



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης  
Πολυτεχνική Σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Ηλεκτρονικής & Υπολογιστών

## Αναφορά Υπολογιστικής Εργασίας στις Χρονοσειρές

**Μάθημα:** Χρονοσειρές

**Συντάκτης:** Θανάσης Χαρισούδης  
ΑΕΜ: 9026

**Διδάσκων:** κ. Δημήτρης Κουγιουμτζής  
Καθηγητής Α.Π.Θ.

# Περιεχόμενα

<b>1 Διερεύνηση Χρονοσειράς Α</b>	<b>1</b>
1.1 Αρχική Ανάλυση . . . . .	1
1.2 Απαλοιφή Στοχαστικής Τάσης . . . . .	3
1.3 Απαλοιφή Στοχαστικής Τάσης & Σταθεροποίηση Διασποράς . . . . .	5
<b>2 Προσαρμογή Γραμμικού Μοντέλου</b>	<b>9</b>
2.1 Προσαρμογή ARIMA(p,d,q) . . . . .	9
2.1.1 Προσαρμογή ARIMA(p,1,q) παίρνοντας τις πρώτες διαφορές . . . . .	9
2.1.2 Προσαρμογή ARMA(p,q) στη στάσιμη χρονοσειρά . . . . .	10
2.2 Διάγνωση καταλληλότητας του μοντέλου ARIMA(0,1,1) . . . . .	11
2.2.1 Τελικό μοντέλο τύπου MA(1) για τη στάσιμη χρονοσειρά . . . . .	11
2.2.2 Διάγνωση Καταλληλότητας & Σφάλματα Προσαρμογής . . . . .	12
<b>3 Διερεύνηση Χρονοσειράς Β</b>	<b>15</b>
3.1 Αρχική Ανάλυση . . . . .	15
3.2 Απαλοιφή Στοχαστικής Τάσης . . . . .	17
3.3 Απαλοιφή Στοχαστικής Τάσης & Σταθεροποίηση Διασποράς . . . . .	19
3.4 Εκτίμηση & Απαλοιφή Εποχικότητας . . . . .	22
3.4.1 Εκτίμηση Εποχικού Όρου . . . . .	22
3.4.2 Απαλοιφή Εποχικού Όρου . . . . .	22
3.4.3 Σύνοψη Διαδικασίας . . . . .	24
3.5 Προσαρμογή ARMA(p,q) . . . . .	24
3.5.1 Προσαρμογή ARMA(p,q) στη στάσιμη χρονοσειρά . . . . .	25
3.6 Διάγνωση καταλληλότητας του μοντέλου ARMA(9,9) . . . . .	26
3.6.1 Τελικό μοντέλο τύπου ARMA(9,9) για τη στάσιμη χρονοσειρά . . . . .	26
3.6.2 Διάγνωση Καταλληλότητας & Σφάλματα Προσαρμογής . . . . .	27
3.7 Διάγνωση καταλληλότητας του μοντέλου ARMA(4,4) . . . . .	29
3.7.1 Τελικό μοντέλο τύπου ARMA(4,4) για τη στάσιμη χρονοσειρά . . . . .	29
3.7.2 Διάγνωση Καταλληλότητας & Σφάλματα Προσαρμογής . . . . .	30
3.7.3 Συμπερασματικά σχόλια . . . . .	32
<b>4 Εντοπισμός Σημείων Αλλαγής της Χρονοσειράς Α</b>	<b>33</b>
4.1 Τρόπος Επιλογής Σημείων Αλλαγής . . . . .	33
4.2 Αρχική Εφαρμογή . . . . .	34
4.3 Επιλογή Βέλτιστων Παραμέτρων . . . . .	36
4.4 Εφαρμογή με Βέλτιστες Παραμέτρους . . . . .	38
4.4.1 Αναπροσαρμογή όταν βρεθεί σημείο αλλαγής . . . . .	38
4.4.2 Αναπροσαρμογή σε κάθε χρονική στιγμή . . . . .	40

4.4.3 Χωρίς Αναπροσαρμογή . . . . .	41
<b>5 Εντοπισμός Σημείων Αλλαγής της Χρονοσειράς B</b>	<b>44</b>
5.1 Αρχική Εφαρμογή . . . . .	44
5.2 Επιλογή Βέλτιστων Hyperparameters . . . . .	46
5.3 Εφαρμογή με Βέλτιστες Παραμέτρους . . . . .	48
5.3.1 Αναπροσαρμογή όταν βρεθεί σημείο αλλαγής . . . . .	48
5.3.2 Αναπροσαρμογή σε κάθε χρονική στιγμή . . . . .	50
5.3.3 Χωρίς Αναπροσαρμογή . . . . .	51
5.3.4 Συμπερασματικά Σχόλια . . . . .	53
<b>6 Εντοπισμός Σημείων Αλλαγής με Μη-Γραμμικά Μοντέλα</b>	<b>54</b>
6.1 Ανακτασκευή του χώρου καταστάσεων . . . . .	54
6.1.1 Παράμετροι ανακατασκευής για τη Χρονοσειρά A . . . . .	54
6.1.1.1 Υστέρηση, $\tau$ . . . . .	54
6.1.1.2 Διάσταση Εμβύθινσης Ελκυστή, $m$ . . . . .	55
6.1.2 Παράμετροι ανακατασκευής για τη Χρονοσειρά B . . . . .	57
6.1.2.1 Υστέρηση, $\tau$ . . . . .	57
6.1.2.2 Διάσταση Εμβύθινσης Ελκυστή, $m$ . . . . .	57
6.2 Προσαρμογή Τοπικού Μοντέλου Κοντινότερων Γειτόνων . . . . .	59
6.2.1 Επιλογή του $K$ για τη στάσιμη χρονοσειρά A . . . . .	59
6.2.2 Επιλογή του $K$ για τη στάσιμη χρονοσειρά B . . . . .	60
6.3 Εφαρμογή Τοπικού Μοντέλου Κοντινότερων Γειτόνων για Εξαγωγή Σημείων Αλλαγής	61
6.3.1 Εφαρμογή στη στάσιμη χρονοσειρά A . . . . .	62
6.3.1.1 Επιλογή Βέλτιστων Παραμέτρων . . . . .	62
6.3.1.2 Εφαρμογή Μεθόδου με Βέλτιστες Παραμέτρους . . . . .	64
6.3.2 Εφαρμογή στη στάσιμη χρονοσειρά B . . . . .	68
6.3.2.1 Επιλογή Βέλτιστων Παραμέτρων . . . . .	68
6.3.2.2 Εφαρμογή Μεθόδου με Βέλτιστες Παραμέτρους . . . . .	70
<b>7 Συμπερασματικά Σχόλια</b>	<b>75</b>

# Κατάλογος Σχημάτων

1.1	Διάγραμμα ιστορίας της χρονοσειράς A, $\{Y_a(t)\}$ . Σε όλα τα διαγράμματα ιστορίας που θα παρουσιαστούν, όπως και σε αυτό, με cyan απεικονίζεται μία προσέγγιση της τάσης ως αποτέλεσμα εφαρμογής moving-average smoothing τάξης 7, MA(7), δηλαδή μετριάζοντας με εβδομαδιαίο «παράθυρο».	2
1.2	Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς A, $r_y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης	2
1.3	Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς A, $\phi_y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης	3
1.4	Διάγραμμα ιστορίας της χρονοσειράς των πρώτων διαφορών, $\{BY_a(t)\}$	4
1.5	Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών, $r_{BY_a}(\tau)$	4
1.6	Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών, $\phi_{BY_a}(\tau)$	5
1.7	Διάγραμμα ιστορίας της χρονοσειράς των τετραγωνικών ριζών, $\{\sqrt{Y_a}(t)\}$	6
1.8	Διάγραμμα ιστορίας της στάσιμης χρονοσειράς, $\{X_a(t)\}$ , ως οι πρώτες διαφορές των τετραγωνικών ριζών της αρχικής χρονοσειράς	7
1.9	Διάγραμμα (δειγματικής) αυτοσυσχέτισης της στάσιμης χρονοσειράς, $r_x(\tau)$	7
1.10	Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της στάσιμης χρονοσειράς, $\phi_x(\tau)$	8
2.1	Διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (2.3), $\{X_{a,res}(t) = \hat{z}(t)\}$	12
2.2	Διάγραμμα δειγματικής αυτοσυσχέτισης της σειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (2.3), $r_{\hat{z}_a}(\tau)$	13
2.3	Διάγραμμα των p-values του στατιστικού ελέγχου ανεξαρτησίας Portmanteau (Ljung & Box test) των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (2.3). Σημειώνεται με διακεκομένη γραμμή το όριο απόφασης όπου όπως φαίνεται η μηδενική υπόθεση $H_0$ (η σειρά των υπολοίπων είναι iid) δεν απόρριπτεται για καμία από τις 30 υστερήσεις.	13
2.4	Διάγραμμα ιστορίας της αρχικής χρονοσειράς προβολών του βίντεο (μπλε) A καθώς και τις προβλέψεις ενός βήματος μπροστά αυτής με βάση το προσαρμοσμένο μοντέλο MA(1) και τη σχέση (2.4) (κόκκινο).	14
3.1	Διάγραμμα ιστορίας της χρονοσειράς B, $\{Y_b(t)\}$ , μαζί με τη καμπύλη MA(7) smoothing	15
3.2	Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς B, $r_y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης	16
3.3	Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς B, $\phi_y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης	16
3.4	Διάγραμμα ιστορίας της χρονοσειράς των πρώτων διαφορών, $\{BY_b(t)\}$	17

3.5 Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών, $r_{BY_b}(\tau)$	18
3.6 Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών, $\phi_{BY_b}(\tau)$	18
3.7 Διάγραμμα ιστορίας της χρονοσειράς των τετραγωνικών ριζών, $\{sqrt(Y_b)(t)\}$	19
3.8 Διάγραμμα ιστορίας της στάσιμης χρονοσειράς, $\{X_b(t)\}$ , ως οι πρώτες διαφορές των τετραγωνικών ριζών της αρχικής χρονοσειράς	20
3.9 Διάγραμμα (δειγματικής) αυτοσυσχέτισης της στάσιμης χρονοσειράς, $r_x(\tau)$	21
3.10 Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της στάσιμης χρονοσειράς, $\phi_x(\tau)$	21
3.11 Πρώτη περίοδος του εποχικού όρου της $\{X_b(t)\}$ , $\tilde{s}_b$	22
3.12 Διάγραμμα ιστορίας της χρονοσειράς $\{X_b(t)\}$ απαλλαγμένης από εποχικότητα, $\{X_{b_{deseasoned}}(t)\}$	23
3.13 Διάγραμμα (δειγματικής) αυτοσυσχέτισης της απαλλαγμένης από εποχικότητα χρονοσειράς, $r_x(\tau)$	23
3.14 Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της απαλλαγμένης από εποχικότητα χρονοσειράς, $\phi_x(\tau)$	24
3.15 Διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (3.8), $\{X_{b_{deseasoned,res}}(t) = \hat{z}(t)\}$	27
3.16 Διάγραμμα δειγματικής αυτοσυσχέτισης της σειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (3.8), $r_{\hat{z}_b}(\tau)$	28
3.17 Διάγραμμα των p-values του στατιστικού ελέγχου ανεξαρτησίας Portmanteau (Ljung & Box test) των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (3.8). Σημειώνεται με διακεκομμένη γραμμή το όριο απόφασης όπου όπως φαίνεται η μηδενική υπόθεση $H_0$ (η σειρά των υπολοίπων είναι iid) δεν απόρριπτεται για καμία από τις 30 υστερήσεις.	28
3.18 Διάγραμμα ιστορίας της αρχικής χρονοσειράς προβολών του βίντεο B (μπλε) καθώς και τις προβλέψεις ενός βήματος μπροστά αυτής με βάση το προσαρμοσμένο μοντέλο ARMA(9, 9) και τη σχέση (3.9) (κόκκινο).	29
3.19 Διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων της προσαρμογής του μοντέλου ARMA(4, 4) της σχέσης (3.11), $\{X_{b_{deseasoned,res(4,4)}}(t) = \hat{z}(t)\}$	30
3.20 Διάγραμμα δειγματικής αυτοσυσχέτισης της σειράς των υπολοίπων της προσαρμογής του μοντέλου ARMA(4, 4) της σχέσης (3.11), $r_{\hat{z}_b}(\tau)$	31
3.21 Διάγραμμα των p-values του στατιστικού ελέγχου ανεξαρτησίας Portmanteau (Ljung & Box test) των υπολοίπων της προσαρμογής του μοντέλου ARMA(4, 4) της σχέσης (3.11). Σημειώνεται με διακεκομμένη γραμμή το όριο απόφασης όπου όπως φαίνεται η μηδενική υπόθεση $H_0$ (η σειρά των υπολοίπων είναι iid) δεν απόρριπτεται για καμία από τις 30 υστερήσεις.	31
3.22 Διάγραμμα ιστορίας της αρχικής χρονοσειράς προβολών του βίντεο B (μπλε) καθώς και τις προβλέψεις ενός βήματος μπροστά αυτής με βάση το προσαρμοσμένο μοντέλο ARMA(4, 4) και τη σχέση (3.12) (κόκκινο).	32
4.1 Τιμές στατιστικού $S_n$ για έως και 5 βήματα μπροστά πρόβλεψη με MA(1) της στάσιμης χρονοσειράς $\{X_a(t)\}$ . Σημειώνονται επίσης το κριτήριο απόφασης, $a$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο). Το $n$ ξεκινάει από το 400 και άρα το $S_n$ δεν ορίζεται για τιμές $n < n_0 = 400$ .	35
4.2 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing)	36

4.3 Αριθμός σημείων αλλαγής, $ MCP $ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\hat{M}_{std}$ του ορίου απόφασης, $a$ . . . . .	37
4.4 NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs), για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\hat{M}_{std}$ του ορίου απόφασης, $a$ . . . . .	37
4.5 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με $MA(1)$ της στάσιμης χρονοσειράς $\{X_a(t)\}$ και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.800, 28.9sec] . . . . .	39
4.6 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$ smoothing) - επιλογή «c» . . . . .	39
4.7 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με $MA(1)$ της στάσιμης χρονοσειράς $\{X_a(t)\}$ και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.802, 86.1sec] . . . . .	40
4.8 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$ smoothing) - επιλογή «b» . . . . .	41
4.9 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με $MA(1)$ της στάσιμης χρονοσειράς $\{X_a(t)\}$ και για επιλογή αναπροσαρμογής «a» (χωρίς αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.789, 27.2sec] . . . . .	42
4.10 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$ smoothing) - επιλογή «a» . . . . .	42
5.1 Τιμές στατιστικού $S_n$ για έως και 5 βήματα μπροστά πρόβλεψη με $ARMA(4, 4)$ της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ . Σημειώνονται επίσης το κριτήριο απόφασης, $a$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο). Το $n$ ξεκινάει από το 400 και άρα το $S_n$ δεν ορίζεται για τιμές $n < n_0 = 400$ . . . . .	45
5.2 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$ smoothing) . . . . .	46
5.3 Αριθμός σημείων αλλαγής, $ MCP $ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\hat{M}_{std}$ του ορίου απόφασης, $a$ . . . . .	47
5.4 NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs), για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\hat{M}_{std}$ του ορίου απόφασης, $a$ . . . . .	47

5.5 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.892, 76.2sec] . . . . .	49
5.6 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «c» . . . . .	49
5.7 Τιμές στατιστικού $S_n$ για έως και 5 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.893, 244.1sec] . . . . .	50
5.8 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «b» . . . . .	51
5.9 Τιμές στατιστικού $S_n$ για έως και 5 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ και για επιλογή αναπροσαρμογής «a» (χωρίς αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.899, 61.1sec] . . . . .	52
5.10 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «a» . . . . .	52
 6.1 Διάγραμμα αμοιβαίας πληροφορίας της στάσιμης χρονοσειράς $\{X_a(t)\}$ για επιλογή της υστέρησης $\tau$ . . . . .	55
6.2 Διάγραμμα διασποράς $x_i - x_{i-1} - x_{i-2}$ της στάσιμης χρονοσειράς $\{X_a(t)\}$ . . . . .	56
6.3 Διάγραμμα μεταβολής των FNNs ως προς τη διάσταση εμβύθινσης, $m$ , για τη στάσιμη χρονοσειρά $\{X_a(t)\}$ . . . . .	56
6.4 Διάγραμμα αμοιβαίας πληροφορίας της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ για επιλογή της υστέρησης $\tau$ . . . . .	57
6.5 Διάγραμμα διασποράς $x_i - x_{i-1} - x_{i-2}$ της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ . . . . .	58
6.6 Διάγραμμα μεταβολής των FNNs ως προς τη διάσταση εμβύθινσης, $m$ , για τη στάσιμη χρονοσειρά $\{X_{b_{deseasoned}}(t)\}$ . . . . .	58
6.7 Διάγραμμα μεταβολής του μέσου NRMSE των σφαλμάτων πρόβλεψης ως προς τον αριθμό των κοντινότερων γειτόνων που συμμετέχουν στη πρόβλεψη, $K$ , για τη στάσιμη χρονοσειρά $\{X_a(t)\}$ . . . . .	60
6.8 Διάγραμμα μεταβολής του μέσου NRMSE των σφαλμάτων πρόβλεψης ως προς τον αριθμό των κοντινότερων γειτόνων που συμμετέχουν στη πρόβλεψη, $K$ , για τη στάσιμη χρονοσειρά $\{X_{b_{deseasoned}}(t)\}$ . . . . .	61
6.9 Αριθμός σημείων αλλαγής της στάσιμης χρονοσειράς A,  MCP , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\bar{M}_{std}$ του ορίου απόφασης, $a$ , για παραμέτρους μη-γραμμικού μοντέλου $m = 5$ , $\tau = 3$ και $K = 16$ . . . . .	62

6.10 NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs) της στάσιμης χρονοσειράς A, για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\bar{J}_{std}$ του ορίου απόφασης, $a$ , για παραμέτρους μη-γραμμικού μοντέλου $m = 5$ , $\tau = 3$ και $K = 16$ . . . . .	63
6.11 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου $K=16$ κοντινότερων γειτόνων της στάσιμης χρονοσειράς $\{X_a(t)\}$ και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.92, 0.6sec] . . . . .	64
6.12 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «c» . . . . .	65
6.13 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου $K=16$ κοντινότερων γειτόνων της στάσιμης χρονοσειράς $\{X_a(t)\}$ και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.934, 0.62sec] . . . . .	65
6.14 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «b» . . . . .	66
6.15 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου $K=16$ κοντινότερων γειτόνων της στάσιμης χρονοσειράς $\{X_a(t)\}$ και για επιλογή αναπροσαρμογής «a» (χωρίες αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.910, 0.61sec] . . . . .	67
6.16 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_a(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «a» . . . . .	67
6.17 Αριθμός σημείων αλλαγής της στάσιμης χρονοσειράς B, $ MCP $ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\bar{J}_{std}$ του ορίου απόφασης, $a$ , για παραμέτρους μη-γραμμικού μοντέλου $m = 5$ , $\tau = 2$ και $K = 13$ . . . . .	68
6.18 NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs) της στάσιμης χρονοσειράς B, για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης, $T$ , και το $\bar{J}_{std}$ του ορίου απόφασης, $a$ , για παραμέτρους μη-γραμμικού μοντέλου $m = 5$ , $\tau = 2$ και $K = 13$ . . . . .	69
6.19 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου $K=13$ κοντινότερων γειτόνων της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης, $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.975, 0.61sec] . . . . .	70

6.20 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «c» . . . . .	71
6.21 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου K=13 κοντινότερων γειτόνων της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης, $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.979, 0.63sec] . . . . .	71
6.22 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «b» . . . . .	72
6.23 Τιμές στατιστικού $S_n$ για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου K=13 κοντινότερων γειτόνων της στάσιμης χρονοσειράς $\{X_{b_{deseasoned}}(t)\}$ και για επιλογή αναπροσαρμογής «a» (χωρίες αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης, $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία $n + T$ (μαύρο) - [NRMSE=0.975, 0.61sec] . . . . .	73
6.24 Διάγραμμα ιστορίας της αρχικής χρονοσειράς $\{Y_b(t)\}$ (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «a» . . . . .	73

# Κατάλογος Πινάκων

2.1 Αναζήτηση Πλέγματος με βάση τη μετρική AIC για διάφορες τιμές των τάξεων ( $p, q$ ). Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο. . . . .	10
2.2 Αναζήτηση Πλέγματος με βάση τη μετρική FPE για διάφορες τιμές των τάξεων ( $p, q$ ). Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο. . . . .	11
3.1 Αναζήτηση Πλέγματος με βάση τη μετρική AIC για διάφορες τιμές των τάξεων ( $p, q$ ). Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο. . . . .	25
3.2 Αναζήτηση Πλέγματος με βάση τη μετρική FPE για διάφορες τιμές των τάξεων ( $p, q$ ). Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο. . . . .	26

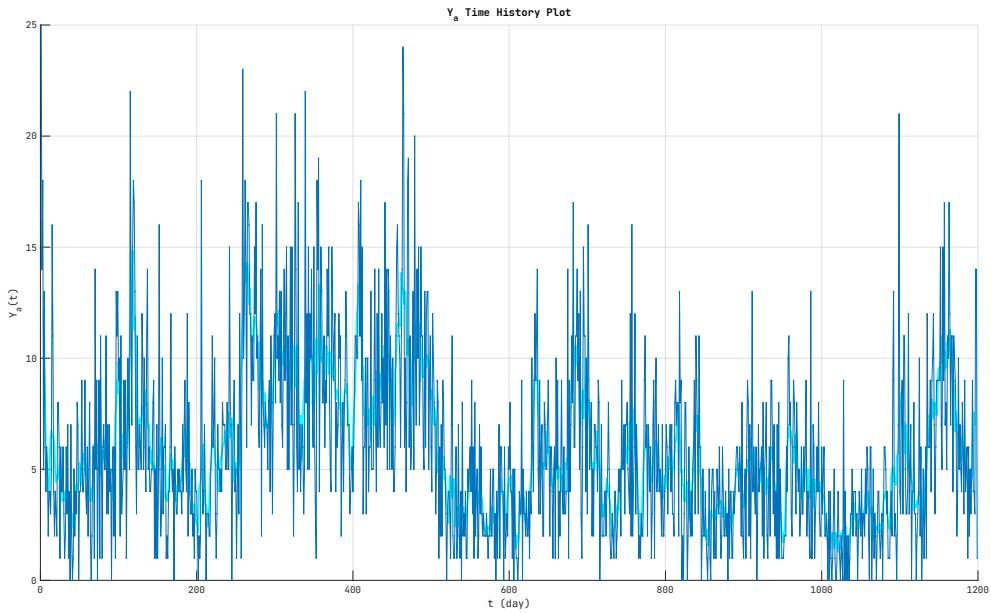
# **Βήμα 1**

## **Διερεύνηση Χρονοσειράς A**

Σκοπός της συγκεκριμένης εργασίας του μαθήματος «Χρονοσειρές» είναι να γίνει ανάλυση κάποιων χρονοσειρών από προβολές βίντεο με σκοπό τη δημιουργία αντιπροσωπευτικών μοντέλων αυτών και κατ' επέκταση, την πρόβλεψή τους. Η ομάδα μου είναι η ομάδα No. 3 και επομένως οι χρονοσειρές που θα αναλυθούν στα πλαίσια αυτής της εργασίας είναι αυτές με δείκτες 3 (η «A») και 13 (η «B»). Ακουλουθεί η αρχική ανάλυση της κάθε μίας χρονοσειράς ξεχωριστά, όπου δίνονται τα διαγράμματα ιστορίας, αυτοσυσχέτισης και μερικής αυτοσυσχέτισης. Στη συνέχεια γίνεται σταθεροποίηση διασποράς και αφαίρεση της τάσης προκειμένου να γίνει στάσιμη η χρονοσειρά, παραθέτοντας και τα αντίστοιχα διαγράμματα της στάσιμης χρονοσειράς που προκύπτει.

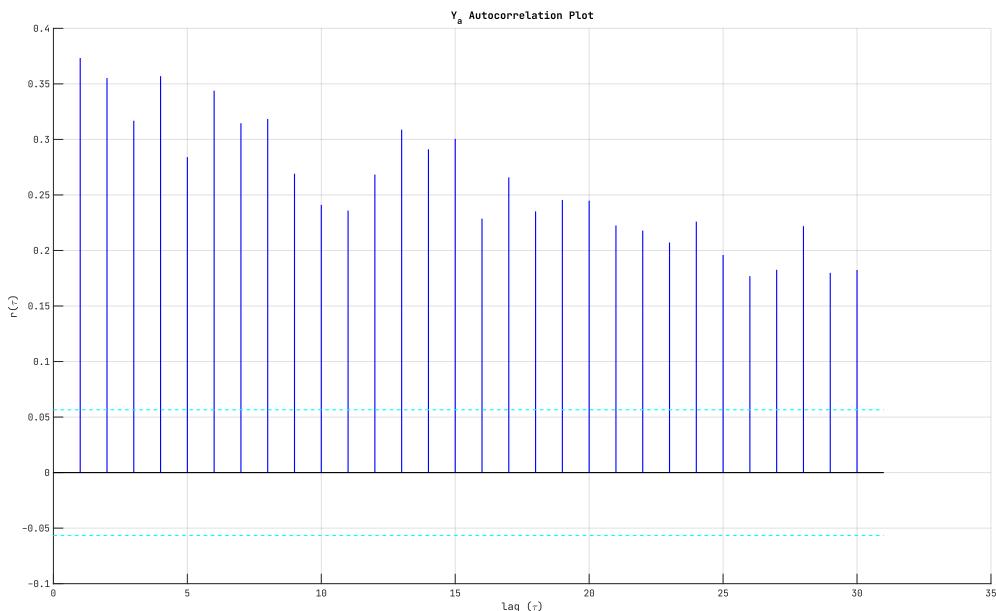
### **1.1 Αρχική Ανάλυση**

Αρχικά, δίνεται το διάγραμμα ιστορίας της χρονοσειράς A,  $\{Y_a(t)\}$ , ακολούθως.

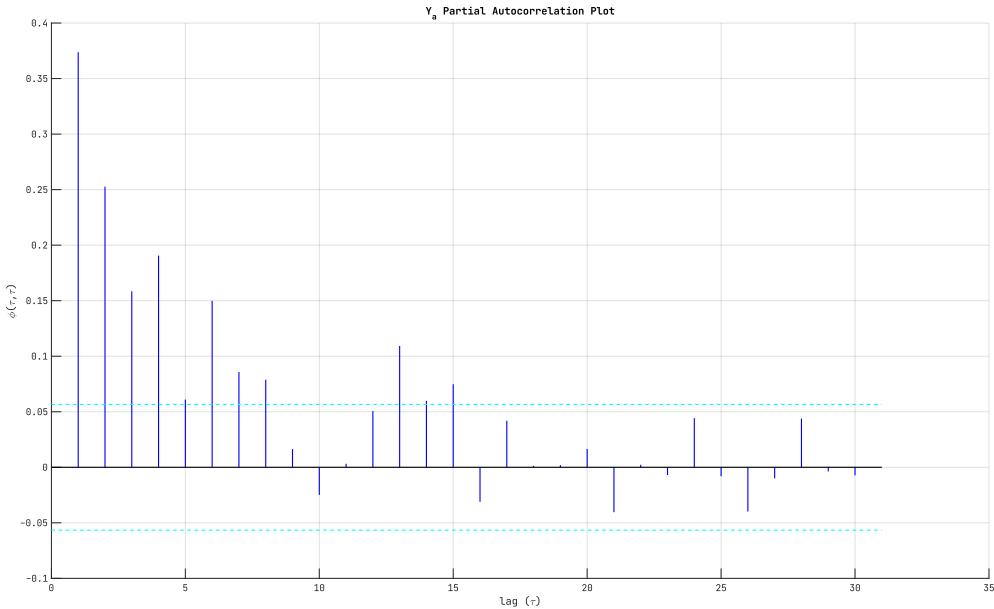


Σχήμα 1.1: Διάγραμμα ιστορίας της χρονοσειράς A, {Y<sub>a</sub>(t)}. Σε όλα τα διαγράμματα ιστορίας που θα παρουσιαστούν, όπως και σε αυτό, με cyan απεικονίζεται μία προσέγγιση της τάσης ως αποτέλεσμα εφαρμογής moving-average smoothing τάξης 7, MA(7), δηλαδή μετριάζοντας με εβδομαδιαίο «παράθυρο».

Από το διάγραμμα ιστορίας της {Y<sub>a</sub>(t)} φαίνεται πως η χρονοσειρά δεν είναι στάσιμη καθώς παρατηρείται κάποια τάση. Αυτό επιβεβαιώνεται και από τα διαγράμματα της (δειγματικής) αυτοσυσχέτισης και (δειγματικής) μερικής αυτοσυσχέτισης που παρατίθενται ακολούθως:



Σχήμα 1.2: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς A, r<sub>y</sub>(τ), μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης



Σχήμα 1.3: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς A,  $\phi_Y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης

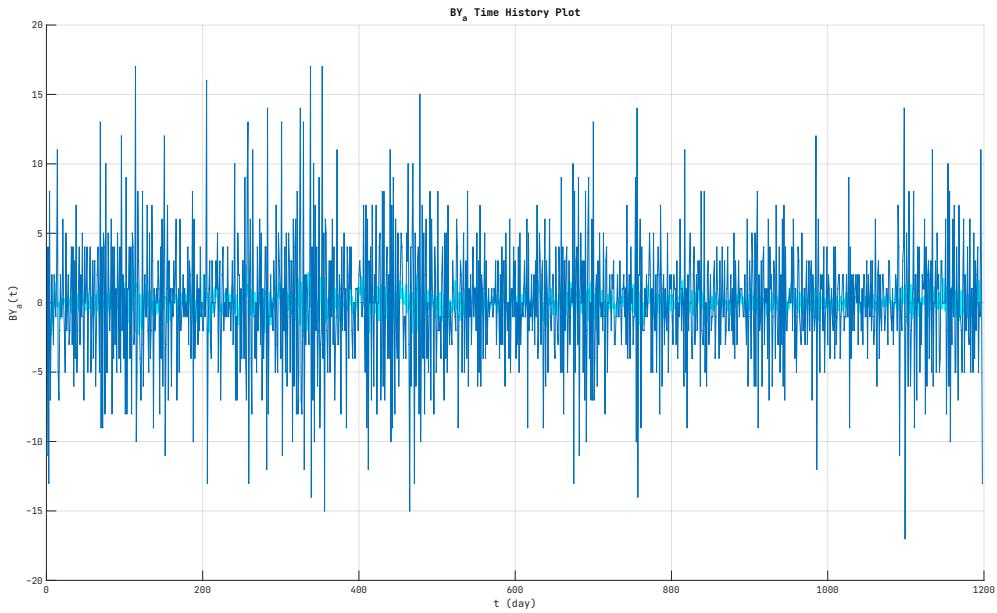
Στα παραπάνω διαγράμματα έχουν σημειωθεί και τα όρια σημαντικότητας για επίπεδο εμπιστοσύνης 95%. Ειδικά στο διάγραμμα αυτοσυσχέτισης φαίνεται έντονα η ύπαρξη τάσης καθώς η αυτοσυσχέτιση έχει υψηλές τιμές και φθίνει πολύ αργά. Η τάση, όπως φαίνεται από το διάγραμμα ιστορίας (σχήμα 1.1), είναι στοχαστική και επομένως για να γίνει στάσιμη η χρονοσειρά θα πρέπει να εφαρμοστεί κάποια μέθοδος απαλοιφής της στοχαστικής τάσης (όπως η μέθοδος των πρώτων διαφορών).

## 1.2 Απαλοιφή Στοχαστικής Τάσης

Για την απαλοιφή της τάσης αρχικά δοκιμάστηκαν οι πρώτες διαφορές. Η χρονοσειρά που προκύπτει λοιπόν θα είναι:

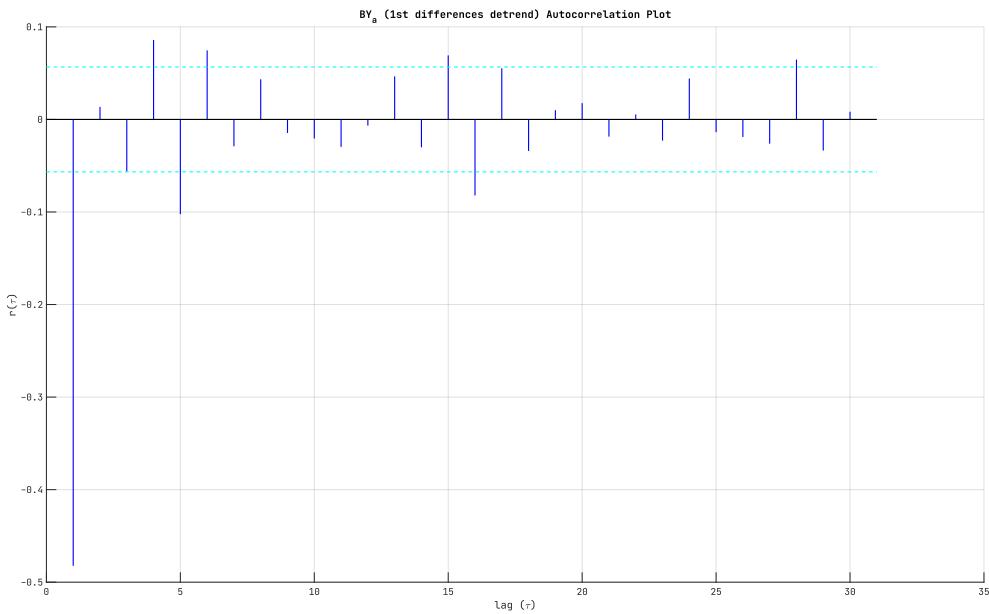
$$BY_a(t) = Y_a(t) - Y_a(t - 1) \quad (1.1)$$

Παρακάτω παρατίθεται το διάγραμμα ιστορίας της  $\{BY_a(t)\}$  όπου φαίνεται ότι ο μετασχηματισμός των πρώτων διαφορών έχει καταφέρει να απαλείψει τη στοχαστική τάση και επομένως δεν υπάρχει η ανάγκη για επανάληψη της διαδικασίας (δηλαδή να πάρω διαφορές δεύτερης τάξης) ή για προσφυγή σε άλλη μέθοδο απαλοιφής στοχαστικής τάσης.

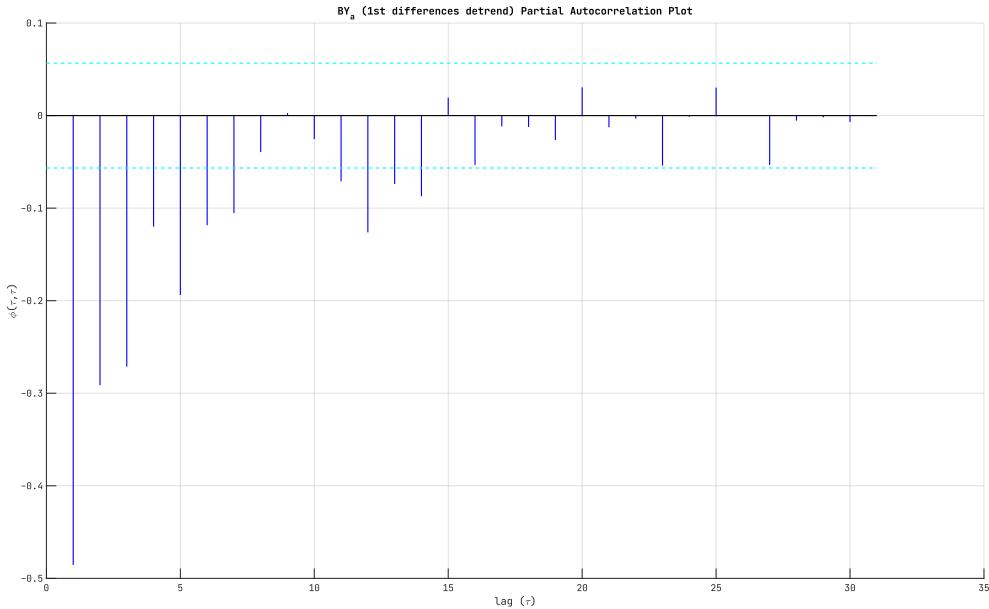


Σχήμα 1.4: Διάγραμμα ιστορίας της χρονοσειράς των πρώτων διαφορών, {BY<sub>a</sub>(t)}

Ο παραπάνω ισχυρισμός περί απαλοιφής τάσης ενισχύεται και από τα ακόλουθα διαγράμματα δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών:



Σχήμα 1.5: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών, r<sub>BY<sub>a</sub></sub>(τ)



Σχήμα 1.6: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών,  $\phi_{BY_a}(\tau)$

όπου αν και δεν φαίνεται η χρονοσειρά να παρουσιάζει αυτοσυσχέτιση λευκού θορύβου επίσης όμως δεν φαίνεται να παρουσιάζει και (στατιστικά) σημαντικές αυτοσυσχετίσεις (σχήμα 1.5) ή αργή πτώση των τιμών.

Σε ότι αφορά τη διασπορά της χρονοσειράς των πρώτων διαφορών, επιστρέφοντας στο διάγραμμα ιστορίας της (σχήμα 1.4) και συγκρίνοντάς το με το διάγραμμα ιστορίας της αρχικής χρονοσειράς (σχήμα 1.1) παρατηρούμε ότι η διασπορά της χρονοσειράς των πρώτων διαφορών φαίνεται να μεταβάλλεται κάπως σχετικά με τις «καμπλύλες» της τάσης της αρχικής χρονοσειράς. Επομένως, ίσως να ήταν σκόπιμο να προσπαθούσαμε να σταθεροποιήσουμε τη διασπορά της προκύπτουσας χρονοσειράς αίροντας την εξάρτησή της από τη τάση της αρχικής, κάτι που αναλύεται ακολούθως.

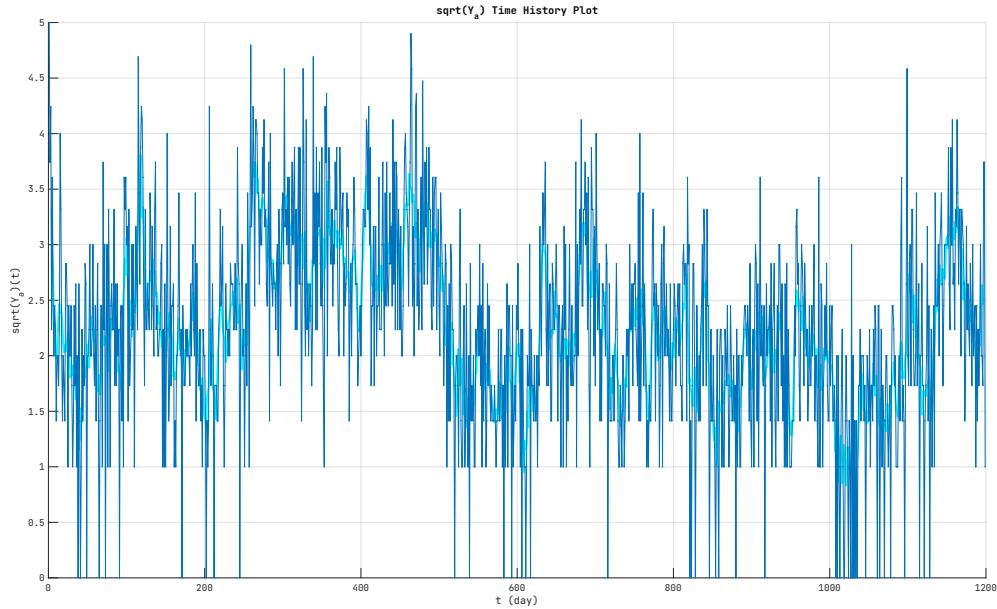
### 1.3 Απαλοιφή Στοχαστικής Τάσης & Σταθεροποίηση Διασποράς

Η απαλοιφή της τάσης (detrending) θεωρούμε πως έχει επιτευχθεί ικανοποιητικά παίρνοντας τις πρώτες διαφορές στην αρχική χρονοσειρά. Ωστόσο, πριν πάρουμε τις πρώτες διαφορές θα μας ενδιέφερε να απαλείψουμε την εξάρτηση της διασποράς της (αρχικής) χρονοσειράς από τη τάση. Για αυτό θα χρησιμοποιηθεί ο μετασχηματισμός των Box & Cox με  $\lambda = 0.5$  αφενός διότι φαίνεται η διασπορά να εξάρταται με ανάλογο (γραμμικό) τρόπο από τη τάση και αφετέρου επειδή η χρονοσειρά περιέχει μηδενικά δείγματα (καμία προβολή του αντίστοιχου βίντεο εκείνη την ημέρα) και άρα οι μετασχηματισμοί λογαρίθμου ή σχετικών μεταβολών δεν αποτελούν πρακτική λύση.

Ο μετασχηματισμός, λοιπόν, που χρησιμοποιήθηκε στην αρχική χρονοσειρά είναι αυτός της τετραγωνικής ρίζας. Δηλαδή, αρχικά η  $\{Y_a(t)\}$  μετασχηματίστηκε σε:

$$\text{sqrt}(Y_a)(t) = \sqrt{Y_a(t)} \quad (1.2)$$

Ακολούθως, δίνεται το διάγραμμα ιστορίας (μαζί με το MA(7) smoothing όπως έχει αναφερθεί παραπάνω) για τη χρονοσειρά των τετραγωνικών ριζών:

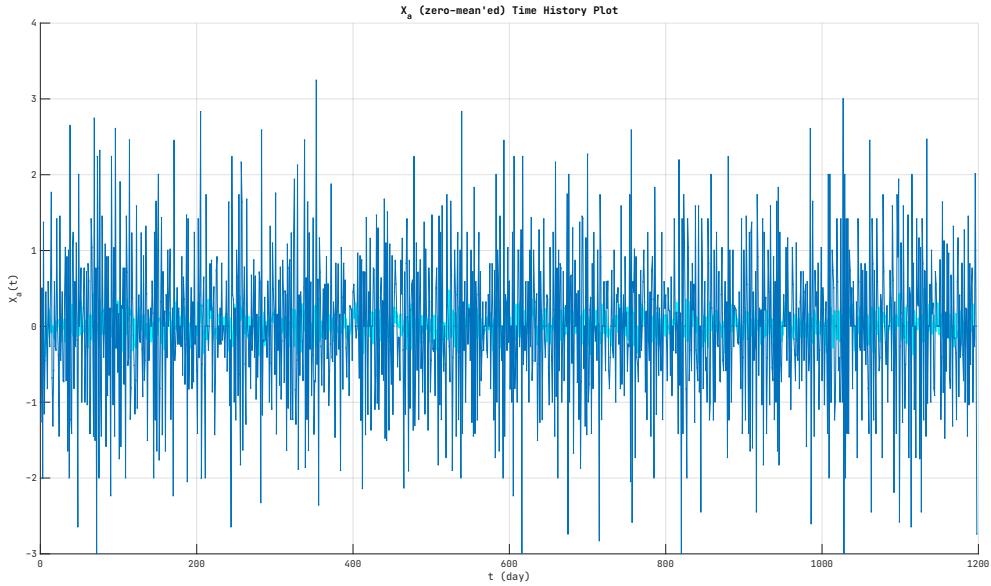


Σχήμα 1.7: Διάγραμμα ιστορίας της χρονοσειράς των τετραγωνικών ριζών,  $\{sqrt(Y_a)(t)\}$

Στη συνέχεια χρησιμοποίηθηκαν οι πρώτες διαφορές για απαλοιφή της τάσης. Συνολικά, λοιπόν, ο μετασχηματισμός που υλοποιήθηκε για μετασχηματισμό της αρχικής χρονοσειράς σε στάσιμη είναι ο ακόλουθος:

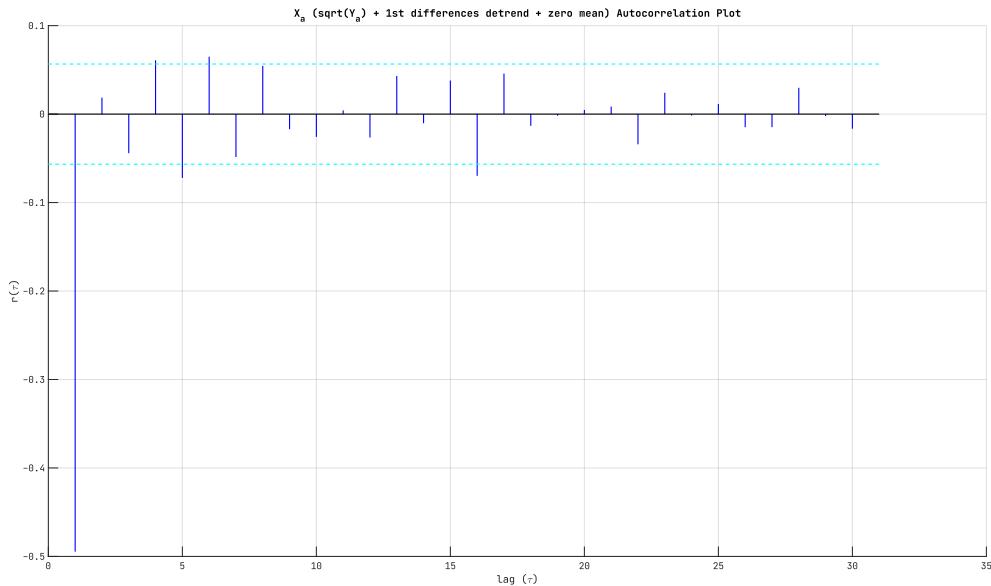
$$X_a(t) = \sqrt{Y_a(t)} - \sqrt{Y_a(t-1)} \quad (1.3)$$

με τη χρονοσειρά  $\{X_a(t)\}$  να είναι η στάσιμη εκδοχή της  $\{Y_a(t)\}$ . Παρακάτω, φαίνεται το διάγραμμα ιστορίας της  $\{X_a(t)\}$  όπου επιβεβαιώνεται η υπόθεσή μας για συσχέτιση της διασποράς με την τάση (αφού πλέον δεν φαίνεται αυτή η «κυματοειδής» μεταβολή της διασποράς). Θεωρούμε, δηλαδή, ότι εφαρμόζοντας το μετασχηματισμό της τετραγωνικής ρίζας στην αρχική χρονοσειρά **πριν** πάρουμε τις πρώτες διαφορές επιτυγχάνουμε την απεξάρτηση της διασποράς από την χρονικής μεταβολή της τάσης. Η χρονοσειρά  $\{X_a(t)\}$  λοιπόν που προκύπτει αφενός δεν έχει εξάρτηση της διασποράς από την τάση και αφετέρου η στοχαστική τάση έχει απαλειφθεί.

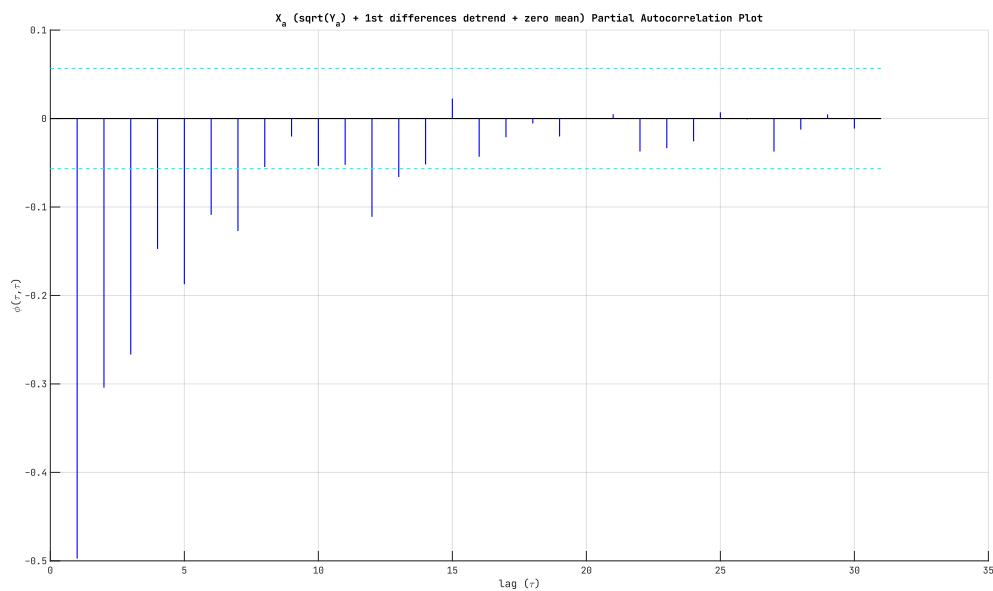


Σχήμα 1.8: Διάγραμμα ιστορίας της στάσιμης χρονοσειράς,  $\{X_a(t)\}$ , ως οι πρώτες διαφορές των τετραγωνικών ριζών της αρχικής χρονοσειράς

Για τη χροοσειρά των πρώτων διαφορών των τετραγωνικών ριζών που καταλήξαμε, παραθέτονται επίσης και τα διαγράμματα δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης με τα όρια σημαντικότητας (για εμπιστοσύνη 95%):



Σχήμα 1.9: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της στάσιμης χρονοσειράς,  $r_x(\tau)$



Σχήμα 1.10: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της στάσιμης χρονοσειράς,  $\phi_x(\tau)$

τα οποία όπως είναι αναμενόμενο είναι σχεδόν ίδια με τα αντίστοιχα διαγράμματα της χρονοσειράς μόνο των πρώτων διαφορών (χωρίς δηλαδή το μετασχηματισμό της τετραγωνικής ρίζας), που δόθηκαν στα σχήματα 1.5 και 1.6 παραπάνω.

## Βήμα 2

### Προσαρμογή Γραμμικού Μοντέλου

Με βάση τα συμπερασμάτα που εξήχθησαν από την ανάλυση στο πρώτο βήμα, το γραμμικό μοντέλο που θα προσαρμόσουμε στη χρονοσειρά A είναι ARIMA με  $d = 1$ , καθώς:

- παίρνοντας τις πρώτες διαφορές (στη μετασχηματισμένη με τετραγωνικές ρίζες αρχική χρονοσειρά), η χρονοσειρά που προκύπτει φαίνεται στα διαγράμματα ιστορίας και αυτοσυσχέτισης να είναι στάσιμη (η τάση έχει απαλειφθεί) και επιπλέον δεν υπάρχει εξάρτηση της διασποράς από τη τάση
- δεν φαίνεται να υπάρχει εποχικότητα και άρα η χρήση ενός ARIMA μοντέλου θα είναι ικανοποιητική. Συγκεκριμένα δοκιμάστηκε η εκτίμηση της εποχικότητας (με μέσους όρους στοιχείων της περιόδου) και οι ακολουθίες που βρέθηκαν για περιόδους 5, 6 και 16 ημερών (καθώς αυτές είναι μόνες στατιστικά σημαντικές τιμές του διαγράμματος αυτοσυσχέτισης - σχήμα 1.9) είναι [0.082,-0.0672,0.021,0.039,-0.075], [0.011,-0.033,-0.017,0.020,-0.106,0.1250] και [0.087,0.224,-0.176,-0.104,0.092,0.214,-0.057,-0.130,0.058,-0.095,0.020,0.046,-0.069,0.219,-0.254,-0.075] αντίστοιχα. Δεδομένου ότι το εύρος τιμών της αντίστοιχης χρονοσειράς είναι περίπου στο [-3, 3] αλλά και ότι οι τιμές της αυτοσυσχέτισεις για τις συγκεκριμένες υστερήσεις είναι οριακά πάνω από το όριο σημαντικότητας, μπορούμε με ασφάλεια να εξάγουμε το συμπέρασμα ότι δεν υπάρχει εποχικός όρος στη χρονοσειρά των πρώτων διαφορών των τετραγωνικών ριζών, δηλαδή στη στάσιμη χρονοσειρά που καταλήξαμε.

#### 2.1 Προσαρμογή ARIMA( $p,d,q$ )

##### 2.1.1 Προσαρμογή ARIMA( $p,1,q$ ) παίρνοντας τις πρώτες διαφορές

Έτσι, συνεχίζοντας την ανάλυση από το τέλος του βήματος 1, μετά την εφαρμογή των πρώτων διαφορών στη χρονοσειρά των τετραγωνικών ριζών έχουμε υλοποιήσει το πρώτο στάδιο προσαρμογής ενός μοντέλου ARIMA( $p,d,q$ ) που είναι η εφαρμογή διαφορών  $d$ -οστής. Άρα εδώ θα είναι  $d = 1$  και η χρονοσειρά που καταλήγουμε είναι η  $\{X_a(t)\}$ .

Το δεύτερο στάδιο είναι η εύρεση των παραμέτρων του μοντέλου ARMA( $p,q$ ), κάτι που αναλύεται στην επόμενη υπο-ενότητα.

## 2.1.2 Προσαρμογή ARMA(p,q) στη στάσιμη χρονοσειρά

Οι παράμετροι  $p$  και  $q$  είναι hyperparameters του μοντέλου και επομένως το πρώτο μας μέλημα είναι να βρούμε το βέλτιστο συνδυσμό των παραμέτρων αυτών ή τάξεων του μοντέλου. Προς το σκοπό αυτό θα κάνουμε grid search για τιμές των τάξεων από 0 (απουσία του αντίστοιχου όρου) έως και 10.

Για την αξιολόγηση του κάθε συνδυασμού χρησιμοποιήθηκαν τα κριτήρια πληροφορίας Akaike (AIC) και Forward Prediction Error (FPE), τα οποία ορίζονται ως εξής:

$$AIC(p, q) = \ln(s_z^2) + \frac{2 \times (p + q)}{n} \quad (2.1)$$

και

$$FPE(p, q) = s_z^2 \times \frac{n + (p + q)}{n - (p + q)} \quad (2.2)$$

όπου  $n$  είναι ο αριθμός των δειγμάτων που χρησιμοποιήθηκαν για την εκτίμηση των ροπών (π.χ για τη δειγματική αυτοσυσχέτιση) και  $s_z^2$  είναι η (δειγματική) διασπορά των σφαλμάτων ή υπολοίπων που προκύπτουν όταν συγκρίνουμε τις τιμές του προσαρμοεσμένου μοντέλου τάξης  $(p, q)$  με τις πραγματικές τιμές της χρονοσειράς (ενν. τη χρονοσειρά  $\{X_a(t)\}$  που προέκυψε ως οι πρώτες διαφορές της χρονοσειράς των τετραγωνικών ριζών της αρχικής χρονοσειράς προβολών του βίντεο A,  $\{Y_a(t)\}$ ).

Παρακάτω παρατίθενται οι τιμές του AIC για τους συνδυασμούς των παραμέτρων  $(p, q) = (0 \dots 10)$ , φυσικά με εξαίρεση το συνδυασμό  $(p, q) = (0, 0)$ :

		$p$										
		0	1	2	3	4	5	6	7	8	9	10
q	0	-0.465	-0.463	-0.462	-0.461	-0.460	-0.459	-0.459	-0.457	-0.458	-0.457	-0.457
	1	-0.211	-0.463	-0.463	-0.462	-0.462	-0.461	-0.460	-0.458	-0.456	-0.457	-0.456
	2	-0.306	-0.462	-0.462	-0.465	-0.463	-0.461	-0.460	-0.457	-0.456	-0.459	-0.458
	3	-0.376	-0.461	-0.462	-0.463	-0.462		-0.458	-0.459	-0.457	-0.458	-0.457
	4	-0.395	-0.46	-0.461	-0.462			-0.463	-0.460	-0.462	-0.457	-0.456
	5	-0.427	-0.460	-0.459	-0.460	-0.459	-0.461	-0.459		-0.455	-0.455	-0.457
	6	-0.436	-0.459	-0.458	-0.459	-0.459	-0.460		-0.463	-0.457	-0.458	-0.455
	7	-0.448	-0.458	-0.456	-0.457	-0.457	-0.457	-0.456	-0.457	-0.462	-0.454	-0.453
	8	-0.449	-0.458	-0.456	-0.459	-0.458	-0.456	-0.456	-0.457	-0.454	-0.462	-0.455
	9	-0.447	-0.456	-0.456	-0.455	-0.456	-0.456	-0.455	-0.455	-0.454	-0.452	-0.456
	10	-0.446	-0.456	-0.454	-0.454	-0.455	-0.454	-0.453	-0.459	-0.456		-0.456

Πίνακας 2.1: Αναζήτηση Πλέγματος με βάση τη μετρική AIC για διάφορες τιμές των τάξεων  $(p, q)$ . Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο.

από όπου φαίνεται πως η χαμηλότερη τιμή του AIC επιτυγχανέται όταν προσαρμοζέται μοντέλο ARMA(0, 1) ή, ισοδύναμα, μοντέλο MA(1). Ακολούθως δίνονται οι τιμές του FPE για τους αντίστοιχους συνδυασμούς τιμών των  $p$  και  $q$ :

	$p$										
	0	1	2	3	4	5	6	7	8	9	10
$q$	0	0.628	0.629	0.630	0.631	0.632	0.632	0.632	0.633	0.632	0.633
	1	0.810	0.629	0.629	0.630	0.630	0.631	0.631	0.632	0.634	0.633
	2	0.737	0.630	0.630	0.628	0.629	0.630	0.631	0.633	0.634	0.632
	3	0.687	0.631	0.630	0.629	0.630		0.633	0.632	0.633	0.632
	4	0.673	0.631	0.631	0.630			0.629	0.631	0.630	0.633
	5	0.652	0.632	0.632	0.631	0.632	0.631	0.632		0.634	0.634
	6	0.647	0.632	0.633	0.632	0.632	0.631		0.630	0.633	0.633
	7	0.639	0.633	0.634	0.633	0.633	0.633	0.634	0.633	0.630	0.635
	8	0.639	0.633	0.634	0.632	0.632	0.634	0.634	0.633	0.635	0.635
	9	0.640	0.634	0.634	0.635	0.634	0.634	0.634	0.634	0.635	0.634
	10	0.640	0.634	0.635	0.635	0.635	0.635	0.636	0.632	0.634	

Πίνακας 2.2: Αναζήτηση Πλέγματος με βάση τη μετρική FPE για διάφορες τιμές των τάξεων ( $p, q$ ). Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο.

Όπως επιβεβαιώνεται και από τους δύο πίνακες παραπάνω, φαίνεται πως από τα γραμμικά μοντέλα καλύτερα προσαρμόζεται το μοντέλο  $MA(1)$ . Ωστόσο, θα θέλαμε το μοντέλο μας να έχει «μνήμη» κάτι που επιτυγχάνεται έαν έχει όρους ανάδρασης ή αυτοπαλινδρόμησης. Κάτι τέτοιο θα ήταν δυνατό μιας και που οι τιμές των κριτηρίων πληροφορίας μόλις που μεταβάλλονται για μοντέλο π.χ.  $ARMA(1, 1)$ . Όμως, όπως αναλυέται και στη συνέχεια, η προσαρμογή μοντέλου κινούμενου μέσου μοναδιαίας τάξης οδηγεί σε αρκετά ικανοποιητικά αποτελέσματα και έτσι θα εμπιστευτούμε τα κριτηρία πληροφορίας.

Πριν προχωρήσουμε στη διάγνωση καταλληλότητας, θα ήταν σκόπιμο να τονιστεί ότι δεν χρησιμοποιήθηκε το NRMSE (πρόβλεψης 1 βήματος μπροστά) ή το  $s_z^2$  των υπολοίπων για επιλογή των τάξεων του μοντέλου καθώς αμφότερα δεν λαμβάνουν υπόψη τους την «πολυπλοκότητα» του μοντέλου και άρα θα μειώνονταν με αύξηση των τάξεων κάτι που οδηγεί σε πιθανό overfitting. Πράγματι, με βάση το NRMSE πρόβλεψης 1 βήματος ο καλύτερος συνδυασμός παραμέτρων θα ήταν  $(p, q) = (10, 10)$  με αντίστοιχο NRMSE 0.744.

## 2.2 Διάγνωση καταλληλότητας του μοντέλου ARIMA(0,1,1)

### 2.2.1 Τελικό μοντέλο τύπου MA(1) για τη στάσιμη χρονοσειρά

Το μοντέλο  $MA(1)$  εκτιμήθηκε από τη συνάρτηση `fitARMA()` και προσαρμόστηκε στη στάσιμη χρονοσειρά (δηλαδή τη χρονοσειρά των διαφορών των τετρ. ριζών,  $\{X_a(t) = \sqrt{Y_a(t)} - \sqrt{Y_a(t-1)}\}$ ), είναι το εξής:

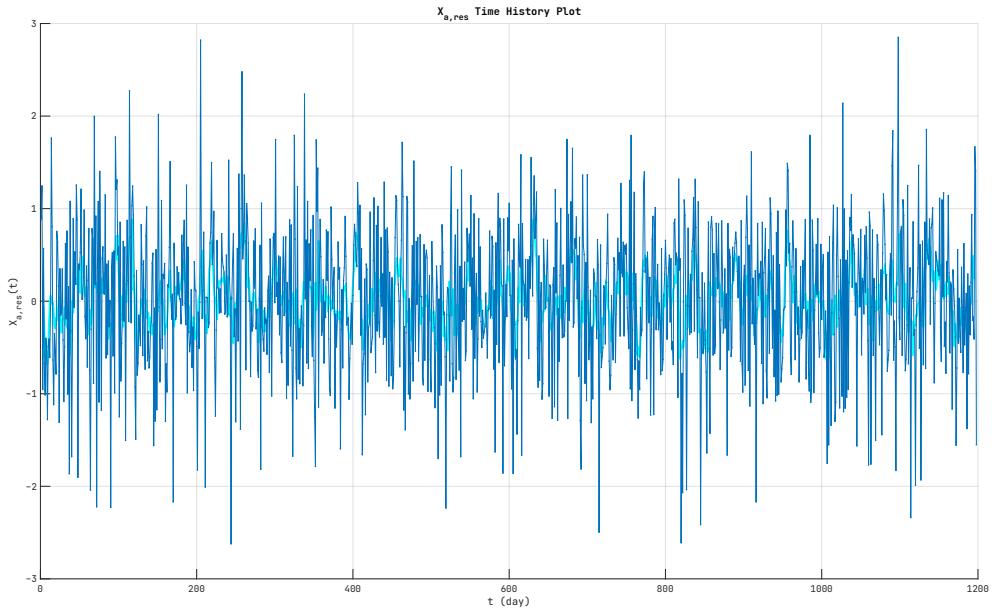
$$x_t = -0.0033 + z_t - 0.8426 z_{t-1}, \quad t = 2, \dots, 1199 \quad (2.3)$$

όπου ο μέσος όρος της στάσιμης χρονοσειράς  $\{X_a(t)\}$  είναι  $\bar{x}_a = -0.0033$ . Ενώ η εκτίμηση της τυπικής απόκλισης των σφαλμάτων ή υπολοίπων προσαρμογής βρέθηκε να είναι ίση με  $s_z = 0.7922$  (εκτίμηση διασποράς ίση με  $s_z^2 = 0.6275$ ). Το αντίστοιχο μοντέλο με το οποίο θα προσεγγίζαμε την αρχική χρονοσειρά των προβολών του βίντεο A, θα είναι:

$$Y_t = \left( \sqrt{Y_{t-1}} - 0.0033 + Z_t - 0.8426 Z_{t-1} \right)^2, \quad t = 2, \dots, 1199 \quad (2.4)$$

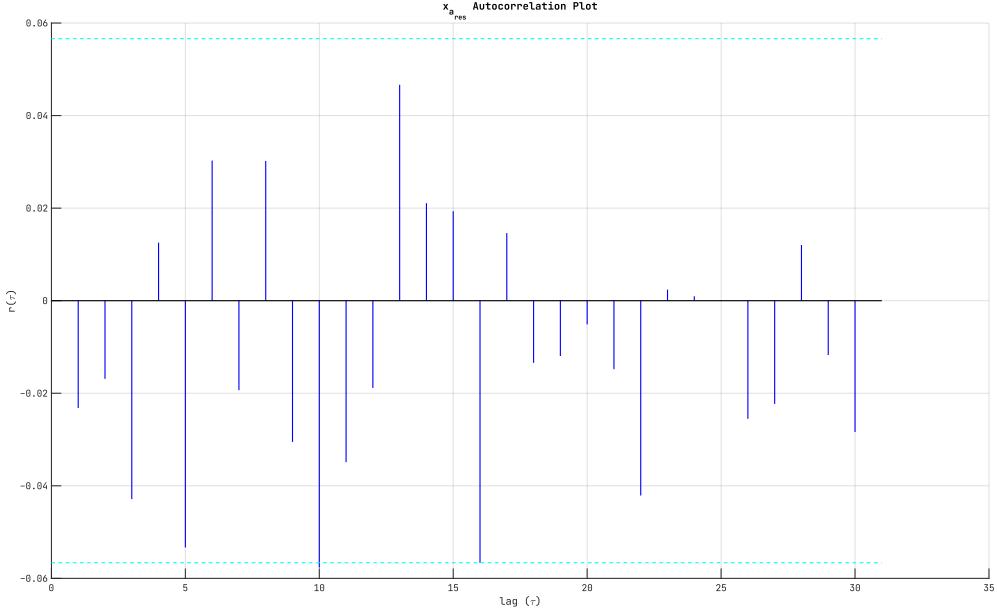
## 2.2.2 Διάγνωση Καταλληλότητας & Σφάλματα Προσαρμογής

Ακολούθως, θα κάνουμε διάγνωση καταλληλότητας του μοντέλου  $MA(1)$  που καταλήξαμε προκειμένου να αποφανθούμε έαν αυτό το μοντέλο αντλεί όλη τη πληροφορία της στάσιμης χρονοσειράς, αφήνοντας ασυσχέτιστα υπόλοιπα (λευκό θόρυβο). Επομένως, θα κάνουμε έλεγχο ανεξαρτησίας στη σειρά των υπολοίπων τόσο με βάση τη δειγματική τους αυτοσυσχέτιση όσο και με τον έλεγχο Portmanteau. Σε πρώτη φάση, όμως, το διάγραμμα ιστορίας της σειράς των υπολοίπων ή σφαλμάτων προσαρμογής δίνεται ακολούθως:

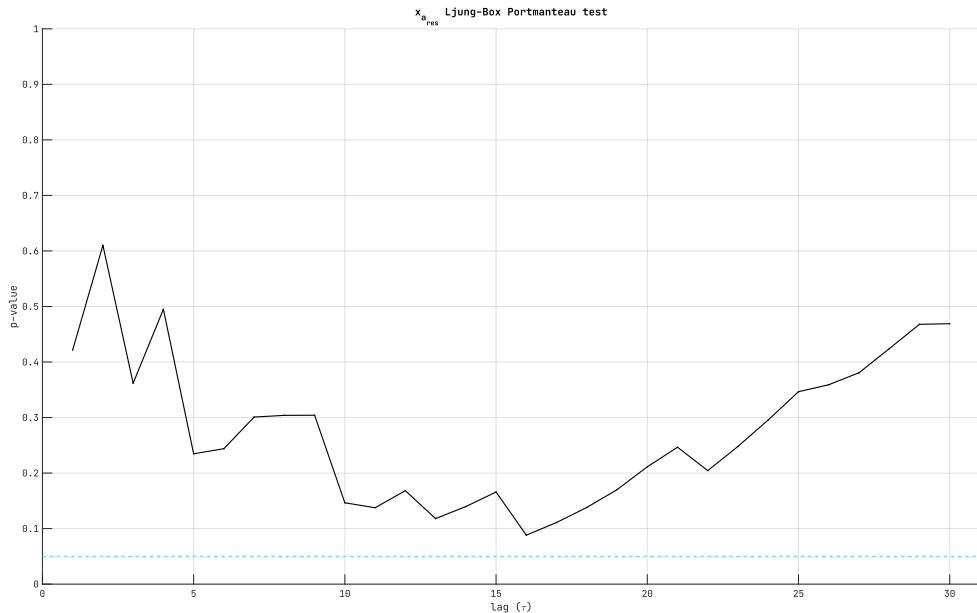


Σχήμα 2.1: Διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (2.3),  $\{X_{a,res}(t) = \hat{z}(t)\}$

Στη συνέχεια δίνονται τα διαγράμματα δειγματικής αυτοσυσχέτισης και p-values του Ljung & Box test για μέγιστη υστέρηση  $\tau$  από 1 έως 30:



Σχήμα 2.2: Διάγραμμα δειγματικής αυτοσυσχέτισης της σειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (2.3),  $r_{\hat{z}_a}(\tau)$



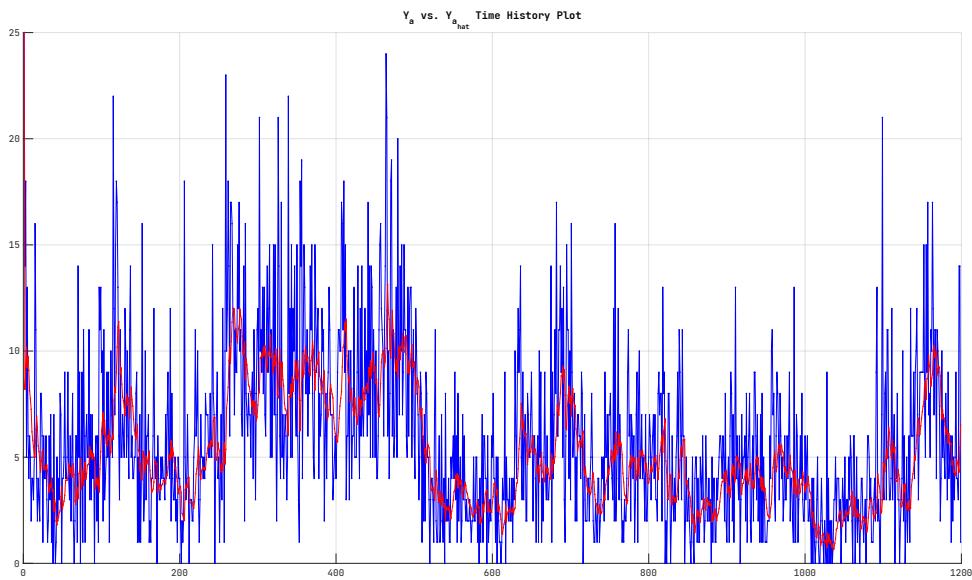
Σχήμα 2.3: Διάγραμμα των p-values του στατιστικού ελέγχου ανεξαρτησίας Portmanteau (Ljung & Box test) των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (2.3). Σημειώνεται με διακεκομένη γραμμή το όριο απόφασης όπου όπως φαίνεται η μηδενική υπόθεση  $H_0$  (η σειρά των υπολοίπων είναι iid) δεν απόρριπτεται για καμία από τις 30 υστερήσεις.

Αμφότερα τα σχήματα 2.2 και 2.3 παραπάνω φανερώνουν ότι η προσαρμογή του  $MA(1)$  μοντέλου της σχέσης (2.3) είναι επιτυχής αφού αφήνει ασυσχέτιστα υπόλοιπα. Αυτό φαίνεται στο διάγραμμα αυτοσυσχετίσεων των υπολοίπων όπου μόνο για υστέρηση 10 η αυτοσυσχέτιση μόλις που ξεπερνάει το όριο σημαντικότητας (κάτι που επιτρέπεται από το επίπεδο ε-

μπιστοσύνης) ενώ όλες οι υπόλοιπες αυτοσυσχετίσεις είναι στατιστικά μηδενικές. Αντίστοιχα αποτελέσματα λαμβάνουμε και από τον έλεγχο ανεξαρτησίας Portmanteau όπου η μηδενική υπόθεση πως η σειρά των υπολοίπων είναι iid δεν απορρίπτεται για καμία από τις 30 υπερήσεις. Με ασφάλεια μπορούμε να πούμε πως η προσαρμογή αφήνει λευκό θόρυβο ως σειρά υπολοίπων.

Τέλος, για λόγους πληρότητας παρεθέτουμε το NRMSE των σφαλμάτων προσαρμογής για πρόβλεψη ενός βήματος μπροστά καθώς και τις ίδιες τις προβλέψεις μαζί με την αρχική χρονοσειρά προβολών του βίντεο A με βάση την σχέση (2.4), παρακάτω:

$$NRMSE(\hat{X}_a, X_a) = 0.7567 \quad (2.5)$$



Σχήμα 2.4: Διάγραμμα ιστορίας της αρχικής χρονοσειράς προβολών του βίντεο (μπλε) A καθώς και τις προβλέψεις ενός βήματος μπροστά αυτής με βάση το προσαρμοσμένο μοντέλο MA(1) και τη σχέση (2.4) (κόκκινο).

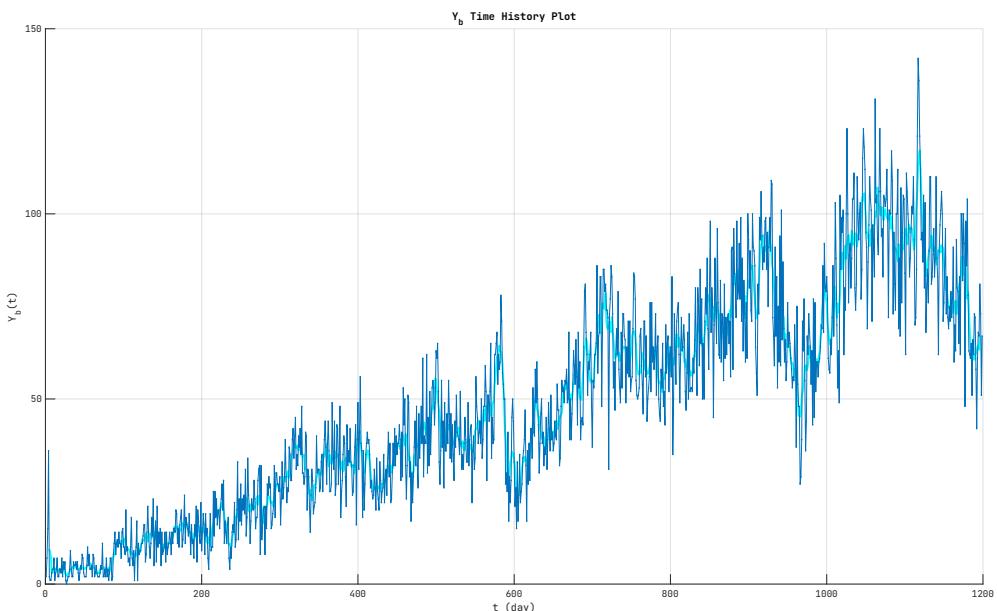
## Βήμα 3

### Διερεύνηση Χρονοσειράς Β

Στις ενότητες και υπο-ενότητες που ακολουθούν γίνεται η ανάλυση της δεύτερης χρονοσειράς ή της «Β». Ακολουθείται ίδια λογική με αυτή της «Α», δηλαδή αρχικά παρατίθενται στοιχεία της αρχικής χρονοσειράς των προβολών, κατόπιν γίνεται σταθεροποίηση διασποράς και απαλοιφή στοχαστικής τάσης. Όστοσο, όπως αναλύεται παρακάτω, η συγκεκριμένη χρονοσειρά φαίνεται να έχει κάποια εποχικότητα η οποία εκτιμάται και απαλείφεται. Τέλος, προσαρμόζονται ARMA μοντέλα στην απαλλαγμένη από εποχικότητα στάσιμη χρονοσειρά και διεξάγονται οι αντίστοιχες διαγνώσεις καταλληλότητας των γραμμικών μοντέλων.

#### 3.1 Αρχική Ανάλυση

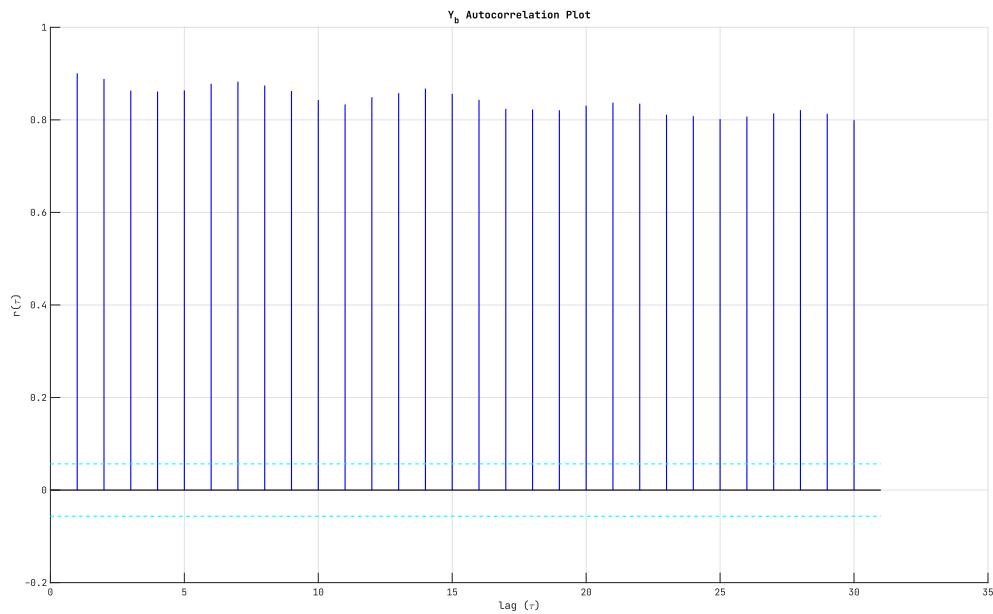
Αρχικά, δίνεται το διάγραμμα ιστορίας της χρονοσειράς Β,  $\{Y_b(t)\}$ , ακολούθως.



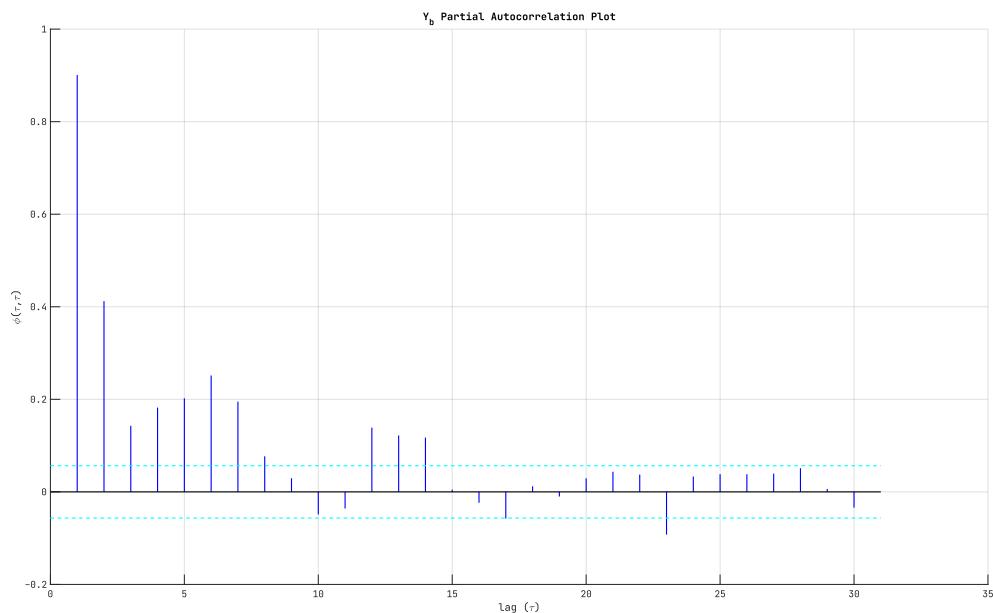
Σχήμα 3.1: Διάγραμμα ιστορίας της χρονοσειράς Β,  $\{Y_b(t)\}$ , μαζί με τη καμπύλη MA(7) smoothing

Από το διάγραμμα ιστορίας της  $\{Y_b(t)\}$  φαίνεται πως η χρονοσειρά δεν είναι στάσιμη καθώς παρατηρείται έντονη αυξητική τάση. Αυτό επιβεβαιώνεται και από τα διαγράμματα της

(δειγματικής) αυτοσυσχέτισης και (δειγματικής) μερικής αυτοσυσχέτισης που παρατίθενται ακολούθως:



Σχήμα 3.2: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς B,  $r_y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης



Σχήμα 3.3: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς B,  $\phi_y(\tau)$ , μαζί με τα όρια σημαντικότητας για 95% επίπεδο εμπιστοσύνης

Στα παραπάνω διαγράμματα έχουν σημειωθεί και τα όρια σημαντικότητας για επίπεδο εμπιστοσύνης 95%. Ειδικά στο διάγραμμα αυτοσυσχέτισης φαίνεται έντονα η ύπαρξη τάσης καθώς η αυτοσυσχέτιση έχει υψηλές τιμές και φθίνει πολύ αργά. Η τάση, όπως φαίνεται από το

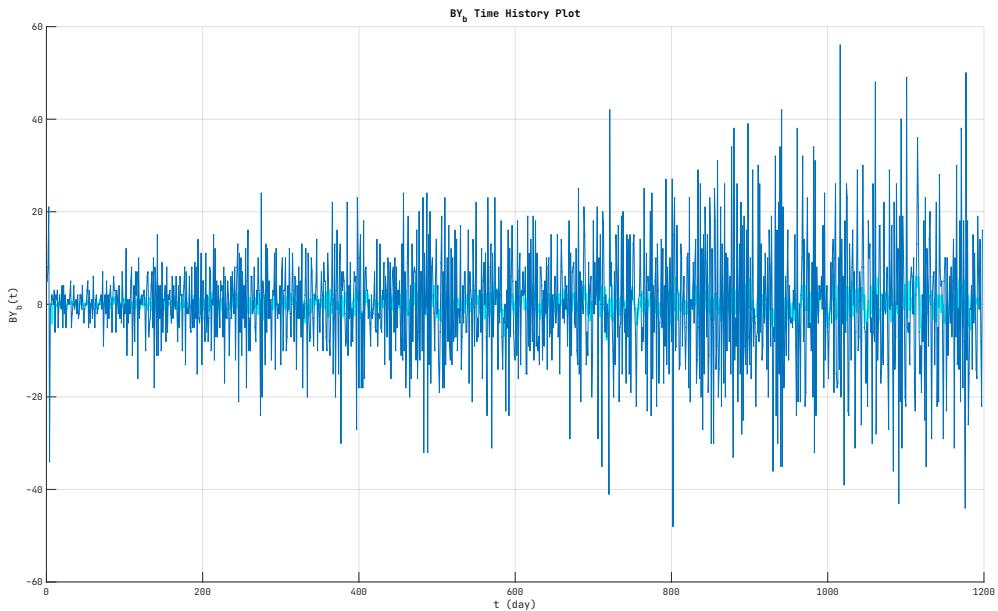
διάγραμμα ιστορίας (σχήμα 3.1), είναι στοχαστική και επομένως για να γίνει στάσιμη η χρονοσειρά θα πρέπει να εφαρμοστεί κάποια μέθοδος απαλοιφής της στοχαστικής τάσης (όπως η μέθοδος των πρώτων διαφορών). Επίσης, αμυδρά φαίνεται στο διάγραμμα αυτοσυσχέτισης και ύπαρξη κάποιας περιοδίκοτητας, όμως αυτή θα φανεί πιο καθαρά παρακάτω.

## 3.2 Απαλοιφή Στοχαστικής Τάσης

Για την απαλοιφή της τάσης αρχικά δοκιμάστηκαν οι πρώτες διαφορές. Η χρονοσειρά που προκύπτει λοιπόν θα είναι:

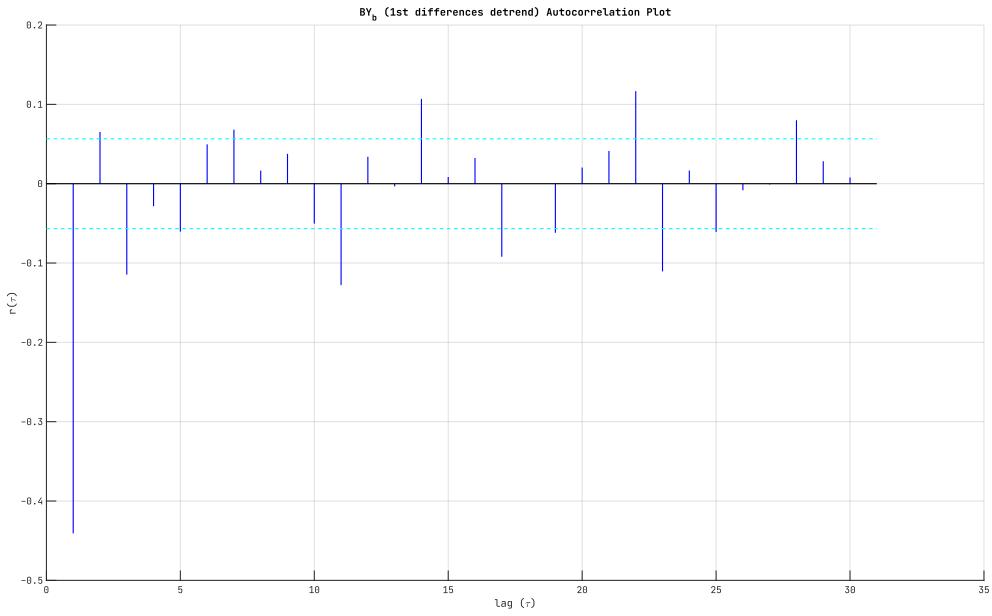
$$BY_b(t) = Y_b(t) - Y_b(t - 1) \quad (3.1)$$

Παρακάτω παρατίθεται το διάγραμμα ιστορίας της  $\{BY_b(t)\}$  όπου φαίνεται ότι ο μετασχηματισμός των πρώτων διαφορών έχει καταφέρει να απαλείψει τη στοχαστική τάση και επομένως δεν υπάρχει η ανάγκη για επανάληψη της διαδικασίας (δηλαδή να πάρω διαφορές δεύτερης τάξης) ή για προσφυγή σε άλλη μέθοδο απαλοιφής στοχαστικής τάσης.

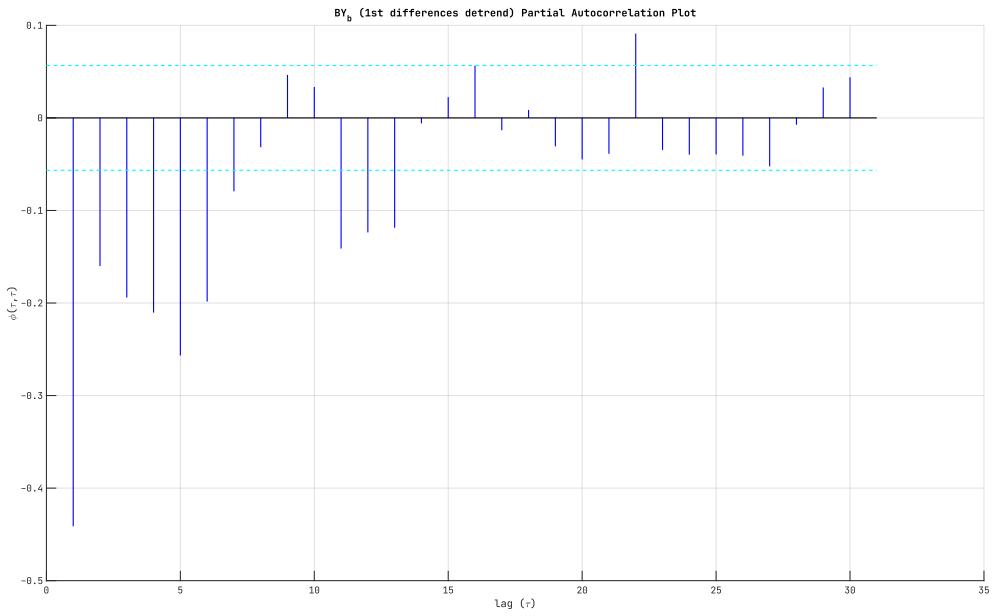


Σχήμα 3.4: Διάγραμμα ιστορίας της χρονοσειράς των πρώτων διαφορών,  $\{BY_b(t)\}$

Ο παραπάνω ισχυρισμός περί απαλοιφής τάσης ενισχύεται και από τα ακόλουθα διαγράμματα δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών:



Σχήμα 3.5: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών,  $r_{BY_b}(\tau)$



Σχήμα 3.6: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της χρονοσειράς των πρώτων διαφορών,  $\phi_{BY_b}(\tau)$

όπου αν και δεν φαίνεται η χρονοσειρά να παρουσιάζει αυτοσυσχέτιση λευκού θορύβου επίσης όμως δεν φαίνεται να παρουσιάζει και (στατιστικά) σημαντικές αυτοσυσχετίσεις (σχήμα 3.5) ή αργή πτώση των τιμών.

Σε ότι αφορά τη διασπορά της χρονοσειράς των πρώτων διαφορών, επιστρέφοντας στο διάγραμμα ιστορίας της (σχήμα 3.4) και συγκρίνοντάς το με το διάγραμμα ιστορίας της αρχικής χρονοσειράς (σχήμα 3.1) παρατηρούμε ότι η διασπορά της χρονοσειράς των πρώτων διαφορών φαίνεται να μεταβάλλεται αρκετά σχετικά με την αυξητική «καμπλύλη» της τάσης της

αρχικής χρονοσειράς. Επομένως, είναι αναγκαίο να προσπαθήσουμε να σταθεροποιήσουμε τη διασπορά της προκύπτουσας χρονοσειράς αίροντας την εξάρτησή της από τη τάση της αρχικής, κάτι που αναλύεται ακολούθως.

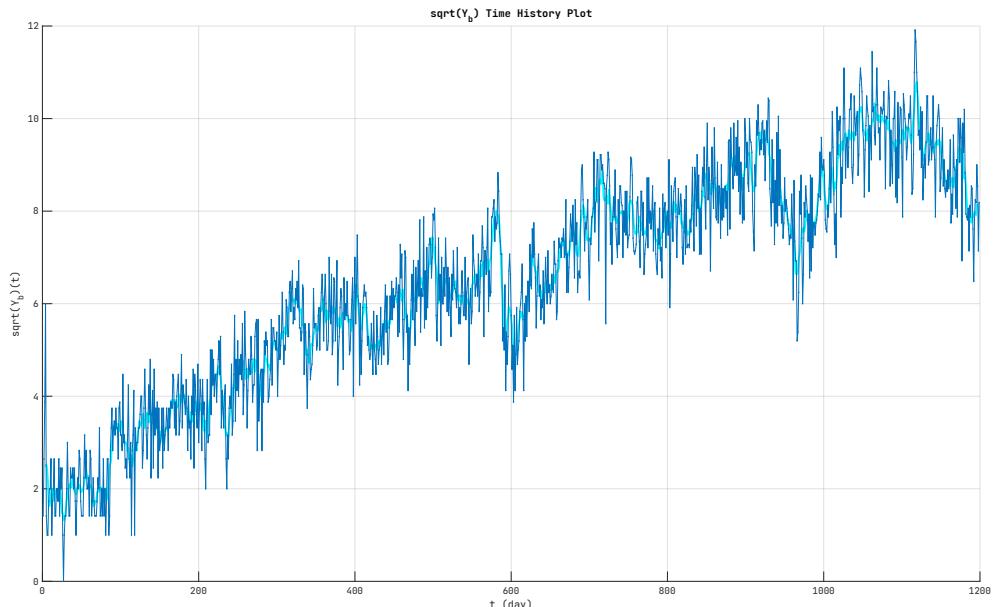
### 3.3 Απαλοιφή Στοχαστικής Τάσης & Σταθεροποίηση Διασποράς

Η απαλοιφή της τάσης (detrending) θεωρούμε πως έχει επιτευχθεί ικανοποιητικά παίρνοντας τις πρώτες διαφορές στην αρχική χρονοσειρά. Ωστόσο, πριν πάρουμε τις πρώτες διαφορές θα μας ενδιέφερε να απαλείψουμε την εξάρτηση της διασποράς της (αρχικής) χρονοσειράς από τη τάση. Για αυτό θα χρησιμοποιηθεί ο μετασχηματισμός των Box & Cox με  $\lambda = 0.5$  για τους ίδιους λόγους που εξηγήθηκαν στο βήμα 1 κατά την επεξεργασία της χρονοσειράς «Α».

Ο μετασχηματισμός, λοιπόν, που χρησιμοποιήθηκε στην αρχική χρονοσειρά Β είναι αυτός της τετραγωνικής ρίζας. Δηλαδή, αρχικά η  $\{Y_b(t)\}$  μετασχηματίστηκε σε:

$$\text{sqrt}(Y_b)(t) = \sqrt{Y_b(t)} \quad (3.2)$$

Ακολούθως, δίνεται το διάγραμμα ιστορίας (μαζί με το MA(7) smoothing όπως έχει αναφερθεί παραπάνω) για τη χρονοσειρά των τετραγωνικών ριζών:

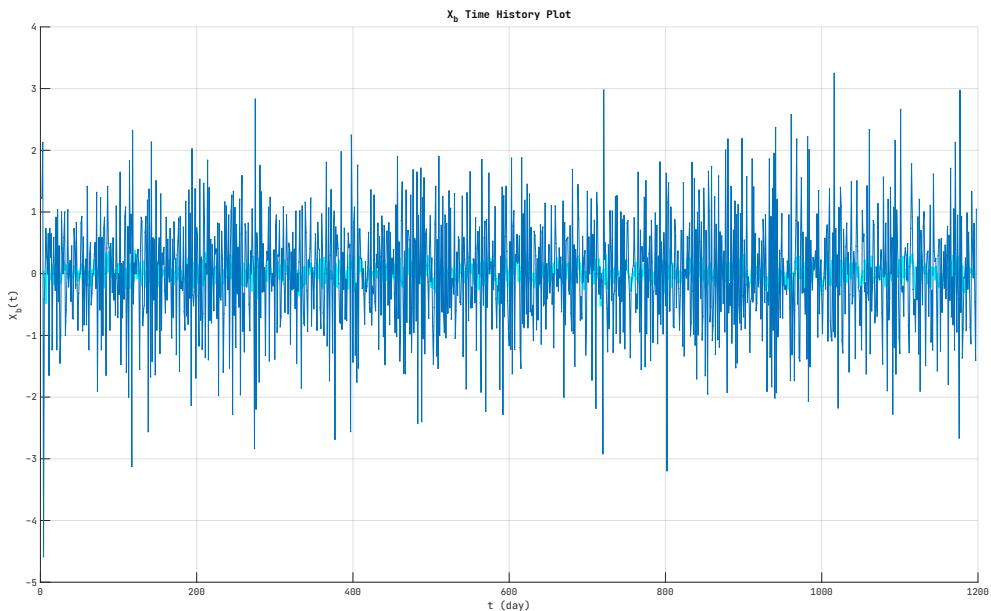


Σχήμα 3.7: Διάγραμμα ιστορίας της χρονοσειράς των τετραγωνικών ριζών,  $\{\text{sqrt}(Y_b)(t)\}$

Στη συνέχεια χρησιμοποίηθηκαν οι πρώτες διαφορές για απαλοιφή της τάσης. Συνολικά, λοιπόν, ο μετασχηματισμός που υλοποιήθηκε για μετασχηματισμό της αρχικής χρονοσειράς σε στάσιμη είναι ο ακόλουθος:

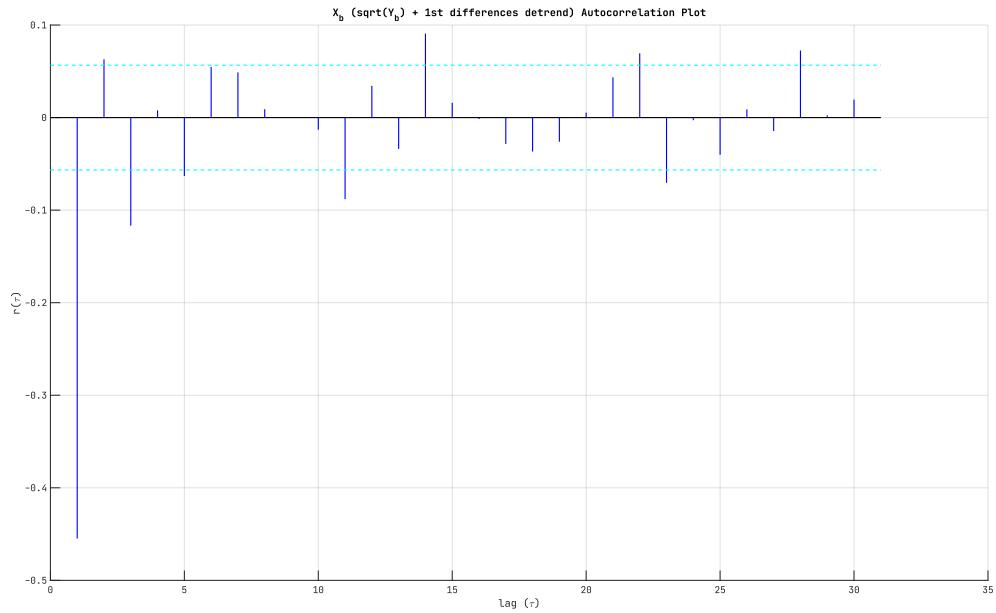
$$X_b(t) = \sqrt{Y_b(t)} - \sqrt{Y_b(t-1)} \quad (3.3)$$

με τη χρονοσειρά  $\{X_b(t)\}$  να είναι η στάσιμη εκδοχή της  $\{Y_b(t)\}$ . Παρακάτω, φαίνεται το διάγραμμα ιστορίας της  $\{X_b(t)\}$  όπου επιβεβαιώνεται η υπόθεσή μας για συσχέτιση της διασποράς με την τάση (αφού πλέον δεν φαίνεται αυτή η «κυματοειδής» μεταβολή της διασποράς). Θεωρούμε, δηλαδή, ότι εφαρμόζοντας το μετασχηματισμό της τετραγωνικής ρίζας στην αρχική χρονοσειρά **πριν** πάρουμε τις πρώτες διαφορές επιτυγχάνουμε την απεξάρτηση της διασποράς από την χρονικής μεταβολή της τάσης. Η χρονοσειρά  $\{X_b(t)\}$  λοιπόν που προκύπτει αφενός δεν έχει εξάρτηση της διασποράς από την τάση και αφετέρου η στοχαστική τάση έχει απαλειφθεί.

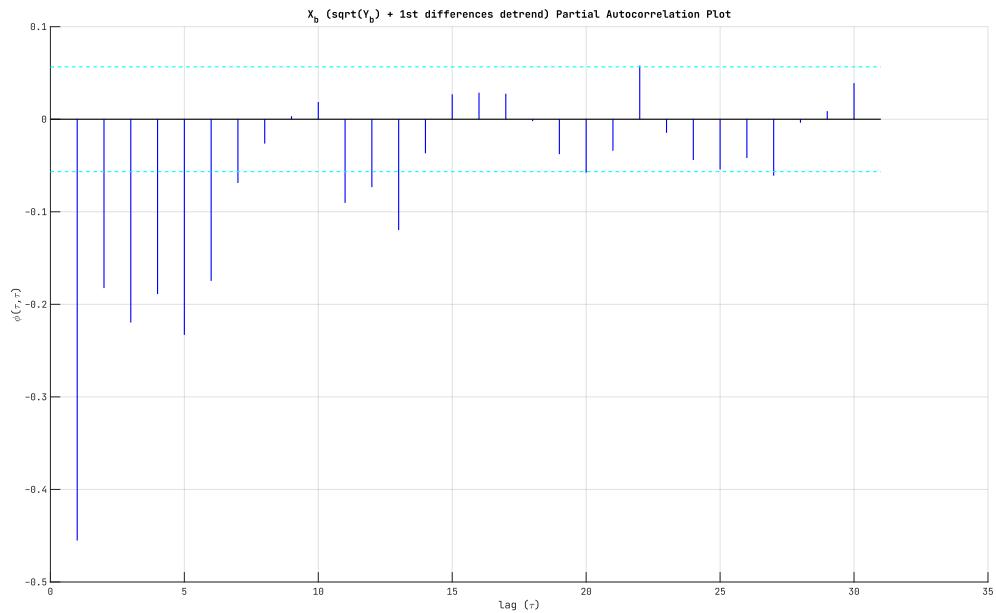


Σχήμα 3.8: Διάγραμμα ιστορίας της στάσιμης χρονοσειράς,  $\{X_b(t)\}$ , ως οι πρώτες διαφορές των τετραγωνικών ριζών της αρχικής χρονοσειράς

Για τη χροσειρά των πρώτων διαφορών των τετραγωνικών ριζών που καταλήξαμε, παραθέτονται επίσης και τα διαγράμματα δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης με τα όρια σημαντικότητας (για εμπιστοσύνη 95%):



Σχήμα 3.9: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της στάσιμης χρονοσειράς,  $r_x(\tau)$



Σχήμα 3.10: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της στάσιμης χρονοσειράς,  $\phi_x(\tau)$

τα οποία όπως είναι αναμενόμενο είναι σχεδόν ίδια με τα αντίστοιχα διαγράμματα της χρονοσειράς μόνο των πρώτων διαφορών (χωρίς δηλαδή το μετασχηματισμό της τετραγωνικής ρίζας), που δόθηκαν στα σχήματα 3.5 και 3.6 παραπάνω.

Από το νέο διάγραμμα αυτοσυσχέτισης της  $\{X_b(t)\}$  (σχήμα 3.9) φαίνεται, ωστόσο, ότι μάλλον υπάρχει κάποιος περιοδικός όρος στη χρονοσειρά, ή εποχικότητα, κάτι που έρχεται σε αντίθεση με τον ισχυρισμό μας περί στασιμότητας. Βλέποντας τις «κορυφές» στο συγκεκριμένο διάγραμμα για υστερήσεις 14 και 28 ημερών να ξεπερνούν τα όρια σημαντικότητας και

άρα να είναι (στατιστικά) μη-μηδενικές, μπορούμε να συμπεράνουμε πως στις προβολές του βίντεο Β μάλλον υπάρχει κάποιος εποχικός όρος διάρκειας 14 ημερών, τον οποίο θα προσπαθήσουμε να απαλείψουμε ακολούθως.

## 3.4 Εκτίμηση & Απαλοιφή Εποχικότητας

### 3.4.1 Εκτίμηση Εποχικού Όρου

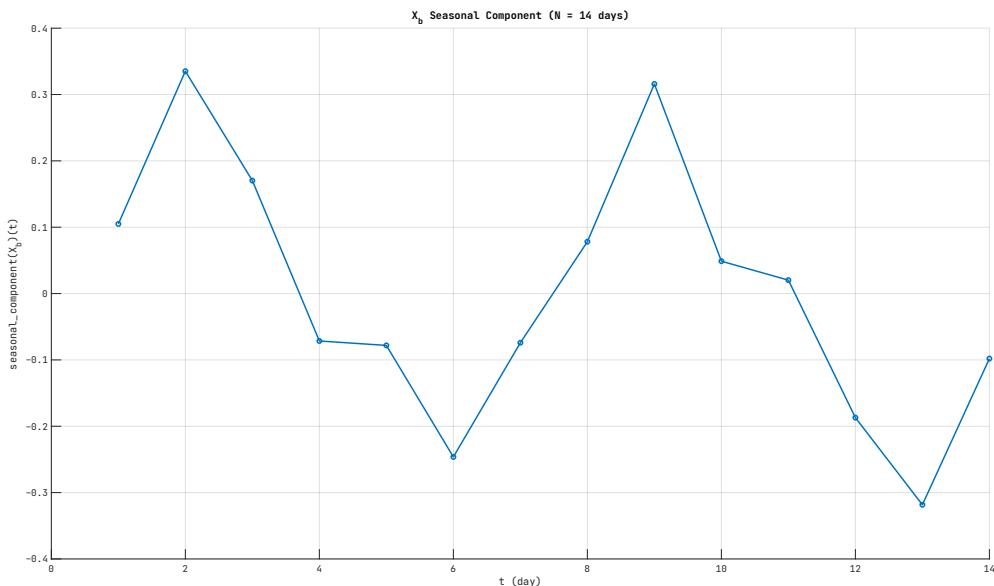
Αρχικά θα πρέπει να εκτιμήσουμε τον περιοδικό όρο της  $\{X_b(t)\}$  διάρκειας 14 ημερών. Υποθέτουμε, δηλαδή πως η  $\{X_b(t)\}$  γράφεται ως εξής:

$$X_b(t) = X_{b_{deseasoned}}(t) + \tilde{s}_b(t) \quad (3.4)$$

όπου η  $\tilde{s}_b$  είναι περιοδική ακολουθία διάρκειας 14 ημερών. Χρησιμοποιήθηκε η μέθοδος του μέσου όρου των στοιχειών της περιοδικής ακολουθίας για την εύρεση του εποχικού όρου. Έτσι, θα είναι:

$$(\tilde{s}_b)_i = \frac{1}{85} \sum_{j=0}^{84} (x_b)_{i+14j}, \quad i = 1, \dots, 14 \quad (3.5)$$

καθώς υπάρχουν 85 περίοδοι στα 1198 δείγματα της  $\{X_b(t)\}$  με μήκος περιόδου 14 δείγματα (ή εποχικότητα 14 ημερών στην αρχική χρονοσειρά προβολών του βίντεο Β). Η πρώτη περίοδος του εποχικού όρου  $\tilde{s}_b$  φαίνεται στο ακόλουθο διάγραμμα:



Σχήμα 3.11: Πρώτη περίοδος του εποχικού όρου της  $\{X_b(t)\}$ ,  $\tilde{s}_b$

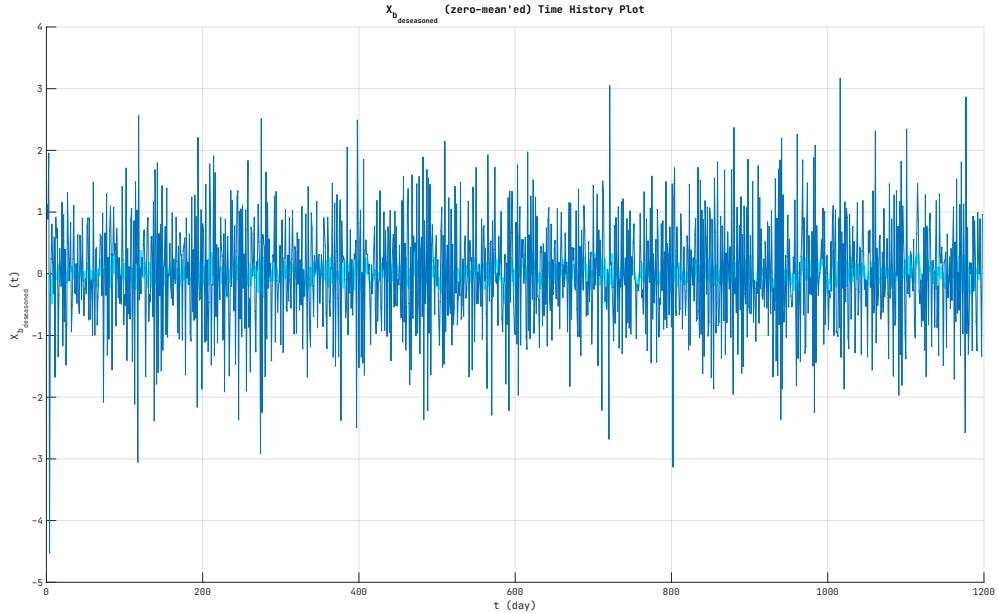
### 3.4.2 Απαλοιφή Εποχικού Όρου

Για την απαλοιφή του εποχικού όρου που εκτιμήθηκε παραπάνω, αφαιρέθηκε κάθε τιμή της περιοδικής ακολουθίας της σχέσης (3.5) από το αντίστοιχο δείγμα της  $\{X_b(t)\}$ , δίνοντας την

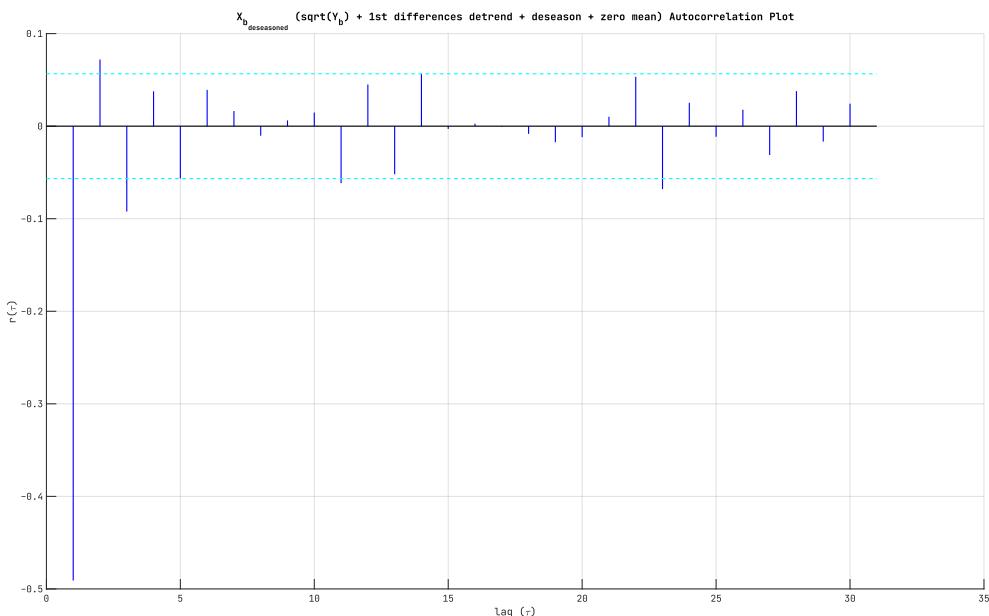
χρονοσειρά χωρίς εποχικότητα,  $\{X_{b_{deseasoned}}(t)\}$ :

$$X_{b_{deseasoned}}(t) = X_b(t) - \tilde{s}_b(t) \quad (3.6)$$

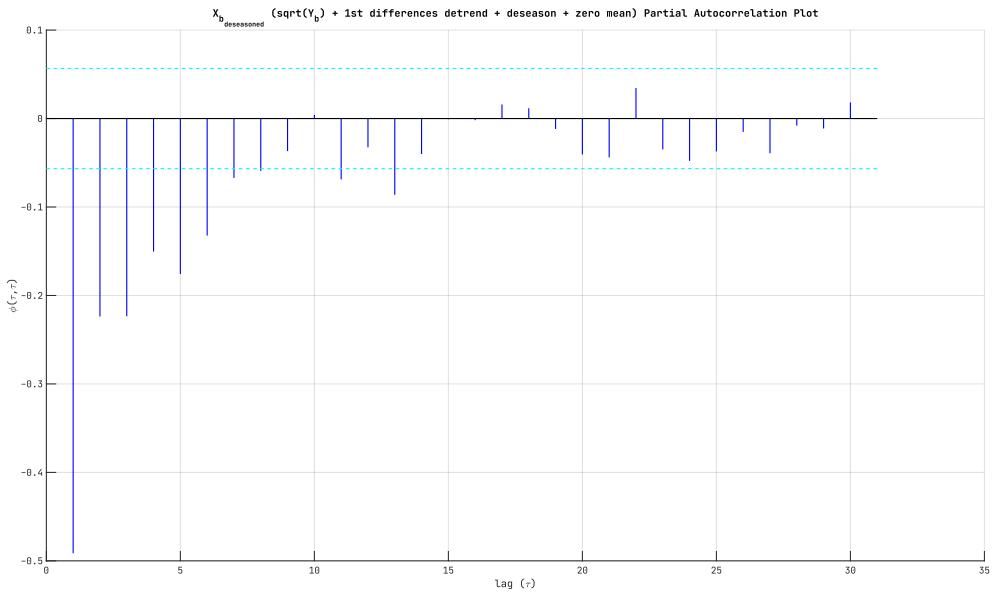
Παρακάτω, παρατίθενται για την τελευταία τα διαγράμματα ιστορίας, δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης:



Σχήμα 3.12: Διάγραμμα ιστορίας της χρονοσειράς  $\{X_b(t)\}$  απαλλαγμένης από εποχικότητα,  $\{X_{b_{deseasoned}}(t)\}$



Σχήμα 3.13: Διάγραμμα (δειγματικής) αυτοσυσχέτισης της απαλλαγμένης από εποχικότητα χρονοσειράς,  $r_x(\tau)$



Σχήμα 3.14: Διάγραμμα (δειγματικής) μερικής αυτοσυσχέτισης της απαλλαγμένης από εποχικότητα χρονοσειράς,  $\phi_x(\tau)$

Από τα παραπάνω διαγράμματα και ειδικά από το διάγραμμα της αυτοσυσχέτισης (σχήμα 3.13) φαίνεται πως αφενός ο εποχικός όρος έχει απαλειφθεί και αφετέρου έχουν μειωθεί και άλλες «κορυφές» της συνάρτησης αυτοσυσχέτισης της απαλλαγμένης από εποχικότητα χρονοσειράς.

### 3.4.3 Σύνοψη Διαδικασίας

Καταλήγουμε επομένως στα εξής για τη χρονοσειρά προβολών του βίντεο B:

Παίρνοντας τις τετραγωνικές ρίζες των τιμών τη αρχικής χρονοσειράς σταθεροποιείται η διασπορά. Ακολούθως, παίρνουμε πρώτες διαφορές στη χρονοσειρά των τετραγωνικών ριζών και έτσι απαλείφεται η στοχαστική τάση. Από το διάγραμμα αυτοσυσχέτισης της τελευταίας παρατηρούμε την ύπαρξη εποχικότητας με περίοδο 14 ημερών. Ακολούθως εκτιμούμε και απαλείφουμε τον εποχικό όρο, καταλήγοντας έτσι στην  $\{X_{b_{deseasoned}}(t)\}$ , σχέση (3.6). Αυτή είναι η **στάσιμη εκδοχή** της  $\{Y_b(t)\}$  στην οποία θα προσαρμόσουμε γραμμικό μοντέλο τύπου ARMA, όπως αναλύεται στην επόμενη ενότητα.

## 3.5 Προσαρμογή ARMA(p,q)

Σε αντιστοιχία με τη διαδικασία που ακουλήθηκε κατά τη προσαρμογή γραμμικού μοντέλου στη χρονοσειρά A, θα επιχειρήσουμε εδώ τη προσαρμογή μοντέλου ARMA και στη χρονοσειρά B.

Έτσι, συνεχίζοντας την ανάλυση από το τέλος της ενότητας 3.4, μετά την εφαρμογή των πρώτων διαφορών στη χρονοσειρά των τετραγωνικών ριζών και την απαλειφή του εποχικού όρου, έχουμε υλοποιήσει το πρώτο στάδιο προσαρμογής ενός μοντέλου ARIMA(p,d,q) που είναι η εφαρμογή διαφορών d-οστής (στη απαλλαγμένη από εποχικότητα χρονοσειρά). Άρα εδώ θα είναι  $d = 1$  και η (στάσιμη) χρονοσειρά που καταλήγουμε είναι η  $\{X_{b_{deseasoned}}(t)\}$ .

Το δεύτερο στάδιο είναι η εύρεση των παραμέτρων του μοντέλου ARMA( $p,q$ ) που θα προσαρμοστεί στη στάσιμη χρονοσειρά, κάτι που αναλύεται στην επόμενη υπο-ενότητα.

### 3.5.1 Προσαρμογή ARMA( $p,q$ ) στη στάσιμη χρονοσειρά

Οι παράμετροι  $p$  και  $q$  είναι hyperparameters του μοντέλου και επομένως το πρώτο μας μέλημα είναι να βρούμε το βέλτιστο συνδυσμό των παραμέτρων αυτών ή τάξεων του μοντέλου. Προς το σκοπό αυτό θα κάνουμε grid search για τιμές των τάξεων από 0 (απουσία του αντίστοιχου όρου) έως και 10.

Για την αξιολόγηση του κάθε συνδυασμού χρησιμοποιήθηκαν τα κριτήρια πληροφορίας Akaike (AIC) και Forward Prediction Error (FPE), όπως αυτά ορίστηκαν στις σχέσεις (2.1) και (2.2) στο βήμα 2.

Παρακάτω παρατίθενται οι τιμές του AIC για τους συνδυασμούς των παραμέτρων  $(p, q) = (0 \dots 10)$ , φυσικά με εξαίρεση το συνδυασμό  $(p, q) = (0, 0)$ :

	$p$										
	0	1	2	3	4	5	6	7	8	9	10
0		-0.562	-0.563	-0.570	-0.569	-0.568	-0.568	-0.567	-0.567	-0.566	-0.566
1	-0.385	-0.563	-0.561	-0.570	-0.570	-0.568	-0.566	-0.571	-0.567		-0.564
2	-0.434	-0.570	-0.570	-0.568	-0.568	-0.571	-0.566	-0.569	-0.568	-0.565	-0.571
3	-0.482	-0.570	-0.568	-0.571	-0.573	-0.573	-0.564	-0.568	-0.567	-0.568	-0.570
4	-0.503	-0.568	-0.568	-0.566	-0.574	-0.572	-0.571	-0.566	-0.568	-0.568	-0.570
5	-0.534	-0.567	-0.566	-0.565	-0.573	-0.571	-0.573	-0.573	-0.573	-0.573	-0.576
6	-0.551	-0.568	-0.571	-0.572	-0.568	-0.568	-0.573	-0.577	-0.575	-0.574	-0.572
7	-0.554	-0.570	-0.569	-0.568	-0.567	-0.570	-0.573	-0.575	-0.574	-0.572	-0.575
8	-0.556	-0.568	-0.568	-0.566	-0.570	-0.577	-0.570	-0.575	-0.574	-0.574	-0.572
9	-0.556	-0.567	-0.566	-0.574	-0.571	-0.570	-0.573	-0.573	-0.575	-0.578	-0.570
10	-0.555	-0.555	-0.567	-0.565	-0.563	-0.571	-0.567	-0.571		-0.571	-0.577

Πίνακας 3.1: Αναζήτηση Πλέγματος με βάση τη μετρική AIC για διάφορες τιμές των τάξεων  $(p, q)$ . Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο.

από όπου φαίνεται πως η χαμηλότερη τιμή του AIC επιτυγχανέται όταν προσαρμοζέται μοντέλο ARMA(0, 1) ή, ισοδύναμα, μοντέλο MA(1). Ακολούθως δίνονται οι τιμές του FPE για τους αντίστοιχους συνδυασμούς τιμών των  $p$  και  $q$ :

	$p$										
	0	1	2	3	4	5	6	7	8	9	10
0		0.570	0.570	0.565	0.566	0.567	0.567	0.567	0.567	0.568	0.568
q	1	0.680	0.569	0.571	0.566	0.566	0.567	0.568	0.565	0.567	
	2	0.648	0.566	0.565	0.566	0.567	0.565	0.568	0.566	0.567	0.569
	3	0.618	0.566	0.566	0.565	0.564	0.564	0.569	0.567	0.567	0.566
	4	0.604	0.567	0.567	0.568	0.563	0.564	0.565	0.568	0.566	0.566
	5	0.586	0.567	0.568	0.568	0.564	0.565	0.564	0.564	0.564	0.564
	6	0.576	0.567	0.565	0.564	0.567	0.567	0.564	0.561	0.563	0.564
	7	0.575	0.566	0.566	0.567	0.567	0.565	0.564	0.563	0.564	0.563
	8	0.573	0.567	0.567	0.568	0.565	0.562	0.565	0.563	0.563	0.565
	9	0.573	0.567	0.568	0.564	0.565	0.565	0.564	0.564	0.563	0.566
	10	0.574	0.574	0.567	0.568	0.569	0.565	0.567	0.565		0.565

Πίνακας 3.2: Αναζήτηση Πλέγματος με βάση τη μετρική FPE για διάφορες τιμές των τάξεων ( $p, q$ ). Τα κενά κελιά σηματοδοτούν ότι για τον αντίστοιχο συνδυασμό τάξεων το προκύπτον ARMA μοντέλο δεν ήταν στάσιμο, αντιστρέψιμο ή και τα δύο.

Όπως επιβεβαιώνεται και από τους δύο πίνακες παραπάνω, φαίνεται πως από τα γραμμικά μοντέλα καλύτερα προσαρμόζεται το μοντέλο ARMA(9, 9). Ωστόσο, οι τάξης αυτές είναι σχετικά μεγάλες κάτι που θα μπορούσε να οδηγήσει σε overfitting του μοντέλου και γενικά xειρότερες προβλέψεις. Παρατηρώντας όμως του παραπάνω πίνακες βλέπουμε ότι η προσαρμογή ενός μοντέλου ARMA(4, 4) οδηγεί σε πολύ κοντινά αποτελέσματα στα κριτηρία πληροφορίας έχοντας και αρκετά μικρότερες τάξεις. Συνεπώς θα ήταν δόκιμο να επιλέξουμε αυτό ως το πιο κατάλληλο γραμμικό μοντέλο για τη χρονοσειρά της προηγούμενης ενότητας εφόσον η προσαρμογή του άφηνε ασυσχέτιστα υπόλοιπα. Επίλεχθηκε η προσαρμογή αμφότερων μοντέλων και η εκ-των-υστέρων σύγκρισή τους και επιλογή του καταλληλότερου, κάτι που αναλύεται ακολούθως.

## 3.6 Διάγνωση καταλληλότητας του μοντέλου ARMA(9,9)

### 3.6.1 Τελικό μοντέλο τύπου ARMA(9,9) για τη στάσιμη χρονοσειρά

Το μοντέλο ARMA(9, 9) που εκτιμήθηκε από τη συνάρτηση `fitARMA()` και προσαρμόστηκε στη στάσιμη χρονοσειρά, δηλαδή την απαλλαγμένη από εποχικότητα χρονοσειρά των διαφορών των τετραγωνικών ριζών

$$\{X_b(t)\} = \left( \sqrt{Y_b(t)} - \sqrt{Y_b(t-1)} \right) - \{\tilde{s}_b(t)\} \quad (3.7)$$

είναι το εξής:

$$\begin{aligned} x_t = & 0.0055 - 0.264x_{t-1} - 0.275x_{t-2} - 0.164x_{t-3} - 0.04x_{t-4} \\ & + 0.166x_{t-5} + 0.121x_{t-6} + 0.228x_{t-7} + 0.838x_{t-8} \\ & + 0.012x_{t-9} + z_t - 0.509z_{t-1} + 0.103z_{t-2} - 0.143z_{t-3} \\ & - 0.077z_{t-4} - 0.26z_{t-5} + 0.052z_{t-6} - 0.144z_{t-7} \\ & - 0.647z_{t-8} + 0.648z_{t-9}, \quad t = 10, \dots, 1199 \end{aligned} \quad (3.8)$$

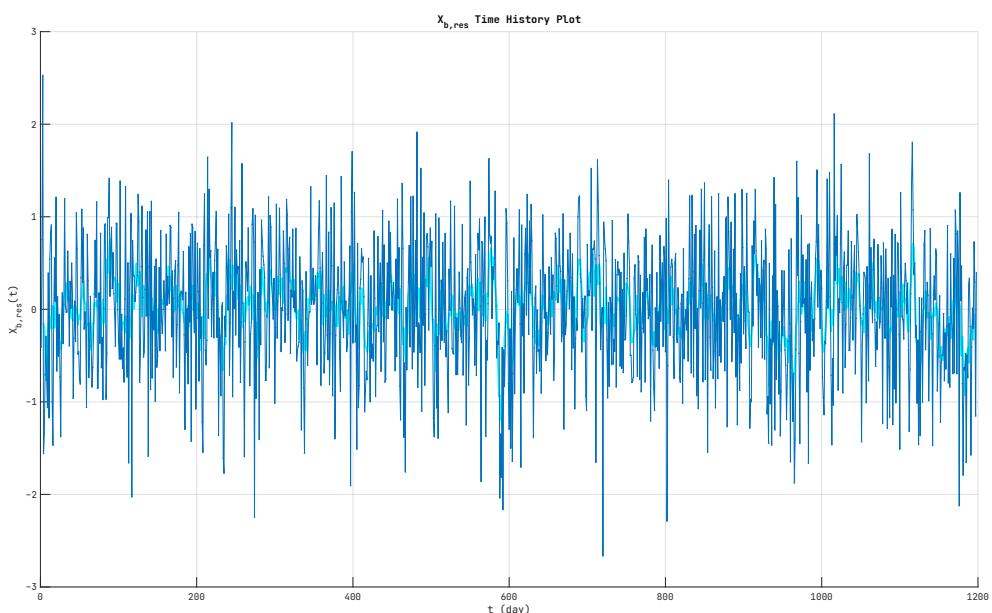
όπου ο μέσος όρος της στάσιμης χρονοσειράς  $\{X_{b_{\text{desasoned}}}(t)\}$  είναι  $\bar{x} = 0.0055$ . Ενώ η εκτίμηση της τυπικής απόκλισης των σφαλμάτων ή υπολοίπων προσαρμογής βρέθηκε να είναι

ίση με  $s_z = 0.7436$  (εκτίμηση διασποράς ίση με  $s_z^2 = 0.5529$ ). Το αντίστοιχο μοντέλο με το οποίο θα προσεγγίζαμε την αρχική χρονοσειρά των προβολών του βίντεο A, λαμβάνοντας υπόψη και τον εποχικό όρο, θα είναι:

$$Y_t = \left( \sqrt{Y_{t-1}} + X_t + \tilde{S}_b(t) \right)^2, \quad t = 10, \dots, 1199 \quad (3.9)$$

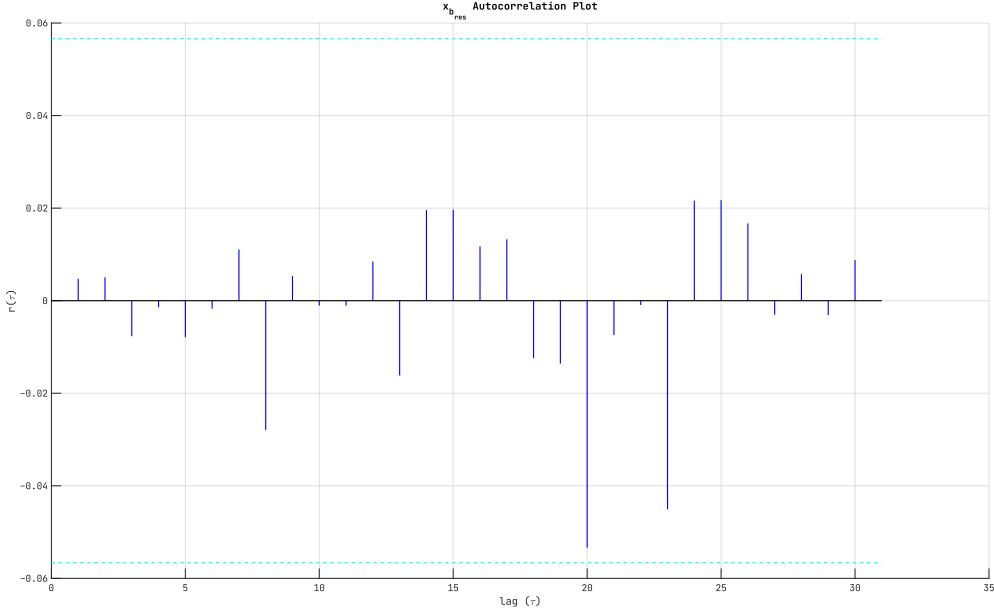
### 3.6.2 Διάγνωση Καταλληλότητας & Σφάλματα Προσαρμογής

Ακολούθως, θα κάνουμε διάγνωση καταλληλότητας του μοντέλου  $ARMA(9, 9)$  που καταλήξαμε προκειμένου να αποφανθούμε έαν αυτό το μοντέλο αντλεί όλη τη πληροφορία της στάσιμης χρονοσειράς, αφήνοντας ασυσχέτιστα υπόλοιπα (λευκό θόρυβο). Επομένως, θα κάνουμε έλεγχο ανεξαρτησίας στη σειρά των υπολοίπων τόσο με βάση τη δειγματική τους αυτοσυσχέτιση όσο και με τον έλεγχο Portmanteau. Σε πρώτη φάση, όμως, το διάγραμμα ιστορίας της σειράς των υπολοίπων ή σφαλμάτων προσαρμογής δίνεται ακολούθως:

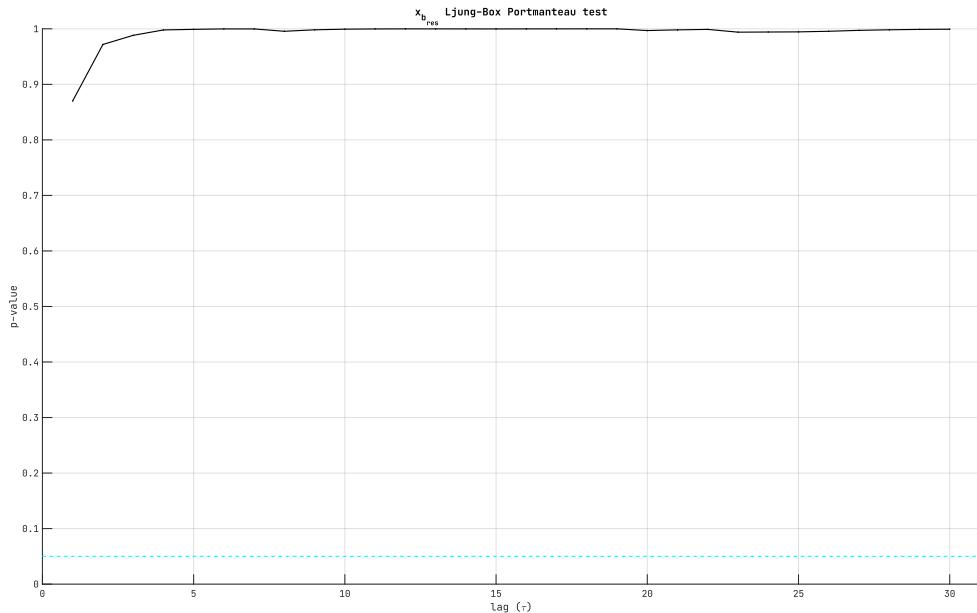


Σχήμα 3.15: Διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (3.8),  $\{X_{b,\text{deseasoned},\text{res}}(t) = \hat{z}(t)\}$

Στη συνέχεια δίνονται τα διαγράμματα δειγματικής αυτοσυσχέτισης και p-values του Ljung & Box test για μέγιστη υστέρηση  $\tau$  από 1 έως 30:



Σχήμα 3.16: Διάγραμμα δειγματικής αυτοσυσχέτισης της σειράς των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (3.8),  $r_{\hat{z}_b}(\tau)$



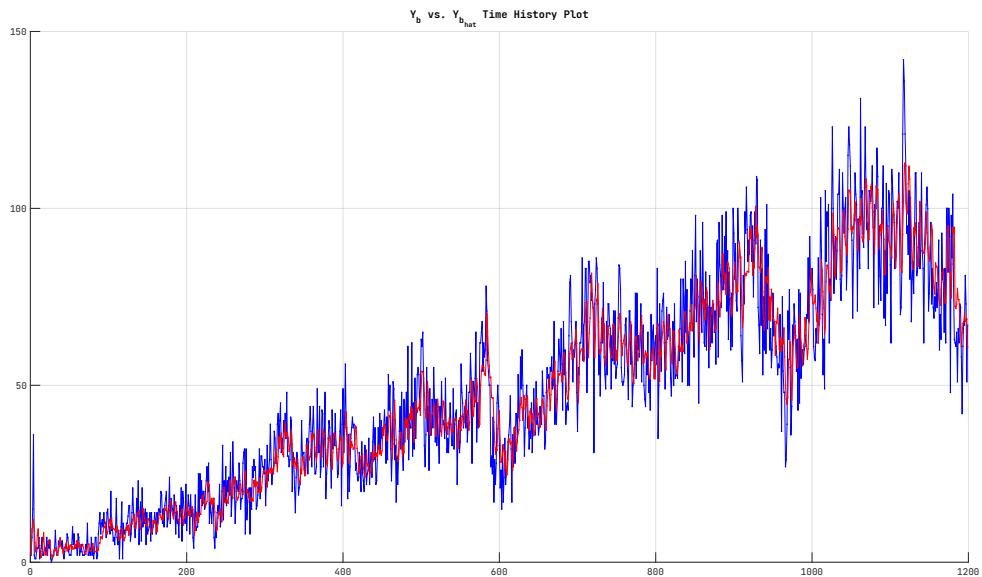
Σχήμα 3.17: Διάγραμμα των p-values του στατιστικού ελέγχου ανεξαρτησίας Portmanteau (Ljung & Box test) των υπολοίπων της προσαρμογής του μοντέλου της σχέσης (3.8). Σημειώνεται με διακεκομένη γραμμή το όριο απόφασης όπου όπως φαίνεται η μηδενική υπόθεση  $H_0$  (η σειρά των υπολοίπων είναι iid) δεν απόρριπτεται για καμία από τις 30 υστερήσεις.

Αμφότερα τα σχήματα 3.16 και 3.17 παραπάνω φανερώνουν ότι η προσαρμογή του ARMA(9, 9) μοντέλου της σχέσης (3.8) είναι επιτυχής αφού αφήνει ασυσχέτιστα υπόλοιπα. Αυτό φαίνεται στο διάγραμμα αυτοσυσχετίσεων των υπολοίπων όπου για καμία υστέρηση η αυτοσυσχέτιση δεν ξεπερνάει το όριο σημαντικότητας, όντας όλες στατιστικά μηδενικές. Αντίστοιχα αποτε-

λέσμστα λαμβάνουμε και από τον έλεγχο ανεξαρτησίας Portmanteau όπου η μηδενική υπόθεση πως η σειρά των υπολοίπων είναι iid δεν απορρίπτεται για καμία από τις 30 υστερήσεις. Με ασφάλεια μπορούμε να πούμε πως η προσαρμογή αφήνει λευκό θόρυβο ως σειρά υπολοίπων.

Τέλος, για λόγους πληρότητας παρεθέτουμε το NRMSE των σφαλμάτων προσαρμογής (στη στάσιμη χρονοσειρά) για πρόβλεψη ενός βήματος μπροστά καθώς και τις ίδιες τις προβλέψεις μαζί με την αρχική χρονοσειρά προβολών του βίντεο B με βάση τη σχέση (3.9), παρακάτω:

$$NRMSE(\hat{X}_{b_{deseasoned}}, X_{b_{deseasoned}}) = 0.7589 \quad (3.10)$$



Σχήμα 3.18: Διάγραμμα ιστορίας της αρχικής χρονοσειράς προβολών του βίντεο B (μπλε) καθώς και τις προβλέψεις ενός βήματος μπροστά αυτής με βάση το προσαρμοσμένο μοντέλο ARMA(9,9) και τη σχέση (3.9) (κόκκινο).

## 3.7 Διάγνωση καταλληλότητας του μοντέλου ARMA(4,4)

### 3.7.1 Τελικό μοντέλο τύπου ARMA(4,4) για τη στάσιμη χρονοσειρά

Το μοντέλο ARMA(4,4) που εκτιμήθηκε από τη συνάρτηση `FitARMA()` και προσαρμόστηκε στη στάσιμη χρονοσειρά, δηλαδή την απαλλαγμένη από εποχικότητα χρονοσειρά των διαφορών των τετραγωνικών ριζών είναι το εξής:

$$\begin{aligned} x_{t(4,4)} &= 0.0055 + 0.806x_{t-1(4,4)} - 0.329x_{t-2(4,4)} - 0.570x_{t-3(4,4)} \\ &\quad + 0.112x_{t-4(4,4)} + z_t - 1.572z_{t-1} + 0.970z_{t-2} \\ &\quad + 0.180z_{t-3} - 0.405z_{t-4}, \quad t = 5, \dots, 1199 \end{aligned} \quad (3.11)$$

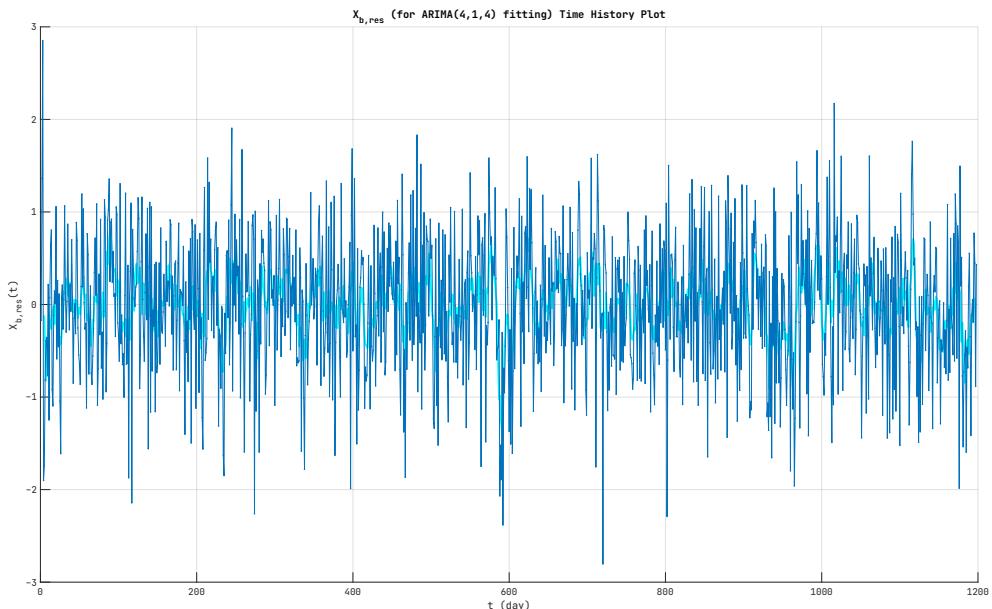
όπου ο μέσος όρος της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  είναι  $\bar{x} = 0.0055$ . Ενώ η εκτίμηση της τυπικής απόκλισης των σφαλμάτων ή υπολοίπων προσαρμογής του ARMA(4,4)

βρέθηκε να είναι ίση με  $s_z = 0.7481$  (εκτίμηση διασποράς ίση με  $s_z^2 = 0.5596$ ), είναι δηλαδή οριακά μεγαλύτερη από τα αντίστοιχα της προσαρμογής  $ARMA(9,9)$ . Το αντίστοιχο μοντέλο με το οποίο θα προσεγγίζαμε την αρχική χρονοσειρά των προβολών του βίντεο A, λαμβάνοντας υπόψη και τον εποχικό όρο, θα είναι:

$$Y_t = \left( \sqrt{Y_{t-1}} + X_{t(4,4)} + \tilde{S}_b(t) \right)^2, \quad t = 5, \dots, 1199 \quad (3.12)$$

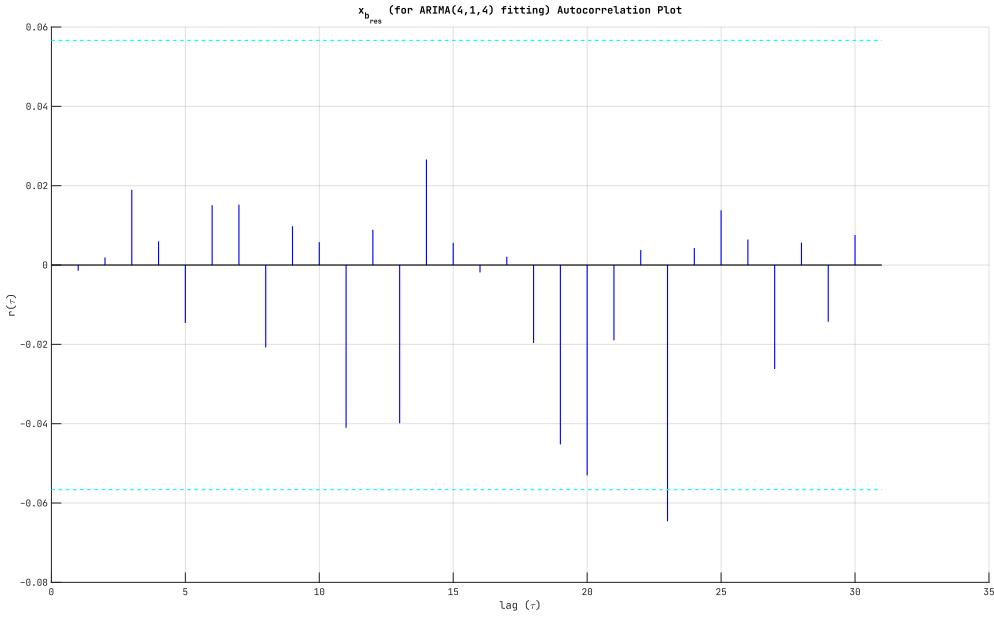
### 3.7.2 Διάγνωση Καταλληλότητας & Σφάλματα Προσαρμογής

Ακολούθως, θα κάνουμε διάγνωση καταλληλότητας του μοντέλου  $ARMA(4,4)$  ακολουθώντας πανομοιότυπη διαδικασία με αυτή της διάγνωσης καταλληλότητας του  $ARMA(9,9)$  μοντέλου, δηλαδή κάνοντας έλεγχο ανεξαρτησίας στη σειρά των υπολοίπων τόσο με βάση τη δειγματική τους αυτοσυσχέτιση όσο και με τον έλεγχο Portmanteau. Σε πρώτη φάση, όμως, το διάγραμμα ιστορίας της σειράς των υπολοίπων ή σφαλμάτων προσαρμογής δίνεται ακολούθως:

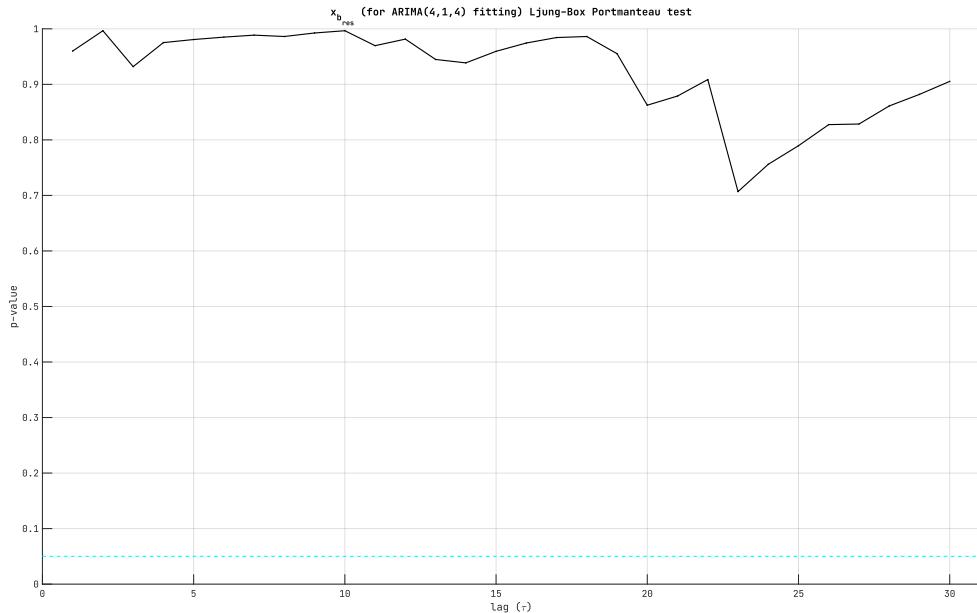


Σχήμα 3.19: Διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων της προσαρμογής του μοντέλου  $ARMA(4,4)$  της σχέσης (3.11),  $\{X_{b_{deseasoned},res(4,4)}(t) = \hat{z}(t)\}$

Στη συνέχεια δίνονται τα διαγράμματα δειγματικής αυτοσυσχέτισης και p-values του Ljung & Box test για μέγιστη υστέρηση  $\tau$  από 1 έως 30:



Σχήμα 3.20: Διάγραμμα δειγματικής αυτοσυσχέτισης της σειράς των υπολοίπων της προσαρμογής του μοντέλου ARMA(4, 4) της σχέσης (3.11),  $r_{\hat{z}_b}(\tau)$



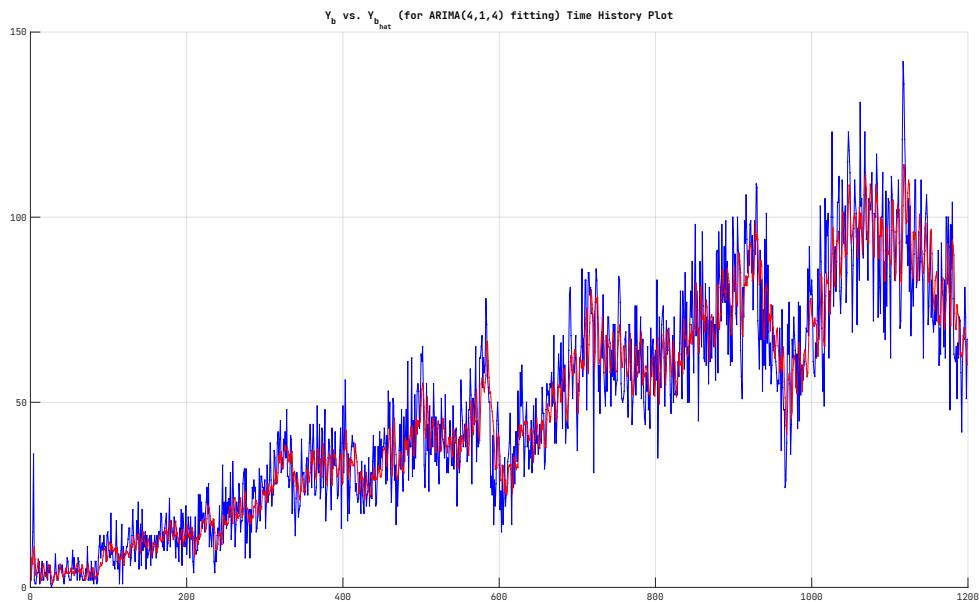
Σχήμα 3.21: Διάγραμμα των p-values του στατιστικού ελέγχου ανεξαρτησίας Portmanteau (Ljung & Box test) των υπολοίπων της προσαρμογής του μοντέλου ARMA(4, 4) της σχέσης (3.11). Σημειώνεται με διακεκομένη γραμμή το όριο απόφασης όπου όπως φαίνεται η μηδενική υπόθεση  $H_0$  (η σειρά των υπολοίπων είναι iid) δεν απόρριπτεται για καμία από τις 30 υστερήσεις.

Αμφότερα τα σχήματα 3.20 και 3.21 παραπάνω φανερώνουν ότι η προσαρμογή του ARMA(4, 4) μοντέλου της σχέσης (3.11) είναι επιτυχής αφού αφήνει ασυσχέτιστα υπόλοιπα. Αυτό φαίνεται στο διάγραμμα αυτοσυσχέτισεων των υπολοίπων όπου μόνο για μία υστέρηση (για τη 23) η αυτοσυσχέτιση ξεπερνάει το όριο σημαντικότητας (κάτι που επιτρέπεται από το επίπεδο

εμπιστοσύνης) ενώ όλες οι υπόλοιπες αυτοσυσχετίσεις είναι στατιστικά μηδενικές. Αντίστοιχα αποτελέσματα λαμβάνουμε και από τον έλεγχο ανεξαρτησίας Portmanteau όπου η μηδενική υπόθεση πως η σειρά των υπολοίπων είναι iid δεν απορρίπτεται για καμία από τις 30 υστερήσεις. Με ασφάλεια μπορούμε να πούμε πως η προσαρμογή αφήνει λευκό θόρυβο ως σειρά υπολοίπων.

Τέλος, για λόγους πληρότητας παρεθέτουμε το NRMSE των σφαλμάτων προσαρμογής του ARMA(4, 4) (στη στάσιμη χρονοσειρά) για πρόβλεψη ενός βήματος μπροστά καθώς και τις ίδιες τις προβλέψεις μαζί με την αρχική χρονοσειρά προβολών του βίντεο B με βάση τη σχέση (3.12), παρακάτω:

$$NRMSE(\hat{X}_{b_{deseasoned}(4,4)}, X_{b_{deseasoned}}) = 0.7712 \quad (3.13)$$



Σχήμα 3.22: Διάγραμμα ιστορίας της αρχικής χρονοσειράς προβολών του βίντεο B (μπλε) καθώς και τις προβλέψεις ενός βήματος μπροστά αυτής με βάση το προσαρμοσμένο μοντέλο ARMA(4, 4) και τη σχέση (3.12) (κόκκινο).

### 3.7.3 Συμπερασματικά σχόλια

Βλέποντας λοιπόν από την ανάλυση που προηγήθηκε ότι η προσαρμογή αμφότερων των γραμμικών μοντέλων ARMA(4, 4) και ARMA(9, 9) οδηγεί σε ασυσχέτιστα υπόλοιπα αντλώντας όλη τη πληροφορία (που μπορεί να αντλήσει ένα γραμμικό μοντέλο) από τη στάσιμη χρονοσειρά,  $\{X_{b_{deseasoned}}(t)\}$ , καταλήγουμε στο μοντέλο ARMA(4, 4) της σχέσης (3.11) ως το πιο κατάλληλο καθώς το γεγονός ότι έχει μικρότερες τάξεις θα μας οδηγήσει σε πιο ασφαλείς προβλέψεις και μείωση της πιθανότητας overfitting.

## Βήμα 4

# Εντοπισμός Σημείων Αλλαγής της Χρονοσειράς A

Παιρνώντας στο δεύτερο μέρος της εργασίας, βήματα 4 έως 6, θα γίνει προσπάθεια ανίχνευσης με αυτόματο τρόπο σημαντικών αλλαγών στις δοθείσες χρονοσειρές. Συγκεκριμένα, ο μέσος όρος των απόλυτων τιμών των σφαλμάτων πρόβλεψης των στάσιμων χρονοσειρών (που προέκυψαν στα βήματα 1 έως 3) για έως και  $T$  βήματα μπροστά, θα χρησιμοποιηθεί για την εξαγωγή σημείων αλλαγής. Για συμφωνία με το αντίστοιχο paper, θα ονομάσουμε τα σημεία αυτά σημεία αλλαγής μέσης τίμης (Mean Change Points ή MCPs).

Στο παρόν βήμα, θα εφαρμόσουμε το κριτήριο εύρεσης σημείων αλλαγής (MCPs) στη χρονοσειρά προβολών του βίντεο A, κάτι που αναλυέται στις ενότητες και υπο-ενότητες που ακολουθούν.

### 4.1 Τρόπος Επιλογής Σημείων Αλλαγής

Σύμφωνα με την εκφώνηση του αντίστοιχου σταδίου της εργασίας, το στατιστικό που θα χρησιμοποιηθεί για την επιλογή των σημείων αλλαγής είναι το εξής:

$$S_n = \frac{1}{T} \sum_{k=1}^T |x_{n+k} - x_n(k)|, \quad n = n_0, \dots, 1199 - T \quad (4.1)$$

όπου  $T$  είναι ο μέγιστος αριθμός βημάτων πρόβλεψης μπροστά, ενώ  $n_0$  είναι ο αριθμός των δειγμάτων «εκπαίδευσης» για προσαρμογή του μοντέλου.

Στο σημείο αυτό πρέπει να διευκρινιστεί ότι τα μοντέλα που θα χρησιμοποιηθούν σε αυτό το βήμα καθώς και στο επόμενο είναι γραμμικά μοντέλα τύπου ARMA τα οποία έχουν προσαρμοστεί στις **στάσιμες χρονοσειρές** που προέκυψαν από την ανάλυση στα βήματα 1 έως 3 που προήγηθηκε.

Επιστρέφοντας στην επιλογή των σημείων αλλαγής (MCPs), αυτή θα γίνεται όποτε το στατιστικό  $S_n$  της σχέχης (4.1) παραπάνω ξεπεράσει μία προκαθορισμένη τιμή,  $a$ . Ακολουθώντας την προσέγγιση που προτείνεται στην εκφώνηση, επιλέγουμε η τιμή αυτή σχετίζεται με την τυπική απόκλιση του training set ως εξής:

$$a = \bar{\mu}_{std} * s_x \quad (4.2)$$

όπου  $\hat{M}_{std}$  είναι βελτιστοποίησμη hyperparameter, ενώ  $s_x$  είναι η τυπική απόκλιση των τιμών της στάσιμης χρονοσειράς που χρησιμοποιήθηκαν για την προσαρμογή του γραμμικού μοντέλου.

Ετσι, όποτε το  $S_n > \alpha$  θα προσθέτουμε το  $n + T$  στα σημεία αλλαγής. Σε αντίθετη περίπτωση το  $n$  θα αυξάνεται κατά 1. Για την εκ' νέου προσαρμογή του γραμμικού μοντέλου καθώς προχωράμε στο χρόνο στη στάσιμη χρονοσειρά υπάρχουν τρείς επιλογές:

- *Επιλογή «a»*: Διατήρηση του μοντέλου που εκτιμήθηκε στις πρώτες  $n_0$  παρατηρήσεις της χρονοσειράς
- *Επιλογή «b»*: Επαναπροσαρμογή του μοντέλου σε κάθε επόμενη χρονική στιγμή  $n$  χρησιμοποιώντας τις  $n_0$  πιο πρόσφατες παρατηρήσεις
- *Επιλογή «c»*: Διατήρηση του μοντέλου έως ότου βρεθεί σημείο αλλαγής και επαναπροσαρμογή του στις  $n_0$  πιο πρόσφατες παρατηρήσεις όταν βρεθεί (και αλλαχτεί η τρέχουσα χρονική στιγμή)

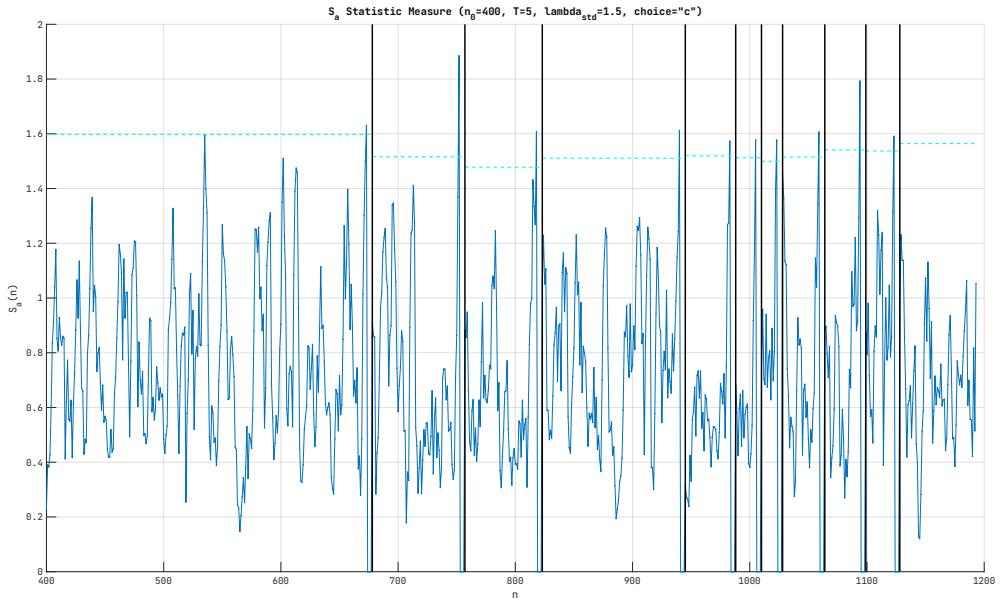
Έχουν υλοποιηθεί και οι τρεις, και άρα η επιλογή αποτελεί και αυτή μια hyperparameter της μεθόδου, ίσως όχι βελτιστοποίησμη αλλά μάλλον επιλέξιμη από πριν.

## 4.2 Αρχική Εφαρμογή

Αρχικά, θα δώσουμε κάποιες τιμές στις hyperparameters της μεθόδου, δηλαδή στο μέγεθος εκπαίδευσης  $n_0$ , στον ορίζοντα πρόβλεψης,  $T$ , στην επιλογή αναπροσαρμογής του μοντέλου και στην παράμετρο  $\hat{M}_{std}$ , και θα τρέξουμε την παραπάνω μέθοδο στη στάσιμη χρονοσειρά που προέκυψε από το βήμα 2,  $\{X_a(t)\}$ , και στο MA(1) μοντέλο που προσαρμόστηκε σε αυτή (σχέση 2.3). Έτσι, για τις παραμέτρους θα έχουμε:

- $n_0 = 400$ : Όπως προτείνεται στην εκφώνηση, το μέγεθος εκπαίδευσης θα είναι 400 παρατηρήσεις
- $T = 5$ : Επίσης δίνεται έμμεσα στην εκφώνηση, το κριτήριο αρχικά θα υπολίζεται για προβλέψεις έως και 5 βήματα εμπρός
- *choice = "c"*: Διατήρηση του μοντέλου έως ότου βρεθεί σημείο αλλαγής και επαναπροσαρμογή του στις 400 πιο πρόσφατες παρατηρήσεις από το  $n + 5$  όταν βρεθεί
- $\hat{M}_{std} = 1.5$ : Όχι σε συμφωνία με την προτεινόμενη τιμή, καθώς αυτή δίνει μόλις ένα σημείο αλλαγής για τις υπόλοιπες επιλεγμένες παραμέτρους. Άρα, κριτήριο επλογής MCPs όταν  $S_n > 1.5 * s_x$

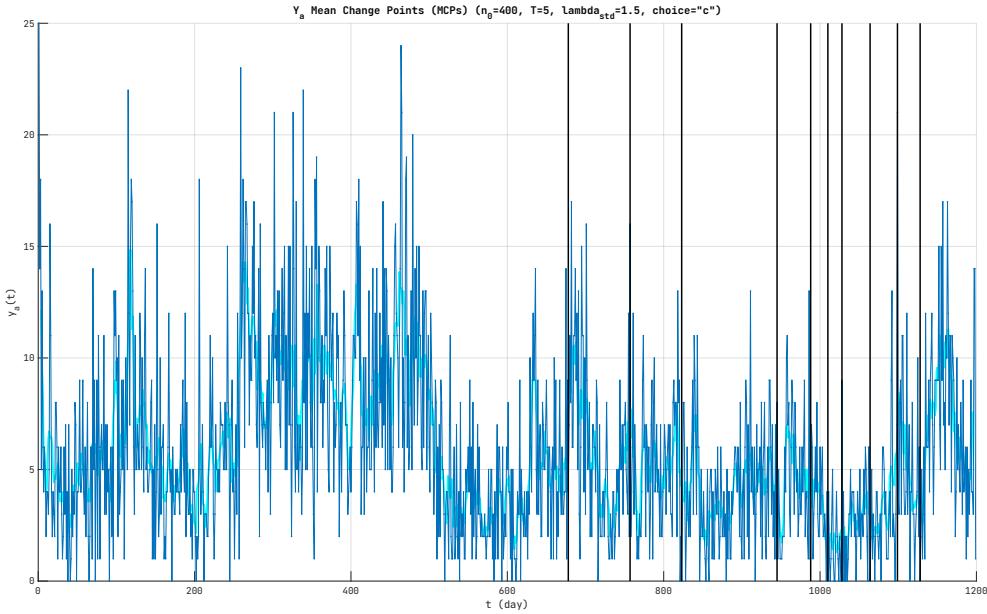
Παρακάτω, φαίνονται τα σημεία αλλαγής που προκύπτουν από την εκτέλεση της μεθόδους με τις παραπάνω παραμέτρους:



Σχήμα 4.1: Τιμές στατιστικού  $S_n$  για έως και 5 βήματα μπροστά πρόβλεψη με MA(1) της στάσιμης χρονοσειράς  $\{X_\alpha(t)\}$ . Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο). Το  $n$  ξεκινάει από το 400 και άρα το  $S_n$  δεν ορίζεται για τιμές  $n < n_0 = 400$ .

Όπως φαίνεται στο παραπάνω διάγραμμα, επειδή έχει χρησιμοποιηθεί η επιλογή «c» για την αναπροσαρμογή του μοντέλου, κάθε φορά που εντοπίζεται σημείο αλλαγής το μοντέλο επανεκτιμάται σε νέο training set (που είναι οι τελευταίες 400 παρατηρήσεις από τη στιγμή  $n + 5$ ) και άρα αλλάζει η δειγματική τυπική απόκλιση των παρατηρήσεων του training set. Αυτό φαίνεται ως αλλαγή στο επίπεδο της cyan διακεκομμένης γραμμής.

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A και κατόπιν ακολουθεί σύντομος σχολιασμός:



Σχήμα 4.2: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_\alpha(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing)

Βλέποντας το πού «πέφτουν» τα σημεία αλλαγής στην αρχική χρονοσειρά προβολών του βιντέο Α παρατηρούμε πως αυτά δεν είναι σε τυχαίες θέσεις. Βρίσκοντα είτε σε κορυφές, δηλαδή σε θέσεις που αριστερά υπάρχει τοπικά αυξητική τάση και δεξιά τοπικά πτωτική (ακόμα και εάν αυτό γίνεται για λίγες παρατηρήσεις), ή σε βυθούς, δηλαδή σε σημεία που αριστερά υπάρχει τοπικά πτωτική τάση ενώ δεξία αυτή αλλάζει και γίνεται αυξητική. Είναι κάπως λογικό οι προβλέψεις μας για τις γειτονιές τέτοιων σημείων (ακόμα και εάν γίνονται μέσω στάσιμων εκδοχών των αρχικών χρονοσειρών) με γραμμικά μοντέλα να εμφανίζουν αρκετά σημαντικά σφάλματα ώστε να σηματοδοτηθούν σημεία αλλαγής.

Ακολουθεί μια πιο ενδελεχής ανάλυση για την επιλογή των hyperparameters που σε αυτήν την υποενότητα έγινε κάπως αυθαίρετα.

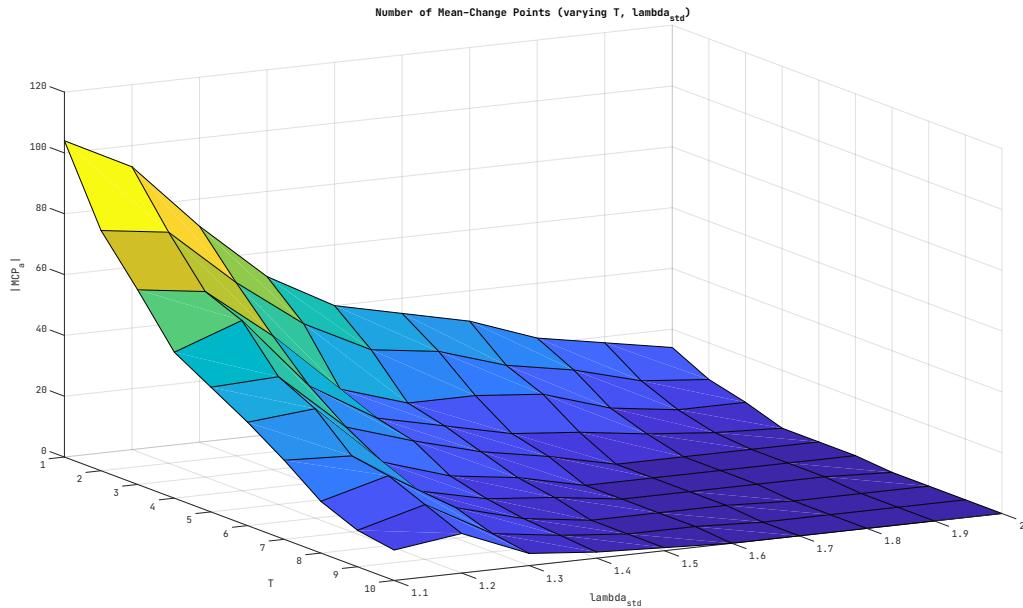
### 4.3 Επιλογή Βέλτιστων Παραμέτρων

Για επιλογή βέλτιστων τιμών στις hyperparameters της μεθόδου και συγκεκριμένα στον ορίζοντα πρόβλεψης,  $T$ , και στο  $\lambda_{std}$  του ορίου απόφασης,  $\alpha$ , θα κάνουμε αναζήτηση πλέγματος ως προς αυτές. Ως μετρικές για αξιολόγηση του κάθε συνδυασμού των παραμέτρων αυτών χρησιμοποιήθηκαν οι εξής δύο (2):

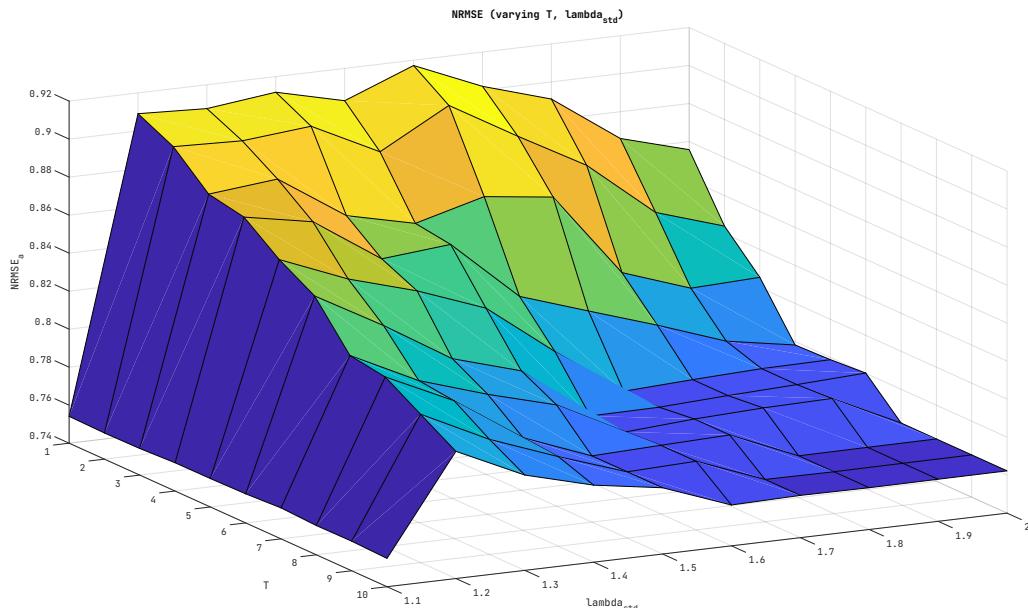
- **Αριθμός MCPs:** Θέλουμε ο αριθμός των σημείων αλλαγής (MCPs) να μην είναι πολύ μεγάλος καθώς κάτι τέτοιο κάνει λιγότερο αξιόπιστη την όλη μέθοδο, αλλά να μην είναι και πολύ μικρός ώστε να υπάρχει νόημα χρήσης της μεθόδου (π.χ. αν έβγαζε ένα ή δύο σημεία αλλαγής τότε μάλλον δεν θα άξιζε η on-line χρήση της μεθόδου για πρόβλεψη αλλαγών της ζήτησης σε εφαρμογές όπως αυτή του δοθέντος paper)
- **NRMSE Προβλέψεων:** Σε αυτή τη μετρική υπολογίζουμε το NRMSE μεταξύ των προβλέψεων που χρησιμοποιήθηκαν για τον εντοπισμό των σημείων αλλαγής και των πραγ-

ματικών τιμών της στάσιμης χρονοσειράς

Αρχικά παραθέτονται σε διαγράμματα τύπου surf τα αποτελέσματα αναζήτησης πλέγματος ως προς τις παραπάνω μετρικές ενώ στη συνέχεια σχολιάζεται ο τρόπος επιλογής προσεγγεστικά βέλτιστων παραμάτρων αλλά και οι τελικές τους τιμές.



Σχήμα 4.3: Αριθμός σημείων αλλαγής,  $|MCP|$ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $a$



Σχήμα 4.4: NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs), για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $a$

Γενικότερα, αναζητούμε τα «γόνατα» στις αντίστοιχες τρισδιάστατες καμπύλες έτσι ώστε πε-

ραιτέρω μεταβολές των αντίστοιχων παραμέτρων να μην είναι πλέον επικερδείς.

Επικεντρώνοντας στο πρώτο διάγραμμα και δεδομένου ότι θέλουμε ο αριθμός των MCPs να μήν είναι πολύ μεγάλος ή πολύ μικρός, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \in [5, 8] \quad \& \quad \lambda_{std} \in [1.3, 1.6] \quad (4.3)$$

Επικεντρώνοντας τώρα στο διάγραμμα των NRMSEs και δεδομένου ότι θέλουμε το NRMSE να είναι κατά το δυνατό μικρό, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \geq 6 \quad \& \quad \lambda_{std} \geq 1.5 \quad (4.4)$$

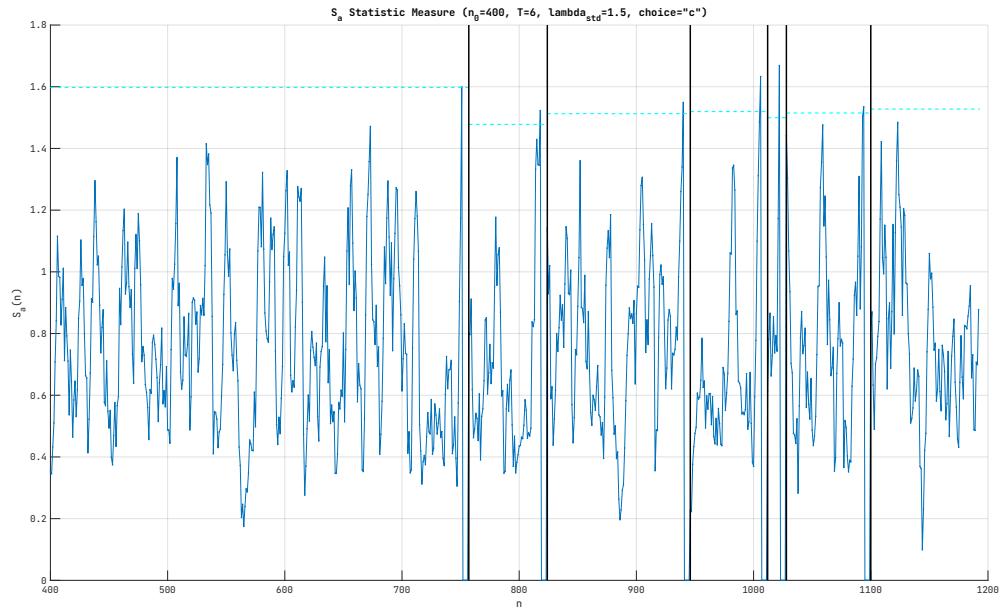
Συνδυάζοντας τις σχέσεις (4.3) και (4.4) παραπάνω καταλήγουμε ότι οι hyperparameters που θα χρησιμοποιηθούν για την εφαρμογή της μεθόδου αυτόματης εύρεσης χρονικών σημείων αλλαγής θα είναι: **T = 6 βήματα** και  **$\lambda_{std} = 1.5$** . Οι αντίστοιχες τιμές του grid search είναι: **|MCP| = 6 MCPs** και **NRMSE = 0.8**.

## 4.4 Εφαρμογή με Βέλτιστες Παραμέτρους

Χρησιμοποιώντας τις επιλεγμένες τιμές για τις hyperparameters της μεθόδου, δηλαδή ορίζοντα πρόβλεψης έως και 6 βημάτων εμπρός,  $T = 6$ , και παράμετρο ορίου απόφασης στο 1.5,  $\lambda_{std} = 1.5$ , θα τρέξουμε την παραπάνω μέθοδο στη στάσιμη χρονοσειρά που προέκυψε από το βήμα 2,  $\{X_a(t)\}$ , και στο MA(1) μοντέλο που προσαρμόστηκε σε αυτή (σχέση 2.3). Παρακάτω, φαίνονται τα σημεία αλλαγής που προκύπτουν από την εκτέλεση της μεθόδους με τις παραπάνω παραμέτρους για κάθε μια από τις επιλογές αναπροσαρμογής του μοντέλου («a», «b» ή «c»).

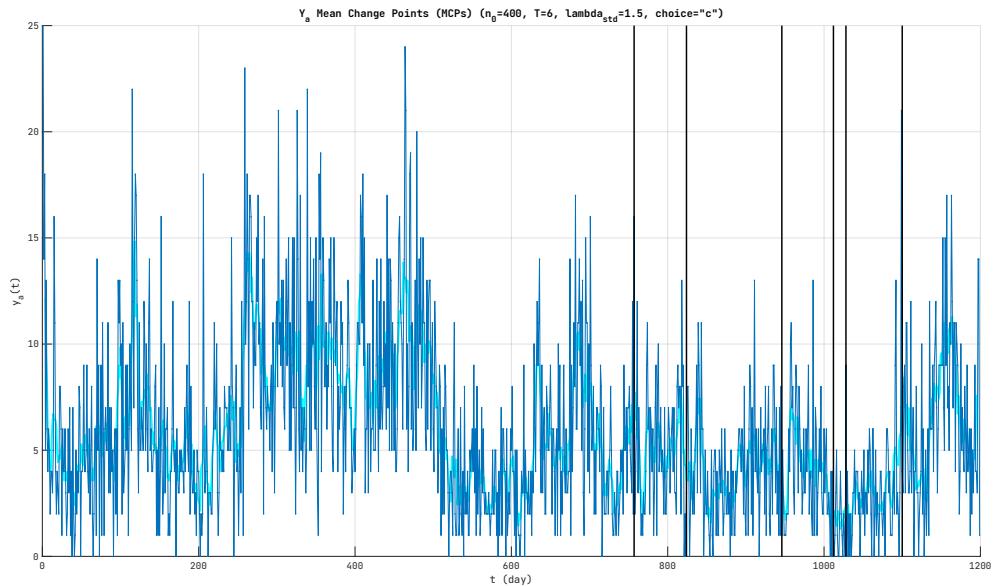
### 4.4.1 Αναπροσαρμογή όταν βρεθεί σημείο αλλαγής

Επιλογή «c»



Σχήμα 4.5: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με  $MA(1)$  της στάσιμης χρονοσειράς  $\{X_a(t)\}$  και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.800, 28.9sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:

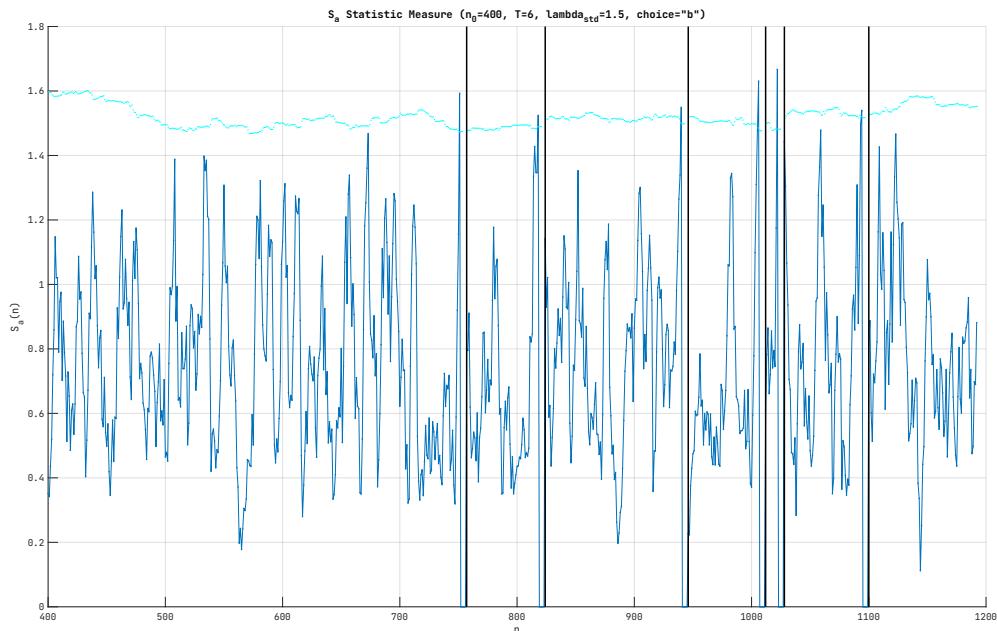


Σχήμα 4.6: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_a(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$  smoothing) - επιλογή «c»

#### 4.4.2 Αναπροσαρμογή σε κάθε χρονική στιγμή

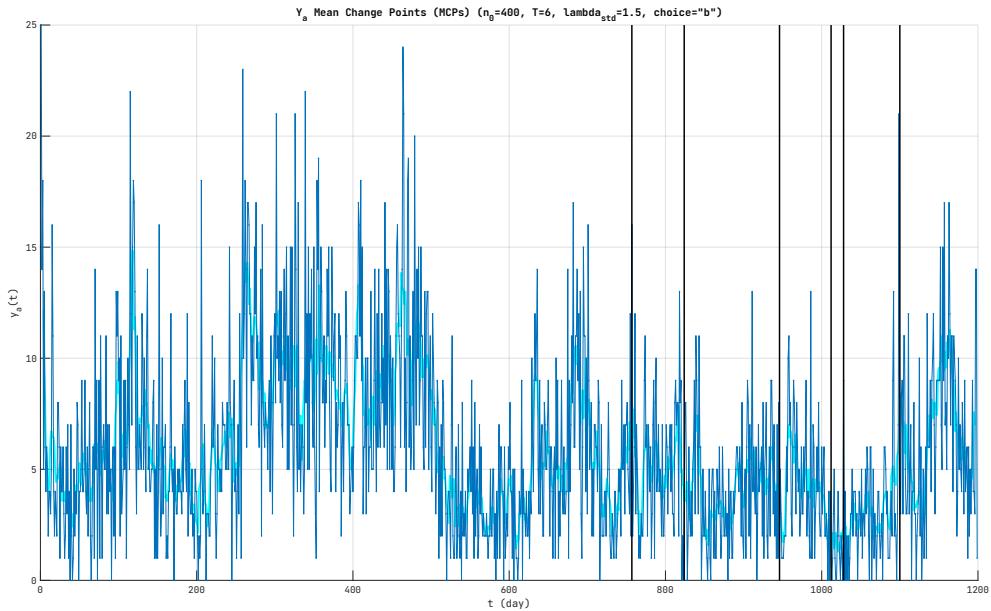
Επιλογή «b»

Τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή) ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 4.7: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με  $MA(1)$  της στάσιμης χρονοσειράς  $\{X_a(t)\}$  και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.802, 86.1sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:



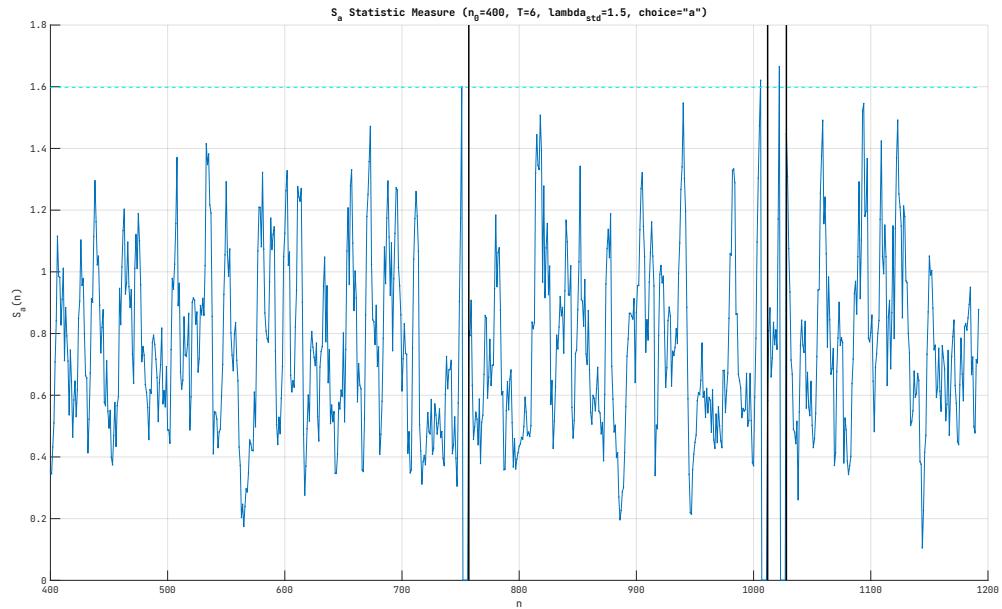
Σχήμα 4.8: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_a(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «b»

Βλέποντας τα τελευταία δύο (2) διαγράμματα και συγκρίνοντάς τα με αυτά της επιλογής «c» κάνουμε τις εξής παρατηρήσεις:

1. Τα σημεία αλλαγής είναι ίδια σε αριθμό (6) και μάλιστα στις ακριβώς ίδιες θέσεις (με εξαίρεση το τελευταίο που έχει μετατοπιστεί 2 χρονικές στιγμές μετά)
2. Το όριο απόφασης (cyan διακεκομμένη γραμμή) φαίνεται να έχει πολύ μεγαλύτερη διακύμανση σε σύγκριση με αυτό της επιλογής «c», κάτι απολύτως αναμενόμενο αφού η αναπροσαρμογή γίνεται πλέον σε κάθε βήμα / χρονική στιγμή και άρα αντίστοιχα η δειγματική τυπική απόκλιση των παρατηρήσεων του training set αλλάζει συνεχώς

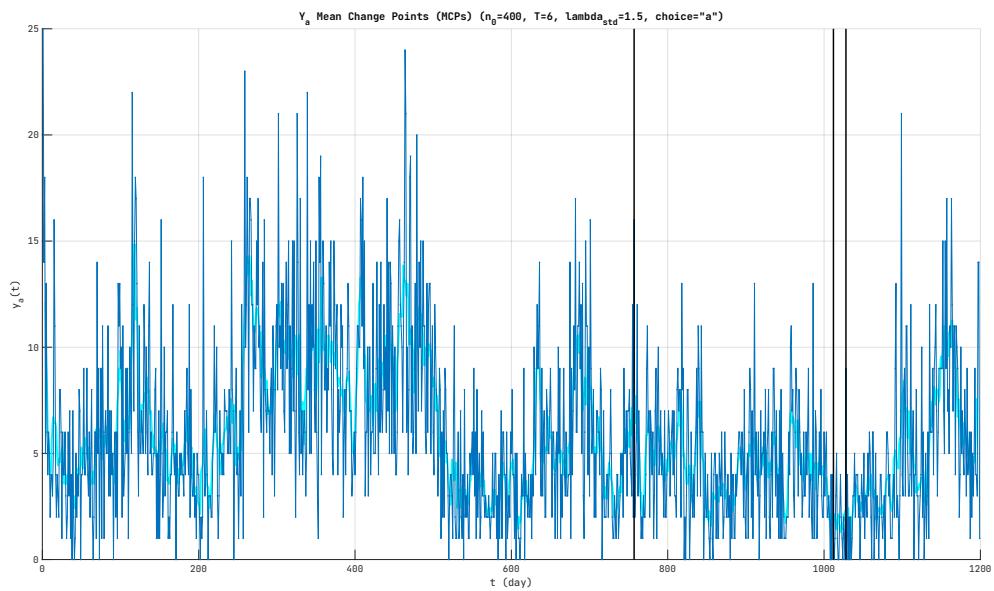
#### 4.4.3 Χωρίς Αναπροσαρμογή

Τέλος, τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «a» (καμία αναπροσαρμογή - διατήρηση του μοντέλου που προσαρμόστηκε στις πρώτες 400 παρατηρήσεις της στάσιμης χρονοσειράς), ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 4.9: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με  $MA(1)$  της στάσιμης χρονοσειράς  $\{X_a(t)\}$  και για επιλογή αναπτροσαρμογής « $a$ » (χωρίς αναπτροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης,  $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.789, 27.2sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:



Σχήμα 4.10: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_a(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$  smoothing) - επιλογή « $a$ »

Βλέποντας τα τελευταία δύο (2) διαγράμματα και συγκρίνοντάς τα με αυτά της επιλογής «c» κάνουμε τις εξής παρατηρήσεις:

1. Τα σημεία αλλαγής είναι λιγότερα, πλέον μόλις τρία (3) αλλά αυτά τα τρία βρίσκονται στις ακριβώς ίδιες θέσεις με τα αντίστοιχα. Αυτό οφείλεται στο ότι το όριο επιλογής παραμένει σταθερό παρόλο που η τυπική απόκλιση των δειγμάτων αλλάζει καθώς προχωράμε στη στάση χρονοσειρά πράγμα που φαίνεται και από το αντίστοιχο διάγραμμα της επιλογής «b»
2. Το όριο απόφασης (cyan διακεκομένη γραμμή), όπως αναφέρθηκε, παραμένει συνεχώς σταθερό, κάτι πιθανότατα μη επιθυμητό

Καταληκτικά, η πιο κατάλληλη επιλογή για αναπροσαρμογή του μοντέλου πρόβλεψης είναι η επιλογή «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής) καθώς συνδυάζει καλή υπολογιστική συμπεριφορά αλλά ταυτόχρονα εξασφαλίζει και ότι το μοντέλο και άρα και το όριο επιλογής σημείων αλλαγής μεταβάλλεται καθώς προχωράμε στη χρονοσειρά και βρίσκουμε σημεία αλλαγής. Πρέπει να σημειωθεί επίσης ότι αυτή η επιλογή χρησιμοποιήθηκε και κατά το grid search των hyperparameters (σχήματα 4.3 και 4.4 παραπάνω).

## Βήμα 5

# Εντοπισμός Σημείων Αλλαγής της Χρονοσειράς B

Παιρνώντας στο βήμα 5, θα επαναλάβουμε τη διαδικασία και ανάλυση του βήματος 4 στη χρονοσειρά προβολών του βίντεο B, για την οποία επίσης θα γίνει προσπάθεια ανίχνευσης σημείων σημαντικών αλλαγών με αυτόματο τρόπο.

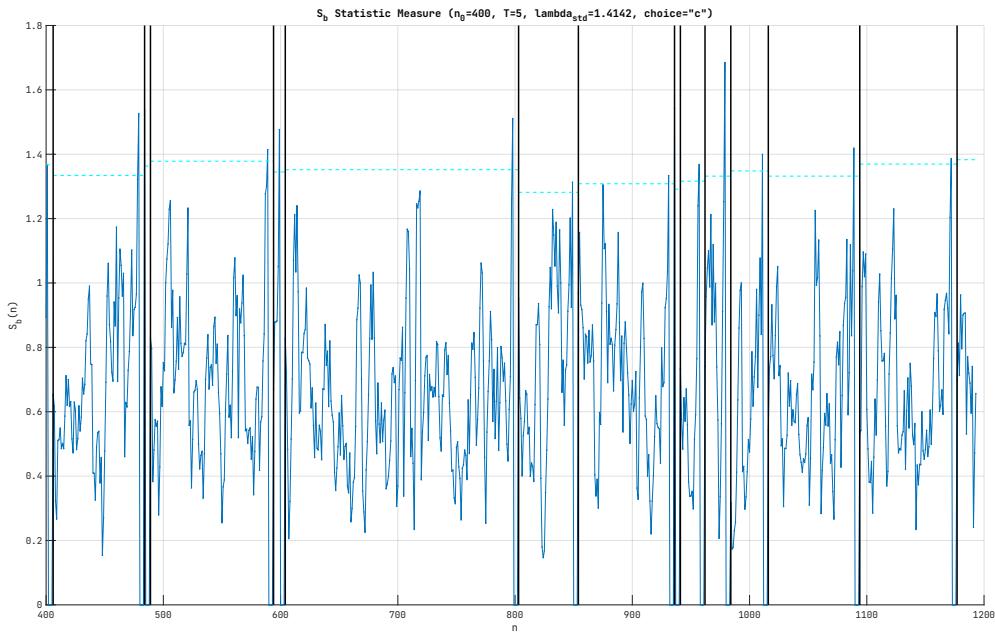
Η μέθοδος και το κριτήριο επιλογής των σημείων αλλαγής παραμένουν τα ίδια με το προηγούμενο βήμα όπως επίσης και οι επιλογές αναπροσαρμογής του μοντέλου κατά τη διάρκεια εφαρμογής της μεθόδου. Επίσης και εδώ έχουν υλοποιηθεί και οι τρεις επιλογές.

### 5.1 Αρχική Εφαρμογή

Αρχικά, θα δώσουμε κάποιες τιμές στις hyperparameters της μεθόδου, δηλαδή στο μέγεθος εκπαίδευσης  $n_0$ , στον ορίζοντα πρόβλεψης,  $T$ , στην επιλογή αναπροσαρμογής του μοντέλου και στην παράμετρο  $\hat{M}_{std}$ , και θα τρέξουμε την παραπάνω μέθοδο στη στάσιμη χρονοσειρά που προέκυψε από το βήμα 3,  $\{X_{b_{deseasoned}}(t)\}$ , και στο ARMA(4, 4) μοντέλο που προσαρμόστηκε σε αυτή (σχέση 3.11). Έτσι, για τις παραμέτρους θα έχουμε:

- $n_0 = 400$ : Όπως προτείνεται στην εκφώνηση, το μέγεθος εκπαίδευσης θα είναι 400 παρατηρήσεις
- $T = 5$ : Επίσης δίνεται έμμεσα στην εκφώνηση, το κριτήριο αρχικά θα υπολίζεται για προβλέψεις έως και 5 βήματα εμπρός
- $choice = "c"$ : Διατήρηση του μοντέλου έως ότου βρεθεί σημείο αλλαγής και επαναπροσαρμογή του στις 400 πιο πρόσφατες παρατηρήσεις από το  $n + 5$  όταν βρεθεί
- $\hat{M}_{std} = \sqrt{2}$ : Όχι σε συμφωνία με την προτεινόμενη τιμή, καθώς αυτή δίνει μόλις ένα σημείο αλλαγής για τις υπόλοιπες επιλεγμένες παραμέτρους, αλλά ως μία τυχαία αρχική επιλογή. Άρα, κριτήριο επλογής MCPs όταν  $S_n > \sqrt{2} * s_x$

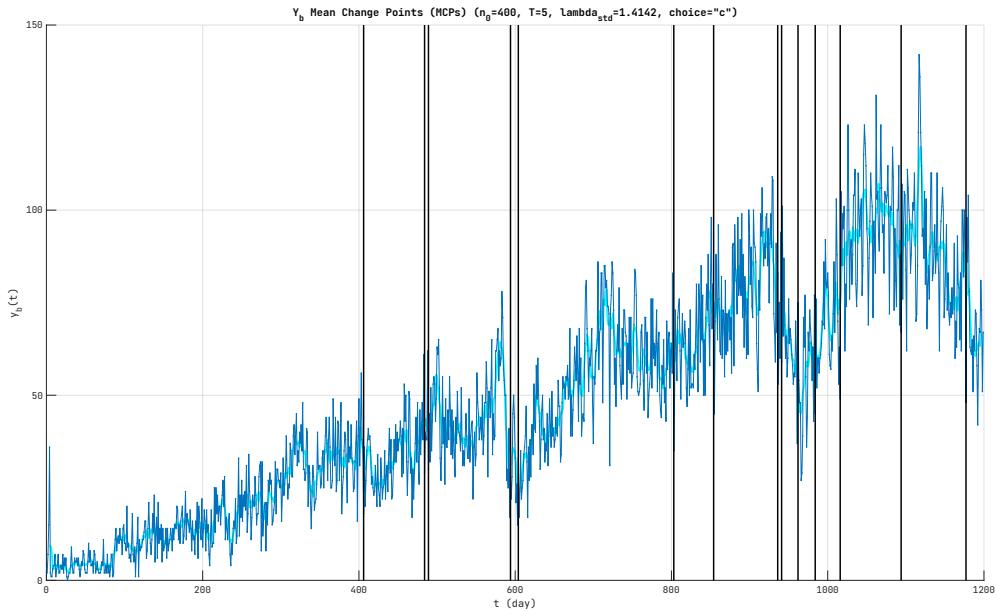
Παρακάτω, φαίνονται τα σημεία αλλαγής που προκύπτουν από την εκτέλεση της μεθόδους με τις παραπάνω παραμέτρους:



Σχήμα 5.1: Τιμές στατιστικού  $S_n$  για έως και 5 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$ . Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο). Το  $n$  ξεκινάει από το 400 και άρα το  $S_n$  δεν ορίζεται για τιμές  $n < n_0 = 400$ .

Όπως φαίνεται στο παραπάνω διάγραμμα, επειδή έχει χρησιμοποιηθεί η επιλογή «c» για την αναπροσαρμογή του μοντέλου, κάθε φορά που εντοπίζεται σημείο αλλαγής το μοντέλο επανεκτιμάται σε νέο training set (που είναι οι τελευταίες 400 παρατηρήσεις από τη στιγμή  $n + 5$ ) και άρα αλλάζει η δειγματική τυπική απόκλιση των παρατηρήσεων του training set. Αυτό φαίνεται ως αλλαγή στο επίπεδο της cyan διακεκομένης γραμμής μεταξύ διαδοχικών σημειών αλλαγής.

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο B και κατόπιν ακολουθεί σύντομος σχολιασμός:



Σχήμα 5.2: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing)

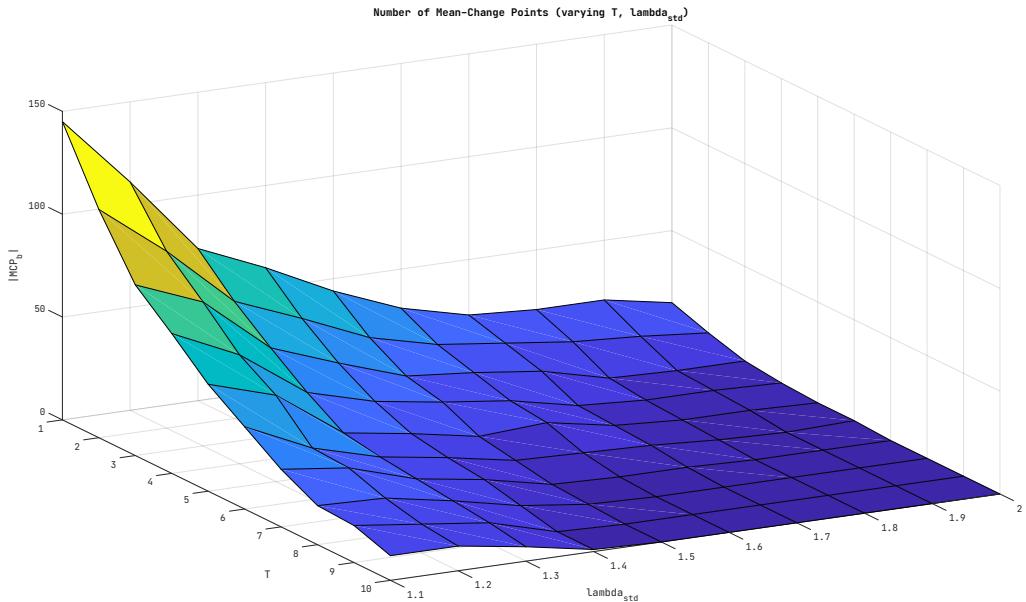
Βλέποντας το πού «πέφτουν» τα 14 σημεία αλλαγής στην αρχική χρονοσειρά προβολών του βιντέο Β παρατηρούμε πως αυτά δεν είναι σε τυχαίες θέσεις. Βρίσκοντα και εδώ (όπως και στην ανάλυση της πρώτης χρονοσειράς) είτε σε κορυφές, δηλαδή σε θέσεις που αριστερά υπάρχει τοπικά αυξητική τάση και δεξιά τοπικά πτωτική (ακόμα και εάν αυτό γίνεται για λίγες παρατηρήσεις), ή σε βυθούς, δηλαδή σε σημεία που αριστερά υπάρχει τοπικά πτωτική τάση ενώ δεξιά αυτή αλλάζει και γίνεται αυξητική. Είναι κάπως λογικό οι προβλέψεις μας για τις γειτονιές τέτοιων σημείων (ακόμα και εάν γίνονται μέσω στάσιμων εκδοχών των αρχικών χρονοσειρών) με γραμμικά μοντέλα να εμφανίζουν αρκετά σημαντικά σφάλματα ώστε να σηματοδοτηθούν σημεία αλλαγής.

Ακολουθεί μια πιο ενδελεχής ανάλυση για την επιλογή των hyperparameters που σε αυτήν την υποενότητα έγινε κάπως αυθαίρετα.

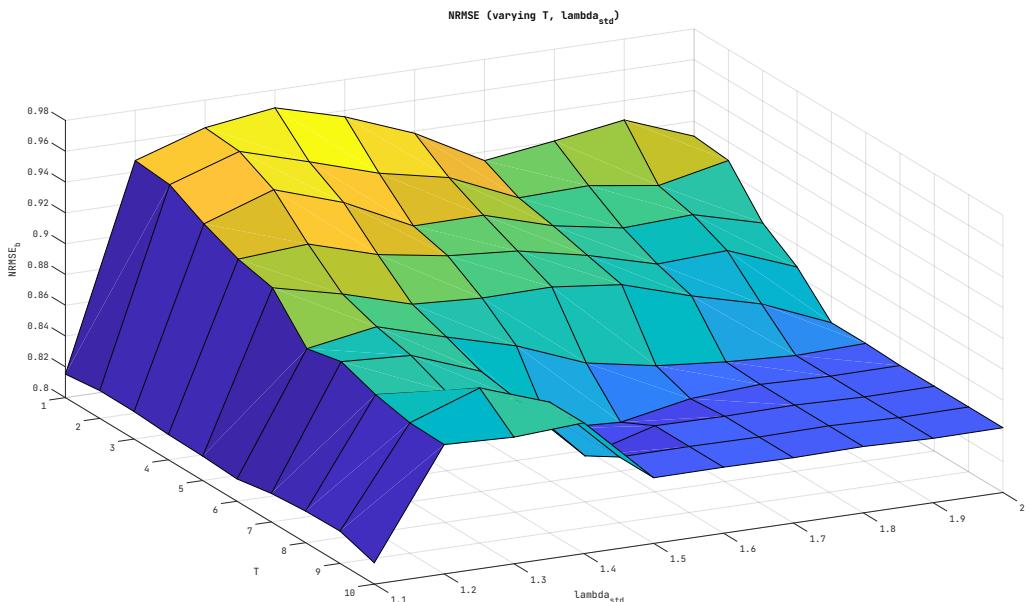
## 5.2 Επιλογή Βέλτιστων Hyperparameters

Για επιλογή βέλτιστων τιμών στις hyperparameters της μεθόδου και συγκεκριμένα στον ορίζοντα πρόβλεψης,  $T$ , και στο  $\lambda_{std}$  του ορίου απόφασης,  $a$ , θα κάνουμε αναζήτηση πλέγματος ως προς αυτές. Ως μετρικές για αξιολόγηση του κάθε συνδυασμού των παραμέτρων αυτών χρησιμοποιήθηκαν οι ίδιες όπως και στη ανάλυση πρώτης χρονοσειράς, δηλαδή ο Αριθμός MCPs και το NRMSE Προβλέψεων.

Αρχικά παραθέτονται σε διαγράμματα τύπου surf τα αποτελέσματα αναζήτησης πλέγματος ως προς τις παραπάνω μετρικές ενώ στη συνέχεια σχολιάζεται ο τρόπος επιλογής προσεγγεστικά βέλτιστων παραμάτρων αλλά και οι τελικές τους τιμές.



Σχήμα 5.3: Αριθμός σημείων αλλαγής,  $|MCP|$ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $\alpha$



Σχήμα 5.4: NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs), για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $\alpha$

Γενικότερα, αναζητούμε τα «γόνατα» στις αντίστοιχες τρισδιάστατες καμπύλες έτσι ώστε περαιτέρω μεταβολές των αντίστοιχων παραμέτρων να μην είναι πλέον επικερδείς.

Επικεντρώνοντας στο πρώτο διάγραμμα και δεδομένου ότι θέλουμε ο αριθμός των MCPs να μήν είναι πολύ μεγάλος ή πολύ μικρός, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστι-

κά):

$$T \in [5, 8] \quad \& \quad \lambda_{std} \in [1.2, 1.5] \quad (5.1)$$

Παρατηρούμε δηλαδή σημαντική μεταβολή στο εύρος της παραμέτρου  $\lambda_{std}$  στη χρονοσειρά B σε σύγκριση με το αντίστοιχο διάγραμμα της A, το οποίο προκύπτει από την απότομη πτώση του διαγράμματος παραπάνω το οποίο με τη σειρά του πιθανότητα οφείλεται στην ύπαρξη μεγάλύτερης τυπικής αποκλίσης στη στάσιμη χρονοσειρά  $X_{b_{deseasoned}}$ .

Επικεντρώνοντας τώρα στο διάγραμμα των NRMSEs και δεδομένου ότι θέλουμε το NRMSE να είναι κατά το δυνατό μικρό, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \geq 5 \quad \& \quad \lambda_{std} \geq 1.5 \quad (5.2)$$

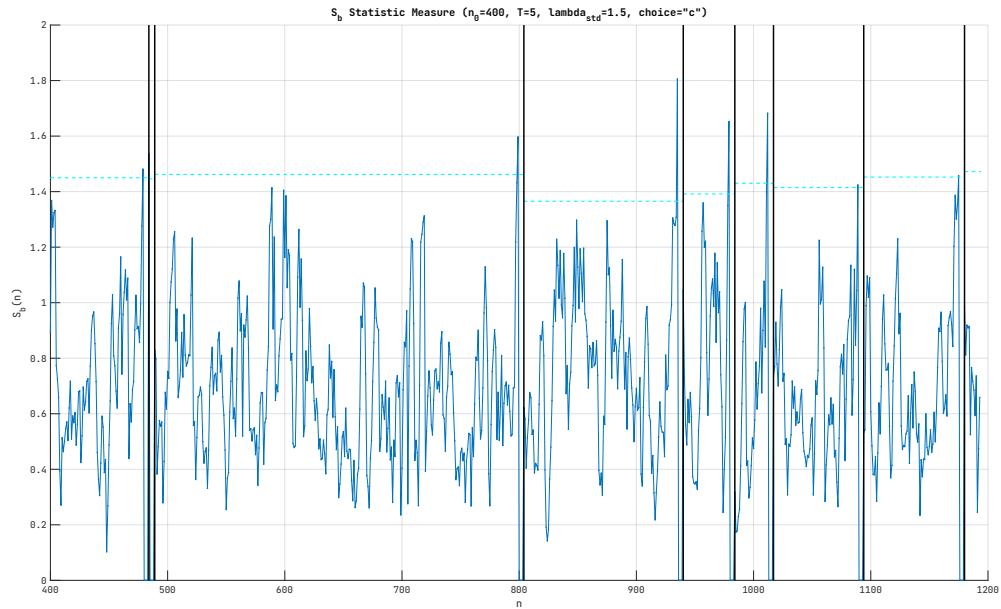
Συνδυάζοντας τις σχέσεις (5.1) και (5.2) παραπάνω καταλήγουμε ότι οι hyperparameters που θα χρησιμοποιηθούν για την εφαρμογή της μεθόδου αυτόματης εύρεσης χρονικών σημείων αλλαγής θα είναι: **T = 5 βήματα** και  **$\lambda_{std} = 1.5$** . Οι αντίστοιχες τιμές του grid search είναι: **|MCP| = 8 MCPs** και **NRMSE = 0.8921**.

### 5.3 Εφαρμογή με Βέλτιστες Παραμέτρους

Χρησιμοποιώντας τις επιλεγμένες τιμές για τις hyperparameters της μεθόδου, δηλαδή ορίζοντα πρόβλεψης έως και 5 βημάτων εμπρός,  $T = 5$ , και παράμετρο ορίου απόφασης στο 1.5,  $\lambda_{std} = 1.5$ , θα τρέξουμε την παραπάνω μέθοδο στη στάσιμη χρονοσειρά που προέκυψε από το βήμα 3,  $\{X_{b_{deseasoned}}(t)\}$ , και στο ARMA(4, 4) μοντέλο που προσαρμόστηκε σε αυτή (σχέση 3.11). Παρακάτω, φαίνονται τα σημεία αλλαγής που προκύπτουν από την εκτέλεση της μεθόδους με τις παραπάνω παραμέτρους για κάθε μια από τις επιλογές αναπροσαρμογής του μοντέλου («a», «b» ή «c»).

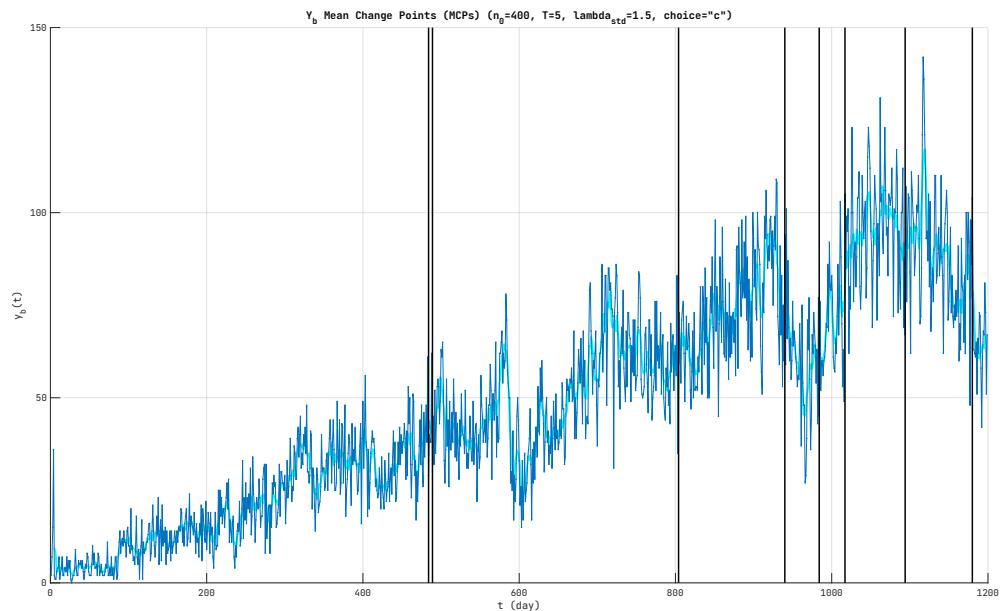
#### 5.3.1 Αναπροσαρμογή όταν βρεθεί σημείο αλλαγής

Επιλογή «c»



Σχήμα 5.5: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.892, 76.2sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο B:

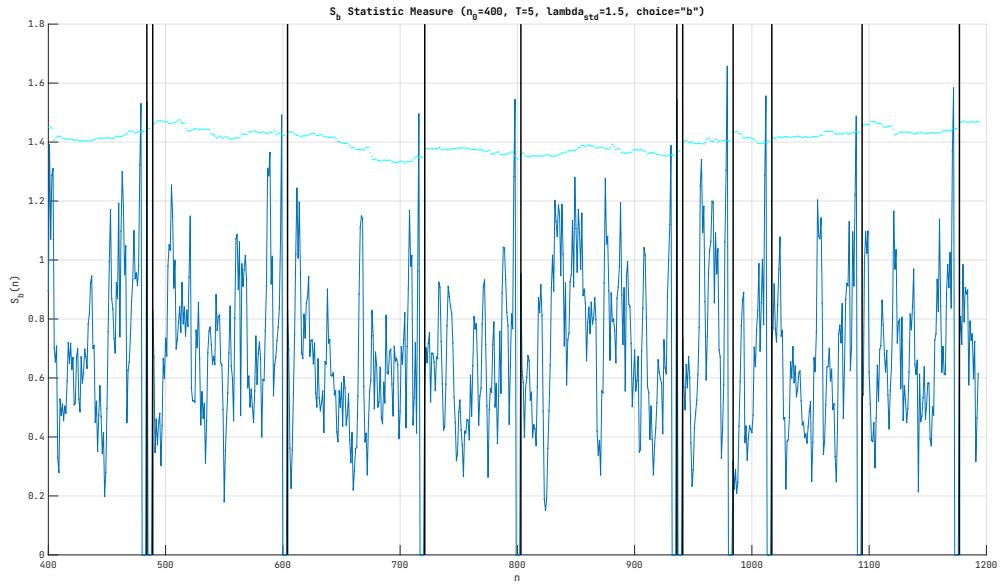


Σχήμα 5.6: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «c»

### 5.3.2 Αναπροσαρμογή σε κάθε χρονική στιγμή

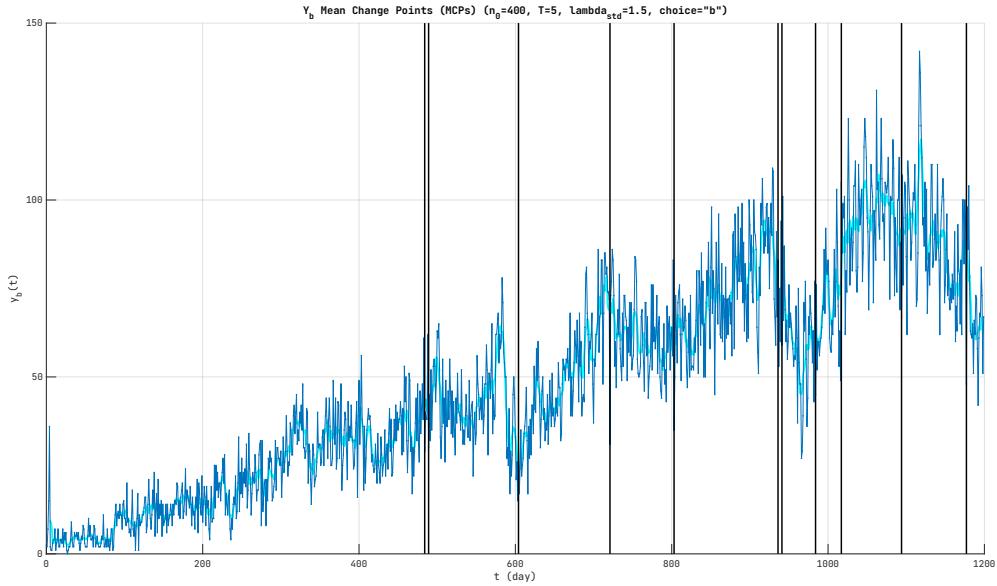
Επιλογή «b»

Τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή) ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 5.7: Τιμές στατιστικού  $S_n$  για έως και 5 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.893, 244.1sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο B:



Σχήμα 5.8: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 ( $MA(7)$  smoothing) - επιλογή «b»

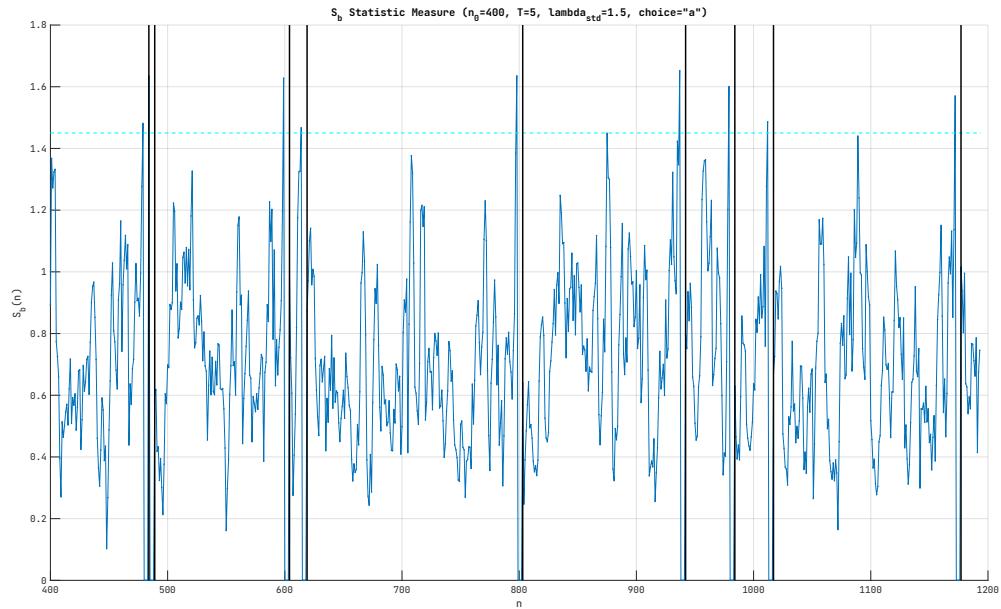
Βλέποντας τα τελευταία δύο (2) διαγράμματα και συγκρίνοντάς τα με αυτά της επιλογής «c» κάνουμε τις εξής παρατηρήσεις:

1. Τα σημεία αλλαγής είναι περισσότερα σε αριθμό (11 πλέον vs. 6 με την επιλογή «c») έχοντας σημεία στις ίδιες θέσεις αλλά προσθέτοντας και επιπλέον σημεία αλλαγής. Αυτό είναι αποτέλεσμα αφενός της κυμάτωσης που εμφανίζεται πλέον στο όριο απόφασης αλλά κυρίως στο ότι πλέον τα σφάλματα  $S_n$  εμφανίζουν περισσότερες «κορυφές».
2. Το όριο απόφασης (cyan διακεκομμένη γραμμή) φαίνεται να έχει πολύ μεγαλύτερη διακύμανση σε σύγκριση με αυτό της επιλογής «c», κάτι απολύτως αναμενόμενο αφού η αναπροσαρμογή γίνεται πλέον σε κάθε βήμα / χρονική στιγμή και άρα αντίστοιχα η δειγματική τυπική απόκλιση των παρατηρήσεων του training set αλλάζει συνεχώς

### 5.3.3 Χωρίς Αναπροσαρμογή

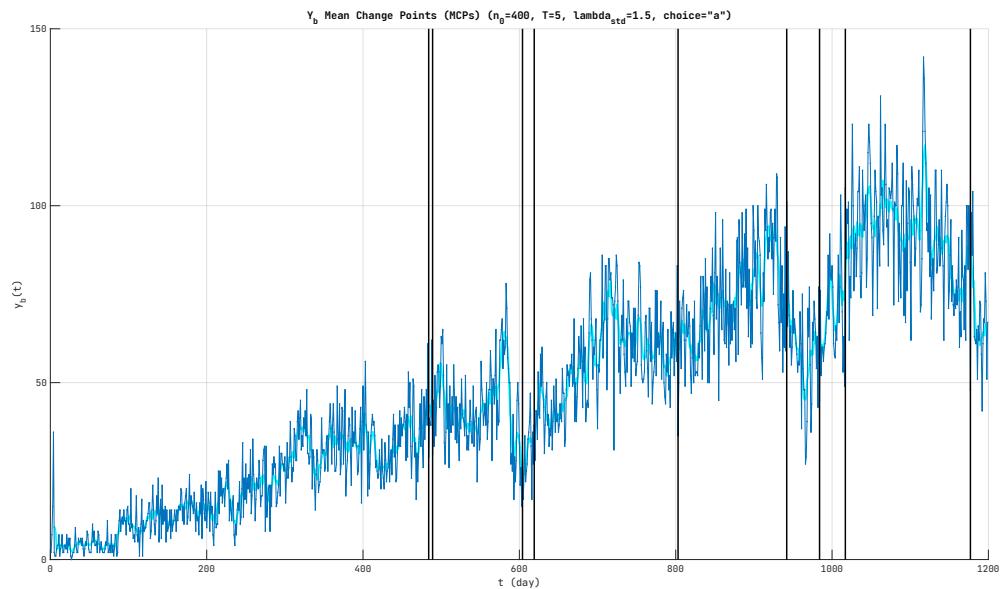
*Επιλογή «a»*

Τέλος, τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «a» (καμία αναπροσαρμογή - διατήρηση του μοντέλου που προσαρμόστηκε στις πρώτες 400 παρατηρήσεις της στάσιμης χρονοσειράς), ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 5.9: Τιμές στατιστικού  $S_n$  για έως και 5 βήματα μπροστά πρόβλεψη με ARMA(4, 4) της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  και για επιλογή αναπροσαρμογής «α» (χωρίς αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.899, 61.1sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο B:



Σχήμα 5.10: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από την ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «α»

Βλέποντας τα τελευταία δύο (2) διαγράμματα και συγκρίνοντάς τα με αυτά της επιλογής «c» κάνουμε τις εξής παρατηρήσεις:

1. Τα σημεία αλλαγής είναι και εδώ περισσότερα κατά 1 (9 vs. 8) αλλά και σε διαφορετικές θέσεις τα 4 από τα 9. Παρόλα αυτά, τα σημεία της επιλογής «b» οδηγούν σε χαμηλότερο NPMSE.
2. Το όριο απόφασης (cyan διακεκομμένη γραμμή), όπως αναφέρθηκε, παραμένει συνεχώς σταθερό, κάτι πιθανότατα μη επιθυμητό

Καταληκτικά, η πιο κατάλληλη επιλογή για αναπροσαρμογή του μοντέλου πρόβλεψης είναι η επιλογή «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής) καθώς συνδυάζει καλή υπολογιστική συμπεριφορά αλλά ταυτόχρονα επιτυγχάνει και το μικρότερο NRMSE των προβλέψεων κατά την εκτέλεση της μεθόδου. Πρέπει να σημειωθεί επίσης ότι αυτή η επιλογή χρησιμοποιήθηκε και κατά το grid search των hyperparameters (σχήματα 5.3 και 5.4 παραπάνω).

### 5.3.4 Συμπερασματικά Σχόλια

Συμπερασματικά και επιγραμματικά, είναι σκόπιμο να αναφερθεί ότι η προσέγγιση εξαγωγής σημείων αλλαγής με την μέθοδο που αναλύθηκε στα βήματα 4-5 παραπάνω, **Θα μπορούσε να χρησιμοποιηθεί για αυτόματο εντοπισμό αυτών των σημείων**. Συγκεκριμένα, με την υπόθεση ότι το  $T$  (ορίζοντας πρόβλεψης) και το  $\bar{M}_{std}$  έχουν κατάλληλα επιλεγεί ενώ το μοντέλο αναπροσαρμόζεται κάθε φορά που βρίσκεται ένα σημείο αλλαγής, τότε αυτή η μέθοδος φαίνεται να επιτυγχάνει την εύρεση σημείων αλλαγής τα οποία - τουλάχιστον οπτικά - φαίνονται να βρίσκονται σε περιοχές που όντως η ζήτηση παρουσιάζει βραχεία έξαρση ή βύθιση.

# Βήμα 6

## Εντοπισμός Σημείων Αλλαγής με Μη-Γραμμικά Μοντέλα

### 6.1 Ανακατασκευή του χώρου καταστάσεων

Για την ανακατασκευή του χώρου καταστάσεων (όπως περιγράφεται στις σημειώσεις του κ. Κουγιουμτζή - σελίδες 90-92) θα χρησιμοποιήσουμε την μέθοδο των υστερήσεων. Επομένως χρειαζόμαστε σημεία  $\mathbf{x}_i \in \mathbb{R}^m$  τα οποία προκύπτουν από τις (μονοδιάστατες) παρατηρήσεις της εκάστοτε χρονοσειράς με βάση τη ακόλουθη σχέση (σχέση 97 σελ. 91):

$$\mathbf{x}_i = [x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau}] \quad (6.1)$$

Το πρώτο πράγμα που απαιτείται να γίνει πριν την προσαρμογή οποιουδήποτε μη-γραμμικού μοντέλου σε χρονοσειρά που βασίζεται στην ύπαρξη συνάρτησης παρατήρησης της τροχίας στο χώρο καταστάσεων, είναι η εκτίμηση των παραμέτρων ανακατασκευής, ως εξής:

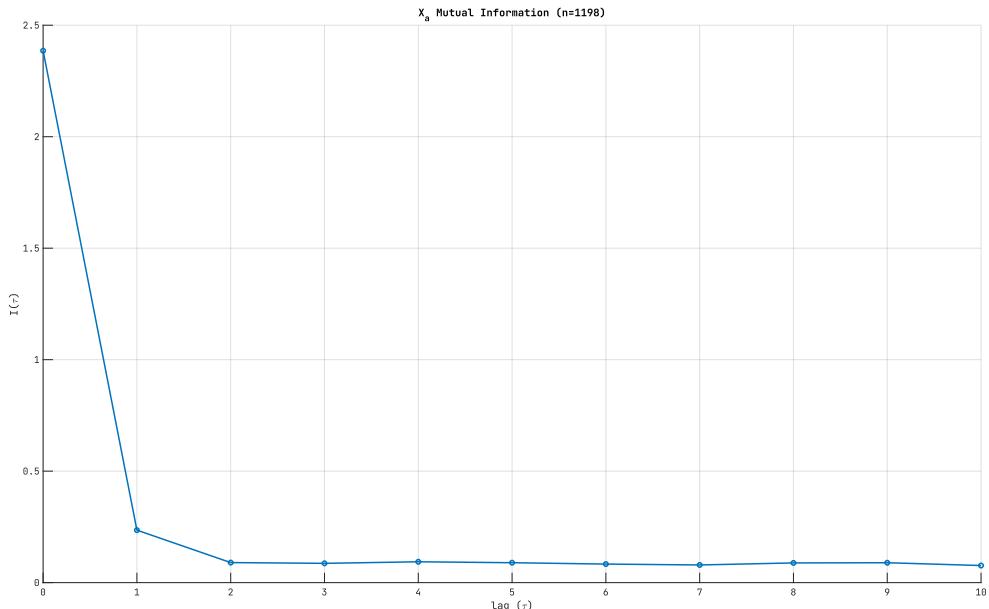
- Διάστασης Εμβύθισης Ελκυστή,  $m$ : Θα οριστεί με βάση τη μέθοδο των ψευδών κοντινότερων γειτόνων (false nearest neighbors - FNN), όπως προτείνεται στη σελ. 92 των σημειώσεων
- Υστέρηση,  $\tau$ : Θα οριστεί από τη συνάρτηση αμοιβαίας πληροφορίας (mutual information -  $I(\tau) = I(x_i, x_{i-\tau})$ ), όπως προτείνεται στις σελ. 91-92 των σημειώσεων

Στο σημείο αυτό, κρίνεται σκόπιμο να τονιστεί ότι η όλη ανάλυση γίνεται για τις στάσιμες χρονοσειρές που προέκυψαν από την ανάλυση των βημάτων 2 και 3. Δηλαδή, αντί για την αρχική χρονοσειρά των προβολών του βίντεο A,  $\{Y_a(t)\}$ , θα ασχοληθούμε με τη στάσιμη εκδοχή της,  $\{X_a(t)\}$ , σχέση (1.3). Εντελώς παρόμοια, αντί για την αρχική χρονοσειρά των προβολών του βίντεο B,  $\{Y_b(t)\}$ , θα ασχοληθούμε με τη στάσιμη εκδοχή της,  $\{X_{b_deseasoned}(t)\}$ , σχέση (3.6).

#### 6.1.1 Παράμετροι ανακατασκευής για τη Χρονοσειρά A

##### 6.1.1.1 Υστέρηση, $\tau$

Για την εύρεση της υστέρησης,  $\tau$ , θα χρησιμοποιήσουμε την συνάρτηση αμοιβαίας πληροφορίας και θα αναζητήσουμε το πρώτο τοπικό της ελάχιστο (ως προς την υστέρηση). Ακολούθως, δίνεται το διάγραμμα της αμοιβαίας πληροφορίας,  $I_a(\tau)$ , και παραθέτεται η επιλεγμένη τιμή για την υστέρηση.

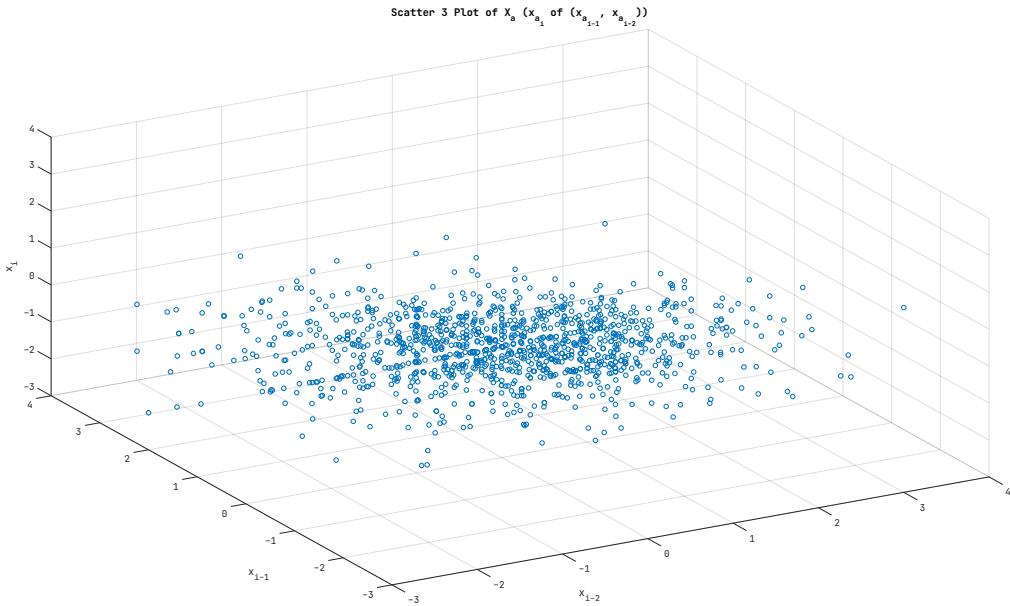


Σχήμα 6.1: Διάγραμμα αμοιβαίας πληροφορίας της στάσιμης χρονοσειράς  $\{X_a(t)\}$  για επιλογή της υστέρησης  $\tau$

Βλέποντας λοιπόν ότι το πρώτο τοπικό ελάχιστο της  $I(\tau)$  είναι για  $\tau = 3$  (θέλει αρκετό ζουμ - στη πράξη χρησιμοποιήθηκε η συνάρτηση `islocalmin()` του MATLAB), θα θεωρήσουμε για τη στάσιμη χρονοσειρά Α ότι **τ=3**.

#### 6.1.1.2 Διάσταση Εμβύθινσης Ελκυστή, m

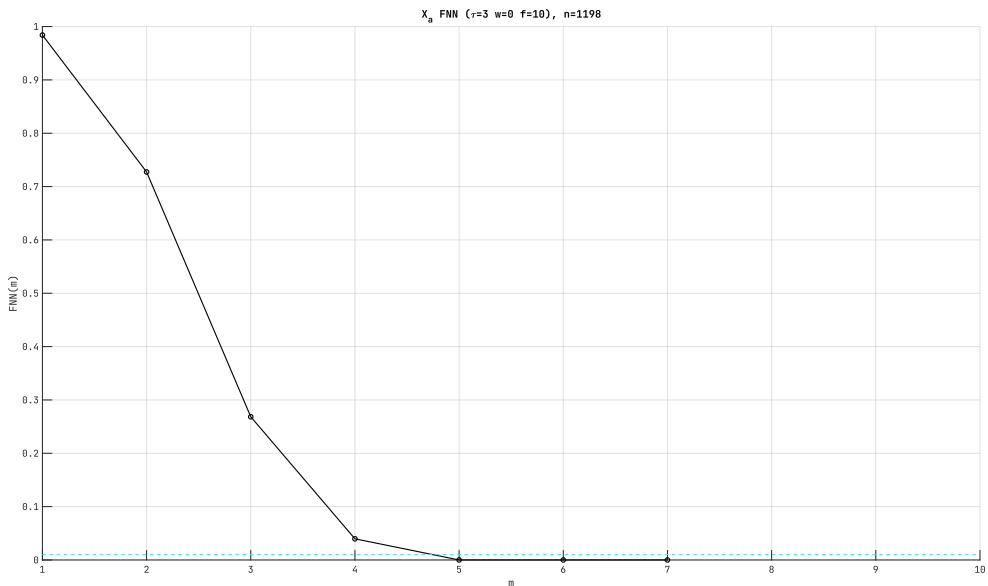
Πριν αναζητήσουμε τη διάσταση εμβύθινσης με τη μέθοδο False Nearest Neighbors θα δούμε το διάγραμμα διασποράς (scatter plot) ώστε να διαπιστώσουμε εάν πρόκειται για χαμηλοδιάστατο ελκυστή.



Σχήμα 6.2: Διάγραμμα διασποράς  $x_i - x_{i-1} - x_{i-2}$  της στάσιμης χρονοσειράς  $\{X_a(t)\}$

Από το παραπάνω διάγραμμα (αν και φαίνεται κάποιο υποσύνολο των σημείων του διαγράμματος να έχει κάποια δομή - μια ευθεία γραμμή που ξεκινάει από το  $x_{i-1} = 3, x_i = -3$ ), δεν φαίνεται να υπάρχει κάποια δομή που να ακολουθούν όλα τα σημεία του scatter plot και επομένως υποψιαζόμαστε ότι ο ελκυστής (εφόσον υπάρχει) θα βρίσκεται σε μεγαλύτερες διαστάσεις (π.χ. 4-D, 5-D κλπ).

Ακολούθως, τρέχουμε τη δοσμένη συνάρτηση `falsenearest()` με παράμετρο `tau=3` και από το διάγραμμα της μεταβολής των FNNs θα καταλήξουμε στην κατάλληλη τιμή του  $m$ :



Σχήμα 6.3: Διάγραμμα μεταβολής των FNNs ως προς τη διάσταση εμβύθινσης,  $m$ , για τη στάσιμη χρονοσειρά  $\{X_a(t)\}$

Στο παραπάνω διάγραμμα έχει σημειωθεί και το όριο επιλογής 1%. Επιλέγουμε λοιπόν για διάσταση εμβύθινσης,  $m=5$  για την στάσιμη χρονοσειρά A.

Σημειώνεται, πως για να τρέξει η παραπάνω συνάρτηση για όλες τις εκτελέσεις έπρεπε να μεγαλώσει η ακτίνα αναζήτησης ψευδών γειτόνων, από 0.1 που ήταν η προεπιλογή σε  $fthres=0.3$ .

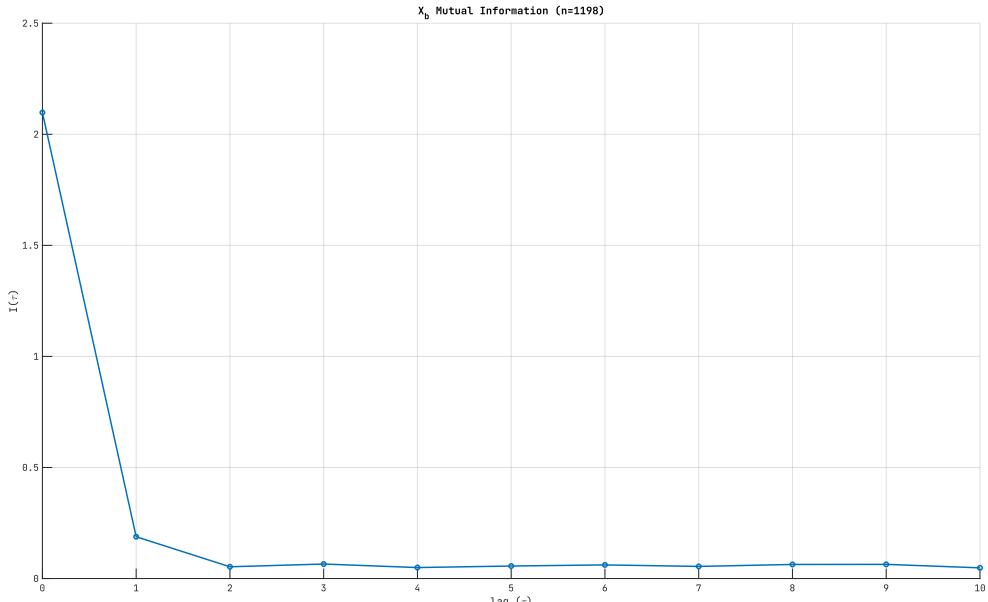
Επομένως, τα σημεία του ανακατασκευασμένου χώρου καταστάσεων θα είναι τα εξής:

$$\mathbf{x}_{a_i} = [x_{a_i}, x_{a_{i-3}}, x_{a_{i-6}}, x_{a_{i-9}}, x_{i-12}] \in \mathbb{R}^5, \quad i = 13, \dots, 1998 \quad (6.2)$$

## 6.1.2 Παράμετροι ανακατασκευής για τη Χρονοσειρά B

### 6.1.2.1 Υστέρηση, $\tau$

Αντίστοιχα με τη χρονοσειρά A, για την εύρεση της υστέρησης στη δεύτερη (στάσιμη) χρονοσειρά, δίνεται το διάγραμμα της αμοιβαίας πληροφορίας,  $I_b(\tau)$ , και παραθέτεται η επιλεγμένη τιμή για την υστέρηση.

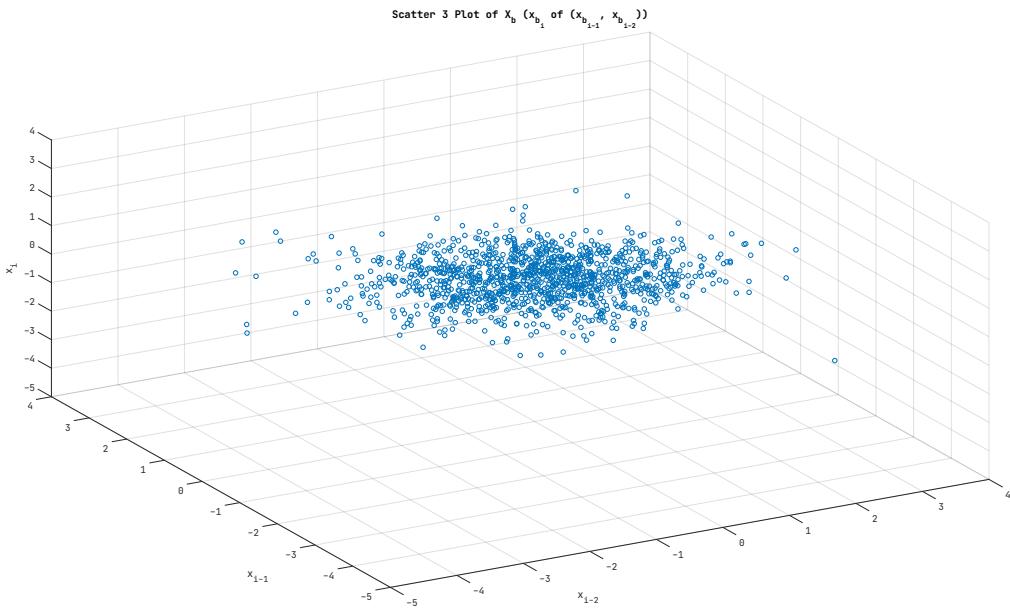


Σχήμα 6.4: Διάγραμμα αμοιβαίας πληροφορίας της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  για επιλογή της υστέρησης  $\tau$

Βλέποντας στο παραπάνω διάγραμμα ότι το πρώτο τοπικό ελάχιστο της  $I(\tau)$  είναι για  $\tau = 2$  θα θεωρήσουμε για τη στάσιμη χρονοσειρά B ότι  $\tau=2$ .

### 6.1.2.2 Διάσταση Εμβύθινσης Ελκυστή, $m$

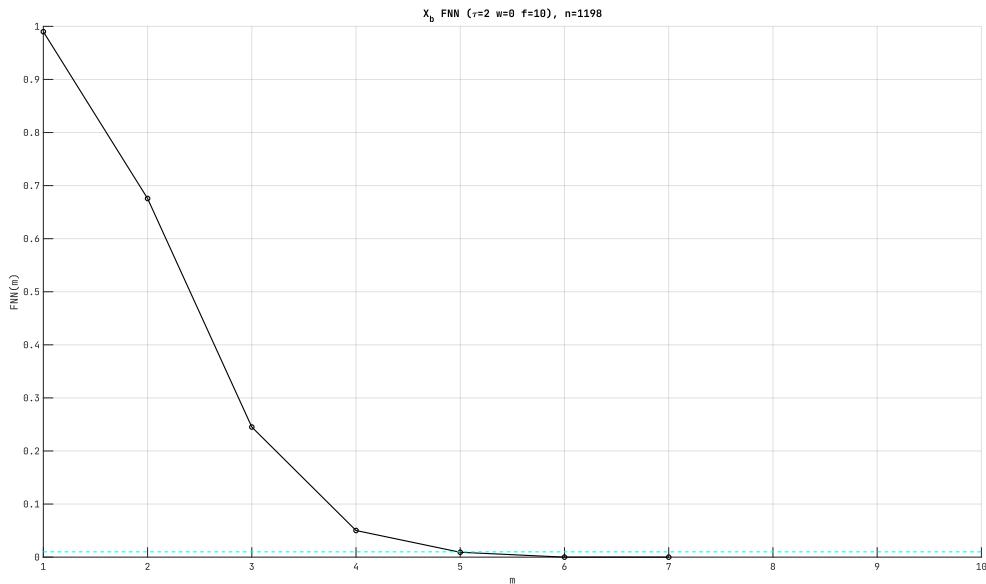
Πριν αναζητήσουμε τη διάσταση εμβύθινσης με τη μέθοδο False Nearest Neighbors θα δούμε το διάγραμμα διασποράς (scatter plot) ώστε να διαπιστώσουμε εάν πρόκειται για χαμηλοδιάστατο ελκυστή.



Σχήμα 6.5: Διάγραμμα διασποράς  $x_i - x_{i-1} - x_{i-2}$  της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$

Από το παραπάνω διάγραμμα δεν φαίνεται να υπάρχει καμία δομή που να ακολουθούν όλα τα σημεία του scatter plot και επομένως υποψιαζόμαστε ότι ο ελκυστής (εφόσον υπάρχει) θα βρίσκεται σε μεγαλύτερες διαστάσεις (π.χ. 4-D, 5-D κλπ). Τα σημεία του scatter plot φαίνονται συγκεντρωμένα σε μία άμορφη περιοχή του τρισδιάστατου χώρου.

Ακολούθως, τρέχουμε τη δοσμένη συνάρτηση `falsenearest()` με παράμετρο `tau=2` και από το διάγραμμα της μεταβολής των FNNs θα καταλήξουμε στην κατάλληλη τιμή του  $m$ :



Σχήμα 6.6: Διάγραμμα μεταβολής των FNNs ως προς τη διάσταση εμβύθινσης,  $m$ , για τη στάσιμη χρονοσειρά  $\{X_{b_{deseasoned}}(t)\}$

Η πρώτη τιμή που πέφτει κάτω από σημειωμένο όριο του 1% είναι και εδώ το 5. Επιλέγου-

με λοιπόν για διάσταση εμβύθινσης,  $m=5$  για την στάσιμη χρονοσειρά A.

Επομένως, τα σημεία του ανακατασκευασμένου χώρου καταστάσεων θα είναι τα εξής:

$$\mathbf{x}_{b_i} = [x_{b_i}, x_{b_{i-2}}, x_{b_{i-4}}, x_{b_{i-6}}, x_{i-8}] \in \mathbb{R}^5, \quad i = 9, \dots, 1998 \quad (6.3)$$

## 6.2 Προσαρμογή Τοπικού Μοντέλου Κοντινότερων Γειτόνων

Για τη πρόβλεψη με μη-γραμμικά μοντέλα θα χρησιμοποιηθεί τοπική πρόβλεψη μέσου όρου, όπου για ένα βήμα εμπρός προβλέψη ορίζεται ως εξής (σχέση 107 - σελ. 100):

$$x_i(1) = \frac{1}{K} \sum_{j=1}^K \mathbf{x}_{i(j)+1} \quad (6.4)$$

όπου  $x_i(1)$  είναι η πρόβλεψη του μοντέλου για το  $(i+1)$ -οστή παρατήρηση της στάσιμης χρονοσειράς,  $x_{i+1}$ , ενώ  $\mathbf{x}_{i(k)}$  είναι ο  $k$ -οστός κοντινότερος γείτονας του σημείου  $\mathbf{x}_i$  στον ανακατασκευασμένο χώρο καταστάσεων.

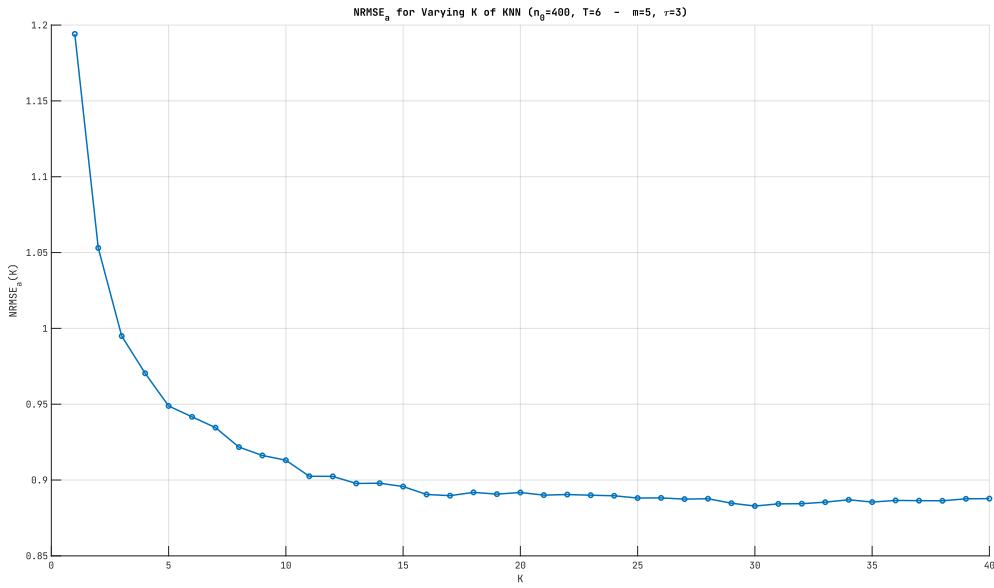
Το μόνο που πρέπει να τεθεί πριν γίνει εφαρμογή του μοντέλου τοπικού μέσου όρου για πρόβλεψη των στάσιμων χρονοσειρών, είναι ο αριθμός των κοντινοτερών γειτόνων,  $K$ . Ακολουθεί ανάλυση για την επιλογή της κατάλληλης τιμής του  $K$  για κάθε μία από τις στάσιμες χρονοσειρές που καταλήξαμε στα βήματα 2 και 3 αντίστοιχα.

### 6.2.1 Επιλογή του $K$ για τη στάσιμη χρονοσειρά A

Θα χρησιμοποιήσουμε την διθείσα συνάρτηση `localpredictmultistep()` για να κάνουμε προβλέψεις έως και  $T$  βήματα εμπρός με παραμέτρους:

- *Ορίζοντας Πρόβλεψης,  $T$ :* Θα χρησιμοποίουμε την βέλτιστη τιμή που καταλήξαμε στο βήμα 4, δηλαδή  $T = 6$
- *Διάσταση Εμβύθινσης,  $m$ :* Θα χρησιμοποίησουμε την τιμή που βρέθηκε από τη μέθοδο FNNs εφαρμοζόμενη σε όλη τη στάσιμη χρονοσειρά A, δηλαδή  $m = 5$
- *Υστέρηση,  $\tau$ :* Θα χρησιμοποήσουμε την τιμή που βρέθηκε από τη συνάρτηση αμοιβαίας πληροφορίας ολόκληρης της στάσιμης χρονοσειράς A, δηλαδή  $\tau = 3$

Παρακάτω εκτελούμε την συνάρτηση `localpredictmultistep()` για κάθε χρονική στιγμή με ορίσματα τις παραπάνω παραμέτρους καθώς και το  $K = 1, \dots, 40$ . Πλοτάρουμε το NRMSE των σφαλμάτων πρόβλεψης και με βάση αυτό επιλέγουμε την καταλληλότερη τιμή για το  $K$ .



Σχήμα 6.7: Διάγραμμα μεταβολής του μέσου NRMSE των σφαλμάτων πρόβλεψης ως προς τον αριθμό των κοντινότερων γειτόνων που συμμετέχουν στη πρόβλεψη,  $K$ , για τη στάσιμη χρονοσειρά  $\{X_a(t)\}$

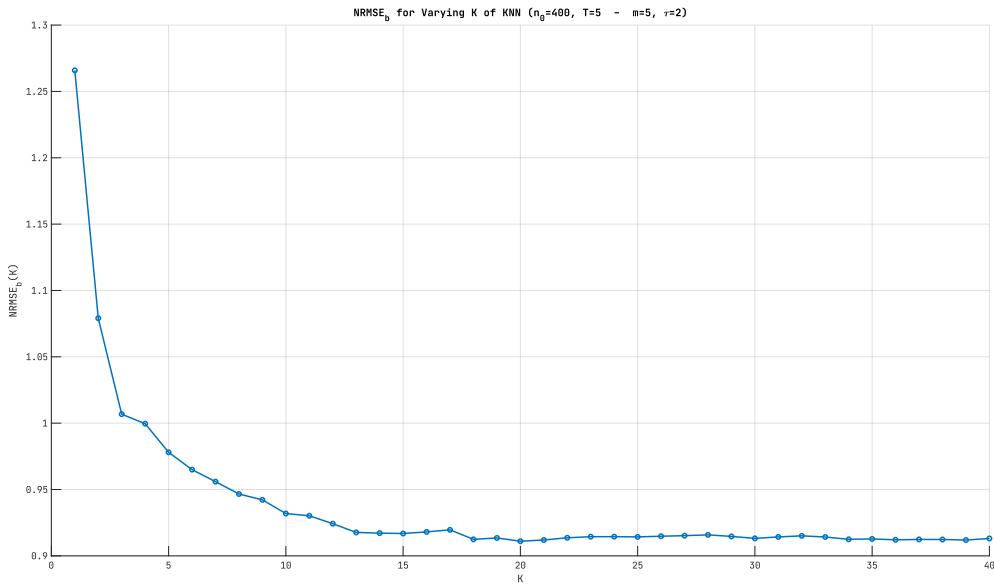
Από το παραπάνω διάγραμμα επιλέγουμε για τιμή για τον αριθμό των κοντινότερων γειτόνων,  $K \in [13, 16]$ , με προτιμότερη τιμή το **K=16**. Παρατηρούμε ότι για μικρές τιμές του  $K$  η πρόβλεψη είναι χειρότερη από το να προβλέπαμε μόνιμα την μέση τιμή ( $NRMSE > 1$ ), ενώ καθώς ανεβαίνει ο αριθμός των γειτόνων που συμμετέχουν στη πρόβλεψη υπάρχει απότομη πτώση αρχικά, κατόπιν ένα «γόνατο» (στο οποίο προσπαθούμε να επιλέξουμε τις τιμές) και τέλος ένα πλατίασμα (plateau) όπου περαιτέρω αύξηση οδηγεί σε μίκρες ή καθόλου αλλαγές του NRMSE.

### 6.2.2 Επιλογή του $K$ για τη στάσιμη χρονοσειρά B

Θα χρησιμοποιήσουμε και πάλι την διθείσα συνάρτηση `localpredictmultistep()` για να κάνουμε προβλέψεις και βρούμε τα σφάλματα πρόβλεψης για έως και  $T$  βήματα εμπρός, με παραμέτρους:

- *Ορίζοντας Πρόβλεψης,  $T$ :* Θα χρησιμοποίουμε την βέλτιστη τιμή που καταλήξαμε στο βήμα 5, δηλαδή  $T = 5$
- *Διάσταση Εμβύθινσης,  $m$ :* Θα χρησιμοποίησουμε την τιμή που βρέθηκε από τη μέθοδο FNNs εφαρμοζόμενη σε όλη τη στάσιμη χρονοσειρά B, δηλαδή  $m = 5$
- *Υστέρηση,  $\tau$ :* Θα χρησιμοποίησουμε την τιμή που βρέθηκε από τη συνάρτηση αμοιβαίας πληροφορίας ολόκληρης της στάσιμης χρονοσειράς B, δηλαδή  $\tau = 2$

Παρακάτω εκτελούμε την συνάρτηση `localpredictmultistep()` για κάθε χρονική στιγμή με ορίσματα τις παραπάνω παραμέτρους καθώς και το  $K = 1, \dots, 40$ . Πλοτάρουμε το NRMSE των σφαλμάτων πρόβλεψης και με βάση αυτό επιλέγουμε την καταλληλότερη τιμή για το  $K$ .



Σχήμα 6.8: Διάγραμμα μεταβολής του μέσου NRMSE των σφαλμάτων πρόβλεψης ως προς τον αριθμό των κοντινότερων γειτόνων που συμμετέχουν στη πρόβλεψη,  $K$ , για τη στάσιμη χρονοσειρά  $\{X_{b_{deseasoned}}(t)\}$

Από το παραπάνω διάγραμμα επιλέγουμε για τιμή για τον αριθμό των κοντινότερων γειτόνων, **K=13**. Το διάγραμμα αυτό έχει παρόμοια μορφή με το αντίστοιχο για τη στάσιμη χρονοσειρά A (σχήμα 6.7) όποτε ισχύουν και εδώ τα αντίστοιχα σχόλια.

### 6.3 Εφαρμογή Τοπικού Μοντέλου Κοντινότερων Γειτόνων για Εξαγωγή Σημείων Αλλαγής

Για την εύρεση των σημείων αλλαγής θα χρησιμοποιηθεί η ίδια μέθοδος που χρησιμοποιήθηκε για τα γραμμικά μοντέλα, σχέση (4.1), επίσης **εφαρμοζόμενη στις στάσιμες χρονοσειρές**,  $\{X_a(t)\}$  και  $\{X_{b_{deseasoned}}(t)\}$  αντίστοιχα. Ωστόσο αντί κάθε φόρα να προσαρμόζεται ένα γραμμικό μοντέλο τύπου ARMA και να γίνονται οι προβλέψεις βάσει αυτού, εδώ θα προσαρμόζεται ένα μη-γραμμικό, τοπικό μοντέλο κοντινότερων γειτόνων, και οι προβλέψεις θα γίνονται βάσει της σχέσης (6.4).

Θα εργαστούμε ξεχωριστά σε κάθε μία από τις στάσιμες χρονοσειες A και B. Για κάθε μία, θα παρουσιάσουμε τα surf plots από τα οποία θα δούμε έαν μπορούμε να εφαρμόσουμε την μέθοδο με τις παραμέτρους ( $T$  και  $\hat{m}_{std}$ ) που επιλέχθηκαν για τα γραμμικά μοντέλα, ή, εάν όχι, ποιες τίμες θα επιλέγαμε για τα μη-γραμμικά μοντέλα. Στη συνέχεια θα τρέξουμε τη μέθοδο για κάθε μία χρησιμοποιώντας τις αντίστοιχες βέλτιστες παραμέτρους, ενώ στο τέλος θα δωθούν τα διάγραμμα τόσο των στάσιμων όσο και των αρχικών χρονοσειρών με τοποθετημένα τα σημεία αλλαγής που εξήχθησαν.

Όπως και για τα γραμμικά, έτσι και εδώ κατά τη διάρκεια του grid search το μοντέλο **αναπροσαρμόζεται με βάση την επιλογή «c»**, δηλαδή παραμένει σταθερό έως ότου βρεθεί σημείο αλλαγής και προσαρμόζεται στις 400 τελευταίες παρατηρήσεις από το σημείο αλλαγής + T.

Η αναπροσαρμογή εδώ σημαίνει απλώς **επαναϋπολογισμός του**  $\alpha = \hat{m}_{std} * s_x$ , καθώς το μη γραμμικό μοντέλο υπολογίζει και κάνει προβλέψεις ταυτόχρονα, ára σε κάθε χρονική στιγμή.

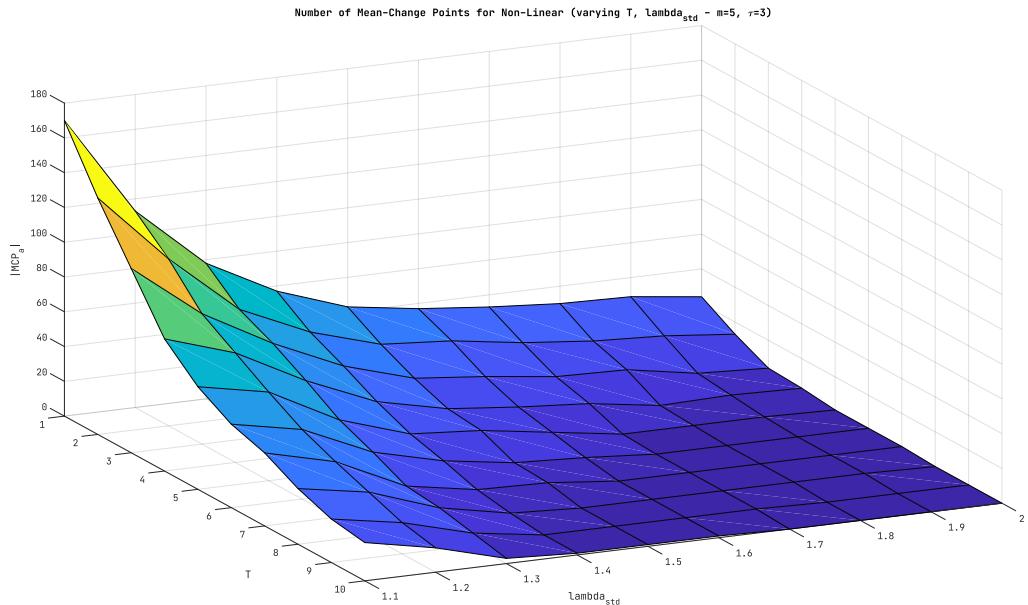
### 6.3.1 Εφαρμογή στη στάσιμη χρονοσειρά A

Για να τρέξει το grid search θέτουμε τη τιμή **K=16 κοντινότερους γείτονες** σύμφωνα με τη τιμή που προέκυψε από το σχήμα 6.7.

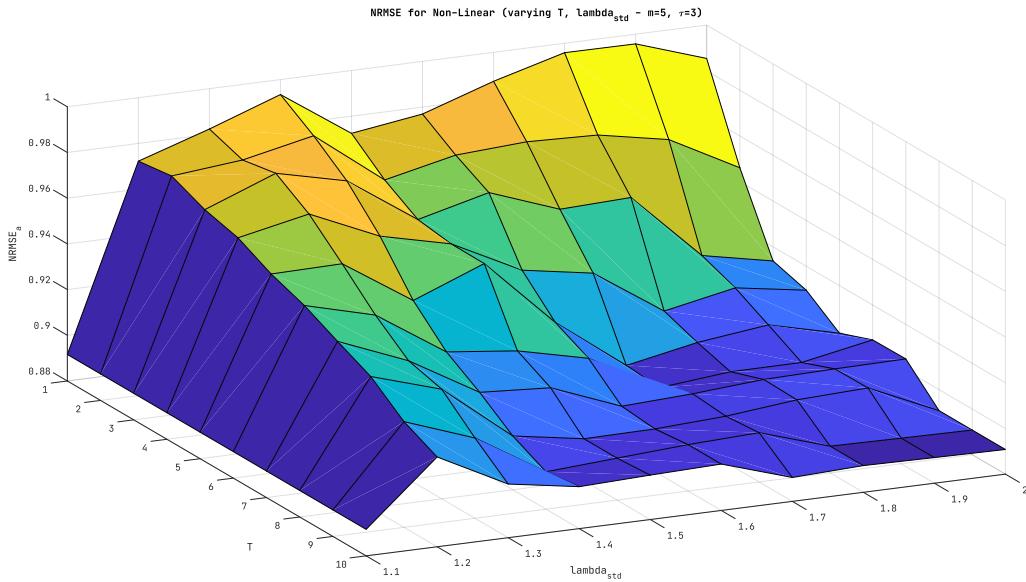
#### 6.3.1.1 Επιλογή Βέλτιστων Παραμέτρων

Για επιλογή βέλτιστων τιμών στις hyperparameters της μεθόδου και συγκεκριμένα στον ορίζοντα πρόβλεψης,  $T$ , και στο  $\lambda_{std}$  του ορίου απόφασης,  $a$ , θα κάνουμε αναζήτηση πλέγματος ως προς αυτές. Ως μετρικές για αξιολόγηση του κάθε συνδυασμού των παραμέτρων αυτών χρησιμοποιήθηκαν και εδώ ο Αριθμός των σημειών αλλαγής MCPs (Θέλουμε να μην είναι πολύ μεγάλος καθώς κάτι τέτοιο κάνει λιγότερο αξιόπιστη την όλη μέθοδο, αλλά να μην είναι και πολύ μικρός ώστε να υπάρχει νόημα χρήσης της μεθόδου), και το NRMSE των προβλέψεων (Θέλουμε κατά το δυνατό μικρότερο).

Αρχικά παραθέτονται σε διαγράμματα τύπου surf τα αποτελέσματα αναζήτησης πλέγματος ως προς τις παραπάνω μετρικές για τη στάσιμη χρονοσειρά  $\{X_a(t)\}$ , ενώ στη συνέχεια σχολιάζεται ο τρόπος επιλογής προσεγγεστικά βέλτιστων παραμάτρων αλλά και οι τελικές τους τιμές.



Σχήμα 6.9: Αριθμός σημείων αλλαγής της στάσιμης χρονοσειράς A,  $|MCP|$ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $a$ , για παραμέτρους μη-γραμμικού μοντέλου  $m = 5$ ,  $\tau = 3$  και  $K = 16$



Σχήμα 6.10: NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs) της στάσιμης χρονοσειράς A, για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $a$ , για παραμέτρους μη-γραμμικού μοντέλου  $m = 5$ ,  $\tau = 3$  και  $K = 16$

Γενικότερα, αναζητούμε τα «γόνατα» στις αντίστοιχες τρισδιάστατες καμπύλες έτσι ώστε περαιτέρω μεταβολές των αντίστοιχων παραμέτρων να μην είναι πλέον επικερδείς.

Επικεντρώνοντας στο πρώτο διάγραμμα και δεδομένου ότι θέλουμε ο αριθμός των MCPs να μήν είναι πολύ μεγάλος ή πολύ μικρός, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \in [6, 10] \quad \& \quad \lambda_{std} \in [1.2, 1.5] \quad (6.5)$$

Επικεντρώνοντας τώρα στο διάγραμμα των NRMSEs και δεδομένου ότι θέλουμε το NRMSE να είναι κατά το δυνατό μικρό, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \geq 5 \quad \& \quad \lambda_{std} \geq 1.3 \quad (6.6)$$

Συνδυάζοντας τις σχέσεις (6.5) και (6.6) παραπάνω καταλήγουμε ότι οι hyperparameters που θα χρησιμοποιηθούν για την εφαρμογή της μεθόδου αυτόματης εύρεσης χρονικών σημείων αλλαγής θα είναι: **T = 6 βήματα** και  **$\lambda_{std} = 1.5$** . Οι αντίστοιχες τιμές του grid search είναι: **|MCP| = 8 MCPs** και **NRMSE = 0.92**.

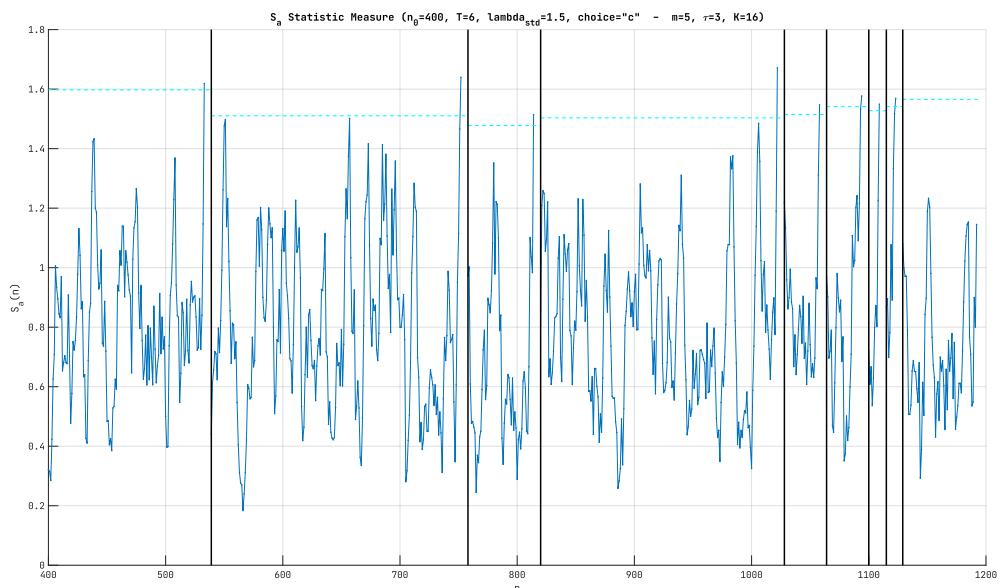
Η τελική επιλογή παραπάνω έγινε με βάση το μικρότερο NRMSE. Τυχάνει λοιπόν στη στάσιμη χρονοσειρά A, οι παράμετροι της μεθόδου να ίδιες για τα γραμμικά και μη-γραμμικά μοντέλα.

Τέλος, φαίνεται πως η χρήση του μη-γραμμικού, τοπικού μοντέλου  $K = 16$  κοντινότερων γειτόνων εμφανίζει μεγαλύτερο NRMSE κατά την εφαρμογή της μεθόδου εξαγωγής σημείων αλλαγής σε σύγκριση με τη χρήση γραμμικού μοντέλου MA(1).

### 6.3.1.2 Εφαρμογή Μεθόδου με Βέλτιστες Παράμετρους

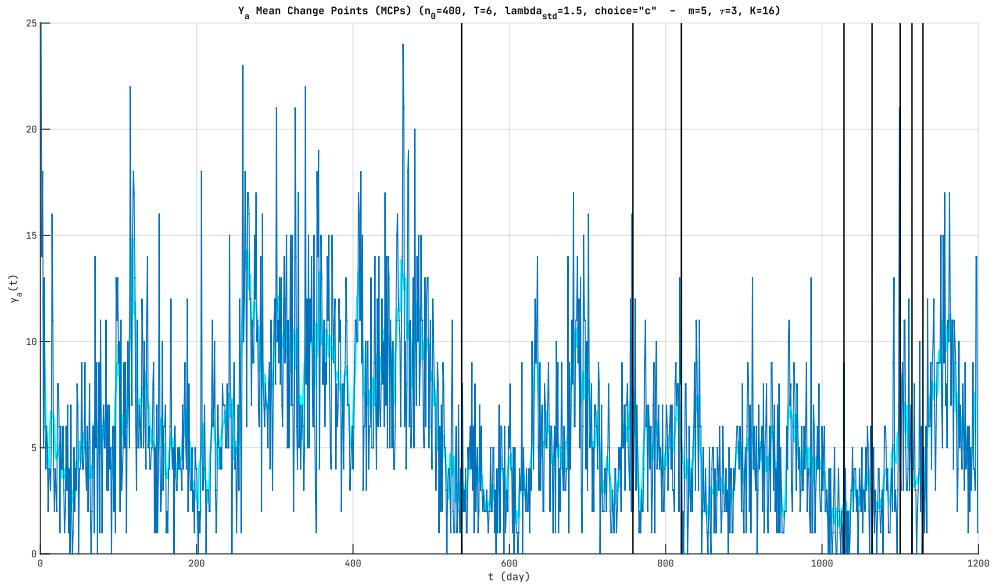
Χρησιμοποιώντας τις επιλεγμένες τιμές για τις hyperparameters της μεθόδου, δηλαδή ορίζοντα πρόβλεψης έως και 6 βημάτων εμπρός,  $T = 6$ , παράμετρο ορίου απόφασης στο 1.5,  $\lambda_{std} = 1.5$ , διάσταση εμβύθινσης,  $m = 5$ , υστέρηση,  $\tau = 3$  και αριθμό κοντινότερων γειτόνων,  $K = 16$ , θα τρέξουμε την παραπάνω μέθοδο στη στάσιμη χρονοσειρά που προέκυψε από το βήμα 1,  $\{X_a(t)\}$ , κάνοντας προβλέψεις με το μη-γραμμικό τοπικό μοντέλο μέσου όρου. Παρακάτω, φαίνονται τα σημεία αλλαγής που προκύπτουν από την εκτέλεση της μεθόδου με τις παραπάνω παραμέτρους για κάθε μια από τις επιλογές αναπροσαρμογής του μοντέλου («a», «b» ή «c»).

**Αναπροσαρμογή όταν βρεθεί σημείο αλλαγής - Επιλογή «c»**



Σχήμα 6.11: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου  $K=16$  κοντινότερων γειτόνων της στάσιμης χρονοσειράς  $\{X_a(t)\}$  και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης,  $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.92, 0.6sec]

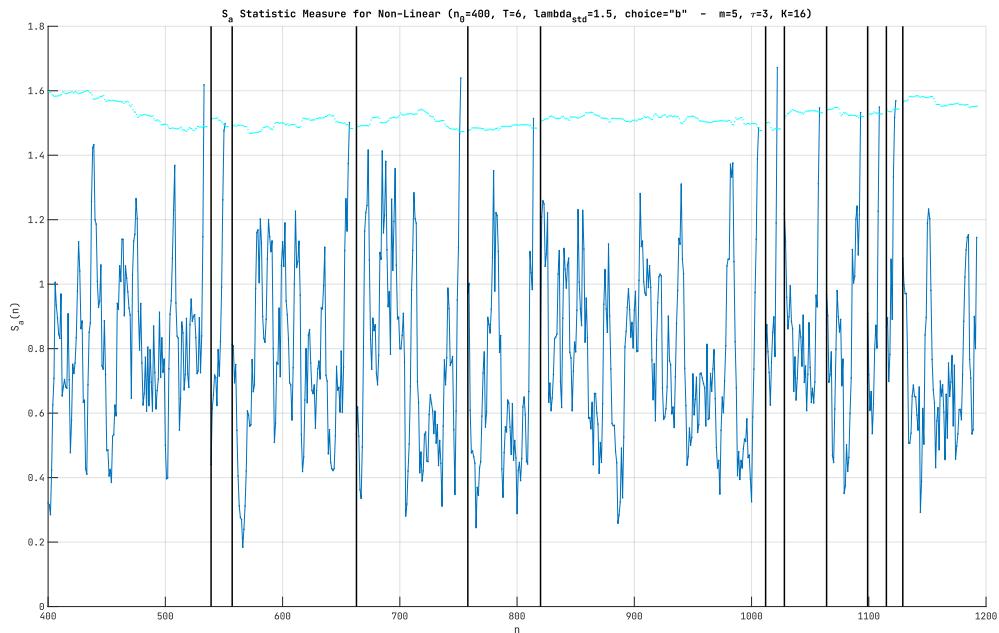
Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:



Σχήμα 6.12: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_a(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «c»

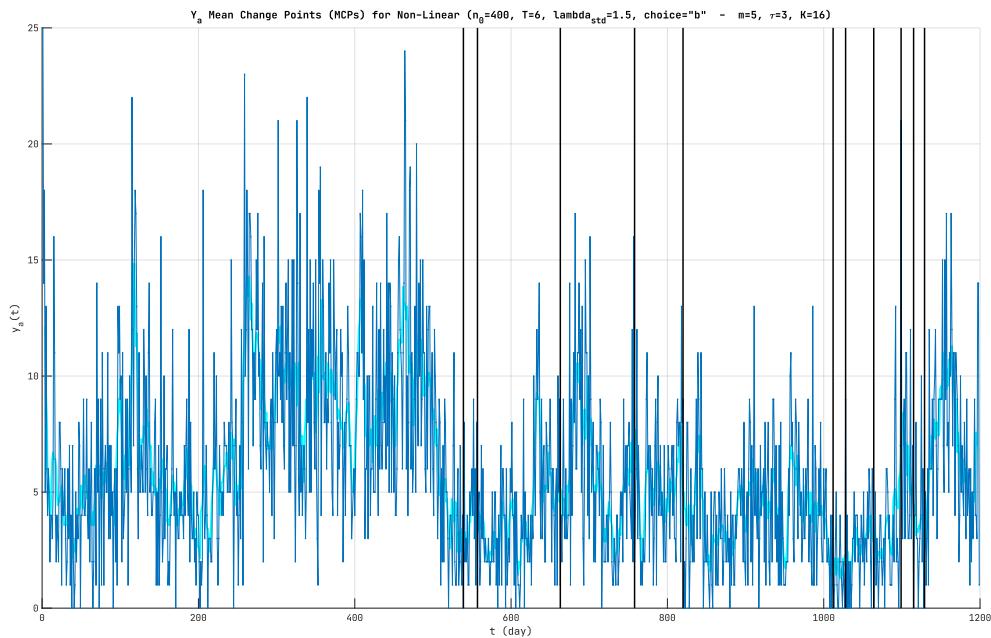
### Αναπροσαρμογή σε κάθε χρονική στιγμή - Επιλογή «b»

Τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή) ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 6.13: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου K=16 κοντινότερων γειτόνων της στάσιμης χρονοσειράς  $\{X_a(t)\}$  και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης,  $a = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.934, 0.62sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:

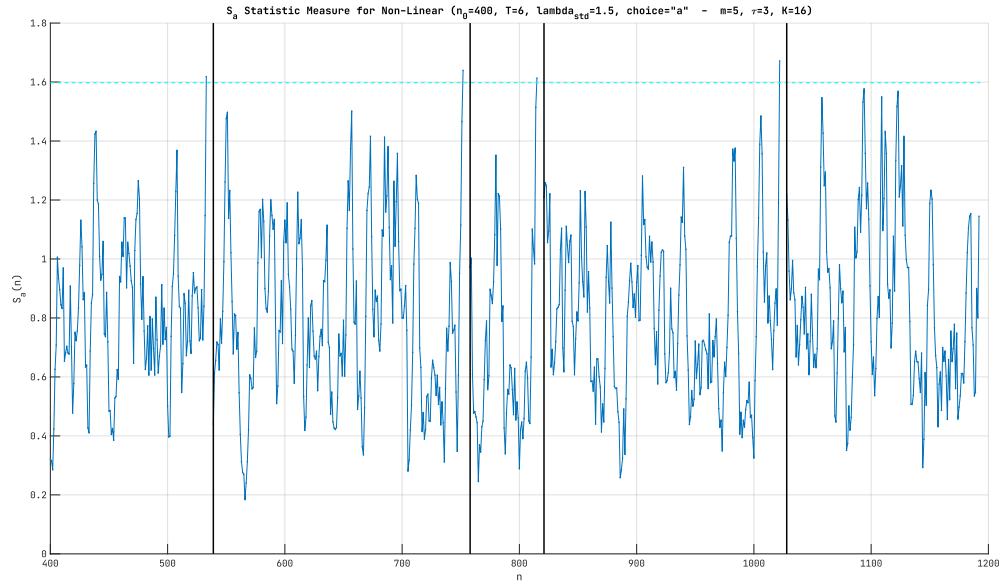


Σχήμα 6.14: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_\alpha(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «b»

Φαίνεται η κυμάτωση του ορίου απόφαση καθώς αυτό επαναϋπολογίζεται σε κάθε χρονική στιγμή. Επίσης, βλέπουμε πως έχουν βγει ακόμα τρία σημεία αλλαγής λόγω αυτής της μεταβολής και πλέον είναι 11 (ενώ πριν ήταν 8). Επίσης υπάρχει μικρή αύξηση του NRMSE.

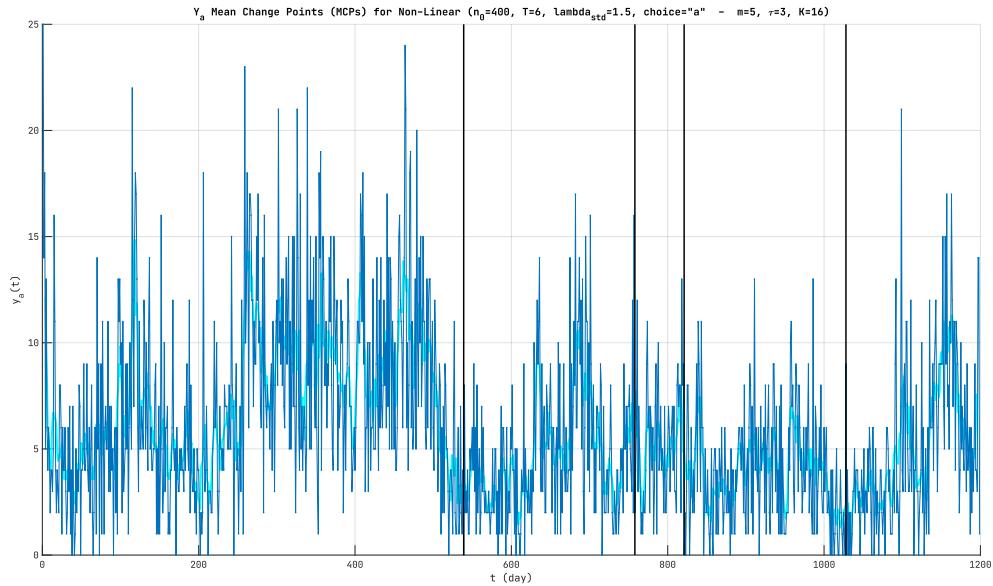
### Χωρίς Αναπροσαρμογή - Επιλογή «a»

Τέλος, τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «a» (καμία αναπροσαρμογή - διατήρηση της τυπικής απόκλισης / ορίου απόφασης που υπολογιστήκε από τις πρώτες 400 παρατηρήσεις της στάσιμης χρονοσειράς), ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 6.15: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου  $K=16$  κοντινότερων γειτόνων της στάσιμης χρονοσειράς  $\{X_\alpha(t)\}$  και για επιλογή αναπροσαρμογής «α» (χωρίες αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.910, 0.61sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:



Σχήμα 6.16: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_\alpha(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «α»

Τα σημεία αλλαγής είναι πολύ λιγότερα σε αριθμό και σε άλλες θέσεις σε σχέση με αυτά της επιλογής «c». Ο χρονος εκτέλεσης είναι ίδιος και το αλλά NRMSE αρκετά καλύτερο.

Γενικότερα όσο μειώνονται τα σημεία αλλαγής που βγάζει η μέθοδος (μέχρι καποιο σημείο) τόσο και το NRMSE των προβλέψεων κατά την εφαρμογή της μεθόδου μειώνεται.

Συγκρίνοντας τα σημεία αλλαγής με τα αντίστοιχα από τη γραμμική πρόβλεψη και ανάλυση του βήματος 4 βλέπουμε ότι **υπάρχει σχετική συνέπεια** στις θέσεις όπου εντοπίζονται σημεία αλλαγής, δηλαδή γύρω από τοπικές «εξάρσεις» ή τοπικές «βυθίσεις» της αρχικής χρονοσειράς προβολών του βίντεο A. Ωστόσο, επειδή **η επιλογή σημείων αλλαγής με γραμμικό μοντέλο πρόβλεψης** αφήνει μικρότερο NRMSE, θα λέγαμε πως **υπερτερεί** έναντι της χρήσης μη-γραμμικού τοπικού μοντέλου κοντινότερων γειτόνων.

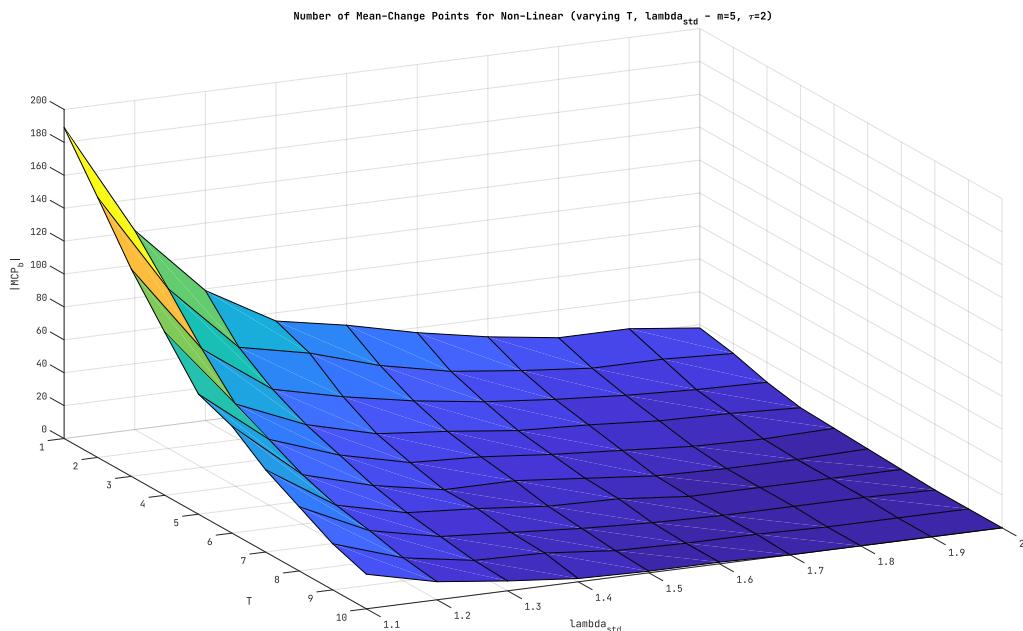
### 6.3.2 Εφαρμογή στη στάσιμη χρονοσειρά B

Για να τρέξει το grid search θέτουμε τη τιμή **K=13 κοντινότερους γείτονες** σύμφωνα με τη τιμή που προέκυψε από το σχήμα 6.8.

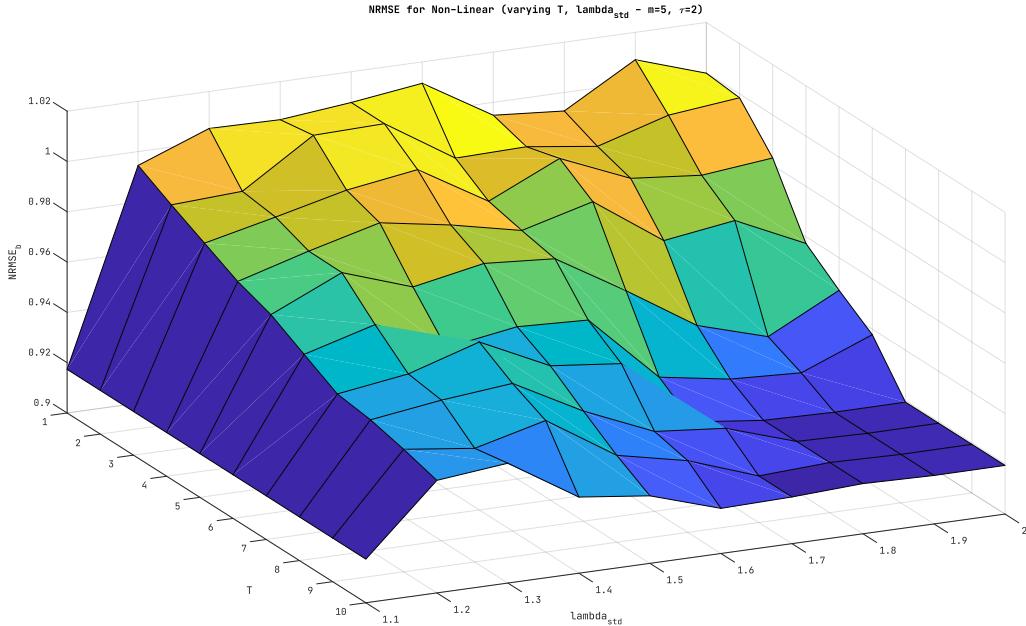
#### 6.3.2.1 Επιλογή Βέλτιστων Παραμέτρων

Για επιλογή βέλτιστων τιμών στις hyperparameters της μεθόδου και συγκεκριμένα στον ορίζοντα πρόβλεψης,  $T$ , και στο  $\lambda_{std}$  του ορίου απόφασης,  $a$ , θα κάνουμε και πάλι αναζήτηση πλέγματος ως προς αυτές. Ως μετρικές για αξιολόγηση του κάθε συνδυασμού των παραμέτρων αυτών χρησιμοποιήθηκαν και εδώ ο Αριθμός των σημειών αλλαγής MCPs και το NRMSE των προβλέψεων.

Αρχικά παραθέτονται σε διαγράμματα τύπου surf τα αποτελέσματα αναζήτησης πλέγματος ως προς τις παραπάνω μετρικές για τη στάσιμη χρονοσειρά  $\{X_{b_{deseasoned}}(t)\}$ , ενώ στη συνέχεια σχολιάζεται ο τρόπος επιλογής προσεγγεστικά βέλτιστων παραμάτρων αλλά και οι τελικές τους τιμές.



Σχήμα 6.17: Αριθμός σημειών αλλαγής της στάσιμης χρονοσειράς B,  $|MCP|$ , που προκύπτουν για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $a$ , για παραμέτρους μη-γραμμικού μοντέλου  $m = 5$ ,  $r = 2$  και  $K = 13$



Σχήμα 6.18: NRMSE προβλέψεων κατά τον υπολογισμό των σημείων αλλαγής (MCPs) της στάσιμης χρονοσειράς B, για κάθε τιμή του πλέγματος αναζήτησης ως πρός τον ορίζοντα πρόβλεψης,  $T$ , και το  $\lambda_{std}$  του ορίου απόφασης,  $\alpha$ , για παραμέτρους μη-γραμμικού μοντέλου  $m = 5$ ,  $\tau = 2$  και  $K = 13$

Γενικότερα, αναζητούμε τα «γόνατα» στις αντίστοιχες τρισδιάστατες καμπύλες έτσι ώστε περαιτέρω μεταβολές των αντίστοιχων παραμέτρων να μην είναι πλέον επικερδείς.

Επικεντρώνοντας στο πρώτο διάγραμμα και δεδομένου ότι θέλουμε ο αριθμός των MCPs να μήν είναι πολύ μεγάλος ή πολύ μικρός, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \in [5, 9] \quad \& \quad \lambda_{std} \in [1.2, 1.5] \quad (6.7)$$

Επικεντρώνοντας τώρα στο διάγραμμα των NRMSEs και δεδομένου ότι θέλουμε το NRMSE να είναι κατά το δυνατό μικρό, θα επιλέγαμε τις ακόλουθες τιμές (προσεγγιστικά):

$$T \geq 6 \quad \& \quad \lambda_{std} \geq 1.5 \quad (6.8)$$

Συνδυάζοντας τις σχέσεις (6.7) και (6.8) παραπάνω καταλήγουμε ότι οι hyperparameters που θα χρησιμοποιηθούν για την εφαρμογή της μεθόδου αυτόματης εύρεσης χρονικών σημείων αλλαγής θα είναι:  $T = 6$  βήματα και  $\lambda_{std} = 1.5$ . Οι αντίστοιχες τιμές του grid search είναι:  $|MCP| = 9$  MCPs και  $NRMSE = 0.975$ .

Η τελική επιλογή παραπάνω έγινε με βάση το μικρότερο NRMSE αλλά χωρίς μεγάλη μείωση των MCPs. Στη στάσιμη χρονοσειρά B, οι παράμετροι της μεθόδου δεν ίδιες για τα γραμμικά και μη-γραμμικά μοντέλα, αφού πλέον οι προβλέψεις γίνονται για έως και 6 βήματα εμπρός (σε σύγκριση με τα γραμμικά όπου ήταν για έως 5 βήματα εμπρός).

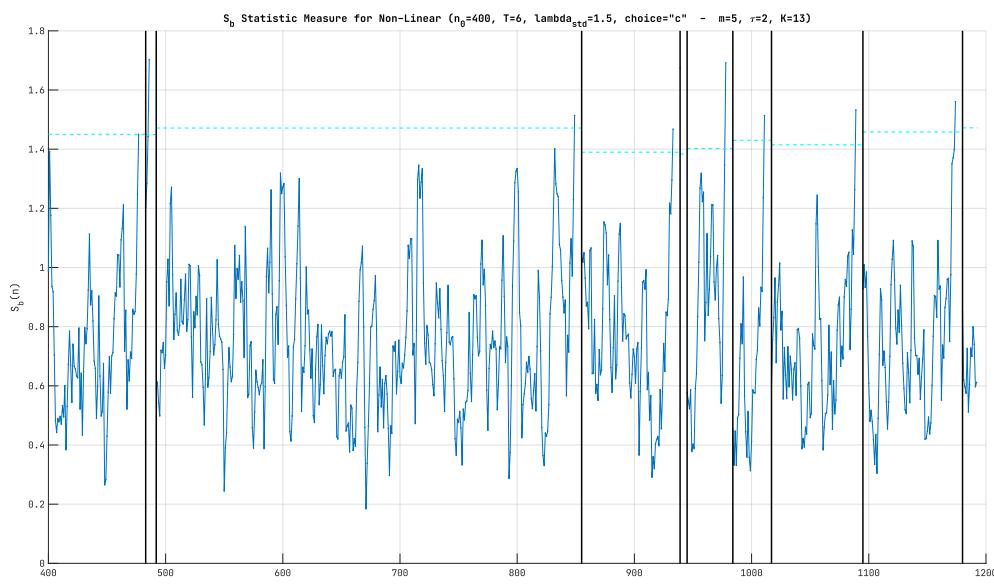
Για ίδιες παραμέτρους με τις βέλτιστες στη γραμμική ανάλυση θα προέκυπταν 13 MCPs (σε σύγκριση με τα 8 στη γραμμική) και NRMSE 0.977 (σε σύγκριση με το 0.892).

Φαίνεται λοιπόν πως και για τη δεύτερη χρονοσειρά, η χρήση του μη-γραμμικού, τοπικού μοντέλου κοντινότερων γειτόνων εμφανίζει μεγαλύτερο NRMSE κατά την εφαρμογή της μεθόδου εξαγωγής σημείων αλλαγής σε σύγκριση με τη χρήση γραμμικού μοντέλου.

### 6.3.2.2 Εφαρμογή Μεθόδου με Βέλτιστες Παράμετρους

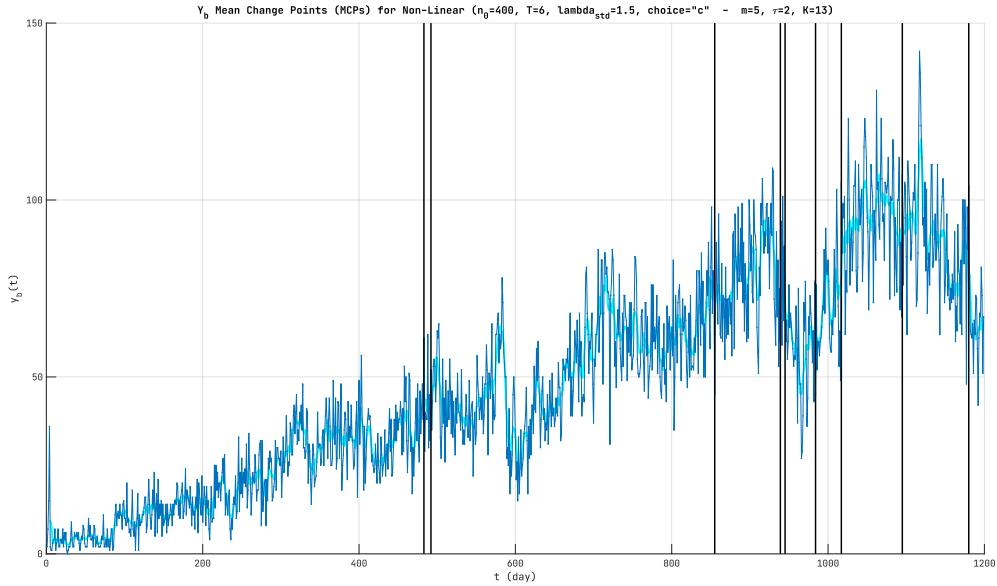
Χρησιμοποιώντας τις επιλεγμένες τιμές για τις hyperparameters της μεθόδου, δηλαδή ορίζοντα πρόβλεψης έως και 6 βημάτων εμπρός,  $T = 6$ , παράμετρο ορίου απόφασης στο 1.5,  $\lambda_{std} = 1.5$ , διάσταση εμβύθινσης,  $m = 5$ , υστέρηση,  $r = 2$  και αριθμό κοντινότερων γειτόνων,  $K = 13$ , θα τρέξουμε την παραπάνω μέθοδο στη στάσιμη χρονοσειρά που προέκυψε από το βήμα 3,  $\{X_{b_{deseasoned}}(t)\}$ , κάνοντας προβλέψεις με το μη-γραμμικό τοπικό μοντέλο μέσου όρου. Παρακάτω, φαίνονται τα σημεία αλλαγής που προκύπτουν από την εκτέλεση της μεθόδου με τις παραπάνω παραμέτρους για κάθε μια από τις επιλογές αναπροσαρμογής του μοντέλου («a», «b» ή «c»).

**Αναπροσαρμογή όταν βρεθεί σημείο αλλαγής - Επιλογή «c»**



Σχήμα 6.19: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου  $K=13$  κοντινότερων γειτόνων της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  και για επιλογή αναπροσαρμογής «c» (αναπροσαρμογή όταν βρεθεί σημείο αλλαγής). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.975, 0.61sec]

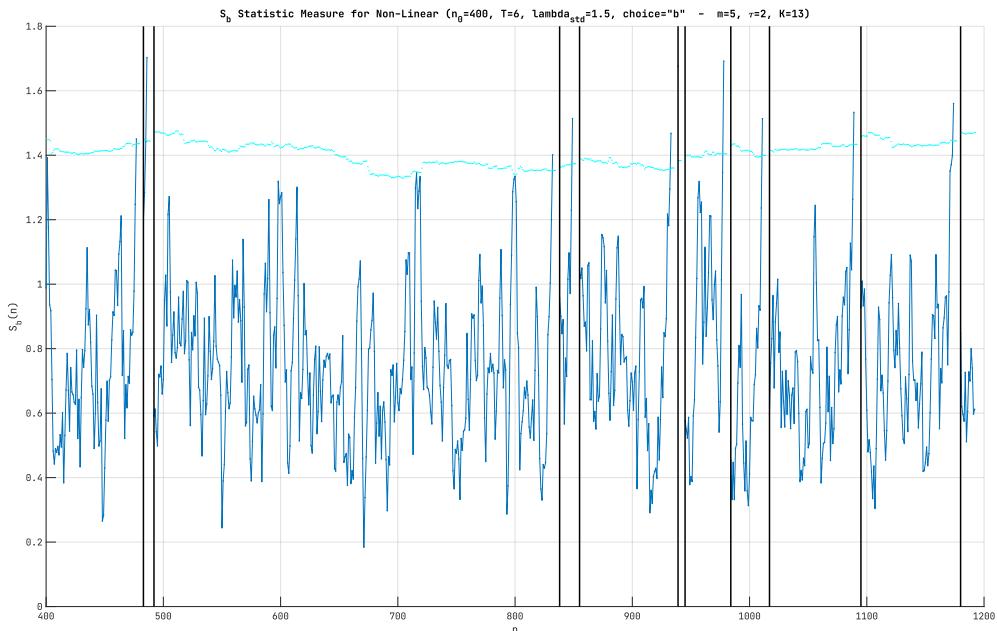
Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:



Σχήμα 6.20: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «c»

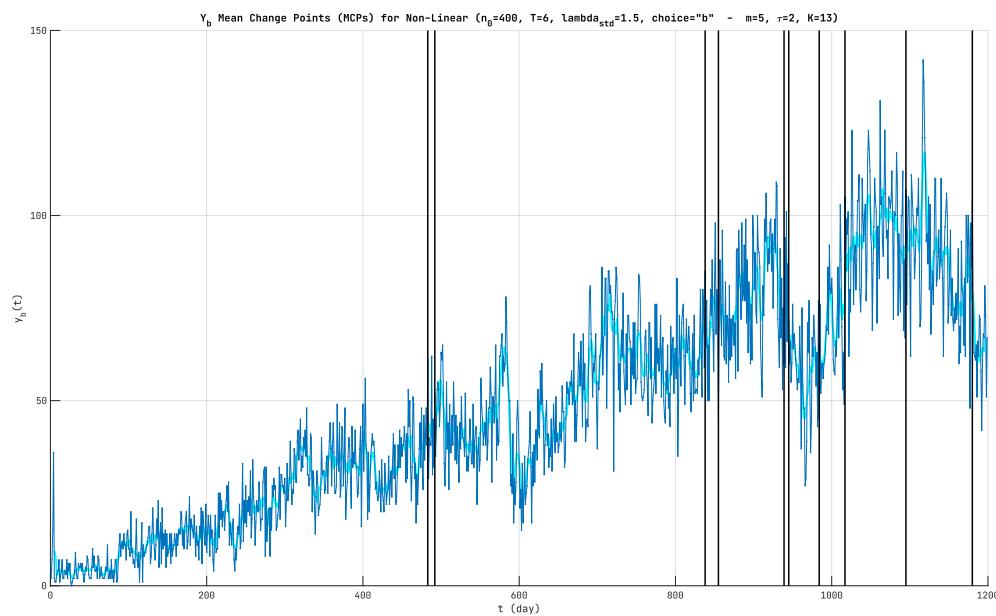
#### Αναπροσαρμογή σε κάθε χρονική στιγμή - Επιλογή «b»

Τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή) ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 6.21: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου K=13 κοντινότερων γειτόνων της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  και για επιλογή αναπροσαρμογής «b» (αναπροσαρμογή σε κάθε χρονική στιγμή). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.979, 0.63sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:

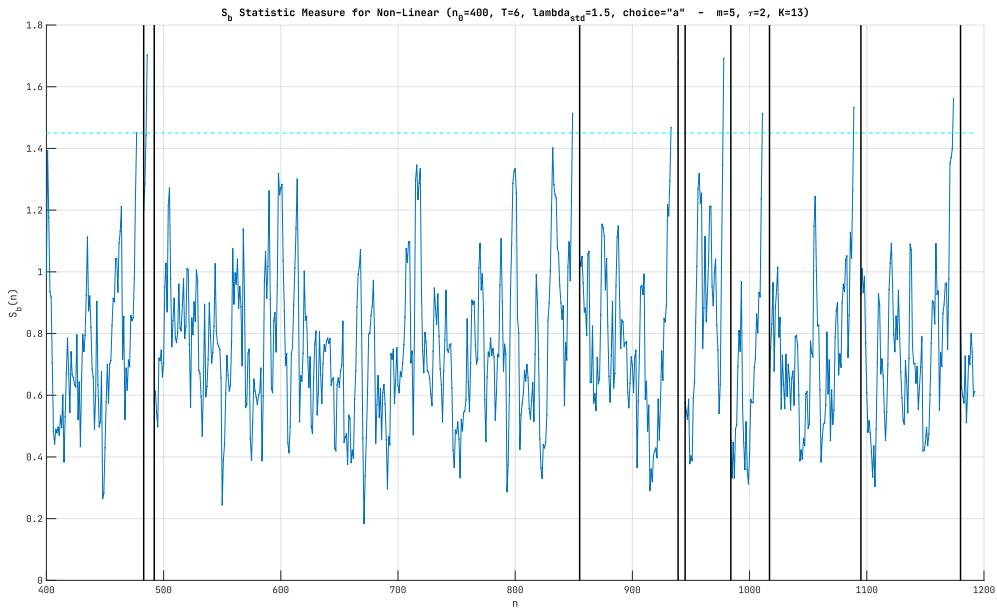


Σχήμα 6.22: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «b»

Φαίνεται η κυμάτωση του ορίου απόφαση καθώς αυτό επαναϋπολογίζεται σε κάθε χρονική στιγμή. Επίσης, βλέπουμε πως έχει βγει ακόμα ένα σημείο αλλαγής λόγω αυτής της μεταβολής και πλέον είναι 10. Επίσης υπάρχει μικρή αύξηση του NRMSE.

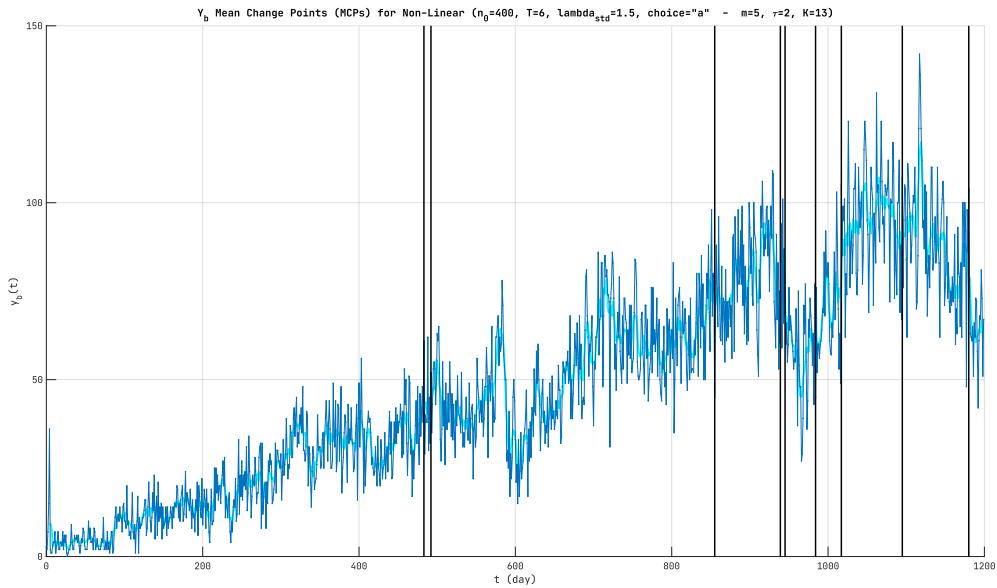
### Χωρίς Αναπροσαρμογή - Επιλογή «a»

Τέλος, τα ίδια διαγράμματα παρουσιάζονται για την επιλογή αναπροσαρμογής «a» (καμία αναπροσαρμογή - διατήρηση της τυπικής απόκλισης / ορίου απόφασης που υπολογιστήκε από τις πρώτες 400 παρατηρήσεις της στάσιμης χρονοσειράς), ενώ ακολουθεί σύντομος σχολιασμός:



Σχήμα 6.23: Τιμές στατιστικού  $S_n$  για έως και 6 βήματα μπροστά πρόβλεψη με τοπική πρόβλεψη μέσου  $K=13$  κοντινότερων γειτόνων της στάσιμης χρονοσειράς  $\{X_{b_{deseasoned}}(t)\}$  και για επιλογή αναπροσαρμογής «α» (χωρίες αναπροσαρμογή). Σημειώνονται επίσης το κριτήριο απόφασης,  $\alpha = 1.5 * s_x$ , (cyan) και φυσικά τα σημεία αλλαγής με έντονες κάθετες γραμμές στα εκάστοτε σημεία  $n + T$  (μαύρο) - [NRMSE=0.975, 0.61sec]

Παρακάτω, τα ίδια σημεία αλλαγής απεικονίζονται στην αρχική χρονοσειρά προβολών του βίντεο A:



Σχήμα 6.24: Διάγραμμα ιστορίας της αρχικής χρονοσειράς  $\{Y_b(t)\}$  (μπλε) μαζί με τα σημεία αλλαγής (μαύρο) που επιλέχθηκαν από τη μη-γραμμική ανάλυση της στάσιμης εκδοχής της με τις βέλτιστες παραμέτρους, καθώς και εκτίμηση της τάσης με φίλτρο κινούμενου μέσου τάξης 7 (MA(7) smoothing) - επιλογή «α»

Τα σημεία αλλαγής είναι τα ίδια σε αριθμό και σε θέσεις με αυτά της επιλογής «c». Ίδιος είναι επίσης ο χρονος εκτέλεσης και το NRMSE.

Γενικότερα, και για τη χρονοσειρά  $B$ , όσο μειώνονται τα σημεία αλλαγής που βγάζει η μέθοδος (μέχρι καποιο σημείο) τόσο και το NRMSE των προβλέψεων κατά την εφαρμογή της μεθόδου μειώνεται. Συγκρίνοντας τα σημεία αλλαγής με τα αντίστοιχα από τη γραμμική πρόβλεψη και ανάλυση του βήματος 4 βλέπουμε ότι **δεν υπάρχει έντονη συνέπεια** στις θέσεις όπου εντοπίζονται σημεία αλλαγής, δηλαδή στο μη-γραμμικό μοντέλο πρόβλεψης δεν είνα πάντα γύρω από τοπικες «εξάρσεις» ή τοπικές «βυθίσεις» της αρχικής χρονοσειράς προβολών του βίντεο  $B$ . Επίσης, επειδή **η επιλογή σημείων αλλαγής με γραμμικό μοντέλο πρόβλεψης** αφήνει μικρότερο NRMSE, θα λέγαμε πως και για τη δεύτερη χρονοσειρά **υπερτερεί** έναντι της χρήσης μη-γραμμικού τοπικού μοντέλου κοντινότερων γειτόνων.

## Βήμα 7

### Συμπερασματικά Σχόλια

Συγκρίνοντας τα αποτελέσματα και σημεία αλλαγής με αυτά του αντίστοιχου paper οι ακόλουθες παρατηρήσεις προκύπτουν:

- Η χρησιμοποίηση **γραμμικών** μοντέλων πρόβλεψης των στάσιμων χρονοσειρών (που έγιναν στάσιμες μετά από τις αντίστοιχες ενέργειες που περιγράφηκαν στα βήματα 1 και 3) οδηγεί εν γένει σε **καλύτερα αποτελέσματα** τόσο σε ότι αφορά τα σφαλμάτα πρόβλεψης όσο και στα σημεία αλλαγής που επιλέγονται από τη μέθοδο, τα οποία είναι σε σχετική συμφωνία και με τις θέσεις που επιλέγονται από τη μέθοδο του προτεινόμενου paper
- Η χρησιμοποίηση **μη-γραμμικών** τοπικών μοντέλων μέσου όρου αν και κάποιες φορές τείνει να δείνει τα ίδια ή κοντινά σημεία αλλαγής, δεν το επιτυγχάνει πάντα και δεδομένου των μεγαλύτερων σφαλμάτων πρόβλεψης τη θεωρούμε **υποδεέστερη**, κάποιες φορές μόλις καλύτερη από τη πρόβλεψη της μέσης τιμής
- Η πρόβλεψη με **γραμμικά** μοντέλα αν και καλύτερη από τα μη-γραμμικά τοπικά μοντέλα μέσου όρου, είναι **σημαντικά πιο αργή** στην εκτέλεσή της κάτι που οφείλεται κυρίως στο ότι πρέπει να εκτιμηθουν οι παράμετροι των εκάστοτε ARMA μοντέλων

Γενικότερα, είμαι πεπεισμένος πως αυτός ο τρόπος επιλογής σημείων αλλαγής να και όχι τόσο πολύπλοκος θεωρητικά και αλγορίθμικά, είναι ικανός να δώσει σωστά σημεία αλλαγής με αυτόματο και ευσταθή τρόπο. Πιθανή βελτίωση θα μπορούσε να αποτελέσει η δοκιμή και άλλων στατιστικών των σφαλμάτων πρόβλεψης, όπως το MSE. Επίσης, και με δεδομένο την ανωτερότητα των γραμμικών μοντέλων στη συγκεκριμένη μέθοδο, θα ήταν σκόπιμο να δοκιμαστεί η πρόβλεψη και με τοπικά γραμμικά μοντέλα πρόβλεψης (*local liner prediction models* στη θέση των τοπικών μοντέλων πρόβλεψης μέσου όρου (*local average prediction models*)).