ΥΔΡΟΣΚΟΠΙΟ

ΠΡΟΓΡΑΜΜΑ STRIDE ΕΛΛΑΣ

ΔΗΜΙΟΥΡΓΙΑ ΕΘΝΙΚΗΣ ΤΡΑΠΕΖΑΣ ΥΔΡΟΛΟΓΙΚΗΣ ΚΑΙ ΜΕΤΕΩΡΟΛΟΓΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΥΠΟΥΡΓΕΙΟ ΒΙΟΜΗΧΑΝΙΑΣ ΕΝΕΡΓΕΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΕΥΘΥΝΣΗ ΥΔΑΤΙΚΟΥ ΔΥΝΑΜΙΚΟΥ ΚΑΙ ΦΥΣΙΚΩΝ ΠΟΡΩΝ

MINISTRY OF INDUSTRY ENERGY AND TECHNOLOGY
WATER AND NATURAL RESOURCES DIRECTORATE

ΜΑΜΑΣΗΣ

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΟΜΟΓΕΝΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ, ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΣΦΑΛΜΑΤΩΝ, ΣΥΣΧΕΤΙΣΗΣ, ΔΙΟΡΘΩΣΗΣ ΑΠΟΚΛΙΣΕΩΝ ΚΑΙ ΣΥΜΠΛΗΡΩΣΗΣ ΕΛΛΕΙΠΟΥΣΩΝ ΤΙΜΩΝ ΧΡΟΝΟΣΕΙΡΩΝ ΔΕΔΟΜΕΝΩΝ (ΒΡΟΧΗ - ΧΙΟΝΙ)

METHODS OF REMOVING INHOMOGENEITIES, DETECTING ERRORS, DEFINING REGRESSION, REMOVING TREND AND FILLING IN MISSING DATA IN TIME SERIES OF RAINFALL AND SNOW

Μ. Γκίνη

M. Ghini

Αριθμός τεύχους 6/2,3
Report number

HYDROSCOPE

STRIDE HELLAS PROGRAMME

DEVELOPMENT OF A NATIONAL DATA BANK FOR HYDROLOGICAL AND METEOROLOGICAL INFORMATION

ΑΘΗΝΑ - ΙΑΝΟΥΑΡΙΟΣ 1993
ATHENS - JANUARY 1993

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

ΠΕΡΙΛΗΨΗ

Στην παρούσα μελέτη περιγράφονται μέθοδοι ανάλυσης ιστορικών χρονοσειρών βροχής και χιονιού ως προς την επιθυμητή ιδιότητα της τυχαιότητας (randomness) και της στασιμότητας (stationarity). Για την ανάλυση χρησιμοποιούνται διάφοροι μέθοδοι ανίχνευσης σφαλμάτων όπως γραφικές μέθοδοι, στατιστικές δοκιμές (tests), κλπ.

Επίσης περιγράφονται μέθοδοι ανίχνευσης της τάσης (trend) και της περιοδικότητας που τυχόν εμφανίζουν οι χρονοσειρές με διάφορες στατιστικές δοκιμές (tests), με στόχο την αφαίρεση και των δύο αυτών μη επιθυμητώνιδιοτήτων.

Αναπτύσσεται η έννοια της συσχέτισης χρονοσειρών καθώς και της αυτοσυσχέτισης δεδομένων της ίδιας χρονοσειράς και μελετώνται οι χρήσιμες πληροφορίες που παρέχουν οι συντελεστές αυτοσυσχέτισης.

Αναπτύσσεται η μέθοδος ελέγχου ομοιογένειας χρονοσειρών (μέθοδος διπλής αθροιστικής καμπύλης) καθώς και η διαδικασία ομογενοποίησης δεδομένων.

Επίσης αναπτύσσεται η μεθοδολογία συμπλήρωσης ελλειπουσών τιμών για χρονοσειρές παρατηρήσεων που αφορούν βροχή και χιόνι. Η χρήση της μεθοδολογίας αυτής είναι απαραίτητη για την προσθήκη δεδομένων σε περιπτώσεις χρονοσειρών με ελλειπή στοιχεία καθώς και για την αντικατάσταση δεδομένων που θεωρούνται αναξιόπιστα. Οι χρονοσειρές μετά τη συμπλήρωσή τους, μπορούν να χρησιμοποιηθούν ως είσοδοι (inputs) στα διάφορα υδρολογικά μοντέλα.

ABSTRACT

In the present study a time series analysis of observed rainfall and snow data is performed. The time series are examined if they satisfy the properties of randomness and stationarity. Different methods of error detection lare used such as graphical methods, statistical tests etc.

Also the time series are examined for trend and periodicity with various Statistical Tests in order to remove both these undesirable properties.

We present the concepts of time series regression and autoregression and the useful information given by the autoregression coefficients.

We also present the method of Double Mass Analysis curve for detecting the possible in homogeneities in a time series and the technique of removing them.

A methodology is also presented for filling in missing data in time series of rainfall and snow. The use of the above methodology is necessary for filling in missing data and for replacing not reliable data. After completion the time series can be used as an input to various hydrological models.

1 ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

1.1 Εισαγωγή

Η παρούσα μελέτη αναφέρεται κυρίως σε χρονοσειρές υδρομετεωρολογικών παραμέτρων, οι οποίες μετρούνται σε ίσα χρονικά διαστήματα (διακριτές).

Ως χρονοσειρά θεωρείται το σύνολο των παρατηρημένων τιμών μιας υδρομετεωρολογικής μεταβλητής, οι οποίες μετρούνται είτε σε ίσα διακριτά χρονικά διαστήματα είτε συνεχώς.

Διακριτές χρονοσειρές είναι δυνατόν να προκύψουν με διάφορους τρόπους. Από μια συνεχή χρονοσειρά εάν ληφθούν μόνο οι τιμές αυτής ανά ίσα χρονικά διαστήματα, προκύπτει μία διακριτή χρονοσειρά. Ενας άλλος τύπος διακριτής χρονοσειράς μπορεί να προκύψει όταν αντί για στιγμιαίες τιμές μίας παραμέτρου μετρούνται αθροιστικά οι τιμές αυτής της παραμέτρου κατά την διάρκεια ίσων χρονικών διαστημάτων (π.χ. ημερήσιες μετρήσεις βροχοπτώσεων).

Η θεωρία της Στατιστικής ασχολείται κυρίως με τυχαίες χρονοσειρές ή τυχαία δείγματα (random time series) από ανεξάρτητες παρατηρήσεις μεταξύ τους. Συνήθως όμως οι παρατηρήσεις σε συνεχή χρονικά διαστήματα δεν είναι ανεξάρτητες μεταξύ τους και τότε στην ανάλυση των χρονοσειρών πρέπει να ληφθεί υπόψη η χρονολογική σειρά των παρατηρουμένων τιμών. Για χρονοσειρές πλήρως εξαρτημένων μεταβλητών, οι μελλοντικές τιμές μπορούν να προβλεφθούν ακριβώς χρησιμοποιώντας τις παλαιότερες τιμές και αυτές οι χρονοσειρές καλούνται Προσδιοριστικές (Ντετερμινιστικές).

Για χρονοσειρές μαερικώς εξαρτημένων μεταβλητών όμως, οι μελλοντικές τιμές μόνον μερικώς μπορούν να προβλεφθούν χρησιμοποιώντας τις παλαιότερες τιμές και αυτές οι χρονοσειρές καλούνται Στοχαστικές.

Οι χρονοσειρές που ικανοποιούν την ιδιότητα της τυχαιότητας (randomness) μελετούνται πιο εύκολα.

Προκειμένου μια χρονοσειρά να θεωρείται τυχαία πρέπει κατ΄αρχήν να ικανοποιεί την ιδιότητα της στασιμότητας (stationarity).

Επειδή το μεγαλύτερο μέρος της Θεωρείας των Πιθανοτήτων αναφέρεται σε στάσιμες χρονοσειρές, κατά την ανάλυση χρονοσειρών συχνά απαιτείται η μετατροπή μη στάσιμων σε στάσιμες χρονοσειρές για να μελετηθούν πιο εύκολα.

Μια χρονοσειρά ονομάζεται στάσιμη (stationary) όταν τα στατιστικά χαρακτηριστικά της είναι ανεξάρτητα του χρόνου. Αυτό σημαινει ότι οι στατιστικές ιδιότητες (μέσος όρος, τυπική απόκλιση,κλπ) μιας στάσιμης χρονοσειράς υπολογιζόμενες από διαφορετικά κομμάτια διατηρούνται σταθερές και δεν αλλάζουν δηλαδή είναι ανεξάρτητες του χρόνου.

Χρονοσειρές που εμφανίζουν τάση (trend), περιοδικότητα (periodicity), άλματα των τιμών (jumbs), ή τιμές που απέχουν πολύ από τις αναμενόμενες (outliers), προφανώς δεν είναι στάσιμες.

Στο παρόν κεφάλαιο αναλύονται οι μέθοδοι ανίχνευσης των στοχαστικών σφάλματων που τυχόν παρουσιάζουν οι χρονοσειρές βροχοπτώσεων ή χιονιού. Στο Κεφάλαιο 2 παρουσιάζεται η μέθοδος εντοπισμού των συστηματικών σφαλμάτων σε χρονοσειρές βροχής ή χιονιού.

Στις επόμενες παραγράφους του παρόντος Κεφαλαίου περιγράφονται αναλυτικά οι γενικές μέθοδοι ανίχνευσης σφαλμάτων, οι μέθοδοι ανίχνευσης τάσεων και περιοδικοτήτων χρονοσειρών, οι μέθοδοι ανίχνευσης τυχαιότητας χρονοσειρών με την βοήθεια αυτοσυσχετογράμματος, καθώς και άλλες μέθοδοι ανίχνευσης τυχαιότητας (randomness) αυτών. Στις παραπάνω μεθόδους περιλαμβάνονται από πολύ απλές τεχνικές όπως η παραγωγή γραφημάτων της εξέλιξης των τιμών κατά την διάρκεια του χρόνου έως πιο σύνθετες όπως οι στατιστικές δοκιμές (tests) για τον έλεγχο της στασιμότητας των χρονοσειρών.

1.2 Γενικές μέθοδοι ανίχνευσης σφαλμάτων

1.2.1 Εκτύπωση Χρονοσειρών σε λίστες και υπολογισμός στατιστικών παραμέτρων αυτών.

Κατ' αρχάς οι υπό εξέταση χρονοσειρές τυπώνονται σε λίστες με παράλληλη διάταξη, όπως για παράδειγμα οι χρονοσειρές βροχοπτώσεων που αφορούν μια ή περισσότερες υδρολογικές λεκάνες τυπώνονται σε λίστες δεδομένων με παράλληλη διάταξη για την ίδια χρονική περίοδο, με σκοπό να εντοπισθούν οι τυχόν μετατοπίσεις (shifts) των τιμών χρονικά. Αυτές οι μετατοπίσεις είναι πιθανόν να οφείλονται στην καταγραφή των παρατηρουμένων τιμών με λάθος ημερομηνία. Παράλληλα εντοπίζονται και οι τιμές που λείπουν και χρειάζεται να συμπληρωθούν καθώς και οι τιμές που είναι έξω από τα όρια της πραγματικότητας και χρειάζεται να διορθωθούν. Στην συνέχεια υπολογίζονται οι στατιστικές παράμετροι των χρονοσειρών όπως ο μέσος όρος, η τυπική απόκλιση, η διασπορά, ο συντελεστής διασποράς, ο συντελεστής αυτοσυσχέτισης καθώς και οι μέγιστες και ελάχιστες τιμές,το ανώτερο και κατώτερο όριο, ο αριθμός δεδομένων κάτω από το κατώτερο όριο και αντίστοιχα ο αριθμός δεδομένων πάνω από το ανώτερο όριο.

Ως ανώτερο και κατώτερο όριο μιας χρονοσειράς ορίζεται η μέγιστη απόκλιση δεδομένων από την μέση τιμή, συνήθως δύο φορές την τυπική απόκλιση περισσότερο ή λιγότερο από τον μέσο όρο και βάσει αυτής της μέγιστης απόκλισης τα δεδομένα χωρίζονται σε αποδεκτά ή μη. Αποδεκτά είναι εκείνα τα δεδομένα τα οποία κείνται μεταξύ του ανώτερου και του κατώτερου ορίου των τιμών, όπως αυτά ορίζονται από τις σχέσεις (1) :

$$x^{+} = m_X + a.S_X \text{ (συνήθως α=2)} \tag{1.1}$$

$$x^{-} = m_X - \beta.S_X \text{ (συνήθως β=2)} \tag{1.2}$$

οπου x^{+}, x^{-} είναι το ανώτερο και κατώτερο όριο αντιστοιχα, α και β παράμετρες, m_X ο μέσος όρος και S_X η τυπική απόκλιση της χρονοσειράς X.

Τα δεδομενα που οι τιμές τους είναι έξω από τα προαναφερθέντα όρια (outliers), είναι εκ πρώτης όψεως μη αποδεκτά και υπόκεινται σε περαιτέρω έλεγχο για την αξιοπιστία τους.

1.2.2 Γραφικές μέθοδοι ανίχνευσης σφαλμάτων.

Πέραν από την εκτύπωση χρονοσειρών σε λίστες, η παρουσίαση αυτών σε γραφήματα δίνει μια γρήγορη οπτική εντύπωση για τα δεδομένα που είναι πιθανόν εσφαλμένα, όπως για παράδειγμα διάφορες μέγιστες ή ελάχιστες τιμές σε χρονικές περιόδους που δεν δικαιολογούνται (π.χ. μέγιστη βροχόπτωση τον Αύγουστο), καθώς και τιμές ασυνήθιστα υψλές ή χαμηλές έξω από τα ανώτερα και κατώτερα όρια που δικαιολογούνται ενδεχομένως ως τυπογραφικά λάθη ή ως σφάλματα μετρήσεων.

Παρατηρώντας τα διαγράμματα χροσειρών είναι δυνατόν να εντοπισθούν τμήματα (κομμάτια) χρονοσειράς με διαφορετικούς μέσους όρους, φαινόμενα τα οποία θα διερευνηθούν περαιτέρω με την Δοκιμή t-student.

Οι κυριότερες γραφικές μέθοδοι που χρησιμοποιούνται για την ανίχνευση σφαλμάτων υδρομετεωρολογικών δεδομένων με σκοπό την αξιολόγηση αυτών είναι οι ακόλουθες:

(α) σχεδιάγραμμα τιμών χρονοσειράς κατά την διάρκεια του χρόνου

(β) σχεδιάγραμμα υπολοίπων χρονοσειράς κατά την διάρκεια του χρόνου

Στις επόμενες παραγράφους αναλύονται οι προαναφερθείσες γραφικές μέθοδοι.

(α) Σχεδιάγραμμα τιμών χρονοσειράς κατά τη διάρκεια του χρόνου.

Δύο ή περισσότερες χρονοσειρές γειτονικών σταθμών σχεδιάζονται στο ίδιο σχεδιάγραμμα για το ίδιο χρονικό διάστημα, το οποίο παρέχει μια γρήγορη οπτική εντύπωση (εικόνα) για την κάθε χρονοσειρά καθώς και την κάθε μια χρονοσειρά συγκριτικά με τις άλλες. Με τον τρόπο αυτό διερευνάται εκτός των άλλων και η μετατόπιση παρατηρήσεων λόγω σφάλματος. (σχήμα 1.1)

(β) Σχεδιάγραμμα υπολοίπων χρονοσειράς κατά την διάρκεια του χρόνου

Χρονοσειρά υπολοίπων (residual series) είναι η χρονοσειρά που προκύπτει από τις διαφορές των παρατηρημένων τιμών από τον μέσο όρο. Σχεδιάζοντας μια χρονοσειρά υπολοίπων (residuals) σχετικά με τον μέσο όρο της χρονοσειράς, όπως φαίνεται στο σχήμα 1.2 αποκτάται μια γρήγορη οπτική εντύπωση όσον αφορά τις υγρές και ξηρές περιόδους.

1.3 Στατιστικές Δοκιμές (tests) για την ανίχνευση τάσης (trend) και περιοδικότητας χρονοσειρών.

1.3.1 Στατιστικές Δοκιμές για την ανίχνευση τάσης (trend) χρονοσειρών.

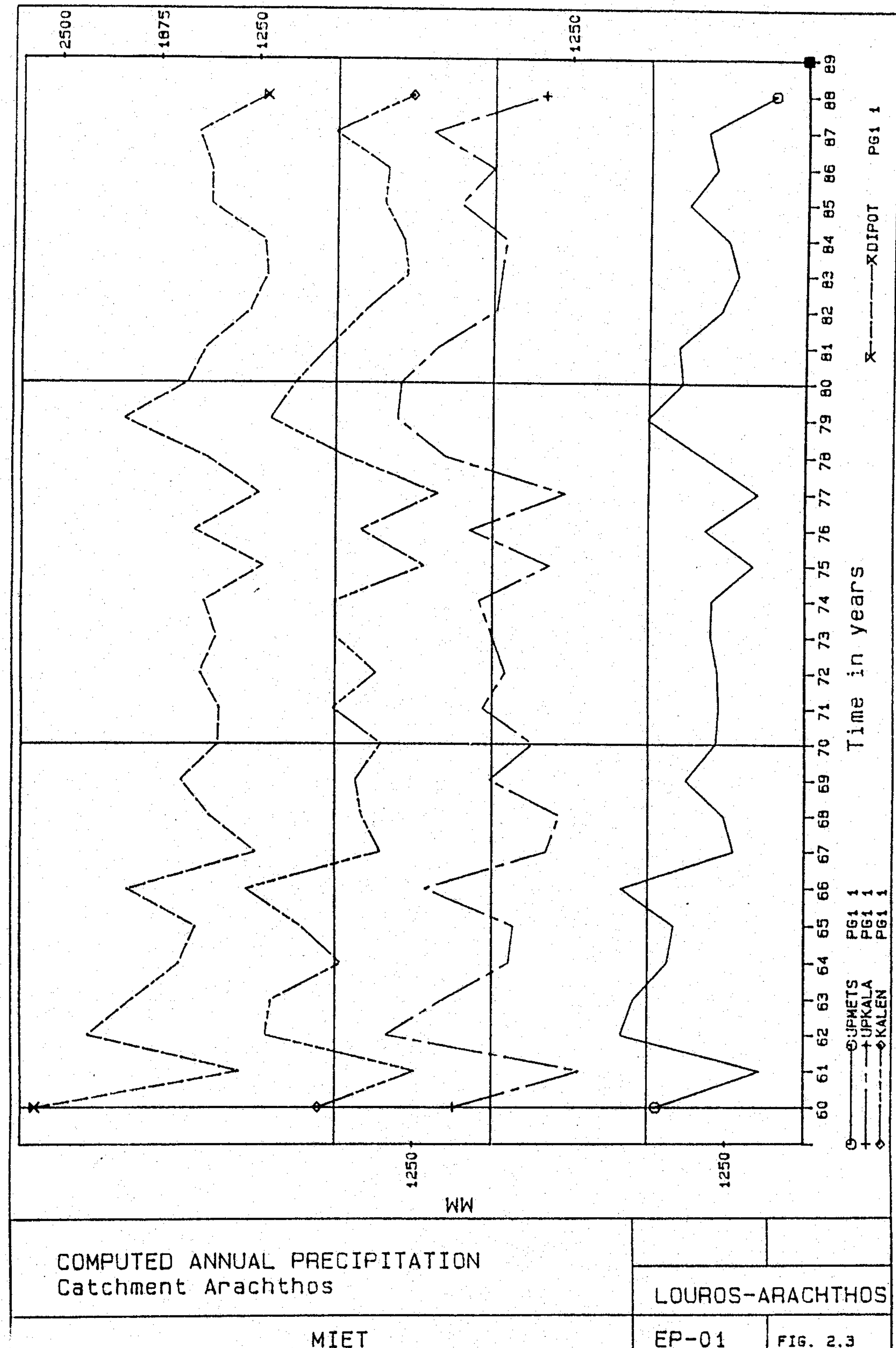
(α) Εισαγωγή

Ενας από τους κύριους στόχους της Ανάλυσης Χρονοσειρών είναι η αφαίρεση από τις χρονοσειρές της τάσης (trend) και της περιοδικότητας (εποχιακή ή άλλου είδους) αυτών εάν υπάρχουν.

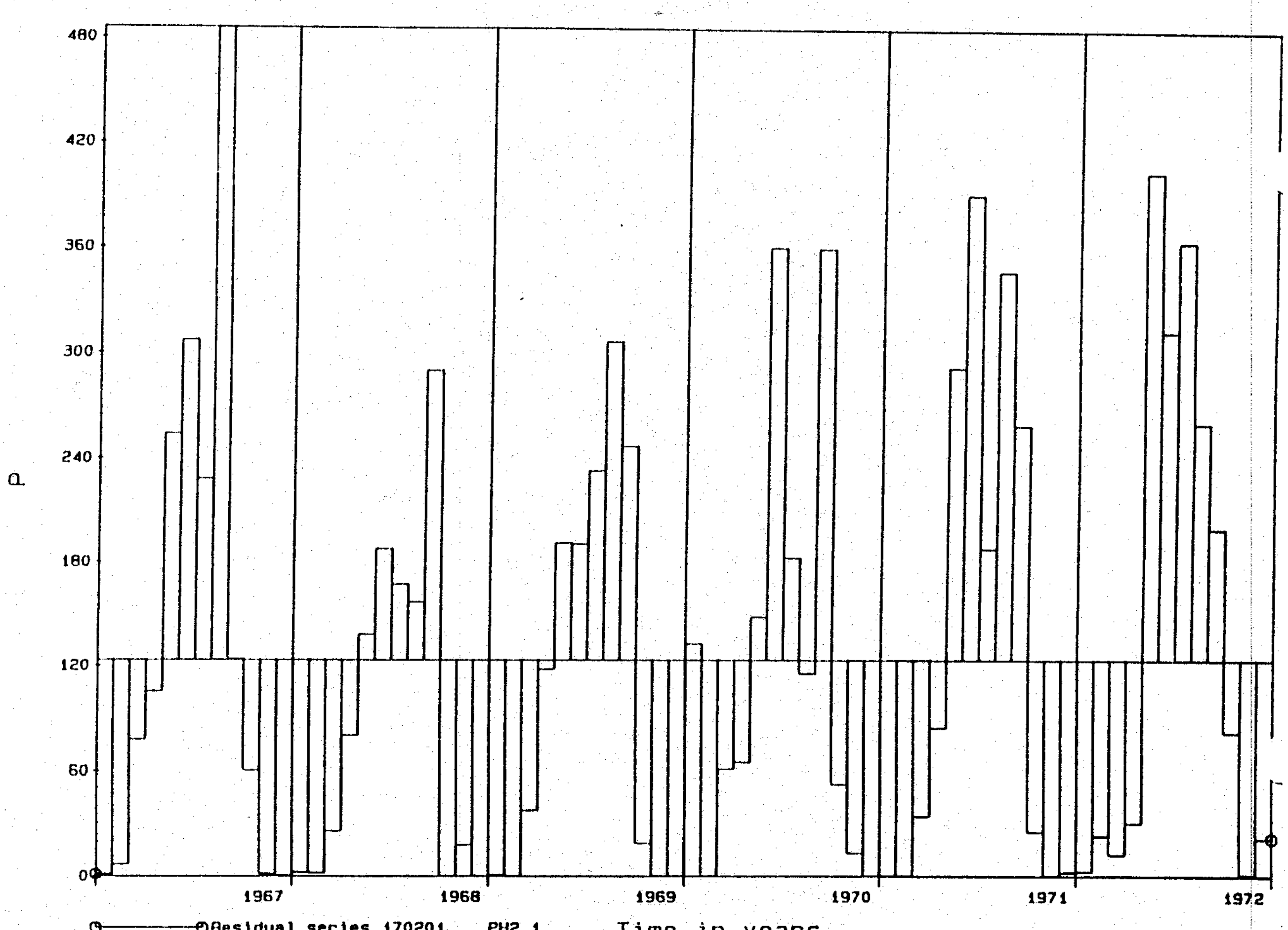
Μια χρονοσειρά παρουσιάζει πτωτική ή ανοδική τάση όταν οι παρατηρημένες τιμές της κατά μέσο όρο είτε αυξάνουν είτε μειώνονται. Αυτού του είδους η συμπεριφορά μιας χρονοσειράς μπορεί να είναι παροδική, όταν η εμφανιζόμενη τάση οφείλεται σε αλλαγές για κάποιο μικρό χρονικό διάστημα, αλλά μπορεί να είναι και μονιμότερη, όταν η παρατηρημένη τάση διαρκεί για ενα μεγάλο χρονικό διάστημα. Στην περίπτωση των υδρομετεωρολογικών παρατηρήσεων αυτή η τάση μπορεί να οφείλεται σε μακράς χρονικής διάρκειας κλιματικές αλλαγές λόγω καταστροφής των τροπικών δασών και απερήμωσης μεγάλων εκτάσεων, στις οποίες προηγουμένως υπήρχε φυτοκάλυψη.

Είναι σκόπιμο να αναφερθεί ότι στην ανάλυση χρονοσειρών για ύπαρξη τάσης, συνιστάται να χρησιμοποιούνται δείγματα ετήσιων παρατηρήσεων, τα οποία είναι απηλλαγμένα από τις εντός του έτους διακυμάνσεις.

Η ύπαρξη τάσης -πτωτικής ή ανοδικής- σε μια χρονοσειρά μόνο για ορισμένες περιπτώσεις είναι προφανής. Συχνά απαιτείται περαιτέρω έλεγχος για να διαπιστωθεί αν τα συστηματικά σφάλματα (suspected systematic effects) που εντοπίζονται σε χρονοσειρές είναι σημαντικά ή όχι, και για το λόγο αυτό υπάρχει διαθέσιμος ένας σημαντικός αριθμός Δοκιμών (tests).



Σχήμα 1.1. Χρονοσειρές ετησίων βροχοπτώσεων στην υδρολογική λεκάνη του Αράχθου (4).



Σχήμα 1.2. Χρονοσειρά υπολοίπων βροχομετρικών παρατηρήσεων (residuals) (1).

Στην παρούσα μελέτη διατίθενται οι ακόλουθες τρεις Δοκιμές :

(i) Δοκιμή Turning Point, η οποία δεν είναι αποτελεσματική για την διερεύνηση τάσης αλλά θεωρείται χρήσιμη πρωταρχική δοκιμή για την τυχαιότητα (randomness) της χρονοσειράς.

(ii) Δοκιμή (Kendall) και

(iii) Δοκιμή γραμμικής συσχέτισης για γραμμική τάση. Οι δύο τελευταίες δοκιμές χρησιμοποιούνται για να ελέγχουν αν υπάρχει τάση στην χρονοσειρά (πτωτική ή ανοδική).

(β) Δοκιμή Turning Point

Σε χρονοσειρά βροχομετρικών παρατηρήσεων X_t, t=1,2,...N, υπάρχει σημείο στροφής (turning point) σε χρόνο t=i, όταν η τιμή της παρατήρησης X_i είναι μεγαλύτερη από τις τιμές και των δύο παρατηρήσεων X_{i-1} και X_{i+1} ή αντίστοιχα μικρότερη και από τις δύό αυτές τιμές. Ως σημείο στροφής (turning point) ορίζεται το σημείο που η χρονοσειρά εμφανίζει τοπικό μέγιστο (peak) ή ελάχιστο (trough). Ο αριθμός των σημείων στροφής N_t μιας στάσιμης χρονοσειράς από ανεξάρτητα γεγονότα ακολουθεί κανονική κατανομή με μέση τιμή μ_t και τυπική απόκλιση σ_t που δίνονται από τις σχέσεις (3) :

$$\mu_t = 2(N-2)/3 \tag{1.3}$$

$$\sigma_t^2 = (16N-29)/90 \tag{1.4}$$

Επειδή πολλά σημεία επαφής ή αντίθετα πολύ λίγα σε μια χρονοσειρά δημιουργούν αμφιβολίες για την τυχαιότητα της χρονοσειράς, για το λόγο αυτό ελέγχεται η ακόλουθη υπόθεση:

H_0 : η χρονοσειρά x_t είναι τυχαία και

H_1 : η χρονοσειρά x_t δεν είναι τυχαία, οπότε εφαρμόζεται ένας αμφίπλευρος έλεγχος.

Υπολογίζεται η απόλυτη τιμή $|u| = |N_t - \mu_t| / \sigma_t$

και στην περίπτωση του αμφίπλευρου ελέγχου η μηδενική υπόθεση περί τυχαιότητας της χρονοσειράς, δεν απορρίπτεται για κάποιο επίπεδο σημαντικότητας, α, όταν ισχύει η σχέση $|u| \leq (1\text{-}a/2)$

(γ) Δοκιμή Kendall

Σε χρονοσειρά βροχομετρικών παρατηρήσεων X_t, t=1,2....N εξετάζονται όλα τα δυνατά ζεύγη αυτής X_i, X_j με $j > i$ και υπολογίζεται ο συνολικός αριθμός p των ζευγών που πληρούν τησχέση $X_j > X_i$. O αριθμός p μεγαλώνει όταν η χρονοσειρά είναι συνεχώς αύξουσα. Για μια χρονοσειρά που δεν παρουσιάζει τάση αποδεικνύεται ότι η αναμενόμενη τιμή του p δίνεται από τη σχέση (3,5) : E(p) = N (N-1) / 4.

Στη συνέχεια υπολογίζεται η ανηγμένη μεταβλητή τ=[4p/N(N-1)]-1 η οποία έχει αναμενόμενη τιμή E(τ) = 0 για μια τυχαία διαδοχή της χρονοσειράς. Η διασπορά του τ δίνεται από την σχέση (5) :

Var(τ) = 2(2N+5)/9N(N-1)

και αποδεικνύεται ότι η κατανομή της παραμέτρου Kendall $\tau/\{var(\tau)\}^{1/2}$ συγκλίνει στην κανονική κατανομή όσο το N μεγαλώνει. Στην περίπτωση του μονόπλευρου ελέγχου η μηδενική υπόθεση H_0 περί μη ύπαρξης πτωτικής/ανοδικής τάσης δεν απορρίπτεται για κάποιο επίπεδο σημαντικότητας, α, όταν ισχύει η σχέση $Z_\alpha \leq \tau/var(\tau)^{1/2}$ ή $\tau/var(\tau)^{1/2} \leq Z_\alpha$ αντίστοιχα, όπου Z_α η ανηγμένη μεταβλητή της κατανομής Gauss για πιθανότητα υπέρβασης α.

(δ) Δοκιμή γραμμικής συσχέτισης για γραμμική τάση (Regression test).

Με την παραδοχή ότι η τάση είναι προσεγγιστικά γραμμική εφαρμόζεται η δοκιμή γραμμικής συσχέτισης για γραμμική τάση. Ετσι η τιμή X_τ της χρονοσειράς βροχομετρικών παρατηρήσεων X_1, X_2, ... X_τ, ...X_N, τη χρονική στιγμή τ μπορεί να εκφραστεί από τη σχέση $X_\tau = x_o + \beta + \xi_\tau$ οπου x_o, β σταθερές, τ ο χρόνος σε διακριτά διαστήματα και ξ_τ ένα στοχαστικό υπόλοιπο με μέση τιμή μηδέν. Η σειρά των στοχαστικών υπολοίπων θεωρείται ότι είναι στάσιμη, διαδοχικά ανεξάρτητη και ακολουθεί την κανονική κατανομή.

Σε μια χρονοσειρά βροχομετρικών παρατηρήσεων αποδεικνύεται ότι (5) :

$$\hat{\beta} = \frac{\sum(t-t_\mu)(x_t - x_\mu)}{\sum(t-t_\mu)^2} \quad \text{και} \quad S_\beta^2 = \frac{\sum(x_t - x_\mu) - \hat{\beta}^2 \sum(t-t_\mu)^2}{(n-2)\sum(t-t_\mu)^2} \qquad (1.5)$$

όπου β και S_β είναι η εκτίμηση του β και της διασποράς του αντίστοιχα ενώ $t_\mu = \Sigma_t/N$, $x_\mu = \Sigma_x/N$, όπου αθροίζονται όλα τα διαθέσιμα στοιχεία. Στην περίπτωση του μονόπλευρου ελέγχου η μηδενική υπόθεση περί μη ύπαρξης πτωτικής/ανοδικής τάσης δεν απορρίπτεται για κάποιο επίπεδο σημαντικότητας, α, όταν ισχύει η σχέση $C_a \leq \beta/S_\beta$ ή $\beta/S_\beta \leq C_a$ αντίστοιχα. Σημειώνεται ότι η κατανομή Student προσεγγίζει την κανονική κατανομή για μεγάλο N, όπου C_a η ανηγμένη μεταβλητή της κατανομής Student για N-2 βαθμούς ελευθερίας και για πιθανότητα υπέρβασης α.

1.3.2 Στατιστικές δοκιμές για την ανίχνευση περιοδικότητας χρονοσειρών.

Η περιοδικότητα που εμφανίζουν χρονοσειρές βροχοπτώσεων είναι κατά κανόνα εποχιακή, που οφείλεται σε εποχιακές αλλαγές στην ίδια χρονιά. Ορισμένες όμως χρονοσειρές παρουσιάζουν κάποια ακανόνιστη υπερετήσια περιοδικότητα, της οποίας τα αίτια είναι άγνωστα.

Η περιοδικότητα των χρονοσειρών είναι φαινόμενο προσδιοριστικό (ντετερμινιστικό) ως προς την συχνότητα, διότι επιβάλλεται στην χρονοσειρά από ένα κυκλικό φυσικό φαινόμενο, όπως η κίνηση της γης γύρω από τον ήλιο ή η κίνηση της σελήνης γύρω από τη γη.

Στη παρούσα μελέτη αναλύονται δύο μέθοδοι ανίχνευσης περιοδικότητας χρονοσειρών. Η πρώτη καλείται μέθοδος της αρμονικής ανάλυσης, κατά την οποία τα περιοδικά φαινόμενα αναπαρίστανται με ένα σύνολο ημιτονοειδών συναρτήσεων, των αρμονικών. Η δεύτερη μέθοδος βασίζεται στη συνάρτηση πυκνότητας φάσματος με βάση το αυτοσυσχετόγραμμα:

(α) Μέθοδος αρμονικής ανάλυσης

Σύμφωνα με τη μέθοδο της αρμονικής ανάλυσης για κάθε χρονοσειρά βροχομετρικών δεδομένων x_1, x_2....x_N, η οποία έχει περιοδική συμπεριφορά, υπάρχει ένας πεπερασμένος αριθμός αρμονικών αυτής x_i, i=1,2,3......L που δίνονται από τη σχέση (3) :

$$x_i = \mu + \sum_{i=1}^{L} \lambda_i \sin\{(2\pi t/T)i + \Phi_i\} + \xi_i, \qquad (1.6)$$

όπου λ_i το πλάτος (ίσο με το μισό ύψος της ημιτονοειδούς καμπύλης κάθε αρμονικής), Φ_i η φάση, μ ο μέσος όρος, i/T η συχνότητα επαναφοράς, T/i, το μήκος κύματος (ίσο με το χρόνο μεταξύ δύο διαδοχικών ακραίων τιμών) και ξ_i ο στοχαστικός όρος.

Στην πράξη συχνά βρίσκουμε ότι οι περιοδικότητες των χρονοσειρών είναι δυνατόν να αναπαρίστανται με μία ή δύο αρμονικές για μηνιαία δεδομένα και με τέσσερις ή έξι αρμονικές για ημερήσια δεδομένα. Σε αυτές τις περιπτώσεις οι άλλες αρμονικές θεωρούνται ότι είναι θόρυβος και συμπεριλαμβάνονται στον στοχαστικό όρο ξ_i. Αντικειμενικά, ο ακριβής αριθμός των αρμονικών που απαιτείται να υπολογισθούν για κάθε περίπτωση δεδομένων δίνεται από την ανάλυση της διασποράς σύμφωνα με Kottegoda, N.T (1980).

(β) Μέθοδος φασματικής ανάλυσης με βάση το αυτοσυσχετόγραμμα.

Μια χρονοσειρά βροχομετρικών παρατηρήσεων μπορεί να θεωρηθεί ότι αποτελείται από ένα συνδυασμό βασικών συχνοτήτων με τις οποίες είναι πιθανόν να συμβούν τα τυχαία γεγονότα (βροχοπτώσεις). Αντικειμενικός σκοπός της φασματικής ανάλυσης είνάι ο υπολογισμός των συχνοτήτων από τις οποίες συνίσταται μία χρονοσειρά και του αντίστοιχου πλάτους καθε συχνότητας. Η χρονοσειρά αποσυντίθεται σε ένα σύνολο βασικών συχνοτήτων, μέσω του φάσματος. Βασικά, το φάσμα είναι δυνατόν να ορισθεί χρησιμοποιώντας την μέθοδο της αρμονικής ανάλυσης όπως αυτή αναλύθηκε στην προηγούμενη παράγραφο.

Εστω N παρατηρήσεις με ίσα χρονικά διαστήματα Δt μεταξύ αυτών, όπου για την βασική περίοδο $p = N.\Delta t$ οι αρμονικές έχουν συχνότητες $i = 1,2,3....p/2$ κύκλους για την περίοδο p. Το διάγραμμα του ημιαθροίσματος των πλατών $(a_i^2+b_i^2)/2$ ως προς την συχνότητα i, όπου a_i και b_i οι τιμές των υπολογισθέντων σταθερών όρων της κάθε αρμονικής, είναι το φάσμα που υπολογίζεται από το δείγμα, ενώ η μαθηματική συνάρτηση που προσαρμόζεται πάνω στα σημεία αυτά καλείται συνάρτηση πυκνότητας φάσματος.

Η συμβατική μέθοδος υπολογισμού του φάσματος περιλαμβάνει την αρμονική ανάλυση του αυτοσυσχετογράμματος. Αυτοσυσχετόγραμμα είναι το διάγραμμα σχεδιασμού των συντελεστών αυτοσυσχέτισης r_K μιας χρονοσειράς συναρτήσει του χρονικού βήματος κ, όπως αυτό αναλύεται λεπτομερώς στον Kottegoda, N.T (1980), κεφάλαιο 4.

Η αρμονική ανάλυση του αυτοσυσχετογράμματος επιτελείται βάσει της διαπίστωσης ότι οι περιοδικότητες που εμφανίζουν οι χρονοσειρές διατηρούνται διά μέσω της συνάρτησης αυτοσχέτισης. Αν και κατά την ανωτέρω ανάλυση η πληροφορία που αφορά την διαφορά φάσης δεν διατηρείται κατά την διαδικασία του υπολογισμού, αυτό δεν είναι σημαντικό για την περίπτωση στάσιμων χρονοσειρών.

1.4 Μέθοδοι ανίχνευσης τυχαιότητας (randomness) χρονοσειρών

1.4.1 Αλλες μέθοδοι ανίχνευσης τυχαιότητας (randomness) χρονοσειρών.

Αλλες μέθοδοι ανίχνευσης τυχαιότητας (randomness) χρονοσειρών που αναλύονται στην παρούσα μελέτη είναι :

(α) αθροιστική καμπύλη υπολοίπων (residuals)

(β) καμπύλη μετακινουμένου μέσου όρου (moving average) (γ) Δοκιμή t-Student

(α) Αθροιστική καμπύλη υπολοίπων (residuals) χρονοσειράς

Η αθροιστική καμπύλη των υπολοίπων (residuals) χρονοσειράς από το μέσο όρο, παρουσιάζει τις αθροιστικές αποκλίσεις των τιμών από το μέσο όρο και είναι ένα αποτελεσματικό εργαλείο για τον εντοπισμό τάσης (trend) της χρονοσειράς,που πιθανόν να οφείλεται σε κλιματικές αλλαγές. Μια χρονοσειρά παρουσιάζει τάση (πτωτική ή ανοδική) όταν οι παρατηρημένες τιμές αυτής κατά μέσο όρο είτε αυξάνουν είτε μειώνονται. Στο Κεφάλαιο 3 αναλύεται με μεγαλύτερη λεπτομέρεια η τάση χρονοσειράς καθώς και τα πιθανά αίτια που την προκαλούν.
Η αθροιστική καμπύλη των υπολοίπων (residuals) υπολογίζεται από την σχέση :

$$Y_{x,i} = Y_{x,i-1} + (X_i - m_x) = \sum_{j=1}^{i} (X_{j-1}/N\, \Sigma X_K) \qquad (1.7)$$

όπου N=αριθμός των παρατηρήσεων της χρονοσειράς X_i.

Η αθροιστική καμπύλη των υπολοίπων ερμηνεύεται ως ακολούθως :

- καμπύλη ανερχόμενη δείχνει μια χρονοσειρά πάνω από το μέσο όρο.
- καμπύλη οριζόντια δείχνει μια χρονοσειρά περίπου στον μέσο όρο.
- καμπύλη κατερχόμενη δείχνει μια χρονοσειρά κάτω από το μέσο όρο.

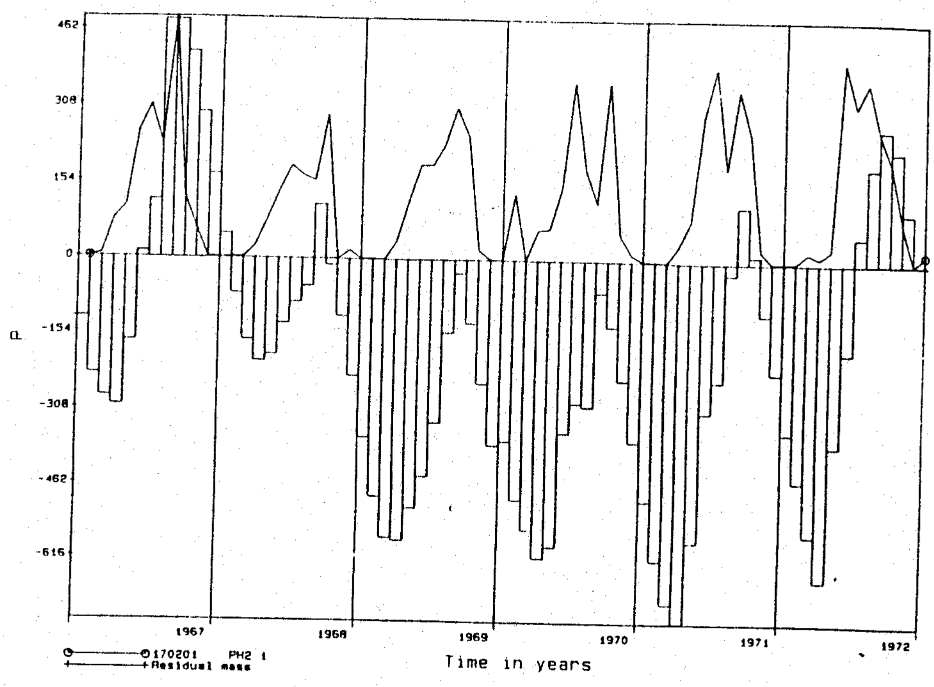
Ενα παραδειγμα αθροιστικής καμπύλης των υπολοίπων φαίνεται στο Σχ. 1.3.

(β) Καμπύλη κινούμενου μέσου όρου (moving average)

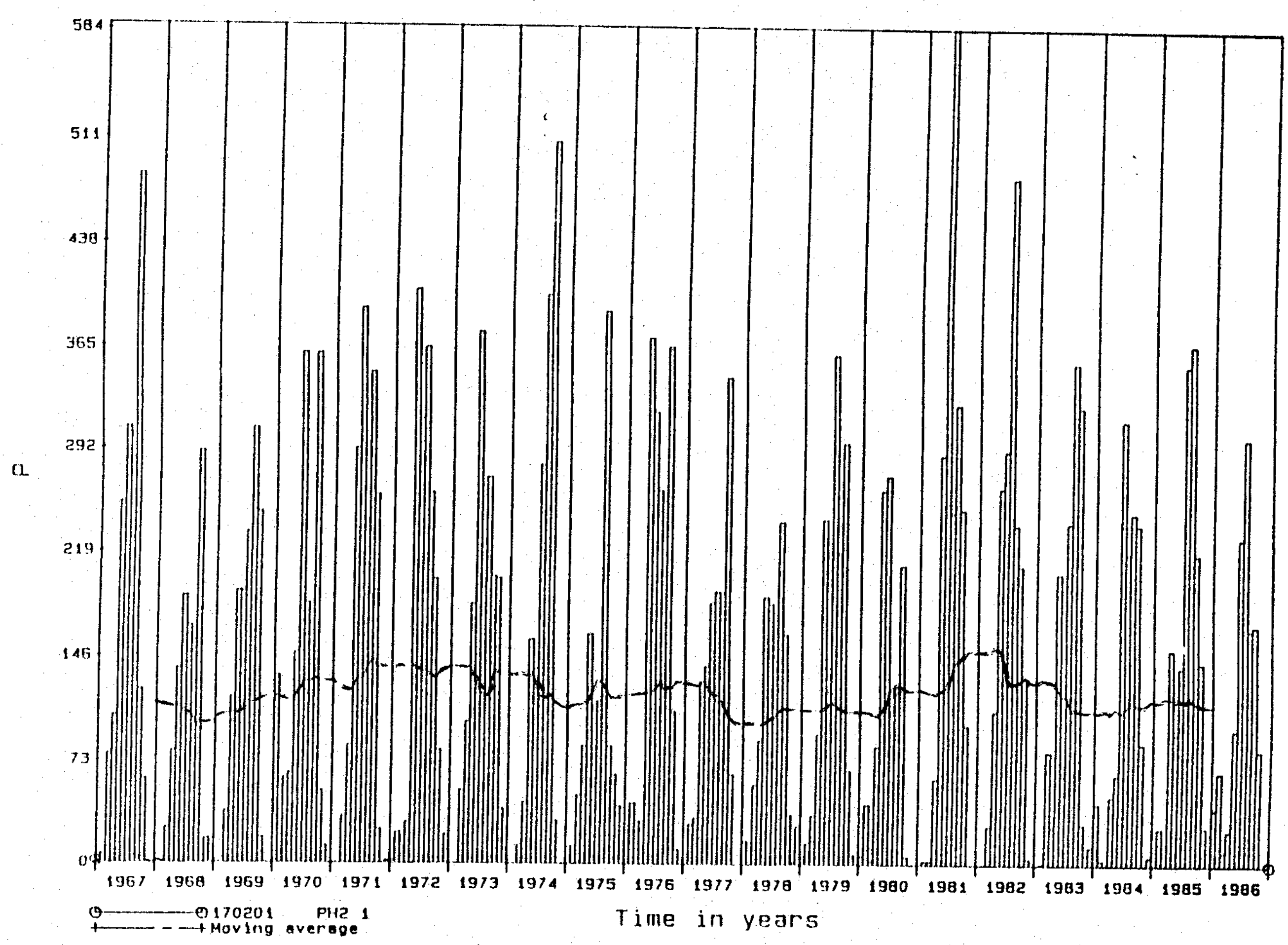
Η χρονοσειρά κινούμενου μέσου όρου (moving average series) Y_t, για την απλούστερη περίπτωση του γραμμικού μετασχηματισμού δίνεται από τη σχέση:

$$Y_t = 1/(2k+1)\sum_{j=-k}^{k} X_{t+j}, \qquad (1.8)$$

όπου X_t είναι η αρχική χρονοσειρά, η οποία μετασχηματίζεται στην Y_t, όπου η τάση (trend) υπολογίζεται για ένα περιορισμένο διάστημα από 2k+1 τιμές και επιλέγεται ένα βέλτιστο k, έτσι ώστε οι μικρές διακυμάνσεις να απομακρύνονται και να επιτυγχάνεται ο εντοπισμός της τάσης (trend).



Σχήμα 1.3. Αθροιστική καμπύλη των υπολοίπων (residuals) (1).



Σχήμα 1.4 Καμπύλη κινούμενου μέσου όρου (moving average) (1).

Οταν η μέση περίοδος της ταλάντωσης είναι 2k+1 τότε η μετασχηματιζόμενη χρονοσειρά Y_t υπολογίζεται σύμφωνα με την προαναφερθείσα σχεση. Οταν η περίοδος είναι 2k τότε η τάση υπολογίζεται επίσης για 2k+1 τιμές, αλλά χρειάζονται διπλά βάρη για όλους τους όρους της εξίσωσης του μετασχηματισμού εκτός από τον πρώτο και τον τελευταίο, όπως φαίνεται στην παρακάτω εξίσωση:

$$Y_t=(X_{t-k}+2X_{t-k+1}+2X_{t-k+2}+...+2X_{t+k-2}+2X_{t+k-1}+X_{t+k})/4k \qquad (1..9)$$

Στο σχήμα 1.4 δίνεται ένα παράδειγμα καμπύλης κινούμενου μέσου όρου, η οποία σχεδιάζεται στο ίδιο διάγραμμα με την αρχική χρονοσειρά (ιστορικά δεδομένα).

(γ) Δοκιμή t-Student

Η δοκιμή t-Student χρησιμοποιείται προκειμένου να ελεγχθεί αν δύο χρονοσειρές βροχοπτώσεων έχουν τον ίδιο μέσο όρο, το οποίο είναι απαραίτητη προϋπόθεση στην περίπτωση συμπλήρωσης ελλειπουσών τιμών μιας χρονοσειράς, από τιμές χρονοσειράς γειτονικού σταθμού για το ίδιο χρονικό διάστημα. Στην περίπτωση που δύο γειτονικές χρονοσειρές, έχουν μεγάλη συσχέτιση αλλά διαφορετικό μέσο όρο, τότε η συμπλήρωση ελλειπουσών τιμών της μιας από την άλλη γίνεται αφού πρώτα γίνει η διόρθωση λόγω διαφοράς μέσου όρου.

Η Δοκιμή t-Student επίσης χρησιμοποιείται προκειμένου να ελεγχθεί αν δύο τμήματα (κομμάτια) της ίδιας χρονοσειράς έχουν διαφορετικό μέσο όρο.

Εστω οι χρονοσειρές βροχομετρικών παρατηρήσεων A_i, (i=1,m) και B_i(i=1,n),για τις οποίες υπολογίζεται ο μέσος όρος δείγματος m_A και m_B αντίστοιχα. Αν μ_A και μ_B είναι οι μέσοι όροι των χρονοσειρών A_i και B_i αντιστοιχα, τότε η μηδενική υπόθεση H_0 περί ισότητας των μέσων όρων εξετάζεται έναντι της υπόθεσης H_1 περί ανισότητας των μέσων όρων των χρονοσειρών.

$H_0 : \mu_A = \mu_B$, και
$H_1 : \mu_A \neq \mu_B$, οπότε γίνεται ένας αμφίπλευρος έλεγχος.

Πρός το σκοπό αυτό υπολογίζεται η απόλυτη τιμή της παραμέτρου t , η οποία υποθέτωντας ότι η μηδενική υπόθεση, H_0 είναι αληθινή ακολουθείμία t-Student κατανομή με N βαθμούς ελευθερίας όπου,

$N = m+X_{n-2}$ για N›10.

$$|t| = |m_A - m_B| / S_{AB} \qquad (1.10)$$

Ο τρόπος με τον οποίο υπολογίζεται η διασπορά S_{AB} εξαρτάται από το αν οι χρονοσειρές A_i και B_i έχουν την ίδια διασπορά. Πρός το σκοπό αυτό ο λόγος των δύο διασπορών Q , εξετάζεται με τη Δοκιμή Fisher (F-test), οπου :

$$Q = S_A^2/S_B^2 \tag{1.11}$$

Εξετάζονται οι ακόλουθες υποθέσεις:

Ho : $\sigma_A^2 = \sigma_B^2$ και

H1 : $\sigma_A^2 \neq \sigma_B^2$, οπότε γίνεται ένας αμφίπλευρος έλεγχος

Η διασπορά S_{AB} υπολογίζεται από τις σχέσεις:

$$s_{AB} = \sqrt{(\frac{1}{m}+\frac{1}{n})\frac{(m-1)s_A^2+(n-1)s_B^2}{m+n-2}}, \tag{1.12}$$

για την περίπτωση που οι χρονοσειρές έχουν την ίδια διασπορά και

$$s_{AB} = \sqrt{\frac{s_A^2}{m}+\frac{s_B^2}{n}}, \tag{1.13}$$

για την περίπτωση που οι χρονοσειρές έχουν διαφορετική διασπορά . Οι βαθμοί ελευθερίας δίνονται από την σχέση:

$$N_{df} = \frac{1}{\frac{\psi^2}{m-1}+\frac{(1-\psi)^2}{n-1}}, \quad \text{όπου} \quad \psi = \frac{\frac{s_A^2}{m}}{\frac{s_A^2}{m}+\frac{s_B^2}{n}} \tag{1.14}$$

Η Δοκιμή πρεπει να πληρεί τις ακόλουθες προϋποθέσεις: $N > 0$, $m > 5$ $n > 5$

1.4.2 Ανίχνευση τυχαιότητας (randomness) χρονοσειρών με την βοήθεια αυτοσυσχετογράμματος.

Η συσχέτιση δύο χρονοσειρών υδρομετεωρολογικών δεδομένων (π.χ. γειτονικών βροχομετρικών σταθμών) ανιχνεύεται με τον συντελεστή συσχέτισης r. Με ανάλογο τρόπο μπορεί να διερευνηθεί και η συσχέτιση που πιθανόν να υπάρχει μεταξύ διαδοχικών παρατηρήσεων της ίδιας χρονοσειράς. Στην περίπτωση αυτή υπολογίζονται οι συντελεστές αυτοσυσχέτισης για διάφορα χρονικά βήματα, οι οποίοι παρέχουν χρήσιμη πληροφορία όσον αφορά την συσχέτιση των δεδομένων της ίδιας χρονοσειράς για διάφορα χρονικά βήματα.

Εστω χρονοσειρά παρατηρημένων τιμών X_1,....X_N, για την οποία σχηματίζουμε τα (N-1) ζεύγη παρατηρήσεων (X_1, X_2), (X_2,X_3).... (X_{N-1}, X_N). Θεωρώντας την πρώτη παρατήρηση κάθε ζεύγους τιμων ως μία μεταβλητή και την δεύτερη παρατήρηση ως άλλη μεταβλητή τότε ο συντελεστής αυτοσυσχέτισης μεταξύ διαδοχικών παρατηρήσεων X_t και X_{t+1} δίνεται από τη σχέση :

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{(N-1)\sum_{t=1}^{N}(x_t - \bar{x})^2 / N}, \quad \text{όπου } \bar{x} = \sum_{t=1}^{N} x_t / N \tag{1.15}$$

Με ανάλογο τρόπο οι συντελεστές αυτοσυσχέτισης r_k, οι οποίοι εκτιμούν την συσχέτιση μεταξύ των παρατηρήσεων X_t και X_{t+k}, που μετρήθηκαν με χρονική διαφορά k δίνονται από τη σχέση :

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{1.16}$$

και καλούνται συντελεστές αυτοσυσχέτισης χρονικού βήματος k.

Ενα χρήσιμο εργαλείο για την ερμηνεία του συνόλου συντελεστών αυτοσυσχέτισης r_k, είναι ένα διάγραμμα, όπου σχεδιάζεται το r_k συναρτήσει του χρονικού βήματος k, το οποίο καλείται αυτοσυσχετόγραμμα. Παρ'ότι η ερμηνεία του αυτοσυσχετογράμματος δεν είναι πάντα εύκολη στην συνέχεια αναλύονται διάφοροι τύποι χρονοσειρών και γενικές διαπιστώσεις όσον αφορά το σχήμα των διαγραμμάτων που δίνουν οι ίδιες οι σειρές καθώς και τα αυτοσυσχετογράμματα αυτών.

<u>(α) Τυχαίες χρονοσειρές (random series).</u>

Μια χρονοσειρά για να είναι τυχαία και να αποτελείται από τελείως ανεξάρτητες παρατηρήσεις μεταξύ τους, αρκεί ο συντελεστής αυτοσυσχέτισης r_k για το χρονικό βήμα k και μεγάλο πληθυσμό παρατηρήσεων N να ικανοποιεί τη σχέση:

$r_k = 0$

Εχει δειχθεί ότι για τυχαία χρονοσειρά, ο συντελεστής αυτοσυσχέτισης r_k ακολουθεί κανονική κατανομή N(0,1/N), έτσι ώστε, εάν μια χρονοσειρά είναι τυχαία 19 από τις 20 τιμές του r_k αναμένονται να κείνται εντός των ορίων $\pm 2/\sqrt{N}$.

Ετσι μια από τις είκοσι τιμές του r_k αναμένεται να είναι σημαντική (significant) ακόμα και όταν η χρονοσειρά είναι τυχαία. Το γεγονός αυτό επιφέρει δυσκολίες στην ερμηνεία του αυτοσυσχετογράμματος διότι πολλές φορές οι συντελεστές αυτοσυσχέτισης παρουσιάζουν ασυνήθεις τιμές, ακόμα και όταν δεν υπάρχει πραγματική αιτία, δηλαδή η χρονοσειρά είναι τυχαία.

(β) Χρονοσειρές με βραχυπρόθεσμη συσχέτιση τιμών (short-term correlation).

Στάσιμες χρονοσειρές συχνά εμφανίζουν βραχυπρόθεσμη συσχέτιση, η οποία χαρακτηρίζεται από σχετικά μεγάλη τιμή του r_1, που ακολουθείται από 2 έως 3 συντελεστές αυτοσυσχέτισης μεγαλύτερους του μηδενός, οι οποίοι διαδοχικά τείνουν να γίνονται μικρότεροι, και οι τιμές του r_k για μεγαλύτερα χρονικά βήματα τείνουν στο 0.

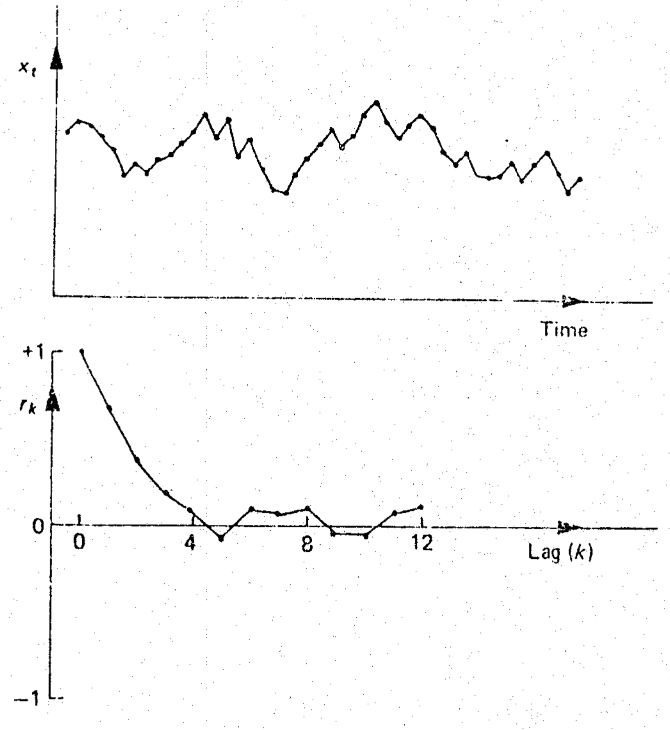
Ενα παράδειγμα αυτού του είδους αυτοσυσχετογράμματος δίνεται στο σχήμα 1.5. Χρονοσειρά, η οποία εμφανίζει τέτοιου είδους αυτοσυσχετόγραμμα, είναι τέτοια ώστε όταν μία παρατήρηση είναι μεγαλύτερη από το μέσο όρο τείνει να ακολουθείται από παρατήρηση επίσης μεγαλύτερη του μέσου όρου και κατ'αναλογία συμβαίνει το ίδιο για παρατήρηση μικρότερη του μέσου όρου.

(γ) Εναλλασσόμενες χρονοσειρές (alternating series)

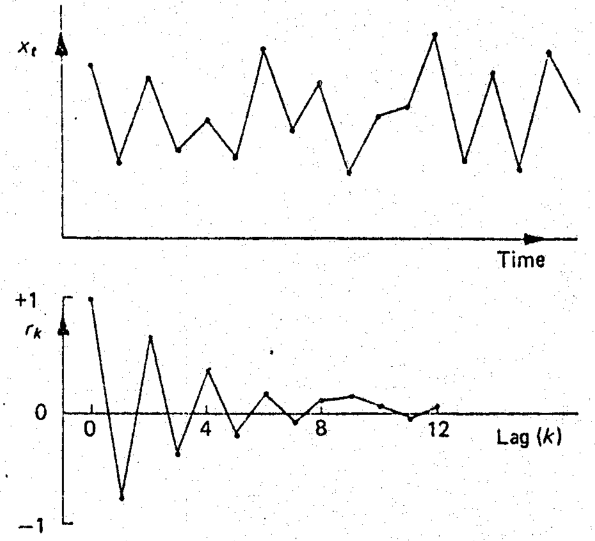
Εάν μια χρονοσειρά έχει την τάση να εναλλάσσεται ως προς τον μέσο όρο, δηλαδή διαδοχικές παρατηρήσεις αυτής εναλλάσσονται από μεγαλύτερη σε μικρότερη του μέσου όρου και αντίστροφα, τότε το πρόσημο των συντελεστών αυτοσυσχέτισης τείνει επίσης να εναλλάσσεται από αρνητικό σε θετικό και αντίστροφα. Η τιμή του r_1 θα είναι αρνητική, ενώ η τιμή του r_2 θα είναι θετική καθώς οι παρατηρήσεις με χρονικό βήμα 2 τείνουν αμφότερες να κείνται από την ίδια πλευρά του μέσου όρου (μεγαλύτερες ή μικρότερες). Μια τυπική εναλλασσόμενη χρονοσειρά μαζί με το αυτοσυσχετόγραμμά της φαίνεται στο σχήμα 1..6

(δ) Μη-στάσιμες χρονοσειρές (non-stationary series).

Οταν μια χρονοσειρά έχει τάση (trend), τότε οι τιμές των συντελεστών αυτοσυσχέτισης r_k δεν τείνουν προς το μηδέν παρά μόνον για πολύ μεγάλες τιμές του χρονικού βήματος k. Αυτό συμβαίνει διότι μετά από μια παρατήρηση μεγαλύτερη του μέσου όρου συνήθως ακολουθεί ένας μεγάλος αριθμός παρατηρήσεων επίσης μεγαλύτερων του μέσου όρου, επειδή ακριβώς η χρονοσειρά παρουσιάζει τάση (trend).



Σχήμα 1.5 Χρονοσειρά βροχομετρικών παρατηρήσεων με βραχυ-πρόθεσμη συσχέτιση τιμών και αυτοσυσχετόγραμμα αυτής (2).



Σχήμα 1.6 Εναλλασσόμενη χρονοσειρά βροχομετρικών παρατηρήσεων και αυτοσυσχετόγραμμα αυτής (2).

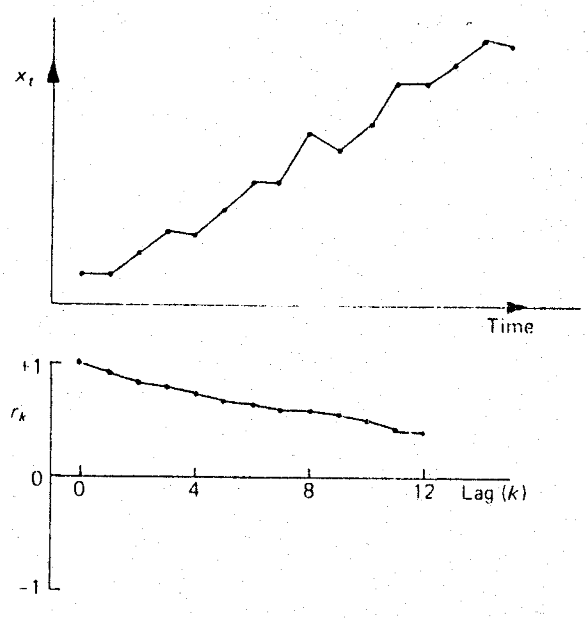
Η χρονοσειρά του σχήματος 1.7 είναι τυπική μη-στάσιμη χρονοσειρά μαζί με το αυτοσυσχετόγραμμα αυτής. Προκειμένου να υπολογισθούν οι συντελεστές αυτοσυσχέτισης r_k για μια τέτοια χρονοσειρά πρέπει πρώτα να αφαιρεθεί η τάση (trend) αυτής.

(ε) Χρονοσειρές με εποχιακή περιοδικότητα (seasonal periodicity).

Οταν μια χρονοσειρά παρουσιάζει εποχιακή περιοδικότητα τότε το αυτοσυσχετόγραμμα αυτής επίσης παρουσιάζει περιοδικότητα της ίδιας συχνότητας. Παράδειγμα για χρονοσειρά μηνιαίων παρατηρήσεων, ο συντελεστής αυτοσυσχέτισης r_6 θα έχει τιμή μεγάλη και αρνητική ενώ ο συντελεστής αυτοσυσχέτισης r_{12} θα έχει τιμή μεγάλη και θετική. Οταν το διάγραμμα μιας χρονοσειράς έχει τη μορφή συνημίτονου τότε και το αυτοσυσχετόγραμμα αυτής ακολουθεί παρόμοια μορφή. Μόνο όταν αφαιρεθεί η εποχιακή περιοδικότητα από μια χρονοσειρά, το αυτοσυσχετόγραμμα μπορεί να δώσει ενδιαφέρουσα πληροφορία.

(στ) Χρονοσειρές με τιμές εκτός ορίων (outliers).

Το αυτοσυσχετόγραμμα χρονοσειράς με μια ή περισσότερες τιμές εκτός των ορίων (ανωτέρου ή κατωτέρου) παρουσιάζει διάφορες διαταραχές.



Σχήμα 1.7 Μη στάσιμη χρονοσειρά βροχομετρικών παρατηρήσεων και αυτοσυσχετόγραμμα αυτής (2).

2 ΟΜΟΓΕΝΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

2.1 Εισαγωγή

Οι χρονοσειρές βροχομετρικών παρατηρήσεων είναι ομογενείς όταν τα γεγονότα προέρχονται από το ίδιο φυσικό φαινόμενο και ανήκουν στον ίδιο πληθυσμό εκφράζοντας τιμές της ίδιας μετρικής ιδιότητας. Οι χρονοσειρές για να είναι στατιστικά επεξεργάσιμες πρέπει να είναι ομογενείς.

Η ανομοιογένεια των χρονοσειρών βροχοπτώσεων, μπορεί να οφείλεται σε μετακίνηση του σταθμού παρατήρησης, σε ανακαίνιση των οργάνων παρατήρησης και σε αλλαγή παρατηρητή με αποτέλεσμα να προκαλείται αλλαγή του υπερετήσιου μέσου βροχοπτώσεων σε μια λεκάνη απορροής.

Για τον εντοπισμό της ανομοιογένειας χρονοσειρών εφαρμόζεται η μέθοδος της διπλής αθροιστικής καμπύλης.

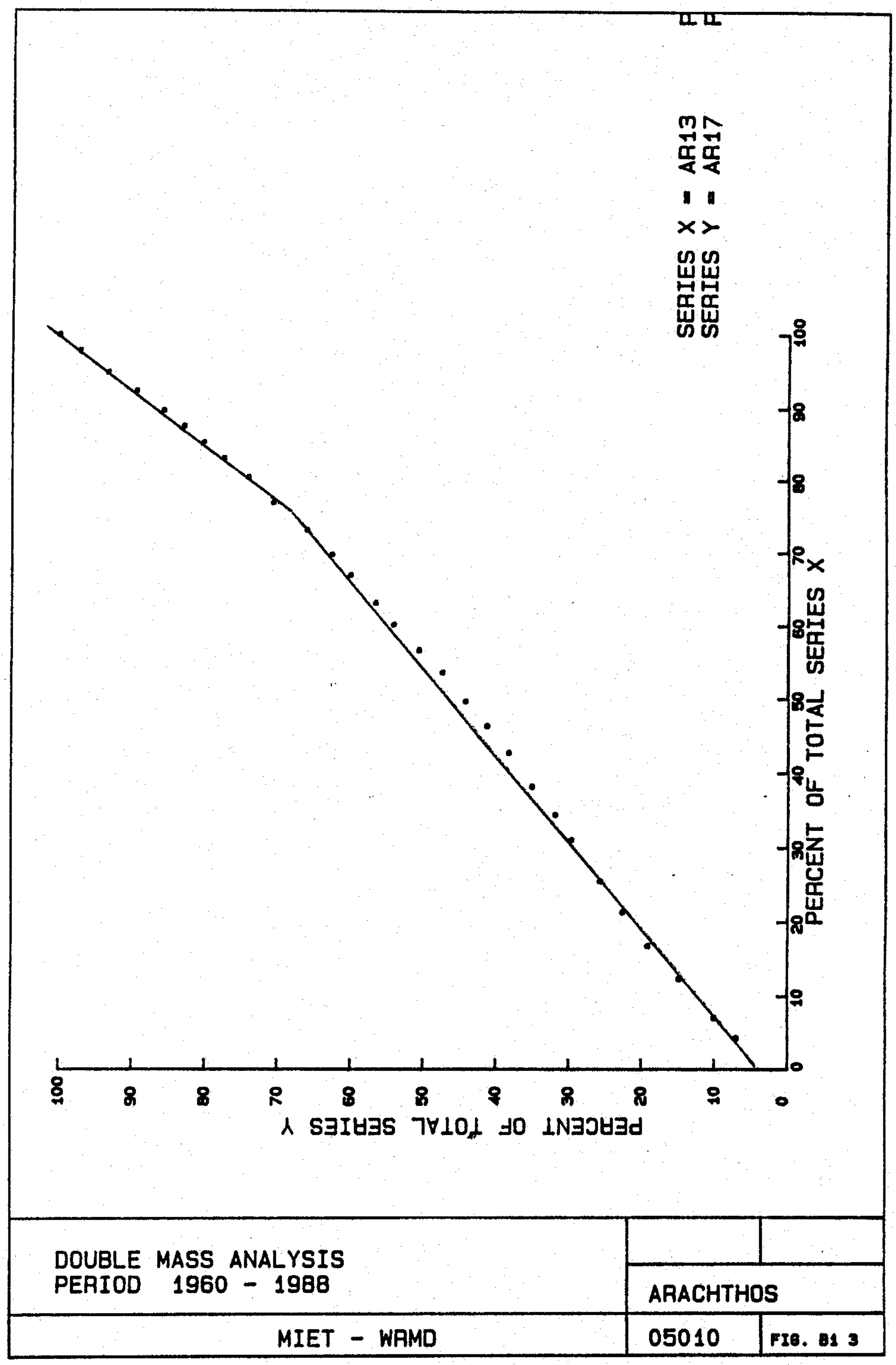
2.2 Μέθοδος Διπλής Αθροιστικής Καμπύλης (Double mass analysis)

Κατά την μέθοδο της διπλής αθροιστικής καμπύλης ελέγχεται η ομοιογένεια του κάθε σταθμού μιας υδρολογικής λεκάνης ως προς τον σταθμό βάσης, του οποίου η ομοιογένεια ειναι εξασφαλισμένη. Σαν σταθμός βάσης της λεκάνης λαμβάνεται ένας υποθετικός σταθμός για τον οποίο είναι διαθέσιμη η χρονοσειρά των μέσων επιφανειακών ετήσιων βροχοπτώσεων της λεκάνης,που θεωρείται ομογενής. Στη συνέχεια περιγράφεται λεπτομερώς η διάδικασία ελέγχου της ομοιογένειας σταθμών με την προαναφερθείσα μέθοδο.

- Ελέγχεται εάν τα δείγματα ετήσιων βροχοπτώσεων των υπό έλεγχο σταθμών και του σταθμού βάσης ακολουθούν την κανονική κατανομή. Ο έλεγχος γίνεται με τη Δοκιμή (test) X^2.

- Υπολογίζεται η χρονοσειρά μέσων ετήσιων βροχοπτώσεων της λεκάνης, σαν συνάρτηση των διαθέσιμων βροχομετρικών παρατηρήσεων όλων των σταθμών της ίδιας λεκάνης και η χρονοσειρά αυτή αποτελεί την καλούμενη βάση.

- Υπολογίζεται ο συντελεστής γραμμικής συσχέτισης r, μεταξύ της υπό έλεγχο χρονοσειράς και της βάσης, για την μεγαλύτερη δυνατή συνεχή περίοδο κοινών παρατηρήσεων. Θεωρείται συνήθως ότι η μέθοδος είναι εφαρμόσιμη εφόσον $r>0{,}70$, το οποίο σημαίνει ότι υπάρχει γραμμική σχέση μεταξύ των δύο χρονοσειρών παρατηρήσεων.



Σχήμα 2.1 Διπλή αθροιστική καμπύλη βροχομετρικών δεδομένων (4).

- Υπολογίζονται τα αθροιστικά ετήσια ύψη βροχοπτώσεων των υπό έλεγχο σταθμών καθώς και του σταθμού βάσης, αρχίζοντας συνήθως από το τελευταίο έτος παρατηρήσεων και προχωρώντας προς τα πίσω. Συντάσσεται διάγραμμα σημείων (σχήμα 2.1) με τετμημένες τα αθροιστικά ύψη βροχοπτώσεων του σταθμού βάσης και τεταγμένες τα αντίστοιχα του υπό έλεγχο σταθμού.

- Από τα σημεία αυτά σχεδιάζεται η βέλτιστη ευθεία (ευθεία ελαχίστων τετραγώνων), που συσχετίζει τα αθροιστικά ύψη μεταξύ των δύο σταθμών. Η ύπαρξη συστηματικού σφάλματος στον υπό έλεγχο βροχομετρικό σταθμό για ορισμένη χρονική περίοδο, εμφανίζεται με θλάση (σπάσιμο) της ευθείας που συσχετίζει τους δύο σταθμούς μεταξύ τους και η θλάση εμφανίζεται στο έτος που αρχίζουν τα συστηματικά σφάλματα του υπό έλεγχο σταθμού.

Στην περίπτωση που η υπό έλεγχο χρονοσειρά είναι ομογενής τότε θα πρέπει τα σημεία που ορίζονται από τα αθροιστικά ύψη βροχοπτώσεων των δύο σταθμών να ακολουθούν γραμμική σχέση.

- Αν κριθεί ότι η προσαρμογή της ευθείας στα σημεία είναι ικανοποιητική και δεν παρατηρηθεί σημαντική θλάση των σημείων της διπλής αθροιστικής, τότε η υπό έλεγχο χρονοσειρά θεωρείται ομογενής και ο έλεγχος σταματάει.

- Διαφορετικά υπολογίζεται η εξίσωση της τεθλασμένης ευθείας παλινδρόμησης, η οποία έχει k τμήματα και k-1 θλάσεις και γίνεται προσπάθεια ώστε ο αριθμός και οι θέσεις των θλάσεων που θα αποφασισθούν να εκφράζουν τις μεταβολές του φυσικού φαινομένου.

- Τέλος πραγματοποιείται η ανόρθωση των σημείων της διπλής αθροιστικής καμπύλης. Πιο αξιόπιστα θεωρούνται τα πιο πρόσφατα στοιχεία και γι αυτό ανορθώνονται τα στοιχεία του κάθε ευθύγραμμου τμήματος με βάση την κλίση του προηγούμενού του.

3 ΣΥΜΠΛΗΡΩΣΗ ΕΛΛΕΙΠΟΥΣΩΝ ΤΙΜΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

3.1 Εισαγωγή

Συμπλήρωση μιας χρονοσειράς βροχομετρικών δεδομένων, καλείται η αύξηση του εύρους αυτής: (α) με την προσθήκη είτε ενδιάμεσων δεδομένων που λείπουν είτε ακραίων δεδομένων στο τέλος ή στην αρχή της χρονοσειράς και (β) με την αντικατάσταση δεδομένων που η τιμή τους κρίνεται αναξιόπιστη.

Η συμπλήρωση των χρονοσειρών γίνεται συνήθως ύστερα από συσχέτιση με χρονοσειρές γειτονικών σταθμών των οποίων οι παρατηρήσεις θεωρούνται αξιόπιστες. Η συσχέτιση αυτή είναι επιτυχής διότι συνήθως υπάρχει ισχυρή συσχέτιση μεταξύ των δειγμάτων γειτονικών βροχομετρικών σταθμών, σε μηνιαία ή ετήσια βάση.

Συμπλήρωση είναι επίσης δυνατόν να γίνει με βάση τα στοχαστικά χαρακτηριστικά του ίδιου σταθμού, δηλαδή **συνθετική χρονοσειρά** για τα κομμάτια της χρονοσειράς που λείπουν, π.χ. με κάποιο μοντέλο Markov.

Στις επόμενες παραγράφους αναλύονται οι διάφοροι μέθοδοι συμπλήρωσης ελλειπουσών τιμών.

3.2 Μέθοδοι συμπλήρωσης ελλειπουσών τιμών

3.2.1 Μέθοδος της μέσης τιμής.

Η πιο απλή μέθοδος συμπλήρωσης ελλείπουσας τιμής του δείγματος περιλαμβάνει την συμπλήρωση αυτής με τη μέση τιμή του δείγματος, αρκει το δείγμα να είναι μεγάλο και να έχει σταθερή μέση τιμή και τυπική απόκλιση.

Στην περίπτωση ελλείπουσας μηνιαίας τιμής (π.χ. Οκτώβριος κάποιου έτους) εκτός από τη μέση μηνιαία τιμή είναι δυνατόν να συμπληρωθεί και από την τιμή του ίδιου μήνα κάποιου άλλου έτους, που συμπεριφέρεται το ίδιο με τον χρόνο που λείπει η μηνιαία τιμή π.χ. τον Οκτώβριο

3.2.2 Μέθοδος της γραμμικής παρεμβολής.

Σε αρκετές περιπτώσεις τα κενά δεδομένων χρονοσειρών είναι δυνατόν να συμπληρωθούν με γραμμική παρεμβολή μεταξύ της τελευταίας τιμής πριν το κενό και της αντίστοιχης πρώτης τιμής μετά το κενό, δεδομένου ότι το διάστημα για το οποίο γινεται η γραμμική παρεμβολή δεν είναι πολύ μεγάλο.

3.2.3 Μέθοδος των αντιστρόφων αποστάσεων.

Εστω Y_t η ελλείπουσα τιμή μιας χρονοσειράς για τον χρόνο t. Τότε η Y_t υπολογίζεται από τον τύπο :

$$Y_t = \sum_{j=1}^{n} A_j X_{j,t} \tag{3.1}$$

όπου A_j οι συντελεστές βάρους οι οποίοι υπολογίζονται από τη σχέση :

$$A_j = (1/D_i^2) \Big/ \sum_{i=1}^{n} 1/D_i^2 \tag{3.2}$$

όπου D_i είναι η απόσταση του σταθμού i από τον σταθμό, ο οποίος πρόκειται να συμπληρωθεί και n είναι ο αριθμός των γειτονικών σταθμών.

3.2.4 Μέθοδος των υπερετήσιων λόγων.

Εστω Y_t η ελλείπουσα τιμή μιας χρονοσειράς για το χρόνο t. Τότε η Y_t υπολογίζεται από τη σχέση :

$$Y_t = (1/n) \sum_{j=1}^{n} (Y/X_j).X_{jt} \tag{3.3}$$

όπου Y και X_j οι μέσες υπερετήσιες τιμές των μηνιαίων ή ετήσιων βροχοπτώσεων του προς συμπλήρωση σταθμού και των j γειτονικών σταθμών αντίστοιχα.

3.2.5 Καμπύλη συσχέτισης χρονοσειρών - Μέθοδος πολλαπλής γραμμικής παλινδρόμησης.

Καμπύλες συσχέτισης χρονοσειρών είναι δυνατόν να χρησιμοποιηθούν για την συμπλήρωση ελλειπουσών τιμών, με την προυπόθεση ότι το σφάλμα προσαρμογής της καμπύλης στα δεδομένα είναι μικρό.

Η καμπύλη συσχέτισης δύο χρονοσειρών μπορεί να είναι είτε ευθεία γραμμή είτε κάθε καμπύλη που περιγράφεται από πολυωνυμική σχέση οποιασδήποτε τάξης. Στην συνέχεια αναλύονται οι εξισώσεις συσχέτισης χρονοσειρών, οι οποίες είναι δυνατόν να χρησιμοποιηθούν στην παρούσα μέθοδο.

Οι ακόλουθοι τύποι εξισώσεων - συναρτήσεων συσχέτισης χρονοσειρών είναι διαθέσιμοι σύμφωνα με Delft Hydraulics (1989), όπου Y_i είναι η εξαρτημένη και X_i οι ανεξάρτητες μεταβλητές αντίστοιχα :

- **πολυωνυμική**

$$Y_i = \sum_{j=0}^{n} C_j X_i^n \qquad (3.4)$$

όπου n ο βαθμός του πολυώνυμου για $n \leq 3$ και C_j συντελεστές,

- **απλή γραμμική**

$Y_i = A + B.X_i$, όπου A, B συντελεστές (3.5)

- **εκθετική**

$Y_i = A\ exp(B.X_i)$, όπου A, B συντελεστές (3.6)

- **εκθετική**

$Y_i = A\ exp(B/X_i)$, όπου A, B συντελεστές (3.7)

- **δύναμη**

$Y_i = A\ X_i^B$, όπου A, B συντελεστές (3.8)

- **λογαριθμική**

$Y_i = A + B\ ln(X_i)$, όπου A, B συντελεστές (3.9)

- **υπερβολική**

$Y_i = A + B/X_i$, όπου A, B συντελεστές (3.10)

- **πολλαπλή γραμμική**

$$Y_i = C_o + \sum_{j=1}^{n} C_j X_{j,i} \qquad (3.11)$$

όπου n, ο αριθμός των χρονοσειρών των ανεξάρτητων μεταβλητών για $n \leq 10$ και C_j συντελεστές.

Ολοι οι προαναφερθέντες συντελεστές των εξισώσεων υπολογίζονται με τη μέθοδο των ελαχίστων τετραγώνων.

Η πολλαπλή γραμμική συσχέτιση χρησιμοποιείται στη μέθοδο πολλαπλής γραμμικής παλινδρόμησης όπου η ελλείπουσα τιμή Y_t για τον χρόνο t, υπολογίζεται από την ακόλουθη σχέση :

$$Y_t = \left|C_1 C_2 \ldots C_{n+1}\right| \begin{vmatrix} 1 \\ Y_1 \\ Y_2 \\ . \\ . \\ Y_{n+1} \end{vmatrix} \tag{3.12}$$

όπου :

$$|C| = (|X|^T |X|^{-1})(|X|^T |Y|) \tag{3.13}$$

|X| είναι το ορθογωνικό μητρώο με n+1 στήλες και m γραμμές,

n είναι ο αριθμός των γειτονικών σταθμών και m είναι ο αριθμός των ετών για τα οποία υπάρχουν αξιόπιστες παρατηρήσεις για τον προς συμπλήρωση σταθμό καθώς και τους γειτονικούς σταθμούς.

Το μητρώο |X| έχει την ακόλουθη μορφή:

$$|X| = \begin{vmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1n+1} \\ 1 & X_{21} & X_{22} & \ldots & X_{2n+1} \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ 1 & X_{m1} & X_{m2} & \ldots & X_{mn+1} \end{vmatrix} \tag{3.14}$$

όπου X_{ij} η τιμή της χρονοσειράς του γειτονικού σταθμού i για το έτος j.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1..0 Delft Hydraulics, Δ/νση Σχεδιασμού και Ανάπτυξns Υδατικών Πόρων, HYMOS-Διαχειριστικό Σύστημα Βάσης Υδρολογικών Δεδομένων (1989)

2.0 Chatfield, C., (1975) The Analysis of Time Series: An Introduction, The University of Bath, Champan and Hall, London.

3.0 Kottegoda, N.T (1980), Stochastic Water Resources Technology, Mc Millan, London.

4.0 Ministry of Industry Energy and Technology, Water and Natural Resources Directorate (1991),"Pilot study for water resources managment of the Louros and Arachthos river basins, Athens.

5.0 Νικολάου Κ., Μαμάσης Ν., Τσακαλίας Γ., Ανυφαντή Χ. (1992), Εκτίμηση και διαχείριση των υδατικών πόρων της Στερεάς Ελλάδας, Τεύχος 5,Βάση και Προγράμματα επεξεργασίας μηνιαίων υδρολογικών δεδομένων, Ε.Μ.Π.

6.0 Ξανθόπουλος Θ. (1972), Εισαγωγή στην Τεχνική Υδρολογία, Εκδοση Ε.Μ.Π., Αθήνα.

7.0 Salas, I.D., Delleur J.W., Yevjevich V. and Lane W.L., (1981), Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado, U.S.A.

8.0 Yevjevich V, (1972), Stochastic Processes in Hydrology, W.R. Publication, Fort Colins, Colorado, U.S.A.