ΥΔΡΟΣΚΟΠΙΟ

ΠΡΟΓΡΑΜΜΑ STRIDE ΕΛΛΑΣ

ΔΗΜΙΟΥΡΓΙΑ ΕΘΝΙΚΗΣ ΤΡΑΠΕΖΑΣ
ΥΔΡΟΛΟΓΙΚΗΣ ΚΑΙ ΜΕΤΕΩΡΟΛΟΓΙΚΗΣ
ΠΛΗΡΟΦΟΡΙΑΣ

HYDROSCOPE

STRIDE HELLAS PROGRAMME

DEVELOPMENT OF A NATIONAL DATA
BANK FOR HYDROLOGICAL AND
METEOROLOGICAL INFORMATION

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΦΥΣΙΚΟ ΤΜΗΜΑ - ΤΟΜΕΑΣ ΜΕΤΕΩΡΟΛΟΓΙΑΣ
και
ΕΘΝΙΚΟ ΑΣΤΕΡΟΣΚΟΠΕΙΟ ΑΘΗΝΩΝ

UNIVERSITY OF ATHENS
DEPARTMENT OF PHYISICS
LABORATORY OF METEOROLOGY
and
NATIONAL OBSERVATORY OF ATHENS

ΣΥΜΠΛΗΡΩΣΗ ΕΛΛΕΙΠΟΥΣΩΝ ΤΙΜΩΝ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕΤΕΩΡΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

FILLING IN MISSING DATA IN METEOROLOGICAL TIME SERIES

Α. Δίπλα

A. Dipla

Επιστ. Υπεύθυνοι: *Γ. Κάλλος (ΕΚΠΑ), Μ. Πετράκης (ΕΑΑ)*

Tech. Supervisors: *G. Kallos (UA), M. Petrakis (NOA)*

Αριθμός τεύχους
Report number **3/2.[redacted]**

ΑΘΗΝΑ - ΑΠΡΙΛΙΟΣ 1993
ATHENS - APRIL 1993

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΙΝΑΚΕΣ

ΣΧΗΜΑΤΑ

ΠΕΡΙΛΗΨΗ

Σκοπός της εργασίας αυτής στα πλαίσια του προγράμματος Υδροσκόπιο, είναι η περιγραφή των στατιστικών μεθόδων για την συμπλήρωση των ελλειπουσών ή αναξιόπιστων τιμών, στα καταγραφέντα μέχρι σήμερα μετεωρολογικά δεδομένα, όπως θερμοκρασία του αέρα, σχετική υγρασία, εξάτμιση, διεύθυνση και ένταση του ανέμου, ηλιοφάνεια, ηλιακή ακτινοβολία,πίεση και νέφωση.

Στο πρώτο κεφάλαιο, γίνεται μια σύντομη ανασκόπηση του είδους των μετεωρολογικών δεδομένων υπό επεξεργασία. Επίσης αναπτύσονται εν περιλήψει μερικές χρήσιμες έννοιες της στατιστικής ανάλυσης που χρησιμοποιούνται στη συνέχεια.

Στο δεύτερο κεφάλαιο, περιγράφονται μερικοί ενδεικτικοί έλεγχοι για την εξακρίβωση βασικών ιδιοτήτων που πρέπει να πληρούν οι εγγραφές, όπως η ομοιογένεια, η τυχαιότητα, η στασιμότητα και η ύπαρξη τάσης.

Στο τρίτο κεφάλαιο περιγράφονται οι προτεινόμενες μέθοδοι, ξεκινώντας από τις απλές περιπτώσεις γραμμικής παρεμβολής έως την στατιστική πρόβλεψη.

ABSTRACT

This is an attempt to describe how to use some very basic statistical methods in order to fill in missing records in time series of meteorological data as air temperature, relative humidity, evaporation, wind data, sunshine duration, global and direct sun radiation, pressure and cloud amount.

This report is set out in three chapters:

The first chapter is a small introduction to the problem, the short description of the data under consideration and the basic ideas of statistical analysis to be used.

The second chapter, describes some basic tests in order to assure some chracteristics of the series as homogeneity, randomness, stationarity and trend.

The third chapter, is concerned with describing in detail the methods for estimating the missing values.

1 ΕΙΣΑΓΩΓΗ

Η συλλογή και καταγραφή των τιμών μετεωρολογικών δεδομένων, η συστηματική τους ανάλυση και η ακριβής επεξεργασία τους, καλύπτουν τις παρουσιαζόμενες απαιτήσεις σε όλους τους έμμεσα και άμμεσα σχετιζόμενους επιστημονικούς κλάδους.

Σημαντική βαρύτητα παρουσιάζει η αξιοπιστία των τιμών των μετεωρολογικών δεδομένων, που συνδέεται άμεσα με το είδος και τον τρόπο συλλογής τους. Οι αυστηροί έλεγχοι κατά την διάρκεια των παρατηρήσεων και των καταγραφών καθώς και η επιβεβαίωση των τιμών με την εφαρμογή απλών μεθόδων, (quality control), αποτελούν το πρώτο βήμα στην αλυσίδα της όλης διαδικασίας.

Ενα από τα συχνά παρουσιαζόμενα προβλήματα των κλασσικών μεθόδων παρατήρησης, (μη αυτοματοποίησης στην συλλογή των τιμών σε απομεμακρυσμένους και δύσκολα προσπελάσιμους σταθμούς), και των αντικειμενικών δυσκολιών που εξυπακούονται, είναι η έλλειψη τιμών στις σειρές των μετεωρολογικών παρατηρήσεων.

Η στατιστική ανάλυση, παρέχει ένα σημαντικό εργαλείο στην επίλυση του ανωτέρω προβλήματος, που ξεκινάει από το στάδιο της ομαδοποίησης των δεδομένων μέχρι την συμπλήρωση των ελλειπουσών τιμών.

1.1 Περιγραφή των υπό θεώρηση μετεωρολογικών δεδομένων.

Αντικείμενο μελέτης αποτελούν οι ακόλουθες ιδιότητες και συνθήκες της ατμόσφαιρας (βλ. Πίνακα 1), σε κάποιο συγκεκριμένο τόπο και κατά την διάρκεια μιας συγκεκριμένης χρονικής στιγμής ή περιόδου.

ΠΙΝΑΚΑΣ 1

Υδρομετεωρολογικές Εφαρμογές[1]

Μεταβλητή	Οργανο	Είδος Εγγραφής	Χρονικό Βήμα Παρατηρ.	Χρονικό Βήμα Αρχειοθ.	Μονάδες
Θερμοκρ. αέρα	Θερμογρά-φος	Συνεχής	Ημερήσιες ταινίες	Ωρα	0c
	Θερμόμετρο	Περιοδική	3 Μετρήσεις / ημέρα	Ημέρα (3 μετρ.)	0c
	Θερμ. Μεγ. Ελαχ.	Περιοδική	1 ζευγ. Μετρ./ημερ.	Ημέρα	0c
Σχετική Υγρασία	Υγρογρά-φος	Συνεχής	Εβδομ. ή ημερήσιες ταινίες	Ωρα	%
	Υγρόμε-τρο	Περιοδική	3-ωρο	3-ωρο	%
Εξάτμιση	Λεκάνη	Περιοδική	Ημέρα	Ημέρα	mm
	PICHE	Περιοδική	Ημέρα	Ημέρα	mm
	Εξατμισιο-γράφος		Εβδομαδι-αίες ταινίες	Ωρα	mm
Ανεμος (διεύθυν-ση, ενταση)	Ανεμογράφος	Συνεχής	Μηνιαίες ταινίες	Ωρα	m/sec, μοίρες
	Ανεμόμετρο	Περιοδική	Ημέρα	Ημέρα	m/sec, μοίρες
Ηλιοφά-νεια	Ηλιογράφος	Συνεχής	Ημερήσιες ταινίες	Ωρα	Ναι/Οχι
Ηλιακή ακτινοβο-λία	Ακτινογράφος	Περιοδική	Ημέρα	Ημέρα	cal/cm^2
Πίεση (Μ.Σ.Θ)	Βαρογράφος	Συνεχής	4-ήμερες ταινίες	Ωρα	mb
	Βαρόμετρο	Περιοδική	3-ωρο	3-ωρο	mb
Νέφωση	Παρατήρη-ση	Περιοδική	42 μετρήσεις/ ημέρα	Ημέρα (42 μετρήσεις)	Ογδοα

[1]Αποτελει τμημα του πινακα που εχει δημοσιευτει στην περιγραφη του προγραμματος Stride με τιτλο: Δημιουργια Εθνικης Τραπεζας Υδρολογικης και Μετεωρολογικης Πληροφοριας.

Τα παραπάνω αναφερόμενα δεδομένα εμφανίζονται στην μορφή χρονοσειρών,(time series), στις οποίες οι καταγραφόμενες τιμές παρουσιάζουν κάποια μορφή εξάρτησης (όχι χρονικής), που είναι και μέρος της αναζήτησης για την επίλυση τού προβλήματος.

Οπως φαίνεται από τον πίνακα, οι καταγραφόμενες τιμές είναι άλλοτε συνεχείς και άλλοτε διακριτές σε ίσα χρονικά διαστήματα. Οι διακριτές μεταβλητές σχετίζονται με την συχνότητα, (αριθμός ωρών, ημερών κ.λ.π), με την οποία το υπό συζήτηση φαινόμενο, (χιόνι, χαλάζι, καταιγίδα κ.λ.π), εμφανίζεται στην διάρκεια κάποιας συγκεκριμένης περιόδου, (μήνας, εποχή, έτος κ.λ π). Οι συνεχείς μεταβλητές, συνήθως καταγράφονται με τέτοιο τρόπο, ώστε μπορούν να πάρουν οποιαδήποτε τιμή σε κάποια συνεχή κλίμακα. Η μέγιστη θερμοκρασία, η ατμοσφαιρική πίεση είναι παραδείγματα τέτοιων μεταβλητών.

Το πρόβλημα εστιάζεται στο παρακάτω σημείο: Υπάρχει περίπτωση, να χρειάζεται να συμπληρωθεί μια τιμή, έχοντας ως δεδομένες τιμές της χρονοσειράς πριν και μετά την συγκεκριμένη τιμή που πρέπει να συμπληρωθεί ή να διορθωθεί. Πολύ συχνά, είναι επιθυμητό να γίνει κάποια πρόβλεψη. Δηλαδή, έχοντας διαθέσιμες τις τιμές μιας χρονοσειράς Χ μέχρι την χρονική στιγμή t_k, ποιά θα είναι η τιμή της χρονοσειράς σε κάποια επομένη στιγμή t_{k+1} ή τήν t_{k+l} όπου $l=1,2....$ Στήν πράξη, υλοποίηση του παραπάνω προβλήματος σημαίνει υπολογισμός της πιθανότητας, να βρίσκεται ή ζητουμένη τιμή μεταξύ ορισμένων ορίων, που εκφράζουν την ακρίβεια της πρόβλεψης.

Οπως είναι γνωστό, στη φύση υπάρχουν δυό μεγάλες κατηγορίες φαινομένων. Τά Ντετερμινιστικά που περιγράφονται απο μαθηματικά μοντέλα και τα οποία επιτρέπουν τον ακριβή υπολογισμό των μεγεθών που υπακούουν τους νόμους του φαινομένου και τα Στοχαστικά ή Στοχαστικές Διαδικασίες, για τα οποία κάποιος ακριβής υπολογισμός της μελλοντικής συμπεριφοράς του φυσικού φαινομένου είναι αδύνατος. Μια στοχαστική διαδικασία θεωρείται ένα στατιστικό φαινόμενο το οποίο εξελίσσεται στο χρόνο σύμφωνα με τους νόμους της πιθανοθεωρίας.

Τά μετεωρολογικά φαινόμενα, τα οποία περιγράφονται μέσω των χρονοσειρών που σχηματίζονται απο τις τιμές των αντιστοίχων μετρήσεων, ανήκουν στην δεύτερη κατηγορία. Ισχύει η βασική υπόθεση ότι, μια συγκεκριμένη χρονοσειρά θεωρείται σαν ένα από τα άπειρα δείγματα(samples), τα οποία θα μπορούσαν να έχουν σχηματισθεί από μια συγκεκριμένη διαδικασία.

1.2 Βασικές αρχές Στατιστικής Ανάλυσης χρονοσειρών στην επεξεργασία Μετεωρολογικών Δεδομένων.

Η απαραίτητη ανάλυση για την εύρεση του στοχαστικού μοντέλου, που θεωρείται κατάλληλο για την περιγραφή της στοχαστικής διαδικασίας, δείγμα (sample) της οποίας θεωρείται η προς επεξεργασία συγκεκριμένη χρονοσειρά, περιλαμβάνει τα ακόλουθα στάδια:

(α) Την <u>διαδικασία αναγνώρισης</u>, κατά την οποία εξετάζεται η υποψηφιότητα διαφόρων μοντέλων, ως κατάλληλα για την συγκεκριμένη χρονοσειρά.

(β) Την προσαρμογή του επιλεγμένου μοντέλου στην συγκεκριμένη χρονοσειρά.

(γ) Τους διαγνωστικούς ελέγχους κατά την διάρκεια των οποίων ελέγχεται η καταλληλότητα του μοντέλου, γίνονται απαραίτητες μετατροπές, υπολογίζεται το λάθος στην επιλογή και αν χρειαστεί επαναλαμβάνεται η διαδικασία από το σημείο (α).

Ακολουθεί μιά συνοπτική περιγραφή των ανωτέρω βημάτων. Από την αλληλεπίδραση θεωρίας και πράξης στην επίλυση συγκεκριμένων προβλημάτων, μια ομάδα μοντέλων, που πιθανα ταιριάζει στα προς επεξεργασία δεδομένα, είναι ήδη γνωστή στην αρχή του προβλήματος και αυτό αποτελεί την αφετηρία της όλης έρευνας. Λόγω του ότι αυτή η αρχικά γνωστή κατηγορία πιθανών μοντέλων είναι εκτεταμένη, είναι απαραίτητο να βρεθούν υποομάδες της κατηγορίας αυτής που να ταιριάζουν περισσότερο στα διαθέσιμα δεδομένα. Σύμφωνα με την θεωρία, η παραπάνω διαδικασία περιλαμβάνει την απόρριψη η την αποδοχή,(η αποδοχή της μιας αναιρεί την ισχύ της άλλης και το αντίθετο), δύο βασικών υποθέσεων, της μηδενικής (null) και της εναλλακτικής (alternative). Η αποδοχή της πρώτης (μηδενικής), επιβεβαιώνει το ότι οι τιμές των δεδομένων πληρούν κάποιες προυποθέσεις επακριβώς καθορισμένες, μέσω των οποίων προσδιορίζεται το υποσύνολο η οι ιδιότητες των μοντέλων που ταιριάζουν στα δεδομένα. Η αποδοχή της δεύτερης (εναλλακτικής), επιβεβαιώνει το ακριβώς αντίθετο. Δηλαδή το ότι μερικά ή ολοκληρωτικά οι συνθήκες της πρώτης υπόθεσης δεν ισχύουν.

Η υλοποίηση των παραπάνω, πραγματοποιείται εξετάζοντας αν η τιμή της αβεβαιότητας του εκτιμητή, (estimator or statistic), στον οποίο υπακούουν οι τιμές βρίσκεται σε κάποιο διάστημα εμπιστοσύνης, (confidence interval), οι τιμές του οποίου συνήθως ορίζονται από την αρχή των υπολογισμών. Είναι σαφές ότι το διάστημα εμπιστοσύνης πρέπει να είναι το μικρότερο δυνατό έτσι ώστε η ακρίβεια των υπολογισμών να είναι η μεγαλύτερη δυνατή.

Δύο είδη λαθών είναι συνδεδεμένα με την ανωτέρω θεώρηση. Τό πρώτο έγκειται στην απόρριψη της πρώτης υπόθεσης, ενώ είναι αληθής (λάθος πρώτου είδους). Με άλλα λόγια η αβεβαιότητα του εκτιμητή, παρά το ότι η εκτίμηση είναι αληθής, βρίσκεται εκτός των τιμών του διαστήματος αποδοχής. Η πιθανότητα νά συμβεί το γεγονός αυτό, δεν πρέπει να υπερβαίνει κάποια όρια τα οποία έχουν τεθεί εκ των προτέρων. Τό δεύτερο, (λάθος δευτέρου είδους), έγκειται στην αποδοχή της πρώτης υπόθεσης, ενώ είναι λάθος. Αυτό σημαίνει ότι η δεύτερη (εναλλακτική) υπόθεση είναι αληθής. Καλώντας β την πιθανότητα να συμβεί ένα λάθος δευτέρου είδους, είναι λογικό ότι η πιθανότητα αποδοχής της εναλλακτικής υπόθεσης ενώ είναι αλήθεια είναι (1-β) και ονομάζεται ισχύς της μεθόδου αναγνώρισης σε σχέση με την εναλλακτική υπόθεση. Ο έλεγχος αναγνώρισης που θα δώσει καλύτερα αποτελέσματα είναι αυτός για τον οποίο ισχύει ότι η ισχύς είναι μέγιστη.

Μία εξ ίσου σημαντική συνθήκη η οποία πρέπει να ικανοποιείται, ώστε να εξασφαλίζεται η επιτυχία της πρώτης φάσης της διαδικασίας αναγνώρισης, είναι η ακόλουθη: Η πιθανότητα να απορριφθεί η πρώτη υπόθεση ενώ είναι αληθής, να μην υπερβαίνει την πιθανότητα αποδοχής της ενώ είναι λάθος. Αν a_0 είναι η πιθανότητα απόρριψης της μηδενικής υπόθεσης ενώ είναι αληθής, τότε $a_0 \leq 1-\beta$. Η a_0 ονομάζεται διάστημα εμπιστοσύνης και τις περισσότερες φορές ισούται μέ

0.05. Μιά επίσης χρήσιμη πιθανότητα είναι αυτή η οποία προσφέρει ένα αντικειμενικό μέτρο του βαθμού ασυμβατότητας της πραγματικής κατάστασης καί της μηδενικής υπόθεσης, (null). Συμβολίζεται με $α_1$. Προκύπτει από την κατανομή πιθανότητας στην οποία υπακούει ο εκτιμητής η μορφή του οποίου υπαγορεύται από την μηδενική υπόθεση, (null).

Είναι κατανοητό από ότι έχει αναφερθεί μέχρι στιγμής, ότι το είδος των ελέγχων της διαδικασίας αναγνώρισης, εξαρτάται από την μορφή του εκτιμητή, ο οποίος έχει άμμεση σχέση με το είδος της εσωτερικής οργάνωσης των δεδομένων. Υπάρχει λοιπόν κάποια κατηγοριοποίηση στούς ελέγχους της διαδικασίας αναγνώρισης, ανάλογα με την μορφή του εκτιμητή.

Αν η εσωτερική οργάνωση είναι ξεκάθαρη, η εκλογή του εκτιμητή έγκειται απλά στην ρύθμιση των παραμέτρων μιας γνωστής κατανομής πιθανότητας, η οποία περιγράφει επιτυχώς την δομή των δεδομένων.

Πρόκειται γιά τήν περίπτωση των **παραμετρικών** ελέγχων. Ισχύει ένας εκ νέου διαχωρισμός σέ δύο υποκατηγορίες. Οταν όλες οι παράμετροι είναι επακριβώς γνωστές εξ αρχής, (πρόκειται για τις <u>απλές περιπτώσεις</u>), κατά τις οποίες, υπάρχει μια πλήρως ορισμένη κατανομή πιθανότητας η οποία ταιριάζει στήν σειρά των δεδομένων. Υπάρχουν όμως περιπτώσεις, (πρόκειται για τις <u>σύνθετες</u>) για τις οποίες ισχύει οτι μια παράμετρος παραμένει μη καθορισμένη κατά την διάρκεια των ελέγχων αναγνώρισης.

Σε μια δεύτερη κατηγορία ανήκουν οι έλεγχοι που γίνονται σε δεδομένα χωρίς καμμιά προφανή εσωτερική οργάνωση και ονομάζονται **μή-παραμετρικοί** έλεγχοι. Υποτίθεται οτι τα δεδομένα αυτά είναι ανεξάρτητα (για την περίπτωση των μετεωρολογικών δεδομένων υποτίθεται χρονική ανεξαρτησία) και οτι υπακούουν στον ίδιο νόμο κατανομής πιθανότητας ο οποίος παραμένει άγνωστος.

Η μελέτη των μετεωρολογικών δεδομένων περιλαμβάνει κατανομές για τον προσδιορισμό των οποίων απαιτούνται έλεγχοι που ανήκουν και στις δυο κατηγορίες που ήδη αναφέρθηκαν. Ενα συγκεκριμένο παράδειγμα ενός παραμετρικού ελέγχου αποτελεί στην περίπτωση της κανονικής κατανομής, ο έλεγχος της ισότητας του μέσου όρου και της διασποράς δύο διαφορετικών σειρών δεδομένων. Σε κάποια άλλη περίπτωση, περιλαμβάνει την επαλήθευση του ότι μια κατανομή δεδομένων είναι ενός συγκεκριμένου τύπου. Ενα παράδειγμα ενός μή παραμετρικού ελέγχου, περιλαμβάνει την επαλήθευση του τυχαίου χαρακτήρα των τιμών των δεδομένων.

Η πλήρης στατιστική ανάλυση των μετεωρολογικών φαινομένων, περιλαμβάνει την επέκταση των ανωτέρω αναφερθέντων στο χώρο των πολλαπλών ματαβλητών. Εξ αιτίας της μη στατιστικής ανεξαρτησίας, των εμπλεκομένων μετεωρολογικών στοιχείων (θερμοκρασία, ένταση ανέμου κ.λ.π), οπότε η κοινή κατανομή πιθανότητας (joint distribution) δεν ισούται με το γινόμενο των κατανομών του καθε μεγέθους, εμφανίζεται η αναγκαιότητα εκτίμησης πολυδιάστατων στατιστικών μοντέλλων για την περιγραφή των κατανομών πιθανοτήτων.

Στην εν λόγω διαδικασία εμπλέκεται και η εμπειρική γνώση ορισμένων χαρακτηριστικών των μεταβλητών. Για παράδειγμα υπάρχει κάποια συσχέτιση μεταξύ χρονοσειρών συγχρόνως καταγραφομένων απο γειτονικούς σταθμούς παρατήρησης. Η εξάρτηση έχει δύο σκέλη. Το πρώτο αναφέρεται σε σχετιζόμενα δεδομένα στον ίδιο σταθμό αλλά για διαφορετικά μεγέθη. Η σχέση εξάρτησης είναι σχέση βασισμένη στη στατιστική συμπεριφορά των χρονοσειρών. Το δεύτερο αναφέρεται στην εξάρτηση του ιδίου μεγέθους μετρούμενο σε διαφορετικούς σταθμούς παρατήρησης. Στην δεύτερη περίπτωση, η σχέση εξάρτησης βασίζεται στα κοινά χαρακτηριστικά της φυσικής διαδικασίας που παρήγαγε την χρονοσειρά.

1.3 Γενικές παρατηρήσεις.

1.3.1 Επεξεργασία διανυσματικών μεγεθών. Ιδιαίτερη προσοχή χρειάζεται η ανάλυση διανυσματικών μεγεθών, αντιπροσωπευτικότερο παράδειγμα των οποίων αποτελεί ο άνεμος. Οπως είναι γνωστό, ένα διάνυσμα περιγράφεται από το μέτρο και την διεύθυνση. Το διάνυσμα τού ανέμου περιγράφεται αντίστοιχα από την διεύθυνση και την ταχύτητα. Στην μετεωρολογία συναντάται και ενα άλλο είδος διανύσματος , η κλίση (gradient). Για παράδειγμα η κλίση (gradient) της πίεσης προς κάποια κατεύθυνση, περιγράφει την μεταβολή πίεσης κατά μήκος της μοναδιαίας οριζόντιας απόστασης, κατά μήκος μιας επιθυμητής διεύθυνσης. Η οποιαδήποτε μορφή στατιστικής ανάλυσης έχει νόημα για τις συνιστώσες του διανύσματος που αντιπροσωπεύει το φυσικό μέγεθος. Στην συνέχεια θα γίνει μια γενική επισκόπηση της τυποποίησης των συμβολισμών και ορισμών που σχετίζονται με την σχετική ανάλυση για την περίπτωση κατά την οποία το διανυσματικό μέγεθος είναι ο άνεμος.

Οι διευθύνσεις που χρησιμοποιούνται για τον προσδιορισμό των συνιστωσών είναι συνήθως η Βόρεια και η Ανατολική. Εστω V το μέτρο της ταχύτητας για οποιδήποτε διάνυσμα ανέμου και Θ η διεύθυνση του, μετρούμενη σύμφωνα με την διεύθυνση των δεικτών του ρολογιού ξεκινώντας απο τον Βορρά. Τότε οι συνιστώσες (Βόρεια, Ανατολική), δίνονται αντίστοιχα από τους παρακάτω τύπους:

$$V_N=V\cos\Theta \qquad \text{και} \qquad V_E=V\sin\Theta \tag{1}$$

Επίσης ισχύουν τα κάτωθι:

$$V^2=V^2\cos^2\Theta+V^2\sin^2\Theta \tag{2}$$

Επιπλέον ισχύει:

$$\tan\Theta=\sin\Theta/\cos\Theta=V_E/V_N \text{ και } \tan\Theta\cong\Theta \text{ για } \Theta \text{ μικρό} \tag{3}$$

Στην περίπτωση πολλών διανυσμάτων ανέμων, ΣV_N και ΣV_E ονομάζονται οι συνιστώσες του αθροίσματος.

Η μέση τιμή ενός διανυσματικού μεγέθους ορίζεται ως τό διάνυσμα με τιμή και διεύθυνση που ορίζονται ως ακολούθως:

$$V_R = \frac{1}{N}\left[\left(\sum V_N\right)^2 + \left(\sum V_E\right)^2\right]^{1/2} \tag{4}$$

$$\tan(a) = \frac{\sum V_E}{\sum V_N} \tag{5}$$

Η διεύθυνση α ονομάζεται επικρατούσα διεύθυνση και η τιμή του διανύσματος που αντιπροσωπεύει την μέση τιμή επικρατών άνεμος.

Τα ανωτέρω επεκτείνονται στον χώρο των τριών διαστάσεων. Οι αρχές της επεξεργασίας δεδομένων ανέμου περιγράφονται λεπτομερώς σε οποιδήποτε βιβλίο Εφαρμοσμένης Στατιστικής Ανάλυσης στην Μετεωρολογία [Brooks and Carrruthers,1953,].

1.3.2 Ελάχιστο πλήθος στοιχείων. Παρακάτω αναφέρεται το ελάχιστο πλήθος εγγραφών που απαιτείται από το κάθε μέγεθος, ωστε οι υπολογισμοί να έχουν κάποια στατιστική σημασία. Πλήρεις πίνακες αναφέρονται στην βιβλιογραφία [Guide to climatological Practices,1984].

(α) Θερμοκρασία αέρα:

ΠΙΝΑΚΑΣ 2

Ελάχιστο πλήθος τιμών θερμοκρασίας αέρα

Μέγιστη (ελάχιστη) θερμ. με περίοδο εμφάνισης 10,25,50,100 χρόνια.	^0C	50 χρόνια
Αθροίσματα θερμοκρασιών εκτός ορισμένων ορίων	^0C	30 χρόνια
Μέση μέγιστη ημερήσια θερμοκρασία στο επίπεδο αναφοράς της θαλάσσης	^0C	30 χρόνια
Μέση ελάχιστη ημερήσια θερμοκρασία στο επίπεδο αναφοράς της θαλάσσης	^0C	30 χρόνια
Μέση ημερήσια θερμοκρασία στο επίπεδο αναφοράς της θαλάσσης	^0C	30 χρόνια
Μέση θερμοκρασία του θερμότερου/ ψυχρότερου μήνα	^0C	20-30 χρόνια

(β) Σχετική Υγρασία:

ΠΙΝΑΚΑΣ 3

Ελάχιστο πλήθος τιμών σχετικής υγρασία

Διάρκεια περιόδων με σχετική υγρασία εκτός ορισμένων ορίων	Ωρα ή %	10 χρόνια

(γ)Εξάτμιση: Υπαρξη τιμών για 10-30 χρόνια αρκεί για τους επιθυμητούς υπολογισμούς.

(δ) Ηλιοφάνεια:

ΠΙΝΑΚΑΣ 4

Ελάχιστο πλήθος τιμών ηλιοφάνειας

Μέσος ορος ημερήσιας διάρκειας φωτεινής ηλιοφάνειας	Ωρα	15-20 χρόνια
Ποσοστιαία συχνότητα εμφάνισης ημερών χωρίς φωτεινή ηλιοφάνεια	%	30 χρόνια
Σχετική διάρκεια ηλιοφάνειας	Ωρα	15-20 χρόνια

(δ) Ηλιακή ακτινοβολία:[Ολική (global)]

ΠΙΝΑΚΑΣ 5

Ελάχιστο πλήθος τιμών ολικής ηλιακής ακτινοβολίας

Μέση ημερήσια τιμή της ολικής ακτινοβολίας σε οριζόντια επιφάνεια	MJm^{-2}	25 χρόνια
Μέγιστη ημερήσια τιμή της ολικής ακτινοβολίας σε οριζόντια επιφάνεια	MJm^{-2}	15-20 χρόνια

(ε) Κατ' ευθείαν (direct):

ΠΙΝΑΚΑΣ 6

Ελάχιστο πλήθος τιμών κατ' ευθείαν ηλιακής ακτινοβολίας

Πλήθος συνεχόμενων ημερών κατά την διάρκεια των οποίων η καταφθάνουσα ακτινοβολία δεν υπερβαίνει κάποιο όριο.	Ημέρα	>= 25 χρόνια

(στ)Ανεμος

Υπαρξη τιμών για 10-20 χρόνια αρκεί για τους επιθυμητούς υπολογισμούς.

2. ΜΕΘΟΔΟΛΟΓΙΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.

2.1 Βασικές ιδιότητες - Σχετικοί έλεγχοι

Πριν την εφαρμογή οποιασδήποτε μεθοδολογίας για την στατιστική ανάλυση των δεδομένων, συνίσταται η εφαρμογή απλών ελέγχων που παρέχουν κατευθυντήρια συμπεράσματα, χρήσιμα κατα την διάρκεια της στατιστικής ανάλυσης των δεδομένων. Δηλαδή, οι εν λόγω χρονοσειρές θα πρέπει να πληρούν βασικές ιδιότητες, ή στην αντίθετη περίπτωση, να γίνουν οι σχετικές διορθώσεις ώστε να ικανοποιούνται οι συνθήκες που προυποθέτουν την ισχύ των ζητουμένων ιδιοτήτων. Οι πλέον βασικές ιδιότητες είναι οι παρακάτω:

(α) Ομοιογένεια

(β) Τυχαιότητα

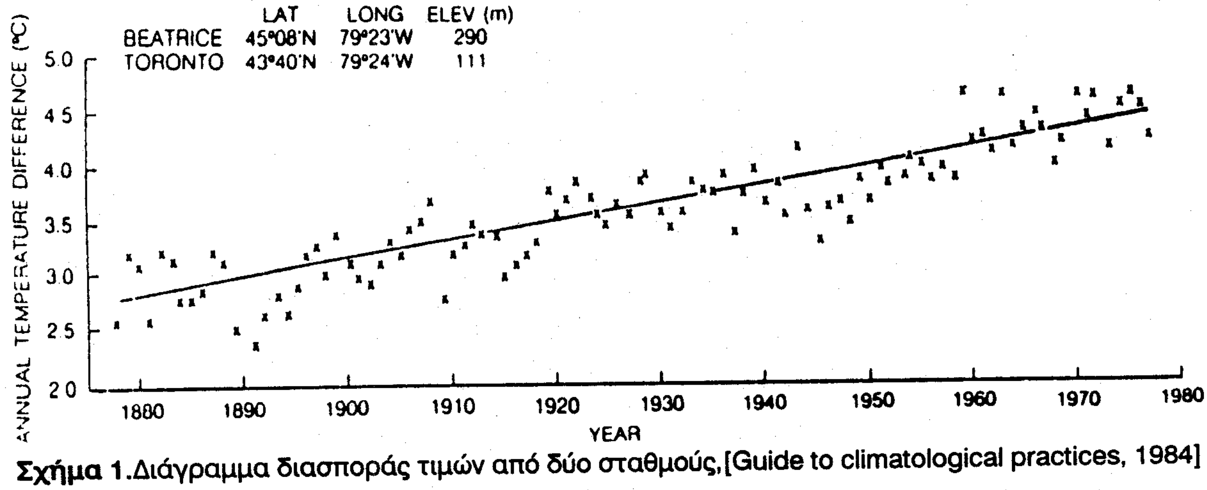
(γ) η μη ύπαρξη Ροπής (trend)

(δ) Στασιμότητα

2.1.1 Ομοιογένεια. Η πρώτη βασική ιδιότητα των καταγραφομένων τιμών είναι η ομοιογένεια. Προυπάρχουσα εμπειρία,[Tech. Note 79 ,WMO-No. 195], βεβαιώνει οτι οποιαδήποτε χρονοσειρά μεγαλύτερη των 10 η 20 χρόνων περιέχει κάποια μορφή ανομοιογένειας. Στήν συνέχεια θα αναφερθούν μέθοδοι προσδιορισμού και διόρθωσης.

(1) Γραφικές Μέθοδοι

(α) Διάγραμμα Διασποράς Τιμών (Scatter Diagram). Αποτελεί μια γραφική μέθοδο, που βοηθά στην γρήγορη εξαγωγή συμπερασμάτων. Υλοποιείται φτιάχνοντας το διάγραμμα των τιμών της υπό εξέταση μεταβλητής (άξονας των τεταγμένων), ως προς το χρόνο(άξονας των τετμημένων). Συχνά, στο ίδιο γράφημα τοποθετούνται δεδομένα απο δύο η περισσότερους σταθμούς. Ενα παράδειγμα τέτοιου διαγράμματος φαίνεται στο παρακάτω σχήμα.



Σχήμα 1.Διάγραμμα διασποράς τιμών από δύο σταθμούς,[Guide to climatological practices, 1984]

(β) Διάγραμμα Residuals. Μια ακόμα γραφική μέθοδος, η οποία παρέχει μια εικόνα της σχέσης των τιμών της χρονοσειράς ως προς τον μέσο όρο των τιμών αυτών. Η καμπύλη που εκφράζει την σχέση αυτή δίνεται από τον ακόλουθο τύπο:

$$y_i = y_{i-1} + (x_i - \overline{x}) = \sum_{j=1}^{i}\left\{x_j - \left(\frac{1}{N}\sum_{k=1}^{N} x_k\right)\right\} \quad (11)$$

όπου N είναι το πλήθος των σημείων της χρονοσειράς.

Η ερμηνεία της μορφής της καμπύλης, παρέχει τα επιθυμητά συμπεράσματα. Σταθερή οριζόντια κλίση, δείχνει κάποια ισορροπία των τιμών ως προς τον μέσο όρο. Αντίστοιχα στην περίπτωση αύξουσας η φθίνουσας κλίσης, υπονοείται η ύπαρξη κάποιας μορφής τάσης (trend).

(2) Μη παραμετρικοί έλεγχοι

Ενας έγκυρος έλεγχος ομοιογένειας περιλαμβάνει μια στατιστική διαδικασία που ξεκινά χρησιμοποιώντας σαν μηδενική (null) υπόθεση το ότι η σειρά είναι ομοιογενής, συνοδευόμενη απο κάποιο κανόνα με την βοήθεια του οποίου υπολογίζεται η πιθανότητα να είναι αληθής η μηδενική (null) υπόθεση. Αν η πιθανότητα να ισχύει η μηδενική υπόθεση είναι μικρή, τότε η χρονοσειρά θεωρείται ανομειογενής. Ο κανόνας παρέχει επιπλέον και το όριο της τιμής της πιθανότητας πέρα απο το οποίο η αρχική υπόθεση της ομοιογένειας πρέπει να απορριφθεί. Πολύ χρήσιμοι στην ανωτέρω διαδικασία είναι οι μη-παραμετρικοί έλεγχοι.

(α) Run tests

Μια ευρέως εφαρμοσμένη διαδικασία μη-παραμετρικού ελέγχου είναι τα run tests. Ο έλεγχος αυτός βασίζεται στην καταμέτρηση του αριθμού **u** των εμφανίσεων ενός γεγονότος. Το γεγονός αυτό, είναι ότι η τιμή των δεδομένων υπερβαίνει η όχι την τιμή μιας μέσης η κεντρικής τιμής, εξετάζοντας κάποιο πίνακα κατανομής των τιμών της προς επεξεργασία χρονοσειράς .

Για την καλύτερη κατανόηση της μεθόδου αναφέρεται το ακόλουθο παράδειγμα. Η χρονοσειρά αποτελείται απο τις μέσες μηνιαίες θερμοκρασίες του μηνός Αυγούστου στην Γενεύη[1].

[1]Tech. Note 81,WMO-199 [5]

ΠΙΝΑΚΑΣ 7

Μέσες μηνιαίες θερμοκρασίες του μηνός Αυγούστου στην Γενεύη.

1927	17.4	B	1942	19.9	A
1928	20.9	A	1943	20.9	A
1929	18.7	B	1944	22.9	A
1930	18.7	B	1945	18.9	B
1931	1.9	B	1946	19.2	A
1932	20.8	A			
1933	20.4	A	1947	22.0	A
1934	17.9	B	1948	18.9	B
1935	18.1	B	1949	20.7	A
1936	18.5	B	1950	19.7	A
			1951	19.5	A
1937	19.5	A	1952	20.3	A
1938	18.6	B	1953	19.8	A
1939	18.6	B	1954	18.3	B
1940	17.9	B	1955	19.3	A
1941	17.8	B	1956	17.5	B

Μελετώντας τις τιμές των θερμοκρασιών στον πινακα 2, η μέση τιμή, που χρησιμοποιείται σαν κατώφλι, βρίσκεται μεταξύ των τιμών 18.9 και 19.2, έστω 19.05. Σημειώνοντας με A και B αντίστοιχα τις μικρότερες η μεγαλύτερες τιμές από το κατώφλι, γίνεται η καταμέτρηση των ακολουθιών που απαρτίζονται από A και B. Στην εξεταζόμενη περίπτωση ειναι u=15.

Η περίπτωση αυξημένου αριθμού runs αντιστοιχεί σε ένδειξη ύπαρξης ταλαντώσεων (oscillations) ενώ μειωμένη τιμή υπονοεί την ύπαρξη κάποιας τάσης (trend), η κάποιας μετατόπισης (shift).

2.1.2 Τυχαιότητα (Randomness). Η πλέον βασική και απαραίτητη προυπόθεση είναι αυτή της **τυχαιότητας**, (randomness), των τιμών της ομογενοποιημένης χρονοσειράς. Τις περισσότερες φορές, οι προς επεξεργασία τιμές παρουσιάζουν στην σύνθεση τους, στοιχεία μη-τυχαιότητας που εκφραζόνται με την παρουσία κάποιας περιοδικότητας (periodicity), τάσης (trend), cycles, εμμονής (persistence), η κάποιου άλλου μη-τυχαίου στοιχείου η κάποιου συνδυασμού των ανωτέρω μορφών. Εφόσον, με την χρήση των καταλλήλων στατιστικών μεθόδων αποδειχθεί ότι η χρονοσειρά περιέχει ολικά η μερικά, κάποια μη-τυχαία χαρακτηριστικά, θα πρέπει πριν τα δεδομένα αυτά χρησιμοποιηθούν για οποιαδήποτε εξαγωγή συμπερασμάτων, να γίνει αφαίρεση αυτών των υπαρχόντων μη-τυχαίων συνθετικών στοιχείων.

Ιδιαίτερη προσοχή χρειάζεται στην εκλογή των μεθόδων ανίχνευσης και αφαίρεσης, διότι ως γνωστόν υπάρχει ισχυρή διαφοροποίηση ανάλογα με το είδος του μη-τυχαίου χαρακτηριστικού της χρονοσειράς. Πιο συγκεκριμένα, οι γενικοί έλεγχοι τυχαιότητας, στην περίπτωση που ανιχνεύουν κάποια μορφή μη-τυχαίου χαρακτηριστικού, πρέπει να ακολουθούνται από ελέγχους αναγνώρισης του είδους της μη-τυχαιότητας. Στην βιβλιογραφία, Tech. Note 79 - WMO-195,[4], εξηγούνται αναλυτικά πλήθος τέτοιων μεθόδων. Οι περισσότερο συχνά χρησιμοποιούμενοι είναι οι παρακάτω αναφερόμενοι.

(1) Serial correlation[2]

Εχοντας τήν σειρά $x_1, x_2 \ldots x_n$, γιά την οποία ισχύει $\sum x_i = 0$, θεωρείται ότι ο εκτιμητής είναι της μορφής

$$R = \sum_{i=1}^{n} x_i x_{i+1} \tag{12}$$

Η μηδενική υπόθεση υπαγορεύει ότι η κατανομή των τιμών της R είναι Gaussian με μέση τιμή καί διασπορά που δίνονται από τους τύπους

$$E(R) = -\frac{\sum x_i^2}{n-1}$$

$$\operatorname{var} R = \frac{\left(\sum x_i^2\right)}{n-1} \tag{13}$$

Η μέθοδος αυτή επεκτείνεται σε πολλά είδη επιμέρους ελέγχων πού περιγράφονται με περισσότερες λεπτομέρειες στην σχετική βιβλιογραφία.

(2) Ο λόγος του VON NEUMANN[3]

Ο αριθμός του Von Neumann ορίζεται:

$$V = \frac{n}{n-1} * \frac{\sum_{i=1}^{n-1} (x_i - x_{i+1})^2}{\sum_{i=1}^{n} (x_i)^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \tag{14}$$

---

[2][Tech. Note143-WMO 415],[3]
[3][Tech. Note79-WMO 195],[4]

οπου n το πλήθος των στοιχείων της σειράς. Για μεγάλη τιμή τού n (όχι λιγότερα των 50), αν πρόκειται για τυχαίες χρονο-τιμές, οι τιμές του V υπακούουν σε μια Κανονική κατανομή με μέση τιμή και διασπορά:

$$E(V) = \frac{2n}{n-1}$$

$$\operatorname{var} V = \frac{4(n-2)}{(n-1)^2} \qquad (15)$$

(3) Student's T-test[4]

Αυτό το είδος ελέγχου προσφέρει, εκτος της τυχαιότητας, χρήσιμα συμπεράσματα στήν ύπαρξη τάσης. Βασίζεται στην σύγκριση των μέσων τιμών της χρονοσειράς σε δύο διαφορετικές περιόδους των εγγραφών, που σύμφωνα με την αρχική υποθεση υπακούουν στην κατανομή Student's T.

2.1.3 Στασιμότητα (Stationarity). Η μελέτη για την ύπαρξη η μη της τυχαιότητας, περιλαμβάνεται στους ελέγχους που εξασφαλίζουν την πληρότητα μιας εξ ίσου σημαντικής ιδιότητας που ονομάζεται στασιμότητα (stationarity).

Η βασική ιδιότητα που ικανοποιούν οι στάσιμες χρονοσειρές είναι αυτή της στατιστικής ισορροπίας, (statistical equilibrium). Στατιστική ισορροπία σημαίνει ότι οι στατιστικές ιδιότητες της διεργασίας είναι ανεξάρτητες από το σημείο το οποίο θεωρείται ως χρονική αρχή. Με άλλα λόγια η κατανομή πιθανότητας είναι αναλλοίωτη ως προς την χρονική μετατόπιση. Σ' αυτή την περίπτωση η μορφή της κατανομής πιθανότητας σε κάποιο χρονικό διάστημα, είναι συνάρτηση των διαφορών χρονικών τιμών και όχι απολύτων τιμών.

Εξ αιτίας της ανεξαρτησίας ως προς την χρονική μετατόπιση, συμπεραίνουμε ότι μπορούν να βγουν συμπεράσματα για το σχήμα και τις ιδιότητες της κατανομής τέτοιου είδους φτιάχνοντας ένα απλό ιστόγραμμα από τις τιμές της χρονοσειράς.

(1) Μελέτη της στασιμότητας με τη χρήση των συναρτήσεων αυτοσυσχέτησης (autocorrelation) και συμμεταβολής (autocovariance). Τέτοιου είδους διαδικασίες περιγράφονται με την βοήθεια του μέσου όρου, της διασποράς και της συνάρτησης αυτοσυσχέτησης η της συνάρτησης φασματικής πυκνότητας (spectral density function). Οι βασικές θεωρητικές αρχές αναπτύσονται στο **Παράρτημα Α**. Το αδύνατο σημείο της μεθόδου βρίσκεται στο ότι όσο μεγαλώνει το πλήθος των τιμών της χρονοσειράς, τόσο αυξάνουν οι συνθήκες που πρέπει να πληρούνται. Το μειονέκτημα αυτό ξεπερνιέται διαλέγοντας την μέθοδο της **φασματικής ανάλυσης** η οποία περιγράφεται σε επόμενο κεφάλαιο.

---

[4][Tech. Note79-WMO 195],[4]

Ας θεωρήσουμε την **πεπερασμένη** χρονοσειρά $z_1,z_2,...,z_N$ N στοιχείων, απο τα οποία θα γίνει κάποια εκτίμηση των τιμών των συναρτήσεων αυτοσυσχέτησης και συμμεταβολής. Ενας απο τους τρόπους εκτίμησης των συντελεστών συσχέτησης και συμμεταβολής που προτείνει η στατιστική ανάλυση, [Box and Jenkins,1976], είναι ο εξής:

$$r_K = \frac{c_K}{c_0} \tag{16}$$

όπου

$$c_K = \frac{1}{N}\sum_{t-1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z}) \qquad k=0,1,2,...K \tag{17}$$

Τα σύμβολα r_K και c_K αναφέρονται στις εκτιμήσεις των ρ_K και γ_K αντοίστοιχα, όπου, όπως έχει σημειώνεται στο Παράρτημα Α, ρ και κ είναι σύμβολα για την αυτοσυσχέτηση και την συμμεταβολή. Το $\bar{z}$ είναι ο μέσος όρος των στοιχείων της χρονοσειράς.

Η αξιοπιστία της εκτίμησης, οπως αναφέρθηκε ανωτέρω, εξαρτάται από το πλήθος των στοιχείων. Μια αρκετά καλή προσέγγιση απαιτεί την ύπαρξη **τουλάχιστον 50 τιμων** για τις οποίες η αυτοσυσχέτιση r_K υπολογίζεται για τιμές κ=0,1,2,...,Κ όπου το Κ δεν υπερβαίνει το N/4, όπου N το πλήθος των στοιχείων της χρονοσειράς.

(ι) Συνθήκη Τυχαιότητας: Χρησιμοποιώντας τις ανωτέρω έννοιες, για μια τυχαία χρονοσειρά ισχύει ότι: Η συνάρτηση αυτοσυσχέτισης είναι 0.

(2) Αρμονική Ανάλυση των Στασίμων Διαδικασιών. Ενας άλλος τρόπος μελέτης μιας χρονοσειράς είναι η μέθοδος της αρμονικής ανάλυσης. Η ιδέα στην οποία στηρίζεται η μέθοδος είναι στην ισχύ της υπόθεσης για την ύπαρξη η μη **περιοδικών στοιχείων** στην χρονοσειρά. Τα στοιχεία αυτά είναι κάποιο ημίτονο γνωστής συχνότητας και αγνώστου μέτρου (amplitude). Το εργαλείο μέσω του οποίου πραγματοποιείται ο έλεγχος της ισχύος της παραπάνω υπόθεσης είναι η γραφική παράσταση των αρμονικών της χρονοσειράς, με τον ακόλουθο τρόπο. Υποθέτουμε οτι η χρονοσειρά αποτελείται απο **περιττό αριθμό στοιχείων**, δηλαδή N=2q+1. Χρησιμοποιώντας το μοντέλο της σειράς Fourier, το κάθε στοιχείο της χρονοσειράς μπορεί να γραφεί ως εξης.

$$z_t = a_0 + \sum_{i=1}^{q}\left(a_i c_{it} + \beta_i s_{it}\right) + e_t \tag{18}$$

οπου τα εμφανιζόμενα μεγέθη στη σχέση (1) ορίζονται,

$$c_{it} = \cos 2\pi f_i t$$
$$s_{it} = \sin 2\pi f_i t$$
$$f_i = \frac{i}{N} \tag{19}$$

Το f_i αντιπροσωπεύει τον $i^{ον}$ αρμονικό της αρχικής συχνότητας 1/N. Οι συντελεστές a_0, a_i, β_i υπολογίζονται ως ακολούθως.

$$a = \bar{z}$$
$$a_i = \frac{2}{N}\sum_{t=1}^{N} z_t c_{it}$$
$$\beta_i = \frac{2}{N}\sum_{t=1}^{N} z_t s_{it} \qquad i=1,2,\ldots,q \tag{20}$$

Η γραφική παράσταση των αρμονικών αποτελείται απο q=(N-1)/2 τιμές που δίνονται απο τον ακόλουθο τύπο,

$$I(f_i) = \frac{N}{2}\left(a_i^2 + b_i^2\right) \qquad i=1,2,\ldots,q \tag{21}$$

όπου $I(f_i)$ καλείται ισχύς (intensity) στην συχνότητα f_i. Ισχύει ότι

$$\frac{1}{N}\sum_{t=1}^{N}(z_t - \bar{z})^2 = \sum_{i=1}^{q} I(f_i) \tag{22}$$

Στην περίπτωση που το N είναι **άρτιος αριθμός**, δηλαδή N=2q, οι παραπάνω σχέσεις (3),(4) ισχύουν για i=1,2,...,q-1. Για i=q ισχύει ότι

$$a_q = \frac{1}{N}\sum_{t=1}^{N}(-1)^t z_t$$
$$b_q = 0$$
$$I(f_q) = N a_q^2 \tag{23}$$

Η αρμονική ανάλυση παρέχει την σχετική πληροφορία για το πώς η διασπορά της χρονοσειράς είναι κατανεμημένη στις αρμονικές συχνότητες.

<u>(ιι) Συνθήκη Τυχαιότητας</u>: Αν πρόκειται για τυχαία χρονοσειρά, χωρίς κανένα ημιτονοειδές συστατικό σήμα (στοιχείο περιοδικότητας), δηλαδή:

$$z_t=a_0+e_t \quad (24)$$

όπου a_0 ισούται με τον μέσο όρο και τα e υπακούουν σε κάποια Κανονική κατανομή, με μέσο όρο μηδέν και διασπορά σ^2 τότε, σε αυτή την περίπτωση, κάθε $I(f_j)$, (ανεξάρτητα απο τα υπόλοιπα q), θα πρέπει να έχει μέση τιμή ίσον με $2\sigma^2$ και θα πρέπει να υπακούει στην κατανομή $\sigma^2\chi^2(2)$, (όπου $\chi^2(2)$ είναι κάποια τυχαία μεταβλητή κατανεμημένη σύμφωνα με την χ^2 κατανομή, έχοντας δύο βαθμούς ελευθερίας).

Συνθήκη Υπαρξης Περιοδικότητας: Αν η χρονοσειρά περιέχει κάποια περιοδικότητα που εκφράζεται με την βοήθεια κάποιου ημιτονοειδούς με συχνότητα f_j, πλάτος A και γωνία φάσης F, δηλαδή:

$$z_t=a_0+a\cos(2f_j\pi t)+\beta\sin(2f_j\pi t)+e_t \quad (25)$$

και $A\sin(F)=a$,$A\cos(F)=\beta$, τότε αθροίζοντας τα τετράγωνα των $I(f_j)$ παρατηρείται γρήγορη ανοδική τάση της τιμής του αθροίσματος, το οποίο αναμένεται να ισούται με

$$2\sigma^2+N(a^2+\beta^2)/2=2\sigma^2+NA^2/2 \quad (26)$$

(3) Φασματική Ανάλυση Στασίμων Χρονοσειρών. Ενα πολύ χρήσιμο εργαλείο στην μελέτη των χρονοσειρών όσον αφορά την ανίχνευση των μη τυχαίων στοιχείων αποτελεί η φασματική ανάλυση, η οποία ουσιαστικά αποτελεί μια επέκταση της μεθόδου της αρμονικής άναλυσης που αναπτύχθηκε ανωτέρω, για τις περιπτώσεις που οι συχνότητες των αρμονικών δεν συνδέονται με το μήκος της χρονοσειράς, όπως συμβαίνει στη μέθοδο της αρμονικής ανάλυσης. Σε αναλογία με την αρμονική ανάλυση, η φασματική ανάλυση δείχνει πώς η διασπορά μιας στοχαστικής διαδικασιάς ειναι κατανεμημένη σε κάποιο συνεχές διάστημα συχνοτήτων. Οι απαραίτητοι ορισμοί των εμπλεκομένων εννοιών παρατίθενται στο **Παράρτημα Β.**

Απο την μορφή του φάσματος είναι δυνατή η εξαγωγή συμπερασμάτων για την φύση των στοιχείων της επεξεργαζομένης χρονοσειράς. Πιο συγκεκριμένα,

(ιι) Συνθήκη Τυχαιότητας: Το φάσμα κάποιας τυχαίας χρονοσειράς εχει τετραγωνική μορφή.

Επιρροή των τιμών των μη τυχαίων χαρακτηριστικών στην μορφή του φάσματος της χρονοσειράς: Απο την θεωρία της φασματικής ανάλυσης και απο την μέχρι τώρα εμπειρία σε χρονοσειρές μετεωρολογικών δεδομένων,[Tech. Note 79 ,WMO-No. 195], εξάγονται τα ακόλουθα συμπεράσματα.

* Αν η περιοδικότητα που εμπεριέχεται στις τιμές της χρονοσειράς είναι **ημιτονοειδούς μορφής** , το φάσμα θα περιέχει κάποια κορυφή (peak), στην θέση της αντίστοιχης συχνότητας.

Στην πρακτική εφαρμογή αυτού του είδους οι περιοδικότητες ανιχνεύονται με την αρμονική ανάλυση.

* Στην περίπτωση **μη-ημιτονοειδούς** περιοδικότητας εμφανίζονται περισσότερες κορυφές ανάλογα με τους αρμονικούς που εμπεριέχει η μορφή της περιοδικότητας.

* Στην περίπτωση της **σχεδόν-περιοδικότητας**, το φάσμα θα περουσιάζει μια hump που θα επεκτείνεται σε ένα ευρύ φάσμα συχνοτήτων.

* Οταν η χρονοσειρά εμπεριέχει εμμονή(**persistence**), που σημαίνει οτι κάθε τιμή επηρεάζεται απο την ακριβώς προηγούμενη της, το φάσμα θα παρουσιάζει παραμόρφωση (distortion) σε ολες τις συχνότητες. Πιο συγκεκριμένα, το πλάτος θα τείνει να μειούται απο τις χαμηλότερες προς τις ψηλότερες συχνότητες.

Οι παραπάνω αναφερόμενες ιδιότητες ειναι ενδεικτικές και δεν καλύπτουν καθ' ολοκληρία τις ιδιότητες και την χρήση της φασματικής ανάλυσης. Περισσότερο λεπτομερειακές αναφορές βρίσκονται στην βιβλιογραφία, [1], [4], [5], [6], [7].

2.1.4 Υπαρξη τάσης (Trend). Τα επόμενα δύο είδη μη παραμετρικών ελέγχων χρησιμοποιούνται στην ανίχνευση ύπαρξης τάσης(trend). Το πρώτο βασίζεται στον υπολογισμό του συντελεστή του Spearman (r_S) και το δεύτερο στον rank συντελεστή του Kendall (t).

(1) Δοκιμή Spearman. Το πρώτο βήμα της μεθόδου περιλαμβάνει την ανακατάταξη των τιμών της χρονοσειράς, έστω x_i,i=1,2,... κατά αύξουσα σειρά οπότε προκύπτει η σειρά y_i. Κατόπιν, κάθε τιμή της σειράς y_i, αντικαθίσταται από ένα νούμερο το οποίο υποδηλώνει την θέση της κάθε τιμής ανάλογα με το μέγεθος της στην καινούργια σειρά y_i. Ο συντελεστής r_S υπολογίζεται βάση του τύπου, [Tech. Note 143-WMO 415]:

$$r_s = 1 - \frac{6}{n\left(n^2-1\right)}\sum\left(y_i - i\right)^2 \tag{27}$$

όπου n το πλήθος των τιμών της χρονοσειράς. Για την επαλήθευση της μηδενικής υπόθεσης, η κατανομή που υπακούει ο παραπάνω συντελεστής είναι κανονική με μέση τιμή και διασποράς που δίνονται απο τους ακόλουθους τύπους:

$$E\left(r_s\right) = 0$$
$$\operatorname{var}\left(r_s\right) = \frac{1}{n-1} \tag{28}$$

Ακολούθως, με βάση μια κανονική κατανομή, υπολογίζεται η τιμή της πιθανότητας a_1, έτσι ώστε:

$$a_1 = P\left(|u| > |u(r_s)|\right)$$
$$u(r_s) = r_s\sqrt{n-1} \qquad (29)$$

Η απόρριψη η αποδοχή της μηδενικής υπόθεσης εξαρτάται απο το αν

$$a_1 > a_0 \qquad \text{η} \qquad a_1 < a_0 \qquad (30)$$

Στην περίπτωση που ανιχνεύεται κάποια τάση, θεωρείται αύξουσα εφόσον $r_S>0$, ενώ φθίνουσα όταν $r_S<0$.

(2) Δοκιμή Kendall. Η μέθοδος προυποθέτει την ανακατάταξη των τιμών και την δημιουργία της σειράς y_i όπως προηγουμένως. Μετά γίνεται η σύγκριση της κάθε τιμής της χρονοσειράς με όλες τις επόμενες τιμές. Για παράδειγμα, η τιμή y_1, συγκρίνεται με όλες τις τιμές αρχίζοντας από την δεύτερη και καταλήγοντας στην νιοστή, η y_2 αρχίζοντας απο την τρίτη και καταλήγοντας στην νιοστή, και ουτω καθεξής. Κατά την διάρκεια της σύγκρισης έστω της y_i, υπολογίζεται ο αριθμός των τιμών της σειράς που υπερβαίνουν την y_i. Ο αριθμός αυτός δηλώνεται ως n_i. Ο εκτιμητής έχει την μορφή

$$t = \sum_i n_i \qquad (31)$$

και υπακούει σε μια κανονική κατανομή με χαρακτηριστικά μεγέθη:

$$E(t) = \frac{n(n-1)}{4}$$
$$\mathrm{var}(t) = \frac{n(n-1)(2n+5)}{72} \qquad (32)$$

Στην περίπτωση αυτή η πιθανότητα a_1, με την βοήθεια των τιμών μιας κανονικής κατανομής, δίνεται από

$$a_1 = P(|u| > |u(t)|)$$
$$u(t) = \frac{[t - E(t)]}{\sqrt{\mathrm{var}(t)}} \qquad (33)$$

Ομοια με την περίπτωση της μεθόδου Spearman, η τιμή της a_1 καθορίζει την αποδοχή η μη της μηδενικής υπόθεσης. Αύξουσα τάση στις τιμές της σειράς, υπάρχει όταν $u(t)>0$, ενώ στην αντίθετη περίπτωση η τάση είναι φθίνουσα.

3. ΜΕΘΟΔΟΙ ΣΥΜΠΛΗΡΩΣΗΣ/ ΑΝΤΙΚΑΤΑΣΤΑΣΗΣ ΤΙΜΩΝ ΤΗΣ ΧΡΟΝΟΣΕΙΡΑΣ

Το πρώτο επίπεδο της διαδικασίας επεξεργασίας των δεδομένων είναι η εφαρμογή των ήδη αναφερθέντων ελέγχων (βλ. 2.1.1) για την ύπαρξη η μη της ομοιογένειας των τιμών. Στην περίπτωση που ανιχνεύεται κάποιας μορφής ανομοιογένεια, αρχίζει η εφαρμογή διαφόρων μεθόδων ομογενοποίησης. Προσοχή αξίζει το γεγονός ότι, ομογενοποίηση των τιμών δεν σημαίνει σχηματισμός κάποιας καινούργιας σειράς τιμών δεδομένων. Αυτό που γίνεται είναι η αναπροσαρμογή κάποιων στατιστικών ιδιοτήτων των τιμών της χρονοσειράς και κάποιες εκτιμήσεις κάποιων εγγραφών που **λείπουν** η είναι **αμφίβολες**.

Πολύ συχνά δύο χρονοσειρές που βρίσκονται σε κάποια συσχέτιση αποτελούνται από διαφορετικό μεγεθος μετρήσεων. Αυτό συμβαίνει διότι υπάρχουν διαφορετικά είδη σταθμών, όπως οι Τοπικοί, και οι Κεντρικοί. Στη πρώτη περίπτωση οι διαθέσιμες χρονοσειρές δεν υπερβαίνουν τα δέκα, είκοσι και σπάνια τα 30 χρόνια, ενώ στην δεύτερη, τα 100 και επιπλέον χρόνια. Εξ αιτίας της συσχέτισης που παρουσιάζουν οι Τοπικοί με τους Κεντρικούς σταθμούς, είναι δυνατή η συμπλήρωση τιμών στην σειρά με το μικρότερο πλήθος στοιχείων, καθώς επίσης και η επιμήκυνση των σειρών αυτών.

Επίσης στις παρακάτω μεθόδους, εμφανίζεται η έννοια του **γειτονικού** σταθμού. Η έννοια του **γειτονικού** σταθμού συνδέεται με την έννοια του μετεωρολογικού συστήματος (meteorological system), και έχει ορισθεί στην βιβλιογραφία ως εξής: Ενας σταθμός Α είναι γειτονικός με κάποιο σταθμό Β, εφόσον ανήκουν στην ακτίνα εμβέλειας (scale) του ίδιου μετεωρολογικού συστήματος. Παραδείγματος χάριν, εάν η ακτίνα εμβέλειας (scale) είναι 1000 km, τότε όλοι οι σταθμοί που ανήκουν στό ίδιο μετεωρολιγικό σύστημα και απέχουν λιγότερο από 1000 km θεωρούνται γειτονικοί.

3.1 Μέθοδοι των Διαφορών/Λόγων.

Δύο θεωρητικά αξιόπιστες καί ευρέως διαδεδομένες μέθοδοι είναι: Η Μέθοδος των Λόγων και η Μέθοδος των Διαφορών. Εχει παρατηρηθεί ότι όταν ταυτόχρονα καταγραφόμενες τιμές από γειτονικούς σταθμούς συγκρίνονται, η διαφορά των τιμών η ο λόγος τους είναι περίπου σταθερός. Αυτό, αν και δεν αληθεύει για πρωτογενείς παρατηρήσεις (single observations), ημερήσια αθροίσματα (daily sums), η ημερήσιους μέσους όρους (daily mean values), συμβαίνει σχεδόν πάντα για τα μηνιαία και ετήσια αθροίσματα και μέσες τιμές (monthly, annual sums and means).

3.1.1 Μέθοδος των Διαφορών. Η μέθοδος εφαρμόζεται κυρίως σε χρονοσειρές θερμοκρασιών και χρησιμοποιεί τις ταυτόχρονες εγγραφές από δύο μετεωρολογικούς σταθμούς. Σ' αυτή την περίπτωση η διαφορά d μεταξύ των τιμών των σταθμών Α και Β ορίζεται ως εξής.

$$d=\frac{\sum(b_i-a_i)}{N} \tag{34}$$

Αν η τιμή b_i (από τις εγγραφές του σταθμού Β) λείπει η είναι αμφίβολη, τότε η εκτιμητέα τιμή που αντικαθιστά η συμπληρώνει την b_i, σε συνάρτηση με τις τιμές a_i που έχουν καταγραφεί στον σταθμό Α, δίνεται από τον ακόλουθο τύπο.

$$b_i = a_i + d \tag{35}$$

Στό παρακάτω παράδειγμα, φαίνεται η χρήση της ανωτέρω μεθοδολογίας. Τά δεδομένα αφορούν δύο σταθμούς στον Νότιο Καναδά[1] ,έστω Α και Β, οι οποιοι απέχουν 200 Km, με αρχεία 30 και 15 χρόνων αντίστοιχα. Οι τιμές φαίνονται στόν παρακάτω πίνακα

ΠΙΝΑΚΑΣ 8

Μέσες Ημερήσιες Θερμοκρασίες γιά τόν Μήνα Ιούλιο (^{0}C)

	Σταθμός Α a_i	Σταθμός Β b_i
1951-1965	20.9	18.1
1941-1970	20.7	?

Το ζητούμενο είναι να υπολογισθεί η ελλείπουσα τιμή. Εφαρμόζοντας τους τύπους (34) και (35) η ζητουμένη θερμοκρασία είναι 17.9 ^{0}C.

3.1.2 Μέθοδος των Λόγων(ratio). Η μέθοδος αυτή εφαρμόζεται κυρίως για χρονοσειρές δεδομένων βροχής, ταχύτητας του ανέμου, ημερήσια ηλιοφάνεια. Σ' αυτή την περίπτωση ο λόγος q των συγχρόνως καταγραφομένων τιμών στους σταθμούς Α και Β ορίζεται ως εξής.

$$q = \frac{\sum b_i}{\sum a_i} \tag{36}$$

Αν η τιμή b_i (από τις εγγραφές του σταθμού Β) λείπει η είναι αμφίβολη, τότε η εκτιμητέα τιμή που αντικαθιστά η συμπληρώνει την b_i, σε συνάρτηση με τις τιμές a_i που έχουν καταγραφεί στον σταθμό Α, δίνεται από τον ακόλουθο τύπο.

$$b_i = a_i * q \tag{37}$$

[1]Guide to climatological practices,[1]

Στό παρακάτω παράδειγμα, φαίνεται η χρήση της παραπάνω μεθοδολογίας. Τά δεδομένα αφορούν δύο σταθμούς στον Νότιο Καναδά[2] ,έστω Α και Β, οι οποιοι απέχουν 200 Χμ., με αρχεία 30 και 15 χρόνων αντίστοιχα. Οι τιμές φαίνονται στόν παρακάτω πίνακα:

ΠΙΝΑΚΑΣ 9

Ολική εναπόθεση (total precipitation) γιά τόν Μήνα Ιούλιο (mm)

	Σταθμός Α a_ι	Σταθμός Β b_ι
1951-1965	66.9	84.3
1941-1970	74.9	?

Απλή εφαρμογή των εξισώσεων (36) και (37) δίνει $q=a_i/b_i=1.26$ και $b_i=94.3$ (mm).

3.2 Μέθοδοι Παρεμβολής- Συσχέτισης

3.2.1 Υπολογισμός Μέσης Τιμής

Η ζητουμένη τιμή εκτιμάται βάσει του <u>μέσου</u> <u>όρου</u> των τιμών του δείγματος.

Στήν περίπτωση που προυπάρχουσα εμπειρία βεβαιώνει <u>παρόμοια</u> συμπεριφορά μέσων μηνιαίων τιμών ενός μεγέθους κατά την διάρκεια μιας χρονικής περιόδου Α (εστω ενός έτους), τότε οποιαδήποτε αμφισβητούμενη τιμή αντικαθίσταται από την αντοίστοιχη μέση μηνιαία τιμή κάποιας αντίστοιχης χρονικής περιόδου Β(ενός άλλου έτους).

Σε άλλες περιπτώσεις, ανάλογα με το μετρούμενο μέγεθος, η τιμή που λείπει τίθεται ίση με την ακριβώς προηγούμενη.

3.2.2 Γραφήματα Συσχέτισης Σειρών σέ Χρονική Μετατόπιση (Relation Curves). Ενα γράφημα συσχέτισης παρουσιάζει τήν συναρτησιακή σχέση στην οποία βρίσκονται δύο σειρές της μορφής

$$Y_t=F(X_{t+t1}) \tag{38}$$

όπου t_1 είναι η χρονική μετατόπιση στην οποία βρίσκονται οι δύο σειρές. Η γραφική παράσταση της παραπάνω σχέσης δίνει οποιαδήποτε τιμή της εξαρτημένης Υ ως προς Χ. Εαν συμβεί κάποια χρονική μετατόπιση των σειρών για $t_1<0$ τότε γίνεται δυνατή η πρόβλεψη τιμών της εξαρτημένης Υ ως προς Χ.

Οι πιο συχνά συναντούμενη μορφή των παραπάνω μαθηματικών συναρτήσεων είναι:

[2]Guide to climatological practices,[1]

* Πολυωνυμική:

$$Y_t=a_0+a_1X_i+a_2(X_i)^2+a_3(X_i)^3 \quad \text{όπου } i=t+t_1 \tag{39}$$

Η παραπάνω μέθοδος είναι δυνατόν να εφαρμοσθεί σε χρονοσειρές προερχόμενες από τον ίδιο η γειτονικούς σταθμούς.

3.2.3 Γραφήματα Συσχέτισης Σειρών. Η βασική ιδέα ομοιάζει με τη παραπάνω αναφερθείσα, με την διαφορά οτι οι χρονοσειρές δεν βρίσκονται σε μετατόπιση. Πιθανές μορφές συναρτήσεων συσχέτισης περιγράφονται στην συνέχεια

α) Πολυωνυμική:

$$Y_i=a_0+a_1X_i+a_2(X_i)^2+a_3(X_i)^3 \tag{40}$$

β) Απλή Γραμμική:

$$Y_i=a+bX_i \tag{41}$$

γ) Εκθετική:

$$Y_i=a\exp(bX_i) \quad \text{ή}$$
$$Y_i=a\exp(b/X_i) \tag{42}$$

δ) Λογαριθμική:

$$Y_i=a+b\ln(X_i) \tag{43}$$

Οι συντελεστές a, b υπολογίζονται με την μέθοδο των ελαχίστων τετραγώνων.

(ε) Cubic Spline

Οι σχετικές εξισώσεις είναι:

$$\frac{x_j-x_{j-1}}{6}y''_{j-1}+\frac{x_{j+1}-x_{j-1}}{3}y''_j+\frac{x_{j+1}-x_j}{6}y''_{j+1}=$$
$$=\frac{y_{j+1}-y_j}{x_{j+1}-x_j}-\frac{y_j-y_{j-1}}{x_j-x_{j-1}} \tag{44}$$

Στις παραπάνω N-2 εξισώσεις προστίθενται οι οριακές συνθήκες, στα σημεία 1 και N, για να εξασφαλισθεί μοναδικότητα στην λύση. Στην προκειμένη περίπτωση, εξισώνονται οι τιμές των πρώτων και δευτέρων παραγώγων στα οριακά σημεία.

Οπως και η προηγούμενη έτσι και αυτή η μέθοδος μπορεί να εφαρμοσθεί σε σειρές καταγραμμένες από τον ίδιο η γειτονικούς σταθμούς.

3.2.4 Πολλαπλή Συσχέτηση Χρονοσειρών

(1) Μέθοδος των fractiles.

Ακολουθεί μια μικρή υπενθύμιση του τρόπου υπολογισμού των fractiles: Οι τιμές μιας χρονοσειράς, n στοιχείων, τοποθετούνται κατά αύξουσα σειρά και ακολούθως γίνεται κάποια καταμέτρηση ενός τμήματος (fraction) των τιμών της χρονοσειράς, με πλήθος n+1. Για παράδειγμα, για να υπολογιθούν, τα quartiles διαιρούνται οι n+1 δια του 4, τα deciles δια του 10, και τα percentiles δια του 100. Στην περίπτωση που απο την διαίρεση προκύπτει μη ακέραιο νούμερο απαιτείται παρεμβολή. Εστω η ακόλουθη σειρά δεδομένων.

ΠΙΝΑΚΑΣ 10[3]

Δεδομένα εναπόθεσης (precipitation, mm)

Ετος	0	1	2	3	4	5	6	7	8	9
193-	108	171	62	67	119	157	23	78	79	85
194-	18	105	48	41	44	133	158	54	72	49
195-	110	100	125	57	206	107	144	58	148	44
196-	154	67	26	189	55	147	68	105	247	71
197-	89	86	52	55	37	153	78	90	126	62

Ανακατανομή κατά αύξουσα σειρά δίνει τον παρακάτω πίνακα:

ΠΙΝΑΚΑΣ 11

18	54	71	105	147
23	55	72	105	148
26	55	78	107	153
37	57	78	108	154
41	58	79	110	157
44	62	85	119	158

[3]Ref [1]

44	62	86	125	171
48	67	89	126	189
49	67	90	133	206
52	68	100	144	247

Προκύπτει ότι το πρώτο quartile είναι το (n+1)/4 = 12.75 μέλος των κατά αύξουσα σειρά ανακατανεμημένων τιμών,το 15ο percentile ισούται με το [15*(n+1)]/100=7.65ο μέλος. Με την μέθοδο της παρεμβολής βρίσκεται ότι το 7.65ο μέλος ισούται με:

[7η τιμή] +0.65*[8η τιμή-7η τιμή] (45)

Εστω, οι χρονοσειρές X και Y, με το μεγαλύτερο και το μικρότερο πλήθος όρων αντίστοιχα. Υποθέτουμε ότι η στατιστική κατανομή των X και Y περιγράφεται πλήρως απο τις παραμέτρους μ και μ_1 αντίστοιχα, καθώς επίσης τις σ και σ_1 και την ενδιάμεση παράμετρο t, η οποία είναι κοινή και για τις δύο χρονοσειρές.

Σύμφωνα με την μέθοδο, τα fractiles της ίδιας τάξης των μεταβλητών X,Y και t, συνδέονται με την σχέση:

$$x = \sigma * t + \mu$$

$$y = \sigma_1 * t + \mu_1 \quad (46)$$

Επίσης,

$$y = a_1 * x + b_1 \quad (47)$$

$$a_1 = \frac{\sigma_1}{\sigma} \quad \text{και} \quad b_1 = \mu_1 - a * \mu = \mu_1 - r * \sigma_1 \quad (48)$$

όπου,

$$r = \frac{\mu}{\sigma} \quad \text{και} \quad a_1^2 = \operatorname{var} y / \operatorname{var} x \quad (49)$$

Δεχόμενοι την ύπαρξη συσχέτισης μεταξύ των δύο σειρών, και συμβολίζοντας με ρ τον συντελεστή συσχέτισης, ισχύει ότι:

$$y - \mu_1 = \rho_1 a_1 (x - \mu) + e \quad (50)$$

όπου e κάποια τυχαία μεταβλητή με μέσο όρο μηδέν και ανεξάρτητη από το χ.

Για τον υπολογισμό κάποιας τιμής της μεταβλητής y, χρησιμοποιώντας σαν σειρά αναφοράς τις τιμές της σειράς χ, γίνεται συνδυασμός των παραπάνω εξισώσεων. Ετσι:

$$t = (x - \mu) / \sigma \tag{50}$$

Αντικαθιστώντας την παραπάνω τιμή του t στην δεύτερη των εξισώσεων (46) επιτυγχάνεται ο υπολογισμός της τιμής της σειράς y.

(2) Μέθοδος της βέλτιστης πρόβλεψης

Η μέθοδος εφαρμόζεται σε χρονοσειρές με την παρακάτω μορφή:

(β.1) χ: είναι η σειρά δεδομένων με το μεγαλύτερο πλήθος στοιχείων (τις περισσότερες φορές προέρχεται από τον κύριο ή κεντρικό σταθμό καταγραφής).

y_1: είναι η σειρά δεδομένων με το μικρότερο πλήθος στοιχείων (προέρχεται από τους γειτονικούς ή τοπικούς σταθμούς).

Οι παράμετροι για την σειρά y_1 υπολογίζονται βάσει της σχέσης συσχέτισης με την σειρά χ ενώ οι ελλείπουσες τιμές της y_1 βάσει των αντίστοιχων τιμών της χ.

(β.2) Τρεις σειρές εμπλέκονται στους υπολογισμούς, η σειρά χ, η y_1 και η y_2 οι οποίες αντιπροσωπεύουν ένα παρελθόν και ένα παρόν δείγμα της Υ αντίστοιχα. Σ' αυτή την περίπτωση, οι παράμετροι των σειρων y_1 και y_2 υπολογίζονται βάσει της σχέσης συσχέτισης με την σειρά χ, ενώ οι ελλείπουσες τιμές της y_2 βάσει των αντίστοιχων τιμών της y_1.

Βασική προυπόθεση για την εφαρμογή αυτής της μεθόδου αποτελεί:

(ι) Οι τιμές $x_\iota, y_\iota, \ldots,\ \iota=1, n_1$ να είναι τυχαίες μεταβλητές (βλ. σχετικό κεφάλαιο για τους απαραίτητους ελέγχους).

(ιι) οι παράμετροι μ και σ (αποτελούν έκφραση της διασποράς και του μέσου όρου) για την σειρά χ να είναι γνωστές ή εύκολα υπολογίσιμες.

Σ' αυτή την περίπτωση απαιτείται ο υπολογισμός των παραμέτρων μ_1 και σ_1 για την μεταβλητή y.

Η μέθοδος της βέλτιστης πρόβλεψης έχει εφαρμογή για χρονοσειρές που πληρούν τις παρακάτω ιδιότητες, οι οποίες συναντώνται πολύ συχνά στα μετεωρολογικά δεδομένα.

(β.3) Η κατανομή των χ,y είναι κανονική.

(β.4) Οι παράμετροι μ και σ και μ_1 και σ_1 υπακούουν σε γραμμικούς εκτιμητές.

Για την περίπτωση (β.1), (β.3) ισχύει ότι:

Συμβολίζοντας με (y_1) την πρόβλεψη της ελλείπουσας τιμής από την χρονοσειρά y_1, και $d(y_1)$ το αντίστοιχο σφάλμα ισχύει:

$$\mathrm{var}(y_1) = \left(1-\rho_1^2\right)\mathrm{var}\hat{\mu}_1 + \left(1-\rho_1\right)^2 \mathrm{var}\hat{\mu}_1 + t^2\left[\left(1-\rho_1^4\right)\mathrm{var}\hat{\sigma}_1 + \left(1-\rho_1^2\right)^2 \mathrm{var}\hat{\sigma}_1\right] \tag{52}$$

Αντίστοιχα, στην περίπτωση (β.1), (β.4) ισχύει ότι:

$$\operatorname{var}(y_1) = \left(1-\rho_1^2\right)\operatorname{var} y_1 + \left(1-\rho_1\right)^2 \operatorname{var}\hat{y}_1 \tag{53}$$

Ο αναλυτικός τρόπος υπολογισμού των μεγεθών που εμφανίζονται στους παραπάνω τύπους αναπτύσσεται στο Παράρτημα Γ.

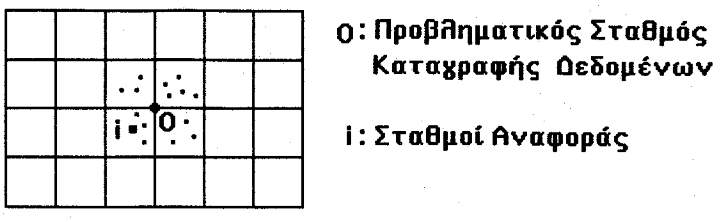
Για την περίπτωση (β.2), (β.3):

$$\begin{aligned}\operatorname{var}(y_2) &= \left\{\begin{array}{l} 1-\rho_2^2 - 2\dfrac{n'}{n_1}\left(\rho_{12} - \rho_1 * \rho_2\right) + \\ \dfrac{t^2}{2}\left[1-\rho_2^4 - 2\dfrac{n'}{n_1}\left(\rho_{12}^2 - \rho_1^2 * \rho_2^2\right)\right] \end{array}\right\}\operatorname{var}\hat{\mu}_2 \\ &+ \left[\left(1-\rho_1^2\right) + \frac{t^2}{2}\left(1-\rho_1^4\right)\right]\operatorname{var}\hat{\mu}_{2(1)} \\ &+ \left[\left(\rho_2-\rho_1\right)^2 + \frac{t^2}{2}\left(\rho_2^2-\rho_1^2\right)^2\right]\operatorname{var}\hat{\hat{\mu}}_2\end{aligned} \tag{54}$$

Τελειώνοντας για την περίπτωση (β.2), (β.4):

$$\begin{aligned}\operatorname{var}(y_2) &= \left\{1-\rho_2^2 - 2\frac{n'}{n_1}\left(\rho_{12} - \rho_1 * \rho_2\right) + \right\}\operatorname{var}\hat{y}_2 \\ &+ \left[\left(1-\rho_1^2\right)\right]\operatorname{var}\hat{y}_{2(1)} + \left(\rho_2-\rho_1\right)^2\operatorname{var}\hat{\hat{y}}_2\end{aligned} \tag{55}$$

(3) Μέθοδος της Βέλτιστης Παρεμβολής (optimal interpolation)[4]



Σχήμα 2. Γραφική αναπαράσταση ενός δικτύου γειτονικών σταθμών.

Οπως φαίνεται από το παραπάνω σχήμα, το ζητούμενο είναι να υπολογισθούν οι τιμές που θα κατεγράφοντο από το σταθμό O, με την βοήθεια των δεδομένων που καταγράφονται από τους γειτονικούς σταθμούς i.

Η μέθοδος βασίζεται στην παρεμβολή με την χρήση συντελεστών βάρους (weights). Εστω f_O και f_i είναι το σύμβολο των τιμών των χρονοσειρών για τους σταθμούς O και i αντίστοιχα. Επίσης έστω $\bar{f}_i$ και f'_i ο κλιματολογικός μέσος όρος (climatological average) και η απόκλιση (deviation). Ισχύει ότι

$$f_i = \bar{f}_i + f'_i \tag{56}$$

Το ζητούμενο είναι ο υπολογισμός του f'_o βάσει των f'_i από N σταθμούς, το οποίο εκφράζεται με την ακόλουθη σχέση:

$$f'_o = \sum_{i=1}^{N} w_i f'_i \tag{57}$$

Οπως είναι προφανές γνώση των w_i δίνει την επιθυμητή λύση στο πρόβλημα. Για τον προσδιορισμό τους εφαρμόζεται η μέθοδος της ελαχιστοποιήσεως του λάθους της παρεμβολής, το οποίο καταλήγει στο ακόλουθο σύστημα **κ** εξισώσεων:

$$\sum_{n=1}^{N} w_n \overline{f'_k f'_n} = \overline{f'_o f'_n} \tag{58}$$

όπου **κ** εκφράζει το πλήθος των w's που πρέπει να προσδιορισθούν αριθμητικά.

Ισχύει ότι:

[4]Ref [9]

$$\sum_{n=1}^{N} w_n \mu_{n,k} = \mu_{o,k} \tag{59}$$

όπου κ=1,2,...Ν. Για την επίλυση του παραπάνω συστήματοςπροαπαιτείται η γνώση των συντελεστών συσχέτισης $\mu_{\iota,\kappa}$ μεταξύ δύο των σταθμών ι,κ.

Συνοψίζοντας η παραπάνω μέθοδος περιλαμβάνει τα εξής στάδια:

(ι) Ξεκινώντας από τις τιμές των παρατηρήσεων f_i, αφαιρείται η τιμή του κλιματολογικού μέσου (climatological mean), πού αντιστοιχεί στον κάθε σταθμό, με αποτέλεσμα τον υπολογισμό των τιμών f'_i.

(ιι) Γνωρίζοντας τις τιμές $\mu_{o1}, \mu_{o2}, \ldots \mu_{oN}, \mu_{12} \ldots \mu_{1N}$ κ.λ.π, υπολογίζονται τα βάρη w_ι.

3.2 Συνθετικές Μέθοδοι συμπλήρωσης ελλειπόντων τμημάτων δεδομένων.

Στην περίπτωση που λείπει ολόκληρο τμήμα δεδομένων, συνίσταται η εφαρμογή κάποιας μεθόδου με αποτέλεσμα την παραγωγή μιας συνθετικής χρονοσειράς που θα συμπληρώνει το τμήμα των εγγραφών που λείπει.

Μια από τις χρησιμοποιούμενες μεθόδους είναι το πρώτης τάξεως autoregressive ή Markov μοντέλλο, το οποίο είναι μιας ειδικής μορφής Γραμμική Στοχαστική Διαδικασία.

3.2.1 Μεθοδος autoregressive χρονοσειρών 1ης ταξεως (Σειρές Markov). Στο Παράρτημα Δ αναφέρονται συνοπτικά μερικά **βασικά στοιχεία** της θεωρίας γραμμικών στοχαστικών διαδικασιών. Μια 1ου βαθμού autoregressive διαδικασία εκφράζεται ως:

$$\tilde{z}_t = \phi_1 * \tilde{z}_{t-1} + a_t = a_t + \phi_1 * a_{t-1} + \phi_1^2 * a_{t-2} + \ldots \tag{60}$$

Εφόσον $-1 < \phi_1 < 1$ ικανοποιείται η συνθήκη της στασιμότητας.

Η **συνάρτηση αυτοσυσχέτισης**: $\rho_\kappa = \phi_1^\kappa \quad \kappa \geq 0$ (61)

Η παραπάνω συνάρτηση φθίνει για ϕ_1 θετικό, ενώ φθίνει εκθετικά προς το μηδέν και ταλαντούται στην αντίθετη περίπτωση. Ισχύει ότι $\rho_1 = \phi_1$.

Η **διασπορά** δίνεται από:

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \phi_1^2} \tag{62}$$

και το **φάσμα** από:

$$p(f) = \frac{2\sigma_a^2}{1 + \phi_1^2 - 2\phi_1 \cos 2\pi f} \qquad 0 \leq f \leq 1/2 \tag{63}$$

Η μέθοδος της Μεγιστης Γειτνίασης υπαγορεύει ότι η συνάρτηση πυκνότητας πιθανότητας για μια σειρά $w_1, w_2, .., w_n$, που μπορεί να περιγραφεί απο κάποια autoregressive διαδικασία 1ης τάξεως, δίνεται από:

$$p(w_n / \phi \sigma_a) = (2\pi\sigma_a^2)^{-n/2} (1-\phi^2)^{1/2} \exp\left[-1/2\sigma_a^2 \left\{ (1-\phi^2) w_1^2 + \sum_{t=2}^{n} (w_t - \phi w_{t-1})^2 \right\} \right] \quad (64)$$

όπου υπάρχουν διάφοροι τρόποι για τον υπολογισμό της παραμέτρου ϕ.

$$\phi = \frac{\sum_{t=2}^{n} w_t w_{t-1}}{\sum_{t=2}^{n-1} w_t^2} \text{ ή } \phi = \frac{\sum_{t=2}^{n} w_t w_{t-1} / (n-1)}{\sum_{t=2}^{n-1} w_t^2 / (n-1)} \text{ ή}$$

$$\phi = \frac{\sum_{t=2}^{n} w_t w_{t-1}}{\sum_{t=1}^{n} w_t^2} \quad (65)$$

3.4 Επίλογος

Η εφαρμογή, από τους ενδιαφερόμενους χρήστες, των μεθόδων που αναφέρθηκαν στα προηγούμενα κεφάλαια, θα γίνει δυνατή με την βοήθεια του κατάλληλου λογισμικού που θα αναπτυχθεί απο το εργαστήριο Μετεωρολογίας στα πλαίσια του προγράμματος Υδροσκόπιο. Σύμφωνα με τις προτεινόμενες προδιαγραφές το λογισμικό θα παρέχει στον χρήστη, την δυνατότητα εύκολης επιλογής της κατάλληλης μεθόδου, ανάλογα με το είδος των προς επεξεργασία τιμών.

ΠΑΡΑΡΤΗΜΑ Α

Ο μέσος όρος της στοχαστικής διαδικασίας μπορεί να περιγραφεί με βάση τις τιμές της χρονοσειράς $z_1, z_2, \ldots, z_N$. Υπολογίζεται δε ως εξής:

$$\bar{z} = \frac{1}{N}\sum_{t=1}^{N} z_t \tag{1}$$

Αντιστοίχα η **διασπορά** της στοχαστικής διαδικασίας υπολογίζεται με βάση τις τιμές της χρονοσειράς ως ακολούθως:

$$\hat{\sigma}_z^2 = \frac{1}{N}\sum_{t=1}^{N} (z_t - \bar{z})^2 \tag{2}$$

Η ιδιότητα της στασιμότητας εξασφαλίζει οτι η κοινή πυκνότητα πιθανότητας (joint probability distibution), $p(z_{t_1}, z_{t_2})$ παραμένει αμετάβλητη για όλες τις τιμές t_1, t_2, και η οποία μπορεί να μορφοποιηθεί φτιάχνοντας ένα διάγραμμα χρησιμοποιώντας ζεύγη τιμών (z_t, z_{t+k}) της χρονοσειράς που απέχουν διάστημα κ (lag k).

Η **συμμεταβολή** (autocovariance) μεταξύ τιμών που απέχουν μεταξύ τους κ χρονικές τιμές ορίζεται ως ακολούθως:

$$\gamma_k = \mathrm{cov}\left[z_t, z_{t+k}\right] = E\left[(z_t - \mu)(z_{t+k} - \mu)\right] \tag{3}$$

Ομοια η **αυτοσθσχέτιση** (autocorrelation) για διάστημα μετατόπισης k οριζεται:

$$\rho_k = \frac{E\left[(z_t - \mu)(z_{t+k} - \mu)\right]}{\sqrt{E\left[(z_t - \mu)^2\right]E\left[(z_{t+k} - \mu)^2\right]}} = \frac{E\left[(z_t - \mu)(z_{t+k} - \mu)\right]}{\sigma_z^2} \tag{4}$$

Ο παρονομαστής απλοποιείται λόγω του ότι η στασιμότητα προυποθέτει ότι η τιμή της διαπoράς είναι η ίδια για χρονικές τιμές που διαφέρουν κ χρονικές μοναδές. Μια ακόμα πιο απλοποιημένη μορφή της (4):

$$\rho_k = \frac{\gamma_k}{\gamma_o} \tag{5}$$

Είναι προφανές ότι $\rho_0 = 1$.

Χρησιμοποιώντας τους παραπάνω ορισμούς για κάθε στάσιμη χρονοσειρά σχηματίζεται ο πίνακας συμμεταβολής (covariance) ο οποίος παρουσιάζει την ακόλουθη μορφή:

$$\Gamma_N = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & \gamma_{N-2} \\ & & \\ \gamma_{N-1} & \gamma_{N-2} & \gamma_0 \end{bmatrix} \qquad (6)$$

$$\Gamma_N = \sigma_Z^2 \begin{bmatrix} 1 & \rho_1 & \rho_{N-1} \\ \rho_1 & 1 & \\ & & \\ \rho_{N-1} & \rho_{N-2} & 1 \end{bmatrix} = \sigma_Z^2 P_N \qquad (7)$$

Ο παραπάνω πίνακας Γ_N ονομάζεται **πίνακας συμμεταβολής** (autocovariance matrix), ο δε P_N **πίνακας αυτοσθσχέτισης** (autocorrelation matrix). Η **συνθήκη στασιμότητας** υπαγορεύει ότι οι δύο αυτοί πίνακες είναι **απόλυτα θετικοί (definite positive)**. Για να συμβαίνει αυτό πρέπει η κύρια ορίζουσα και οι κύριες υποορίζουσες(principal minors) να είναι θετικές. Για παράδειγμα στην περίπτωση N=2, δηλαδή πρόκειται για την χρονοσειρά z_1,z_2 πρέπει να ισχύει ότι: $\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} \rangle 0$

Η ερμηνεία των ανωτέρω αναφερομένων συντελεστών διευκολύνεται με την χρήση των συναρτήσεων αυτοσθσχέτισης και συμμεταβολή (autocovariance) που βασίζεται στους συντελεστές αυτοσθσχέτισης και συμμεταβολή (autocovariance). Πιο συγκεκριμένα, η γραφική παράσταση γ_K ως προς κ ονομάζεται **συνάρτηση συμμεταβολής**. Ομοια η γραφική παράσταση ρ_K ως προς κ ονομάζεται **συνάρτηση αυτοσθσχέτισης**. Η συνάρτηση αυτοσυσχέτισης είναι συμμετρική ως προς το μηδέν, οπότε αρκεί η γραφική αναπαράσταση του τμήματος που βρίσκεται στο θετικό μέρος του συστήματος συντεταγμένων. Ισχύει η ακόλουθη σημαντική σχέση για τα ανωτέρω μεγέθη:

$$\gamma_\kappa = \rho_\kappa \sigma_Z^2 \qquad (8).$$

Μια πολύ σημαντική ιδιότητα των στασίμων στοχαστικών διαδικασιών είναι οτι κάθε **γραμμικός συνδυασμός** που προκύπτει είναι επίσης στάσιμος.

Μια κατηγορία στοχαστικών διαδικασιών που χρησιμοποιείται ευρέως για την αναπαράσταση μετεωρολογικών δεδομένων,στην περίπτωση μέσων ετησίων, εποχιακών ακόμα και μηνιαίων παρατηρήσεων [Tech. Note 79 ,WMO-No. 195], είναι οι Gaussian (Normal). Ειναι γνωστό ότι για την περιγραφή μιας Κανονικής (Normal) κατανομής αρκεί η γνώση των ροπών (moments) πρώτης και δεύτερης τάξης. Εξ αιτίας αυτού,η ύπαρξη ενός σταθερού μέσου όρου και γνώση του πίνακα συμμεταβολής (autocovariance matrix) Γ_N για όλα τα N αρκεί για να βεβαιώσει την ιδιότητα της στασιμότητας μιας κανονικής διεργασίας.

Στην πρακτική εφαρμογή των ανωτέρω εννοιών εμπλέκονται διάφοροι περιορισμοί που συνδέονται άμμεσα με τον πεπερασμένο αριθμό των στοιχείων της προς επεξεργασία χρονοσειράς. Σαν αποτέλεσμα, οι **υπολογισμοί** καταλήγουν να είναι **εκτιμήσεις** των ανωτέρω μεγεθών. Οπως είναι φυσικό, η αξιοπιστία της εκτίμησης εκφράζεται με τον υπολογισμό του αντίστοιχου σφάλματος, το οποίο χρησιμοποιείται σαν εργαλείο **αποδοχής** ή **απόρριψης** της εκτίμησης.

ΠΑΡΑΡΤΗΜΑ Β

Στην συνέχεια δίνονται οι απαραίτητοι ορισμοί των χρησιμοποιουμένων εννοιών.

(1) Φάσμα (Sample Spectrum):

Ορίζεται ως ακολούθως

$$I(f) = \frac{2}{N}\left(a_f^2 + b_f^2\right) \qquad \text{οπου } 0<=f<=0.5 \qquad (1)$$

και συνδέεται με την συνάρτηση autocovariance ως εξής:

$$I(f) = 2\left\{c_0 + 2\sum_{k=1}^{N-1} c_k \cos 2\pi fk\right\} \qquad \text{οπου } 0<=f<=0.5 \qquad (2)$$

Οι αδυναμία αυτού του είδους του φάσματος να περιγράψει στάσιμες διαδικασίες στις οποίες οι αλλαγές στην συχνότητα,το πλάτος και την φάση είναι τυχαίες, ο ανωτέρω ορισμός επεκτείνεται στην έννοια του φάσματος ισχύος(power spectrum).

(2) Φάσμα Ισχύος (Power Spectrum)

Ορίζεται ως εξής:

$$p(f) = \lim_{N\to\infty} E[I(f)] = 2\left\{\gamma_0 + 2\sum_{k=1}^{N-1} \gamma_k \cos 2\pi fk\right\} \qquad 0<=f<=0.5 \qquad (3)$$

Το φάσμα ισχύος δείχνει την κατανομή της διασποράς της στοχαστικής διαδικασίας σε κάποιο συνεχές πεδίο συχνοτήτων.

(3) Συνάρτηση Φάσματος Ισχύος (Spectrum Density Function)

Υπάρχει η ανάλογη εξάρτηση του φάσματος με την συνάρτηση αυτοσυσχέτισης, η οποία εκφράζεται ως ακολούθως:

$$g(f) = \frac{p(f)}{\sigma_Z^2} = 2\left\{1 + 2\sum_{k=1}^{\infty} \rho_k \cos 2\pi fk\right\} \qquad \text{όπου } 0<=f<=0.5 \qquad (4)$$

Η παραπάνω συνάρτηση έχει τις ιδιότητες της συνάρτησης πυκνότητας πιθανότητας, το οποίο αποτελεί σημαντικό στοιχείο για την διαδικασία της στατιστικής ανάλυσης στοχαστικών διαδικασιών.

Στην πράξη χρησιμοποιείται η παρακάτω μορφή του φασματος ισχύος, η οποία λύνει κάποια προβλήματα διακυμάνσεων των τιμών του φάσματος, τα οποία οφείλονται στο περιοσμένο πλάτος του φάσματος των συχνοτήτων.

$$\hat{p}(f) = 2\left\{c_0 + 2\sum_{k=1}^{N-1} \lambda_k c_k \cos 2\pi f k\right\} \tag{5}$$

οπου λ_K ονομαζεται διάστημα μετατόπισης (lag window) και ανάλογα με τις τιμές που παίρνει αυξάνει το πλάτος του πεδίου των συχνοτήτων με αποτέλεσμα την εξομάλυνση της μορφής του εκτιμητέου φάσματος.

ΠΑΡΑΡΤΗΜΑ Γ

(1) Περίπτωση κατά την οποία οι χρονοσειρές X,Y υπακούουν σε κανονική κατανομή δύο μεταβλητών.

Οπως έχει ήδη αναφερθεί, το ζητούμενο είναι η πρόβλεψη των τιμών των μ_1 και σ_1 (παράμετροι που περιγράφουν την σειρά Y) και ρ_1 (συντελεστή συσχέτησης των σειρών X,Y), βασιζόμενη στην γνώση των παραμέτρων μ και σ (παράμετροι που περιγράφουν την σειρά X).

Ο εκτιμητής του μ_1, συμβολίζεται με $\mu_1{}^*$ και δίνεται από τον παρακάτω τύπο:

$$\mu_1^* = \hat{\mu}_1 - \hat{\rho}_1\left(\frac{\hat{\sigma}_1}{\hat{\sigma}}\right)(\hat{\mu}-\mu) \qquad (1)$$

όπου,

$$\hat{\mu} = \sum x_i / n_1 \qquad (2)$$

$$\hat{\mu}_1 = \sum y_i / n_1 \qquad (3)$$

$$\hat{\sigma}^2 = \sum (x_i - \hat{\mu})^2 / n_1 \qquad (4)$$

$$\hat{\sigma}_1^2 = \sum (y_i - \hat{\mu}_1)^2 / n_1 \qquad (5)$$

$$\hat{\rho}_1^2 = \sum (x_i - \hat{\mu})(y_i - \hat{\mu}_1) / (\hat{\sigma} * \hat{\sigma}_1) \qquad (6)$$

Ομοιως οι εκτιμητές των ζητουμένων τιμών των παραμέτρων σ_1 και ρ_1 δίνονται από τους παρακάτω τύπους.

$$\sigma_1^* = \hat{\sigma}_1 (\sigma / \hat{\sigma})^{\hat{\rho}_1^2} \qquad (7)$$

$$\rho_1^* = \hat{\rho}_1 (\sigma / \hat{\sigma})^{(1-\hat{\rho}_1^2)} \qquad (8)$$

Οι διασπορές και συμμεταβολές (variances and covariances) δίνονται από τους παρακάτω τύπους.

$$\text{var}\,\mu_1^* = \left(1 - \hat{\rho}_1^2\right) \text{var}\,\hat{\mu}_1 \qquad (9)$$

$$\text{var}\,\sigma_1^* = \left(1 - \hat{\rho}_1^4\right) \text{var}\,\hat{\sigma}_1 \qquad (10)$$

$$\operatorname{var}\rho_1^* = \frac{2-\hat{\rho}_1^2}{2}\operatorname{var}\hat{\rho}_1 \tag{11}$$

$$\operatorname{cov}\left(\mu_1^*, \sigma_1^*\right) = 0 \tag{12}$$

$$\operatorname{cov}\left(\mu_1^*, \rho_1^*\right) = 0 \tag{13}$$

$$\operatorname{cov}\left(\sigma_1^*, \rho_1^*\right) = \left(1-\rho_1^2\right)\operatorname{cov}\left(\hat{\sigma}_1, \hat{\rho}_1\right) \tag{14}$$

Στην επίλυση πραγματικών προβλημάτων, οι τιμές των μ και σ δεν είναι δεδομένες αλλά πρέπει να εκτιμηθούν βάσει των τιμών της σειράς Χ, που χρησιμοποιείται σαν σειρά αναφοράς. Συμβολίζουμε με $\hat{\mu}$ και $\hat{\sigma}$ τις εκτιμήσεις των ανωτέρω παραμέτρων.

Θέλοντας να υπολογίσουμε και το λάθος της εκτίμησης καταλήγουμε στην χρήση των παρακάτω τύπων:

$$\begin{aligned} d\mu_1^* &= d\hat{\mu}_1' + \rho_1 a_1 d\hat{\mu} \\ d\sigma_1^* &= d\hat{\sigma}_1' + \rho_1^2 a_1 d\hat{\sigma} \end{aligned} \tag{15}$$

Οι $\hat{\mu}_1'$ και $\hat{\sigma}_1'$ είναι τυχαίες μεταβλητές ανεξάρτητες των $\hat{\mu}$ και $\hat{\sigma}$ των οποίων οι διασπορές και συμμεταβολές υπακούουν τις εξισώσεις (9)-(14).

(2) Περίπτωση κατά την οποία οι εκτιμητές των παραμέτρων μ και σ είναι γραμμικοί.

Υποθέτοντας ότι x_j και y_j $j=1,\ldots,n_1$ είναι οι τιμές των χρονοσειρών Χ και Υ σε αύξουσα σειρά, οι εκτιμητές των παραμέτρων μ και σ δίνονται από τους ακόλουθους τύπους:

$$\hat{\sigma} = \sum \lambda_j x_j, \quad \hat{\sigma}_1 = \sum \lambda_j y_j, \quad \hat{\mu} = \sum v_j x_j, \quad \hat{\mu} = \sum v_j y_j \tag{16}$$

Στις παραπάνω σχέσεις οι τιμές των συντελεστών λ_j και v_j είναι πλήρως καθορισμένες για κάθε n_1.

Στην περίπτωση που τα μ και σ για την χρονοσειρά Χ είναι ήδη γνωστά και ο εκτιμητής του συντελεστή συσχέτησης υπολογίζεται βάσει της σχέσης [Παράρτημα Γ, (6)], τότε οι εκτιμητές των μ_1, σ_1 ορίζονται με τον ακόλουθο πιο ακριβή τρόπο:

$$\sigma_1^* = \hat{\sigma}_1(\sigma/\hat{\sigma})^{\hat{\rho}_1}, \quad \mu_1^* = \hat{\mu}_1 - \hat{\rho}_1 a_1^*(\hat{\mu}-\mu), \quad a_1^* = \sigma_1^*/\sigma \tag{17}$$

Τα στατιστικά χαρακτηριστικά των εκτιμητών δίνονται παρακάτω:

$$\operatorname{var}\sigma_1^* = \left(1-\rho_1^2\right)\operatorname{var}\hat{\sigma}_1, \quad \operatorname{var}\mu_1^* = \left(1-\rho_1^2\right)\operatorname{var}\hat{\mu}_1,$$

$$\operatorname{cov}\left(\sigma_1^*,\mu_1^*\right) = \left(1-\rho_1^2\right)\operatorname{cov}\left(\hat{\sigma}_1,\hat{\mu}_1\right) \tag{18}$$

Εξ άλλου, αν οι τιμές σ και μ αποτελούν εκτιμήσεις των $\hat{\sigma}$ και $\hat{\mu}$ για n, αντί n_1, παρατηρήσεις της σειράς X , οι σχέσεις (17) εξακολουθούν να ισχύουν. Σ' αυτή την περίπτωση τα λάθος της εκτίμησης εκφράζεται ως εξής:

$$d\sigma_1^* = d\hat{\sigma}_1' + \rho_1 a_1 d\hat{\sigma}, \quad d\mu_1^* = d\hat{\mu}_1' + \rho_1 a_1 d\hat{\mu} \tag{19}$$

Ισχύει ότι οι $\hat{\sigma}_1'$ και $\hat{\mu}_1'$ είναι τυχαίες μεταβλητές των οποίων οι ροπές δίνονται από τις σχέσεις (18). Η μορφή των ζητουμένων εκτιμητών μετατρέπεται σε:

$$\begin{aligned}
\operatorname{var}\sigma_1^* &= \left(1-\rho_1^2\right)\operatorname{var}\hat{\sigma}_1 + \rho_1^2 a_1^2 \operatorname{var}\hat{\sigma} \\
\operatorname{var}\mu_1^* &= \left(1-\rho_1^2\right)\operatorname{var}\hat{\mu}_1 + \rho_1^2 a_1^2 \operatorname{var}\hat{\mu} \\
\operatorname{cov}\left(\sigma_1^*,\mu_1^*\right) &= \left(1-\rho_1^2\right)\operatorname{cov}\left(\hat{\sigma}_1,\hat{\mu}_1\right) + \rho_1^2 a_1^2 \operatorname{cov}\left(\hat{\sigma},\hat{\mu}\right)
\end{aligned} \tag{20}$$

Μια άλλη μορφή των εκτιμητών, οταν $\sigma=\sigma_1$, είναι η ακόλουθη:

$$\begin{aligned}
\operatorname{var}\sigma_1^* &= \left(1-\rho_1^2\right)\operatorname{var}\hat{\sigma}_1 + \rho_1^2 \operatorname{var}\hat{\sigma} \\
\operatorname{var}\mu_1^* &= \left(1-\rho_1^2\right)\operatorname{var}\hat{\mu}_1 + \rho_1^2 \operatorname{var}\hat{\mu} \\
\operatorname{cov}\left(\sigma_1^*,\mu_1^*\right) &= \left(1-\rho_1^2\right)\operatorname{cov}\left(\hat{\sigma}_1,\hat{\mu}_1\right) + \rho_1^2 \operatorname{cov}\left(\hat{\sigma},\hat{\mu}\right)
\end{aligned} \tag{21}$$

Εξ άλλου θέτοντας,

$$y^* = \sigma_1^* t + \mu_1^*, \quad \hat{y} = \hat{\sigma}_1 t + \hat{\mu}_1, \quad \hat{y} = \hat{\sigma}_1 t + \hat{\mu}_1 \tag{22}$$

χρησιμοποιώντας τις εξισώσεις (21)

$$\text{var}\,y^{*} = \left(1-\rho_1^2\right)\text{var}\,\hat{y} + \rho_1^2\,\text{var}\,\hat{y}_1 \tag{23}$$

Εξ αιτίας του ότι οι διασπορές και συνδιασπορές των γραμμικών εκτιμητών είναι αντιστρόφως ανάλογες του αριθμού των τιμών των σειρών, ισχύει:

$$n_1\,\text{var}\,\hat{y} = n\text{var}\,\hat{y} \tag{24}$$

Η σχέση (23) δίνει,

$$\text{var}\,y^{*} = \left[1-\rho_1^2\left(1-\frac{n_1}{n}\right)\right]\text{var}\,\hat{y} \tag{26}$$

Καταλήγοντας, η ακρίβεια υπολογισμού του fractile όπως υπολογίζεται από την σχέση (20), είναι ισοδύναμη αν για τον υπολογισμό χρησιμοποιηθεί κάποια σειρά παρατηρήσεων της Y, με μήκος $\kappa^{*}n_1$, όπου κ δίνεται:

$$\kappa = \frac{1}{\left[1-\rho_1^2\left(1-\frac{n_1}{n}\right)\right]} \tag{27}$$

ΠΑΡΑΡΤΗΜΑ Ε

Στον παρακάτω πίνακα έχει γίνει μια προσπάθεια τυποποίησης του τρόπου χρήσης των διαφόρων μεθόδων συμπλήρωσης. Οπου δεν αναφέρεται συγκεκριμένο είδος μετεωρολογικών καταγραφών ενοείται ότι δεν υπάρχει περιορισμός στην χρήση. Οι στατιστικές μέθοδοι, δεν αναγράφονται στον πίνακα λόγω της εφαρμογης τους ανεξάρτητα του είδους των καταγραφών.

ΠΙΝΑΚΑΣ 12

Είδος μετεωρολογικών καταγραφών - Μέθοδοι επεξεργασίας

Είδος μετεωρολογικών καταγραφών		**Επίπεδο καταχώρησης**	**Μέθοδος επεξεργασίας**
	Γ	Πρωτογενείς τιμές (single observations)	Διάγραμμα Διασποράς τιμών (Scatter Diagram)
	Α	Πρωτογενείς τιμές	Διάγραμμα Residuals
	Α	Πρωτογενείς τιμές	Run Tests
Εφαρμόζεται κυρίως σε χρονοσειρές Βροχής, Ταχύτητας ανέμου, ημερήσιας ηλιοφάνειας	Β	Μηνιαία, Ετήσια Αθροίσματα και μέσες τιμές	Μέθοδος των Λόγων
Εφαρμόζεται κυρίως σε χρονοσειρές θερμοκρασιών	Β	Μηνιαία, Ετήσια Αθροίσματα και Μέσες τιμές	Μέθοδος των Διαφορών
	Α,Β	Μέσες τιμές	Υπολογισμός της μέσης τιμής

ΠΙΝΑΚΑΣ 12 (συνέχεια)

	Γ	Ωριαίες ή ημερήσιες τιμές	Γραφήματα συσχέτησης σειρών σε μετατόπιση ή μη (Γραμμική Παρεμβολή, Cubic Spline κλπ)

όπου A: Δεδομένα προερχόμενα από έναν σταθμό

B: Δεδομένα προερχόμενα από γειτονικούς σταθμούς

Γ: (A OR B)

ΣΥΜΒΟΛΙΣΜΟΙ

(α) Ελληνικά Σύμβολα

α	Πιθανότητα απόρριψης της μηδενικής απόρριψης
ρ	Συντελεστής συσχέτησης

(β) Λατινικά Σύμβολα

X,Y	Τυχαίες μεταβλητές
R	Μορφή κάποιου εκτιμητή
N, n, n_1	Πλήθος στοιχείων της χρονοσειράς
V	Von Neumann Ratio
E(V)	Μέση τιμή μιας κατανομής V
var(V)	Διασπορά μιας κατανομής V
r	Συντελεστής συσχέτησης
c	Συντελεστής συμμεταβολής
a	Συντελεστής σειράς Fourier

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

[1] World meteorological Organization, July 1984, Second Edition, Supplement No 1, *Guide to Climatological Practices*

[2] Γκίνη Μ.,ΥΒΕΤ-Διεύθυνση Υδατικού Δυναμικού και Φυσικών Πόρων, Υδροσκόπιο, Αριθμός Τεύχους 6/2, 6/3, Νοέμβριος 1992

[3] Sneyers R., 1975, *Sur l' analyse statistique des series d' observations*, Organisation Meteorologique Mondiale,Tech. Note 143-WMO 415, Geneve.

[4] World Meteorological Organization, *Climatic Change*, Tech. Note 79-WMO 195, Geneva.

[5] World Meteorological Organization, *Some methods of climatological analysis*,Tech. Note 81 - WMO 199, Geneva

[6] A note on climatological normals , World Meteorological Organization, Tech. Note 84-WMO 208, Geneva

[7] G.Box and G.M. Jenkins,1976, *Time series analysis, forecasting and control,* by HOLDEN-DAY, INC.

[8] Brooks C.E.P., Carruthers N., *Handbook of statistical methods in meteorology* ,H.M. Stationary Office, London U.K.

[9] Morel, P., *Space and time meteorological data analysis and initialization in Dynamic Meteorology.*