



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ
ΠΕΡΙΒΑΛΛΟΝΤΟΣ

Διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής
στην ετήσια βροχόπτωση από μία παγκόσμια βάση
δεδομένων

Διπλωματική εργασία

Ηλιοπούλου Θεανώ

Επιβλέπων : Δημήτρης Κουτσογιάννης, Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2013

Ευχαριστίες

Τελειώνοντας τη διπλωματική μου εργασία και συνάμα την φοιτητική μου πορεία νιώθω την ανάγκη να ευχαριστήσω τους ανθρώπους που είχαν καθοριστική συμβολή και στα δύο.

Έτσι, ευχαριστώ θερμά τον επιβλέποντά μου, Καθηγητή Δημήτρη Κουτσογιάννη, αφενός για το χρόνο, τις συμβουλές, το ενδιαφέρον και τη διορατική καθοδήγηση που μου προσέφερε κατά την εκπόνηση αυτής της εργασίας και αφετέρου, γιατί με την παρουσία του ως καθηγητή και ανθρώπου στη σχολή των πολιτικών μηχανικών μου ενέπνευσε το ενδιαφέρον για τη μελέτη των Υδατικών Πόρων και ειδικά για τη Στοχαστική Υδρολογία.

Ευχαριστώ εξίσου θερμά το Σίμωνα Παπαλεξίου, υποψήφιο διδάκτορα, για την καθοριστική βοήθειά του στην εκπόνηση αυτής της εργασίας, τις ιδέες του και το χρόνο που αφιέρωσε.

Ευχαριστώ επίσης τη Σοφία, με την οποία μας συνδέει κοινή πορεία κατά τη διάρκεια των τελευταίων δύο χρόνων, για τις συμβουλές της, τη συμπαράστασή της αλλά και την πρακτική βοήθεια που μου προσέφερε και είχε πάντα τη διάθεση να προσφέρει όποτε το χρειαζόμουν.

Τέλος, η συμπαράσταση και ψυχολογική υποστήριξη της οικογένειάς μου και των δικών μου ανθρώπων καθόλη τη διάρκεια της παρούσας διπλωματικής αλλά και των δύσκολων στιγμών της φοιτητικής μου πορείας είναι ανεκτίμητη.

Θεανώ Ηλιοπούλου

Μάρτιος 2013

Περιεχόμενα

ΕΥΧΑΡΙΣΤΙΕΣ.....	iii
ΠΕΡΙΕΧΟΜΕΝΑ.....	v
ΠΕΡΙΛΗΨΗ.....	ix
ABSTRACT.....	xi
1. ΕΙΣΑΓΩΓΗ.....	1
1.1 ΓΕΝΙΚΑ.....	1
1.2 ΑΝΤΙΚΕΙΜΕΝΟ ΕΡΓΑΣΙΑΣ.....	2
1.3 ΔΙΑΡΘΡΩΣΗ ΕΡΓΑΣΙΑΣ.....	3
2. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΝΝΟΙΑ ΤΗΣ ΧΡΟΝΙΚΗΣ ΕΞΑΡΤΗΣΗΣ.....	5
2.1 Η ΕΝΝΟΙΑ ΤΗΣ ΑΠΛΗΣ ΣΤΟΧΑΣΤΙΚΗΣ ΑΝΕΛΙΞΗΣ.....	5
2.2 ΑΠΛΑ ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ.....	8
2.2.1 ΜΟΝΤΕΛΟ ΛΕΥΚΟΥ ΘΟΡΥΒΟΥ.....	8
2.2.2 ΜΟΝΤΕΛΟ AR(1).....	9
2.3 ΑΠΟΚΛΙΣΕΙΣ ΑΠΟ ΤΑ ΑΠΛΑ ΜΟΝΤΕΛΑ.....	11
3. ΤΟ ΦΑΙΝΟΜΕΝΟ ΤΗΣ ΜΑΚΡΟΠΡΟΘΕΣΜΗΣ ΕΜΜΟΝΗΣ.....	17
3.1 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ.....	17
3.2 ΜΑΘΗΜΑΤΙΚΟΣ ΟΡΙΣΜΟΣ.....	18
3.3 ΕΠΙΔΡΑΣΗ ΣΤΗΝ ΕΚΤΙΜΗΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ.....	20
3.4 ΜΟΝΤΕΛΟ ΔΙΑΤΗΡΗΣΗΣ ΤΗΣ ΕΜΜΟΝΗΣ.....	26
3.5 ΑΛΓΟΡΙΘΜΟΣ ΑΝΑΠΑΡΑΓΩΓΗΣ ΚΛΑΣΜΑΤΙΚΟΥ ΓΚΑΟΥΣΙΑΝΟΥ ΘΟΡΥΒΟΥ.....	28
3.6 ΦΥΣΙΚΗ ΕΡΜΗΝΕΙΑ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ.....	29
4. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΜΑΚΡΟΠΡΟΘΕΣΜΗΣ ΕΜΜΟΝΗΣ.....	31
4.1 ΤΟ ΖΗΤΗΜΑ ΤΗΣ ΕΥΡΩΣΤΙΑΣ.....	31

4.2 Η ΜΕΘΟΔΟΣ R/S.....	33
4.3 Η ΜΕΘΟΔΟΣ ΤΗΣ ΣΥΝΑΘΡΟΙΣΜΕΝΗΣ ΔΙΑΣΠΟΡΑΣ.....	35
4.4 Η ΜΕΘΟΔΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ ΒΑΣΙΣΜΕΝΗ ΣΤΗΝ ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ (LSSD).....	36
4.5 ΜΕΘΟΔΟΣ ΤΩΝ ΑΠΟΛΥΤΩΝ ΤΙΜΩΝ ΤΗΣ ΣΥΝΑΘΡΟΙΣΜΕΝΗΣ ΔΙΑΣΠΟΡΑΣ.....	39
4.6 ΜΕΘΟΔΟΣ ΤΟΥ HIGUCHI.....	40
4.7 ΜΕΘΟΔΟΣ ΤΟΥ ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΟΣ.....	41
4.8 ΜΕΘΟΔΟΣ ΤΟΥ PENG (DFA).....	42
4.9 ΕΚΤΙΜΗΤΡΙΕΣ ΤΟΥ WHITTLE.....	43
4.10 ΜΕΘΟΔΟΣ ΚΥΜΑΤΙΚΗΣ ΑΝΑΛΥΣΗΣ (WAVELET ANALYSIS).....	46
4.11 ΑΛΛΕΣ ΜΕΘΟΔΟΙ.....	48
4.12 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΜΕΛΕΤΩΝ ΓΙΑ ΤΗΝ ΥΠΑΡΞΗ ΜΑΚΡΟΠΡΟΘΕΣΜΗΣ ΕΜΜΟΝΗΣ ΣΤΟ ΦΑΙΝΟΜΕΝΟ ΤΗΣ ΒΡΟΧΟΠΤΩΣΗΣ.....	48
5. ΕΦΑΡΜΟΓΗ ΣΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ.....	51
5.1 ΠΕΡΙΓΡΑΦΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ.....	51
5.2 ΣΥΜΠΛΗΡΩΣΗ ΔΕΔΟΜΕΝΩΝ.....	54
5.3 ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΤΕΛΙΚΩΝ ΣΤΑΘΜΩΝ.....	59
5.4 ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΥ ΣΥΝΑΘΡΟΙΣΜΕΝΗΣ ΔΙΑΣΠΟΡΑΣ.....	65
5.4.1 ΕΠΙΔΡΑΣΗ ΠΑΡΑΜΕΤΡΩΝ ΣΤΗΝ ΕΚΤΙΜΗΣΗ.....	69
5.5 ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΥ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ ΒΑΣΙΣΜΕΝΗ ΣΤΗΝ ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ (LSSD).....	71
5.5.1 ΣΥΓΚΡΙΣΗ ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ.....	73
5.6 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΚΟΙΝΟΥ ΣΥΝΤΕΛΕΣΤΗ HURST.....	75
5.7 ΕΛΕΓΧΟΣ ΔΟΜΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ.....	77
6. ΣΥΜΠΕΡΑΣΜΑΤΑ.....	89

6.1 ΣΥΝΟΨΗΣ ΜΕΛΕΤΗΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	89
6.2 ΣΥΜΠΕΡΑΣΜΑΤΑ.....	91
6.3 ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ.....	92
7. ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ.....	93
ΠΑΡΑΡΤΗΜΑ Α.....	103
ΠΑΡΑΡΤΗΜΑ Β.....	117

Περίληψη

Η διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής σε γεωφυσικά δεδομένα συγκεντρώνει το ενδιαφέρον των επιστημόνων τις τελευταίες δεκαετίες μετά την ανακάλυψη του φαινομένου από τον Hurst (1951). Εδώ μελετάται η υπόθεση ύπαρξης μακροπρόθεσμης εμμονής στην ετήσια βροχόπτωση. Για το σκοπό αυτό χρησιμοποιήθηκαν ημερήσια δεδομένα βροχόπτωσης από την παγκόσμια βάση δεδομένων GHCN (Global Historical Climatology Network), τα οποία μετά από επεξεργασία και συμπλήρωση χρησιμοποιήθηκαν για τον υπολογισμό του συντελεστή μακροπρόθεσμης εμμονής H . Επιλέχθηκαν 1265 σταθμοί, βάσει κριτηρίων καταγραφών και ποιότητας, οι οποίοι εξετάστηκαν επιπρόσθετα και ως προς ορισμένα χαρακτηριστικά της βροχόπτωσης. Για την εκτίμηση της μακροπρόθεσμης εμμονής, εφαρμόστηκαν η μέθοδος της συναθροισμένης διασποράς και η μέθοδος ελαχίστων τετραγώνων βασισμένη στην τυπική απόκλιση LSSD (Least Squares based on Standard Deviation). Επίσης, έγινε έλεγχος της υπόθεσης ύπαρξης κοινού συντελεστή H στα δεδομένα. Για την πληρέστερη μελέτη της δομής εξάρτησης μελετήθηκε η δομή της αυτοσυσχέτισης των δεδομένων και πραγματοποιήθηκε έλεγχος για τη σχέση της με δομή AR(1). Τα αποτελέσματα συνηγορούν υπέρ της ύπαρξης ενδείξεων μακροπρόθεσμης εμμονής στα 3/4 των σταθμών, με διαφοροποιήσεις ως προς την ισχύ της. Εκτενέστερη ανάλυση χρειάζεται πιθανόν μελλοντικά για την αξιοποίηση μεγαλύτερων χρονοσειρών που θα είναι διαθέσιμες τότε.

Abstract

High scientific interest is raised around the hypothesis testing of long range dependence structure in geophysical data since Hurst discovered the phenomenon in 1951. Herein, we examine the hypothesis of long range dependence in the annual rainfall structure. A global data base of daily rainfall data GHCN (Global Historical Climatology Network) is processed along with the infilling of missing values. Based on length and quality criteria, 1265 rainfall stations are chosen for the estimation. Statistical properties of rainfall are also derived from the stations. In order to estimate the Hurst coefficient, the aggregated variance method as well as the LSSD (Least Squares based on Standard Deviation) method are applied to the data. Hypothesis testing for the existence of a common Hurst coefficient is conducted. In addition, the study of the autocorrelation function and the comparison with an AR(1) structure are performed as they provide a deeper insight on the matter. The results are in favor of the existence of long range dependence for the 3/4 of rainfall stations, although there exist considerable differences among them in the intensity of the phenomenon. In the future, the longer existing data lengths will perhaps enable a more certain estimation.

1. Εισαγωγή

1.1 Γενικά

Η μελέτη των γεωφυσικών διαδικασιών υπό το πρίσμα στοχαστικής ανάλυσης και συγκεκριμένα η αντιμετώπιση των γεωφυσικών μεταβλητών ως στοχαστικών ανελίξεων είναι ένα πεδίο που συγκεντρώνει επιστημονικό ενδιαφέρον κατά τη διάρκεια των τελευταίων δεκαετιών. Είναι πλέον γνωστό ότι τα γεωφυσικά φαινόμενα χαρακτηρίζονται από εγγενή τυχαιότητα και επομένως δεν μπορούν να περιγραφούν ντετερμινιστικά. Έχει παρατηρηθεί επίσης η ύπαρξη σημαντικής χρονικής εξάρτησης στη δομή πολλών εξ αυτών, γεγονός που καθιστά τους νόμους της κλασικής στατιστικής η οποία βασίζεται στην αρχή της ανεξαρτησίας των παρατηρήσεων, άτοπους και παραπλανητικούς. Μεγάλο ενδιαφέρον παρουσιάζει η διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής σε γεωφυσικά δεδομένα καθώς επηρεάζει δραματικά την εκτίμηση των στατιστικών χαρακτηριστικών τους, τη δυνατότητα «πρόβλεψης» μελλοντικών παρατηρήσεων και τελικά την προσομοίωση των αντίστοιχων φαινομένων.

Οι μελέτες προς αυτή την κατεύθυνση συναντούν συχνά το πρόβλημα της έλλειψης επαρκούς μήκους δείγματος λόγω του σχετικά πρόσφατου επιστημονικού ενδιαφέροντος για την καταγραφή μετρήσεων φαινομένων. Το φαινόμενο της μακροπρόθεσμης εμμονής απαιτεί καταγραφές αρκετών ετών για να μελετηθεί ικανοποιητικά και δεν υπάρχουν πολλά διαθέσιμα δείγματα επαρκούς μήκους. Ένα δεύτερο πρόβλημα που συναντάται είναι η δυσκολία επεξεργασίας δεδομένων που διατίθενται στο διαδίκτυο σε μορφές μη φιλικές προς τα συνήθη προγράμματα μαθηματικών εργαλείων ενώ πολύ συχνά η παρουσία προβληματικών παρατηρήσεων και ελλιπών τιμών θέτει ερωτήματα για την αξιοπιστία των δεδομένων και την ορθή διαχείρισή τους. Σε αυτά τέλος προστίθεται το ήδη σύνθετο πρόβλημα της ορθής επιλογής και εφαρμογής της μεθόδου εκτίμησης του φαινομένου.

Υπάρχουν αρκετές δημοσιευμένες εργασίες που έχουν ήδη ασχοληθεί με την διερεύνηση της υπόθεσης μακροπρόθεσμης εμμονής σε μια σειρά μεταβλητών όπως η απορροή ποταμών, η θερμοκρασία, η ατμοσφαιρική πίεση, η ταχύτητα του ανέμου καθώς και τοπικά μετεωρολογικά φαινόμενα (π.χ. El Nino). Τις περισσότερες φορές αφορούν τοπικά δείγματα ή επεξεργασία μετρήσεων ορισμένων σταθμών και σπανιότερα έχουν παγκόσμιο χαρακτήρα. Θεωρείται λοιπόν ότι η παρούσα εργασία συμβάλει στη διεύρυνση του δείγματος των μελετών που ασχολούνται με την ύπαρξη του φαινομένου στις φυσικές διαδικασίες και προσθέτει σημαντικά συμπεράσματα τόσο για τη μεθοδολογία που χρησιμοποιήθηκε όσο και για τα τελικά αποτελέσματα.

1.2 Αντικείμενο εργασίας

Σκοπός της διπλωματικής είναι η διερεύνηση ύπαρξης μακροπρόθεσμης εμμονής στην ετήσια βροχόπτωση χρησιμοποιώντας μια παγκόσμια βάση δεδομένων. Δεδομένα από 3477 σταθμούς αναλύονται ως προς την ποιότητά τους σε συνάρτηση του αριθμού των κενών τους και του μήκους των καταγραφών τους. Προτείνεται μια μεθοδολογία επεξεργασίας κατά την οποία χρησιμοποιείται ήδη δημοσιευμένη μέθοδος συμπλήρωσης ελλιπών τιμών. Επιλέγονται βάσει κριτηρίων 1265 σταθμοί οι οποίοι εξετάζονται στατιστικά και ως προς ορισμένα χαρακτηριστικά της βροχόπτωσης. Εφαρμόζονται δύο μέθοδοι εκτίμησης της εμμονής στα δεδομένα και αναλύονται στατιστικά τα συμπεράσματα κάθε μίας. Πραγματοποιείται επίσης σύγκριση των αποτελεσμάτων των δύο μεθόδων. Στη συνέχεια επιχειρείται εύρεση πιθανής αντιπροσωπευτικής τιμής του συντελεστή Hurst για το σύνολο των καταγραφών. Τέλος, μελετάται αναλυτικότερα η δομή της αυτοσυσχέτισης με εκτίμηση των αυτοσυσχετίσεων μικρών τάξεων και πραγματοποίηση στατιστικού τεστ για τον έλεγχο πιθανής δομής AR(1). Απώτερος στόχος είναι η σύνθεση μιας συνολικής εικόνας για τη δομή αυτοσυσχέτισης που παρουσιάζουν τα δεδομένα της βροχόπτωσης λαμβάνοντας υπόψη την αβεβαιότητα της εκτίμησης.

1.3 Διάρθρωση της εργασίας

Η εργασία αποτελείται από 7 κεφάλαια, συμπεριλαμβανομένης και της εισαγωγής και από δύο παραρτήματα.

Στο δεύτερο κεφάλαιο παρουσιάζονται βασικές έννοιες της στοχαστικής υδρολογίας καθώς και δύο απλά στοχαστικά μοντέλα στα οποία ανατρέχει η ανάλυση αργότερα.

Στο τρίτο κεφάλαιο το ενδιαφέρον εστιάζεται στην ανάλυση του φαινομένου της μακροπρόθεσμης εμμοής μέσω της μαθηματικής του προσέγγισης, των αποτελεσμάτων της επίδρασης καθώς και των μοντέλων που το αναπαράγουν. Εξηγούνται βασικές έννοιες οι οποίες στη συνέχεια χρησιμοποιούνται στην εφαρμογή στα δεδομένα.

Στο τέταρτο κεφάλαιο πραγματοποιείται βιβλιογραφική επισκόπηση των πιο διαδεδομένων μεθόδων εκτίμησης της εμμοής. Ειδικά αναλύονται και εξηγούνται οι μέθοδοι που επιλέχθηκαν για την ανάλυση. Γίνεται επίσης μια ενδεικτική βιβλιογραφική επισκόπηση των μελετών που έχουν ασχοληθεί με τη διερεύνηση της μακροπρόθεσμης εμμοής στη βροχόπτωση.

Στο πέμπτο κεφάλαιο παρουσιάζεται η μεθοδολογία που εφαρμόστηκε στα πραγματικά δεδομένα και τα αποτελέσματα όλων των αναλύσεων που έγιναν.

Στο έκτο κεφάλαιο γίνεται η σύνοψη των αποτελεσμάτων και εξάγονται τα τελικά συμπεράσματα και οι προτάσεις της μελέτης.

Στο έβδομο κεφάλαιο δίνεται η λίστα με τις βιβλιογραφικές αναφορές.

Στο παράρτημα Α υπάρχει λεπτομερέστερη ανάλυση των σταθμών που παρουσίαζαν εντονότερα το φαινόμενο μακροπρόθεσμης εμμοής μέσω διαγραμμάτων και χαρακτηριστικών πινάκων.

Στο παράρτημα Β δίνονται όλοι οι κώδικες που χρησιμοποιήθηκαν για την ανάλυση.

2. Εισαγωγή στη θεωρία της χρονικής εξάρτησης

Στο κεφάλαιο αυτό εξηγούνται βασικές έννοιες της στοχαστικής υδρολογίας και περιγράφονται δύο απλά στοχαστικά μοντέλα, ο Λευκός Θόρυβος και το μοντέλο AR(1), στα οποία αναφέρεται στη συνέχεια η ανάλυση. Επίσης, παρουσιάζονται παραδείγματα αποκλίσεων πραγματικών δεδομένων από τα απλά μοντέλα καθώς και ορισμένες βιβλιογραφικές αναφορές σχετικά με την παρατήρηση αντίστοιχων συμπεριφορών (μακροπρόθεσμης εμμονής) σε πολλούς κλάδους της επιστήμης.

2.1 Η έννοια της απλής στοχαστικής ανέλιξης

Στοχαστική ανέλιξη X_i ονομάζεται ένα σύνολο τυχαίων μεταβλητών σε ακολουθία $\{X_1, X_2, X_3, \dots, X_i\}$ όπου i είναι παράμετρος που παίρνει τιμές από ένα κατάλληλο σύνολο T το οποίο συνήθως παριστάνει χρόνο. Η υλοποίηση μιας στοχαστικής ανέλιξης, δηλαδή ένα σύνολο παρατηρήσεων $x(i)$ της $X(i)$ για μεταβαλλόμενο χρόνο i ονομάζεται χρονοσειρά. Η στοχαστική ανέλιξη χαρακτηρίζεται από ένα σύνολο στατιστικών παραμέτρων με κυριότερες τις εξής:

- Μέση τιμή :

$$\mu := E[X_i] \quad (2.1)$$

- Τυπική απόκλιση:

$$\sigma := \sqrt{\text{Var}[X_i]} = \sqrt{E[(X_i - \mu)^2]} \quad (2.2)$$

- Αυτοσυνδιασπορά της ανέλιξης :

$$\gamma_j = Cov[X_i, X_{i+j}] = E[(X_i - \mu)(X_{i+j} - \mu)] \quad (2.3)$$

- Συντελεστή αυτοσυσχέτισης :

$$\rho(i; j) = \frac{Cov[X_i, X_{i+j}]}{\sqrt{Var[X_i]Var[X_{i+j}]}} \quad (2.4)$$

Ο συντελεστής αυτοσυσχέτισης (2.4) ανήκει στο διάστημα $[-1,1]$ με την τιμή 0 να δηλώνει ασυσχέτιστα δεδομένα και τις τιμές $\{-1,1\}$ να υποδηλώνουν απόλυτη αρνητική ή θετική συσχέτιση αντίστοιχα. Η τάξη που μελετάται στις υδρολογικές εφαρμογές αφορά την κλίμακα ενδιαφέροντος. Για παράδειγμα, για την μελέτη της ετήσιας απορροής ενός ταμιευτήρα η χρονική κλίμακα συνηθίζεται να είναι ένα ή περισσότερα χρόνια. Αντίστοιχα, προσομοίωση βροχοπτώσεων σε μικρή κλίμακα θα σήμαινε τιμές για την κλίμακα ενδιαφέροντος από μερικά λεπτά έως και ώρες (Koutsoyiannis 2002).

Στη γενική περίπτωση, οι στατιστικές παράμετροι μιας στοχαστικής ανέλιξης μεταβάλλονται με τη μεταβολή του χρόνου. Μια ειδική κατηγορία ανελιξεων είναι οι στάσιμες ανελιξεις στις οποίες δεν υπάρχει μεταβολή των στατιστικών χαρακτηριστικών με την πάροδο του χρόνου. Μια στοχαστική ανέλιξη λέγεται *στάσιμη με την ευρεία έννοια* αν η μέση τιμή της είναι σταθερή και η αυτοσυνδιασπορά της εξαρτάται μόνο από τη διαφορά του χρόνου. Μια στάσιμη στοχαστική ανέλιξη είναι *εργοδική* αν κάθε παράμετρος της κατανομής μπορεί να προσδιοριστεί από μια απλή δειγματοσυνάρτηση της ανέλιξης. Οι παραπάνω δύο ιδιότητες επιτρέπουν την εύκολη προσομοίωση φαινομένων και για αυτό γίνεται συνήθως η παραδοχή ότι ισχύουν.

Ανέλιξη στη συναθροισμένη κλίμακα

Έστω k η τιμή της χρονικής κλίμακας όπου k θετικός ακέραιος. Η συναθροισμένη ανέλιξη στην κλίμακα k ορίζεται ως εξής :

$$Z_i^{(k)} := \sum_{l=(i-1)k+1}^{ik} X_l \quad (2.5)$$

Έτσι, για $k = 1$, $Z_i^{(1)} = X_i$, για $k = 2$, $Z_1^{(2)} = X_1 + X_2$, $Z_2^{(2)} = X_3 + X_4$ κ.ο.κ..

Οι στατιστικές ιδιότητες της συναθροισμένης ανέλιξης $Z_i^{(k)}$ προκύπτουν από τις αντίστοιχες της X_i . Συγκεκριμένα, είναι:

- Μέση τιμή

$$E [Z_i^{(k)}] = k\mu \quad (2.6)$$

- αυτοσυνδιασπορά

$$\gamma_j^{(k)} = cov [Z_i^{(k)}, Z_{i+j}^{(k)}] = \sum_{l=1}^k \sum_{m=jk+1}^{(j+1)k} \gamma_{m-l} \quad j = 0, \pm 1, \pm 2, \dots \quad (2.7)$$

- αυτοσυσχέτιση

$$\rho_j^{(k)} := Corr [Z_i^{(k)}, Z_{i+j}^{(k)}] = \frac{\gamma_j^{(k)}}{\gamma_0^{(k)}} \quad j = 0, 1, 2, \dots \quad (2.8)$$

Η αυτοσυνδιασπορά συνδέεται με το δυναμικό φάσμα της ανέλιξης το οποίο στη γενική περίπτωση είναι ο διακριτός μετασχηματισμός Fourier (ή αντίστροφος μετασχηματισμός Fourier) :

$$s_Y^{(k)}(\omega) := 2\gamma_0^{(k)} + 4 \sum_{j=1}^{\infty} \gamma_j^{(k)} \cos(2\pi j\omega) = 2 \sum_{j=-\infty}^{\infty} \gamma_j^{(k)} \cos(2\pi j\omega) \quad (2.9)$$

Υποθέτουμε στην σχέση (2.9) ότι η συχνότητα ω παίρνει τιμές στο διάστημα $[0, 1/2]$ και επομένως η αυτοσυνδιασπορά $\gamma_j^{(k)}$ ορίζεται με βάση το μετασχηματισμό Fourier ως εξής :

$$s_Y^{(k)}(\omega) \gamma_j^{(k)} = \int_0^{1/2} s_Y^{(k)}(\omega) \cos(2\pi j\omega) d\omega \quad (2.10)$$

2.2 Απλά στοχαστικά μοντέλα

2.2.1 Μοντέλο Λευκού Θορύβου

Στο μοντέλο του Λευκού Θορύβου διαφορετικές ανελιξεις X_i είναι ανεξάρτητες τυχαίες μεταβλητές έτσι ώστε $\gamma_j = 0$ (και $\rho_j = 0$) για $j \neq 0$. Τότε η συναθροισμένη ανέλιξη έχει τα εξής χαρακτηριστικά:

- μέση τιμή:

$$E [Z_i^{(k)}] = kE[X_i] \quad (2.11)$$

- διασπορά:

$$\gamma_0^{(k)} := \text{Var} [Z_i^{(k)}] = k\gamma_0 \quad (2.12)$$

- αυτοσυνδιασπορά – αυτοσυσχέτιση:

$$\gamma_j^{(k)} = 0, \quad \rho_j^{(k)} = 0 \quad (2.13)$$

- φάσμα ισχύος :

$$s_\gamma^{(k)}(\omega) = 2\gamma_0^{(k)} \quad (2.14)$$

2.2.2 Μοντέλο AR(1)

Το μοντέλο AR(1) ή αλλιώς αλυσίδες Markov λόγω του ρώσου μαθηματικού που το ανακάλυψε στις αρχές του 20^{ου} αιώνα, είναι το απλούστερο στοχαστικό μοντέλο με διατήρηση κάποιας εξάρτησης. Συγκεκριμένα, η εξάρτηση αφορά στην εκτίμηση της κατάστασης (i) μιας στοχαστικής ανέλιξης με βάση την αμέσως προηγούμενη κατάσταση ($i-1$). Υποθέτοντας μια στάσιμη στοχαστική ανέλιξη X_i το μοντέλο στη βασική κλίμακα εκφράζεται ως εξής:

$$X_i = \rho X_{i-1} + V_i \quad (2.15)$$

όπου ρ είναι ο συντελεστής αυτοσυσχέτισης 1^{ης} τάξης και V_i η ακολουθία λευκού θορύβου με μέση τιμή $(1-\rho)\mu$ και διασπορά $(1-\rho^2)\gamma_0$. Η διαδικασία είναι Μαρκοβιανή επειδή η εξάρτηση της μεταβλητής X_i από τη μεταβλητή X_{i-1}

επαρκεί για να προσδιορισθεί πλήρως η εξάρτηση του μέλλοντος από το παρελθόν. Η αυτοσυσχέτιση της X_i είναι:

$$\rho_j := \text{corr}[X_i, X_{i+j}] = \rho^{|j|} \quad (2.16)$$

Για τη συναθροισμένη ανέλιξη μπορεί να δειχθεί (Koutsoyiannis 2002) ότι έχει τα εξής χαρακτηριστικά:

- διασπορά:

$$\gamma_0^{(k)} := \gamma_0 \frac{\kappa(1 - \rho^2) - 2\rho(1 - \rho^k)}{(1 - \rho)^2} \quad j \geq 1 \quad (2.17)$$

- αυτοσυνδιασπορά:

$$\gamma_j^{(k)} = \gamma_0 \frac{\rho^{kj-k+1} (1 - \rho^k)^2}{(1 - \rho)^2} \quad j \geq 1 \quad (2.18)$$

- αυτοσυσχέτιση:

$$\rho_j^{(k)} = \rho_1^{(k)} \rho^{k(j-1)} \quad \text{όπου} \quad \rho_1^{(k)} = \frac{\rho(1 - \rho^k)^2}{\kappa(1 - \rho^2) - 2\rho(1 - \rho^k)} \quad j \geq 1 \quad (2.19)$$

- φάσμα ισχύος:

$$\frac{s_Y^{(k)}(\omega)}{\gamma_0^{(k)}} = 2 + 4\rho_1^{(k)} \frac{\cos(2\pi\omega) - \rho^k}{1 + \rho^{2k} - 2\rho^k \cos(2\pi\omega)} \quad (2.20)$$

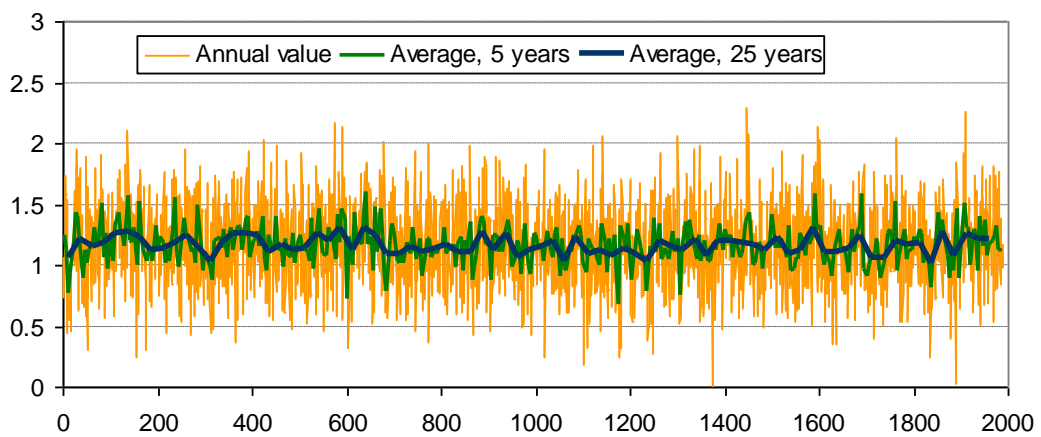
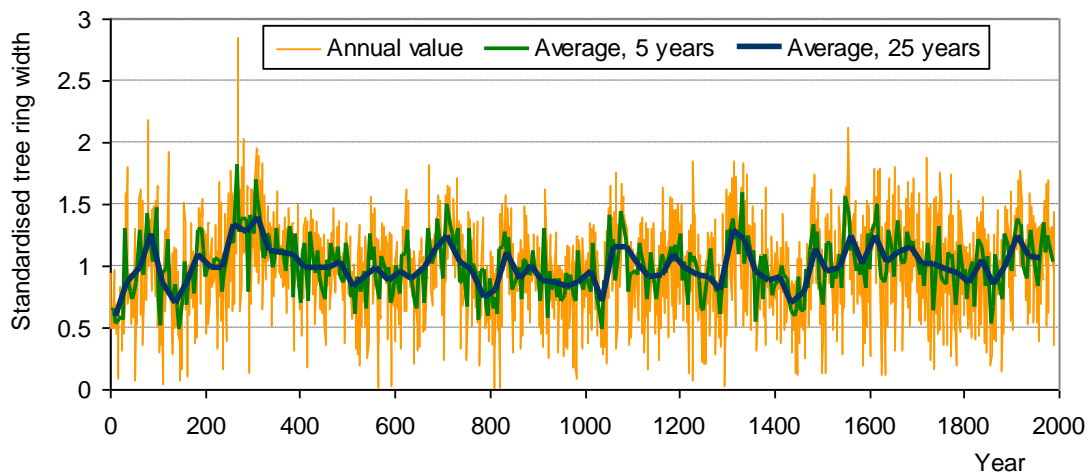
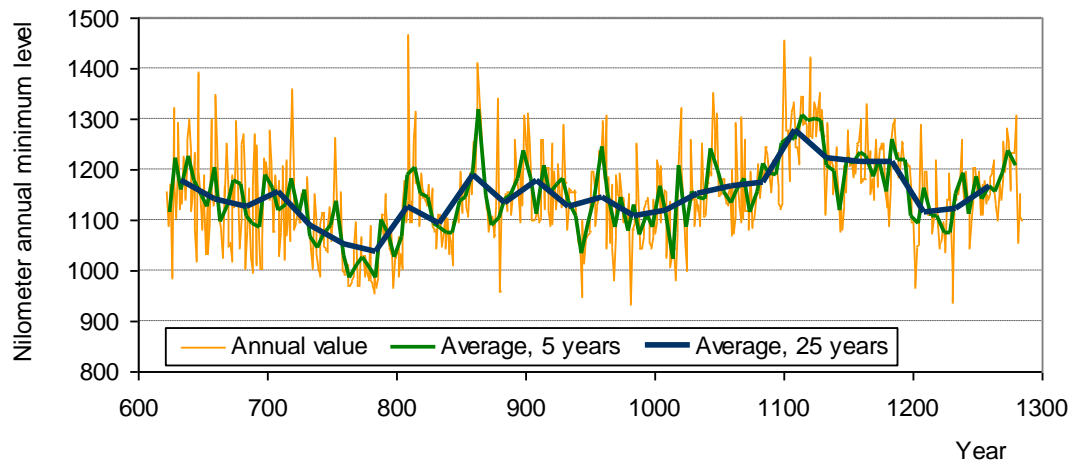
Αξίζει να σημειωθεί ότι από τη σύγκριση της σχέσης (2.19) με τη σχέση (2.16) προκύπτει ότι η συναθροισμένη ανέλιξη AR(1) δεν είναι ανέλιξη Markov και αποτελεί πιο πολύπλοκη δομή αυτοσυσχέτισης. Η απλή ανέλιξη AR(1) είναι δηλαδή ανέλιξη AR(1)/Markov μόνο στη βασική της κλίμακα.

Για μεγάλες τιμές της κλίμακας k παρατηρούμε από τη σχέση (2.17) ότι ο αριθμητής του κλάσματος επηρεάζεται περισσότερο από τον πρώτο όρο και η διασπορά της συναθροισμένης ανέλιξης γίνεται ανάλογη της κλίμακας k όπως και στην περίπτωση του λευκού θορύβου. Επίσης, μικραίνει η τιμή $\rho_1^{(k)}$ καθώς και η τιμή $\rho_j^{(k)}$ (σχέση 2.19). Φαίνεται από τη σχέση 2.20 ότι η τιμή του φάσματος γίνεται $\frac{s\gamma^{(k)}(\omega)}{\gamma_0^{(k)}} = 2$, ίδια δηλαδή με την τιμή του φάσματος στο λευκό θόρυβο. Τελικά, φαίνεται ότι αν η μελετώμενη ανέλιξη είναι τύπου Markov στη βασική κλίμακα τείνει στο λευκό θόρυβο όσο αυξάνεται η τιμή της κλίμακας συνάθροισης.

2.3 Αποκλίσεις από τα απλά μοντέλα

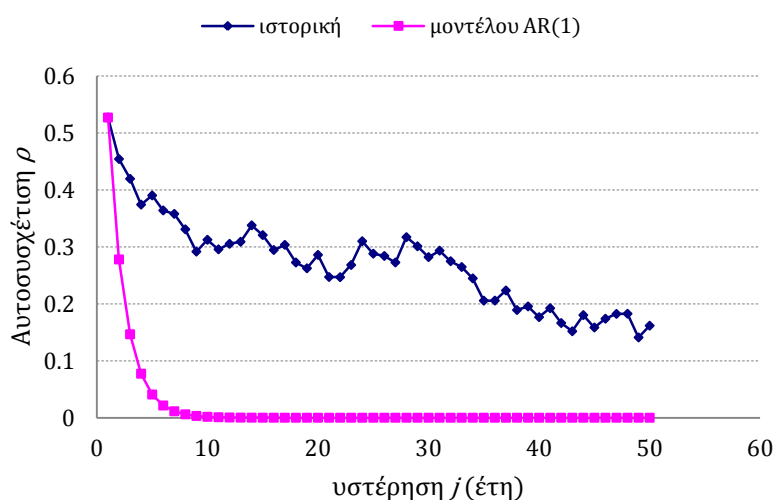
Από τις αρχές του 20^{ου} ήδη αιώνα στατιστικολόγοι και φυσικοί επιστήμονες είχαν παρατηρήσει αποκλίσεις στη δομή της αυτοσυσχέτισης από τα απλά στοχαστικά μοντέλα του λευκού θορύβου και των ανελιξεων Markov. Η υπόθεση της ανεξαρτησίας των παρατηρήσεων που μέχρι τότε ήταν πολύ κοινή στην κλασική στατιστική είχε αρχίσει να διαφαίνεται πως ήταν περισσότερο η εξαίρεση παρά ο κανόνας. Υπάρχουν πολλές αναφορές στην ύπαρξη εμμονής σε αστρονομικά, χημικά και γεωφυσικά δεδομένα (Jan Beran 1992) από επιστήμονες της εποχής που περιέγραφαν το φαινόμενο προτού γίνει εφικτή η μαθηματική του διατύπωση από το Βρετανό μηχανικό Hurst το 1951. Ο Hurst μελέτησε τη χρονοσειρά της ετήσιας ελάχιστης στάθμης του ποταμού Νείλου για την οποία υπήρχαν καταγραφές από το έτος 622 μέχρι 1284 μ.Χ. και παρατήρησε την τάση ομαδοποίησης <ομοειδών> παρατηρήσεων. Η χρονοσειρά των δεδομένων του ποταμού Νείλου αποτελεί μια από τις μελετημένες χρονοσειρές για το θέμα λόγω της συμβολής της στην ανακάλυψη του φαινομένου αλλά και του μεγάλου πλήθους καταγραφών της που δεν συνηθίζεται σε γεωφυσικά δεδομένα. Στο σχήμα 2.1 φαίνεται το διάγραμμα της χρονοσειράς της ετήσιας στάθμης του ποταμού Νείλου, το διάγραμμα της χρονοσειράς των τυποποιημένων δεντροδακτύλιων της περιοχής Mammoth στην Γιούτα καθώς και μια χρονοσειρά λευκού θορύβου με τα ίδια στατιστικά

χαρακτηριστικά για σύγκριση. Παρατηρούμε ότι οι διακυμάνσεις της συναθροισμένης ανέλιξης ειδικά για $k = 25$ είναι πολύ μεγαλύτερες στην ιστορική χρονοσειρά παρά στη χρονοσειρά λευκού θορύβου με τα ίδια στατιστικά χαρακτηριστικά. Αξίζει να σημειωθεί ότι ενώ τμήματα και των δύο ιστορικών χρονοσειρών μπορούν να παρερμηνευθούν ως ντετερμινιστική τάση ανόδου ή καθόδου, η συνολική εικόνα οδηγεί στην παρατήρηση ότι πρόκειται μάλλον για τυχαίες διακυμάνσεις σε πολλαπλές κλίμακες.

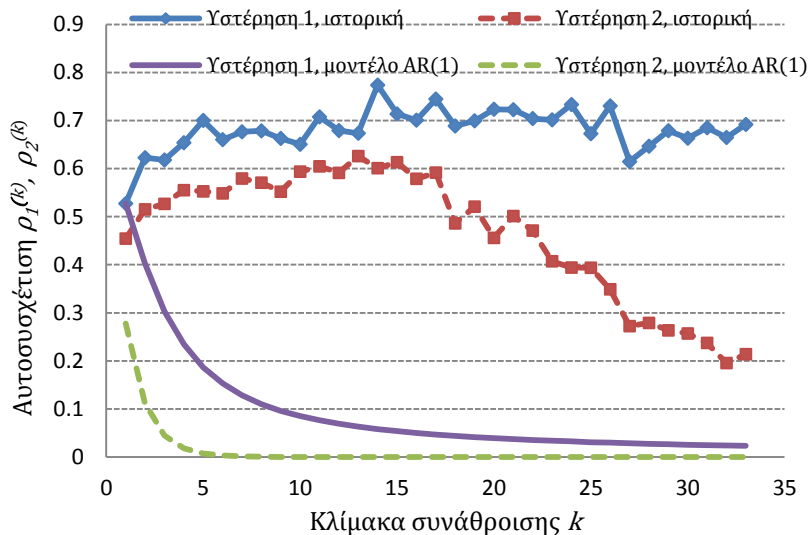


Σχήμα 2.1 Διάγραμμα της χρονοσειράς της ετήσιας στάθμης του ποταμού Νείλου, της χρονοσειράς των τυποποιημένων δεντροδακτυλίων της περιοχής Mammoth στη Γιούτα και μίας χρονοσειράς λευκού θορύβου με τα ίδια στατιστικά χαρακτηριστικά . Πηγή: Koutsoyiannis 2002

Ένα άλλο παράδειγμα απόκλισης από απλά στοχαστικά μοντέλα προκύπτει από την παρατήρηση της δομής της αυτοσυσχέτισης που παρουσιάζει το δείγμα παλαιοκλιματικών θερμοκρασιακών ανωμαλιών του Βορείου Ημισφαιρίου περιόδου 1000-1991 AD. Η σειρά ανακατασκευάστηκε χρησιμοποιώντας «θερμοκρασιακά ευαίσθητα» παλαιοκλιματικά δεδομένα όπως δακτυλίους δέντρων, κοράλλια, δείγματα πάγου και ιστορικά δεδομένα (Jones et al. 1998). Όπως φαίνεται στο σχήμα 2.2 το απλό μοντέλο AR(1) αποτυγχάνει να διατηρήσει τη δομή της αυτοσυσχέτισης καθώς η τελευταία φθίνει εκθετικά και μηδενίζεται για υστέρηση μεγαλύτερη των 7 ετών. Στο σχήμα 2.2 φαίνεται επίσης ότι ενώ οι αυτοσυσχετίσεις των δύο πρώτων τάξεων της συναθροισμένης ανέλιξης παραμένουν σχεδόν σταθερές χρονικά για την ιστορική χρονοσειρά, οι αντίστοιχες που προκύπτουν από το μοντέλο AR(1) συγκλίνουν στον μηδέν όσο αυξάνεται η κλίμακα συνάθροισης.



Σχήμα 2.2 Διάγραμμα της αυτοσυσχέτισης σε συνάρτηση με την τιμή της υστέρησης για την ιστορική χρονοσειρά και σε σύγκριση την αντίστοιχη θεωρητική AR(1) μοντέλου.



Σχήμα 2.3 Διάγραμμα της αυτοσυσχέτισης των δύο πρώτων τάξεων σε συνάρτηση με την κλίμακα συνάθροισης για τη συναθροισμένη ιστορική χρονοσειρά και σε σύγκριση με την αντίστοιχη θεωρητική AR(1) μοντέλου.

Έκτοτε έχουν γίνει πολλές συστηματικές μελέτες στις οποίες παρατηρήθηκε το φαινόμενο σε ποικίλους επιστημονικούς κλάδους. Χαρακτηριστικές είναι οι μελέτες για την επίδραση του φαινομένου στον κλάδο της οικονομίας και μακρο-οικονομίας και με έμφαση συχνά στις τιμές των μετοχών (Orcutt and Irwin 1948; Ding, Granger, and Engle 1993; Lux 1996; Henry and Zaffaroni 2002; Cajueiro and Tabak 2004; Carbone, Castelli, and Stanley 2004; Mikosch and Starica 2004; Grech and Pamu\la 2008) ενώ έχει βρει εφαρμογή και σε κοινωνικές επιστήμες (Lipset 1988; Asur et al. 2011). Το μεγαλύτερο ενδιαφέρον παρουσιάζει όμως η ύπαρξη του φαινομένου σε γεωφυσικές μεταβλητές όπως : η απορροή ποταμών (Eltahir 1996; Puckridge, Walker, and Costelloe 2000; Koutsoyiannis 2003; Blender and Fraedrich 2006; Koscielny-Bunde et al. 2006; Daniels 2007), η ταχύτητα του ανέμου (Bakker and Van den Hurk 2012), η θερμοκρασία (Bloomfield 1992; Zhang et al. 2000; Santer et al. 2000; Fraedrich and Blender 2003; William Rea, Reale, and Brown 2011), οι κλιματικοί δείκτες της βόρειας ατλαντικής ταλάντωσης (Stephenson, Pavan, and Bojariu 2000; Wanner et al. 2001). Αξίζει να σημειωθεί ότι δεδομένα ηλεκτρονικών δικτύων έχουν μελετηθεί επίσης ως προς την ύπαρξη τάσεων

(J. Beran et al. 1995; Abry and Veitch 1998; Kettani and Gubner 2002; Karagiannis, Molle, and Faloutsos 2004; Stoen et al. 2005) ενώ έχει διαπιστωθεί και εμφάνιση του φαινομένου στην κλίμακα του χώρου (Kendall 1943; Kaiser 1984). Τα τελευταία χρόνια το ενδιαφέρον της επιστημονικής κοινότητας έχει αναζωπυρωθεί για το φαινόμενο του Hurst λόγω της σύνδεσής του με τις κλιματικές αλλαγές (Evans 1996; Koutsoyiannis 2003).

3. Το φαινόμενο της μακροπρόθεσμης εμμονής

Στο κεφάλαιο αυτό γίνεται μια σύντομη ανάλυση και παρουσίαση των βασικών εννοιών που συνδέονται με το φαινόμενο της μακροπρόθεσμης εμμονής. Περιγράφεται το βασικό μοντέλο που τη διατηρεί το οποίο χρησιμοποιείται και στη συνέχεια και επίσης τονίζονται οι επιδράσεις που έχει στην επίδραση των στατιστικών παραμέτρων. Η αρνητική μεροληψία που εισάγει στην εκτίμηση της τυπικής απόκλισης και της αυτοσυσχέτισης είναι ζητήματα που αφορούν ιδιαίτερα την εφαρμογή στα πραγματικά δεδομένα.

3.1 Ιστορική αναδρομή

Οι πρώτες αναφορές σε φαινόμενα που εμφάνιζαν επαναλαμβανόμενη συμπεριφορά σε μακρά κλίμακα πρέπει να αποδοθούν στους αρχαίους Αιγύπτιους για τις παρατηρήσεις της συμπεριφοράς του ποταμού Νείλου. Πράγματι, ήταν γνωστό από τους αρχαίους χρόνους ότι ο ποταμός παρουσίαζε μεγάλες περιόδους ξηρασίας ακολουθούμενες από περιόδους έντονων πλημμυρών. Μια πολλή σχηματική καταγραφή του φαινομένου εντοπίζεται και στη Βίβλο με τη χαρακτηριστική αναφορά «Εφτά χρόνια μεγάλης αφθονίας έρχονται στη γη της Αιγύπτου όμως θα ακολουθήσουν εφτά χρόνια μεγάλης πείνας» (Γένεση 41,29-31). Αν και δεν υπάρχουν επίσημες καταγραφές για αυτά τα χρόνια η μελέτη των ύστερων παρατηρήσεων κατά τη διάρκεια των χρόνων 622-1281 μ.Χ. δείχνει ότι είναι λογικό να υποθέσει κανείς μια αντίστοιχη συμπεριφορά του ποταμού και στο παρελθόν.

Το φαινόμενο της εμμονής ήταν γνωστό στους επιστημονικούς κύκλους πολύ πριν γίνει εφικτό να περιγραφεί μαθηματικά. Εμπειρικά είχε παρατηρηθεί ότι οι αυτοσυσχετίσεις γεγονότων απομακρυσμένων στο χώρο ή στο χρόνο έφθιναν προς το μηδέν με πολύ αργότερο ρυθμό από αυτόν που θα περίμενε κανείς με βάση την υπόθεση της ανεξαρτησίας ή τη δομή Markov.

Η ανάλυση των ιστορικών στοιχείων του ποταμού Νείλου που προήλθαν από μετρήσεις κατά τη διάρκεια των χρόνων 622-1281 οδήγησε στην ανακάλυψη του ομώνυμου φαινομένου από το Hurst (1951).

Ο Hurst προσπάθησε να βρει έναν τρόπο να κανονικοποιήσει τη ροή του ποταμού καθώς επιχειρούσε να μελετήσει την ιδανική αποθηκευτική ικανότητα ενός ταμιευτήρα. Κατέληξε να κατασκευάσει ένα διάγραμμα γνωστό ως μέθοδος R/S (Rescaled Adjusted Range), τα αποτελέσματα του οποίου έρχονταν σε αντίθεση με αυτά που θα περίμενε κανείς έως τότε από τις δομές Markov. Η ίδια στατιστική συμπεριφορά είχε διαπιστωθεί περίπου 10 χρόνια νωρίτερα από τον Kolmogorov (1940) στη συμπεριφορά της τύρβης. Εισηγήθηκε για να τη μελετήσει ένα νέο μαθηματικό μοντέλο γνωστό ως στοχαστική ανέλιξη αυτοόμοια ή μοντέλο απλής ομοιοθεσίας (Simple Scaling Stochastic Process, SSS). Προς αναγνώρισή του η μακροπρόθεσμη εμμογή, εκτός από φαινόμενο Hurst αναφέρεται συχνά και ως δυναμική Hurst – Kolmogorov (HK Dynamics).

Οι εμπειρικές παρατηρήσεις του Hurst ώθησαν τον Mandelbrot και τους συνεργάτες του (1968) να αναπτύξουν το μοντέλο του Κλασματικού Γκαουσιανού Θορύβου, του πρώτου στην ουσία στατιστικού μοντέλου που αναπαρήγαγε μακροπρόθεσμη εμμογή. Μάλιστα ο Mandelbrot ορμώμενος από τη βιβλική φράση ονόμασε το φαινόμενο «φαινόμενο του Ιωσήφ» (J. Beran 1994).

3.2 Μαθηματικός ορισμός

Σύμφωνα με τον Beran(1994) οι παρακάτω δύο ορισμοί της μακροπρόθεσμης εμμογής είναι μαθηματικά ισοδύναμοι:

- *Ορισμός με βάση την αυτοσυσχέτιση*

Αν η X_t είναι μια στάσιμη ακολουθία για την οποία ισχύει : Υπάρχει ένας πραγματικός αριθμός α που ανήκει στο διάστημα $(0,1)$ και μια σταθερά $c_p > 0$ τέτοια ώστε:

$$\lim_{\kappa \rightarrow \infty} \frac{\rho(\kappa)}{c_p * \kappa^{-\alpha}} = 1 \quad (3.1)$$

τότε η X_t είναι στάσιμη ακολουθία με μακροπρόθεσμη εμμογή.

Χρησιμοποιείται επίσης και η παράμετρος $H = 1 - \frac{\alpha}{2}$, γνωστή ως συντελεστής του Hurst, αντί για το α . Μακροπρόθεσμη εμμονή εμφανίζεται για $1/2 < H < 1$. Για τη σχέση των αυτοσυσχετίσεων ισχύει:

$$\rho(k) \approx c_\rho |k|^{-\alpha} \quad (3.2)$$

καθώς το $|k|$ τείνει στο άπειρο και το c_ρ είναι μια ορισμένη θετική σταθερά.

- *Ορισμός με βάση τη φασματική πυκνότητα*

Η γνώση των αυτοσυνδιασπορών ισοδυναμεί με τη γνώση της φασματικής πυκνότητας f . Επομένως η εξάρτηση μεγάλου εύρους μπορεί να ορισθεί επίσης διατυπώνοντας μια υπόθεση για τη φασματική πυκνότητα ως εξής: Αν η X_t είναι μια στάσιμη ακολουθία για την οποία ισχύει: Υπάρχει ένας πραγματικός αριθμός β που ανήκει στο διάστημα $(0,1)$ και μια σταθερά $c_f > 0$, τέτοια ώστε:

$$\lim_{\lambda \rightarrow \infty} \frac{f(\lambda)}{c_f * |\lambda|^{-\beta}} = 1 \quad (3.3)$$

τότε η X_t είναι στάσιμη ακολουθία με μακροπρόθεσμη εμμονή.

Είναι σημαντικό να παρατηρήσουμε ότι οι δύο αυτοί ορισμοί (3.1) και (3.3) ισχύουν μόνο ασυμπτωτικά. Η πληροφορία που μας δίνουν αφορά τη συνολική συμπεριφορά των αυτοσυσχετίσεων καθώς η υστέρηση τείνει στο άπειρο. Γενικά δηλαδή αν έχουμε μια στάσιμη ακολουθία με αυτοσυσχετίσεις που τείνουν στο μηδέν με πολύ αργό ρυθμό τότε την ονομάζουμε μια στάσιμη ακολουθία με μακροπρόθεσμη εμμονή. Σε αυτό το γενικό πλαίσιο όμως δεν προσδιορίζονται οι συσχετίσεις για καμία ορισμένη πεπερασμένη υστέρηση. Αυτό που προσδιορίζεται είναι μόνο ο ρυθμός της σύγκλισης και όχι το απόλυτο μέγεθος της. Κάθε μεμονωμένη συσχέτιση μπορεί να είναι ιδιαίτερα μικρή. Μόνο η σύγκλιση των συσχετίσεων είναι αργή.

- Ορισμός με βάση την τυπική απόκλιση

Η μελέτη της συμπεριφοράς της τυπικής απόκλισης στη συναθροισμένη ανέλιξη μπορεί να χρησιμεύσει ως ένας απλούστερος ορισμός της μακροπρόθεσμης εμμοής (Koutsoyiannis 2002). Είναι γνωστό ότι στην κλασική στατιστική ισχύει:

$$\sigma^{(k)} = \frac{\sigma}{\sqrt{k}} \quad (3.4)$$

όπου $\sigma^{(k)}$ η τυπική απόκλιση της συναθροισμένης ανέλιξης $X_i^{(k)}$ και k η κλίμακα συνάθροισης. Σε χρονοσειρές που παρουσιάζουν μακροπρόθεσμη εμμοή αντί της σχέσης (3.4) παρατηρείται ένας γενικότερος νόμος της μορφής :

$$\sigma^{(k)} = \frac{\sigma}{k^{1-H}} \quad (3.5)$$

όπου H είναι μια σταθερά που για θετικά συσχετισμένες ανεξίξεις παίρνει τιμές στο διάστημα $0,5 - 1$. Η τιμή $H = 0,5$ δηλώνει ανεξαρτησία, ενώ τιμές $H < 0,5$, δηλώνουν αντι-εμμοή. Η ισχύς της σχέσης μπορεί να διαπιστωθεί με τη βοήθεια ενός διπλού λογαριθμικού διαγράμματος της δειγματικής τυπικής απόκλισης $s^{(k)}$ συναρτήσει της κλίμακας k , η κλίση του οποίου αντιστοιχεί στην τιμή $H - 1$.

3.3 Επίδραση στην εκτίμηση στατιστικών παραμέτρων

Μια από τις πιο σημαντικές επιπτώσεις της συμπεριφοράς εμμοής αφορά την τυπική στατιστική εκτίμηση. Η δυναμική Hurst-Kolmogorov επιφέρει δραματική αύξηση των διαστημάτων εμπιστοσύνης στην εκτίμηση στατιστικών παραμέτρων θέσης όπως η μέση τιμή και υψηλή αρνητική μεροληψία στην εκτίμηση παραμέτρων διασποράς όπως η τυπική απόκλιση. Είναι δυνατή όμως η εκτίμηση της μεροληψίας και της αβεβαιότητας των παραμέτρων μέσα από το πρίσμα της δυναμικής Hurst-Kolmogorov (Koutsoyiannis 2003; Koutsoyiannis and Montanari 2007; Koutsoyiannis 2011a).

- Εκτίμηση μέσης τιμής

Η μέση τιμή $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ είναι μια αμερόληπτη εκτιμήτρια ανεξάρτητα από το είδος της ανέλιξης, δηλαδή $E[\bar{X}] = \mu$. Η διασπορά της στην κλασική στατιστική είναι:

$$Var[\bar{X}] = \frac{\sigma^2}{n} \quad (3.6)$$

Η (3.6) όμως δεν ισχύει υπό την υπόθεση ύπαρξης εμμοής αλλά ισχύει αντίστοιχα (J. Beran 1994) :

$$Var[\bar{X}] = \frac{\sigma^2}{n^{2-2H}} \quad (3.7)$$

Σημειώνεται ότι η τυπική απόκλιση, η ρίζα δηλαδή της διασποράς είναι το τυπικό σφάλμα στην εκτίμηση της αληθινής μέσης τιμής από μια ιστορική χρονοσειρά και αποτελεί ένα βολικό δείκτη για την αβεβαιότητα στην εκτίμηση. Για $H= 0,5$ οι σχέσεις (3.6) και (3.7) δίνουν το ίδιο τυπικό σφάλμα που είναι αντιστρόφως ανάλογο της τετραγωνικής ρίζας του μήκους της χρονοσειράς. Το ισοδύναμο μήκος μιας σειράς με εμμοή ώστε να ισούται το τυπικό της σφάλμα με αυτό που προκύπτει υπό την υπόθεση κλασικής στατιστικής υπολογίζεται (Koutsoyiannis and Montanari 2007): $n' = n^{2(1-H)}$. Όσο η τιμή του H μεγαλώνει τόσο αυξάνεται η διαφορά μεταξύ των δύο σχέσεων και τα αποτελέσματα της επίδρασης. Έτσι, παρότι η δειγματική μέση τιμή είναι μια αμερόληπτη εκτιμήτρια της μέσης τιμής και για τις δύο περιπτώσεις, η διασπορά της, που εκφράζει την αβεβαιότητα στην εκτίμησή της είναι πολύ υψηλότερη στην περίπτωση ύπαρξης εμμοής. Επίσης, καθώς αυξάνεται το τυπικό σφάλμα σε όρους τυπικής απόκλισης το αντίστοιχο ισοδύναμο «απαιτούμενο» μήκος χρονοσειράς για να εξισωθεί το σφάλμα με αυτό της κλασικής στατιστικής αυξάνεται εκθετικά.

- Εκτίμηση τυπικής απόκλισης

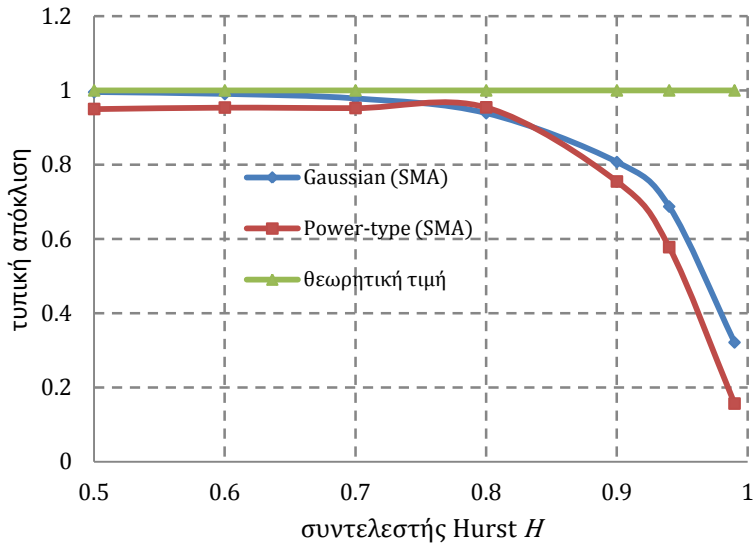
Η τυπική εκτιμήτρια της τυπικής απόκλισης :

$$S = \frac{1}{n - n^{2H-1}} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3.8)$$

είναι σχεδόν αμερόληπτη στην κλασική στατιστική, όμως υπό την επίδραση της συμπεριφοράς Hurst-Kolmogorov εμφανίζει αρνητική μεροληψία που μπορεί να εκτιμηθεί προσεγγιστικά σε $-\frac{1}{2(n^{2-2H})}$. Στο σχήμα 3.1 φαίνεται το αποτέλεσμα μιας προσομοίωσης κατά την οποία παρήχθησαν 10.000 συνθετικές χρονοσειρές μήκους δείγματος 100 χρόνων που διατηρούσαν το συντελεστή Hurst με χρήση του μοντέλου SMA (Koutsoyiannis 2002) για δύο διαφορετικές κατανομές, μία γκαουσιανή (κανονική) και μία δυναμικού τύπου, προερχόμενη από τον εξής μετασχηματισμό της κανονικής :

$$g(z) = \beta \left(\frac{\exp(z^2)^{\frac{\gamma_1 \gamma_2}{2}} - 1}{\gamma_2} \right)^{\frac{1}{\gamma_1}}$$

Παρατηρούμε ότι όσο αυξάνεται η τιμή του συντελεστή τόσο μεγαλώνει η αρνητική μεροληψία στην εκτίμηση της τυπικής απόκλισης. Ειδικά για τιμές $H > 0,8$ φαίνεται ότι η κλασική εκτιμήτρια υποεκτιμά σημαντικά την τυπική απόκλιση η οποία για πολύ υψηλές τιμές του H (π.χ. $H = 0,97$) εμφανίζεται πλασματικά έως και 5 φορές μικρότερη.



Σχήμα 3.1 Διάγραμμα της τυπικής απόκλισης σε συνάρτηση με την τιμή του συντελεστή Hurst για δύο μοντέλα SMA και σε σύγκριση με τη θεωρητική τιμή.

Η διασπορά της τυπικής απόκλισης στην κλασική στατιστική για κανονική κατανομή της X_i υπολογίζεται:

$$Var[S] \approx \frac{\sigma^2}{2(n-1)} \quad (3.9)$$

ενώ στην περίπτωση ύπαρξης δομής HK βρέθηκε (Koutsoyiannis 2003):

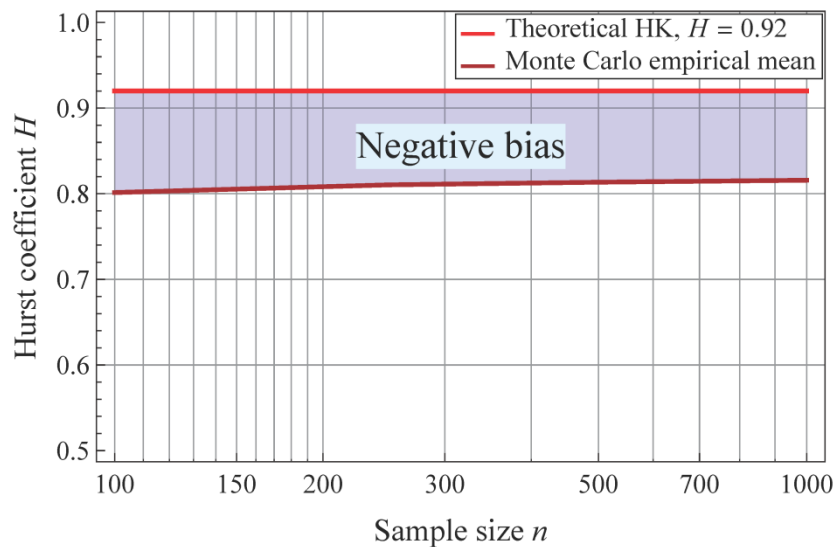
$$Var[\tilde{S}] \approx \frac{(0,1n + 0,8)^{\lambda(H)} \sigma^2}{2(n-1)} \quad \text{όπου } \lambda(H) := 0,088(4H^2 - 1)^2 \quad (3.10)$$

Και οι δύο σχέσεις αναφέρονται σε κανονική κατανομή της X_i . Μπορεί να δειχθεί ότι η διασπορά που προκύπτει από τη σχέση (3.10) είναι μεγαλύτερη από την αντίστοιχη της σχέσης (3.9) εκτός από την περίπτωση $H=0,5$ όπου συμπίπτουν.

Η χρησιμοποίηση της σχέσης (3.9) οδηγεί σε υποεκτίμηση της διασποράς, μικρή για τις μικρές χρονικές κλίμακες αλλά συνεχώς αυξανόμενη καθώς μεγαλώνει η τιμή της κλίμακας. Αυτό οδηγεί και σε υποεκτίμηση του φαινομένου Hurst σε μικρά δείγματα. Συγκεκριμένα, σε ένα λογαριθμικό διάγραμμα της τυπικής απόκλισης συναρτήσει της κλίμακας η κλίση της καμπύλης δεν είναι σταθερή αλλά μειώνεται με την αύξηση της κλίμακας k .

Έτσι, όχι μόνο υποεκτιμάται ο συντελεστής Hurst αλλά μπορεί να οδηγήσει και στο λανθασμένο συμπέρασμα ότι σε μεγάλες κλίμακες ο συντελεστής πλησιάζει την τιμή 0,5 που υποδηλώνει απουσία του φαινομένου (Koutsoyiannis 2003).

Στο σχήμα 3.2 φαίνεται πως για υψηλή τιμή του συντελεστή Hurst ($H = 0,92$) η μέθοδος του απλού κλιμακογράμματος που βασίζεται στη κλασική εκτίμηση της τυπικής απόκλισης οδηγεί σε σημαντική αρνητική μεροληψία στην εκτίμηση ακόμα και για σχετικά μεγάλες τιμές του μήκους δείγματος.



Σχήμα 3.2 Διάγραμμα της εμπειρικής τιμής του συντελεστή Hurst συναρτήσει του μήκους δείγματος για θεωρητική τιμή του $H = 0,92$ (Papalexίου, Koutsoyiannis, and Montanari 2010)

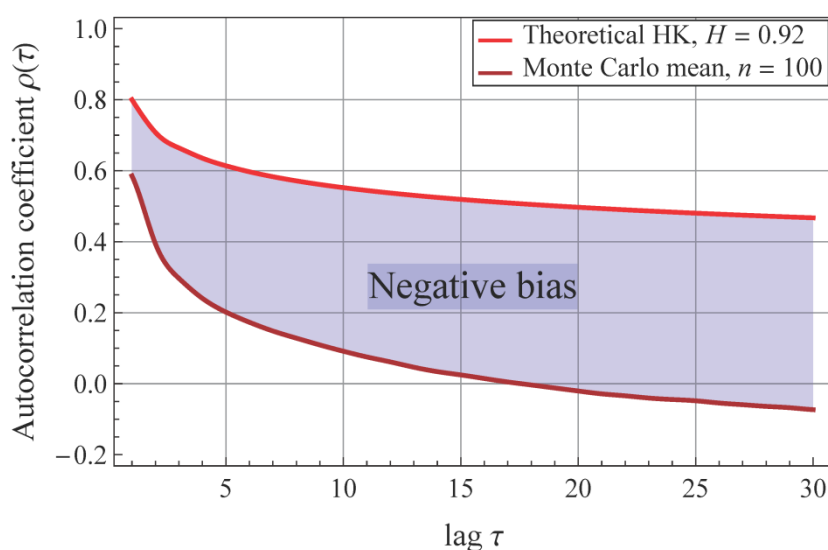
- Εκτίμηση της αυτοσυσχέτισης

Η τυπική εκτιμήτρια της αυτοσυσχέτισης ρ_1 με βάση την κλασική στατιστική είναι:

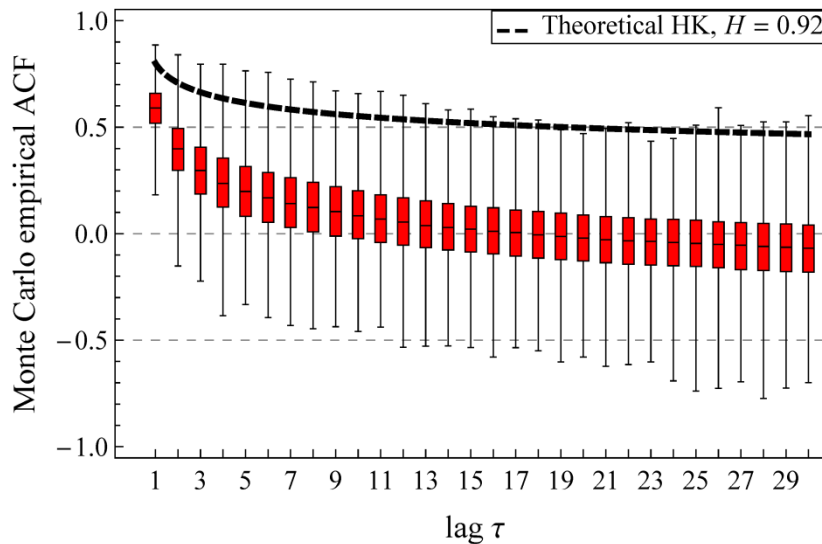
$$R_l := \frac{1}{(n-1)S^2} \sum_{i=1}^{n-1} (X_i - \bar{X})(X_{i+1} - \bar{X}) \quad (3.11)$$

που ενώ είναι σχεδόν αμερόληπτη υπό την υπόθεση κλασικής στατιστικής προσδίδει αρνητική μεροληψία περίπου ίση με $-\frac{1/\rho_l-1}{n^2-2H-1}$ (-79%) στην περίπτωση ύπαρξης εμμονής.

Στο σχήμα 3.3 φαίνεται η αυξανόμενη με την υστέρηση αρνητική μεροληψία στην εκτίμηση της αυτοσυσχέτισης για υψηλή τιμή του συντελεστή Hurst $H = 0,92$. Στο σχήμα 3.4 φαίνεται επίσης και η μεγάλη αβεβαιότητα στην εκτίμηση αφού για τιμές της υστέρησης μεγαλύτερες του 3 η θεωρητική τιμή είναι εκτός του 80% του διαστήματος εμπιστοσύνης της εμπειρικής εκτίμησης. Είναι αξιοσημείωτο δε ότι ενώ οι θεωρητικές τιμές της αυτοσυσχέτισης ακόμα και για υστέρηση 25 είναι μεγαλύτερες ή ίσες της τιμής 0,5, για υστέρηση μεγαλύτερη του 13 το 50% των εμπειρικών εκτιμήσεων είναι μικρότερο του 0, κάτι που δηλώνει παντελή έλλειψη συσχέτισης.



Σχήμα 3.3 Διάγραμμα της θεωρητική τιμής της αυτοσυσχέτισης συναρτήσει της υστέρησης για $H = 0,92$ σε σύγκριση με την εμπειρική της εκτίμησης ως μέση τιμή Monte Carlo προσομοιώσεων FGN. (Papalexiou, Koutsoyiannis, and Montanari 2010)



Σχήμα 3.4 Box-plots της εμπειρικής αυτοσυσχέτισης μέσω Monte Carlo προσομοιώσεων FGN για διάφορες τιμές της υστέρησης και σύγκριση με τη θεωρητικές τιμές μοντέλου FGN με $H = 0,92$. (Papalexίου, Koutsoyiannis, and Montanari 2010)

- *Επίδραση στην «πρόβλεψη»*

Είναι γνωστό ότι στην κλίμακα 1 η πρόβλεψη της τιμής μια ανέλιξης για ένα χρονικό βήμα μπροστά είναι περισσότερο δύσκολη στην περίπτωση μιας τελείως τυχαίας διαδικασίας. Η εξάρτηση βελτιώνει λοιπόν την προβλεψιμότητα για ένα χρονικό βήμα μπροστά. Παρόλα αυτά σύμφωνα με τον Koutsoyiannis (2010) η ύπαρξη μακροπρόθεσμης εμμονής δυσχεραίνει σημαντικά την πρόβλεψη για μεγάλες χρονικές κλίμακες, ενώ αντίθετα η αντι-εμμονή την βελτιώνει. Τέτοιου είδους προβλέψεις είναι οι κλιματικού τύπου, όπου μελετάται συνήθως ο μέσος όρος 30 χρόνων.

3.4 Μοντέλα που διατηρούν την εμμονή

Το βασικό μοντέλο διατήρησης της εμμονής είναι οι ανελιξεις απλής ομοιοθεσίας (ή στάσιμα αυξήματα αυτό-όμοιων ανελιξεων), όπως ο κλασματικός γκαουσιανός θόρυβος.

Οι ανελίξεις αυτές εισήχθησαν για πρώτη φορά σε θεωρητική βάση από τον Kolmogorov(1941). Οι στατιστικοί αγνοούσαν τη στατιστική σημασία τους πριν ο Mandelbrot και Wallis τις εισάγουν στα κείμενά τους (1968).

Κλασματική γκαουσιανή κίνηση και θόρυβος

Έστω ότι η X_t είναι μια γκαουσιανή ακολουθία οπότε η κατανομή της καθορίζεται πλήρως από τη μέση τιμή και τη διασπορά της. Άρα για κάθε τιμή του H στο $(0,1)$ υπάρχει μία ακριβώς γκαουσιανή ακολουθία X_i η οποία είναι η στάσιμη ανέλιξη της ομοιόθετης διαδικασίας Y_t . Τότε η X_i ονομάζεται κλασματικός γκαουσιανός θόρυβος και η Y_t κλασματικός θόρυβος του Brown (fractional brownian motion) και συμβολίζεται με $B_H(t)$.

Ο κλασματικός γκαουσιανός θόρυβος που αναπαράγει την ομοιοθεσία περιγράφεται μαθηματικά ως εξής :

$$Z_i^{(k)} - \mu =_d \left(\frac{k}{l}\right)^{H-1} (Z_j^{(l)} - \mu) \quad (3.12)$$

Στην παραπάνω σχέση το σύμβολο $=_d$ δείχνει την ισότητα στην από κοινού κατανομή των δύο μελών. Ο συντελεστής H είναι μια θετική σταθερά ($0 < H < 1$) γνωστή ως συντελεστής του Hurst. Η σχέση (3.2) ισχύει για κάθε ακέραιο i και j (στάσιμη ανέλιξη) και για οποιαδήποτε χρονική κλίμακα k και l . Θέτοντας $i = j = l = 1$ παίρνουμε :

$$\gamma_0^{(k)} := Var [Z_i^{(k)}] = k^{2H} \gamma_0 \quad (3.13)$$

Φαίνεται ότι η τυπική απόκλιση $\sigma^{(k)} := \left(\gamma_0^{(k)}\right)^{\frac{1}{2}}$ είναι εκθετική συνάρτηση της κλίμακας k με εκθέτη H . Σύμφωνα με τον Koutsoyiannis (2002) η συνάρτηση αυτοσυσχέτισης είναι ανεξάρτητη της κλίμακας k και δίνεται από τη σχέση:

$$\rho_j^{(k)} = \rho_j = \frac{1}{2} [(j+1)^{2H} + (j-1)^{2H}] - j^{2H}, \quad j > 0 \quad (3.14)$$

Η σχέση προσεγγίζεται πολύ ικανοποιητικά εκτός από την περίπτωση πολύ μικρών j ως εξής:

$$\rho_j^{(k)} = \rho_j = H(2H - 1)j^{2H-2}, \quad j > 0 \quad (3.15)$$

που δείχνει ότι η αυτοσυσχέτιση είναι εκθετική συνάρτηση της υστέρησης. Αντίστοιχα, για το φάσμα μπορεί ναδειχτεί ότι είναι εκθετική συνάρτηση της συχνότητας ω :

$$\frac{s_\gamma^{(k)}(\omega)}{\gamma_0^{(k)}} \approx 4(1 - H)(2\omega)^{1-2H} \quad (3.16)$$

Παρατηρούμε ότι όπως οι ανελίξεις AR(1) χρησιμοποιούν μια μοναδική παράμετρο ρ για να εκφράσουν τη δομή της αυτοσυσχέτισης, το μοντέλο FGN χρησιμοποιεί αντίστοιχα το συντελεστή αυτοσυσχέτισης H . Επίσης με μια απλή σύγκριση των βασικών στατιστικών παραμέτρων των δύο μοντέλων διαπιστώνει κανείς εύκολα ότι οι σχέσεις του μοντέλου FGN είναι αρκετά απλούστερες του απλού AR(1) μοντέλου.

3.5 Αλγόριθμος αναπαραγωγής Κλασματικού Γκαουσιανού Θορύβου

Ένας αρκετά απλός τρόπος αναπαραγωγής της απλής ομοιοθετικής ανέλιξης είναι μέσω της λογικής των τυχαίων διακυμάνσεων πολλαπλής κλίμακας. Ο Koutsoyiannis(2002) δείχνει ότι το άθροισμα τριών εκθετικών συναρτήσεων της χρονικής υστέρησης προσεγγίζει ικανοποιητικά την συνάρτηση αυτοσυσχέτισης μιας απλής ομοιοθετικής ανέλιξης στην ετήσια χρονική κλίμακα. Συγκεκριμένα, η ανέλιξη παράγεται ως το άθροισμα τριών ανελίξεων AR(1) από τη σχέση:

$$X_i = A_i + B_i + C_i \quad (3.17)$$

Οι συντελεστές αυτοσυσχέτισης των τριών ανελιξεων για υστέρηση 1 οι οποίοι δίνουν την καλύτερη προσέγγιση (σε όρους μέσου τετραγωνικού σφάλματος) της εξίσωσης (3.17) προκύπτουν με αριθμητική αναζήτηση:

$$\rho = 1,52 (H - 0,5)^{1,32} \quad (3.18)$$

$$\varphi = 0,953 - 7,69(1 - H)^{3,85} \quad (3.19)$$

$$\xi = \begin{cases} 0,932 + 0,087 H, & H \leq 0,76 \\ 0,993 + 0,007 H, & H > 0,76 \end{cases} \quad (3.20)$$

Η διασπορά των ανελιξεων είναι: $(1 - c_1 - c_2)\gamma_0$, $c_1 \gamma_0$, $c_2 \gamma_0$ αντίστοιχα. Οι συντελεστές c_1 και c_2 υπολογίζονται έτσι ώστε η συνάρτηση αυτοσυσχέτισης του αθροίσματος των τριών ανελιξεων να ταυτίζεται με την θεωρητική αυτοσυσχέτιση της ανέλιξης απλής ομοιοθεσίας, για υστέρηση 1 και 100. Η συνάρτηση αυτοσυσχέτισης του αθροίσματος είναι:

$$\rho_j = (1 - c_1 - c_2)\rho + c_1\varphi^j + c_2\xi^j \quad (3.21)$$

3.6 Φυσική ερμηνεία του φαινομένου

Διάφορες ερμηνείες έχουν δοθεί για την φυσική θεμελίωση του φαινομένου επικεντρωμένες κυρίως στην έννοια ενός φυσικού συστήματος που «θυμάται» και δρα συντονισμένα ή και ενίοτε ντετερμινιστικά. Η συμπεριφορά της εμμονής έχει αποδοθεί λοιπόν είτε στη συντονισμένη δράση διαφορετικών στοιχείων ενός κοινού συστήματος, είτε σε έναν φυσικό μηχανισμό «μνήμης» είτε τέλος, έχει ταυτιστεί με την ύπαρξη ντετερμινιστικών τάσεων. Μια απλούστερη ερμηνεία προτείνεται από τον Koutsoyiannis (2002) και βασίζεται αντίθετα στην έννοια της «απουσίας μνήμης». Μέσα στο πλαίσιο ενός κλίματος που αλλάζει κατά τρόπο ακανόνιστο σε όλες τις χρονικές κλίμακες, το φαινόμενο αποτελεί το αποτέλεσμα τυχαίων διακυμάνσεων σε διαφορετικές χρονικές

κλίμακες. Μπορεί να δειχθεί ότι η σύνθεση δύο ανεξίτηων Markov οι οποίες χαρακτηρίζονται από τυχαίες διακυμάνσεις της μέσης τιμής τους σε μικρότερη και μεγαλύτερη χρονική κλίμακα αντίστοιχα, οδηγεί στην προσέγγιση ενός μοντέλου FGN. Είναι προφανές άλλωστε ότι αυτή η ιδέα είναι ο αλγόριθμος υλοποίησης του Κλασματικού Γκαουσιανού Θορύβου που διατηρεί την εμμονή. Μία υπέρθεση τριών ανεξίτηων AR(1) με τα επιθυμητά χαρακτηριστικά πλησιάζει πολύ ικανοποιητικά μια ανέλιξη FGN.

Η παραπάνω εξήγηση διαφοροποιείται από την έννοια της μη-στασιμότητας. Η ερμηνεία με βάση μια «μη στάσιμη συμπεριφορά» της μέσης τιμής προϋποθέτει την ύπαρξη μιας ντετερμινιστικής συνάρτησης που θα ορίζει τη μέση τιμή σαν συνάρτηση του χρόνου. Παρόλο που η ύπαρξη τάσεων συναντάται στην υδρολογία θεωρείται ότι ακολουθούν ένα ακανόνιστο τυχαίο μοτίβο και όχι ένα ντετερμινιστικά ορισμένο. Η έννοια άλλωστε περιττεύει στην εξήγηση αφού δείχνεται ότι μια στάσιμη ανέλιξη μπορεί να προσεγγίσει το φαινόμενο.

Τέλος, ο Koutsoyiannis (2011b) έδειξε ότι η εφαρμογή της αρχής της μέγιστης εντροπίας σε ένα στοχαστικό σύστημα μπορεί να καταλήξει στη συμπεριφορά μακροπρόθεσμης εμμονής.

4. Βιβλιογραφική επισκόπηση μεθόδων εκτίμησης της εμμοής

Παρουσιάζονται οι βασικές μέθοδοι εκτίμησης της εμμοής όπως εντοπίζονται στη βιβλιογραφία, δύο από τις οποίες χρησιμοποιούνται τελικά στην εφαρμογή στα πραγματικά δεδομένα (Κεφάλαιο 5). Σκοπός του κεφαλαίου δεν είναι η πλήρης και αναλυτική βιβλιογραφική επισκόπηση όλων των μεθόδων αλλά κυρίως η συνοπτική περιγραφή των βασικότερων εξ αυτών ώστε να δοθεί μια γενική ιδέα για το είδος και τη φύση των μεθοδολογιών του προβλήματος. Για όσες γραφικές μεθόδους ήταν αναπτυγμένες στο ελεύθερο λογισμικό R (πακέτο fARMA) παρουσιάζεται και διάγραμμα εφαρμογής τους σε μια συνθετική χρονοσειρά FGN μήκους 10.000 τιμών και θεωρητικού συντελεστή $H = 0,7$. Στο τελευταίο υποκεφάλαιο περιγράφονται τα αποτελέσματα κάποιων συναφών μελετών ως προς το φαινόμενο της βροχόπτωσης με αναφορά στη μέθοδο που χρησιμοποιήθηκε κάθε φορά.

4.1 Το ζήτημα της ευρωστίας

Οι βασικές υποθέσεις των μεθόδων εκτίμησης της εμμοής είναι εξής (Beran 1994):

1. Η χρονοσειρά που μελετάται είναι στάσιμη
2. Η χρονοσειρά που μελετάται ακολουθεί κανονική κατανομή ή τουλάχιστον αποτελεί γραμμική ανέλιξη
3. Η φασματική πυκνότητα ανήκει σε μια γνωστή παραμετρική τάξη πυκνοτήτων $f(x;\theta)$

Αυτές οι υποθέσεις είναι μαθηματικά συμφέρουσες. Η πρώτη μας επιτρέπει να φτιάξουμε ένα παρσιμόνιο μοντέλο και είναι προϋπόθεση για την ισχύ συγκεκριμένων μαθηματικών ιδιοτήτων (π.χ. εργοδικότητα). Η μέση τιμή είναι συνεχής και οι συνδιασπορές εξαρτώνται από την υστέρηση μεταξύ δύο διαδοχικών τιμών μόνο.

Η υπόθεση της κανονικότητας σημαίνει ότι η χρονοσειρά καθορίζεται απόλυτα από τις δύο πρώτες ροπές. Η συνάρτηση πιθανοφάνειας είναι συνάρτηση μόνο της μέσης τιμής και των συνδιασπορών. Η τελευταία υπόθεση περιορίζει την εκτίμηση της συνάρτησης φασματικής πυκνότητας στην εκτίμηση ενός διανύσματος πεπερασμένων διαστάσεων θ .

Η μαθηματική απλότητα είναι χρήσιμη όχι μόνο για θεωρητικούς λόγους (αρχή της παρσιμονίας) αλλά και γιατί ευνοεί μια ενστικτώδη ερμηνεία των αποτελεσμάτων της ανάλυσης χρονοσειρών. Παρόλα αυτά είναι απαραίτητο να ξέρουμε σε ποιο βαθμό αυτές οι υποθέσεις ισχύουν ή όχι. Τα βασικά ερωτήματα που τίθενται είναι τα εξής :

- Ισχύουν οι παραπάνω υποθέσεις ή τουλάχιστον μια καλή προσέγγιση τους; Αν όχι τι είδους αποκλίσεις παρατηρούνται;
- Πόσο αξιόπιστα είναι τα αποτελέσματα των μεθόδων όταν παραβιάζονται οι υποθέσεις;
- Υπάρχουν λιγότερο ευαίσθητες μέθοδοι σε αποκλίσεις από τις αρχικές υποθέσεις που να είναι όμως αρκετά αξιόπιστες όταν οι υποθέσεις τηρούνται;
- Τι επίδραση έχει στο μοντέλο η ύπαρξη ιδιαίτερα ακραίων τιμών;

Σύμφωνα με το Beran(1992) υπάρχουν παραπάνω από έναν τρόποι με τους οποίους μπορεί να αποκλίνει η πραγματικότητα από τις υποθέσεις του μοντέλου και δεν είναι δυνατό να προφυλαχθεί κανείς από όλες τις πιθανές αποκλίσεις. Στη βιβλιογραφία υπάρχουν εύρωστα μοντέλα που έχουν αναπτυχθεί για συγκεκριμένες αποκλίσεις κανένα όμως που να ισχύει για όλες τις περιπτώσεις.

Είναι συνήθως δύσκολο να αποφασίσει κανείς με βεβαιότητα ποιο μοντέλο είναι το σωστό και το πιθανότερο είναι, ωστόσο ότι κανένα μοντέλο δεν είναι απόλυτα κατάλληλο. Επομένως, ο στόχος της επεξεργασίας δεδομένων είναι κυρίως να βρει κανείς το μοντέλο που εξυπηρετεί τον επιδιωκόμενο στόχο καλύτερα παρά το απόλυτα σωστό μοντέλο.

Κριτήρια προς αυτή την κατεύθυνση είναι μερικά από τα εξής:

- Η ικανοποιητική προσαρμογή στα δεδομένα
- Η αξιόπιστη πρόβλεψη μελλοντικών παρατηρήσεων
- Επαρκώς ακριβείς εκτιμήτριες παραμέτρων
- Η απλότητα
- Η δυνατότητα φυσικής εξήγησης
- Η αρχή της parsimonίας
- Κλπ..

Γενικά, η επίτευξη ικανοποιητικής βεβαιότητας είναι ένα δύσκολο εγχείρημα που όμως διευκολύνεται με την ύπαρξη αρκετών παρατηρήσεων. Πάντως ένα από τα κυριότερα επιχειρήματα υπέρ της χρήσης μοντέλων μακροπρόθεσμης εμμονής είναι η αρχή της parsimonίας. Για πεπερασμένα μεγέθη δειγμάτων η προσθήκη μιας μόνο παραμέτρου H επιτρέπει την προσομοίωση ενός φαινομένου που θα απαιτούσε την ύπαρξη πολλών παραμέτρων εάν επρόκειτο να προσομοιωθεί με μοντέλα βραχυπρόθεσμης εξάρτησης.

4.2 Η μέθοδος R/S

Για μια χρονοσειρά X_i ($i = 1, 2, \dots$) με μερικό άθροισμα $Y(n) = \sum_{i=1}^n X_i$ και δειγματική διασπορά $S^2(n) := 1/n \sum_{i=1}^n X_i^2 - (1/n)^2 Y(n)^2$ η R/S μέθοδος ή μέθοδος εύρους σε προσαρμοσμένη κλίμακα (rescaled adjusted range) ορίζεται ως εξής (Murad S. Taqqu, Teverovsky, and Willinger 1995) :

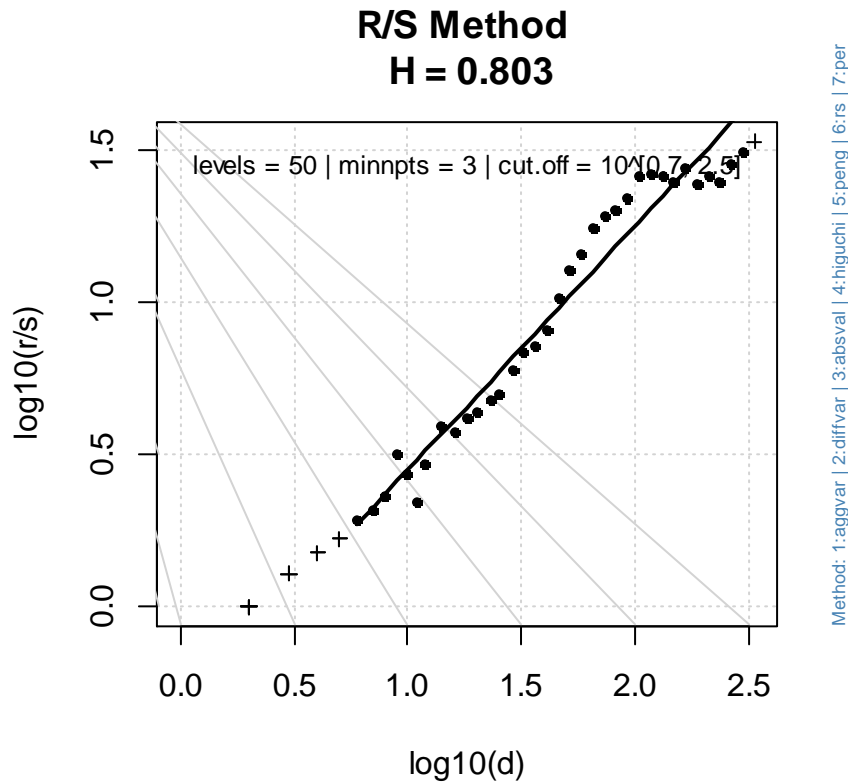
$$\frac{R}{S}(n) := \frac{1}{S(n)} \left[\max_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) - \min_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) \right] \quad (4.1)$$

Για Κλασματικό Γκαουσιανό Θόρυβο ισχύει:

$E[R/S(n)] \sim C_H n^H$ καθώς το $n \rightarrow \infty$, όπου C_H θετική πεπερασμένη σταθερά ανεξάρτητη του n .

Για να καθοριστεί το H χρησιμοποιείται η ακόλουθη διαδικασία. Μια σειρά μήκους N διαιρείται σε K τμήματα μεγέθους N/K το καθένα. Ύστερα, για κάθε υστέρηση n υπολογίζεται η ποσότητα $R(k_i, n)/S(k_i, n)$ αρχίζοντας από σημεία $k_i = \frac{iN}{K} + 1, i = 1, 2, \dots$ έτσι ώστε $k_i + n \leq N$. Για τιμές του n μικρότερες από N/K προκύπτουν K διαφορετικές εκτιμήσεις του $R(n)/S(n)$. Για τιμές του n που πλησιάζουν N προκύπτουν λιγότερες τιμές έως και 1 όταν $n \geq N - N/K$.

Στη συνέχεια δημιουργείται το λογαριθμικό διάγραμμα της $R(k_i, n) / S(k_i, n)$ σε συνάρτηση με το n . Η παράμετρος H εκτιμάται προσαρμόζοντας μια γραμμή στα σημεία του διαγράμματος. Καθώς κάθε βραχύχρονη εξάρτηση συνήθως προκαλεί μια μεταβατική ζώνη στο κάτω μέρος του τέλους του διαγράμματος, το μέρος αυτό καλό είναι να αποκλείεται από την εκτίμηση του H . Συνήθως, το πάνω μέρος του διαγράμματος δεν χρησιμοποιείται επίσης γιατί υπάρχουν πολύ λίγα σημεία σε εκείνο το τμήμα που δεν επαρκούν για αξιόπιστες εκτιμήσεις. Οι ενδιάμεσες τιμές του n χρησιμοποιούνται για την εκτίμηση του H .



Σχήμα 4.1 Διάγραμμα της μεθόδου R/S στο R για συνθετική χρονοσειρά FGN με $H = 0,7$

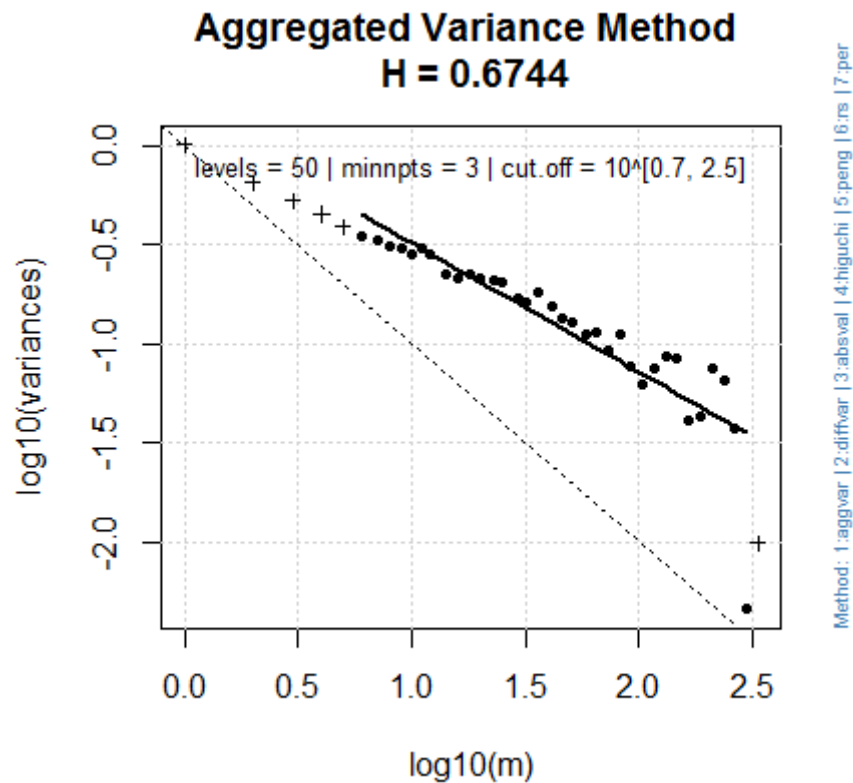
4.3 Μέθοδος της συναθροισμένης διασποράς

Η αρχική χρονοσειρά X_i ($i = 1, 2, \dots$) χωρίζεται σε τμήματα μήκους m από κάθε ένα από τα οποία προκύπτει ένας μέσος όρος :

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)*m+1}^{km} X(i) \quad (4.2)$$

όπου k ο αύξων αριθμός του τμήματος, $k = 1, 2, \dots$. Στη συνέχεια υπολογίζεται η δειγματική διασπορά των στοιχείων που έχουν προκύψει σε κάθε κλίμακα m . Η διαδικασία επαναλαμβάνεται για διαφορετικές τιμές της κλίμακας m και δημιουργείται λογαριθμικό γράφημα με τιμές στον άξονα των y τους λογαρίθμους της δειγματικής διασποράς και στον άξονα των x τους λογαρίθμους των τιμών αντίστοιχων τιμών της κλίμακας m . Ο συντελεστής Hurst προκύπτει από την κλίση του διαγράμματος προσαυξημένη κατά 1.

Η μέθοδος στηρίζεται στον ορισμό της μακροπρόθεσμης εμμονής όπως περιγράφηκε στο εδάφιο 3.2.2. Η κλίση είναι πιο απότομη για περιπτώσεις βραχυπρόθεσμης εξάρτησης ή ανεξάρτητων δεδομένων και αναμένεται να παίρνει τιμές γύρω στο -0,5.



Σχήμα 4.2 Διάγραμμα της μεθόδου συναθροισμένης διασποράς στο R για συνθετική χρονοσειρά FGN με $H = 0,7$

4.4 Μέθοδος ελαχίστων τετραγώνων βασισμένη στην τυπική απόκλιση (Least Squares based on Standard Deviation, LSSD)

Μετά από συστηματική μελέτη μέσω ανάλυσης Monte Carlo ο Koutsoyiannis (2003) βρήκε μια εκτιμήτρια S της τυπικής απόκλισης σ , περίπου αμερόληπτη για γνωστό συντελεστή Hurst και για ανέλιξη X_i που ακολουθεί την κανονική κατανομή η οποία είναι:

$$\tilde{S} := \sqrt{\frac{n - \frac{1}{2}}{n - n^{(2H-1)}}} S = \sqrt{\frac{n - \frac{1}{2}}{(n-1)(n - n^{(2H-1)})}} \sqrt{\sum_{i=1}^n (X_i - X_i^{(n)})^2} \quad (4.3)$$

$$\tilde{S}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - X_i^{(n)})^2 \quad (4.4)$$

όπου $X_i^{(n)}$ η μέση τιμή του δείγματος της συναθροισμένης ανέλιξης (εξίσωση). Ο αλγόριθμος στηρίζεται στις κλασικές δειγματικές εκτιμήτριες $s^{(k)}$ της τυπικής απόκλισης $\sigma^{(k)}$ για τιμές της κλίμακας k από 1 έως $k' := n/10$. Η μέγιστη τιμή της κλίμακας k' επιλέχθηκε έτσι ώστε η $s^{(k)}$ να εκτιμάται από τουλάχιστον 10 τιμές.

Συνδυάζοντας τις σχέσεις (4.3) και (4.4) και υποθέτοντας ότι $E[\tilde{S}] = \sigma$, παίρνουμε :

$$s^{(k)} \approx c_k(H) k^H \sigma \quad \text{όπου } c_k(H) := \sqrt{\frac{n/k - (n/k)^{2H-1}}{n/k - 1/2}} \quad (4.5)$$

Αν αγνοήσουμε στην εξίσωση (4.5) τον όρο $c_k(H)$, είναι δυνατή η εκτίμηση του H και της τυπικής απόκλισης σ κάνοντας γραμμική παλινδρόμηση μεταξύ του $\ln s^{(k)}$ και του $\ln k$. Αυτή η παλινδρόμηση είναι ένας γνωστός αλγόριθμος (Koutsoyiannis 2002), που εισάγει όμως αρνητική μεροληψία στην εκτίμηση του H και της τυπικής απόκλισης σ όταν οι τιμές του H είναι υψηλές. Προκειμένου να μειωθεί η μεροληψία χρησιμοποιείται η εκτίμηση του Hurst από την παλινδρόμηση για τον προσδιορισμό του $c_k(H)$ και στη συνέχεια γίνεται νέα παλινδρόμηση του μεταξύ του $\ln s^{(k)}$ και του $\ln k$. Η διαδικασία αυτή συνεχίζεται μέχρι να προκύψει σύγκλιση.

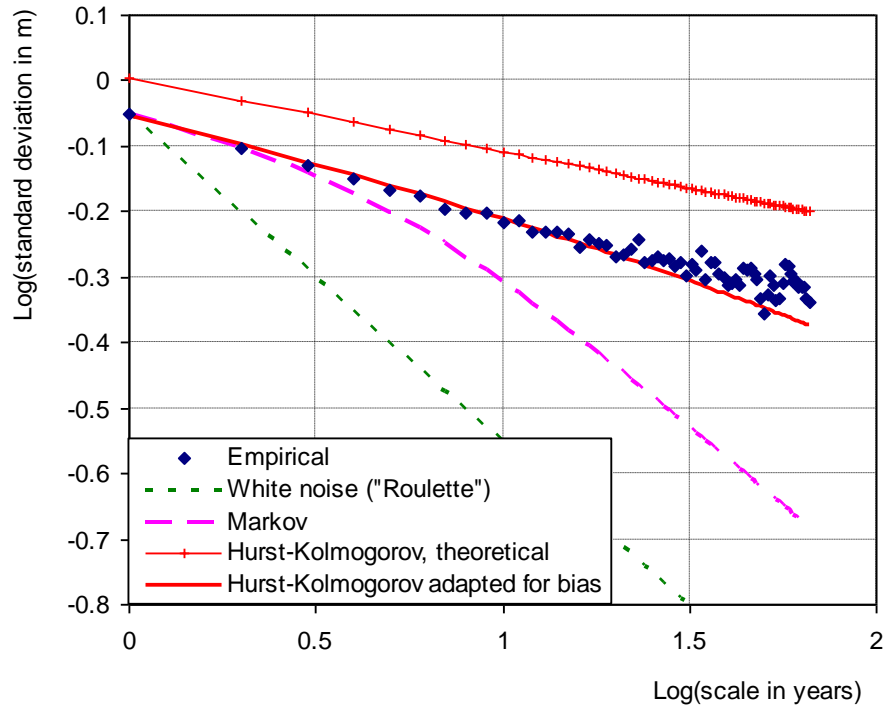
Μια πιο συστηματική προσέγγιση για την εκτίμηση των δύο παραμέτρων είναι ο αλγόριθμος ελαχιστοποίησης του σφάλματος $e^2(\sigma, H)$. Πρόκειται για τη

μέθοδο ελαχίστων τετραγώνων βασισμένη στην τυπική απόκλιση (LSSD) που αναπτύχθηκε από τους Tyralis and Koutsoyiannis (2011):

$$\begin{aligned}
 e^2(\sigma, H) &:= \sum_{K=1}^{K'} \frac{[\ln E[S^{(k)}] - \ln s^{(k)}]^2}{k^p} + \frac{H^{q+1}}{q+1} & (4.6) \\
 &= \sum_{K=1}^{K'} \frac{[\ln \sigma + H \ln \kappa + \ln c_k(H) - \ln s^{(K)}]^2}{k^p} + \frac{H^{q+1}}{q+1}
 \end{aligned}$$

Όπου $1/k^p$ ο συντελεστής βάρους στο σφάλμα κάθε κλίμακας. Για τιμή του p ίση με 0 οι συντελεστές βάρους είναι ίσοι για όλες τις κλίμακες ενώ για $p = 1, 2, \dots$ μειούμενοι συντελεστές βάρους προσαρμόζονται σε αυξανόμενες κλίμακες καθώς το μικρότερο μέγεθος δείγματος στις μεγάλες κλίμακες αυξάνει την αβεβαιότητα. Μετά από Monte Carlo πειράματα βρέθηκε ότι η τιμή του p ίση με 2 δίνει πιο αξιόπιστες εκτιμήσεις. Ο όρος $\frac{H^{q+1}}{q+1}$ είναι ένας όρος ποινής που για υψηλή τιμή του q (πχ. 50) απορρίπτει την τιμή $\hat{H} = 1$ εξαναγκάζοντάς την σε ελαφρώς χαμηλότερες τιμές. Έτσι, αποτρέπονται οι τιμές της τυπικής απόκλισης που τείνουν στο άπειρο.

Στο σχήμα 4.3 φαίνεται η πολύ καλύτερη προσαρμογή του τροποποιημένου κλιμακογράμματος στα εμπειρικά αποτελέσματα.



Σχήμα 4.3 Κλιμακόγραμμα της χρονοσειράς ελάχιστης στάθμης του ποταμού Νείλου

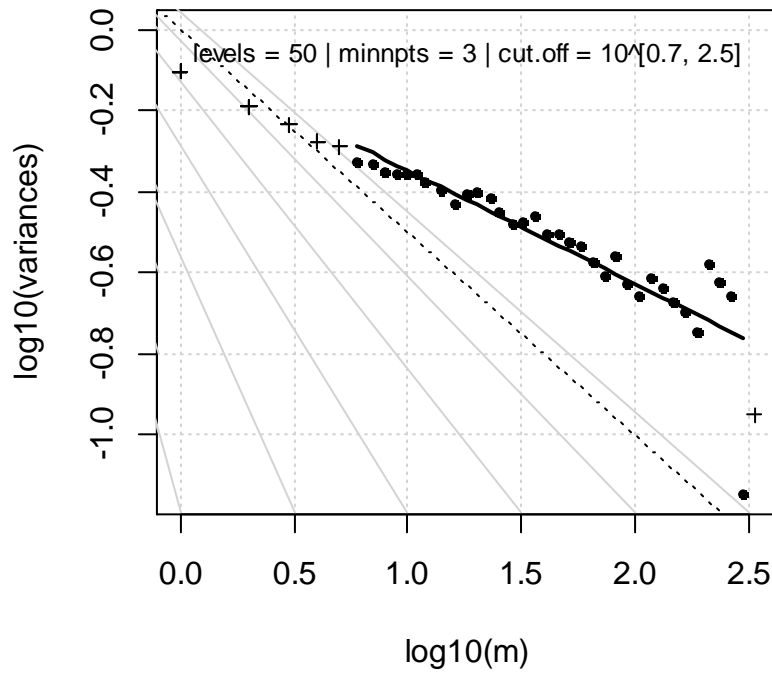
4.5 Μέθοδος των απολύτων τιμών της συναθροισμένης σειράς

Η μέθοδος είναι παρόμοια με εκείνη της συναθροισμένης ανέλιξης. Τα δεδομένα χωρίζονται με τον ίδιο τρόπο σύμφωνα με τη σχέση (4.2). Αντί για τον υπολογισμό της δειγματικής διασποράς υπολογίζεται το σύνολο των απολύτων τιμών της συναθροισμένης σειράς, δηλαδή

$$\frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k)| \quad (4.7)$$

Στη συνέχεια δημιουργείται το λογαριθμικό διάγραμμα της (4.7) συναρτήσει της κλίμακας m . Αν υπάρχει εξάρτηση σε μακρά κλίμακα τότε η κλίση της ευθείας αναμένεται να είναι $H - 1$.

Absolute Moment - No. 1 H = 0.7203



Σχήμα 4.4 Διάγραμμα της μεθόδου απολύτων τιμών στο R για συνθετική χρονοσειρά FGN με $H = 0,7$

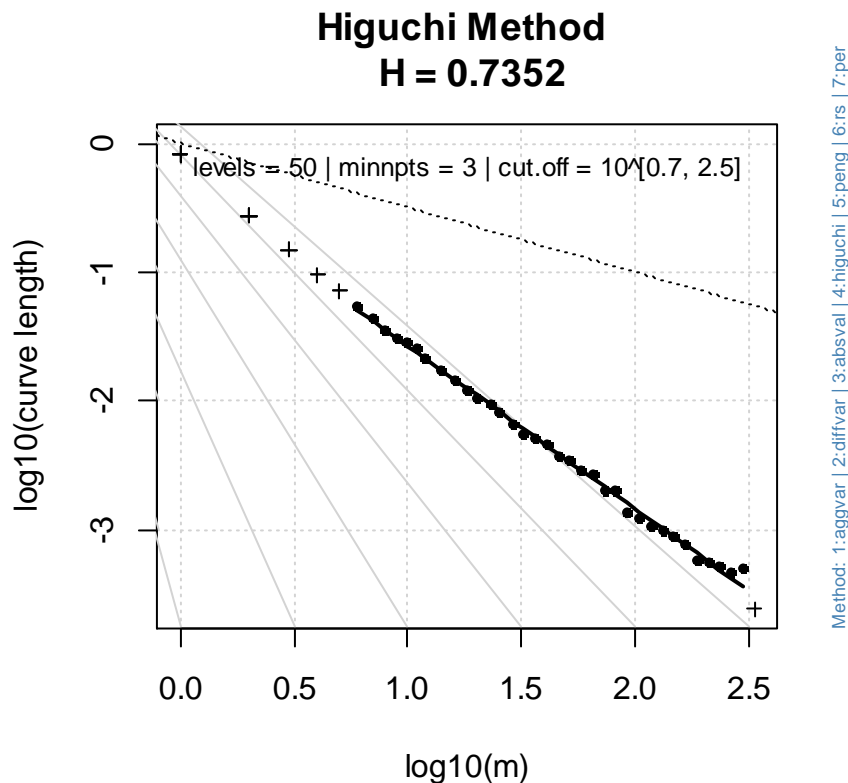
4.6 Μέθοδος του Higuchi

Η μέθοδος αυτή προτάθηκε από τον Higuchi(1988). Περιλαμβάνει τον υπολογισμό του μήκους μιας διαδρομής (path) και την εύρεση της κλασματικής της διάστασης D .

Υπολογίζονται πρώτα τα μερικά αθροίσματα $Y(n) = \sum_{i=1}^n X_i$ της αρχικής χρονοσειράς X_i ($i = 1, \dots, N$) παράγοντας για παράδειγμα κλασματική γκαουσιανή κίνηση από κλασματικό γκαουσιανό θόρυβο. Στη συνέχεια βρίσκεται το κανονικοποιημένο μήκος της καμπύλης:

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^m \left[\frac{N-i}{m} \right]^{-1} \sum_{k=1}^{\lfloor \frac{N-i}{m} \rfloor} |Y(i+km) - Y(i+(k-1)m)| \quad (4.8)$$

όπου N είναι το μήκος της χρονοσειράς, m είναι το μέγεθος του τμήματος (block size) και το σύμβολο $[]$ σημαίνει τη μεγαλύτερη ακέραια συνάρτηση. Τότε $EL(m) \sim c_H m^{-D}$, όπου $D = 2 - H$. Επομένως, ένα λογαριθμικό διάγραμμα του $L(m)$ συναρτήσει του m θα πρέπει να δώσει μια ευθεία γραμμή με κλίση $D = 2 - H$ (Murad S. Taqqu, Teverovsky, and Willinger 1995).

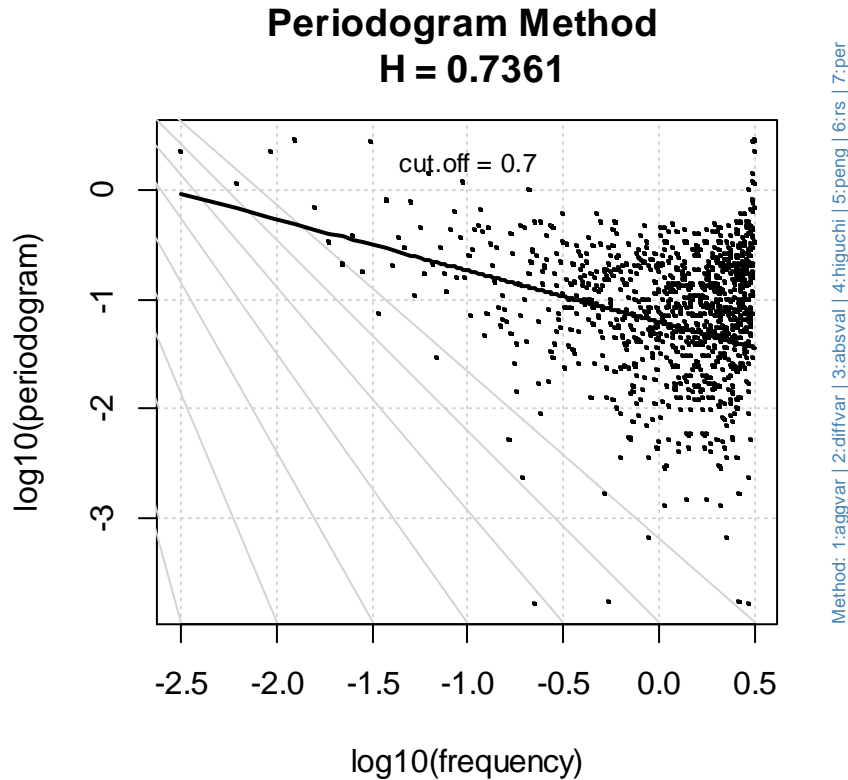


Σχήμα 4.5 Διάγραμμα της μεθόδου Higuchi στο R για συνθετική χρονοσειρά FGN με $H = 0,7$

4.7 Περιοδόγραμμα

Υπολογίζεται πρώτα η ποσότητα $I(\lambda) = \frac{1}{2\pi N} \left| \sum_{j=1}^N X_j e^{ij\lambda} \right|^2$ όπου λ η συχνότητα, N ο αριθμός των όρων στη σειρά και X_j τα δεδομένα. Επειδή $I(\lambda)$ είναι μια εκτιμήτρια της φασματικής πυκνότητας, μια σειρά με μακροπρόθεσμη εμμονή αναμένεται να έχει ένα περιοδόγραμμα που είναι ανάλογο του $|\lambda|^{1-2H}$ κοντά στην αρχή. Επομένως, μια γραμμική παλινδρόμηση του λογαρίθμου του

περιοδογράμματος στο λογάριθμο της συχνότητας λ αναμένεται να δώσει ένα συντελεστή $1-2H$ που αποτελεί μια προσέγγιση της παραμέτρου H .

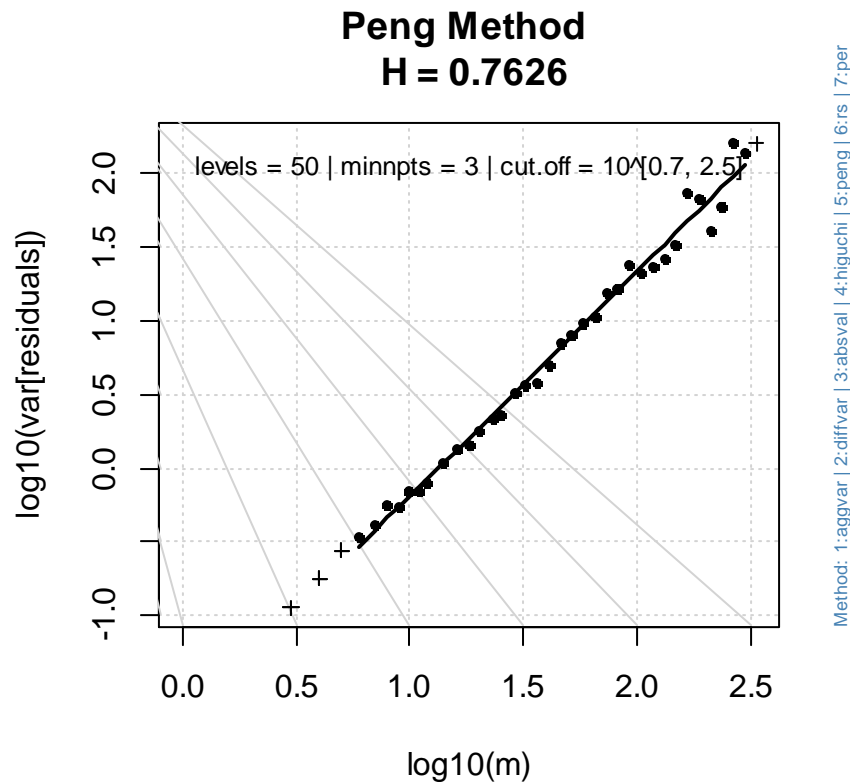


Σχήμα 4.6 Διάγραμμα της μεθόδου περιοδογράμματος στο R για συνθετική χρονοσειρά FGN με $H = 0,7$

4.8 Μέθοδος του Peng (ή Detrended Fluctuation Analysis, DFA)

Η μέθοδος αυτή χρησιμοποιήθηκε από τους Peng et al. (1994) και περιλαμβάνει διάφορα στάδια. Αρχικά, τα δεδομένα χωρίζονται σε τμήματα μεγέθους m . Ύστερα, τα μερικά αθροίσματα $Y(i), i = 1, 2, \dots, m$ υπολογίζονται για κάθε τμήμα. Μια γραμμή ελαχίστων τετραγώνων προσαρμόζεται στο $Y(i)$ και μετά υπολογίζεται η δειγματική διασπορά των υπολοίπων της παλινδρόμησης. Αυτή η διαδικασία επαναλαμβάνεται για κάθε τμήμα και στο τέλος βρίσκεται ο μέσος

όρος από τις παραγόμενες δειγματικές διασπορές. Καθώς τα τμήματα είναι ίδιου μεγέθους η μέθοδος είναι ισοδύναμη με τον υπολογισμό της διασποράς σε ολόκληρη τη σειρά. Σε λογαριθμικό διάγραμμα της διασποράς συναρτήσει της τιμής m , η κλίση ισούται με $2H$ (Murad S. Taqqu, Teverovsky, and Willinger 1995).



Σχήμα 4.7 Διάγραμμα της μεθόδου Peng στο R για συνθετική χρονοσειρά FGN με $H = 0,7$

4.9 Εκτιμήτριες του Whittle

Οι εκτιμήτριες του Whittle βασίζονται στο περιοδόγραμμα. Περιλαμβάνουν τη συνάρτηση:

$$Q(n) := \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda;n)} d\lambda \quad (4.9)$$

όπου $I(\lambda)$ είναι το περιοδόγραμμα, $f(\lambda; n)$ η φασματική πυκνότητα στη συχνότητα λ και το n είναι το μητρώο των άγνωστων παραμέτρων. Σύμφωνα με τον Taqqu και Teverovsky(1997) οι εκτιμήτριες μπορεί να πάρουν τρεις μορφές.

- *Βασική εκτιμήτρια του Whittle*

Η εκτιμήτρια του Whittle είναι η τιμή του n η οποία ελαχιστοποιεί τη συνάρτηση Q . Για την περίπτωση του Κλασματικού Γκαουσιανού Θορύβου η παράμετρος n αντιστοιχεί στην παράμετρο H . Αυτή η εκτιμήτρια απαιτεί περισσότερο χρόνο για τον υπολογισμό της δίνει όμως τη δυνατότητα για εκτίμηση ορίων εμπιστοσύνης. Υποθέτει επίσης ότι η παραμετρική μορφή της φασματικής πυκνότητας είναι γνωστή. Στην περίπτωση που το μοντέλο προσομοιώνει ακριβώς την εξεταζόμενη σειρά αυτή η υπόθεση οδηγεί σε μια πολύ καλή εκτίμηση του H . Στην αντίθετη περίπτωση όμως μπορεί να δώσει λανθασμένα αποτελέσματα (M. S. Taqqu and Teverovsky 1997).

- *Συναθροισμένη μέθοδος του Whittle*

Αποτελεί μια παραλλαγή της τυπικής μεθόδου που χει μπορεί να αυξήσει την ευρωστία της εκτίμησης του H . Εφαρμόζεται για μεγάλο μήκος σειρών. Αρχικά η σειρά συναθροίζεται έτσι ώστε να προκύψει η εξής (σχέση 4.2):

$$X_i^{(m)} := \frac{1}{m} \sum_{j=m(i-1)+1}^{mi} X_j$$

Η βραχύτερη σειρά αυξάνει την τυπική απόκλιση της εκτιμήτριας. Όμως η ποσότητα $m^{1-H} X_i^{(m)}$ δεν συγκλίνει καλά στο Κλασματικό Γκαουσιανό Θόρυβο εκτός και αν η κλίμακα συνάθροισης m πάρει μεγάλη τιμή και η ύπαρξη μακροπρόθεσμης εμμονής είναι έντονη(M. S. Taqqu and Teverovsky 1997).

- *Τοπική εκτιμήτρια του Whittle*

Πρόκειται για μία ημιπαραμετρική εκτιμήτρια με την έννοια ότι προσδιορίζει μόνο την παραμετρική μορφή της φασματικής πυκνότητας όταν το n είναι πολύ κοντά στο μηδέν:

$$f(v) \sim G(H)|v|^{1-2H} \quad (4.10)$$

Καθώς το $|v| \rightarrow 0$.

Βασίζεται και αυτή στο περιοδόγραμμα αλλά ο υπολογισμός της περιλαμβάνει την εκτίμηση μιας πρόσθετης παραμέτρου m που είναι ακέραιος και για την οποία ισχύει: $m < \frac{N}{2}$ και $1/m + m/N \rightarrow 0$ για $N \rightarrow \infty$.

Για φασματική πυκνότητα που ικανοποιεί τη σχέση (4.10) το ανάλογο της συνάρτησης Q είναι

$$Q(G, H) := \frac{1}{m} \sum_{j=1}^m \left(\frac{I(v_j)}{G v_j^{1-2H}} + \log G v_j^{1-2H} \right) \quad (4.11)$$

όπου $v_j = 2\pi j/N$. Περισσότερες πληροφορίες υπάρχουν στις μελέτες των Taqqu και Teverovsky (1997; 1998).

❖ Σύγκριση των εκτιμητριών Whittle

Ακόμα και αν η εξεταζόμενη σειρά είναι «ιδανική» δηλαδή FGN θα υπάρχει πάντα μία θεωρητική μεροληψία. Αυτό γιατί η θεωρητική φασματική πυκνότητα είναι μόνο ανάλογη του v^{1-2H} καθώς το v τείνει στο μηδέν και επομένως για κάθε πεπερασμένου μήκους σειρά, όπου η ποσότητα m/N δεν είναι αμελητέα, η πραγματική τιμή του \hat{H} μπορεί να διαφέρει από την H_0 ακόμα και αν το περιοδόγραμμα αντικατασταθεί από τη φασματική πυκνότητα.

Οι Taqqu et Teverovsky(1997) καταλήγουν ότι για να επιτευχθεί ένας ικανοποιητικός βαθμός ακρίβειας πρέπει να χρησιμοποιηθεί μια σχετικά μικρή τιμή του m . Το τίμημα όμως είναι μια μεγαλύτερη τυπική απόκλιση οπότε τα αποτελέσματα στη μεροληψία και στην τυπική απόκλιση πρέπει να ισορροπηθούν.

Στην ίδια μελέτη πραγματοποιούν μια αναλυτική σύγκριση των μεθόδων και καταλήγουν στα παρακάτω.

Η προτιμώμενη μέθοδος εξαρτάται κυρίως από το πόσο καλά είναι γνωστή η ακριβής μορφή της μελετώμενης σειράς. Αν η σειρά είναι FGN τότε η σωστά προσδιορισμένη εκτιμήτρια του Whittle δίνει τη μικρότερη μεροληψία. Όταν η σειρά είναι μικρή ($N=100$) η συναθροισμένη μέθοδος δεν έχει εφαρμογή και η τοπική εκτιμήτρια είναι αναξιόπιστη γιατί τα συστηματικά λάθη είναι πολύ μεγάλα. Αν παρόλα αυτά η σειρά είναι FGN η τοπική εκτιμήτρια με $m = N/4$ είναι αποδεκτή γιατί δίνει μια καλύτερη εκτίμηση από μια λάθος προσδιορισμένη βασική εκτιμήτρια.

Όταν η σειρά είναι επαρκώς μεγάλη ($N=10.000$) τόσο η συναθροισμένη μέθοδος όσο και η τοπική δίνουν πολύ καλά αποτελέσματα με την τελευταία να αποδίδει ελαφρώς καλύτερα. Και οι δύο πάντως είναι πολύ πιο εύρωστες από την κλασική εκτιμήτρια καθώς αποδίδουν για αυτό το μήκος σειράς μόνο ελαφρώς χειρότερα από μια σωστά προσδιορισμένη Whittle. Το γεγονός επίσης ότι δεν χρειάζεται να προσδιορισθεί το μοντέλο για να χρησιμοποιηθούν αποτελεί συγκριτικό πλεονέκτημα. Ένα μειονέκτημα της συναθροισμένης και τοπικής μεθόδου είναι η ανάγκη προσεκτικής επιλογής της επιπλέον παραμέτρου m . Μια προσέγγιση είναι να δημιουργηθεί ένα διάγραμμα της εκτιμήτριας συναρτήσεως του m και να επιλεγεί το m από μια περιοχή όπου η εκτιμήτρια παραμένει σχετικά σταθερή.

4.10 Μέθοδος Κυματικής Ανάλυσης (Wavelet analysis)

Τα κυμάτια είναι ένα εργαλείο που χρησιμοποιείται στην ανάλυση του σήματος (signal analysis) και είναι παρόμοια σε λογική με το μετασχηματισμό Fourier. Η βασική διαφορά είναι ότι ο κυματικός μετασχηματισμός γίνεται τόσο στο χρόνο όσο και στις συχνότητες ενώ ο μετασχηματισμός Fourier αφορά μόνο τον τομέα των συχνοτήτων.

Υπάρχουν δύο βασικές συναρτήσεις στην κυματική ανάλυση, η συνάρτηση κλίμακας φ_0 και η μητρική ψ_0 που προκύπτει από την πρώτη. Η μητρική συνάρτηση πρέπει να ικανοποιεί τη σχέση $\int \psi_0(t)dt = 0$ και για τον αντίστοιχο μετασχηματισμό Fourier να ισχύει $|\Psi_0(v)| \sim v^N, v \rightarrow 0$ όπου N ένας θετικός ακέραιος που ονομάζεται αριθμός των εξαφανιζόμενων ροπών του κυματίου σύμφωνα με I. Daubechies (1992).

Ο συντελεστής $|d_x(j,k)|^2$ μετράει την ποσότητα της ενέργειας στο αναλυόμενο σήμα τη χρονική στιγμή $2^j k$ και συχνότητα $2^{-j}v_0$ όπου v_0 είναι μια αυθαίρετη αναφορά συχνότητας που έχει καθοριστεί από την επιλογή του ψ_0 . Έχει προταθεί από τους Arby, Goncalves και Flandrin(1993; 1995) ότι μια χρήσιμη φασματική εκτιμήτρια μπορεί να σχεδιασθεί από τον υπολογισμό ενός μέσου όρου χρόνου (time average) του $|d_x(j,k)|^2$ σε μια δεδομένη κλίμακα, δηλαδή:

$$\hat{\Gamma}_x(2^{-j}v_0) = \frac{1}{n_j} \sum_k |d_x(j,k)|^2 \quad (4.12)$$

Οι Arby και Veitch (1998) δείχνουν ότι μια αμερόληπτη εκτιμήτρια μπορεί να προκύψει από τη γραμμική συσχέτιση του $\log_2(\hat{\Gamma}_x(2^{-j}v_0))$ ως προς το j :

$$\log_2(\hat{\Gamma}_x(2^{-j}v_0)) = \log_2\left(\frac{1}{n_j} \sum_k |d_x(j,k)|^2\right) = (2\hat{H} - 1)j + \hat{c} \quad (4.13)$$

όπου το \hat{c} εκτιμάει την ποσότητα $\log_2(c_f \int |v|^{(1-2H)} |\Psi_0(v)|^2 dv)$ εφόσον το ολοκλήρωμα $\int |v|^{(1-2H)} |\Psi_0(v)|^2 dv$ συγκλίνει.

Σύμφωνα με τους Arby και Veitch (1998) υπάρχουν λόγοι για τους οποίους η κυματική ανάλυση πλεονεκτεί έναντι κάποιων εκ των ήδη γνωστών μεθόδων εκτίμησης της μακροπρόθεσμης εμμονής. Έχει ήδη αναφερθεί ότι οι παραδοσιακές απλές μέθοδοι εκτίμησης του συντελεστή Hurst παρουσιάζουν μεγάλο βαθμό μεροληψίας. Παρόλα αυτά οι μέθοδοι που προκύπτουν από τη γκαουσιανή μέθοδο μέγιστης πιθανοφάνειας μπορούν να δώσουν ασυμπτωτικά αμερόληπτες εκτιμήτριες. Όμως απαιτούν την a priori επιλογή μιας

παραμετρικής οικογένειας μοντέλων και επιπλέον για μεγάλα σετ δεδομένων προκύπτουν υπολογιστικές πολυπλοκότητες και μεγάλες ανάγκες υπολογιστικής μνήμης (J. Beran 1994). Σε αντίθεση με τα παραπάνω, η κυματική ανάλυση είναι ένα εργαλείο που μελετά τη μακροπρόθεσμη εξάρτηση κατευθείαν μέσω συντελεστών που προκύπτουν από τη συνδυασμένη σε χρόνο και κλίμακα αποδόμηση. Έτσι, λίγες υποθέσεις πρέπει να γίνουν για το υποκείμενο μοντέλο. Δεδομένου ότι υπάρχουν ενδείξεις μακράς μνήμης, η κυματική ανάλυση δίνει ημιπαραμετρικές αμερόληπτες εκτιμήτριες για την εκτίμησή της. Ο υπολογισμός της εκτιμήτριας μπορεί να γίνει επίσης με έναν απευθείας υπολογιστικό αλγόριθμο με χαμηλή πολυπλοκότητα, αντί για αλγόριθμο βελτιστοποίησης, κάτι που επιτρέπει την ανάλυση πολύ μεγάλων σετ δεδομένων (Abry and Veitch 1998).

4.11 Άλλες μέθοδοι

Αναφέρθηκαν ήδη συνοπτικά οι περισσότεροι γνωστές μέθοδοι εκτίμησης της μακροπρόθεσμης εμμονής όπως εντοπίζονται στη βιβλιογραφία. Για περαιτέρω ανάλυση και σύγκριση των μεθόδων υπάρχουν αρκετές αναφορές επίσης (Weron 2002; William Rea et al. 2009; W. Rea et al. 2012).

Υπάρχουν βέβαια και άλλες μέθοδοι που δεν περιγράφηκαν ή βελτιώσεις των ήδη υπάρχουσών μεθόδων. Χαρακτηριστικά αναφέρονται η υπολογιστικά δυσκολότερη μέθοδος της μέγιστης πιθανοφάνειας (McLeod and Hipel 1978) και η βελτιωμένη μέθοδος R/S (Mielniczuk and Wojdyło 2007).

4.12 Βιβλιογραφική επισκόπηση μελετών για την ύπαρξη μακροπρόθεσμης εμμονής στο φαινόμενο της βροχόπτωσης

Η ύπαρξη μακροπρόθεσμης εμμονής στο φαινόμενο της βροχόπτωσης έχει μελετηθεί ιδιαίτερα ως προς τις λεπτές χρονικές κλίμακες. Οι Κουτσογιάννης, Παπαλεξίου και Montanari (2011) μελετώντας ένα δείγμα 7 γεγονότων καταιγίδας (μετρήσεων ανά 10s) του πανεπιστημίου της Iowa διαπίστωσαν με τη χρήση της μεθόδου του κλιμακογράμματος ότι ο ελάχιστος συντελεστής H

που παρατηρήθηκε ήταν 0,79. Υποθέτοντας δε ότι τα επεισόδια είναι εκφάνσεις της ίδιας στοχαστικής ανέλιξης βρέθηκε ότι ο κοινός συντελεστής H ισούται με 0,94. Χρησιμοποιώντας το ίδιο δείγμα δεδομένων και με τη χρήση της μεθόδου κυματικής ανάλυσης, οι Kumar και Foufoula-Georgiou (1997) έδειξαν ότι υπάρχουν κυρίαρχες και διακριτές κλίμακες διακύμανσης στην παρατηρημένη βροχόπτωση, σε αντίθεση με την υπόθεση της αμεταβλητότητας ως προς την κλίμακα (scale invariance). Οι Bove, Pelino και De Leonibus (2006) μελετώντας μετρήσεις βροχόπτωσης 15min από 60 αυτοματοποιημένους σταθμούς της Ιταλίας κατέληξαν με τη χρήση της μεθόδου R/S στην ύπαρξη συντελεστή $H = 0,7$.

Βέβαια η συμπεριφορά κλίμακας στη βροχόπτωση υψηλής ανάλυσης είναι διαφορετικό φαινόμενο από τη συμπεριφορά κλίμακας στη βροχόπτωση μεγαλύτερων κλιμάκων. Οι Matsoukas και Islam (2000) μελέτησαν 9 χαρακτηριστικούς βροχομετρικούς σταθμούς των Η.Π.Α με τη μέθοδο DFA (detrended fluctuation analysis) και βρήκαν ότι υπάρχουν δύο διαφορετικές περιοχές στη δομή της αυτοσυσχέτισης, μία που χαρακτηρίζει χρονικά διαστήματα μιας ώρας έως 5 ημερών, και μία για χρονικά διαστήματα από 10 ημέρες μέχρι χρόνια. Για χρονική ανάλυση μικρότερη των 5 ημερών κατέληξαν στην ύπαρξη κλασματικού θορύβου (FGN), ενώ για μεγαλύτερα διαστήματα η δομή της αυτοσυσχέτισης γινόταν ασθενέστερη χωρίς όμως να μπορεί να χαρακτηριστεί λευκός θόρυβος.

Οι Ramachandra και Bhattacharya(1999) μελέτησαν ένα μικρό δείγμα υδρολογικών σειρών μεταξύ των οποίων δύο μηνιαίων και δύο ετήσιων χρονοσειρών βροχόπτωσης με τη μέθοδο R/S και με έλεγχο υπόθεσης για βραχυπρόθεσμη εξάρτηση και πλήρη ανεξαρτησία. Κατέληξαν ότι υπήρχαν λίγες ενδείξεις για ύπαρξη μακροπρόθεσμης εμμονής σε μηνιαία δεδομένα (στα οποία είχε αφαιρεθεί η επίδραση της περιοδικότητας) και ακόμα λιγότερες σε ετήσια, αν και το αποτέλεσμα στη δεύτερη περίπτωση κρίθηκε αβέβαιο λόγω του μικρού μήκους δείγματος και της αδυναμίας της μεθόδου. Οι Pellieter και Turcotte (1997) μελετώντας με τη βοήθεια της φασματικής ανάλυσης 49 σταθμούς ετήσιας βροχόπτωσης καταγραφών άνω των 128 ετών παρατήρησαν ότι δεν υπήρχε ένδειξη εμμονής για χρονικές κλίμακες μικρότερες των 10 ετών,

σε αντίθεση με τα αντίστοιχα ευρήματα τους για τους δεντροδακτυλίους και την απορροή ποταμών. Για μεγαλύτερες όμως κλίμακες βρήκαν συντελεστή H ίσο με 0,75. Σε μια πιο εκτεταμένη μελέτη, οι Koutsoyiannis και Langousis (2011) μελετώντας παγκόσμια μηνιαία δεδομένα βροχόπτωσης από δορυφόρο βρήκαν με τη μέθοδο του κλιμακογράμματος συντελεστή Hurst ίσο με 0,7.

Όπως φαίνεται από τα παραπάνω, υπάρχουν λίγες βιβλιογραφικές αναφορές στην διερεύνηση μακροπρόθεσμης εμμονής στην βροχόπτωση και ακόμα λιγότερες στην ετήσια. Τα αποτελέσματα συχνά είναι δύσκολο να αξιολογηθούν λόγω των διαφορετικών μεθόδων αλλά και των διαφορετικών δεδομένων (συνήθως περιορισμένων) που ερευνήθηκαν. Το μεγαλύτερο μέρος της σχετικής βιβλιογραφίας επίσης αφορά στην διερεύνηση και ταυτοποίηση ντετερμινιστικών τάσεων στην βροχόπτωση (Hanssen-Bauer and Førland 1998; Zhang et al. 2000; Liu et al. 2005; Jiang, Su, and Hartmann 2007; Lennartz and Bunde 2011), προσέγγιση όμως που αμφισβητείται ως προς την ορθότητά της (Koutsoyiannis and Montanari 2007; Koutsoyiannis 2010; Koutsoyiannis 2011a).

5. Εφαρμογή σε πραγματικά δεδομένα

5.1 Περιγραφή βάσης δεδομένων

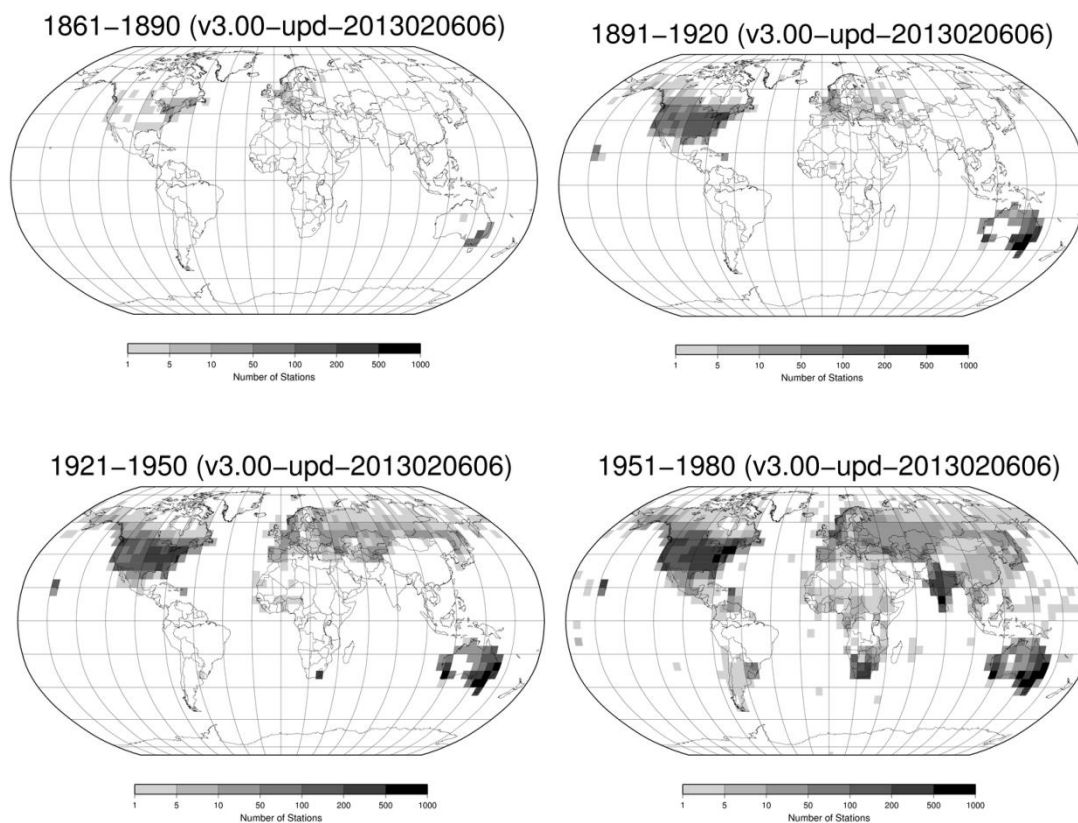
Τα δεδομένα που μελετήθηκαν προήλθαν από την επεξεργασία ημερήσιων τιμών βροχόπτωσης από το Παγκόσμιο Ιστορικό Δίκτυο Κλιματολογίας – Ημερήσια Βάση Δεδομένων (Global Historical Climatology Network, version 2.60, <http://www.ncdc.noaa.gov/oa/climate/ghcn-daily>) το οποίο περιέχει δεδομένα από σταθμούς σε όλο τον κόσμο. Συγκεκριμένα, το GHCN-daily αποτελεί μια ενσωματωμένη βάση δεδομένων που περιέχει ημερήσια κλιματικά αρχεία ελεγμένης ποιότητας. Περιέχει δεδομένα από περίπου 75.000 σταθμούς από 180 χώρες που περιλαμβάνουν γεωφυσικές μετρήσεις όπως μέγιστη και ελάχιστη θερμοκρασία, ολική ημερήσια βροχόπτωση, χιονόπτωση και ύψος χιονιού. Περίπου όμως το 2/3 των σταθμών καταγράφει μόνο τη βροχόπτωση. Το μήκος των αρχείων καθώς και η χρονική περίοδος των καταγραφών ποικίλλουν από λιγότερο από ένα χρόνο μέχρι περισσότερα από 175 χρόνια.

Το δείγμα δεδομένων υπόκειται σε ελέγχους ποιότητας που λαμβάνουν χώρα περίπου κάθε σαββατοκύριακο κατά τη διάρκεια των οποίων τα δεδομένα ανακατασκευάζονται από τις πηγές δεδομένων που τα συνθέτουν. Η βάση ενημερώνεται έτσι όταν είναι δυνατό από μια ποικιλία νέων πληροφοριών όταν οι τελευταίες είναι διαθέσιμες και έχουν επίσης έχουν υποστεί ελέγχους ποιότητας. Κάθε δείγμα συνοδεύεται από ένα αρχείο στο οποίο ορίζεται η ποιότητα των δεδομένων και αξιολογείται η πηγή προέλευσης με χρήση «σημαιών ποιότητας». Αυτή η βάση δεδομένων είναι επίσης η επίσημη βάση δεδομένων που χρησιμοποιεί το U.S. Cooperative Observer data.

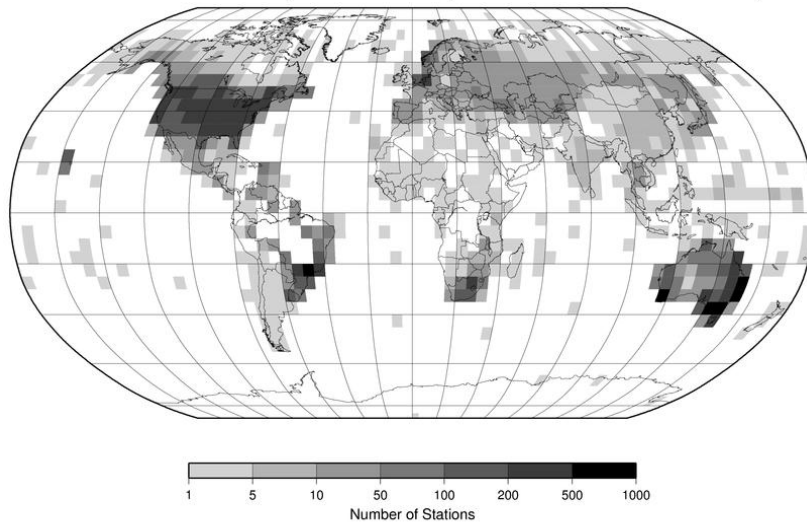
Τα δεδομένα έχουν συλλεχθεί μέσω επαφών με αντιπροσώπους των Εθνικών Μετεωρολογικών και Υδρολογικών Κέντρων (National Meteorological and Hydrological Centers) παγκοσμίως και προέρχονται από τέσσερις βασικές πηγές : α) την βάση δεδομένων Η.Π.Α (U.S Collection), β) τη διεθνή βάση

δεδομένων εκτός Η.Π.Α (International Collection), γ) κρατικά δεδομένα (Government Exchange Data) και δ) ημερήσια δεδομένα (Global Summary of the Day) που μεταφέρονται μέσω του δικτύου GTS (Global Telecommunications System).

Η πυκνότητα του δικτύου σταθμών είναι γενικά μεγαλύτερη στις Η.Π.Α, στον Καναδά και στην Αυστραλία και μικρότερη σε περιοχές όπως η Αφρική, η Νότια Αμερική ή η Ανταρκτική. Κάποιοι σταθμοί παρόλα αυτά προέρχονται από τη Βραζιλία, την Ινδία και τη Νότια Αφρική. Η χρονική εξέλιξη του δικτύου των σταθμών είναι ιδιαίτερα έντονη. Χαρακτηριστικό είναι ότι πριν το 1890 ο αριθμός των σταθμών που κατέγραφαν τη βροχόπτωση αποτελούσε μόνο το 8,9% του μεγίστου αριθμού σταθμών του επόμενου αιώνα. Παρακάτω φαίνεται η χρονική πορεία της γεωγραφικής επέκτασης του συνόλου των βροχομετρικών σταθμών.



1981–2010 (v3.00–upd–2013020606)



Σχήμα 5.1 Χρονική εξέλιξη του δικτύου των βροχομετρικών σταθμών. Πηγή:
<http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/>

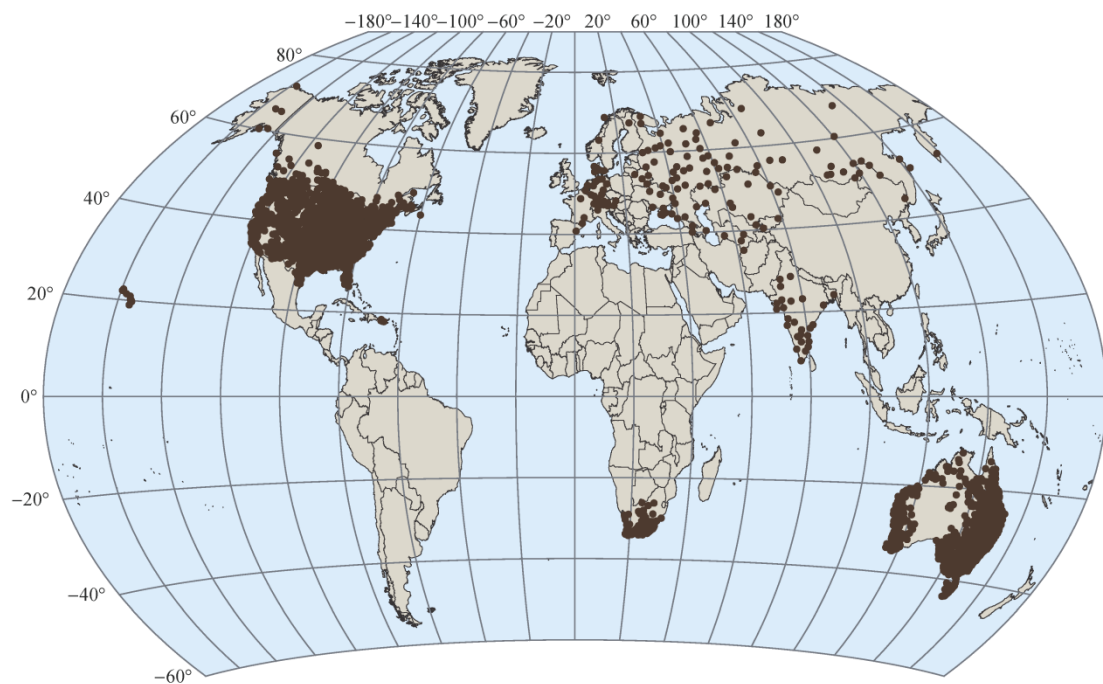
Η διαδικασία της ενσωμάτωσης νέων δεδομένων στο δίκτυο περιλαμβάνει γενικά τρία στάδια: α) το στάδιο της αρχικής διαλογής δεδομένων βάσει προϋποθέσεων (π.χ. ύπαρξη επαρκών γεωγραφικών και ιστορικών πληροφοριών για το σταθμό) β) το στάδιο της ταξινόμησης σταθμού είτε ως τοποθεσία για την οποία ήδη υπάρχουν δεδομένα είτε ως νέα τοποθεσία και γ) το στάδιο του συγκερασμού των σταθμών με χρήση ιεραρχικής ταξινόμησης βάσει κριτηρίων ποιότητας των πηγών προέλευσης των δεδομένων. Σε περιπτώσεις αλληλοεπικάλυψης των δεδομένων χρησιμοποιούνται δεδομένα από την πιο αξιόπιστη πηγή. Επίσης, δεδομένα από μια νέα πηγή για την ίδια τοποθεσία μπορεί να συγκριθούν με τα υπάρχοντα δεδομένα και ανάλογα με το βαθμό ταύτισης και την απόσταση μεταξύ των δύο πηγών, να δημιουργηθεί ή όχι νέα βάση δεδομένων για αυτήν την πηγή.

Αξίζει να σημειωθεί ότι το GHCN-Daily δεν περιέχει διορθώσεις για τη μεροληψία που προκύπτει από ιστορικές αλλαγές στα όργανα και τις πρακτικές μέτρησης. Επίσης, οι σταθμοί δεν ήταν δυνατό να ικανοποιήσουν όλα τα απαιτούμενα κριτήρια για την παρατήρηση του κλίματος καθώς οι περισσότεροι έχουν ως στόχο την ικανοποίηση μετρητικών αναγκών για την γεωργία, την

υδρολογία, την πρόγνωση καιρού, την αεροπορία κλπ. Περισσότερες πληροφορίες για τη μεθοδολογία της βάσης δεδομένων μπορεί κανείς να βρει στην ιστοσελίδα <http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/> καθώς και στη βιβλιογραφία (Durre, Menne, and Vose 2008; Durre et al. 2010; Menne et al. 2012).

5.2 Συμπλήρωση δεδομένων

Σε πρώτη φάση επεξεργασίας χρησιμοποιήθηκαν σταθμοί που είχαν καταγραφές άνω των 100 ετών και δεν τους έλειπε πάνω από το 20% των δεδομένων τους για τα οποία η αμφισβητούμενη ποιότητα ήταν μικρότερη του 0,1%. Προέκυψαν έτσι 3.477 σταθμοί.



Σχήμα 5.2 Γεωγραφική απεικόνιση των υπό εξέταση σταθμών

Για την εκτίμηση του Hurst ενδιαφέρει στην περίπτωση αυτής της διπλωματικής η επεξεργασία των ετήσιων τιμών κάθε σταθμού. Είναι απαραίτητη δηλαδή η εξαγωγή της ετήσιας τιμής από δεδομένα που περιέχουν ημερήσιες. Λόγω της ελλιπούς φύσης όμως των δεδομένων ήταν αναγκαίο να αποφασιστούν κάποια κριτήρια απόρριψης ή σε ευνοϊκότερες περιπτώσεις συμπλήρωσης ετών χωρίς να καταστραφεί η δομή της αυτοσυσχέτισης εφόσον υπάρχει. Σε δεύτερη φάση έγινε περαιτέρω διερεύνηση των αποτελεσμάτων και αποφασίστηκε η μεθοδολογία απόρριψης προβληματικών σταθμών ώστε να προκύψουν τα τελικά δεδομένα προς επεξεργασία.

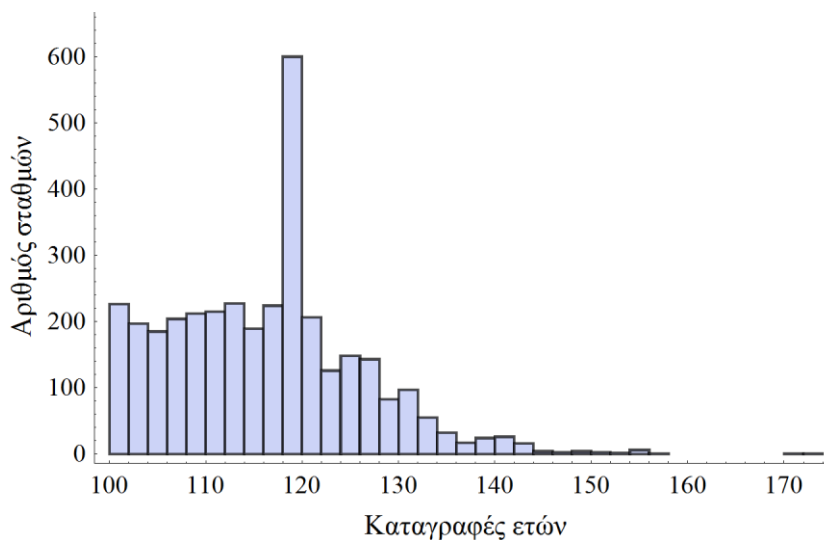
Αποφασίστηκε η χρησιμοποίηση μιας απλής αλλά αποτελεσματικής μεθόδου συμπλήρωσης που στηρίζεται στη μεθοδολογία που ανέπτυξε ο Παππάς (2010). Σύμφωνα με τη δημοσιευμένη του μελέτη, σε περιπτώσεις χρονοσειρών που παρουσιάζουν δυναμική Hurst-Kolmogorov είναι προτιμότερη η συμπλήρωση μεμονωμένων κενών με βάση τον τοπικό μέσο όρο των γειτονικών τιμών. Με βάση αυτή τη θεώρηση η μεθοδολογία που αναπτύχθηκε έχει ως εξής:

1. Αρχικά διορθώθηκαν όλες οι προβληματικές τιμές δηλαδή οι αδικαιολόγητα μεγάλες τιμές σε όλους τους σταθμούς και αντικαταστάθηκαν με κενό.
2. Στη συνέχεια μέσα σε κάθε σταθμό έγινε χωρισμός σε έτη και σε όποιες περιπτώσεις έλειπε λιγότερο από το 5,5% των τιμών, δηλαδή λιγότερες από 20 μέρες υπολογίστηκε η μέση τιμή. Χρησιμοποιήθηκαν επομένως τιμές από έτη που είχαν από 345 μέρες και πάνω. Η τιμή που προέκυψε είναι ο μέσος όρος του έτους και στη συνέχεια ονομάζεται *μέση ημερήσια τιμή έτους* (mm/d). Είναι χαρακτηριστική τιμή για το έτος επίσης αφού με πολλαπλασιασμό επί των ημερών του έτους προκύπτει η ετήσια τιμή (mm/yr).
3. Τέλος αντικαταστάθηκαν όσα μεμονωμένα κενά υπήρχαν με το μέσο όρο της επόμενης και προηγούμενης ετήσιας τιμής εφόσον δεν ήταν ελλιπείς.

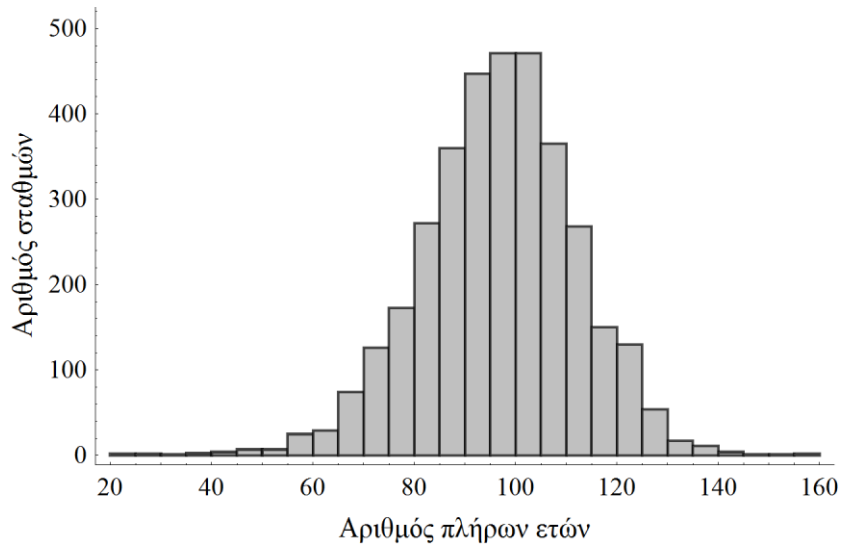
Στη συνέχεια έγινε επεξεργασία σε επίπεδο σταθμού. Συγκεκριμένα από τους 3477 σταθμούς προέκυψαν:

- 558 σταθμοί που είχαν 100 συνεχόμενα χρόνια (χωρίς κενά στις ετήσιες τιμές)
- 1474 σταθμοί που είχαν πάνω από 100 ετήσιες τιμές ανεξαρτήτως της παρεμβολής κενών
- 8 σταθμοί στους οποίους δεν παρατηρήθηκε κανένα κενό στις ετήσιες τιμές
- 1265 σταθμοί στους οποίους υπήρχαν 100 ετήσιες τιμές (ανεξαρτήτως της παρεμβολής κενών) και το ποσοστό των ελλειπουσών τιμών δεν υπερέβαινε το 15% των τιμών του σταθμού.

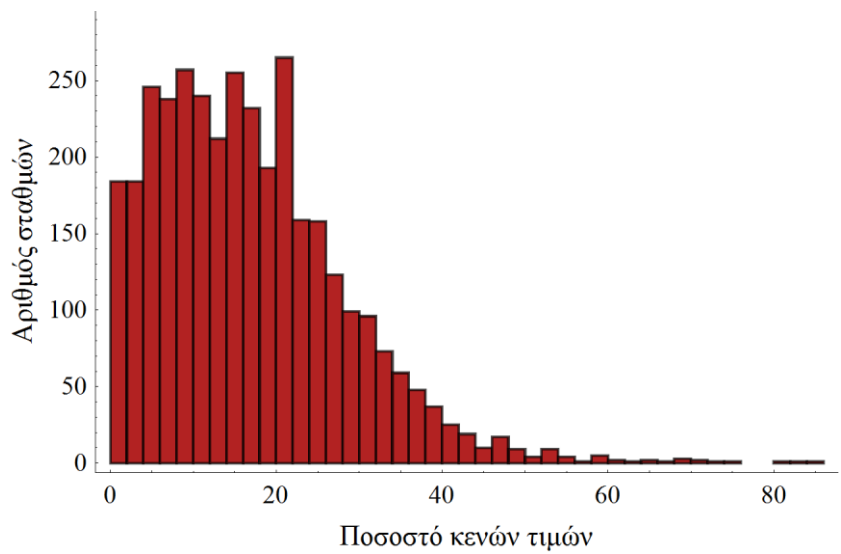
Παρακάτω παρατίθενται ιστογράμματα που δείχνουν το στατιστικό προφίλ των υπό εξέταση σταθμών:



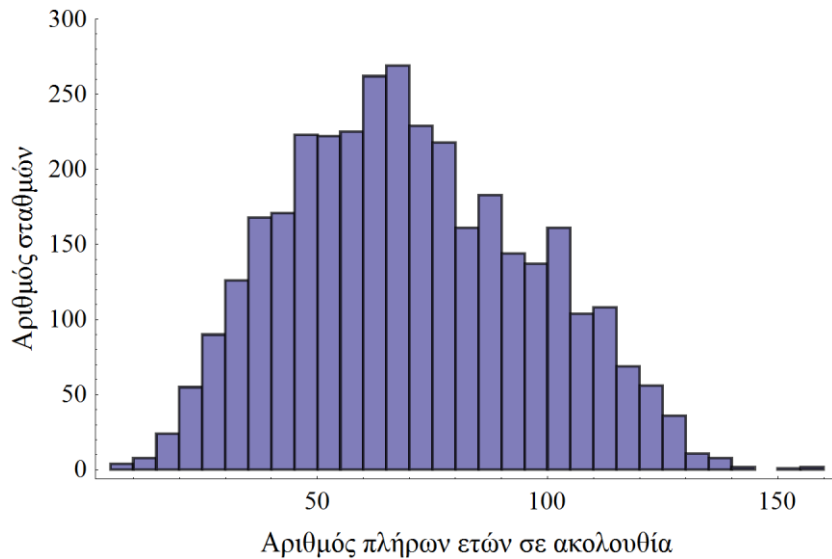
Σχήμα 5.3 Ιστόγραμμα κατανομής των σταθμών με βάση τις καταγραφές ετών που περιέχουν



Σχήμα 5.4 Ιστόγραμμα κατανομής των σταθμών με βάση τον αριθμό των ετήσιων τιμών που δεν είναι ελλειπείς



Σχήμα 5.5 Ιστόγραμμα κατανομής των σταθμών με βάση το ποσοστό των κενών τιμών τους προς το σύνολο των καταγραφών



Σχήμα 5.6 Ιστόγραμμα κατανομής των σταθμών με βάση τη μέγιστη συνεχόμενη ακολουθία ετήσιων τιμών χωρίς ελλείψεις

Όπως φαίνεται στο σχήμα 5.3 όλοι οι σταθμοί έχουν καταγραφές άνω των 100 χρόνων όπως ήταν αναμενόμενο από την αρχική τους επιλογή. Η πλειοψηφία των σταθμών (95%) έχει καταγραφές από 100 έως και 138 χρόνια ενώ μόλις δύο σταθμοί ξεπερνούν τα 160 χρόνια. Παρόλα αυτά το σχήμα 5.4 είναι ενδεικτικό της ποιότητας των σταθμών αφού δείχνει πόσες από αυτές τις τιμές υπήρχαν ή ήταν δυνατό να συμπληρωθούν. Η κατανομή τους πλησιάζει στην κανονική κατανομή με μέση τιμή τα 96 χρόνια και δείχνει ότι περίπου οι μισοί μόνο από τους σταθμούς έχουν πάνω από 96 συμπληρωμένες τιμές ετών. Σύμφωνα και με το σχήμα 5.5 το ποσοστό των κενών στην πλειοψηφία των σταθμών (95%) κινείται ανάμεσα στο 2 - 43% των συνολικών τιμών, ενώ το μέσο ποσοστό ελλিপών τιμών είναι περίπου 17%. Τέλος, όσον αφορά τη μέγιστη ακολουθία ετών χωρίς κενά φαίνεται ότι περίπου το 16% των σταθμών ξεπερνά τα 100 χρόνια. Η μέση τιμή των ετών μιας πλήρους χρονοσειράς είναι περίπου 70 χρόνια, ενώ το 95% των τιμών βρίσκεται στο διάστημα 24 - 122 χρόνια.

Η επιλογή των σταθμών στηρίχθηκε σε τρία κριτήρια: α) την ύπαρξη ενός επαρκούς αριθμού ετών για την εκτίμηση του Hurst β) την αξιοπιστία του σταθμού και γ) την προσαρμογή της αυστηρότητας των κριτηρίων ώστε να προκύψει ικανός στατιστικά αριθμός σταθμών για την εξαγωγή

συμπερασμάτων. Συγκεκριμένα επιλέχθηκαν σταθμοί που είχαν τουλάχιστον 100 μη κενές ετήσιες τιμές και στους οποίους το ποσοστό των ελλিপών τιμών δεν ήταν πολύ υψηλό. Από τους παραπάνω συνδυασμούς λοιπόν επιλέχτηκαν οι 1265 σταθμοί που είχαν 100 μη κενές τιμές και τους έλειπε ποσοστό λιγότερο από το 15% των καταγραφών του σταθμού. Η ύπαρξη ορισμένων ενδιάμεσων κενών αντιμετωπίστηκε στο στάδιο της εφαρμογής της μεθόδου.

5.3 Στατιστική επεξεργασία τελικών σταθμών

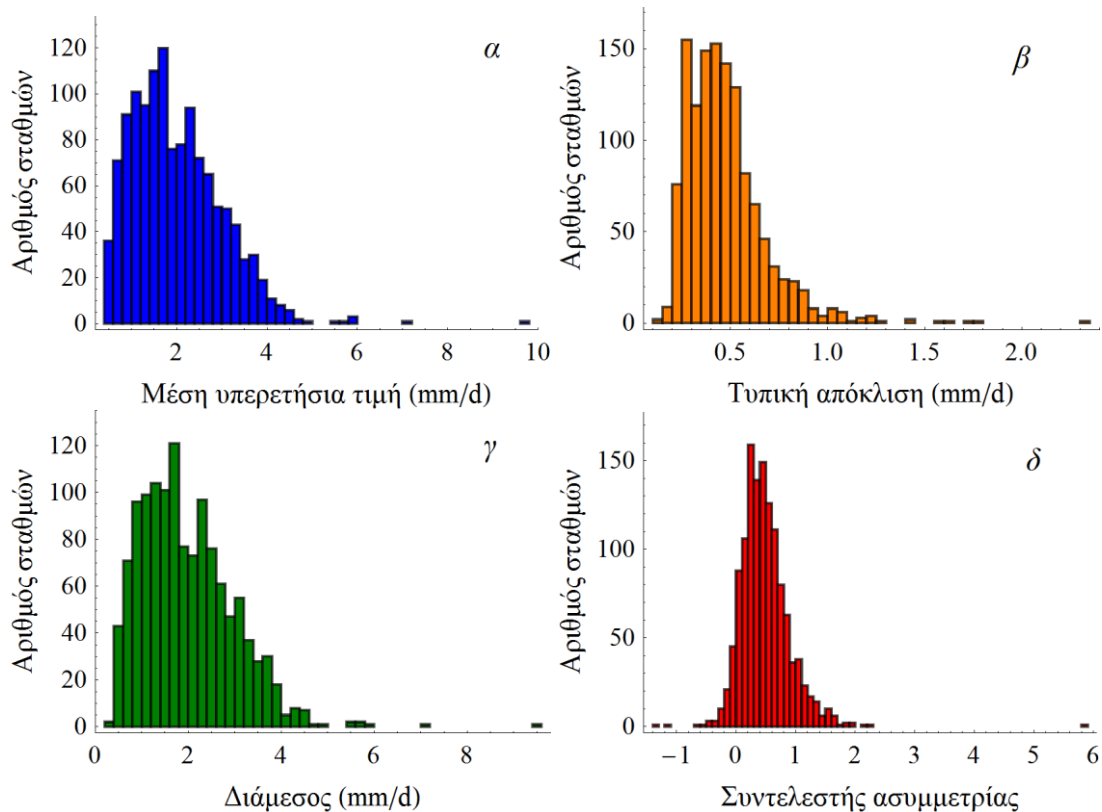
Παρακάτω δίνονται οι πίνακες και τα διαγράμματα που προέκυψαν από τη στατιστική επεξεργασία των 1265 σταθμών που πληρούσαν τα προαναφερθέντα κριτήρια. Μετά την αρχική επεξεργασία και συμπλήρωση των δεδομένων, κάθε σταθμός περιείχε έναν αριθμό *μέσων ημερήσιων τιμών* (mm/d) που η κάθε μία είχε προκύψει ως μέσος όρος του αντίστοιχου έτους. Έτσι, για κάθε σταθμό υπολογίστηκε η μέση τιμή του δείγματος του, δηλαδή η *μέση υπερετήσια τιμή* (mm/d). Υπολογίστηκαν επίσης η τυπική απόκλιση, η διάμεσος και ο συντελεστής ασυμμετρίας του δείγματος των μέσων ημερήσιων τιμών έτους, καθώς και οι ακραίες παρατηρήσεις του δείγματος, δηλαδή η ελάχιστη τιμή της μέσης ημερήσιας βροχόπτωσης έτους του σταθμού και αντίστοιχα, η μέγιστη τιμή. Για κάθε ένα από αυτά τα μεγέθη έγινε στατιστική ανάλυση σε επίπεδο δείγματος σταθμών, που περιλαμβάνει τον υπολογισμό της μέσης τιμής τους, της τυπικής απόκλισης, της ελάχιστης τιμής, των βασικών ποσοστημορίων και της μέγιστης τιμής. Τα αποτελέσματα της ανάλυσης παρουσιάζονται στον πίνακα 5.1 ενώ στα σχήματα 5.7 και 5.9 φαίνονται οι κατανομές των μεγεθών.

Σημειώνεται ότι η αναμενόμενη μεροληψία στην κλασική στατιστική εκτίμηση παραμέτρων για χρονοσειρές που παρουσιάζουν μακροπρόθεσμη εμμονή δεν είναι ιδιαίτερα σημαντική εδώ λόγω του χαμηλού συντελεστή Hurst που εκτιμήθηκε (5.4).

Πίνακας 5.1 Στατιστικά χαρακτηριστικά στατιστικών μεγεθών των σταθμών

	Μέση τιμή	Τυπική απόκλιση	Διάμεσος	Συντελεστής ασυμμετρίας	Ελάχιστη τιμή	Μέγιστη τιμή
Μέση τιμή	1,998	0,474	1,963	0,483	0,996	3,461
Τυπική απόκλιση	0,99	0,205	0,98	0,413	0,654	1,536
Ελάχιστη τιμή	0,41	0,132	0,365	-1,39	0	0,881
Q_{2,5}	0,591	0,223	0,555	-0,141	0,106	1,459
Q₂₅	1,22	0,333	1,205	0,225	0,48	2,327
Q₅₀	1,82	0,438	1,794	0,434	0,858	3,239
Q₇₅	2,61	0,557	2,576	0,682	1,421	4,271
Q_{97,5}	4,056	0,993	3,993	1,395	2,378	6,772
Μέγιστη τιμή	9,696	2,315	9,577	5,837	4,863	16,15

Στον πίνακα φαίνεται ότι η μέση τιμή της μέσης υπερετήσιας βροχόπτωσης όλων των σταθμών είναι 1,998 mm/d, δηλαδή περίπου 729 mm/yr. Η τυπική απόκλιση της μέσης υπερετήσιας τιμής του δείγματος των σταθμών 0,99 mm/d είναι ενδεικτική των μεγάλων διαφορών που εντοπίζονται σε παγκόσμια κλίμακα, όπως επίσης και το γεγονός ότι το 95% των τιμών βρίσκεται στο διάστημα 0,59 – 4,056 mm/d. Η μέση τιμή της διαμέσου και τα υπόλοιπα στατιστικά της χαρακτηριστικά είναι επίσης αρκετά κοντά στις αντίστοιχες εκτιμήσεις για τη μέση τιμή. Το μεγάλο εύρος των τιμών της ελάχιστης και της μέγιστης βροχόπτωσης είναι ενδεικτικό επίσης της μεγάλης διασποράς που χαρακτηρίζει το φαινόμενο σε παγκόσμια κλίμακα.

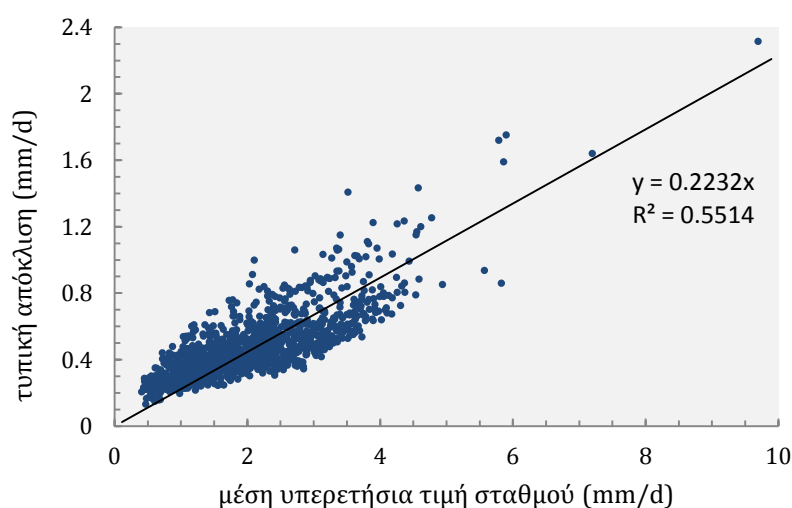


Σχήμα 5.7 Κατανομή σταθμών για τα εξής στατιστικά μεγέθη: *σχήμα α)* μέση υπερετήσια τιμή (mm/d), *σχήμα β)* τυπική απόκλιση μέσης ημερήσιας τιμής έτους (mm/d), *σχήμα γ)* διάμεσος μέσης ημερήσιας τιμής έτους (mm/d), *σχήμα δ)* συντελεστής ασυμμετρίας μέσης ημερήσιας τιμής έτους

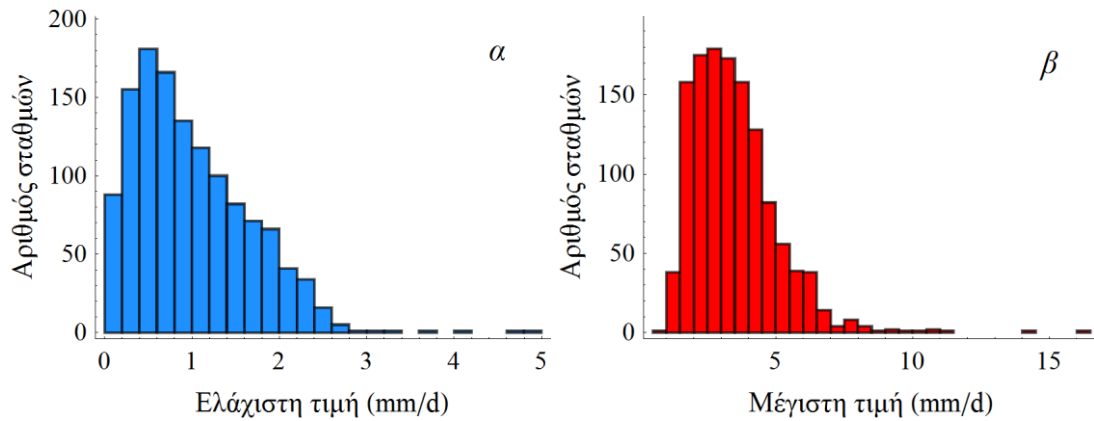
Στο σχήμα 5.7 φαίνονται πιο εποπτικά ορισμένα από τα αποτελέσματα του πίνακα όπως το γεγονός ότι η κατανομή της μέσης τιμής πλησιάζει την κατανομή της διαμέσου. Παρατηρούμε ότι μόνο δύο σταθμοί έχουν μέση υπερετήσια βροχόπτωση άνω των 9 mm/d ενώ περίπου 200 σταθμοί έχουν μέση υπερετήσια βροχόπτωση μικρότερη ή ίση του 1 mm/d. Συγκεκριμένα, η μέγιστη μέση υπερετήσια βροχόπτωση έτους ανήκει στον σταθμό Innisfail που βρίσκεται στην Αυστραλία, ενώ η ελάχιστη παρατηρήθηκε στο σταθμό Woomera (Arcoona) που βρίσκεται επίσης στην Αυστραλία. Σημειώνεται ότι η πόλη Innisfail που ανήκει στο Queensland της Αυστραλίας είναι γνωστή ως μια από τις πιο βροχερές πόλεις της Αυστραλίας η οποία χαρακτηρίζεται από τροπικό κλίμα με έντονη υγρασία και ζέστη και μεγάλες διαφορές βροχόπτωσης ανάμεσα στο χειμώνα και στο καλοκαίρι.

Αντίθετα, στην πόλη Woomera επικρατεί κλίμα ερήμου με ζεστά και ξηρά καλοκαίρια και κρύους χειμώνες.

Η τυπική απόκλιση των σταθμών είναι περίπου για τους μισούς μικρότερη του 0,5 ενώ σε ορισμένες περιπτώσεις παρατηρείται τυπική απόκλιση μεγαλύτερη του 1,5 όπως είναι αναμενόμενο σε περιοχές με μεγάλη κλιματική ποικιλομορφία (π.χ τροπικά κλίματα). Συγκεκριμένα, η μεγαλύτερη τυπική απόκλιση παρατηρήθηκε στο σταθμό Innisfail στην Αυστραλία, ενώ η ελάχιστη στο σταθμό Sunnyside στις Η.Π.Α. Το γεγονός ότι η μέγιστη τυπική απόκλιση παρατηρήθηκε στο σταθμό που παρουσίασε τη μέγιστη υπερετήσια τιμή είναι αναμενόμενο καθώς όπως φαίνεται και στο διάγραμμα 5.8 υπάρχει αναλογική σχέση ανάμεσα στην αύξηση της μέσης υπερετήσιας τιμής (mm/d) και στην αύξηση της τυπικής απόκλισης του δείγματος των μέσων ημερήσιων τιμών (mm/d).



Σχήμα 5.8 Διάγραμμα της τυπικής απόκλισης της μέσης ημερήσιας τιμής έτους (mm/d) συναρτήσει της μέσης υπερετήσιας τιμής (mm/d) κάθε σταθμού.



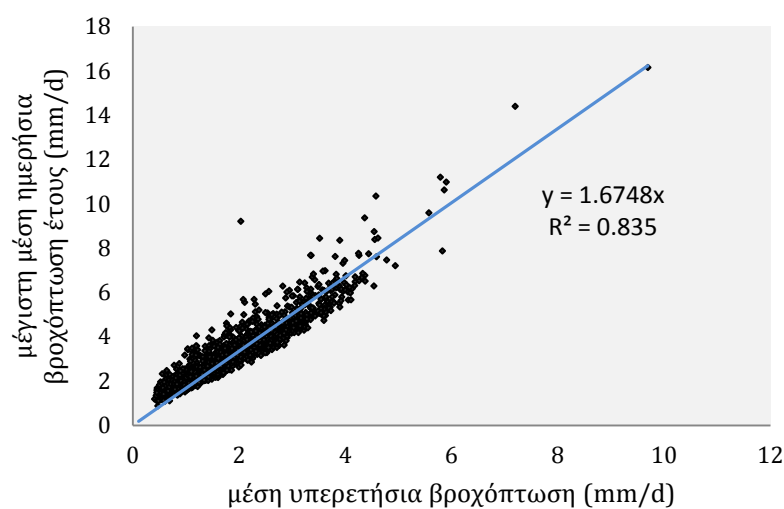
Σχήμα 5.9 Κατανομή σταθμών για ακραία μεγέθη: α) ελάχιστη τιμή της μέσης ημερήσιας τιμής έτους (mm/d), β) μέγιστη τιμή της μέσης ημερήσιας τιμής έτους (mm/d)

Στο σχήμα 5.9 φαίνεται το σημαντικά μεγαλύτερο εύρος που παρουσιάζει η κατανομή της μέγιστης τιμής σε σύγκριση με την ελάχιστη. Περίπου στο 50% των σταθμών έχει παρατηρηθεί ελάχιστη μέση ημερήσια τιμή μικρότερη του 0,85 mm/d, δηλαδή περίπου 310mm το έτος. Σε 80 σταθμούς έχουν παρατηρηθεί τιμές μηδενικές ή πολύ κοντά στο 0, δηλαδή ολική ξηρασία κατά το έτος, ενώ σε λιγότερους η ελάχιστη τιμή που παρατηρήθηκε είναι αρκετά μεγαλύτερη της μέσης υπερετήσιας του συνόλου των σταθμών. Η μικρότερη ελάχιστη τιμή που παρατηρήθηκε ισούται με 0 mm/d, δηλαδή ολική ξηρασία κατά το έτος και ανήκει στο σταθμό Hay (Illilawa) στην Αυστραλία, ενώ η μεγαλύτερη ελάχιστη τιμή παρατηρήθηκε στο σταθμό Innisfail στην Αυστραλία, στον οποίο έχει παρατηρηθεί επίσης και η μέγιστη μέση υπερετήσια βροχόπτωση.

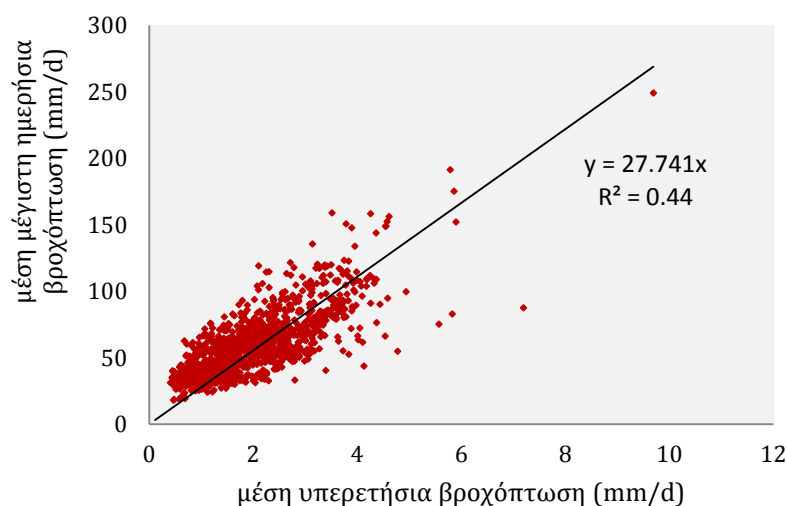
Η μέγιστη τιμή έχει πολύ μεγαλύτερη διασπορά αποτελεσμάτων όπως φαίνεται και στον πίνακα. Το 95% των παρατηρήσεων ανήκει στο διάστημα 1,46 – 6,77 mm/d, ενώ σε ένα σταθμό έχει παρατηρηθεί βροχόπτωση μεγαλύτερη των 16mm/d, δηλαδή 5844 mm το έτος. Πρόκειται για το σταθμό Innisfail στην Αυστραλία που έχει ήδη σχολιαστεί. Ο σταθμός με τη μικρότερη μέγιστη τιμή είναι ο σταθμός Sunnyside που ανήκει στις Η.Π.Α και στον οποίο παρατηρήθηκε μέγιστη μέση ημερήσια βροχόπτωση ίση με 0,88 mm/d.

Στο διάγραμμα 5.10 φαίνεται η αναλογική σχέση που υπάρχει ανάμεσα στη μέγιστη μέση ημερήσια τιμή έτους (χαρακτηριστική της μέγιστης ετήσιας

τιμής) και στην μέση υπερετήσια τιμή του σταθμού. Στο διάγραμμα 5.11 φαίνεται επίσης και η σχέση που υπάρχει ανάμεσα στη μέση μέγιστη ημερήσια τιμή έτους και στη μέση υπερετήσια τιμή. Για το σκοπό αυτό υπολογίστηκε για κάθε σταθμό η μέση τιμή όλων των μέγιστων ημερήσιων τιμών (και όχι μέσων ημερήσιων έτους) που παρατηρήθηκαν σε κάθε έτος. Φαίνεται εξάλλου και η αύξηση της αβεβαιότητας της σχέσης καθώς επιχειρείται η προσέγγιση της τιμής σε επίπεδο ημέρας.



Σχήμα 5.10 Διάγραμμα της μέγιστης μέσης ημερήσιας τιμής έτους (mm/d) συναρτήσει της μέσης υπερετήσιας τιμής (mm/d) κάθε σταθμού.



Σχήμα 5.11 Διάγραμμα της μέσης μέγιστης ημερήσιας τιμής έτους (mm/d) συναρτήσει της μέσης υπερετήσιας τιμής (mm/d) κάθε σταθμού.

5.4 Εφαρμογή μεθόδου συναθροισμένης διασποράς

Υπάρχουν αρκετές μέθοδοι για την εκτίμηση του συντελεστή Hurst, οι κυριότερες από τις οποίες περιγράφονται στο κεφάλαιο 4. Στη συγκεκριμένη μελέτη, επιλέχθηκε η μέθοδος της συναθροισμένης διασποράς λόγω μιας σειράς πλεονεκτημάτων. Είναι εύκολα κατανοητή και επιτρέπει καλύτερη αντίληψη της συμπεριφοράς εμμονής και των συνεπειών που αυτή έχει. Είναι απλή στην προσομοίωση και το γεγονός ότι εμπλέκει μία μόνο παράμετρο, την τυπική απόκλιση, καθιστά δυνατή την πιθανοτική διερεύνηση των εννοιών που χρησιμοποιεί και επομένως τη δημιουργία ενός πλαισίου στατιστικής ανάλυσης. Σε αυτή τη θεώρηση βασίστηκε εξάλλου και η διερεύνηση των επιδράσεων που έχει στα στατιστικά χαρακτηριστικά (3.3). Διατηρείται έτσι μια συνέπεια στη μελέτη ανάμεσα στη θεωρητική ανάλυση του φαινομένου και στα στατιστικά εργαλεία που χρησιμοποιήθηκαν για την ανάλυση των αποτελεσμάτων. Το μειονέκτημα της αρνητικής μεροληψίας που επιβάλλει στην εκτίμηση του Hurst είναι ήδη γνωστό, δίνεται όμως η δυνατότητα να ελεγχθεί η επίδραση του με τη χρήση της τροποποιημένης μεθόδου συναθροισμένης διασποράς LSSD (4.4). Η μέθοδος παράγει επίσης αποδεκτά αποτελέσματα καθώς οι εκτιμήσεις κυμαίνονται στο διάστημα (0,1). Τέλος, αναφορικά με τις μεθόδους που εμφανίζονται να πλεονεκτούν στις βιβλιογραφικές αναφορές όπως οι εκτιμήτριες του Whittle ή η κυματική ανάλυση σημειώνεται ότι οι δημοσιευμένες εργασίες και επομένως τα θετικά αποτελέσματα, αφορούν μεγάλα μήκη δείγματος άνω των 1000 τιμών. Επίσης, οι αλγόριθμοι είναι δυσκολότεροι υπολογιστικά και νοηματικά και οι υπάρχοντες αλγόριθμοι υπολογισμού (fArma, R package) δίνουν αμφισβητούμενα αποτελέσματα και απαιτείται περαιτέρω διερεύνηση της αξιοπιστίας τους.

Για την εκτίμηση του συντελεστή Hurst έγινε χρήση της μεθόδου συναθροισμένης διασποράς όπως αυτή περιγράφεται στην ενότητα 4.3. Για το σκοπό αυτό αναπτύχθηκε κώδικας στο Mathematica στον οποίο η μέθοδος τροποποιήθηκε κατάλληλα για να ληφθεί υπόψη και η παρουσία κενών τιμών. Αρχικά παρήχθη η μέση συναθροισμένη ανέλιξη (σχέση 4. 2) για κάθε κλίμακα κ με τους εξής περιορισμούς:

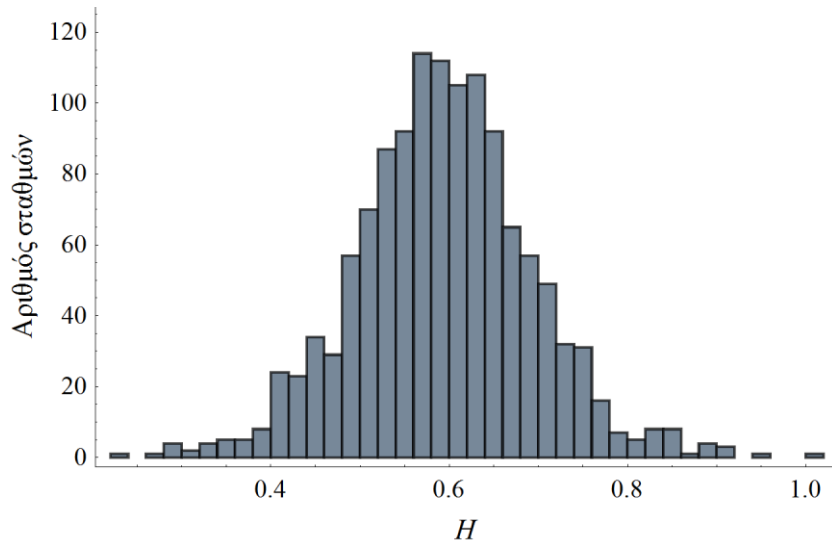
- στην κλίμακα 2 υπολογίστηκε η μέση τιμή εφόσον υπήρχαν και οι 2 ετήσιες τιμές
- για τις κλίμακες $k \geq 3$ υπολογίστηκε η μέση τιμή εφόσον υπήρχαν τουλάχιστον 3 μη κενές ετήσιες τιμές στην κλίμακα

Η διαδικασία συνεχίστηκε μέχρι τη μέγιστη κλίμακα που ορίστηκε ως ο λόγος του μήκους της χρονοσειράς προς τον επιθυμητό ελάχιστο αριθμό n στοιχείων που θα προέκυπταν στην τελευταία κλίμακα: $\max k = \text{μήκος χρονοσειράς} / n$. Μετά υπολογίστηκε η τυπική απόκλιση των στοιχείων που προέκυψαν από κάθε κλίμακα και δημιουργήθηκε το λογαριθμικό διάγραμμα της τυπικής απόκλισης συναρτήσει της κλίμακας. Η κλίση του διαγράμματος βρίσκεται προσαρμόζοντας μια ευθεία η οποία ορίζεται με πρώτο σημείο την επιλεγείσα τιμή $\min k$ στον άξονα των τεταγμένων. Ο συντελεστής προκύπτει από την κλίση του διαγράμματος προσαυξημένη κατά 1: $H = (\text{κλίση}) + 1$.

Για επιλογή παραμέτρων : $\min k = 1$ και $n = 10$ από τη στατιστική επεξεργασία των σταθμών προκύπτει ένα δείγμα συντελεστών Hurst με τα εξής χαρακτηριστικά:

Πίνακας 5.2 Στατιστικά χαρακτηριστικά εκτιμήτριας Hurst μεθόδου συναθροισμένης διασποράς

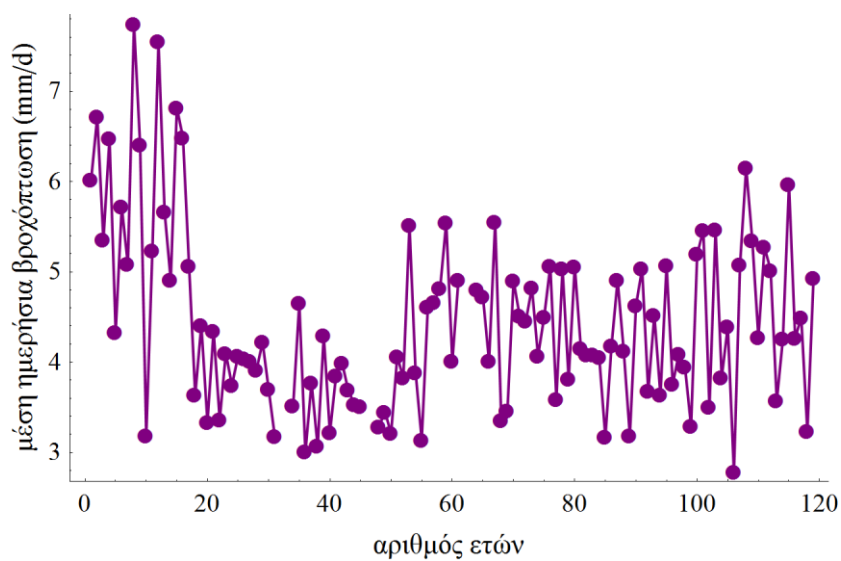
Ελάχιστη τιμή	$Q_{2,5}$	Q_{25}	Διάμεσος	Q_{75}	$Q_{97,5}$	Μέγιστη τιμή	Μέση τιμή	Τυπική απόκλιση
0,23	0,402	0,532	0,594	0,653	0,799	0,99	0,594	0,01



Σχήμα 5.12 Ιστόγραμμα κατανομής του H για τις δεδομένες τιμές των παραμέτρων

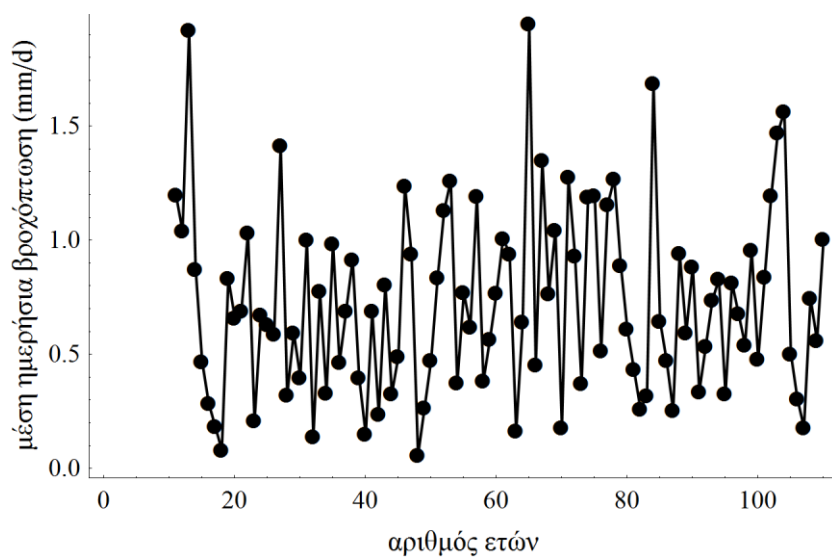
Όπως φαίνεται στον πίνακα 5.2 και στο σχήμα 5.12 οι σταθμοί παρουσιάζουν ενδείξεις ύπαρξης μακράς μνήμης καθώς το 75% των σταθμών έχει $H \geq 0,532$, μόνο το 25% όμως παρουσιάζει $H \geq 0,653$ δηλαδή σημαντική δομή αυτοσυσχέτισης. Συγκεκριμένα, βρέθηκε ότι το 85% των σταθμών έχει H που υπερβαίνει την τιμή 0,5 αλλά το 50% έχει $H \geq 0,594$, δηλαδή σημαντικότερη δομή αυτοσυσχέτισης. Πολύ ισχυρή δομή αυτοσυσχέτισης, $H \geq 0,799$ παρουσιάζει το 2,5% των σταθμών. Ένα ποσοστό της τάξης του 15% των σταθμών εμφανίζει ωστόσο χρονική ανεξαρτησία ως προς το φαινόμενο της βροχόπτωσης. Για διάστημα εμπιστοσύνης 95% οι τιμές κυμαίνονται από 0,4-0,8, υπάρχει δηλαδή σημαντική αβεβαιότητα στην εκτίμηση. Γενικά η κατανομή του H πλησιάζει την κανονική κατανομή με μέση τιμή 0,594 όπως φαίνεται στο σχήμα 5.5. Παρακάτω παρουσιάζονται ενδεικτικά διαγράμματα των χρονοσειρών χαρακτηριστικών σταθμών με υψηλό και χαμηλό συντελεστή H αντίστοιχα.

- Σταθμός με $H = 0,85$

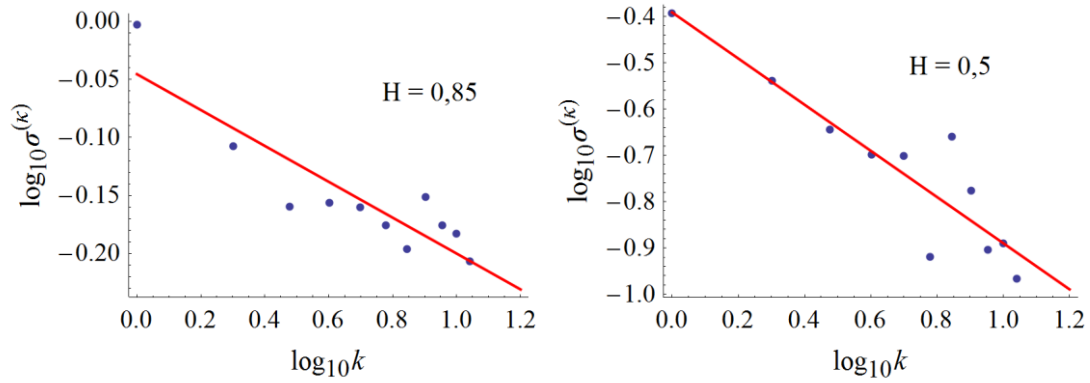


Σχήμα 5.13 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

- Σταθμός με $H = 0,50$



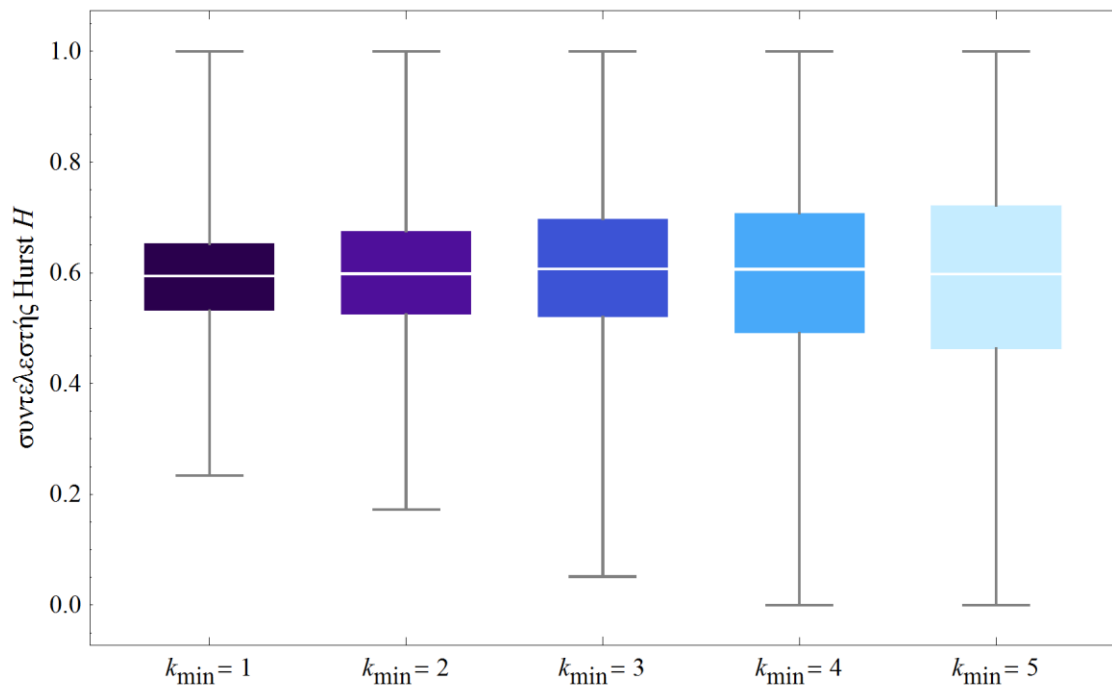
Σχήμα 5.14 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με χαμηλό συντελεστή H .



Σχήμα 5.15 Κλιμακογράμματα των δύο σταθμών

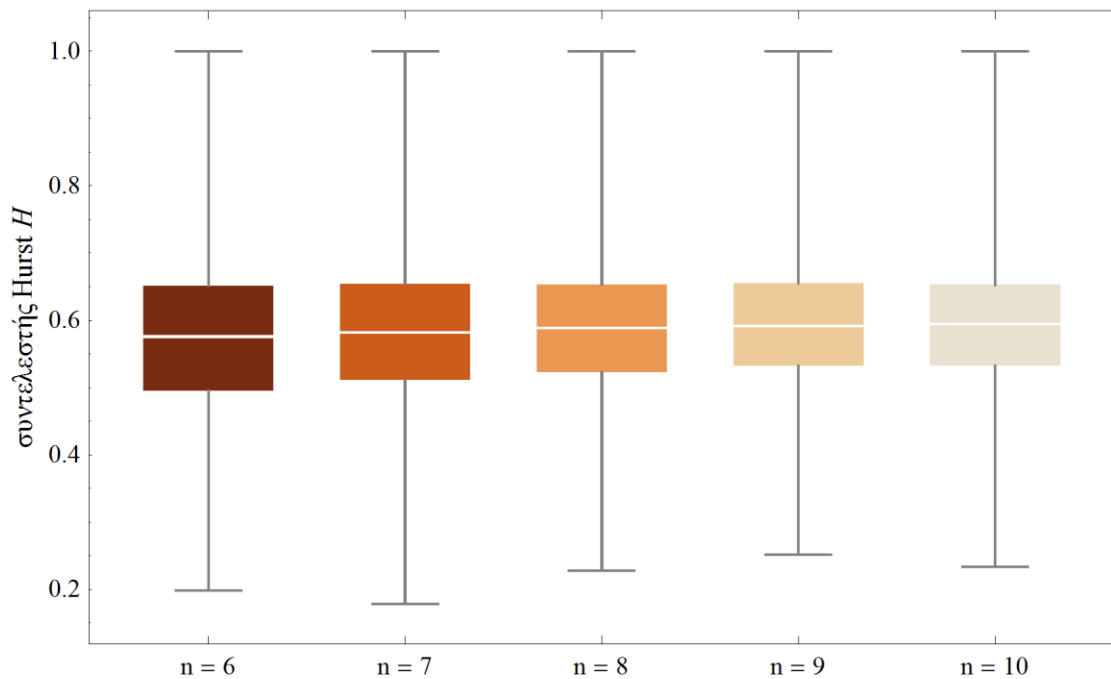
5.4.1 Επίδραση παραμέτρων στην εκτίμηση

Προκειμένου να ελεγχθεί η επίδραση της επιλογής της ελάχιστης και της μέγιστης κλίμακας για την προσαρμογή της ευθείας από την κλίση της οποίας προκύπτει ο συντελεστής Hurst πραγματοποιούνται δοκιμές για διάφορες λογικές τιμές αυτών των μεγεθών. Για την περίπτωση της μέγιστης κλίμακας χρησιμοποιείται ο ελάχιστος αριθμός των στοιχείων που επιλέγουμε να προκύψουν σε αυτήν : $\max \kappa = \text{μήκος χρονοσειράς} / n$.



Σχήμα 5.16 Box-plots για την κατανομή του H για τιμές της ελάχιστης κλίμακας 1 - 5.

Όπως φαίνεται στο σχήμα 5.16 ορθώς επιλέχτηκε ελάχιστη κλίμακα 1 στους υπολογισμούς καθώς όσο αυξάνεται η τιμή της ελάχιστης κλίμακας η διάμεσος παραμένει η ίδια, αυξάνεται όμως η διασπορά και άρα η αβεβαιότητα της εκτίμησης. Το γεγονός επίσης ότι η τιμή της διαμέσου δεν διαφοροποιείται με την «αφαίρεση» των πρώτων κλιμάκων αποτελεί στοιχείο υπέρ της ύπαρξης μακροπρόθεσμης εμμονής καθώς το κλιμακόγραμμα της Markov δομής εμφανίζει σημείο αλλαγής της κλίσης της ευθείας μετά τις πρώτες κλίμακες, όπως φαίνεται και στο σχήμα 4.3.



Σχήμα 5.17 Box-plots για την κατανομή του H για τιμές του ελάχιστου αριθμού των στοιχείων της μέγιστης κλίμακας 6 – 10.

Φαίνεται από το σχήμα 5.17 ότι όσο μειώνεται ο αριθμός των στοιχείων στη μέγιστη κλίμακα αυξάνεται η διασπορά της εκτίμησης άρα και η αβεβαιότητα της ενώ η διάμεσος παραμένει σταθερή. Άρα και σε αυτήν την περίπτωση η επιλογή της τιμής $n = 10$ οδηγεί σε περισσότερο αξιόπιστο αποτέλεσμα.

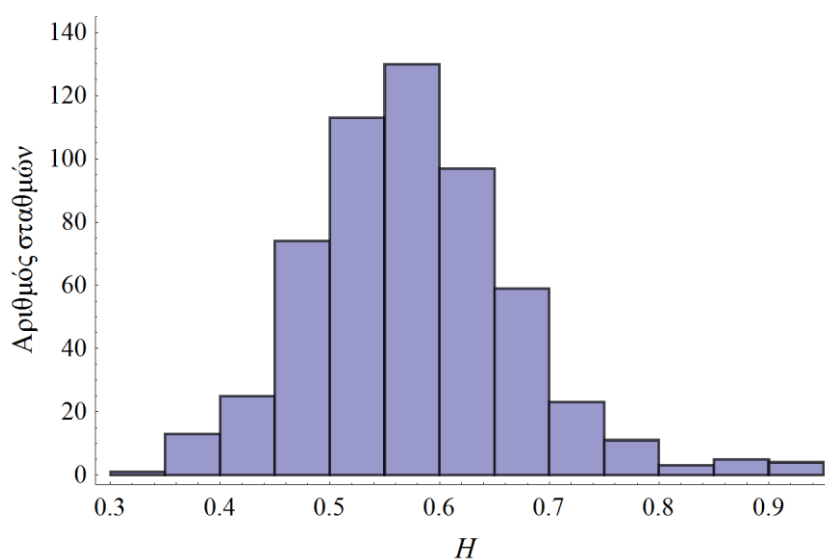
5.5 Εφαρμογή μεθόδου ελαχίστων τετραγώνων βασισμένη στην τυπική απόκλιση (LSSD, Least Squares based on Standard Deviation)

Για την εκτίμηση του Hurst έγινε επίσης εφαρμογή της μεθόδου 4.4. που βασίζεται στο κλιμακόγραμμα και εκτιμά ταυτόχρονα το συντελεστή Hurst και τη διασπορά με σκοπό τη μείωση της μεροληψίας. Χρησιμοποιήθηκε κώδικας στη Visual Basic (Tyralis and Koutsoyiannis 2011). Ο κώδικας της μεθόδου δεν είχε πρόβλεψη κενών τιμών και για αυτό δοκιμάστηκαν αρχικά οι 558 χρονοσειρές του δείγματος που είχαν πάνω από 100 συνεχόμενες τιμές χωρίς κενά και οι οποίες αποτελούν το 44% του δείγματος των χρονοσειρών, ποσοστό

επαρκές για την εξαγωγή συμπεράσματος. Τα αποτελέσματα φαίνονται παρακάτω.

Πίνακας 5.3 Στατιστικά χαρακτηριστικά εκτιμήτριας Hurst μεθόδου LSSD

Ελάχιστη τιμή	$Q_{2,5}$	Q_{25}	Διάμεσος	Q_{75}	$Q_{97,5}$	Μέγιστη τιμή	Μέση τιμή	Τυπική απόκλιση
0,334	0,398	0,52	0,572	0,635	0,794	0,916	0,577	0,094



Σχήμα 5.18 Ιστόγραμμα κατανομής του H τροποποιημένης μεθόδου

Όπως φαίνεται στον Πίνακα 5.3 και στο σχήμα 5.18 το 75% των σταθμών έχει $H \geq 0,52$, ενώ το 25% παρουσιάζει εντονότερη δομή αυτοσυσχέτισης με $H \geq 0,635$. Συγκεκριμένα, βρέθηκε ότι το 79% των σταθμών έχει H που υπερβαίνει την τιμή 0,5 αλλά το 50% έχει $H \geq 0,572$, δηλαδή σημαντικότερη δομή αυτοσυσχέτισης. Πολύ ισχυρή δομή αυτοσυσχέτισης, $H \geq 0,794$ παρουσιάζει το 2,5% των σταθμών. Ένα ποσοστό της τάξης του 25% των σταθμών εμφανίζει

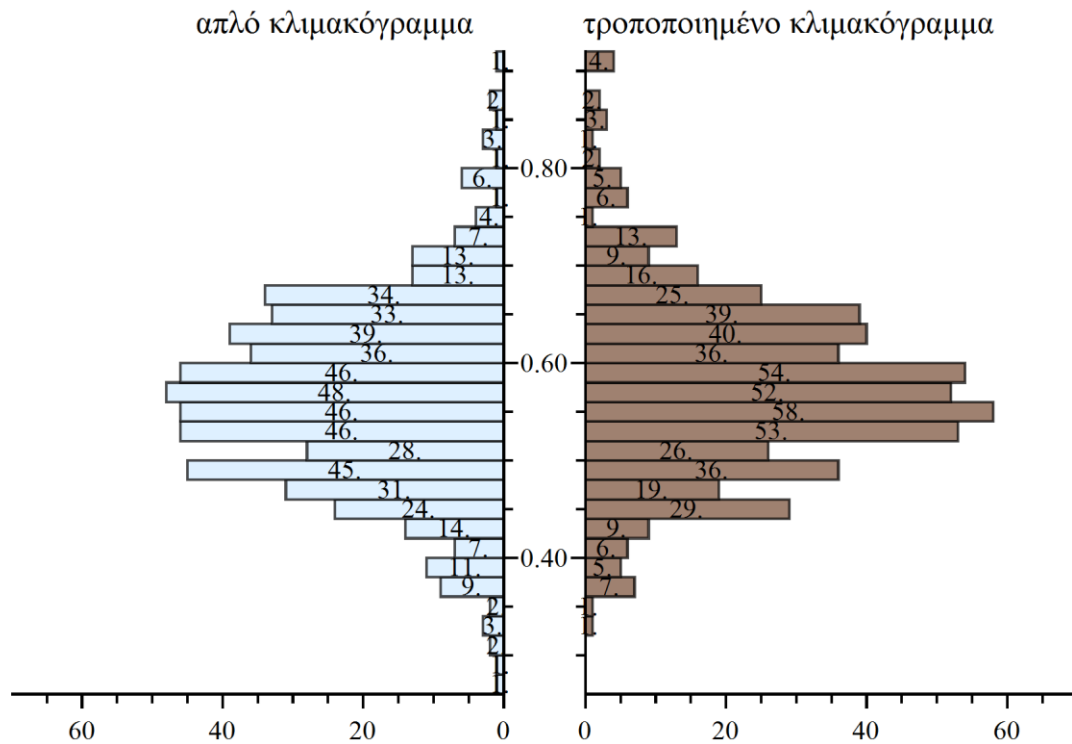
ωστόσο χρονική ανεξαρτησία ως προς το φαινόμενο της βροχόπτωσης. Για διάστημα εμπιστοσύνης 95% οι τιμές κυμαίνονται από 0,4-0,79, υπάρχει δηλαδή σημαντική αβεβαιότητα στην εκτίμηση. Για λόγους σύγκρισης επαναλήφθηκε η μέθοδος του απλού κλιμακογράμματος για τις ίδιες 558 χρονοσειρές. Τα αποτελέσματα φαίνονται στην επόμενη παράγραφο.

5.5.1 Σύγκριση των δύο μεθόδων

Ο υπολογισμός του συντελεστή Hurst επαναλήφθηκε για τις 558 χρονοσειρές με την πρώτη μέθοδο και παρουσιάζονται οι συγκριτικοί πίνακες και διαγράμματα για τα δύο αποτελέσματα.

Πίνακας 5.4 Σύγκριση στατιστικών χαρακτηριστικών δειγμάτων από τις δύο μεθόδους εκτίμησης

	Απλή μέθοδος συναθροισμένης ανέλιξης	Μέθοδος LSSD (διόρθωση μεροληψίας)
Μέση τιμή	0,564	0,576
Τυπική απόκλιση	0,099	0,094
Ελάχιστη τιμή	0,276	0,334
Q_{2,5}	0,369	0,398
Q₂₅	0,496	0,52
Διάμεσος	0,564	0,572
Q₇₅	0,63	0,635
Q_{97,5}	0,782	0,794
Μέγιστη τιμή	0,904	0,916



Σχήμα 5.19 Συνδυασμένο ιστόγραμμα κατανομής του H για τις δύο μεθόδους

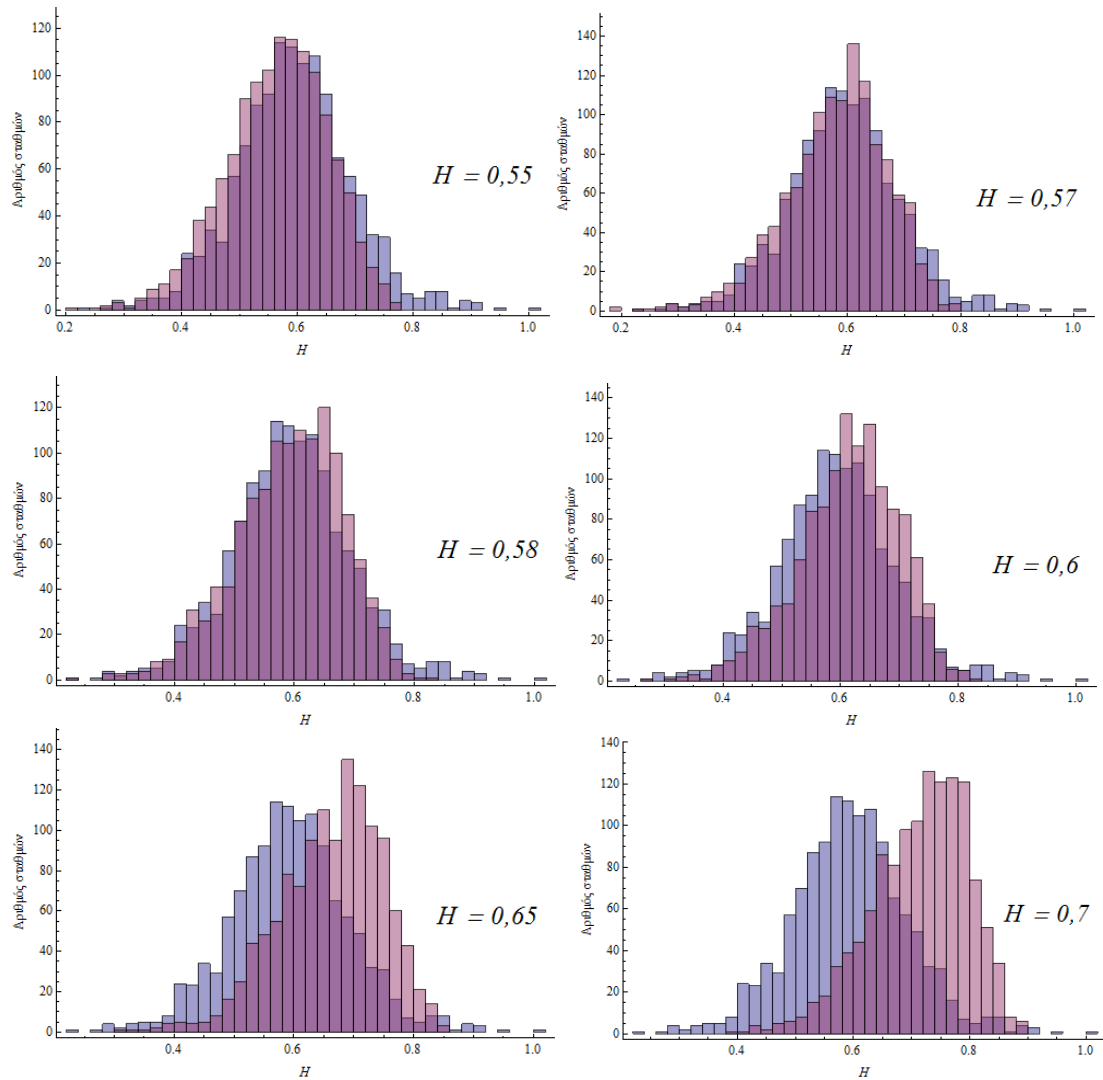
Όπως φαίνεται στον Πίνακα 5.4 και στο σχήμα 5.19 οι δύο μέθοδοι έχουν μικρές αποκλίσεις. Συγκεκριμένα φαίνεται ότι η απόκλιση της μέσης τιμής είναι περίπου 1%. Φαίνεται ότι η απλή μέθοδος εφαρμοζόμενη στα συγκεκριμένα δεδομένα υποεκτιμά ελαφρά το H καθώς οι εκτιμήσεις της είναι 1-2% μικρότερες από τις αντίστοιχες της τροποποιημένης μεθόδου για τα βασικά ποσοστημόρια της κατανομής (25%, 50%, 75%). Η τροποποιημένη μέθοδος δεν βελτίωσε ιδιαίτερα την εκτίμηση και αυτό οφείλεται στο γεγονός ότι η εκτιμώμενη τιμή του Hurst είναι σχετικά μικρή και δεν επιβάλλει μεγάλη μεροληψία. Η τροποποιημένη μέθοδος βέβαια οδηγεί σε λιγότερες εκτιμήσεις στο διάστημα $H < 0,4$ και σε κάποιες περισσότερες στο διάστημα $H > 0,75$. Δεδομένης της αβεβαιότητας της εκτίμησης και του μικρού μεγέθους της απόκλισης δεν κρίνεται απαραίτητη η επανάληψη των υπολογισμών για το σύνολο των σταθμών. Το αποτέλεσμα από την εφαρμογή της μεθόδου στο 44% των σταθμών οδηγεί στο συμπέρασμα ότι η εκτίμηση του Hurst στο σύνολο των

σταθμών με την μέθοδο 4.3 (απλό κλιμακόγραμμα) έχει αρνητική μεροληψία της τάξης 1-2%.

5.6 Έλεγχος της υπόθεσης κοινού συντελεστή Hurst

Στην παράγραφο αυτή επιχειρείται η διερεύνηση της υπόθεσης ύπαρξης ενός κοινού συντελεστή Hurst του οποίου η θεωρητική κατανομή ταιριάζει με την εμπειρική. Για το σκοπό αυτό παρήχθησαν στο Excel 1265 συνθετικές χρονοσειρές, όσες δηλαδή και οι ιστορικές, με τη μέθοδο των τυχαίων διακυμάνσεων πολλαπλής κλίμακας (Koutsoyiannis 2002). Το δείγμα των συνθετικών χρονοσειρών σχεδιάστηκε έτσι ώστε να διατηρούν τη μέση τιμή και τυπική απόκλιση του ιστορικού δείγματος, ενώ επεξεργάστηκαν ώστε να αναπαράγει κάθε μία τον ίδιο αριθμό ετών με την αντίστοιχη ιστορική. Ο θεωρητικός συντελεστής Hurst που προσεγγίζει καλύτερα την εμπειρική κατανομή βρέθηκε με τις δοκιμές που φαίνονται παρακάτω.

■ ιστορικές χρονοσειρές ■ συνθετικές χρονοσειρές



Σχήμα 5.20 Δοκιμές H για βέλτιστη ταύτιση των κατανομών H των συνθετικών και ιστορικών χρονοσειρών.

Φαίνεται (σχήμα 5.20) ότι οι θεωρητικές κατανομές για $H = 0,57$ και $H = 0,58$ προσεγγίζουν καλύτερα την εμπειρική κατανομή με την πρώτη να εμφανίζει τις μικρότερες αποκλίσεις στους μικρότερους συντελεστές H και τη δεύτερη να προσεγγίζει καλύτερα την κατανομή για τιμές $H \geq 0,7$. Βέβαια οι τιμές $H \geq 0,75$ που αποτελούν το 6% του δείγματος δεν εξηγούνται ικανοποιητικά από καμία από τις δύο κατανομές ενώ ένα 2,5% των χρονοσειρών που έχουν $H \geq 0,8$ δεν προσεγγίζεται καθόλου από καμία από τις δύο θεωρητικές. Η αύξηση του θεωρητικού συντελεστή φαίνεται επίσης να

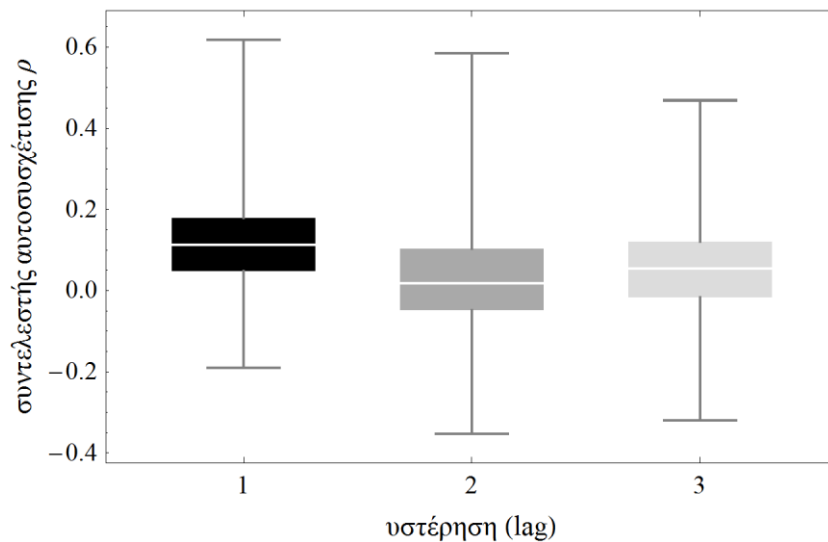
βελτιώνει ελάχιστα την προσέγγιση αυτού του τμήματος συγκριτικά με το πόσο αποκλίνει το υπόλοιπο τμήμα της κατανομής. Παρόλα αυτά η ικανοποιητική ταύτιση του υπόλοιπου μέρους της κατανομής επιτρέπει την υπόθεση ότι υπάρχει ένας κοινός συντελεστής Hurst που αφορά στην πλειοψηφία των μελετώμενων βροχοπτώσεων σε παγκόσμιο επίπεδο με ορισμένες βέβαια εξαιρέσεις που αποκλίνουν από την κατανομή του και εγείρουν κάποιες επιφυλάξεις ως προς την καταλληλότητά του.

5.7 Έλεγχος δομής αυτοσυσχέτισης

Επειδή ο συντελεστής Hurst που υπολογίστηκε δεν είναι αρκετά υψηλός για το σύνολο των σταθμών κρίνεται χρήσιμη η εκτίμηση των τριών πρώτων αυτοσυσχετίσεων για την πληρέστερη διερεύνηση της δομής της αυτοσυσχέτισης. Παρουσιάζονται παρακάτω τα αποτελέσματα της εκτίμησης:

Πίνακας 5.5 Στατιστικά χαρακτηριστικά 3 πρώτων αυτοσυσχετίσεων

Στατιστικά χαρακτηριστικά	Αυτοσυσχέτιση 1 ^{ης} τάξεως	Αυτοσυσχέτιση 2 ^{ης} τάξεως	Αυτοσυσχέτιση 3 ^{ης} τάξεως
Μέση τιμή	0,119	0,032	0,054
Τυπική απόκλιση	0,113	0,116	0,109
Ελάχιστη τιμή	-0,191	-0,353	-0,319
$Q_{2,5}$	-0,095	-0,163	-0,154
Q_{25}	0,048	-0,048	-0,017
Q_{50}	0,113	0,018	0,054
Q_{75}	0,179	0,103	0,121
$Q_{97,5}$	0,37	0,286	0,273
Μέγιστη τιμή	0,618	0,585	0,469

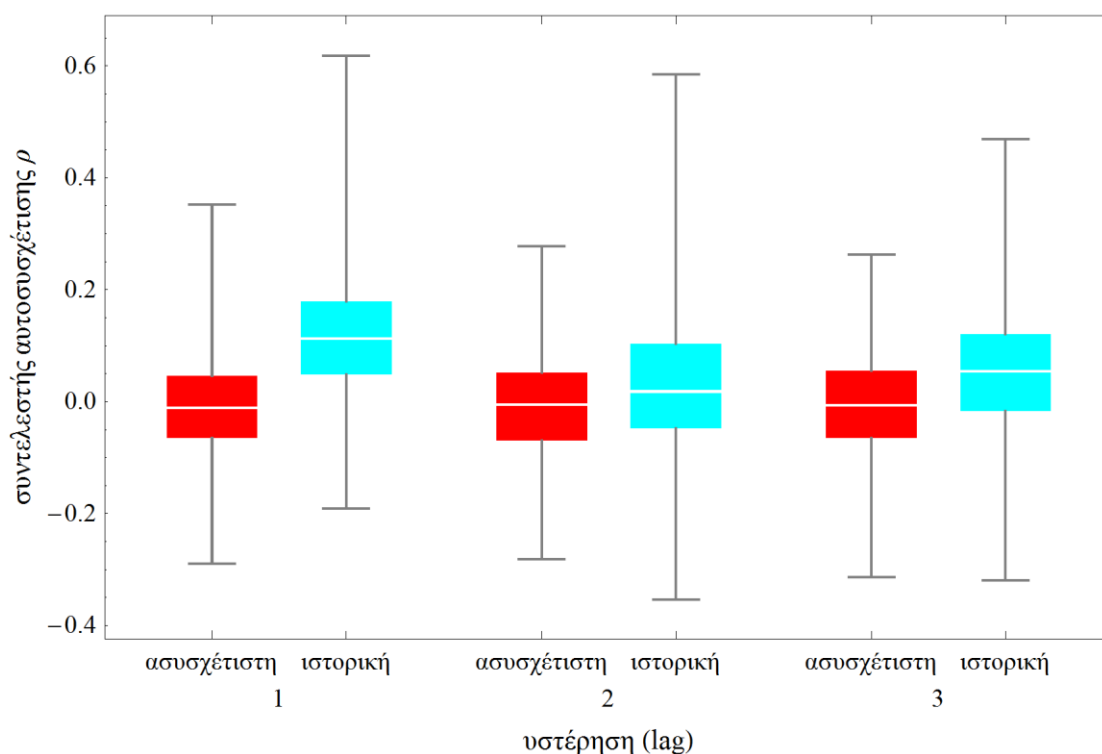


Σχήμα 5.21 Box-plots για την κατανομή των τριών πρώτων αυτοσυσχετίσεων.

Όπως φαίνεται στον πίνακα 5.5 και στο σχήμα 5.21 οι αυτοσυσχετίσεις είναι αρκετά χαμηλές και φθίνουν με την αυξανόμενη τάξη όπως ήταν αναμενόμενο. Συγκεκριμένα, φαίνεται ότι το 50% μόνο των αυτοσυσχετίσεων πρώτης τάξης είναι πάνω από την τιμή 0,113 ενώ για διάστημα εμπιστοσύνης 95% οι τιμές κυμαίνονται από -0,095 έως 0,37. Οι αντίστοιχες τιμές για δεύτερη και τρίτη τάξη είναι μικρότερες όπως φαίνεται. Είναι αξιοσημείωτο όμως το γεγονός ότι η αυτοσυσχέτιση 3^{ης} τάξης εμφανίζεται μεγαλύτερη ως προς τη μέση τιμή από την αυτοσυσχέτιση 2^{ης} τάξης. Η τυπική απόκλιση είναι περίπου η ίδια για τα τρία μεγέθη. Σημειώνεται πάντως ότι η εκτίμηση της αυτοσυσχέτισης ενέχει σημαντική αρνητική μεροληψία όπως εξηγήθηκε στην ενότητα 3.3 και άρα οι πραγματικές τιμές αναμένονται μεγαλύτερες από αυτές που εκτιμήθηκαν.

❖ *Σύγκριση με αποτελέσματα τυχαίας κατανομής*

Λόγω των χαμηλών τιμών στην κατανομή της αυτοσυσχέτισης και της ύπαρξης πλήθους αρνητικών τιμών διερευνάται αρχικά η σχέση με την κατανομή της αυτοσυσχέτισης που θα προέκυπτε από 1265 τυχαία δείγματα κανονικής κατανομής μήκους 100-180 τιμών (όσων και των χρόνων της ιστορικής κατανομής). Τα αποτελέσματα φαίνονται στο σχήμα:



Σχήμα 5.22 Box-plots για σύγκριση της ιστορικής κατανομής με κατανομή αυτοσυσχέτισης ασυσχέτιστων δεδομένων

Σημειώνεται ότι δεν έχει νόημα ο υπολογισμός της κατανομής της αυτοσυσχέτισης 2^{ης} και 3^{ης} τάξεως σε ασυσχέτιστα (τυχαία) δεδομένα και για αυτό όπως παρατηρούμε δεν αλλάζει η μορφή της κατανομής της σε σχέση με την 1^η τάξη. Παρόλα αυτά χρησιμοποιείται στο διάγραμμα για την καλύτερη εποπτεία της σχέσης των αυτοσυσχετίσεων 2^{ης} και 3^{ης} τάξης με την κατανομή αυτοσυσχέτισης τυχαίων δεδομένων.

Από το σχήμα φαίνεται ότι σε κάθε περίπτωση η διάμεσος για τα ιστορικά δεδομένα έχει μεγαλύτερη τιμή από την αντίστοιχη για τυχαία δεδομένα. Πιο συγκεκριμένα η κατανομή της αυτοσυσχέτισης 1^η τάξεως έχει τη μεγαλύτερη απόκλιση συγκριτικά με τις 2 άλλες από την τυχαία κατανομή. Για διάστημα εμπιστοσύνης 95% οι αυτοσυσχετίσεις των τυχαίων δειγμάτων παίρνουν τιμές στο διάστημα $-0,175 - 0,173$ ενώ για το ίδιο διάστημα εμπιστοσύνης οι τιμές των ιστορικών δεδομένων κυμαίνονται από $-0,09 - 0,37$. Το εύρος επίσης παρουσιάζει σημαντική διαφορά αφού είναι αρκετά μεγαλύτερο προς τις

θετικές παρατηρήσεις στην ιστορική κατανομή με μέγιστη τιμή 0,62 σε αντιδιαστολή με την μέγιστη τιμή 0,29 της τυχαίας χρονοσειράς. Για τη δεύτερη και την τρίτη τάξη αυτοσυσχέτισης παρατηρούμε επίσης αποκλίσεις αν και μικρότερες. Παρόλα αυτά οι κατανομές και στις δύο περιπτώσεις είναι μετατοπισμένες προς τις θετικές τιμές, σε σχέση με τα ασυσχέτιστα δεδομένα.

❖ *Σύγκριση με θεωρητική δομή AR(1)*

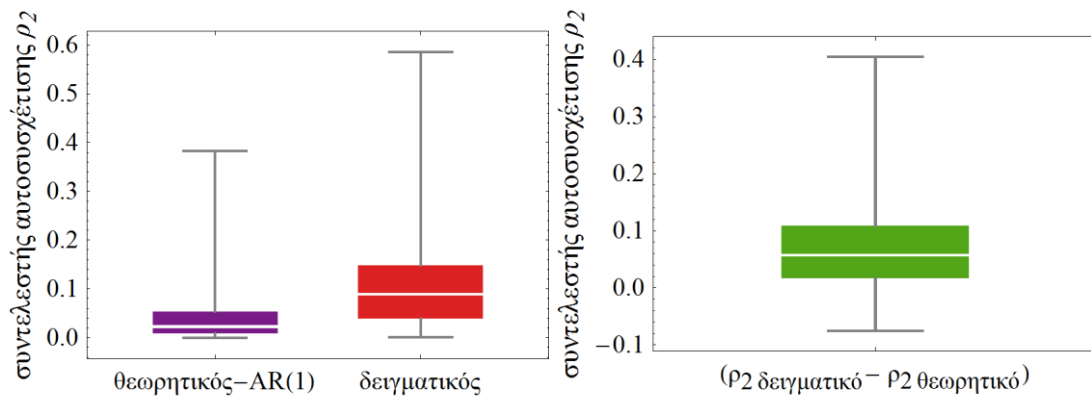
Προκειμένου να ελεγχθεί η υπόθεση δομής Markov στα δεδομένα χρησιμοποιείται για όσες αυτοσυσχετίσεις μόνο ήταν θετικές το εξής στατιστικό τεστ:

- Υπολογίζεται η θεωρητική αυτοσυσχέτιση 2^{ης} τάξεως με βάση τη δομή Markov ως : $\rho_2 = \rho_1^2$ με ρ_1 την τιμή που προέκυψε από τα ιστορικά δεδομένα
- Συγκρίνεται η θεωρητική τιμή ρ_2 που προέκυψε με την ήδη υπολογισμένη τιμή εμπειρική τιμή ρ_2
- Αν η εμπειρική τιμή είναι μεγαλύτερη από τη θεωρητική τότε, δεδομένης και της αρνητικής μεροληψίας στον υπολογισμό της εμπειρικής τιμής, απορρίπτεται η υπόθεση δομής Markov. Τα δεδομένα εμφανίζουν δηλαδή μια ισχυρότερη δομή αυτοσυσχέτισης.

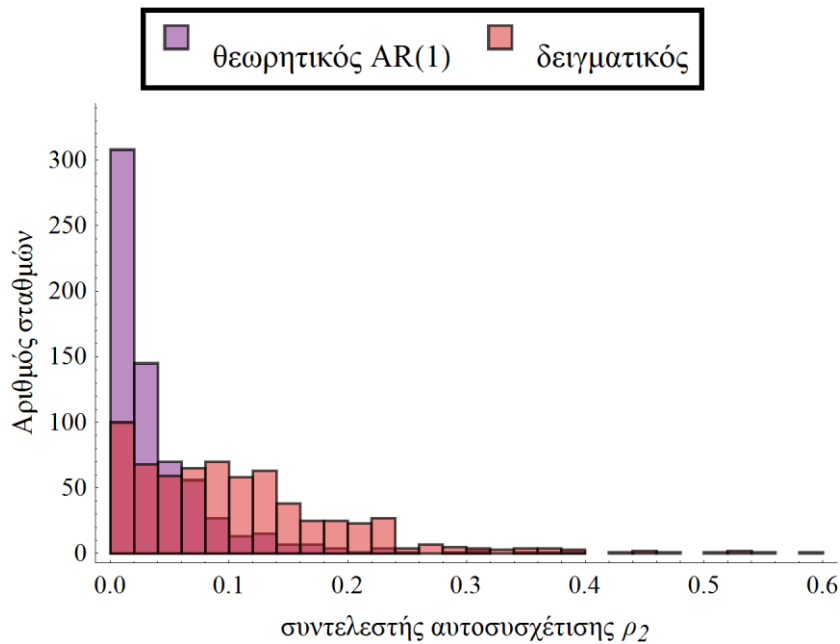
Παρακάτω δίνονται τα αποτελέσματα της ανάλυσης από την πραγματοποίηση του στατιστικού τεστ για τη σχέση του θεωρητικού και του δειγματικού συντελεστή αυτοσυσχέτισης 2^{ης} τάξης. Σημειώνεται ότι τα αποτελέσματα αφορούν μόνο τους σταθμούς για τους οποίους οι αντίστοιχες αυτοσυσχετίσεις ήταν θετικές που αποτελούν περίπου το 52% των σταθμών. Στους υπόλοιπους σταθμούς που η αυτοσυσχέτιση 2^{ης} ή 3^{ης} τάξης ήταν αρνητική δεν είχε νόημα η σύγκριση με AR(1) δομή.

Πίνακας 5.6 Στατιστική σύγκριση των δύο συντελεστών

	Θεωρητικός συντελεστής αυτοσυσχέτισης 2 ^{ης} τάξης	Δειγματικός συντελεστής αυτοσυσχέτισης 2 ^{ης} τάξης
Μέση τιμή	0,04	0,108
Τυπική απόκλιση	0,051	0,092
Ελάχιστη τιμή	0,000	0,001
Q_{2.5}	0,000	0,004
Q₂₅	0,009	0,039
Διάμεσος	0,023	0,09
Q₇₅	0,054	0,148
Q_{97.5}	0,17	0,358
Μέγιστη τιμή	0,382	0,585



Σχήμα 5.23 Box-plots για τη σύγκριση της κατανομής του θεωρητικού συντελεστή αυτοσυσχέτισης 2^{ης} τάξης με βάση την δομή AR(1) και του δειγματικού συντελεστή από την ανάλυση των δεδομένων



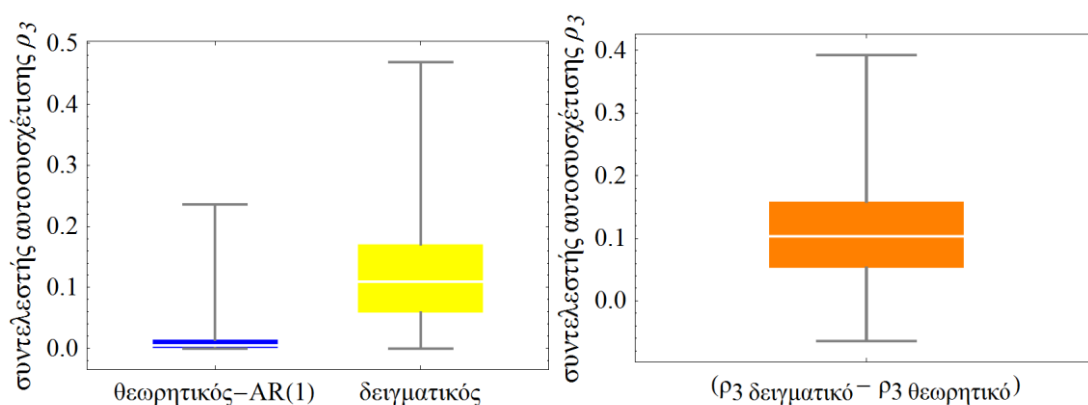
Σχήμα 5.24 Ιστόγραμμα σύγκρισης της κατανομής του θεωρητικού συντελεστή αυτοσυσχέτισης 2^{ης} τάξης με βάση την AR(1) δομή και του δειγματικού συντελεστή από την ανάλυση των δεδομένων

Όπως φαίνεται στον πίνακα 5.6 και στο σχήμα 5.23 ο συντελεστής αυτοσυσχέτισης 2^{ης} τάξης, αν και χαμηλός, είναι σημαντικά μεγαλύτερος από αυτόν που θα προέκυπτε από μία δομή AR(1). Η μέση τιμή του πρώτου είναι 0,108 ενώ του θεωρητικού μόλις 0,04 και σημαντικές διαφορές φαίνονται και στις τιμές των διαστημάτων εμπιστοσύνης. Στο διάγραμμα της διαφοράς τους φαίνεται εξάλλου ότι στην πλειονότητα των σταθμών (περίπου 85%) η διαφορά είναι θετική, δηλαδή ο δειγματικός συντελεστής υπερβαίνει τον θεωρητικά αναμενόμενο της AR(1) δομής. Στο σχήμα 5.24 φαίνεται πιο εποπτικά η σύγκριση των δύο κατανομών και παρατηρείται εύκολα ότι η πλειοψηφία των εκτιμήσεων (75%) με βάση την AR(1) δομή συγκεντρώνεται στο εύρος 0 - 0,05, γεγονός αναμενόμενο λόγω των ήδη μικρών σχετικά αυτοσυσχετίσεων 1^{ης} τάξης με βάση τις οποίες εκτιμήθηκαν. Οι εμπειρικές όμως τιμές κατανέμονται διαφορετικά σε μεγαλύτερο εύρος με το 75% των τιμών να συγκεντρώνεται στο εύρος 0- 0,15.

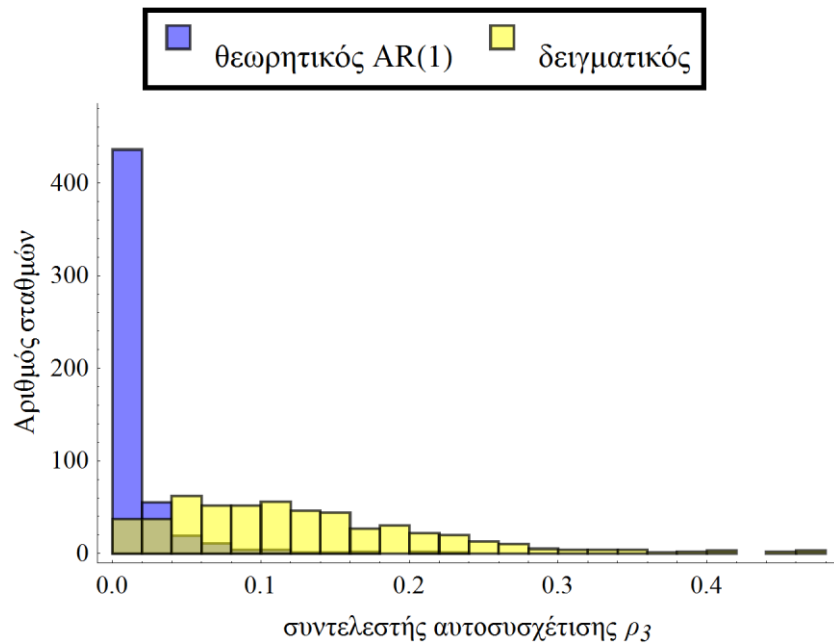
Παρακάτω ακολουθούν τα αποτελέσματα της αντίστοιχης ανάλυσης για από την πραγματοποίηση του στατιστικού τεστ για τη σχέση του θεωρητικού και του δειγματικού συντελεστή αυτοσυσχέτισης 3^{ης} τάξης.

Πίνακας 5.7 Στατιστική σύγκριση των δύο συντελεστών

	Θεωρητικός συντελεστής αυτοσυσχέτισης 3 ^{ης} τάξης	Δειγματικός συντελεστής αυτοσυσχέτισης 3 ^{ης} τάξης
Μέση τιμή	0,014	0,125
Τυπική απόκλιση	0,027	0,087
Ελάχιστη τιμή	0,000	0,000
Q_{2.5}	0,000	0,009
Q₂₅	0,001	0,059
Διάμεσος	0,005	0,110
Q₇₅	0,015	0,171
Q_{97.5}	0,088	0,348
Μέγιστη τιμή	0,236	0,469



Σχήμα 5.25 Box-plots για τη σύγκριση της κατανομής του θεωρητικού συντελεστή αυτοσυσχέτισης 3^{ης} τάξης με βάση την δομή AR(1) και του δειγματικού συντελεστή από την ανάλυση των δεδομένων

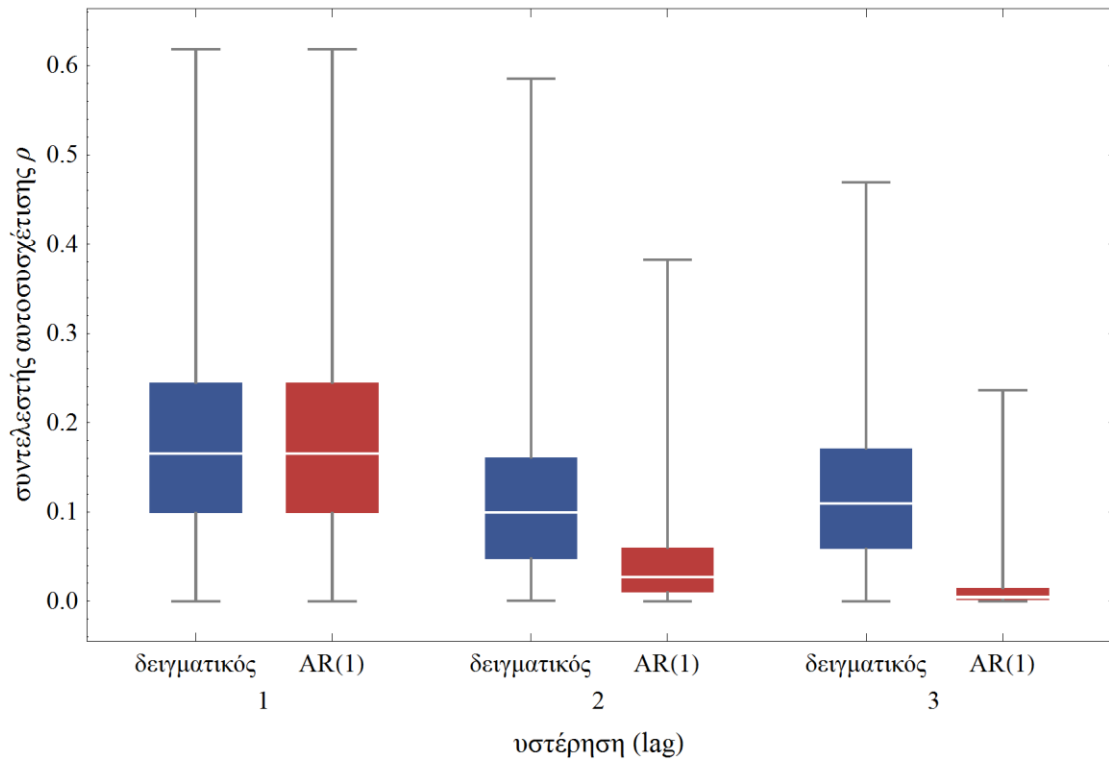


Σχήμα 5.26 Ιστόγραμμα σύγκρισης της κατανομής του θεωρητικού συντελεστή αυτοσυσχέτισης 3^{ης} τάξης με βάση την AR(1) δομή και του δειγματικού συντελεστή από την ανάλυση των δεδομένων

Όπως φαίνεται στον πίνακα 5.7 και στο σχήμα 5.25 ο συντελεστής αυτοσυσχέτισης 3^{ης} τάξης, αν και χαμηλός, είναι σημαντικά μεγαλύτερος από αυτόν που θα προέκυπτε από μία δομή AR(1). Η μέση τιμή του πρώτου είναι 0,125 ενώ του θεωρητικού μόλις 0,014 και πολύ σημαντικές διαφορές φαίνονται και στις τιμές των διαστημάτων εμπιστοσύνης. Στο διάγραμμα της διαφοράς τους φαίνεται εξάλλου ότι στην σχεδόν στο σύνολο των σταθμών (97%) η διαφορά είναι θετική, δηλαδή ο δειγματικός συντελεστής υπερβαίνει τον θεωρητικά αναμενόμενο της AR(1) δομής. Στο σχήμα 5.26 φαίνεται πιο εποπτικά η σύγκριση των δύο κατανομών και η διαφορά είναι περισσότερο έντονη και από την περίπτωση του συντελεστή 2^{ης} τάξης. Παρατηρείται εύκολα ότι η πλειοψηφία των εκτιμήσεων (75%) με βάση την AR(1) δομή συγκεντρώνεται στο εύρος 0 - 0.015, γεγονός αναμενόμενο λόγω των ήδη μικρών σχετικά αυτοσυσχετίσεων 2^{ης} τάξης με βάση τις οποίες εκτιμήθηκαν. Οι εμπειρικές όμως τιμές κατανέμονται διαφορετικά, σε αρκετά μεγαλύτερο εύρος, με το 75% των τιμών να συγκεντρώνεται στο διάστημα 0- 0,17. Φαίνεται στο διπλό ιστόγραμμα η πολύ διαφορετική εικόνα των δύο κατανομών με την

εμπειρική να διασπείρεται προς γενικά μεγαλύτερες τιμές και την θεωρητική να συγκεντρώνεται κοντά σε μηδενικές.

Από την ανάλυση και των δύο περιπτώσεων προκύπτει γενικά ότι η εμπειρική δομή αυτοσυσχέτισης διαφοροποιείται σημαντικά από την δομή AR(1). Αυτό παρουσιάζεται εποπτικά στο σχήμα 5.27 για την περίπτωση που και οι 3 συντελεστές αυτοσυσχέτισης είναι θετικοί. Συνυπολογίζοντας μάλιστα το γεγονός ότι η εμπειρικά εκτιμώμενη αυτοσυσχέτιση παρουσιάζει αρνητική μεροληψία όπως αναλύθηκε στην ενότητα 3.3, το αποτέλεσμα είναι σημαντικότερο καθώς η απόκλιση είναι θεωρητικά ακόμα μεγαλύτερη. Αξίζει να σημειωθεί επίσης ότι η αυτοσυσχέτιση 3^{ης} τάξης εμφανίζεται στο ίδιο επίπεδο με αυτή της 2^{ης} τάξης, γεγονός τελείως αντίθετο από τη συμπεριφορά ενός AR(1) μοντέλου όπου η αυτοσυσχέτιση φθίνει εκθετικά με την τάξη. Η διατήρηση των τιμών της αυτοσυσχέτισης σε ίδια επίπεδα παρά την αύξηση της υστέρησης μπορεί να θεωρηθεί γενικά ένδειξη μακροπρόθεσμης εμμονής.



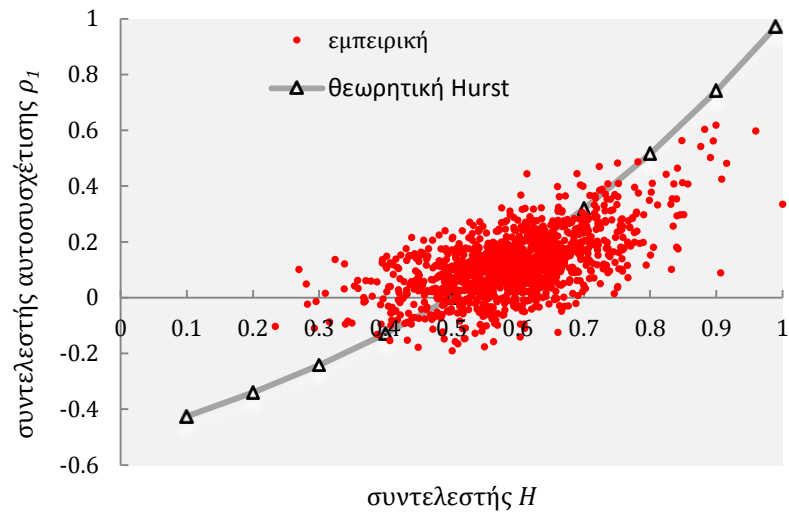
Σχήμα 5.27 Box-plots για τη σύγκριση της κατανομής της δειγματικής αυτοσυσχέτισης των 3^{ων} πρώτων τάξεων με την κατανομή της θεωρητικής AR(1) δομής.

❖ *Σύγκριση με θεωρητική αυτοσυσχέτιση μοντέλου απλής ομοιοθεσίας*

Για μοντέλο απλής ομοιοθεσίας η θεωρητική αυτοσυσχέτιση είναι ανεξάρτητη της κλίμακας συνάθροισης k και προκύπτει ως εξής (σχέση 3.14):

$$\rho_j^{(k)} = \rho_j = \frac{1}{2} [(j+1)^{2H} + (j-1)^{2H}] - j^{2H}, \quad j > 0$$

Με βάση τη σχέση 3.14 προκύπτουν οι θεωρητικές τιμές της αυτοσυσχέτισης 1^{ης} τάξης για διαδοχικές τιμές του συντελεστή Hurst H , όπως φαίνονται στο σχήμα 5.28. Στο ίδιο σχήμα παρουσιάζονται τα εμπειρικά ζεύγη παρατηρήσεων αυτοσυσχέτισης 1^{ης} τάξης και συντελεστή Hurst H . Λαμβάνοντας υπόψη τη σημαντική αρνητική μεροληψία που παρουσιάζει η εκτίμηση της αυτοσυσχέτισης, ιδίως για τιμές $H > 0,8$, φαίνεται ότι η δομή της αυτοσυσχέτισης είναι συνεπής με τη θεωρητική δομή ενός μοντέλου FGN.



Σχήμα 5.28 Διάγραμμα συντελεστή αυτοσυσχέτισης ρ_1 συναρτήσει της τιμής του συντελεστή H . Σύγκριση με τις θεωρητικές τιμές μοντέλου FGN.

6. Συμπεράσματα

6.1 Σύνοψη μελέτης και αποτελεσμάτων

Σκοπός της παρούσας διπλωματικής ήταν η επεξεργασία μιας παγκόσμιας βάσης δεδομένων βροχόπτωσης για τη διερεύνηση της ύπαρξης ή όχι του φαινομένου Hurst ή αλλιώς μακροπρόθεσμης εμμονής, στην ετήσια βροχόπτωση. Χρησιμοποιήθηκαν ημερήσια δεδομένα από το Παγκόσμιο Ιστορικό Δίκτυο Κλιματολογίας – Ημερήσια Βάση Δεδομένων (Global Historical Climatology Network, version 2.60) και συγκεκριμένα, 3477 σταθμοί οι οποίοι πληρούσαν κάποια αρχικά κριτήρια ποιότητας και καταγραφών. Για όσα έτη είχαν πλήρεις τιμές άνω των 345 ημερών εξήχθη η μέση τιμή και στη συνέχεια ακολουθήθηκε μια απλή μέθοδος συμπλήρωσης ετών. Σε δεύτερη φάση επεξεργασίας μετά την αρχική συμπλήρωση επιλέχθηκαν 1265 τελικοί σταθμοί βάσει κριτηρίων καταλληλότητας που ορίστηκαν για τον υπολογισμό του Hurst. Για την πιθανή αξία που έχει η μελέτη μιας μεγάλης βάσης δεδομένων παγκόσμιας κλίμακας, το δείγμα των σταθμών αναλύθηκε επιπρόσθετα και ως προς τα στατιστικά χαρακτηριστικά της ετήσιας βροχόπτωσης (μέση τιμή, διάμεσος, τυπική απόκλιση, ασυμμετρία, ακραίες τιμές) και τα αποτελέσματα παρουσιάζονται σε πίνακες και διαγράμματα του κεφαλαίου 5.

Για τον υπολογισμό του συντελεστή Hurst επιλέχθηκε η μέθοδος του κλιμακογράμματος (συναθροισμένης διασποράς), η οποία αναπτύχθηκε αλγοριθμικά στο Wolfram Mathematica με πρόβλεψη για την περίπτωση κενών τιμών. Η ανάλυση του συνόλου των σταθμών έδειξε ότι υπάρχουν ενδείξεις μακροπρόθεσμης εμμονής καθώς το 85% των σταθμών έχει $H > 0.5$. Παρόλα αυτά η κατανομή του H στο σύνολο των σταθμών είναι τέτοια ώστε μόνο το 50% να έχει $H > 0,59$, δηλαδή σημαντικότερη δομή αυτοσυσχέτισης, ενώ το 25% μόνο παρουσιάζει $H > 0,65$. Το 2,5% των σταθμών παρουσίασε πολύ έντονη δομή με $H > 0,8$. Για διάστημα εμπιστοσύνης 95% οι τιμές κυμαίνονται στο εύρος 0,4-0,8, γεγονός που δείχνει αφενός τη μεγάλη διασπορά των αποτελεσμάτων, αφετέρου όμως το γεγονός ότι ασυμμετρία γύρω από την τιμή $H = 0,5$ είναι θετική.

Η ανάλυση της επίδρασης της παραμέτρου ελάχιστης κλίμακας για την προσαρμογή της ευθείας συνηγορεί επίσης υπέρ της ύπαρξης εμμονής για λόγους που αναλύθηκαν στην ενότητα 5.4.

Εφαρμόστηκε επίσης η μέθοδος LSSD στο 44% των σταθμών που είχαν συνεχόμενα χρόνια χωρίς κενά. Η μέθοδος υπολογίζει αμερόληπτα το συντελεστή Hurst, διορθώνοντας τη μεροληψία που προκύπτει από την κλασική εκτίμηση της τυπικής απόκλισης. Για λόγους σύγκρισης οι δύο μέθοδοι εφαρμόστηκαν στο ίδιο δείγμα δεδομένων και βρέθηκε ότι στο συγκεκριμένο δείγμα η απλή μέθοδος υποεκτιμά το συντελεστή κατά 1-2%. Η μικρή αυτή μεροληψία είναι αναμενόμενη και στην εκτίμηση του Hurst στο σύνολο των σταθμών.

Ελέγχθηκε η υπόθεση ενός κοινού συντελεστή Hurst του οποίου η θεωρητική κατανομή ταιριάζει με την δειγματική με παραγωγή συνθετικών χρονοσειρών. Βρέθηκε ότι η καλύτερη προσαρμογή επιτυγχάνεται με θεωρητικό συντελεστή 0,58, παρόλο που αποκλείεται κατά αυτόν τον τρόπο το ποσοστό των σταθμών με $H > 0,8$ (2,5%).

Τέλος, διερευνήθηκε η δομή της αυτοσυσχέτισης ως σημαντικό εργαλείο για την πληρέστερη μελέτη της δομής εξάρτησης των δεδομένων. Έχοντας υπόψη τη σημαντική αρνητική μεροληψία που ενέχει η κλασική εκτίμηση της αυτοσυσχέτισης, εκτιμήθηκαν οι αυτοσυσχετίσεις των τριών πρώτων τάξεων που αν και μικρές διαφοροποιούνται από την κατανομή αυτοσυσχέτισης τυχαίων δεδομένων. Για όσους σταθμούς είχαν θετικές και τις 3 τιμές αυτοσυσχετίσεων, δηλαδή για το 52% των σταθμών, πραγματοποιήθηκε στατιστικό τεστ για τη σύγκριση με τη δομή AR(1) και βρέθηκε ότι τα δεδομένα παρουσιάζουν ισχυρότερη δομή εξάρτησης και άρα η υπόθεση Markov απορρίπτεται. Η σχέση της εμπειρικής αυτοσυσχέτισης με την εμπειρική εκτίμηση του H διερευνήθηκε και φαίνεται ότι προσεγγίζει αρκετά καλά (δεδομένης και της αρνητικής μεροληψίας στην εκτίμηση της αυτοσυσχέτισης) τη θεωρητική σχέση που υπάρχει σε ένα μοντέλο FGN.

6.2 Συμπεράσματα

Από την ανάλυση της μεγάλης βάσης δεδομένων βροχόπτωσης προκύπτει ότι υπάρχουν ενδείξεις ύπαρξης μακροπρόθεσμης εμμονής στην ετήσια βροχόπτωση σε παγκόσμια κλίμακα, όπως αναλύθηκε προηγουμένως. Βέβαια ένα ποσοστό της τάξης του 15% εμφανίζει χρονική ανεξαρτησία και ένα 15% εμφανίζει ιδιαίτερα χαμηλές τιμές του συντελεστή και δεν μπορεί να αποκλειστεί το ενδεχόμενο χρονικής ανεξαρτησίας. Ο θεωρητικός συντελεστής 0,58 φαίνεται ότι αντιπροσωπεύει την κατανομή στην πλειονότητα της εκτός ορισμένων υψηλών τιμών. Η δομή της αυτοσυσχέτισης (αν και εκτιμημένης με αρνητική μεροληψία) είναι συνεπής προς τα αποτελέσματα της εκτίμησης αφού στο 52% των περιπτώσεων απορρίπτεται η δομή Markov, ενώ στο σύνολο παρουσιάζει συνέπεια με το μοντέλο FGN.

Για την πληρέστερη κατανόηση του προβλήματος και των αποτελεσμάτων είναι αναγκαίο να ληφθούν υπόψη ορισμένες παράμετροι αβεβαιότητας. Η σημαντικότερη είναι ίσως το μήκος των καταγραφών το οποίο αν και μεγάλο σε σύγκριση με τα συνήθη δείγματα δεδομένων, δεν είναι απόλυτα ικανοποιητικό για την διερεύνηση μακροπρόθεσμης εμμονής, ενός φαινομένου η δομή του οποίου αναπτύσσεται στο χρόνο. Η χαμηλή τιμή επομένως που εντοπίζεται σε πολλούς σταθμούς είναι χαρακτηριστική μόνο για το διάστημα των περίπου 100 χρόνων που μελετήθηκαν και η χρονική ανεξαρτησία που παρατηρήθηκε είναι πιθανό να οφείλεται σε κάποιες περιπτώσεις στην έλλειψη δεδομένων. Επίσης, είναι σημαντική η αβεβαιότητα η οποία οφείλεται σε λάθη μετρήσεων ή ακόμα και σε επεξεργασία μετρήσεων σε επίπεδο σταθμού για λόγους ομογενοποίησης των δεδομένων. Παρότι η επεξεργασία των δεδομένων είναι δυνατό να οδηγήσει σε διαστρέβλωση της πραγματικότητας, αποτελούσε κατά το παρελθόν ίσως και σήμερα συνήθη τακτική ορισμένων υπηρεσιών για τη «βελτίωση» των μετρήσεών τους. Επιπρόσθετα, η βάση δεδομένων δεν περιλαμβάνει πολλούς σταθμούς από ορισμένες χαρακτηριστικές περιοχές του πλανήτη όπως η Αφρική ή η Ασία, για λόγους ιστορικούς και οικονομικούς. Το γεγονός ότι οι σταθμοί με

τον υψηλότερο συντελεστή Hurst ανήκουν σε περιοχές όπως η Η.Π.Α και η Αυστραλία, οι οποίες έχουν συνεισφέρει το μεγαλύτερο ποσοστό σταθμών δημιουργεί το ερώτημα σχετικά με το αν μια διευρυμένη βάση δεδομένων οδηγούσε σε αύξηση του συντελεστή. Τέλος, η υψηλή διασπορά των αποτελεσμάτων είναι ενδεικτική της αβεβαιότητας της εκτίμησης και των μεγάλων διαφορών που εντοπίζονται ακόμα και σε επίπεδο ηπείρου.

6.3 Προτάσεις για μελλοντική έρευνα

Στη βιβλιογραφία αναφέρονται πολλές διαφορετικές μέθοδοι με πλεονεκτήματα και μειονεκτήματα για την εκτίμηση του συντελεστή Hurst. Οι περισσότερες υπάρχουν ήδη αναπτυγμένες σε διάφορα λογισμικά όπως το στατιστικό πρόγραμμα R, απαιτούν όμως μελέτη για την αξιοπιστία τους, που λόγω χρόνου δεν επιτεύχθηκε στην παρούσα διπλωματική. Παρόλα αυτά δεν αναμένεται να βελτιώσουν την εκτίμηση για λόγους που εξηγήθηκαν ήδη στο κεφάλαιο 5.

Περισσότερο σημαντική θα ήταν η επανάληψη των αναλύσεων με διαφορετικά κριτήρια επιλογής σταθμών και με έλεγχο της παραμέτρου των καταγραφών για τη διερεύνηση της επίδρασης του μήκους δείγματος στην εκτίμηση του Hurst. Θα ήταν πιθανόν σημαντική επίσης η διερεύνηση μιας άλλης βάσης δεδομένων βροχόπτωσης παγκόσμιας κλίμακας, ή και της ίδια βάσης διευρυμένης.

Τέλος, λόγω της φύσης του προβλήματος αυτό που θα είχε περισσότερη αξία θα ήταν η επανάληψη των υπολογισμών σε βάθος χρόνου, όταν θα ήταν διαθέσιμα πολύ μεγαλύτερα μήκη καταγραφών για τους σταθμούς.

7. Βιβλιογραφικές αναφορές

A.Ramachandra, and D. Bhattacharya. 1999. 'Hypothesis Testing for Long-term Memory in Hydrologic Series'. *Journal of Hydrology* 216 (3-4) (March 22): 183-196. doi:10.1016/S0022-1694(99)00005-0.

Abry, P., P. Goncalves, and P. Flandrin. 1993. 'Wavelet-based Spectral Analysis of 1/f Processes'. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference On*, 3:237-240. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=319479.

Abry, P., P. Goncalves, and P. Flandrin. 1995. 'Wavelets, Spectrum Analysis and 1/f Processes'. *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG*:-: 15-15.

Abry, P., and D. Veitch. 1998. 'Wavelet Analysis of Long-range-dependent Traffic'. *Information Theory, IEEE Transactions On* 44 (1) (January): 2 -15. doi:10.1109/18.650984.

Asur, S., B. A. Huberman, G. Szabo, and C. Wang. 2011. 'Trends in Social Media: Persistence and Decay'. In *5th International AAAI Conference on Weblogs and Social Media*.

Bakker, Alexander, and Bart van den Hurk. 2012. 'Estimation of Persistence and Trends in Geostrophic Wind Speed for the Assessment of Wind Energy Yields in Northwest Europe'. *Climate Dynamics*: 1-16. doi:10.1007/s00382-011-1248-1.

Beran, J. 1994. *Statistics for Long-memory Processes*. Vol. 61. Chapman & Hall/CRC.

Beran, J., R. Sherman, M.S. Taqqu, and W. Willinger. 1995. 'Long-range Dependence in Variable-bit-rate Video Traffic'. *Communications, IEEE Transactions On* 43 (234) (April): 1566 -1579. doi:10.1109/26.380206.

Beran, Jan. 1992. 'Statistical Methods for Data with Long-Range Dependence'. *Statistical Science* 7 (4) (November 1): 404-416.

Blender, Richard, and Klaus Fraedrich. 2006. 'Long-term Memory of the Hydrological Cycle and River Runoffs in China in a High-resolution Climate Model'. *International Journal of Climatology* 26 (12) (October 1): 1547–1565. doi:10.1002/joc.1325.

Bloomfield, P. 1992. 'Trends in Global Temperature'. *Climatic Change* 21 (1): 1–16.

Bove, R., V. Pelino, and L. De Leonibus. 2006. 'Complexity in Rainfall Phenomena'. *Communications in Nonlinear Science and Numerical Simulation* 11 (6): 678–684.

Cajueiro, Daniel O, and Benjamin M Tabak. 2004. 'The Hurst Exponent over Time: Testing the Assertion That Emerging Markets Are Becoming More Efficient'. *Physica A: Statistical Mechanics and Its Applications* 336 (3–4) (May 15): 521–537. doi:10.1016/j.physa.2003.12.031.

Carbone, A., G. Castelli, and H. E. Stanley. 2004. 'Time-dependent Hurst Exponent in Financial Time Series'. *Physica A: Statistical Mechanics and Its Applications* 344 (1): 267–271.

Daniels, J. Michael. 2007. 'Flood Hydrology of the North Platte River Headwaters in Relation to Precipitation Variability'. *Journal of Hydrology* 344 (1–2) (September 30): 70–81. doi:10.1016/j.jhydrol.2007.06.020.

Daubechies, I. 1992. *Ten Lectures on Wavelets*. Vol. 61. SIAM. <http://epubs.siam.org/doi/pdf/10.1137/1.9781611970104.fm>.

Ding, Z., C. W. J. Granger, and R. F. Engle. 1993. 'A Long Memory Property of Stock Market Returns and a New Model'. *Journal of Empirical Finance* 1 (1): 83–106.

Durre, Imke, Matthew J. Menne, Byron E. Gleason, Tamara G. Houston, and Russell S. Vose. 2010. 'Comprehensive Automated Quality Assurance of Daily Surface Observations'. *Journal of Applied Meteorology and Climatology* 49 (8): 1615–1633.

Durre, Imke, Matthew J. Menne, and Russell S. Vose. 2008. 'Strategies for Evaluating Quality Assurance Procedures'. *Journal of Applied Meteorology and Climatology* 47 (6): 1785–1791.

Eltahir, E.A.B. 1996. 'El Nino and the Natural Variability in the Flow of the Nile River'. *Water Resources Research* 32 (1): 131–137.

Evans, T. E. 1996. *The Effects of Changes in the World Hydrological Cycle on Availability of Water Resources*. Vol. 15. John Wiley & Sons, Chichester, UK.

Fraedrich, K., and R. Blender. 2003. 'Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models'. *Physical Review Letters* 90 (10): 108501.

Grech, D., and G. Pamula. 2008. 'The Local Hurst Exponent of the Financial Time Series in the Vicinity of Crashes on the Polish Stock Exchange Market'. *Physica A: Statistical Mechanics and Its Applications* 387 (16): 4299–4308.

Hanssen-Bauer, I., and E. J. Førland. 1998. 'Long-term Trends in Precipitation and Temperature in the Norwegian Arctic: Can They Be Explained by Changes in Atmospheric Circulation Patterns?' *Climate Research* 10: 143–153.

Henry, M., and P. Zaffaroni. 2002. 'The Long Range Dependence Paradigm for Macroeconomics and Finance'. *Available at SSRN*.

Higuchi, T. 1988. 'Approach to an Irregular Time Series on the Basis of the Fractal Theory'. *Physica D: Nonlinear Phenomena* 31 (2): 277–283.

Jiang, Tong, Buda Su, and Heike Hartmann. 2007. 'Temporal and Spatial Trends of Precipitation and River Flow in the Yangtze River Basin, 1961–2000'. *Geomorphology* 85 (3): 143–154.

Jones, P. D., K. R. Briffa, T. P. Barnett, and S. F. B. Tett. 1998. 'Millennial Temperature Reconstructions'. *IGBP PAGES/world Data center-A for Paleoclimatology Data Contribution Series* 39.

Kaiser, N. 1984. 'On the Spatial Correlations of Abell Clusters'. *The Astrophysical Journal* 284: L9–L12.

Karagiannis, T., M. Molle, and M. Faloutsos. 2004. 'Long-range Dependence Ten Years of Internet Traffic Modeling'. *Internet Computing, IEEE* 8 (5): 57–64.

Kendall, M. G. 1943. 'Oscillatory Movements in English Agriculture'. *Journal of the Royal Statistical Society* 106 (2) (January 1): 91–124. doi:10.2307/2980373.

Kettani, H., and J.A. Gubner. 2002. 'A Novel Approach to the Estimation of the Hurst Parameter in Self-similar Traffic'. In *Local Computer Networks, 2002. Proceedings. LCN 2002. 27th Annual IEEE Conference On*, 160 – 165. doi:10.1109/LCN.2002.1181780.

Kolmogorov, A. 1941. 'The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers'. In *Akademiia Nauk SSSR Doklady*, 30:301–305.

Koscielny-Bunde, E., J. W. Kantelhardt, P. Braun, A. Bunde, and S. Havlin. 2006. 'Long-term Persistence and Multifractality of River Runoff Records: Detrended Fluctuation Studies'. *Journal of Hydrology* 322 (1): 120–137.

Koutsoyiannis, D. 2002. 'The Hurst Phenomenon and Fractional Gaussian Noise Made Easy'. *Hydrological Sciences Journal* 47 (4): 573–595.

Koutsoyiannis, D. 2003. 'Climate Change, the Hurst Phenomenon, and Hydrological Statistics'. *Hydrological Sciences Journal* 48 (1): 3–24.

Koutsoyiannis, D. 2010. 'A Random Walk on Water'. *Hydrology and Earth System Sciences* 14: 585–601.

Koutsoyiannis, D. 2011a. 'Hurst-Kolmogorov Dynamics and Uncertainty1'. *JAWRA Journal of the American Water Resources Association* 47 (3): 481–495.

Koutsoyiannis, D. 2011b. 'Hurst-Kolmogorov Dynamics as a Result of Extremal Entropy Production'. *Physica A: Statistical Mechanics and Its Applications* 390 (8): 1424–1432.

Koutsoyiannis, D., and A. Langousis. 2011. 'Precipitation'. In *Treatise on Water Science*, edited by P. Wilderer and S. Uhlenbrook, 2:27–78. Academic. <http://www.itia.ntua.gr/en/docinfo/999/>.

Koutsoyiannis, D., and A. Montanari. 2007. 'Statistical Analysis of Hydroclimatic Time Series: Uncertainty and Insights'. *Water Resources Research* 43 (5): W05429.

Kumar, Praveen, and Efi Foufoula-Georgiou. 1997. 'Wavelet Analysis for Geophysical Applications'. *Reviews of Geophysics* 35 (4): 385–412.

Lennartz, Sabine, and Armin Bunde. 2011. 'Distribution of Natural Trends in Long-term Correlated Records: A Scaling Approach'. *Physical Review E* 84 (2): 021129.

Lipset, S. M. 1988. *Revolution and Counterrevolution: Change and Persistence in Social Structures*. Transaction Publishers.

Liu, Binhui, Ming Xu, Mark Henderson, and Ye Qi. 2005. 'Observed Trends of Precipitation Amount, Frequency, and Intensity in China, 1960–2000'. *Journal of Geophysical Research* 110 (D8): D08103.

Lux, T. 1996. 'Long-term Stochastic Dependence in Financial Prices: Evidence from the German Stock Market'. *Applied Economics Letters* 3 (11): 701–706.

Mandelbrot, Benoit B., and John W. Van Ness. 1968. 'Fractional Brownian Motions, Fractional Noises and Applications'. *SIAM Review* 10 (4): 422–437.

Matsoukas, Christos, Shafiqul Islam, and Ignacio Rodriguez-Iturbe. 2000. 'Detrended Fluctuation Analysis of Rainfall and Streamflow Time Series'. *Journal of Geophysical Research* 105 (D23): 29165–29172.

McLeod, Angus Ian, and Keith William Hipel. 1978. 'Preservation of the Rescaled Adjusted Range, 1: a Reassessment of the Hurst Phenomenon'. *Water Resources Research* 14 (3): 491–508.

Menne, Matthew J., Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston. 2012. 'An Overview of the Global Historical Climatology Network-daily Database'. *Journal of Atmospheric and Oceanic Technology* 29 (7): 897–910.

Mielniczuk, J., and P. Wojdyła. 2007. 'Estimation of Hurst Exponent Revisited'. *Computational Statistics & Data Analysis* 51 (9) (May 15): 4510–4525. doi:10.1016/j.csda.2006.07.033.

Mikosch, T., and C. Starica. 2004. 'Nonstationarities in Financial Time Series, the Long-range Dependence, and the IGARCH Effects'. *Review of Economics and Statistics* 86 (1): 378–390.

Orcutt, G. H., and J. O. Irwin. 1948. 'A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States, 1919-1932'. *Journal of the Royal Statistical Society. Series B (Methodological)* 10 (1) (January 1): 1–53.

Papalexiou, S. M., D. Koutsoyiannis, and A. Montanari. 2013. 'Mind the Bias!'

Papalexiou, Simon-Michael, Demetris Koutsoyiannis, and Alberto Montanari. 2011. 'Can a Simple Stochastic Model Generate Rich Patterns of Rainfall Events?' *Journal of Hydrology*.

Pelletier, Jon D., and Donald L. Turcotte. 1997. 'Long-range Persistence in Climatological and Hydrological Time Series: Analysis, Modeling and Application to Drought Hazard Assessment'. *Journal of Hydrology* 203 (1–4) (December 31): 198–208. doi:10.1016/S0022-1694(97)00102-9.

Peng, C. K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. 1994. 'Mosaic Organization of DNA Nucleotides'. *Physical Review E* 49 (2): 1685.

Puckridge, J.T., K.F. Walker, and J.F. Costelloe. 2000. 'Hydrological Persistence and the Ecology of Dryland Rivers'. *River Research and Applications* 16 (5): 385–402.

Rea, W., L. Oxley, M. Reale, and J. Brown. 2012. 'Not All Estimators Are Born Equal: The Empirical Properties of Some Estimators of Long Memory'. *Mathematics and Computers in Simulation*.

Rea, William, Les Oxley, Marco Reale, and Jennifer Brown. 2009. 'Estimators for Long Range Dependence: An Empirical Study'. *arXiv Preprint arXiv:0901.0762*. <http://arxiv.org/abs/0901.0762>.

Rea, William, Marco Reale, and Jennifer Brown. 2011. 'Long Memory in Temperature Reconstructions'. *Climatic Change* 107 (3): 247–265. doi:10.1007/s10584-011-0068-y.

Santer, B. D., J. S. Boyle, J. J. Hnilo, K. E. Taylor, T. M. L. Wigley, D. Nychka, D. J. Gaffen, and D. E. Parker. 2000. 'Statistical Significance of Trends and Trend Differences in Layer-average Atmospheric Temperature Time Series'. *Journal of Geophysical Research* 105 (D6): 7337–7356.

Stephenson, D. B., V. Pavan, and R. Bojariu. 2000. 'Is the North Atlantic Oscillation a Random Walk?' *International Journal of Climatology* 20 (1): 1–18.

Stoev, S., M. S. Taqqu, C. Park, and J. S. Marron. 2005. 'On the Wavelet Spectrum Diagnostic for Hurst Parameter Estimation in the Analysis of Internet Traffic'. *Computer Networks* 48 (3): 423–445.

Taqqu, M. S., and V. Teverovsky. 1997. 'Robustness of Whittle-type Estimators for Time Series with Long-range Dependence'. *Communications in Statistics. Stochastic Models* 13 (4): 723–757.

Taqqu, M. S., and V. Teverovsky. 1998. 'On Estimating the Intensity of Long-range Dependence in Finite and Infinite Variance Time Series'. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* 177: 218.

Taqqu, Murad S., Vadim Teverovsky, and Walter Willinger. 1995. *Estimators for Long-Range Dependence: An Empirical Study*.

Tyrallis, H., and D. Koutsoyiannis. 2011. 'Simultaneous Estimation of the Parameters of the Hurst–Kolmogorov Stochastic Process'. *Stochastic Environmental Research and Risk Assessment* 25 (1): 21–33.

Wanner, H., S. Brönnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, and E. Xoplaki. 2001. 'North Atlantic Oscillation–concepts and Studies'. *Surveys in Geophysics* 22 (4): 321–381.

Weron, Rafał. 2002. 'Estimating Long-range Dependence: Finite Sample Properties and Confidence Intervals'. *Physica A: Statistical Mechanics and Its Applications* 312 (1): 285–299.

Zhang, X., L. A. Vincent, W. D. Hogg, and A. Niitsoo. 2000. 'Temperature and Precipitation Trends in Canada During the 20th Century'. *Atmosphere-Ocean* 38 (3): 395–429.

Παππάς, Χ. 2010. 'Βέλτιστη συμπλήρωση ελλιπών υδρομετεωρολογικών δεδομένων με χρήση γειτονικών χρονικά παρατηρήσεων'. ΤΥΠΕ. <http://www.itia.ntua.gr/el/docinfo/1065/>.

Παράρτημα Α

Ι. Στατιστική επεξεργασία σταθμών με $H > 0.8$

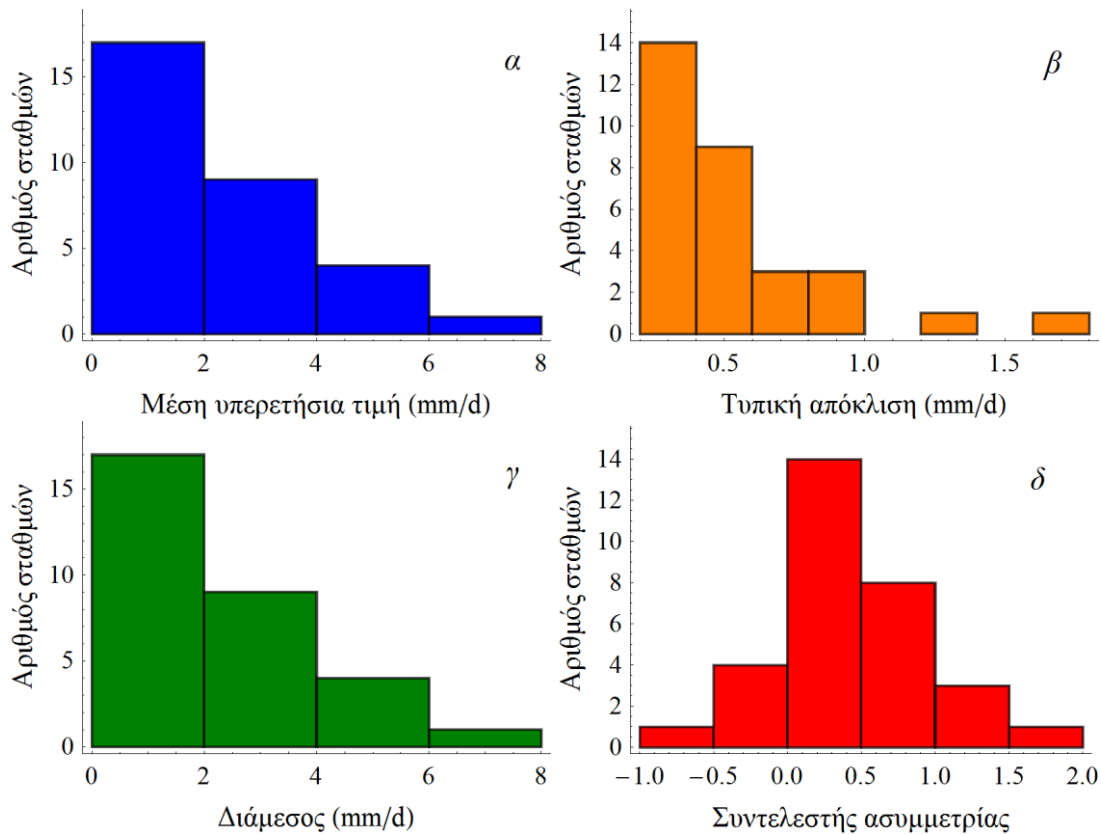
Βρέθηκαν 31 σταθμοί συνολικά με $H > 0,8$ οι οποίοι προέρχονται από τις εξής χώρες :

- 8 από Ρωσία
- 7 από Αυστραλία
- 7 από Η.Π.Α
- 4 από Καναδά
- 2 από Γερμανία
- 1 από Λιθουανία
- 1 από Αυστρία
- 1 από Ελβετία

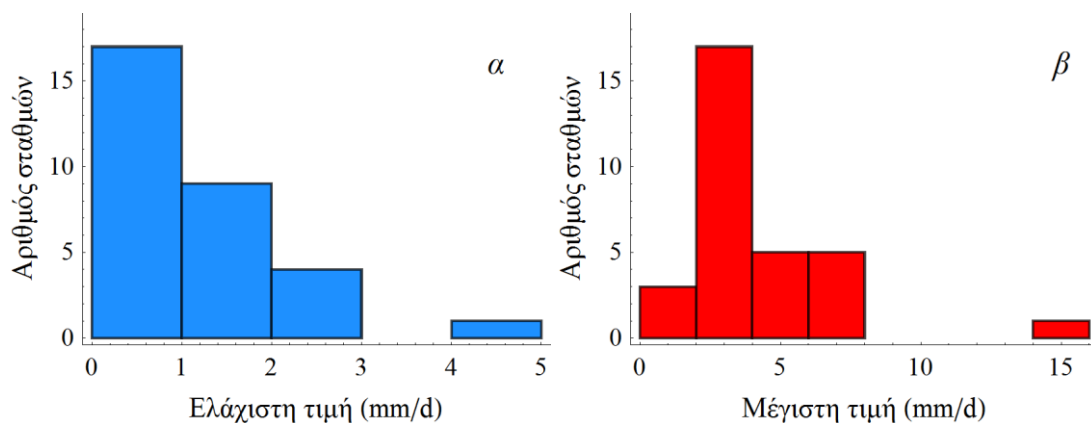
Παρακάτω δίνονται τα βασικά στατιστικά χαρακτηριστικά που προέκυψαν από την ανάλυση τους ως προς το φαινόμενο της βροχόπτωσης καθώς και τα διαγράμματα της κατανομής βασικών στατιστικών μεγεθών, κατ' αντιστοιχία της ενότητας 5.3. Σημειώνεται ότι σε αυτή την περίπτωση λόγω του υψηλού συντελεστή Hurst του δείγματος υπάρχει αρνητική μεροληψία στην εκτίμηση των μεγεθών όπως αναλύθηκε στην ενότητα 3.3. Συγκεκριμένα, η μέση τιμή είναι αμερόληπτη όχι όμως και η διασπορά της η οποία εμφανίζεται μικρότερη από την πραγματική. Αρνητική μεροληψία υπάρχει επίσης στην εκτίμηση της τυπικής απόκλισης καθώς και στην τιμή του συντελεστή ασυμμετρίας.

Πίνακας Ι. Στατιστικά χαρακτηριστικά στατιστικών μεγεθών των σταθμών

	Μέση τιμή	Τυπική απόκλιση	Διάμεσος	Συντελεστής ασυμμετρίας	Ελάχιστη τιμή	Μέγιστη τιμή
Μέση τιμή	2,324	0,535	2,3	0,43	1,228	3,975
Τυπική απόκλιση	1,42	0,314	1,41	0,543	0,933	2,576
Ελάχιστη τιμή	0,923	0,257	0,877	-0,557	0,158	1,596
Q_{2,5}	0,923	0,257	0,877	-0,557	0,158	1,596
Q₂₅	1,392	0,302	1,387	0,106	0,559	2,267
Q₅₀	1,696	0,407	1,724	0,256	0,908	3,206
Q₇₅	2,802	0,61	2,763	0,794	1,50	5,111
Q_{97,5}	7,198	1,64	7,092	1,558	4,615	14,398
Μέγιστη τιμή	7,198	1,64	7,092	1,558	4,615	14,398



Σχήμα Ι.α Κατανομή σταθμών για τα εξής στατιστικά μεγέθη: *σχήμα α*) μέση υπερετήσια τιμή (mm/d), *σχήμα β*) τυπική απόκλιση μέσης ημερήσιας τιμής έτους (mm/d), *σχήμα γ*) διάμεσος μέσης ημερήσιας τιμής έτους (mm/d), *σχήμα δ*) συντελεστής ασυμμετρίας μέσης ημερήσιας τιμής έτους



Σχήμα Ι.β Κατανομή σταθμών για ακραία μεγέθη: *α*) ελάχιστη τιμή μέσης ημερήσιας τιμής έτους, *β*) μέγιστη τιμή μέσης ημερήσιας τιμής έτους.

II. Παρουσίαση και ανάλυση των 10 σταθμών με το υψηλότερο συντελεστή Hurst

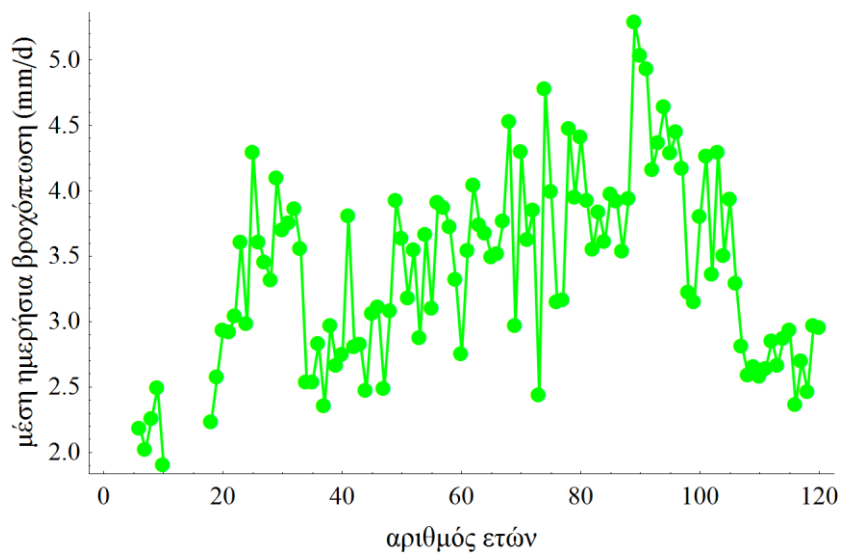
Σταθμός 1

$H = 0,959$

Κωδικός : ASN00094041

Τοποθεσία : Maatsuyker Island Lighthouse

Χώρα : Αυστραλία



Σχήμα II.1 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

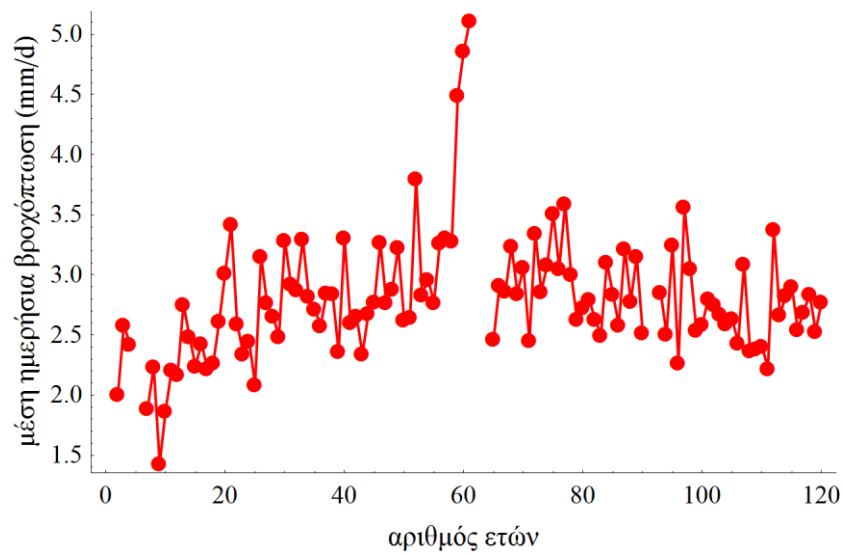
Σταθμός 2

$H = 0,915$

Κωδικός : CA001090660

Τοποθεσία : Barkerville

Χώρα : Καναδάς



Σχήμα II.2 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

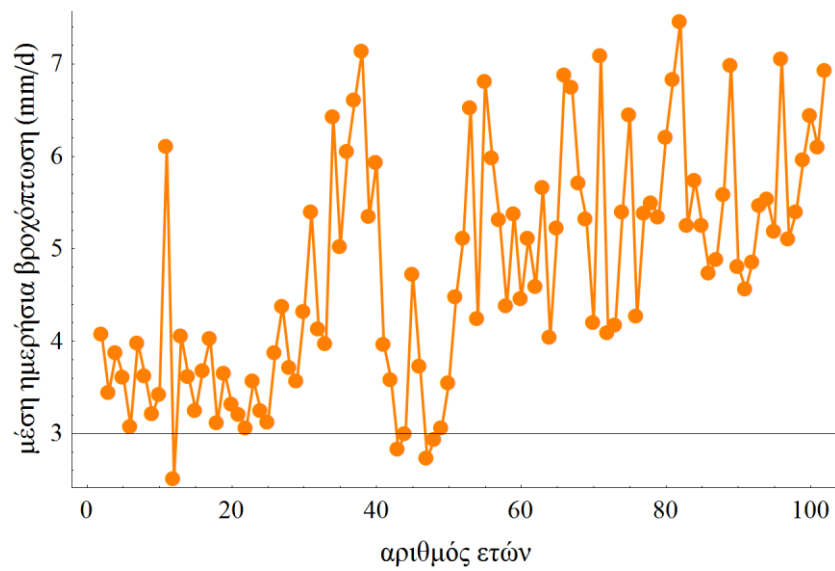
Σταθμός 3

$H = 0,895$

Κωδικός : GM000004155

Τοποθεσία : Zugspitze

Χώρα : Γερμανία



Σχήμα Π.3 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

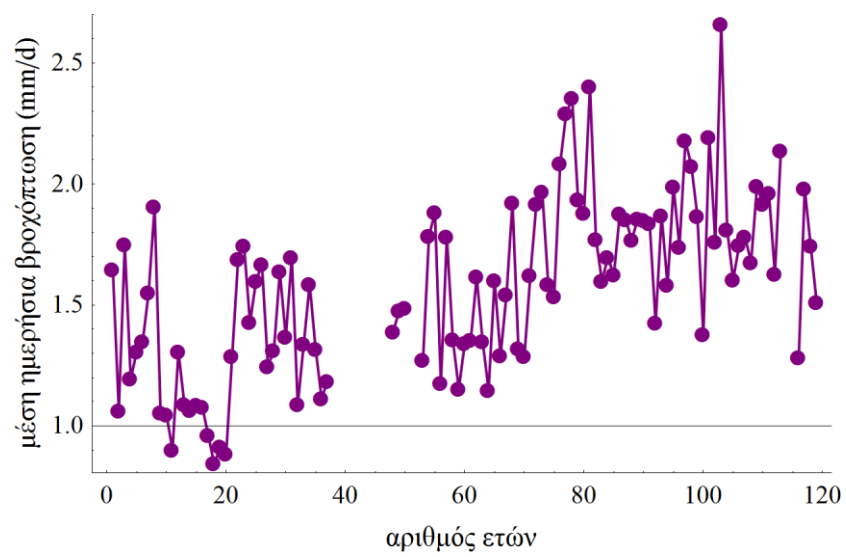
Σταθμός 4

$H = 0,876$

Κωδικός : RSM00022837

Τοποθεσία : Vytegra

Χώρα : Ρωσία



Σχήμα II.4 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

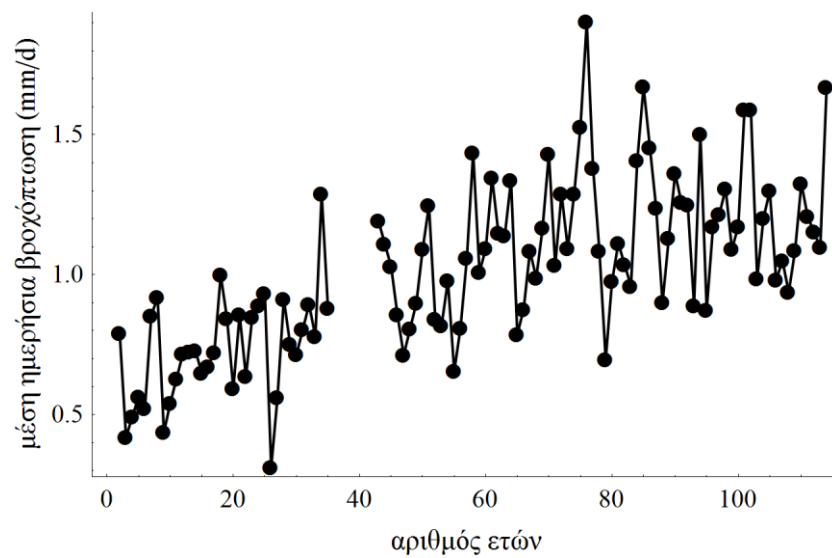
Σταθμός 5

$H = 0,882$

Κωδικός : RSM00023330

Τοποθεσία : Salehard

Χώρα : Ρωσία



Σχήμα II.5 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

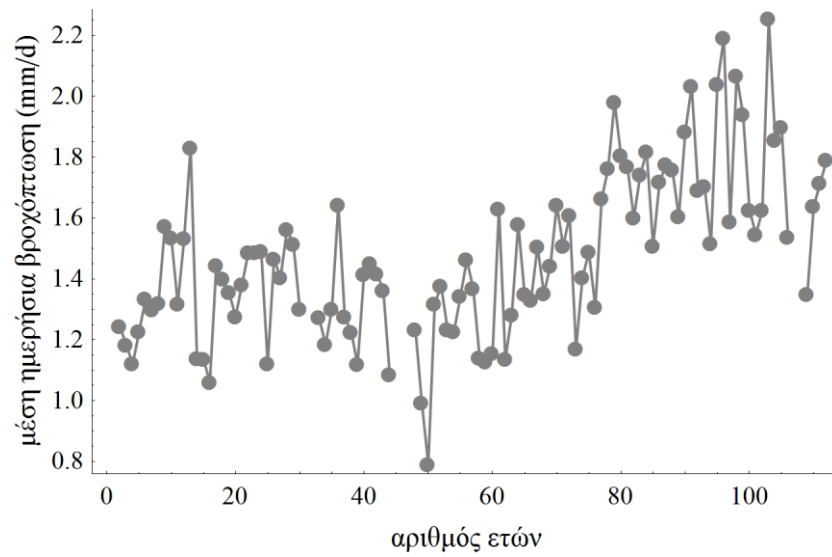
Σταθμός 6

$H = 0,899$

Κωδικός : RSM00023711

Τοποθεσία : Troicko – Pecherskoe

Χώρα : Ρωσία



Σχήμα II.6 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

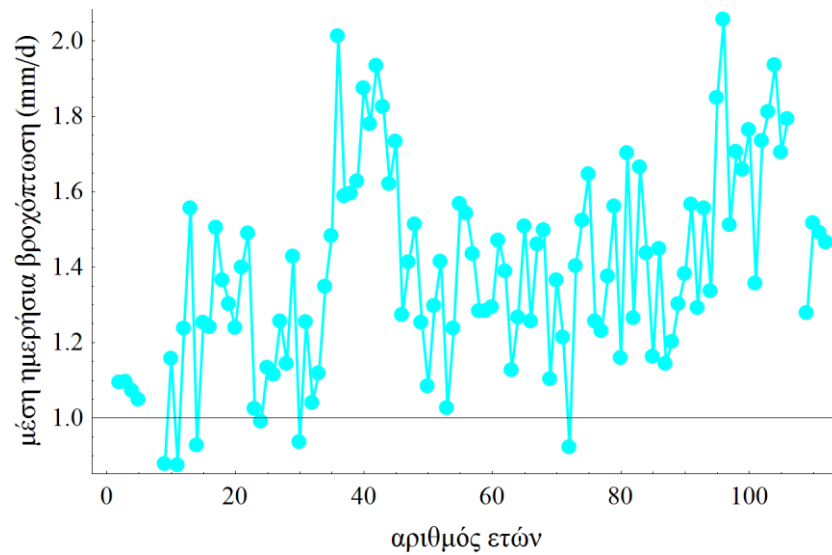
Σταθμός 7

$H = 0,891$

Κωδικός : RSM00023804

Τοποθεσία : Syktyvkar

Χώρα : Ρωσία



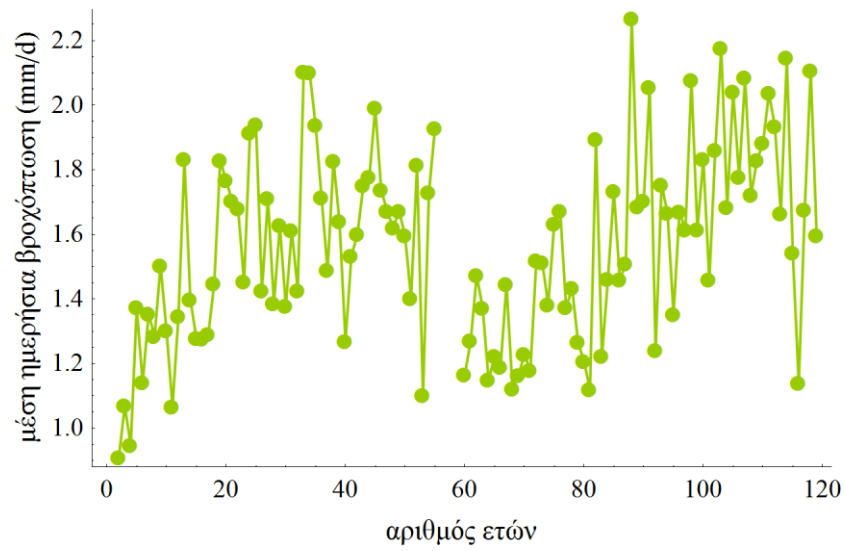
Σχήμα II.7 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

Σταθμός 8

$H = 0,857$

Κωδικός : RSM00027199

Χώρα : Ρωσία



Σχήμα II.8 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H.

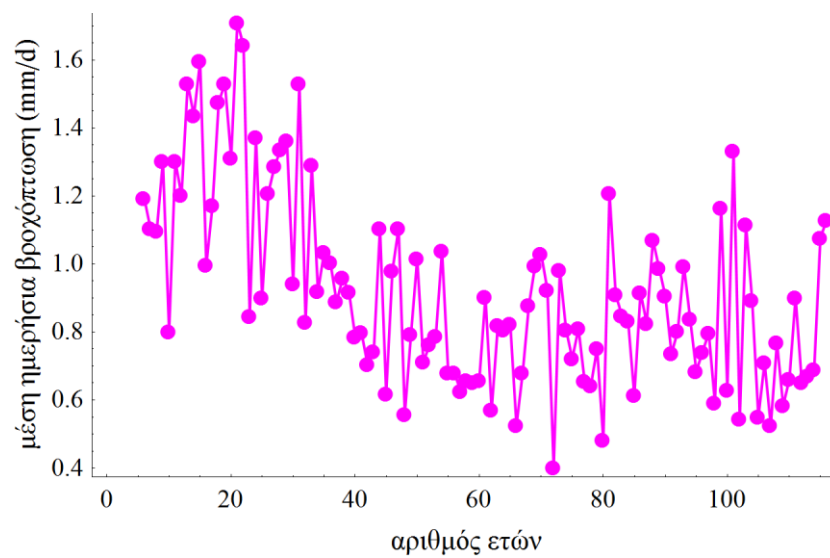
Σταθμός 9

$H = 0,908$

Κωδικός : USC00242409

Τοποθεσία : Dillion University of Montana Western

Χώρα : Η.Π.Α



Σχήμα II.9 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H.

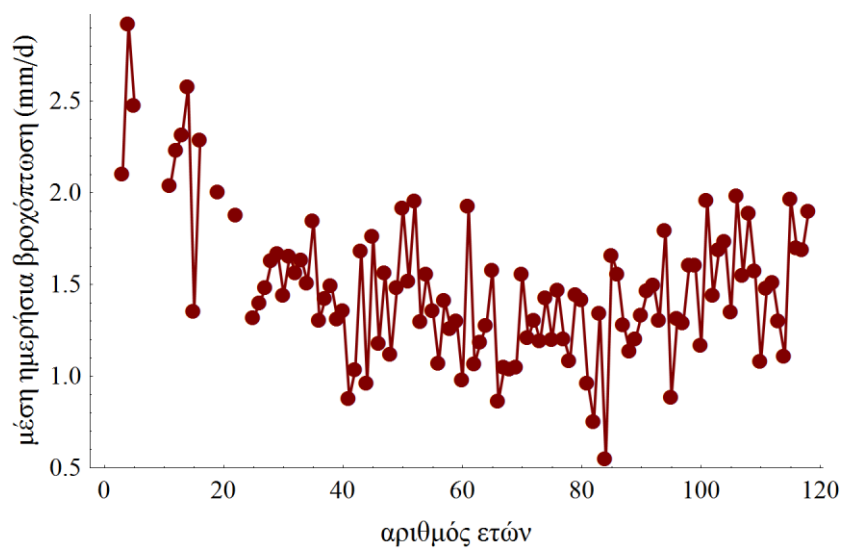
Σταθμός 10

$H = 0,99$

Κωδικός : USC00390022

Τοποθεσία : Aberdeen

Χώρα : Η.Π.Α



Σχήμα Π.10 Διάγραμμα χρονοσειράς βροχόπτωσης σταθμού με υψηλό συντελεστή H .

Πίνακας II. Στατιστικά χαρακτηριστικά 10 σταθμών με μεγαλύτερο συντελεστή *H*.

Σταθμός	1	2	3	4	5	6	7	8	9	10
Συντ/ής H	0,959	0,915	0,895	0,876	0,882	0,899	0,891	0,857	0,908	0,99
Μέση τιμή	3,397	2,802	4,778	1,577	1,012	1,479	1,392	1,57	0,925	1,485
Τυπική απόκλιση	0,716	0,52	1,254	0,366	0,302	0,273	0,262	0,301	0,283	0,393
Ελάχιστη τιμή	1,908	1,431	2,516	0,843	0,312	0,79	0,877	0,908	0,401	0,549
Q_{0.25}	2,832	2,5	3,688	1,307	0,806	1,282	1,232	1,353	0,704	1,213
Q_{0.50}	3,457	2,755	4,73	1,602	0,978	1,45	1,368	1,596	0,877	1,445
Q_{0.75}	3,914	3,051	5,59	1,851	1,208	1,638	1,558	1,766	1,104	1,672
Μέγιστη τιμή	5,292	5,112	7,464	2,659	1,903	2,254	2,058	2,267	1,709	2,923
ρ₁	0,597	0,481	0,562	0,542	0,603	0,618	0,502	0,407	0,425	0,335
ρ₂	0,525	0,421	0,517	0,524	0,445	0,469	0,457	0,362	0,585	0,39
ρ₃	0,445	0,249	0,467	0,466	0,412	0,456	0,36	0,393	0,469	0,257

Παράρτημα Β

Κώδικες Mathematica

Συμπλήρωση και στατιστική επεξεργασία σταθμών

```
SetDirectory["C:\\diplwmatikh\\PRCPData100yr"];
files=FileNames[];
Info=ParallelTable[
DailyData=Import[files[[stNo]]/.{y_,m_,d_,val_,mf_,qf:"G"|"X",sf_}>:{y,m,d,"",mf,qf
,sf};
    StID=StringDrop[files[[stNo]],-4];
DailyDataPerYear=Split[DailyData[[All,{1,4}]],#1[[1]]==#2[[1]]&][[All,All,2]];
    YearLst=Union[DailyData[[All,1]]];
    AnnualMeanValues=If[Length[#]>=
345,Mean[#,""]&/@DeleteCases[#,""]&/@DailyDataPerYear];
    AnnualMeanValuesFilled=ReplaceRepeated[AnnualMeanValues,{left___x_"
",y_right__}>:{left,x,Mean[{x,y}],y,right}/;x!=""&&y!=""};

    RL=Length[YearLst];
    ValsNo=Count[AnnualMeanValuesFilled,_?NumberQ];
    MisValsNo=Count[AnnualMeanValuesFilled,""];
    MVPCT=Round[N[100MisValsNo/(ValsNo+MisValsNo)],.1];
    LongestSeqValsNo=AnnualMeanValuesFilled/.{a___,Longest[b__Real],c___}>
Length[{b}];
    {StID,RL,ValsNo,MisValsNo,MVPCT,LongestSeqValsNo},
    {stNo,Length[files]}]
```

Επιλογή 1265 σταθμών

```
SetDirectory["C:\diplwmatikh\Info"];
InfoFile2=Import["InfoFile2.csv"];
files=StringJoin[{#, ".csv"}]&/@Cases[Drop[InfoFile2,1],{StId_,RL_,ValsNo_,MisValsNo_,MVPCT_,LongestSeqValsNo_/;LongestSeqValsNo>= 100}][[All,1]];
Table[ SetDirectory["C:\\diplwmatikh\\PRCPData100yr"];
DailyData=Import[files[[stNo]]]/.{y_,m_,d_,val_,mf_,qf:"G"|"X",sf_}>:{y,m,d,"",mf,qf,sf};
DailyDataPerYear=Split[DailyData[[All,{1,4}]],#1[[1]]==#2[[1]]&][[All,All,2]];
YearLst=Union[DailyData[[All,1]]];
AnnualMeanValues=If[Length[#]>=345,Mean[#,""]&/@(DeleteCases[#, ""]&/@DailyDataPerYear);
AnnualMeanValuesFilled=ReplaceRepeated[AnnualMeanValues,{left___,x_,"",y_,right___}>:{left,x,Mean[{x,y}],y,right}/;x!=""&&y!=""};
TS=Thread[{YearLst,AnnualMeanValuesFilled}];
SetDirectory["C:\\diplwmatikh\\finaldata2-345meres"];
Export[files[[stNo]],TS,
{stNo,Length[files]}
```

Στατιστική επεξεργασία επιλεγμένων σταθμών

```
SetDirectory["C:\\diplwmatikh\\finaldata2-345meres-1220"];

files=FileNames[];
Infomean=Table[data=Import[files[[stNo]]][[All,2]];data2=DeleteCases[data,""];
Mean[data2],{stNo,1,Length[files]};
Mean1=Mean[Infomean]
Sd1=StandardDeviation[Infomean]
Min1=Min[Infomean]
Q1=Quantile[Infomean,{0.025,0.25,0.50,0.75,0.975}]
Max1=Max[Infomean]
```

Υπολογισμός συντελεστή Hurst σταθμών με ελλειπείς τιμές

```
HurstExponentII[data_,minScale_,maxScale_] := Block[{pnts,H,x},
```

```
pnts=Table[  
  {Log[10,k],  
   Log[10,StandardDeviation[  
     DeleteCases[  
       Which[  
         k==1,DeleteCases[data,""],  
         k==2,If[Length[#]>= 2,Mean[#, ""]&/@DeleteCases[Partition[data,k], "", {2}],  
         k>=3,If[Length[#]>= 3,Mean[#, ""]&/@DeleteCases[Partition[data,k], "", {2}]  
       ],  
       ""]  
     ]}],  
  {k,maxScale}];
```

```
H=1+Fit[Drop[pnts,minScale-1],{1,x},x][[2,1]];
```

```
Which[0<H<1,H,H<0,0,H>1,1]
```

```
];
```

Υπολογισμός συντελεστή Hurst σταθμών πλήρων ετών

```
HurstExponentII[data_,minScale_,maxScale_]:=Block[{pnts,H,x},
```

```
pnts=Table[  
  {Log[10,k],  
  Log[10,StandardDeviation[Mean/@  
  Partition[data,k]]  
  ]},  
  {k,maxScale}];
```

```
H=1+Fit[Drop[pnts,minScale-1],{1,x},x][[2,1]];
```

```
Which[0<H<1,H,H<0,0,H>1,1]
```

```
];
```

Σύγκριση συνθετικών χρονοσειρών με ιστορικών

```
result2=Table[SetDirectory["C:\\diplwmatikh\\sun8etikes xronoseires-  
0.65"];data=Flatten[Import[data22[[stNo]]]];SetDirectory["C:\\diplwmatikh\\In  
fo"];
```

```
list=Flatten[Import["Info4.csv"]];data23=Take[data,list[[stNo]]];HurstExponentI  
I[data,1,Floor[Length[data]/10]],{stNo,1,1265}];
```

Υπολογισμός αυτοσυσχετίσεων

```
Correl[data_lag_]:=Block[{standardizedSet,productsSet},
```

```
dataVals=Cases[data,_?NumberQ];
```

```
standardizedSet=(data-Mean[dataVals])/StandardDeviation[dataVals];
```

```
productsSet=Cases[Drop[standardizedSet,lag]Drop[standardizedSet,-  
lag],_?NumberQ];
```

```
Total[productsSet]/(Length[productsSet]-1)
```

```
];
```

```
ACF[data_maxLag_]:=Table[Correl[data,lag],{lag,maxLag}];
```

```
Test=Cases[Info2,{r1_/;r1>0,r2_/;r2>0,_}];
```

```
rtet=Test[[All,1]]^2;
```

```
r2=Test[[All,2]];
```

```
dif=r2-rtet;
```

```
Test2=Cases[Info2,{r1_/;r1>0,r2_/;r2>0,r3_/;r3>0}];
```

```
rtrit=Test2[[All,1]]^3;
```

```
r3=Test2[[All,3]];
```

```
dif2=r3-rtrit;
```