



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και
Συστημάτων Αποφάσεων

Σύγκριση Απόδοσης Συμβατικών
Παλινδρομικών Μοντέλων και
Μοντέλων Πρόβλεψης Χρονοσειρών με
Χρονική Συνάθροιση σε Δεδομένα
Ενεργειακής Κατανάλωσης Σ.Η.Ε.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ορέστης Α. Αλμπανούδης

Υπεύθυνος : Ευάγγελος Σπηλιώτης
Υποψήφιος Διδάκτωρ Ε.Μ.Π.

Επιβλέπων : Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και
Συστημάτων Αποφάσεων

Σύγκριση Απόδοσης Συμβατικών
Παλινδρομικών Μοντέλων και
Μοντέλων Πρόβλεψης Χρονοσειρών με
Χρονική Συνάθροιση σε Δεδομένα
Ενεργειακής Κατανάλωσης Σ.Η.Ε.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ορέστης Α. Αλμπανούδης

Υπεύθυνος : Ευάγγελος Σπηλιώτης
Υποψήφιος Διδάκτωρ Ε.Μ.Π.

Επιβλέπων : Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18^η Μαρτίου 2016.

.....
Βασίλειος Ασημακόπουλος

.....
Ιωάννης Ψαρράς

.....
Δημήτριος Ασκούνης

.....

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

Copyright © Αλμπανούδης Ορέστης, 2016
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.

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο των ερευνητικών δραστηριοτήτων του Εργαστηρίου της Μονάδας Προβλέψεων και Στρατηγικής κατά το ακαδημαϊκό έτος 2015-2016. Η μονάδα υπάγεται στον τομέα Συστημάτων Απόφασης της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα να εκφράσω τις θερμές και ειλικρινείς μου ευχαριστίες στον Καθηγητή της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου κ. Βασίλειο Ασημακόπουλο για την ευκαιρία που μου έδωσε να ασχοληθώ με το αντικείμενο των προβλέψεων ενεργειακής κατανάλωσης.

Επίσης θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου κ. Ευάγγελο Σπηλιώτη για τη συνεχή παρακολούθηση της πορείας της διπλωματικής μου εργασίας, την καθοδήγησή του και το ενδιαφέρον που έδειξε καθ' όλη τη διάρκεια εκπόνησης της εργασίας.

Αλμπανούδης Ορέστης

Αθήνα, Μάρτιος 2016

Περίληψη

Η παρούσα διπλωματική στοχεύει στη διερεύνηση της καταληλότητας μοντέλων πρόβλεψης σε ιεραρχίες χρονοσειρών κατανάλωσης ηλεκτρικής ενέργειας έχοντας ως στόχο να εξασφαλίσουμε την καλύτερη δυνατή ακρίβεια πρόβλεψης σε όλα τα επίπεδα αυτών λαμβάνοντας ταυτόχρονα υπόψη ότι οι επιμέρους προβλέψεις των επιπέδων θα πρέπει να είναι «συμφιλιωμένες». Με την έννοια «συμφιλιωμένες» εννοούμε, ότι για να έχουν οι ιεραρχικές προβλέψεις λογική υπόσταση, απαιτείται οι προβλέψεις των από κάτω επιπέδων να συναθροίζονται σε αυτές των ανωτέρων. Για να εξακριβώσουμε την βέλτιστη προσέγγιση ανά ορίζοντα πρόβλεψης και ανά επίπεδο συνάθροισης, χρησιμοποιούμε δύο διαφορετικές οικογένειες μεθόδων, αυτά των αιτιοκρατικών μοντέλων πρόβλεψης και των μεθόδων πρόβλεψης χρονοσειρών. Επιπλέον εξετάζονται τρεις διαφορετικές προσεγγίσεις ιεραρχικής συμφιλίωσης προβλέψεων καθώς και μέθοδοι χρονικής συνάθροισης χρονοσειρών για την επίτευξη του καλύτερου δυνατού αποτελέσματος.

Αρχικά, πραγματοποιείται διεξοδική βιβλιογραφική επισκόπηση των διαφόρων μεθόδων πρόβλεψης ενεργειακής ζήτησης. Στη συνέχεια, παρουσιάζονται αναλυτικά τα μοντέλα πρόβλεψης χρονοσειρών με χρονική συνάθροιση καθώς και τα μοντέλα πολλαπλής γραμμικής παλινδρόμησης που θα χρησιμοποιηθούν στη συγκεκριμένη εργασία. Στο πλαίσιο των μοντέλων παλινδρόμησης, πραγματοποιείται αναλυτική περιγραφή των μεθόδων επιλογής παλινδρομητών, ενώ αναλύεται η έννοια της συσχέτισης καθώς και του δείκτη συσχέτισης R^2 . Έπειτα ορίζεται η έννοια της χρονικής συνάθροισης χρονοσειρών, και παρουσιάζεται ο αλγόριθμος της πολλαπλής χρονικής συνάθροισης MAPA. Ακόμα, παρουσιάζονται οι τρεις διαφορετικές ιεραρχικές προσεγγίσεις Bottom-Up, Top-Down και Optimal. Εν συνεχεία, αναπτύσσονται τα βασικά χαρακτηριστικά των χρονοσειρών και παρατίθενται αναλυτικοί στατιστικοί δείκτες καθώς και στατιστικοί δείκτες ακρίβειας προβλέψεων.

Στο τελευταίο κομμάτι της εργασίας, γίνεται μελέτη εφαρμογής των μεθόδων πρόβλεψης χρονοσειρών με χρονική συνάθροιση και των παλινδρομικών μοντέλων σε ενεργειακά δεδομένα. Η υλοποίηση των μεθόδων πρόβλεψης γίνεται με το λογισμικό R, ενώ τα ιεραρχικά δομημένα δεδομένα που χρησιμοποιήθηκαν είναι τα δεδομένα του διαγωνισμού GEFC 2012 και αναφέρονται σε ωριαίες καταναλώσεις ενέργειας 21 διαφορετικών περιοχών της Αμερικής. Με βάση τα αποτελέσματα, προκύπτουν συγκριτικά συμπεράσματα σχετικά με τις μεθόδους πρόβλεψης, τον ορίζοντα πρόβλεψης, τις ιεραρχικές προσεγγίσεις και τα επίπεδα της ιεραρχίας.

Περιεχόμενα

| | |
|--|----------|
| Πρόλογος | iv |
| Περίληψη | iv |
| Περιεχόμενα..... | iv |
| i | |
| Κατάλογος Σχημάτων | x |
| Κατάλογος Πινάκων | xi |
| | |
| Κεφάλαιο 1 : Εισαγωγή..... | 1 |
| 1.1 Αντικείμενο της εργασίας | 1 |
| 1.2 Δομή της εργασίας | 2 |
| Κεφάλαιο 2 : Σημασία Ενεργειακής Κατανάλωσης | 5 |
| 2.1 Σημασία της ενεργειακής κατανάλωσης σε παγκόσμιο επίπεδο | 5 |
| 2.2 Ενεργειακή απόδοση κτιρίων..... | 7 |
| Κεφάλαιο 3 : Μέθοδοι Πρόβλεψης Ενεργειακής Ζήτησης | 9 |
| 3.1 Γενικά για τις προβλέψεις | 9 |

| | | |
|---|---|-----------|
| 3.2 | Κατηγορίες Μεθόδων Πρόβλεψης Ενεργειακής Ζήτησης | 9 |
| 3.2.1 | Μηχανικές Μέθοδοι..... | 9 |
| 3.2.2 | Στατιστικές Μέθοδοι..... | 12 |
| 3.2.3 | Νευρωνικά Δίκτυα..... | 13 |
| 3.2.4 | Μηχανές Διανυσμάτων Υποστήριξης..... | 18 |
| 3.2.5 | Γκρι μοντέλα | 19 |
| Κεφάλαιο 4 : Μοντέλα Πρόβλεψης | | 21 |
| 4.1 | Κατηγορίες Μοντέλων Πρόβλεψης..... | 21 |
| 4.1.1 | Κριτικές Προβλέψεις | 21 |
| 4.1.2 | Στατιστικές Προβλέψεις | 21 |
| 4.2 | Μοντέλα Εκθετικής εξομάλυνσης | 22 |
| 4.2.2 | Τύποι Μοντέλων Εξομάλυνσης | 23 |
| 4.2.3 | Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing) | 24 |
| 4.2.4 | Μοντέλο εξομάλυνσης γραμμικής τάσης (Holt Exponential Smoothing)..... | 26 |
| 4.2.5 | Μοντέλο εξομάλυνσης μη γραμμικής τάσης | 28 |
| 4.2.6 | Εποχιακή εξομάλυνση..... | 29 |
| 4.3 | Μοντέλα Γραμμικής Παλινδρόμησης | 32 |
| 4.3.1 | Απλή Γραμμική Παλινδρόμηση | 33 |
| 4.3.2 | Ο Συντελεστής γραμμικής συσχέτισης..... | 34 |
| 4.3.3 | Ο συντελεστής R^2 (coefficient of determination) | 34 |
| 4.3.4 | Η εξίσωση παλινδρόμησης ως μοντέλο | 35 |
| 4.3.5 | Πολλαπλή Γραμμική Παλινδρόμηση..... | 36 |
| 4.3.6 | Διαδικασία επιλογής ανεξάρτητων μεταβλητών..... | 37 |
| 4.3.7 | Επιλογή βέλτιστου μοντέλου | 38 |
| 4.3.8 | Υπολογισμός συντελεστών παλινδρόμησης | 41 |
| 4.3.9 | Πολλαπλή συσχέτιση και ο συντελεστής R^2 | 41 |
| 4.3.10 | Ο στατιστικός δείκτης F (F-test) | 42 |
| 4.3.11 | Οι στατιστικοί δείκτες t (t-test) για τους συντελεστές..... | 42 |
| 4.3.12 | Έλεγχος των υπόλοιπων σφαλμάτων (Residual Errors) | 43 |
| 4.3.13 | Βασικές υποθέσεις στην πολλαπλή παλινδρόμηση και πολυσυγγραμικότητα..... | 444 |
| 4.3.14 | Εφαρμογή της πολλαπλής παλινδρόμησης στην πράξη..... | 45 |
| Κεφάλαιο 5 : Συνάθροιση Προβλέψεων | | 49 |
| 5.1 | Μέθοδοι Συνάθροισης Προβλέψεων..... | 49 |

| | | |
|--|---|-----------|
| 5.2 | Χρονική Συνάθροιση Προβλέψεων Χρονοσειρών | 49 |
| 5.2.1 | Κλασική Χρονική Συνάθροιση (Temporal Aggregation)..... | 49 |
| 5.2.2 | Πολλαπλή Χρονική Συνάθροιση (Multiple Temporal Aggregation)..... | 50 |
| 5.2.3 | Πολλαπλή Χρονική Συνάθροιση Προβλέψεων Συνιστωσών της Χρονοσειράς (Multi Aggregation Prediction Algorithm MAPA)..... | 50 |
| 5.3 | Προβλέψεις Ιεραρχικών Χρονοσειρών | 61 |
| 5.3.1 | Γενικές Ιεραρχικές Προβλέψεις..... | 61 |
| 5.3.2 | Μέθοδοι ιεραρχικής συνάθροισης | 63 |
| 5.3.3 | Bottom-up προβλέψεις..... | 64 |
| 5.3.4 | Top-down προβλέψεις | 65 |
| 5.3.5 | Middle-out Προβλέψεις | 66 |
| 5.3.6 | Optimal combination forecasts for hierarchical time series | 66 |
| Κεφάλαιο 6 : Στατιστική ανάλυση χρονοσειρών και δείκτες σφαλμάτων | | 73 |
| 6.1 | Χρονοσειρές (Timeseries)..... | 73 |
| 6.1.1 | Ποιοτικά χαρακτηριστικά των χρονοσειρών..... | 73 |
| 6.2 | Βασική στατιστική ανάλυση..... | 75 |
| 6.2.1 | Μέση τιμή (Average) | 75 |
| 6.2.2 | Μέγιστη και ελάχιστη τιμή (Maximum και Minimum) | 75 |
| 6.2.3 | Τυπική απόκλιση (Standard Deviation)..... | 75 |
| 6.2.4 | Διακύμανση (Variance) | 75 |
| 6.2.5 | Συνδιακύμανση (Covariance) | 75 |
| 6.2.6 | Συντελεστής γραμμικής συσχέτισης (Linear Correlation Coefficient) | 76 |
| 6.2.7 | Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient) | 76 |
| 6.2.8 | Συντελεστής μεταβλητότητας (Coefficient of Variation) | 76 |
| 6.2.9 | Μέση τιμή διαστήματος μεταξύ ζητήσεων (Intermittent Demand Interval).. | 77 |
| 6.3 | Στατιστική ανάλυση ακρίβειας προβλέψεων | 77 |
| 6.3.1 | Μέσο σφάλμα (Mean Error)..... | 78 |
| 6.3.2 | Μέσο απόλυτο σφάλμα (Mean Absolute Error) | 78 |
| 6.3.3 | Μέσο τετραγωνικό σφάλμα (Mean Squared Error)..... | 78 |
| 6.3.4 | Ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Squared Error) | 78 |
| 6.3.5 | Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error) | 79 |
| 6.3.6 | Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error) | 79 |
| 6.3.7 | Relative Measures | 80 |

| | | |
|--|--|------------|
| 6.3.8 | Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error)... | 80 |
| 6.3.9 | Theil's U-Statistic | 81 |
| 6.3.10 | Percentage Better..... | 81 |
| 6.3.11 | Ρυθμός ανάπτυξης | 82 |
| Κεφάλαιο 7 : Ενεργειακές προβλέψεις μέσω μοντέλου πρόβλεψης χρονοσειρών με χρονική συνάθροιση | | 83 |
| 7.1 | Μεθοδολογία πρόβλεψης χρονοσειρών με χρονική συνάθροιση | 83 |
| 7.1.1 | Ιδέα της μεθόδου | 83 |
| 7.1.2 | Εφαρμογή μεθόδου πρόβλεψης χρονοσειρών με χρονική συνάθροιση | 83 |
| 7.1.3 | Παραγωγή Ιεραρχικών Προβλέψεων..... | 86 |
| 7.1.4 | Υλοποίηση στην R..... | 88 |
| 7.2 | Case Study : Παραγωγή Προβλέψεων μέσω μοντέλου πρόβλεψης χρονοσειρών με χρονική συνάθροιση | 90 |
| 7.2.1 | Επισκόπηση Προβλήματος και Συλλογή Δεδομένων..... | 90 |
| 7.2.2 | Εφαρμογή μεθόδου μοντέλου πρόβλεψης χρονοσειρών με χρονική συνάθροιση | 91 |
| 7.2.3 | Παραγωγή Ιεραρχικών Προβλέψεων..... | 91 |
| 7.2.4 | Επικύρωση και Αξιολόγηση Προβλέψεων | 94 |
| 7.2.5 | Αποτελέσματα | 95 |
| Κεφάλαιο : 8 Ενεργειακές προβλέψεις μέσω μοντέλου πολλαπλής παλινδρόμησης .. | | 99 |
| 8.1 | Μεθοδολογία κατασκευής μοντέλου πολλαπλής γραμμικής παλινδρόμησης..... | 99 |
| 8.1.1 | Ιδέα της μεθόδου | 99 |
| 8.1.2 | Κατασκευή μοντέλων παλινδρόμησης για όλα τα στοιχεία της ιεραρχίας | 99 |
| 8.1.3 | Παραγωγή Προβλέψεων | 101 |
| 8.1.4 | Υλοποίηση στην R..... | 102 |
| 8.2 | Case Study : Παραγωγή Προβλέψεων μέσω μοντέλου πολλαπλής γραμμικής παλινδρόμησης | 103 |
| 8.2.1 | Επισκόπηση Προβλήματος και Συλλογή Δεδομένων..... | 103 |
| 8.2.2 | Εφαρμογή της Μεθόδου Πολλαπλής Γραμμικής Παλινδρόμησης | 104 |
| 8.2.3 | Παραγωγή Ιεραρχικών Προβλέψεων..... | 107 |
| 8.2.4 | Επικύρωση και Αξιολόγηση Προβλέψεων | 109 |
| 8.2.5 | Αποτελέσματα | 109 |
| Κεφάλαιο 9 : Συμπεράσματα | | 115 |
| Κεφάλαιο 10 : Βιβλιογραφία | | 119 |
| 10.1 | Βιβλία-Σημειώσεις..... | 119 |

| | | |
|------------------------|------------------------------------|------------|
| 10.2 | Ερευνητικές δημοσιεύσεις..... | 119 |
| ΠΑΡΑΡΤΗΜΑ | | 122 |
| (I) | Σφάλματα s-MAPE μεθόδου MLR..... | 125 |
| (II) | Σφάλματα s-MAPE μεθόδου MAPA | 122 |

Κατάλογος Σχημάτων

| | |
|-------------|---|
| Σχήμα 2.1 : | Διάγραμμα ετήσια ενεργειακής κατανάλωσης, εκπομπών CO ₂ και πληθυσμού σε περίοδο δύο δεκαετιών |
| Σχήμα 2.2 : | Κατανάλωση ανά χρήση για διάφορους τύπους κτιρίων (Πηγή: EIA)..... |
| Σχήμα 4.1 : | Βασικές κατηγορίες μοντέλων εξομάλυνσηςως συνδυασμοί 4 μοντέλων τάσης και 3 εποχιακών μοντέλων (Πηγή : Gardner, 1987)..... |
| Σχήμα 4.2 : | Διάγραμμα ροής μεθόδου «Forward Selection»..... |
| Σχήμα 4.3 : | Διάγραμμα ροής μεθόδου «Backward Selection»..... |
| Σχήμα 5.1 : | Μηνιαία χρονοσειρά για διαφορετικά επίπεδα χρονικής συνάθροισης χωρίς επικάλυψη (Πηγή: N. Kourentzes, F. Petropoulos and J. R. Trapero, 2014, improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting)..... |
| Σχήμα 5.2: | Διάγραμμα ροής του προτεινόμενου αλγόριθμου πρόβλεψης πολλαπλών συναθροίσεων (MAPA) (Πηγή:International Institute of Forecaster Published by Elsevier B.V.)..... |

| | |
|--|--|
| Σχήμα 5.3 : Διαγράμματα χρονοσειρών και φάσματα ισχύος για $k=1,3,5,12$ (Πηγή: <i>International Journal of Forecasting</i> 30 (2014)/ N. Kourentzes et al.)..... | |
| Σχήμα 5.4 : Πρόβλεψη συνιστωσών στην περίπτωση αθροιστικής διαμόρφωσης..... | |
| Σχήμα 5.5 : Εκτιμώμενη συνιστώσα τάσης για $k = (1, 4, 8)$ και στην αρχική χρονική κλίμακα (Πηγή : <i>International Journal of Forecasting</i> 30 (2014)/ N. Kourentzes et al.)..... | |
| Σχήμα 5.6 : Συνιστώσες επιπέδου και τάσης για $k = (1, 4, 8)$ και οι προκύπτουσες συνδυασμένες συνιστώσες (Πηγή : <i>International Journal of Forecasting</i> (2014)/ N. Kourentzes et al.)..... | |
| Σχήμα 5.7 : Παράδειγμα Ιεραρχίας..... | |
| Σχήμα 5.8 : Σχηματική απεικόνιση ιεραρχίας και των προσεγγίσεων Top-Down, Bottom-Up & Middle-Out..... | |
| Σχήμα 7.1 : Διάγραμμα ροής μεθόδου πρόβλεψης χρονοσειρών με χρονική συνάθροιση..... | |
| Σχήμα 7.2 : Διάγραμμα Παραγωγής Προβλέψεων για τους τρεις χρονικούς ορίζοντες και για τις 3 ιεραρχικές προσεγγίσεις..... | |
| Σχήμα 7.3 : Ιεραρχία εξεταζόμενου προβλήματος..... | |
| Σχήμα 8.1 : Διάγραμμα ροής μεθόδου «Forward Selection»..... | |
| Σχήμα 8.2 : Διάγραμμα Παραγωγής Προβλέψεων για τους τρεις χρονικούς ορίζοντες και για τις 3 ιεραρχικές προσεγγίσεις..... | |
| Σχήμα 8.3 : Ιεραρχία εξεταζόμενου προβλήματος..... | |

Κατάλογος Πινάκων

| | |
|--|--|
| Πίνακας 6.1 : Παρατηρήσεις χρονοσειράς προς στατιστική ανάλυση..... | |
| Πίνακας 6.2 : Παρατηρήσεις και προβλέψεις χρονοσειράς προς στατιστική ανάλυση ακρίβειας προβλέψεων..... | |
| Πίνακας 6.3 : Παρατηρήσεις και διαφορετικές προβλέψεις χρονοσειράς προς στατιστική ανάλυση ακρίβειας προβλέψεων..... | |
| Πίνακας 7.1 : Σφάλματα s-MAPE πάνω επιπέδου ιεραρχίας μεθόδου MAPA για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 7.2 : Σφάλματα s-MAPE κάτω επιπέδου ιεραρχίας μεθόδου MAPA για τους τρεις ορίζοντες πρόβλεψης..... | |

| | |
|--|--|
| Πίνακας 7.3 : Μέσα σφάλματα s-MAPE ιεραρχίας μεθόδου MAPA για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 7.4 : Λόγος Σφαλμάτων sMAPE πάνω επιπέδου ιεραρχίας μεθόδων MAPA/SES για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 7.5 : Λόγος Σφαλμάτων sMAPE κάτω επιπέδου ιεραρχίας μεθόδων MAPA/SES για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 7.6 : Λόγος μέσων σφαλμάτων sMAPE ιεραρχίας μεθόδων MAPA/SES για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 8.1 : Παλινδρομητές που επιλέχθηκαν για τα μοντέλα..... | |
| Πίνακας 8.2 : Τιμή Δείκτη R^2 με την προσθήκη κάθε νέου παλινδρομητή στα μοντέλα παλινδρόμησης..... | |
| Πίνακας 8.3 : Σφάλματα s-MAPE πάνω επιπέδου ιεραρχίας μεθόδου MLR για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 8.4: Σφάλματα s-MAPE κάτω επιπέδου ιεραρχίας μεθόδου MLR για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 8.5 : Μέσα σφάλματα s-MAPE ιεραρχίας μεθόδου MLR για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 8.6 : Λόγος σφαλμάτων s-MAPE πάνω επιπέδου ιεραρχίας μεθόδων MLR/SES για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 8.7 : Λόγος σφαλμάτων s-MAPE κάτω επιπέδου ιεραρχίας μεθόδων MLR/SES για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας 8.8 : Λόγος μέσων σφαλμάτων s-MAPE πάνω ιεραρχίας μεθόδων MLR/SES για τους τρεις ορίζοντες πρόβλεψης..... | |
| Πίνακας (i) : Σφάλματα sMAPE μεθόδου MLR για οριζόντια πρόβλεψης : 24 h..... | |
| Πίνακας (ii) : Σφάλματα sMAPE μεθόδου MLR για οριζόντια πρόβλεψης : 72h..... | |
| Πίνακας (iii) : Σφάλματα sMAPE μεθόδου MLR για οριζόντια πρόβλεψης : 168 h..... | |
| Πίνακας (iv) : Σφάλματα sMAPE μεθόδου MAPA για οριζόντια πρόβλεψης : 24 h..... | |
| Πίνακας (v) : Σφάλματα sMAPE μεθόδου MAPA για οριζόντια πρόβλεψης : 72..... | |
| Πίνακας (vi) : Σφάλματα sMAPE μεθόδου MAPA για οριζόντια πρόβλεψης : 168 h..... | |

Κεφάλαιο 1 : Εισαγωγή

1.1 Αντικείμενο της εργασίας

Η πρόβλεψη της ενεργειακής κατανάλωσης είναι ιδιαίτερα σημαντική, καθώς ο άνθρωπος πάντα καλούταν να λαμβάνει αποφάσεις για το μέλλον, τόσο υποσυνείδητα όσο και στο πλαίσιο του οργανωμένου προγραμματισμού. Αυτό συμβαίνει λόγω της αβεβαιότητας που σχετίζεται με την εξέλιξη των αντικειμένων και την πορεία τους στη πάροδο του χρόνου. Σημειώνεται ότι, όσο πιο πολύπλοκα είναι τα αντικείμενα, τόσο αυξάνει η τυχαιότητα και είναι δυσκολότερο να πραγματοποιηθούν προβλέψεις χωρίς τη βοήθεια συστηματικών μεθόδων και τεχνικών πρόβλεψης. Η εκτίμηση του μέλλοντος σε έναν ικανοποιητικό βαθμό, γίνεται ακόμα πιο επιτακτική ανάγκη σε εφαρμογές που ενέχουν οικονομικές και περιβαλλοντικές επιπτώσεις, όπως στην περίπτωση της κατανάλωσης ενέργειας.

Αντικείμενο της παρούσας εργασίας είναι να ερευνήσουμε την καταλληλότητα μοντέλων πρόβλεψης σε ιεραρχίες χρονοσειρών κατανάλωσης ηλεκτρικής ενέργειας έχοντας ως στόχο να εξασφαλίσουμε την καλύτερη δυνατή ακρίβεια πρόβλεψης σε όλα τα επίπεδα αυτών λαμβάνοντας ταυτόχρονα υπόψη ότι οι επιμέρους προβλέψεις των επιπέδων θα πρέπει να είναι «συμφιλωμένες», με την έννοια ότι για να έχουν λογική υπόσταση απαιτείται οι προβλέψεις των από κάτω επιπέδων να συναθροίζονται σε αυτές των ανωτέρων. Για να εξακριβώσουμε την βέλτιστη προσέγγιση ανά ορίζοντα πρόβλεψης και επίπεδο συνάθροιση, χρησιμοποιούμε δύο διαφορετικές οικογένειες μεθόδων, αυτά των αιτιοκρατικών μοντέλων πρόβλεψης και των μεθόδων πρόβλεψης χρονοσειρών. Επιπλέον εξετάζονται τρεις διαφορετικές προσεγγίσεις ιεραρχικής συμφιλίωσης προβλέψεων καθώς και μέθοδοι χρονικής συνάθροισης χρονοσειρών για την επίτευξη του καλύτερου δυνατού αποτελέσματος.

Αναφορικά με την οικογένεια των αιτιοκρατικών μοντέλων, αναπτύσσουμε μοντέλα πολλαπλής γραμμικής παλινδρόμησης, σύμφωνα με τα οποία συσχετίζουμε την εξαρτημένη μεταβλητή της ενεργειακής ζήτησης με ανεξάρτητες μεταβλητές. Χρησιμοποιώντας ιστορικά στοιχεία, «καταρτίζουμε» αυτά τα μοντέλα πρόβλεψης, ούτως ώστε να είναι κατάλληλα για παραγωγή προβλέψεων. Κάθε στοιχείο της ιεραρχίας, στο οποίο θέλουμε να παράγουμε προβλέψεις, διαθέτει διαφορετικά χαρακτηριστικά κι επηρεάζεται με διαφορετικό τρόπο από τις ανεξάρτητες παραμέτρους (π.χ. θερμοκρασία), γεγονός που μας οδηγεί στη μοντελοποίηση της ενεργειακής ζήτησης με βάση τις συγκεκριμένες παραμέτρους για κάθε στοιχείο της ιεραρχίας ξεχωριστά. Κατασκευάζοντας τα μοντέλα παλινδρόμησης είμαστε σε θέση να παράγουμε μελλοντικές εκτιμήσεις για διάφορους χρονικούς ορίζοντες.

Αναφορικά με τις μεθόδους πρόβλεψης χρονοσειρών, εφαρμόζουμε μεθόδους εκθετικής εξομάλυνσης χρονοσειρών για την παραγωγή προβλέψεων. Επιπλέον, εξετάζουμε την επίδραση της χρονικής συνάθροισης σε μεθόδους εκθετικής εξομάλυνσης για την παραγωγή προβλέψεων μεγαλύτερης ακρίβειας. Μέσω αυτής της μεθόδου, στοχεύουμε στη διάκριση των διαφορετικών συνιστωσών των χρονοσειρών (π.χ. τάση, επίπεδο, εποχιακότητα) και στην παραγωγή προβλέψεων μέσω του συνδυασμού προβλέψεων των συνιστωσών αυτών για διάφορους χρονικούς ορίζοντες.

Έχοντας στη διάθεση μας τις δύο προαναφερθείσες μεθόδους πρόβλεψης, παράγουμε ιεραρχικές προβλέψεις με βάση τρεις διαφορετικές ιεραρχικές προσεγγίσεις που θα παρουσιαστούν αναλυτικά στα επόμενα κεφάλαια. Εξετάζουμε την συμπεριφορά αυτών μεταξύ τους και συγκριτικά ανάμεσα στις δύο μεθόδους πρόβλεψης ανά οριζόντια πρόβλεψη. Με βάση συγκεκριμένους στατιστικούς δείκτες ελέγχουμε την ακρίβεια και αξιοπιστία των προβλέψεων κι εξάγουμε χρήσιμα συμπεράσματα.

1.2 Δομή της εργασίας

Στο δεύτερο κεφάλαιο της παρούσας εργασίας, πραγματοποιείται μια επισκόπηση της σημασίας της ενεργειακής ζήτησης και συμπεριφοράς κτιρίων σε παγκόσμιο επίπεδο. Αναδεικνύεται η ανάγκη για την κατανόηση της ενεργειακής συμπεριφοράς των κτιρίων λόγω της άρρηκτης σύνδεσης της με οικονομικές υποχρεώσεις και περιβαλλοντικές επιπτώσεις. Επίσης παρουσιάζονται τα οφέλη που προκύπτουν από την ανάλυση της ενεργειακής συμπεριφοράς και την παραγωγή προβλέψεων σχετικά με μεγέθη ενεργειακής ζήτησης.

Στο τρίτο κεφάλαιο, πραγματοποιείται μια εκτενής αναφορά στις κατηγορίες μεθόδων ενεργειακής πρόβλεψης. Αναλύονται οι βασικές κατηγορίες μεθόδων πρόβλεψης και παρέχεται, με βάση την βιβλιογραφία, πληθώρα παραδειγμάτων εφαρμογής των μεθόδων αυτών σε ένα μεγάλο εύρος προβλημάτων.

Το τέταρτο κεφάλαιο αναφέρεται ολοκληρωτικά στα μοντέλα πρόβλεψης που θα χρησιμοποιηθούν στη συγκεκριμένη εργασία. Πιο συγκεκριμένα το τέταρτο κεφάλαιο ξεκινά με μια διάκριση των προβλέψεων σε στατιστικές και κριτικές. Στη συνέχεια τα στατιστικά μοντέλα πρόβλεψης διακρίνονται σε αιτιοκρατικά μοντέλα επί μιας σειράς δεδομένων και σε στατιστικά μοντέλα πρόβλεψης χρονοσειρών.

Στα πλαίσια των μοντέλων πρόβλεψης χρονοσειρών αναλύονται οι εκθετικές μέθοδοι εξομάλυνσης σταθερού επιπέδου, γραμμικής τάσης, εκθετικής και φθίνουσας τάσης και παρουσιάζεται ο συνδυασμός αυτών με τα τρία μοντέλα εποχιακότητας : μηδενική, αθροιστική και πολλαπλασιαστική εποχιακότητα. Παρουσιάζονται αναλυτικά οι διαφοροποιήσεις και τα χαρακτηριστικά των μεθόδων αυτών, ενώ παράλληλα αναδεικνύεται η σημασία χρήσης κατάλληλου μοντέλου εποχιακότητας για την παραγωγή αξιόπιστων προβλέψεων.

Στα πλαίσια των αιτιοκρατικών μοντέλων, αναλύονται μοντέλα παλινδρόμησης επί μιας σειράς δεδομένων. Αρχικά παρουσιάζεται η έννοια της γραμμικής παλινδρόμησης, η οποία γενικεύεται, εν συνεχεία, στην έννοια της πολλαπλής γραμμικής παλινδρόμησης. Σύμφωνα με την τελευταία, ελέγχεται η συσχέτιση του προβλεπόμενου μεγέθους με ανεξάρτητες παραμέτρους, για τις οποίες διαθέτουμε μια σειρά ιστορικών δεδομένων. Η συσχέτιση ελέγχεται με βάση τον δείκτη R^2 , ο οποίος θα παρουσιαστεί αναλυτικά. Η επιλογή του βέλτιστου μοντέλου επιτυγχάνεται μέσω του δείκτη BIC. Μέσω της μεθόδου της πολλαπλής γραμμικής παλινδρόμησης στοχεύουμε στην εύρεση και μοντελοποίηση κάποιου συγκεκριμένου πρότυπου συμπεριφοράς. Ο όρος μοντελοποίηση αναφέρεται στον υπολογισμό των παραμέτρων παλινδρόμησης του μοντέλου χρησιμοποιώντας τα ιστορικά δεδομένα (διαδικασία “training” του μοντέλου). Παρουσιάζεται ο τρόπος υπολογισμού κι

ελέγχου αυτών (μέσω στατιστικών δεικτών), μέσω των οποίων θα καταστεί το μοντέλο κατάλληλο για την παραγωγή προβλέψεων για διαφορους ορίζοντες πρόβλεψης. Τέλος παρουσιάζονται κάποιες προϋποθέσεις σχετικά με την πολλαπλή παλινδρόμηση, οι οποίες θα πρέπει να πληρούνται προκειμένου να ξεπεραστούν κάποιοι κίνδυνοι (πιθανή πολυσυγγραμικότητα) και να παραχθούν αξιόπιστες προβλέψεις.

Το πέμπτο κεφάλαιο αφορά την συνάθροιση των προβλέψεων. Γίνεται διάκριση ανάμεσα στην χρονική συνάθροιση, η οποία ανφέρεται στην παραγωγή πολλαπλών χρονοσειρών εναλλακτικών συχνοτήτων από μια χρονοσειρά και στην ιεραρχική συνάθροιση, η οποία ανφέρεται στην οργάνωση των χρονοσειρών σε μια ιεραρχική δομή βάσει των διακριτών χαρακτηριστικών τους.

Στα πλαίσια της χρονικής συνάθροισης, παρουσιάζεται, αρχικά, η χρονική συνάθροιση, ενώ στη συνέχεια μελετάται διεξοδικά η πολλαπλή χρονική συνάθροιση προβλέψεων συνιστωσών της χρονοσειράς (MAPA) στα πλαίσια της πολλαπλής χρονικής συνάθροισης. Η συγκεκριμένη μέθοδος θα εφαρμοσθεί για την παραγωγή προβλέψεων μέσω χρονικής συνάθροισης συνιστωσών της χρονοσειράς (Multi Aggregation Prediction Algorithm MAPA) και συνίσταται στα βήματα της συνάθροισης, της πρόβλεψης, και του συνδυασμού που περιγράφονται αναλυτικά. Παρουσιάζονται διάφορα πλεονεκτήματα, απόρροιες της συγκεκριμένης προσέγγισης με βασικότερα όλων το κέρδος στην ακρίβεια πρόβλεψης, καθώς και το κέρδος του συνδυασμού των προβλέψεων.

Στα πλαίσια της ιεραρχικής συνάθροισης, παρουσιάζονται τρεις διαφορετικές προσεγγίσεις ιεραρχικών προβλέψεων, που θα εφαρμοσθούν στην παρούσα εργασία. Αυτές οι προσεγγίσεις είναι οι Bottom-Up, Top-Down και Optimal. Παρουσιάζονται τα χαρακτηριστικά των προσεγγίσεων συγκριτικά και παρέχεται πλήθος παραδειγμάτων εφαρμογής αυτών στην βιβλιογραφία. Επιπρόσθετα, εξετάζονται πιθανές υπολογιστικές παγίδες και ο τρόπος αντιμετώπισης αυτών.

Το έκτο κεφάλαιο ασχολείται αποκλειστικά με χρονοσειρές και σφάλματα. Αρχικά διακρίνονται τα ποιοτικά χαρακτηριστικά των χρονοσειρών και ακολουθούν η βασική στατιστική ανάλυση και η στατιστική ανάλυση ακρίβειας των προβλέψεων. Στο κεφάλαιο αυτό, περιέχονται όλα τα σφάλματα και οι στατιστικοί δείκτες που θα χρησιμοποιηθούν στη συνέχεια για την εκτίμηση και την αξιολόγηση της ακρίβειας των προβλέψεων.

Το έβδομο και το όγδοο κεφάλαιο αποτελούν Case-Studies των μεθόδων της εργασίας. Παρουσιάζουν τις μεθοδολογίες που εφαρμόστηκαν πάνω σε αληθινά δεδομένα ενεργειακής ζήτησης από 20 διαφορετικές περιοχές της Αμερικής, τα οποία συλλέχθηκαν από τον παγκόσμιο διαγωνισμό ενεργειακής κατανάλωσης που έλαβε χώρα στο Vancouver του Καναδά το 2012. Στόχος των 2 κεφαλαίων είναι η ανάδειξη της αποτελεσματικότητας των μεθόδων πρόβλεψης με βάση τις τρεις διαφορετικές ιεραρχικές προσεγγίσεις που αναφέρθηκαν και η αξιολόγησή αυτών υπό ρεαλιστικές συνθήκες, παρατηρώντας, δηλαδή, τις αποκλίσεις των προβλέψεων από τα πραγματικά δεδομένα.

Στο τελευταίο κεφάλαιο της εργασίας, εξάγονται τα κύρια συμπεράσματα με βάση τα αποτελέσματα των πειραμάτων και προτείνονται θέματα για μελέτη και ανάλυση στο μέλλον.

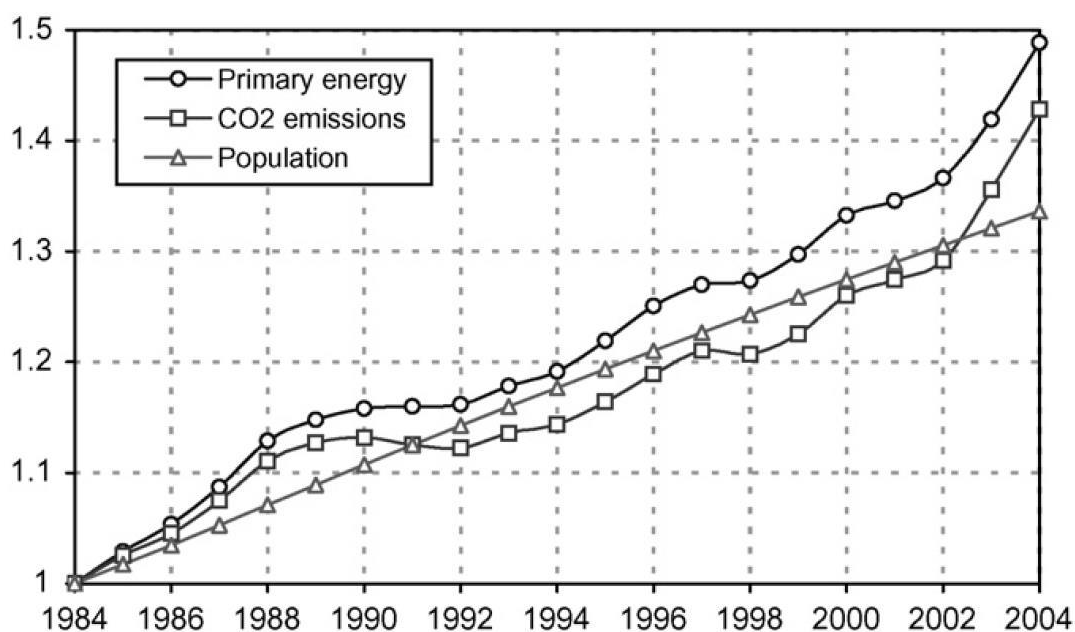
Κεφάλαιο 2 : Σημασία Ενεργειακής Κατανάλωσης

2.1 Σημασία της ενεργειακής κατανάλωσης σε παγκόσμιο επίπεδο

Τα τελευταία χρόνια, παρατηρείται συνεχώς αυξανόμενη χρήση ηλεκτρικής ενέργειας σε παγκόσμιο επίπεδο. Η αύξηση του πληθυσμού σε συνδυασμό με το γεγονός ότι οι άνθρωποι δραστηριοποιούνται ολοένα και περισσότερο εντός κτιρίων προκειμένου να παράγουν προϊόντα ή υπηρεσίες και να ικανοποιούν ανάγκες τους αποτελεί το βασικό αίτιο αυτής της αυξανόμενης ζήτησης και χρήσης ηλεκτρικής ενέργειας. Παράλληλα, ο σύγχρονος τρόπος ζωής, κυρίως στις αναπτυγμένες χώρες, όπου οι άνθρωποι επιδιώκουν να κατασκευάζουν και να ζουν σε κτίρια με ιδιαίτερες τεχνολογικές παροχές και ανέσεις συντελεί σ'αυτήν την ανοδική τάση της χρήσης ηλεκτρικής ενέργειας.

Ωστόσο, η συνεχώς αυξανόμενη χρήση ηλεκτρικής ενέργειας δημιουργεί προβληματισμό στην παγκόσμια κοινότητα αναφορικά με θέματα τροφοδότησης, εξάντλησης ενεργειακών αποθεμάτων και περιβαλλοντικών συνεπειών, όπως το φαινόμενο θερμοκηπίου, η τρύπα του όζοντος και οι ευρύτερες κλιματικές αλλαγές.

Όπως παρουσιάζεται στο σχήμα 1.1, κατά την περίοδο δύο δεκαετιών, από το 1984 ως το 2004, παρατηρήθηκε αύξηση της ενεργειακής ζήτησης κατά 49% και αύξηση των εκπομπών διοξειδίου του άνθρακα κατά 43%. Υπολογίζεται μάλιστα ότι τα δύο αυτά μεγέθη αυξάνονται ετησίως με ένα μέσο ρυθμό 2% και 1,8% αντίστοιχα.

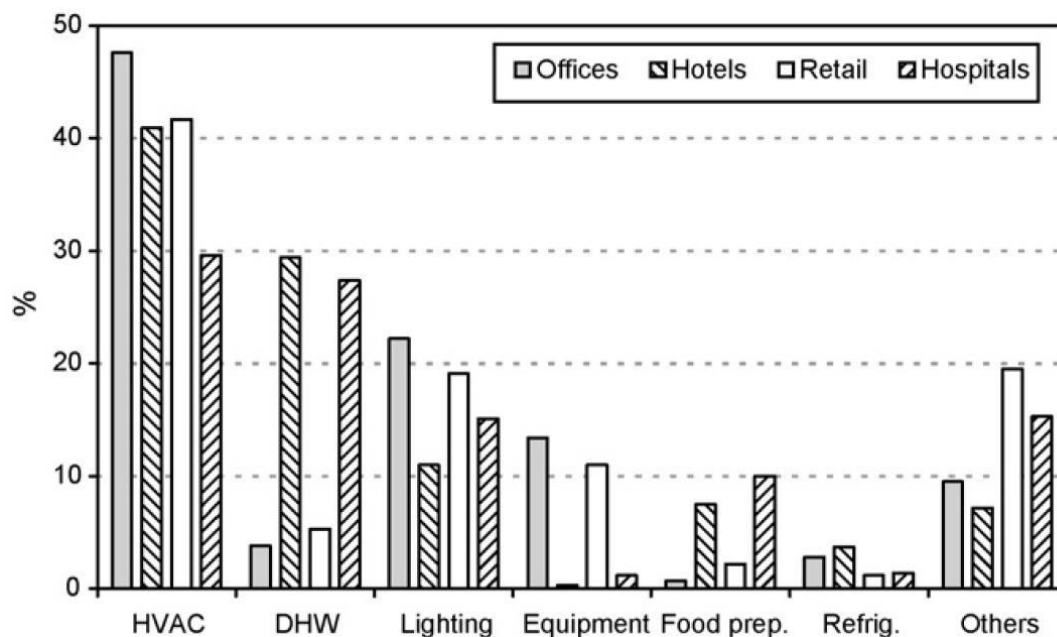


Σχήμα 2.1 : Διάγραμμα ετήσια ενεργειακής κατανάλωσης, εκπομπών CO2 και πληθυσμού σε περίοδο δύο δεκαετιών

Η παγκόσμια συνεισφορά των εμπορικών κτιρίων και των κατοικιών στη συνολική χρήση ενέργειας αυξάνεται με σταθερό ρυθμό. Σήμερα, αυτή η συνεισφορά υπολογίζεται σε ποσοστά μεταξύ του 20% και του 30% της συνολικής ενεργειακής χρήσης στις αναπτυγμένες χώρες ενώ έχει ήδη ξεπεράσει άλλους νευραλγικούς τομείς όπως η βιομηχανία και οι μεταφορές. Στην Ευρώπη, η κατανάλωση ενέργειας από τα κτίρια

αποτελεί περίπου το 40 % της συνολικής κατανάλωσης ενέργειας και «ευθύνεται» για το 36% της συνολικής εκπομπής διοξειδίου του άνθρακα.

Τα συστήματα θέρμανσης και κλιματισμού (HVAC) είναι τα πλέον απαιτητικά σε επίπεδο ενεργειακής κατανάλωσης και είναι υπεύθυνα για το 10-20% της τελικής ενεργειακής χρήσης στις αναπτυγμένες χώρες. Στις Η.Π.Α. ,η κατανάλωση ενέργειας από τα συστήματα θέρμανσης και κλιματισμού (HVAC) συνιστά το 50% της κτιριακής και το 20% της συνολικής κατανάλωσης ενέργειας.



Σχήμα 2.2 :Κατανάλωση ανά χρήση για διάφορους τύπους κτιρίων. (Πηγή: EIA)

Αποτελεί, συνεπώς, πρόβλημα κομβικής σημασίας η πρόβλεψη της ενεργειακής κατανάλωσης, προκειμένου να επιτευχθεί βελτίωση της ενεργειακής απόδοσης, η οποία θα οδηγήσει τόσο σε εξοικονόμηση ενέργειας, όσο και σε μείωση των περιβαλλοντικών επιπτώσεων.

2.2 Ενεργειακή απόδοση κτιρίων

Το σύστημα ενέργειας των κτιρίων είναι πολύπλοκο καθώς ποικίλουν οι τύποι των ενεργειακών φορτίων καθώς και των κτιρίων. Στην βιβλιογραφία οι κύριες κατηγορίες ενεργειακών φορτίων που λαμβάνονται υπόψη είναι τα φορτία θέρμανσης/ψύξης, θέρμανσης νερού και κατανάλωσης ηλεκτρικού ρεύματος. Αντίστοιχα, οι κύριες κατηγορίες των κτιρίων που λαμβάνονται υπόψη είναι γραφεία, οικιακά κτίρια και μηχανικά κτίρια που ποικίλουν σε έκταση από ένα δωμάτιο έως ένα μεγάλο συγκρότημα.

Η ενεργειακή συμπεριφορά ενός κτιρίου επηρεάζεται από πολλούς παράγοντες. Σημαντικότεροι εξ' αυτών είναι οι καιρικές συνθήκες και οι συνθήκες περιβάλλοντος (βαρύνουσα σημασία έχουν τα ξηρά κλίματα και οι υψηλές θερμοκρασίες), η δομή και οι διαστάσεις των κτιρίων, τα ενεργειακά φορτία (όπως ο φωτισμός, τα συστήματα θέρμανσης, εξαερισμού και ψύξης), η συμπεριφορά αυτών των φορτίων καθώς και το πρόγραμμα λειτουργίας τους. Άλλοι παράγοντες που επηρεάζουν την ενεργειακή συμπεριφορά ενός κτιρίου σχετίζονται με την κατασκευή των κτιρίων και τις φυσικές ιδιότητες των υλικών που χρησιμοποιήθηκαν κατά την κατασκευή τους καθώς και τη συμπεριφορά αυτών των υλικών στην διάρκεια του χρόνου και υπο διάφορες συνθήκες.

Η πολυπλοκότητα που παρουσιάζει το πρόβλημα πρόβλεψης ενεργειακής ζήτησης καθιστά την ακριβή ενεργειακή πρόβλεψη πολύ δύσκολη. Τα τελευταία χρόνια έχει προταθεί μεγάλος αριθμός λεπτομερών καθώς και απλοποιημένων μεθόδων επεξεργασίας δεδομένων και πρόβλεψης και έχει εφαρμοστεί σε μεγάλο εύρος προβλημάτων. Αυτές οι μέθοδοι είναι έργο έρευνας που διεξάγεται τόσο κατά τον σχεδιασμό νέων κτιρίων, όσο και κατά την μετασκευή υπαρχόντων κτιρίων. Η έρευνα αυτή εκτείνεται από την ανάλυση του τύπου των κτιρίων έως τη μοντελοποίηση σε τοπικό και εθνικό επίπεδο. Αξίζει να σημειωθεί, ότι με βάση τις παραπάνω μεθόδους πρόβλεψης, μπορούν να γίνουν προβλέψεις στο σύνολο ενός κτιρίου ή σε ένα τμήμα αυτού με την λεπτομερή ανάλυση κάθε παράγοντα επιρροής ή με τον υπολογισμό της κατανάλωσης λαμβάνοντας υπόψη διάφορους άλλους σημαντικούς παράγοντες.

Έχουν αναπτυχθεί συγκεκριμένα μοντέλα πρόβλεψης, τα οποία περιλαμβάνουν μεθόδους επεξεργασίας, απλοποιημένες μεθόδους, μηχανικές μεθόδους, στατιστικές μεθόδους και μεθόδους τεχνητής νοημοσύνης.

Κεφάλαιο 3 : Μέθοδοι Πρόβλεψης Ενεργειακής Ζήτησης

3.1 Γενικά για τις προβλέψεις

Βασικό αντικείμενο της επιστήμης των προβλέψεων αποτελεί η συστηματοποίηση της παραγωγής και η ερμηνεία τους. Με τον όρο πρόβλεψη αναφερόμαστε στη διαδικασία αξιοποίησης της διαθέσιμης γνώσης για την παραγωγή ισχυρισμών για γεγονότα, των οποίων τα πραγματικά αποτελέσματα δεν έχουν ακόμα παρατηρηθεί. Η πρόκληση που προβάλλει ο στοχαστικός χαρακτήρας του προβλήματος αλλά και η ανάγκη του ανθρώπου για διορατικότητα στο βάθος του χρόνου καθιστούν τον τομέα των προβλέψεων απαραίτητο.

Δεν υπάρχει αμφιβολία ότι το ενδιαφέρον και η σημασία της πρόβλεψης έχει γνωρίσει ραγδαία άυξηση τα τελευταία τριάντα χρόνια. Το ενδιαφέρον αυτό για την πρόβλεψη προκύπτει τόσο από την ακαδημαϊκή έρευνα όσο και από την επιχειρησιακή έρευνα όπου οι πρακτικά ασχολούμενοι με τον τομέα της πρόβλεψης καλούνται να λάβουν κάποια απόφαση αντιμετωπίζοντας την αβεβαιότητα του μέλλοντος. Η αντίληψη αυτής της κατάστασης αβεβαιότητας γίνεται όλο και πιο έντονη και έχει επιβάλει μια πιο συστηματική και προσεκτική έρευνα του μέλλοντος. Αυτή την χρονική στιγμή, η μεγαλύτερη πρόκληση στον τομέα της πρόβλεψης είναι να γίνει η διαδικασία των προβλέψεων όσο το δυνατόν πιο χρήσιμη και αποδοτική.

Οι Goosijer και Hyndman (2005) έκαναν μια επισκόπηση αναφορικά με την έρευνα που έχει διεξαχθεί στο πεδίο των προβλέψεων τα τελευταία τριάντα χρόνια και υποστηρίζουν πως παρά το μεγάλο βαθμό προόδου, υπάρχουν αρκετές επιστημονικές περιοχές που χρήζουν περαιτέρω έρευνας και ανάπτυξης.

3.2 Κατηγορίες Μεθόδων Πρόβλεψης Ενεργειακής Ζήτησης

3.2.1 Μηχανικές Μέθοδοι

Πρόκειται για μεθόδους που έχουν αναπτυχθεί επαρκώς τα τελευταία πενήντα χρόνια. Οι μηχανικές μέθοδοι χρησιμοποιούν φυσικές αρχές για να υπολογίσουν θερμικά δυναμικά και ενεργειακή συμπεριφορά είτε στο επίπεδο ενός κτιρίου συνολικά είτε για κάποια υποσύνολα του κτιρίου. Αυτές οι μέθοδοι μπορούν σε γενικές γραμμές να ταξινομηθούν σε δύο κατηγορίες : τις λεπτομερείς ολοκληρωμένες μεθόδους και τις απλοποιημένες μεθόδους.

3.2.1.1 Ολοκληρωμένες Μηχανικές Μέθοδοι

Οι ολοκληρωμένες μέθοδοι χρησιμοποιούν φυσικές συναρτήσεις επεξεργασίας ή θερμικά δυναμικά για να υπολογίσουν βήμα-βήμα την ενεργειακή κατανάλωση όλων των στοιχείων των κτιρίων με εισόδους πληροφορίες για το κτίριο και το πειβάλλον του όπως οι εξωτερικές κλιματικές συνθήκες, η κατασκευή του κτιρίου, οι διαστάσεις του, η λειτουργία του, το πρόγραμμα ρυθμού χρησιμοποίησης και ο εξοπλισμός θέρμανσης / εξαερισμού / ψύξης. Συγκεκριμένα για τα φορτία θέρμανσης, εξαερισμού και ψύξης, η ISO έχει αναπτύξει μια συγκεκριμένη μέθοδο για τον υπολογισμό της ενεργειακής κατανάλωσης για την θέρμανση, τον εξαερισμό και την ψύξη ενός κτιρίου ή ενός χώρου αυτού.

Επίσης, έχουν αναπτυχθεί εκατοντάδες εργαλεία λογισμικού για την αξιολόγηση της ενεργειακής απόδοσης, των ανανεώσιμων πηγών ενέργειας και την βιωσιμότητα αυτών.

Τέτοια προγράμματα είναι τα εξής : DOE-2,Energy Plus,BLAST,ESP-r. Παρά το γεγονός, ότι αυτά τα προσομοιωτικά προγράμματα επεξεργασίας είναι αποτελεσματικά και ακριβή, στην πράξη υπάρχουν κάποιες δυσκολίες. Εξ' αιτίας του γεγονότος ότι αυτά τα προγράμματα βασίζονται σε φυσικές αρχές για την επίτευξη μιας ακριβούς προσομοίωσης, απαιτούνται ως είσοδοι πληροφορίες σχετικά με παραμέτρους που αφορούν στο κτίριο ή στις καιρικές συνθήκες. Από τη μία πλευρά, αυτές οι πληροφορίες δεν είναι διαθέσιμες τουλάχιστον προς το παρόν,δηλαδή δεν μπορεί να συλλεχθεί μία τέτοια πληροφορία για κάθε χώρο ενός μεγάλου κτιρίου. Αυτή η έλλειψη σε ακριβείς εισόδους θα οδηγήσει σε χαμηλής ακρίβειας προσομοίωση. Από την άλλη πλευρά, η χρήση αυτών των μεθόδων απαιτεί κουραστική και εξειδικευμένη εργασία και είναι οικονομικά μη αποδοτική. Για τον λόγο αυτό, κάποιοι ερευνητές έχουν προτείνει απλούστερα μοντέλα για να παρέχουν εναλλακτική λύση σε συγκεκριμένες εφαρμογές.

3.2.1.2 Απλοποιημένες Μηχανικές Μέθοδοι

Ο Al-Homoud ερεύνησε δύο απλοποιημένες μεθόδους. Η μία μέθοδος είναι η «degree day», όπου μοναδική παράμετρος που αναλύεται είναι ο βαθμός της ημέρας. Αυτή η μέθοδος σταθερής κατάστασης είναι κατάλληλη για εκτίμηση ενεργειακής κατανάλωσης μικρών κτιρίων,όπου κυριαρχεί ο παράγοντας της ενέργειας που αφορά στις βασικές οικιακές καταναλώσεις. Η άλλη μέθοδος είναι η «bin», γνωστή επίσης και ως μέθοδος συχνότητας θερμοκρασίας, η οποία μπορεί να χρησιμοποιηθεί για την μοντελοποίηση μεγάλων κτιρίων όπου τα εσωτερικά παραγόμενα φορτία κυριαρχούν ή τα φορτία δεν εξαρτώνται γραμμικά από την διαφορά εσωτερικής/εξωτερικής θερμοκρασίας.

Οι καιρικές συνθήκες συνιστούν σημαντικό παράγοντα για την εκτίμηση της ενεργειακής κατανάλωσης και αναλύονται σε ταχύτητα ανέμου, θερμοκρασία, ηλιακή ακτινοβολία, υγρασία και ποικίλουν ανάλογα την εποχή. Έχουν εκπονηθεί συγκεκριμένες έρευνες με στόχο την απλοποίηση του παράγοντα του καιρού στους υπολογισμούς ενεργειακών καταναλώσεων κτιρίων.

Οι White και Reichmuth , το 1996, χρησιμοποίησαν τις μέσες μηνιαίες θερμοκρασίες για να προβλέψουν τις μηνιαίες ενεργειακές καταναλώσεις κτιρίων. Αυτή η πρόβλεψη ήταν πιο ακριβής από τις καθιερωμένες διαδικασίες οι οποίες συνήθως χρησιμοποιούν λίστα θερμοκρασιών. Οι Westphal και Lamberts, το 2004, παρήγαγαν πρόβλεψη για το ετήσιο φορτίο θέρμανσης και ψύξης ενός μη οικιστικού κτιρίου βασιζόμενοι σε κάποιες καιρικές παραμέτρους συμπεριλαμβανομένων : του μηνιαίου μέσου όρου, των μεγίστων κι ελαχίστων θερμοκρασιών, την ατμοσφαιρική πίεση, την σκίαση από τα σύννεφα και τη σχετική υγρασία. Τα αποτελέσματα τους έδειξαν ικανοποιητική ακρίβεια σε χαμηλής μαζικότητας κτίρια σε σύγκριση με εργαλεία προσομοίωσης (BLAST,ESP,DOE-2).

Οι Yao και Steemers, το 2005, ανέπτυξαν μια απλοποιημένη μέθοδο πρόβλεψης της ημερήσιας κατανάλωσης ενέργειας με στόχο τον σχεδιασμό ενός συστήματος ανανεώσιμων μορφών ενέργειας για οικιστικά κτίρια. Η συνολική κατανάλωση ορίζεται ως το άθροισμα καταναλώσεων πολλών στοιχείων όπως : κατανάλωση ενέργειας για ζεστό νερό, θέρμανση, ψύξη ,οικιακές συσκευές κλπ. Για κάθε στοιχείο κατανάλωσης ενέργειας εφαρμόστηκε μια συγκεκριμένη μέθοδος μοντελοποίησης. Για παράδειγμα, για την μοντελοποίηση των ηλεκτρικών συσκευών, χρησιμοποιήθηκαν οι μέσες τιμές κατανάλωσης από μια μεγάλη

λίστα στατιστικών δεδομένων. Για τη μοντελοποίηση της θέρμανσης / ψύξης χρησιμοποιήθηκε ένα απλοποιημένο φυσικό μοντέλο. Η μέθοδος αυτή παράγει πρόβλεψη ενεργειακής κατανάλωσης για μια εποχή κάθε φορά καθώς οι μέσες τιμές, τις οποίες λαμβάνει ως εισόδους, ποικίλουν ανάλογα την εποχή. Υιοθετώντας αυτή την ιδέα της διαχώρισης και συνάθροισης, ο Rice απλοποίησε τον υπολογισμό κατανάλωσης κάθε υποσυνόλου του κτιρίου για να εξηγήσει το σύστημα κατανάλωσης ενέργειας του κτιρίου συνολικά. Στην έρευνα τους, το 2010, με τίτλο «Updating the ASHRAE/ACCa Residential Heating and Cooling Load Calculation Procedures and Data» οι Barnaby και Spittler πρότειναν μία απλοποιημένη και απτή μέθοδο με παράγοντα το οικιακό φορτίο. Οι καταναλώσεις των διαφόρων φορτίων υπολογίστηκαν και αξιολογήθηκαν χωριστά και στη συνέχεια αθροίστηκαν.

Οι Wang και Xu, το 2006, απλοποίησαν τα φυσικά χαρακτηριστικά των κτιρίων για να εφαρμόσουν προβλέψεις. Για το περιβάλλον των κτιρίων, οι παράμετροι του μοντέλου καθορίστηκαν με την χρήση εύκολα διαθέσιμων φυσικών πληροφοριών σχετικά με την ανάλυση συχνοτικών χαρακτηριστικών. Για τους διαφορους εσωτερικούς χώρους χρησιμοποίησαν ένα θερμικό δίκτυο από συγκεντρωμένη θερμική μάζα για να αναπαραστήσουν την εσωτερική συγκέντρωση.

Το 2001 ο Yik χρησιμοποίησε λεπτομερή εργαλεία προσομοίωσης για να εξάγει τα φορτία ψύξης για διαφορετικούς τύπους κτιρίων. Ένα απλοποιημένο μοντέλο, συνδυασμός των προσομοιωτικών αποτελεσμάτων προτάθηκε για τον καθορισμό του άμεσου φορτίου ψύξης ενός κτιρίου.

Ο Al-Homoud υποστήριξε ότι, αν στόχος είναι η μελέτη τάσεων, η σύγκριση συστημάτων ή η σύγκριση και ανάλυση εναλλακτικών συστημάτων, τότε οι απλοποιημένες μέθοδοι είναι επαρκείς. Αντίθετα, για μια λεπτομερή ενεργειακή ανάλυση κτιρίου και υποσυνόλων κτιρίων καθώς και για αναλύσεις κόστους στον κύκλο ζωής, τα ολοκληρωμένα εργαλεία κρινονται πιο κατάλληλα.

3.2.1.3 Η σημασία της βαθμονόμησης

Η βαθμονόμηση είναι ένα σημαντικό θέμα αναφορικά με την προσομοίωση κατανάλωσης ενέργειας σε κτίρια. Ρυθμίζοντας τις εισόδους προσεκτικά, η ενεργειακή συμπεριφορά ενός κτιρίου μπορεί να ταιριάζει απόλυτα με την συμπεριφορά ενός κτιρίου στην πραγματικότητα. Ο Pan, το 2007, συνόψισε τη βαθμονομημένη προσομοίωση ως μέθοδο ενεργειακής ανάλυσης ενός κτιρίου και την αφήρησε για να αναλύσει την κατανάλωση ενέργειας ενός μεγάλου εμπορικού κτιρίου. Μετά απο επαρκή βήματα επαναλαμβανόμενης βαθμονόμησης, αυτό το μοντέλο ενέργειας έδειξε υψηλή ακρίβεια στην πρόβλεψη της πραγματικής κατανάλωσης ενέργειας συγκεκριμένων κτιρίων. Ωστόσο, η βαθμονόμηση είναι μια διαδικασία απαιτητική σε χρόνο και απαιτεί εξειδικευμένη εργασία αποδεικνύοντας ότι η παραγωγή πρόβλεψης με ένα λεπτομερές μηχανικό μοντέλο αποτελεί εγχείρημα υψηλής πολυπλοκότητας.

3.2.2 Στατιστικές Μέθοδοι

Τα στατιστικά μοντέλα παλινδρόμησης απλά συσχετίζουν την κατανάλωση ενέργειας ή άλλους ενεργειακούς δείκτες με τις μεταβλητές που αυτοί επηρεάζουν. Αυτά τα εμπειρικά μοντέλα αναπτύσσονται με βάση ιστορικά δεδομένα απόδοσης, γεγονός που σημαίνει, ότι πριν την χρήση αυτών μοντέλων πρέπει πρώτα να συλλέξουμε ιστορικά δεδομένα. Μεγάλο τμήμα της έρευνας σχετικά με τα μοντέλα παλινδρόμησης έχει διεξαχθεί κατά τα τρία ακόλουθα προβλήματα. Το πρώτο είναι να προβλεφθεί η κατανάλωση ενέργειας σύμφωνα με μία ή περισσότερες καιρικές παραμέτρους. Το δεύτερο είναι να προβλεφεί κάποιος χρήσιμος δείκτης ενέργειας. Το τρίτο είναι να εκτιμηθούν κάποιες σημαντικές παράμετροι σχετικές με την κατανάλωση ενέργειας, όπως ο δείκτης απώλειας συνολικής ενέργειας, η συνολική θερμοχωρητικότητα και ο παράγοντας κέρδους, που είναι χρήσιμες στην ανάλυση θερμικής συμπεριφοράς των κτιρίων ή τμημάτων αυτών.

Σε ορισμένα απλοποιημένα μηχανικά μοντέλα, η παλινδρόμηση χρησιμοποιείται για να συσχετιστεί η κατανάλωση ενέργειας με τις κλιματικές μεταβλητές.

Οι Bauer και Scartezini, (1998), πρότειναν μια μέθοδο παλινδρόμησης για να χειριστούν ταυτόχρονα τους υπολογισμούς θέρμανσης και ψύξης, χρησιμοποιώντας ως μεταβλητές τα εσωτερικά και ηλιακά κέρδη. Το 2005, ο Ansari υπολόγισε το φορτίο ψύξης ενός κτιρίου ως άθροισμα των φορτίων ψύξης του κάθε υποσυνόλου του κτιρίου ξεχωριστά. Κάθε φορτίο ψύξης υποσυνόλου του κτιρίου είναι μια απλή συνάρτηση παλινδρόμησης, με παλινδρομητή την διαφορά θερμοκρασίας μεταξύ του εσωτερικού και του εξωτερικού χώρου του υποσυνόλου του κτιρίου.

Ο Dhar, το 1998, μοντελοποίησε το φορτίο ψύξης και θέρμανσης σε εμπορικά κτίρια με μοναδική μεταβλητή καιρού την εξωτερική θερμοκρασία.

Ένα νέο μοντέλο βασισμένο στις σειρές Fourier θερμοκρασίας προτάθηκε για να εκφράσει τη μη γραμμική εξάρτηση των φορτίων ψύξης και θέρμανσης από την θερμοκρασία και τον χρόνο. Αν επίσης είναι διαθέσιμα τα στοιχεία υγρασίας και τα ηλιακά δεδομένα, προτάθηκε η χρήση του μοντέλου γενικευμένης σειράς Fourier, λόγω της μηχανικής σχετικότητας και της μεγαλύτερης δυνατότητας πρόβλεψης. Λαμβάνοντας ως μοναδική μεταβλητή για την ανάπτυξη μοντέλου τη θερμοκρασία οι Lei και Hu, το 2009, αξιολόγησαν μοντέλα παλινδρόμησης για την πρόβλεψη εξοικονομούμενης ενέργειας από τη μετασκευή κτιρίων με γραφεία σε περιοχές καυτού καλοκαιριού και και κρύου χειμώνα. Έδειξαν ότι ένα γραμμικό μοντέλο μιας μεταβλητής είναι επαρκές και πρακτικό για τη μοντελοποίηση της κατανάλωσης ενέργειας σε συνθήκες ζεστού καλοκαιριού και κρύου χειμώνα.

Ο Μα, το 2010, ολοκλήρωσε πολλές μεθόδους γραμμικής παλινδρόμησης και αυτοπαλινδρόμησης για να προβλέψει μηνιαία κατανάλωση ενέργειας για μεγάλα κτίρια δημόσιας κλίμακας. Ο Cho, στην έρευνα του (2004), χρησιμοποίησε μοντέλο παλινδρόμησης με ημερήσιες, εβδομαδιαίες και τριμηνιαίες παρατηρήσεις παρατηρώντας σφάλμα πρόβλεψης της ετήσιας κατανάλωσης ενέργειας 100%, 30 % και 6% αντίστοιχα. Τα αποτελέσματα της έρευνας έδειξαν ότι το μέγεθος της περιόδου μέτρησης επηρεάζει σε σημαντικό βαθμό τα μοντέλα παλινδρόμησης που εξαρτώνται από την θερμοκρασία.

Όσον αφορά στην πρόβλεψη του δείκτη ενέργειας, ο Lam, το 2010, χρησιμοποιώντας την ανάλυση βασικής συνιστώσας, ανέπτυξε έναν κλιματικό δείκτη Z αναφορικά με την ηλιακή ακτινοβολία και την θερμοκρασία υγρού και ξηρού βολβού. Έδειξε ότι ο δείκτης Z έχει την ίδια τάση με το προσομοιωμένο φορτίο ψύξης / θέρμανσης / εξαερισμού και την κατανάλωση ενέργειας του κτιρίου. Η τάση αυτή ελήφθη από την ανάλυση του συσχετισμού της γραμμικής παλινδρόμησης. Το μοντέλο αναπτύχθηκε με βάση στοιχεία που συλλέχθηκαν από το 1979 ως το 2007

Ο Ghiaus ,το 2006, ανέπτυξε ένα μοντέλο ισχυρής παλινδρόμησης για τη συσχέτιση της απώλειας ενέργειας θέρμανσης και της θερμοκρασίας ξηρού βολβού χρησιμοποιώντας το εύρος από το 1^ο ως το 3^ο τεταρτημόριο, που δίνει τη σχέση των 2 μεταβλητών.

Οι Himenez και Heras χρησιμοποίησαν, το 2005, το αυτο-παλινδρομικό μοντέλο με εισόδους για να υπολογίσουν τους δείκτες U και Z των στοιχείων ενός κτιρίου.

Ο Kimbara ανέπτυξε ένα αυτο-παλινδρομικό μοντέλο ολοκληρωμένου κινητού μέσου όρου για να εφαρμόσει on-line πρόβλεψη. Το μοντέλο βασίζεται στα πρόσφατα στοιχεία φορτίου και χρησιμοποιούνται για να προβλέψει τα φορτία της επόμενης ημέρας.

Μοντέλα ARIMA με εξωτερικές εισόδους έχουν επίσης εφαρμοστεί σε ορισμένες εφαρμογές με στόχο την πρόβλεψη και τον έλεγχο της ζήτησης της ηλεκτρικής ενέργειας αιχμής εμπορικών κτιρίων, καθώς επίσης και την πρόβλεψη ζήτησης ενέργειας διαφόρων κτιρίων. Οι Newsham και Birt έδωσαν ιδιαίτερη έμφαση στην επίδραση της πληρότητας, η οποία μπορεί προφανώς να αυξήσει την ακρίβεια του μοντέλου.

Οι Aydinalp-Koksal και Ugursal, το 2008, πρότειναν την εξέταση μιας μεθόδου παλινδρόμησης (Conditional Demand Analysis) για την πρόβλεψη της κατανάλωσης ενέργειας κτιρίων σε εθνικό επίπεδο. Σε πειραματικές συγκρίσεις τους, τα μοντέλα CDA έδειξαν ικανότητα ακριβούς πρόβλεψης, αντίστοιχη με τα νευρωνικά δίκτυα και τις μηχανικές μεθόδους, έχοντας, παράλληλα το προτέρημα της εύκολης ανάπτυξης και χρήσης.

Ωστόσο, το μειονέκτημα του μοντέλου CDA είναι η έλλειψη λεπτομέρειας και ευελιξίας και το γεγονός ότι απαιτεί μια μεγάλη ποσότητα πληροφοριών εισόδου. Το μοντέλο CDA χρησιμοποιήθηκε στις πρώτες αναλύσεις οικιακής κατανάλωσης ενέργειας.

3.2.3 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα τεχνητής νοημοσύνης είναι τα πιο ευρέως χρησιμοποιούμενα μοντέλα τεχνητής νοημοσύνης σε εφαρμογές πρόβλεψης κατανάλωσης ενέργειας κτιρίων. Αυτός ο τύπος μοντέλου είναι κατάλληλος για την επίλυση μη γραμμικών προβλημάτων και αποτελεί μια αποδοτική προσέγγιση στην πολύπλοκη αυτή εφαρμογή. Κατά τα τελευταία είκοσι χρόνια, οι ερευνητές έχουν εφαρμόσει τη μέθοδο των δικτύων τεχνητής νοημοσύνης, προκειμένου να αναλύσουν την συνολική κατανάλωση ενέργειας σε διάφορους τύπους κτιρίων υπό διάφορες συνθήκες, το φορτίο ψύξης / θέρμανσης, την κατανάλωση ηλεκτρικής ενέργειας, τη λειτουργία και τη βελτιστοποίηση λειτουργίας υπο-στοιχείων του κτιρίου και την εκτίμηση παραμέτρων χρήσης.

Το 2006, ο Καλογήρου έκανε μια ανασκόπηση των νευρωνικών δικτύων τεχνητής νοημοσύνης σε εφαρμογές ενέργειας σε κτιρια όπως τα ηλιακά συστήματα θέρμανσης νερού, η ηλιακή ακτινοβολία, η ταχύτητα του ανέμου, η κατανομή της ροής του αέρα μέσα σε ένα δωμάτιο, η θερμοκρασία του εσωτερικού αέρα, η πρόβλεψη κατανάλωσης ηλεκτρικής ενέργειας και η ανάλυση συστημάτων θέρμανσης, ψύξης και εξαερισμού. Ο Καλογήρου χρησιμοποίησε νευρωνικά δίκτυα διάδοσης προς τα πίσω, προκειμένου να προβλέψει το απαιτούμενο φορτίο θέρμανσης των κτιρίων. Το μοντέλο εξετάστηκε σε δεδομένα κατανάλωσης ενέργειας απο 225 κτίρια τα οποία ποικίλαν απο μικρά κτίρια έως κτίρια μεγάλης έκτασης. Οι Ekici και Aksoy χρησιμοποίησαν το ίδιο μοντέλο για να προβλέψουν τα φορτία θέρμανσης κτιρίων σε τρία κτίρια .Ο έλεγχος και οι δοκιμές των συνόλων δεδομένων υπολογίστηκαν με την χρήση της προσέγγισης των πεπερασμένων διαφορών της παροδικής κατάστασης της μονοδιάστατης θερμικής αγωγιμότητας.

Ο Olofsson προέβλεψε την ετήσια ζήτηση ενέργειας θέρμανσης μικρών κτιρίων οικογενειών στα βόρεια της Σουηδίας. Αργότερα, οι Olofsson και Andersen ανέπτυξαν ένα νευρωνικό δίκτυο, το οποίο παράγει μακροχρόνιες προβλέψεις ζήτησης ενέργειας (ετήσια ζήτηση ενέργειας θέρμανσης) βασιζόμενο σε βραχυπρόθεσμα δεδομένα μετρήσεων (συνήθων 2-5 εβδομάδων) με υψηλό ρυθμό πρόβλεψης για κτίρια στα οποία διαμένει μία οικογένεια.

Το 2009, ο Yokohama χρησιμοποίησε ένα νευρωνικό δίκτυο πίσω διάδοσης για να προβλεψει την ζήτηση ενέργειας για την ψύξη ενός κτιρίου. Στο έργο προτάθηκε μια παγκόσμια μέθοδος βελτιστοποίησης, που ονομάζεται «Modal Trimming Method» για τον προσδιορισμό των παραμέτρων του μοντέλου. Ο Kreider κατέγραψε τα αποτελέσματα ενός επαναλαμβανόμενου νευρωνικού δικτύου βασισμένο σε δεδομένα ωριαίας κατανάλωσης ενέργειας για να προβλεψει τις ανάγκες ενέργειας θέρμανσης και ψύξης στο μέλλον, γνωρίζοντας μόνο τον καιρό και την ζώνη ώρας. Με βάση το ίδιο επαναλαμβανόμενο νευρωνικό δίκτυο, οι Ben-Nakhi και Mahmoud προέβλεψαν το φορτίο ψύξης τριών κτιρίων με γραφεία. Τα δεδομένα φορτίων ψύξης απο το 1997 ως το 2000 χρησιμοποιήθηκαν για την εξάσκηση του μοντέλου και τα δεδομένα για το 2001 χρησιμοποιήθηκαν για τον έλεγχο του μοντέλου.

Ο Καλογήρου χρησιμοποίησε νευρωνικά δίκτυα για την πρόβλεψη της ενεργειακής κατανάλωσης παθητικών ηλιακών κατοικιών, όπου δεν χρησιμοποιούνται μηχανικές και ηλεκτρικές συσκευές θέρμανσης.

Λαμβάνοντας υπόψη την επίδραση του καιρού στην κατανάλωση ενέργειας σε διαφορετικές περιοχές, οι Yan και Yao χρησιμοποίησαν ένα νευρωνικό δίκτυο για να προβλέψουν το φορτίο θέρμανσης και ψύξης ενός κτιρίου σε διάφορες κλιματικές ζώνες που αντιπροσωπεύονται από τον βαθμό θέρμανσης ημέρας και τον βαθμό ψύξης ημέρας. Το νευρωνικό δίκτυο δοκιμάστηκε με αυτές τις δύο μετρήσεις ενέργειας ως τμήματα των μεταβλητών εισόδου.

Κατά την εφαρμογή πρόβλεψης χρήσης ηλεκτρικής ενέργειας, σε μια πρώιμη μελέτη χρησιμοποιήθηκαν με επιτυχία νευρωνικά δίκτυα για την πρόβλεψη της ωριαίας κατανάλωσης ηλεκτρικής ενέργειας καθώς και ενέργειας για ζεστό νερό σε ένα κτίριο με μηχανικές δραστηριότητες

Οι Nizami και Al-Garni δοκίμασαν ένα απλό νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης για να συσχετίσουν την ηλεκτρική κατανάλωση ενέργειας με τις καιρικές συνθήκες.

Οι González και Zamarrero προέβλεψαν βραχυπρόθεσμα ηλεκτρικά φορτία με ένα ειδικό νευρωνικό δίκτυο που ανατροφοδοτείται με τις εξόδους του. Αντιθέτως, ο Azadeh προέβλεψε τη μακροπρόθεσμη μέση ετήσια κατανάλωση ηλεκτρικής ενέργειας σε βιομηχανίες παραγωγής και έδειξε ότι το νευρωνικό δίκτυο είναι εφαρμόσιμο στο συγκεκριμένο και σε συναφή πρόβλημα, όπου η κατανάλωση ενέργειας παρουσιάζει μεγάλη διακύμανση. Ο Wong χρησιμοποίησε ένα νευρωνικό δίκτυο για να προβλέψει την κατανάλωση ενέργειας σε κτίρια γραφείων με έλεγχο του φωτισμού της ημέρας σε υποτροπικά κλίματα. Οι έξοδοι του μοντέλου περιλαμβάνουν καθημερινή χρήση ηλεκτρικής ενέργειας για ψύξη, θέρμανση, ηλεκτρικό φωτισμό και το συνολικό κτίριο. Τα Τ.Ν.Δ. χρησιμοποιούνται επίσης για την ανάλυση και βελτιστοποίηση της συμπεριφοράς των συνιστωσών του συστήματος ενέργειας, ως επί το πλείστον για τα συστήματα θέρμανσης, ψύξης, εξαερισμού.

Το 2006, ο Hou προέβλεψε το φορτίο κλιματισμού σε ένα κτίριο, το οποίο αποτελεί το κλειδί για τον βέλτιστο έλεγχο του συστήματος θέρμανσης, ψύξης και εξαερισμού. Ο Lee χρησιμοποίησε ένα γενικό νευρωνικό δίκτυο παλινδρόμησης για την ανίχνευση και τη διάγνωση σφαλμάτων στη μονάδα διαχείρισης του αέρα ενός κτιρίου.

Ο Aydinalp έδειξε ότι το νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί για την εκτίμηση της κατανάλωσης ενέργειας για τον φωτισμό και την ψύξη χώρων καθώς επίσης ότι αποτελεί ένα καλό μοντέλο για να εκτιμηθούν οι επιδράσεις των κοινωνικο-οικονομικών παραγόντων στην κατανάλωση αυτή στον οικιακό τομέα του Καναδά. Στην επέκταση του έργου, αναπτύχθηκαν μοντέλα νευρωνικών δικτύων για να εκτιμηθεί επιτυχώς η απαιτούμενη ενέργεια για θέρμανση χώρου και παροχή ζεστού νερού.

Ο Yalcintas χρησιμοποίησε νευρωνικά δίκτυα για την πρόβλεψη της χρήσης ενέργειας από ένα κτίριο σε ένα τροπικό κλίμα. Αργότερα, χρησιμοποίησε ένα νευρωνικό δίκτυο εμπροσθόδρομης τροφοδοσίας τριών στρωμάτων για να προβλέψει την εξοικονόμηση ενέργειας κατά την μετασκευή εξοπλισμού. Ο Gouda χρησιμοποίησε ένα νευρωνικό δίκτυο πολυεπίπεδης εμπρόσθιας τροφοδότησης για να προβλέψει την εσωτερική θερμοκρασία χρησιμοποιώντας εύκολα μετρήσιμες εισροές οι οποίες περιλαμβάνουν την εξωτερική θερμοκρασία, την ηλιακή ακτινοβολία, την θέση της βαλβίδας θέρμανσης και την εσωτερική θερμοκρασία του κτιρίου.

Οι παράμετροι ενεργειακής απόδοσης κτιρίων μπορούν να εκτιμηθούν με νευρωνικά δίκτυα. Οι ερευνητές υπολόγισαν τον συνολικό συντελεστή απώλειας θερμότητας, τη συνολική θερμοχωρητικότητα και τον συντελεστή κέρδους που είναι ιδιαίτερα σημαντικός για μια αξιόπιστη πρόβλεψη της ζήτησης ενέργειας. Η μέθοδος βασίζεται στην ανάλυση ενός μοντέλου νευρωνικού δικτύου το οποίο δοκιμάζεται σε απλά δεδομένα : εσωτερική / εξωτερική διαφορά θερμοκρασίας, παρεχόμενη θερμότητα και δωρεάν διαθέσιμη θερμότητα.

Ο Kreider ανέφερε τα αποτελέσματα των επαναλαμβανόμενων νευρωνικών δικτύων σε δεδομένα ωριαίας κατανάλωσης ενέργειας. Παρουσίασε, επίσης, τα αποτελέσματα σχετικά

με την εύρεση της θερμική αντίστασης και της θερμική χωρητικότητας C κτιρίων από δίκτυα που εκπαιδεύονται σε δεδομένα κτιρίων.

Ο Zmeureanu πρότεινε μια μέθοδο νευρωνικών δικτύων γενικής παλινδρόμησης για την αξιολόγηση του Συντελεστή Απόδοσης (COP) των υφιστάμενων μονάδων στον τελευταίο όροφο κτιρίων.

Ο Yalcintas παρουσίασε μια τεχνική συγκριτικής αξιολόγησης που βασίζοταν σε νευρωνικά δίκτυα για την ενέργεια κτιρίων σε τροπικά κλίματα, επικεντρώνοντας στην πρόβλεψη ενός σταθμισμένου δείκτη ενεργειακής χρήσης. Τα επιλεγμένα κτίρια βρίσκονταν σε μεγάλη ποικιλία.

Τα δεδομένα εισόδου για την εκπαίδευση του μοντέλου μπορούν να ληφθούν από επί-τόπου μέτρηση, έρευνα, συλλογή χρέωσης ή προσομοίωση. Τα πρώτα δεδομένα μπορεί να έχουν θόρυβο ή να περιλαμβάνουν άχρηστες μεταβλητές, ως εκ τούτου, μπορούν να απαλοιοφούν και να μειωθούν πριν από την ανάπτυξη του μοντέλου. Υπάρχει πολλή έρευνα όσον αφορά τις τεχνικές προ-επεξεργασίας δεδομένων.

Οι González και Zamarrero εκτίμησαν τα βραχυπρόθεσμα φορτία ηλεκτρικής ενέργειας με χρήση δύο φάσεων των νευρωνικών δικτύων. Το πρώτο στρώμα προβλέπει κλιματικές μεταβλητές, ενώ το δεύτερο προβλέπει τη χρήση της ενέργειας, το οποίο λαμβάνει ως εισόδους τις έξοδους του πρώτου στρώματος. Η ίδια τεχνολογία δύο φάσεων χρησιμοποιήθηκε επίσης από τον Yokoyama στην πρόβλεψη φορτίου ψύξης. Η τάση και οι περιοδικές αλλαγές απομακρύνθηκαν πρώτα από τα δεδομένα, και τα τροποποιημένα δεδομένα χρησιμοποιήθηκαν ως κύρια είσοδος για το μοντέλο «κατάρτισης». Πρόσθετες εισροές, συμπεριλαμβανομένης της θερμοκρασίας, του αέρα και της σχετικής υγρασίας, λήφθηκαν υπ' όψιν για να εξαχθούν προβλεπόμενες τιμές. Οι επιπτώσεις τους σχετικά με την πρόβλεψη της ζήτησης ενέργειας ερευνήθηκαν επίσης στη συγκεκριμένη εργασία.

Οι Ben-Nakhi και Mahmud προέβλεψαν την ενέργεια ψύξης κτιρίου, της επόμενης ημέρας, και το μοντέλο καταρτίστηκε με βάση τη μοναδική μεταβλητή της εξωτερικής θερμοκρασίας ξηρού βολβού. Οι Ekici και Aksoy προέβλεψαν τα θερμικά φορτία κτιρίων χωρίς να λαμβάνονται υπόψη οι κλιματικές μεταβλητές. Τα δίκτυα καταρτίστηκαν βασιζόμενα σε τρεις μόνο εισόδους: βαθμός διαφάνειας, προσανατολισμός κτιρίου και πάχος της μόνωσης.

Οι Kreider και Χάμπερλ προέβλεψαν το προσεχές μέλλον έχοντας ως είσοδο τα πιο πρόσφατα δεδομένα. Για μακροπρόθεσμες προβλέψεις, χρησιμοποίησαν επαναλαμβανόμενα νευρωνικά δίκτυα. Ο Γιανγκ χρησιμοποίησε αθροιστικές μεθόδους και μεθόδους «συρόμενου παράθυρου» για την κατάρτιση νευρωνικών δικτύων με σκοπό την on-line πρόβλεψη ενέργειας κτιρίου. Το «συρόμενο παράθυρο» περιορίζει τα δείγματα εισόδου σε ένα μικρό εύρος.

Ο Olofsson χρησιμοποίησε PCA για τη μείωση της διάστασης των μεταβλητών πριν από την πρόβλεψη της ετήσιας ζήτησης θέρμανσης. Σε μεταγενέστερη εργασία του, πέτυχε μακροπρόθεσμη πρόβλεψη της ζήτησης ενέργειας χρησιμοποιώντας βραχυπρόθεσμα μετρούμενα δεδομένα. Ο Kubota χρησιμοποίησε γενετικό αλγόριθμο για την εξαγωγή

μεταβλητών και διαλογή μετρήσεων, και, στη συνέχεια νευρωνικά δίκτυα αναπτύχθηκαν για την πρόβλεψη του ενεργειακού φορτίου κτιρίου. Εδώ η εξαγωγή μεταβλητών σημαίνει μετάφραση των πρωτότυπων πληροφοριών σε χρήσιμη πληροφορία που χρησιμοποιείται ως είσοδος στο νευρωνικό δίκτυο.

Ο Hui χρησιμοποίησε την θεωρία συνόλων και ένα νευρωνικό δίκτυο με στόχο την πρόβλεψη του φορτίου κλιματισμού. Εφαρμόστηκε η θεωρία συνόλου προκειμένου να βρεθούν σχετικοί παράγοντες που επηρεάζουν το φορτίο, οι οποίοι χρησιμοποιήθηκαν ως είσοδοι σε ένα νευρωνικό δίκτυο για την πρόβλεψη του φορτίου κλιματισμού.

Ο Kusiak πρόβλεψε το καθημερινό φορτίο ατμού κτιρίων από ένα νευρωνικό δίκτυο με σύνολο πέντε μεθόδους Perceptrons Multi-Layer (MLPs) καθώς σε αρκετές μελέτες περιπτώσεων, υπερτερεί έναντι 9 άλλων αλγορίθμων εξόρυξης δεδομένων, συμπεριλαμβανομένων της CART, CHAID, exhaustive CHAID, boosting tree, MARSplines, random forest, SVM, MLP και k-NN. Για την επιλογή μεταβλητών χρησιμοποιήθηκαν μία μήτρα συντελεστών συσχέτισης και ο αλγόριθμος «boosting tree». Ο Karataşu ερεύνησε πως οι στατιστικές μέθοδοι μπορούν να βελτιώσουν τα μοντέλα νευρωνικών δικτύων στην πρόβλεψη του ωριαίου ενεργειακού φορτίου. Οι στατιστικές μέθοδοι, όπως ο έλεγχος υποθέσεων, τα κριτήρια και η επαλήθευση πληροφοριών εφαρμόστηκαν τόσο στην επιλογή μοντέλου όσο και στην προ-επεξεργασία των εισόδων. Τα πειραματικά αποτελέσματα έδειξαν ότι η ακρίβεια της πρόβλεψης είναι συγκρίσιμη με τα καλύτερα αποτελέσματα που καταγράφονται στην βιβλιογραφία.

Οι έξοδοι των νευρωνικών δικτύων μπορεί να μην είναι οι προσδοκώμενες, οπότε ο Kajl πρότεινε μια λογική για εκ των υστέρων επεξεργασία των αποτελεσμάτων των νευρωνικών δικτύων. Σύμφωνα με την λογική αυτή δίνεται η δυνατότητα στον χρήστη να καθορίσει την επίδραση διαφόρων παραμέτρων του κτιρίου, σχετικά με την ετήσια και μηνιαία κατανάλωση ενέργειας. Στην βιβλιογραφία έχουν διεξαχθεί συγκρίσεις μεταξύ των νευρωνικών δικτύων και άλλων μοντέλων πρόβλεψης. Ο Azadeh έδειξε ότι τα νευρωνικά δίκτυα είναι εφαρμόσιμα στην πρόβλεψη της ετήσιας κατανάλωσης ενέργειας σε κατασκευαστικές βιομηχανίες, όπου η κατανάλωση ενέργειας παρουσιάζει μεγάλη διακύμανση. Σύμφωνα με την ανάλυση διακύμανσης τα νευρωνικά δίκτυα είναι ανώτερα, ως μέθοδος πρόβλεψης, του μοντέλου της συμβατικής μη γραμμικής παλινδρόμησης. Ο Aydinlar έδειξε ότι τα νευρωνικά δίκτυα μπορούν να έχουν υψηλότερη απόδοση πρόβλεψης από μηχανικά μοντέλα όσον αφορά την εκτίμηση κατανάλωσης ενέργειας λόγω εγκαταστάσεων, φωτισμού και ψύξης χώρου και των επιπτώσεων των κοινωνικο-οικονομικών παραμέτρων στην Καναδική οικιακή κατανάλωση ενέργειας.

Αργότερα πραγματοποιήθηκε σύγκριση των νευρωνικών δικτύων με την ανάλυση κριτηρίου λόγου και η έρευνα έδειξε ότι οι δύο μέθοδοι έχουν και οι δύο υψηλή απόδοση πρόβλεψης με την διαφορά ότι η δεύτερη μέθοδος παρουσιάζει ευκολία χρήσης και ανάπτυξης. Ο Neto συνέκρινε τις μηχανικές μεθόδους με τα νευρωνικά δίκτυα αναφορικά με την πρόβλεψη της κατανάλωσης ενέργειας κτιρίων και έδειξε ότι και οι 2 μέθοδοι έχουν υψηλή ακρίβεια πρόβλεψης ενώ υπερτερούν τα νευρωνικά δίκτυα στις βραχυπρόθεσμες προβλέψεις.

3.2.4 Μηχανές Διανυσμάτων Υποστήριξης

Χρησιμοποιούνται ολοένα και περισσότερο στην έρευνα και στην βιομηχανία, είναι ιδιαίτερα αποιτελεσματικά μοντέλα για την επίλυση μη γραμμικών προβλημάτων ακόμα και στην περίπτωση διάθεσης λίγων μόνο δεδομένων «εκπαίδευσης». Τα τελευταία πέντε χρόνια, έχουν διεξαχθεί πολλές έρευνες πάνω στην πρόβλεψη της κατανάλωσης ενέργειας κτιρίων με τα συγκεκριμένα μοντέλα. Ο Dong αρχικά εφάρμοσε τα συγκεκριμένα μοντέλα για να εξάγει μηνιαίες πρόβψεις κατανάλωσης ενέργειας τεσσάρων κτιρίων σε τροπικές περιοχές.

Το μοντέλο εκπαιδεύτηκε σε δεδομένα τριών ετών και το προκύπτον μοντέλο εφαρμόστηκε σε δεδομένα ενός έτους για να εκτιμήσει την χρησιμότητα του ιδιοκτήτη κατά το έτος αυτό και είχε καλή απόδοση. Ο Lai εφάρμοσε αυτό το μοντέλο στην ετήσια κατανάλωση ηλεκτρικής ενέργειας ενός κτιρίου. Οι μεταβλητές περιλαμβάνουν τις κλιματικές διακυμάνσεις. Στα πειράματα, η απόδοση εξήχθη από τα δεδομένα ενός έτους και εξετάστηκε στην συνέχεια για την συμπεριφορά δεδομένων τριών μηνών. Εξετάστηκε το μοντέλο σε ημερήσια δεδομένα με στόχο την επιβεβαίωση της σταθερότητας αυτής της προσέγγισης σε μικρά χρονικά διαστήματα. Επίσης, πρόσθεσαν διαταραχή σε ένα συγκεκριμένο κομμάτι της απόδοσης του ιστορικού και χρησιμοποίησαν το μοντέλο για την ανίχνευση της διαταραχής εξετάζοντας την αλλαγή των βαρών των συνιστωσών.

Ο Li χρησιμοποίησε τη συγκεκριμένη μέθοδο για να κάνει πρόβλεψη του ωριαίου φορτίου ψύξης ενός κτιρίου με γραφεία. Η απόδοση της παλινδρόμησης με διανύσματα στήριξης, είναι καλύτερη συγκριτικά με τα νευρωνικά δίκτυα διάδοσης προς τα πίσω.

Οι Χου και Lian χρησιμοποίησαν τα συγκεκριμένα μοντέλα για την πρόβλεψη του φορτίου ψύξης και απέδειξαν ότι είναι καλύτερα από το μοντέλο ARIMA.

Με βάση όλα τα παραπάνω προκύπτει ότι τα συγκεκριμένα μοντέλα είναι ιδιαίτερα αποδοτικά για την πρόβλεψη ωριαίας και μηνιαίας κατανάλωσης ενέργειας. Ωστόσο τα μοντέλα εφαρμόστηκαν σε μικρό αριθμό κτιρίων και το έργο τους εστιάστηκε στη μελλοντική πρόβλεψη κατανάλωσης ενέργειας. Οι Zhao και Magoules εκπαιδύσαν το μοντέλο σε δεδομένα θέρμανσης πολλών κτιρίων με στόχο να εξετάσουν την επίδραση των χαρακτηριστικών των διαφόρων κτιρίων στο μοντέλο αυτό και το δοκίμασαν σε δεδομένα φορτίων ενός καινούριου κτιρίου. Δεδομένου ότι το μέγεθος των δεδομένων εκπαίδευσης από πολλαπλά κτίρια είναι πολύ μεγάλο, καθίσταται η διαδικασία εκπαίδευσης μοντέλου εξαιρετικά αργή. Ως εκ τούτου, εφαρμόστηκε μοντέλο παράλληλων SVM για την επιτάχυνση της εκπαίδευσης του μοντέλου.

Ο Li πρόβλεψε την ετήσια κατανάλωση ενέργειας με νευρωνικά δίκτυα πίσω διάδοσης, RBF νευρωνικά δίκτυα, νευρωνικά δίκτυα παλινδρόμησης και SVMs. Έδειξε ότι τα μοντέλα νευρωνικών δικτύων παλινδρόμησης και SVMs παρουσιάζουν καλύτερη εφαρμογή σε αυτό το πρόβλημα σε σχέση με άλλα μοντέλα, με τα δεύτερα να παρουσιάζουν την καλύτερη απόδοση από όλα τα μοντέλα. Τα μοντέλα εκπαιδεύτηκαν σε δεδομένα 59 κτιρίων και δοκιμάστηκαν σε 9 κτίρια. Οι Liang και Du παρουσίασαν μια οικονομικά αποδοτική μέθοδο διάγνωσης και ανίχνευσης βλάβης για συστήματα ψύξης, εξαερισμού, θέρμανσης κτιρίων συνδυάζοντας φυσικά μοντέλα και μοντέλα SVMs.

Με τη χρήση ενός ταξινομητή SVM τεσσάρων στρωμάτων, η κανονική κατάσταση και τρεις πιθανές βλάβες μπορούν να αναγνωριστούν γρήγορα και με ακρίβεια με ένα μικρό αριθμό δειγμάτων εκπαίδευσης.

Αναφορικά με την προεπεξεργασία και την εκ των υστέρων επεξεργασία των μοντέλων έχει διεξαχθεί συγκεκριμένη έρευνα. Οι Magoules και Zhao μείωσαν τις μεταβλητές για την εκπαίδευση των μοντέλων πρόβλεψης βασιζόμενοι σε πρακτικές θεωρήσεις και στην αξιολόγηση των αποτελεσμάτων δύο αριθμητικών μεθόδων : την επιλογή του συντελεστή αυτοσυσχέτισης και της κλίσης παλινδρόμησης.

3.2.5 Γκρει μοντέλα

Όταν η πληροφορία ενός συστήματος είναι εν μέρει γνωστή, καλούμε το σύστημα αυτό σύστημα Γκρέι. Το μοντέλο Γκρέι μπορεί να χρησιμοποιηθεί για να γίνει ανάλυση της ενεργειακής συμπεριφοράς κτιρίων όταν υπάρχουν ελλιπή ή αβέβαια στοιχεία. Έχει γίνει πολύ λίγη δουλειά πάνω σε αυτό το μοντέλο. Το 1999, ο Wang εφάρμοσε ένα τέτοιο μοντέλο για να προβλέψει το σύστημα θέρμανσης ενός κτιρίου. Η ακρίβεια πρόβλεψης προέκυψε αρκετά υψηλή. Ο Guo χρησιμοποίησε ένα βελτιωμένο ανάλογο μοντέλο, για την πρόβλεψη της ενεργειακής κατανάλωσης θερμοσιφωνων και αντλιών θερμότητας σε οικίες. Αξιολόγησε την επίδραση του διαστήματος δειγμάτων στην ακρίβεια πρόβλεψης διαπιστώνοντας ότι το βέλτιστο διάστημα είναι 4 εβδομάδες. Το μοντέλο αυτό απαιτεί ελάχιστα δεδομένα και το σφάλμα πρόβλεψης περιορίζεται σε ένα φυσιολογικό εύρος. Ο Zhou έκανε on-line πρόβλεψη του φορτίου ψύξης με την ενσωμάτωση δύο ενοτήτων πρόγνωσης του καιρού σε ένα απλοποιημένο θερμικό μοντέλο φορτίου. Η μία είναι η σχετική πρόβλεψη θερμοκρασίας / υγρασίας που επιτυγχάνεται χρησιμοποιώντας ένα τροποποιημένο μοντέλο Γκρέι, η άλλη είναι η πρόβλεψη της ηλιακής ακτινοβολίας χρησιμοποιώντας ένα μοντέλο παλινδρόμησης. Τα πειραματικά αποτελέσματα έδειξαν ότι η απόδοση του απλοποιημένου μοντέλου θερμικού δικτύου βελτιώνεται όσο χρησιμοποιούνται τα προβλεπόμενα δεδομένα καιρού από το πρώτο στοιχείο στην διαδικασία εκπαίδευσης.

Κεφάλαιο 4 : Μοντέλα Πρόβλεψης

4.1 Κατηγορίες Μοντέλων Πρόβλεψης

Η παραγωγή προβλέψεων επιτυγχάνεται με την αξιοποίηση της διαθέσιμης γνώσης και εμπειρίας και αφορά μελλοντικά γεγονότα που δεν έχουν παρατηρηθεί ακόμα.

4.1.1 Κριτικές Προβλέψεις

Στην περίπτωση που η παραγωγή προβλέψεων επιτυγχάνεται με τη συμβολή ειδικών και εμπειρογνομόνων, αναφερόμαστε σε κριτικές προβλέψεις (judgmental forecasting). Τα κριτικά μοντέλα πρόβλεψης μπορούν να αναλυθούν στην κριτική πρόβλεψη και στην πρόβλεψη στόχου. Η κριτική πρόβλεψη μπορεί να βασίζεται στη γνώση, την κρίση και την πείρα ενός ατόμου ή μιας ομάδας ατόμων χωρίς την απαίτηση δεδομένων όπως γίνεται στη στατιστική πρόβλεψη. Ακόμη, μπορεί να λαμβάνει υπόψη ειδικά γεγονότα, να αντισταθμίζει τις ανεπάρκειες και να εξισορροπεί τις ανομοιογένειες των καταγεγραμμένων δεδομένων. Καθίσταται έτσι κατάλληλη σε περιπτώσεις που περισσότερο από τεχνοοικονομικούς παράγοντες έχουμε ηθικούς και όταν είναι επιθυμητός ο έλεγχος του μεγέθους του οποίου η ζήτηση προβλέπεται, ενώ το μειονέκτημά της είναι η προκατάληψη των ατόμων στη κρίση των οποίων στηρίζονται τα αποτελέσματά της.

Από την άλλη, η πρόβλεψη στόχου είναι ουσιαστικά η επιθυμητή πορεία της χρονοσειράς που θέλει η επιχείρηση να έχει και όχι τόσο η πραγματική πρόβλεψή της. Με ανάλυση των ιστορικών δεδομένων του μεγέθους που μελετάται υπολογίζεται ο ρυθμός ανάπτυξής του για το παρελθόν διάστημα. Στη συνέχεια, η εφαρμογή των στατιστικών προβλέψεων δίνει μια εκτίμηση της πορείας της χρονοσειράς και πλέον μένει ο χειριστής των δεδομένων να τροποποιήσει το μελλοντικό ρυθμό ανάπτυξης, σύμφωνα με την επιθυμητή μελλοντική πορεία της επιχείρησης. Η τελική πρόβλεψη είναι συνεπώς εδώ η πρόβλεψη του διευθυντή ανάπτυξης (χειριστή) και περιέχει αρκετή αισιοδοξία και μεροληψία, καθώς καθορίζει τους στόχους της επιχείρησης.

4.1.2 Στατιστικές Προβλέψεις

Στην περίπτωση που η διαδικασία παραγωγής προβλέψεων μπορεί να συστηματοποιηθεί μέσω στατιστικών μεθόδων πρόβλεψης, με στόχο την αυτόματη παραγωγή προβλέψεων, αναφερόμαστε σε στατιστικές προβλέψεις.

Οι στατιστικές προβλέψεις αναφέρονται στην εφαρμογή στατιστικών μοντέλων χρονοσειρών ή αιτιοκρατικών μοντέλων επί μιας σειράς δεδομένων με στόχο τη συστηματική παραγωγή προβλέψεων (Πετρόπουλος, Ασημακόπουλος, 2011). Η χρήση του μοντέλου χρονοσειρών παράγει ικανοποιητικά αποτελέσματα στις περιπτώσεις που η εξέλιξη της χρονοσειράς γίνει με βάση τις τρέχουσες συνθήκες (εφόσον οι προβλέψεις στηρίζονται στις ιστορικές παρατηρήσεις). Ένα σημαντικό μειονέκτημα του μοντέλου χρονοσειρών είναι η μη συσχέτιση του προβλεπόμενου μεγέθους με παράγοντες που πιθανότατα το επηρεάζουν σημαντικά.

Στις μεθόδους χρονοσειρών συγκαταλέγονται οι μέθοδοι αποσύνθεσης, εξομάλυνσης, οι αιτιοκρατικές μέθοδοι και οι αυτοπαλινδρομούμενες μέθοδοι κινητού μέσου όρου (Autoregressive Moving Average). Στην παρούσα εργασία θα ασχοληθούμε αποκλειστικά με τις στατιστικές μεθόδους πρόβλεψης χρονοσειρών.

4.2 Μοντέλα Εκθετικής εξομάλυνσης

4.2.1 Γενικά για μεθόδους εκθετικής εξομάλυνσης

Σε πολλές περιπτώσεις η ανάπτυξη και η εφαρμογή κάποιας πολύπλοκης μεθόδου πρόβλεψης δεν είναι πρακτική καθώς απαιτείται επεξεργαστική ισχύς και σημαντικός χρόνος. Υπάρχει ανάγκη, λοιπόν, για μια τεχνική η οποία να μπορεί να εφαρμοστεί εύκολα και να παρέχει ικανοποιητικές προβλέψεις σε βραχυπρόθεσμο ορίζοντα. Σε τέτοιες περιπτώσεις, χρησιμοποιείται συχνά μια κλάση μεθόδων πρόβλεψης, γνωστές ως μέθοδοι εξομάλυνσης. Στα πλαίσια αυτών των μεθόδων, γίνεται ουσιαστικά χρήση των ιστορικών δεδομένων για τον προσδιορισμό μιας εξομαλυμένης τιμής για τις χρονοσειρές. Στη συνέχεια, αυτή η εξομαλυμένη τιμή προεκτείνεται, ώστε να αποτελέσει πρόβλεψη για τη μελλοντική τιμή της χρονοσειράς.

Η βασική ιδέα των μεθόδων εξομάλυνσης είναι ότι υπάρχει ένα λανθάνον πρότυπο συμπεριφοράς (underlying pattern), το οποίο ακολουθούν οι τιμές των μεταβλητών πρόβλεψης και ότι οι ιστορικές παρατηρήσεις της κάθε μεταβλητής αντιπροσωπεύουν αυτό το πρότυπο καθώς και τυχαίες διακυμάνσεις. Στόχος αυτών των μεθόδων πρόβλεψης είναι να διακρίνουν τις τυχαίες αποκλίσεις και το βασικό πρότυπο, εξομαλύνοντας τα ιστορικά δεδομένα. Επιτυγχάνεται, δηλαδή, ελαχιστοποίηση της τυχειότητας που παρουσιάζεται στην ιστορική ακολουθία με αποτέλεσμα η πρόβλεψη να βασίζεται στον εξομαλυμένο τρόπο συμπεριφοράς των δεδομένων.

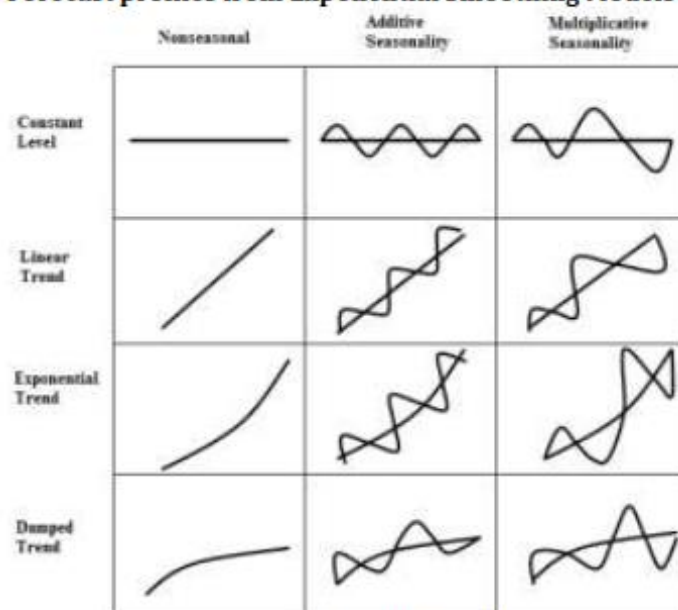
Η εκθετική εξομάλυνση είναι μία μέθοδος πρόβλεψης, η οποία προεκτείνει στο μέλλον στοιχεία του προτύπου των ιστορικών δεδομένων όπως τάσεις, και εποχιακούς κύκλους. Οι προβλέψεις υπολογίζονται μετά από εξομάλυνση των ιστορικών δεδομένων προκειμένου να ελαχιστοποιηθεί η τυχειότητα. Οι μέθοδοι εκθετικής εξομάλυνσης έχουν ως βασική αρχή το γεγονός, ότι όσο πιο πρόσφατα είναι τα δεδομένα, τόσο πιο πολλή η πληροφορία που εμπεριέχεται σε αυτά. Για το λόγο αυτό, δίνεται μεγαλύτερη βαρύτητα στα πρόσφατα δεδομένα και η φθίνει εκθετικά καθώς αναφερόμαστε σε δεδομένα παλιότερων χρονικών περιόδων.

Οι μέθοδοι της εκθετικής εξομάλυνσης αναπτύχθηκαν στις αρχές της δεκαετίας του 1950, γνώρισαν ευρεία εφαρμογή με την επανάσταση της πληροφορικής τη δεκαετία του 1960 και αποτελούν έως και σήμερα ένα σημαντικό εργαλείο παραγωγής μαζικών προβλέψεων. Η δημοτικότητα των μεθόδων αυτών οφείλεται στην απλότητα των μοντέλων που υιοθετούν, τις περιορισμένες απαιτήσεις τους σε αποθήκευση δεδομένων και το μειωμένο υπολογιστικό φόρτο. Εμπειρικές μελέτες έχουν αναδείξει τα ικανοποιητικά ποσοστά ακρίβειας σε σχέση με άλλες πιο πολύπλοκες μεθόδους πρόβλεψης. Αυτό ερμηνεύεται από το γεγονός ότι οι μέθοδοι εκθετικής εξομάλυνσης δεν επηρεάζονται από τις ιδιομορφίες των προτύπων των δεδομένων ή από περιστασιακά ακραίες τιμές που τυγχάνει να εμφανίζονται.

4.2.2 Τύποι Μοντέλων Εξομάλυνσης

Τα μοντέλα εξομάλυνσης κατηγοριοποιούνται ανάλογα με την γενική μορφή της γραφικής αναπαράστασης των ιστορικών δεδομένων με οριζόντιο άξονα τον χρόνο. Οι βασικές κατηγορίες παρουσιάζονται στο παρακάτω σχήμα, όπου τέσσερα μοντέλα τάσης (constant level, linear level, exponential level, damped level) συνδυάζονται με τρία εποχιακά μοντέλα (no seasonality, additive seasonality, multiplicative seasonality).

Graph 2 - Forecast profiles from Exponential Smoothing Models by Gardner



Source: Gardner (1987, p.175).

Σχήμα 4.1 : Βασικές κατηγορίες μοντέλων εξομάλυνσηςως συνδυασμοί 4 μοντέλων τάσης και 3 εποχιακών μοντέλων (Πηγή : Gardner, 1987)

Το μοντέλο σταθερού επιπέδου (constant level) υποθέτει την απουσία τάσης από τα δεδομένα. Θεωρείται, δηλαδή, ότι οι χρονοσειρές έχουν ένα σχετικά σταθερό μέσο όρο και η πρόβλεψη προκύπτει από την προέκταση μιας οριζόντιας ευθείας γραμμής. Τα μοντέλα σταθερού επιπέδου χρησιμοποιούνται ευρέως για προβλέψεις ενός βήματος (one-step-ahead forecasting) ή για προβλέψεις χρονοσειρών που χαρακτηρίζονται από αυξημένο θόρυβο ή τυχαιότητα.

Το μοντέλο γραμμικής τάσης (linear trend) αποτελεί την πιο δημοφιλή παραδοχή. Οι προβλέψεις προκύπτουν απλά από την προέκταση μιας ευθείας γραμμής για οποιαδήποτε χρονική περίοδο στο μέλλον.

Το μοντέλο εκθετικής τάσης (exponential trend) χρησιμοποιείται για προβλέψεις, όπου το προβλεπόμενο μέγεθος τείνει να αυξάνει συνεχώς και περισσότερο.

Το μοντέλο της φθίνουσας τάσης (damped trend) χρησιμοποιείται για προβλέψεις, όπου το προβλεπόμενο μέγεθος τείνει να αυξάνει συνεχώς αλλά με ολόένα και μικρότερο ρυθμό καθώς επιμηκύνεται ο χρονικός ορίζοντας πρόβλεψης.

Πολλές εμπειρικές μελέτες έχουν δείξει ότι τα μοντέλα γραμμικής και εκθετικής τάσης οδηγούν σε υπεραισιόδοξες προβλέψεις και, ακολούθως, σε υψηλές τιμές του στατιστικού δείκτη mean error (bias) καθώς επιμηκύνεται ο χρονικός ορίζοντας πρόβλεψης. Ειδικά στις περιπτώσεις μεσοπρόθεσμων και μακροπρόθεσμων προβλέψεων, το μοντέλο της φθίνουσας τάσης αποτελεί την βέλτιστη επιλογή. Σύμφωνα με το μοντέλο αυτό μειώνεται βαθμιαία το μέγεθος κατά το οποίο αυξάνονται οι τιμές της χρονοσειράς σε κάθε χρονική περίοδο.

Αναφορικά με τα μοντέλα εποχιακότητας που παρουσιάζονται στο παραπάνω σχήμα, στο προσθετικό μοντέλο εποχιακότητας (additive seasonality), το εύρος των εποχιακών διακυμάνσεων θεωρείται σταθερό και ανεξάρτητο από οποιαδήποτε αύξηση στις τιμές της χρονοσειράς. Αντίστοιχα στο πολλαπλασιαστικό μοντέλο εποχιακότητας (multiplicative seasonality), το οποίο είναι και το επικρατέστερο στην πράξη, οι εποχιακές διακυμάνσεις είναι ανάλογες του ύψους τιμών των δεδομένων, δηλαδή, καθώς η τάση αυξάνει, οι εποχιακές διακυμάνσεις γίνονται μεγαλύτερες.

4.2.3 Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing)

Το μοντέλο σταθερού επιπέδου, το οποίο αναφέρεται και ως απλή εκθετική εξομάλυνση (SES) περιγράφεται από τις παρακάτω εξισώσεις :

$$e_t = Y_t - F_t \quad (4.1)$$

$$S_t = S_{t-1} + \alpha \cdot e_t \quad (4.2)$$

$$F_{t+1} = S_t \quad (4.3)$$

Το e δηλώνει το σφάλμα (απόκλιση πραγματικής τιμής από πρόβλεψη), το S δηλώνει το επίπεδο, το F την πρόβλεψη και το t την χρονική στιγμή. Η παράμετρος α αποτελεί το συντελεστή εξομάλυνσης της μεθόδου και μπορεί να λάβει τιμές στο διάστημα $[0,1]$.

4.2.3.1 Αρχικό Επίπεδο

Όπως εύκολα γίνεται αντιληπτό από την αναδρομικότητα του μοντέλου, προκειμένου να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης αρχικά και, κατόπιν, των ζητούμενων σημειακών προβλέψεων δεδομένου χρονικού ορίζοντα, θα πρέπει να οριστεί ένα αρχικό επίπεδο S_0 . Ως αρχικό επίπεδο μπορεί να χρησιμοποιηθεί:

- Ο μέσος όρος όλων των παρατηρήσεων
- Ο μέσος όρος των n πρώτων παρατηρήσεων
- Η πρώτη παρατήρηση
- Το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης.

Η επιλογή μιας από τις εναλλακτικές σχετίζεται με τα ποιοτικά χαρακτηριστικά της ενίστε χρονοσειράς. Αν επιλεχθεί μη αντιπροσωπευτικό των δεδομένων αρχικό επίπεδο, τότε αυτό θα έχει αρνητική επίδραση στις τιμές των προβλέψεων. Από τις παραπάνω εξισώσεις γίνεται σαφές ότι το αρχικό επίπεδο που θα επιλεγεί αποτελεί ουσιαστικά και την πρώτη πρόβλεψη. Επιπρόσθετα, το αρχικό επίπεδο επηρεάζει και την επιλογή του συντελεστή εξομάλυνσης.

4.2.3.2 Συντελεστής Εξομάλυνσης

Ο καθορισμός του βέλτιστου συντελεστή εξομάλυνσης επηρεάζεται από δύο παράγοντες οι οποίοι αλληλοεξαρτώνται. Ο πρώτος παράγοντας συνίσταται στο ποσοστό θορύβου στην χρονοσειρά. Όσο περισσότερος θόρυβος υπάρχει στα δεδομένα της χρονοσειράς, τόσο μικρότερη θα πρέπει να είναι η τιμή του συντελεστή εξομάλυνσης ώστε να αποφεύγεται η υπερβόλική αντίδραση στον θόρυβο. Ο δεύτερος παράγοντας αφορά στη σταθερότητα του μέσου όρου της χρονοσειράς. Αν ο μέσος όρος μεταβάλλεται, ο συντελεστής εξομάλυνσης θα πρέπει να είναι μεγάλος ώστε οι προβλέψεις να παρακολουθούν τις μεταβολές που παρουσιάζουν τα δεδομένα. Αντίθετα αν ο μέσος όρος των δεδομένων είναι σταθερός, θα πρέπει να επιλέγεται μικρός συντελεστής εξομάλυνσης.

Ο καθορισμός της βέλτιστης παραμέτρου εξομάλυνσης είναι μια υπολογιστική διαδικασία που δύναται να πραγματοποιηθεί σχετικά άμεσα και αυτόματα από σύγχρονα υπολογιστικά συστήματα.

Η πιο διαδεδομένη μέθοδος υπολογισμού της βέλτιστης παραμέτρου έγκειται στην γραμμική αναζήτηση της παραμέτρου που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (MSE), έχοντας εξαρχής θέσει την ακρίβεια της αναζήτησης. Για παράδειγμα θέτοντας ακρίβεια στο εκατοστό και δεδομένου του πεδίου τιμών [0,1], στο οποίο η παραμέτρος α μπορεί να λάβει τιμές, η διαδικασία αναζήτησης ολοκληρώνεται μετά από 101 επαναλήψεις. Σε κάθε επανάληψη υπολογίζεται η στατιστική ακρίβεια του μοντέλου με την τρέχουσα παράμετρο εξομάλυνσης μέσω του δείκτη MSE και γίνεται ορισμός της βέλτιστης παραμέτρου σε περίπτωση που υπολογιστεί ελάχιστη ποσότητα του δείκτη MSE.

Μια δεύτερη μέθοδος επιλογής βέλτιστης παραμέτρου παρουσιάζεται στην παρακάτω διαδικασία και βασίζεται στην προϋπόθεση της ύπαρξης ενός ακριβώς τοπικού ελαχίστου της γραφικής παράστασης του σε συνάρτηση με την παράμετρο εξομάλυνσης α στο διάστημα (0,1). Αρχικά υπολογίζεται η τιμή του σφάλματος για δύο τιμές του συντελεστή εξομάλυνσης : τις τιμές $\alpha = 0,33$ και $\alpha = 0,67$ και επιλέγεται η βέλτιστη. Έπειτα υπολογίζεται η τιμή του συγκεκριμένου σφάλματος για τιμές $\pm 0,17$ γύρω από το βέλτιστο α και επιλέγεται η βέλτιστη τιμή. Αυτή η διαδικασία εκτελείται επαναληπτικά μέχρι το σημείο που η μεταβολή του σφάλματος γίνει μικρότερη από 1% ενώ τα αντίστοιχα διαστήματα μειώνονται σταδιακά σε $\pm 0,08, \pm 0,04, \pm 0,02, κ.ο.κ. .$

Όταν η παράμετρος εξομάλυνσης λάβει τιμή κοντά στη μονάδα, αυτό σημαίνει ότι το σφάλμα της τελευταίας περιόδου διαδραματίζει σημαντικό ρόλο στον υπολογισμό της επόμενης σημειακής πρόβλεψης. Σε αντίθετη περίπτωση, η νέα πρόβλεψη θα μεταβληθεί σε πολύ μικρό βαθμό σε σχέση με την τελευταία περίοδο. Στην ακραία περίπτωση όπου $\alpha=1$, το μοντέλο ταυτίζεται με την απλοϊκή μέθοδο (Naive), ενώ στην περίπτωση που $\alpha=0$, τότε κάθε πρόβλεψη είναι ίση με το αρχικό επίπεδο.

4.2.3.3 Εναλλακτικός τρόπος υπολογισμού των προβλέψεων

Από τις αρχικές εξισώσεις του μοντέλου έχουμε :

$$F_{t+1} = S_t = S_{t-1} + \alpha \cdot e_t$$

$$\begin{matrix} F_t = S_{t-1} \\ \longrightarrow \end{matrix} F_{t+1} = F_t + \alpha \cdot e_t$$

$$\begin{matrix} e_t = Y_t - F_t \\ \longrightarrow \end{matrix} F_{t+1} = F_t + \alpha \cdot (Y_t - F_t)$$

$$\text{Άρα : } F_{t+1} = \alpha \cdot Y_t + (1-\alpha) \cdot F_t \quad (4.4)$$

Σύμφωνα με την σχέση 4.4, η πρόβλεψη αποτελεί έναν γραμμικό συνδυασμό της προηγούμενης παρατήρησης και της προηγούμενης πρόβλεψης, με βάρη που καθορίζονται από την παράμετρο εξομάλυνσης. Συλλογίζομενοι πως στον καθορισμό της πρόβλεψης F_t συμμετέχει με βάρος $(1-\alpha)$ η πρόβλεψη F_{t-1} και με βάρος α η πραγματική παρατήρηση Y_{t-1} , συμπεραίνουμε πως τελικά η πρόβλεψη εξαρτάται από όλες τις προηγούμενες παρατηρήσεις με βάρη που φθίνουν εκθετικά αλλά και από την αρχική πρόβλεψη, που όπως ήδη επισημάναμε είναι το επιλεγθέν αρχικό επίπεδο.

Η συμπεριφορά του μοντέλου σταθερού επιπέδου θυμίζει τη συμπεριφορά ενός θερμοστάτη. Για κάθε τιμή της χρονοσειράς, υπολογίζεται το σφάλμα πρόβλεψης. Αν το σφάλμα είναι θετικό (δηλαδή η προηγούμενη πρόβλεψη ήταν μικρότερη της πραγματικής τιμής), τότε η πρόβλεψη για την επόμενη χρονική περίοδο αυξάνεται. Αν το σφάλμα είναι αρνητικό (δηλαδή η προηγούμενη πρόβλεψη ήταν μεγαλύτερη της πραγματικής τιμής), τότε η πρόβλεψη μειώνεται. Με τον τρόπο αυτό, τα σφάλμα χρησιμοποιούνται για να καθοδηγήσουν τις προβλέψεις προς ένα συγκεκριμένο στόχο, το πραγματικό επίπεδο της χρονοσειράς. Τέλος, στην περίπτωση που ζητάμε περισσότερες απο μία σημειακές προβλέψεις, τότε όλες αυτές τίθενται ίσες με την τελευταία υπολογισμένη πρόβλεψη καθώς το μοντέλο θεωρεί την ύπαρξη ενός σταθερού επιπέδου.

4.2.4 Μοντέλο εξομάλυνσης γραμμικής τάσης (Holt Exponential Smoothing)

Το μοντέλο εξομάλυνσης για γραμμική τάση είναι μία επέκταση του μοντέλου απλής εκθετικής εξομάλυνσης, η οποία δύναται επιπρόσθετα να διαχειριστεί τη συνιστώσα της παρατηρούμενης τάσης. Η επέκταση αυτή προτάθηκε από τον Holt το 1957. Το μοντέλο εξομάλυνσης γραμμικής τάσης περιγράφεται από τις παρακάτω σχέσεις:

$$e_t = Y_t - F_t \quad (4.5)$$

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t \quad (4.6)$$

$$T_t = T_{t-1} + \beta \cdot e_t \quad (4.7)$$

$$F_{t+m} = S_t + m \cdot T_t \quad (4.8)$$

Το e δηλώνει το σφάλμα (απόκλιση πραγματικής τιμής από πρόβλεψη), το S δηλώνει το επίπεδο, το T την τάση, το F την πρόβλεψη, το t την χρονική στιγμή και το m το χρονικό ορίζοντα της πρόβλεψης.

Η παράμετρος α αποτελεί το συντελεστή εξομάλυνσης της μεθόδου για το επίπεδο ενώ η παράμετρος β αποτελεί το συντελεστή εξομάλυνσης της μεθόδου για την τάση. Οι δύο παράμετροι λαμβάνουν τιμές στο διάστημα $[0,1]$.

4.2.4.1 Αρχικό επίπεδο και αρχική τάση

Όπως εύκολα γίνεται αντιληπτό από την αναδρομικότητα του μοντέλου, προκειμένου να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης αρχικά και, κατόπιν, των ζητούμενων σημειακών προβλέψεων δεδομένου χρονικού ορίζοντα, θα πρέπει να οριστεί ένα αρχικό επίπεδο S_0 και μία αρχική τάση T_0 . Ως αρχική τάση T_0 μπορεί να χρησιμοποιηθεί:

- ο Διαφορά δεύτερης και πρώτης παρατήρησης ($Y_2 - Y_1$)
- ο Διαφορά n -οστής και πρώτης παρατήρησης διαιρεμένης με $(n-1)$
- ο Σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης

Η επιλογή μιας από τις εναλλακτικές σχετίζεται με τα ποιοτικά χαρακτηριστικά της ενίοτε χρονοσειράς. Αν επιλεγεί μη αντιπροσωπευτικό των δεδομένων αρχικό επίπεδο ή τάση, τότε αυτό θα έχει αρνητική επίδραση στις τιμές των προβλέψεων. Στην περίπτωση μαζικής πρόβλεψης σε μεγάλο πλήθος χρονοσειρών συνίσταται η εφαρμογή γραμμικής παλινδρόμησης με ανεξάρτητη μεταβλητή τον χρόνο.

4.2.4.2 Συντελεστές εξομάλυνσης

Η διαδικασία εύρεσης του βέλτιστου συνδυασμού τιμών για τους συντελεστές εξομάλυνσης βασίζεται στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE). Η πιο διαδεδομένη μέθοδος καθορισμού των βέλτιστων συντελεστών εξομάλυνσης είναι η γραμμική αναζήτηση αυτών που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα, έχοντας εξαρχής θέσει την ακρίβεια της πρόβλεψης. Για παράδειγμα θέτοντας ακρίβεια στο εκατοστό και δεδομένου του πεδίου τιμών $[0,1]$, στο οποίο οι παραμέτρους α, β μπορούν να λάβουν τιμές, η διαδικασία αναζήτησης ολοκληρώνεται μετά από $101 \cdot 101 = 10201$ επαναλήψεις. Σε κάθε επανάληψη υπολογίζεται η στατιστική ακρίβεια του μοντέλου με τον τρέχοντα συνδυασμό παραμέτρων εξομάλυνσης α, β μέσω του δείκτη MSE και γίνεται ορισμός του βέλτιστου συνδυασμού παραμέτρων σε περίπτωση που υπολογιστεί ελάχιστη ποσότητα του δείκτη MSE.

Μια δεύτερη μέθοδος επιλογής βέλτιστης παραμέτρου παρουσιάζεται στην παρακάτω διαδικασία και βασίζεται στην προϋπόθεση της ύπαρξης ενός ακριβώς τοπικού ελαχίστου της γραφικής παράστασης του σε συνάρτηση με την παράμετρο εξομάλυνσης α στο διάστημα $(0,1)$. Αρχικά υπολογίζεται η τιμή του σφάλματος για δύο τιμές του συντελεστή εξομάλυνσης: τις τιμές $\alpha = 0,33$ και $\alpha = 0,67$ και επιλέγεται η βέλτιστη. Έπειτα υπολογίζεται η τιμή του συγκεκριμένου σφάλματος για τιμές $\pm 0,17$ γύρω από το βέλτιστο α και επιλέγεται η βέλτιστη τιμή. Αυτή η διαδικασία εκτελείται επαναληπτικά μέχρι το σημείο που η μεταβολή του σφάλματος γίνει μικρότερη από 1% ενώ τα αντίστοιχα διαστήματα μειώνονται σταδιακά σε $\pm 0,08, \pm 0,04, \pm 0,02, κ. ο. κ. .$

Μια δεύτερη μέθοδος επιλογής βέλτιστης παραμέτρου παρουσιάζεται στην παρακάτω διαδικασία. Αρχικά υπολογίζεται η τιμή του σφάλματος για τέσσερις συνδυασμούς παραμέτρων εξομάλυνσης: τις τιμές $\alpha = 0,33$ ή $\alpha = 0,67$ και $\beta = 0,33$ ή $\beta = 0,67$. Επιλέγεται ο βέλτιστος συνδυασμός. Έπειτα υπολογίζεται η τιμή του συγκεκριμένου σφάλματος για τιμές $\pm 0,17$ γύρω από το βέλτιστο συνδυασμό και επιλέγεται και πάλι ο βέλτιστος συνδυασμός. Αυτή η διαδικασία εκτελείται επαναληπτικά μέχρι το σημείο που η μεταβολή του σφάλματος γίνει μικρότερη από 1% ενώ τα αντίστοιχα διαστήματα μειώνονται σταδιακά σε $\pm 0,08, \pm 0,04, \pm 0,02, κ. ο. κ. .$

Συνήθως η βέλτιστη τιμή του συντελεστή για την τάση β είναι μικρότερη από την τιμή του συντελεστή α για το επίπεδο. Ο λόγος που συμβαίνει αυτό είναι ότι η τιμή της τάσης για κάθε περίοδο είναι συνήθως πολύ μικρότερη από την τιμή του επιπέδου. Αν ρυθωθούν το ίδιο ποσοστό σφάλματος και στο επίπεδο και στην τάση τότε οι προβλέψεις που θα προκύψουν μπορεί να είναι ασταθείς.

4.2.5 Μοντέλο εξομάλυνσης μη γραμμικής τάσης

Το μοντέλο της γραμμικής τάσης, όπως παρατηρήθηκε και προηγουμένως, παρουσιάζει μια τάση για υπερεκτίμηση (θετική προκατάληψη) ιδιαίτερα όταν εφαρμόζεται σε μεσοπρόθεσμες ή μακροπρόθεσμες προβλέψεις. Για το λόγο αυτό, οι Gardner και McKenzie το 1985, περιέγραψαν πως το μοντέλο της γραμμικής τάσης μπορεί να μεταβληθεί κατάλληλα ώστε να μπορεί να προσαρμόζεται και στις μη γραμμικές τάσεις. Ουσιαστικά αυτή η μεταβολή επιτυγχάνεται με την προσθήκη μιας παραμέτρου, η οποία ελέγχει το ρυθμό αύξησης των τιμών τους. Αυτή η παράμετρος διόρθωσης της τάσης (trend-modification parameter) συμβολίζεται με ϕ . Οι εξισώσεις που περιγράφουν το μοντέλο μη γραμμικής τάσης είναι οι ακόλουθες :

$$e_t = Y_t - F_t \quad (4.9)$$

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t \quad (4.10)$$

$$T_t = T_{t-1} + \beta \cdot e_t \quad (4.11)$$

$$F_{t+m} = S_t + \sum_{i=1}^m \varphi^i \cdot T_t \quad (4.12)$$

Οι εξισώσεις αυτές είναι πανομοιότυπες με αυτές του γραμμικού μοντέλου πλήν της τελευταίας, σύμφωνα με την οποία πραγματοποιείται ένα μη γραμμικός υπολογισμός της τάσης με βάση την παράμετρο εξομάλυνσης ϕ , αντί της γραμμικής αύξησης της τάσης μέσω του συντελεστή m , που προτείνεται στην εξομάλυνση γραμμικής τάσης.

Η παράμετρος ϕ σε αντιδιαστολή με τις παραμέτρους α, β μπορεί να λάβει θετικές τιμές χωρίς κάποιο άνω όριο. Κρίνεται, ωστόσο, πολύ σημαντική η επιβολή άνω και κάτω ορίων ανάλογα με την εκάστοτε περίπτωση. Ανάλογα με την τιμή που παίρνει η παράμετρος ϕ , το μοντέλο της μη γραμμικής τάσης μπορεί να λάβει τις ακόλουθες μορφές :

- Αν $\phi=0$, προκύπτει το μοντέλο της απλής εκθετικής εξομάλυνσης, καθώς η τάση δεν έχει πλέον καμία επίδραση στον καθορισμό των στατιστικών σημειακών προβλέψεων.
- Αν $0 < \phi < 1$, τότε προκύπτει το μοντέλο της φθίνουσας τάσης (damped exponential smoothing), το οποίο θεωρείται καταλληλότερο όταν καλούμαστε να παράγουμε μακροπρόθεσμες προβλέψεις, αφού χαρακτηρίζεται από έλλειψη τάσης για υπεραισιοδοξία.
- Αν $\phi=1$, τότε προκύπτει το μοντέλο της γραμμικής τάσης, καθώς στην εξίσωση υπολογισμού της πρόβλεψης, το άθροισμα αντικαθίσταται από το γινόμενο της μεταβλητής χρονικού ορίζοντα m και της προηγούμενης τάσης T_t .
- Αν $\phi > 1$, τότε προκύπτει το μοντέλο της εκθετικής εξομάλυνσης, το οποίο χαρακτηρίζεται από μεγάλη θετική προκατάληψη και η χρήση του συνίσταται μόνο

στις περιπτώσεις πρόβλεψης ζήτησης στην αρχή του κύκλου ζωής ενός προϊόντος ή μια υπηρεσίας.

Η χρήση του μοντέλου εξομάλυνσης μη γραμμικής τάσης εγκυμονεί κινδύνους λόγω της περίπτωσης παραγωγής προκατειλημμένων προβλέψεων, γεγονός που πιθανότητα οφείλεται σε λανθασμένη βελτιστοποίηση του συντελεστή ϕ . Το μοντέλο εκθετικής τάσης μπορεί να αποκλειστεί αν επιλέξουμε το ϕ να λαμβάνει τιμές στο διάστημα $[0,1]$. Εμπειρικές ερευνητικές μελέτες έχουν αναδείξει τη μεγαλύτερη ακρίβεια προβλέψεων του μοντέλου εξομάλυνσης μη γραμμικής τάσης σε σχέση με το μοντέλο εξομάλυνσης γραμμικής τάσης. Γενικά το μοντέλο μη γραμμικής τάσης δίνει ικανοποιητικά αποτελέσματα σε περιπτώσεις που είναι αδύνατη η επιλογή κάποιου συγκεκριμένου μοντέλου για την περιγραφή κάθε χρονοσειράς ή σε περιπτώσεις που απαιτείται ένα αυτοματοποιημένο σύστημα για παραγωγή προβλέψεων μεγάλου πλήθους χρονοσειρών (batch forecasting), ειδικά αν οι ορίζοντες πρόβλεψης των αποθεμάτων είναι μεγαλύτεροι της μίας χρονικής περιόδου. Συμπερασματικά, όσο μεγαλύτερος είναι ο απαιτούμενος ορίζοντας πρόβλεψης, τόσο πλεονεκτεί το μοντέλο μη γραμμικής τάσης από τα άλλα μοντέλα εκθετικής εξομάλυνσης.

4.2.5.1 Αρχικό επίπεδο και αρχική τάση

Όπως και στα μοντέλα εξομάλυνσης που μελετήθηκαν προηγουμένως, προκειμένου να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης αρχικά και, κατόπιν, των ζητούμενων σημειακών προβλέψεων δεδομένου χρονικού ορίζοντα, θα πρέπει να οριστεί ένα αρχικό επίπεδο S_0 και μία αρχική τάση T_0 . Συνίσταται η εφαρμογή γραμμικής παλινδρόμησης με ανεξάρτητη μεταβλητή τον χρόνο για τον προσδιορισμό των S_0 και T_0 .

4.2.5.2 Συντελεστές εξομάλυνσης

Η διαδικασία εύρεσης του βέλτιστου συνδυασμού τιμών για τους συντελεστές εξομάλυνσης βασίζεται στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE). Προτείνεται η γραμμική αναζήτηση των συντελεστών εξομάλυνσης που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα, έχοντας εξαρχής θέσει την ακρίβεια της πρόβλεψης. Για παράδειγμα θέτοντας ακρίβεια στο εκατοστό και δεδομένου του πεδίου τιμών $[0,1]$, στο οποίο οι παραμέτροι α, β, ϕ μπορούν να λάβουν τιμές, η διαδικασία αναζήτησης ολοκληρώνεται μετά από $101 \cdot 101 \cdot 101 = 1.030.301$ επαναλήψεις. Σε κάθε επανάληψη υπολογίζεται η στατιστική ακρίβεια του μοντέλου με τον τρέχοντα συνδυασμό παραμέτρων εξομάλυνσης α, β, ϕ μέσω του δείκτη MSE και γίνεται ορισμός του βέλτιστου συνδυασμού παραμέτρων σε περίπτωση που υπολογιστεί ελάχιστη ποσότητα του δείκτη MSE.

4.2.6 Εποχιακή εξομάλυνση

Τα μοντέλα εξομάλυνσης που περιγράφηκαν στις προηγούμενες παραγράφους είναι μη εποχιακά και είναι κατάλληλα για χρονοσειρές που εμφανίζουν μόνο πρότυπα τάσεων. Αν, ωστόσο, η παρουσία της εποχιακής συνιστώσας είναι έντονη στα ιστορικά δεδομένα, τα εποχιακά μοντέλα θα πρέπει να επεκταθούν με την προσθήκη ενός κατάλληλου εξομαλυμένου εποχιακού παράγοντα για κάθε χρονική περίοδο σε διάστημα ενός έτους. Ο ρόλος του παράγοντα αυτού είναι η διόρθωση των προβλέψεων ανάλογα με την αναμενόμενη εποχιακή διακύμανση.

Αν οι εποχιακές διακυμάνσεις είναι προσθετικού χαρακτήρα, ο εποχιακός παράγοντας ορίζεται ως η διαφορά κάθε σημείου της χρονοσειράς και του μέσου όρου των τιμών της

χρονοσειράς για όλο το έτος. Με την χρήση του προσθετικού εποχιακού παράγοντα, προκύπτουν οι αποεποχικοποιημένες τιμές των δεδομένων, δηλαδή οι τιμές από τις οποίες έχει αφαιρεθεί η εποχιακή συνιστώσα. Ισχύει, δηλαδή :

$$\text{πραγματικές τιμές-εποχιακός παράγοντας} = \text{αποεποχικοποιημένες τιμές} \quad (4.13)$$

Μπορούμε, φυσικά, να επαναφέρουμε την εποχιακότητα εφαρμόζοντας τη σχέση:

$$\text{αποεποχικοποιημένες τιμές} + \text{εποχιακός παράγοντας} = \text{πραγματικές τιμές}$$

Αν οι εποχιακές διακυμάνσεις είναι πολλαπλασιαστικού χαρακτήρα, ο εποχιακός παράγοντας ορίζεται ως ο λόγος κάθε τιμής της χρονοσειράς προς το μέσο όρο των τιμών της χρονοσειράς για όλο το έτος. Για την αποεποχικοποίηση σε αυτή την περίπτωση χρησιμοποιείται η ακόλουθη σχέση :

$$\frac{\text{πραγματικές τιμές}}{\text{εποχιακός παράγοντας}} = \text{αποεποχικοποιημένες τιμές} \quad (4.14)$$

Για την επαναφορά της εποχιακότητας εφαρμόζουμε, αντίστοιχα την παρακάτω σχέση :

$$\text{αποεποχικοποιημένες τιμές} \cdot \text{εποχιακός παράγοντας} = \text{πραγματικές τιμές}$$

4.2.6.1 Μοντέλο Winters

Στη συνέχεια περιγράφεται το μοντέλο εκθετικής εξομάλυνσης σταθερού επιπέδου με πολλαπλασιαστική εποχιακότητα (μοντέλο Winters), το οποίο περιγράφεται από τις ακόλουθες εξισώσεις :

$$e_t = Y_t - F_t \quad (4.15)$$

$$S_t = S_{t-1} + \frac{\alpha \cdot e_t}{I_{t-p}} \quad (4.16)$$

$$I_t = I_{t-p} + \frac{\gamma \cdot e_t}{S_t} \quad (4.17)$$

$$F_{t+m} = S_t \cdot I_{t-p+m} \quad (4.18)$$

Στις παραπάνω εξισώσεις, με I συμβολίζεται ο εξομαλυμένος εποχιακός παράγοντας. Αν ένα έτος αποτελείται από p περιόδους (periods per year), τότε υπάρχουν p εποχιακοί παράγοντες, ένας για κάθε περίοδο t . Ο εποχιακός παράγοντας I_t αναφέρεται στην περίοδο t , ενώ ο εποχιακός παράγοντας I_{t-p} αναφέρεται στον εποχιακό παράγοντα πριν από p περιόδους (ή πριν ένα έτος). Σε σχέση με το μοντέλο σταθερού επιπέδου, η τιμή που υπολογίζεται για το επίπεδο της χρονοσειράς είναι αποεποχικοποιημένη και μεταβάλλεται κατά ένα ποσοστό α του αποεποχικοποιημένου σφάλματος πρόβλεψης. Η πρόβλεψη προκύπτει από το γινόμενο του αποεποχικοποιημένου επιπέδου και του κατάλληλου εποχιακού παράγοντα για την χρονική περίοδο στο μέλλον, στην οποία θα αναφέρεται η πρόβλεψη. Επιπρόσθετα, εισάγεται μια νέα εξίσωση αναφορικά με την εξομάλυνση των εποχιακών παραγόντων, οι οποίοι μεταβάλλονται κατά ένα ποσοστό γ του λόγου του σφάλματος πρόβλεψης προς το αποεποχικοποιημένο επίπεδο για τη συγκεκριμένη χρονική περίοδο με συντελεστή εξομάλυνσης γ που λαμβάνει τιμές στο διάστημα $[0,1]$.

4.2.6.2 Μοντέλο Holt-Winters

Στη συνέχεια περιγράφεται το μοντέλο εκθετικής εξομάλυνσης γραμμικής τάσης με πολλαπλασιαστική εποχιακότητα (μοντέλο Holt-Winters), το οποίο περιγράφεται από τις ακόλουθες εξισώσεις :

$$e_t = Y_t - F_t \quad (4.19)$$

$$S_t = S_{t-1} + T_{t-1} + \frac{\alpha \cdot e_t}{I_{t-p}} \quad (4.20)$$

$$T_t = T_{t-1} + \frac{\beta \cdot e_t}{I_{t-p}} \quad (4.21)$$

$$I_t = I_{t-p} + \frac{\gamma \cdot e_t}{S_t} \quad (4.22)$$

$$F_{t+m} = (S_t + m \cdot T_t) \cdot I_{t-p+m} \quad (4.23)$$

4.2.6.3 Αρχικοποιήσεις και συντελεστές εξομάλυνσης

Όπως και στα λοιπά μοντέλα εκθετικής εξομάλυνσης, απαιτούνται συγκεκριμένες αρχικοποιήσεις αναφορικά με το επίπεδο, την τάση (στην περίπτωση του γραμμικού μοντέλου) και τους p αρχικούς εποχιακούς παράγοντες. Προκειμένου να εκτιμήσουμε τους αρχικούς εποχιακούς παράγοντες, απαιτούνται τουλάχιστον p παρατηρήσεις, οι οποίες αντιστοιχούν σε ένα ολόκληρο έτος. Σε αντίθεση με τα άλλα μοντέλα εξομάλυνσης, θα αρχικοποιήσουμε το επίπεδο και την τάση στην περίοδο p και η διαδικασία υπολογισμού του μοντέλου πρόβλεψης θα ξεκινήσει από την περίοδο $p+1$.

Για τον υπολογισμό του αρχικού επιπέδου θα χρησιμοποιήσουμε τον ίδιο αριθμό παρατηρήσεων p , εξομαλύνοντας με αυτόν τον τρόπο τη συνιστώσα της εποχιακότητας:

$$S_p = \frac{1}{p} \cdot (Y_1 + Y_2 + \dots + Y_p) \quad (4.24)$$

Για την αρχικοποίηση της τάσης θα χρησιμοποιήσουμε δύο ολόκληρες περιόδους, ως εξής :

$$T_p = \frac{1}{p} \cdot \left(\frac{Y_{p+1} - Y_1}{p} + \frac{Y_{p+2} - Y_2}{p} + \dots + \frac{Y_{p+p} - Y_p}{p} \right) \quad (4.25)$$

Οι αρχικές τιμές για τους εποχιακούς παράγοντες προκύπτουν από το λόγο των παρατηρήσεων του πρώτου έτους προς τη μέση τιμή αυτών, ως εξής :

$$I_1 = \frac{Y_1}{S_p}, I_2 = \frac{Y_2}{S_p}, \dots, I_p = \frac{Y_p}{S_p} \quad (4.26)$$

Ένας εναλλακτικός τρόπος αρχικοποίησης θα μπορούσε να είναι η χρήση των συντελεστών επιπέδου και τάσης από το μοντέλο της απλής γραμμικής παλινδρόμησης (ευθεία ελαχίστων τετραγώνων) αναφορικά με το αρχικό επίπεδο και αρχική τάση και η χρήση των εποχιακών συντελεστών οι οποίοι προκύπτουν από την εφαρμογή της κλασικής μεθόδου αποσύνθεσης της χρονοσειράς στις συνιστώσες για τις αρχικές τιμές των εποχιακών παραγόντων. Τέλος, οι συντελεστές εξομάλυνσης για τα μοντέλα της εποχιακής εξομάλυνσης επιλέγονται και πάλι με γνώμονα την ελαχιστοποίηση του σφάλματος MSE που αναφέρεται στη σύγκριση των πραγματικών δεδομένων και του μοντέλου πρόβλεψης.

4.3 Μοντέλα Γραμμικής Παλινδρόμησης

Στον τομέα της στατιστικής, η γραμμική παλινδρόμηση συνίσταται στην προσέγγιση για την μοντελοποίηση της σχέσης ανάμεσα σε μία εξαρτημένη μεταβλητή Y και σε μία ή περισσότερες ανεξάρτητες μεταβλητές που συμβολίζονται με X . Στην περίπτωση που έχουμε μια ανεξάρτητη μεταβλητή, η προσέγγιση αυτή καλείται απλή γραμμική παλινδρόμηση. Για περισσότερες από μία ανεξάρτητες μεταβλητές, η διαδικασία ονομάζεται πολλαπλή γραμμική παλινδρόμηση. Τονίζεται ότι ο όρος «πολλαπλή γραμμική παλινδρόμηση» δεν σχετίζεται με τον όρο «πολυμεταβλητή γραμμική παλινδρόμηση». Ο πρώτος ορισμός έγκειται στην διαδικασία συσχέτισης μιας εξαρτημένης βαθμωτής μεταβλητής Y με περισσότερες της μίας ανεξάρτητες μεταβλητές X_i . Ο δεύτερος ορισμός έγκειται στη συσχέτιση πολλών εξαρτημένων μεταβλητών Y_i με τις ανεξάρτητες μεταβλητές X_i .

Στην γραμμική παλινδρόμηση, οι σχέσεις μοντελοποιούνται με την χρήση γραμμικών συναρτήσεων (μοντέλων) πρόβλεψης, των οποίων οι άγνωστες παράμετροι μοντελοποίησης υπολογίζονται από τα δεδομένα. Αυτά τα μοντέλα καλούνται γραμμικά μοντέλα. Συνήθως, η μέση τιμή της εξαρτημένης μεταβλητής Y , δεδομένης της X (ή των X_i), υποτίθεται να είναι μια συνάρτηση συσχετισμού της X , ενώ πιο σπάνια, η διάμεση τιμή ή κάποια άλλη ποσοστωση της κατανομής της Y υποτίθεται να είναι μια γραμμική συνάρτηση της X .

Όπως συμβαίνει σε όλους τους τύπους παλινδρόμησης, η γραμμική παλινδρόμηση εστιάζει στην κατανομή πιθανότητας y , δεδομένης της X , και όχι στην απο κοινού κατανομή πιθανότητας των y και X , αντικείμενο της ανάλυσης πολλών μεταβλητών. Η γραμμική παλινδρόμηση ήταν ο πρώτος τύπος της ανάλυσης παλινδρόμησης που μελετήθηκε διεξοδικά και εφαρμόστηκε εκτενώς σε πρακτικές εφαρμογές. Αυτό οφείλεται στο γεγονός ότι τα μοντέλα που εμφανίζουν γραμμική εξάρτηση από τις άγνωστες παραμέτρους τους παρουσιάζουν καλύτερη εφαρμογή από μοντέλα τα οποία εμφανίζουν μη γραμμική εξάρτηση από τις παραμέτρους τους, καθώς επίσης και στο γεγονός ότι είναι ευκολότερος ο προσδιορισμός των στατιστικών ιδιοτήτων των προκυπτουσών εκτιμητριών.

Η γραμμική παλινδρόμηση έχει πολλές πρακτικές εφαρμογές. Οι περισσότερες εφαρμογές εμπίπτουν σε μία από τις ακόλουθες δύο γενικές κατηγορίες:

Στην περίπτωση που στοχεύουμε στην παραγωγή προβλέψεων ή στην μείωση του σφάλματος πρόβλεψης, μπορούμε να χρησιμοποιήσουμε γραμμική παλινδρόμηση με σκοπό να κατασκευάσουμε ένα μοντέλο πρόβλεψης ικανό να εφαρμοστεί στα σύνολα παρατηρήσεων των y και X . Αφού αναπτύξουμε το προαναφερθέν μοντέλο, δεδομένης κάποιας επιπρόσθετης τιμής της ανεξάρτητης μεταβλητής X μπορούμε να προβλέψουμε την τιμή της εξαρτημένης μεταβλητής y .

Δεδομένης μια μεταβλητής y κι ενός πλήθους πιθανών μεταβλητών συσχέτισης X_1, X_2, \dots, X_p μπορούμε να εφαρμόσουμε γραμμική παλινδρόμηση προκειμένου να προσδιορίσουμε το βαθμό συσχετισμού μεταξύ των y και X_i καθώς και κατά πόσο κάποια υποσύνολα των X_i περιέχουν περιττες πληροφορίες σχετικά με την y .

4.3.1 Απλή Γραμμική Παλινδρόμηση

Η παρακάτω εξίσωση περιγράφει τη σχέση της απλής γραμμικής παλινδρόμησης :

$$Y=a+b \cdot X \quad (4.27)$$

Με a συμβολίζεται η τεταγμένη του σημείου τομής της ευθείας με τον άξονα των εξαρτημένων μεταβλητών, ενώ με b συμβολίζεται η κλίση της ευθείας. Η ανεξάρτητη μεταβλητή συμβολίζεται με X (στην περίπτωση πρόβλεψης χρονοσειρών είθισται να είναι ο χρόνος), ενώ η εξαρτημένη μεταβλητή (δηλαδή η τιμή πρόβλεψης) συμβολίζεται με Y . Οι τιμές των συντελεστών a και b υπολογίζονται με βάση την αρχή των ελαχίστων τετραγώνων, επιλέγονται δηλαδή οι συντελεστές που ελαχιστοποιούν το άθροισμα των τετραγώνων των διαφορών των πραγματικών τιμών Y_i από τις προβλεπόμενες σε κάθε χρονική περίοδο \hat{Y}_i , όπως φαίνεται και στην επόμενη σχέση:

$$(a,b) \mid \min \left[\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \right] \quad (4.28)$$

Με βάση τη λογική της ελαχιστοποίησης της απόστασης των πραγματικών παρατηρήσεων Y_i από τη βέλτιστη γραμμή παλινδρόμησης προκύπτουν οι εξισώσεις υπολογισμού των συντελεστών a και b , όπως παρουσιάζονται στη συνέχεια :

$$b = \frac{\frac{\sum_{i=1}^n Y_i X_i}{n} - \bar{X}\bar{Y}}{\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2} = \frac{\sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (4.29) \quad \text{και}$$

$$a = \bar{Y} - b \cdot \bar{X} \quad (4.30)$$

Με \bar{X} και \bar{Y} συμβολίζονται οι μέσες τιμές των μεταβλητών X και Y αντίστοιχα. Ο αριθμός των παρατηρήσεων, βάσει των οποίων υπολογίζεται η ευθεία παλινδρόμησης, συμβολίζεται με n .

Η κλίση της ευθείας παλινδρόμησης ισούται με το γινόμενο της συσχέτισης των μεταβλητών X και Y και του λόγου των τυπικών αποκλίσεων σ_Y και σ_X , σύμφωνα με την παρακάτω σχέση :

$$b = r_{XY} \cdot \frac{\sigma_Y}{\sigma_X} \quad (4.31)$$

Η πρόβλεψη με χρήση της μεθόδου της απλής γραμμικής παλινδρόμησης δίνει μια καλή εικόνα της μέσης και της μακροπρόθεσμης συμπεριφοράς του υπό μελέτη μεγέθους.

Σε περίπτωση που η σχέση ανάμεσα σε δύο μεταβλητές, την εξαρτημένη και την ανεξάρτητη, δεν είναι γραμμική, μπορεί και πάλι να εφαρμοστεί η μέθοδος της απλής γραμμικής παλινδρόμησης, αφού πρώτα γίνει μετασχηματισμός της σχέσης των δύο μεταβλητών σε γραμμική.

4.3.2 Ο Συντελεστής γραμμικής συσχέτισης

Ακόμη εφόσον η απλή γραμμική παλινδρόμηση στηρίζεται στην ύπαρξη σχέσης ανάμεσα στις δύο μεταβλητές, μας ενδιαφέρει ο προσδιορισμός του βαθμού συσχέτισής τους. Αυτό το πετυχαίνουμε με τη χρήση του συντελεστή γραμμικής συσχέτισης των δύο μεταβλητών, ο οποίος λαμβάνει τιμές στο διάστημα $[-1, 1]$.

- Για μηδενική τιμή δεν υπάρχει καμία σχέση ανάμεσα στις δύο μεταβλητές.
- Για τιμή ± 1 υπάρχει απόλυτη συσχέτιση ανάμεσα στις δύο μεταβλητές.

Γενικά για θετικές τιμές του δείκτη οι μεταβλητές αυξάνονται ή μειώνονται συγχρόνως, ενώ για αρνητικές τιμές του δείκτη συσχέτισης, οι αυξομειώσεις τους είναι αντίθετης κατεύθυνσης.

4.3.3 Ο συντελεστής R^2 (coefficient of determination)

Η συσχέτιση των τιμών \hat{Y} , οι οποίες προκύπτουν από την εξίσωση της ευθείας παλινδρόμησης και των πραγματικών τιμών Y συμβολίζεται με R . Στην πράξη, αυτή η συσχέτιση χρησιμοποιείται στην τετραγωνική της μορφή, συνεπώς προκύπτει ο συντελεστής R^2 . Λόγω της τετραγωνικής του μορφής, ο συντελεστής αυτός είναι πάντα θετικός και παίρνει τιμές στο διάστημα $[0,1]$. Αντιπροσωπεύει το ποσοστό της διακύμανσης της μεταβλητής Y , η οποία ερμηνεύεται από την ευθεία παλινδρόμησης. Με άλλα λόγια, ο συντελεστής R^2 ορίζεται ως ο λόγος της διακύμανσης των τιμών \hat{Y} προς τη διακύμανση των τιμών Y , δηλαδή:

$$R^2 = \frac{\text{διακύμανση των τιμών } \hat{Y}}{\text{διακύμανση των τιμών } Y} \quad (4.32)$$

Λαμβάνοντας υπ' όψιν το γεγονός ότι οι τιμές \hat{Y} προκύπτουν από την εξίσωση της ευθείας παλινδρόμησης, ο συντελεστής R^2 μπορεί, επίσης, να εκφραστεί ως εξής:

$$R^2 = \frac{\text{ερμηνευθείσα διακύμανση των τιμών } Y}{\text{συνολική διακύμανση των τιμών } Y} \quad (4.33)$$

, δηλαδή:

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = r^2_{XY} \quad (4.34)$$

Για κάθε τιμή Y_i υπάρχει μια συνολική διακύμανση $(Y_i - \bar{Y})$, η οποία αντιπροσωπεύει το πόσο απέχει η τιμή Y_i από το μέσο όρο των τιμών Y . Η συνολική αυτή διακύμανση μπορεί να διαμεριστεί σε δύο τμήματα :

Τη μη ερμηνευθείσα διακύμανση $(Y_i - \hat{Y}_i)$, η οποία αντιπροσωπεύει το πόσο απέχει η τιμή \hat{Y}_i της εξίσωσης παλινδρόμησης από την πραγματική τιμή Y_i .

Την ερμηνευθείσα διακύμανση $(\hat{Y}_i - \bar{Y})$, η οποία αντιπροσωπεύει το πόσο απέχει η τιμή \hat{Y}_i της εξίσωσης παλινδρόμησης από το μέσο όρο των τιμών Y .

Ισχύει δηλαδή η σχέση :

$$(Y_i - \bar{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) \quad (4.35)$$

(συνολική διακύμανση) = (μη ερμηνευθείσα διακύμανση) + (ερμηνευθείσα διακύμανση)

4.3.4 Η εξίσωση παλινδρόμησης ως μοντέλο

Η εξίσωση παλινδρόμησης $\hat{Y}_i = a + b \cdot X_i$, όπως και κάθε άλλη μορφή εξίσωσης, μπορεί να θεωρηθεί ως ένα αφηρημένο μοντέλο, το οποίο αναπαριστά κάποια εκδοχή της πραγματικότητας. Το μοντέλο αυτό απλοποιεί την πραγματικότητα και την αναπαριστά με βάση την αλληλεπίδραση δύο μονο παραγόντων. Ωστόσο, η πραγματικότητα είναι πιο περίπλοκη. Η εξαρτημένη προβλεπόμενη μεταβλητή μπορεί να επηρεάζεται από περισσότερους παράγοντες. Σε κάθε προσπάθεια μοντελοποίησης υπάρχουν οι εξής δύο επιλογές. Η πρώτη έγκειται στην κατασκευή ενός απλού μοντέλου το οποίο ενδέχεται να μην αντιπροσωπεύει τέλεια την πραγματικότητα ενώ η δεύτερη έγκειται στην κατασκευή ενός πολύπλοκου μοντέλου, το οποίο είναι πιο αξιόπιστο αλλά συγχρόνως παρουσιάζει δυσκολίες στην ανάπτυξη και τον χειρισμό του. Ακόμη και αν μπορούσε να αναπτυχθεί το πιο περίπλοκο μοντέλο, θα υπήρχε πάλι κάποιο κομμάτι της πραγματικότητας το οποίο δεν θα αντιπροσωπευόταν σε αυτό. Ο αριθμός των παραγόντων στα πραγματικά φαινόμενα είναι άπειρος. Παράλληλα, είναι ιδιαίτερα επίπονη η συλλογή δεδομένων για όλους τους παράγοντες οι οποίοι μπορεί να επηρεάζουν το προβλεπόμενο μέγεθος.

Προκειμένου να ληφθεί υπ' όψιν το γεγονός ότι ένα μέρος της πραγματικής διαδικασίας δεν μπορεί να ερμηνευθεί από το μοντέλο παλινδρόμησης, χρησιμοποιείται ο όρος e για να δηλώσει τις διακυμάνσεις (αποκλίσεις), οι οποίες δεν ερμηνεύονται από το μοντέλο. Συνεπώς, η εξίσωση παλινδρόμησης διαμορφώνεται ως εξής: $\hat{Y}_i = a + b \cdot X_i + e$

Ο όρος e σπάνια χρειάζεται να υπολογιστεί. Η θεωρητική του σημασία είναι ότι πρόβλεψη μπορεί να διαφέρει από την υπολογισμένη τιμή $\hat{Y}_i = a + b \cdot X_i$ κατά μια ποσότητα e (ουσιαστικά δείχνει το σφάλμα της πρόβλεψης), η οποία μπορεί να εκτιμηθεί βάσει της θεωρίας των πιθανοτήτων. Είναι προφανές, ότι το σφάλμα αυτό διαφέρει από μοντέλο σε μοντέλο. Θεωρητικά, όσο περισσότερες μεταβλητές εισάγονται στο μοντέλο, τόσο μικρότερο το διάστημα των τιμών οι οποίες περιλαμβάνονται στο e . Στην πράξη, υπάρχει κάποιο όριο στον αριθμό των μεταβλητών, οι οποίες θα χρησιμοποιηθούν, εφ' όσον αυτές εισάγουν μεγαλύτερη πολυπλοκότητα και υψηλότερο κόστος εφαρμογής. Επομένως, η βέλτιστη λύση θα ήταν η εισαγωγή του μικρότερου δυνατού αριθμού μεταβλητών για τις οποίες το διάστημα τιμών του e είναι το ελάχιστο δυνατό.

Αναφορικά με την εξίσωση παλινδρόμησης $\hat{Y}_i = a + b \cdot X_i + e$, το e θα πρέπει να πληροί τις ακόλουθες συνθήκες:

- Η μέση τιμή του e πρέπει να ισούται με μηδέν. Παράγοντες, οι οποίοι δεν επηρεάζουν το Y δεν περιλαμβάνονται στην εξίσωση παλινδρόμησης. Ωστόσο, οι επιδράσεις τους πρέπει να είναι αντίθετων κατευθύνσεων, ούτως ώστε να τείνουν να εξουδετερώσουν η μία την άλλη.
- Ο όρος σφάλματος e πρέπει να είναι τυχαία μεταβλητή. Σε κάποια συγκεκριμένη χρονική περίοδο, είναι πιθανόν ορισμένοι παράγοντες, οι οποίοι δεν περιλαμβάνονται στην εξίσωση, να επηρεάσουν το προβλεπόμενο μέγεθος Y περισσότερο από άλλους. Εφ' όσον, όμως, οι τιμές e που αντιστοιχούν σε κάθε περίοδο είναι τυχαίες και όχι το αποτέλεσμα κάποιου συστηματικού προτύπου, η επίδραση τους στην υπολογισμένη τιμή του Y μπορεί να καθοριστεί με βάση την

θεωρία πιθανοτήτων. Θα πρέπει, γενικά, τα σφάλματα να είναι ανεξάρτητα το ένα από το άλλο.

- Ο όρος των σφαλμάτων e θα πρέπει να ακολουθεί την κανονική κατανομή. Αυτό είναι συνέπεια του μεγάλου πλήθους παραγόντων που επηρεάζουν το Y και δεν περιλαμβάνονται στην εξίσωση παλινδρόμησης. Σε αυτήν την περίπτωση, είναι περισσότερο πιθανόν ότι οι ακραίες αποκλίσεις θα αλληλοεξουδετερώνονται και θα εμφανίζονται πολύ σπάνια, ενώ στην πλειοψηφία των περιπτώσεων, τα σφάλματα θα συγκεντρώνονται γύρω από τη μέση τιμή. Ένα τέτοιο πρότυπο οδηγεί σε κανονική κατανομή.

4.3.5 Πολλαπλή Γραμμική Παλινδρόμηση

Όπως αναφέρθηκε προηγουμένως, σε περιπτώσεις που απαιτούνται περισσότερες από μια ανεξάρτητες μεταβλητές, το μοντέλο γραμμικής παλινδρόμησης μπορεί να γενικευθεί μέσω της τεχνικής της πολλαπλής παλινδρόμησης προκειμένου να συμπεριληφθούν σε αυτό όλες οι μεταβλητές που επηρεάζουν την τιμή της μεταβλητής πρόβλεψης.

Στην πολλαπλή παλινδρόμηση υπάρχει εξαρτημένη μεταβλητή, της οποίας η τιμή πρέπει να προβλεφθεί βάσει των τιμών δύο ή περισσότερων ανεξάρτητων μεταβλητών. Έτσι, η γενική μορφή της πολλαπλής παλινδρόμησης είναι:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e \quad (4.36)$$

Στην παραπάνω εξίσωση, οι μεταβλητές X_1, X_2, \dots, X_k εκφράζουν τις ανεξάρτητες μεταβλητές ενώ η μεταβλητή Y εκφράζει την εξαρτημένη μεταβλητή. Οι συντελεστές $b_0, b_1, b_2, \dots, b_k$ είναι σταθερές παράμετροι ενώ το e δηλώνει τον τυχαίο παράγοντα, ο οποίος θεωρείται κανονικά κατανεμημένος γύρω από το μηδέν.

Η εξίσωση της πολλαπλής γραμμικής παλινδρόμησης είναι γραμμική ως προς τους συντελεστές. Κάθε συντελεστής b_i έχει εκθέτη ίσο με τη μονάδα, γεγονός που εξασφαλίζει την γραμμικότητα. Οι τιμές των συντελεστών αυτών μπορούν να προκύψουν με εφαρμογή της μεθόδου ελαχίστων τετραγώνων. Το σχήμα της συνάρτησης που συνδέει τις ανεξάρτητες με τις εξαρτημένες μεταβλητές δεν είναι εύκολο να περιγραφεί. Στην περίπτωση μιας ανεξάρτητης μεταβλητής, το σχήμα της συνάρτησης είναι μια ευθεία γραμμή. Στην περίπτωση δύο ανεξάρτητων μεταβλητών, η Y παριστάνεται στο επίπεδο που σχηματίζουν οι δύο ανεξάρτητες μεταβλητές. Στην περίπτωση περισσότερων από δύο μεταβλητές, η Y παριστάνεται σε υπερεπίπεδο (επιφάνεια με περισσότερες από δύο διαστάσεις).

Στην πράξη, η διαδικασία της πολλαπλής παλινδρόμησης αποσκοπεί στον προσδιορισμό των αγνώστων παραμέτρων συντελεστές $b_0, b_1, b_2, \dots, b_k$ του μοντέλου και της διακύμανσης του τυχαίου παράγοντα, δεδομένου ενός συγκεκριμένου συνόλου δεδομένων, στο οποίο μπορεί να εφαρμοστεί η μέθοδος ελαχίστων τετραγώνων. Συνεπώς η μορφή του στατιστικού μοντέλου παλινδρόμησης είναι :

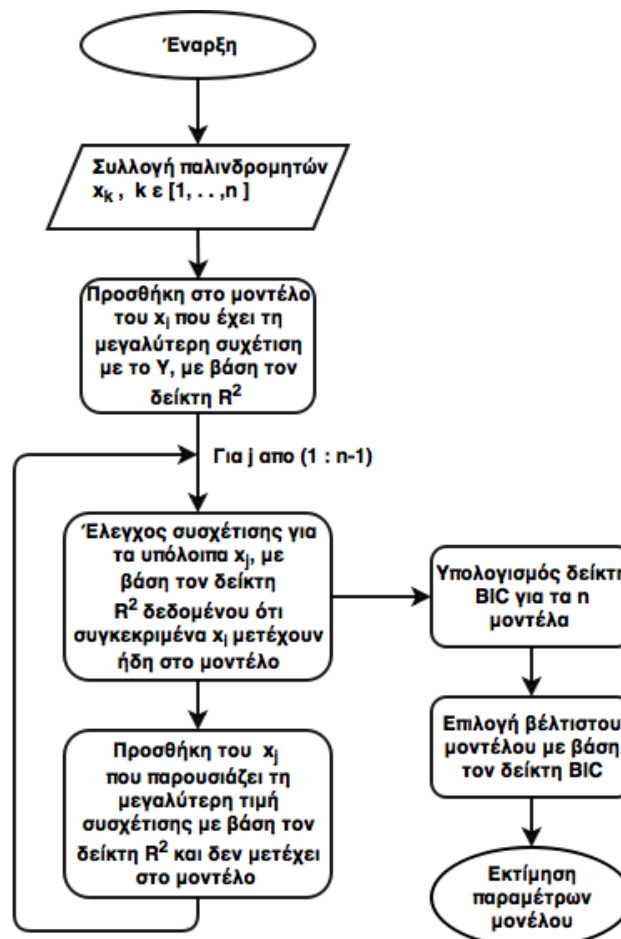
$$Y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_kX_{k,i} + e_i \quad (4.37)$$

4.3.6 Διαδικασία επιλογής ανεξάρτητων μεταβλητών

Υπάρχουν αρκετές μεθοδολογίες, οι οποίες προσεγγίζουν με βηματικό τρόπο την ανάπτυξη ενός μοντέλου παλινδρόμησης, έτσι ώστε να μην είναι απαραίτητος ο εξαντλητικός έλεγχος όλων των διαθέσιμων δεδομένων. Τέτοιου είδους μεθοδολογίες είναι και αυτές που ανήκουν στην οικογένεια της κλασσικής βηματικής ανάλυσης παλινδρόμησης (stepwise regression analysis). Πιο αναλυτικά, υπάρχουν και χρησιμοποιούνται σχετικά συχνά σε διάφορες εφαρμογές, η προς τα εμπρός επιλογή (forward selection), η προς τα πίσω απαλοιφή (backward elimination) και η βηματική παλινδρόμηση (stepwise regression).

4.3.6.1 Forward selection

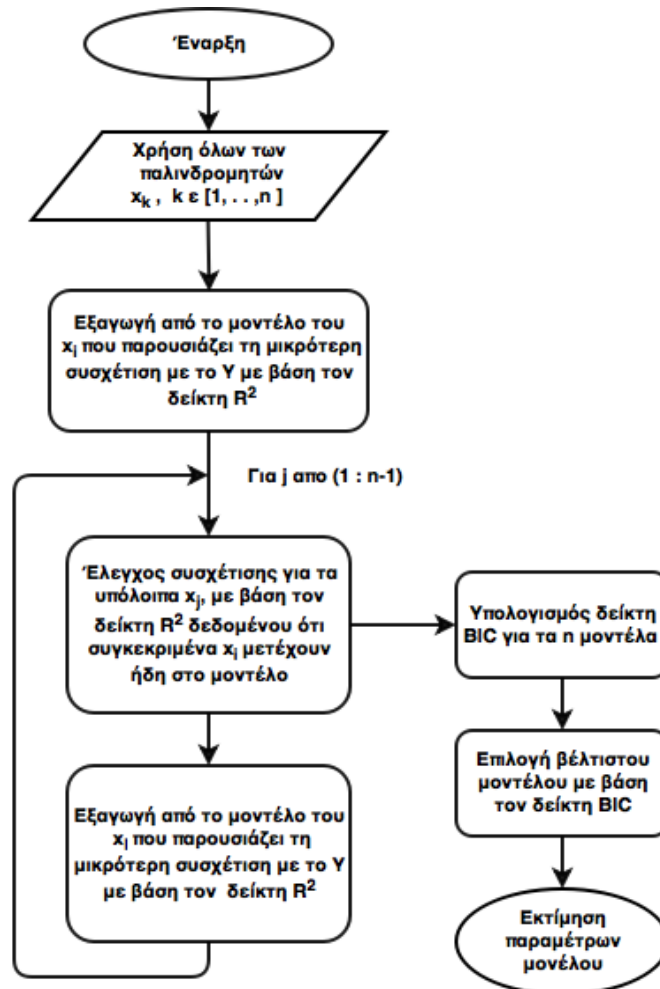
Η προς τα εμπρός επιλογή ξεκινά από τη μη συμμετοχή καμιάς ανεξάρτητης μεταβλητής στο μοντέλο παλινδρόμησης και βήμα-βήμα εξετάζεται η συμμετοχή ή όχι κάποιας από τις διαθέσιμες μεταβλητές σε αυτό. Η εισαγωγή αυτή στηρίζεται, σε κάθε βήμα (εκτός του αρχικού), στο γεγονός ότι είναι γνωστό ότι ήδη κάποιες συγκεκριμένες μεταβλητές συμμετέχουν σε αυτό και η συμμετοχή μιας άλλης πραγματοποιείται μέσω στατιστικών υποθέσεων, οι οποίες επιβεβαιώνονται ή όχι με βάση ένα προκαθορισμένο κατώφλι αποδοχής (π.χ. χρησιμοποίηση του δείκτη R^2 ή της στατιστικής F).



Σχήμα 4.2 : Διάγραμμα ροής μεθόδου «Forward Selection»

4.3.6.2 Backward selection

Στην περίπτωση της απαλοιφής προς τα πίσω, λειτουργούμε με την αντίθετη λογική της προς τα εμπρός επιλογής. Η διαδικασία ξεκινά με την υπόθεση ότι όλες οι διαθέσιμες μεταβλητές συμμετέχουν στο μοντέλο παλινδρόμησης, οπότε σε κάθε επόμενο βήμα, εξετάζεται, πάλι με στατιστικές υποθέσεις και ένα προκαθορισμένο κατώφλι απόρριψης, η εξαγωγή μιας μεταβλητής από το μοντέλο.



Σχήμα 4.3 : Διάγραμμα ροής μεθόδου «Backward Selection»

4.3.7 Επιλογή βέλτιστου μοντέλου

Έχοντας σχηματίσει n μοντέλα παλινδρόμησης με διαφορετικό πλήθος παλινδρομητών (που επιλέγονται σύμφωνα με κάποια από τις διαδικασίες επιλογής παλινδρομητών που περιγράφηκαν στην προηγούμενη ενότητα), χρειαζόμαστε κάποιον δείκτη σχετικής ποιότητας των n στατιστικών μοντέλων για ένα δοσμένο σύνολο δεδομένων, ο οποίος θα μας οδηγήσει στην επιλογή του βέλτιστου μοντέλου παλινδρόμησης. Η σχετική ποιότητα των στατιστικών μοντέλων δίνεται με χρήση κάποιου κριτηρίου πληροφορίας. Ακολουθεί η αναλυτική περιγραφή δύο τέτοιων κριτηρίων : του κριτηρίου πληροφορίας Akaike (AIC) και του κριτηρίου πληροφορίας Bayes (BIC).

4.3.7.1 Κριτήριο Πληροφορίας AIC (Akaike Information Criterion)

Ο δείκτης AIC αποτελεί ένα δείκτη σχετικής ποιότητας των στατιστικών μοντέλων για ένα δοσμένο σύνολο πληροφοριών. Για δεδομένη λίστα μοντέλων για τα δεδομένα, ο δείκτης αυτός εκτιμά την απόδοση κάθε μοντέλου σχετικά με κάθε άλλο μοντέλο.

Βασίζεται στη θεωρία δεδομένων και προσφέρει μια σχετική εκτίμηση των χαμένων πληροφοριών όταν χρησιμοποιείται ένα συγκεκριμένο δοσμένο μοντέλο για την αναπαράσταση της διαδικασίας παραγωγής δεδομένων. Συμβιβάζει την καλή δυνατότητα εφαρμογής ενός μοντέλου και την πολυπλοκότητα του. Ως σχετικός δείκτης εξετάζει τη σχετική και όχι την απόλυτη απόδοση των μοντέλων. Δίνεται από την ακόλουθη σχέση :

$$AIC = 2k - 2\ln(L) \quad (4.38)$$

,όπου:

L : μέγιστη τιμή συνάρτησης πιθανότητας και

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

Όσο μεγαλύτερη είναι η τιμή του δείκτη AIC, τόσο καλύτερη σχετική απόδοση παρουσιάζει το μοντέλο. Συνεπώς, ο AIC :

- ο «επιβραβεύει» την καλή εφαρμογή του μοντέλου (σύμφωνα με τη συνάρτηση πιθανότητας)
- ο περιλαμβάνει «ποινή», η οποία είναι αύξουσα συνάρτηση του αριθμού των εκτιμώμενων παραμέτρων

Η ποινή αυτή είναι ανασταλτικός παράγοντας για το overfitting. Σημειώνεται ότι, όσο αυξάνεται ο αριθμός των παραμέτρων στο μοντέλο, σχεδόν πάντα βελτιώνεται η δυνατότητα εφαρμογής του μοντέλου.

Έστω ότι μια άγνωστη διαδικασία f παράγει δεδομένα κι έστω ότι έχουμε στη διάθεση μας δύο υποψήφια μοντέλα g_1 και g_2 . Εάν γνωρίζαμε την f , θα μπορούσαμε να διαλέξουμε το μοντέλο που θα εμφάνιζε τις ελάχιστες χαμένες πληροφορίες (ελάχιστη απόκλιση), έχοντας πρώτα υπολογίσει τις αποκλίσεις Kullback-Leibler μέσω του τύπου :

$$D_{KL_i} = \int_{-\infty}^{\infty} \log \frac{f(x)}{g_i(x)} dx \quad (4.39)$$

Μη γνωρίζοντας την f , δεν είναι δυνατόν να στηριχτεί με βεβαιότητα ποιο μοντέλο είναι καταλληλότερο. Με βάση τον δείκτη AIC, ωστόσο, πόσο περισσότερη ή λιγότερη πληροφορία χάνεται μεταξύ των g_1 και g_2 .

4.3.7.1.1 Εφαρμογή AIC

Έστω R υποψήφια μοντέλα, τα οποία ελέγχονται πάνω σε συγκεκριμένα δεδομένα ως προς την ελάχιστη απώλεια πληροφοριών. Έστω $AIC_1, AIC_2, \dots, AIC_R$ οι τιμές του δείκτη για τα μοντέλα και AIC_{min} η ελάχιστη τιμή του δείκτη που εμφανίζεται. Τότε, ο όρος $e^{\frac{AIC_{min} - AIC_i}{2}}$

αποτυπώνει τη σχετική πιθανότητα το i -μοντέλο να ελαχιστοποιεί την απώλεια πληροφοριών.

Αν το μοντέλο είναι γραμμικό και τα σφάλματα ακολουθούν κανονική κατανομή, ισχύει :

$$AIC_L = AIC + \frac{2k(k+1)}{n-k-1} \quad (4.40)$$

,όπου:

L : μέγιστη τιμή συνάρτησης πιθανότητας και

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

Με βάση τον παραπάνω τύπο, γίνεται αντιληπτό, ότι όταν το n δεν είναι πολλές φορές μεγαλύτερο του k² και για κάθε επιπλέον παράμετρο, αυξάνεται η πιθανότητα να επιλεχθεί μοντέλο που έχει πάρα πολλές παραμέτρους (overfitting). Υποθέτωντας ότι τα σφάλματα ακολουθούν κανονική κατανομή, προκύπτει :

$$AIC = n \cdot \ln(\text{MSE}) + 2 \cdot k \quad (4.41)$$

4.3.7.2 Κριτήριο Πληροφορίας BIC (Bayesian Information Criterion)

Ο δείκτης BIC αποτελεί ένα δείκτη επιλογής μοντέλου για ένα δοσμένο σύνολο πληροφοριών. Βασίζεται στη θεωρία της συνάρτησης πιθανότητας. Όσο μικρότερος υπολογιστεί ο δείκτης για ένα μοντέλο, τόσο καταλληλότερο θεωρείται το μοντέλο. Η βασική διαφορά του συγκεκριμένου δείκτη με τον δείκτη AIC είναι ότι εισάγει μεγαλύτερη ποινή για κάθε επιπλέον παράμετρο που λαμβάνεται υπόψιν στο μοντέλο προβλεψής. Υπολογίζεται με βάση τον ακόλουθο τύπο:

$$BIC = -2 \cdot \ln(\hat{L}) + k \cdot \ln(n) \quad (4.42)$$

,όπου:

L : μέγιστη τιμή συνάρτησης πιθανότητας και

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

4.3.7.2.1 Εφαρμογή BIC

Το ολοκλήρωμα της συνάρτησης πιθανότητας $P(x|\theta, m)$ μετρά την εκ των προτέρων κατανομή πιθανότητας $P(\theta|m)$ επί των παραμέτρων θ του μοντέλου m για δεδομένο x και υπολογίζεται ως εξής :

$$BIC = -2 \cdot \ln(P(x|m)) \approx 2 \cdot \ln(\hat{L}) + k \cdot (\ln(n) - 2\ln(2\pi)) \quad (4.43)$$

,όπου:

x : παρατηρήσεις

θ : παράμετροι μοντέλου

n : μέγεθος δείγματος

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

L : μέγιστη τιμή συνάρτησης πιθανοφάνειας μοντέλου $\hat{L} = P(x|\hat{\theta}, m)$ και

$\hat{\theta}$: τιμές παραμέτρων που μεγιστοποιούν τη συνάρτηση πιθανότητας

Περιορισμοί :

- Θα πρέπει $n \gg k$
- Ο δείκτης BIC δεν μπορεί να διαχειριστεί σύνθετες και μεγάλες συλλογές μοντέλων

Υποθέτοντας ότι τα σφάλματα ακολουθούν κανονική κατανομή, προκύπτει :

$$BIC = n \cdot \ln(MSE) + k \cdot \ln(n) \quad (4.44)$$

4.3.8 Υπολογισμός συντελεστών παλινδρόμησης

Έστω ότι το μοντέλο παλινδρόμησης περιέχει k ανεξάρτητες μεταβλητές. Τότε:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e \quad (4.45)$$

Για κάθε διαφορετικό διάνυσμα παρατηρήσεων χρησιμοποιείται μια ξεχωριστή τιμή του δείκτη i , όπως παρατηρείται στην παρακάτω εξίσωση :

$$Y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_kX_{k,i} + e_i = \hat{Y}_i + e_i \quad (4.46)$$

,όπου \hat{Y}_i είναι μια εκτίμηση της τιμής της μεταβλητής Y , η οποία βασίζεται στις τιμές των X_1, X_2, \dots, X_k . Επομένως, το σφάλμα e_i ισούται με $e_i = Y_i - \hat{Y}_i$.

Εφαρμόζοντας τη μέθοδο ελαχίστων τετραγώνων, υπολογίζουμε το ελάχιστο άθροισμα των τετραγώνων των σφαλμάτων e_i , δηλαδή :

$$(b_0, b_1, \dots, b_k) \mid \min [\sum_{i=1}^n e_i^2] \quad (4.47)$$

Όμως,

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - b_0 - b_1X_{1,i} - b_2X_{2,i} - \dots - b_kX_{k,i})^2 \quad (4.48)$$

Προκειμένου να προσδιορίσουμε τους άγνωστους συντελεστές b_0, b_1, \dots, b_k , οι οποίοι ελαχιστοποιούν την παραπάνω ποσότητα, υπολογίζουμε τις μερικές παραγώγους αυτής για κάθε έναν από τους συντελεστές και θέτοντας αυτές ίσες με το μηδέν, λύνουμε ένα γραμμικό σύστημα k εξισώσεων με k αγνώστους.

4.3.9 Πολλαπλή συσχέτιση και ο συντελεστής R^2

Η συσχέτιση ανάμεσα στην πραγματική τιμή της μεταβλητής Y και στην υπολογισμένη τιμή \hat{Y} με βάση την εξίσωση παλινδρόμησης δίνεται από την εξίσωση :

$$R_{Y\hat{Y}} = \frac{n \cdot \sum_{i=1}^n (Y_i \cdot \hat{Y}_i) - (\sum_{i=1}^n Y_i) \cdot (\sum_{i=1}^n \hat{Y}_i)}{\sqrt{n \cdot \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2} \cdot \sqrt{n \cdot \sum_{i=1}^n \hat{Y}_i^2 - (\sum_{i=1}^n \hat{Y}_i)^2}} \quad (4.49)$$

Το τετράγωνο του $R_{Y\hat{Y}}$ καλείται coefficient of determination. Το $R_{Y\hat{Y}}$ είναι γνωστό ως συντελεστής πολλαπλής συσχέτισης και εκφράζει τη συσχέτιση ανάμεσα στην εξαρτημένη μεταβλητή Y και την εκτίμηση της \hat{Y} με βάση τις ανεξάρτητες μεταβλητές. Για τον υπολογισμό του R^2 χρησιμοποιείται η ίδια εξίσωση που χρησιμοποιείται και στην περίπτωση της απλής γραμμικής παλινδρόμησης :

$$R^2 = \frac{\text{ερμηνευθείσα διακύμανση των τιμών } Y}{\text{συνολική διακύμανση των τιμών } Y} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (4.50)$$

Ωστόσο, στην προηγούμενη εξίσωση δε λαμβάνεται υπ' όψιν ο αριθμός των ανεξάρτητων μεταβλητών και ο αριθμός του συνόλου των παρατηρήσεων. Προκειμένου να ξεπεραστεί αυτό το πρόβλημα, υπολογίζεται ένας «διορθωμένος» συντελεστής R^2 από την εξίσωση:

$$\hat{R}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1} \quad (4.51)$$

Ο νέος διορθωμένος συντελεστής \hat{R}^2 εκφράζει το ποσοστό της διασποράς της μεταβλητής Y που ερμηνεύεται από τις ανεξάρτητες μεταβλητές X_1, X_2, \dots, X_k . Η διαφορά $(n-1)$ εκφράζει τους συνολικούς βαθμούς ελευθερίας της συνολικής διακύμανσης του μοντέλου ενώ ο όρος $(n-k-1)$ εκφράζει τους βαθμούς ελευθερίας της ερμηνευθείσας διακύμανσης.

4.3.10 Ο στατιστικός δείκτης F (F-test)

Ο στατιστικός δείκτης F αποτελεί ένα μέτρο της σημαντικότητας του μοντέλου παλινδρόμησης και υπολογίζεται από αντίστοιχες, όπως στην απλή παλινδρόμηση, εξισώσεις :

$$F = \frac{\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{k}}{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}} \quad (4.52)$$

Αν η μη ερμηνευθείσα διακύμανση (διακύμανση σφαλμάτων) είναι μεγάλη, τότε αντίστοιχα και ο παρονομαστής της παραπάνω εξίσωσης είναι μεγάλος και ο δείκτης F γίνεται μικρότερος, γεγονός που σημαίνει ότι το μοντέλο παλινδρόμησης δεν είναι επιτυχημένο. Αντίθετα, αν η ερμηνευθείσα διακύμανση, που εμφανίζεται στον αριθμητή, είναι σχετικά μεγαλύτερη, τότε και ο δείκτης F είναι μεγαλύτερος.

Όπως φαίνεται και στο δεύτερο σκέλος της παραπάνω εξίσωσης υπάρχει στενή σχέση ανάμεσα στο συντελεστή R^2 και στο στατιστικό δείκτη F.

4.3.11 Οι στατιστικοί δείκτες t (t-test) για τους συντελεστές

Αφού πρώτα ελεγχθεί η συνολική σημαντικότητα του μοντέλου παλινδρόμησης, είναι συχνά χρήσιμο να εξεταστεί η σημαντικότητα καθενός από τους συντελεστές παλινδρόμησης. Στην περίπτωση της πολλαπλής παλινδρόμησης, ο στατιστικός δείκτης t για κάθε συντελεστή αποτελεί εκτίμηση της σημαντικότητας του συντελεστή αυτού με την παρουσία όλων των άλλων ανεξάρτητων μεταβλητών. Για κάθε συντελεστή παλινδρόμησης b_j , μπορεί να οριστεί ένα τυπικό σφάλμα ως μέτρο της σταθερότητας του συντελεστή και με βάση την υπόθεση της κανονικότητας του μοντέλου παλινδρόμησης, ο δείκτης t ακολουθεί την t-κατανομή με $(n-k-1)$ βαθμούς ελευθερίας και δίνεται από τον παρακάτω τύπο :

$$t_{b_j} = \frac{b_j}{SE_{b_j}} \quad (4.53)$$

Υπολογίζοντας τον δείκτη t για κάθε συντελεστή του μοντέλου παλινδρόμησης, υπολογίζεται η σημαντικότητά του, μέσω της σύγκρισης της τιμής του συντελεστή αυτού με το μηδέν, τιμή για την οποία η αντίστοιχη ανεξάρτητη μεταβλητή δε συνεισφέρει στην

πρόβλεψη της εξατημένης μεταβλητής Y , με δεδομένη την παρουσία των άλλων ανεξάρτητων μεταβλητών.

Στο σημείο αυτό αξίζει να σημειωθούν δύο βασικά θέματα σχετικά με τους στατιστικούς δείκτες των συντελεστών παλινδρόμησης. Πρώτον, η σταθερότητα των συντελεστών παλινδρόμησης εξαρτάται από τη συσχέτιση των ανεξάρτητων μεταβλητών. Για δύο ανεξάρτητες μεταβλητές X_1 και X_2 , όσο μεγαλύτερη είναι η μεταξύ τους συσχέτιση τόσο πιο ασταθείς θα είναι οι δυο συντελεστές b_1 και b_2 που θα υπολογιστούν για τις μεταβλητές αυτές. Δεύτερον, στην πρακτική μορφή του μοντέλου παλινδρόμησης οι συντελεστές b_0 έως b_k είναι όλοι τυχαίες μεταβλητές, δηλαδή οι τιμές τους κυμαίνονται από δείγμα σε δείγμα, ενώ ακολουθούν μια κατανομή πιθανότητας. Συνεπώς, είναι δυνατόν να υπολογισθούν οι συσχετίσεις ανάμεσα στους συντελεστές.

4.3.12 Έλεγχος των υπόλοιπων σφαλμάτων (Residual Errors)

Η μελέτη των υπόλοιπων σφαλμάτων αναφέρεται στη μελέτη των σφαλμάτων προσαρμογής του μοντέλου στα πραγματικά δεδομένα και είναι ιδιαίτερα σημαντική για να αποφαιστεί η καταλληλότητα ενός μοντέλου πρόβλεψης. Αν τα σφάλματα είναι επαρκώς τυχαία, τότε το μοντέλο μπορεί να θεωρηθεί ικανοποιητικό. Αν τα σφάλματα ακολουθούν οποιοδήποτε πρότυπο, τότε το μοντέλο αδυνατεί να εκμεταλλευτεί όλη τη συστηματική πληροφορία που εμπεριέχεται στα δεδομένα. Μερικές από τις πιο πιθανές αναλύσεις των σφαλμάτων είναι οι ακόλουθες:

(α) διαγραμματική αναπαράσταση των σφαλμάτων για οπτική επισκόπηση και εύρεση της κατανομής που ακολουθούν

(β) μελέτη της αυτοσυσχέτισης των υπολοίπων σφαλμάτων

(γ) υπολογισμός του στατιστικού δείκτη Durbin-Watson

Ο στατιστικός δείκτης Durbin-Watson δίνεται από την παρακάτω εξίσωση :

$$DW = \frac{\sum_{t=2}^N (e_t - e_{t-1})^2}{\sum_{t=1}^N e_t^2} \quad (4.54)$$

Στον αριθμητή εμφανίζονται οι διαφορές ανάμεσα σε διαδοχικά σφάλματα, ενώ ο παρονομαστής ισούται με το άθροισμα των τετραγωνικών σφαλμάτων. Σε κάθε συνδυασμό αριθμού παρατηρήσεων, αριθμού συντελεστών παλινδρόμησης και επιπέδου εμοιστοσύνης, αντιστοιχεί ένα ζευγάρι αριθμητικών τιμών DW_L και DW_U . Ανάλογα με την υπολογισμένη τιμή του στατιστικού δείκτη, τα σφάλματα του εκάστοτε μοντέλου παλινδρόμησης χαρακτηρίζονται ως:

- Σημαντικά θετικά συσχετισμένα, αν $DW \leq DW_L$
- Ασυσχετίστα, αν $DW_U \leq DW \leq 4 - DW_L$
- Σημαντικά αρνητικά συσχετισμένα, αν $DW \geq 4 - DW_L$

Αν $DW_L \leq DW \leq DW_U$ ή $DW_U \leq DW \leq 4 - DW_L$ τότε δεν μπορεί να εξαχθεί ασφαλές συμπέρασμα από το στατιστικό δείκτη Durbin-Watson σχετικά με την τυχαιότητα των σφαλμάτων.

4.3.13 Βασικές υποθέσεις στην πολλαπλή παλινδρόμηση και πολυσυγγραμμικότητα

Σύμφωνα με τους Μακρυδάκη, Wheelright και Hyndman (1998), κάθε φορά που εφαρμόζεται στην πράξη η πολλαπλή παλινδρόμηση, πραγματοποιούνται τέσσερις βασικές υποθέσεις.

Η πρώτη εξ αυτών αφορά στην ύπαρξη γραμμικής σχέσης ανάμεσα στην εξαρτημένη και τις ανεξάρτητες μεταβλητές. Στην περίπτωση που δεν ικανοποιείται η υπόθεση αυτή, μετασχηματίζονται οι ανεξάρτητες μεταβλητές σε νέες μεταβλητές που εμφανίζουν γραμμική σχέση με την εξαρτημένη μεταβλητή Y .

Η δεύτερη υπόθεση αφορά στην ομοσκεδαστικότητα,στη σταθερή, δηλαδή, διακύμανση των σφαλμάτων παλινδρόμησης. Ο αντίστοιχος όρος για την έλλειψη σταθερής διακύμανσης είναι ετεροσκεδαστικότητα. Αυτή η υπόθεση δηλώνει ότι τα σφάλματα πρόβλεψης θα πρέπει να είναι σταθερα για όλο το εύρος των παρατηρήσεων.

Η τρίτη υπόθεση έγκειται στην ανεξαρτησία των υπολοίπων σφαλμάτων. Αυτό, πρακτικά, σημαίνει ότι η τιμή του κάθε υπολοίπου είναι ανεξάρτητη από τις τιμές των προηγούμενων και των επόμενων. Όταν δεν ικανοποιείται αυτή η υπόθεση, υπάρχει σειριακή συσχέτιση (ή αυτοσυσχέτιση) ανάμεσα σε διαδοχικές τιμές των υπολοίπων σφαλμάτων. Προκειμένου να προσδιοριστεί η ανεξαρτησία των υπολοίπων μπορεί να εφαρμοστεί είτε γραφική παράσταση των τιμών των υπολοίπων, είτε ελεγχος των προσήμων, είτε υπολογισμός του στατιστικού δείκτη Durbin-Watson. Όταν παρατηρείται εξάρτηση των υπολοίπων είναι πιθανόν να έχει παραλειφθεί κάποια ανεξάρτητη μεταβλητή ή να μην υπάρχει γραμμική σχέση ανάμεσα στις μεταβλητές της εξίσωσης παλινδρόμησης. Σε αυτήν την περίπτωση, η εξίσωση δεν αποδίδει πλήρως το βασικό λανθάνον πρότυπο (underlying pattern) των δεδομένων και τα υπόλοιπα σφάλματα, τα οποία δεν είναι τυχαία, αντιπροσωπεύουν κάποιο τμήμα του βασικού προτύπου. Αν το πρότυπο αυτό μπορεί να αποδοθεί πλήρως από την εξίσωση παλινδρόμησης, τότε είναι δυνατόν να επιτευχθεί μεγαλύτερη ακριβεια πρόβλεψης. Αναφορικά με την εξάλειψη της αυτοσυσχέτισης των υπολοίπων εφαρμόζονται δύο διαδικασίες:

Είναι πιθανόν να χρειάζεται και άλλη ανεξάρτητη μεταβλητή προκειμένου να ληφθούν υπ' όψιν μεταβολές της εξαρτημένης μεταβλητής που δεν μπορούσαν να ερμηνευτούν από τις υπάρχουσες ανεξάρτητες μεταβλητές και οδηγούσαν σε συστηματικά μη τυχαία σφάλματα.

Είναι πιθανόν να έχει χρησιμοποιηθεί στην εξίσωση παλινδρόμησης λανθασμένος συναρτησιακός τύπος (π.χ. : γραμμικός αντί για εκθετικός). Στην περίπτωση που δεν μπορεί να επινοηθεί μια νέα μεταβλητή ή που δεν μπορεί να μετασχηματιστεί μια υπάρχουσα μεταβλητή, η οποία να εξαλείφει την αυτοσυσχέτιση, μπορούμε να προβούμε στην μέθοδο των πρώτων διαφορών. Με βάση τη μέθοδο αυτή, αφού υπολογιστούν οι πρώτες διαφορές για κάθεμία από τις μεταβλητές της εξίσωσης παλινδρόμησης, οι συντελεστές της εξίσωσης παλινδρόμησης μπορούν να επαναυπολογιστούν χρησιμοποιώντας τις διαφορές αυτές ως παρατηρήσεις.

Η τέταρτη και τελευταία βασική υπόθεση έγκειται στην κανονική διασπορά που θα πρέπει να εμφανίζουν οι τιμές των υπολοίπων σφαλμάτων αν επιχειρήσουμε να τις αναπαραστήσουμε γραφικά. Η υπόθεση αυτή γενικά δεν είναι δεσμευτική καθώς τα υπόλοιπα αντιπροσωπεύουν την επίδραση (σχετικά ασήμαντη) ενός μεγάλου αριθμού παραγόντων στην τιμή της εξαρτημένης μεταβλητής.

Ένας επιπρόσθετος «κίνδυνος» που εμφανίζεται κάθε φορά που εφαρμόζουμε πολλαπλή παλινδρόμηση είναι η πιθανότητα πολυσυγγραμικότητας. Η πολυσυγγραμικότητα δημιουργείται όταν δύο ή περισσότερες ανεξάρτητες μεταβλητές εμφανίζουν ισχυρή συσχέτιση. Στην περίπτωση εμφάνιση πολυσυγγραμικότητας, το αποτέλεσμα είναι υπερβολικά μεγάλοι αριθμοί, των οποίων ο χειρισμός είναι δύσκολος. Η πολυσυγγραμικότητα αποτελεί ένα ιδιαίτερα συχνό πρόβλημα σε οικονομικά και επιχειρησιακά δεδομένα εξαιτίας του υψηλού βαθμού συσχέτισης που εμφανίζουν οι διάφοροι παράγοντες που επιλέγονται ως ανεξάρτητες μεταβλητές. Συνεπώς, το γεγονός αυτό θα πρέπει να λαμβάνεται σοβαρά υπόψη κατά την επιλογή των ανεξάρτητων μεταβλητών καθώς και κατά τη συλλογή δεδομένων. Ο στόχος είναι η χρησιμοποίηση ανεξάρτητων μεταβλητών, οι οποίες δεν είναι ισχυρά συσχετισμένες. Ένας εμπειρικός κανόνας προς έλεγχο τούτου είναι η απόλυτη τιμή της συσχέτισης να μην υπερβαίνει την τιμή 0,7. Στην περίπτωση που υπάρχει υπέρβαση αυτού του ορίου, παρέχεται πλεονάζουσα πληροφορία από τις ανεξάρτητες μεταβλητές, η οποία δεν μπορεί να αξιοποιηθεί με αποτέλεσμα να μην βελτιώνει την ερμηνευτική δύναμη της παλινδρόμησης.

4.3.14 Εφαρμογή της πολλαπλής παλινδρόμησης στην πράξη

Η πολλαπλή παλινδρόμηση είναι μία μέθοδος πρόβλεψης που βασίζεται στην αναγνώριση των παραγόντων που επηρεάζουν την προβλεπόμενη μεταβλητή. Ίσως το μεγαλύτερο πλεονέκτημα της μεθόδου αυτής είναι το γεγονός ότι μπορεί να θεωρηθεί και να διερευνηθεί ένας σημαντικός αριθμός διαφορετικών σχέσεων. Στη συνέχεια περιγράφεται μια διαδικασία κατά την οποία θεωρούνται κάποιες σχέσεις και χρησιμοποιώντας παλινδρόμηση καθορίζεται ποια από αυτές είναι η καταλληλότερη σε μια συγκεκριμένη περίπτωση.

Διατύπωση του προβλήματος

Η διατύπωση αρχίζει με την περιγραφή του αντικειμένου σχετικά με το οποίο πρέπει να ληφθεί μια απόφαση και με την αναγνώριση της μεταβλητής ή και των μεταβλητών που θα πρέπει να προβλεφθούν. Το αποτέλεσμα αυτού του σταδίου της διαδικασίας είναι ο καθορισμός της εξαρτημένης και των ανεξάρτητων μεταβλητών. Όπως γίνεται αντιληπτό, το συγκεκριμένο βήμα της διαδικασίας είναι κρίσιμης σημασίας και θα πρέπει να βασίζεται στην εμπειρία ανθρώπων που σχετίζονται με την περιοχή ενδιαφέροντος.

Επιλογή οικονομικών και άλλων σχετικών δεικτών

Εκτός από τις ανεξάρτητες μεταβλητές που καθορίστηκαν στο προηγούμενο βήμα, θα πρέπει να εξεταστεί το ενδεχόμενο να συμπεριληφθούν στην εξίσωση παλινδρόμησης και άλλοι πιθανοί παράγοντες που επιδρούν στην τιμή της εξαρτημένης μεταβλητής. Η απόφαση θα πρέπει να λαμβάνεται με βάση τη διαθεσιμότητα δεδομένων τόσο για παρελθούσες όσο και για μελλοντικές περιόδους, στις οποίες θα αναφέρεται η πρόβλεψη.

Τέτοιοι παράγοντες προς έλεγχο, είναι συχνά η δραστηριότητα κάποιας επιχείρησης και οι οικονομικές σειρές που σχετίζονται με την οικονομία και την βιομηχανία.

Αρχική δοκιμαστική εφαρμογή της πολλαπλής παλινδρόμησης

Η αρχική εφαρμογή θα πρέπει να περιέχει όλα τα δεδομένα σχετικά με τις ανεξάρτητες και εξαρτημένες μεταβλητές και μετασχηματισμούς μεταβλητών στην περίπτωση που κάποιες σχέσεις δεν παρουσιάζουν γραμμικότητα. Το συγκεκριμένο στάδιο μπορεί να περιλαμβάνει και ελέγχους κάποιων πιθανών εξισώσεων παλινδρόμησης σχετικά με τα αποτελέσματα που παράγουν. Ίσως το πιο χρήσιμο στοιχείο που απορρέει από το συγκεκριμένο στάδιο της διαδικασίας είναι ο προκύπτων πίνακας απλών συσχετίσεων που θα χρησιμοποιηθεί στο επόμενο στάδιο.

Μελέτη του πίνακα απλών συσχετίσεων

Η προσεκτική επιλογή των μεταβλητών (ή των μετασχηματισμών αυτών) που θα συμπεριληφθούν στην εξίσωση παλινδρόμησης αποτελεί βασική προϋπόθεση για την παραγωγή προβλέψεων υψηλής ακρίβειας. Θα πρέπει να επιλέγονται ανεξάρτητες μεταβλητές των οποίων οι συσχετίσεις δεν είναι μεγαλύτερες από 0,7 ή μικρότερες από -0,7. Επίσης θα πρέπει να επιλέγονται μεταβλητές οι οποίες συμβάλλουν στην ερμηνευτική ικανότητα της εξίσωσης παλινδρόμησης. Οι συντελεστές συσχέτισης ανάμεσα στις ανεξάρτητες και εξαρτημένες μεταβλητές που επιλέχθηκαν θα πρέπει να είναι σημαντικά μεγαλύτεροι του μηδενός. Κατά την ολοκλήρωση αυτού του βήματος της διαδικασίας θα πρέπει να έχουμε καταλήξει σε πέντε ή έξι εναλλακτικές εξισώσεις παλινδρόμησης οι οποίες θα ελεγχθούν ενδελεχώς στα επόμενα βήματα.

Επιλογή της εξίσωσης παλινδρόμησης

Για κάθε εξίσωση παλινδρόμησης, από αυτές που καταλήξαμε στο προηγούμενο βήμα, υπολογίζονται οι συντελεστές, τα τυπικά σφάλματα της πρόβλεψης και οι στατιστικοί δείκτες σημαντικότητας. Όταν βρεθεί η εξίσωση παλινδρόμησης της οποίας οι ανεξάρτητες μεταβλητές επηρεάζουν στατιστικά σημαντικά την εξαρτημένη μεταβλητή, επιχειρείται η αύξηση της τιμής του R^2 εισάγοντας και άλλες ανεξάρτητες μεταβλητές, ελέγχοντας κάθε φορά αν η σημαντικότητα της εξίσωσης παραμένει ικανοποιητική.

Παρατηρώντας την τιμή του R^2

Όταν όλοι οι συντελεστές παλινδρόμησης έχουν βρεθεί στατιστικά σημαντικοί και το τυπικό σφάλμα της πρόβλεψης θεωρείται αποδεκτό, εξετάζουμε την τιμή του R^2 . Ο συντελεστής R^2 εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής, το οποίο ερμηνεύεται από την εξίσωση παλινδρόμησης. Αν το ποσοστό αυτό είναι μικρό, γίνεται σαφές ότι η εξίσωση δεν μπορεί να χρησιμοποιηθεί για αξιόπιστες προβλέψεις.

Έλεγχος της εγκυρότητας των υποθέσεων για την παλινδρόμηση

Όταν θα έχει προσδιοριστεί μια ικανοποιητική εξίσωση, ελέγχεται αν ικανοποιούνται οι τέσσερις βασικές προϋποθέσεις που αναφέρθηκαν προηγουμένως. Στην περίπτωση που δεν ικανοποιούνται, θα πρέπει να γίνουν οι κατάλληλες ενέργειες ώστε να αποφευχθούν οι παραβιάσεις των υποθέσεων ή να αναπτυχθούν και να εξεταστούν πρόσθετες εξισώσεις παλινδρόμησης. Στο σημείο αυτό, αξίζει να σημειωθεί ότι υψηλές τιμές του R^2 δεν έχουν καμία αξία τη στιγμή που ο στατιστικός δείκτης Durbin-Watson δεν κυμαίνεται στο κατάλληλο διάστημα τιμών. Παρομοίως, παραβιάσεις των άλλων υποθέσεων ενδέχεται να

οδηγήσουν σε προβλήματα αξιοπιστίας της εξίσωσης, των οποίων,ωστόσο, η βαρύτητα δεν είναι τόσο μεγάλη όσο στην περίπτωση που ο δείκτης Durbin-Watson δηλώνει ότι τα υπόλοιπα σφάλματα δεν είναι τυχαία.

Προετοιμασία της Πρόβλεψης

Όταν προσδιοριστεί η εξίσωση παλινδρόμησης για την οποία ισχύουν τα εξής :

1. Οι συντελεστές παλινδρόμησης είναι στατιστικά σημαντικοί
2. Η τιμή του R^2 είναι επαρκώς υψηλή
3. Ικανοποιούνται οι 4 βασικές προϋποθέσεις

, τότε συμπεραίνουμε πως η εξίσωση αυτή μπορεί να χρησιμοποιηθεί για πραγματοποίηση προβλέψεων. Θα πρέπει να ελεγχθούν τα διαστήματα εμπιστοσύνης και η ακρίβεια των προβλεπόμενων τιμών για τις ανεξάρτητες μεταβλητές. Όπως είναι φυσικό, οι περισσότερες προβλέψεις βασίζονται σε εκτιμώμενες τιμές των ανεξάρτητων μεταβλητών παρά σε πραγματικές τιμές. Γίνεται σαφές ότι θα πρέπει να καθοριστεί η εγκυρότητα τους, καθώς αν οι προβλέψεις των ανεξάρτητων μεταβλητών είναι λανθασμένες, τότε και η πρόβλεψη της εξαρτημένης μεταβλητής Y είναι πού πιθανό να είναι λανθασμένη. Τέλος, αξίζει να σημειωθεί ότι η ανάπτυξη εξισώσεων παλινδρόμησης, εκτός από την παραγωγή προβλέψεων, επιτρέπει την καλύτερη ερμηνεία και κατανόηση διαφόρων πιθανών καταστάσεων και φαινομένων.

Κεφάλαιο 5 : Συνάθροιση Προβλέψεων

5.1 Μέθοδοι Συνάθροισης Προβλέψεων

Σύμφωνα με το ερευνητικό έργο των Babai, Ali και Νικολόπουλου (2012), υπάρχουν δύο διαφορετικοί τύποι συνάθροισης. Ο πρώτος τύπος ορίζει ότι η χρονική συνάθροιση αναφέρεται στη συνάθροιση χρόνου χωρίς επικαλύψεις (non-overlapping) μιας συγκεκριμένης χρονοσειράς. Με τον τρόπο αυτό, μια χρονοσειρά υψηλής συχνότητας μετατρέπεται σε χαμηλότερης συχνότητας χρονοσειρά σύμφωνα με το επιλεγμένο επίπεδο συνάθροισης. Για παράδειγμα, μπορούμε να εξάγουμε την ετήσια ζήτηση από μια μηνιαία χρονοσειρά ζήτησης επιλέγοντας επίπεδο συνάθροισης ίσο με 12 περιόδους. Ο δεύτερος τύπος ορίζει ότι η ιεραρχική συνάθροιση αναφέρεται στη συνάθροιση των ζητήσεων (bottom-up) πολλαπλών αντικειμένων (π.χ. αποθεματικών μονάδων) με στόχο των σχηματισμό οικογενειών προϊόντων

5.2 Χρονική Συνάθροιση Προβλέψεων Χρονοσειρών

Η χρονική συνάθροιση αναφέρεται στην παραγωγή μιας χρονοσειράς χαμηλής συχνότητας (π.χ. τριμηνιαία ή ετήσια) από μια χρονοσειρά (π.χ. μηνιαία), οι τιμές της οποίας παρατηρούνται σε υψηλότερη συχνότητα, μέσω μιας κατάλληλης διαδικασίας συνάθροισης των τιμών της. Η διαδικασία της χρονικής συνάθροισης μπορεί να εφαρμόζεται σκόπιμα στα δεδομένα, στα πλαίσια της επεξεργασίας τους, ή να προκύπτει φυσικά κατά τη συλλογή των τιμών της χρονοσειράς. Η δεύτερη περίπτωση εμφανίζεται κατά τη δειγματοληψία μεταβλητών ροής, όταν δεν καθίσταται δυνατή μια αναλυτική δειγματοληψία.

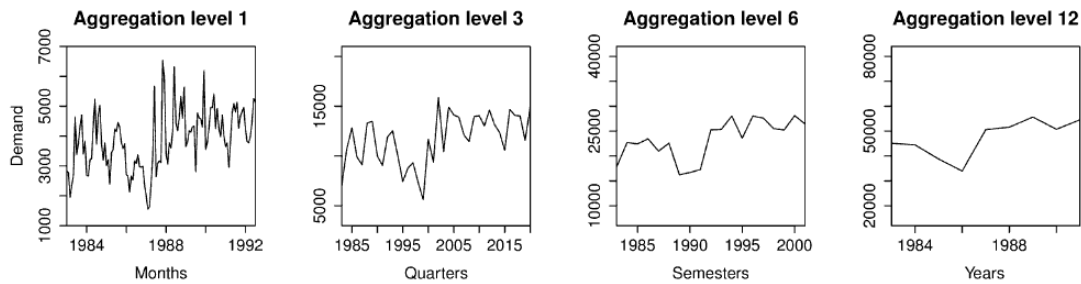
5.2.1 Κλασική Χρονική Συνάθροιση (Temporal Aggregation)

Η διαδικασία της χρονικής συνάθροισης συνδέεται με τις διαδικασίες δειγματοληψίας. Η εφαρμογή χρονικής συνάθροισης επί των δεδομένων εμπεριέχει τον κίνδυνο απώλειας πληροφορίας, αφού τα δεδομένα παρατηρούνται σε υψηλότερο επίπεδο. Στόχος της χρονικής συνάθροισης είναι ο εντοπισμός του επιπέδου συνάθροισης, στο οποίο παρέχεται η βέλτιστη πληροφορία με τον ελάχιστο θόρυβο και κατόπιν η παραγωγή προβλέψεων με βάση το συγκεκριμένο επίπεδο συνάθροισης.

Η χρονική συνάθροιση παρουσιάζει πολλές ομοιότητες με τη συνάθροιση διαφορετικών χρονοσειρών. Έχουν διεξαχθεί διάφορες μελέτες αναφορικά με την αξιοποίηση πληροφορίας από ανώτερα επίπεδα συνάθροισης και έχουν δείξει την προοπτική τέτοιων τεχνικών (Andrawis, Atiya, and El-Shishiny, 2011; Dekker, M., van Donselaar, K., and Ouwehand, 2004). Επίσης έχει επισημανθεί η σημασία του σαφούς προσδιορισμού ενός κατάλληλου επιπέδου συνάθροισης, το οποίο εξαρτάται από τις απαιτήσεις της εκάστοτε εφαρμογής και το οποίο επηρεάζει την αποδοτικότητα των διαδικασιών πρόβλεψης (Zotteri, Kalchschmidt, and Caniato, 2005).

Οι διαφορετικές συχνότητες των δεδομένων μπορούν να κρύβουν ή να αποκαλύπτουν τα διάφορα χαρακτηριστικά των χρονοσειρών. Κατά την εξέταση χρονοσειρών, οι τυχαίες διακυμάνσεις και τα εποχιακά χαρακτηριστικά είναι πιο εμφανή σε ημερήσια, εβδομαδιαία ή μηνιαία δεδομένα. Κάνοντας χρήση της χρονικής συνάθροισης χωρίς επικάλυψη (non-overlapping temporal aggregation) μπορούμε να προβούμε στην κατασκευή χρονοσειρών

χαμηλότερης συχνότητας. Το επίπεδο συνάθροισης αναφέρεται στη συχνότητα κατά τη διαδικασία της χρονικής συνάθροισης. Η αύξηση του επιπέδου συνάθροισης οδηγεί σε μία χρονοσειρά χαμηλότερης συχνότητας. Ταυτόχρονα, αυτή η διαδικασία λειτουργεί ως φίλτρο, που «λειαίνει» τα χαρακτηριστικά υψηλής συχνότητας, παρέχοντας, με τον τρόπο αυτό, μια καλύτερη προσέγγιση των μακροπρόθεσμων χαρακτηριστικών των δεδομένων, όπως το επίπεδο, την τάση και κύκλο. Στο παρακάτω σχήμα παρουσιάζεται η μορφή της χρονοσειράς για διάφορα επίπεδα συνάθροισης. Η αρχική μηνιαία χρονοσειρά κυριαρχείται από την εποχιακή συνιστώσα, ενώ η ετήσια σειρά (12^ο επίπεδο συνάθροισης) κυριαρχείται από μια μετατόπιση του επιπέδου και μια αδύναμη τάση.



Σχήμα 5.1 : Μηνιαία χρονοσειρά για διαφορετικά επίπεδα χρονικής συνάθροισης χωρίς επικάλυψη (Πηγή: N. Kourentzes, F. Petropoulos and J. R. Trapero, 2014, *improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting*)

Διαδικασία αντίστροφη της συνάθροισης είναι η διάσπαση των συναθροισμένων δεδομένων (disaggregation), σύμφωνα με κάποιον αλγόριθμο διάσπασης ώστε από μια χρονοσειρά χαμηλής συχνότητας (π.χ. τριμηνιαία ή ετήσια) να προκύψει μια χρονοσειρά υψηλής συχνότητας (π.χ. μηνιαία). Με τον τρόπο αυτό, κάποια δεδομένα υψηλής συχνότητας μπορούν να συναθροιστούν σε υψηλότερο επίπεδο, στη συνέχεια να γίνει κάποιου είδους επεξεργασία των δεδομένων (π.χ. προέκταση στο μέλλον) και κατόπιν μέσω κάποιας διαδικασίας διάσπασης να επιστρέψουμε στο χαμηλότερο επίπεδο παρατήρησης.

5.2.2 Πολλαπλή Χρονική Συνάθροιση (Multiple Temporal Aggregation)

Η πολλαπλή χρονική συνάθροιση αναφέρεται στην παραγωγή πολλαπλών χρονοσειρών εναλλακτικών συχνοτήτων από μια χρονοσειρά, οι τιμές της οποίας παρατηρούνται σε μια συγκεκριμένη (υψηλότερη συχνότητα). Μέσω των διαφορετικών συχνοτήτων των δεδομένων, αποκαλύπτονται ή κρύβονται διαφορετικά χαρακτηριστικά των δεδομένων. Η παραγωγή των πολλαπλών χρονοσειρών εναλλακτικών συχνοτήτων επιτυγχάνεται με κατάλληλη επιλογή του επιπέδου συνάθροισης.

5.2.3 Πολλαπλή Χρονική Συνάθροιση Προβλέψεων Συνιστωσών της Χρονοσειράς (Multi Aggregation Prediction Algorithm MAPA)

Στόχος κάθε αποδοτικής μεθόδου πρόβλεψης είναι η παραγωγή προβλέψεων υψηλής ακρίβειας. Για την επίτευξη αυτής της ζητούμενης ακρίβειας, σπουδαίο ρόλο παίζει ο εντοπισμός και η επιλογή του κατάλληλου μοντέλου πρόβλεψης χρονοσειρών. Το συγκεκριμένο πρόβλημα είναι ιδιαίτερα πολύπλοκο και απαιτεί αρκετή ανάλυση και

κόστος. Η επιλογή κατάλληλων παραμέτρων για κάποιο μοντέλο πρόβλεψης χρονοσειρών είναι ιδιαίτερα απαιτητική διαδικασία. Πιο συγκεκριμένα, οι σημερινές επιχειρήσεις και οργανισμοί έχουν ανάγκη να παράγουν πληθώρα προβλέψεων για διάφορες χρονοσειρές. Η έλλειψη, ωστόσο, στατιστικών δεδομένων και η αδυναμία επεξεργασίας αυτών καθιστά το πρόβλημα της πρόβλεψης ακόμα δυσκολότερο, όπως παρατήρησε ο Hughes στην έρευνα του το 2006.

Προτείνεται, λοιπόν, στην βιβλιογραφία ένας αλγόριθμος που στοχεύει στην αύξηση της ακρίβειας των προβλέψεων ενώ παράλληλα μετριάξει τη σημασία της επιλογής κατάλληλου μοντέλου πρόβλεψης.

Σύμφωνα με αυτόν τον αλγόριθμο, παράγονται πολλαπλές χρονοσειρές από τα αρχικά δεδομένα της χρονοσειράς, χρησιμοποιώντας χρονική συνάθροιση (temporal aggregation). Οι νέες χρονοσειρές που προκύπτουν εκφράζουν και τονίζουν διαφορετικά χαρακτηριστικά της αρχικής χρονοσειράς καθώς η χρονική συνάθροιση συμβάλλει στην επισήμανση ή στην μείωση της σημασίας σημάτων που προέρχονται από συνιστώσες διαφορετικών χρονοσειρών. Σε κάθε χρονοσειρά εφαρμόζεται η κατάλληλη εκθετική μέθοδος εξομάλυνσης και παράγονται προβλέψεις για τις συνιστώσες κάθε χρονοσειράς.

Αυτή η διαδικασία μας επιτρέπει τον τελικό συνδυασμό των συνιστωσών των χρονοσειρών διαφορετικής χρονικής συνάθροισης και την παραγωγή των τελικών προβλέψεων. Με αυτήν την προσέγγιση, επιτυγχάνεται καλύτερη πρόβλεψη των συνιστωσών των διαφορετικών χρονοσειρών μέσω της χρονικής συνάθροισης, ενώ μειώνεται η σημασία επιλογής μοντέλου αφού έχουμε συνδυασμό των προβλέψεων.

Στηριζόμενοι στην έρευνα των Andrawis, Atiya και el-Shishiny το 2011, παρατηρούμε πως είναι δυνατόν να δοθεί έμφαση σε διαφορετικά χαρακτηριστικά των χρονοσειρών μέσω της μετατροπής των αρχικών δεδομένων σε δεδομένα εναλλακτικών συχνοτήτων χρόνου. Σύμφωνα με την έρευνα αυτή, προτείνεται η συνάθροιση μιας χρονοσειράς σε πολλαπλές χαμηλότερες συχνότητες. Για παράδειγμα, μια μηνιαία χρονοσειρά γίνεται δημηνιαία, τριμηνιαία, κ.ο.κ. Σε κάθε επίπεδο συνάθροισης, τονίζεται ή εξασθενείται η σημασία διαφορετικών χαρακτηριστικών των χρονοσειρών. Στα χαμηλότερα επίπεδα συνάθροισης, όπου ασχολούμαστε με χρονοσειρές υψηλής συχνότητας, δίνεται έμφαση στις περιοδικές συνιστώσες της χρονοσειράς, όπως η εποχιακότητα. Καθώς αυξάνεται το επίπεδο συνάθροισης, τα σήματα υψηλής συχνότητας φιλτράρονται με αποτέλεσμα να μετριάζεται η σημασία τους ενώ δίνεται έμφαση στις συνιστώσες χαμηλής συχνότητας των δεδομένων όπως το επίπεδο και η τάση της χρονοσειράς. Αναμένουμε, δηλαδή, να εντοπίσουμε τα εποχιακά χαρακτηριστικά μιας χρονοσειράς στα χαμηλά επίπεδα συνάθροισης (δεδομένα υψηλής συχνότητας). Το αντίθετο ισχύει για το επίπεδο και την τάση, συνιστώσες οι οποίες τονίζονται στα υψηλά επίπεδα συνάθροισης (δεδομένα χαμηλής συχνότητας). Συνεπώς, κάνοντας χρήση της παραπάνω διαδικασίας συνάθροισης, διευκολύνεται ο προσδιορισμός, η επιλογή και η παραμετροποίηση των προγνωστικών μοντέλων.

Μετά την χρονική συνάθροιση, παράγονται προβλέψεις σε κάθε επίπεδο συνάθροισης χρησιμοποιώντας την κατάλληλη μέθοδο πρόβλεψης. Εξάγουμε προβλέψεις, κάνοντας χρήση μεθόδων εκθετικής εξομάλυνσης, οι οποίες παρέχουν το πλαίσιο για την

μοντελοποίηση όλων των τύπων των χρονοσειρών, είτε παρουσιάζουν εποχιακότητα, επίπεδο ή τάση. Ο υπολογισμός πολυάριθμων προβλέψεων σε διαφορετικές συχνότητες οδηγεί αναπόφευκτα στην χρήση κάποιων συστημάτων συνδυασμού των προβλέψεων αυτών. Ακριβώς σε αυτό το στάδιο του συνδυασμού, παρέχεται η δυνατότητα περαιτέρω βελτίωσης της ακρίβειας και της αξιοπιστίας των προκυπτουσών προβλέψεων.

Αντί του άμεσου συνδυασμού των προβλέψεων που προκύπτουν από διαφορετικά επίπεδα συνάθροισης, προτείνεται η ανάλυση/αποσύνθεση των προβλέψεων αυτών σε προβλέψεις συνιστωσών των χρονοσειρών και ο εν-συνεχεία συνδυασμός των αναλυτικών αυτών προβλέψεων. Η προτεινόμενη αυτή διαδικασία υπαγορεύεται από τις πιθανές διαφορές στους τρόπους με τους οποίους εμφανίζονται ή μεταλλάσσονται οι συνιστώσες της χρονοσειράς στα διάφορα επίπεδα συνάθροισης. Για παράδειγμα, αν από μια μηνιαία χρονοσειρά δημιουργήσουμε την ετήσια μέσω χρονικής συνάθροισης, η προκύπτουσα χρονοσειρά δεν θα έχει καμία εποχιακότητα. Αν, ωστόσο, συνδυάσουμε τις μηνιαίες και τις ετήσιες προβλέψεις, τότε θα παρατηρήσουμε το εποχιακό σήμα της χρονοσειράς. Συνεπώς, συνδυασμός δεδομένων με εποχιακότητα θα πρέπει να γίνεται μόνο σε εκείνα τα επίπεδα συνάθροισης, στα οποία παρουσιάζεται αυτή η εποχιακότητα, ενώ ο συνδυασμός δεδομένων με τάση και επίπεδο μπορεί να πραγματοποιηθεί σε οποιοδήποτε επίπεδο συνάθροισης. Επιπρόσθετα, αν συνδυάσουμε τις προβλέψεις των επιπέδων συνάθροισης αντί των συνιστωσών των χρονοσειρών δεν θα είμαστε σε θέση να διακρίνουμε τα επιθυμητά στοιχεία των χρονοσειρών που προκύπτουν από κάθε επίπεδο συνάθροισης που πρόκειται να συνδυαστεί.

Τα πλεονεκτήματα-κλειδιά του προτεινόμενου πλαισίου είναι τα εξής:

- Μετριάζεται η σημασία επιλογής του μοντέλου, αξιακό πρόβλημα της μοντελοποίησης χρονοσειρών.
- Εκμεταλλεύεται το κέρδος της ακριβούς πρόβλεψης από την χρονική συνάθροιση καθώς και το κέρδος του συνδυασμού των προβλέψεων .
- Καθιστά την χρήση της εκθετικής εξομάλυνσης ,μέθοδος η οποία είναι καθιερωμένη, καλά μελετημένη, και εφαρμοσμένη σε πολλά προγνωστικά συστήματα υποστήριξης.
- Είναι απλό επιτρέποντας έτσι την διαφάνεια και την άμεση εφαρμογή του σε οργανισμούς και επιχειρήσεις.

Η χρονική συνάθροιση έχει μελετηθεί στο πλαίσιο των μοντέλων ARIMA και διαδικασιών GARCH. Παρά το γεγονός ότι δεν υπάρχει κάποιο θεωρητικό υπόβαθρο (σύμφωνα με την έρευνα του Wei από το 1990), έχει αποδειχτεί εμπειρικά ότι η συνάθροιση λειτουργεί εξαιρετικά καλά για μοντέλα ARIMA αναφορικά με την ακρίβεια των προβλέψεων και την μείωση της μεροληψίας.

Οι συνέπειες της χρονικής συνάθροισης και της συστηματικής δειγματοληψίας αναφορικά με το μήκος υστερήσης (lag length) (Brewer,1973) και τις δυναμικές σχέσεις μεταξύ των μεταβλητών (Weiss 1984) είναι ένα θέμα που έχει μελετηθεί στην βιβλιογραφία. Έχουν προκύψει χρήσιμα συμπεράσματα σχετικά με την σταθερότητα και των μοντέλων ARIMA

και διαδικασιών GARCH. Επιπλέον, οι Drost και Nijman (1993) έδειξαν ότι στην περίπτωση μοντέλων GARCH με μία μεταβλητή, οι παράμετροι διασποράς του μοντέλου χαμηλής συχνότητας εξαρτώνται γενικά από τη μέση, διακύμανση, και την παράμετρο κύρτωσης του μοντέλου υψηλής συχνότητας.

Ο Haffner μελέτησε την χρονική συνάθροιση στα πλαίσια πολυμεταβλητών διαδικασιών GARCH (1,1) και κατέληξε στο συμπέρασμα ότι τα δυναμικά διαδικασιών συνάθροισης μπορούν να ληφθούν με έναν απλό τρόπο. Ως αποτέλεσμα, σήμερα, έχει πραγματοποιηθεί πολύ εμπειρικό και θεωρητικό έργο πάνω στα μοντέλα ARIMA και στις διαδικασίες GARCH στα πλαίσια της χρονικής συνάθροισης.

Από την άλλη πλευρά, η βιβλιογραφία σχετικά με τις επιδράσεις της χρονικής συνάθροισης με μοντέλα εκθετικής εξομάλυνσης είναι πολύ πιο περιορισμένη, παρά το γεγονός ότι αυτή μέθοδος εφαρμόζεται ευρέως στην πράξη (Gardner 2006).

Οι Σπιθουράκης, Πετρόπουλος, Babai, Νικολόπουλος και Ασημακόπουλος (2011) εξέτασαν την αποτελεσματικότητα ενός πλαισίου χρονικής συνάθροισης για τεχνικές πρόβλεψης που χρησιμοποιούνται ευρέως. Τα αποτελέσματα, με βάση τα μηνιαία στοιχεία από τον M3- Διαγωνισμό (Μακριδάκης & Hibon, 2000), έδειξαν σημαντικά βελτιωμένα ακρίβεια πρόβλεψης για τις μεθόδους Naive, SES και Theta (Ασημακόπουλος & Νικολόπουλος, 2000). Οι Αθανασόπουλος, Hyndman, Song και Wu (2011) έδειξαν, ότι στον τομέα του τουρισμού οι προβλέψεις που παρήχθησαν μέσω χρονικής συνάθροισης δεδομένων υψηλής συχνότητας ήταν πιο ακριβείς από τις προβλέψεις που παρήχθησαν απευθείας από τα δεδομένα χαμηλής συχνότητας. Ήταν οι πρώτοι που διερεύνησαν εμπειρικά τον αντίκτυπο της χρονικής συνάθροισης στην απόδοση της οικογένειας μοντέλων εκθετικής εξομάλυνσης, αν και η έρευνα τους περιοριζόταν στην συνάθροιση μηνιαίων και τριμηνιαίων δεδομένων σε ετήσια βάση. Αξιολόγηση των προβλέψεων τους έγινε μόνο στις χαμηλότερες συχνότητες (ετήσια δεδομένα). Στην βιβλιογραφία έχει επίσης μελετηθεί η εφαρμογή χρονικής συνάθροισης σε διακοπώμενα δεδομένα. Τα συμπεράσματα που προκύπτουν είναι ότι η χρονική συνάθροιση μέχρι ενός σημείου είναι ευεργετική καθώς η τυχαιότητα εξομαλύνεται ενώ παράλληλα ενισχύονται τα μακροπρόθεσμα χαρακτηριστικά της χρονοσειράς.

Ένα από τα κυριότερα θέματα που σχετίζονται με την χρονική συνάθροιση είναι η επιλογή του βέλτιστου επιπέδου συνάθροισης για ένα σύνολο χρονοσειρών ή για κάθε χρονοσειρά ξεχωριστά καθώς δεν υπάρχει κάποια απόδειξη ότι το επίπεδο αυτό πρέπει να είναι καθολικό σε ένα σύνολο χρονοσειρών. Σε αυτήν την κατεύθυνση, ο Νικολόπουλος (2011) προτείνει τον καθορισμό του επιπέδου συνάθροισης ως τον ορίζοντα πρόβλεψης συν μια περίοδο εξέτασης και αυτό το εύρημα δείχνει μια καλή ακρίβεια πρόβλεψης. Ο Σπιθουράκης (2011), στην προσπάθεια του να εντοπίσει εμπειρικά το βέλτιστο επίπεδο συνάθροισης για κάθε χρονοσειρά, χρησιμοποίησε διάφορα κριτήρια σε δείγματα (in-sample criteria) και διαπίστωσε σημαντικές βελτιώσεις για επίπεδα συγκέντρωσης έως και 12 περιόδους.

Η προτεινόμενη μέθοδος MAPA βαίζεται σε αυτά τα κριτήρια. Ωστόσο δεν επικεντρώνεται σε ένα προκαθορισμένο επίπεδο συνάθροισης λόγω της απουσίας θεωρητικού υπόβαθρου για την επιλογή του βέλτιστου επιπέδου. Ενώ η μέθοδος ADIDA χρησιμοποιεί ένα ενιαίο

καθολικό επίπεδο συνάθροισης, η μέθοδος MAPA λαμβάνει υπ' όψιν της την χρονοσειρά με πολλαπλά επίπεδα συνάθροισης ταυτόχρονα, εκμεταλλεύομενη κατ' αυτόν τον τρόπο τα πλεονεκτήματα τόσο από κάθε χρονοσειρά που έχει προκύψει με συνάθροιση (για τα πολλαπλά επίπεδα συνάθροισης), όσο και από την αρχική χρονοσειρά. Δεδομένου ότι δεν έχει επιλεγεί κάποιο ενιαίο επίπεδο συνάθροισης, παράγονται διαφορετικές προβλέψεις για το ίδιο αντικείμενο. Οι προκύπτουσες πληροφορίες μπορούν να συνδυαστούν σε μια ενιαία πρόβλεψη. Αυτή η ενιαία πρόβλεψη αποτελεί την εξοδο MAPA.

5.2.3.1 Συνδυασμός προβλέψεων για διαφορετικά επίπεδα συνάθροισης

Ο συνδυασμός θεωρείται ευρέως επωφελής για τις προβλέψεις όπως έχουν υποστηρίξει στην βιβλιογραφία οι Bates & Granger (1969), ο Clemen (1989) και οι Μακριδάκης & Winkler (1983). Λαμβάνοντας το μέσο όρο στατιστικών προσεγγίσεων, οδηγούμαστε στην βελτίωση της προγνωστικής ακρίβειας. Επίσης, ο συνδυασμός οδηγεί σε χαμηλότερα ποσοστά αβεβαιότητας, καθώς μειώνεται η διακύμανση των σφαλμάτων πρόβλεψης (Hibon & Ευγενίου, 2005). Πολλές πρόσφατες μελέτες έχουν εξετάσει εξελιγμένες διαδικασίες επιλογής βάρους για το συνδυασμό προβλέψεων (HE & Xu (2005), Kolassa (2011), Taylor (2008)). Ωστόσο, οι απλές προσεγγίσεις θεωρούνται σταθερές και αποδίδουν αρκετά καλά σε σχέση με πιο πολύπλοκες μεθόδους (Clemen (1989), Hibon & Ευγενίου (2005), Χοσέ & Winkler (2008), Timmermann, (2006)). Σύμφωνα με τον Kolassa (2011), παρατηρείται μικρή συνέναιση στην βιβλιογραφία σχετικά με το εάν θα έπρεπε να λαμβάνουμε τον μέσο όρο ή την τιμή median, στα πλαίσια της καλύτερης απόδοσης των προβλέψεων.

Ο συνδυασμός των προβλέψεων που παράγονται από διαφορετικές συχνότητες των ίδιων δεδομένων παρέχει την δυνατότητα να ληφθούν άμεσα διαφορετικές πληροφορίες των δυναμικών των δεδομένων. Ο Cholette (1982) ήταν ο πρώτος που ερεύνησε τα πλεονεκτήματα αυτά τροποποιώντας τις μηνιαίες προβλέψεις που παρήχθησαν από ένα μοντέλο ARIMA μέσω της ενσωμάτωσης κριτικών προβλέψεων από εμπειρογνώμονες ειδικούς που χρησιμοποίησαν δεδομένα διαφορετικών συχνοτήτων.

Με το θέμα του συνδυασμού προβλέψεων από μοντέλα ARIMA ασχολήθηκαν αργότερα και οι Trabelsi and Hillmer (1989), Casals, Jerez, and Sotoca (2009) και απέδειξαν θεωρητικά ότι η συνάθροιση μπορεί να επιφέρει βελτίωση της προγνωστικής ακρίβειας μέσω του κατάλληλου συνδυασμού προβλέψεων που παράγονται από δεδομένα διαφορετικής συχνότητας.

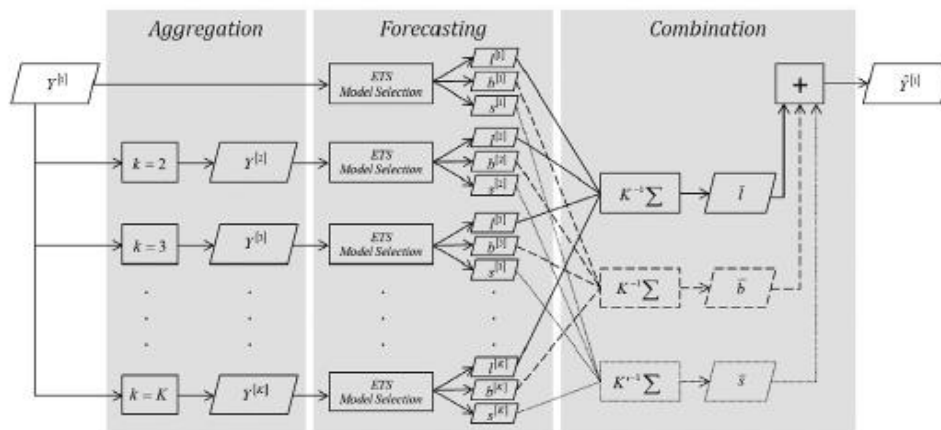
Ο Andrawis (2011) ερεύνησε τα οφέλη του συνδυασμού βραχυπρόθεσμων και μακροπρόθεσμων προβλέψεων. Χρησιμοποιώντας μηνιαίες και ετήσιες χρονοσειρές, ερεύνησε τη βέλτιστη μέθοδο συνδυασμού των προβλέψεων. Προκειμένου να καταστεί εφικτός ο συνδυασμός, πρώτα απεποχικοποίησε τη μηνιαία χρονοσειρά. Κατέληξε στο συμπέρασμα ότι ο συνδυασμός μπορεί να είναι επωφελής για την ακρίβεια των προβλέψεων. Τέλος, ο Granger (1993) είχε ισχυριστεί, νωρίτερα, ότι η χρονική συνάθροιση πρέπει να εφαρμόζεται με σύνεση, ιδιαίτερα στην περίπτωση της ενσωμάτωσης κριτικών προβλέψεων καθώς ενδέχεται να έχει διάφορες συνέπειες και επιδράσεις στις τελικές προβλέψεις.

5.2.3.2 Παραγωγή Προβλέψεων μέσω Χρονικής Συνάθροισης Συνιστωσών της Χρονοσειράς (Multi Aggregation Prediction Algorithm MAPA)

Το συγκεκριμένο πλαίσιο παραγωγής προβλέψεων μπορεί να διαχωριστεί στα εξής τρία στάδια:

- (α) το στάδιο της συνάθροισης
- (β) το στάδιο της πρόβλεψης
- (γ) το στάδιο του συνδυασμού

Μια οπτική επισκόπηση των 3 βημάτων της MAPA παρουσιάζεται στο παρακάτω σχήμα:



Σχήμα 5.2 : Διάγραμμα ροής του προτεινόμενου αλγόριθμου πρόβλεψης πολλαπλών συναθροίσεων (MAPA) (Πηγή:International Institute of Forecasters.Published by Elsevier B.V.)

5.2.3.2.1 Στάδιο της συνάθροισης

Έστω Y_t μια χρονοσειρά με παρατηρήσεις y_t (όπου $t=1, \dots, n$) που λαμβάνονται με μια δεδομένη συχνότητα. Μπορούμε να συναθροίσουμε αυτή την χρονοσειρά θεωρώντας διαδοχικά σύνολα τιμών της αρχικής χρονοσειράς ομαδοποιημένα σε σύνολα μήκους k . Συμβολίζουμε τις χρονικά συναθροισμένες χρονοσειρές με $Y^{[k]}$ με παρατηρήσεις $y_i^{[k]}$ και $i=1, 2, \dots, n/k$ έτσι ώστε :

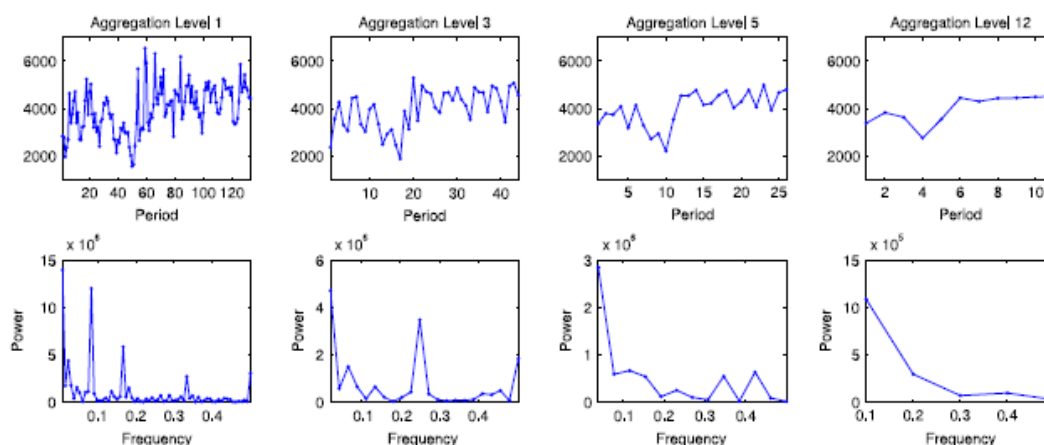
$$y_i^{[k]} = \sum_{t=1+(i-1)k}^{ik} y_t \quad (5.1)$$

Προφανώς για $i=1$ προκύπτει $y_i^{[k]} = y_t$. Θεωρούμε ως τελεστή συνάθροισης τον αριθμητικό μέσο όρο. Η προκύπτουσα χρονοσειρά $Y^{[k]}$ έχει (n/k) παρατηρήσεις. Αν το υπόλοιπο της διαίρεσης n/k είναι μη μηδενικό για ένα συγκεκριμένο επίπεδο συνάθροισης k , τότε κάποιες παρατηρήσεις δεν θα συμπεριληφθούν κατά τον σχηματισμό συνόλων χρονικών συναθροίσεων (aggregation buckets). Στην περίπτωση αυτή αφαιρούμε από την αρχική χρονοσειρά $(n-n/k)$ παρατηρήσεις ξεκινώντας από την αρχή ούτως ώστε να σχηματίσουμε πλήρη σύνολα συνάθροισης.

Θεωρητικά, μπορούμε να συναθροίσουμε μια χρονοσειρά για όλα τα $k=2, \dots, n$. Ωστόσο, στην πράξη είναι προτιμότερο να κάνουμε συνάθροιση για $k \ll n$, ούτως ώστε να αφήσουμε

ένα επαρκές δείγμα για παραμετροποίηση της μεθόδου πρόβλεψης που εφαρμόζεται στην χρονοσειρά $Y^{[k]}$.

Στο παρακάτω σχήμα παρουσιάζεται ένα παράδειγμα όπου μια μηνιαία χρονοσειρά συναθροίζεται για $k=1,3,5,12$. Το σχήμα περιλαμβάνει τα διαγράμματα των προκυπτουσών χρονοσειρών και τα αντίστοιχα φάσματα ισχύος. Για $k=1$ διακρίνεται ξεκάθαρα ένα εποχιακό σήμα στο φάσμα ισχύος. Το ίδιο συμβαίνει για $k=3$ με τη διαφορά ότι η διακύμανση που παρουσιάζεται πλέον είναι μικρότερη και με σαφή αλλαγή της συχνότητας. Για υψηλότερα επίπεδα συνάθροισης, η εποχιακή συνιστώσα δεν μπορεί να παρατηρηθεί ούτε στο διάγραμμα της χρονοσειράς, ούτε στο φάσμα ισχύος.



Σχήμα 5.3 : Διαγράμματα χρονοσειρών και φάσματα ισχύος για $k=1,3,5,12$
(Πηγή : *International Journal of Forecasting* 30 (2014)/ N. Kourentzes et al.)

5.2.3.2.2 Στάδιο της πρόβλεψης

Μετά τον σχηματισμό των συναθροισμένων χρονοσειρών $Y^{[k]}$, εφαρμόζουμε τα κατάλληλα μοντέλα πρόβλεψης για κάθε χρονοσειρά. Είναι επιθυμητό να λάβουμε υπ' όψιν τις διάφορες συνιστώσες των χρονοσειρών, που εμφανίζονται σε κάθε επίπεδο συνάθροισης, και όχι τις προβλέψεις, έτσι ώστε να είναι διακριτά τα επίπεδα συγέντρωσης στα οποία παρατηρείται η συνιστώσα της εποχιακότητας. Επιπλέον, εξετάζοντας τις συνιστώσες των χρονοσειρών απλοποιείται ο συνδυασμός των προβλέψεων καθώς ο συνδυασμός συνιστωσών διαφορετικών επιπέδων συνάθροισης ή ο συνδυασμός των εποχιακών συνιστωσών μπορεί να γίνει απ' ευθείας χωρίς κάποια ειδική προ-επεξεργασία. Από την άλλη, η σημασία του συνδυασμού προβλέψεων των συνιστωσών επιπέδου, τάσης και εποχιακότητας δεν είναι σαφής. Για παράδειγμα, ο συνδυασμός μιας μη εποχιακής πρόβλεψης και μιας εποχιακής συνιστώσας θα είχε ως αποτέλεσμα την κατά το ήμισυ εμφάνιση του εποχιακού σήματος. Προηγούμενες μελέτες έχουν ασχοληθεί με αυτό το ζήτημα μέσω της εποχιακής προσαρμογής των αρχικών δεδομένων θέτοντας με αυτόν τον τρόπο μια ντετερμινιστική εποχιακότητα η οποία μπορεί να είναι ακατάλληλη ή να χρήζει χαρακτηρισμού ως αθροιστική ή πολλαπλασιαστική.

Η μέθοδος πρόβλεψης εκθετικής εξομάλυνσης παρέχει τη δυνατότητα της άμεσης «άντλησης» των συνιστωσών του επιπέδου, της τάσης και της εποχιακότητας από το προσαρμοσμένο μοντέλο.

Τα μοντέλα εκθετικής εξομάλυνσης εντοπίζουν και εξομαλύνουν τις υφιστάμενες διαρθρωτικές συνιστώσες της χρονοσειράς : επίπεδο, τάση και εποχιακότητα. Οι συνιστώσες αυτές μπορούν να αλληλεπιδρούν μεταξύ τους είτε με αθροιστικό είτε με πολλαπλαστικό τρόπο. Η τάση μπορεί να είναι γραμμική ή αποσβενύμενη. Για κάθε χρονοσειρά πρέπει να επιλέγονται οι κατάλληλες συνιστώσες. Ο Hyndman (2002) πρότεινε την εφαρμογή του κριτηρίου πληροφoρίας Akaike για την «αυτόματη» επιλογή του κατάλληλου τύπου εξομάλυνσης, παρουσιάζοντας αποτελέσματα ικανοποιητικής ακρίβειας.

Με βάση το προαναφερθέν κριτήριο, εφαρμόζεται μια μέθοδος εξομάλυνσης σε κάθε χρονικά συναθροισμένη χρονοσειρά $Y^{[k]}$. Ενδιαφερόμαστε για το τελευταίο διάλυσμα κατάστασης $x_i^{[k]}$ της εκθετικής εξομάλυνσης. Το διάλυσμα κατάστασης αποτελείται από τις ανανεωμένες τιμές κάθε συνιστώσας των χρονοσειρών : επίπεδο I_i , τάση b_i και εποχιακότητα s_i , όπου είναι διαθέσιμη, δηλαδή $x_i^{[k]} = (I_i, b_i, s_i, s_{i-1}, \dots, s_{i-S+1})'$. Χρησιμοποιώντας αυτές τις πληροφορίες και γνωρίζοντας τον τύπο της εξομάλυνσης που εφαρμόστηκε, μπορούμε να παράγουμε προβλέψεις για οποιονδήποτε ορίζοντα πρόβλεψης επιθυμούμε. Σημειώνεται ότι για $k > 1$, ο επιθυμητός χρονικός ορίζοντας για κάθε επίπεδο συνάθροισης ρυθμίζεται ως $h^{[k]} = (h/k)$. Ο ορίζοντας πρόβλεψης για την αρχική χρονοσειρά είναι $h^{[1]} = h$. Η εποχιακή συνιστώσα μπορεί να εξαφανίζεται και να επανεμφανίζεται για διαφορετικά επίπεδα συνάθροισης. Εάν η χρονοσειρά $Y^{[1]}$ έχει περίοδο εποχιακότητας S , τότε η συγκεκριμένη συνιστώσα θα ληφθεί ξανά υπ όψιν αν ο λόγος S/k είναι ακέραιος αριθμός και εφ' όσον $k < S$. Για παράδειγμα, αν η αρχική εποχιακή περιοδικότητα είναι $S=12$ για τα μηνιαία δεδομένα, τότε η εποχιακότητα θα μοντελοποιηθεί ξανά για $k=2, 3, 4, 6$. Πιθανές μη ακέραιες εποχιακότητες δεν λαμβάνονται υπ' όψιν λόγω της αδυναμίας των μεθόδων εξομάλυνσης και των άλλων συμβατικών μεθόδων πρόβλεψης να μοντελοποιήσει μη ακέραιες εποχιακότητες. Οι αριθμητικές τιμές των αθροιστικών και πολλαπλασιαστικών συνιστωσών του $x_i^{[k]}$ δεν είναι άμεσα συγκρίσιμες. Είναι πιθανό να μπορεί να εφαρμοστεί μια αθροιστική εξομάλυνση σε ένα επίπεδο συνάθροισης και μια πολλαπλασιαστική μέθοδος σε ένα άλλο επίπεδο συνάθροισης. Αυτό, ωστόσο, θα αύξανε την πολυπλοκότητα του μεταγενέστερου συνδυασμού των συνιστωσών όλων των χρονικά συναθροισμένων χρονοσειρών $Y^{[k]}$. Στις περιπτώσεις αυτές προτείνεται η μετατροπή των πολλαπλασιαστικών συνιστωσών σε αθροιστικές.

Χρησιμοποιώντας τις συνιστώσες του διανύσματος κατάστασης $x_i^{[k]}$, κατασκευάζουμε προβλέψεις για κάθε επίπεδο I_i , τάση b_i και εποχιακότητα s_i , όπου είναι διαθέσιμη και τις μετατρέπουμε σε αθροιστικές όπου είναι απαραίτητο με βάση τους τύπους του παρακάτω πίνακα.

| Trend | Seasonal | | |
|-------|---|--|---|
| | N | A | M |
| N | $I_{i+h} = I_i$ | $I_{i+h} = I_i$ $s_{i-S+h} = s_{i-S+h}$ | $I_{i+h} = I_i$ $s_{i-S+h} = (s_{i-S+h} - 1)I_{i+h}$ |
| A | $I_{i+h} = I_i$ $b_{i+h} = hb_i$ | $I_{i+h} = I_i$ $b_{i+h} = hb_i$ $s_{i-S+h} = s_{i-S+h}$ | $I_{i+h} = I_i$ $b_{i+h} = hb_i$ $s_{i-S+h} = (s_{i-S+h} - 1)(I_{i+h} + b_{i+h})$ |
| Ad | $I_{i+h} = I_i$ $b_{i+h} = \sum_{j=1}^h \phi^j b_i$ | $I_{i+h} = I_i$ $b_{i+h} = \sum_{j=1}^h \phi^j b_i$ $s_{i-S+h} = s_{i-S+h}$ | $I_{i+h} = I_i$ $b_{i+h} = \sum_{j=1}^h \phi^j b_i$ $s_{i-S+h} = (s_{i-S+h} - 1)(I_{i+h} + b_{i+h})$ |
| M | $I_{i+h} = I_i$ $b_{i+h} = (b_i^h - 1)I_{i+h}$ | $I_{i+h} = I_i$ $b_{i+h} = (b_i^h - 1)I_{i+h}$ $s_{i-S+h} = s_{i-S+h}$ | $I_{i+h} = I_i$ $b_{i+h} = (b_i^h - 1)I_{i+h}$ $s_{i-S+h} = (s_{i-S+h} - 1)(I_{i+h} + b_{i+h})$ |
| Md | $I_{i+h} = I_i$ $b_{i+h} = (b_i \sum_{j=1}^h \phi^j - 1)I_{i+h}$ | $I_{i+h} = I_i$ $b_{i+h} = (b_i \sum_{j=1}^h \phi^j - 1)I_{i+h}$ $s_{i-S+h} = s_{i-S+h}$ | $I_{i+h} = I_i$ $b_{i+h} = (b_i \sum_{j=1}^h \phi^j - 1)I_{i+h}$ $s_{i-S+h} = (s_{i-S+h} - 1)(I_{i+h} + b_{i+h})$ |

Σχήμα 5.4 : Πρόβλεψη συνιστωσών στην περίπτωση αθροιστικής διαμόρφωσης

Οι τύποι του παραπάνω πίνακα προκύπτουν τετριμμένα από τη διαμόρφωση προβλέψεων εκθετικής εξομάλυνσης πολλαπλών βημάτων μπροστά και βασίζονται στην άθροιση της πληροφορίας κάθε επιπέδου στην πολλαπλασιαστική τάση και εποχιακότητα. Σημειώνεται ότι η βελτιστοποίηση των παραμέτρων και ο υπολογισμός των ανανεωμένων καταστάσεων πραγματοποιούνται σύμφωνα με την αρχική διατύπωση εκθετικής εξομάλυνσης (Hyndman 2008), ανεξάρτητα από το κάθε επίπεδο συνάθροισης. Όταν οι συνιστώσες μετατραπούν σε αθροιστικές, οι παράμετροι του μοντέλου σε κάθε επίπεδο δεν είναι πλέον χρήσιμες. Εφ' εξής χρησιμοποιούνται μόνο οι ισοδύναμες αθροιστικές τιμές των συνιστωσών σε κάθε σημείο t . Η μετατροπή αυτή χρησιμοποιείται μόνο για την παραγωγή των εκτός δείγματος προβλέψεων των συνιστωσών.

Παρά το γεγονός ότι η φύση της αλληλεπίδρασης του σφάλματος με τις υπόλοιπες συνιστώσες (αθροιστικές ή πολλαπλασιαστικές) είναι κρίσιμη για τη σωστή βελτιστοποίηση της εξομάλυνσης και για την ανανέωση του διανύματος κατάστασης της, δεν επηρεάζει τις εκτός δείγματος προβλέψεις των συνιστωσών και ο παραπάνω πίνακας μπορεί να χρησιμοποιηθεί και στις δύο περιπτώσεις. Όταν οι συνιστώσες μετατραπούν σε αθροιστικές μπορούμε να προχωρήσουμε στο επόμενο βήμα του αλγόριθμου.

Οι προβλεπόμενες τιμές της $Y^{[k]}$ προκύπτουν με απλή άθροιση των προβλεπόμενων συνιστωσών, δηλαδή :

$$\hat{Y}_{i+h[k]}^{[k]} = \hat{I}_{i+h[k]}^{[k]} + \hat{D}_{i+h[k]}^{[k]} + \hat{S}_{i-S+h[k]}^{[k]} \quad (5.2)$$

Η παραγωγή προβλέψεων διαφορετικών επιπέδων συνάθροισης για μια δοσμένη χρονική περίοδο έχει ως έξοδο διαφορετικές μήκους τιμών. Για παράδειγμα, αν η αρχική χρονοσειρά $Y^{[1]}$ έχει μηνιαία συχνότητα, τότε η πρόβλεψη για έναν χρόνο μπροστά απαιτεί 12 τιμές προβλέψεων ενώ για τη συναθροισμένη χρονοσειρά $Y^{[12]}$ απαιτείται μόνο μία τιμή για την ίδια πρόβλεψη. Προκειμένου να διευκολυνθεί ο συνδυασμός των προβλέψεων, είναι επιθυμητό να έχουμε όλες τις συναθροισμένες προβλέψεις εκφρασμένες στην χρονική κλίμακα της αρχικής χρονοσειράς $Y^{[1]}$. Για κάθε τιμή $z_i^{[k]}$ αυτό επιτυγχάνεται κάνοντας χρήση της παρακάτω σχέσης :

$$z_i^{[k]} = \sum_{j=1}^k \omega \cdot z_i^{[k]} , \text{ όπου } t = 1, 2, \dots, n \text{ και } i = \lceil t/k \rceil \quad (5.3)$$

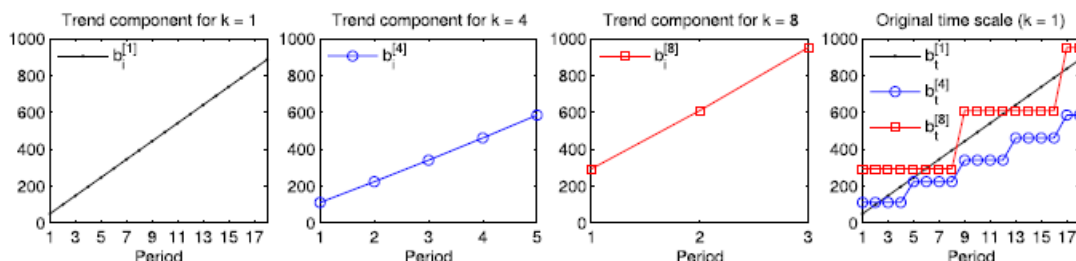
Ουσιαστικά, η παραπάνω εξίσωση διαμοιράζει την i -τιμή σε k παρατηρήσεις στην αρχική χρονική βάση λειτουργώντας ως τμηματικά συνεχής παρεμβολή.

Αυτή η προσέγγιση είναι πλήρως ισοδύναμη με τον απευθείας υπολογισμό σημειακών προβλέψεων στο επίπεδο συνάθροισης και την εν-συνεχεία αναφορά των προβλέψεων στο αρχικό επίπεδο συνάθροισης μέσω επιμερισμού με ίσα βάρη $\omega_j = 1/k$ (Νικολόπουλος (2011), Spinthourakis (2011)).

Εφαρμόζοντας την παραπάνω εξίσωση σε όλα τα $I_i^{[k]}$, $b_i^{[k]}$ και $s_i^{[k]}$ οδηγεί σε έως k διαφορετικές εκτιμήσεις των συνιστωσών της χρονοσειράς στην αρχική χρονική βάση. Η παρακάτω εικόνα απεικονίζει την εκτιμώμενη συνιστώσα της τάσης της χρονοσειράς.

Η συνιστώσα της τάσης που εκτιμάται σε επίπεδα συγκέντρωσης $k = (1, 4, 8)$ εμφανίζεται και στην αρχική χρονική κλίμακα. Σημειώνεται ότι οι εκτιμώμενες συνιστώσες της τάσης για

$k = (1, 4, 8)$ έχουν διαφορετική κλίση και πλήθος παρατηρήσεων και μπορούν να είναι αθροιστικές ή πολλαπλασιαστικές, αποσβενύμενες ή όχι. Κάνοντας χρήση των σχέσεων του πίνακα και της σχέσης $z_t^{[k]} = \sum_{j=1}^k \omega \cdot z_t^{[k]}$ οι εκτιμώμενες συνιστώσες μετατρέπονται σε αθροιστικές και επαναφέρονται στην χρονική κλίμακα της αρχικής χρονοσειράς.



Σχήμα 5.5 : Εκτιμώμενη συνιστώσα τάσης για $k = (1, 4, 8)$ και στην αρχική χρονική κλίμακα (Πηγή : *International Journal of Forecasting* 30 (2014)/ N. Kourentzes et al.)

Σε κάποιες περιπτώσεις, κάποιες συνιστώσες εκθετικής εξομάλυνσης δεν εφαρμόζονται σε όλα τα επίπεδα συνάθροισης. Για παράδειγμα, η εκθετική εξομάλυνση μπορεί να εντοπίσει τη συνιστώσα της τάσης σε κάποια επίπεδα συνάθροισης ενώ σε κάποια άλλα όχι. Αν δεν υπάρχει η εκτίμηση κάποιας συγκεκριμένης συνιστώσας σε κάποιο επίπεδο συνάθροισης, τότε οι προβλέψεις αυτής της συνιστώσας θέτονται ίσες με μηδέν. Η συλλογιστική που κρύβεται πίσω από αυτή την επιλογή είναι ότι αν κάποια συνιστώσα δεν λαμβάνεται υπ όψιν, τότε αυτή η συνιστώσα δεν μπορεί να προσδιοριστεί στην χρονοσειρά και δεν έχει κάποιον αντίκτυπο στις προβλέψεις. Αυτό μεταφράζεται σε μηδενική αθροιστική συνεισφορά της συγκεκριμένης συνιστώσας στις προβλέψεις.

Συνοψίζοντας, στο συγκεκριμένο στάδιο του αλγορίθμου MAPA εφαρμόζονται μοντέλα εκθετικής εξομάλυνσης σε κάθε επίπεδο συνάθροισης και εξάγονται οι συνιστώσες του συγκεκριμένου επιπέδου. Αυτές οι συνιστώσες τροποποιούνται κατάλληλα ώστε να είναι αθροιστικές και εκφρασμένες στην αρχική χρονική κλίμακα. Τελικά η έξοδος αυτού του βήματος του αλγορίθμου είναι σύνολα συνιστωσών επιπέδου τάσης και εποχιακότητας από κάθε επίπεδο συνάθροισης, έτοιμα να συνδυαστούν στο τρίτο και τελευταίο στάδιο του αλγορίθμου MAPA, που περιγράφεται παρακάτω.

5.2.3.2.3 Στάδιο του συνδυασμού

Στο τρίτο και τελευταίο στάδιο του αλγορίθμου MAPA συνδυάζονται οι συνιστώσες που υπολογίστηκαν προηγουμένως με στόχο την παραγωγή εκτιμήσεων για την τάση, το επίπεδο και την εποχιακότητα, οι οποίες μεταγενέστερα θα χρησιμοποιηθούν για την παραγωγή των τελικών προβλέψεων. Δεδομένου ότι η συνάθροιση των χρονοσειρών έγινε κάνοντας χρήση του αριθμητικού μέσου όρου, δεν χρειάζεται περαιτέρω ρύθμιση του επιπέδου των συνιστωσών που προέκυψαν για τα διάφορα επίπεδα συνάθροισης.

Στην βιβλιογραφία έχει αποδειχτεί ότι δύο μέθοδοι συνδυασμού που αποδίδουν ικανοποιητικά είναι ο μη σταθμισμένος μέσος όρος και η διάμεση τιμή (Agnew (1985), Clemen (1989), Clements & Hendry (2007), Miller, Clemen και Winkler (1992), Palm & Zellner (1992), Timmermann (2006)).

Στην περίπτωση του μη σταθμισμένου μέσου όρου, οι συνιστώσες συνδυάζονται ως εξής:

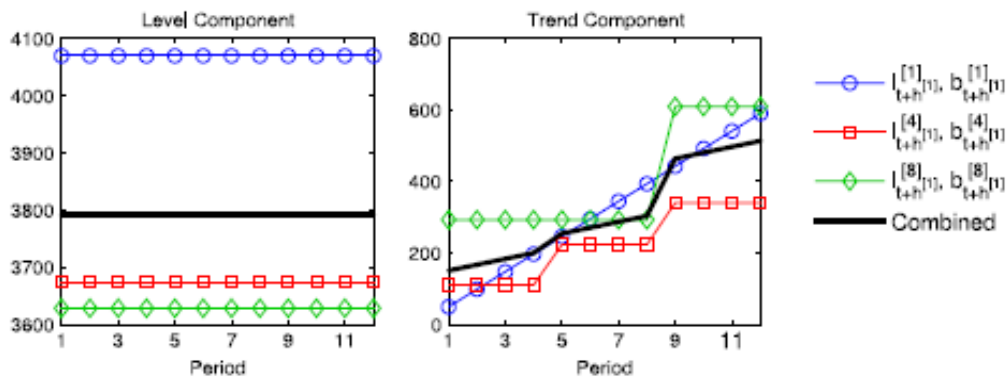
$$\hat{I}_{t+h[1]}^{[k]} = K^{-1} \sum_{k=1}^K (I_{t+h[1]}^{[k]}) \quad (5.4)$$

$$\hat{b}_{t+h[1]}^{[k]} = K^{-1} \sum_{k=1}^K (b_{t+h[1]}^{[k]}) \quad (5.5)$$

$$\hat{S}_{t+h[1]}^{[k]} = K'^{-1} \sum_{k=1}^{K'} (S_{t+h[1]}^{[k]}) \quad , \text{αν } (S/k) \in \square \text{ και } k < S \quad (5.6)$$

Στις παραπάνω σχέσεις, K είναι το μέγιστο επίπεδο συνάθροισης και K' είναι είναι το πλήθος των επιπέδων συνάθροισης, στα οποία διακρίνεται η συνιστώσα της εποχιακότητας.

Ο συνδυασμός των συνιστωσών παρουσιάζεται στο παρακάτω σχήμα:



Σχήμα 5.6 : Συνιστώσες επιπέδου και τάσης για $k = (1, 4, 8)$ και οι προκύπτουσες συνδυασμένες συνιστώσες (Πηγή : *International Journal of Forecasting* 30 (2014)/ N. Kourentzes et al.)

Παρουσιάζονται οι εκτιμήσεις για τις συνιστώσες μιας χρονοσειράς με τάση για επίπεδα συνάθροισης $k = (1, 4, 8)$ και για τις συνδυασμένες προκύπτουσες συνιστώσες. Παρά το γεγονός ότι στα παραδείγματα που παρουσιάζονται, διακρίνονται ολιγαριθμα επίπεδα συνάθροισης, προτείνεται η κατασκευή συνδυασμένων συνιστωσών ΜΑΡΑ από πολλαπλά συνεχόμενα επίπεδα συνάθροισης. Με τον τρόπο αυτό, επιτυγχάνεται καλύτερη εξομάλυνση των συνδυασμένων συνιστωσών. Εναλλακτικά, ερευνούμε την χρήση της ενδιάμεσης τιμής αντί του μέσου όρου για την κατασκευή των $\hat{I}_{t+h[1]}^{[k]}$, $\hat{b}_{t+h[1]}^{[k]}$, $\hat{S}_{t+h[1]}^{[k]}$. Η ενδιάμεση τιμή είναι πιο ισχυρή σε ακραίες τιμές. Ως εκ τούτου, επηρεάζεται λιγότερο από ανεπαρκώς εκτιμημένες συνιστώσες.

Αν η εκθετική εξομάλυνση που εφαρμόζεται σε διαφορετικά επίπεδα συγκέντρωσης προσδιορίσει διαφορετικές συνιστώσες, η ενδιάμεση τιμή θα ευνοήσει την πιο «δημοφιλή/συχνά παρατηρούμενη» συνιστώσα. Αντίθετα η μέση τιμή λαμβάνει εξίσου υπ' όψιν όλες τις συνιστώσες.

Για την παραγωγή, λοιπόν, των τελικών προβλέψεων για $h^{[1]}$ βήματα μπροστα (ορίζοντας πρόβλεψης της αρχικής χρονοσειράς), απλά προσθέτουμε τις συνδυασμένες συνιστώσες ως εξής:

$$\hat{y}_{t+h[1]}^{[1]} = \hat{I}_{t+h[1]} + \hat{b}_{t+h[1]} + \hat{S}_{t-S+h[1]} \quad (5.7)$$

Αξίζει να σημειωθεί, ότι αν κάποια συνιστώσα λείπει και για τα K επίπεδα συνάθροισης, τότε η συμβολή της στην πρόβλεψη $\hat{y}_{t+h[1]}^{[1]}$ είναι μηδενική. Για παραδειγμα, σε μια μη εποχιακή χρονοσειρά, η προβλεπόμενη εποχιακή συνιστώσα είναι πάντα μηδέν για κάθε επίπεδο k . Προφανώς το ίδιο ισχύει και για την συνδυασμένη εποχιακή συνιστώσα $\hat{s}_{t-s+h[1]}$.

5.3 Προβλέψεις Ιεραρχικών Χρονοσειρών

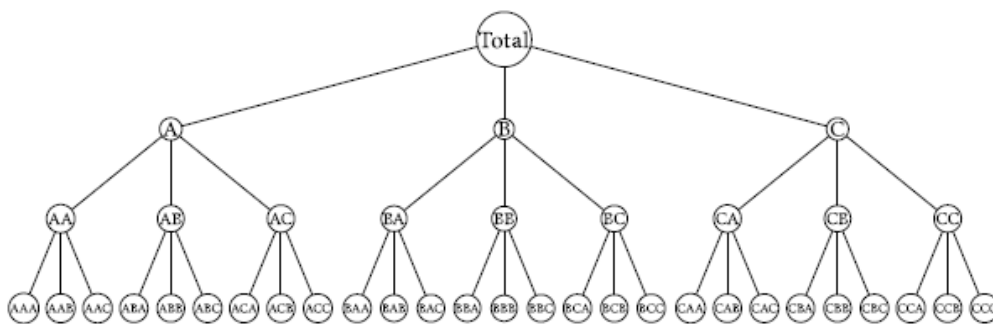
Στον τομέα των οικονομικών και των επιχειρήσεων, παρουσιάζονται συχνά εφαρμογές, οι οποίες απαιτούν οι προβλέψεις πολλών χρονοσειρών, που σχετίζονται μεταξύ τους, να οργανωνονται σε μια ιεραρχική δομή βάσει των διακριτών χαρακτηριστικών τους (π.χ. προϊόντα, γεωγραφία ή κάποια άλλα χαρακτηριστικά). Το γεγονός αυτό δημιούργησε την ανάγκη να βρεθεί ένας τρόπος με τον οποίο να γεφυρώνονται οι προβλέψεις των διαφόρων επιπέδων κατά μήκος της ιεραρχίας χρονοσειρών, με την προϋπόθεση ότι οι προβλέψεις αθροίζονται κατάλληλα μεταξύ των επιπέδων.

5.3.1 Γενικές Ιεραρχικές Προβλέψεις

Έστω μια ιεραρχία πολλών επιπέδων. Το επίπεδο 0 δηλώνει την πλήρως συναθροισμένη σειρά, το επίπεδο 1 το πρώτο επίπεδο διαχωρισμού, το επίπεδο 2 το δεύτερο επίπεδο διαχωρισμού κ.ο.κ. ως το επίπεδο K που αποτελεί το πιο αναλυτικό επίπεδο χρονοσειρών.

Χρησιμοποιείται μια αλληλουχία γραμμάτων για τον εντοπισμό συγκεκριμένης σειράς και του επίπεδο διαχωρισμού. Για παράδειγμα: Ο δείκτης «A» δηλώνει την σειρά A στο επίπεδο 1, ο δείκτης «AF» δηλώνει την σειρά F στο 2^ο επίπεδο εντός της σειράς A στο 1^ο επίπεδο. Η σειρά AFC υποδηλώνει σειρά C στο 3^ο επίπεδο εντός της σειράς AF στο 2^ο επίπεδο και ούτω καθεξής.

Για να είμαστε συγκεκριμένοι, ας υποθέσουμε ότι είχαμε τρία επίπεδα στην ιεραρχία με κάθε ομάδα σε κάθε επίπεδο να αποτελείται από τρεις σειρές. Στην περίπτωση αυτή, $K = 3$ και η ιεραρχία έχει τη δομή δέντρου φαίνεται παρακάτω:



Σχήμα 5.7 : Παράδειγμα Ιεραρχίας

Έστω ότι οι παρατηρήσεις καταγράφονται σε χρόνους $t = 1, 2, \dots, N$, και ότι μας ενδιαφέρει πρόβλεψη κάθε σειράς σε κάθε επίπεδο σε χρόνους $t = n + 1, n + 2, \dots, N + h$. Για λόγους ευκολίας, χρησιμοποιούμε τον συμβολισμό X για να αναφερθούμε σε μια γενική σειράς στο εσωτερικό της ιεραρχίας. Οι παρατηρήσεις σχετικά με την σειρά X γράφονται ως $Y_{X,t}$. Έτσι,

$Y_{AF, t}$ είναι η τιμή της σειράς AF στο χρόνο t . Χρησιμοποιούμε τον συμβολισμό Y_t για το σύνολο όλων των σειρών στο χρόνο t . Ως εκ τούτου:

$$Y_t = \sum_i Y_{i,t}, \quad Y_{i,t} = \sum_j Y_{ij,t}, \quad Y_{ij,t} = \sum_k Y_{ijk,t}, \quad Y_{ijk,t} = \sum_l Y_{ijkl,t} \quad (5.8)$$

Έτσι, οι παρατηρήσεις σε υψηλότερα επίπεδα μπορούν να ληφθούν αθροίζοντας την κατωτέρη σειρά. Αν η μεταβλητή m_i χαρακτηρίζει το συνολικό αριθμό των σειρών στο επίπεδο i , $i = 0, 1, 2, \dots, K$, τότε $m_i > m_{i-1}$ και ο συνολικός αριθμός των σειρών στην ιεραρχία είναι $m = m_0 + m_1 + m_2 + \dots + m_K$. Στο παραπάνω παράδειγμα, $m_i = 3$ και $m = 40$. Θα ήταν βολικό να εργαστούμε με εκφράσεις διανυσμάτων και μητρών. Έστω ότι $Y_{i,t}$ δηλώνει το διάνυσμα όλων των παρατηρήσεων σε επίπεδο i και χρόνο t και $Y_t = [Y_t, Y_{1,t}, \dots, Y_{K,t}]$.

Σημειώνεται ότι:

$$Y_t = S \cdot Y_K \quad (5.9)$$

,όπου S είναι μία " αθροίζουσα " μήτρα τάξης $m \times m_K$ που χρησιμοποιείται για τη συνάθροιση των σειρών κατώτερου επιπέδου. Στο παραπάνω παράδειγμα, ισχύει ότι :

$$Y_t = \sum (Y_t, Y_{a,t}, Y_{b,t}, Y_{c,t}, Y_{AA,t}, \dots, Y_{CC,t}, Y_{AAA,t}, Y_{AAB,t}, \dots, Y_{CCC,t}) \quad (5.10)$$

και η μήτρα συνάθροισης είναι τάξης 40×27 και δίνεται ως:

$$S = \begin{bmatrix} 11111111111111111111111111111111 \\ 1111111110000000000000000000 \\ 0000000001111111110000000000 \\ 0000000000000000000111111111 \\ 1110000000000000000000000000 \\ 0000111000000000000000000000 \\ \vdots \\ 0000000000000000000000000111 \\ 1000000000000000000000000000 \\ 0100000000000000000000000000 \\ \vdots \\ 00000000000000000000000000001 \end{bmatrix}$$

Ο βαθμός της μήτρας S είναι m_K . Είναι σαφές ότι η μήτρα S μπορεί να τμηματοποιηθεί ανάλογα τα επίπεδα της ιεραρχίας. Η κορυφαία σειρά είναι ένα μοναδιαίο διάνυσμα μήκους m_K και το κάτω τμήμα είναι ένας τετραγωνικός πίνακας $m_K \times m_K$. Τα μεσαία τμήματα του S είναι διαγώνιοι τετραγωνικοί πίνακες.

Ας υποθέσουμε ότι πρώτα υπολογίζουμε τις προβλέψεις για κάθε σειρά σε κάθε επίπεδο και έχουμε τις προβλέψεις βάσης m για κάθε μία από τις περιόδους $n + 1, \dots, n + h$, με βάση τις διαθέσιμες πληροφορίες μέχρι και του χρόνου n . Συμβολίζουμε αυτές τις προβλέψεις βάσης με $Y_{X,n}(h)$, όπου το X σημαίνει η σειρά που προβλέπεται. Έτσι, η έκφραση $Y_n(h)$ υποδηλώνει την πρόβλεψη κατά h -βήμα μπροστά του συνόλου των σειρών, η έκφραση $Y_{A,n}(h)$ δηλώνει την πρόβλεψη της σειράς A , η έκφραση $Y_{AC,n}(h)$ υποδηλώνει την πρόβλεψη της σειράς AC , και ούτω καθεξής. Έστω $\hat{Y}_n(h)$ το διάνυσμα που αποτελείται

από αυτές τις προβλέψεις στην ίδια σειρά με το Y_t . Τότε όλες οι υπάρχουσες μέθοδοι ιεραρχικής πρόβλεψης μπορούν να γραφτούν ως :

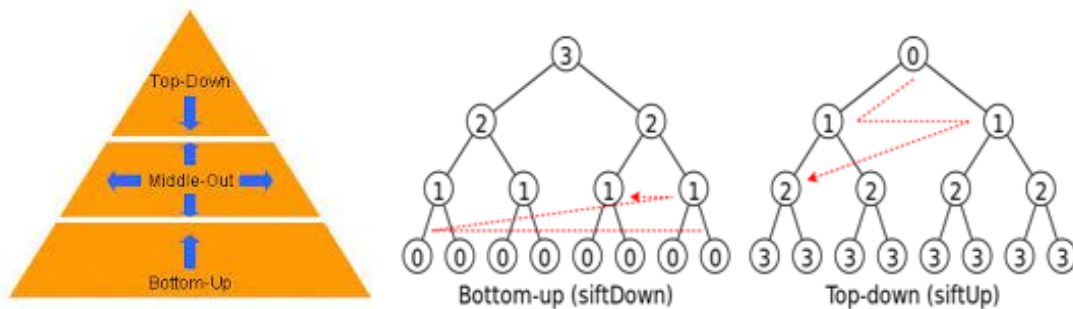
$$\tilde{Y}_n(h) = SP \hat{Y}_n(h) \quad (5.11)$$

, για έναν κατάλληλα επιλεγμένο πίνακα P διαστάσεων $m_k \times m$. Αυτό προκύπτει από το γεγονός ότι όλες οι υπάρχουσες μέθοδοι αποτελούνται από κάποιον γραμμικό συνδυασμό αυτών των προβλέψεων. Αυτοί οι γραμμικοί συνδυασμοί είναι σύμφωνοι με το γεγονός ότι οι προβλέψεις των χαμηλότερων επιπέδων αθροίζονται για να δώσουν προβλέψεις ανώτερων επιπέδων. Η επίδραση της μήτρας P είναι να εξαγάγει και να συνδυάσει τα βασικά στοιχεία των προβλέψεων $\hat{Y}_n(h)$ οι οποίες εν συνεχεία αθροίζονται μέσω του S για να δώσουν τις τελικές αναθεωρημένες ιεραρχικές προβλέψεις $\tilde{Y}_n(h)$.

5.3.2 Μέθοδοι ιεραρχικής συνάθροισης

Οι περισσότερες υπάρχουσες προσεγγίσεις της ιεραρχικής πρόβλεψης συνήθως περιλαμβάνουν μία μέθοδο «top-down», μία μέθοδο «bottom-up» ή ένα συνδυασμό αυτών των δύο μεθόδων. Η μέθοδος «bottom-up» έγκειται στην πρόβλεψη όλων των αποσυνθεμένων χρονοσειρών του χαμηλότερου επιπέδου ιεραρχίας και στην απλή σύνθεση αυτών με στόχο να ληφθούν προβλέψεις για τα ανώτερα επίπεδα της ιεραρχίας.

Στην πράξη, πολλές επιχειρήσεις συνδυάζουν αυτές τις μεθόδους (δίνοντας αυτό που μερικές φορές ονομάζεται "middle-out" μέθοδος), όπου οι προβλέψεις λαμβάνονται για κάθε σειρά σε ένα ενδιάμεσο επίπεδο της ιεραρχίας, και στη συνέχεια πραγματοποιείται σύνθεση για τη λήψη προβλέψεων σε υψηλότερα επίπεδα και απόσυνθεση για τη λήψη προβλέψεων σε χαμηλότερα επίπεδα. Καμία από αυτές τις μεθόδους δεν λαμβάνει υπόψη την υπάρχουσα συσχετιστική δομή της ιεραρχίας, και επιπλέον δεν είναι εύκολο να λάβουμε διαστήματα πρόβλεψης για τις προβλέψεις μας με καμία από αυτές τις μεθόδους.



Σχήμα 5.8 : Σχηματική απεικόνιση ιεραρχίας και των προσεγγίσεων Top-Down, Bottom-Up & Middle-Out

Φυσικά, είναι δυνατόν να προβλεφθούν ανεξάρτητα όλες οι χρονοσειρές σε όλα τα επίπεδα της ιεραρχίας με το μειονέκτημα ότι οι προβλέψεις των υψηλότερων επιπέδων δεν είναι ίσες με το άθροισμα των προβλέψεων των χαμηλότερων επιπέδων ιεραρχίας. Συνεπώς αν επιλεγθεί αυτή η μέθοδος, διεξάγονται εν συνεχεία κάποιες προσαρμογές ώστε να εξασφαλιστεί ότι οι προβλέψεις συναθροίζονται κατάλληλα. Συνήθως πρόκειται για ad-hoc προσαρμογές.

Στην βιβλιογραφία υπάρχουν έρευνες που εξετάζουν συγκριτικά τις μεθόδους «bottom-up» και «top-down». Από το 1960 οι Grunfeld και Griliches ισχυρίστηκαν ότι τα αναλυτικά δεδομένα είναι επιρρεπή σε λάθη και επομένως η μέθοδος «top-down» μπορεί να παρουσιάζει πιο ακριβείς προβλέψεις. Παρόμοια συμπεράσματα αντλούνται αργότερα από τις έρευνες των Fogarty et al. (1990) Narasimhan (1995) και Fliedner (1999) που υποστήριξαν, ότι οι επιδόσεις των συναθροισμένων προβλέψεων είναι καλύτερες με δεδομένα συναθροισμένα σε επίπεδα. Από την άλλη πλευρά, οι Orcutt et al. (1968) και Edwards και Orcutt (1969) υποστήριξαν ότι η απώλεια πληροφοριών είναι σημαντική κατά την συσσωμάτωση και ως εκ τούτου η μέθοδος bottom-up δίνει πιο ακριβείς προβλέψεις.

Οι Tiao και Guttman (1980) και Kohn (1982), χρησιμοποίησαν θεωρητικά επιχειρήματα για να αποδείξουν ότι η αποτελεσματικότητα της συνάθροισης εξαρτάται από τη δομή συνδιασποράς των χρονοσειρών. Ο Shing (1993) κατέδειξε ότι δεν υπάρχει ενιαία ανωτερότητα μιας μεθόδου έναντι της άλλης. Οι Fliedner και Lawrence (1995) κατέληξε στο συμπέρασμα ότι οι τρέχουσες επίσημες ιεραρχικές τεχνικές πρόβλεψης δεν έχουν κανένα πλεονέκτημα σε σχέση με ορισμένες άτυπες στρατηγικές της ιεραρχικής πρόβλεψης. Ο Kahn (1998) πρότεινε ότι είναι καιρός να συνδυάσει τις υπάρχουσες μεθοδολογίες, ώστε να μπορούμε να «επωφεληθούμε» από τα πλεονεκτήματα και των δύο μεθόδων. Η έρευνα του Fliedner (2001) συνοψίζει τις χρήσεις και τις κατευθυντήριες γραμμές εφαρμογής των ιεραρχικών πρόβλεψης.

5.3.3 Bottom-up προβλέψεις

Η μέθοδος αυτή χρησιμοποιείται ευρέως και προϋποθέτει την παραγωγή προβλέψεων στο κατώτατο επίπεδο της ιεραρχίας και στη συνέχεια τη συνάθροιση των σειρών του χαμηλότερου επιπέδου προς τα πάνω ώστε να προκύψουν οι προβλέψεις και για τις συνολικές σειρές και την πλήρη ιεραρχία.

Οι bottom-up προβλέψεις λαμβάνονται χρησιμοποιώντας :

$$P = [0_{mk \times (mk \times mk)} \mid I_{mk}] \quad (5.12)$$

,όπου $0_{l \times k}$ είναι μηδενική μήτρα διαστάσεων $l \times k$ και I_k είναι μοναδιαία μήτρα ταυτότητα μήκους $k \times k$. Στην περίπτωση αυτή, ο πίνακας P εξάγει μόνο τις προβλέψεις χαμηλότερου επιπέδου από το $\hat{Y}_n(h)$ οι οποίες στη συνέχεια αθροίζονται από την S για να δώσουν τις bottom-up προβλέψεις. Στην Bottom Up μεθοδολογία επομένως, οι προβλέψεις του κατώτατου επιπέδου ταυτίζονται με τις βασικές.

Αν υποθέσουμε ότι οι ανεξάρτητες προβλέψεις είναι αμερόληπτες ,δηλαδή :

$$E [\hat{Y}_n(h)] = E[Y_n(h)] \quad (5.13)$$

και ότι θέλουμε οι αναθεωρημένες ιεραρχικές προβλέψεις να είναι επίσης αμερόληπτες, τότε θα πρέπει να απαιτήσουμε : $[\check{Y}_n(h)] = E[Y_n(h)] = S \cdot E[Y_{k,n}(h)]$. (5.14)

Έστω ότι $\beta_n(h) = E [Y_{k, n+h} \mid Y_1, \dots, Y_n]$ είναι η μέση τιμή των μελλοντικών αξιών του κάτω επιπέδου k . Τότε $E [\check{Y}_n(h)] = SPE [Y_n(h)] = SPS\beta_n(h)$. Συνεπώς, η αμεροληψία των

αναθεωρημένων προβλέψεων θα παρέχεται αρκεί να τηρείται η συνθήκη : $SPS = S$. Αυτό ισχύει για τη μέθοδο bottom-up με τον πίνακα P να δίνεται από την (5.12).

Το μεγαλύτερο πλεονέκτημα της μεθόδου αυτής είναι ότι δεν έχουμε απώλεια δεδομένων από τη συνάθροιση. Ο υψηλός θόρυβος όμως στις σειρές του κατώτατου επιπέδου κάνει την καλή επιλογή μοντέλου δύσκολη για το επίπεδο αυτό.

5.3.4 Top-down προβλέψεις

Η μέθοδος Top-down περιλαμβάνει αρχικά παραγωγή βασικών προβλέψεων για την συνολική χρονοσειρά y_t στην κορυφή της ιεραρχίας και τον εν-συνεχεία επιμερισμό της προς τα χαμηλότερα επίπεδα.

Όταν προκύψουν οι προβλέψεις και των χαμηλότερων επιπέδων μπορούμε με τη χρήση του πίνακα άθροισης να παράγουμε προβλέψεις και για τα υπόλοιπα ενδιάμεσα επίπεδα. Στην μεθοδολογία Top Down επομένως, οι προβλέψεις του ανώτατου επιπέδου ταυτίζονται με τις βασικές.

Οι top-down προβλέψεις λαμβάνονται χρησιμοποιώντας :

$$P = [p \quad | \quad O_{mk \times (mk \times mk)}] \quad (5.15)$$

,όπου $P=[P_1, P_2, \dots, P_{mk}]$ είναι ένα διάνυσμα αναλογιών που αθροίζονται σε μία. Στην συγκεκριμένη περίπτωση, ο ρόλος του πίνακα P είναι να διανείμει την συναθροισμένη πρόβλεψη σε προβλέψεις χαμηλότερου επιπέδου. Διαφορετικές μέθοδοι top-down πρόβλεψης οδηγούν σε διαφορετικό πίνακα αναλογιών P.

Αν υποθέσουμε ότι οι ανεξάρτητες προβλέψεις είναι αμερόληπτες ,δηλαδή :

$$E [\hat{Y}_n (h)] = E [Y_n (h)] \quad (5.16)$$

και ότι θέλουμε οι αναθεωρημένες ιεραρχικές προβλέψεις να είναι επίσης αμερόληπτες, τότε θα πρέπει να απαιτήσουμε : $E [\check{Y}_n (h)] = E [Y_n (h)] = S \cdot E [Y_{k,n}(h)]$. (5.17)

Έστω ότι $\beta_n(h) = E [Y_{k, n+h} | Y_1, \dots, Y_n]$ είναι η μέση τιμή των μελλοντικών αξιών του κάτω επιπέδου K. Τότε $E [\check{Y}_n(h)] = SPE [Y_n (h)] = SPS\beta_n (h)$. Συνεπώς, η αμεροληψία των αναθεωρημένων προβλέψεων θα παρέχεται αρκεί να τηρείται η συνθήκη :

$$S \cdot P \cdot S = S \quad (5.18)$$

Χρησιμοποιώντας τη μέθοδο top-down με P να δίνεται από την (5.15), διαπιστώνουμε ότι το $SPS \neq S$ για κάθε επιλογή του P. Έτσι, η μέθοδος top-down δεν μπορεί να δώσει αμερόληπτες προβλέψεις, ακόμη και αν οι ανεξάρτητες προβλέψεις είναι αμερόληπτες.

Έστω ότι η διακύμανση των ανεξάρτητων προβλέψεων βάσης, $Y_n (h)$, να δοθεί από Σ_h . Στη συνέχεια, η διακύμανση των αναθεωρημένων προβλέψεων δίνεται από

$$\text{Var} [\check{Y}_n(h)] = SP\Sigma_h P'S' \quad (5.19)$$

Έτσι, μπορούμε να λάβουμε τα διαστήματα πρόβλεψης σχετικά με τις αναθεωρημένες προβλέψεις με την προϋπόθεση ότι η Σ_h μπορεί να εκτιμηθεί αξιόπιστα. Η έκφραση (5.18),

ισχύει για όλες τις υπάρχουσες μεθόδους που μπορούν να εκφραστούν με την έκφραση (5.9), συμπεριλαμβανομένων των μεθόδων bottom-up, top-down και middle-out.

Η συνηθέστερη προσέγγιση τέτοιου τύπου επιμερίζει τη σειρά του κορυφαίου επιπέδου με βάρη, τα οποία βασίζονται σε ιστορικά δεδομένα και προκύπτουν με τον παρακάτω τρόπο. Η κάθε αναλογία (βάρος) w_j αντιστοιχεί στο μέσο όρο των ιστορικών αναλογιών του κατώτατου επιπέδου $y_{j,t}$ για την περίοδο $t = 1, \dots, T$ σε σχέση με τη συνολική αθροιστική σειρά Y_t .

$$\text{Μέση ιστορική αναλογία (βάρος) ανά επίπεδο: } w_j = \frac{1}{T} \sum_{t=1}^T \frac{Y_{j,t}}{Y_t} \quad , \quad j = 1, \dots, m_k \quad (5.20)$$

Εναλλακτικά, μπορούμε να κάνουμε χρήση αναλογιών (βαρών) με βάση τους ιστορικούς

$$\text{μέσους όρους σύμφωνα με τον τύπο : } w'_j = \frac{\sum_{t=1}^T \frac{Y_{j,t}}{T}}{\sum_{t=1}^T \frac{Y_t}{T}} \quad , \quad j = 1, \dots, m_k \quad (5.21)$$

Το σημαντικότερο πλεονέκτημα των top-down μεθόδων είναι η απλότητα και αμεσότητα της εφαρμογής τους, αφού απαιτείται μοντελοποίηση μόνο για τη συνολική σειρά στην κορυφή της ιεραρχίας. Γενικά η προσέγγιση αυτή παράγει ικανοποιητικά σε ακρίβεια αποτελέσματα, ειδικά για δείγματα με μικρό πλήθος παρατηρήσεων. Από την άλλη το μεγαλύτερο μειονέκτημα της μεθόδου αυτής είναι η απώλεια δεδομένων κατά την άθροιση και το γεγονός ότι δεν είμαστε σε θέση να αντιληφθούμε και να εκμεταλλευθούμε χαρακτηριστικά όπως οι δυναμικές και τα special events στην κάθε σειρά ξεχωριστά.

5.3.5 Middle-out Προβλέψεις

Η προσέγγιση αυτή είναι ένας συνδυασμός των δύο προηγούμενων. Αρχικά, επιλέγεται το «μεσαίο» επίπεδο και παράγονται προβλέψεις για όλες τις σειρές που βρίσκονται στο επίπεδο αυτό. Για τις σειρές άνω του μεσαίου επιπέδου, οι προβλέψεις προκύπτουν χρησιμοποιώντας μια bottom-up προσέγγιση, αθροίζοντας προς τα πάνω τις προβλέψεις του μεσαίου επιπέδου. Για τις σειρές χαμηλότερα του μεσαίου επιπέδου, οι προβλέψεις προκύπτουν με επιμερισμό των βασικών προβλέψεων μεσαίου επιπέδου προς τα χαμηλότερα επίπεδα.

5.3.6 Optimal combination forecasts for hierarchical time series

Σε πολλές εφαρμογές, υπάρχουν πολλαπλές χρονοσειρές, οι οποίες οργανώνονται ιεραρχικά και μπορούν να συναθροιστούν σε ομάδες σε διάφορα επίπεδα με βάση τα προϊόντα, τη γεωγραφία ή κάποια άλλα χαρακτηριστικά. Αυτές οι ιεραρχικές χρονοσειρές είναι κοινώς προβλέψεις που χρησιμοποιούν είτε την «top down» είτε την «bottom-up» μέθοδο. Στην βιβλιογραφία έχουν προταθεί διάφορες προσεγγίσεις ως προς την ιεραρχική πρόβλεψη που παρέχουν βέλτιστες προβλέψεις σχετικά με τις παραγόμενες προβλέψεις των μεθόδων «top down» και «bottom-up».

Μία προσέγγιση βασίζεται στην πρόβλεψη όλων των σειρών ανεξάρτητα σε όλα τα επίπεδα της ιεραρχίας και, στη συνέχεια, ο συνδυασμός των προβλέψεων αυτών χρησιμοποιώντας ένα μοντέλο παλινδρόμησης. Οι προκύπτουσες αναθεωρημένες προβλέψεις προστίθενται

κατάλληλα σε όλη την ιεραρχία, είναι αμερόληπτες και παρουσιάζουν ελάχιστη διακύμανση μεταξύ όλων των συνδυασμένων προβλέψεων σύμφωνα με κάποιες απλές υποθέσεις.

Στον κόσμο των επιχειρήσεων και της οικονομίας, υπάρχουν συχνά εφαρμογές που απαιτούν προβλέψεις πολλών σχετικών χρονοσειρών οργανωμένων σε μία ιεραρχική δομή με βάση παράγοντες, όπως το προϊόν και η γεωγραφική θέση. Αυτό έχει οδηγήσει στην ανάγκη για κατάλληλο συνδυασμό των προβλέψεων σε όλα τα επίπεδα ιεραρχίας στην ανάγκη δηλαδή να συνοψιστούν (συντεθούν) κατάλληλα όλες οι προβλέψεις σε όλη την ιεραρχία. Στην βιβλιογραφία προτείνεται μία στατιστική μέθοδος για προβλέψεις ιεραρχικών χρονοσειρών, η οποία :

- (α) παρέχει σημειακές προβλέψεις που μπορούν να συνοψιστούν σε όλα τα επίπεδα ιεραρχίας
- (β) επιτρέπει συσχετισμούς και αλληλεπιδράσεις των χρονοσειρών σε όλα τα επίπεδα ιεραρχίας
- (γ) παρέχει εκτιμήσεις της αβεβαιότητας των προβλέψεων, οι οποίες μπορούν να συνοψιστούν σε όλα τα επίπεδα ιεραρχίας
- (δ) είναι επαρκώς ευέλικτη ώστε:
 - (1) να μπορούν να ενσωματωθούν ad-hoc προσαρμογές
 - (2) να μπορούν να συμπεριληφθούν πληροφορίες σχετικά με τις επιμέρους χρονοσειρές
 - (3) να μπορούν να συμπεριληφθούν σημαντικές συμπληρωματικές μεταβλητές

Τα διάφορα συστατικά της ιεραρχίας μπορούν να αλληλεπιδρούν με ποικίλες και πολύπλοκους τρόπους. Μια αλλαγή σε κάποια χρονοσειρά σε ένα επίπεδο, μπορεί να έχει επιπτώσεις στις άλλες χρονοσειρές στο ίδιο επίπεδο, καθώς και σε σειρές σε υψηλότερα και χαμηλότερα επίπεδα. Με την μοντελοποίηση ολόκληρης της ιεραρχίας των χρονοσειρών ταυτόχρονα, θα λάβουμε βελτιωμένες προβλέψεις των συστατικών χρονοσειρών.

5.3.6.1 Βελτιστοποίηση προβλέψεων χρησιμοποιώντας παλινδρόμηση

Μπορούμε να γράψουμε τις ανεξάρτητες προβλέψεις βάσης ως :

$$Y_n(h) = S\beta_n(h) + \varepsilon_h \quad (5.22)$$

όπου $\beta_n(h) = E[Y_{K,n+h} | Y_1, \dots, Y_n]$ είναι η άγνωστη μέση τιμή του κάτω επιπέδου K και ε_h έχει μηδενική μέση τιμή και μήτρα συνδιακύμανσης $\text{Var}(\varepsilon_h) = \Sigma_h$. Αυτό υποδηλώνει ότι μπορούμε να εκτιμήσουμε τη μέση τιμή $\beta_n(h)$ μέσω της εξίσωσης παλινδρόμησης (5.22), και ως εκ τούτου να ληφθούν προβλέψεις για όλα τα επίπεδα ιεραρχίας.

Αν η Σ_h ήταν γνωστή, θα μπορούσαμε να χρησιμοποιήσουμε την γενικευμένη εκτίμηση ελαχίστων τετραγώνων για να υπολογίσουμε την αμερόληπτη εκτίμηση των $\beta_n(h)$ με την ελάχιστη διακύμανση ως :

$$\hat{\beta}_n(h) = (S' \Sigma_h^+ S)^{-1} S' \Sigma_h^+ \hat{Y}_n(h) \quad (5.23)$$

,όπου η Σ_h^+ είναι ο γενικευμένος αντίστροφος πίνακας Σ_h κατά Moore-Penrose. Χρησιμοποιούμε τον γενικευμένο αντίστροφο, επειδή η Σ_h είναι συχνά (σχεδόν) μη αντιστρέψιμος πίνακας λόγω της συσσωμάτωσης που εμπλέκεται στην Y_n . Αυτό οδηγεί στις ακόλουθες αναθεωρημένες προβλέψεις:

$$\check{Y}_n(h) = S \hat{\beta}_n(h) = SP \hat{Y}_n(h) \quad (5.24)$$

,όπου $P = (S' \Sigma_h^+ S)^{-1} S' \Sigma_h^+$

Σαφώς, αυτό ικανοποιεί την αμεροληψία σύμφωνα με την σχέση (5.13). Η διακύμανση των προβλέψεων αυτών δίνεται από:

$$\text{Var} [\check{Y}_n(h)] = S(S' \Sigma_h^+ S)^{-1} S' \quad (5.25)$$

Η δυσκολία αυτής της μεθόδου είναι ότι απαιτεί γνώση του Σ_h , ή τουλάχιστον μια καλή εκτίμηση του. Σε μια μεγάλη ιεραρχία, με χιλιάδες σειρές, αυτό μπορεί να μην είναι δυνατό. Ωστόσο, μπορούμε να απλοποιήσουμε σημαντικά τους υπολογισμούς με την παραδοχή ότι το σφάλμα μπορεί να εκφραστεί ως $\varepsilon_h \approx S\varepsilon_{k,h}$

όπου $\varepsilon_{k,h}$ είναι το σφάλμα πρόβλεψης στο κάτω επίπεδο K . Δηλαδή, υποθέτουμε ότι τα λάθη ικανοποιούν την ίδια περιορισμό συνάθροισης με τα αρχικά δεδομένα (5.8).

Η υπόθεση αυτή θα ισχύει υπό την προϋπόθεση ότι οι προβλέψεις επίσης ικανοποιούν περίπου τον περιορισμό συνάθροισης, ο οποίος θα πρέπει να ισχύει για οποιοδήποτε λογικό σύνολο προβλέψεων. Σε ορισμένες περιπτώσεις, είναι δυνατόν τα σφάλματα να ικανοποιούν την έκφραση ακριβώς. Για παράδειγμα, εάν μία μέθοδος γραμμικής πρόβλεψης (π.χ., ένα μοντέλο ARIMA) με σταθερές παραμέτρους χρησιμοποιείται για όλες τις σειρές, τότε τα σφάλματα θα είναι ακριβώς αθροιστικά με αυτόν τον τρόπο. Έτσι, η υπόθεση είναι μια λογική προσέγγιση του τι πραγματικά συμβαίνει. Η προσέγγιση οδηγεί στο αποτέλεσμα $\Sigma_h \approx S\Omega_h S'$, όπου $\Omega_h = \text{Var}(\varepsilon_{k,h})$. Είμαστε τώρα έτοιμοι να δηλώσουμε το κύριο αποτέλεσμα μας.

Θεώρημα 1: Έστω $Y = S\beta_h + \varepsilon$ με $\text{Var}(\varepsilon) = \Sigma_h = S\Omega_h S'$ και S μία «αθροίζουσα μήτρα». Στη συνέχεια, η γενικευμένη εκτίμηση ελαχίστων τετραγώνων του β που λαμβάνεται με τη χρήση της γενικευμένης αντιστροφής Moore-Penrose είναι ανεξάρτητη από την Ω_h :

$$\hat{\beta}_n(h) = (S' \Sigma_h^+ S)^{-1} S' \Sigma_h^+ Y_n(h) = (S'S)^{-1} S' Y \quad (5.26)$$

με μήτρα διακύμανσης $\text{Var}(\beta) = \Omega_h$. Επιπλέον, αυτή είναι η γραμμική αμερόληπτη εκτίμηση με την ελάχιστη διακύμανση.

Απόδειξη:

Γράφουμε $\Sigma_h = BC$, όπου $B = S\Omega_h$ και $C = S'$. Στη συνέχεια, ο κατά Moore-Penrose γενικευμένος αντίστροφος της Σ_h είναι:

$$\Sigma_h^+ = C'(CC')^{-1}(B'B)^{-1}B' = S(S'S)^{-1}(\Omega_h'S'S\Omega_h)^{-1}\Omega_h'S' \quad (5.27)$$

Τότε: $(S'\Sigma_h^+S)^{-1}S'\Sigma_h^+ = (S'S)^{-1}S'$

Η διακύμανση λαμβάνεται με αντικατάσταση της (9) σε $(S'\Sigma_h^+S)^{-1}$. Οι Tian και Wiens (2006, Θεώρημα 3) δείχνουν ότι ο εκτιμητής GLS θα είναι αμερόληπτη εκτίμησή ελάχιστης διακύμανσης, αν και μόνο αν:

$$SS^+\Sigma_h^+\Sigma_h(1-SS^+) = 0 \quad (5.28)$$

,όπου: $S^+ = (S'S)^{-1}S'$

Χρησιμοποιώντας την (5.27), είναι εύκολο να δείξουμε ότι η προϋπόθεση αυτή ισχύει. Αυτό το αξιοσημείωτο αποτέλεσμα δείχνει ότι μπορούμε να χρησιμοποιήσουμε OLS αντί του GLS κατά τον υπολογισμό των αναθεωρημένων προβλέψεων μας, χωρίς να υπάρχει ανάγκη εκτίμησης του πίνακα συνδιασποράς. Έτσι διευκολύνεται ο υπολογισμός ειδικά για μεγάλες μήτρες συνδιακύμανσης. Δηλαδή, χρησιμοποιούμε:

$$\check{Y}_n(h) = (S'S)^{-1}S'\hat{Y}_n(h) \quad \text{και} \quad P = (S'S)^{-1}S' \quad (5.29)$$

Η μήτρα συνδιακύμανσης των αναθεωρημένων προβλέψεων είναι:

$$\text{Var}(\check{Y}_n(h)) = \Sigma_h \quad (5.30)$$

Ως εκ τούτου, τα διαστήματα πρόβλεψης εξακολουθούν να απαιτούν την εκτίμηση του Σ_h . Η εξίσωση (5.29) δείχνει ότι, με την παραδοχή $\Sigma_h = S\Omega_hS'$, ο βέλτιστος συνδυασμός των ανεξάρτητων προβλέψεων βάσης είναι ανεξάρτητος από τα δεδομένα. Για μια απλή ιεραρχία με ένα μόνο επίπεδο διαχωρισμού ($K = 1$) και με m_1 κόμβους στο επίπεδο 1, τα βάρη δίνονται ως:

$$S(S'S)^{-1}S' = (m_1 + 1)^{-1} \begin{pmatrix} m_1 & 1 & 1 & \dots & \dots & 1 \\ 1 & m_1 & -1 & -1 & \dots & -1 \\ 1 & -1 & m_1 & -1 & & -1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & -1 & & -1 & \ddots & -1 \\ 1 & -1 & -1 & \dots & -1 & m_1 \end{pmatrix}.$$

Ομοίως, μπορούν να ληφθούν σύνολα βαρών για οποιαδήποτε συγκεκριμένη ιεραρχία, ανεξάρτητα από τα δεδομένα που παρατηρήθηκαν.

5.3.6.2 Υπολογιστικές παγίδες και αντιμετώπιση

Η κύρια δυσκολία στην εφαρμογή της μεθόδου πρόβλεψης που δίδεται από την (5.290), είναι ότι η μήτρα S μπορεί να είναι πολύ μεγάλη, και έτσι και η απαιτούμενη υπολογιστική ισχύς για την εύρεση του αντίστροφου του S μπορεί να είναι απαγορευτική. Ακόμη και η κατασκευή της μήτρας S μπορεί να καταστεί ιδιαίτερα δύσκολη για τις μεγάλες ιεραρχίες. Προτείνονται τρεις λύσεις σε αυτό το πρόβλημα.

Αρχικά, από τα $m_k \times m$ στοιχεία στη μήτρα S , υπάρχουν μόνο $m_k \times K$ μη μηδενικά στοιχεία. Δεδομένου ότι το K είναι συνήθως πολύ μικρότερο από το m , η S είναι μια αραιή μήτρα. Μπορούμε να χρησιμοποιήσουμε αραιή αποθήκευση και αριθμητική μήτρας (π.χ., Duff et al., 2002), για να εξοικονομήσουμε μνήμη και υπολογιστικό χρόνο. Στη συνέχεια, μπορούμε να χρησιμοποιήσουμε τον αλγόριθμο των Ng και Peyton (1993), όπως εφαρμόζεται από τους Koenker και Ng (2010) για να λύσουμε την εξίσωση (5.29). Αυτή η μέθοδος λειτουργεί καλά για μέτριες ιεραρχίες, αλλά για μερικές από τις μεγάλες ιεραρχίες που συναντάμε στην πράξη, ακόμη και αυτή η μέθοδος είναι ακατάλληλη.

Για πολύ μεγάλες ιεραρχίες, μπορεί να χρησιμοποιηθεί μια επαναληπτική προσέγγιση με βάση τη μέθοδο του Lanczos (1950, 1952). Αυτή η μέθοδος αναπτύχθηκε κυρίως για την επίλυση ενός συστήματος γραμμικών εξισώσεων, αλλά στη συνέχεια επεκτάθηκε από τους Golub και Kahan (1965) με στόχο να συνεργαστεί με προβλήματα ελαχίστων τετραγώνων με χαμηλότερο bidiagonalization της μήτρας συντελεστή. Με βάση τις μεθόδους αυτές, οι Paige και Saunders (1982) ανέπτυξαν έναν επαναληπτικό αλγόριθμο για την επίλυση αραιών γραμμικών προβλημάτων ελαχίστων τετραγώνων. Λόγω της επαναληπτικής διαδικασίας, τα αποτελέσματα είναι μια προσέγγιση για την άμεση λύση, αλλά σχεδόν σε όλες τις περιπτώσεις στις οποίες έχουμε χρησιμοποιήσει τη μέθοδο, η διαφορά είναι αμελητέα.

Η τρίτη προσέγγιση βασίζεται σε μια επανα-παραμετροποίηση του μοντέλου παλινδρόμησης. Αντί να χρησιμοποιείται παράμετρος που αντιπροσωπεύει κάθε σειρά του κάτω επίπεδου, χρησιμοποιείται μια παράμετρος για κάθε σειρά στην ιεραρχία, αλλά επιβάλλονται κάποιοι περιορισμοί μηδενικού αθροίσματος για την αποφυγή υπερ-παραμετροποίησης. Το νέο διάνυσμα παραμέτρων συμβολίζεται με $\phi_n(h)$ και έχει στοιχεία στην ίδια σειρά με το διάνυσμα Y_t . Κάθε παράμετρος μετρά την σχετική συμβολή ενός επιπέδου στις σειρές επίπεδου κάτω από αυτό. Στο παράδειγμα που παρουσιάζεται στο Σχ. 1, ισχύει :

$$\phi_n(h) = [\mu_{T,h}, \mu_{A,h}, \mu_{B,h}, \mu_{C,h}, \mu_{AA,h}, \mu_{AB,h}, \dots, \mu_{CC,h}, \mu_{AAA,h}, \mu_{AAB,h}, \dots, \mu_{CC,h}]' \quad (5.31)$$

$$\text{και} \quad \beta_{AAB,h} = \mu_{T,h} + \mu_{A,h} + \mu_{AA,h} + \mu_{AAB,h} \quad (5.32)$$

Ομοίως, οι άλλες τιμές β υπολογίζονται αθροίζοντας τους κόμβους συνάθροισης που συμβάλλουν σε αυτόν τον κόμβο του κατώτερου επίπεδου. Τότε μπορούμε να γράψουμε:

$$\beta_n(h) = S' \phi_n(h), \text{ έτσι ώστε:} \quad \hat{Y}_n(h) = SS' \phi_n(h) + \epsilon_h \quad (5.33)$$

Λόγω της υπερ-παραμετροποίησης, επιβάλλονται περιορισμοί που, για κάθε διάσπαση, το άθροισμα των εκτιμώμενων παραμέτρων είναι ίσο με 0, δηλαδή:

$$\sum_i \hat{\mu}_{i,h} = 0, \quad \sum_j \hat{\mu}_{ij,h} = 0, \quad \sum_k \hat{\mu}_{ijk,h} = 0, \quad \sum_l \hat{\mu}_{ijkl,h} = 0 \quad (5.34)$$

Αυτή η τυποποίηση μπορεί να θεωρηθεί ένα μοντέλο ANOVA με παραμέτρους που εκτιμώνται με τη μέθοδο ελαχίστων τετραγώνων. Το πλεονέκτημα αυτού του μοντέλου είναι ότι οι εκτιμητές μπορούν να επανεκφραστούν με το συνηθισμένο στυλ ANOVA, που περιλαμβάνει αθροίσεις αλλά όχι αναστροφές μητρώων. Η προσέγγιση λειτουργεί για κάθε ιεραρχία, αλλά, όταν η ιεραρχία δεν είναι ισορροπημένη, οι εξισώσεις είναι αρκετά περίπλοκες. Κατά συνέπεια, θα παρέχουμε μόνο τις εξισώσεις εκτίμησης για ισορροπημένες ιεραρχίες (δηλαδή, μέσα σε ένα δεδομένο επίπεδο, κάθε κόμβος έχει έναν ίσο αριθμό σειρών από κάτω).

Για να αναφερόμαστε σε αυτούς τους εκτιμητές, θα πρέπει να εισάγουμε κάποιο νέο συμβολισμό. Έστω X ένας κόμβος στην ιεραρχική δομή και m_x είναι ο αριθμός των σειρών στο υπο-δέντρο με επικεφαλής τον κόμβο X (συμπεριλαμβανομένου του X). Επίσης, έστω j_x είναι η σειρά του $Y_n(h)$ που περιέχει $Y_{x,h}$ και έστω ℓ_x είναι το επίπεδο στο οποίο εμφανίζεται ο κόμβος X , δηλαδή, $\ell_x \in \{0, 1, \dots, K\}$ και τέλος έστω $X^{(j)}$ είναι ο κόμβος, ο οποίος είναι j επίπεδα πάνω από το X , 1_k είναι ένα διάνυσμα σειρά μηδενικών με μοναδικό 1 στην k -οστή στήλη, και $b_x = \text{diag}[0, \dots, 0, 1, \dots, 1]$, όπου τα πρώτα J_{x-1} διαγώνια στοιχεία είναι 0 και κάθε στοιχείο μετά από αυτό είναι 1. Για παράδειγμα, στην ιεραρχία που φαίνεται στο Σχ. 5.1, $m_{BC} = 4$, $J_{BC} = 10$, $\ell_{BC} = 2$, και $BC^{(1)} = B$.

Έτσι, για μια ισορροπημένη ιεραρχία, οι ANOVA εκτιμητές για τις παραμέτρους είναι:

$$\hat{\mu}_{T,h} = \frac{1}{(1_1 SS')'(1_1 SS')} \{1_1 SS' \hat{Y}_n(h)\} \quad (5.35)$$

$$\hat{\mu}_{X,h} = \frac{1}{(1_{j_x} SS')'(1_{j_x} SS')} \{1_{j_x} SS' \hat{Y}_n(h) - \sum_{i=1}^{\ell_x} (1_{j_x} SS') (1_{j_x} SS' b_{X^{(i)}})' \hat{\mu}_{X^{(i)},h}\}, \text{ όπου } \ell_x \geq 1 \quad (5.36)$$

Στην πράξη, χρησιμοποιούμε κανονικά την προσέγγιση αραιής μήτρας για μικρές έως μέτριου μεγέθους ιεραρχίες, την προσέγγιση ANOVA για τις μεγάλες ισορροπημένες ιεραρχίες, και την επαναληπτική προσέγγιση για τις μεγάλες ανισόρροπες ιεραρχίες. Παρά το ότι είναι δυνατόν να λάβουμε εξισώσεις ANOVA για ασύμμετρες ιεραρχίες, είναι δύσκολο να πράξουμε το ίδιο εντελώς γενικά, ενώ είναι ευκολότερο να χρησιμοποιηθεί η επαναληπτική προσέγγιση των εξισώσεων ANOVA για μια συγκεκριμένη περίπτωση.

5.3.6.1 Συμπεράσματα

Προτείνεται μια νέα στατιστική μέθοδος για την ιεραρχική πρόβλεψη χρονολογικών σειρών, που επιτρέπει την βέλτιστοποίηση σημειακών παραγόμενων προβλέψεων που «συμφιλιώνονται» σε όλα τα επίπεδα της ιεραρχίας.

Ένα χρήσιμο χαρακτηριστικό της μεθόδου μας είναι ότι **οι ανεξάρτητες προβλέψεις βάσης μπορεί να προέρχονται από οποιοδήποτε μοντέλο, ή μπορεί να είναι επικριτικές προβλέψεις. Έτσι, στις προβλέψεις, μπορούν να ενσωματωθούν ad hoc προσαρμογές, να συμπεριληφθούν πληροφορίες σχετικά με τις επιμέρους σειρές καθώς και να**

συμπεριληφθούν σημαντικοί συμ-παράγοντες. Η μέθοδος δεν θέτει περιορισμούς σχετικά με το πώς δημιουργούνται οι αρχικές προβλέψεις βάσης. Η διαδικασία προβλέπει απλώς ένα μέσο για τον βέλτιστο συνδυασμό των προβλέψεων, έτσι ώστε να συγκεντρώνονται κατάλληλα σε όλη την ιεραρχία.

Ένα αξιοσημείωτο χαρακτηριστικό των αποτελεσμάτων της μεθόδου είναι ότι **οι σημειακές προβλέψεις είναι ανεξάρτητες από τους συσχετισμούς μεταξύ των σειρών**. Ενώ αυτό μπορεί αρχικά να φαίνεται αντι-δαισθητικό, είναι μια φυσική συνέπεια της μεθόδου, αν υποθεθεί ότι τα σφάλματα πρόβλεψης σε όλη την ιεραρχία αθροίζονται συνολικά με τον ίδιο τρόπο με τα παρατηρούμενα σύνολα δεδομένων, η οποία είναι μια λογική προσέγγιση στην πραγματικότητα. Φυσικά, **η διακυμάνση των προβλέψεων θα εξαρτηθεί σε μεγάλο βαθμό από τις συσχετίσεις μεταξύ των σειρών, και έτσι η μήτρα συνδιακύμανσης είναι ακόμα απαραίτητη προκειμένου να παραχθούν διαστήματα πρόβλεψης**.

Ένα άλλο εκπληκτικό αποτέλεσμα είναι ότι τα **βέλτιστα βάρη συνδυασμού εξαρτώνται μόνο από την ιεραρχική δομή και όχι από τα παρατηρούμενα δεδομένα**. Και πάλι, αυτό προκύπτει από την υπόθεση σχετικά με το συνυπολογισμό των σφαλμάτων των προβλέψεων. Εξαιτίας αυτού, είναι δυνατόν να προσδιοριστεί ο συνδυασμός βάρων μόνο μία φορά, και στη συνέχεια να εφαρμόζεται καθώς κάθε νέο σύνολο των παρατηρήσεων γίνεται διαθέσιμο. Αυτό εξοικονομεί πολύ υπολογιστικό χρόνο.

Οι προσομοιώσεις και εμπειρικές εφαρμογές αποδεικνύουν ότι η βέλτιστη μέθοδος συνδυασμού που προτείνεται ξεπερνά τις υπάρχουσες μεθόδους για την ιεραρχική πρόβλεψη δεδομένων. Γι 'αυτό και προτείνεται η έγκρισή της για τη συνήθη χρήση στις επιχειρήσεις και στη βιομηχανία, όπου χρειάζεται να γίνει ιεραρχική πρόβλεψη χρονοσειρών.

Κεφάλαιο 6 : Στατιστική ανάλυση χρονοσειρών και δείκτες σφαλμάτων

6.1 Χρονοσειρές (Timeseries)

Οι χρονοσειρές αποτελούν ένα σύνολο διαδοχικών παρατηρήσεων της τιμής κάποιου φυσικού ή άλλου μεγέθους. Ένα σημαντικό χαρακτηριστικό των χρονοσειρών είναι το γεγονός ότι οι διαδοχικές αυτές παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους. Όταν οι διαδοχικές παρατηρήσεις είναι εξαρτημένες, οι μελλοντικές τιμές μπορούν να προσδιοριστούν ακριβώς από τις προηγούμενες. Μια τέτοια διαδικασία ονομάζεται ντετερμινιστική. Αυτό δεν συμβαίνει με τις πραγματικές χρονοσειρές καθώς το μέλλον καθορίζεται μόνο μερικώς από το παρελθόν. Στην περίπτωση αυτή, θεωρείται ότι οι χρονοσειρές αντιπροσωπεύουν στοχαστικές διαδικασίες. Με βάση τα παραπάνω, καθίσταται σαφές ότι τα ντετερμινιστικά μοντέλα είναι εκείνα που περιγράφουν την διαχρονική εξέλιξη κάποιου φυσικού μεγέθους, για το οποίο υπάρχει πλήρης γνώση των παραγόντων που το επηρεάζουν. Ωστόσο, στην πραγματικότητα τα περισσότερα μεγέθη επηρεάζονται και από τον παράγοντα «τύχη», ο οποίος αντιπροσωπεύει μια στατιστική μεταβλητή. Τα μοντέλα που περιέχουν τον τυχαίο παράγοντα, ονομάζονται στοχαστικά.

6.1.1 Ποιοτικά χαρακτηριστικά των χρονοσειρών

Οι παραδοσιακές μέθοδοι ανάλυσης των χρονοσειρών ασχολούνται κυρίως με την αποσύνθεση της διακύμανσης της χρονοσειράς σε τέσσερα συστατικά της στοιχεία:

- Την τάση
- Την κυκλικότητα
- Την εποχιακότητα
- Τις μη κανονικές διακυμάνσεις

Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη τόσο για την εφαρμογή της κλασικής μεθόδου αποσύνθεσης, όσο και για την ανάλυση των ποιοτικών χαρακτηριστικών της χρονοσειράς.

Μελετώντας το γράφημα μιας χρονοσειράς στο πεδίο του χρόνου, προκύπτουν τα βασικά ποιοτικά χαρακτηριστικά της χρονοσειράς όπως η τάση, η εποχιακότητα, η κυκλικότητα και οι ασυνέχειες. Οι ασυνέχειες αποτελούν κατά μία έννοια υποσύνολο της έννοιας των μη κανονικών διακυμάνσεων.

Η *τάση* ορίζεται ως μια «μακροπρόθεσμη» μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς. Μιλώντας για τάση θα πρέπει κανείς να έχει στην διάθεση του έναν ικανό αριθμό παρατηρήσεων και να εκτιμήσει το κατάλληλο μήκος της περιόδου μέσα στην οποία θα αναζητήσει την ύπαρξη τάσης.

Η *κυκλικότητα* αντιπροσωπεύει μια «κυματοειδή» μεταβολή που οφείλεται σε ειδικές εξωγενείς συνθήκες και εμφανίζεται κατά περιόδους. Οι περίοδοι δεν είναι απαραίτητα σταθερές και το μήκος τους είναι κατά κανόνα μεγαλύτερο του έτους. Κυκλικότητα εμφανίζουν οι χρονοσειρές των περισσότερων οικονομικών μεγεθών όπως το Α.Ε.Π. . Οι κυκλικές αυτές μεταβολές είναι αποτέλεσμα των γενικότερων οικονομικών συνθηκών που χαρακτηρίζονται από ανόδους και υφέσεις και είναι γνωστές ως επιχειρηματικός κύκλος.

Η *εποχιακότητα* ορίζεται ως μια περιοδική διακύμανση που έχει σταθερό και μικρότερο του έτους μήκος. Η διακύμανση αυτή είναι συνήθως κατανοητή και προβλέψιμη. Πρόκειται συνήθως για μεταβολές που είναι άμεσα εξηγήσιμες και θεωρούνται κανονικές καθώς επαναλαμβάνονται κατά τον ίδιο ακριβώς τρόπο με τον χρόνο. Συνεπώς είναι εύκολο να μετρηθούν και να απομονωθούν έτσι ώστε να προκύψουν τελικά τα απεποχικοποιημένα δεδομένα.

Ασυνέχειες ονομάζονται οι απομονωμένες παρατηρήσεις που εμφανίζονται στο γράφημα κάποιας χρονοσειράς ως απότομες αλλαγές στο πρότυπο συμπεριφοράς της και δεν θα προορούσαν να έχουν προβλεφθεί από την ιστορία της. Αυτές οι αλλαγές δύνανται να έχουν παροδικό ή μόνιμο χαρακτήρα.

Στην περίπτωση που έχουν παροδικό χαρακτήρα ονομάζονται «*outliers*» ή «*special events*» και η επίδραση τους στην χρονοσειρά έχει μικρή χρονική διάρκεια. Για την ερμηνεία τέτοιων παρατηρήσεων απαιτείται θεωρητική γνώση, κριτική ικανότητα και κοινή λογική. Ένα outlier μπορεί να αντιπροσωπεύει μια ασυνήθιστη παρατήρηση που οφείλεται σε κάποιο εξαιρετικό και αδύνατο να προβλεφθεί γεγονός. Χαρακτηριστικό παράδειγμα αποτελεί η πτώση της παραγωγής λόγω κάποια απεργίας.

Στην περίπτωση που οι παρατηρούμενες ασυνέχειες έχουν μόνιμο χαρακτήρα, ονομάζονται «*level-shifts*» καθώς εμφανίζονται ως απότομες αλλαγές στο μέσο επίπεδο τιμών μιας χρονοσειράς. Χαρακτηριστικό παράδειγμα αποτελεί η μείωση των πωλήσεων μιας επιχείρησης εξαιτίας της εμφάνισης κάποιας ανταγωνίστριας επιχείρησης.

Μη κανονικές διακυμάνσεις θεωρούνται εκείνες που απομένουν όταν όλα τα υπόλοιπα συστατικά στοιχεία της χρονοσειράς (τάση, εποχιακότητα και κυκλικότητα) έχουν απομονωθεί. Οι διακυμάνσεις αυτές μπορεί να αντιπροσωπεύουν μια εντελώς τυχαία μεταβλητή που εκφράζει τον τυχαίο παράγοντα μιας στοχαστικής διαδικασίας. Εναλλακτικά μπορεί να αντιπροσωπεύουν κάποια ασυνέχεια (outlier) που σχετίζεται με κάποιο εξαιρετικό γεγονός.

6.2 Βασική στατιστική ανάλυση

| t | Δεδομένα |
|----------|-----------|
| 1 | Y_1 |
| 2 | Y_2 |
| 3 | Y_3 |
| \vdots | \vdots |
| n-2 | Y_{n-2} |
| n-1 | Y_{n-1} |
| n | Y_n |

Πίνακας 6.1 : Παρατηρήσεις χρονοσειράς προς στατιστική ανάλυση

6.2.1 Μέση τιμή (Average)

Αναφέρεται σε απλό γραμμικό μέσο όρο των τιμών των παρατηρήσεων και δείχνει το επίπεδο γύρω από το οποίο κινούνται οι πραγματικές τιμές της χρονοσειράς. Υπολογίζεται ως εξής:

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i \quad (6.1)$$

6.2.2 Μέγιστη και ελάχιστη τιμή (Maximum και Minimum)

Αναφέρονται στις ακραίες τιμές της χρονοσειράς και αποτελούν μία εκτίμηση της διακύμανσης των δεδομένων αλλά και της τυχαιότητας που αυτά περιέχουν.

6.2.3 Τυπική απόκλιση (Standard Deviation)

Αναφέρεται συχνά και ως διασπορά και εκφράζει το βαθμό κατά τον οποίο οι παρατηρήσεις είναι διεσπαρμένες γύρω από τη μέση τιμή. Υπολογίζεται σύμφωνα με τους παρακάτω τύπους:

$$\sigma_{\text{δείγματος}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad (6.2)$$

$$\sigma_{\text{πληθυσμού}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}} \quad (6.3)$$

6.2.4 Διακύμανση (Variance)

Ορίζεται ως το τετράγωνο της τυπικής απόκλισης

6.2.5 Συνδιακύμανση (Covariance)

Υποθέτοντας πως έχουμε δύο διακριτές τυχαίες μεταβλητές X και Y, ορίζεται ένα μέτρο σχέσης μεταξύ των δύο περιοχών δεδομένων, ώστε να καθοριστεί κατά πόσο οι δύο μεταβλητές μεταβάλλονται ανάλογα (θετική συνδιακύμανση) ή αντιστρόφως ανάλογα (αρνητική συνδιακύμανση) ή κατά πόσο οι τιμές των δύο συνόλων είναι ασυσχέτιστες μεταξύ τους (μηδενική συνδιακύμανση). Στις χρονοσειρές είναι συνήθης ο υπολογισμός της συνδιακύμανσης των δεδομένων σε σχέση με τον αύξοντα αριθμό της χρονικής περιόδου. Υπολογίζεται ως εξής:

$$\text{Cov}(X,Y) = \frac{1}{n} \cdot \sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})] \quad (6.4)$$

6.2.6 Συντελεστής γραμμικής συσχέτισης (Linear Correlation Coefficient)

Αναφέρεται και ως συντελεστής του Pearson και ακφράζει την συγκέντρωση των σημείων ενός διαγράμματος διασποράς γύρω από την ευθεία παλινδρόμησης, δίνοντας ένα μέτρο της γραμμικής συσχέτισης μεταξύ των δύο μεταβλητών. Ο συντελεστής παίρνει τιμές στο διάστημα $[-1,1]$ και ανάλογα με την τιμή του έχουμε :

- Αν $r = \pm 1$, υπάρχει τέλεια γραμμική συσχέτιση.
- Αν $-0,3 < r < 0,3$, δεν υπάρχει γραμμική συσχέτιση.

Αυτό, όμως, δεν σημαίνει ότι δεν υπάρχει άλλου είδους συσχέτιση μεταξύ των δύο μεταβλητών.

- Αν $-0,5 < r \leq -0,3$ ή $0,3 \leq r < 0,5$, υπάρχει ασθενής γραμμική συσχέτιση.
- Αν $-0,7 < r \leq -0,5$ ή $0,5 \leq r < 0,7$, υπάρχει μέση γραμμική συσχέτιση.
- Αν $-0,8 < r \leq -0,7$ ή $0,7 \leq r < 0,8$, υπάρχει ισχυρή γραμμική συσχέτιση.
- Αν $-1 < r \leq -0,8$ ή $0,8 \leq r < 1$, υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

Όπως και στην περίπτωση της συνδιακύμανσης, η συσχέτιση γίνεται συνήθως θεωρώντας ως μια μεταβλητή τον χρόνο. Υπολογίζεται ως εξής:

$$r_{XY} = \frac{\sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (6.5)$$

6.2.7 Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient)

Εκφράζει τη συσχέτιση μεταξύ παρατηρήσεων της ίδιας μεταβλητής με χρονική υστέρηση k περιόδους. Ο συντελεστής παίρνει τιμές στο διάστημα $[0,1]$. Εάν η τιμή του είναι κοντά στο 0, δηλώνει μηδενική συσχέτιση των παρατηρήσεων χρονικής υστέρησης k , ενώ αν είναι κοντά στο 1 δηλώνει μεγάλη συσχέτιση. Ο συγκεκριμένος συντελεστής είναι πολύ χρήσιμος σε περιπτώσεις που θέλουμε να αντοπίσουμε αλγεβρικά την εποχιακή συμπεριφορά μιας σειράς παρατηρήσεων. Υπολογίζεται ως εξής:

$$ACF_k = \frac{\sum_{i=1+k}^n [(Y_i - \bar{Y}) \cdot (Y_{i-k} - \bar{Y})]}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (6.6)$$

6.2.8 Συντελεστής μεταβλητότητας (Coefficient of Variation)

Είναι ένα κανονικοποιημένο μέτρο της διασποράς των παρατηρήσεων ενός δείγματος ή ενός πληθυσμού. Σε σύγκριση με την απλή τυπική απόκλιση, έχει το πλεονέκτημα ότι είναι απαλλαγμένο από την επίδραση του επιπέδου των παρατηρήσεων, αλλά δεν μπορεί να υπολογισθεί όταν η μέση τιμή ισούται με μηδέν. Υπολογίζεται ως εξής:

$$C_v = \frac{\sigma}{\bar{Y}} \cdot 100 (\%) \quad (6.7)$$

6.2.9 Μέση τιμή διαστήματος μεταξύ ζητήσεων (Intermittent Demand Interval)

Αναφέρεται σε χρονοσειρές διακοπτόμενης ζήτησης και εκφράζει τη μέση τιμή των αποστάσεων διαδοχικών περιόδων με μη μηδενική τιμή. Αν η τιμζοντας πρόβλεψή του δείκτη προκύψει ίση με τη μονάδα, τότε αναφερόμαστε σε χρονοσειράσυνεχούς ζήτησης, ενώ στην περίπτωση διακοπτόμενης ζήτησης παίρνει τιμές μεγαλύτερες της μονάδας. Μεγάλη τιμή του δείκτη συνεπάγεται μεγάλα, κατα μέσο όσο, μεσοδιαστήματα μεταξύ μη μηδενικών παρατηρήσεων.

6.3 Στατιστική ανάλυση ακρίβειας προβλέψεων

Αντίθετα με τη βασική στατιστική ανάλυση, όπου απαιτείται μονάχα η πραγματική σειρά των παρατηρήσεων, στην στατιστική ανάλυση ακρίβειας προβλέψεων ανήκουν στατιστικοί δείκτες, για τον υπολογισμό των οποίων απαιτείται μια σειρά πρόβλεψης, η οποία προκύπτει από την εφαρμογή κάποιας μεθόδου πρόβλεψης επί της πραγματικής χρονοσειράς. Η εφαρμογή μιας μεθόδου πρόβλεψης έχει ως αποτέλεσμα τον υπολογισμό των τιμών F_1 έως F_{n+h} από τις τιμές Y_1 έως Y_n .

Μ' αυτόν τον τρόπο, το διάλυμα F δηλώνει τις προβλέψεις πλήθους $n+h$, όπου n είναι το πλήθος των τιμών της πραγματικής χρονοσειράς και h ο ζητούμενος ορίζοντας πρόβλεψης. Το διάλυμα Y αποτελεί τα πραγματικά δεδομένα και, ανάλογα με την περίπτωση, μπορεί να αναφέρεται στα αρχικά μη επεξεργασμένα δεδομένα, στα επεξεργασμένα δεδομένα (αναφορικά με μηδενικές τιμές και ημερολογιακές προσαρμογές) ή ακόμα και στην αποεποχικοποιημένη χρονοσειρά. Το σύνολο των τιμών πρόβλεψης αποτελείται από δύο μέρη. Οι πρώτες n παρατηρήσεις αυτού (όπου υπάρχουν παράλληλα διαθέσιμα και τα αντίστοιχα πραγματικά δεδομένα) καλείται προσαρμογή του μοντέλου πρόβλεψης (forecast model fitting), ενώ οι υπόλοιπες αποτελούν τις μελλοντικές προβλέψεις. Ορισμές μέθοδοι πρόβλεψης ξεκινούν την παραγωγή του μοντέλου πρόβλεψης αφήνοντας κάποιες αρχικές περιόδους κενές, λόγω αρχικοποιήσεων.

Ορίζεται ως σφάλμα πρόβλεψης η διαφορά μεταξύ πραγματικής τιμής και πρόβλεψης για μια περίοδο: $e_i = Y_i - F_i$.

| t | Δεδομένα | Πρόβλεψη |
|-----|-----------|-----------|
| 1 | Y_1 | F_1 |
| 2 | Y_2 | F_2 |
| 3 | Y_3 | F_3 |
| ⋮ | ⋮ | ⋮ |
| n-2 | Y_{n-2} | F_{n-2} |
| n-1 | Y_{n-1} | F_{n-1} |
| n | Y_n | F_n |
| n+1 | | F_{n+1} |
| ⋮ | | ⋮ |
| n+h | | F_{n+h} |

Πίνακας 6.2 : Παρατηρήσεις και προβλέψεις χρονοσειράς προς στατιστική ανάλυση ακρίβειας προβλέψεων

Είναι προφανές πως, προτού γίνουν διαθέσιμες οι πραγματικές τιμές Y_{n+1} έως Y_{n+h} , η τιμή του σφάλματος μπορεί να υπολογιστεί μόνο για τις τιμές που έχουμε υπολογίσει το μοντέλο πρόβλεψης (in-sample error). Όταν, όμως, έχει παρέλθει κατάλληλο χρονικό διάστημα και τα δεδομένα έχουν γίνει διαθέσιμα, μπορούμε να υπολογίσουμε τόσο το σφάλμα του μοντέλου πρόβλεψης, όσο και το πραγματικό σφάλμα (out-of-sample error). Οι παρακάτω στατιστικοί δείκτες σφάλματος είναι εκφρασμένοι έτσι ώστε να αναφέρονται στο σφάλμα του μοντέλου πρόβλεψης, αλλά με κατάλληλες αλλαγές στις περιόδους που συμμετέχουν για τον υπολογισμό αυτών, μπορούν να αναφερθούν στην ακρίβεια των προβλέψεων πέραν της περιόδου n , δεδομένης της ύπαρξης νέων πραγματικών δεδομένων.

6.3.1 Μέσο σφάλμα (Mean Error)

Υπολογίζεται από τον απλο προσημασμένο μέσο όρο των σφαλμάτων και εκφράζει ένα μέτρο συστηματικότητας του σφάλματος. Όσο η τιμή αυτού είναι κοντά στο μηδέν, τόσο τα σφάλματα είναι τυχαία και όχι συστηματικά. Αν ο δείκτης παίρνει θετικές τιμές, δηλώνει απαισιοδοξία στις προβλέψεις, μιας και οι προβλέψεις ήταν κατά μέσο όρο μικρότερες των πραγματικών τιμών. Από την άλλη, αρνητικές τιμές του δείκτη δηλώνουν αισιοδοξία. Συχνά ο δείκτης αναφέρεται και ως bias και υπολογίζεται από τον παρακάτω τύπο :

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i) \quad (6.8)$$

6.3.2 Μέσο απόλυτο σφάλμα (Mean Absolute Error)

Εκφράζει ένα μέτρο της ακρίβειας της πρόβλεψης έναντι των πραγματικών τιμών διατηρώντας τις μονάδες μέτρησης της αρχικής χρονοσειράς. Δηλώνει ένα μέσο μέτρο της αστοχίας της πρόβλεψης, χωρίς να δίνεται έμφαση στην κατεύθυνση της πρόβλεψης. Όσο μεγαλύτερη είναι η τιμή του δείκτη, τόσο μικρότερη προκύπτει η ακρίβεια της μεθόδου που εφαρμόστηκε. Υπολογίζεται ως εξής:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i| \quad (6.9)$$

6.3.3 Μέσο τετραγωνικό σφάλμα (Mean Squared Error)

Όπως και το απόλυτο σφάλμα, είναι ένα μέτρο της ακρίβειας πρόβλεψης, το οποίο όμως δίνει πολύ μεγαλύτερο βάρος στα μεγάλα σφάλματα (δεδομένου ότι τα σφάλματα τετραγωνίζονται) και μικρότερο βάρος στα μικρά σφάλματα. Χρησιμοποιείται ευρέως για τον υπολογισμό των βέλτιστων παραμέτρων εξομάλυνσης. Υπολογίζεται ως εξής:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2 \quad (6.10)$$

6.3.4 Ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Squared Error)

Υπολογίζεται άμεσα από την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος. Έχει τις ίδιες ιδιότητες με αυτό, αλλά είναι εκφρασμένο στις μονάδες της αρχικής χρονοσειράς. Δίνεται ως:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2} \quad (6.11)$$

6.3.5 Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error)

Ορισμένες φορές είναι πιο χρήσιμος ο υπολογισμός των σφαλμάτων πρόβλεψης σε καθαρά ποσοστιαία μορφή. Αυτό, για παράδειγμα, θα ήταν χρήσιμο όταν θέλουμε να συγκρίνουμε την ακρίβεια μιας μεθόδου πρόβλεψης που έχει εφαρμοσθεί σε περισσότερες από μία χρονοσειρές, όπου η καθεμία έχει διαφορετικό επίπεδο μέσης τιμής. Επίσης, είναι πολύ χρήσιμο, όταν οι πραγματικές τιμές Y είναι ιδιαίτερα υψηλές. Το μέσο απόλυτο ποσοστιαίο σφάλμα είναι εκφρασμένο επί τοις εκατό και λαμβάνει τιμές μεγαλύτερες ή ίσες του μηδενούς, με τις μικρότερες τιμές να υποδηλώνουν και καλύτερη απόδοση της μεθόδου πρόβλεψης. Δίνεται από την ακόλουθη εξίσωση:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| \cdot 100 (\%) \quad (6.12)$$

Ο δείκτης αυτός δεν μπορεί να εφαρμοσθεί σε χρονοσειρές διακοπτόμενης ζήτησης (δηλαδή χρονοσειρές με μεγάλο πλήθος μηδενικών), καθώς καταλήγει σε απροσδιοριστία.

6.3.6 Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error)

Αποτελεί μια παραλλαγή του μέσου απόλυτου ποσοστιαίου σφάλματος και υπολογίζεται ως εξής:

$$\text{sMAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\left(\frac{Y_i + F_i}{2}\right)} \right| \cdot 100 (\%) = \frac{1}{n} \sum_{i=1}^n \left| \frac{2 \cdot (Y_i - F_i)}{(Y_i + F_i)} \right| \cdot 100 (\%) \quad (6.13)$$

Σε αντίθεση με το μέσο απόλυτο σφάλμα, το απόλυτο του σφάλματος δε διαιρείται απλώς με την πραγματική τιμή, αλλά με το ημιάθροισμα της πραγματικής τιμής και της πρόβλεψης. Με την αλλαγή αυτή, ο δείκτης αυτός αποκτά και πάνω όριο και πλέον μπορεί να πάρει τιμές στο διάστημα [0%,200%]. Παρ' όλα αυτά, υπάρχει ένα πρόβλημα με το συγκεκριμένο δείκτη, καθώς δεν είναι τόσο συμμετρικός όσο δηλώνει το όνομα του. Οι αισιόδοξες και οι απαισιόδοξες προβλέψεις δεν μεταχειρίζονται αντίστοιχα και αυτό γίνεται εμφανές από το παρακάτω παράδειγμα :

- Αισιόδοξη πρόβλεψη: $Y_t=100$ και $F_t=110$, τότε $\text{sMAPE}= 4,76\%$
- Απαισιόδοξη πρόβλεψη: $Y_t=100$ και $F_t=90$, τότε $\text{sMAPE}= 5,26\%$

6.3.7 Relative Measures

Σε πολλές περιπτώσεις εφαρμόζουμε περισσότερες από μία μεθόδους πρόβλεψης σε μία χρονοσειρά κι υπολογίζοντας για όλες τις μεθόδους κάποιον από τους παραπάνω δείκτες σφάλματος αποφασίζουμε για την καταλληλότητα της μεθόδου για τα συγκεκριμένα δεδομένα, βάσει της σωστής προσαρμογής του μοντέλου.

| Δεδομένα | Πρόβλεψη 1 (Benchmark) | Πρόβλεψη 2 | Πρόβλεψη 3 |
|-----------|---------------------------|-------------|-------------|
| Y_1 | F^1_1 | F^2_1 | F^3_1 |
| Y_2 | F^1_2 | F^2_2 | F^3_2 |
| Y_3 | F^1_3 | F^2_3 | F^3_3 |
| \vdots | \vdots | \vdots | \vdots |
| Y_{n-2} | F^1_{n-2} | F^2_{n-2} | F^3_{n-2} |
| Y_{n-1} | F^1_{n-1} | F^2_{n-1} | F^3_{n-1} |
| Y_n | F^1_n | F^2_n | F^3_n |
| | F^1_{n+1} | F^2_{n+1} | F^3_{n+1} |
| | \vdots | \vdots | \vdots |
| | F^1_{n+h} | F^2_{n+h} | F^3_{n+h} |

Πίνακας 6.3 : Παρατηρήσεις και διαφορετικές προβλέψεις χρονοσειράς προς στατιστική ανάλυση ακρίβειας προβλέψεων

Η σύγκριση μπορεί να γίνει βάσει κάποια μεθόδου που θα θεωρηθεί ως μέθοδος αναφοράς (benchmark). Με τον τρόπο αυτό ορίζεται το σχετικό σφάλμα ως εξής:

$$RelMAE = \frac{MAE_i}{MAE_{benchmark}} \quad (6.14)$$

Στη συνέχεια ακολουθεί η παρουσίαση δύο δεικτών, που βασίζονται στη σχετική σύγκριση μεθόδων που μόλις περιγράφηκε. Αξίζει, όμως να σημειωθεί πως τα σχετικά σφάλματα ορίζονται γενικά και επιτρέπουν την χρήση οποιασδήποτε μεθόδου ως μέθοδο αναφοράς, αλλά και την χρήση οποιουδήποτε δείκτη.

6.3.8 Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error)

Οι Hydmann και Koehler (2006), προκειμένου να αντιμετωπίσουν τις περιπτώσεις απροσδιοριστίας των δεικτών MAPE και sMAPE αλλά και να δώσουν την ίδια βαρύτητα στα μικρά και τα μεγάλα σφάλματα, σε αντίθεση με τους δείκτες MSE και RMSE, πρότειναν τον ακόλουθο στατιστικό δείκτη για τη μέτρηση της ακρίβειας πρόβλεψης:

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - F_i|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|} \quad (6.15)$$

Η παραπάνω εξίσωση μοιάζει αρκετά με το μέσο απόλυτο σφάλμα (αριθμητής), η οποία, όμως, είναι κανονικοποιημένη με τη μέση τιμή των διαφορών πρώτου βαθμού της χρονοσειράς. Όμως, ο παρονομαστής, ουσιαστικά αποτελεί το μέσο απόλυτο σφάλμα του μοντέλου πρόβλεψης, όταν ως μέθοδος πρόβλεψης έχει εφαρμοστεί η μέθοδος Naive

(αφελής ή απλοϊκή μέθοδος), η οποία δίνει ως πρόβλεψη για την επόμενη περίοδο την πραγματική τιμή της τρέχουσας περιόδου. Όταν ο δείκτης υπολογιστεί μικρότερος της μονάδας, τότε συμπεραίνουμε πως η μέθοδος πρόβλεψης που εφαρμόστηκε έχει κατά μέσο όρο καλύτερη απόδοση από την πιο απλή μέθοδο πρόβλεψης (Naive), ενώ το αντίθετο συμβαίνει αν προκύψει τιμή μεγαλύτερη της μονάδας. Οι Hydmann και Koehler προτείνουν πως οι δείκτες που βασίζονται σε κανονικοποιημένα σφάλματα πρέπει να καθιερωθούν ως μόνιμη προσέγγιση στη σύγκριση ακρίβειας προβλέψεων μεταξύ χρονοσειρών διαφορετικών επιπέδων. Και αυτό, διότι διαθέτουν μια κανονικοποίηση που είναι εύκολα κατανοητή και ευρέως εφαρμόσιμη. Η μόνη περίπτωση, κατά την οποία ο συγκεκριμένος δείκτης θα δώσει απροσδιοριστά σφάλματα είναι όταν όλες οι τιμές των ιστορικών δεδομένων είναι ίσες μεταξύ τους.

6.3.9 Theil's U-Statistic

Ο Theil (1966) πρότεινε ένα δείκτη σφάλματος, ο οποίος συνδυάζει χαρακτηριστικά σχετικών σφαλμάτων (όπου γίνεται άμεση σύγκριση με μια μέθοδο αναφοράς όπως στο MAE), δίνοντας παράλληλα μεγαλύτερο βάρος στα μεγαλύτερα σφάλματα (όπως το MSE), καθώς μεγάλα σφάλματα για μια επιχείρηση συνεπάγονται και ισοδύναμα μεγάλη ζημιά. Ο στατιστικός αυτός δείκτης επιτρέπει τη συγκριτική συσχέτιση της ακρίβειας των προβλέψεων με τις προβλέψεις της μεθόδου Naive και ταυτόχρονα τετραγωνίζει τα σφάλματα, ούτως ώστε να δίνεται μεγαλύτερο βάρος στις μεγάλες αποκλίσεις. Η μαθηματική διατύπωση του δείκτη έχει ως εξής:

$$U = \sqrt{\frac{\sum_{i=2}^n \left(\frac{|Y_i - F_i|}{|Y_{i-1}|}\right)^2}{\sum_{i=2}^n \left(\frac{|Y_i - Y_{i-1}|}{|Y_{i-1}|}\right)^2}} \quad (6.16)$$

- Αν $U=1$, τότε η μέθοδος Naive είναι εξίσου ακριβής με την μέθοδο πρόβλεψης που εφαρμόστηκε.
- Αν $U < 1$, τότε η μέθοδος πρόβλεψης που εφαρμόστηκε έχει καλύτερη απόδοση από πλευράς ακρίβειας σε σχέση με τη μέθοδο Naive (όσο μικρότερη τιμή, τόσο καλύτερη απόδοση).
- Αν $U > 1$, τότε η μέθοδος πρόβλεψης που εφαρμόστηκε έχει χειρότερη απόδοση από πλευράς ακρίβειας σε σχέση με τη μέθοδο Naive, οπότε δεν υπάρχει λόγος να εφαρμοσθεί (όσο μεγαλύτερη τιμή, τόσο χειρότερη απόδοση).

6.3.10 Percentage Better

Μια συσχετιστική προσέγγιση, που επιτρέπει την σύγκριση διαφορετικών μεθόδων που έχουν εφαρμοσθεί σε μεγάλο πλήθος χρονοσειρών διαφορετικών επιπέδων είναι η χρήση του ποσοστού, για το οποίο μια μέθοδος είναι ακριβέστερη έναντι μιας μεθόδου αναφοράς για ένα συγκεκριμένο δείκτη σφάλματος. Για παράδειγμα, κάνοντας χρήση του δείκτη MAPE, θα είχαμε:

$$\text{Percentage Better} = \text{mean}(\{I\{\text{MAPE}_i < \text{MAPE}_b\}\}) \cdot 100 (\%) \quad (6.17)$$

Το μειονέκτημα του παραπάνω δείκτη είναι ότι δε δίνει καμία αίσθηση του επιπέδου βελτίωσης. Για παράδειγμα, έστω ότι θέλαμε να αξιολογήσουμε τη μέθοδο 1 έναντι της μεθόδου 2 για την ακρίβεια τους στην πρόβλεψη 10 χρονοσειρών. Υπάρχει περίπτωση η μέθοδος 1 να είναι ελάχιστα καλύτερη από τη μέθοδο 2 για τις 9 από τις 10 χρονοσειρές και πολύ χειρότερη στην 10^η χρονοσειρά. Έτσι, ακόμα και αν η τιμή του συγκεκριμένου δείκτη υπολογιστεί 90%, αυτό δεν σημαίνει ότι η μέθοδος 2 δεν είναι η πλέον κατάλληλη.

6.3.11 Ρυθμός ανάπτυξης

Ο δείκτης του ρυθμού ανάπτυξης (growth rate) αποτελεί ένα μέτρο της αυξητικής ή φθίνουσας πορείας μιας σειράς δεδομένων για ένα συγκεκριμένο χρονικό διάστημα. Εκφράζεται σε ποσοστιαία μορφή και συνήθως αναφέρεται στη σύγκριση του ύψους των δεδομένων του τελευταίου έτους σε σχέση με τα υπόλοιπα διαθέσιμα δεδομένα. Η μαθηματική έκφραση του ρυθμού ανάπτυξης έχει ως εξής:

$$\text{Growth Rate} = \frac{\frac{1}{ppy} \sum_{i=n-ppy+1}^n Y_i - \frac{1}{n-ppy} \sum_{i=1}^{n-ppy} Y_i}{\frac{1}{n-ppy} \sum_{i=1}^{n-ppy} Y_i} \cdot 100 (\%) \quad (6.18)$$

,όπου Y το διάνυσμα των n παρατηρήσεων και ppy το πλήθος των περιόδων στο μήκος ενός έτους (για παράδειγμα, ppy=12 αν τα δεδομένα αφορούν μηνιαίες παρατηρήσεις).

Κεφάλαιο 7 : Ενεργειακές προβλέψεις μέσω μοντέλου πρόβλεψης χρονοσειρών με χρονική συνάθροιση

7.1 Μεθοδολογία πρόβλεψης χρονοσειρών με χρονική συνάθροιση

7.1.1 Ιδέα της μεθόδου

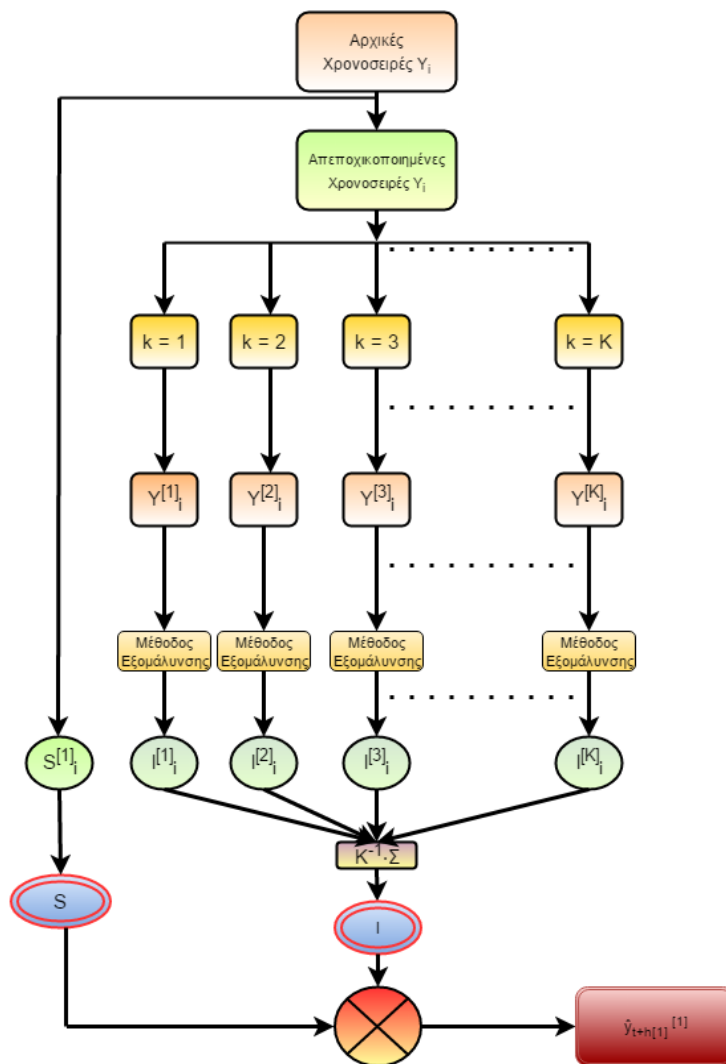
Η πολλαπλή χρονική συνάθροιση αναφέρεται στην παραγωγή πολλαπλών χρονοσειρών εναλλακτικών συχνοτήτων από μια χρονοσειρά, οι τιμές της οποίας παρατηρούνται σε μια συγκεκριμένη (υψηλότερη) συχνότητα. Μέσω των διαφορετικών συχνοτήτων των δεδομένων, αποκαλύπτονται ή κρύβονται διαφορετικά χαρακτηριστικά των δεδομένων. Για παράδειγμα, στα υψηλά επίπεδα συνάθροισης (χαμηλή συχνότητα) αναμένεται να αναδεικνύονται τα μακροχρόνια χαρακτηριστικά της χρονοσειράς, όπως η τάση και το επίπεδο, ενώ στα χαμηλά (υψηλή συχνότητα) τα βραχυχρόνια χαρακτηριστικά της, όπως η εποχιακότητα. Η παραγωγή των πολλαπλών χρονοσειρών εναλλακτικών συχνοτήτων επιτυγχάνεται με κατάλληλη επιλογή του επιπέδου συνάθροισης, ενώ η πρόβλεψη παράγεται στο πλέον κατάλληλο επίπεδο (temporal aggregation), ή σε συνδυασμό επιπέδων (multiple temporal aggregation).

7.1.2 Εφαρμογή μεθόδου πρόβλεψης χρονοσειρών με χρονική συνάθροιση

Η ακριβής μέθοδος που ακολουθήσαμε απεικονίζεται στο διάγραμμα ροής 7.1.πο ακολουθεί. Το πρώτο βήμα του πλαισίου που ακολουθήσαμε αναφέρεται στην απεποχικοποίηση των χρονοσειρών όλων των στοιχείων της ιεραρχίας και τον υπολογισμό της εποχιακής συνιστώσας της κάθε χρονοσειράς μέσω των κατάλληλων δεικτών εποχιακότητας.

Στο δεύτερο βήμα της μεθόδου γίνεται η συνάθροιση των απεποχικοποιημένων χρονοσειρών και εν συνεχεία εφαρμόζεται η κατάλληλη μέθοδος εξομάλυνσης σε κάθε χρονοσειρά. Λόγω της φύσης των δεδομένων αυτή είναι η μέθοδος εκθετικής εξομάλυνσης σταθερού επιπέδου (SES).

Μέσω της SES παράγονται προβλέψεις συνιστώσας επιπέδου για κάθε επίπεδο συνάθροισης, και στη συνέχεια υπολογίζεται ο μέσος όρος αυτών, που αποτελεί τις προβλέψεις συνιστώσας επιπέδου, στις οποίες πολλαπλασιάζονται οι δείκτες εποχιακότητας για να δώσουν τις τελικές προβλέψεις.



Σχήμα 7.1 : Διάγραμμα ροής μεθόδου πρόβλεψης χρονοσειρών με χρονική συνάθροιση

Παράγονται προβλέψεις σύμφωνα με τρεις διαφορετικές ιεραρχικές προσεγγίσεις. Επιπρόσθετα σημειώνουμε ότι εφαρμόζουμε τη μέθοδο rolling horizon για την παραγωγή προβλέψεων μεγαλύτερης αξιοπιστίας.

Ελέγχεται η ακρίβεια των προβλέψεων με βάση τον στατιστικό δείκτη s-MAPE και η απόδοση της μεθόδου συγκριτικά με τη μέθοδο απλής εκθετικής εξομάλυνσης SES, που λαμβάνεται ως benchmark στην εργασία αυτή.

Ακολουθεί αναλυτική περιγραφή των βημάτων.

7.1.2.1 Απαλοιφή εποχιακότητας

Γενικά η απαλοιφή της τάσης ή της εποχιακότητας (γενικά περιοδικότητας) γίνεται όταν δε μας ενδιαφέρει να μελετήσουμε τις μεταβολές στη χρονοσειρά που οφείλονται σε τάσεις ή περιοδικότητα γιατί θεωρούμε ότι δημιουργούνται από άλλους παράγοντες που δε σχετίζονται με το σύστημα που θέλουμε να διερευνήσουμε ή περιγράψουμε. Για προβλέψεις, είτε συμπεριλαμβάνουμε την τάση και περιοδικότητα στο μοντέλο πρόβλεψης, είτε εκτιμούμε το μοντέλο στη χρονοσειρά που προκύπτει αφαιρώντας την τάση ή περιοδικότητα και στις προβλέψεις του μοντέλου αυτού προσθέτουμε την τάση και περιοδικότητα για να πάρουμε την πρόβλεψη του παρατηρούμενου μεγέθους.

Αντίθετα, σε κάποιες εφαρμογές μπορεί η πληροφορία που θέλουμε να αντλήσουμε από τη χρονοσειρά να είναι ακριβώς η τάση ή η περιοδικότητα (ή και τα δύο). Τότε η ανάλυση της χρονοσειράς περιορίζεται στην εκτίμηση της τάσης ή της περιοδικότητας και θεωρεί την υπόλοιπη πληροφορία στις παρατηρήσεις χωρίς καμιά σημασία ή τυχαία.

Συνοψίζοντας, μια χρονοσειρά Y_t μπορεί σε κάθε χρονική στιγμή t να αναλυθεί στις συνιστώσες τάσης και περιοδικότητας, δηλαδή για τη μεταβλητή θεωρούμε το μοντέλο $Y_t = f(T_t, S_t, I_t)$, όπου :

- T_t είναι η τάση ως συνάρτηση του χρόνου, δηλαδή το αργά μεταβαλλόμενο μέσο επίπεδο τιμών της
- S_t είναι η συνιστώσα εποχιακότητας ή περιοδικότητας για κάποια περίοδο d
- I_t είναι το υπόλοιπο, που συγκεντρώνει την πληροφορία (αν υπάρχει) για τη δυναμική του συστήματος απαλλαγμένο από τάση και περιοδικότητα (επίπεδο).

Λόγω έλλειψης τάσης των δεδομένων, δεν εξετάζουμε την συνιστώσα της τάσης. Συνεπώς η διαδικασία απεποχικοποίησης των δεδομένων έγκειται στην απαλοιφή της εποχιακότητας από τα αρχικά δεδομένα. Επιλέγοντας το πολλαπλασιαστικό μοντέλο, αυτό επιτυγχάνεται μέσω της διαίρεσης των αρχικών παρατηρήσεων με τους δείκτες εποχιακότητας που υπολογίζονται για κάθε περιοχή ξεχωριστά με βάση τα ιστορικά δεδομένα.

Στόχος μας είναι να απαλοίσουμε την εποχιακότητα από τα αρχικά δεδομένα, να επεξεργαστούμε τα απεποχικοποιημένα δεδομένα με κάποια μέθοδο εξομάλυνσης, να προεκτείνουμε στο μέλλον το πρότυπο συμπεριφοράς των απεποχικοποιημένων δεδομένων και κατόπιν να συνδυάσουμε τις προβλέψεις αυτές με τις προβλέψεις της συνιστώσας της εποχιακότητας.

7.1.2.2 Χρονική συνάθροιση απεποχικοποιημένων δεδομένων

Οι απεποχικοποιημένες χρονοσειρές συναθροίζονται για $k = 1, \dots, 168$. Ελέγχονται 168 επίπεδα συνάθροισης με βάση τον αλγόριθμο MAPA, που περιγράφηκε στην ενότητα 5.2.3. Δημιουργούνται, με τον τρόπο αυτό, 168 συναθροισμένες απεποχικοποιημένες χρονοσειρές στις οποίες θα εφαρμόσουμε εκθετική εξομάλυνση σταθερού επιπέδου.

7.1.2.3 Εκθετική εξομάλυνση απεποχικοποιημένων χρονοσειρών και παραγωγή προβλέψεων συνιστώσας επιπέδου

Σε κάθε συναθροισμένη χρονοσειρά εφαρμόζουμε την κατάλληλη μέθοδο εκθετικής εξομάλυνσης. Λόγω έλλειψης τάσης αλλά και εποχιακότητας από τα δεδομένα, επιλέγεται η μέθοδος σταθερού επιπέδου από την οικογένεια των μεθόδων εξομάλυνσης, ως μέθοδος εξομάλυνσης των χρονοσειρών. Παράγονται προβλέψεις συνιστώσας επιπέδου για τις 168 χρονοσειρές των διαφορετικών επιπέδων συνάθροισης, όπως περιγράφηκε στην ενότητα 4.2.3, και στη συνέχεια υπολογίζεται ο μέσος όρος αυτών.

7.1.2.4 Παραγωγή προβλέψεων εποχιακής συνιστώσας

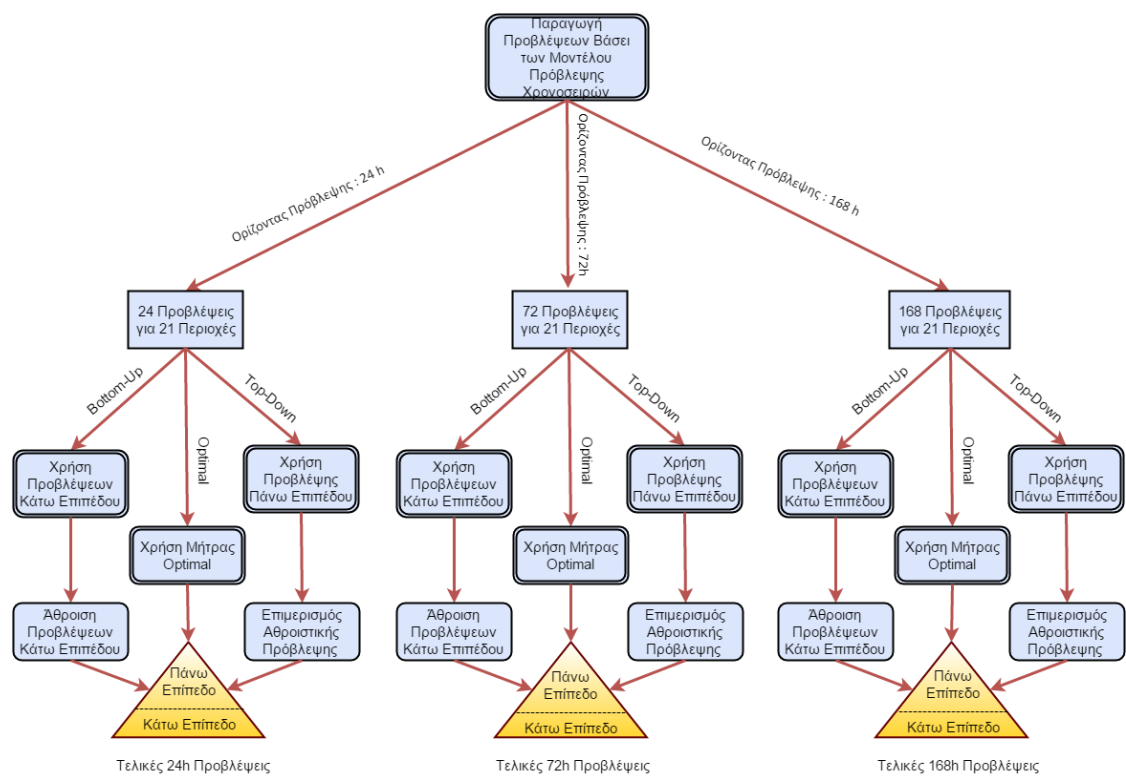
Σχετικά με την συνιστώσα της εποχιακότητας, επιλέγουμε να χρησιμοποιούμε εξωτερική εποχιακή συνιστώσα και όχι την εποχιακότητα που προτείνεται στο πλαίσιο του αλγόριθμου MAPA στην ενότητα 5.2.3. Ο λόγος που πράττουμε κατ' αυτόν τον τρόπο έγκειται στο γεγονός ότι με την εξωτερική εποχιακότητα που επιλέγουμε, αναδεικνύουμε τα χαρακτηριστικά της εποχιακότητας των δεδομένων της αρχικής συχνότητας του δείγματος. Εάν επιλέγαμε το μέσο όρο των συναθροισμένων συνιστωσών των εποχιακοτήτων προκειμένου να εκτιμήσουμε την εποχιακότητα στο μέλλον έστω με ορίζοντα πρόβλεψης 24 ωρών, θα λαμβάναμε προβλέψεις εποχιακότητας που θα είχαν συνεκτιμήσει την ιστορική εποχιακότητα ανώτερων επίπεδων συνάθροισης (π.χ. τριμηνιαία εποχιακότητα). Αυτό θα μας οδηγούσε σε παραγωγή προβλέψεων μικρότερης ακρίβειας. Συνεπώς επιλέγουμε αναδρομική εξωτερική εποχιακότητα, η οποία ουσιαστικά υπολογίζει δείκτες εποχιακότητας από το πιο πρόσφατο δείγμα παρατηρήσεων μήκους ίσου με τον ορίζοντα πρόβλεψης. Για να προσαρμόσουμε την εποχιακότητα στις απεποχικοποιημένες προβλέψεις, απλά πολλαπλασιάζουμε τις προβλέψεις αυτές με τους κατάλληλους δείκτες εποχιακότητας του πιο πρόσφατου δείγματος. Με τον όρο καταλληλότητα εννοούμε ότι αν για παράδειγμα θέλουμε να προβλέψουμε για την ημέρα Τετάρτη 24 ώρες μπροστά με πρώτη παρατήρηση την 01.00, η 5^η πρόβλεψη, για παράδειγμα θα αναφέρεται στις 5.00 της Τετάρτης και θα πρέπει να ενσωματώνει τον εποχιακό δείκτη του πιο πρόσφατου αντίστοιχου δείγματος, δηλαδή τον δείκτη που υπολογίστηκε με βάση την παρατήρηση της πιο πρόσφατης Τετάρτης (προηγούμενη εβδομάδα) την ίδια ώρα.

7.1.2.5 Συνδυασμός προβλέψεων συνιστωσών

Ο συνδυασμός των προβλέψεων συνιστωσών επιπέδου και εποχιακότητας γίνεται με βάση το πολλαπλασιαστικό μοντέλο $Y [t] = I[t] * S[t] * e [t]$. Λόγω μηδενικής τάσης, ο συνδυασμός των προβλέψεων των συνιστωσών έγκειται στον πολλαπλασιασμό του επιπέδου με την εξωτερική εποχιακότητα που έχουμε υπολογίσει με βάση την διαδικασία που περιγράφηκε.

7.1.3 Παραγωγή Ιεραρχικών Προβλέψεων

Στη συνέχεια, γίνεται η παραγωγή προβλέψεων για όλες τις συνολικές και επιμέρους καταναλώσεις, για τον χρονικό ορίζοντα που μας ενδιαφέρει και μέσω των τριών ιεραρχικών μεθόδων, που παρουσιάστηκαν στην ενότητα 4.3, σχηματίζονται οι τελικές προβλέψεις, σε όλα τα επίπεδα της ιεραρχίας. Εν συνεχεία, οι προβλέψεις αυτές αξιοποιούνται ως είσοδος για τις τρεις διαφορετικές ιεραρχικές προσεγγίσεις, μέσω των οποίων προκύπτουν οι τελικές προβλέψεις ανά επίπεδο ιεραρχίας. Για την Bottom-Up προσέγγιση, απαιτούνται μόνο οι προβλέψεις του κατώτατου επιπέδου, ενώ για την Top-Down απαιτείται μόνο εκείνη της συνολικής αθροιστικής κατανάλωσης. Για την Optimal προσέγγιση απαιτούνται όλες οι προβλέψεις.



Σχήμα 7.2 : Διάγραμμα Παραγωγής Προβλέψεων για τους τρεις χρονικούς ορίζοντες και για τις 3 ιεραρχικές προσεγγίσεις

7.1.4 Υλοποίηση στην R

Στην παρούσα εργασία χρησιμοποιήθηκε η γλώσσα προγραμματισμού R. Η R παρέχεται με βιβλιοθήκες και πακέτα που προσφέρουν μια μεγάλη ποικιλία στατιστικών και γραφικών μεθόδων, στις οποίες περιλαμβάνονται γραμμικά και μη γραμμικά μοντέλα, κλασικές στατιστικές δοκιμές, ανάλυση χρονοσειρών, ταξινόμηση, ομαδοποίηση, και άλλα. Η R είναι εύκολα επεκτάσιμη μέσω βιβλιοθηκών και πακέτων που μπορεί κανείς να «κατεβάσει» και να εγκαταστήσει. Η γλώσσα προγραμματισμού R καθώς και οι επεκτάσεις που χρησιμοποιήθηκαν είναι διαθέσιμες στην ιστοσελίδα :

<https://cran.r-project.org/bin/windows/base/>

Παρακάτω παρουσιάζονται κάποιες εντολές, συναρτήσεις και πακέτα συναρτήσεων που χρησιμοποιήθηκαν στην παρούσα μεθοδολογία την υλοποίηση συγκεκριμένων διαδικασιών.

- Το πακέτο συναρτήσεων `forecast` είναι ένα γενικό πακέτο συναρτήσεων για την πρόβλεψη από χρονοσειρές ή από μοντέλα πρόβλεψης χρονοσειρών. Η συνάρτηση `forecast` επικαλείται συγκεκριμένες μεθόδους που εξαρτώνται από την κατηγορία του πρώτου ορίσματος. Για παράδειγμα, η συνάρτηση `forecast.ts` παράγει εκτιμήσεις, χρησιμοποιώντας ETS μοντέλα (αν τα δεδομένα δεν είναι εποχιακά ή η εποχιακή περίοδος είναι 12 ή λιγότερο) . Συντάσσεται με βάση τα εξής ορίσματα : `forecast(object, h, subset, robust, weights, find.frequency, allow.multiplicative.trend, ...)` , όπου :
 - `object` : μια χρονοσειρά ή ένα μοντέλο χρονοσειρών για την οποία, ή για το οποίο θέλουμε να παράγουμε προβλέψεις
 - `h` : ορίζοντας πρόβλεψη
 - `robust`: Αν αυτή η λογική μεταβλητή είναι αληθής, τότε η συνάρτηση είναι ισχυρή σε τιμές που λείπουν και ακραίες τιμές στην χρονοσειρά.
 - `find.frequency` : Αν αυτή η λογική μεταβλητή είναι αληθής, τότε η συνάρτηση ορίζει το κατάλληλο χρονικό διάστημα ως περίοδο, εάν τα δεδομένα είναι παρατηρήσεις άγνωστης περιόδου.
 - `allow.multiplicative.trend` : Αν αυτή η λογική μεταβλητή είναι αληθής, τότε επιτρέπονται μοντέλα ETS με πολλαπλασιαστική τάση. Αλλιώς, επιτρέπονται μόνο μοντέλα ETS με αθροιστική τάση ή χωρίς καθόλου τάση.
- Η συνάρτηση `decompose()` αποσυνθέτει μια χρονοσειρά σε συνιστώσες εποχιακότητας, τάσης και ακανόνιστων συστατικών που χρησιμοποιούν κινητούς μέσους όρους. Εφαρμόζεται σε αθροιστική ή πολλαπλασιαστική εποχιακή συνιστώσα. Πρόκειται για μια συνάρτηση του βασικού πακέτου "stats" της R. Συντάσσεται με βάση τα εξής ορίσματα : `decompose(x, type = c("additive", "multiplicative"), filter = NULL)` , όπου :
 - `x` : μια χρονοσειρά
 - `type` : ο τύπος της εποχιακής συνιστώσας
 - `filter` : ένα διάνυσμα συντελεστών με αντίστροφη χρονολογική σειρά

Το αθροιστικό μοντέλο προϋποθέτει $Y[t] = I[t] + S[t] + e[t]$, ενώ το πολλαπλασιαστικό : $Y[t]=I[t]*S[t]*e[t]$.

Η συνάρτηση καθορίζει πρώτα τη συνιστώσα της τάσης χρησιμοποιώντας έναν κινούμενο μέσο όρο (αν το φίλτρο είναι NULL, χρησιμοποιείται ένα συμμετρικό παράθυρο με ίσα βάρη), και την αφαιρεί από τη χρονοσειρά. Στη συνέχεια, η εποχιακή περίοδος υπολογίζεται από το μέσο όρο για όλες τις περιόδους. Τέλος, η συνιστώσα σφάλματος καθορίζεται αφαιρώντας την τάση και την εποχιακή συνιστώσα από την αρχική χρονοσειρά. Η διαδικασία αυτή λειτουργεί καλά, μόνο εάν το x καλύπτει έναν ακέραιο αριθμό πλήρων χρονικών περιόδων.

- Το πακέτο συναρτήσεων MAPA αποτελεί, ουσιαστικά, έναν αλγόριθμο για την εκτίμηση και την παραγωγή προβλέψεων MAPA εντός και εκτός δείγματος. Συντάσσεται με τα εξής ορίσματα :

`mapa (y, rpy, fh = rpy, ifh = 1, minimumAL = 1, maximumAL = rpy, comb = "mean", paral = 0, display = 0, outplot = 1, hybrid = TRUE, model = "ZZZ", conf.lvl = NULL)` , όπου:

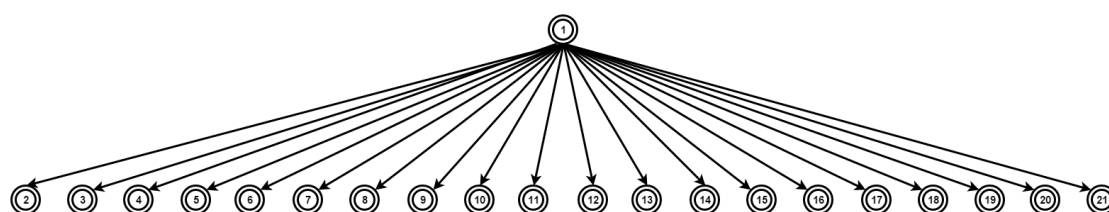
- y : in-sample δείγμα παρατηρήσεων της χρονοσειράς.
- rpy : Περίοδοι σε μια εποχή των χρονοσειρών στη συχνότητα της δειγματοληψίας. Αν το δείγμα in-sample είναι μια χρονοσειρά, τότε η τιμή αυτής της μεταβλητής λαμβάνεται από τη συχνότητα του.
- fh : ορίζοντας πρόβλεψης (απο προεπιλογή η τιμή του ορίζοντα πρόβλεψης είναι ίση με την τιμή rpy)
- Lower aggregation level (Default = 1), Lowest aggregation level (Default = 1) : επιλεγμένα ελάχιστα επίπεδα συνάθροισης
- `comb = "mean"` : τελεστής συνδυασμού (μπορεί να είναι μέσος όρος ή ενδιάμεση τιμή)
- `model` : Επιτρέπονται μόνο συγκεκριμένα είδη του ETS σε κάθε επίπεδο συγκέντρωσης. Αυτό ακολουθεί παρόμοια κωδικοποίηση με τη λειτουργία του ETS. Το πρώτο γράμμα αναφέρεται στο είδος σφάλματος ("A", "M" ή "Z"). Το δεύτερο γράμμα αναφέρεται στον τύπο τάσης ("N", "A", "Ad", "M", "MD" ή "Z") και το τρίτο γράμμα αναφέρεται στον τύπο εποχιακότητα ("N", "A", "M" ή "Z"). Τα γράμματα σημαίνουν: "N" = κανένα, "A" = πρόσθετική, "M" = πολλαπλασιαστική και "Z" = επιλέγεται αυτόματα. Το γράμμα "d" για την τάση συνεπάγεται απόσβεση. Εξ ορισμού επιλέγεται το μοντέλο = "ZZZ". Αν λόγω του περιορισμού του δείγματος ETS δεν μπορεί να υπολογιστεί σε επίπεδο συγκέντρωσης για το επιλεγμένο μοντέλο, τότε δεν γίνεται εκτίμηση για το συγκεκριμένο επίπεδο.
- `conf.lvl` : διάνυσμα του επιπέδου εμπιστοσύνης για τα διαστήματα πρόβλεψης. Λαμβάνει τιμές στο (0,1). Αν `conf.lvl == NULL` τότε δεν υπολογίζονται διαστήματα εμπιστοσύνης. Ο υπολογισμός των διαστημάτων πρόβλεψης βασίζεται στην εμπειρικό υπολογισμό MSE πολλαπλών βημάτων μπροστά.

(Πηγή : Kourentzes N., Petropoulos F., Trapero J.R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies, *International Journal of Forecasting*, 30(2), 291-302)

7.2 Case Study : Παραγωγή Προβλέψεων μέσω μοντέλου πρόβλεψης χρονοσειρών με χρονική συνάθροιση

7.2.1 Επισκόπηση Προβλήματος και Συλλογή Δεδομένων

Η μέθοδος πρόβλεψης χρονοσειρών με χρονική συνάθροιση που περιγράφηκε στην παραπάνω ενότητα εφαρμόζεται σε ένα case study για 20 διαφορετικές περιοχές της Αμερικής. Στόχος ήταν να δοκιμαστεί η αποτελεσματικότητα της μεθόδου σε μια ιεραρχική δομή που απαρτίζεται από 21 χρονοσειρές, οι οποίες αναλύονται σε 20 επιμέρους χρονοσειρές για κάθε περιοχή ξεχωριστά και 1 χρονοσειρά για όλες τις περιοχές συνολικά. Ακολουθεί το ιεραρχικό διάγραμμα του προβλήματος:



Σχήμα 7.3 : Ιεραρχία εξεταζόμενου προβλήματος

Σύμφωνα με το παραπάνω σχήμα, η ιεραρχική δομή αποτελείται από 2 επίπεδα. Οι 20 χρονοσειρές, που προκύπτουν από τις 20 επιμέρους περιοχές, αποτελούν το κάτω επίπεδο της ιεραρχίας, ενώ η 1^η χρονοσειρά, που αποτελεί και το πάνω επίπεδο της ιεραρχίας, είναι το άθροισμα των 20 επιμέρους χρονοσειρών του κάτω επιπέδου. Τα δεδομένα που χρησιμοποιήθηκαν για την εφαρμογή της μεθόδου προέρχονται από τον διαγωνισμό GEFC 2012 (Global Energy Forecasting Competition). Προκειται για ωριαίες παρατηρήσεις ενεργειακής ζήτησης (σε kWh) που εκτείνονται από τις 1/1/2004 και ώρα 01:00 έως τις 30/6/2008 και ώρα 00:00. Οι παρατηρήσεις αυτές συνοδεύονται από αντίστοιχες παρατηρήσεις ώρας, ημέρας, μήνα και έτους. Από τις συνολικά 39.408 παρατηρήσεις που έχουμε στην διάθεση μας για κάθε στοιχείο της ιεραρχίας, ανάλογα με τον ορίζοντα πρόβλεψης, θεωρούμε ως άγνωστο το δείγμα πρόσφατων παρατηρήσεων, μήκους διπλάσιου του ορίζοντα πρόβλεψης. Εξασκούμε το μοντέλο πρόβλεψης χρονοσειρών με βάση το μειωμένο δείγμα παρατηρήσεων και παράγουμε προβλέψεις για το δείγμα παρατηρήσεων που θεωρήσαμε άγνωστο.

7.2.2 Εφαρμογή μεθόδου μοντέλου πρόβλεψης χρονοσειρών με χρονική συνάθροιση

Το πρώτο στάδιο της προτεινόμενης μεθοδολογίας είναι το στάδιο της συνάθροισης. Με βάση τις αρχικές ωριαίες χρονοσειρές, δημιουργούνται οι συναθροισμένες χρονοσειρές χαμηλότερης συχνότητας. Σε κάθε μία από αυτές εφαρμόζουμε μέθοδο εξομάλυνσης σταθερού επιπέδου. Εξάγουμε, με τον τρόπο αυτό τις συνιστώσες της κάθε εξομαλυμένης χρονοσειράς. Λαμβάνουμε υπ' όψιν μόνο τις συνιστώσες των επιπέδων καθώς οι χρονοσειρές μας δεν έχουν τάση και περιοδικότητα.

Η μέθοδος αποεποχικοποίησης που ακολουθήθηκε στην παρούσα διαδικασία ήταν αυτή της απλής πολλαπλασιαστικής μεθόδου, μέσω της παρακάτω συνάρτησης :

`Decompose (time_series, type = c ("multiplicative"), filter = NULL).`

Η παράμετρος `type` ορίζει το μοντέλο της αποεποχικοποίησης, στη συγκεκριμένη περίπτωση επιλέγεται το πολλαπλασιαστικό μέσω του ορίσματος "multiplicative". Η παράμετρος `filter` με όρισμα `NULL` κάνει χρήση κινητών και κεντρικών κινητών μέσων όρων, εύρους ανάλογου της εποχιακότητας που έχουμε ορίσει για τη χρονοσειρά.

Στη συνέχεια, μέσω του συνδυασμού των προβλέψεων των συνιστωσών (με βάση το πρότυπο της πολλαπλασιαστικής εποχιακότητας που περιγράφηκε στην προηγούμενη ενότητα), γίνεται η παραγωγή προβλέψεων για όλες τις συνολικές και επιμέρους καταναλώσεις, για τον χρονικό ορίζοντα που μας ενδιαφέρει και μέσω των τριών ιεραρχικών μεθόδων, που παρουσιάστηκαν στην ενότητα 4.3, σχηματίζονται οι τελικές προβλέψεις, σε όλα τα επίπεδα της ιεραρχίας. Εν συνεχεία, οι προβλέψεις αυτές αξιοποιούνται ως είσοδος για τις τρεις διαφορετικές ιεραρχικές προσεγγίσεις, μέσω των οποίων προκύπτουν οι τελικές προβλέψεις ανά επίπεδο ιεραρχίας. Για την Bottom-Up προσέγγιση, απαιτούνται μόνο οι προβλέψεις του κατώτατου επιπέδου, ενώ για την Top-Down απαιτείται μόνο εκείνη της συνολικής αθροιστικής κατανάλωσης. Για την Optimal προσέγγιση απαιτούνται όλες οι προβλέψεις.

7.2.3 Παραγωγή Ιεραρχικών Προβλέψεων

Έχοντας στην διάθεση μας πλήρεις προβλέψεις για όλες τις χρονοσειρές, όλων των επιπέδων προχωρήσαμε στην παραγωγή των τελικών προβλέψεων ανά επίπεδο, μέσω των τριών διαφορετικών ιεραρχικών προσεγγίσεων.

Στην περίπτωση του συγκεκριμένου προβλήματος, έχουμε : $m = m_0 + m_1 = 20+1=21$, όπου m_0 και m_1 είναι τα αντίστοιχα πλήθη των επιμέρους ενεργειακών ζητήσεων του κάτω επιπέδου της ιεραρχίας και της συναθροισμένης ενεργειακής ζήτησης του πάνω επιπέδου. Η συναθροισμένη χρονοσειρά δίνεται ως :

$$Y_t = \sum_{i=2}^{21} Y_{i,t} \quad (7.1)$$

Ορίζεται η μήτρα συνάθροισης S διαστάσεων 21×20 , για την οποία ισχύει : $Y_t = S \cdot Y_{k,t}$ (7.2)

7.2.4 Επικύρωση και Αξιολόγηση Προβλέψεων

Ελέγχεται η ακρίβεια των προβλέψεων και η απόδοση της μεθόδου συγκριτικά με τη μέθοδο απλής εκθετικής εξομάλυνσης SES, που λαμβάνεται ως benchmark στην εργασία αυτή, με βάση τον στατιστικό δείκτη s-MAPE, που παρουσιάστηκε στην ενότητα 6.3.6. Η ακρίβεια των προβλέψεων ελέγχεται ανά επίπεδο ιεραρχίας, ανά ορίζοντα πρόβλεψης (24h, 72h, 168h) και ανά ιεραρχική προσέγγιση.

Εφαρμόζουμε τη μέθοδο «rolling horizon» για τον υπολογισμό της απόδοσης της μεθόδου προβλέψεων με μεγαλύτερης ακρίβεια Σύμφωνα με τη μέθοδο αυτή, θεωρούμε μεταβλητό (κινούμενο) σημείο εκκίνησης για τις προβλέψεις. Αρχικά παράγουμε προβλέψεις έχοντας στη διάθεση μας N-2h ιστορικές παρατηρήσεις και προβλέπουμε h τιμές μπροστά, όπου h : ορίζοντας πρόβλεψης. Συγκρίνουμε με τις πραγματικές τιμές και υπολογίζουμε σφάλμα s-MAPE. Ο λόγος που επιλέγουμε N-2· h παρατηρήσεις ως δείγμα εξάσκησης της μεθόδου και 2· h ως δείγμα ελέγχου, είναι διότι προκειμένου να εφαρμόσουμε τη συγκεκριμένη μέθοδο, χρειαζόμαστε δείγμα με διπλάσιο μήκος παρατηρήσεων σε σχέση με το μέγιστο ορίζοντα πρόβλεψης h. Στη συνέχεια, θεωρούμε ότι έχουμε στη διάθεση μας N+1 παρατηρήσεις και πάλι προβλέπουμε h τιμές μπροστά με σημείο εκκίνησης N+2. Συγκρίνουμε και πάλι με τις πραγματικές τιμές και υπολογίζουμε το αντίστοιχο σφάλμα. Υλοποιούμε αυτή την διαδικασία έως ότου το σημείο εκκίνησης φτάσει να πάρει την τιμή (N-h), όπου θα παράγουμε προβλέψεις για το τελευταίο διάστημα παρατηρήσεων [N-h,N]. Με τον τρόπο αυτό παράγουμε προβλέψεις περισσότερες από μια φορά με μεταβλητό σημείο εκκίνησης και ελέγχουμε το σφάλμα της μεθόδου πρόβλεψης ως το μέσο όρο των σφαλμάτων πρόβλεψης για κάθε διαφορετικό σημείο εκκίνησης. Ο αναδρομικός τύπος που δίνει τον αριθμό εκτελέσεων της μεθόδου είναι : $v = (2 \cdot h_{max}) - h + 1$. (7.8)

Στην περίπτωση του συγκεκριμένου προβλήματος, από τις 39.408 παρατηρήσεις που έχουμε στη διάθεση μας για κάθε στοιχείο της ιεραρχίας, θεωρούμε το δείγμα των πρώτων (39.408-336)=39.072 προς εκπαίδευση της μεθόδου πρόβλεψης και το δείγμα των τελευταίων 336 παρατηρήσεων προς έλεγχο των προβλέψεων. Ο λόγος που επιλέγουμε 336 παρατηρήσεις ως δείγμα ελέγχου, είναι διότι προκειμένου να εφαρμόσουμε τη συγκεκριμένη μέθοδο, χρειαζόμαστε δείγμα με διπλάσιο μήκος παρατηρήσεων σε σχέση με το μέγιστο ορίζοντα πρόβλεψης h=168. Ξεκινάμε να παράγουμε προβλέψεις έχοντας στη διάθεση μας (39408-336)=39072 ιστορικές παρατηρήσεις και προβλέπουμε h τιμές μπροστά. Το τελευταίο βήμα που παράγουμε προβλέψεις είναι όταν φτάσουμε να έχουμε στην διάθεση μας N-h παρατηρήσεις.

Τελικά για ορίζοντα πρόβλεψης 168h, λαμβάνουμε ως σφάλμα της μεθόδου το μέσο όρο σφαλμάτων 169 εκτελέσεων, για ορίζοντα πρόβλεψης 72h, λαμβάνουμε ως σφάλμα της μεθόδου το μέσο όρο σφαλμάτων 265 εκτελέσεων και για ορίζοντα πρόβλεψης 24h, λαμβάνουμε ως σφάλμα της μεθόδου το μέσο όρο σφαλμάτων 313 εκτελέσεων.

7.2.5 Αποτελέσματα

Οι πίνακες 7.1, 7.2, 7.3 που ακολουθούν, δίνουν τα σφάλματα ανά επίπεδο και τα μέσα σφάλματα ανα ιεραρχική προσέγγιση και ανά χρονικό ορίζοντα πρόβλεψης.

| Ιεραρχική Μέθοδος Ορίζοντα Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|--------------------------------------|-------------------|------------------|-----------------|
| 24 h | 13,72 | 13,75 | 13,74 |
| 72 h | 14,74 | 14,77 | 14,77 |
| 168 h | 14,74 | 14,77 | 14,77 |

Πίνακας 7.1 : Σφάλματα s-MAPE πάνω επιπέδου ιεραρχίας μεθόδου MAPA για τους τρεις ορίζοντες πρόβλεψης

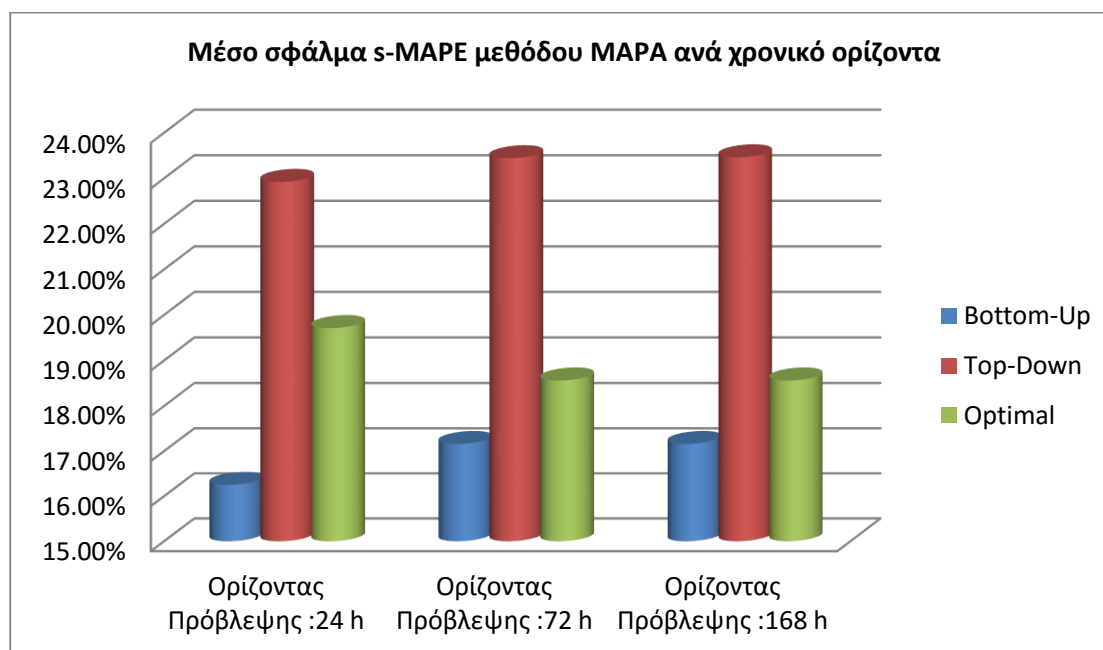
| Ιεραρχική Μέθοδος Ορίζοντα Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|--------------------------------------|-------------------|------------------|-----------------|
| 24 h | 16,38 | 23,38 | 17,93 |
| 72 h | 17,28 | 23,87 | 18,85 |
| 168 h | 17,29 | 23,89 | 18,74 |

Πίνακας 7.2 : Σφάλματα s-MAPE κάτω επιπέδου ιεραρχίας μεθόδου MAPA για τους τρεις ορίζοντες πρόβλεψης

| Ιεραρχική Μέθοδος Ορίζοντα Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|--------------------------------------|-------------------|------------------|-----------------|
| 24 h | 16,25 | 22,92 | 19,71 |
| 72 h | 17,16 | 23,44 | 18,65 |
| 168 h | 17,16 | 23,46 | 18,56 |

Πίνακας 7.3 : Μέσα σφάλματα s-MAPE ιεραρχίας μεθόδου MAPA για τους τρεις ορίζοντες πρόβλεψης

Με βάση τους πίνακες 7.1, 7.2, 7.3, φτιάχνουμε το διάγραμμα του σχήματος 7.4, που απεικονίζει τα μέσα σφάλματα s-MAPE της μεθόδου MAPA ανά χρονικό ορίζοντα.



Σχήμα 7.4 : Μέσο σφάλμα s-MAPE μεθόδου MAPA ανά χρονικό ορίζοντα

Παρατηρούμε ότι και στους τρεις χρονικούς ορίζοντες πρόβλεψης, η Bottom-Up προσέγγιση παρουσιάζει την μεγαλύτερη ακρίβεια, γεγονός που αναμενόταν καθώς η συγκεκριμένη προσέγγιση λαμβάνει περισσότερο υπόψιν τα διαφορετικά ιδιαίτερα χαρακτηριστικά των επιμέρους στοιχείων του κάτω επιπέδου. Για τον ίδιο ακριβώς λόγο, την χειρότερη ακρίβεια προβλέψεων, σε όλες τις μεθόδους πρόβλεψης, παρουσιάζει η Top-Down προσέγγιση. Η προσέγγιση Optimal παρουσιάζει βελτιωμένη ακρίβεια πρόβλεψης σε σχέση με την Top-Down, αλλά δεν πλησιάζει την ακρίβεια της Bottom-Up προσέγγισης. Στο σημείο αυτό αξίζει να σημειώσουμε, ότι η αυξημένη τιμή του μέσου σφάλματος s-MAPE στις προσεγγίσεις Top-Down και Optimal, οφείλεται σε μεγάλο βαθμό, στα ιδιαίτερα υψηλά σφάλματα που παρουσιάζονται σε συγκεκριμένα στοιχεία της ιεραρχίας, όπως μπορεί κανείς να παρατηρήσει στους τρεις πίνακες του παραρτήματος (I). Για παράδειγμα, τα μεγαλύτερα σφάλματα παρουσιάζονται στα στοιχεία (5) και (9) της ιεραρχίας. Οι συγκεκριμένες περιοχές, είτε παρουσιάζουν πολύ μικρές ενεργειακές ζητήσεις σε σχέση με τις άλλες περιοχές (η περιοχή 5 έχει ωριαίες παρατηρήσεις περίπου 500kWh έναντι 5.000-60.000KWh των υπόλοιπων περιοχών), είτε αποτελούν μονάδα παραγωγής (περιοχή 9), οπότε η ζήτηση τους δεν εξαρτάται τόσο από τα ιστορικά δεδομένα αλλά από τις εκάστοτε ανάγκες παραγωγής ενέργειας από την συγκεκριμένη περιοχή. Όταν επιχειρούμε, λοιπόν, να παράγουμε προβλέψεις για τις συγκεκριμένες περιοχές με κάποια διαδικασία επιμερισμού (Top-Down ή Optimal), είναι πιθανόν να προκύψουν πολύ άστοχες προβλέψεις (ακόμη και αρνητικές) λόγω της ανομοιογένειας των μεγεθών. Οι συγκεκριμένες προβλέψεις συμβάλλουν στην αύξηση του μέσου δείκτη s-Mape, ιδιαίτερα για τις προσεγγίσεις Top-Down και Optimal, που παρουσιάζουν μέσα σφάλματα 22,92% και 19,71% αντίστοιχα έναντι 16,25% για την προσέγγιση Bottom-Up. Η τιμή του αντίστοιχου ενδιάμεσου δείκτη s-Mape για τις προσεγγίσεις Top-Down και Optimal είναι περίπου 20,2% και 17,5% αντίστοιχα.

Στους πίνακες 7.4, 7.5, 7.6 που ακολουθούν, παρουσιάζεται ο λόγος των σφαλμάτων πρόβλεψης των μεθόδων MAPA και SES ανά επίπεδο ιεραρχίας, ανά ιεραρχική προσέγγιση και ανά χρονικό ορίζοντα πρόβλεψης.

| Λόγος σφαλμάτων πάνω επιπέδου s-MAPE (MAPA/SES) Οριζοντίας Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|--|-------------------|------------------|-----------------|
| 24 h | 0,933 | 0,932 | 0,932 |
| 72 h | 0,946 | 0,947 | 0,947 |
| 168 h | 0,918 | 0,921 | 0,921 |

Πίνακας 7.4 : Λόγος Σφαλμάτων sMAPE πάνω επιπέδου ιεραρχίας μεθόδων MAPA/SES για τους τρεις οριζοντες πρόβλεψης

| Λόγος σφαλμάτων κάτω επιπέδου s-MAPE (MAPA/SES) Οριζοντίας Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|--|-------------------|------------------|-----------------|
| 24 h | 0,892 | 0,989 | 0,857 |
| 72 h | 0,900 | 1,001 | 0,862 |
| 168 h | 0,877 | 1,011 | 0,847 |

Πίνακας 7.5 : Λόγος Σφαλμάτων sMAPE κάτω επιπέδου ιεραρχίας μεθόδων MAPA/SES για τους τρεις οριζοντες πρόβλεψης

| Λόγος μέσων σφαλμάτων s-MAPE (MAPA/SES) Οριζοντιας Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|--|-------------------|------------------|-----------------|
| 24 h | 0,893 | 0,988 | 0,955 |
| 72 h | 0,902 | 1,000 | 0,864 |
| 168 h | 0,878 | 1,009 | 0,850 |

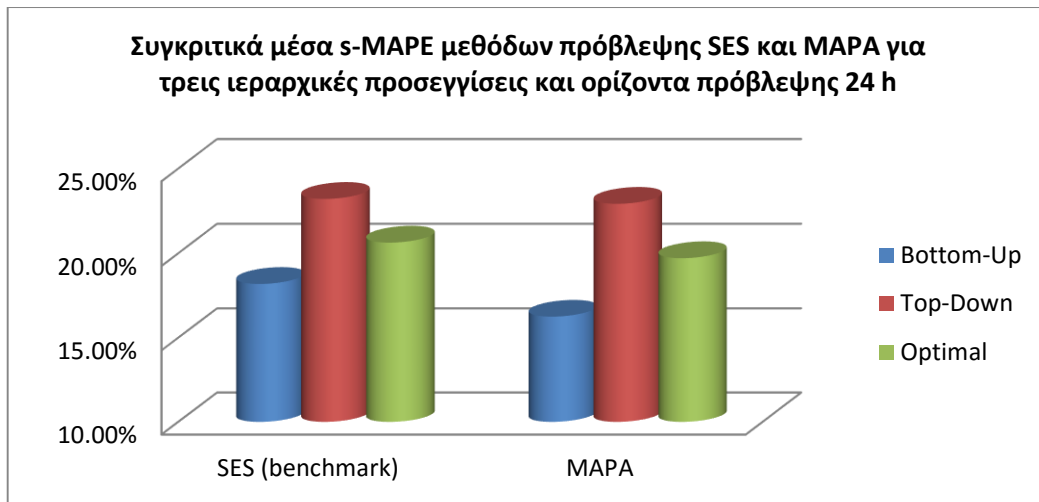
Πίνακας 7.6 : Λόγος μέσων σφαλμάτων sMAPE ιεραρχίας μεθόδων MAPA/SES για τους τρεις οριζοντιες πρόβλεψης

Επίσης, όπως φαίνεται από τους πίνακες 7.1, 7.2, 7.3, παρατηρείται, εν γένει, μικρότερη ακρίβεια προβλέψεων, καθώς αυξάνεται ο χρονικός οριζοντας. Πιο συγκεκριμένα, τα σφάλματα πρόβλεψης s-MAPE αυξάνονται από 0,6% έως 1% για τις ιεραρχικές προσεγγίσεις Bottom-Up και Top-Down, καθώς ο οριζοντας πρόβλεψης αυξάνεται από 24h σε 168h. Το γεγονός αυτό συνδέεται με την αυξημένη αβεβαιότητα που παρουσιάζει η πρόβλεψη σε μεγάλους οριζοντιες πρόβλεψης σε σχέση με τους μικρούς οριζοντιες πρόβλεψης. Αυτή είναι μια παρατήρηση που αφορά όλες τις μεθόδους πρόβλεψης και συνήθως όλες τις ιεραρχικές προσεγγίσεις. Συνεπώς, η σχετική συμπεριφορά των μεθόδων ως προς την ακρίβεια πρόβλεψης στους διαφορετικούς οριζοντιες πρόβλεψης διατηρείται. Συνεπώς, ελέγχουμε τη μέθοδο MAPA συγκριτικά με τις μεθόδους SES (benchmark) και MLR (Κεφάλαιο 8) ανά επίπεδο ιεραρχίας για έναν οριζοντια πρόβλεψης (έστω 24h) και για τις τρεις ιεραρχικές προσεγγίσεις. Τα συμπεράσματα ισχύουν και για τους τρεις οριζοντιες πρόβλεψης.

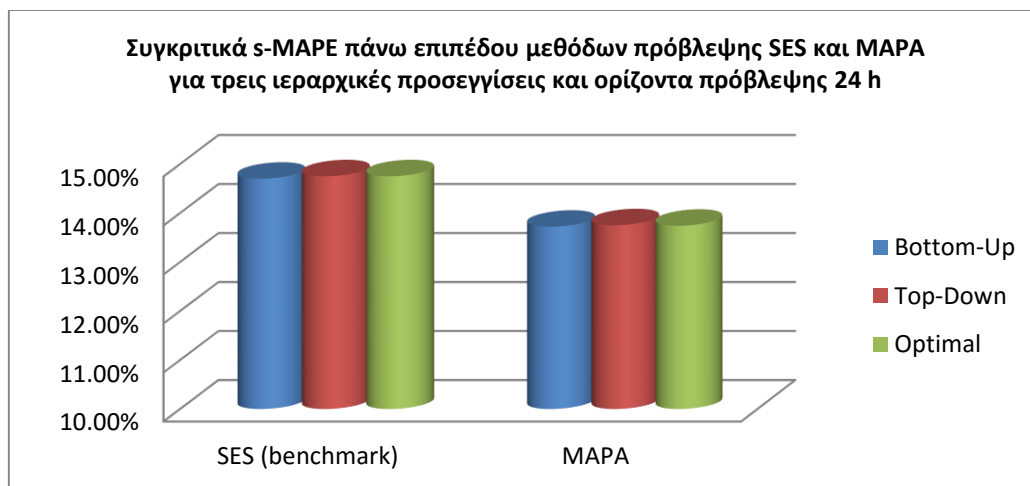
Παρατηρούμε, ότι η μέθοδος MAPA παρουσιάζει μεγαλύτερη ακρίβεια πρόβλεψης σε σχέση με τη μέθοδο SES, την οποία θεωρούμε μέθοδο αναφοράς (benchmark). Πιο συγκεκριμένα για οριζοντια πρόβλεψης 24h, η μέθοδος MAPA παρουσιάζει σε σχέση με τη μέθοδο SES :

- Μέσο κέρδος στην ακρίβεια πρόβλεψης όλης της ιεραρχίας :
 - στην προσέγγιση bottom-up 10,7%
 - στην προσέγγιση top-down 1,2%
 - στην προσέγγιση optimal 4,5%
- Μέσο κέρδος στην ακρίβεια πρόβλεψης πάνω επιπέδου της ιεραρχίας :
 - στην προσέγγιση bottom-up 6,7%
 - στην προσέγγιση top-down 6,8%
 - στην προσέγγιση optimal 6,8%
- Μέσο κέρδος στην ακρίβεια πρόβλεψης κάτω επιπέδου της ιεραρχίας :
 - στην προσέγγιση bottom-up 10,8%
 - στην προσέγγιση top-down 1,1%
 - στην προσέγγιση optimal 14,3%

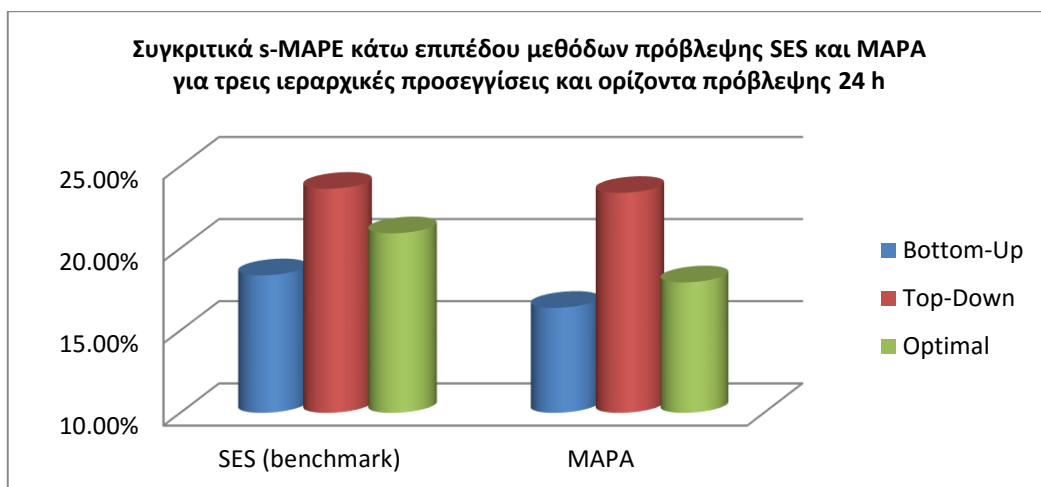
Ακολουθούν τα συγκριτικά διαγράμματα των σφαλμάτων των μεθόδων MAPA και SES για τις τρεις ιεραρχικές προσεγγίσεις και για οριζοντια πρόβλεψης 24h.



Σχήμα 7.5 : Συγκριτικά μέσα s-MAPE μεθόδων πρόβλεψης SES και MAPA για τρεις ιεραρχικές προσεγγίσεις και οριζόντια πρόβλεψη 24 h



Σχήμα 7.6 : Συγκριτικά μέσα s-MAPE πάνω επιπέδου μεθόδων πρόβλεψης SES και MAPA για τρεις ιεραρχικές προσεγγίσεις και οριζόντια πρόβλεψη 24 h



Σχήμα 7.7 : Συγκριτικά s-MAPE κάτω επιπέδου μεθόδων πρόβλεψης SES και MAPA για τρεις ιεραρχικές προσεγγίσεις και οριζόντια πρόβλεψη 24 h

Κεφάλαιο : 8 Ενεργειακές προβλέψεις μέσω μοντέλου πολλαπλής παλινδρόμησης

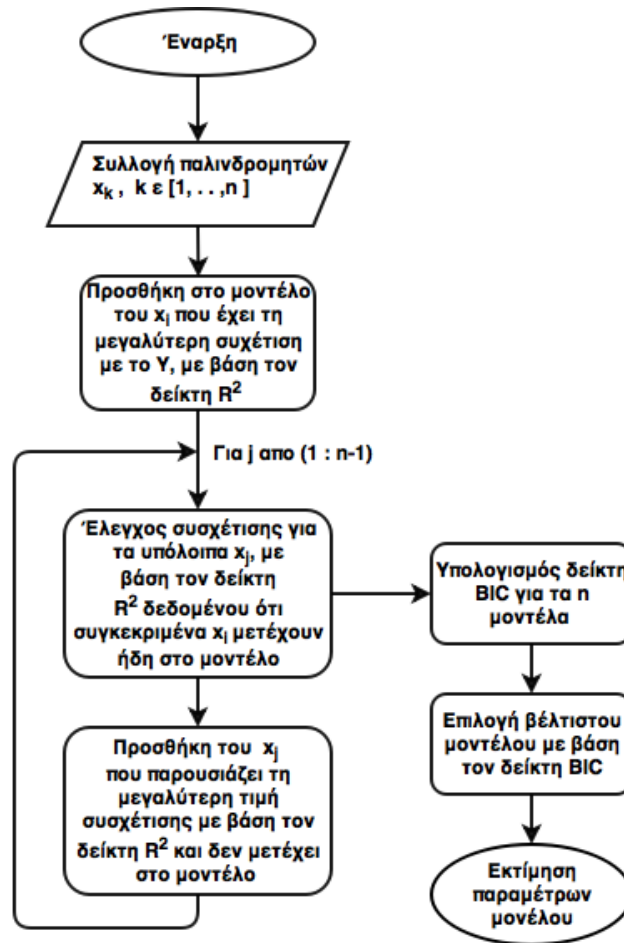
8.1 Μεθοδολογία κατασκευής μοντέλου πολλαπλής γραμμικής παλινδρόμησης

8.1.1 Ιδέα της μεθόδου

Στόχος του συγκεκριμένου μοντέλου είναι η μοντελοποίηση της διαδικασίας της πρόβλεψης μιας σύνθετης μεταβλητής, όπως η κατανάλωση ενέργειας, μέσω εξειδικευμένων μεταβλητών. Το παραμετρικό μοντέλο βασίζεται στην υπόθεση ύπαρξης κάποιας σταθερής σχέσης ανάμεσα στην προβλεπόμενη μεταβλητή και σε ορισμένες ανεξάρτητες παραμέτρους που την επηρεάζουν. Στην περίπτωση του μοντέλου αυτού, η συνάρτηση που περιγράφει το σύστημα προκύπτει από μία ορισμένη διαδικασία, σε αντίθεση με το μοντέλο χρονοσειρών στο οποίο η συνάρτηση προσδιορίζεται από το πρότυπο που έχει οριστεί από την υπόθεση. Μειονέκτημα του αιτιοκρατικού μοντέλου είναι η ανάγκη για περισσότερα δεδομένα, λόγω των διαφορετικών ανεξάρτητων μεταβλητών και μάλιστα πολλές φορές για μεγάλα χρονικά διαστήματα πριν το τρέχον. Ακόμη, οι παράγοντες που επηρεάζουν όλες τις ανεξάρτητες μεταβλητές επιδρούν και στο μοντέλο αυτό. Κατά τη διαδικασία της μοντελοποίησης λοιπόν, στόχος είναι να εντοπιστεί το πλήθος των επιπλέον βαθμών ελευθερίας από ανεξάρτητες μεταβλητές που ανταποκρίνονται καλύτερα στις μεταβολές της χρονοσειράς και την περιγράφουν όσο το δυνατόν πιο αποτελεσματικά, όπως περιγράφηκε στην ενότητα 3.3.

8.1.2 Κατασκευή μοντέλων παλινδρόμησης για όλα τα στοιχεία της ιεραρχιάς

Στοχεύουμε στην κατασκευή και επιλογή του βέλτιστου μοντέλου παλινδρόμησης για κάθε στοιχείο της ιεραρχίας με βάση τους n παλινδρομητές που προκύπτουν από τις δεδομένες μεταβλητές εισόδου. Η κατασκευή αποδοτικών και αξιόπιστων μοντέλων παλινδρόμησης συνίσταται στην επιλογή βέλτιστων παλινδρομητών για κάθε μοντέλο παλινδρόμησης για κάθε στοιχείο της ιεραρχικής δομής, όπως παρουσιάζεται στο παρακάτω σχήμα. Αυτό επιτυγχάνεται μέσω της διαδικασίας της προς τα εμπρός επιλογής, όπως περιγράφηκε στην ενότητα 3.3. Η προς τα εμπρός επιλογή ξεκινά από τη μη συμμετοχή καμιάς ανεξάρτητης μεταβλητής στο μοντέλο παλινδρόμησης και βήμα-βήμα εξετάζεται η συμμετοχή ή όχι κάποιας από τις διαθέσιμες μεταβλητές σε αυτό. Η εισαγωγή αυτή στηρίζεται, σε κάθε βήμα (εκτός του αρχικού), στο γεγονός ότι είναι γνωστό ότι ήδη κάποιες συγκεκριμένες μεταβλητές συμμετέχουν σε αυτό και η συμμετοχή μιας άλλης ελέγχεται βάσει της τιμής του δείκτη R^2 . Με τον τρόπο αυτό δημιουργούνται n πιθανά μοντέλα για κάθε στοιχείο της ιεραρχίας και επιλέγεται το βέλτιστο με βάση τον δείκτη BIC, ο οποίος παρουσιάστηκε στην ενότητα 3.3.

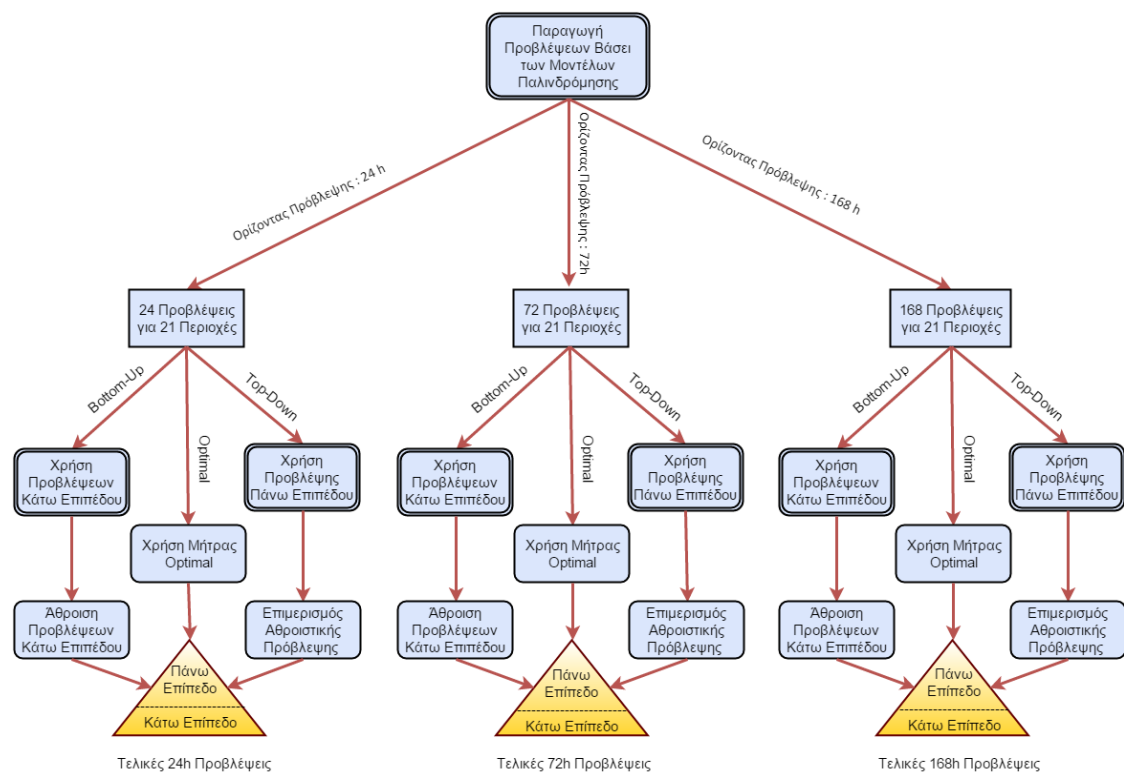


Σχήμα 8.1 : Διάγραμμα ροής μεθόδου «Forward Selection»

Αφού έχουμε ορίσει κατάλληλα τους παλινδρομητές των μοντέλων πολλαπλής γραμμικής παλινδρόμησης για κάθε στοιχείο της ιεραρχίας, προσδιορίζουμε τις παραμέτρους του μοντέλου με την διαδικασία που περιγράφηκε στην ενότητα 3.3.6.

8.1.3 Παραγωγή Προβλέψεων

Στη συνέχεια, μέσω των μοντέλων αυτών, γίνεται η παραγωγή προβλέψεων για όλες τις συνολικές και επιμέρους καταναλώσεις, για τον χρονικό ορίζοντα που μας ενδιαφέρει και μέσω των τριών ιεραρχικών μεθόδων, που παρουσιάστηκαν στην ενότητα 4.3, σχηματίζονται οι τελικές προβλέψεις, σε όλα τα επίπεδα της ιεραρχίας. Εν συνεχεία, οι προβλέψεις αυτές αξιοποιούνται ως είσοδος για τις τρεις διαφορετικές ιεραρχικές προσεγγίσεις, μέσω των οποίων προκύπτουν οι τελικές προβλέψεις ανά επίπεδο ιεραρχίας. Για την Bottom-Up προσέγγιση, απαιτούνται μόνο οι προβλέψεις του κατώτατου επιπέδου, ενώ για την Top-Down απαιτείται μόνο εκείνη της συνολικής αθροιστικής κατανάλωσης. Για την Optimal προσέγγιση απαιτούνται όλες οι προβλέψεις.



Σχήμα 8.2 : Διάγραμμα Παραγωγής Προβλέψεων για τους τρεις χρονικούς ορίζοντες και για τις 3 ιεραρχικές προσεγγίσεις

8.1.4 Υλοποίηση στην R

Όπως και στη μεθοδολογία του προηγούμενου κεφαλαίου, χρησιμοποιήσαμε την γλώσσα προγραμματισμού R. Παρακάτω παρουσιάζονται κάποιες εντολές, συναρτήσεις και πακέτα συναρτήσεων που χρησιμοποιήθηκαν στην παρούσα μεθοδολογία την υλοποίηση συγκεκριμένων διαδικασιών.

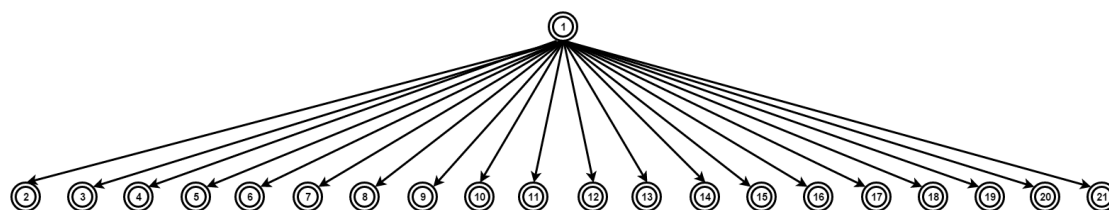
- Η συνάρτηση `lm()` αντιπροσωπεύει τη συνάρτηση ενός γραμμικού μοντέλου, όπως δηλαδή ενός γραμμικού μοντέλου παλινδρόμησης. Πρόκειται για μια συνάρτηση του βασικού πακέτου “stats” της R. Συντάσσεται με βάση τα εξής ορίσματα :
lm(formula, data, subset, weights, ...) , όπου :
 - *formula* : ο τύπος που περιγράφει το μοντέλο, για παράδειγμα $x \sim y$
 - *data* : δεδομένα μεταβλητών στο μοντέλο
 - *subset* : υποσύνολο των παρατηρήσεων που θα χρησιμοποιηθούν στη διαδικασία προσαρμογής του μοντέλου
 - *weights* : τα βάρη που θα χρησιμοποιηθούν στη διαδικασία προσαρμογής

- Το πακέτο συναρτήσεων `forecast` είναι ένα γενικό πακέτο συναρτήσεων για την πρόβλεψη από χρονοσειρές ή από μοντέλα πρόβλεψης χρονοσειρών. Η συνάρτηση `forecast` επικαλείται συγκεκριμένες μεθόδους που εξαρτώνται από την κατηγορία του πρώτου ορίσματος. Για παράδειγμα, η συνάρτηση `forecast.ts` παράγει εκτιμήσεις, χρησιμοποιώντας ETS μοντέλα (αν τα δεδομένα δεν είναι εποχιακά ή η εποχιακή περίοδος είναι 12 ή λιγότερο) . Συντάσσεται με βάση τα εξής ορίσματα :
forecast(object, h, subset, robust, weights, find.frequency, allow.multiplicative.trend, ...) , όπου :
 - *object* : μια χρονοσειρά ή ένα μοντέλο χρονοσειρών για την οποία, ή για το οποίο θέλουμε να παράγουμε προβλέψεις
 - *h* : ορίζοντας πρόβλεψης
 - *robust*: Αν αυτή η λογική μεταβλητή είναι αληθής, τότε η συνάρτηση είναι ισχυρή σε τιμές που λείπουν και ακραίες τιμές στην χρονοσειρά.
 - *find.frequency* : Αν αυτή η λογική μεταβλητή είναι αληθής, τότε η συνάρτηση ορίζει το κατάλληλο χρονικό διάστημα ως περίοδο, εάν τα δεδομένα είναι παρατηρήσεις άγνωστης περιόδου.
 - *allow.multiplicative.trend* : Αν αυτή η λογική μεταβλητή είναι αληθής, τότε επιτρέπονται μοντέλα ETS με πολλαπλασιαστική τάση. Αλλιώς, επιτρέπονται μόνο μοντέλα ETS με αθροιστική τάση ή χωρίς καθόλου τάση.

8.2 Case Study : Παραγωγή Προβλέψεων μέσω μοντέλου πολλαπλής γραμμικής παλινδρόμησης

8.2.1 Επισκόπηση Προβλήματος και Συλλογή Δεδομένων

Η μέθοδος της πολλαπλής γραμμικής παλινδρόμησης που περιγράφηκε στην παραπάνω ενότητα εφαρμόζεται σε ένα case study για 20 διαφορετικές περιοχές της Αμερικής. Στόχος ήταν να δοκιμαστεί η αποτελεσματικότητα της μεθόδου σε μια ιεραρχική δομή που απαρτίζεται από 21 χρονοσειρές, οι οποίες αναλύονται σε 20 επιμέρους χρονοσειρές για κάθε περιοχή ξεχωριστά και 1 χρονοσειρά για όλες τις περιοχές συνολικά. Ακολουθεί το ιεραρχικό διάγραμμα του προβλήματος:



Σχήμα 8.3 : Ιεραρχία εξεταζόμενου προβλήματος

Σύμφωνα με το παραπάνω σχήμα, η ιεραρχική δομή αποτελείται από 2 επίπεδα. Οι 20 χρονοσειρές, που προκύπτουν από τις 20 επιμέρους περιοχές, αποτελούν το κάτω επίπεδο της ιεραρχίας, ενώ η 1^η χρονοσειρά, που αποτελεί και το πάνω επίπεδο της ιεραρχίας, είναι το άθροισμα των 20 επιμέρους χρονοσειρών του κάτω επιπέδου. Τα δεδομένα που χρησιμοποιήθηκαν για την εφαρμογή της μεθόδου προέρχονται από τον διαγωνισμό GEFC 2012. Προκειται για ωριαίες παρατηρήσεις ενεργειακής ζήτησης (σε kWh) που εκτείνονται από τις 1/1/2004 και ώρα 01:00 έως τις 30/6/2008 και ώρα 00:00. Οι παρατηρήσεις αυτές συνοδεύονται από αντίστοιχες παρατηρήσεις ώρας, ημέρας, μήνα και έτους. Τέλος, χρησιμοποιήθηκαν, ως δεδομένα θερμοκρασίας, παρατηρήσεις θερμοκρασίας για 11 διαφορετικές ζώνες θερμοκρασίας, οι οποίες εκτείνονταν από τις 1/1/2004 και ώρα 01:00 έως τις 30/6/2008 και ώρα 00:00. Από τις συνολικά 39.408 παρατηρήσεις που έχουμε στην διάθεση μας για κάθε στοιχείο της ιεραρχίας, ανάλογα με τον ορίζοντα πρόβλεψης, θεωρούμε ως άγνωστο το δείγμα πρόσφατων παρατηρήσεων, μήκους διπλάσιου του ορίζοντα πρόβλεψης. Εξασκούμε το μοντέλο πρόβλεψης χρονοσειρών με βάση το μειωμένο δείγμα παρατηρήσεων και παράγουμε προβλέψεις για το δείγμα παρατηρήσεων που θεωρήσαμε άγνωστο.

8.2.2 Εφαρμογή της Μεθόδου Πολλαπλής Γραμμικής Παλινδρόμησης

Το μοντέλο που χρησιμοποιήθηκε στην παρούσα εργασία βασίζεται στην ιδέα της πολλαπλής γραμμικής παλινδρόμησης. Η ιδέα βασίστηκε στο benchmark μοντέλο του Global Energy Forecasting Competition 2012. Η κατανάλωση ηλεκτρικής ενέργειας ορίζεται ως μια συνάρτηση του τύπου :

$$E_i = a_0 + a_1 \cdot X_{i,1} + a_2 \cdot X_{i,2} + \dots + a_k \cdot X_{i,k} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (8.1)$$

, όπου :

- Y_i είναι η τιμή της εξαρτημένης μεταβλητής για την i παρατήρηση.
- $X_{i,1}, X_{i,2}, \dots, X_{i,k}$ είναι οι τιμές των ανεξάρτητων μεταβλητών στην i παρατήρηση (υποτίθενται γνωστές σταθερές).
- Τα ε_i είναι ανεξάρτητα $N(0, \sigma^2)$.

Στην περίπτωση αυτή a_0 είναι η $E(Y)$ για $X_1 = X_2 = \dots = X_k = 0$ ενώ το a_i ($i = 1, 2, \dots, k$) δείχνει την μεταβολή της $E(Y)$ όταν η μεταβλητή X_i αυξηθεί κατά μια μονάδα ενώ όλες οι άλλες ανεξάρτητες μεταβλητές παραμένουν σταθερές.

Ως παράμετροι/παλινδρομητές που επηρεάζουν την προβλεπόμενη εναργειακή ζήτηση θεωρήθηκαν οι ανεξάρτητες μεταβλητές ώρας, ημέρας, μήνα, έτους, θερμοκρασίας. Ελέγχουμε την συσχέτιση της ενεργειακής ζήτησης με τις παραπάνω μεταβλητές καθώς και με τις μεταβλητές που προκύπτουν ως δυνάμεις των παραπάνω μεταβλητών (π.χ. T^2) ή ως γενόμενα αυτών. Στο σημείο αυτό αξίζει να σημειωθεί, ότι είμαστε σε θέση να συσχετίσουμε την εξαρτημένη μεταβλητή με κάποια παράμετρο υψωμένο σε δύναμη ή με κάποιο γινόμενο μεταβλητών χωρίς να «παραβιαστεί» η γραμμικότητα του μοντέλου καθώς διατηρείται η γραμμική σχέση των συντελεστών παλινδρόμησης.

Ελέγχουμε ως παλινδρομητές για τα υπό-κατασκευή μοντέλα παλινδρόμησης για κάθε στοιχείο της ιεραρχίας τις εξής μεταβλητές :

- (a) Μεταβλητή Ωριαίας Τρέχουσας Θερμοκρασίας T, T^2, T^3
- (b) Μεταβλητή Μήνα (μια μεταβλητή 12 πιθανών τιμών που αντιπροσωπεύουν τους μήνες του έτους)
- (c) Μεταβλητή Ημέρας (μια μεταβλητή 2 πιθανών τιμών ανάλογα με το αν η ημέρα είναι καθημερινή ή όχι)
- (d) Μεταβλητή Ώρας (μια μεταβλητή 24 πιθανών τιμών που αντιπροσωπεύουν τις 24 ώρες μιας ημέρας)
- (e) Αλληλεπιδράσεις των παραπάνω χαρακτηριστικών :
 - (Μεταβλητή Ώρας) * (Μεταβλητή Ημέρας)
 - (Μεταβλητή Θερμοκρασίας) * (Μεταβλητή Μήνα)
 - (Μεταβλητή Θερμοκρασίας)² * (Μεταβλητή Μήνα)
 - (Μεταβλητή Θερμοκρασίας)³ * (Μεταβλητή Μήνα)
 - (Μεταβλητή Θερμοκρασίας) * (Μεταβλητή Ώρας)
 - (Μεταβλητή Θερμοκρασίας)² * (Μεταβλητή Ώρας)
 - (Μεταβλητή Θερμοκρασίας)³ * (Μεταβλητή Ώρας)

Δεν ελέγχουμε ως ανεξάρτητες μεταβλητές τις δυνάμεις των μεταβλητών ώρας, ημέρας, μήνα, έτους καθώς αυτή η υπόθεση θα οδηγούσε σε λογικό σφάλμα και αναξιόπιστες προβλέψεις. Αντίθετα, συσχετίζουμε την εξαρτημένη μεταβλητή με το τετράγωνο ή τον κύβο της μεταβλητής της θερμοκρασίας, καθώς αυτή η προσέγγιση συντελεί στην πιθανά καλύτερη προσαρμογή του μοντέλου αναφορικά με τις μεταβολές της θερμοκρασίας. Για να συσχετίσουμε την ενεργειακή ζήτηση με την θερμοκρασία σε κάθε στοιχείο της ιεραρχικής δομής, υπολογίσαμε το μέσο όρο των ωριαίων παρατηρήσεων θερμοκρασίας για τις 11 ενεργειακές ζώνες και εξαγάγαμε μια νέα χρονοσειρά παρατηρήσεων που εκτείνονταν από τις 1/1/2004 και ώρα 01:00 έως τις 30/6/2008 και ώρα 00:00.

Επιπρόσθετα, η μοντελοποίηση της μεταβλητής της ημέρας επιτυγχάνεται μέσω μιας μεταβλητής 2 πιθανών τιμών, η οποία ουσιαστικά διακρίνει τις ημέρες σε εργάσιμες και μη εργάσιμες. Η επιλογή αυτή βασίζεται στο γεγονός ότι η ενεργειακή ζήτηση καθορίζεται ικανοποιητικά από ακριβώς αυτή την διάκριση (εργάσιμη ημέρα ή μη) και όχι από τον προσδιορισμό της ακριβούς ημέρας (π.χ. Δευτέρα ή Τρίτη).

Έτσι, για παράδειγμα για την περιοχή 13 το μοντέλο παλινδρόμησης προέκυψε ως εξής. Ως πρώτος παλινδρομητής επιλέχθηκε η μεταβλητή της θερμοκρασίας καθώς παρουσίασε τη μεγαλύτερη συσχέτιση με την εξαρτημένη μεταβλητή σύμφωνα με τον δείκτη R^2 .

1^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ώρα (h) | Ημέρα (d) | Μήνας (m) | Έτος (y) | Θερμοκρασία (T) |
|-----------------------|---------|-----------|-----------|----------|-----------------|
| Τιμή R^2 | 0,091 | 0,00003 | 0,016 | 0,006 | 0,155 |

2^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ώρα (hour) | Ημέρα (day) | Μήνας (month) | Έτος (year) | Θερμοκρασία· Έτος (T· y) | Θερμοκρασία· Μήνας (T· m) | Θερμοκρασία· Ημέρα (T· d) | Θερμοκρασία· Ώρα (T· h) | Θερμοκρασία ² (T ²) |
|-----------------------|------------|-------------|---------------|-------------|--------------------------|---------------------------|---------------------------|-------------------------|--|
| Τιμή R^2 | 0,291 | 0,155 | 0,155 | 0,160 | 0,159 | 0,156 | 0,155 | 0,317 | 0,525 |

3^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ώρα (h) | Ημέρα (d) | Μήνας (m) | Έτος (y) | Θερμοκρασία ³ (T ³) |
|-----------------------|---------|-----------|-----------|----------|--|
| Τιμή R^2 | 0,650 | 0,525 | 0,526 | 0,533 | 0,550 |

4^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ημέρα (d) | Μήνας (m) | Έτος (y) | Θερμοκρασία ³ (T ³) | Θερμοκρασία· Ώρα (T· h) |
|-----------------------|-----------|-----------|----------|--|-------------------------|
| Τιμή R^2 | 0,650 | 0,651 | 0,657 | 0,666 | 0,650 |

5^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ημέρα (d) | Μήνας (m) | Έτος (y) | Θερμοκρασία· Ώρα (T· h) |
|-----------------------|-----------|-----------|----------|-------------------------|
| Τιμή R^2 | 0,666 | 0,668 | 0,672 | 0,650 |

6^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ημέρα (d) | Μήνας (m) | Έτος · Ώρα (y · h) | Έτος · Θερμοκρασία (y · T) | Θερμοκρασία · Ώρα (T · h) |
|-----------------------|-----------|-----------|--------------------|----------------------------|---------------------------|
| Τιμή R ² | 0,672 | 0,675 | 0,672 | 0,676 | 0,673 |

7^{ος} Παλινδρομητής

| Πιθανοί Παλινδρομητές | Ημέρα (d) | Μήνας (m) | Έτος · Ώρα (y · h) | Θερμοκρασία · Ώρα (T · h) |
|-----------------------|-----------|-----------|--------------------|---------------------------|
| Τιμή R ² | 0,676 | 0,679 | 0,676 | 0,676 |

Έτσι η γενική μορφή του μοντέλου για την περιοχή 13 ορίζεται ως :

$$E_i = a_0 + a_1 \cdot T_i + a_2 \cdot T_i^2 + a_3 \cdot h_i + a_4 \cdot T_i^3 + a_5 \cdot y_i + a_6 \cdot (y_i \cdot T_i) + a_7 \cdot m_i + \varepsilon_i \quad (8.2)$$

Κατ' αυτόν τον τρόπο κατασκευάζουμε τα μοντέλα παλινδρόμησης για κάθε στοιχείο της ιεραρχίας. Με βάση τα μοντέλα πρόβλεψης, παράγουμε προβλέψεις με ορίζοντα πρόβλεψης 24h, 72h και 168h.

Οι παλινδρομητές καθώς και η σειρά με την οποία επιλέγονται αυτοί για κάθε μοντέλο παρουσιάζεται στον παρακάτω πίνακα.

| Στοιχείο Ιεραρχίας Παλινδρομητές | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|----------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1 ^{ος} | h | h | h | T | h | h | h | h | h | y | h |
| 2 ^{ος} | T | T | T | T | T | T | T | T | T | m | y |
| 3 ^{ος} | T ² | T ² | T ² | T ² | T ² | T ² | T ² | T ² | T ² | y · m | d |
| 4 ^{ος} | T ³ | T ³ | T ³ | T ³ | T ³ | T ³ | T ³ | T ³ | T ³ | h | m |
| 5 ^{ος} | y | y | y | y · T | h · T | y | y | y | h · T | y · h | m · h |
| 6 ^{ος} | y · T | y · T | y · T | m | m | y · T | y · T | y · T | m | T | T |
| 7 ^{ος} | h · T | y · h | y · h | y · T | m · T | h · T | y · h | h · T | y | T ² | T ² |
| Στοιχείο Ιεραρχίας Παλινδρομητές | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | |
| 1 ^{ος} | h | T | h | h | h | h | h | h | h | h | |
| 2 ^{ος} | y | T ² | T | T | T | y | T | T | T | T | |
| 3 ^{ος} | m | h | T ² | T ² | T ² | T | T ² | T ² | T ² | T ² | |
| 4 ^{ος} | y · h | T ³ | T ³ | T ³ | T ³ | T ² | d | T ³ | d | T ³ | |
| 5 ^{ος} | m · h | y | m | m | y | T ³ | y | y | y | y | |
| 6 ^{ος} | y · m | y · T | y | m · t | y · T | m | T ³ | h · t | T ³ | h · T | |
| 7 ^{ος} | t | m | h · T | h · T | h · T | y · T | h · T | m | d · T | m | |

Πίνακας 8.1 : Παλινδρομητές που επιλέχθηκαν για τα μοντέλα παλινδρόμησης

Οι αντίστοιχες τιμές του δείκτη R², μετά την προσθήκη του εκάστοτε παλινδρομητή, παρουσιάζονται στον πίνακα που ακολουθεί.

| | | | | | | | | | | | |
|--------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Στοιχείο / Παλινδρομητές | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 ^{ος} | 0,1139 | 0,1206 | 0,1217 | 0,1948 | 0,1213 | 0,1237 | 0,1217 | 0,1241 | 0,0087 | 0,2628 | 0,1089 |
| 2 ^{ος} | 0,1455 | 0,1504 | 0,1538 | 0,4050 | 0,1906 | 0,1584 | 0,1538 | 0,2600 | 0,0207 | 0,2991 | 0,1312 |
| 3 ^{ος} | 0,7028 | 0,5825 | 0,5956 | 0,5766 | 0,6590 | 0,6094 | 0,5956 | 0,6298 | 0,0737 | 0,3839 | 0,1343 |
| 4 ^{ος} | 0,7306 | 0,6126 | 0,6255 | 0,5883 | 0,6845 | 0,6394 | 0,6255 | 0,6786 | 0,0906 | 0,4062 | 0,1343 |
| 5 ^{ος} | 0,7541 | 0,6298 | 0,6431 | 0,5950 | 0,6901 | 0,6550 | 0,6431 | 0,7073 | 0,0924 | 0,4063 | 0,1398 |
| 6 ^{ος} | 0,7580 | 0,6330 | 0,6465 | 0,6016 | 0,6919 | 0,6585 | 0,6465 | 0,7164 | 0,0943 | 0,4084 | 0,1411 |
| 7 ^{ος} | 0,7589 | 0,6331 | 0,6466 | 0,6027 | 0,6923 | 0,6587 | 0,6466 | 0,7184 | 0,0943 | 0,5108 | 0,7149 |
| Στοιχείο / Παλινδρομητές | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | |
| 1 ^{ος} | 0,1278 | 0,1546 | 0,0852 | 0,0717 | 0,0573 | 0,0855 | 0,0839 | 0,0779 | 0,1223 | 0,0605 | |
| 2 ^{ος} | 0,1425 | 0,5253 | 0,1130 | 0,0763 | 0,1190 | 0,1059 | 0,1438 | 0,1144 | 0,1790 | 0,0749 | |
| 3 ^{ος} | 0,1445 | 0,6496 | 0,7102 | 0,6905 | 0,7544 | 0,1164 | 0,7376 | 0,7120 | 0,6211 | 0,4680 | |
| 4 ^{ος} | 0,1500 | 0,6662 | 0,7257 | 0,7092 | 0,7720 | 0,7573 | 0,7584 | 0,7307 | 0,6549 | 0,4807 | |
| 5 ^{ος} | 0,1502 | 0,6721 | 0,7339 | 0,7146 | 0,7769 | 0,7664 | 0,7740 | 0,7421 | 0,6783 | 0,4899 | |
| 6 ^{ος} | 0,1507 | 0,6758 | 0,7407 | 0,7226 | 0,7801 | 0,7708 | 0,7740 | 0,7476 | 0,6783 | 0,4975 | |
| 7 ^{ος} | 0,1511 | 0,6790 | 0,7447 | 0,7266 | 0,7812 | 0,7737 | 0,7789 | 0,7528 | 0,6835 | 0,5022 | |

Πίνακας 8.2 : Τιμή Δείκτη R^2 με την προσθήκη κάθε νέου παλινδρομητή στα μοντέλα παλινδρόμησης

8.2.3 Παραγωγή Ιεραρχικών Προβλέψεων

Έχοντας στην διάθεση μας πλήρεις προβλέψεις για όλες τις χρονοσειρές, όλων των επιπέδων προχωρήσαμε στην παραγωγή των τελικών προβλέψεων ανά επίπεδο, μέσω των τριών διαφορετικών ιεραρχικών προσεγγίσεων.

Στην περίπτωση του συγκεκριμένου προβλήματος, έχουμε : $m = m_0 + m_1 = 20+1=21$ (8.3)

,όπου m_0 και m_1 είναι τα αντίστοιχα πλήθη των επιμέρους ενεργειακών ζητήσεων του κάτω επιπέδου της ιεραρχίας και της συναθροισμένης ενεργειακής ζήτησης του πάνω επιπέδου.

Η συναθροισμένη χρονοσειρά δίνεται ως : $Y_t = \sum_{i=2}^{21} Y_{i,t}$ (8.4)

Ορίζεται η μήτρα συνάθροισης S διαστάσεων 21×20 , για την οποία ισχύει : $Y_t = S \cdot Y_{k,t}$.

Μήτρα Συνάθροισης

$$S_{[21 \times 20]} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (8.5)$$

Ακολουθούν οι μήτρες σύμφωνα με τις οποίες υπολογίσαμε τις τελικές προβλέψεις για κάθε μία από τις ιεραρχικές μεθόδους.

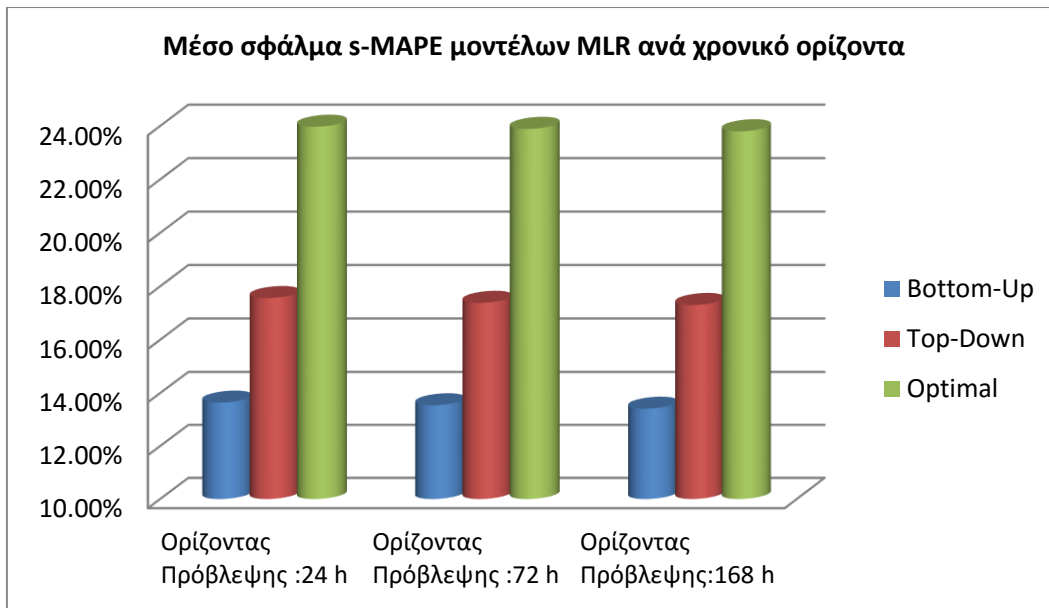
Bottom-Up

$$P_{[20 \times 21]} = [0_{[20 \times 1]} \mid I_{[20 \times 20]}] = \begin{pmatrix} 0 \mid 100000000000000000000 \\ 0 \mid 010000000000000000000 \\ 0 \mid 001000000000000000000 \\ 0 \mid 000100000000000000000 \\ 0 \mid 000010000000000000000 \\ 0 \mid 000001000000000000000 \\ 0 \mid 000000100000000000000 \\ 0 \mid 000000010000000000000 \\ 0 \mid 000000001000000000000 \\ 0 \mid 000000000100000000000 \\ 0 \mid 000000000010000000000 \\ 0 \mid 000000000001000000000 \\ 0 \mid 000000000000100000000 \\ 0 \mid 000000000000010000000 \\ 0 \mid 000000000000001000000 \\ 0 \mid 000000000000000100000 \\ 0 \mid 000000000000000010000 \\ 0 \mid 000000000000000001000 \\ 0 \mid 000000000000000000100 \\ 0 \mid 000000000000000000010 \\ 0 \mid 000000000000000000001 \end{pmatrix} \quad (8.6)$$

Top-Down

$$P_{[20 \times 21]} = [W_{[20 \times 1]} \mid 0_{[20 \times 20]}] = \begin{pmatrix} w_2 \mid 000000000000000000000 \\ w_3 \mid 000000000000000000000 \\ w_4 \mid 000000000000000000000 \\ w_5 \mid 000000000000000000000 \\ w_6 \mid 000000000000000000000 \\ w_7 \mid 000000000000000000000 \\ w_8 \mid 000000000000000000000 \\ w_9 \mid 000000000000000000000 \\ w_{00} \mid 000000000000000000000 \\ w_{00} \mid 000000000000000000000 \\ w_{02} \mid 000000000000000000000 \\ w_{03} \mid 000000000000000000000 \\ w_{04} \mid 000000000000000000000 \\ w_{05} \mid 000000000000000000000 \\ w_{06} \mid 000000000000000000000 \\ w_{07} \mid 000000000000000000000 \\ w_{08} \mid 000000000000000000000 \\ w_{09} \mid 000000000000000000000 \\ w_{20} \mid 000000000000000000000 \\ w_{20} \mid 000000000000000000000 \end{pmatrix} \quad (8.7)$$

,όπου w_i : τα βάρη επιμερισμού, για τα οποία ισχύει : $\sum_{i=2}^{21} w_i = 1$ (8.8)



Σχήμα 8.4 : Μέσο σφάλμα s-MAPE μεθόδου MLR ανά χρονικό ορίζοντα

Παρατηρούμε ότι και στους τρεις χρονικούς ορίζοντες πρόβλεψης, η Bottom-Up προσέγγιση παρουσιάζει την μεγαλύτερη ακρίβεια, γεγονός που αναμενόταν καθώς η συγκεκριμένη προσέγγιση λαμβάνει περισσότερο υπ'όψιν τα διαφορετικά ιδιαίτερα χαρακτηριστικά των επιμέρους στοιχείων του κάτω επιπέδου. Για τον ίδιο ακριβώς λόγο, η Top-Down προσέγγιση παρουσιάζει χειρότερη μέση ακρίβεια προβλέψεων σε σχέση με την Bottom-Up, σε όλες τις μεθόδους πρόβλεψης. Η προσέγγιση Optimal παρουσιάζει την χειρότερη μέση ακρίβεια πρόβλεψης. Στο σημείο αυτό αξίζει να σημειώσουμε, ότι η αυξημένη τιμή του μέσου σφάλματος s-MAPE στις προσεγγίσεις Top-Down και Optimal (ιδιαίτερα στην προσέγγιση Optimal) οφείλεται σε μεγάλο βαθμό, στα ιδιαίτερα υψηλά σφάλματα που παρουσιάζονται σε συγκεκριμένα στοιχεία της ιεραρχίας, όπως μπορεί κανείς να παρατηρήσει στους τρεις πίνακες του παραρτήματος (II). Για παράδειγμα, τα μεγαλύτερα σφάλματα παρουσιάζονται στα στοιχεία (5) και (9) της ιεραρχίας. Οι συγκεκριμένες περιοχές, είτε παρουσιάζουν πολύ μικρές ενεργειακές ζητήσεις σε σχέση με τις άλλες περιοχές (η περιοχή 5 έχει ωριαίες παρατηρήσεις περίπου 500kWh έναντι 5.000-60.000kWh των υπόλοιπων περιοχών), είτε αποτελούν μονάδα παραγωγής (περιοχή 9), οπότε η ζήτηση τους δεν εξαρτάται τόσο από τα ιστορικά δεδομένα αλλά από τις εκάστοτε ανάγκες παραγωγής ενέργειας από την συγκεκριμένη περιοχή. Όταν επιχειρούμε, λοιπόν, να παράγουμε προβλέψεις για τις συγκεκριμένες περιοχές με κάποια διαδικασία επιμερισμού (Top-Down ή Optimal), είναι πιθανόν να προκύψουν πολύ άστοχες προβλέψεις (ακόμη και αρνητικές) λόγω της ανομοιογένειας των μεγεθών. Οι συγκεκριμένες προβλέψεις συμβάλλουν στην αύξηση του μέσου δείκτη s-Mape, ιδιαίτερα για τις προσεγγίσεις Top-Down και Optimal, που παρουσιάζουν μέσα σφάλματα 17,56% και 23,97% αντίστοιχα έναντι 13,64% για την προσέγγιση Bottom-Up. Η τιμή του αντίστοιχου ενδιάμεσου δείκτη s-Mape για τις προσεγγίσεις Top-Down και Optimal είναι περίπου 12,3 % και 10,1% αντίστοιχα. Στους πίνακες 8.6, 8.7, 8.8 που ακολουθούν, παρουσιάζεται ο λόγος των σφαλμάτων πρόβλεψης των μεθόδων MLR και SES ανά επίπεδο ιεραρχίας, ανά ιεραρχική προσέγγιση και ανά χρονικό ορίζοντα πρόβλεψης.

| Λόγος σφαλμάτων s-MAPE (MLR/SES) Οριζοντιας Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|---|-------------------|------------------|-----------------|
| 24 h | 0,46 | 0,46 | 0,46 |
| 72 h | 0,44 | 0,44 | 0,44 |
| 168 h | 0,42 | 0,43 | 0,42 |

Πίνακας 8.6 : Λόγος σφαλμάτων s-MAPE πάνω επιπέδου ιεραρχίας μεθόδων MLR/SES για τους τρεις οριζοντιες πρόβλεψης

| Λόγος σφαλμάτων s-MAPE (MLR/SES) Οριζοντιας Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|---|-------------------|------------------|-----------------|
| 24 h | 0,76 | 0,77 | 1,19 |
| 72 h | 0,72 | 0,72 | 1,13 |
| 168 h | 0,70 | 0,75 | 1,11 |

Πίνακας 8.7 : Λόγος σφαλμάτων s-MAPE κάτω επιπέδου ιεραρχίας μεθόδων MLR/SES για τους τρεις οριζοντιες πρόβλεψης

| Λόγος σφαλμάτων s-MAPE (MLR/SES) Οριζοντιας Πρόβλεψης | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
|---|-------------------|------------------|-----------------|
| 24 h | 0,75 | 0,76 | 1,16 |
| 72 h | 0,71 | 0,74 | 1,11 |
| 168 h | 0,69 | 0,74 | 1,09 |

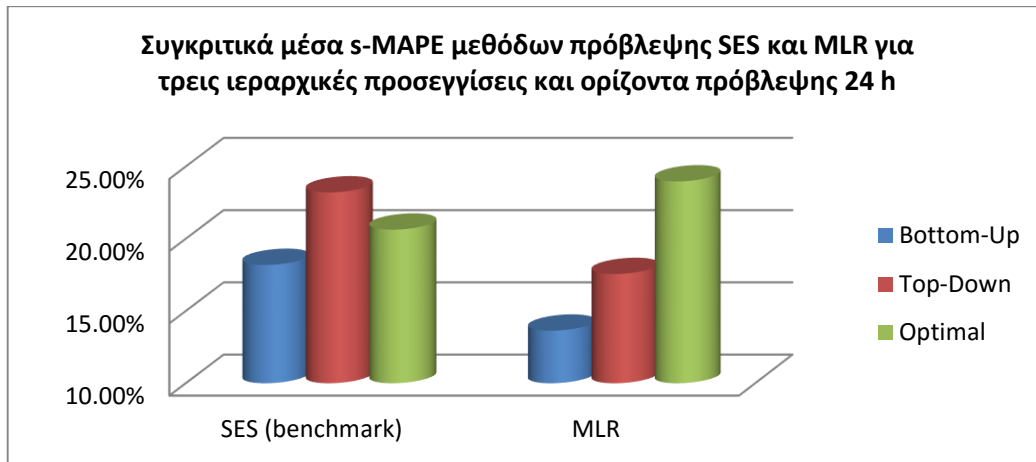
Πίνακας 8.8 : Λόγος μέσων σφαλμάτων s-MAPE πάνω ιεραρχίας μεθόδων MLR/SES για τους τρεις οριζοντιες πρόβλεψης

Επίσης, όπως φαίνεται από τους πίνακες 8.3, 8.4, 8.5, παρατηρείται, ότι ο χρονικός οριζοντιας πρόβλεψης έχει μικρή ως αμελητέα επιρροή στα σφάλματα. Συνεπώς, η σχετική συμπεριφορά των μεθόδων ως προς την ακρίβεια πρόβλεψης στους διαφορετικούς οριζοντιες πρόβλεψης διατηρείται. Συνεπώς, ελέγχουμε τη μέθοδο MLR συγκριτικά με τις μεθόδους SES (benchmark) και MAPA (Κεφάλαιο 7) ανά επίπεδο ιεραρχίας για έναν οριζοντιας πρόβλεψης (έστω 24h) και για τις τρεις ιεραρχικές προσεγγίσεις. Τα συμπεράσματα ισχύουν και για τους τρεις οριζοντιες πρόβλεψης.

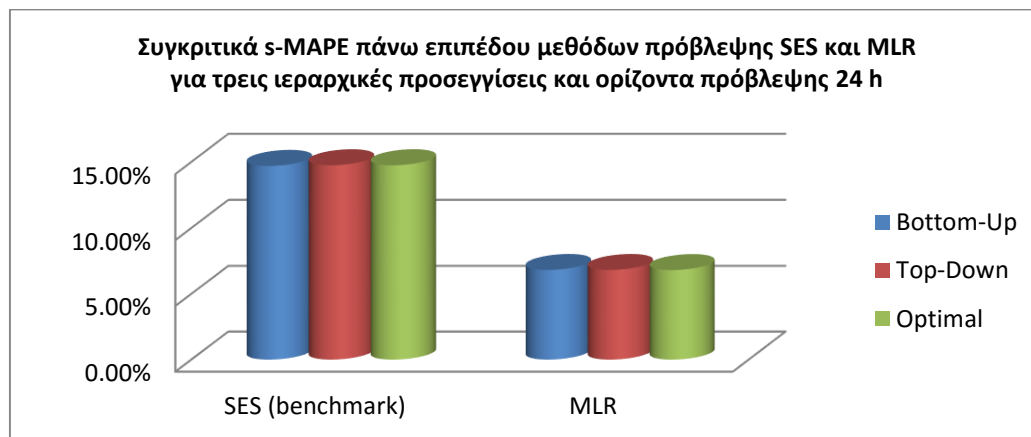
Παρατηρούμε, ότι η μέθοδος MLR παρουσιάζει μεγαλύτερη ακρίβεια πρόβλεψης σε σχέση με τη μέθοδο SES, την οποία θεωρούμε μέθοδο αναφοράς (benchmark). Πιο συγκεκριμένα για οριζοντιας πρόβλεψης 24h, η μέθοδος MLR παρουσιάζει σε σχέση με τη μέθοδο SES :

- Μέσο κέρδος στην ακρίβεια πρόβλεψης όλης της ιεραρχίας :
 - στην προσέγγιση bottom-up 25%
 - στην προσέγγιση top-down 24%
- Μέσο κέρδος στην ακρίβεια πρόβλεψης πάνω επιπέδου της ιεραρχίας :
 - στην προσέγγιση bottom-up 54%
 - στην προσέγγιση top-down 54%
- Μέσο κέρδος στην ακρίβεια πρόβλεψης κάτω επιπέδου της ιεραρχίας :
 - στην προσέγγιση bottom-up 24%
 - στην προσέγγιση top-down 23%

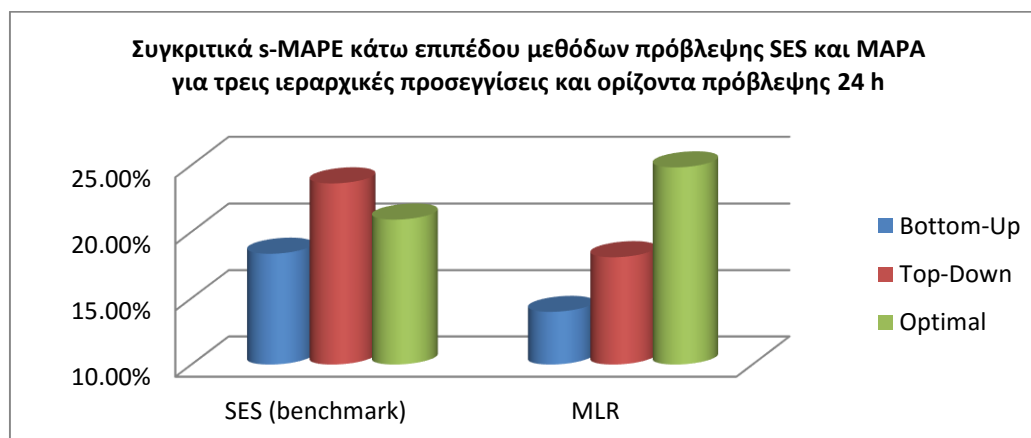
Ακολουθούν τα συγκριτικά διαγράμματα των σφαλμάτων των μεθόδων MLR και SES για τις τρεις ιεραρχικές προσεγγίσεις και για οριζόντα πρόβλεψη 24h.



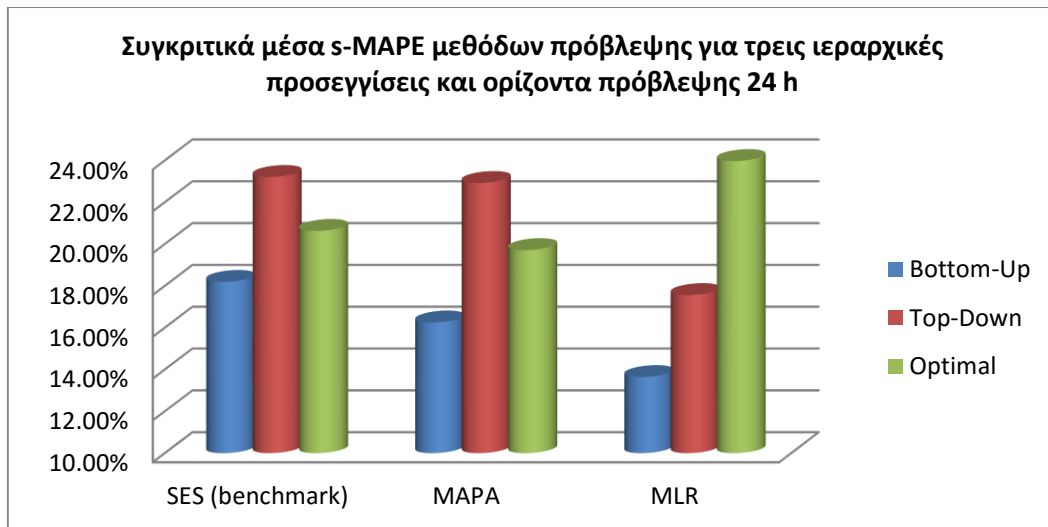
Σχήμα 8.5 : Συγκριτικά μέσα s-MAPE μεθόδων πρόβλεψης SES και MLR για τρεις ιεραρχικές προσεγγίσεις και οριζόντα πρόβλεψη 24 h



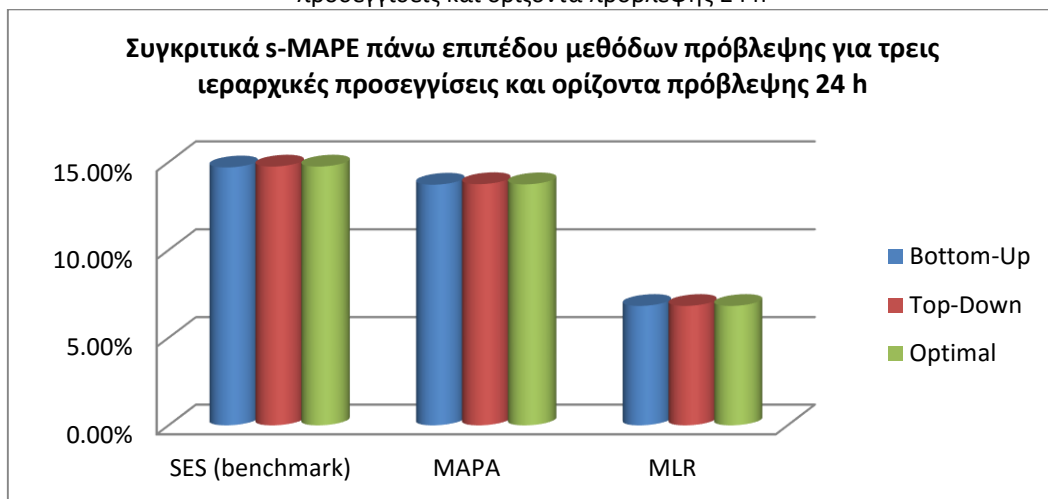
Σχήμα 8.6 : Συγκριτικά μέσα s-MAPE πάνω επιπέδου μεθόδων πρόβλεψης SES και MLR για τρεις ιεραρχικές προσεγγίσεις και οριζόντα πρόβλεψη 24 h



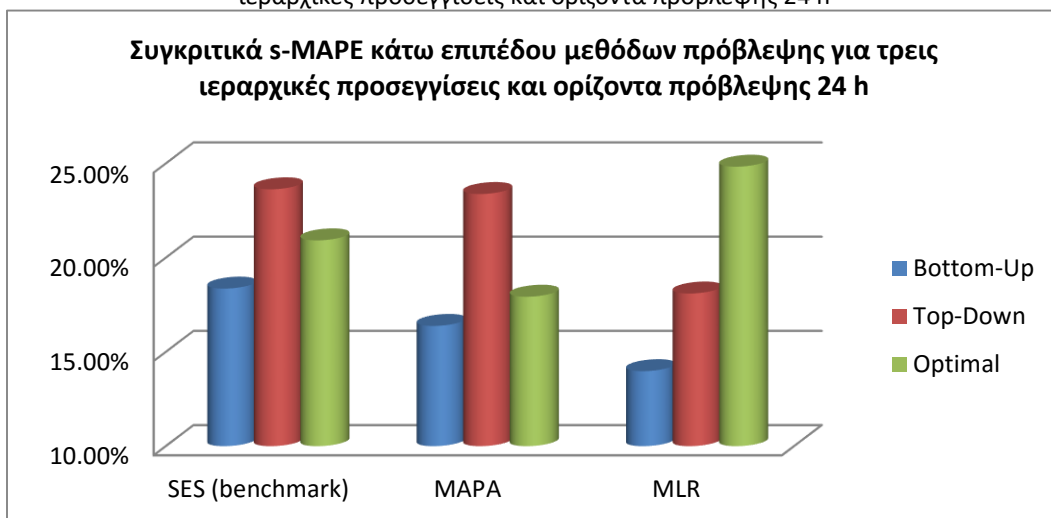
Σχήμα 8.7 : Συγκριτικά s-MAPE κάτω επιπέδου μεθόδων πρόβλεψης SES και MAPA για τρεις ιεραρχικές προσεγγίσεις και οριζόντα πρόβλεψη 24 h



Σχήμα 8.8 : Συγκριτικά μέσα s-MAPE μεθόδων πρόβλεψης SES και MLR για τρεις ιεραρχικές προσεγγίσεις και οριζόντια πρόβλεψη 24 h



Σχήμα 8.9 : Συγκριτικά μέσα s-MAPE πάνω επιπέδου μεθόδων πρόβλεψης SES και MLR για τρεις ιεραρχικές προσεγγίσεις και οριζόντια πρόβλεψη 24 h



Σχήμα 8.10 : Συγκριτικά s-MAPE κάτω επιπέδου μεθόδων πρόβλεψης SES και MAPA για τρεις ιεραρχικές προσεγγίσεις και οριζόντια πρόβλεψη 24 h

Κεφάλαιο 9 : Συμπεράσματα

9.1 Συνοπτικά αποτελέσματα εργασίας

Στην παρούσα διπλωματική εργασία, εξετάσθηκε η χρήση δύο βασικών μεθόδων πρόβλεψης για την παραγωγή ενεργειακών προβλέψεων σε ιεραρχικά δομημένες χρονοσειρές ενεργειακής ζήτησης. Πιο συγκεκριμένα, ερευνήθηκε η απόδοση μοντέλων πολλαπλής γραμμικής παλινδρόμησης (MLR) και μεθόδων πρόβλεψης χρονοσειρών με χρονική συνάθροιση (MAPA) σε ιεραρχικά δομημένα ενεργειακά δεδομένα, σε σχέση με την απόδοση της μεθόδου SES, η οποία θεωρείται μέθοδος αναφοράς. Τα ενεργειακά δεδομένα, πάνω στα οποία εφαρμόσαμε τις μεθόδους πρόβλεψης χωρίστηκαν σε δεδομένα εκπαίδευσης, με βάση τα οποία εξασκήσαμε τις μεθόδους πρόβλεψης και σε δεδομένα εξιολόγησης των μεθόδων, με βάση τα οποία αξιολογήσαμε τις μεθόδους..

Στο πλαίσιο μεθόδων πρόβλεψης χρονοσειρών με χρονική συνάθροιση (MAPA), αρχικά απεποχικοποιήσαμε τις χρονοσειρές όλων των στοιχείων της ιεραρχίας με βάση το πολλαπλασιαστικό πρότυπο εποχιακότητας. Στη συνέχεια, εφαρμόσαμε χρονική συνάθροιση στις χρονοσειρές και εξάγαμε απεποχικοποιημένες συναθροισμένες χρονοσειρές για τα διάφορα επίπεδα συνάθροισης. Εφαρμόσαμε εκθετική εξομάλυνση σταθερού επιπέδου (λόγω της φύσης των δεδομένων) στις χρονοσειρές αυτές και παρήχθησαν προβλέψεις της συνιστώσας του επιπέδου, ως μέσος όρος των προβλέψεων της συνιστώσας του επιπέδου όλων των συναθροισμένων απεποχικοποιημένων χρονοσειρών. Οι προβλέψεις αυτές της συνιστώσας του επιπέδου, συνδυάστηκαν με τις προβλέψεις της συνιστώσας της εποχιακότητας σύμφωνα με το πολλαπλασιαστικό πρότυπο εποχιακότητας για να δώσουν τις τελικές ενεργειακές προβλέψεις. Οι προβλέψεις της συνιστώσας της εποχιακότητας παρήχθησαν εξωτερικά με βάση τις ιστορικές παρατηρήσεις της συνιστώσας της εποχιακότητας των αρχικών χρονοσειρών.

Στο πλαίσιο των μοντέλων πολλαπλής γραμμικής παλινδρόμησης, αρχικά ορίσαμε όλους τους πιθανούς παλινδρομητές. Στόχος μας ήταν να συσχετίσουμε την εξαρτημένη μεταβλητή της ενεργειακής ζήτησης με διάφορες ανεξάρτητες μεταβλητές, όπως η θερμοκρασία, η ώρα, η μέρα και ο μήνας, για τις οποίες διαθέταμε ιστορικά στοιχεία. Για κάθε στοιχείο της ιεραρχίας συσχετίσαμε την ενεργειακή ζήτηση με τους διάφορους παλινδρομητές και κάνοντας χρήση του δείκτη R^2 , καταρτίσαμε υποψήφια μοντέλα παλινδρόμησης με αυτούς τους παλινδρομητές που εμφάνισαν τη μεγαλύτερη συσχέτιση με την εξαρτημένη μεταβλητή. Ανάλογα με το πλήθος των παλινδρομητών, δημιουργήθηκε αντίστοιχος αριθμός υποψήφια μοντέλων για κάθε στοιχείο της ιεραρχίας και επιλέχθηκε το βέλτιστο για κάθε στοιχείο σύμφωνα με το κριτήριο πληροφορίας BIC. Ουσιαστικά, η κατασκευή των παλινδρομικών μοντέλων έγκειται στην κατάλληλη επιλογή παλινδρομητών και στον υπολογισμό των παραμέτρων των μοντέλων. Με βάση τις παραμέτρους του μοντέλου και θεωρώντας τις αντίστοιχες εισόδους των παλινδρομητών, παρήγαμε εκτιμήσεις ενεργειακής ζήτησης για όλα τα στοιχεία της ιεραρχίας.

Και για τις δύο μεθόδους ελέχθηκαν 3 διαφορετικές ιεραρχικές προσεγγίσεις, η Bottom-Up, η Top-Down και η Optimal, ενώ παρήχθησαν προβλέψεις για τρεις διαφορετικούς ορίζοντες (24h, 72h, 168h). Συγκρίναμε την ακρίβεια των μεθόδων πρόβλεψης με τη μέθοδο αναφοράς αλλά και μεταξύ τους ανά ορίζοντα πρόβλεψης, ιεραρχικά επίπεδα και

ιεραρχικές προσεγγίσεις κάνοντας χρήση του στατιστικού δείκτη s-MAPE. Για την καλύτερη εκτίμηση της ακρίβειας των μεθόδων πρόβλεψης, εφαρμόσαμε τη μέθοδο “rolling horizon”.

9.2 Συμπεράσματα

Τα συμπεράσματα, που προέκυψαν στη διπλωματική αυτή εργασία μπορούν να κατηγοριοποιηθούν ανάλογα με τη μέθοδο πρόβλεψης, τον ορίζοντα πρόβλεψης, τις ιεραρχικές προσεγγίσεις και τα επίπεδα της ιεραρχίας.

Γενικά, παρατηρούμε μια υπεροχή της μεθόδου πολλαπλής γραμμικής παλινδρόμησης (MLR) έναντι της μεθόδου πρόβλεψης χρονοσειρών με χρονική συνάθροιση (MAPA) σε όλους τους ορίζοντες πρόβλεψης, σε όλα τα επίπεδα ιεραρχίας και για τις ιεραρχικές προσεγγίσεις Bottom-Up και Top-Down. Αν και παρατηρήθηκαν τα οφέλη της χρονικής συνάθροισης, αναφορικά με τη μέθοδο MAPA, αυτά δεν ήταν αρκετά, προκειμένου να υπερκαλύψει η ακρίβεια της συγκεκριμένης μεθόδου την ακρίβεια της μεθόδου MLR. Και οι δύο εξεταζόμενες μέθοδοι παρουσιάζουν μεγαλύτερη ακρίβεια πρόβλεψης σε σχέση με τη μέθοδο αναφοράς (SES) σε όλα τα επίπεδα της ιεραρχίας και για όλους τους χρονικούς ορίζοντες. Παρατίθενται, ως ενδεικτικά, τα κέρδη 10,8% και 25% των μεθόδων MAPA και MLR αντίστοιχα, σε σχέση με τη SES στην ακρίβεια απρόβλεψης, για όλη την ιεραρχία συνολικά, για την προσέγγιση bottom-up (που αποτελεί και τη βέλτιστη γενικά) και για ορίζοντα πρόβλεψης 24h.

Αναφορικά με τον ορίζοντα πρόβλεψης, παρατηρήθηκε μικρή ως και αμελητέα επίδραση του παράγοντα αυτού στην ακρίβεια των προβλέψεων. Γενικά, όσο αυξάνει ο ορίζοντας πρόβλεψης, αυξάνει και η αβεβαιότητα, οπότε αναμένονται μεγαλύτερα σφάλματα και συνεπώς μικρότερη ακρίβεια στις προβλέψεις. Αυτή η τάση παρατηρήθηκε στη μέθοδο αναφοράς και στη μέθοδο MAPA (η ακρίβεια μειώθηκε κατά 1% καθώς ο ορίζοντας πρόβλεψης αυξήθηκε από 24h σε 168h), ενώ στη μέθοδο MLR η ακρίβεια μειώθηκε ελαφρώς μεταξύ των οριζόντων 24h και 72h, ενώ μειώθηκε ελαφρώς μεταξύ των οριζόντων 72h και 168h, συμπεριφορά που οφείλεται πιθανότητα στην τυχαιότητα των δεδομένων.

Αναφορικά με τις τρεις διαφορετικές προσεγγίσεις, στις περισσότερες των περιπτώσεων η ιεραρχική προσέγγιση Bottom-Up παρουσιάζει τη μεγαλύτερη ακρίβεια πρόβλεψης. Είναι η καταλληλότερη ιεραρχική προσέγγιση για την πρόβλεψη του κάτω επιπέδου της ιεραρχίας ανεξαρτήτως της μεθόδου και του χρονικού ορίζοντα. Από την άλλη πλευρά, η μέθοδος Top-Down παρουσιάζει μειωμένη ακρίβεια πρόβλεψης στην πρόβλεψη του κάτω επιπέδου της ιεραρχίας. Οι δύο τελευταίες παρατηρήσεις ήταν ουσιαστικά αναμενόμενες, καθώς έχει υποστηριχθεί στην βιβλιογραφία η ανωτερότητα της μεθόδου bottom-up στα ενεργειακά ιεραρχικά δομημένα δεδομένα, που παρουσιάζουν ιδιαίτερη διακύμανση λόγω της ποικιλίας των φορτίων. Η ανωτερότητα αυτή έγκειται στο γεγονός ότι η συγκεκριμένη προσέγγιση λαμβάνει υπ’ όψιν της τα επιμέρους χαρακτηριστικά του κάθε στοιχείου του κάτω επιπέδου.

Τέλος, αναφορικά με τα επίπεδα της ιεραρχίας, οι προσεγγίσεις Optimal καθώς και Top-Down κρίνονται εξίσου κατάλληλες με τη μέθοδο Bottom-up, βάσει της ακρίβειας που παρουσιάζουν, για την παραγωγή προβλέψεων για το άνω επίπεδο της ιεραρχίας.

Αντίθετα, για το κάτω επίπεδο, όπως προαναφέρθηκε βέλτιστη μέθοδος θεωρείται η Bottom-Up. Σχετικά με την καλή ακρίβεια που παρουσιάζει η μέθοδος Optimal για τα μεσαία επίπεδα της ιεραρχίας, δεν είμασταν σε θέση να τη διαπιστώσουμε λόγω της ιεραρχικής δομής των δεδομένων προς εξέταση.

9.3 Μελλοντικές Προεκτάσεις

Με βάση όσα παρατηρήθηκαν στο πλαίσιο αυτής της διπλωματικής εργασίας, προτείνονται οι εξής προεκτάσεις για το μέλλον :

- Σύγκριση των δύο μεθόδων προβλεψης έχοντας ως μέθοδο αναφοράς το μοντέλο Holt-Winters. Το συγκεκριμένο μοντέλο, που περιγράφηκε στην ενότητα 4.2.3, έχει χρησιμοποιηθεί στην βιβλιογραφία σε ενεργειακά δεδομένα και έχει δείξει ότι εφαρμόζεται καλά σε ενεργειακά δεδομένα, λόγω της «διπλής» εποχιακότητας που χρησιμοποιεί.
- Σχεδιασμός σε προγραμματιστικό περιβάλλον αυτοματοποιημένου ολοκληρωμένου συστήματος, το οποίο να εφαρμόζει τις μεθοδολογίες πρόβλεψης που εξετάστηκαν και να δέχεται ως είσοδο την ιεραρχική δομή της διάταξης για την οποία επιθυμούμε να παραχθούν προβλέψεις, τις επιμέρους χρονοσειρές καθώς και τον ορίζοντα πρόβλεψης. Έξοδος του συστήματος θα είναι οι τελικές προβλέψεις ανά ιεραρχική μέθοδο.
- Παραγωγή προβλέψεων για περισσότερους και μεγαλύτερους χρονικούς ορίζοντες, προκειμένου να αναδειχθεί η συμπεριφορά των μεθόδων σε μακροπρόθεσμους ορίζοντες συγκριτικά με τους βραχυπρόθεσμους και μεσοπρόθεσμους ορίζοντες που εφαρμόστηκαν σε αυτή την βιβλιογραφία.
- Χρήση αθροιστικού μοντέλου εποχιακότητας ή μοντέλου εποχιακότητας STL, αντί του πολλαπλασιαστικού που χρησιμοποιήθηκε στην εργασία αυτή. Το μοντέλο STL κάνει επίσης χρήση εναλλακτικής αθροιστικής εποχιακότητας κι έχει αναδειχθεί στην βιβλιογραφία η ευστάθεια του.
- Επαλήθευση των αποτελεσμάτων σε μεγαλύτερα σετ δεδομένων και σε μεγαλύτερες ιεραρχίες δεδομένων, προκειμένου να αναπτυχθεί μελέτη για την επιλογή της κατάλληλης ιεραρχικής μεθόδου ανάμεσα στις τρεις ιεραρχικές προσεγγίσεις που εξετάστηκαν.

Κεφάλαιο 10 : Βιβλιογραφία

10.1 Βιβλία-Σημειώσεις

- (1) Πετρόπουλος, Φ., Ασημακόπουλος, Ε. (2011).Επιχειρησιακές Προβλέψεις, εκδόσεις : Συμμετρία, Αθήνα
- (2) Hyndman, R.J., Athanasopoulos, G. (2013).Forecasting: principles and practice, publ. : OTexts
- (3) Makridakis, S. G., Wheelwright, S.C., Hyndman, R.J. (1997). Forecasting: Methods and Applications, Wiley 3rd edition
- (4) Taylor, J.K., Cihon C., Chapman, Hall (2004).Statistical Techniques for Data Analysis, Chapman and Hall/CRC 2nd edition
- (5) Gelman, A., Hill, J. (2007).Data Analysis Using Regression and Multilevel/Hierarchical Models”, Cambridge University Press 1st edition
- (6) Δ. Κουγιουμτζής (2012). Ανάλυση Χρονοσειρών, Σημειώσεις για «Στατιστική και Μοντελοποίηση» του Τμήματος Μαθηματικών ΑΠΘ
- (7) Frees, E.W. (2010) . Regression Modeling with Actuarial and Financial Applications”, Cambridge University Press, 1st edition

10.2 Ερευνητικές δημοσιεύσεις

- (1) Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renewable and Sustainable Energy Reviews 33 (1) 102–109
- (2) Zhao,H., Magoulès, F. (2012). A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews 16 (1), 3586– 3592
- (3) Hyndman, R.J., Koehlerb, A.B., Snydera, R.D., Grosea, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting 18 (3) 439–454
- (4) Suganthi, L., Samuel, A.A. (2011). Energy models for demand forecasting—A review, Renewable and Sustainable Energy Reviews 16 (2) 1223– 1240
- (5) Everette S., Gardner Jr. (2006). Exponential smoothing: The state of the art—Part II, International Journal of Forecasting 22 (4) 637– 666
- (6) Hong, T., Pinson, P., Fan, S. (2013). Global Energy Forecasting Competition 2012, International Journal of Forecasting, 30 (2), 357–363
- (7) Spithourakis, G., Petropoulos, F., Babai, M., Nikolopoulos, K., Assimakopoulos, V. (2011). Improving the Performance of Popular Supply Chain Forecasting Techniques, Supply Chain Forum An International Journal Vol. 12 (1)

- (8) Athanasopoulos, G., Hyndman, R.J., Song, H., Wu, D. (2011). The tourism forecasting competition, *International Journal of Forecasting*, 27 (3), 822-844
- (9) Weiss, A. (1984). Systematic Sampling and Temporal Aggregation in Time Series Models, *Journal of Econometrics*, 26 (3), 271-281, North-Holland
- (10) Athanasopoulos, G., Ahmed, R., Hyndman, R.J. (2009). Hierarchical forecasts for Australian domestic tourism, *International Journal of Forecasting*, 25 (1), 146–166
- (11) Kourentzes, N., Petropoulos, F., Trapero, J. (2013). Improving forecasting by estimating time series structural components across multiple frequencies, *International Journal of Forecasting*, 30 (2), 291–302
- (12) Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G., Shang, H.L. (2011). Optimal combination forecasts for hierarchical time series, *Computational Statistics and Data Analysis* 55 (1), 2579–2589
- (13) ZiedBabai, M., Ali, M., Nikolopoulos, K. (2012). Impact of temporal aggregation on stock control performance of intermittent demand estimators : Empirical analysis, *Omega*, 40 (6), 713-721
- (14) Andrawis, R., Atiya, A., El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting, *International Journal of Forecasting*, 27 (3), 870–886
- (15) Viswanathant, S., Widiarta, H. (2008). Forecasting aggregate time series with intermittent subaggregate components: top-down versus bottom-up forecasting, *IMA Journal of Management Mathematics*, 19 (1), 275–287
- (16) Fliedner, E., Lawrence, B., (1995). Forecasting system parent group formation : An empirical application of cluster analysis, *Journal of Operations Management*, 12 (2), 119-130
- (17) Tiao, G.C., Guttman, I., (1980). Forecasting Contemporaneous Aggregates of Multiple Time Series, *Journal of Econometrics*, 12 (2), 219-230, North-Holland Publishing Company
- (18) Gross, C.W., Sohl, J.E. (1995). Disaggregation Methods to Expedite Product Line Forecasting, *Journal of Forecasting*, 29 (6), 233-254
- (19) Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation, *Computers & Operations Research*, 26 (10-11)
- (20) Zotteri, G., Kalchschmidt, M., Caniato, F. (2005). The impact of aggregation level on forecasting performance, *Int. J. Production Economics* , 93–94, 479–491
- (21) Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past, *Journal of Econometrics*, 18 (6), 337-349. North-Holland Publishing Company

- (22) Silvestrini, A., Veredas, D. (2008). Temporal Aggregation of Univariate and Multivariate Time Series Models : A Survey, *Journal of Economic Surveys*, 22 (3), 458–497
- (23) Mohammadipour, M., Boylan, J. (2011). Forecast horizon aggregation in integer autoregressive moving average (INARMA) models, *Omega* 40 (6), 703–712
- (24) Tashman, L.J. (2000). Out-of-sample tests of forecasting accuracy : an analysis and review, *International Journal of Forecasting*, 16 (4), 437-450
- (25) R.J. Hyndman, (2006) Fpp: An R package for Data for Forecasting: principles and practice, *International Journal of Forecasting*, 22, (3), 443-473
- (26) R.J. Hyndman, Forecast : An R package for “Forecasting functions for time series and linear models, *Computational Statistics & Data Analysis*, 9
- (27) L.Lombard, J. Ortiz, C. Pout (2008), A review on buildings energy consumption information, *Energy and Buildings*, 40 (3), 394-398
- (28) Aranda, A., Ferreira, G., Mainar-Toledo, M., Scarpellini, S. & Sastresa, E. L. (2012)., Multiple regression models to predict the annual energy consumption in the Spanish banking sector, *Energy and Buildings*, 49, . 380 – 387.
- (29) Armstrong, J.S. (2014). Combining forecasts, *International Journal of Forecasting*, 30 (1), Boston, MA: Kluwer Academic Publishing, 43-54
- (30) Zhao, X., Magoules, F. (2012), A review on the prediction of building energy consumption, *Renewable and Sustainable Energy Reviews*, 16 (6), 3586-3592

ΠΑΡΑΡΤΗΜΑ

(I) Σφάλματα s-MAPE μεθόδου ΜΑΡΑ

| Δείκτης sMAPE (%) για τις 3 Ιεραρχικές Προσεγγίσεις (Ορίζοντας Πρόβλεψης : 24h) | | | |
|---|-------------------|------------------|-----------------|
| Ιεραρχική Μέθοδος Στοιχείο Ιεραρχίας | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
| 1 | 13,72 | 13,75 | 13,74 |
| 2 | 19,52 | 20,21 | 19,71 |
| 3 | 10,92 | 13,30 | 10,93 |
| 4 | 10,89 | 13,33 | 10,90 |
| 5 | 9,38 | 24,73 | 38,24 |
| 6 | 20,75 | 28,74 | 21,10 |
| 7 | 11,25 | 13,83 | 11,26 |
| 8 | 10,89 | 13,33 | 10,90 |
| 9 | 15,74 | 24,07 | 16,67 |
| 10 | 25,21 | 33,90 | 25,25 |
| 11 | 13,42 | 86,17 | 13,42 |
| 12 | 17,43 | 19,50 | 17,43 |
| 13 | 20,67 | 24,33 | 20,68 |
| 14 | 12,12 | 19,97 | 12,27 |
| 15 | 22,53 | 24,37 | 22,67 |
| 16 | 19,52 | 18,60 | 19,56 |
| 17 | 21,20 | 22,81 | 21,30 |
| 18 | 14,46 | 13,53 | 14,53 |
| 19 | 19,47 | 20,43 | 19,48 |
| 20 | 20,52 | 20,23 | 20,55 |
| 21 | 11,75 | 12,18 | 11,77 |
| Μέσος Όρος | 16,25 | 22,92 | 19,71 |

Πίνακας (iv) : Σφάλματα sMAPE μεθόδου ΜΑΡΑ για ορίζοντα πρόβλεψης : 24 h

| Δείκτης sMAPE (%) για τις 3 Ιεραρχικές Προσεγγίσεις (Ορίζοντας Πρόβλεψης : 72 h) | | | |
|--|-------------------|------------------|-----------------|
| Ιεραρχική Μέθοδος Στοιχείο Ιεραρχίας | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
| 1 | 14,74 | 14,77 | 14,77 |
| 2 | 20,01 | 20,66 | 20,18 |
| 3 | 11,65 | 13,73 | 11,66 |
| 4 | 11,62 | 13,76 | 11,63 |
| 5 | 9,93 | 24,84 | 39,17 |
| 6 | 22,25 | 29,31 | 22,59 |
| 7 | 12,01 | 14,28 | 12,02 |
| 8 | 11,62 | 13,76 | 11,63 |
| 9 | 16,86 | 24,58 | 17,67 |
| 10 | 24,97 | 33,99 | 25,02 |
| 11 | 14,24 | 85,95 | 14,26 |
| 12 | 18,49 | 20,09 | 18,51 |
| 13 | 21,88 | 24,71 | 21,90 |
| 14 | 13,34 | 20,356 | 13,47 |
| 15 | 23,55 | 24,70 | 23,69 |

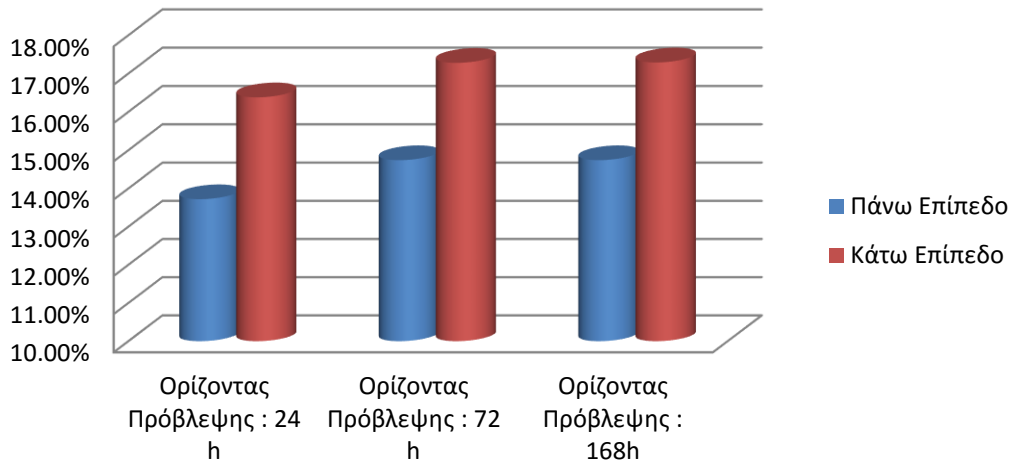
| | | | |
|-------------------|--------------|--------------|--------------|
| 16 | 21,03 | 19,64 | 21,08 |
| 17 | 22,15 | 23,22 | 22,26 |
| 18 | 15,10 | 14,40 | 15,17 |
| 19 | 20,38 | 21,23 | 20,40 |
| 20 | 21,53 | 21,03 | 21,57 |
| 21 | 13,05 | 13,24 | 13,06 |
| Μέσος Όρος | 17,16 | 23,44 | 18,65 |

Πίνακας (v) : Σφάλματα sMAPE μεθόδου ΜΑΡΑ για ορίζοντα πρόβλεψης : 72

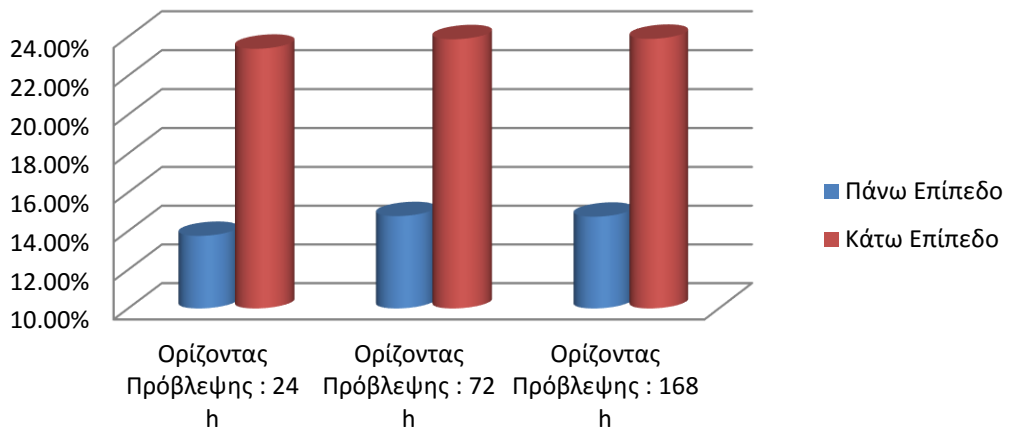
| Δείκτης sMAPE (%) για τις 3 Ιεραρχικές Προσεγγίσεις (Ορίζοντας Πρόβλεψης : 168h) | | | |
|--|-------------------|------------------|-----------------|
| Ιεραρχική Μέθοδος Στοιχείο Ιεραρχίας | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
| 1 | 14,74 | 14,77 | 14,77 |
| 2 | 20,13 | 20,57 | 20,28 |
| 3 | 11,51 | 14,37 | 11,53 |
| 4 | 11,50 | 14,40 | 11,51 |
| 5 | 9,96 | 25,44 | 37,21 |
| 6 | 22,23 | 29,06 | 22,57 |
| 7 | 11,89 | 14,87 | 11,90 |
| 8 | 11,50 | 14,40 | 11,51 |
| 9 | 17,15 | 24,52 | 17,84 |
| 10 | 24,37 | 33,58 | 24,42 |
| 11 | 14,00 | 84,64 | 14,04 |
| 12 | 18,04 | 19,61 | 18,06 |
| 13 | 21,14 | 24,05 | 21,16 |
| 14 | 13,46 | 20,80 | 13,60 |
| 15 | 23,94 | 24,53 | 24,08 |
| 16 | 21,81 | 19,76 | 21,85 |
| 17 | 22,15 | 22,65 | 22,27 |
| 18 | 15,12 | 14,35 | 15,20 |
| 19 | 20,58 | 21,18 | 20,59 |
| 20 | 21,83 | 21,00 | 21,87 |
| 21 | 13,39 | 14,09 | 13,40 |
| Μέσος Όρος | 17,16 | 23,46 | 18,56 |

Πίνακας (vi) : Σφάλματα sMAPE μεθόδου ΜΑΡΑ για ορίζοντα πρόβλεψης : 168 h

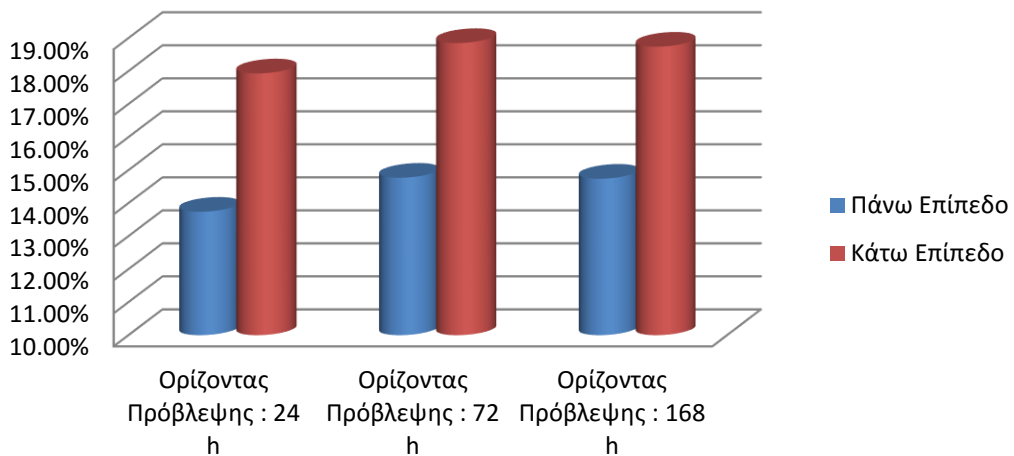
Σφάλμα s-MAPE μεθόδου MAPA Bottom-Up ανά χρονικό ορίζοντα



Μέσο σφάλμα s-MAPE μεθόδου MAPA Top-Down ανά χρονικό ορίζοντα



Μέσο σφάλμα s-MAPE μεθόδου MAPA Optimal ανά χρονικό ορίζοντα



(II) Σφάλματα s-MAPE μεθόδου MLR

| Δείκτης sMAPE (%) για τις 3 Ιεραρχικές Προσεγγίσεις (Ορίζοντας Πρόβλεψης : 24h) | | | |
|---|-------------------|------------------|-----------------|
| Ιεραρχική Μέθοδος Στοιχείο Ιεραρχίας | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
| 1 | 6,82 | 6,82 | 6,81 |
| 2 | 11,87 | 12,32 | 12,95 |
| 3 | 7,28 | 8,21 | 7,18 |
| 4 | 7,27 | 8,24 | 7,17 |
| 5 | 13,66 | 22,49 | 169,71 |
| 6 | 13,86 | 20,98 | 26,24 |
| 7 | 7,34 | 8,40 | 7,25 |
| 8 | 7,27 | 8,24 | 7,17 |
| 9 | 13,14 | 17,66 | 48,53 |
| 10 | 25,414 | 37,20 | 25,72 |
| 11 | 60,84 | 88,46 | 62,41 |
| 12 | 7,97 | 13,68 | 7,71 |
| 13 | 24,59 | 16,59 | 23,884 |
| 14 | 10,68 | 16,80 | 12,49 |
| 15 | 12,55 | 17,08 | 16,86 |
| 16 | 8,89 | 10,17 | 9,11 |
| 17 | 11,12 | 13,80 | 13,57 |
| 18 | 7,55 | 9,00 | 10,03 |
| 19 | 9,47 | 11,34 | 9,41 |
| 20 | 9,78 | 12,24 | 10,08 |
| 21 | 9,05 | 9,02 | 9,05 |
| Μέσος Όρος | 13,64 | 17,56 | 23,97 |

Πίνακας (i) : Σφάλματα sMAPE μεθόδου MLR για ορίζοντα πρόβλεψης : 24 h

| Δείκτης sMAPE (%) για τις 3 Ιεραρχικές Προσεγγίσεις (Ορίζοντας Πρόβλεψης : 72h) | | | |
|---|-------------------|------------------|-----------------|
| Ιεραρχική Μέθοδος Στοιχείο Ιεραρχίας | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
| 1 | 6,85 | 6,87 | 6,86 |
| 2 | 11,29 | 12,22 | 12,13 |
| 3 | 7,06 | 7,82 | 6,91 |
| 4 | 7,04 | 7,85 | 6,91 |
| 5 | 13,92 | 22,69 | 170,93 |
| 6 | 14,03 | 21,27 | 26,41 |
| 7 | 7,14 | 8,04 | 7,00 |
| 8 | 7,04 | 7,85 | 6,91 |
| 9 | 13,30 | 17,68 | 48,83 |
| 10 | 25,632 | 37,02 | 25,80 |
| 11 | 60,51 | 88,02 | 62,23 |
| 12 | 7,81 | 13,25 | 7,59 |
| 13 | 23,68 | 15,84 | 22,99 |
| 14 | 10,93 | 16,87 | 12,53 |
| 15 | 12,16 | 16,47 | 16,41 |
| 16 | 8,81 | 10,17 | 9,02 |

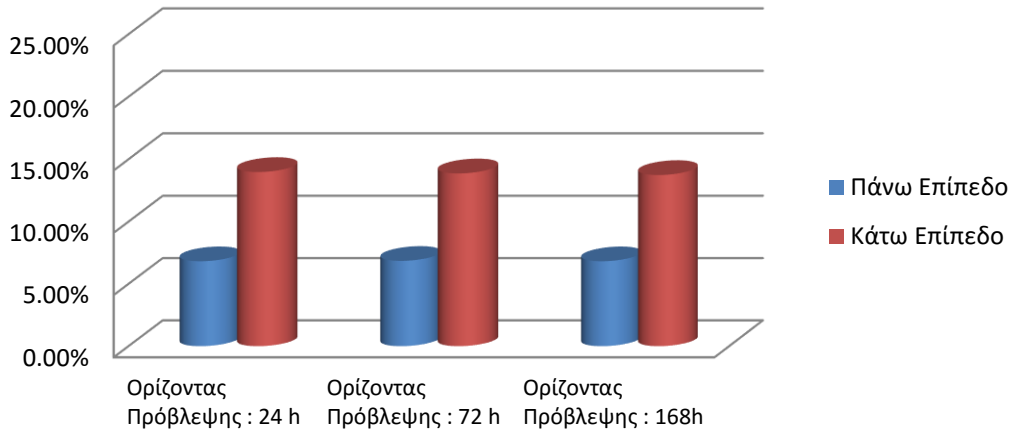
| | | | |
|-------------------|--------------|--------------|--------------|
| 17 | 11,02 | 13,48 | 13,60 |
| 18 | 7,51 | 8,90 | 9,99 |
| 19 | 9,58 | 11,44 | 9,52 |
| 20 | 9,80 | 12,09 | 10,073 |
| 21 | 9,21 | 9,07 | 9,14 |
| Μέσος Όρος | 13,54 | 17,38 | 23,89 |

Πίνακας (ii) : Σφάλματα sMAPE μεθόδου MLR για ορίζοντα πρόβλεψης : 72h

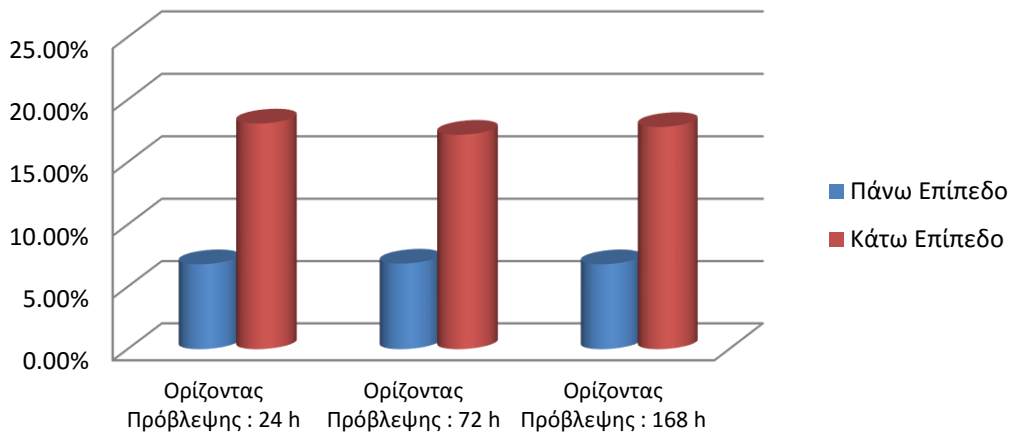
| Δείκτης sMAPE (%) για τις 3 Ιεραρχικές Προσεγγίσεις (Ορίζοντας Πρόβλεψης : 168 h) | | | |
|---|-------------------|------------------|-----------------|
| Ιεραρχική Μέθοδος Στοιχείο Ιεραρχίας | Μέθοδος Bottom-Up | Μέθοδος Top-Down | Μέθοδος Optimal |
| 1 | 6,96 | 6,92 | 6,92 |
| 2 | 10,71 | 12,27 | 11,32 |
| 3 | 7,23 | 7,90 | 7,03 |
| 4 | 7,22 | 7,93 | 7,02 |
| 5 | 13,87 | 22,45 | 169,86 |
| 6 | 13,56 | 20,95 | 26,63 |
| 7 | 7,28 | 8,11 | 7,08 |
| 8 | 7,22 | 7,93 | 7,02 |
| 9 | 12,81 | 17,10 | 50,08 |
| 10 | 25,08 | 36,19 | 25,22 |
| 11 | 59,99 | 87,37 | 61,91 |
| 12 | 7,89 | 13,38 | 7,72 |
| 13 | 23,17 | 16,03 | 22,53 |
| 14 | 10,69 | 16,52 | 12,09 |
| 15 | 12,28 | 16,66 | 16,57 |
| 16 | 8,60 | 10,02 | 8,73 |
| 17 | 10,88 | 13,39 | 13,56 |
| 18 | 7,60 | 9,02 | 9,96 |
| 19 | 9,51 | 11,69 | 9,43 |
| 20 | 9,55 | 12,25 | 9,77 |
| 21 | 9,56 | 9,24 | 9,33 |
| Μέσος Όρος | 13,41 | 17,30 | 23,80 |

Πίνακας (iii) : Σφάλματα sMAPE μεθόδου MLR για ορίζοντα πρόβλεψης : 168 h

Σφάλμα s-MAPE Bottom-Up προβλέψεων των επιπέδων βάσει μεθόδου MLR ανά χρονικό ορίζοντα



Σφάλμα s-MAPE Top-Down προβλέψεων των επιπέδων βάσει μεθόδου MLR ανά χρονικό ορίζοντα



Σφάλμα s-MAPE Optimal προβλέψεων των επιπέδων βάσει μεθόδου MLR ανά χρονικό ορίζοντα

