Τμήμα Πολιτικών Μηχανικών
Πανεπιστήμιο Θεσσαλίας

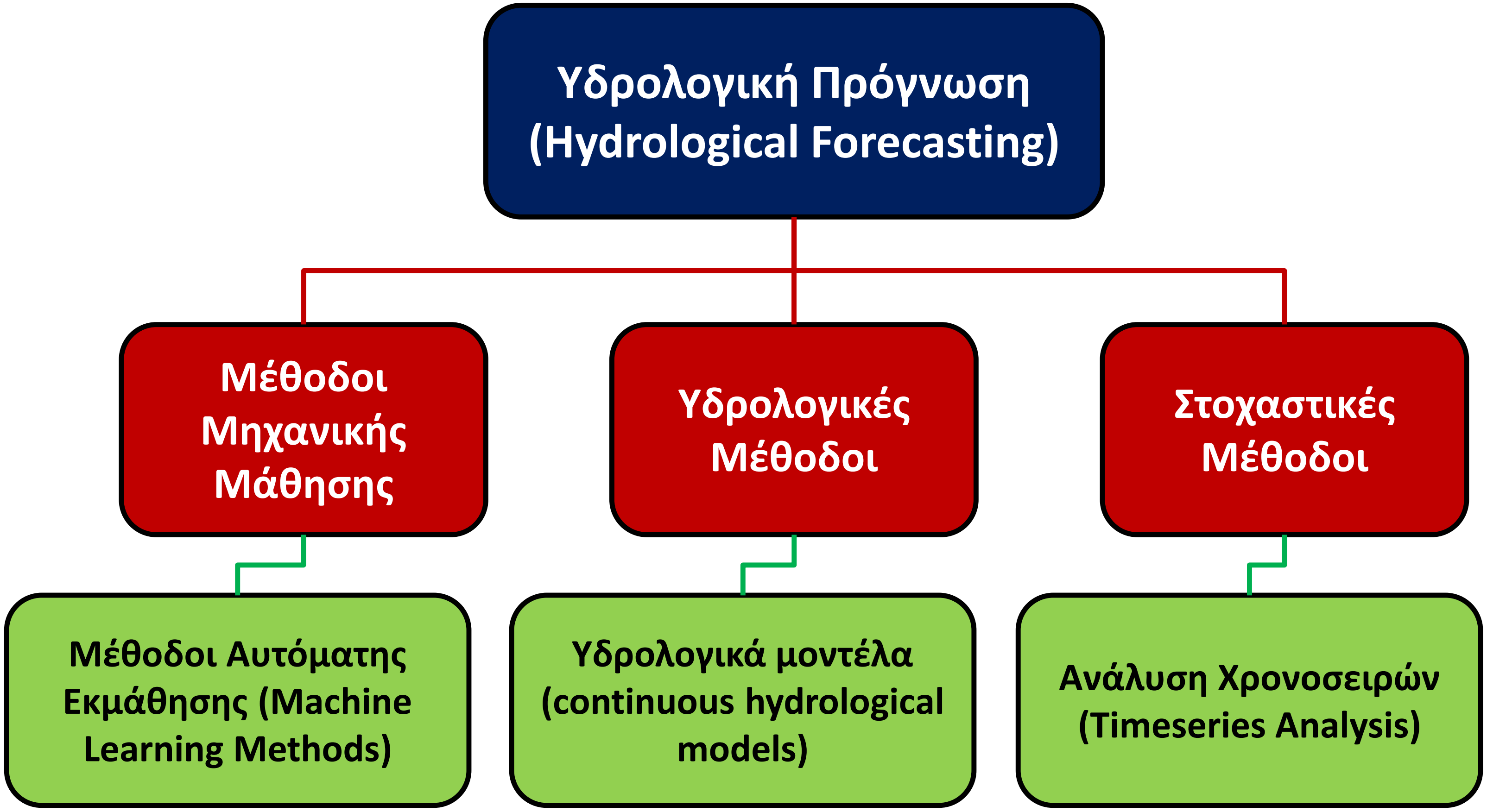
ΥΔΡΟΛΟΓΙΚΗ ΠΡΟΣΟΜΟΙΩΣΗ ΚΑΙ ΠΡΟΓΝΩΣΗ

Ενότητα 3: Υδρολογική πρόγνωση

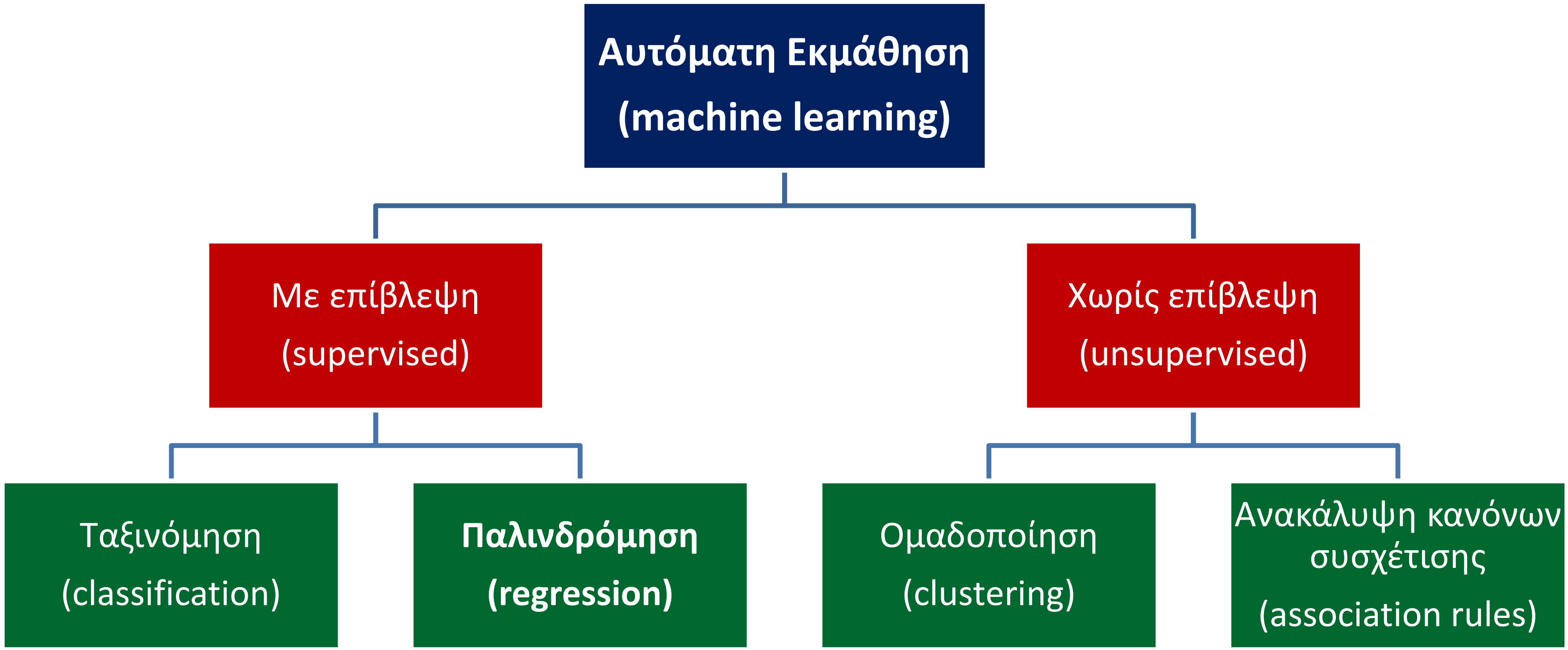
3.1. Ανάλυση Χρονοσειρών

Καθ. Αθανάσιος Λουκάς
Εργαστήριο Υδρολογίας και Ανάλυσης Υδατικών Συστημάτων
Τμήμα Πολιτικών Μηχανικών
Πολυτεχνική Σχολή

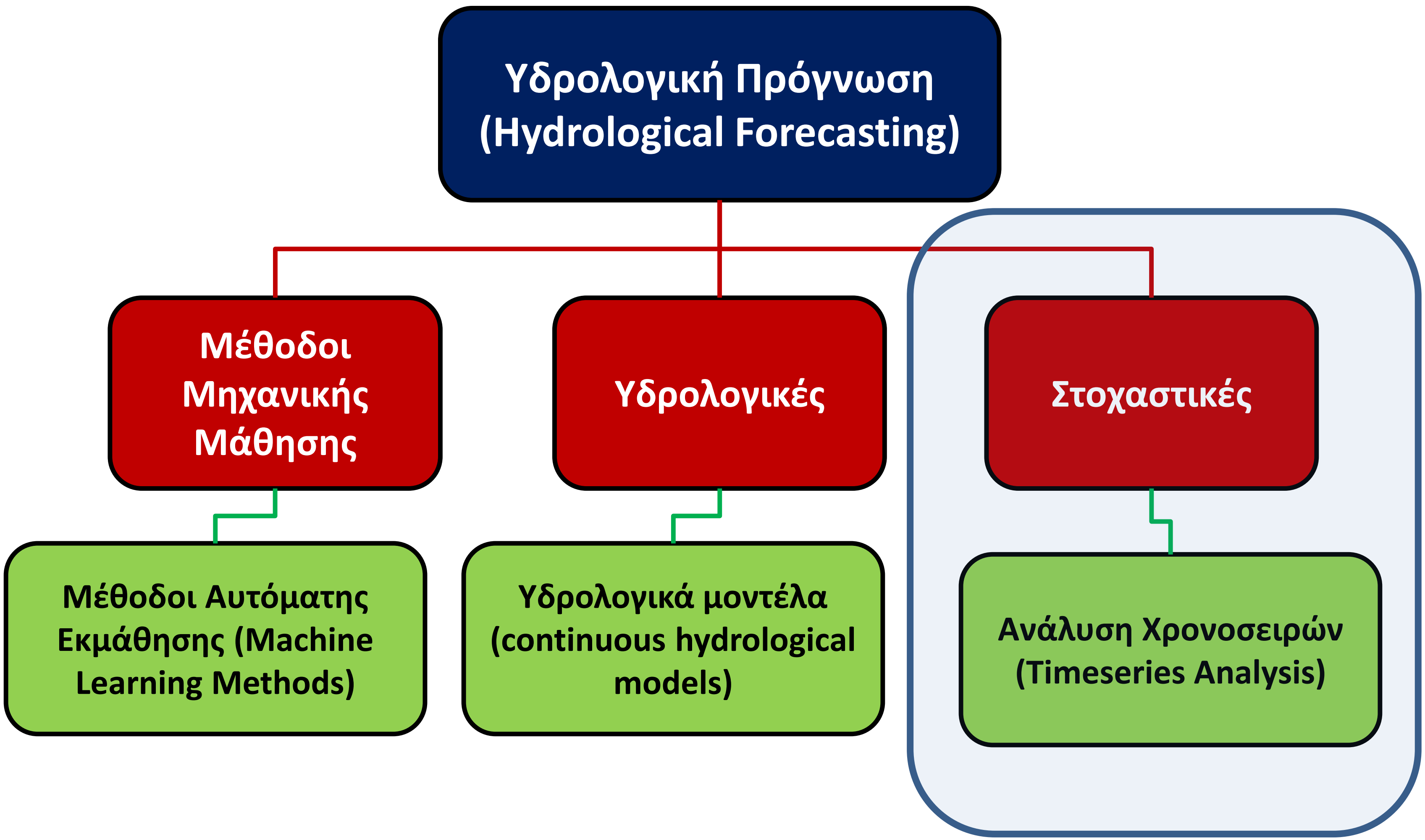
ΜΕΘΟΔΟΙ ΥΔΡΟΛΟΓΙΚΗΣ ΠΡΟΓΝΩΣΗΣ



ΚΑΤΗΓΟΡΙΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



ΜΕΘΟΔΟΙ ΥΔΡΟΛΟΓΙΚΗΣ ΠΡΟΓΝΩΣΗΣ



ΑΝΑΛΥΣΗ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ (Time Series Analysis)

- ***Χρήση:***
  - Δημιουργία συνθετικών χρονοσειρών
  - Πρόγνωση (forecasting) υδρολογικών γεγονότων
  - Εύρεση τάσεων και αλλαγών σε υδρολογικές χρονοσειρές
  - Συμπλήρωση κενών στις χρονοσειρές και επέκταση χρονοσειρών

- ***Περιεχόμενο ενότητας:***
  - Στοχαστική δομή υδρολογικών χρονοσειρών
  - Αρχές ανάλυσης χρονοσειρών
  - Στοχαστικά μοντέλα για απλές και πολλαπλές χρονοσειρές
  - Ανελίξεις (Αλυσίδες) Markov
  - Μέθοδοι συμπλήρωσης κενών και επέκταση χρονοσειρών
  - Μέθοδοι δημιουργίας δεδομένων στοχαστικής προσομοίωσης

ΑΝΑΛΥΣΗ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ
Time Series Analysis

Μελέτη των δεδομένων
(Γραφική παράσταση, οπτικός έλεγχος)

↓

Αναγνώριση μοντέλου
(Αναγνώριση της τάξης του μοντέλου, αναζήτηση μοντέλων αποδεκτών)

↓

Εκτίμηση παραμέτρων του μοντέλου

↓

Έλεγχος του μοντέλου
(Από όλα τα αποδεκτά επιλέγεται το "καλύτερο")

↓

Χρησιμοποίηση του μοντέλου για πρόβλεψη

ΑΝΑΛΥΣΗ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

- ***Υδρολογική χρονοσειρά***: οι διατεταγμένες σε αυστηρή χρονική ακολουθία πραγματοποιήσεις (μετρήσεις) μιας υδρολογικής μεταβλητής
- Οι **χρονοσειρές** διακρίνονται σε:
  - Διακριτές ή διακεκριμένες (discrete)
    - π.χ. ημερήσιες, μηνιαίες τιμές παροχής σε μία θέση μέτρησης είναι διακεκριμένες χρονοσειρές
  - Συνεχείς (continuous)
    - π.χ. συνεχείς τιμές της παροχής για μια σειρά ετών

- Τα χρονικά ισοδιαστήματα μεταξύ των τιμών μιας διακεκριμένης χρονοσειράς έχουν πολύ μεγάλη σημασία όταν πρόκειται για την κατασκευή μαθηματικού ομοιώματος της υδρολογικής μεταβλητής

ΑΝΑΛΥΣΗ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

- ***Υδρολογική διαδικασία (hydrologic process)***: χαρακτηρίζεται κάθε υδρολογικό φαινόμενο, που δείχνει μια συνεχή αλλαγή στο χρόνο, στο χώρο, στην επιφάνεια ή πάνω σε μια γραμμή
  - π.χ. η βροχή και η απορροή μεταβάλλονται χωρικά και χρονικά σε μια λεκάνη απορροής , όπου επίσης μεταβάλλονται η εξατμισοδιαπνοή κλπ.

- Οι υδρολογικές διαδικασίες και η λεκάνη απορροής αποτελούν στο σύνολο τους ένα δυναμικό σύστημα μεταβλητό στο χρόνο

ΚΑΤΗΓΟΡΙΕΣ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

1. **Απλές χρονοσειρές (single timeseries):** Χρονοσειρά υδρολογικής μεταβλητής σε μία θέση (ονομάζονται και μονομεταβλητές χ.σ. – univariate t.s.)
   - π.χ. Μέτρηση βροχόπτωσης σε 5 σταθμούς σε μία λεκάνη απορροής
     - Κάθε μία καταγραφή είναι απλή χ.σ.
2. **Πολλαπλές χρονοσειρές (multiple timeseries):** Σετ απλών χρονοσειρών (ονομάζονται και πολυμεταβλητές χ.σ. – multivariate t.s.)
   - Χρονοσειρές σε διαφορετικές θέσεις ή διαφορετικών παραμέτρων στην ίδια θέση
     - Σε μία θέση π.χ. μέτρηση
       - α) παροχής,
       - β) στάθμης,
       - γ) θερμοκρασία νερού και
       - δ) φερτές ύλες

ΚΑΤΗΓΟΡΙΕΣ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

3. **Χρονοσειρές μη συσχετισμένες και συσχετισμένες (uncorrelated – correlated)**
   - Αν σε μία απλή χρονοσειρά x(t) το x στο χρόνο t εξαρτάται γραμμικά από τις τιμές του x στο χρόνο t-k, η χρονοσειρά x(t) ονομάζεται αυτοσυσχετισμένη, σειριακά συσχετισμένη, ή συσχετισμένη στο χρόνο (autocorrelated, serially correlated, correlated in time)
   - Διαφορετικά η χρονοσειρά x(t) είναι μη συσχετισμένη ή ανεξάρτητη
   - Αυτοσυσχέτιση (autocorrelation) ή η εξάρτηση σε μια χ.σ. ροής ξεκινά από ότι η επιφάνεια, το έδαφος και ο υδροφόρος ορίζοντας κάνει το νερό να παραμένει στο σύστημα της λεκάνης απορροής για αρκετό διάστημα
     - π.χ. λεκάνη απορροής με σημαντική αποθήκευση. Λίμνες, έλη ή παγετώνες είναι σημαντικά αυτοσυσχετισμένες
     - Αντίθετα, μηνιαίες ή ετήσιες υετοπτώσεις ή χ.σ. μεγίστων ετήσιων παροχών δεν είναι σημαντικά αυτοσυσχετισμένες
   - Αν για δύο χ.σ. η μεταβλητή της δεύτερης είναι γραμμικά συσχετισμένη (εξαρτάται) από τις τιμές της πρώτης για το χρόνο t-k τότε οι χ.σ. είναι διασυσχετισμένες (cross correlated)
     - π.χ. χ.σ. μέσης επιφανειακής βροχόπτωσης και απορροής

ΚΑΤΗΓΟΡΙΕΣ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

4. **Διακοπτόμενες χρονοσειρές (intermittent timeseries):** Όταν η υδρολογική μεταβλητή παίρνει μηδενικές και μη μηδενικές τιμές
   - π.χ. βροχόπτωση αλλά για μικρό χρονικό βήμα (π.χ. ωριαίο)
     - Για μεγάλο χρονικό βήμα π.χ. εβδομαδιαίο, μηνιαίο → συνήθως μη διακοπτόμενο
   - Άλλο παράδειγμα: χ.σ. παροχής σε ξηρές ή ημίξηρες περιοχές
5. **Χρονοσειρές απαρίθμησης (counting timeseries):** Η μεταβλητή τους προκύπτει από την απαρίθμηση των συμβάντων
   - π.χ. ημέρες με βροχή σε ένα μήνα
6. **Χρονοσειρές με καθορισμένο σταθερό χρονικό βήμα ή μη σταθερό χρονικό βήμα (Regularly and Irregularly spaced timeseries):**
   - Συνήθως χ.σ. με σταθερό χρονικό βήμα
     - Μετρήσεις ποιότητας σε μη σταθερό χρονικό βήμα
     - Συνήθως χ.σ. με σταθερό χρονικό βήμα
   - Για την ανάλυση χρονοσειρών → Σταθερό χρονικό βήμα

ΚΑΤΗΓΟΡΙΕΣ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

7. **Χρονικά σταθερές και μη σταθερές χρονοσειρές (stationary and nonstationary timeseries):** Χρονικά σταθερή χρονοσειρά
   - Ελεύθερη από τάσεις, αλλαγές ή περιοδικότητα
   - Η μέση τιμή και η διακύμανση παραμένουν σταθερά με το χρόνο → stationary
   - Η μέση τιμή και η διακύμανση δεν παραμένουν σταθερά με το χρόνο → nonstationary
     - Γενικά ετήσιες χ.σ. υποτίθεται ότι είναι σταθερές
     - Για μικρότερο χρονικό βήμα από ένα έτος → συνήθως η χ.σ. είναι nonstationary

Η χρονολογική σειρά (time series)

Είναι ένα **δείγμα τιμών** **y_1, y_2, …….., y_t,** όπου ο δείκτης t = 1, 2, …..,T παριστάνει έτη, μήνες, μέρες κ.ο.κ. ή χρονικά διαστήματα (5 έτη, 3 μήνες , 2 εβδομάδες κ.ο.κ.).

- **Θεωρούμε** ότι οι παρατηρήσεις **y_1, y_2, …….., y_T** είναι **συγκεκριμένες τιμές ή συγκεκριμένες πραγματοποιήσεις των τυχαίων μεταβλητών**

  Y_1, Y_2, ….., Y_T

  που είναι μέρος **μιας άπειρης σειράς (ακολουθίας) τυχαίων μεταβλητών.**

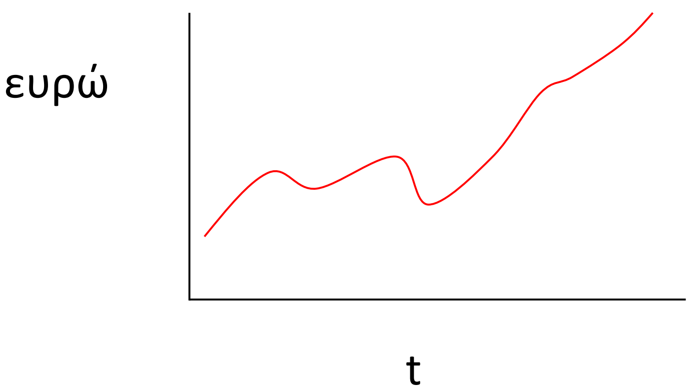
- Η άπειρη αυτή ακολουθία των τυχαίων μεταβλητών ονομάζεται **στοχαστική ή τυχαία διαδικασία (stochastic process) ή στοχαστική ανέλιξη** και συνήθως παριστάνεται ως **$\{Y_t\}$.**

ΠΑΡΑΔΕΙΓΜΑΤΑ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ:

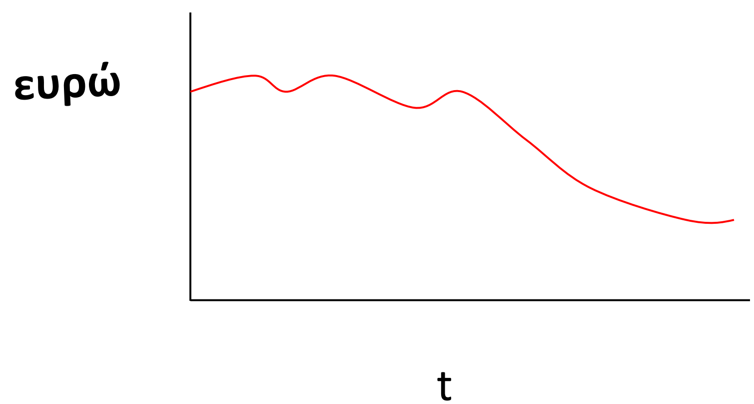
Εμπόριο (marketing)

- πωλήσεις ανά μήνα
- δαπάνες για διαφημίσεις

διαφημίσεις



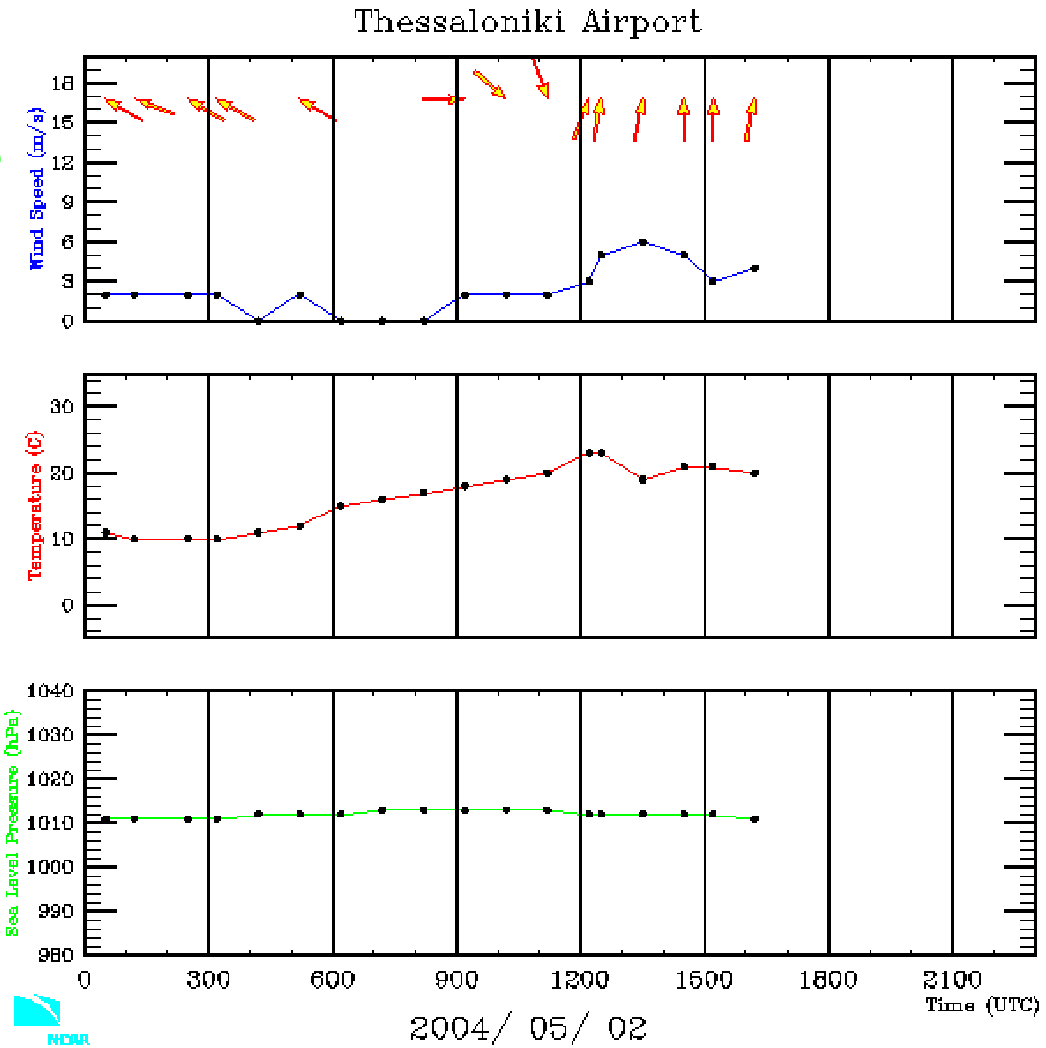
πωλήσεις



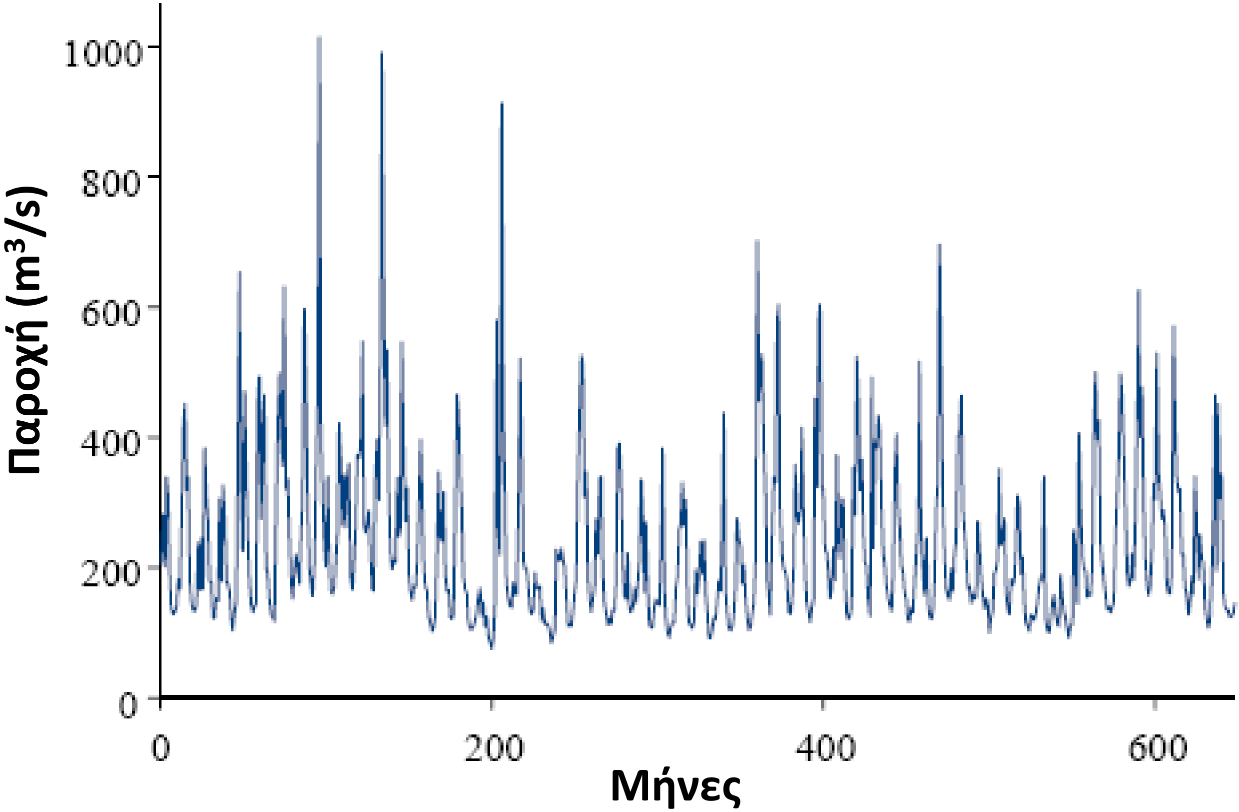
Ατμοσφαιρική φυσική

ταχύτητα ανέμου
(διάνυσμα !)

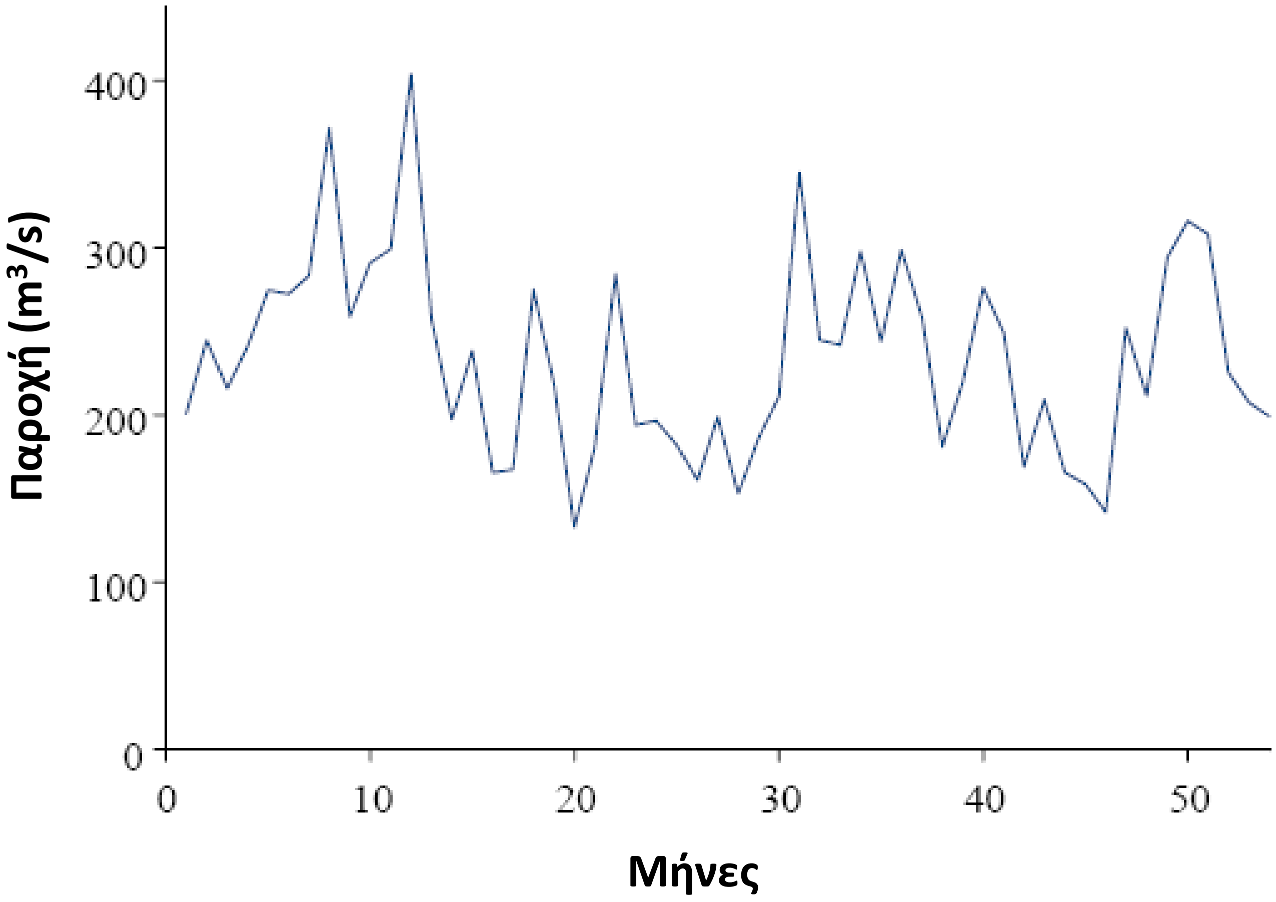
θερμοκρασία



Μηνιαίες Παροχές Ποταμού



Ετήσιες Παροχές Ποταμού



ΔΟΜΗ ΚΑΙ ΙΔΙΟΤΗΤΕΣ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

- ***Μνήμη υδρολογικής μεταβλητής*** : η εξάρτηση από προηγούμενες τιμές της ακολουθίας της χρονοσειράς ή μιας άλλης χρονοσειράς (συγγενικής)
- ***Ομοιογένεια*** : Μια χρονοσειρά είναι ομοιογενής στο χρόνο αν κάθε γεγονός έχει την ίδια ευκαιρία να συμβεί όλες τις φορές.
  - Ανομοιογένεια στο χρόνο, σε ένα συγκεκριμένο σημείο του χώρου, μπορεί να προέλθει από φυσικές ή ανθρωπογενείς επεμβάσεις στη σειρά.
    - Φυσικά φαινόμενα (π.χ. σεισμοί, πλημμύρες)
    - Φυσικές αλλαγές μπορούν να προέλθουν στην
      - Τάση (trend),
      - Περιοδικότητα (periodicity),
      - Εμμονή (persistence) της χρονοσειράς
    - Ανθρωπογενείς επεμβάσεις (π.χ κατασκευή φραγμάτων)
  - Ανομοιογενής μεταβλητή σημαίνει ότι τα δείγματά τις δεν ανήκουν στον ίδιο πληθυσμό

ΔΟΜΗ ΚΑΙ ΙΔΙΟΤΗΤΕΣ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

- ***Χωρική Ομοιογένεια*** υπάρχει όταν δύο ή περισσότερες χρονοσειρές στο χώρο έχουν στατιστικές παραμέτρους που δεν διαφέρουν πολύ μεταξύ τους, δηλαδή προέρχονται από τον ίδιο πληθυσμό.

- ***Μονιμότητα ή Στασιμότητα (stationarity)***
  - Μόνιμες ή στάσιμες (stationary) χρονοσειρές
  - Μη μόνιμες ή μη στάσιμες (non stationary)χρονοσειρές

Ορισμός Χρονικής σειράς

Χρονική σειρά είναι μια οικογένεια τυχαίων μεταβλητών X_t, $t \in T$, όπου T χρονική περίοδος ή υποσύνολο του χώρου, συνεχές ή διακριτό.

- X_t: η παρατήρηση σε χρόνο t
- Μονάδες χρόνου: έτος, μήνας, εποχή, ώρα,
- $\forall$ t το X_t θεωρείται τυχαία μεταβλητή.

Η Χρονική σειρά σαν Στοχαστική διαδικασία

Στοχαστική διαδικασία καλείται κάθε στατιστικό φαινόμενο που εκτυλίσσεται μέσα στο χρόνο. Οι χρονικές σειρές μπορούν να θεωρηθούν στοχαστικές διαδικασίες με πεπερασμένο πλήθος παρατηρήσεων.

Βασικές ιδιότητες Χρονοσειρών

Έστω $X_{1;}$ X_2, ..., X_n παρατηρήσεις μιας χρονοσειράς και F(Xi, X_2,..., X_n) η κοινή κατανομή τους.

Η μελέτη της κοινής κατανομής $F(X_1, X_2, ..., X_n)$ είναι δύσκολη. Συνήθως χρησιμοποιούνται οι ροπές πρώτης και δεύτερης τάξης.

α) Στασιμότητα

Μια χρονική σειρά (Χ-Σ) θα λέγεται **αυστηρά στάσιμη** ή **στάσιμη πρώτης τάξης** αν

$$F(X_{t_1}, X_{t_2}, ..., X_{t_n}) = F(X_{t_1+k}, X_{t_2+k}, ..., X_{t_n+k})$$

για οποιαδήποτε n-άδα $t_1, t_2, ..., t_n$ και οποιοδήποτε ακέραιο k.

Βασικές ιδιότητες Χρονοσειρών

α) Στασιμότητα

Μια χρονική σειρά θα λέγεται **στάσιμη δεύτερης τάξης** αν

α) $EX_t = \mu \;\kappa\alpha\iota\; Var(X_t) = \sigma_X^2 \;\forall t \in T$

β) $Cov(X_t, X_s) = \gamma_{|t-s|} \;\forall t, s \in T$ ή

$Cov(X_t, X_{t+k}) = \gamma_k \;\forall t \in T, \forall k \in N$

δηλαδή έχει συνδιασπορά σταθερή, **ανεξάρτητη του t**, η οποία **εξαρτάται μόνο από τη χρονική διαφορά των παρατηρήσεων** .

k = υστέρηση (lag)

Βασικές ιδιότητες Χρονοσειρών

β) Αντιστρεψιμότητα

Μια χ.σ. θα λέγεται αντιστρέψιμη εάν μπορεί να εκφρασθεί κατά μοναδικό τρόπο βάσει ενός "τύπου", τον οποίο μπορούμε να χρησιμοποιήσουμε για πρόβλεψη.

Παρατήρηση: Κάθε στάσιμη χρονική σειρά είναι αντιστρέψιμη, ενώ κάθε αντιστρέψιμη δεν είναι στάσιμη.

Μη στασιμότητα σε μια Χρονοσειρά

Τα δεδομένα πραγματοποίηση μιας διαδικασίας

$$Z_t = m_t + s_t + X_t$$

m_t: Συνιστώσα τάσης (trend component)

s_t : Συνιστώσα εποχικότητας (seasonal component)

X_t: Συνιστώσα τυχαίου θορύβου στάσιμη (random noise component)

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ όπου $EX_t = 0$

α) Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Τρόποι εξομάλυνσης

Μέθοδος 1η: Απομάκρυνση της τάσης με εξομάλυνση της σειράς

α) Εξομάλυνση με τη βοήθεια κινούμενου μέσου

β) Εκθετική εξομάλυνση

Μέθοδος 2η: Απομάκρυνση της τάσης με προσαρμογή πολυωνύμου

Μέθοδος 3η: Μέθοδος Box & Jenkins

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ όπου $EX_t = 0$

α) Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Μέθοδος 1α: Εξομάλυνση με τη βοήθεια κινούμενου μέσου

Κινούμενος μέσος μιας σειράς, $Z_t : t = 1, 2, ..., n$ είναι μια νέα σειρά m_t που ορίζεται από τη σχέση

$$m_t = \sum_{j=-p}^{p} w_j z_{t+j} \quad t = p+1, \ldots, n-p$$

όπου p θετικός ακέραιος και w_j βάρη θετικά, τέτοια ώστε $\sum w_j = 1$ και $w_j = w_{-j}$.

Ο αριθμός 2p+1 ονομάζεται **τάξη του κινούμενου μέσου**.

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ *όπου* $EX_t = 0$

Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Μέθοδος 1α: Εξομάλυνση με τη βοήθεια κινούμενου μέσου

- Οι παρατηρήσεις στην αρχή και στο τέλος δεν ορίζονται, συνήθως χρησιμοποιούνται οι μη εξομαλυμένες
- Συνήθως χρησιμοποιούνται βάρη ίσα με ένα, οπότε ο κινούμενος μέσος, ονομάζεται **απλός κινούμενος (ή κυλιόμενος) μέσος**
- Η συνήθης τάξη που χρησιμοποιείται είναι η 3 ή 5
- Ο συνήθης τρόπος για να χρησιμοποιηθούν άνισα βάρη, είναι να εφαρμόσουμε δύο ή τρεις διαδοχικές φορές έναν απλό κινούμενο μέσο
- Σαν κινούμενος μέσος μπορεί να ληφθεί και ο μέσος όρος p προηγούμενων παρατηρήσεων δηλαδή $m_t = \frac{1}{p}\sum_{j=t-p}^{t-1} z_j$

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ όπου $EX_t = 0$

Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Μέθοδος 1α: Εξομάλυνση με τη βοήθεια κινούμενου μέσου

Παράδειγμα : ο απλός κινούμενος μέσος για τάξη 3 είναι:

$$m_t = (z_{t-1} + z_t + z_{t+1})/3$$

ο απλός κινούμενος μέσος για τάξη 5 είναι:

$$m_t = (z_{t-2} + z_{t-1} + z_t + z_{t+1} + z_{t+2})/5$$

ο κινούμενος μέσος τάξης 3 για άνισα βάρη, χρησιμοποιώντας

π. χ. δύο διαδοχικές εφαρμογές απλού κινούμενου μέσου είναι:

$$m_t = \quad (z_{t-2} + z_{t-1} + z_t)/3 + (z_{t-1} + z_t + z_{t+1})/3 + (z_t + z_{t+1} + z_{t+2})/3 \quad /3 =$$
$$= (z_{t-2} + 2z_{t-1} + 3z_t + 2z_{t+1} + z_{t+2})/9$$

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ όπου $EX_t = 0$

α) Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Μέθοδος 1$^{\beta}$: Εκθετική Εξομάλυνση

Για οποιοδήποτε α∈[0, 1], οι μονόπλευροι κινούμενοι μέσοι $\hat{m}_t,\ t=1,2,..,n$ που ορίζονται από την επαναληπτική σχέση:

$$\hat{m}_t = a z_t + (1-a)\hat{m}_{t-1} \quad t=2,\dots,n \quad \kappa\alpha\iota \quad \hat{m}_1 = z_1$$

αποτελούν την **εκθετικά εξομαλυμένη** σειρά της δοθείσης σειράς z_t .

- για $t \geq 2$ $\quad \hat{m}_t = \sum_{j=0}^{t-2} a(1-a)^j z_{t-j} + (1-a)^{t-1} z_1$

που είναι ο κινούμενος μέσος των $z_t, z_{t-1}, \dots\dots$ με βάρη που ελαττώνονται εκθετικά

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ όπου $EX_t = 0$

α) Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Μέθοδος 2η: Απομάκρυνση της τάσης με προσαρμογή πολυωνύμου

Χρησιμοποιείται όταν υπάρχει τάση πολυωνυμικής μορφής. Στα δεδομένα (t, z_t) προσαρμόζεται ένα πολυώνυμο ως προς t.

$$m_t = \sum_{i=0}^{p} b_i t^i$$

και τα b_i εκτιμώνται από τα δεδομένα συνήθως με μεθόδους ελαχίστων τετραγώνων $(b = (x'x)^{-1}x'y)$.

- για p=1 ή 2 εξομαλύνονται απλές τάσεις.

Απομάκρυνση της τάσης μετά την εξομάλυνση: τα υπόλοιπα (*residuals*) $Y_t = Z_t - \hat{m}_t$ δεν έχουν τάση

Προσαρμογή μοντέλου μετά την απομάκρυνση της τάσης:

Προσαρμόζεται μοντέλο στα υπόλοιπα που προκύπτουν μετά την απομάκρυνση της τάσης.

Μοντέλο μόνο με τάση: $Z_t = m_t + X_t$ όπου $EX_t = 0$

α) Εκτίμηση και απομάκρυνση της τάσης όταν δεν υπάρχει εποχικότητα (όχι δόμηση μοντέλου)

Μέθοδος 3η: Μέθοδος Box & Jenkins

- Λήψη διαφορών πρώτης τάξης $Y_t = \nabla z_t = z_t - z_{t-1}$ ή
- Λήψη διαφορών δεύτερης τάξης $Y_t = \nabla^2 z_t = z_t - 2z_{t-1} + z_{t-2}$

Προσαρμογή μοντέλου στις πρώτες ή δεύτερες διαφορές

Παρατήρηση: Με κάθε διαφόριση χάνουμε μία παρατήρηση

β) Μη στασιμότητα ως προς τη μεταβλητότητα

Η σταθεροποίηση της μεταβλητότητας επιτυγχάνεται με διάφορους μετασχηματισμούς οι κυριότεροι των οποίων είναι:

- ο λογαριθμικός μετασχηματισμός (X_t= log(z_t))
- ο μετασχηματισμός της τετραγωνικής ρίζας (X_t= sqrt(z_t))

- Παρατήρηση: Εάν υπάρχουν παρατηρήσεις αρνητικές, προσθέτουμε μια σταθερά σ’ όλες τις παρατηρήσεις, ώστε να γίνουν θετικές.

γ) Εποχική μεταβλητότητα

Μία χρονική σειρά λέγεται **εποχική (seasonal)** με εποχικότητα s, όταν παρουσιάζει την ίδια συμπεριφορά ανά s χρονικά διαστήματα.

- Απαλοιφή εποχικότητας:
  - **Εποχικές διαφορές πρώτης τάξης:** $Y_t = (1 - B^s)z_t = z_t - z_{t-s}$
  - **Εποχικές διαφορές δεύτερης τάξης:** $Y_t = (1 - B^s)^2 z_t = z_t - 2z_{t-s} + z_{t-2s}$

Ιδιότητες της συνδιασποράς

Αν $c_1, c_2, .., c_m$ και $d_1, d_2, .., d_n$ σταθερές, τότε

$$\operatorname{cov}\left[\sum_{i=1}^{m} c_i z_t, \sum_{j=1}^{n} d_j z_S\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j \operatorname{cov}(z_{t_i}, z_{S_j})$$

και

$$Var\left[\sum_{i=1}^{m} c_i z_t\right] = \sum_{i=1}^{m} c_i^2\, Var(z_t) + 2\sum_{i=2}^{m}\sum_{j=1}^{i-1} \operatorname{cov}(z_{t_i}, z_{t_j})$$

Ιδιότητες της συνδιασποράς

Παράδειγμα : Δίνεται η παρακάτω χρονική σειρά

$$\boxed{x_t = \mathrm{A}\sin(\omega t + \theta)}$$

όπου Α τ.μ. με μ=0 και $\sigma_A^2 = 1$, θ τ.μ. τέτοια ώστε θ ~ U(-π, π) και Α και θ ανεξάρτητες τ.μ. Είναι η σειρά στάσιμη;

$$\mathrm{E}x_t = \mathrm{E}\ \ \mathrm{A}\sin(\omega t + \theta)\ \ = \mathrm{EAE}(\sin\omega t\cos\theta + \cos\omega t\sin\theta) = 0$$

$$Cov(x_t, x_{t+k}) = \mathrm{E}(x_t x_{t+k}) - \mathrm{E}x_t\mathrm{E}x_{t+k} = \mathrm{E}\ \ \mathrm{A}^2\sin(\omega t + \theta)\sin\ \ \omega(t+k) + \theta\ \ =$$

$$\ldots\ldots\ldots\ldots\ldots = \frac{1}{2}\cos(\omega k)$$

Δηλαδή μέση τιμή σταθερή και συνδιασπορά που εξαρτάται μόνο από το k. Η σειρά λοιπόν είναι στάσιμη.

Ιδιότητες της συνδιασποράς

Παράδειγμα : Έστω z_t μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν εναλλάξ μια N(0,1) και μια δίτιμη διακριτή ομοιόμορφη κατανομή που παίρνει τις τιμές +1 και -1 με πιθανότητα ½.

Άρα $Ez_t = 0$, $Var(z_t) = Ez_t^2 = 1$, $\forall t$ και

$$E(z_t, z_{t+k}) = \begin{cases} 0 & k \neq 0 \\ 1 & k = 0 \end{cases}$$

$$\rho_k = \frac{\operatorname{cov}(z_t, z_{t+k})}{\sqrt{\operatorname{var} z_t \operatorname{var} z_{t+k}}} = \begin{cases} 0 & k \neq 0 \\ 1 & k = 0 \end{cases} \Rightarrow$$

z_t στάσιμη β΄ τάξης.

Συναρτήσεις αυτοδιασποράς και αυτοσυσχέτισης

Αν $\{X_t : t \in 0, \pm 1, \pm 2, \ldots.\}$ τότε $\mu_t = E(X_t)$ για $t = 0, \pm 1, \ldots..$

Τα μ_t γενικά είναι διαφορετικά, όταν όμως η χ.σ. είναι στάσιμη, $\mu_t = \mu \ \forall t$.

Συνάρτηση θεωρητικής αυτοδιασποράς υστέρησης k

$$\gamma_k = \mathrm{E}(x_t - \mu)(x_{t+k} - \mu) = \mathrm{cov}(x_t, x_{t+k}) \quad k = 0, \pm 1, \pm 2, \ldots.$$

Συνάρτηση δειγματικής αυτοδιασποράς υστέρησης k

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad k = 0, \pm 1, \pm 2, \ldots.$$

Συναρτήσεις αυτοδιασποράς και αυτοσυσχέτισης

Συνάρτηση θεωρητικής αυτοσυσχέτισης υστέρησης k

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\mathrm{cov}(x_t, x_{t+k})}{\sqrt{\mathrm{var}\, x_t \ \mathrm{var}\, x_{t+k}}} \qquad k = 0, \pm 1, \pm 2, \ldots.$$

ρ_k =συντελεστής συσχέτισης μεταξύ παρατηρήσεων που απέχουν k χρονικές στιγμές

Συνάρτηση δειγματικής αυτοσυσχέτισης υστέρησης k

$$r_k = \frac{c_k}{c_0} \qquad k = 0, \pm 1, \pm 2, \ldots.$$

Ιδιότητες:

1. $\gamma_0 = \mathrm{var}(x_t)$ $\qquad \rho_0 = 1$
2. $\gamma_k = \gamma_{-k}$ $\qquad \rho_k = \rho_{-k}$
3. $|\gamma_k| \leq \sqrt{\gamma_0^2}$ $\qquad |\rho_k| \leq 1$

Συναρτήσεις αυτοδιασποράς και αυτοσυσχέτισης

- ρ_k κοντά στο 1 σημαίνει υψηλή θετική γραμμική συσχέτιση
- ρ_k = 0 συνεπάγεται ανεξαρτησία των παρατηρήσεων
- ρ_k κοντά στο -1 σημαίνει υψηλή αρνητική συσχέτιση

Ορίζουμε:

$$\Gamma_N = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N-1} & \gamma_{N-2} & \cdots & \gamma_0 \end{pmatrix} = \sigma_X^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \cdots & \cdots & 1 \end{pmatrix} = \sigma_X^2 \mathbf{P}_N$$

Γ_N = πίνακας αυτοδιακυμάνσεων ή πίνακας αυτοδιασπορών

P_N = πίνακας αυτοσυσχετίσεων

Γ_N, P_N πίνακες συμμετρικοί, θετικά ορισμένοι

Συναρτήσεις αυτοδιασποράς και αυτοσυσχέτισης

Πράγματι: αν $L_t = \sum_{i=1}^{n} l_i x_{t_i}$ όπου L_t τ.μ., l_i σταθερές, x_{t_i} τ.μ. που

αντιστοιχούν στις χρονικές στιγμές $t_1, t_2, \ldots, t_n$ τότε

$$0 \leq Var(L_t) = \sum_{i=1}^{n}\sum_{j=1}^{n} l_i l_j \operatorname{cov}(x_{t_i}, x_{t_j}) = \sum_{i=1}^{n}\sum_{j=1}^{n} l_i l_j \gamma_{|i-j|} = l' \Gamma_N l$$

Επειδή P_N θετικά ορισμένος προκύπτει:

- για n=2 $\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0$ ή $-1 < \rho_1 < 1$

- για n=3 $-1 < \rho_1 < 1$

  $-1 < \rho_2 < 1$

  $-1 < \dfrac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1$

Correlogram *= το γράφημα των* $\boldsymbol{\rho_k}$ *ή* $\boldsymbol{r_k}$

Συναρτήσεις αυτοδιασποράς και αυτοσυσχέτισης

Αποδεικνύεται ότι $r_k \sim N(\rho_k, Var(r_k))$, όπου

$$Var(r_k) = \frac{1}{N}\sum_{m=-\infty}^{\infty}(\rho_m^2 + \rho_{m+k}\rho_{m-k} + 2\rho_k^2\rho_m^2 - 4\rho_k\rho_m\rho_{m-k}) \quad \text{(Bartlett 1946)}$$

Για N μεγάλο $Var(r_k) \simeq \frac{1}{\mathrm{N}}\sum_{m=-\infty}^{m=\infty}\rho_m^2$

Για τυχαία διαδικασία $Var(r_k) \simeq \frac{1}{N}$

Έλεγχος σημαντικότητας των r_k :

H_0: $\rho_k = 0$ η H_0 απορρίπτεται αν r_k έξω από το διάστημα $\pm\frac{2}{\sqrt{N}}$

H_1: $\rho_k \neq 0$

Η συνάρτηση αυτοσυσχέτισης (Autocorrelation Function - ACF)

- Ο συντελεστής συσχέτισης ρ_s ανάμεσα στην Y_t και την Y_{t+S} ονομάζεται και **συντελεστής αυτοσυσχέτισης (autocorellation coefficient)**.

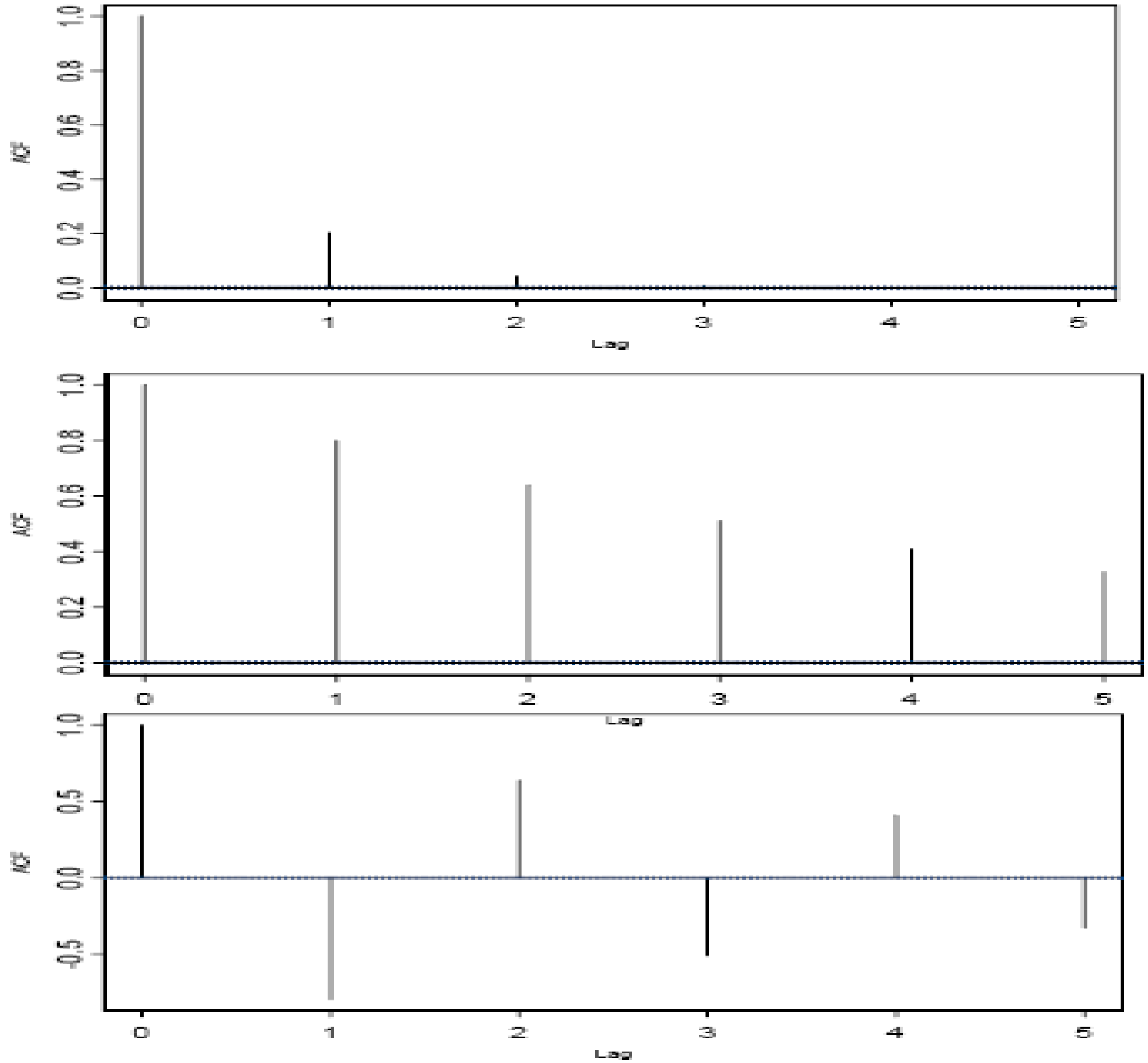
$$\rho_s = \frac{Cov(Y_t, Y_{t+s})}{\sqrt{V(Y_t)V(Y_{t+s})}}$$

Ισχύει ότι: $\rho_s = \rho_{-s}$

Συνάρτηση αυτοσυσχέτισης (ACF - Autocorellation Function):
ονομάζεται **η σχέση που υπάρχει ανάμεσα στον ρs και στο s**

και **διάγραμμα αυτοσυσχέτισης (Correlogram)** η γραφική παράσταση αυτής

Η συνάρτηση αυτοσυσχέτισης (Autocorrelation Function - ACF)



Γράφημα της ACF

Στατιστικοί έλεγχοι του συντελεστή αυτοσυσχέτισης

- Στην κατηγορία των στατιστικών ελέγχων που αφορούν το συντελεστή αυτοσυσχέτισης και σχετίζονται με την **στασιμότητα ή μη στασιμότητα** μιας χρονολογικής σειράς έχουμε τους ακόλουθους ελέγχους:

- **Ο έλεγχος του Bartlett (Bartlett Test)**

- **Η Q Στατιστική των Box – Pierce (Box – Pierce Test)**

- **Η Στατιστική των Ljung – Box (Ljung – Box Statistic)**

• Ο έλεγχος του Bartlett (Bartlett Test)

Ο έλεγχος του Bartlett (Bartlett Test) στηρίζεται στην ακόλουθη **υπόθεση**:

- Αν η χρονολογική σειρά είναι **στάσιμη** τότε οι **συντελεστές αυτοσυσχέτισης ρ_s** του δείγματος **ακολουθούν προσεγγιστικά την κανονική κατανομή** με **μέσο μηδέν και διακύμανση 1/T (T το μέγεθος του δείγματος)*.**

  [* Για μεγάλα δείγματα οι συντελεστές αυτοσυσχέτισης ρ_s του δείγματος ακολουθούν **την κανονική κατανομή** με μέσο μηδέν και διακύμανση 1/T (T το μέγεθος του δείγματος)].

Θεωρούμε την **μηδενική υπόθεση**:

$H_0 : \rho_s = 0$ (Η χρονολογική σειρά είναι **στάσιμη**)

Η υπόθεση H_0 ελέγχεται με τη στατιστική: $t_s = \frac{\rho_s}{\sqrt{\frac{1}{T}}}$

Για α = 5% και για T > 30 η κρίσιμη τιμή του t_α είναι (- ή +)1,96 κατά συνέπεια η H_0 απορρίπτεται για

$$\boldsymbol{t_s = \rho_s\sqrt{T}} < -1{,}96 \text{ ή } \boldsymbol{t_s = \rho_s\sqrt{T}} > 1{,}96$$

• Q Στατιστική (Box – Pierce Test)

Ο έλεγχος της στασιμότητας μιας χρονολογικής σειράς μπορεί να γίνει και με τη **στατιστική Q των Box – Pierce.**

Θεωρούμε την **μηδενική υπόθεση**

$H_0 : \rho_1 = \rho_2 = \ldots = \rho_ς = 0$ (Η χρονολογική σειρά είναι στάσιμη)

Η **στατιστική Q των Box – Pierce** που χρησιμοποιείται για τον έλεγχο της H_0 είναι (ρ*s οι δειγματικές αυτοσυσχετίσεις των καταλοίπων):

$$Q = T \sum_{s=1}^{m} \rho{*}_s^{\,2}$$

Όπου m το μήκος της χρονικής υστέρησης, ρ_s οι δειγματικοί συντελεστές αυτοσυσχέτισης των παρατηρήσεων, Τ ο αριθμός των παρατηρήσεων.

Η **στατιστική Q** ακολουθεί **την X^2 κατανομή με m βαθμούς ελευθερίας.**
Για δεδομένο επίπεδο σημαντικότητας α:
Αν $Q > X^2(\alpha, m)$ η H_0 απορρίπτεται και η χρονολογική σειρά **δεν είναι στάσιμη**

• Η Στατιστική των Ljung – Box (Ljung – Box Statistic)

- Η στατιστική των **Ljung – Box,** ακολουθεί την κατανομή $X^2(\alpha,m)$ και δίνει καλύτερα αποτελέσματα από την Q στατιστική όταν έχουμε μικρά δείγματα **T < 30.**

Θεωρούμε την **μηδενική υπόθεση** H_0: $\rho_s = 0$
(Η χρονολογική σειρά είναι **στάσιμη)**

Η στατιστική που χρησιμοποιείται είναι:

$$LB = T(T+2)\sum_{s=1}^{m}\frac{{p^*_s}^2}{T-s}$$

Για δεδομένο **επίπεδο σημαντικότητας α**:

- Αν LB > $X^2(\alpha, m)$ απορρίπτεται η H_0, άρα η χρονολογική σειρά **δεν είναι στάσιμη**

Συνάρτηση μερικής αυτοσυσχέτισης

Η **συνάρτηση μερικής αυτοσυσχέτισης PACF**, ορίζεται από τη σχέση:

$$\varphi_{\kappa\kappa} = Corr(x_t x_{t+k} / x_{t+1}, \cdots, x_{t+k-1}) \quad k = 1, 2, \ldots$$

δηλ. είναι ο συντελεστής συσχέτισης της διδιάστατης κατανομής των x_t και x_{t+K} με συνθήκη στις $x_{t+1}, \cdots, x_{t+k-1}$.

Υπολογισμός συνάρτησης μερικής αυτοσυσχέτισης

Έστω το μοντέλο παλινδρόμησης:

$$z_{t+k} = \varphi_{k1} z_{t+k-1} + \varphi_{k2} z_{t+k-2} + \cdots + \varphi_{kk} z_t + e_{t+k}$$

στο οποίο η φ_{kj} είναι η j παράμετρος παλινδρόμησης και e_{t+k} κανονικά κατανεμημένα σφάλματα ασυσχέτιστα με την z_{t+k-j} για $j \geq k$, τότε:

$$E(z_{t+k-j} z_{t+k}) = \varphi_{k1} E(z_{t+k-j} z_{t+k-1}) + \cdots + \varphi_{kk} E(z_{t+k-j} z_t) + E(z_{t+k-j} e_{t+k}) \text{ ή}$$

$$\gamma_j = \varphi_{k1} \gamma_{j-1} + \varphi_{k2} \gamma_{j-2} + \cdots + \varphi_{kk} \gamma_{j-k} \text{ ή}$$

$$\rho_j = \varphi_{k1} \rho_{j-1} + \varphi_{k2} \rho_{j-2} + \cdots + \varphi_{kk} \rho_{j-k}$$

Συνάρτηση μερικής αυτοσυσχέτισης

Υπολογισμός συνάρτησης μερικής αυτοσυσχέτισης (συν.)

Για j=1, 2,....., k παίρνουμε το παρακάτω σύστημα:

$$\rho_1 = \varphi_{k1}\rho_0 + \varphi_{k2}\rho_1 + \cdots + \varphi_{kk}\rho_{k-1}$$
$$\rho_2 = \varphi_{k1}\rho_1 + \varphi_{k2}\rho_0 + \cdots + \varphi_{kk}\rho_{k-2}$$
$$\vdots$$
$$\rho_k = \varphi_{k1}\rho_{k-1} + \varphi_{k2}\rho_{k-2} + \cdots + \varphi_{kk}\rho_0$$

Χρησιμοποιώντας τον κανόνα του Grammer διαδοχικά για k=1,2,… υπολογίζουμε τις φ_{kk}:

- για k=1 $\boldsymbol{\varphi_{11} = \rho_1}$
- για k=2 $\varphi_{22} = \dfrac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$

για k=3 $\varphi_{33} = \dfrac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$

Συνάρτηση μερικής αυτοσυσχέτισης

Υπολογισμός συνάρτησης μερικής αυτοσυσχέτισης (συν.)

- και τέλος $$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

Αποδεικνύεται ότι: $\hat{\varphi}_{kk} \sim N(\varphi_{kk}, \frac{1}{N})$

Έλεγχος σημαντικότητας των φ_{kk} :

$H_0: \ \varphi_{kk} = 0$

$H_1: \ \varphi_{kk} \neq 0$

η H_0 απορρίπτεται αν φ_{kk} έξω από το διάστημα $\left(-\frac{2}{\sqrt{N}}, \frac{2}{\sqrt{N}}\right)$

Η συνάρτηση μερικής αυτοσυσχέτισης (Partial Autocorrelation Function - PACF)

Η **συνάρτηση της μερικής αυτοσυσχέτισης** ορίζεται ανάμεσα στην Y_t και στην Y_{t-s} και αναφέρεται στην συσχέτιση ανάμεσά τους, **όταν έχουν αφαιρεθεί οι γραμμικές επιδράσεις των ενδιαμέσων μεταβλητών**

$$Y_{t-1},\ Y_{t-2},\ \ldots,\ Y_{t-(s-1)}$$

Ο συντελεστής μερικής αυτοσυσχέτισης p_{ss} είναι ο συντελεστής μερικής παλινδρόμησης της μεταβλητής Y_{t-s} στο υπόδειγμα:

$$Y_t = p_{1s}Y_{t-1} + \ldots\ldots + p_{ss}Y_{t-s} + \varepsilon_t$$

➢Για μια **αυτοπαλίνδρομη διαδικασία τάξης p η συνάρτηση μερικής αυτοσυσχέτισης είναι μηδέν για s > p.**

Ο συντελεστής μερικής αυτοσυσχέτισης ρ_{ss} είναι σημαντικός, δηλαδή όχι μηδέν,

Όταν $t_s = \left| p_{ss}\sqrt{T} \right|$ **> 2***

(Θεωρούμε α = 5%, βλέπε έλεγχο του Bartlett για συντελεστή αυτοσυσχέτισης, * <u>το 2 παίρνεται προσεγγιστικά στο 1,96</u>)

✓Με τον **έλεγχο σημαντικότητας των συντελεστών μερικής αυτοσυσχέτισης ρ_{ss} μπορεί να καθορισθεί η τάξη μιας AR διαδικασίας.**

Δηλαδή εξετάζοντας την ακολουθία των τιμών t_s για $s = 1, 2, \ldots\ldots$επιλέγεται ως **τάξη της σειράς αυτή που αντιστοιχεί στην τελευταία σημαντική τιμή του t_s.**

Για παράδειγμα:

αν η τελευταία σημαντική τιμή του t_s είναι για $s = 2$, δηλαδή **ο συντελεστής ρ_{22} είναι σημαντικός,** ενώ **ο συντελεστής ρ_{33} δεν είναι σημαντικός**, τότε **η τάξη ρ του υποδείγματος είναι το 2.**

Τεχνικές ανάλυσης χρονολογικών σειρών υδρολογικών και μετεωρολογικών μεταβλητών

- **Η κλασική μέθοδος ανάλυσης**

και

- **Η στοχαστική μέθοδος ανάλυσης**

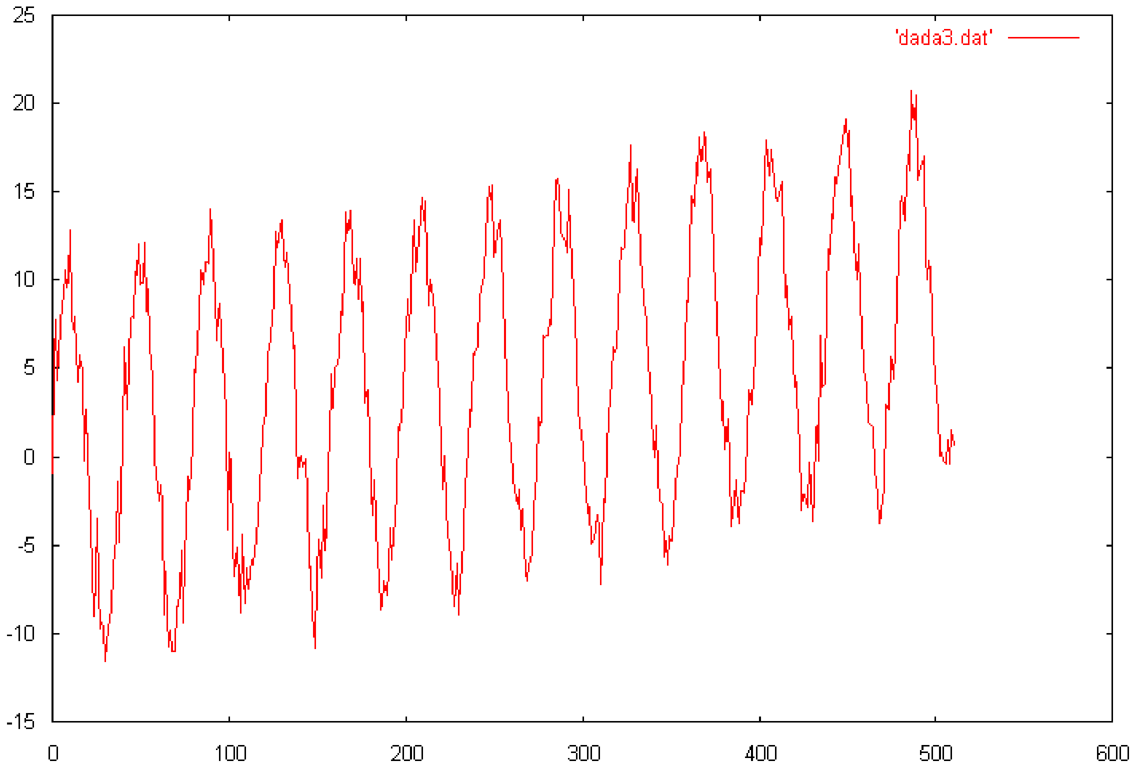
  (στοχαστικά υποδείγματα ή ομοιώματα)

Κλασική μέθοδος ανάλυσης

Στην **κλασική μέθοδο ανάλυσης** μια χρονολογική σειρά $\{x_t\}$ αναλύεται στις ακόλουθες **συνιστώσες:**

- **Τάση (T_t),**
- **Εποχιακές διακυμάνσεις (S_t)**
- **Μη – ομαλές κινήσεις (Y_t)**

$$X_t = T_t + S_t + Y_t$$

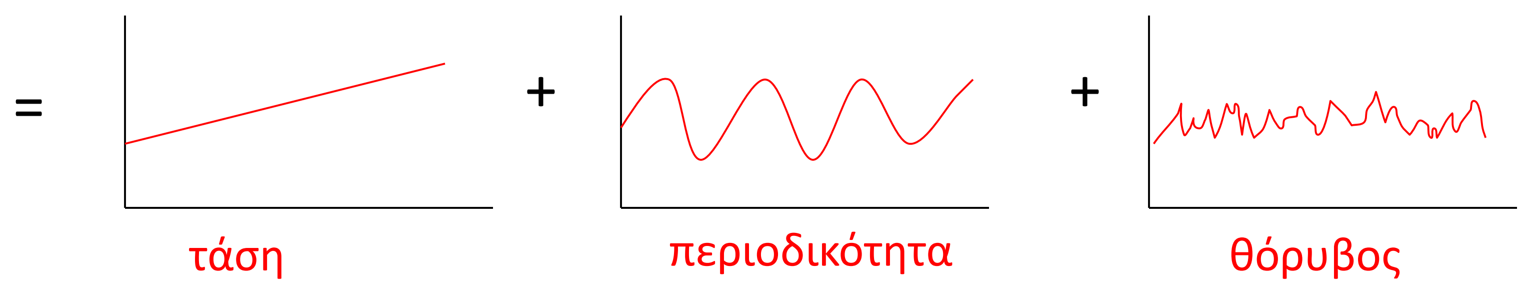


Κλασική μέθοδος ανάλυσης

Συνιστώσες:

- **τάση,**
- **περιοδικότητα,**
- **θόρυβος**

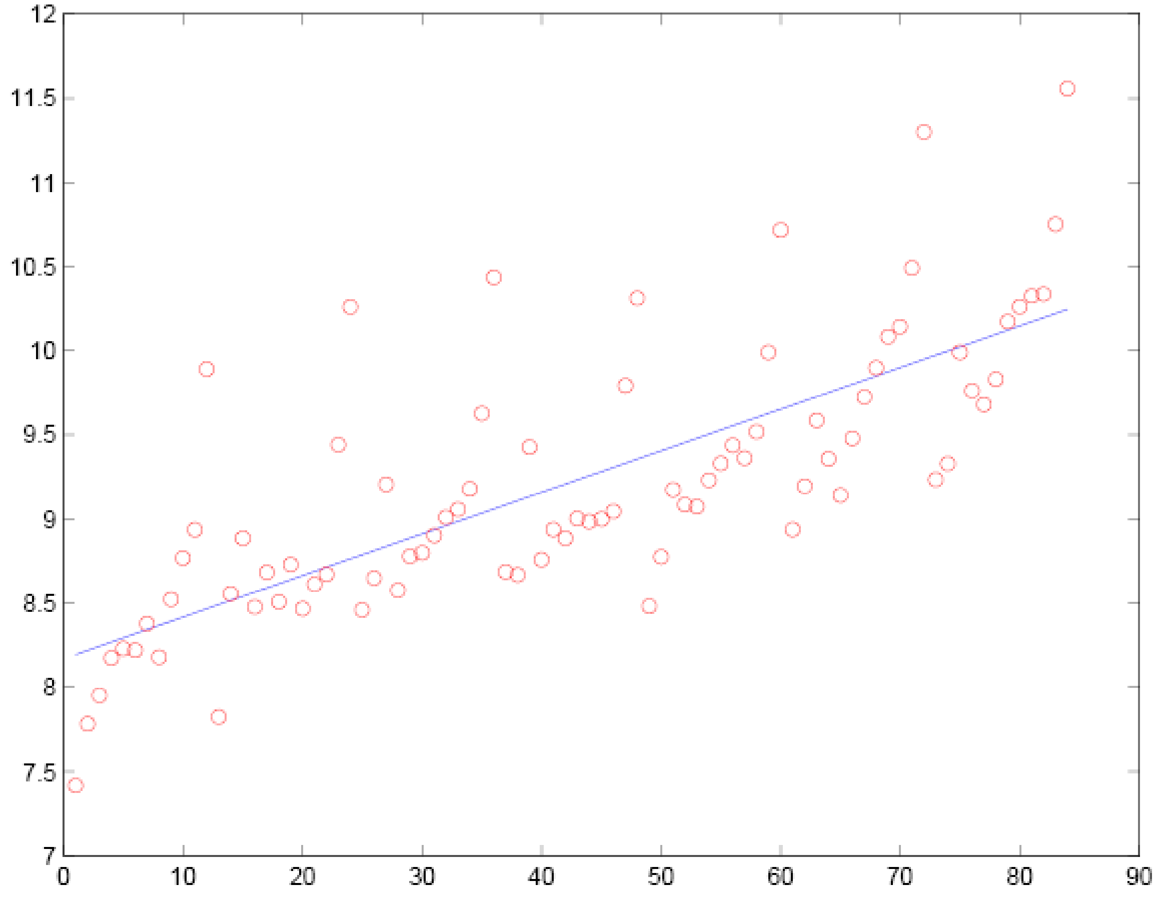
Σκοπός: **απομόνωση της κάθε συνιστώσας**



έτσι μπορούμε να χαρακτηρίσουμε πιο καθαρά την κάθε συνιστώσα

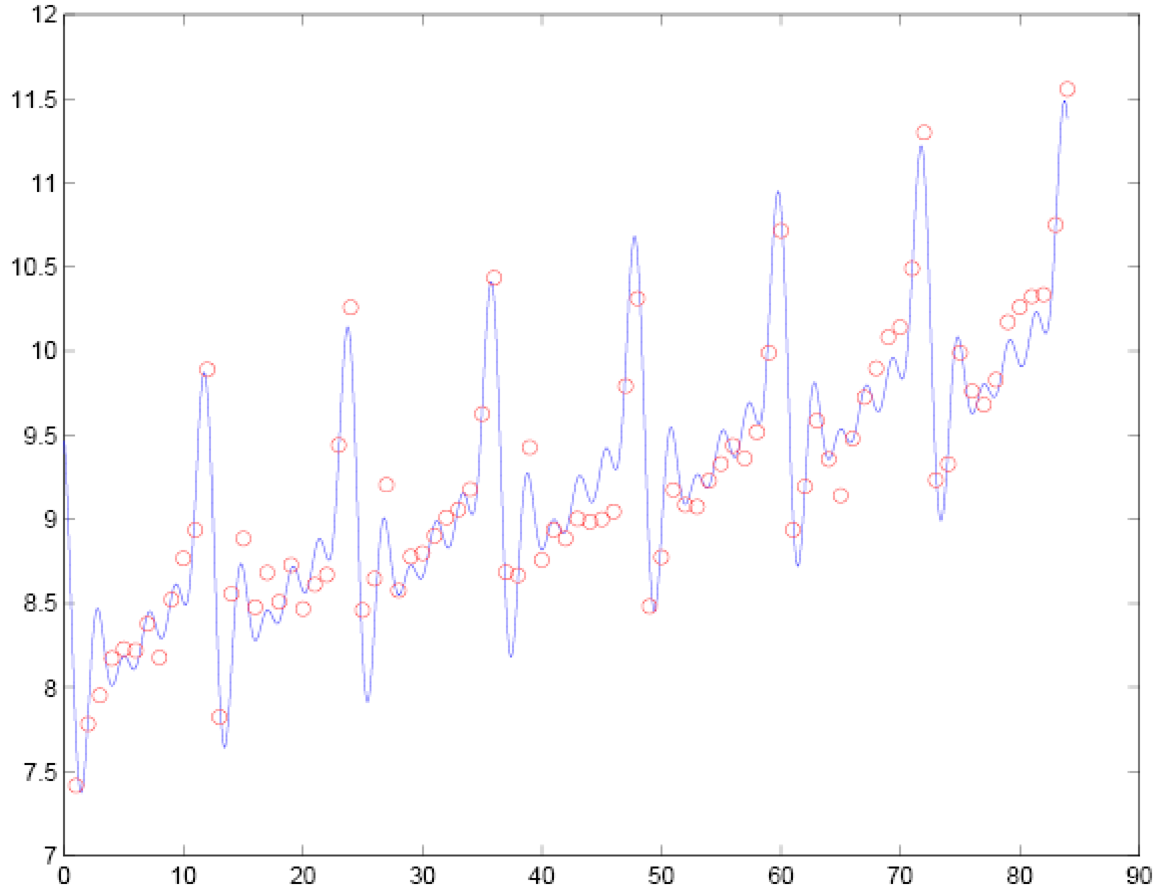
Κλασική μέθοδος ανάλυσης

Τάση Υποθετικής Χρονολογικής Σειράς

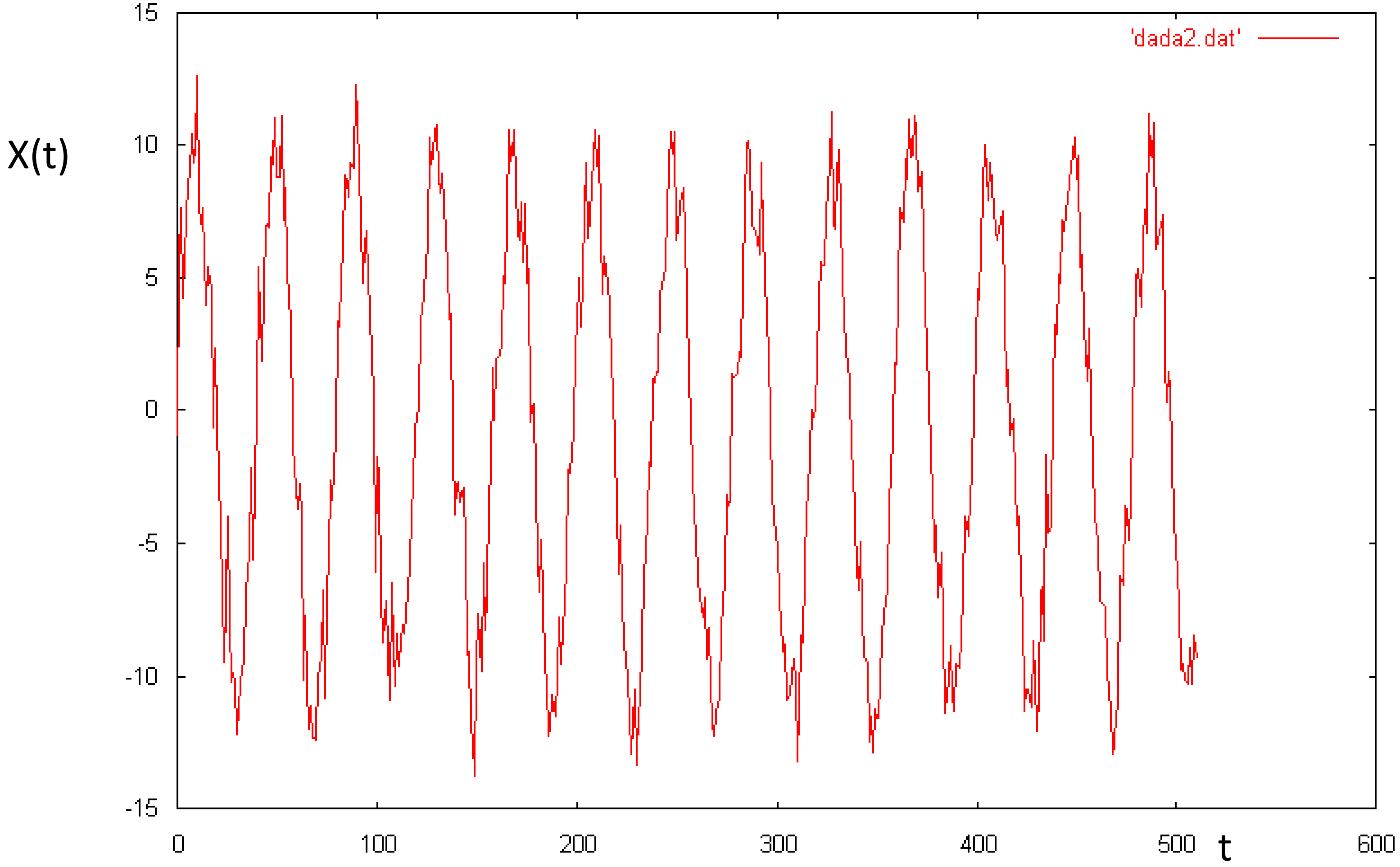


Κλασική μέθοδος ανάλυσης

Τάση και Εποχικότητα Υποθετικής Χρονολογικής Σειράς



Παράδειγμα



Ιδιότητες ?

- Κάπως περιοδική
- Έχει θόρυβο

Μέσος όρος **μ** χρονολογικής σειράς

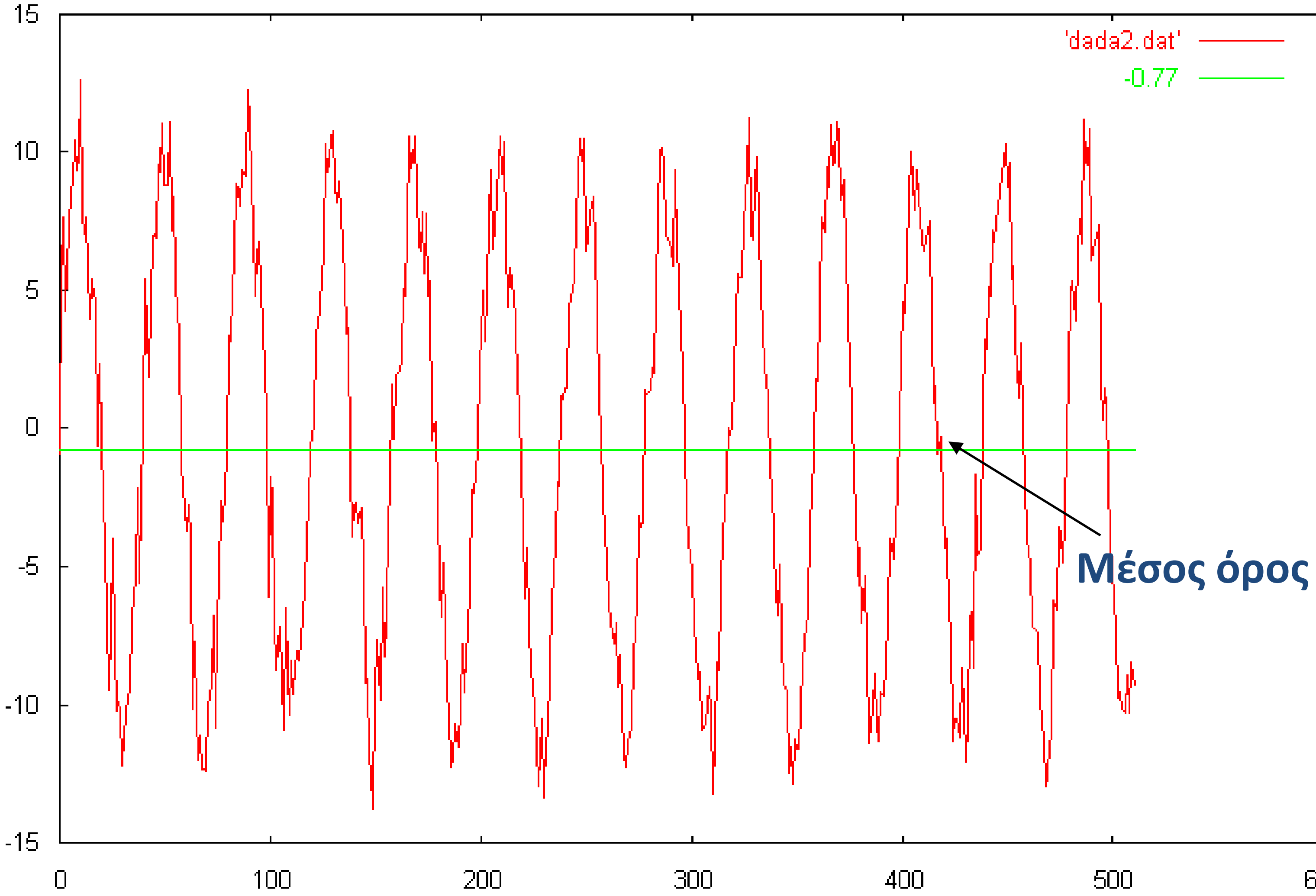
Έστω η ΧΣ

Ορισμός του μέσου όρου **μ** χρονολογικής σειράς

$$\mu := \frac{1}{N}\sum_{i=1}^{N} X(t_i)_N$$

δηλ. η μέση τιμή όλων των τιμών της ΧΣ, όπου η σειρά των $X(t_i)$ δεν παίζει ρόλο !

μ= - 0.77



Διακύμανση σ^2 της χρονολογικής σειράς

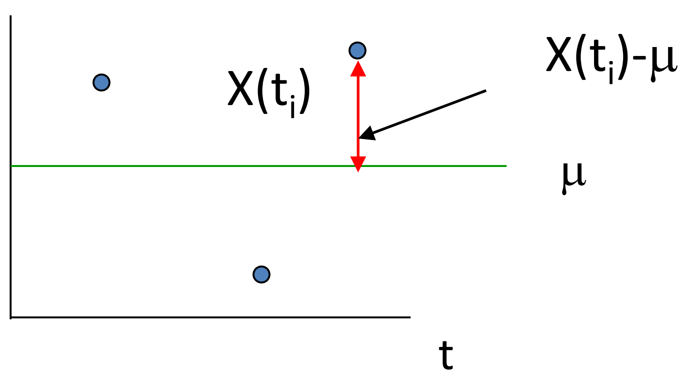
Έστω η ΧΣ

Ορισμός της διακύμανσης της χρονολογικής σειράς

$$\sigma^2 := \frac{1}{N-1}\sum_{i=1}^{N}\left(X(t_i)-\mu\right)^2$$

$$X(t_i),\ i = 1, 2, 3, ..., N$$

Μέσος όρος των αποκλίσεων από τη μέση τιμή στο τετράγωνο



Τυπική απόκλιση σ (standard deviation) της χρονολογικής σειράς

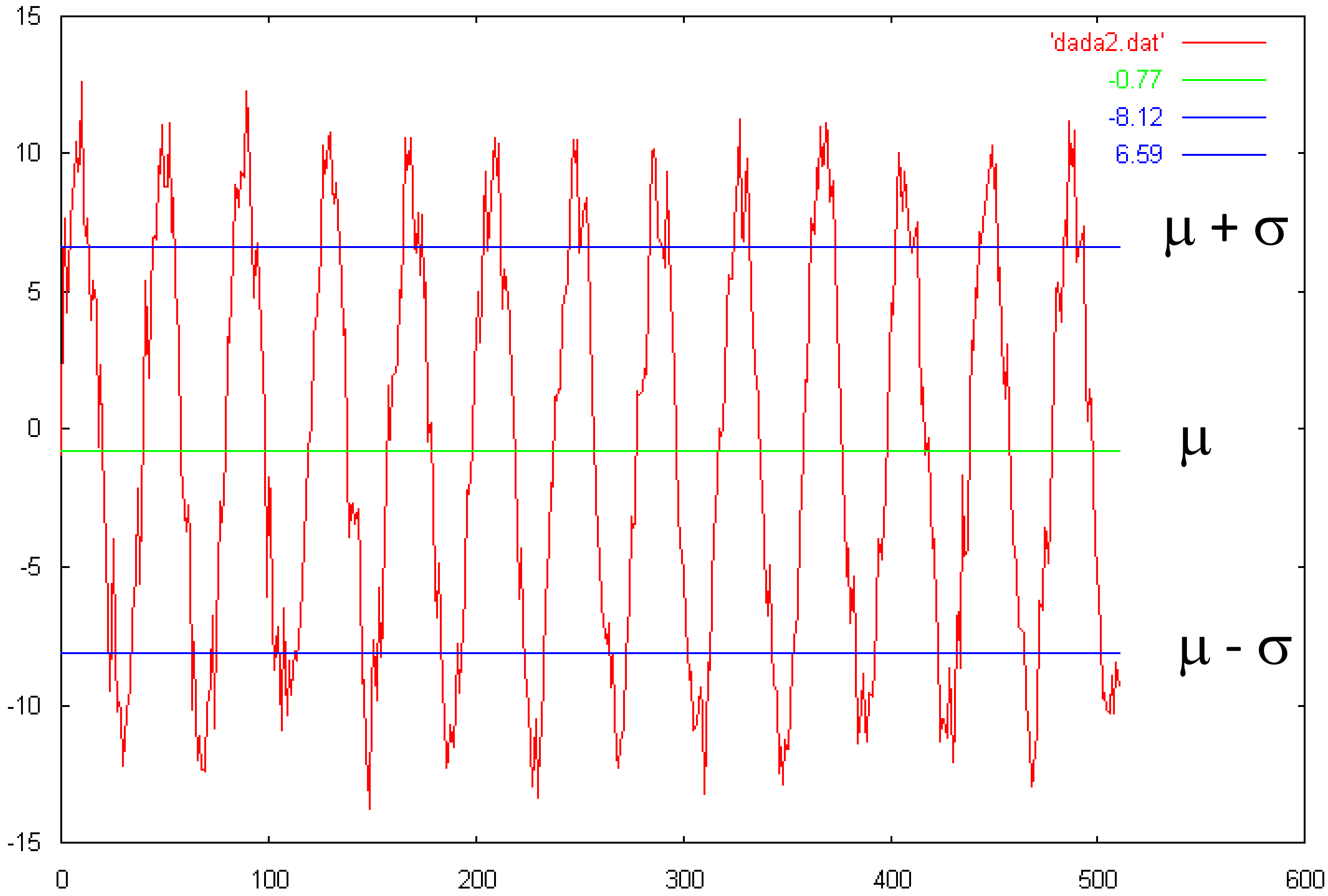
- Η διασπορά έχει μαθηματικά πλεονεκτήματα, δηλ. στην στατιστική θεωρία

- Πιο διαισθητική είναι η τυπική απόκλιση **σ** (standard deviation)

$$\sigma = \sqrt{\sigma^2} := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(X(t_i)-\mu\right)^2}$$

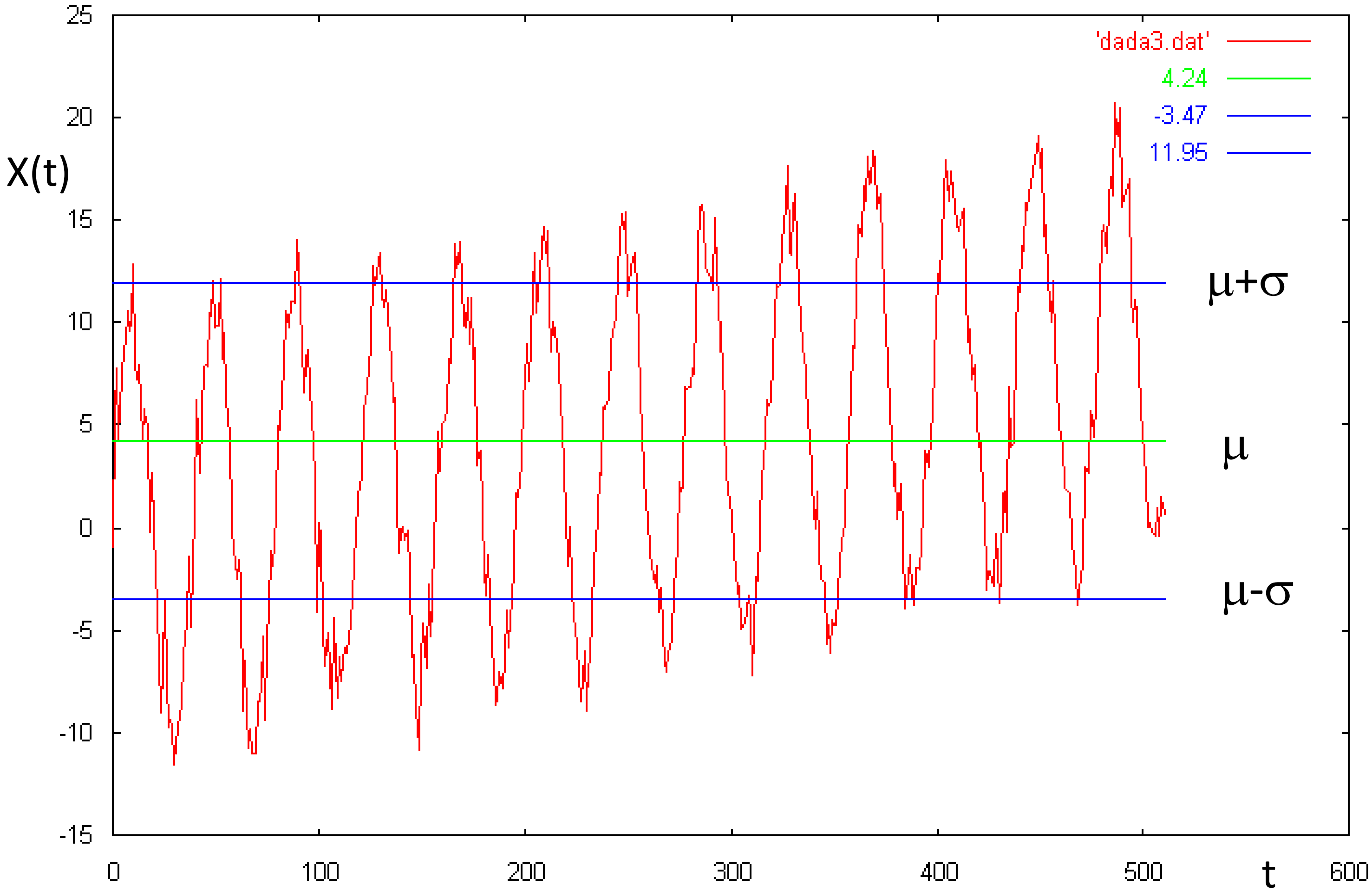
- μέση απόκλιση από την μέση τιμή

Παράδειγμα: $\sigma^2 = 54.10$, $\sigma = 7.36$, $\mu = -0.77$



Μεταξύ μ-σ και μ+σ βρίσκονται τα περισσότερα σημεία της ΧΣ,
Το διάστημα αυτό μας δίνει τη διακύμανση των τιμών της ΧΣ

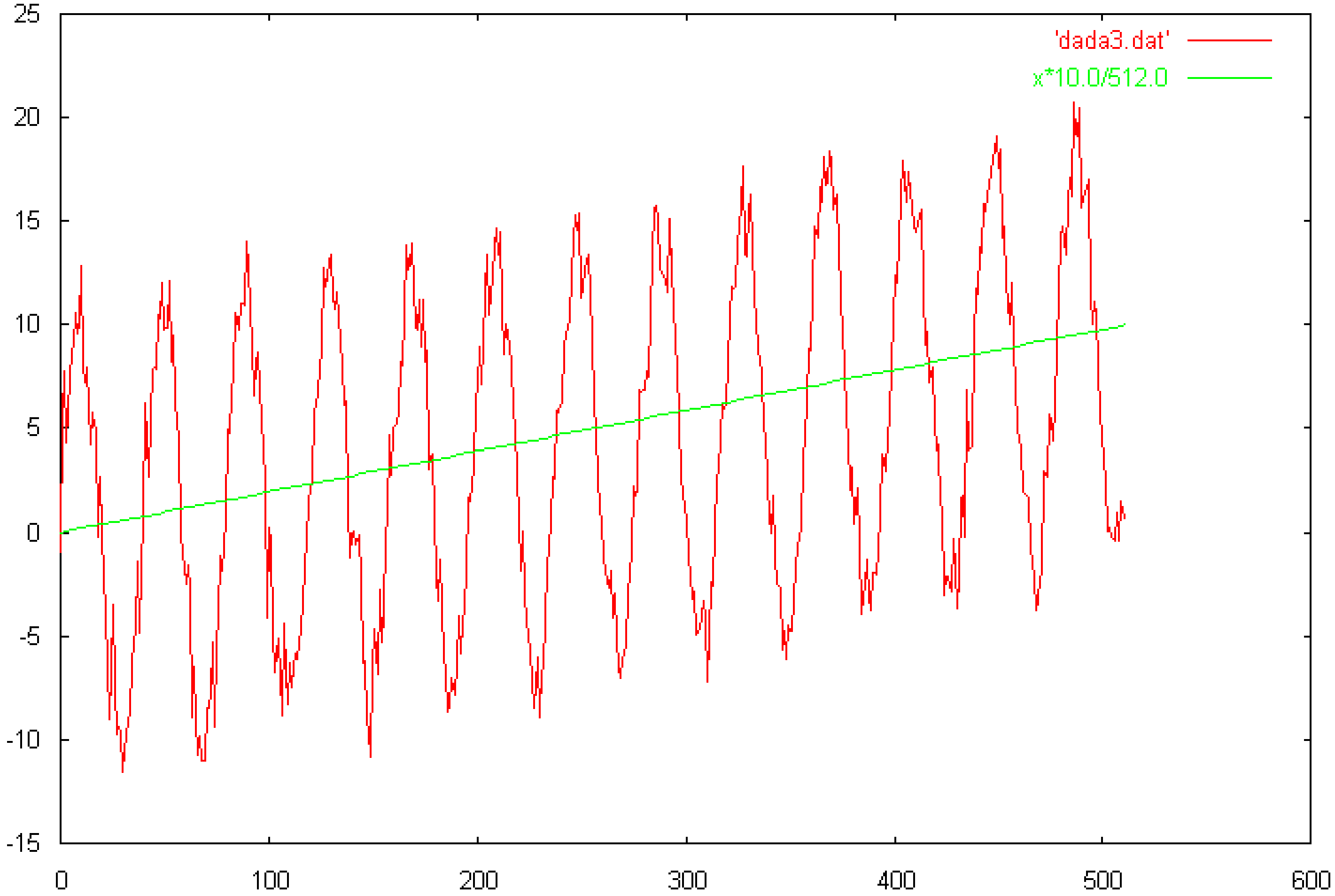
Παράδειγμα: $\mu = 4.24$, $\sigma^2 = 59.50$, $\sigma = 7.71$



μ και σ δεν δίνουν καλή περιγραφή της πραγματικής μέσης τιμής εδώ

Λόγος: υπάρχει μια τάση (trend), δηλ. σαν να αλλάζει το πραγματικό μ στο χρόνο: $\mu = \mu(t)$

Χρονολογική σειρά με τάση (trend)



Στασιμότητα (stationarity)

- <u>διαισθητικός ορισμός</u>:
μια ΧΣ είναι στάσιμη αν δεν υπάρχει συστηματική αλλαγή του μέσου όρου και της διασποράς στο χρόνο

- <u>Αυτό σημαίνει:</u> τάση = μη - στασιμότητα

- η στασιμότητα είναι προϋπόθεση για τα περισσότερα εργαλεία της ΑΧΣ (π.χ. αυτο-συσχέτιση, φασματική ανάλυση)
- χρειάζονται εργαλεία μετατροπής μη-στάσιμων σε στάσιμες ΧΣ !

Μέθοδος της ανάλυσης (decomposition)

- βασικός σκοπός: να διαχωρίσουμε και να απομονώσουμε τα διάφορα χαρακτηριστικά μιας χρονοσειράς:

κυρίως τα μη-στάσιμα από τα στάσιμα χαρακτηριστικά

- Εργαλεία: εξομάλυνση (smoothing), προσαρμογή (fitting), αφαίρεση, διαφορά …

Ανάλυση Περιοδικότητας

- Η **περιοδικότητα** συνήθως συνυπάρχει με τη στοχαστική συνιστώσα (και με προσδιοριστικές συνιστώσες) → δυσκολία ανίχνευσης
  - ***Ανίχνευση***
    - Εμπειρικοί τρόποι με διάφορα στατιστικά τεστ (Schuster, Walker and Fisher)
    - Αναλυτικοί τρόποι
      - Ανάλυση αυτοσυσχέτισης
      - Φασματική ανάλυση
    - Με τις μεθόδους ανίχνευσης βρίσκει κανείς τη βασική περίοδο λ, τη συχνότητα: $\omega = 2\pi / \lambda$, και όλες τις σημαντικές αρμονικές της j = 1,2,…δηλαδή τις περιοδικότητες με πολλαπλάσια συχνότητα της βασικής: $\omega j = 2\pi j / \lambda$, όπου j θετικός ακέραιος αριθμός, που συνυπάρχουν με τη βασική και που πρέπει και αυτές να αφαιρεθούν

Προσομοίωση Περιοδικότητας

- Μετά την ανίχνευση → ***προσομοίωση*** : υπολογισμός εύρος και φάση της κάθε περιοδικότητας
  - Υπολογιστικό ομοίωμα για προσομοίωση συνολικής περιοδικής συνιστώσας
  - Υπολογισμός παραμέτρων μιας περιοδικής συνιστώσας χρονοσειράς
    - Μέθοδος ελαχίστων τετραγώνων
      - π.χ. Συνημιτονοειδής συνάρτηση: Acosωt + Bsinωt Αναλυτικοί τρόποι (προτεινόμενοι)
      - Ανάλυση αυτοσυσχέτισης
      - Φασματική ανάλυση
    - Με τις μεθόδους ανίχνευσης βρίσκει κανείς τη βασική περίοδο *λ*, τη συχνότητα: $\omega = 2\pi / \lambda$, και όλες τις σημαντικές αρμονικές της j = 1,2,…δηλαδή τις περιοδικότητες με πολλαπλάσια συχνότητα της βασικής: $\omega j = 2\pi j / \lambda$, όπου j θετικός ακέραιος αριθμός, που συνυπάρχουν με τη βασική και που πρέπει και αυτές να αφαιρεθούν

Αναγνώριση Εμμονής σε υδρολογικές χρονοσειρές

Εμμονή

- Τα διαδοχικά στοιχεία μιας χρονοσειράς συνδέονται μεταξύ τους με κάποιο τρόπο που «εμμένει» και έχει σαν αποτέλεσμα τη μη τυχαιότητά της.
  - Τα υγρά χρόνια τείνουν να συμβούν κατά ομάδες και ομοίως τα ξηρά.

Έστω τώρα $x(t)$ είναι μια σειρά με T στοιχεία σε ίσα χρονικά διαστήματα από το χρόνο $t = 1$ έως $t = T$, π.χ. ετήσιοι όγκοι νερού εισρέοντες μέσα σε ταμιευτήρα.

Ορίζεται το άθροισμα:

$$x^*(t) = \sum_{u=1}^{t} x(u) \tag{4.11}$$

π.χ. το νερό που μπαίνει στον ταμιευτήρα για τα t χρόνια.

Υποτίθεται ότι το νερό αποφορτίζεται με σταθερό ρυθμό: $\frac{1}{T}x^*(T)$, έτσι, το ποσό που αποφορτίζεται σε t χρόνια θα είναι: $\frac{t}{T}x^*(T)$. Ορίζεται σαν περίσσεια ή έλλειμμα στο t χρόνο η σχέση:

$$D(t) = x^*(t) - \frac{t}{T}x^*(T) \tag{4.12}$$

και υπολογίζεται η:

$$\begin{aligned} &D(t+u) - D(t) - \frac{u}{s}\left[D(t+s) - D(t)\right] = \Delta(u) = \\ &x^*(t+u) - x^*(t) - \frac{u}{s}\left[x^*(t+s) - x^*(t)\right] \end{aligned} \tag{4.13}$$

Σαν εύρος στο διάστημα (t, s) ορίζεται:

$$\begin{aligned} &R_{t,s} = \max_{0 \le u \le s} \Delta(u) - \min_{0 \le u \le s} \Delta(u) = \\ &\max_{0 \le u \le s}\left\{x^*(t+u) - x^*(t) - \frac{u}{s}\left[x^*(t+s) - x^*(t)\right]\right\} - \\ &\min_{0 \le u \le s}\left\{x^*(t+u) - x^*(t) - \frac{u}{s}\left[x^*(t+s) - x^*(t)\right]\right\} \end{aligned} \tag{4.14}$$

Υπολογίζεται η σκέδαση του δείγματος από το χρόνο $(t+1)$ μέχρι $(t+s)$ ως εξής:

$$S_{t,s}^2 = s^{-1}\sum_{u=t+1}^{t+s} x^2(u) - \left[s^{-1}\sum_{u=t+1}^{t+s} x(u)\right]^2 \tag{4.15}$$

$$R_s = \frac{R_{t,s}}{S_{t,s}} = Ks^H \qquad H > 0.5 \tag{4.16}$$

Το H είναι τυπικά περίπου 0.72, ενώ για μια κανονική ανεξάρτητη σειρά:

$$E(R_s) = \sqrt{\frac{\pi}{2}}s^{0.5} \tag{4.17}$$

- Ο υπολογισμός της εμμονής γίνεται με εκτίμηση του δείκτη H με χάραξη των ζευγαριών $R_{t,s}$ και $S_{t,s}$ για $s \geq 3$ από το δείγμα. Η κλίση της ευθείας σε διπλό λογαριθμικό χαρτί είναι το H

Παράδειγμα υπολογισμού εμμονής

Πίνακας 4.1: Μηνιαίες και ετήσιες παροχές σε $\mu^3 \delta\lambda^{-1}$ του ποταμού Αχελώου στη θέση του Υ.Η.Σ. Κρεμαστών.

α/α	1	2	3	4	5	6	7	8	9	10	11	12	Ετ.
1	50.0	168.6	270.0	275.0	268.5	170.0	162.5	132.8	78.5	54.4	24.2	40.0	140.0
2	122.1	139.2	150.7	234.6	215.4	159.6	149.3	70.0	44.3	33.6	27.6	24.1	118.9
3	41.5	180.2	251.3	263.5	267.1	161.5	117.5	138.1	132.7	82.7	23.6	24.3	139.6
4	47.3	90.2	103.0	190.0	234.5	231.5	220.0	132.1	80.0	40.0	28.1	20.0	117.3
5	39.3	130.0	210.0	235.0	277.0	188.0	205.1	102.8	60.0	34.1	37.1	50.0	129.7
6	90.8	181.0	178.0	192.0	354.8	246.1	205.7	118.2	75.0	51.1	21.4	22.7	143.8
7	51.1	175.0	188.6	222.7	201.5	107.3	103.5	89.0	78.0	37.3	25.3	39.0	109.2
8	98.3	125.0	183.1	207.6	209.1	255.0	222.5	166.2	34.4	25.0	20.0	35.0	131.4
9	56.9	164.7	206.7	275.8	100.1	127.5	180.0	115.7	64.9	42.2	30.5	38.7	117.2
10	59.5	194.3	290.9	413.0	288.1	194.2	193.2	136.7	65.8	32.5	25.9	28.5	160.0
11	42.6	65.0	289.2	143.1	133.6	113.7	106.3	77.5	42.8	30.0	20.1	18.5	90.2
12	33.1	148.1	162.1	104.3	163.2	330.0	241.0	110.1	51.5	31.2	24.1	34.1	118.2
13	100.2	502.0	383.6	432.0	500.1	261.9	229.2	197.8	110.7	57.8	45.4	34.6	235.1
14	77.3	57.7	288.1	71.0	151.1	190.1	186.8	86.2	91.0	53.2	40.0	37.0	108.7
15	54.1	108.1	274.7	188.5	136.7	202.4	287.8	180.8	80.5	46.5	25.1	20.1	126.1
16	22.9	199.0	383.7	372.1	224.5	163.4	124.0	101.0	60.6	26.5	19.6	21.7	143.1
17	43.0	326.0	371.0	266.0	101.9	88.3	150.9	94.3	50.8	41.0	30.0	30.8	133.4
18	38.0	39.6	192.4	326.0	271.2	186.8	146.2	88.0	69.3	32.1	30.0	30.0	120.5
19	32.9	72.5	236.8	226.4	345.3	253.1	148.2	118.5	52.0	33.6	30.0	34.8	130.9
20	38.0	108.3	406.0	326.0	224.3	218.0	159.9	88.1	43.0	36.8	30.0	30.0	142.3
21	41.8	76.9	198.2	241.8	181.7	347.1	233.0	105.6	48.2	34.1	30.0	34.7	131.0
22	31.3	148.4	166.3	131.8	196.9	227.7	214.8	143.2	48.6	39.5	33.6	28.1	117.2
23	151.6	125.2	62.3	110.0	225.0	207.8	222.1	163.4	50.2	37.7	30.0	30.0	119.5
24	68.8	103.1	302.7	109.9	255.4	162.9	297.7	186.4	63.5	35.1	30.0	35.9	136.8
25	119.8	185.8	115.6	67.4	101.6	163.5	132.5	68.7	42.6	54.6	31.1	30.0	92.6
26	46.2	90.5	174.5	87.8	121.6	104.8	151.9	85.8	41.4	32.1	30.0	30.0	82.9
27	40.7	166.0	365.0	160.1	152.2	81.7	72.3	50.9	30.0	30.0	30.0	30.0	100.6

Ο υπολογισμός του εύρους $R_{t,s}$ και της τυπικής απόκλισης $S_{t,s}$ για κάθε επί μέρους δείγμα (t,s) εκτιμάται με βάση τις εξ.(4.14) και (4.15) αντίστοιχα. Σύμφωνα με την εξ.(4.16) ισχύει:

$$\log R_s = \log K + H \log s$$

Η κλίση της ευθείας παλινδρόμησης που αντιστοιχεί στα σημεία ($\log s$, $\log R_s$) δίνει την εκτίμηση H του συντελεστή Hurst.

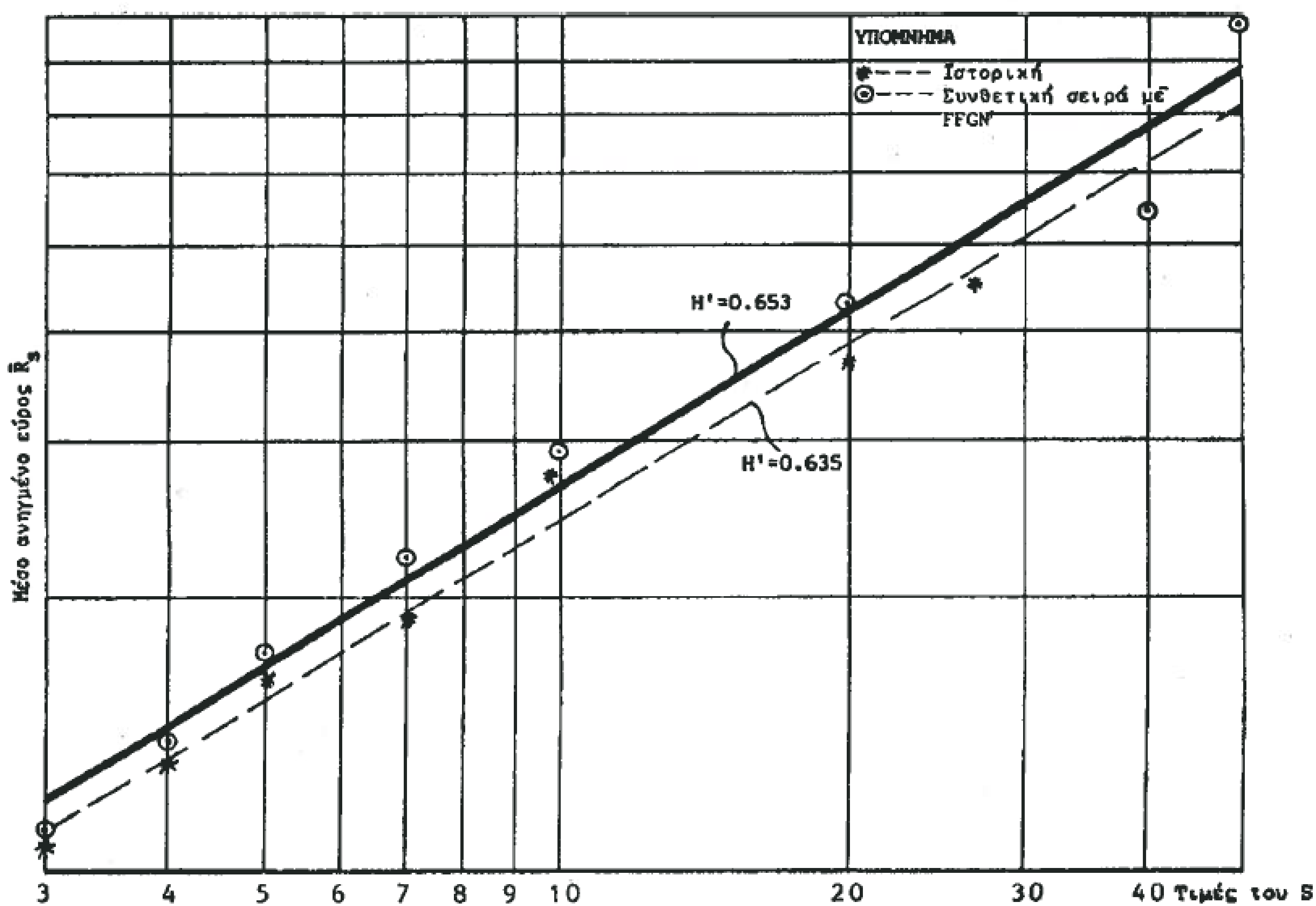
Για την όσο το δυνατό καλύτερη εκτίμηση του H τίθεται το ζήτημα της κατάλληλης εκλογής των τιμών των t και s. Όσον αφορά τους αρχικούς χρόνους t έχουν προταθεί διάφορες τιμές (Kottegoda, 1980), αλλά αναφέρονται σε μεγάλα δείγματα. Εδώ έχουν εκλεγεί τιμές του t ανά 2 έτη, δηλαδή t = 2,4,6,8,10,12,14,16,18,20, καθώς και η αρχή του δείγματος t=1. Οι 11 αυτές τιμές του t θεωρούνται ικανοποιητικές για τον υπολογισμό της μέσης τιμής $\overline{R}_s$. Τιμές $t > 20$ δεν ελήφθησαν γιατί θεωρήθηκε ότι δεν έχει νόημα το μέγεθος $t + s$ να υπερβαίνει το μέγεθος ολόκληρου του δείγματος (N = 27) για πολλές τιμές του s. Πάντως η εκλογή των χρόνων t ενέχει κάποιο βαθμό αυθαιρεσίας και υπόκειται σε παραπέρα μελέτη. Η εκλογή των μεγεθών μικρότερου δείγματος s υπαγορεύεται από την ανάγκη μείωσης της αβεβαιότητας στον υπολογισμό του H. Δεδομένου ότι η προς εκτίμηση συνάρτηση είναι εκθετική, παρουσιάστηκε το πρόβλημα να ληφθούν τόσο μικρές όσο και μεγάλες τιμές του s, ώστε η εκτίμηση της κλίσης της ευθείας παλινδρόμησης να είναι όσο το δυνατό ακριβέστερη. Έτσι ελήφθησαν οι μικρές τιμές του s (δηλαδή s= 3,4,5) να ισαπέχουν κατά μονάδα, ενώ για τις υπόλοιπες τιμές επιδιώχθηκε να ισαπέχουν περίπου σε λογαριθμική κλίμακα γεγονός που αποδείχθηκε ότι δίνει καλύτερα αποτελέσματα σύμφωνα με την ανάλυση ευαισθησίας που έγινε. Εκλέγονται αρχικά οι τιμές s = 3,4,5,7,10,20 καθώς και ολόκληρο το δείγμα s = N = 27. Η τιμή s = 2 δίνει πάντοτε για το ανηγμένο εύρος διακύμανσης τιμή ίση με $1/\sqrt{2}$ και συνεπώς δεν συνιστά πληροφορία από το δείγμα.

Πίνακας 4.2: Ανάλυση ευαισθησίας H στις τιμές s.

s	H	r
3,4,5,7,10,20	0,652471	0,989482
3,4,5,7,10,20,27	0,635292	0,993011
3,4,5,7,10,15,20,27	0,632827	0,993397
3,4,5,7,10,20,27	0,611297	0,992163

Πίνακας 4.3: Εκτίμηση ανηγμένου εύρους $R_{t,s}$ και μέσης τιμής του R_s

Τιμές του s	Τιμές του t											Μέση τιμή
	1	2	4	6	8	10	12	14	16	18	20	$\bar{R}_s$
3	1.155	1.146	1.021	1.080	1.092	1.059	1.151	1.021	1.044	1.015	1.031	1.074
4	.983	1.227	1.556	1.201	1.214	1.491	1.487	1.418	1.352	1.249	1.582	1.341
5	1.165	1.196	1.564	1.921	1.515	1.603	1.729	1.305	1.868	1.920	1.030	1.665
7	1.404	1.872	1.800	1.744	2.029	1.991	2.186	2.975	1.444	2.093	2.662	1.927
10	2.475	2.332	2.502	2.492	2.491	2.494	2.707	3.397	2.712	3.883	0.000	2.749
20	3.545	3.567	3.599	3.488	4.005	0.000	0.000	0.000	0.000	0.000	0.000	3.641
27	4.476	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4.476



Σχήμα 4.3. Υπολογισμός συντελεστή Hurst H για την ιστορική και τη συνθετική σειρά παροχών στα Κρεμαστά.

Αναγνώριση Τάσεων σε υδρολογικές χρονοσειρές

Στόχοι

- Λόγοι για αναγνώριση τάσεων/αλλαγών σε υδρολογικές χρονοσειρές.
- Οπτικά εργαλεία για αναγνωριστικές αναλύσεις δεδομένων (exploratory data analysis, EDA).
- Κατανόηση βασικών θεωριών στην στατιστική ανάλυση τάσεων και αλλαγών σε δεδομένα χρονοσειρών.
- Επίδειξη των μεθόδων με παραδείγματα στατιστικών μεθόδων.
- Παρουσίαση λογισμικών πακέτων για στατιστική ανάλυση χρονοσειρών.

Οι μέθοδοι που θα παρουσιαστούν μπορούν να εφαρμοστούν σε οποιαδήποτε χρονοσειρές, αλλά το ενδιαφέρον κυρίως εστιάζεται σε ετήσια δεδομένα απορροών, βροχόπτωσης και άλλων ετήσιων υδρομετεωρολογικών μεταβλητών.

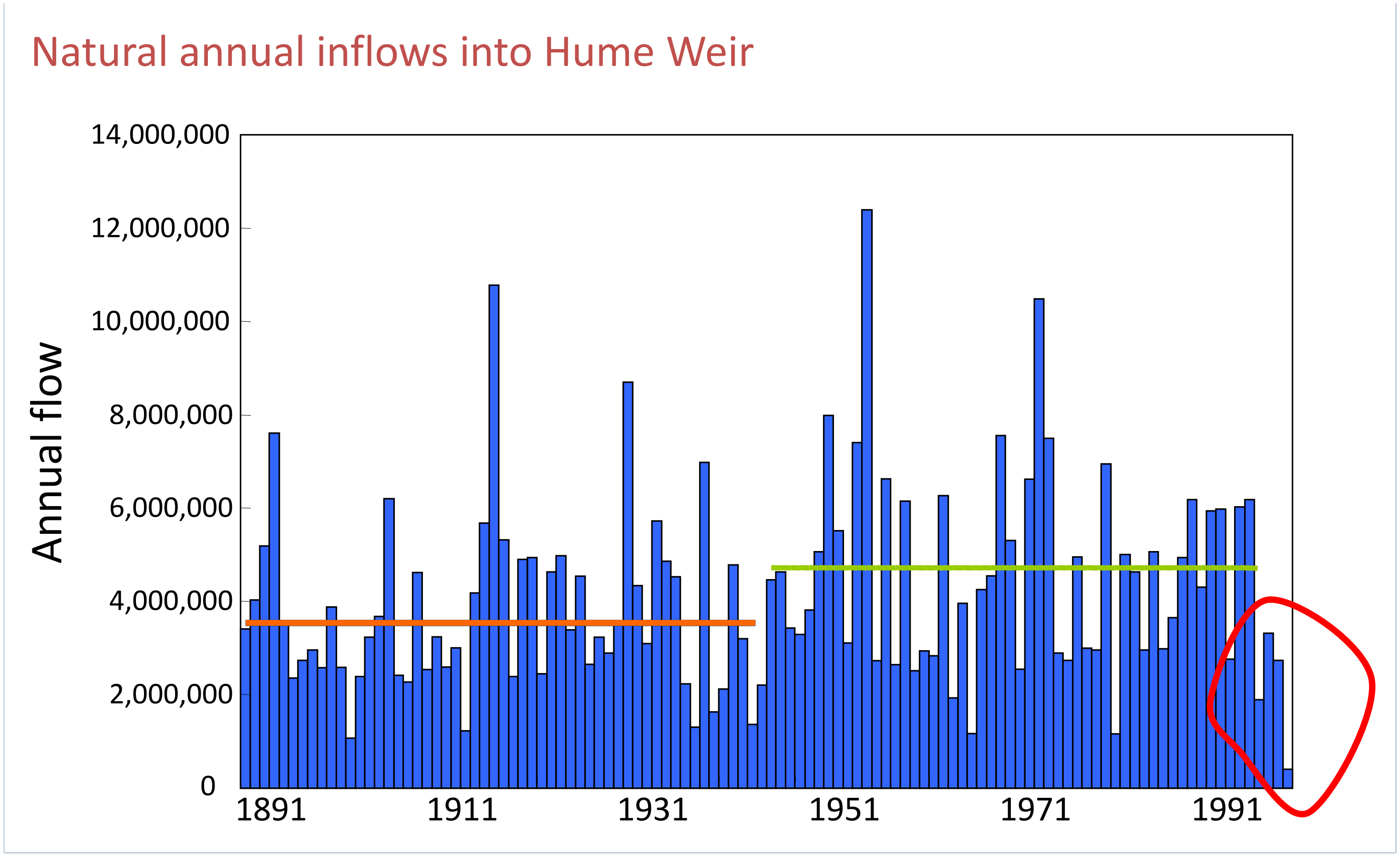
Γιατί ανιχνεύουμε τάσεις/αλλαγές σε υδρολογικές χρονοσειρές?

Πολλά υδατικά συστήματα έχουν σχεδιαστεί και λειτουργούν με βάση την υπόθεση της μονιμότητας (στασιμότητας) (stationary hydrology). Εάν η υπόθεση της μονιμότητας δεν είναι έγκυρη, τα τωρινά συστήματα μπορούν να αλλάξουν και να υπερ-σχεδιαστούν.

Οι τάσεις/αλλαγές σε υσρολογικές χρονοσειρές μπορούν να προκληθούν από:

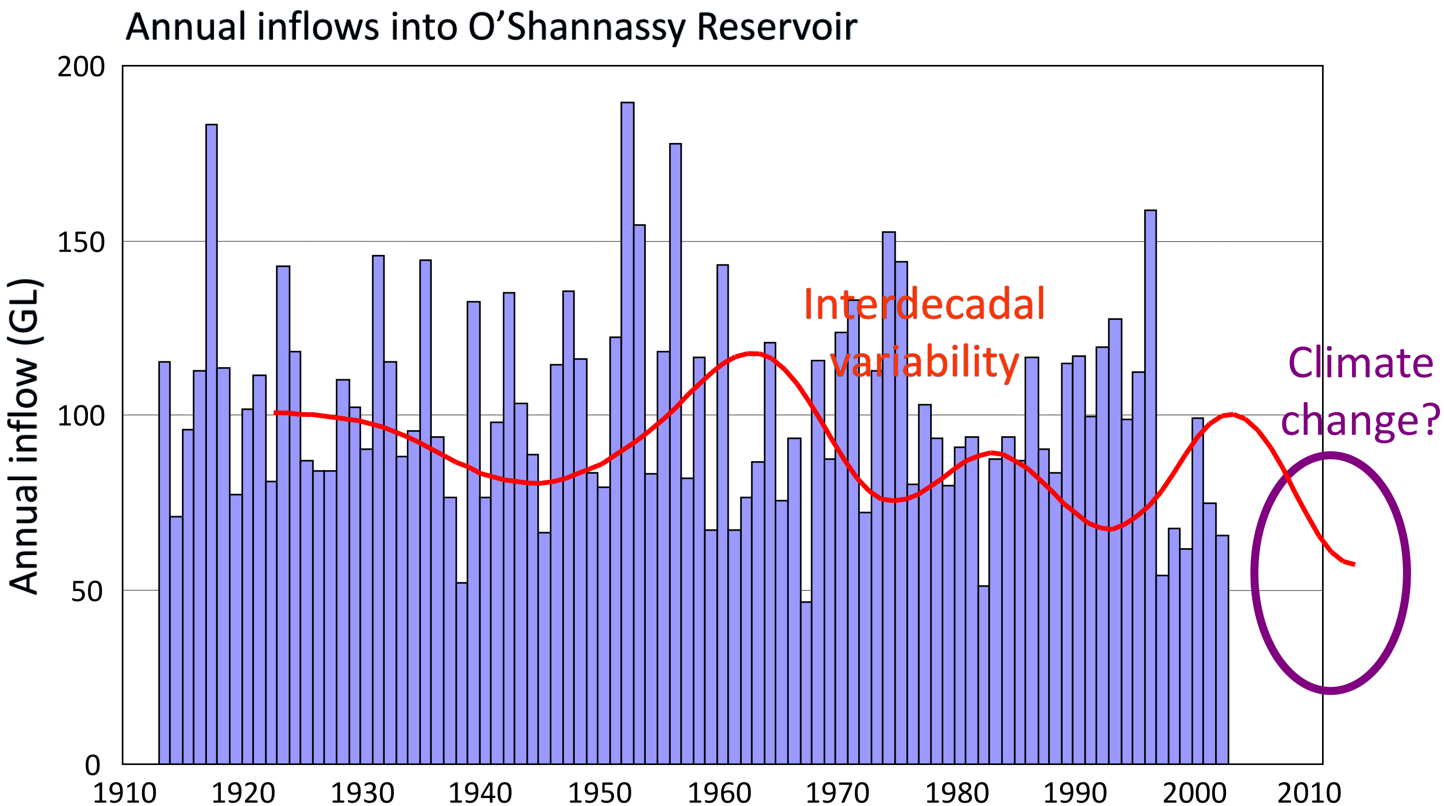
- Κλιματική αλλαγή λόγω αύξησης της συσγκέντρωσης των αερίων του θερμοκηπίου
- Αλλαγές χρήσεων γης (αστικοποίηση, καλλιεργούμενα γεωργικά είδη)
- Διαχειριστικές αλλαγές
- …

Κλιματική αλλαγή σε απορροές?



- Εμφανής αλλαγή (άλμα) στο μέσο της χρονοσειράς (κλιματική αλλαγή?)
- Πολλαπλά άλματα στο μέσο (inter-decadal variability?)

Το υδροκλίμα πάντα αλλάζει σε διάφορες χρονικές κλίμακες



- Seasonal
- Inter-annual
- **ENSO (3-7 years)**
- Interdecadal (20-30 years)
- "Climate change"

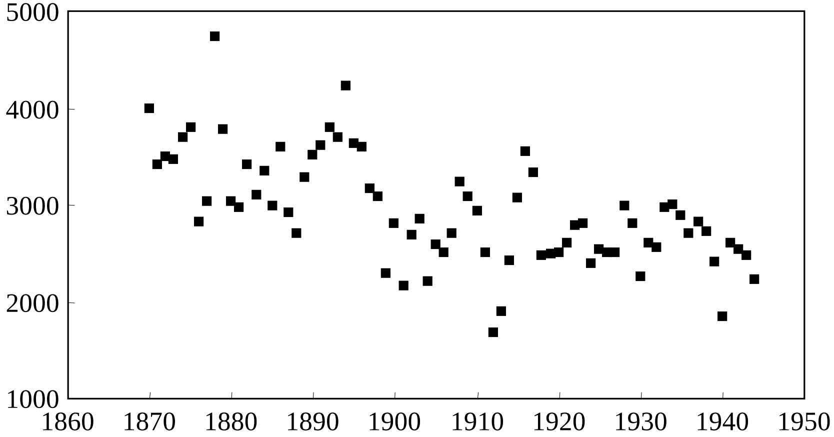
Exploratory data analysis (EDA)

- Exploratory data analysis (EDA) (Τεχνικές αναγνώρισης και ανάλυσης δεδομένων) χρησιμοποιεί γραφήματα για εξερεύνηση, κατανόηση και παρουσίαση δεδομένων. EDA είναι μία επαναληπτική διαδικασία όπου γραφήματα σχεδιάζονται και φιλτράρονται ώστε οι σημαντικές ιδιότητες και χαρακτηριστικά των δεδομένων να αναγνωρίζονται.

- EDA είναι βασικό στοιχείο οποιασδήποτε στατιστικής ανάλυσης. Η EDA αναδεικνύει τα γνωρίσματα των δεδομένων πιο πολύ από τα περιληπτικά στατιστικά (summary statistics) ή τα στατιστικά επίπεδα εμπιστοσύνης (statistical significance levels). Αυτό γιατί το ανθρώπινο μυαλό και το οπτικό σύστημα είναι πολύ ικανό στην αναγνώριση και επεξήγηση προτύπων. Μία σωστή EDA μπορεί να απαλείψει την ανάγκη για τυπικές στατιστικές αναλύσεις.

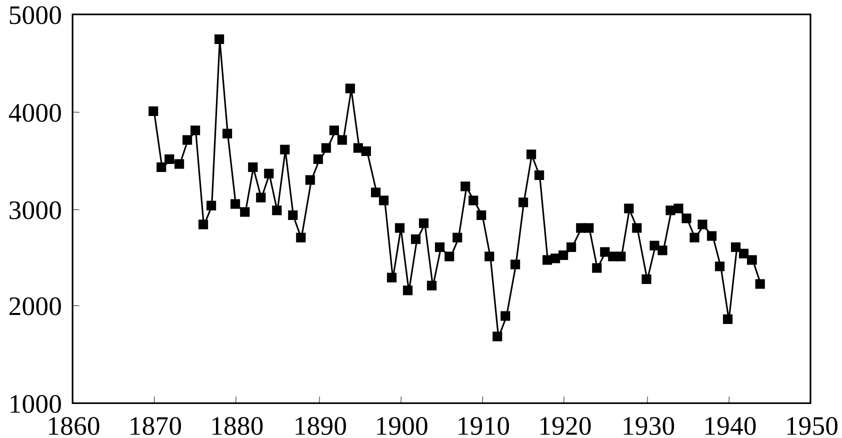
- EDA μπορεί να αναγνωρίσει έκτοπες ή εξωκείμενες τιμές δεδομένων, ανεξαρτησία/ αυτοσυσχέτιση και αλλαγές σε θέσεις δεδομένων

Παράδειγμα γραφικής ανάλυσης χρονοσειρών

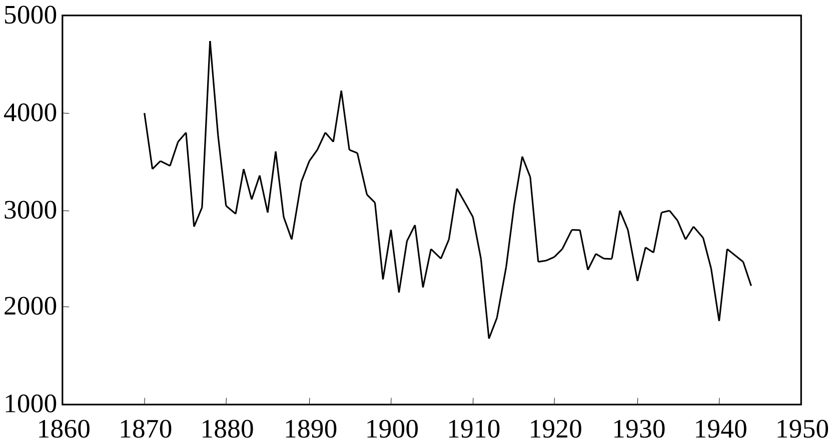


Scatter plot

Annual River Nile flows
at Aswan in cumecs

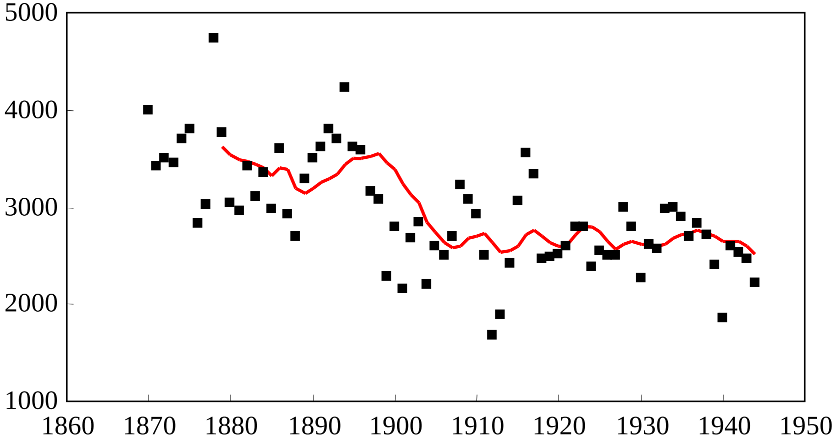


Line plot with data points

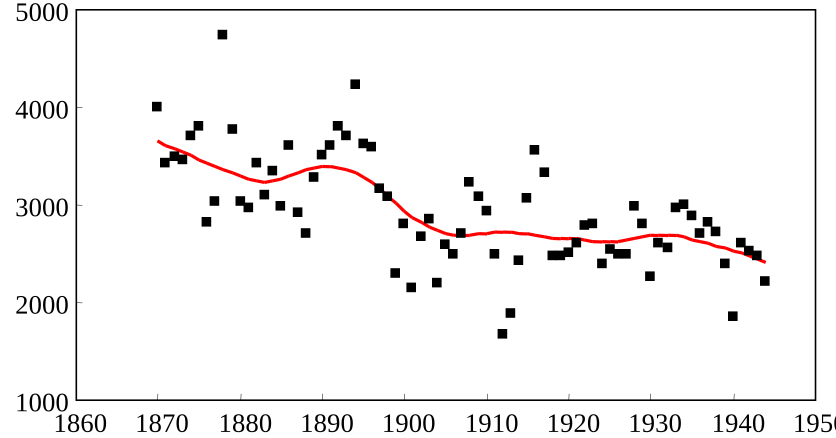


Line plot without data points

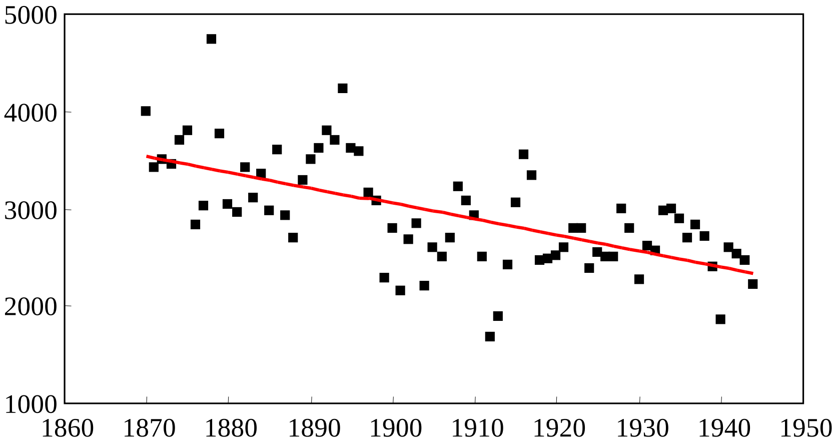
Παραδείγματα εφαρμογής τάσεων σε χρονοσειρές



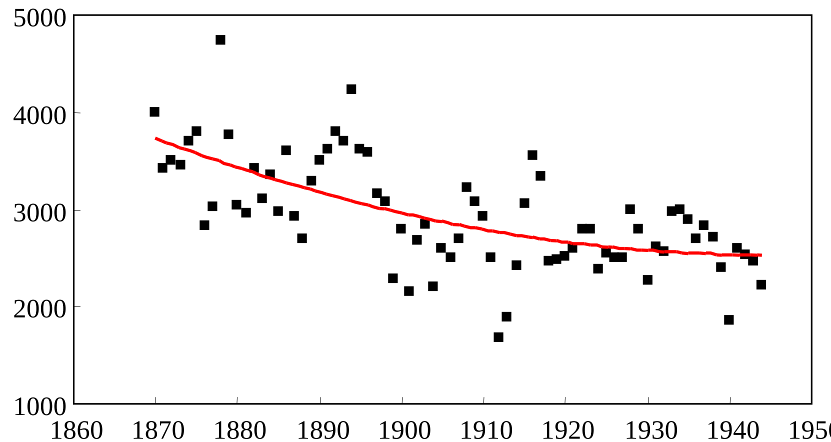
10-point moving average



Lowess smooth



Linear regression



Polynomial regression

Λογισμικά για γραφική ανάλυση δεδομένων και ανάλυση χρονοσειρών

- Excel
- Mathematica
- MATLAB
- MINITAB
- SAS
- SPlus
- SPSS
- Systat
- R-project

Στατιστικές Μέθοδοι
για αναγνώριση τάσεων και αλλαγών σε χρονοσειρές

Αλλαγές σε χρονοσειρές μπορούν να συμβούν σε σταθερό χρόνο (μία τάση), ξαφνικά (a step-change) ή σε πιο σύνθετες μορφές. Μπορούν να επηρεάσουν το μέσο, τη διάμεσο, τη διασπορά ή άλλα στατιστικά χαρακτηριστικά των δεδομένων.

Παρουσίαση στατιστικών μεθόδων για αναγνώριση τάσης (trend), απότομης αλλαγής βήματος (step change), διαφορές στο μέσο/διάμεσο ανάμεσα σε δύο περιόδους και τυχαιότητας (randomness) σε υδρολογικές χρονοσειρές

Οι στατιστικές μέθοδοι βασίζονται στο WMO/UNESCO WCP Expert Workshop on “Detecting Trend and Other Changes in Hydrological Data” and the CRCCH “Hydrological Recipes”.

Kundzewicz, Z.W. and Robson, A. (Editors) (2000) Detecting Trend and Other Changes in Hydrological Data. World Climate Program – Water, WMO/UNESCO, WCDMP-45, WMO/TD 1013, Geneva, 157 pp.

Grayson, R.B., Argent, R.M., Nathan, R.J., McMahon, T.A. and Mein, R. (1996) Hydrological Recipes: Estimation Techniques in Australian Hydrology. Cooperative Research Centre for Catchment Hydrology, Australia, 125 pp.

Parametric and non-parametric tests

- Most statistical tests assume that the time series data are independent and identically distributed.

- Parametric tests also assume that the time series data and the errors (deviations from the trend) follow a particular distribution. Most parametric tests assume that the data are normally distributed. Parametric tests are useful as they also quantify the change in the data (e.g., change in mean or gradient of trend). Parametric tests are generally more powerful than non-parametric tests.

- Non-parametric tests are generally distribution-free. They detect trend/change, but do not quantify the size of the trend/change. They are very useful because most hydrologic time series data are not normally distributed.

Statistical tests

Tests for trend

- Mann-Kendall (non-parametric)
- Spearman's Rho (non-parametric)
- Linear Regression (parametric)

Tests for step change in mean/median

- Distribution Free CUSUM (non-parametric)

Tests for difference in mean/median in two different data periods

- Rank-Sum (non-parametric)
- Student's t-test (parametric)

Tests for randomness

- Median Crossing (non-parametric)
- Turning Points (non-parametric)
- Rank Difference (non-parametric)
- Autocorrelation (parametric)

Cautionary Words

- Must have good data and must understand data (via exploratory data analysis).
- Must understand statistical test and the assumptions.
- A statistical test provides evidence, not proof.
- Significance is not the same as importance (e.g., a change may be detected, but the size of the change may be so small that it is of no importance).
- If H_0 is rejected, the reason for the trend/change must be investigated.

Tests for trend

Mann-Kendall (non-parametric)

- This method tests whether there is a trend in the time series data. It is a non-parametric test.
  The n time series values (X_1, X_2, X_3, ….., X_n) are replaced by their relative ranks (R_1, R_2, R_3, ….., Rn) (starting at 1 for the lowest up to n).

The test statistic S is:

$$S = \sum_{i=1}^{n-1} \left[\sum_{j=i+1}^{n} \operatorname{sgn}(R_j - R_i) \right]$$

where sgn(x) = 1 for x > υ
sgn(x) = 0 for x = 0
sgn(x) = -1 for x < 0

If the null hypothesis Ho is true, then S is approximately normally distributed with: $\mu = 0$
$\sigma = n\,(n-1)\,(2n+5)\,/\,18$

The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):
$z = |\,S\,| \,/\, \sigma^{0.5}$

A positive value of S indicates that there is an increasing trend and vice versa.

Tests for trend

Spearman's Rho Test

- This is a rank-based test that determines whether the correlation between two variables is significant. In trend analysis, one variable is taken as the time itself (years) and the other as the corresponding time series data.
- Like the Mann-Kendall Test, the n time series values are replaced by their ranks.
- The test statistic ρ_s is the correlation coefficient, which is obtained in the same way as the usual sample correlation coefficient, but using ranks:

$$\rho_s = S_{xy} / (S_x S_y)^{0.5}$$

where

$$S_x = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$S_y = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

For large samples, the quantity ρs $(n-1)^{0.5}$ is approximately normally distributed with mean of 0 and variance of 1 (critical test statistic values for various significance levels can be obtained from normal probability tables).

Tests for trend

Linear Regression Test

- This is a parametric test that assumes that the data are normally distributed. It tests whether there is a linear trend by examining the relationship between time (x) and the variable of interest (y).
- The regression gradient is estimated by:

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

and the intercept is estimated as:

$$a = \bar{y} - b\,\bar{x}$$

The test statistic S is:

$$S = b / \sigma$$

$$\text{where } \sigma = \sqrt{\frac{12\sum_{i=1}^{n}(y_i - a - bx_i)}{n(n-2)(n^2-1)}}$$

The test statistic S follows a Student-t distribution with n-2 degrees of freedom under the null hypothesis (critical test statistic values for various significance levels can be obtained from Student's t statistic tables).

The linear regression test assumes that the data are normally distributed and that the errors (deviations from the trend) are independent and follows the same normal distribution with zero mean.

Tests for step change in mean/median

Distribution Free CUSUM Test

- This method tests whether the means in two parts of a record are different (for an unknown time of change). It is a non-parametric test (distribution free).
- Given a time series data (x_1, x_2, x_3, ….., x_n), the test statistic is defined as:

$$V_k = \sum_{i=1}^{k} \operatorname{sgn}(x_i - x_{median}) \qquad k = 1,2,3,\ldots\ldots,n$$

where sgn(x) = 1 for x > 0

sgn(x) = 0 for x = 0

sgn(x) = -1 for x < 0

x_{median} is the median value of the x_i data set.

The distribution of V_k follows the Kolmogorov-Smirnov two-sample statistic (KS = (2/n) max|Vk|) with the critical values of max|Vk| given by:

α = 0.10	1.22√n
α = 0.05	1.36√n
α = 0.01	1.63√n

A negative value of V_k indicates that the latter part of the record has a higher mean than the earlier part and vice versa.

Tests for difference in mean/median in two different data periods

Rank-Sum Test

- This method tests whether the medians in two different periods are different. It is a nonparametric test.
- To compute the rank-sum test statistic:

  Rank all the data, from 1 (smallest) to N (largest). In the case of ties (equal data values), use the average of ranks;

  Compute a statistic S as the sum of ranks of the observations in the smaller group (the number of observations in the smaller group is denoted as n, and the number of observations in the larger group is denoted as m); and

  Compute the theoretical mean and standard deviation of S under H_o for the entire sample

$$\mu = n\,(N + 1) / 2$$
$$\sigma = [n\,m\,(N + 1) / 12]^{0.5}$$

- The standardised form of the test statistic Z_{rs} is computed as:

$Z_{rs} = (S - 0.5 - \mu) / \sigma$	if	$S > \mu$
$Z_{rs} = 0$	if	$S = \mu$
$Z_{rs} = \lvert S + 0.5 - \mu \rvert / \sigma$	if	$S < \mu$

- Z_{rs} is approximately normally distributed, and the critical test statistic values for various significance levels can be obtained from normal probability tables.

Tests for difference in mean/median in two different data periods

Student's t Test

- This method tests whether the means in two different periods are different. The test assumes that the data are normally distributed.
- The Student's t test statistic t is (critical test statistic values for various significance levels can be obtained from Student's t statistic tables):

$$t = \frac{(\overline{x} - \overline{y})}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

where x and y are the means of the first and second periods respectively, and m and n are the number of observations in the first and second periods respectively, and S is the sample standard deviation (of the entire m and n observations).

Tests for randomness

Median Crossing Test

- The n time series values are replaced by 0 if $x_i < x_{median}$ and by 1 if $x_i > x_{median}$.
- If the time series data come from a random process, then m (the number of times 0 is followed by 1 or 1 is followed by 0) is approximately normally distributed with:

$$\mu = (n - 1) / 2$$
$$\sigma = (n - 1) / 4$$

- The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = | (m - \mu) | / \sigma^{0.5}$$

Tests for randomness

Turning Points Test

- The n time series values are assigned 1 if $x_{i-1} < x_i > x_{i+1}$ or $x_{i-1} > x_i < x_{i+1}$, otherwise they are assigned as 0.
- The number of times 1 appears (m*) is approximately normally distributed with:

$$\mu = 2\,(n - 2) / 3$$

$$\sigma = (16n - 29) / 90$$

- The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = |\, m^* - \mu \,| / \sigma^{0.5}$$

Tests for randomness

Rank Difference Test

- The n time series values are replaced by their relative ranks starting at 1 for the lowest up to n.
- The statistic U is the sum of the absolute rank differences between successive ranks:

$$U = \sum_{i=2}^{n} \left| R_i - R_{i-1} \right|$$

For large n, U is normally distributed with:

$\mu = (n + 1)(n - 1) / 3$

$\sigma = (n - 2)(n + 1)(4n - 7) / 90$

- The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$z = | U - \mu | / \sigma^{0.5}$

Tests for randomness

Autocorrelation Test

- The lag-one autocorrelation coefficient is calculated as:

$$r_1 = \frac{\left[\sum_{i=1}^{n-1}(x_i - \overline{x})(x_{i+1} - \overline{x})\right]}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]}$$

If the time series data come from a random process, then the expected value and variance of r_1 are:

$$E(r_1) = -1 / n$$

$$Var(r_1) = (n^3 - 3n^2 + 4) / [n^2 (n^2 - 1)]$$

- The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = | r_1 - E(r_1) | / Var(r_1)^{0.5}$$

Βιβλιογραφία

Μιμίκου, Μ.Α., 2006. «Τεχνολογία Υδατικών Πόρων», 3η Έκδοση, Α. Παπασωτηρίου & Σία ΟΕ.

Kundzewicz, Z.W. and Robson, A. (Editors) (2000) Detecting Trend and Other Changes in Hydrological Data. World Climate Program – Water, WMO/UNESCO, WCDMP-45, WMO/TD 1013, Geneva, 157 pp.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Θεσσαλίας**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης