ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΤΜΗΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ
ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ»

Εύα - Στυλιανή Στείρου

Διερεύνηση των σφαλμάτων από ομογενοποίηση σε υδροκλιματικές χρονοσειρές με μακροπρόθεσμη εμμονή

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων Καθηγητής: Δημήτρης Κουτσογιάννης

Αθήνα, Μάρτιος 2013

ΠΡΟΛΟΓΟΣ

Η εκτίμηση των κλιματικών μεταβολών είναι ένα ζήτημα εξαιρετικά σημαντικό από επιστημονική, κοινωνική και οικονομική άποψη. Ως προς την επιστημονική, συγκεκριμένα, πλευρά του είναι ένα πεδίο συνεχιζόμενης για δεκαετίες συζήτησης σε έντονους συχνά τόνους. Ένα από τα κεντρικά ζητήματα της συζήτησης αυτής είναι η ομογενοποίηση των υδροκλιματικών δεδομένων, δηλαδή η επεξεργασία τους για την αφαίρεση σφαλμάτων διαφόρων τύπων που τα χαρακτηρίζουν.

Η ομογενοποίηση υδροκλιματικών δεδομένων και συγκεκριμένα χρονοσειρών θερμοκρασίας και βροχόπτωσης είναι το αντικείμενο που πραγματεύεται η παρούσα Μεταπτυχιακή εργασία. Η εργασία αυτή αποτελεί ουσιαστικά συνέχεια της προηγούμενης Διπλωματικής μου εργασίας που εκπονήθηκε μεταξύ Νοέμβρη 2010 - Οκτώβρη 2011 και έθιξε ανάμεσα σε άλλα ζητήματα την αγνόηση της μακροπρόθεσμης εμμονής, ενός σημαντικού χαρακτηριστικού των υδροκλιματικών δεδομένων, κατά την ομογενοποίησή τους. Η αλήθεια είναι ότι όταν ξεκίνησα να ασχολούμαι με το θέμα αυτό, δεν περίμενα ότι θα ήταν τόσο εκτεταμένο και πολύπλοκο, αλλά και ζήτημα το οποίο θα μπορούσε να προκαλέσει τόσο πολλές αντιδράσεις.

Η οριστική απόφαση για την ενασχόληση με την ομογενοποίηση υδροκλιματικών δεδομένων με μακροπρόθεσμη εμμονή λήφθηκε μετά από παρουσίαση των αποτελεσμάτων της Διπλωματικής μου εργασίας στο συνέδριο της EGU, τον Απρίλιο 2012. Το ενδιαφέρον που συγκέντρωσε η παρουσίαση αυτή και η κριτική που δέχτηκε, θετική αλλά και αρνητική, ήταν κάτι που δεν περίμενα, αλλά βοήθησε να ξεκαθαριστούν σημαντικές παράμετροι που διαπραγματεύεται η παρούσα Μεταπτυχιακή εργασία και σε σημαντικό βαθμό και η κατεύθυνσή της.

Ολοκληρώνοντας λοιπόν την εργασία αυτή θέλω να ευχαριστήσω όσους συνεισέφεραν στο τελικό αποτέλεσμα, ιδιαίτερα τον Victor Venema, με θετικά ή κριτικά σχόλια. Τα δεύτερα ήταν και τα πιο εποικοδομητικά για τη βελτίωση της μεθοδολογίας που ακολουθήθηκε και των αποτελεσμάτων που προέκυψαν. Ευχαριστώ επίσης θερμά και τα μέλη της εξεταστικής μου επιτροπής, την Αναπλ. Καθ. Νυμφοδώρα Παπασιώπη και τον Επικ. Καθ. Νίκο Μαμάση για τις παρατηρήσεις στο τελικό κείμενο και τα ενθαρρυντικά τους σχόλια.

Ένα μεγάλο ευχαριστώ επίσης οφείλω σε όσους με στήριξαν κατά τη διάρκεια της εργασίας μου, την οικογένειά μου, φίλους και συμφοιτητές.

Πάνω απ’όλα θέλω να ευχαριστήσω θερμά τον Καθ. Δημήτρη Κουτσογιάννη, ο οποίος επέβλεψε την εργασία αυτή όπως και την προηγούμενη και μου έδωσε τη δυνατότητα να ασχοληθώ με ουσιαστικό τρόπο με ένα ζήτημα με μεγάλο κοινωνικό ενδιαφέρον. Είναι ο εμπνευστής της διερεύνησης αυτής και η συνεργασία μας άλλαξε πραγματικά τον τρόπο που σκέφτομαι. Ελπίζω πραγματικά η διερεύνηση που έγινε και συνοψίζεται στην παρούσα εργασία να χρησιμεύσει στην καλύτερη κατανόηση και εκτίμηση των κλιματικών μεταβολών.

Ε. Σ.

Μάρτης 2013

ΠΕΡΙΕΧΟΜΕΝΑ

Πρόλογος i
Περίληψη iv
Extended Abstract v
1 Εισαγωγή 1
1.1 Κίνητρα και αντικείμενο της εργασίας 1
1.2 Διάρθρωση της εργασίας 3
2 Ανομογένειες υδροκλιματικών δεδομένων 4
2.1 Τυπολογία ανομογενειών χρονοσειρών θερμοκρασίας και βροχοπτώσεων 5
2.1.1 Σφάλματα θερμομέτρων/καταγραφών (measurement bias) 6
2.1.2 Μεταβολές συνθηκών μετρήσεων 6
2.1.3 Μεταβολή του ρυθμού δειγματοληψίας ή/και διαφορές στη μέθοδο υπολογισμού ημερήσιων τιμών 9
3 Ομογενοποίηση υδροκλιματικών δεδομένων 11
3.1 Υποκειμενικές (subjective) μέθοδοι ομογενοποίησης 12
3.1.1 Διπλή αθροιστική καμπύλη 13
3.2 Αντικειμενικές (objective) μέθοδοι ομογενοποίησης 15
3.2.1 Μέθοδος των Easterling and Peterson (1995) 16
3.2.2 Standard Normal Homogeneity Test for single shifts 19
3.2.3 Standard Normal Homogeneity Test for trends 21
3.2.4 Pairwise Homogenization Algorithm 23
4 Αξιολόγηση μεθόδων ομογενοποίησης – Προηγούμενες προσεγγίσεις 26
4.1 Αξιολόγηση μεθόδων ομογενοποίησης με χρήση πραγματικών δεδομένων 26
4.2 Αξιολόγηση μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων 27
4.2.1 Μοντέλα παραγωγής συνθετικών δεδομένων 28
5 Προηγούμενες προσεγγίσεις ομογενοποίησης με συνθετικά δεδομένα – Ανάλυση και κριτική προσέγγιση 31
5.1 Προηγούμενες προσεγγίσεις 31
5.2 Κατηγοριοποίηση χαρακτηριστικών των προσεγγίσεων 32
5.3 Παράμετροι διερεύνησης 33
5.4 Επισκόπηση των μεθόδων με βάση τα επιλεγμένα κριτήρια 33
5.4.1 Συγκριτική αξιολόγηση διαφορετικών μεθόδων 34
5.4.2 Ανάπτυξη και έλεγχος καινούριας μεθόδου ή τροποποίηση/εφαρμογή μιας προϋπάρχουσας μεθόδου 51
5.5 Σύγκριση - σχολιασμός προηγούμενων προσεγγίσεων 55
5.5.1 Χαρακτηριστικά συνθετικών δεδομένων 56
5.5.2 Μέθοδοι ομογενοποίησης 57
5.5.3 Διερεύνηση αυτοσυσχέτισης διαφοράς χρονοσειρών ελεγχόμενης – αναφοράς 58
6 Μεθοδολογία 60
6.1 Επιλογή μεθόδων ομογενοποίησης 61
6.2 Γενικοί περιορισμοί και προϋποθέσεις των μεθόδων 62

6.3 Στατιστικά χαρακτηριστικά και παράμετροι των συνθετικών χρονοσειρών 62
6.3.1 Συντελεστής Hurst 63
6.3.2 Συντελεστής ετεροσυσχέτισης 63
6.3.3 Σύστημα χρονοσειρών αναφοράς 64
6.3.4 Μήκος συνθετικών χρονοσειρών 66
6.3.5 Ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς 66
6.4 Παραγωγή συνθετικών χρονοσειρών 67
6.4.1 Αλγόριθμος παραγωγής συνθετικών χρονοσειρών με μακροπρόθεσμη εμμονή 68
6.4.2 Παραγωγή τελικών συσχετισμένων συνθετικών χρονοσειρών με μακροπρόθεσμη εμμονή 69
6.5 Εφαρμογή σε θερμοκρασίες 71
6.5.1 Επιλογή τρόπου εφαρμογής μεθόδου SNHT 71
6.5.2 Συνδυασμοί παραμέτρων 73
6.6 Εφαρμογή σε βροχοπτώσεις 74
6.6.1 Επιλογή τρόπου εφαρμογής Διπλής αθροιστικής καμπύλης 74
6.6.2 Αλγόριθμος διγραμμικής προσαρμογής 75
6.6.3 Συνδυασμοί παραμέτρων 78
7 Ανάπτυξη λογισμικού για παραγωγή και επεξεργασία χρονοσειρών 80
7.1 Κύριες συναρτήσεις 80
7.2 Βοηθητικές συναρτήσεις 84
8 Αποτελέσματα ομογενοποίησης σε δεδομένα θερμοκρασίας 92
8.1 Αποτελέσματα συστήματος 1/1 92
8.2 Αποτελέσματα συστήματος 3/7 94
8.3 Αποτελέσματα συστήματος 2/2 95
8.4 Αποτελέσματα συστήματος 3/3 96
8.5 Αποτελέσματα συστήματος 3/10 97
8.6 Αποτελέσματα συστήματος 3/20 98
8.7 Συγκριτική παρουσίαση αποτελεσμάτων διερεύνησης βασικών παραμέτρων 99
8.8 Διερεύνηση επιρροής μήκους χρονοσειρών 101
8.9 Διερεύνηση επιρροής ελάχιστης απόστασης ανομογενειών μεταξύ τους και από τα άκρα της χρονοσειράς 102
8.10 Διόρθωση χρονοσειρών 103
9 Αποτελέσματα ομογενοποίησης σε δεδομένα βροχόπτωσης 108
9.1 Αποτελέσματα συστήματος 1/1 108
9.2 Αποτελέσματα συστήματος 3/1 110
9.3 Διερεύνηση της παραμέτρου k 112
10 Συμπεράσματα και Προτάσεις 114
10.1 Συμπεράσματα 114
10.2 Προτάσεις 116
Αναφορές 117
ΠΑΡΑΡΤΗΜΑ Α 123
ΠΑΡΑΡΤΗΜΑ Β 124

ΠΕΡΙΛΗΨΗ

Οι καταγραφές υδροκλιματικών δεδομένων περιέχουν ανομογένειες, δηλαδή σφάλματα διαφόρων τύπων, συνηθέστερα άλματα, τα οποία εντοπίζονται και διορθώνονται με στατιστικές κυρίως μεθόδους ομογενοποίησης, συνήθως σε σύγκριση με γειτονικές χρονοσειρές. Η ομογενοποίηση αποτελεί αντικείμενο διαμάχης, επειδή υπάρχει υποψία ότι εισάγει συστηματικά σφάλματα τα οποία αλλοιώνουν την κλιματική πληροφορία.

Στα πλαίσια της εργασίας αυτής διερευνήθηκαν συστηματικά οι διάφορες μέθοδοι ομογενοποίησης δεδομένων θερμοκρασίας και βροχόπτωσης και οι μελέτες αξιολόγησής τους στη βιβλιογραφία. Διαπιστώθηκε ότι οι μέθοδοι και οι αξιολογήσεις τους μέχρι σήμερα βασίζονται σε δύο υποθέσεις: Πρώτον ότι τα υδροκλιματικά δεδομένα χαρακτηρίζονται από ασθενή ή και καθόλου αυτοσυσχέτιση, χωρίς να λαμβάνεται υπόψη ένα στοχαστικό χαρακτηριστικό των δεδομένων, η μακροπρόθεσμη εμμονή, δηλαδή μια δομή αυτοσυσχέτισης που διατηρείται στο χρόνο και προσδιορίζεται από το συντελεστή Hurst. Και δεύτερον, ότι η διαφορά θερμοκρασίας και ο λόγος βροχόπτωσης γειτονικών σταθμών αποτελούν σειρά τυχαίων αριθμών, ενώ αντίθετα αποδεικνύεται ότι είναι συσχετισμένα.

Για να διερευνηθούν τα αποτελέσματα της ομογενοποίησης σε δεδομένα με μακροπρόθεσμη εμμονή, εφαρμόστηκαν, με χρήση της μεθόδου Monte Carlo δύο κλασικές μέθοδοι ομογενοποίησης, οι SNHT και Διπλή αθροιστική καμπύλη, σε ετήσια ομογενή συνθετικά δεδομένα θερμοκρασίας και βροχόπτωσης, αντίστοιχα.

Διαπιστώθηκε ότι οι δύο μέθοδοι ομογενοποίησης οδηγούν σε στατιστικά μη σημαντικό αριθμό χρονοσειρών με πλασματικά άλματα για τιμή του συντελεστή Hurst $H = 0.5$, που αντιστοιχεί σε ασυσχέτιστα δεδομένα. Αύξηση όμως του συντελεστή Hurst ($0.5 < H < 1$) οδηγεί σε αύξηση του ποσοστού των χρονοσειρών με πλασματικά άλματα

Διαπιστώθηκε επίσης ότι ο αριθμός και ο τρόπος χρήσης των χρονοσειρών αναφοράς επηρεάζει επίσης πολύ το ποσοστό των χρονοσειρών στις οποίες εντοπίζονται ψευδή άλματα.

Από την εφαρμογή της μεθόδου SNHT προέκυψε ότι το ποσοστό χρονοσειρών με ψευδή άλματα φαίνεται να επηρεάζεται από το μήκος των χρονοσειρών και την ελάχιστη απόσταση ανάμεσα στις ανομογένειες, όχι όμως από τη συσχέτιση ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες.

Τέλος, από διόρθωση των αλμάτων που εντοπίζονται από το SNHT φαίνεται ότι η μέθοδος δεν οδηγεί σε αλλοίωση των τάσεων μεταβολών θερμοκρασίας, καθώς τα ποσοστά των χρονοσειρών στις οποίες αυξάνονται και μειώνονται οι τάσεις είναι σχεδόν ίσα. Αντίθετα, η διόρθωση αλμάτων τείνει να μειώνει το συντελεστή Hurst για αρχικά δεδομένα με μέτρια έως ισχυρή εμμονή ($H > 0.65$), όχι όμως για $H < 0.65$.

EXTENDED ABSTRACT

Introduction

Hydroclimatic time series contain inhomogeneities, which are errors introduced by replacements and calibration of instruments, station relocations, changes in the environment of the stations, etc. The most common expression of inhomogeneities is shifts between two parts of time series. The identification and correction of inhomogeneities is called homogenization and is usually done with statistical methods which compare a candidate station with one or more neighbouring reference stations, assuming that they belong to the same climatological region and they reflect the same weather and climate variations.

The homogenization of hydroclimatic data, mainly of temperature and precipitation time series, is a procedure of great importance and also a controversial subject because of its implications in the estimations of climate change. This study focuses on a generally ignored by the homogenization community, though important characteristic of hydroclimatic time series and of its effects on homogenization, the long-term persistence of hydroclimatic data, and has two components: (a) a literature review and (b) a computational approach.

1. Literature review

A systematic study of the scientific literature was made in order to examine types and causes of inhomogeneities, to identify and classify the existing homogenization methods, understand their stochastic background and to evaluate their output. This literature review focused mainly on previous evaluation studies of homogenization methods with synthetic data.

A main result of this study is that existing homogenization methods generally ignore the long-term persistence of hydroclimatic data expressed by the Hurst coefficient and examine only first-order autoregressive series (AR(1)) or series of identically and independently distributed Gaussian errors. No systematic studies of the relationship between the Hurst coefficient and the homogenization results have been identified.

It was also found that homogenization methods assume that the series of temperature differences or precipitation ratios between reference and candidate stations constitute series of random numbers (e.g. white noise). However a basic stochastic analysis indicates that the difference and ratio series reproduce the autocorrelation function of the reference and candidate series assuming that both have the same autocorrelation structure.

2. Computational approach

A computational approach based on Monte Carlo simulations permitted to understand and evaluate the behaviour of selected classical homogenization methods as a function of various parameters:

a) the Hurst coefficient of hydroclimatic time series (tested values H={0.50, 0.55, ..., 0.90}),

b) the cross-correlation coefficient between candidate and reference time series (tested values: ρ_{XY}={0.5, 0.6, 0.7, 0.8, 0.9, 0.95}),
c) the number of reference time series and the way they were used to locate shifts (reference systems - shown in Table 1),
d) the length of the time series (50 and 100 years), and
e) the minimum distance between possible inhomogeneities or between inhomogeneities and the edge of series (tested values 5 and 10 years).

The percentage of time series with false alarms was regarded as the critical factor for this evaluation.

Table 1. *Characteristics of the homogenization tests used.*
Number of common shifts: minimum number of times a shift must be identified by different reference series so that it can be considered significant, L: length of simulated time series, MD: minimum distance between possible inhomogeneities or between inhomogeneities and the edge of series

Symbol	Number of reference series	Number of common shifts	Comments
1/1	1	1	L:100, MD:5
3/7	7	3	L:100, MD:5
3/1	3	1	An artificial time series is constructed as the mean of the simulated reference series L:100, MD:5
2/2	2	2	L:100, MD:5
3/3	3	3	L:100, MD:5
3/10	10	3	L:100, MD:5
3/20	20	3	L:100, MD:5
3/7-50y	7	3	L:50, MD:5
3/7–j10	7	3	L:100, MD:10

For temperatures, the homogenization method selected was SNHT for shifts (Alexandersson and Moberg, 1997) in combination with all systems of reference series summarized in Table 1 except 3/1 (see Figure 1). For multiple reference series a pairwise comparison of the candidate series with all reference series was applied. The SNHT was applied using a cutting algorithm described by Domonkos (2011a) and a 95% confidence level. For the precipitation data the Double Mass Curve (Kohler, 1949, Searcy and Hardison,1960) was selected. The original method is subjective and involves the identification of the main inhomogeneity of the time series on a graph. An objective (automated) version was developed using a piecewise linear algorithm based on the least squares approach. The reference series systems used were the most commonly applied 1/1 and 3/1.

The synthetic series with long-term persistence were simulated using a multiple time-scale fluctuation approach proposed by Koutsoyiannis (2002) and following the normal distribution. Temperature data were generated with zero mean and unit standard deviation and precipitation data with a mean of 1000mm and a standard deviation of 300mm. For every candidate series one or multiple correlated reference series with the same characteristics were generated (see Table 1). All simulations and computations were based on original Matlab codes.

3. Results and conclusions

Some main conclusions of this study are summarized in Figures 1, 2, 3 and 4.

a) For time series with H=0.5 (i.e. characterized by white noise), the false alarm rate is not significant (below 5%), which is expected because of the design of the homogenization methods, but the percentage of series with false alarms increases with H.
b) The number of reference series and of the minimum number required to locate shifts in the time series greatly affects the percentage of series with false alarms.

Furthermore, some more conclusions can be extracted concerning the application of the SNHT to temperature data and the Double mass curve to precipitation data.

For the temperature data (see Figures 1 and 2) it can be assumed that:

a) For a Hurst coefficient $H \geq 0.85$, common shifts located by all (minimum number 3) reference time series correspond to percentage of series with false alarms higher than 5% and tend to indicate a real inhomogeneity (e.g. systems 2/2 and 3/3 in Figure 1).
b) In the case of a common shift identified by some of the reference series only, this may only correspond to a false alarm (e.g. Figure 1). In such cases a possible inhomogeneity must be confirmed by analysis of the reference time series.
c) The cross-correlation coefficient between reference and candidate series does not seem to influence the percentage of time series with false alarms.
d) The percentage of series with false alarms for time series with length 50 years was lower than the percentage for 100 years (Figure 1).
e) The minimum distance between possible inhomogeneities or between inhomogeneities and the edge of series influences but not greatly the percentage of series with false alarms. A minimum distance of 10 years leads to a lower percentage than a minimum distance of 5 years (Figure 1).
f) For the case of a single reference series, corrections of the located shifts were applied. These corrections led to a similar percentage of series with increased and decreased trend after the homogenization. Therefore it seems that SNHT does not introduce significant changes in the temperature trends.
g) For $H < 0.65$ the percentage of series with an increased Hurst coefficient after homogenization is similar to that with a lower Hurst coefficient. For $H > 0.65$ there is a different case. The percentage of series with an increased Hurst coefficient exceeds that of

series with a lower Hurst coefficient. This difference increases with the increase of the initial Hurst coefficient of the time series.

For the precipitation data (see Figures 3 and 4) it can be assumed that:

a) For all values of the Hurst coefficient examined, the percentage of series with false alarms decreases with the increase of the ratio of the slopes of the two lines of the Double Mass Curve.
b) Application of the Double Mass Curve with a reference time series produced by three time series (3/1) tends to decrease the percentage of false alarms in comparison to the application of the method with a single reference series (1/1).
c) For the system 1/1 and all the parameters examined a slope ratio 1.5 corresponds to a percentage of series with false alarms lower than 5%. For the system 3/1 the same ratio is 1.3. These values seem to be indicative of a real inhomogeneity.

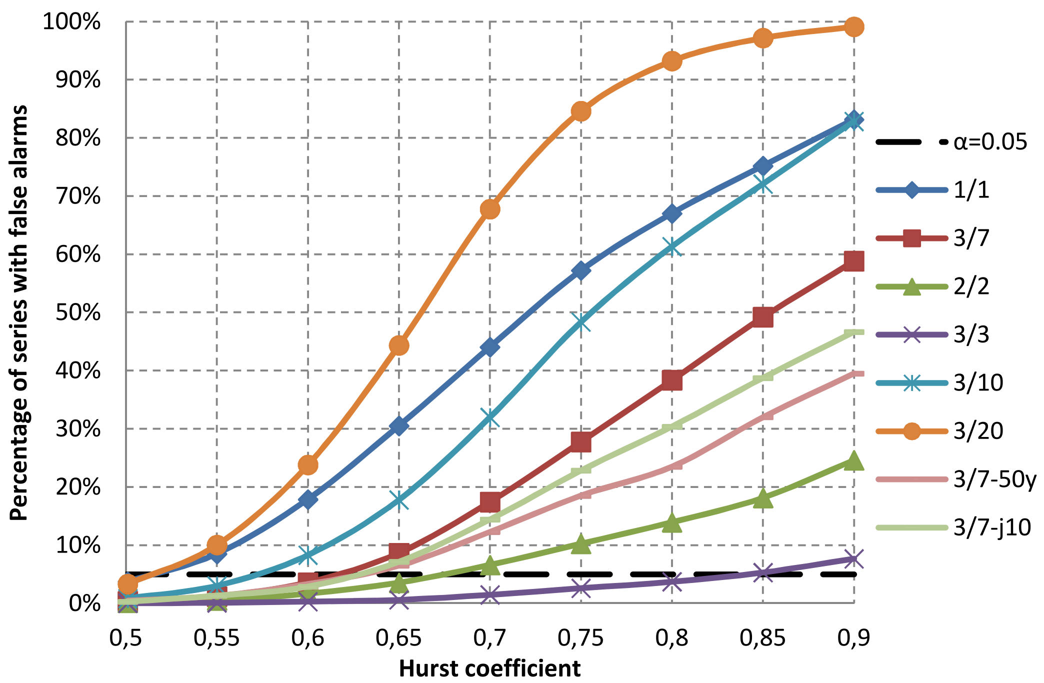


Figure 1. *Percentage of series with false alarms as a function of the Hurst coefficient for various combinations of reference series analysed in Table 1. Results are shown for a typical cross-correlation coefficient between reference and candidate series,* $\rho_{XY}=0.8$.

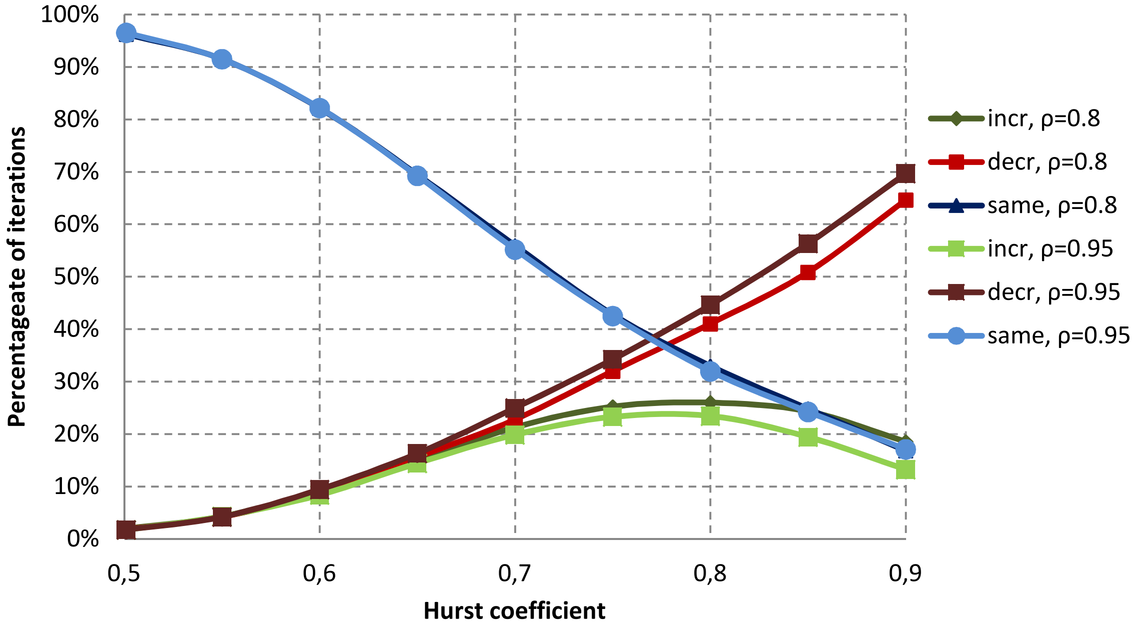


Figure 2. *Influence of the SNHT homogenization in the Hurst coefficient of the homogenized times series. Results are shown for a typical cross-correlation coefficient between reference and candidate series,* ρ_{XY}*=0.8 and for* ρ_{XY}*=0.95 that led to slightly different results.*

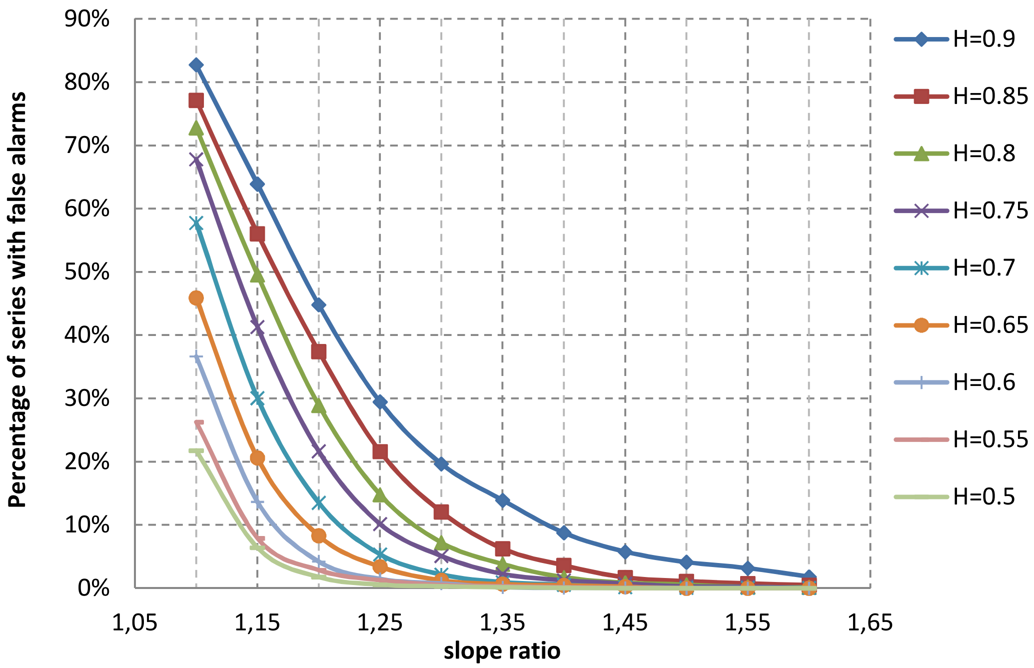


Figure 3.*Percentage of series with false alarms as a function of the Hurst coefficient and the slope ratio of the lines of the Double Mass Curve for the reference system 1/1 analysed in Table 1. Results are shown for a typical cross-correlation coefficient between reference and candidate series,* ρ_{XY}*=0.8.*

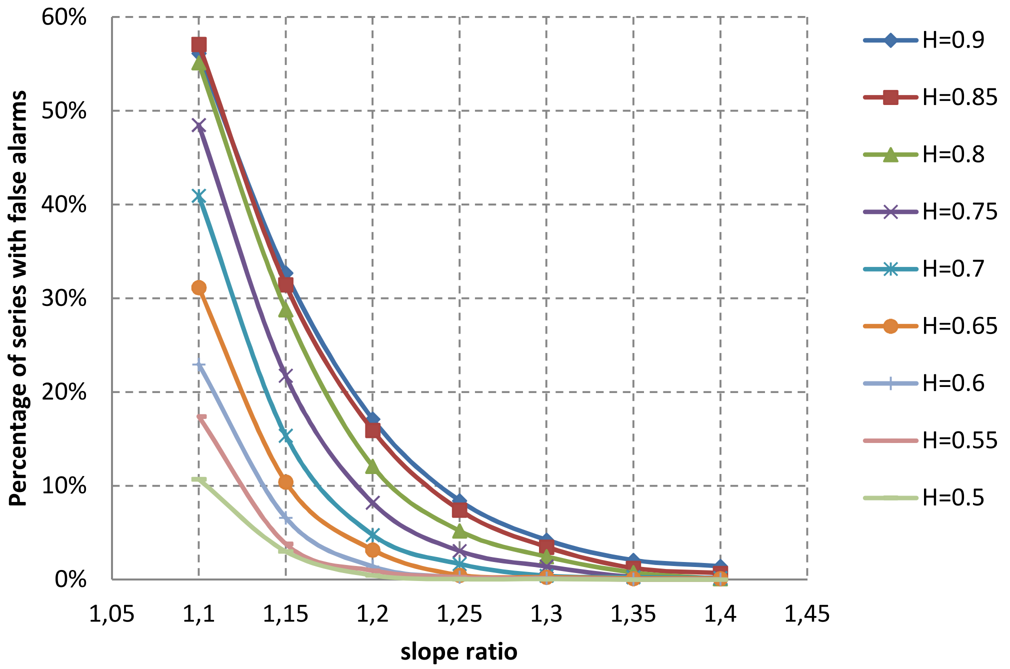


Figure 4.*Percentage of series with false alarms as a function of the Hurst coefficient and the slope ratio of the lines of the Double Mass Curve for the reference system 3/1 analysed in Table 1. Results are shown for a typical cross-correlation coefficient between reference and candidate series,* ρ_{XY}*=0.8.*

1. ΕΙΣΑΓΩΓΗ

1.1. Κίνητρα και αντικείμενο της εργασίας

Η κλιματική μεταβολή αποτελεί ένα πεδίο διεθνούς πολιτικής διαμάχης, ένα αντικείμενο με σημαντική οικονομική διάσταση λόγω των φόρων ρύπων, αλλά και ένα πεδίο έρευνας αλλά και έντονης διαφωνίας μεταξύ των επιστημόνων, κυρίως όσον αφορά την εκτίμηση της αύξησης της θερμοκρασίας σε παγκόσμια κλίμακα και των μεθόδων που επιτρέπουν την εκτίμηση αυτή.

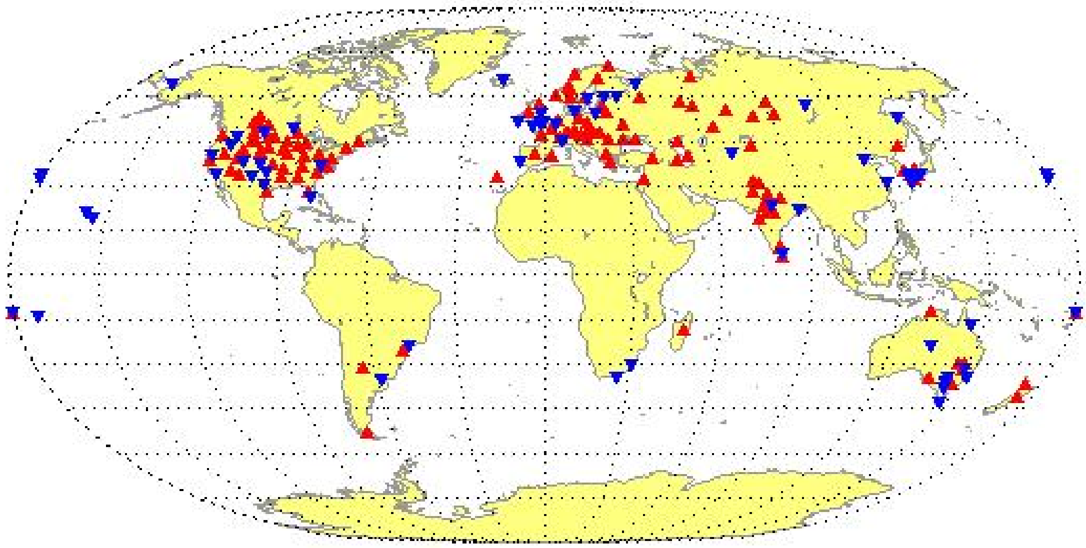
Οι εκτιμήσεις μεταβολής του κλίματος βασικά προκύπτουν από υδροκλιματικά δεδομένα όπως η μέση μηνιαία και ετήσια θερμοκρασία και η συνολική μηνιαία και ετήσια βροχόπτωση που μετρώνται στην επιφάνεια του εδάφους (π.χ. IPCC, 2007), αλλά και η ατμοσφαιρική πίεση, η ηλιοφάνεια (π.χ. Auer *et al*., 2007) κ.ά.

Τα υδροκλιματικά δεδομένα στην αρχική τους μορφή (ανεπεξέργαστες καταγραφές), περιέχουν σφάλματα τα οποία οφείλονται σε λάθη των παρατηρητών, αλλαγές τοποθεσίας των σταθμών, αλλαγές οργάνων και πολυάριθμους ακόμα παράγοντες (Alexandersson and Moberg, 1997). Τα σφάλματα αυτά προφανώς επηρεάζουν τις εκτιμήσεις των μεταβολών του κλίματος. Για το λόγο αυτό, εδώ και κάποιες δεκαετίες γίνεται προσπάθεια εντοπισμού και απομάκρυνσής τους από τα δεδομένα με στατιστικές κυρίως μεθόδους ομογενοποίησης ώστε να εξασφαλιστούν δεδομένα ομογενή, απαλλαγμένα δηλαδή από σφάλματα - ανομογένειες (Venema *et al*., 2012).

Είναι προφανές ότι η αξιοπιστία των συμπερασμάτων που προκύπτουν από τη μελέτη των υδροκλιματικών δεδομένων εξαρτάται από το βαθμό ομογένειας των χρησιμοποιούμενων χρονοσειρών (Aguilar *et al*., 2003). Πρόσφατες όμως μελέτες (Στείρου, 2011, Steirou and Koutsoyiannis, 2012), οι οποίες εστιάστηκαν στις μεθόδους ομογενοποίησης και τα αποτελέσματά τους, εντόπισαν κάποια σημεία που δημιουργούν προβληματισμούς γύρω από τη διαδικασία της ομογενοποίησης.

Για παράδειγμα, από συστηματική δειγματοληπτική έρευνα σε παγκόσμια κλίμακα προέκυψε ότι στα 2/3 των χρονοσειρών θερμοκρασίας εδάφους μήκους τουλάχιστον 100 ετών, η ομογενοποίηση τείνει να διογκώνει τις τάσεις αύξησης της θερμοκρασίας και να μειώνει τις τάσεις μείωσης της ή να τις μετατρέπει σε αυξητικές. Εφόσον τα στοιχεία αυτά είναι αντιπροσωπευτικά, η ομογενοποίηση τείνει να εμφανίζει ένα αυξημένο επίπεδο παγκόσμιας κλιματικής μεταβολής (Σχήμα 1.1).

Επιπλέον, διαπιστώθηκε ότι η ομογενοποίηση μέχρι τώρα έχει βασιστεί στην υπόθεση ότι τα κλιματικά δεδομένα χαρακτηρίζονται από ανεξαρτησία ή βραχυπρόθεσμη εμμονή, αγνοώντας ένα χαρακτηριστικό των υδρολογικών και κλιματικών δεδομένων, τη μακροπρόθεσμη εμμονή, δηλαδή μια δομή αυτοσυσχέτισης που διατηρείται στο χρόνο (Koutsoyiannis, 2002). Τα δεδομένα με μακροπρόθεσμη εμμονή παρουσιάζουν τυχαίες διακυμάνσεις γύρω από τη μέση τιμή τους. Οι μέθοδοι ομογενοποίησης είναι πιθανόν να χαρακτηρίζουν τις διακυμάνσεις αυτές ως σφάλματα και να τις διορθώνουν.



Σχήμα 1.1. *Γεωγραφική κατανομή δείγματος σταθμών μέτρησης θερμοκρασίας στους οποίους έχει γίνει ομογενοποίηση. Με κόκκινα και μπλε τρίγωνα σημειώνονται οι σταθμοί στους οποίους η ομογενοποίηση αύξησε ή μείωσε αλγεβρικά, αντίστοιχα, το ρυθμό αύξησης της θερμοκρασίας. Η ομογενοποίηση φαίνεται ότι τείνει να αυξάνει τα επίπεδα κλιματικής αλλαγής (Στείρου, 2011).*

Από τις παραπάνω παρατηρήσεις προκύπτει ένα σημαντικό ερώτημα το οποίο χρήζει περαιτέρω έρευνας: Κάτω από ποιες συνθήκες μπορεί να γίνεται διόρθωση μιας χρονοσειράς με μακροπρόθεσμη εμμονή στην οποία έχουν εντοπιστεί ανομογένειες με βάση στατιστικές μεθόδους, και τι επιπτώσεις μπορεί να έχει μια τέτοια διόρθωση, τόσο στις θερμοκρασίες, όσο και σε άλλα υδροκλιματικά δεδομένα;

Το παραπάνω ουσιώδες ερώτημα έχει ελάχιστα απασχολήσει τη διεθνή βιβλιογραφία. Μόνο κάποιες μεμονωμένες εφαρμογές μεθόδων ομογενοποίησης έχουν γίνει για την κατανόηση της επίδρασης της ομογενοποίησης σε δεδομένα με μακροπρόθεσμη εμμονή (Koutsoyiannis and Langousis, 2011, Rust *et al.*, 2008, Στείρου, 2011).

Συμβολή στην απάντηση του παραπάνω ερωτήματος αποτελεί και η συγκεκριμένη Μεταπτυχιακή εργασία. Η εργασία αυτή εκπονήθηκε στα πλαίσια του ΔΠΜΣ «Επιστήμη και Τεχνολογία Υδατικών Πόρων», του Εθνικού Μετσόβιου Πολυτεχνείου και εστιάστηκε σε ετήσια δεδομένα θερμοκρασίας και βροχόπτωσης, τα οποία είναι βασικές παραμέτροι στη μελέτη και τη διαχείριση των υδατικών πόρων και από τους κυριότερους τύπους δεδομένων στους οποίους εφαρμόζονται στατιστικές μέθοδοι ομογενοποίησης.

Ως προς τη μεθοδολογία, η έρευνα επικεντρώθηκε στην εφαρμογή μεθόδων ομογενοποίησης σε συνθετικά δεδομένα, με στόχο να διερευνηθεί η σχέση ανάμεσα στον εντοπισμό πλασματικών - ψευδών αλμάτων από τις μεθόδους και διάφορα χαρακτηριστικά των συνθετικών δεδομένων και των μεθόδων ομογενοποίησης, με κυριότερο το συντελεστή Hurst των χρονοσειρών, ο οποίος είναι ενδεικτικός της μακροπρόθεσμης εμμονής.

Στα συνθετικά δεδομένα θερμοκρασίας και βροχόπτωσης εφαρμόστηκαν δύο από τις σημαντικότερες μεθόδους ομογενοποίησης, το Standard Normal Homogeneity Test for single shifts (SNHT) και η Διπλή Αθροιστική καμπύλη, αντίστοιχα.

Η επιλογή των μεθόδων αυτών και του τρόπου εφαρμογής τους στις συνθετικές χρονοσειρές, καθώς και των χαρακτηριστικών των χρονοσειρών έγινε μετά από συστηματική επισκόπηση της βιβλιογραφίας γύρω από τις ανομογένειες, τις μεθόδους ομογενοποίησης και την αξιολόγηση των μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων.

Τα συμπεράσματα της παρούσας Μεταπτυχιακής εργασίας και κυρίως η σχέση ανάμεσα στο συντελεστή Hurst των χρονοσειρών και τα πλασματικά άλματα, μπορεί να έχουν κάποια χρησιμότητα και για την επεξεργασία διαφόρων τύπων υδροκλιματικών δεδομένων, εκτός από τα δεδομένα θερμοκρασίας και βροχόπτωσης. Μπορεί επίσης να αποδειχτούν σημαντικά για τη βελτίωση της υπάρχουσας μεθοδολογίας ομογενοποίησης, με αποτέλεσμα πιο αξιόπιστη εκτίμηση των κλιματικών μεταβολών σε τοπική και παγκόσμια κλίμακα.

1.2. Διάρθρωση της εργασίας

Η εργασία αυτή περιλαμβάνει συνολικά 9 Κεφάλαια και 2 Παραρτήματα. Το παρόν Κεφάλαιο 1 αποτελεί την Εισαγωγή. Τα Κεφάλαια 2 - 5 εστιάζονται και αξιολογούν την υπάρχουσα βιβλιογραφία σχετικά με την ομογενοποίηση. Στα Κεφάλαια 6 - 10 περιγράφεται η μεθοδολογία που αναπτύχθηκε για την αξιολόγηση μεθόδων ομογενοποίησης με συνθετικά δεδομένα και παρουσιάζονται τα αποτελέσματά της, καθώς και γενικότερα συμπεράσματα.

Συγκεκριμένα, στα Κεφάλαια 2 και 3 γίνεται επισκόπηση και κατηγοριοποίηση των ανομογενειών και των μεθόδων ομογενοποίησης, αντίστοιχα. Στο Κεφάλαιο 3 παρουσιάζονται επίσης αναλυτικά μερικές από τις σημαντικότερες και ευρύτερα χρησιμοποιούμενες μεθόδους ομογενοποίησης. Στο Κεφάλαιο 4 αναλύονται οι τρόποι αξιολόγηση της ομογενοποίησης με ιδιαίτερο βάρος στην αξιολόγηση με χρήση συνθετικών δεδομένων. Στο Κεφάλαιο 5 παρατίθενται, συγκρίνονται και σχολιάζονται οι προηγούμενες αξιολογήσεις της ομογενοποίησης με χρήση συνθετικών δεδομένων. Από τη συγκριτική αυτή αξιολόγηση προκύπτει καθαρά το κενό που υπάρχει στην εφαρμογή μεθόδων ομογενοποίησης σε συνθετικά δεδομένα με μακροπρόθεσμη εμμονή.

Στο Κεφάλαιο 6 γίνεται αναλυτική περιγραφή και δικαιολόγηση της μεθοδολογίας που επιλέχτηκε να ακολουθηθεί. Στο Κεφάλαιο 7 περιγράφονται οι συναρτήσεις που χρησιμοποιήθηκαν για τη διερεύνηση της ομογενοποίησης και δίνονται διαγράμματα ροής τους. Στα Κεφάλαια 8 και 9 παρατίθενται τα αποτελέσματα της εφαρμογής της ομογενοποίησης στα δεδομένα θερμοκρασίας και βροχόπτωσης, αντίστοιχα. Στο Κεφάλαιο 10 παρατίθενται τα συμπεράσματα από τη σύγκριση των αποτελεσμάτων στα δεδομένα θερμοκρασίας και βροχόπτωσης μεταξύ τους και με προηγούμενες προσεγγίσεις και δίνονται κάποιες πιθανές κατευθύνσεις για συνέχιση της έρευνας στο συγκεκριμένο αντικείμενο.

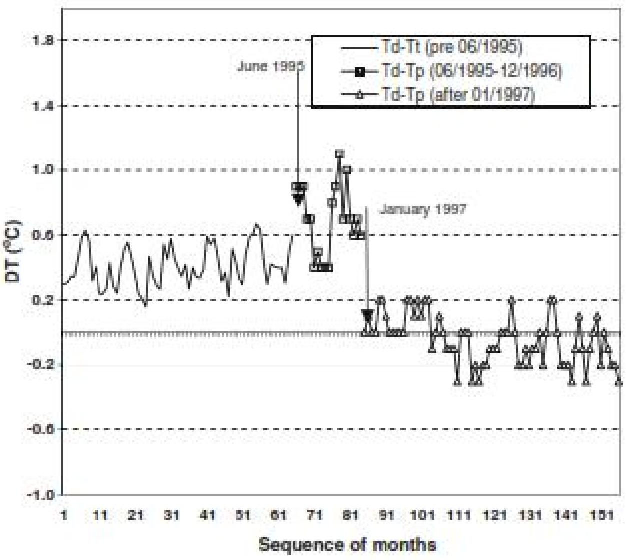
Στο Παράρτημα Α δίνονται οι κρίσιμες τιμές της μεθόδου SNHT for single shifts για τον εντοπισμό αλμάτων. Τέλος, στο Παράρτημα Β παρατίθενται οι κώδικες που χρησιμοποιήθηκαν για τον έλεγχο των μεθόδων ομογενοποίησης.

2. ΑΝΟΜΟΓΕΝΕΙΕΣ ΥΔΡΟΚΛΙΜΑΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Οι χρονοσειρές υδροκλιματικών δεδομένων περιγράφουν φαινόμενα διαφορετικά (θερμοκρασία, βροχόπτωση, πίεση, ηλιοφάνεια κ.ά.) και προκύπτουν από διαφορετικού τύπου μετρήσεις. Μετρήσεις παίρνονται από την επιφάνεια της γης, από όργανα πάνω σε σκάφη, από δορυφόρους κλπ και εκφράζουν καταγραφές από πολύ διαφορετικές πηγές (IPCC, 2007). Διαφοροποιήσεις υπάρχουν ακόμα και στις παραδοσιακές μετρήσεις κοντά στην επιφάνεια του εδάφους: οι τρόποι και οι συνθήκες των μετρήσεων (τύπος οργάνων, απόσταση οργάνων από την επιφάνεια του εδάφους κλπ) μεταβάλλονται από τόπο σε τόπο και από εποχή σε εποχή. Οι διαφοροποιήσεις αυτές προκαλούν ανομογένειες στα δεδομένα.

Με τον όρο **ανομογένεια** χαρακτηρίζονται συνήθως ατέλειες ή πιθανές ατέλειες των μετρήσεων, καθώς επίσης και τμήματα χρονοσειρών ή και ολόκληρες χρονοσειρές που δεν θεωρούνται αντιπροσωπευτικές των κλιματικών μεταβολών της θερμοκρασίας μιας ευρύτερης περιοχής ως συνάρτηση του χρόνου (Aguilar *et al.*, 2003).

Χαρακτηριστικό παράδειγμα χρονοσειράς με ανομογένειες αποτελεί το Σχήμα 2.1. Τα δεδομένα προέρχονται από το Αστεροσκοπείο Αθηνών. Η εικονιζόμενη χρονοσειρά είναι η διαφορά των μέσων μηνιαίων θερμοκρασιών από συμβατικό όργανο (υδραργυρικό θερμόμετρο) και από πιο σύγχρονο όργανο (αρχικά θερμογράφος συνεχούς καταγραφής και στη συνέχεια ηλεκτρονικό θερμόμετρο). Η αντικατάσταση του θερμογράφου έγινε τον Ιούνιο 1995, ενώ τον Ιανουάριο 1997 εφαρμόστηκε μια διόρθωση στο νέο όργανο. Οι δύο αυτές αλλαγές προκάλεσαν δύο εμφανείς μεταβολές στη μέση τιμή της χρονοσειράς (Founda *et al.*, 2009).



Σχήμα 2.1. *Χρονοσειρά διαφορών (DT) από το Αστεροσκοπείο Αθηνών ανάμεσα στις μέσες μηνιαίες θερμοκρασίες που προκύπτουν μετρήσεις με συμβατικό και σύγχρονο όργανο. Φαίνονται δύο ανομογένειες λόγω αντικατάστασης και ρύθμισης του σύγχρονου οργάνου Πηγή: Founda et al. (2009).*

Ο ορισμός των **ομογενών χρονοσειρών** που δίνεται στη βιβλιογραφία (π.χ. Easterling and Peterson, 1995) είναι ο ορισμός που δόθηκε από τους Conrad and Pollak (1962): χρονοσειρές στις οποίες οι μεταβολές οφείλονται αποκλειστικά σε μεταβολές των μετεωρολογικών συνθηκών και του κλίματος.

Οι Conrad and Pollak (1962) αναφέρουν ακόμα ότι μια χρονοσειρά κλιματικών δεδομένων είναι ομογενής συγκριτικά με μια σύγχρονη χρονοσειρά σε μια άλλη τοποθεσία αν η διαφορά τους (για τις χρονοσειρές θερμοκρασίας) ή ο λόγος τους (για χρονοσειρές βροχόπτωσης) είναι μια σειρά τυχαίων αριθμών (δηλαδή λευκός θόρυβος). Διερεύνηση της υπόθεσης αυτής θα γίνει αργότερα, στην Παράγραφο 5.5.

Η αξία των υδροκλιματικών δεδομένων στην εξαγωγή συμπερασμάτων εξαρτάται από το βαθμό ομογένειας των χρησιμοποιούμενων χρονοσειρών (Venema *et al*., 2012). Για το λόγο αυτό η αντιμετώπιση των προβλημάτων που προκαλούνται από τις ανομογένειες έχει απασχολήσει τους ερευνητές και έχει οδηγήσει

- σε μεθόδους εντοπισμού των ανομογενειών (homogeneity tests), και
- σε διόρθωση ή και απαλοιφή των αντίστοιχων στοιχείων (δεδομένων).

Μια τυπολογική ταξινόμηση των περιπτώσεων ανομογενειών των χρονοσειρών θερμοκρασίας και βροχοπτώσεων περιγράφεται στην Παράγραφο 2.1. Η ταξινόμηση αυτή προέκυψε μετά από επισταμένη μελέτη της σχετικής βιβλιογραφίας. Στην Παράγραφο 2.2 γίνεται ανάλυση της διαδικασίας με την οποία διορθώνονται οι ανομογένειες των υδροκλιματικών δεδομένων και ταξινόμηση των κυριότερων μεθόδων ομογενοποίησης.

2.1 Τυπολογία ανομογενειών χρονοσειρών θερμοκρασίας και βροχοπτώσεων

Στη διεθνή βιβλιογραφία έχουν γίνει διάφορες τυπολογικές προσεγγίσεις των ανομογενειών των χρονοσειρών θερμοκρασίας εδάφους και βροχοπτώσεων. Με βάση αυτές (π.χ. Easterling and Peterson, 1995, http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html) προτείνεται διάκριση των ανομογενειών σε κατηγορίες:

(α) ως προς τις αιτίες των ανομογενειών, και

(β) ως προς τη μορφή με την οποία εμφανίζονται οι ανομογένειες στις μετρήσεις.

Συγκεκριμένα, μπορεί να προταθεί ο παρακάτω διαχωρισμός που αναλύεται και τεκμηριώνεται στις επόμενες παραγράφους:

α) ως προς τις αιτίες:

1. σφάλματα οργάνων/καταγραφών (measurement bias),
2. μεταβολές συνθηκών μετρήσεων, ή
3. μεταβολή του ρυθμού δειγματοληψίας ή/και διαφορές στη μέθοδο υπολογισμού ημερήσιων τιμών.

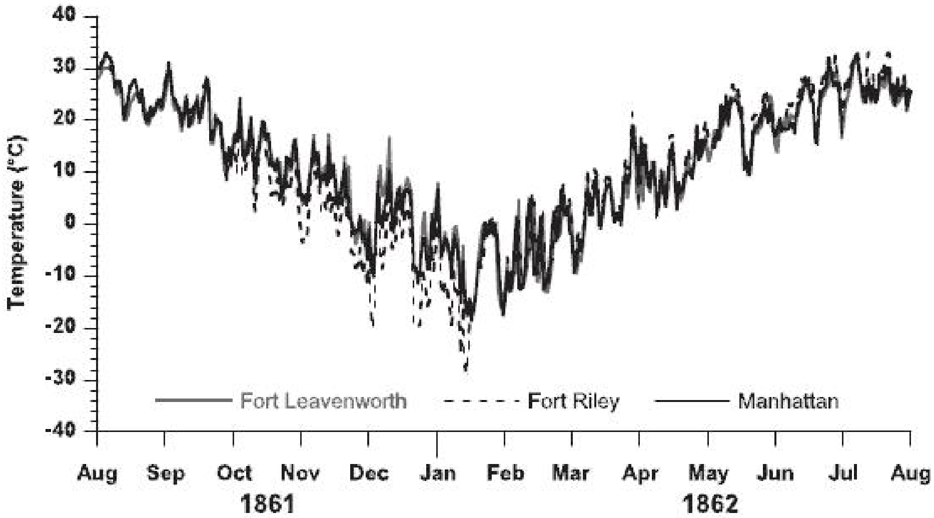
β) ως προς τη μορφή:

1. άλματα (breaks/shifts/jumps),
2. μεταβολές κλίσης (trends), ή
3. μεμονωμένες μη αντιπροσωπευτικές τιμές (outliers).

2.1.1 Σφάλματα οργάνων/καταγραφών (measurement bias)

Οι περιπτώσεις αυτές αφορούν:

1. Προβλήματα ρύθμισης-διακρίβωσης (calibration) ή σφάλματα οργάνων. Οι ανομογένειες σε αυτή την περίπτωση έχουν συνήθως τη μορφή συστηματικών σφαλμάτων σε μια ολόκληρη περίοδο καταγραφών (Σχήμα 2.2).



Σχήμα 2.2. *Σύγκριση των μέσων ημερήσιων θερμοκρασιών από τρεις σταθμούς του Kansas των ΗΠΑ: Fort Leavenworth, Fort Riley και Manhattan από τους Burnette et al. (2010). Ανάμεσα στο Σεπτέμβριο 1861 και τον Ιανουάριο 1862 παρατηρείται μια απόκλιση ανάμεσα στους σταθμούς στα Fort Riley και Manhattan, περιοχές που απέχουν μόνο 6.5 km. Η απόκλιση αυτή που φτάνει τους 13.28^o C εικάζεται από τους Burnette et al. ότι προκλήθηκε από απορρύθμιση ή ζημιά θερμομέτρου.*

2. Υποκειμενικά σφάλματα καταγραφών των μετρήσεων. Αυτά αφορούν είτε μετρήσεις που καταγράφηκαν εξ’αρχής λανθασμένα, είτε αντιγράφηκαν λανθασμένα, είτε καταγράφηκαν με χρονική διαφορά από την αναφερόμενη, π.χ. αντί για 9 το βράδυ, νωρίτερα, είτε ακόμα καταγράφτηκαν μετά από αυθαίρετη εκτίμηση του χειριστή του οργάνου (Peterson *et al.*, 1998). Σφάλματα μετρήσεων θεωρούνται επίσης και οι επαναλήψεις ίδιων γεγονότων (Menne *et al.*, 2009).

2.1.2 Μεταβολές συνθηκών μετρήσεων

Διακρίνονται τρεις περιπτώσεις:

1) Ασυνέχειες στη διαδικασία των μετρήσεων, κυρίως αντικατάσταση των οργάνων με όργανα ίδιου ή συνηθέστερα άλλου τύπου και διαφορετική ρύθμιση (calibration) των οργάνων, εσφαλμένη ή και κανονική.

Οι ασυνέχειες που συνδέονται με την αλλαγή της τεχνολογίας των οργάνων είναι πολύ συνήθεις και θεωρούνται από τις πιο σημαντικές. Συνήθως έχουν τη μορφή απότομων αλμάτων που επηρεάζουν τη μέση τιμή της χρονοσειράς (Easterling and Peterson, 1995).

Ένα χαρακτηριστικό παράδειγμα εικονίζεται στο Σχήμα 2.1.

Στις ΗΠΑ, σύμφωνα με τους Quayle *et al*. (1991), σημαντική αλλαγή των οργάνων μέτρησης θερμοκρασίας συνέβη στα μέσα και τέλη της δεκαετίας του 1980, με την αντικατάσταση των υδραργυρικών θερμομέτρων (liquid-in-glass- LIG thermometers) με καταγραφικά όργανα μεγίστου-ελαχίστου (thermistor Maximum- Minimum Temperature System- MMTS).

2) Αλλαγή τύπου ή μετατόπιση του μετεωρολογικού κλωβού

Η αλλαγή του τύπου μετεωρολογικού κλωβού αφορά το μέγεθος, τη διαπερατότητα (τύπος διέλευσης αέρα) και την ανακλαστικότητα του κλωβού. Η μετατόπιση προκαλεί μεταβολή στο ύψος, τον προσανατολισμό και τη θέση του μετεωρολογικού κλωβού. Οι δύο αυτοί τύποι αλλαγών έχουν ως αποτέλεσμα να εισάγεται στα καταγραφόμενα δεδομένα ένα συστηματικό σφάλμα.

Συνήθως το σφάλμα αυτό είναι πιο έντονο στην ελάχιστη και μέγιστη καταγραφόμενη θερμοκρασία λόγω αυξημένης ή μειωμένης έκθεσης στον ήλιο ή βόρειους ανέμους.

Χαρακτηριστικό είναι το παράδειγμα της Κεντρικής Ευρώπης που περιγράφεται από τους Böhm *et al. (*2010). Την περίοδο 1850-70 έγινε στην Κεντρική Ευρώπη αντικατάσταση των μετεωρολογικών κλωβών με καινούριους. Σύμφωνα με τους Böhm *et al.* (2010), οι παλιοί κλωβοί οδηγούσαν σε διαφοροποίηση της «πραγματικής» θερμοκρασίας εμφανίζοντας φαινομενικά μεγαλύτερες τιμές τις ζεστές περιόδους (έως και 2.5^{o} C) και φαινομενικά πιο κρύες τις κρύες (περίπου 0.5^{o} C). Το φαινόμενο αυτό ονομάστηκε “Early Instrumental Bias”.

Στις ΗΠΑ σημαντική αλλαγή των μετεωρολογικών κλωβών έγινε στα τέλη της δεκαετίας του 1980. Οι αρχικοί ξύλινοι κλωβοί (Cotton Region Shelter- CRS) αντικαταστάθηκαν από κυλινδρικούς πλαστικούς κλωβούς (Quayle *et al*., 1991). Ταυτόχρονα με την αλλαγή αυτή, συντελέστηκε και αλλαγή των οργάνων μέτρησης θερμοκρασίας, όπως αναφέρεται και παραπάνω. Οι δύο τύποι κλωβών εικονίζονται στο Σχήμα 2.3.

3) Μεταβολές των τοπικών μετεωρολογικών συνθηκών στη γειτονιά του μετεωρολογικού σταθμού.

Οι μεταβολές αυτές οφείλονται:

- σε αλλαγή χρήσης της γης στον περιβάλλοντα χώρο

Οι αλλαγές αυτές μπορεί να περιλαμβάνουν: καλλιέργειά ή εκχέρσωση μιας περιοχής με αποτέλεσμα διαφορετική επίδραση του εδάφους στη θερμοκρασία του αέρα σε χαμηλά ύψη, κατασκευή ή απομάκρυνση κατασκευών, αστικοποίηση ή βιομηχανοποίηση αγροτικών περιοχών (urbanization) κ.ά

- μείωση της έκθεσης των οργάνων στην ηλιακή ακτινοβολία ή σε ρεύματα αέρα

Η αλλαγή μπορεί να οφείλεται σε σκίαση λόγω ανάπτυξης ψηλών δέντρων πλησίον του καταγραφικού οργάνου, ανέγερσης κτιρίων (Alexandersson and Moberg, 1997) κ.ά.

- αύξηση της έκθεσης των οργάνων στην ηλιακή ακτινοβολία ή σε ρεύματα αέρα

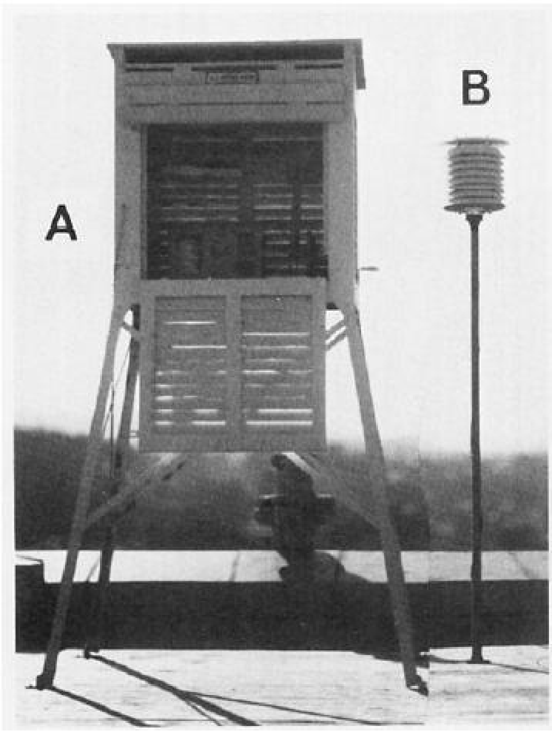
Αντίστοιχα με πριν, μπορεί να αφαιρεθεί με κάποιο τρόπο ο παράγοντας που εμποδίζει την έκθεση του σταθμού στις καιρικές συνθήκες (π.χ. κόψιμο δέντρων).

Από τους παραπάνω τρεις παράγοντες, οι δύο τελευταίοι (αύξηση και μείωση της έκθεσης σε ηλιακή ακτινοβολία και ρεύματα αέρα) σχετίζονται με φαινόμενα αύξησης ή μείωσης της καταγραφόμενης ελάχιστης ή μέγιστης θερμοκρασίας.

Ο παράγοντας του ανέμου επηρεάζει σημαντικά και τη μετρούμενη βροχόπτωση. Αλλαγές στο περιβάλλον του σταθμού που επηρεάζουν τον άνεμο στο σταθμό εισάγουν άλματα στα δεδομένα βροχόπτωσης (Beaulieu *et al*., 2007).

Οι μεταβολές των τοπικών μετεωρολογικών συνθηκών στη γειτονιά του μετεωρολογικού σταθμού όταν είναι απότομες θεωρείται ότι εισάγουν άλματα, ενώ όταν είναι βαθμιαίες (π.χ. αστικοποίηση) θεωρείται ότι εισάγουν κυρίως γραμμικές τάσεις στα δεδομένα.

Οι μεταβολές που περιγράφτηκαν όσον αφορά στις συνθήκες μέτρησης και στο περιβάλλον των σταθμών μέτρησης θερμοκρασίας δεν είναι καθόλου σπάνιες. Αντίθετα είναι εξαιρετικά διαδομένες. Εκτός από σταθμούς σε αγροτικές περιοχές, ελάχιστοι σταθμοί π.χ. δεν έχουν επηρεαστεί από την αλλαγή του άμεσου περιβάλλοντός τους (π.χ. Fall *et al*., 2011).



Σχήμα 2.3. *Εικονίζονται (A) χαρακτηριστικός ξύλινος κλωβός (CRS) που χρησιμοποιείτο μέχρι τη δεκαετία του 1980 στις ΗΠΑ και (B) πλαστικός κλωβός που αντικατέστησε τον αρχικό, φωτογραφημένοι στην ίδια κλίμακα. Πηγή: Quayle et al. (1991).*

2.1.3. Μεταβολή του ρυθμού δειγματοληψίας ή/και διαφορές στη μέθοδο υπολογισμού ημερήσιων τιμών

Οι μηνιαίες τιμές θερμοκρασίας και βροχόπτωσης προκύπτουν από τις αντίστοιχες ημερήσιες τιμές. Η διαδικασία με την οποία προκύπτουν οι ημερήσιες τιμές θερμοκρασίας και βροχόπτωσης διαφέρουν μεταξύ τους. Η ημερήσια βροχόπτωση συνήθως σημειώνεται με παρατήρηση του οργάνου κάθε 24 ώρες. Αντίθετα, η μέση ημερήσια θερμοκρασία προκύπτει συνήθως από κάποιες μετρήσεις ανά συγκεκριμένο διάστημα ή από θερμόμετρα μεγίστου – ελαχίστου.

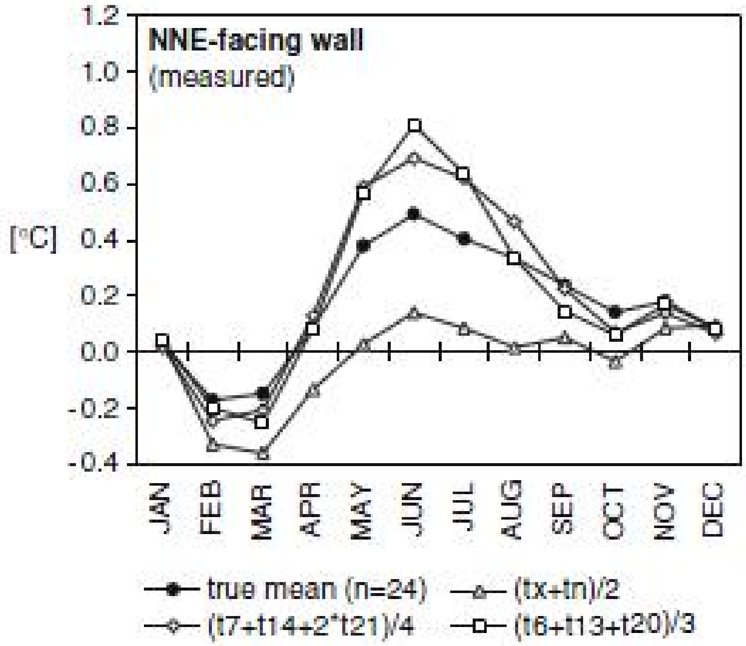
Οι ώρες δειγματοληψίας και ο αλγόριθμος υπολογισμού της μέσης ημερήσιας θερμοκρασίας διαφέρουν από εποχή σε εποχή και από χώρα σε χώρα, ακόμα και ανάμεσα σε δύο πόλεις τις ίδιας χώρας (Moberg and Bergström, 1997). Οι διαφορετικοί αυτοί υπολογισμοί λόγω της ετερογένειάς τους εισάγουν ανομογένειες στα δεδομένα.

Στον Πίνακα 2.1 παρουσιάζονται μερικοί από τους βασικότερους αλγόριθμους που έχουν χρησιμοποιηθεί ή και χρησιμοποιούνται ακόμα για τον υπολογισμό των μέσων ημερήσιων τιμών θερμοκρασίας, από τις οποίες προκύπτουν στη συνέχεια οι μέσες μηνιαίες τιμές. Παραδείγματα διαφορετικών μέσων μηνιαίων θερμοκρασιών που προκύπτουν από χρήση διαφορετικών τέτοιων αλγορίθμων φαίνονται στα Σχήματα 2.4 και 2.5.

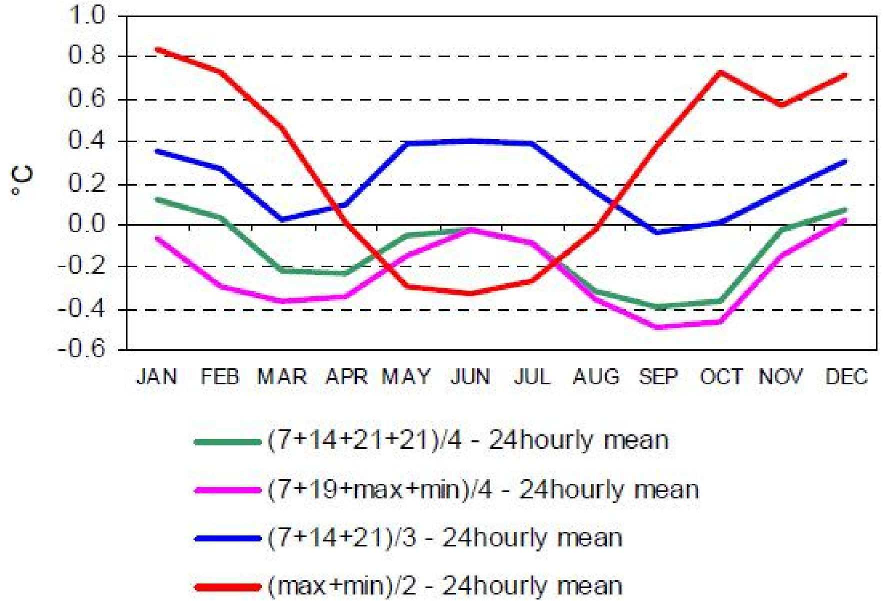
Πίνακας 2.1. *Αλγόριθμοι που έχουν χρησιμοποιηθεί σε διάφορα μέρη παγκοσμίως για τον υπολογισμό μέσων ημερήσιων θερμοκρασιών. Βιβλιογραφικές αναφορές:*
(1): Karl et al., 1986, (2): Böhm et al., 2010,(3): Moberg and Bergström, 1997,
(4): Founda et al., 2009

ΜΕΤΡΗΣΗ	**ΕΚΤΙΜΗΣΗ ΜΕΣΗΣ ΗΜΕΡΗΣΙΑΣ ΘΕΡΜΟΚΡΑΣΙΑΣ**	**ΒΙΒΛΙΟΓΡΑΦΙΑ**
t_{max}, t_{min}	$t= (t_{max}+t_{min})/2$	(1)
t_{07}, t_{14}, t_{21}	$t= (t_{07}+t_{14}+2\times t_{21})/4$	(2),(3)
t_{06}, t_{13}, t_{20}	$t= (t_{06}+t_{13}+ t_{20})/3$	(3)
t_{08}, t_{14}, t_{21}	$t= (t_{08}+t_{14}+5\times t_{21})/7$	(3)
t_{01}, t_{02},..., t_{23}, t_{24}	$t= (t_{01}+t_{02}+...+t_{24})/24$	(3),(4)

Αξίζει να αναφερθεί ότι ανομογένειες μπορεί να προκύπτουν ακόμα και όταν χρησιμοποιείται η ίδια μέθοδος υπολογισμού των μέσων ημερήσιων τιμών, λόγω διαφοράς ώρας ή λόγω της χρονικής στιγμής που θεωρείται ότι ξεκινάει το νέο 24ωρο π.χ. τα μεσάνυχτα ή στη δύση του ήλιου (Karl *et al.*, 1986).



Σχήμα 2.4. *Διαφορές στις μέσες μηνιαίες τιμές στο σταθμό Kremsmünster της Αυστρίας που προκύπτουν από τέσσερις διαφορετικούς αλγόριθμους υπολογισμού των μέσων ημερήσιων τιμών. Οι τιμές έχουν προκύψει από τη διαφορά των μετρήσεων σε παραδοσιακό και καινούριο κλωβό με χρήση σύγχρονων οργάνων. Πηγή: Böhm et al. (2010).*



Σχήμα 2.5. *Επιρροή διαφορετικών τρόπων υπολογισμού ημερήσιων μέσων θερμοκρασιών στον υπολογισμό μέσων μηνιαίων τιμών για το σταθμό Puchberg της Αυστρίας (1987-1996). Πηγή: Aguilar et al. (2003).*

3. ΟΜΟΓΕΝΟΠΟΙΗΣΗ ΥΔΡΟΚΛΙΜΑΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Η αναγνώριση ανομογενειών στα κλιματικά δεδομένα (θερμοκρασίες, βροχοπτώσεις κλπ) κατά το 19^{ο} αιώνα (early instrumental period) οδήγησε στις πρώτες απόπειρες αφαίρεσής τους με χρήση παράλληλων μετρήσεων οργάνων. Στις αρχές του 20^{ου} αιώνα, έγιναν προσπάθειες από τον Conrad για ομογενοποίηση των χρονοσειρών με χρήση των κριτηρίων των Heidke, Helmert και Abbe (Venema *et al.*, 2012).

Στα μέσα του εικοστού αιώνα ξεκινάει η ανάπτυξη πρωτότυπων μεθόδων ομογενοποίησης των κλιματικών χρονοσειρών με την ανάπτυξη και την ευρεία χρήση της Διπλής αθροιστικής καμπύλης (Kohler, 1949, Venema *et al.*, 2012). Από τότε μέχρι σήμερα έχουν προταθεί δεκάδες διαφορετικές προσεγγίσεις της ομογενοποίησης από ερευνητές σε όλο τον κόσμο. Μάλιστα σήμερα χρησιμοποιούνται ευρέως περίπου 20 διαφορετικές μέθοδοι ομογενοποίησης και αν λάβουμε υπόψη τις διαφορές στην πρακτική εφαρμογή τους, ο αριθμός αυτός γίνεται ακόμα μεγαλύτερος (Domonkos 2011a).

Οι ανομογένειες, όπως αναφέρθηκε και στο Κεφάλαιο 2, θεωρείται ότι εισάγονται στα δεδομένα με τη μορφή:

- μεμονωμένων μη αντιπροσωπευτικών τιμών,
- αλμάτων (shifts), δηλαδή μεταβολών στη μέση τιμή τμημάτων της χρονοσειράς, ή
- μεταβολών κλίσης (trends).

Οι μέθοδοι ομογενοποίησης έχουν ως στόχο να εντοπίσουν και να αφαιρέσουν τους παραπάνω τύπους ανομογενειών. Στο παρελθόν η επεξεργασία των κλιματικών δεδομένων εφαρμοζόταν συνηθέστερα σε δύο ή περισσότερα στάδια (π.χ. USHCN V1, http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html), τα οποία περιλάμβαναν:

1. ανίχνευση και αφαίρεση χονδροειδών σφαλμάτων - ακραίων τιμών (outliers),
2. αναγωγή δεδομένων σε τυποποιημένες συνθήκες μέτρησης (τυποποιημένους χρόνους παρατήρησης, όργανα αναφοράς κλπ) και διόρθωση καταγεγραμμένων αλλαγών, και
3. χρήση κάποιας στατιστικής μεθόδου για τον εντοπισμό της θέσης και τη διόρθωση καταγεγραμμένων ή μη αλμάτων ή διαφορών κλίσεων.

Μια πιο μοντέρνα προσέγγιση (π.χ. Menne and Williams, 2009) είναι η ταυτόχρονη αντιμετώπιση των βημάτων 2 και 3 με χρήση συνήθως μιας στατιστικής μεθόδου ομογενοποίησης. Πρόσφατη επίσης προσέγιση αποτελεί το σπάσιμο της χρονοσειράς σε τμήματα όταν εντοπίζεται κάποια ανομογένεια (scalpel method) και η αντιμετώπιση των τμημάτων ως ξεχωριστές χρονοσειρές (Rohde *et al*., 2012).

Όπως αναφέρεται και παραπάνω, μέχρι σήμερα έχουν αναπτυχθεί πολλές διαφορετικές μέθοδοι ομογενοποίησης και με διαφορετικούς στόχους, μπορούν όμως γενικά να καταταχθούν σε δύο κύριες κατηγορίες (Beaulieu *et al.*, 2007):

- **Υποκειμενικές μέθοδοι (subjective methods)**: η θέση της ανομογένειας εντοπίζεται με γυμνό μάτι σε ένα γράφημα (ακόμα και αν στη συνέχεια εφαρμόζονται στατιστικές μέθοδοι).

- **Αντικειμενικές μέθοδοι (objective methods)**: ο εντοπισμός των ανομογενειών δεν εξαρτάται από την κρίση του χρήστη.

Τα τελευταία χρόνια προτιμώνται οι αντικειμενικές μέθοδοι ομογενοποίησης, οι οποίες έχουν το πλεονέκτημα έναντι των υποκειμενικών, της αυτοματοποίησης της διαδικασίας εντοπισμού των ανομογενειών, χωρίς τη χρήση γραφημάτων. Αυτό φαίνεται από τη χρήση αντικειμενικών μεθόδων από μεγάλες ερευνητικές ομάδες όπως της NOAA (Menne and Williams, 2009), αλλά και κατά κύριο λόγο στις αξιολογήσεις της ομογενοποίησης (π.χ. Ducré-Robitaille *et al*., 2003).

Μια δεύτερη κατηγοριοποίηση που μπορεί να γίνει στις μεθόδους ομογενοποίησης είναι ανάμεσα σε απόλυτες και σχετικές - συγκριτικές (Sahin and Cigizoglu, 2010).

- **Απόλυτες μέθοδοι (absolute methods)**: εντοπίζουν ανομογένειες στα δεδομένα μιας χρονοσειράς με βάση μόνο τα δεδομένα της χρονοσειράς αυτής.
- **Σχετικές - συγκριτικές μέθοδοι (relative methods**): εντοπίζουν ανομογένειες συγκρίνοντας τις ελεγχόμενες χρονοσειρές (candidate series) με γειτονικές χρονοσειρές αναφοράς (reference series).

Οι σχετικές μέθοδοι ομογενοποίησης θεωρούν ότι γειτονικοί σταθμοί εκτίθενται στο ίδιο κλιματικό σήμα και γι'αυτό οι διαφορές/λόγοι τους μπορούν να χρησιμοποιηθούν για τον εντοπισμό ανομογενειών (Conrad and Pollak, 1962). Πιο παλιά συνηθιζόταν από τις γειτονικές χρονοσειρές στην ελεγχόμενη να συντίθεται μία χρονοσειρά αναφοράς, η οποία και να συγκρίνεται με την ελεγχόμενη χρονοσειρά. Τα τελευταία χρόνια, κυρίως λόγω της δυσκολίας εύρεσης ομογενών χρονοσειρών αναφοράς, συνηθίζεται η σύγκριση της ελεγχόμενης χρονοσειράς με όλες τις γειτονικές κατά ζεύγη (pairwise fashion, π.χ. Menne and Williams, 2009). Γενικά οι σχετικές μέθοδοι ομογενοποίησης χρησιμοποιούνται περισσότερο από τις απόλυτες, καθώς θεωρούνται πιο αξιόπιστες, εφόσον βέβαια το δίκτυο σταθμών είναι αρκετά πυκνό και υπάρχει καλή συσχέτιση μεταξύ των εξεταζόμενων σταθμών και των σταθμών αναφοράς (Sahin and Cigizoglu, 2010).

Στο εξής, όταν αναφερόμαστε σε σχετικές μεθόδους ομογενοποίησης θα συμβολίζουμε με X_i τις χρονοσειρές αναφοράς και με Y_i τις ελεγχόμενες χρονοσειρές.

3.1. Υποκειμενικές (subjective) μέθοδοι ομογενοποίησης

Οι κυριότερες υποκειμενικές μέθοδοι ομογενοποίησης που εμφανίζονται στη βιβλιογραφία είναι:

1. Διπλή αθροιστική καμπύλη
2. Γραφική εξέταση συσσωρευτικών αποκλίσεων
3. Γραφική εξέταση των χρονοσειρών των διαφορών ή των αναλογιών
4. Γραφική εξέταση φιλτραρισμένων δεδομένων

Οι μέθοδοι αυτές συγκεντρώθηκαν και παρουσιάστηκαν από τους Beaulieu *et al*. (2007) και είναι όλες σχετικές – συγκριτικές (χρησιμοποιούν δηλαδή χρονοσειρές αναφοράς) με εξαίρεση τη γραφική εξέταση φιλτραρισμένων δεδομένων, η οποία είναι απόλυτη.

Η σημαντικότερη και πιο ευρέως διαδεδομένη είναι η Διπλή αθροιστική καμπύλη, η οποία και θα περιγραφεί αναλυτικά στην Παράγραφο 3.1.1. Κλασική (και με ευρεία χρήση μέχρι σήμερα, π.χ. Auer *et al*., 2005) μέθοδος ομογενοποίησης είναι και η γραφική εξέταση συσσωρευτικών αποκλίσεων (Craddock, 1979).

3.1.1. Διπλή αθροιστική καμπύλη

Η Διπλή αθροιστική καμπύλη αναπτύχθηκε στα μέσα του εικοστού αιώνα (Kohler, 1949) και χρησιμοποιήθηκε ευρύτατα για την ομογενοποίηση υδροκλιματικών δεδομένων, κυρίως βροχοπτώσεων και απορροών (π.χ. Searcy and Hardison,1960).

Έστω y_i οι τιμές ετήσιας βροχόπτωσης στο σταθμό που ελέγχεται για ανομοιογένειες και x_i οι τιμές της ετήσιας βροχόπτωσης σε γειτονικό σταθμό αναφοράς (ή η μέση τιμή από περισσότερους από έναν σταθμούς αναφοράς στην ίδια περιοχή). Για τη δημιουργία της Διπλής αθροιστικής καμπύλης υπολογίζουμε τις τιμές Σy_i και Σx_i σύμφωνα με τις Σχέσεις 3.1 και 3.2:

$$\sum y_i = y_1 + y_2 + ... + y_i \quad (3.1)$$

$$\sum x_i = x_1 + x_2 + ... + x_i \quad (3.2)$$

όπου i=1,2,…,ν και ν το κοινό μήκος των χρονοσειρών αναφοράς και εξεταζόμενης.

Η Διπλή αθροιστική καμπύλη είναι το γράφημα συσσωρευτικών αθροισμάτων των ετήσιων βροχοπτώσεων Σy_i στο σταθμό που ελέγχεται (Σχέση 3.1) έναντι των συσσωρευτικών αθροισμάτων των ετήσιων βροχοπτώσεων Σx_i (Σχέση 3.2) για την ίδια χρονική περίοδο.

Αν οι σταθμοί βρίσκονται σε κοντινή απόσταση και σε περιοχή με ίδιες κλιματικές συνθήκες, οι ολικές ετήσιες τιμές των βροχοπτώσεων αναμένεται να συσχετίζονται μεταξύ τους. Εφόσον συντρέχουν οι προηγούμενοι παράγοντες, οι αθροιστικές τιμές Σy_i και Σx_i αναμένεται να έχουν γραμμική σχέση. Απόκλιση από την αναλογία θεωρείται συστηματικό σφάλμα (ανομογένεια), το οποίο πρέπει να διορθωθεί. Η απόκλιση αυτή συνήθως εμφανίζεται με τη μορφή αλλαγής της κλίσης των σημείων του διαγράμματος (Dingman, 1994, Koutsoyiannis and Langousis, 2011). Η διαδοχική άθροιση των σημείων x_i και y_i ώστε να υπολογιστούν οι τιμές Σx_i και Σy_i μπορεί να γίνει είτε από την παλαιότερη προς την πιο πρόσφατη ημερομηνία (Searcy and Hardison, 1960, Dingman, 1994) είτε από την πιο πρόσφατη ημερομηνία προς την παλαιότερη (Chang and Lee, 1974, Koutsoyiannis and Langousis, 2011). Ένα παράδειγμα της δεύτερης μεθόδου εμφανίζεται στο Σχήμα 3.5.

Η αλλαγή κλίσης που εντοπίζεται στη Διπλή αθροιστική καμπύλη θεωρείται συνήθως σημαντική και διορθώνεται αν έχει μήκος ίσο ή μεγαλύτερο από πέντε χρόνια και εφόσον:

1. συνδέεται με κάποια καταγεγραμμένη αλλαγή στο αρχείο του σταθμού, και
2. είναι στατιστικά σημαντική.

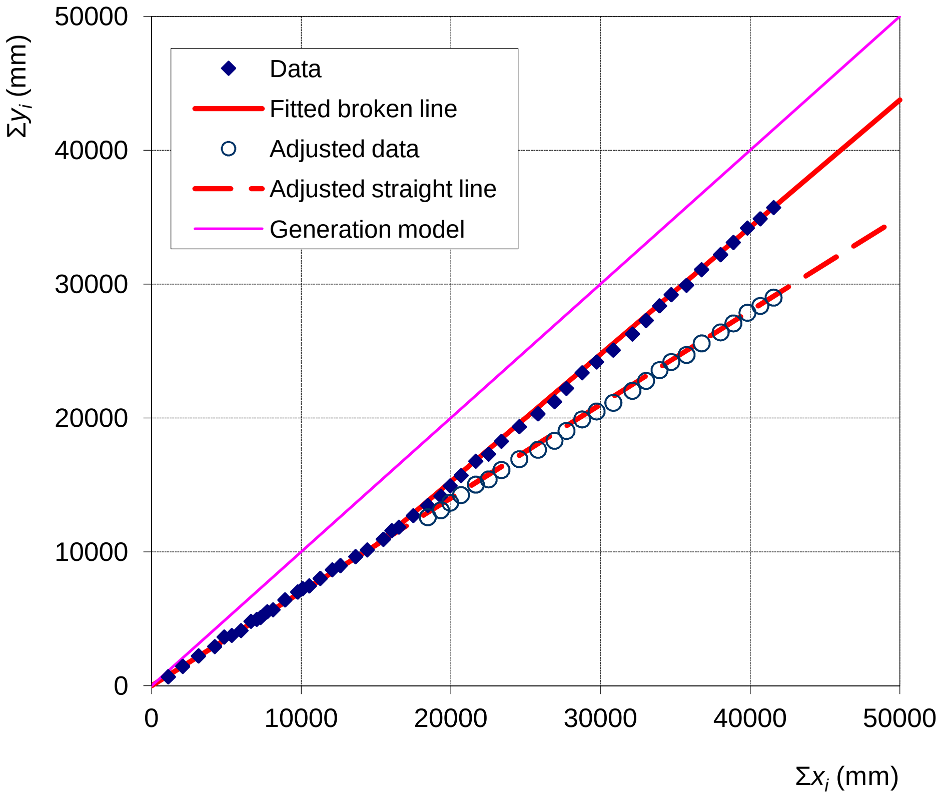
Η στατιστική σημαντικότητα των αλλαγών κλίσης μπορεί να εκτιμηθεί με χρήση τεχνικών που ονομάζονται αναλύσεις μεταβλητότητας - συμμεταβλητότητας (Searcy and Hardison 1960). Οι αναλύσεις αυτές δεν θεωρούνται αξιόπιστες από τον Dingman (1994), καθώς θεωρεί ότι η διαδικασία της άθροισης εισάγει ένα υψηλό βαθμό συσχέτισης ανάμεσα στις διαδοχικές τιμές.

Η διόρθωση των αλλαγών κλίσης στη διπλή αθροιστική καμπύλη γίνεται συνήθως με θεώρηση ότι οι πιο πρόσφατες μετρήσεις είναι πιο σωστές και η ανομογένεια βρίσκεται στα παλιότερα δεδομένα (Dingman, 1994, Koutsoyiannis and Langousis, 2011). Τα δεδομένα της περιόδου πριν την αλλαγή κλίσης πολλαπλασιάζονται με τον όρο

$\lambda = m/m'$, όπου m: κλίση τη χρονική περίοδο μετά την αλλαγή

m΄: κλίση τη χρονική περίοδο πριν την αλλαγή

Στο Σχήμα 3.1 φαίνεται ένα παράδειγμα διόρθωσης μιας συνθετικής χρονοσειράς ετήσιων βροχοπτώσεων από τους Koutsoyiannis and Langousis (2011) με χρήση της Διπλής αθροιστικής καμπύλης. Το διάγραμμα αυτό τονίζει ένα σημαντικό πρόβλημα στο οποίο εστιάζεται η παρούσα Μεταπτυχιακή εργασία και εξετάζεται με λεπτομέρεια στα επόμενα Κεφάλαια.



Σχήμα 3.1. *Παράδειγμα εντοπισμού και διόρθωσης, με χρήση της Διπλής αθροιστικής καμπύλης, πλασματικής ανομογένειας σε ομογενή ετήσια συνθετική χρονοσειρά βροχόπτωσης με μακροπρόθεσμη εμμονή. Ο εντοπισμός της πλασματικής ανομογένειας έγινε με σύγκριση με επίσης ομογενή συνθετική χρονοσειρά αναφοράς. Πριν τη διόρθωση οι κλίσεις στο αριστερό (νεότερες τιμές) και το δεξί (παλιότερες τιμές) τμήμα της χρονοσειράς είναι αντίστοιχα m=0.737 και m΄=0.95. Πηγή: Koutsoyiannis and Langousis (2011).*

Παράχθηκαν δύο ομογενείς χρονοσειρές, X (αναφοράς) και Y (εξεταζόμενη), με επιλεγμένα χαρακτηριστικά (μήκος 50 έτη, μέση τιμή 1000mm, τυπική απόκλιση 250mm, συντελεστή Hurst H=0.75 και μεταξύ τους συσχέτιση 0.82). Η διαδικασία ομογενοποίησης της εξεταζόμενης χρονοσειράς Y εμφάνισε μια ψευδή ανομογένεια (false alarm) στον όρο 25 με αποτέλεσμα να διορθωθεί η χρονοσειρά Y, αν και ήταν εξαρχής ομογενής.

Παραλλαγή της μεθόδου αποτελεί η μέθοδος των Chang and Lee (1974) οι οποίοι πρότειναν μια αντικειμενική (δηλαδή με εύρεση της μεγαλύτερης αλλαγής κλίσης με κάποιο αυτόματο τρόπο και όχι με οπτική επισκόπηση του γραφήματος) εκδοχή της Διπλής αθροιστικής καμπύλης με ενσωμάτωση της ανάλυσης μεταβλητότητας - συμμεταβλητότητας.

3.2. Αντικειμενικές (objective) μέθοδοι ομογενοποίησης

Οι αντικειμενικές μέθοδοι ομογενοποίησης ακολουθούν γενικά μια πιο μοντέρνα λογική από τις υποκειμενικές μεθόδους, καθώς επιτρέπουν την ευκολότερη επεξεργασία των δεδομένων χωρίς τη χρήση γραφημάτων. Οι αντικειμενικές μέθοδοι ομογενοποίησης που αναπτύχθηκαν σύμφωνα με τους Beaulieu *et al*. (2007) και Venema *et al*. (2012) βασίζονται στις εξής τεχνικές:

- κλασικά στατιστικά τεστ (π.χ. Buishand 1982, 1984, Alexandersson, 1986, Karl and Williams, 1987, Jaruskova, 1996, Alexandersson and Moberg, 1997, Wang *et al*., 2007, Wang, 2008 κ.ά.),
- μοντέλα γραμμικής παλινδρόμησης (π.χ. Easterling and Peterson, 1995, Vincent, 1998), ή
- Μπεϋζιανές προσεγγίσεις (π.χ. Lee and Heghinian, 1977, Ouarda et al., 1999, Perreault et al., 1999, 2000).

Η πρώτη από τις παραπάνω κατηγορίες μεθόδων βασίζεται στη λογική των κλασικών στατιστικών τεστ, κυρίως των Student-t και F-test. Τα τεστ αυτά έχουν στόχο να εκτιμήσουν αν δύο τμήματα μιας χρονοσειράς έχουν διαφορετικό μέσο όρο (Student-t test) και συνεπώς εκφράζουν ένα άλμα (offset) ή αντιστοιχούν σε χρονοσειρές με διαφορετικές τυπικές αποκλίσεις (F-test). Τα Student-t και F-test μπορεί να χρησιμοποιούνται και τα ίδια ως μέθοδοι ομογενοποίησης (Karl and Williams, 1987, Štěpánek and Zahradníček, 2009) ή επικουρικά σε κάποιες περιπτώσεις για να αξιολογηθεί η στατιστική σημαντικότητα κάποιων μεταβλητών που περιλαμβάνουν οι μέθοδοι ομογενοποίησης (π.χ. Easterling and Peterson, 1995).

Οι παραπάνω κατηγορίες μεθόδων εντοπίζουν συνήθως τη **σημαντικότερη ανομογένεια** στα δεδομένα που ελέγχουν. Σύμφωνα με τους Venema *et al*. (2012), οι οποίοι έκαναν την πιο πρόσφατη αξιολόγηση μεθόδων ομογενοποίησης (βλ. Κεφάλαιο 5), η πιο σύγχρονη τάση στην προσπάθεια ομογενοποίησης των υδροκλιματικών δεδομένων είναι η ανάπτυξη μεθόδων οι οποίες εντοπίζουν και διορθώνουν **πολλαπλές ανομογένειες** με χρήση μη ομογενών χρονοσειρών αναφοράς (π.χ. Szentimrey, 1999, Mestre, 1999, Caussinus and Mestre, 2004, Menne and Williams, 2009). Στο πλαίσιο αυτό, πολλές μέθοδοι που είχαν αναπτυχθεί για τον εντοπισμό μιας μόνο ανομογένειας χρησιμοποιούνται για τον εντοπισμό πολλαπλών ανομογενειών. Αυτό επιτυγχάνεται με διαδοχικό «σπάσιμο» της χρονοσειράς κάθε φορά που εντοπίζεται κάποια πιθανή ανομογένεια και εφαρμογή εκ νέου της μεθόδου στα τμήματα που προκύπτουν (Moberg and Alexandersson, 1997, Beaulieu *et al.*, 2007, Domonkos, 2011a).

Οι περισσότερες αντικειμενικές μέθοδοι ομογενοποίησης που χρησιμοποιούνται σήμερα εντοπίζουν ένα ή περισσότερα άλματα (Domonkos, 2011a). Λίγες από τις μεθόδους που υπάρχουν μπορούν να εντοπίσουν και ανομογένειες με μορφή κλίσης μέσα στα δεδομένα (π.χ. Alexandersson and Moberg, 1997, Menne and Williams, 2009).

Μερικές από τις σημαντικότερες και πιο ευρέως χρησιμοποιούμενες αντικειμενικές μεθόδους ομογενοποίησης υδροκλιματικών δεδομένων παρατίθενται και αναλύονται στις Παραγράφους 3.2.1-3.2.4. Οι μέθοδοι αυτές είναι σχετικές - συγκριτικές και επιλέχτηκαν με βάση:

- την ευρύτητα της εφαρμογής τους στην πράξη (π.χ. η μέθοδος των Easterling and Peterson (1995) χρησιμοποιήθηκε για την ομογενοποίηση χιλιάδων σταθμών της NOAA), και
- την αναφορά, χρήση και αξιολόγησή τους από μεγάλο πλήθος βιβλιογραφικών πηγών (π.χ. Wijngaard et al., 2003, Sahin and Cigizoglu, 2010, Venema *et al.*, 2012).

Οι μέθοδοι αυτές είναι:

1. Easterling and Peterson test (Easterling and Peterson, 1995)
2. Standard Normal Homogeneity Test (SNHT) for single shifts (Alexandersson, 1986, Alexandersson and Moberg, 1997)
3. Standard Normal Homogeneity Test (SNHT) for trends (Alexandersson and Moberg, 1997)
4. Pairwise Homogenization Algorithm (PHA, Menne and Williams, 2009)

3.2.1. Μέθοδος των Easterling and Peterson (1995)

Η μέθοδος των Easterling and Peterson (1995) είναι μια σχετική μέθοδος ομογενοποίησης και χρησιμοποιεί για τον εντοπισμό των πιθανών ανομογενειών γραμμικό μοντέλο παλινδρόμησης. Οι σταθμοί αναφοράς επιλέγονται ώστε να έχουν υψηλή ετεροσυσχέτιση με τον εξεταζόμενο σταθμό. Αναλυτικά, η διαδικασία που ακολουθείται κατά τη μέθοδο είναι η παρακάτω:

1) Αρχικά, γίνεται η επιλογή των σταθμών αναφοράς (reference stations) από τους γειτονικούς σταθμούς του ελεγχόμενου σταθμού (candidate series). Για κάθε γειτονικό σταθμό εξετάζεται η συσχέτιση των ετήσιων μεταβολών της θερμοκρασίας (dT/dt) με αυτές του εξεταζόμενου σταθμού. Από το σύνολο των πιθανών σταθμών αναφοράς επιλέγονται οι πέντε καλύτερα συσχετισμένοι σταθμοί με τον εξεταζόμενο σταθμό και κατασκευάζεται μια μοναδική συνθετική σειρά αναφοράς (reference series) που θεωρείται ότι εκφράζει τις πραγματικές κλιματικές μεταβολές και χρησιμοποιείται στην όλη ανάλυση.

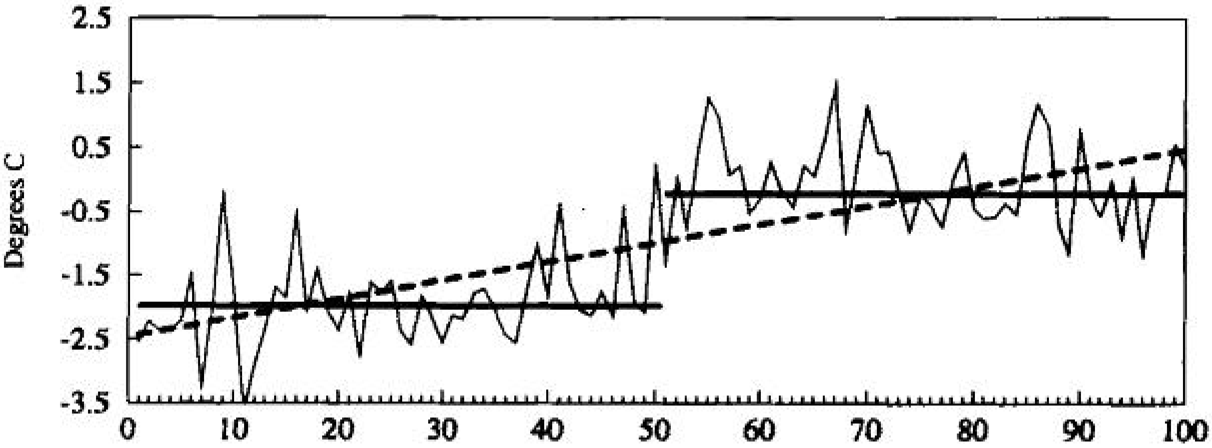
2) Δημιουργείται η χρονοσειρά των διαφορών μεταξύ της συνθετικής σειράς αναφοράς και της ελεγχόμενης χρονοσειράς. Η χρονοσειρά των διαφορών απεικονίζεται σε διάγραμμα, με τεταγμένη τις διαφορές και τετμημένη το χρόνο.

3) Στη χρονοσειρά των διαφορών προσαρμόζεται ευθεία γραμμή και υπολογίζεται το άθροισμα των τετραγώνων των υπολοίπων των διαφορών των αντίστοιχων τιμών (residual sum of squares- RSS_1) όλης της χρονοσειράς.

4) Στη συνέχεια ακολουθείται μια επαναληπτική διαδικασία για την εύρεση των σημείων στα οποία εμφανίζονται ανομογένειες: Για καθένα από τα έτη i της χρονοσειράς, όπου i=1,2,...,ν, η χρονοσειρά χωρίζεται σε δύο τμήματα: ένα πριν το έτος i και ένα μετά το έτος i. Σε καθένα από τα δύο τμήματα προσαρμόζεται μια ευθεία ελαχίστων τετραγώνων και υπολογίζεται το άθροισμα των τετραγώνων των υπολοίπων (RSS). Με πρόσθεση των δύο RSS προκύπτει το RSS_2. Έτσι, σε κάθε έτος i της χρονοσειράς αντιστοιχίζεται μια τιμή RSS_2.

5) Από όλα τα έτη της χρονοσειράς, η θέση i στην εξεταζόμενη χρονοσειρά στην οποία αντιστοιχεί το ελάχιστο RSS_2 σημειώνεται ως θέση πιθανής ασυνέχειας.

Στο Σχήμα 3.2 παρουσιάζεται ένα παράδειγμα εντοπισμού πιθανού άλματος από τη μέθοδο των Easterling and Peterson (1995). Εικονίζεται η προσαρμογή τριών ευθειών στα δεδομένα, σε όλο το μήκος της χρονοσειράς, πριν την πιθανή ανομογένεια και μετά την πιθανή ανομογένεια.



Σχήμα 3.2. *Παράδειγμα εύρεσης ασυνέχειας σε συνθετικά δεδομένα από τους Easterling and Peterson (1995) με τη μέθοδο που ανέπτυξαν. Εικονίζεται χρονοσειρά διαφορών ανάμεσα σε συνθετικό ζεύγος χρονοσειράς αναφοράς - εξεταζόμενης, με γραμμική παρεμβολή για όλη τη χρονοσειρά και ξεχωριστά για τα τμήματα πριν και μετά την ασυνέχεια. Η ασυνέχεια έχει μέγεθος* 2^o *C.*

6) Για να ελεγχθεί η στατιστική σημαντικότητα της προσαρμογής με δύο καμπύλες (two-phase fit), χρησιμοποιείται η ακόλουθη αναλογία πιθανοφάνειας (likelihood ratio statistic) από τον Solow (1987):

$$U = \left[\left(RSS_1 - RSS_2\right)/3\right]/\left[RSS_2/\left(n-4\right)\right] \quad (3.3)$$

όπου η μεταβλητή U θεωρείται ότι ακολουθεί κατανομή F με 3 και n-4 βαθμούς ελευθερίας.

7) Η στατιστική σημαντικότητα της διαφοράς των δύο μέσων, πριν και μετά την ασυνέχεια, ελέγχεται και με ένα δεύτερο τρόπο, με χρήση του Student *t*-test.

8) Αν με μία τουλάχιστον από τις παραπάνω δύο τεχνικές (βήματα 6 και 7) έχει προκύψει ότι η ασυνέχεια είναι στατιστικά σημαντική, τότε αυτή δεν διορθώνεται αμέσως, αλλά σημειώνεται ως *πιθανή* ασυνέχεια.

9) Αφού έχουν επισημανθεί οι πιθανές ασυνέχειες, η στατιστική σημαντικότητα των ασυνεχειών ελέγχεται με μια επιπλέον τεχνική που ονομάζεται διαδικασία αναδιάταξης πολλαπλούς απόκρισης - multiresponse permutation procedures (MRPP, Mielke *et al*., 1981).

Η **μέθοδος MRPP** είναι μια μη-παραμετρική μέθοδος η οποία εκτιμά αν ο χωρισμός ενός συνόλου δεδομένων σε δύο υποσύνολα με βάση κάποια συγκεκριμένα χαρακτηριστικά είναι στατιστικά σημαντικός. Η μέθοδος συγκρίνει τις Ευκλείδιες αποστάσεις ανάμεσα στα μέλη καθεμιάς από τις δύο ομάδες με τις αποστάσεις ανάμεσα σε όλα τα μέλη των δύο ομάδων ώστε να εκτιμήσει εάν οι διαφορές μεταξύ των δύο ομάδων είναι τυχαίες.

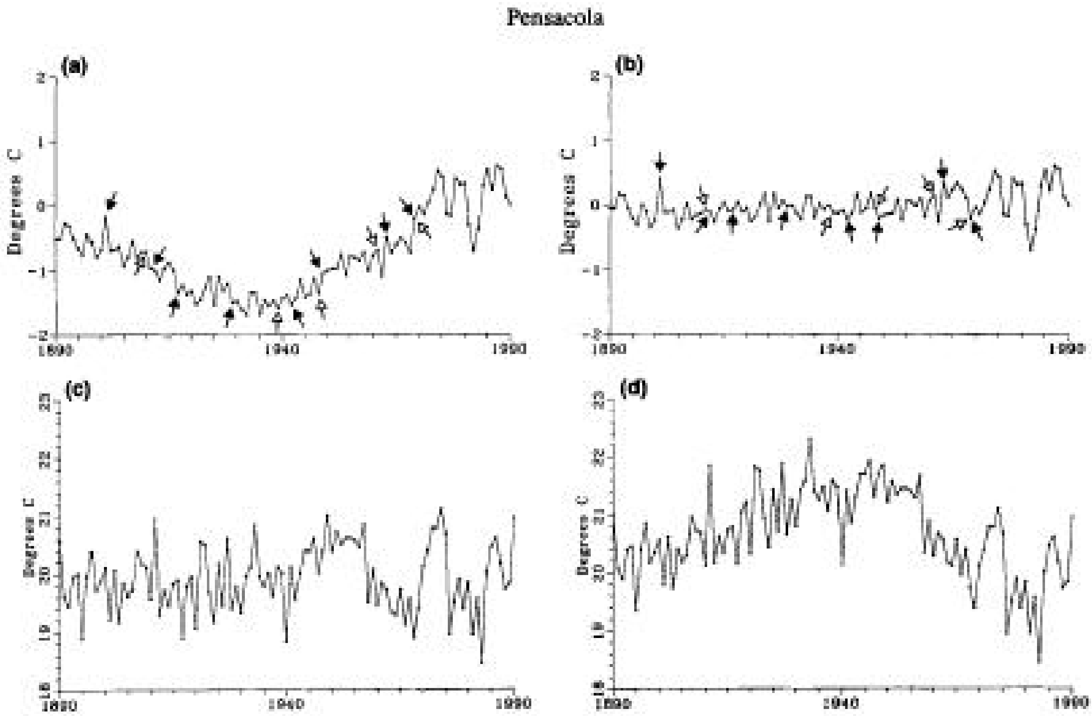
Στην περίπτωση της μεθόδου των Easterling-Peterson, οι δύο ομάδες αποτελούνται από τις τιμές της χρονοσειράς διαφορών πριν (D_b) και μετά (D_a) την πιθανή ασυνέχεια και οι αποστάσεις είναι οι διαφορές ανάμεσα σε οποιεσδήποτε δύο τιμές της χρονοσειράς διαφορών. Η μέθοδος εφαρμόζεται λαμβάνοντας διάστημα 12 ετών και από τις δύο πλευρές μιας πιθανής ασυνέχειας και ελέγχοντας τη στατιστική σημαντικότητά της με επίπεδο εμπιστοσύνης 95% εκτός αν εμφανίζεται άλλη πιθανή ασυνέχεια σε μικρότερο χρονικό διάστημα, οπότε και επιλέγεται μικρότερο διάστημα.

Περισσότερες λεπτομέρειες σε σχέση με τον τρόπο εφαρμογής της μεθόδου MRPP δίνονται από τους Easterling and Peterson (1995) και Mielke *et al*. (1981).

10) Εφ’όσον εντοπιστούν σημαντικές ασυνέχειες, γίνεται διόρθωση χρησιμοποιώντας διαστήματα 12 ετών. Για παράδειγμα έστω ότι δύο ασυνέχειες εντοπίζονται στις χρονιές 31 και 61 μιας χρονοσειράς 100 ετών. Η πρώτη διόρθωση θα γίνει για την πιο πρόσφατη ασυνέχεια. Θα υπολογιστούν οι μέσες τιμές των χρονοσειρών για τα χρόνια 49-60 και 61-72. Η διόρθωση που θα προκύψει από τη διαφορά των δύο μέσων θα εφαρμοστεί στη χρονοσειρά από το χρόνο 1 μέχρι το χρόνο 60. Στη συνέχεια θα γίνει η διόρθωση της δεύτερης ασυνέχιας στη χρονιά 31 με τον ίδιο τρόπο.

11) Η διορθωμένη σειρά που θα προκύψει εξετάζεται εκ νέου για τυχόν εναπομείνασες ασυνέχειες με την διαδικασία που έχει περιγραφεί. Η διόρθωση των ασυνεχειών γίνεται όπως στο βήμα 10. Η διαδικασία επαλαμβάνεται μέχρις ότου να μην εντοπίζονται άλλες ασυνέχειες, οπότε και η σειρά θεωρείται ομογενοποιημένη.

Ένα παράδειγμα ομογενοποίησης χρονοσειρών με τη μέθοδο των Easterling and Peterson (1995) παρουσιάζεται παραπάνω στο Σχήμα 3.3.



__Σχήμα 3.3.__ Παραδείγμα ομογενοποίησης με χρήση της μεθόδου Easterling-Peterson ετήσιας χρονοσειράς από την Pensacola, Florida των ΗΠΑ.
(a) ανεπεξέργαστη χρονοσειρά διαφορών μεταξύ σειράς αναφοράς και εξεταζόμενης,
(b) επεξεργασμένη χρονοσειρά διαφορών,
(c) ανεπεξέργαστη εξεταζόμενη χρονοσειρά, και
(d) επεξεργασμένη εξεταζόμενη χρονοσειρά
Με τα μαύρα βέλη επισημαίνονται οι ασυνέχειες που αναγνωρίστηκαν με τη μέθοδο Easterling-Peterson και με τα άσπρα βέλη οι πιθανές ασυνέχειες που αναγνωρίστηκαν από τα ιστορικά δεδομένα κάθε σταθμού. Πηγή: Easterling and Peterson (1995).

3.2.2. Standard Normal Homogeneity Test for single shifts

Η μέθοδος αυτή αρχικά δημιουργήθηκε για ανάλυση δεδομένων βροχοπτώσεων, αλλά εφαρμόζεται και σε δεδομένα θερμοκρασιών. Αν και είναι συγκριτική έχει χρησιμοποιηθεί και σε απόλυτη μορφή (π.χ. Sahin and Cigizoglu, 2010).

Στη συνήθη της μορφή είναι σχετική – συγκριτική: Αρχικά δημιουργείται από την εξεταζόμενη (candinate) χρονοσειρά Y και τις περιβάλουσες χρονοσειρές αναφοράς X μια χρονοσειρά Q στην οποία μετά από κανονικοποίηση εφαρμόζεται η μέθοδος.

Στις παρακάτω σχέσεις, με Y_i δηλώνεται μια συγκεκριμένη τιμή της εξεταζόμενης χρονοσειράς (ετήσια συναθροισμένη βροχόπτωση ή μέση ετήσια θερμοκρασία) τη χρονιά i. Με X_j δηλώνεται μία από τις περιβάλουσες reference series (η j σε σειρά από σύνολο k) και με X_{ji} μία συγκεκριμένη τιμή από αυτό το σταθμό. Με ρ_j συμβολίζεται ο συντελεστής ετεροσυσχέτισης ανάμεσα στην εξεταζόμενη χρονοσειρά και σε μια σειρά από έναν από τους περιβάλλοντες σταθμούς.

Για τον εντοπισμό των σχετικών ανομογενειών χρησιμοποιούνται αναλογίες (παραδοσιακά χρησιμοποιούμενες σε χρονοσειρές βροχοπτώσεων, Σχέση 3.4) ή διαφορές (για τα θερμοκρασιακά δεδομένα, Σχέση 3.5).

$$Q_i = Y_i / \left\{ \left[\sum_{j=1}^{k} \circ_j^2 X_{ji} \bar{Y} / \bar{X}_j \right] / \sum_{j=1}^{k} \circ_j^2 \right\} \qquad (3.4)$$

$$Q_i = Y_i - \left\{ \sum_{j=1}^{k} \circ_j^2 \left[X_{ji} - \bar{X}_j + \bar{Y} \right] / \sum_{j=1}^{k} \circ_j^2 \right\} \qquad (3.5)$$

Οι μέσες τιμές των Y και X_j υπολογίζονται για μια κοινή περίοδο για ολους τους σταθμούς, δηλαδή για όλα τα j=1,…,k.

Το SNHT στη μορφή που το εξετάζουμε εφαρμόζεται σε κανονικοποιημένες χρονοσειρές όπως παρακάτω. Οι τυπικές αποκλίσεις έχουν (n-1) βαθμούς ελευθερίας.

$$Z_i = (Q_i - \bar{Q}) / s_Q \qquad (3.6)$$

Εναλλακτικά η μέθοδος μπορεί να εφαρμοστεί και κατευθείαν στη χρονοσειρά Y (απόλυτη μέθοδος). Η χρονοσειρά Z δηλαδή προκύπτει από κανονικοποίηση της Y.

Εφαρμογή του SNHT for single shifts:

Η ύπαρξη ενός μοναδικού άλματος στην εξεταζόμενη χρονοσειρά Y μπορεί να εκφραστεί με μια μηδενική υπόθεση (H_0) και μια εναλλακτική υπόθεση (H_1):

$$H_0 : Z_i \in N(0,1) \quad i \in \{1,...,n\} \qquad (3.7)$$

$$H_1 : \begin{cases} Z_i \in N(\text{«}_1, 1) & i \in \{1,...,a\} \\ Z_i \in N(\text{«}_2, 1) & i \in \{a+1,...,n\} \end{cases} \qquad (3.8)$$

όπου α το πιο πιθανό σημείο ύπαρξης άλματος και n το πλήθος των όρων της χρονοσειράς.

Η στατιστική τιμή T_0 ελέγχου (test value) που χρησιμοποιεί το SNHT ορίζεται από τη Σχέση:

$$T_{\max}^{s} = \max_{1 \le a \le n-1} \left\{ T_a^s \right\} = \max_{1 \le a \le n-1} \left[a\bar{z}_1^2 + (n-a)\bar{z}_2^2 \right] \qquad (3.9)$$

Με $\bar{z}_1$ συμβολίζεται η μέση τιμή των όρων της χρονοσειράς πριν την ασυνέχεια και με $\bar{z}_2$ η μέση τιμή των όρων της χρονοσειράς μετά την ασυνέχεια.

Οι κρίσιμες τιμές του όρου $T_{\max}$ κατά τους Alexandersson and Moberg (1997) παρουσιάζονται στον Πίνακα 3.2.

Σύμφωνα με τους άνω συγγραφείς, αν γνωρίζουμε ότι υπάρχει μόνο μια πιθανή θέση ασυνέχειας, μπορεί αντί εφαρμοστεί απευθείας το t-test στη χρονοσειρά Q_i, χωρίς προηγουμένως να γίνει κανονικοποίηση της χρονοσειράς.

Για να γίνει διόρθωση της εξεταζόμενης χρονοσειράς υπολογίζονται οι τιμές

$$\bar{q}_1 = \pm_Q \bar{z}_1 + \bar{Q} \qquad \text{και} \qquad \bar{q}_2 = \pm_Q \bar{z}_2 + \bar{Q} \qquad (3.10)$$

Αν επιθυμούμε να διορθώσουμε την περίοδο {1,2....α} τότε:

- Για δεδομένα βροχοπτώσεων, οι τιμές Q_i της περιόδου {1,2....α} πολλαπλασιάζονται με την τιμή $¢ = \overline{q}_2 / \overline{q}_1$ (3.11)
- Για δεδομένα θερμοκρασιών, στις τιμές Q_i της περιόδου {1,2....α} προστίθεται η τιμή $¢ = \overline{q}_2 - \overline{q}_1$ (3.12)

Αν τα δεδομένα παριέχουν μόνο ένα άλμα στη μέση τιμή, τότε με τη διαδικασία αυτή λαμβάνουμε μια ομογενοποιημένη χρονοσειρά όπου όλα τα δεδομένα αναφέρονται στην πιο πρόσφατη κατάσταση μέτρησης.

Η παραπάνω μεθοδολογία περιγράφει τη διόρθωση ενός μόνο άλματος από την εξεταζόμενη χρονοσειρά. Σε πολλές περιπτώσεις όμως οι χρονοσειρές που ελέγχονται περιέχουν παραπάνω από ένα άλματα ή και διαφορές κλίσεων. Κάποιοι τρόποι εφαρμογής της μεθόδου SNHT for single shifts που έχουν προταθεί για τη διόρθωση τέτοιων περιπτώσεων αναφέρονται στο τέλος της Παραγράφου 3.2.4 μετά από μια συνοπτική παρουσίαση της (συμπληρωματικής) μεθόδου Standard Normal Homogeneity Test for trends.

3.2.3. Standard Normal Homogeneity Test for trends

Η κατασκευή της χρονοσειράς Z είναι κοινή για το SNHT for single shifts και το SNHT for trends. Όμως η μηδενική (H_0) και η εναλλακτική υπόθεση (H_1) παίρνουν στην περίπτωση του SNHT for trends την παρακάτω μορφή:

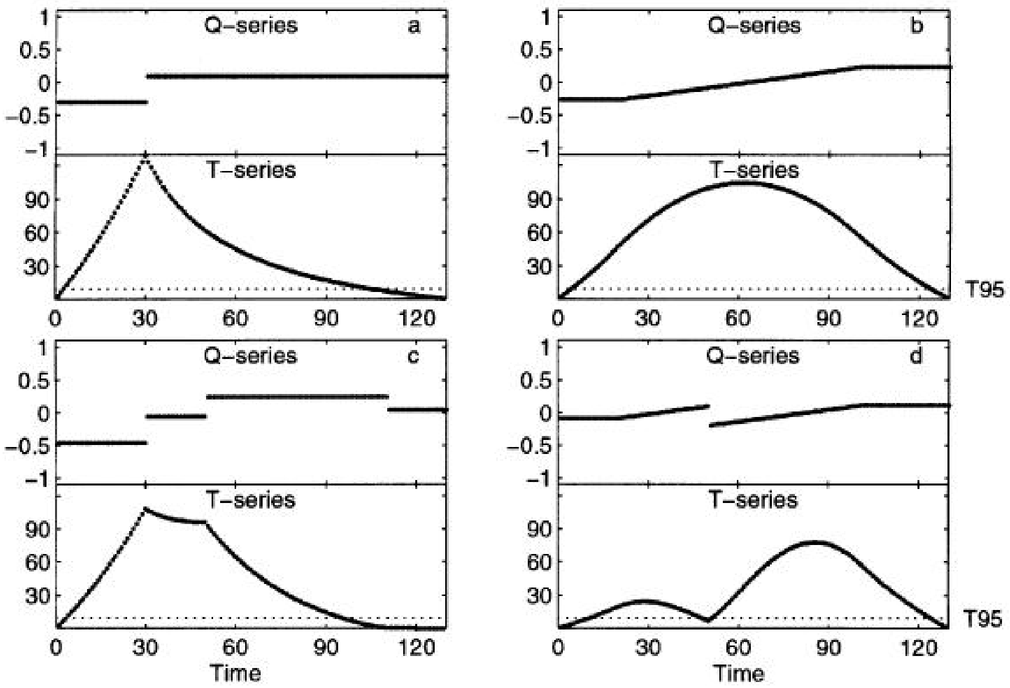
$$H_0 : Z_i \in N(0,1) \quad i \in \{1,...,n\} \tag{3.13}$$

$$H_1 : \begin{cases} Z_i \in N(«_1,1) & i \in \{1,...,a\} \\ Z_i \in N(«_1 + (1-a)(«_2 - «_1)(b-a),1) & i \in \{a+1,...,b\} \\ Z_i \in N(«_2,1) & i \in \{b+1,...,n\} \end{cases} \tag{3.14}$$

όπου α το πιο πιθανό σημείο μεταβολής της τάσης, b το πιο πιθανό σημείο στο οποίο ξαναμηδενίζεται η τάση και n το πλήθος των όρων της χρονοσειράς.

Το SNHT for trends εφαρμόζεται με παρόμοιο τρόπο με το SNHT for single shifts, μόνο που η σχέση που το περιγράφει είναι αρκετά πιο περίπλοκη. Η μεταβλητή T_{max} σ' αυτή την περίπτωση είναι συνάρτηση του σημείου α, στο οποίο θεωρείται ότι ξεκινάει η μεταβολή της τάσης και του σημείου b, στο οποίο θεωρείται ότι σταματάει. Αναλυτικά, οι σχέσεις παρουσιάζονται από τους Alexandersson and Moberg (1997).

Η ιδεατή μορφή των συναρτήσεων Q και T ανάλογα με τους διάφορους τύπους ανομογενειών παρουσιάζεται στο Σχήμα 3.4.



Σχήμα 3.4. *Ιδεατές μορφές των συναρτήσεων Q και T ανάλογα με τους διάφορους τύπους ανομογενειών.Πηγή: Alexandersson and Moberg (1997).*
(a)Απλό άλμα, (b) Αλλαγή τάσης,   (c) Τρία ξεχωριστά άλματα, και
(d) Αλλαγή τάσης που διακόπτεται από άλμα

Στον Πίνακα 3.1 παρουσιάζονται οι κρίσιμες τιμές του όρου T_{max} του SNHT για διάφορα μήκη n των χρονοσειρών όπως υπολογίστηκαν από τους Alexandersson and Moberg (1997) με προσομοίωση 20000 χρονοσειρών με τυχαίους αριθμούς από κανονική κατανομή. Οι τιμές του όρου T_{crit} υπολογίστηκαν και από τους Khaliq and Ouarda (2007), με χρήση πολύ μεγάλου αριθμού (ενός εκατομμυρίου) προσομοιώσεων Monte Carlo και για περισσότερα μήκη χρονοσειρών από τους Alexandersson and Moberg (1997).

Πίνακας 3.1. *Κρίσιμες τιμές του όρου T_{max} του SNHT ως συνάρτηση των όρων n της χρονοσειράς. Πηγή: Alexandersson and Moberg (1997)*

n	10	20	30	40	50	70	80	90	100	150	250
T_{90}	5.05	6.10	6.65	7.00	7.25	7.40	7.55	7.80	7.85	8.05	8.35
T_{95}	5.70	6.95	7.65	8.10	8.45	8.65	8.80	9.05	9.15	9.35	9.70
$T_{97.5}$	6.25	7.80	8.65	9.25	9.65	9.85	10.2	10.3	10.4	10.8	11.2

Η μέθοδος SNHT θεωρείται ότι δεν δίνει καλά αποτελέσματα στην αρχή και το τέλος των χρονοσειρών, γι’ αυτό και εφαρμογή της στα σημεία αυτά πρέπει να αποφεύγεται.

Κάποιες κατευθύνσεις για την εφαρμογή των μεθόδων Standard Normal Homogeneity Test for single shifts και Standard Normal Homogeneity Test for trends δίνονται από τους Moberg and Alexandersson (1997) και Domonkos (2011a). Από τον Domonkos (2011a) προτείνονται δύο εκδοχές για εντοπισμό αλμάτων μόνο:

1) Όταν εντοπίζεται μια ανομογένεια, η χρονοσειρά διαιρείται σε δύο τμήματα και τα τμήματα αυτά εξετάζονται για περαιτέρω ανομογένειες (cutting algorithm). Η διαδικασία αυτή επαναλαμβάνεται στα τμήματα της χρονοσειράς που προκύπτουν κάθε φορά από το σπάσιμο, εφόσον το μήκος των τμημάτων είναι επαρκές για περαιτέρω εξέταση.

2) Στη δεύτερη εκδοχή της μεθόδου εντοπίζονται οι θέσεις των πιθανών ανομογενειών με τη μέθοδο 1. Στη συνέχεια καθεμιά από τις πιθανές ανομογένειες ξαναελέγχεται με χρήση του τμήματος της χρονοσειράς που περιέχει μόνο την ανομογένεια αυτή και οριοθετείται από μία άλλη ανομογένεια, πιθανή ανομογένεια ή την άκρη της χρονοσειράς. Με αυτή τη διαδικασία κάποιες ανομογένειες βρίσκονται μη σημαντικές και δεν διορθώνονται.

Περισσότερες λεπτομέρειες πάνω στη μεθοδολογία του Domonkos (2011a) δίνονται στην Παράγραφο 5.4.1.

Οι Moberg and Alexandersson (1997) εφάρμοσαν και τις δύο μεθόδους (για τάσεις και άλματα) σε δύο χρονοσειρές ετήσιων θερμοκρασιών. Οι πιθανές τάσεις που εντοπίζονταν θεωρούνταν σημαντικές αν επηρέαζαν ένα διάστημα μεγαλύτερο των 5 ετών, αλλιώς επιλεγόταν ως στατιστικά σημαντικό το άλμα που προέκυπτε από το Standard Normal Homogeneity Test for single shifts. Στη συνέχεια έσπαγαν τη χρονοσειρά σε δύο τμήματα στη θέση της ανομογένειας και εξέταζαν τα δύο αυτά τμήματα εκ νέου και με τις δύο μεθόδους. Η διαδικασία αυτή επαναλήφθηκε μέχρι να μην εντοπιστούν άλλες ανομογένειες ή να προκύψουν τμήματα με μήκος μικρότερο από δύο έτη (μη ελέγξιμα). Οι ανομογένειες που προέκυψαν ελεγχτηκαν ξανά για στατιστική σημαντικότητα και διορθώθηκαν από την τελευταία (πιο πρόσφατη προς την πρώτη. Οι Moberg and Alexandersson (1997) δίνουν αναλυτική περιγραφή της διαδικασίας διόρθωσης.

3.2.4. Pairwise Homogenization Algorithm

Η μέθοδος ομογενοποίησης των Menne and Williams (2009) βασίζεται σε μια σειρά προγενέστερες μεθόδους, όπως στη μέθοδο της Vincent (1998) και το SNHT.

1) Επιλέγονται οι 100 κοντινότεροι σταθμοί με βάση το συντελεστή συσχέτισής τους με την εξεταζόμενη χρονοσειρά (πρέπει να είναι μεγαλύτερος από 0.5). Ο συντελεστής συσχέτισης υπολογίζεται από τις διαφορές διαδοχικών μηνών ως εξής:

$Corr(X_t - X_{t-1}, Y_t - Y_{t-1})$

Αυτό σύμφωνα με τους συγγραφείς γίνεται για να ελαχιστοποιηθεί η επιρροή αλμάτων στη συσχέτιση ανάμεσα στους σταθμούς.

2) Από τους 100 σταθμούς που επιλέγονται στο πρώτο βήμα εντοπίζονται οι 40 σταθμοί με την καλύτερη συσχέτιση με την εξεταζόμενη χρονοσειρά. Από τους 40 αυτούς σταθμούς γίνεται μια λίστα με τους σταθμούς που καλύπτουν κάθε έτος της χρονοσειράς αναφοράς (αφού οι σταθμοί δεν έχουν το ίδο μήκος ούτε καλύπτουν την ίδια ακριβώς χρονική περίοδο). Το επιθυμητό είναι για κάθε έτος να προκύψουν τουλάχιστον 7 σταθμοί αναφοράς.

3) Ο αλγόριθμος κρίνει κατά πόσο ένας επιπλέον από τους 60 σταθμούς που δεν έχουν χρησιμοποιηθεί προσθέτει πληροφορίες για τις χρονικές περιόδους που με την παραπάνω διαδικασία έχουν λιγότερους από 7 σταθμούς αναφοράς. Εφόσον βρεθεί ένας τέτοιος σταθμός που να καλύπτει κάποια από τις περιόδους αυτές αντικαθιστά τον λιγότερο καλά

συσχετισμένο από τους 40 σταθμούς που είχαν επιλεγεί με την προϋπόθεση να μην απορρίπτονται σταθμοί με δεδομένα για άλλες περιόδους με λίγους (7 ή λιγότερους) σταθμούς αναφοράς. Τελικά προκύπτει μια λίστα με τουλάχιστον 7 σταθμούς αναφοράς για κάθε χρονική περίοδο της εξεταζόμενης χρονοσειράς.

4) Δημιουργείται η χρονοσειρά των διαφορών D μεταξύ καθεμιάς από τις χρονοσειρές αναφοράς Y και της υποψήφιας χρονοσειράς X, δηλαδή:

$$D=X-Y \tag{3.15}$$

Οι χρονοσειρές X καιY είναι μηνιαίες και σύμφωνα με τους συγγραφείς έχουν τη μορφή

$$X = \mu_{\nu}^{X} + \beta^{X}(mT+\nu) + \delta_{X}^{mT+\nu} + \varepsilon_{X}^{mT+\nu}, \quad Y = \mu_{\nu}^{Y} + \beta^{Y}(mT+\nu) + \delta_{Y}^{mT+\nu} + \varepsilon_{Y}^{mT+\nu} \tag{3.16}$$

όπου μ: η μέση απόκλιση από τη μέση τιμή για κάθε χρονοσειρά, β: η κλίση της ευθείας που προσαρμόζεται στη χρονοσειρά, T=12: οι μήνες του έτους, $\nu \in \{0, \dots, 12\}$: ο δείκτης του μήνα, m: το έτος, δ_t: το συστηματικό σφάλμα (άλμα λόγω αλλαγών στο σταθμό) , και ε_t: τυχαίο σφάλμα με μέση τιμή μηδέν.

5) Ο εντοπισμός των πιθανών ανομογενειών γίνεται με χρήση μιας παραλλαγής του SNHT (Alexandersson and Moberg, 1997). Αντί για δύο μοντέλα εντοπισμού πιθανών αλμάτων και διαφορών κλίσεων, οι Menne and Williams ανέπτυξαν πέντε διαφορετικά μοντέλα τα οποία παρουσιάζονται στον Πίνακα 3.2. Οι συγγραφείς χρησιμοποιούν επίπεδο σημαντικότητας 5% (α=0.05). Για να λάβει υπόψη η μέθοδος πολλές πιθανές ανομογένειες η χρονοσειρά των διαφορών θεωρείται ότι αποτελείται από K αριθμό τμημάτων, καθένα από τα οποία οριοθετείται από δύο ανομογένειες c_{k-1} και c_k. Η μέθοδος SNHT λαμβάνει επομένως την παρακάτω μορφή:

$$H_0 : D_t \in N(\mu_k, \sigma^2), c_{k-1} \le t \le c_k \tag{3.17}$$

$$H_1 : \begin{cases} D_t \in N(\mu_1, \sigma^2), & c_{k-1}+1 \le t \le c \\ D_t \in N(\mu_2, \sigma^2), & c+1 \le t \le c_k \end{cases} \tag{3.18}$$

Αρχικά ο αλγόριθμος ξεκινάει με c_0=1 c_k=n και σπάει τη χρονοσειρά σε δύο τμήματα σε κάθε θέση ανομογένειας που εντοπίζει. Τελικά προκύπτει μια λύση K τμημάτων, οριοθετημένα από K-1 πιθανές ανομογένειες.

6) Όλες οι κοινές περίοδοι της εξεταζόμενης χρονοσειράς και των χρονοσειρών αναφοράς εξετάζονται και με τα παραπάνω πέντε μοντέλα και επιλέγεται εκείνο το μοντέλο που προσαρμόζεται καλύτερα στα δεδομένα (π.χ. απλό άλμα - μοντέλο M3). Το κριτήριο για την επιλογή του μοντέλου είναι η ελαχιστοποίηση του κριτηρίου Μπεϋσιανής πληροφορίας (BIC, Schwarz 1978). Κατά την επιλογή του μοντέλου είναι δυνατό να απορριφθεί κάποια από τις ανομογένειες που ενοπίστηκαν, π.χ. επειδή αρχικά είχε υποτεθεί ένα άλμα στη μέση ενός τμήματος, ενώ μπορεί να εφαρμόζεται καλύτερα μια γραμμική τάση σε όλο το τμήμα.

7) Ελέγχεται σε ποια από τις δύο χρονοσειρές X και Y βρίσκεται η ανομογένεια μέσω σύγκρισης των πιθανών ανομογενειών με τις άλλες χρονοσειρές.

Πίνακας 3.2. *Παραλλαγή του SNHT με πέντε διαφορετικά μοντέλα για τον εντοπισμό διαφορών μέσων τιμών και κλίσεων. Πηγή: Menne and Williams (2009)*

Model	Description	Schematic of model	Number of parameters p required to fit model
M1	$D_t = \mu + \varepsilon_t$		1
M2	$D_t = \mu + \beta t + \varepsilon_t$		2
M3	$D_t = \begin{cases} \mu_1 + \varepsilon_t, t \le c \\ \mu_2 + \varepsilon_t, t > c \end{cases}$		3
M4	$D_t = \begin{cases} \mu_1 + \beta t + \varepsilon_t, t \le c \\ \mu_2 + \beta t + \varepsilon_t, t > c \end{cases}$		4
M5	$D_t = \begin{cases} \mu_1 + \beta_1 t + \varepsilon_t, t \le c \\ \mu_2 + \beta_2 t + \varepsilon_t, t > c \end{cases}$		5

8) Από τα προηγούμενα βήματα της μεθόδου δεν έχει προσδιοριστεί με ακρίβεια η θέση κάθε πιθανής ασυνέχειας. Από κάθε εφαρμογή της μεθόδου σε ένα ζεύγος X -Y προκύπτει και μια πιθανή θέση με το αντίστοιχο διάστημα εμπιστοσύνης. Καθώς οι θέσεις αυτές είναι πολύ κοντινές προκύπτει τελικά ένα σύνολο γειτονικών ή και επικαλυπτόμενων διαστημάτων, μέσα στο οποίο βρίσκεται η πιθανότερη θέση της ασυνέχειας. Η θέση αυτή τελικά αποφασίζεται είτε από τα ιστορικά δεδομένα του σταθμού είτε από το πιο σημείο το οποίο βρίσκεται μέσα στα διαστήματα εμπιστοσύνης για όλες τα ζεύγη χρονοσειρών-αναφοράς και στο οποίο εντοπίστηκε ασυνέχεια τις περισσότερες φορές.

9) Εφόσον μια ανομογένεια σε μια χρονοσειρά σε σύγκρισή με τουλάχιστον τρεις χρονοσειρές αναφοράς, η ανομογένεια αυτή διορθώνεται. Καθεμιά από τις πιθανές διορθώσεις υπολογίζεται από ομογενή τμήματα των χρονοσειρών αναφοράς που να καλύπτουν ένα χρονικό διάστημα τουλάχιστον 24 μηνών πριν και με΄τα την ανομογένεια (Δύο ασυνέχειες μέσα στο διάστημα αυτό των 24 μηνών θεωρούνται ως μία ανομογένεια και γίνεται μία διόρθωση για το συνδυασμό τους). Η τελική διόρθωση προκύπτει από τη μεσαία από τις υπολογισμένες διορθώσεις, καθώς η κατανομή των διορθώσεων δεν είναι απαραίτητα συμμετρική (εναλλακτικά μπορεί να χρησιμοποιηθεί μια διαφορετική μέθοδος που να λαμβάνει υπόψη την ασυμμετρία).

4. ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΑΞΙΟΛΟΓΗΣΗΣ ΜΕΘΟΔΩΝ ΟΜΟΓΕΝΟΠΟΙΗΣΗΣ

Όπως αναφέρθηκε και στο Κεφάλαιο 3, σήμερα, για τον εντοπισμό και τη διόρθωση των ανομογενειών από τα υδροκλιματικά δεδομένα, χρησιμοποιούνται ευρέως περίπου 20 διαφορετικές μέθοδοι ομογενοποίησης και αν λάβουμε υπόψη τις διαφορές στην πρακτική εφαρμογή τους, ο αριθμός αυτός γίνεται ακόμα μεγαλύτερος (Domonkos 2011a).

Ήδη από τα πρώτα χρόνια εφαρμογής των μεθόδων ομογενοποίησης παρουσιάστηκε η ανάγκη να αξιολογηθεί η συμπεριφορά τους, όσον αφορά:

1. την ικανότητά τους να εντοπίζουν και να διορθώνουν **πραγματικές ανομογένειες** στα δεδομένα (hit rate),
2. τις **ψευδείς – πλασματικές ανομογένειες** που εντοπίζουν και διορθώνουν (false alarms), και
3. την ικανότητά τους να αποκαθιστούν τα πραγματικά στατιστικά χαρακτηριστικά των ομογενών χρονοσειρών.

Υποπερίπτωση του ψευδούς εντοπισμού ανομογενειών αποτελεί και η υπερβολική ευαισθησία κάποιων μεθόδων στον εντοπισμό ανομογενειών σε κάποιο συγκεκριμένο τμήμα της χρονοσειράς, π.χ. στα άκρα ή στο μέσο της χρονοσειράς. Χαρακτηριστική περίπτωση μεθόδου που αναφέρεται στη βιβλιογραφία ότι τείνει να εντοπίζει πολλές ψευδείς - πλασματικές ανομογένειες στα άκρα των χρονοσειρών (overdetection at the edges) είναι το SNHT (Wang, 2008).

Για την αξιολόγηση της συμπεριφοράς των μεθόδων ομογενοποίησης έχουν αναπτυχθεί δύο κύριες μεθοδολογίες (π.χ. Venema *et al*., 2012) οι οποίες αναλύονται στις Παραγράφους 4.1 και 4.2:

1. εφαρμογή σε **πραγματικά - παρατηρημένα δεδομένα**, ομογενή ή και με ανομογένειες, και
2. εφαρμογή σε **συνθετικά δεδομένα**, ομογενή ή και με ανομογένειες.

4.1. Αξιολόγηση μεθόδων ομογενοποίησης με χρήση πραγματικών δεδομένων

Η βιβλιογραφία γύρω από την ομογενοποίηση των υδροκλιματικών δεδομένων είναι πολύ μεγάλη και μεγάλο τμήμα της αφορά την εφαρμογή μεθόδων ομογενοποίησης σε πραγματικά – παρατηρημένα δεδομένα ή/και την αξιολόγηση των μεθόδων ομογενοποίησης μέσα από την εφαρμογή αυτή (π.χ. Moberg and Alexandersson, 1997, Wijngaard *et al.*, 2003, Venema *et al*., 2012).

Για την αξιολόγηση με βάση πραγματικά δεδομένα εξετάζεται βασικά η ικανότητά των μεθόδων ομογενοποίησης να εντοπίζουν και να διορθώνουν τις πραγματικές ανομογένειες των δεδομένων. Αυτό συνήθως γίνεται με εξέταση της συμπεριφοράς των μεθόδων στον εντοπισμό καταγεγραμμένων ανομογενειών, δηλαδή ανομογενειών οι οποίες συμπίπτουν χρονικά με κάποια αλλαγή που έχει συμβεί στο σταθμό ή το περιβάλλον του και έχει καταγραφεί στο αρχείο - μεταδεδομένα του σταθμού. (π.χ. Easterling and Peterson, 1995).

Το μέγεθος των καταγεγραμμένων ανομογενειών συνήθως δεν είναι γνωστό από τα μεταδεδομένα ή άλλα στοιχεία που συγκεντρώνονται και εκτιμάται συνήθως από τις μεθόδους ομογενοποίησης. Συχνά όμως τα μεταδεδομένα μπορεί να δίνουν πληροφορίες και για το πρόσημο του άλματος, αν είναι δηλαδή θετικό ή αρνητικό (Easterling and Peterson, 1995).

Η αξιολόγηση των μεθόδων ομογενοποίησης με χρήση πραγματικών δεδομένων μπορεί να γίνει και με άλλους τρόπους, όπως π.χ. με σύγκριση των αποτελεσμάτων της ομογενοποίησης σε πραγματικά και συνθετικά δεδομένα (Venema *et al.*, 2012).

4.2.Αξιολόγηση μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων

Η αξιολόγηση μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων είναι ένας εξαιρετικά ενδιαφέρον τομέας, καθώς παρέχει μια επιπλέον δυνατότητα σε σύγκριση με την ομογενοποίηση με πραγματικά δεδομένα: επιτρέπει εκτός από τη διερεύνηση της ικανότητας των μεθόδων στον εντοπισμό γνωστών ανομογενειών, να γίνει εκτίμηση του ποσοστού ψευδούς εντοπισμού σε ομογενή ή μη ομογενή δεδομένα από τις μεθόδους ομογενοποίησης. Παράλληλα επιτρέπει να διερευνηθεί με ακρίβεια η ικανότητα των μεθόδων στον εντοπισμό αλμάτων αλλά και στην αποκατάσταση των αρχικών στατιστικών χαρακτηριστικών των χρονοσειρών (π.χ. γραμμικές τάσεις).

Αυτό συμβαίνει γιατί τα δεδομένα παράγονται με ελεγχόμενο τρόπο, οπότε είναι γνωστό αν είναι ομογενή ή μη, τι είδους και μεγέθους ανομογένειες μπορεί να έχουν και σε ποιες θέσεις των χρονοσειρών βρίσκονται οι ανομογένειες αυτές. Το στοιχείο αυτό αποτελεί και το μεγαλύτερο πλεονέκτημα της μεθοδολογίας αυτής έναντι της χρήσης πραγματικών δεδομένων.

Αν και τα συνθετικά δεδομένα υπερέχουν λοιπόν των πραγματικών ως προς την πληρότητα της διαδικασίας αξιολόγησης, υστερούν ωστόσο σε ένα άλλο τομέα: τη ρεαλιστικότητα.

Αυτό είναι και το βασικό μειονέκτημα της μεθοδολογίας αυτής. Η παραγωγή των συνθετικών χρονοσειρών (εξεταζόμενων και αναφοράς) γίνεται κάθε φορά από κάποιο μοντέλο, λιγότερο ή περισσότερο κοντά στα πραγματικά δεδομένα. Ο βαθμός ομοιότητας των συνθετικών δεδομένων με τα πραγματικά προφανώς επηρεάζει τα αποτελέσματα της ομογενοποίησης και τη σημασία τους για την εξαγωγή συμπερασμάτων.

Από την αρχή της χρήσης προσομοιωμένων δεδομένων μέχρι σήμερα, έχει γίνει προσπάθεια οι βάσεις συνθετικών δεδομένων που δημιουργούνται να γίνονται όλο και πιο ρεαλιστικές (Venema *et al.*, 2012). Αυτό γίνεται κυρίως με εισαγωγή ρεαλιστικών ανομογενειών και σε πολύ μικρότερο βαθμό με χρήση διαφορετικών μοντέλων για την παραγωγή των ομογενών χρονοσειρών.

Αναλυτική παρουσίαση των προηγούμενων προσεγγίσεων που κάνουν αξιολόγηση μεθόδων ομογενοποίησης με συνθετικά δεδομένα γίνεται στο Κεφάλαιο 5. Για την καλύτερη κατανόηση των προσεγγίσεων αυτών αλλά και της μεθοδολογίας που ακολουθήθηκε στην παρούσα εργασία (Κεφάλαιο 6), στην Παράγραφο 4.2.1 γίνεται μια παρουσίαση βασικών μοντέλων παραγωγής συνθετικών δεδομένων.

4.2.1. Μοντέλα παραγωγής συνθετικών δεδομένων

Δύο βασικά μοντέλα για την παραγωγή συνθετικών δεδομένων τα οποία χρησιμοποιούνται ευρέως στις αξιολογήσεις της ομογενοποίησης (βλ. Κεφάλαιο 5) είναι το μοντέλο του λευκού θορύβου και το μοντέλο αυτοπαλινδρόμησης τάξης 1 (AR(1)). Ρεαλιστικότερο από τα μοντέλα αυτά είναι το μοντέλο των δεδομένων με μακροπρόθεσμη εμμονή (Koutsoyiannis, 2002) το οποίο και χρησιμοποιήθηκε και στην πορεία της εργασίας αυτής.

Τα τρία αυτά μοντέλα περιγράφουν στάσιμες στοχαστικές ανελίξεις σε διακριτό χρόνο. Στην περιγραφή των βασικών χαρακτηριστικών των μοντέλων που ακολουθεί, η ανέλιξη αυτή σε διακριτό χρόνο δηλώνεται ως X_i, με το δείκτη i=1,2, …, να υποδηλώνει τις διακριτές μονάδες χρόνου (π.χ. μήνες, χρόνια κ.ά.).

Α) Μοντέλο λευκού θορύβου

Ο λευκός θόρυβος είναι μια ακολουθία από τυχαίες μεταβλητές X_i, οι οποίες έχουν μέση τιμή μηδέν, σταθερή τυπική απόκλιση και είναι ασυσχέτιστες. Δηλαδή:

$$« = E[X_i] = 0 \qquad (4.1)$$

$$E[X_i^2] = ±^2, \text{ και} \qquad (4.2)$$

$$E[X_i X_{i+j}] = 0 \qquad (4.3)$$

όπου τ: χρονική υστέρηση

Η συνάρτηση αυτοσυσχέτισης της τυχαίας μεταβλητής X_i είναι:

$$°_j = \begin{cases} 1, & j = 0 \\ 0, & j \neq 0 \end{cases} \qquad (4.4)$$

Η τυχαία μεταβλητή του λευκού θορύβου ακολουθεί δεδομένη κατανομή η οποία μπορεί να είναι κανονική (Γκαουσιανός λευκός θόρυβος), ομοιόμορφη κ.ά.

Β) Μοντέλο αυτοπαλινδρόμησης τάξης 1

Το μοντέλο αυτοπαλινδρόμησης τάξης 1, το οποίο συμβολίζεται ως AR(1), είναι ένα κλασικό μοντέλο περιγραφής δεδομένων με βραχεία – ασθενή μνήμη. Εναλλακτικά ονομάζεται και μοντέλο Markov σε διακριτό χρόνο.

Ανέλιξη Markov λέγεται μια ανέλιξη $X(t)$ στην οποία, αν είναι γνωστό το παρόν, το μέλλον δεν εξαρτάται από το παρελθόν αλλά μόνο από το παρόν. Χαρακτηριστικό των ανελίξεων Markov είναι η εκθετική μείωση της αυτοσυσχέτισης με τη χρονική υστέρηση j.

Το μοντέλο AR(1) περιγράφεται από τη Σχέση:

$$X_i = \rho X_i + V_i \qquad (4.5)$$

όπου ρ: συντελεστής αυτοσυσχέτισης για υστέρηση 1 ($-1<\rho<1$), και

V_i: ακολουθία ανεξάρτητων ομοιόμορφα κατανεμημένων τυχαίων μεταβλητών με μέση τιμή $(1-\rho)\mu$ και μεταβλητότητα $(1-\rho^2)\gamma_0$.

Ο συντελεστής αυτοσυσχέτισης της ανέλιξης X_i (χαρακτηριστικός ανέλιξης Markov) είναι:

$$\rho_j := corr[X_i, X_{i+j}] = \rho^{|j|} \qquad (4.6)$$

Ως μέτρο της βραχυπρόθεσμης μνήμης για τα μοντέλα Markov χρησιμοποιείται ο συντελεστής αυτοσυσχέτισης για υστέρηση 1 (ρ_1).

Γ) Μοντέλο με μακροπρόθεσμη εμμονή

Ένα σημαντικό χαρακτηριστικό που παρατηρείται στα υδροκλιματικά δεδομένα είναι η διατήρηση της αυτοσυσχέτισης για μεγάλες τιμές της υστέρησης και της κλίμακας συνάθροισης. Το φαινόμενο αυτό έχει πολλές εναλλακτικές ονομασίες όπως μακροπρόθεσμη εμμονή, φαινόμενο Hurst, δυναμική Hurst-Kolmogorov και μακρά μνήμη (Koutsoyiannis, 2002) και δεν αναπαράγεται από τα κλασικά στατιστικά μοντέλα όπως ο λευκός θόρυβος και το μοντέλο AR(1). Στην παρούσα εργασία θα αναφέρεται ως μακροπρόθεσμη εμμονή ή δυναμική Hurst-Kolmogorov.

Μοντέλο αναπαραγωγής της μακροπρόθεσμης εμμονής προτάθηκε από τον Mandelbrot (1965), το οποίο είναι γνωστό ως κλασματικός Γκαουσιανός θόρυβος ή αλλιώς ανέλιξη απλής ομοιοθεσίας. Ο κλασματικός Γκαουσιανός θόρυβος είναι στάσιμη ανέλιξη που ορίζεται από τη Σχέση:

$$Z_i^{(k)} - k\mu =_d \left(\frac{k}{l}\right)^H \left(Z_j^{(l)} - l\mu\right) \qquad (4.7)$$

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

i: (οποιοσδήποτε) όρος της ανέλιξης,

j: χρονική υστέρηση (λαμβάνει οποιαδήποτε τιμή),

k, l: οποιεσδήποτε χρονικές κλίμακες (ακέραιοι αριθμοί),

$Z_i^{(k)}$: συναθροισμένη ανέλιξη σε κλίμακα k,

μ: η μέση τιμή του Z_i^k, και

H: σταθερά ($0 < H < 1$) γνωστή ως συντελεστής (ή εκθέτης) Hurst.

Για τη συναθροισμένη ανέλιξη ισχύει:

$$Z_i^{(k)} = \sum_{l=(i-1)k+1}^{ik} X_l \qquad (4.8)$$

Εφόσον η ανέλιξη είναι στάσιμη, για i=j=l=1 προκύπτει η τυπική απόκλιση ως συνάρτηση δύναμης της κλίμακας συνάθροισης *k*:

$$¥_0^{(k)} = k^{2H} ¥_0 \qquad (4.9)$$

Η συνάρτηση αυτοσυσχέτισης της ανέλιξης δίνεται από τη Σχέση 4.10 και είναι ανεξάρτητη της κλίμακας συνάθροισης k:

$$°_j^k = °_j = (1/2)[(j+1)^{2H} + (j-1)^{2H}] - j^{2H}, \qquad j > 0 \qquad (4.10)$$

Από τα παραπάνω προκύπτει ότι το μοντέλο αναπαραγωγής της μακροπρόθεσμης εμμονής ορίζεται μέσω των χαρακτηριστικών του (Σχέσεις 4.7-4.10). Στη βιβλιογραφία υπάρχει πλήθος αλγορίθμων που προσπαθούν να αναπαράξουν το μοντέλο αυτό. Κάποιες απλές προσεγγίσεις που προτείνονται από τον Koutsoyiannis (2002) είναι:

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

Ως μέτρο της μακροπρόθεσμης εμμονής χρησιμοποιείται ο συντελεστής ή εκθέτης Hurst (Σχέση 4.7). Ο συντελεστής Hurst μπορεί να λάβει τιμές από το σύνολο (0, 1), όμως φυσική σημασία έχουν τιμές $H \geq 0.5$. Για τιμές του συντελεστή Hurst $H < 0.5$ η συνάρτηση αυτοσυσχέτισης (Σχέση 4.10) παίρνει αρνητικές τιμές, μια περίπτωση που δεν συναντάται στα υδροκλιματικά δεδομένα (Koutsoyiannis, 2002).

5. ΠΡΟΗΓΟΥΜΕΝΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΟΜΟΓΕΝΟΠΟΙΗΣΗΣ ΜΕ ΣΥΝΘΕΤΙΚΑ ΔΕΔΟΜΕΝΑ – ΑΝΑΛΥΣΗ ΚΑΙ ΚΡΙΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

Η χρήση συνθετικών δεδομένων είναι από τους βασικούς τρόπους ελέγχου της απόδοσης των μεθόδων ομογενοποίησης. Για το λόγο αυτό αποφασίστηκε να γίνει μια συστηματική βιβλιογραφική επισκόπηση των μελετών που έχουν γίνει μέχρι στιγμής και αφορούν μεθόδους ομογενοποίησης και εφαρμογή τους σε συνθετικά δεδομένα. Από την επισκόπηση αυτή προέκυψαν τρεις διαφορετικοί τύποι μελετών που εστιάζονται σε :

Α) Συγκριτική αξιολόγηση μεθόδων

Β) Ανάπτυξη και έλεγχο καινούριας μεθόδου ή τροποποίηση/εφαρμογή μιας προϋπάρχουσας μεθόδου

Γ) Εύρεση κρίσιμων τιμών (στατιστικά σημαντικών τιμών εντοπισμού ανομογενειών)

Από τις παραπάνω κατηγορίες μελετών σχετικές με την αξιολόγηση μεθόδων είναι οι δύο πρώτες κατηγορίες και στις κατηγορίες αυτές επικεντρώνεται αυτό το Κεφάλαιο.

Αναζητήθηκαν και έγινε προσπάθεια να γίνει όσο το δυνατόν πληρέστερη βιβλιογραφική επισκόπηση των συγκριτικών μελετών με συνθετικά δεδομένα γύρω από την ομογενοποίηση, θεωρώντας ότι αποτυπώνουν ικανοποιητικά την πορεία της μεθοδολογίας μέχρι σήμερα.

Αντίθετα, δεν έγινε προσπάθεια να συγκεντρωθούν όλες οι μελέτες που αφορούν μεμονωμένες μεθόδους. Από την κατηγορία αυτή αναφέρονται μόνο κάποιες μελέτες που εντοπίστηκαν και η μεθοδολογία των οποίων θεωρήσαμε ότι διαφοροποιείται από αυτή στις συγκριτικές μελέτες ή προσφέρει πληροφορίες – δίνει έναυσμα για περαιτέρω διερεύνηση.

Στις Παραγράφους 5.1 - 5.5 γίνεται παρουσίαση, κατηγοριοποίηση και κριτική προσέγγιση των μελετών που εντοπίστηκαν μετά από βιβλιογραφική επισκόπηση. Με βάση τα αποτελέσματα της διαδικασίας αυτής προέκυψαν τα βασικά στοιχεία της μεθοδολογίας που ακολουθήθηκε στην παρούσα Μεταπτυχιακή εργασία (Κεφάλαιο 6).

5.1. Προηγούμενες προσεγγίσεις

Όπως αναφέρθηκε πιο πάνω, η διερεύνησή μου εστιάζεται σε δύο τύπους μελετών:

Α) Μελέτες στις οποίες έγινε συγκριτική αξιολόγηση διαφορετικών μεθόδων

Στην κατηγορία αυτή εντοπίστηκαν συνολικά 12 δημοσιεύσεις:

1. Buishand (1982)
2. Easterling and Peterson (1992)
3. Easterling and Peterson (1995)
4. Lubes-Niel *et al*. (1998)
5. Ducré-Robitaille *et al*. (2003)
6. Menne and Williams (2005)
7. DeGaetano (2006)
8. Reeves *et al*. (2007)

9. Beaulieu *et al.* (2008)
10. Domonkos (2008)
11. Domonkos (2011a)
12. Venema *et al.* (2012)

Από τις παραπάνω μελέτες, οι περισσότερες κάνουν χρήση συνθετικών δεδομένων θερμοκρασιών. Εξαίρεση αποτελούν οι μελέτες των Buishand (1982), Lubes-Niel *et al.* (1998) και Beaulieu *et al.* (2008) που κάνουν χρήση συνθετικών δεδομένων βροχοπτώσεων και η μελέτη των Venema *et al.* (2012) που κάνει χρήση και των δύο τύπων δεδομένων.

Β) Μελέτες στις οποίες έγινε ανάπτυξη και έλεγχος καινούριας μεθόδου ή τροποποίηση /εφαρμογή μιας προϋπάρχουσας μεθόδου

Οι δημοσιεύσεις που εντοπίστηκαν και επιλέχτηκαν να εξεταστούν αναλυτικά είναι:

1. Wang (2008)
2. Rust *et al*. (2008)

Αξίζει να αναφερθεί ότι από τους Easterling and Peterson (1995) έγινε ανάπτυξη μιας νέας μεθόδου ομογενοποίησης και σύγκρισή της με άλλες μεθόδους. Καθώς όμως είναι μια κλασική σύγκριση μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων και ενέπνευσε πολλές επόμενες συγκριτικές μελέτες (Venema *et al.*, 2012), κατατάχθηκε στην πρώτη κατηγορία. Στην πρώτη κατηγορία κατατάξαμε επίσης και τη μελέτη των Reeves *et al.* (2007) αν και επίσης γίνεται εφαρμογή μιας καινούριας μεθόδου, λόγω της σημασίας της στη βιβλιογραφία ως συγκριτική μελέτη.

Κοινό στοιχείο όλων των μελετών που εντοπίστηκαν είναι η χρήση της μεθόδου Monte Carlo, δηλαδή προσομοιώσεων οι οποίες περιλαμβάνουν τυχαίους αριθμούς.

5.2. Κατηγοριοποίηση χαρακτηριστικών των προσεγγίσεων

Οι μελέτες των μεθόδων ομογενοποίησης μπορούν να κατηγοριοποιηθούν με πολλούς διαφορετικούς τρόπους με βάση κάποια βασικά χαρακτηριστικά τους. Ενδεικτικά, κάποιες κατηγοριοποιήσεις που μπορεί να γίνουν είναι με βάση:

- ποιες μέθοδοι ομογενοποίησης επιλέχτηκαν να εφαρμοστούν στα συνθετικά δεδομένα,
- το είδος των μεθόδων ομογενοποίησης που επιλέχτηκαν (σχετικών/απόλυτων και υποκειμενικών/αντικειμενικών),
- στην περίπτωση της σχετικής ομογενοποίησης τον αριθμό των χρονοσειρών αναφοράς που επιλέχτηκαν και τον τρόπο που χρησιμοποιήθηκαν,
- το μοντέλο με το οποίο έγινε παραγωγή των συνθετικών δεδομένων (π.χ. λευκός θόρυβος, μοντέλο με αυτοσυσχέτιση κλπ.),
- το μήκος και η χρονική κλίμακα (ετήσια, μηνιαία) των συνθετικών δεδομένων που χρησιμοποιήθηκαν,
- τις ανομογένειες (τύπος, μέγεθος, θέση) που εισάχθηκαν στα συνθετικά δεδομένα,
- τα κριτήρια αποτελεσματικότητας των μεθόδων ομογενοποίησης που επιλέχθηκαν, και

- τα αποτελέσματα που προέκυψαν από κάθε ανάλυση ως προς την καταλληλότητα ή μη των μεθόδων ομογενοποίησης.

5.3. Παράμετροι διερεύνησης

Επειδή η παρούσα Μεταπτυχιακή εργασία εστιάστηκε στη θεωρητική δομή των μεθόδων ομογενοποίησης και κατά πόσο αυτή είναι κατάλληλη για τα υδροκλιματικά δεδομένα, από τα χαρακτηριστικά των αναλύσεων που παρατίθενται στην Παράγραφο 5.2, επιλέχτηκε να διερευνήσουμε:

- τις μεθόδους ομογενοποίησης που έχουν εφαρμοστεί σε συνθετικά δεδομένα,
- το είδος των μεθόδων αυτών (σχετικών/απόλυτων και υποκειμενικών/αντικειμενικών),
- στην περίπτωση της σχετικής ομογενοποίησης τον αριθμό των χρονοσειρών αναφοράς και τον τρόπο που χρησιμοποιήθηκαν,
- το μοντέλο με το οποίο έγινε παραγωγή των συνθετικών δεδομένων, και
- το ποσοστό των χρονοσειρών στις οποίες εντοπίζονται ψευδή άλματα ως κριτήριο αποτελεσματικότητας των μεθόδων ομογενοποίησης

Για λόγους πληρότητας γίνεται μια σύντομη αναφορά στις επιλογές των διαφόρων ερευνητών σε σχέση με τη χρήση ομογενών ή μη ομογενών συνθετικών δεδομένων. Όμως δεν γίνεται ανάλυση του είδους, της κατανομής και του μεγέθους των ανομογενειών που έχουν εισαχθεί μετά την προσομοίωση των δεδομένων, καθώς αποτελούν διαφορετικό πεδίο μελέτης. Μελέτη της κατανομής των ανομογενειών έχει γίνει π.χ. από τους Menne and Williams (2005). Επιπλέον δεν γίνεται γενικά αναφορά στα αποτελέσματα των αναλύσεων, πέρα από τα κριτήρια που σχετίζονται με τον εντοπισμό ψευδών - πλασματικών αλμάτων, παρά μόνο ένα γενικό σχόλιο σχετικά με το κατά πόσο βρέθηκαν από τους μελετητές που τις χρησιμοποίησαν κατάλληλες για τα δεδομένα στα οποία εφαρμόστηκαν.

Τέλος, μια σύντομη αναφορά γίνεται σε σχόλια ή συμπεράσματα των συγγραφέων, τα οποία θεωρήθηκαν χρήσιμα για την περαιτέρω πορεία της παρούσας Μεταπτυχιακής εργασίας.

5.4. Επισκόπηση των μεθόδων με βάση τα επιλεγμένα κριτήρια

Όπως αναφέρθηκε και στην Παράγραφο 5.1, διερευνήθηκαν συνολικά 12 δημοσιεύσεις στις οποίες έγινε συγκριτική αξιολόγηση διαφορετικών μεθόδων ομογενοποίησης και 2 δημοσιεύσεις στις οποίες έγινε ανάπτυξη και έλεγχος καινούριας μεθόδου ή τροποποίηση /εφαρμογή μιας προϋπάρχουσας μεθόδου.

Σε όλες τις μελέτες αυτές για την παραγωγή συνθετικών δεδομένων χρησιμοποιούνται δεδομένα από μοντέλα στάσιμα ή μη και με ή χωρίς αυτοσυσχέτιση. Τα βασικά χαρακτηριστικά των μοντέλων συνοψίζονται στην Παράγραφο 4.2.1.

Για ομοιομορφία των αναλύσεων και αποτελεσμάτων, στην παρούσα εργασία υιοθετήθηκε ο συμβολισμός X για τις χρονοσειρές αναφοράς και Y για τις ελεγχόμενες χρονοσειρές. Ο συμβολισμός αυτός όμως πολλές φορές οδηγεί σε ασάφειες σε σχέση με τη διαφορετική προσέγγιση κάθε συγγραφέα. Για το λόγο αυτό, όπου θεωρήθηκε αναγκαίο διατηρήθηκε ο συμβολισμός που έχει επιλέξει κάθε συγγραφέας. Τα διαφορετικά σύμβολα όμως που χρησιμοποιούνται συνοδεύονται από επεξήγηση, ώστε να αποφευχθούν παρανοήσεις.

Ο συμβολισμός των συγγραφέων διατηρήθηκε και για τις μεθόδους ομογενοποίησης που εξετάστηκαν. Μικρές διαφοροποιήσεις πάντως παρατηρούνται σ'αυτό τον τομέα, καθώς οι συγγραφείς διατηρούν γενικά το συμβολισμό προηγούμενων μελετών για λόγους σύγκρισης.

5.4.1. Συγκριτική αξιολόγηση διαφορετικών μεθόδων

1. **Buishand (1982)**

Πρόκειται για μία από τις πρώτες συγκρίσεις μεθόδων ομογενοποίησης με συνθετικά δεδομένα. Ο συγγραφέας εφάρμοσε στα ίδια συνθετικά δεδομένα τρεις κλασικές απόλυτες μεθόδους ομογενοποίησης για δεδομένα βροχόπτωσης:

- Von Neumann ratio test
- Cumulative deviations (Γραφική εξέταση συσσωρευτικών αποκλίσεων, Craddock, 1979)
- Worsley's likelihood ratio test (Worsley, 1979)

Τα συνθετικά δεδομένα που χρησιμοποιήθηκαν ήταν 1999 χρονοσειρές από 30 ψευδοτυχαίους αριθμούς, οι οποίοι παράχθηκαν από κανονική κατανομή. Στις χρονοσειρές αυτές προστέθηκε ένα μη αρνητικό άλμα $\Delta \geq 0$ σε δύο εναλλακτικές θέσεις m, όπου m=15 ή m=5 και ελέγθηκε η πιθανότητα εντοπισμού άλματος.

Από τη σύγκριση ο συγγραφέας κατέληξε ότι οι μέθοδοι Worsley's likelihood ratio test και Cumulative deviations έχουν καλύτερη συμπεριφορά στον εντοπισμό αλμάτων από το Von Neumann ratio test για μοντέλο μιας μόνο αλλαγής στη μέση τιμή.

2. **Easterling and Peterson (1992)**

Η μελέτη βασίστηκε σε αξιολόγηση 8 αντικειμενικών σχετικών μεθόδων ομογενοποίησης με χρήση ετήσιων συνθετικών δεδομένων μήκους 100 ετών. Οι μέθοδοι που επιλέχτηκαν ήταν:

- SNHT (Alexandersson, 1986)
- Potter's technique (Potter, 1981)
- t test
- Reg Y-hat (regression approach, Easterling and Peterson, 1992)
- Reg avg Y-hat (regression approach, Easterling and Peterson, 1992)
- Double mass analysis (Kohler, 1949)
- CUSUM (Cumulative sum technique, van Dobben de Bruyn, 1968)
- Two-phase regression (Solow, 1987)

Για την παραγωγή των χρονοσειρών χρησιμοποιήθηκε μια προσέγγιση τύπου Monte Carlo. Παράχθηκαν 1000 ζεύγη χρονοσειρών αναφοράς-ελεγχόμενων, με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1)). Τα ζεύγη αυτά χρονοσειρών αναφοράς-ελεγχόμενων συσχετίστηκαν μεταξύ τους πολλαπλασιάζοντας τις κανονικοποιημένες (ετήσιες) τιμές της χρονοσειράς αναφοράς με μια σταθερά με τιμή μεταξύ 0.5 και 2 (ανάλογα με τον επιθυμητό βαθμό συσχέτισης) και προσθέτοντας την τιμή αυτή στην αντίστοιχη κανονικοποιημένη τιμή (αντίστοιχο έτος) της ελεγχόμενης χρονοσειράς. Σταθερές 0.5-2 αντιστοιχούν σε συντελεστή ετεροσυσχέτισης 0.45-0.9 περίπου.

Οι ελεγχόμενες χρονοσειρές που προέκυψαν από την παραπάνω διαδικασία κανονικοποιήθηκαν ξανά και προστέθηκε μια μέση τιμή 23.0. Αντίστοιχα μια μέση τιμή 25.0 προστέθηκε στις χρονοσειρές αναφοράς.

Οι συγγραφείς έκαναν έλεγχο των μεθόδων στον εντοπισμό αλμάτων. Για το λόγο αυτό προστέθηκε στις ελεγχόμενες χρονοσειρές ένα μοναδικό άλμα μεγέθους 0.5-2.0 από τον όρο όρο 65 μέχρι το τέλος των χρονοσειρών.

Από την εφαρμογή των μεθόδων στα συνθετικά δεδομένα προέκυψε ότι το SNHT είχε την καλύτερη συμπεριφορά στον εντοπισμό του άλματος, ακολουθούμενο από τη μέθοδο του Potter, ενώ η μέθοδος Two-phase regression δεν εντόπισε τα άλματα καθώς αναζητά διαφορές κλίσεων στα δεδομένα.

Εκτός από τις χρονοσειρές με ένα άλμα οι συγγραφείς έλεγξαν τις μεθόδους του Potter και το SNHT σε χρονοσειρές με 1-3 άλματα. Οι μέθοδοι είχαν παρόμοια συμπεριφορά στον εντοπισμό αλμάτων, με το SNHT να υπερτερεί ελαφρά.

3. Easterling and Peterson (1995)

Η συγκριτική μελέτη των Easterling and Peterson (1995) ίσως ήταν η πρώτη αξιολόγηση μεθόδων ομογενοποίησης με χρήση χρονοσειρών με πολλαπλά άλματα που δημοσιεύτηκε σε περιοδικό με κριτές και ενέπνευσε τα επόμενα χρόνια μια σειρά από μελέτες γύρω από την αποτελεσματικότητα της σχετικής ομογενοποίησης (Venema *et al.*, 2012).

Η μελέτη βασίστηκε σε αξιολόγηση 2 αντικειμενικών μεθόδων ομογενοποίησης με χρήση ετήσιων συνθετικών δεδομένων μήκους 100 ετών. Οι συγγραφείς συνέκριναν τη συμπεριφορά μιας νέας μεθόδου που ανέπτυξαν, με τη μέθοδο SNHT που είχε την καλύτερη συμπεριφορά κατά την αξιολόγηση από τους Easterling and Peterson (1992).

Αξιολογήθηκαν συγκεκριμένα οι μέθοδοι:

- SNHT (Alexandersson, 1986),
- Easterling and Peterson test (Easterling and Peterson, 1995).

Οι συνθετικές χρονοσειρές παράχθηκαν με τον ίδιο τρόπο που χρησιμοποιήθηκε από τους Easterling and Peterson (1995), με συντελεστή όμως ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες 0.75 (τυπική τιμή κατά τους συγγραφείς) και πολλαπλές ανομογένειες αντί για μία μοναδική. Συνολικά εξετάστηκαν 5 διαφορετικές περιπτώσεις χρονοσειρών με 1-5 ανομογένειες. Για κάθε περίπτωση έγιναν 1000 προσομοιώσεις Monte Carlo, παράχθηκαν δηλαδή 1000 ζεύγη χρονοσειρών αναφοράς-ελεγχόμενων.

Στις χρονοσειρές αυτές μετά την ομογενοποίηση εξετάστηκε:

- το ποσοστό των ανομογενειών που εντοπίστηκαν, και

- το ποσοστό των ψευδών ανομογενειών που εντόπισαν οι μέθοδοι.

Οι συγγραφείς συμπέραναν ότι η μέθοδος του Alexandersson είναι καλύτερη για χρήση σε χρονοσειρές με ένα μοναδικό άλμα, ενώ η μέθοδος που αναπτύχθηκε από τους ίδιους είναι καλύτερη για χρονοσειρές με πολλαπλά άλματα.

Από τα αποτελέσματα που εμφανίζουν οι συγγραφείς προκύπτει ότι οι λανθασμένοι εντοπισμοί ανομογενειών ως ποσοστό των ολικών (πραγματικών και ψευδών) εντοπισμών των δύο μεθόδων, κυμαίνονται για τις 5 διαφορετικές περιπτώσεις δεδομένων με ανομογένειες που εξετάστηκαν:

- από 26 μέχρι 51% για το SNHT, και
- από 23 μέχρι 63% για τη μέθοδο των Easterling and Peterson (1995).

Τα υψηλά αυτά ποσοστά λανθασμένων εντοπισμών ανομογενειών σε όλες τις περιπτώσεις ελάχιστα σχολιάστηκαν από τους συγγραφείς.

4. Lubes-Niel *et al*. (1998):

Η μελέτη αυτή βασίστηκε σε αξιολόγηση 5 μεθόδων ομογενοποίησης:

- rank correlation test (Kendall and Stuart, 1943, WMO, 1966, Olaniran, 1991)
- Pettitt's test (Pettitt, 1979),
- Buishand's test (Buishand, 1982,1984),
- Lee and Heghinian's bayesian procedure (Lee and Heghinian, 1977), και
- Hubert and Carbonnel's segmentation procedure for hydrometeorological series (Hubert and Carbonnel, 1987, 1993, Hubert *et al.*, 1989),

με χρήση ετήσιων συνθετικών δεδομένων βροχόπτωσης μήκους 50 ετών. Τα προσομοιωμένα δεδομένα επιλέχθηκε να έχουν μέση τιμή 1500 και τυπική απόκλιση 300, συνήθεις τιμές στη Δυτική Αφρική εκτός της Σαχάρας. Παράχθηκαν 5 τύποι δεδομένων:

- ανεξάρτητα με κανονική κατανομή,
- ανεξάρτητα με κατανομή λογαριθμοκανονική,
- ανεξάρτητα με κατανομή γάμμα,
- συσχετισμένα με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1)), και
- ανεξάρτητα με κανονική κατανομή και προσθήκη γραμμικής τάσης.

Για καθεμία από τις 5 περιπτώσεις έγιναν 100 προσομοιώσεις. Ανάλογα με τη μέθοδο που εφαρμόστηκε προστέθηκαν σε κάποιες ομάδες των παραπάνω χρονοσειρών άλματα στη μέση τιμή ή απότομες αλλαγές στη μεταβλητότητα.

Για να διερευνηθεί αν τα αποτελέσματα της μελέτης επηρεάστηκαν από τη μέση τιμή και την τυπική απόκλιση που επιλέχτηκαν αρχικά, έγινε εφαρμογή των μεθόδων ομογενοποίησης σε ανεξάρτητα δεδομένα με κανονική κατανομή, με συντελεστή μεταβλητότητας 0.13, μέση τιμή 2600 και τυπική απόκλιση 340, τυπικές τιμές ετήσιων βροχοπτώσεων στη περιοχή του Καμερούν.

Οι συγγραφείς δεν συνέκριναν τις μεθόδους μεταξύ τους και έβγαλαν κυρίως συμπεράσματα για τους τύπους ανομογενειών που μπορεί να εντοπίσει η κάθε μέθοδος.

5. Ducré-Robitaille *et al*. (2003):

Η μελέτη αυτή βασίστηκε σε αξιολόγηση 8 αντικειμενικών μεθόδων ομογενοποίησης με χρήση ετήσιων συνθετικών δεδομένων μήκους 100 ετών. Οι μέθοδοι αυτές ήταν:

- SNHT without trend (Alexandersson, 1986),
- SNHT with trend (Alexandersson and Moberg, 1997),
- MLR (multiple linear regression, Vincent, 1998),
- TPR (two-phase regression, Easterling and Peterson, 1995),
- WRS (Wilcoxon rank-sum, Karl and Williams, 1987),
- ST (sequential testing for equality of means, Gullett *et al*., 1990),
- Bayesian approach without reference series (Ouarda *et al*., 1999; Perreault *et al*., 1999, 2000), και
- Bayesian approach with reference series (Ouarda *et al*., 1999; Perreault *et al*., 1999, 2000).

Για τη μελέτη χρησιμοποιήθηκαν τρεις τύποι συνθετικών δεδομένων: ομογενή, με ένα άλμα και με τυχαίο αριθμό αλμάτων.

Σε πρώτη φάση παράχθηκαν οι ελεγχόμενες (candidate) χρονοσειρές με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1)), από κανονική κατανομή και με μέση τιμή μηδέν και μεταβλητότητα 1, σύμφωνα με τη Σχέση:

$$X_t = 0.1X_{t-1} + e_t \quad e_t \sim N(0, 1) \qquad (5.1)$$

Οι χρονοσειρές αυτές χωρίστηκαν σε τρεις ομάδες. Η πρώτη ομάδα διατηρήθηκε ομογενής, στη δεύτερη ομάδα προστέθηκε ένα άλμα, ενώ στην τρίτη ομάδα προστέθηκε τυχαίος αριθμός αλμάτων.

Στη συνέχεια για την καθεμία από τις εξεταζόμενες χρονοσειρές παράχθηκαν τρεις χρονοσειρές αναφοράς Y_t με χρήση του ίδιου μοντέλου όπως και στις εξεταζόμενες χρονοσειρές (Σχέση 4.1). Καθεμία από τις χρονοσειρές αναφοράς συσχετίστηκε με την ελεγχόμενη χρονοσειρά (πριν την εισαγωγή των ανομογενειών) πολλαπλασιάζοντας την ελεγχόμενη χρονοσειρά με 1.5 και προσθέτοντάς τη στη χρονοσειρά αναφοράς σύμφωνα με τη Σχέση 5.2.

$$Y_t = 1.5X_t + 0.1Y_{t-1} + e_t \quad e_t \sim N(0, 1), \qquad (5.2)$$

όπου Y_t είναι η συνθετική χρονοσειρά αναφοράς.

Καθεμία από τις χρονοσειρές αναφοράς κανονικοποιήθηκε ώστε να έχει μηδενική μέση τιμή και μοναδιαία μεταβλητότητα. Η τελική χρονοσειρά αναφοράς που χρησιμοποιήθηκε για τον έλεγχο των μεθόδων ομογενοποίησης προέκυψε ως η μέση τιμή των τριών χρονοσειρών αναφοράς που παράχθηκαν με την παραπάνω διαδικασία. Η ετεροσυσχέτιση ανάμεσα στις εξεταζόμενες και τις χρονοσειρές αναφοράς ήταν της τάξης του 0.8.

Οι μέθοδοι ομογενοποίησης εφαρμόστηκαν στις ομάδες χρονοσειρών αναφοράς- ομογενών ελεγχόμενων και χρονοσειρών αναφοράς- ελεγχόμενων μετά από εισαγωγή ενός ή περισσότερων αλμάτων. Εξετάστηκαν η ικανότητα των μεθόδων στον εντοπισμό αλμάτων και στον εντοπισμό ομογενών χρονοσειρών. Ο εντοπισμός ομογενών χρονοσειρών εκφράστηκε από το ποσοστό χρονοσειρών με άλμα που βρέθηκαν (θα το αναφέρουμε ως ποσοστό ψευδών εντοπισμών) ενώ ήταν ομογενείς.

Στον εντοπισμό αλμάτων καλύτερη συμπεριφορά είχε η μέθοδος SNHT. Η ίδια μέθοδος μαζί με το MLR και την Bayesian with reference είχαν και τα μικρότερα ποσοστά ψευδών εντοπισμών ανομογενειών όταν εφαρμόστηκαν σε ομογενείς χρονοσειρές (1.2%, 3.6% και 6.8% αντίστοιχα). Οι υπόλοιπες μέθοδοι είχαν ποσοστά ψευδών εντοπισμών μεγαλύτερο από 5% που όπως αναφέρεται και από τους συγγραφείς,είναι το ανεκτό ποσοστό. Πολύ μεγάλα ποσοστά σφαλμάτων παρουσίασαν τα SNHT with trend, TPR, και WRS (13.3%, 41.3%, και 56.3%, αντίστοιχα).

6. Menne and Williams (2005):

Από τους συγγραφείς επιλέχτηκε να αξιολογηθούν οι 3 μέθοδοι ομογενοποίησης που από τη μελέτη των Ducré-Robitaille et al. (2003) βρέθηκε να έχουν την καλύτερη συμπεριφορά:

- SNHT (standard normal homogeneity test without trend, Alexandersson, 1986),
- MLR (multiple linear regression, Vincent, 1998), και
- Bayesian approach with reference series (Ouarda *et al*., 1999; Perreault *et al*., 1999, 2000).

Και οι τρεις μέθοδοι εφαρμόστηκαν στα δεδομένα με χρήση χρονοσειρών αναφοράς και μόνο για την εύρεση αλμάτων και όχι τάσεων. Επιλέχτηκε να χρησιμοποιηθούν για κάθε εξεταζόμενη χρονοσειρά m=5 χρονοσειρές αναφοράς. Οι μέθοδοι χρησιμοποιήθηκαν με χρήση ενός ημι-ιεραρχικού αλγόριθμου ο οποίος σπάει σε τμήματα τη χρονοσειρά κάθε φορά που εντοπίζεται ένα άλμα και αναζητά στα τμήματα που προκύπτουν νέα άλματα. Κάθε φορά που ένα τμήμα χωρίζεται στα δύο τα άλματα που εντοπίστηκαν στο προηγούμενο «σπάσιμο» επαναξιολογούνται για τη στατιστική σημαντικότητά τους.

Οι συγγραφείς χρησιμοποίησαν διαφορετικά κριτήρια για την αξιολόγηση των μεθόδων, ανάμεσα στα οποία και οι ψευδείς εντοπισμοί ανομογενειών.

Για την αξιολόγηση των μεθόδων παράχθηκαν 7 σύνολα 1000 χρονοσειρών, μήκους 100 ετών, με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1)), από κανονική κατανομή και με μέση τιμή μηδέν και μεταβλητότητα 1, σύμφωνα με τη Σχέση:

$$x_{t+1}-\mu=\varphi(x_t-\mu)+e_{t+1} \quad e \sim N(0, 1) \qquad (5.3)$$

όπου μ είναι η μέση τιμή της χρονοσειράς (σε αυτή την περίπτωση 0), φ ο συντελεστής αυτοσυσχέτισης και e ο όρος που αντιστοιχεί στο τυχαίο σφάλμα.

Σε κάθε χρονοσειρά ο συντελεστής αυτοσυσχέτισης επιλέχτηκε τυχαία από ένα δείγμα συντελεστών αυτοσυσχέτισης για υστέρηση ένα (αντιστοιχεί σε ένα έτος) που προέκυψε από το δίκτυο σταθμών United State Historical Climatology Network (USHCN).

Οι 7 ομάδες χρονοσειρών που προέκυψαν, αν και σχεδόν κανονικοποιημένες, κανονικοποιήθηκαν ξανά. Για να δημιουργηθούν δίκτυα ετεροσυσχετισμένων χρονοσειρών, για καθεμία από τις 1000 χρονοσειρές κάθε ομάδας φτιάχτηκαν m=5 αντίγραφα. Σε καθεμία από τις (m + 1)=6 χρονοσειρές που προέκυψαν προστέθηκε μια σταθερά ίση με το διπλάσιο του συντελεστή ετεροσυσχέτισης που υπολογίστηκε σε παρατηρημένες χρονοσειρές δημιουργώντας με αυτό τον τρόπο δίκτυα 6 σταθμών, 1 ελεγχόμενης και m=5 αναφοράς, συσχετισμένων μεταξύ τους στον ίδιο βαθμό περίπου.

Στις χρονοσειρές που προέκυψαν από την παραπάνω διαδικασία προστέθηκαν μηδέν, ένα, δύο, ή τυχαίος αριθμός αλμάτων, ανάλογα με την ομοάδα στην οποία βρίσκονταν. Μια διαφορά από προηγούμενες μελέτες ήταν ότι προστέθηκαν άλματα όχι μόνο στις εξεταζόμενες χρονοσειρές, αλλά και στις χρονοσειρές αναφοράς.

Από τις m χρονοσειρές αναφοράς που δημιουργήθηκαν για κάθε ελεγχόμενη χρονοσειρά σχηματίστηκε μία συνθετική χρονοσειρά αναφοράς με 3 διαφορετικούς τρόπους που περιγράφονται από τους συγγραφείς. Στη συνέχεια, εφαρμόστηκαν οι 3 μέθοδοι ομογενοποίησης με διάφορους συνδυασμούς των 3 διαφορετικών τύπων χρονοσειρών αναφοράς. Κατά την εφαρμογή της ομογενοποίησης αναζητήθηκαν άλματα ξεχωριστά από κάθε μέθοδο ή κοινά άλματα από 2 μεθόδους. Συνολικά εξετάστηκαν πολλοί διαφορετικοί συνδυασμοί μεθόδων και χρονοσειρών αναφοράς.

Όσον αφορά τους ψευδείς εντοπισμούς ανομογενειών εξετάστηκαν η πιθανότητα ψευδούς εντοπισμού (probability of false detection, F) και το ποσοστό ψευδών εντοπισμών (false alarm rate), FAR), τα οποία ορίζονται ως :

$$F=b/(b+d) \quad (5.4)$$

$$FAR=b/(a+b) \quad (5.5)$$

όπου a: ο αριθμός των πραγματικών αλμάτων που εντοπίζονται,

b: ο αριθμός των ψευδών αλμάτων που εντοπίζονται, και

d: ο αριθμός των μηνών, ετών κλπ στα οποία δεν εντοπίζεται ούτε υπάρχει πραγματικό άλμα.

Οι συγγραφείς συμπέραναν ότι για ομογενείς χρονοσειρές αναφοράς και μεμονωμένη εφαρμογή των 3 μεθόδων με όλους τους συνδυασμούς χρονοσειρών αναφοράς, το ποσοστό ψευδών εντοπισμών είναι τις περισσότερες φορές μεγαλύτερο από 5%. Μεγάλη μείωση του ποσοστού ψευδούς εντοπισμού παρατηρήθηκε αντίθετα για εντοπισμό αλμάτων από τουλάχιστον δύο διαφορετικές μεθόδους. Επίσης για ομογενείς χρονοσειρές αναφοράς η επιλογή της μεθόδου ομογενοποίησης αποδείχτηκε πιο σημαντική από τον τρόπο κατασκευής της χρονοσειράς αναφοράς.

Για μη ομογενείς χρονοσειρές αναφοράς το ποσοστό ψευδών εντοπισμών αυξήθηκε πολύ για όλους τους τρόπους χρήσης των μεθόδων ομογενοποίησης και των χρονοσειρών αναφοράς. Πιο σημαντικός παράγοντας για τη μείωση αυτού του ποσοστού αποδείχτηκε ο τρόπος παραγωγής της χρονοσειράς αναφοράς από τη μέθοδο ομογενοποίησης. Καλύτερος τρόπος χρήσης των χρονοσειρών αναφοράς αποδείχτηκε η συνδυασμένη χρήση δύο μεθόδων για την εύρεση αλμάτων.

7. **DeGaetano (2006):**

Ο συγγραφέας επέλεξε να αξιολογήσει 7 μεθόδους ομογενοποίησης, με βάση ετήσια συνθετικά δεδομένα θερμοκρασίας. Οι μέθοδοι που εξετάστηκαν ήταν:

- SNHT without trend (Alexandersson, 1986),
- POTT (Potter's method, Potter, 1981),
- MLR (multiple linear regression, Vincent, 1998),
- TPR (two-phase regression, Easterling and Peterson, 1995),
- BAYE (Bayesian test, Perreault *et al*., 1999, 2000),
- PMETA (Parametric metadata-based test, Karl and Williams, 1987), και
- NMETA (Nonparametric metadata-based test, Allen and DeGaetano, 2000).

Για να παραχθούν οι συνθετικές χρονοσειρές επιλέχτηκαν αρχικά 5 σύνολα σταθμών αναφοράς – εξεταζόμενων από το National Climatic Data Center (NCDC), με κριτήριο να είναι όσο το δυνατόν πιο αντιπροσωπευτικά των κλιματικών συνθηκών των ΗΠΑ.

Οι χρονοσειρές αναφοράς επιλέχτηκαν ώστε να ελαχιστοποιούν τη γεωγραφική απόσταση από τον εξεταζόμενο σταθμό, διατηρώντας παράλληλα μια συσχέτιση με το σταθμό αυτό μεγαλύτερη ή ίση με 0.70. Η προσομοίωση των συνθετικών χρονοσειρών έγινε με τέτοιο τρόπο ώστε να διατηρούνται οι ετεροσυσχετίσεις ανάμεσα στις εξεταζόμενες χρονοσειρές και τις χρονοσειρές αναφοράς και οι μεταβλητότητες των πραγματικών σταθμών. Τυπική τιμή ετεροσυσχέτισης ήταν της τάξης του 0.8. Αυτή είναι και η βασική διαφοροποίηση της μελέτης αυτής από προηγούμενες. Ο συγγραφέας εξέτασε χρονοσειρές με διαφορετικά μήκη.

Οι ομάδες ομογενών χρονοσειρών αναφοράς-εξεταζόμενων παράχθηκαν σύμφωνα με τη Σχέση 5.6:

$$Z_t=[\Phi]z_{t-1}+[B]e_t \qquad (5.6)$$

όπου Z_t είναι ένα διάνυσμα πέντε στάσιμων κανονικοποιημένων μεταβλητών, που αντιπροσωπεύουν την εξεταζόμενη χρονοσειρά και τις τέσσερις χρονοσειρές αναφοράς, [B] είναι ένα μητρώο 5×5 με στοιχεία τους συντελεστές ετεροσυσχέτισης ανάμεσα στις πραγματικές χρονοσειρές, e_t είναι ένα διάνυσμα ανεξάρτητων γκαουσιανών τυχαίων αριθμών και είναι $[\Phi]z_{t-1}$ ένας όρος αυτοσυσχέτισης.

Οι συντελεστές αυτοσυσχέτισης για υστέρηση 1 των παρατηρημένων χρονοσειρών μετά την αφαίρεση της τάσης τους (detrended) θεωρήθηκαν από τον DeGaetano (2006) μικρές, οπότε ο όρος $[\Phi]z_{t-1}$ απαλείφθηκε από την Εξίσωση 4.4.

Στις ελεγχόμενες χρονοσειρές που προέκυψαν από την παραπάνω διαδικασία προστέθηκαν πολλαπλά άλματα σε συγκεκριμένες θέσεις της χρονοσειράς.

Συμπληρωματικά στην παραπάνω μελέτη, έγινε εφαρμογή των μεθόδων ομογενοποίησης από το συγγραφέα σε μη στάσιμες χρονοσειρές, με προσθήκη γραμμικής τάσης στις ομογενείς συνθετικές χρονοσειρές, σύμφωνα με τη Σχέση 5.7.

$$Z_t= Z_t +tb \qquad (5.7)$$

Οι μέθοδοι ομογενοποίησης εφαρμόστηκαν στα δεδομένα πριν και μετά την εισαγωγή ανομογενειών και γραμμικής τάσης.

Ο συγγραφέας εξέτασε στην ανάλυση αυτή πολλές παραμέτρους για να συγκρίνει τις μεθόδους ομογενοποίησης, όπως το μήκος των χρονοσειρών, το ποσοστό χρονοσειρών με ένα άλμα που εντοπίζουν, την ικανότητα των μεθόδων στον εντοπισμό πολλαπλών αλμάτων, το ποσοστό ψευδών εντοπισμών κ.ά.

Ανάλογα με το συνδυασμό των παραμέτρων κάποιες μέθοδοι είχαν καλύτερη ή χειρότερη συμπεριφορά, πάντως δεν φάνηκε κάποια να υπερτερεί σημαντικά. Όσον αφορά το ποσοστό ψευδών εντοπισμών, εξετάστηκαν ομογενείς χρονοσειρές μήκους 50 ετών και διερευνήθηκε σε τι ποσοστό τους εντοπίστηκε τουλάχιστον μία ανομογένεια.

Καλύτερη συμπεριφορά εκτιμήθηκε ότι είχαν οι μέθοδοι BAYE, POTT, NMETA και SNHT με ποσοστά μικρότερα από 5%. Όλες οι άλλες μέθοδοι παρουσίασαν ποσοστά μεγαλύτερα από το ανεκτό 5%, με τη μέθοδο PMETA να έχει τη χειρότερη συμπεριφορά με 17.1% σε μια σειρά προσομοιώσεων.

8. Reeves et al. (2007)

Στη μελέτη αυτή αξιολογήθηκαν 8 απόλυτες μέθοδοι ομογενοποίησης (χωρίς χρήση χρονοσειρών αναφοράς):

- SNH (Standard normal homogeneity test, Alexandersson, 1986)
- NPW (Nonparametric SNH test, Pettit, 1979)
- XLW (Two-phase regression of Wang, Wang, 2003)
- LR (TPR of Lund and Reeves, Lund and Reeves, 2002)
- GNL (New generalised method of this study)
- MLV (Method of Vincent, Vincent, 1998)
- AIC (Akaike's information criteria)
- SBC (Sawa's Bayes criteria)

Για την αξιολόγηση των μεθόδων παράχθηκαν 5 σύνολα 10000 ετήσιων χρονοσειρών, μήκους 100 ετών. Τα 5 αυτά σύνολα παράχθηκαν σύμφωνα με τα 5 μοντέλα που παρουσιάζονται στον Πίνακα 5.1, δηλαδή με θεώρηση επαλληλίας γραμμικού μοντέλου και σφαλμάτων ανεξάρτητων και κανονικά κατανεμημένων. Τα μοντέλα M1 και M2 θεωρούνται ομογενή, ενώ τα μοντέλα M3-M5 μη ομογενή (παρουσία αλμάτων και αλλαγών κλίσεων).

Η μέση τιμή μ των δεδομένων επιλέχτηκε να είναι μηδέν και τα σφάλματα ε_t προσομοιώθηκαν ως λευκός θόρυβος N(0,1). Στα μοντέλα M3-M5 η παράμετρος c (προσομοίωση άλματος) πήρε τιμές 50, 65, ή 80, η παράμετρος Δ τιμές 0.5, 1, ή 2 και οι κλίσεις β_1 και β_2 τιμές 0.005, 0.01, ή 0.02. Η ένδειξη άλματος I είναι μονάδα για $t > c$ και μηδέν $t \leq c$. Οι διαφορές στη σταθερά c αντιπροσωπεύουν άλματα κοντά στο μέσο των χρονοσειρών και ελαφρώς μακρύτερα, εφόσον οι χρονοσειρές έχουν μήκος 100 έτη. Οι τρεις επιλογές για τις κλίσεις β_1 και β_2 αντιπροσωπεύουν μικρή, μεσαία και μεγάλη επίδραση της μεταβλητότητας.

Από την εφαρμογή των μεθόδων στα ομογενή δεδομένα M1 και M2 οι συγγραφείς κατέληξαν ότι δεν υπάρχει κάποια μέθοδος που να υπερτερεί των άλλων. Οι πρώτες έξι μέθοδοι παρουσιάζουν περίπου 5% πιθανότητα να εντοπίσουν κάποια ψευδή ανομογένεια, έχουν δηλαδή ικανοποιητική συμπεριφορά, ενώ η μέθοδος AIC έχει πολύ κακή συμπεριφορά με 70% πιθανότητα εντοπισμού ψευδούς άλματος για το μοντέλο Μ1 και 90% για το μοντέλο Μ2. Η μέθοδος SBC εμφανίζει αρκετά ικανοποιητική συμπεριφορά (ποσοστό ψευδών εντοπισμών 6.7%) για το μοντέλο Μ1 και κακή συμπεριφορά (32%) για το μοντέλο Μ2.

Για τα μη ομογενή δεδομένα των μοντέλων Μ3-Μ5, οι μέθοδοι SNH και NPW είχαν την καλύτερη συμπεριφορά στον εντοπισμό αλμάτων, αλλά όχι άλλων παραμέτρων. Για δεδομένα με κλίσεις, το οποίο ερμηνεύτηκε από τους συγγραφείς ως ένδειξη χρήσης «κακών» χρονοσειρών αναφοράς, καλύτερη συμπεριφορά είχαν οι μέθοδοι XLW και SBC.

Πίνακας 5.1. *Μοντέλα που χρησιμοποιήθηκαν από τους Reeves et al. (2007) για την παραγωγή συνθετικών χρονοσειρών*

Model 1 (M1)	$Y_t = \mu$	$+ \varepsilon_t$
Model 2 (M2)	$Y_t = \mu + \beta_1 t$	$+ \varepsilon_t$
Model 3 (M3)	$Y_t = \mu + \Delta I(t > c)$	$+ \varepsilon_t$
Model 4 (M4)	$Y_t = \mu + \beta_1 t + \Delta I(t > c)$	$+ \varepsilon_t$
Model 5 (M5)	$Y_t = \mu + \beta_1 t + \Delta I(t > c) + \beta_2 t I(t > c)$	$+ \varepsilon_t$

9. **Beaulieu *et al.* (2008):**

Από τους συγγραφείς έγινε αξιολόγηση 8 αντικειμενικών μεθόδων ομογενοποίησης με χρήση ετήσιων συνθετικών δεδομένων βροχόπτωσης. Οι μέθοδοι αυτές ήταν:

- SNHT (Alexandersson, 1986, Khaliq and Ouarda, 2007),
- MREG (multiple regression, Vincent, 1998),
- REG2 (two-phase regression, Easterling and Peterson, 1995, Lund and Reeves, 2002),
- BIVT (bivariate test, Maronna and Yohai, 1978, Potter, 1981),
- WILS (sequential Wilcoxon test, Karl and Williams, 1987, Lanzante, 1996, Ducré-Robitaille et al., 2003),
- STUS (sequential t-test, Gullett et al., 1990),
- JARU (Jaruskova's method, Jaruskova, 1996), και
- BAYE (Bayesian approach, Rasmussen, 2001).

Οι μέθοδοι αυτές επιλέχτηκε να μην λαμβάνουν υπόψη ανομογένειες στα πρώτα και τελευταία 10 έτη των χρονοσειρών, καθώς οι συγγραφείς θεώρησαν ότι το μέγεθος των ανομογενενειών είναι λογικό να υπολογίζεται με τουλάχιστον 10 έτη σε κάθε πλευρά της ανομογένειας.

Η παραγωγή των συνθετικών χρονοσειρών για τον έλεγχο των παραπάνω μεθόδων έγινε με βάση την τεχνική που χρησιμοποιήθηκε από τους *Easterling and Peterson* (1992). Δημιουργήθηκαν βάσεις δεδομένων με συνθετικές χρονοσειρές με τα ίδια στατιστικά χαρακτηριστικά (μέση τιμή, τυπική απόκλιση, αυτοσυσχέτιση για υστέρηση 1 και χωρική ετεροσυσχέτιση) με τυπικές ετήσιες χρονοσειρές βροχόπτωσης καλής ποιότητας (μεγάλο μήκος και λίγα κενά) από το νότιο και κεντρικό τμήμα της επαρχείας του Κεμπέκ στον Καναδά. Όλοι οι πραγματικοί σταθμοί που επιλέχτηκαν πέρασαν το Shapiro-Wilk normality test, υποστηρίζοντας την επιλογή των συγγραφέων να χρησιμοποιήσουν κανονική κατανομή στην παραγωγή των δεδομένων.

Σε ένα πρώτο βήμα παράχθηκαν ομογενείς χρονοσειρές με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1)) σύμφωνα με τη Σχέση 5.8:

$$z_i = \varphi_1 z_{i-1} + e_i \quad (5.8)$$

όπου φ_1 ο συντελεστής αυτοσυσχέτισης για υστέρηση 1 και e_i ένα υπόλοιπο με κανονική κατανομή, μέση τιμή μηδέν και μεταβλητότητα $1-\varphi_1^2$. Η μέση τιμή και η μεταβλητότητα της πραγματικής ετήσιας βροχόπτωσης εισάχθηκαν στη χρονοσειρά $\mathbf{z}_i$.

Τα στατιστικά χαρακτηριστικά που αναπαράχθηκαν ήταν: α) μέση ολική ετήσια βροχόπτωση των σταθμών 1089 mm, β) τυπική απόκλιση 142 mm και γ) αυτοσυσχέτιση για υστέρηση 1 0.02. Οι τιμές αυτές ήταν οι μέσες τιμές από τους πραγματικούς σταθμούς που επιλέχτηκαν. Αν και η αυτοσυσχέτιση θεωρήθηκε μικρή, οι συνθετικές χρονοσειρές επιλέχτηκε να αναπαραχθούν με χρήση μοντέλου αυτοπαλινδρόμησης και όχι ως ανεξάρτητα από κανονική κατανομή, ώστε να αντιπροσωπεύουν όσο το δυνατόν περισσότερο τις πραγματικές χρονοσειρές.

Συνολικά παράχθηκαν 10000 ομογενείς χρονοσειρές, 5000 με μήκος 60 έτη και 5000 με 100 έτη.

Εκτός από την παραπάνω βάση δεδομένων με ομογενείς χρονοσειρές, δημιουργήθηκαν από τους Beaulieu *et al.* (2008) τέσσερις ακόμα βάσεις δεδομένων με μη ομογενείς χρονοσειρές:

- 50000 χρονοσειρές στις οποίες προστέθηκε ένα άλμα (25000 με μήκος 60 έτη και 25000 με μήκος 100 έτη),
- 30000 χρονοσειρές στις οποίες προστέθηκαν περισσότερα του ενός άλματα (15000 με 2 άλματα και 15000 με 3 άλματα),
- 10000 χρονοσειρές με προσθήκη κλίσης, και
- 10000 χρονοσειρές με αλλαγή της μεταβλητότητας.

Για καθεμά από τις παραπάνω χρονοσειρές στην ομογενή τους μορφή (πριν την προσθήκη ανομογενειών), παράχθηκαν τρεις χρονοσειρές αναφοράς σε δύο βήματα:

Αρχικά παράχθηκαν τρεις ομογενείς, ασυσχέτιστες μεταξύ τους χρονοσειρές w, με τον ίδιο τρόπο με τις εξεταζόμενες χρονοσειρές z, δηλαδή σύμφωνα με τη Σχέση 5.9, και με τα ίδια στατιστικά χαρακτηριστικά.

$$w_i = \varphi_1 w_{i-1} + e_i \quad (5.9)$$

Στη συνέχεια εισάχθηκε μια δομή αυτοσυσχέτισης ανάμεσα στην εξεταζόμενη χρονοσειρά και τις χρονοσειρές αναφοράς, σύμφωνα με τη Σχέση 5.10.

$$w_i = \psi z_i + w_i \quad (5.10)$$

όπου ο όρος z_i αντιπροσωπεύει την κανονικοποιημένη συνολική βροχόπτωση της εξεταζόμενης χρονοσειράς τη χρονιά i, ο όρος w_i την κανονικοποιημένη συνολική βροχόπτωση μιας χρονοσειράς αναφοράς τη χρονιά i και ο όρος ψ το συντελεστή συσχέτισης ανάμεσα στην εξεταζόμενη και τη χρονοσειρά αναφοράς.

Οι χρονοσειρές αναφοράς w που προέκυψαν από την παραπάνω διαδικασία κανονικοποιήθηκαν, ώστε να έχουν μέση τιμή μηδέν και μοναδιαία τυπική απόκλιση.

Ο συντελεστής ψ αποφασίστηκε να λάβει την τιμή 0.7 ώστε να αναπαράγει (χωρική) ετεροσυσχέτιση 0.55. Η τιμή αυτή επιλέχτηκε επειδή ο μέσος συντελεστής ετεροσυσχέτισης στην ομάδα των επιλεγμένων πραγματικών σταθμών οι οποίοι απείχαν μεταξύ τους κατά μέσο όρο 300km ήταν 0.55. Η απόσταση 300km είναι αρκετά μεγάλη και οφείλεται στο αραιό δίκτυο σταθμών στην περιοχή του Quebec.

Στις ομογενείς και μη ομογενείς χρονοσειρές που προσομοιώθηκαν, εφαρμόστηκαν στη συνέχεια οι μέθοδοι ομογενοποίησης που επιλέχτηκαν από τους συγγραφείς. Από την εφαρμογή στα ομογενή δεδομένα μήκους 60 και 100 ετών υπολογίστηκαν τα ποσοστά ψευδών εντοπισμών (Πίνακες 4. και 4.).

Τα ποσοστά αυτά ήταν μικρότερα από 5% με εξαίρεση τη μέθοδο BAYE (21.2%). Τα μικρά ποσοστά αυτά αποδόθηκαν από τους συγγραφείς σε ένα βαθμό στο γεγονός ότι δεν έλαβαν υπόψη άλματα κοντά στην άκρη των χρονοσειρών. Με πρόσθεση των ψευδών αλμάτων στα άκρα οι συγγραφείς υποστηρίζουν ότι το ποσοστό αυτό θα ήταν περίπου 5%. Η μέθοδος BAYE έδωσε τόσο μεγάλο ποσοστό ψευδών αλμάτων γιατί αναζητά υποχρεωτικά άλμα στις χρονοσειρές που ελέγχει.

Ακόμα το ποσοστό των ψευδών εντοπισμών στις χρονοσειρές μήκους 60 έτη ήταν μικρότερο σχεδόν για όλες τις μεθόδους από το ποσοστό για χρονοσειρές μήκους 100 έτη.

Όσον αφορά τη συνολική απόδοση των μεθόδων δηλαδή τον εντοπισμό ομογενών χρονοσειρών βροχοπτώσεων ή αλμάτων, τρεις μέθοδοι είχαν παρόμοια απόδοση: οι BIVT, JARU και SNHT.

Η μέθοδος MREG είχε καλή απόδοση στον εντοπισμό ομογενών χρονοσειρών και ενός μοναδικού άλματος, όχι όμως πολλαπλών αλμάτων. Η μέθοδος BAYE είχε καλή συμπεριφορά στον εντοπισμό ενός ή περισσότερων αλμάτων, αλλά εντόπισε παράλληλα και πολλά ψευδή άλματα. Οι μέθοδοι STUS και REG2 είχαν καλή απόδοση στον εντοπισμό ομογενών χρονοσειρών αλλά όχι αλμάτων. Γενικά δεν υπήρξε κάποια μέθοδος που να υπερτερεί των άλλων σε όλους τους τομείς και οι συγγραφείς πρότειναν τη συνδυασμένη χρήση διαφορετικών μεθόδων για τον εντοπισμό των πραγαμτικών ανομογενειών με παράλληλη μείωση των ψευδών εντοπισμών. Τέλος, καμία μέθοδος δεν φάνηκε ευαίσθητη στην αλλαγή της τυπικής απόκλισης των χρονοσειρών.

10. Domonkos (2008):

Από το συγγραφέα εξετάστηκαν 17 εκδοχές από 11 αντικειμενικές μεθόδους ομογενοποίησης, οι οποίες χρησιμοποιούνται ευρέως στις επεξεργασίες κλιματικών δεδομένων. Οι 7 μέθοδοι είναι:

- tta (t-test, Ducré-Robitaille *et al*., 2003)
- ttb (t-test, Kyselý and Domonkos, 2006)
- Buia (Buishand-test, maximum of the absolute values of accumulated anomalies, Buishand, 1982)
- Buib (Buishand-test, difference between maximum and minimum values of accumulated anomalies, Buishand, 1982)
- SNHa (Standard Normal Homogeneity Test for shifts only, Alexandersson, 1986)
- WRS (Wilcoxon Rank Sum test, Karl and Williams, 1987)
- MLR (Multiple Linear Regression, Vincent, 1998)
- Baya (Bayesian test, Ducré-Robitaille et al., 2003, with serial correlation analysis, Sneyers, 1999)
- Bayb (Bayesian test, Ducré-Robitaille et al., 2003, with penalised maximum likelihood method for calculating number of change-points, Caussinus and Lyazrhi, 1997, Mestre, 2004)
- Pett (Pettitt-test, Pettitt, 1979)
- M-K (Mann-Kendall test, Aesawy and Hasanean, 1998)
- Mest (method of Mestre, Mestre, 2004)
- Mesb (method of Mestre with parameterised minimum unit-length)
- SNHT (Standard Normal Homogeneity Test for shifts and trends, Alexandersson and Moberg, 1997)
- East (Easterling-Peterson test, Easterling and Peterson, 1995)
- MASH, (Multiple Analysis of Series for Homogenisation, Szentimrey, 1999)
- MASb, (Multiple Analysis of Series for Homogenisation with parameterised minimum unitlength)

Από τις παραπάνω μεθόδους δύο (MLR και η μεταγενέστερη εκδοχή του Standard Normal Homogeneity Test- SNHT) μπορούν να εντοπίσουν απότομες αλλαγές (άλματα) και σταδιακές αλλαγές (κλίσεις). Οι υπόλοιπες μέθοδοι εντοπίζουν μόνο άλματα στις ελεγχόμενες χρονοσειρές.

Για να παραχθούν όσο το δυνατόν πιο ρεαλιστικά συνθετικά δεδομένα υπολογίστηκαν κάποια στατιστικά χαρακτηριστικά (ροπές, κατανομή, αυτοσυσχέτιση) πραγματικών χρονοσειρών και επίσης μέσα χαρακτηριστικά εντοπισμένων ανομογενειών. Τα πραγματικά δεδομένα αποτελούνταν από 215 χρονοσειρές θερμοκρασίας και 112 χρονοσειρές βροχόπτωσης. Οι χρονοσειρές ήταν ετήσιες ή μηνιαίες και είχαν μήκος 98-100 χρόνια. Οι περισσότερες προήλθαν από το δίκτυο σταθμών της Ουγγρικής Μετεωρολογικής Υπηρεσίας. Από τις πραγματικές χρονοσειρές επιλέχτηκαν εκείνες μόνο που είχαν αυτοσυσχέτιση μεγαλύτερη από 0.4.

Μετά από αρκετές προσομοιώσεις και εξέταση των στατιστικών χαρακτηριστικών των χρονοσειρών, ο συγγραφέας κατέληξε ότι οι εξεταζόμενες χρονοσειρές μπορούν να μοντελοποιηθούν ικανοποιητικά με α) στάσιμο λευκό θόρυβο, β) άλματα σε ορισμένα σημεία, γ) σταδιακές γραμμικές αλλαγές, δ) ανομογένειες τύπου πλατφόρμας, δηλ. δύο διαδοχικά άλματα ίδιου μεγέθους αλλά αντίθετης κατεύθυνσης και ε) ανομογένειες τύπου πλατφόρμας αλλά μικρής διάρκειας.

Το μήκος των προσομοιωμένων χρονοσειρών ήταν σε όλες τις περιπτώσεις 100 χρόνια.

Στις προσοσμοιωμένες χρονοσειρές αφαιρέθηκαν τιμές εκτός του ορίου 4 τυπικών αποκλίσεν και αντικαταστάθηκαν με τη μέση τιμή των χρονοσειρών. Αυτό έγινε γιατί παρατηρήθηκε ότι στις πραγματικές χρονοσειρές τιμές έξω από το κατώφλι αυτό ήταν εξαιρετικά σπάνιες.

Οι 17 μέθοδοι ομογενοποίησης εφαρμόστηκαν στα συνθετικά δεδομένα σε συνδυασμό με ένα ευρύ φάσμα δύο παραγόντων:

- της ελάχιστης απόστασης ανάμεσα σε ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς (u), και
- του επίπεδου εμπιστοσύνης της μεθόδου (c).

Ο αριθμός των προσομοιωμένων χρονοσειρών για κάθε συνδυασμό παραγόντων ήταν 1000.

Ο συγγραφέας έλεγξε την ποιότητα της αποτελεσματικότητα των μεθόδων ομογενοποίησης με μια σύνθεση διαφορετικών παραμέτρων που αφορούν το πόσο καλά εντοπίζουν τις ανομογένειες, κατά πόσο διατηρούν τα αρχικά χαρακτηριστικά των αρχικών χρονοσειρών κ.ά.

Καλύτερες μέθοδοι αποδείχτηκαν σύμφωνα με το κριτήριο του Domonkos (2008) οι Mesb, MASb, Mest, Bayb, MLR, MASH και SNHa. Η βέλτιστη αποτελεσματικότητα όλων των μεθόδων προέκυψε θετική, κάτι που σύμφωνα με το συγγραφέα σημαίνει ότι εφαρμογή οπιασδήποτε από τις μεθόδους αυτές βελτιώνει την ποιότητα των υδροκλιματικών χρονοσειρών (όχι όμως ότι δεν υπάρχουν ψειδείς εντοπισμοί ανομογενειών).

Ένα συμπέρασμα του συγγραφέα που αξίζει να αναφέρουμε είναι ότι οι πιο πολύπλοκες μέθοδοι δεν αποδεικνύονται πάντα πιο αποτελεσματικές από μεθόδους πιο απλές. Για παράδειγμα, η παλιότερη και απλούστερη εκδοχή του SNHT (Alexandersson, 1986), η οποία εντοπίζει μόνο άλματα στα δεδομένα βρέθηκε από τον Domonkos (2008) να έχει καλύτερη συμπεριφορά από τη μεταγενέστερη εκδοχή του η οποία αναζητά άλματα και αλλαγές κλίσεων.

11. Domonkos (2011a):

Από το συγγραφέα εξετάστηκαν 7 σχετικές μέθοδοι ομογενοποίησης, οι οποίες χρησιμοποιούνται ευρέως στις επεξεργασίες κλιματικών δεδομένων. Οι 7 μέθοδοι είναι:

- MLR (Multiple Linear Regression, Vincent, 1998),
- PMT (Penalised Maximal t test, Wang *et al*., 2007),
- SNH (Standard Normal Homogeneity Test for shifts only, Alexandersson, 1986),
- SNHT (Standard Normal Homogeneity Test for shifts and trends, Alexandersson and Moberg, 1997),
- tts (t test, Ducré-Robitaille et al., 2003),
- C-M (Caussinus and Mestre, 2004), και
- MASH (Szentimrey, 1999).

Χρησιμοποιήθηκαν δύο εκδοχές του SNH (SNH1 και SNH2):

Στο SNH1 όταν εντοπίζεται μια ανομογένεια, η χρονοσειρά διαιρείται σε δύο τμήματα και τα τμήματα αυτά εξετάζονται για περαιτέρω ανομογένειες (cutting algorithm). Η διαδικασία αυτή επαναλαμβάνεται στα τμήματα της χρονοσειράς που προκύπτουν κάθε φορά από το σπάσιμο, εφόσον το μήκος των τμημάτων είναι επαρκές για περαιτέρω εξέταση. Το κατώφλι αυτό λήφθηκε από τον Domonkos (2011a) 10 έτη, σύμφωνα με τη συνήθη σύσταση (Easterling and Peterson, 1995, Lanzante, 1996). Ένας επιπλέον περιορισμός στην ανάλυση του Domonkos (2011a) ήταν ότι η ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς λήφθηκε 5 έτη. Η ίδια λογική ακολουθήθηκε και στους αλγορίθμους MLR και PMT.

Στο SNH2 και το SNHT χρησιμοποιήθηκε ο ημι-ιεραρχικός αλγόριθμος που συστήνεται από τους Moberg and Alexandersson (1997). Ο αλγόριθμος αυτός εντοπίζει τις θέσεις των πιθανών ανομογενειών με τη μέθοδο που χρησιμοποιείται και στο SNH1. Στη συνέχεια καθεμιά από τις πιθανές ανομογένειες ξαναελέγχεται με χρήση του τμήματος της χρονοσειράς που περιέχει μόνο την ανομογένεια αυτή και οριοθετείται από μία άλλη ανομογένεια, πιθανή ανομογένεια ή την άκρη της χρονοσειράς. Η διαδικασία αυτή στοχεύει να ελαχιστοποιήσει την επίδραση μιας ανομογένειας στον εντοπισμό και υπολογισμό άλλων ανομογενειών όταν σε μια χρονοσειρά υπάρχουν παραπάνω από μία ανομογένειες.

Επομένως, με βάση τα παραπάνω εξετάστηκαν από τον Domonkos (2011a) 8 διαφορετικές μέθοδοι ομογενοποίησης (6 διαφορετικές μέθοδοι και 2 εκδοχές της ίδιας μεθόδου). Οι μέθοδοι εφαρμόστηκαν για διάστημα εμπιστοσύνης 95%. Στο SNHT, οι εντοπιζόμενες ανομογένειες θεωρήθηκαν τάσεις όταν η διάρκεια της αλλαγής ήταν τουλάχιστον 5 έτη και άλματα στην αντίθετη περίπτωση. Τέλος, πριν την εφαρμογή των μεθόδων έγινε αντικατάσταση των τιμών οι απόλυτες τιμές των οποίων ξεπερνούσαν τις 4 τυπικές αποκλίσεις των χρονοσειρών με τη μέση τιμή.

Οι μέθοδοι ομογενοποίησης εφαρμόστηκαν συνολικά σε 6 βάσεις δεδομένων (A-F) με 10000 χρονοσειρές των 100 ετών η καθεμία. Οι χρονοσειρές αυτές δεν ήταν αναφοράς και ελεγχόμενες όπως στις περισσότερες παρόμοιες μελέτες. Αντίθετα, προσομοιώθηκαν κατευθείαν χρονοσειρές οι οποίες θεωρήθηκαν ως διαφορές ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες.

Οι 6 βάσεις δεδομένων στις οποίες εφαρμόστηκαν οι μέθοδοι ομογενοποίησης παράχθηκαν με το ίδιο μοντέλο λευκού θορύβου (ανεξάρτητα δεδομένα) με τυπική απόκλιση 1.

Το μοντέλο των 6 βάσεων συνθετικών χρονοσειρών ήταν μια επαλληλία λευκού θορύβου (W) με τυπική απόκλιση 1 και συσσωρευτικών επιπτώσεων ανομογενειών (H), οπότε κάθε έτος x_j των προσομοιωμένων χρονοσειρών μπορεί να εκφραστεί σύμφωνα με τη Σχέση 5.11.

$$x_j = w_j + h_j, \qquad j = [1, 2, ..., n] \qquad (5.11)$$

Οι βάσεις δεδομένων διαφοροποιούνταν μεταξύ τους λόγω των διαφορετικών χαρακτηριστικών των ανομογενειών στην καθεμία (αριθμός, είδος, θέση). Στις βάσεις E και F έγινε μια ακόμα διαφοροποίηση από το μοντέλο λευκού θορύβου που χρησιμοιήθηκε. Από την προσομοίωση των ανομογενειών προέκυψαν χρονοσειρές με αυτοσυσχέτιση για υστέρηση 1 μεγαλύτερη από 0.4. Η αυτοσυσχέτιση αυτή ερμηνεύτηκε από τον Domonkos (2011a) ως ενδεικτική του μη ομογενούς χαρακτήρα της προσομοιωμένης χρονοσειράς των διαφορών ανάμεσα στη χρονοσειρά αναφοράς και την ελεγχόμενη. Απλούστερη βάση δεδομένων ήταν η A με 1 μόνο άλμα σε κάθε χρονοσειρά. Οι βάσεις B-D είχαν πολλαπλά άλματα με διαφορετικά χαρακτηριστικά και οι βάσεις E-F πολλαπλές ανομογένειες διαφορετικού τύπου (άλματα, τάσεις κλπ).

Για την αξιολόγηση των μεθόδων ομογενοποίησης χρησιμοποιήθηκαν 4 κριτήρια που σχετίζονταν με την ικανότητα εντοπισμού πραγματικών ανομογενειών, το ποσοστό ψευδών εντοπισμών και την εκτίμηση γραμμικών τάσεων. Το ποσοστό ψευδών εντοπισμών, το οποίο και μας ενδιαφέρει δίνεται από τη Σχέση 5.12.

$$F_{\sim} = \frac{S_F}{S_R + S_F} \qquad (5.12)$$

όπου S_R άθροισμα των σωστών εντοπισμών και S_F άθροισμα των λανθασμένων εντοπισμών.

Από την εφαρμογή των μεθόδων ομογενοποίησης στα συνθετικά δεδομένα ο Domonkos (2011a) συμπέρανε ότι προκύπτει γενικά καλή συμπεριφορά όλων των μεθόδων στον εντοπισμό ανομογενειών, κυρίως στον εντοπισμό ενός άλματος, με εξαίρεση το tts που δεν είχε ικανοποιητική συμπεριφορά.

Ενδιαφέρον παρουσιάζουν τα αποτελέσματα σχετικά με το ποσοστό ψευδούς εντοπισμού ανομογενειών. Τα καλύτερα αποτελέσματα προέκυψαν από τη βάση δεδομένων με ένα μόνο άλμα (A), ενώ και για τη βάση D όλες οι μέθοδοι ομογενοποίησης έδωσαν ποσοστά μικρότερα ή ίσα περίπου με 5%. Στις βάσεις B, C και E τα ποσοστά κυμαίνονται μεταξύ 5-20% περίπου, ενώ πολύ μεγάλα ποσοστά παρατηρούνται στη βάση F (30-45% περίπου). Η πολύ μεγάλη διαφορά της βάσης F αποδίδεται από το συγγραφέα στις χρονικά έμμονες ανομογένειες που περιέχει, οι οποίες εμποδίζουν τον εντοπισμό των πραγματικών ανομογενειών, αυξάνοντας ως εκ τούτου το δείκτη των ψευδών εντοπισμών.

Όσον αφορά τις διαφορετικές μεθόδους, καλύτερα αποτελέσματα σύμφωνα με το συγγραφέα στις βάσεις A-C έδωσαν οι μέθοδοι PMT και SNH και στις βάσεις D-F οι μέθοδοι C-M και MASH. Για το συμπέρασμα αυτό έλαβε υπόψη του το ποσοστό ψευδών εντοπισμών αλλά και την ισχύ του εντοπισμού ανομογενειών από τις 8 διαφορετικές μεθόδους. Κρίνοντας ότι οι βάσεις E και F ήταν οι πιο ρεαλιστικές, ο συγγραφέας κατέληξε ότι οι μέθοδοι ομογενοποίησης με την καλύτερη συμπεριφορά ήταν οι C-M και MASH.

Όσον αφορά τη σύγκριση των μεθόδων SNH1 και SNH2, αξίζει να αναφερθεί ότι τα αποτελέσματα SNH2 ήταν ελάχιστα καλύτερα από αυτά του SNH1, με μια διαφορά όμως απογοητευτικά μικρή σύμφωνα με το συγγραφέα (τα σημεία στα διαγράμματα εντοπισμού ανομογενειών και ποσοστού ψευδούς εντοπισμού ανομογενειών σχεδόν ταυτίζονται).

12. Venema *et al.* (2012):

Η μελέτη αυτή είναι η πιο πρόσφατη και η μεγαλύτερου εύρους αξιολόγηση μεθόδων ομογενοποίησης που έχει γίνει μέχρι σήμερα. Συνολικά αξιολογήθηκαν 25 διαφορετικές εκδοχές των μεθόδων ομογενοποίησης που παρουσιάζονται στον Πίνακα 5.2. Όλες οι μέθοδοι ήταν σχετικές, έκαναν δηλαδή χρήση χρονοσειρών αναφοράς, εκτός από την PMFred (Wang, 2008) που ήταν απόλυτη.

Για τη μελέτη χρησιμοποιήθηκαν τριών ειδών μηνιαίες βάσεις δεδομένων, με (α) πραγματικά, (β) υποκατάστατα (surrogate), και (γ) συνθετικά δεδομένα θερμοκρασίας και βροχόπτωσης.

Πίνακας 5.2. *Μέθοδοι ομογενοποίησης που αξιολογήθηκαν από τους Venema et al. (2012). Για πολλές από τις μεθόδους εξετάστηκαν πάνω από μία διαφορετικές εκδοχές.*

	Comparison		Detection		References
Method	Comparison	Time step	Search	Criterion	
MASH	Multiple references	Annual, parallel monthly	Exhaustive	Statistical test (MLR)	Szentimrey (2007, 2008)
PRODIGE	Pairwise, human synthesis	Annual, parallel monthly	DP	Penalized Likelihood	Caussinus and Mestre (2004)
USHCN	Pairwise, automatic synthesis	Serial monthly	HBS	Statistical test (MLR)	Menne et al. (2009)
AnClim	Reference series	Annual, parallel monthly	HBS, moving window	Statistical test	Štepanek et al. (2009)
Craddock	Pairwise, human synthesis	Serial monthly	Visual	Visual	Craddock (1979); Brunetti et al. (2006)
RhtestV2	Reference series or absolute	Serial monthly	Stepwise	Statistical test (modified Fisher)	Wang (2008)
SNHT	Reference series	Annual	HBS	Statistical test (MLR)	Alexandersson and Moberg (1997)
Climatol	Reference series	Parallel monthly	HBS, moving window	Statistical test	Guijarro (2011)
ACMANT	Reference series	Annual, joint seasonal	DP	Penalized Likelihood	Domonkos et al. (2011)

DP = Dynamic programming (optimization method); HBS = (semi-)hierarchic binairy splitting; MLR = Maximum Likelihood Ratio test.

Τα υποκατάστατα δεδομένα είναι ουσιαστικά συνθετικά δεδομένα τα οποία αναπαράγουν κάποια στατιστικά χαρακτηριστικά πραγματικών δεδομένων. Στη μελέτη αυτή παράχθηκαν με χρήση του αλγορίθμου Iterative Amplitude Adjusted Fourier Transform Algorithm (IAAFT), τροποποιημένο από τους Venema *et al*. (2006b). Ο αλγόριθμος αυτός ξεκινά από λευκό θόρυβο και αναπαράγει την κατανομή και το φάσμα Fourier των πραγματικών δεδομένων.

Για να δημιουργηθεί η βάση υποκατάστατων δεδομένων παράχθηκαν δίκτυα 5, 9 ή 15 σταθμών μήκους 100 ετών (1900 μέχρι 1999). Τα υποκατάστατα δεδομένα βασίστηκαν σε ομογενοποιημένα δεδομένα θερμοκρασίας και βροχόπτωσης χωρίς κενά ή με εκτιμημένες τιμές για τα υπάρχοντα κενά. Τα δεδομένα θερμοκρασίας προήλθαν από την Αυστρία, τη Γαλλία και την περιοχή της Καταλονίας και τα δεδομένα βροχόπτωσης από την Αυστρία και τη Γαλλία.

Οι χρονοσειρές βροχόπτωσης που χρησιμοποιήθηκαν δεν περιείχαν μηδενικές τιμές. Πριν την παραγωγή των υποκατάστατων δεδομένων αφαιρέθηκαν η μέση τιμή και η γραμμική τους τάση και περιορίστηκαν σε διάρκεια 100 χρόνων. Στα δεδομένα θερμοκρασίας πριν την παραγωγή υποκατάστατων αφαιρέθηκαν η μέση τιμή και η εποχικότητα. Οι μέσες τιμές και η εποχικότητα των δεδομένων θερμοκρασίας προστέθηκαν ξανά στα δεδομένα μετά τη δημιουργία υποκατάστατων. Κάποιες χρονοσειρές θερμοκρασίας είχαν μήκος μικρότερο των 100 ετών και επεκτάθηκαν με επαναλαμβανόμενη αντιγραφή τους (mirroring) μέχρι να συμπληρωθεί 100ετία.

Η βάση συνθετικών δεδομένων δημιουργήθηκε από τη βάση υποκατάστατων δεδομένων ως εξής:

Δημιουργήθηκε μια χρονοσειρά M (ο συμβολισμός δικός μας για καλύτερη κατανόηση) ως μέση τιμή κάθε δικτύου 5, 9 ή 15 σταθμών των υποκατάστατων δεδομένων. Στη συνέχεια υπολογίστηκε για κάθε διαφορετική χρονοσειρά του δικτύου η διαφορά (για θερμοκρασιακά δεδομένα) ή λόγος της (για δεδομένα βροχόπτωσης) από τη χρονοσειρά M. Οι σχετικές αυτές χρονοσειρές μετατράπηκαν σε χρονοσειρές Γκαουσιανού λευκού θορύβου διατηρώντας τη μέση τιμή, την τυπική απόκλιση και το μητρώο ετεροσυσχετίσεών τους. Στη συνέχεια τους προστέθηκε η μέση τιμή του δικτύου. Τέλος, τυχόν αρνητικές τιμές στα δεδομένα βροχόπτωσης αντικαταστάθηκαν με μηδενικές.

Στα τεχνητά υποκατάστατα και συνθετικά δεδομένα που προέκυψαν με την παραπάνω διαδικασία προστέθηκαν γνωστές ανομογένειες (άλματα και αλλαγές τάσεων), ακραίες τιμές, κενά και μια παγκόσμια μη γραμμική τάση. Κάθε δίκτυο σταθμών είχε διαφορετική παγκόσμια τάση ώστε να μην διευκολύνεται η ομογενοποίηση των δεδομένων.

Τα πραγματικά, υποκατάστατα και συνθετικά δεδομένα θερμοκρασίας και βροχόπτωσης ομογενοποιήθηκαν από πολλούς διαφορετικούς χρήστες αλγορίθμων ομογενοποίησης. Η μελέτη ήταν «τυφλή», δηλαδή οι χρήστες δεν γνώριζαν τις θέσεις και τα μεγέθη των ανομογενειών. Η ομογενοποίηση έγινε με χρήση μιας συνθετικής χρονοσειράς αναφοράς από τις υπάρχουσες η οποία να θεωρείται ομογενής (π.χ. SNHT), η μη ομογενής (π.χ. MASH) ή με απευθείας σύγκριση των χρονοσειρών κατά ζεύγη (π.χ. USHCN, PRODIGE). Για τις θερμοκρασίες και τις βροχοπτώσεις λήφθηκαν, αντίστοιχα, διαφορές και λόγοι ανάμεσα στη χρονοσειρά αναφοράς και την ελεγχόμενη.

Για την αξιολόγηση της συμπεριφοράς των μεθόδων ομογενοποίησης χρησιμοποιήθηκαν αρκετά διαφορετικά κριτήρια:

α) Χρησιμοποιήθηκαν κριτήρια που σχετίζονταν με τη χρονική συμβατότητα των ομογενοποιημένων δεδομένων με τα αρχικά ομογενή δεδομένα. Τα κριτήρια αυτά βασίστηκαν στον υπολογισμό της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) των διαφορών ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες, στις οποίες όμως χρονοσειρές αφαιρέθηκε πριν τον υπολογισμό των διαφορών η μέση τιμή (centered RMSE, CRMSE). Ο υπολογισμός του CRMSE έγινε για διάφορες χρονικές κλίμακες (μηνιαία, ετήσια κ.ά.).

β) Επιπλέον εξετάστηκαν οι διαφορές των γραμμικών τάσεων ανάμεσα στα αρχικά ομογενή και τα ομογενοποιημένα δεδομένα και κριτήρια που σχετίζονταν με το ποσοστό των πραγματικών και ψευδών ανομογενειών που εντόπισαν οι μέθοδοι ομογενοποίησης.

Συγκεκριμένα όσον αφορά τους ψευδείς εντοπισμούς ανομογενειών εξετάστηκε η πιθανότητα ψευδούς εντοπισμου (probability of false detection, POFD), η οποία ορίζεται ως :

$$POFD=b/(b+d) \qquad (5.13)$$

όπου b: ο αριθμός των ψευδών ανομογενειών που εντοπίζονται, και

d: ο αριθμός των μηνών, ετών κλπ στα οποία δεν εντοπίζεται ούτε υπάρχει πραγματικό άλμα.

Η εξαγωγή των συμπερασμάτων από την παραπάνω μελέτη βασίστηκε κυρίως στα αποτελέσματα της ομογενοποίησης των υποκατάστατων δεδομένων τα οποία θεωρήθηκαν πιο ρεαλιστικά από τα συνθετικά. Καλύτερες μέθοδοι από την τυφλή μελέτη με βάση συνδυασμό των κριτηρίων αξιολόγησης θεωρήθηκαν οι Craddock, MASH και PRODIGE. Εκ των υστέρων βελτίωση όμως της μεθόδου ACMANT με βάση τα αποτελέσματα της ομογενοποίησης έδωσε συνολικά τα καλύτερα αποτελέσματα. Χειρότερη μέθοδος, η οποία έκανε τα δεδομένα πιο ανομογενή αποδείχτηκε η PMFred, η μόνη απόλυτη μέθοδος που αξιολογήθηκε. Λόγω του αποτελέσματος αυτού οι συγγραφείς συνιστούν την εφαρμογή απόλυτης ομογενοποίησης μόνο σε συνδυασμό με μεταδεδομένα.

Όσον αφορά τα ποσοστά ψευδούς εντοπισμού όλες οι μέθοδοι είχαν καλή συμπεριφορά με ποσοστά μικρότερα ή ίσα με 5%. Εξαίρεση αποτέλεσε η μέθοδος MASH με ποσοστό περίπου 10%. Οι συγγραφείς αποδίδουν το ποσοστό αυτό στην πολύ μεγάλη ικανότητα της μεθόδου στον εντοπισμό αλμάτων στα μηνιαία δεδομένα, κάτι που το οδηγεί και σε μεγαλύτερο αριθμό ψευδών εντοπισμών.

Όσον αφορά τους διαφορετικούς τύπους δεδομένων, τα λιγότερο ρεαλιστικά συνθετικά δεδομένα αποδείχτηκαν πιο εύκολο να ομογενοποιηθούν και είχαν 15% μικρότερο σφάλμα στην εκτίμηση της γραμμικής τάσης. Τέλος, η ομογενοποίηση δεδομένων θερμοκρασίας έδωσε πολύ καλύτερα αποτελέσματα από την ομογενοποίηση δεδομένων βροχόπτωσης, στα οποία σε πολλές περιπτώσεις η ομογενοποίηση οδήγησε σε χειρότερης ποιότητας δεδομένα από τα αρχικά ανομογενή. Με βάση τα συμπεράσματα αυτά οι συγγραφείς συνιστούν πρώτον, να μη γίνεται ομογενοποίηση σε μηνιαία αλλά σε ετήσια δεδομένα βροχόπτωσης, και δεύτερον, περαιτέρω διερεύνηση των αποτελεσμάτων της ομογενοποίησης σε δεδομένα βροχόπτωσης.

5.4.2. Ανάπτυξη και έλεγχος καινούριας μεθόδου ή τροποποίηση/εφαρμογή μιας προϋπάρχουσας μεθόδου

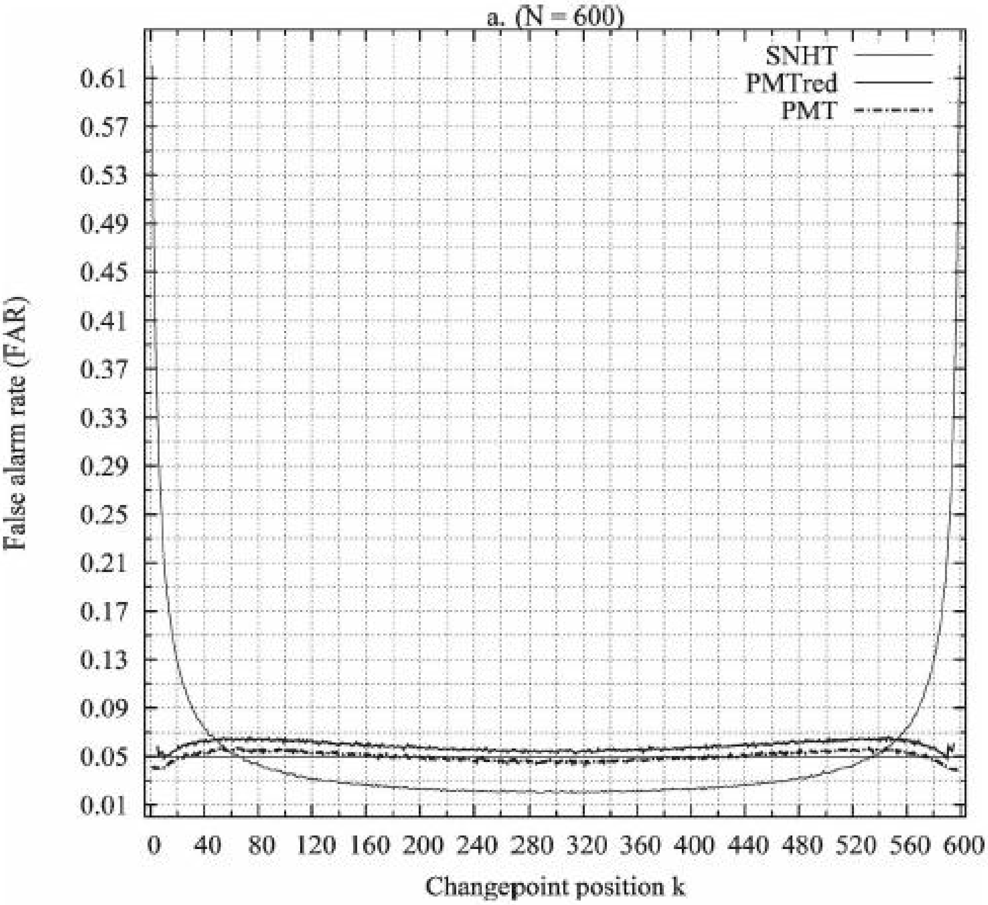
1. **Wang (2008)**

Η συγκεκριμένη μελέτη έγινε από τη συγγραφέα για να τροποποιήσει δύο μεθόδους που αναπτύχθηκαν από την ίδια, το PMT και το PMF, ώστε να λαμβάνουν υπόψη την αυτοσυσχέτιση των δεδομένων. Οι τροποποιημένες μέθοδοι ονομάστηκαν PMTred και PMFred.

Από τη συγκεκριμένη δημοσίευση θα αναφερθούμε μόνο σε ένα σημείο που κρίθηκε σημαντικό, δηλαδή τον έλεγχο των μεθόδων PMTred και PMFred σε συνθετικά δεδομένα και συγκεκριμένα τα ποσοστά ψευδών εντοπισμών που εντοπίστηκαν. Θα επικεντρωθούμε στη μέθοδο PMTred, καθώς έγινε σύγκρισή της με μια από τις συνηθισμένες μεθόδους ομογενοποίησης, το SNHT.

Η μέθοδος PMTred ελέγθηκε σε σύγκριση με τις μεθόδους PMT και SNHT. Για τον έλεγχο έγινε προσομοίωση $(N - 2) \times (N_{min} + 1) \times 10000$ ομογενών, μηνιαίων, ανεξάρτητων μεταξύ τους χρονοσειρών μήκους 600 μηνών (50 έτη), με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1)). N_{min} είναι το ελάχιστο τμήμα της χρονοσειράς που θεωρεί η μέθοδος. Το μήκος N των χρονοσειρών λήφθηκε N=600. Ο συντελεστής αυτοσυσχέτισης για υστέρηση 1 του μοντέλου λήφθηκε φ=0.1925.

Από την παραπάνω προσομοίωση υπολογίστηκε το ποσοστό ψευδούς εντοπισμού (False alarm rate - FAR) των τριών μεθόδων συναρτήσει των σημείων στα οποία εντοπίστηκαν ανομογένειες (Σχήμα 5.1). Το κριτήριο FAR εκτιμήθηκε ως ο λόγος των συνολικών στατιστικά σημαντικών αλμάτων που βρέθηκαν από την κάθε μέθοδο προς τον ολικό βαθμό προσομοιώσεων.



Σχήμα 5.1. *Κατανομή του ποσοστού ψευδούς εντοπισμού (FAR) της μεθόδου PMTred συγκριτικά με τις αντίστοιχες κατανομές των μεθόδων PMT και SNHT. Πηγή: Wang (2008).*

Από το Σχήμα 5.1 φαίνεται ότι τα PMTred και PMT είχαν παρόμοια συμπεριφορά με ομοιόμορφη κατανομή του FAR για εντοπισμό αλμάτων σε όλα τα σημεία k των χρονοσειρών. Το PMTred είχε ελαφρά καλύτερη συμπεριφορά από το PMT, όμως η συγγραφέας αναφέρει ότι η συμπεριφορά του μπορεί να βελτιωθεί με κάποιες υποκειμενικές επιλογές του χρήστη της μεθόδου. Το SNHT είχε καλύτερη συμπεριφορά στο μέσο των χρονοσειρών από τις δύο άλλες μεθόδους, αλλά πολύ κακή συμπεριφορά στα άκρα των χρονοσειρών με πολύ μεγάλη αύξηση του FAR. Από το Σχήμα 5.1 φαίνεται ότι πρέπει να αποφεύγεται εφαρμογή του στους 60 μήνες (αντιστοιχούν σε 5 έτη) στην αρχή και το τέλος των μηνιαίων χρονοσειρών (FAR>5%).

Ακολούθως παράχθηκαν 10000 ακόμα χρονοσειρές με χρήση μοντέλου αυτοπαλινδρόμησης τάξης 1 (AR(1) για καθεμία από τις περιπτώσεις: 1) φ=0 και N=600, 2) φ=0.1925 και N=600, και 3) φ=0.1366 και N=1200 και έγινε και πάλι υπολογισμός του FAR. Η συμπεριφορά του PMT, αλλά και του PMTred κρίθηκαν ικανοποιητικές.

Η Wang (2008) έκανε έλεγχο και της μεθόδου PMFred, η οποία θεωρεί δεδομένα με γραμμική τάση. Επιλέχτηκαν 43 τιμές του συντελεστή φ: {-0.200, -0.150, -0.100, -0.050, 0, 0.025, ..., 0.0950}. Για καθεμία από τις τιμές αυτές παράχθηκαν 10000000 ανεξάρτητες μεταξύ τους χρονοσειρές με χρήση μοντέλου AR(1), στις οποίες προστέθηκε γραμμική τάση β=0.01. Η συμπεριφορά της μεθόδου σε όρους FAR βρέθηκε από τη συγγραφέα καλή (περίπου 5%).

2. Rust *et al*. (2008)

Η συγκεκριμένη δημοσίευση παρουσιάζει μεγάλο ενδιαφέρον γιατί είναι η μόνη που βρέθηκε στη βιβλιογραφία και διερευνά τη σχέση της ομογενοποίησης και της μακροπρόθεσμης εμμονής των χρονοσειρών. Οι συγγραφείς χρησιμοποίησαν τη μέθοδο ομογενοποίησης που παρουσιάζεται από τους Caussinus and Mestre (2004), την οποία στη συνέχεια αναφέρεται ως CM.

Για τα δεδομένα θερμοκρασίας θεωρήθηκε ότι οι χρονοσειρές αποτελούν επαλληλίες: α) της επίδρασης του κλίματος μ_t, β) της επίδρασης του σταθμού ν_t, και γ) τυχαίου λευκού θορύβου.

Ο όρος μ_t, θεωρήθηκε ότι έχει μέση τιμή μηδέν, εξετάστηκαν επομένως θερμοκρασιακές ανωμαλίες (anomalies, διαφορές από τη μέση τιμή μιας περιόδου). Η επίδραση του σταθμού ν_t θεωρήθηκε σταθερή για αξιόπιστες χρονοσειρές και με άλμα για μη αξιόπιστες χρονοσειρές.

Οι διαφορές ανάμεσα σε χρονοσειρές αναφοράς και ελεγχόμενες θεωρήθηκε ότι είναι ανεξάρτητες και με κανονική κατανομή, σύμφωνα με προηγούμενες προσεγγίσεις, π.χ. Alexandersson (1986).

Οι συγγραφείς εφάρμοσαν τη μέθοδο CM σε μηνιαία συνθετικά δεδομένα τα οποία αναπαράγουν το συντελεστή Hurst και τη χωρική ετεροσυσχέτιση πραγματικών χρονοσειρών. Οι πραγματικές χρονοσειρές που επιλέχτηκαν ήταν 24 σταθμοί μέγιστης θερμοκρασίας από την επιφάνεια του εδάφους (surface air temperature) στη Γαλλία. Τα δεδομένα αυτά ήταν ήδη ομογενοποιημένα από τους Caussinus and Mestre (2004). Κάποιοι από τους σταθμούς αυτούς περιείχαν δεδομένα από τα τέλη του 19^ου^ αιώνα και ήταν σε μορφή ανωμαλιών. Η ομογενοποίηση των σταθμών διόρθωσε άλματα που εντοπίστηκαν αλλά δεν αναφέρονταν στα μεταδεδομένα των σταθμών.

Ο συντελεστής Hurst των αρχικών και ομογενοποιημένων πραγματικών σταθμών εκτιμήθηκε με δύο μεθόδους:

- με μια παραμετρική προσέγγιση η οποία βασίζεται στην ευέλικτη κατηγορία μοντέλων FARIMA[p; d; q], (συγκεκριμένα χρησιμοποιήθηκαν τα μοντέλα FARIMA[1; d; 0] και FARIMA[1; d; 1]), και

- με ανάλυση των διακυμάνσεων μετά την αφαίρεση της τάσης (detrended fluctuation analysis, DFA), η οποία είναι ευρυστική προσέγγιση.

Η δεύτερη μέθοδος θεωρήθηκε λιγότερο ακριβής αλλά χρησιμοποιήθηκε λόγω της συχνής χρήσης της στις γεωεπιστήμες. Οι μέσες τιμές του συντελεστή Hurst από τους 24 σταθμούς πριν και μετά την ομογενοποίηση παρουσιάζονται στον Πίνακα 5.3. Στις περισσότερες περιπτώσεις ο συντελεστής Hurst μειώθηκε από την ομογενοποίηση, π.χ. στην περίπτωση του FARIMA[1; d; 0] μειώθηκε στο 80% των σταθμών.

Πίνακας 5.3. *Μέση τιμή των εκτιμήσεων του συντελεστή Hurst για τις 24 αρχικές και ομογενοποιημένες χρονοσειρές θερμοκρασιακών ανωμαλιών που χρησιμοποίησαν οι Rust et al. (2008). Παρουσιάζονται εκτιμήσεις δύο διαφορετικών τύπων μοντέλων.*

	FARIMA		DFA		
Series	$[1, d, 0]$	$[1, d, 1]$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$
Raw	0.657(0.008)	0.707(0.012)	0.675	0.707	0.751
Corrected	0.619(0.008)	0.641(0.011)	0.634	0.651	0.690

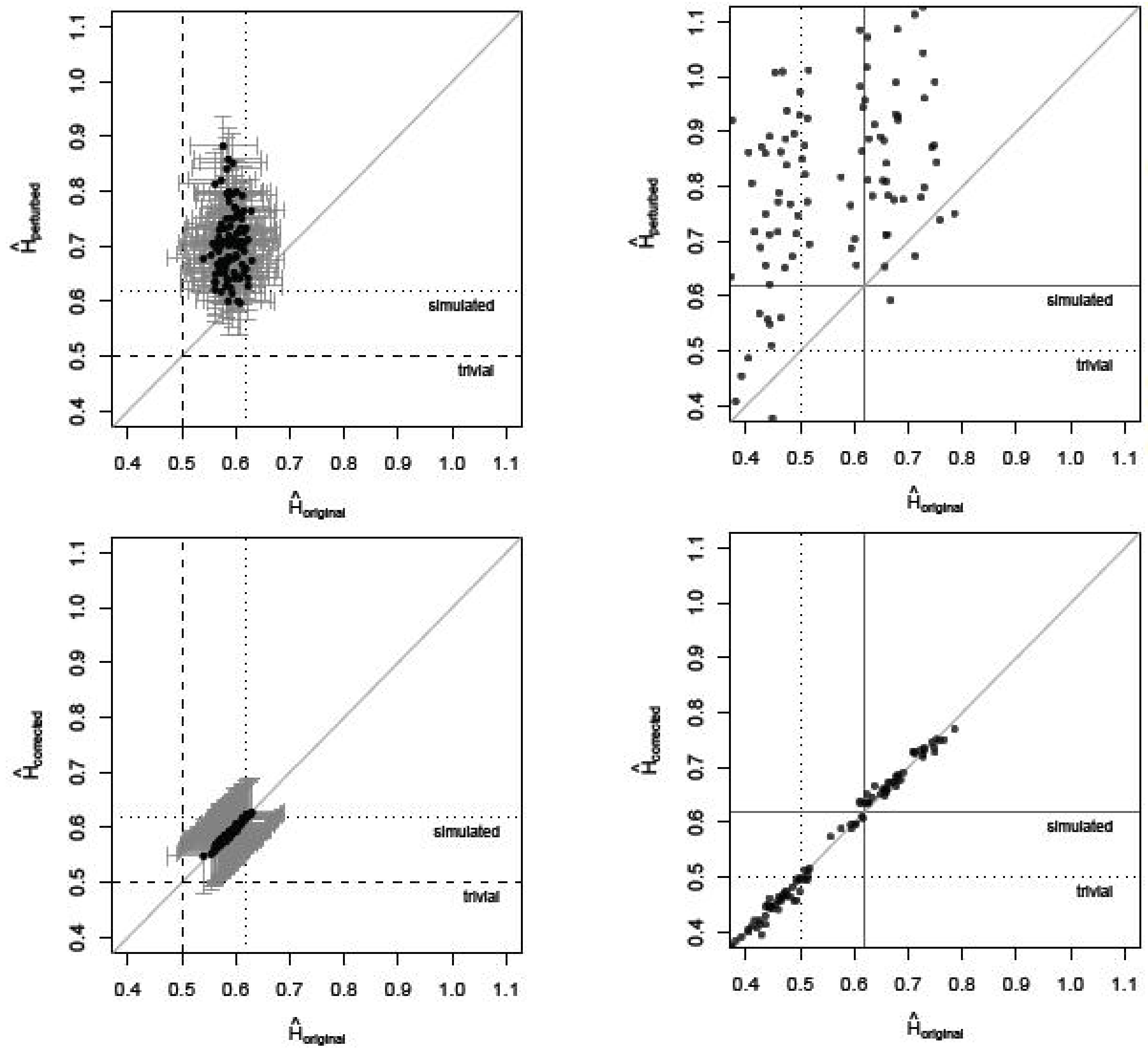
Με βάση τα ομογενοποιημένα πραγματικά δεδομένα έγινε προσομοίωση συνθετικών δεδομένων με μοντέλο FARIMA[1,d,0] και παραμέτρους d=0.119, a=0.061 και σ_n= 3.414 (παρόμοιες τιμές με αυτές των πραγματικών χρονοσειρών ανωμαλιών). Στις παραμέτρους αυτές αντιστοιχεί, σύμφωνα με τους συγγραφείς, συντελεστής Hurst 0.62 (βλ. ομογενοποιημένα δεδομένα του Πίνακα 5.3). Στη συνέχεια προστέθηκε ένα περιοδικό σήμα το οποίο να αναπαριστά τον ετήσιο κύκλο.

Το μήκος των χρονοσειρών ήταν 150 έτη, δηλαδή 12×150 μήνες.

Για κάθε χρονοσειρά που παράχθηκε με αυτό τον τρόπο έγιναν 9 «αντίγραφα» και προστέθηκε σε καθεμία από τις 10 χρονοσειρές που προέκυψαν λευκός θόρυβος με τυπική απόκλιση 0.25, ώστε να δημιουργηθεί ένα δίκτυο συσχετισμένων σταθμών.

Στις τελικές χρονοσειρές που προέκυψαν από την παραπάνω διαδικασία προστέθηκαν άλματα σε τυχαία αλλά γνωστά σημεία. Οι μη ομογενείς χρονοσειρές ομογενοποιήθηκαν με τη μέθοδο CM θεωρώντας τα σημεία των αλμάτων γνωστά.

Για την ανάλυση των αποτελεσμάτων από τις αρχικές, μη ομογενείς και ομογενοποιημένες χρονοσειρές αφαιρέθηκε ο ετήσιος κύκλος και υπολογίστηκε ο συντελεστής Hurst με χρήση μοντέλου FARIMA[1,d,0] και εκτιμητή βασισμένο σε DFA. Τα αποτελέσματα των εκτιμήσεων αυτών φαίνονται στο Σχήμα 5.2. Με βάση τα αποτελέσματα αυτά, οι Rust *et al.* (2008) συμπεραίνουν ότι οι ανομογένειες αυξάνουν το συντελεστή Hurst των χρονοσειρών και η ομογενοποίηση αποκαθιστά την αρχική του τιμή.



Σχήμα 5.2. *Εκτιμήσεις του συντελεστή Hurst των συνθετικών χρονοσειρών που χρησιμοποιήθηκαν από τους Rust et al. (2008), με δύο διαφορετικές μεθόδους: μοντέλο FARIMA[1; d; 0] αριστερά και μοντέλο DFA δεξιά. Στα γραφήματα πάνω φαίνεται η εκτίμηση του συντελεστή Hurst στα αρχικά ομογενή δεδομένα έναντι των αρχικών δεδομένων μετά από προσθηκη ανομογενειών, ενώ στα γραφήματα κάτω η εκτίμηση του συντελεστή Hurst στα αρχικά ομογενή δεδομένα έναντι των τελικών ομογενοποιημένων δεδομένων. Και με τους δύο τρόπους εκτίμησης φαίνεται η ομογενοποίηση να αποκαθιστά τους αρχικούς συντελεστές Hurst των συνθετικών χρονοσειρών. Πηγή: Rust et al. (2008) μετά από τροποποίηση.*

5.5. Σύγκριση - σχολιασμός προηγούμενων προσεγγίσεων

Από την επισκόπηση των μελετών των Παραγράφων 5.4.1 και 5.4.2 μπορούν να εξαχθούν κάποια βασικά συμπεράσματα σε σχέση με τις μεθόδους που χρησιμοποιούνται για την ομογενοποίηση των δεδομένων, τον τρόπο εφαρμογής τους, τα συνθετικά δεδομένα με βάση τα οποία ελέγχονται και τα κριτήρια με τα οποία αξιολογούνται.

5.5.1. Χαρακτηριστικά συνθετικών δεδομένων

Από τις μελέτες που εξετάστηκαν 3 μόνο αφορούν δεδομένα βροχόπτωσης (Buishand, 1982, Lubes-Niel *et al.*, 1998, Beaulieu *et al.*, 2008). Οι υπόλοιπες χρησιμοποιούν συνθετικά δεδομένα θερμοκρασίας, ενώ μόνο από τους Venema *et al.* (2012) εξετάζονται και οι δύο περιπτώσεις δεδομένων. Το στοιχείο αυτό δείχνει και τη μεγάλη σημασία που δίνεται τις τελευταίες δεκαετίες στα δεδομένα θερμοκρασίας και την επεξεργασία τους.

Η κλίμακα των δεδομένων που εξετάζεται από τους περισσότερους ερευνητές είναι ετήσια, ενώ από το 2008 και μετά παρατηρείται μια στροφή στην εφαρμογή των μεθόδων ομογενοποίησης και την αξιολόγησή τους σε μηνιαία κλίμακα. Το μήκος συνθετικών χρονοσειρών που επιλέγεται τις περισσότερες φορές είναι 100 έτη. Πάντως εφαρμογή των μεθόδων σε χρονοσειρές διαφορετικού μήκους, με όλους τους υπόλοιπους παράγοντες σταθερούς, έδωσε διαφορετικά αποτελέσματα. Για παράδειγμα, ομογενοποίηση χρονοσειρών βροχόπτωσης μήκους 60 και 100 ετών από τους Beaulieu *et al.* (2008) έδωσε μικρότερα ποσοστά ψευδών εντοπισμών για μήκος 60 έτη σε σχέση με 100.

Το μοντέλο με το οποίο παράγονται τα δεδομένα είναι κοινό σε όλες σχεδόν τις μελέτες που βρέθηκαν με μόνη εξαίρεση τους Rust *et al.* (2008). Οι Rust *et al.* (2008) εφαρμόζουν τη μέθοδο ομογενοποίησης των Caussinus and Mestre (2004) σε μη ομογενή συνθετικά δεδομένα από μοντέλο FARIMA[1,d,0] με αρχικό συντελεστή Hurst (ενδεικτικό της μακροπρόθεσμης εμμονής των δεδομένων), δηλαδή πριν την προσθήκη ανομογενειών H=0.62. Σε όλες τις υπόλοιπες περιπτώσεις που εξετάστηκαν, τα συνθετικά δεδομένα παράγονται με χρήση μοντέλου AR(1), λευκού θορύβου (δεδομένα χωρίς αυτοσυσχέτιση) και στην περίπτωση των υποκατάστατων δεδομένων από τους Venema *et al.* (2012) με αναπαραγωγή του φάσματος πραγματικών χρονοσειρών, χωρίς όμως να αναφέρεται ότι αναπαράγεται η μακροπρόθεσμη εμμονή.

Όλοι σχεδόν οι συγγραφείς επίσης παράγουν συνθετικά δεδομένα από κανονική κατανομή. Ασύμμετρες κατανομές χρησιμοποιούνται από τους Lubes-Niel *et al.* (1998) για την παραγωγή δεδομένων βροχοπτώσεων.

Οι προσομοιωμένες χρονοσειρές που παράγονται για την αξιολόγηση των μεθόδων ομογενοποίησης είναι συνήθως ομάδες χρονοσειρών αναφοράς – ελεγχόμενων ή απευθείας διαφορές τους. Γενικά στη χρήση χρονοσειρών αναφοράς παρατηρήθηκε ότι κυρίαρχη είναι η κλασική στατιστική αντίληψη που θεωρεί δεδομένα χωρίς αυτοσυσχέτιση. Από όλους τους ερευνητές, ακόμα και από τους Rust *et al.* (2008), οι οποίοι ήταν και οι μόνοι που εξέτασαν δεδομένα με μακροπρόθεσμη εμμονή, επικρατεί η αντίληψη που εκφράστηκε από τους Conrad and Pollak (1962) ότι μια χρονοσειρά κλιματικών δεδομένων (θερμοκρασιών ή βροχόπτωσης) είναι ομογενής σε σχέση με μια σύγχρονή της χρονοσειρά σε μια άλλη τοποθεσία αν οι διαφορές των θερμοκρασιών ή οι λόγοι των βροχοπτώσεων συνθέτουν μια χρονοσειρά τυχαίων αριθμών, δηλαδή λευκού θορύβου (Karl and Williams, 1987).

Τέλος, όσον αφορά το συντελεστή ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες, όλοι οι συγγραφείς χρησιμοποιούν τιμές από 0.45-0.9. Τυπική τιμή του συντελεστή ετεροσυσχέτισης θεωρείται από πολλούς η τιμή ρ_{XY}=0.8 (Easterling and Peterson, 1995, Ducré-Robitaille *et al.*, 2003, DeGaetano, 2006).

5.5.2. Μέθοδοι ομογενοποίησης

Στις μελέτες που παρουσιάζονται στην Παράγραφο 5.4 συγκρίνονται και αξιολογούνται διαφορετικές μέθοδοι ομογενοποίησης σε συνθετικά δεδομένα με διαφορετικά χαρακτηριστικά, οπότε δεν είναι εύκολο να καταλήξει κανείς σε σχέση με το ποια μέθοδος φαίνεται να έχει καλύτερη απόδοση.

Οι περισσότερες μέθοδοι που εξετάζονται είναι σχετικές, κάνουν χρήση δηλαδή χρονοσειρών αναφοράς. Ελάχιστες είναι οι περιπτώσεις απόλυτων μεθόδων (π.χ. Buishand, 1982, Reeves *et al.*, 2007).

Η μέθοδος που έχει αξιολογηθεί πιο πολλές φορές είναι με διαφορά το SNHT για άλματα μόνο (Alexandersson, 1986), ακολουθούμενο από τη μέθοδο MLR (Multiple Linear Regression, Vincent, 1998). Αρκετές φορές έχουν συμπεριληφθεί στην αξιολόγηση και η μέθοδος του Potter (1981) και διάφορες εκδοχές Bayesian tests (π.χ. Perreault *et al.*, 1999, 2000, Ducré-Robitaille et al., 2003).

Στις περισσότερες περιπτώσεις, οι μέθοδοι εφαρμόζονται για επίπεδο εμπιστοσύνης 95%, το οποίο αντιστοιχεί σε πιθανότητα εντοπισμού άλματος 0.95. Με βάση το επίπεδο εμπιστοσύνης αυτό, οι περισσότεροι μελετητές συμπεραίνουν ότι τίθεται ένα πλαίσιο και για τον αποδεκτό ποσοστό ψευδών εντοπισμών περίπου στο 5%, άποψη που φαίνεται λογική.

Για τον έλεγχο της απόδοσης των μεθόδων ομογενοποίησης υπάρχουν πολλά διαφορετικά κριτήρια. Τα κριτήρια αυτά επιλέγονται κάθε φορά ανάλογα με το στόχο του χρήστη της μεθόδου, π.χ. αν τον ενδιαφέρει περισσότερο ο εντοπισμός των περισσότερων πραγματικών αλμάτων, η επαναφορά της αρχικής γραμμικής τάσης των χρονοσειρών ή η μείωση των λανθασμένων εντοπισμών. Συνήθως από τους ερευνητές επιλέγεται ένας συνδυασμός των ανομογενειών που εντοπίζουν οι μέθοδοι και των ψευδών εντοπισμών. Αρκετοί ερευνητές δεν εξετάζουν κάποιο κριτήριο που σχετίζεται με τους εντοπισμούς πλασματικών ανομογενειών.

Καλύτερη απόδοση σε αξιολογήσεις με λίγες ανομογένειες στα δεδομένα φαίνεται να έχει η μέθοδος SNHT για άλματα μόνο, καθώς είναι πάντα ανάμεσα στις μεθόδους με τα καλύτερα αποτελέσματα. Η μέθοδος αυτή έχει μάλιστα πολύ καλύτερη συμπεριφορά από τη μεταγενέστερη εκδοχή της που αναζητά και διαφορές τάσεων. Στις πιο σύγχρονες αξιολογήσεις (Domonkos, 2011a, Venema *et al.*, 2012) με πιο πολύπλοκες βάσεις δεδομένων με πολλές και διαφορετικές ανομογένειες, καλύτερη απόδοση φαίνεται να έχουν οι μέθοδοι MASH, Craddock, PRODIGE, ACMANT και η μέθοδος των Caussinus and Mestre (2004).

Από την επισκόπηση των δημοσιεύσεων των Παραγράφων 5.4.1 και 5.4.2 προκύπτει ότι σε όρους ψευδών εντοπισμών, κάποιες βασικές μέθοδοι που εξετάζονται από αρκετούς ερευνητές, όπως το SNHT για άλματα και η μέθοδος της Vincent (1998), όταν εφαρμόζονται σε δεδομένα από απλό μοντέλο (π.χ. AR(1)) και με λίγες ανομογένειες με απλή μορφή (π.χ. μικρός αριθμός αλμάτων), παρουσιάζουν γενικά καλά αποτελέσματα (ο δείκτης FAR είναι κοντά στο 5%). Εξαίρεση αποτελεί η μελέτη των Easterling and Peterson (1995), με πολύ μεγάλα ποσοστά ψευδών εντοπισμών και σε κάποιες περιπτώσεις η μελέτη των Menne and Williams (2005).

Σε πιο σύνθετες περιπτώσεις, δηλαδή για χρήση ανομογενών χρονοσειρών αναφοράς ή/και για χρήση απλών μοντέλων με πολλές και σύνθετες ανομογένειες, το ποσοστό ψευδών εντοπισμών αυξάνεται κατά πολύ. Για παράδειγμα, στη μελέτη του Domonkos (2011a), μεγαλύτερο ποσοστό ψευδών ανομογενειών εντοπίζεται στις βάσεις E και F με σύνθετες ανομογένειες, που θεωρούνται από το συγγραφέα και οι πιο ρεαλιστικές. Βασική εξαίρεση αποτελεί η πιο πρόσφατη αξιολόγηση των μεθόδων ομογενοποίησης από τους Venema *et al.* (2012), στην οποία με εξαίρεση τη μέθοδο MASH, όλες οι άλλες μέθοδοι είχαν ποσοστά ψευδών εντοπισμών μικρότερα από 5%.

Οι περισσότερες από τις μεθόδους ομογενοποίησης που αξιολογήθηκαν αναπτύχθηκαν για τον εντοπισμό ενός μοναδικού άλματος στις χρονοσειρές, π.χ. SNHT. Λίγες μόνο αναζητούν ταυτόχρονα πολλά διαφορετικά άλματα, π.χ. MASH. Οι μέθοδοι για ένα μόνο άλμα εφαρμόζονται συνήθως με χρήση:

- αλγορίθμου κοπής (cutting algorithm), ο οποίος κόβει τη χρονοσειρά σε τμήματα κάθε φορά που εντοπίζει ένα άλμα και επανεξετάζει τα τμήματα για επιπλέον άλματα ή
- ημι-ιεραρχικού αλγόριθμου (semihierarchic algorithm), ο οποίος λειτουργεί όπως και ο αλγόριθμος κοπής στο σπάσιμο των χρονοσειρών, αλλά κάθε φορά που ένα τμήμα χωρίζεται στα δύο, τα άλματα που εντοπίστηκαν στο προηγούμενο «σπάσιμο» επαναξιολογούνται για τη στατιστική σημαντικότητά τους.

Οι περισσότεροι συγγραφείς που κάνουν χρήση τέτοιων αλγορίθμων θεωρούν ότι ο ημι-ιεραρχικός αλγόριθμος είναι πιο αποτελεσματικός, κάτι που όμως σε αρκετές περιπτώσεις δεν επαληθεύτηκε (π.χ. κατά την εφαρμογή δύο εκδοχών του SNHT για άλματα μόνο από τον Domonkos (2011a).

Τέλος, οι περισσότεροι ερευνητές λαμβάνουν ελάχιστη απόσταση ανάμεσα στις ανομογένειες και στις ανομογένειες και την άκρη της χρονοσειράς 5 έτη. Διαφορετική απόσταση 10 έτη προτείνεται από τους Beaulieu *et al.* (2008). Γενικά όπως φαίνεται και από τη μελέτη της Wang (2008), η παράμετρος αυτή είναι πολύ σημαντική αφού υπάρχουν μέθοδοι, όπως το SNHT που εντοπίζουν πλασματικές ανομογένειες στα άκρα των χρονοσειρών πάνω από το ανεκτό όριο 5%.

5.6. Διερεύνηση αυτοσυσχέτισης διαφοράς χρονοσειρών ελεγχόμενης - αναφοράς

Από τη σύγκριση των προηγούμενων προσεγγίσεων αξιολόγησης της ομογενοποίησης με συνθετικά δεδομένα, προέκυψε η πολύ σημαντική διαπίστωση ότι μια βασική υπόθεση που γίνεται από τους χρήστες μεθόδων ομογενοποίησης μέχρι σήμερα είναι ότι η διαφορά δύο γειτονικών χρονοσειρών αναφοράς - ελεγχόμενης χαρακτηρίζεται από λευκό θόρυβο, δεν παρουσιάζει δηλαδή αυτοσυσχέτιση.

Η ισχύς αυτής της υπόθεσης μπορεί να διερευνηθεί με ένα απλό παράδειγμα:

Θεωρούμε δύο χρονοσειρές X_t και Y_t που εκφράζουν δύο γειτονικούς σταθμούς μέτρησης θερμοκρασίας με μέση τιμή μηδέν και μοναδιαία τυπική απόκλιση και μία χρονοσειρά Z_t η οποία προκύπτει από τη Σχέση:

$$Z_t = X_t - Y_t \qquad (5.14)$$

Μπορούμε να θεωρήσουμε χωρίς βλάβη της γενικότητας ότι η Y_t συσχετίζεται με τη X_t με τη Σχέση:

$$Y_t = \alpha X_t + \sqrt{1-\alpha^2} V_t \qquad (5.15)$$

όπου V_t ακολουθία λευκού θορύβου (βλ. Παράγραφο 4.2.1) και α μη αρνητικός συντελεστής.

Έστω ότι οι δύο χρονοσειρές X_t, Y_t έχουν ίδια δομή αυτοσυσχέτισης, δηλαδή

$$Cov[X_t, X_{t-\tau}] = Cov[Y_t, Y_{t-\tau}] = \gamma_\tau \qquad (5.16)$$

όπου τ η χρονική υστέρηση.

Επίσης είναι: $Var[X_t] = Var[Y_t] = \gamma_0 = 1$ (5.17)

οπότε για το συντελεστή αυτοσυσχέτισης ρ των χρονοσειρών X_t, Y_t ισχύει:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \gamma_\tau \qquad (5.18)$$

Λαμβάνοντας αναμενόμενες τιμές προκύπτει για τις δύο χρονοσειρές:

$$E[X_t Y_{t-\tau}] = E[\alpha X_t X_{t-\tau}] + E[X_t \sqrt{1-\alpha^2} V_{t-\tau}] = a\gamma_\tau \qquad (5.19)$$

Λαμβάνοντας αναμενόμενες τιμές για τη χρονοσειρά Z_t έχουμε:

$$E[Z_t Z_{t-\tau}] = Cov[Z_t, Z_{t-\tau}] = E[(X_t - Y_t)(X_{t-\tau} - Y_{t-\tau})] =$$

$$= E[X_t X_{t-\tau}] + E[Y_t Y_{t-\tau}] - E[X_t Y_{t-\tau}] - E[X_{t-\tau} Y_t] =$$

$$= 2\gamma_\tau - 2\alpha\gamma_\tau = 2(1-\alpha)\gamma_\tau \qquad (5.20)$$

Από τη Σχέση 5.20, προκύπτει για μηδενική υστέρηση η (σταθερή) μεταβλητότητα της χρονοσειράς Z_t:

$$Var[Z_t] = 2(1-\alpha)\gamma_0 \qquad (5.21)$$

οπότε ο συντελεστής αυτοσυσχέτισης r της χρονοσειράς Z_t για υστέρηση τ είναι:

$$r_\tau = \frac{Cov[Z_t, Z_{t-\tau}]}{\sqrt{Var[Z_t]Var[Z_{t-\tau}]}} = \frac{2(1-\alpha)\gamma_\tau}{2(1-\alpha)\gamma_0} = \frac{\gamma_\tau}{\gamma_0} = \rho_\tau \qquad (5.22)$$

Παρατηρούμε δηλαδή ότι η χρονοσειρά Z_t όχι μόνο παρουσιάζει αυτοσυσχέτιση, δηλαδή δεν χαρακτηρίζεται από λευκό θόρυβο, όπως έχει διατυπωθεί από τους Conrad and Pollak (1962), αλλά έχει και την ίδια δομή αυτοσυσχέτισης με τις χρονοσειρές X_t και Y_t.

Μηδενισμό της αυτοσυσχέτισης έχουμε μόνο αν $\rho_\tau=0$, δηλαδή για μηδενική αυτοσυσχέτιση των χρονοσειρών X_t και Y_t, ή ισοδύναμα μόνο για την περίπτωση που οι χρονοσειρές X_t και Y_t χαρακτηρίζονται από λευκό θόρυβο.

6. ΜΕΘΟΔΟΛΟΓΙΑ

Από την παρουσίαση της βιβλιογραφίας στα προηγούμενα Κεφάλαια προκύπτει ότι:

Α) Οι μέθοδοι ομογενοποίησης εφαρμόζονται συστηματικά σε υδροκλιματικά δεδομένα. Πολύ συχνή είναι η εφαρμογή τους σε ετήσια και μηνιαία δεδομένα θερμοκρασίας και βροχόπτωσης.

Β) Η αξιολόγηση της ομογενοποίησης γίνεται με δύο κυρίως τρόπους: εφαρμογή σε πραγματικά και σε συνθετικά δεδομένα.

Γ) Οι προηγούμενες μελέτες αξιολόγησης της ομογενοποίησης με χρήση συνθετικών δεδομένων κάνουν χρήση συνθετικών δεδομένων χωρίς αυτοσυσχέτιση (λευκός θόρυβος) ή με βραχεία μνήμη (π.χ. μοντέλο AR(1)). Μόνη εξαίρεση αποτελεί η δημοσίευση των Rust *et al*. (2008), στην οποία όμως διερευνάται μόνο μια συγκεκριμένα περίπτωση συντελεστή Hurst H=0.62.

Από τις παραπάνω παρατηρήσεις προκύπτει ένα βασικό κενό όσον αφορά τη μεθοδολογία που έχει χρησιμοποιηθεί μέχρι σήμερα για την αξιολόγηση των μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων: δεν έχει γίνει μέχρι στιγμής μια ολοκληρωμένη μελέτη των αποτελεσμάτων της ομογενοποίησης σε συνθετικά δεδομένα με **μακροπρόθεσμη εμμονή**.

Η ανάγκη ελέγχου δεδομένων με μακροπρόθεσμη εμμονή επισημαίνεται μάλιστα και στην τελευταία προσπάθεια που έχει γίνει για αξιολόγηση των μεθόδων ομογενοποίησης από τους Venema *et al*. (2012). Προκύπτει επομένως μια ουσιαστική ανάγκη την οποία η παρούσα Μεταπτυχιακή εργασία φιλοδοξεί να καλύψει σε όσο το δυνατόν μεγαλύτερο βαθμό.

Η όλη προσέγγιση που ακολουθήθηκε βασίστηκε στην εφαρμογή επιλεγμένων μεθόδων ομογενοποίησης σε συνθετικές χρονοσειρές που χαρακτηρίζονται από εμμονή και εξετάστηκε σε τι συχνότητα και υπό ποιές προϋποθέσεις εντοπίζονται σε αυτές ψευδείς ανομογένειες.

Πιο συγκεκριμένα, η όλη μεθοδολογική προσέγγιση:

Α) Βασίστηκε στην παραγωγή ομογενών συνθετικών χρονοσειρών. Ως εκ τούτου το κριτήριο με το οποίο αξιολογήθηκαν οι μέθοδοι ομογενοποίησης είναι το ποσοστό των περιπτώσεων που εντοπίζονται ψευδείς ανομογένειες (false alarm rate).

Β) Εστιάστηκε σε ετήσια δεδομένα θερμοκρασιών και βροχοπτώσεων. Η ομογενοποίηση μηνιαίων δεδομένων, η οποία είναι και νεότερη στη βιβλιογραφία, αποτελεί ένα πιθανό μελλοντικό αντικείμενο έρευνας.

Γ) Εστιάστηκε σε δύο επιλεγμένες μεθόδους ομογενοποίησης, μία για τα θερμοκρασιακά δεδομένα και μία για τα δεδομένα βροχόπτωσης. Οι μέθοδοι αυτές είναι σχετικές – συγκριτικές, χρησιμοποιούν δηλαδή χρονοσειρές αναφοράς για τον εντοπισμό των ανομογενειών.

Δ) Βασίστηκε σε στοχαστικό μοντέλο προσομοίωσης (Monte Carlo). Η μέθοδος αυτή χρησιμοποιήθηκε και σε όλες τις περιπτώσεις αξιολόγησης μεθόδων ομογενοποίησης με συνθετικά δεδομένα που αναφέρθηκαν στο Κεφάλαιο 5. Η προσομοίωση έγινε με χρήση γλώσσας προγραμματισμού Matlab.

Ε) Έκανε χρήση συνθετικών χρονοσειρών με συγκεκριμένα στατιστικά χαρακτηριστικά (μήκος, κατανομή, μέση τιμή, τυπική απόκλιση, συντελεστής Hurst, συντελεστής ετεροσυσχέτισης, πληρότητα). Λεπτομέρειες σε σχέση με την επιλογή των χαρακτηριστικών αυτών δίνονται στην Παράγραφο 6.3.

ΣΤ) Έκανε χρήση ήδη ανεπτυγμένων μοντέλων για την παραγωγή χρονοσειρών με συγκεκριμένο συντελεστή Hurst και την εκτίμηση του συντελεστή αυτού. Περισσότερες λεπτομέρειες δίνονται στις Παραγράφους 6.3.1 και 6.4.

Ζ) Εξέτασε σε ένα βαθμό την επιρροή της ομογενοποίησης στην κλίση και το συντελεστή Hurst των αρχικών χρονοσειρών.

Το θεωρητικό και μεθοδολογικό υπόβαθρο της διερεύνησης (επεξήγηση ορολογίας, επιλογή μεθόδων και χαρακτηριστικών των δεδομένων) αναλύεται στις Παραγράφους 6.1-6.4. Στις Παραγράφους 6.5 και 6.6 περιγράφεται η εφαρμογή των μεθόδων ομογενοποίησης στα υδροκλιματικά δεδομένα.

6.1. Επιλογή μεθόδων ομογενοποίησης

Οι μέθοδοι που επιλέχτηκε να εφαρμοστούν στα συνθετικά δεδομένα θερμοκρασίας και βροχόπτωσης προέκυψαν από τη βιβλιογραφική επισκόπηση που περιγράφεται στα Κεφάλαια 2-5. Όπως αναφέρθηκε και στη εισαγωγή του Κεφαλαίου 6, επιλέχτηκαν δύο μέθοδοι ομογενοποίησης:

1. η μέθοδος **SNHT** για τα θερμοκρασιακά δεδομένα, και
2. η **Διπλή αθροιστική καμπύλη** για τα δεδομένα βροχόπτωσης.

Το SNHT είναι μία από τις κλασικές μεθόδους ομογενοποίησης δεδομένων θερμοκρασίας και βροχόπτωσης. Έχει χρησιμοποιηθεί ευρέως παγκοσμίως (π.χ. Sahin and Cigizoglu, 2010), παραλλαγές της μεθόδου χρησιμοποιούνται για εντοπισμό πολλαπλών αλμάτων (Domonkos 2011a) και έχει εφαρμοστεί πολλές φορές σε συνθετικά δεδομένα σε διάφορες μελέτες αξιολόγησης των μεθόδων ομογενοποίησης (βλ. Παράγραφο 5.5). Επιπλέον νεότερες μέθοδοι όπως το PHA βασίζονται στο SNHT και το χρησιμοποιούν για τον εντοπισμό αλμάτων (Menne and Williams, 2009). Τα αποτελέσματα επομένως της εφαρμογής της μεθόδου αυτής σε δεδομένα με μακροπρόθεσμη εμμονή παρουσιάζουν μεγάλο ενδιαφέρον και μπορούν να συγκριθούν με τις προηγούμενες προσεγγίσεις.

Η Διπλή αθροιστική καμπύλη είναι μια μέθοδος που επίσης έχει χρησιμοποιηθεί ευρέως στην ομογενοποίηση υδροκλιματικών δεδομένων. Τα τελευταία χρόνια δεν προτιμάται από τους κλιματολόγους, κυρίως επειδή στη βασική της μορφή είναι υποκειμενική (Easterling and Peterson, 1995). Παρόλα αυτά είναι η βασική μέθοδος ομογενοποίησης που αναφέρεται σε υδρολογικά βιβλία (π.χ. Dingman, 1994). Επιλέχτηκε να εφαρμοστεί σε πρώτη φάση στα δεδομένα βροχόπτωσης, ενώ σαν μελλοντική προέκταση της παρούσας Μεταπτυχιακής εργασίας θα μπορούσε να εφαρμοστεί και μια πιο πρόσφατα ανεπτυγμένη μέθοδος.

6.2. Γενικοί περιορισμοί και προϋποθέσεις των μεθόδων

Η μηδενική υπόθεση της μεθόδου SNHT, αν και χρησιμοποιείται για την ομογενοποίηση δεδομένων με αυτοσυσχέτιση, είναι ότι τα δεδομένα στα οποία εφαρμόζεται είναι ανεξάρτητα με κανονική κατανομή (βλ. Σχέση 3.7).

Η Διπλή αθροιστική καμπύλη θεωρεί ότι τα συσσωρευτικά αθροίσματα δεδομένων γειτονικών σταθμών (εφόσον δεν υπάρχει π.χ. κάποια σημαντική αλλαγή υψομέτρου) είναι μεταξύ τους ανάλογα. Δεν γίνεται κάποια άλλη υπόθεση για τα στατιστικά χαρακτηριστικά των δεδομένων (Searcy and Hardison, 1960).

6.3. Στατιστικά χαρακτηριστικά και παράμετροι των συνθετικών χρονοσειρών

Για να αξιολογηθεί η συμπεριφορά των δύο μεθόδων ομογενοποίησης που επιλέχτηκαν, έγινε μια πολυπαραμετρική προσέγγιση της ομογενοποίησης. Επιλέχτηκαν κάποιες σταθερές και κάποιες μεταβαλλόμενες παράμετροι.

Όλα τα συνθετικά δεδομένα που χρησιμοποιήθηκαν προέκυψαν καταρχάς από στάσιμο στοχαστικό μοντέλο, παράχθηκαν δηλαδή ανελίξεις στις οποίες δεν υπάρχει μεταβολή των στατιστικών χαρακτηριστικών με την πάροδο του χρόνου.

Η κατανομή με την οποία επιλέχτηκε να γίνει παραγωγή των συνθετικών δεδομένων, προέκυψε με βάση την ετήσια χρονική κλίμακα των δεδομένων αυτών. Στις μέσες ετήσιες θερμοκρασίες προσαρμόζεται ικανοποιητικά η κανονική κατανομή. Οι ετήσιες βροχοπτώσεις, αντίθετα παρουσιάζουν συχνά μια ασυμμετρία, κυρίως σε ξηρά κλίματα. Σε υγρότερα κλίματα έχουμε πιο ικανοποιητική προσαρμογή της κανονικής κατανομής στα ετήσια δεδομένα. Αντίθετα, σε δεδομένα βροχόπτωσης μικρότερης κλίμακας, π.χ. μηνιαίας, ωριαίας κλπ η κατανομή της βροχόπτωσης γίνεται όλο και πιο ασύμμετρη (Koutsoyiannis, 2005).

Λαμβάνοντας υπόψη τα παραπάνω, αποφασίστηκε να παραχθούν δεδομένα από κανονική κατανομή και στις δύο περιπτώσεις, των θερμοκρασιών και των βροχοπτώσεων. Στα δεδομένα βροχόπτωσης επιλέχτηκε μεγάλη μέση τιμή (>900mm) και τυπική απόκλιση, ώστε να προσομοιάζονται καλύτερα δεδομένα από περιοχή με υγρό κλίμα.

Στην επιλογή μέσης τιμής και τυπικής απόκλισης για τα δεδομένα θερμοκρασίας, κύριο ρόλο έπαιξε η μέθοδος ομογενοποίησης SNHT που επιλέχτηκε. Στο SNHT γίνεται κανονικοποίηση των δεδομένων, οπότε η μέση τιμή και η τυπική απόκλιση των χρονοσειρών δεν θα επηρέαζαν το αποτέλεσμα της ομογενοποίησης. Για το λόγο αυτό οι χρονοσειρές θερμοκρασίας παράχθηκαν απευθείας με μέση τιμή μηδέν και μοναδιαία τυπική απόκλιση.

Οι τιμές που επιλέχτηκαν για τη μέση τιμή και την τυπική απόκλιση θερμοκρασιών και βροχοπτώσεων φαίνονται στον Πίνακα 6.1.

Πίνακας 6.1. *Μέση τιμή και τυπική απόκλιση των συνθετικών δεδομένων θερμοκρασιών και βροχοπτώσεων*

Στατιστικά χαρακτηριστικά	Μέση τιμή	Τυπική απόκλιση
Μέση ετήσια θερμοκρασία	0 (^oC)	1 (^oC)
Ολική ετήσια βροχόπτωση	1000 (mm)	300 (mm)

Οι μεταβαλλόμενες παράμετροι των συνθετικών δεδομένων που επιλέχτηκαν μπορούν να χωριστούν σε δύο κατηγορίες: βασικές και δευτερεύουσες.

Βασικές παράμετροι:

1. Συντελεστής Hurst,
2. Συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες, και
3. Αριθμός και τρόπος χρήσης των χρονοσειρών αναφοράς.

Δευτερεύουσες παράμετροι:

1. Μήκος των συνθετικών χρονοσειρών, και
2. Ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς.

Η διερεύνηση έγινε σε ενότητες με βάση διαφορετικούς συνδυασμούς των παραπάνω παραμέτρων. Στις ενότητες αυτές οι δευτερεύουσες παράμετροι παρέμειναν σταθερές εκτός από κάποιες περιπτώσεις στις οποίες θα αναφερθεί και θα δικαιολογηθεί μια διαφορετική επιλογή.

Παρουσίαση και δικαιολόγηση του συνόλου τιμών των βασικών και δευτερευουσών παραμέτρων γίνεται στις Παραγράφους 5.1.4.1 έως 5.1.4.5.

6.3.1. Συντελεστής Hurst

Επιλέξαμε να εξετάσουμε τιμές που να καλύπτουν επαρκώς το εύρος [0.5, 1), το οποίο έχει φυσική σημασία, όπως αναφέρθηκε στην Παράγραφο 4.2.1. Οι τιμές που επιλέχτηκαν είναι όλες οι τιμές από H=0.5 μέχρι και H=0.9 με βήμα 0.05, συνολικά δηλαδή 9 διαφορετικές τιμές.

Δηλαδή: H={0.50, 0.55, ..., 0.90}.

6.3.2. Συντελεστής ετεροσυσχέτισης

Τα συνθετικά δεδομένα που παράχθηκαν για τον έλεγχο της ομογενοποίησης ήταν ομογενή. Επομένως δεν αντιμετωπίστηκε πρόβλημα λανθασμένης εκτίμησης του συντελεστή ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και την εξεταζόμενη λόγω της παρουσίας ανομογενειών.

Σε κάποιες περιπτώσεις ομογενοποίησης, π.χ. εφαρμογής του SNHT σε πραγματικά δεδομένα από τους Moberg and Alexandersson (1997), λόγω της ύπαρξης ανομογενειών στα δεδομένα, έγινε υπολογισμός του συντελεστή ετεροσυσχέτισης από τις χρονοσειρές του ρυθμού μεταβολής της θερμοκρασίας $\frac{dT}{dt}$, ώστε να μειωθεί η ευαισθησία του υπολογιζόμενου συντελεστή ετεροσυσχέτισης σε τυχόν ανομογένειες.

Στην περίπτωσή μας όμως, λόγω της παραγωγής ομογενών χρονοσειρών, ο συντελεστής ετεροσυσχέτισης υπολογίστηκε από τις αρχικές - ανεπεξέργαστες (raw) χρονοσειρές αναφοράς (X) και ελεγχόμενες (Y) σύμφωνα με τον ορισμό του (Σχέση 3.1).

$$\circ_{XY} := \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} = \frac{\pm_{XY}}{\pm_X \pm_Y} \qquad (-1 \leq \circ_{XY} \leq 1) \qquad (6.1)$$

Για να είναι όσο το δυνατόν πιο πλήρης η διερεύνηση, επιλέχτηκε ένα μεγάλο φάσμα συντελεστών ετεροσυσχέτισης ανάμεσα στις ελεγχόμενες χρονοσειρές και τις χρονοσειρές αναφοράς. Η μικρότερη τιμή επιλέχτηκε να είναι 0.5, σύμφωνα με τη μεθοδολογία των Easterling and Peterson (1995) και Beaulieu *et al*. (2008).

Οι υπόλοιπες τιμές του συντελεστή ετεροσυσχέτισης λήφθηκαν με βήμα 0.1, μέχρι την τιμή 0.9. Για να διερευνηθεί και η περίπτωση πολύ υψηλής συσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες αποφασίστηκε να εξεταστεί συμπληρωματικά και η τιμή 0.95. Συνολικά δηλαδή, στην παρούσα Μεταπτυχιακή εργασία, ο συντελεστής ετεροσυσχέτισης μπορούσε να πάρει τιμές από ένα σύνολο 6 τιμών:

ρ_{XY}={0.5, 0.6, 0.7, 0.8, 0.9, 0.95}.

6.3.3 Σύστημα χρονοσειρών αναφοράς

Όπως αναλύθηκε στο Κεφάλαιο 3, στις σχετικές μεθόδους ομογενοποίησης γίνεται έλεγχος μιας χρονοσειράς σε σχέση με μία, δύο, τρεις ή περισσότερες χρονοσειρές αναφοράς ή και τη σύνθεση ενός αριθμού χρονοσειρών αναφοράς.

Με τον όρο σύστημα χρονοσειρών αναφοράς θα θεωρείται στα επόμενα Κεφάλαια ο αριθμός των χρονοσειρών αναφοράς τις οποίες συγκρίνουμε με την ελεγχόμενη, ο τρόπος με τον οποίο γίνεται η σύγκριση (κατά ζεύγη ή με σύνθεση μιας χρονοσειράς αναφοράς) και ο ελάχιστος αριθμός χρονοσειρών στις οποίες εντοπίζεται κοινό άλμα ώστε το άλμα αυτό να θεωρηθεί «πραγματικό».

Έγινε προσπάθεια να καλυφτεί ένα μεγάλο φάσμα αυτών των περιπτώσεων-συστημάτων ώστε να φανεί με ποιο τρόπο η επιλογή και χρήση των χρονοσειρών αναφοράς επηρεάζει τα αποτελέσματα της ομογενοποίησης, διογκώνοντας ή αντίθετα μικραίνοντας πιθανά προβλήματα των μεθόδων.

Συνολικά αποφασίστηκε να εξεταστούν 9 συστήματα χρονοσειρών αναφοράς (Πίνακας 6.2).

Πίνακας 6.2. *Περιγραφή και συμβολισμός συστημάτων χρονοσειρών αναφοράς που εξετάστηκαν*

Συμβολισμός	**Περιγραφή συστήματος**
1/1	χρήση 1 χρονοσειράς αναφοράς και εντοπισμός αλμάτων
3/7	χρήση 7 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων σε 3 χρονοσειρές
3/1	παραγωγή 1 «τεχνητής» χρονοσειράς αναφοράς ως σύνθεση 3 χρονοσειρών αναφοράς και εντοπισμός αλμάτων
2/2	χρήση 2 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων και στις 2 χρονοσειρές
3/3	χρήση 3 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων και στις 3 χρονοσειρές
3/10	χρήση 10 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων σε 3 χρονοσειρές
3/20	χρήση 20 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων σε 3 χρονοσειρές
3/7-50 years*	χρήση 7 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων σε 3 χρονοσειρές, με μήκος όμως χρονοσειρών 50 έτη αντί για 100
3/7–j10*	χρήση 7 χρονοσειρών αναφοράς και εντοπισμός κοινών αλμάτων σε 3 χρονοσειρές, με ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς 10 έτη αντί για 5

**Αν και ουσιαστικά πρόκειται για το σύστημα 3/7 με αλλαγή δευτερευουσών παραμέτρων, θεωρείται ξεχωριστό σύστημα για λόγους ευκολίας της επισκόπησης.*

Κατά την αξιολόγηση των μεθόδων δεν εξετάστηκαν όλα τα συστήματα και στις δύο μεθόδους ομογενοποίησης. Αντίθετα, έγινε επιλογή συστημάτων για κάθε μέθοδο ανάλογα με τις συνήθεις τεχνικές που προκύπτουν από τη βιβλιογραφία.

Οι λόγοι που επιλέχτηκαν να εξεταστούν τα συγκεκριμένα συστήματα είναι ποικίλοι:

Το σύστημα 1/1 επιλέχτηκε γιατί αποτελεί τη θεωρητική βάση της εφαρμογής των υπολοίπων περιπτώσεων. Οι μέθοδοι ομογενοποίησης θεωρούν συνήθως κάποιο ζεύγος χρονοσειρών αναφοράς-ελεγχόμενης, το οποίο ελέγχουν για ανομογένειες. Ο αριθμός των ζευγών (ανάλογα με τον αριθμό των χρονοσειρών αναφοράς) τα οποία ελέγχουν για κοινές ανομογένειες και ο τρόπος που προκύπτουν οι χρονοσειρές αναφοράς δεν σχετίζεται συνήθως με το θεωρητικό τμήμα της μεθόδου αλλά με την πρακτική εφαρμογή της.

Το σύστημα 3/7 αντλεί τη σημασία του από τη βιβλιογραφία. Κατά την επιλογή των σταθμών αναφοράς στη μέθοδο PHA (Menne and Williams, 2009), μία από τις σημαντικότερες μεθόδους ομογενοποίησης σήμερα, η οποία κάνει και χρήση του SNHT, προκύπτει μια λίστα με τουλάχιστον 7 σταθμούς αναφοράς για κάθε χρονική περίοδο της εξεταζόμενης χρονοσειράς. Πραγματική ανομογένεια θεωρείται εκείνη που εντοπίζεται από τη μέθοδο σε τουλάχιστον 3 χρονοσειρές αναφοράς (βλ. Παράγραφο 3.2.4). Στο στάδιο 4 θα διερευνηθούν διαφορετικές 4 εκδοχές και/ή προεκτάσεις της περίπτωσης 3/7.

Το τρίτο σύστημα είναι η κατασκευή δηλαδή μιας χρονοσειράς αναφοράς ως σύνθεση 3 χρονοσειρών αναφοράς (3/1). Η κατασκευή μιας χρονοσειράς αναφοράς από τις γειτονικές σειρές της ελεγχόμενης (ως μέση τιμή ή με κάποιο συντελεστή βαρύτητας) είναι μια συνήθης πρακτική στην ομογενοποίηση χρονοσειρών, καθώς επιτρέπει την ελαχιστοποίηση της επιρροής των ανομογενειών στις χρονοσειρές αναφοράς (Searcy and Hardison, 1960).

Γενικά η επιλογή τουλάχιστον 3 χρονοσειρών αναφοράς για τον εντοπισμό και τη διόρθωση ανομογενειών (κοινό στοιχείο των 3/7 και 3/1) είναι συνήθης στη βιβλιογραφία (Ducré-Robitaille *et al*., 2003,Vincent, 1998, Beaulieu *et al*., 2008). Για να διερευνηθεί περαιτέρω κατά πόσο είναι λογική η επιλογή αυτή, επιλέχτηκε να εξεταστούν τα συστήματα 3/10 και 3/20.

Διερεύνηση στην ίδια κατεύθυνση με εξέταση όμως και μιας επιπλέον παραμέτρου (μήκος χρονοσειρών, ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς) γίνεται και με τα συστήματα χρονοσειρών αναφοράς 3/7-50 years και 3/7–j10. Περισσότερες πληροφορίες για τις παραμέτρους αυτές δίνονται στις Παραγράφους 6.3.4 και 6.3.5.

6.3.4. Μήκος συνθετικών χρονοσειρών

Το μήκος των συνθετικών χρονοσειρών περιλαμβάνεται στις δευτερεύουσες παραμέτρους, και παρέμεινε σταθερό και ίσο με 100 έτη σχεδόν σε όλους τους συνδυασμούς παραγόντων που εξετάστηκαν. Η τιμή αυτή του μήκους είναι μια συνήθης επιλογή στη βιβλιογραφία (π.χ. Caussinus and Mestre, 2004, Easterling and Peterson, 1995, Moberg and Alexandersson, 1997 κ.ά.).

Εκτός από την τιμή 100 έτη αποφασίστηκε να εξεταστεί και μια εναλλακτική τιμή 50 έτη, για να να διερευνηθεί αν το μήκος των χρονοσειρών επηρεάζει τη συμπεριφορά των μεθόδων ομογενοποίησης.

6.3.5. Ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς

Ένα πρόβλημα στατιστικής σημαντικότητας των αποτελεσμάτων αποτελεί η ελάχιστη απόσταση μεταξύ δύο ανιχνευόμενων ανομογενειών ή η απόσταση ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς, γιατί προφανώς δεν έχει νόημα να ανιχνευτούν π.χ. δύο ανομογένειες σε τρία χρόνια ή μια ανομογένεια στο 99^{ο} έτος μιας χρονοσειράς με 100 έτη.

Μια συνήθης πρακτική στη βιβλιογραφία είναι η επιλογή του εμπειρικού ορίου των 5 ετών ως συνθήκη για να θεωρηθεί μια ανομογένεια στατιστικά σημαντική. Το όριο αυτό λαμβάνεται συνήθως εμπειρικά (π.χ. Easterling and Peterson, 1995), αν και έχουν γίνει κάποιες μελέτες για την ευαισθησία μεθόδων ομογενοποίησης στα άκρα των χρονοσειρών, όπως η μελέτη της Wang (2008) σε μηνιαία δεδομένα, η οποία αναλύεται στο Κεφάλαιο 5. Η τιμή αυτού του ορίου πάντως είναι σημαντικό ζήτημα, καθώς πολλές από τις ευρέως χρησιμοποιούμενες μεθόδους ομογενοποίησης είναι γνωστό ότι εντοπίζουν μεγαλύτερο από το ανεκτό ποσοστό ψευδών ανομογενειών στα άκρα τους. Δεν θα έπρεπε επομένως να βασίζεται κυρίως σε εμπειρικές τιμές.

Η ελάχιστη απόσταση ανάμεσα σε δύο ανιχνευόμενες ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς σ΄αυτή τη μελέτη περιλαμβάνεται στις δευτερεύουσες παραμέτρους, επομένως παρέμεινε σταθερή σχεδόν σε όλους τους συνδυασμούς παραγόντων που εξετάστηκαν. Η προεπιλεγμένη τιμή της ήταν 5 έτη, σύμφωνα και με προηγούμενες πρακτικές που λήφθηκαν ως κατευθυντήριες (π.χ. Dingman, 1994, Domonkos, 2011a).

Εναλλακτική τιμή που εξετάστηκε κατά την εφαρμογή των μεθόδων ήταν 10 έτη, τιμή πιο λογική σύμφωνα με τους Beaulieu *et al*. (2008).

6.4. Παραγωγή συνθετικών χρονοσειρών

Οι συνθετικές χρονοσειρές παράχθηκαν με χρήση κατάλληλου αλγορίθμου ο οποίος αναπαράγει την επιθυμητή μέση τιμή, την τυπική απόκλιση και το συντελεστή Hurst. Ο αριθμός των παραγόμενων χρονοσειρών κάθε φορά ήταν m+1, όπου m ο αριθμός των χρονοσειρών αναφοράς (άρα m+1 το άθροισμα χρονοσειρών αναφοράς και ελεγχόμενης).

Οι m+1 χρονοσειρές που παράγονταν ήταν αρχικά ασυσχέτιστες μεταξύ τους, αφού είχαν παραχθεί αυτόνομα. Για να προσομειωθεί όμως το δίκτυο χρονοσειρών αναφοράς-ελεγχόμενης, στη συνέχεια συσχετίζονταν μεταξύ τους (με επιλεγμένο κάθε φορά συντελεστή συσχέτισης).

Σε αρκετές περιπτώσεις γινόταν μετά την παραγωγή των χρονοσειρών αναφοράς εκτίμηση του συντελεστή Hurst κάθε χρονοσειράς. Η διαδικασία αυτή περιέχει μια αβεβαιότητα και σε πολλές περιπτώσεις προκύπτει συντελεστής πολύ διαφορετικός από αυτόν βάσει του οποίου παράχθηκε η χρονοσειρά. Για το λόγο αυτό σε κάποιες περιπτώσεις συνθετικών χρονοσειρών προστέθηκε μια ρήτρα όσον αφορά τις εκτιμούμενες τιμές του συντελεστή Hurst.

Ο αλγόριθμος παραγωγής χρονοσειρών με μακροπρόθεσμη εμμονή που χρησιμοποιήθηκε παρουσιάζεται στην Παράγραφο 6.4.1. Η συσχέτιση των παραγόμενων χρονοσειρών, η διαδικασία της ρήτρας και η παραγωγή των τελικών χρονοσειρών που χρησιμοποιήθηκαν στην ομογενοποίηση παρουσιάζονται στην Παράγραφο 6.4.2.

Οι αλγόριθμοι που χρησιμοποιήθηκαν για την παραγωγή των συνθετικών χρονοσειρών, όπως και όλες οι επαναληπτικές διαδικασίες που έγιναν στα πλαίσια της Μεταπτυχιακής αυτής εργασίας, γράφτηκαν σε γλώσσα προγραμματισμού Matlab. Αναλυτική παρουσίαση τους και παράθεση των διαγραμμάτων ροής γίνεται στο Κεφάλαιο 7.

6.4.1. Αλγόριθμος παραγωγής συνθετικών χρονοσειρών με μακροπρόθεσμη εμμονή

Στη βιβλιογραφία υπάρχει πλήθος αλγορίθμων που προσπαθούν να αναπαράγουν συνθετικές χρονοσειρές με δυναμική ΗΚ. Όμως, πολλοί από αυτούς δεν είναι εύκολοι ούτε στην κατανόηση αλλά ούτε και στην εφαρμογή (Koutsoyiannis, 2002).

Ο αλγόριθμος που επιλέχτηκε να χρησιμοποιηθεί για τη δημιουργία συνθετικών χρονοσειρών με μακροπρόθεσμη εμμονή, βασίζεται στην προσέγγιση τυχαίων διακυμάνσεων πολλαπλής κλίμακας και προτάθηκε από τον Koutsoyiannis (2002). Ο αλγόριθμος αυτός είναι απλός και γρήγορος στην εφαρμογή και με ικανοποιητική ακρίβεια για τιμές υστέρησης μέχρι και 1000.

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

$$X_i = A_i + B_i + C_i \quad (6.2)$$

με συντελεστές αυτοσυσχέτισης για υστέρηση 1, αντίστοιχα

$$^\circ = 1.52(H - 0.5)^{1.32}, \quad (6.3)$$

$$\ddot{} = 0.953 - 7.69(1 - H)^{3.85}, \text{ και} \quad (6.4)$$

$$\P = \begin{cases} 0.932 + 0.087\text{H}, & \text{H} \le 0.76 \\ 0.993 + 0.007\text{H}, & \text{H} > 0.76 \end{cases} \quad (6.5)$$

και διασπορές

$$Var[A_i] = (1 - c_1 - c_2)¥_0 \quad (6.6)$$

$$Var[B_i] = c_1¥_0 \quad (6.7)$$

$$Var[C_i] = c_2¥_0 \quad (6.8)$$

Οι σταθερές c_1 και c_2 εκτιμώνται με τρόπο ώστε η αυτοσυσχέτιση του αθροίσματος των τριών ανελίξεων

$$^\circ_j = (1 - c_1 - c_2)^{\circ\, j} + c_1^{\ddot{}\, j} + c_2\P^{\,j} \quad (6.9)$$

να ταυτίζεται με τη θεωρητική αυτοσυσχέτιση του κλασματικού Γκαουσιανού θορύβου για υστέρηση 1 και 100. Με βάση τα ανωτέρω, για τον υπολογισμό των c_1 και c_2 λύνουμε το σύστημα:

$$\begin{cases} ^\circ_1 = (1 - c_1 - c_2)^{\circ\, 1} + c_1^{\ddot{}\, 1} + c_2\P^{1} \\ ^\circ_{100} = (1 - c_1 - c_2)^{\circ\, 100} + c_1^{\ddot{}\, 100} + c_2\P^{100} \end{cases} \quad (6.10)$$

Η αυτοσυσχέτιση του κλασματικού Γκαουσιανού θορύβου για υστέρηση j δίνεται από τη Σχέση 6.11.

$$^\circ_j = (1/2)[(j+1)^{2H} + (j-1)^{2H}] - j^{2H}, \qquad j > 0 \quad (6.11)$$

6.4.2. Παραγωγή τελικών συσχετισμένων συνθετικών χρονοσειρών με μακροπρόθεσμη εμμονή

Από τον προηγούμενο αλγόριθμο προκύπτει ένας αριθμός m+1 ως άθροισμα χρονοσειρών αναφοράς και ελεγχόμενης, όπου m ο αριθμός των χρονοσειρών αναφοράς. Οι χρονοσειρές αυτές είναι ανεξάρτητες μεταξύ τους. Για να συσχετιστούν οι χρονοσειρές αυτές, στη γενική περίπτωση (m > 1), χρησιμοποιήθηκε το παρακάτω μοντέλο:

$$Y = AX \qquad (6.12)$$

όπου X οι αρχικές ανεξάρτητες χρονοσειρές και Y οι συσχετισμένες χρονοσειρές και A ένα μητρώο σταθερών όρων.

Για παράδειγμα, στην περίπτωση που θέλουμε να δημιουργήσουμε 3 χρονοσειρές αναφοράς και την ελεγχόμενη, η Σχέση 6.12 παίρνει τη μορφή:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \qquad (6.13)$$

Τα στοιχεία του μητρώου A μπορούν να βρεθούν με υπολογισμούς από το μητρώο συσχετίσεων των χρονοσειρών Y.

$$Y = AX$$

$$YY^T = AXX^T A^T$$

Οι χρονοσειρές X είναι ανεξάρτητες μεταξύ τους, επομένως:

$$E[X, X^T] = Cov[X, X^T] = A \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & \dots \\ \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

Οπότε $E[Y, Y^T] = AE[X, X^T]A^T = AIA^T$

Όμως $E[Y, Y^T] = Cov[Y, Y^T]$ οπότε

$$Cov[Y, Y^T] = AA^T = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1\nu} \\ \rho_{21} & 1 & \dots & \dots \\ \dots & \dots & \dots & \rho_{\nu-1,\nu} \\ \rho_{\nu 1} & \dots & \rho_{\nu,\nu-1} & 1 \end{bmatrix}, \qquad (6.14)$$

όπου ρ_{ij} οι συντελεστές ετεροσυσχέτισης ανάμεσα στις χρονοσειρές Y_i και Y_j.

Για λόγους απλοποίησης της ανάλυσης και ευκολότερης ερμηνείας των αποτελεσμάτων, θεωρήσαμε ότι οι χρονοσειρές αναφοράς και η ελεγχόμενη χρονοσειρά συσχετίζονται μεταξύ τους με τον ίδιο βαθμό συσχέτισης. Επομένως κάναμε την παραδοχή ότι

$$\rho_{12} = \rho_{13} = \dots = \rho_{1\nu} = \rho_{21} = \rho_{\nu 1} = \dots = \rho_{ij} \qquad (6.15)$$

Για παράδειγμα στην περίπτωση 3 χρονοσειρών αναφοράς και για συντελεστή συσχέτισης ρ_{XY} = 0.9 είναι:

$$Cov[Y, Y^T] = AA^T = \begin{bmatrix} 1 & 0.9 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 & 0.9 \\ 0.9 & 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 0.9 & 1 \end{bmatrix}$$

Το σύστημα που προκύπτει από την παραπάνω διαδικασία είναι της μορφής

$$AA^T = B$$

Η επίλυση του συστήματος αυτού για την εύρεση του μητρώου A έγινε με χρήση της μεθόδου Cholesky.

Σε ένα από τα συστήματα χρονοσειρών αναφοράς που επιλέξαμε να μελετήσουμε, το 3/1, έγινε σύνθεση των τριών συσχετισμένων χρονοσειρών αναφοράς Y_1, Y_2, Y_3 που παράξαμε σε μία $Y_{\text{ref final}}$. Καθώς και οι τρεις χρονοσειρές αναφοράς ήταν συσχετισμένες με την ελεγχόμενη χρονοσειρά με τον ίδιο τρόπο, πήραμε κατευθείαν τη μέση τιμή τους. Δηλαδή:

$$Y_{ref,final} = \frac{(Y_1 + Y_2 + Y_3)}{3} \tag{6.16}$$

Αν οι χρονοσειρές αναφοράς είχαν διαφορετικό βαθμό συσχέτισης με την ελεγχόμενη χρονοσειρά, πιο λογικό θα ήταν να ληφθεί κάθε χρονοσειρά αναφοράς, ανάλογα με τη συσχέτισή της με την ελεγχόμενη, με κάποιο συντέλεστή βαρύτητας.

Η διαδικασία που περιγράφτηκε για την παραγωγή συσχετισμένων χρονοσειρών ακολουθήθηκε για αριθμό χρονοσειρών m > 1.

Για m = 1 ακολουθήθηκε μια λίγο διαφορετική τεχνική, κυρίως για λόγους σύγκρισης με κάποια παραδείγματα συνθετικών χρονοσειρών που περιγράφονται από την Στείρου (2011).

Από τις χρονοσειρές X και Y παράχθηκε στη συνέχεια μια χρονοσειρά W σύμφωνα με την Εξίσωση 6.17.

$$W = \frac{X + \kappa Y}{\alpha} \tag{6.17}$$

Ο συντελεστής κ προέκυψε σύμφωνα με τις παρακάτω σχέσεις:

$$Cov[W, Y] = E\left[\frac{X' + \kappa Y'}{\alpha} Y'\right] = E\left[\frac{X'Y' + \kappa Y'^2}{\alpha}\right] = \frac{\kappa}{\alpha} Var[Y] = \frac{\kappa}{\alpha} \sigma_y^2 \tag{6.18}$$

Από τη Σχέση 6.18 και βάσει του ορισμού του συντελεστή ετεροσυσχέτισης προκύπτει

$$\rho = \frac{Cov[W, Y]}{\sigma_W \cdot \sigma_Y} \quad \rightarrow \quad \rho = \frac{\kappa \cdot \sigma_Y}{\alpha \cdot \sigma_W} \tag{6.19}$$

Όμως $\sigma_W^2 = E\left[\left(\frac{X' + \kappa Y'}{\alpha}\right)^2\right] = E\left[\frac{1}{\alpha^2} X'^2 + 2\frac{\kappa}{\alpha} X'Y' + \frac{\kappa^2}{\alpha^2} Y'^2\right] = \frac{1}{\alpha^2}\sigma_X^2 + \frac{\kappa^2}{\alpha^2}\sigma_Y^2$ (6.20)

Από τη Σχέση 6.19 με ύψωση στο τετράγωνο προκύπτει

$$\circ^2 = \frac{\copyright^2 \cdot \pm_Y{}^2}{a^2 \cdot \pm_W{}^2} \qquad \rightarrow \qquad \pm_W^2 = \frac{\copyright^2}{a^2}\frac{\pm_Y^2}{\circ^2} \tag{6.21}$$

Από τις Σχέσεις 6.20 και 6.21 έχουμε:

$$\frac{1}{a^2}\pm_X^2 + \frac{\copyright^2}{a^2}\pm_Y^2 = \frac{\copyright^2}{a^2}\frac{\pm_Y^2}{\circ^2} \qquad \rightarrow \qquad \pm_X^2 + \copyright^2\pm_Y^2 = \frac{\copyright^2\pm_Y^2}{\circ^2} \qquad \rightarrow \qquad \circ^2\pm_X^2 + \circ^2\copyright^2\pm_Y^2 = \copyright^2\pm_Y^2 \rightarrow$$

$$\copyright^2 = \frac{\circ^2 \cdot \pm_X^2}{\pm_Y^2 - \circ^2 \cdot \pm_Y^2} \tag{6.22}$$

Ο συντελεστής λ προέκυψε με τη δέσμευση $\pm_W^2 = 1$ από τη Σχέση 6.20

$$a^2 = \pm_X^2 + \copyright^2 \cdot \pm_Y^2 = \frac{1}{1-\circ^2}\pm_X^2 \tag{6.23}$$

Για τη χρονοσειρά W που προέκυπτε από την παραπάνω διαδικασία γινόταν στη συνέχεια εκτίμηση του συντελεστή Hurst. Η εκτίμηση αυτή έγινε με τη μέθοδο ελαχίστων τετραγώνων βασισμένη σε εκτιμητή μεταβλητότητας (LSV, least squares based on variance) η οποία προτάθηκε από τους Tyralis and Koutsoyiannis (2011).

Η εκτίμηση αυτή περιέχει μια αβεβαιότητα. Επιπλέον η χρονοσειρά W, αν και προκύπτει γραμμικά από τις X, Y, είχε συχνά διαφορετικό συντελεστή Hurst από τον επιθυμητό. Για τους λόγους αυτούς τέθηκε στο συντελεστή Hurst μία ρήτρα. Αν H η επιθυμητή τιμή και H15 η πραγματική τιμή του συντελεστή Hurst της χρονοσειράς W, τότε για να γίνει δεκτή η W έπρεπε να ισχύει:

$$\begin{cases} H15 \le H + 0.05 \\ H15 \ge H - 0.05 \end{cases} \tag{6.24}$$

Στις περιπτώσεις που δεν ικανοποιούνταν οι παραπάνω περιορισμοί, η χρονοσειρά W παραγόταν από την αρχή και γινόταν ξανά εκτίμηση του συντελεστή Hurst. Η διαδικασία αυτή επαναλαμβανόταν μέχρι να προκύψει χρονοσειρά που να ικανοποιεί τη ρήτρα που θέσαμε.

6.5 Εφαρμογή σε θερμοκρασίες

6.5.1. Επιλογή τρόπου εφαρμογής μεθόδου SNHT

Η μέθοδος SNHT έχει πολλούς τρόπους εφαρμογής (βλ. Παράγραφο 3.2.2). Μπορεί να εφαρμοστεί είτε ως απόλυτη είτε ως σχετική μέθοδος, δηλαδή με χρήση χρονοσειρών αναφοράς. Επιπλέον μπορεί να εντοπίζονται μία ή περισσότερες ανομογένειες, άλματα ή και αλλαγές κλίσης.

Στην παρούσα Μεταπτυχιακή εργασία επιλέχτηκε να εξεταστεί μια εκδοχή της μεθόδου σχετική, η οποία να εντοπίζει πολλαπλά άλματα. Δύο τέτοιες μεθοδολογίες (SNH1 και SNH2) δίνονται από τον Domonkos (2011a). Παρουσίαση των δύο αλγορίθμων δίνεται στην Παράγραφο 5.3. Για λόγους πληρότητας επαναλαμβάνουμε και εδώ τα βασικά τους στοιχεία.

Το SNH1 ξεκινάει ελέγχοντας όλη τη χρονοσειρά για την πιο σημαντική ανομογένεια. Εφόσον αυτή εντοπιστεί και είναι στατιστικά σημαντική, η χρονοσειρά διαιρείται σε δύο τμήματα και τα τμήματα αυτά εξετάζονται για περαιτέρω ανομογένειες (cutting algorithm). Η διαδικασία αυτή επαναλαμβάνεται στα τμήματα της χρονοσειράς που προκύπτουν κάθε φορά από το σπάσιμο, εφόσον το μήκος των τμημάτων είναι επαρκές για περαιτέρω εξέταση. Το κατώφλι αυτό λήφθηκε από τον Domonkos (2011a) 10 έτη. Η ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς λήφθηκε από τον Domonkos (2011a) 5 έτη.

Η εκδοχή SNH2 διαφέρει ελαφρά από το SNH1. Στη μέθοδο αυτή εντοπίζονται οι θέσεις των πιθανών ανομογενειών με τη μέθοδο που χρησιμοποιείται και στο SNH1. Στη συνέχεια καθεμιά από τις πιθανές ανομογένειες ξαναελέγχεται με χρήση του τμήματος της χρονοσειράς που περιέχει μόνο την ανομογένεια αυτή και οριοθετείται από μία άλλη ανομογένεια, πιθανή ανομογένεια ή την άκρη της χρονοσειράς.

Από τις δύο μεθοδολογίες αρχικά φάνηκε λογικό να προτιμηθεί η SNH2, αφού φαίνεται να μειώνει την αλληλεπίδραση διαφορετικών ανομογενειών κατά τον υπολογισμό των διορθώσεων. Παρουσιάζει όμως παράλληλα και κάποια μειονεκτήματα, τα οποία εντοπίστηκαν από μια αρχική διερευνητική εφαρμογή και των δύο μεθοδολογιών:

- Είναι πιο πολύπλοκη από το SNH1.
- Προέκυψαν αρκετές περιπτώσεις κατά τις οποίες η σημαντικότερη ανομογένεια που εντοπιζόταν στο πρώτο βήμα εφαρμογής της μεθόδου SNH2, δεν βρισκόταν στη συνέχεια στατιστικά σημαντική.

Κρίνοντας ότι το δεύτερο στοιχείο κυρίως είναι αρκετά προβληματικό, αποφασίστηκε να γίνει χρήση τελικά της μεθόδου SNH1. Ένα δεύτερο στοιχείο που συνηγόρησε στην επιλογή της εκδοχής SNH1 είναι η πολύ μικρή διαφορά στα αποτελέσματα των δύο μεθόδων κατά την εφαρμογή σε συνθετικά δεδομένα από τον Domonkos (2011a).

Στο εξής, εφόσον επιλέχτηκε να εφαρμοστεί μια μοναδική εκδοχή της μεθόδου SNHT, αυτή θα αποκαλείται με τη γενική ονομασία της μεθόδου.

Η μέθοδος SNHT μπορεί να εφαρμοστεί για πολλά επίπεδα σημαντικότητας. Εδώ κάναμε χρήση της συνήθους πρακτικής για **επίπεδο εμπιστοσύνης 95%**. Συμπερασματικά, **ποσοστό ψευδούς εντοπισμού** ανομογενειών από τη μέθοδο **μεγαλύτερο από 5%** θεωρήθηκε προβληματικό. Εξετάστηκε επομένως, για διάφορες περιπτώσεις χρονοσειρών, αν και κατά πόσο οι ψευδείς εντοπισμοί υπερβαίνουν το ποσοστό 5%.

Οι τιμές της test value T_{crit} λήφθηκαν από τους Khaliq and Ouarda (2007), οι οποίοι υπολόγισαν τις κρίσιμες τιμές για το SNHT με χρήση πολύ μεγάλου αριθμού (ενός εκατομμυρίου) προσομοιώσεων Monte Carlo και για 108 διαφορετικά μήκη χρονοσειρών. Οι τιμές αυτές είχαν υπολογιστεί και από τους Alexandersson and Moberg (1997), για λίγα όμως μήκη χρονοσειρών και με μικρότερο αριθμό προσομοιώσεων.

Παρά την εξαγωγή κρίσιμων τιμών για πολλά μήκη δείγματος, η μελέτη των Khaliq and Ouarda (2007) δεν καλύπτει όλο το εύρος τιμών μεταξύ 10 και 100 ετών, το οποίο μελετάμε. Έτσι για κάποια μήκη οι τιμές T_{crit} υπολογίστηκαν με γραμμική παρεμβολή. Ο πίνακας με τις κρίσιμες τιμές T_{crit} (προτεινόμενες από τους Khaliq and Ouarda (2007) και υπολογισμένες) παρατίθεται στο Παράρτημα Α.

6.5.2. Συνδυασμοί παραμέτρων

Η μέθοδος SNHT, με τη μορφή που περιγράψαμε στην Παράγραφο 6.5.1, αποφασίστηκε να εφαρμοστεί σε επιλεγμένα συστήματα χρονοσειρών αναφοράς για όλες τις τιμές των βασικών παραμέτρων Η και ρ_{XY}. Δηλαδή για:

Η={0.50, 0.55, ..., 0.90} και ρ_{XY}={0.5, 0.6, 0.7, 0.8, 0.9, 0.95}.

Η διερεύνηση αποφασίστηκε να γίνει σε 3 ενότητες:

1. Εντοπισμός αλμάτων στα συστήματα χρονοσειρών αναφοράς 1/1, 3/7, 2/2, 3/3, 3/10 και 3/20 για διερεύνηση των βασικών παραμέτρων.
2. Εντοπισμός αλμάτων στα συστήματα χρονοσειρών αναφοράς 3/7-50 years και 3/7-j10 για (μικρότερης κλίμακας) διερεύνηση των δευτερευουσών παραμέτρων.
3. Εφαρμογή των διορθώσεων που υπολογίστηκαν στην ενότητα 1 για το σύστημα χρονοσειρών αναφοράς 1/1 και υπολογισμός της τάσης και του συντελεστή Hurst των χρονοσειρών πριν και μετά τις διορθώσεις.

Αν σε κάποιο βήμα αποδεικνυόταν ότι κάποια παράμετρος δεν επηρέαζε τα αποτελέσματα αποφασίστηκε στα επόμενα βήματα να παραμείνει σταθερή. Για κάθε διαφορετικό συνδυασμό παραμέτρων έγιναν 10000 επαναλήψεις.

Επίσης στην ενότητα 3 έγινε διόρθωση των αλμάτων που υπολογίστηκαν μόνο από τη μέθοδο SNHT στη βασική περίπτωση 1/1, δηλαδή από τη χρήση 1 χρονοσειράς αναφοράς. Δεν έγινε διόρθωση των κοινών αλμάτων που προέκυψαν από τις άλλες περιπτώσεις χρήσης των χρονοσειρών αναφοράς, π.χ. την 3/7, για ένα πρακτικό λόγο: από τις διαφορετικές μεθόδους ομογενοποίησης προέκυπταν κοινά μεν άλματα, αλλά το μέγεθος και θέση των οποίων δεν ταυτίζονταν απόλυτα. Για τον υπολογισμό των δύο αυτών στοιχείων έπρεπε επομένως να γίνουν κάποιες επιπλέον παραδοχές και να αναπτυχθεί μια επιπλέον μεθοδολογία, το οποίο αποφασίστηκε να μη γίνει στα πλαίσια αυτής της Μεταπτυχιακής εργασίας, αλλά θα μπορούσε να αποτελέσει μια μελλοντική προέκτασή της.

6.6 Εφαρμογή σε βροχοπτώσεις

6.6.1. Επιλογή τρόπου εφαρμογής Διπλής αθροιστικής καμπύλης

Η Διπλή αθροιστική καμπύλη είναι παραδοσιακά υποκειμενική μέθοδος ομογενοποίησης, καθώς στην κλασική της μορφή γίνεται εύρεση της μεγαλύτερης αλλαγής κλίσης με οπτική επισκόπηση του γραφήματος. Μια αντικειμενική παραλλαγή της μεθόδου με εύρεση της μεγαλύτερης αλλαγής κλίσης με κάποιο αυτόματο τρόπο προτάθηκε από τους Chang and Lee (1974), με παράλληλη ενσωμάτωση της ανάλυσης μεταβλητότητας-συμμεταβλητότητας για τον καθορισμό της στατιστικής σημαντικότητας των ανομογενειών.

Στην παρούσα Μεταπτυχιακή εργασία έγινε εφαρμογή της Διπλής αθροιστικής καμπύλης σε μερικές χιλιάδες χρονοσειρές. Επομένως ήταν πρακτικά αδύνατο να εφαρμοστεί στην υποκειμενική της μορφή.

Αποφασίστηκε επομένως να αναπτυχθεί ένα μοντέλο στην ίδια λογική με τους Chang and Lee (1974), το οποίο να προσαρμόζει στα δεδομένα δύο ευθείες διαφορετικής κλίσης και να εντοπίζει το σημείο με τη μεγαλύτερη αλλαγή κλίσης.

Το μοντέλο που αναπτύχθηκε βασίζεται στην αρχή των ελαχίστων τετραγώνων και κάνει χρήση αναλυτικής επίλυσης και επαναληπτικής διαδικασίας.

Συγκεκριμένα η μεθοδολογία που υιοθετήθηκε ακολουθεί τα εξής βήματα:

1. Υπολογισμός των συσσωρευτικών αθροισμάτων των ετήσιων βροχοπτώσεων Σx_i και Σy_i στο σταθμό αναφορας και τον ελεγχόμενο, αντίστοιχα, για την ίδια χρονική περίοδο. Η διαδοχική άθροιση των σημείων x_i και y_i ώστε να υπολογιστούν οι τιμές Σx_i και Σy_i γίνεται σύμφωνα με τους δηλαδή από την παλαιότερη προς την πιο πρόσφατη ημερομηνία.
2. Εύρεση με αναλυτικό τρόπο, για όλες τις τετμημένες των σημείων της Διπλής αθροιστικής καμπύλης, των κλίσεων των δύο καμπυλών που προσαρμόζονται καλύτερα στα δεδομένα (με κάποιες προϋποθέσεις που θα αναφερθούν παρακάτω).
3. Εύρεση του σημείου της Διπλής αθροιστικής καμπύλης στο οποίο αντιστοιχεί η μεγαλύτερη αλλαγή κλίσης.

Στην παραπάνω μεθοδολογία δεν συμπεριλαμβάνεται η αναλύση μεταβλητότητας-συμμεταβλητότητας, λαμβάνοντας υπόψη και τις επιφυλάξεις του Dingman (1994) σε σχέση με την εφαρμοσιμότητα της μεθόδου λόγω της άθροισης των δεδομένων των χρονοσειρών.

Επίσης, οι ανομογένειες που εντοπίστηκαν κατά την εφαρμογή της Διπλής αθροιστικής καμπύλης δεν διορθώθηκαν. Η μελέτη επικεντρώθηκε στη διερεύνηση του λόγου των κλίσεων των δύο ευθειών που προσαρμόζονται στα δεδομένα κατά την εφαρμογή της μεθόδου.

6.6.2. Αλγόριθμος διγραμμικής προσαρμογής

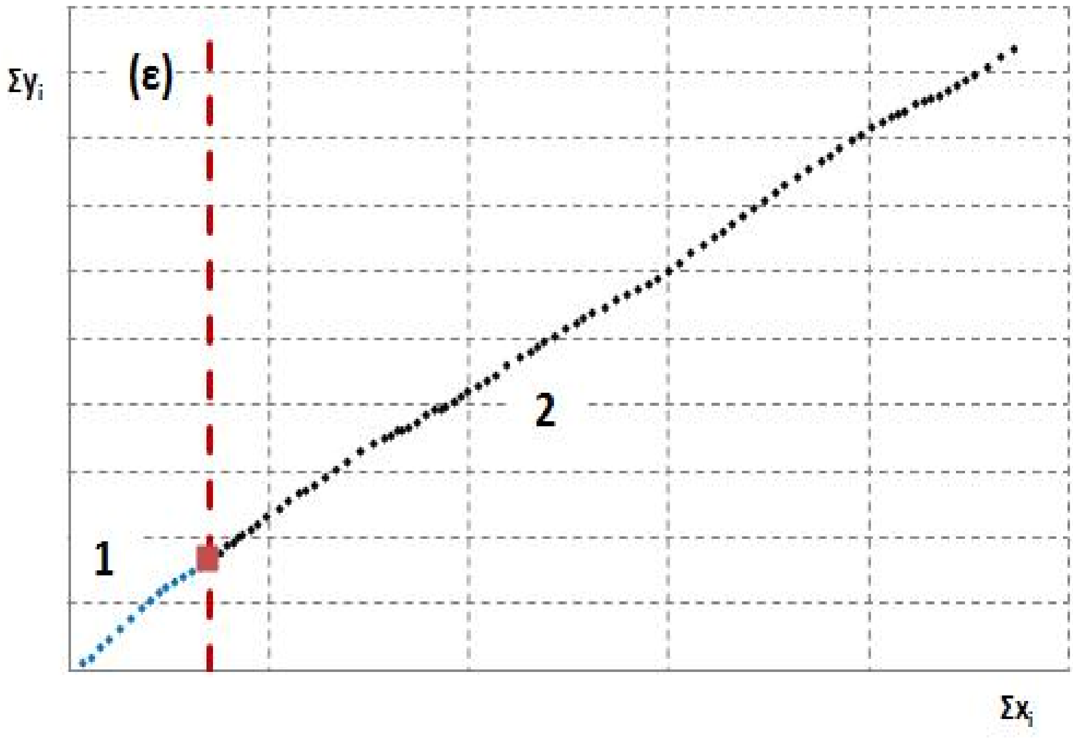
Αναπτύχθηκε ένας αλγόριθμος ο οποίος:

Α) Λαμβάνει ως δεδομένα όλα τα σημεία Σx_i και Σy_i που προκύπτουν από την άθροιση των δεδομένων της χρονοσειράς αναφοράς και της ελεγχόμενης.

Β) Στη συνέχεια θεωρεί ένα σημείο από αυτά. Η τετμημένη του σημείου αυτού ορίζει μια κατακόρυφη ευθεία ε, η οποία διαχωρίζει τα σημεία της καμπύλης σε δύο ενότητες (Σχήμα 6.1).

Γ) Προσαρμόζει στις δύο ενότητες σημείων δύο ευθείες που τέμνονται στη ευθεία ε και υπολογίζει το άθροισμα των τετραγώνων των αποκλίσεων των τεταγμένων όλων των σημείων από τα αντίστοιχα σημεία των δύο ευθειών. Η ευθεία που προσαρμόζεται στο πρώτο τμήμα πρέπει να περνάει από την αρχή των αξόνων, να είναι δηλαδή της μορφής $y=\alpha x$.

Δ) Επαναλαμβάνει τη διαδικασία για όλα τα σημεία και επιλέγει το σημείο που οδηγεί στο ελάχιστο άθροισμα των τετραγώνων των αποκλίσεων.



Σχήμα 6.1. *Διαχωρισμός των σημείων (Σx_i, Σy_i) που προκύπτουν από τα συσσωρευτικά αθροίσματα της χρονοσειράς αναφοράς και της ελεγχόμενης σε δύο ενότητες, στις οποίες προσαρμόζονται οι δύο ευθείες της Διπλής αθροιστικής καμπύλης.*

Ε) Για το σημείο αυτό επιστρέφει σαν αποτέλεσμα τις κλίσεις των δύο ευθειών που προσαρμοζονται στα δεδομένα και τη θέση του σημείου στην αρχική χρονοσειρά (δηλαδή όχι με αντίστροφη χρονολογική σειρά).

Ένα παράδειγμα συνθετικών δεδομένων με προσαρμογή της Διπλής αθροιστικής καμπύλης φαίνεται στο Σχήμα 6.2. Φαίνονται οι δύο ευθείες που προσαρμόζονται καλύτερα στα δεδομένα με κόκκινο χρώμα. Στο σημείο $A(x_0, y_0)$ παρατηρείται αλλαγή κλίσης, η οποία ερμηνεύεται ως ύπαρξη ανομογένειας. Με μπλε χρώμα φαίνεται η προέκταση της ευθείας y=αx. Τα δεδομένα θα θεωρούνταν ομογενή αν με τη μέθοδο των ελαχίστων τετραγώνων προσαρμοζόταν στο δεύτερο τμήμα η μπλε ευθεία.

Για να γίνει προσαρμογή των δύο ευθειών στα δεδομένα, πρέπει να υπολογιστούν οι κλίσεις τους α και β και οι συντεταγμένες των σημείων Α και Β.

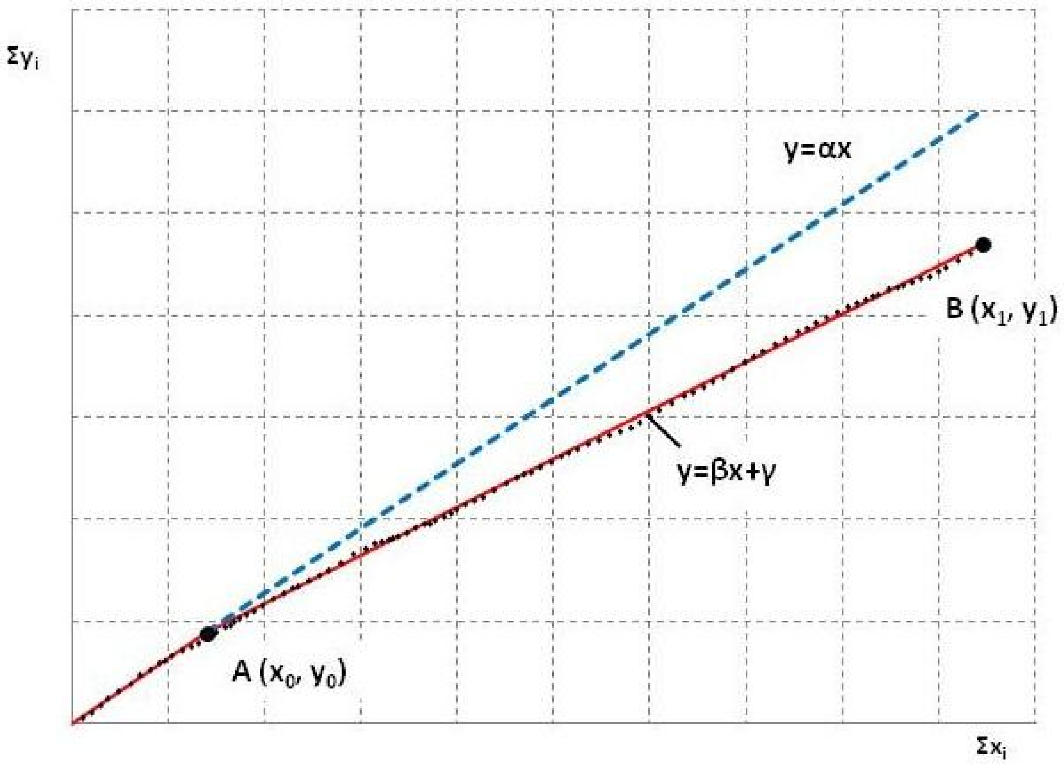
Η τετμημένη του σημείου Β x_1, είναι γνωστή (τετμημένη του 100 αθροίσματος). Η τετμημένη του Α είναι επίσης γνωστή (αρχική θεώρηση). Οι τεταγμένες των δύο σημείων υπολογίζονται με αναλυτικό τρόπο από την ελαχιστοποίηση των τετραγώνων. Τέλος, οι κλίσεις των δύο ευθειών α και β υπολογίζονται συναρτήσει των συντεταγμένων των Α και Β.

Η διαδικασία αυτή περιγράφεται αναλυτικά παρακάτω:

Για την πρώτη ευθεία είναι:

$$y = \sim x \qquad \rightarrow \qquad \sim = \frac{y_0}{x_0} \tag{6.25}$$

Η εξίσωση της ευθείας παίρνει τη μορφή: $$y = \frac{y_0}{x_0} x \tag{6.26}$$



Σχήμα 6.2. *Παράδειγμα προσαρμογής της Διπλής αθροιστικής καμπύλης σε δεδομένα βροχόπτωσης (κόκκινη γραμμή). Με μπλε διακεκομμένη γραμμή φαίνεται η ευθεία που θα έπρεπε να προσαρμοστεί στα δεδομένα του τμήματος ΑΒ ώστε να θεωρούνται από τη μέθοδο ομογενή.*

Για τη δεύτερη ευθεία της μορφής $y = \ x + ¥$, είναι:

$$\begin{cases} y_1 = \ x_1 + ¥ \\ y_0 = \ x_0 + ¥ \end{cases},$$ οπότε η κλίση της υπολογίζεται από τη Σχέση:

$$= \frac{y_1 - y_0}{x_1 - x_0} \tag{6.27}$$

Η παράμετρος γ συναρτήσει των x_0, x_1, y_0, y_1 παίρνει τη μορφή:

$$¥ = \frac{y_1 - y_0}{x_1 - x_0} x + y_0 - \frac{y_1 - y_0}{x_1 - x_0} x_0 \tag{6.28}$$

Το ζήτούμενο στον αλγόριθμο είναι η ελαχιστοποίηση των τετραγώνων των κατακόρυφων αποστάσεων των σημείων (Σx_i, Σy_i) από τις ευθείες που προσαρμόζουμε. Δηλαδή το άθροισμα: $\sum \left(y_i - \hat{y}_i \right)^2 = \min$

Η ποσότητα επομένως που πρέπει να ελαχιστοποιηθεί παίρνει με βάση τις Σχέσεις 6.26, 6.27 και 6.28 την παρακάτω μορφή:

$$Q = \sum_{x \le x_0} \left(y_i - \frac{y_0}{x_0} x_i \right)^2 + \sum_{x > x_0} \left(y_i - \frac{y_1 - y_0}{x_1 - x_0} x_i - y_0 + \frac{y_1 - y_0}{x_1 - x_0} x_0 \right)^2 \tag{6.29}$$

Οι δύο μεταβλητές στην ποσότητα Q είναι οι τεταγμένες , y_0 και y_1, οπότε για να ελαχιστοιήσουμε την Q πρέπει να μηδενιστούν οι μερικές της παράγωγοι ως προς y_0 και y_1. Δηλαδή πρέπει να λυθεί το σύστημα:

$$\begin{cases} \dfrac{\partial Q}{\partial y_0} = 0 \\ \dfrac{\partial Q}{\partial y_1} = 0 \end{cases} \tag{6.30}$$

Από την επίλυση του συστήματος προκύπτουν οι ζητούμενες τεταγμένες y_0 και y_1 (Σχέσεις 6.31 και 6.32).

$$y_0 = \frac{\frac{D}{C}\frac{x_1 - x_0}{x_0} \sum_{x \le x_0} x_i y_i - \frac{D}{C} \sum_{x > x_0} x_i y_i + \frac{D}{C} x_1 \sum_{x > x_0} y_i - \sum_{x > x_0} x_i y_i - x_0 \sum_{x > x_0} y_i}{m x_0 - \sum_{x > x_0} x_i + \frac{x_1 - x_0}{x_0^2} \frac{D}{C} \sum_{x \le x_0} x_i^2 - \frac{D}{C} \sum_{x > x_0} x_i + \frac{D}{C} x_1 m} \tag{6.31}$$

$$y_1 = y_0 + \frac{(x_1 - x_0)(m x_0 y_0 + \sum_{x > x_0} x_i y_i - x_0 \sum_{x > x_0} y_i - y_0 \sum_{x > x_0} x_i}{D} \tag{6.32}$$

όπου: $$C = x_0 \sum_{x > x_0} x_i - m x_0 x_1 - \sum_{x > x_0} x_i^2 + x_1 \sum_{x > x_0} x_i \tag{6.33}$$

$$D = m x_0^2 + \sum_{x > x_0} x_i^2 - 2 x_0 \sum_{x > x_0} x_i \tag{6.34}$$

Στη συνέχεια, οι κλίσεις α και β υπολογίζονται από τις Σχέσεις 6.25 και 6.27, αντίστοιχα.

6.6.3. Συνδυασμοί παραμέτρων

Ομοίως με τη μέθοδο SNHT, αποφασίστηκε να εφαρμοστεί η Διπλή αθροιστική καμπύλη σε επιλεγμένα συστήματα χρονοσειρών αναφοράς για όλες τις τιμές των βασικών παραμέτρων H και ρ_{XY}. Δηλαδή για:

H={0.50, 0.55, ..., 0.90} και ρ_{XY}={0.5, 0.6, 0.7, 0.8, 0.9, 0.95}.

Βασική διαφορά με το SNHT ήταν ότι τυχόν παράμετροι που βρέθηκαν να μην επηρεάζουν τα αποτελέσματα του SNHT θα λαμβάνονταν και εδώ σταθεροί.

Η διερεύνηση αποφασίστηκε να γίνει σε 2 ενότητες:

1. Εντοπισμός αλμάτων στα συστήματα χρονοσειρών αναφοράς 1/1 και 3/1 για διερεύνηση των βασικών παραμέτρων.
2. Για δύο συντελεστές Hurst H=0.75 και H=0.50 και συντελεστή ετεροσυσχέτισης ρ_{XY}=0.8, σε συνδυασμό με τα συστήματα χρονοσειρών αναφοράς 1/1 και 3/1, για τη διερεύνηση της κατανομής του λόγου των κλίσεων α και β των δύο ευθειών που αποτελούν τη Διπλή αθροιστική καμπύλη.

Εξετάστηκαν τα δύο μόνο συστήματα χρονοσειρών αναφοράς 1/1 και 3/1 γιατί η χρήση μιας χρονοσειράς αναφοράς ή μιας συνθετικής χρονοσειράς από τις γειτονικές χρονοσειρές της ελεγχόμενης είναι οι πιο συνήθεις κατά την εφαρμογή της Διπλής αθροιστικής καμπύλης (Searcy and Hardison, 1960).

Στην πρώτη ενότητα προσαρμόστηκαν σε καθεμία από τις 2000 προσομοιώσεις δεδομένων βροχόπτωσης που έγιναν οι δύο ευθείες της Διπλής αθροιστικής καμπύλης και υπολογίστηκαν οι κλίσεις τους. Στη συνέχεια εξετάστηκε τι ποσοστό των 2000 προσομοιώσεων αντιστοιχεί σε λόγο λ των κλίσεων α και β, μεγαλύτερο από κάποια τιμή, όπου:

$$\mathrm{a} = \begin{cases} \frac{\sim}{\quad}, & \sim > \\ \\ \frac{\quad}{\sim}, & \sim < \end{cases} \qquad (6.35)$$

Οι τιμές του λόγου λ που εξετάστηκαν ήταν οι {1.1, 1.15, …, 1.6}, δηλαδή συνολικά 11 τιμές.

Με αυτό τον τρόπο σχεδιάστηκαν καμπύλες του ποσοστού ψευδούς εντοπισμού συναρτήσει του συντελεστή Hurst και του λόγου λ των κλίσεων α και β.

Στη δεύτερη ενότητα εξετάστηκαν μόνο οι συντελεστές H=0.75, ο οποίος έχει παρατηρηθεί σε πραγματικές χρονοσειρές (Koutsoyiannis and Langousis, 2011) και H=0.50, ο οποίος αντιστοιχεί στη συνήθη θεώρηση ασυσχέτιστων δεδομένων.

Εφόσον δεν έγινε ευρείας κλίμακας διερεύνηση, επιλέχτηκε να εξεταστεί μία μόνο τιμή του συντελεστή ετεροσυσχέτισης, ρ_{XY}=0.8. Η τιμή αυτή δίνεται στη βιβλιογραφία ως μια τυπική τιμή για το συντελεστή συσχέτισης ανάμεσα σε χρονοσειρές αναφοράς και ελεγχόμενες (π.χ. Ducré-Robitaille *et al*., 2003, DeGaetano, 2006, Vincent, 1998, Easterling and Peterson, 1995).

Για καθένα από τους 4 συνδυασμούς παραμέτρων της ενότητας 2 φτιάχτηκαν διαγράμματα του όρου k (Σχέση 6.36), ο οποίος ορίζεται ως ο νεπέριος λογάριθμος του λόγου των κλίσεων α και β των δύο ευθειών της Διπλής αθροιστικής καμπύλης. Ο λογάριθμος επιλέχτηκε για λόγους συμμετρίας.

$$k = \ln\left(\frac{\tilde{\ }}{\ }\right) \qquad (6.36)$$

Για κάθε διαφορετικό συνδυασμό παραμέτρων κατά την εφαρμογή της Διπλής αθροιστικής καμπύλης έγιναν 2000 επαναλήψεις. Οι επαναλήψεις αυτές είναι λιγότερες από το SNHT (10000 επαναλήψεις). Η απόφαση να χρησιμοποιηθεί μικρότερο δείγμα πάρθηκε γιατί σε διερευνητική εφαρμογή της μεθόδου σε διάφορα μεγέθη δείγματος η μέθοδος συνέκλινε για 2000 επαναλήψεις. Προτιμήθηκε επομένως το δείγμα των 2000 αφού μείωνε κατά πολύ τον απαιτούμενο χρόνο για τους υπολογισμούς.

7. ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ ΓΙΑ ΠΑΡΑΓΩΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΧΡΟΝΟΣΕΙΡΩΝ

Στο Κεφάλαιο αυτό παρουσιάζονται οι συναρτήσεις σε γλώσσα Matlab που αναπτύχθηκαν για την εφαρμογή της μεθοδολογίας που περιγράφεται στο Κεφάλαιο 6.

Συνολικά αναπτύχθηκαν τρεις κύριες συναρτήσεις για την αξιολόγηση των μεθόδων ομογενοποίησης, δύο για τη μέθοδο SNHT και μία για τη Διπλή αθροιστική καμπύλη. Στη μέθοδο SNHT χρησιμοποιήθηκαν δύο διαφορετικές συναρτήσεις, καθώς εφαρμόστηκε ελαφρώς διαφορετική μεθοδολογία για χρήση μίας ή πολλαπλών χρονοσειρών αναφοράς (βλ. Παράγραφο 6.4.2.).

Εκτός από τις κύριες συναρτήσεις, αναπτύχθηκαν συμπληρωματικά 11 βοηθητικές συναρτήσεις για τη δημιουργία συνθετικών χρονοσειρών αναφοράς και ελεγχόμενων, την εφαρμογή των μεθόδων ομογενοποίησης με συγκεκριμένο τρόπο και τη διόρθωση των χρονοσειρών σε κάποιες περιπτώσεις. Επιπλέον, για την εκτίμηση του συντελεστή Hurst των χρονοσειρών, χρησιμοποιήθηκε ως βοηθητική συνάρτηση, συνάρτηση που προτάθηκε από τους Tyralis and Koutsoyiannis (2011) και βασίζεται στη μέθοδο ελαχίστων τετραγώνων με εκτιμητή μεταβλητότητας (LSV, least squares based on variance). Συνολικά χρησιμοποιήθηκαν 12 βοηθητικές συναρτήσεις.

Το λογισμικό που αναπτύχθηκε έγινε προσπάθεια να είναι όσο πιο γενικό γίνεται ώστε να μπορεί να καλύψει με αυτόματο τρόπο περιπτώσεις χρονοσειρών με διαφορετικά χαρακτηριστικά (π.χ. μήκος, συντελεστής ετεροσυσχέτισης κ.ά.) και διαφορετικών τρόπων εφαρμογής των μεθόδων ομογενοποίησης.

Οι κώδικες αναλυτικά για τις κύριες και βοηθητικές συναρτήσεις που αναπτύχθηκαν παρατίθενται στο Παράρτημα Β. Για ευκολότερη όμως επισκόπηση παρατίθενται στις επόμενες Παραγράφους περιγραφές και διαγράμματα ροής των κύριων (Παράγραφος 7.1) και βοηθητικών συναρτήσεων (Παράγραφος 7.2) που χρησιμοποιήθηκαν.

7.1. Κύριες συναρτήσεις

Όπως αναφέρθηκε και παραπάνω, αναπτύχθηκαν οι ακόλουθες 3 κύριες συναρτήσεις για τον έλεγχο της μεθόδου SNHT και της Διπλής αθροιστικής καμπύλης (Πίνακας 7.1).

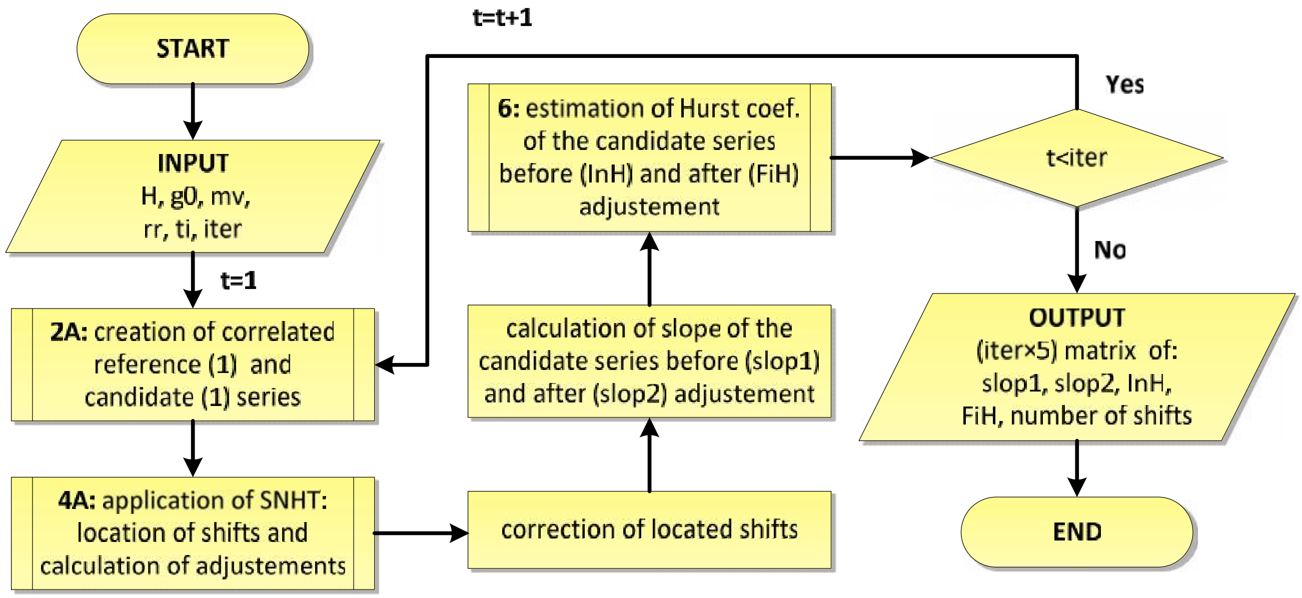
Πίνακας 7.1. *Κύριες συναρτήσεις που αναπτύχθηκαν με τους αντίστοιχους συμβολισμούς τους*

Συμβολισμός	Συνάρτηση
I	CheckSNHT_multiple
II	CheckSNHT_Mref
III	CheckDMass1

I) Συνάρτηση CheckSNHT_multiple:

Εφαρμόζει τη μέθοδο SNHT σε χρονοσειρές αναζητώντας πολλαπλά άλματα με χρήση μίας χρονοσειράς αναφοράς. Μετά τον εντοπισμό αλμάτων διορθώνει τις χρονοσειρές και υπολογίζει τις τάσεις και τους συντελεστές Hurst των χρονοσειρών πριν και μετά την ομογενοποίηση και τον αριθμό των αλμάτων που εντόπισε το SNHT.

Οι επιμέρους διαδικασίες γίνονται με τη χρήση βοηθητικών συναρτήσεων, περιγραφή των οποίων γίνεται στην Παράγραφο 7.2. Το διάγραμμα ροής της συνάρτησης CheckSNHT_multiple παρουσιάζεται στο Σχήμα 7.1.



Σχήμα 7.1. *Διάγραμμα ροής συνάρτησης CheckSNHT_multiple, η οποία εφαρμόζει τη μέθοδο SNHT σε συνθετικές χρονοσειρές αναζητώντας πολλαπλά άλματα με χρήση μιας χρονοσειράς αναφοράς*

Η συνάρτηση CheckSNHT_multiple λαμβάνει ως εισόδους τις μεταβλητές:

H: επιθυμητός συντελεστής Hurst των χρονοσειρών
g0: μεταβλητότητα των χρονοσειρών
mv: μεταβλητότητα των χρονοσειρών
rr: συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς-ελεγχόμενες
ti: μήκος των χρονοσειρών
iter: αριθμός προσομοιώσεων Monte Carlo

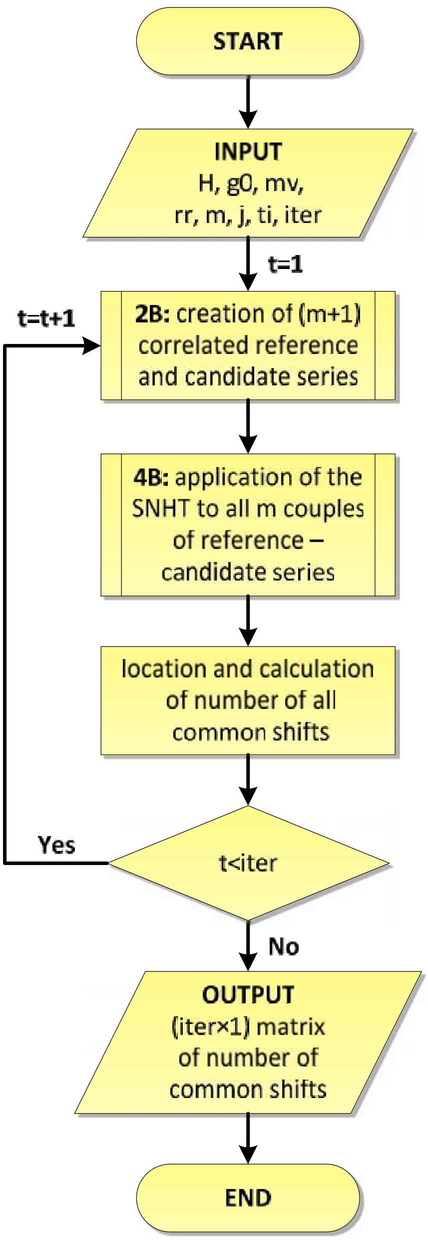
Ως έξοδοι επιστρέφονται:

slop1: Η γραμμική τάση της ελεγχόμενης χρονοσειράς πριν την ομογενοποίηση
slop2: Η γραμμική τάση της ελεγχόμενης χρονοσειράς μετά την ομογενοποίηση
InH: Ο συντελεστής Hurst της ελεγχόμενης χρονοσειράς πριν την ομογενοποίηση
FiH: Ο συντελεστής Hurst της ελεγχόμενης χρονοσειράς μετά την ομογενοποίηση
alm: Ο αριθμός των αλμάτων που εντοπίστηκαν

II) Συνάρτηση CheckSNHT_Mref:

Εφαρμόζει τη μέθοδο SNHT σε χρονοσειρές αναζητώντας πολλαπλά άλματα με χρήση πολλαπλών χρονοσειρών αναφοράς. Επιστρέφει τον αριθμό των αλμάτων που εντόπισε το SNHT.

Το διάγραμμα ροής της συνάρτησης CheckSNHT_Mref παρουσιάζεται στο Σχήμα 7.2.



Σχήμα 7.2. *Διάγραμμα ροής συνάρτησης CheckSNHT_Mref, η οποία εφαρμόζει τη μέθοδο SNHT σε συνθετικές χρονοσειρές αναζητώντας πολλαπλά άλματα με χρήση πολλαπλών χρονοσειρών αναφοράς*

Η συνάρτηση CheckSNHT_Mref λαμβάνει ως εισόδους τις μεταβλητές:

H: επιθυμητός συντελεστής Hurst των χρονοσειρών
g0: μεταβλητότητα των χρονοσειρών
mv: μέση τιμή των χρονοσειρών
rr: συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς-ελεγχόμενες
m: αριθμός χρονοσειρών αναφοράς
j: κατώφλι – αριθμός ταυτόχρονων αλμάτων που εμφανίζονται στις χρονοσειρές αναφοράς ώστε να θεωρηθεί ένα άλμα στατιστικά σημαντικό
ti: μήκος των χρονοσειρών
iter: αριθμός προσομοιώσεων Monte Carlo

Ως έξοδος επιστρέφεται:

alm: μητρώο αριθμού κοινών αλμάτων, διάσταση ανάλογα με αριθμό προσομοιώσεων

III) Συνάρτηση CheckDMass1:

Εφαρμόζει τη Διπλή αθροιστική καμπύλη σε χρονοσειρές αναζητώντας τη σημαντικότερη ανομογένεια με χρήση μίας χρονοσειράς αναφοράς ή πολλαπλών αρχικών χρονοσειρών αναφοράς από τις οποίες συντίθεται μία τελική. Επιστρέφει τις κλίσεις των ευθειών της Διπλής αθροιστικής καμπύλης και τη θέση της σημαντικότερης ανομογένειας.

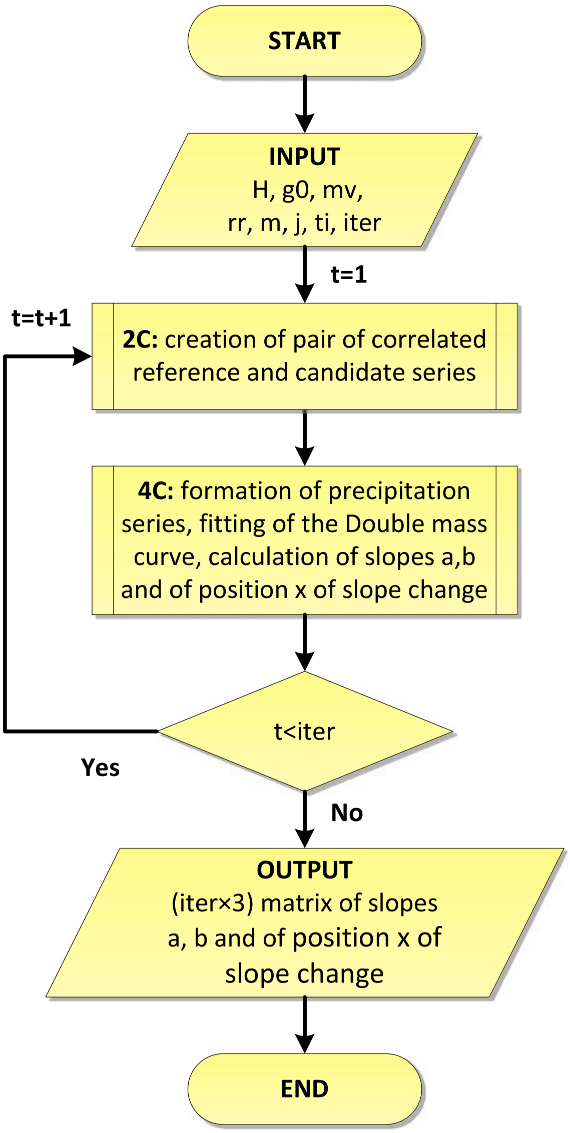
Η συνάρτηση CheckDMass1 λαμβάνει ως εισόδους τις μεταβλητές:

H: επιθυμητός συντελεστής Hurst των χρονοσειρών
g0: μεταβλητότητα των χρονοσειρών
mv: μέση τιμή των χρονοσειρών
rr: συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς-ελεγχόμενες
m: αριθμός χρονοσειρών αναφοράς
ti: μήκος των χρονοσειρών
iter: αριθμός προσομοιώσεων Monte Carlo

Ως έξοδοι επιστρέφονται:

a: Κλίση της πρώτης ευθείας της Διπλής αθροιστικής καμπύλης
b: Κλίση της δεύτερης ευθείας της Διπλής αθροιστικής καμπύλης
x: Θέση της χρονοσειράς στην οποία γίνεται αλλαγή κλίσης της Διπλής αθροιστικής καμπύλης

Το διάγραμμα ροής της συνάρτησης CheckDMass1 παρουσιάζεται στο Σχήμα 7.3.



Σχήμα 7.3. *Διάγραμμα ροής συνάρτησης CheckDMass1, η οποία εφαρμόζει τη μέθοδο SNHT σε συνθετικές χρονοσειρές αναζητώντας πολλαπλά άλματα με χρήση πολλαπλών χρονοσειρών αναφοράς*

7.2. Βοηθητικές συναρτήσεις

Όπως αναφέρθηκε και παραπάνω, αναπτύχθηκαν 11 βοηθητικές συναρτήσεις για επιμέρους διαδικασίες των κύριων συναρτήσεων που περιγράφτηκαν στην Παράγραφο 7.1, ενώ ως βοηθητική χρησιμοποιήθηκε και συνάρτηση που προτάθηκε από τους Tyralis and Koutsoyiannis (2011).

Οι βοηθητικές συναρτήσεις χωρίστηκαν με βάση τη διαδικασία που επιτελούν σε 7 κατηγορίες από τις οποίες προέκυψε και ο συμβολισμός της καθεμίας. Το σύνολο των βοηθητικών συναρτήσεων που χρησιμοποιήθηκαν με τους συμβολισμούς τους παρουσιάζεται στον Πίνακα 7.2 και αναλύεται παρακάτω.

Η συνάρτηση 6, η οποία εκτιμά το συντελεστή Hurst των χρονοσειρών, δεν παρουσιάζεται εδώ, καθώς αναλύεται από τους Tyralis and Koutsoyiannis (2011). Όσον αφορά τη συνάρτηση 7, ο τρόπος επιλογής κρίσιμων τιμών για το SNHT και παράθεση των τιμών αυτών γίνεται στο Παράρτημα Α, οπότε δεν κρίθηκε απαραίτητο να γίνει ανάλυση και σε αυτή την Παράγραφο.

Πίνακας 7.2. *Βοηθητικές συναρτήσεις που χρησιμοποιήθηκαν με τους αντίστοιχους συμβολισμούς τους*

Κατηγορία	**Περιγραφή κατηγορίας**	**Συμβολισμός**	**Συνάρτηση**
1	Δημιουργία ανεξάρτητων χρονοσειρών με μακροπρόθεσμη εμμονή	1	hurst_3AR1
2	Δημιουργία συσχετισμένων χρονοσειρών αναφοράς – ελεγχόμενων	2A	series_ref_cand
		2B	creation_series_ref_cand
		2C	series_ref_cand_DMass
3	Εφαρμογή μεθόδου ομογενοποίησης για εντοπισμό μιας ανομογένειας	3A	SNHT_breaking_pos
		3B	piecewise_linear
		3C	DMC_mult_ref_cand
4	Εφαρμογή μεθόδου ομογενοποίησης για εντοπισμό πολλαπλών ανομογενειών	4A	multiple_shifts2
		4B	multiple_shifts_mref2
5	Διόρθωση χρονοσειρών	5	SCor
6	Εκτίμηση συντελεστή Hurst	6	lsv
7	Επιλογή κρίσιμων τιμών για το SNHT	7	Tcritical

1) Συνάρτηση hurst_3AR1

Δημιουργεί μία συνθετική χρονοσειρά με επιθυμητό συντελεστή Hurst.

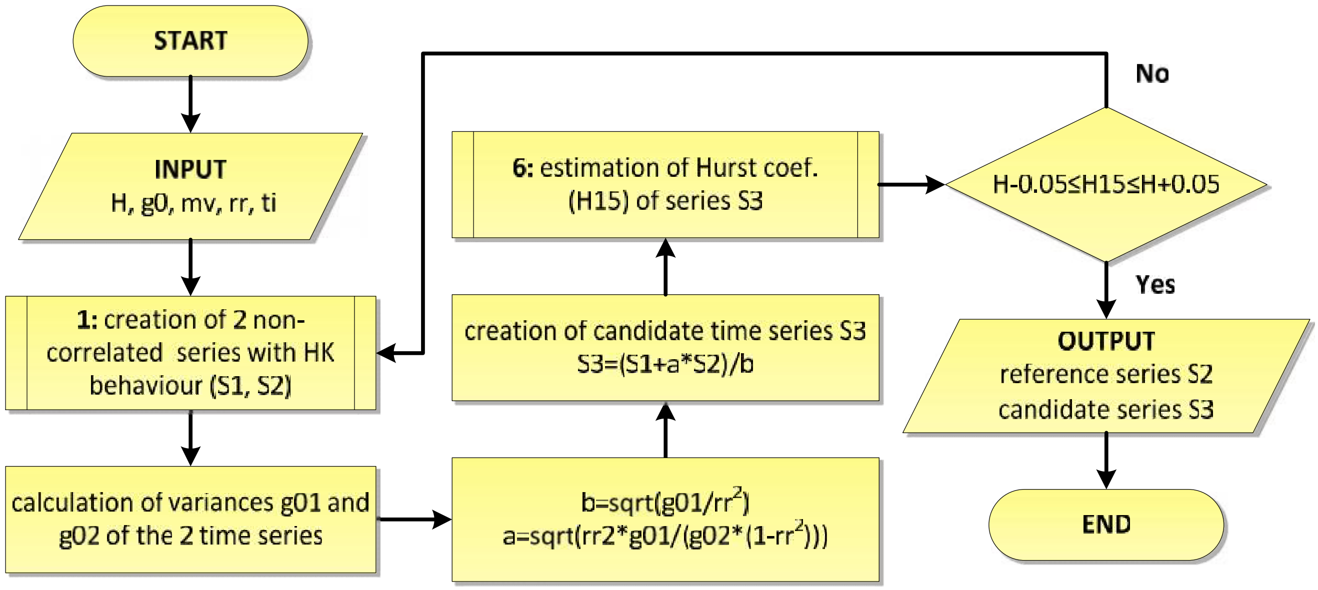
Η συνάρτηση series ref_cand λαμβάνει ως εισόδους τις μεταβλητές:
H: επιθυμητός συντελεστής Hurst της χρονοσειράς
g0: μεταβλητότητα της χρονοσειράς
mv: μέση τιμή της χρονοσειράς
ti: μήκος της χρονοσειράς

Ως έξοδος επιστρέφεται:
XT: συνθετική χρονοσειρά με επιθυμητό συντελεστή Hurst

Αναλυτική περιγραφή της διαδικασίας που ακολουθεί η συνάρτηση παρουσιάζεται στην Παράγραφο 6.4.1, οπότε δεν κρίθηκε απαραίτητο να παρατεθούν εδώ περισσότερα στοιχεία.

2A) Συνάρτηση series_ref_cand:

Δημιουργεί ένα ζεύγος χρονοσειρών αναφοράς - ελεγχόμενης, συσχετισμένων μεταξύ τους. Το διάγραμμα ροής της παρουσιάζεται στο Σχήμα 7.4.



Σχήμα 7.4. *Διάγραμμα ροής συνάρτησης series_ref_cand, η οποία δημιουργεί ένα ζεύγος συσχετισμένων χρονοσειρών αναφοράς- ελεγχόμενης*

Η συνάρτηση series_ref_cand λαμβάνει ως εισόδους τις μεταβλητές:

H: επιθυμητός συντελεστής Hurst των χρονοσειρών
g0: μεταβλητότητα των χρονοσειρών
mv: μέση τιμή των χρονοσειρών
rr: συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς-ελεγχόμενες
ti: μήκος των χρονοσειρών

Ως έξοδοι επιστρέφονται:

S2: χρονοσειρά αναφοράς
S3: ελεγχόμενη χρονοσειρά

2B) <u>Συνάρτηση creation_series_ref_cand</u>:

Δημιουργεί ένα μητρώο (m+1) συσχετισμένων μεταξύ τους χρονοσειρών, m χρονοσειρές αναφοράς και 1 ελεγχόμενη χρονοσειρά.

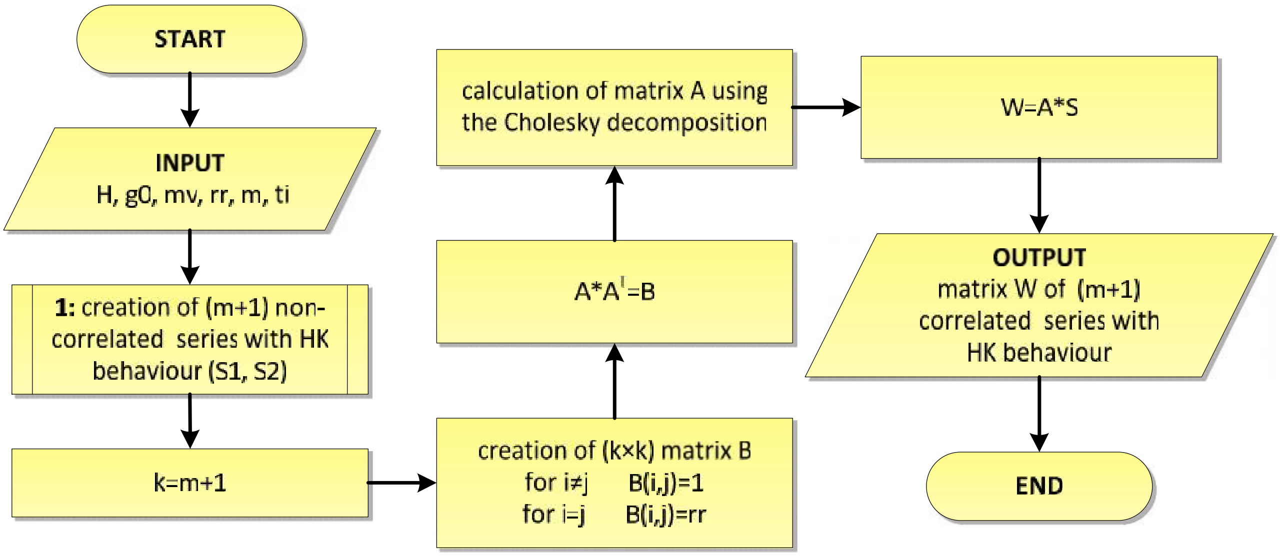
Η συνάρτηση λαμβάνει ως εισόδους τις μεταβλητές:

H: επιθυμητός συντελεστής Hurst των χρονοσειρών
g0: μεταβλητότητα των χρονοσειρών
mv: μέση τιμή των χρονοσειρών
rr: συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς-ελεγχόμενες
m: αριθμός χρονοσειρών αναφοράς
ti: μήκος των χρονοσειρών

Ως έξοδος επιστρέφεται:

W: μητρώο (m+1) συσχετισμένων μεταξύ τους χρονοσειρών

Το διάγραμμα ροής της συνάρτησης creation_series_ref_cand παρουσιάζεται στο Σχήμα 7.5.



Σχήμα 7.5. *Διάγραμμα ροής συνάρτησης creation_series_ref_cand, η οποία δημιουργεί (m+1) συσχετισμένες μεταξύ τους χρονοσειρές αναφοράς - ελεγχόμενη*

2C) Συνάρτηση series_ref_cand_DMass:

Δημιουργεί ένα ζεύγος χρονοσειρών αναφοράς - ελεγχόμενης, συσχετισμένων μεταξύ τους. Ο τρόπος παραγωγής της τελικής χρονοσειράς αναφοράς εξαρτάται από το πλήθος των αρχικών χρονοσειρών αναφοράς (είσοδος). Αν επιλεγεί μία αρχική χρονοσειρά αναφοράς είναι και η τελική. Αν επιλεγούν περισσότερες χρονοσειρές αναφοράς, η τελική χρονοσειρά προκύπτει ως μέση τιμή των χρονοσειρών αυτών.

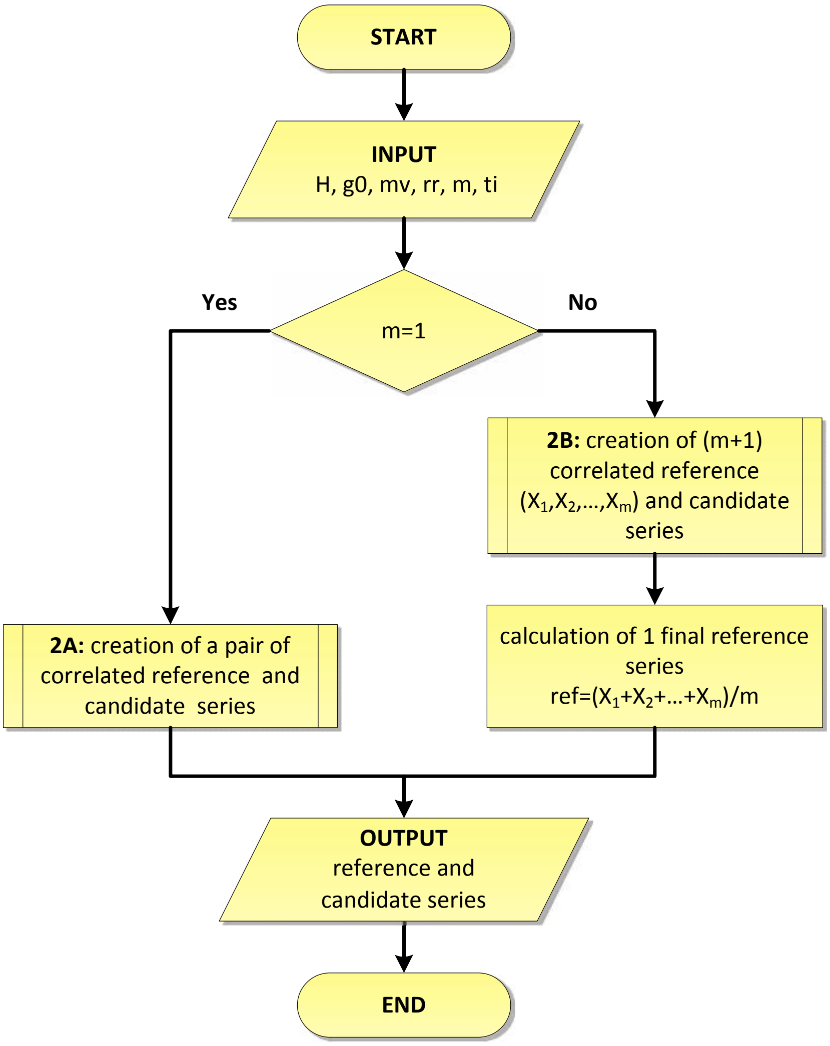
Το διάγραμμα ροής της συνάρτησης series_ref_cand_DMass παρουσιάζεται στο Σχήμα 7.6.

Η συνάρτηση λαμβάνει ως εισόδους τις μεταβλητές:

H: επιθυμητός συντελεστής Hurst των χρονοσειρών
g0: μεταβλητότητα των χρονοσειρών
mv: μέση τιμή των χρονοσειρών
rr: συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς-ελεγχόμενες
m: αριθμός χρονοσειρών αναφοράς
ti: μήκος των χρονοσειρών

Ως έξοδοι επιστρέφονται:

ref: χρονοσειρά αναφοράς
Y: ελεγχόμενη χρονοσειρά



Σχήμα 7.6. *Διάγραμμα ροής συνάρτησης series_ref_cand_DMass, η οποία δημιουργεί ένα ζεύγος συσχετισμένων χρονοσειρών αναφοράς - ελεγχόμενης*

3A) <u>Συνάρτηση SNHT_breaking_pos</u>:

Εφαρμόζει σε ζεύγος χρονοσειρών αναφοράς – ελεγχόμενης τη μέθοδο SNHT for single shifts, λαμβάνοντας ως κρίσιμες τιμές τις τιμές που παρατίθενται στο Παράρτημα Α. Τυχόν εντοπισθέν άλμα στις πρώτες 5 ή 5 τελευταίες τιμές της ελεγχόμενης χρονοσειράς δεν θεωρείται στατιστικά σημαντικό. Στην περίπτωση που η υπορουτίνα δεν εντοπίζει στατιστικά σημαντικά άλματα επιστρέφει τιμές (-999.9, -999.9).

Αναλυτική περιγραφή της μεθόδου γίνεται στις Παραγράφους 3.2.2 και 6.5.1, οπότε αναφέρουμε εδώ μόνο τις εισόδους και εξόδους της υπορουτίνας.

Ως εισόδους λαμβάνει τις μεταβλητές:

X: χρονοσειρά αναφοράς

Y: ελεγχόμενη χρονοσειρά ίδιου μήκους με τη χρονοσειρά αναφοράς

Ως έξοδοι επιστρέφονται:

x0: θέση σημαντικότερου στατιστικά σημαντικού άλματος στη χρονοσειρά

adj: διόρθωση άλματος

3B) Συνάρτηση piecewise_linear:

Προσαρμόζει σε δύο χρονοσειρές που αποτελούνται από τα συσσωρευτικά αθροίσματα των χρονοσειρών αναφοράς και ελεγχόμενης τις δύο ευθείες που αποτελούν τη Διπλή αθροιστική καμπύλη. Αυτό γίνεται ελαχιστοποιώντας το άθροισμα των τετραγώνων των αποστάσεων ανάμεσα στις ευθείες που προσαρμόζονται και τα σημεία που προκύπτουν από τα συσσωρευτικά αθροίσματα. Το σημείο αλλαγής κλίσης θεωρείται γνωστό.

Αναλυτική περιγραφή της μεθόδου γίνεται στις Παραγράφους 3.1.1 και 6.6.1, οπότε αναφέρουμε εδώ μόνο τις εισόδους και εξόδους της συνάρτησης.

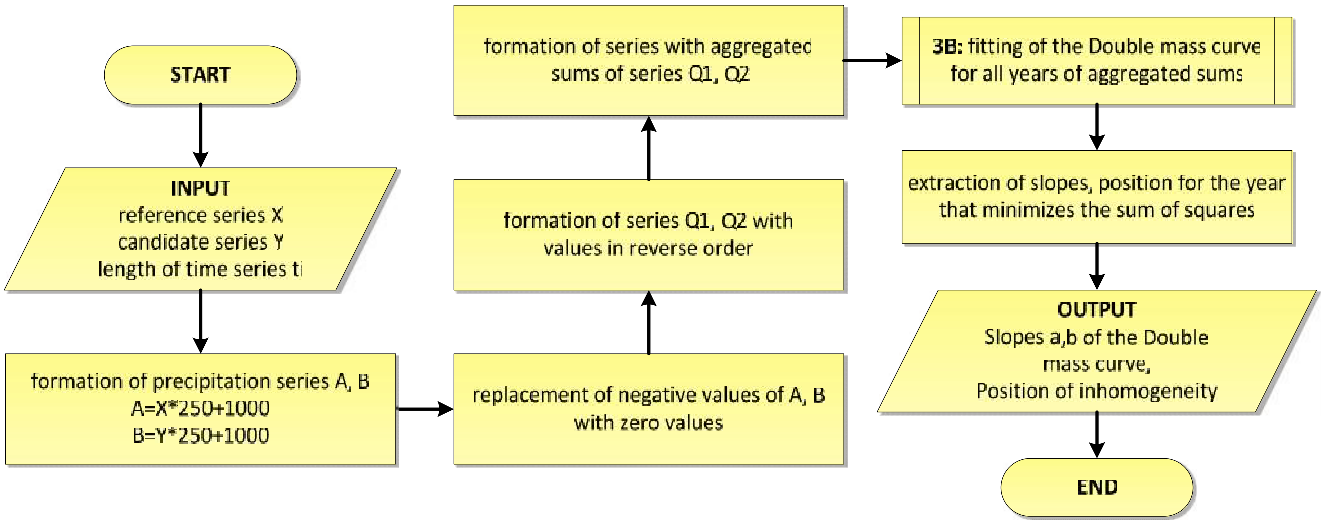
Ως εισόδους λαμβάνει τις μεταβλητές:
S1: χρονοσειρά συσσωρετικών αθροισμάτων της χρονοσειράς αναφοράς
S2: χρονοσειρά συσσωρετικών αθροισμάτων της ελεγχόμενης χρονοσειράς

Ως έξοδοι επιστρέφονται:
a: κλίση της πρώτης ευθείας της Διπλής αθροιστικής καμπύλης
b: κλίση της δεύτερης ευθείας της Διπλής αθροιστικής καμπύλης
SRes: άθροισμα τετραγώνων των αποκλίσεων

3C) Συνάρτηση DMCmult_ref_cand:

Από τις αρχικές χρονοσειρές αναφοράς – ελεγχόμενη με μέση τιμή μηδέν και μοναδιαία τυπική απόκλιση, δημιουργεί χρονοσειρές βροχόπτωσης και προσαρμόζει τη Διπλή αθροιστική καμπύλη.

Το διάγραμμα ροής της συνάρτησης DMCmult_ref_cand παρουσιάζεται στο Σχήμα 7.7.



Σχήμα 7.7. *Διάγραμμα ροής συνάρτησης DMCmult_ref_cand, η οποία δημιουργεί χρονοσειρές βροχόπτωσης και τους προσαρμόζει τη Διπλή αθροιστική καμπύλη*

Ως εισόδους η συνάρτηση DMCmult_ref_cand λαμβάνει τις μεταβλητές:
X: χρονοσειρά αναφοράς με μέση τιμή μηδέν και μοναδιαία τυπική απόκλιση
Y: ελεγχόμενη χρονοσειρά με μέση τιμή μηδέν και μοναδιαία τυπική απόκλιση

Ως έξοδοι επιστρέφονται:

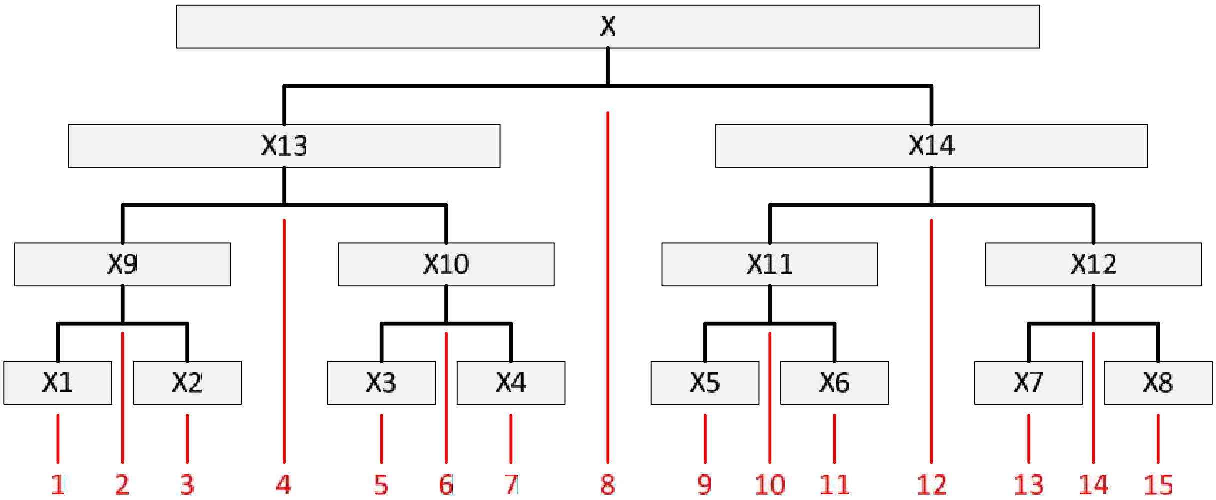
a: κλίση της πρώτης ευθείας της Διπλής αθροιστικής καμπύλης

b: κλίση της δεύτερης ευθείας της Διπλής αθροιστικής καμπύλης

x: θέση σημαντικότερης ανομογένειας (στην αρχική χρονοσειρά)

4Α) Συνάρτηση multiple_shifts2:

Η συνάρτηση αυτή εφαρμόζει τη μέθοδο SNHT σε ελεγχόμενες χρονοσειρές συγκρίνοντάς τις με χρονοσειρές αναφοράς. Κάθε φορά που εντοπίζεται στατιστικά σημαντικό άλμα οι χρονοσειρά αναφοράς – ελεγχόμενη χωρίζονται σε δύο τμήματα με σημείο κοπής το σημείο της ανομογένειας. Τα τμήματα που προκύπτουν επανελέγχονται για στατιστικά σημαντικές ανομογένειες. Συνολικά ελέγχονται μέχρι 15 ανομογένειες σε κάθε χρονοσειρά, όπως φαίνεται και στο Σχήμα 7.8. Από τη διαδικασία αυτή οι χρονοσειρές χωρίζονται σε τμήματα, με μέγιστο αριθμό των τμημάτων αυτών τα 8.



Σχήμα 7.8. *Συμβολική παράσταση της διαδικασίας που εφαρμόζει η συνάρτηση multiple_shifts2. Το SNHT εφαρμόζεται σε χρονοσειρές σπάζοντάς τις σε επιμέρους τμήματα κάθε φορά που εντοπίζεται σημαντικό άλμα. Τα επιμέρους τμήματα ξαναελέγχονται για ανομογένειες. Με κόκκινο χρώμα φαίνονται τα 15 πιθανά άλματα που αναζητά η συνάρτηση. Η αρίθμηση των αλμάτων γίνεται με βάση την απόστασή τους από την αρχή της χρονοσειράς. Εδώ παρουσιάζεται η διαδικασία διαίρεσης της χρονοσειράς αναφοράς X. Ίδια διαδικασία με σπάσιμο στα ίδια σημεία ακολουθείται και για την ελεγχόμενη χρονοσειρά Y.*

Ως εισόδους η συνάρτηση multiple_shifts2 λαμβάνει τις μεταβλητές:

X: χρονοσειρά αναφοράς

Y: ελεγχόμενη χρονοσειρά

Ως έξοδος επιστρέφεται:

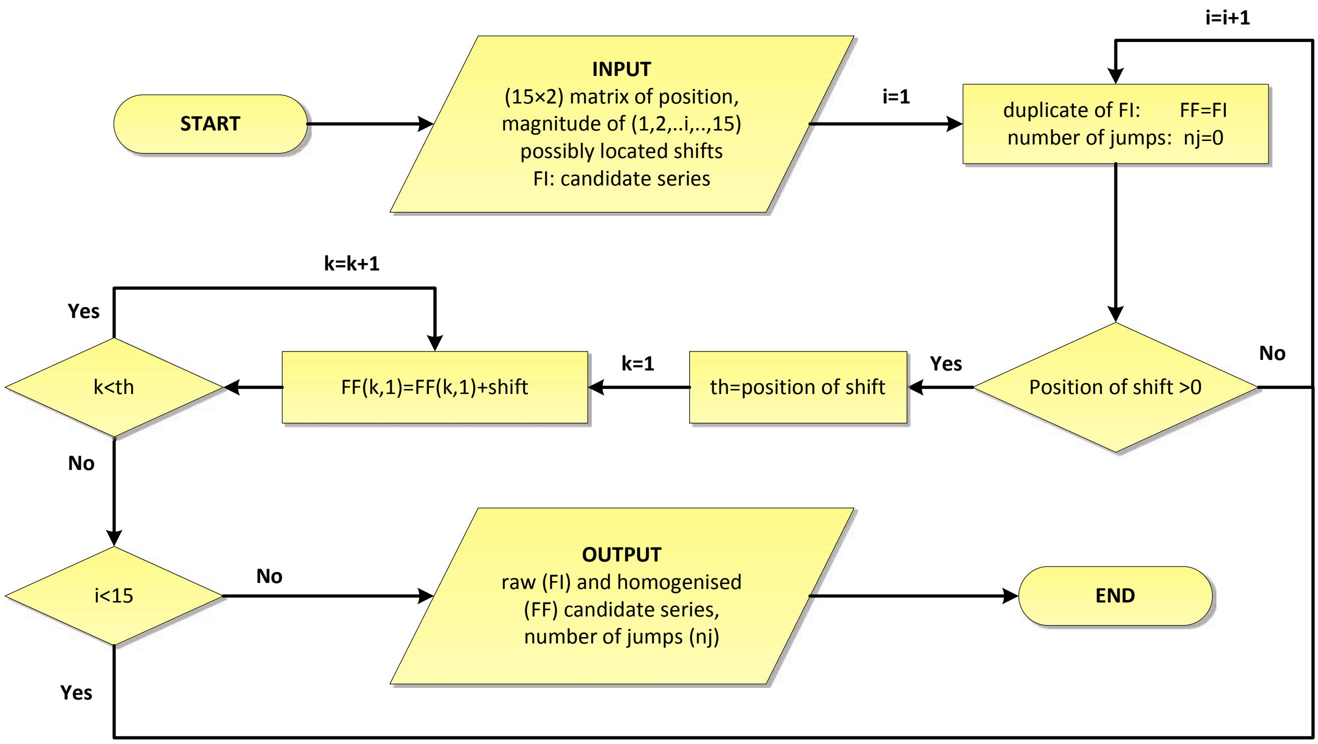
Uresult: μητρώο (15×2) με τη θέση και τη διόρθωση που αντιστοιχεί στα 15 πιθανά άλματα που αναζητά η συνάρτηση

4Β) Συνάρτηση multiple_shifts_mref2:

Ταυτίζεται με τη συνάρτηση multiple_shifts2 σε όλα εκτός από τις εξόδους. Μετά τη δημιουργία του μητρώου Uresult με τα πιθανά άλματα, αποδίδει στα έτη της ελεγχόμενης χρονοσειράς τα στατιστικά άλματα που εντόπισε με ένδειξη 77, αν τους αντιστοιχεί θετική διόρθωση και 88, αν τους αντιστοιχεί αρνητική. Στα υπόλοιπα έτη αποδίδεται η τιμή μηδέν.

5) Συνάρτηση SCor

Εφαρμόζει διορθώσεις για όλα τα στατιστικά σημαντικά άλματα που έχουν εντοπιστεί σε μια χρονοσειρά. Το διάγραμμα ροής της συνάρτησης παρουσιάζεται στο Σχήμα 7.9.



Σχήμα 7.9. *Διάγραμμα ροής συνάρτησης Scor, η οποία διορθώνει έως και 15 πιθανά άλματα σε μια ελεγχόμενη χρονοσειρά*

Ως εισόδους η συνάρτηση SCor λαμβάνει τις μεταβλητές:

InSeries: μητρώο που αποτελείται από τη χρονοσειρά αναφοράς και την ελεγχόμενη

Uresult: μητρώο (15×2) με τη θέση και τη διόρθωση που αντιστοιχεί σε 15 πιθανά άλματα

Ως έξοδοι επιστρέφονται:

FI: ελεγχόμενη χρονοσειρά πριν την ομογενοποίηση

FF: ελεγχόμενη χρονοσειρά μετά την ομογενοποίηση

Njumps: αριθμός αλμάτων που διορθώθηκαν

8. ΑΠΟΤΕΛΕΣΜΑΤΑ ΟΜΟΓΕΝΟΠΟΙΗΣΗΣ ΣΕ ΔΕΔΟΜΕΝΑ ΘΕΡΜΟΚΡΑΣΙΑΣ

Τα αποτελέσματα της εφαρμογής της μεθοδολογίας που εκτέθηκε στο Κεφάλαιο 6 και ειδικότερα η διερεύνηση εάν συνθετικές χρονοσειρές θερμοκρασίας που δεν περιέχουν ανομογένειες εμφανίζουν εικονικές με την εφαρμογή ομογενοποίησης και για διαφορετικούς συνδυασμούς παραμέτρων (στατιστικών χαρακτηριστικών των χρονοσειρών, αριθμού χρονοσειρών αναφοράς κ.ά.) αναλύονται στις επόμενες Παραγράφους. Για την ομογενοποίηση των συνθετικών δεδομένων θερμοκρασίας χρησιμοποιήθηκε η μέθοδος SNHT. Η διερεύνηση έγινε σε 3 ενότητες.

Η μέθοδος SNHT εφαρμόστηκε καταρχάς στην ενότητα 1 εξετάζοντας μόνο τις βασικές παραμέτρους H και ρ_{XY}. Στη συνέχεια, στην ενότητα 2 εξετάστηκαν συμπληρωματικά και οι δευτερεύουσες παράμετροι (μήκος χρονοσειρών και ελάχιστη απόσταση ανάμεσα σε δύο ανομογένειες ή ανάμεσα σε μια ανομογένεια και την άκρη της χρονοσειράς). Τέλος, στην ενότητα 3 έγινε εφαρμογή των διορθώσεων που υπολογίστηκαν στην ενότητα 1 για το σύστημα χρονοσειρών αναφοράς 1/1 και υπολογισμός της τάσης και του συντελεστή Hurst των χρονοσειρών πριν και μετά τις διορθώσεις.

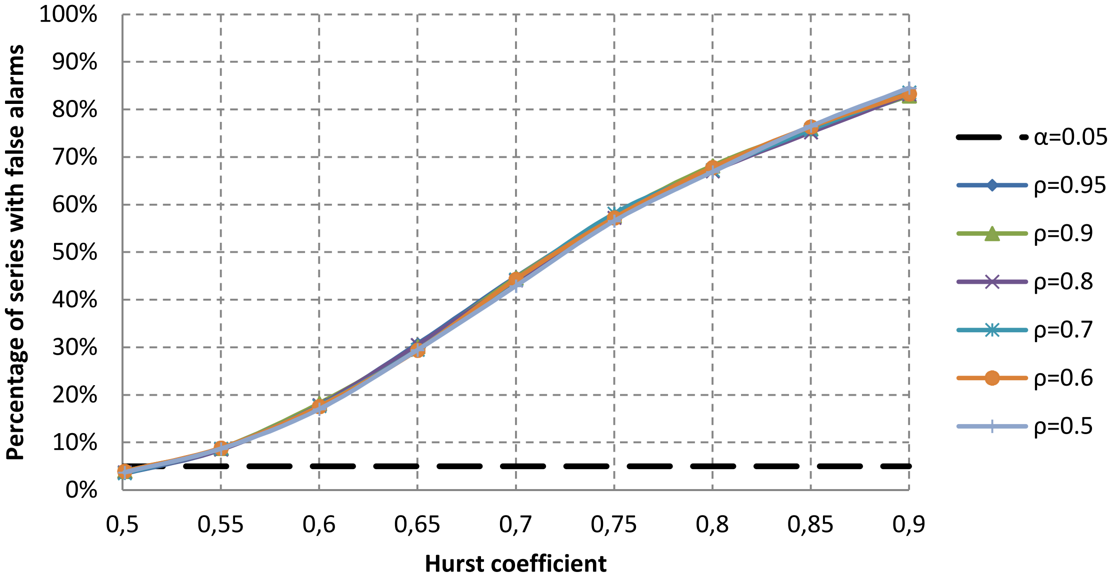
Αναλυτικά τα αποτελέσματα για όλα τα επιλεγμένα συστήματα χρονοσειρών αναφοράς (βλ. Πίνακα 6.2) και τη διόρθωση των χρονοσειρών σε κάποιες περιπτώσεις και σύγκριση των αποτελεσμάτων παρουσιάζονται στις Παραγράφους 8.1-8.10.

8.1. Αποτελέσματα συστήματος 1/1

Διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT με 1 χρονοσειρά αναφοράς και για όλους τους δυνατούς συνδυασμούς του συντελεστή Hurst και του συντελεστή ετεροσυσχέτισης που επιλεχτηκαν στις Παραγράφους 6.3.1 και 6.3.2, δηλαδή H={0.50, 0.55, ..., 0.90} και ρ_{XY}={0.5, 0.6, 0.7, 0.8, 0.9, 0.95}. Συνολικά εξετάστηκαν 9 × 6 = 54 περιπτώσεις. Για καθεμία από τις περιπτώσεις αυτές έγιναν 10000 προσομοιώσεις Monte Carlo και υπολογίστηκε το ποσοστό των χρονοσειρών στο οποίο η μέθοδος SNHT εντοπίζει πλασματικά άλματα. Τα αποτελέσματα αυτά συνοψίζονται στον Πίνακα 8.1 και το Σχήμα 8.1.

Πίνακας 8.1. *Ποσοστό χρονοσειρών με πλασματικά άλματα για 54 συνδυασμούς συντελεστή Hurst και συντελεστή ετεροσυσχέτισης με χρήση μίας χρονοσειράς αναφοράς (σύστημα χρονοσειρών αναφοράς 1/1). Με σκίαση σημειώνονται τα αποτελέσματα για ρ_{XY}=0.8, μια τυπική τιμή ετεροσυσχέτισης θεωρούμενη ως αντιπροσωπευτική.*

ρ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.95	3.5%	8.5%	17.8%	30.7%	44.8%	57.5%	68.0%	75.8%	82.9%
0.9	4.0%	8.8%	18.2%	30.2%	44.6%	57.7%	68.2%	76.0%	82.9%
0.8	3.7%	8.5%	17.9%	30.5%	44.1%	57.2%	67.0%	75.2%	83.2%
0.7	3.6%	8.7%	17.7%	29.5%	44.3%	58.1%	67.3%	75.8%	83.6%
0.6	4.0%	8.8%	17.6%	29.4%	44.3%	57.2%	67.8%	76.4%	83.3%
0.5	3.7%	8.7%	17.0%	29.4%	42.9%	56.6%	66.9%	76.5%	84.5%



Σχήμα 8.1. *Σχηματική απεικόνιση των ποσοστών χρονοσειρών με πλασματικά άλματα για χρήση μιας χρονοσειράς αναφοράς του Πίνακα 8.1. Τα αποτελέσματα για όλους τους συντελεστές ετεροσυσχέτισης ταυτίζονται σχεδόν μεταξύ τους.*

Ένα πρώτο συμπέρασμα που προκύπτει από τον Πίνακα 8.1 είναι ότι το ποσοστό χρονοσειρών με ψευδή άλματα δεν επηρεάζεται σχεδόν καθόλου από την ετεροσυσχέτιση ανάμεσα στη χρονοσειρά αναφοράς και την ελεγχόμενη. Η παρατήρηση αυτή φαίνεται περισσότερο στο Σχήμα 8.1 όπου οι καμπύλες για διαφορετικούς συντελεστές ετεροσυσχέτισης πρακτικά ταυτίζονται μεταξύ τους.

Ένα δεύτερο συμπέρασμα που προκύπτει από τον Πίνακα 8.1 αφορά τη δυναμική Hurst-Kolmogorov των χρονοσειρών. Σημαντικά ποσοστά χρονοσειρών με ψευδή άλματα (άνω του 8%) εντοπίζονται για όλες τις τιμές Hurst με εξαίρεση την περίπτωση H=0.5, που αντιστοιχεί σε δεδομένα χωρίς αυτοσυσχέτιση (λευκός θόρυβος).

Αυτό είναι αναμενόμενο γιατί η μέθοδος SNHT θεωρεί ότι ομογενή δεδομένα είναι τα ανεξάρτητα δεδομένα που ακολουθούν κατανομή N(0, 1). Αντίθετα, σε όλες τις άλλες τιμές της παραμέτρου H που αντιστοιχούν σε δεδομένα με αυτοσυσχέτιση, παρατηρούμε ότι η μέθοδος εκλαμβάνει μεταβολές τιμών των ομογενών χρονοσειρών ως ανομογένειες. Το ποσοστό των χρονοσειρών με ψευδή άλματα αυξάνεται σχεδόν γραμμικά με την αύξηση του συντελεστή Hurst, κάτι λογικό αφού αύξηση του συντελεστή Hurst σημαίνει α) αύξηση της απόστασης του μοντέλου των δεδομένων από το μοντέλο των ανεξάρτητων δεδομένων και β) δεδομένα που τείνουν να δημιουργούν συσσωματώματα θετικών και αρνητικών τιμών εφόσον η μέση τιμή στα δεδομένα που εξετάζουμε είναι μηδέν. Τα συσσωματώματα αυτά η μέθοδος τα ερμηνεύει ως άλματα (ανομογένειες).

Ενα σημείο που πρέπει να τονιστεί αφορά το ποσοστό χρονοσειρών με ψευδή άλματα για την περιοχή H=0.6-0.7, στην οποία θεωρείται από τους Rust *et al*. (2008) ότι κινούνται τα υδροκλιματικά δεδομένα. Όπως προκύπτει από τον Πίνακα 8.1., στην περιοχή αυτή εμφανίζονται ποσοστά ψευδούς εντοπισμού 17-45% (Πίνακας 8.1), δηλαδή πολύ μεγαλύτερα από το ανεκτό επίπεδο 5%.

8.2. Αποτελέσματα συστήματος 3/7

Διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος σε 3 από τις 7 χρονοσειρές και για όλους τους δυνατούς συνδυασμούς του συντελεστή Hurst και του συντελεστή ετεροσυσχέτισης που επιλεχτηκαν στις Παραγράφους 6.3.1 και 6.3.2, δηλαδή H={0.50, 0.55, ..., 0.90} και ρ_{XY}={0.5, 0.6, 0.7, 0.8, 0.9, 0.95}. Συνολικά εξετάστηκαν 9 × 6 = 54 περιπτώσεις.

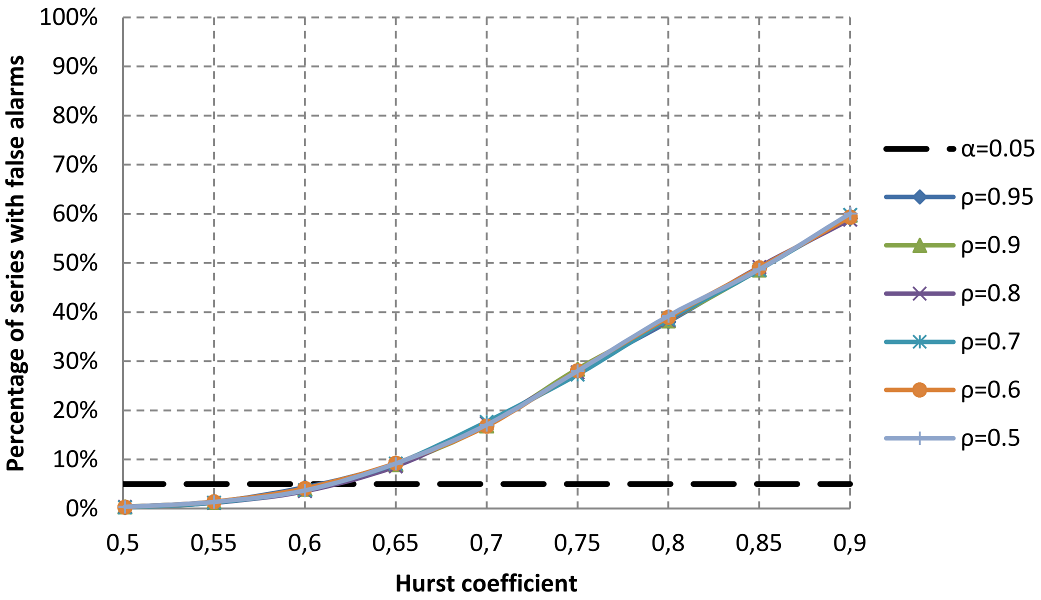
Τα ποσοστά χρονοσειρών με ψευδή άλματα για καθεμία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 54 περιπτώσεις συνοψίζονται στον Πίνακα 8.2 και το Σχήμα 8.2.

Πίνακας 8.2. *Ποσοστό χρονοσειρών με πλασματικά άλματα για 54 συνδυασμούς συντελεστή Hurst και συντελεστή ετεροσυσχέτισης με χρήση επτά χρονοσειρών αναφοράς και θεώρηση άλματος όταν αυτό είναι κοινό σε τρεις χρονοσειρές (σύστημα χρονοσειρών αναφοράς 3/7). Με σκίαση σημειώνονται τα αποτελέσματα για* ρ_{XY}*=0.8, μια τυπική τιμή ετεροσυσχέτισης θεωρούμενη ως αντιπροσωπευτική*

ρ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.95	0.3%	1.4%	4.4%	9.1%	16.9%	28.0%	38.1%	48.9%	59.1%
0.9	0.4%	1.4%	4.0%	9.0%	16.9%	28.4%	38.3%	48.6%	59.8%
0.8	0.4%	1.2%	3.6%	8.7%	17.4%	27.8%	38.4%	49.2%	58.9%
0.7	0.3%	1.3%	3.8%	9.2%	17.7%	27.3%	38.6%	48.6%	59.8%
0.6	0.3%	1.4%	4.2%	9.3%	16.8%	28.1%	39.0%	49.1%	59.4%
0.5	0.4%	1.4%	3.7%	9.2%	17.1%	28.0%	39.3%	48.7%	60.1%

Από την εξέταση του Πίνακα 8.2 προκύπτει ότι, όπως και για το σύστημα 1/1, το ποσοστό χρονοσειρών με ψευδή άλματα δεν φαίνεται να επηρεάζεται από την ετεροσυσχέτιση ανάμεσα στη χρονοσειρά αναφοράς και την ελεγχόμενη. Όπως δε φαίνεται και στο Σχήμα 8.2, οι καμπύλες που προέκυψαν για τους συντελεστές ετεροσυσχέτισης που επιλέχτηκαν πρακτικά ταυτίζονται μεταξύ τους.

Επομένως, εφόσον δεν φαίνεται να υπάρχει γενικά μια συστηματική σχέση ανάμεσα στον συντελεστή ετεροσχέτισης και το ποσοστό χρονοσειρών με ψευδή άλματα, στις επόμενες εφαρμογές της ομογενοποίησης, καθώς και στη διερεύνηση της Διπλής αθροιστικής καμπύλης η παράμετρος αυτή διατηρήθηκε σταθερή και ίση με 0.8. Η τιμή αυτή,όπως αναφέρεται και στην Παράγραφο 6.6.3, δίνεται στη βιβλιογραφία ως μια τυπική τιμή για το συντελεστή συσχέτισης ανάμεσα σε χρονοσειρές αναφοράς και ελεγχόμενες (π.χ. Ducré-Robitaille *et al*., 2003, DeGaetano, 2006, Vincent, 1998, Easterling and Peterson, 1995).



Σχήμα 8.2. *Σχηματική απεικόνιση των ποσοστών χρονοσειρών με πλασματικά άλματα του Πίνακα 8.2 για χρήση επτά χρονοσειρών αναφοράς και θεώρηση άλματος όταν αυτό είναι κοινό σε τρεις χρονοσειρές. Τα αποτελέσματα για όλους τους συντελεστές ετεροσυσχέτισης ταυτίζονται σχεδόν μεταξύ τους.*

Όσον αφορά τη δυναμική Hurst-Kolmogorov των χρονοσειρών, με χρήση του συστήματος εξισώσεων 3/7, η μέθοδος SNHT φαίνεται να έχει καλή συμπεριφορά για H=0.5-0.6, δηλαδή για δεδομένα ανεξάρτητα ή με σχετικά ασθενή αυτοσυσχέτιση. Όμως για H > 0.6 το ποσοστό των χρονοσειρών με ψευδή άλματα αυξάνεται απότομα και φτάνει την (πολύ μεγάλη) τιμή 60% για H=0.9.

Στην περιοχή H=0.6-0.7, στην οποία θεωρείται από τους Rust *et al*. (2008) ότι κινούνται τα υδροκλιματικά δεδομένα, προκύπτουν με χρήση του συστήματος 3/7 ποσοστά χρονοσειρών με ψευδείς εντοπισμούς 4-17.5% περίπου (Πίνακας 8.2).

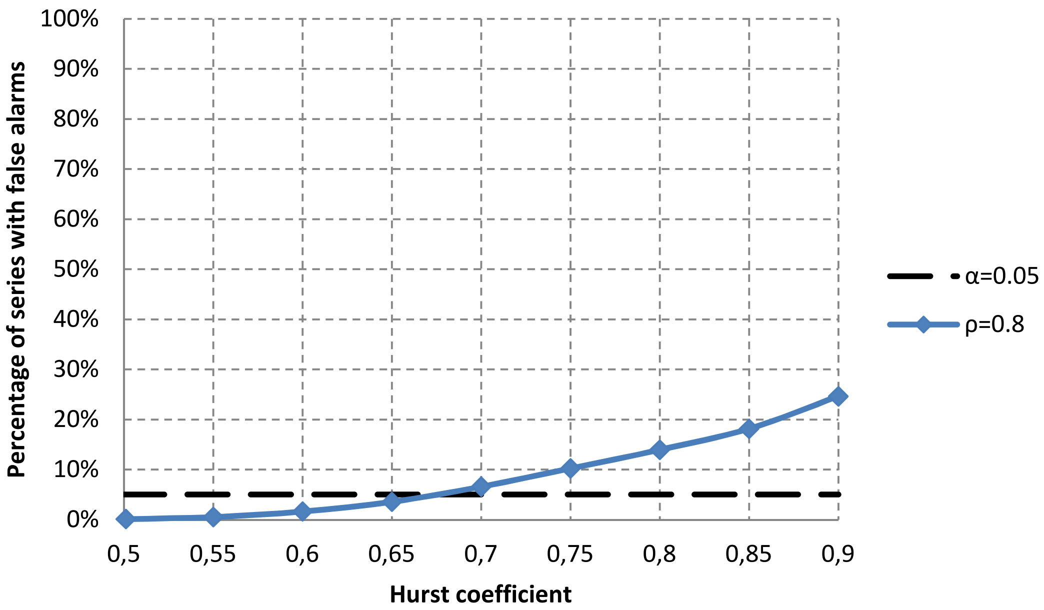
8.3. Αποτελέσματα συστήματος 2/2

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, όπως τεκμηριώθηκε πιο πάνω (Παράγραφος 8.2), διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος και στις 2 χρονοσειρές αναφοράς και για όλους τους δεκτούς συνδυασμούς του συντελεστή Hurst δηλαδή H={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν 9 περιπτώσεις.

Τα ποσοστά χρονοσειρών με ψευδή άλματα για καθεμία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 9 περιπτώσεις συνοψίζονται στον Πίνακα 8.3 και το Σχήμα 8.3.

Πίνακας 8.3. *Ποσοστό χρονοσειρών με πλασματικά άλματα για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8, με χρήση δύο χρονοσειρών αναφοράς και θεώρηση κοινών αλμάτων (σύστημα χρονοσειρών αναφοράς 2/2)*

ρ \ Η	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.8	0.1%	0.5%	1.7%	3.6%	6.6%	10.3%	14.0%	18.2%	24.6%



Σχήμα 8.3. *Σχηματική απεικόνιση των ποσοστών χρονοσειρών με πλασματικά άλματα του Πίνακα 8.3 για συντελεστή ετεροσυσχέτισης 0.8, χρήση δύο χρονοσειρών αναφοράς και θεώρηση κοινών αλμάτων.*

Η μόνη παράμετρος που μεταβάλλεται στην παραπάνω διερεύνηση είναι ο συντελεστής Hurst. Με χρήση του συστήματος εξισώσεων 2/2, η μέθοδος SNHT φαίνεται να έχει καλή συμπεριφορά για H=0.5-0.7. Για $H \geq 0.7$ το ποσοστό των χρονοσειρών με ψευδή άλματα αυξάνεται αρκετά, χωρίς όμως να ξεπεράσει στη δυσμενέστερη περίπτωση την τιμή 25%.

Στην περιοχή H=0.6-0.7, τα ποσοστά χρονοσειρών με ψευδή άλματα διατηρούνται σε ανεκτά επίπεδα σχεδόν σε όλες τις περιπτώσεις: 1.7-6.6% (Πίνακας 8.3).

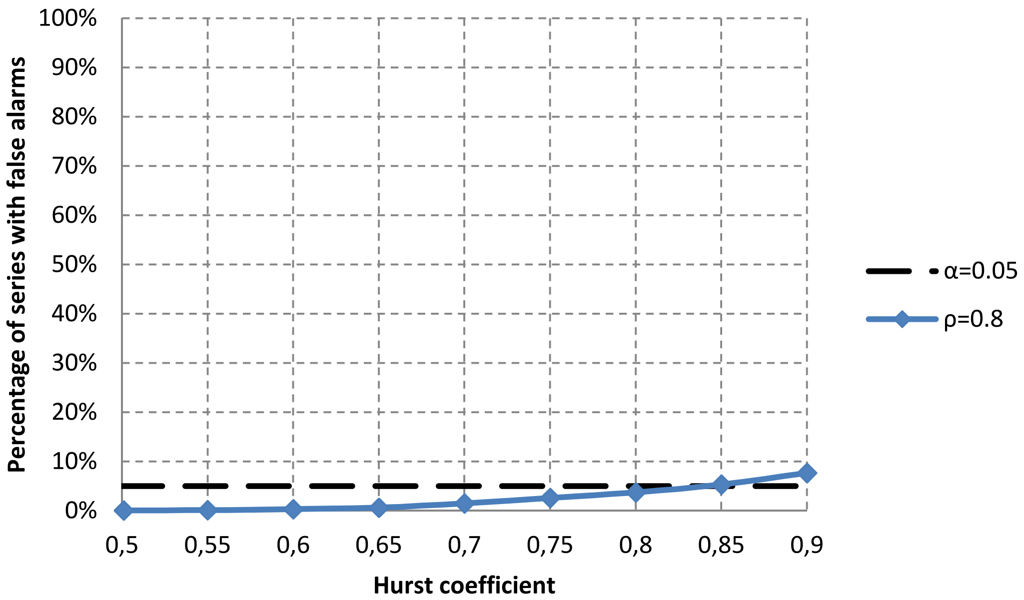
8.4. Αποτελέσματα συστήματος 3/3

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, όπως τεκμηριώθηκε πιο πάνω (Παράγραφος 8.2), διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος και στις 3 χρονοσειρές αναφοράς και για όλους τους δεκτούς συνδυασμούς του συντελεστή Hurst δηλαδή H={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν 9 περιπτώσεις.

Τα ποσοστά των χρονοσειρών με ψευδή άλματα για καθεμία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 9 περιπτώσεις συνοψίζονται στον Πίνακα 8.4 και το Σχήμα 8.4.

Πίνακας 8.4. *Ποσοστό χρονοσειρών με ψευδή εντοπισμό ανομογενειών για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8, με χρήση τριών χρονοσειρών αναφοράς και θεώρηση κοινών αλμάτων (σύστημα χρονοσειρών αναφοράς 3/3)*

ρ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.8	0.1%	0.1%	0.3%	0.6%	1.5%	2.6%	3.7%	5.3%	7.7%



Σχήμα 8.4. *Σχηματική απεικόνιση των ποσοστών χρονοσειρών με ψευδή εντοπισμό ανομογενειών του Πίνακα 8.4 για συντελεστή ετεροσυσχέτισης 0.8, χρήση τριών χρονοσειρών αναφοράς και θεώρηση κοινών αλμάτων.*

Με χρήση του συστήματος εξισώσεων 3/3, η μέθοδος SNHT φαίνεται να έχει καλή συμπεριφορά για όλο σχεδόν το εύρος των εξεταζόμενων συντελεστών Hurst. Υπέρβαση της τιμής ψευδούς εντοπισμού ανομογενειών 5% γίνεται για $H \geq 0.85$. Σε καμία από τις εξεταζόμενες περιπτώσεις το ποσοστό αυτό δεν ξεπέρασε το 10%.

Στην περιοχή H=0.6-0.7, τα ποσοστά ψευδούς εντοπισμού διατηρούνται σε πολύ χαμηλά επίπεδα: 0.3-1.5% (Πίνακας 8.4).

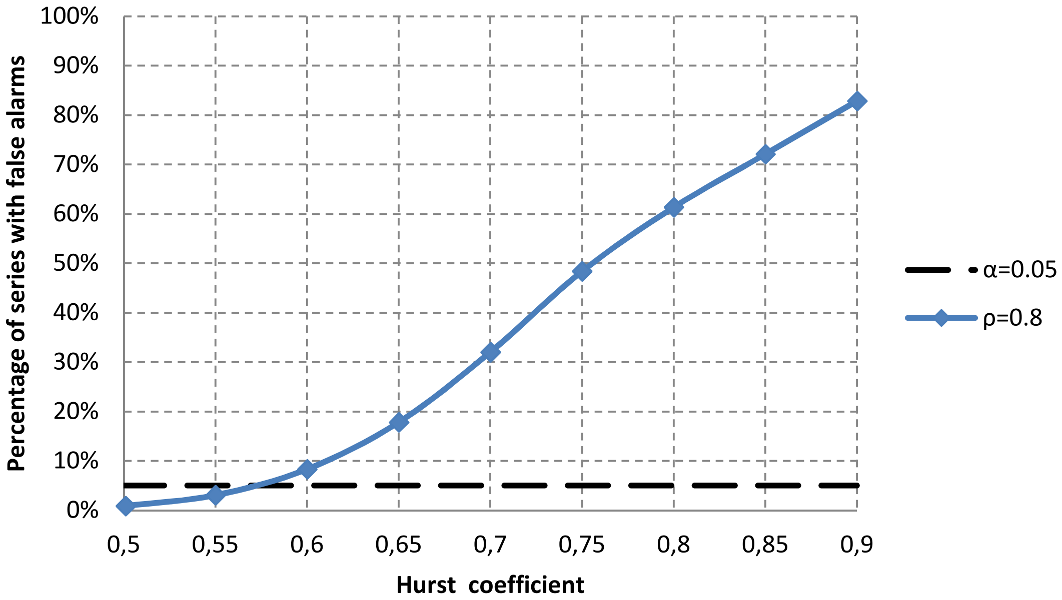
8.5. Αποτελέσματα συστήματος 3/10

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, όπως τεκμηριώθηκε πιο πάνω (Παράγραφος 8.2), διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος σε 3 από τις 10 χρονοσειρές αναφοράς και για όλους τους δεκτούς συνδυασμούς του συντελεστή Hurst δηλαδή H={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν 9 περιπτώσεις.

Τα ποσοστά χρονοσειρών με ψευδή άλματα για καθεμία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 9 περιπτώσεις συνοψίζονται στον Πίνακα 8.5 και το Σχήμα 8.5.

Πίνακας 8.5. *Ποσοστό χρονοσειρών με ψευδή άλματα για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8, με χρήση δέκα χρονοσειρών αναφοράς και θεώρηση άλματος όταν αυτό είναι κοινό σε τρεις χρονοσειρές (σύστημα χρονοσειρών αναφοράς 3/10)*

ρ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.8	0.9%	3.0%	8.3%	17.8%	32.0%	48.4%	61.3%	72.1%	82.8%



Σχήμα 8.5. *Σχηματική απεικόνιση των ποσοστών χρονοσειρών με ψευδή άλματα του Πίνακα 8.5 για συντελεστή ετεροσυσχέτισης 0.8, χρήση δέκα χρονοσειρών αναφοράς και θεώρηση άλματος όταν αυτό είναι κοινό σε τρεις χρονοσειρές.*

Με χρήση του συστήματος εξισώσεων 3/10, η μέθοδος SNHT φαίνεται να έχει καλή συμπεριφορά για H=0.5-0.55, δηλαδή για δεδομένα ανεξάρτητα ή με σχετικά ασθενή αυτοσυσχέτιση. Όμως για H > 0.55 το ποσοστό των χρονοσειρών με ψευδή άλματα αυξάνεται απότομα και φτάνει την (πολύ μεγάλη) τιμή 83% περίπου για H=0.9.

Στην περιοχή H=0.6-0.7, τα ποσοστά χρονοσειρών με ψευδείς εντοπισμούς είναι αρκετά υψηλά: 8.3-32% (Πίνακας 8.5).

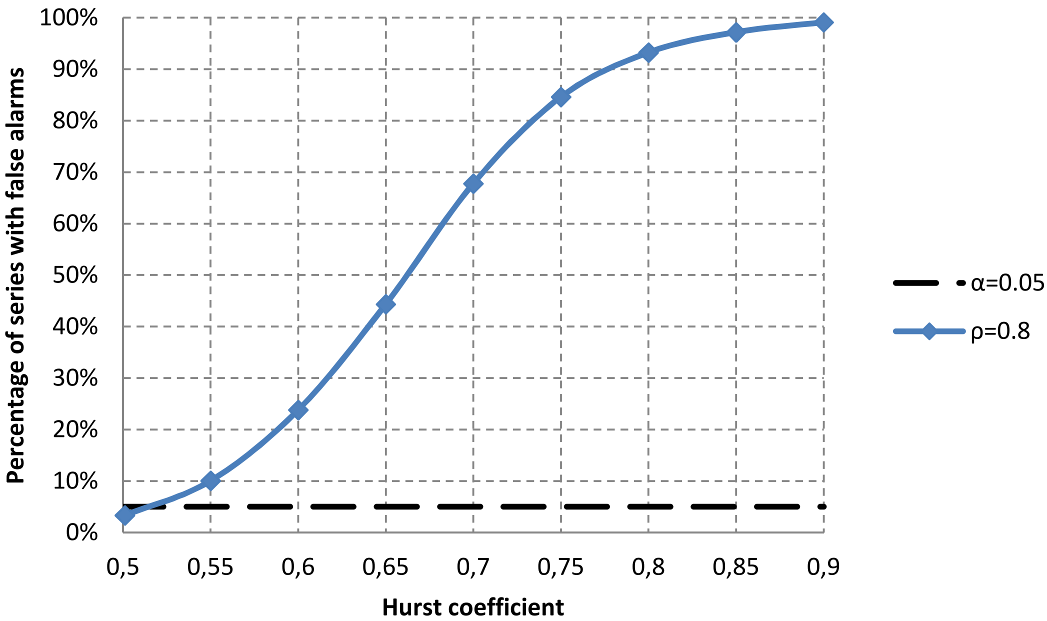
8.6. Αποτελέσματα συστήματος 3/20

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, όπως τεκμηριώθηκε πιο πάνω (Παράγραφος 8.2), διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος σε 3 από τις 20 χρονοσειρές αναφοράς και για όλους τους δεκτούς συνδυασμούς του συντελεστή Hurst δηλαδή H={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν 9 περιπτώσεις.

Τα ποσοστά χονοσειρών με ψευδή άλματα για καθεμία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 9 περιπτώσεις συνοψίζονται στον Πίνακα 8.6 και το Σχήμα 8.6.

Πίνακας 8.6. *Ποσοστό χρονοσειρών με ψευδή άλματα για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8, με χρήση είκοσι χρονοσειρών αναφοράς και θεώρηση άλματος όταν αυτό είναι κοινό σε τρεις χρονοσειρές (σύστημα χρονοσειρών αναφοράς 3/20)*

ρ \ Η	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.8	3.4%	10.1%	23.8%	44.4%	67.8%	84.6%	93.2%	97.2%	99.1%



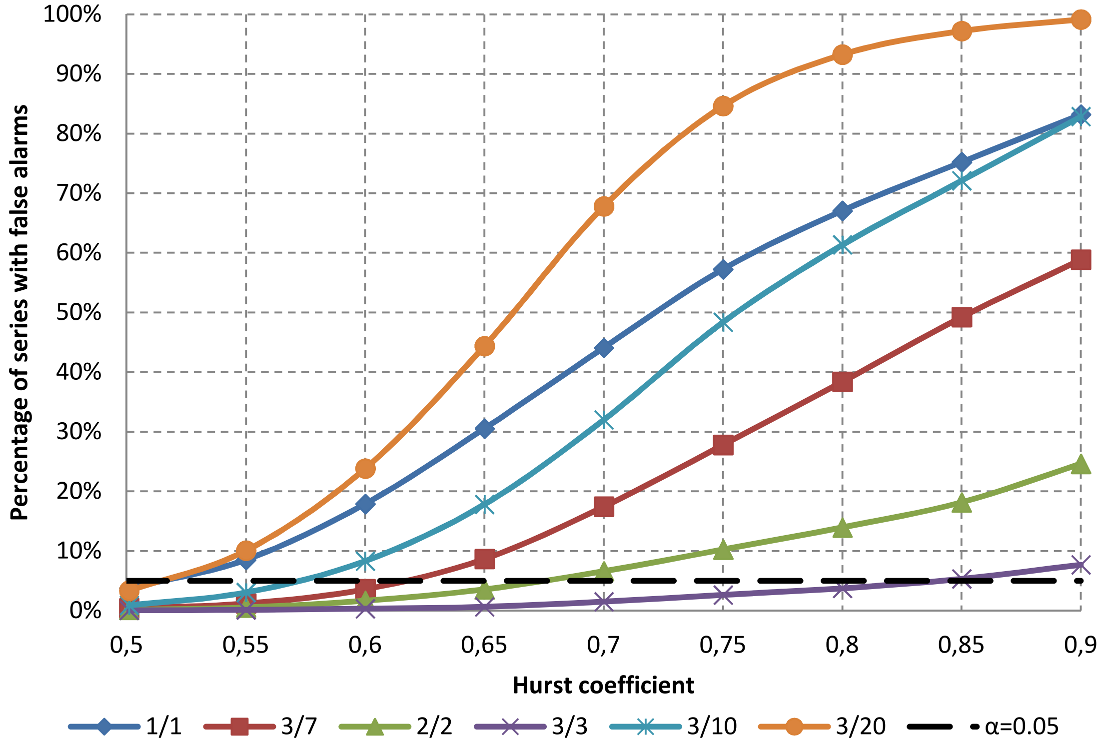
Σχήμα 8.6. *Σχηματική απεικόνιση των ποσοστών χρονοσειρών με ψευδή άλματα του Πίνακα 8.6 για συντελεστή ετεροσυσχέτισης 0.8, χρήση είκοσι χρονοσειρών αναφοράς και θεώρηση άλματος όταν αυτό είναι κοινό σε τρεις χρονοσειρές.*

Με χρήση του συστήματος εξισώσεων 3/20, η μέθοδος SNHT έχει καλή συμπεριφορά μόνο για Η=0.5, δηλαδή για δεδομένα ανεξάρτητα. Σε όλες τις περιπτώσεις δεδομένων με αυτοσυσχέτιση (Η>0.5) το ποσοστό ψευδών εντοπισμών ανομογενειών ξεπερνάει το όριο 5%. Η αύξηση των ψευδών εντοπισμών με αύξηση του συντελεστή Hurst είναι πολύ μεγάλη. Μάλιστα στη δυσμενέστερη περίπτωση που εξετάστηκε, δηλαδή για Η=0.9, βρέθηκαν ψευδείς ανομογένειες στο σύνολο σχεδόν των χρονοσειρών (99.1%).

Στην περιοχή Η=0.6-0.7, τα ποσοστά χρονοσειρών με ψευδείς εντοπισμούς είναι πολύ υψηλά: 24-68% περίπου (Πίνακας 8.6).

8.7. Συγκριτική παρουσίαση αποτελεσμάτων διερεύνησης βασικών παραμέτρων

Στο Σχήμα 8.7 γίνεται μια συγκριτική παρουσίαση των διαγραμμάτων των ποσοστών χρονοσειρών με ψευδείς εντοπισμούς των Σχημάτων 8.1-8.6. Εφόσον δεν παρατηρήθηκαν διαφορές συναρτήσει των διαφορετικών συντελεστών ετεροσυσχέτισης που εξετάστηκαν, παρουσιάζονται εδώ μόνο τα αποτελέσματα για $ρ_{XY}$=0.8, για τα συστήματα 1/1 και 3/7.



Σχήμα 8.7. *Συγκριτική παρουσίαση γραφημάτων 7.1-7.6. Όλα τα αποτελέσματα που παρουσιάζονται είναι για συντελεστή ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και την ελεγχόμενη $ρ_{XY}$=0.8. Ανεκτό επίπεδο ψευδών εντοπισμών θεωρείται το 5%. Καλύτερη συμπεριφορά φαίνεται να έχει η μέθοδος 3/3 (εντός ορίων για τις περισσότερες τιμές Η).*

Από το Σχήμα 8.7 φαίνεται ότι ο συντελεστής Hurst επηρεάζει τον εντοπισμό ψευδών ανομογενειών από τη μέθοδο SNHT. Τιμές Η=0.5 (μηδενική υπόθεση της μεθόδου SNHT) σε όλες τις περιπτώσεις που διερευνήθηκαν οδήγησαν σε ανεκτό ποσοστό χρονοσειρών με ψευδείς εντοπισμούς (<5%), όπως ήταν και το αναμενόμενο, ενώ αντίθετα, τιμές υψηλές (Η=0.8-0.9) οδήγησαν σε όλες τις περιπτώσεις σε μεγάλα ποσοστά ψευδών εντοπισμών. Εξαίρεση αποτελεί η 3/3, η οποία δεν εμφάνισε τιμές >8%. Δεύτερη καλύτερη συμπεριφορά είχε το σύστημα 2/2, με ποσοστά εντός ορίων για Η<0.7, ακολουθούμενη από το 3/7 με ποσοστά εντός ορίων για Η<0.6.

Για χρήση περισσότερων χρονοσειρών από 7 και εντοπισμό κοινών αλμάτων σε 3 τουλάχιστον χρονοσειρές, τα ποσοστά χρονοσειρών με ψευδείς εντοπισμούς ήταν >5% για όλες τις περιπτώσεις πλην των ανεξάρτητων δεδομένων (Η=0.5). Το ίδιο ισχύει και για χρήση μίας μόνο χρονοσειράς αναφοράς. Ενδιαφέρον επίσης παρουσιάζει το γεγονός ότι για είκοσι χρονοσειρές αναφοράς (3/20) η μέθοδος δίνει πολύ μεγάλο ποσοστό ψευδούς εντοπισμού ακόμα και για πολύ ασθενή εμμονή (Η=0.55), ενώ αγγίζει το 100% για Η>0.8.

Από τις προηγούμενες παρατηρήσεις φαίνεται ότι η μέθοδος SNHT έχει καλή συμπεριφορά όσον αφορά τον εντοπισμό ψευδών ανομογενειών όταν οι ανομογένειες εντοπίζονται σε σύγκριση με όλες τις χρονοσειρές που εξετάζονται. Επιπλέον φαίνεται ότι η συνήθης χρήση τριών χρονοσειρών αναφοράς για τον εντοπισμό ανομογενειών είναι επαρκής. Σε χρονοσειρές με μεγάλες τιμές συντελεστή Hurst, θα ήταν ίσως καλύτερο να χρησιμοποιούνται για λόγους ασφάλειας τέσσερις χρονοσειρές αναφοράς.

Αρκετά διαφορετικά αποτελέσματα προκύπτουν για χρήση μιας χρονοσειράς αναφοράς, εξέταση δηλαδή του θεωρητικού μοντέλου της μεθόδου. Η μέθοδος φαίνεται ανεπαρκής για όλες τις περιπτώσεις δεδομένων με μακροπρόθεσμη εμμονή που εξετάστηκαν, καθώς για όλους τους συνδυασμούς παραμέτρων προέκυψαν ποσοστά ψευδών εντοπισμών μεγαλύτερο από τα ανεκτά. Μη ικανοποιητική είναι και η συμπεριφορά της μεθόδου όταν αναζητούνται κοινά άλματα σε κάποιες μόνο από τις υπάρχουσες χρονοσειρές αναφοράς.

Μπορούμε να παρατηρήσουμε πάντως πως ανεξάρτητα από τη θεωρητική βάση της μεθόδου, ο τρόπος που αυτή εφαρμόζεται επηρεάζει κατά πολύ τα ποσοστά ψευδών ανομογενειών που εντοπίζει.

8.8. Διερεύνηση επιρροής μήκους χρονοσειρών

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, όπως τεκμηριώθηκε πιο πάνω (Παράγραφος 8.2), διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος σε 3 από τις 7 χρονοσειρές αναφοράς, για μήκος χρονοσειρών 50 έτη και για όλους τους δεκτούς συνδυασμούς του συντελεστή Hurst δηλαδή H={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν 9 περιπτώσεις.

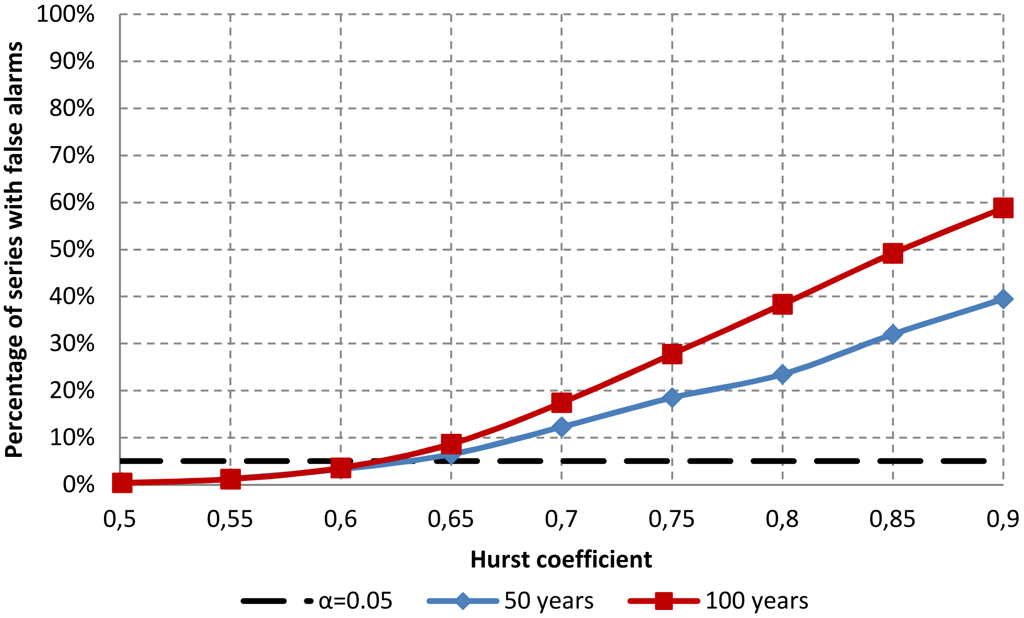
Τα ποσοστά χρονοσειρών με ψευδή άλματα για κάθε μία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 9 περιπτώσεις συνοψίζονται στον Πίνακα 8.7 και το Σχήμα 8.8. Στο Σχήμα 8.8 γίνεται μια συγκριτική παρουσίαση των αποτελεσμάτων του συστήματος 3/7-50 years με το σύστημα 3/7 (δηλαδή για μήκος χρονοσειρών 100 έτη).

Πίνακας 8.7. *Ποσοστό χρονοσειρών με ψευδή άλματα για χρονοσειρές μήκους 50 ετών, με χρήση του συστήματος εξισώσεων αναοράς 3/7, για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8. Τα αποτελέσματα συγκρίνονται με τα αντίστοιχα αποτελέσματα του Σχήματος 8.2 για χρονοσειρές μήκους 100 ετών*

ρ \ H	**0.501**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.8	0.5%	1.3%	3.3%	6.5%	12.3%	18.5%	23.5%	32.0%	39.5%

Τα αποτελέσματα για μήκη χρονοσειρών 50 και 100 έτη ταυτίζονται για H=0.5-0.6. Στην περιοχή αυτή παρατηρείται ικανοποιητική συμπεριφορά της μεθόδου SNHT. Για H>0.6 τα αποτελέσματα αρχίζουν να αποκλίνουν, με το ποσοστό χρονοσειρών με ψευδείς εντοπισμούς για μήκος χρονοσειρών 50 έτη να είναι σταθερά μικρότερο από το αντίστοιχο ποσοστό για 100 έτη και την ψαλίδα να μεγαλώνει με αύξηση του συντελεστή Hurst.

Γενικά για μήκος 50 έτη το SNHT δίνει καλύτερα αποτελέσματα, χωρίς όμως αυτά να είναι εντός των αποδεκτών ορίων.



Σχήμα 8.8. *Παρουσιάζεται το ποσοστό χρονοσειρών με ψευδή άλματα για χρονοσειρές μήκους 50 ετών, για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8 (μπλε χρώμα) με χρήση του συστήματος 3/7 σε σχέση με τα αποτελέσματα του Σχήματος 7.2 για χρονοσειρές μήκους 100 ετών (κόκκινο χρώμα). Για χρονοσειρές μικρότερου μήκους φαίνεται να μειώνεται το ποσοστό ψευδών εντοπισμών.*

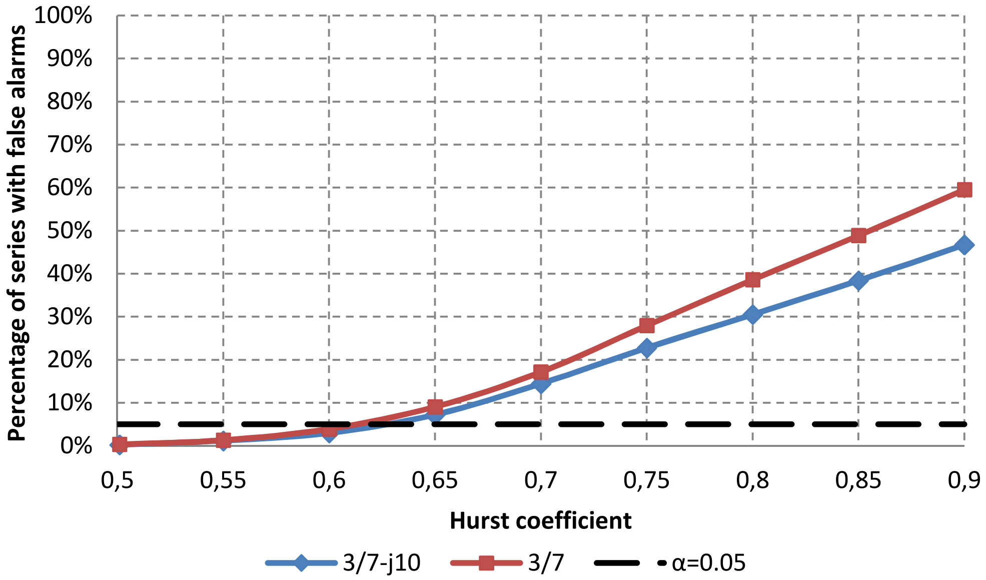
8.9. Διερεύνηση επιρροής ελάχιστης απόστασης ανομογενειών μεταξύ τους και από τα άκρα της χρονοσειράς

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, όπως τεκμηριώθηκε πιο πάνω (Παράγραφος 8.2), διερευνήθηκε ο εντοπισμός ψευδών αλμάτων σε ομογενείς χρονοσειρές με βάση την εφαρμογή της μεθόδου SNHT για εμφάνιση κοινού άλματος σε 3 από τις 7 χρονοσειρές αναφοράς, ελάχιστη απόσταση ανομογενειών μεταξύ τους και από τα άκρα της χρονοσειράς 10 έτη (αντί για 5) και για όλους τους δεκτούς συνδυασμούς του συντελεστή Hurst δηλαδή H={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν 9 περιπτώσεις.

Τα ποσοστά χρονοσειρών με ψευδή άλματα για καθεμία από τις 10000 προσομοιώσεις που αντιστοιχούν στις άνω 9 περιπτώσεις συνοψίζονται στον Πίνακα 8.8 και το Σχήμα 8.9. Στο Σχήμα 8.9 γίνεται μια συγκριτική παρουσίαση των αποτελεσμάτων του συστήματος 3/7- j10 με το σύστημα 3/7 (δηλαδή για ελάχιστη απόσταση ανομογενειών μεταξύ τους και από τα άκρα της χρονοσειράς 5 έτη).

Πίνακας 8.8. *Ποσοστό χρονοσειρών με ψευδή άλματα για χρονοσειρές μήκους 100 ετών, με χρήση του συστήματος εξισώσεων αναφοράς 3/7-j10, για 9 διαφορετικές τιμές του συντελεστή Hurst και για συντελεστή ετεροσυσχέτισης 0.8.*

ρ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
0.8	0.2%	1.2%	3.0%	7.2%	14.4%	22.7%	30.5%	38.4%	46.7%



Σχήμα 8.9. *Παρουσιάζεται το ποσοστό χρονοσειρών με ψευδή άλματα για χρονοσειρές μήκους 100 ετών, για 9 διαφορετικές τιμές του συντελεστή Hurst, για συντελεστή ετεροσυσχέτισης 0.8 και ελάχιστη απόσταση 10 έτη (μπλε χρώμα) με χρήση του συστήματος 3/7 σε σχέση με τα αποτελέσματα του Σχήματος 8.2 για χρονοσειρές μήκους 100 ετών (κόκκινο χρώμα). Για μεγαλύτερη απόσταση ανομογενειών το ποσοστό ψευδών εντοπισμών μειώνεται, όχι όμως σημαντικά.*

Το SNHT είναι μια μέθοδος που είναι ευαίσθητη σε ανομογένειες κοντά στα άκρα των χρονοσειρών (overdetection at the edges, π.χ. Wang, 2008). Λαμβάνοντας υπόψη το χαρακτηριστικό του αυτό αναμέναμε να υπάρχει αρκετά μεγάλη διαφορά στα αποτελέσματα της εφαρμογής της μεθόδου με ελάχιστη απόσταση ανομογενειών μεταξύ τους και από τα άκρα της χρονοσειράς 5 ή 10 έτη.

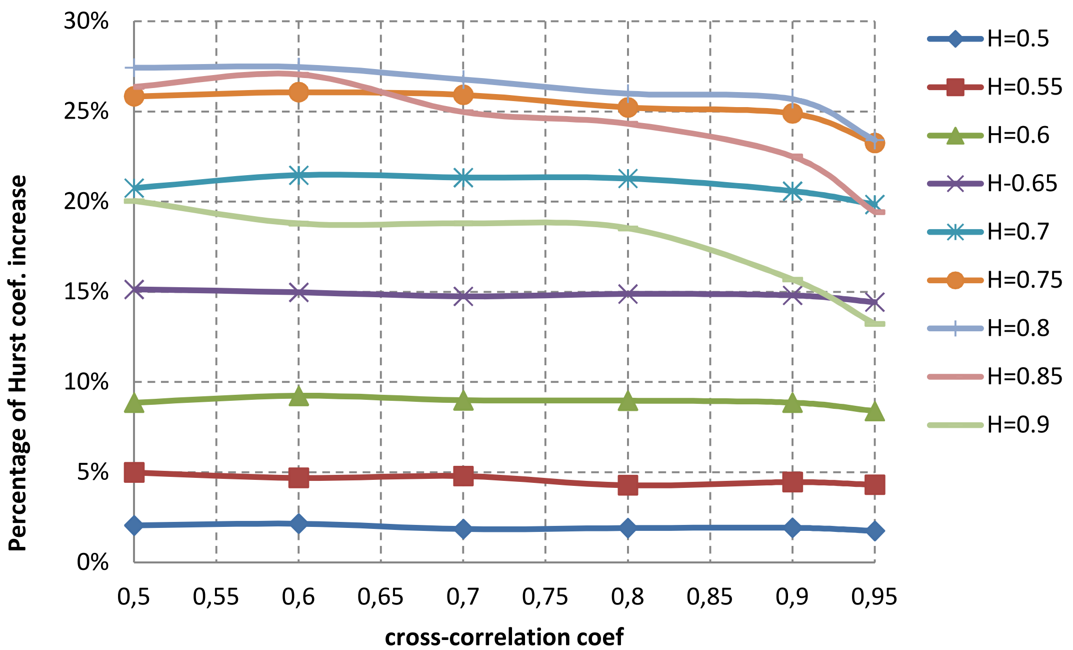
Τα αποτελέσματα για τη συντηρητικότερη εφαρμογή του SNHT είναι πράγματι ελαφρώς καλύτερα από τα αποτελέσματα της εφαρμογής της μεθόδου με θεώρηση πολύ κοντινών ανομογενειών. Η διαφορά όμως αυτή, η οποία φαίνεται στο Σχήμα 8.9 δεν είναι τόσο μεγάλη όσο αναμέναμε.

8.10. Διόρθωση χρονοσειρών

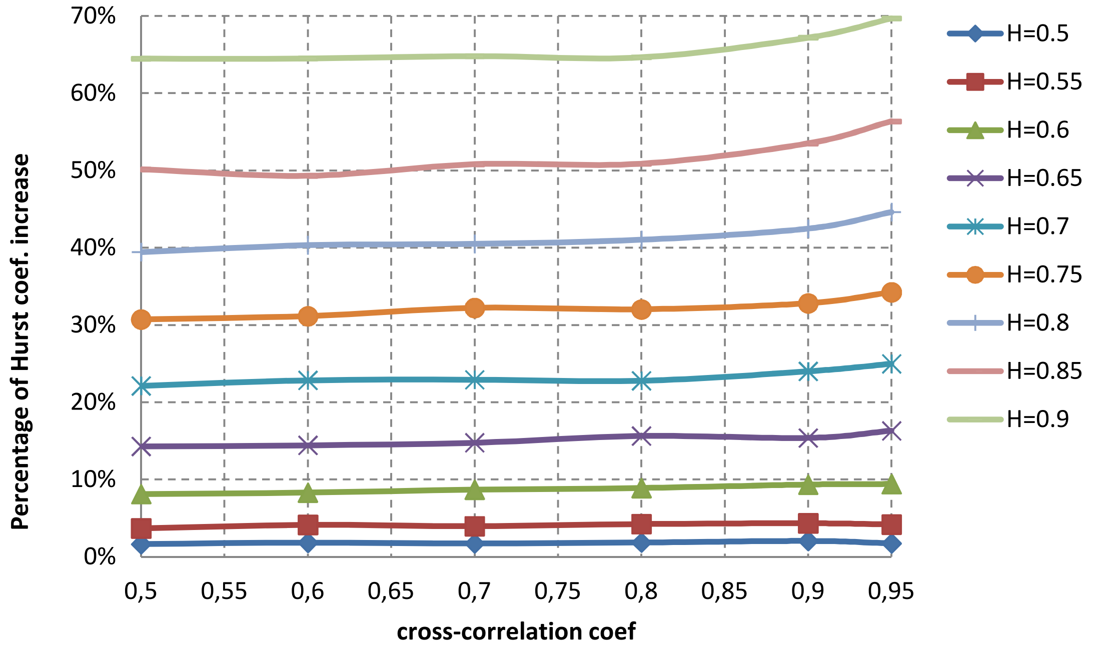
Η διόρθωση χρονοσειρών, όπως αναλύθηκε στην Παράγραφο 5.5.2, αφορά μόνο την περίπτωση του συστήματος χρονοσειρών αναφοράς 1/1, ώστε να είναι εύκολος ο υπολογισμός διορθώσεων από τις ανομογένειες που εντοπίζει το SNHT. Το SNHT εφαρμόστηκε στα ζεύγη συνθετικών χρονοσειρών αναφοράς- ελεγχόμενης, διορθώθηκαν οι ανομογένειες που εντοπίστηκαν και εξετάστηκαν η τάση και ο συντελεστής Hurst των χρονοσειρών πριν και μετά την ομογενοποίηση.

Όσον αφορά την τάση των χρονοσειρών, η μέθοδος SNHT έδωσε συμμετρικά αποτελέσματα. Δηλαδή σε όλες τις περιπτώσεις που εξετάστηκαν το ποσοστό των χρονοσειρών στις οποίες αυξήθηκε η τάση ήταν σχεδόν ίσο με το ποσοστό των χρονοσειρών στις οποίες η τάση μειώθηκε. Δεν μπορεί επομένως να βγει κάποιο απευθείας συμπέρασμα σε σχέση με τον τρόπο που η ομογενοποίηση μπορεί να επηρεάζει την εκτίμηση της μεταβολής των θερμοκρασιών παγκοσμίως.

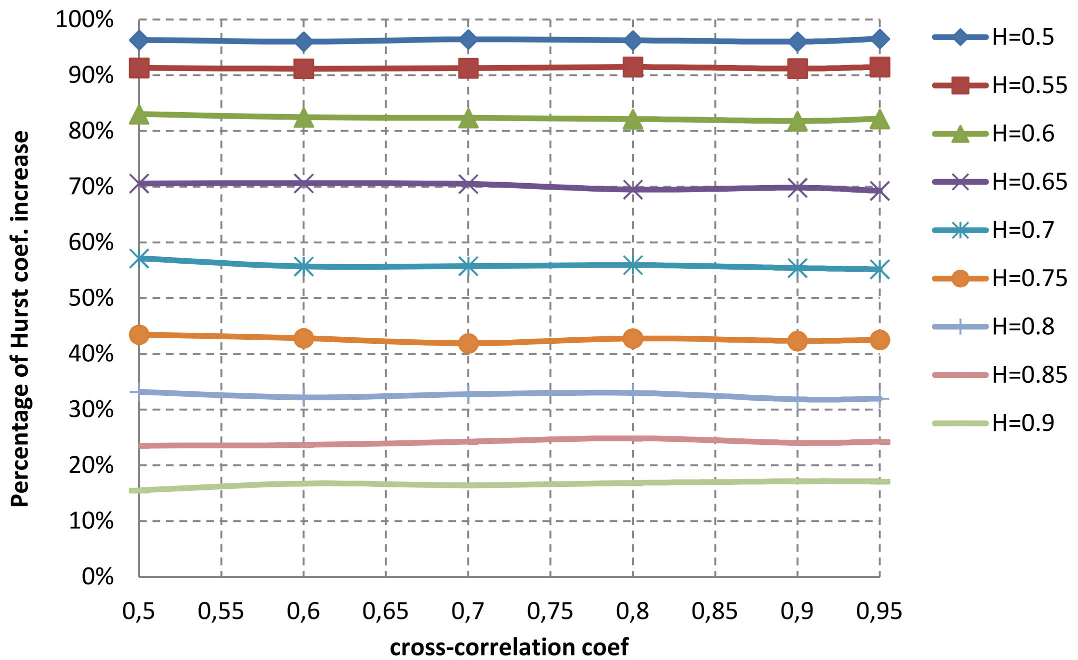
Τα αποτελέσματα σε σχέση με το συντελεστή Hurst ήταν αρκετά διαφορετικά. Το ποσοστό των χρονοσειρών στις οποίες αυξήθηκε ή μειώθηκε ο συντελεστής Hurst φαίνεται να μεταβάλλεται συναρτήσει του αρχικού συντελεστή Hurst των χρονοσειρών. Αντίθετα, δεν φαίνεται να υπάρχει κάποια επιρροή του συντελεστή ετεροσυσχέτισης. Όπως φαίνεται στα Σχήματα 8.10-8.12, το ποσοστό των χρονοσειρών στις οποίες αυξήθηκε, μειώθηκε ή έμεινε σταθερός ο συντελεστής Hurst μετά την ομογενοποίηση είναι σχεδόν σταθερό για όλους τους συντελεστές ετεροσυσχέτισης που εξετάστηκαν. Μια μικρή διαφορά, κυρίως στο ποσοστό των χρονοσειρών στις οποίες παρατηρήθηκε αύξηση του συντελεστή Hurst μπορεί να επισημανθεί για το συντελεστή ετεροσυσχέτισης ρ_{XY}=0.95.



Σχήμα 8.10. *Ποσοστό των χρονοσειρών στις οποίες αυξήθηκε ο συντελεστής Hurst μετά την ομογενοποίηση με τη μέθοδο SNHT, συναρτήσει του συντελεστή ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες πριν την ομογενοποίηση*



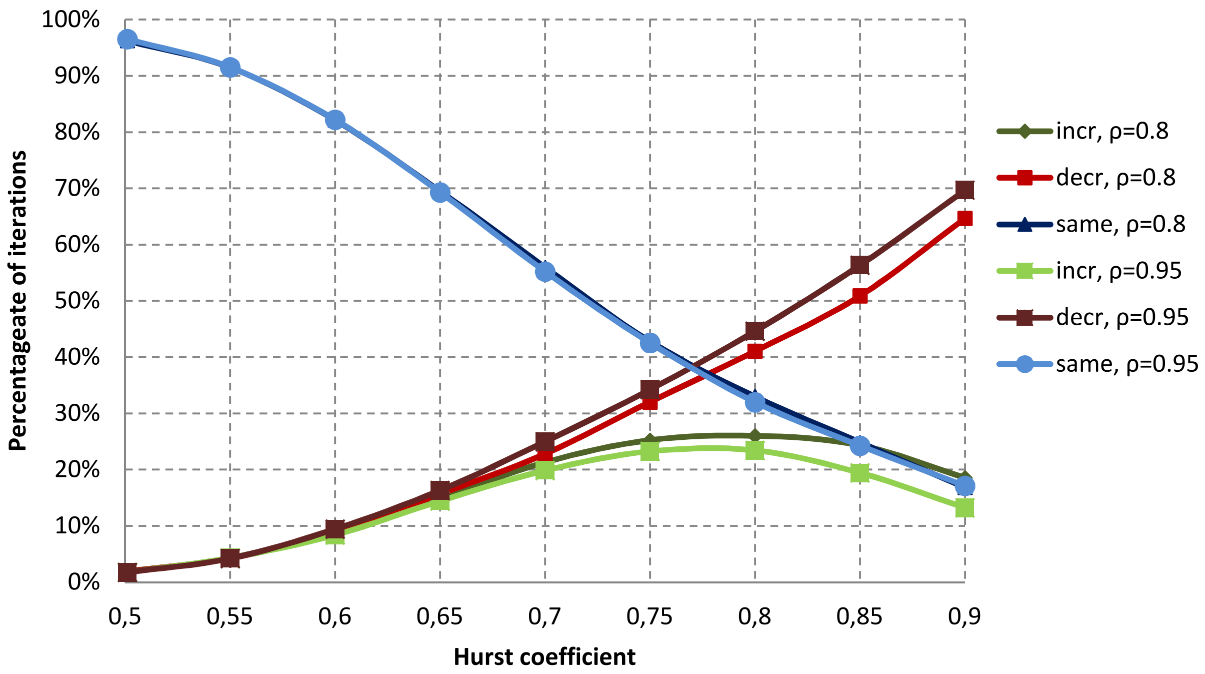
Σχήμα 8.11. *Ποσοστό των χρονοσειρών στις οποίες μειώθηκε ο συντελεστής Hurst μετά την ομογενοποίηση με τη μέθοδο SNHT, συναρτήσει του συντελεστή ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες πριν την ομογενοποίηση*



Σχήμα 8.12. *Ποσοστό των χρονοσειρών στις οποίες ο συντελεστής Hurst παρέμεινε σταθερός μετά την ομογενοποίηση με τη μέθοδο SNHT, συναρτήσει του συντελεστή ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες πριν την ομογενοποίηση*

Στο Σχήμα 8.13 φαίνονται τα ποσοστά των χρονοσειρών στις οποίες ο συντελεστής Hurst αυξήθηκε, μειώθηκε ή έμεινε σταθερός συναρτήσει του αρχικά επιλεγμένου συντελεστή Hurst. Εφόσον για ρ_{XY}=0.5-0.9 είχαμε παρόμοια αποτελέσματα όσον αφορά την επιρροή της ομογενοποίησης στο συντελεστή Hurst, στο Σχήμα 8.13 εικονίζονται δύο περιπτώσεις συντελεστών ετεροσυσχέτισης:

- η περίπτωση ρ_{XY}=0.8, η οποία επιλέχτηκε να παρουσιαστεί και στα διαγράμματα του ποσοστού ψευδών εντοπισμών ανομογενειών, και
- η περίπτωση ρ_{XY}=0.95, η οποία δίνει λίγο διαφορετικά αποτελέσματα από όλες τις υπόλοιπες.



Σχήμα 8.13. *Επιρροή της ομογενοποίησης με χρήση της μεθόδου SNHT στο συντελεστή Hurst των ελεγχόμενων χρονοσειρών*

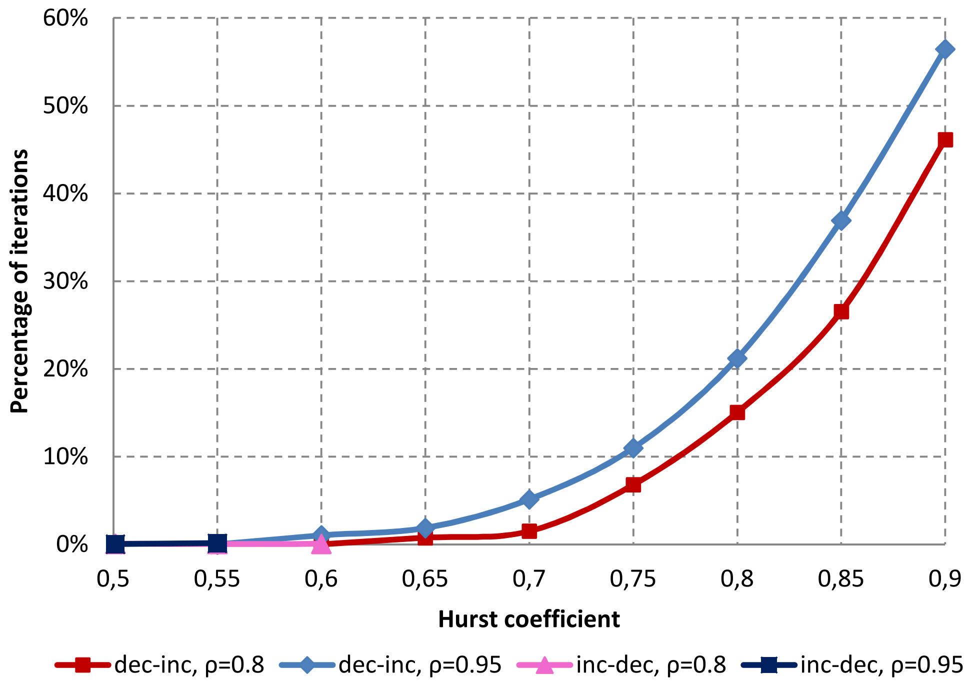
Από το Σχήμα 8.14 μπορούμε να παρατηρήσουμε τα εξής:

1. Το ποσοστό των χρονοσειρών στις οποίες ο συντελεστής Hurst παραμένει σταθερός μειώνεται με αύξηση του αρχικού συντελεστή Hurst.
2. Για σχετικά ασθενή μακροπρόθεσμη εμμονή των αρχικών χρονοσειρών (Η<0.65) το ποσοστό των χρονοσειρών στις οποίες ο συντελεστής Hurst αυξάνεται με την ομογενοποίηση είναι σχεδόν ίδιο με το ποσοστό των χρονοσειρών στις οποίες μειώνεται.
3. Για Η>0.65 το ποσοστό των χρονοσειρών στις οποίες η ομογενοποίηση μειώνει το συντελεστή Hurst ξεπερνάει το ποσοστό των χρονοσειρών στις οποίες τον αυξάνει, με την ψαλίδα να ανοίγει συνεχώς.

Η αυξανόμενη διαφορά ανάμεσα στις χρονοσειρές στις οποίες μειώνεται και αυξάνεται ο συντελεστής Hurst φαίνεται καλύτερα στο Σχήμα 8.14, στο οποίο εικονίζονται οι διαφορές των ποσοστών των δύο αυτών αποτελεσμάτων στις χρονοσειρές.

Για ασθενή μνήμη το ποσοστό των χρονοσειρών που αυξάνεται ο Hurst είναι ελαφρά μεγαλύτερο από ποσοστό των χρονοσειρών που μειώνεται γι’αυτό και εικονίζονται οι διαφορές ποσοστό αύξησης-ποσοστό μείωσης. Πρακτικά βέβαια για H<0.65, δεν φαίνεται η ομογενοποίηση να επηρεάζει την εμμονή των χρονοσειρών προς κάποια συγκεκριμένη κατεύθυνση. Αυτό μάλλον συμβασίνει γιατί τα δεδομένα είναι σχετικά κοντά στα ανεξάρτητα δεδομένα που θεωρεί το SNHT ως μηδενική υπόθεση.

Αντίθετα, για H>0.65, η εμμονή γίνεται ισχυρή και η μέθοδος τη θεωρεί σφάλμα και τη διορθώνει. Τείνει δηλαδή στις περισσότερες περιπτώσεις να φέρει τα δεδομένα πιο κοντά στα ανεξάρτητα της μηδενικής υπόθεσης.



Σχήμα 8.14. *Διαφορά των ποσοστών των χρονοσειρών στις οποίες μειώνεται και αυξάνεται ο συντελεστής Hurst μετά από ομογενοποίηση με τη μέθοδο SNHT*

9. ΑΠΟΤΕΛΕΣΜΑΤΑ ΟΜΟΓΕΝΟΠΟΙΗΣΗΣ ΣΕ ΔΕΔΟΜΕΝΑ ΒΡΟΧΟΠΤΩΣΗΣ

Τα αποτελέσματα της εφαρμογής της μεθοδολογίας που εκτέθηκε στο Κεφάλαιο 6 και ειδικότερα η διερεύνηση εάν συνθετικές χρονοσειρές βροχόπτωσης που δεν περιέχουν ανομογένειες εμφανίζουν πλασματικές με την εφαρμογή ομογενοποίησης και για διαφορετικούς συνδυασμούς παραμέτρων (στατιστικών χαρακτηριστικών των χρονοσειρών, αριθμού χρονοσειρών αναφοράς κ.ά.) αναλύονται στις επόμενες Παραγράφους. Για την ομογενοποίηση των συνθετικών δεδομένων βροχόπτωσης χρησιμοποιήθηκε η Διπλή αθροιστική καμπύλη. Η διερεύνηση έγινε σε 2 ενότητες.

Η Διπλή αθροιστική καμπύλη εφαρμόστηκε καταρχάς στην ενότητα 1 εξετάζοντας μόνο τη βασική παράμετρο Η. Ο συντελεστής ετεροσυσχέτισης ρ_{XY} θεωρήθηκε σταθερός και ίσος με 0.8, σύμφωνα με την τεκμηρίωση της Παραγράφου 8.2. Στη συνέχεια, στην ενότητα 2 έγινε μια διερεύνηση της παραμέτρου κ, η οποία ορίζεται ως η απόκλιση από τη μονάδα του λόγου των κλίσεων των δύο ευθειών που προσαρμόζονται στα δεδομένα για τη δημιουργία της Διπλής αθροιστικής καμπύλης (Σχέση 6.36).

Αναλυτικά τα αποτελέσματα της διερεύνησης των ψευδών εντοπισμών για τα δύο συστήματα αναφοράς δίνονται στις Παραγράφους 9.1 και 9.2, ενώ στην Παράγραφο 9.3 παρουσιάζονται τα αποτελέσματα από τη διερεύνηση της παραμέτρου k=ln(α/β).

9.1. Αποτελέσματα συστήματος 1/1

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, εφαρμόστηκε η Διπλή αθροιστική καμπύλη σε ζεύγος ομογενών χρονοσειρών αναφοράς – ελεγχόμενης (σύστημα 1/1) για όλες τις τιμές του συντελεστή Hurst που επιλέχτηκαν, δηλαδή Η={0.50, 0.55, ..., 0.90}. Συνολικά εξετάστηκαν δηλαδή 9 συνδυασμοί παραμέτρων, για καθέναν από τους οποίους έγιναν 2000 προσομοιώσεις.

Από τις προσομοιώσεις που έγιναν υπολογίστηκαν τα ποσοστά των χρονοσειρών με πλασματικές ανομογένειες συναρτήσει των εξεταζόμενων συντελεστών Hurst και του λόγου λ των κλίσεων των δύο ευθειών της Διπλής αθροιστικής καμπύλης. Τα αποτελέσματα αυτά συνοψίζονται στον Πίνακα 9.1 και το Σχήμα 9.1.

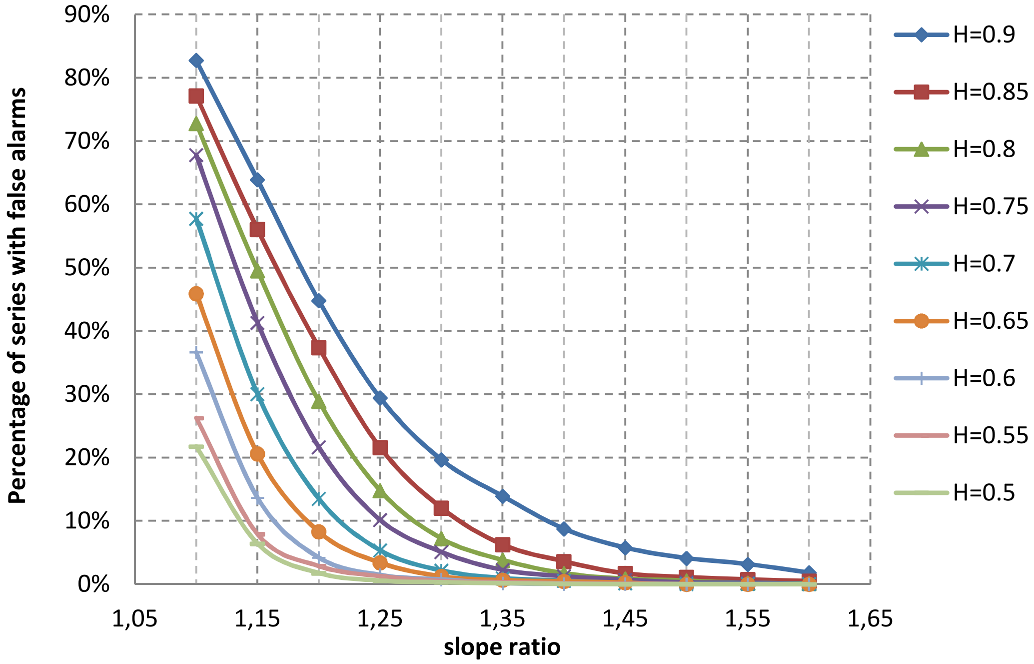
Από την εξέταση του Πίνακα 9.1 προκύπτει ότι το ποσοστό των χρονοσειρών με ψευδή άλματα αυξάνεται:

- με αύξηση του συντελεστή Hurst, και
- με μείωση του λόγου λ των κλίσεων.

Μη ανεκτά ποσοστά χρονοσειρών με ψευδείς ανομογένειες θεωρήθηκαν ποσοστά μεγαλύτερα από 5% (γραμμοσκιασμένα στον Πίνακα 9.1).

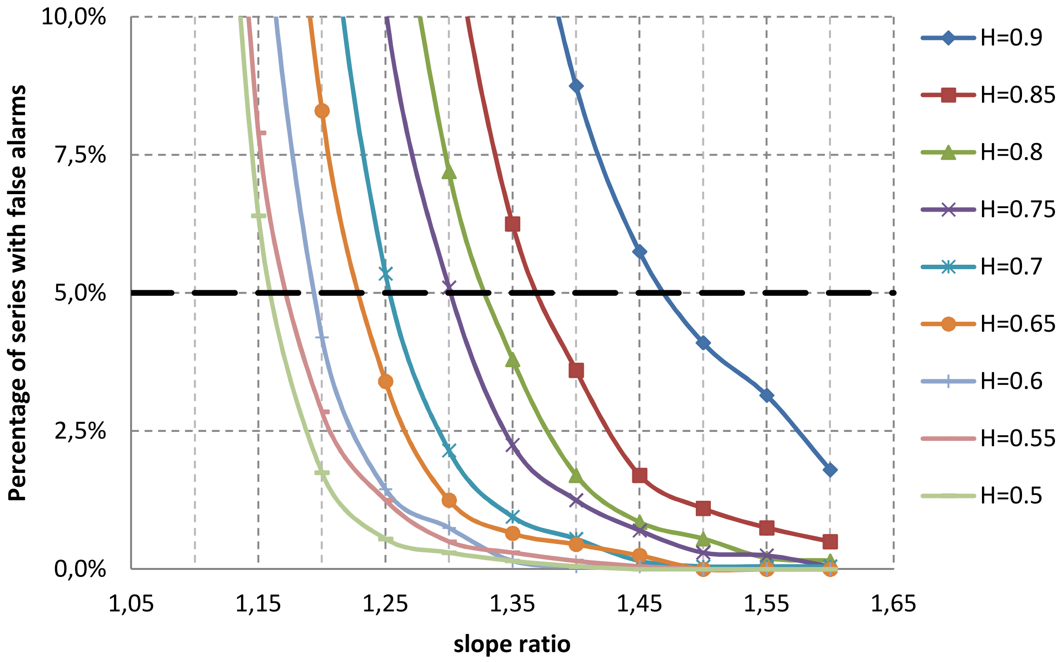
Πίνακας 9.1. *Ποσοστά χρονοσειρών στις οποίες η Διπλή αθροιστική καμπύλη εντοπίζει πλασματικές ανομογένειες με χρήση μιας χρονοσειράς αναφοράς συσχετισμένης με την ελεγχόμενη με συντελεστή ετεροσυσχέτισης ρ_{XY}=0.8. Τα αποτελέσματα παρουσιάζονται συναρτήσει του συντελεστή Hurst των χρονοσειρών και του λόγου λ των κλίσεων των δύο ευθειών της Διπλής αθροιστικής καμπύλης. Με σκίαση σημειώνονται τα ποσοστά ψευδούς εντοπισμού που είναι πάνω από το ανεκτό όριο 5%.*

λ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
1.10	21.8%	26.3%	36.7%	45.9%	57.8%	67.8%	72.8%	77.2%	82.8%
1.15	6.4%	7.9%	13.7%	20.6%	30.1%	41.3%	49.6%	56.1%	63.9%
1.20	1.8%	2.9%	4.2%	8.3%	13.5%	21.7%	28.9%	37.4%	44.8%
1.25	0.6%	1.3%	1.5%	3.4%	5.4%	10.2%	14.8%	21.6%	29.5%
1.30	0.3%	0.5%	0.8%	1.3%	2.2%	5.1%	7.2%	12.1%	19.7%
1.35	0.2%	0.3%	0.2%	0.7%	1.0%	2.3%	3.8%	6.3%	13.9%
1.40	0.1%	0.2%	0.1%	0.5%	0.6%	1.3%	1.7%	3.6%	8.8%
1.45	0.0%	0.1%	0.1%	0.3%	0.2%	0.7%	0.9%	1.7%	5.8%
1.50	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	0.6%	1.1%	4.1%
1.55	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	0.2%	0.8%	3.2%
1.60	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.5%	1.8%



Σχήμα 9.1. *Ποσοστά χρονοσειρών με πλασματικά άλματα για εφαρμογή της Διπλής αθροιστικής καμπύλης σε δεδομένα με συντελεστή ετεροσυσχέτισης ρ_{XY}=0.8, επιλεγμένους συντελεστές Hurst και χρήση του συστήματος 1/1. Φαίνεται καθαρά μια μετατόπιση των καμπυλών προς μεγαλύτερα ποσοστά ψευδών εντοπισμών για αύξηση του συντελεστή Hurst.*

Όπως φαίνεται από τον Πίνακα 9.1 και το Σχήμα 9.2, για H ≤ 0.6 απόπειρα να διορθωθούν ανομογένειες που αντιστοιχούν σε λόγο μικρότερο από 1.2, οδηγεί σε ποσοστό ψευδών εντοπισμών μεγαλύτερο από 5%. Ο λόγος που φαίνεται να μπορεί να διορθωθεί με δεδομένο μόνο την κλίση της Διπλής αθροιστικής καμπύλης αυξάνεται για αύξηση του συντελεστή Hurst. Για H=0.9, που είναι και η μεγαλύτερη τιμή του συντελεστή Hurst που εξετάστηκε, ο λόγος αυτός φτάνει σχεδόν το 1.5. Διόρθωση δηλαδή δεδομένων με λόγο κλίσεων μικρότερο από 1.5 αντιστοιχεί σε πιθανότητα λανθασμένης διόρθωσης ≥ 5%.



Σχήμα 9.2. *Λεπτομέρεια του Σχήματος 9.1. Τα ποσοστά των χρονοσειρών με πλασματικές ανομογένειες παρουσιάζονται σε σύγκριση με το ανεκτό ποσοστό 5%.*

9.2. Αποτελέσματα συστήματος 3/1

Θεωρώντας το συντελεστή ετεροσυσχέτισης σταθερό και ίσο με 0.8, εφαρμόστηκε η Διπλή αθροιστική καμπύλη σε ζεύγος ομογενών χρονοσειρών αναφοράς – ελεγχόμενης (σύστημα 1/1) για όλες τις τιμές του συντελεστή Hurst που επιλέχτηκαν, δηλαδή H={0.50, 0.55, ..., 0.90}. Η χρονοσειρά αναφοράς ήταν «τεχνητή» και προέκυψε ως σύνθεση 3 προσομοιωμένων χρονοσειρών αναφοράς. Συνολικά εξετάστηκαν δηλαδή 9 συνδυασμοί παραμέτρων, για καθέναν από τους οποίους έγιναν 2000 προσομοιώσεις.

Από τις προσομοιώσεις που έγιναν υπολογίστηκαν τα ποσοστά των χρονοσειρών με πλασματικές ανομογένειες συναρτήσει των εξεταζόμενων συντελεστών Hurst και του λόγου λ των κλίσεων των δύο ευθειών της Διπλής αθροιστικής καμπύλης. Τα αποτελέσματα αυτά συνοψίζονται στον Πίνακα 9.2 και το Σχήμα 9.3.

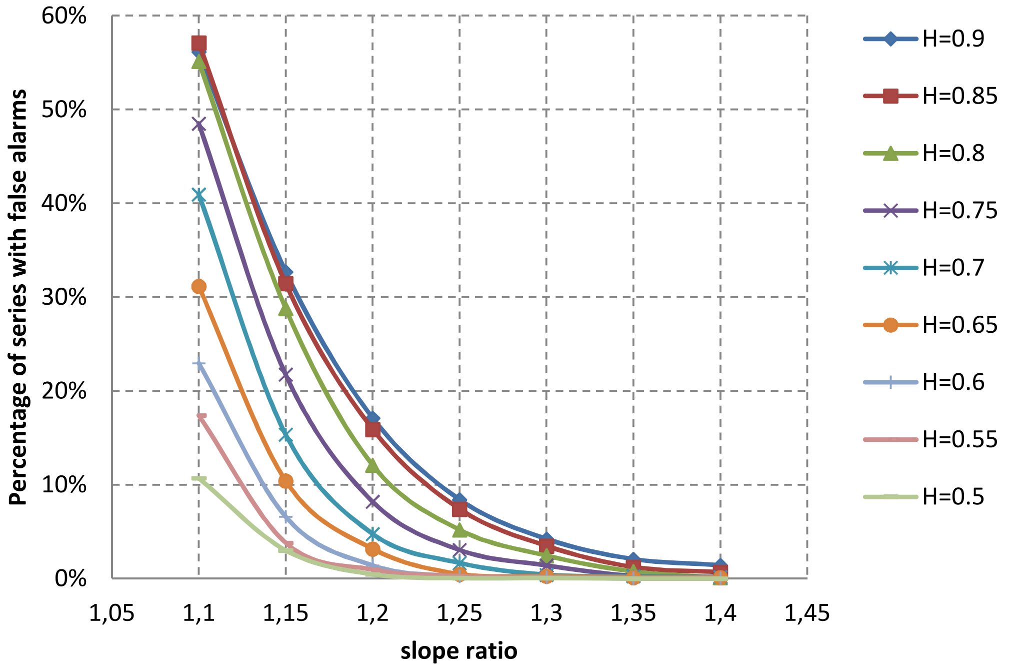
Από την εξέταση του Πίνακα 9.2 προκύπτει ότι το ποσοστό των χρονοσειρών με ψευδή άλματα, ομοίως με το σύστημα 1/1, αυξάνεται:

- με αύξηση του συντελεστή Hurst, και
- με μείωση του λόγου λ των κλίσεων.

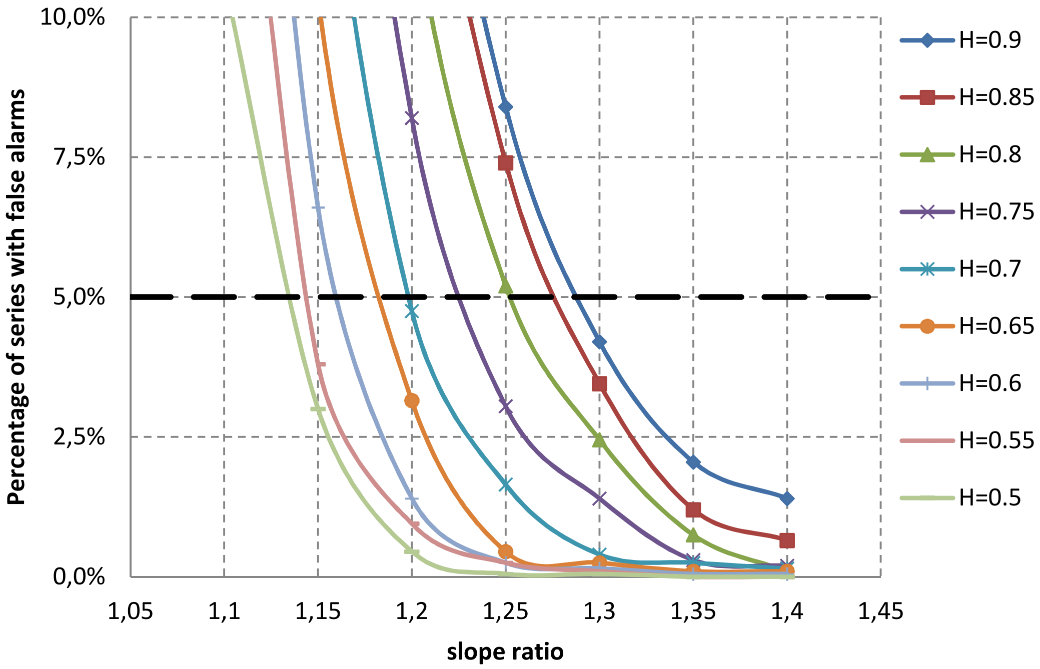
Μη ανεκτά ποσοστά χρονοσειρών με ψευδείς ανομογένειες θεωρήθηκαν ποσοστά μεγαλύτερα από 5% (γραμμοσκιασμένα στον Πίνακα 9.2).

Πίνακας 9.2. *Ποσοστά χρονοσειρών στις οποίες η Διπλή αθροιστική καμπύλη εντοπίζει πλασματικές ανομογένειες με χρήση του συστήματος χρονοσειρών αναφοράς 3/1 και συντελεστή ετεροσυσχέτισης των προσομοιωμένων χρονοσειρών* $\rho_{XY}=0.8$*. Τα αποτελέσματα παρουσιάζονται συναρτήσει του συντελεστή Hurst των χρονοσειρών και του λόγου λ των κλίσεων των δύο ευθειών της Διπλής αθροιστικής καμπύλης. Με σκίαση σημειώνονται τα ποσοστά ψευδούς εντοπισμού που είναι πάνω από το ανεκτό όριο 5%.*

λ \ H	**0.5**	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**	**0.85**	**0.9**
1.10	10.7%	17.4%	23.0%	31.2%	41.0%	48.5%	55.2%	57.1%	56.2%
1.15	3.0%	3.8%	6.6%	10.4%	15.4%	21.8%	28.8%	31.5%	32.7%
1.20	0.5%	1.0%	1.4%	3.2%	4.8%	8.2%	12.1%	15.9%	17.1%
1.25	0.1%	0.3%	0.3%	0.5%	1.7%	3.1%	5.2%	7.4%	8.4%
1.30	0.1%	0.1%	0.2%	0.3%	0.4%	1.4%	2.5%	3.5%	4.2%
1.35	0.0%	0.0%	0.1%	0.1%	0.3%	0.3%	0.8%	1.2%	2.1%
1.40	0.0%	0.0%	0.1%	0.1%	0.2%	0.2%	0.1%	0.7%	1.4%



Σχήμα 9.3. *Ποσοστά χρονοσειρών με πλασματικά άλματα για εφαρμογή της Διπλής αθροιστικής καμπύλης σε δεδομένα με συντελεστή ετεροσυσχέτισης* $\rho_{XY}=0.8$*, επιλεγμένους συντελεστές Hurst και χρήση του συστήματος 3/1. Φαίνεται καθαρά μια μετατόπιση των καμπυλών προς μεγαλύτερα ποσοστά ψευδών εντοπισμών για αύξηση του συντελεστή Hurst*



Σχήμα 9.4. *Λεπτομέρεια του Σχήματος 9.3. Τα ποσοστά των χρονοσειρών με πλασματικές ανομογένειες παρουσιάζονται σε σύγκριση με το ανεκτό ποσοστό 5%.*

Όπως φαίνεται από τον Πίνακα 9.2 και το Σχήμα 9.4, η συμπεριφορά της μεθόδου για σύνθεση μίας χρονοσειράς αναφοράς από τρεις είναι σαφώς βελτιωμένη από την περίπτωση χρήσης μίας μόνο χρονοσειράς αναφοράς (1/1), καθώς διόρθωση της ίδιας αλλαγής κλίσης αντιστοιχεί σε μικρότερο ποσοστό λανθασμένου εντοπισμού.

Για το σύστημα 3/1 φαίνεται επίσης ασφαλές να διορθώνονται λόγοι κλίσεων ≥ 1.3, καθώς αντιστοιχούν σε πιθανότητα λανθασμένου εντοπισμού < 5% (αντίστοιχος λόγος για το σύστημα 1/1, 1.5).

Για H=0.5-0.55 (δεδομένα ανεξάρτητα ή με ασθενή εμμονή) ο λόγος των κλίσεων που αντιστοιχεί σε σφάλμα μικρότερο από 5% είναι 1.15. Για αύξηση του συντελεστή Hurst σε H=0.6-0.7 ο λόγος γίνεται 1.2, για H=0.75, 1.25 και τελικά για πολύ υψηλή συσχέτιση H=0.8-0.9, 1.3.

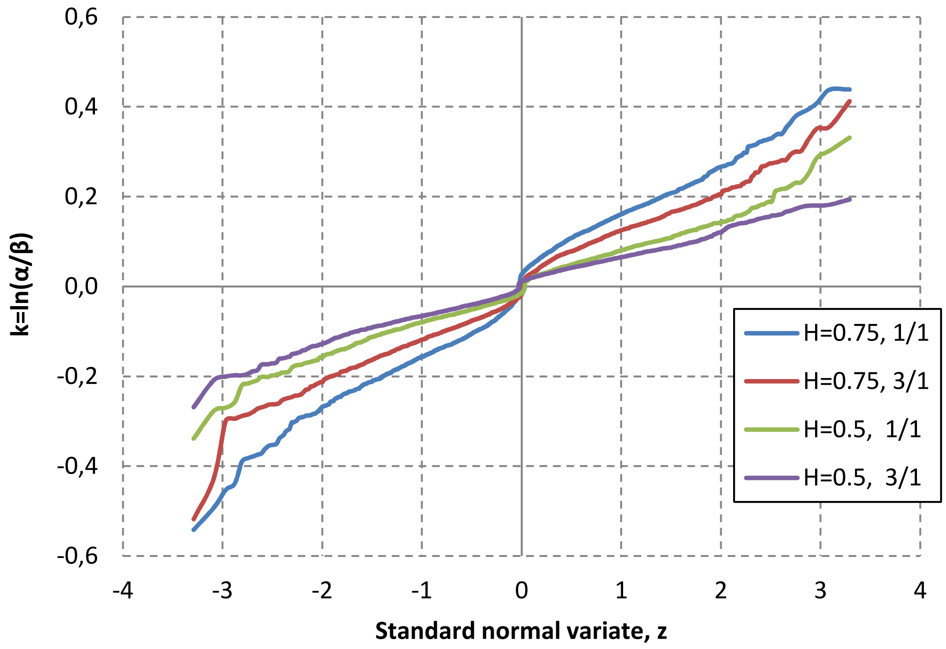
9.3. Διερεύνηση της παραμέτρου k

Για τις δύο περιπτώσεις των συστημάτων αναφοράς 1/1 και 3/1, συντελεστή ετεροσυσχέτισης ανάμεσα στις προσομοιωμένες χρονοσειρές ρ_{XY}=0.8 και δύο τιμές του συντελεστή Hurst H=0.75 και H=0.5, υπολογίστηκε η κατανομή της παραμέτρου k, η οποία

ορίζεται ως: $k = \ln(\tilde{—})$

όπου α, β: οι κλίσεις των ευθειών της Διπλής αθροιστικής καμπύλης.

Ο υπολογισμός έγινε από ένα σύνολο 2000 προσομοιώσεων Monte Carlo. Τα αποτελέσματα για τους τέσσερις συνδυασμούς παραμέτρων παρουσιάζονται στα Σχήμα 9.5.



Σχήμα 9.5. *Σύγκριση των κατανομών του όρου k=ln(α/β) για συντελεστή Hurst H=0.5 (ασυσχέτιστα δεδομένα) και H=0.75 (δεδομένα με μακροπρόθεσμη εμμονή) και χρήση συστημάτων αναφοράς 1/1 και 3/1. Ο υπολογισμός των κατανομών έγινε με 2000 προσομοιώσεις Monte Carlo.*

Από τη σύγκριση των κατανομών που παρουσιάζονται στο Σχήμα 9.5, φαίνεται καθαρά κάτι που παρατηρήθηκε και στις Παραγράφους 9.1 και 9.2. Ότι δηλαδή μια τιμή της αλλαγής κλίσης στη Διπλή αθροιστική καμπύλη έχει μεγαλύτερη πιθανότητα να εντοπιστεί:

- σε χρονοσειρές με μακροπρόθεσμη εμμονή (H > 0.5) απ'ότι σε ασυσχέτιστα δεδομένα,
- για χρήση μίας χρονοσειράς αναφοράς απ'ότι για σύνθεση μιας τεχνητής χρονοσειράς αναφοράς από τρεις προσομοιωμένες χρονοσειρές.

Για συντελεστή Hurst H=0.75 και χρήση μίας χρονοσειράς αναφοράς, τιμές -0.3 ≤ z ≤ 0.3 οι οποίες αντιστοιχούν σε λόγο κλίσεων λ ≤ 1.35 περίπου φαίνονται να είναι αρκετά συχνές.

Για συντελεστή Hurst H=0.75 και χρήση του συστήματος 3/1, λογικές είναι τιμές του όρου -0.2 ≤ z ≤ 0.2 που αντιστοιχούν σε λόγο κλίσεων λ ≤ 1.2 περίπου.

Τέλος, για συντελεστή Hurst H=0.5 (ασυσχέτιστα δεδομένα) και για τα δύο συστήματα 1/1 και 3/1, λογικές είναι τιμές του όρου -0.1 ≤ z ≤ 0.1 που αντιστοιχούν σε λόγο κλίσεων λ ≤ 1.1 περίπου.

10. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

10.1. Συμπεράσματα

Με βάση μια εκτενή βιβλιογραφική επισκόπηση διερευνήθηκε το πρόβλημα της ομογενοποίησης υδροκλιματικών δεδομένων, του εντοπισμού δηλαδή και της διόρθωσης σφαλμάτων που χαρακτηρίζουν τις αντίστοιχες χρονοσειρές. Συγκεκριμένα, εξετάστηκαν και ταξινομήθηκαν οι ανομογένειες και οι υπάρχουσες μεθόδοι ομογενοποίησης δεδομένων θερμοκρασίας και βροχόπτωσης (Κεφάλαια 2, 3). Παρουσιάστηκαν επίσης αναλυτικά και σχολιάστηκαν οι προηγούμενες προσεγγίσεις της αξιολόγησης μεθόδων ομογενοποίησης με χρήση συνθετικών δεδομένων (Κεφάλαιο 4). Από την επισκόπηση αυτή προέκυψαν τα εξής προκαταρκτικά συμπεράσματα στα οποία βασίστηκε η υπολογιστική διερεύνηση της ομογενοποίησης που ειναι το βασικό αντικείμενο της παρούσας Μεταπτυχιακής εργασίας:

Α1. Από τις πιο σημαντικές μεθόδους ομογενοποίησης μέχρι σήμερα είναι το SNHT για άλματα. Η μέθοδος αυτή, αν και πλέον χρησιμοποιείται λιγότερο από πιο σύγχρονες μεθόδους, έχει ωστόσο εμπνεύσει και αποτελεί τη θεωρητική βάση για πρόσφατα ανεπτυγμένες μεθόδους που χρησιμοποιούνται ευρύτατα, όπως το PHA (βλ. Κεφάλαιο 3). Επίσης έχει συμπεριληφθεί στις συγκριτικές μελέτες τις περισσότερες φορές από όλες τις μεθόδους ομογενοποίησης και έχει αξιολογηθεί ως η μέθοδος με την καλύτερη συμπεριφορά σε βάσεις δεδομένων με λίγα άλματα.

Α2. Η αξιολόγηση μεθόδων ομογενοποίησης με βάση συνθετικά δεδομένα έχει επικεντρωθεί μέχρι σήμερα στην εφαρμογή τους σε δεδομένα θερμοκρασίας και σε πολύ μικρότερο βαθμό σε δεδομένα βροχόπτωσης, κάτι λογικό αφού το σημαντικότερο στοιχείο των κλιματικών μεταβολών που μελετώνται είναι η άνοδος της θερμοκρασίας.

Α3. Μέχρι στιγμής δεν έχει γίνει μια συστηματική έρευνα της ομογενοποίησης σε συνθετικά δεδομένα με ένα σημαντικό χαρακτηριστικό που εντοπίζεται σε πολλούς τύπους υδροκλιματικών δεδομένων: τη μακροπρόθεσμη εμμονή.

Α4. Στις μεθόδους και αξιολογήσεις της ομογενοποίησης θεωρείται ότι η διαφορά ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες για τα δεδομένα θερμοκρασίας ή ο λόγος τους για τα δεδομένα βροχόπτωσης είναι μια σειρά τυχαίων αριθμών (χωρίς αυτοσυσχέτιση).

Το τελευταίο αυτό συμπέρασμα είναι κρίσιμο γιατί σ’αυτό βασίζονται οι μέθοδοι ομογενοποίησης. Όμως, μια προκαταρκτική ανάλυση έδειξε ότι η διαφορά δύο γειτονικών χρονοσειρών αναπαράγει τη συνάρτηση αυτοσυσχέτισής τους και δεν οδηγεί σε χρονοσειρά ασυσχέτιστων μεταβλητών (Παράγραφος 5.6).

Βάσει των συμπερασμάτων της βιβλιογραφικής επισκόπησης διερευνήθηκε επίσης η συμπεριφορά των μεθόδων SNHT για άλματα και Διπλής αθροιστικής καμπύλης στον εντοπισμό πλασματικών ανομογενειών σε ετήσια ομογενή συνθετικά δεδομένα θερμοκρασίας και βροχόπτωσης, αντίστοιχα, με μακροπρόθεσμη εμμονή. Εξετάστηκαν διάφορες παράμετροι που μπορεί να επηρεάζουν την ομογενοποίηση, με κυριότερο το συντελεστή Hurst των χρονοσειρών. Για το SNHT διερευνήθηκε επίσης η επιρροή της διόρθωσης των αλμάτων που εντοπίζονται στα χαρακτηριστικά των χρονοσειρών. Για να χρησιμοποιηθεί η Διπλή αθροιστική καμπύλη, αυτοματοποιήθηκε με χρήση αλγορίθμου διγραμμικής προσαρμογής με βάση τα ελάχιστα τετράγωνα. Η δημιουργία των τεχνητών χρονοσειρών έγινε με τη μέθοδο Monte Carlo. Για τα δεδομένα θερμοκρασίας έγιναν 10000 προσοσμοιώσεις για κάθε συνδυασμό παραμέτρων που ελέγθηκε και για τα δεδομένα βροχόπτωσης 2000 προσομοιώσεις. Από τη διερεύνηση, η οποία έγινε με ανάπτυξη συναρτήσεων σε περιβάλλον Matlab, μπορούν να εξαχθούν, σε σύγκριση και με τις προηγούμενες προσεγγίσεις, κάποια βασικά συμπεράσματα για τις δύο μεθόδους που χρησιμοποιήθηκαν αλλά και για την ομογενοποίηση υδροκλιματικών δεδομένων γενικότερα.

Κάποια γενικά συμπεράσματα που προκύπτουν από το σύνολο των προσομοιώσεων είναι:

Β1. Οι μέθοδοι ομογενοποίησης που εξετάστηκαν οδηγούν σε στατιστικά μη σημαντικό αριθμό χρονοσειρών με πλασματικά άλματα για τιμή του συντελεστή Hurst H = 0.5, που αντιστοιχεί σε ασυσχέτιστα δεδομένα (Σχήματα 8.7, 9.2, 9.4). Το συμπέρασμα αυτό είναι σε συμφωνία και με προηγούμενες προσεγγίσεις της ομογενοποίησης.

Β2. Το ποσοστό χρονοσειρών που οι μέθοδοι ομογενοποίησης εντοπίζουν ψευδή άλματα φαίνεται να αυξάνεται με αύξηση του συντελεστή Hurst των χρονοσειρών (Σχήματα 8.7, 9.1, 9.3). Η παρατήρηση αυτή επιβεβαιώθηκε από όλους τους συνδυασμούς δεδομένων και μεθόδων που εξετάστηκαν.

Β3. Ο αριθμός και ο τρόπος χρήσης των χρονοσειρών αναφοράς επηρεάζει πολύ το ποσοστό των χρονοσειρών στις οποίες εντοπίζονται ψευδή άλματα (Σχήματα 8.7, 9.1, 9.3).

Συγκεκριμένα για τα δεδομένα θερμοκρασίας και τη μέθοδο SNHT για άλματα προκύπτουν κάποια επιπλέον συμπεράσματα:

Γ1. Κοινό άλμα σε όλες τις χρονοσειρές αναφοράς (για τουλάχιστον τρεις χρονοσειρές αναφοράς) είναι, εκτός από περιπτώσεις δεδομένων με πολύ μεγάλο συντελεστή Hurst ($H \geq 0.85$) ενδεικτικό πραγματικής ανομογένειας.

Γ2. Στην περίπτωση κοινού άλματος σε μερικές μόνο από τις χρονοσειρές αναφοράς υπάρχει μεγάλη πιθανότητα αυτό να είναι πλασματικό. Επομένως πριν από τη διόρθωσή του πρέπει να γίνεται διερεύνηση της συμπεριφοράς του συστήματος χρονοσειρών αναφοράς που επιλέχτηκε.

Γ3. Ο συντελεστής ετεροσυσχέτισης ανάμεσα στις χρονοσειρές αναφοράς και τις ελεγχόμενες δεν φαίνεται να επηρεάζει το ποσοστό χρονοσειρών με πλασματικά άλματα.

Γ4. Το ποσοστό χρονοσειρών με πλασματικά άλματα φαίνεται να αυξάνεται με το μήκος των χρονοσειρών, όπως προέκυψε από ανάλυση σε χρονοσειρές μήκους 50 και 100 χρόνων

Γ5. Το ποσοστό των χρονοσειρών με πλασματικά άλματα διογκώνεται όσο μειώνεται το ελάχιστο εξεταζόμενο μήκος των τμημάτων των χρονοσειρών κυρίως στα άκρα, κάτι που συμφωνεί και με προηγούμενες εφαρμογές της μεθόδου στη βιβλιογραφία. Θεώρηση τμημάτων μήκους 5 ετών οδηγεί σε πολύ μεγαλύτερο χρονοσειρών με πλασματικά άλματα απ’ότι μήκος 10 ετών.

Γ6. Σε όλες τις περιπτώσεις διόρθωσης των αλμάτων που εντοπίστηκαν, το ποσοστό των χρονοσειρών στις οποίες αυξήθηκε η γραμμική τάση ήταν περίπου ίσο με αυτό των χρονοσειρών στις οποίες μειώθηκε. Επομένως, η συγκεκριμένη μέθοδος ομογενοποίησης δεν φαίνεται να έχει κάποια άμεση επιρροή στις τάσεις μεταβολής της θερμοκρασίας.

Γ7. Για $H < 0.65$ διόρθωση των αλμάτων που εντοπίζει η μέθοδος οδηγεί σε ίδια περίπου ποσοστά χρονοσειρών που αυξάνεται ή μειώνεται ο συντελεστής Hurst. Αντίθετα, για δεδομένα με $H > 0.65$, δηλαδή ισχυρότερη εμμονή, το ποσοστό των χρονοσειρών στις οποίες μειώνεται ο Hurst από την ομογενοποίηση ξεπερνάει το ποσοστό αυτών στις οποίες αυξάνεται, με την ψαλίδα να ανοίγει για αύξηση του Hurst των αρχικών χρονοσειρών. Φαίνεται δηλαδή ότι οι μέθοδοι με μηδενική υπόθεση ανεξάρτητα δεδομένα θεωρούν την εμμονή σφάλμα και τη διορθώνουν.

Για τα δεδομένα βροχόπτωσης και εφαρμογή της Διπλής αθροιστική καμπύλης διαπιστώθηκε ότι:

Δ1. Το ποσοστό των χρονοσειρών με πλασματικά άλματα μειώνεται με την αύξηση του λόγου των κλίσεων των δύο ευθειών για κάθε τιμή του συντελεστή Hurst.

Δ2. Η παραγωγή μίας χρονοσειράς αναφοράς από τρεις χρονοσειρές μειώνει το ποσοστό χρονοσειρών με πλασματικά άλματα σε σχέση με τη χρήση μίας μόνο χρονοσειράς αναφοράς.

Δ3. Το όριο στον λόγο των κλίσεων για το οποίο το ποσοστό των χρονοσειρών με πλασματικά άλματα είναι <5% (άρα η διόρθωση είναι ασφαλής) είναι 1.5 για μία χρονοσειρά αναφοράς και 1.3 για παραγωγή μίας χρονοσειράς αναφοράς από τρεις χρονοσειρές.

10.2. Προτάσεις

Η ανωτέρω έρευνα κάλυψε ένα μεγάλο εύρος παραμέτρων και μεθόδων. Όμως, θα ήταν σκόπιμο στο μέλλον να επεκταθεί:

α) σε μεθόδους ομογενοποίησης που έχουν πρόσφατα αρχίσει να χρησιμοποιούνται ευρέως όπως οι: PHA (Menne and Williams, 2009), MASH (Szentimrey, 1999) και η μέθοδος των Caussinus and Mestre (2004),

β) σε μηνιαία δεδομένα που έχουν πρόσφατα αρχίσει να αποτελούν αντικείμενο έρευνας, και

γ) για την περίπτωση των βροχοπτώσεων, σε δεδομένα που προκύπτουν όχι από κανονική αλλά με χρήση ασύμμετρης κατανομής (π.χ. γάμμα ή λογαριθμοκανονική).

Τέλος, θα ήταν ενδιαφέρουσα η διερεύνηση της αξιοπιστίας των μεθόδων ομογενοποίησης, όχι μόνο με βάση τον εντοπισμό πλασματικών αλμάτων σε ομογενή δεδομένα όπως στην παρούσα έρευνα, αλλά με εισαγωγή ανομογενειών (αλμάτων ή και αλλαγών τάσης) σε συνθετικές χρονοσειρές με εμμονή και διερεύνηση του κατά πόσο οι μέθοδοι ομογενοποίησης μπορούν να ανιχεύσουν τη θέση και το μέγεθος των «πραγματικών» αυτών ανομογενειών.

ΑΝΑΦΟΡΕΣ

Στείρου, Ε. (2011) ‘*Διερεύνηση και αξιολόγηση μεθόδων ομογενοποίησης υδροκλιματικών δεδομένων’*, Διπλωματική εργασία, 156 σελίδες, Τομέας Υδατικών Πόρων και Περιβάλλοντος – Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα.

Aesawy, A.M., Hasanean, H.M. (1998) ‘Annual and seasonal climatic analysis of surface air temperature variations at six southern Mediterranean stations’, *Theor. Appl. Climatol*. 61,55-68.

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J. (2003) ‘Guidelines on climate metadata and homogenization’, *World Meteorological Organization*, WMO-TD No. 1186, WCDMP No. 53, Geneva, Switzerland, p. 55.

Alexandersson, H. (1986) ‘A homogeneity test applied to precipitation data’, *J. Climatol.,* 6, 661475.

Alexandersson H., Moberg A. (1997) ‘Homogenization of Swedish temperature data Part I: homogeneity test for linear trends’, *Int J Climatol* 17:25–34.

Allen, R. J., and A. T. DeGaetano (2000) ‘A method to adjust long-term temperature extreme series for nonclimatic inhomogeneities’, *J. Climate,* 13, 3680–3695.

Auer I, Böhm R, Jurkovic A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones PD, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin J-M, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajic-Capka M, Zaninovic K, Majstorovic Z, Nieplova E. (2007) ‘HISTALP—Historical instrumental climatological surface time series of the greater Alpine region 1760–2003’, *Int J Climatol* 27:17–46.

Auer I, Bohm R, Jurkovic A, Orlik A, Potzmann R, Schöner W, Ungersböck M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones P, Efthymiadis D, Mestre O, Moisselin JM, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajic-Capka M, Zaninovic K, Majstorovic Z, Szalai S, Szentimrey T, Mercalli L. (2005) ‘A new instrumental precipitation dataset for the Greater Alpine Region for the period 1800–2002’. *International Journal of Climatology* 25: 139–166.

Beaulieu C., Ouarda T. B. M. J., Seidou O. (2007) ‘Synthèse des techniques d'homogénéisation des séries climatiques et analyse d’applicabilité aux séries de précipitations’, *Hydrological Sciences Journal*, 52:1, 18-37.

Beaulieu, C., Seidou, O., Ouarda, T. B. M. J., Zhang, X., Boulet, G., and Yagouti, A. (2008) ‘Intercomparison of homogenization techniques for precipitation data’, *Water Resour. Res.*, 44, W02425, doi:10.1029/2006WR005615.

Böhm, R., Jones, P.D., Hiebl, J., Frank, D., Brunetti, M., Maugeri, M. (2010) ‘The early instrumental warm-bias: A solution for long central European temperature series 1760-2007’, *Climatic Change*, 101 (1), pp. 41-67.

Buishand TA. (1982) 'Some methods for testing the homogeneity of rainfall records', *Journal of Hydrology* 58: 11–27.

Buishand, T. A. (1984) 'Tests for detecting a shift in the mean of hydrological time series', *J. Hydrol.*, 73, 51–69.

Burnette, D.J., Stahle, D.W., Mock, C.J. (2010) 'Daily-mean temperature reconstructed for Kansas from early instrumental and modern observations', *Journal of Climate*, 23 (6), pp. 1308-1333.

Caussinus, H. and Lyazrhi, F. (1997) 'Choosing a linear model with a random number of change-points and outliers', *Ann. Inst. Statist. Math*., 49/4, 761-775.

Caussinus, H. and Mestre, O. (2004) 'Detection and correction of artificial shifts in climate series', Appl. Statist., 53, 405–425.

Chang. M., and R. Lee (1974) 'Objective double-mass analysis', *Water Resources Research*, 10: 1123- 1 126.

Conrad, V. and Pollak, C. (1962) 'Methods in Climatology', *Harvard University Press*, Cambridge, MA, 459 pp.

Craddock, J.M. (1979) 'Methods of comparing annual rainfall records for climatic purposes', *Weather*, 34: 332-346.

DeGaetano, A. T. (2006) 'Attributes of several methods for detecting discontinuities in mean temperature series', *J. Climate*, 19, 838–853.

Dingman SL (1994) *Physical Hydrology*. Englewood Cliffs, NJ: Prentice Hall.

Domonkos, P. (2008) 'Testing of homogenisation methods: purposes, tools and problems of implementation', *Proceedings of the 5th Seminar and Quality Control in Climatological Databases*, edited by: Lakatos, M., Szentimrey, T., Bihari, Z. and Szalai, S., WCDMPNo. 71, WMO/TD-NO. 1493, 126–145.

Domonkos, P. (2011a) 'Efficiency evaluation for detecting inhomogeneities by objective homogenization methods', *Theor. Appl. Climatol*., 105, 455–467, doi:10.1007/s00704-011-0399-7.

Ducré-Robitaille, J.-F., Vincent, L. A., and Boulet, G.(2003) 'Comparison of techniques for detection of discontinuities in temperature series', *Int. J. Climatol*., 23, 1087–1101.

Gullett DW, Vincent L, Sajecki PJF. (1990) 'Testing homogeneity in temperature series at Canadian climate stations', *CCC report 90-4*, Climate Research Branch, Meteorological Service of Canada, Ontario, Canada.

IPCC (2007) Summary for Policymakers, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge: Cambridge University.

Easterling, D. R. and Peterson, T. C. (1992) 'Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review', *Preprints: Fifth International Meeting on Statistical Climatology,* 22-26 June 1992 Toronto, pp J28-J32.

Easterling, D.R., and T.C. Peterson (1995) 'A new method for detecting undocumented discontinuities in climatological time series', *International Journal of Climatology*, **15** (4), 369-377.

Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. Christy, and R.A. Pielke Sr. (2011) 'Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends', *J. Geophys. Res., in press*. Copyright (2011) American Geophysical Union.

Founda, D., Kambezidis, H.D., Petrakis, M., Zanis, P., Zerefos, C. (2009) 'A correction of the recent air-temperature record at the historical meteorological station of the National Observatory of Athens (NOA) due to instrument change', *Theoretical and Applied Climatology*, 97 (3-4), pp. 385-389.

Hubert P., Carbonnel J.P. (1987) 'Approche statistique de l'aridification de l'Afrique de l'Ouest', *J. Hydrol.,* 95,165-183.

Hubert P., Carbonnel J.P. (1993) 'Segmentation des séries annuelles de debits de grands fleuves africains', *Bulletin de liaison du CIEH., 92*, 3-10.

Hubert P., Carbonnel J.P., Chaouche A. (1989) 'Segmentation des séries hydrométéorologiques - Application à des séries de précipitations et de débits de l'Afrique de l'Ouest', *J. Hydrol., 110,* 349-367.

Jaruskova, D. (1996) 'Change-point detection in meteorological measurement', *Mon Weath. Rev.* 124, 1535–1543.

Karl, T.R., and C.N. Williams Jr. (1987) 'An approach to adjusting climatological time series for discontinuous inhomogeneities', *J. Climate Appl. Meteor*., 26, 1744-1763.

Karl, T.R., C.N. Williams, Jr., P.J. Young, and W.M. Wendland (1986) 'A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States', *J. Climate Appl. Meteor.*, 25, 145-160.

Kendall S.M.and Stuart A., 1943 (1977 edn) *The Advanced Theory of Statistics,* Charles Griffin, London, UK.Vol. 1, 2 & 3, 472 p, 723 p, 585 p.

Khaliq, M.N., Ouarda, T.B.M.J. (2007) 'A Note on the Critical Values of Standard Normal Homogeneity Test (SNHT)', *International Journal of Climatology* 27 (5), 681-687.

Kohler, M. A. (1949) 'Double-mass analysis for testing the consistency of records and for making adjustments', *B. Am. Meteorol. Soc*., 30, 188–189.

Koutsoyiannis, D. (2002) 'The Hurst phenomenon and fractional Gaussian noise made easy', Hydrological Sciences-Journal-des Sciences Hydrologiques, Vol. 47, Apr, p.573-595.

Koutsoyiannis, D. (2005) 'Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling', *Hydrological Sciences Journal*, 50 (3), 381–404.

Koutsoyiannis, D., and A. Langousis (2011) 'Precipitation', *Treatise on Water Science*, edited by P. Wilderer and S. Uhlenbrook, 2, 27–78, Academic Press, Oxford.

Kyselý, J. and Domonkos, P. (2006) ‘Recent increase in persistence of atmospheric circulation over Europe: comparison with long-term variations since 1881’, *Int. J. Climatol.* 26, 461-483.

Lanzante, J. R. (1996) ‘Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data’, *Int. J. Climatol., 16*, 1197-1226.

Lee, A. F. S. & Heghinian, S. M. (1977) ‘A shift of the mean level in a sequence of independent normal random variables: a Bayesian approach’, *Technometrics,* **19**, 503–506.

Lubes-Niel, H., Masson, J. M., Paturel, J. E. & Servat, E. (1998) ‘Variabilité climatique et statistiques. Étude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité de chroniques’, *Rev. Sci. Eau* **11**, 383–408.

Lund, R., and J. Reeves (2002) ‘Detection of undocumented changepoints: A revision of the two-phase regression model’, *J. Climate,* 15, 2547–2554.

Mandelbrot, B. B. (1965) ‘Une class de processus stochastiques homothetiques a soi: Application a la loi climatologique de H. E. Hurst’, C. R. Hebd. Seances Acad. Sci., 260, 3284–3277.

Maronna R, Yohai VJ. (1978) ‘A bivariate test for the detection of a systematic change in mean’, *Journal of American Statistics Assocation* 73: 640–645.

Menne, M. J. and Williams, C. N. Jr.(2005) ‘Detection of undocumented changepoints using multiple test statistics and composite reference series’, *J. Climate*, 18, 4271–4286.

Menne, M.J., and C.N. Williams Jr. (2009) ‘Homogenization of temperature series via pairwise comparisons’, Journal of Climate, 22(7), 1700-1717.

Menne, M.J., C.N. Williams, and R.S. Vose (2009) ‘The United States Historical Climatology Network Monthly Temperature Data - Version 2’, *Bulletin of the American Meteorological Society*, 90, 993-1107.

Mestre, O. (1999) ‘Step-by-step procedures for choosing a model with change-points’, In *Proceedings of the second seminar for homogenisation of surface climatological data*, Budapest, Hungary, WCDMP-No.41, WMO-TD No. 962, 15–26.

Mestre, O. (2004) ‘Detecting multiple change-points in a gaussian sample using dynamic Programming’, In: *Fourth Seminar for Homogenization and Quality Control in Climatological Databases.* WMO, WCDMP-No. 56, 89-92.

Mielke, P. W., Berry, K. J. and Brier, G. W. (1981) ‘Application of multi-response permutation procedure for examining seasonal changes in monthly mean sea-level pressure patterns’, *Mon. Wea. Rev.,* 109, 12&126.

Moberg, A. and Alexandersson, H. (1997) ‘Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1961’, *Int. J. Climatol*, 17, 35–54.

Moberg, A. and Bergström, H. (1997) ‘Homogenization of Swedish temperature data. Part III: The long temperature records from Uppsala and Stockholm’, *International Journal of Climatology*, 17 (7), pp. 667-699.

Olaniran O.J. (1991) Evidence of climatic change in Nigeria based on annual series of rainfall of different daily amounts, 1919-1985. *Climatic change., 19,* 319-341

Ouarda, T. B. M. J., Rasmussen, P. F., Cantin, J.-F., Bobée, B., Laurence, R., Hoang, V. D. & Barabé, G. (1999) 'Identification d'un réseau hydrométrique pour le suivi des modifications climatiques dans la province de Québec', *Rev. Sci. Eau* **12**, 425–448.

Perreault, L., Haché, M., Slivitzsky, M. & Bobée, B. (1999) 'Detection of changes in precipitation and runoff over eastern Canada and US using a Bayesian approach', *Stochast. Environ. Res. Risk Assesment* **13**, 201–216.

Perreault, L., Parent É., Bernier J., Bobée B. & Slivitzky M. (2000) 'Retrospective multivariate Bayesian change-point analysis: a simultaneous single change in the mean of several hydrological sequences', *Stochast. Environ. Res. Risk Assessment* **14**, 243–261.

Peterson, T.C., R. Vose, R. Schmoyer, and V. Razuvaev (1998) 'Global Historical Climatology Network (GHCN) quality control of monthly temperature data', *International Journal of Climatology,* 18 (11), 1169-1179.

Pettitt AN. (1979), 'A non-parametric approach to the change-point detection', *Applied Statistics* 28: 126–135.

Potter KW. (1981) 'Illustration of a new test for detecting a shift in mean precipitation series', *Monthly Weather Review* 109: 2040–2045.

Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes (1991) 'Effects of recent thermometer changes in the Cooperative Station Network" *Bull. Amer. Meteor. Soc.,* 72, 1718–1723.

Rasmussen, P. (2001) 'Bayesian estimation of change points using the general linear model', *Water Resour. Res., 37*, 2723-2731.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. (2007) 'A review and comparison of changepoint detection techniques for climate data', *J. Appl. Meteorol. Climatol*., 46, 900–915.

Rohde R, Muller RA, Jacobsen R, Muller E, Perlmutter S, et al. (2012) 'A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011', *Geoinfor Geostat: An Overview* 1:1.

Rust, H.W., O. Mestre, and V.K.C. Venema (2008) 'Less jumps, less memory: homogenized temperature records and long memory', JGR-Atmospheres, 113, D19110, doi: 10.1029/2008JD009919.

Sahin, S., Cigizoglu, H.K. (2010) 'Homogeneity analysis of Turkish meteorological data set', *Hydrological Processes*, 24 (8), pp. 981-992.

Schwarz, G. (1978) 'Estimating the dimension of a model', *Ann. Stat.,* 6, 461–464.

Searcy, J.K. and Hardison, C.H. (1960) 'Double-mass curves' In: *Manual of Hydrology*: Part 1, General Surface Water Techniques, U.S. Geol. Surv., Water-Supply Pap., 1541-B: Washington, D.C., 31-59.

Solow, A. (1987) 'Testing for climatic change: an application of the two-phase regression model', J. *Clim. Appl. Meteorol.*, 26, 1401-1405.

Steirou, E., and D. Koutsoyiannis (2012) ‘Investigation of methods for hydroclimatic data homogenization’, *European Geosciences Union General Assembly 2012, Geophysical Research Abstracts,* Vol. 14, Vienna, 956-1.

Štěpánek, P. and P. Zahradníček (2009): 1st Joint DMCSEE-JRS Workshop on Drought Monitoring, Ljubljana, Slovenia, 21.-25. Sept. 2009.

Szentimrey, T. (1999) ‘Multiple Analysis of Series for homogenization (MASH)’, *Proceedings of the second seminar for homogenization of surface climatological data*, Budapest, Hungary, WMO, WCDMP-No. 41, 27–46.

Tyralis, H., and D. Koutsoyiannis (2011) ‘Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process’, *Stochastic Environmental Research & Risk Assessment*, 25 (1), 21–33.

van Dobben de Bruyn, C.S. (1968) ‘Cumulative Sum Tests: Theory and Practice, Griffin’s Statistical Monographs and Courses, No. 24, London, Griffin.

Venema, V. K. C., O. Mestre, E. Aguilar, I. Auer, J.A. Guijarro, P. Domonkos, G. Vertacnik, T. Szentimrey, P. Stepanek, P. Zahradnicek, J. Viarre, G. Müller-Westermeier, M. Lakatos, C.N. Williams, M.J. Menne, R. Lindau, D. Rasol, E. Rustemeier, K. Kolokythas, T. Marinova, L. Andresen, F. Acquaotta, S. Fratianni, S. Cheval, M. Klancar, M. Brunetti, C. Gruber, M. Prohom Duran, T. Likso, P. Esteban, and T. Brandsma, (2012) ‘Benchmarking homogenization algorithms for monthly data’, *Climates of the Past*, 8, 89-115, doi:10.5194/cp-8-89-2012.

Vincent LA. (1998) ‘A technique for the identification of inhomogeneities in Canadian temperature series’, *Journal of Climate* 11: 1094–1104.

Wang, X. L. (2003) Comments on “Detection of undocumented changepoints: A revision of the two-phase regression model.” *J. Climate,* 16**,** 3383–3385.

Wang, X. L. L. (2008) ‘Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test’, J. Appl. Meteor. Climatol., 47, 2423–2444.

Wang, X. L. L., Q. H. Wen, and Y. Wu (2007) ‘Penalized maximal *t* test for detecting undocumented mean change in climate data series’, *J. Appl. Meteor. Climatol.,* 46, 916–931, doi:10.1175/JAM2504.1.

Wijngaard JB, Klein Tank AMG, Können GP. (2003) ‘Homogeneity of 20th century European daily temperature and precipitation series’, *International Journal of Climatology* 23: 679–692.

WMO (World Meteorological Organization) (1966) Climatic change by a working group of the commission for climatology. W.M.O. 195, TP 100, Tech, note n° 79. WMO Geneva, Switzerland.

Ιστοσελίδες:

http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html

ΠΑΡΑΡΤΗΜΑ Α

Παρατίθενται οι κρίσιμες τιμές T_{crit} της στατιστικής τιμής ελέγχου (test value) T_0 της μεθόδου SNHT for single shifts σύμφωνα με τους Khaliq and Ouarda (2007), για χρονοσειρές μήκους 10-100 έτη.

Οι Khaliq and Ouarda (2007) υπολόγισαν τις T_{crit} για τα μεγέθη δείγματος n που εικονίζονται στον παρακάτω πίνακα με μαύρο χρώμα. Οι υπόλοιπες τιμές (εικονίζονται με κόκκινο χρώμα) υπολογίστηκαν με γραμμική παρεμβολή από την προηγούμενη και την επόμενη τιμή.

n	Tcrit	n	Tcrit	n	Tcrit
10	5.637	41	8.1825	71	8.829
11	5.8525	42	8.214	72	8.844
12	6.068	43	8.2435	73	8.8585
13	6.235	44	8.273	74	8.873
14	6.402	45	8.302	75	8.8855
15	6.538	46	8.331	76	8.898
16	6.674	47	8.3565	77	8.912
17	6.7865	48	8.382	78	8.926
18	6.899	49	8.407	79	8.9385
19	6.994	50	8.432	80	8.951
20	7.089	51	8.456	81	8.9635
21	7.173	52	8.48	82	8.976
22	7.257	53	8.502	83	8.9885
23	7.3285	54	8.524	84	9.001
24	7.4	55	8.545	85	9.0135
25	7.4645	56	8.566	86	9.026
26	7.529	57	8.586	87	9.0365
27	7.586	58	8.606	88	9.047
28	7.643	59	8.6265	89	9.057
29	7.695	60	8.647	90	9.067
30	7.747	61	8.665	91	9.0785
31	7.794	62	8.683	92	9.09
32	7.841	63	8.7	93	9.1
33	7.8855	64	8.717	94	9.11
34	7.93	65	8.7345	95	9.1185
35	7.9695	66	8.752	96	9.127
36	8.009	67	8.768	97	9.137
37	8.045	68	8.784	98	9.147
38	8.081	69	8.799	99	9.157
39	8.116	70	8.814	100	9.167
40	8.151				

ΠΑΡΑΡΤΗΜΑ Β

Παρατίθενται οι συναρτήσεις που αναπτύχθηκαν στα πλαίσια της εργασίας αυτής για τον έλεγχο, με χρήση συνθετικών δεδομένων, των μεθόδων ομογενοποίησης SNHT και Διπλή αθροιστική καμπύλη. Πληροφορίες και διαγράμματα ροής για τις παρακάτω συναρτήσεις παρατίθενται στο Κεφάλαιο 7.

Κύριες συναρτήσεις

I) Συνάρτηση CheckSNHT_multiple:

```
function apotelesma6=CheckSNHT_multiple(H,g0,mv,rr,ti,iter)

% description: location and correction of jumps identified by
% 1 reference series

%   H:    Hurst coefficient
%   g0:   variance
%   mv:   mean value
%   rr:   cross-correlation coefficient
%   ti:   length of time series
%   iter: number of iterations


apotelesma6=zeros(iter,5);

for t=1:iter

% subroutine 2A - creation of correlated reference and candidate
series
apotelesma2=series_ref_cand(H,g0,mv,rr,ti);
X=apotelesma2(:,1);
Y=apotelesma2(:,2);

% subroutine 4A - application of the SNHT to all couples of
% reference-candidate series
Uresult=multiple_shifts2(X,Y);

% subroutine 5 - correction of located shifts
apotelesma4=Scor(apotelesma2,Uresult);

% calculation of initial and final slope  slop1, slop2
x=zeros(length(apotelesma2),1);
x(:)=1:100;

y1=apotelesma4(:,1);
p1=polyfit(x,y1,1);
slop1=p1(1);

y2=apotelesma4(:,2);
```

```
p2=polyfit(x,y2,1);
slop2=p2(1);

% subroutine 6 - estimation of initial and final Hurst coef. InH,
FiH
x1=apotelesma4(:,1);
H1 = lsv(x1);

x2=apotelesma4(:,2);
H2 = lsv(x2);

alm=apotelesma4(1,3);

apotelesma6(t,:)=[slop1 slop2 H1 H2 alm];

end

end
```

II) Συνάρτηση CheckSNHT_Mref:

```
function apotelesmaMULT2=CheckSNHT_Mref(H,g0,mv,rr,m,j,ti,iter)

% description: location of common jumps identified by multiple
reference series

%   H:    Hurst coefficient
%   g0:   variance
%   mv:   mean value
%   rr:   cross-correlation coefficient
%   m:    number of reference series
%   j:    threshold - number of simultaneous jumps in order to be
considered significant
%   ti:   length of time series
%   iter: number of iterations


apotelesmaMULT2=zeros(iter,1);

for t=1:iter

% subroutine 2B - creation of (m+1) correlated reference and
candidate series
Mref=creation_series_ref_cand(H,g0,mv,rr,m,ti);

k=m+1;

% extraction of reference series
X=zeros(ti,k-1);

for i=1:k-1
    X(:,i)=Mref(:,i);
end

% extraction of candidate series
```

```
Y1=Mref(:,k);

% subroutine 4B - application of the SNHT to all m couples of
reference-candidate series
pos=zeros(ti,m);

for l=1:m
    pos(:,l)=multiple_shifts_mref2(X(:,l),Y1,ti);
end


% identification of all common jumps
Njumps=0;
d=0;

n=0;
while n<ti-2;
    n=n+1;
    comj=pos(n:n+2,:);

    ncj1=length(find(comj==77));
    ncj2=length(find(comj==88));

    if ncj1>=j || ncj2>=j
        Njumps=Njumps+1;
        d=1;
    end
    if d==1
        n=n+2;
        d=0;
    end

end

apotelesmaMULT2(t,:)=Njumps;

end

end
```

III) Συνάρτηση CheckDMass1:

```
function DoubleMass1=CheckDMass1(H,g0,mv,rr,m,ti,iter)

% description: identification of the most significant inhomogeneity
of a
% precipitation time series using the Double mass curve and a single
% reference series or 1 final reference series calculated by
multiple
% initial reference series

%   H:    Hurst coefficient
%   g0:   variance
%   mv:   mean value
%   rr:   cross-correlation coefficient
```

```
%   m:    number of reference series
%   ti:   length of time series
%   iter: number of iterations


DoubleMass1=zeros(iter,3);

for t=1:iter

    % subroutine 2C - production of reference and candidate series
    apotelesma2=series_ref_cand_DMass(H,g0,mv,rr,m,ti);

    X=apotelesma2(:,1);
    Y=apotelesma2(:,2);

    % subroutine 4C - formation of  precipitation series,
    % double mass procedure, calculation of trends, position of
change
    DMC_mult=DMCmult_ref_cand(X,Y,ti);

    % trends before and after the most significant inhomogeneity,
position
    % of change
    a=DMC_mult(1,1);
    b=DMC_mult(1,2);
    x=DMC_mult(1,3);

    DoubleMass1(t,:)=[a b x];

end

end
```

Βοηθητικές συναρτήσεις

1) <u>Συνάρτηση hurst_3AR1</u>

```
function apotelesma1=hurst_3AR1(H,g0,mv,ti)

% Subroutine 1: creation of time series with HK behaviour. Method
proposed by Koutsoyiannis (2002)


% calculation of autocorrelation coef. for lag 1 of the 3 AR(1)
series
r=1.52*(H-0.5)^1.32;
f=0.953-7.69*(1-H)^3.85;
if H<=0.76
    ksi=0.932+0.087*H;
else
    ksi=0.993+0.007*H;
end
```

```
% calculation of autocorrelation coef. for lag 1 and 100 of
fractional Gaussian noise
r1=0.5*((1+1)^(2*H)+(1-1)^(2*H))-1^(2*H);
r100=0.5*((100+1)^(2*H)+(100-1)^(2*H))-100^(2*H);


% calculation of parameters c1, c2
a11=f-r;
a12=ksi-r;
a21=f^100-r^100;
a22=ksi^100-r^100;

A=[a11 a12;a21 a22];

b1=r1-r;
b2=r100-r^100;

B=[b1;b2];

C=A\B;

% selection of the mean value of the 3 AR(1) series
m1=0;
m2=0;
m3=0;

% calculation of variances of the 3 AR(1) series
var1=(1-C(1,1)-C(2,1))*g0;
var2=C(1,1)*g0;
var3=C(2,1)*g0;

varv1=(1-r^2)*var1;
varv2=(1-f^2)*var2;
varv3=(1-ksi^2)*var3;


% creation 3 AR(1) series of length ti
W1=zeros(ti+1,1);
W1(1,1)=norminv(rand(1),m1,sqrt(var1));

W2=zeros(ti+1,1);
W2(1,1)=norminv(rand(1),m2,sqrt(var2));

W3=zeros(ti+1,1);
W3(1,1)=norminv(rand(1),m3,sqrt(var3));

XT=zeros(ti,1);

for i=2:ti+1
    W1(i,1)=W1(i-1,1)*r+norminv(rand(1),m1,sqrt(varv1));
    W2(i,1)=W2(i-1,1)*f+norminv(rand(1),m2,sqrt(varv2));
    W3(i,1)=W3(i-1,1)*ksi+norminv(rand(1),m3,sqrt(varv3));
end

% sum of the 3 independent AR(1) processes and of the selected mean
value
```

```
for i=1:ti
    XT(i,1)=W1(i+1,1)+W2(i+1,1)+W3(i+1,1)+mv;
end

apotelesma1=XT;

end
```

2A) Υπορουτίνα series ref_cand:

```
function apotelesma2=series_ref_cand(H,g0,mv,rr,ti)

% Subroutine 2A: formation of a pair of correlated reference-
candidate
% series (S2,S3)

for i=1:100

% subroutine 1 - creation of 2 non-correlated time series S1 and S2
% with HK behaviour
S1=hurst_3AR1(H,g0,mv,ti);
S2=hurst_3AR1(H,g0,mv,ti);

% creation of the candidate series S3 with HK behaviour as a linear
% function of S1, S2
sS1=std(S1);
sS2=std(S2);

g01=sS1^2;
g02=sS2^2;

b=sqrt(g01/(1-rr^2));
a=sqrt(rr^2*g01/(g02*(1-rr^2)));

S3=(S1+a.*S2)./b;

% subroutine 6: estimation of Hurst coef. of series S3
x15=S3;
H15 = lsv(x15);

% examination if the Hurst coef. of the candidate series is out of
bounds
if H15<=H+0.05 && H15>=H-0.05, break, end;

end;

if H15>H+0.05 || H15<H-0.05, error('hurst coef. out of bounds');
end;

apotelesma2=[S2 S3];

end
```

2B) Συνάρτηση creation_series ref_cand:

```
function Mref=creation_series_ref_cand(H,g0,mv,rr,m,ti)

% Subroutine 2B: creation of a matrix of (m+1) correlated series W
from a
% matrix of (m+1) non-correlated series S. Procedure: W=A*S


% creation of independent reference and candidate series - matrix S
k=m+1;

S=zeros(ti,k);

% subroutine 1 - creation of (m+1) series with HK behaviour
for i=1:k
    S(:,i)=hurst_3AR1(H,g0,mv,ti);
end

% creation of matrix A
B=zeros(k,k);
B(:)=rr;
for i=1:k
    for j=1:k
        if i==j, B(i,j)=1; end;
    end
end

A = chol(aa,'lower');

% creation of correlated reference and candidate series - matrix W
W=zeros(ti,k);
R=zeros(ti,k);
D=zeros(ti,1);

for i=1:k
    for j=1:k
        R(:,j)=A(i,j).*S(:,j);
    end

    for l=1:ti
        D(l,1)=sum(R(l,:));
    end

    W(:,i)=D;

end

Mref=W;

end
```

2C) Συνάρτηση series_ref_cand_DMass:

```
function Series_Dmass=series_ref_cand_DMass(H,g0,mv,rr,m,ti)
```

```
% Subroutine 2C: creation of a pair of correlated reference and
candidate
% series

Ser=zeros(ti,2);

if m==1

    % subroutine 2A - creation of a pair of correlated reference and
    % candidate series
    Ser=series_ref_cand(H,g0,mv,rr,ti);

elseif m>=2

    % subroutine 2B - creation of a matrix of (m+1) correlated
series,
    % m reference  and 1 candidate series
    Mref=creation_series_ref_cand(H,g0,mv,rr,m,ti);
    X=Mref(:,1:m);
    Y=Mref(:,m+1);
    ref=zeros(ti,1);

    % creation of 1 final reference series calculated by the m
initial
    % reference series
    for j=1:ti
        ref(j,1)=sum(X(j,:)/m);
    end

    Ser=[ref Y];
end

Series_Dmass=Ser;

end
```

3A) Συνάρτηση SNHT_breaking_pos:

```
function result_p=SNHT_breaking_pos(X,Y)

% Subroutine 3A: A candidate series (Y) is tested using the SNHT for
the
% existence of a statistically significant shift at a 95%
significance
% level, in comparison to a reference time series (X)

% creation of time series Q from the pair of reference-candidate
time
% series
Q0=Y-X+mean(X)-mean(Y);

% calculation of mean value, standard deviation and length of the
series Q
mQ0=mean(Q0);
sq=std(Q0);
```

```
n=length(Q0);

% examination if the length of the time series Q is adequate for the
test
if n<10
    apot=[-999.9,-999.9];

elseif n>=10

% normalisation of time series Q
Z0=(Q0-mQ0)./sq;

z1=zeros(n-1,1);
z2=zeros(n-1,1);
T0=zeros(n-1,1);

% Calculation of test value T0
for i=1:n-1
        z1(i,1)=mean(Z0(1:i));
        z2(i,1)=mean(Z0(i+1:end));

        T0(i,1)=i.*z1(i,1).^2+(n-i).*z2(i,1).^2;
end

% identification of maximum value of T0 (mT0)
mT0=max(T0);

% location of position x0 of mT0
x0=find(T0==mT0);

% subroutine 7: critical values of the SNHT for multiple series
lengths at
% a significance level 95%
Tcr=Tcritical(n);

% examination if mT0 is statistically significant and calculation of
% adjustement
if (mT0>Tcr)
    if (x0>5 && x0<(n-5))

        mq1=sq.*mean(Z0(1:x0))+mQ0;
        mq2=sq.*mean(Z0(x0+1:end))+mQ0;
        adj=mq2-mq1;

        apot=[x0,adj];
    else
        apot=[-999.9,-999.9];
    end

else

    apot=[-999.9,-999.9];

end

end
```

```
% Output: if the shift is significant, location x0 of maximum value,
% adjustement
result_p=apot;

end
```

3B) Συνάρτηση piecewise_linear:

```
function proc_pl=piecewise_linear(S1,S2,le)

% Subroutine 3B: Double mass curve procedure

k=length(S1);

% extraction of reference (X) and candidate (Y) series before and
after the
% possible inhomogeneity
X1=S1(1:le,1);
X2=S1(le+1:k,1);
Y1=S2(1:le,1);
Y2=S2(le+1:k,1);

m=k-le;
n=le;
x0=S1(le);
x1=S1(k);

Sx1=sum(S1(1:le));
Sx2=sum(S1(le+1:k));
Sy2=sum(S2(le+1:k));
Sx1_2=sum(X1.*X1);
Sx2_2=sum(X2.*X2);
xy1=sum(X1.*Y1);
xy2=sum(X2.*Y2);

% calculation of "help" values
cc=x0*Sx2-m*x0*x1-Sx2_2+x1*Sx2;
aa=m*x0^2+Sx2_2-2*x0*Sx2;


y01=aa*xy1*(x1-x0)/(x0*cc)-aa*xy2/cc+aa*x1*Sy2/cc-xy2+x0*Sy2;
y02=m*x0-Sx2+(x1-x0)*aa*Sx1_2/(x0^2*cc)-aa*Sx2/cc+aa*x1*m/cc;

y0=y01/y02;

y11=(x1-x0)*(m*x0*y0+xy2-x0*Sy2-y0*Sx2);

y1=y0+y11/aa;

a=y0/x0;
b=(y1-y0)/(x1-x0);
g=y0-((y1-y0).*x0)./(x1-x0);

% least squares
```

```
K1=a.*X1;
Res1=(Y1-K1).^2;
sR1=sum(Res1);

K2=b.*X2+g;
Res2=(Y2-K2).^2;
sR2=sum(Res2);

% sum of least squares
SRes=sR1+sR2;

proc_pl=[a b SRes];

end
```

3C) Συνάρτηση DMCmult_ref_cand:

```
function DMC_mult=DMCmult_ref_cand(X,Y,ti)

% Subroutine 3C: formation of  precipitation series, fitting of the
Double mass curve

A=X*250+1000;
B=Y*250+1000;

% replacement of negative precipitation values with zero values

for g=1:ti
    if A(g)<0;A(g)=0;end
    if B(g)<0;B(g)=0;end
end

% formation of aggregated values with reverse order

Q1=flipud(A);
Q2=flipud(B);

S1=cumsum(Q1,1);
S2=cumsum(Q2,1);

% identification of the most significant trend change

kliseis=zeros(ti-9,3);

for i=1:ti-9
    kliseis(i,:)=piecewise_linear(S1,S2,i+4);
end

MR=min(kliseis(:,3));
loc=find(kliseis(:,3)==MR);

% output: slopes of the Double mass curve, position of inhomogeneity
a=kliseis(loc,1);
b=kliseis(loc,2);
x=ti-loc-4;
```

```
DMC_mult=[a b x];

end
```

4A) Συνάρτηση multiple_shifts2:

```
function Uresult=multiple_shifts2(X,Y)

% Subroutine 4A: location of shifts identified by successive
application
% of the SNHT every time a shifts is located the time series is cut
in the
% place of the jump and both parts are retested


bpoint=zeros(15,2);
bpoint(:,:)=-999.9;

bpoint(8,1:2)=SNHT_breaking_pos(X,Y);
x8=bpoint(8,1);

% breaking the time series in two parts, 1/2 & 2/2, if a significant
shift
% is identified

if bpoint(8,1)>0

    % first part of the time series

    XN13=X(1:x8,1);
    YN13=Y(1:x8,1);

    XN14=X(x8+1:end,1);
    YN14=Y(x8+1:end,1);

    bpoint(4,1:2)=SNHT_breaking_pos(XN13,YN13);
    x4=bpoint(4,1);

    if bpoint(4,1)>0

        % breaking the time series in two parts, 1/4 & 2/4, because
of the
        % identification of a significant shift

        XN9=XN13(1:x4,1);
        YN9=YN13(1:x4,1);

        XN10=XN13(x4+1:end,1);
        YN10=YN13(x4+1:end,1);

        bpoint(2,1:2)=SNHT_breaking_pos(XN9,YN9);
        x2=bpoint(2,1);
```

```
        % breaking the time series in two parts, 1/8 & 2/8, because
of the
        % identification of a significant shift

        if bpoint(2,1)>0

            XN1=XN9(1:x2,1);
            YN1=YN9(1:x2,1);

            XN2=XN9(x2+1:end,1);
            YN2=YN9(x2+1:end,1);

            bpoint(1,1:2)=SNHT_breaking_pos(XN1,YN1);
            bpoint(3,1:2)=SNHT_breaking_pos(XN2,YN2);

            if bpoint(3,1)>0
            bpoint(3,1)=bpoint(3,1)+x2;
            end

        end

        bpoint(6,1:2)=SNHT_breaking_pos(XN10,YN10);
        x6=bpoint(6,1);

        if bpoint(6,1)>0

            % breaking the time series in two parts, 3/8 & 4/8,
because of the
            % identification of a significant shift

            bpoint(6,1)=bpoint(6,1)+x4;

            XN3=XN10(1:x6,1);
            YN3=YN10(1:x6,1);

            XN4=XN10(x6+1:end,1);
            YN4=YN10(x6+1:end,1);

            bpoint(5,1:2)=SNHT_breaking_pos(XN3,YN3);
            if bpoint(5,1)>0
            bpoint(5,1)=bpoint(5,1)+x4;
            end
            bpoint(7,1:2)=SNHT_breaking_pos(XN4,YN4);
            if bpoint(7,1)>0
            bpoint(7,1)=bpoint(7,1)+bpoint(6,1);
            end

        end

    end


    % second part of the time series

    bpoint(12,1:2)=SNHT_breaking_pos(XN14,YN14);
    x12=bpoint(12,1);
```

```
    if bpoint(12,1)>0

        % breaking the time series in two parts, 3/4 & 4/4, because
of the
        % identification of a significant shift

        bpoint(12,1)=bpoint(12,1)+x8;

        XN11=XN14(1:x12,1);
        YN11=YN14(1:x12,1);

        XN12=XN14(x12+1:end,1);
        YN12=YN14(x12+1:end,1);

        bpoint(10,1:2)=SNHT_breaking_pos(XN11,YN11);
        x10=bpoint(10,1);

        % breaking the time series in two parts, 5/8 & 6/8, because
of the
        % identification of a significant shift

        if bpoint(10,1)>0

            bpoint(10,1)=bpoint(10,1)+x8;

            XN5=XN11(1:x10,1);
            YN5=YN11(1:x10,1);

            XN6=XN11(x10+1:end,1);
            YN6=YN11(x10+1:end,1);

            bpoint(9,1:2)=SNHT_breaking_pos(XN5,YN5);
            if bpoint(9,1)>0
            bpoint(9,1)=bpoint(9,1)+x8;
            end
            bpoint(11,1:2)=SNHT_breaking_pos(XN6,YN6);
            if bpoint(11,1)>0
            bpoint(11,1)=bpoint(11,1)+bpoint(10,1);
            end

        end

        bpoint(14,1:2)=SNHT_breaking_pos(XN12,YN12);
        x14=bpoint(14,1);

        if bpoint(14,1)>0

            % breaking the time series in two parts, 7/8 & 8/8,
because of the
            % identification of a significant shift

            bpoint(14,1)=bpoint(14,1)+bpoint(12,1);

            XN7=XN12(1:x14,1);
            YN7=YN12(1:x14,1);
```

```
            XN8=XN12(x14+1:end,1);
            YN8=YN12(x14+1:end,1);

            bpoint(13,1:2)=SNHT_breaking_pos(XN7,YN7);
            if bpoint(13,1)>0
            bpoint(13,1)=bpoint(13,1)+bpoint(12,1);
            end
            bpoint(15,1:2)=SNHT_breaking_pos(XN8,YN8);
            if bpoint(15,1)>0
            bpoint(15,1)=bpoint(15,1)+bpoint(14,1);
            end

        end

    end

end

Uresult=bpoint;

end
```

4B) Συνάρτηση multiple_shifts_mref2:

```
function MRshifts2=multiple_shifts_mref2(X,Y,ti)

% Subroutine 4B: location of shifts identified by successive
application
% of the SNHT every time a shifts is located the time series is cut
in the
% place of the jump and both parts are retested - attribution of
shifts to
% the candidate series

bpoint=zeros(15,2);
bpoint(:,:)=-999.9;

bpoint(8,1:2)=SNHT_breaking_pos(X,Y);
x8=bpoint(8,1);


% breaking the time series in two parts, 1/2 & 2/2, if a significant
shift
% is identified

if bpoint(8,1)>0

    % first part of the time series

    XN13=X(1:x8,1);
    YN13=Y(1:x8,1);

    XN14=X(x8+1:end,1);
    YN14=Y(x8+1:end,1);
```

```
    bpoint(4,1:2)=SNHT_breaking_pos(XN13,YN13);
    x4=bpoint(4,1);

    if bpoint(4,1)>0

        % breaking the time series in two parts, 1/4 & 2/4, because
of the
        % identification of a significant shift

        XN9=XN13(1:x4,1);
        YN9=YN13(1:x4,1);

        XN10=XN13(x4+1:end,1);
        YN10=YN13(x4+1:end,1);

        bpoint(2,1:2)=SNHT_breaking_pos(XN9,YN9);
        x2=bpoint(2,1);

        % breaking the time series in two parts, 1/8 & 2/8, because
of the
        % identification of a significant shift

        if bpoint(2,1)>0

            XN1=XN9(1:x2,1);
            YN1=YN9(1:x2,1);

            XN2=XN9(x2+1:end,1);
            YN2=YN9(x2+1:end,1);

            bpoint(1,1:2)=SNHT_breaking_pos(XN1,YN1);
            bpoint(3,1:2)=SNHT_breaking_pos(XN2,YN2);

            if bpoint(3,1)>0
            bpoint(3,1)=bpoint(3,1)+x2;
            end

        end

        bpoint(6,1:2)=SNHT_breaking_pos(XN10,YN10);
        x6=bpoint(6,1);

        if bpoint(6,1)>0

            % breaking the time series in two parts, 3/8 & 4/8,
because of the
            % identification of a significant shift

            bpoint(6,1)=bpoint(6,1)+x4;

            XN3=XN10(1:x6,1);
            YN3=YN10(1:x6,1);

            XN4=XN10(x6+1:end,1);
            YN4=YN10(x6+1:end,1);
```

```
            bpoint(5,1:2)=SNHT_breaking_pos(XN3,YN3);
            if bpoint(5,1)>0
            bpoint(5,1)=bpoint(5,1)+x4;
            end
            bpoint(7,1:2)=SNHT_breaking_pos(XN4,YN4);
            if bpoint(7,1)>0
            bpoint(7,1)=bpoint(7,1)+bpoint(6,1);
            end

        end

    end


    % second part of the time series

    bpoint(12,1:2)=SNHT_breaking_pos(XN14,YN14);
    x12=bpoint(12,1);

    if bpoint(12,1)>0

        % breaking the time series in two parts, 3/4 & 4/4, because
of the
        % identification of a significant shift

        bpoint(12,1)=bpoint(12,1)+x8;

        XN11=XN14(1:x12,1);
        YN11=YN14(1:x12,1);

        XN12=XN14(x12+1:end,1);
        YN12=YN14(x12+1:end,1);

        bpoint(10,1:2)=SNHT_breaking_pos(XN11,YN11);
        x10=bpoint(10,1);

        % breaking the time series in two parts, 5/8 & 6/8, because
of the
        % identification of a significant shift

        if bpoint(10,1)>0

            bpoint(10,1)=bpoint(10,1)+x8;

            XN5=XN11(1:x10,1);
            YN5=YN11(1:x10,1);

            XN6=XN11(x10+1:end,1);
            YN6=YN11(x10+1:end,1);

            bpoint(9,1:2)=SNHT_breaking_pos(XN5,YN5);
            if bpoint(9,1)>0
            bpoint(9,1)=bpoint(9,1)+x8;
            end
            bpoint(11,1:2)=SNHT_breaking_pos(XN6,YN6);
```

```
            if bpoint(11,1)>0
            bpoint(11,1)=bpoint(11,1)+bpoint(10,1);
            end

        end

        bpoint(14,1:2)=SNHT_breaking_pos(XN12,YN12);
        x14=bpoint(14,1);

        if bpoint(14,1)>0

            % breaking the time series in two parts, 7/8 & 8/8,
because of the
            % identification of a significant shift

            bpoint(14,1)=bpoint(14,1)+bpoint(12,1);

            XN7=XN12(1:x14,1);
            YN7=YN12(1:x14,1);

            XN8=XN12(x14+1:end,1);
            YN8=YN12(x14+1:end,1);

            bpoint(13,1:2)=SNHT_breaking_pos(XN7,YN7);
            if bpoint(13,1)>0
            bpoint(13,1)=bpoint(13,1)+bpoint(12,1);
            end
            bpoint(15,1:2)=SNHT_breaking_pos(XN8,YN8);
            if bpoint(15,1)>0
            bpoint(15,1)=bpoint(15,1)+bpoint(14,1);
            end

        end

    end

end


% attribution of positive and negative shifts to the time series

shifts=zeros(ti,1);

for i=1:15
    if bpoint(i,1)>0
        p=bpoint(i,1);
        if bpoint(i,2)>0
            shifts(p)=77;
        elseif bpoint(i,2)<0
            shifts(p)=88;
        end

    end
end

MRshifts2=shifts;
```

```
end
```

5) <u>Συνάρτηση SCor</u>

```
function apotelesma4=Scor(InSeries,Uresult)

% Subroutine 5: correction of located shifts at candidate series

% extraction of candidate series
FI(:,1)=InSeries(:,2);
FF(:,1)=FI(:,1);

nj=0;

% correction of 1-15 possible shifts
for i=1:15
    if Uresult(i,1)>0
        th=Uresult(i,1);
        nj=nj+1;
        for k=1:th
            FF(k,1)=FF(k,1)+Uresult(i,2);
        end
    end
end

Njumps=zeros(length(FI),1);
Njumps(1,1)=nj;

% output: raw and adjusted candidate time series, number of
corrected
% shifts

apotelesma4=[FI FF Njumps];

end
```