



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών

Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και
Συστημάτων Αποφάσεων

**Θεωρητική Ανάλυση και Εμπειρική Μελέτη Εφαρμογής
Μεθοδολογίας Προβλέψεων ADIDA σε Δεδομένα Συνεχούς
Ζήτησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Π. Σπιθουράκης

Επιβλέπων: Βασίλειος Ασημακόπουλος,
Καθηγητής Ε.Μ.Π

Υπεύθυνος: Φώτιος Πετρόπουλος,
Υποψήφιος Διδάκτωρ Ε.Μ.Π

Αθήνα, Οκτώβριος 2011



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και
Συστημάτων Αποφάσεων

**Θεωρητική Ανάλυση και Εμπειρική Μελέτη Εφαρμογής
Μεθοδολογίας Προβλέψεων ADIDA σε Δεδομένα Συνεχούς
Ζήτησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Π. Σπιθουράκης

Επιβλέπων: Βασίλειος Ασημακόπουλος,
Καθηγητής Ε.Μ.Π

Υπεύθυνος: Φώτιος Πετρόπουλος,
Υποψήφιος Διδάκτωρ Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την^η Οκτωβρίου 2011.

.....
Β. Ασημακόπουλος,
Καθηγητής Ε.Μ.Π

.....
Ι. Ψαρράς,
Καθηγητής Ε.Μ.Π

.....
Γ. Μέντζας,
Καθηγητής Ε.Μ.Π

Αθήνα, Οκτώβριος 2011

.....
Γεώργιος Π. Σπιθουράκης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Γεώργιος Π. Σπιθουράκης
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Θέμα: Θεωρητική ανάλυση και εμπειρική μελέτη εφαρμογής μεθοδολογίας προβλέψεων ADIDA σε δεδομένα συνεχούς ζήτησης

Η εργασία μελετά θεωρητικά και εμπειρικά την εφαρμογή μεθοδολογίας προβλέψεων ADIDA, μίας προσέγγισης συνάθροισης-διάσπασης για παραγωγή προβλέψεων που προορίζεται για δεδομένα διακοπτόμενης ζήτησης. Η απόδοση της αξιολογείται με έμφαση στην πρόβλεψη δεδομένων συνεχούς ζήτησης.

Εισαγωγικά, παρουσιάζονται έννοιες και εφαρμογές των προβλέψεων. Αναπτύσσονται τα χαρακτηριστικά των χρονοσειρών και η διαδικασία αποσύνθεσης σε επιμέρους συνιστώσες. Μετά από την εκτενή καταγραφή μεθόδων πρόβλεψης, με έμφαση στην εκθετική εξομάλυνση, παρουσιάζονται σφάλματα μέτρησης της ακρίβειας και σκιαγραφείται, πραγματοποιώντας μία ιστορική αναδρομή, η σκοπιμότητα της αξιολόγησης μεταξύ των μεθόδων πρόβλεψης μέσω διαγωνισμών.

Μέσω θεωρητικής ανάλυσης, η μεθοδολογία διασπάται σε απλούστερα επιμέρους στάδια, για τα οποία παρέχεται ξεχωριστή μελέτη. Ως προς την πρακτική εφαρμογή, δείχνονται τα θετικά αποτελέσματα στην πρόβλεψη χρονοσειρών διακοπτόμενης ζήτησης. Αναγνωρίζονται τα κενά στην πειραματική εφαρμογή σε δεδομένα συνεχούς ζήτησης και έπειτα ελέγχεται η εξάρτηση της απόδοσης της μεθοδολογίας από μηχανισμούς διαχείρισης της εποχικότητας.

Στη συνέχεια, εξετάζεται θεωρητικά η επίδραση του επιπέδου συνάθροισης και δοκιμάζεται πειραματικά η βελτιωτική απόδοση της μεθοδολογίας, όπου ελέγχονται οι προσεγγίσεις χρήσης κοινού στο σύνολο των δεδομένων ή ξεχωριστού ανά χρονοσειρά επιπέδου συνάθροισης. Για τη δεύτερη προσέγγιση, εξετάζονται τεχνικές εκτίμησης του βέλτιστου επιπέδου και υπολογίζονται τα κατώτερα όρια σφάλματος. Επιπροσθέτως, μετά από λεπτομερή βιβλιογραφική έρευνα, περιγράφονται και μελετώνται εμπειρικά διάφοροι αλγόριθμοι διάσπασης.

Στο τέλος, τα συμπεράσματα συνοψίζονται αποκαλύπτοντας το δυναμικό της μεθοδολογίας για μείωση των σφαλμάτων πρόβλεψης και δίνονται οι κατευθύνσεις για περαιτέρω έρευνα στα ερωτήματα που προκύπτουν.

Λέξεις κλειδιά: προβλέψεις, χρονική συνάθροιση, διάσπαση, βελτιστοποίηση εντός δείγματος, αξιολόγηση, μεθοδολογικό πλαίσιο προβλέψεων

Abstract

Thesis: Theoretical analysis and empirical study of the application of ADIDA forecasting methodology to non-intermittent demand data

This study considers theoretical and empirical aspects of the application of ADIDA methodology, an aggregation-disaggregation approach to forecasting intermittent demand data. Its performance is assessed with emphasis on its use for forecasting non-intermittent demand data.

Initially, concepts and applications of forecasting are presented. The characteristics of time series and the process of decomposition into individual components are analyzed. After the description of forecasting methods, with emphasis on exponential smoothing, errors for accuracy measurement are presented and the purpose of the evaluation of forecasting methods is outlined, through a retrospection of forecasting competitions.

Through theoretical analysis, the methodology is broken down into simpler sub-stages, which are studied separately. With respect to practical application, positive results for predicting intermittent demand time series are shown. Scientific gaps in the experimental use on non-intermittent demand data are detected and the dependence of the methodology performance on mechanisms to handle seasonality is inquired.

Subsequently, the effect of aggregation level is theoretically examined and the improving effect of the methodology is experimentally inspected by testing two different approaches: using a fixed aggregation level across series or using separate aggregation level per series. For the second approach, techniques to estimate the optimal level are studied and the lower error bounds are calculated. Furthermore, after a thorough search in literature, various disaggregation algorithms are empirically studied.

In the end, the conclusions are summarized, revealing the potential of the methodology for forecast error reduction and directions for further research are proposed.

Keywords: forecasting, temporal aggregation, disaggregation, in-sample optimization, evaluation, forecasting framework

Πρόλογος

Θα ήθελα θερμότατα να ευχαριστήσω τον Καθηγητή Βασίλειο Ασημακόπουλο για την επιστημονική και ηθική υποστήριξη που αφειδώς μου παρείχε μέσω της συνεργασίας μας και της επίβλεψής του σε όλα τα στάδια της εκπόνησης της παρούσας εργασίας. Αποτελεί για εμένα πηγή έμπνευσης και παράδειγμα προς μίμηση με τις πολύπλευρες γνώσεις του, το καινοτόμο και διερευνητικό του πνεύμα, τις πολύτιμες και εύστοχες συμβουλές του και την πάραυτα ανθρώπινη στάση του.

Παράλληλα, θα ήθελα εκ βαθέων να ευχαριστήσω τους Καθηγητές Ι. Ψαρρά και Γ. Μέντζα, καθώς θεωρώ ιδιαίτερη τιμή για εμένα την συμμετοχή τους στην επιτροπή εξέτασης της διπλωματικής εργασίας.

Ιδιαίτερες ευχαριστίες πρέπει να αποδοθούν στον Καθηγητή Κ. Νικολόπουλο, χωρίς τις ανεκτίμητες συμβουλές και παραινέσεις του οποίου θα ήταν αδύνατη η έμπνευση και εκπόνηση μίας τέτοιας εργασίας.

Για την πιστή του συμπαράσταση και βοήθεια και τις εποικοδομητικές του προτάσεις και συμβουλές, ευχαριστώ ένθερμα τον υποψήφιο Διδάκτορα Φώτιο Πετρόπουλο. Οι οργανωτικές του ικανότητες στάθηκαν πολύτιμες στην προσπάθειά μου για την εκπόνηση της εργασίας, στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων & Στρατηγικής (Τομέας Ηλεκτρικών και Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο).

Ευχαριστώ επίσης τα υπόλοιπα μέλη της Μονάδας, Ήρα, Φωτεινή, Αχιλλέα και Όμηρο, για τη φιλική εμπύχωση και ηθική υποστήριξη που ανιδιοτελώς προσέφεραν.

Τέλος, ευχαριστώ την αγαπημένη μου οικογένεια για τα ηθικά εφόδια που μου παρείχε, ώστε να μπορώ να σταθώ στα πόδια μου και να ανταποκριθώ στις προκλήσεις και τις δυσκολίες.

Γεώργιος Π. Σπιθουράκης,

Αθήνα, Οκτώβριος 2011

Θέμα: Θεωρητική ανάλυση και εμπειρική μελέτη εφαρμογής μεθοδολογίας προβλέψεων ADIDA σε δεδομένα συνεχούς ζήτησης

Πίνακας Περιεχομένων

Περίληψη	5
Abstract.....	6
Πρόλογος	7
Πίνακας Περιεχομένων	9
Ευρετήριο Πινάκων	11
Ευρετήριο Εικόνων.....	11
ΚΕΦΑΛΑΙΟ 1: Εκτενής Περίληψη	13
1.1 Εισαγωγή στις Προβλέψεις	13
1.2 Περιγραφή και Θεωρητική Ανάλυση της Μεθοδολογίας ADIDA.....	14
1.3 Εφαρμογή της Μεθοδολογίας ADIDA.....	15
1.4 Επίπεδο Συνάθροισης	17
1.5 Αλγόριθμοι Διάσπασης	18
1.6 Συμπεράσματα και Προεκτάσεις.....	19
ΚΕΦΑΛΑΙΟ 2: Εισαγωγή στις Προβλέψεις.....	21
2.1 Γενικά.....	21
2.2 Χαρακτηριστικά των Χρονοσειρών	22
2.3 Μοντέλα Προβλέψεων.....	25
2.4 Σφάλματα και Ακρίβεια των Προβλέψεων	31
2.5 Διαγωνισμοί Προβλέψεων	33
ΚΕΦΑΛΑΙΟ 3: Περιγραφή και Θεωρητική Ανάλυση της Μεθοδολογίας ADIDA.....	37
3.1 Χρονική Συνάθροιση Δεδομένων	37
3.2 Η Μεθοδολογία Προβλέψεων ADIDA.....	39
3.3 Θεωρητική Ανάλυση των Σταδίων της ADIDA.....	42
3.3.1 Συνάθροιση (Aggregation)	42
3.3.2 Πρόβλεψη (Forecasting).....	46
3.3.3 Διάσπαση (Disaggregation)	47
3.3.4 Συνολική Διάταξη.....	49
ΚΕΦΑΛΑΙΟ 4: Εφαρμογή της Μεθοδολογίας ADIDA	51
4.1 Δεδομένα Διακοπτόμενης Ζήτησης.....	51

4.2	Δεδομένα Συνεχούς Ζήτησης	52
4.3	Μεθοδολογία ADIDA και Αποεποχικοποίηση.....	57
ΚΕΦΑΛΑΙΟ 5: Επίπεδο Συνάθροισης.....		59
5.1	Σημασία της Επιλογής του Επιπέδου Συνάθροισης	59
5.2	Κοινό Βέλτιστο Επίπεδο Συνάθροισης σε Σύνολο Χρονοσειρών	62
5.2.1	Μηνιαία Δεδομένα	62
5.2.2	Τριμηνιαία Δεδομένα.....	66
5.2.3	Ετήσια Δεδομένα	69
5.2.4	Άλλα Δεδομένα	72
5.2.5	Σύνοψη Αποτελεσμάτων και Συμπεράσματα	75
5.3	Βέλτιστο Επίπεδο Συνάθροισης Ξεχωριστά ανά Χρονοσειρά.....	78
5.3.1	Μέσο Τετραγωνικό Σφάλμα – MSE	79
5.3.2	Κριτήριο Πληροφορίας του Bayes – BIC	80
5.3.3	Κριτήριο Πληροφορίας του Akaike – AIC.....	81
5.3.4	Μηνιαία Δεδομένα	81
5.3.5	Τριμηνιαία Δεδομένα.....	84
5.3.6	Ετήσια Δεδομένα	87
5.3.7	Άλλα Δεδομένα	90
5.3.8	Σύνοψη Αποτελεσμάτων και Συμπεράσματα	92
ΚΕΦΑΛΑΙΟ 6: Αλγόριθμοι Διάσπασης.....		99
6.1	Βασικοί Αλγόριθμοι Διάσπασης	99
6.1.1	Ίσα Βάρη/Equal Weights (EQW).....	100
6.1.2	Προηγούμενα βάρη/PRevious Weights (PRW)	100
6.1.3	Μέσος όρος των βαρών/AVerage Weights (AVW).....	101
6.1.4	Εξισώσεις Yule-Walker (YW).....	101
6.2	Επίδραση του Αλγορίθμου Διάσπασης.....	107
ΚΕΦΑΛΑΙΟ 7: Συμπεράσματα και Προεκτάσεις		111
7.1	Συμπεράσματα.....	111
7.2	Προεκτάσεις.....	114
Βιβλιογραφικές Αναφορές		117
Παράρτημα Κώδικα Matlab.....		121
P.1	Μεθοδολογία ADIDA και Διαδικασίες Πειραμάτων	121
P.2	Μέθοδοι Πρόβλεψης	124
P.3	Συνάθροιση και Αλγόριθμοι Αποσύνθεσης	129
P.4	Σφάλματα και Τεχνικές Εκτίμησης του Βέλτιστου Επιπέδου Συνάθροισης ..	131

Π.5 Λοιπές Βοηθητικές Συναρτήσεις.....	132
--	-----

Ευρετήριο Πινάκων

Πίνακας 2.1	35
Πίνακας 2.2	35
Πίνακας 5.1	76
Πίνακας 5.2	77
Πίνακας 5.3	77
Πίνακας 5.4	78
Πίνακας 5.5	93
Πίνακας 5.6	93
Πίνακας 5.7	94
Πίνακας 5.8	95
Πίνακας 5.9	95
Πίνακας 5.10	95
Πίνακας 5.11	96
Πίνακας 5.12	96

Ευρετήριο Εικόνων

Εικόνα 3.1	42
Εικόνα 3.2	44
Εικόνα 4.1	54
Εικόνα 4.2	55
Εικόνα 4.3	56
Εικόνα 4.4	58
Εικόνα 5.1	63
Εικόνα 5.2	64
Εικόνα 5.3	64
Εικόνα 5.4	65
Εικόνα 5.5	65

Εικόνα 5.6	66
Εικόνα 5.7	67
Εικόνα 5.8	68
Εικόνα 5.9	68
Εικόνα 5.10	69
Εικόνα 5.11	70
Εικόνα 5.12	71
Εικόνα 5.13	71
Εικόνα 5.14	72
Εικόνα 5.15	72
Εικόνα 5.16	73
Εικόνα 5.17	74
Εικόνα 5.18	74
Εικόνα 5.19	75
Εικόνα 5.20	75
Εικόνα 5.21	82
Εικόνα 5.22	82
Εικόνα 5.23	83
Εικόνα 5.24	83
Εικόνα 5.25	85
Εικόνα 5.26	85
Εικόνα 5.27	86
Εικόνα 5.28	86
Εικόνα 5.29	88
Εικόνα 5.30	88
Εικόνα 5.31	89
Εικόνα 5.32	89
Εικόνα 5.33	90
Εικόνα 5.34	91
Εικόνα 5.35	91
Εικόνα 5.36	92
Εικόνα 6.1	107
Εικόνα 6.2	108
Εικόνα 6.3	108
Εικόνα 6.4	109

ΚΕΦΑΛΑΙΟ 1: Εκτενής Περίληψη

1.1 Εισαγωγή στις Προβλέψεις

Αρχικά, περιγράφεται το αντικείμενο και ο σκοπός της επιστήμης των προβλέψεων και σκιαγραφούνται περιληπτικά τα σημαντικότερα εργαλεία που χρησιμοποιεί για την επίτευξη των διάφορων στόχων της. Έμφαση δίνεται στους ποικίλους και πολύπλευρους τομείς όπου οι προβλέψεις βρίσκουν εφαρμογή, με ιδιαίτερη μνεία για τις ευρύτερα διεπιστημονικές εφαρμογές. Στη συνέχεια αναγνωρίζεται η επιτακτική ανάγκη για νέες εξελίξεις για την εξυπηρέτηση των σύγχρονων αναγκών μέσω της βελτίωσης των παρεχόμενων προβλέψεων και την διάπλαση μίας σφαιρικής και πληροφορημένης άποψης πάνω στις δυνατότητες και τους περιορισμούς τους.

Η μελέτη αυτή εξετάζει την επέκταση της χρήση του μεθοδολογικού πλαισίου προβλέψεων χρονοσειρών διακοπτόμενης ζήτησης ADIDA σε χρονοσειρές συνεχούς ζήτησης. Η έρευνα παρουσιάζει μία θεωρητική ανάλυση της μεθοδολογίας σε επιμέρους στάδια και προχωρά σε εμπειρική αξιολόγηση της αποδοτικότητάς της με πειραματική εφαρμογή στο σύνολο των δεδομένων του διαγωνισμού προβλέψεων M3. Για την βαθύτερη κατανόηση πτυχών της μεθοδολογίας, αλλά και των προβλέψεων γενικότερα, κρίνεται σκόπιμη η συγκεντρωτική παράθεση των σημαντικότερων σημείων σχετικά με τις χρονοσειρές, τα μοντέλα προβλέψεων και την αξιολόγησή τους μέσω των σφαλμάτων και των διαγωνισμών προβλέψεων.

Αναλυτικότερα, ορίζεται η οντότητα της χρονοσειράς και περιγράφονται συνοπτικά τα σπουδαιότερα χαρακτηριστικά που μπορεί να διαθέτει. Η εισαγωγή αυτή θέτει τις βάσεις για την παραγωγή προβλέψεων μέσω του μοντέλου των χρονοσειρών. Ακολούθως, εστιάζουμε στην προσέγγιση αποσύνθεσης των χρονοσειρών σε βασικές συνιστώσες, που αντιστοιχούν στα ποιοτικά χαρακτηριστικά τους. Συγκεκριμένα, παρουσιάζονται μέσω της περιγραφής των χαρακτηριστικών τους γνωρισμάτων οι παραδοσιακές συνιστώσες της τάσης, της εποχικότητας, της κυκλικότητας και της τυχαιότητας, στις οποίες οι χρονοσειρές μπορούν να διαχωριστούν μέσω διαδικασιών αποσύνθεσης. Ως τέτοια διαδικασία αναφέρεται η μέθοδος της κλασσικής αποσύνθεσης, ενώ αναφορά γίνεται για τα προσθετικά και πολλαπλασιαστικά μοντέλα, καθώς και υβριδικά μοντέλα, ενώ στη συνέχεια περιγράφεται η διαδικασία πρόβλεψης μέσω αποσύνθεσης. Έπεται η επεξήγηση της διάκρισης μεταξύ μεταβλητών ροής και μεταβλητών αποθέματος και μία βασική περιγραφή των πρακτικών επιπτώσεων του διαχωρισμού αυτού. Η καταγραφή των χαρακτηριστικών των χρονοσειρών κλείνει με την επισήμανση της σημαντικότητας διάκρισης μεταξύ μεταβλητών διακοπτόμενης και συνεχούς

ζήτησης, όπου παρατίθενται τα βασικότερα ποιοτικά χαρακτηριστικά των χρονοσειρών διακοπτόμενης ζήτησης και τα αριθμητικά μέτρα της ασυνέχειας των παρατηρήσεων (intermittency).

Ακολουθεί έπειτα μία διεξοδική περιγραφή πλήθους παραδοσιακών και σύγχρονων κοινά χρησιμοποιούμενων στατιστικών μεθόδων και τεχνικών πρόβλεψης, μετά την παρουσίαση των κύριων σημείων του μοντέλου των χρονοσειρών και του αιτιοκρατικού μοντέλου πρόβλεψης. Ξεκινώντας πρώτα με την απλοϊκή μέθοδο (Naïve), δίνεται ιδιαίτερη βαρύτητα στην γενική κατηγορία των μεθόδων εκθετικής εξομάλυνσης (σταθερού επιπέδου, γραμμικής και μη γραμμικής τάσης), όπου περιγράφεται αναλυτικά η σημασία των παραμέτρων εξομάλυνσης και των αρχικοποιήσεων των μεθόδων και παρέχονται βασικές διαδικασίες επιλογής των παραμέτρων. Ύστερα, αναφέρονται τα μοντέλα παλινδρόμησης, η χρήση νευρωνικών δικτύων, τα μοντέλα ARIMA, συνδυαστικές μέθοδοι και η μέθοδος Theta, ενώ, στο τέλος, μνεία γίνεται στις κριτικές μεθόδους πρόβλεψης.

Εξηγείται ο τρόπος αξιολόγησης της προσαρμογής των μοντέλων προβλέψεων και της ακρίβειας των ίδιων των προβλέψεων μέσω στατιστικών μέτρων σφάλματος. Ειδικότερα, διευκρινίζεται η διαφορά μεταξύ in-sample και out-of-sample σφαλμάτων και δίνονται οι τύποι και ορισμένες χρήσιμες πληροφορίες για το μέσο τετραγωνικό σφάλμα (MSE), τη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE), το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) και το συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (sMAPE).

Το κεφάλαιο καταλήγει στον ορισμό και τη σκοπιμότητα των διαγωνισμών προβλέψεων και αναλύει βασικές μεθοδολογίες που χρησιμοποιούνται. Πραγματοποιείται μία διεξοδική καταγραφή της διαθέσιμης βιβλιογραφίας για τους διαγωνισμούς προβλέψεων, εμμένοντας ιδιαίτερα στη σειρά διαγωνισμών M. Εστιάζουμε στον διαγωνισμό M3-Competition και την αναλυτική περιγραφή του συνόλου των δεδομένων του, καθώς αυτά χρησιμοποιούνται αργότερα στο πειραματικό τμήμα της παρούσας μελέτης.

1.2 Περιγραφή και Θεωρητική Ανάλυση της Μεθοδολογίας ADIDA

Στο κεφάλαιο αυτό, η μεθοδολογία ADIDA περιγράφεται και αναλύεται θεωρητικά σε επιμέρους στάδια, τα οποία μελετώνται ξεχωριστά. Λόγω της ιδιαίτερης συσχέτισης της χρονικής συνάθροισης με τη μεθοδολογία ADIDA, περιγράφεται αρχικά ως διαδικασία και αμέσως μετά επιχειρείται μία εκτενής βιβλιογραφική μελέτη των μέχρι στιγμής σχετικών επιστημονικών εξελίξεων. Το μεγαλύτερο μέρος της βιβλιογραφίας αναφέρεται στη συνάθροιση δεδομένων συνεχούς ζήτησης, ενώ ένα μικρότερο καλύπτει πτυχές των δεδομένων διακοπτόμενης ζήτησης. Έπειτα, παρουσιάζεται η αντίστροφη διαδικασία της

χρονικής συνάθροισης, η διάσπαση των συναθροισμένων δεδομένων, και περιγράφεται ένα βασικό πλαίσιο εφαρμογής των διαδικασιών αυτών.

Στη συνέχεια περιγράφεται και αναλύεται η ίδια η μεθοδολογία προβλέψεων ADIDA. Μετά από μία συνοπτική καταγραφή των οφελών που προσφέρει για την πρόβλεψη χρονοσειρών διακοπτόμενης ζήτησης, παρουσιάζονται τα βασικά βήματα εφαρμογής της διαδικασίας, η οποία μπορεί να εφαρμοστεί σε χρονοσειρά οποιουδήποτε τύπου ζήτησης. Ακολουθεί η εξήγηση των συμβολισμών που υιοθετούνται κοινώς για τις διάφορες παραμετροποιήσεις που μπορεί να λάβει η διαδικασία. Στο σημείο αυτό παρουσιάζεται επίσης μία εναλλακτική προσέγγιση της μοντελοποίησης της μεθοδολογίας. Αναλύοντας τη χρονική συνάθροιση ως έναν απλό εμπρός κινητό μέσο όρο ακολουθούμενο από μία μονάδα υποδειγματοληψίας και τη διάσπαση ως υπερδειγματοληψία ακολουθούμενη από έναν, σταθμισμένο εν γένει, πίσω κινητό μέσο όρο, καταλήγουμε σε μία συστημική προσέγγιση της συνολικής διαδικασίας (εικόνα 3.1), η οποία υπενθυμίζει πολυρυθμικά συστήματα επεξεργασίας σήματος.

Τα επιμέρους στάδια της μεθοδολογίας ADIDA αναλύονται ξεχωριστά, τόσο στο πεδίο του χρόνου, όσο και στο πεδίο της συχνότητας. Παρουσιάζονται εκτενώς διάφορες ενδιαφέρουσες ιδιότητες του απλού μέσου όρου και της εξάρτησής του από το επιλεγμένο επίπεδο συνάθροισης, ενώ αργότερα παρουσιάζεται η διαδικασία της δειγματοληψίας, ώστε να ολοκληρωθεί η θεωρητική περιγραφή του βήματος της συνάθροισης. Εφιστάται η προσοχή στο πρόβλημα της φασματικής αναδίπλωσης, η οποία σχετίζεται με την υποδειγματοληψία και μπορεί να προσθέσει ανεπιθύμητο θόρυβο στα δεδομένα. Για το λόγο αυτό περιγράφονται οι συνθήκες αποφυγής της και γίνεται αναλογία της χρονικής συνάθροισης με τη διαδικασία αποδεκατισμού.

Μετά από μία σύντομη αναφορά στην ευελιξία επιλογής μεθόδου πρόβλεψης στα πλαίσια της μεθοδολογίας, έπεται η θεωρητική ανάλυση του σταδίου της διάσπασης. Αναλύεται η έννοια της υπερδειγματοληψίας και της δράσης των σταθμισμένων κινητών μέσων όρων, με έμφαση στην περίπτωση των ίσων βαρών. Το στάδιο της διάσπασης στη συνέχεια παρουσιάζεται ως παράδειγμα διαδικασίας παρεμβολής. Στο τέλος, όλα τα θεωρητικά αποτελέσματα συνοψίζονται ώστε να αποκτηθεί μία πλήρης εικόνα για τη συνολική διάταξη.

1.3 Εφαρμογή της Μεθοδολογίας ADIDA

Η εφαρμογή της μεθοδολογίας ADIDA σε πραγματικά δεδομένα αποτελεί το βασικό αντικείμενο του κεφαλαίου αυτού. Αναλυτικότερα, μελετάται η εφαρμογή σε δεδομένα διακοπτόμενης και συνεχούς ζήτησης και στο τέλος εξετάζεται μια

ιδιαίτερη περίπτωση εφαρμογής της μεθοδολογίας, όπου τα δεδομένα είναι απευθείας διαθέσιμα μόνο στο συναθροισμένο επίπεδο.

Για την εφαρμογή στα δεδομένα διακοπτόμενης ζήτησης, το κύριο βάρος πέφτει στη μελέτη των Nikolopoulos et al. (2011), όπου η μεθοδολογία είχε δοκιμαστεί σε πραγματικά δεδομένα αποθεμάτων. Το σύνολο των δεδομένων ελέγχου περιγράφεται ως προς τα χαρακτηριστικά του, καθώς και η διαδικασία εφαρμογής της μεθοδολογίας. Στη συνέχεια, αναφέρονται μερικά από τα σημαντικότερα ευρήματα, όπως η επίδραση του επιπέδου συνάθροισης και η σπουδαιότητα της βελτιστοποίησης του. Παρατίθεται μία επιχειρησιακή ευριστική προσέγγιση για την επιλογή ενός εμπειρικά βέλτιστου επιπέδου. Έπεται η παρουσίαση των μεθόδων πρόβλεψης που ελέγχθηκαν και η συμπεριφορά της μεθοδολογίας ως μηχανισμός αυτοβελτίωσης των μεθόδων πρόβλεψης. Επιπλέον, αναφορά γίνεται στους αλγορίθμους διάσπασης που μελετήθηκαν και τα σχετικά αποτελέσματα που εξήχθησαν για την απόδοσή τους.

Όσον αφορά στα δεδομένα συνεχούς ζήτησης, αναγνωρίζεται αρχικά το κενό στην πρακτική εφαρμογή και εξηγούνται οι λόγοι που παροτρύνουν στην πειραματική εφαρμογή της μεθοδολογίας σε τέτοιου τύπου δεδομένα. Κατόπιν, ασχολούμαστε με το πρόβλημα επιλογής του κατάλληλου χρονικού επιπέδου καταγραφής και πρόβλεψης των δεδομένων. Έπεται η διευκρίνιση της διάκρισης μεταξύ τεχνητής (ελεγχόμενης) και αυτόματης (μη ελεγχόμενης) χρονικής συνάθροισης των δεδομένων. Σε όλα τα σημεία τονίζεται η σχετικότητα της μεθοδολογίας ADIDA, τόσο ως νέας μεθοδολογίας πρόβλεψης, όσο και ως μοντέλου που ερμηνεύει ήδη εμπειρικά χρησιμοποιούμενες πρακτικές. Επισημαίνεται η χρήση των δεδομένων του διαγωνισμού M3 για την εμπειρική εφαρμογή της μεθοδολογίας και καταγράφεται και περιγράφεται εκτενώς η σειρά των βημάτων της εφαρμοζόμενης πειραματικής διαδικασίας: αποεποχικοποίηση, χρήση μεθοδολογίας ADIDA για πρόβλεψη, επανεποχικοποίηση, επανάληψη για το σύνολο των χρονοσειρών και υπολογισμοί σφαλμάτων και λοιπών αποτελεσμάτων. Δίνονται σε πρώτη φάση παραδείγματα μεταβολής των σφαλμάτων λόγω της εφαρμογής της μεθοδολογίας σε μεμονωμένες χρονοσειρές, τα οποία δίνουν ποικίλα αποτελέσματα. Αμέσως μετά όμως δίνεται επιπλέον παράδειγμα μείωσης των μέσων σφαλμάτων πρόβλεψης σε σύνολο δεδομένων χάρη στην εφαρμογή της μεθοδολογίας, αποτέλεσμα που αιτιολογεί το ενδιαφέρον της έρευνας αυτής και παρουσιάζει το δυναμικό της μεθοδολογίας. Κλείνοντας, ιδιαίτερη αναφορά γίνεται για την σπουδαιότητα της κατάλληλης επιλογής του επιπέδου συνάθροισης.

Στην τελευταία υποενότητα εξετάζεται η περίπτωση που τα δεδομένα είναι απευθείας και αποκλειστικά διαθέσιμα στο συναθροισμένο επίπεδο, οπότε δεν παρέχεται η δυνατότητα αποεποχικοποίησης. Το παράδειγμα της προηγούμενης ενότητας επεκτείνεται, αποκαλύπτοντας και πάλι την εν δυνάμει βελτιωτική δράση

της μεθοδολογίας. Ελέγχεται η επίδραση της μετέπειτα επανεποχικοποίησης των παραγόμενων προβλέψεων και ερμηνεύεται η ικανότητα της χρονικής συνάθροισης να απομακρύνει ενίοτε την εποχικότητα των δεδομένων. Τέλος, θέτονται οι πρακτικοί περιορισμοί μίας τέτοιας προσέγγισης και εκφράζονται οι απαραίτητες προειδοποιήσεις για την εγκυρότητα των αποτελεσμάτων.

1.4 Επίπεδο Συνάθροισης

Το κεφάλαιο αυτό ασχολείται με το επίπεδο συνάθροισης της χρονοσειράς, την επίδρασή του στα αποτελέσματα της μεθοδολογίας και τη σημασία της επιλογής του. Επικεντρωνόμαστε σε δύο εναλλακτικές προσεγγίσεις, τη χρήση κοινού επιπέδου για όλες τις χρονοσειρές ή την χρήση ξεχωριστού επιπέδου ανά χρονοσειρά. Για κάθε προσέγγιση εξετάζονται θεωρητικά ζητήματα και πραγματοποιείται πειραματική μελέτη για κάθε χρονική κατηγορία δεδομένων και μέθοδο πρόβλεψης.

Με αναφορά στη σπουδαιότητα επιλογής ενός κατάλληλου επιπέδου συνάθροισης, αναγνωρίζεται η αδυναμία εφαρμογής του εμπειρικού κανόνα των Nikolopoulos et al. (2011) για τα δεδομένα συνεχούς ζήτησης. Στην προσπάθεια απόκτησης ενόρασης για τη λειτουργία του επιπέδου συνάθροισης, αυτό αποκαλύπτεται ότι ταυτίζεται με το μήκος των κινητών μέσων όρων των σταδίων της ADIDA. Εξετάζεται ως συνέπεια της ταύτισης αυτής η επίδραση του επιπέδου συνάθροισης στην δυνατότητα των ΚΜΟ για μείωση του θορύβου και εξομάλυνση των δεδομένων. Μία επιπλέον κυρίαρχη επίδραση εντοπίζεται λόγω της σύνδεσης του επιπέδου συνάθροισης με την μονάδα υποδειγματοληψίας. Βάσει του θεωρήματος δειγματοληψίας, αποδεικνύεται η ύπαρξη μίας ανώτατης τιμής, πάνω από την οποία φαινόμενα φασματικής αναδίπλωσης καταστρέφουν την πληροφορία του αρχικού σήματος. Οι διαφορετικοί τρόποι επίδρασης του επιπέδου συνάθροισης ανακεφαλαιώνονται και ερμηνεύονται τα χαρακτηριστικά που πρέπει να πληροί το βέλτιστο επίπεδο.

Μελετάται αρχικά ο απλούστερος τρόπος επιλογής του επιπέδου συνάθροισης, που είναι η χρήση κοινού επιπέδου για ολόκληρο το σύνολο των χρονοσειρών. Παραθέτονται τα θετικά σημεία και οι διευκολύνσεις αυτής της προσέγγισης, αλλά και οι περιορισμοί που σχετίζονται με την ομοιογένεια ή όχι των δεδομένων. Για την αντικειμενική αποτίμηση της αποδοτικότητας της προσέγγισης αυτής, πραγματοποιείται πειραματική μελέτη επί των δεδομένων του διαγωνισμού M3, όπου μετράται η out-of-sample ακρίβεια των προβλέψεων. Η εφαρμογή γίνεται χωριστά για κάθε χρονική κατηγορία δεδομένων και μέθοδο πρόβλεψης. Ακολουθεί η συγκεντρωμένη καταγραφή των αποτελεσμάτων, τα οποία ποικίλουν δραστικά ανάλογα με την εκάστοτε εφαρμογή και βρίσκονται σε συμφωνία με τα αντίστοιχα

των Nikolopoulos et al. (2011). Τα θετικότερα ευρήματα αφορούν τα μηνιαία δεδομένα, όπου παρατηρούνται οι σημαντικότερες μειώσεις σφάλματος.

Στη συνέχεια, εξετάζεται η επιλογή επιπέδου συνάρτησης ξεχωριστά ανά χρονοσειρά. Παρουσιάζεται ένα γενικό πλαίσιο εκτίμησης της βέλτιστης τιμής και ο τρόπος με τον οποίο μπορούμε να υπολογίσουμε το βέλτιστο αποτέλεσμα σε όρους σφάλματος αν είχαμε εξ αρχής γνώση των *a posteriori* βέλτιστων επιπέδων συνάθροισης. Παρουσιάζονται τρεις ευρέως χρησιμοποιούμενες μέθοδοι εκτίμησης, μέσω της ελαχιστοποίησης του MSE, του BIC ή του AIC. Για το κάθε κριτήριο δίνονται απαραίτητες πληροφορίες για την χρήση και τη συμπεριφορά του. Επαναλαμβάνεται η πειραματική μελέτη και για την προσέγγιση αυτή. Και πάλι, εξετάζονται χωριστά οι χρονικές κατηγορίες δεδομένων, οι μέθοδοι πρόβλεψης και οι τεχνικές εκτίμησης του βέλτιστου επιπέδου. Τα αποτελέσματα, τα οποία εμφανίζουν και για την προσέγγιση αυτή μεγάλη ποικιλία, συγκεντρώνονται και συνοψίζονται. Οι μεθοδολογία έδωσε γενικά ικανοποιητικά αποτελέσματα μείωσης σφάλματος, ωστόσο η εκ των προτέρων γνώση των *a posteriori* βέλτιστων αποκαλύπτει ακόμα πιο ευρέα περιθώρια περαιτέρω βελτιώσεων. Τα περιθώρια αυτά κάνουν την εφαρμογή και διερεύνηση της προσέγγισης αυτής ιδιαίτερα δελεαστικές. Αν και οι υπόλοιπες, υλοποιήσιμες τεχνικές εκτίμησης έδωσαν βελτιώσεις, δεν κατάφεραν να προσεγγίσουν πλήρως την ποιότητα των *a posteriori* αποτελεσμάτων.

1.5 Αλγόριθμοι Διάσπασης

Βασικό αντικείμενο του κεφαλαίου αυτού αποτελούν οι αλγόριθμοι διάσπασης της χρονοσειράς και η επίδρασή τους στην αποτελεσματικότητα της μεθοδολογίας. Ακολουθώντας την επισκόπηση ορισμένων βασικών μεθόδων, η πειραματική εφαρμογή αποκαλύπτει την συμπεριφορά κάθε μίας στα μέσα σφάλματα των μεθόδων.

Πρώτα, εισάγεται η έννοια του διανύσματος των βαρών διάσπασης και η ανάγκη ικανοποίησης της συνθήκης κανονικοποίησης και παρουσιάζεται ένα γενικής χρήσης απλό πλαίσιο διάσπασης μίας χρονοσειράς. Παρατίθενται τα σημαντικότερα ευρήματα της βιβλιογραφικής έρευνας για την διάσπαση χρονοσειρών σε λεπτότερα χρονικά επίπεδα και περιγράφονται οι προς μελέτη μέθοδοι των ίσων, προηγούμενων και μέσω βαρών, καθώς και μία τροποποιημένη έκδοση της στατιστικής μεθόδου των Yule-Walker. Για κάθε μέθοδο περιγράφεται ο τρόπος υπολογισμού των βαρών, ενώ για τον προτεινόμενο τροποποιημένο αλγόριθμο Yule-Walker εξηγείται το μαθηματικό πλαίσιο που χρησιμοποιείται.

Για την αξιολόγηση των αλγορίθμων διάσπασης πραγματοποιείται εμπειρική μελέτη για κάθε κατηγορία δεδομένων, μέθοδο πρόβλεψης και αλγόριθμο

διάσπασης. Τα αποτελέσματα ταυτίζονται ακριβώς με εκείνα της έρευνας των Nikolopoulos et al. (2011). Στις περισσότερες περιπτώσεις, διατηρείται η σειρά απόδοσης των αλγορίθμων, με τον αλγόριθμο των ίσων βαρών να προηγείται, ακολουθούμενο από τον αλγόριθμο μέσον βαρών και την τροποποιημένη μέθοδο Yule-Walker, ενώ ο αλγόριθμος των προηγούμενων βαρών έδωσε τα χειρότερα αποτελέσματα.

1.6 Συμπεράσματα και Προεκτάσεις

Ως επίλογος της εργασίας, παρουσιάζονται συνοπτικά τα σημαντικότερα συμπεράσματα και δίνονται κατευθύνσεις για μελλοντική έρευνα. Επισκοπούνται οι λόγοι που οδήγησαν στην εκπόνηση της συγκεκριμένης εργασίας και παρουσιάζονται τα σπουδαιότερα ευρήματα αναφορικά με την καλή απόδοση της μεθοδολογίας, τις δυνατότητες γενίκευσής της, την θεωρητική ανάλυση σε επιμέρους στάδια που πραγματοποιήθηκε και τη λειτουργία της ως μηχανισμού αυτοβελτίωσης των μεθόδων πρόβλεψης. Βάσει της ανάλυσης σε στάδια, συνοψίζονται οι σημαντικότερες επιδράσεις του επιπέδου συνάθροισης και τα αποτελέσματα των δύο προσεγγίσεων ως προς την επιλογή του επιπέδου (κοινό ή ξεχωριστά ανά χρονοσειρά). Συμπεριλαμβάνονται τα συμπεράσματα για τις τεχνικές εκτίμησης του βέλτιστου επιπέδου συνάθροισης ανά χρονοσειρά και των αλγορίθμων διάσπασης.

Για την συνέχιση της έρευνας, δίνονται οι απαραίτητες κατευθύνσεις και τα σημεία που απαιτούν επιπλέον διερεύνηση. Τα επισημασμένα πεδία αφορούν τη θεωρητική ανάλυση, αλλά και την εμπειρική επανάληψη της έρευνας με διαφορετικά σύνολα δεδομένων, μη δοκιμασμένες μεθόδους πρόβλεψης, νέες τεχνικές εκτίμησης του βέλτιστου επιπέδου συνάθροισης και νέους αλγορίθμους διάσπασης. Ιδιαίτερη σημασία δίνεται στο επίπεδο συνάθροισης, το οποίο πρέπει να ελεγχθεί πειραματικά και θεωρητικά ως προς την ύπαρξη βέλτιστης τιμής. Στο τέλος, υπενθυμίζονται οι λόγοι που καθιστούν επιτακτική την ανάγκη της έρευνας στην κατεύθυνση αυτή, με σκοπό τον εμπλουτισμό των γνώσεων πάνω στις προβλέψεις και την βελτίωση των πρακτικών εφαρμογών τους.

ΚΕΦΑΛΑΙΟ 2: Εισαγωγή στις Προβλέψεις

2.1 Γενικά

Η επιστήμη των προβλέψεων αποτελεί έναν υβριδικό κλάδο των μαθηματικών, της στατιστικής και της επιχειρησιακής έρευνας. Βασικός στόχος της είναι η παραγωγή ποσοτικών ή ποιοτικών προβλέψεων σχετικά με την εξέλιξη κάποιων φαινομένων ή μεταβλητών στο μέλλον. Για το σκοπό αυτό χρησιμοποιεί μαθηματικά εργαλεία, στατιστικές, εμπειρικές και οργανωτικές μεθόδους για την αξιοποίηση της διαθέσιμης πληροφορίας, αλλά και την αποτίμηση της ακρίβειας των παραγόμενων εκτιμήσεων. Παράλληλα, ασχολείται με την ερμηνεία και τη βέλτιστη πρακτική αξιοποίηση των προβλέψεων για την λήψη αποτελεσματικών αποφάσεων προς την κατεύθυνση της επίτευξης στρατηγικών στόχων.

Οι τεχνικές προβλέψεων βρίσκουν εφαρμογή σε ευρύ πλήθος διεπιστημονικών πεδίων και περιστάσεων της ανθρώπινης δραστηριότητας. Αναλυτικότερα, αποτελούν ένα πολυδύναμο εργαλείο των επιχειρηματιών για την πρόβλεψη κέρδους, κόστους, αποθήκευσης, ζήτησης, κλπ. Οικονομικοί αναλυτές και επενδυτές βασίζονται σε αυτές για τη μεγιστοποίηση του κέρδους και την ελαχιστοποίηση του κινδύνου σε οικονομικές και χρηματοοικονομικές εφαρμογές, μέσω της πρόβλεψης τιμών των οικονομικών τιμών, των τιμών μετοχών κλπ. Στον τομέα του τουρισμού, βασικού προϊόντος της χώρας μας, επιτρέπουν την εκτίμηση των συνολικών αφίξεων τουριστών, τις διανυκτερεύσεις ανά προορισμό και άλλα σχετικά μεγέθη. Προβλέψεις κυκλοφοριακού φορτίου, καταληψιμότητας και τροχαίων ατυχημάτων διευκολύνουν τους συγκοινωνιολόγους και τους υπεύθυνους των αντίστοιχων υπουργείων στον συγκοινωνιακό και οδικό σχεδιασμό. Ακαδημαϊκοί, κρατικοί αρμόδιοι και ανεξάρτητες οργανώσεις τις χρησιμοποιούν στη μελέτη της εξέλιξης των δημογραφικών μεγεθών, της κατανομής του πλούτου, της διάδοσης των ασθενειών, των επιπέδων εγκληματικότητας και άλλων κοινωνικών παραμέτρων που επηρεάζουν άμεσα την ποιότητα της ανθρώπινης ζωής. Ολοένα σημαντικότερη διαφαίνεται η σημασία τους σε θέματα σχετικά με το περιβάλλον και τη διαχείριση της ενέργειας, όπως η πρόβλεψη ενεργειακής ζήτησης, η μετεωρολογία και η ρύπανση. Τα παραπάνω σαφώς δεν αποτελούν μία διεξοδική απαρίθμηση των διαφόρων τομέων εφαρμογής των προβλέψεων, ιδιαίτερα αφού η επιστημονική εξέλιξη και η διεπιστημονική εξέταση των ζητημάτων δημιουργεί διαρκώς νέους τομείς εφαρμογής.

Νέες εφαρμογές και ανάγκη βελτίωσης της ποιότητας των εκτιμήσεων των παραδοσιακών τεχνικών προβλέψεων καθιστούν επιτακτική την ανάγκη της συνεχούς και καινοτόμου έρευνας στον τομέα των προβλέψεων. Αναλυτικότερα, ορισμένα μόνο από τα αντικείμενα μελέτης είναι η ανάπτυξη νέων μεθόδων και

μεθοδολογικών πλαισίων προβλέψεων για την εξυπηρέτηση σύγχρονων αναγκών και την βελτίωση της ακρίβειας των προβλέψεων, η συγκριτική μελέτη της επίδοσης των διαφόρων μεθόδων πρόβλεψης σε διαφορετικά σύνολα δεδομένων και ο έλεγχος και η αναθεώρηση αντιλήψεων και πρακτικών.

Στα πλαίσια της μελέτης αυτής εξετάζεται η επέκταση της χρήση του μεθοδολογικού πλαισίου προβλέψεων χρονοσειρών διακοπτόμενης ζήτησης ADIDA σε χρονοσειρές συνεχούς ζήτησης. Η μέθοδος και διάφορες παραλλαγές της αναλύονται θεωρητικά και ελέγχεται η επίδοσή τους σε πραγματικά δεδομένα, προερχόμενα από τον ακαδημαϊκό διαγωνισμό προβλέψεων M3. Τα αποτελέσματα αντιπαραβάλλονται με τα αντίστοιχα από άλλες μεθόδους, ενώ τίγονται ζητήματα βελτιστοποίησης του προτεινόμενου πλαισίου, μέσω καταλληλότερης επιλογής των παραμέτρων του.

2.2 Χαρακτηριστικά των Χρονοσειρών

Οι χρονοσειρές αποτελούνται από σύνολα διαδοχικών παρατηρήσεων που προκύπτουν από την καταγραφή, τυπικά κατά ίσα διαστήματα, της τιμής κάποιου μεταβαλλόμενου φυσικού ή άλλου μεγέθους. Βασικός σκοπός της επιστήμης των προβλέψεων είναι η εξαγωγή χρήσιμων συμπερασμάτων και εκτιμήσεων για την εξέλιξη του μετρούμενου μεγέθους σε μελλοντικές στιγμές, καθώς και η εκτίμηση του διαστήματος εμπιστοσύνης των προβλέψεων. Κατά την παραδοσιακή ανάλυση των χρονοσειρών, αναγνωρίζονται ορισμένα ποιοτικά και ποσοτικά χαρακτηριστικά, από τα οποία εξαρτάται η επιλογή της κατάλληλης μεθόδου πρόβλεψης. Μία ευρέως χρησιμοποιούμενη μεθοδολογία για την παραγωγή προβλέψεων είναι το μοντέλο των χρονοσειρών, κατά το οποίο η αρχική χρονοσειρά υφίσταται μία επεξεργασία μέσω στατιστικών και μαθηματικών μεθόδων ώστε να παραχθεί ένα μοντέλο πρόβλεψης, ικανό να επεκταθεί στο μέλλον. Το μοντέλο των χρονοσειρών προϋποθέτει ότι οι ιστορικές τιμές της χρονοσειράς εμπεριέχουν πληροφορία για την εξέλιξή της στο μέλλον. Η προϋπόθεση αυτή δεν ικανοποιείται απολύτως από τις πραγματικές χρονοσειρές λόγω της τυχαιότητας και της εξάρτησής τους από εξωτερικούς παράγοντες.

Υπάρχουν πολλά μεθοδολογικά πλαίσια επεξεργασίας και πρόβλεψης των χρονοσειρών. Ειδικότερα, θα αναφερθούμε στην αντιμετώπιση των χρονοσειρών ως το αποτέλεσμα της σύνθεσης διακριτών συνιστωσών, η κάθε μία από τις οποίες αντιστοιχεί σε διαφορετικά ποιοτικά χαρακτηριστικά της αρχικής χρονοσειράς. Παραδοσιακά αναγνωρίζονται οι ακόλουθες συνιστώσες: η τάση, η εποχικότητα, η κυκλικότητα και η τυχαιότητα.

Η τάση (trend – T) ορίζεται διαισθητικά ως μία μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς προς κάποια κατεύθυνση. Ανάλογα

με το αν οι τιμές της χρονοσειράς αυξάνονται ή μειώνονται μακροπρόθεσμα, κάνουμε, αντίστοιχα, λόγο για αύξουσα ή φθίνουσα τάση. Επιπλέον, μπορούμε να διακρίνουμε μεταξύ γραμμικών, εκθετικών, λογαριθμικών και γενικότερα μη γραμμικών τάσεων. Καθώς απαιτείται γενικά ένας επαρκώς μεγάλος αριθμός παρατηρήσεων, χρειάζεται ιδιαίτερη προσοχή στην ταυτοποίηση της τάσης, αφού μπορεί να αποτελεί, για παράδειγμα, κύκλο με μακρύ διάστημα μεταβολής. Η αναγνώριση και απομόνωση της συνιστώσας της τάσης εξαρτάται άμεσα από την έννοια που προσδίδουμε στην «μακροπρόθεσμη» μεταβολή και τις ειδικότερες απαιτήσεις.

Η εποχικότητα (seasonality – S) αντιπροσωπεύει μία περιοδική διακύμανση των τιμών της χρονοσειράς με σταθερή περίοδο, τυπικά μικρότερη του έτους. Οφείλεται συνήθως σε συγκεκριμένους και άμεσα κατανοητούς παράγοντες (π.χ. αύξηση των πωλήσεων των παγωτών το καλοκαίρι) που επηρεάζουν με σταθερό τρόπο την πορεία της χρονοσειράς. Για το λόγο αυτό, η συνιστώσα της εποχικότητας είναι εύκολο γενικά να αναγνωριστεί και να απομονωθεί, ενώ η παρουσιάζει ιδιαίτερη ευχρηστία στον χειρισμό και την πρόβλεψή της, λόγω της κανονικότητάς της.

Η κυκλικότητα (cycle – C) αποτελεί μία κυματοειδή περιοδική ή ημι-περιοδική μεταβολή του επιπέδου της χρονοσειράς, με μήκος κατά κανόνα μεγαλύτερο του έτους. Η κυκλικότητα οφείλεται σε ειδικούς εξωγενείς παράγοντες και μπορεί να απαιτείται η συλλογή ενός πολύ μεγάλου όγκου δεδομένων για την ασφαλή εξακρίβωσή της. Τυπικό παράδειγμα κύκλου αποτελεί η συμπεριφορά πλήθους οικονομικών δεικτών και λοιπών μεγεθών που προκαλείται από την εναλλαγή μεταξύ διαδοχικών περιόδων άνθησης και ύφεσης (επιχειρηματικός κύκλος). Λόγω του μεγάλου μήκους του, αποτελεί αντικείμενο ενδιαφέροντος κυρίως των πολύ μακροπρόθεσμων προβλέψεων, ενώ η επίδρασή του εξασθενεί στους βραχυπρόθεσμους ορίζοντες.

Η τυχαιότητα (randomness – R) αντιστοιχεί στην εναπομείνασα συνιστώσα, εάν αφαιρεθούν από τη χρονοσειρά όλα τα υπόλοιπα συστατικά στοιχεία (τάση, εποχικότητα, κυκλικότητα). Αποτελεί τον εξ ορισμού απροσδόκητο παράγοντα, ο οποίος τυπικά δεν μπορεί να προβλεφθεί βάσει της ιστορικότητας των παρατηρήσεων. Μπορεί να αντιπροσωπεύει μία εντελώς τυχαία μεταβλητή ή να εκφράζει μία ασυνέχεια που σχετίζεται με κάποιο εξαιρετικό γεγονός. Τέτοιες ασυνέχειες είναι οι ακραίες παρατηρήσεις (outliers) και οι αλλαγές επιπέδου (level-shifts).

Οι συνιστώσες μπορεί να συνδυάζονται για την παραγωγή της αρχικής χρονοσειράς είτε προσθετικά είτε πολλαπλασιαστικά. Το πολλαπλασιαστικό μοντέλο αποτελεί μία συχνή παραδοχή, καθώς φαίνεται αποτελεσματικά να

περιγράφει ένα ευρύ πλήθος πραγματικών μεταβλητών. Ταυτόχρονα, υπάρχουν και υβριδικά μοντέλα που αποτελούν συνδυαστικές και εξειδικευμένες προσεγγίσεις.

Η αναγνώριση και απομόνωση των συνιστωσών μίας χρονοσειράς γίνεται βάσει κάποιας διαδικασίας αποσύνθεσης. Η πιο δημοφιλής και στην πράξη συχνά χρησιμοποιούμενη διαδικασία διαχωρισμού των συνιστωσών είναι η κλασική μέθοδος της αποσύνθεσης, η οποία μπορεί να αναζητηθεί σε οποιοδήποτε γενικό σύγγραμμα προβλέψεων (π.χ. Makridakis, Wheelwright, and Hyndman, 1998). Η απομόνωση της κάθε συνιστώσας μπορεί να υποβοηθήσει ποικιλοτρόπως τη διαδικασία πρόβλεψης. Για παράδειγμα, η εξέλιξη κάθε συνιστώσας μπορεί να προβλεφθεί ανεξάρτητα και η τελική πρόβλεψη να προκύψει με ανασύνθεση των επιμέρους, μία διαδικασία που αναφέρεται ως πρόβλεψη μέσω αποσύνθεσης. Συχνά στην πράξη, από την αρχική σειρά απομονώνεται και αφαιρείται η συνιστώσα της εποχικότητας, καθώς αποτελεί το στοιχείο με την μεγαλύτερη κανονικότητα, οπότε προκύπτει η αποεποχικοποιημένη χρονοσειρά. Η πληροφορία της εποχικότητας αποθηκεύεται με τη μορφή δεικτών εποχικότητας, ίσους στο πλήθος με το πλήθος των περιόδων ανά έτος της χρονοσειράς. Από την αποεποχικοποιημένη χρονοσειρά δομείται ένα από μοντέλο πρόβλεψης, το οποίο προεκτείνεται για την παραγωγή αποεποχικοποιημένων προβλέψεων. Στη συνέχεια, οι προβλέψεις επανεποχικοποιούνται πολλαπλασιάζοντάς τις με τους αντίστοιχους δείκτες εποχικότητας, υπό την προϋπόθεση χρήσης του πολλαπλασιαστικού μοντέλου.

Μία σημαντική διάκριση που προέρχεται από τους τομείς της οικονομίας και της λογιστικής, αλλά μπορεί με φυσικό τρόπο να επεκταθεί και σε κάθε άλλη κατηγορία δεδομένων, είναι αυτή μεταξύ των μεταβλητών ροής και των μεταβλητών αποθέματος (Harrison, 2008). Συγκεκριμένα, μεταβλητή ροής (flow variable) καλείται κάθε μεταβλητή της οποίας η τιμή ορίζεται σχετικά με κάποιο χρονικό διάστημα, ενώ δεν μπορεί να αποκτήσει συγκεκριμένο νόημα σε κάποια μεμονωμένη χρονική στιγμή. Εξ ορισμού, λοιπόν, τέτοιου τύπου μεταβλητές ορίζονται συνήθως σε μονάδες ως προς το χρόνο. Η εξάρτηση αυτή μπορεί να δηλώνεται άμεσα ή να υπονοείται σιωπηλά. Τυπικά παραδείγματα αποτελούν το (ετήσιο) ακαθάριστο εθνικό προϊόν, οι πωλήσεις ενός προϊόντος σε κάποιο χρονικό διάστημα, κλπ. Αντιθέτως, μεταβλητή αποθέματος (stock variable) ονομάζεται μία μεταβλητή της οποίας η τιμή μπορεί να οριστεί σε κάθε μεμονωμένη χρονική στιγμή. Για παράδειγμα, τέτοιου τύπου μεταβλητές εμφανίζονται κατά κανόνα στη διαχείριση των αποθεμάτων σε αποθήκες, εξ ου προέρχεται και η ονομασία «μεταβλητές αποθέματος». Άλλες συνηθισμένες περιπτώσεις αποτελούν οι τιμές των μετοχών, η ατμοσφαιρική θερμοκρασία, κλπ.

Χρονοσειρές στις οποίες εμφανίζονται περίοδοι μηδενικών τιμών ονομάζονται σειρές διακοπτόμενης ζήτησης. Η διακοπτόμενη ζήτηση αναφέρεται

ακριβώς στην ιδιότητα αυτή, να μεσολαβούν δηλαδή περίοδοι μηδενικής ζήτησης, όχι υποχρεωτικά σταθερού μήκους ή ισοκατανεμημένες. Σε τέτοιου είδους χρονοσειρές ενδιαφέρει όχι μόνο το ύψος της ζήτησης αλλά και το πότε αυτή εμφανίζεται. Η διακύμανση των μη μηδενικών παρατηρήσεων και η διακύμανση των μηδενικών μεσοδιαστημάτων συνδυάζονται και οδηγούν γενικά σε χρονοσειρές υψηλότερης διακύμανσης και, συνεπώς, σε μεγαλύτερη δυσκολία κατά την πρόβλεψη. Χρονοσειρές διακοπτόμενης ζήτησης προκύπτουν τυπικά από τις μετρήσεις μεταβλητών αποθέματος σε μονάδες διαχείρισης αποθεμάτων (stock keeping units – SKU) και ανταλλακτικών (spare parts).

Ένα σημαντικό χαρακτηριστικό των χρονοσειρών διακοπτόμενης ζήτησης είναι η ασυνέχεια των παρατηρήσεων (intermittency), η οποία αναφέρεται στο πόσο συχνά ή σπανιότερα εμφανίζονται μηδενικές παρατηρήσεις. Ένα σημαντικό μέτρο που επιτρέπει την ποσοτικοποίηση της ασυνέχειας αυτής είναι η μέση τιμή του διαστήματος μεταξύ διαδοχικών ζητήσεων (Intermittent Demand Interval –IDI), η οποία μπορεί να λάβει τιμές μεγαλύτερες ή ίσες τις μονάδας. Προφανώς, η οριακή περίπτωση που ισχύει $IDI = 1$ εμφανίζεται όταν η σειρά δεν περιέχει καμία μηδενική παρατήρηση, οπότε αναφερόμαστε σε σειρά συνεχούς ζήτησης.

2.3 Μοντέλα Προβλέψεων

Οι μέθοδοι πρόβλεψης μπορεί να είναι είτε στατιστικές είτε κριτικές και έχουν ως κοινό σκοπό την επέκταση των διαθέσιμων δεδομένων μιας μεταβλητής στο μέλλον, ώστε να αποκτήσουμε εκτιμήσεις των μελλοντικών τιμών της.

Από τις δύο κατηγορίες, η πρώτη παρουσιάζει ιδιαίτερο μαθηματικό ενδιαφέρον. Οι στατιστικές μέθοδοι πρόβλεψης (Statistical Forecasting Methods) χρησιμοποιούν μαθηματικές και στατιστικές τεχνικές για την μοντελοποίηση και προέκταση των δεδομένων. Δύο είναι οι κυρίαρχες προσεγγίσεις των στατιστικών μεθόδων πρόβλεψης:

- Το μοντέλο χρονοσειρών (time series model), σύμφωνα με το οποίο δεχόμαστε την βασική υπόθεση ότι η τιμή της μεταβλητής μεταβάλλεται σύμφωνα με ένα λανθάνον σταθερό πρότυπο, το οποίο επαναλαμβάνεται στο χρόνο. Η πρόβλεψη γίνεται εφικτή με την αναγνώριση, συνήθως με στατιστικό τρόπο, και την προέκταση του εν λόγω προτύπου. Φυσικά, η προσέγγιση αυτή προϋποθέτει την διαθεσιμότητα ιστορικών δεδομένων της τιμής τους υπό πρόβλεψη μεγέθους.
- Το αιτιοκρατικό ή επεξηγηματικό μοντέλο (casual relationship or explanatory model), στο πλαίσιο του οποίου υποθέτουμε ότι το υπό

πρόβλεψη μέγεθος αποτελεί μία εξαρτημένη μεταβλητή που συνδέεται μέσω μιας σταθερής σχέσης με ορισμένες άλλες παραμέτρους που το επηρεάζουν (ανεξάρτητες μεταβλητές). Η πρόβλεψη υλοποιείται προσδιορίζοντας τις τιμές των ανεξάρτητων μεταβλητών (π.χ. από μετρήσεις, εκτιμήσεις ή προβλέψεις) και εφαρμόζοντας την εν λόγω σχέση.

Ακολουθεί μία συνοπτική περιγραφή κάποιων από τις περισσότερο διαδεδομένες μεθόδους, με ιδιαίτερη έμφαση στις μεθόδους εκθετικής εξομάλυνσης, καθώς το μεγαλύτερο μέρος της μελέτης που θα ακολουθήσει κάνει χρήση μεθόδων από την κατηγορία αυτή:

- Οι μέθοδοι εκθετικής εξομάλυνσης είναι μέθοδοι που αναπτύχθηκαν στις αρχές της δεκαετίας του 1950, ακολουθούν την προσέγγιση του μοντέλου των χρονοσειρών και έχουν αναδειχθεί ανάμεσα στις πιο δημοφιλείς μεθόδους για επιχειρηματική εφαρμογή. Τυπικά παραδείγματα συνιστούν οι παρακάτω μέθοδοι:

- Η απλοϊκή ή αφελής μέθοδος (Naïve) αποτελεί αδιαμφισβήτητα την απλούστερη στατιστική μέθοδο, παρατήρηση από την οποία προκύπτει και η ονομασία της. Η κάθε νέα τιμή του μοντέλου πρόβλεψης ισούται με την ακριβώς προηγούμενη διαθέσιμη πραγματική παρατήρηση.

$$F_t = Y_{t-1}$$

Το μοντέλο παράγει επίπεδες προβλέψεις, αφού όλες οι παραγόμενες προβλέψεις ισούνται ακριβώς με την τελευταία διαθέσιμη πραγματική παρατήρηση. Διακρίνονται δύο παραλλαγές της μεθόδου, η Naïve1 και η Naïve2, ανάλογα με το αν η πρόβλεψη εφαρμόζεται στην εποχικοποιημένη ή την αποεποχικοποιημένη χρονοσειρά, αντίστοιχα.

- Η απλή εκθετική εξομάλυνση σταθερού επιπέδου (Single Exponential Smoothing – SES), η οποία χρησιμοποιεί μία αρχική εκτίμηση του επιπέδου της χρονοσειράς S_t , την οποία διορθώνει μέσω ανάδρασης του σφάλματος e_t . Κάθε νέα τιμή του μοντέλου πρόβλεψης ισούται με την ακριβώς προηγούμενη τιμή του επιπέδου της χρονοσειράς.

$$\begin{aligned}e_t &= Y_t - F_t \\S_t &= S_{t-1} + ae_t \\F_{t+1} &= S_t\end{aligned}$$

ή, ισοδύναμα,

$$F_{t+1} = aY_t + (1 - a)F_t$$

Δεν υπάρχει μηχανισμός ανανέωσης των προβλέψεων, άρα οι προβλέψεις που παράγει το μοντέλο είναι επίσης επίπεδες, όλες ίσες με την τελευταία τιμή του επιπέδου της χρονοσειράς.

Η παράμετρος a που εμφανίζεται στον τύπο ανανέωσης του επιπέδου ονομάζεται συντελεστής εξομάλυνσης του επιπέδου και λαμβάνει τιμές στο διάστημα $[0,1]$. Η τιμή του a επηρεάζει την ευαισθησία του μοντέλου στην μεταβολή των τιμών της χρονοσειράς. Υψηλότερες τιμές επιφέρουν ταχύτερες αλλαγές στο μοντέλο, δίνοντας μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις. Επομένως, η ύπαρξη ισχυρού θορύβου επιβάλλει την επιλογή χαμηλής τιμής για το a , ώστε το μοντέλο να αντιστέκεται στις τυχαίες μεταβολές. Ωστόσο, προσοχή πρέπει να δοθεί ώστε να είναι ταυτόχρονα αρκετά υψηλός ώστε να μπορεί το μοντέλο πρόβλεψης να παρακολουθήσει τυχόν αλλαγές του επιπέδου της χρονοσειράς, αλλιώς υπάρχει ο κίνδυνος της χρονικής υστέρησης του μοντέλου έναντι της χρονοσειράς. Για την ακραία τιμή $a = 1$ είναι εμφανές ότι η μέθοδος ταυτίζεται με τη Naïve, ενώ για $a = 0$ η πρόβλεψη είναι ίση με το αρχικό επίπεδο S_0 . Συνήθης πρακτική είναι ο συντελεστής εξομάλυνσης a να επιλέγεται ώστε να ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (MSE) του μοντέλου πρόβλεψης. Καθώς δεν υπάρχει λύση του παραπάνω προβλήματος ελαχιστοποίησης σε κλειστή μορφή, είναι ανάγκη η βελτιστοποίηση να γίνει αριθμητικά. Ο απλούστερος τρόπος είναι με την διαδοχική εφαρμογή μοντέλων με γραμμικά αυξανόμενο a στο διάστημα $[0,1]$ με βήμα που προσδιορίζει την ακρίβεια της αναζήτησης και την τελική επιλογή εκείνου που επιστρέφει το μικρότερο MSE.

Για την εκτίμηση του αρχικού επιπέδου χρησιμοποιείται συνήθως το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης ή ο μέσος όρος των παρατηρήσεων από την αρχή μέχρι κάποιο σημείο της χρονοσειράς. Ακραίες περιπτώσεις αποτελούν η επιλογή της πρώτης παρατήρησης ή του μέσου όρου του συνόλου των παρατηρήσεων. Η κατάλληλη επιλογή καθορίζεται από τα ποιοτικά χαρακτηριστικά της χρονοσειράς.

➤ Η εκθετική εξομάλυνση γραμμικής τάσης (Holt Exponential Smoothing), η οποία επεκτείνει το μοντέλο της SES, θεωρώντας επιπλέον την ύπαρξη μίας γραμμικής τάσης T_t στη χρονοσειρά. Οι προβλέψεις παύουν να

είναι επίπεδες και ανανεώνονται με την τελευταία υπολογισθείσα τιμή της τάσης.

$$\begin{aligned} e_t &= Y_t - F_t \\ S_t &= S_{t-1} + T_{t-1} + \alpha e_t \\ T_t &= T_{t-1} + \beta e_t \\ F_{t+m} &= S_t + mT_t \end{aligned}$$

Πέραν του συντελεστή εξομάλυνσης του επιπέδου α , για τον οποίο ισχύουν ακριβώς τα ίδια με την περίπτωση της SES, η μέθοδος εισάγει επιπλέον το συντελεστή εξομάλυνσης της τάσης β . Λαμβάνει επίσης τιμές στο διάστημα $[0,1]$ και εκφράζει την ευκολία (υψηλές τιμές) ή τη δυσκολία (χαμηλές τιμές) με την οποία μεταβάλλεται η γραμμική τάση της χρονοσειράς. Για την επιλογή των βέλτιστων α και β μπορεί να εφαρμοστεί παρόμοια προσέγγιση με πριν: ο διδιάστατος χώρος που ορίζεται από τα α και β πλεγματοποιείται, για κάθε σημείο υπολογίζεται το σφάλμα MSE της μεθόδου και επιλέγεται τελικά ο συνδυασμός των α και β που επιστρέφουν τη χαμηλότερη του τιμή.

Το αρχικό επίπεδο προσδιορίζεται από κάποια από τις ήδη αναφερθείσες επιλογές, ενώ η αρχική τάση μπορεί να ληφθεί είτε από την κλίση του μοντέλου γραμμικής παλινδρόμησης είτε από την διαφορά της ν -στης και πρώτης παρατήρησης διαιρεμένης με $\nu - 1$.

➤ Η εκθετική εξομάλυνση μη γραμμικής τάσης (Gardner, and McKenzie, 1985) που αποτελεί γενίκευση της Holt για μη γραμμική τάση. Οι προβλέψεις ανανεώνονται με την τάση πολλαπλασιασμένη επί έναν συντελεστή φ .

$$\begin{aligned} e_t &= Y_t - F_t \\ S_t &= S_{t-1} + T_{t-1} + \alpha e_t \\ T_t &= T_{t-1} + \beta e_t \\ F_{t+m} &= S_t + \sum_{i=1}^m \varphi^i T_t \end{aligned}$$

Η παράμετρος φ ονομάζεται παράμετρος διόρθωσης της τάσης και λαμβάνει μη αρνητικές τιμές. Για διάφορες τιμές της, το μοντέλο παρουσιάζει διαφορετικές ιδιότητες:

✓ Για $\varphi = 0$, ταυτίζεται με τη απλή εκθετική εξομάλυνση.

- ✓ Για $0 < \varphi < 1$, ορίζει το μοντέλο εξομάλυνσης φθίνουσας τάσης (Damped Exponential Smoothing). Αποτελεί συχνά ένα περισσότερο ρεαλιστικό μοντέλο από το αντίστοιχο της γραμμικής τάσης, αφού δεν επιτρέπει στις προβλέψεις να αυξάνουν ασταμάτητα, αλλά κάνει την τάση να «σβήνει» με την πάροδο του χρόνου.
- ✓ Για $\varphi = 1$, το μοντέλο ταυτίζεται με τη γραμμική εκθετική εξομάλυνση.
- ✓ Για $\varphi > 1$, προκύπτει το μοντέλο εκθετικής τάσης. Στις περισσότερες περιπτώσεις το μοντέλο δίνει υπεραισιόδοξες προβλέψεις που αποκλίνουν ραγδαία, για το λόγο αυτό χρησιμοποιείται σπάνια, συνήθως για να μοντελοποιηθεί η έναρξη κάποιας διαδικασίας, όπως η εισχώρηση στην αγορά ενός νέου προϊόντος.

Η επιλογή των παραμέτρων εξομάλυνσης γίνεται στην περίπτωση αυτή με βελτιστοποίηση σε έναν τρισδιάστατο πλεγματοποιημένο χώρο, όπου δεν εξετάζονται τιμές του φ μεγαλύτερες τις μονάδας, ώστε να μην ληφθούν υπόψη τα εκθετικά μοντέλα. Η βελτιστοποίηση με τρεις παραμέτρους, λόγω της συνδυαστικής έκρηξης, είναι περισσότερο υπολογιστικά απαιτητική και χρονοβόρα.

- Το μοντέλα παλινδρόμησης (regression models) συνιστούν τη βασική υλοποίηση της αιτιοκρατικής προσέγγισης. Στη μέθοδο της γραμμικής παλινδρόμησης αναζητείται μία γραμμική σχέση που εκφράζει την εξαρτημένη μεταβλητή (υπό πρόβλεψη μέγεθος) με τις εξαρτημένες μεταβλητές. Περιπτώσεις όπου τα μεγέθη σχετίζονται με μη γραμμική σχέση αντιμετωπίζονται με κατάλληλους μετασχηματισμούς κάποιων από αυτά, όπως η λογαρίθμιση, ώστε οι προκύπτουσες ποσότητες να υπακούουν σε μία γραμμική σχέση. Αν στο μοντέλο υπεισέρχεται μία μόνο εξαρτημένη μεταβλητή, κάνουμε λόγο για απλή γραμμική παλινδρόμηση, ειδάλλως (αν δηλαδή χρησιμοποιούνται παραπάνω της μίας μεταβλητές) μιλάμε για πολλαπλή γραμμική παλινδρόμηση. Το πλήθος των μεταβλητών που χρησιμοποιούνται στο μοντέλο καθορίζεται βάσει του βαθμού συσχέτισης που παρουσιάζουν οι υπό θεώρηση ανεξάρτητες μεταβλητές με την εξαρτημένη ποσότητα, καθώς και το κατά πόσον η συμπερίληψη μίας επιπλέον ανεξάρτητης μεταβλητής οδηγεί σε ακριβέστερο μοντέλο, δηλαδή εξυπηρετεί στην ερμηνεία μεγαλύτερου μέρους της συνολικής διακύμανσης των δεδομένων ή, ισοδύναμα, οδηγεί σε μοντέλο με μικρότερα σφάλματα. Ο προσδιορισμός των συντελεστών με τους οποίους συμμετέχει κάθε ανεξάρτητη μεταβλητή στο μοντέλο καθορίζεται σύμφωνα με την αρχή των ελαχίστων τετραγώνων. Η απλούστερη

περίπτωση μοντέλου παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση όπου ο χρόνος χρησιμοποιείται ως μοναδική ανεξάρτητη μεταβλητή.

- Μία εναλλακτική προσέγγιση της παλινδρόμησης είναι μέσω της χρήσης νευρωνικών δικτύων (Neural Networks) και τεχνικών υπολογιστικής νοημοσύνης, οπότε εξεζητημένοι αλγόριθμοι υπολογιστικής νοημοσύνης χρησιμοποιούνται απευθείας για την παραγωγή προβλέψεων. Ως είσοδο ενός μοντέλου νευρωνικών δικτύων χρησιμοποιούνται και πάλι κάποιες ανεξάρτητες μεταβλητές ή χαρακτηριστικά εξαγμένα από αυτές, ωστόσο το μοντέλο καλύπτει και περιπτώσεις μη γραμμικών συσχετίσεων με την εξαρτημένη μεταβλητή. Εναλλακτικά, τέτοιες τεχνικές μπορούν να χρησιμοποιηθούν βοηθητικά στην διαδικασία της πρόβλεψης, όπως, παραδείγματος χάριν για την επιλογή των παραμέτρων κάποιας στατιστικής παραμετρικής μεθόδου πρόβλεψης.

- Τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων (Auto-Regressive Integrated Moving Average – ARIMA) (Box, Jenkins, and Reinsel, 1994) είναι στοχαστικά μαθηματικά μοντέλα που εμπλέκουν γραμμικούς συνδυασμούς κυρίως των προηγούμενων εμφανισθεισών τιμών της χρονοσειράς και του τυχαίου παράγοντα (τυχαίο σφάλμα ή σφάλμα πρόβλεψης).

- Συχνά είναι αποδοτικός ο συνδυασμός μεθόδων πρόβλεψης όπου οι προβλέψεις διαφορετικών μεθόδων μπορούν να συνδυάζονται γραμμικά με κατάλληλα βάρη, ώστε να προκύπτουν νέες προβλέψεις που χάρη στα συνδυασμένα χαρακτηριστικά των επιμέρους μεθόδων είναι περισσότερο ακριβείς και σθεναρές.

Μία νέα και πολλά υποσχόμενη μέθοδος πρόβλεψης (έδωσε τα ακριβέστερα αποτελέσματα στον διαγωνισμό πρόβλεψης M3) είναι η μέθοδος Theta. Πρόκειται για μια μεθοδολογία που βασίζεται στην τροποποίηση των τοπικών καμπυλοτήτων της χρονοσειράς (Assimakopoulos and Nikolopoulos, 2000). Αποτελεί εφαρμογή της πρόβλεψης μέσω αποσύνθεσης και του συνδυασμού των προβλέψεων, αφού η αρχική χρονοσειρά αποσυντίθεται σε ένα πλήθος γραμμών Theta (theta lines) οι οποίες προεκτείνονται ανεξάρτητα και στη συνέχεια οι προβλέψεις συνδυάζονται με κατάλληλα βάρη.

- Στον αντίποδα των στατιστικών μεθόδων πρόβλεψης βρίσκονται οι κριτικές μέθοδοι πρόβλεψης (Judgmental Forecasting Methods), οι οποίες είναι μη μαθηματικές μέθοδοι που αξιοποιούν την ανθρώπινη διαίσθηση, πληροφόρηση και κριτική ικανότητα. Ενδεχομένως να χρησιμοποιούν τα αποτελέσματα κάποιας μαθηματικής μεθόδου ως βάση τους.

2.4 Σφάλματα και Ακρίβεια των Προβλέψεων

Ο υπολογισμός στατιστικών δεικτών σφάλματος είναι απαραίτητος για την εξακρίβωση της ποιότητας των παραγόμενων προβλέψεων. Το σφάλμα μίας περιόδου ορίζεται ως η διαφορά μεταξύ πραγματικής τιμής και της αντίστοιχης πρόβλεψης:

$$e_t = Y_t - F_t$$

Κάνουμε λόγο για σφάλμα μοντέλου πρόβλεψης (in-sample error), αναφερόμενοι σε οποιοδήποτε σφάλμα χρησιμοποιεί αποκλειστικά τις εξ αρχής διαθέσιμες παρατηρήσεις που χρησιμοποιήθηκαν για την κατασκευή του μοντέλου πρόβλεψης. Τέτοια σφάλματα δείχνουν την προσαρμογή του χρησιμοποιούμενου μοντέλου στα διαθέσιμα δεδομένα και χρησιμοποιούνται συχνά για την επιλογή μεταξύ μοντέλων ή την επιλογή των παραμέτρων ενός μοντέλου πρόβλεψης. Από την άλλη, μιλάμε για πραγματικό σφάλμα προβλέψεων (out-of-sample error) όταν το σφάλμα υπολογίζεται συγκρίνοντας τις προβλέψεις με τις πραγματικές τιμές του προβλεπόμενου μεγέθους, όταν αυτές γίνουν διαθέσιμες. Αυτού του είδους τα σφάλματα χρησιμοποιούνται για την εκ του αποτελέσματος αξιολόγηση του μοντέλου. Αν παρατηρηθούν σημαντικές αποκλίσεις, μπορεί να χρειαστεί να αναθεωρηθεί το μοντέλο ή να ληφθούν άλλες διορθωτικές ενέργειες. Το παραπάνω σφάλμα μίας περιόδου αποτελεί βάση για περισσότερους δείκτες σφαλμάτων, που φέρουν εξειδικευμένες πληροφορίες.

Το μέσο τετραγωνικό σφάλμα (Mean Squared Error – MSE) υπολογίζεται ως ο μέσος όρος των τετραγώνων των σφαλμάτων. Χρησιμοποιείται ευρέως ως μέτρο σύγκρισης της εφαρμογής των μοντέλων πρόβλεψης διαφορετικών μεθόδων και ιδιαίτερα στην επιλογή των βέλτιστων παραμέτρων των μεθόδων εξομάλυνσης, οπότε ως βέλτιστες τιμές επιλέγονται εκείνες που το ελαχιστοποιούν.

$$MSE = \frac{1}{N} \sum_{t=1}^N e_t^2 = \frac{1}{N} \sum_{t=1}^N (Y_t - F_t)^2$$

Εξαιτίας του τετραγωνισμού των σφαλμάτων προσδίδεται αρκετά μεγαλύτερη βαρύτητα στα μεγαλύτερα σφάλματα, ενώ το MSE μετράται στις μονάδες της αρχικής χρονοσειράς υψωμένες στο τετράγωνο. Προκειμένου να εκφράσουμε το σφάλμα στις ίδιες μονάδες με αυτές της αρχικής χρονοσειράς, πρέπει να θεωρήσουμε τη ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error – RMSE), η οποία έχει κατά τα άλλα τις ίδιες ιδιότητες με το MSE και εκφράζει την ενεργό τιμή των σφαλμάτων:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N e_t^2} = \sqrt{\frac{1}{N} \sum_{t=1}^N (Y_t - F_t)^2}$$

Τα παραπάνω σφάλματα είναι μη σχετικά, με την έννοια ότι εξαρτώνται από το επίπεδο τιμών της εκάστοτε χρονοσειράς, με αποτέλεσμα να δυσχεραίνεται η άμεση σύγκριση της εφαρμογής μίας μεθόδου πρόβλεψης σε διαφορετικές χρονοσειρές. Σε τέτοιες περιπτώσεις είναι αναγκαίος ο υπολογισμός των σφαλμάτων σε καθαρά ποσοστιαία μορφή. Αυτό μπορεί να εξασφαλιστεί με τη χρήση του μέσου απόλυτου ποσοστιαίου σφάλματος (Mean Absolute Percentage Error – MAPE). Το MAPE αποτελεί εκφράζεται ως ποσοστό επί τοις εκατό και λαμβάνει μη φραγμένες προς τα επάνω μη αρνητικές τιμές. Επιπλέον, συμπεριφέρεται με καλό τρόπο ακόμα και όταν οι πραγματικές τιμές είναι ιδιαίτερα υψηλές.

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|Y_t - F_t|}{|Y_t|} 100\%$$

Παρατηρούμε ότι ο παραπάνω τύπος καταλήγει σε απροσδιόριστες μορφές (διαίρεση με το μηδέν) αν εφαρμοστεί σε σειρές διακοπτόμενης ζήτησης, λόγω των μηδενικών παρατηρήσεων. Μία τροποποίηση που διορθώνει τη συμπεριφορά αυτή είναι η χρήση του συμμετρικού μέσου απόλυτου ποσοστιαίου σφάλματος (Symmetric Mean Absolute Percentage Error – sMAPE), όπου κάθε όρος του αθροίσματος σταθμίζεται πλέον με το ημιάθροισμα της πραγματικής τιμής και της πρόβλεψης:

$$sMAPE = \frac{1}{N} \sum_{t=1}^N \frac{|Y_t - F_t|}{\left| \frac{F_t + Y_t}{2} \right|} 100\%$$

Ως αποτέλεσμα, το sMAPE είναι κάτω και άνω φραγμένο, λαμβάνοντας τιμές στο διάστημα [0%, 200%]. Παρά τον χαρακτηρισμό του ως συμμετρικό, ωστόσο, ο δείκτης αυτός έχει δειχθεί ότι μεταχειρίζεται ανόμοια τα αισιόδοξα και τα απαισιόδοξα σφάλματα.

2.5 Διαγωνισμοί Προβλέψεων

Οι διαγωνισμοί προβλέψεων είναι εμπειρικές συγκριτικές μελέτες στον τομέα των προβλέψεων. Στο πλαίσιο τους πραγματοποιείται η σύγκριση της απόδοσης ενός πλήθους διαφορετικών μεθόδων πρόβλεψης, μέσω της μαζικής εφαρμογής τους σε ένα σύνολο χρονοσειρών διαφορετικών χαρακτηριστικών. Η απόδοση ποσοτικοποιείται μέσω του υπολογισμού κάποιου δείκτη σφάλματος, κατάλληλου ώστε να έχει νόημα η αντιπαράθεση μεταξύ διαφορετικών χρονοσειρών και μεθόδων. Οι συμμετέχοντες σε έναν διαγωνισμό καλούνται να υποβάλλουν στους διοργανωτές του ένα ζητούμενο πλήθος προβλέψεων, το οποίο προκύπτει μέσω της υλοποίησης κάποιας μεθόδου πρόβλεψης. Η υποβολή των προβλέψεων από ειδικευμένους ερευνητές ενισχύει την αντικειμενικότητα των διαγωνισμών. Ανάλογα με την εμβέλεια του διαγωνισμού, εξετάζονται: απλές ή στατιστικά πολυπλοκότερες μέθοδοι πρόβλεψης · ερευνητικά καινοτόμες μέθοδοι, όπως νευρωνικά δίκτυα (neural systems) και έξυπνα συστήματα (expert systems) · εμπορικά πακέτα προβλέψεων · κριτικές προβλέψεις, κλπ. Σε κάθε περίπτωση, τα σφάλματα διαφόρων απλών ή κοινώς αποδεκτών μεθόδων χρησιμοποιούνται ως σημεία αναφοράς (benchmarks). Μετά τη δημοσίευση των τελικών αποτελεσμάτων, τα δεδομένα του διαγωνισμού γίνονται ελεύθερα διαθέσιμα στην ακαδημαϊκή κοινότητα, με σκοπό την προώθηση της επιστημονικής έρευνας, όπως, για παράδειγμα, για την αντιπαράθεση της απόδοσης μίας νέας μεθόδου πρόβλεψης με τα αποτελέσματα του διαγωνισμού.

Οι διαγωνισμοί προβλέψεων βοηθούν στην εμπειρική μελέτη των ιδιοτήτων των διαφόρων μεθόδων πρόβλεψης και παίζουν σημαντικό ρόλο στη διαμόρφωση στάσης απέναντι στις εξεταζόμενες μεθόδους, αλλά και την επιστήμη των προβλέψεων γενικότερα.

Μια από τις πρώτες μελέτες εμπειρικής σύγκρισης μεθόδων πρόβλεψης έγινε την δεκαετία του 1970 από τους Reid, Newbold και Granger, ωστόσο εξετάστηκε ένας περιορισμένος αριθμός μεθόδων.

Ακολούθησε από τους Μακρυδάκη και Hibon (1979) η προσπάθεια σύγκρισης μεγάλου αριθμού μεθόδων επί μεγάλου πλήθους χρονοσειρών (111) από διάφορες κατηγορίες (δεδομένα από εταιρίες, βιομηχανίες, μακροοικονομία). Η παρατήρηση ότι οι απλούστερες μέθοδοι εξομάλυνσης δίνουν συχνά καλύτερα αποτελέσματα από στατιστικά πολυπλοκότερες μεθόδους προκάλεσε αμφισβήτηση ως προς την αντικειμενικότητα της έρευνας αυτής.

Σε απάντηση στις αμφιβολίες, διοργανώθηκε ο M-Competition (Makridakis et al., 1982), ο οποίος αποτελούσε μία επανάληψη της προηγούμενης μελέτης με μεγαλύτερο πλήθος μεθόδων (15 με 9 επιπλέον παραλλαγές τους) επί μεγαλύτερου συνόλου χρονοσειρών (1001) από περισσότερες κατηγορίες. Μελετήθηκαν επίσης

περισσότερα κριτήρια απόδοσης. Η καθοριστική όμως διαφορά ήταν η παροχή των προβλέψεων από ανεξάρτητους ειδικευμένους ερευνητές, πράγμα που καθιστά τον M-Competition τον πρώτο διαγωνισμό προβλέψεων, σύμφωνα με την περιγραφή που δόθηκε προηγουμένως. Η συμμετοχή των ειδικευμένων ερευνητών εγγυήθηκε την αντικειμενικότητα του διαγωνισμού, τα συμπεράσματα του οποίου ήταν σε συμφωνία με την αντίστοιχη μελέτη του 1979.

Σε συμπλήρωση και επέκταση, ακολούθησε αργότερα ο M2-Competition (Makridakis et al., 1993), στα πλαίσια του οποίου εξετάστηκαν ιδιαίτερα οι κριτικές προβλέψεις πάνω σε χρονοσειρές προερχόμενες από επιχειρήσεις, η τιμές των οποίων γίνονταν διαθέσιμες σε μία βάση πραγματικού χρόνου (real-time basis).

Ο M3-Competition (Makridakis et al., 2000) είναι ο μεγαλύτερος διαγωνισμός πρόβλεψης που έχει διοργανωθεί μέχρι στιγμής. Αποτελεί επανάληψη μέρους και επέκταση των προηγούμενων διαγωνισμών M και M2. Το σύνολο των δεδομένων που εξετάστηκαν αποτελείται από 3003 διαφορετικές χρονοσειρές από 6 κατηγορίες διαφορετικών τύπων δεδομένων (micro, industry, macro, economics, demographics, other) και σε 4 διαφορετικά χρονικά διαστήματα παρατήρησης (ετήσιες, τριμηνιαίες, μηνιαίες, άλλες) (πίνακας 2.1). Όλες οι τιμές ήταν αυστηρά θετικές, ιδιότητα που χρειαζόταν να πληρούται και από τις υποβληθείσες προβλέψεις. Επιπλέον, απαραίτητη κρίθηκε η θέσπιση ενός ελάχιστου αριθμού παρατηρήσεων ανά κατηγορία διαφορετικού χρονικού διαστήματος παρατήρησης (πίνακας 2.2), ώστε να εξασφαλιστεί η ύπαρξη επαρκούς πλήθους παρατηρήσεων για την κατασκευή των μοντέλων πρόβλεψης. Τα δεδομένα των χρονοσειρών είναι διαθέσιμα για κάθε ερευνητική δραστηριότητα και ο διαγωνισμός αποτέλεσε σημείο αναφοράς πάρα πολλών επιστημονικών δημοσιεύσεων.

Συνέχεια των διαγωνισμών M, M2 και M3 αποτελεί ο επερχόμενος M4-Competition, όπου εξετάζονται παραπάνω από 10000 χρονοσειρές δεδομένων.

Παράλληλα με τη σειρά διαγωνισμών M έχουν γίνει και άλλες παρόμοιες μελέτες, όπως οι διαγωνισμοί NN3 (2006-2007), NN4 (2009) και NN5 (2008) που διοργανώθηκαν από τον S.Crone και περιλάμβαναν μελέτη κυρίως μεθόδων πρόβλεψης που χρησιμοποιούσαν νευρωνικά μοντέλα ή, γενικότερα, υπολογιστική νοημοσύνη.

Όπως είναι λοιπόν εμφανές, εξελίξεις στον τομέα των προβλέψεων, νέες μέθοδοι, τάσεις, προσεγγίσεις και απόψεις δημιουργούν διαρκώς την ανάγκη για νέους διαγωνισμούς.

		Χρονική Περίοδος				Σύνολο
		Ετήσιες	Τριμηνιαίες	Μηνιαίες	Άλλες	
Τύπος Δεδομένων	Micro	146	204	474	4	828
	Industry	102	83	334	-	519
	Macro	83	336	312	-	731
	Economics	58	76	145	29	308
	Demographic	245	57	111	-	413
	Other	11	-	52	141	204
	Σύνολο	645	756	1428	174	3003

Πίνακας 2.1

	Αριθμός παρατηρήσεων		
	Ελάχιστος	Μέσος όρος	Ζητούμενος αριθμός προβλέψεων
Ετήσιες	14	19	6
Τριμηνιαίες	16	44	8
Μηνιαίες	48	115	18
Άλλες	60	63	8

Πίνακας 2.2

ΚΕΦΑΛΑΙΟ 3: Περιγραφή και Θεωρητική Ανάλυση της Μεθοδολογίας ADIDA

3.1 Χρονική Συνάθροιση Δεδομένων

Μία βασική έννοια που θα μελετηθεί στα πλαίσια αυτής της εργασίας είναι εκείνη της χρονική συνάθροισης των δεδομένων (temporal aggregation), η οποία αναφέρεται στην παραγωγή μίας χρονοσειράς χαμηλής συχνότητας (π.χ. τριμηνιαία) από μία άλλη χρονοσειρά, η τιμές της οποίας παρατηρούνται σε υψηλότερη συχνότητα (π.χ. μηνιαία), μέσω μίας κατάλληλης διαδικασίας συνάθροισης των τιμών της. Η χρονική συνάθροιση μπορεί να εφαρμόζεται σκόπιμα στα δεδομένα, στα πλαίσια της επεξεργασίας τους, ή να προκύπτει φυσικά, κατά τη συλλογή των τιμών της χρονοσειράς. Το δεύτερο συμβαίνει συχνά κατά τη δειγματοληψία μεταβλητών ροής, όταν δεν είναι εφικτή μία λεπτότερη δειγματοληψία. Όπως θα αποδειχθεί και αργότερα, η χρονική συνάθροιση συνδέεται στενά με τις διαδικασίες δειγματοληψίας. Η εφαρμογή χρονικής συνάθροισης επί των δεδομένων εμπεριέχει πάντα τον κίνδυνο της απώλειας της πληροφορίας, αφού τα δεδομένα παρατηρούνται σε υψηλότερο επίπεδο. Οι παρατηρήσεις αυτές προβάλλουν την ανάγκη μελέτης της επίδρασης της χρονικής συνάθροισης.

Αρκετές μελέτες σχετικά με την χρονική συνάθροιση δεδομένων συνεχούς ζήτησης έχουν πραγματοποιηθεί. Στα γενικότερα πλαίσια της έρευνας της επίδρασης της χρονικής συνάθροισης στις ιδιότητες της χρονοσειράς, ενδιαφέρει ιδιαίτερα η περιγραφή τυχόν προβλημάτων που μπορούν να προκύψουν εξαιτίας της συνάθροισης, όπως η λανθασμένη αναγνώριση του επιπέδου της χρονοσειράς (Granger, and Siklos, 1995), ή τα ζητήματα μείωσης της ακρίβειας των προβλέψεων σε ορισμένα δομικά μοντέλα χρονοσειρών (González, 1992). Η επίδραση της συνάθροισης στην απόδοση των προβλέψεων έχει μελετηθεί θεωρητικά και εμπειρικά (Zellner, and Tobias, 2000). Αναλυτικότερα, θεωρώντας μοντέλα ζήτησης από τη γενικότερη κατηγορία των μοντέλων ARIMA αποτελέσματα έχουν εξαχθεί τόσο για επικαλυπτόμενη όσο και μη επικαλυπτόμενη συνάθροιση για μοντέλα ARIMA, ARIMAX (ARIMA με εξωγενείς/exogenous μεταβλητές) (Weiss, 1984), ARMA-GARCH (Auto-Regressive Moving Average – Generalized Auto-Regressive Conditional Heteroscedastic) και VARMA (Vector ARMA) (Marcellino, 1999). Επιπλέον, αποτελέσματα υπάρχουν και για μοντέλα ARFIMA (Auto-Regressive Fractionally Integrated Moving Averages) (Man, and Tiao, 2006; Tsai and Chan, 2005). Στην μελέτη των Silvestrini και Veredas (2008) μπορούν να βρεθούν συγκεντρωμένες οι πιο σύγχρονες εξελίξεις στη συνάθροιση των δεδομένων.

Η χρονική συνάθροιση παρουσιάζει πολλές ομοιότητες με τη συνάθροιση μεταξύ διαφορετικών χρονοσειρών. Διάφορες μελέτες έχουν ασχοληθεί με την αξιοποίηση πληροφορίας από ανώτερα επίπεδα συνάθροισης μέσω συνδυασμού προβλέψεων και έχουν δείξει την προοπτική τέτοιων τεχνικών (Andrawis, Atiya, and El-Shishiny, 2011; Dekker, M., van Donselaar, K., and Ouwehand, 2004). Έχει επίσης επισημαίνεται η σημασία του σαφούς προσδιορισμού ενός κατάλληλου επιπέδου συνάθροισης, το οποίο εξαρτάται από τις εκάστοτε εφαρμογή και τις απαιτήσεις και επηρεάζει δραστικά την αποδοτικότητα των διαδικασιών προβλέψεων (Zotteri, Kalchschmidt, and Caniato, 2005). Σε έρευνες σύγκρισης της απόδοσης απλών προβλεπτών συναθροισμένων χρονοσειρών έχει επίσης τεθεί το ερώτημα του κατάλληλου επιπέδου στο οποίο πρέπει να γίνει η πρόβλεψη (Koreisha, and Fang, 2004). Για το λόγο αυτό έχουν προταθεί κριτήρια επιλογής μεταξύ της διεξοδικής προσέγγισης και της συναθροισμένης προσέγγισης πρόβλεψης (Zotteri, and Kalchschmidt, 2007).

Σχετικά με την χρονική συνάθροιση δεδομένων διακοπτόμενης ζήτησης, ένας μικρός μονάχα αριθμός μελετών έχει πραγματοποιηθεί. Το ζήτημα θίγεται μερικώς μόνο και κατά εμπειρικό τρόπο από τους Willemain et al. (1994). Εξετάζοντας μονάχα τις μεθόδους SES και Croston, βρέθηκαν στοιχεία βελτίωσης της ακρίβειας πρόβλεψης (ειδικότερα, του MAPE) μέσω της χρονικής συνάθροισης ημερησίων δεδομένων διακοπτόμενης ζήτησης σε επίπεδο εβδομάδας. Ωστόσο, η πληρέστερη ανάλυση στον τομέα αυτό πραγματοποιήθηκε από τους Brännäs, Hellström, and Nordström (2002), οι οποίοι μελέτησαν την χρονική συνάθροιση ακέραιων αυτοπαλλινδρομικών διαδικασιών (Integer Auto-Regressive – INAR). Ειδικότερα, δείχθηκε ότι ενώ η επικαλυπτόμενη συνάθροιση διαδικασιών όπως το μοντέλο INAR(1) διατηρεί την μορφή του μοντέλου, η περίπτωση μη επικαλυπτόμενης συνάθροισης είναι πολυπλοκότερη. Για παράδειγμα, μη επικαλυπτόμενη συνάθροιση ενός μοντέλου INAR(1) δημιουργεί ένα μοντέλο παρόμοιο με INARMA(1,1), αλλά με συσχετισμένους όρους κινητού μέσου όρου. Τέλος, το ζήτημα της χρονικής συνάθροισης αναφέρεται διεξοδικά από τους Nikolopoulos et al. (2011), στην εμπειρική πρόταση και ανάλυση της προσέγγισης ADIDA.

Διαδικασία αντίστροφη της χρονικής συνάθροισης αποτελεί η διάσπαση (disaggregation) των συναθροισμένων δεδομένων, σύμφωνα με κάποιον αλγόριθμο διάσπασης, ώστε από μία χρονοσειρά χαμηλής συχνότητας (π.χ. τριμηνιαία) να προκύψει μία άλλη χρονοσειρά υψηλότερης συχνότητας (π.χ. μηνιαία). Τέτοιες μέθοδοι διάσπασης θα παρουσιαστούν αναλυτικότερα, αργότερα στην εργασία. Έτσι, τα αρχικά δεδομένα μπορούν να συναθροιστούν σε υψηλότερο επίπεδο, στη συνέχεια να πραγματοποιηθεί κάποιου είδους επεξεργασία (π.χ. προέκταση στο

μέλλον) και κατόπιν μέσω μίας διαδικασίας διάσπασης να επιστρέψουμε στο χαμηλότερο επίπεδο παρατήρησης.

3.2 Η Μεθοδολογία Προβλέψεων ADIDA

Στην παραπάνω λογική στηρίζεται η μεθοδολογία ADIDA (Aggregate-Disaggregate Intermittent Demand Approach/Προσέγγιση Συνάθροισης-Διάσπασης Διακοπτόμενης Ζήτησης) (Nikolopoulos et al, 2011), η οποία αποτελεί ένα πλαίσιο για την παραγωγή προβλέψεων για χρονοσειρές διακοπτόμενης ζήτησης. Ειδικότερα, η μεθοδολογία αυτή έκανε δυνατή την εφαρμογή στην περίπτωση χρονοσειρών διακοπτόμενης ζήτησης μεθόδων πρόβλεψης που τυπικά χρησιμοποιούνται σε σειρές συνεχούς ζήτησης, εμπλουτίζοντας την εργαλειοθήκη από τις διαθέσιμες μεθόδους που μπορούν να χρησιμοποιηθούν. Το πλαίσιο αυτό έδωσε ικανοποιητικά αποτελέσματα, μειώνοντας τις τιμές μέτρων σφάλματος όταν δοκιμάστηκε για την πρόβλεψη χρονοσειρών αποθεμάτων, και παρουσιάστηκε ως ένας μηχανισμός αυτοβελτίωσης των υπάρχοντων μεθόδων πρόβλεψης. Συνοπτικά, η μεθοδολογία μπορεί να αναλυθεί σε τέσσερα βασικά στάδια:

1. Συλλογή των αρχικών δεδομένων, στο κατώτερο επίπεδο παρατήρησης.
2. Συνάθροιση των δεδομένων σε κάποιο επίπεδο συνάθροισης L .
3. Εφαρμογή κάποιας μεθόδου F επέκτασης στο μέλλον (extrapolation) της συναθροισμένης χρονοσειράς για την παραγωγή προβλέψεων.
4. Διάσπαση των προβλέψεων στο αρχικό επίπεδο παρατήρησης, σύμφωνα με κάποιον αλγόριθμο διάσπασης D .

Αναλυτικότερα, το πρώτο βήμα αναφέρεται στην συγκέντρωση των δεδομένων που συνθέτουν τη χρονοσειρά που επιθυμούμε να προβλέψουμε. Στη συνέχεια, οι παρατηρήσεις διαιρούνται από το τέλος της χρονοσειράς προς την αρχή διαιρούνται σε διαδοχικά τμήματα (time buckets) μήκους L . Αθροίζοντας το περιεχόμενο καθενός από τα τμήματα αυτά, προκύπτει η αντίστοιχη τιμή της συναθροισμένης χρονοσειράς, στο υψηλότερο επίπεδο παρατήρησης. Αν το πλήθος των παρατηρήσεων της αρχικής χρονοσειράς δεν είναι πολλαπλάσιο του L , τότε οι εναπομείνουσες $N \bmod L$ αρχικές παρατηρήσεις δεν χρησιμοποιούνται. Κάνουμε λόγο για μη-επικαλυπτόμενη συνάθροιση, αφού κάθε τιμή της αρχικής χρονοσειράς αντιστοιχίζεται σε μοναδική χρονοθυρίδα. Κατά το τρίτο βήμα, εφαρμόζεται κάποια μέθοδος πρόβλεψης, οπότε προκύπτουν προβλέψεις σε ένα υψηλότερο επίπεδο παρατήρησης από το επιθυμητό. Αυτές ανάγονται κατά το τελικό στάδιο στο αρχικό επίπεδο παρατήρησης της χρονοσειράς, μέσω της διάσπασής τους, η οποία γίνεται

συνήθως με τη χρήση κατάλληλων βαρών. Μία ειδική περίπτωση, και η απλούστερη όλων, είναι αυτή της χρήσης ίσων βαρών, την οποία συμβολίζουμε ως EQW (Equal Weights). Τότε τα βάρη ισούνται προφανώς με $1/L$, ώστε το άθροισμά τους να είναι μονάδα. Αργότερα θα ερευνήσουμε και εναλλακτικά βάρη διάσπασης.

Η παραπάνω μεθοδολογία επιφέρει τη μείωση του intermittency των σειρών διακοπτόμενης ζήτησης κατά το στάδιο της συνάθροισης, ωστόσο δεν υπεισέρχεται πουθενά η προϋπόθεση οι σειρές να ανήκουν στην κατηγορία αυτή. Με άλλα λόγια, η ίδια μεθοδολογία μπορεί να εφαρμοστεί αυτούσια και σε δεδομένα συνεχούς ζήτησης και την αποτελεσματικότητα μίας τέτοιας πρακτικής καλούμαστε να ελέγξουμε στο πλαίσιο της παρούσας εργασίας.

Δεδομένης της εξάρτησης της συγκεκριμένης εφαρμογής της μεθόδου από το επιλεγμένο επίπεδο συνάθροισης L , τη μέθοδο πρόβλεψης F και τον αλγόριθμο διάσπασης D , συμβολίζουμε τη μέθοδο ως:

$$ADIDA(L, F, D)$$

Έτσι, π.χ., η ADIDA που μετασχηματίζει μηνιαία δεδομένα σε τριμηνιαία ($L = 3$), χρησιμοποιεί απλή εκθετική εξομάλυνση (SES) ως μέθοδο πρόβλεψης και τελικά διασπάει τις προβλέψεις με ίσα βάρη συμβολίζεται ως:

$$ADIDA(3, SES, EQW)$$

Παρατηρούμε ότι η μεθοδολογία ADIDA δεν συσχετίζεται αποκλειστικά ούτε με κάποια συγκεκριμένη μέθοδο πρόβλεψης F ούτε με κάποιο δεδομένο αλγόριθμο διάσπασης D . Παρομοίως, μπορεί να επιλεγεί οποιοδήποτε επιτρεπτό επίπεδο συνάθροισης L , απόφαση που επηρεάζει καθοριστικά τη συγκεκριμένη εφαρμογή της μεθοδολογίας. Επομένως, η ADIDA αποτελεί ένα γενικευμένο πλαίσιο ενός εναλλακτικού τρόπου για την παραγωγή προβλέψεων.

Ακόμα μία ενδιαφέρουσα παρατήρηση είναι πως οποιαδήποτε μέθοδος πρόβλεψης F μπορεί να αναχθεί σε μία ειδική περίπτωση της ADIDA με μοναδιαίο επίπεδο συνάθροισης, δηλαδή:

$$F \equiv ADIDA(1, F, -)$$

όπου η χρήση της παύλας στη θέση του αλγορίθμου διάσπασης υποδεικνύει πως δεν απαιτείτε διάσπαση των προβλέψεων.

Ένας εναλλακτικός τρόπος για να μοντελοποιήσουμε τη διαδικασία της ADIDA είναι μέσω της αξιοποίησης της παρατήρησης ότι η χρονική συνάθροιση μπορεί να αναλυθεί ως ένας απλός κινητός μέσος όρος ακολουθούμενος από μία

διαδικασία υποδειγματοληψίας, ενώ η διάσπαση αναλύεται ως υπερδειγματοληψία και κατόπιν εφαρμογή κινητού μέσου όρου, οπότε καταλήγουμε σε μία συστημική προσέγγιση.

Μπορούμε να εκφράσουμε την χρονική συνάθροιση σε επίπεδο L μέσω του τύπου:

$$y[n] = \sum_{k=0}^{L-1} x[nL + k]$$

Σύμφωνα με τον παραπάνω τύπο, αθροίζουμε τις L επόμενες παρατηρήσεις, δηλαδή η άθροιση είναι «εμπρός-άθροιση». Αυτό προϋποθέτει ότι οι επόμενες παρατηρήσεις είναι ήδη διαθέσιμες και κρατούνται κάπου αποθηκευμένες. Ο ίδιος τύπος μπορεί να διασπαστεί σε δύο επιμέρους στάδια υπολογισμού:

$$d[n] = \sum_{k=0}^{L-1} x[n + k]$$

$$y[n] = d[nL]$$

Πράγματι, αναγνωρίζουμε στα δύο παραπάνω στάδια τους τύπους του κινητού μέσου όρου μήκους L και της υποδειγματοληψίας κατά L , αντίστοιχα. Η πρώτη σχέση, που εκφράζει το άθροισμα L διαδοχικών παρατηρήσεων, μπορεί να ιδωθεί ως ένας κινητός μέσος όρος μήκους L και σταθερών βαρών $1/L$, πολλαπλασιασμένος με το μήκος L . Η υποδειγματοληψία πραγματοποιείται απλά κρατώντας κάθε L -στη παρατήρηση της σειράς d . Στη συνέχεια, κάποια μέθοδος πρόβλεψης χρησιμοποιείται ώστε από τη συναθροισμένη σειρά y να προκύψει ένα μοντέλο πρόβλεψης \hat{y} , το οποίο μπορεί να προεκταθεί προς το μέλλον. Κατά αντίστοιχο τρόπο, το στάδιο της διάσπασης μπορεί επίσης να επιμεριστεί σε δύο βήματα υπολογισμού:

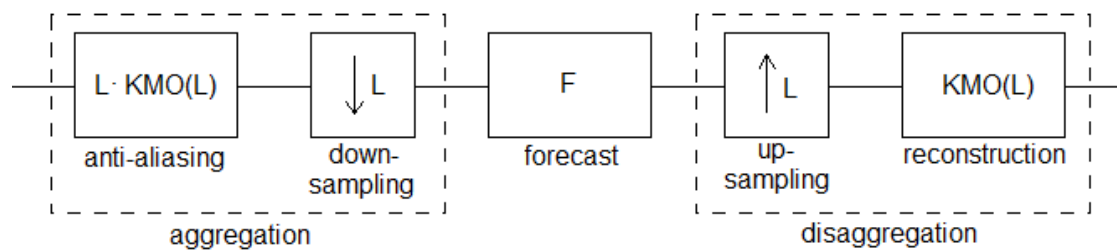
$$\hat{d}[n] = \begin{cases} \hat{y}\left[\frac{n}{L}\right] & \frac{n}{L} \in \mathbb{Z} \\ 0 & \text{αλλιώς} \end{cases}$$

$$\hat{x}[n] = \sum_{k=1}^L w_k \hat{d}[n - k]$$

όπου τα $w_k, k = 1, \dots, L$ αποτελούν τα βάρη με τα οποία γίνεται η διάσπαση. Το πρώτο βήμα της διάσπασης εκφράζει απλή υπερδειγματοληψία κατά L της

χρονοσειράς, μία διαδικασία, δηλαδή, κατά την οποία αυξάνουμε το πλήθος των παρατηρήσεων της χρονοσειράς, προσθέτοντας $L - 1$ μηδενικά μετά από κάθε διαθέσιμη παρατήρηση. Κατόπιν, ακολουθεί ένας «πίσω» σταθμισμένος κινητός μέσος όρος μήκους L , ο οποίος καταμερίζει τις προβλέψεις στο κατώτερο επίπεδο παρατήρησης. Ο προσδιορισμός «πίσω» αναφέρεται στο γεγονός πως χρησιμοποιούνται τιμές από τις προηγούμενες χρονικές στιγμές.

Συνδυάζοντας τις παραπάνω περιγραφές των επιμέρους σταδίων της διαδικασίας ADIDA, προκύπτει μία συστημική προσέγγιση της εικόνας 3.1. Η προσέγγιση αυτή διευκολύνει τη μελέτη της διαδικασίας, διασπώντας την σε επιμέρους απλούστερα στάδια, για τα οποία διατίθεται εκτεταμένη βιβλιογραφική μελέτη. Παράλληλα, η συστημική προσέγγιση αποκαλύπτει ομοιότητες της μεθοδολογίας με τις πολύ-ρυθμικές τεχνικές επεξεργασίας σήματος.



Εικόνα 3.1

3.3 Θεωρητική Ανάλυση των Σταδίων της ADIDA

Στο σημείο αυτό θα επιχειρηθεί μία θεωρητική ανάλυση των επιμέρους σταδίων της ADIDA. Η μελέτη θα γίνει τόσο στο πεδίο του χρόνου, όσο και στο πεδίο της συχνότητας, με χρήση κατάλληλων μετασχηματισμών.

3.3.1 Συνάθροιση (Aggregation)

Στο πρώτο στάδιο της διαδικασίας, δομείται η συναθροισμένη χρονοσειρά. Όπως δείχτηκε ήδη, αρχικά υπολογίζεται ένας εμπρός απλός κινητός μέσος όρος μήκους L (ΚΜΟ(L)), ο οποίος πολλαπλασιάζεται με έναν παράγοντα L και κατόπιν διέρχεται από μία μονάδα υποδειγματοληψίας ρυθμού L .

Στο πεδίο του χρόνου, ο εμπρός ΚΜΟ(L) περιγράφεται μαθηματικά από τον τύπο:

$$y[n] = \frac{1}{L} \sum_{k=0}^{L-1} x[n+k]$$

Εφαρμόζοντας τον μετασχηματισμό Z , μπορούμε να πάρουμε μία νέα έκφραση για τον κινητό μέσο όρο:

$$Y(z) = \frac{1}{L} \sum_{k=0}^{L-1} z^k X(z) = z^L \frac{1}{L} \sum_{k=1}^L z^{-k} X(z)$$

Το z^k δεν είναι τίποτε άλλο παρά ο τελεστής προήγησης κατά k . Αντίστοιχα, το z^{-k} εκφράζει τον τελεστή καθυστέρησης κατά k . Συνεπώς, από την παραπάνω έκφραση είναι εμφανές ότι η διαδικασία της άθροισης γίνεται προς τα εμπρός, το οποίο απαιτεί την εξ αρχής ύπαρξη αποθηκευμένων δεδομένων. Εναλλακτικά, σε περίπτωση που επιθυμούμε η διαδικασία να εκτελείται on-line, απευθείας με την απόκτηση των δεδομένων, είναι απαραίτητη η χρήση ενός buffer (όπως εκφράζεται από τον όρο z^L) χωρητικότητας L παρατηρήσεων, ώστε αργότερα να χρησιμοποιηθεί ένας εμπρός κινητός μέσος όρος ($\frac{1}{L} \sum_{k=1}^L z^{-k} X(z)$).

Ο ανωτέρω μετασχηματισμός Z κάνει εμφανές πως μπορούμε να αντιμετωπίσουμε τον κινητό μέσο όρο ως ένα γραμμικό φίλτρο. Αντικαθιστώντας $z = e^{j\omega}$, προκύπτει εύκολα η συνάρτηση μεταφοράς του φίλτρου κινητού μέσου όρου:

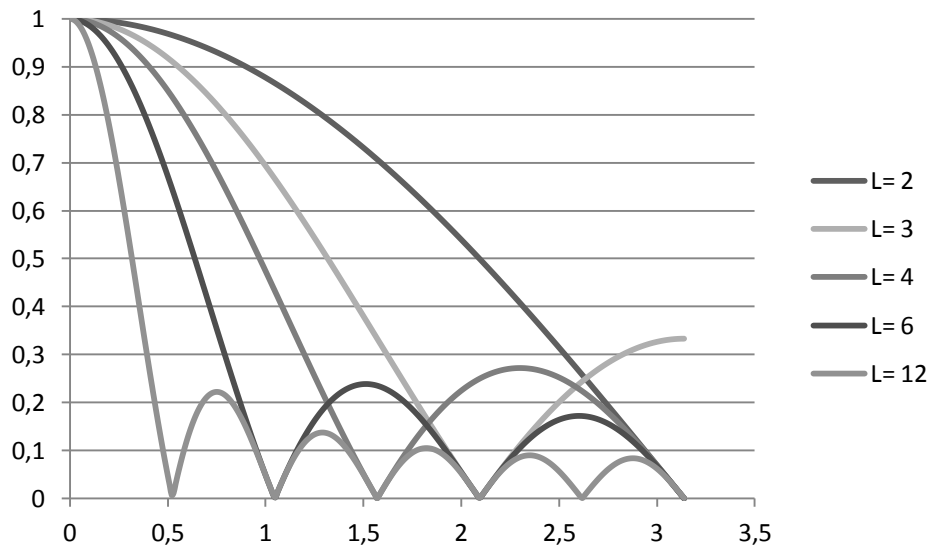
$$H(e^{j\omega}) = \frac{X(e^{j\omega})}{Y(e^{j\omega})} = \frac{\sin\left(\frac{\omega L}{2}\right)}{L \sin\left(\frac{\omega}{2}\right)} e^{j\omega \frac{(L-1)}{2}}$$

Πρόκειται ουσιαστικά για ένα βαθυπερατό φίλτρο, του οποίου η απόκριση συχνότητας εξαρτάται από την τάξη L (εικόνα 3.2). Όσο αυξάνεται η τάξη του φίλτρου, τόσο ενισχύεται και η βαθυπερατή του συμπεριφορά. Είναι ωστόσο εμφανές ότι η απόκριση απέχει πολύ από αυτή ενός ιδανικού φίλτρου, αφού οι $(L-1) \text{ div } 2$ το πλήθος πλευρικού λοβοί επιτρέπουν τη διέλευση και μέρους από τις υψηλότερες συχνότητες, με αυξανόμενη βέβαια εξασθένηση. Οι λοβοί αυτοί παρουσιάζουν μέγιστα στις συχνότητες που αποτελούν λύσεις της υπερβατικής εξίσωσης:

$$\frac{\tan\left(\frac{\omega L}{2}\right)}{L \tan\left(\frac{\omega}{2}\right)} = 1$$

Η απόκριση συχνότητας παρουσιάζει $L-1$ μηδενισμούς, στις συχνότητες $\omega = \frac{2k\pi}{L}, k = 1, \dots, L-1$. Ο πρώτος μηδενισμός παρατηρείται στη συχνότητα

$\omega = \frac{2\pi}{L}$, η οποία μπορεί να θεωρηθεί προσεγγιστικά ως το εύρος ζώνης του φίλτρου απλού κινητού μέσου όρου. Δεδομένων των συχνοτήτων μηδενισμού, τα σημεία όπου παρατηρούνται τα μέγιστα των λοβών μπορούν να προσεγγιστούν ως το μέσον μεταξύ δύο διαδοχικών μηδενισμών, δηλαδή από τις συχνότητες $\omega = \frac{(2k+1)\pi}{L}$. Τότε η εξασθένιση του k -στού λοβού στο μέγιστό του (χειρότερη περίπτωση) θα προσεγγίζεται ως $\left| L \sin\left(\frac{(2k+1)\pi}{2L}\right) \right|^{-1}$. Σε κάθε περίπτωση, η μηδενική συχνότητα διέρχεται χωρίς εξασθένιση (μοναδιαίο κέρδος), άρα ο ΚΜΟ δεν μεταβάλλει το επίπεδο (μέση τιμή) της αρχικής χρονοσειράς. Το φίλτρο χαρακτηρίζεται από γραμμική φάση ($\frac{j\omega(L-1)}{2}$), άρα δεν μεταβάλλεται η μορφή της φιλτραρισμένης χρονοσειράς.



Εικόνα 3.2

Αντίστοιχα, στο πεδίο του χρόνου η υποδειγματοληψία (down-sampling) ρυθμού L μίας χρονοσειράς x εκφράζεται από τη σχέση:

$$y[n] = x[nL]$$

Αν η χρονοσειρά x προκύπτει μέσω της δειγματοληψίας μίας μεταβλητής με περίοδο δειγματοληψίας T_s , η υποδειγματοληψία επιφέρει μεταβολή του ρυθμού δειγματοληψίας (αντίστροφο της περιόδου δειγματοληψίας), ώστε η νέα περίοδος δειγματοληψίας της να είναι $T'_s = LT_s$.

Η περιγραφή της διαδικασίας της υποδειγματοληψίας στο πεδίο της συχνότητας είναι περισσότερο πολύπλοκη. Απαραίτητη για τη μελέτη της είναι η γνώση της διακριτής συνάρτησης παλμοσειράς, η οποία ορίζεται ως:

$$II_L[n] = \frac{1}{L} \sum_{k=0}^{L-1} e^{-\frac{j2k\pi n}{L}} = \begin{cases} 1 & \frac{n}{L} \in \mathbb{Z} \\ 0 & \text{αλλιώς} \end{cases}$$

Η παραπάνω συνάρτηση είναι μηδενική παντού, εκτός από τα σημεία για τα οποία ισχύει $\frac{n}{L} \in \mathbb{Z}$. Ξεκινούμε και πάλι θεωρώντας το μετασχηματισμό Z της υπό μελέτη σχέσης:

$$\begin{aligned} Y(z) &= \sum_{n=-\infty}^{\infty} y[n]z^{-n} = \sum_{n=-\infty}^{\infty} x[nL]z^{-n} = \sum_{m=\dots, -L, 0, L, \dots} x[m]z^{-\frac{m}{L}} \\ &= \sum_{m=-\infty}^{\infty} x[m]II_L[m]z^{-\frac{m}{L}} = \sum_{m=-\infty}^{\infty} x[m] \left(\frac{1}{L} \sum_{k=0}^{L-1} e^{-\frac{j2k\pi m}{L}} \right) z^{-\frac{m}{L}} \\ &= \frac{1}{L} \sum_{k=0}^{L-1} \left[\sum_{m=-\infty}^{\infty} x[m]z^{-\frac{m}{L}} e^{-\frac{j2k\pi m}{L}} \right] = \frac{1}{L} \sum_{k=0}^{L-1} X \left(z^{\frac{1}{L}} e^{-\frac{j2k\pi}{L}} \right) \end{aligned}$$

Θέτοντας $z = e^{j\omega}$ προκύπτει ο μετασχηματισμός Laplace για τον τελεστή της υποδειγματοληψίας:

$$Y(e^{j\omega}) = \frac{1}{L} \sum_{k=0}^{L-1} X \left(e^{j\frac{\omega - 2k\pi}{L}} \right)$$

Με άλλα λόγια, η υποδειγματοληψία επιδρά στο φάσμα κλιμακώνοντας (επεκτείνοντας) κατά L τον άξονα των συχνοτήτων, μεταβάλλοντας το πλάτος του φάσματος κατά ένα συντελεστή $1/L$ και δημιουργώντας φασματικά αντίγραφα που έχουν υποστεί ολίσθηση κατά $\frac{2k\pi}{L}$.

Ένα πρόβλημα που συνδέεται με την διαδικασία της υποδειγματοληψίας είναι αυτό της φασματικής αναδίπλωσης (aliasing). Το πρόβλημα αυτό παρατηρείται όταν τα φασματικά αντίγραφα επικαλύπτονται, με αποτέλεσμα την άθροισή τους στο διάστημα επικάλυψης και την αλλαγή του φασματικού περιεχομένου του σήματος. Αν προκύψει φασματική αναδίπλωση, δεν είναι εφικτή η ακριβής ανακατασκευή της αρχικής σειράς, αφού η αθροιστική αλληλοεπικάλυψη των φασματικών αντιγράφων έχει προσθέσει θόρυβο. Αν η υποδειγματοληψία έχει γίνει με σωστό τρόπο ώστε να μην συμβαίνει φασματική αναδίπλωση, η αρχική χρονοσειρά μπορεί πάντα να ανακτηθεί, με τη χρήση ενός βαθυπερατού φίλτρου για την αποκοπή των φασματικών αντιγράφων. Η αναδίπλωση μπορεί να

αποφευχθεί αν το σήμα που υφίσταται την δειγματοληψία είναι βαθυπερατό, δηλαδή το φάσμα του είναι μηδενικό πάνω από κάποια κρίσιμη συχνότητα.

Η κρίσιμη συχνότητα αυτή βρίσκεται σύμφωνα με το θεώρημα της δειγματοληψίας του Shannon, που επιβάλλει ο ρυθμός δειγματοληψίας να είναι τουλάχιστον διπλάσιος της μέγιστης συχνότητας f_{max} που περιέχεται στα δεδομένα. Τα δεδομένα του πραγματικού κόσμου περιέχουν θόρυβο, ο οποίος συχνά σχετίζεται με υψηλές συχνότητες, που όμως δεν έχουν υψηλή ισχύ. Έτσι μπορεί αντί της συχνότητας f_{max} να χρησιμοποιηθεί εναλλακτικά μία εκτίμηση του εύρους ζώνης B των πραγματικών δεδομένων. Τελικά, η φασματική αναδίπλωση αποφεύγεται αν η νέα περίοδος δειγματοληψίας υπακούει στην ανίσωση:

$$T'_s \leq \frac{1}{2B}$$

Η εφαρμογή απλού κινητού μέσου όρου πριν από την υποδειγματοληψία, ο οποίος αποτελεί τελικά βαθυπερατό φιλτράρισμα, βοηθά στον περιορισμό του εύρους ζώνης B και συνεπώς στην αποτροπή φαινομένων φασματικής αναδίπλωσης. Για την ακρίβεια, είναι συνήθης πρακτική η μονάδα υποδειγματοληψίας να προηγείται από ένα βαθυπερατό φίλτρο, για τον σκοπό αυτό. Για το λόγο ακριβώς αυτό, ονομάζεται anti-aliasing φίλτρο.

Η διάταξη ενός βαθυπερατού anti-aliasing φίλτρου ακολουθούμενου από τη μονάδα υποδειγματοληψίας χρησιμοποιείται συχνά για την μείωση του ρυθμού των δεδομένων, υλοποιώντας μία διαδικασία που ονομάζεται αποδεκατισμός (decimation).

3.3.2 Πρόβλεψη (Forecasting)

Δεδομένου ότι η μέθοδος ADIDA είναι ουσιαστικά ένα μεθοδολογικό πλαίσιο, δεν υπάρχει κανένας περιορισμός σχετικά με τη μέθοδο πρόβλεψης που χρησιμοποιείται για την προέκταση της χρονοσειράς στο μέλλον. Ωστόσο, μέθοδοι χρονοσειρών, όπως οι μέθοδοι εκθετικής εξομάλυνσης ή η μέθοδος Theta, παρουσιάζουν μεγαλύτερη ευκολία και αμεσότητα στην εφαρμογή. Επιπλέον, οι μαθηματικές τους εκφράσεις είναι επαρκώς απλές ώστε να επιτρέπουν την θεωρητική ανάλυση της μεθοδολογίας από την αρχή της μέχρι το τέλος.

3.3.3 Διάσπαση (Disaggregation)

Για τη διάσπαση του μοντέλου πρόβλεψης και των προβλέψεων, αρχικά πραγματοποιείται υπερδειγματοληψία ρυθμού L και στη συνέχεια υπολογίζεται ένας πίσω σταθμισμένος κινητός μέσος όρος μήκους L (ΣΚΜΟ(L)).

Στο πεδίο του χρόνου, η υπερδειγματοληψία (up-sampling) ρυθμού L μίας χρονοσειράς x εκφράζεται ως:

$$y[n] = \begin{cases} x\left[\frac{n}{L}\right] & \frac{n}{L} \in \mathbb{Z} \\ 0 & \text{αλλιώς} \end{cases}$$

Αν η χρονοσειρά x προκύπτει μέσω της δειγματοληψίας μίας μεταβλητής με περίοδο δειγματοληψίας T_s' , η υποδειγματοληψία επιφέρει μεταβολή του ρυθμού δειγματοληψίας (αντίστροφο της περιόδου δειγματοληψίας), ώστε η νέα περίοδος δειγματοληψίας της να είναι $T_s = \frac{T_s'}{L}$. Άρα η διαδικασία υπερδειγματοληψίας εξουδετερώνει την επίδραση στο ρυθμό των δεδομένων της υποδειγματοληψίας κατά τη συνάθροιση.

Για την περιγραφή της διαδικασίας της υπερδειγματοληψίας στο πεδίο της συχνότητας ξεκινούμε θεωρώντας το μετασχηματισμό Z της υπό μελέτη σχέσης:

$$Y(z) = \sum_{n=-\infty}^{\infty} y[n]z^{-n} = \sum_{n=\dots, -L, 0, L, \dots} x\left[\frac{n}{L}\right]z^{-n} = \sum_{m=-\infty}^{\infty} x[m]z^{-mL} = X(z^L)$$

Θέτοντας $z = e^{j\omega}$ προκύπτει ο μετασχηματισμός Laplace για τον τελεστή της υπερδειγματοληψίας:

$$Y(e^{j\omega}) = X(e^{j\omega L})$$

Με άλλα λόγια, η υποδειγματοληψία επιδρά στο φάσμα κλιμακώνοντας (συμπύσσοντας) κατά $1/L$ τον άξονα των συχνοτήτων. Η υπερδειγματοληψία δεν σχετίζεται με την εμφάνιση φαινομένων φασματικής αναδίπλωσης.

Στη συνέχεια, με τη χρήση ενός πίσω σταθμισμένου κινητού μέσου όρου μήκους L (ΣΚΜΟ(L)) ολοκληρώνεται η διαδικασία της διάσπασης.

Στο πεδίο του χρόνου, ο πίσω ΣΚΜΟ(L) περιγράφεται μαθηματικά από τον τύπο:

$$y[n] = \sum_{k=1}^L w_k x[n-k]$$

Το άθροισμα των βαρών διάσπασης $w_k, k = 1, \dots, L$ πρέπει να ισούται με τη μονάδα, ώστε ο παραπάνω τύπος να εκφράζει σταθμισμένο μέσο όρο:

$$\sum_{k=1}^L w_k = 1$$

Εφαρμόζοντας τον μετασχηματισμό Z , μπορούμε να πάρουμε μία νέα έκφραση για τον σταθμισμένο κινητό μέσο όρο:

$$Y(z) = \sum_{k=1}^L w_k z^{-k} X(z)$$

Η χρήση του συντελεστή καθυστέρησης z^{-k} δηλώνει ότι η διαδικασία της άθροισης γίνεται προς τα πίσω. Ο ανωτέρω μετασχηματισμός Z κάνει εμφανές πως μπορούμε να αντιμετωπίσουμε τον κινητό μέσο όρο ως ένα γραμμικό φίλτρο. Αντικαθιστώντας $z = e^{j\omega}$, προκύπτει εύκολα η συνάρτηση μεταφοράς του φίλτρου σταθμισμένου κινητού μέσου όρου:

$$H(e^{j\omega}) = \frac{X(e^{j\omega})}{Y(e^{j\omega})} = \sum_{k=1}^L w_k e^{-j\omega k}$$

Με αντιστροφή του μετασχηματισμού Laplace βρίσκουμε την κρουστική απόκριση του φίλτρου:

$$h[n] = \sum_{k=1}^L w_k \delta[n-k] = \begin{cases} w_n & n = 1, \dots, L \\ 0 & \text{αλλιώς} \end{cases}$$

Οι ιδιότητες του φίλτρου εξαρτώνται προφανώς από τα βάρη διάσπασης $w_k, k = 1, \dots, L$. Μία ειδική περίπτωση, προκύπτει όταν χρησιμοποιούνται ίσα βάρη, που για να ικανοποιούν τον περιορισμό του αθροίσματος των βαρών πρέπει να ισούνται με $1/L$, δηλαδή ο πίσω ΣΚΜΟ(L) εκφυλίζεται σε πίσω απλό ΚΜΟ(L). Τότε, η συνάρτηση μεταφοράς του φίλτρου απλοποιείται ως:

$$H(e^{j\omega}) = \frac{\sin\left(\frac{\omega L}{2}\right)}{L \sin\left(\frac{\omega}{2}\right)} e^{-\frac{j\omega(L+1)}{2}}$$

Παρατηρώντας την ομοιότητα της παραπάνω σχέσης με την αντίστοιχη για τον πίσω ΚΜΟ(L), διαπιστώνουμε ότι τα πλάτη της απόκρισης συχνότητας των δύο φίλτρων ταυτίζονται και παριστάνονται στην εικόνα 3.2. Επομένως, πρόκειται επίσης για ένα βαθυπερατό φίλτρο, για το οποίο ισχύουν οι παρατηρήσεις που έχουν γίνει ως τώρα. Το φίλτρο χαρακτηρίζεται από γραμμική φάση ($\frac{j\omega(L+1)}{2}$), άρα δεν μεταβάλλεται η μορφή της φιλτραρισμένης χρονοσειράς.

Η διάταξη μίας μονάδας αύξησης του ρυθμού των δεδομένων ακολουθούμενης από ένα βαθυπερατό φίλτρο χρησιμοποιείται συχνά για την αύξηση του ρυθμού των δεδομένων που έχουν προηγουμένως υποστεί υποδειγματοληψία, υλοποιώντας μία διαδικασία που ονομάζεται παρεμβολή (interpolation). Το βαθυπερατό φίλτρο χρησιμοποιείται τότε για την απομόνωση του φασματικού αντιγράφου που περιέχει την αρχική πληροφορία. Υπό μία τέτοια σκοπιά, το βαθυπερατό φίλτρο θεωρείται ως ένα φίλτρο ανακατασκευής (reconstruction).

3.3.4 Συνολική Διάταξη

Η συνολική διάταξη που αντιπροσωπεύει τη συστημική προσέγγιση της μεθοδολογίας ADIDA βρίσκεται σε τέλεια αναλογία με τις διατάξεις πολύ-ρυθμικών τεχνικών επεξεργασίας σήματος. Συγκεκριμένα, αποτελεί μία διάταξη υποβιβασμού του ρυθμού δειγματοληψίας, επεξεργασίας σήματος και επαναφοράς στον αρχικό ρυθμό δεδομένων. Ειδικότερα, ο εμπρός ΚΜΟ παίζει το ρόλο ενός βαθυπερατού anti-aliasing φίλτρου, ώστε να αποφευχθούν τα φαινόμενα φασματικής αναδίπλωσης που ενδέχεται να εισάγει η μονάδα υποβιβασμού του ρυθμού των δεδομένων. Στη συνέχεια, τα δεδομένα υφίστανται επεξεργασία, η οποία στην παρούσα περίπτωση είναι η προέκτασή τους μέσω πρόβλεψης στο μέλλον, και κατόπιν επαναφέρονται στον αρχικό τους ρυθμό. Ακολουθεί ο πίσω ΚΜΟ, ο οποίος αποτελεί ένα βαθυπερατό φίλτρο ανακατασκευής, αποκόπτοντας τα φασματικά αντίγραφα.

Η αναλογία αυτή παρέχει βαθύτερη κατανόηση στις διαδικασίες που συντελούνται κατά την εφαρμογή της μεθοδολογίας ADIDA, επιτρέπει την ενοποιημένη μαθηματική αντιμετώπισή της από την αρχή μέχρι το τέλος, ενώ διευκολύνει την θεωρητική μελέτη αντιστοιχώντας τη μεθοδολογία στην εκτενώς μελετημένη διάταξη της πολυρυθμικής επεξεργασίας.

ΚΕΦΑΛΑΙΟ 4: Εφαρμογή της Μεθοδολογίας ADIDA

4.1 Δεδομένα Διακοπτόμενης Ζήτησης

Στην ίδια μελέτη στην οποία περιγράφηκε αρχικά η μεθοδολογία ADIDA, επιβεβαιώθηκε με εμπειρικό τρόπο η καλή της απόδοση στην παραγωγή προβλέψεων (Nikolopoulos et al., 2011). Η εφαρμογή της έγινε σε δεδομένα διακοπτόμενης ζήτησης, τύπος δεδομένων που άλλωστε έδωσε την αρχική ανάγκη και έμπνευση για την δημιουργία της μεθόδου. Το σύνολο των δεδομένων αποτελείτο από 5000 χρονοσειρές ζήτησης από μονάδες διατήρησης αποθεμάτων της Βασιλικής Πολεμικής Αεροπορίας (Royal Air Force – RAF) της Αγγλίας, καταγεγραμμένες σε ένα χρονικό διάστημα 7 ετών (συνολικά 84 παρατηρήσεις από τον Ιανουάριο του 1996 μέχρι και τον Δεκέμβριο του 2002). Οι ερευνητές κατακράτησαν τις τελευταίες 24 μηνιαίες παρατηρήσεις (δύο έτη) και εφάρμοσαν στην υπόλοιπη χρονοσειρά κυλιόμενη πρόβλεψη μίας περιόδου μέχρις ότου αποτιμηθούν οι εκτιμήσεις των παραπάνω 24 σημείων με χρήση της μεθοδολογίας $ADIDA(L, F, D)$ για διαφορετικούς συνδυασμούς των παραμέτρων της μεθόδου L , F και D . Τα σφάλματα που υπολογίστηκαν από την αντιπαραβολή των προβλέψεων και των πραγματικών τιμών χρησιμοποιήθηκαν για την εξαγωγή συμπερασμάτων προς τις ιδιότητες της μεθόδου.

Ήδη από την μελέτη των Nikolopoulos et al. (2011), διαπιστώθηκε πως η επιλογή του επιπέδου συνάθροισης είναι καθοριστικής σημασίας για την βέλτιστη απόδοση της μεθόδου. Συνεπώς, τέθηκε το πρόβλημα της αναγνώρισης ενός βέλτιστου επιπέδου συνάθροισης, το οποίο, όπως φάνηκε, είναι διαφορετικό ανά χρονοσειρά, αφού, σύμφωνα και με τη θεωρία, το επίπεδο αυτό θα σχετίζεται άμεσα με τα διαφορετικά στατιστικά χαρακτηριστικά κάθε χρονοσειράς. Δεδομένης ωστόσο της ομοιογένειας του δείγματος των χρονοσειρών, οι οποίες είχαν κοινή προέλευση, προτάθηκε η χρήση ενός κοινού επιπέδου συνάθροισης στο σύνολο των χρονοσειρών. Για την περίπτωση της μεθόδου Naïve, αναγνωρίστηκε ένα βέλτιστο επίπεδο συνάθροισης μήκους 9 περιόδων. Ωστόσο, η συσχέτιση της βέλτιστης τιμής από το δεδομένο σύνολο δεδομένων και η παρατήρηση ότι λίγες εκ του συνόλου των χρονοσειρών είχαν πραγματικά βέλτιστο το συγκεκριμένο επίπεδο δεν επιτρέπουν τη γενίκευση του συμπεράσματος αυτού και για άλλα σύνολα δεδομένων.

Παρότι η θεωρητική απόδειξη ύπαρξης και εξεύρεση ενός βέλτιστου επιπέδου συνάθροισης ανά χρονοσειρά δεν μελετήθηκε παρά αφέθηκε για περαιτέρω έρευνα, προτάθηκε μία επιχειρησιακή ευριστική προσέγγιση. Ειδικότερα, καλά αποτελέσματα φαίνεται να προκύπτουν θέτοντας το επίπεδο συνάθροισης ίσο με τον χρόνο παράδοσης (Lead Time – LT) προσαυξημένο κατά ένα

χρονικό διάστημα επιθεώρησης (Review – R), δηλαδή ίσο με $LT + R$. Ο χρόνος παράδοσης αναφέρεται στο διάστημα που μεσολαβεί μεταξύ της παραγγελίας ενός τεμαχίου της αποθήκης μέχρι την παραλαβή του, ενώ το διάστημα επιθεώρησης τίθεται τυπικά ίσο με τη μονάδα. Η χρήση της ευριστικής τιμής $LT + R$ στηρίζεται από την παρατήρηση ότι αθροιστικά σφάλματα που αντιστοιχούν σε αυτό ακριβώς το διάστημα απαιτούνται για τη λήψη αποφάσεων ελέγχου αποθέματος.

Ως προς τον αλγόριθμο προέκτασης των δεδομένων στο μέλλον, εφαρμόστηκαν οι μέθοδοι πρόβλεψης Naïve και SBA (Syntetos, and Boylan, 2001), εκ των οποίων η δεύτερη αποτελεί μία βελτίωση της μεθόδου πρόβλεψης δεδομένων διακοπτόμενης ζήτησης Croston (Croston, 1972). Μέσω της εφαρμογής τους στα πλαίσια της μεθοδολογίας ADIDA φάνηκε να βελτιώνεται την ακρίβεια και των δύο μεθόδων, για αυτό η μεθοδολογία χαρακτηρίστηκε ως ένας μηχανισμός αυτοβελτίωσης των μεθόδων πρόβλεψης (self-improving mechanism). Παρατηρήθηκε επίσης ότι η συναθροισμένη χρονοσειρά έχει λιγότερες ή καθόλου μηδενικές παρατηρήσεις, οπότε γίνεται εφικτή η εφαρμογή μεθόδων από τη ευρύτερη εργαλειοθήκη των μεθόδων πρόβλεψης δεδομένων συνεχούς ζήτησης.

Τέλος, δοκιμάστηκαν τρεις αλγόριθμοι διάσπασης της συναθροισμένης πρόβλεψης, όλοι βασισμένοι στη χρήση βαρών στάθμισης. Συγκεκριμένα, ελέγχθηκαν οι περιπτώσεις των ίσων (EQW), των προηγούμενων (PRW) και των μέσων βαρών (AVW). Η περίπτωση EQW έδωσε σταθερά τα μικρότερα σφάλματα πρόβλεψης, γεγονός που αποδόθηκε στη στασιμότητα των χρονοσειρών ζήτησης που μελετήθηκαν. Οι άλλες δύο περιπτώσεις απέτυχαν να δώσουν εξίσου μικρά σφάλματα, με την περίπτωση PRW να δίνει τα χειρότερα αποτελέσματα.

4.2 Δεδομένα Συνεχούς Ζήτησης

Μία διεξοδική αναζήτηση στη σχετική βιβλιογραφία αποκαλύπτει ότι μέχρι στιγμής δεν έχει πραγματοποιηθεί εις γνώση του γράφοντος κάποια θεωρητική ή εμπειρική μελέτη που να αφορά στην εξειδικευμένη εφαρμογή της μεθόδου ADIDA σε δεδομένα συνεχούς ζήτησης, παρά μόνο μελέτες σε επιμέρους ζητήματα, όπως η χρονική συνάθροιση ή η χρονική διάσπαση σε χαμηλότερο επίπεδο παρατήρησης, οι οποίες όμως εξετάζουν τα θέματα αυτά ανεξάρτητα από τα πλαίσια της εν λόγω μεθοδολογίας.

Βασικό αντικείμενο της μελέτης αυτής αποτελεί η εξακρίβωση της ποιότητας των παραγόμενων προβλέψεων όταν η μεθοδολογία ADIDA εφαρμόζεται σε δεδομένα συνεχόμενης ζήτησης, δυνατότητα που παρέχεται από την παρατήρηση ότι σε κανένα στάδιο της μεθοδολογίας δεν προϋποτίθεται η διακοπτόμενη φύση της προς πρόβλεψη χρονοσειράς. Πέραν της μελέτης της μεθοδολογίας ως ενός νέου τρόπου υπολογισμού προβλέψεων, υπάρχει μία πρακτική όψη του ζητήματος

που καθιστά ενδιαφέρουσα και χρήσιμη τη μελέτη της, λόγω της σύνδεσής της με κοινά και διαισθητικά χρησιμοποιούμενες πρακτικές, ιδιαίτερα στο επιχειρησιακό τομέα, όπως θα παρουσιαστεί ευθύς αμέσως.

Ένα γνωστό πρόβλημα που καλούνται να συμβιβάσουν όσοι τηρούν ιστορικό της μεταβολής διαφόρων μεταβλητών σχετίζεται με το κατάλληλο επίπεδο καταγραφής των δεδομένων. Συχνά η λεπτομερέστερη δειγματοληψία είναι ασύμφορη, λόγω του αυξημένου κόστους που ενδεχομένως συνεπάγεται, ή ακόμα και αδύνατη, όταν δεν αντιμετωπίζεται σύγχρονη καταμέτρηση της τιμής αλλά συλλογή των τιμών από ιστορικές πηγές (π.χ. δημοσιευμένα δεδομένα). Σε τέτοιες περιπτώσεις συχνά δεν υπάρχουν διαθέσιμα δεδομένα σε ένα επιθυμητό κατώτερο επίπεδο παρατήρησης (π.χ. εβδομαδιαίο) παρά μόνο σε κάποιο υψηλότερο επίπεδο (π.χ. μηνιαίο). Οι τιμές αυτές αντιστοιχούν στην συναθροισμένη χρονοσειρά, ακριβώς όπως θα είχε εκείνη προκύψει αν συνειδητά εφαρμόζαμε τη μεθοδολογία ADIDA, όταν το μετρούμενο μέγεθος αποτελεί μία μεταβλητή ροής. Στη συνέχεια, προκειμένου να αποκτήσουμε προβλέψεις για επόμενες τιμές στο κατώτερο επίπεδο παρατήρησης (εβδομαδιαίο), η διαισθητική πρακτική που συνήθως εφαρμόζεται είναι να προεκτείνουμε την συναθροισμένη χρονοσειρά και στη συνέχεια να διασπάσουμε με κάποιον τρόπο τις παραχθείσες προβλέψεις που βρίσκονται στο υψηλό επίπεδο (μηνιαίο), έτσι ώστε να επιστρέψουμε στο κατώτερο επίπεδο παρατήρησης. Δηλαδή, η παραγωγή προβλέψεων σε ένα λεπτομερέστερο χρονικά επίπεδο μπορεί να αναλυθεί ως μία διαδικασία ADIDA, στην οποία το μέρος της χρονικής συνάθροισης συμβαίνει αυτόματα κατά την διαδικασία δειγματοληψίας, ενώ τα υπόλοιπα στάδια της μεθόδου εφαρμόζονται τεχνητά/συνειδητά από τον υπεύθυνο για την παραγωγή της πρόβλεψης.

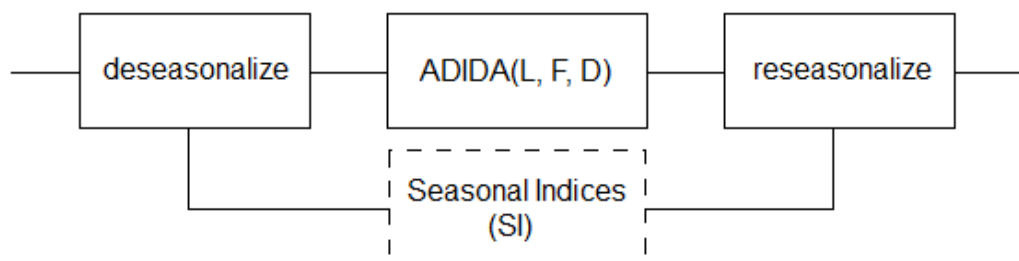
Στην παρούσα μελέτη θα ασχοληθούμε κατά κύριο λόγο με μία περισσότερο ελεγχόμενη εκδοχή της μεθόδου, στην οποία όλα τα στάδια εφαρμόζονται με τεχνητό/συνειδητό τρόπο, ως τμήματα της διαδικασίας παραγωγής προβλέψεων. Μία τέτοια προσέγγιση επιτρέπει την προεπεξεργασία της αρχικής χρονοσειράς, όπως η αποεποχικοποίησή της, η οποία φυσικά είναι αδύνατη στην περίπτωση που η χρονική συνάθροιση γίνεται αυτόματα.

Η απόδοση της μεθόδου σε δεδομένα συνεχόμενης ζήτησης θα ελεγχθεί μέσω της εφαρμογής της για πρόβλεψη των χρονοσειρών του συνόλου των δεδομένων που χρησιμοποιήθηκαν στον διαγωνισμό M3. Θα χρησιμοποιηθεί μία παρόμοια διαδικασία με εκείνη του διαγωνισμού, ώστε τα αποτελέσματα να είναι άμεσα συγκρίσιμα. Η διαδικασία αυτή έχει ως εξής:

Μελετώνται ξεχωριστά οι σειρές από τις διαφορετικές κατηγορίες διαστημάτων δειγματοληψίας (ετήσιες, τριμηνιαίες, μηνιαίες και άλλες). Μπορούμε επίσης να επιβάλλουμε τον επιπλέον διαχωρισμό σε κατηγορίες ανάλογα με την προέλευση της χρονοσειράς (μικροοικονομία, δημογραφικά κλπ). Επιλέγεται μία

κατηγορία και επεξεργαζόμαστε σειριακά τις χρονοσειρές της. Από κάθε σειρά αποκρύπτονται οι τελευταίες παρατηρήσεις, ίσου πλήθους με εκείνες που είχαν κρατηθεί κρυφές κατά τη διεξαγωγή του διαγωνισμού M3, σύμφωνα με τον πίνακα 2.2. Το εναπομείναν τμήμα της σειράς αποεποχικοποιείται. Οι εποχικοί δείκτες που χρησιμοποιούνται είναι αυτοί που είχαν δοθεί και κατά την διεξαγωγή του διαγωνισμού M3 και είχαν προκύψει εφαρμόζοντας τη διαδικασία της κλασικής αποσύνθεσης στις χρονοσειρές που ικανοποιούσαν έναν απλό έλεγχο εποχικότητας κατάλληλου μήκους. Κατόπιν, εφαρμόζεται η μεθοδολογία πρόβλεψης ADIDA για την παραγωγή προβλέψεων δεδομένου ορίζοντα (θα τίθεται ίσος με το κατά περίπτωση πλήθος κρυφών παρατηρήσεων, εκτός αν αναφέρεται αλλιώς), επιλέγοντας κάθε φορά κατάλληλα το επίπεδο συνάθροισης, την τεχνική προέκτασης και τον αλγόριθμο διάσπασης. Επειδή τα δεδομένα του διαγωνισμού ήταν αυστηρώς θετικά μεγέθη, τυχόν αρνητικές τιμές του μοντέλου πρόβλεψης αντικαθίστανται με την τιμή που αντιστοιχεί σε πρόβλεψη με χρήση της Naïve. Στη συνέχεια, το μοντέλο πρόβλεψης επανεποχικοποιείται σύμφωνα με τους προηγούμενους δείκτες εποχικότητας. Σχηματικά η διαδικασία παρουσιάζεται στην εικόνα 4.1.

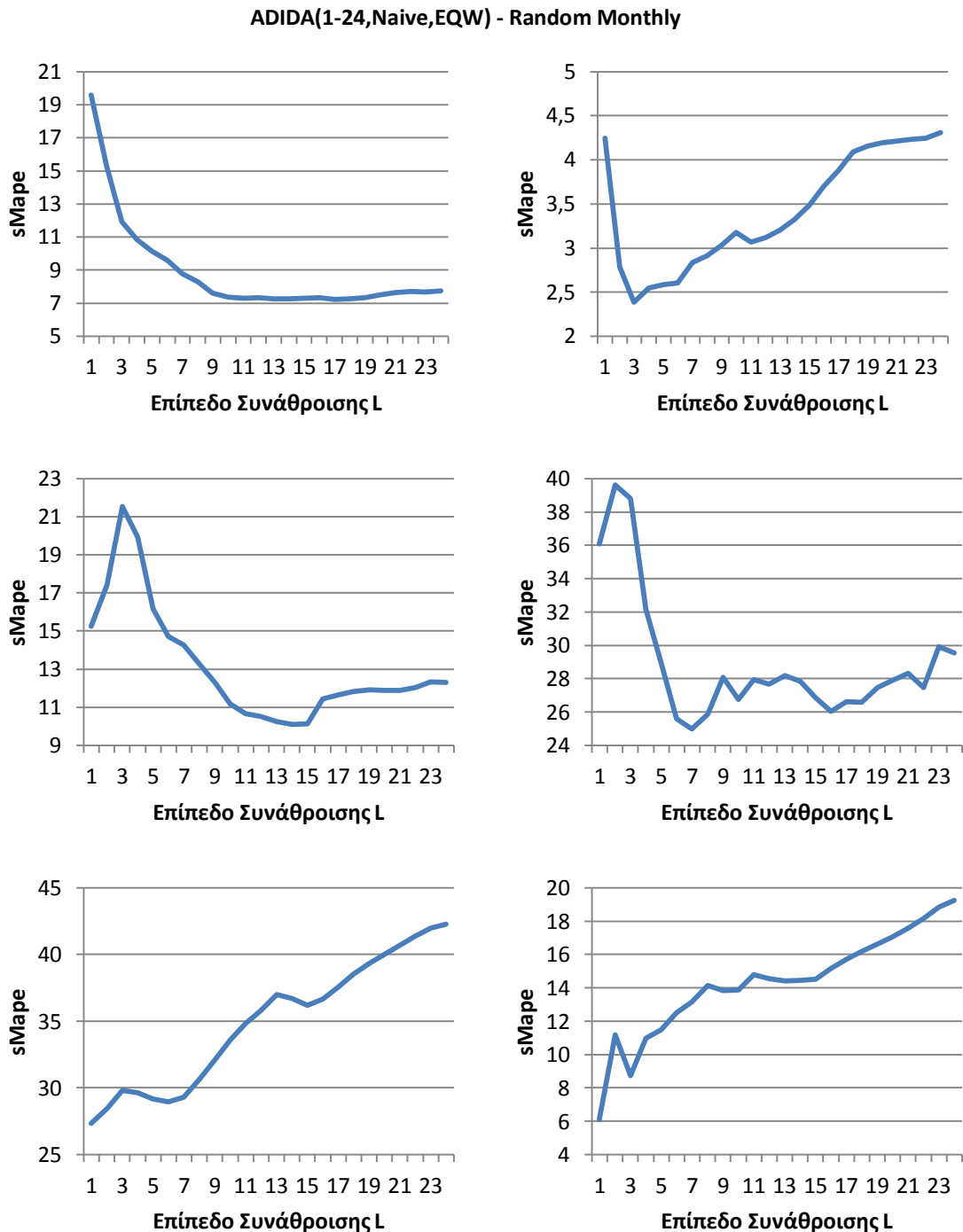
Στο τέλος, οι παραχθείσες προβλέψεις συγκρίνονται με τις αντίστοιχες πραγματικές τιμές για τον υπολογισμό κατάλληλων δεικτών σφάλματος, με ιδιαίτερο ενδιαφέρον στα out-of-sample σφάλματα, αφού αυτά μετρούν την ακρίβεια της πρόβλεψης. Από τους διάφορους δείκτες σφάλματος που θα μπορούσαν να μελετηθούν, θα εστιάσουμε την προσοχή μας στο symmetric MAPE, καθώς είναι έτσι δυνατή η άμεση σύγκριση μεταξύ διαφορετικών χρονοσειρών, ενώ αποτελεί ταυτόχρονα έναν από τους βασικούς δείκτες που μελετήθηκαν στον M3-Competition. Η συμπεριφορά του μέσου όρου των δεικτών αυτών ανά κατηγορία χρονοσειρών θα χρησιμοποιηθεί για την εξαγωγή συμπερασμάτων ως προς τις ιδιότητες της μεθοδολογίας.



Εικόνα 4.1

Στην εικόνα 4.2 παρουσιάζεται ενδεικτικά η μεταβολή του sMAPE για κάποιες τυχαίες χρονοσειρές της κατηγορίας των μηνιαίων χρονοσειρών

συναρτήσεως του επιπέδου συνάθροισης (στα όρια [1,24]). Σε όλες τις περιπτώσεις έχει χρησιμοποιηθεί η μέθοδος πρόβλεψης Naïve και ο αλγόριθμος διάσπασης με ίσα βάρη, εφαρμόζουμε δηλαδή ADIDA(1-24, Naïve, EQW).

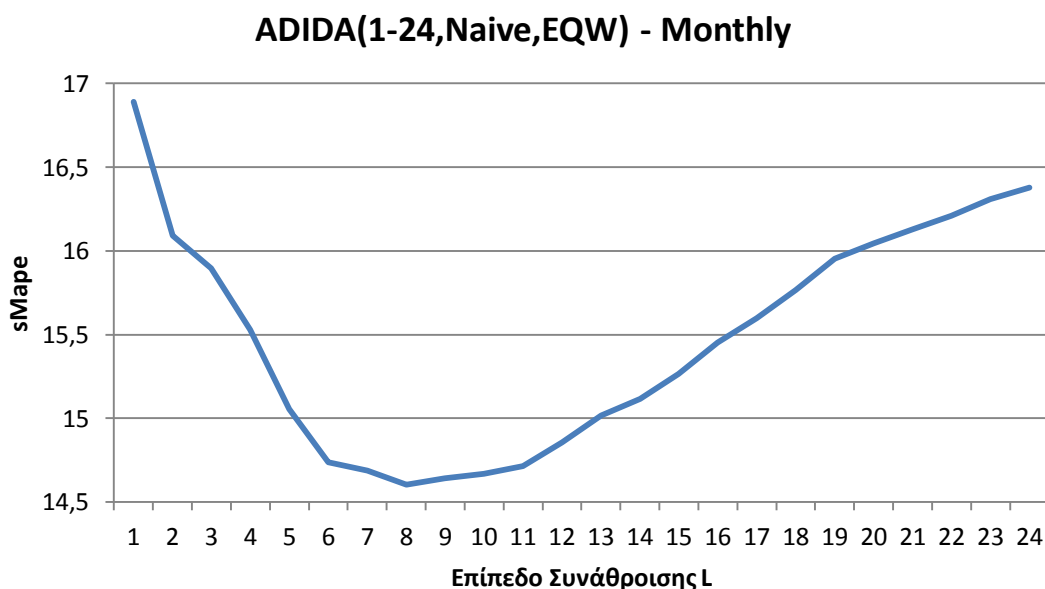


Εικόνα 4.2

Η συμπεριφορά της μεθοδολογίας φαίνεται να ποικίλει ανάλογα με την περίπτωση. Κάποιες φορές παρουσιάζει βελτίωση, δηλαδή υπάρχει κάποια μεγαλύτερη τιμή του L για την οποία το σφάλμα είναι μικρότερο από την τιμή για

$L = 1$ (υπενθυμίζεται ότι τότε η μεθοδολογία ταυτίζεται με απλή εφαρμογή της ενσωματωμένης μεθόδου πρόβλεψης). Άλλοτε η τιμή του σφάλματος μειώνεται σταθερά μέχρι κάποια ελάχιστη τιμή που παρατηρείται για κάποιο βέλτιστο επίπεδο συνάθροισης, ενώ άλλοτε φαίνεται να ακολουθείται ένα περισσότερο τυχαίο πρότυπο. Τέλος, υπάρχουν περιπτώσεις για τις οποίες δεν παρατηρείται καμία βελτίωση ως προς το σφάλμα.

Αυτή η ποικιλομορφία ως προς τη συμπεριφορά οφείλεται στα διαφορετικά χαρακτηριστικά των χρονοσειρών. Μελετώντας ταυτόχρονα ένα μεγαλύτερο σύνολο χρονοσειρών μπορούμε να πάρουμε μία περισσότερο συνεπή εικόνα που αποκαλύπτει τη συμπεριφορά της μεθοδολογίας στην μέση περίπτωση. Σχεδιάζουμε λοιπόν το διάγραμμα της εικόνας 4.3 όπου απεικονίζεται η μεταβολή του μέσου sMAPE για στο σύνολο των χρονοσειρών της κατηγορίας των μηνιαίων χρονοσειρών.



Εικόνα 4.3

Από το διάγραμμα είναι άμεσα εμφανής μία μείωση του μέσου σφάλματος από το 16,89% (αντιστοιχεί στην τιμή 16,91% που είχε επιτύχει η εφαρμογή της μεθόδου στον διαγωνισμό προβλέψεων) για μοναδιαίο επίπεδο συνάθροισης στο 14,60% (μεταβολή -2,29%) για επίπεδο συνάθροισης ίσο με 8. Το σφάλμα μειώνεται αρχικά ραγδαία μέχρι το επίπεδο $L = 8$ και στη συνέχεια αυξάνεται πιο αργά, ωστόσο, ακόμα και για $L = 24$, η τιμή του κυμαίνεται κάτω από εκείνη για $L = 1$. Για τα συγκεκριμένα δεδομένα, η μεθοδολογία λειτούργησε πράγματι στη μέση περίπτωση ως ένας μηχανισμός βελτίωσης της μεθόδου πρόβλεψης Naive.

Οι εικόνες 4.2 και 4.3 φανερώνουν απευθείας την εξάρτηση της απόδοσης της μεθοδολογίας από την επιλογή του επιπέδου συνάθροισης. Η επιλογή του ώστε

να βρίσκεται κοντά στη βέλτιστη τιμή είναι καθοριστική για την ακρίβεια της μεθόδου. Παράλληλα, υπάρχει η ανάγκη να εξεταστεί αν τα παραπάνω προκαταρκτικά συμπεράσματα μπορούν να γενικευθούν και για τη χρήση διαφορετικών μεθόδων πρόβλεψης, καθώς και ποια είναι η καταλληλότερη επιλογή του αλγορίθμου διάσπασης. Απευθυνόμαστε στα ειδικότερα ζητήματα αυτά ξεχωριστά στα κεφάλαια που ακολουθούν.

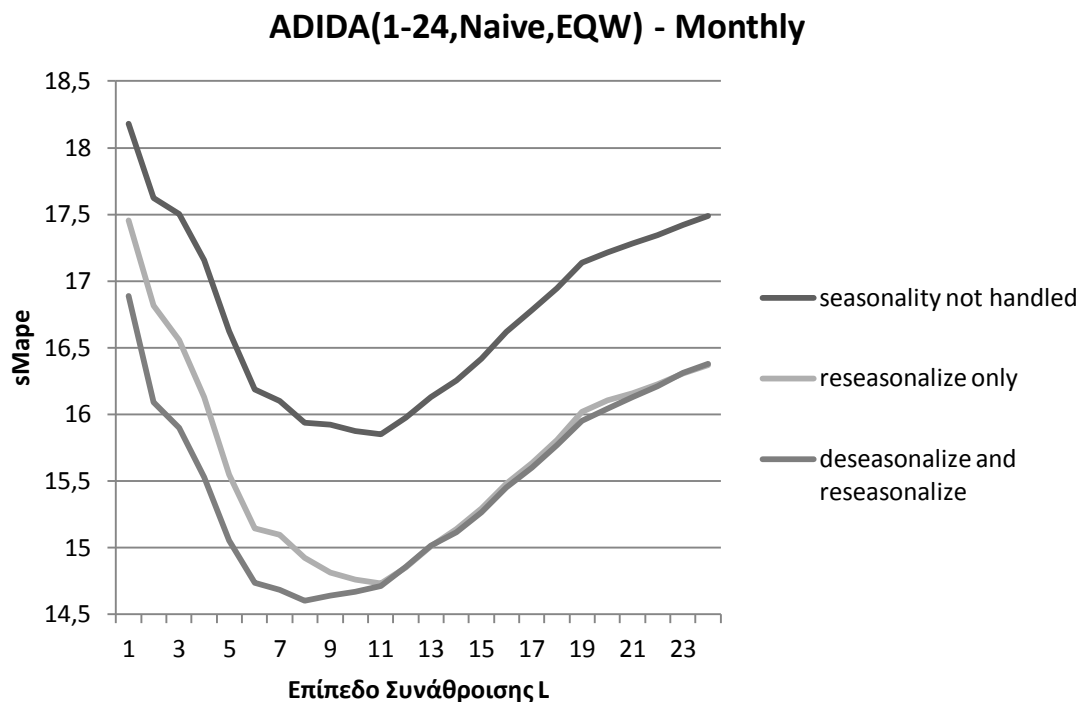
4.3 Μεθοδολογία ADIDA και Αποεποχικοποίηση

Κλείνουμε το κεφάλαιο αυτό με μία αναφορά στην περίπτωση που τα διαθέσιμα δεδομένα βρίσκονται σε επίπεδο παρατήρησης ανώτερο από το επίπεδο πρόβλεψης, οπότε (αναφερόμενοι σε μεταβλητές ροής) το στάδιο της συνάθροισης της χρονοσειράς υλοποιείται αυτόματα κατά τη διαδικασία της δειγματοληψίας. Όπως ήδη αναφέρθηκε, στην περίπτωση αυτή είναι φυσικά αδύνατη η προεπεξεργασία της αρχικής χρονοσειράς για την απαλοιφή της εποχικότητας, αφού η συνάθροιση γίνεται μη ελεγχόμενα και διατίθεται απευθείας μόνο η συναθροισμένη χρονοσειρά.

Μελετώντας την απόδοση της μεθοδολογίας αν τα υπόλοιπα στάδιά της εφαρμοστούν από το σημείο αυτό και ύστερα, παρατηρούμε και πάλι ότι υπάρχει μείωση του σφάλματος από την τιμή 18,18% σε 15,85% (μεταβολή -2,33%) για $L = 11$. Άρα η βελτιωτική δράση της μεθοδολογίας φαίνεται να ισχύει και για την εφαρμογή της και σε εποχικές χρονοσειρές. Ωστόσο, η καμπύλη του sMAPE παραμένει μονίμως σε υψηλότερο επίπεδο από την αντίστοιχη περίπτωση όπου έχει εφαρμοστεί εκατέρωθεν της μεθοδολογίας αποεποχικοποίηση και επανεποχικοποίηση (εικόνα 4.4). Η μορφή της καμπύλης διατηρείται σε αξιοπρόσεκτο βαθμό και η επίδραση του σχήματος χειρισμού της εποχικότητας φαίνεται να προκαλεί απλά μία κατακόρυφη μετακίνησή της.

Ωστόσο, η απόδοση της μεθόδου στην περίπτωση της μη ελεγχόμενης συνάθροισης μπορεί να βελτιωθεί, αρκεί να υπάρχουν διαθέσιμες αρκετά ακριβείς εκτιμήσεις για την εποχικότητα της αρχικής (μη συναθροισμένης) χρονοσειράς, με τη μορφή δεικτών εποχικότητας. Αν, παρά την απουσία ενός σταδίου αποεποχικοποίησης κατά την προεπεξεργασία, επανεποχικοποιήσουμε με τους διαθέσιμους δείκτες τις παραχθείσες προβλέψεις, τότε παρατηρούμε ότι για επαρκώς μεγάλες τιμές του επιπέδου συνάθροισης, πάνω από την τιμή της εποχικότητας πιο συγκεκριμένα, τα σφάλματα προσεγγίζουν εκείνα της περίπτωσης της πλήρως ελεγχόμενης (με πλαισίωση από διαδικασία αποεποχικοποίησης-επανεποχικοποίησης) εφαρμογής της μεθοδολογίας (εικόνα 4.4). Η συμπεριφορά αυτή ερμηνεύεται από την λειτουργία του κινητού μέσου όρου στο στάδιο της

συνάθροισης που, για μήκη μεγαλύτερα της εποχικότητας της χρονοσειράς, αποκόπτει την εποχική συνιστώσα.



Εικόνα 4.4

Η συγκεκριμένη παρατήρηση μπορεί να δώσει εύρωστα και αξιόπιστα αποτελέσματα υπό την προϋπόθεση ότι χρησιμοποιείτε επίπεδο συνάθροισης μεγαλύτερο της εποχικότητας, ειδάλλως ελλοχεύει ο κίνδυνος επανεποχικοποίησης μίας ήδη εποχικής χρονοσειράς. Στο σημείο αυτό πρέπει να επισημανθεί ωστόσο ότι γενικά, όπως θα δειχθεί και στη συνέχεια, το βέλτιστο επίπεδο συνάθροισης δεν λαμβάνει συνήθως πολύ μεγάλες τιμές. Επίσης, τίθεται το ζήτημα της ποιότητας των διαθέσιμων δεικτών, καθώς στο παραπάνω διάγραμμα χρησιμοποιήθηκαν οι πραγματικοί εποχικοί δείκτες που εξήχθησαν από την αρχική χρονοσειρά. Στην πραγματικότητα, δεν θα μπορούσαμε να έχουμε τους ακριβείς αυτούς δείκτες, παρά μόνο συγκριτικές και στατιστικές εκτιμήσεις τους, η ακρίβεια των οποίων θα επηρέαζε το τελικό αποτέλεσμα.

ΚΕΦΑΛΑΙΟ 5: Επίπεδο Συνάθροισης

5.1 Σημασία της Επιλογής του Επιπέδου Συνάθροισης

Όπως φάνηκε από τα παραδείγματα του προηγούμενου κεφαλαίου, η επιλογή του επιπέδου συνάθροισης L είναι μέγιστης σημασίας στην καλή συμπεριφορά της μεθοδολογίας. Επίσης, διαφαίνεται η ύπαρξη ενός βέλτιστου επιπέδου συνάθροισης, διαφορετικού ανά χρονοσειρά, το οποίο, σύμφωνα και με τη θεωρία, θα πρέπει να εξαρτάται με κάποιον τρόπο από τα χαρακτηριστικά της χρονοσειράς. Συνεπώς, προκύπτει η ανάγκη της εξεύρεσης ενός κατάλληλου τρόπου επιλογής του L , έτσι ώστε να προκύπτουν ακριβείς και συνεπείς προβλέψεις.

Δυστυχώς, η εμπειρική προσέγγιση που προτάθηκε για την περίπτωση ελέγχου αποθέματος (Nikolopoulos et al, 2011), να τίθεται δηλαδή το L ίσο με $LT + R$ όπου LT ο χρόνος παράδοσης τεμαχίου (Lead Time) και R ένα διάστημα επιθεώρησης (Review, τυπικά θεωρείται μοναδιαίο), δεν μπορεί να έχει εφαρμογή στην γενικότερη περίπτωση χρονοσειρών οποιουδήποτε τύπου, με την μέγιστη δυσχέρεια να εμφανίζεται στην περίπτωση χρονοσειρών συνεχόμενης ζήτησης. Συνεπώς, είναι ανάγκη να μελετηθούν εναλλακτικοί τρόποι επιλογής ενός κατάλληλου επιπέδου συνάθροισης.

Μία προφανής επίδραση του επιπέδου συνάθροισης στις ιδιότητες της μεθοδολογίας σχετίζεται με την ταύτισή του με το μήκος των δύο ΚΜΟ που χρησιμοποιούνται: του απλού ΚΜΟ(L) κατά τη συνάθροιση και του σταθμισμένου ΚΜΟ(L) κατά τη διάσπαση. Αύξηση του L συνεπάγεται αύξηση του βαθμού εξομάλυνσης που προκαλούν οι δύο ΚΜΟ. Ισοδύναμα είδαμε ότι ειδικότερα στην περίπτωση του απλού ΚΜΟ (στον οποίο επίσης μεταπίπτει και ο ΣΚΜΟ για διάσπαση EQW) η εφαρμογή του αποτελεί γραμμικού φιλτράρισμα μέσω ενός μη ιδανικού βαθυπερατού φίλτρου, του οποίου η (κανονικοποιημένη) συχνότητα αποκοπής ισούται με $2\pi/L$.

Αυτή η αυξανόμενη με το L εξομάλυνση σχετίζεται με μία αυξανόμενη δυνατότητα απόρριψης του θορύβου που ενυπάρχει στη χρονοσειρά με τη μορφή τυχαίων αυξομειώσεων. Πράγματι, υπό την υπόθεση ότι αντιμετωπίζουμε λευκό γκαουσιανό αθροιστικό θόρυβο μηδενικής μέσης τιμής, ο θόρυβος που παραμένει μετά το γραμμικό φιλτράρισμα από έναν ΚΜΟ(L) είναι ομοειδής (αφού το άθροισμα μεταβλητών που ακολουθούν κανονική κατανομή ακολουθεί επίσης κανονική κατανομή). Ο μέσος όρος του φιλτραρισμένου θορύβου είναι επίσης μηδενικός, όπως φαίνεται από τον τύπο:

$$\mu_y = \sum_{k=1}^L w_k \mu_x = 0$$

Από την άλλη μεριά, η διακύμανση του φιλτραρισμένου θορύβου σχετίζεται με τη διακύμανση του αρχικού θορύβου μέσω ενός πολλαπλασιαστικού συντελεστή, όπως προκύπτει από τη σχέση:

$$\sigma_y^2 = \sum_{k=1}^L w_k^2 \sigma_x^2$$

Με άλλα λόγια, ισχύει η παρακάτω σχέση μεταξύ των τυπικών αποκλίσεων:

$$\sigma_y = \sigma_x \sqrt{\sum_{k=1}^L w_k^2}$$

Και για την ειδική περίπτωση των ίσων βαρών $w_k = \frac{1}{L}$, ο φιλτραρισμένος θόρυβος έχει τυπική απόκλιση:

$$\sigma_y = \frac{1}{\sqrt{L}} \sigma_x$$

Για το συντελεστή που σταθμίζει την αρχική διακύμανση προκύπτει άμεσα από την γενικευμένη τριγωνική ανισότητα, ότι είναι πάντοτε μικρότερος ή ίσος της μονάδας:

$$\sum_{k=1}^L w_k^2 \leq \left(\sum_{k=1}^L w_k \right)^2 = 1$$

Επομένως, προκειμένου για λευκό γκαουσιανό αθροιστικός θόρυβος μηδενικής μέσης τιμής, μετά την εφαρμογή ενός ΚΜΟ(L) προκύπτει και πάλι λευκός γκαουσιανός αθροιστικός θόρυβος μηδενικής μέσης τιμής αλλά μειωμένης διακύμανσης κατά το συντελεστή $\sqrt{\sum_{k=1}^L w_k^2} \leq 1$, ο οποίος εξαρτάται από τα βάρη στάθμισης του ΚΜΟ, w_k . Φυσικά, η επιλογή πολύ υψηλού επιπέδου συνάθροισης μπορεί να οδηγήσει σε ισχυρή εξομάλυνση που πέραν του θορύβου ίσως καταπιέσει συνιστώσες της χρονοσειράς που φέρουν πληροφορία. Η κατάλληλη

επιλογή από την άποψη αυτή εξαρτάται από την έννοια που δίνουμε στον θόρυβο και τον επιθυμητό βαθμό εξομάλυνσης.

Εναλλακτικά, με αναφορά στην εικόνα 3.1, όπου περιγράφεται διαγραμματικά η συστημική προσέγγιση της μεθόδου, και σύμφωνα με όσα αναφέρθηκαν για την υποδειγματοληψία, παρατηρούμε ότι η επιλογή του L μπορεί να ιδωθεί και ως επιλογή του ελάχιστου ρυθμού δειγματοληψίας. Όντως, αν η αρχική χρονοσειρά δειγματοληπτείται με περίοδο T_s , κατά τη χρονική συνάθροιση η περίοδος δειγματοληψίας μεταβάλλεται σε $T'_s = LT_s$, δηλαδή ο ρυθμός δειγματοληψίας μειώνεται κατά τον παράγοντα L . Η χρήση της υποδειγματοληψίας σχετίζεται με την εμφάνιση φαινομένων φασματικής αναδίπλωσης, η επίδραση των οποίων μετριάζεται από την εφαρμογή του κινητού μέσου όρου που λειτουργεί ως φίλτρο anti-aliasing.

Η φασματική αναδίπλωση αποτρέπεται σύμφωνα με το θεώρημα δειγματοληψίας όταν ισχύει $T'_s \leq \frac{1}{2B}$, όπου B είναι το εύρος ζώνης της φιλτραρισμένης από τον απλό ΚΜΟ χρονοσειράς. Με άλλα λόγια, όπως ορίζεται από την παρακάτω ανίσωση, υφίσταται μία μέγιστη τιμή του L , πάνω από την οποία παρατηρείται φασματική αναδίπλωση που επηρεάζει την αρχική χρονοσειρά, προσθέτοντας ανεπιθύμητο θόρυβο:

$$L \leq \frac{1}{2BT_s}$$

Μέχρι αυτό το σημείο έχουμε παρακολουθήσει δύο που σχετίζονται με την επιλογή του επιπέδου συνάθροισης L . Από τη μία, αυξανόμενη τιμή του συνδέεται με αυξανόμενη ικανότητα απόρριψης του θορύβου, μέσω της εξομαλυντικής δράσης των δύο ΚΜΟ. Απαιτείται ιδιαίτερη προσοχή κατά την επιλογή, καθώς μία υψηλή τιμή μπορεί να καταπιέσει χρήσιμες διακυμάνσεις της χρονοσειράς. Από την άλλη, υψηλές τιμές ενδεχομένως να οδηγήσουν σε παραβίαση του θεωρήματος της δειγματοληψίας και την προσθήκη ανεπιθύμητου θορύβου λόγω της φασματικής αναδίπλωσης. Αν και οι δύο διαφορετικοί τρόποι επίδρασης του επιπέδου συνάθροισης φαίνονται να αλληλοσυγκρούονται στις χαμηλότερες τιμές και οι δύο συνηγορούν στη διαπίστωση ότι πολύ μεγάλες τιμές έχουν αρνητική επίπτωση στην απόδοση της μεθόδου, καθώς προκαλούν υπερβολική εξομάλυνση και/ή εισάγουν θόρυβο από φασματική αναδίπλωση.

Στη συνέχεια του κεφαλαίου θα ασχοληθούμε με δύο διαφορετικές προσεγγίσεις: τη χρήση κοινού επιπέδου συνάθροισης για ένα σύνολο χρονοσειρών ή την εξειδικευμένη επιλογή για κάθε μία χρονοσειρά ξεχωριστά.

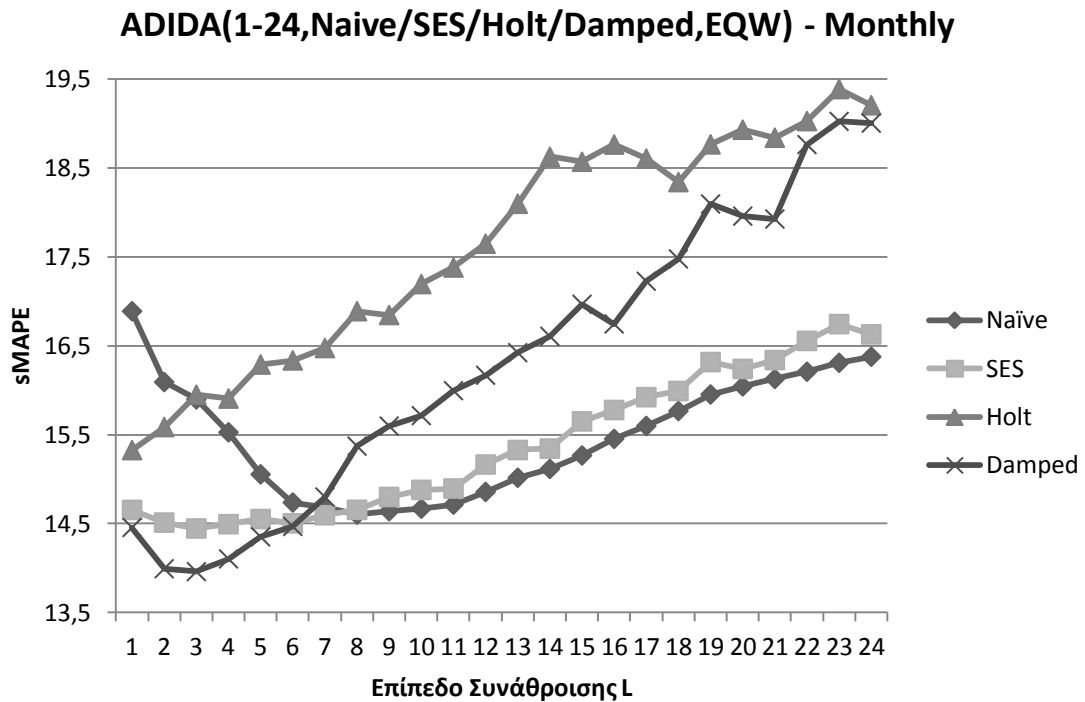
5.2 Κοινό Βέλτιστο Επίπεδο Συνάθροισης σε Σύνολο Χρονοσειρών

Ο απλούστερος τρόπος εφαρμογής της μεθοδολογίας ADIDA σε μεγάλα σύνολα χρονοσειρών είναι με τη χρήση κοινού επιπέδου συνάθροισης σε όλες τις χρονοσειρές. Έτσι, δεν είναι απαραίτητη η επιλογή διαφορετικού βέλτιστου επιπέδου συνάθροισης ανά χρονοσειρά, παρά αρκεί αν επιλεγεί ένα κοινό βέλτιστο επίπεδο για όλο το σύνολο των δεδομένων προς πρόβλεψη. Αν και η προσέγγιση αυτή θεωρητικά θα λειτουργεί αρκετά αποδοτικά σε ομογενή σύνολα δεδομένων, όπου δηλαδή όλες οι χρονοσειρές έχουν κοινή προέλευση και παρόμοια φύση, τα συμπεράσματα είναι μάλλον δύσκολο να γενικευθούν σε ετερογενή σύνολα. Ακόμα όμως και στην περίπτωση αυτή παρουσιάζει ιδιαίτερο ενδιαφέρον η μελέτη της χρήσης ενός κοινού και σταθερού L παρά την ανομοιογένεια των δεδομένων.

Στη συνέχεια, εφαρμόζουμε τη μεθοδολογία ADIDA στα δεδομένα του διαγωνισμού M3 και μετράμε την out-of-sample ακρίβεια των προβλέψεων με χρήση του δείκτη σφάλματος sMAPE. Η εφαρμογή γίνεται χωριστά για κάθε κατηγορία χρονικής δειγματοληψίας των δεδομένων (μηνιαία, τριμηνιαία, ετήσια, άλλα). Σε κάθε κατηγορία υπολογίζουμε ακριβώς το σφάλμα ακριβώς στο σύνολο του πλήθους των προβλέψεων που απαιτήθηκαν και από τον διαγωνισμό M3 (πίνακας 2.2). Το κάθε υποσύνολο είναι εσωτερικά ανομοιογενές, αφού περιέχει χρονοσειρές από ένα πλήθος τομέων ενδιαφέροντος (μικροοικονομία, βιομηχανία, μακροοικονομία, οικονομικά, δημογραφικά και λοιπά δεδομένα). Ελέγχουμε ένα πλήθος συνδυασμών διαφορετικών μεθόδων πρόβλεψης και επιπέδων συνάθροισης. Υπενθυμίζουμε ότι το επίπεδο συνάθροισης $L = 1$ ταυτίζεται με το αποτέλεσμα που προκύπτει χωρίς τη χρήση της μεθοδολογίας. Σε όλες τις περιπτώσεις χρησιμοποιείται η μέθοδος διάσπασης με ίσα βάρη.

5.2.1 Μηνιαία Δεδομένα

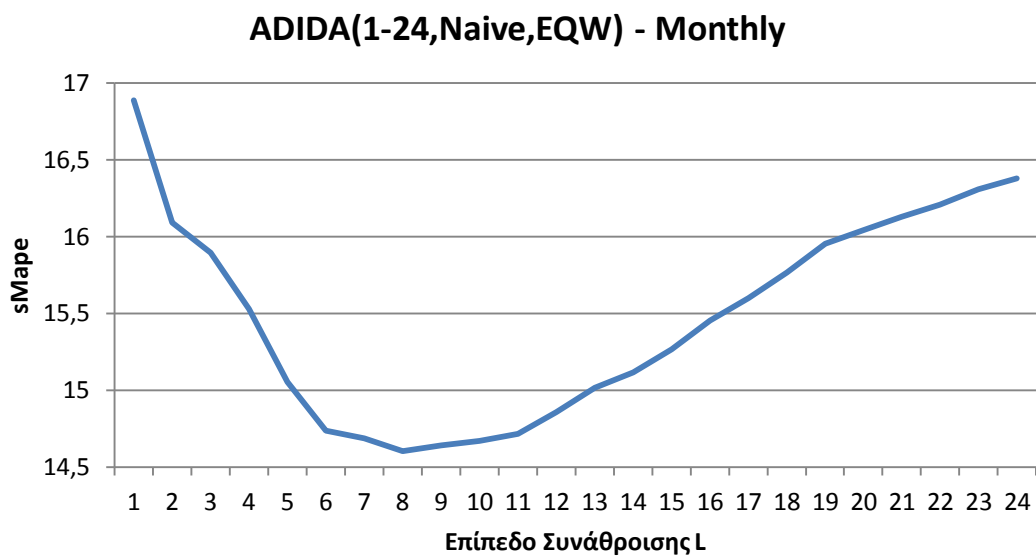
Στο συγκεντρωτικό διάγραμμα της εικόνας 5.1 απεικονίζουμε την επίδραση της εφαρμογής μεθοδολογίας ADIDA στο μέσο sMAPE out-of-sample σφάλμα πρόβλεψης ορίζοντα 18 περιόδων για διαφορετικές τιμές του κοινού επιπέδου συνάθροισης στο υποσύνολο των μηνιαίων δεδομένων. Ειδικότερα, χαράσσονται οι καμπύλες που αντιστοιχούν στην χρήση διαφορετικών μεθόδων προέκτασης και για όλες τις τιμές του επιπέδου συνάθροισης μέχρι την τιμή $L = 24$ (δύο χρόνια). Υπενθυμίζεται ότι η τιμή $L = 1$ αντιστοιχεί στην απλή εφαρμογή της μεθόδου, χωρίς χρήση της μεθοδολογίας ADIDA. Αποτελεί συνεπώς ένα σαφές επίπεδο αναφοράς για την αξιολόγηση της αποδοτικότητας της μεθόδου. Σε όλες τις εξεταζόμενες περιπτώσεις χρησιμοποιείται η μέθοδος διάσπασης με ίσα βάρη.



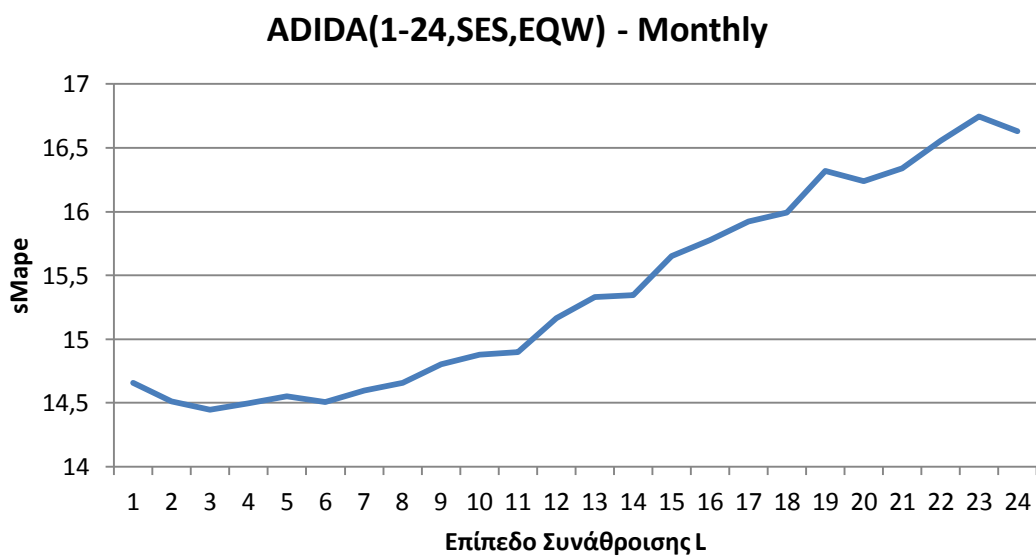
Εικόνα 5.1

Παρατηρούμε ότι η συμπεριφορά της μεθοδολογίας ως προς την απόδοσή της δίνει ποικίλα αποτελέσματα, ανάλογα με τη χρησιμοποιούμενη μέθοδο πρόβλεψης. Η σημαντική παρατήρηση είναι η επισήμανση των τιμών του επιπέδου συνάθροισης για τις οποίες το μέσο σφάλμα πέφτει υπό της στάθμης για $L = 1$, δηλαδή τις τιμές του επιπέδου συνάθροισης για τις οποίες η μεθοδολογία έχει πράγματι βελτιωτική δράση στην ακρίβεια πρόβλεψης, παρά την χρήση ενός κοινού επιπέδου για το ανομοιογενές σύνολο των μηνιαίων δεδομένων.

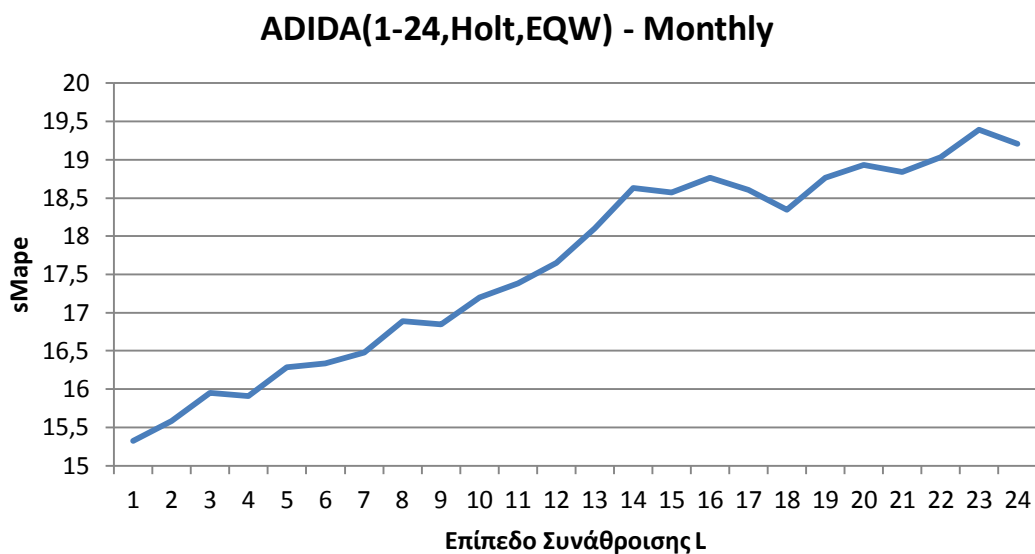
Στα ακόλουθα διαγράμματα απεικονίζονται μεμονωμένα τα μέσα sMAPE σφάλματα που προκύπτουν από τη χρήση διαφορετικών μεθόδων, ώστε να είναι περισσότερο εύκολα ορατή η εξάρτηση της ακρίβειας πρόβλεψης με το χρησιμοποιούμενο επίπεδο συνάθροισης.



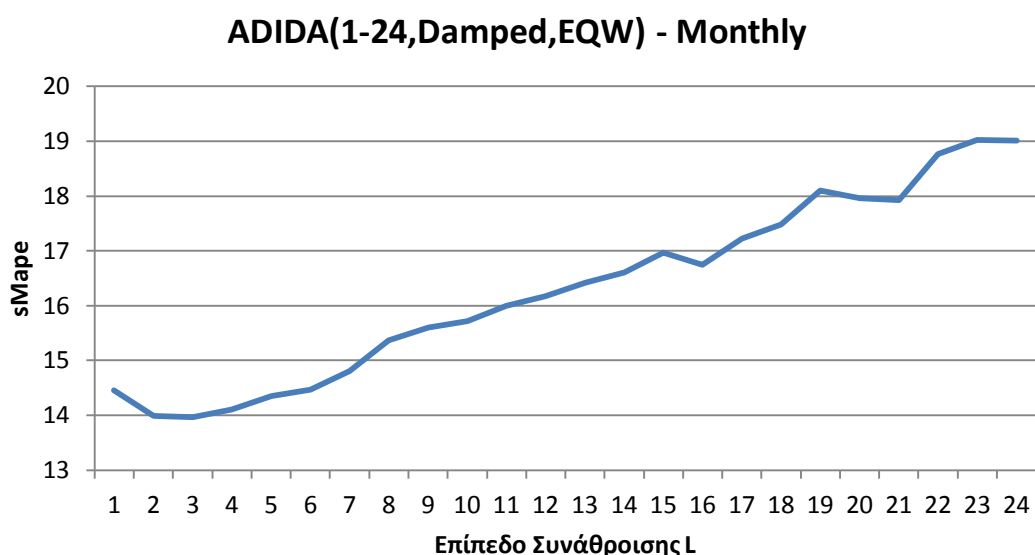
Εικόνα 5.2



Εικόνα 5.3



Εικόνα 5.4



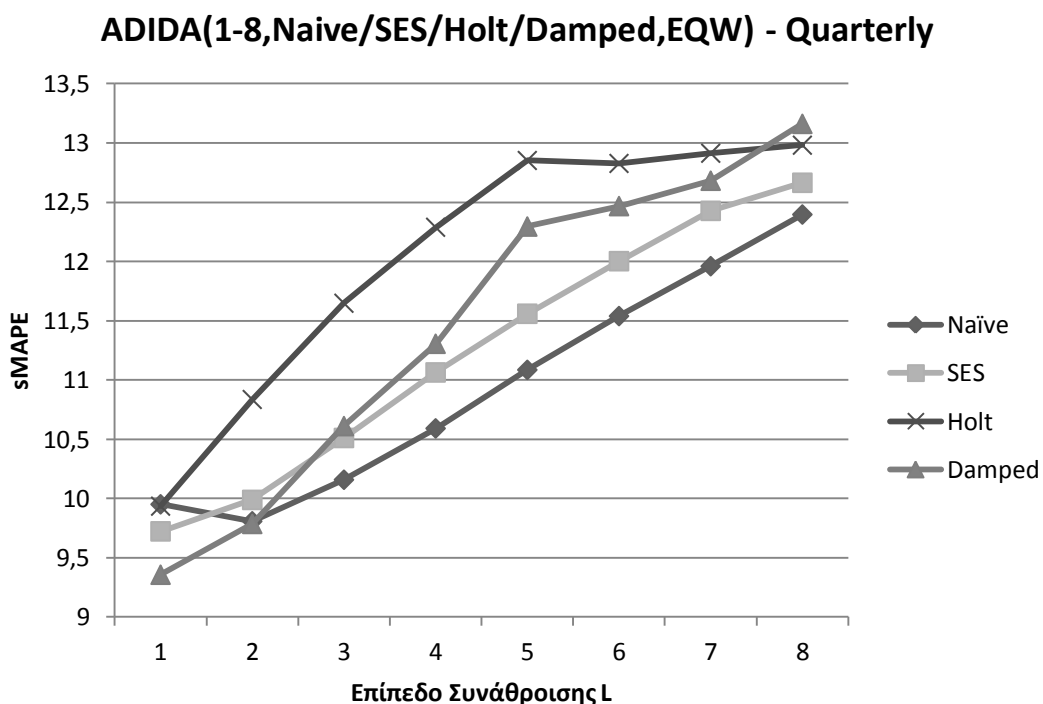
Εικόνα 5.5

Πράγματι, τα μεμονωμένα διαγράμματα απεικονίζουν με σαφή τρόπο την δραστική εξάρτηση του αποτελέσματος εφαρμογής της μεθοδολογίας από την εκάστοτε υλοποιούμενη μέθοδο. Για τις περισσότερες εξεταζόμενες μεθόδους παρουσιάζεται σε κάποιο βαθμό βελτίωση της ακρίβειας των προβλέψεων, ενώ υπάρχουν περιπτώσεις (π.χ. Holt) όπου η αύξηση του επιπέδου συνάθροισης συνδέεται με μόνιμη επιδείνωση. Το ποσοστό της βελτίωσης ή επιδείνωσης ποικίλει κατά περίπτωση, με αποτέλεσμα να μην διατηρείται η σειρά ταξινόμησης των μεθόδων σύμφωνα με το παραγόμενο σφάλμα. Γενικά, τυχόν βελτίωση εμφανίζεται για μικρές σχετικά τιμές του επιπέδου συνάθροισης και το σφάλμα σταδιακά

μειώνεται μέχρι ένα κατώτερο σημείο. Για παράδειγμα, με χρήση της Naïve έχουμε το βέλτιστο αποτέλεσμα για $L = 8$. Μετά την τιμή στην οποία παρατηρείται το ελάχιστο του σφάλματος, αρχίζει μία σταδιακή χειροτέρευση του ποσοστού βελτίωσης, η οποία συχνά μπορεί να οδηγήσει στην επιδείνωση του σφάλματος για μεγάλες τιμές του επιπέδου συνάθροισης. Τέλος, η σχετικά ομαλή μορφή των καμπύλων συνηγορεί υπέρ της πιθανής ύπαρξης ενός βέλτιστου επιπέδου συνάθροισης.

5.2.2 Τριμηνιαία Δεδομένα

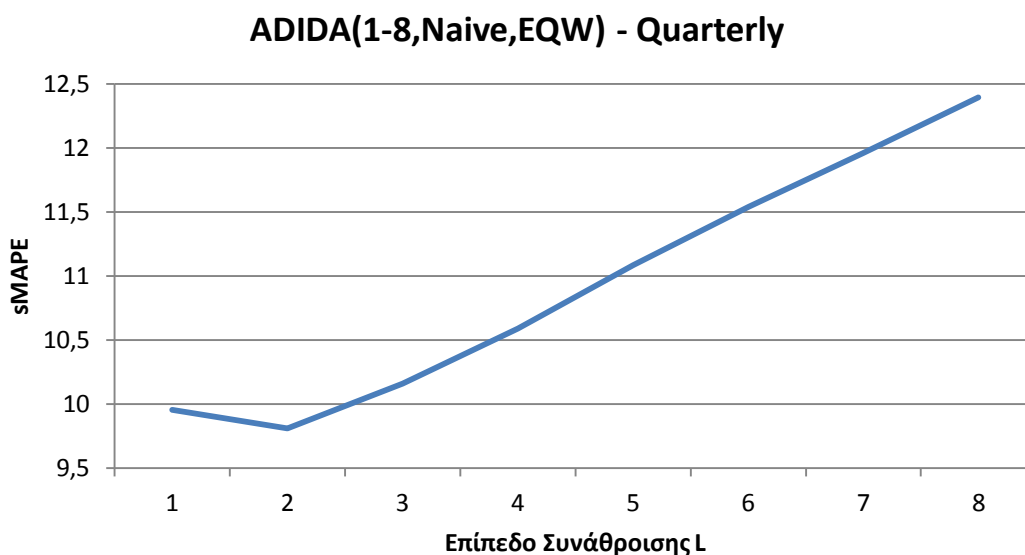
Στο ακόλουθο διάγραμμα της εικόνας 5.6 παριστούμε συγκεντρωτικά την επίδραση της εφαρμογής μεθοδολογίας ADIDA στο μέσο sMAPE out-of-sample σφάλμα πρόβλεψης ορίζοντα 8 περιόδων για διαφορετικές τιμές του κοινού επιπέδου συνάθροισης στο υποσύνολο των τριμηνιαίων δεδομένων. Συγκεκριμένα, παρουσιάζονται οι σχετικές καμπύλες που προκύπτουν από την εφαρμογή ποικίλων μεθόδων πρόβλεψης, για όλες τις τιμές του επιπέδου συνάθροισης μέχρι την τιμή $L = 8$ (δύο χρόνια) και για διάσπαση με ίσα βάρη. Στο σημείο αυτό κρίνεται σκόπιμη η υπενθύμιση ότι η τιμή $L = 1$ αντιστοιχεί στην απλή εφαρμογή της μεθόδου, χωρίς χρήση της μεθοδολογίας ADIDA. Για το λόγο αυτό, αποτελεί μία αναμφισβήτητη βάση αναφοράς για την αξιολόγηση της απόδοσης της μεθοδολογίας.



Εικόνα 5.6

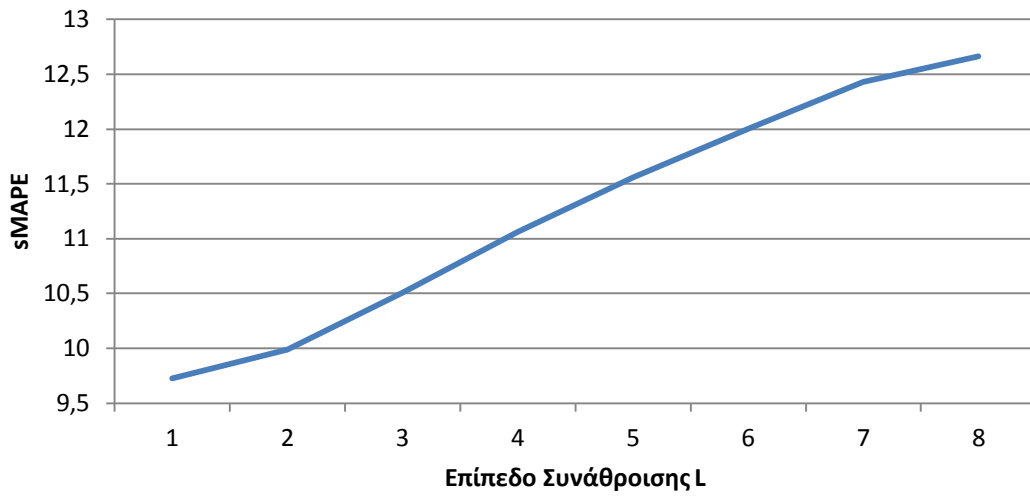
Η συμπεριφορά της μεθοδολογίας ως προς την ακρίβεια των προβλέψεων της δίνει ανόμοια αποτελέσματα, που εξαρτώνται άμεσα από τη χρησιμοποιούμενη μέθοδο προέκτασης. Αναζητώντας τιμές του επιπέδου συνάθροισης για τις οποίες το μέσο σφάλμα πέφτει υπό της στάθμης για $L = 1$, δηλαδή τις τιμές του επιπέδου συνάθροισης για τις οποίες η μεθοδολογία έχει όντως βελτιωτική δράση στην ακρίβεια της πρόβλεψης, μόνο η Naïve παρουσιάζει μία μικρή βελτίωση για $L = 2$. Το αποτέλεσμα αυτό έρχεται σε αντίθεση με τα αποτελέσματα για το σύνολο των μηνιαίων δεδομένων.

Για ευκρινέστερη παρατήρηση της εξάρτησης από το χρησιμοποιούμενο επίπεδο συνάθροισης, στα επόμενα διαγράμματα παρουσιάζονται μεμονωμένες οι καμπύλες των μέσων sMAPE σφαλμάτων που προκύπτουν με τη χρήση διαφορετικών μεθόδων.



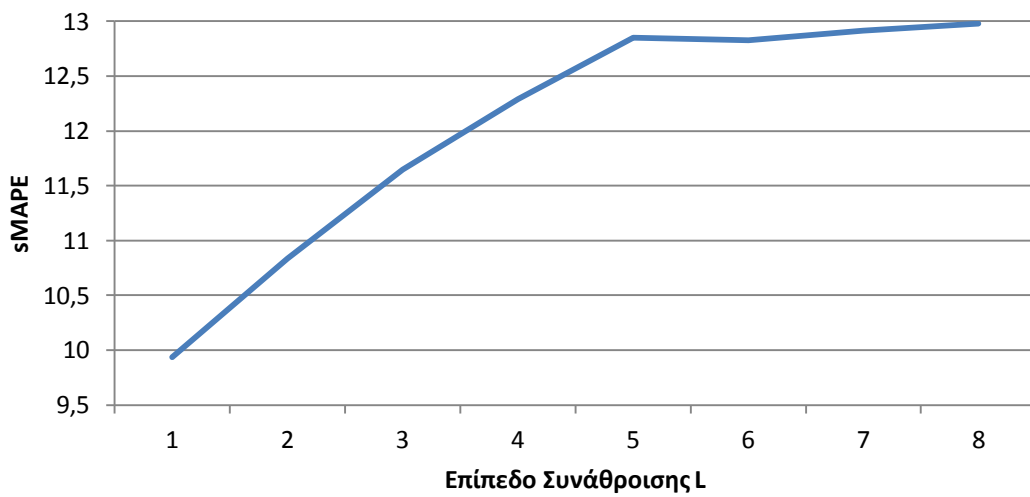
Εικόνα 5.7

ADIDA(1-8,SES,EQW) - Quarterly

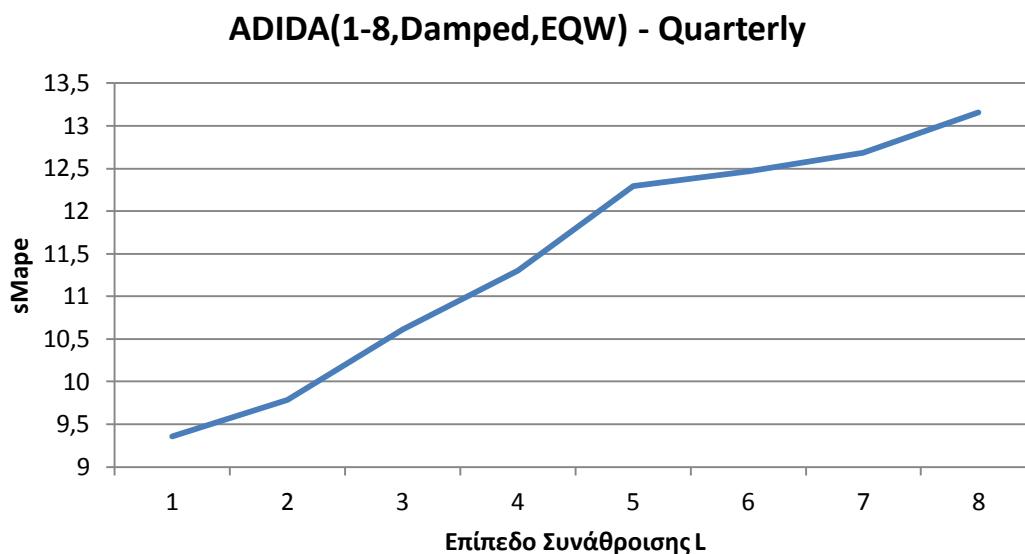


Εικόνα 5.8

ADIDA(1-8,Holt,EQW) - Quarterly



Εικόνα 5.9

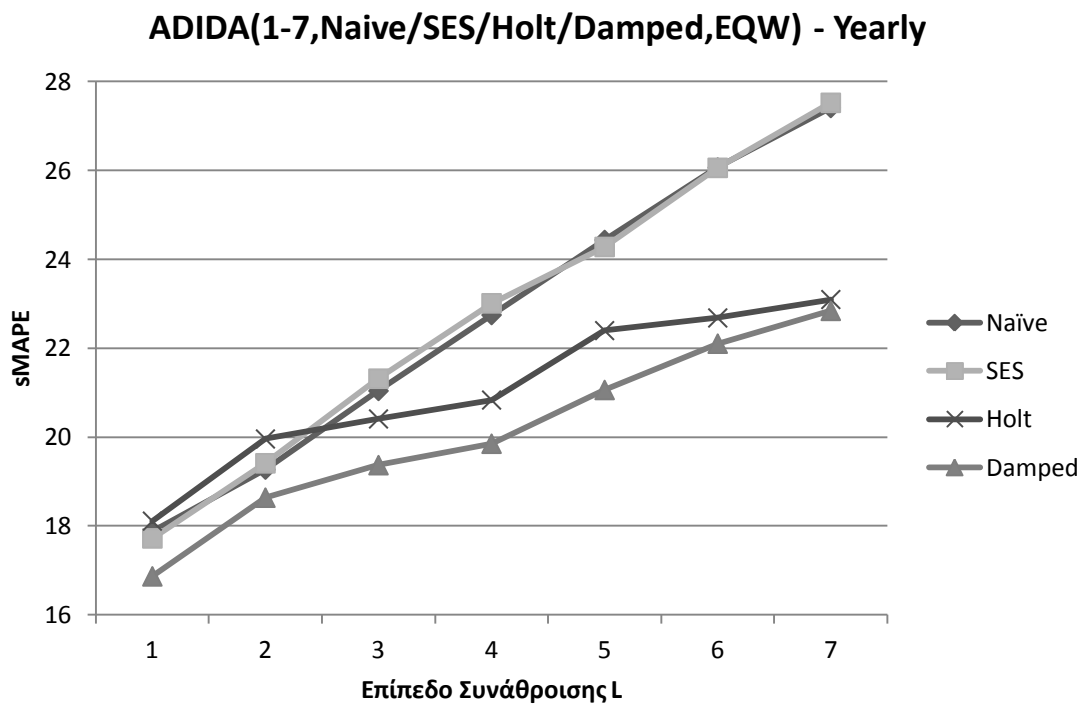


Εικόνα 5.10

Για τις περισσότερες εξεταζόμενες μεθόδους παρουσιάζεται μόνιμη επιδείνωση της ακρίβειας των προβλέψεων, με εξαίρεση της περίπτωσης της Naïve που οδηγεί σε μία μικρή βελτίωση, ώστε να προσεγγίζει την περίπτωση της εφαρμογής μεθόδου SES εκτός των πλαισίων της μεθοδολογίας ADIDA.

5.2.3 Ετήσια Δεδομένα

Στο συγκεντρωτικό διάγραμμα της εικόνας 5.11 απεικονίζεται η επίδραση της εφαρμογής μεθοδολογίας ADIDA στο μέσο sMAPE out-of-sample σφάλμα πρόβλεψης ορίζοντα 6 περιόδων για διαφορετικές τιμές του κοινού επιπέδου συνάθροισης στο υποσύνολο των ετήσιων δεδομένων. Ειδικότερα, εμφανίζονται οι καμπύλες για ποικίλες μεθόδους προέκτασης στο μέλλον, για όλες τις τιμές του επιπέδου συνάθροισης μέχρι την τιμή $L = 7$ (έτη) και διάσπαση με ίσα βάρη. Τα αποτελέσματα για την τιμή $L = 1$, που αντιστοιχεί σε απλή εφαρμογή της μεθόδου χωρίς χρήση μεθοδολογίας ADIDA, αποτελεί ένα επίπεδο αναφοράς για την αξιολόγηση της αποδοτικότητας της μεθόδου.

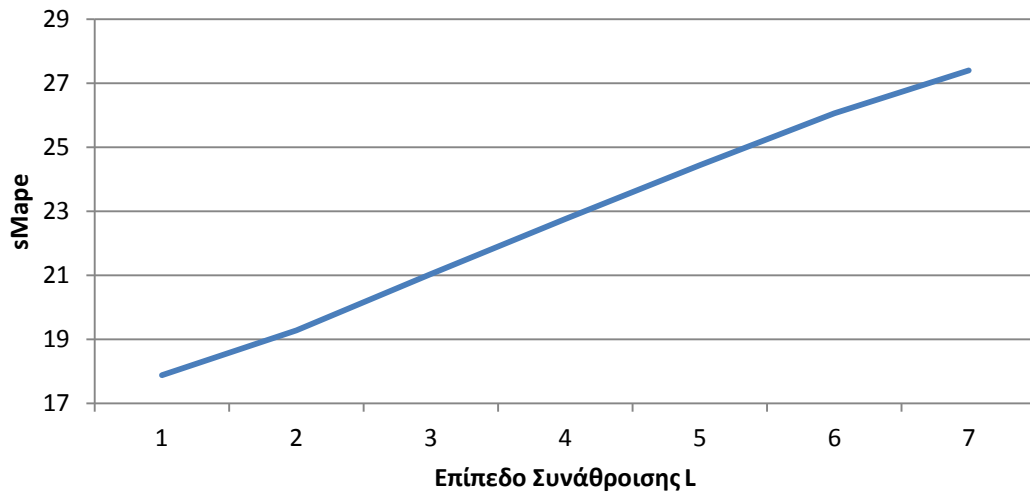


Εικόνα 5.11

Για όλες τις χρησιμοποιούμενες μεθόδους πρόβλεψης, παρουσιάζεται μία σταθερή επιδείνωση της ακρίβειας των παραγόμενων προβλέψεων. Σε καμία από τις εξεταζόμενες περιπτώσεις το μέσο σφάλμα δεν πέφτει υπό της στάθμης για $L = 1$, το οποίο θα υποδείκνυε τη βελτιωτική δράση της μεθοδολογίας στην ακρίβεια της πρόβλεψης για το συγκεκριμένο σύνολο των ετήσιων δεδομένων.

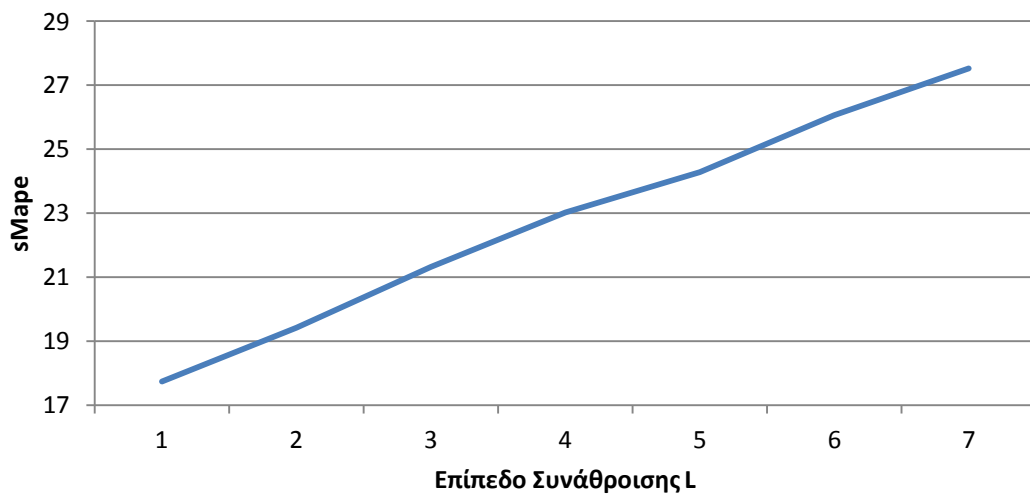
Στα ακόλουθα διαγράμματα παρουσιάζονται μεμονωμένα οι καμπύλες εξάρτησης των μέσων sMAPE σφαλμάτων από το χρησιμοποιούμενο επίπεδο συνάθροισης για τη χρήση διαφορετικών μεθόδων πρόβλεψης.

ADIDA(1-7,Naive,EQW) - Yearly

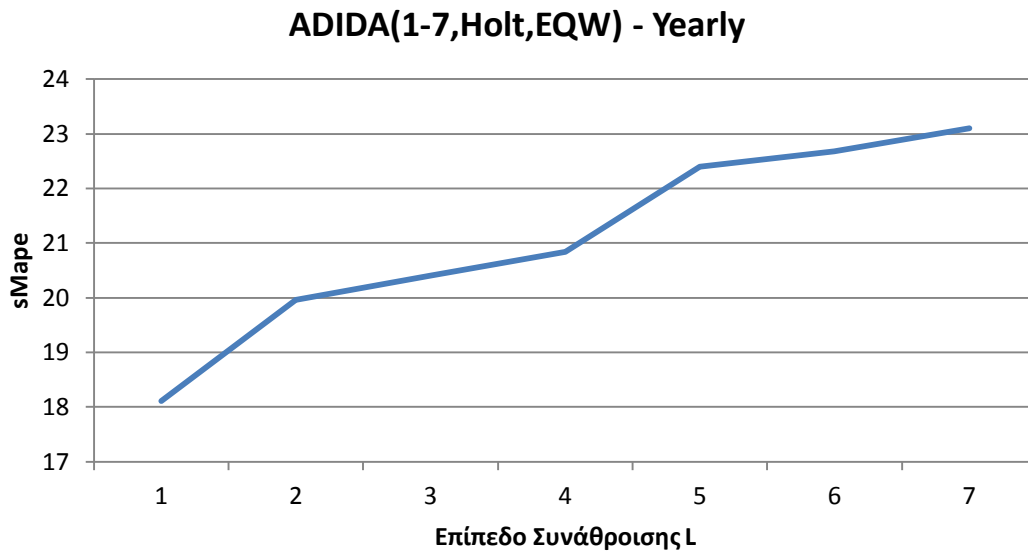


Εικόνα 5.12

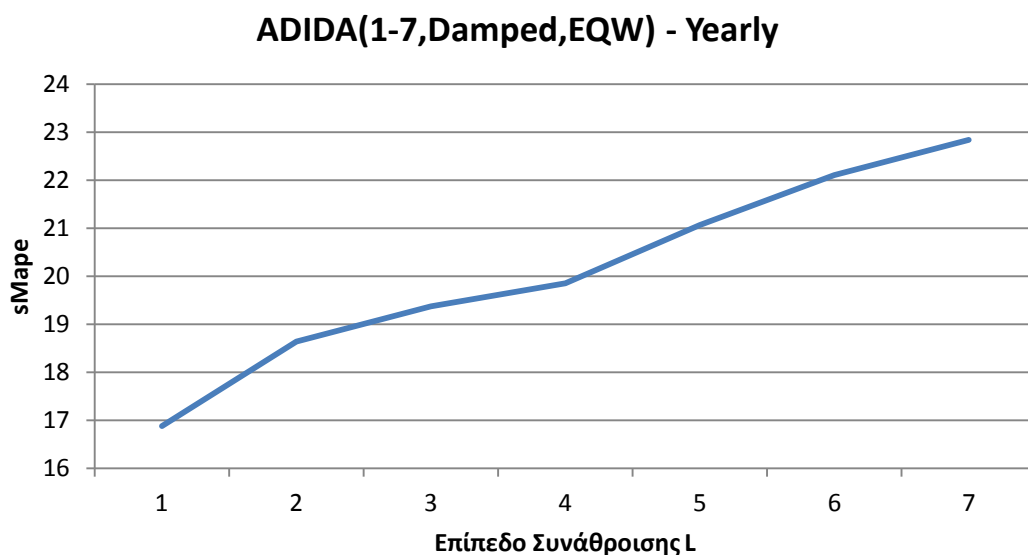
ADIDA(1-7,SES,EQW) - Yearly



Εικόνα 5.13



Εικόνα 5.14

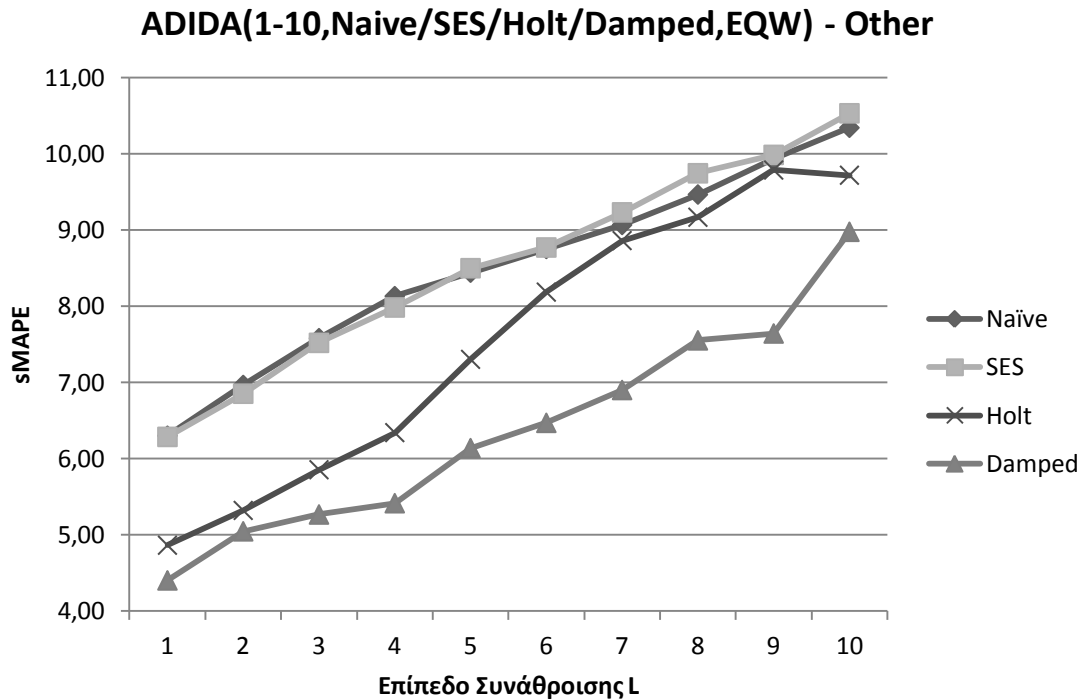


Εικόνα 5.15

5.2.4 Άλλα Δεδομένα

Στο συγκεντρωτικό διάγραμμα της εικόνας 5.16 απεικονίζουμε την επίδραση της εφαρμογής μεθοδολογίας ADIDA στο μέσο sMAPE out-of-sample σφάλμα πρόβλεψης ορίζοντα 8 περιόδων για διαφορετικές τιμές του κοινού επιπέδου συνάθροισης στο υποσύνολο των υπόλοιπων δεδομένων. Ειδικότερα, χαράσσονται οι καμπύλες που αντιστοιχούν στην χρήση διαφορετικών μεθόδων προέκτασης και για όλες τις τιμές του επιπέδου συνάθροισης μέχρι την τιμή $L = 10$. Χρησιμοποιούνται ίσα βάρη για τη διάσπαση των συναθροισμένων προβλέψεων. Η

απόδοση της μεθοδολογίας μπορεί να μετρηθεί με σύγκριση του σφάλματος σε κάθε επίπεδο με την αντίστοιχη τιμή για $L = 1$, η οποία αντιστοιχεί στην απλή εφαρμογή της μεθόδου, χωρίς χρήση της μεθοδολογίας ADIDA.

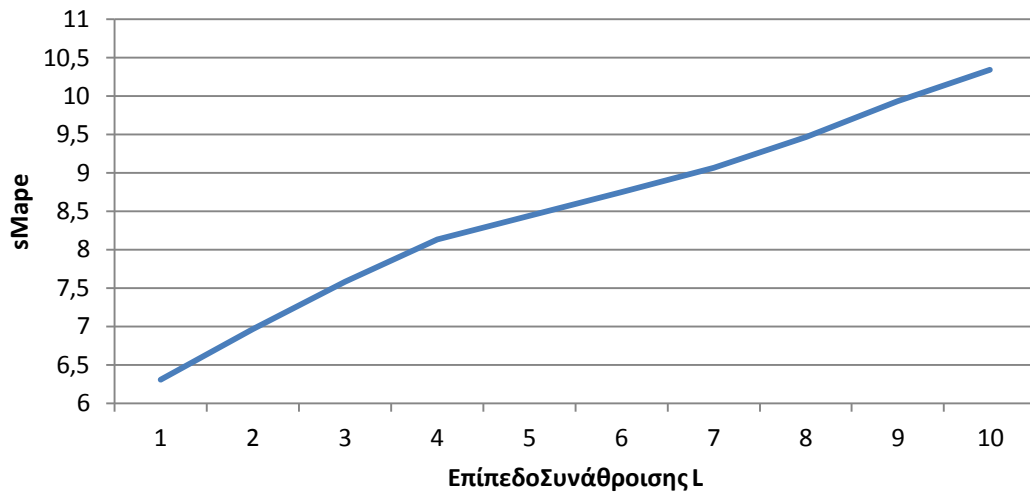


Εικόνα 5.16

Όμοια με την περίπτωση των ετήσιων δεδομένων, παρουσιάζεται μία σταθερή επιδείνωση της ακρίβειας των παραγόμενων προβλέψεων για όλες τις χρησιμοποιούμενες μεθόδους προέκτασης. Σε καμία από τις εξεταζόμενες περιπτώσεις το μέσο σφάλμα δεν πέφτει υπό της στάθμης του σφάλματος για $L = 1$, το οποίο θα υποδείκνυε βελτιωτική επίδραση της μεθοδολογίας στην ακρίβεια της πρόβλεψης για το συγκεκριμένο σύνολο των υπόλοιπων δεδομένων.

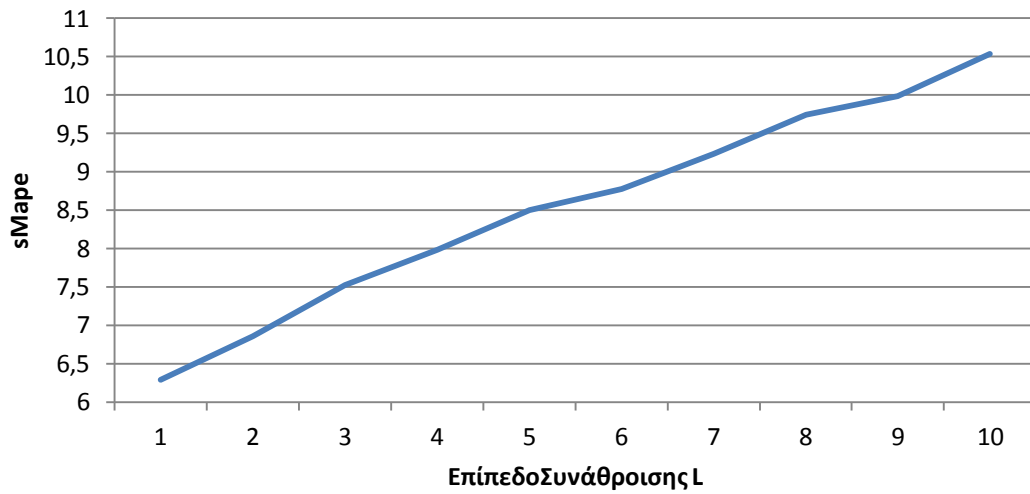
Για λόγους ευκρίνειας, οι καμπύλες εξάρτησης των μέσων sMAPE σφαλμάτων από το χρησιμοποιούμενο επίπεδο συνάθροισης για τη χρήση διαφορετικών μεθόδων πρόβλεψης απεικονίζονται μεμονωμένα στα επόμενα διαγράμματα.

ADIDA(1-10,Naive,EQW) - Other

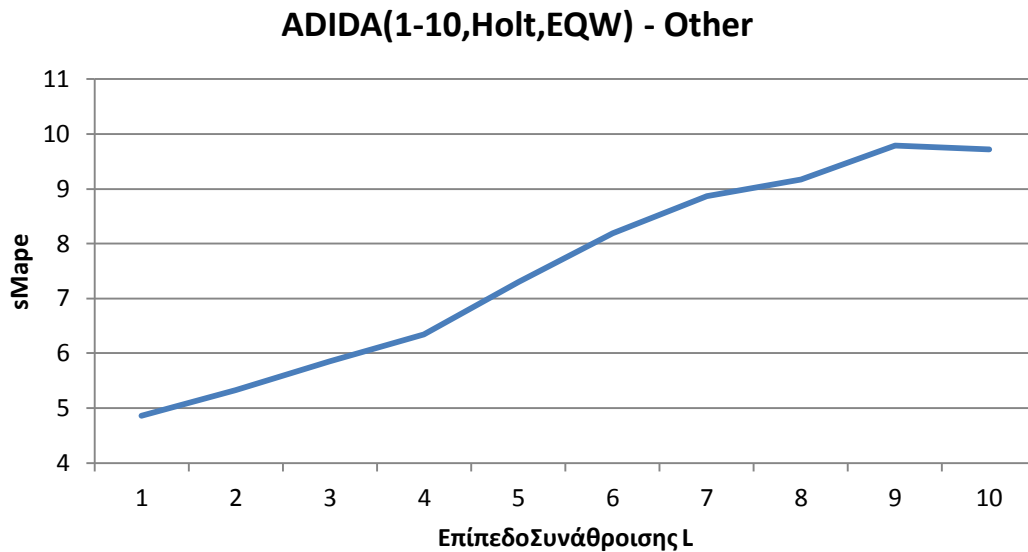


Εικόνα 5.17

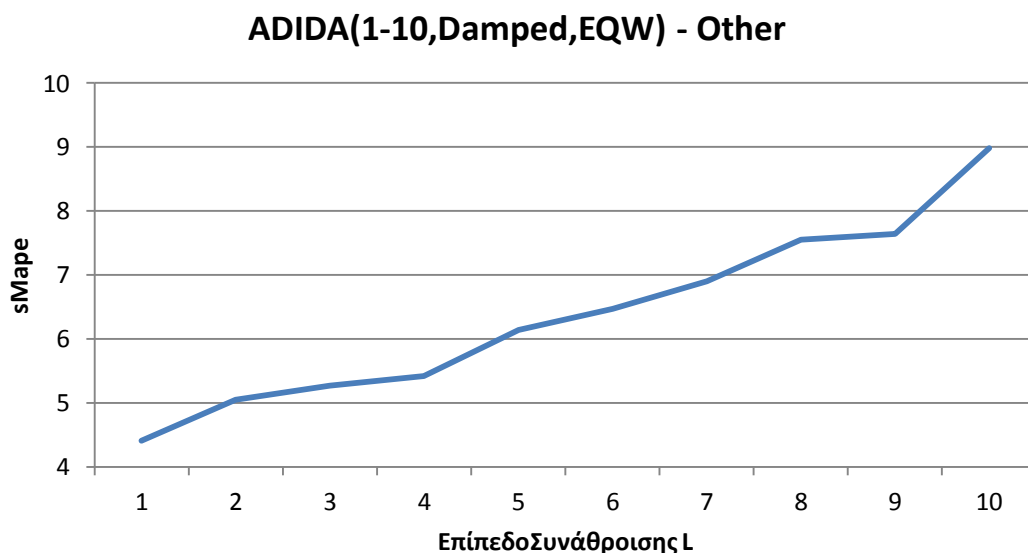
ADIDA(1-10,SES,EQW) - Other



Εικόνα 5.18



Εικόνα 5.19



Εικόνα 5.20

5.2.5 Σύνοψη Αποτελεσμάτων και Συμπεράσματα

Τα αποτελέσματα από την εφαρμογή της μεθοδολογίας ποικίλουν δραστικά ανάλογα με την κατηγορία των δεδομένων (μηνιαία, τριμηνιαία, ετήσια, άλλα), τις εξεταζόμενες μεθόδους πρόβλεψης και το επιλεγμένο επίπεδο συνάθροισης. Ανάλογα με τις παραπάνω επιλογές, παρατηρούμε μία ευρεία διακύμανση στην αποδοτικότητα της μεθόδου, που κυμαίνεται μεταξύ σημαντικών μειώσεων των σφαλμάτων για τις βέλτιστες επιλογές και μεγάλων επιδεινώσεων για μη κατάλληλους συνδυασμούς παραμέτρων της μεθόδου.

Τα σημαντικότερα ευρήματα προέρχονται από το σύνολο των μηνιαίων δεδομένων. Αποτελέσματα για την εφαρμογή της μεθοδολογίας στο σύνολο αυτό περιγράφονται συνοπτικά στον πίνακα 5.1. Συγκεκριμένα, περιλαμβάνεται μία στήλη για με τα υπολογισμένα σφάλματα των μεθόδων στον M3 διαγωνισμό και μία στήλη με τα αντίστοιχα σφάλματα από την επανεκτέλεση των πειραμάτων του διαγωνισμού, η οποία αντιστοιχεί στην εφαρμογή της μεθοδολογίας ADIDA για μοναδιαίο επίπεδο συνάθροισης. Τυχόν αποκλίσεις οφείλονται σε διαφορετικές αρχικοποιήσεις και βελτιστοποιήσεις, αν και τα αποτελέσματα είναι παραπλήσια. Η μικρότερη τιμή του σφάλματος ανάμεσα στις δύο εκτελέσεις εικονίζεται σε έντονα γράμματα και αποτελεί το επίπεδο σφάλματος κάτω από το οποίο θεωρούμε ότι η μεθοδολογία προσέφερε βελτίωση. Δύο επιπλέον στήλες περιγράφουν τα βέλτιστα αποτελέσματα σε όρους σφάλματος από την εφαρμογή της μεθοδολογίας ADIDA. Ειδικότερα, στη μία στήλη σημειώνεται η ελάχιστη τιμή που επιτεύχθηκε με χρήση της μεθοδολογίας και στην επόμενη το αντίστοιχο βέλτιστο επίπεδο συνάθροισης για το οποίο παρατηρήθηκε. Τα αποτελέσματα απεικονίζονται σε σκιασμένο φόντο όταν εμφανίζεται βελτίωση. Τέλος, κάθε γραμμή είναι αφιερωμένη και σε διαφορετική υπό εξέταση μέθοδο. Τα αποτελέσματα όλων των μεθόδων έχουν προκύψει για διάσπαση με ίσα βάρη.

Μηνιαία δεδομένα/Ορίζοντας 1-18

Method	sMAPE (%)			
	M3	Replication (L=1)	Best ADIDA	Best L
Naive2	16,91	16,89	14,60	8
Single	15,32	14,65	14,45	3
Holt	15,36	15,33	15,33	1
Dampen	14,59	14,46	13,96	3

Πίνακας 5.1

Όπως φαίνεται στον παραπάνω πίνακα, υπάρχουν περιπτώσεις στις οποίες η μεθοδολογία οδηγεί σε σημαντική βελτίωση του σφάλματος. Ειδικότερα, με χρήση Naïve στα πλαίσια της μεθοδολογίας, το σφάλμα βελτιώνεται από το 16,89% στο 14,60% για επίπεδο συνάθροισης $L = 8$. Η νέα τιμή του σφάλματος είναι μειωμένη κατά 13,55% και άμεσα συγκρίσιμη με το επίπεδο σφάλματος της Damped. Το αποτέλεσμα αυτό είναι ενδιαφέρον, δεδομένης της ταχύτητας και της απλότητας της μεθόδου Naïve, η οποία δεν απαιτεί τη βελτιστοποίηση παραμέτρων, εξαλείφοντας την πιθανότητα κακής βελτιστοποίησης η οποία συνδέεται οι παραμετρικές μέθοδοι. Μικρότερη μείωση του σφάλματος (1,42%) παρατηρείται με χρήση της μεθόδου SES. Για τη μέθοδο Damped, μία μείωση 3,44% φέρνει το σφάλμα σε παρόμοια επίπεδα με εκείνα των νικητριών μεθόδων του M3

διαγωνισμού (π.χ. 13,85% για τη Theta). Από την άλλη μεριά, καμία βελτίωση δεν παρατηρείται στην περίπτωση της μεθόδου Holt. Οι βελτιώσεις αυτές, όπου εμφανίζονται, είναι ιδιαίτερα σημαντικές και επιβεβαιώνουν τον χαρακτηρισμό της μεθοδολογίας ADIDA ως μηχανισμού αυτοβελτίωσης των μεθόδων πρόβλεψης. Αξιοπρόσεκτο είναι ότι οι βελτιώσεις εμφανίζονται παρά τη χρήση κοινού επιπέδου συνάθροισης στο εξαιρετικά ανομοιογενές σύνολο των μηνιαίων δεδομένων.

Ακολούθως, συγκεντρώνουμε τα αντίστοιχα στοιχεία για το σύνολο των τριμηνιαίων δεδομένων στο πίνακα 5.2.

Τριμηνιαία δεδομένα/Ορίζοντας 1-8

Method	sMAPE (%)			
	M3	Replication (L=1)	Best ADIDA	Best L
Naive2	9,95	9,95	9,81	2
Single	9,72	9,72	9,72	1
Holt	10,67	9,93	9,93	1
Dampen	9,33	9,36	9,36	1

Πίνακας 5.2

Στην περίπτωση των τριμηνιαίων δεδομένων, δεν παρουσιάζονται εξίσου θεαματικά αποτελέσματα. Μία μικρή μείωση του σφάλματος (1,44%) εμφανίζεται μόνο για χρήση της μεθόδου Naïve. Για τις υπόλοιπες μεθόδους δεν παρατηρείται κάποια βελτίωση.

Στη συνέχεια παραθέτουμε τα αντίστοιχα στοιχεία για το σύνολο των ετήσιων και υπόλοιπων δεδομένων στους πίνακες 5.3 και 5.4, αντίστοιχα.

Ετήσια δεδομένα/Ορίζοντας 1-6

Method	sMAPE (%)			
	M3	Replication (L=1)	Best ADIDA	Best L
Naive2	17,88	17,88	17,88	1
Single	17,82	17,72	17,72	1
Holt	19,27	18,11	18,11	1
Dampen	17,18	16,87	16,87	1

Πίνακας 5.3

Άλλα δεδομένα/Ορίζοντας 1-8

Method	sMAPE (%)			
	M3	Replication (L=1)	Best ADIDA	Best L
Naive2	6,30	6,30	6,30	1
Single	6,29	6,29	6,29	1
Holt	4,81	4,86	4,86	1
Dampen	4,61	4,40	4,40	1

Πίνακας 5.4

Για τα ετήσια και τα υπόλοιπα δεδομένα, η εφαρμογή της μεθοδολογίας δεν επιφέρει καμία βελτίωση ως προς τα σφάλματα.

Συμπερασματικά, η εφαρμογή της μεθοδολογίας με χρήση κοινού επιπέδου συνάθροιση L στο σύνολο των δεδομένων παρουσιάζει αποτελέσματα που ποικίλουν, ανάλογα με την κατηγορία ως προς το διάστημα δειγματοληψίας των δεδομένων και την επιλεγμένη μέθοδο πρόβλεψης. Ιδιαίτερα αξιοπρόσεκτο είναι το γεγονός ότι βελτιώσεις εμφανίζονται παρά την εξαιρετική ανομοιογένεια των διαθέσιμων δεδομένων, αφού περιέχουν χρονοσειρές από πλήθος διαφορετικών κατηγοριών. Η εκπληκτική ομοιότητα της καμπύλης του της εικόνας 5.2, για τη εφαρμογή της μεθόδου Naïve στα μηνιαία δεδομένα, με την αντίστοιχη καμπύλη της μελέτης για δεδομένα διακοπτόμενης ζήτησης υποστηρίζει τα αποτελέσματα των Nikolopoulos et al. (2011) και γενικεύει τις ιδιότητες της μεθοδολογίας ADIDA ως μηχανισμού αυτοβελτίωσης και για την περίπτωση των δεδομένων συνεχούς ζήτησης. Ιδιαίτερα σημαντικές δυνατότητες βελτίωσης της απόδοσης προκύπτουν για τις περιπτώσεις δεδομένων συχνής δειγματοληψίας (μηνιαία δεδομένα). Τα περιθώρια της βελτίωσης μειώνονται όσο αντιμετωπίζονται δεδομένα αραιότερης δειγματοληψίας (π.χ. τριμηνιαία δεδομένα), ενώ τελικά πρακτικά εξαλείφονται (ετήσια δεδομένα). Γενικά, η μεθοδολογία παρουσιάζει δυνατότητες βελτίωσης των τεχνικών προβλέψεων. Η μεγαλύτερες μειώσεις σφάλματος εμφανίζονται για την μέθοδο Naïve. Διαισθητικά λοιπόν μπορούμε να παρατηρήσουμε ότι η βελτίωση που προσφέρει η μεθοδολογία είναι μεγαλύτερη για τις εξ αρχής λιγότερο αποδοτικές μεθόδους.

5.3 Βέλτιστο Επίπεδο Συνάθροισης Ξεχωριστά ανά Χρονοσειρά

Εναλλακτικά στη χρήση ενός κοινού επιπέδου συνάθροισης για όλο το σύνολο δεδομένων, η μεθοδολογία μπορεί επίσης να εφαρμοστεί ώστε να επιλέγεται διαφορετικό επίπεδο συνάθροισης ανά χρονοσειρά. Στην περίπτωση αυτή, η απλούστερη προσέγγιση για την μαθηματική διατύπωση και επίλυση του προβλήματος της επιλογής είναι ως εξής: Δεδομένων χρονοσειράς y , μοντέλου

πρόβλεψης \hat{y} και κριτηρίου J , επιθυμούμε να βρούμε το επίπεδο συνάθροισης L για το οποίο ελαχιστοποιείται το κριτήριο J . Συμβολικά, έχουμε λοιπόν το πρόβλημα:

$$L = \arg \min_L \{J(y, \hat{y})\} \text{ με } y \rightarrow ADIDA(L, F, D) \rightarrow \hat{y}$$

Ως προς το κριτήριο J , αυτό πρέπει να αποτελεί ένα μέτρο σφάλματος της προσαρμογής του μοντέλου ή κάποιο εναλλακτικό κριτήριο προσαρμογής. Στα πλαίσια της παρούσας εργασίας θα μελετηθούν οι περιπτώσεις όπου ως κριτήριο λαμβάνεται: το μέσο τετραγωνικό σφάλμα (MSE), το κριτήριο πληροφορίας του Bayes (Bayesian Information Criterion – BIC) και το κριτήριο πληροφορίας του Akaike (Akaike Information Criterion – AIC). Και στις τρεις περιπτώσεις τα κριτήρια υπολογίζονται στο σύνολο των διαθέσιμων δεδομένων του in-sample.

Γενικά, με χρήση μόνο των διαθέσιμων (in-sample) δεδομένων μπορούν να εκτιμηθούν οι βέλτιστες τιμές του επιπέδου συνάθροισης μέσω κάποιας κατάλληλης μεθόδου εκτίμησης. Η απόδοση της μεθόδου εκτίμησης κρίνεται από το κατά πόσον η εκτιμώμενη βέλτιστη τιμή ταυτίζεται με την πραγματική βέλτιστη τιμή, η οποία εξαρτάται από τα out-of-sample δεδομένα. Θα αναφερόμαστε στην διαδικασία εκτίμησης της τιμής με χρήση μόνο του in-sample τμήματος της χρονοσειράς ως εκ των προτέρων εκτίμηση (a priori), αφού πραγματοποιείται μόνο με τα μέχρι στιγμής διαθέσιμα δεδομένα. Αντίθετα, μπορούμε να πραγματοποιήσουμε μία εκ των υστέρων (a posteriori) εκτίμηση του βέλτιστου επιπέδου συνάθροισης λαμβάνοντας υπόψη και τις out-of-sample παρατηρήσεις. Ωστόσο, αν και έτσι υπολογίζουμε ουσιαστικά το πραγματικό βέλτιστο, η πρακτική αυτή δεν είναι υλοποιήσιμη σε πραγματικές περιπτώσεις, καθώς απαιτεί δεδομένα που δεν έχουν γίνει ακόμα διαθέσιμα. Η σύγκριση a priori εκτιμήσεων και a posteriori πραγματικών τιμών μπορεί να χρησιμοποιηθεί για την αποτίμηση της απόδοσης της μεθόδου εκτίμησης. Επίσης, οι a posteriori πραγματικές τιμές αποτελούν ένα σαφές άνω όριο της επιτεύξιμης βελτίωσης που μπορεί να επιφέρει η εφαρμογή της μεθοδολογίας.

5.3.1 Μέσο Τετραγωνικό Σφάλμα – MSE

Το κριτήριο αυτό δεν αποτελεί τίποτε άλλο παρά τη γνωστή έκφραση του μέσου τετραγωνικού σφάλματος υπολογισμένου σε όλο το μήκος των in-sample παρατηρήσεων της χρονοσειράς.

$$MSE = \frac{1}{N} \sum_{t=1}^N e_t^2 = \frac{1}{N} \sum_{t=1}^N (Y_t - \hat{Y}_t)^2$$

Το μέσο τετραγωνικό σφάλμα αποτελεί ένα από τα ευρύτερα χρησιμοποιούμενα μέτρα της προσαρμογής ενός μοντέλου στα διαθέσιμα δεδομένα. Η απλή του μορφή προσφέρεται για εκτενή μαθηματική ανάλυση, ενώ έχει αποτελέσει τη βάση για πολλά άλλα πολυπλοκότερα μέτρα. Όπως μαρτυρεί και ο τύπος του μέσου τετραγωνικού σφάλματος, αυτό σχετίζεται στενά με την διακύμανση των σφαλμάτων.

Το μέσο τετραγωνικό σφάλμα μπορεί να χρησιμοποιηθεί για την επιλογή μεταξύ στατιστικών μοντέλων, επιλέγοντας το μοντέλο που επιστρέφει την μικρότερη τιμή τους σφάλματος. Υπό την έννοια αυτή, το MSE αποτελεί μέτρο του βαθμού στον οποίο τα υπό εξέταση μοντέλα ερμηνεύουν τα δεδομένα.

5.3.2 Κριτήριο Πληροφορίας του Bayes – BIC

Το κριτήριο αυτό αποτελεί θεμέλιο της στατιστικής επιλογής μοντέλων. Συχνά αναφέρεται και ως κριτήριο του Schwarz (Schwarz, 1978) και συνδέεται στενά με το κριτήριο πληροφορίας του Akaike. Χρησιμοποιείται για την επιλογή μεταξύ διαφορετικών στατιστικών μοντέλων, επιλέγοντας κάθε φορά το μοντέλο εκείνο που επιστρέφει τη χαμηλότερη τιμή του κριτηρίου.

Ουσιαστικά, αντιστοιχεί στο κριτήριο της μέγιστης πιθανοφάνειας με έναν επιπρόσθετο όρο ποινής που αυξάνεται με την τάξη του μοντέλου k , στη βάση της φειδωλότητας ως προς τις παραμέτρους του μοντέλου. Μοντέλα υψηλής τάξης τείνουν να υπερπροσαρμόζονται στα δεδομένα εκπαίδευσής τους, με αποτέλεσμα να υστερούν στην δυνατότητα γενίκευσης. Επιπλέον, τα μικρότερα μοντέλα έχουν λιτότερες υπολογιστικές απαιτήσεις. Η ποινή που αποδίδεται για το μέγεθος του μοντέλου είναι μεγαλύτερη στο BIC από ό,τι στο AIC. Υπό την προϋπόθεση ότι τα σφάλματα του μοντέλου είναι κανονικά και ανεξάρτητα, το BIC μπορεί να υπολογιστεί (εκτός μίας σταθεράς, η οποία όμως δεν επηρεάζει τη διαδικασία επιλογής) από τον τύπο:

$$BIC = N \ln(MSE) + k \ln(N)$$

Η παραπάνω έκφραση κάνει χρήση του μέσου τετραγωνικού σφάλματος. Και πάλι, το N αποτελεί το πλήθος των in-sample παρατηρήσεων, ενώ η τάξη k του μοντέλου, δηλαδή το πλήθος των παραμέτρων, είναι ίσο με το L , αφού ακριβώς τόσες παρατηρήσεις της χρονοσειράς συμμετέχουν στο μοντέλο ADIDA, μέσω των ΚΜΟ.

5.3.3 Κριτήριο Πληροφορίας του Akaike – AIC

Το AIC (Akaike, 1974) συνιστά ένα κριτήριο παρόμοιο με το BIC, που όμως προσδίδει μεγαλύτερη ποινή στα μοντέλα υψηλότερης τάξης. Η ποινή αποτρέπει την υπερπροσαρμογή ενός μεγάλου μοντέλου στα δεδομένα εκπαίδευσης, κατάσταση που υποβαθμίζει τις ικανότητες γενίκευσης του μοντέλου. Η λειτουργία του κριτηρίου στηρίζεται στην έννοια της εντροπίας της πληροφορίας, προσφέροντας πρακτικά ένα σχετικό μέτρο της χαμένης πληροφορίας από τη χρήση ενός μοντέλου για την περιγραφή πραγματικών δεδομένων. Και πάλι με την προϋπόθεση των κανονικών και ανεξάρτητων υπολοίπων του μοντέλου, το AIC εκφράζεται μέσω της σχέσης:

$$AIC = N \ln(MSE) + 2k$$

Δεδομένου ενός συνόλου δεδομένων, ορισμένα υποψήφια μοντέλα διατάσσονται σύμφωνα με τις επιστρεφόμενες τιμές του κριτηρίου. Στη συνέχεια, επιλέγονται ως καλύτερα τα μοντέλα με την χαμηλότερη επιστρεφόμενη τιμή του κριτηρίου.

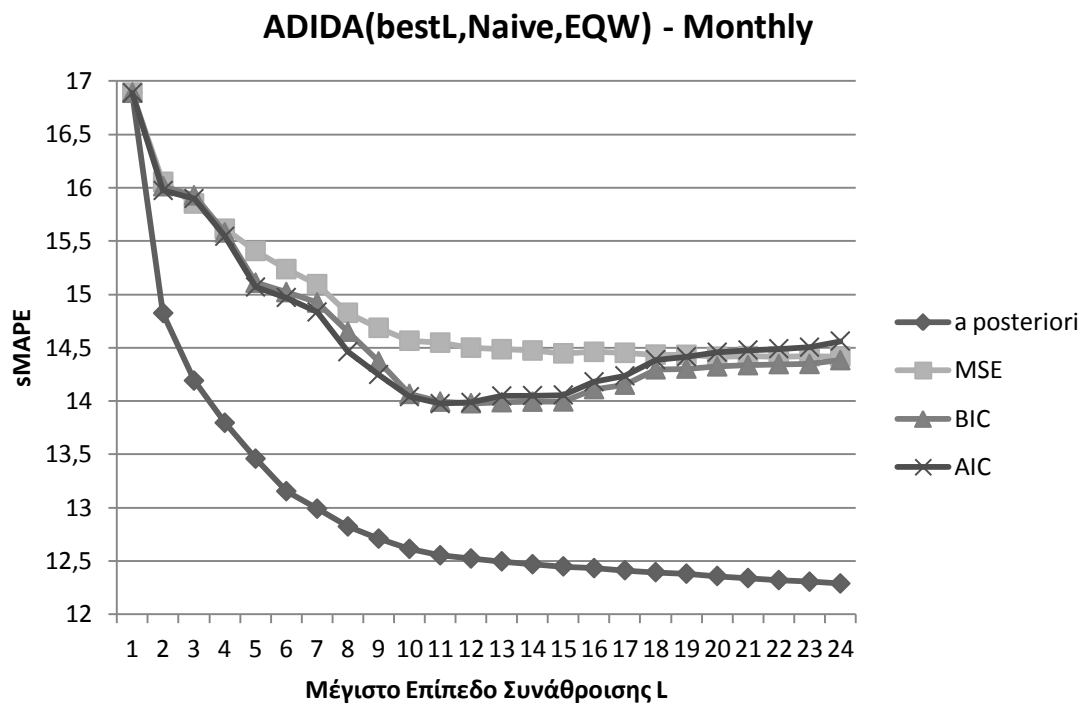
5.3.4 Μηνιαία Δεδομένα

Στα ακόλουθα διαγράμματα απεικονίζεται η απόδοση σε όρους μέσου sMAPE out-of-sample σφάλματος πρόβλεψης ορίζοντα 18 περιόδων για τις διάφορες τεχνικές επιλογής του επιπέδου συνάθροισης L . Τα αποτελέσματα αναφέρονται στα μηνιαία δεδομένα. Ο οριζόντιος άξονας αντιπροσωπεύει την ανώτερη τάξη επιπέδου συνάθροισης μέχρι την οποία εξετάζουμε τα μοντέλα ADIDA για την επιλογή του κατά περίπτωση βέλτιστου.

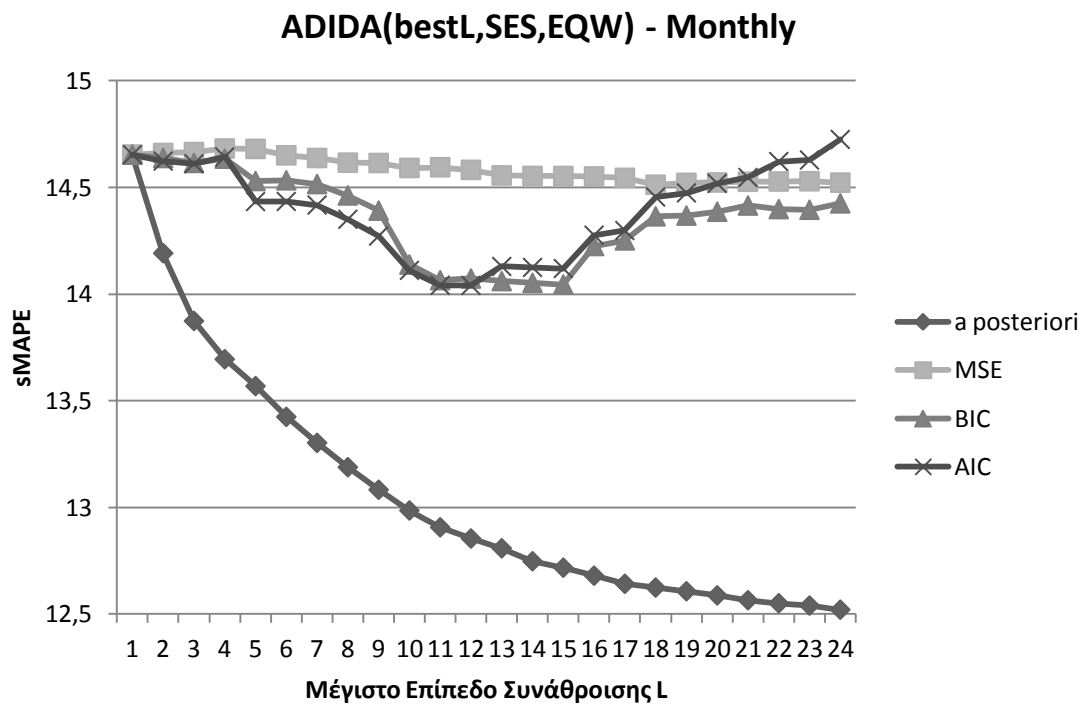
Διαφορετικές καμπύλες αντιστοιχούν σε διαφορετικές τεχνικές επιλογής του βέλτιστου επιπέδου, το οποίο υπολογίζεται εξατομικευμένα για κάθε χρονοσειρά. Συγκεκριμένα, μελετώνται οι τεχνικές επιλογής βασιζόμενες στην ελαχιστοποίηση των MSE, BIC και AIC. Παράλληλα, μία επιπλέον γραμμή αντιστοιχεί στο σφάλμα που θα είχαμε αν γνωρίζαμε απευθείας το a posteriori βέλτιστο για κάθε χρονοσειρά επίπεδο συνάθροισης. Η καμπύλη αυτή υποδεικνύει και την οριακή τιμή στην οποία μπορεί να βελτιωθεί το σφάλμα.

Παρέχονται ξεχωριστά διαγράμματα για κάθε εξεταζόμενη μέθοδο πρόβλεψης. Η απόδοση της μεθοδολογίας μπορεί να αξιολογηθεί με αναφορά στο παρατηρούμενο σφάλμα για μοναδιαίο επίπεδο συνάθροισης, αφού αντιστοιχεί σε απλή χρήση της μεθόδου πρόβλεψης χωρίς εφαρμογή της μεθοδολογίας ADIDA. Σε

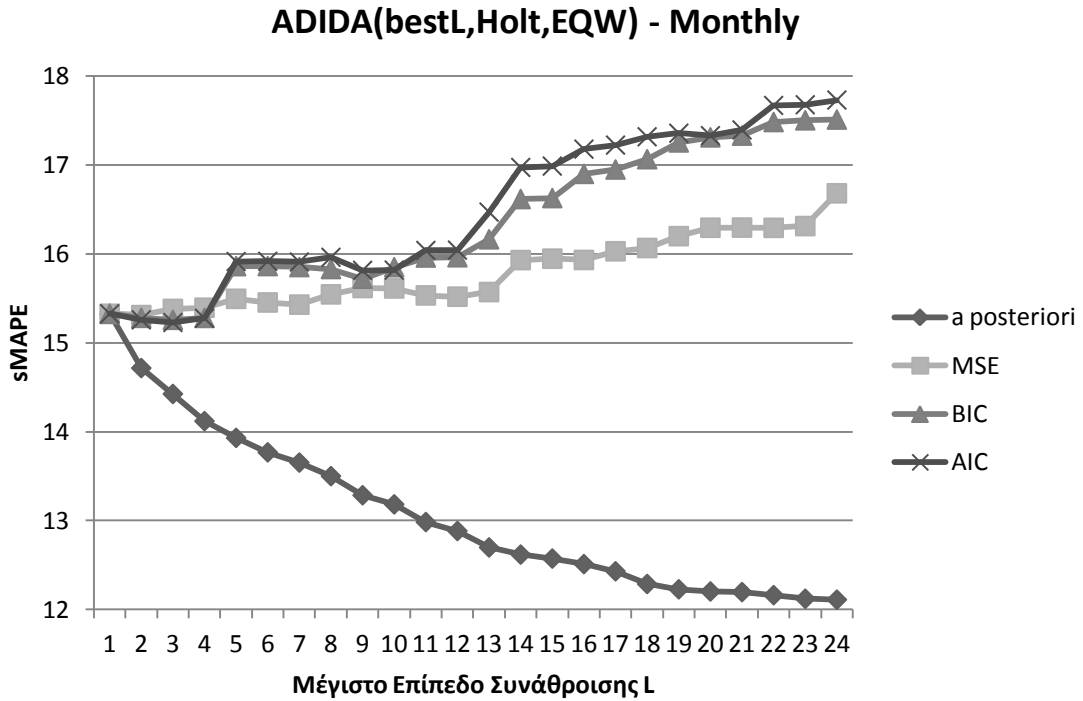
όλες τις εξεταζόμενες περιπτώσεις χρησιμοποιείται η μέθοδος διάσπασης με ίσα βάρη.



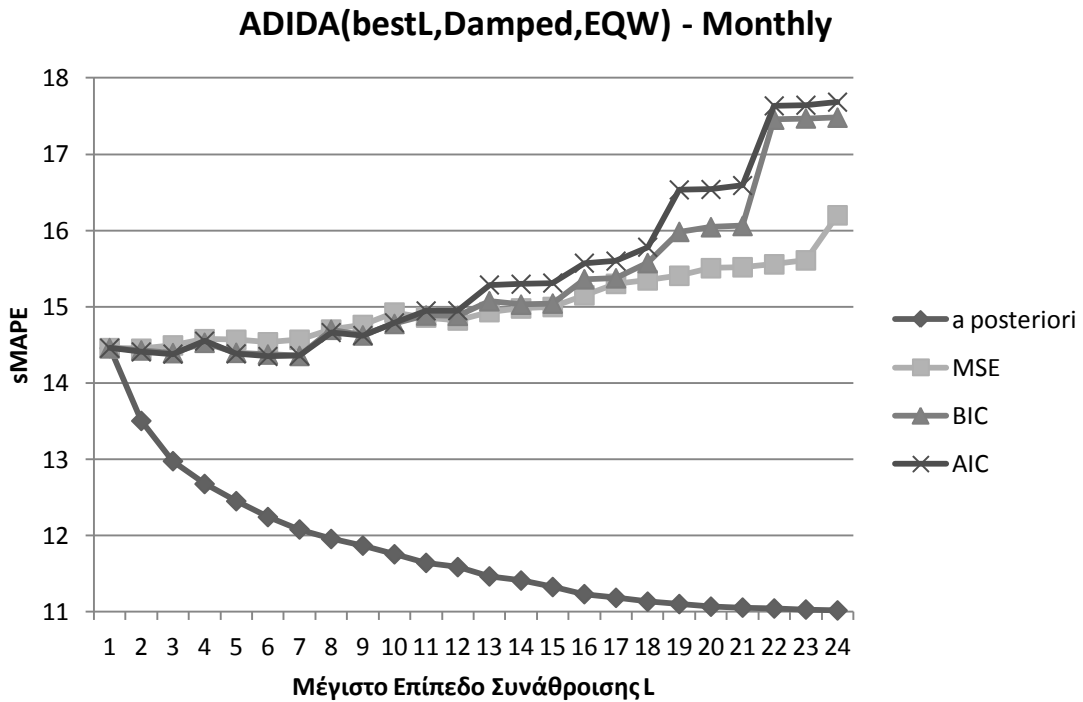
Εικόνα 5.21



Εικόνα 5.22



Εικόνα 5.23



Εικόνα 5.24

Το εμφανέστερο κοινό σημείο όλων των διαγραμμάτων είναι η συμπεριφορά της καμπύλης που αντιστοιχεί στην πλήρη εξ αρχής γνώση των a posteriori βέλτιστων επιπέδων συνάθροισης. Η καμπύλες αυτές πάντα μειώνονται,

αρχικά γρήγορα και μετά βραδύτερα και τελικά δείχνουν να συγκλίνουν προς κάποια οριακή τιμή, η οποία αποτελεί και η οριακά ελάχιστη τιμή σφάλματος που μπορεί να επιτευχθεί μέσω της μεθοδολογίας.

Οι γραμμές των τριών εξεταζόμενων τεχνικών a posteriori εκτίμησης απέχουν αρκετά από την οριακή καμπύλη, το οποίο σημαίνει ότι υπάρχουν σαφή περιθώρια περαιτέρω βελτίωσης. Οι τρεις τεχνικές δίνουν μεταβαλλόμενα αποτελέσματα, ανάλογα με το πλήθος των μοντέλων που λαμβάνονται υπόψη, και η κατάταξή τους σε όρους απόδοσης δεν παραμένει σταθερή.

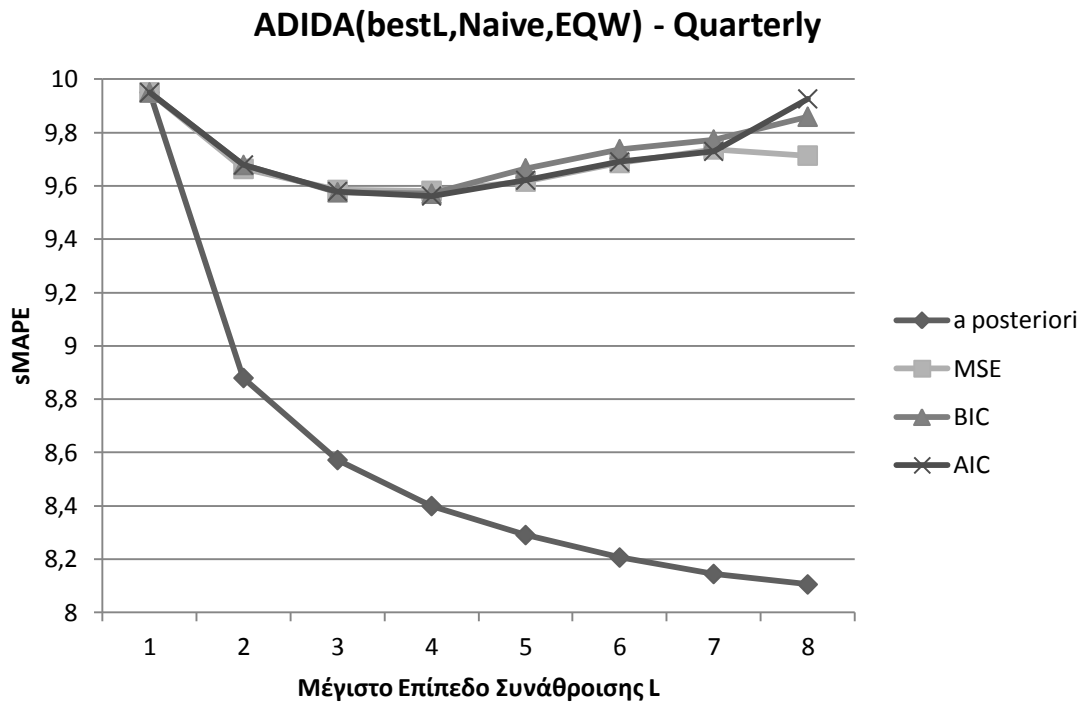
Για την μέθοδο Naïve οι τρεις τεχνικές παρουσιάζουν μία ικανοποιητική συμπεριφορά, παρόμοια με εκείνη της a posteriori καμπύλης, με μία υψηλότερη όμως τιμή σφάλματος σύγκλισης. Ωστόσο, οι καμπύλες των BIC και AIC εμφανίζουν ένα βύθισμα υπό της τιμής αυτής. Παρόμοια, αλλά όχι τόσο ικανοποιητικά αποτελέσματα προκύπτουν για τη μέθοδο SES. Στις μεθόδους Holt και Damped, η τρεις τεχνικές αποτυγχάνουν να ακολουθήσουν τη συμπεριφορά της a posteriori καμπύλης.

5.3.5 Τριμηνιαία Δεδομένα

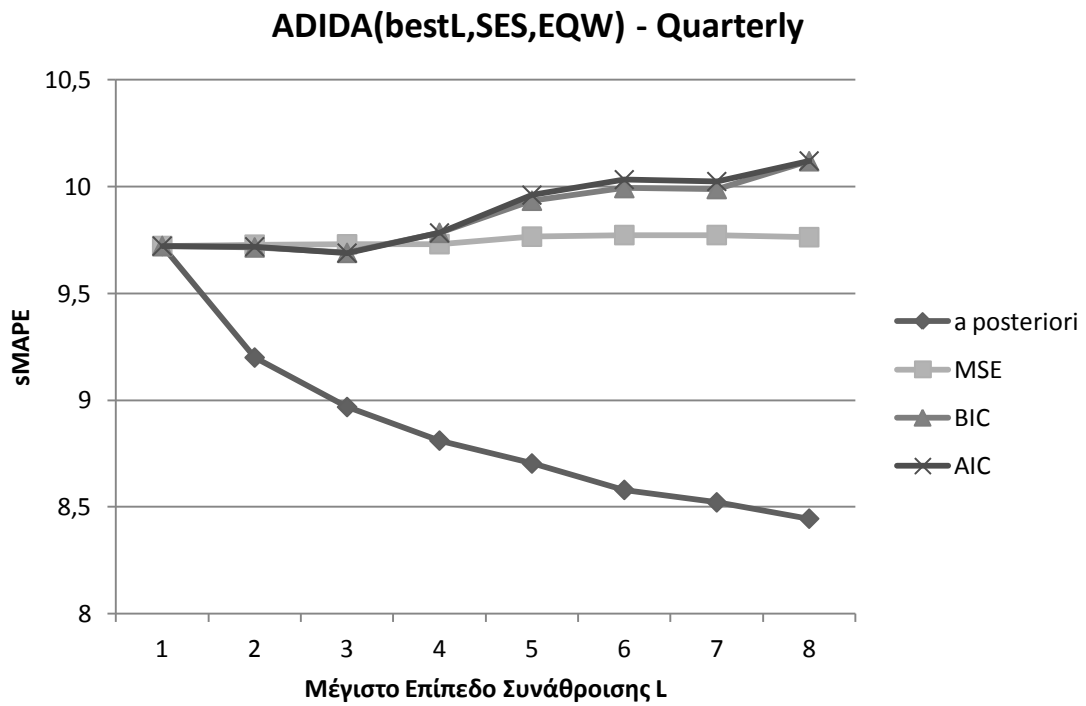
Στα επόμενα διαγράμματα απεικονίζεται η συμπεριφορά του μέσου sMAPE out-of-sample σφάλματος πρόβλεψης ορίζοντα 8 περιόδων για τις διάφορες τεχνικές εκτίμησης του βέλτιστου επιπέδου συνάθροισης L . Τα αποτελέσματα αναφέρονται στα τριμηνιαία δεδομένα. Ο οριζόντιος άξονας αντιπροσωπεύει και πάλι το επίπεδο συνάθροισης του μοντέλου ανώτερης τάξης μέχρι το οποίο εξετάζουμε για την επιλογή του κατά περίπτωση βέλτιστου.

Εξετάζονται οι τεχνικές επιλογής που στηρίζονται στην ελαχιστοποίηση των MSE, BIC και AIC. Διαφορετικές καμπύλες αντιστοιχούν σε διαφορετικές τεχνικές επιλογής βέλτιστου επιπέδου, που εκτιμάται ξεχωριστά για κάθε χρονοσειρά. Μία επιπλέον γραμμή αντιστοιχεί στο σφάλμα που εξασφαλίζει η πλήρης γνώση του a posteriori βέλτιστου για κάθε χρονοσειρά και υποδεικνύει και την οριακή τιμή μέχρι την οποία μπορεί να βελτιωθεί το σφάλμα.

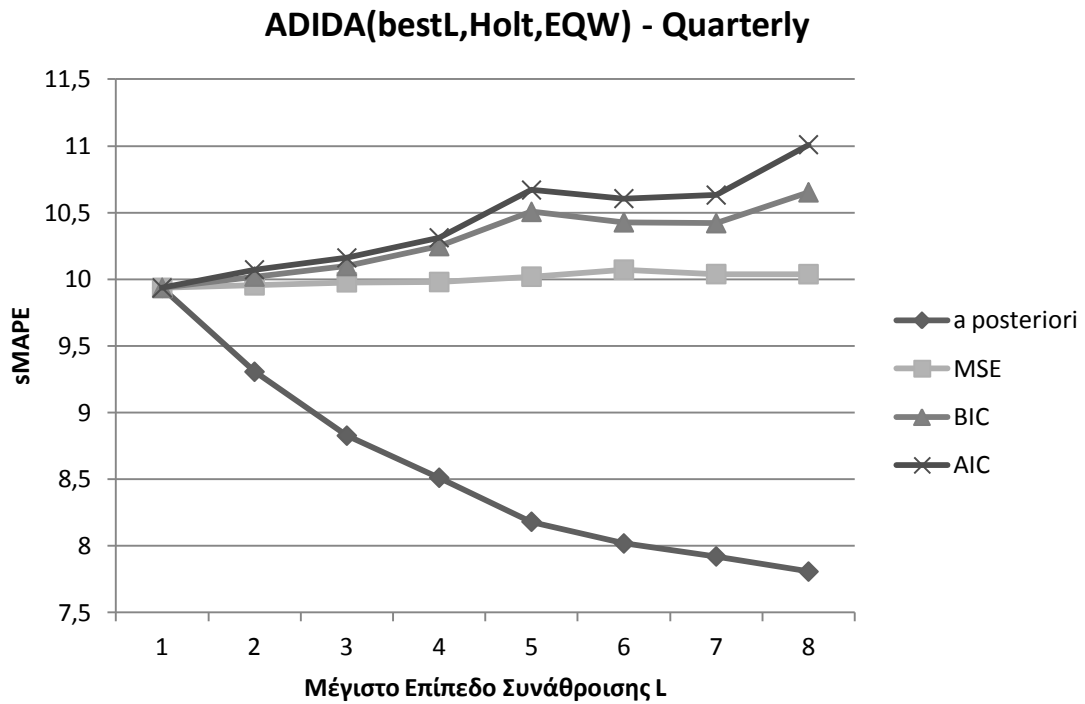
Η αποδοτικότητα της μεθοδολογίας μπορεί να αξιολογηθεί με σύγκριση με το μετρούμενο σφάλμα για επίπεδο συνάθροισης $L = 1$, αφού αυτό αντιστοιχεί σε απλή χρήση της μεθόδου πρόβλεψης χωρίς εφαρμογή της μεθοδολογίας ADIDA. Στις εξεταζόμενες περιπτώσεις χρησιμοποιείται η μέθοδος διάσπασης με ίσα βάρη. Σχεδιάζονται ξεχωριστά διαγράμματα για κάθε εξεταζόμενη μέθοδο πρόβλεψης.



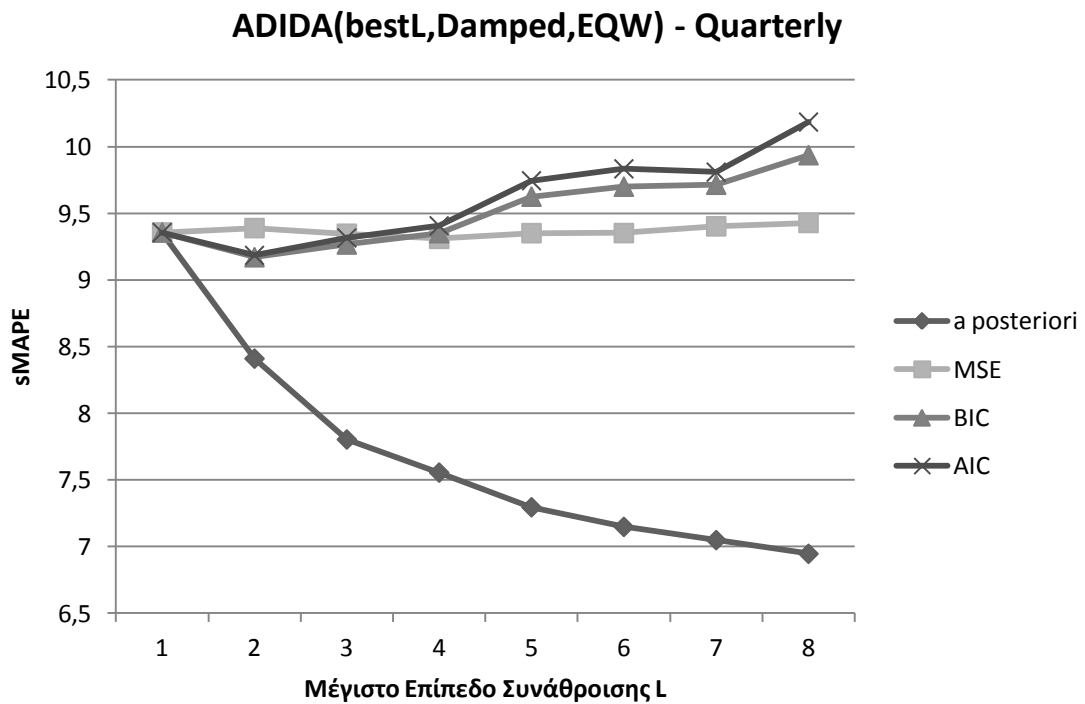
Εικόνα 5.25



Εικόνα 5.26



Εικόνα 5.27



Εικόνα 5.28

Και πάλι, οι καμπύλες που αντιστοιχούν στην πλήρη εξ αρχής γνώση των a posteriori βέλτιστων επιπέδων συνάθροισης πάντα μειώνονται με φθίνοντα ρυθμό και τελικά φαίνονται να συγκλίνουν προς κάποια οριακή τιμή, η οποία αποτελεί και

η οριακά ελάχιστη τιμή σφάλματος που μπορεί να επιτευχθεί μέσω της εφαρμογής της μεθοδολογίας.

Σε σχέση με τα μηνιαία δεδομένα, οι γραμμές των τριών εξεταζόμενων τεχνικών *a priori* εκτίμησης απέχουν ακόμα περισσότερο από την οριακή καμπύλη, το οποίο σημαίνει ότι υπάρχουν σαφή περιθώρια περαιτέρω βελτίωσης. Οι συμπεριφορά των καμπυλών των τριών τεχνικών μεταβάλλεται ανάλογα με το πλήθος των μοντέλων που λαμβάνονται υπόψη και η κατάταξή τους ως προς τα μετρούμενα σφάλματα δεν παραμένει σταθερή.

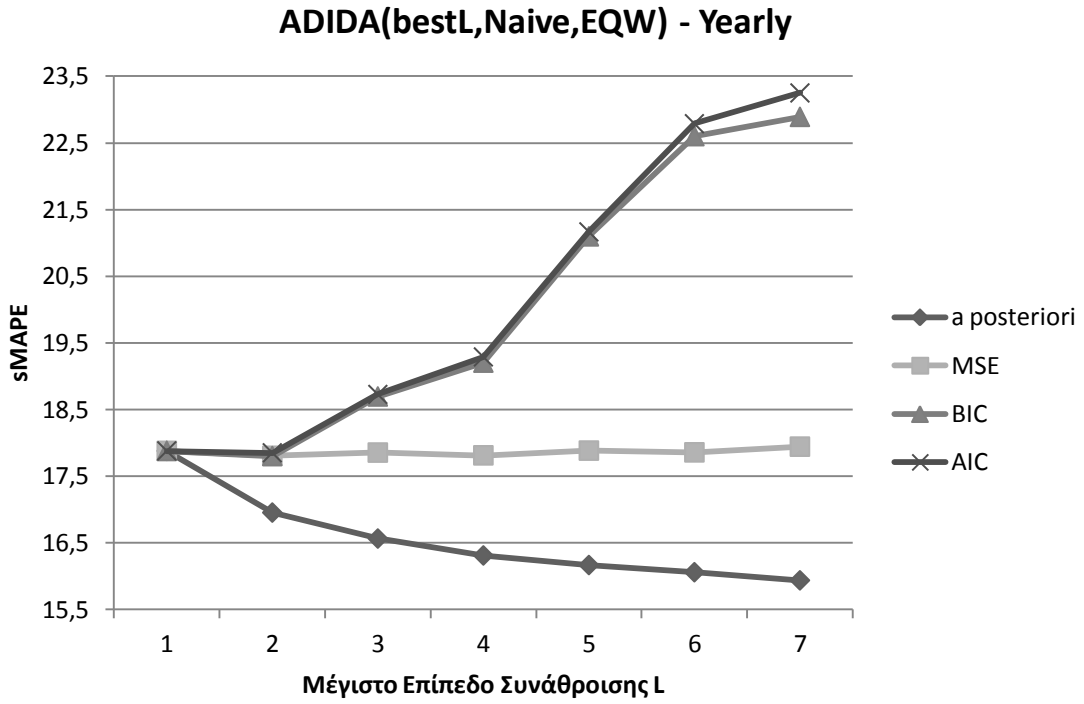
Η καλύτερη συμπεριφορά παρουσιάζεται στη μέθοδο *Naïve*, όπου οι τρεις τεχνικές παρουσιάζουν μία σχετικά καλή συμπεριφορά, που θυμίζει εκείνη της *a posteriori* καμπύλης, κινούμενη όμως σε υψηλότερα επίπεδα. Στις υπόλοιπες μεθόδους, τα αποτελέσματα, παρότι στιγμιαία και σποραδικά προκαλούνται μειώσεις του σφάλματος, δεν είναι εξίσου ικανοποιητικά.

5.3.6 Ετήσια Δεδομένα

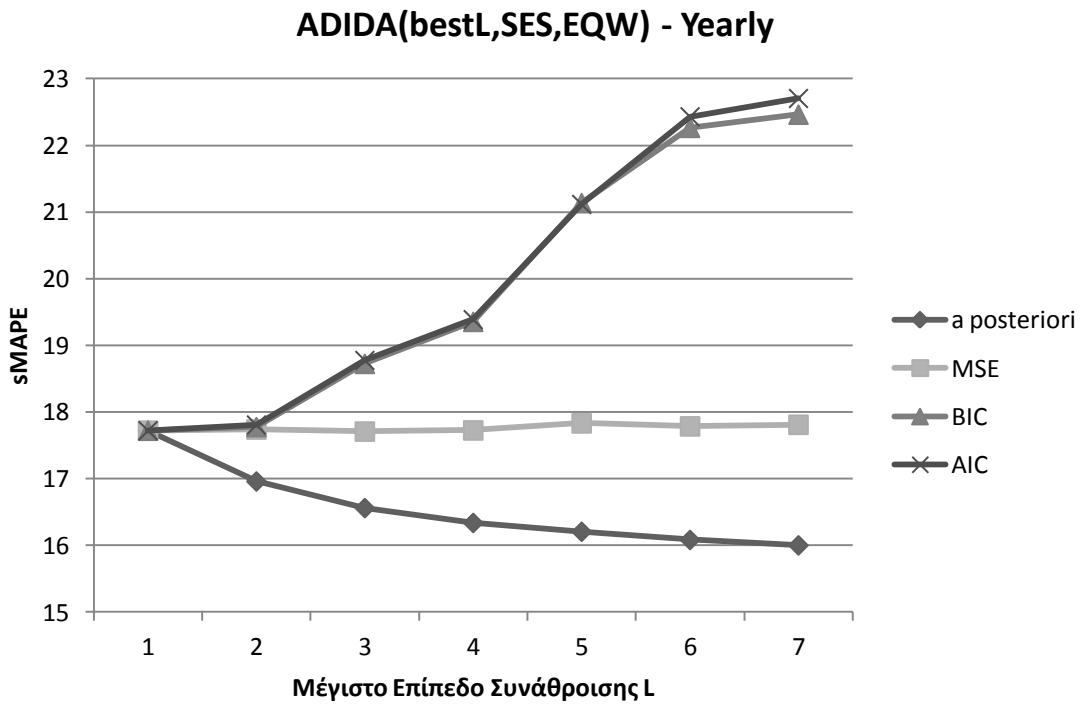
Παρόμοια διαγράμματα παραθέτουμε και για τα ετήσια δεδομένα. Στα ακόλουθα διαγράμματα απεικονίζεται το μέσο *sMAPE out-of-sample* σφάλμα πρόβλεψης ορίζοντα 6 περιόδων για διάφορες τεχνικές επιλογής του επιπέδου συνάθροισης L (ελαχιστοποίηση των *MSE*, *BIC*, *AIC*). Ο οριζόντιος άξονας αντιπροσωπεύει το ανώτερο επίπεδο συνάθροισης μέχρι το οποίο εξετάζουμε τα μοντέλα *ADIDA* για την καταλληλότητά τους.

Επιπλέον, μία καμπύλη παριστά το σφάλμα αν γνωρίζαμε εκ των προτέρων το *a posteriori* βέλτιστο επίπεδο συνάθροισης για κάθε χρονοσειρά και αποτελεί την οριακή τιμή στην οποία μπορεί να ελαττωθεί το σφάλμα.

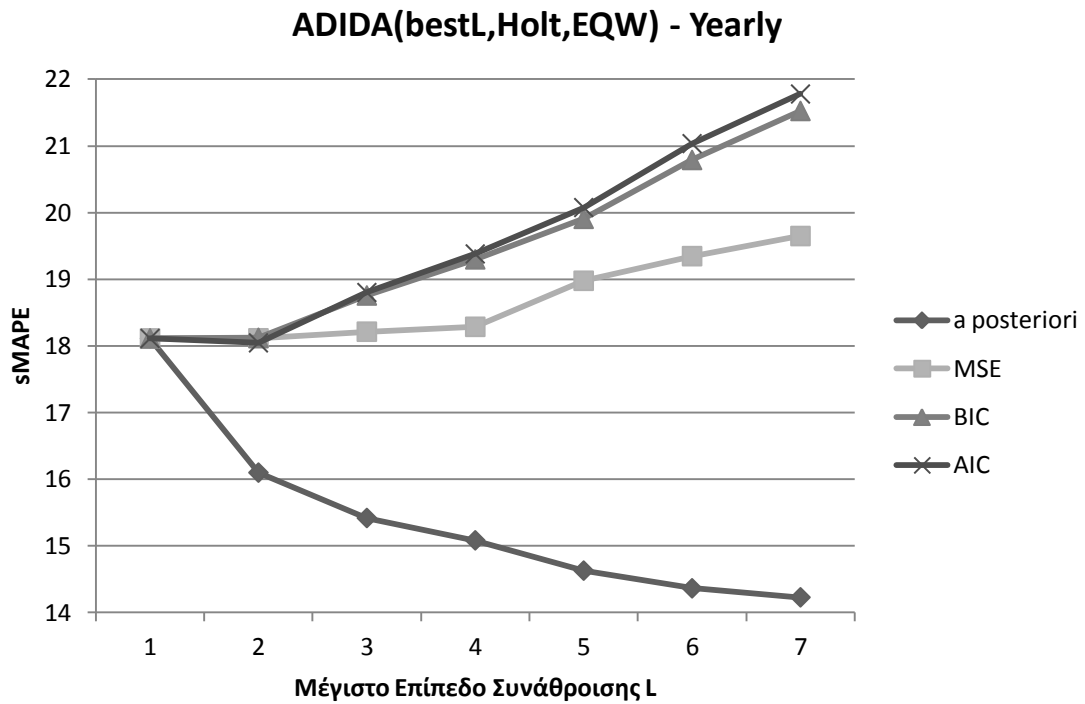
Η απόδοση της μεθοδολογίας μπορεί να αξιολογηθεί συγκρίνοντας με το παρατηρούμενο σφάλμα για το μοναδιαίο επίπεδο συνάθροισης, που αντιστοιχεί στην απλή χρήση της μεθόδου πρόβλεψης χωρίς εφαρμογή της μεθοδολογίας *ADIDA*. Παντού χρησιμοποιείται η μέθοδος διάσπασης με ίσα βάρη. Παρέχονται ξεχωριστά διαγράμματα για κάθε εξεταζόμενη μέθοδο πρόβλεψης.



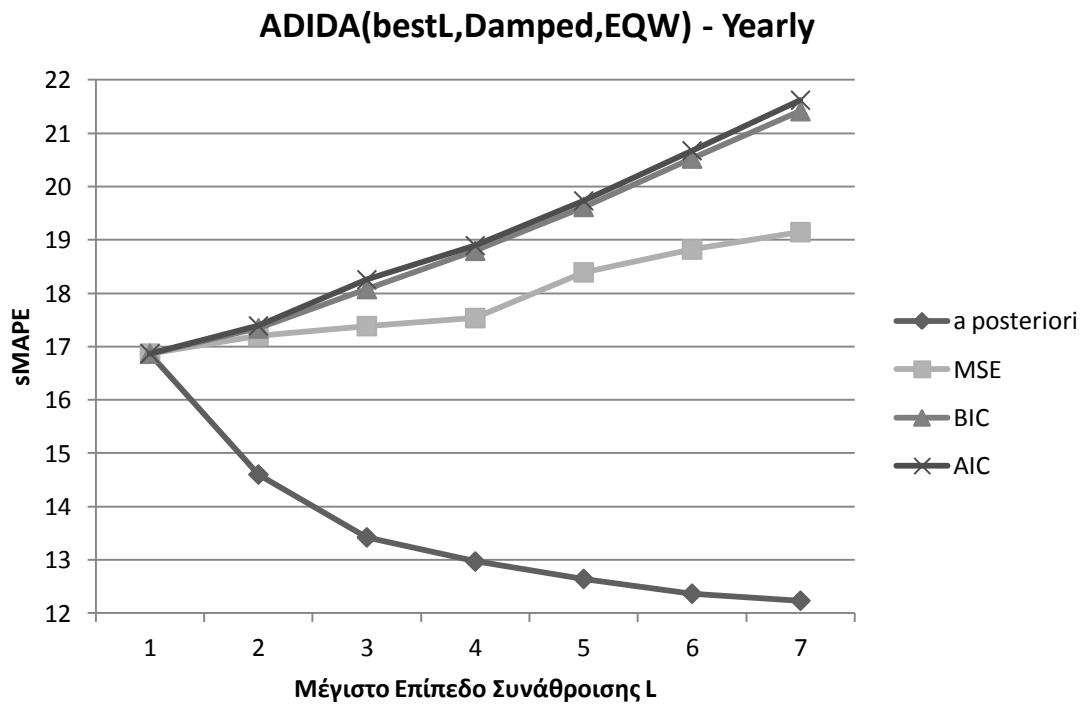
Εικόνα 5.29



Εικόνα 5.30



Εικόνα 5.31



Εικόνα 5.32

Ενώ, οι καμπύλες που αντιστοιχούν στην ακριβή εκ των προτέρων γνώση των a posteriori βέλτιστων επιπέδων συνάθροισης έχουν φθίνουσα και συγκλίνουσα σε μία οριακή τιμή πορεία, οι εκτιμήσεις μέσω των διαφόρων

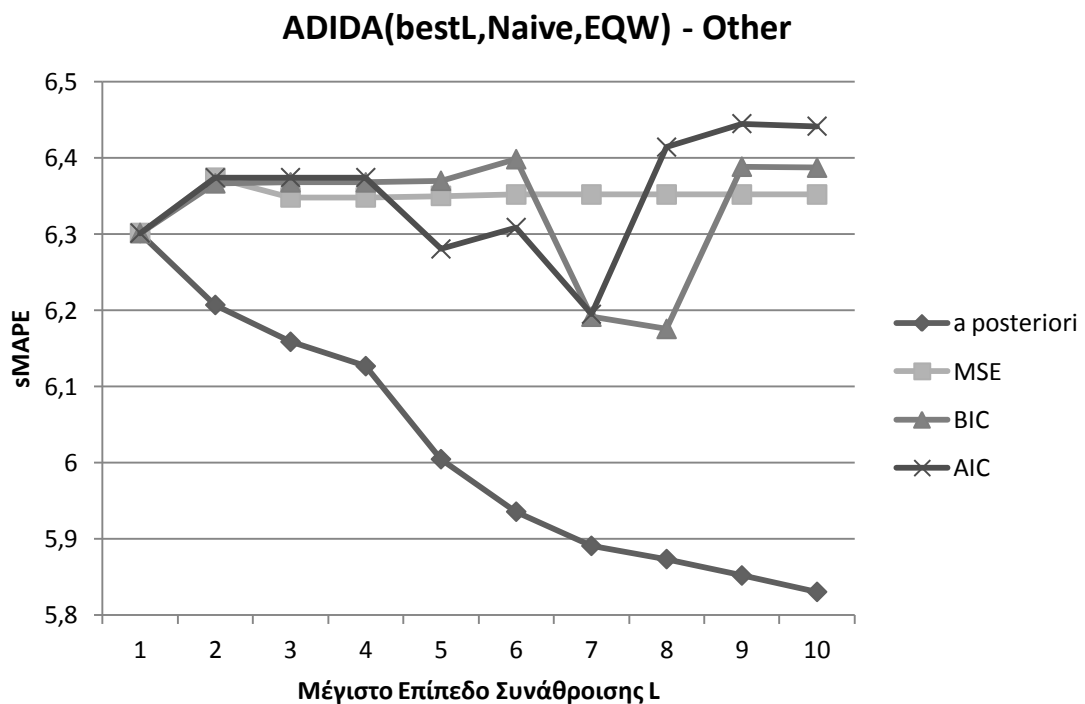
τεχνικών παρουσιάζουν μόνο λίγα σημεία μείωσης τους σφάλματος και τελικά συχνά αποκλίνουν όταν εξετάζονται και μοντέλα πολύ μεγάλης τάξης.

5.3.7 Άλλα Δεδομένα

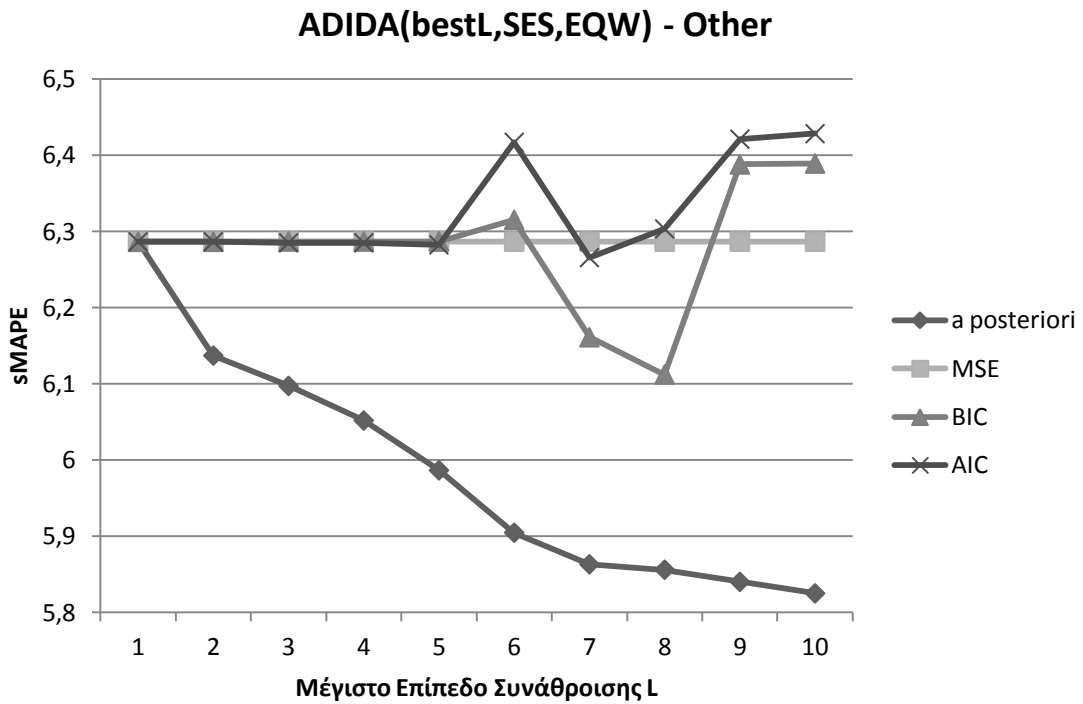
Στα επόμενα γραφήματα παρουσιάζεται η μεταβολή του μέσου sMAPE out-of-sample σφάλματος πρόβλεψης των υπολοίπων δεδομένων και για ορίζοντα 8 περιόδων συναρτήσει της μέγιστης τάξης επιπέδου συνάθροισης μέχρι την οποία λαμβάνονται υπόψη τα μοντέλα ADIDA για την επιλογή ενός βέλτιστου επιπέδου.

Εξετάζονται οι τεχνικές επιλογής μέσω ελαχιστοποίησης των MSE, BIC και AIC, για κάθε μία από τις οποίες χαράσσεται διαφορετική γραμμή. Μία επιπλέον γραμμή αναφέρεται στο σφάλμα για ακριβή εκ των προτέρων γνώση των a posteriori βέλτιστων επιπέδων συνάθροισης. Η καμπύλη αυτή υποδεικνύει και την οριακή τιμή στην οποία μπορεί να βελτιωθεί το σφάλμα.

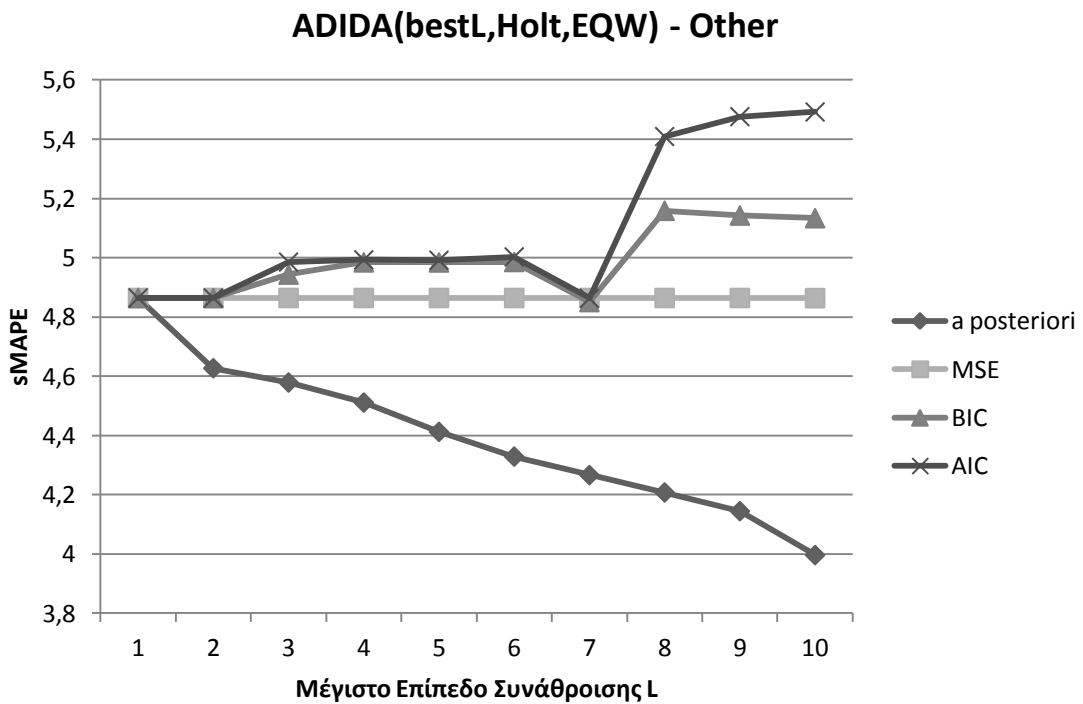
Παρέχονται ξεχωριστά διαγράμματα για κάθε εξεταζόμενη μέθοδο πρόβλεψης. Το παρατηρούμενο σφάλμα για μοναδιαίο επίπεδο συνάθροισης, που αντιστοιχεί σε απλή χρήση της μεθόδου πρόβλεψης χωρίς εφαρμογή της μεθοδολογίας ADIDA, αποτελεί στάθμη αναφοράς για την αξιολόγηση της απόδοσης της μεθοδολογίας. Σε όλες τις εξεταζόμενες περιπτώσεις χρησιμοποιείται η μέθοδος διάσπασης με ίσα βάρη.



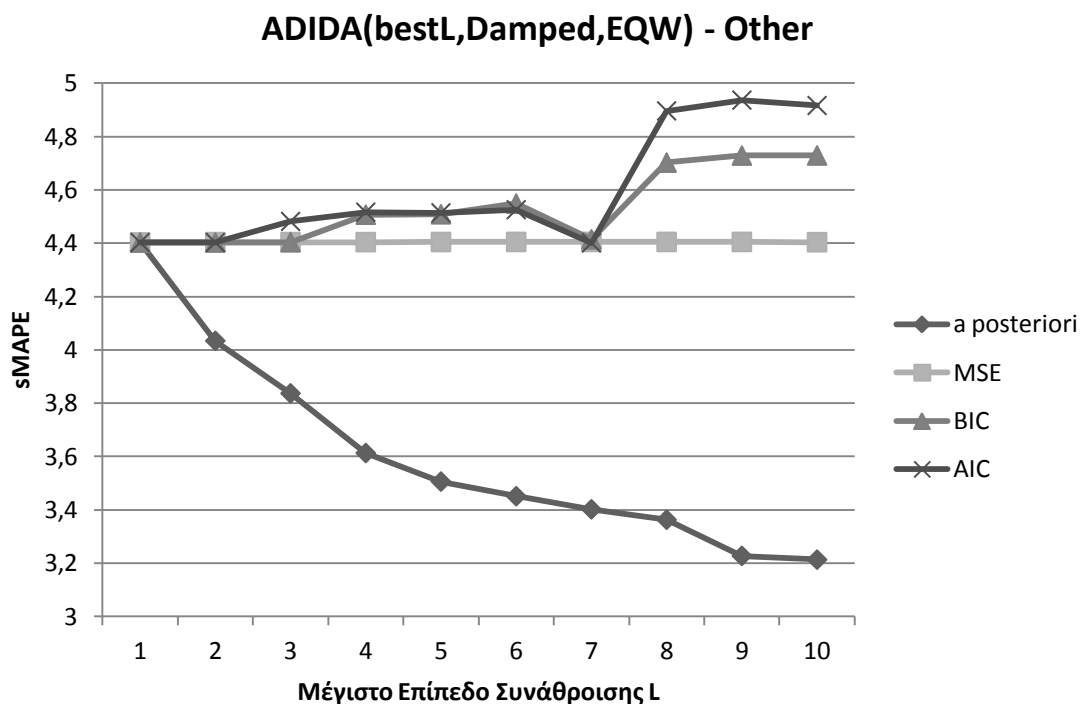
Εικόνα 5.33



Εικόνα 5.34



Εικόνα 5.35



Εικόνα 5.36

Η a posteriori καμπύλη μειώνεται με γενικά φθίνοντα ρυθμό και φαίνεται να συγκλίνει σε κάποια οριακή τιμή. Τα πρότυπα που ακολουθούν οι υπόλοιπες καμπύλες είναι περισσότερο περίπλοκα. Αν και γενικά απέχουν αρκετά από την οριακή καμπύλη και δεν ακολουθούν την πορεία της, υπάρχουν σημεία που παρουσιάζονται μικρές βελτιώσεις του σφάλματος.

5.3.8 Σύνοψη Αποτελεσμάτων και Συμπεράσματα

Τα αποτελέσματα από την εφαρμογή της μεθοδολογίας ποικίλουν δραστικά ανάλογα με την κατηγορία των δεδομένων (μηνιαία, τριμηνιαία, ετήσια, άλλα), τις εξεταζόμενες μεθόδους πρόβλεψης και την επιλεγμένη τεχνική εκτίμησης του βέλτιστου επιπέδου συνάθροισης. Ανάλογα με τις παραπάνω επιλογές, παρατηρούμε μία ευρεία διακύμανση στην αποδοτικότητα της μεθόδου, που κυμαίνεται μεταξύ σημαντικών μειώσεων των σφαλμάτων για τις βέλτιστες επιλογές και μεγάλων επιδεινώσεων για ακατάλληλους συνδυασμούς παραμέτρων της μεθόδου.

Στην ενότητα αυτή τα σημαντικότερα ευρήματα θα συγκεντρωθούν σε ζεύγη συνοπτικών πινάκων για κάθε κατηγορία δεδομένων. Αναλυτικότερα, ο πρώτος πίνακας περιλαμβάνεται δύο στήλες με τα σφάλματα των μεθόδων σύμφωνα με το διαγωνισμό M3 και την επανεκτέλεση των πειραμάτων του διαγωνισμού, αντίστοιχα, η οποία αντιστοιχεί στην εφαρμογή της μεθοδολογίας ADIDA για

μοναδιαίο επίπεδο συνάθροισης. Τυχόν αποκλίσεις οφείλονται σε διαφορετικές αρχικοποιήσεις και βελτιστοποιήσεις, αν και τα αποτελέσματα είναι παραπλήσια. Η μικρότερη εκ των δύο τιμών του σφάλματος βρίσκεται σε έντονα γράμματα και αποτελεί το επίπεδο αναφοράς του σφάλματος υπό του οποίου θεωρούμε ότι η μεθοδολογία προσέφερε βελτίωση. Επιπλέον στήλες περιγράφουν για τις διαφορετικές τεχνικές εκτίμησης των βέλτιστων επιπέδων συνάθροισης τα βέλτιστα αποτελέσματα σε όρους σφάλματος από την εφαρμογή της μεθοδολογίας ADIDA. Καλύπτονται όλες οι μελετημένες περιπτώσεις, με την a posteriori περίπτωση να αποτελεί το κατώτερο όριο μέχρι το οποίο μπορεί να επιτευχθεί ελάττωση του σφάλματος. Τα αποτελέσματα απεικονίζονται σε σκιασμένο φόντο όταν εμφανίζεται βελτίωση. Σε έναν συνοδευτικό πίνακα εμφανίζονται με παρόμοια δομημένο τρόπο τα αντίστοιχα επίπεδα συνάθροισης στα οποία παρατηρήθηκαν τα ελάχιστα σφάλματα κατά περίπτωση. Επιπλέον, κάθε γραμμή είναι αφιερωμένη και σε διαφορετική υπό εξέταση μέθοδο. Τα αποτελέσματα όλων των μεθόδων έχουν προκύψει για διάσπαση με ίσα βάρη.

Μηνιαία δεδομένα/Ορίζοντας 1-18

Method	sMAPE (%)					
	M3	Replication (L=1)	A posteriori	MSE	BIC	AIC
Naive2	16,91	16,89	12,29	14,41	13,98	13,97
Single	15,32	14,65	12,52	14,51	14,04	14,04
Holt	15,36	15,33	12,11	15,31	15,26	15,23
Dampen	14,59	14,46	11,02	14,45	14,35	14,35

Πίνακας 5.5

Method	Best max L			
	A posteriori	MSE	BIC	AIC
Naive2	24+	20	12	11
Single	24+	18	15	12
Holt	24+	2	3	3
Dampen	24+	2	7	6

Πίνακας 5.6

Τα αποτελέσματα για τα μηνιαία δεδομένα συνοψίζονται στους πίνακες 5.5 και 5.6. Είναι σαφές ότι σε για όλες τις τεχνικές εκτίμησης και όλες τις μεθόδους πρόβλεψης υπήρξε ελάττωση του σφάλματος σε κάποιο βαθμό. Με γνώση της a posteriori πραγματικής τιμής, η οποία αντιστοιχεί σε μία ιδανική και απολύτως ακριβή τεχνική εκτίμησης, υπάρχουν σημαντικά περιθώρια μείωσης, αρκετά κάτω από τα αντίστοιχα ελάχιστα καταγεγραμμένα σφάλματα του διαγωνισμού M3

(13,85% για τη μέθοδο Theta). Για παράδειγμα, με τη μέθοδο Damped και τέλεια εκτίμηση, το σφάλμα μπορεί να μειωθεί στο 11,02%. Ακόμα όμως και με τις υπόλοιπες, πραγματοποιησιμες τεχνικές εκτίμησης, υπήρξαν σημαντικές βελτιώσεις. Για παράδειγμα, μια ποσοστιαία μείωση του σφάλματος κατά 17,29% παρατηρείται για τη μέθοδο πρόβλεψης Naïve και εκτίμηση του βέλτιστου επιπέδου συνάθροισης μέσω του AIC.

Πέραν της a posteriori περίπτωσης, για την οποία τα σφάλματα μειώνονται με την αύξηση του πλήθους των υπό εξέταση μοντέλων και συγκλίνουν σε μία ελάχιστη τιμή, οι υπόλοιπες τεχνικές εμφάνισαν βέλτιστα για διάφορες μέγιστες τιμές της τάξης των μοντέλων.

Τα αποτελέσματα για τα τριμηνιαία δεδομένα συνοψίζονται στους πίνακες 5.7 και 5.8. Σε αρκετές περιπτώσεις καταγράφηκε βελτιωτική επίδραση της μεθοδολογίας στο σφάλμα πρόβλεψης. Η γνώση της a posteriori πραγματικής τιμής, που αντιστοιχεί σε μία ιδανική, απολύτως ακριβή τεχνική εκτίμησης, εξασφαλίζει σημαντικά περιθώρια μείωσης, αρκετά κάτω από τα αντίστοιχα ελάχιστα καταγεγραμμένα σφάλματα του διαγωνισμού M3 (8,96% για τη μέθοδο Theta). Για παράδειγμα, με τη μέθοδο Damped και τέλεια εκτίμηση, το σφάλμα μπορεί να μειωθεί στο 6,94%. Για όλες τις μεθόδους, πέραν τις Holt, οι υπόλοιπες, πραγματοποιησιμες τεχνικές εκτίμησης οδήγησαν σε κάποια μείωση του σφάλματος. Για παράδειγμα, μια ποσοστιαία μείωση κατά 3,92% παρατηρείται επίσης για το συνδυασμό μεθόδου πρόβλεψης Naïve και εκτίμησης του βέλτιστου επιπέδου συνάθροισης με χρήση του AIC.

Και πάλι, μόνο η a posteriori περίπτωση παρουσιάζει αδιαμφισβήτητη συνέπεια ως προς το βέλτιστο μέγιστο επίπεδο μέχρι το οποίο πρέπει να θεωρήσουμε τα εξεταζόμενα μοντέλα.

Τριμηνιαία δεδομένα/Ορίζοντας 1-8

Method	sMAPE (%)					
	M3	Replication (L=1)	A posteriori	MSE	BIC	AIC
Naive2	9,95	9,95	8,11	9,58	9,57	9,56
Single	9,72	9,72	8,45	9,72	9,69	9,68
Holt	10,67	9,93	7,81	9,93	9,93	9,93
Dampen	9,33	9,36	6,94	9,31	9,17	9,19

Πίνακας 5.7

Method	Best max L			
	A posteriori	MSE	BIC	AIC
Naive2	8+	4	4	4
Single	8+	1	3	3
Holt	8+	1	1	1
Dampen	8+	4	2	2

Πίνακας 5.8

Τα αποτελέσματα για τα ετήσια δεδομένα συνοψίζονται στους πίνακες 5.9 και 5.10. Αν και οι πραγματοποιήσιμες τεχνικές εκτίμησης παρουσιάζουν μικρές βελτιώσεις και σε ορισμένους μόνο συνδυασμούς, η γνώση της a posteriori πραγματικής τιμής, που αναφέρεται σε μία ιδανική τεχνική εκτίμησης, υπάρχουν σημαντικά περιθώρια βελτίωσης, αρκετά κάτω από τα αντίστοιχα ελάχιστα καταγεγραμμένα σφάλματα του διαγωνισμού M3 (16,42% για τη μέθοδο RBF). Για παράδειγμα, με τη μέθοδο Dampred και τέλεια εκτίμηση, το σφάλμα μπορεί να μειωθεί στο 12,24%.

Και πάλι, για τις υλοποιήσιμες τεχνικές εκτίμησης δεν διαφαίνεται κάποια συνέπεια στην βέλτιστη μέγιστη τάξη μέχρι την οποία πρέπει να εξεταστούν τα μοντέλα ADIDA. Μόνο στην a posteriori περίπτωση, λόγω της σύγκλισης, εμφανίζεται η κοινή συμπεριφορά σε όλες τις μεθόδους, ότι συμπερίληψη μεγαλύτερων μοντέλων οδηγεί σε οριακά καλύτερα αποτελέσματα.

Ετήσια δεδομένα/Ορίζοντας 1-6

Method	sMAPE (%)					
	M3	Replication (L=1)	A posteriori	MSE	BIC	AIC
Naive2	17,88	17,88	15,94	17,81	17,80	17,85
Single	17,82	17,72	16,00	17,71	17,72	17,72
Holt	19,27	18,11	14,23	18,11	18,11	18,05
Dampen	17,18	16,87	12,24	16,87	16,87	16,87

Πίνακας 5.9

Method	Best max L			
	A posteriori	MSE	BIC	AIC
Naive2	7+	2	2	2
Single	7+	3	1	1
Holt	7+	1	1	2
Dampen	7+	1	1	1

Πίνακας 5.10

Τα αποτελέσματα για τα υπόλοιπα δεδομένα συνοψίζονται στους πίνακες 5.11 και 5.12. Παρότι οι υλοποιήσιμες τεχνικές εκτίμησης εμφανίζουν μικρές μόνο βελτιώσεις και σε ορισμένους μονάχα συνδυασμούς, η a posteriori περίπτωση αποκαλύπτει περιθώρια για σημαντική βελτίωση, αρκετά κάτω από τα αντίστοιχα ελάχιστα καταγεγραμμένα σφάλματα του διαγωνισμού M3 (4,28% για τη μέθοδο ARARMA). Για παράδειγμα, με τη μέθοδο Dampred και τέλεια εκτίμηση, το σφάλμα μπορεί να μειωθεί στο 3,21%.

Μόνο η a posteriori περίπτωσης, λόγω της σύγκλισης, παρουσιάζει κοινή συμπεριφορά σε όλες τις μεθόδους, πως συμπερίληψη υψηλότερης τάξης μοντέλων οδηγεί σε οριακά καλύτερα αποτελέσματα. Και πάλι, για τις υλοποιήσιμες τεχνικές εκτίμησης δεν διαφαίνεται κάποια συνέπεια στην βέλτιστη μέγιστη τάξη μέχρι την οποία πρέπει να εξεταστούν τα μοντέλα.

Άλλα δεδομένα/Ορίζοντας 1-8

Method	sMAPE (%)					
	M3	Replication (L=1)	A posteriori	MSE	BIC	AIC
Naive2	6,30	6,30	5,83	6,30	6,18	6,19
Single	6,29	6,29	5,83	6,29	6,11	6,27
Holt	4,81	4,86	4,00	4,86	4,85	4,86
Dampen	4,61	4,40	3,21	4,40	4,40	4,40

Πίνακας 5.11

Method	Best max L			
	A posteriori	MSE	BIC	AIC
Naive2	10+	1	8	7
Single	10+	1	8	7
Holt	10+	1	7	1
Dampen	10+	1	1	1

Πίνακας 5.12

Συμπερασματικά, τα παραπάνω αποτελέσματα, και κυρίως εκείνα που σχετίζονται με την a posteriori εκτίμηση του επιπέδου, αποδεικνύουν ότι είναι εξαιρετικά αποδοτική η χρήση ξεχωριστού επιπέδου συνάθροισης ανά χρονοσειρά. Ειδικότερα, η a posteriori καμπύλες αποκαλύπτουν ότι η μεθοδολογία μπορεί να επιφέρει σημαντικές μειώσεις των σφαλμάτων, υπό την προϋπόθεση της απόκτησης ακριβών εκτιμήσεων των βέλτιστων επιπέδων συνάθροισης ανά χρονοσειρά. Σχετικά με τη σύγκλιση της καμπύλης, αφού η εξέταση ολοένα μεγαλύτερων μοντέλων επιδρά με φθίνοντα τρόπο στο τελικό αποτέλεσμα, επιβεβαιώνεται πειραματικά το θεωρητικά αναμενόμενο, ότι το βέλτιστο επίπεδο

συνάθροισης είναι μικρότερο μίας συγκεκριμένης, πεπερασμένης τιμής, η οποία καθορίζεται από το θεώρημα δειγματοληψίας και τα χαρακτηριστικά της χρονοσειράς. Οι υλοποιήσιμες τεχνικές (MSE, BIC και AIC) που εξετάστηκαν φάνηκαν να προκαλούν ικανοποιητικές μειώσεις σφάλματος σε δεδομένα υψηλότερων ρυθμών δειγματοληψίας, ωστόσο δεν είναι εξίσου αποδοτικές για αραιότερα καταγεγραμμένα δεδομένα, ούτε για τις διαφορετικές μεθόδους πρόβλεψης. Από τις τρεις μεθόδους, η MSE οδήγησε γενικά στις μικρότερες μειώσεις σφάλματος, ενώ τα αποτελέσματα για τις εκτιμήσεις BIC και AIC ήταν άμεσα συγκρίσιμα. Ωστόσο, η MSE παρουσίασε μεγάλη σταθερότητα ως προς την αύξηση του πλήθους των μοντέλων που λαμβάνονται υπόψη, ακόμα και όταν τα BIC και AIC φάνηκαν να παρουσιάζουν σημαντική απόκλιση από το επίπεδο αναφοράς του μοναδιαίου επιπέδου συνάθροισης. Παρά την φθίνουσα πορεία της *a posteriori* καμπύλης, οι μέθοδοι που μελετήθηκαν αδυνατούν συχνά να παρουσιάσουν πάντα μία αντίστοιχη και συνεπή συμπεριφορά μείωσης του σφάλματος. Καταδεικνύεται έτσι η ανάγκη της ανάπτυξης αποτελεσματικότερων μεθόδων εκτίμησης, ώστε να επιτυγχάνεται μία συμπεριφορά κοντινότερη σε εκείνη της *a posteriori* καμπύλης. Η προσέγγιση του ξεχωριστού επιπέδου συνάθροισης ανά χρονοσειρά φαίνεται περισσότερο υποσχόμενη από εκείνη του κοινού επιπέδου, λόγω κυρίως των μεγάλων περιθωρίων βελτίωσης που μαρτυρούν οι φθίνουσες και συγκλίνουσες *a posteriori* καμπύλες.

ΚΕΦΑΛΑΙΟ 6: Αλγόριθμοι Διάσπασης

6.1 Βασικοί Αλγόριθμοι Διάσπασης

Η παραχθείσα πρόβλεψη είναι εκφρασμένη στο συναθροισμένο επίπεδο, οπότε για την αναγωγή της στο αρχικό επίπεδο παρατήρησης της χρονοσειράς είναι συνήθως αναγκαία η εφαρμογή κάποιου αλγορίθμου διάσπασης.

Ο πιο απλός τρόπος διάσπασης συνίσταται στην χρήση L κατάλληλων βαρών διάσπασης w_k , τα οποία μπορούν συγκεντρωτικά να γραφούν σε μορφή διανύσματος:

$$w = [w_1 \quad \dots \quad w_L]^T$$

Όπως έχει αναλυθεί ήδη σε προηγούμενη ενότητα, η διάσπαση μπορεί να γίνει σε δύο στάδια, πρώτα με την προσθήκη $L - 1$ μηδενικών παρατηρήσεων μετά από κάθε πραγματική τιμή της συναθροισμένης χρονοσειράς ώστε να προκύψει μία σειρά διακοπτόμενης ζήτησης και στη συνέχεια με το γραμμικό φιλτράρισμα της σειράς αυτής με το διάνυσμα βαρών. Δηλαδή, αν x είναι η συναθροισμένη σειρά στην οποία έχουμε ήδη παρεμβάλει τεχνητά τις μηδενικές παρατηρήσεις, η διασπασμένη σειρά προκύπτει:

$$\hat{y}[n] = \sum_{k=1}^L w_k x[n - k]$$

Για να εκφράζει η παραπάνω σχέση ένα σταθμισμένο κινητό μέσο όρο είναι απαραίτητο τα βάρη να ικανοποιούν τη συνθήκη κανονικοποίησης:

$$\sum_{k=1}^L w_k = 1$$

Συνεπώς, γίνεται εύκολα αντιληπτό ότι το πρόβλημα της διάσπασης της συναθροισμένης χρονοσειράς συνίσταται στον τρόπο επιλογής του διανύσματος βαρών w . Σε μία γενικότερη εκδοχή, τα βάρη μπορεί να παρουσιάζουν εξάρτηση από το χρόνο, ωστόσο τέτοιοι αλγόριθμοι δεν θα μελετηθούν εδώ.

Τρεις απλές μέθοδοι διάσπασης είχαν μελετηθεί στα πλαίσια της έρευνας των Nikolopoulos et al. (2011): η μέθοδος ίσων (EQW), προηγούμενων (PRW) και μέσων (AVW) βαρών, οι οποίες περιγράφονται στη συνέχεια. Η μελέτη έδειξε ότι για το συγκεκριμένο σύνολο δεδομένων η απλούστερη μέθοδος των ίσων βαρών

δίνει τα καλύτερα αποτελέσματα. Οι Lisman, and Sandee (1964) είχαν προτείνει μία μέθοδο παραγωγής τριμηνιαίων χρονοσειρών από ετήσια δεδομένα, που όμως δεν μπορεί να διασπάσει τα πρώτο και το τελευταίο έτος. Αργότερα, εναλλακτικές μέθοδοι έκαναν εφικτή την επιπλέον απόκτηση διασπασμένων τιμών για την αρχή και το τέλος της χρονοσειράς (Boot, Feibes, and Lisman 1967). Οι εξελίξεις στον τομέα είναι διαρκείς με προτάσεις νέων μεθόδων, όπως διαδικασιών μονο-μεταβλητής (Wei, and Stram, 1990) και δι-μεταβλητής (Hodgess, and Wei, 2000) διάσπασης, αλλά και συγκεντρωτικές βιβλιογραφικές και συγκριτικές μελέτες για τις μέχρι στιγμή σημαντικότερες μεθόδους (Chan, 1993; Feijó, Caro, and Quintana, 2003). Στις μελέτες αυτές, υπολογίζονται αποτελέσματα μέσω προσομοιώσεων για τις διάφορες μεθόδους διάσπασης, οι οποίες συγκρίνονται μεταξύ τους, και γίνονται προτάσεις για την καταλληλότητα εφαρμογής τους.

6.1.1 Ίσα Βάρη/Equal Weights (EQW)

Ο απλούστερος αλγόριθμος διάσπασης είναι ο ισοβαρής διαχωρισμός. Προκειμένου να ισχύει η συνθήκη κανονικοποίησης, τα ίσα βάρη πρέπει να ισούνται με:

$$w_k = \frac{1}{L}, k = 1, \dots, L$$

Για τα δεδομένα των Nikolopoulos et al. (2011), ο αλγόριθμος των ίσων βαρών έδωσε το καλύτερο αποτέλεσμα. Το αποτέλεσμα αυτό αποδόθηκε στη στασιμότητα των χρονοσειρών που χρησιμοποιήθηκαν.

Ένα σημαντικό χαρακτηριστικό του αλγορίθμου αυτού είναι ότι τα βάρη δεν παρουσιάζουν καμία εξάρτηση από τα στατιστικά χαρακτηριστικά των χρονοσειρών, αλλά εξαρτώνται μόνο από το επίπεδο συνάθροισης, έτσι ώστε να ικανοποιείται σε κάθε περίπτωση η συνθήκη κανονικοποίησης. Η μορφή που παίρνει το φίλτρο ανακατασκευής του σταδίου διάσπασης είναι πάντα αυτή ενός βαθυπερατού φίλτρου τύπου κινητού μέσου όρου.

6.1.2 Προηγούμενα βάρη/PRevious Weights (PRW)

Επιλέγονται τα βάρη σύμφωνα με τους λόγους των τελευταίων L μη συναθροισμένων παρατηρήσεων προς το άθροισμά τους, το οποίο φυσικά αποτελεί και την τελευταία διαθέσιμη συναθροισμένη παρατήρηση. Αυτή η επιλογή εξασφαλίζει αυτόματα την ικανοποίηση της συνθήκης κανονικοποίησης και συνδέει

τα βάρη με τις στατιστικές ιδιότητες των τελευταίων παρατηρήσεων της χρονοσειράς.

Ωστόσο ο αλγόριθμος αυτός έδωσε τα λιγότερο ικανοποιητικά αποτελέσματα στα δεδομένα των Nikolopoulos et al. (2011). Αυτό μπορεί να ερμηνευθεί βάσει του γεγονότος ότι δεν επαναλαμβάνονται υποχρεωτικά οι ίδιες αναλογίες τιμών και στο επόμενο παράθυρο συνάθροισης.

Επιπλέον, ανάλογα με τα υπολογισμένα βάρη, το φίλτρο ανακατασκευής του σταδίου διάσπασης μπορεί να μην έχει βαθυπερατή συμπεριφορά και επομένως να μην είναι κατάλληλο για την ανακατασκευή του μοντέλου πρόβλεψης στο κατώτερο χρονικό επίπεδο.

6.1.3 Μέσος όρος των βαρών/Average Weights (AVW)

Σύμφωνα με τον αλγόριθμο των μέσων βαρών, τα βάρη επιλέγονται χωρίζοντας αρχικά τη χρονοσειρά σε τμήματα των L παρατηρήσεων, ξεκινώντας ανάποδα, από το τέλος προς την αρχή. Σε κάθε διάστημα υπολογίζονται οι λόγοι, όπως στην περίπτωση των προηγούμενων βαρών, και στο τέλος ως βάρη διάσπασης λαμβάνονται οι μέσοι όροι των αντίστοιχων λόγων. Αν το μήκος της χρονοσειράς δεν είναι ακέραιο πολλαπλάσιο του L , τυχόν εναπομείνουσες παρατηρήσεις παραμένουν αχρησιμοποίητες. Ο αλγόριθμος των μέσων βαρών εξασφαλίζει αυτόματα την ικανοποίηση της συνθήκης κανονικοποίησης, ενώ ταυτόχρονα συνδέει τα βάρη με τις τιμές της χρονοσειράς καθαυτές.

Από τη μελέτη των Nikolopoulos et al. (2011), ο αλγόριθμος AVW έδωσε μέτρια αποτελέσματα, με απόδοση που κυμάνθηκε ανάμεσα σε εκείνες των μεθόδων EQW και PRV. Η καλύτερη απόδοση σε σχέση με την PRV εξηγείται από το γεγονός ότι λαμβάνεται υπόψη η πληροφορία των λόγων των τιμών του συνόλου της χρονοσειράς και όχι μόνο των τελευταίων παρατηρήσεων. Έτσι, τα βάρη είναι περισσότερο αντιπροσωπευτικά για το σύνολο της χρονοσειράς.

Ωστόσο, ούτε ο αλγόριθμος αυτός εγγυάται ότι το φίλτρο ανακατασκευής του σταδίου διάσπασης θα έχει βαθυπερατή συμπεριφορά και συνεπώς μπορεί να είναι ακατάλληλο για την ανακατασκευή του μοντέλου πρόβλεψης στο κατώτερο χρονικό επίπεδο.

6.1.4 Εξισώσεις Yule-Walker (YW)

Στο σημείο αυτό θα προτείνουμε και θα μελετήσουμε έναν επιπλέον αλγόριθμο διάσπασης που δεν εμφανίστηκε στην αρχική μελέτη της μεθοδολογίας. Η προσέγγιση αυτή θα λαμβάνει περισσότερο υπόψη τις στατιστικές ιδιότητες της

χρονοσειράς (με τη μορφή συσχετίσεων) για την παραγωγή των σχετικών βαρών. Η χρήση του σταθμισμένου κινητού μέσου όρου για τη διάσπαση των προβλέψεων αποτελεί μία μορφή που θυμίζει αντίστοιχες εκφράσεις που εμφανίζονται σε προβλήματα γραμμικής πρόβλεψης και αυτό-συσχετιστικών μοντέλων. Και στις δύο περιπτώσεις, τα βάρη μπορούν να προσδιοριστούν με την επίλυση των εξισώσεων Yule-Walker. Με τον τρόπο αυτό, λαμβάνονται κατά την επιλογή των βαρών υπόψη τα στατιστικά χαρακτηριστικά (συσχετίσεις) της αρχικής χρονοσειράς.

Από μαθηματική σκοπιά, το πρόβλημα μπορεί να περιγραφεί ως πρόβλημα επιλογής κατάλληλου διανύσματος βαρών, ώστε μετά το φιλτράρισμα με αυτό μίας σειράς x , αυτή να παρουσιάζει τη μέγιστη δυνατή ομοιότητα με μία άλλη χρονοσειρά-πρότυπο y . Η έννοια της ομοιότητας μπορεί να πάρει συγκεκριμένη υπόσταση με τη χρήση ενός κριτηρίου σφάλματος, όπως το MSE, το οποίο αποτελεί μέτρο της ομοιότητας υπό την έννοια των ελαχίστων τετραγώνων:

$$MSE(w) = E[(y[n] - \hat{y}[n])^2] = E \left[\left(y[n] - \sum_{k=1}^L w_k x[n-k] \right)^2 \right]$$

Στον παραπάνω τύπο, με $E[\]$ δηλώνεται ο στατιστικός τελεστής της αναμενόμενης τιμής. Συνεπώς, είναι πλέον εμφανές ότι βρισκόμαστε αντιμέτωποι με το πρόβλημα ελαχιστοποίησης:

$$w = \arg \min_w \{MSE(w)\}$$

Λύση στο παραπάνω πρόβλημα ελαχιστοποίησης μπορεί να βρεθεί με τη χρήση των πρώτων παραγώγων ως προς τα βάρη και επιλύοντας το σύστημα εξισώσεων:

$$\frac{\partial MSE(w)}{\partial w_m} = 0, \quad m = 1, \dots, L$$

Η μερική παράγωγος ως προς κάποιο βάρος w_m υπολογίζεται ως:

$$\begin{aligned}
\frac{\partial MSE(w)}{\partial w_m} &= -E \left[2 \left(y[n] - \sum_{k=1}^L w_k x[n-k] \right) x[n-m] \right] \\
&= 2E \left[\sum_{k=1}^L w_k x[n-k] x[n-m] \right] - 2E[y[n]x[n-m]] \\
&= 2 \sum_{k=1}^L w_k E[x[n-k]x[n-m]] - 2E[y[n]x[n-m]], \\
& \quad m = 1, \dots, L
\end{aligned}$$

Στην παραπάνω έκφραση εμφανίζονται οι γνωστές συναρτήσεις αυτοσυσχέτισης της χρονοσειράς x και της ετεροσυσχέτισής της x με την y :

$$\psi_{x,x}(k, m) = E[x[n-k]x[n-m]] = E[x[n]x[n-(m-k)]] = \psi_{x,x}(m-k)$$

$$\psi_{x,y}(m) = E[y[n]x[n-m]]$$

Οι συσχετίσεις αυτές είναι τα στατιστικά χαρακτηριστικά των χρονοσειρών που χρησιμοποιούνται από τον αλγόριθμο αυτό για την επιλογή των βαρών. Μπορούμε να υπολογίσουμε εκτιμήσεις των τιμών των παραπάνω συσχετίσεων αντικαθιστώντας τον τελεστή της στατιστικά αναμενόμενης τιμής $E[\cdot]$ με τον δειγματικό μέσο όρο, οπότε:

$$\begin{aligned}
\psi_{x,x}(k, m) &= \frac{1}{N} \sum_{n=1}^N x[n-k]x[n-m] = \frac{1}{N} \sum_{n=1}^N x[n]x[n-(m-k)] = \psi_{x,x}(m-k) \\
\psi_{x,y}(m) &= \frac{1}{N} \sum_{n=1}^N y[n]x[n-m]
\end{aligned}$$

Παρατηρούμε λοιπόν ότι το σύστημα των εξισώσεων που σχηματίζεται υπό την απαίτηση κάθε μερική παράγωγος του MSE να μηδενίζεται είναι ένα γραμμικό σύστημα εξισώσεων ως προς τα άγνωστα βάρη και που μπορεί να γραφεί περισσότερο συνοπτικά ως:

$$\sum_{k=1}^L w_k \psi_{x,x}(k, m) = \psi_{x,y}(m), \quad m = 1, \dots, L$$

ή, ακόμα συνοπτικότερα, σε μορφή πίνακα ως:

$$\begin{bmatrix} \psi_{x,x}(1,1) & \dots & \psi_{x,x}(L,1) \\ \vdots & \ddots & \vdots \\ \psi_{x,x}(1,L) & \dots & \psi_{x,x}(L,L) \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} \psi_{x,y}(1) \\ \vdots \\ \psi_{x,y}(L) \end{bmatrix}$$

$$Aw = b$$

Οι εξισώσεις του παραπάνω συστήματος αναφέρονται συχνά στη βιβλιογραφία ως εξισώσεις Yule-Walker. Έχουν αναπτυχθεί αρκετοί αλγόριθμοι για την επίλυση των εξισώσεων αυτών. Η λύση μπορεί να βρεθεί με αντιστροφή του πίνακα (που όμως είναι αργή και μπορεί να οδηγήσει σε αριθμητικά σφάλματα), με αριθμητικές μεθόδους, όπως η μέθοδος της απαλοιφής Gauss, της LUV παραγοντοποίησης, κλπ. Ανεξαρτήτως πάντως του ειδικότερου αλγορίθμου που χρησιμοποιείται για την εύρεση της λύσης, αυτή μπορεί σε κάθε περίπτωση να γραφεί ως:

$$w = A^{-1}b$$

Δυστυχώς, η προσέγγιση αυτή έχει αγνοήσει μέχρι στιγμής τη συνθήκη κανονικοποίησης, της οποίας η ικανοποίηση δεν εξασφαλίζεται με κάποιον τρόπο.

Μία διαισθητικά απλή προσέγγιση θα ήταν η απευθείας κανονικοποίηση των βαρών που έχουν εξευρεθεί, διαιρώντας την τιμή κάθε βάρους με το άθροισμα των βαρών, $sum(w)$. Ωστόσο, αν η κανονικοποίηση αυτή επιφέρει μεγάλες αλλαγές στα βάρη, τότε αυτά ενδέχεται να απομακρυνθούν πολύ από τη λύση του γραμμικού συστήματος των εξισώσεων Yule-Walker, οπότε θα πάψουν να είναι βέλτιστα υπό κάποια έννοια.

Η επιπλοκή αυτή επιβάλλει να ληφθεί υπόψη εξ αρχής η συνθήκη κατά την κατάστρωση των εξισώσεων, οπότε καλούμαστε να λύσουμε ένα πρόβλημα ελαχιστοποίησης υπό συνθήκη:

$$w = \arg \min_w \{MSE(w) | sum(w) = 1\}$$

Για τη λύση θα χρησιμοποιήσουμε τη μέθοδο των πολλαπλασιαστών Lagrange. Καταφεύγουμε λοιπόν στη χρήση της Lagrangian συνάρτησης:

$$L(w, \lambda) = MSE(w) + \lambda \left(\sum_{k=1}^L w_k - 1 \right)$$

Η παράμετρος $\lambda \in \mathbb{R}$ αποτελεί έναν λαγκρανζιανό πολλαπλασιαστή, ο οποίος θα προσδιοριστεί στην πορεία. Η λύση για τα βάρη προκύπτει τώρα επιλύοντας το επεκταμένο σύστημα εξισώσεων:

$$\begin{aligned}\frac{\partial L(w, \lambda)}{\partial w_m} &= 0, \quad m = 1, \dots, L \\ \frac{\partial L(w, \lambda)}{\partial \lambda} &= 0\end{aligned}$$

Οι μερικές παράγωγοι ισούνται με:

$$\begin{aligned}\frac{\partial L(w, \lambda)}{\partial w_m} &= 2 \sum_{k=1}^L w_k \psi_{x,x}(k, m) - 2\psi_{x,y}(m) + \lambda, \quad m = 1, \dots, L \\ \frac{\partial L(w, \lambda)}{\partial \lambda} &= \sum_{k=1}^L w_k - 1\end{aligned}$$

Παρατηρούμε λοιπόν ότι το σύστημα των εξισώσεων αποτελείται από ένα γραμμικό σύστημα ως προς τα βάρη, που εμπλέκει την άγνωστη παράμετρο λ , και μια συνθήκη.

$$\begin{aligned}\sum_{k=1}^L w_k \psi_{x,x}(k, m) &= \psi_{x,y}(m) - \frac{\lambda}{2}, \quad m = 1, \dots, L \\ \sum_{k=1}^L w_k &= 1\end{aligned}$$

Ορίζοντας ένα διάστημα μήκους L που κάθε του τιμή ισούται με την ομάδα, μπορούμε να πάρουμε μία βολικότερη στον χειρισμό έκφραση για το παραπάνω σύστημα:

$$\begin{aligned}Aw &= b - \frac{\lambda}{2}u, \quad u = [1 \quad \dots \quad 1]^T \\ \text{sum}(w) &= 1\end{aligned}$$

Αν η παράμετρος λ ήταν γνωστή, η λύση θα βρισκόταν εύκολα ως:

$$w = A^{-1} \left(b - \frac{\lambda}{2}u \right) = A^{-1}b - \frac{\lambda}{2}A^{-1}u$$

Η παραπάνω λύση θα ικανοποιούσε τη συνθήκη κανονικοποίησης. Αντικαθιστώντας τη λύση στη συνθήκη προσδιορίζουμε τον πολλαπλασιαστή λ :

$$\text{sum}(A^{-1}b) - \frac{\lambda}{2} \text{sum}(A^{-1}u) = 1$$

$$\lambda = 2 \frac{\text{sum}(A^{-1}b) - 1}{\text{sum}(A^{-1}u)}$$

Πλέον, απομένει μονάχα η λύση του τελικού συστήματος, το οποίο εκφράζει τις τροποποιημένες εξισώσεις Yule-Walker, ώστε να λαμβάνεται υπόψη ο περιορισμός της κανονικοποίησης, και γράφεται τελικά ως:

$$Aw = b - \frac{\text{sum}(A^{-1}b) - 1}{\text{sum}(A^{-1}u)} u$$

Τα ίδια με πριν ισχύουν και για την επίλυση του παραπάνω συστήματος, που δίνει λύση:

$$w = A^{-1} \left(b - \frac{\text{sum}(A^{-1}b) - 1}{\text{sum}(A^{-1}u)} u \right) = A^{-1}b - \frac{\text{sum}(A^{-1}b) - 1}{\text{sum}(A^{-1}u)} A^{-1}u$$

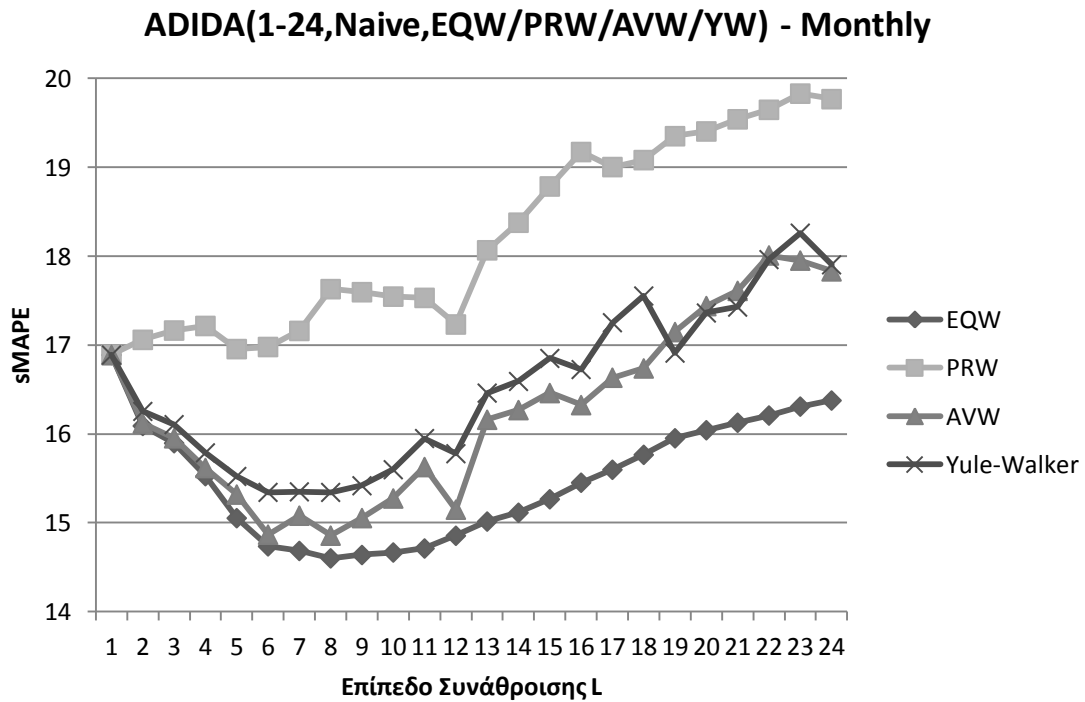
Η διαδικασία της επίλυσης μπορεί να απλοποιηθεί ελαφρώς επιλύοντας δύο επιμέρους συστήματα και εφαρμόζοντας υπέρθεση στις λύσεις, αντί να επιλυθεί και ένα τρίτο σύστημα. Τα δύο επιμέρους συστήματα είναι τα $Aw_1 = b$ και $Aw_2 = u$. Από το πρώτο σύστημα, το οποίο είναι το σύστημα που προκύπτει αν αγνοήσουμε τον περιορισμό κανονικοποίησης, ο υπολογισμός του w_1 μας εξασφαλίζει και τον όρο $\text{sum}(A^{-1}b) = \text{sum}(w_1)$. Από το δεύτερο σύστημα, ο υπολογισμός του w_2 μας εξασφαλίζει και τον όρο $\text{sum}(A^{-1}u) = \text{sum}(w_2)$. Η συνολική λύση προκύπτει από την υπέρθεση:

$$w = w_1 - \frac{\text{sum}(w_1) - 1}{\text{sum}(w_2)} w_2$$

Παρατηρούμε ότι αν εξασφαλίζεται απευθείας $\text{sum}(w_1) = 1$, το w_1 ταυτίζεται με την τελική λύση και δεν απαιτείται η επίλυση του δεύτερου γραμμικού συστήματος. Επίσης, αν $\text{sum}(w_1) \approx 1$, μπορούμε και πάλι να παραλείψουμε την επίλυση του δεύτερου συστήματος και αντί αυτού να ακολουθήσουμε την προσέγγιση της κανονικοποίησης του διανύσματος των βαρών, ανάλογα με τις απαιτήσεις μας σε ακρίβεια και τους διαθέσιμους πόρους.

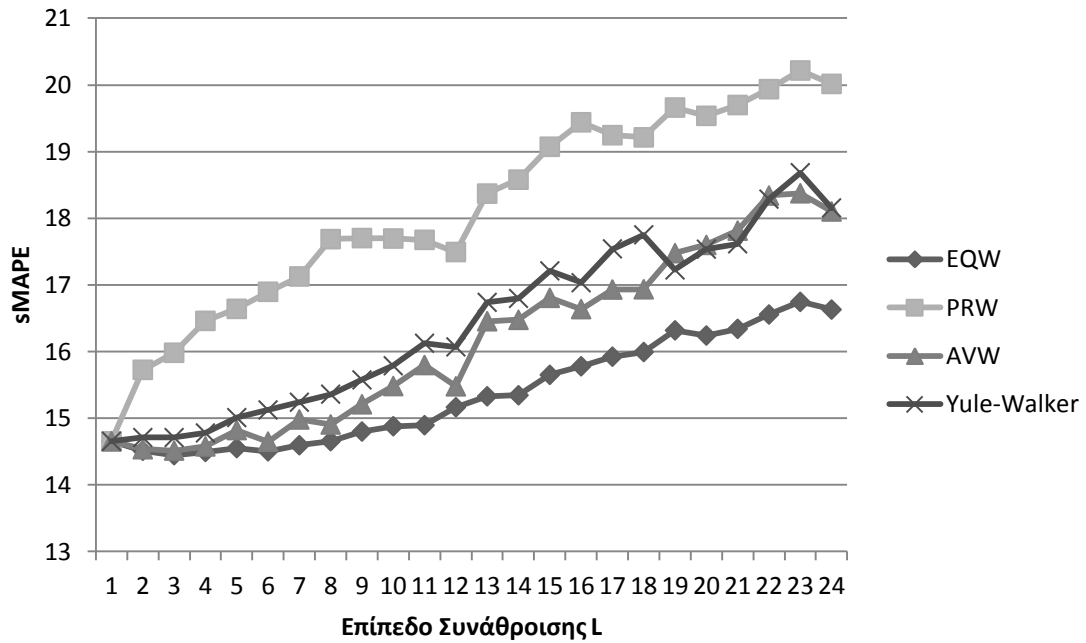
6.2 Επίδραση του Αλγορίθμου Διάσπασης

Για την εμπειρική μελέτη της επίδρασης του αλγορίθμου διάσπασης στην απόδοση της μεθόδου σε όρους μείωσης σφάλματος, στα ακόλουθα διαγράμματα παρουσιάζουμε το μέσο sMAPE out-of-sample σφάλμα πρόβλεψης 18 περιόδων για το σύνολο των μηνιαίων δεδομένων, με εφαρμογή της μεθοδολογίας ADIDA με κοινό επίπεδο συνάθροισης για όλες τις χρονοσειρές, συναρτήσει του κοινού επιπέδου συνάθροισης L . Για κάθε υπό μελέτη αλγόριθμο διάσπασης χαράσσεται διαφορετική καμπύλη, ενώ διαφορετικά γραφήματα σχεδιάζονται για διαφορετικές μεθόδους πρόβλεψης.



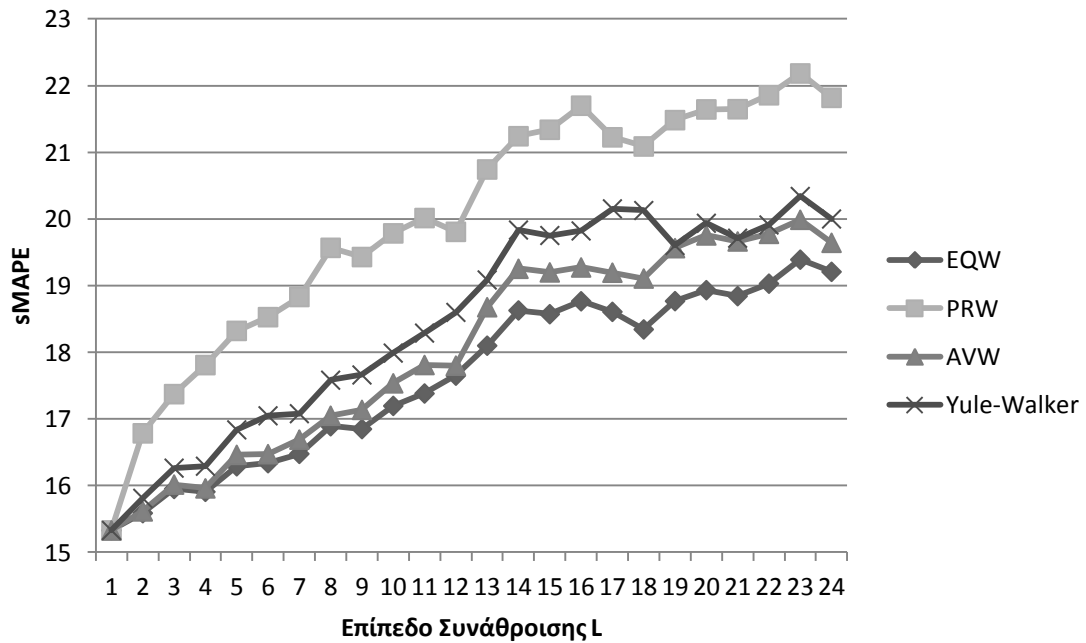
Εικόνα 6.1

ADIDA(1-24,SES,EQW/PRW/AVW/YW) - Monthly

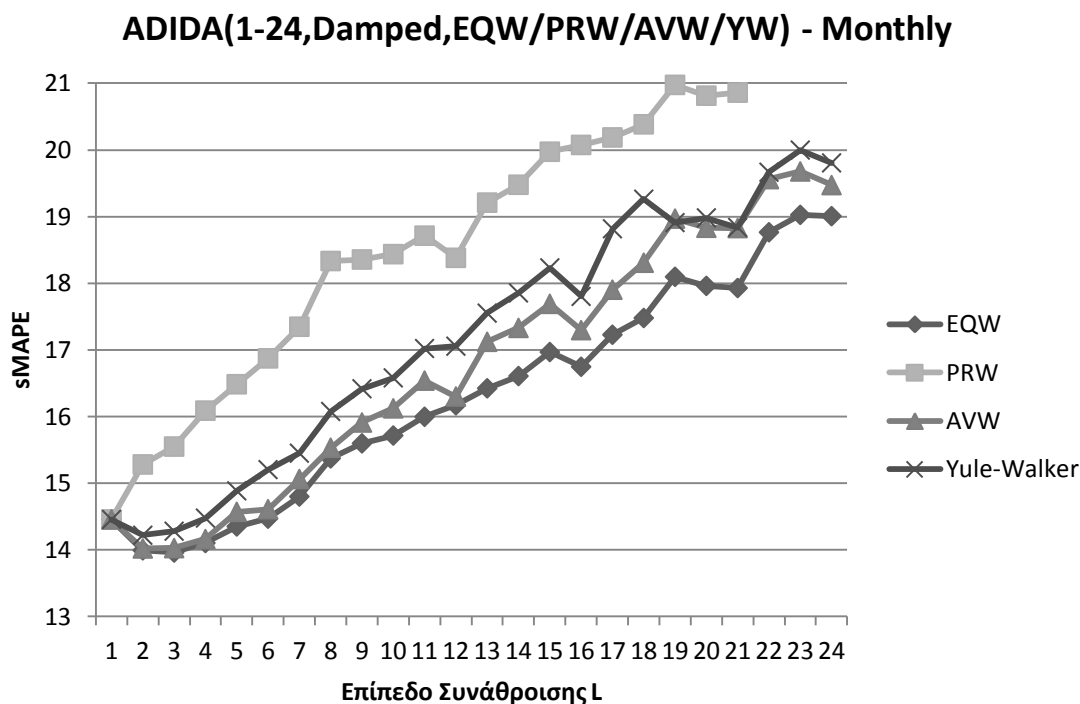


Εικόνα 6.2

ADIDA(1-24,Holt,EQW/PRW/AVW/YW) - Monthly



Εικόνα 6.3



Εικόνα 6.4

Σε όλα τα παραπάνω γραφήματα, σε γενικές γραμμές διατηρείται η διάταξη τεχνικών διάσπασης ως προς τα παραγόμενα σφάλματα των. Για το συγκεκριμένο σύνολο δεδομένων, η μέθοδος των ίσων βαρών (EQW) παρουσιάζει την καλύτερη συμπεριφορά, αφού αντιστοιχεί στην κατώτερη καμπύλη σφάλματος. Ακολουθεί η μέθοδος των μέσων βαρών (AVW) και αμέσως μετά η μέθοδος των εξισώσεων Yule-Walker. Η μέθοδος των προηγούμενων βαρών έχει τα χειρότερα αποτελέσματα, αφού οδηγεί μάλλον σε αύξηση παρά μείωση του σφάλματος.

Τα αποτελέσματα ταυτίζονται ακριβώς με εκείνα της έρευνας των Nikolopoulos et al. (2011). Μάλιστα, οι καμπύλες για τη μέθοδο Naïve (εικόνα 6.1) παρουσιάζουν αξιοπρόσεκτη ομοιότητα με τις αντίστοιχες καμπύλες της μελέτης. Η υπεροχή της μεθόδου EQW είχε τότε αποδοθεί στη στασιμότητα των δεδομένων που μελετήθηκαν, χαρακτηριστικό όμως που δεν υπάρχει στο σύνολο των μηνιαίων δεδομένων του διαγωνισμού M3. Συνεπώς, η αιτία της υπεροχής αυτής πρέπει να αναζητηθεί αλλού. Η ομοιότητα των αποτελεσμάτων υποδεικνύει δυνατότητες γενίκευσης των ευρημάτων και σε διαφορετικά σύνολα δεδομένων.

Η καλύτερη απόδοση της μεθόδου EQW έναντι των υπόλοιπων μεθόδων παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς αποτελεί τον μαθηματικά απλούστερο τρόπο επιλογής των βαρών. Επιπλέον, τα βάρη της EQW δεν συνδέονται με κανέναν τρόπο με τα στατιστικά χαρακτηριστικά της χρονοσειράς. Οι υπόλοιπες μέθοδοι (AVW, Yule-Walker, PRW), οι οποίες είναι πολυπλοκότερες και χρησιμοποιούν τις διαθέσιμες τιμές της χρονοσειράς, δίνουν φτωχότερα αποτελέσματα.

ΚΕΦΑΛΑΙΟ 7: Συμπεράσματα και Προεκτάσεις

7.1 Συμπεράσματα

Η παρούσα μελέτη αποτελεί μία επισταμένη προσπάθεια να συγκεντρωθούν και να συμπληρωθούν θεωρητικές αναλύσεις και εμπειρικά αποτελέσματα από την εφαρμογή της μεθοδολογίας πρόβλεψης ADIDA, με έμφαση στην χρήση της σε δεδομένα συνεχόμενης ζήτησης. Ταυτόχρονα, αποτελεί την πρώτη μελέτη της χρήσης της εν λόγω μεθοδολογίας, η οποία αναπτύχθηκε για τον χειρισμό δεδομένων διακοπτόμενης ζήτησης, για πρόβλεψη δεδομένων μη διακοπτόμενης φύσης, με σκοπό την εξακρίβωση της χρησιμότητας μίας τέτοιας πρακτικής. Στην ενότητα αυτή θα επιχειρηθεί η συνοπτική συγκέντρωση των κυριότερων ευρημάτων και συμπερασμάτων που προέκυψαν.

Η βιβλιογραφική μελέτη αποκάλυψε το πλήθος των επιδράσεων που επιφέρει η χρονική συνάθροιση στις χρονοσειρές και τα μοντέλα ερμηνείας και πρόβλεψής τους. Δείχθηκε ότι η μεθοδολογία ADIDA μπορεί να αναλυθεί σε επιμέρους απλούστερα και ευρέως μελετημένα στάδια. Ειδικότερα, αποδείχθηκε ότι η μεθοδολογία συνίσταται ουσιαστικά στην εν σειρά εφαρμογή ενός απλού εμπρός κινητού μέσου όρου, μίας μονάδας υποδειγματοληψίας, μίας μεθόδου προέκτασης, μίας μονάδας υπερδειγματοληψίας και ενός σταθμισμένου πίσω κινητού μέσου όρου (εικόνα 3.1). Η διάταξη στην οποία ανήχθη η μεθοδολογία παρουσιάζει εκπληκτικές ομοιότητες με τις διατάξεις πολυρυθμικής επεξεργασίας σήματος, για τις οποίες διατίθεται εκτενής βιβλιογραφία.

Ένα βασικό εύρημα είναι ότι όλα τα ευρήματα της εμπειρικής εφαρμογής βρίσκονται σε πλήρη συμφωνία με τα αποτελέσματα της μελέτης των Nikolopoulos et al. (2011). Επομένως, τα συμπεράσματά της επαληθεύονται, ενισχύονται και γενικεύονται και για την περίπτωση δεδομένων συνεχόμενης ζήτησης. Συνεπώς, η μεθοδολογία φαίνεται να λειτουργεί γενικότερα ως ένας μηχανισμός αυτοβελτίωσης της ακρίβειας των μεθόδων πρόβλεψης, ανεξαρτήτως του είδους της ζήτησης, διακοπτόμενης ή συνεχούς.

Η μεθοδολογία φάνηκε ότι μπορεί να έχει βελτιωτική επίδραση στις περιπτώσεις αυτόματης (μη ελεγχόμενης) συνάθροισης κατά τη διαδικασία συλλογής δεδομένων μεταβλητών ροής, οπότε τα υπόλοιπα στάδια μπορούν να εφαρμοστούν με τεχνητό (ελεγχόμενο) τρόπο για την παραγωγή ακριβέστερων προβλέψεων. Παρότι στην μη ελεγχόμενη περίπτωση δεν υπάρχει η δυνατότητα προεπεξεργασίας με αποεποχικοποίηση των δεδομένων, από την εικόνα 4.4 μπορούμε να συμπεράνουμε ότι ενδεχομένως η εκ των υστέρων επανεποχικοποίηση να βελτιώνει το αποτέλεσμα. Η δυνατότητα αυτή παρέχεται για υψηλότερα επίπεδα αυτόματης συνάθροισης, αφού τότε ο αρχικός ΚΜΟ προκαλεί

αποκοπή της συνιστώσας της εποχικότητας. Προσοχή όμως χρειάζεται ώστε να μην εφαρμοστεί η εποχικότητα σε μη αποεποχικοποιημένη σειρά. Επιπλέον περιορισμοί που συνδέονται με μία τέτοια προσέγγιση σχετίζονται με την ευκολία πρακτικής απόκτησης δεικτών εποχικότητας χωρίς την συλλογή του συνόλου των τιμών σε ένα κατώτερο επίπεδο παρατήρησης. Η πρακτική χρησιμότητα τότε της μεθοδολογίας είναι ενδεχομένως σημαντική όταν διατίθενται μονάχα απευθείας συναθροισμένα δεδομένα, είτε λόγω υψηλού κόστους λεπτότερης δειγματοληψίας είτε επειδή η πηγή των δεδομένων δεν μπορεί να τα παράσχει σε κατώτερο επίπεδο.

Ιδιαίτερη έμφαση δόθηκε στη σημασία, την επίδραση και την μελέτη των τρόπων επιλογής του επιπέδου συνάθροισης για τη μεθοδολογία, παρουσιάστηκε ότι, σε θεωρητικό επίπεδο, η αύξηση του επιπέδου συνάθροισης ταυτίζεται με αύξηση του μήκους των ΚΜΟ των σταδίων της ADIDA και, επομένως, με μία συνεπακόλουθη αύξηση των ικανοτήτων απόρριψης θορύβου και του βαθμού εξομάλυνσης που προκαλεί η μεθοδολογία. Παράλληλα, όμως, αποδείχθηκε ότι η αύξηση του επιπέδου συνάθροισης πάνω από μία ορισμένη τιμή, που εξαρτάται από το φασματικό περιεχόμενο της χρονοσειράς, σχετίζεται με την παρεμβολή επιπλέον θορύβου και τη στρέβλωση της χρονοσειράς μέσω του φαινομένου της φασματικής αναδίπλωσης (aliasing). Επιπλέον, πολύ μεγάλες τιμές του επιπέδου συνάθροισης επιφέρουν υπερβολική εξομάλυνση, καταπιέζοντας τελικά την πληροφορία της χρονοσειράς. Συνεπώς, στην επιλογή ενός επιπέδου συνάθροισης πρέπει να αποφεύγονται οι πολύ μεγάλες τιμές, όπου η έννοια της υψηλής τιμής ορίζεται σε συνάρτηση με τα φασματικά χαρακτηριστικά της χρονοσειράς. Τα παραπάνω συμπεράσματα δίνουν ενόραση στην επίδραση του επιπέδου συνάθροισης και αποτελούν υποδείξεις για την κατάλληλη επιλογή του.

Μελετήθηκαν δύο κύριες προσεγγίσεις για την επιλογή του επιπέδου συνάθροισης: η χρήση ενός κοινού επιπέδου συνάθροισης για το σύνολο των χρονοσειρών και η χρήση ξεχωριστού επιπέδου συνάθροισης ανά χρονοσειρά. Για την επιλογή κοινού επιπέδου συνάθροισης καταλήξαμε στο συμπέρασμα ότι οδηγεί σε μειώσεις των σφαλμάτων για ορισμένες μόνο μεθόδους πρόβλεψης. Επιπλέον, βρήκαμε ότι ενώ είναι αρκετά βελτιωτική για σύνολα δεδομένων που έχουν δειγματοληπτηθεί σε υψηλότερους ρυθμούς (π.χ. μηνιαία δεδομένα), η βελτίωση γίνεται ασθενέστερη για μικρότερους ρυθμούς δειγματοληψίας (π.χ. τριμηνιαία δεδομένα) και τελικά εξαλείφεται για ακόμα αραιότερη δειγματοληψία (π.χ. ετήσια δεδομένα). Η ομαλότητα των καμπυλών του μέσου σφάλματος συναρτήσει του επιπέδου συνάθροισης υποδεικνύει την ενδεχόμενη ύπαρξη ενός βέλτιστου κοινού επιπέδου συνάθροισης. Ωστόσο, θεωρείται ότι η τεχνική του κοινού επιπέδου είναι περισσότερο αποδοτική για ομοιογενή σύνολα δεδομένων, όπου δηλαδή όλες οι χρονοσειρές προέρχονται από κοινή πηγή και περιγράφουν σχετικά μεγέθη.

Βάσει των ευρημάτων, στην περίπτωση ανομοιογενών συνόλων δεδομένων, όπως τα δεδομένα του διαγωνισμού M3 που χρησιμοποιήθηκαν στη μελέτη αυτή, προτείνεται ως αποδοτικότερη η επιλογή ξεχωριστού βέλτιστου επιπέδου συνάθροισης ανά χρονοσειρά. Δείχθηκε ότι η πρότερη γνώση των εκ των υστέρων βέλτιστων τιμών του επιπέδου συνάθροισης για κάθε χρονοσειρά μπορεί να επιφέρει ιδιαίτερα σημαντικές μειώσεις στα επίπεδα του σφάλματος, αρκετές ποσοστιαίες μονάδες κάτω από τα αντίστοιχα σφάλματα των μεθόδων που αναδείχθηκαν νικήτριες στον διαγωνισμό M3. Αξιοπρόσεκτη ήταν επίσης η συμπεριφορά των καμπυλών του μέσου σφάλματος συναρτήσει του μέγιστου επιπέδου συνάθροισης, μέχρι το οποίο ελέγχονταν τα μοντέλα ADIDA για την επιλογή του βέλτιστου. Οι εν λόγω καμπύλες σφάλματος μειώνονται σε όλες τις χρονικές κατηγορίες δεδομένων και για όλες τις μεθόδους πρόβλεψης, αποκαλύπτοντας το δυναμικό της μεθοδολογίας ADIDA για βελτίωση των χρησιμοποιούμενων μεθόδων πρόβλεψης. Επιπλέον, ο φθίνων ρυθμός μείωσης του σφάλματος με την αύξηση του πλήθους των συγκρινόμενων μοντέλων και η ασυμπτωτική σύγκλιση της καμπύλης σε οριακές τιμές βρίσκονται σε απόλυτη συμφωνία με τη θεωρητική παρατήρηση ότι «μεγάλες» τιμές του επιπέδου συνάθροισης προκαλούν μάλλον επιδεινώση παρά βελτίωση. Έτσι, καθώς τα μεγαλύτερα μοντέλα δεν επιλέγονται, οι καμπύλες σφάλματος επηρεάζονται ολοένα και λιγότερο από την συμπερίληψή τους στη σύγκριση μεταξύ των μοντέλων.

Η απευθείας απόκτηση των *a posteriori* βέλτιστων τιμών του επιπέδου συνάθροισης, δεν είναι άμεσα πραγματοποιήσιμη. Αντί αυτής, είμαστε υποχρεωμένοι να χρησιμοποιήσουμε κάποια υλοποιήσιμη τεχνική εκτίμησης των βέλτιστων τιμών. Η ακριβής γνώση των *a posteriori* βέλτιστων αντιστοιχεί τότε στο αποτέλεσμα μίας τέλει μεθόδου εκτίμησης. Από την εξέταση των μεθόδων ελαχιστοποίησης των MSE, BIC και AIC αποκαλύφθηκε ότι, παρότι και οι τρεις μελετηθείσες μέθοδοι αποτυγχάνουν να ακολουθήσουν με ακρίβεια την μειούμενη και συγκλίνουσα πορεία της *a posteriori* καμπύλης, μπορούν να συνδεθούν ορισμένες φορές με σημαντικά οφέλη σε όρους σφάλματος, στις περισσότερες περιπτώσεις σημαντικότερα από εκείνα που εξασφαλίζει η χρήση κοινού επιπέδου συνάθροισης. Για παράδειγμα, με εκτίμηση μέσω του BIC ή του AIC το σφάλμα της Naïve για τα μηνιαία δεδομένα μειώνεται στα επίπεδα της νικήτριας μεθόδου Theta για τη συγκεκριμένη χρονική κατηγορία του διαγωνισμού M3. Γενικά, το κριτήριο MSE παρουσίασε τη φτωχότερη απόδοση, ενώ τα κριτήρια BIC και AIC έδωσαν άμεσα συγκρίσιμα αποτελέσματα. Ωστόσο, οι εκτιμήσεις των BIC και AIC οδήγησαν κατά τη θεώρηση μεγάλου αριθμού μοντέλων συχνά σε αποκλίσεις του σφάλματος από την τιμή αναφοράς του μοναδιαίου επιπέδου συνάθροισης, τη στιγμή που η εκτίμηση μέσω του MSE προκάλεσε μόνο μικρή επιδεινώση. Σε κάθε περίπτωση

όμως, η *a posteriori* καμπύλη υποδεικνύει ότι υπάρχουν επιπλέον περιθώρια βελτίωσης των μεθόδων εκτίμησης για την περαιτέρω μείωση του σφάλματος.

Στα πλαίσια της παρούσας μελέτης παρουσιάστηκαν και δοκιμάστηκαν εμπειρικά τέσσερις αλγόριθμοι διάσπασης με χρήση βαρών: ίσα (EQW), προηγούμενα (PRW) και μέσα (AVW) βάρη, καθώς και βάρη προσδιορισμένα μέσω των εξισώσεων Yule-Walker, οι οποίες τροποποιήθηκαν κατάλληλα για το σκοπό αυτό. Η εμπειρική εφαρμογή στα μηνιαία δεδομένα έδωσε σταθερά αποτελέσματα για όλες τις μεθόδους και όμοια με εκείνα των Nikolopoulos et al. (2011), επιτρέποντας να γίνουν κάποιες γενικεύσεις. Συγκεκριμένα, η απλούστερη μέθοδος EQW φαίνεται να δίνει μονίμως τα καλύτερα αποτελέσματα μεταξύ των μεθόδων που εξετάστηκαν. Ακολουθεί ο αλγόριθμος AVW και παραπλήσια αποτελέσματα προκύπτουν και με τον αλγόριθμο των Yule-Walker. Ο αλγόριθμος PRW φάνηκε να έχει τη χειρότερη απόδοση, οδηγώντας μάλλον σε επιδείνωση του σφάλματος παρά βελτίωση.

Συμπερασματικά, η κατάλληλη εφαρμογή της μεθοδολογίας, δηλαδή με τους κατάλληλους συνδυασμούς επιπέδου συνάθροισης, μεθόδου πρόβλεψης και αλγορίθμου διάσπασης, φαίνεται να μπορεί να επιφέρει ικανές μειώσεις του σφάλματος των προβλέψεων. Οι μειώσεις του σφάλματος σε πλήθος περιπτώσεων και τα υψηλά περιθώρια περαιτέρω βελτίωσης καθιστούν ιδιαίτερα ελκυστική την συνέχιση της μελέτης στον τομέα των μεθοδολογιών πρόβλεψης συνάθροισης-διάσπασης, με χαρακτηριστικό παράδειγμα τη μεθοδολογία ADIDA.

7.2 Προεκτάσεις

Επιπλέον των συμπερασμάτων και λοιπόν ευρημάτων που προκύπτουν από τη μελέτη αυτή, αποκαλύπτονται επίσης αρκετά ζητήματα που απαιτούν περαιτέρω διερεύνηση. Τα ζητούμενα αυτά πρέπει να αναλυθούν και να γίνουν προσπάθειες για εξεύρεση λύσεων και απαντήσεων, με σκοπό την προώθηση της επιστημονικής έρευνας και κατανόησης στον τομέα των προβλέψεων.

Αρχικά, πρέπει να αξιοποιηθούν και να εμπλουτιστούν τα ευρήματα της θεωρητικής ανάλυσης. Επιπρόσθετη μελέτη των σταδίων της διαδικασίας της μεθοδολογίας μπορεί να δώσει περισσότερες πληροφορίες για τις συνολικές ιδιότητές της και την κατάλληλη επιλογή των παραμέτρων. Ειδικότερα, μπορούν να διερευνηθούν σε μεγαλύτερο βάθος οι επιδράσεις του φαινομένου της φασματικής αναδίπλωσης και οι προεκτάσεις της ομοιότητας της μεθοδολογίας με τις πολυρυθμικές μεθόδους επεξεργασίας σήματος.

Περαιτέρω έρευνα θα πρέπει να συμπεριλάβει την επίδραση και άλλων μεθόδων πρόβλεψης, που ως τώρα έχουν δοκιμαστεί μόνο εκτός των πλαισίων της μεθοδολογίας ADIDA (όπως η μέθοδος Theta, Assimakopoulos and Nikolopoulos,

2000). Επιπροσθέτως, η επαλήθευση των αποτελεσμάτων αυτής της μελέτης σε διαφορετικά σύνολα δεδομένων, οιασδήποτε φύσης, αποτελεί χρήσιμη άσκηση, αλλά ταυτόχρονα μπορεί να αποκαλύψει πολλές λεπτομέρειες για τις δυνατότητες γενίκευσης των ευρημάτων και τη λειτουργία της μεθοδολογίας ως μηχανισμού αυτοβελτίωσης. Ειδικότερα, θα πρέπει να καλυφθούν περιπτώσεις που δεν μελετήθηκαν στην εργασία αυτή, όπως λεπτομερέστερα καταγεγραμμένα δεδομένα (εβδομαδιαία, ημερήσια, ωριαία κλπ). Επιπλέον, ιδιαίτερο ενδιαφέρον παρουσιάζει η μελέτη της αλληλεπίδρασης της διαδικασίας με διαδικασίες χειρισμού της εποχικότητας και η περισσότερο επικεντρωμένη έρευνα σε περιπτώσεις όπου τα δεδομένα διατίθενται απευθείας σε συναθροισμένη μορφή.

Τα εμπειρικά αποτελέσματα υποδεικνύουν ότι ίσως υφίσταται ένα βέλτιστο επίπεδο συνάθροισης. Αυτό ισχύει τόσο για το κοινό επίπεδο σε σύνολο χρονοσειρών, όσο και για το ξεχωριστό επίπεδο ανά χρονοσειρά. Ωστόσο περαιτέρω έρευνα απαιτείται για να εξεταστεί το ζήτημα σε μεγαλύτερο βάθος. Το ζήτημα αυτό σχετίζεται άμεσα με το γενικότερο ερώτημα του κατάλληλου χρονικού επιπέδου παραγωγής των προβλέψεων. Η ύπαρξη τυχόν βέλτιστων επιπέδων συνάθροισης πρέπει πρώτα να αποδειχθεί και στη συνέχεια τα επίπεδα αυτά να αναγνωριστούν με εξειδικευμένες τεχνικές που θα λαμβάνουν υπόψη όλα τα απαραίτητα χαρακτηριστικά και τις λανθάνουσες ιδιότητες εκάστοτε χρονοσειράς. Πρέπει να εξετασθεί η σχέση των βέλτιστων επιπέδων με τα ποιοτικά και ποσοτικά χαρακτηριστικά, όπως η στασιμότητα, η διακύμανση, το είδος δεδομένων της χρονοσειράς, κλπ.

Για την περίπτωση χρήσης κοινού επιπέδου συνάθροισης για το σύνολο των χρονοσειρών, πειράματα πρέπει να γίνουν με χρήση τόσο ομοιογενών (π.χ. ξεχωριστά σε κάθε κατηγορία δεδομένων του M3) και ανομοιογενών συνόλων για την αξιολόγηση της αποδοτικότητας μίας τέτοιας πρακτικής. Από την άλλη, για την προσέγγιση χρήσης ξεχωριστού επιπέδου συνάθροισης ανά χρονοσειρά, είναι απαραίτητη η ανεύρεση αποδοτικότερων μηχανισμών *a priori* εκτίμησης του βέλτιστου επιπέδου συνάθροισης, ώστε να επιτυγχάνονται αποτελέσματα εγγύτερα στα θεωρητικά *a posteriori* βέλτιστα. Τελικά, πρέπει να δοθούν τεκμηριωμένες απαντήσεις για τη φύση των δεδομένων που απαιτούν τη μία ή την άλλη αντιμετώπιση.

Τέλος, ως προς τη διάσπαση των δεδομένων, οφείλουμε να αναζητήσουμε τους παράγοντες που οδηγούν στην υπεροχή του αλγορίθμου των ίσων βαρών έναντι των υπολοίπων μεθόδων που μελετήθηκαν και, αν είναι δυνατόν, να αναπτύξουμε περισσότερο εκλεπτυσμένους αλγορίθμους που θα αποδίδουν καλύτερα.

Εν κατακλείδι, τα παραπάνω σημεία αποτελούν καταγραφή ενός συνόλου μόνο ζητημάτων που χρήζουν επισταμένης μελέτης και έρευνας. Το θεωρητικό

ενδιαφέρον, τα θετικά εμπειρικά αποτελέσματα και η διεπιστημονική εν δυνάμει χρησιμότητα που παρουσιάζει η μεθοδολογία πρόβλεψης ADIDA, καθώς και η διαρκής ανάγκη για εκσυγχρονισμό και ανανέωση στην επιστήμη των προβλέψεων ώστε να παρακολουθούνται στενά οι ρυθμοί της υπόλοιπης επιστημονικής προόδου, επιτάσσουν την συνέχιση της έρευνας για την απάντηση στο πλήθος των νέων ερωτημάτων που προκύπτουν.

Βιβλιογραφικές Αναφορές

- Akaike, H. (1974). 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716-723.
- Andrawis, R.R., Atiya, A.F., and El-Shishiny, H. (2011). 'Combination of long term and short term forecasts, with application to tourism demand forecasting', *International Journal of Forecasting*, Vol. 27, No. 3, pp. 870-886.
- Assimakopoulos, V., and Nikolopoulos, N. (2000). 'The theta model: a decomposition approach to forecasting', *International Journal of Forecasting*, Vol. 16, No. 4, pp. 521-530.
- Boot, J.C.G., Feibes, W., and Lisman, J.H.C (1967). 'Further Methods of Derivation of Quarterly Figures from Annual Data', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 6, No. 1.
- Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.), New Jersey: Prentice Hall.
- Brännäs, K. Hellström, J., and Nordström, J. (2002). 'A new approach to modelling and forecasting monthly guest nights in hotels', *International Journal of Forecasting*, Vol. 18, No. 1, pp. 19-30.
- Chan, W.S. (1993). 'Disaggregation of Annual Time-series Data to Quarterly Figures: A Comparative Study', *Journal of Forecasting*, Vol. 12, pp. 677-688.
- Croston, J.D. (1972). 'Forecasting and Stock Control for Intermittent Demands', *Operational Research Quarterly*, Vol. 23, No. 3, pp. 289-303.
- Dekker, M., van Donselaar, K., and Ouwehand, P. (2004). 'How to use aggregation and combined forecasting to improve seasonal demand forecasts', *International Journal of Production Economics*, Vol. 90, No. 2, pp. 151-167.
- Feijóo, S.R., Caro, A.R., and Quintana, D.D. (2003). 'Methods for quarterly disaggregation without indicators; a comparative study using simulation', *Computational Statistics & Data Analysis*, Vol. 43, No. 1, pp. 63-78.
- Gardner, E.S., and McKenzie, E. (1985). 'Forecasting trends in time series', *Management Science*, Vol. 31, pp. 1237-1246.
- González, P. (1992). 'Temporal Aggregation and Systematic Sampling in Structural Time-series Models', *Journal of Forecasting*, Vol. 11, pp. 271-281.
- Granger, C.W.J., and Siklos, P.L. (1995). 'Systematic sampling, temporal aggregation, seasonal adjustment, and cointegration theory and evidence', *Journal of Econometrics*, Vol. 66, No. 1-2, pp. 357-369.
- Harrison, G.W. (2008). 'Stocks and flows', *The New Palgrave Dictionary of Economics* (2nd ed.), Vol. 4, pp. 506-509.
- Hodgess, E.M., and Wei, W.W.S. (2000). 'Temporal disaggregation of stationary bivariate time series', Vol. 321, No. 1-3, pp. 175-196.

Koreisha, S.G., and Fang, Y. (2004). 'Updating ARMA Predictions for Temporal Aggregates', *Journal of Forecasting*, Vol. 23, pp. 275-296.

Lisman, J.H.C, and Sandee, J. (1964). 'Derivation of Quarterly Figures from Annual Data', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 13, No. 2.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982). 'The accuracy of extrapolation (time series) methods: results of a forecasting competition', *Journal of Forecasting*, Vol. 1, pp. 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., and Simmons, L.F. (1993). 'The M2-Competition: A real-time judgmentally based forecasting study', *International Journal of Forecasting*, Vol. 9, No. 1, pp. 5-22.

Makridakis, S., and Hibon, M. (1979). 'Accuracy of forecasting: an empirical investigation (with discussion)', *Journal of the Royal Statistical Society A*, Vol. 142, pp. 97-145.

Makridakis, S., and Hibon, M. (2000). 'The M3-Competition: results, conclusions and implications', *International Journal of Forecasting*, Vol. 16, No. 4, pp. 451-476.

Makridakis, S., Wheelwright, S.C., and Hyndman, R.J. (1998). *Forecasting: Methods and Applications* (3rd ed.), New York: John Wiley and Sons.

Man, K.S., and Tiao, G.C. (2006). 'Aggregation effect and forecasting temporal aggregates of long memory processes', *International Journal of Forecasting*, Vol. 22, pp. 267-281.

Marcellino, M. (1999). 'Some Consequences of Temporal Aggregation in Empirical Analysis', *Journal of Business & Economic Statistics*, Vol.17, No. 1, pp. 129-136.

Nikolopoulos, K., Syntetos, A.A., Boylan, J.E., Petropoulos, F., and Assimakopoulos, V. (2011). 'An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis', *Journal of Operational Research Society*, Vol. 62, pp. 544-554.

Schwarz, G.E. (1978). 'Estimating the dimension of a model', *Annals of Statistics*, Vol. 6, No. 2, pp. 461-464.

Silvestrini, A., and Veredas, D. (2008). *Temporal aggregation of univariate and multivariate time series models: a survey*. Temi di discussione (Working papers) 685 (August 2008), Banca d'Italia.

Syntetos, A.A., and Boylan, J.E. (2001). 'On the bias of intermittent demand estimates', *International Journal of Production Economics*, Vol. 71, pp. 457-466.

Tsai, H., and Chan, K.S. (2005). 'Temporal aggregation of stationary and non-stationary discrete-time processes', *Journal of Time Series Analysis*, Vol. 26, pp. 613-624.

Wei, W.W.S., and Stram, D.O. (1990). 'Disaggregation of time series models', *Journal of the Royal Statistical Society. Series B3*, Vol. 52, No. 3, pp. 453-467.

Weiss, A.A. (1984). 'Systematic sampling and temporal aggregation in time series models', *Journal of Econometrics*, Vol. 26, No. 3, pp. 271-281.

Willemain, T.R., Smart, C.N., and Schwarz, H.F. (1994). 'Forecasting intermittent demand in manufacturing : A comparative evaluation of Croston's method', *International Journal of Forecasting*, Vol. 10, No. 4, pp. 529-538.

Zotteri, G., and Kalchschmidt, M. (2007). 'A model for selecting the appropriate level of aggregation in forecasting processes', *International Journal of Production Economics*, Vol. 108, No. 1-2, pp. 74-83.

Zotteri, G., Kalchschmidt, M., and Caniato, F. (2005). 'The impact of aggregation level on forecasting performance', *International Journal of Production Economics*, Vol. 93-94, pp. 479-491.

Zellner, A., and Tobias, J. (2000). 'A Note on Aggregation, Disaggregation and Forecasting Performance', *Journal of Forecasting*, Vol. 19, pp. 457-469.

Παράρτημα Κώδικα Matlab

Για την εκτέλεση των πειραμάτων και τον υπολογισμό των αποτελεσμάτων, αναπτύχθηκαν κατάλληλα υποπρογράμματα στη γλώσσα μαθηματικού προγραμματισμού Matlab. Για την πληρότητα της εργασίας, παρατίθεται στο παράρτημα αυτού το σύνολο του ανεπτυγμένου κώδικα, ο οποίος περιλαμβάνει συναρτήσεις υλοποίησης των διαφόρων μεθόδων πρόβλεψης, των αλγορίθμων συνάθροισης και διάσπασης, της ίδιας της μεθοδολογίας ADIDA, των τεχνικών επιλογής του βέλτιστου επιπέδου συνάθροισης, καθώς και διαδικασίες εκτέλεσης των πειραμάτων. Ο κώδικας είναι διαθέσιμος και σε συνοδευτικό CD.

Π.1 Μεθοδολογία ADIDA και Διαδικασίες Πειραμάτων

```
function [y] = adida( x, horizon, L, F, w)
%-----
%Εφαρμόζει τη μεθοδολογία πρόβλεψης ADIDA
%----Είσοδοι----
%x:        διάνυσμα χρονοσειράς προς πρόβλεψη
%horizon:  ορίζοντα πρόβλεψης
%L:        επίπεδο συνάθροισης
%F:        μέθοδος πρόβλεψης
%          -'ses'
%          -'holt'
%          -'damped'
%          -'lrl'
%          -'theta'
%w:        αλγόριθμος διάσπασης
%          -'EQW'
%          -'PRW'
%          -'AVW'
%          -'YW1' (χωρίς κανονικοποίηση)
%          -'YW2' (με εκ των υστέρων κανονικοποίηση)
%          -'YW3 ή 'YW' (με ενσωματωμένη κανονικοποίηση)
%----Εξοδοι----
%y:        διάνυσμα μοντέλου πρόβλεψης
%-----
hor=floor((horizon-1)/L)+1;
%-----aggregate
[A offset]=aggregate(x,L);
%-----forecast
if strcmp(F,'naive')
    A2=naive(A,hor);
else
    %-----calculate level and trend (if needed)
    [Level Trend]=lr(A);
    if strcmp(F,'ses')
        A2=seslinear(A, 0:0.01:1, Level, hor);
    elseif strcmp(F,'holt')
        A2=holtlinear(A, 0:0.01:1, 0:0.01:1, Level, Trend, hor);
    elseif strcmp(F,'damped')
        %par=0:0.01:1;
        %A2=dampedlinear(A,par,par,par,Level,Trend,hor);
        A2=dampednonlinear(A, 7, Level, Trend, hor);
    end
end
```

```

elseif strcmp(F,'lrl')
    A2=lrmmodel( Level, Trend, length(A)+hor);
elseif strcmp(F,'theta')
    A2=theta( A, Level, Trend, hor);
end
end
%-----disaggregate
if strcmp(w,'EQW')
    w=EQW(L);
elseif strcmp(w,'PRW')
    w=PRW(A,x,L);
elseif strcmp(w,'AVW')
    w=AVW(A,x,L);
elseif strcmp(w,'YW1')
    w=YW1(A,x,L);
elseif strcmp(w,'YW2')
    w=YW2(A,x,L);
elseif strcmp(w,'YW3') || strcmp(w,'YW')
    w=YW3(A,x,L);
end
y=disaggregate(A2,L,w,offset);
%-----drop unwanted forecasts
y=y(1:end-(L*hor-horizon));
end

%-----
%ADIDA_UseSameL
%Χρήση κοινού επιπέδου συνάθροισης για
%όλες τις χρονοσειρές
%-----
load Monthly.mat;TimeseriesData=monthlyTimeseries;
SIData=monthlySI;Obs=monthlyObs;PPY=12;horizon=18;hidden=18;

%load Quarterly.mat;TimeseriesData=quarterlyTimeseries;
%SIData=quarterlySI;Obs=quarterlyObs;PPY=4;horizon=8;hidden=8;

%load Other.mat;TimeseriesData=otherTimeseries;
%SIData=num2cell(ones(size(TimeseriesData,2)));Obs=otherObs;
%PPY=1;horizon=8;hidden=8;

%load Yearly.mat;TimeseriesData=yearlyTimeseries;
%SIData=num2cell(ones(size(TimeseriesData,2)));
%Obs=yearlyObs;PPY=1;horizon=6;hidden=6;

Lsearch=1:10;
isearch=1:length(TimeseriesData);
tic;
for i=isearch
    %-----select a series
    ts=TimeseriesData{i};
    SI=SIData{i};
    N=length(ts);
    %-----keep insample (hide outsample)
    x=ts(1:N-hidden);
    %-----deseasonalize
    d=deseasonalize(x,SI,0);
    %-----apply adida methodology
    for L=Lsearch

```

```

    [i L]
    %-----adida(x,horizon,AggregationLevel,
    %ForecastMethod,DisaggregationAlgorithmOrWeights)
    D= adida( d, horizon, L, 'naive', 'EQW');
    %-----remove zero values
    D(N-hidden+find(D(N-hidden+1:N-hidden+horizon)<0))=d(end);
    %-----reseasonalize
    y=reseasonalize(D,SI,0);
    %figure(); plot(ts);hold on;plot(y,'r');
    %-----calculate outsample errors
    error(i-min(isearch)+1,L-min(Lsearch)+1) = smape(ts(N-
hidden+1:N-hidden+horizon),y(N-hidden+1:N-hidden+horizon));
    end
    bestL(i-min(isearch)+1)=find(error(i-
min(isearch)+1,:)==min(error(i-min(isearch)+1,:),1));
end
toc
if max(size(Lsearch)>1)
    figure();
    plot(Lsearch,mean(error,1),'b');
end
min(mean(error,1))
%-----

%-----
%ADIDA_UseDifferentL
%Χρήση διαφορετικού επιπέδου συνάθροισης
%ανά χρονοσειρά
%-----

load Monthly.mat;TimeseriesData=monthlyTimeseries;
SIData=monthlySI;Obs=monthlyObs;PPY=12;horizon=18;hidden=18;

%load Quarterly.mat;TimeseriesData=quarterlyTimeseries;
%SIData=quarterlySI;Obs=quarterlyObs;PPY=4;horizon=8;hidden=8;

%load Other.mat;TimeseriesData=otherTimeseries;
%SIData=num2cell(ones(size(TimeseriesData,2)));Obs=otherObs;
%PPY=1;horizon=8;hidden=8;

%load Yearly.mat;TimeseriesData=yearlyTimeseries;
%SIData=num2cell(ones(size(TimeseriesData,2)));
%Obs=yearlyObs;PPY=1;horizon=6;hidden=6;

for kkk=10:10 %μέγιστες τάξεις μοντέλων που θα εξεταστούν

Lsearch=1:kkk;
isearch=1:length(TimeseriesData);

forecastmethod='damped';
disaggregation='EQW';
hidden2=0;
tic;
for i=isearch
    %-----select a series
    ts=TimeseriesData{i};
    SI=SIData{i};
    N=length(ts);
    %-----keep insample (hide outsample)

```

```

x=ts(1:N-hidden-hidden2);
%-----deseasonalize
d=deseasonalize(x,SI,0);
%-----apply adida methodology
for L=Lsearch
    errorcalstart=mod(length(d),L)+1;
    [i L]
    %-----adida(x,horizon,AggregationLevel,
    %ForecastMethod,DisaggregationAlgorithmOrWeights)
    D=adida(d, hidden2, L, forecastmethod, disaggregation);
    %-----remove zero values
    if hidden2>0
        D(N-hidden+find(D(N-hidden+1:N-
hidden+horizon)<0))=d(end);
    end
    %-----reseasonalize
    y=reseasonalize(D,SI,0);
    %figure(); plot(ts);hold on; plot(y,'r');
    %-----calculate outsample errors
    error(i-min(isearch)+1,L-
min(Lsearch)+1)=mse(ts(errorcalstart:N-hidden),y(errorcalstart:N-
hidden));
    %error(i-min(isearch)+1,L-
min(Lsearch)+1)=bic(ts(errorcalstart:N-hidden),y(errorcalstart:N-
hidden),L);
    %error(i-min(isearch)+1,L-
min(Lsearch)+1)=aic(ts(errorcalstart:N-hidden),y(errorcalstart:N-
hidden),L);
    end
    bestL(i-min(isearch)+1)=Lsearch(find(error(i-
min(isearch)+1,:)==min(error(i-min(isearch)+1,:),1)));
    %-----keep insample (hide outsample)
    x=ts(1:N-hidden);
    %-----deseasonalize
    d=deseasonalize(x,SI,0);
    %-----adida(x,horizon,AggregationLevel,
    %ForecastMethod,DisaggregationAlgorithmOrWeights)
    D=adida(d, horizon, bestL(i-min(isearch)+1), forecastmethod,
disaggregation);
    %-----remove zero values
    D(N-hidden+find(D(N-hidden+1:N-hidden+horizon)<0))=d(end);
    %-----reseasonalize
    y=reseasonalize(D,SI,0);
    %figure(); plot(ts);hold on; plot(y,'r');
    %-----calculate outsample errors
    errorbest(i-min(isearch)+1)=smape(ts(N-hidden+1:N-
hidden+horizon),y(N-hidden+1:N-hidden+horizon));
end
toc
mean(errorbest)
a(:,kkk)=errorbest';
end

```

Π.2 Μέθοδοι Πρόβλεψης

```

function [ y ] = naive( x, hor )
%-----
%Εφαρμόζει τη μέθοδο πρόβλεψης Naive
%----Είσοδοι----

```

```

%x:   διάνυσμα χρονοσειράς προς πρόβλεψη
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι-----
%y:   διάνυσμα μοντέλου πρόβλεψης
%-----
    N=length(x);
    y=zeros(1,N+hor);
    y(1:N+1)=[x(1) x];
    y((N+2):end)=y(N+1);
end

function [ y ] = ses( x, a, S0, hor)
%-----
%Εφαρμόζει τη μέθοδο πρόβλεψης SES
%----Είσοδοι-----
%x:   διάνυσμα χρονοσειράς προς πρόβλεψη
%a:   παράμετρος εξομάλυνσης α
%S0:  αρχικό επίπεδο
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι-----
%y:   διάνυσμα μοντέλου πρόβλεψης
%-----
    N=length(x);
    A=length(a);
    a=reshape(a,A,1);
    y=zeros(A,N+hor);
    S=zeros(A,N+hor);
    e=zeros(A,N+hor);
    y(:,1)=S0;
    e(:,1)=x(1)-y(:,1);
    S(:,1)=S0+a.*e(:,1);
    for i=2:N
        y(:,i)=S(:,i-1);
        e(:,i)=x(:,i)-y(:,i);
        S(:,i)=S(:,i-1)+a.*e(:,i);
    end
    y(:,(N+1):end)=a.*x(N)+(1-a).*y(:,N);
end

function [ y ] = holt( x, a, b, S0, T0, hor )
%-----
%Εφαρμόζει τη μέθοδο πρόβλεψης Holt
%----Είσοδοι-----
%x:   διάνυσμα χρονοσειράς προς πρόβλεψη
%a:   παράμετρος εξομάλυνσης α
%b:   παράμετρος εξομάλυνσης β
%S0:  αρχικό επίπεδο
%T0:  αρχική τάση
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι-----
%y:   διάνυσμα μοντέλου πρόβλεψης
%-----
    N=length(x);
    A=length(a);
    B=length(b);
    M=A*B;
    a=reshape(a,A,1);
    b=reshape(b,B,1);
    [a b]=combineab(a,b);

    y=zeros(M,N+hor);

```

```

S=S0;
T=T0;
y(:,1)=S0+T0;

for i=1:N
    Sold=S;
    Told=T;
    S=a(:).*x(i)+(1-a(:)).*(Sold+Told);
    T=b(:).*(S-Sold)+(1-b(:)).*Told; %version 1 (makridakis)
    %T=b(:).*(x(i)-Sold)+(1-b(:)).*Told; %version 2
    y(:,i+1)=S+T;
end
for i=N+2:N+hor
    y(:,i)=y(:,i-1)+T;
end

end

function [ y ] = damped( x, a, b, f, S0, T0 ,hor)
%-----Είσοδοι-----
%Εφαρμόζει τη μέθοδο πρόβλεψης Damped
%----Είσοδοι----
%x: διάνυσμα χρονοσειράς προς πρόβλεψη
%a: παράμετρος εξομάλυνσης α
%b: παράμετρος εξομάλυνσης β
%f: παράμετρος εξομάλυνσης φ
%S0: αρχικό επίπεδο
%T0: αρχική τάση
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι-----
%y: διάνυσμα μοντέλου πρόβλεψης
%-----

A=length(a);
B=length(b);
F=length(f);
N=length(x);
M=A*B*F;

a=reshape(a,A,1);
b=reshape(b,B,1);
f=reshape(f,F,1);
[a b f]=combineabf(a,b,f);

y=zeros(M,N+hor);

S=S0;
T=T0;
y(:,1)=S0+f(:).*T0;

for i=1:N
    Sold=S;
    Told=T;
    S=Sold+Told+a(:).*(x(i)-y(:,i));
    T=b(:).*(S-Sold)+(1-b(:)).*f(:).*Told;%version 1 (makridakis)
    %T=Told+b(:).*(x(i)-y(:,i)); %version 2
    y(:,i+1)=S+f(:).*T;
end
for i=N+2:N+hor
    y(:,i)=y(:,i-1)+(f(:).^ (i-N)).*T;
end

```

```

    end
end

function [ y besta ] = seslinear( x, a, S0, hor)
%-----
%Εφαρμόζει τη (γραμμικά) βελτιστοποιημένη ως
%προς τις παραμέτρους μέθοδο πρόβλεψης SES
%----Είσοδοι----
%x: διάνυσμα χρονοσειράς προς πρόβλεψη
%a: διάνυσμα παραμέτρων εξομάλυνσης α
%   που εξετάζονται για τη βελτιστοποίηση
%S0: αρχικό επίπεδο
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι-----
%y: διάνυσμα μοντέλου πρόβλεψης
%besta: βέλτιστη παράμετρος εξομάλυνσης α
%-----

    y=ses(x, a, S0, 1);
    error=mse(x,y(:,1:end-1));
    i= find(error==min(error),1);
    besta=a(i);
    y=ses( x, besta, S0, hor);
end

function [ y besta bestb ] = holtlinear( x, a, b, S0, T0, hor )
%-----
%Εφαρμόζει τη (γραμμικά) βελτιστοποιημένη ως
%προς τις παραμέτρους μέθοδο πρόβλεψης Holt
%----Είσοδοι----
%x: διάνυσμα χρονοσειράς προς πρόβλεψη
%a: διάνυσμα παραμέτρων εξομάλυνσης α
%   που εξετάζονται για τη βελτιστοποίηση
%b: διάνυσμα παραμέτρων εξομάλυνσης β
%   που εξετάζονται για τη βελτιστοποίηση
%S0: αρχικό επίπεδο
%T0: αρχική τάση
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι-----
%y: διάνυσμα μοντέλου πρόβλεψης
%besta: βέλτιστη παράμετρος εξομάλυνσης α
%bestb: βέλτιστη παράμετρος εξομάλυνσης β
%-----

    y=holt(x, a, b, S0, T0, 1);
    error=mse(x,y(:,1:end-1));
    i= find(error==min(error),1);
    [a b]=combineab(a,b);
    besta=a(i);
    bestb=b(i);
    y=holt( x, besta, bestb, S0, T0, hor);
end

function [y besta bestb bestf]=dampedlinear(x, a, b, f, S0, T0, hor)
%-----
%Εφαρμόζει τη (γραμμικά) βελτιστοποιημένη ως
%προς τις παραμέτρους μέθοδο πρόβλεψης Damped
%----Είσοδοι----
%x: διάνυσμα χρονοσειράς προς πρόβλεψη
%a: διάνυσμα παραμέτρων εξομάλυνσης α
%   που εξετάζονται για τη βελτιστοποίηση
%b: διάνυσμα παραμέτρων εξομάλυνσης β
%   που εξετάζονται για τη βελτιστοποίηση

```

```

%f:   διάνυσμα παραμέτρων εξομάλυνσης φ
%     που εξετάζονται για τη βελτιστοποίηση
%S0:  αρχικό επίπεδο
%T0:  αρχική τάση
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι----
%y:   διάνυσμα μοντέλου πρόβλεψης
%besta: βέλτιστη παράμετρος εξομάλυνσης α
%bestb: βέλτιστη παράμετρος εξομάλυνσης β
%bestf: βέλτιστη παράμετρος εξομάλυνσης φ
%-----
y=damped(x, a, b, f, S0, T0, 0);
for i=1:size(y,1)
    y(i,:)=y(i,)-x;
end
y=y.^2;
error = mean(y,2);
i= find(error==min(error),1);
[a b f]=combineabf(a,b,f);
besta=a(i);
bestb=b(i);
bestf=f(i);
y=damped( x, besta, bestb, bestf, S0, T0, hor);
end

function [y besta bestb bestf] = dampednonlinear(x, n, S0, T0, hor)
%-----
%Εφαρμόζει τη (μη γραμμικά) βελτιστοποιημένη ως
%προς τις παραμέτρους μέθοδο πρόβλεψης Damped
%----Είσοδοι----
%x:   διάνυσμα χρονοσειράς προς πρόβλεψη
%n:   πλήθος επαναλήψεων διαδικασίας βελτιστοποίησης
%S0:  αρχικό επίπεδο
%T0:  αρχική τάση
%hor: ορίζοντας πρόβλεψης
%----Εξοδοι----
%y:   διάνυσμα μοντέλου πρόβλεψης
%besta: βέλτιστη παράμετρος εξομάλυνσης α
%bestb: βέλτιστη παράμετρος εξομάλυνσης β
%bestf: βέλτιστη παράμετρος εξομάλυνσης φ
%-----
a=[0.33 0.667];
b=[0.33 0.667];
f=[0.33 0.667];
dV=a(1)/2;
for k=1:n
    y=damped(x, a, b, f, S0, T0, 0);
    y=y(:,1:size(x,2));
    for i=1:size(y,1)
        y(i,:)=y(i,)-x;
    end
    y=y.^2;
    error = mean(y,2);
    i= find(error==min(error),1);
    [a b f]=combineabf(a,b,f);
    besta=a(i);
    bestb=b(i);
    bestf=f(i);
    a=[besta-dV besta+dV];
    b=[bestb-dV bestb+dV];
    f=[bestf-dV bestf+dV];
end

```



```

        dV=dV/2;
    end
    y=damped( x, besta, bestb, bestf, S0, T0, hor);
end

function [ x ] = lrmodel( Level, Trend, N )
%-----
%Εφαρμόζει πρόβλεψη επεκτείνοντας την ευθεία με τις
%δεδομένες παραμέτρους (τυπικά, ευθεία ελαχίστων τετραγώνων)
%----Είσοδοι----
%Level: το επίπεδο της ευθείας
%Trend: η τάση της ευθείας
%N:      μήκος του μοντέλου πρόβλεψης
%----Εξοδοι-----
%x: διάνυσμα μοντέλου πρόβλεψης
%-----
    x=Level+Trend*(1:N);
end

```

Π.3 Συνάθροιση και Αλγόριθμοι Αποσύνθεσης

```

function [ y, offset ] = aggregate( x, L )
%-----
%Προκαλεί χρονική συνάθροιση των δεδομένων
%----Είσοδοι----
%x:      διάνυσμα αρχικής χρονοσειράς
%L:      επίπεδο συνάθροισης
%----Εξοδοι-----
%y:      διάνυσμα συναθροισμένης χρονοσειράς
%offset: πλήθος παρατηρήσεων της αρχικής συμπεριφοράς που δεν
%        θα συμμετάσχουν στο μοντέλο ADIDA
%-----
    N=length(x);
    %αποκοπή πρώτων παρατηρήσεων που πλεονάζουν
    offset=mod(N,L);
    x=x(1+offset:end);
    %υπολογισμός αθροισμάτων
    y=filter(ones(1,L),1,x);
    %κρατάμε κάθε L-στό άθροισμα
    y=downsample(y,L, L-1);
end

function [ y ] = disaggregate( x, L, w, offset )
%-----
%Προκαλεί χρονική συνάθροιση των δεδομένων
%----Είσοδοι----
%x:      διάνυσμα συναθροισμένης χρονοσειράς
%L:      επίπεδο συνάθροισης
%w:      διάνυσμα βαρών διάσπασης
%offset: πλήθος αχρησιμοποίητων αρχικών παρατηρήσεων
%----Εξοδοι-----
%y:      διάνυσμα διασπασμένης χρονοσειράς
%-----
    x=upsample(x,L);
    y=filter(w,1,x);
    y=[ones(1,offset)*y(1) y];
end

function [ w ] = EQW( L )

```

```

%-----
%Υπολογίζει τα βάρη της μεθόδου EQW
%----Είσοδοι----
%L: το επίπεδο συνάθροισης
%----Εξοδοι-----
%w: το διάνυσμα των βαρών
%-----
w=ones(1,L)/L;
end

function [ w ] = PRW( ag, or, L)
%-----
%Υπολογίζει τα βάρη της μεθόδου PRW
%----Είσοδοι----
%ag: συναθροισμένη χρονοσειρά
%or: αρχική χρονοσειρά
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%w: το διάνυσμα των βαρών
%-----
w=or(end-L+1:end)/ag(end);
end

function [ w ] = AVW( ag, or, L)
%-----
%Υπολογίζει τα βάρη της μεθόδου AVW
%----Είσοδοι----
%ag: συναθροισμένη χρονοσειρά
%or: αρχική χρονοσειρά
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%w: το διάνυσμα των βαρών
%-----
N=length(or);
offset=mod(N,L);
or=or(1+offset:end);
buf=buffer(or,L);
w=buf./ repmat(ag,L,[]);
w=mean(w,2)';
end

function [ w A ] = YW1( ag, or, L)
%-----
%Υπολογίζει τα βάρη της μεθόδου YW (χωρίς κανονικοποίηση)
%----Είσοδοι----
%ag: συναθροισμένη χρονοσειρά
%or: αρχική χρονοσειρά
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%w: το διάνυσμα των βαρών
%A: ο πίνακας του συστήματος της μεθόδου
%-----
N=length(or);
offset=mod(N,L);
y=or(1+offset:end);
N=length(y);

x=upsample(ag,L);
x=x(1:N);

```

```

yxk=xcorr(y,x,L);
yxk=yxk(L+2:end);
xmxk=xcorr(x,x,L);
xmxk=xmxk(2:end-1);

A=zeros(L,L);
for j=1:L
    A(:,j)=xmxk(L+1-j:2*L-j)';
end
b=yxk';
w=A\b;
end

function [ w ] = YW2( ag, or, L)
%-----
%Υπολογίζει τα βάρη της μεθόδου YW (με εκ των υστέρων κανονικοποίηση)
%----Είσοδοι----
%ag: συναθροισμένη χρονοσειρά
%or: αρχική χρονοσειρά
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%w: το διάνυσμα των βαρών
%-----
w=YW1(ag,or,L);
w=w/sum(w);
end

function [ w ] = YW3( ag, or, L)
%-----
%Υπολογίζει τα βάρη της μεθόδου YW (με ενσωματωμένη κανονικοποίηση)
%----Είσοδοι----
%ag: συναθροισμένη χρονοσειρά
%or: αρχική χρονοσειρά
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%w: το διάνυσμα των βαρών
%-----
u=ones(L,1);
[w1 A]=YW1(ag,or,L);
w2=A\u;
l=2*(sum(w1)-1)/sum(w2);
w=w1-l/2*w2;
end

```

Π.4 Σφάλματα και Τεχνικές Εκτίμησης του Βέλτιστου Επιπέδου Συνάθροισης

```

function [ error ] = smape( x,y )
%-----
%Υπολογίζει το SMAPE σφάλμα
%----Είσοδοι----
%x: διάνυσμα πραγματικών τιμών
%y: διάνυσμα εκτιμώμενων τιμών
%----Εξοδοι-----
%error: το υπολογισμένο SMAPE σφάλμα
%-----
error= 2*mean(abs((y-x)./(y+x)))*100;
end

```

```

function [ error ] = mse( x,y )
%-----
%Υπολογίζει το MSE σφάλμα
%----Είσοδοι----
%x: διάνυσμα πραγματικών τιμών
%y: διάνυσμα εκτιμώμενων τιμών
%----Εξοδοι-----
%error: το υπολογισμένο MSE σφάλμα
%-----
x=repmat(x,size(y,1),1);
error = mean((y-x).^2,2);
end

```

```

function [ a ] = bic( x, y, L )
%-----
%Υπολογίζει την τιμή του BIC
%----Είσοδοι----
%x: διάνυσμα πραγματικών τιμών
%y: διάνυσμα εκτιμώμενων τιμών
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%a: τιμή του κριτηρίου BIC
%-----
N=length(x);
error=mse(x,y);
a=N*reallog(error)+L*reallog(N);
end

```

```

function [ a ] = aic( x, y, L )
%-----
%Υπολογίζει την τιμή του AIC
%----Είσοδοι----
%x: διάνυσμα πραγματικών τιμών
%y: διάνυσμα εκτιμώμενων τιμών
%L: επίπεδο συνάθροισης
%----Εξοδοι-----
%a: τιμή του κριτηρίου AIC
%-----
n=length(x);
error=mse(x,y);
a=n*reallog(error)+2*L;
end

```

Π.5 Λοιπές Βοηθητικές Συναρτήσεις

```

function [ y ] = deseasonalize( x, SI, offset )
%-----
%Αφαιρεί την εποχικότητα από μία χρονοσειρά
%----Είσοδοι----
%x: διάνυσμα εποχικοποιημένης χρονοσειράς
%SI: διάνυσμα δεικτών εποχικότητας
%offset: μετατόπιση πρώτου δείκτη σε σχέση
% με την πρώτη περίοδο της χρονοσειράς
%----Εξοδοι-----
%y: διάνυσμα αποεποχικοποιημένης
%-----
seasonality=length(SI);

```

```

    y=x;
    for i=1:length(x)
        y(i)=x(i)/SI(mod(i-1+offset,seasonality)+1);
    end
end

```

```

function [ y ] = reseasonalize( x, SI, offset )
%-----
%Εφαρμόζει εποχικότητα σε χρονοσειρά
%----Είσοδοι----
%x:      διάνυσμα αποεποχικοποιημένης χρονοσειράς
%SI:     διάνυσμα δεικτών εποχικότητας
%offset: μετατόπιση πρώτου δείκτη σε σχέση
%        με την πρώτη περίοδο της χρονοσειράς
%----Εξοδοι----
%y:      διάνυσμα επανεποχικοποιημένης
%-----
    seasonality=length(SI);
    y=x;
    for i=1:length(x)
        y(i)=x(i)*SI(mod(i-1+offset,seasonality)+1);
    end
end

```

```

function [ Level Trend ] = lr( x )
%-----
%Υπολογίζει τις παραμέτρους της ευθείας ελαχίστων τετραγώνων
%----Είσοδοι----
%x:      διάνυσμα χρονοσειράς προς πρόβλεψη
%----Εξοδοι----
%Level:  το επίπεδο της ευθείας
%Trend:  η τάση της ευθείας
%-----
    t=1:length(x);
    xbar=mean(x);
    tbar=mean(t);
    Trend=sum((x-xbar).*(t-tbar))/sum((t-tbar).^2);
    Level=xbar-Trend*tbar;
end

```

```

function [ a b ] = combineab( a, b )
%-----
%Υπολογίζει συνδυασμούς παραμέτρων για χρήση στη βελτιστοποίηση
%----Είσοδοι----
%a:      διάνυσμα παραμέτρων εξομάλυνσης α
%b:      διάνυσμα παραμέτρων εξομάλυνσης β
%----Εξοδοι----
%a:      συνδυασμένο διάνυσμα παραμέτρων εξομάλυνσης α
%b:      συνδυασμένο διάνυσμα παραμέτρων εξομάλυνσης β
%-----
    A=length(a);
    B=length(b);
    a=upsample(a,B);
    a=filter(ones(1,B),1,a);
    b=repmat(b,1,A);
end

```

```

function [ a b f ] = combineabf( a, b, f )
%-----
%Υπολογίζει συνδυασμούς παραμέτρων για χρήση στη βελτιστοποίηση

```

```

%----Είσοδοι----
%a:  διάνυσμα παραμέτρων εξομάλυνσης α
%b:  διάνυσμα παραμέτρων εξομάλυνσης β
%f:  διάνυσμα παραμέτρων εξομάλυνσης φ
%----Εξοδοι-----
%a:  συνδυασμένο διάνυσμα παραμέτρων εξομάλυνσης α
%b:  συνδυασμένο διάνυσμα παραμέτρων εξομάλυνσης β
%f:  συνδυασμένο διάνυσμα παραμέτρων εξομάλυνσης φ
%-----
A=length(a);
B=length(b);
F=length(f);
[a b]=combineab(a,b);

a=upsample(a,F);
a=filter(ones(1,F),1,a);
b=upsample(b,F);
b=filter(ones(1,F),1,b);

f=repmat(f,1,A*B);
end

```