



ΑΡΙΣΤΟΤΕΛΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΟΝΙΚΗΣ

Υπολογιστική Εργασία  
«Ανάλυση χρονοσειρών από θερμοπίδακα»

# Χρονοσειρές

στο MATLAB

Φεβρουάριος 2023



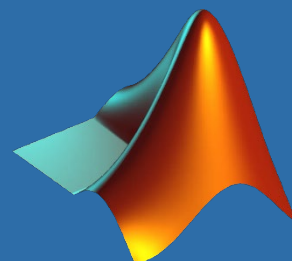
ΑΕΜ: 10265

ΣΤΑΥΡΟΣ ΚΙΗΓΜΑΣ

ΑΕΜ: 10120

ΙΩΑΝΝΗΣ-ΜΑΡΙΟΣ  
ΚΑΡΑΓΙΑΝΝΗΣ

Φοιτητές ΤΗΜΜΥ ΑΠΘ



Πρώτο Στάδιο - Γραμμική ανάλυση για τα έτη 1989,2000,2011 .....	3
1).....	3
2).....	6
Δεύτερο Στάδιο - Γραμμική και Μη-γραμμική ανάλυση για το έτος που σας αντιστοιχεί .....	8
1.Σχεδιαγράμματα χρονοσειρών .....	8
2.Στατιστικός έλεγχος ανεξαρτησίας .....	9
3.Εκτίμηση υστέρησης $\tau$ που δίνει το κριτήριο της αμοιβαίας πληροφορίας .....	9
4.Εκτίμηση της διάστασης εμβύθισης $m$ που σας δίνει το κριτήριο των ψευδών κοντινότερων γειτόνων (false nearest neighbors) χρησιμοποιώντας ως υστέρηση αυτή που βρήκατε παραπάνω .....	10
5. Πρόβλεψη με τοπικό μοντέλο μέσου όρου και τοπικό γραμμικό μοντέλο για υστέρηση $\tau$ και διάσταση εμβύθισης $m$ που εκτιμήσατε παραπάνω (και τον ίδιο αριθμό γειτονικών σημείων).....	11
6. Σχεδιάγραμμα των προβλέψεων ενός βήματος μπροστά. Σύγκριση με τις αντίστοιχες προβλέψεις του γραμμικού μοντέλου (από το πρώτο μέρος της εργασίας). Σύγκριση του γραμμικού και μη-γραμμικού μοντέλου ως προς την πρόβλεψη για ένα βήμα μπροστά με βάση το $nrmse$ .....	13
7. Εκτίμηση της διάστασης συσχέτισης για διαστάσεις $m=1,...,10$ , χρησιμοποιώντας την υστέρηση από το κριτήριο της αμοιβαίας πληροφορίας με κατάλληλα σχήματα .....	14
Σχήμα 1. Αυτοσυσχέτιση eruption_1989 .....	3
Σχήμα 2. Αυτοσυσχέτιση eruption_2000 .....	4
Σχήμα 3. Αυτοσυσχέτιση eruption2011 .....	5
Σχήμα 4. Μερική Αυτοσυσχέτιση eruption1989.....	6
Σχήμα 5. Σχεδιάγραμμα eruption2005 για $n=4915$ .....	8
Σχήμα 6. Σχεδιάγραμμα eruption2005 για $n=500$ .....	8
Σχήμα 7. Αμοιβαία Πληροφορία για χρονοσειρά $n=4194$ .....	9
Σχήμα 8.Αμοιβαία Πληροφορία για χρονοσειρά $n=500$ .....	10
Σχήμα 9b. Nearest Neighbor για χρονοσειρά $n=500$ .....	10
Σχήμα 10. Μοντέλο τοπικού μέσου όρου για $n=4914$ .....	11
Σχήμα 11. Μοντέλο τοπικού μέσου όρου για $n=500$ .....	12
Σχήμα 12. Προβλέψεις και για τα 2 μοντέλα για $n=4914$ .....	13
Σχήμα 13. Προβλέψεις και για τα 2 μοντέλα για $n=500$ .....	13
Σχήμα 14. Διάγραμμα εκτίμησης διάστασης συσχέτισης για $n=4914$ .....	14
Σχήμα 15. Διάγραμμα εκτίμησης διάστασης συσχέτισης για $n=500$ .....	14
Σχήμα 16. Correlation integral για $n=4914$ .....	15
Σχήμα 17. Correlation integral για $n=500$ .....	15

Σχήμα 18. Διάγραμμα κλίσης για $n=500$ .....	15
Σχήμα 19. Διάγραμμα κλίσης για $n=4914$ .....	15
Σχήμα 20. Αυτοσυσχέτιση για δείγματα $n=500$ .....	16
Σχήμα 21. Αυτοσυσχέτιση για δείγματα $n=4914$ .....	16

Πίνακας 1 NRMSE για προβλέψεις τοπικού γραμμικού μοντέλου.....	12
--	----

## Πρώτο Στάδιο - Γραμμική ανάλυση για τα έτη 1989,2000,2011

Μας ζητήθηκε να εξετάσουμε αν το σύστημα που δημιουργεί τις εξάρσεις, και ειδικότερα την αναμονή των εξάρσεων, παραμένει το ίδιο στις διαφορετικές περιόδους, και συγκεκριμένα, για το 1989, το 2000 και το 2011 (διαφορά κατά 11 έτη).

Αρχικά επιλέξαμε τυχαία, για τις χρονοσειρές **eruption2000** και **eruption2011**, 298 δείγματα (κάθε φορά που τρέχει ο κώδικας τα δείγματα είναι διαφορετικά) ώστε οι χρονοσειρές να είναι ίδιου μήκους.

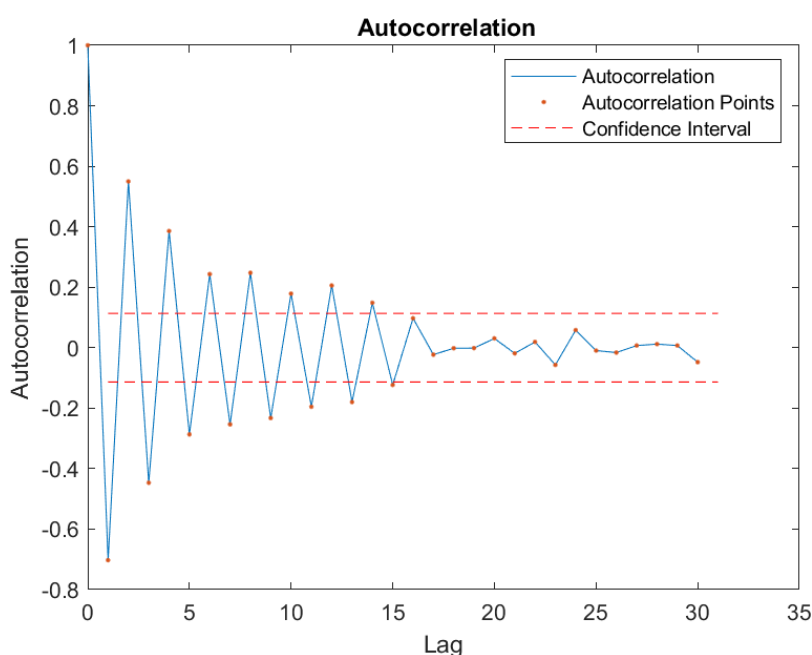
1)

Για να καταλάβουμε αν μία χρονοσειρά είναι λευκός θόρυβος πρέπει να δούμε την αυτοσυσχέτισή της.

Έχοντας χρονοσειρές 298 δειγμάτων επιλέξαμε να εξετάσουμε την αυτοσυσχέτιση για το 10% των δειγμάτων, δηλαδή για 30 υστερήσεις.

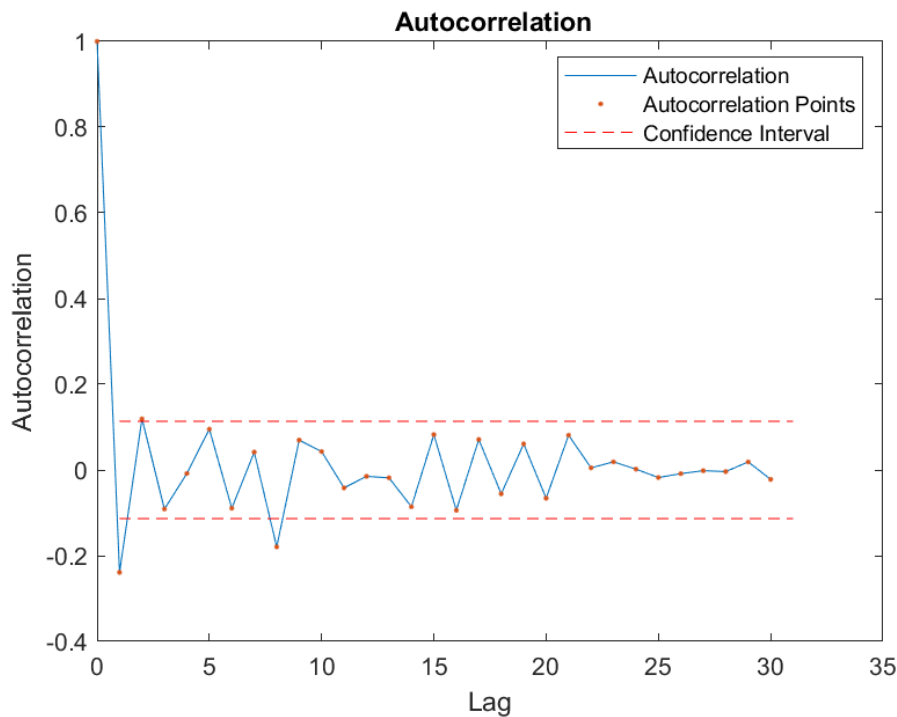
Θεωρήσαμε πως το επίπεδο σημαντικότητας,  $\alpha = 0.05$  και άρα, το διάστημα εμπιστοσύνης να είναι το  $R = \left[ \frac{-1.96}{\sqrt{298}}, \frac{1.96}{\sqrt{298}} \right]$

Όπως βλέπουμε από το παρακάτω γράφημα η χρονοσειρά **eruption1989** δεν είναι λευκός θόρυβος γιατί υπάρχουν σημαντικές αυτοσυσχετίσεις (έξω από το διάστημα εμπιστοσύνης).



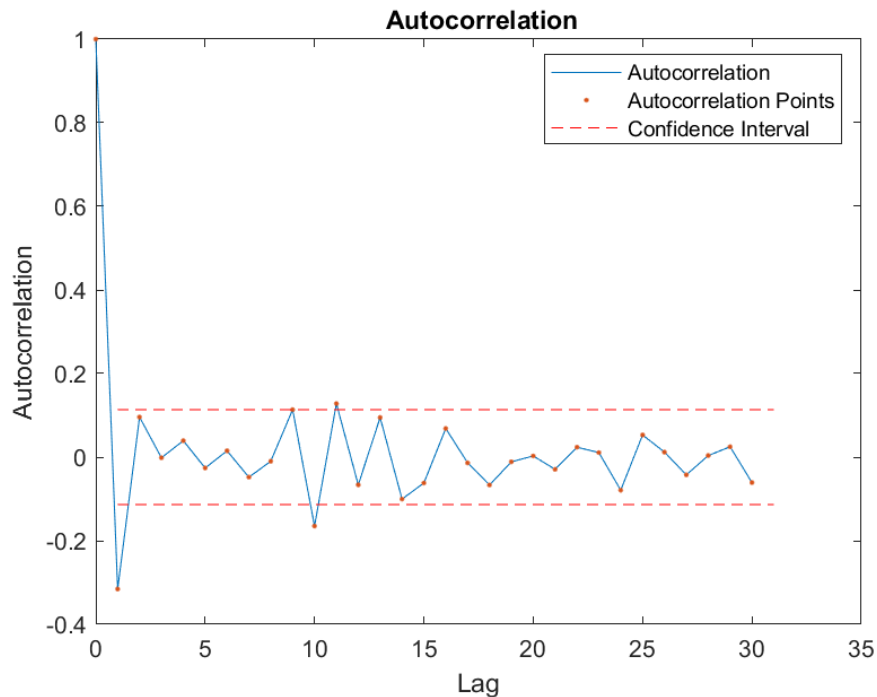
Σχήμα 1. Αυτοσυσχέτιση eruption\_1989

Συνεχίζοντας στην αυτοσυσχέτιση της χρονοσειράς **eruption2000** παρατηρούμε πως μόνο η 1<sup>η</sup> τιμή βρίσκεται εκτός του διαστήματος εμπιστοσύνης ενώ η 2<sup>η</sup> είναι οριακά σημαντική. Η 8<sup>η</sup> φαίνεται να είναι έξω από τα όρια σημαντικότητας, αλλά αφού προηγούνται 5 ασήμαντες αυτοσυσχετίσεις, δε τη λαμβάνουμε υπόψιν. Βάσει των παραπάνω η **eruption2000** είναι λευκός θόρυβος



Σχήμα 2. Αυτοσυσχέτιση eruption\_2000

Τέλος, κοιτώντας την αυτοσυσχέτιση της **eruption2011** βλέπουμε επίσης πως είναι λευκός θόρυβος, μια και μόνο η 1<sup>η</sup> αυτοσυσχέτιση είναι στατιστικά σημαντική ενώ οι υπόλοιπες όχι (για τους λόγους που αναφέρθηκαν και για την **eruption2000**)



Σχήμα 3. Αυτοσυσχέτιση eruption2011

**Παρατήρηση 1:** Τα αποτελέσματα για τις 3 χρονοσειρές δεν είναι ίδια το οποίο φανερώνει ότι ο «μηχανισμός» δημιουργίας των εξάρσεων είναι διαφορετικός σε αυτές τις περιόδους.

**Παρατήρηση 2:** Επαναλάβουμε για διαφορετικά διαστήματα, 298 δειγμάτων, των χρονοσειρών **eruption2000**, **eruption2011** και τα συμπεράσματά μας ήταν τα ίδια.

**Παρατήρηση 3:** Εκτελέσαμε την παραπάνω διαδικασία και για  $n = 298$ , ώστε να ελέγξουμε αν υπόθεσή μας, πως δε χρειαζόμαστε όλα τα δείγματα (αλλά μόνο το 10%) είναι σωστή. Εν τέλει παρατηρήσαμε πως τα συμπεράσματα ήταν ίδια.

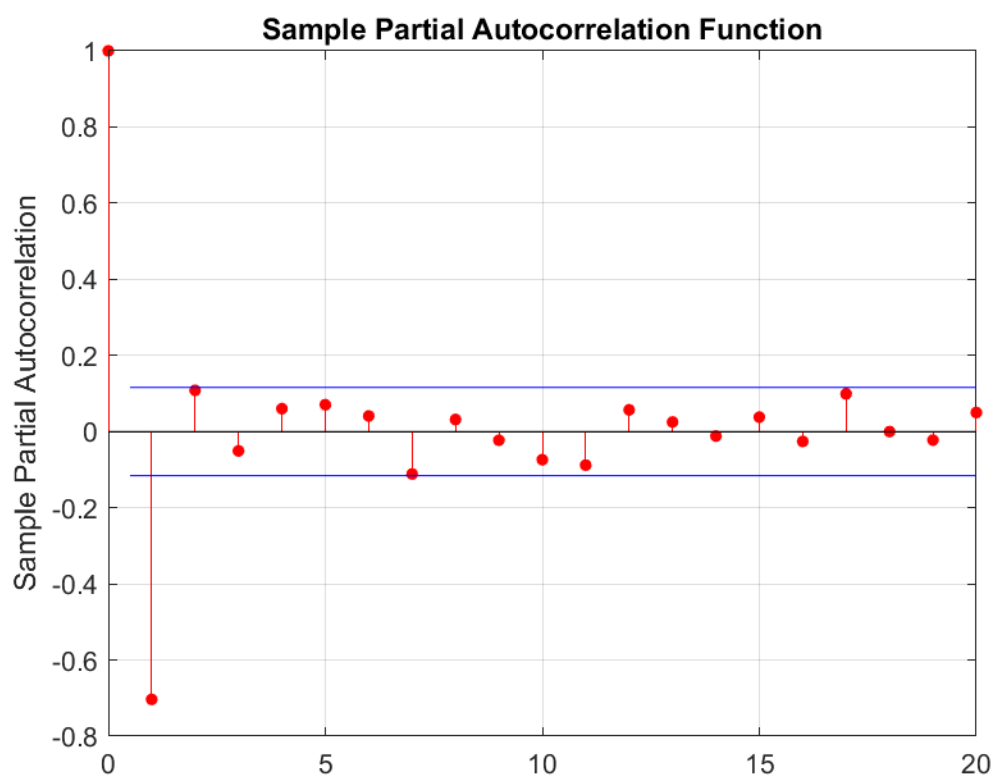
2)

Σε συνέχεια του ερωτήματος 1) επιτρέπεται να κάνουμε προσαρμογή μοντέλου και πρόβλεψη μόνο για τη χρονοσειρά **eruption1989**.

Για την εύρεση κατάλληλου μοντέλου θεωρήσαμε σωστό να δούμε και την μερική αυτοσυσχέτιση της εν λόγω χρονοσειράς.

Χρησιμοποιήσαμε τη συνάρτηση **parcorr()** (γραμμή 81 στον κώδικα) που βρίσκεται στο πακέτο **econometric modeler** του MATLAB.

Αυτό που παρατηρούμε είναι πως έχουμε μόνο την 1<sup>η</sup> μερική αυτοσυσχέτιση να είναι στατιστικά σημαντική ενώ η 2<sup>η</sup> να είναι οριακά σημαντική. Σύμφωνα με τη θεωρία, μπορούμε να προσαρμόσουμε ένα γραμμικό αυτοπαλίνδρομο μοντέλο τάξης 1 (AR(1)) στη χρονοσειρά μας.



Σχήμα 4. Μερική Αυτοσυσχέτιση eruption1989

Στη συνέχεια του ερωτήματος χρησιμοποιήσαμε τη συνάρτηση **fitARMA()** ( $p=1, q=0$ ) για να προβλέψουμε για 1, 2 και 3 χρονικά βήματα μπροστά. Για να δούμε αν η πρόβλεψή μας είναι καλή κοιτάμε την κανονικοποίηση της ρίζας του μέσου τετραγωνικού σφάλματος (nrmse). Οι τιμές αυτές είναι:

Βήμα 1: nrmse=0.7111

Βήμα 2: nrmse=0.8358

Βήμα 3: nrmse=0.8985

Αυτό υποδηλώνει πως η προβλέψεις μας δεν είναι καλές αφού οι τιμές του nrmse είναι κοντά στο 1.



Δεύτερο Στάδιο - Γραμμική και Μη-γραμμική ανάλυση για το έτος που σας αντιστοιχεί

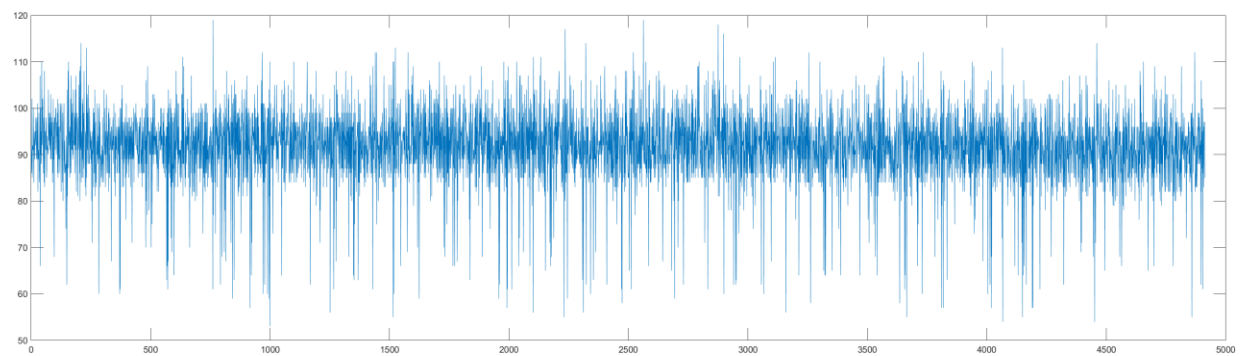
Το έτος που μας αντιστοιχεί είναι :  $(10120 + 10265) \bmod 10 = 5$

Άρα η χρονοσειρά ***eruption2005***.

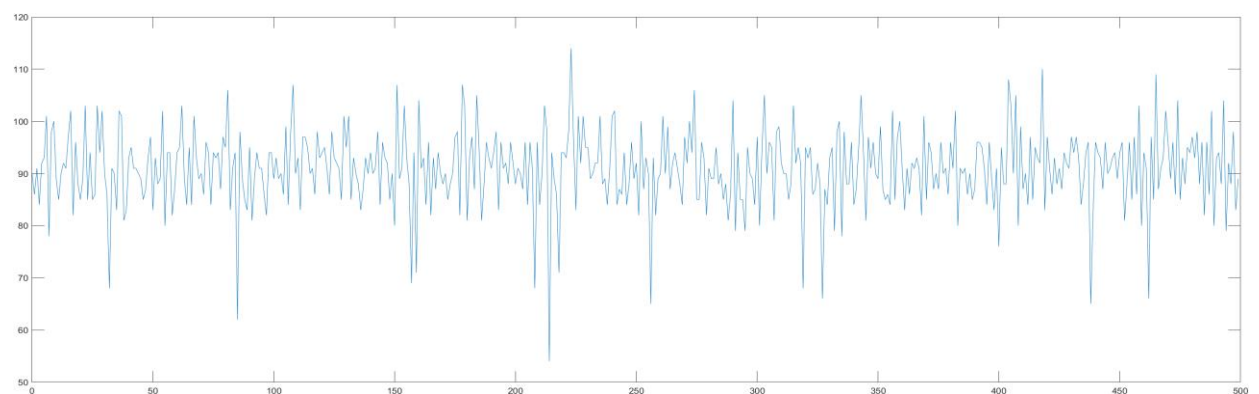
Το αρχείο του κώδικα για αυτό το στάδιο είναι το ***Code\_2.m***.

Όπως ζητείται σχηματίσαμε δύο χρονοσειρές. Την ***original\_array\_2005*** (χρονοσειρά με όλες τις παρατηρήσεις) και την ***part\_of\_array\_2005*** (χρονοσειρά με τις 500 παρατηρήσεις)

### 1. Σχεδιαγράμματα χρονοσειρών



Σχήμα 5. Σχεδιάγραμμα eruption2005 για n=4915



Σχήμα 6. Σχεδιάγραμμα eruption2005 για n=500

## 2. Στατιστικός έλεγχος ανεξαρτησίας

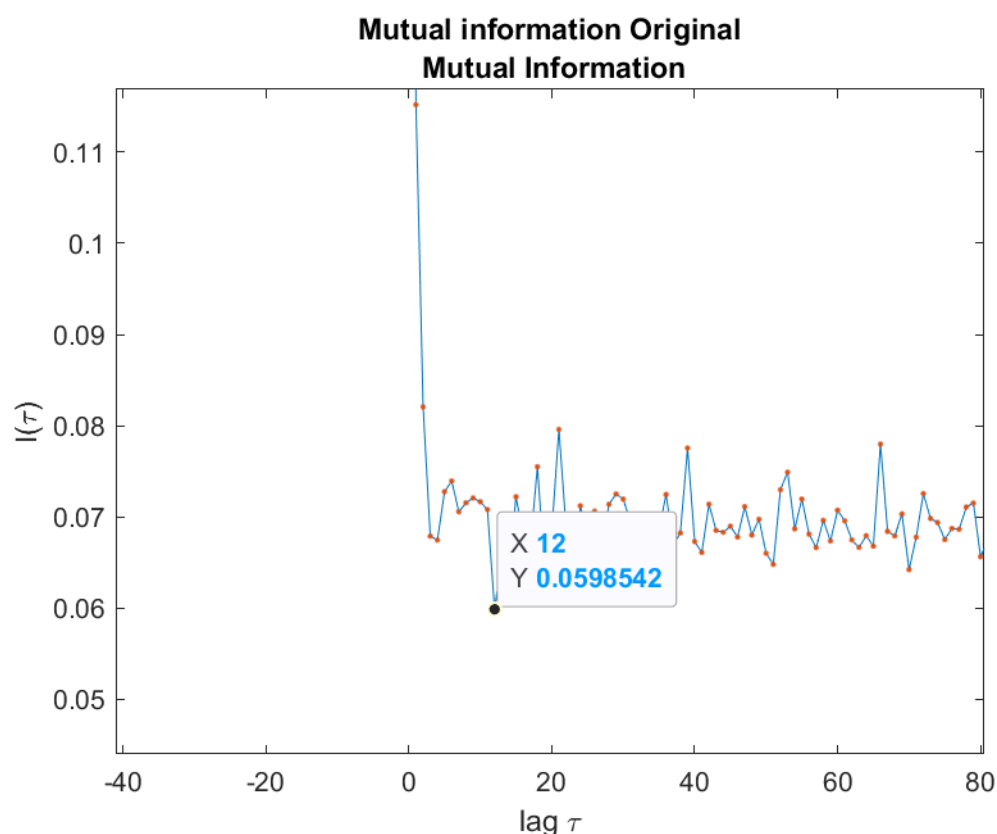
Ονομάζοντας **hV\_1**, **hV\_2** (**portmanteauLB()**) συνάρτηση τους πίνακες που καθορίζουν αν θα απορρίψουμε την  $H_0$  ( $H_0: \rho_\tau = 0$ ) για τη χρονοσειρά για  $n = 4915$  και για  $n = 500$  αντίστοιχα παρατηρούμε στον **hV\_1** ότι η  $H_0$  απορρίπτεται για όλες τις υστερήσεις. Ενώ στον **hV\_2** απορρίπτεται για όλες τις υστερήσεις εκτός από την πρώτη.

Θεωρητικά θα μπορούσαμε να βάλουμε τις τιμές για τον κάθε πίνακα αλλά είναι πολύ μεγάλος ο αριθμός των υστερήσεων και δεν είναι πρακτικό.

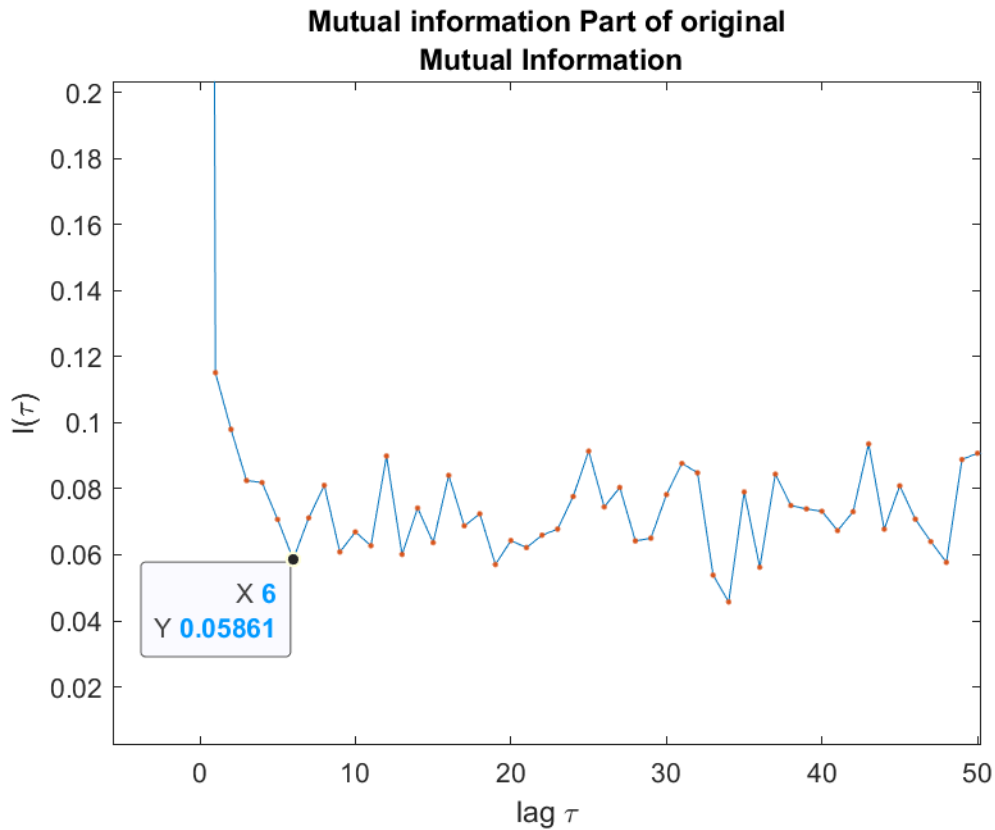
## 3. Εκτίμηση υστέρησης $\tau$ που δίνει το κριτήριο της αμοιβαίας πληροφορίας

Χρησιμοποιώντας το κριτήριο της αμοιβαίας πληροφορίας (**mutualinformation()**) θα επιλέξουμε κατάλληλο  $\tau$  για κάθε μία από τις χρονοσειρές. Ειδικότερα, θα δούμε για ποια τιμή υστέρησης παρουσιάζεται τοπικό ελάχιστο.

Σε συνέχεια των παραπάνω βλέπουμε για τη χρονοσειρά **original\_array\_2005** ότι το τοπικό ελάχιστο παρουσιάζεται για  $\tau = 12$  ενώ για τη χρονοσειρά **part\_of\_array\_2005** παίρνουμε  $\tau = 6$



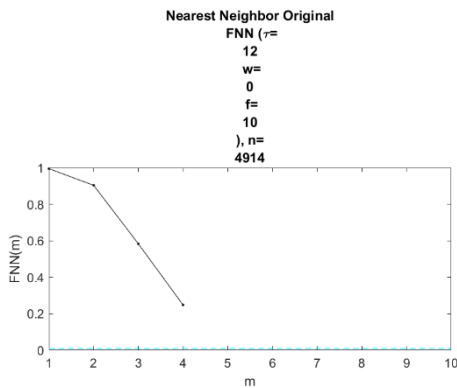
Σχήμα 7. Αμοιβαία Πληροφορία για χρονοσειρά  $n=4194$



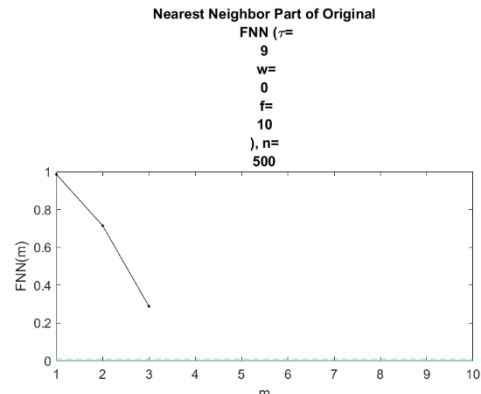
Σχήμα 8. Αμοιβαία Πληροφορία για χρονοσειρά  $n=500$

4. Εκτίμηση της διάστασης εμπύθισης  $m$  που σας δίνει το κριτήριο των ψευδών κοντινότερων γειτόνων (false nearest neighbors) χρησιμοποιώντας ως υστέ-ρηση αυτή που βρήκατε παραπάνω

Έχοντας τα συγκεκριμένα  $\tau$  θα χρησιμοποιήσουμε τη συνάρτηση ***falsenearest()*** και θα επιλέξουμε τις Default τιμές για τα ορίσματα *escape* και *theiler*. Η εκτίμηση μας για το  $m$  θα προκύψει από τα παρακάτω διαγράμματα κοιτώντας για ποιο  $m$  τερματίζει ο αλγόριθμος. Όπως φαίνεται για τη χρονοσειρά ***original\_array\_2005***  $m = 4$  και για τη χρονοσειρά ***part\_of\_array\_2005***  $m = 3$



Σχήμα 9a. Nearest Neighbor για χρονοσειρά  $n=4914$



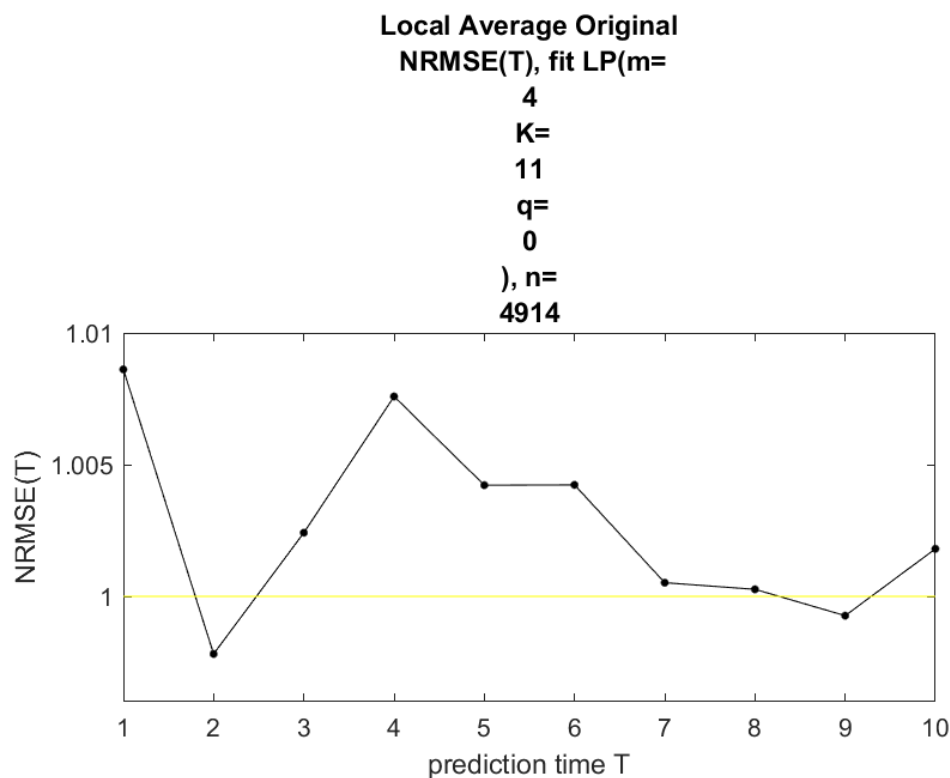
Σχήμα 9b. Nearest Neighbor για χρονοσειρά  $n=500$

5. Πρόβλεψη με τοπικό μοντέλο μέσου όρου και τοπικό γραμμικό μοντέλο για υστέρηση  $\tau$  και διάσταση εμπύθισης  $m$  που εκτιμήσατε παραπάνω (και τον ίδιο αριθμό γειτονικών σημείων)

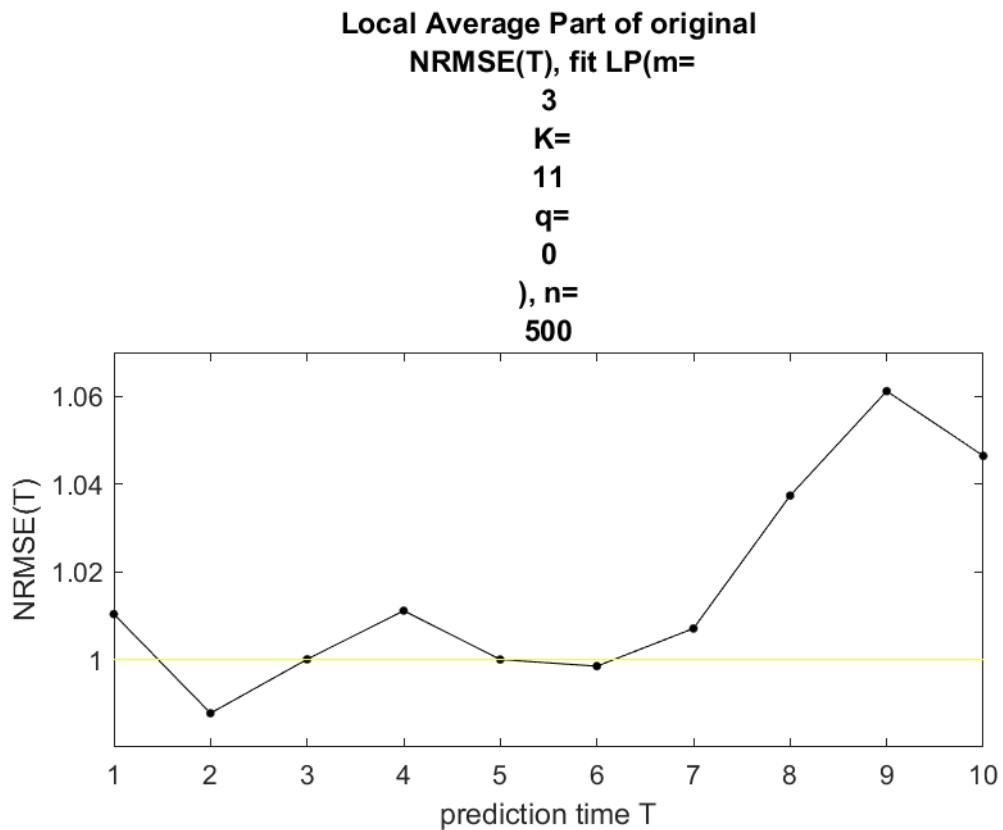
Μέσω της ***localfitnrmse()***, για  $\tau$  και  $m$  που αναφέρθηκαν, όπως επίσης και για  $K = 11$ ,  $T = 10$  προβλέψαμε και για τις 2 χρονοσειρές με το μοντέλο τοπικού μέσου όρου ( $q=0$ ) και με το τοπικό γραμμικό μοντέλο ( $q=4, n=4914$  |  $q=3, n=500$ ).

Αυτό που φαίνεται, βάσει του NRMSE για:

- Μοντέλο τοπικού μέσου όρου, η πρόβλεψη του μοντέλου είναι κοντά στο 1. Άρα είναι, περίπου, το ίδιο αν κάναμε πρόβλεψη με το μέσο όρο (και στις 2 χρονοσειρές).



Σχήμα 10. Μοντέλο τοπικού μέσου όρου για  $n=4914$



Σχήμα 11. Μοντέλο τοπικού μέσου όρου για  $n=500$

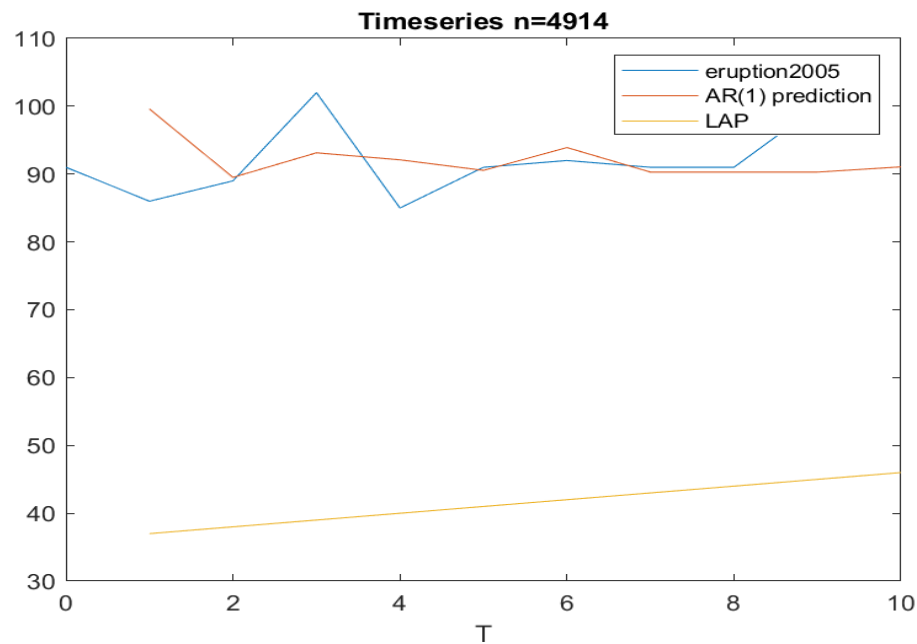
- Τοπικό γραμμικό μοντέλο, οι προβλέψεις είναι πολύ κακές ( $\text{NRMSE} \gg 1$ ) το οποίο δείχνει ότι η προσαρμογή ενός μη-γραμμικού μοντέλου δεν είναι καλή επιλογή. Μας προμηνύει ότι το σύστημα ίσως να είναι γραμμικό

Πίνακας 1 NRMSE για προβλέψεις τοπικού γραμμικού μοντέλου

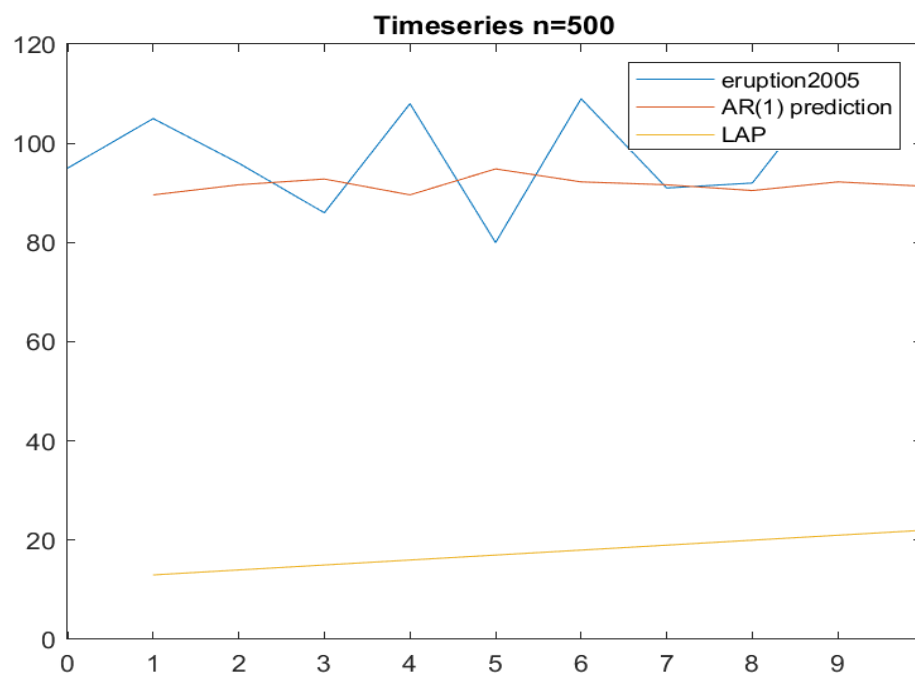
	NRMSE n=4914	NRMSE n=500
$\tau=1$	24,3776	33.8631
$\tau=2$	1,2253e+03	2.0352e+03
$\tau=3$	4.9436e+04	1.1163e+05
$\tau=4$	2.2030e+06	8.3352e+06
$\tau=5$	9.8779e+07	4.6378e+08
$\tau=6$	4.6813e+09	1.7241e+10
$\tau=7$	2.2422e+11	9.3980e+11
$\tau=8$	1.0740e+13	5.1648e+13
$\tau=9$	5.1469e+14	2.8361e+15
$\tau=10$	3.7407e+16	3.4116e+17

6. Σχεδιάγραμμα των προβλέψεων ενός βήματος μπροστά. Σύγκριση με τις αντίστοιχες προβλέψεις του γραμμικού μοντέλου (από το πρώτο μέρος της εργασίας). Σύγκριση του γραμμικού και μη-γραμμικού μοντέλου ως προς την πρόβλεψη για ένα βήμα μπροστά με βάση το  $nrmse$

Βλέποντας ότι το τοπικό γραμμικό μοντέλο αποτυγχάνει εντελώς στην προσαρμογή θα κάνουμε πρόβλεψη για ένα βήμα μπροστά μόνο με το μοντέλο τοπικού μέσου όρου. Όπως βλέπουμε στα παρακάτω διαγράμματα υπάρχουν οι προβλέψεις για ένα βήμα μπροστά και από το γραμμικό αλλά και το μη γραμμικό μοντέλο.



Σχήμα 12. Προβλέψεις και για τα 2 μοντέλα για  $n=4914$



Σχήμα 13. Προβλέψεις και για τα 2 μοντέλα για  $n=500$

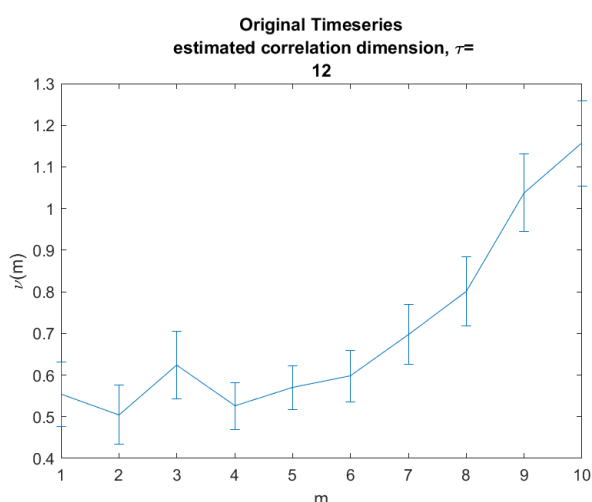
Χωρίς να δούμε τις τιμές του NRMSE, διαισθητικά, εξάγουμε το συμπέρασμα ότι το γραμμικό μοντέλο είναι καλύτερο έναντι του μη-γραμμικού

7. Εκτίμηση της διάστασης συσχέτισης για διαστάσεις  $m=1,...,10$ , χρησιμοποιώντας την υστέρηση από το κριτήριο της αμοιβαίας πληροφορίας με κατάλληλα σχήματα

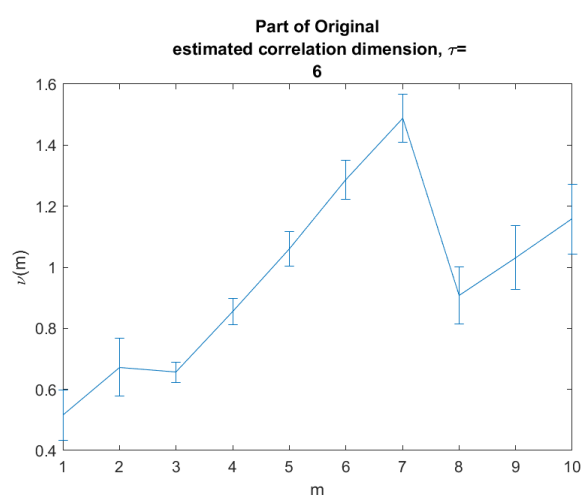
Χρησιμοποιήσαμε τη συνάρτηση `correlationdimension()`.

Γενικά, από τη θεωρία του χάους γνωρίζουμε πως η διάσταση συσχέτισης είναι ένα μέτρο της πολυπλοκότητας ή της μορφοκλασματικής διάστασης ενός συστήματος. Περιγράφει τον τρόπο με τον οποίο ο αριθμός των σημείων που απαιτούνται για την κάλυψη ενός χώρου αυξάνεται ανάλογα με το μέγεθος αυτού του χώρου.

Στην περίπτωση μας το διάγραμμα διάστασης συσχέτισης φαίνεται να μεγαλώνει εκθετικά (και για τις 2 χρονοσειρές). Αυτό σημαίνει ότι καθώς το  $m$  μεγαλώνει, αυξάνεται και η πολυπλοκότητα του συστήματος με επιταχυνόμενο ρυθμό. Αυτό είναι ένα χαρακτηριστικό χαοτικής συμπεριφοράς που καθιστά το σύστημα αιτιοκρατικό.

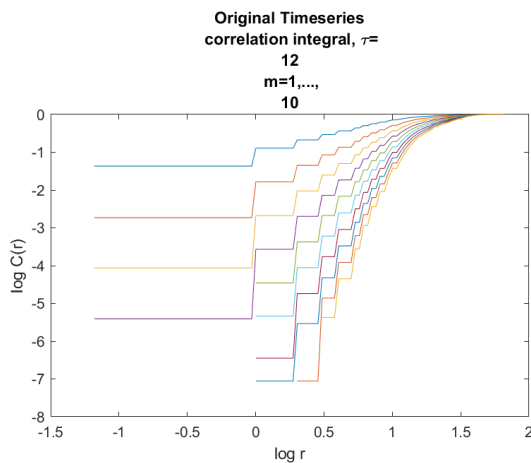


Σχήμα 14. Διάγραμμα εκτίμησης διάστασης συσχέτισης για  $n=4914$

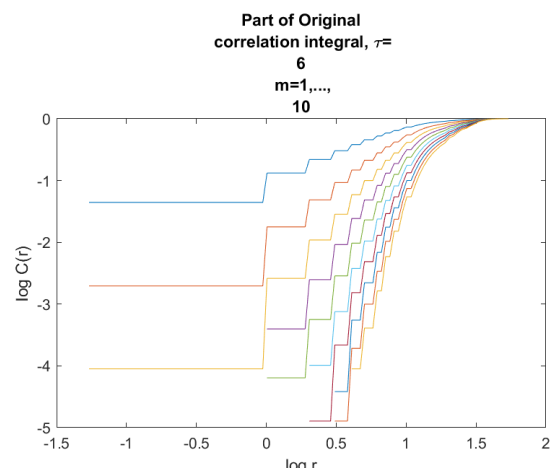


Σχήμα 15. Διάγραμμα εκτίμησης διάστασης συσχέτισης για  $n=500$

Για  $r$  μικρό έχουμε σταθερές κλίσεις ενώ για  $r$  μεγάλο δε διατηρείται η κλιμάκωση.



Σχήμα 16. Correlation integral για  $n=4914$

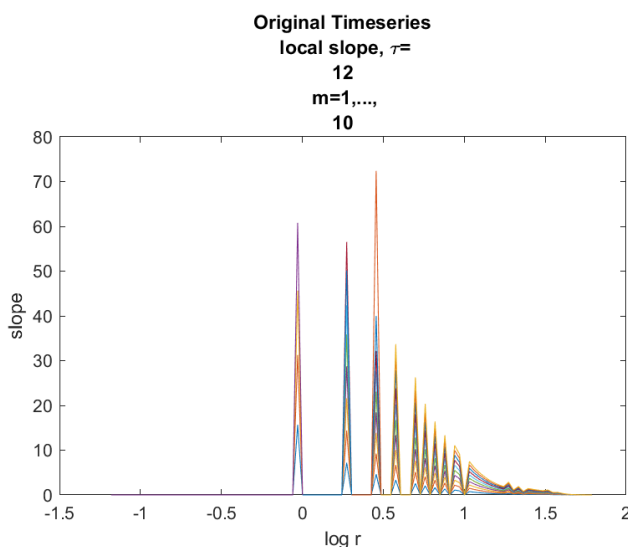


Σχήμα 17. Correlation integral για  $n=500$

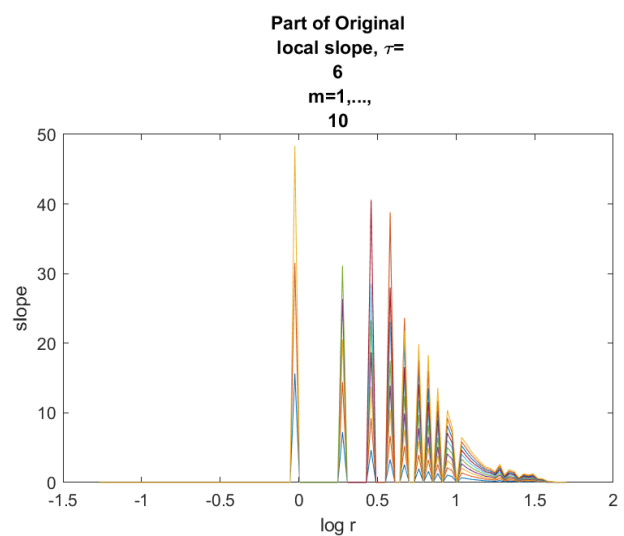
Η κλίση του αντίστοιχου διαγράμματος κλίσης δίνει πληροφορίες για τη μορφοκλασματική διάσταση του συστήματος. Εάν η κλίση είναι σταθερή, σημαίνει ότι το σύστημα είναι μορφοκλασματικό με κάποια ορισμένη διάσταση. Ωστόσο, εάν η κλίση αλλάζει καθώς αλλάζει η ακτίνα των κύκλων, υποδηλώνει ότι το σύστημα δεν είναι ένα απλό μορφοκλασματικό, αλλά έχει μια πιο περίπλοκη δομή.

Μια θετική κλίση στο αντίστοιχο διάγραμμα κλίσης υποδηλώνει ότι το σύστημα έχει υψηλή συσχέτιση, ενώ μια αρνητική κλίση δείχνει ότι το σύστημα είναι αντισυσχετισμένο. Μια κλίση μηδέν υποδηλώνει ότι το σύστημα δεν είναι συσχετισμένο.

Από το διάγραμμα της παραγώγου φαίνεται πως όσο το  $r$  μεγαλώνει, συγκλίνει στο 0, δείγμα μη-συσχετίσεων (κλίση μηδέν)



Σχήμα 19. Διάγραμμα κλίσης για  $n=4914$

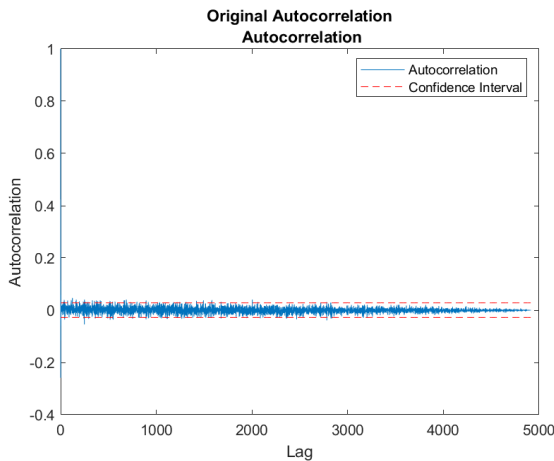


Σχήμα 18. Διάγραμμα κλίσης για  $n=500$

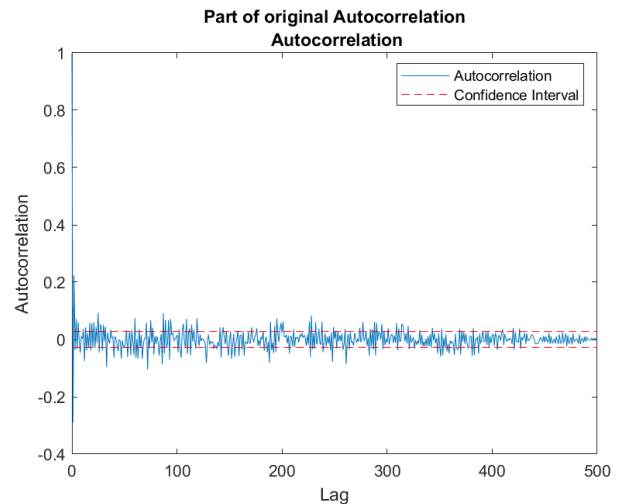


Σημαντικό μέτρο για τον καθορισμό ότι ένα σύστημα είναι χαοτικό ή στοχαστικό είναι η αυτοσυσχέτιση.

Στη χρονοσειρά για  $n=4914$ , βλέπουμε πως παρουσιάζεται μία «πιο στοχαστική» δομή από αυτή των 500 δειγμάτων (παρουσιάζονται περισσότερες στατιστικά σημαντικές αυτοσυσχετίσεις στη 2<sup>η</sup>).



Σχήμα 21. Αυτοσυσχέτιση για δείγματα  $n=4914$



Σχήμα 20. Αυτοσυσχέτιση για δείγματα  $n=500$

## **Συμπεράσματα**

Τα παραπάνω αποτελέσματα δε μας δίνουν μία ξεκάθαρη εικόνα για την περίπτωση που το σύστημα είναι χαοτικό η στοχαστικό. Με ευκολία ωστόσο μπορούμε να πούμε ότι ταιριάζει περισσότερο η προσέγγιση ενός γραμμικού μοντέλου αφού τα μη-γραμμικά μοντέλα παρουσίασαν μεγαλύτερο σφάλμα από τα γραμμικά. Θα μπορούσαμε να πούμε πως το σύστημα μας έχει μια περισσότερο στοχαστική μορφή για περισσότερες παρατηρήσεις, όταν όμως μειώνεται η αριθμός των δειγμάτων παρουσιάζεται σαν «λίγο πιο αιτιοκρατικό». Από τη στιγμή που θεωρούμε ότι όσο περισσότερα δείγματα έχουμε, τόσο το καλύτερο ως προς την αναλυτικότητα του συστήματος θεωρούμε το σύστημα στοχαστικό.