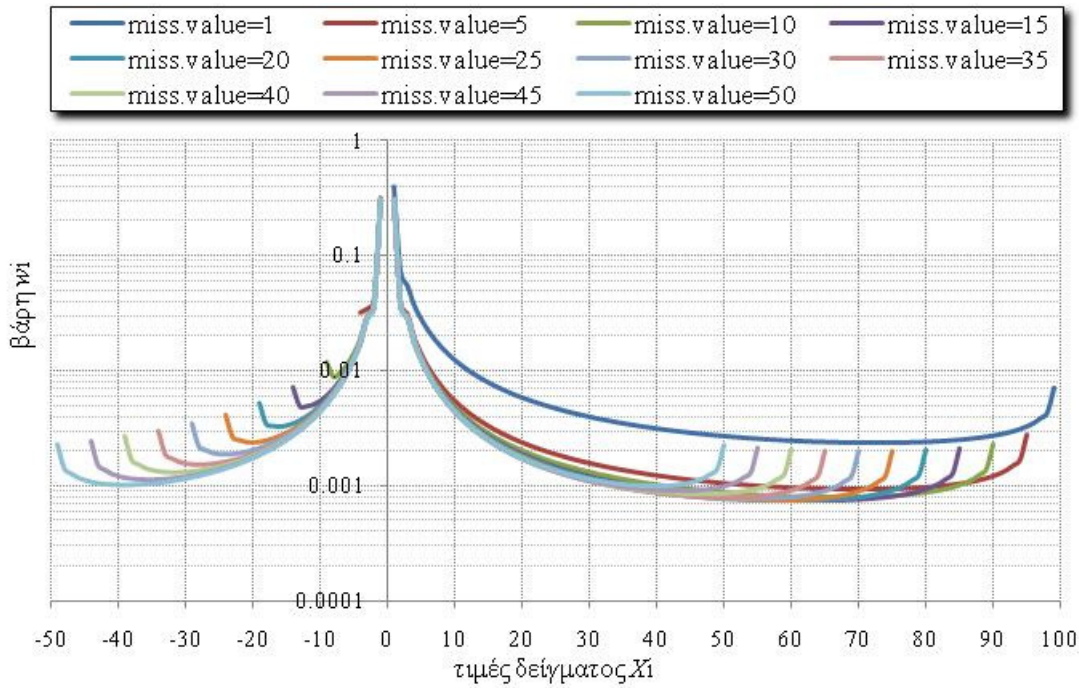



Διάγραμμα 0.29 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.9.

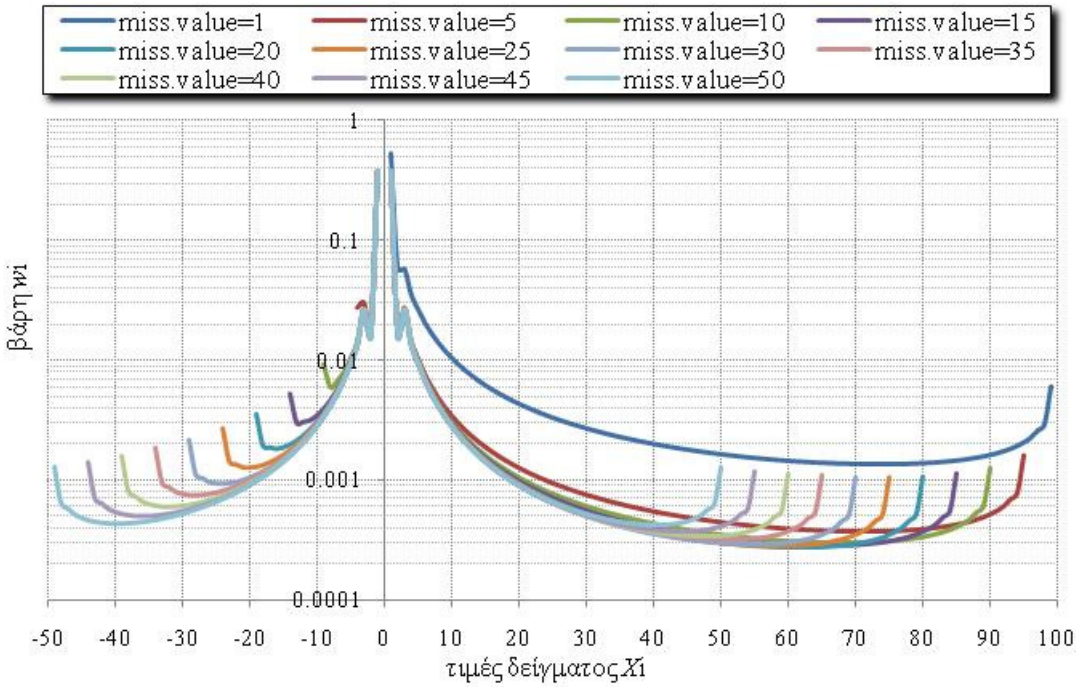
- Για δείγμα μεγέθους 100 τιμών έχουμε:




Διάγραμμα 0.30 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.7.




Διάγραμμα 0.31 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.8.



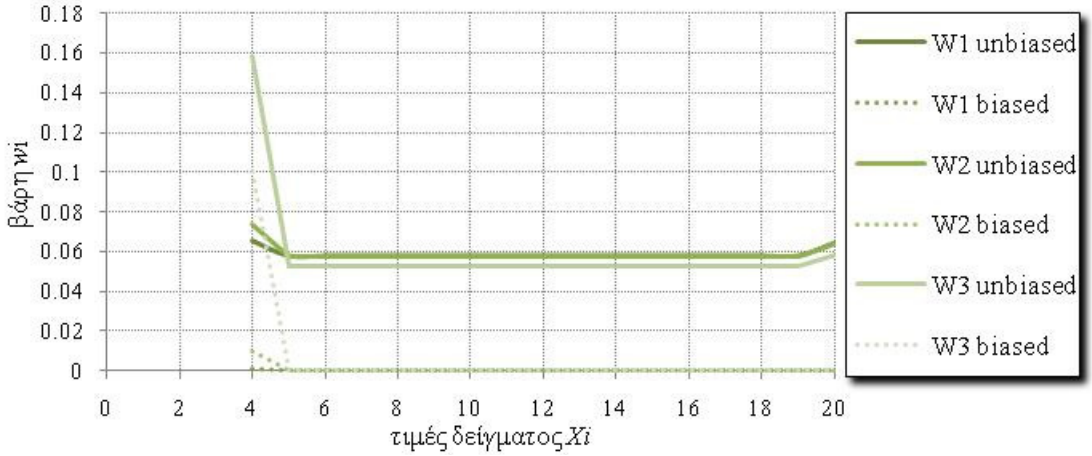

Διάγραμμα 0.32 Σταθμισμένα βάρη $w_i$ για τις διάφορες τιμές του δείγματος, ανάλογα με την θέση της ελλείπουσας τιμής, για συντελεστή Hurst $H$=0.9.

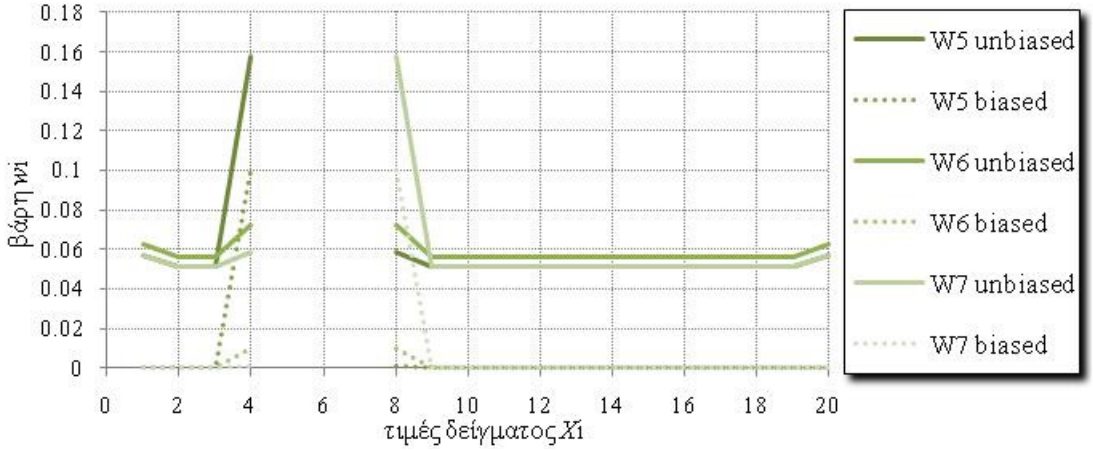
## ΠΑΡΑΡΤΗΜΑ F

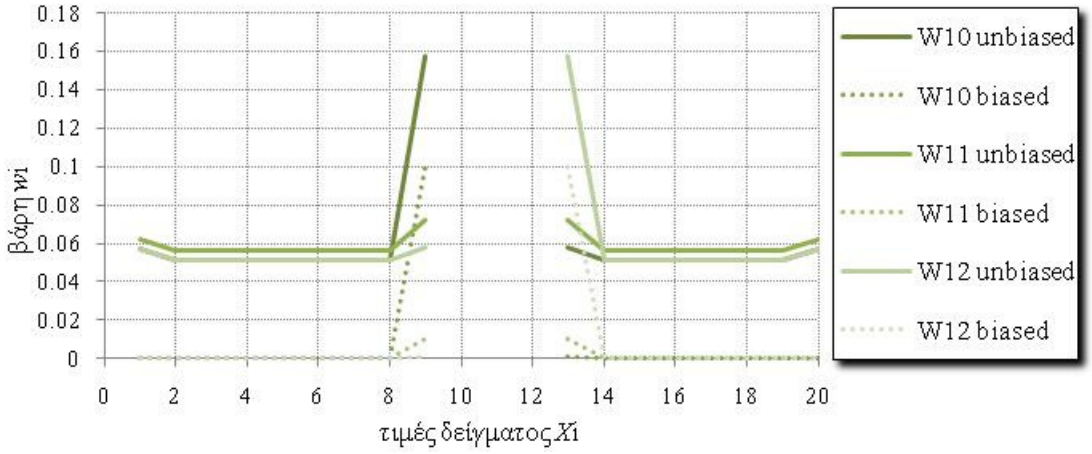
Συμπλήρωση 3 συνεχόμενων ελλειπουσών τιμών με σταθμισμένα βάρη $w_i$ σε υδρομετεωρολογικές χρονοσειρές τύπου Markov.

Ανάλογα με το συντελεστή αυτοσυσχέτισης για υστέρηση 1 ($\rho_1$) κατασκευάστηκαν τα πιο κάτω διαγράμματα που παρουσιάζουν τα σταθμισμένα βάρη των τιμών της χρονοσειράς, ανάλογα με την θέση της ελλείπουσας τιμής.

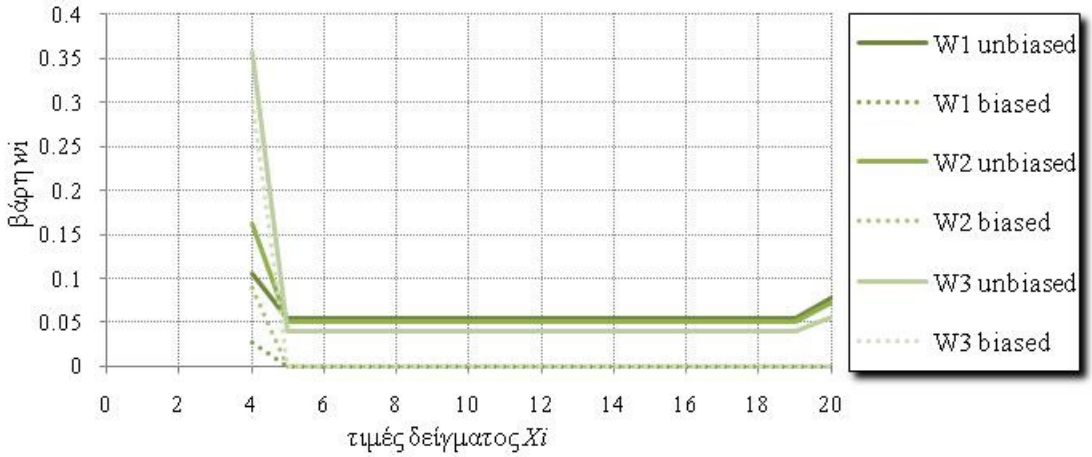
- Για δείγμα μεγέθους 20 τιμών και $\rho_1$=0.1 έχουμε:

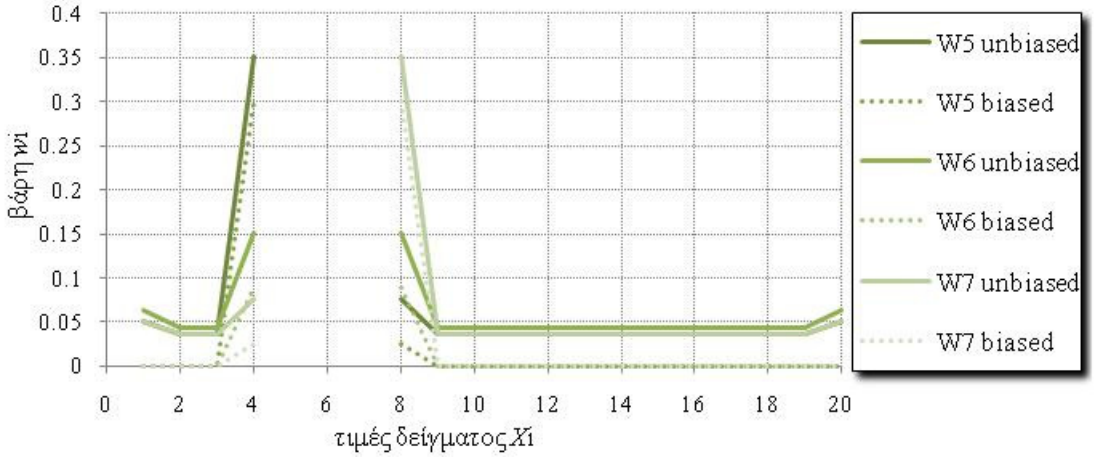


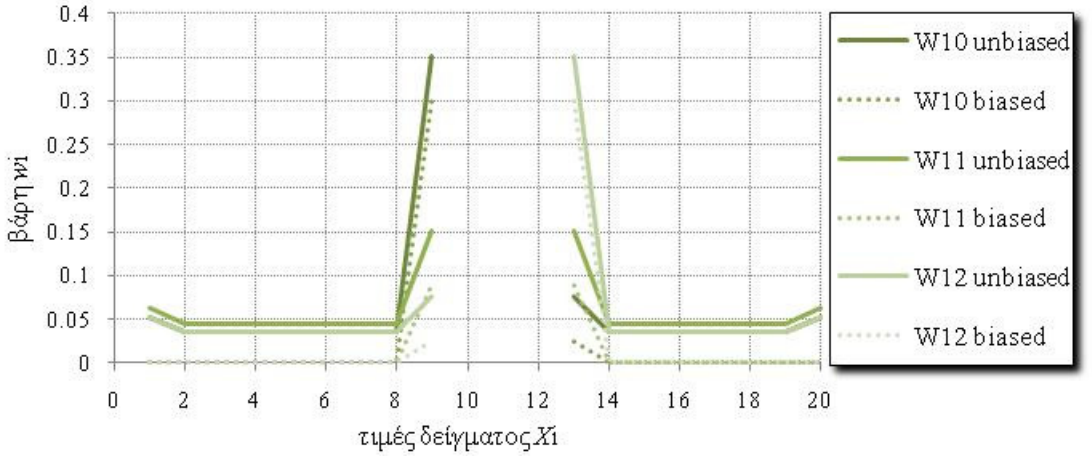



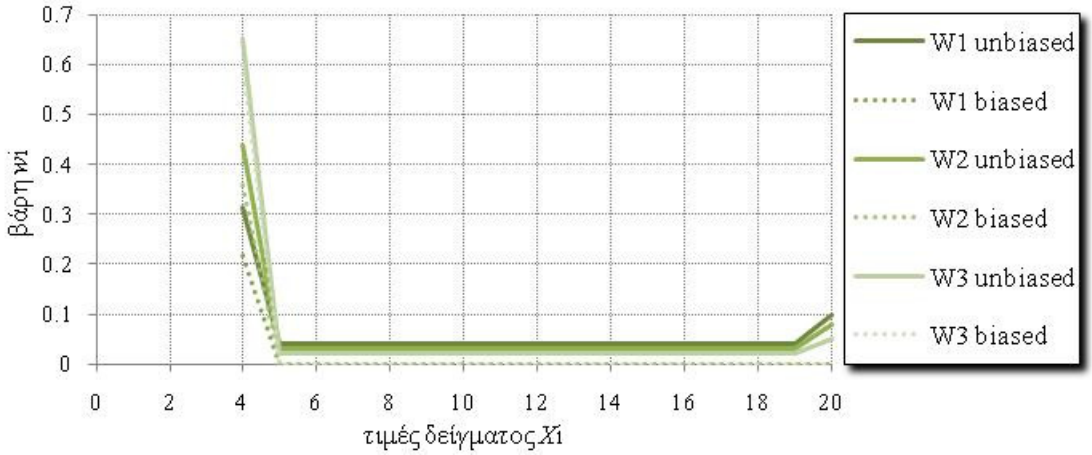
- Για δείγμα μεγέθους 20 τιμών και $\rho_1$=0.3 έχουμε:

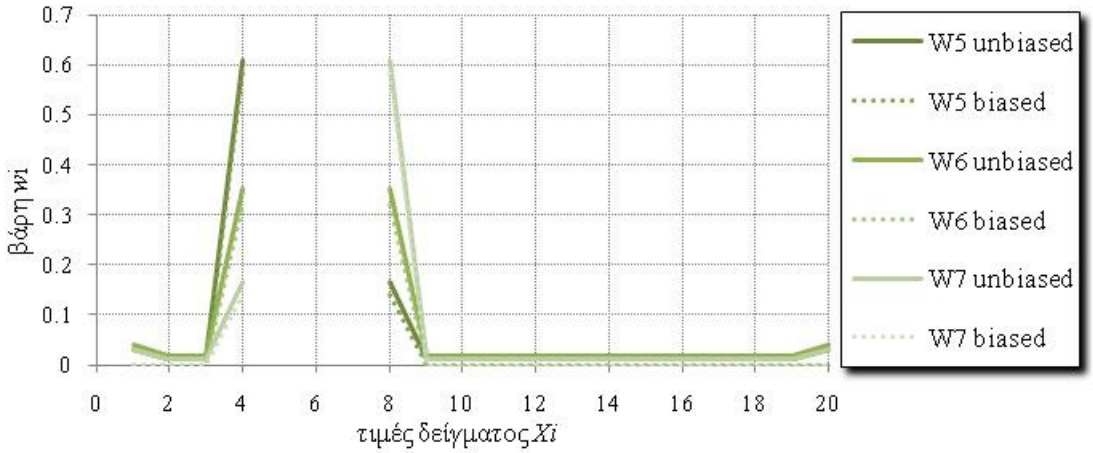


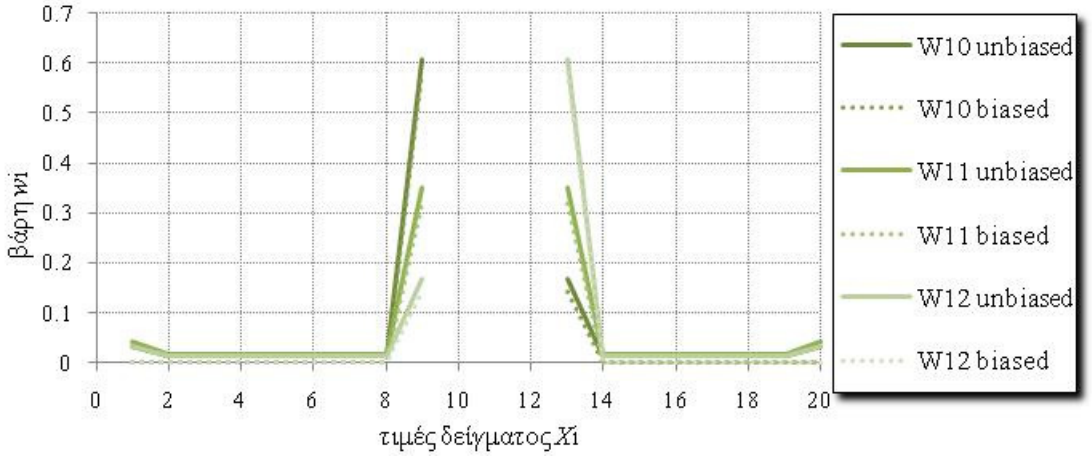



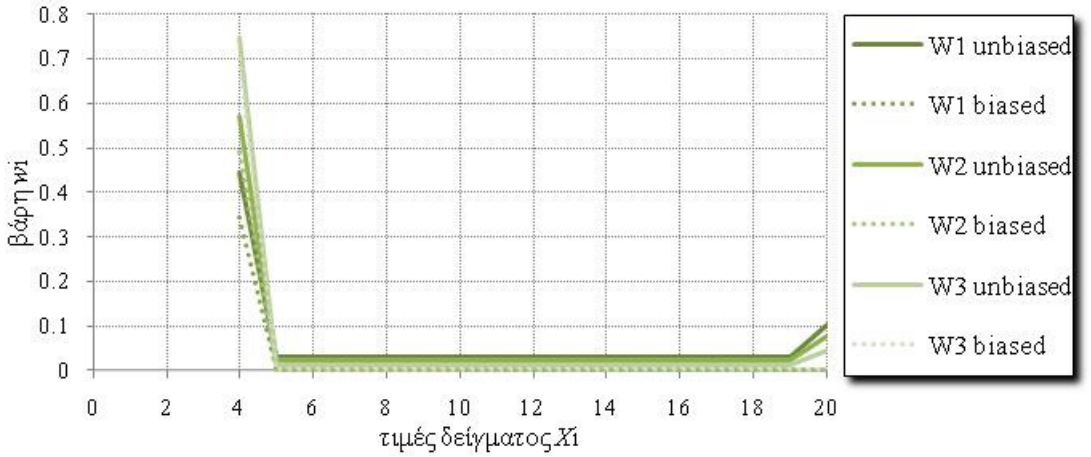
- Για δείγμα μεγέθους 20 τιμών και $\rho_1$=0.6 έχουμε:

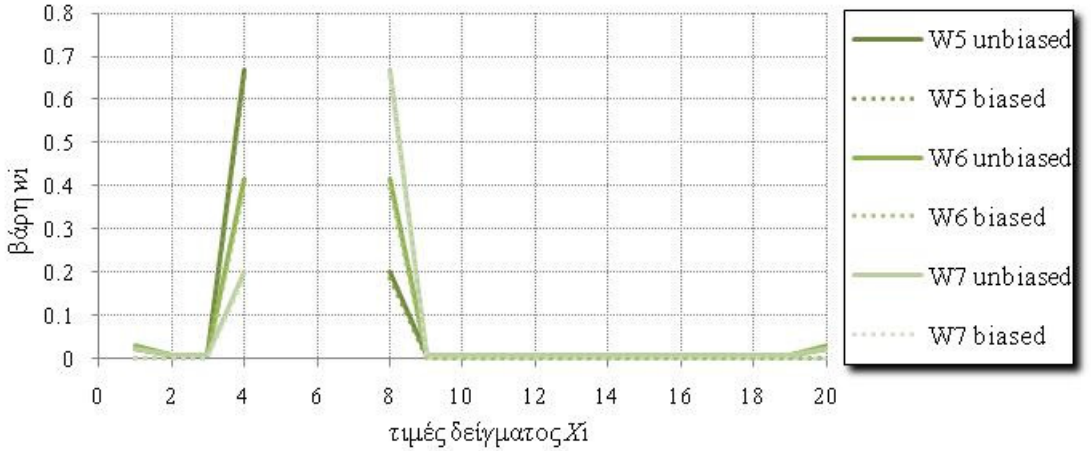


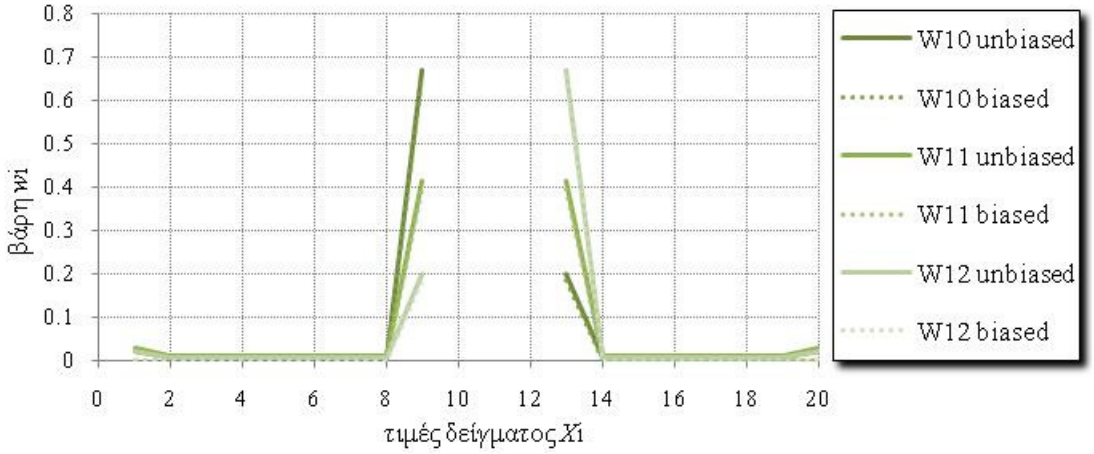



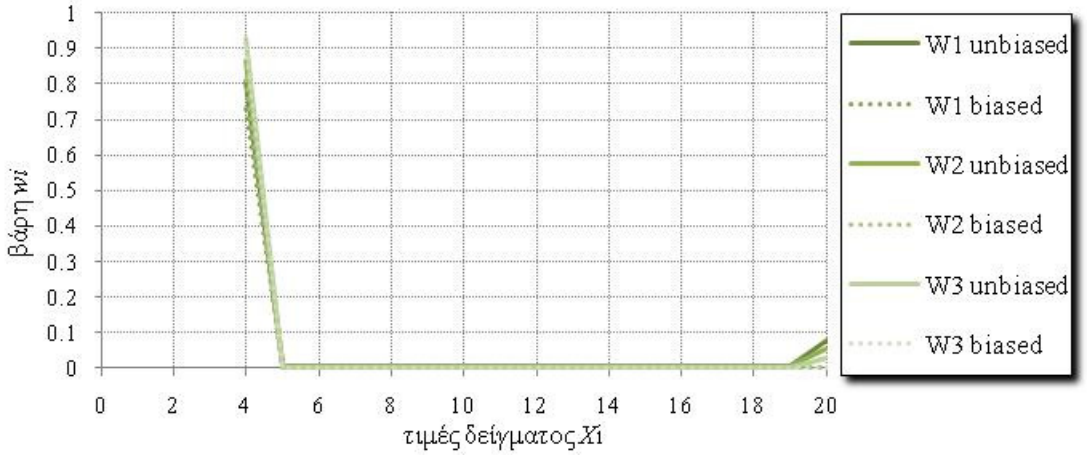
- Για δείγμα μεγέθους 20 τιμών και $\rho_1$=0.7 έχουμε:

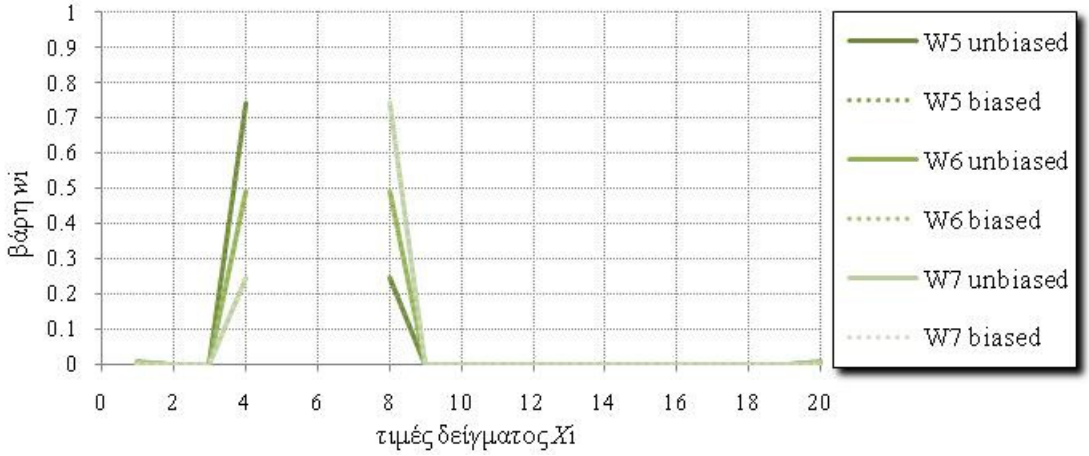


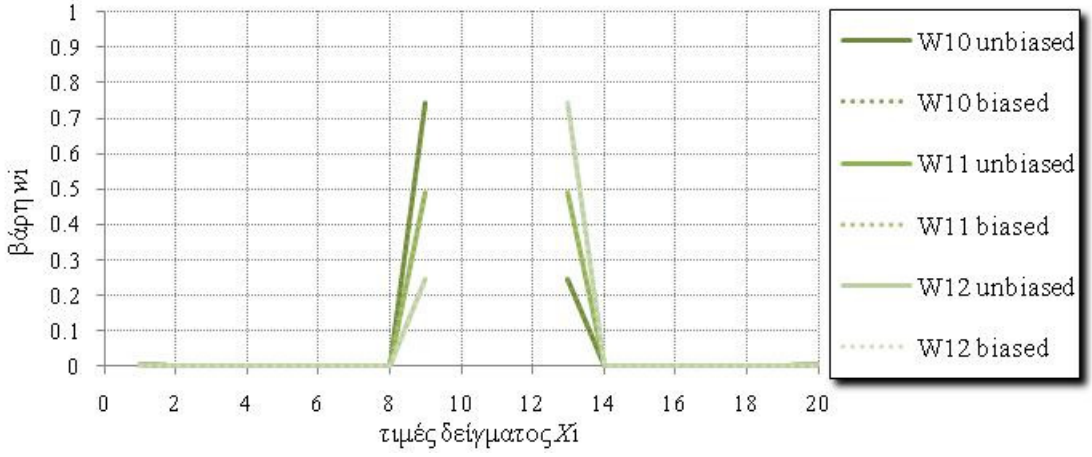



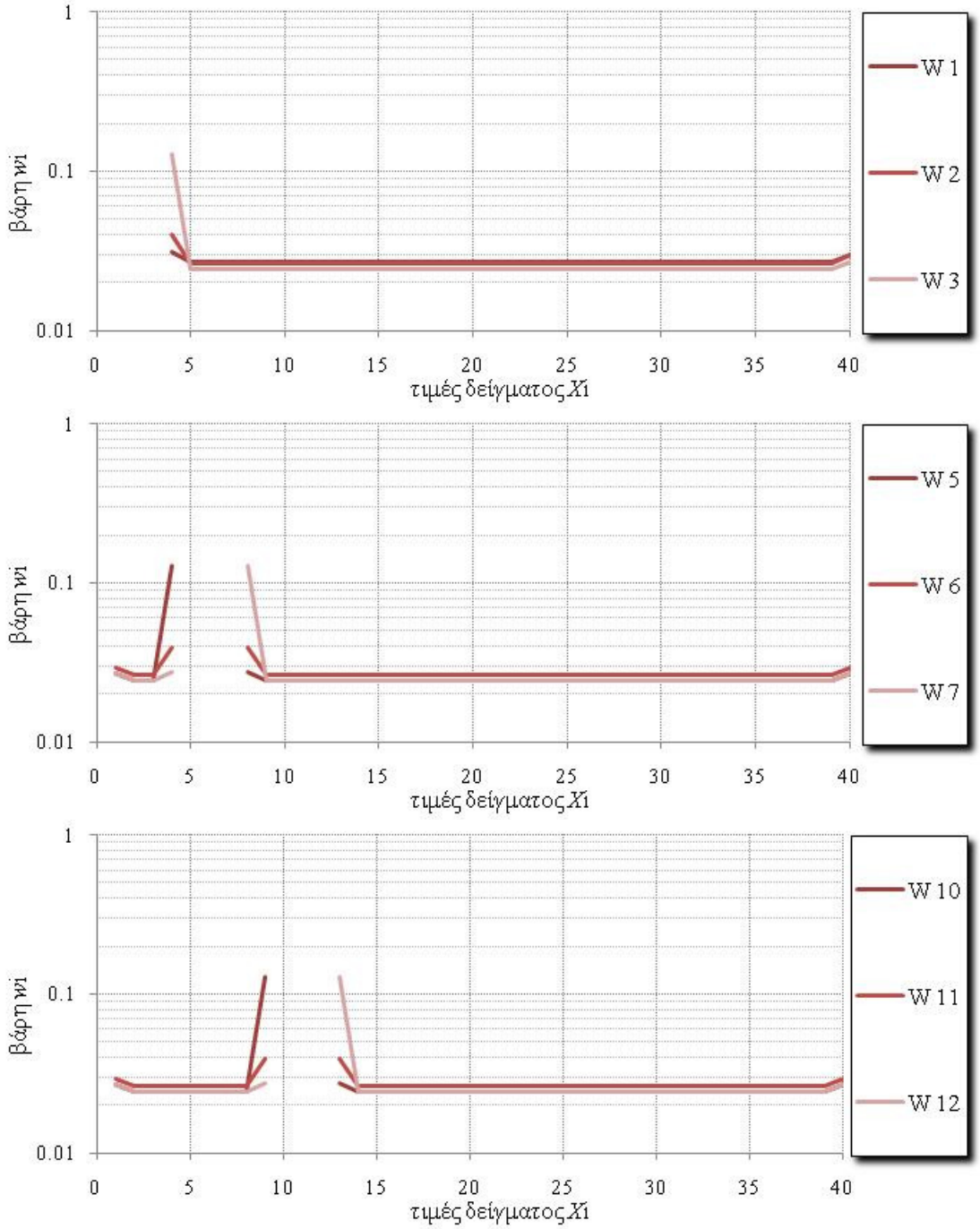
- Για δείγμα μεγέθους 20 τιμών και $\rho_1$=0.9 έχουμε:

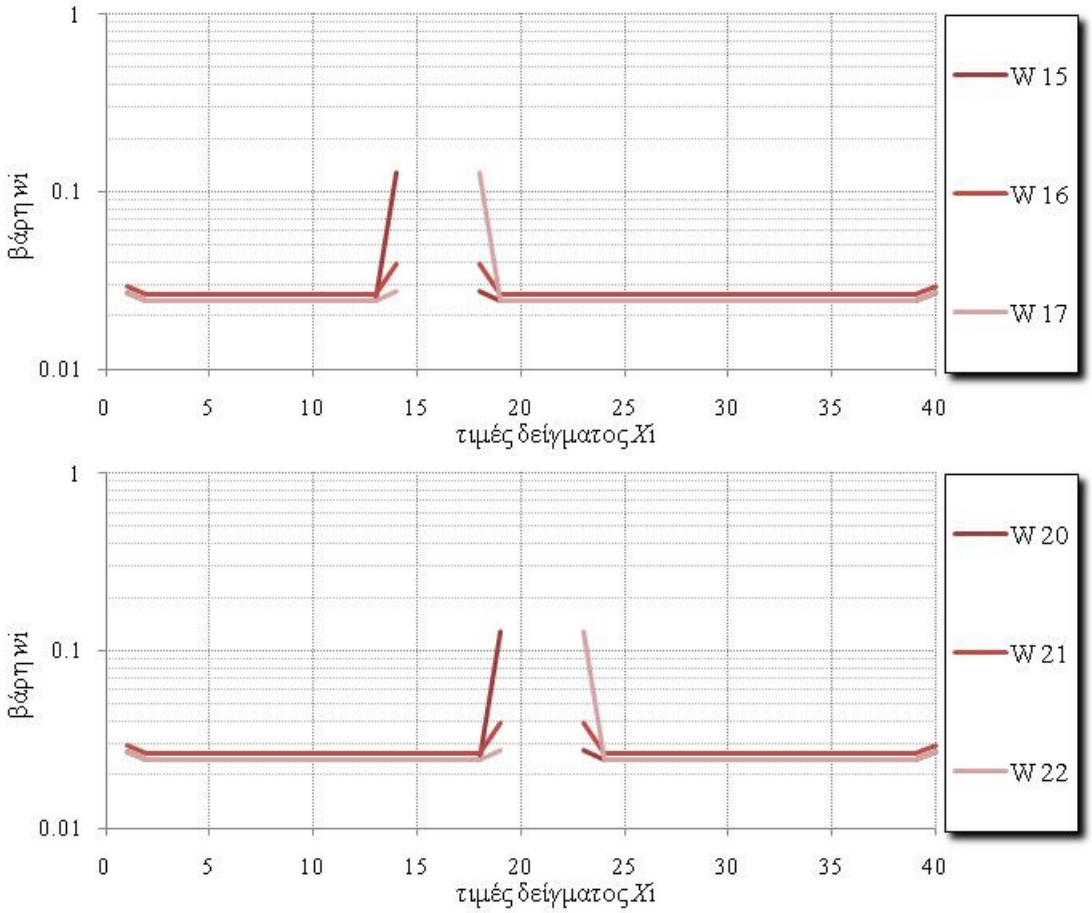


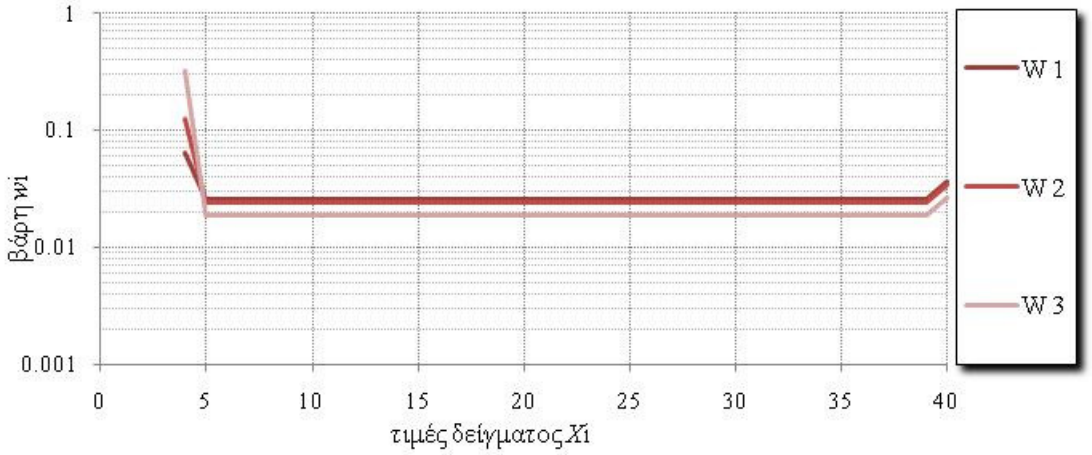



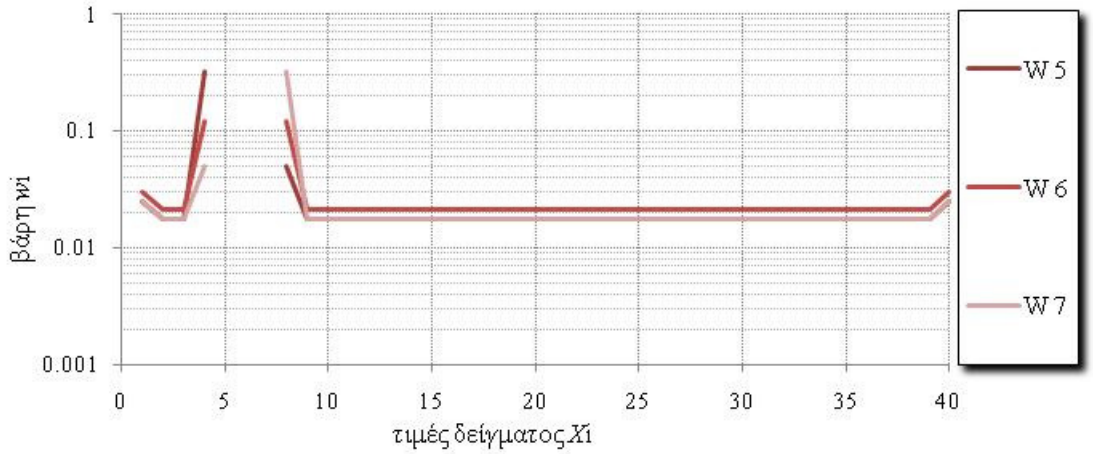
- Για δείγμα μεγέθους 40 τιμών και $\rho_1$=0.1 έχουμε:

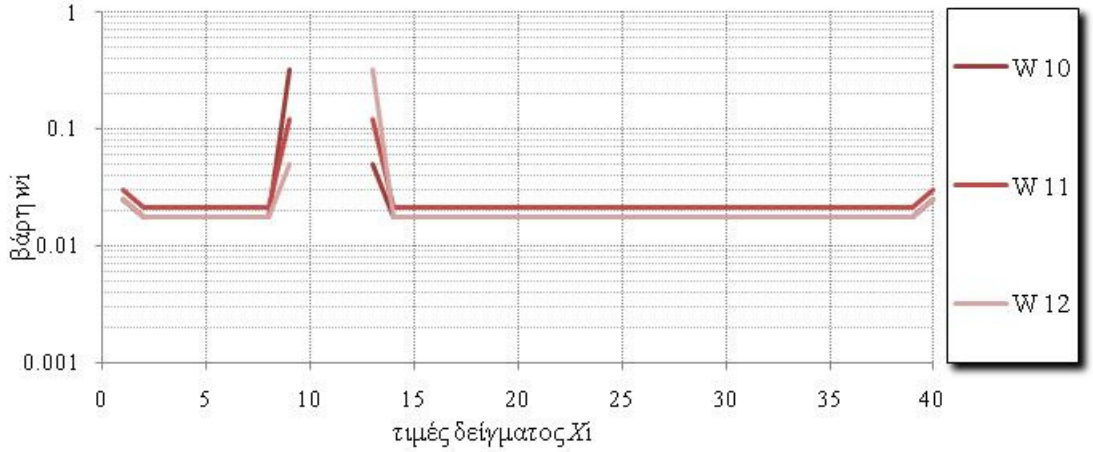


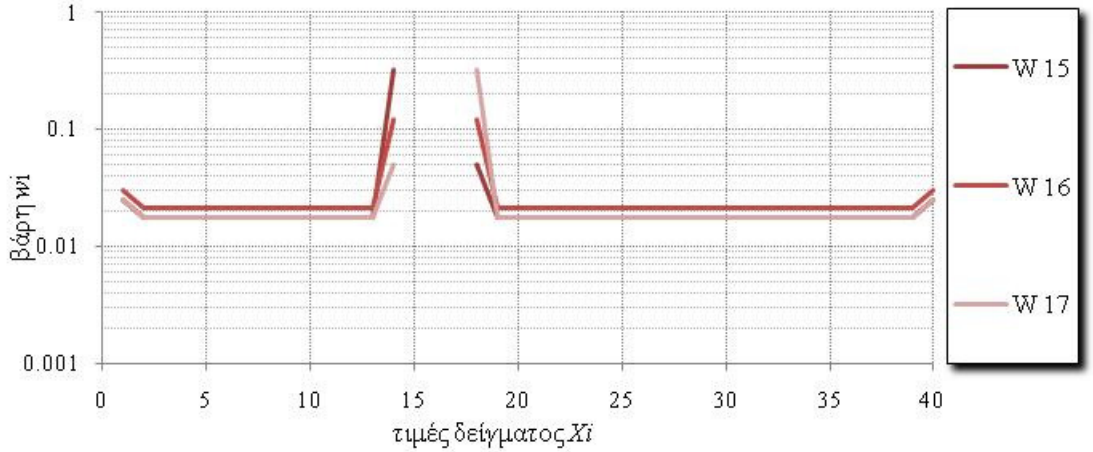


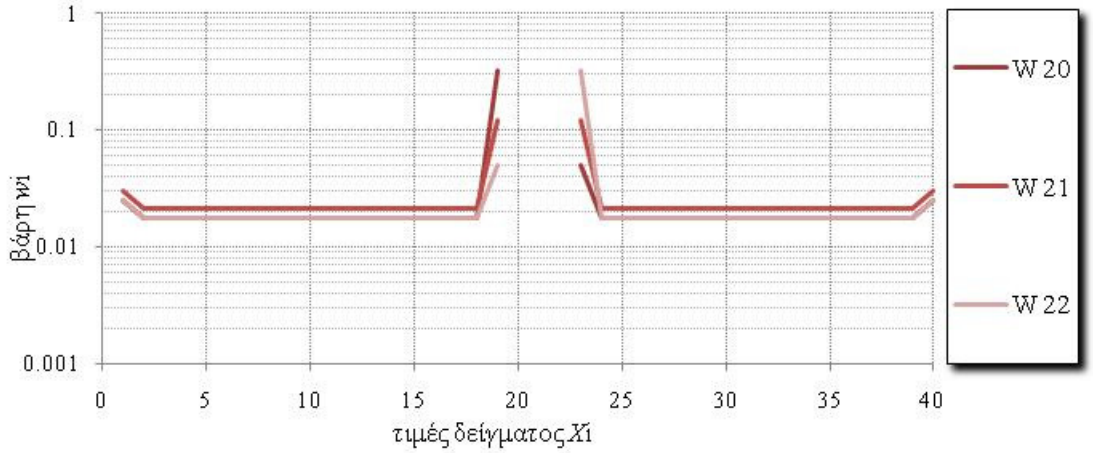

- Για δείγμα μεγέθους 40 τιμών και $\rho_1$=0.3 έχουμε:



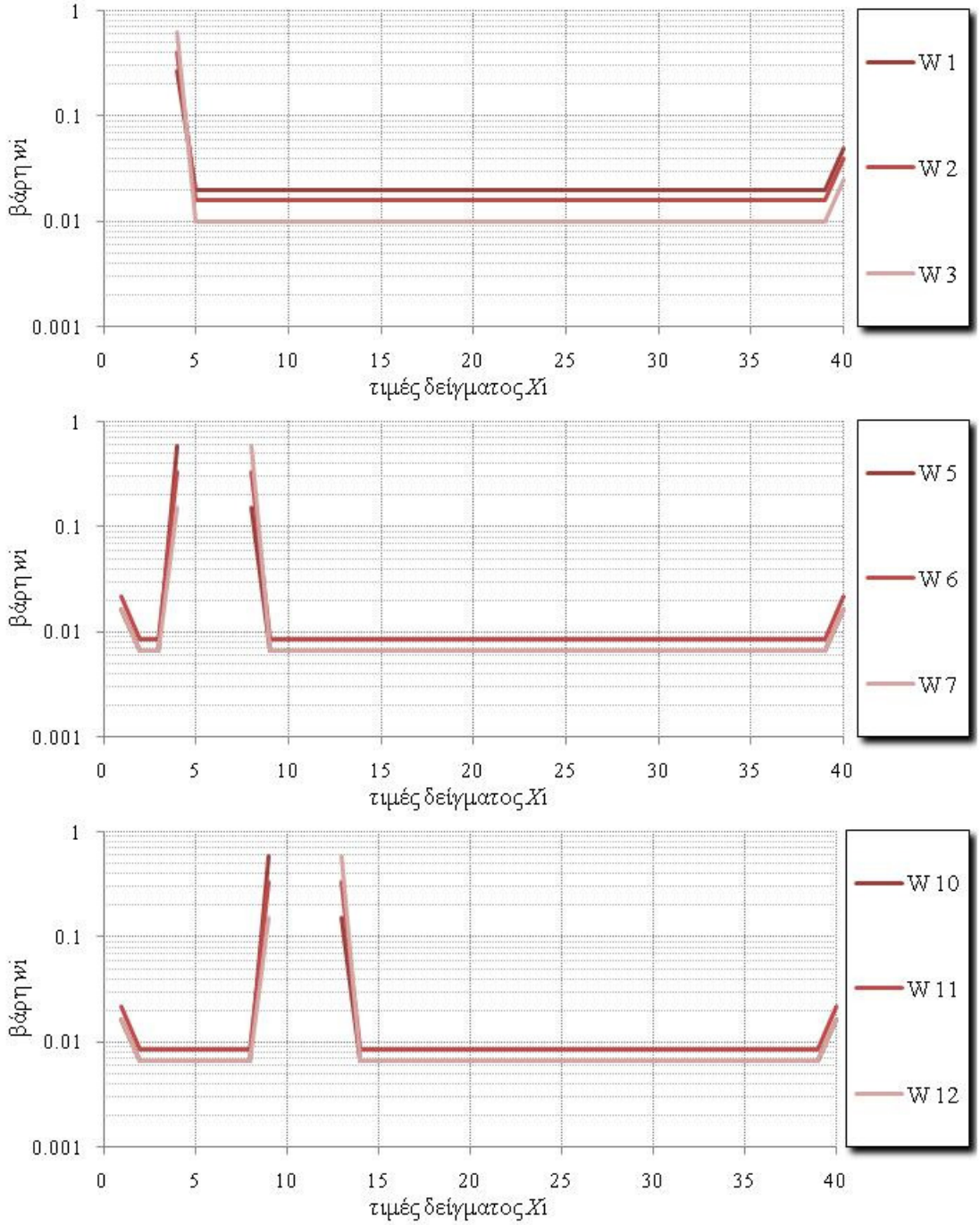

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 5
W 6
W 7

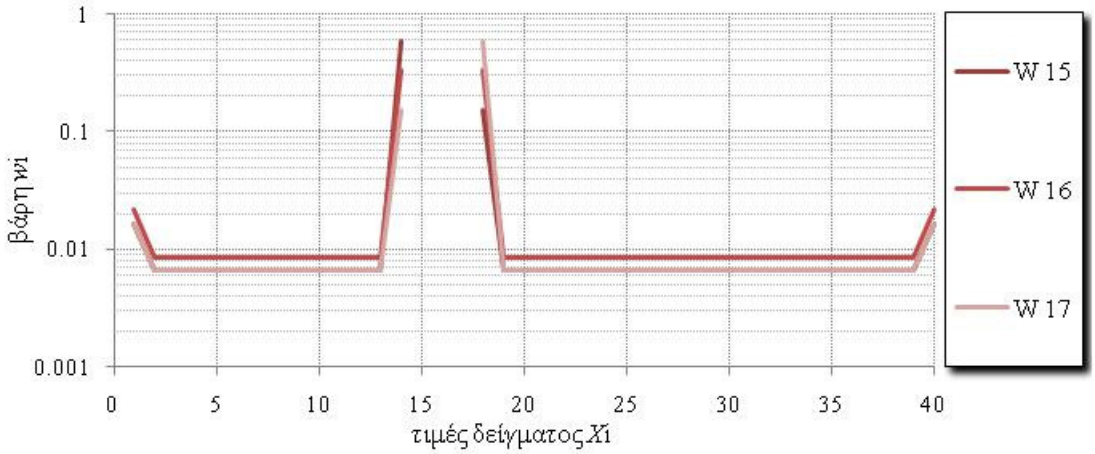

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 10
W 11
W 12

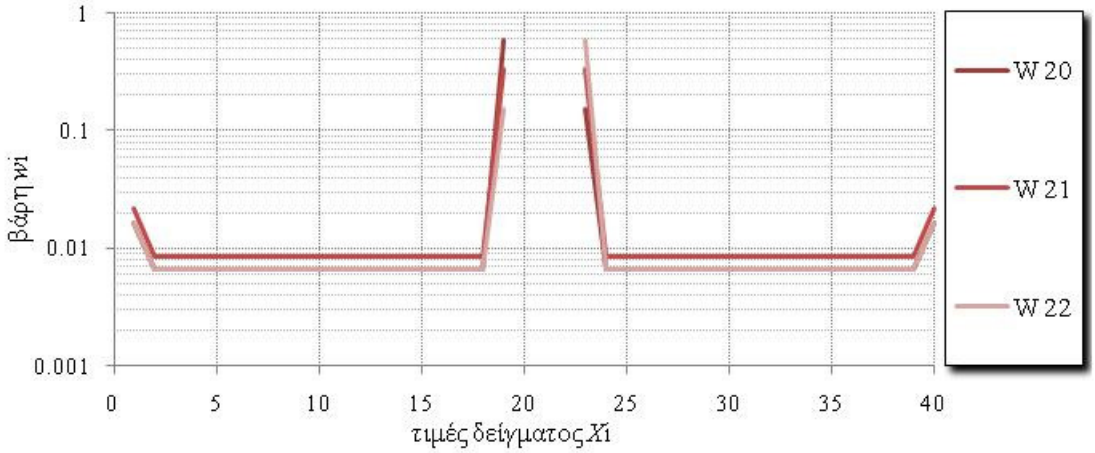

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 15
W 16
W 17


1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 20
W 21
W 22

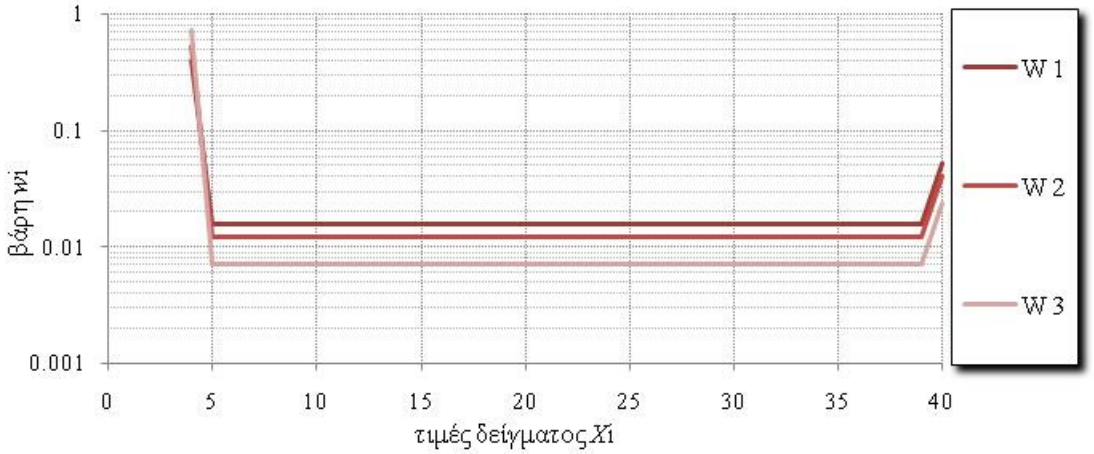
- Για δείγμα μεγέθους 40 τιμών και $\rho_1$=0.6 έχουμε:

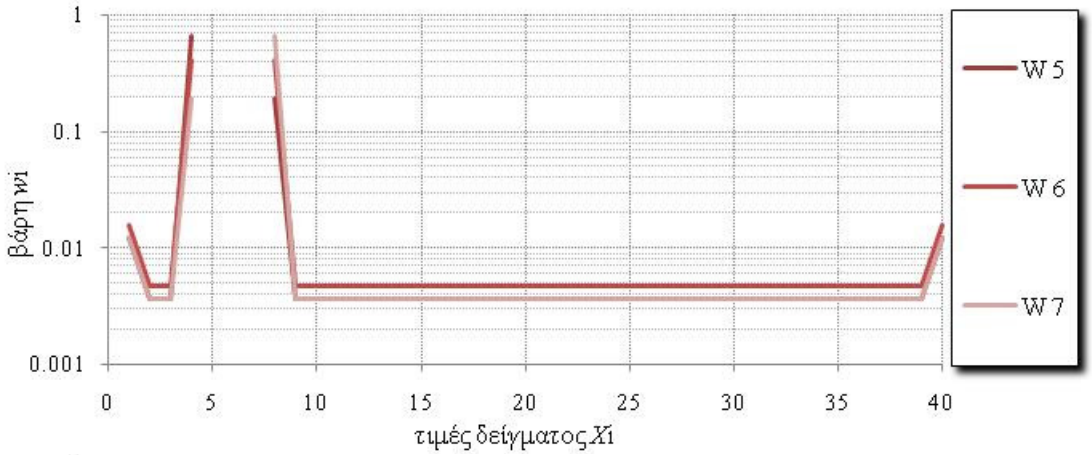


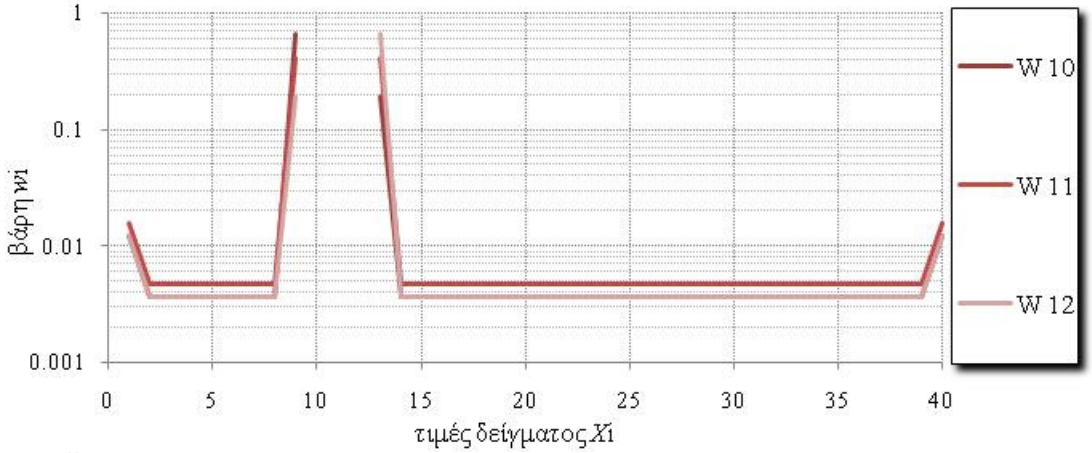


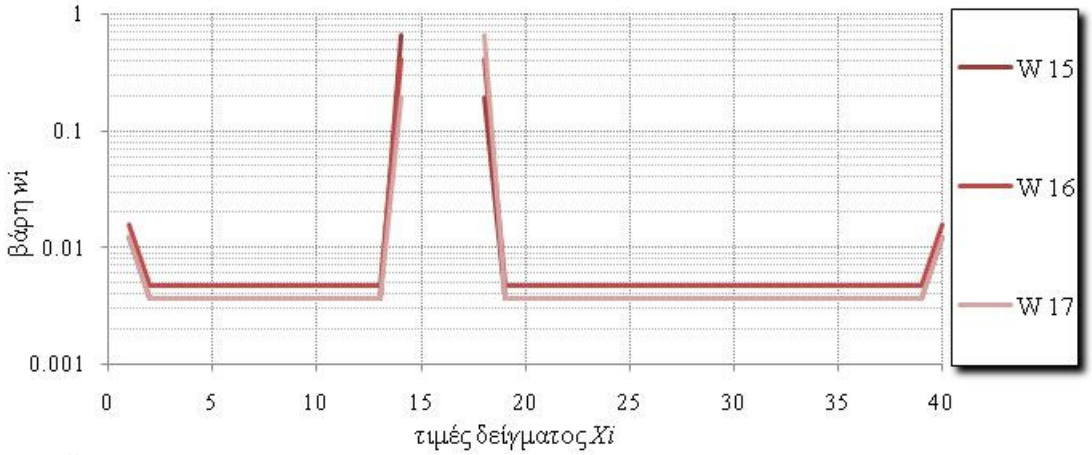



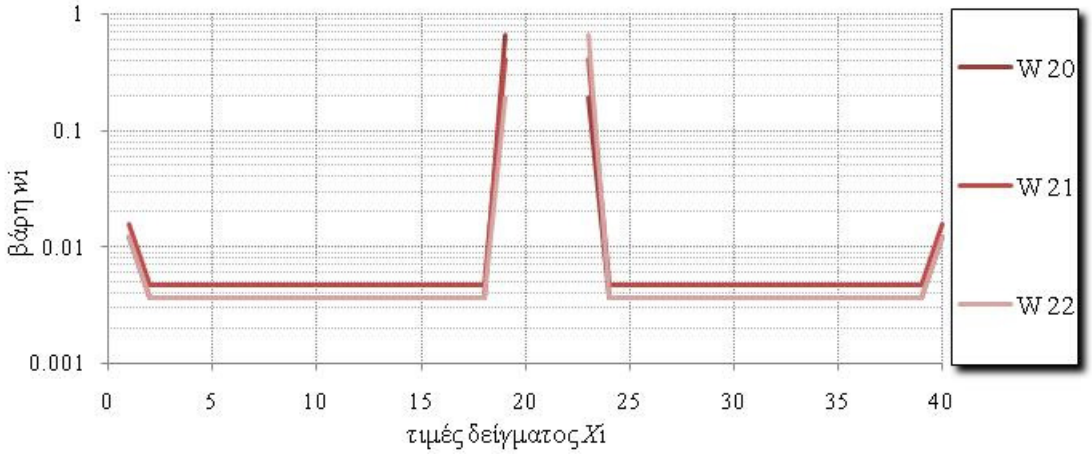

- Για δείγμα μεγέθους 40 τιμών και $\rho_1$=0.7 έχουμε:



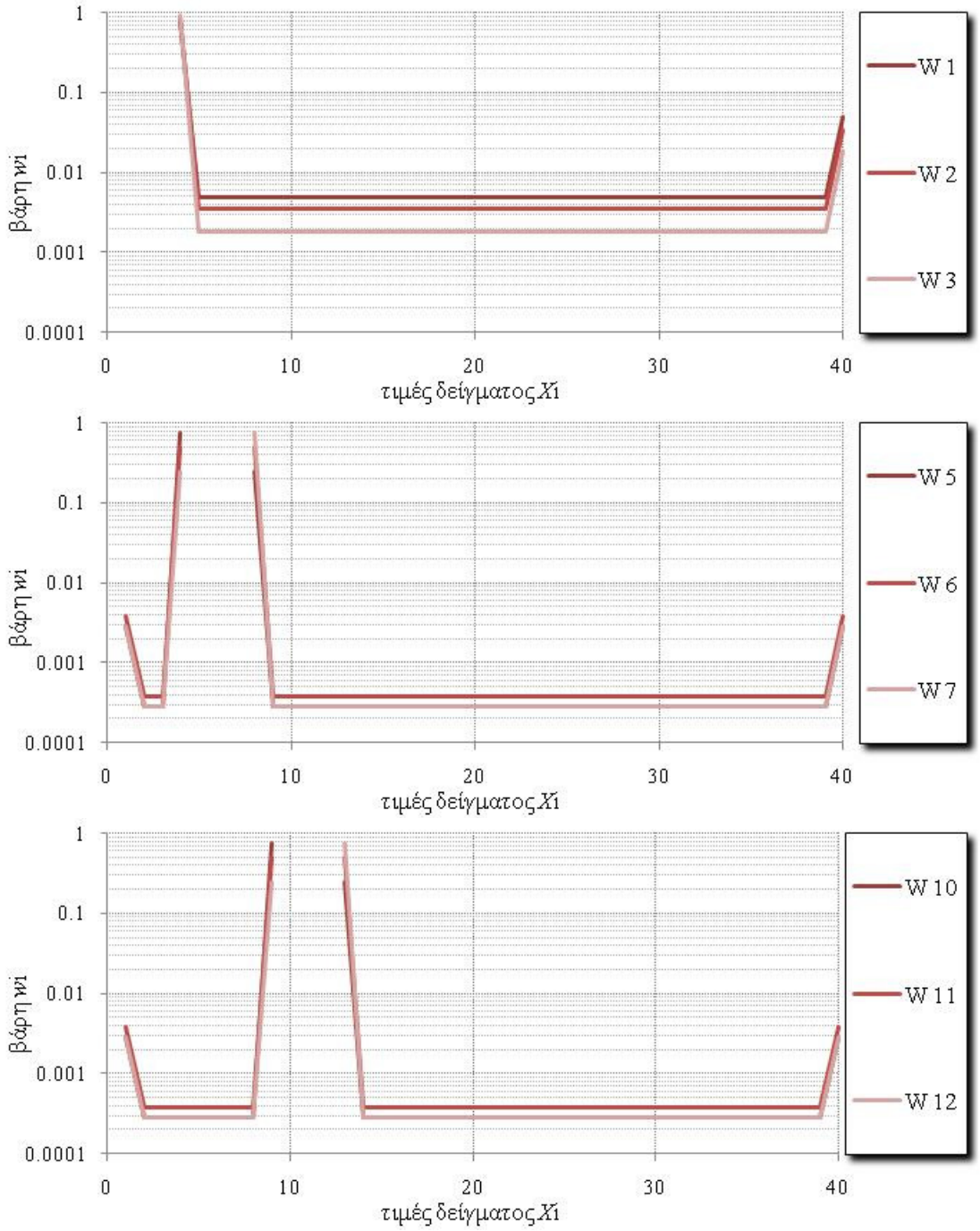

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 5
W 6
W 7

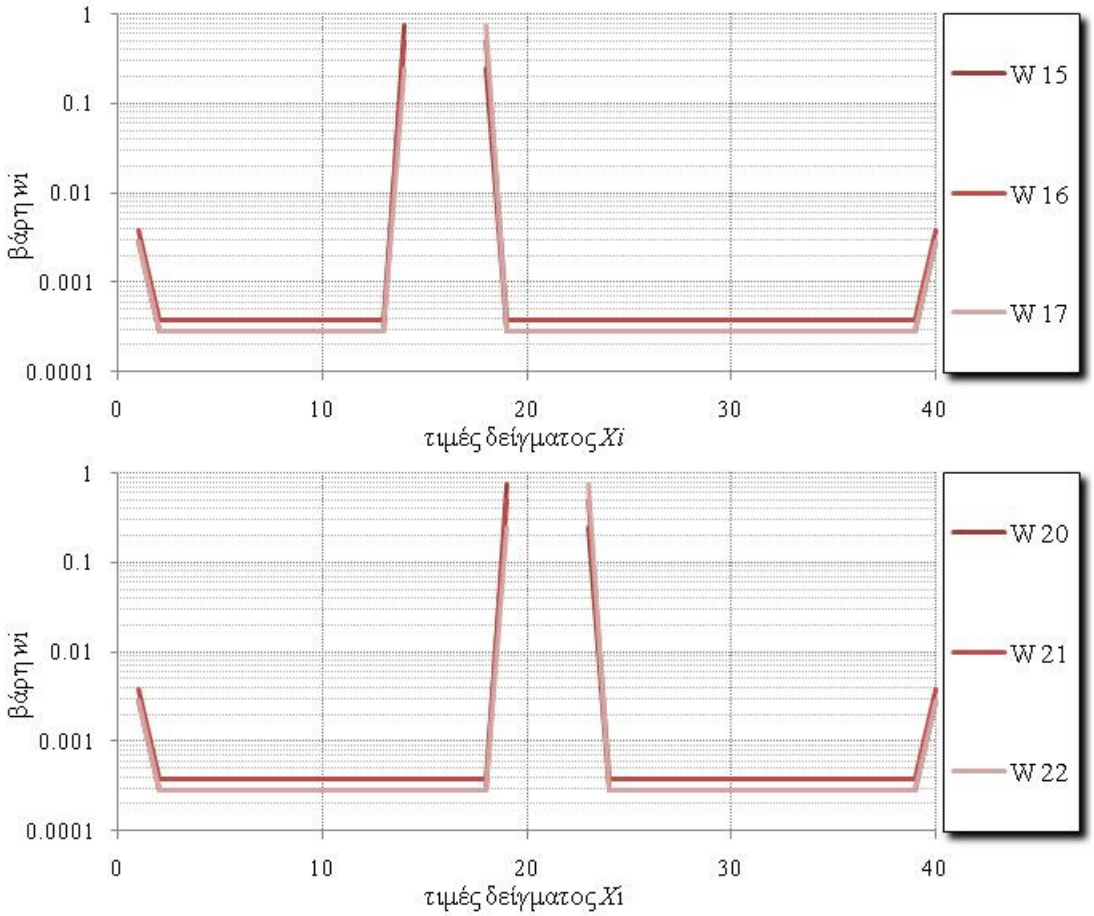

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 10
W 11
W 12


1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 15
W 16
W 17

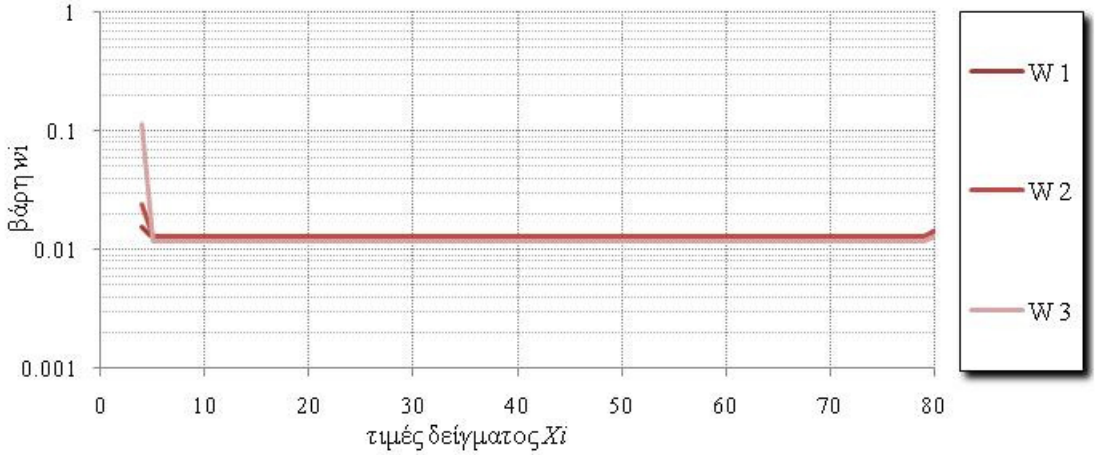

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 20
W 21
W 22

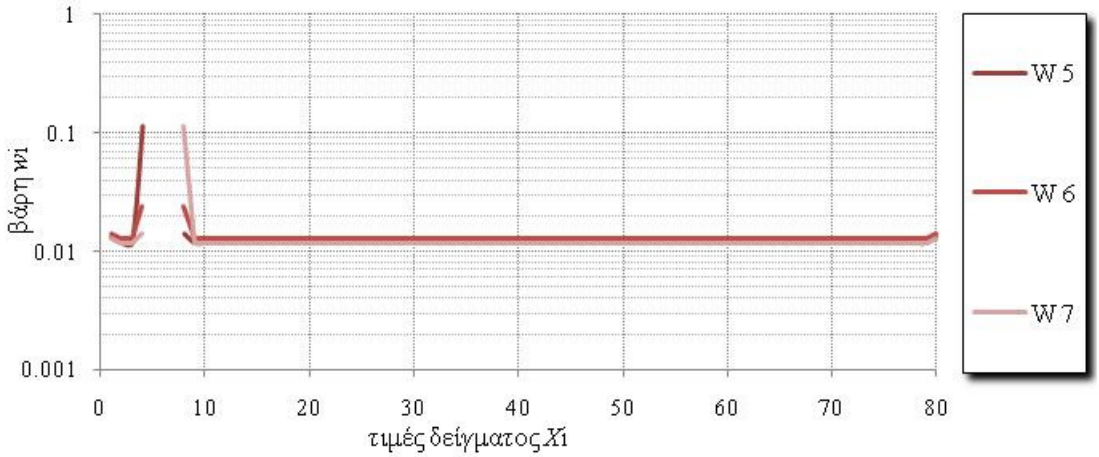
- Για δείγμα μεγέθους 40 τιμών και $\rho_1$=0.9 έχουμε:

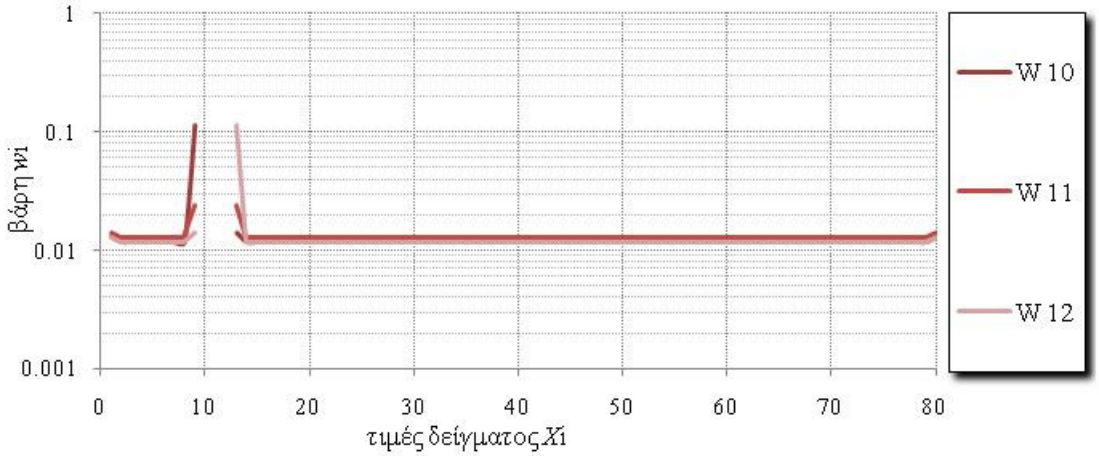


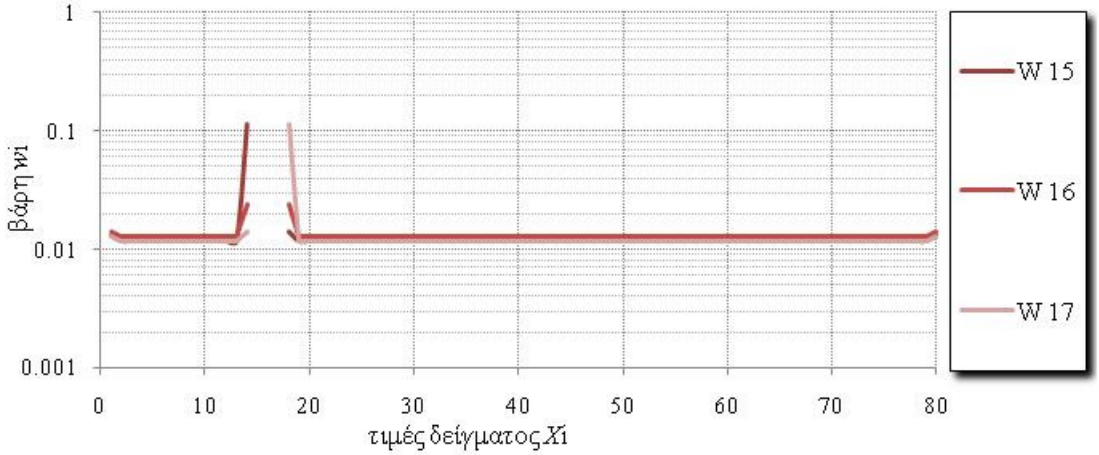


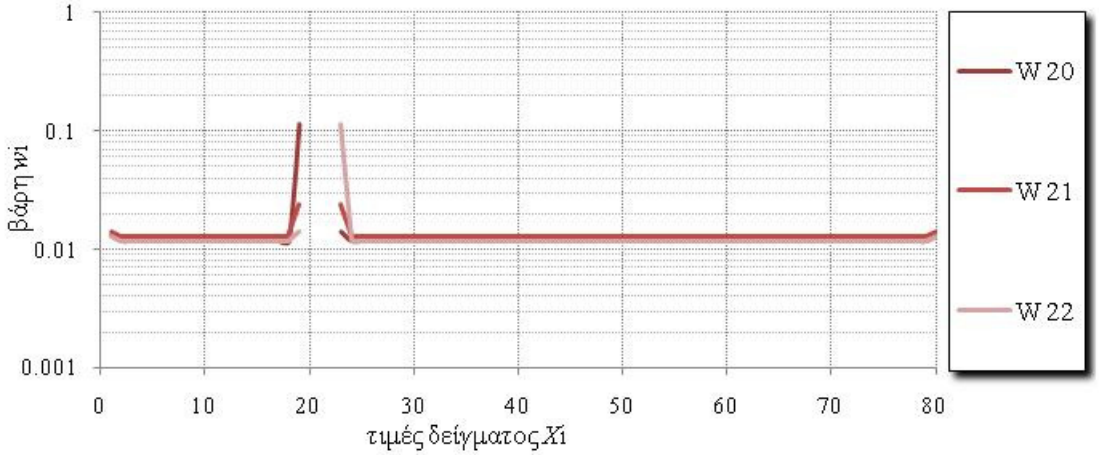

- Για δείγμα μεγέθους 80 τιμών και $\rho_1$=0.1 έχουμε:

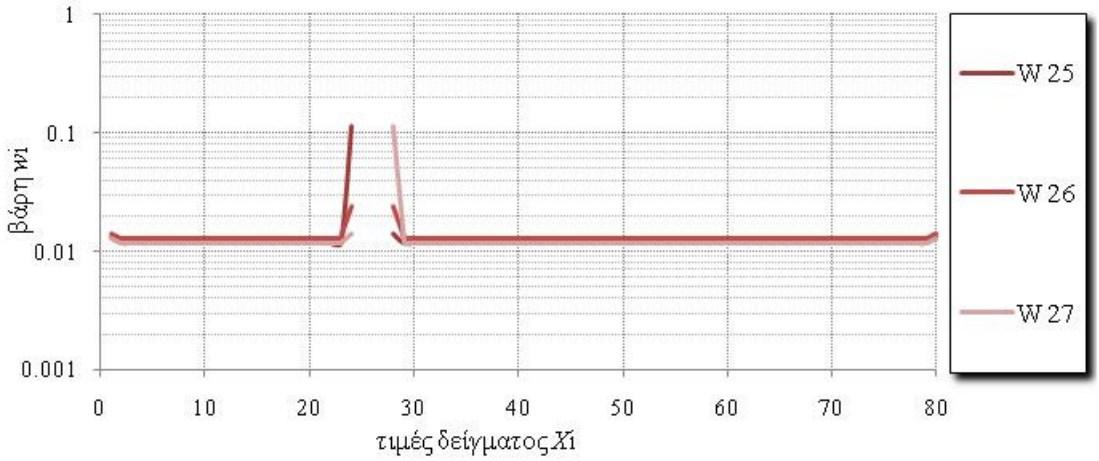


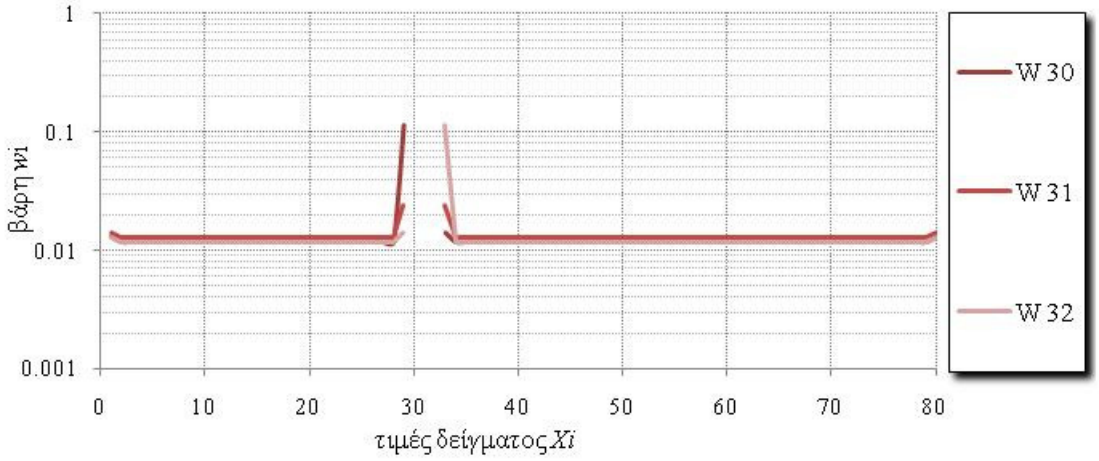

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 5
W 6
W 7

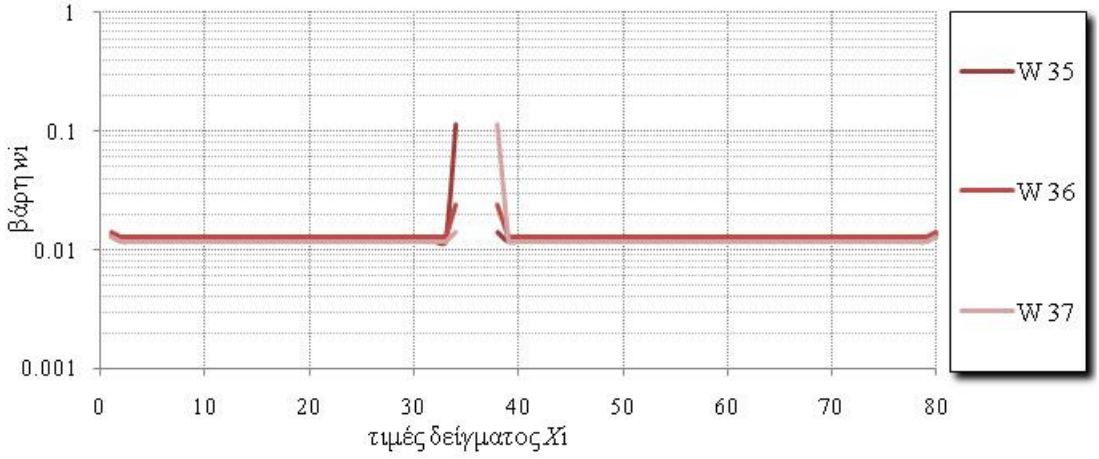

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12

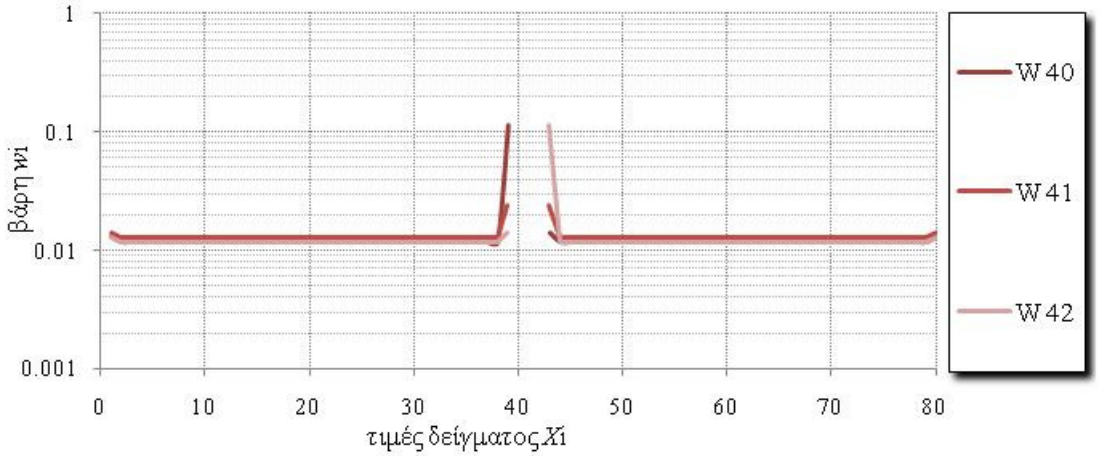

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17


1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

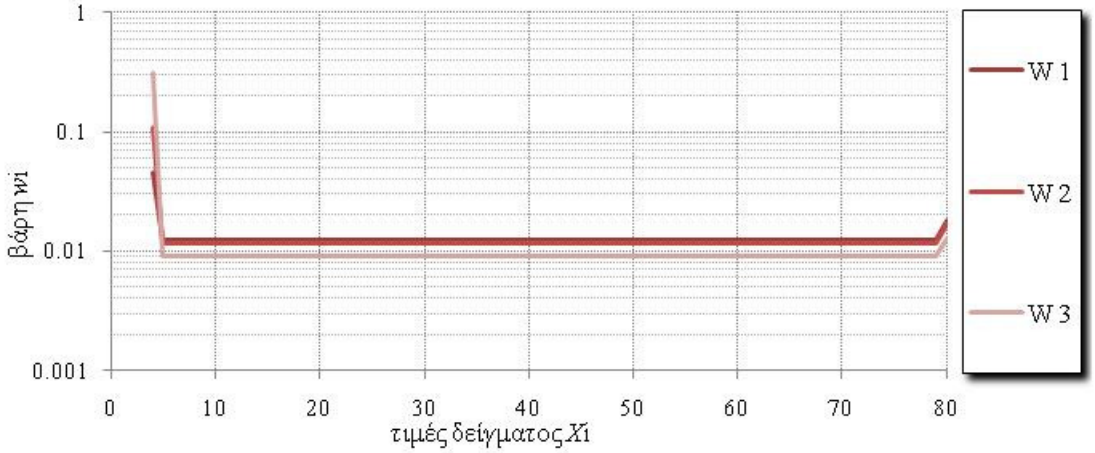

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

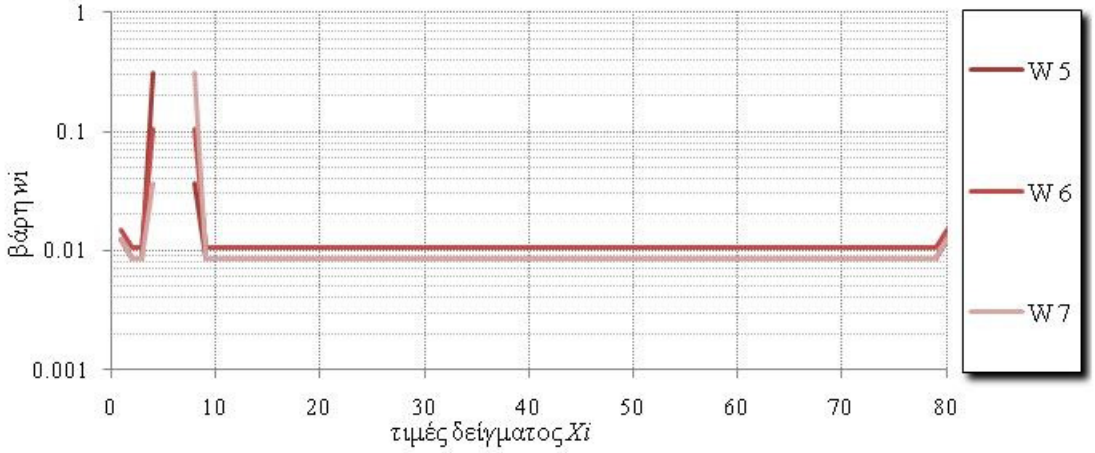

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

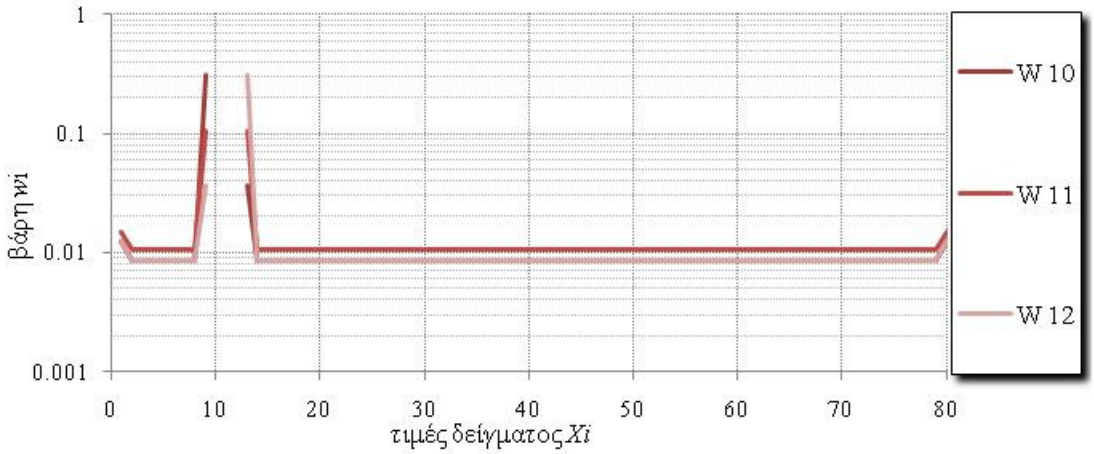

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 35
W 36
W 37

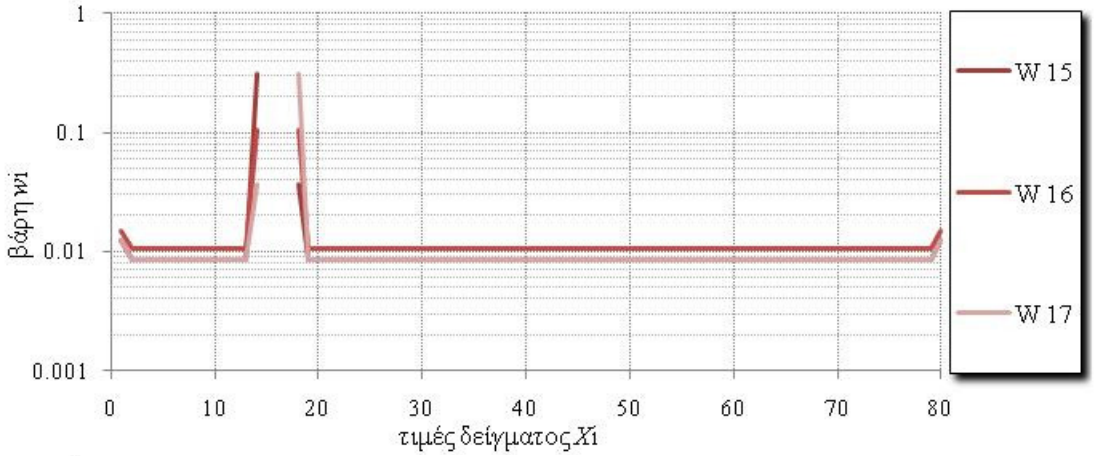


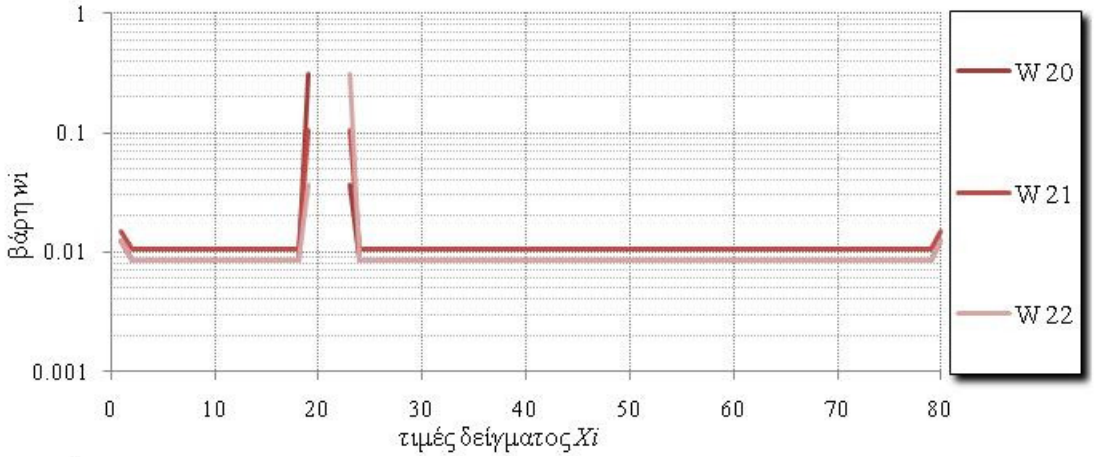

- Για δείγμα μεγέθους 80 τιμών και $\rho_1$=0.3 έχουμε:

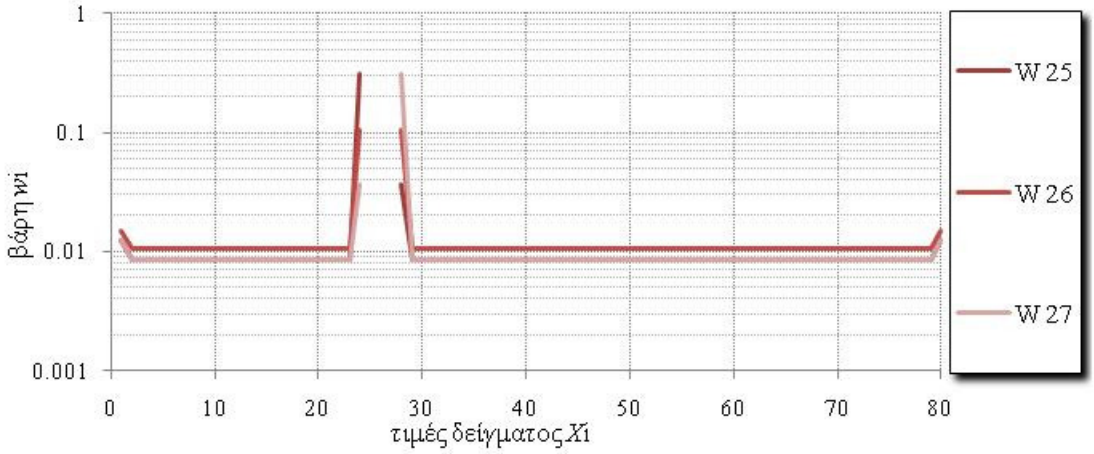


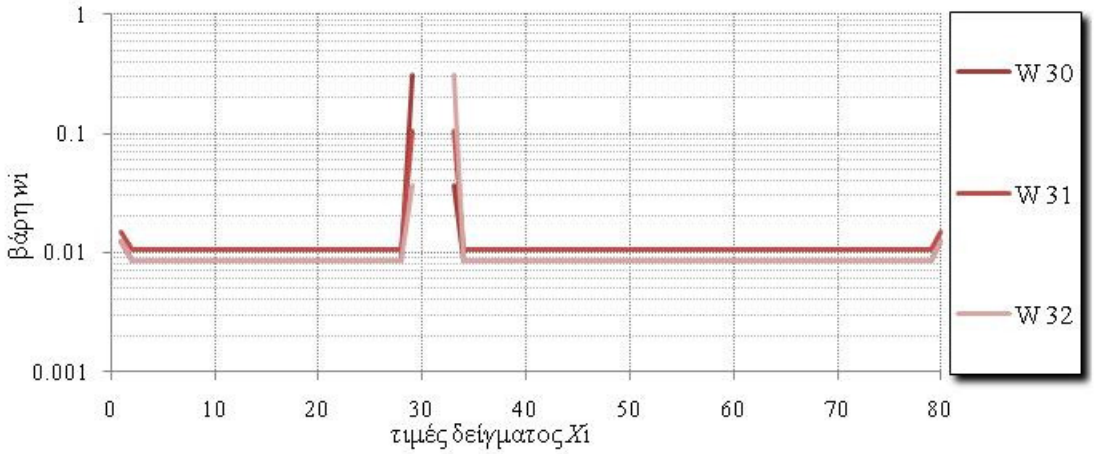



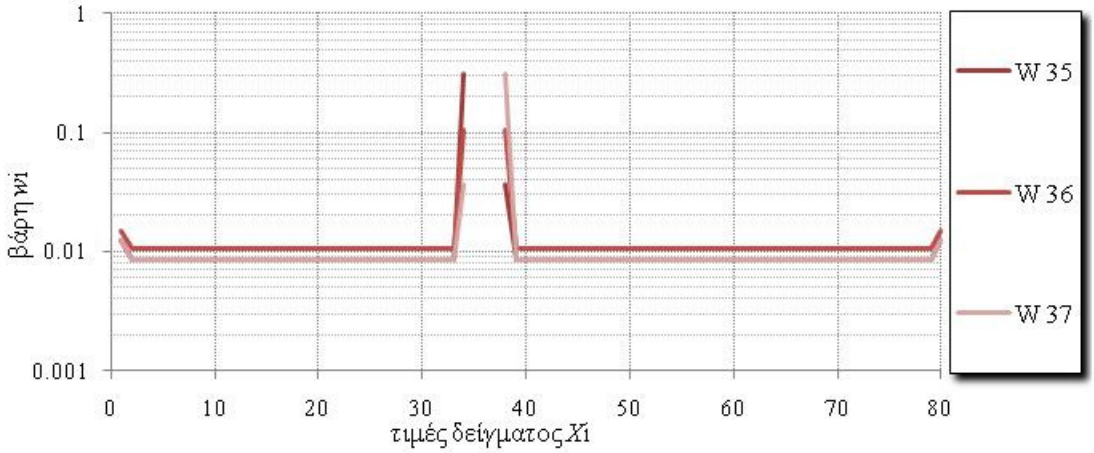

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12

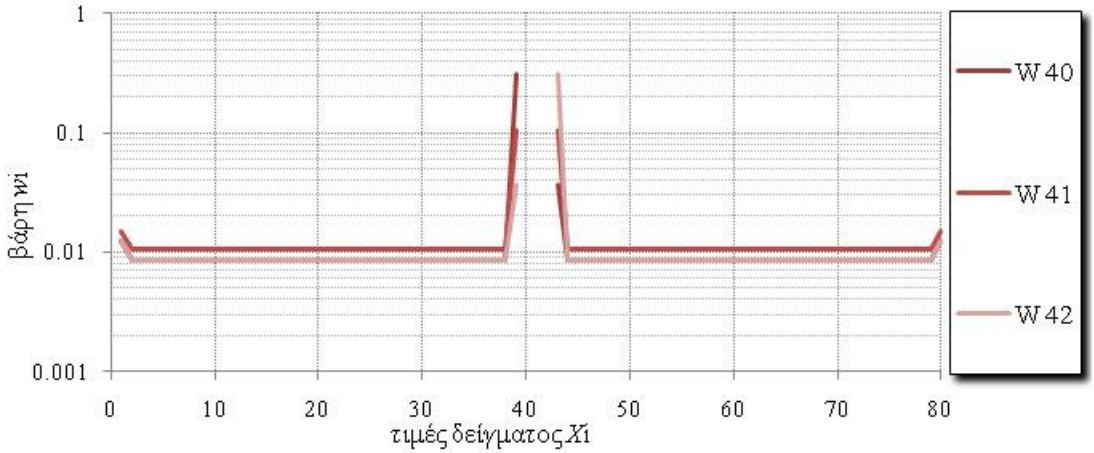

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17


1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

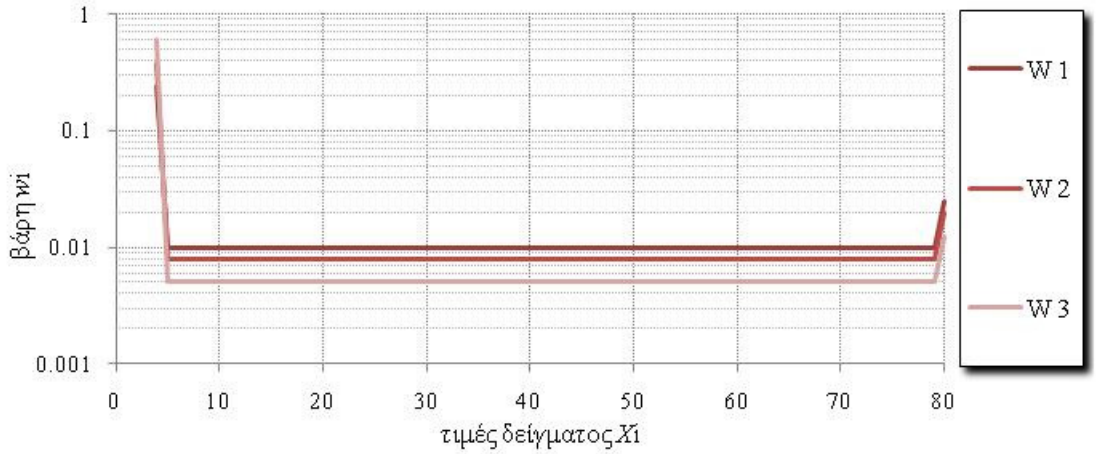

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

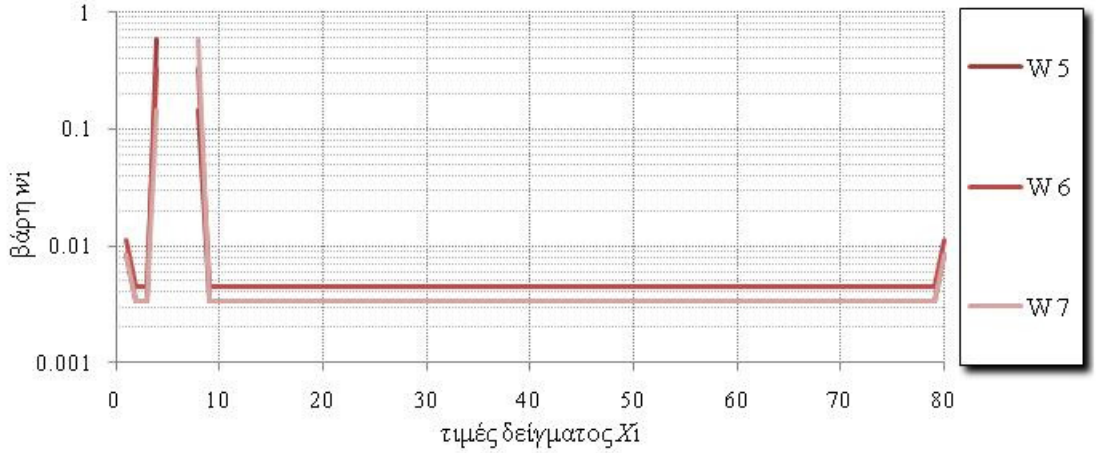

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

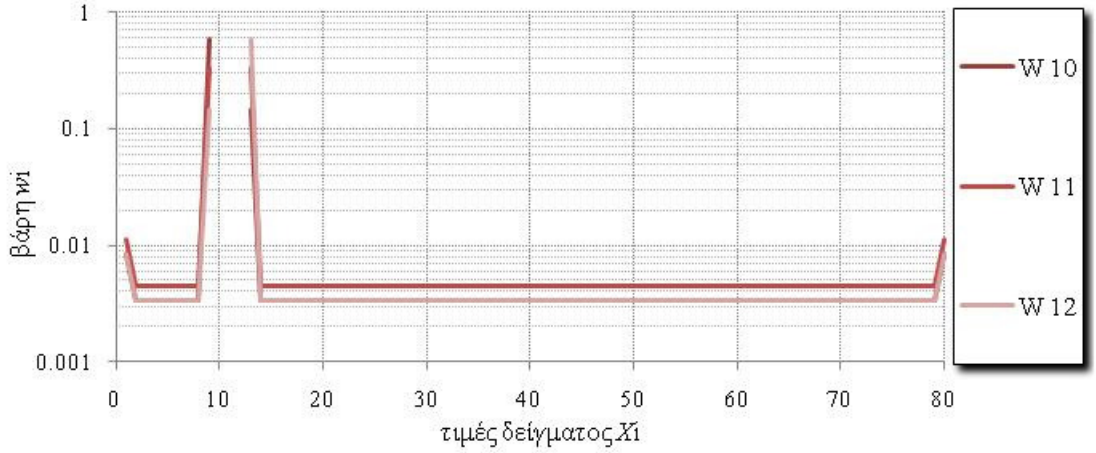

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 35
W 36
W 37

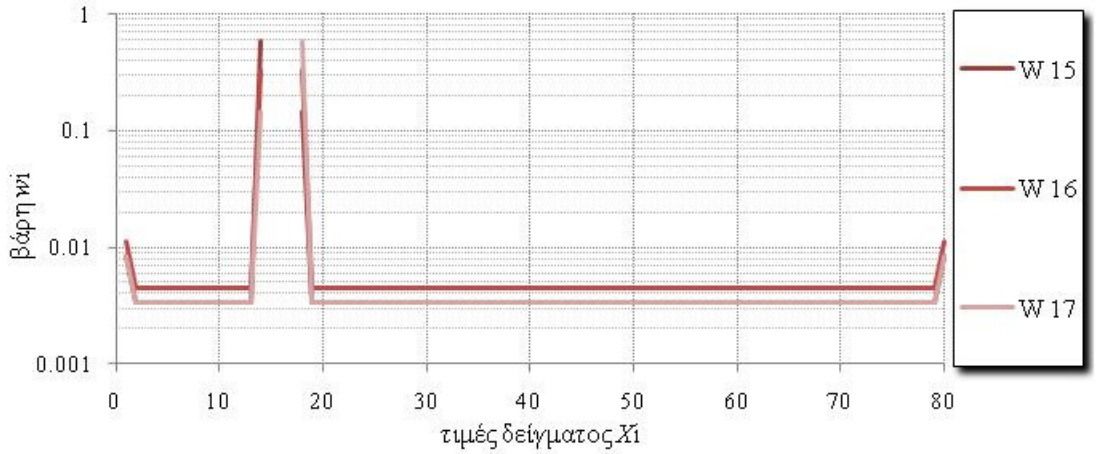

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 40
W 41
W 42

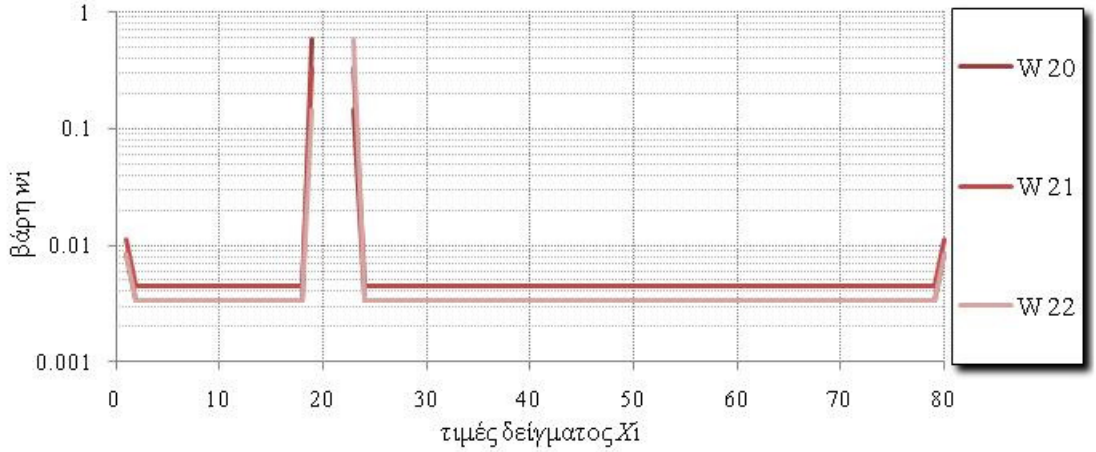
- Για δείγμα μεγέθους 80 τιμών και $\rho_1$=0.6 έχουμε:

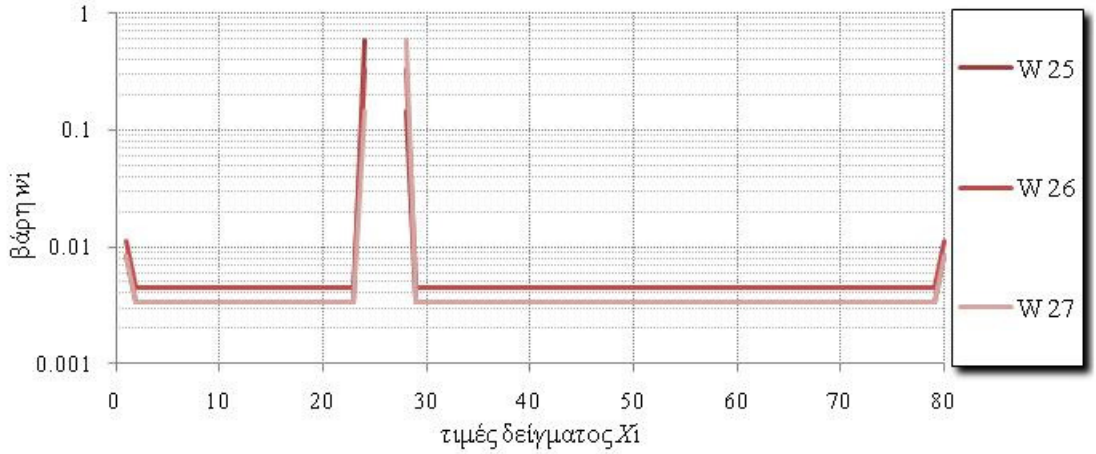


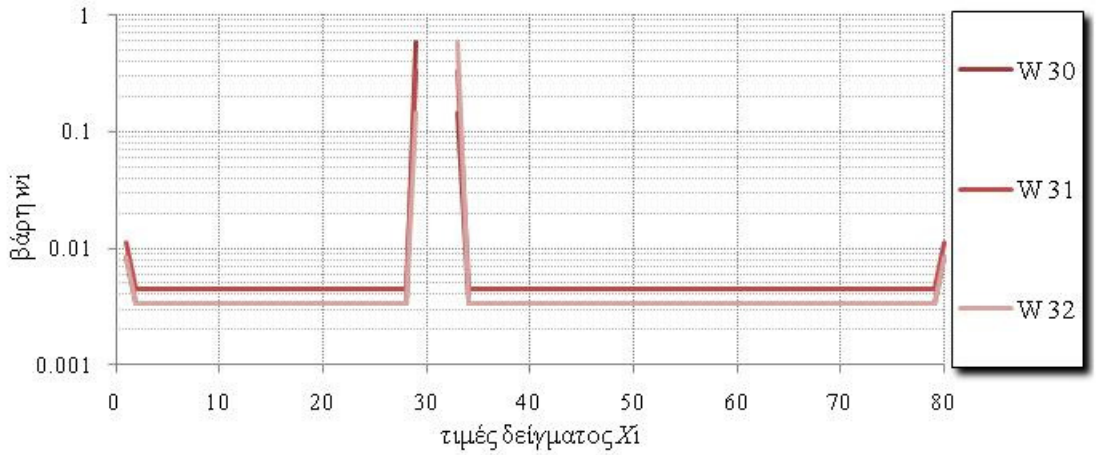



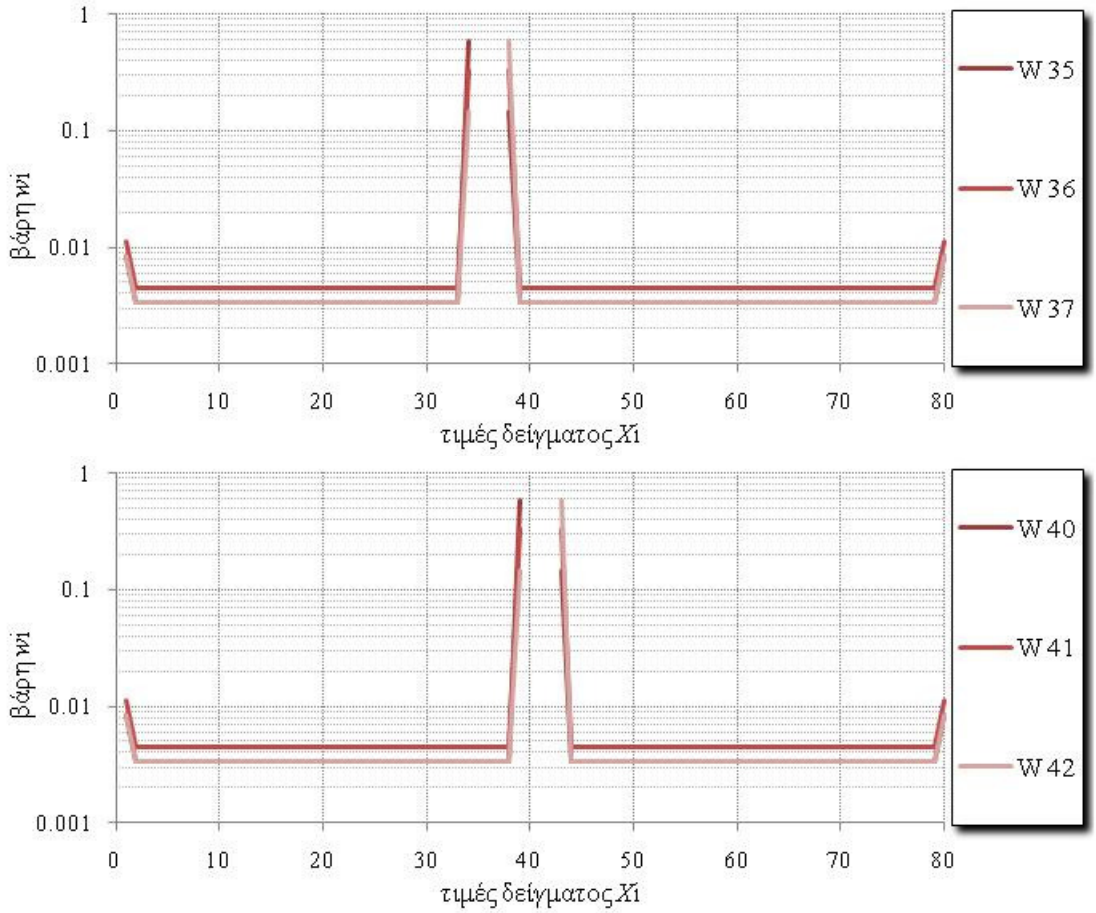




1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17

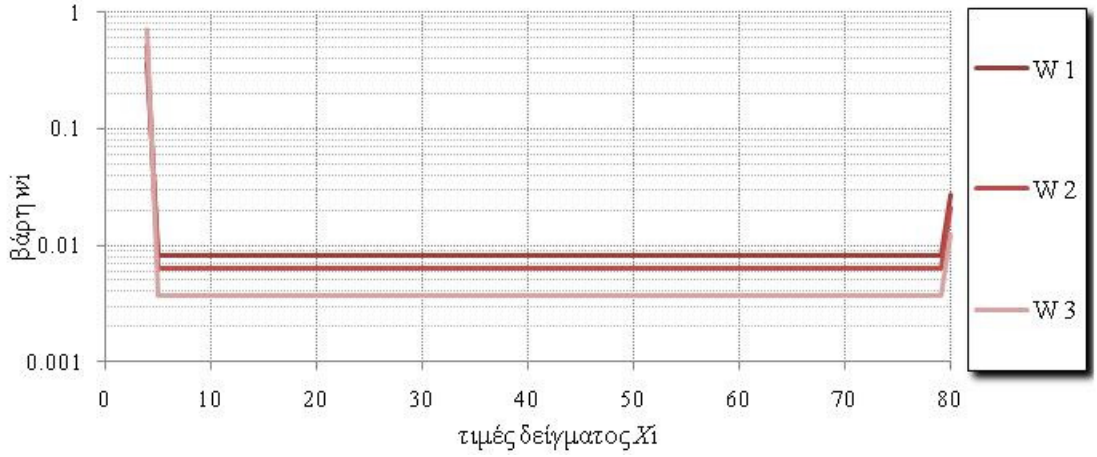

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

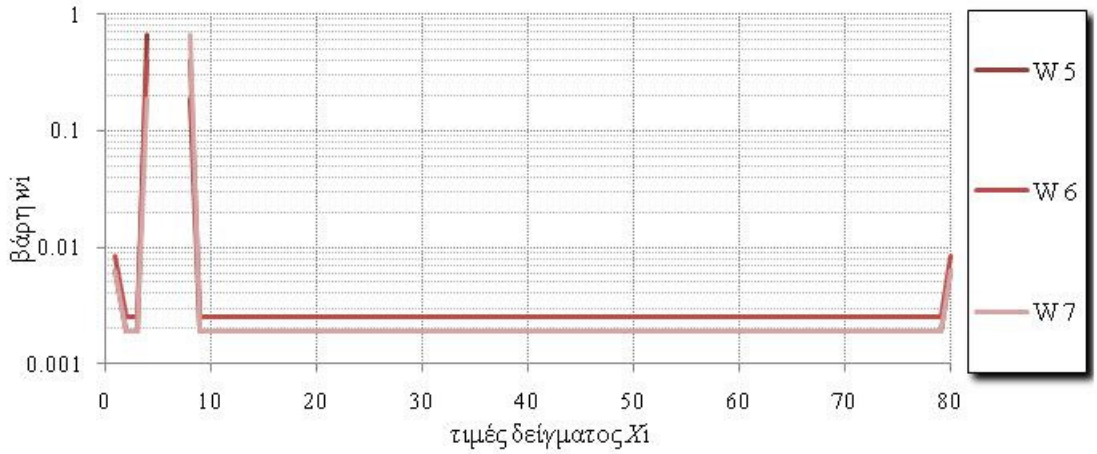

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

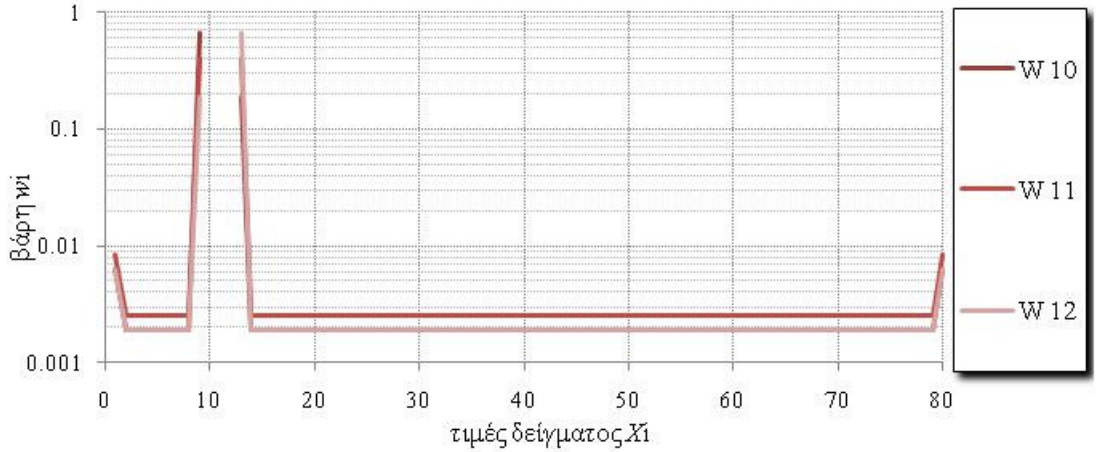

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

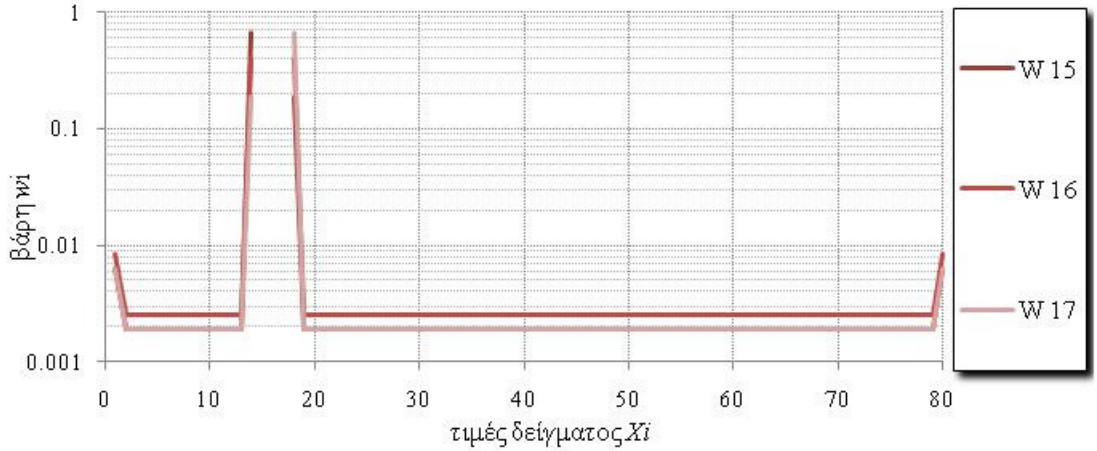


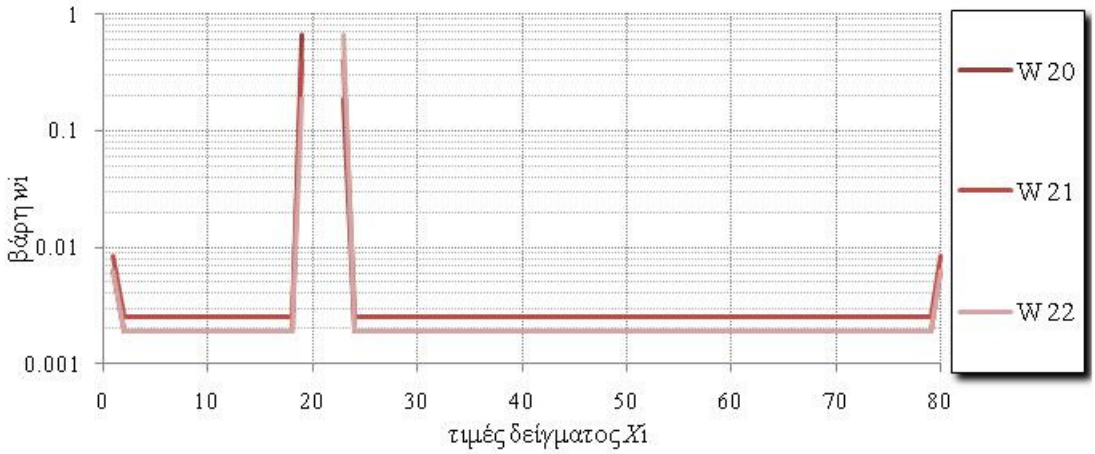

- Για δείγμα μεγέθους 80 τιμών και $\rho_1$=0.7 έχουμε:

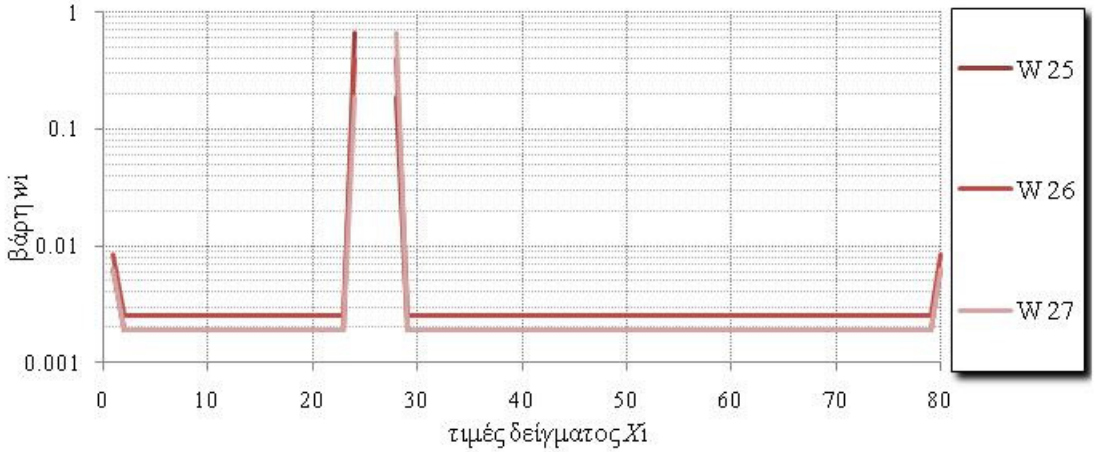


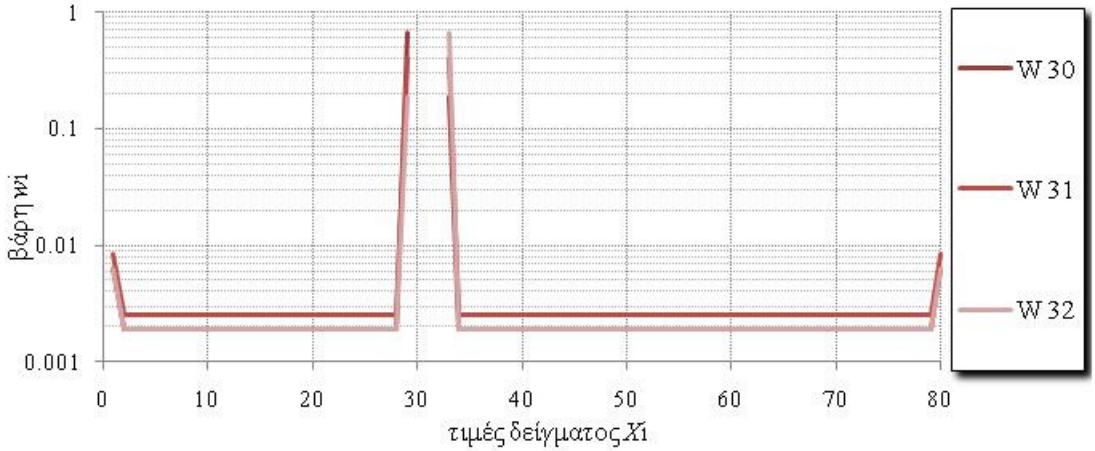

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 5
W 6
W 7

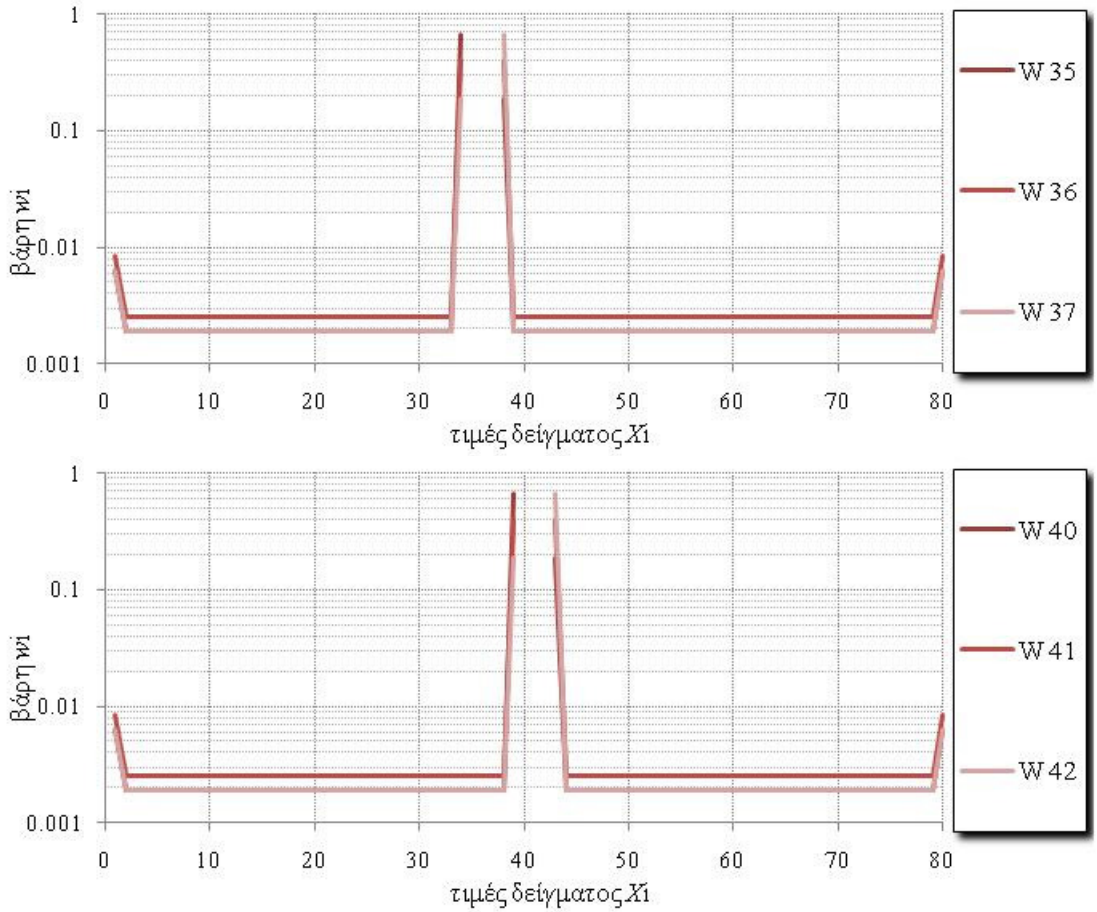

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12


1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17

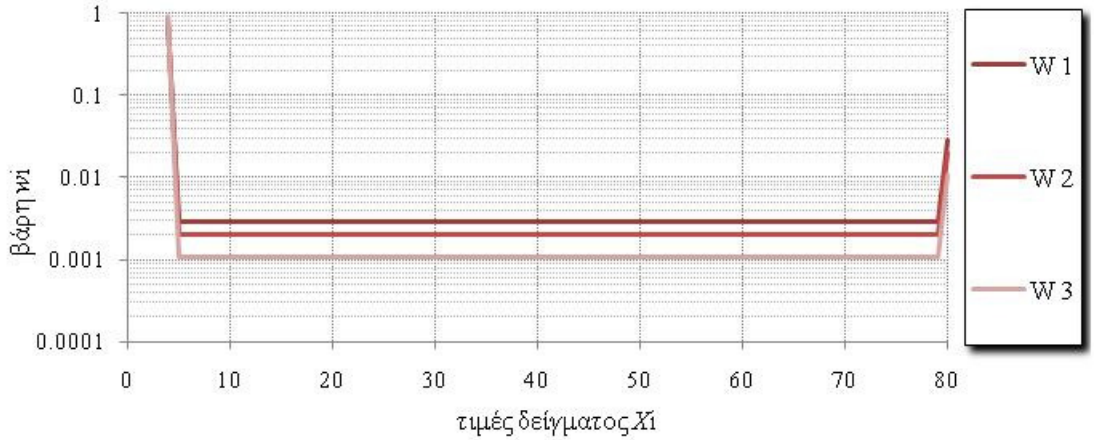

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

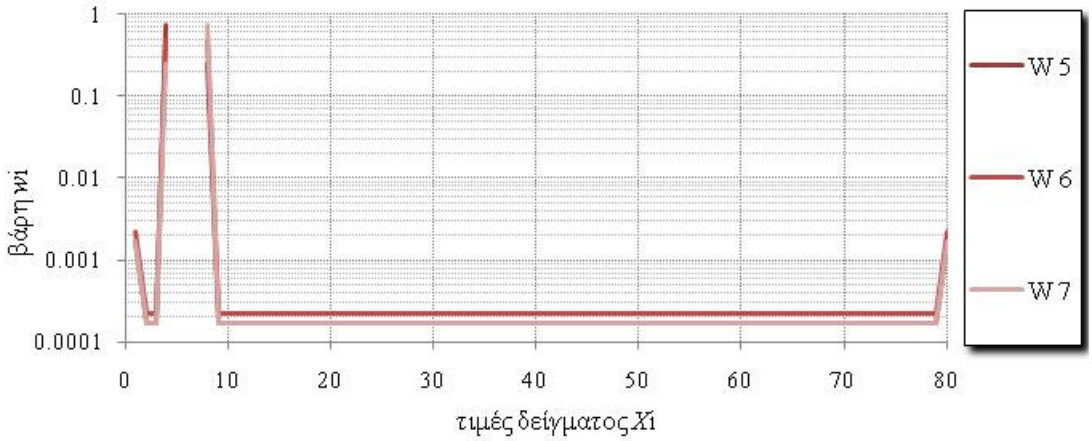

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

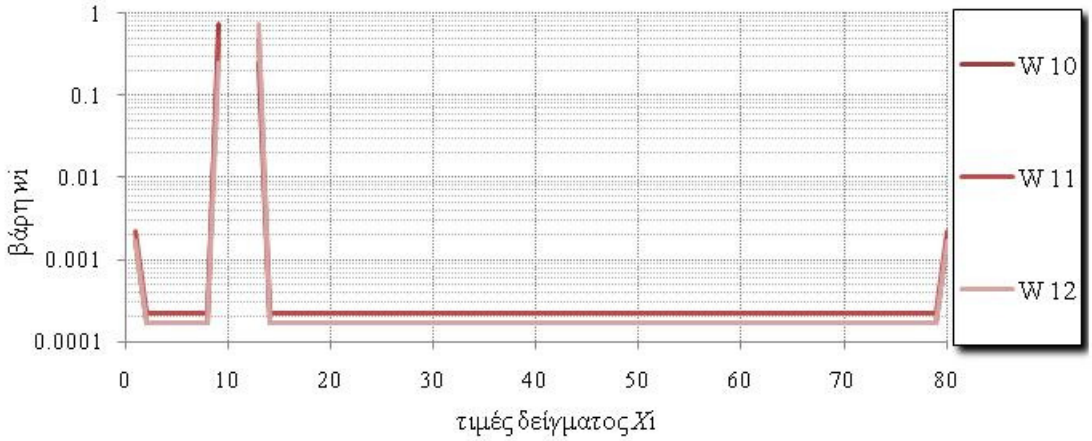

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

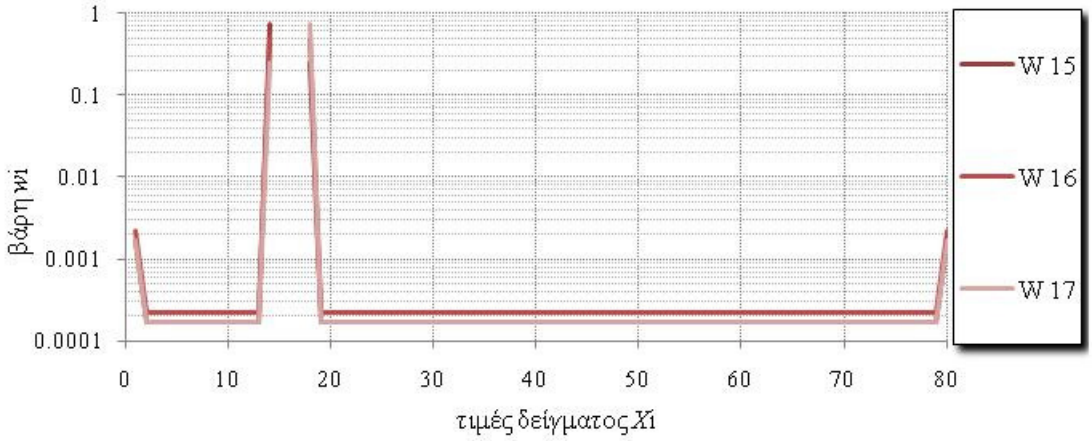


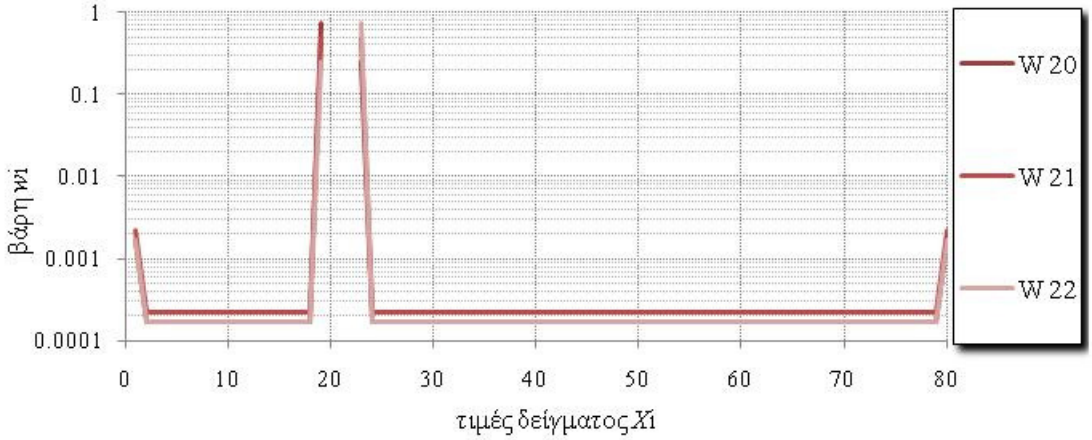

- Για δείγμα μεγέθους 80 τιμών και $\rho_1$=0.9 έχουμε:

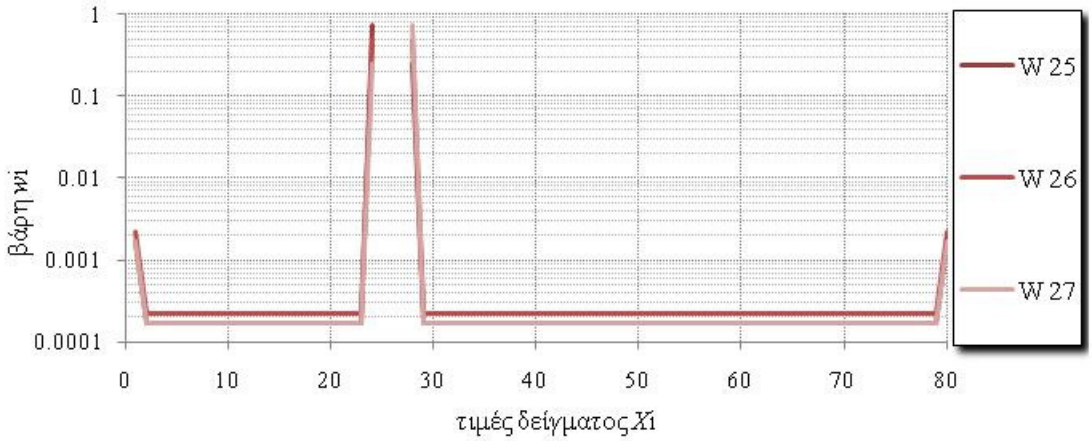


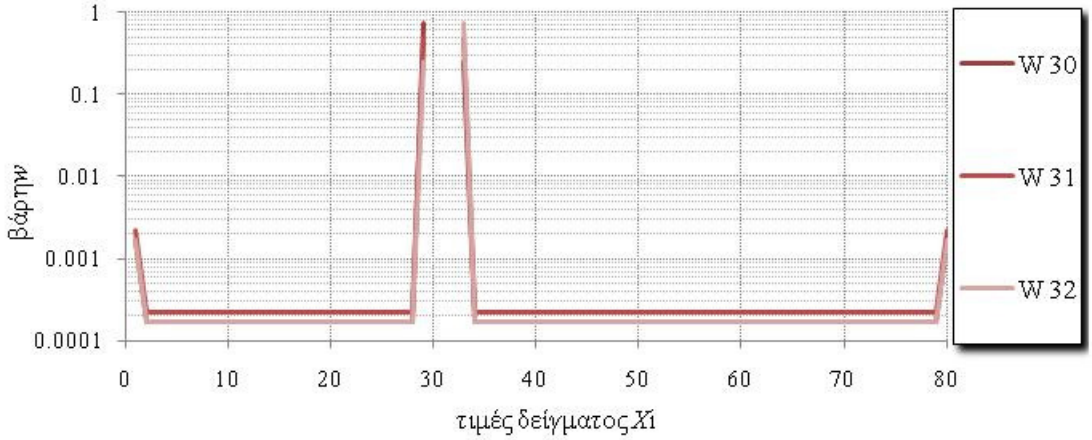

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 5
W 6
W 7

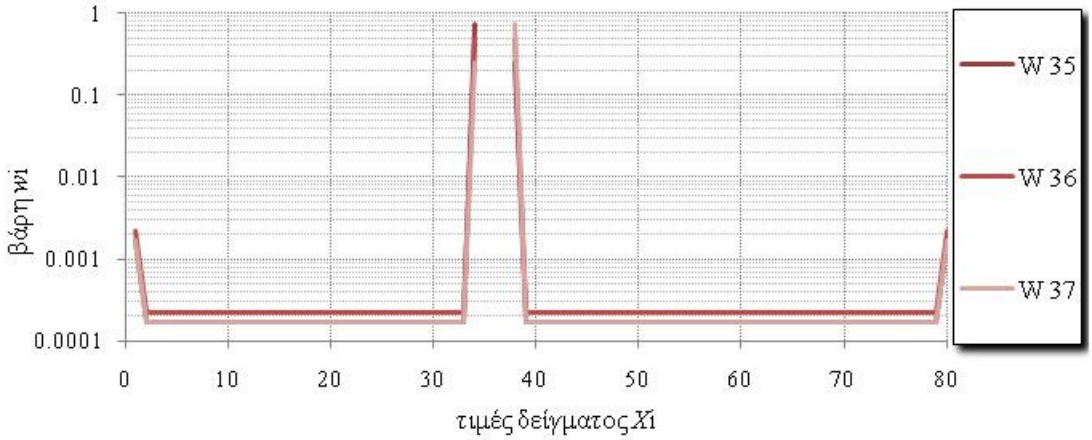

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12

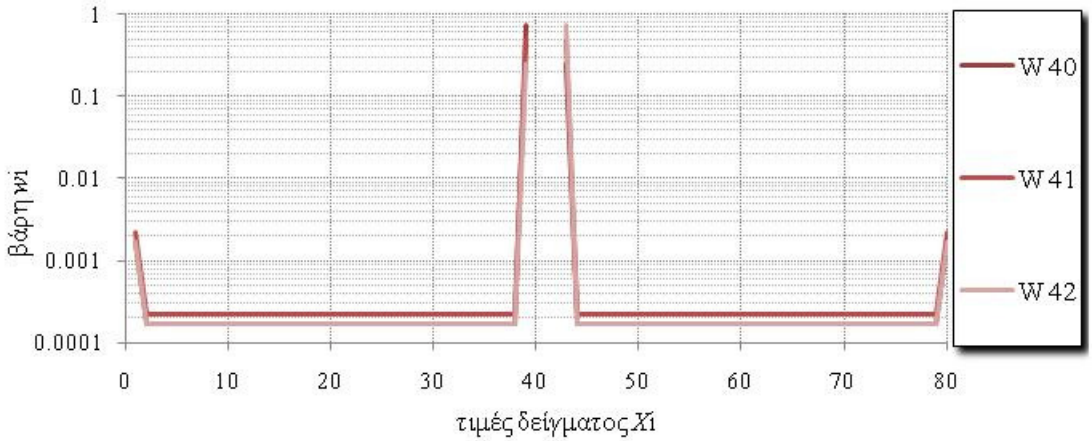

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17


βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

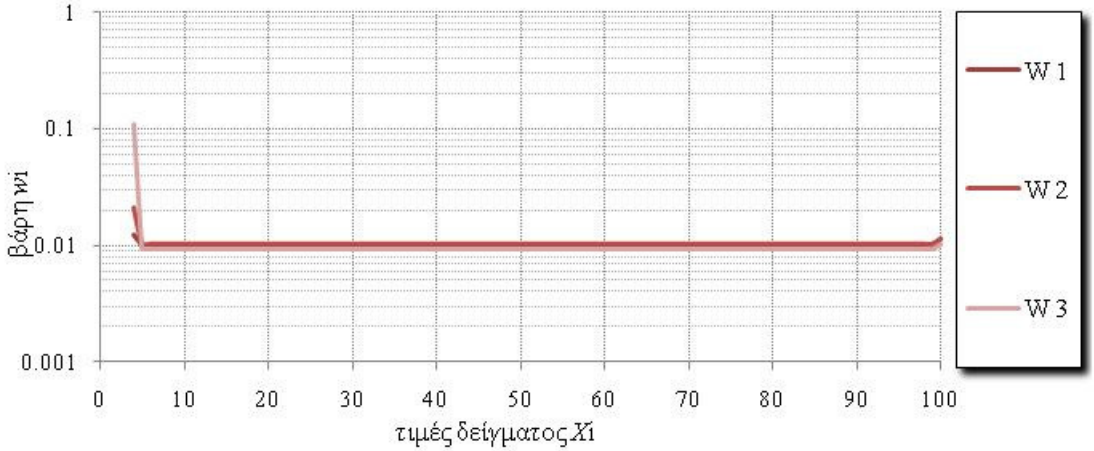

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

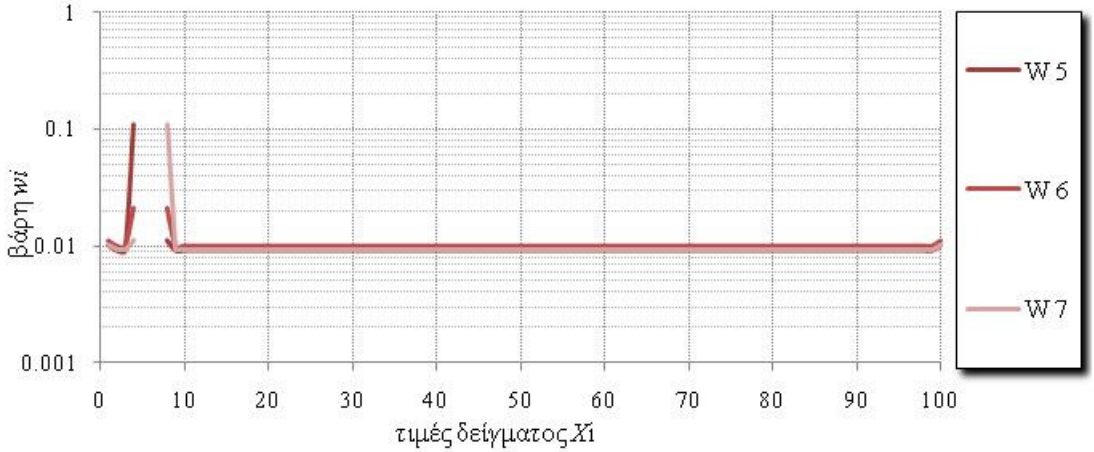

1
0.1
0.01
0.001
0.0001
βάρη w
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

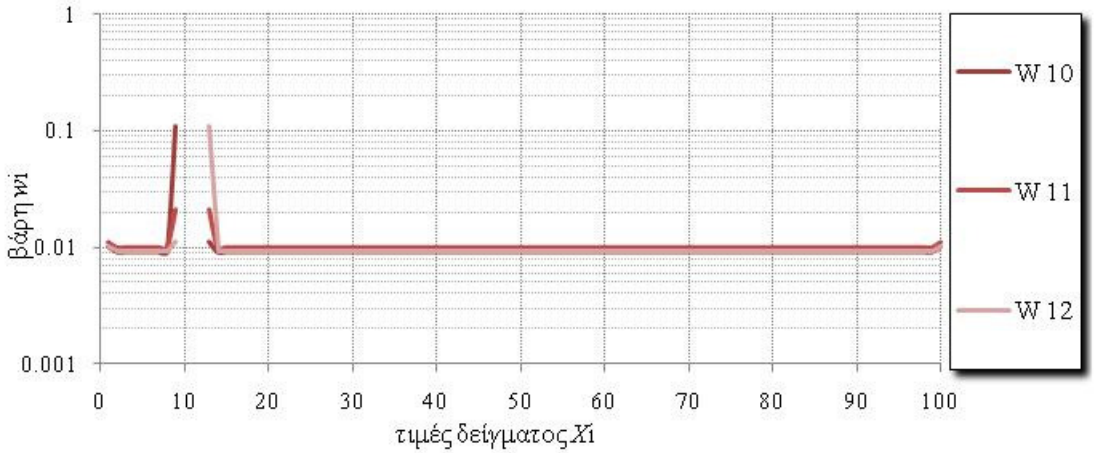

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 35
W 36
W 37

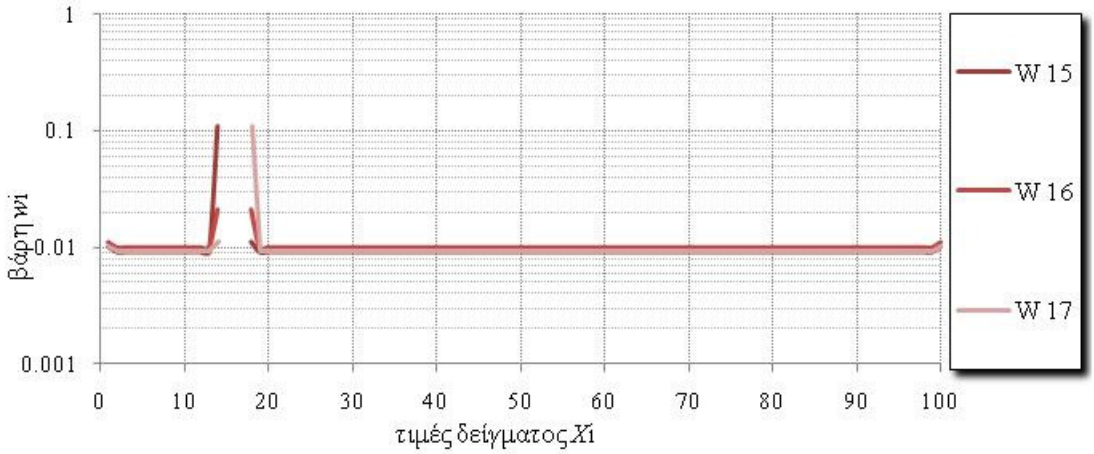

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 40
W 41
W 42

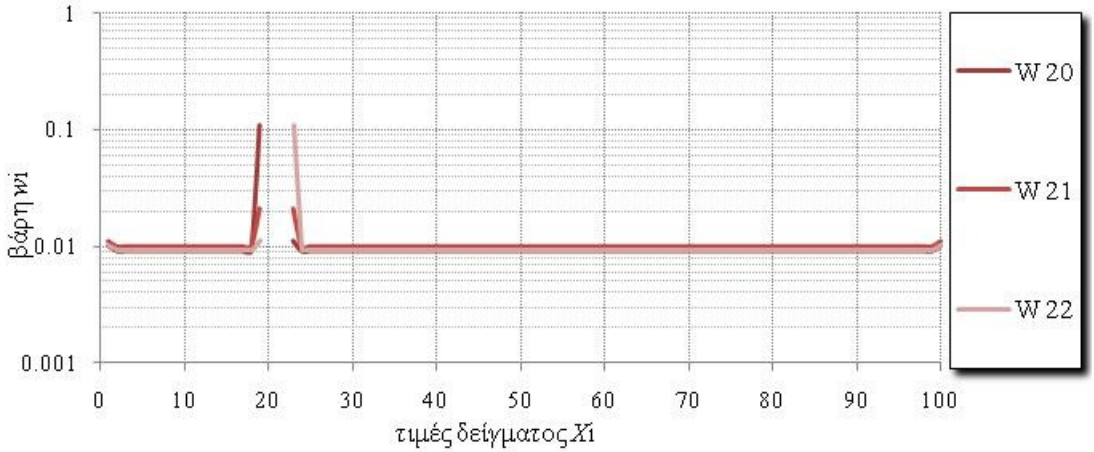
- Για δείγμα μεγέθους 100 τιμών και $\rho_1$=0.1 έχουμε:

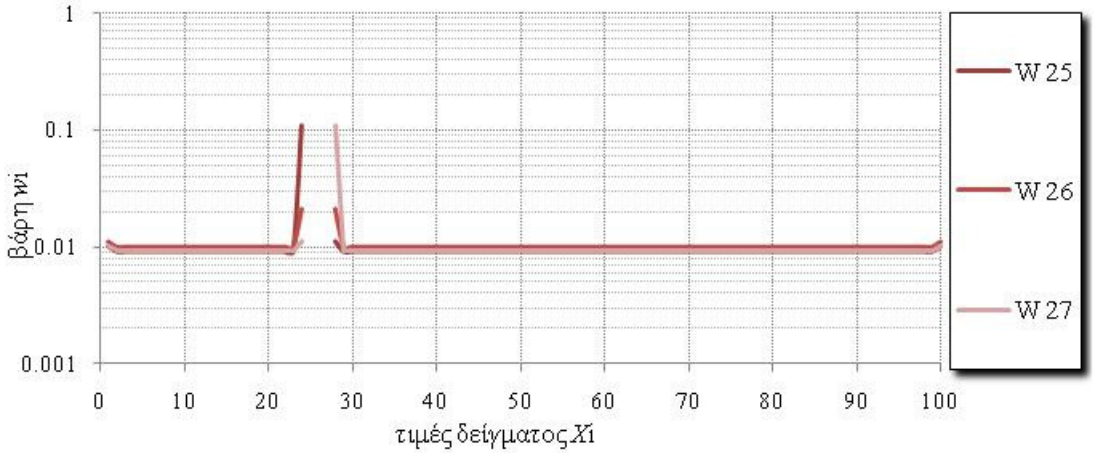


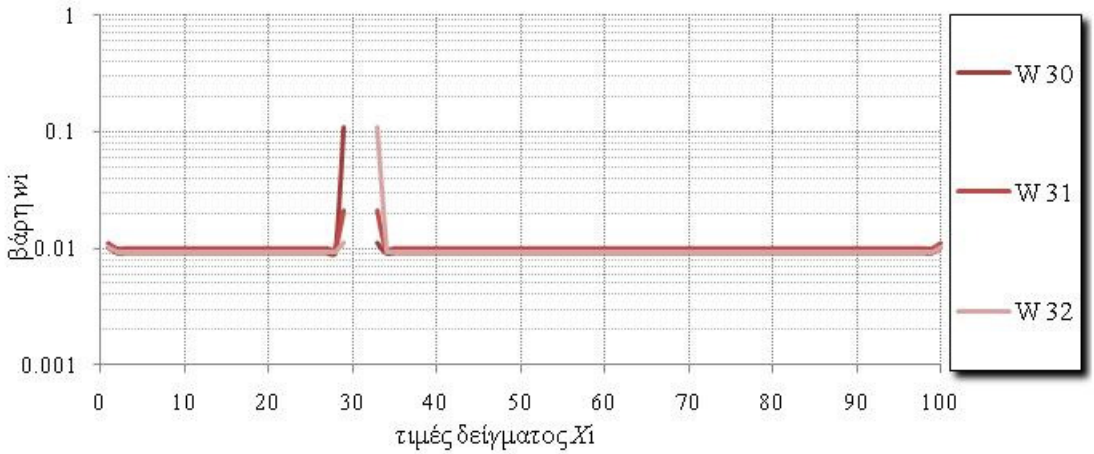



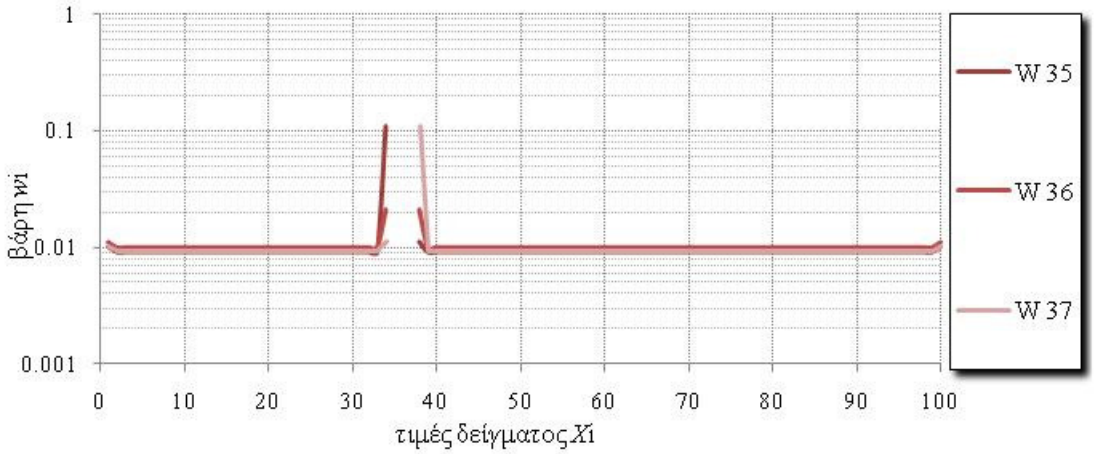



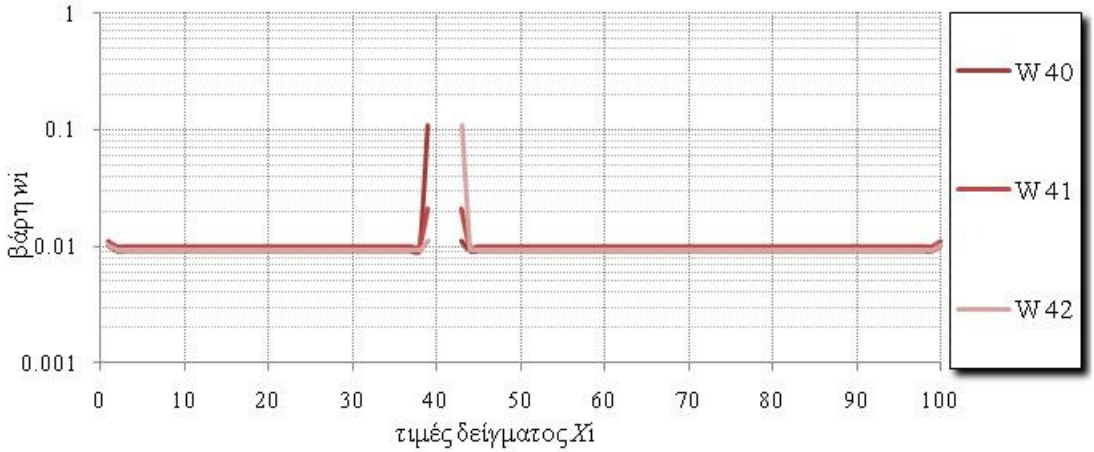

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 15
W 16
W 17

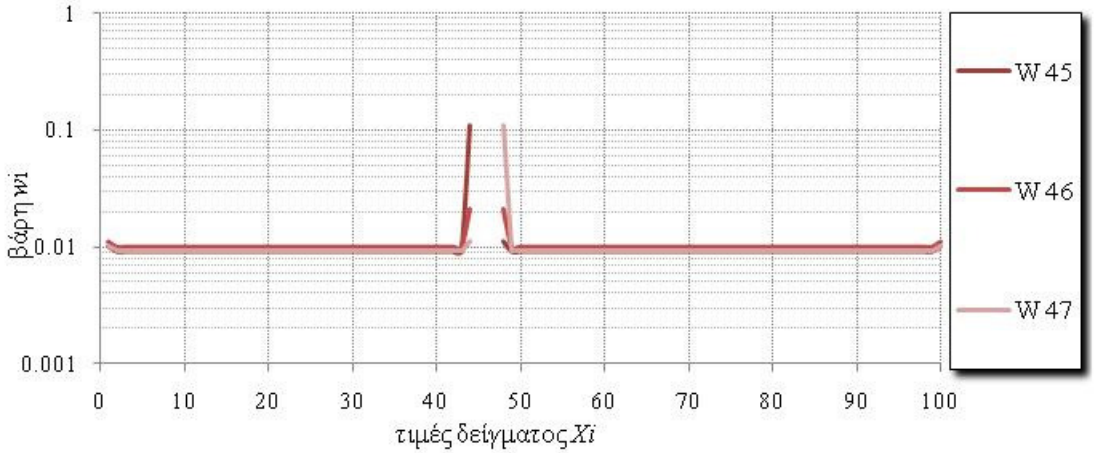

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 20
W 21
W 22

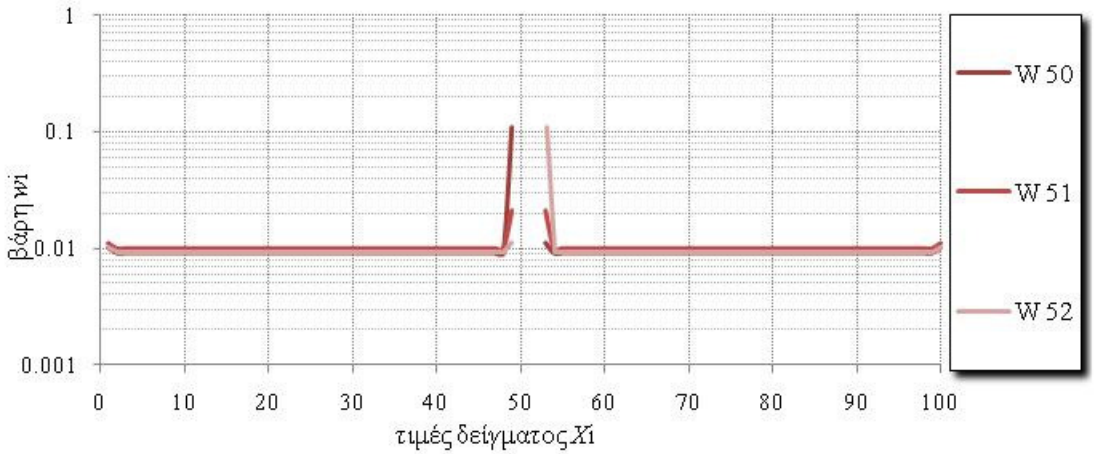

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 30
W 31
W 32

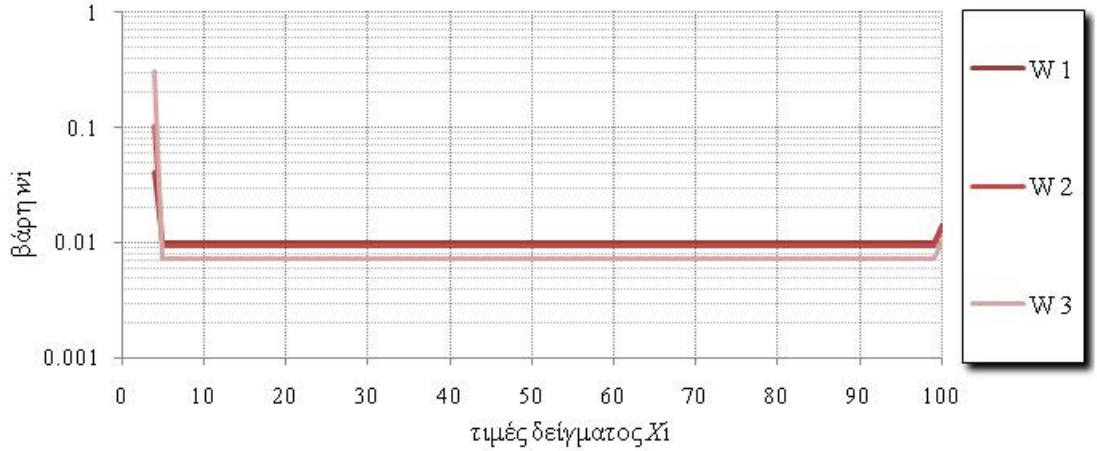

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 35
W 36
W 37

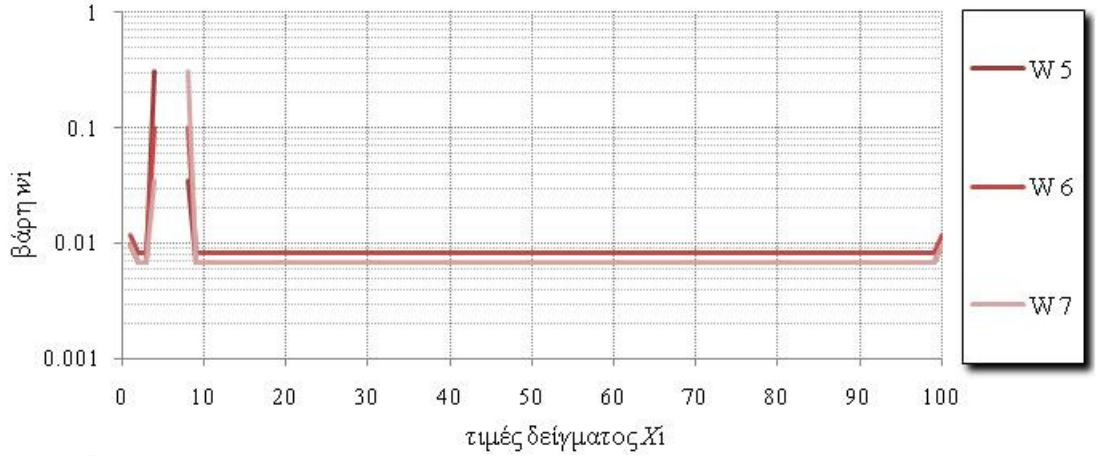

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 40
W 41
W 42

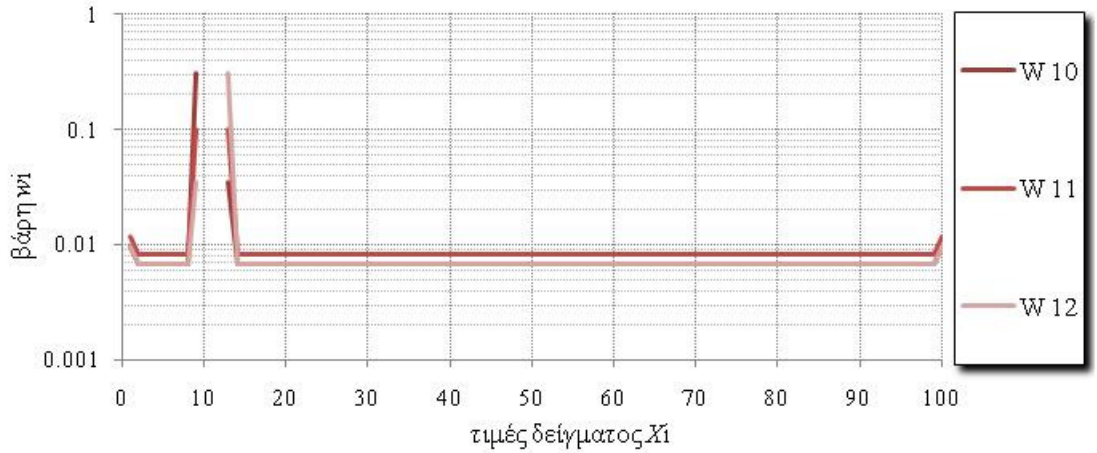

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 45
W 46
W 47

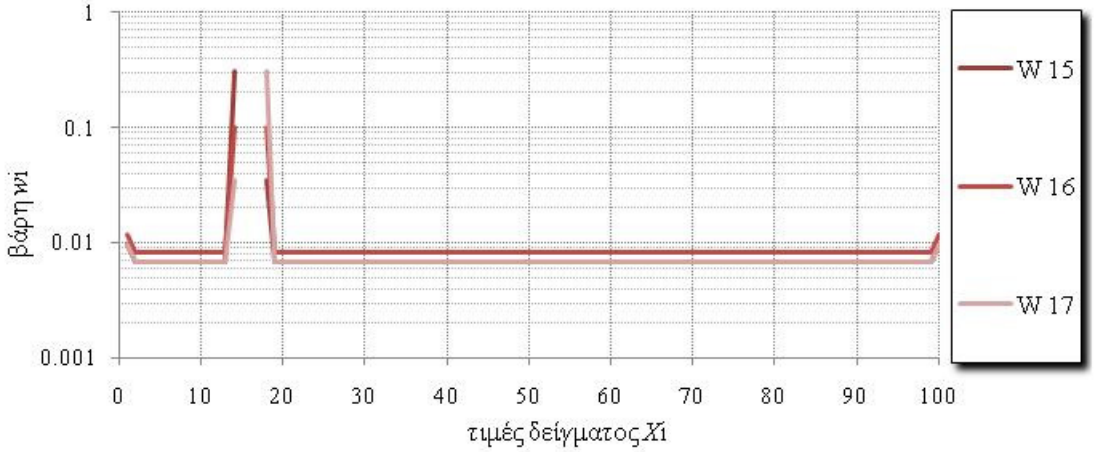

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 50
W 51
W 52

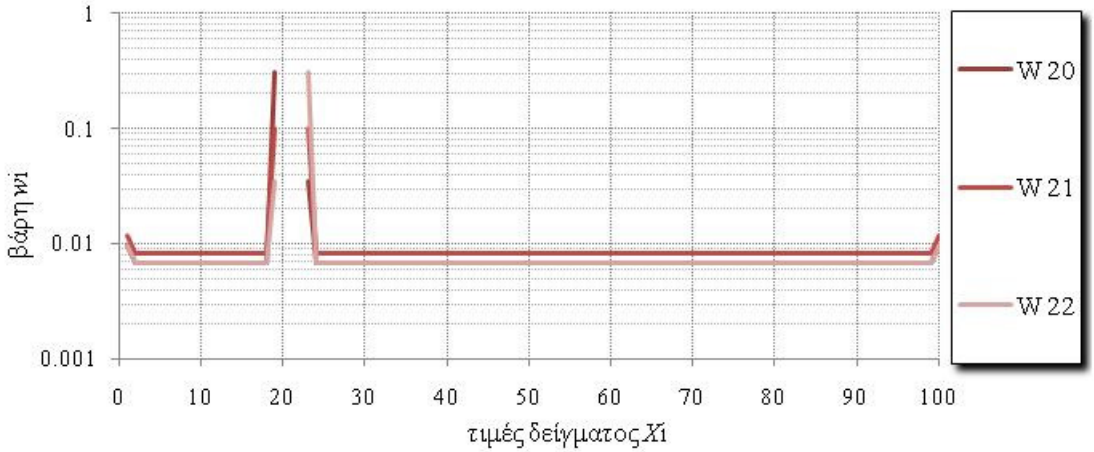
- Για δείγμα μεγέθους 100 τιμών και $\rho_1$=0.3 έχουμε:

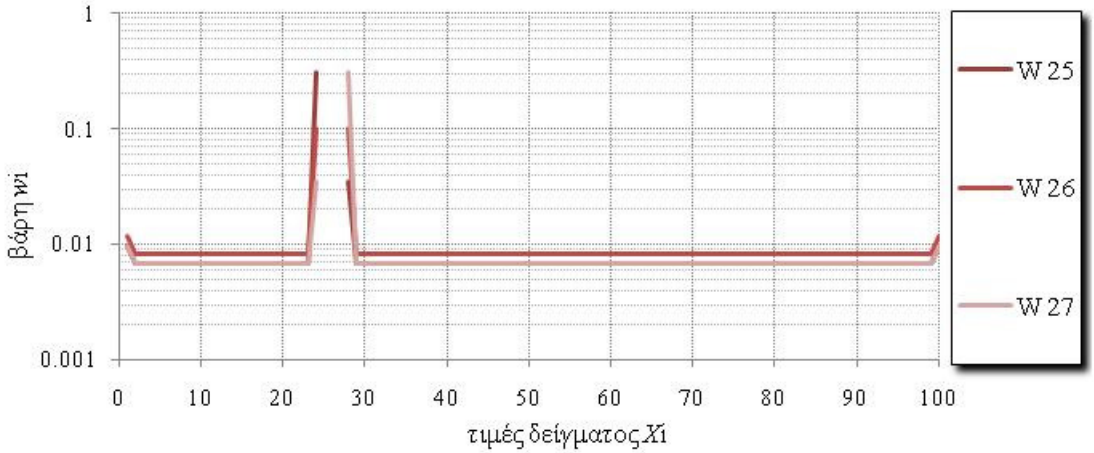


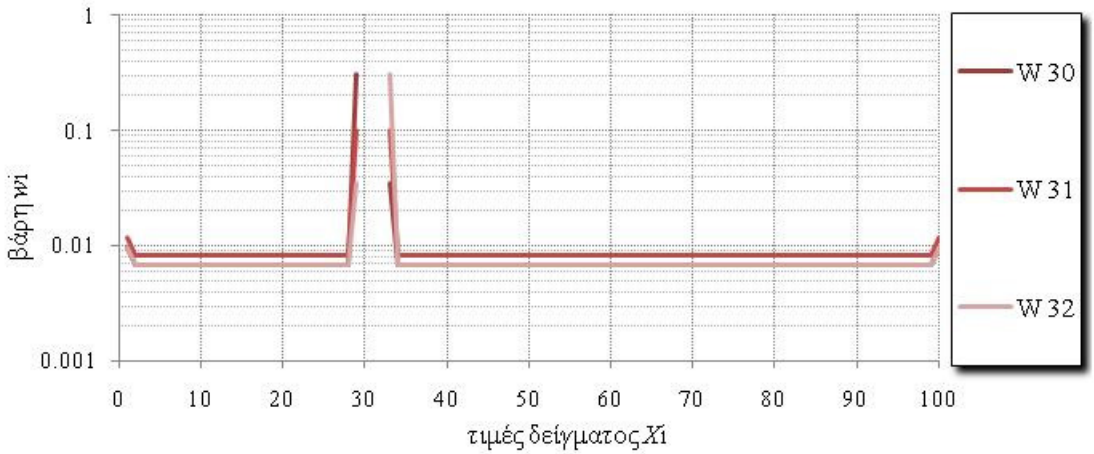



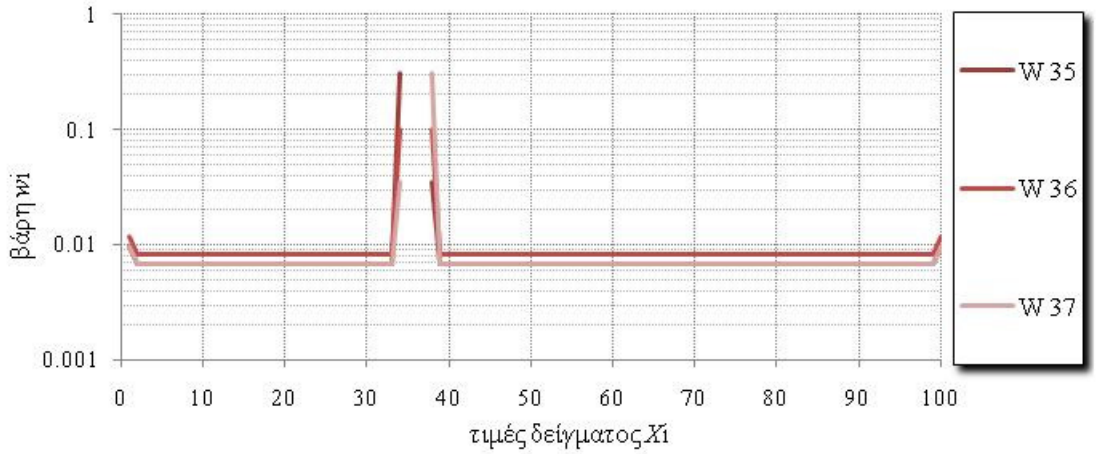



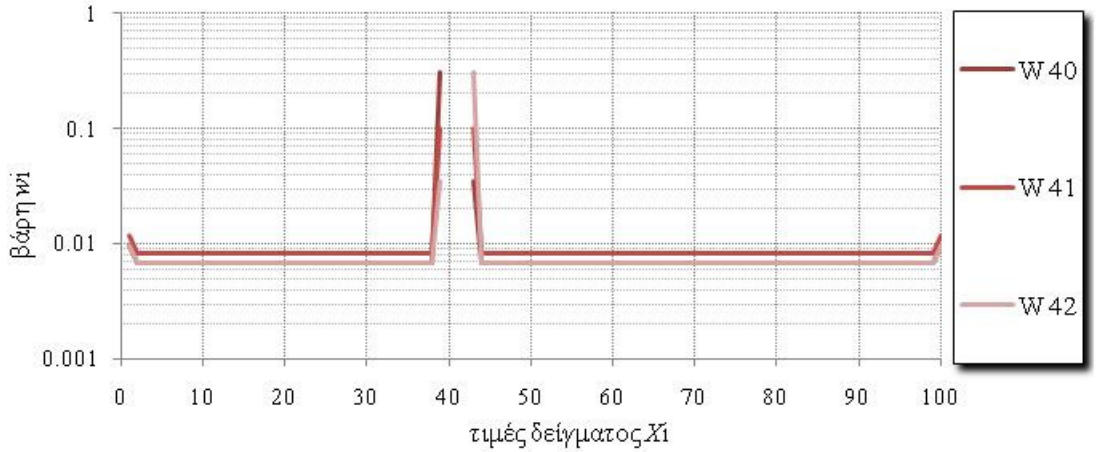

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 15
W 16
W 17

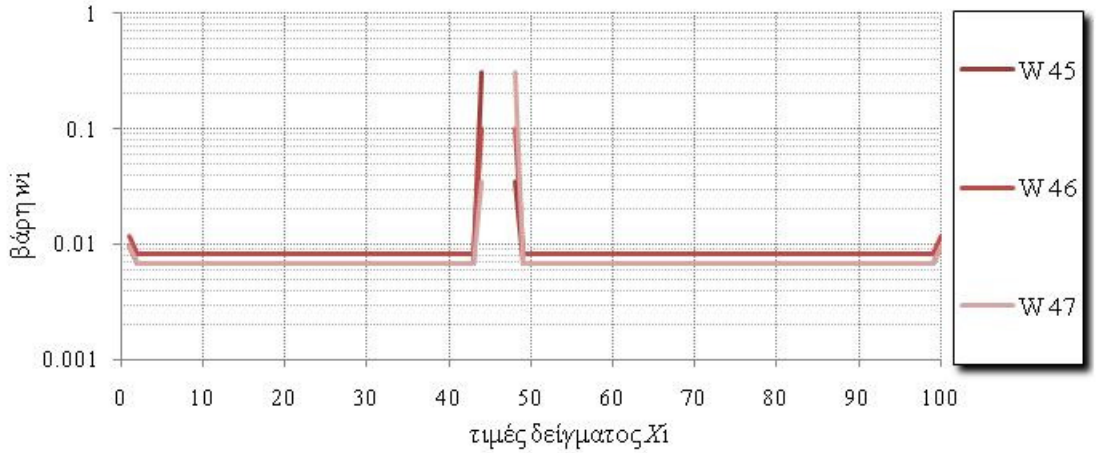

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 20
W 21
W 22

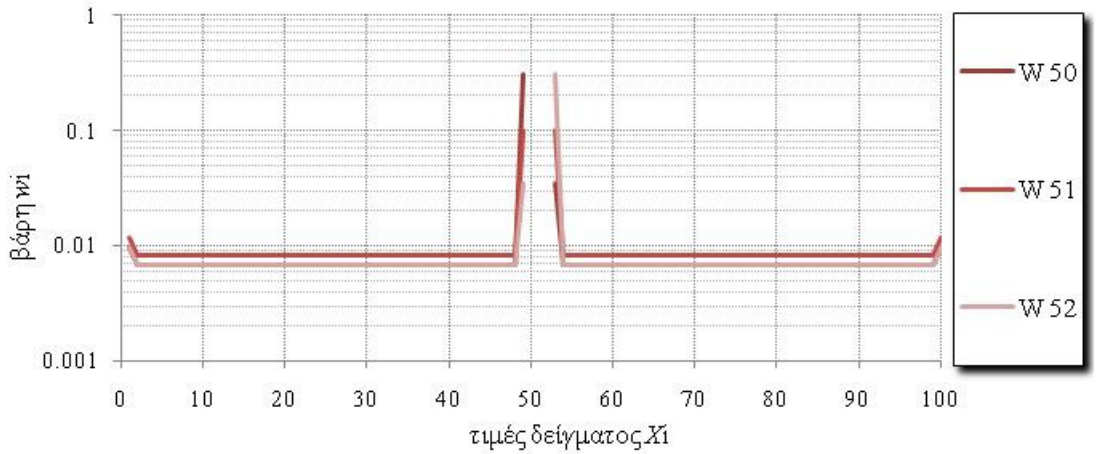

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 30
W 31
W 32

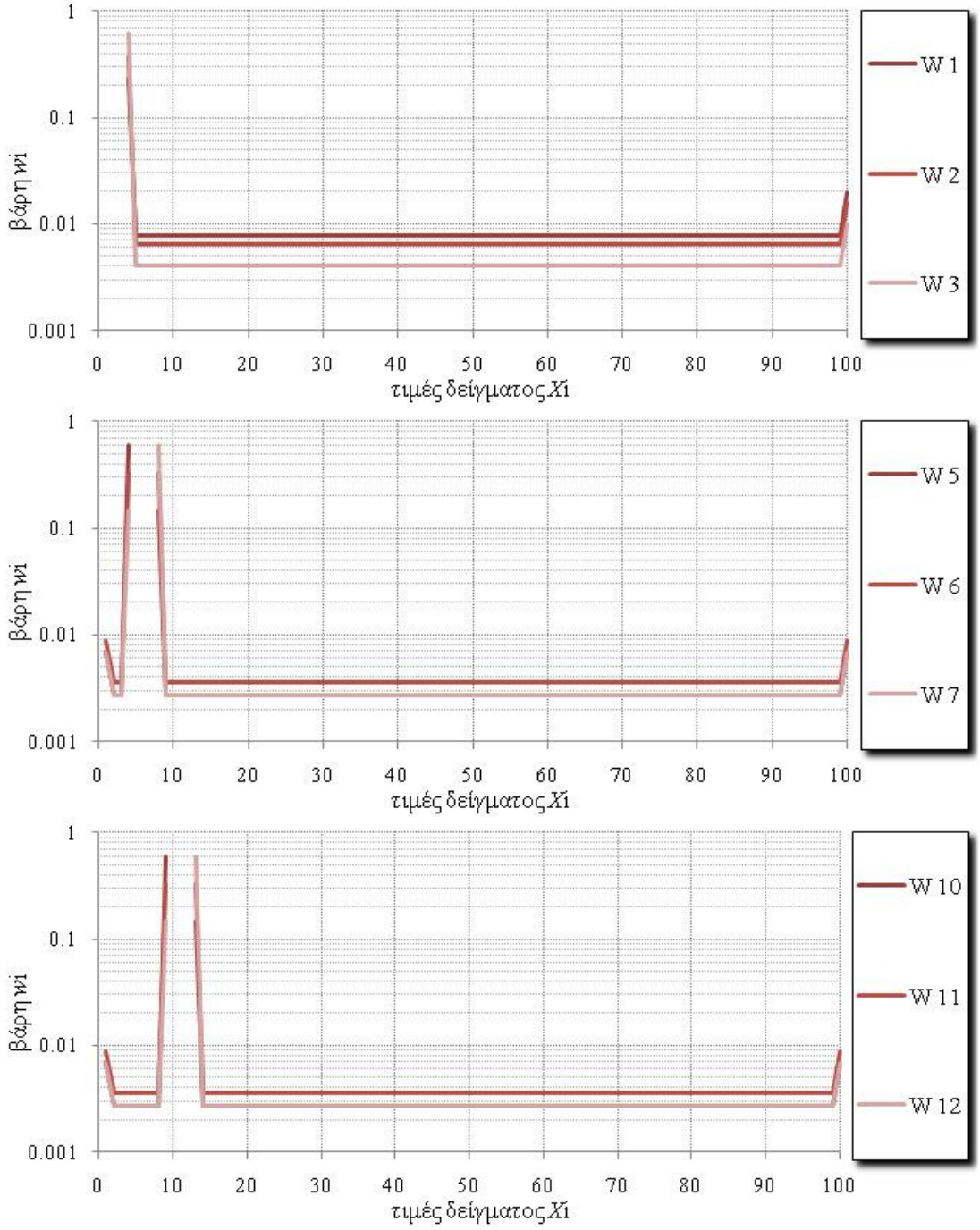

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 35
W 36
W 37

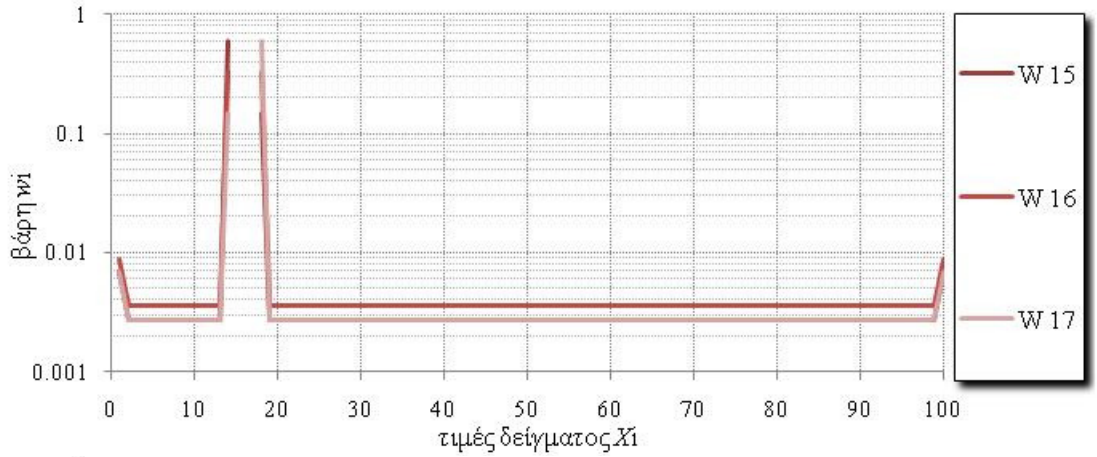

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 40
W 41
W 42

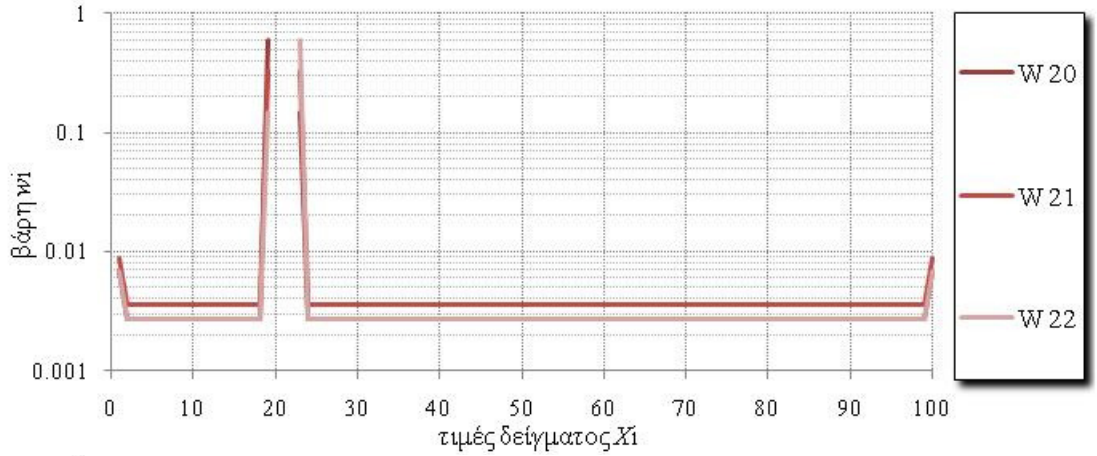

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 45
W 46
W 47

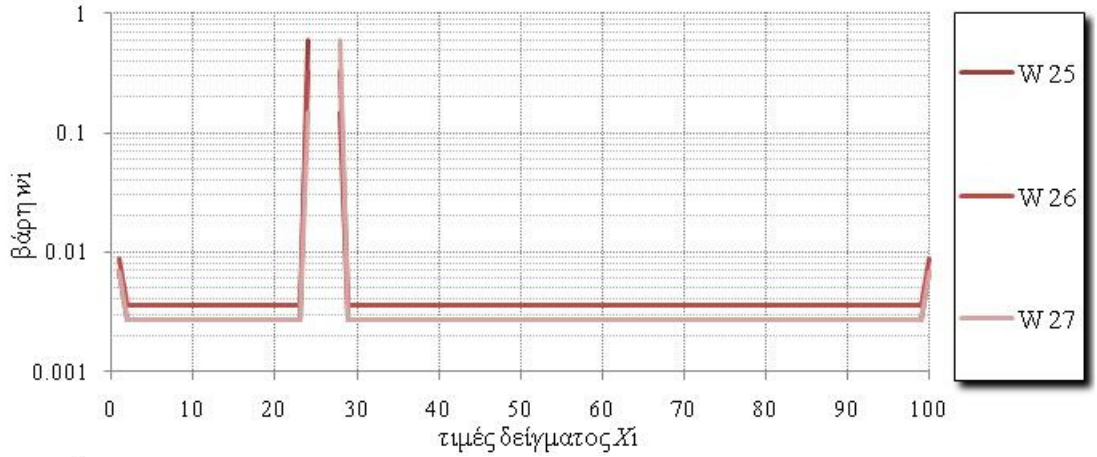

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 50
W 51
W 52

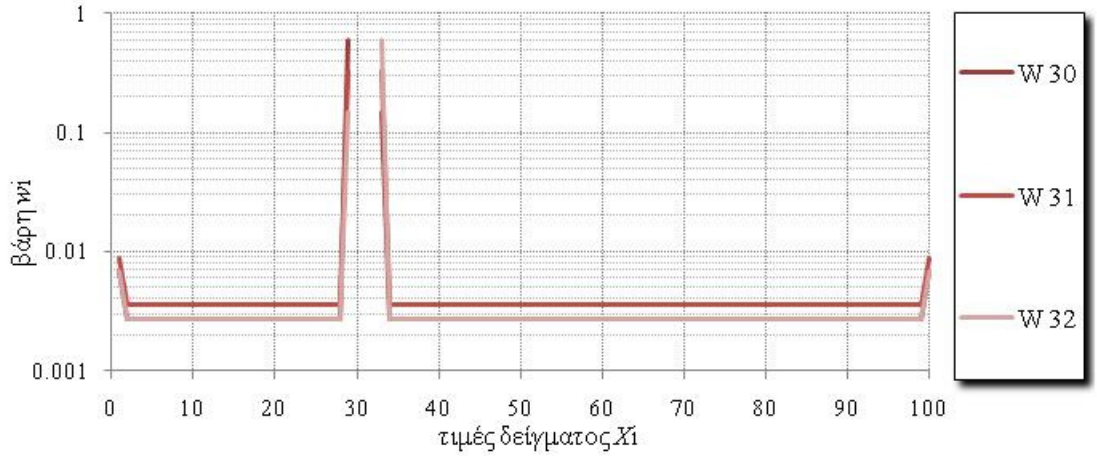
- Για δείγμα μεγέθους 100 τιμών και $\rho_1$=0.6 έχουμε:

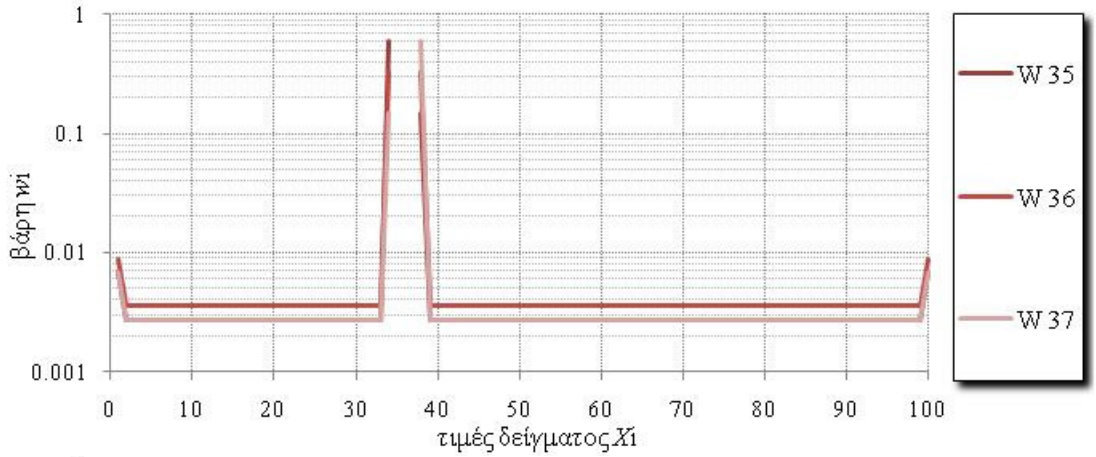


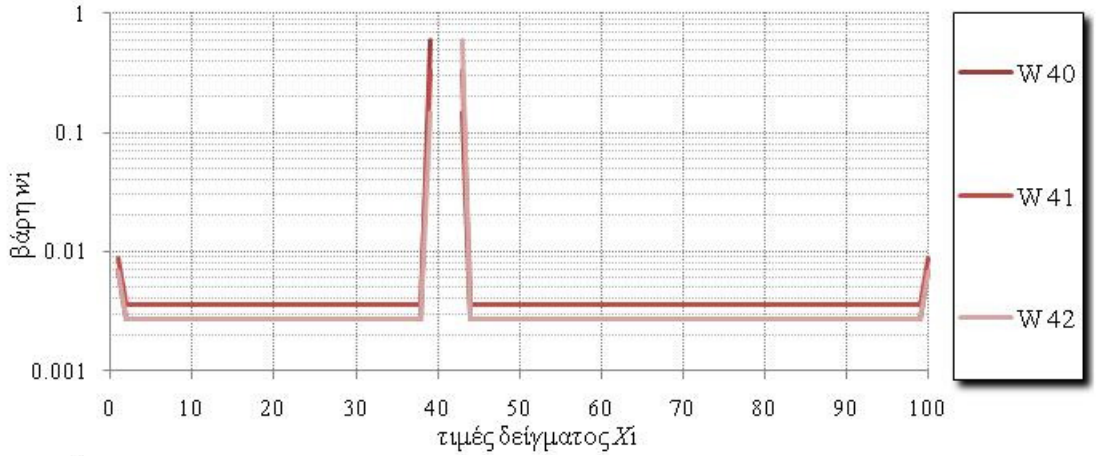

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 15
W 16
W 17

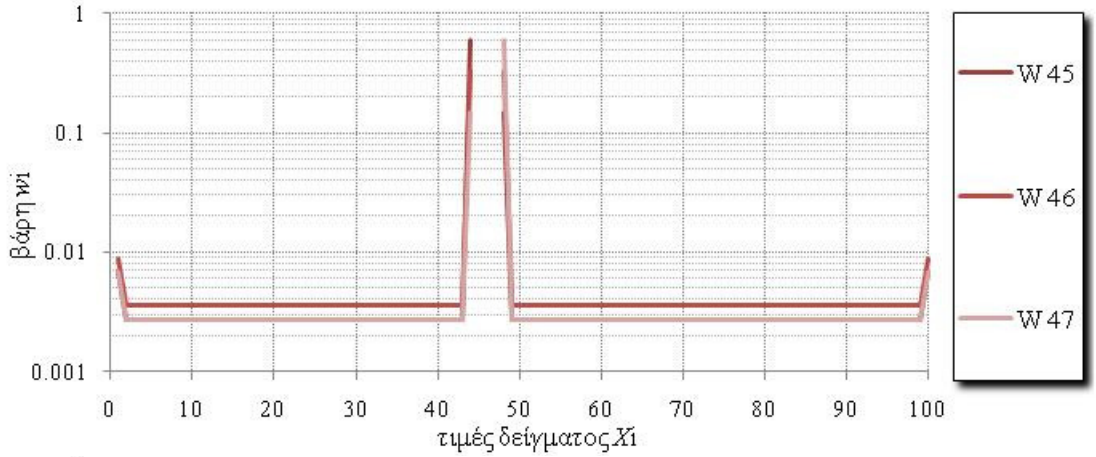

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 20
W 21
W 22

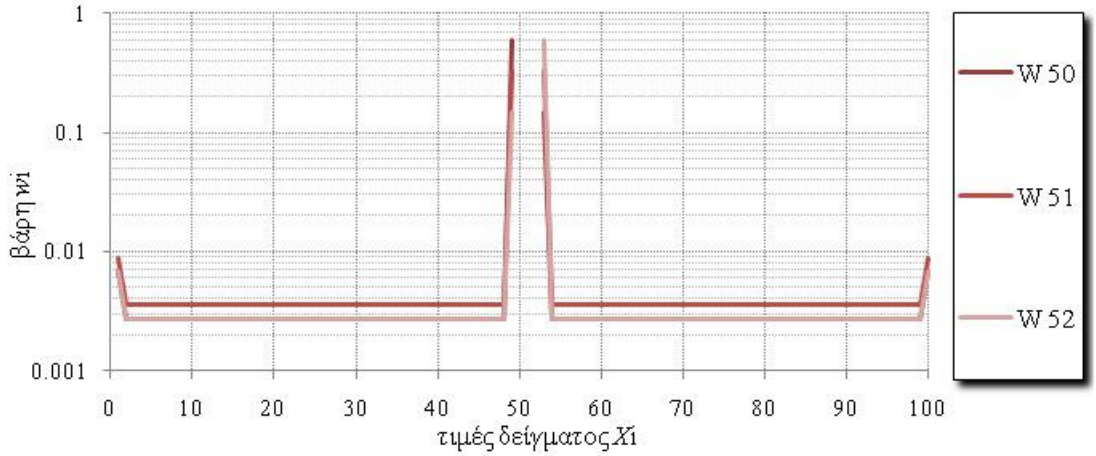

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 30
W 31
W 32

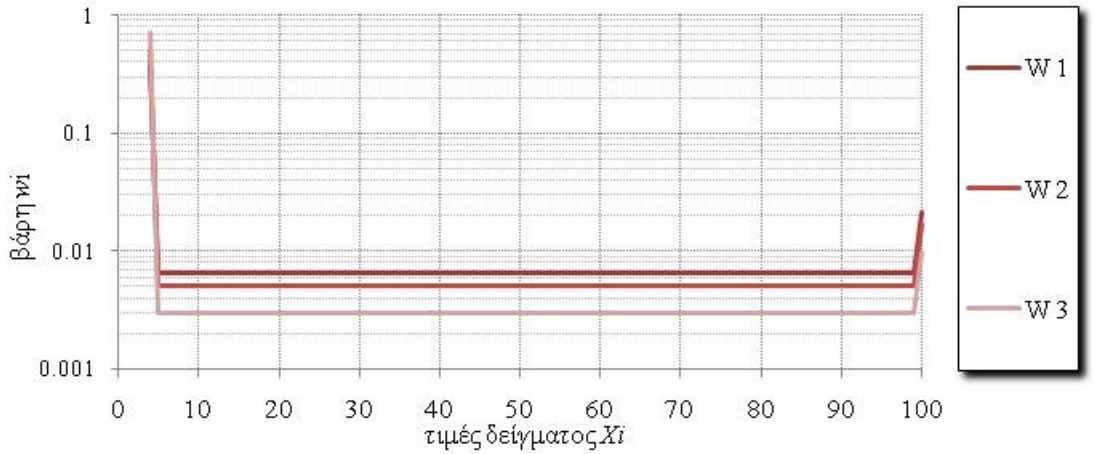

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 35
W 36
W 37

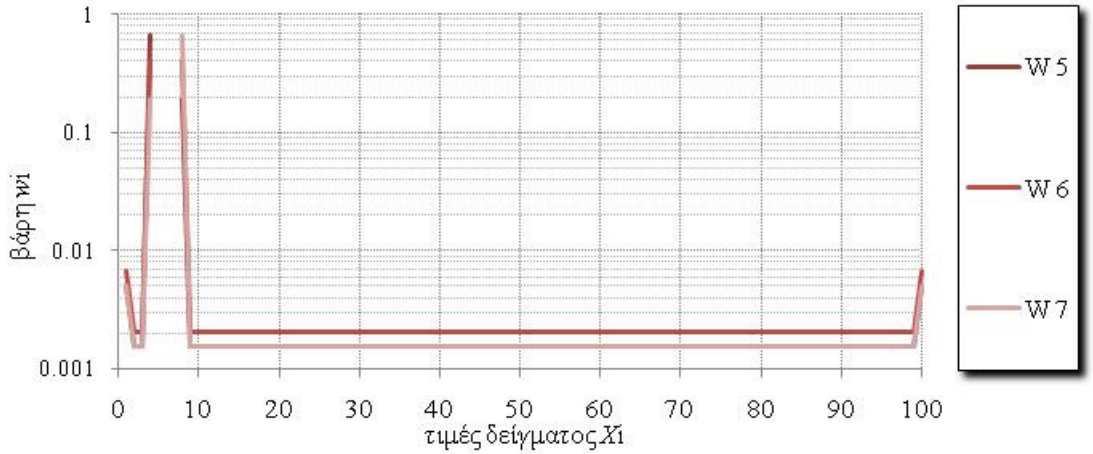

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 40
W 41
W 42

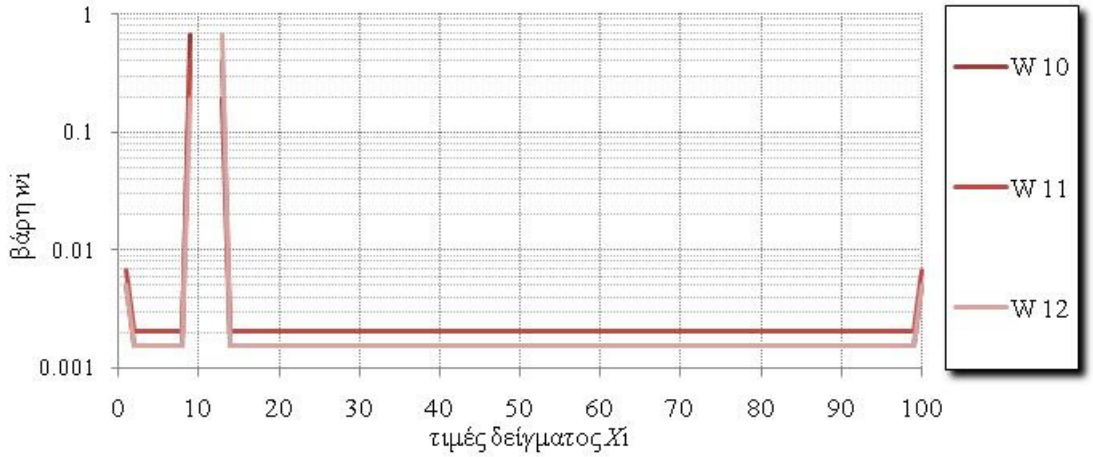

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 45
W 46
W 47

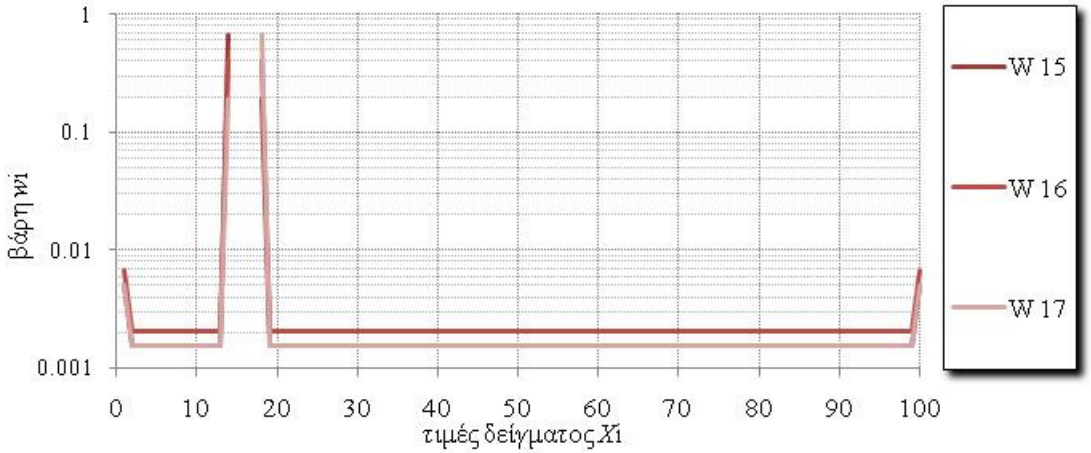

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 50
W 51
W 52

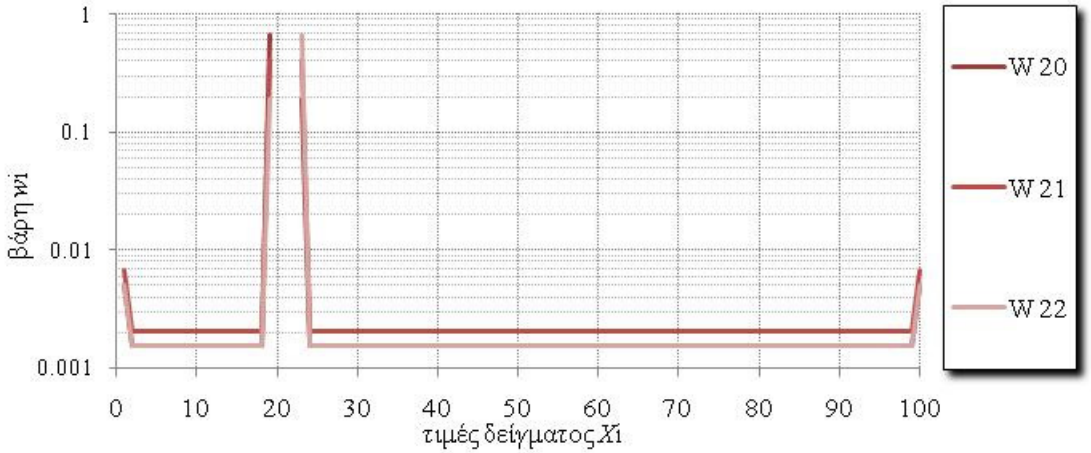
- Για δείγμα μεγέθους 100 τιμών και $\rho_1$=0.7 έχουμε:

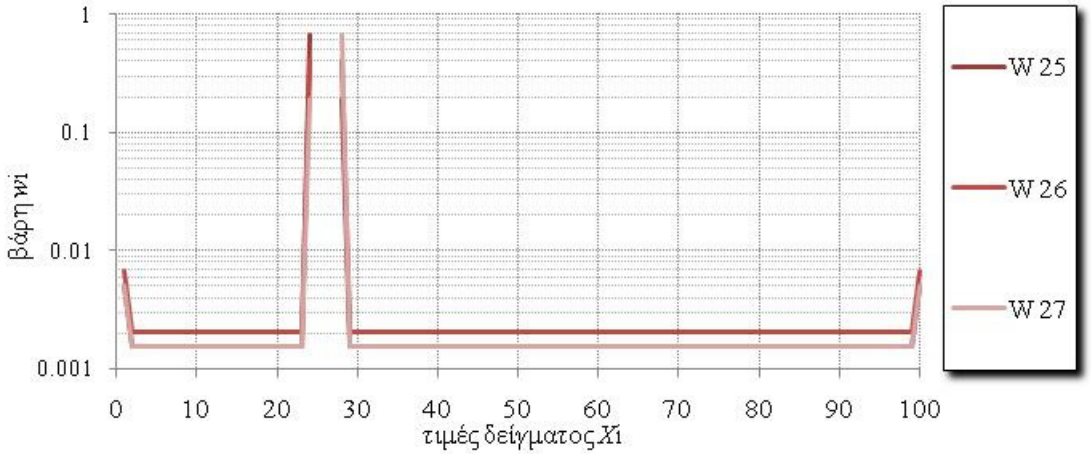


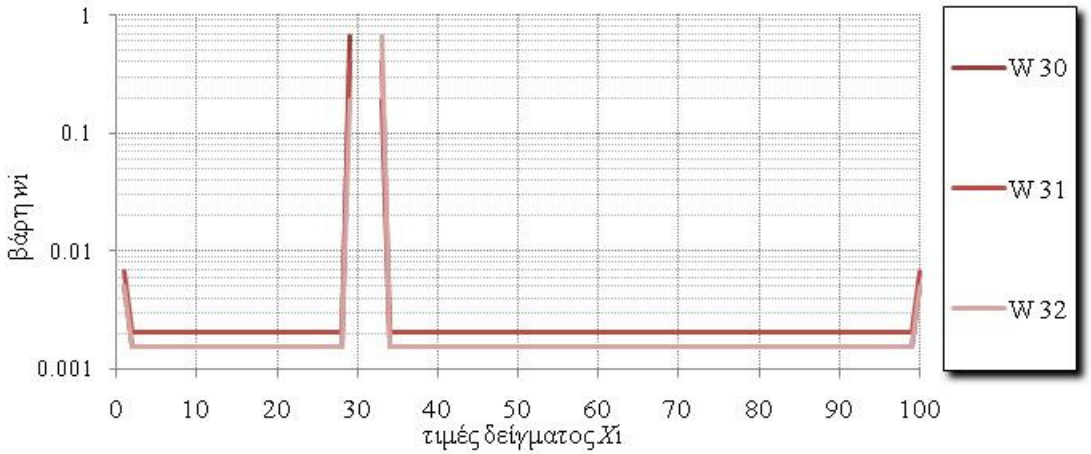



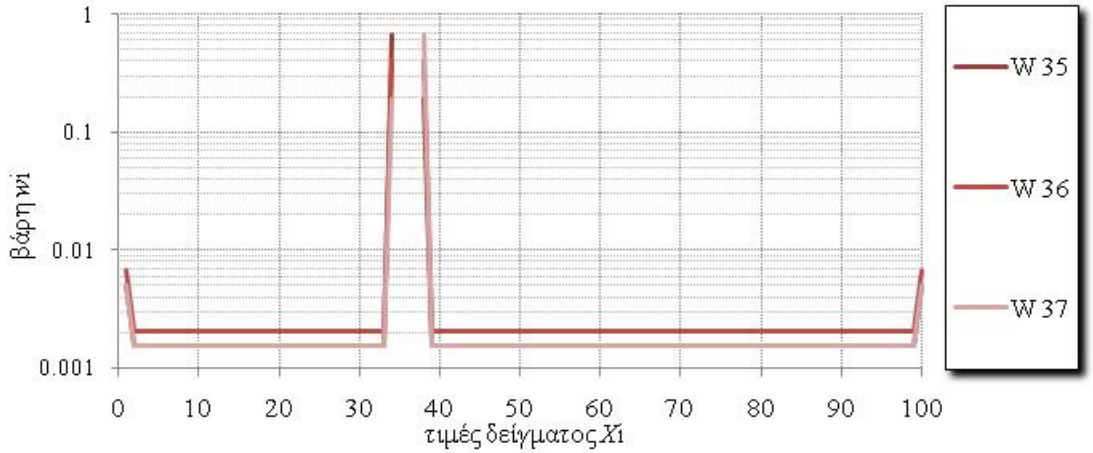



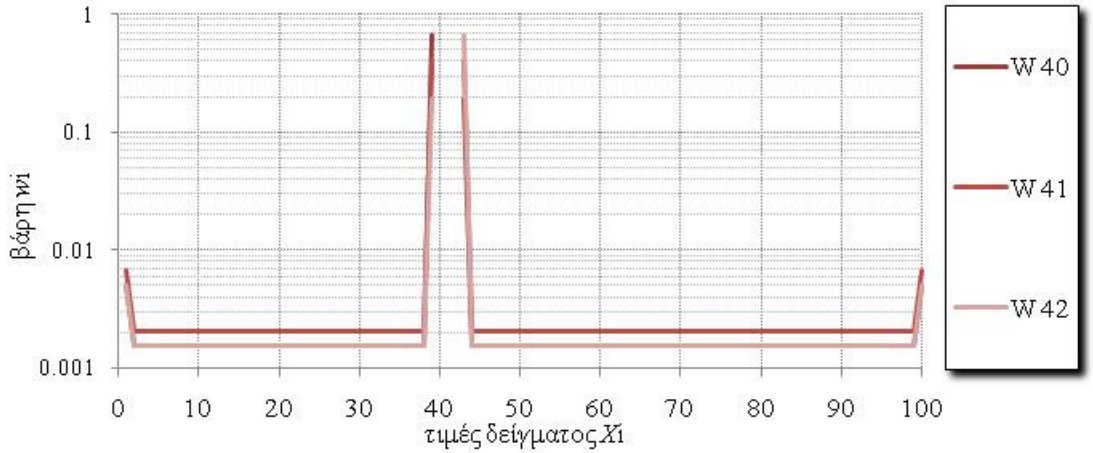

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 15
W 16
W 17

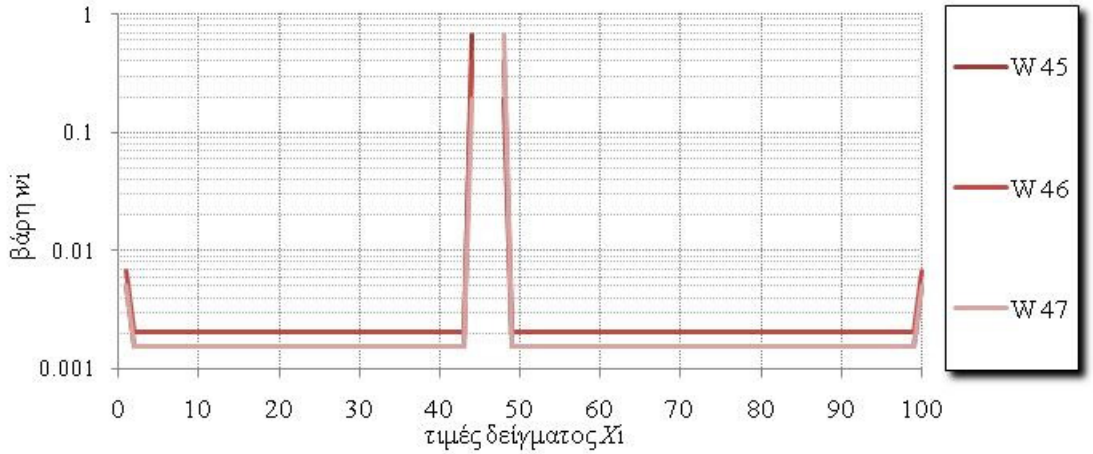

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 20
W 21
W 22

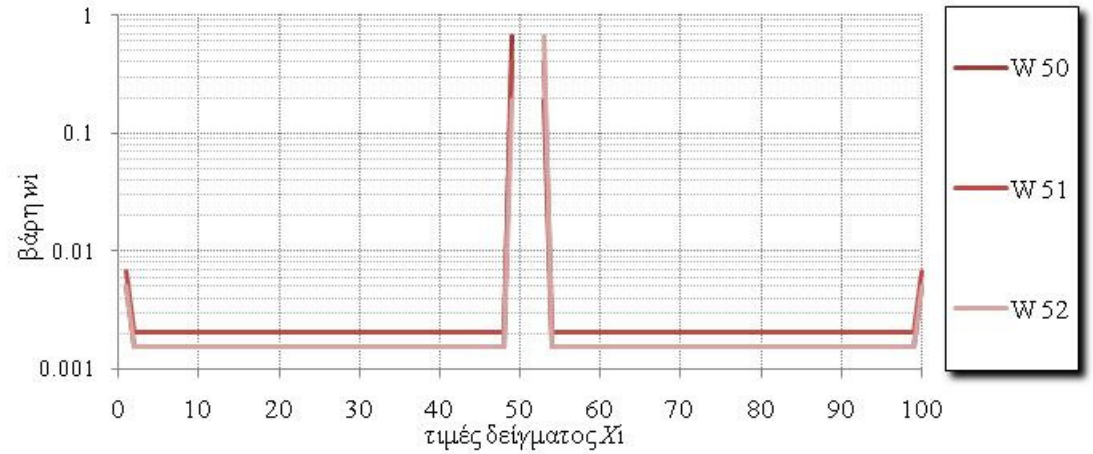

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 30
W 31
W 32

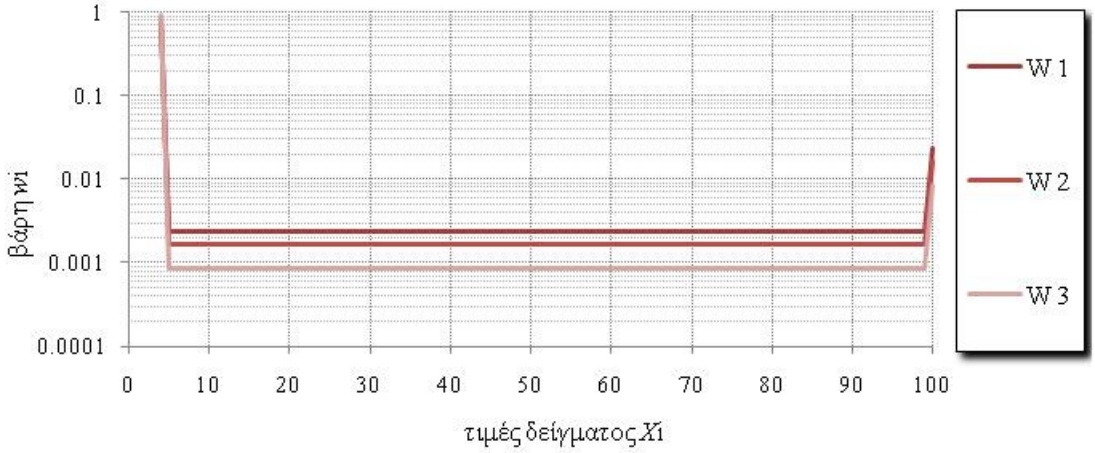

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 35
W 36
W 37

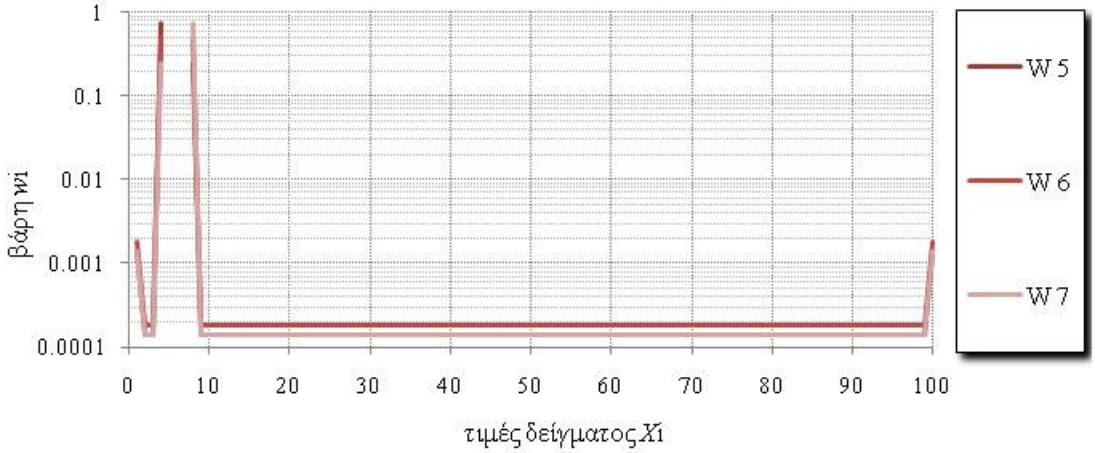

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 40
W 41
W 42

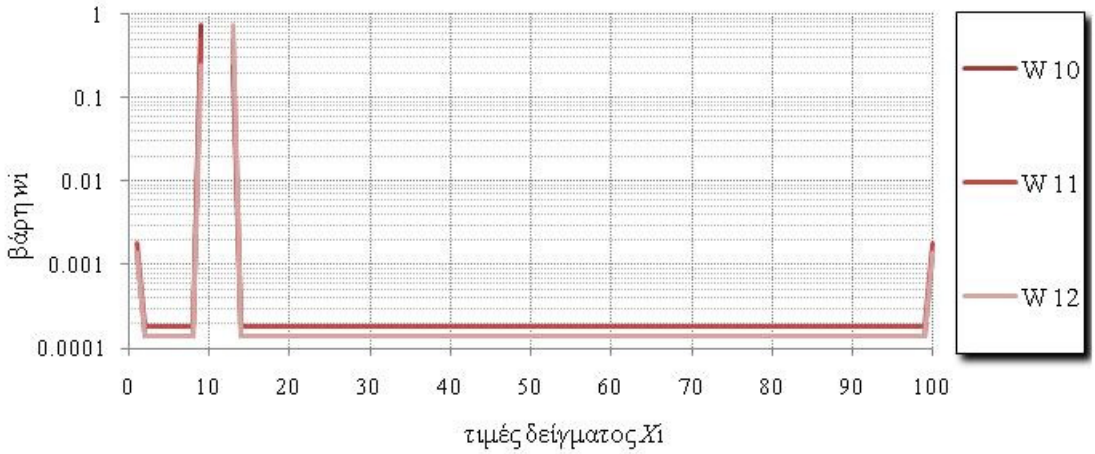

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 45
W 46
W 47

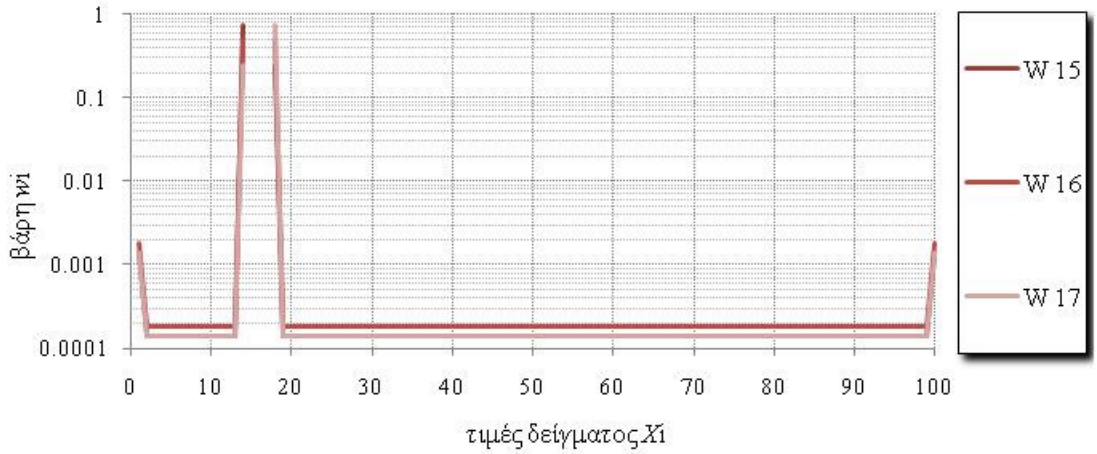

1
0.1
0.01
0.001
βάρη wi
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 50
W 51
W 52

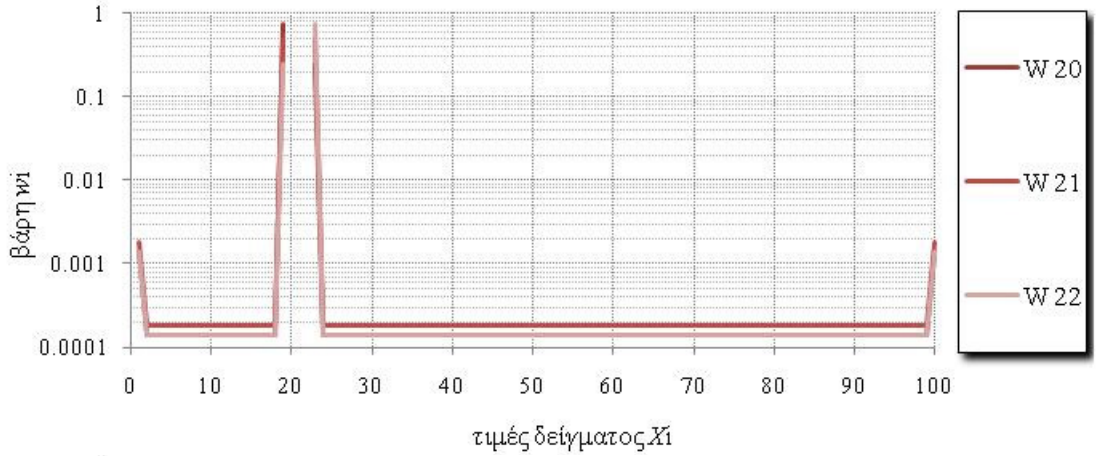
- Για δείγμα μεγέθους 100 τιμών και $\rho_1$=0.9 έχουμε:

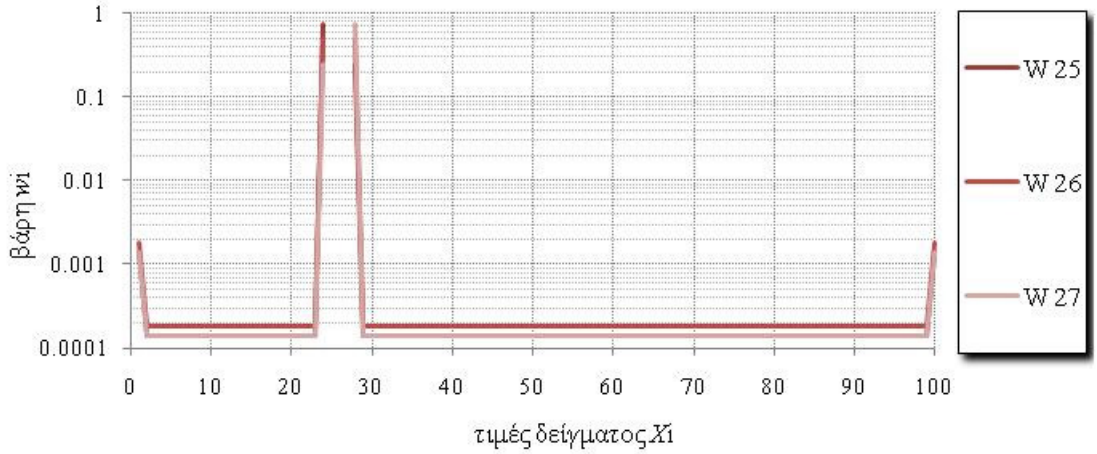


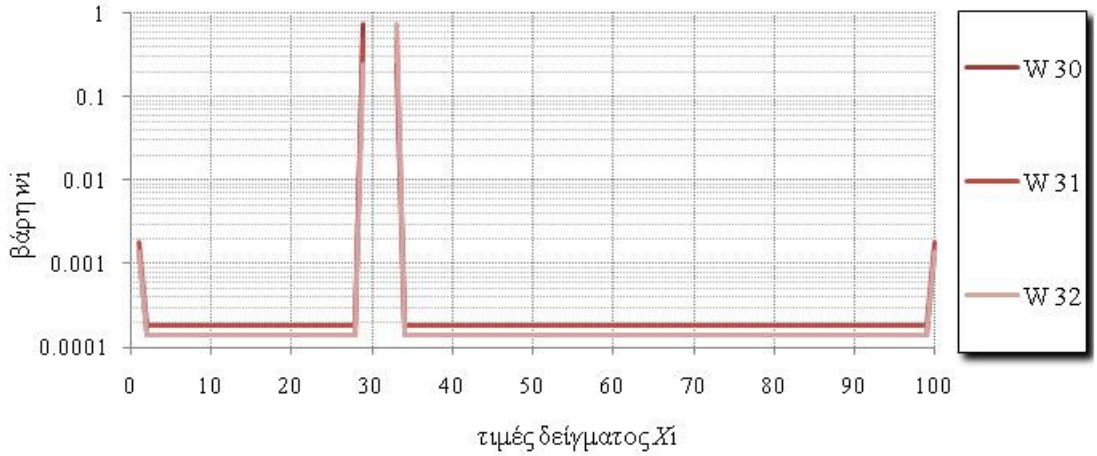



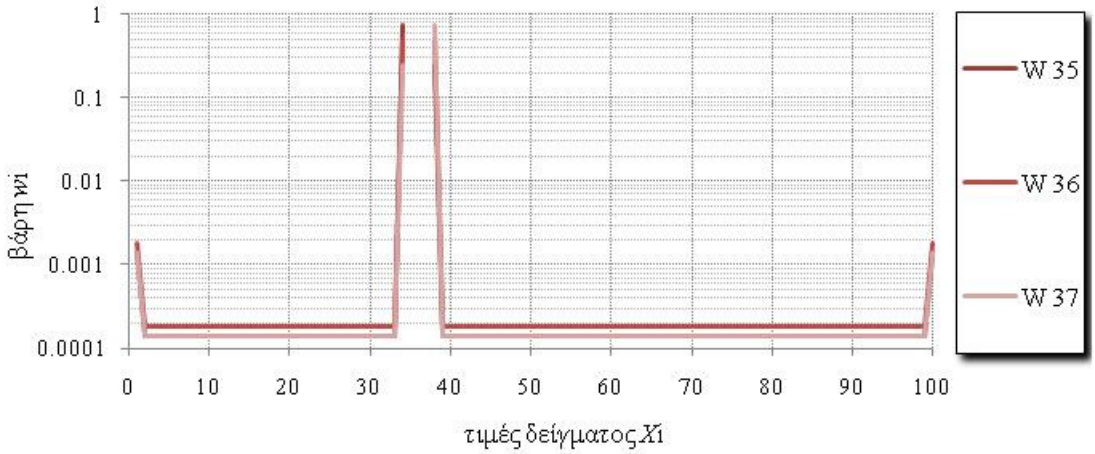



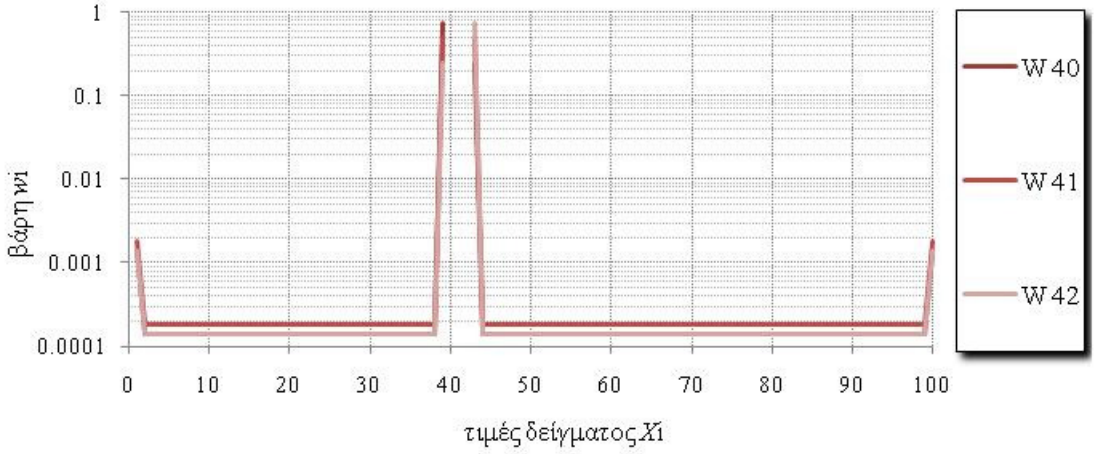

βάρη wi
1
0.1
0.01
0.001
0.0001
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 15
W 16
W 17

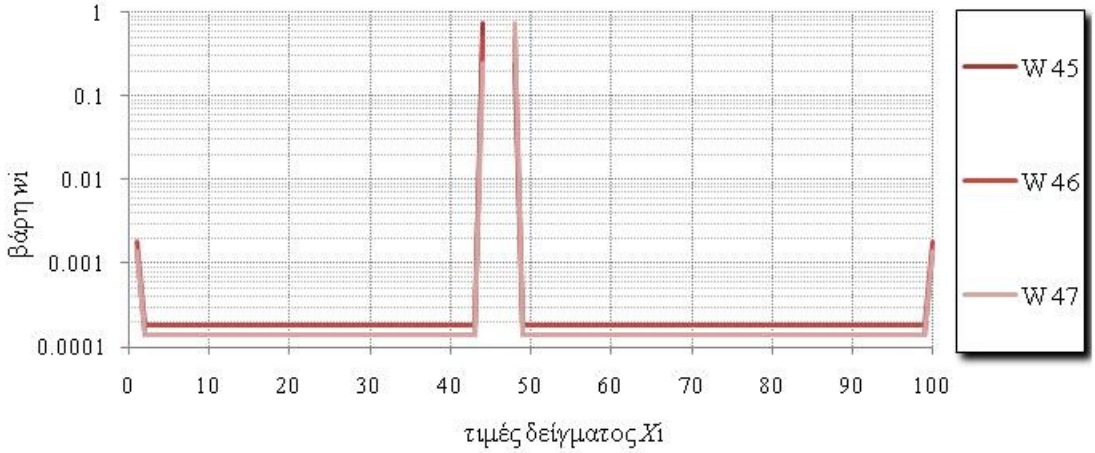

βάρη wi
1
0.1
0.01
0.001
0.0001
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 20
W 21
W 22

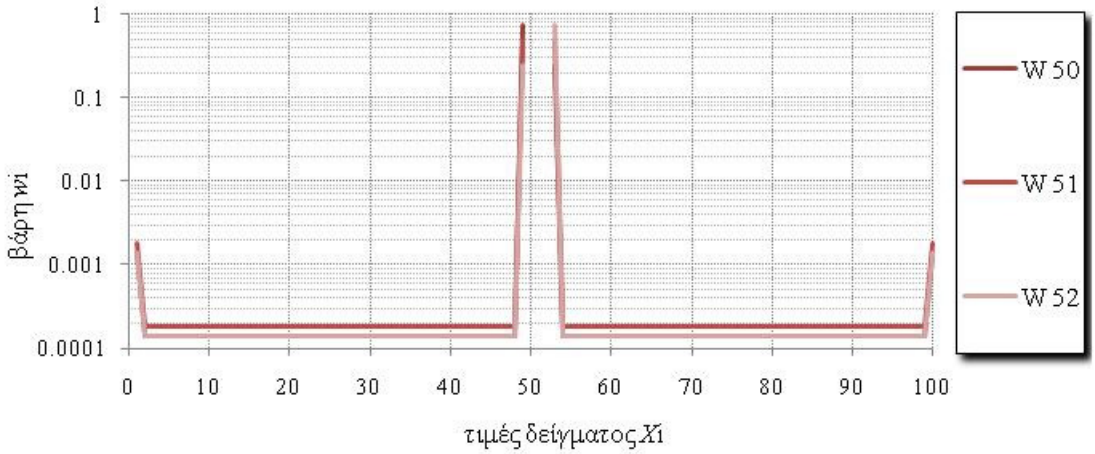

βάρη wi
1
0.1
0.01
0.001
0.0001
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 25
W 26
W 27


βάρη wi
1
0.1
0.01
0.001
0.0001
0 10 20 30 40 50 60 70 80 90 100
τιμές δείγματος Xi
W 30
W 31
W 32


βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 35
W 36
W 37


βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 40
W 41
W 42


βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 45
W 46
W 47

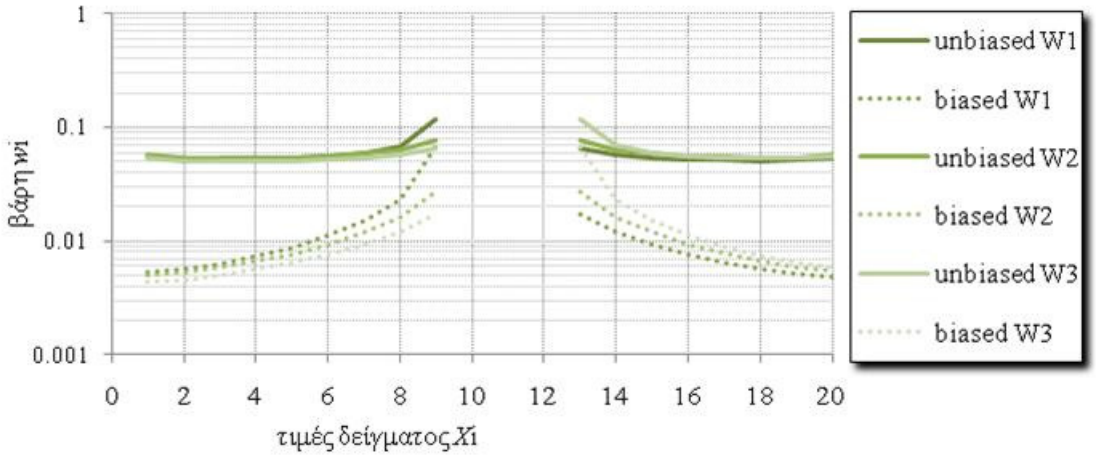

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 50
W 51
W 52

## ΠΑΡΑΡΤΗΜΑ G

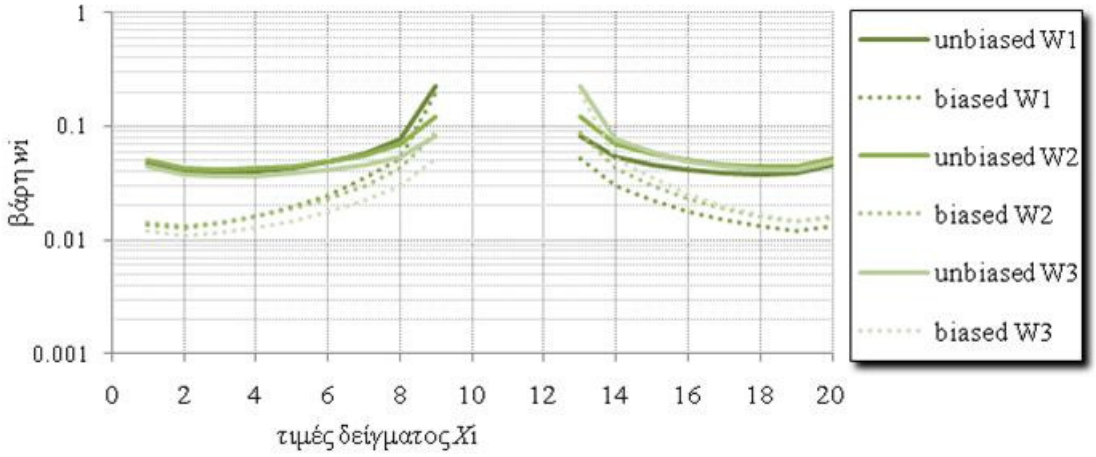
Συμπλήρωση τριών συνεχόμενων ελλειπουσών τιμών με σταθμισμένα βάρη $w_i$ σε υδρομετεωρολογικές χρονοσειρές που παρουσιάζουν δυναμική Hurst-Kolmogorov.

Ανάλογα με το συντελεστή Hurst ($H$), κατασκευάστηκαν τα πιο κάτω διαγράμματα που παρουσιάζουν τα σταθμισμένα βάρη των τιμών της χρονοσειράς, ανάλογα με την θέση της ελλείπουσας τιμής.

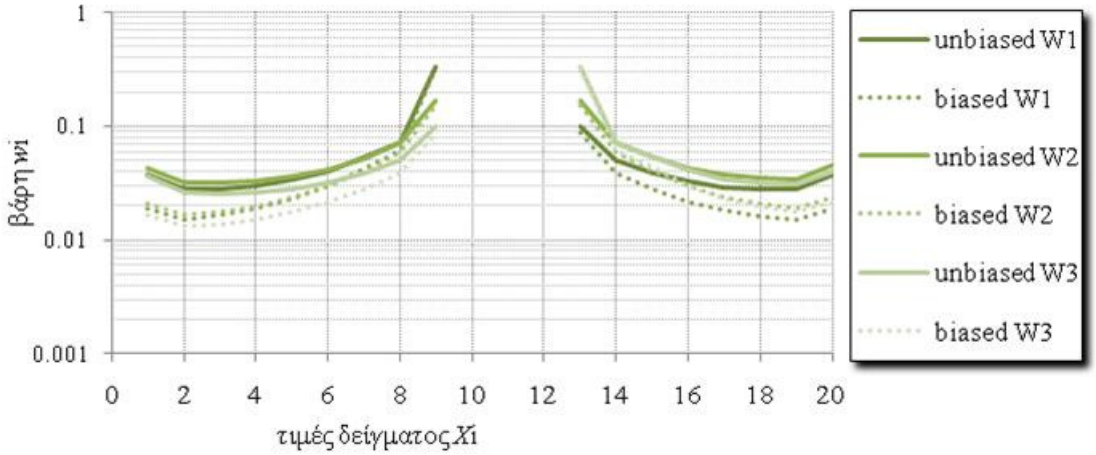
- Για δείγμα μεγέθους 20 τιμών, αν λείπουν οι $10^η$, $11^η$ και $12^η$ τιμή, οι αμερόληπτες (συνεχείς γραμμές) και οι μεροληπτικές (διακεκομμένες γραμμές) τιμές των βαρών $w_i$ για $H$=0.55 είναι:



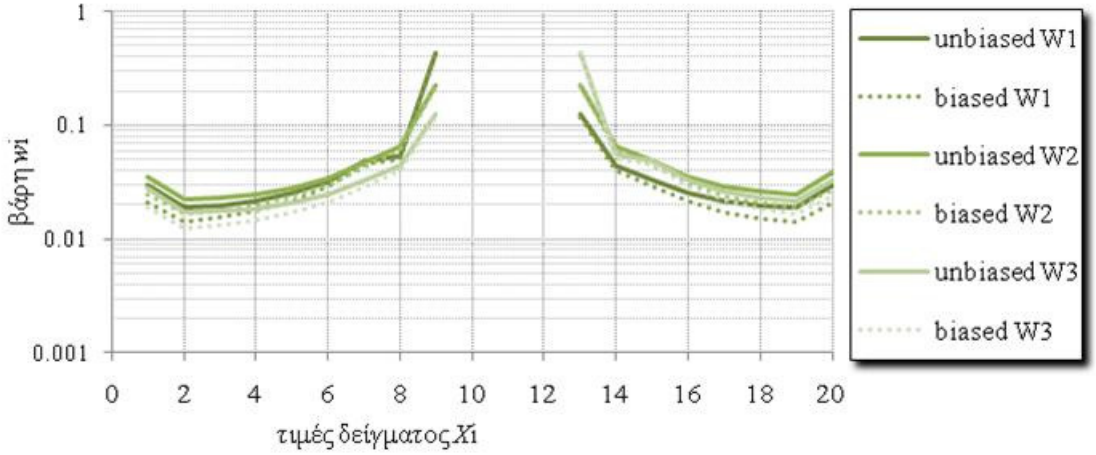

- Για δείγμα μεγέθους 20 τιμών, αν λείπουν οι $10^η$, $11^η$ και $12^η$ τιμή, οι αμερόληπτες (συνεχείς γραμμές) και οι μεροληπτικές (διακεκομμένες γραμμές) τιμές των βαρών $w_i$ για $H$=0.65 είναι:



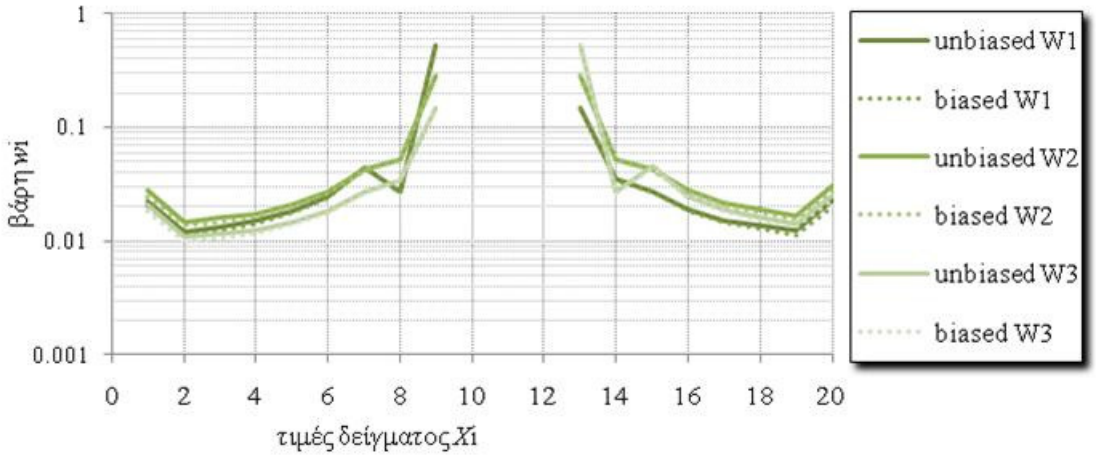
- Για δείγμα μεγέθους 20 τιμών, αν λείπουν οι $10^{η}$, $11^{η}$ και $12^{η}$ τιμή, οι αμερόληπτες (συνεχείς γραμμές) και οι μεροληπτικές (διακεκομμένες γραμμές) τιμές των βαρών $w_i$ για $H$=0.75 είναι:



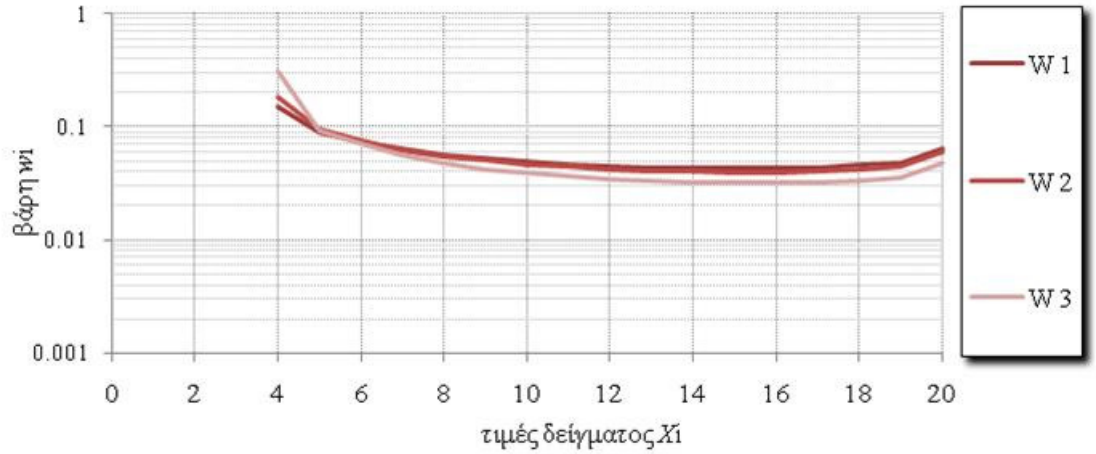

- Για δείγμα μεγέθους 20 τιμών, αν λείπουν οι $10^{η}$, $11^{η}$ και $12^{η}$ τιμή, οι αμερόληπτες (συνεχείς γραμμές) και οι μεροληπτικές (διακεκομμένες γραμμές) τιμές των βαρών $w_i$ για $H$=0.85 είναι:

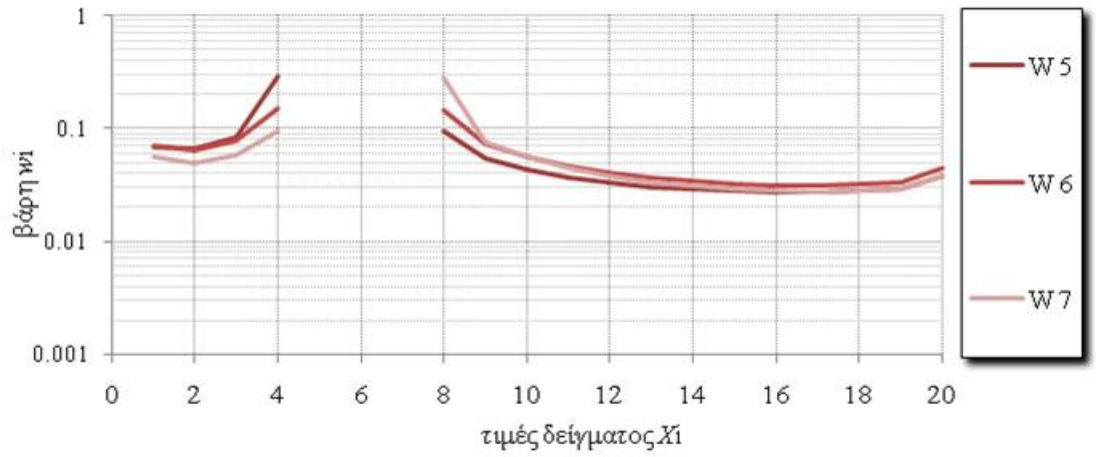


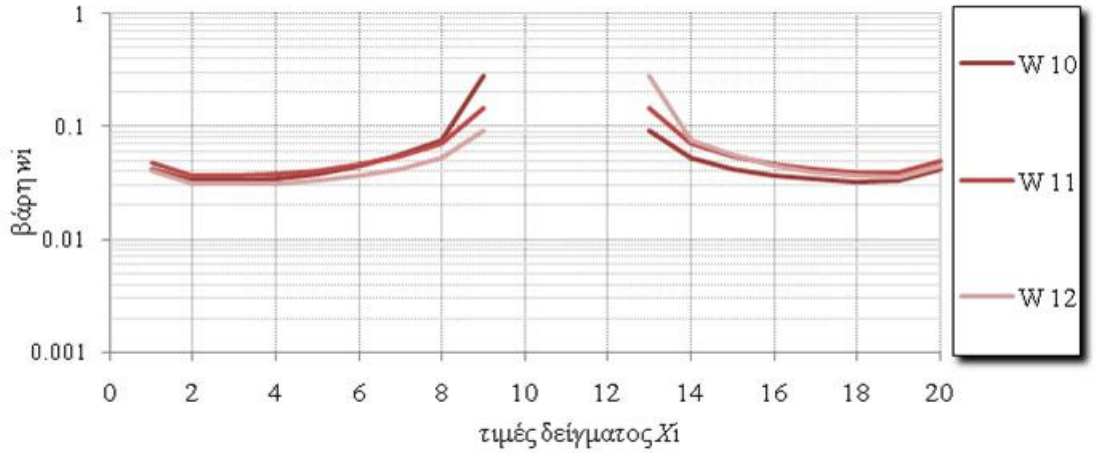

- Για δείγμα μεγέθους 20 τιμών, αν λείπουν οι $10^{η}$, $11^{η}$ και $12^{η}$ τιμή, οι αμερόληπτες (συνεχείς γραμμές) και οι μεροληπτικές (διακεκομμένες γραμμές) τιμές των βαρών $w_i$ για $H$=0.95 είναι:



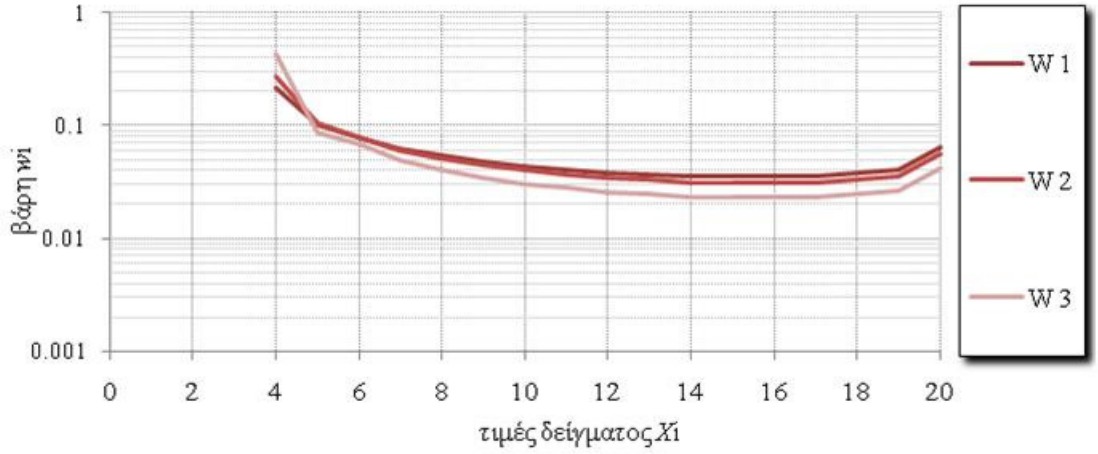

- Για δείγμα μεγέθους 20 τιμών και $H$=0.7 έχουμε:

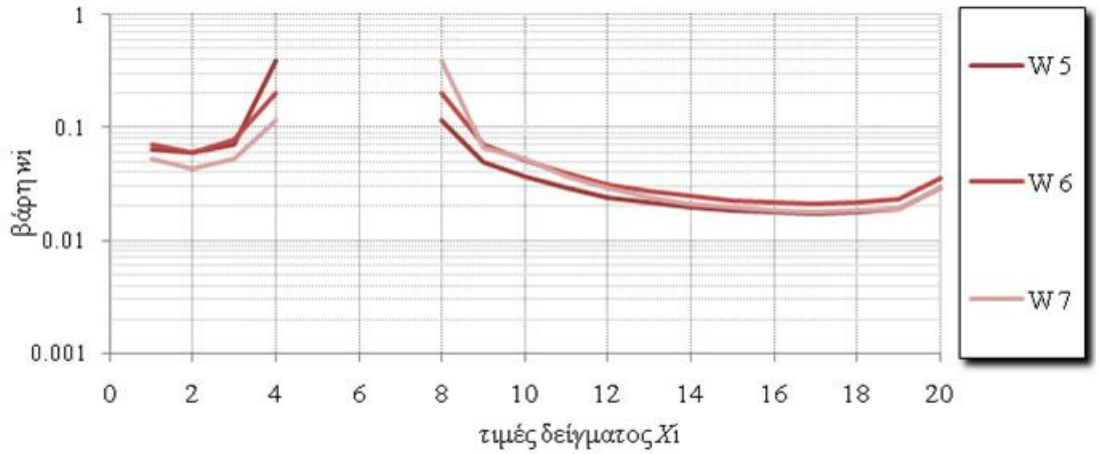


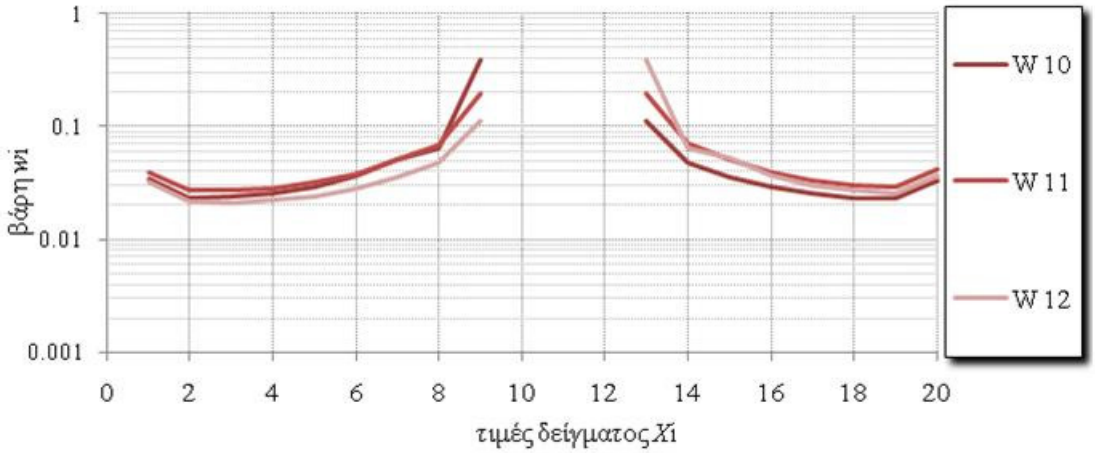





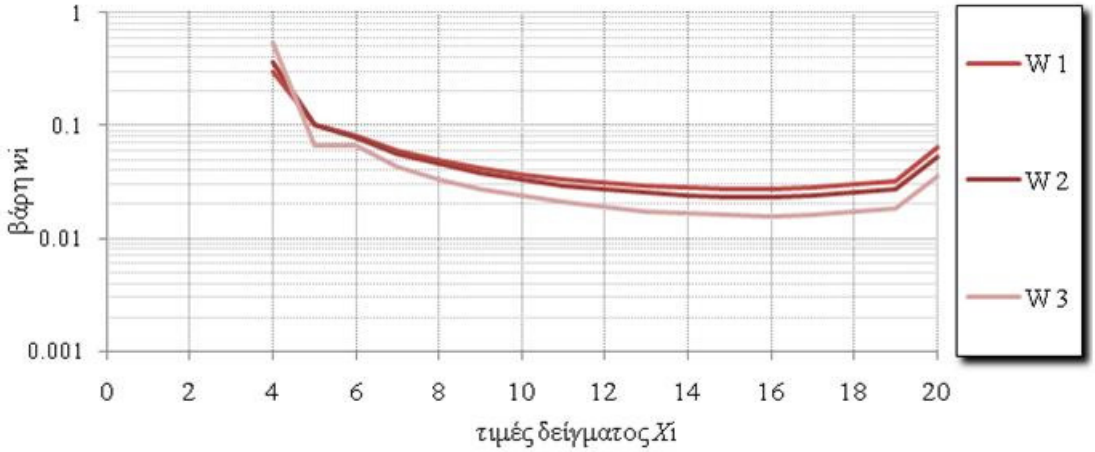

- Για δείγμα μεγέθους 20 τιμών και *H*=0.8 έχουμε:

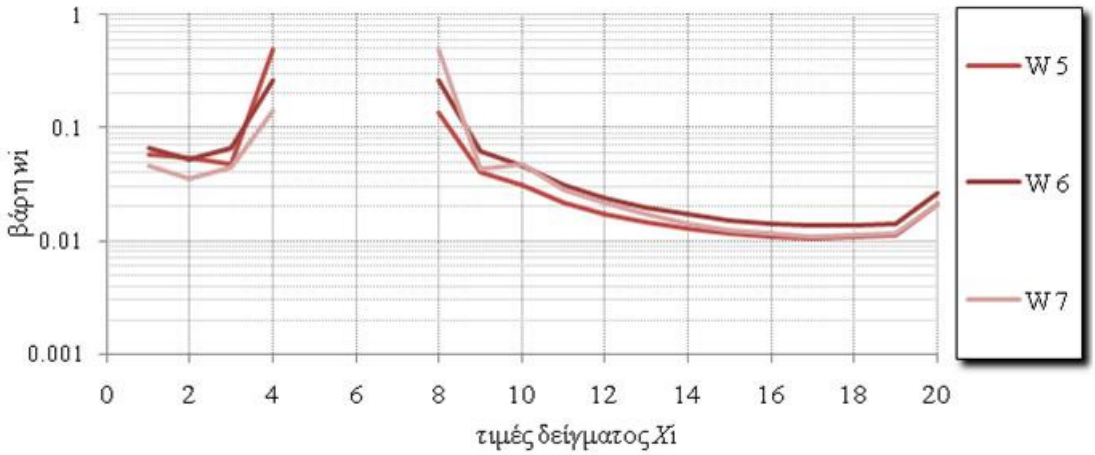


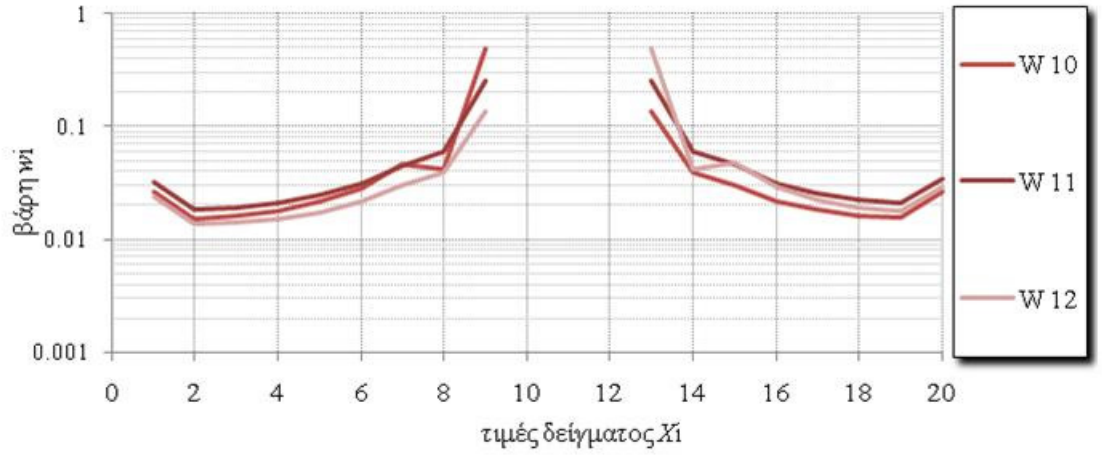





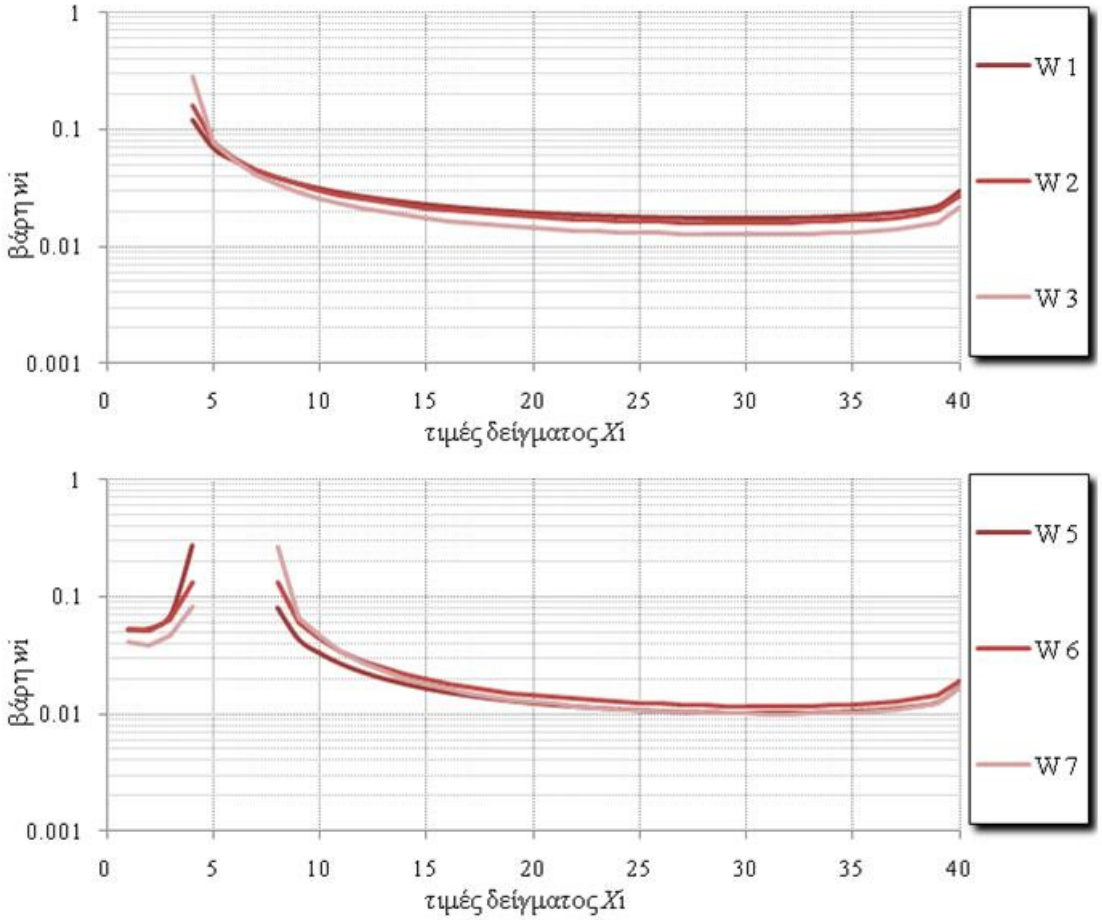

- Για δείγμα μεγέθους 20 τιμών και *H*=0.9 έχουμε:

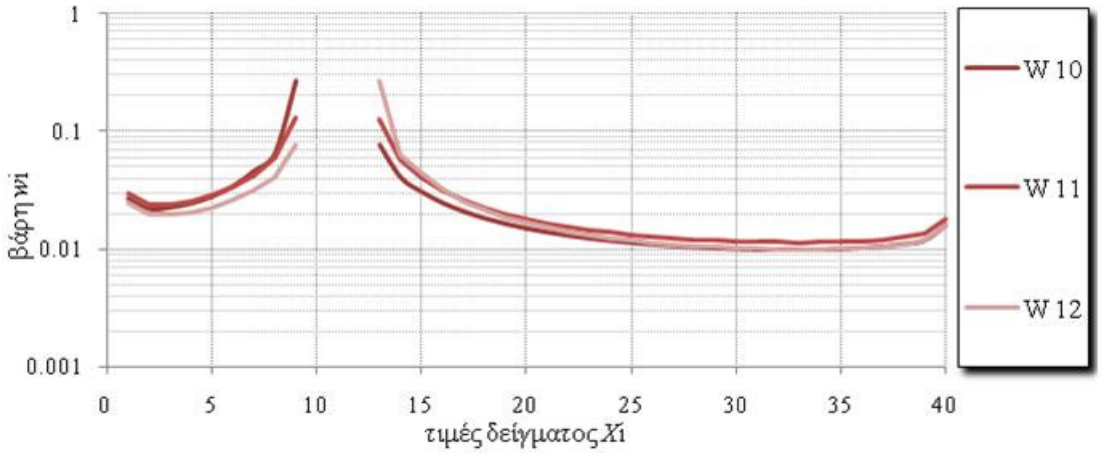


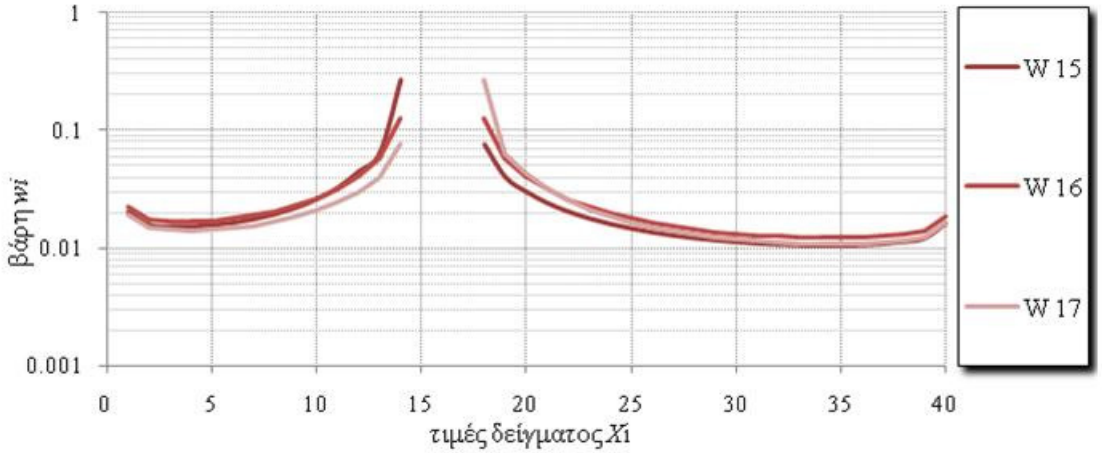



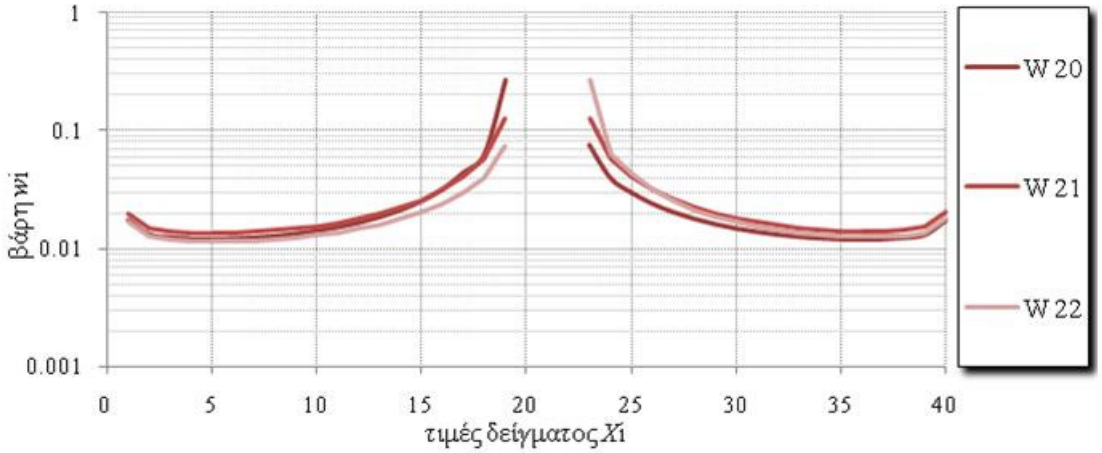



- Για δείγμα μεγέθους 40 τιμών και $H$=0.7 έχουμε:

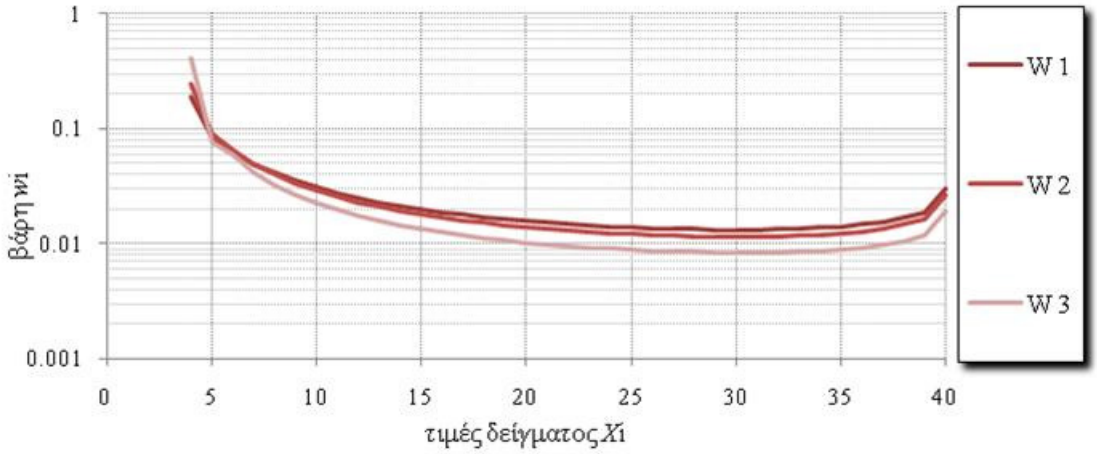


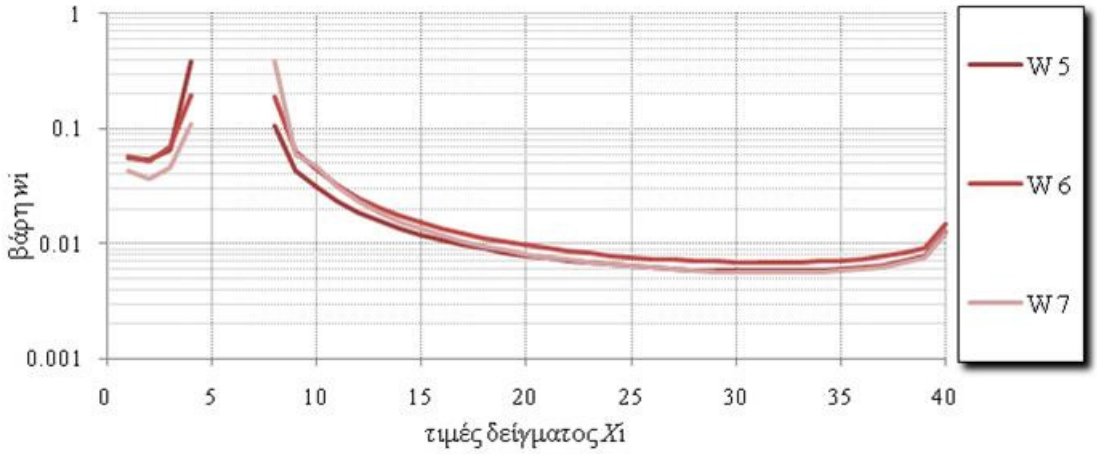

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 10
W 11
W 12

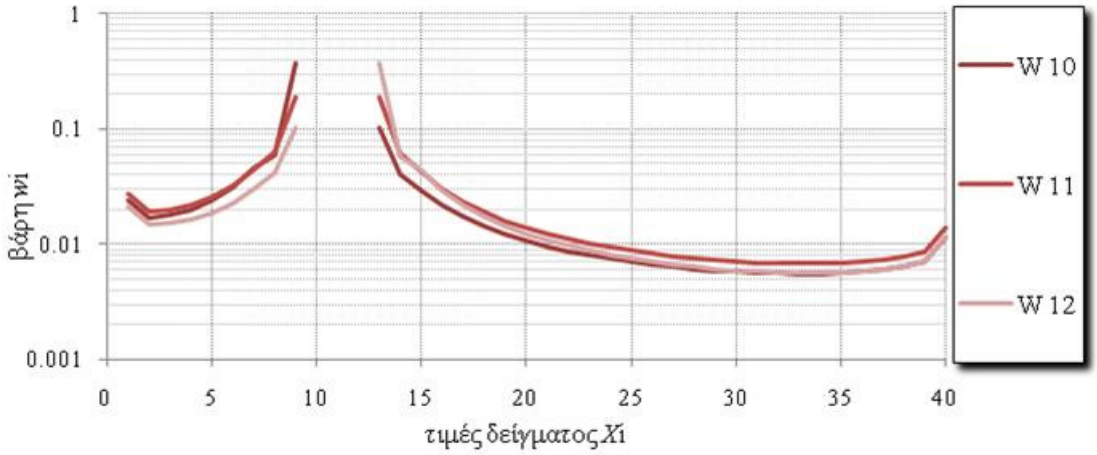

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 15
W 16
W 17

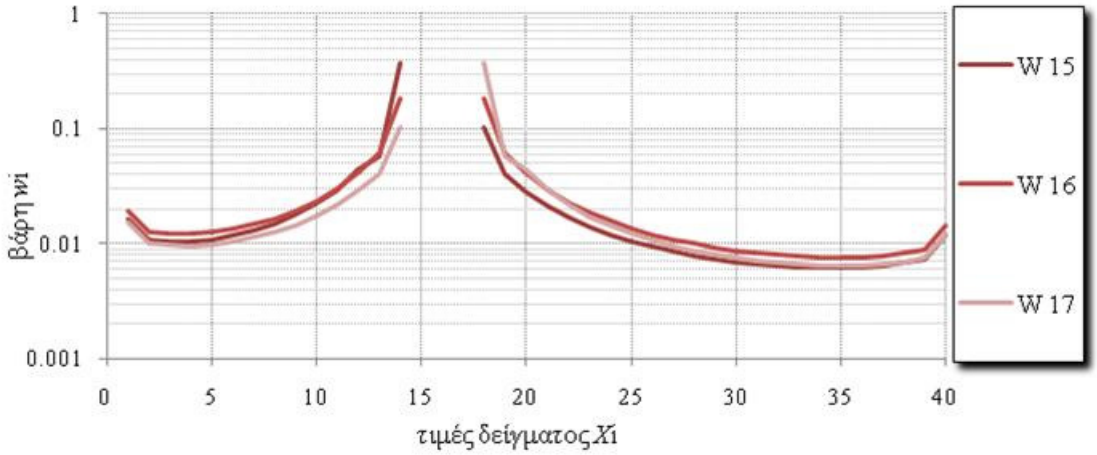

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 20
W 21
W 22

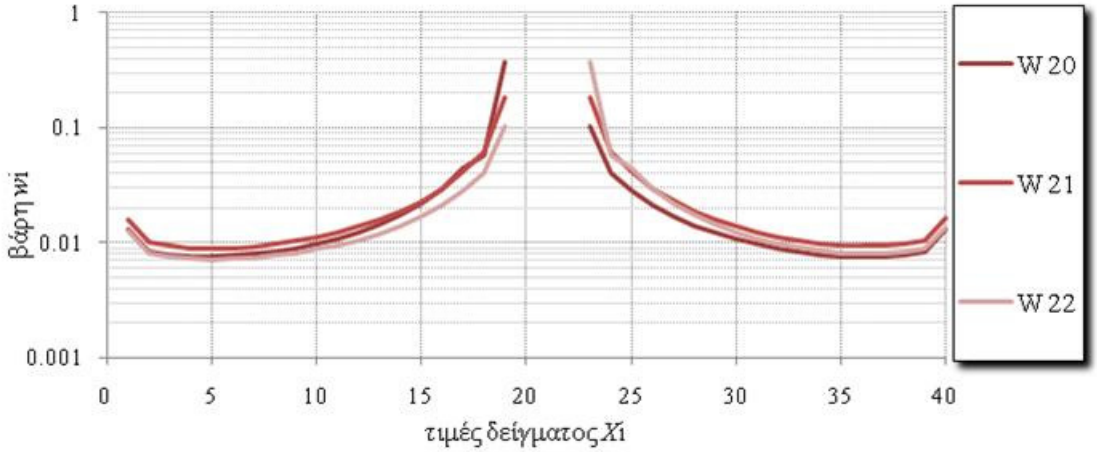
- Για δείγμα μεγέθους 40 τιμών και $H$=0.8 έχουμε:



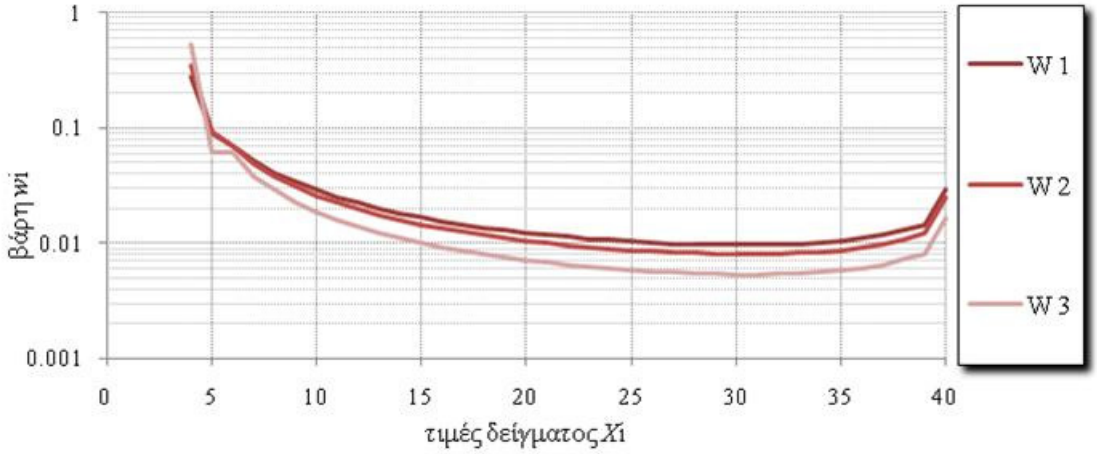



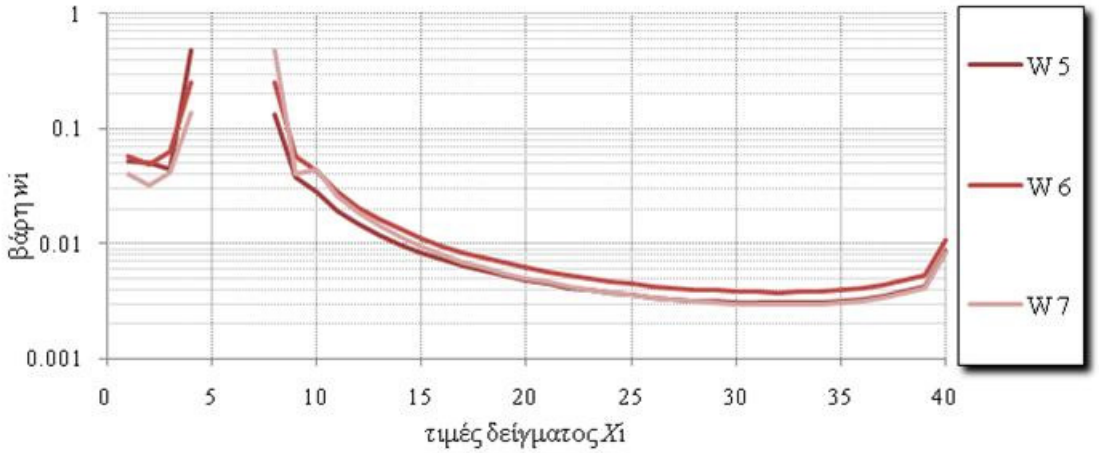



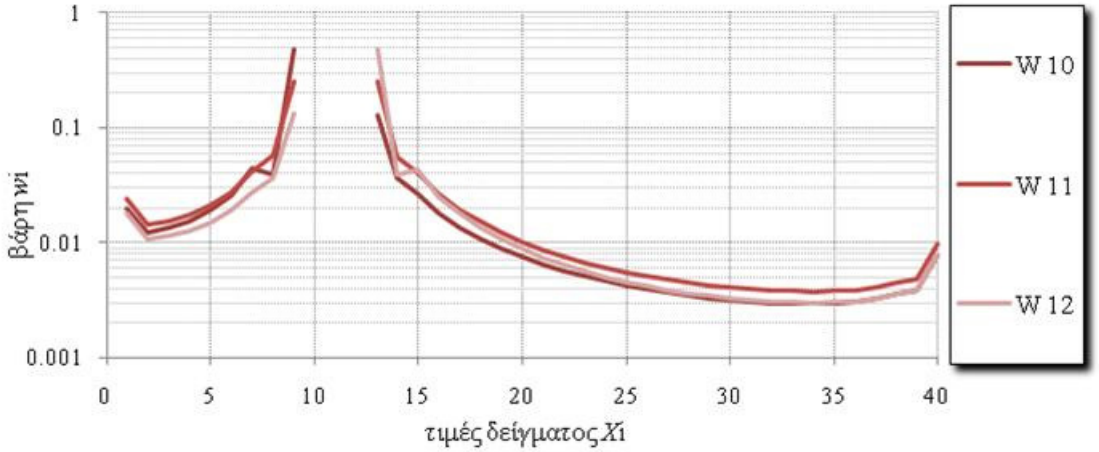


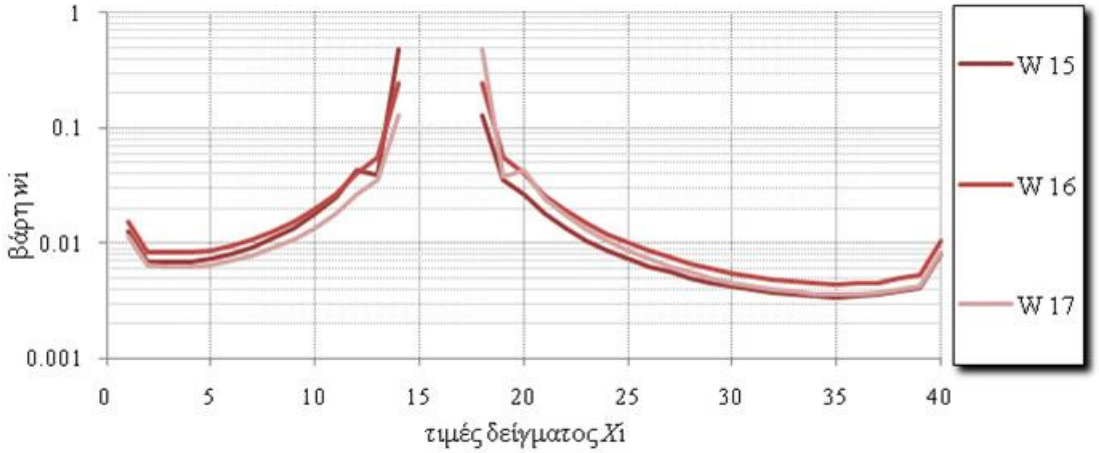



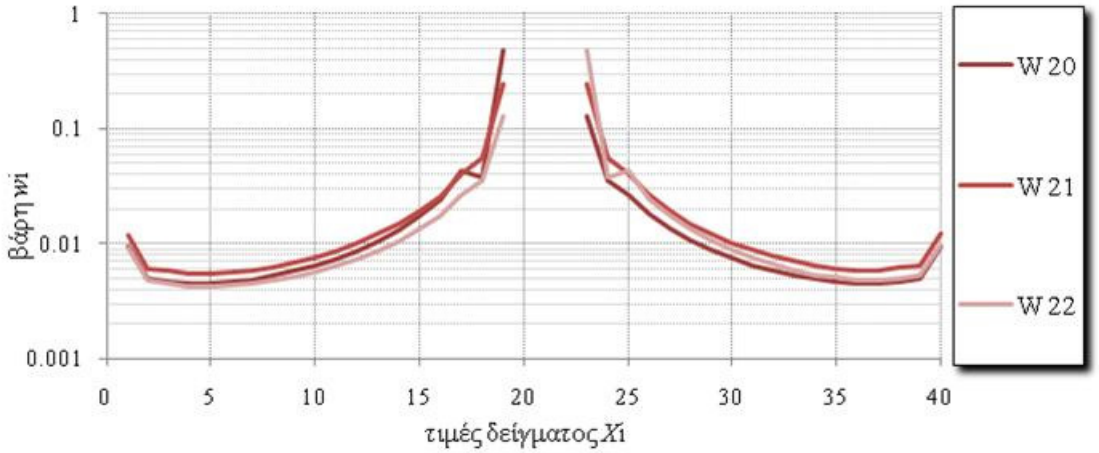

- Για δείγμα μεγέθους 40 τιμών και $H$=0.9 έχουμε:



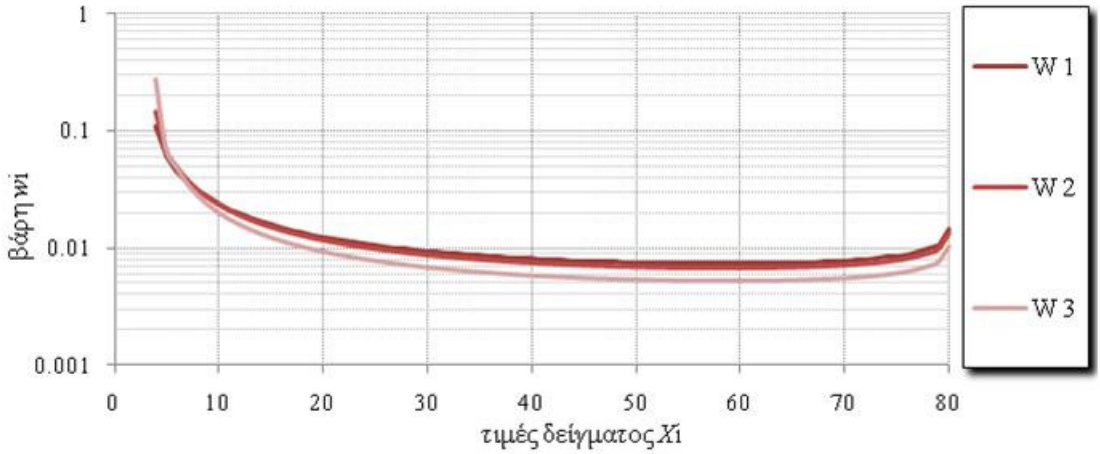

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 5
W 6
W 7

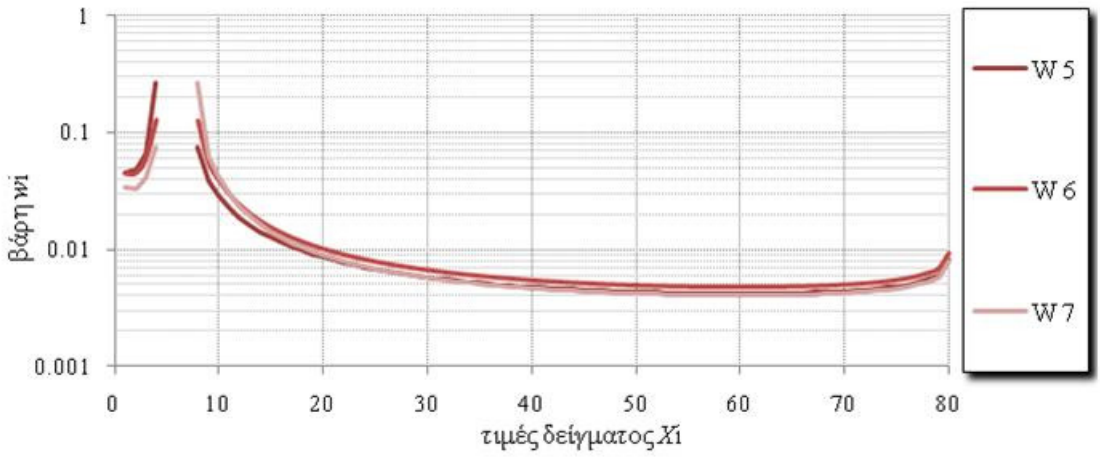

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 10
W 11
W 12

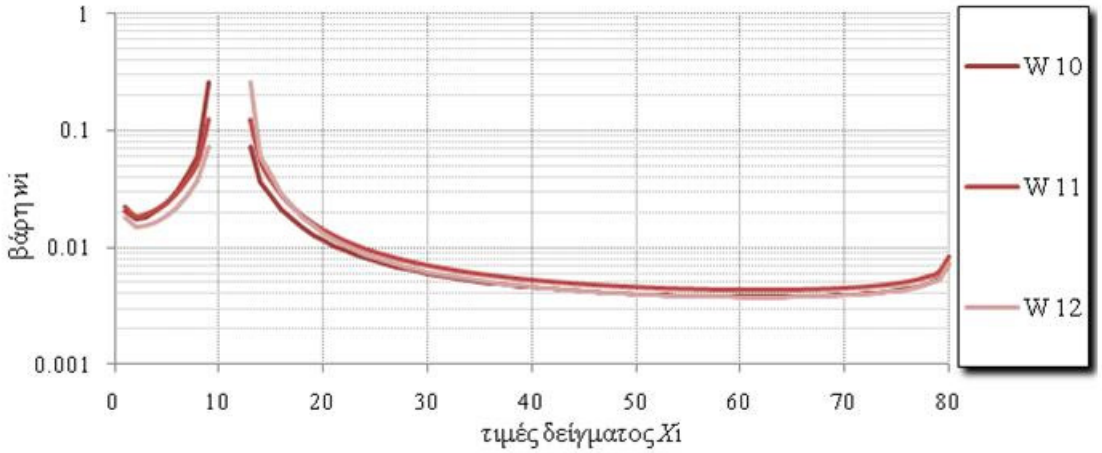

1
0.1
0.01
0.001
βάρη wi
0
5
10
15
20
25
30
35
40
τιμές δείγματος Xi
W 15
W 16
W 17

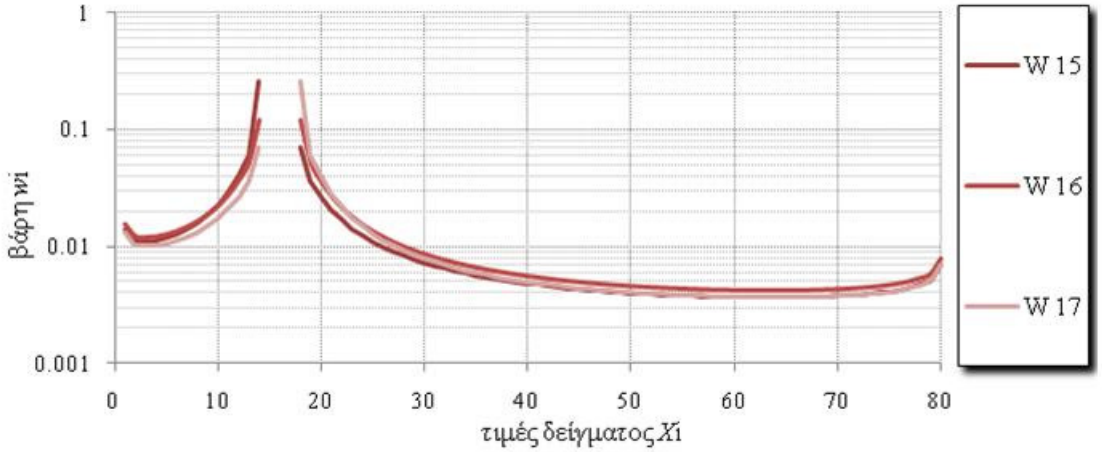


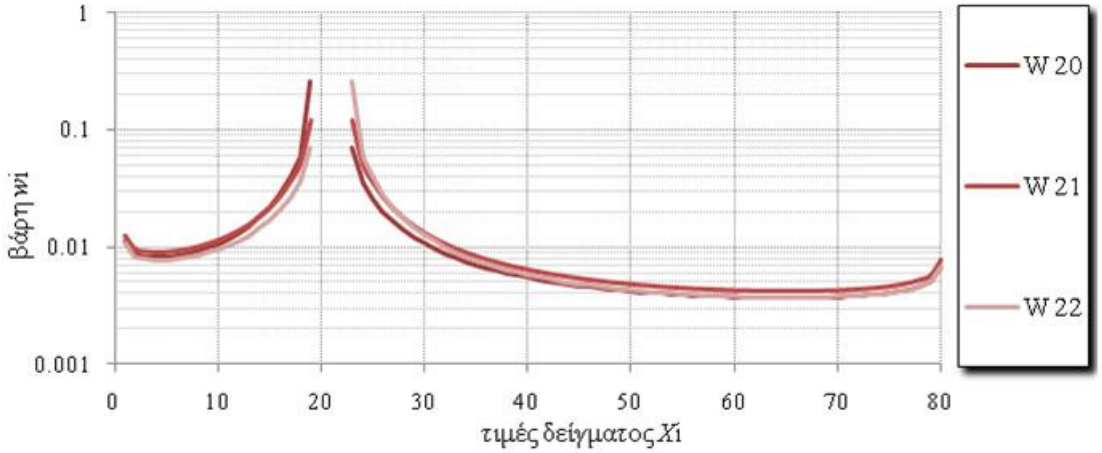

- Για δείγμα μεγέθους 80 τιμών και $H$=0.7 έχουμε:

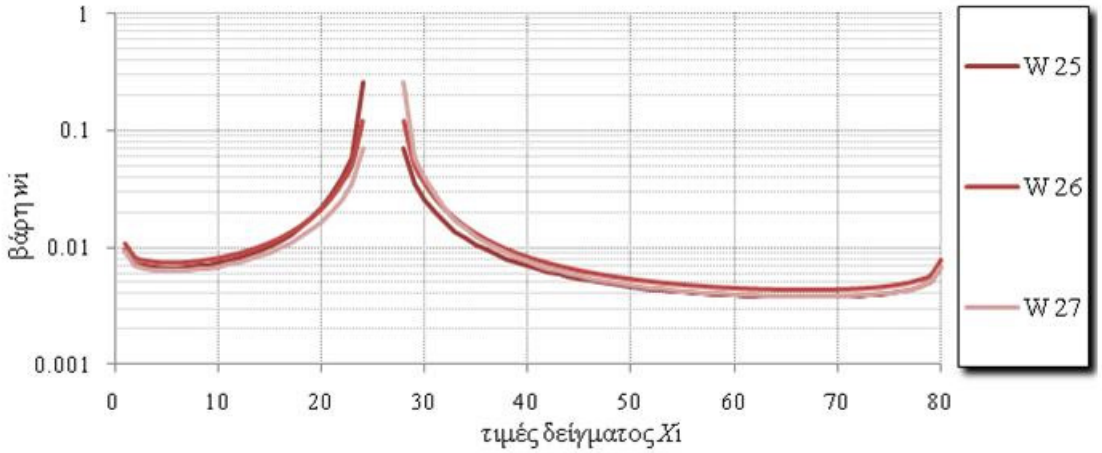


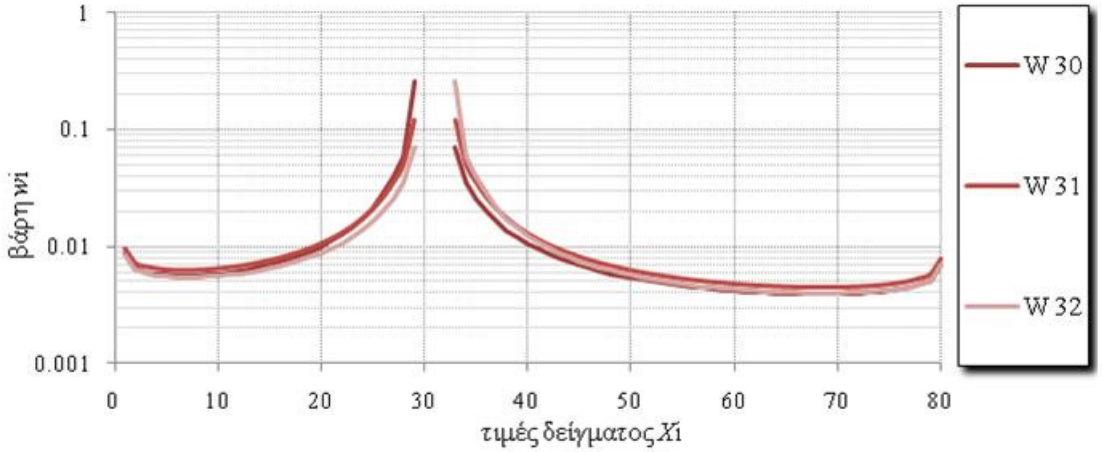



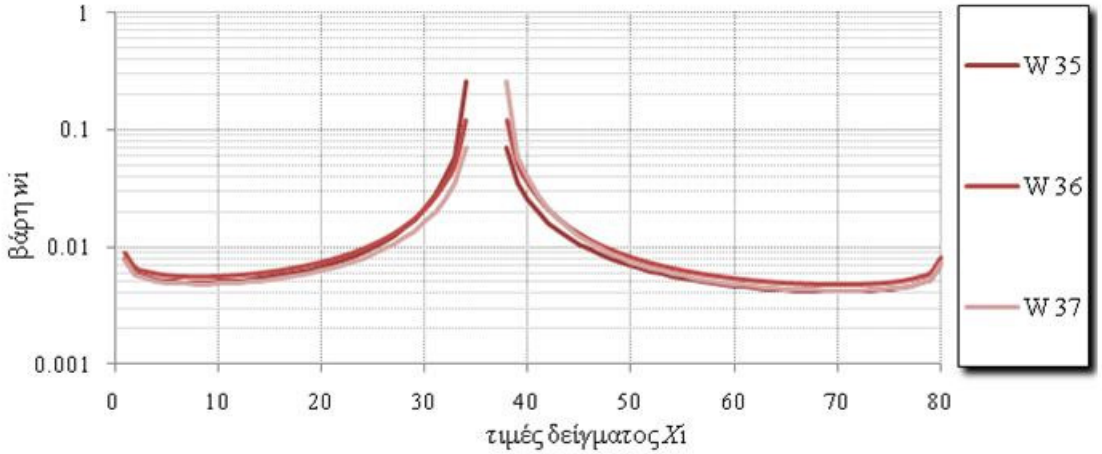

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12

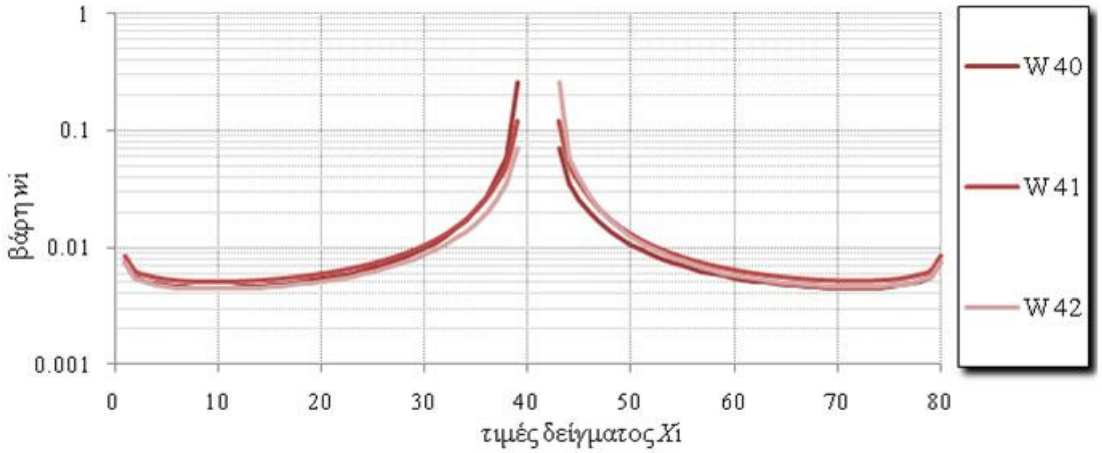

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17


1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

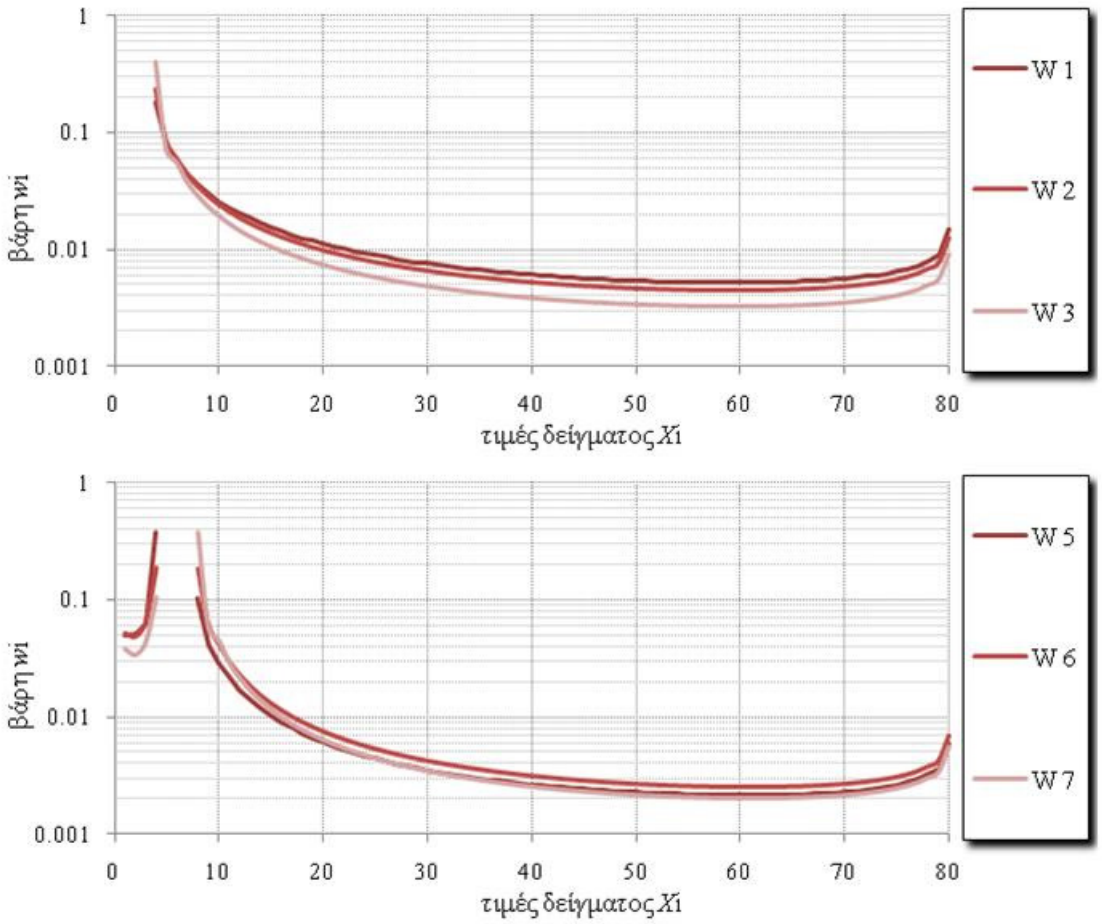

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

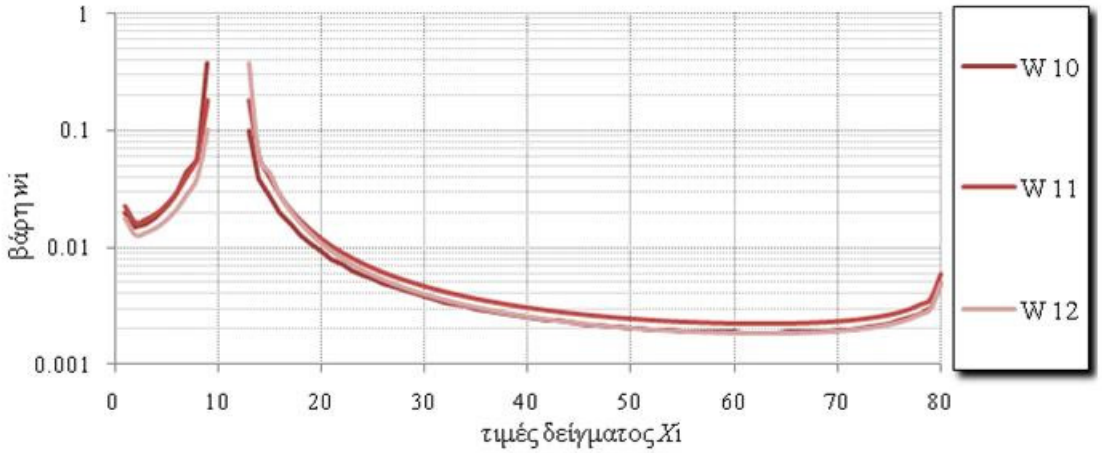

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

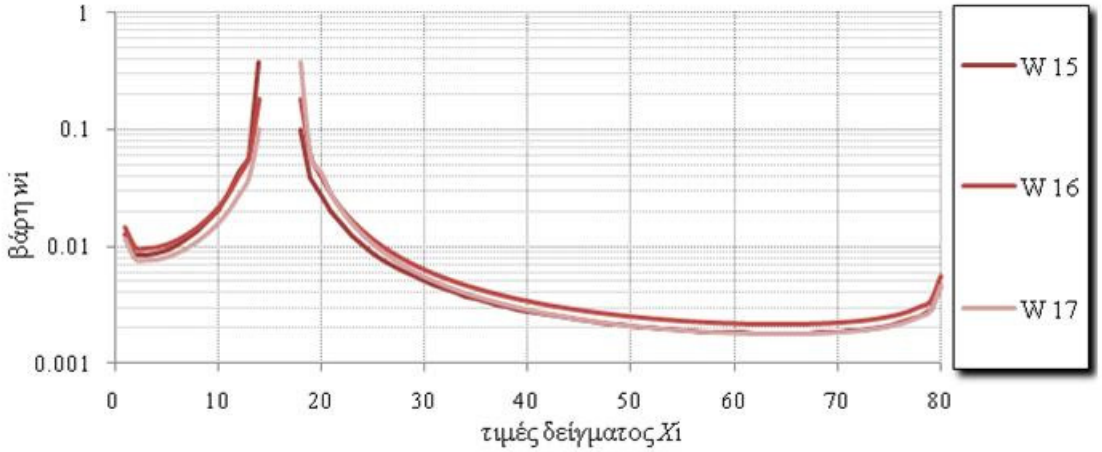

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 35
W 36
W 37

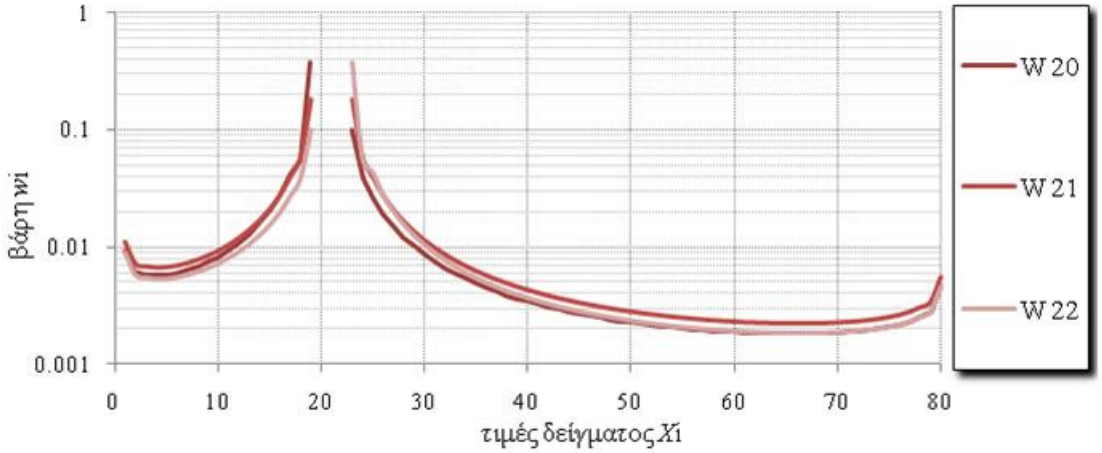


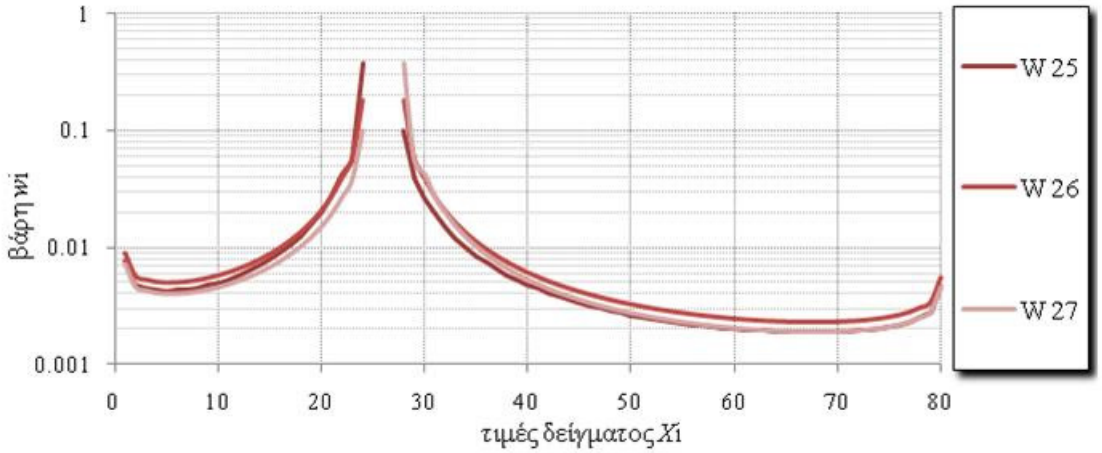

- Για δείγμα μεγέθους 80 τιμών και $H$=0.8 έχουμε:

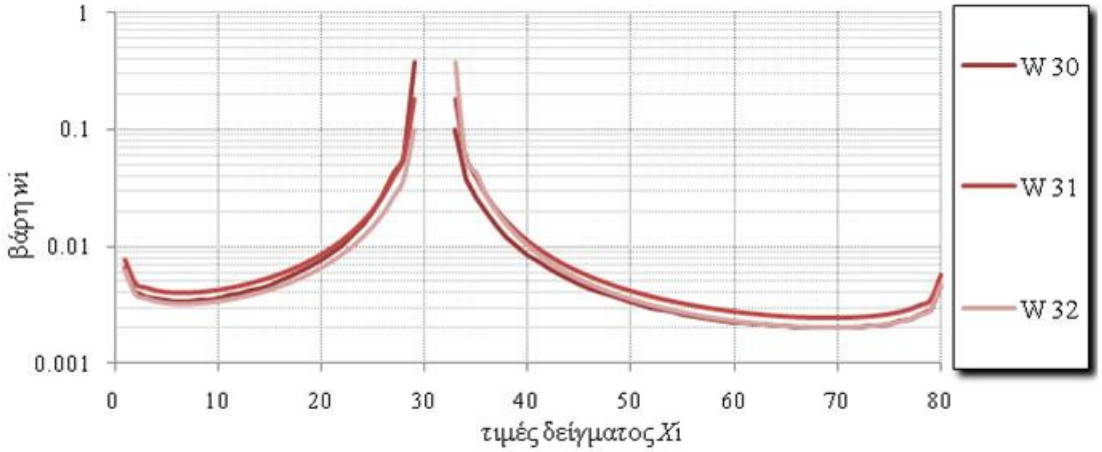


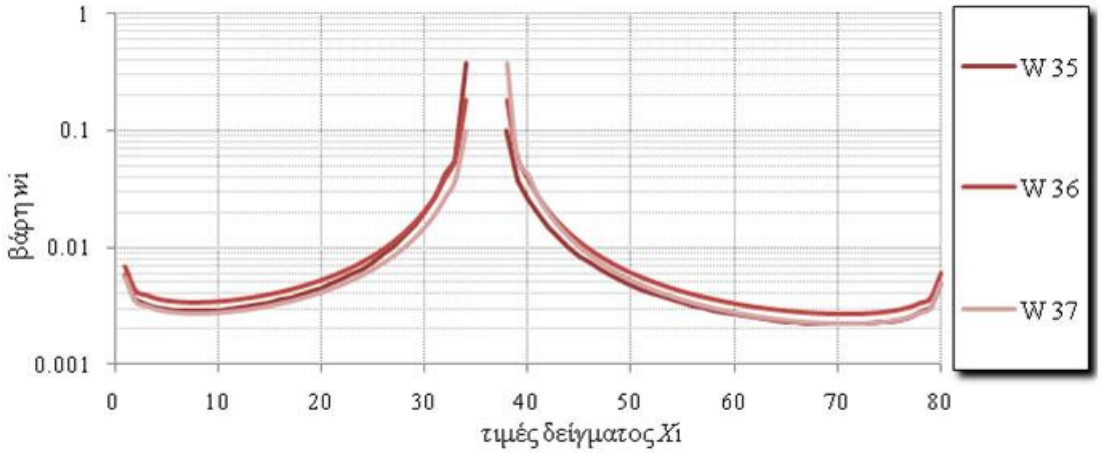

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12

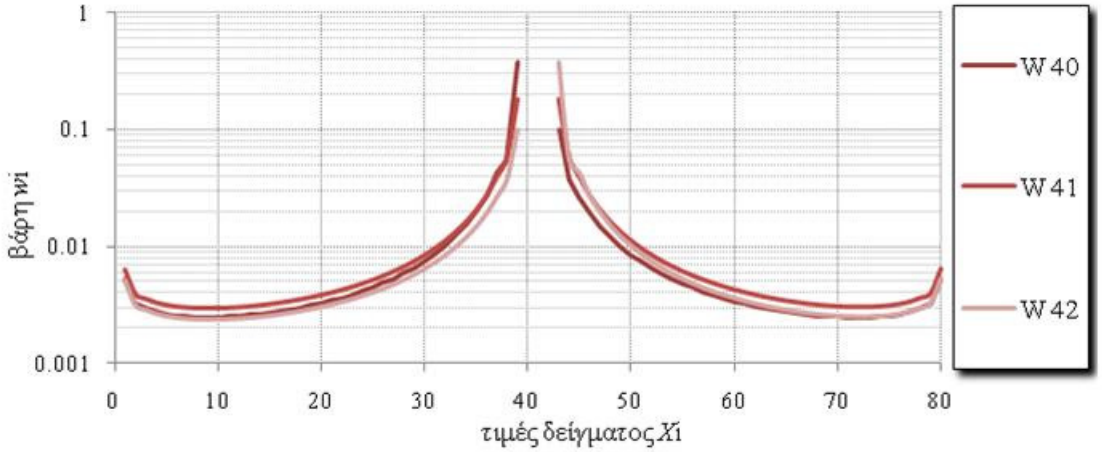

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17


1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

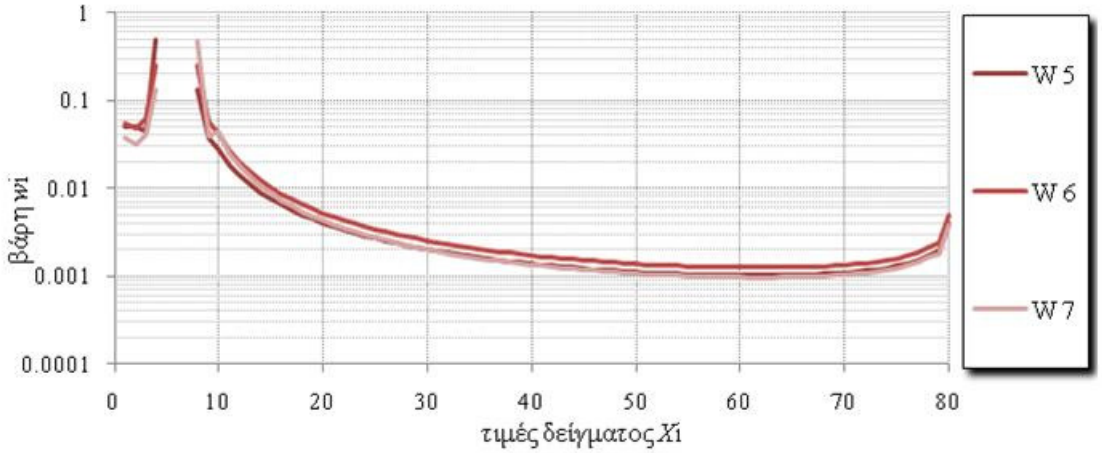

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

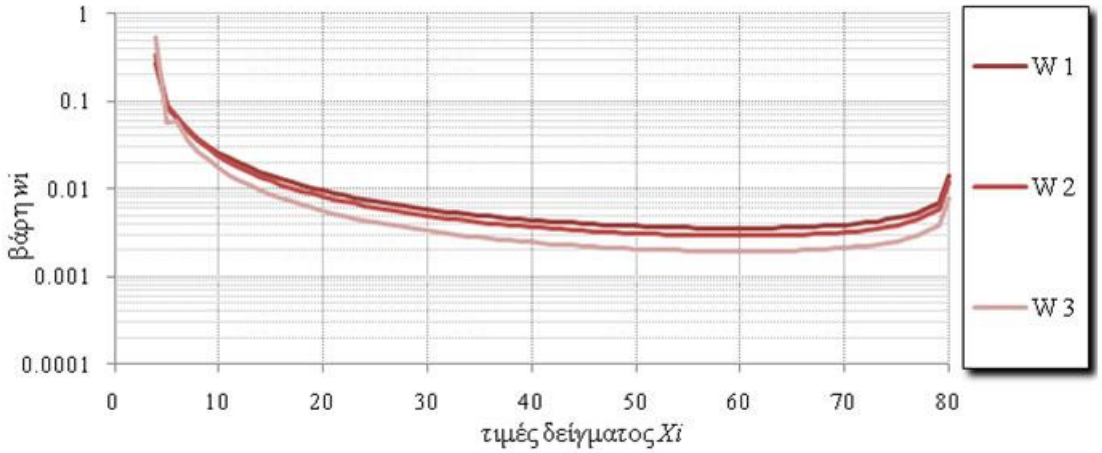

1
0.1
0.01
0.001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

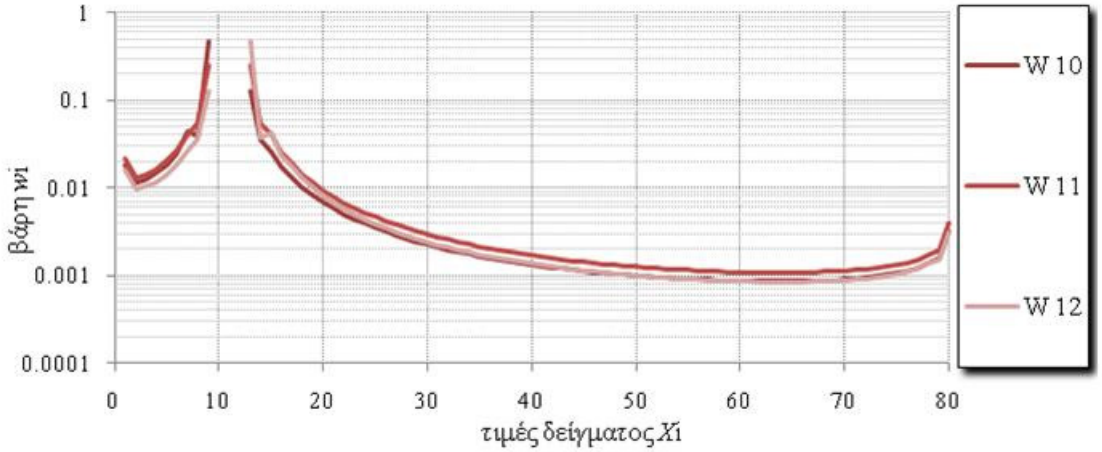


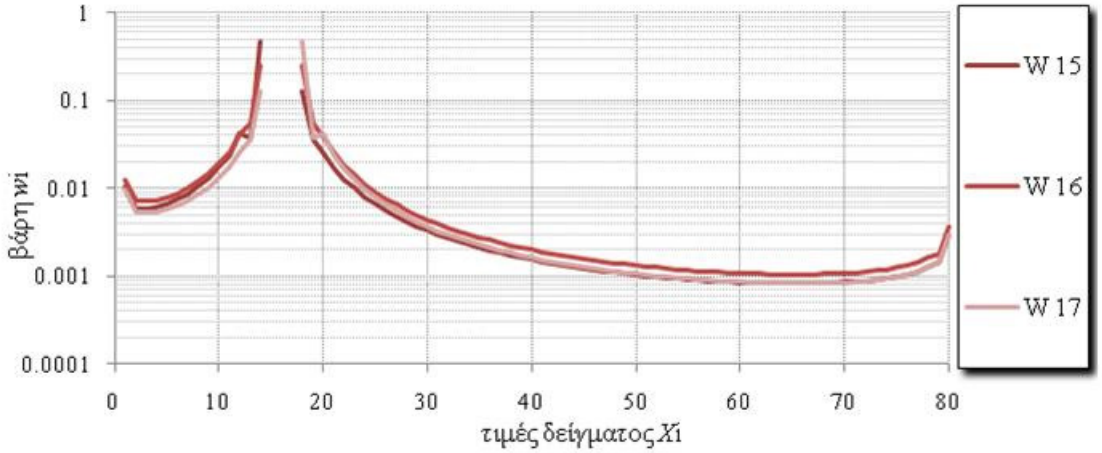



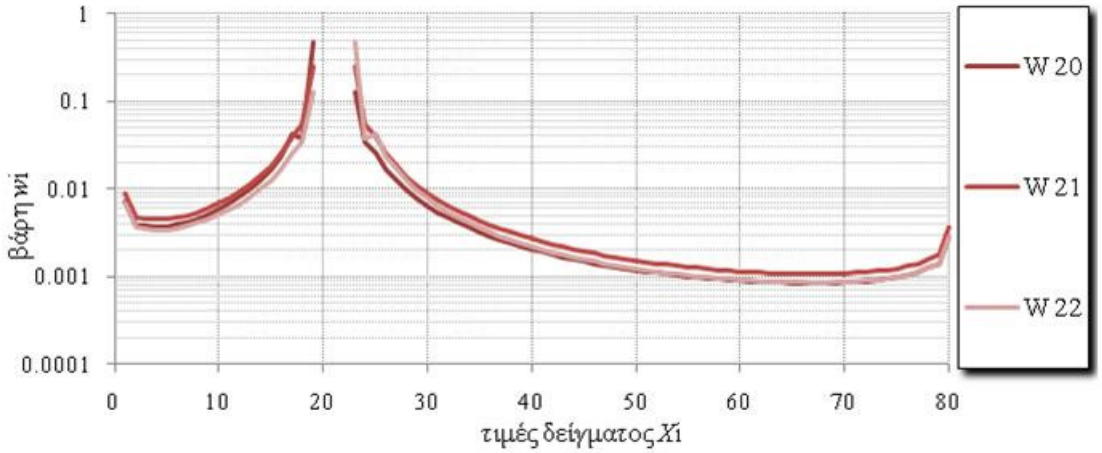

- Για δείγμα μεγέθους 80 τιμών και *H*=0.9 έχουμε:

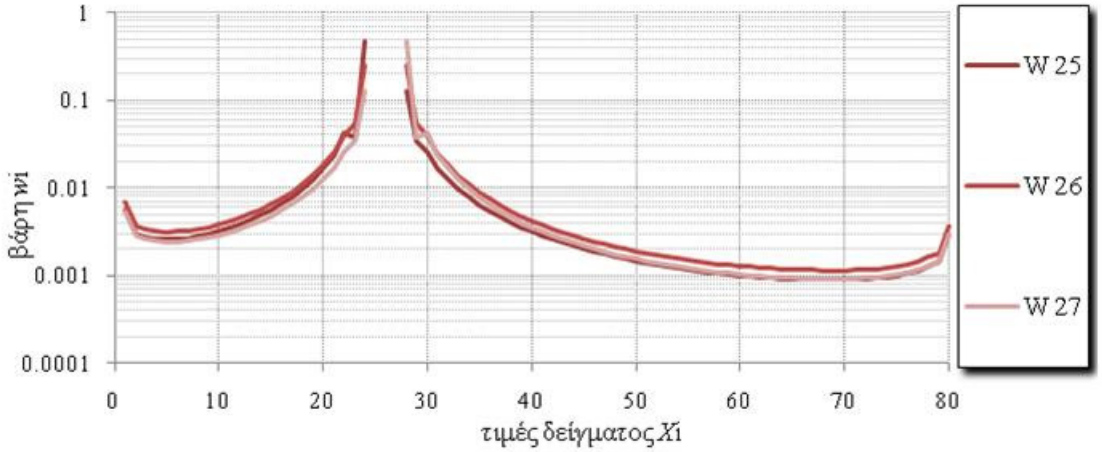


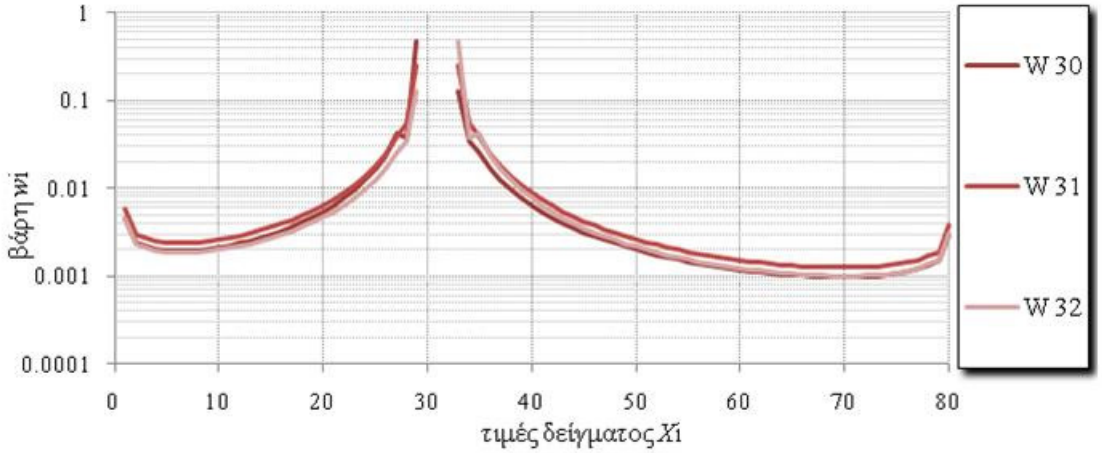

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 1
W 2
W 3

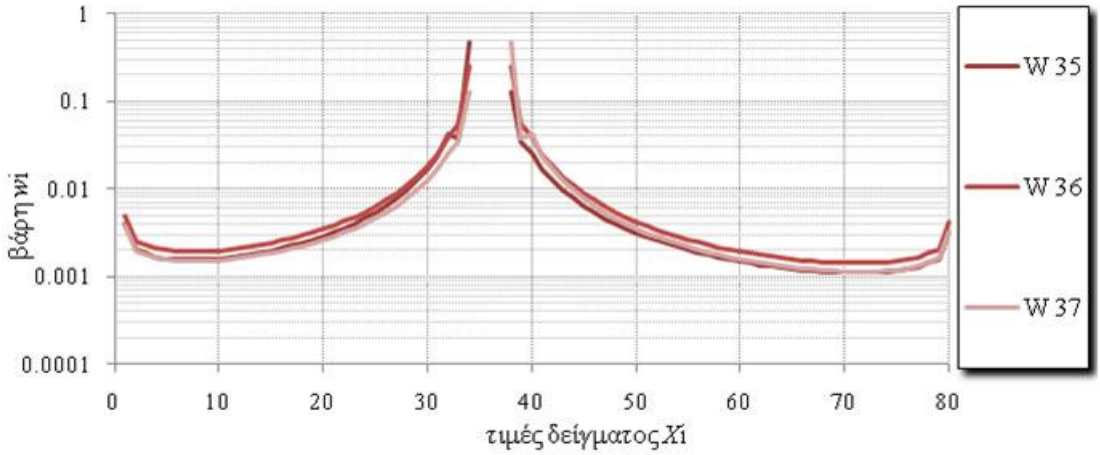

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 10
W 11
W 12

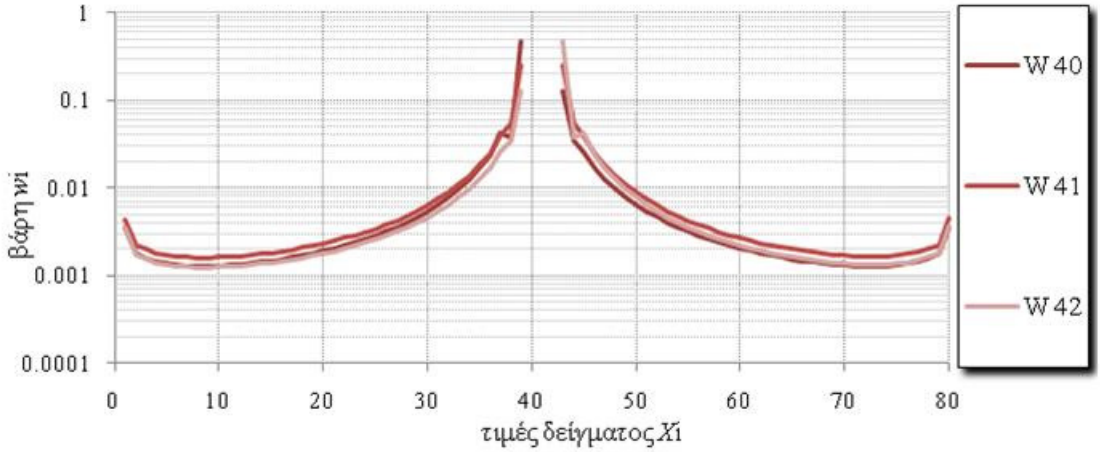

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 15
W 16
W 17


βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 20
W 21
W 22

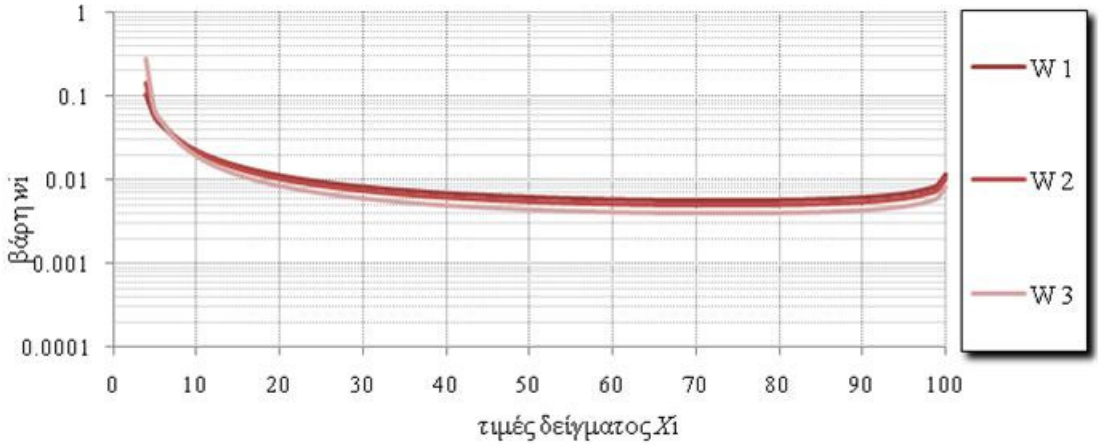

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 25
W 26
W 27

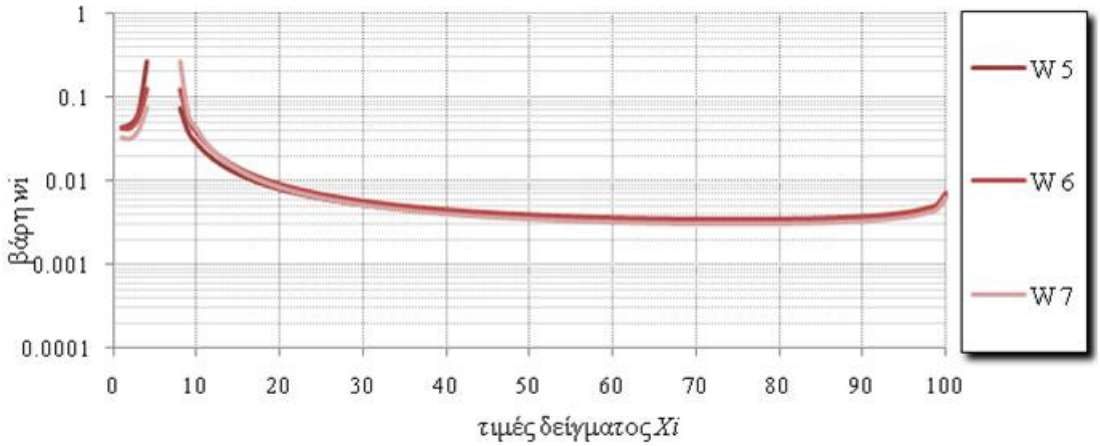

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
τιμές δείγματος Xi
W 30
W 31
W 32

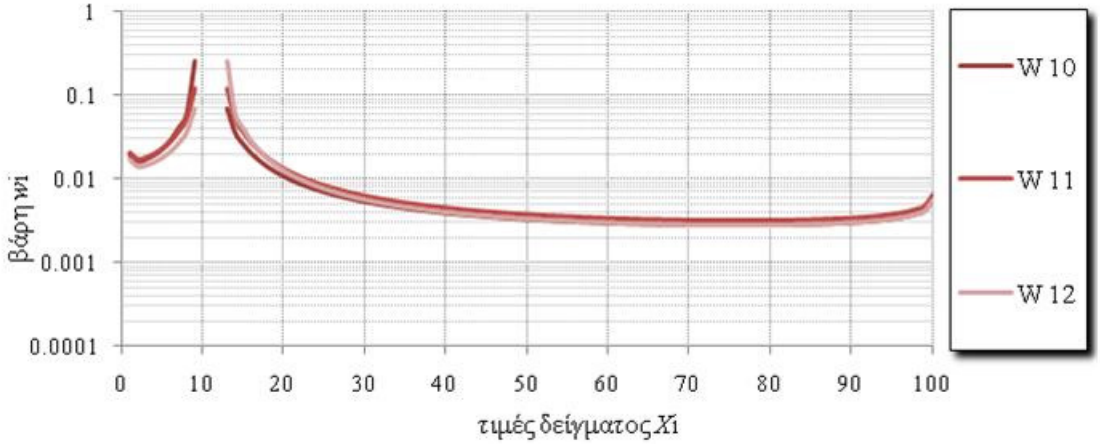


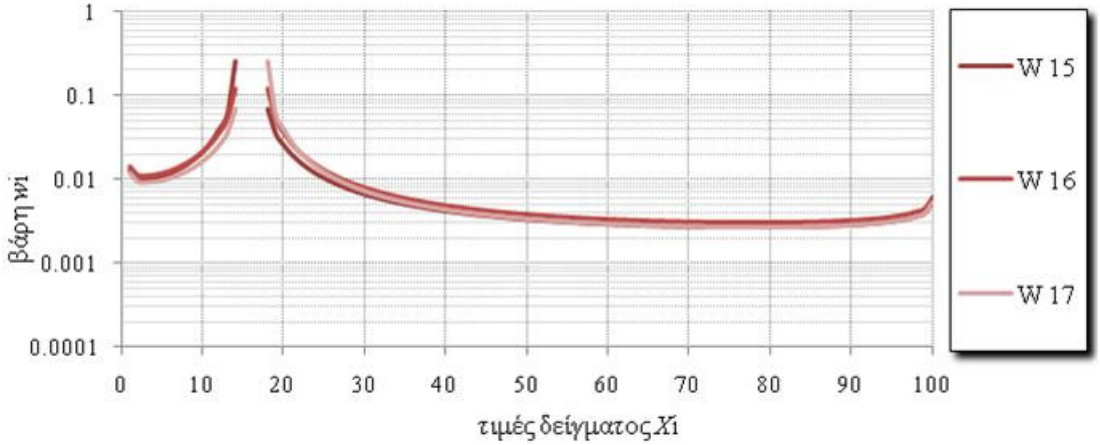



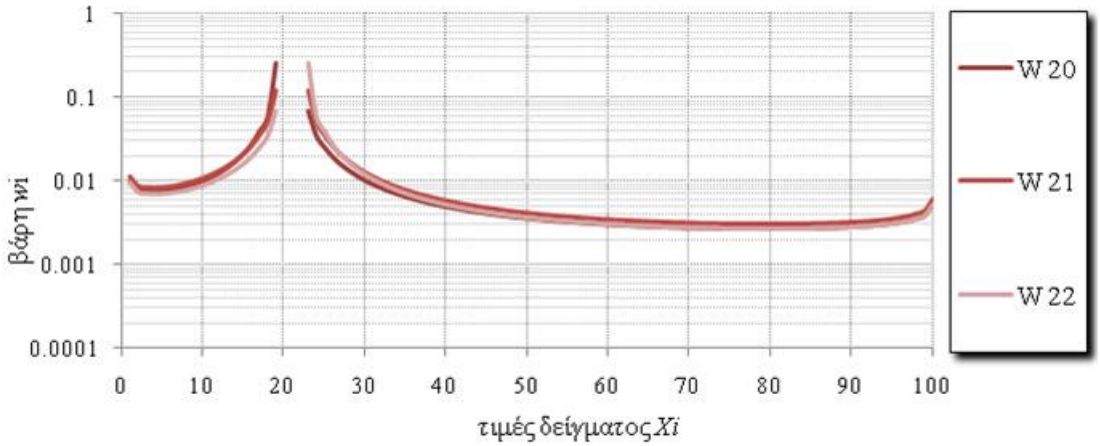

- Για δείγμα μεγέθους 100 τιμών και $H$=0.7 έχουμε:

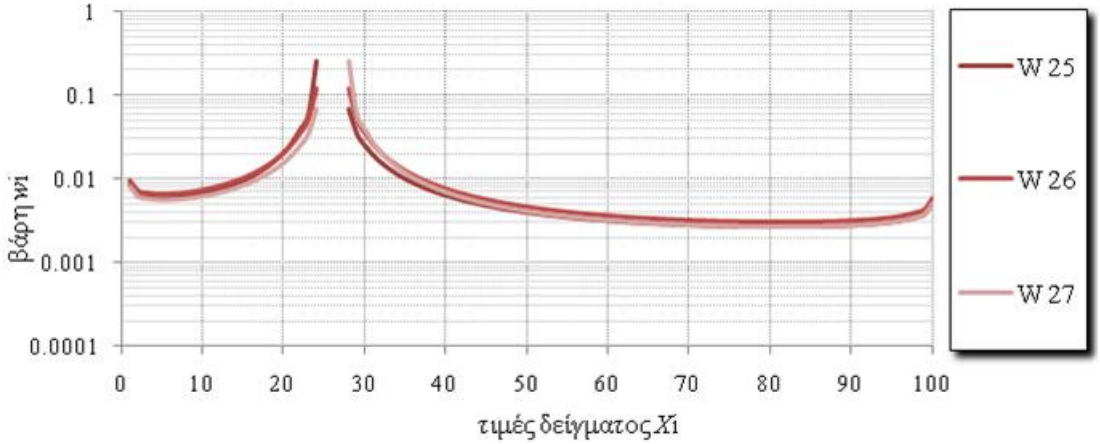


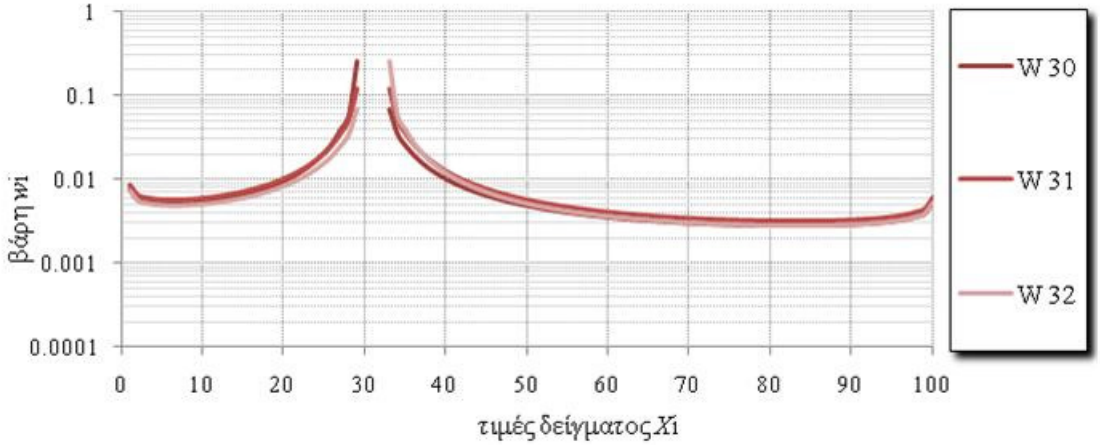

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 5
W 6
W 7

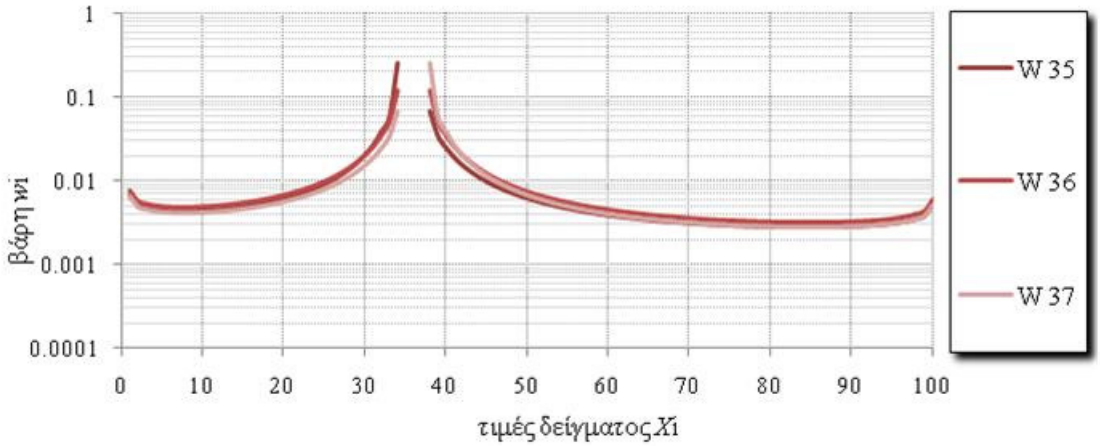

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 10
W 11
W 12

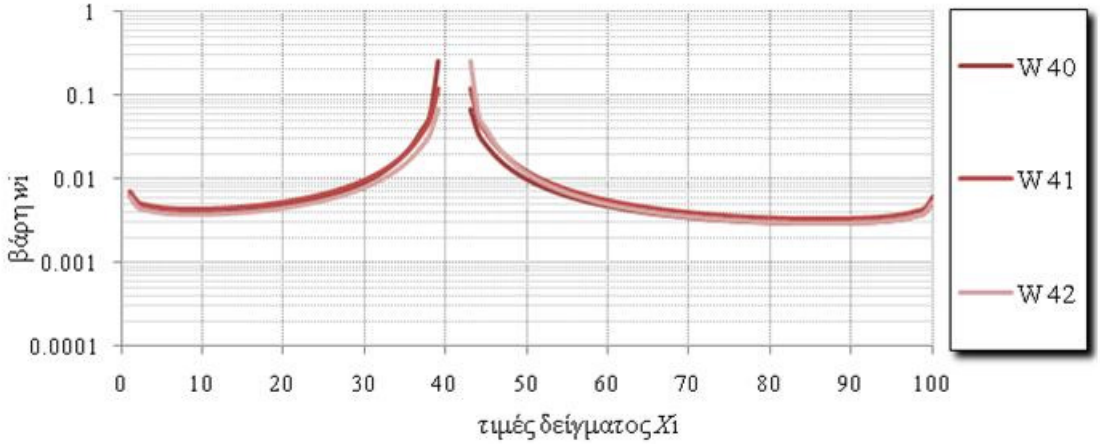

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 15
W 16
W 17

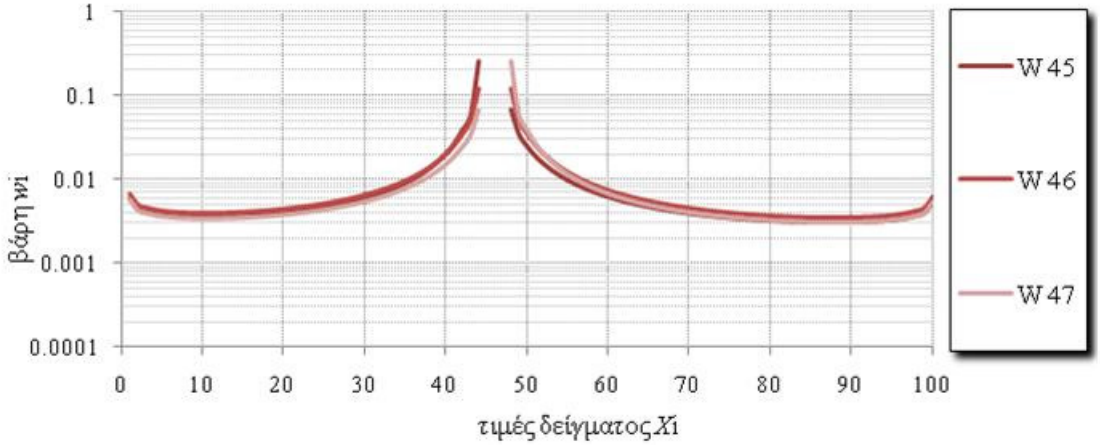

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 20
W 21
W 22

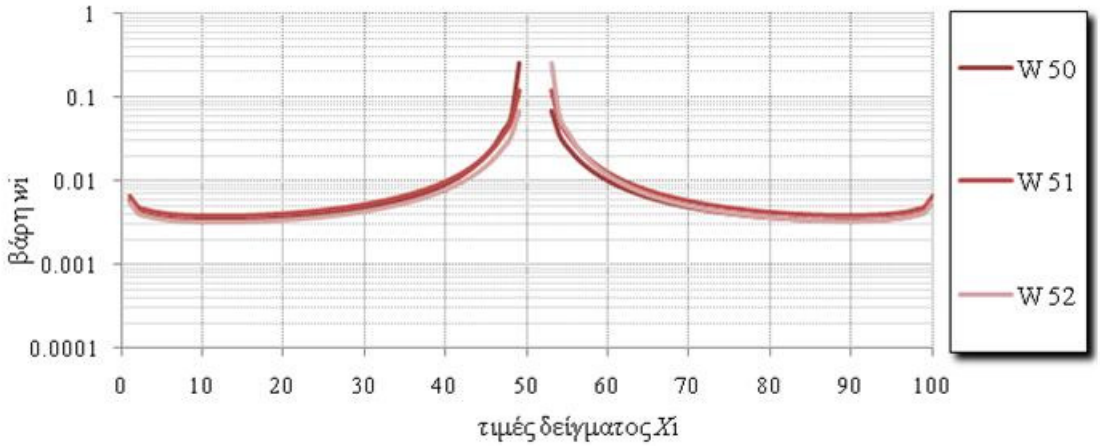

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 30
W 31
W 32

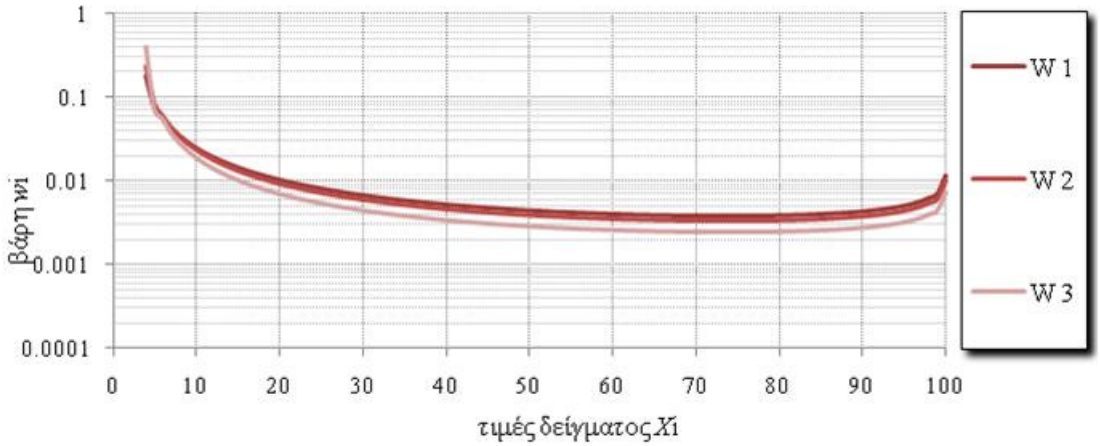

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 35
W 36
W 37

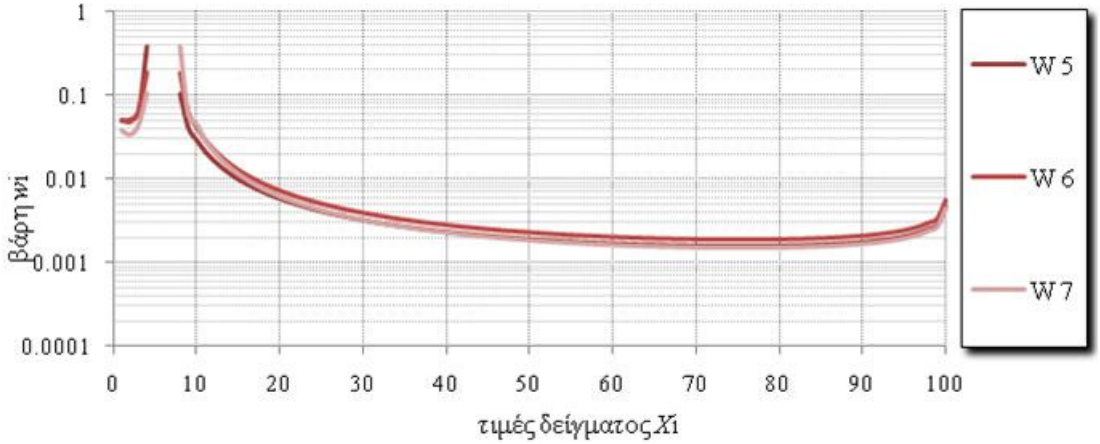

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 40
W 41
W 42

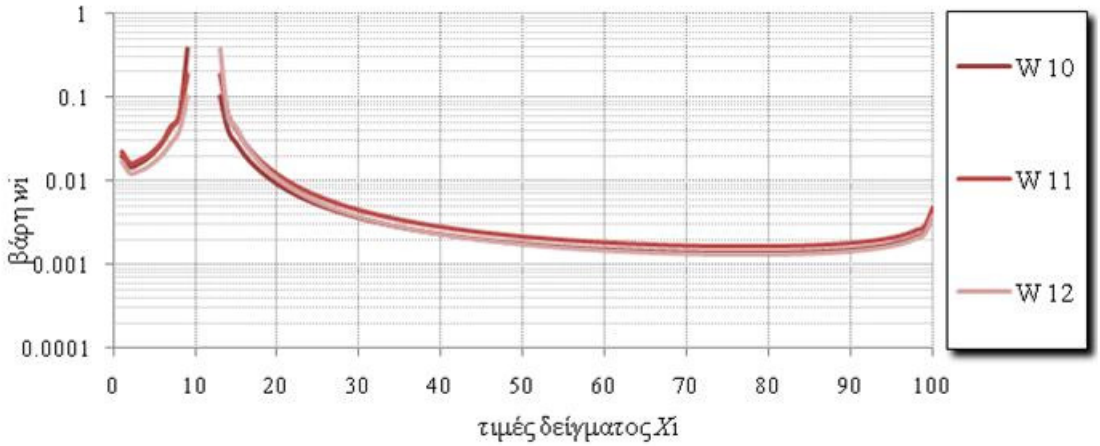

βάρη wi
1
0.1
0.01
0.001
0.0001
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 45
W 46
W 47

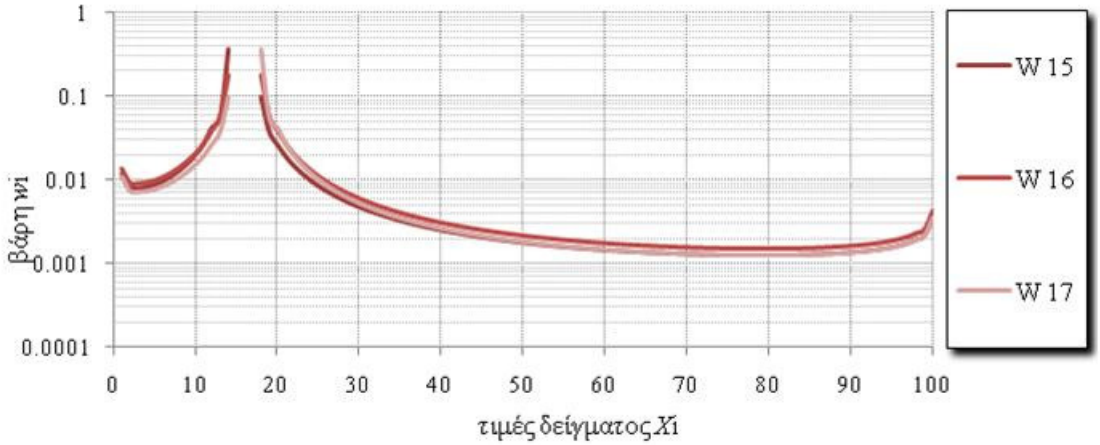


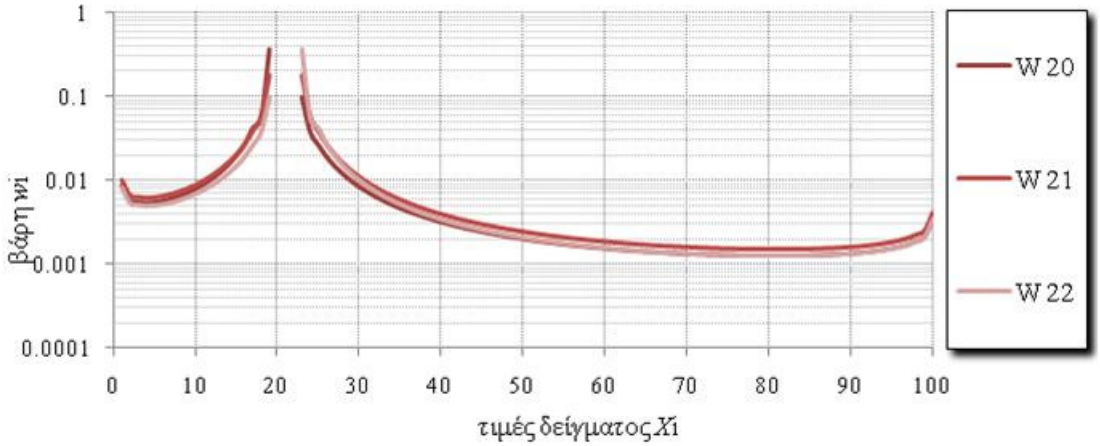

- Για δείγμα μεγέθους 100 τιμών και $H$=0.8 έχουμε:

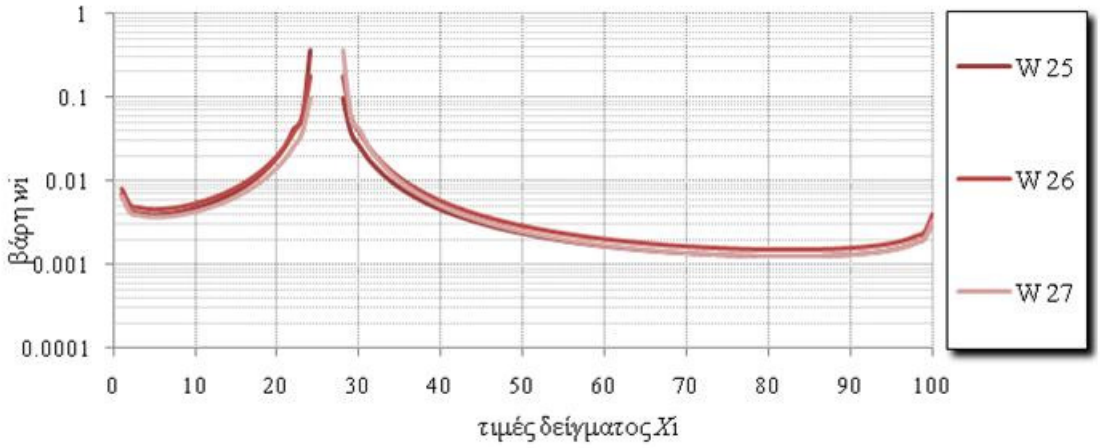


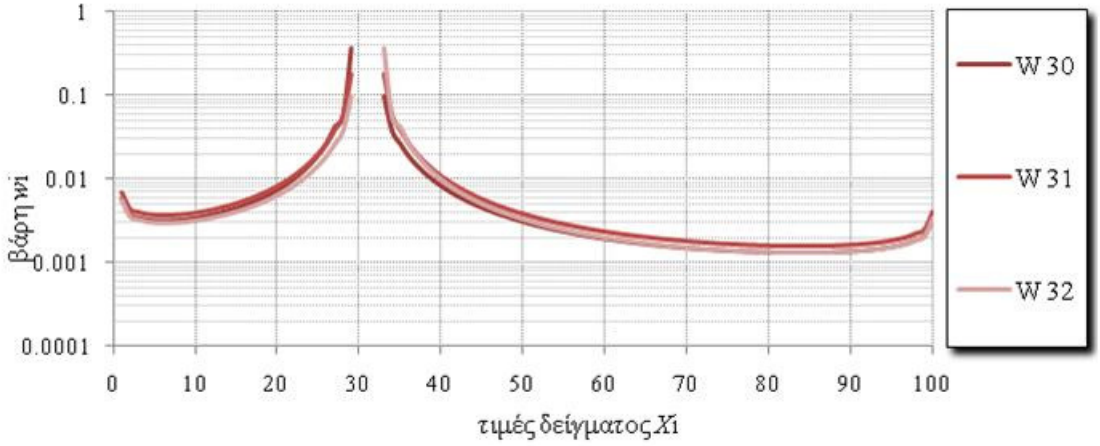



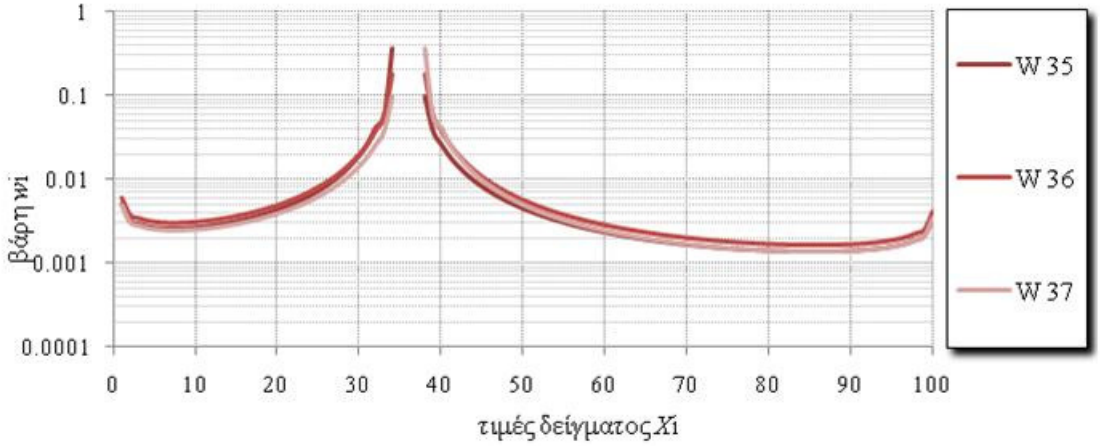

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 10
W 11
W 12

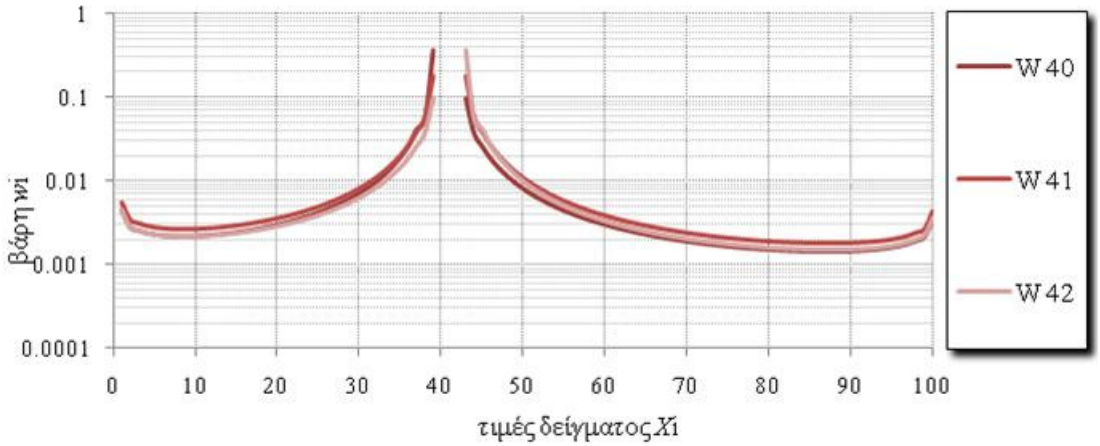

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 15
W 16
W 17

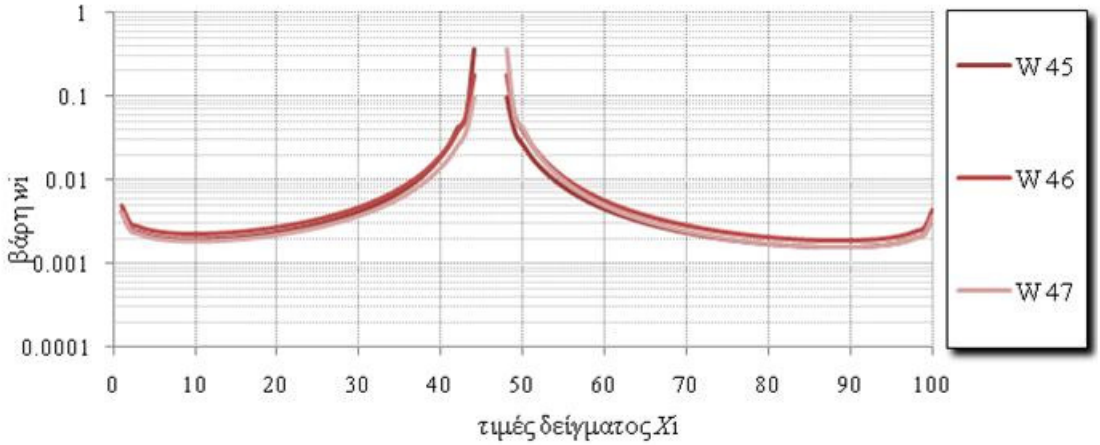

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 20
W 21
W 22

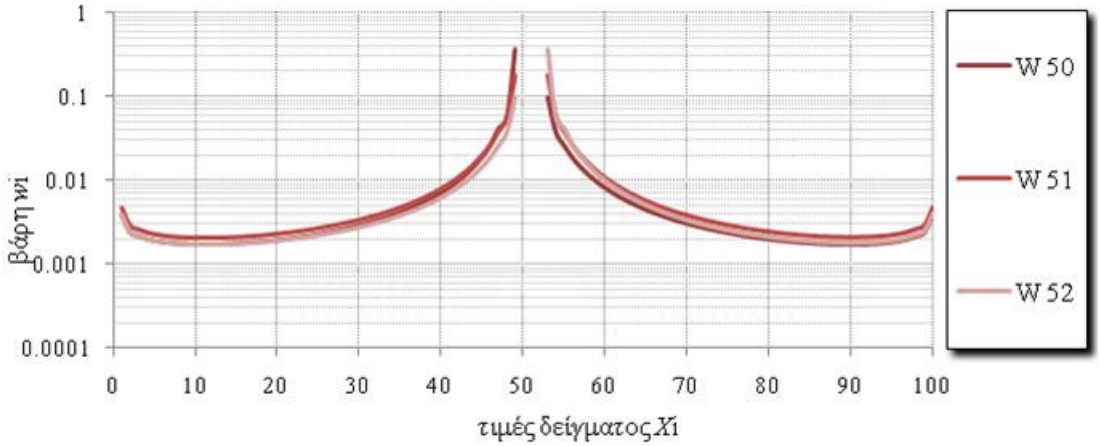

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 30
W 31
W 32

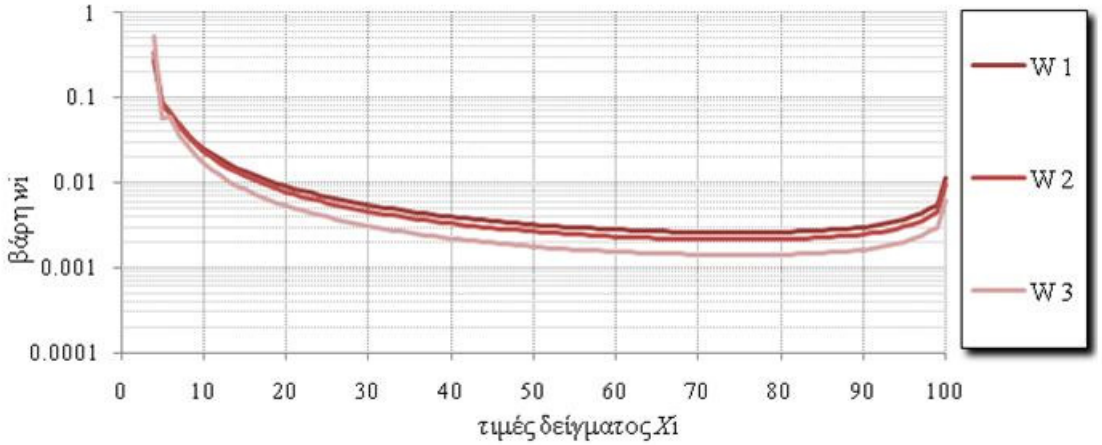

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 35
W 36
W 37

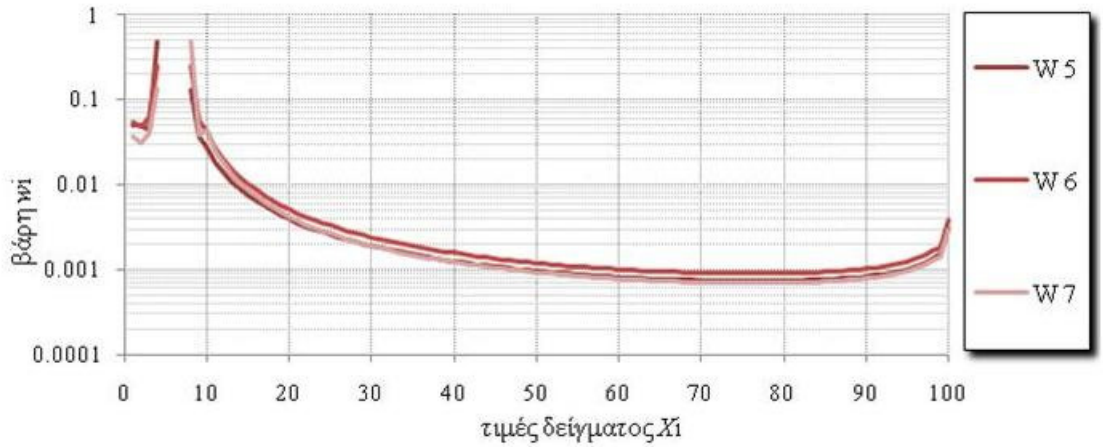

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 40
W 41
W 42

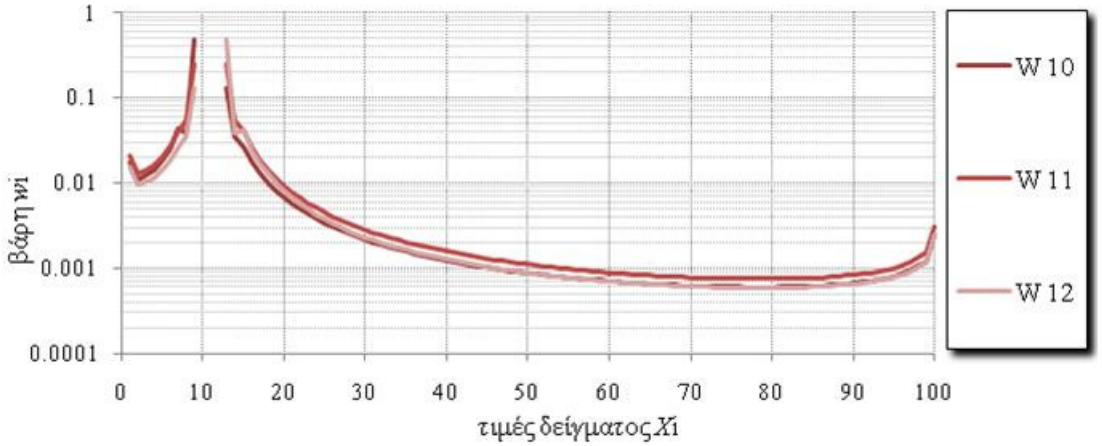

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 45
W 46
W 47

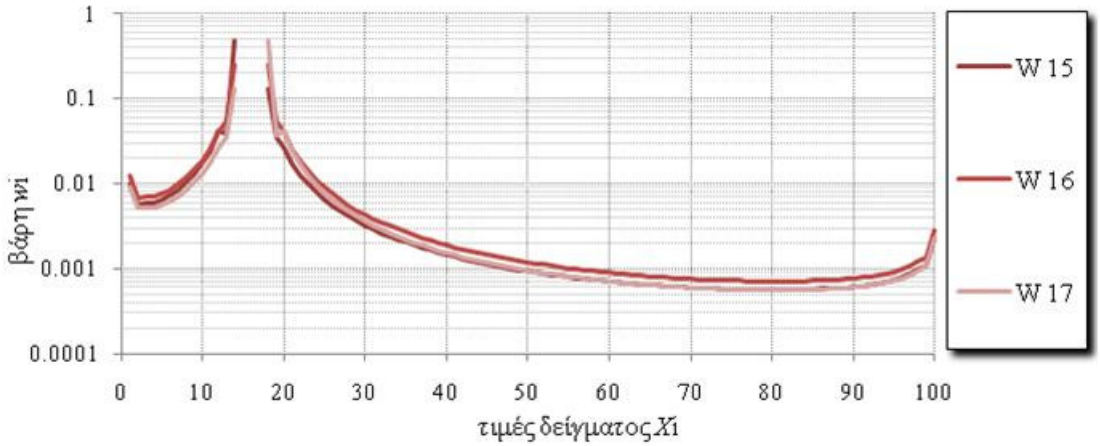

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 50
W 51
W 52

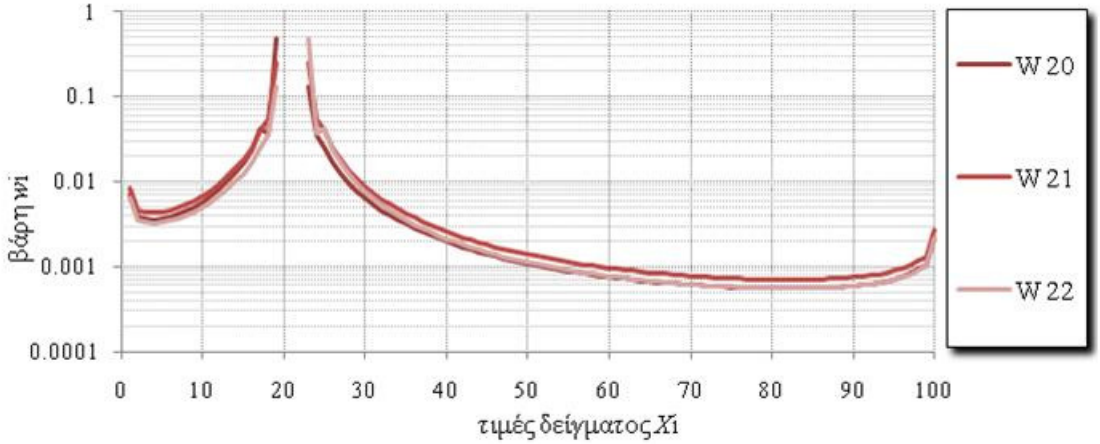
- Για δείγμα μεγέθους 100 τιμών και $H$=0.9 έχουμε:

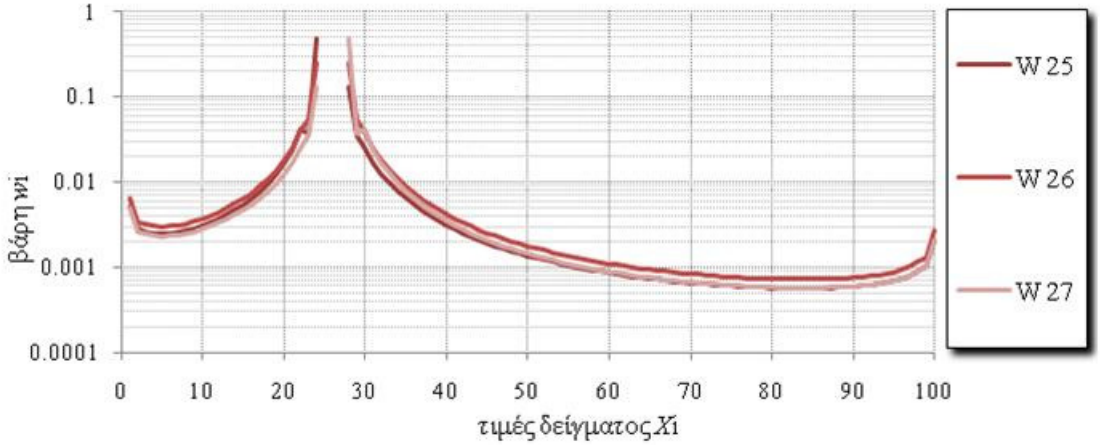


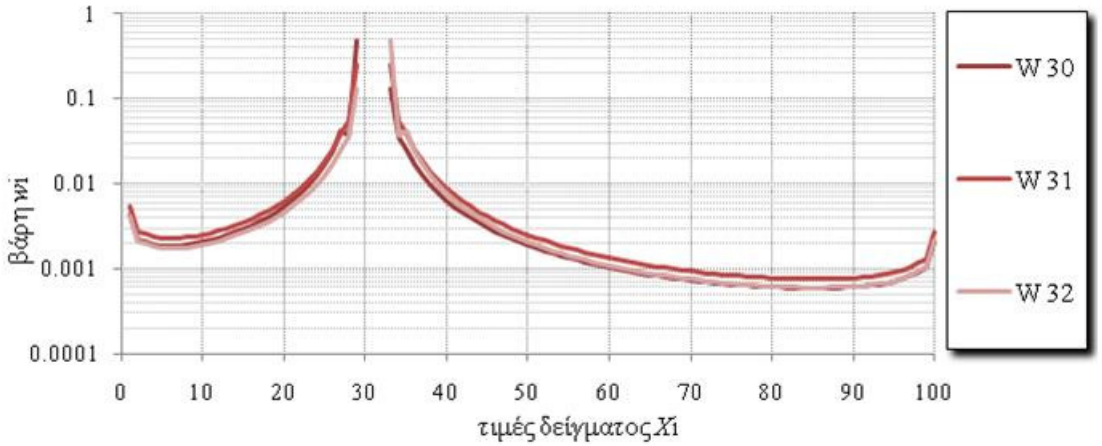



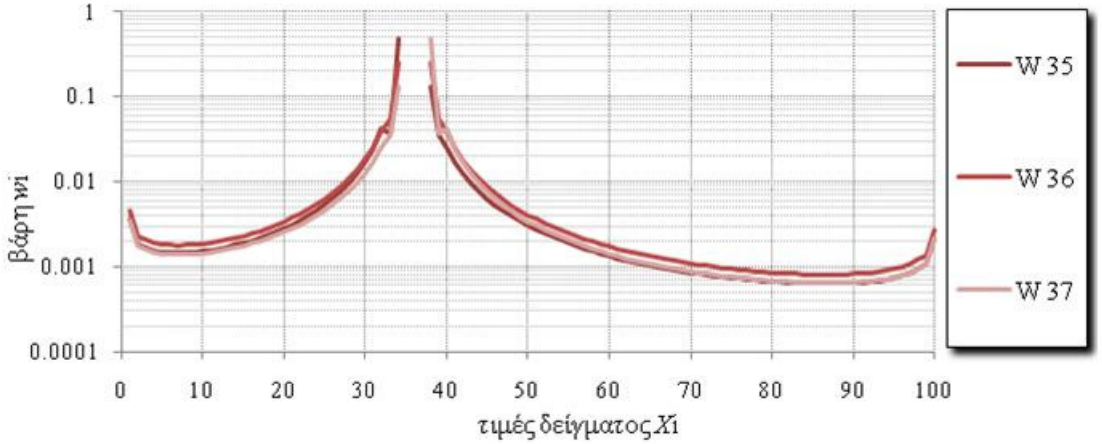



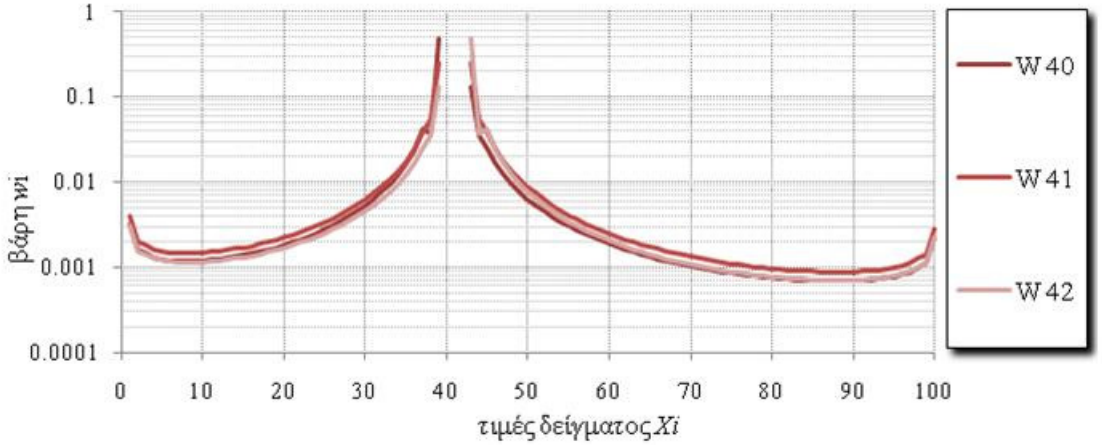

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 15
W 16
W 17

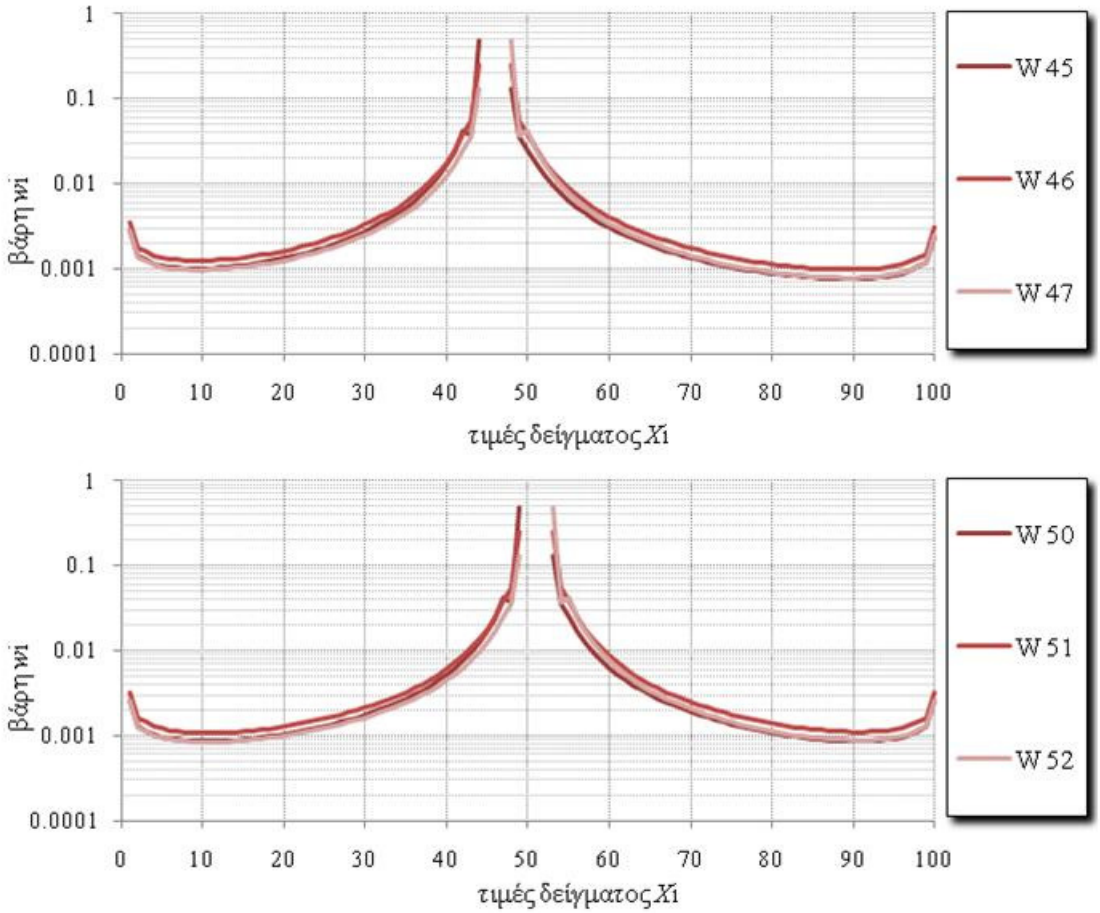

1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 20
W 21
W 22


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 25
W 26
W 27


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 30
W 31
W 32


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 35
W 36
W 37


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 40
W 41
W 42


1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 45
W 46
W 47
1
0.1
0.01
0.001
0.0001
βάρη wi
0
10
20
30
40
50
60
70
80
90
100
τιμές δείγματος Xi
W 50
W 51
W 52