

Χρονοσειρές
Εργασία μαθήματος 2024-2025

ΟΜΑΔΑ 12
Μαυρογεωργιάδη Χριστίνα 10439
Πάυλου Δημήτριος 10315

Φεβρουάριος 2025

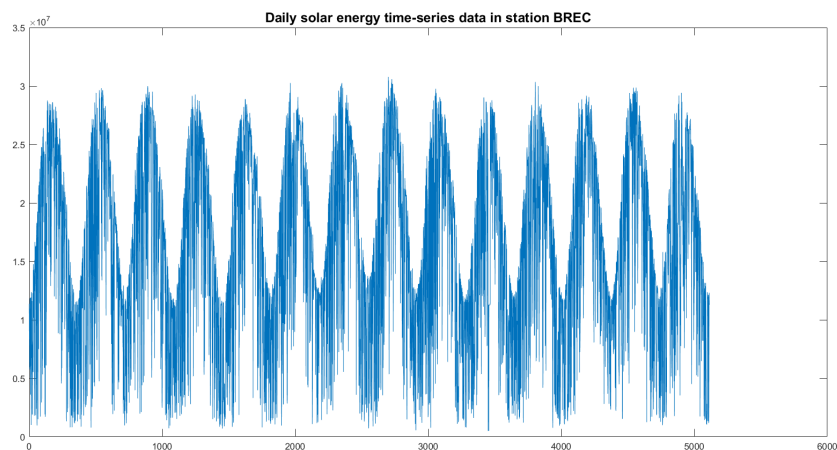
Περιεχόμενα

1	Εισαγωγή	1
2	Προεπεξεργασία δεδομένων	2
3	Απαλοιφή εποχικότητας	2
4	Στατιστικός έλεγχος ανεξαρτησίας	4
5	Προσαρμογή γραμμικού μοντέλου χρονοσειράς	5
6	Πρόβλεψη με γραμμικό μοντέλο χρονοσειράς	9
7	Πρόβλεψη με επέκταση της περιοδικής συνάρτησης	10
8	Υπολογισμός Seasonal NRMSE (SNRMSE)	11
9	Προσαρμογή μη-γραμμικού μοντέλου χρονοσειράς	11
9.1	Εκτίμηση υστέρησης τ	11
9.2	Εκτίμηση διάστασης εμβύθισης m	12
9.3	Τοπικό γραμμικό μοντέλο	13
9.4	Τοπικό μοντέλο μέσου όρου	14
10	Πρόβλεψη με μη-γραμμικό μοντέλο χρονοσειράς	16
11	Σύγκριση μοντέλων	17
12	Συμπεράσματα	18
13	Διάρθρωση κώδικα	18

1 Εισαγωγή

Η παρούσα εργασία αφορά την ανάλυση χρονοσειράς και την προσαρμογή γραμμικού και μη-γραμμικού μοντέλου με σκοπό την πρόβλεψη. Τα δεδομένα αφορούν την συνολική ημερήσια εισερχόμενη ηλιακή ενέργεια (J/m^2) σε 98 σταθμούς, για το χρονικό διάστημα 1 Ιανουαρίου 1994 έως 31 Δεκεμβρίου 2007. Θα μελετήσουμε τις μετρήσεις που σχετίζονται με τον 12ο σταθμό (ομάδα 12), δηλαδή τον σταθμό BREC.

Παρακάτω απεικονίζεται η εν λόγω χρονοσειρά.



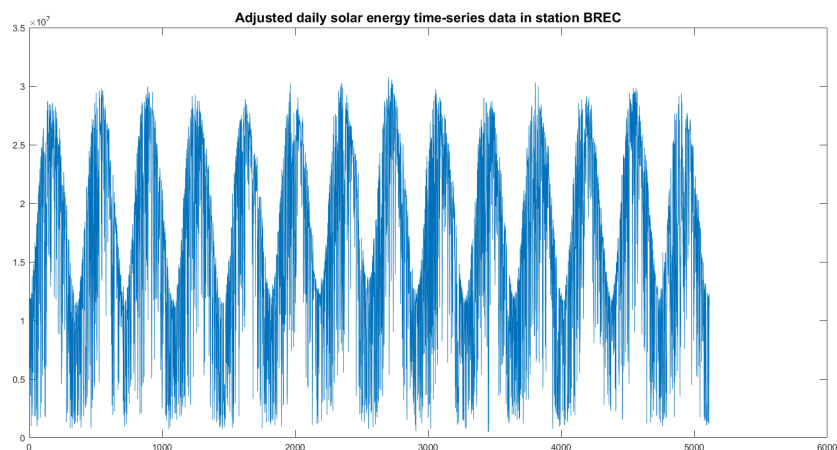
Σχήμα 1: Μη-στάσιμη χρονοσειρά

Παρατηρούμε ότι η χρονοσειρά εμφανίζει εποχικότητα, η οποία υποθέτουμε ότι είναι ετήσια με μήκος περιόδου $s \approx 365$. Επιπλέον, υποθέτουμε ότι η χρονοσειρά δεν χαρακτηρίζεται από τάση.

2 Προεπεξεργασία δεδομένων

Κατά τη μελέτη της χρονοσειράς, διαπιστώσαμε, εφαρμόζοντας πρώτες διαφορές, ότι υπάρχουν μετρήσεις που παραμένουν σταθερές επί συνεχόμενες ημέρες, λόγω σφαλμάτων στη μέτρηση. Αυτό συμβαίνει για 13 παρατηρήσεις της χρονοσειράς. Εξαλείφουμε τις εσφαλμένες παρατηρήσεις, πραγματοποιώντας εκτίμηση της τιμής τους ως εξής: η εκτιμώμενη τιμή υπολογίζεται ως η μέση τιμή των μετρήσεων του προηγούμενου και του επόμενου έτους. Η εκτίμηση αυτή λαμβάνει υπόψη την περιοδικότητα της χρονοσειράς.

Η χρονοσειρά μετά την επεξεργασία απεικονίζεται παρακάτω, χωρίς να είναι διακριτή βέβαια η διαφορά στις τιμές λόγω του μεγάλου πλήθους παρατηρήσεων.



Σχήμα 2: Μη-στάσιμη χρονοσειρά μετά την προεπεξεργασία

3 Απαλοιφή εποχικότητας

Πρώτο μας βήμα κατά την ανάλυση της χρονοσειράς είναι η απαλοιφή της εποχικότητας. Όπως αναφέραμε προηγουμένως, θεωρούμε περίοδο ίση με $d \approx 365$. Για την απαλοιφή της εποχικότητας χρησιμοποιούμε την παρακάτω μέθοδο:

- Έστω s_t , $t = 1, \dots, d$ είναι τα στοιχεία της περιοδικής συνάρτησης στη διάρκεια μιας περιόδου. Έχουμε

$$y_t = s_t + x_t$$

όπου y_t είναι η μη-στάσιμη χρονοσειρά.

- Τα στοιχεία αυτά εκτιμώνται ως

$$\hat{s}_i = \frac{1}{k} \sum_{j=1}^k y_{i+jd}$$

όπου $k = \lfloor n/d \rfloor$ είναι το πλήθος των περιόδων που εμφανίζονται στην χρονοσειρά.

- Υπολογίζουμε τη μέση τιμή των περιοδικών στοιχείων

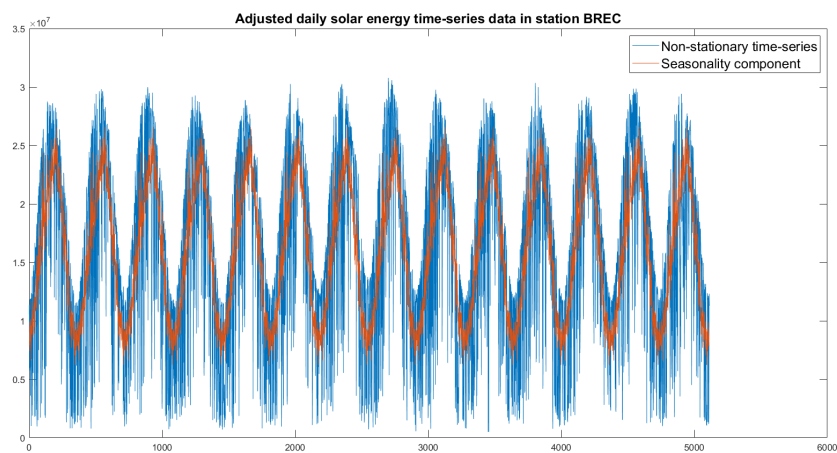
$$\bar{s} = \frac{1}{d} \sum_{i=1}^d \hat{s}_i$$

- Αφαιρούμε τη μέση τιμή \bar{s} από τα περιοδικά στοιχεία που υπολογίσαμε:

$$\hat{s}_i \leftarrow \hat{s}_i - \bar{s}$$

έτσι ώστε η περιοδική χρονοσειρά να έχει μηδενική μέση τιμή.

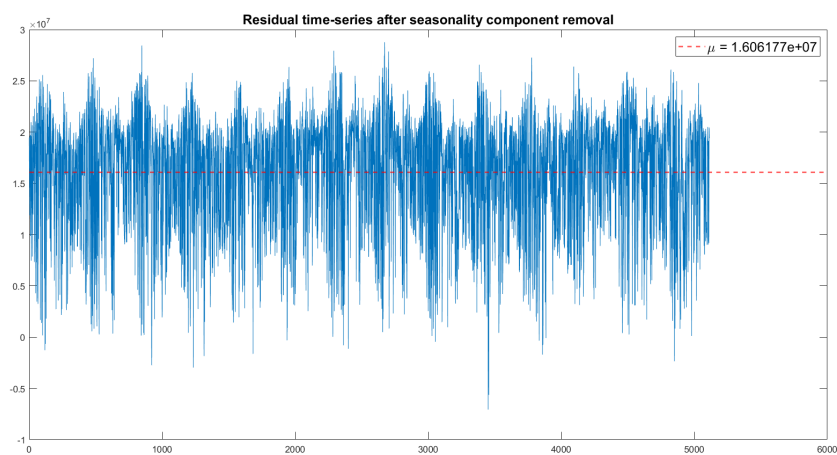
Αφού έχουμε ολοκληρώσει την εκτίμηση των περιοδικών στοιχείων της χρονοσειράς, επαναλαμβάνουμε τα εκτιμώμενα στοιχεία για k περιόδους, έτσι ώστε να προκύψει περιοδική χρονοσειρά μήκους n .



Σχήμα 3: Μη-στάσιμη χρονοσειρά και εποχικό στοιχείο

Στο παραπάνω σχήμα, έχουμε προσθέσει στην περιοδική χρονοσειρά τη μέση τιμή των υπολοίπων για λόγους απεικόνισης.

Για να εξαλείψουμε την εποχικότητα, αφαιρούμε από την αρχική μη-στάσιμη χρονοσειρά την περιοδική χρονοσειρά και προκύπτει η παρακάτω:



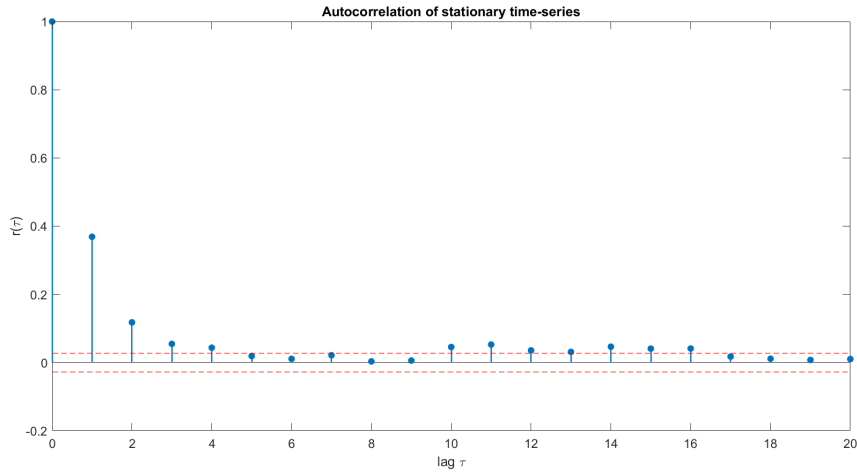
Σχήμα 4: Στάσιμη χρονοσειρά

Η στάσιμη χρονοσειρά x_t έχει μέση τιμή ίση με $\mu = 1.606177e + 07$, σύμφωνα με την δειγματική εκτίμηση.

4 Στατιστικός έλεγχος ανεξαρτησίας

Πριν προχωρήσουμε στην διερεύνηση κάποιου γραμμικού μοντέλου για την χρονοσειρά, πρέπει να εξετάσουμε αν η χρονοσειρά x_t μπορεί να αποτελεί λευκό θόρυβο. Στην περίπτωση που η χρονοσειρά είναι λευκός θόρυβος, δεν έχει νόημα να προσαρμόσουμε κάποιο γραμμικό μοντέλο, αφού χαρακτηρίζεται από πλήρη τυχαιότητα και μη-προβλεψιμότητα.

Υπολογίζουμε την αυτοσυσχέτιση της χρονοσειράς για υστερήσεις έως και 20 ημέρες. Η συνάρτηση αυτοσυσχέτισης απεικονίζεται στο παρακάτω σχήμα.



Σχήμα 5: Συνάρτηση αυτοσυσχέτισης

Οι κόκκινες διακεκομμένες γραμμές αναπαριστούν τα όρια της περιοχής απόρριψης στατιστικής σημαντικότητας της αυτοσυσχέτισης. Τα όρια αυτά είναι περίπου ίσα με $\pm 2/\sqrt{n}$ για επίπεδο σημαντικότητας $\alpha = 0.05$. Σύμφωνα με τα όρια αυτά, η αυτοσυσχέτιση είναι στατιστικά σημαντική για τις υστερήσεις: 1, 2, 3, 4, 10, 11, 12, 13, 14, 15, 16.

Το ποσοστό των υστερήσεων για τις οποίες η συσχέτιση είναι σημαντική είναι ίσο με 0.55, συνεπώς με βάση αυτόν τον έλεγχο απορρίπτουμε τη μηδενική υπόθεση ότι η χρονοσειρά παράγεται από στοχαστική διαδικασία λευκού θορύβου.

Θα εφαρμόσουμε, επιπλέον, τον στατιστικό έλεγχο Portmanteau, ο οποίος ελέγχει τη σημαντικότητα της r_τ για ένα εύρος υστερήσεων συγχρόνως.

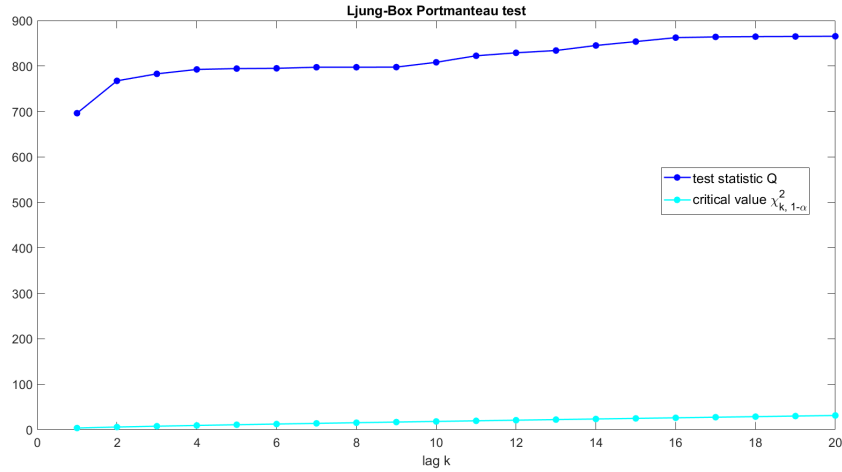
Υπολογίζουμε το στατιστικό

$$Q = n(n+2) \sum_{\tau=1}^k r_\tau^2 / (n - \tau)$$

για κάποια μέγιστη υστέρηση k .

Το στατιστικό ελέγχου Q ακολουθεί χ_k^2 κατανομή. Μπορούμε να απορρίψουμε την μηδενική υπόθεση ότι η χρονοσειρά είναι λευκός θόρυβος, με επίπεδο σημαντικότητας α εάν

$$Q > \chi_{k,1-\alpha}^2$$



Σχήμα 6: Στατιστικός έλεγχος Portmanteau

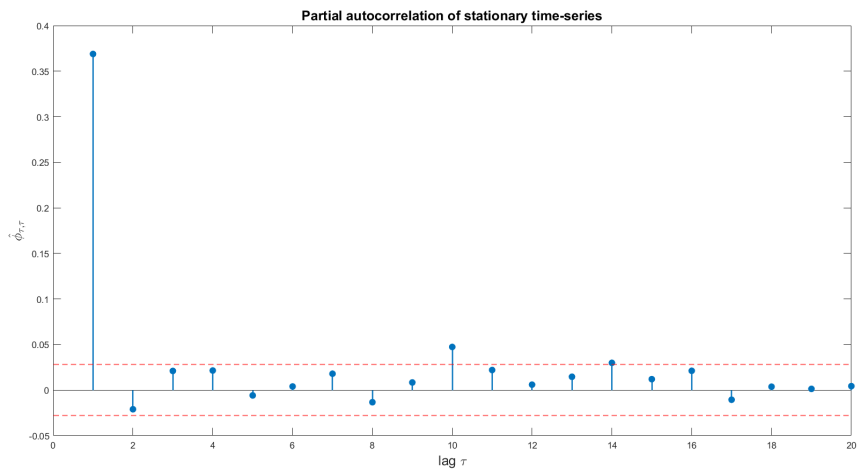
Η τιμή του στατιστικού είναι αρκετά υψηλότερη της κρίσιμης τιμής για όλες τις τιμές της μέγιστης υστέρησης k . Δηλαδή, η μηδενική υπόθεση του ελέγχου απορρίπτεται για όλες τις τιμές του k που εξετάστηκαν.

Συμπερασματικά, δεν μπορούμε να δεχτούμε ότι η στάσιμη χρονοσειρά προέρχεται από στοχαστική διαδικασία λευκού θορύβου.

5 Προσαρμογή γραμμικού μοντέλου χρονοσειράς

Στην ενότητα αυτή, θα διερευνήσουμε το καταλληλότερο γραμμικό μοντέλο που προσαρμόζεται στη στάσιμη χρονοσειρά. Συγκεκριμένα, επιθυμούμε να εκτιμήσουμε την κατάλληλη τάξη p, q ενός $ARMA(p, q)$ μοντέλου που προσαρμόζεται στη χρονοσειρά. Στην περίπτωση ενός $ARMA$ μοντέλου, η μερική αυτοσυσχέτιση και η αυτοσυσχέτιση δεν επιτρέπουν τον προσδιορισμό των τάξεων του μοντέλου. Για το σκοπό αυτό, θα χρησιμοποιήσουμε τα κριτήρια πληροφoρίας AIC και FPE για συνδυασμούς τιμών των p, q .

Θα παραθέσουμε, πρώτα, το διάγραμμα της συνάρτησης μερικής αυτοσυσχέτισης της στάσιμης χρονοσειράς, για λόγους πληρότητας της ανάλυσης.



Σχήμα 7: Συνάρτηση μερικής αυτοσυσχέτισης

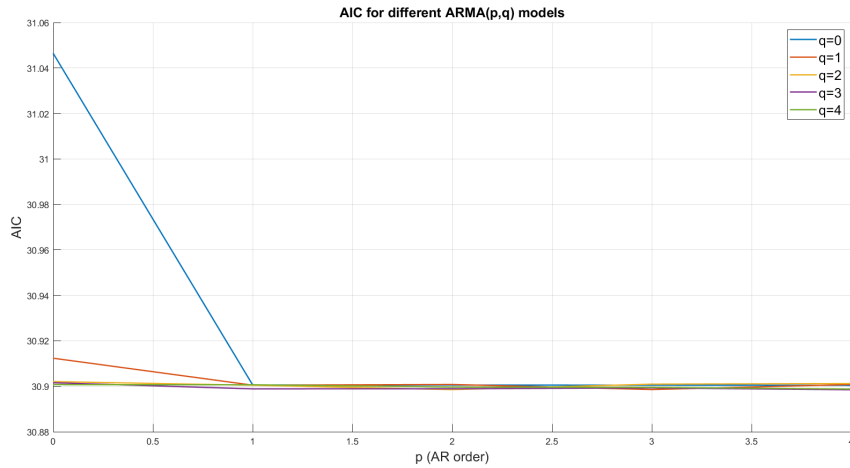
Παρατηρούμε ότι η συνάρτηση μερικής αυτοσυσχέτισης είναι στατιστικά σημαντική για τις υστερήσεις: 1, 10 και 14. Με βάση αυτή την παρατήρηση, θα μπορούσαμε να θεωρήσουμε κατάλληλο ένα $AR(1)$ μοντέλο, αν επιλέξουμε να αγνοήσουμε τη σημαντικότητα για τις πολύ υψηλές υστερήσεις.

Στη συνέχεια, προσαρμόζουμε διαφορετικά $ARMA(p, q)$ μοντέλα για συνδυασμούς των εξής τιμών των p, q :

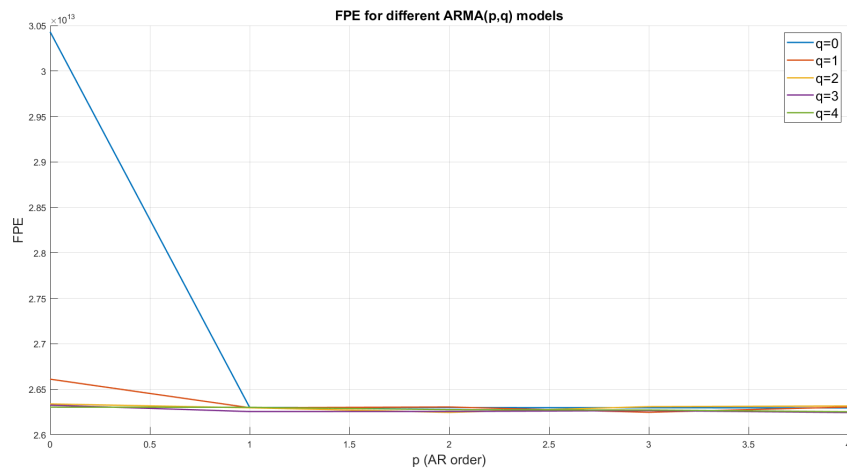
- $p = 0, 1, 2, 3, 4$
- $q = 0, 1, 2, 3, 4$

και υπολογίζουμε τα αντίστοιχα κριτήρια πληροφορίας.

Προκύπτουν τα παρακάτω διαγράμματα:



Σχήμα 8: Κριτήριο πληροφορίας AIC



Σχήμα 9: Κριτήριο πληροφορίας FPE

Παραθέτουμε, επίσης, τις τιμές του AIC για τα ζεύγη τιμών p, q στον παρακάτω πίνακα, με ακρίβεια 4ου δεκαδικού ψηφίου.

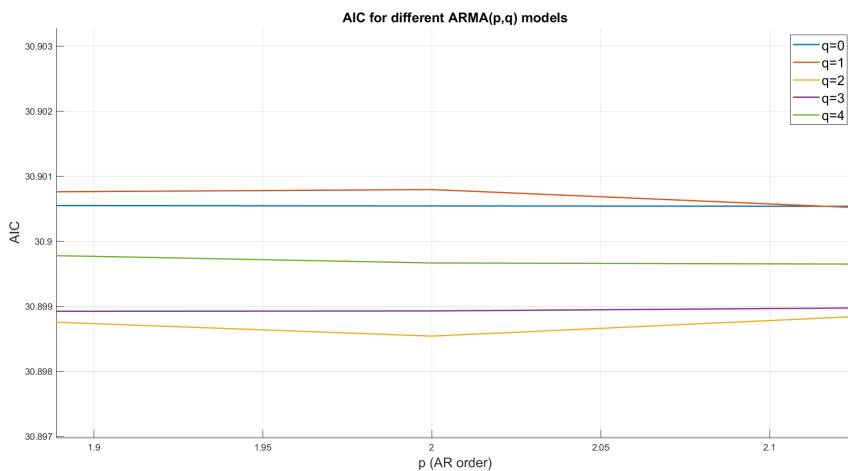
	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$
$p = 0$	31.0465	30.9124	30.9021	30.9016	30.9007
$p = 1$	30.9006	30.9005	30.9005	30.8989	30.9007
$p = 2$	30.9005	30.9008	30.8985	30.8989	30.8997
$p = 3$	30.9005	30.8986	30.9009	30.8993	30.8995
$p = 4$	30.9004	30.9008	30.9012	30.8985	30.8988

Πίνακας 1: Κριτήριο πληροφορίας AIC

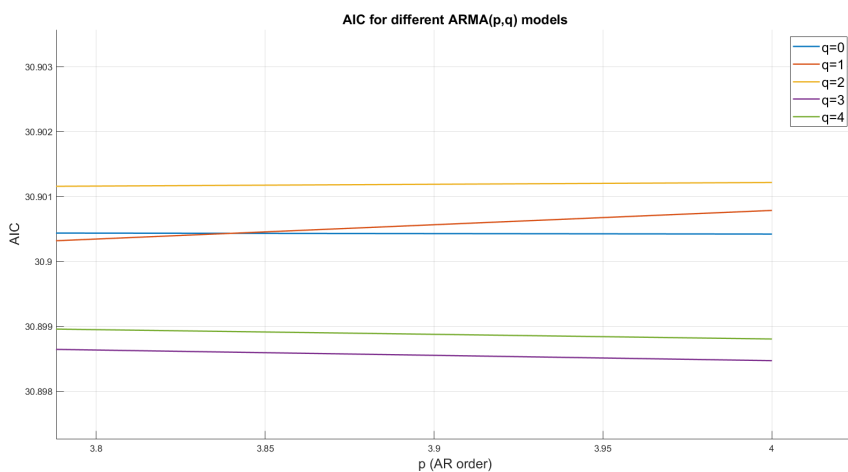
Διαπιστώνουμε, γενικά, πως δεν υπάρχει κάποιο ευδιάκριτο ολικό ελάχιστο του AIC που επιτυγχάνεται για κάποιο ζεύγος p, q . Σύμφωνα με τον πίνακα, με ακρίβεια 10^{-4} , η ελάχιστη τιμή του AIC είναι 30.8985 και επιτυγχάνεται

για τους συνδυασμούς $(p, q) = (2, 2)$ και $(p, q) = (4, 3)$. Επιπλέον, η δεύτερη κοντινότερη ελάχιστη τιμή είναι 30.8986 και επιτυγχάνεται για τον συνδυασμό $(p, q) = (3, 1)$.

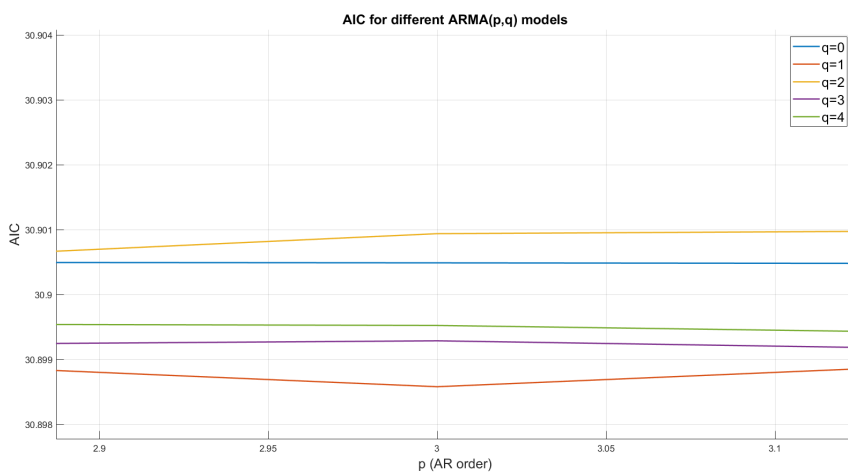
Μεγεθύνοντας το διάγραμμα του AIC , βλέπουμε:



(α') Ελάχιστη τιμή στο $(p, q) = (2, 2)$



(β') Ελάχιστη τιμή στο $(p, q) = (4, 3)$



(γ') Ελάχιστη τιμή στο $(p, q) = (3, 1)$

Σχήμα 10: Κριτήριο πληροφορίας AIC σε μεγέθυνση

Γενικά, δεν είναι ξεκάθαρο ποιες τάξεις του μοντέλου να επιλέξουμε. Αποφεύγουμε την προσαρμογή μοντέλων με μεγαλύτερη τάξη, διότι μπορεί να οδηγήσει σε υπερ-προσαρμογή του μοντέλου στο θόρυβο (overfitting). Θα επιλέξουμε, λοιπόν, το μοντέλο $ARMA(2, 2)$ για να κάνουμε μελλοντικές προβλέψεις, το οποίο έχει το ελάχιστο AIC και την μικρότερη πολυπλοκότητα ($p + q = 4$).

Το μοντέλο $ARMA(2, 2)$ για τη στάσιμη χρονοσειρά x_t ορίζεται ως:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + z_t - \theta_1 z_{t-1} - \theta_2 z_{t-2}, \quad z_t \sim WN(0, \sigma_z^2)$$

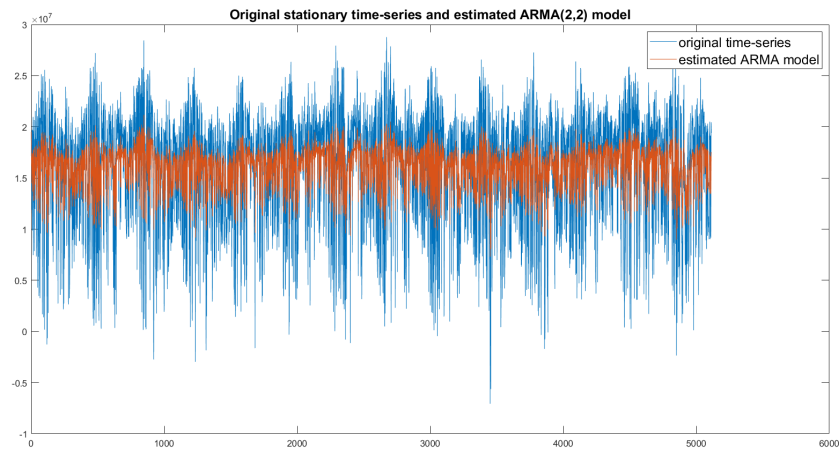
Οι συντελεστές του μοντέλου εκτιμήθηκαν ως:

$$\hat{\phi}_0 = 1.7607e + 05, \quad \hat{\phi}_1 = 1.2547, \quad \hat{\phi}_2 = -0.2657$$

$$\hat{\theta}_1 = 0.88, \quad \hat{\theta}_2 = 0.0929$$

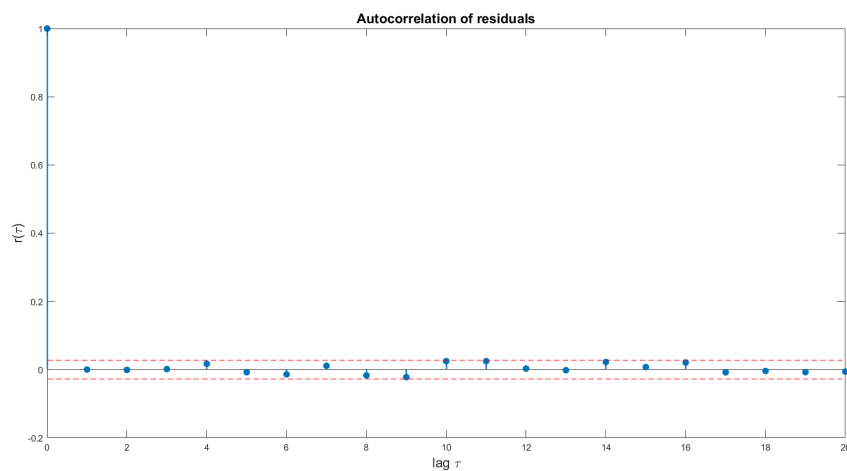
ενώ η διασπορά του θορύβου εκτιμήθηκε ως:

$$s_z^2 = 5.1210863e + 06$$



Σχήμα 11: Προσαρμοσμένο μοντέλο $ARMA(2, 2)$

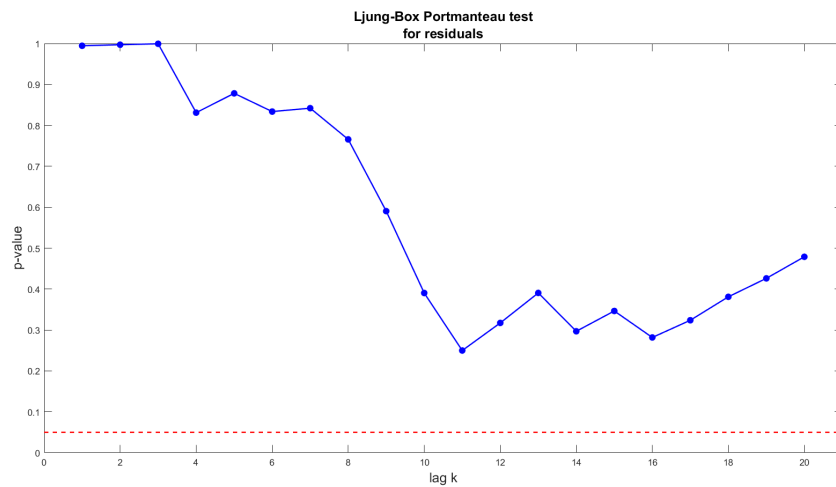
Πριν προχωρήσουμε στις προβλέψεις, είναι σημαντικό να επιβεβαιώσουμε την καταλληλότητα του μοντέλου, ελέγχοντας ότι τα υπολοίπα της προσαρμογής προέρχονται από διαδικασία λευκού θορύβου.



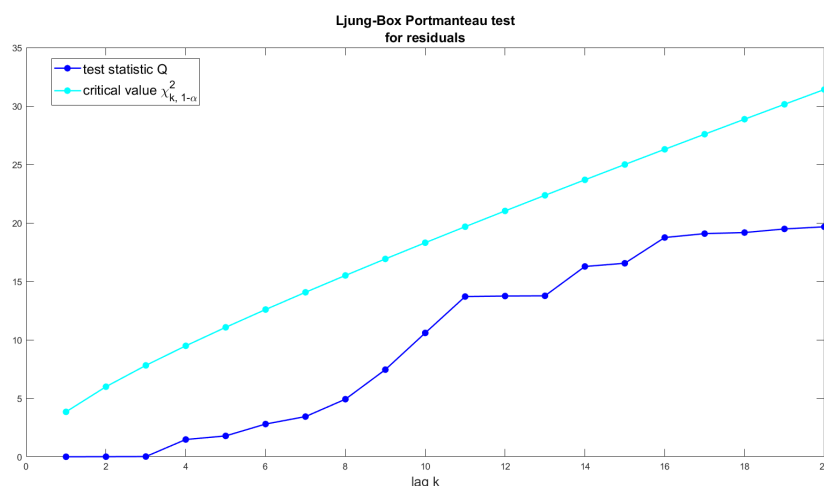
Σχήμα 12: Συνάρτηση αυτοσυσχέτισης υπολοίπων του μοντέλου $ARMA(2, 2)$

Μελετώντας τη συνάρτηση αυτοσυσχέτισης των υπολοίπων, βλέπουμε ότι οι τιμές της αυτοσυσχέτισης είναι στατιστικά μη-σημαντικές για όλες τις υστερήσεις. Αυτό μας υποδεικνύει ότι η χρονοσειρά των υπολοίπων είναι λευκός

θόρυβος. Στη συνέχεια, εκτελούμε τον έλεγχο ανεξαρτησίας Portmanteau για τη χρονοσειρά των υπολοίπων. Όπως γίνεται φανερό από τα παρακάτω σχήματα, ο έλεγχος δεν απορρίπτει τη μηδενική υπόθεση λευκού θορύβου, καθώς η τιμή του στατιστικού Q για όλες τις μέγιστες υστερήσεις k είναι κάτω από την κρίσιμη τιμή της κατανομής. Συμπεραίνουμε, λοιπόν, πως το μοντέλο $ARMA(2, 2)$ είναι κατάλληλο.



Σχήμα 13: p -τιμές ελέγχου Portmanteau



Σχήμα 14: Στατιστικό ελέγχου Portmanteau

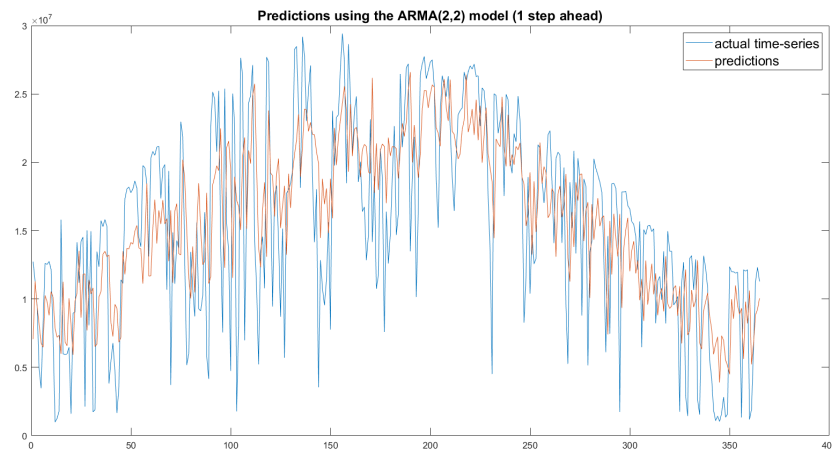
Σχήμα 15: Στατιστικός έλεγχος ανεξαρτησίας υπολοίπων του μοντέλου $ARMA(2, 2)$

6 Πρόβλεψη με γραμμικό μοντέλο χρονοσειράς

Θεωρούμε το σύνολο εκπαίδευσης ως τις τιμές τις χρονοσειράς τα έτη 1994-2006 και το σύνολο ελέγχου ως τις τιμές της χρονοσειράς το έτος 2007. Επιθυμούμε να κάνουμε προβλέψεις για όλες τις ημέρες του έτους 2007 για $T = 1, \dots, 30$ βήματα μπροστά. Οι προβλέψεις αφορούν την αρχική μη-στάσιμη χρονοσειρά.

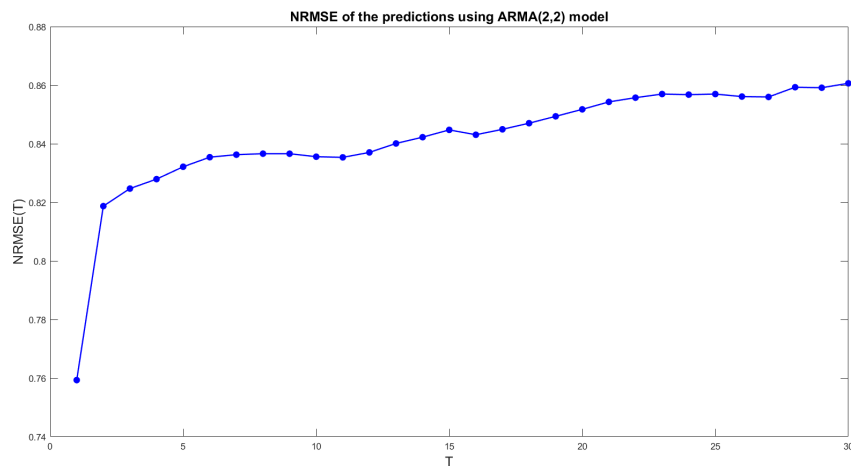
Αρχικά, χρησιμοποιούμε την χρονοσειρά του συνόλου εκπαίδευσης για να εκτιμήσουμε το στοιχείο εποχικότητας της χρονοσειράς. Επεκτείνουμε την περιοδική συνάρτηση, έτσι ώστε να καλύψει τις τιμές της χρονοσειράς που βρίσκονται στο σύνολο ελέγχου και αφαιρούμε την εποχικότητα από την αρχική χρονοσειρά. Στη συνέχεια, χρησιμοποιούμε το μοντέλο $ARMA(2, 2)$ για να κάνουμε προβλέψεις στο σύνολο ελέγχου με τη στάσιμη χρονοσειρά που προκύπτει για $T = 1, \dots, 30$ βήματα μπροστά. Τέλος, στις προβλέψεις της στάσιμης χρονοσειράς προσθέτουμε την επέκταση της περιοδικής συνάρτησης που εκτιμήσαμε, έτσι ώστε να προκύψουν οι τελικές προβλέψεις.

Στο παρακάτω σχήμα, φαίνονται οι προβλέψεις για το έτος 2007 για ένα βήμα μπροστά.



Σχήμα 16: Προβλέψεις στο σύνολο ελέγχου για 1 βήμα μπροστά

Υπολογίζουμε το NRMSE των προβλέψεων για $T = 1, \dots, 30$ βήματα μπροστά και προκύπτει το παρακάτω διάγραμμα:



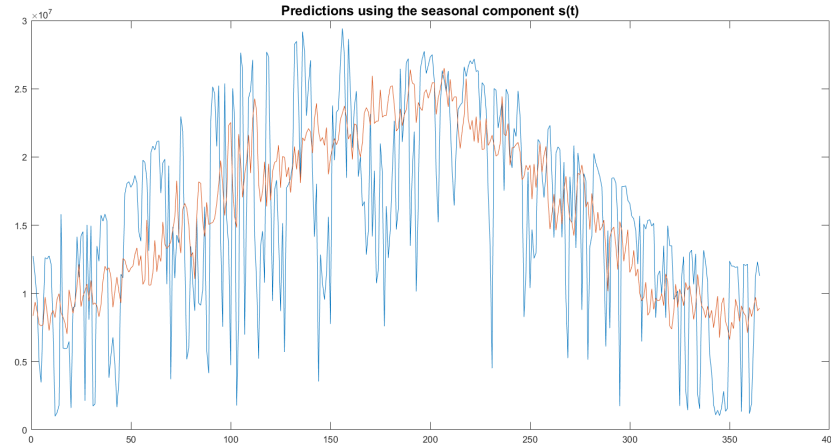
Σχήμα 17: NRMSE προβλέψεων σε συνάρτηση με το T

Για 1 βήμα μπροστά, έχουμε $NRMSE(1) = 0.759306$. Όσο αυξάνεται ο χρονικός ορίζοντας πρόβλεψης T , το NRMSE αυξάνεται και συγκλίνει σε μία σταθερή τιμή, καθώς η πρόβλεψη του γραμμικού μοντέλου για τη στάσιμη χρονοσειρά συγκλίνει στη μέση τιμή της.

7 Πρόβλεψη με επέκταση της περιοδικής συνάρτησης

Για την πρόβλεψη των τιμών της χρονοσειράς στο σύνολο ελέγχου, χρησιμοποιούμε την επέκταση της περιοδικής συνάρτησης που εκτιμήσαμε από το σύνολο εκπαίδευσης στο προηγούμενο βήμα. Σε αυτή προσθέτουμε τη μέση τιμή της χρονοσειράς των υπολοίπων που υπολογίζεται από το σύνολο εκπαίδευσης. Οι προβλέψεις που προκύπτουν για 1 βήμα μπροστά φαίνονται στο παρακάτω σχήμα.

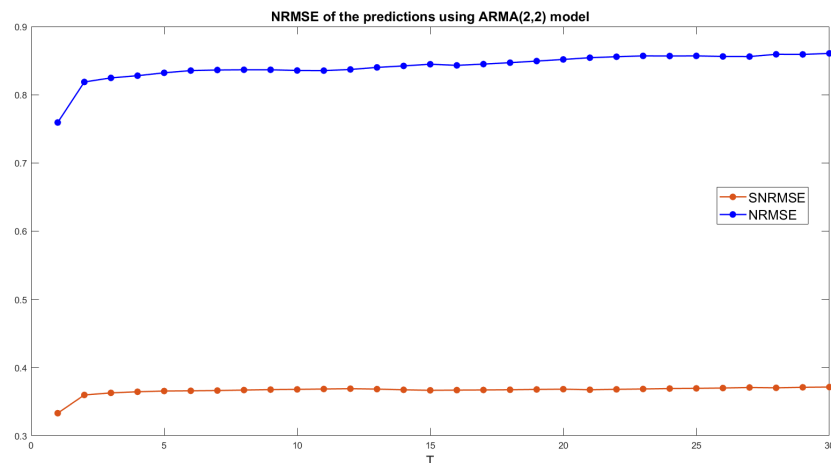
Το NRMSE που υπολογίστηκε για αυτές τις προβλέψεις είναι ίσο με 0.830344. Συγκρίνοντας τα NRMSE για τις προβλέψεις του γραμμικού μοντέλου και του μοντέλου περιοδικής συνάρτησης, παρατηρούμε ότι το μοντέλο περιοδικής συνάρτησης έχει αυξημένο NRMSE. Αυτό συμβαίνει διότι το μοντέλο αυτό δε λαμβάνει υπόψη του τις αυτοσυσχετίσεις μεταξύ των υστερήσεων της χρονοσειράς.



Σχήμα 18: Προβλέψεις στο σύνολο ελέγχου με επέκταση εποχικότητας

8 Υπολογισμός Seasonal NRMSE (SNRMSE)

Για το γραμμικό μοντέλο πρόβλεψης, θα υπολογίσουμε ένα νέο στατιστικό, το SNRMSE. Ο υπολογισμός του στατιστικού αυτού είναι παρόμοιος με το NRMSE, με την διαφορά ότι στη θέση της δειγματικής μέσης τιμής τίθεται η τιμή της εκτίμησης της περιοδικής συνάρτησης. Παρουσιάζουμε στο παρακάτω διάγραμμα τις τιμές των NRMSE και SNRMSE για το γραμμικό μοντέλο.



Σχήμα 19: NRMSE και SNRMSE για το γραμμικό μοντέλο σε συνάρτηση με το T

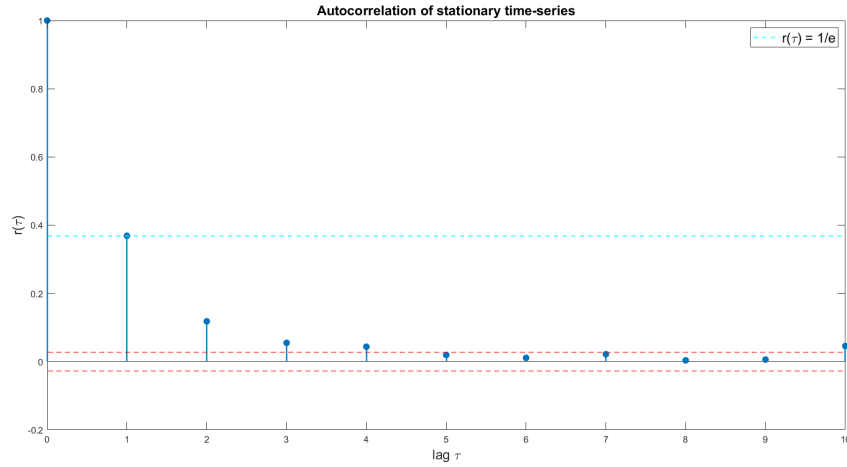
Παρατηρούμε ότι οι τιμές του SNRMSE είναι χαμηλότερες από αυτές του NRMSE για όλα τα βήματα πρόβλεψης T . Αυτό μπορεί να αιτιολογηθεί από το γεγονός ότι η μέση τιμή μιας χρονοσειράς με εποχικότητα δεν είναι αντιπροσωπευτικό στατιστικό μέτρο για την χρονοσειρά, σε αντίθεση με τις τιμές της περιοδικής συνάρτησης.

9 Προσαρμογή μη-γραμμικού μοντέλου χρονοσειράς

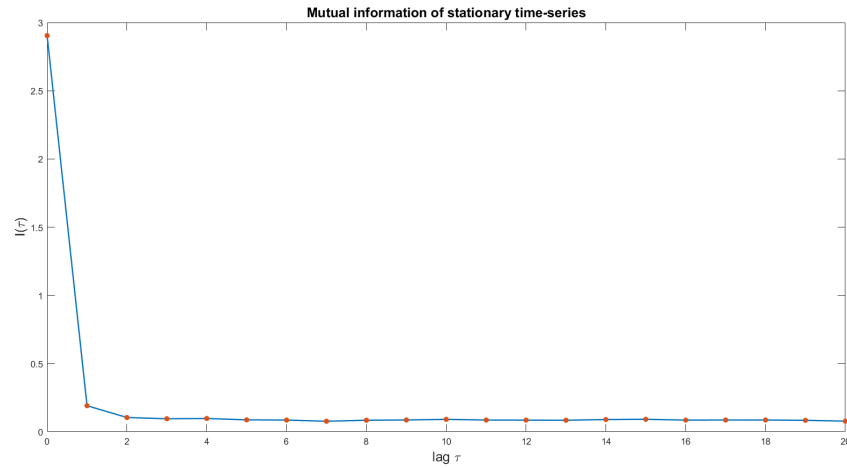
Επιθυμούμε να προσαρμόσουμε κάποιο μη-γραμμικό μοντέλο στη στάσιμη χρονοσειρά. Για το σκοπό αυτό, θα υποθέσουμε ότι η χρονοσειρά προέρχεται από κάποιο μη-γραμμικό δυναμικό σύστημα και θα χρησιμοποιήσουμε ανακατασκευή του χώρου καταστάσεων. Για την ανακατασκευή είναι απαραίτητο να εκτιμήσουμε τις παραμέτρους τ και m , δηλαδή την υστέρηση και την διάσταση εμβύθισης.

9.1 Εκτίμηση υστέρησης τ

Για την εκτίμηση της υστέρησης τ θα σχηματίσουμε τα διαγράμματα της συνάρτησης αυτοσυσχέτισης και της αμοιβαίας πληροφορίας σε συνάρτηση με την υστέρηση για τη στάσιμη χρονοσειρά.



Σχήμα 20: Συνάρτηση αυτοσυσχέτισης σε σχέση με την υστέρηση

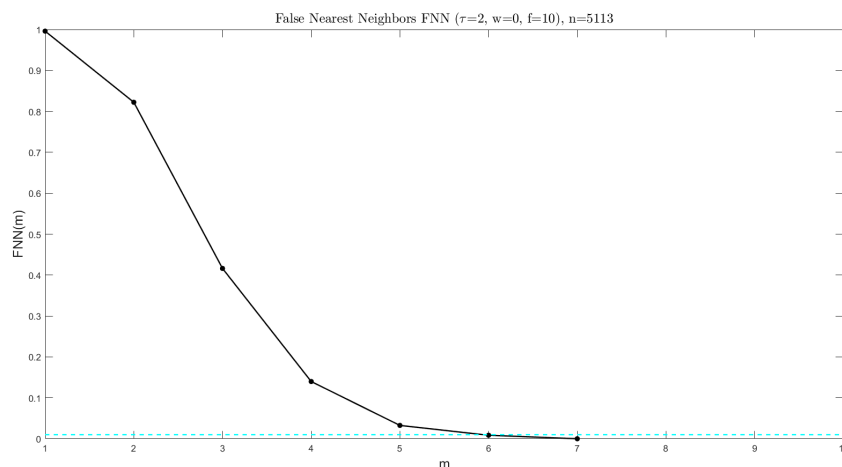


Σχήμα 21: Συνάρτηση αμοιβαίας πληροφορίας σε σχέση με την υστέρηση

Από το διάγραμμα της συνάρτησης αυτοσυσχέτισης, θα επιλέγαμε την πρώτη υστέρηση στην οποία η αυτοσυσχέτιση βρίσκεται κάτω από την τιμή $r(\tau) = 1/e$, δηλαδή την υστέρηση $\tau = 2$. Επιπλέον, μελετώντας το διάγραμμα της αμοιβαίας πληροφορίας, πρέπει να επιλέξουμε την πρώτη υστέρηση στην οποία παρουσιάζεται τοπικό ελάχιστο. Εδώ η αμοιβαία πληροφορία δε φαίνεται να παρουσιάζει κάποιο ευδιάκριτο τοπικό ελάχιστο. Ωστόσο, είναι φανερό ότι μειώνεται με υψηλότερο ρυθμό μέχρι την υστέρηση $\tau = 2$, ενώ μετά την υστέρηση 2 μειώνεται με πολύ μικρό ρυθμό και παραμένει σχεδόν στο ίδιο επίπεδο. Συνεπώς, σύμφωνα με την αμοιβαία πληροφορία θα επιλέγαμε, επίσης, την υστέρηση $\tau = 2$.

9.2 Εκτίμηση διάστασης εμβύθισης m

Για την εκτίμηση της διάστασης εμβύθισης, χρησιμοποιούμε τη μέθοδο των Ψευδών Κοντινότερων Γειτόνων (False Nearest Neighbors- FNN). Είναι σημαντικό να τονίσουμε ότι η επιλογή του m επηρεάζεται από την επιλογή που έχουμε κάνει για την υστέρηση και είναι ίση με $\tau = 2$.



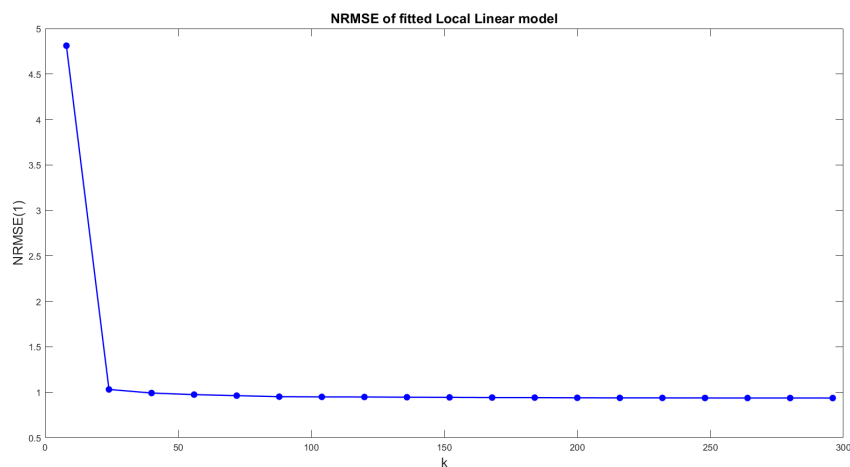
Σχήμα 22: Ποσοστό Ψευδών Πλησιέστερων Γειτόνων σε συνάρτηση με την διάσταση εμφύθισης

Επιλέγουμε την πρώτη τιμή του m η οποία οδηγεί σε ποσοστό ψευδών πλησιέστερων γειτόνων κάτω του 1%, δηλαδή $m = 6$.

Αφού έχουμε εκτιμήσεις τις παραμέτρους της ανακατασκευής ($\tau = 2, m = 6$), θα δοκιμάσουμε να προσαρμόσουμε ένα τοπικό μοντέλο στη στάσιμη χρονοσειρά. Συγκεκριμένα, θα προσαρμόσουμε τοπικό γραμμικό μοντέλο (Local Linear Model) και τοπικό μοντέλο μέσου όρου (Local Average Model). Και στις δύο περιπτώσεις, χρειάζεται να εκτιμήσουμε το πλήθος K των κοντινότερων γειτονικών σημείων, που χρησιμοποιούνται για την πρόβλεψη του τοπικού μοντέλου σε κάποιο σημείο της χρονοσειράς. Για το σκοπό αυτό, θα εκτιμήσουμε πολλαπλά μοντέλα για διαφορετικές τιμές του K και θα επιλέξουμε εκείνη την τιμή που ελαχιστοποιεί το NRMSE.

9.3 Τοπικό γραμμικό μοντέλο

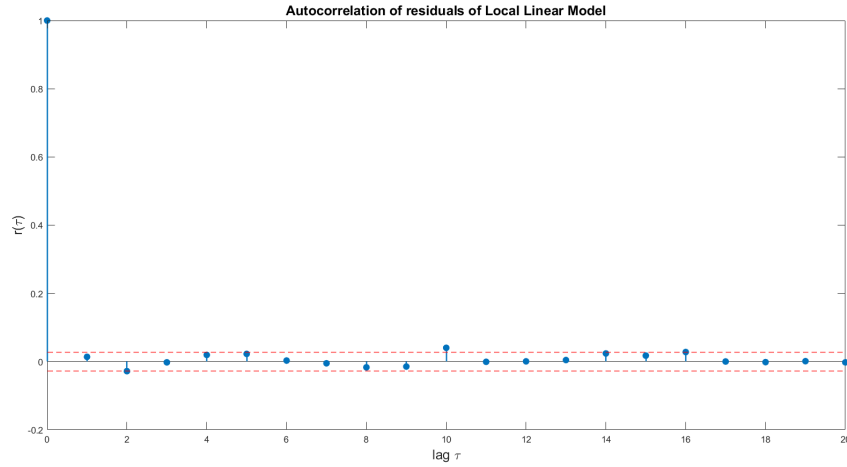
Κατά την προσαρμογή του τοπικού γραμμικού μοντέλου, υπολογίστηκε η χαμπύλη του NRMSE σε συνάρτηση με το k για εύρος τιμών 8 έως 300, με βήμα 16.



Σχήμα 23: NRMSE σε συνάρτηση με το πλήθος πλησιέστερων γειτόνων k για το τοπικό γραμμικό μοντέλο

Παρατηρούμε ότι η χαμπύλη του NRMSE για 1 βήμα μπροστά αρχικά μειώνεται με μεγαλύτερο ρυθμό και στη συνέχεια μειώνεται με πιο αργό ρυθμό. Συνεπώς, δεν υπάρχει ένα ξεκάθαρο τοπικό ελάχιστο της χαμπύλης. Επιλέγουμε την πρώτη τιμή του k η οποία οδηγεί σε τιμή $NRMSE < 0.94$.

Προσαρμόζουμε τοπικό γραμμικό μοντέλο με τιμή $k = 200$ και υπολογίζουμε τα υπόλοιπα της προσαρμογής.

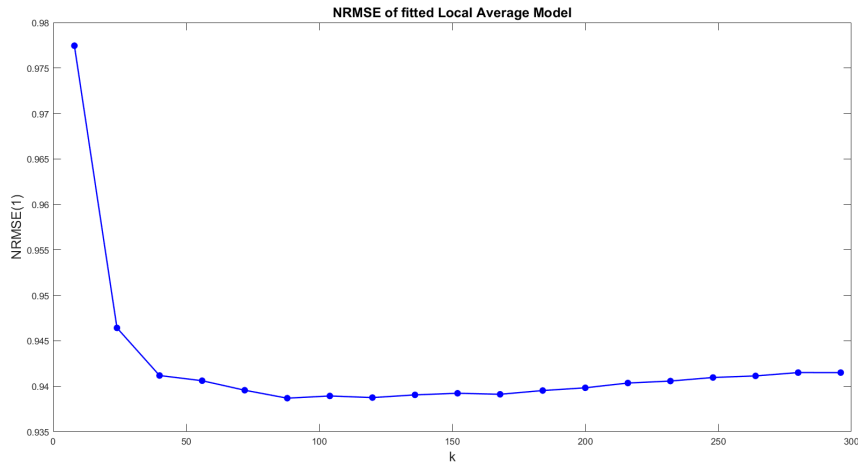


Σχήμα 24: Συνάρτηση αυτοσυσχέτισης του τοπικού γραμμικού μοντέλου

Από τον στατιστικό ελέγχο Portmanteau (Σχήμα 30) για τα υπόλοιπα της προσαρμογής, διαπιστώνουμε ότι το στατιστικό ελέγχου Q βρίσκεται κάτω από την κρίσιμη τιμή της κατανομής για μέγιστη υστέρηση έως και 9. Συνεπώς, μπορούμε να υποθέσουμε ότι η χρονοσειρά των υπολοίπων είναι προσεγγιστικά λευκός θόρυβος.

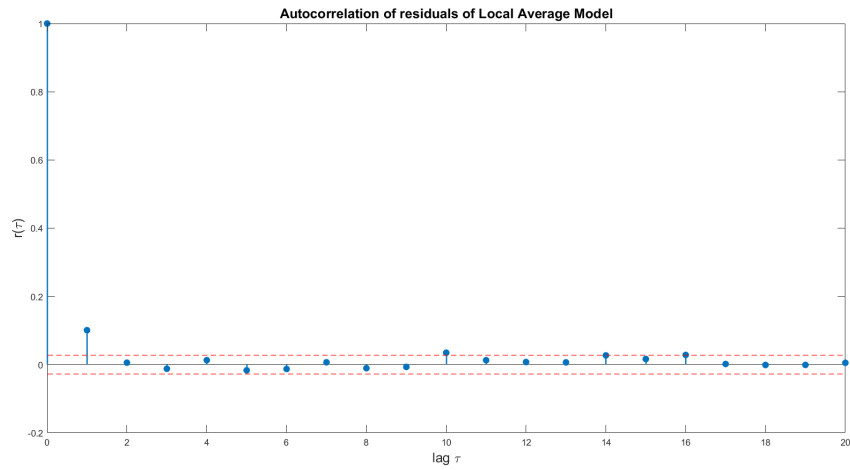
9.4 Τοπικό μοντέλο μέσου όρου

Προσαρμόζουμε, επιπλέον, τοπικό μοντέλο μέσου όρου για τις ίδιες τιμές του k και προκύπτει η παρακάτω καμπύλη του NRMSE για ένα 1 βήμα μπροστά.



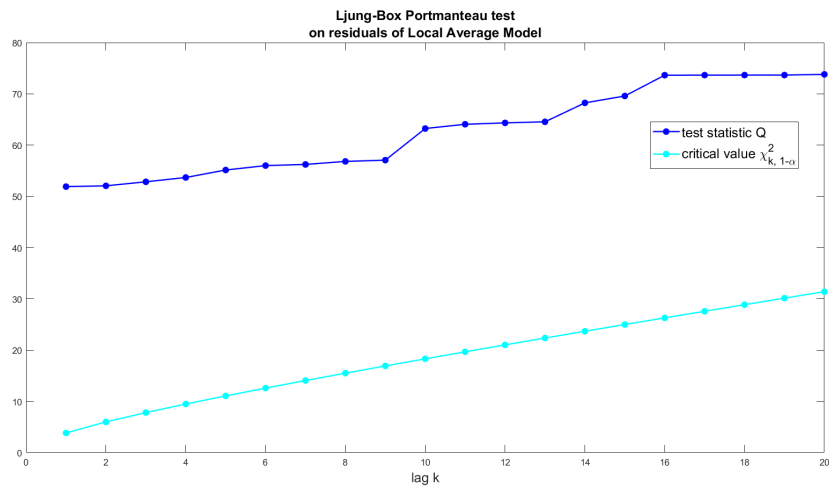
Σχήμα 25: NRMSE σε συνάρτηση με το πλήθος πλησιέστερων γειτόνων k για το τοπικό μοντέλο μέσου όρου

Η καμπύλη του $NRMSE(1)$ παρουσιάζει ένα τοπικό ελάχιστο για την τιμή $k = 88$. Προσαρμόζουμε το τοπικό γραμμικό μοντέλο για αυτή την τιμή του k και υπολογίζουμε τα υπόλοιπα της προσαρμογής.

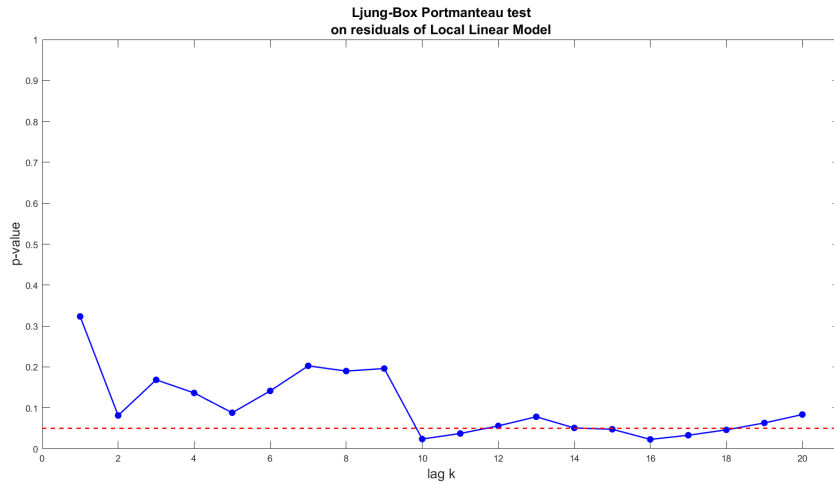


Σχήμα 26: Συνάρτηση αυτοσυσχέτισης του τοπικού μοντέλου μέσου όρου

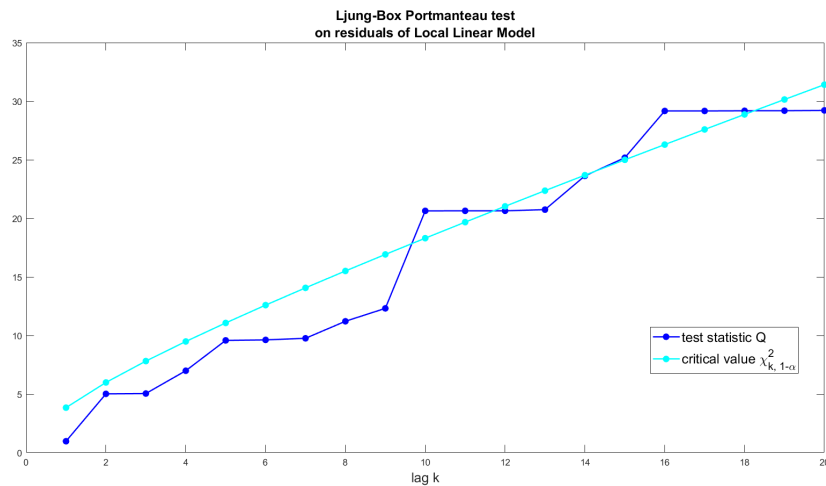
Παρατηρούμε ότι η τιμή της αυτοσυσχέτισης είναι σημαντική για $\tau = 1$. Εφαρμόζοντας, επιπλέον, τον στατιστικό έλεγχο Portmanteau (Σχήμα 27), διαπιστώνουμε ότι η τιμή του στατιστικού ελέγχου Q βρίσκεται πάνω από την κρίσιμη τιμή για κάθε τιμή της μέγιστης υστέρησης k . Αυτό σημαίνει ότι ο έλεγχος απορρίπτει την μηδενική υπόθεση πως η χρονοσειρά των υπολοίπων είναι λευκός θόρυβος. Συνεπώς, το τοπικό μοντέλο μέσου όρου που προσαρμόσαμε δεν είναι κατάλληλο για τη χρονοσειρά.



Σχήμα 27: Στατιστικός έλεγχος ανεξαρτησίας υπολοίπων του τοπικού μοντέλου μέσου όρου



Σχήμα 28: p -τιμές ελέγχου Portmanteau



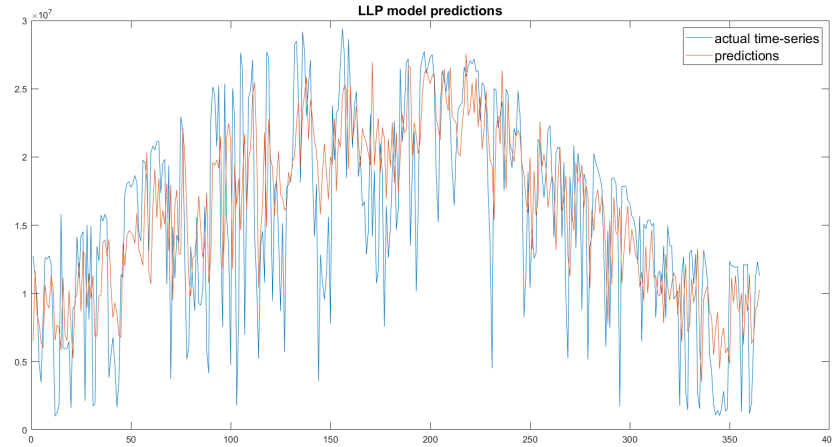
Σχήμα 29: Στατιστικό ελέγχου Portmanteau

Σχήμα 30: Στατιστικός έλεγχος ανεξαρτησίας υπολοίπων του τοπικού γραμμικού μοντέλου

10 Πρόβλεψη με μη-γραμμικό μοντέλο χρονοσειράς

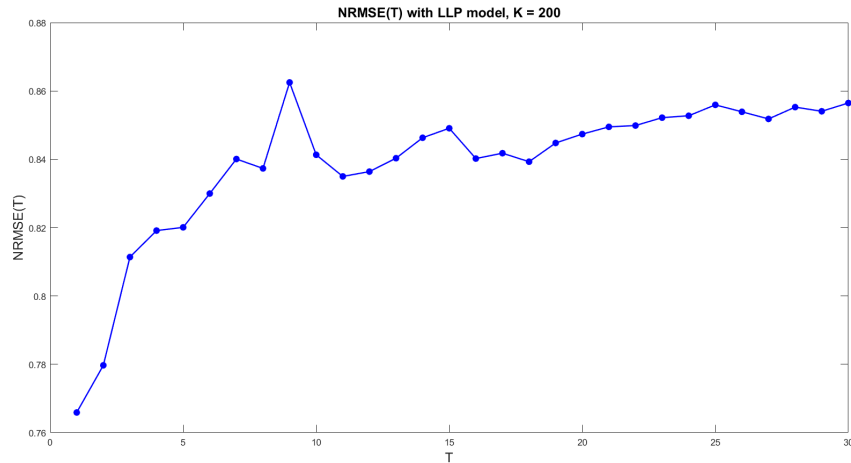
Σκοπός μας είναι να προβλέψουμε τις τιμές της χρονοσειράς για το έτος 2007. Ακολουθούμε την ίδια διαδικασία όπως και για την πρόβλεψη με γραμμικό μοντέλο, δηλαδή χωρίζουμε το σύνολο των παρατηρήσεων της χρονοσειράς σε σύνολο εκπαίδευσης και ελέγχου και απαλείφουμε την εποχικότητα με βάση την εκτίμηση από το σύνολο εκπαίδευσης. Στη συνέχεια, χρησιμοποιούμε το τοπικό γραμμικό μοντέλο που βρέθηκε κατάλληλο, με $k = 200$, για να εξάγουμε τις προβλέψεις για $T = 1, \dots, 30$ βήματα μπροστά.

Στο παρακάτω σχήμα, φαίνονται οι προβλέψεις για το έτος 2007 για ένα βήμα μπροστά.



Σχήμα 31: Προβλέψεις στο σύνολο ελέγχου με τοπικό γραμμικό μοντέλο για 1 βήμα μπροστά

Υπολογίζουμε το NRMSE των προβλέψεων για $T = 1, \dots, 30$ βήματα μπροστά και προκύπτει το παρακάτω διάγραμμα:



Σχήμα 32: NRMSE προβλέψεων σε συνάρτηση με το T

Για ένα βήμα μπροστά, έχουμε $NRMSE(1) = 0.765903$. Όπως και στην περίπτωση του γραμμικού μοντέλου, το NRMSE αυξάνεται καθώς αυξάνεται το T , καθώς οι προβλέψεις χειροτερεύουν για μεγαλύτερο χρονικό ορίζοντα πρόβλεψης.

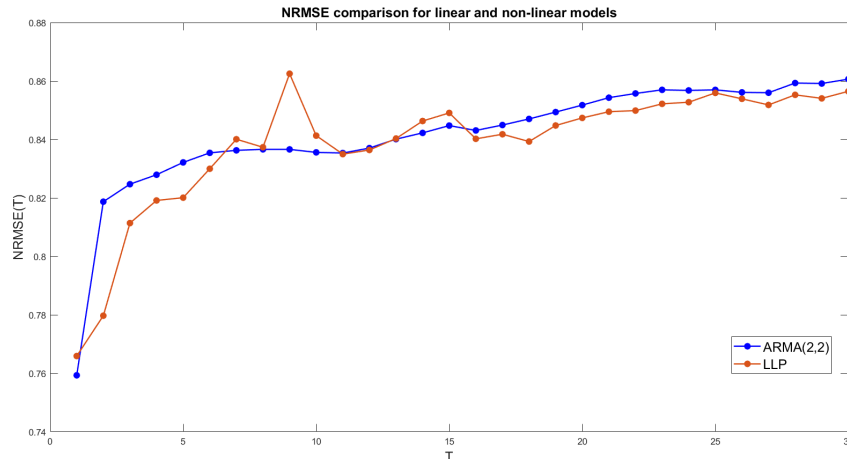
11 Σύγκριση μοντέλων

Θα συγκρίνουμε τα τρία μοντέλα που χρησιμοποιήσαμε για την πρόβλεψη της χρονοσειράς: το γραμμικό μοντέλο, το μοντέλο περιοδικής συνάρτησης και το μη-γραμμικό μοντέλο.

Παραθέτουμε έναν πίνακα που παρουσιάζει το $NRMSE(1)$ για τα τρία μοντέλα. Παρατηρούμε ότι το γραμμικό και μη-γραμμικό μοντέλο έχουν πολύ κοντινές τιμές του στατιστικού, ενώ το μοντέλο περιοδικής συνάρτησης έχει σχετικά υψηλότερη τιμή. Δηλαδή, το μοντέλο περιοδικής συνάρτησης παρουσιάζει τη χειρότερη απόδοση στις προβλέψεις, δεδομένου ότι δεν λαμβάνει υπόψη τις συσχετίσεις, γραμμικές ή μη-γραμμικές, που μπορεί να υπάρχουν στη χρονοσειρά μετά την αφαίρεση της περιοδικής συνάρτησης. Επιπρόσθετα, το γραμμικό και μη-γραμμικό μοντέλο έχουν παρόμοια απόδοση στις προβλέψεις για τις περισσότερες υστερήσεις, όπως απεικονίζεται και στο Σχήμα 33, όπου φαίνεται πως οι δύο καμπύλες ακολουθούν σχεδόν η μία την άλλη. Καταληκτικά, σημειώνουμε ότι κανένα από τα μοντέλα δεν επιτυγχάνει ικανοποιητικές προβλέψεις, καθώς όλα χαρακτηρίζονται από σχετικά υψηλές τιμές του NRMSE για όλα τα βήματα πρόβλεψης.

	NRMSE(1)
γραμμικό μοντέλο $ARMA(2, 2)$	0.759306
μοντέλο περιοδικής συνάρτησης	0.830344
τοπικό γραμμικό μοντέλο $\tau = 2, m = 6, k = 200$	0.765903

Πίνακας 2: Στατιστικό NRMSE για τα μοντέλα πρόβλεψης



Σχήμα 33: Σύγκριση NRMSE για το γραμμικό και μη-γραμμικό μοντέλο

12 Συμπεράσματα

Από την παραπάνω ανάλυση, μπορούμε να συμπεράνουμε αρχικά πως το σύστημα που παράγει τη χρονοσειρά είναι **περιοδικό** με περίοδο περίπου 365 ημερών, όπως φαίνεται από απλή παρατήρηση της χρονοσειράς. Στη συνέχεια, εφόσον το μοντέλο ARMA δεν προσαρμόζεται καλά στην χρονοσειρά, καταλήγουμε στο ότι είναι πιθανόν να υπάρχουν μη-γραμμικές συσχετίσεις μεταξύ των παρατηρήσεων της χρονοσειράς, τις οποίες το γραμμικό μοντέλο δεν καταφέρνει να συλλάβει. Με βάση αυτήν την παρατήρηση, μπορούμε να θεωρήσουμε ότι το σύστημα είναι **μη-γραμμικό**. Παρόλα αυτά, το μη-γραμμικό μοντέλο (τοπικό γραμμικό) που προσαρμόσαμε στη χρονοσειρά αποδείχτηκε επίσης μη ικανοποιητικό στις προβλέψεις και παρουσίασε παρόμοια συμπεριφορά με το ARMA στην καμπύλη του NRMSE. Έτσι, μπορούμε να πούμε ότι το σύστημα έχει **υψηλή πολυπλοκότητα** σε σύγκριση με αυτή που χρησιμοποιήθηκε και πιθανότατα **υψηλή διάσταση**. Τέλος, με βάση τα παραπάνω, υποθέτουμε ότι το σύστημα είναι **αιτιοκρατικό**, δεδομένων των γραμμικών και μη-γραμμικών συσχετίσεων στη χρονοσειρά.

13 Διάρθρωση κώδικα

Για την ανάλυση χρησιμοποιήθηκε η έκδοση 2023a του MATLAB.

Τα ερωτήματα 1-6 που αφορούν τη γραμμική ανάλυση της χρονοσειράς απαντώνται στο αρχείο **linear_analysis.m**. Τα ερωτήματα 7-8 που αφορούν τη μη-γραμμική ανάλυση της χρονοσειράς απαντώνται στο αρχείο **nonlinear_analysis.m**.

Επιπλέον, υπάρχουν οι εξής βοηθητικές συναρτήσεις:

- **autocorrelation.m**: εκτίμηση αυτοσυσχέτισης
- **acf2pacf.m**: εκτίμηση μερικής αυτοσυσχέτισης
- **portmanteauLB.m**: στατιστικός έλεγχος ανεξαρτησίας
- **seasonal_components.m**: εκτίμηση εποχικότητας
- **remove_seasonality.m**: εκτίμηση και απαλοιφή εποχικότητας
- **fitARMA.m**: προσαρμογή γραμμικού μοντέλου ARMA
- **predictARMAnrmse.m**: πρόβλεψη με γραμμικό μοντέλο ARMA

- **nrmse.m**: υπολογισμός στατιστικού NRMSE
- **snrmse.m**: υπολογισμός στατιστικού SNRMSE
- **mutual_information.m**: υπολογισμός αμοιβαίας πληροφορίας σε συνάρτηση με την υστέρηση
- **falsenearest.m**: μέθος των ψευδών κοντινότερων γειτόνων για την εκτίμηση της διάστασης εμβύθισης
- **AddNoise.m**: βοηθητική συνάρτηση για τη συνάρτηση falsenearest.m
- **localfitnrmse.m**: προσαρμογή τοπικού μοντέλου
- **localpredictnrmse.m**: πρόβλεψη με τοπικό μοντέλο