



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ**

**Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων
Αποφάσεων**

**ΠΡΟΒΛΕΨΗ ΗΛΕΚΤΡΙΚΗΣ ΚΑΤΑΝΑΛΩΣΗΣ ΣΕ
ΣΥΜΠΛΕΓΜΑ ΕΜΠΟΡΙΚΩΝ ΚΤΙΡΙΩΝ ΜΕ
ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΧΡΟΝΙΚΗΣ ΣΥΝΑΘΡΟΙΣΗΣ
ΔΕΔΟΜΕΝΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΕΩΡΓΙΟΣ ΝΙΚ. ΦΑΚΙΝΟΣ

Επιβλέπων: Βασίλειος Ασημακόπουλος

Καθηγητής, ΕΜΠ

Υπεύθυνος: Ευάγγελος Ν. Σπιλιώτης

Αθήνα, Ιούλιος 2016



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων
Αποφάσεων

ΠΡΟΒΛΕΨΗ ΗΛΕΚΤΡΙΚΗΣ ΚΑΤΑΝΑΛΩΣΗΣ ΣΕ ΣΥΜΠΛΕΓΜΑ ΕΜΠΟΡΙΚΩΝ ΚΤΙΡΙΩΝ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΧΡΟΝΙΚΗΣ ΣΥΝΑΘΡΟΙΣΗΣ ΔΕΔΟΜΕΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΕΩΡΓΙΟΣ ΝΙΚ. ΦΑΚΙΝΟΣ

Επιβλέπων: Βασίλειος Ασημακόπουλος

Καθηγητής, ΕΜΠ

Υπεύθυνος: Ευάγγελος Ν. Σπιλιώτης

Εγκρίθηκε απο την 3^η μελή επιτροπή την Ιουλίου 2016

Αθήνα, Ιούλιος 2016

.....

ΓΕΩΡΓΙΟΣ ΝΙΚ. ΦΑΚΙΝΟΣ

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

Copyright © Γεώργιος Νικ. Φακίνος, 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 τους συγγραφείς.

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

ΠΡΟΛΟΓΟΣ

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής κατά το ακαδημαϊκό έτος 2015-2016. Η μονάδα υπάγεται στον Τομέα Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων της Σχολής Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ, του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Βασίλη Ασημακόπουλο για την ευκαιρία που μου έδωσε να ασχοληθώ με το αντικείμενο των προβλέψεων ενεργειακής κατανάλωσης σε ένα σύμπλεγμα κτιρίων τόσο σε επίπεδο μαθήματος όσο και σε επίπεδο διπλωματικής εργασίας. Ακόμα, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Ιωάννη Ψαρρά και τον Επίκουρο Καθηγητή κ. Δημήτριο Ασκούνη για την τιμή που μας έκαναν να συμμετάσχουν στην επιτροπή εξέτασης της εργασίας.

Τέλος θα ήθελα να ευχαριστήσω τον Ευάγγελο Ν. Σπηλιώτη για την βοήθεια και την καθοδήγηση του σε όλη τη διάρκεια εκτέλεσης της εργασίας. Η συνεχή βοήθειά του, οι διορθώσεις και οι συμβουλές του ήταν καθοριστικής σημασίας για την περάτωση της μελέτης αυτής.

Γεώργιος Νικ. Φακίνος

Αθήνα, Ιούλιος 2016

ΠΕΡΙΛΗΨΗ

Στην παρούσα εργασία εξετάζεται η πρόβλεψη ηλεκτρικής κατανάλωσης κτιρίων με χρήση μεθόδων χρονικής συνάθροισης (*temporal aggregation*), της μετάπτωσης δηλαδή των αρχικών δεδομένων σε πολλαπλές χαμηλότερες συχνότητες και την παραγωγή προβλέψεων σε κάθε μία εξ αυτών ξεχωριστά. Σκοπός είναι να εξακριβωθεί σε ποιο βαθμό οι εν λόγω μέθοδοι μπορούν να βοηθήσουν στην επίτευξη ακριβέστερων και πιο αμερόληπτων προβλέψεων εκμεταλλευόμενοι την επιπλέον πληροφορία που κρύβεται σε κάθε συχνότητα που μελετάται.

Η μελέτη γίνεται σε σύμπλεγμα κτιρίων μιας τράπεζας και λόγω του ότι για το σύνολο των κτιρίων διαθέτουμε δεδομένα τόσο ανά ενεργειακή χρήση όσο και για το σύνολο των επιμέρους κτιρίων της τράπεζας, απαιτείται η συμφιλίωση των επιμέρους προβλέψεων που θα παραχθούν προκειμένου οι προβλέψεις στα χαμηλότερα επίπεδα της ιεραρχίας να αθροίζονται σε αυτές των υψηλότερων. Απαιτείται δηλαδή η χρήση μεθόδων ιεραρχικής συνάθροισης δεδομένων (*hierarchical aggregation*).

Στόχος λοιπόν της εργασίας αυτής είναι να καλύψει το κενό που υπάρχει στην βιβλιογραφία όσον αφορά την παραγωγή προβλέψεων με τον συνδυασμό μεθόδων χρονικής και ιεραρχικής συνάθροισης. Με ποιους τρόπους μπορεί να γίνει αυτό αλλά και ποιες θα είναι οι επιπτώσεις στα χαρακτηριστικά των προβλέψεων είναι κάποια από τα ερωτήματα που θα απαντηθούν στην εργασία αυτή.

Εισαγωγικά, παρουσιάζεται η έννοια της ενεργειακής κατανάλωσης, ποια η ανάγκη και τα βήματα που έχουν γίνει για την μελέτη και την μείωση αυτής καθώς και την σημασία που έχει για τις σημερινές κοινωνίες. Στην συνέχεια αναλύουμε τις μεθόδους προβλέψεων, τα μοντέλα χρονοσειρών καθώς και τα χαρακτηριστικά τους αλλά και την ανάλυση που γίνεται προτού ακολουθήσει η διαδικασία της πρόβλεψης. Περιγράφεται, με παράλληλη αναφορά σε βιβλιογραφικά στοιχεία, τις μεθόδους συνάθροισης που θα χρησιμοποιηθούν καθώς και τους δείκτες σφαλμάτων που θα υπολογιστούν για την αξιολόγηση των μεθόδων.

Τέλος περιγράφεται αναλυτικά η μεθοδολογία που θα χρησιμοποιηθεί στη μελέτη και αναλύεται περιληπτικά το προγραμματιστικό περιβάλλον *RStudio*, το οποίο χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων. Προκειμένου να ελέγξουμε την επίδραση των μεθόδων στην αποτελεσματικότητα και την ακρίβεια των προβλέψεων θα γίνει ένα σενάριο μελέτης σε πέντε κτίρια μιας τράπεζας.

Λέξεις Κλειδιά: Τεχνικές Προβλέψεων, Χρονική Συνάθροιση Χρονοσειρών, Ιεραρχικές Προβλέψεις, Ενεργειακές Προβλέψεις,

ABSTRACT

This paper examines the prediction electricity consumption buildings using temporal aggregation methods, i.e. the transition of the original data into multiple lower frequencies produce estimates in each of them separately. The aim is to ascertain to what extent these methods can help to achieve more accurate and unbiased estimations, taking advantage of the extra information that is hidden in each frequency studied.

The study was done in a bank building complex and because the data we have are organized per energy use as well as per individual buildings of the bank the reconciliation of individual predictions is required to produce predictions at the lower levels of hierarchy summed in these higher. That requires the use of hierarchical data aggregation methods (hierarchical aggregation).

So the aim of this is to fill the gap that exists in the literature regarding production forecasts by combining temporal and hierarchical aggregation methods. In what ways can this be done and what will be the impact on the forecasts features are some of the questions that will be answered in this paper.

What does energy consumption mean, what is the need and the steps taken for the study and this reduction and its importance for today's societies. Then we analyze the forecast methods, time series models and their characteristics and the analysis is done before the forecasting process. Aggregation methods and error indicators, that are used for the evaluation of methods, are then in depth analyzed.

Finally we describe the methodology used in the study and briefly analyze the programming environment RStudio, which was used to export the results. To test the effect of the methods concerning the efficiency and accuracy of the forecast, a case study in five buildings of a bank is executed.

Keywords: Forecasting Methods, Temporal Aggregation, Hierarchical Aggregation, Energy Forecast,

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΡΟΛΟΓΟΣ	5
ΠΕΡΙΛΗΨΗ	6
ABSTRACT	7
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	9
ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ	11
1.1 ΔΟΜΗ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ	13
1.2 ΕΝΕΡΓΕΙΑΚΗ ΚΑΤΑΝΑΛΩΣΗ	14
1.3 ΠΡΟΒΛΕΨΗ ΕΝΕΡΓΕΙΑΚΗΣ ΚΑΤΑΝΑΛΩΣΗΣ.....	19
1.4 ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ ΕΝΕΡΓΕΙΑΚΗΣ ΚΑΤΑΝΑΛΩΣΗΣ	19
1.4.1 ΜΗΧΑΝΙΚΕΣ ΜΕΘΟΔΟΙ (<i>Engineering Methods</i>)	20
1.4.2 ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ (<i>Statistical Methods</i>).....	21
1.4.3 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (ANNs)/ ΜΗΧΑΝΗΜΑΤΑ ΥΠΟΣΤΗΡΙΞΗΣ ΔΙΑΝΥΣΜΑΤΩΝ (SVMs)	22
1.4.4 ΜΟΝΤΕΛΑ GREY	23
ΚΕΦΑΛΑΙΟ 2: ΜΟΝΤΕΛΑ ΧΡΟΝΟΣΕΙΡΩΝ	25
2.1 ΜΟΝΤΕΛΟ ΧΡΟΝΟΣΕΙΡΩΝ	25
2.2 ΒΑΣΙΚΑ ΒΗΜΑΤΑ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΠΡΟΒΛΕΨΗΣ	25
2.3 ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΕΩΝ	26
2.3.1 ΝΑΪΒΕ Η ΑΠΛΟΪΚΗ ΜΕΘΟΔΟΣ	26
2.3.2 ΜΕΘΟΔΟΙ ΕΚΘΕΤΙΚΗΣ ΕΞΟΜΑΛΥΝΣΗΣ.....	26
2.3.3 ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (<i>Regression Models</i>).....	30
2.3.4 ΜΕΘΟΔΟΣ ΘΗΤΑ	33
2.3.5 ΜΟΝΤΕΛΑ ARIMA (<i>Autoregressive Integrated Moving Average</i>)	34
ΚΕΦΑΛΑΙΟ 3: ΜΕΘΟΔΟΙ ΣΥΝΑΘΡΟΙΣΗΣ ΔΕΔΟΜΕΝΩΝ (AGGREGATION)	37
3.1 ΜΕΘΟΔΟΣ ΣΥΝΑΘΡΟΙΣΗΣ	37
3.2 CROSS-SECTIONAL/ CONTEMPORANEOUS AGGREGATION	38
3.3 ΧΡΟΝΙΚΗ ΣΥΝΑΘΡΟΙΣΗ (<i>TEMPORAL AGGREGATION OR FLOW SCHEME</i>)	39
3.3.1 ΚΛΑΣΣΙΚΗ ΧΡΟΝΙΚΗ ΣΥΝΑΘΡΟΙΣΗ ΜΗ ΕΠΙΚΑΛΥΠΤΟΜΕΝΗ (<i>CLASSIC TEMPORAL AGGREGATION NONOVERLAPPING</i>)	40
3.3.2 ΚΛΑΣΣΙΚΗ ΧΡΟΝΙΚΗ ΣΥΝΑΘΡΟΙΣΗ ΕΠΙΚΑΛΥΠΤΟΜΕΝΗ (<i>CLASSIC TEMPORAL AGGREGATION OVERLAPPING</i>)	41
3.3.3 ΜΑΡΑ (<i>MULTI AGGREGATION PREDICTION ALGORITHM</i>)	41
ΚΕΦΑΛΑΙΟ 4: ΑΚΡΙΒΕΙΑ ΤΩΝ ΠΡΟΒΛΕΨΕΩΝ	45
4.1 ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ	45
4.2 ΣΤΑΤΙΣΤΙΚΟΙ ΔΕΙΚΤΕΣ ΣΦΑΛΜΑΤΩΝ	45
4.2.1 ΜΕΣΟ ΣΦΑΛΜΑ (<i>Mean Error</i>)	45
4.2.2 ΜΕΣΟ ΑΠΟΛΥΤΟ ΣΦΑΛΜΑ (<i>Mean Absolute Error</i>)	46
4.2.3 ΜΕΣΟ ΤΕΤΡΑΓΩΝΙΚΟ ΣΦΑΛΜΑ (<i>Mean Squared Error</i>).....	46
4.2.4 ΡΙΖΑ ΜΕΣΟΥ ΤΕΤΡΑΓΩΝΙΚΟΥ ΣΦΑΛΜΑΤΟΣ (<i>Root Mean Squared Error</i>).....	47
4.2.5 ΜΕΣΟ ΑΠΟΛΥΤΟ ΠΟΣΟΣΤΙΑΙΟ ΣΦΑΛΜΑ (<i>Mean Absolute Percentage Error</i>)	47
4.2.6 ΣΥΜΜΕΤΡΙΚΟ ΜΕΣΟ ΑΠΟΛΥΤΟ ΠΟΣΟΣΤΙΑΙΟ ΣΦΑΛΜΑ (<i>Symmetric Mean Absolute Percentage Error</i>)	48
4.2.7 ΜΕΣΟ ΑΠΟΛΥΤΟ ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΟ ΣΦΑΛΜΑ (<i>Mean Absolute Scaled Error</i>)	49
ΚΕΦΑΛΑΙΟ 5: ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ	53

5.1 ΜΕΘΟΔΟΛΟΓΙΑ	53
5.1.1 ΕΙΣΑΓΩΓΗ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ	54
5.1.2 ΔΗΜΙΟΥΡΓΙΑ ΠΙΝΑΚΩΝ ΚΑΙ ΜΕΤΑΤΡΟΠΗ ΔΕΔΟΜΕΝΩΝ	55
5.1.3 ΑΠΟΣΥΝΘΕΣΗ ΧΡΟΝΟΣΕΙΡΑΣ	56
5.1.4 ΕΠΙΛΟΓΗ ΜΕΘΟΔΟΥ ΚΑΙ ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ	57
5.1.5 ΠΑΡΑΓΩΓΗ ΤΩΝ ΤΕΛΙΚΩΝ ΠΡΟΒΛΕΨΕΩΝ	59
5.1.6 ΣΥΜΠΛΗΡΩΣΗ ΠΙΝΑΚΩΝ	61
5.2 ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΜΕΘΟΔΟΛΟΓΙΚΗΣ ΠΡΟΣΕΓΓΙΣΗΣ	62
ΚΕΦΑΛΑΙΟ 6: CASE-STUDY ΣΕ ΣΥΜΠΛΕΓΜΑ ΚΤΗΡΙΩΝ ΤΡΑΠΕΖΑΣ.....	65
6.1 ΣΥΛΛΟΓΗ ΚΑΙ ΑΝΑΛΥΣΗ ΣΤΟΙΧΕΙΩΝ	65
6.2 ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ	73
6.3 ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ (<i>RStudio</i>).....	84
ΚΕΦΑΛΑΙΟ 7: ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΠΡΟΕΚΤΑΣΕΙΣ	89
7.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΕΡΓΑΣΙΑΣ.....	89
7.2 ΠΡΟΕΚΤΑΣΕΙΣ	90
ΒΙΒΛΙΟΓΡΑΦΙΑ	93

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Παγκόσμια Κατανάλωση Ενέργειας	14
Εικόνα 2 Κατανομή Ενεργειακής Κατανάλωσης	15
Εικόνα 3 Γρήγορα Μεταβαλλόμενη Ζήτηση	42
Εικόνα 4 Αργά Μεταβαλλόμενη Ζήτηση	42
Εικόνα 5 Μεθοδολογία MAPA	43
Εικόνα 6 Χαρακτηριστικά Χρονοσειράς	56
Εικόνα 7 Mapasimple	57
Εικόνα 8 Μεθοδολογική Προσέγγιση	62
Εικόνα 9 Διάρθρωση Δεδομένων.....	65
Εικόνα 10 Συνολική Κατανάλωση Κτιρίων (Total).....	66
Εικόνα 11 Συνολική Κατανάλωση 1ου Κτιρίου (Sum1).....	66
Εικόνα 12 Συνολική Κατανάλωση 2ου Κτιρίου (Sum2).....	67
Εικόνα 13 Συνολική Κατανάλωση 3ου Κτιρίου (Sum3).....	67
Εικόνα 14 Συνολική Κατανάλωση 4ου Κτιρίου (Sum4).....	68
Εικόνα 15 Συνολική Κατανάλωση 5ου Κτιρίου (Sum5).....	68
Εικόνα 16 Κατανάλωση ανά τμήμα του 1ου κτιρίου	69
Εικόνα 17 Μέσο Σφάλμα sMAPE,για την SES σε συνδυασμό με την MAPA, για 7 οριζόντες και για κάθε επίπεδο ξεχωριστά.....	76
Εικόνα 18 Μέσο Σφάλμα ME για όλους τους οριζόντες και για όλα τα επίπεδα	77
Εικόνα 19 Μέσο Σφάλμα RMSE για όλους τους οριζόντες και για όλα τα επίπεδα	77
Εικόνα 20 Μέσο Σφάλμα MAE για όλους τους οριζόντες και για όλα τα επίπεδα.....	78
Εικόνα 21 Μέσο Σφάλμα MAsE για όλους τους οριζόντες και για όλα τα επίπεδα	78
Εικόνα 22 sMAPE για SES και MAPA για όλα τα επίπεδα και για όλους τους οριζόντες πρόβλεψης ...	79
Εικόνα 23 sMAPE, MAPA over SES (bottom-up, top-down, optimal)	80
Εικόνα 24 MPE, MAPA over SES (bottom-up, top-down, optimal)	Error! Bookmark not defined.
Εικόνα 25 sMAPE, MAPA over SES (f).....	Error! Bookmark not defined.
Εικόνα 26 MPE, MAPA over SES (f)	81
Εικόνα 27 sMAPE, MAPA over SES (k)	82
Εικόνα 28 MPE, MAPA over SES (k).....	82
Εικόνα 29 Βιβλιοθήκη MAPA	84
Εικόνα 30 Συνάρτηση mapasimple	85

Εικόνα 31 Βιβλιοθήκη forecast	85
Εικόνα 32 Συνάρτηση accuracy	86
Εικόνα 33 Δείκτες Σφαλμάτων	86
Εικόνα 34 Περιβάλλον R-Studio	87

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 ΔΟΜΗ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ

Στο πρώτο κεφάλαιο γίνεται μια περιγραφή της ενεργειακής κατανάλωσης κτιρίων. Ποια η σημασία της για τις σημερινές κοινωνίες, τι ποσοστό της συνολικής κατανάλωσης περιλαμβάνει, πως μπορεί να μειωθεί και ποια βήματα έχουν γίνει μέχρι τώρα. Επίσης γίνεται μια περιληπτική αναφορά στις μέχρι τώρα μεθόδους πρόβλεψης ενεργειακής κατανάλωσης που χρησιμοποιούνται. Ποιες είναι αυτές, από ποιους έχουν χρησιμοποιηθεί και ποια τα αποτελέσματά τους.

Στο δεύτερο κεφάλαιο αναλύονται τα μοντέλα χρονοσειρών καθώς και οι μέθοδοι προβλέψεων. Περιγράφονται τα βασικά βήματα της διαδικασίας των προβλέψεων και οι βασικές μέθοδοι προβλέψεων που χρησιμοποιούνται και είναι γνωστές στην βιβλιογραφία. Η μέθοδος *Naïve*, οι μέθοδοι εκθετικής εξομάλυνσης, τα μοντέλα παλινδρόμησης, τα μοντέλα *ARIMA* καθώς και η μέθοδος *Theta* είναι κάποια από τα βασικά και πιο γνωστά μοντέλα προβλέψεων.

Στο επόμενο κεφάλαιο περιγράφονται οι μέθοδοι συνάθροισης που θα χρησιμοποιηθούν. Αναλύεται βιβλιογραφικά η χρονική μέθοδος συνάθροισης και οι κυριότερες μορφές της καθώς και η ιεραρχική συνάθροιση και με ποιους τρόπους γίνεται. Επίσης γίνεται πλήρης αναφορά και περιγραφή του αλγορίθμου *MAPA* που χρησιμοποιείται στην παρούσα μελέτη για την πραγματοποίηση της χρονικής συνάθροισης των δεδομένων.

Στο τέταρτο κεφάλαιο περιγράφονται οι δείκτες σφαλμάτων που χρησιμοποιούνται για την αξιολόγηση των μεθόδων προβλέψεων. Τι δείχνει ο κάθε δείκτης και με ποιον τρόπο μπορεί να βρεθεί. Για την παρούσα μελέτη δεν είναι αναγκαίοι όλοι οι δείκτες ωστόσο για λόγους πληρότητας αναφέρονται και αναλύονται όλοι στο κεφάλαιο αυτό.

Στα επόμενα δύο κεφάλαια αναλύεται η μεθοδολογία που θα χρησιμοποιηθεί, πραγματοποιείται το σενάριο μελέτης στα κτίρια της τράπεζας και παρουσιάζονται οι πίνακες σφαλμάτων που προκύπτουν. Επίσης περιγράφεται περιληπτικά το πρόγραμμα ανάλυσης δεδομένων που χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων, *RStudio*.

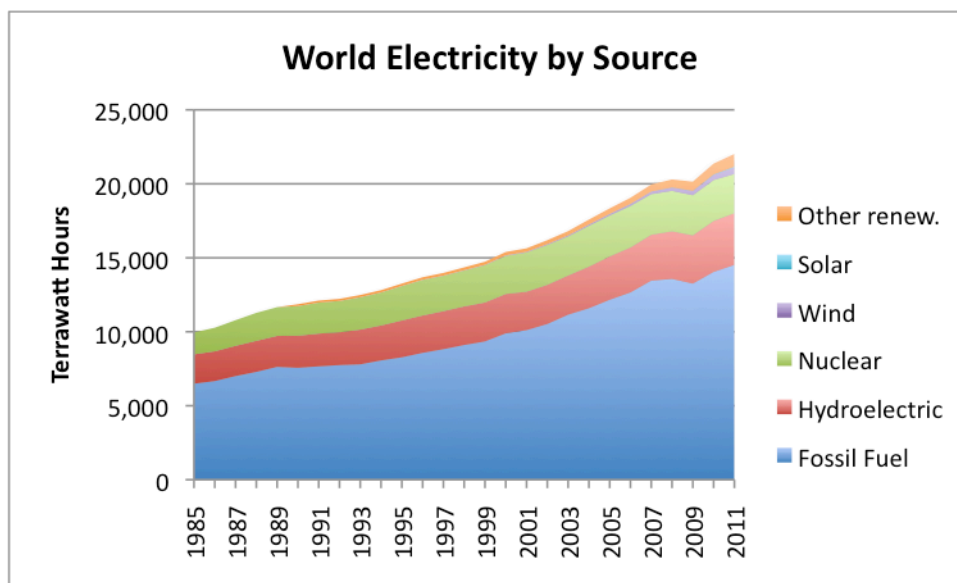
Τέλος στο έβδομο κεφάλαιο παρατίθεται περιληπτικά τα συμπεράσματα που προκύπτουν από την μελέτη και την εξαγωγή των αποτελεσμάτων με τον συνδυασμό των μεθόδων συνάθροισης.

1.2 ΕΝΕΡΓΕΙΑΚΗ ΚΑΤΑΝΑΛΩΣΗ

Ένα θέμα μείζονος σημασίας για τις σημερινές κοινωνίες είναι η εξάντληση των ενεργειακών αποθεμάτων καθώς και το περιβαλλοντικό αντίκτυπο που έχει η συνεχώς αυξανόμενη ενεργειακή κατανάλωση σε παγκόσμιο επίπεδο. Η αύξηση του παγκόσμιου πληθυσμού, των αναγκών του, των προϊόντων και των υπηρεσιών έχει εκτοξεύσει την ενεργειακή ζήτηση σε υψηλά επίπεδα και η αύξηση αυτή αναμένεται να συνεχιστεί και στο μέλλον.

Για την εξασφάλιση της αποτελεσματικής λειτουργίας του ενεργειακού τομέα των χωρών πρωταρχικό μέτρο αποτελεί η ορθή διαχείριση της ενέργειας. Η ορθή διαχείριση της ενέργειας έχει ως σκοπό την αποδοτική χρήση της, χωρίς την μείωση των παραγωγικών μεγεθών μιας χώρας και χωρίς να θυσιάσει η ποιότητα των προϊόντων ή των υπηρεσιών καθώς και την εναρμόνισή της με τις σημερινές περιβαλλοντικές απαιτήσεις.

Τα παρακάτω γραφήματα δείχνουν χαρακτηριστικά την αύξηση της ενεργειακής κατανάλωσης παγκοσμίως καθώς και τους τομείς όπου καταναλώνεται:



Εικόνα 1 Παγκόσμια Κατανάλωση Ενέργειας

Η παραπάνω εικόνα (Εικόνα 1) παρουσιάζει την αύξηση της παγκόσμιας κατανάλωσης της ενέργειας από το 1985 έως το 2011, χωρίζοντας την παράλληλα ανάλογα με το είδος της ενέργειας. Στην παρούσα διπλωματική εργασία μας ενδιαφέρει η αύξηση της ενεργειακής κατανάλωσης και όχι τόσο το είδος της ενέργειας.



Εικόνα 2 Κατανομή Ενεργειακής Κατανάλωσης

Η παραπάνω εικόνα (Εικόνα 2) παρουσιάζει τα ποσοστά κατανάλωσης ενέργειας κατανεμημένα στους τρεις κύριους τομείς της κοινωνίας της Ευρωπαϊκής Ένωσης. Το 32% της συνολικής κατανάλωσης οφείλεται στις μεταφορές, το 25% οφείλεται στην βιομηχανία, το 40% οφείλεται σε κτήρια (σπίτια, καταστήματα κ.α.) και το υπόλοιπο 3% που δεν παρουσιάζεται στο διάγραμμα οφείλεται σε άλλους τομείς όπως την γεωργία και την αλιεία. Η μείωση της κατανάλωσης ενέργειας και η εξάλειψη της σπατάλης ενέργειας αποκτούν διαρκώς αυξανόμενη σημασία στην ΕΕ. Οι ηγέτες της ΕΕ αποφάσισαν το 2007 να θέσουν ως στόχο για το 2020 τη μείωση της ετήσιας κατανάλωσης ενέργειας στην Ένωση κατά 20%. Τα μέτρα ενεργειακής απόδοσης αναγνωρίζονται όλο και περισσότερο ως μέσο όχι μόνο για την επίτευξη βιώσιμου ενεργειακού εφοδιασμού, τη μείωση των εκπομπών αερίων θερμοκηπίου, τη βελτίωση της ασφάλειας του εφοδιασμού και τη μείωση των δαπανών για εισαγωγές, αλλά και για την προαγωγή της ανταγωνιστικότητας των ευρωπαϊκών οικονομιών. Το Ευρωπαϊκό Συμβούλιο της 20ής και 21ης Μαρτίου 2014 τόνισε την αποτελεσματικότητα της ενεργειακής απόδοσης όσον αφορά τη μείωση του κόστους ενέργειας και της ενεργειακής εξάρτησης. Η ΕΕ έχει θέσει ελάχιστα πρότυπα ενεργειακής απόδοσης και κανόνες σχετικά με την επισήμανση και την οικολογική σχεδίαση των προϊόντων, των υπηρεσιών και των υποδομών. Τα μέτρα αυτά αποσκοπούν στη βελτίωση της απόδοσης σε όλα τα στάδια της ενεργειακής αλυσίδας, από τον ενεργειακό εφοδιασμό έως τη χρήση της ενέργειας από τους καταναλωτές.

Είναι λοιπόν απαραίτητη η μελέτη και ο περιορισμός της ενεργειακής κατανάλωσης, όσο είναι δυνατόν, από πλευράς χωρών. Τα οφέλη που μπορούν να προκύψουν από την μείωση της ενεργειακής κατανάλωσης είναι πολλά. Επιγραμματικά αναφέρονται κάποια για λόγους πληρότητας:

- Μείωση του κόστους παραγωγής τελικών προϊόντων
- Μείωση των εκπομπών των αερίων ρύπων και προσαρμογή των βιομηχανιών στις υπό εφαρμογή σχετικές Κοινοτικές Οδηγίες
- Θετική συμβολή στο ενεργειακό ισοζύγιο της κάθε χώρας
- Μείωση της εξάρτησης της χώρας από εισαγόμενα καύσιμα και αντίστοιχη εξοικονόμηση συναλλάγματος

- Επίτευξη των στόχων και των δεσμεύσεων της χώρας, που έχουν τεθεί για τη μείωση των εκπομπών των αερίων του θερμοκηπίου σε διεθνές επίπεδο.

Η ενεργειακή κατανάλωση μετρείται σε kWh και η σημασία της είναι τόσο μεγάλη που αρκετές χώρες έχουν δεσμευθεί (Πρωτόκολλο του Κιότο) να μειώσουν, σε καθορισμένο χρονικό διάστημα, την ενεργειακή τους κατανάλωση και τις εκπομπές. Η Ευρωπαϊκή Ένωση μάλιστα έχει εκδώσει **οδηγία για την ενεργειακή απόδοση** (2012/27/ΕΕ), η οποία τέθηκε σε ισχύ τον Δεκέμβριο του 2012. Σύμφωνα με την οδηγία, τα κράτη μέλη υποχρεούνται να θέσουν ενδεικτικούς εθνικούς στόχους σχετικά με την ενεργειακή απόδοση για το 2020, με βάση την πρωτογενή ή την τελική κατανάλωση ενέργειας. Η οδηγία θέτει επίσης νομικά δεσμευτικούς κανόνες για τους τελικούς χρήστες και τους προμηθευτές ενέργειας. Τα κράτη μέλη είναι ελεύθερα να καθιστούν τις ελάχιστες αυτές απαιτήσεις πιο αυστηρές, στην προσπάθειά τους να εξοικονομηθεί ενέργεια. Η οδηγία περιλαμβάνει, μεταξύ άλλων, τις ακόλουθες απαιτήσεις:

- την ανακαίνιση των κτιρίων που είναι στην ιδιοκτησία των κεντρικών δημόσιων διοικήσεων σε ποσοστό τουλάχιστον 3% του συνολικού εμβαδού δαπέδου, για κάθε έτος από το 2014 και μετά, και την αγορά κτιρίων, υπηρεσιών και προϊόντων με υψηλή ενεργειακή απόδοση, έτσι ώστε ο δημόσιος τομέας να ηγηθεί της προσπάθειας·
- τη χάραξη μακροπρόθεσμων εθνικών στρατηγικών προκειμένου να προωθηθούν οι επενδύσεις για την ανακαίνιση κατοικιών και εμπορικών κτιρίων και τη θέσπιση εθνικών καθεστώτων επιβολής της υποχρέωσης ενεργειακής απόδοσης ή ισοδύναμων μέτρων που να εξασφαλίζουν ετήσια εξοικονόμηση ενέργειας 1,5% για τους τελικούς καταναλωτές·
- την αξιολόγηση, έως το τέλος του 2015, του δυναμικού υλοποίησης της συμπαράγωγής υψηλής απόδοσης και της αποδοτικής τηλεθέρμανσης και τηλεψύξης σε όλα τα κράτη μέλη·
- τη διεξαγωγή υποχρεωτικών τακτικών ενεργειακών ελέγχων στις μεγάλες επιχειρήσεις, τουλάχιστον ανά τέσσερα χρόνια, με εξαίρεση τις επιχειρήσεις που διαθέτουν πιστοποιημένα ενεργειακά και περιβαλλοντικά συστήματα·
- την ανάπτυξη ευφών δικτύων και μετρητών και την παροχή ακριβούς πληροφόρησης στους λογαριασμούς κατανάλωσης ενέργειας, προκειμένου να ενδυναμωθούν οι καταναλωτές και να ενθαρρυνθεί η πιο αποδοτική κατανάλωση ενέργειας.

Το Νοέμβριο του 2013, η Επιτροπή εξέδωσε μια ανακοίνωση και επτά κατευθυντήρια σημειώματα σχετικά με την υλοποίηση των διάφορων πτυχών της οδηγίας για την ενεργειακή απόδοση.

Η οδηγία 2002/91/ΕΚ για την ενεργειακή απόδοση των κτιρίων (ειδικότερα τη μόνωση, τον κλιματισμό και τη χρήση ανανεώσιμων πηγών ενέργειας) παρέχει μια μέθοδο υπολογισμού της ενεργειακής απόδοσης κτιρίων και προβλέπει ελάχιστες απαιτήσεις για τα νέα και τα υφιστάμενα μεγάλα κτίρια, καθώς και ενεργειακή πιστοποίηση. Την 1η Φεβρουαρίου 2012, η οδηγία καταργήθηκε από την οδηγία αναδιατύπωσης 2010/31/ΕΕ, που τέθηκε σε ισχύ τον Ιούλιο του 2010. Κύριος στόχος της οδηγίας αναδιατύπωσης ήταν ο εξορθολογισμός ορισμένων διατάξεων της προηγούμενης οδηγίας και η ενίσχυση των απαιτήσεων ενεργειακής απόδοσης σε σχέση με:

- το κοινό γενικό πλαίσιο για μια μεθοδολογία υπολογισμού της συνολικής ενεργειακής απόδοσης κτιρίων και κτιριακών μονάδων·
- την εφαρμογή ελάχιστων απαιτήσεων για την ενεργειακή απόδοση νέων κτιρίων και νέων κτιριακών μονάδων, ορίζοντας, για παράδειγμα, ότι έως τις 31 Δεκεμβρίου 2020 όλα τα νέα κτίρια πρέπει να έχουν σχεδόν μηδενικό ενεργειακό ισοζύγιο·
- την εφαρμογή ελάχιστων απαιτήσεων για την ενεργειακή απόδοση, ειδικότερα όσον αφορά: υφιστάμενα κτίρια, δομικά στοιχεία στα οποία γίνεται μεγάλη κλίμακας

ανακαίνιση, και τεχνικά συστήματα κτιρίων όταν εγκαθίστανται, αντικαθίστανται ή αναβαθμίζονται·

- την ενεργειακή πιστοποίηση κτιρίων ή κτιριακών μονάδων, την τακτική επιθεώρηση των συστημάτων θέρμανσης και κλιματισμού στα κτίρια, και ανεξάρτητα συστήματα ελέγχου για τα πιστοποιητικά ενεργειακής απόδοσης και τις εκθέσεις ελέγχου.

Η οδηγία αναδιατύπωσης καθορίζει ελάχιστες απαιτήσεις, αλλά τα κράτη μέλη μπορούν να προωθήσουν ή να θεσπίσουν περαιτέρω μέτρα. Ως συνέχεια της οδηγίας αναδιατύπωσης, η Επιτροπή δημοσιοποίησε τον Απρίλιο του 2013 έκθεση για την αξιολόγηση της αποτελεσματικότητας της τρέχουσας οικονομικής υποστήριξης στην ενεργειακή απόδοση των κτιρίων (COM(2013)0225). Η έκθεση αυτή είχε επίσης στόχο να βοηθήσει τα κράτη μέλη να υλοποιήσουν μια απαίτηση που καθορίζεται στην οδηγία για την ενεργειακή απόδοση, σχετικά με τη χάραξη, έως τον Απρίλιο του 2014, μακροπρόθεσμης στρατηγικής για την κινητοποίηση επενδύσεων στην ανακαίνιση των κτιρίων σε εθνικό επίπεδο. Μία άλλη έκθεση που δημοσιεύθηκε το Φεβρουάριο του 2014 παρέιχε τεχνικές οδηγίες σχετικά με τη χρηματοδότηση της ενεργειακής ανακαίνισης των κτιρίων με πόρους της πολιτικής συνοχής.

Στο ψήφισμά του της 15ης Δεκεμβρίου 2010 σχετικά με την αναθεώρηση του σχεδίου δράσης για την ενεργειακή απόδοση (T7-0485/2010), το Ευρωπαϊκό Κοινοβούλιο κατέστησε σαφές ότι θα πρέπει να εγκριθεί ένας δεσμευτικός στόχος για την ενεργειακή απόδοση, της τάξης τουλάχιστον του 20% έως το 2020. Ζητούσε επίσης να γίνει αναθεώρηση της οδηγίας για τις ενεργειακές υπηρεσίες το 2011, η οποία να περιλαμβάνει διευρυμένο χρονικό πλαίσιο έως το 2020 και κριτική αξιολόγηση των εθνικών σχεδίων δράσης για την ενεργειακή απόδοση και της υλοποίησής τους.

Σε προηγούμενο ψήφισμά του, της 6ης Μαΐου 2010, σχετικά με την κινητοποίηση των τεχνολογιών της πληροφορίας και της επικοινωνίας για τη διευκόλυνση της μετάβασης προς ενεργειακά αποδοτική οικονομία με χαμηλές εκπομπές άνθρακα (T7-0153/2010), το Ευρωπαϊκό Κοινοβούλιο τόνιζε ότι απαιτούνται σημαντικές επενδύσεις τόσο για έρευνα και ανάπτυξη όσο και για την αξιοποίηση των υφιστάμενων τεχνολογιών, προκειμένου να εξασφαλισθεί η μετάβαση σε μια ενεργειακά αποδοτική οικονομία με χαμηλές εκπομπές άνθρακα. Πρότεινε να παρέχουν τα κράτη μέλη κίνητρα τόσο για δημόσιες όσο και για ιδιωτικές επενδύσεις στην ενεργειακή απόδοση· σημείο εκκίνησης θα αποτελούσε η ένταξη της ενεργειακής απόδοσης στη σχολική εκπαίδευση. Τόνιζε επίσης ότι οι ενημερωτικές εκστρατείες ευρείας κλίμακας για να γίνουν κατανοητά τα οφέλη των έξυπνων μετρητών και των προγραμμάτων ΤΠΕ έχουν αποφασιστική σημασία για να επιτευχθεί η υποστήριξη του κοινού.

Κατά τη διαδικασία έγκρισης της οδηγίας αναδιατύπωσης για την ενεργειακή απόδοση των κτιρίων, το Ευρωπαϊκό Κοινοβούλιο τάχθηκε υπέρ μιας ισχυρότερης και περισσότερο φιλόδοξης ρύθμισης. Για παράδειγμα, επέμεινε ότι όλα τα κτίρια θα πρέπει να παρουσιάζουν ήδη τουλάχιστον μηδενικό ισοζύγιο ενέργειας έως τις 31 Δεκεμβρίου 2016 (COD/2008/223). Όσον αφορά την αναδιατύπωση της οδηγίας για την επισήμανση της κατανάλωσης ενέργειας (COD/2008/222), το Κοινοβούλιο εξασφάλισε ότι το κείμενο θα αναφέρεται ρητά τόσο σε προϊόντα τα οποία πράγματι καταναλώνουν ενέργεια (και επομένως έχουν άμεσο αντίκτυπο) όσο και σε προϊόντα τα οποία δεν καταναλώνουν τα ίδια ενέργεια, αλλά έχουν έμμεσο αντίκτυπο στην εξοικονόμηση ενέργειας. Επιπλέον, το Κοινοβούλιο ενίσχυσε τον βαθμό στον οποίο πρέπει να περιλαμβάνονται πληροφορίες σχετικά με την ενεργειακή απόδοση σε διαφημίσεις και σε τεχνικό προωθητικό υλικό.

Το 2012, το Κοινοβούλιο διαδραμάτισε κεντρικό ρόλο στη διαπραγμάτευση της οδηγίας για την ενεργειακή απόδοση (COD/2011/0172) και εξασφάλισε τη διατήρηση των απαιτήσεων για τις εθνικές στρατηγικές ανακαίνισης κτιρίων και τις υποχρεωτικές ενεργειακές επιθεωρήσεις για τις μεγάλες επιχειρήσεις στον τελικό συμβιβασμό που συμφωνήθηκε με το Συμβούλιο.

Κατάφερε επίσης να διατηρηθεί μια τροπολογία με την οποία ζητείται η θέσπιση κανόνων για μηχανισμούς αντίδρασης που θα δίνουν στους καταναλωτές ενέργειας τη δυνατότητα να προσαρμόζουν τη χρήση ενέργειας στις συνθήκες εφοδιασμού και με τον τρόπο αυτό να μειώνουν τους λογαριασμούς κατανάλωσης ενέργειας.

Το Κοινοβούλιο ενέκρινε πρόσφατα ψήφισμα σχετικά με την εφαρμογή και τον αντίκτυπο των μέτρων για την ενεργειακή απόδοση στο πλαίσιο της πολιτικής για τη συνοχή (2013/2038(INI)), χαιρετίζοντας τις νέες ευκαιρίες που προσφέρουν τα διαρθρωτικά και επενδυτικά ταμεία και το Ταμείο Συνοχής, ιδίως στον κατασκευαστικό τομέα. Το Κοινοβούλιο ζήτησε να ληφθούν μέτρα για την αύξηση της ευαισθητοποίησης και τη διάδοση των πληροφοριών και τόνισε την ανάγκη για ανάπτυξη ικανοτήτων και τεχνική βοήθεια στον τομέα αυτό.

Στο ψήφισμά του σχετικά με ένα πλαίσιο για τις πολιτικές που αφορούν το κλίμα και την ενέργεια με χρονικό ορίζοντα το έτος 2030 (2013/2135(INI)), το Κοινοβούλιο ζήτησε να τεθεί ένας δεσμευτικός στόχος ενεργειακής απόδοσης στην ΕΕ για το 2030 της τάξης του 40% με συνοδευτικούς επιμέρους εθνικούς στόχους. Επανέλαβε επίσης την παλαιότερη έκκλησή του να τεθεί ένας δεσμευτικός στόχος για το 2020

Στην παρούσα διπλωματική εργασία θα γίνει μια προσπάθεια πρόβλεψης κατανάλωσης ενέργειας σε κάποιο ΣΗΕ (Σύστημα Ηλεκτρικής Ενέργειας) και σε ένα Σύμπλεγμα Κτηρίων χρησιμοποιώντας κατάλληλες μεθόδους και προγραμματιστικά εργαλεία (R Studio). Στην συνέχεια θα γίνει ανάλυση των δεδομένων που θα προκύψουν από την πρόβλεψη χρησιμοποιώντας τα αποτελέσματα του προγράμματος.

1.3 ΠΡΟΒΛΕΨΗ ΕΝΕΡΓΕΙΑΚΗΣ ΚΑΤΑΝΑΛΩΣΗΣ

Στο παρελθόν έχουν γίνει αρκετές προσπάθειες πρόβλεψης ενεργειακής κατανάλωσης σε διάφορους τομείς με χρήση διαφορετικών μοντέλων πρόβλεψης. Στην παρούσα ενότητα θα γίνει μια γενικότερη ανασκόπηση για τις προβλέψεις όσον αφορά την περιοχή της ενεργειακής κατανάλωσης, τις διάφορες μεθόδους που υπάρχουν, την χρήση της καθεμίας καθώς και τα συμπεράσματα που έχουν προκύψει από προηγούμενες εργασίες στο κομμάτι αυτό.

1.4 ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ ΕΝΕΡΓΕΙΑΚΗΣ ΚΑΤΑΝΑΛΩΣΗΣ

Στην παρούσα ενότητα περιγράφονται οι διάφορες μέθοδοι πρόβλεψης ενεργειακής κατανάλωσης. Η ενεργειακή κατανάλωση ενός κτιρίου είναι αποτέλεσμα πολλών παραγόντων, όπως της θερμοκρασίας, του καιρού, των χαρακτηριστικών του κτιρίου και άλλων πολλών παραγόντων. Η ύπαρξη όλων αυτών των παραγόντων καθιστά αρκετά πολύπλοκη την πρόβλεψη και την ανάλυση της ενεργειακής κατανάλωσης των κτιρίων και πολλές φορές αρκετά χρονοβόρα και με υψηλό κόστος. Οι κύριες μέθοδοι που χρησιμοποιούνται για την πρόβλεψη και την ανάλυση της ενεργειακής κατανάλωσης είναι πέντε (Hai-xiang Zhao 2011 [\[1\]](#)) και αναλύονται στις παρακάτω υποενότητες:

1. Μηχανικές Μέθοδοι
2. Στατιστικές Μέθοδοι
3. Νευρωνικά Δίκτυα
4. Μηχανές Υποστήριξης Διανυσμάτων
5. Μοντέλα Grey

Κάθε μέθοδος έχει αναλυθεί και αναπτυχθεί για συγκεκριμένους σκοπούς. Η καθεμία έχει μειονεκτήματα και πλεονεκτήματα και η χρήση τους γίνεται ανάλογα την κατάσταση την οποία θέλουμε να εκτιμήσουμε. Παρακάτω αναλύονται η καθεμία ξεχωριστά με τα πλεονεκτήματα και τα μειονεκτήματά της καθώς και τις περιπτώσεις όπου η καθεμία είναι περισσότερο χρήσιμη ή όχι. Στην παρούσα διπλωματική δεν θα γίνει χρήση όλων των μεθόδων, ωστόσο για λόγους πληρότητας τις παρουσιάζουμε συνοπτικά όλες.

1.4.1 ΜΗΧΑΝΙΚΕΣ ΜΕΘΟΔΟΙ (*Engineering Methods*)

Οι μηχανικές μέθοδοι είναι το πρώτο είδος μεθόδων για τον υπολογισμό και την ανάλυση της ενεργειακής κατανάλωσης ενός κτιρίου. Πρόκειται για μεθόδους που χρησιμοποιούν τις φυσικές αρχές για να υπολογίσουν την θερμοδυναμική και την ενεργειακή συμπεριφορά ενός κτιρίου ή ενός συγκεκριμένου στοιχείου ενός κτιρίου. Οι μηχανικές μέθοδοι χωρίζονται σε δύο κύριες υποκατηγορίες, τις απλοϊκές μεθόδους και τις λεπτομερείς ολοκληρωμένες μεθόδους. Δεν υπάρχει κάποιο εμφανές όριο μεταξύ των δύο υποκατηγοριών, αντιθέτως μπορούμε να κάνουμε την ανάλυση μας με την απλοϊκή μέθοδο χρησιμοποιώντας και εργαλεία των ολοκληρωμένων μεθόδων.

Οι λεπτομερείς ολοκληρωμένες μέθοδοι χρησιμοποιούν πολύ επεξεργασμένες φυσικές λειτουργίες και θερμοδυναμική για να κάνουν βήμα προς βήμα υπολογισμό της ενεργειακής κατανάλωσης των κτιρίων. Το κύριο μειονέκτημα των μεθόδων αυτών είναι ότι για την χρήση των εργαλείων τους χρειαζόμαστε αρκετές λεπτομέρειες των κτιρίων και παραμέτρους του περιβάλλοντος ώστε να παραχθεί το τελικό αποτέλεσμα. Τα δεδομένα αυτά αρκετές φορές δεν είναι διαθέσιμα με αποτέλεσμα τα αποτελέσματά μας να μην είναι όσο ακριβή θα θέλαμε. Ένα άλλο πολύ σημαντικό μειονέκτημα των μεθόδων αυτών είναι το πολύ υψηλό κόστος εφαρμογής τους. Για την χρήση των εργαλείων τους χρειάζονται εξειδικευμένα άτομα με γνώση του αντικειμένου, ενώ απαιτείται αρκετός χρόνος για την σύνθεση των δεδομένων και την παραγωγή των τελικών αποτελεσμάτων. Κάποια πολύ γνωστά προγράμματα των μεθόδων αυτών είναι τα DOE-2, EnergyPlus, BLAST, ESP-r.

Όσον αφορά τις απλοϊκές μεθόδους υπάρχουν δύο κύριες κατηγορίες, η *degree day method* και η *bin or temperature frequency method*. Η πρώτη είναι κατάλληλη για τον υπολογισμό της ενεργειακής κατανάλωσης σε μικρά κτίρια ενώ η δεύτερη είναι κατάλληλη για την μοντελοποίηση μεγάλων κτιριακών μονάδων με σκοπό τον υπολογισμό της ενεργειακής κατανάλωσης τους.

Η σωστή επιλογή των παραπάνω μεθόδων έχει να κάνει με διάφορους παράγοντες όπως αναφέρει ο Al-Homoud το 2001 [2], όπως η ακρίβεια, η ευαισθησία, η ταχύτητα και το κόστος εκμάθησης και χρήσης τους, η διαθεσιμότητα των απαιτούμενων δεδομένων, η ποιότητα των αποτελεσμάτων κ.α. Ο Al-Homoud θέλοντας να μετρήσει την ενεργειακή κατανάλωση κάποιων κτιρίων χρησιμοποίησε τόσο απλοϊκές όσο και λεπτομερείς ολοκληρωμένες μεθόδους με βάση τους παραπάνω παράγοντες. Τα αποτελέσματα της έρευνάς του δείχνουν ότι στην περίπτωση που θέλουμε απλά να μελετήσουμε την τάση ή να συγκρίνουμε συστήματα τότε οι απλοϊκές μέθοδοι ανάλυσης είναι αρκετές, αντιθέτως αν θέλουμε μια λεπτομερή ενεργειακή ανάλυση των κτιρίων καθώς και το κόστος ανάλυσης τότε οι λεπτομερείς ολοκληρωμένες μέθοδοι με την χρήση υπολογιστικών συστημάτων είναι καταλληλότερες.

1.4.2 ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ (*Statistical Methods*)

Οι στατιστικές μέθοδοι συσχετίζουν την ενεργειακή κατανάλωση των κτιρίων με τις μεταβλητές που την επηρεάζουν. Τα κυριότερα μοντέλα είναι η *Naïve*, η *Theta*, τα Μοντέλα Παλινδρόμησης, τα Μοντέλα *ARIMA* και οι μέθοδοι εκθετικής εξομάλυνσης. Αυτά τα εμπειρικά μοντέλα σχεδιάζονται με τη χρήση ιστορικών στοιχείων απόδοσης, επομένως προτού γίνει η εφαρμογή τους πρέπει πρώτα να γίνει η συλλογή αρκετών στοιχείων. Η εφαρμογή των μεθόδων αυτών είναι αρκετά εύκολη αλλά δεν έχουν τόσο καλή ακρίβεια αποτελεσμάτων και δεν είναι τόσο εύκαμπτα όσο τα προηγούμενα μοντέλα. Κατά την εφαρμογή των μοντέλων το κυριότερο πρόβλημα είναι ο υπολογισμός σημαντικών παραμέτρων της ενεργειακής χρήσης, όπως η συνολική απώλεια θερμότητας που είναι χρήσιμη για την ανάλυση της θερμικής συμπεριφοράς των κτιρίων. Ο A. Kimbara το 1995 [3] χρησιμοποίησε ένα μοντέλο *ARIMA* για την πρόβλεψη του απαιτούμενου φορτίου ενός συστήματος θέρμανσης. Ο Kimbara αρχικά εφάρμοσε το μοντέλο σε παρελθοντικά δεδομένα και στη συνέχεια το χρησιμοποίησε για την πρόβλεψη του φορτίου της επόμενης ημέρας. Για την αξιολόγηση του μοντέλου έγινε καταγραφή του φορτίου θέρμανσης/ψύξης στην Osaka το 1987. Τα αποτελέσματα της έρευνας έδειξαν πολύ καλή συμφωνία των δεδομένων που προβλέφθηκαν, με τα μοντέλα *ARIMA*, με τα πραγματικά δεδομένα που συλλέχθηκαν.

Αργότερα το 1998 ο Aj. Hoffman [4] πρόβλεψε την μέγιστη ενεργειακή ζήτηση (peak demand) σε εμπορικά κτίρια χρησιμοποιώντας μοντέλα *ARIMAX*. Αρχικά ο Hoffman συνειδητοποίησε την κυκλικότητα ανά εβδομάδα που χαρακτήριζε τα δεδομένα και για τον λόγο αυτόν χρησιμοποίησε εποχικό μοντέλο με περίοδο μιας εβδομάδας με επαρκή ιστορικά δεδομένα ζήτησης. Τα δεδομένα συλλέχθηκαν σε περίοδο χρόνου όπου δεν είχε εφαρμοστεί κανένας δυναμικός έλεγχος. Ο Hoffman χρησιμοποίησε στο μοντέλο του μια εξωγενή παράμετρο η οποία ήταν μια κατάσταση on/off ώστε να εξαλείψει την επίδραση των *shedtable loads*, των φορτίων δηλαδή που θα επηρεάσουν την ζήτηση σε περίπτωση που η τροφοδοσία είναι ανεπαρκής. Η ανάλυση του κατέληξε σε πολύ καλά αποτελέσματα σε σχέση με τα πραγματικά δεδομένα και απέδειξε πως η σωστή πρόβλεψη της συνολικής μελλοντικής ζήτησης μπορεί να χρησιμοποιηθεί ώστε να παραμείνει η μέγιστη ζήτηση (peak demand) σε θεμιτά όρια.

Το 2008 οι M. Aydinalp-Koksal και V.I. Ugursal [5] θέλησαν να συγκρίνουν την απόδοση ενός μοντέλου βασιζόμενο στην παλινδρόμηση με τις μηχανικές μεθόδους και τα νευρωνικά δίκτυα τα οποία αναλύονται παρακάτω. Σκοπός της μελέτης τους ήταν η πρόβλεψη της ενεργειακής κατανάλωσης σε εθνικό επίπεδο. Η χώρα που επιλέχθηκε για την πρόβλεψη ήταν ο Καναδάς, καθώς έχει δεσμευθεί νομικά να μειώσει την ενεργειακή κατανάλωση του. Το μοντέλο που χρησιμοποίησαν ονομάζεται Ανάλυση υπό Όρους Ζήτησης (*CDA*) το οποίο είχε προταθεί το 1980 από τους Pati M. και Pati C. Το *CDA* που παρουσιάστηκε το 2008 αποτελούνταν από τρία μοντέλα, το ηλεκτρικό μοντέλο, το μοντέλο φυσικού αερίου και το μοντέλο πετρελαίου. Χρησιμοποιήθηκαν δυο βάσεις δεδομένων του 1993, για την πραγματοποίηση των προβλέψεων, οι οποίες προέρχονταν από καταναλώσεις νοικοκυριών και συστημάτων θέρμανσης και ψύξης. Τα αποτελέσματα έδειξαν ότι το μοντέλο *CDA* πραγματοποίησε προβλέψεις αντίστοιχης ακρίβειας με αυτή των νευρωνικών δικτύων και των μηχανικών μεθόδων με αρκετά μεγαλύτερη ευκολία στην εφαρμογή του αλλά με μικρότερη ευελιξία όσον αφορά την αξιολόγηση κοινωνικό-οικονομικών παραγόντων και σεναρίων εξοικονόμησης ενέργειας.

1.4.3 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (ANNs)/ ΜΗΧΑΝΗΜΑΤΑ ΥΠΟΣΤΗΡΙΞΗΣ ΔΙΑΝΥΣΜΑΤΩΝ (SVMs)

Το νευρωνικό δίκτυο είναι ένα δίκτυο από απλούς υπολογιστικούς κόμβους (νευρώνες), διασυνδεδεμένους μεταξύ τους. Η ιδέα του προέρχεται από το Κεντρικό Νευρικό Σύστημα, το οποίο και προσπαθεί να προσομοιώσει. Οι υπολογιστικοί κόμβοι αποτελούν τα δομικά συστατικά του δικτύου και με βάση τις εισόδους που δέχονται εκτελούν κάποιους υπολογισμούς και παράγουν μια έξοδο η οποία αποτελεί είσοδο κάποιου άλλου κόμβου ή έξοδο στο περιβάλλον. Υπάρχουν τριών ειδών νευρώνων. Οι νευρώνες εισόδου μεσολαβούν απλά ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδο τους με το αντίστοιχο συναπτικό βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η εγγενής ικανότητα μάθησης. Ως μάθηση μπορεί να οριστεί η σταδιακή βελτίωση της ικανότητας του δικτύου να επιλύει κάποιο πρόβλημα (π.χ. η σταδιακή προσέγγιση μίας συνάρτησης). Η μάθηση επιτυγχάνεται μέσω της εκπαίδευσης, μίας επαναληπτικής διαδικασίας σταδιακής προσαρμογής των παραμέτρων του δικτύου (συνήθως των βαρών και της πόλωσης του) σε τιμές κατάλληλες ώστε να επιλύεται με επαρκή επιτυχία το προς εξέταση πρόβλημα. Αφού ένα δίκτυο εκπαιδευτεί, οι παράμετροί του συνήθως «παγώνουν» στις κατάλληλες τιμές και από εκεί κι έπειτα είναι σε λειτουργική κατάσταση. Το ζητούμενο είναι το λειτουργικό δίκτυο να χαρακτηρίζεται από μία ικανότητα γενίκευσης: αυτό σημαίνει πως δίνει ορθές εξόδους για εισόδους καινοφανείς και διαφορετικές από αυτές με τις οποίες εκπαιδεύτηκε. Είναι τα πιο ευρέως χρησιμοποιούμενα μοντέλα τεχνητής νοημοσύνης σε εφαρμογές για πρόβλεψη της ενέργειας των κτιρίων. Είναι καλά και αποτελεσματικά για την επίλυση μη γραμμικών προβλημάτων, ενώ χρησιμοποιούνται επίσης για να αναλύσουν και να επεξεργαστούν τη συμπεριφορά στοιχείων του κτιρίου, κυρίως για συστήματα HVAC. Με την χρήση των μοντέλων αυτών καθίσταται δυνατός ο υπολογισμός παραμέτρων ενεργειακής κατανάλωσης κτιρίων. Και τα νευρωνικά δίκτυα και τα μηχανικά μοντέλα είναι αρκετά ακριβή για πρόβλεψη ενεργειακής κατανάλωσης κτιρίων, ωστόσο τα ANNs είναι καλύτερα για βραχυπρόθεσμη περίοδο πρόβλεψης. Τα μοντέλα αυτά χρειάζονται μεγάλο αριθμό ιστορικών δεδομένων απόδοσης για τον σχεδιασμό τους και την επικύρωση των αποτελεσμάτων τους και είναι αρκετά περίπλοκα στην εφαρμογή τους. Το 2000 οι Kalogirou S. και M.Bojic [6] χρησιμοποίησαν τα νευρωνικά δίκτυα με σκοπό να προβλέψουν την ενεργειακή κατανάλωση των παθητικών ηλιακών κτιρίων. Οι προβλέψεις τους πραγματοποιήθηκαν για το χειμώνα και το καλοκαίρι με αποδεκτή ακρίβεια. Τα αποτελέσματα της έρευνας τους έχουν μεγάλη σημασία καθώς δίνουν την δυνατότητα σε σχεδιαστές κτιρίων και συστημάτων να κάνουν γρήγορες δοκιμές, με ελάχιστα δεδομένα εισόδου με σκοπό να εκτιμήσουν την κατάσταση και τα οφέλη που θα προκύψουν. Το κυριότερο πλεονέκτημα τους είναι ο χρόνος περάτωσης των προβλέψεων.

Τα μηχανήματα υποστήριξης διανυσμάτων είναι μοντέλα που χρησιμοποιούνται όλο και περισσότερο στο πεδίο της έρευνας και των βιομηχανιών. Είναι αρκετά αποτελεσματικά για την επίλυση μη γραμμικών προβλημάτων ακόμα και αν έχουμε λίγα δεδομένα στην διάθεσή μας. Και τα μοντέλα αυτά όπως και τα νευρωνικά δίκτυα χρειάζονται μεγάλο πλήθος ιστορικών δεδομένων απόδοσης και η εφαρμογή τους είναι αρκετά περίπλοκη. Το 2005 οι B.Dong και C.Cao [7] χρησιμοποίησαν τα SVM για να προβλέψουν την ενεργειακή κατανάλωση των κτιρίων σε τροπικές περιοχές. Σκοπός της έρευνας τους ήταν να εξετάσουν αν μπορούν να

εφαρμοστούν τα μηχανήματα αυτά για την πραγματοποίηση προβλέψεων στον τομέα της ενεργειακής κατανάλωσης. Για τον σκοπό αυτόν επιλέχθηκε τυχαία τέσσερα εμπορικά κτίρια στη Σιγκαπούρη. Τα αποτελέσματα της μελέτης τους έδειξαν ότι τα *SVM* έχουν καλύτερη απόδοση από τα *ANNs* και από τα *ARIMA Models*. Παρατηρήθηκε επίσης ότι η απόδοση τους είναι αρκετά καλή όταν πρόκειται για πρόβλεψη ωριαίων και μηνιαίων τιμών ενέργειας.

1.4.4 ΜΟΝΤΕΛΑ GREY

Ένα σύστημα ονομάζεται σύστημα *Grey* όταν οι πληροφορίες για το σύστημα αυτό είναι μερικώς γνωστές. Τέτοια μοντέλα χρησιμοποιούνται για την ανάλυση της ενεργειακής συμπεριφοράς των κτιρίων μόνο σε περίπτωση όπου τα δεδομένα είναι ελλιπή ή αβέβαια. Δεν υπάρχουν αρκετά δεδομένα για την απόδοση των μοντέλων αυτών, λίγοι τα έχουν χρησιμοποιήσει και από την μέχρι τώρα ανάλυση τα αποτελέσματα που έχουν προκύψει είναι επιεικώς καλά. Το 1999 οι Wang X και Chen Z [8] χρησιμοποίησαν ένα τέτοιο μοντέλο για να πραγματοποιήσουν προβλέψεις για το σύστημα θέρμανσης κάποιων κτηρίων με αρκετά καλή απόδοση. Απέδειξαν ότι το μοντέλο είναι εφαρμόσιμο για την πρόβλεψη της ενεργειακής κατανάλωσης των κτιρίων και η ακρίβεια του είναι αρκετά υψηλή. Οι παράγοντες που επηρέαζαν την κατανάλωση και την εφαρμογή του μοντέλου ήταν τόσο εξωτερικοί όσο και εσωτερικοί. Το κλίμα, η θερμοκρασία, η υγρασία και η ακτινοβολία του ηλίου ήταν κάποιο από τους εξωτερικούς παράγοντες ενώ ο εξοπλισμός, ο φωτισμός και οι εργάτες ήταν κάποιοι από τους εσωτερικούς παράγοντες.

ΚΕΦΑΛΑΙΟ 2: ΜΟΝΤΕΛΑ ΧΡΟΝΟΣΕΙΡΩΝ

2.1 ΜΟΝΤΕΛΟ ΧΡΟΝΟΣΕΙΡΩΝ

Το μοντέλο χρονοσειρών βασίζεται στην ιδέα ότι η μεταβολή της τιμής του μεγέθους ακολουθεί ένα συγκεκριμένο λανθάνον πρότυπο που επαναλαμβάνεται στο χρόνο και παραμένει σταθερό. Η συνάρτηση που περιγράφει το σύστημα είναι προκαθορισμένη και προσδιορίζεται από το πρότυπο που υποθέτουμε ότι ακολουθούν τα ιστορικά δεδομένα.

Οι μέθοδοι χρονοσειρών που χρησιμοποιούμε είναι τρεις: η αποσύνθεση (*decomposition*), η εξομάλυνση (*smoothing*) και οι αυτοπαλινδρομικές μέθοδοι κινητού μέσου όρου (*autoregressive moving average ARMA*). Τα κύρια πλεονεκτήματα που καθιστούν συχνή την χρήση των μεθόδων των χρονοσειρών είναι τα εξής:

- δεν υπάρχει πάντα η δυνατότητα να συσχετίσουμε ένα μεταβαλλόμενο μέγεθος με κάποιους παράγοντες και πολύ περισσότερο να προσδιορίσουμε τον τρόπο αλληλεπίδρασης τους,
- γιατί σε πολλές περιπτώσεις ενδιαφερόμαστε να προσδιορίσουμε μόνο το τι θα συμβεί και όχι το γιατί
- και γιατί το κόστος που απαιτείται στην πρώτη περίπτωση είναι πολύ μικρότερο από αυτό που απαιτείται στην δεύτερη

2.2 ΒΑΣΙΚΑ ΒΗΜΑΤΑ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΠΡΟΒΛΕΨΗΣ

- Καθορισμός του προβλήματος: αποτελεί αρκετές φορές το πιο δύσκολο μέρος της διαδικασίας. Πρέπει να γίνει απολύτως σαφές τι θέλουμε να προβλέψουμε αλλά και πώς και από ποιους θα χρησιμοποιηθούν οι παραχθείσες προβλέψεις.
- Συλλογή των δεδομένων: πρέπει να δαπανήσουμε αρκετό χρόνο για την ορθή συλλογή και συντήρηση των δεδομένων. Εκτός των μετρήσιμων αριθμητικών δεδομένων (πωλήσεις, κέρδος κλπ.) κρίνεται απαραίτητη η συλλογή και αξιοποίηση της κριτικής εμπειρίας και συσσωρευμένης γνώσης των εργαζομένων στην επιχείρηση για την εξεταζόμενη χρονική περίοδο.
- Προετοιμασία χρονοσειρών: στο βήμα αυτό πρέπει να εξομαλύνουμε την σειρά δεδομένων (πιθανά λανθάνοντα πρότυπα, σημαντική τάση ή εποχικότητα και ασυνήθιστες τιμές) ώστε να είναι έτοιμη για την εφαρμογή μοντέλων πρόβλεψης.

- Επιλογή μεθόδων πρόβλεψης: στο βήμα αυτό επιτυγχάνεται η ορθή επιλογή κατάλληλων μοντέλων πρόβλεψης, αλλά και η επιλογή των παραμέτρων τους, βάσει των ιστορικών δεδομένων, αλλά και των ιδιοτήτων της κάθε μεθόδου.
- Χρήση και αξιολόγηση των μοντέλων πρόβλεψης. Το επιλεγμένο μοντέλο χρησιμοποιείται για να παραχθούν οι ζητούμενες προβλέψεις. Η αξιολόγηση και η μέτρηση της ακρίβειας των προβλέψεων επιτυγχάνεται με εξειδικευμένους στατιστικούς δείκτες. Η παρακολούθηση του προτύπου της χρονοσειράς και των σφαλμάτων της πρόβλεψης συμβάλλει στην έγκαιρη αντιμετώπιση διαρθρωτικών αλλαγών και στην εξάλειψη προκατάληψης στις τελικές προβλέψεις.

2.3 ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΕΩΝ

2.3.1 ΝΑΪΒΕ Η ΑΠΛΟΪΚΗ ΜΕΘΟΔΟΣ

Η μέθοδος *Naive* ή απλοϊκή μέθοδος, όπως ονομάζεται, είναι η πιο απλή μέθοδος πρόβλεψης. Δεν χρησιμοποιείται στην καθημερινότητα για προβλέψεις αλλά ως σημείο αναφοράς για άλλες πιο πολύπλοκες μεθόδους.

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

Όπως φαίνεται και από τον τύπο της μεθόδου, η πρόβλεψη που παράγει για μια χρονική στιγμή t είναι στην ουσία η παρατήρηση της ακριβώς προηγούμενης χρονικής περιόδου $t-1$.

2.3.2 ΜΕΘΟΔΟΙ ΕΚΘΕΤΙΚΗΣ ΕΞΟΜΑΛΥΝΣΗΣ

Αποτελούν τις πιο δημοφιλείς μεθόδους προβλέψεων μεταξύ των επιχειρηματιών λόγω της ευκολίας τους, της ελάχιστης απαίτησης σε υπολογιστικό χρόνο και της απαίτησης σχετικά

λίγων παρατηρήσεων προκειμένου να παραχθούν οι προβλέψεις. Οι πιο γνωστές μέθοδοι εκθετικής εξομάλυνσης είναι:

- Απλή Εκθετική Εξομάλυνση Σταθερού Επιπέδου (*Simple Exponential Smoothing*)
- Εκθετική Εξομάλυνση Γραμμικής Τάσης (*Holt Exponential Smoothing*)
- Εκθετική Εξομάλυνση μη Γραμμικής ή Φθίνουσας Τάσης (*Damped Exponential Smoothing*)

Οι μέθοδοι εξομάλυνσης είναι κατάλληλες κυρίως για βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις μεγάλου όγκου χρονοσειρών. Οι μέθοδοι αυτές παρουσιάζουν ικανοποιητικά ποσοστά ακριβείας σε σχέση με πιο πολύπλοκες μεθόδους πρόβλεψης καθώς δεν επηρεάζονται από περιστασιακά εμφανιζόμενες ακραίες τιμές ή από τις ιδιομορφίες των προτύπων των δεδομένων, ανάλογα πάντα με τον συντελεστή εξομάλυνσης.

2.3.2.1 ΑΠΛΗ ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ (*Simple Exponential Smoothing SES*)

Το μοντέλο της *SES* χρησιμοποιείται για προβλέψεις ενός βήματος ή όταν οι χρονοσειρές χαρακτηρίζονται από αυξημένο θόρυβο ή τυχαιότητα. Επίσης το μοντέλο της *SES* υποθέτει την απουσία τάσης από τα δεδομένα. Οι εξισώσεις που περιγράφουν το μοντέλο είναι οι εξής:

$$e_t = Y_t - F_t$$

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

$$F_{t+1} = S_t$$

Όπου το e δηλώνει το σφάλμα, το S το επίπεδο, το F την πρόβλεψη και το α δηλώνει τον συντελεστή εξομάλυνσης. Η επιλογή του συντελεστή εξομάλυνσης εξαρτάται από το ποσοστό θορύβου στην χρονοσειρά και από την σταθερότητα του μέσου όρου της χρονοσειράς. Γενικά όσο περισσότερο θόρυβο έχουμε τόσο μικρότερος θα πρέπει να είναι ο συντελεστής εξομάλυνσης που επιλέγουμε για να μην έχουμε υπερβολική αντίδραση στον θόρυβο ενώ αντίθετα όσο μεταβάλλεται ο μέσος όρος της χρονοσειράς τόσο ο συντελεστής θα πρέπει να είναι μεγαλύτερος ώστε οι προβλέψεις μας να ακολουθούν τις μεταβολές των δεδομένων.

Όσον αφορά το σφάλμα πρόβλεψης αυτό που μπορούμε να πούμε είναι ότι για σφάλμα θετικό τότε η πρόβλεψη για την επόμενη χρονική περίοδο αυξάνεται, ενώ αν το σφάλμα πρόβλεψης είναι αρνητικό τότε η πρόβλεψη για την επόμενη χρονική περίοδο μειώνεται.

Με βάση τους τύπους της μεθόδου παρατηρούμε ότι πρέπει να οριστεί ένα αρχικό επίπεδο S_0 ώστε να μπορέσουμε να εφαρμόσουμε την διαδικασία για να πραγματοποιήσουμε τις προβλέψεις. Το αρχικό αυτό επίπεδο είναι συνήθως:

- 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

Επιλέγουμε μια από τις παραπάνω εναλλακτικές ανάλογα με τα ποιοτικά χαρακτηριστικά της χρονοσειράς μας.

2.3.2.2 ΜΟΝΤΕΛΟ ΓΡΑΜΜΙΚΗΣ ΤΑΣΗΣ (*Holt Exponential Smoothing*)

Το μοντέλο αυτό αποτελεί την πιο δημοφιλή παραδοχή. Οι προβλέψεις προκύπτουν από την προέκταση μιας ευθείας γραμμής για οποιαδήποτε χρονική περίοδο στο μέλλον. Το μοντέλο αυτό είναι μια προέκταση του προηγούμενου με την διαφορά ότι μπορεί να διαχειριστεί τη συνιστώσα της τάσης. Οι εξισώσεις που περιγράφουν το μοντέλο είναι οι εξής:

$$e_t = Y_t - F_t$$

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

$$T_t = T_{t-1} + \beta \cdot e_t$$

$$F_{t+m} = S_t + m \cdot T_t$$

Όπου το e είναι το σφάλμα της πρόβλεψης, S το επίπεδο, T η τάση, F η πρόβλεψη, m ο χρονικός ορίζοντας της πρόβλεψης ενώ οι παράμετροι α και β είναι οι συντελεστές εξομάλυνσης του επιπέδου και της τάσης αντίστοιχα.

Στο μοντέλο αυτό ορίζουμε αρχικό επίπεδο S_0 όπως και στο προηγούμενο ενώ τώρα ορίζουμε και αρχική τάση T_0 ως:

- Διαφορά δεύτερης και πρώτης παρατήρησης ($Y_2 - Y_1$)
- Διαφορά n -στης και πρώτης παρατήρησης διαιρεμένης με $n-1$
- Η σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης

Στην μέθοδο *Holt* η επιλογή των συντελεστών εξομάλυνσης α , β βασίζεται στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος *MSE*. Συνήθως η βέλτιστη τιμή του συντελεστή β για την τάση είναι μικρότερη από την τιμή του συντελεστή α για το επίπεδο.

2.3.2.3 ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ ΜΗ ΓΡΑΜΜΙΚΗΣ Ή ΦΘΙΝΟΥΣΑΣ ΤΑΣΗΣ (*Damped*)

Αποτελεί την καλύτερη επιλογή στις περιπτώσεις μεσοπρόθεσμων και μακροπρόθεσμων προβλέψεων. Με την εφαρμογή του μοντέλου αυτού ουσιαστικά μειώνουμε βαθμιαία το μέγεθος κατά το οποίο αυξάνονται οι τιμές της χρονοσειράς κάθε χρονική περίοδο. Το μοντέλο αυτό είναι μια παραλλαγή της *Holt* έτσι ώστε να μπορεί να προσαρμόζεται και στις μη γραμμικές τάσεις με την προσθήκη μιας παραμέτρου διόρθωσης της τάσης, η οποία ελέγχει το ρυθμό αύξησης των τιμών τους. Οι εξισώσεις που περιγράφουν το μοντέλο είναι οι εξής:

$$e_t = Y_t - F_t$$

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

$$T_t = \phi \cdot T_{t-1} + \beta \cdot e_t$$

$$F_{t+m} = S_t + \sum_{i=1}^m \phi^i \cdot T_t$$

Από την τιμή που θα λάβει η παράμετρος ϕ θα εξαρτηθεί τελικά το μοντέλο πρόβλεψης. Διακρίνονται οι εξής περιπτώσεις:

- $\phi = 0$ τότε προκύπτει το μοντέλο της απλής εκθετικής εξομάλυνσης
- $0 < \phi < 1$ τότε προκύπτει το μοντέλο της φθίνουσας τάσης
- $\phi = 1$ τότε προκύπτει το μοντέλο της γραμμικής τάσης
- $\phi > 1$ τότε προκύπτει το μοντέλο της εκθετικής τάσης

Όσο μεγαλύτερος είναι ο απαιτούμενος ορίζοντας πρόβλεψης τόσο πλεονεκτεί το μοντέλο μη γραμμικής τάσης από τα άλλα μοντέλα εκθετικής εξομάλυνσης.

2.3.3 ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (*Regression Models*)

Τα μοντέλα παλινδρόμησης μελετούν τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής με συγκεκριμένες ανεξάρτητες μεταβλητές. Ελέγχουμε δηλαδή με ποιο τρόπο μεταβάλλεται μια εξαρτημένη μεταβλητή αν αλλάξει μια ανεξάρτητη μεταβλητή και όλες οι άλλες παραμένουν ίδιες. Τα μοντέλα παλινδρόμησης είναι τα εξής:

- Απλή Γραμμική Παλινδρόμηση
- Πολλαπλή Γραμμική Παλινδρόμηση

Τα μοντέλα αυτά θεωρούνται κατάλληλα για παραγωγή μακροπρόθεσμων προβλέψεων. Παρόλο που η παλινδρόμηση χρησιμοποιείται ευρέως σε θέματα προβλέψεων, ο κύριος ρόλος χρησιμοποίησής της είναι η ανάλυση και η κατανόηση των σχέσεων μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών.

Μπορούμε να βγάλουμε συμπεράσματα σχετικά με το ποιες μεταβλητές είναι συσχετισμένες επαρκώς αλλά και τη μορφή της συσχέτισης τους. Τα μοντέλα παλινδρόμησης περιλαμβάνουν τις εξής μεταβλητές:

- οι άγνωστες παράμετροι συσχέτισης που δηλώνονται ως β (διάνυσμα)
- οι ανεξάρτητες μεταβλητές X (διάνυσμα)
- η εξαρτώμενη μεταβλητή Y

Ο τύπος που συσχετίζει τις παραπάνω μεταβλητές και περιγράφει τα μοντέλα παλινδρόμησης είναι ο κάτωθι:

$$E(Y / X) = f(X, \beta)$$

2.3.3.1 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (*Simple Linear Regression*)

Στο μοντέλο αυτό υποθέτουμε μια γραμμική σχέση μεταξύ της μεταβλητής πρόβλεψης (εξαρτημένης) και μιας άλλης μεταβλητής (ανεξάρτητης). Η απλή γραμμική παλινδρόμηση εκφράζει την σχέση μεταξύ των δυο παραπάνω μεταβλητών με την εξίσωση μιας ευθείας γραμμής:

$$\hat{Y}_i = a + b \cdot X_i$$

Το b όπως ξέρουμε δηλώνει την κλίση της ευθείας και το a δηλώνει την τιμή της εξαρτημένης μεταβλητής όταν υποθέσουμε ότι η ανεξάρτητη τιμή παίρνει την τιμή 0. Στις περιπτώσεις πρόβλεψης χρονοσειρών επιλέγουμε ως ανεξάρτητη μεταβλητή τον χρόνο, δηλαδή τις περιόδους.

Για την εύρεση των τιμών των a, b χρησιμοποιούμε τη μέθοδο των ελαχίστων τετραγώνων. Σκοπός μας είναι να επιλέξουμε τους δείκτες αυτούς που ελαχιστοποιούν τα σφάλματα πρόβλεψης. Τους δείκτες δηλαδή εκείνους που ελαχιστοποιούν τις διαφορές των πραγματικών τιμών από τις τιμές που προκύπτουν από την εξίσωση παλινδρόμησης. Με την χρήση της μεθόδου των ελαχίστων τετραγώνων προκύπτουν οι εξής τύποι:

$$a = \bar{Y} - b \cdot \bar{X}$$

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

Όπου \bar{Y} και \bar{X} οι μέσες τιμές των διανυσμάτων Y, X και n ο αριθμός των παρατηρήσεων.

2.3.3.2 ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (*Multivariate Linear Regression*)

Το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης αποτελεί γενίκευση του μοντέλου απλής γραμμικής παλινδρόμησης σε περιπτώσεις όπου απαιτούνται περισσότερες από μία ανεξάρτητες μεταβλητές. Όπως και παραπάνω έχουμε μία εξαρτώμενη μεταβλητή που όμως τώρα εξαρτάται από περισσότερες από μία ανεξάρτητες μεταβλητές. Η εξίσωση του μοντέλου είναι η εξής:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_k \cdot X_k + e$$

Όπου η μεταβλητή Y εκφράζει την εξαρτημένη μεταβλητή, τα X_1, X_2, \dots, X_k τις ανεξάρτητες μεταβλητές, τα b_0, b_1, \dots, b_k είναι σταθερές παράμετροι και το e είναι ο τυχαίος παράγοντας, ο οποίος είναι κανονικοποιημένος γύρω από το 0.

Και στην περίπτωση του μοντέλου πολλαπλής παλινδρόμησης οι συντελεστές b υπολογίζονται με την μέθοδο των ελαχίστων τετραγώνων, ωστόσο στην συγκεκριμένη περίπτωση δεν είναι

εύκολο να σχεδιαστεί η σχέση της εξαρτημένης με τις ανεξάρτητες μεταβλητές καθώς πλέον δεν αναφερόμαστε σε μια ευθεία αλλά σε υπερεπίπεδο.

2.3.4 ΜΕΘΟΔΟΣ ΘΗΤΑ

Η μέθοδος αυτή βασίζεται στην τροποποίηση των τοπικών καμπυλοτήτων της χρονοσειράς. Η τροποποίηση αυτή πραγματοποιείται με μια παράμετρο θ . Η μέθοδος *Theta* διαχωρίζει την αρχική χρονοσειρά σε δυο ή περισσότερες γραμμές *Theta*. Η κάθε γραμμή προεκτείνεται ξεχωριστά και μετά συνδυάζουμε τις παραγόμενες προβλέψεις για να προκύψει η τελική πρόβλεψη. Η μέθοδος *Theta* μπορεί να θεωρηθεί ως μια εναλλακτική προσέγγιση της αποσύνθεσης και μια επέκταση της θεωρίας των γραμμικών συνδυασμών παραγόμενων προβλέψεων. Η εφαρμογή της μεθόδου ακολουθεί 6 βήματα:

1. Έλεγχος εποχικότητας
2. Αποεποχικοποίηση
3. Αποσύνθεση
4. Πρόβλεψη
5. Συνδυασμός
6. Εποχικοποίηση

Από την γραμμή $\theta = 0$ εξασφαλίζεται η μακροπρόθεσμη τάση ενώ η ύπαρξη της γραμμής $\theta = 2$ εξασφαλίζει την αξιοποίηση της βραχυπρόθεσμης πληροφορίας. Σύμφωνα με τους δημιουργούς της μεθόδου (B. Ασημακόπουλος και Κ. Νικολόπουλος 2000 [9]) οι βασικές της ιδιότητες είναι οι εξής:

- Κάθε γραμμή *Theta* (Y^θ) προκύπτει από την επίλυση ενός γραμμικού συστήματος εξισώσεων του οποίου οι συντελεστές εξαρτώνται από τα δεδομένα της αρχικής χρονοσειράς και την τιμή της παραμέτρου θ .
- Κάθε γραμμή *Theta* διατηρεί τη μέση τιμή και την κλίση παλινδρόμησης της αρχικής χρονοσειράς.
- $Y_t^{\theta=0} \equiv LRL_t$ (*LRL*: *Linear regression line* ή ευθεία γραμμικής παλινδρόμησης)
- $Y_t = \frac{1}{2} \cdot (Y_t^{\theta=1-\alpha} + Y_t^{\theta=1+\alpha})$

Οι εξισώσεις που περιγράφουν το μοντέλο της *Theta* είναι οι εξής:

$$Y_t^\theta = \theta \cdot Y_t''$$

$$Y_t'' = Y_t - 2 \cdot Y_{t-1} + Y_{t-2}$$

2.3.5 ΜΟΝΤΕΛΑ ARIMA (*Autoregressive Integrated Moving Average*)

Τα μοντέλα *ARIMA* είναι στοχαστικά μαθηματικά μοντέλα με τα οποία προσπαθούμε να περιγράψουμε τη διαχρονική εξέλιξη κάποιου φυσικού μεγέθους. Αναφέρονται συχνά ως *ARIMA(p,d,q)* και εμπεριέχουν τον τυχαίο παράγοντα, τις τιμές του μεγέθους, οι οποίες εμφανίστηκαν σε προηγούμενες χρονικές στιγμές, και ίσως κάποιους άλλους στοχαστικούς παράγοντες.

Χρησιμοποιούνται ευρέως γιατί βρίσκουν εφαρμογή στη μελέτη πολλών μεγεθών και δίνουν μια καλή εικόνα της διαχρονικής τους συμπεριφοράς, καθώς και ικανοποιητικά αποτελέσματα στην πρόβλεψη των μελλοντικών τιμών του μεγέθους. Στα μοντέλα αυτά οι προβλέψεις βασίζονται αποκλειστικά στις προηγούμενες τιμές και στα πρότυπα συμπεριφοράς που εμφανίζονται στην υπό εξέταση της χρονοσειράς.

Πλήρης περιγραφή και μελέτη των μοντέλων *ARIMA* έχει γίνει από τους G. Box, G. Jenkins 1970 [\[10\]](#) των οποίων τα ονόματα έχουν ταυτιστεί με τα μοντέλα αυτά. Χρησιμοποίησαν δεδομένα από κλίβανο αερίου για να εφαρμόσουν το μοντέλο τους. Σύμφωνα με τους Box, Jenkins τα στάδια της προσέγγισης τους είναι τρία. Αρχικά έχουμε την αναγνώριση και την επιλογή του μοντέλου αφού πρώτα επιβεβαιώσουμε ότι οι μεταβλητές είναι στατικές. Στην συνέχεια γίνεται υπολογισμός των παραμέτρων με την χρήση προγραμματιστικών αλγορίθμων και τέλος γίνεται έλεγχος του μοντέλου βάση αποτελεσμάτων. Η προσέγγιση των Box, Jenkins χαρακτηρίστηκε ως προβληματική από τους Commandeur, Koopman το 2007 [\[11\]](#), οι οποίοι υποστήριξαν ότι στο οικονομικό και κοινωνικό πεδίο τα δεδομένα δεν είναι ποτέ στατικά. Επομένως ο αναλυτής πρέπει πρώτα να αναρωτηθεί μέχρι πόσο μπορεί να κυμαίνεται η μεταβλητότητα των δεδομένων ώστε να θεωρούνται στατικά και στη συνέχεια να ακολουθήσει τα επόμενα στάδια.

ΚΕΦΑΛΑΙΟ 3: ΜΕΘΟΔΟΙ ΣΥΝΑΘΡΟΙΣΗΣ ΔΕΔΟΜΕΝΩΝ (Aggregation)

3.1 ΜΕΘΟΔΟΣ ΣΥΝΑΘΡΟΙΣΗΣ

Η συνάθροιση δεδομένων (*aggregation*) είναι μια διαδικασία την οποία ακολουθούμε προκειμένου να αποφύγουμε την ασυνέχεια των δεδομένων διακοπτόμενης ζήτησης αλλά και προκειμένου να αναδείξουμε τα ποιοτικά χαρακτηριστικά της χρονοσειράς μας στα διαφορετικά επίπεδα χρόνου. Υπάρχουν τρία είδη συνάθροισης δεδομένων:

1. ιεραρχική συνάθροιση (*cross-sectional or contemporaneous aggregation*)
2. χρονική συνάθροιση (*temporal aggregation or flow scheme*)
3. forecast horizon aggregation

Βασική παράμετρο της διαδικασίας αυτής αποτελεί το επίπεδο συνάθροισης (*aggregation level*) που θα επιλέξουμε. Επιλέγουμε κάθε φορά τέτοιο επίπεδο συνάθροισης που αναδεικνύει το ποιοτικό χαρακτηριστικό, της χρονοσειράς, που επιθυμούμε. Στην περίπτωση που το επίπεδο συνάθροισης k δεν χωράει ακριβώς στο πλήθος n των παρατηρήσεων τότε αφαιρούμε τις παρατηρήσεις που μένουν και κάνουμε τις προβλέψεις με τις υπόλοιπες παρατηρήσεις. Οι παρατηρήσεις που αφαιρούνται είναι οι παλαιότερες χρονολογικά, καθώς είναι αυτές των οποίων η απώλεια δεν θα επηρεάσει σε μεγάλο βαθμό τις προβλέψεις μας.

Το 2005 οι G. Zotteri και M. Kalchschmidt [\[12\]](#) μελέτησαν την επίδραση που έχει το επίπεδο συνάθροισης στην επίδοση των προβλέψεων καθώς και τον τρόπο με τον οποίο επιλέγεται το κατάλληλο επίπεδο. Τα αποτελέσματα δείχνουν ότι η αλλαγή του επιπέδου μπορεί να έχει εξίσου θετικές αλλά και αρνητικές συνέπειες στην επίδοση της εκάστοτε μεθόδου. Σύμφωνα με τους G. Zotteri και M. Kalchschmidt δεν υπάρχει κάποιος συγκεκριμένος τρόπος επιλογής του κατάλληλου επιπέδου συνάθροισης αλλά αντίθετα θα πρέπει κάθε φορά να μελετούνται διάφοροι παράγοντες που χαρακτηρίζουν και επηρεάζουν τα δεδομένα μας. Συγκεκριμένα παράγοντες όπως το πλήθος των διαθέσιμων πληροφοριών καθώς και η ανομοιογένεια που παρουσιάζουν πρέπει να ληφθούν σοβαρά υπόψη στην επιλογή του επιπέδου. Αποδεικνύουν ότι όσο μεγαλύτερη είναι η ανομοιογένεια της ζήτησης τόσο πιο δύσκολο είναι να εφαρμοστεί η μέθοδος της συνάθροισης. Παρουσιάζουν μια μέθοδο η οποία συνδυάζει τα δεδομένα με βάση τον βαθμό των ομοιοτήτων που παρουσιάζουν. Τα αποτελέσματα της μεθόδου ήταν αρκετά καλά και παρόλο που σχεδιάστηκε για τα δεδομένα της συγκεκριμένης έρευνας, πιστεύουν ότι μπορεί να γενικευτεί και να χρησιμοποιηθεί και σε άλλες αντίστοιχες περιπτώσεις.

3.2 ΙΕΡΑΡΧΙΚΗ ΣΥΝΑΘΡΟΙΣΗ (CROSS-SECTIONAL/ CONTEMPORANEOUS AGGREGATION)

Με την τεχνική της ιεραρχικής συνάθροισης (*cross-sectional aggregation*) πραγματοποιούμε τις προβλέψεις μας ξεχωριστά για κάθε χρονοσειρά και στην συνέχεια τις συνδυάζουμε για να την παραγωγή μιας τελικής πρόβλεψης (N. Kourentzes και F. Petropoulos 2014 [13]). Παράδειγμα αυτής της μεθόδου είναι η παραγωγή ξεχωριστών προβλέψεων για κάποια παρεμφερή προϊόντα και στην συνέχεια ο συνδυασμός τους για την δημιουργία μιας συνολικής πρόβλεψης .

Η τεχνική της ιεραρχικής συνάθροισης μπορεί να εφαρμοστεί στην περίπτωση όπου έχουμε ιεραρχική πρόβλεψη (*Hierarchical Forecasting*), η οποία αποτελείται από δυο ξεχωριστές διαδικασίες πρόβλεψης: την *top-down* και την *bottom-up*. Το 2009 οι G. Athanasopoulos και R.A. Ahmed [14] μελέτησαν την ιεραρχική φύση των χρονοσειρών της τουριστικής ζήτησης και παράγαν βραχυπρόθεσμες προβλέψεις για τον εγχώριο τουρισμό στην Αυστραλία. Τα δεδομένα και οι προβλέψεις τους είναι οργανωμένα σε ιεραρχία με βάση την γεωγραφική περιοχή και τον σκοπό του ταξιδιού και η ανάλυση έγινε με τη χρήση πέντε μορφών ιεραρχικής πρόβλεψης. Οι δύο ακολουθούν την τεχνική *top-down*, μια την τεχνική *bottom-up*, μια νέας μορφής *top-down* όπου οι *top-level* προβλέψεις χωρίζονται σύμφωνα με τις προβλεπόμενες αναλογίες των *lower-level* σειρών και τέλος ένας συνδυασμός των παραπάνω. Χρησιμοποιώντας τις μεθόδους αυτές πραγματοποιήθηκαν λεπτομερές προβλέψεις για τον εγχώριο τουρισμό στην Αυστραλία. Η αξιολόγηση της απόδοσης των προβλέψεων έδειξε ότι οι δύο τελευταίες μέθοδοι ήταν οι πιο αποδοτικές για τις ιεραρχίες τουρισμού που χρησιμοποιήθηκαν. Το κυριότερο πλεονέκτημα των δύο νέων προσεγγίσεων είναι η σύλληψη των διαφόρων χαρακτηριστικών των σειρών. Για την αξιολόγηση των επιδόσεων των νέων μεθόδων πραγματοποιήθηκε μια *out of sample* αξιολόγηση και το τεστ *DieHold-Mariano*. Τα αποτελέσματα των προβλέψεων των δυο τελευταίων μεθόδων ήταν αρκετά καλά και η σύγκρισή τους δεν έδειξε κάποια να υπερτερεί της άλλης.

Όσον αφορά την επιλογή μεταξύ *top-down* και *bottom-up* έχουν γίνει αρκετές μελέτες και τόσο οι απόψεις όσο και τα αποτελέσματα διίστανται για το ποια από τις δύο είναι καλύτερη. Πολλοί είναι εκείνοι που έχουν επιχειρηματολογήσει υπέρ της *bottom-up* όπως επίσης πολλοί είναι και εκείνοι που απέδειξαν ότι η *top-down* δίνει καλύτερα αποτελέσματα. Κάποιες έρευνες μάλιστα απέδειξαν ότι καμιά δεν είναι καλύτερη από την άλλη αλλά η επιλογή τους εξαρτάται από τα δεδομένα που έχουμε και από τα αποτελέσματα τα οποία θέλουμε να πετύχουμε.

Το 1976 οι D. Dunn και W. Williams [15] απέδειξαν την υπεροχή της *bottom-up* σε σχέση με την *top-down* μέθοδο σε δεδομένα διακοπόμενης ζήτησης καταλήγοντας σε πιο ακριβείς προβλέψεις σχετικά με την τηλεφωνική ζήτηση. Το τελικό τους συμπέρασμα ήταν ότι τα αποτελέσματα των προβλέψεων μας θα είναι πιο ακριβή και πιο σωστά όταν χρησιμοποιούμε την *bottom-up* ειδικά σε περιπτώσεις όπου τα δεδομένα ποικίλουν και έχουν μεγάλες διαφορές μεταξύ τους.

Το 1989 ο D. F. D'Attilio [16] επιχειρηματολόγησε υπέρ της μεθόδου *top-down*, σε σχέση με την *bottom-up*. Θεωρεί ότι η μέθοδος *bottom-up* είναι μια καλή επιλογή για να πραγματοποιήσει κανείς τις προβλέψεις του και έχει αρκετά καλά αποτελέσματα, ωστόσο δεν δίνει αρκετή σημασία στην τάση που μπορεί να χαρακτηρίζει τα δεδομένα. Κατά τον D. F.

D'Attilio η τάση παίζει πολύ σημαντικό ρόλο στις προβλέψεις και μπορεί να έχει μεγάλη επίδραση στα αποτελέσματα της μεθόδου αλλά και στην αξιολόγηση της ποιότητας των προβλέψεων.

Το 1988 οι A.B. Schwarzkopf και R.J. Tersine [17] έκαναν μια έρευνα σχετικά με το ποια μέθοδο είναι καλύτερη από τις δύο. Για να το πετύχουν αυτό πραγματοποίησαν τις προβλέψεις τους και με τις δύο μεθόδους και στην συνέχεια αξιολόγησαν την συνολική απόδοση των προβλέψεων. Η συνολική απόδοση εξαρτάται από τρεις κύριους παράγοντες, την ακρίβεια, την προκατάληψη και την ευρωστία. Όσον αφορά την ακρίβεια των προβλέψεων, η μέθοδος *top-down* παρουσίασε καλύτερα αποτελέσματα αλλά ήταν χειρότερη ως προς την προκατάληψη σε σχέση με την *bottom-up*. Στο θέμα της ευρωστίας η *top-down* μειώνει την επίδραση του τυχαίου λάθους στη πρόβλεψη αλλά εισάγει μια περίπλοκη αλληλεπίδραση μεταξύ της προκατάληψης και των ακραίων γεγονότων. Το συμπέρασμα λοιπόν των A.B. Schwarzkopf και R.J. Tersine είναι πως οι δύο μέθοδοι μπορούν να έχουν επιθυμητά αποτελέσματα όταν χρησιμοποιηθούν σε κατάλληλες συνθήκες. Η *top-down* για παράδειγμα είναι απαραίτητη όταν θέλουμε να διακρίνουμε συγκεκριμένα μοτίβα στην ζήτηση μεμονωμένων αγαθών.

Το 2011 οι R.J. Hyndman και R.A. Ahmed [18] πρότειναν μια άλλη μέθοδο προβλέψεων για ιεραρχικές χρονοσειρές, πιο αποτελεσματική από τις *top-down* και *bottom-up*. Η μέθοδος τους ονομάζεται *optimal* και βασίζεται σε ανεξάρτητες προβλέψεις σε όλα τα επίπεδα ιεραρχίας και στην συνέχεια χρησιμοποιείται κάποιο μοντέλο παλινδρόμησης για να συνδυάσει βέλτιστα και να συμφιλιώσει τις προβλέψεις αυτές. Η μέθοδος αυτή παρέχει σημείο σε σημείο πρόβλεψη οι οποίες συμφιλιώνονται σε όλα τα επίπεδα ιεραρχίας, επιτρέπει τους συσχετισμούς και τις αλληλεπιδράσεις μεταξύ των σειρών σε κάθε επίπεδο της ιεραρχίας και είναι αρκετά ευέλικτη σε σχέση με τις *top-down* και *bottom-up*. Για την πραγματοποίηση της έρευνας χρησιμοποιήθηκαν δεδομένα τουριστικής ζήτησης στην Αυστραλία τα οποία ήταν χωρισμένα ανάλογα με την περιοχή και το σκοπό του ταξιδιού. Ένα θετικό της μεθόδου αυτής είναι ότι δεν μας δεσμεύει στην επιλογή του μοντέλου πρόβλεψης για την πραγματοποίηση των αρχικών προβλέψεων.

3.3 ΧΡΟΝΙΚΗ ΣΥΝΑΘΡΟΙΣΗ (TEMPORAL AGGREGATION OR FLOW SCHEME)

Η δεύτερη μορφή συνάθροισης είναι η χρονική συνάθροιση δεδομένων (*temporal aggregation or flow scheme*), όπου μετατρέπουμε μια υψηλής συχνότητας χρονοσειρά σε μια χαμηλής συχνότητας και χωρίζεται σε δυο κατηγορίες, την *nonoverlapping* και την *overlapping*. Στην πρώτη περίπτωση, οι χρονοσειρές χωρίζονται σε διαδοχικές μη επικαλυπτόμενες χρονικές περιοχές όπου το μήκος της κάθε περιοχής ισούται με το επίπεδο συνάθροισης. Αντίθετα η δεύτερη περίπτωση είναι παρόμοια με την τεχνική του κινούμενου παραθύρου όπου το μέγεθος του παραθύρου είναι ίσο με το επίπεδο συνάθροισης. Σε κάθε περίοδο, το παράθυρο κινείται ένα βήμα μπροστά, έτσι ώστε η παλαιότερη παρατήρηση να βγει εκτός και να συμπεριληφθεί η νεότερη.

3.3.1 ΚΛΑΣΣΙΚΗ ΧΡΟΝΙΚΗ ΣΥΝΑΘΡΟΙΣΗ ΜΗ ΕΠΙΚΑΛΥΠΤΟΜΕΝΗ (CLASSIC TEMPORAL AGGREGATION NONOVERLAPPING)

Χρησιμοποιώντας την τεχνική της χρονικής συνάθροισης (*temporal aggregation*) μετατρέπουμε μια υψηλής συχνότητας χρονοσειρά σε μια χαμηλής συχνότητας (Κ. Nikolopoulos και Α.Α. Syntetos 2011 [19]). Με την τεχνική αυτή έχουμε καλύτερη εκτίμηση των χαρακτηριστικών κάθε χρονοσειράς καθώς και μείωση της επίδρασης του μοντέλου πρόβλεψης που χρησιμοποιούμε. Σε χαμηλά επίπεδα συνάθροισης φαίνονται περισσότερο περιοδικά χαρακτηριστικά της χρονοσειράς, όπως η εποχικότητα. Όσο αυξάνονται τα επίπεδα συνάθροισης δίνουμε περισσότερη βάση σε χαμηλόσυχνα χαρακτηριστικά της χρονοσειράς, όπως στο επίπεδο και στη τάση της χρονοσειράς.

Σκοπός μας είναι η ταυτοποίηση, η επιλογή και ο υπολογισμός των παραμέτρων των μοντέλων πρόβλεψης. Αφού γίνει η χρονική συνάθροιση, παράγουμε προβλέψεις για κάθε επίπεδο συνάθροισης χρησιμοποιώντας το κατάλληλο μοντέλο πρόβλεψης.

Ο Β. Abraham το 1982 [20] χρησιμοποίησε την τεχνική της χρονικής συνάθροισης (*temporal aggregation*) για να πραγματοποιήσει κάποιες προβλέψεις, θέλοντας να δει την επίδραση της στην αποτελεσματικότητα των προβλέψεων αλλά και την επίδραση των αρχικών δεδομένων που χρησιμοποιούμε. Τα συμπεράσματα του δεν μπορεί να είναι απόλυτα καθώς χρησιμοποίησε μικρό πλήθος χρονοσειρών, ωστόσο κατέληξε στο συμπέρασμα ότι η χρονική συνάθροιση επιδρά θετικά στην αποτελεσματικότητα του μοντέλου πρόβλεψης (ARIMA).

Το 2012 οι Μ.Ζ. Babai και Μ.Μ. Ali [21] εφάρμοσαν την μέθοδο της χρονικής συνάθροισης σε δεδομένα με διακοπτόμενη ζήτηση. Η διαχείριση της διακοπτόμενης ζήτησης και η πραγματοποίηση των προβλέψεων είναι αρκετά δύσκολη λόγω της παρουσίας διαστημάτων μηδενικών τιμών. Ένας τρόπος διαχείρισης της δυσκολίας αυτής και της πραγματοποίησης των προβλέψεων είναι η χρήση της μεθόδου συνάθροισης. Τα οφέλη της μεθόδου προέκυψαν εμπειρικά από τις προβλέψεις που πραγματοποιήθηκαν για δεδομένα της Royal Air Force (UK). Τα αποτελέσματα δείχνουν ότι η χρήση της συνάθροισης οδηγεί σε μεγαλύτερη ακρίβεια από την κλασική μέθοδο πρόβλεψης. Με την χρήση της συνάθροισης μειώθηκαν ή εξαφανίστηκαν τελείως οι περίοδοι μηδενικών τιμών, περισσότερα μοντέλα προβλέψεων έγιναν διαθέσιμα για χρήση και δεν υπάρχει λόγος να ξαναγίνει χωρισμός των προβλέψεων γιατί στην συγκεκριμένη περίπτωση ενδιαφερόμαστε για την συνολική πρόβλεψη.

3.3.2 ΚΛΑΣΣΙΚΗ ΧΡΟΝΙΚΗ ΣΥΝΑΘΡΟΙΣΗ ΕΠΙΚΑΛΥΠΤΟΜΕΝΗ (CLASSIC TEMPORAL AGGREGATION OVERLAPPING)

Όσον αφορά την τεχνική της επικαλυπτόμενης χρονικής συνάθροισης (*overlapping temporal aggregation*) έχουν γίνει λίγες μελέτες για την επίδραση της. Το 1992 οι K.H. Luiz και A.M. Pedro [22] ανέλυσαν τις επιδράσεις της σε μοντέλα *ARIMA* και κατέληξαν στο συμπέρασμα ότι το μοντέλο από *ARIMA(p,d,q)* μετατρέπεται σε *ARIMA(P,d,Q)*. Αντίστοιχες μελέτες έγιναν και από τους K. Brannas και J. Hellstrom το 2002 [23] σε μοντέλα *INARMA(1, 0, 0)* που παρέμειναν τα ίδια και από τους M. Mohammadpour και J.E. Boylan το 2012 [24] σε μοντέλα *INARMA(p, 0, q)* όπου και αυτά παρέμειναν τα ίδια.

3.3.3 MAPA (MULTIPLE AGGREGATION PREDICTION ALGORITHM)

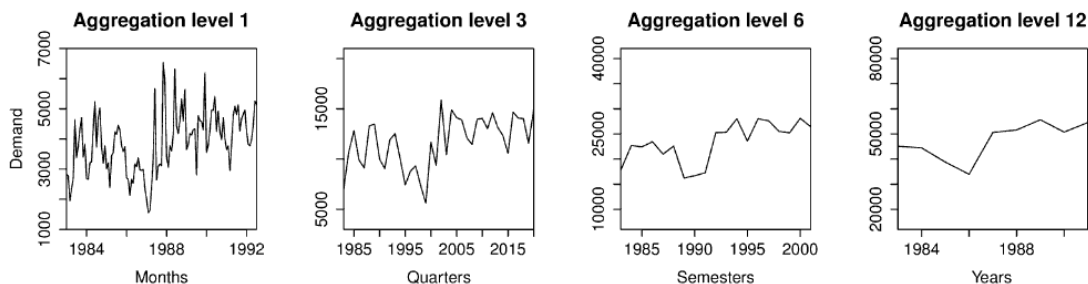
Η μεθοδολογία *MAPA* χρησιμοποιεί την μέθοδο της συνάθροισης και τον συνδυασμό των προβλέψεων. Έχει προταθεί από τους N. Kourentzes και F. Petropoulos 2014 [13] και [25]. Η φιλοσοφία της μεθόδου είναι να χρησιμοποιήσουμε πολλά επίπεδα συνάθροισης για την δοσμένη χρονοσειρά, τονίζοντας κάθε φορά διαφορετικό ποιοτικό χαρακτηριστικό της, και να πραγματοποιήσουμε τις προβλέψεις μας για το κάθε επίπεδο συνάθροισης χρησιμοποιώντας κατάλληλη μέθοδο πρόβλεψης. Σε χαμηλά επίπεδα συνάθροισης τονίζονται χαρακτηριστικά όπως η εποχικότητα ενώ όσο αυξάνεται το επίπεδο τονίζονται χαρακτηριστικά όπως η τάση και το επίπεδο της χρονοσειράς.

Στην συνέχεια, αντί να συνδυάσουμε απευθείας τις προβλέψεις μας από τα διαφορετικά επίπεδα συνάθροισης, *σπάμε* τις χρονοσειρές ανάλογα με τα ποιοτικά χαρακτηριστικά τους και τα συνδυάζουμε. Στο σημείο αυτό πρέπει να προσέχουμε ο συνδυασμός που θα γίνει να περιέχει τα ποιοτικά χαρακτηριστικά που θέλουμε και η επιλογή του επιπέδου συνάθροισης να μην οδηγεί σε απώλεια δεδομένων. Εάν συνδυάσω για παράδειγμα μια εποχικοποιημένη πρόβλεψη με μια μη εποχικοποιημένη θα καταλήξω σε μια πρόβλεψη που θα έχει την μισή εποχικότητα από την αρχική. Με την μέθοδο αυτή καταφέρνουμε:

- μείωση της επίδρασης της μεθόδου που χρησιμοποιούμε για την πρόβλεψη
- συνδυάζει τα πλεονεκτήματα ακριβείας και από την συνάθροιση και από την συνδυασμένη πρόβλεψη
- είναι πιο ακριβής

Η εφαρμογή της μεθόδου είχε αρκετά πλεονεκτήματα στις μεθόδους *Naive*, *SES* και *Theta* ωστόσο δεν έχει μελετηθεί πλήρως η επίπτωση της στα μοντέλα εκθετικής εξομάλυνσης. Μέχρι ένα σημείο είναι αρκετά κερδοφόρα καθώς βελτιώνει την τυχαιότητα και τις μακροχρόνιες χρονοσειρές. Κυριότερο θέμα είναι η σωστή επιλογή του επιπέδου της συνάθροισης καθώς δεν υπάρχει κάποιο οικουμενικό επίπεδο το οποίο να δρα θετικά σε οποιαδήποτε χρονοσειρά.

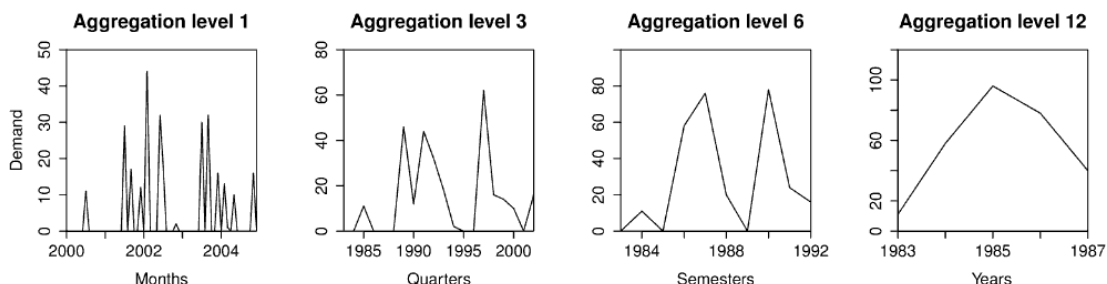
Fast-moving demand



Εικόνα 3 Γρήγορα Μεταβαλλόμενη Ζήτηση

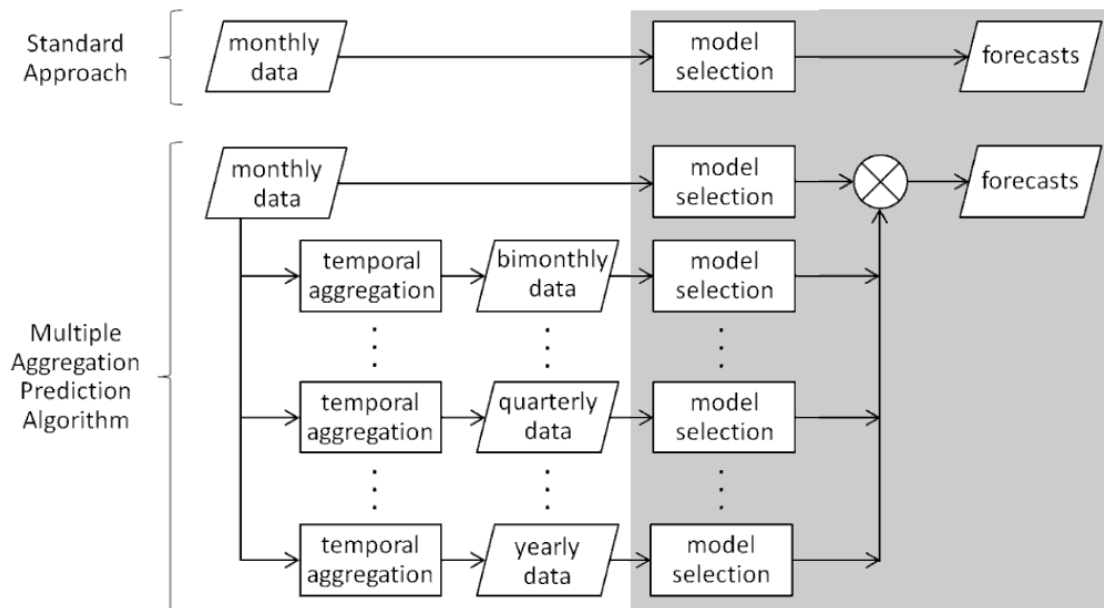
Το παραπάνω διάγραμμα δείχνει πως επηρεάζεται η χρονοσειρά (*fast moving demand*) καθώς αυξάνω κάθε φορά το επίπεδο συνάθροισης, μειώνοντας έτσι την συχνότητα. Η αρχική χρονοσειρά έχει εμφανές το στοιχείο της εποχικότητας ενώ η τελική, με επίπεδο συνάθροισης 12, έχει μια μετατόπιση επιπέδου και μια ασθενή τάση. Το στοιχείο της εποχικότητας δεν είναι πλέον τόσο εμφανές.

Slow-moving demand



Εικόνα 4 Αργά Μεταβαλλόμενη Ζήτηση

Αντίστοιχα και σε αυτό το διάγραμμα φαίνεται η σημασία του επιπέδου συνάθροισης για την χρονοσειρά (*slow moving demand*). Ιδιαίτερα στην περίπτωση αυτή, όπου έχουμε διακοπτόμενη ζήτηση, η αλλαγή επιπέδου συνάθροισης οδηγεί σε μείωση εξαφάνιση των περιόδων μηδενικής ζήτησης οπότε μπορούμε πλέον να εφαρμόσουμε συμβατικές μεθόδους πρόβλεψης.



Εικόνα 5 Μεθοδολογία MAPA σε παράλληλη σύγκριση με την απλή προσέγγιση

Η μεθοδολογία της MAPA φαίνεται στο παραπάνω διάγραμμα σε παράλληλη σύγκριση με την απλή μέθοδο. Όπως έχει προαναφερθεί γίνεται ο συνδυασμός των προβλέψεων για κάθε επίπεδο συνάθροισης χρησιμοποιώντας κάθε φορά το μοντέλο πρόβλεψης που κρίνουμε καταλληλότερο.

Η μεθοδολογία όπως αναφέρθηκε και παραπάνω αποτελείται από τρία στάδια: την συνάθροιση, την πρόβλεψη και τον συνδυασμό των αποτελεσμάτων. Όσον αφορά το κομμάτι της συνάθροισης μεγάλη προσοχή χρειάζεται γενικά στην επιλογή του επιπέδου συνάθροισης. Ωστόσο στην μεθοδολογία αυτή δεν μας επηρεάζει καθώς η εφαρμογή της απαιτεί η συνάθροιση να γίνεται σε κάθε επίπεδο ξεχωριστά. Στην συνέχεια επιλέγουμε κατάλληλο μοντέλο πρόβλεψης για κάθε επίπεδο συνάθροισης και πραγματοποιούμε τις προβλέψεις μας. Το τελικό στάδιο της μεθοδολογίας, ο συνδυασμός δηλαδή των προβλέψεων, είναι λίγο πιο πολύπλοκος καθώς προτού συνδυάσουμε τις προβλέψεις μας πρέπει πρώτα να τις επαναφέρουμε στο αρχικό επίπεδο συνάθροισης με την χρήση του κατάλληλου συντελεστή. Έχουν γίνει αρκετές προσπάθειες για να βρεθεί ο καταλληλότερος συντελεστής, ωστόσο έχει παρατηρηθεί πειραματικά ότι κρατώντας απλά το μέσο όρο των προβλέψεων είναι αρκετό και επιφέρει αρκετά αξιόπιστα και ακριβή αποτελέσματα.

Ένα πρόβλημα που παρουσιάζεται εφαρμόζοντας το αλγόριθμο MAPA είναι η περίπτωση της εποχικότητας. Είδαμε παραπάνω στην χρονοσειρά με επίπεδο συνάθροισης 1 ότι το στοιχείο της εποχικότητας είναι αρκετά εμφανές, ενώ χρησιμοποιώντας επίπεδο συνάθροισης 12 χάνεται το στοιχείο της εποχικότητας. Είναι λοιπόν λογικό ότι αν στο τέλος αθροίσουμε τις προβλέψεις μας θα προκύψουν δεδομένα στα οποία θα απουσιάζει το στοιχείο της εποχικότητας. Για να αποφευχθεί το πρόβλημα αυτό πρέπει ο συνδυασμός που θα κάνουμε να γίνει στα ποιοτικά χαρακτηριστικά της κάθε χρονοσειράς και όχι στις προβλέψεις.

ΚΕΦΑΛΑΙΟ 4: ΑΚΡΙΒΕΙΑ ΤΩΝ ΠΡΟΒΛΕΨΕΩΝ

4.1 ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

Η γνώση μιας μεθόδου πρόβλεψης δεν είναι αρκετή για την πραγματοποίηση των προβλέψεων μας σε μια χρονοσειρά. Πρέπει πρώτα να επιλέξουμε την καταλληλότερη μέθοδο πρόβλεψης και στην συνέχεια να προχωρήσουμε στην πραγματοποίηση των προβλέψεων. Ένα από τα σημαντικότερα κριτήρια για την επιλογή μιας μεθόδου είναι η επίδοση της στην στατιστική ανάλυση. Υπάρχουν δύο είδη στατιστικών αναλύσεων, η βασική στατιστική ανάλυση η οποία περιλαμβάνει βασικούς στατιστικούς δείκτες και δεν θα μας απασχολήσει στην συγκεκριμένη διπλωματική και η στατιστική ανάλυση ακρίβειας προβλέψεων η οποία πραγματοποιείται χρησιμοποιώντας διάφορους δείκτες που αναλύονται παρακάτω.

4.2 ΣΤΑΤΙΣΤΙΚΟΙ ΔΕΙΚΤΕΣ ΣΦΑΛΜΑΤΩΝ

Για την στατιστική ανάλυση των προβλέψεων χρησιμοποιούμε κάποιους στατιστικούς δείκτες σφαλμάτων. Γενικά το σφάλμα σε μια πρόβλεψη είναι η διαφορά των πραγματικών τιμών από τις προβλεπόμενες. Η αξιολόγηση της μεθόδου όσον αφορά τα σφάλματα μπορεί να γίνει είτε *in-sample*, δηλαδή σε παρατηρήσεις που ήδη διαθέτουμε, είτε *out-sample*, δηλαδή πραγματοποιούμε την πρόβλεψη και αφού γίνουν σε εμάς γνωστά τα πραγματικά δεδομένα τότε ελέγχουμε τα σφάλματα. Συνήθως χρησιμοποιείται η *out-sample* αξιολόγηση γιατί μας ενδιαφέρει περισσότερο η επίδοση τις πρόβλεψης στα νέα δεδομένα.

Οι κυριότεροι δείκτες σφαλμάτων που χρησιμοποιούνται για την αξιολόγηση μια μεθόδου πρόβλεψης αναλύονται στις επόμενες ενότητες του κεφαλαίου.

4.2.1 ΜΕΣΟ ΣΦΑΛΜΑ (*Mean Error*)

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

Εκφράζει ένα μέτρο της συστηματικότητας του σφάλματος και υπολογίζεται εύκολα από τον παραπάνω τύπο. Ο δείκτης αυτός αναφέρεται στην βιβλιογραφία ως *bias* και δηλώνει την προκατάληψη της μεθόδου ανάλογα με το πρόσημο του δείκτη. Για αρνητικές τιμές του δείκτη

έχουμε αισιοδοξία της μεθόδου πρόβλεψης αφού οι προβλεπόμενες τιμές είναι μεγαλύτερες των πραγματικών ενώ οι θετικές τιμές του δείκτη δηλώνουν απαισιοδοξία στις προβλέψεις αφού οι πραγματικές τιμές είναι μεγαλύτερες των προβλεπόμενων.

4.2.2 ΜΕΣΟ ΑΠΟΛΥΤΟ ΣΦΑΛΜΑ (*Mean Absolute Error*)

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

Ο δείκτης αυτός εκφράζει ένα μέτρο της ακρίβειας της πρόβλεψης χωρίς να λαμβάνει υπόψιν την προκατάληψη της μεθόδου πρόβλεψης. Η ακρίβεια της πρόβλεψης χειροτερεύει με την αύξηση της τιμής του δείκτη.

4.2.3 ΜΕΣΟ ΤΕΤΡΑΓΩΝΙΚΟ ΣΦΑΛΜΑ (*Mean Squared Error*)

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

Όπως και το μέσο απόλυτο σφάλμα έτσι και το μέσο τετραγωνικό σφάλμα εκφράζει ένα μέτρο της ακρίβειας της πρόβλεψης χωρίς να λαμβάνει υπόψιν την προκατάληψη της μεθόδου πρόβλεψης. Η διαφορά των δύο δεικτών φαίνεται στον τύπο υπολογισμού τους. Το μέσο τετραγωνικό σφάλμα λόγω του τετραγώνου που υπάρχει στον τύπο δίνει πολύ περισσότερη

έμφαση στα μεγάλα σφάλματα και πολύ μικρότερη στα μικρά. Το μέσο τετραγωνικό σφάλμα χρησιμοποιείται ευρέως σε περιπτώσεις όπου θέλουμε να κάνουμε εξομάλυνση των χρονοσειρών μας και πρέπει να υπολογίσουμε τις βέλτιστες παραμέτρους.

4.2.4 ΡΙΖΑ ΜΕΣΟΥ ΤΕΤΡΑΓΩΝΙΚΟΥ ΣΦΑΛΜΑΤΟΣ (*Root Mean Squared Error*)

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

Ο δείκτης αυτός υπολογίζεται ως η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος και έχει τις ίδιες ιδιότητες με αυτόν. Η διαφορά του έγκειται στο γεγονός ότι η ρίζα του μέσου τετραγωνικού σφάλματος είναι εκφρασμένη στις μονάδες της αρχικής χρονοσειράς.

4.2.5 ΜΕΣΟ ΑΠΟΛΥΤΟ ΠΟΣΟΣΤΙΑΙΟ ΣΦΑΛΜΑ (*Mean Absolute Percentage Error*)

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

Ο δείκτης αυτός λαμβάνει τιμές μεγαλύτερες ή ίσες του μηδενός. Όσο μικρότερη η τιμή τόσο καλύτερη η ακρίβεια των προβλέψεων. Η χρήση του δείκτη αυτού αποφεύγεται σε χρονοσειρές διακοπτόμενες ζήτησης γιατί η ύπαρξη πολλών μηδενικών οδηγεί σε απροσδιοριστία. Το μεγάλο πλεονέκτημα του υπολογισμού των σφαλμάτων πρόβλεψης σε

ποσοστιαία μορφή είναι το γεγονός ότι μπορούμε να συγκρίνουμε τα σφάλματα μιας μεθόδου πρόβλεψης που έχει εφαρμοστεί σε περισσότερες χρονοσειρές με διαφορετικό επίπεδο μέσης τιμής. Ο δείκτης είναι επίσης χρήσιμος καθώς διευκολύνει τους υπολογισμούς όταν οι πραγματικές τιμές είναι πολύ υψηλές.

4.2.6 ΣΥΜΜΕΤΡΙΚΟ ΜΕΣΟ ΑΠΟΛΥΤΟ ΠΟΣΟΣΤΙΑΙΟ ΣΦΑΛΜΑ (*Symmetric Mean Absolute Percentage Error*)

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

Το συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα αποτελεί μια παραλλαγή του παραπάνω δείκτη (*MAPE*) με την διαφορά που όπως φαίνεται από τον τύπο έχει πάνω όριο. Ο δείκτης αυτός παίρνει τιμές από 0 έως 200%. Το μειονέκτημα του δείκτη αυτού είναι ότι δεν διαχειρίζεται τις αισιόδοξες και τις απαισιόδοξες προβλέψεις αντίστοιχα. Δεν είναι δηλαδή απόλυτα συμμετρικός όπως δηλώνει το όνομα του.

4.2.7 ΜΕΣΟ ΑΠΟΛΥΤΟ ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΟ ΣΦΑΛΜΑ (*Mean Absolute Scaled Error*)

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

Ο δείκτης αυτός δημιουργήθηκε από τους *Hyndman* και *Koehler* ώστε να αντιμετωπιστούν οι περιπτώσεις απροσδιοριστίας των παραπάνω δεικτών αλλά και με σκοπό να δοθεί ίδια βαρύτητα τόσο στα μικρά όσο και στα μεγάλα σφάλματα. Από τον τύπο παρατηρούμε ότι ο αριθμητής θυμίζει το μέσο απόλυτο σφάλμα ενώ ο παρονομαστής είναι το μέσο απόλυτο σφάλμα όταν χρησιμοποιείται ως μέθοδο πρόβλεψης η *Naïve*. Συμπεραίνουμε λοιπόν από αυτά ότι αν η τιμή του *MAsE* είναι μικρότερη της μονάδας τότε η μέθοδος πρόβλεψης που χρησιμοποιήθηκε έχει κατά μέσο όρο καλύτερη απόδοση από την *Naïve* ενώ αν η τιμή είναι μεγαλύτερη της μονάδας τότε έχει χειρότερη απόδοση.

Η χρήση των παραπάνω στατιστικών δεικτών έγινε το 1982 στον Μ διαγωνισμό από τους *S. Makridakis* και *M. Hibon* [26]. Για τις ανάγκες του διαγωνισμού χρησιμοποιήθηκαν 1001 χρονοσειρές, οι οποίες παρόλο που δεν επιλέχθηκαν τυχαία κάλυπταν αρκετά μεγάλο εύρος περιπτώσεων, από εταιρίες, βιομηχανίες και κράτη. Οι μέθοδοι προβλέψεων που χρησιμοποιήθηκαν στον διαγωνισμό αξιολογήθηκαν από πέντε στατιστικούς δείκτες με σημαντικότερους το *MAPE* και το *MSE*. Τα αποτελέσματα του διαγωνισμού έδειξαν ότι σε μηνιαία δεδομένα, όπου η τυχαιότητα παίζει σημαντικό ρόλο, οι απλές μέθοδοι παράγουν καλύτερα αποτελέσματα από τις εξειδικευμένες και πιο λεπτομερείς μεθόδους. Αντίθετα σε ετήσια δεδομένα οι εξειδικευμένες μέθοδοι αποδίδουν καλύτερα. Όσον αφορά δεδομένα τα οποία χαρακτηρίζονται από εποχικότητα και τα δύο είδη μεθόδων παρήγαγαν αποτελέσματα με εξίσου μεγάλη ακρίβεια.

Το 2000 οι *S. Makridakis* και *M. Hibon* [27] επανέλαβαν την έρευνα τους με τον διαγωνισμό *M3* χρησιμοποιώντας πάλι αντίστοιχους στατιστικούς δείκτες για την αξιολόγηση των μεθόδων που χρησιμοποίησαν. Στον διαγωνισμό αυτόν χρησιμοποιήθηκαν 3003 χρονοσειρές καλύπτοντας ετήσια, τριμηνιαία και μηνιαία δεδομένα ώστε να προκύψουν όσο το δυνατόν πιο αξιόπιστα συμπεράσματα. Στα αποτελέσματα του διαγωνισμού ξεχώρισε η *Theta*, η οποία παρόλη την ευκολία στην χρήση της είχε αρκετά καλή απόδοση. Επίσης αποδείχθηκε ότι, σε περίπτωση ετήσιων δεδομένων η σε περιπτώσεις όπου ισχύουν κάποιες ειδικές συνθήκες, κάποιες νέες μέθοδοι μπορούν να έχουν καλύτερα αποτελέσματα από τις ήδη υπάρχουσες και θα πρέπει να χρησιμοποιούνται.

Παρόμοια διαδικασία ακολούθησαν και οι *G. Athanasopoulos* και *R.J. Hyndman* το 2011 [28] στον διαγωνισμό που πραγματοποίησαν σε τουριστικά δεδομένα. Ο διαγωνισμός περιλάμβανε πάνω από 1300 χρονοσειρές, με ετήσια, τριμηνιαία και μηνιαία δεδομένα. Για την αξιολόγηση των μεθόδων χρησιμοποιήθηκαν και εδώ στατιστικοί δείκτες όπως το *MAPE* και το *MAsE*. Τα αποτελέσματα που προέκυψαν δείχνουν ότι για ετήσια δεδομένα υπερτερεί

η *Naïve* ενώ για μηνιαία δεδομένα πολύ καλή απόδοση είχε τόσο η *Damped* όσο και η *Theta*. Αποδείχθηκε επίσης ότι με βάση τον δείκτη *MAsE* τα μοντέλα *ARIMA* ήταν τα πιο αποδοτικά.

ΚΕΦΑΛΑΙΟ 5: ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

5.1 ΜΕΘΟΔΟΛΟΓΙΑ

Σε παραπάνω κεφάλαια εξηγήσαμε με λεπτομέρεια τις δυο κυριότερες μορφές συνάθροισης δεδομένων, την χρονική και την ιεραρχική. Σκοπός μας εδώ είναι να συνδυάσουμε τις δυο αυτές μορφές ώστε να βελτιώσουμε την απόδοση και την ακρίβεια των προβλέψεων. Έχει αποδειχθεί ότι η ιεραρχική συνάθροιση συμβάλλει στην συμφιλίωση των προβλέψεων και στην βελτίωση της απόδοσης τους. Μπορεί να πραγματοποιηθεί με πολλούς τρόπους, ωστόσο εδώ θα χρησιμοποιήσουμε τους πιο γνωστούς στην βιβλιογραφία *top-down*, *bottom-up* και *optimal*. Αντίστοιχα η χρονική βελτιώνει την ακρίβεια των προβλέψεων και μειώνει την προκατάληψη τους. Για την εφαρμογή της θα χρησιμοποιήσουμε τον αλγόριθμο *MAPA* που έχουμε αναλύσει σε παραπάνω κεφάλαιο. Για την αξιολόγηση των μεθόδων μας χρησιμοποιούμε ως μέθοδο αναφοράς την *Naïve* και την *SES (Simple Exponential Smoothing)*. Στο σημείο αυτό θα εξηγήσουμε αναλυτικά τη μεθοδολογία και τα βήματα που ακολουθήθηκαν καθ' όλη τη διάρκεια της μελέτη καθώς και τους λόγους που επιλέξαμε συγκεκριμένες μεθόδους ή συναρτήσεις στο πρόγραμμα.

5.1.1 ΕΙΣΑΓΩΓΗ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Η πρώτη ενέργεια που κάνουμε για την μελέτη αυτή είναι να εισάγουμε τα δεδομένα που έχουμε στο πρόγραμμα. Συνήθως βρίσκονται σε μορφή αρχείου *txt* ή *csv*. Στην συνέχεια πρέπει να γίνει λεπτομερής ανάλυση αυτών των δεδομένων ώστε να μπορούμε να αποφασίσουμε αργότερα τι μεθόδους θα χρειαστούμε και πως θα κινηθούμε γενικότερα για να εκτελέσουμε την μελέτη. Τι δεδομένα έχουμε; Πόσα στον αριθμό; Ποια είναι τα χαρακτηριστικά τους; Ποια είναι η συχνότητα τους; Πρέπει από την ανάλυση που θα γίνει να είμαστε σε θέση να απαντήσουμε σε τέτοιας μορφής ερωτήματα ώστε να είμαστε σίγουροι ότι έχουμε λάβει υπόψιν μας ότι χρειάζεται για να προχωρήσουμε στην επιλογή κατάλληλων μοντέλων και μεθόδων για την επίλυση τους.

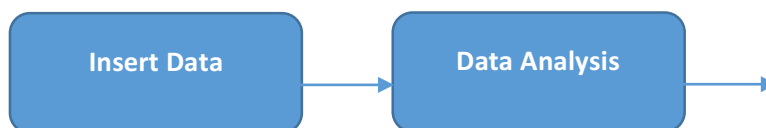
Αφού συγκεντρώσουμε τα δεδομένα πρέπει να τα κατατάξουμε σε ένα ιεραρχικό μοντέλο για να μπορούμε να διακρίνουμε τα επίπεδα και ποια δεδομένα κατατάσσονται σε ποιο επίπεδο. Η σημασία αυτού θα φανεί αργότερα όταν θα θελήσουμε να εφαρμόσουμε κάποια μέθοδο ιεραρχικής συνάθροισης (*top-down*, *bottom-up*, *optimal*). Ο αριθμός των επιπέδων έχει καθοριστική σημασία σε αρκετά σημεία της εργασίας, όπως στον υπολογισμό των συντελεστών που θα χρησιμοποιήσουμε στην *optimal*.

Το είδος των δεδομένων που έχουμε στην διάθεση μας είναι ένας ακόμη παράγοντας καθοριστικής σημασίας. Πρέπει να διακρίνουμε αν τα διαθέσιμα δεδομένα είναι ωριαία, ημερήσια, εβδομαδιαία ή μηνιαία και ετήσια. Το είδος αυτό θα καθορίσει την συχνότητα που θα επιλέξουμε καθώς και τους οριζόντες πρόβλεψης που θα χρησιμοποιήσουμε αργότερα για να παράγουμε τις προβλέψεις.

Αναπαριστώντας τις παρατηρήσεις που έχουμε γραφικά πρέπει να είμαστε σε θέση να διακρίνουμε τα ποιοτικά τους χαρακτηριστικά. Αν υπάρχει τυχαιότητα, εποχικότητα και τάση πρέπει να το γνωρίζουμε ώστε να κάνουμε αν χρειαστεί αποσύνθεση των δεδομένων και να προκύψουν οι αντίστοιχοι δείκτες.

Τέλος παρατηρώντας τις τιμές που έχουν οι παρατηρήσεις πρέπει να είμαστε σε θέση να διακρίνουμε το χρονικό εύρος στο οποίο είναι πιθανό να έχει δημιουργηθεί η τιμή αυτή. Αν για παράδειγμα έχουμε στην διάθεση μας μια πολύ υψηλή τιμή κατανάλωσης ενέργειας μπορούμε να θεωρήσουμε ότι πρόκειται για κάποια πρωινή στιγμή μιας καθημερινής εργάσιμης ημέρας καθώς μελετάμε ένα σύμπλεγμα κτιρίων τράπεζας. Είναι πολύ λογικό για παράδειγμα το πρωί μιας Δευτέρας η τιμή να είναι αρκετά μεγάλη καθώς υπάρχει μεγάλη κατανάλωση μετά από ένα Σαββατοκύριακο.

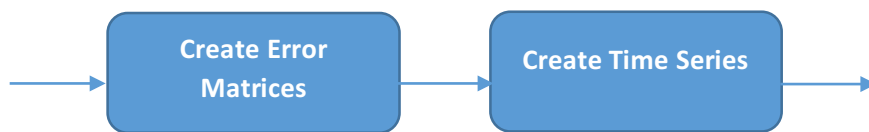
Η ανάλυση των δεδομένων είναι πολύ σημαντική πριν από κάθε μελέτη και κρίνεται αναγκαία καθώς βασικές επιλογές που θα γίνουν στην αρχή και κατά την διάρκεια της μελέτης εξαρτώνται από την ανάλυση αυτήν. Η λάθος επιλογή μιας μεθόδου λόγω ελλιπής ανάλυσης μπορεί να επιφέρει λάθος ή μη επιθυμητά αποτελέσματα.



5.1.2 ΔΗΜΙΟΥΡΓΙΑ ΠΙΝΑΚΩΝ ΚΑΙ ΜΕΤΑΤΡΟΠΗ ΔΕΔΟΜΕΝΩΝ

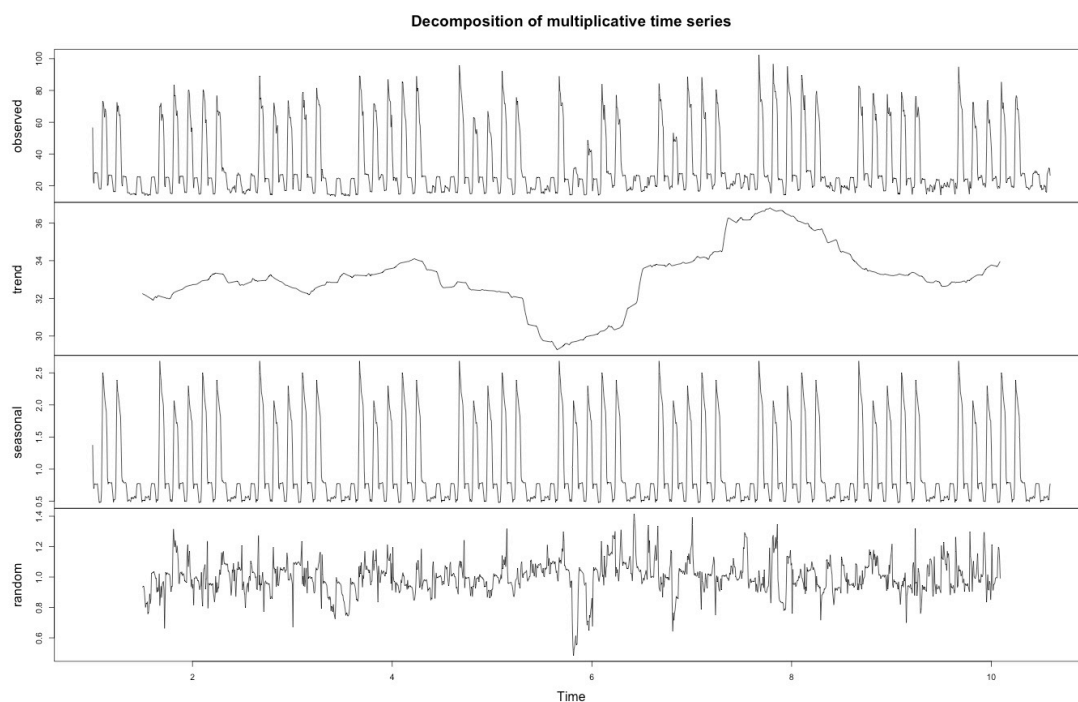
Αφού ολοκληρώσουμε την ανάλυση των δεδομένων μας μπορούμε να προχωρήσουμε στην δημιουργία πινάκων για να συγκεντρώσουμε τα σφάλματα των μεθόδων. Δημιουργούμε όσους πίνακες χρειαζόμαστε, τους οποίους αργότερα θα γεμίζουμε, όσο τρέχει το πρόγραμμα, με τα σφάλματα που θα χρησιμοποιήσουμε για την σύγκριση των μεθόδων. Η δημιουργία γίνεται στην αρχή του προγράμματος και πρέπει προηγουμένως να έχουμε καθορίσει τους δείκτες των σφαλμάτων που θα χρησιμοποιήσουμε καθώς και τον αριθμό των επαναλήψεων του προγράμματος. Εδώ πρέπει να προσέξουμε το μέγεθος των πινάκων να είναι όσος είναι ο αριθμός των σφαλμάτων επι τους οριζόντες πρόβλεψης που έχουμε ορίσει διαφορετικά υπάρχει πιθανότητα να χάσουμε δεδομένα ή σε περίπτωση που δημιουργήσουμε μεγαλύτερους πίνακες το πρόγραμμα θα βάλει από μόνο του την τιμή 0 σε όσες θέσεις δεν έχουμε γεμίσει εμείς.

Στη συνέχεια πρέπει να μετατρέψουμε τα δεδομένα μας σε χρονοσειρές για να μπορούμε να δουλέψουμε με τις συναρτήσεις και να πραγματοποιήσουμε τις προβλέψεις μας. Όπως αναφέρθηκε και παραπάνω τα δεδομένα, που έχουμε στη διάθεση μας και τα έχουμε εισάγει στο πρόγραμμα, είναι πιθανότητα σε μια μορφή αρχείου τύπου *csv* ή *txt*. Σε αυτή τη μορφή δεν είναι επεξεργάσιμα από τον χρήστη και δεν μπορούν να αναπαρασταθούν σε γραφική παράσταση. Στην μετατροπή των δεδομένων σε χρονοσειρές ιδιαίτερη σημασία έχει η επιλογή συχνότητας των δεδομένων καθώς μια λάθος επιλογή μπορεί να οδηγήσει στην απώλεια ή αλλαγή ποιοτικών χαρακτηριστικών της χρονοσειράς όπως η εποχικότητα. Στο σημείο αυτό επιλέγουμε επίσης τα δεδομένα που θα έχουμε διαθέσιμα για την αξιολόγηση των μεθόδων. Ανάλογα με την μέθοδο που θα χρησιμοποιήσουμε, όπως θα δούμε και παρακάτω, παίρνουμε διαφορετικό κάθε φορά αριθμό διαθέσιμων δεδομένων.



5.1.3 ΑΠΟΣΥΝΘΕΣΗ ΧΡΟΝΟΣΕΙΡΑΣ

Αφού τελειώσουμε με την μετατροπή όλων των δεδομένων σε χρονοσειρές είμαστε έτοιμοι να προχωρήσουμε με την παραγωγή προβλέψεων μέσω της χρονικής συνάθροισης. Η χρονική συνάθροιση θα γίνει με την χρήση της βιβλιοθήκης *MAPA*. Από την βιβλιοθήκη αυτή θα χρησιμοποιήσουμε την συνάρτηση *mapasimple* κατά την οποία πρέπει να προσέξουμε τις παραμέτρους που θα χρησιμοποιήσουμε. Το πρώτο που πρέπει να προσέξουμε είναι τα ποιοτικά χαρακτηριστικά της χρονοσειράς μου. Δεδομένου ότι οι χρονοσειρές που έχουμε χαρακτηρίζονται από εποχικότητα πρέπει να κάνουμε αποσύνθεση της χρονοσειράς και να αφαιρέσουμε τον παράγοντα της εποχικότητας από τα δεδομένα μας, προτού προβούμε στην προβλέψεις. Η παράληψη του συγκεκριμένου βήματος θα έχει αρνητικές επιπτώσεις στις προβλέψεις καθώς “εκτοξεύει” τα αποτελέσματα των προβλέψεων. Παρακάτω φαίνεται ενδεικτικά πως είναι μια χρονοσειρά μετά την αποεποχικοποίηση. Βλέπουμε στην εικόνα ότι παρουσιάζονται τόσο τα ποιοτικά χαρακτηριστικά της χρονοσειράς (*trend, seasonal, random*) όσο και η υπό εξέταση χρονοσειρά (*observed*). Μπορούμε τώρα με κατάλληλες εντολές να χρησιμοποιήσουμε όποιο ποιοτικό χαρακτηριστικό της χρονοσειρές θελήσουμε.



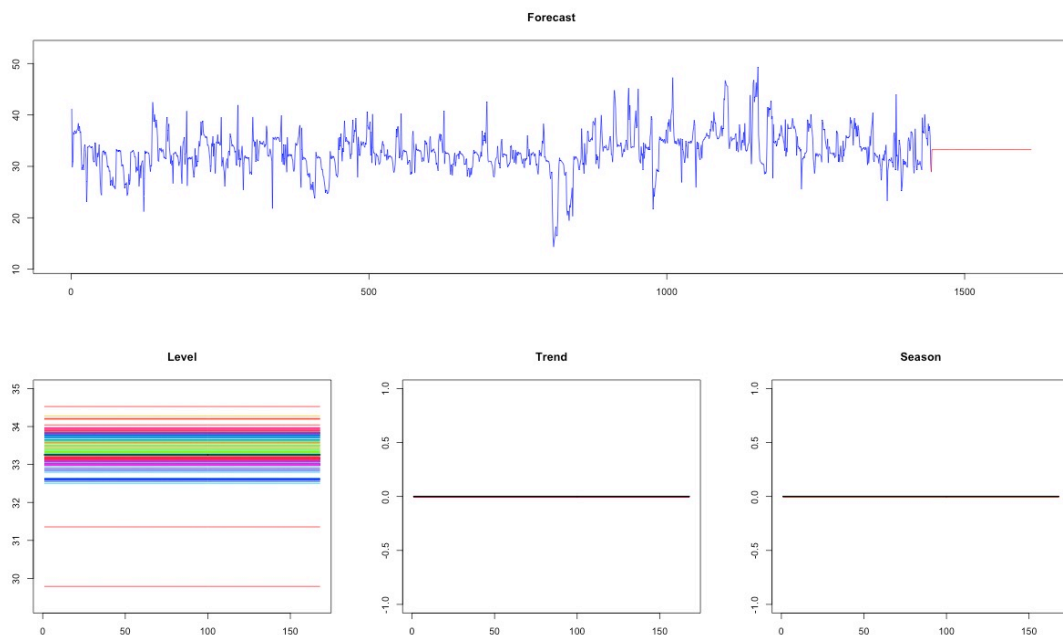
Εικόνα 6 Χαρακτηριστικά Χρονοσειράς



5.1.4 ΕΠΙΛΟΓΗ ΜΕΘΟΔΟΥ ΚΑΙ ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ

Αφού τελειώσουμε με την αποεποχικοποίηση των δεδομένων μπορούμε να συνεχίσουμε με την παραγωγή των προβλέψεων με την συνάρτηση *mapasimple*. Όπως αναφέραμε και παραπάνω, για την πραγματοποίηση των προβλέψεων πρέπει να δώσουμε ιδιαίτερη προσοχή στις παραμέτρους που δέχεται η συνάρτηση *mapasimple*. Η πρώτη σημαντική παράμετρος που δέχεται είναι τα δεδομένα, τα οποία όπως θα δούμε παρακάτω διαφέρουν σε πλήθος ανάλογα με τον τρόπο που επιλέγουμε να κάνουμε τις προβλέψεις. Η δεύτερη κύρια παράμετρος είναι ο ορίζοντας πρόβλεψης ή *forecast horizon*, ο οποίος μεταβάλλεται όπως και το πλήθος των δεδομένων και τέλος είναι η μέθοδος προβλέψεων που θα χρησιμοποιήσω.

Παρακάτω βλέπουμε την χρονοσειρά μαζί με τις προβλέψεις και τα ποιοτικά χαρακτηριστικά της. Φαίνεται ξεχωριστά το επίπεδο, η τάση και εποχικότητα της χρονοσειράς.



Εικόνα 7 Mapasimple

Όσον αφορά τις μεθόδους πρόβλεψης (*model=()*) που έχουμε στην διάθεσή μας έχει γίνει πλήρη αναφορά στο δεύτερο κεφάλαιο. Η επιλογή της μεθόδου εξαρτάται τόσο από την χρονοσειρά που έχουμε και τα χαρακτηριστικά όσο και από το τι θέλω να πετύχω και να μελετήσω. Η μέθοδος που επιλέχθηκε για την μελέτη αυτή είναι η *SES (Simple Exponential Smoothing)*. Για πλήρη ανάλυση της μεθόδου μπορεί ο αναγνώστης να ανατρέξει στο δεύτερο κεφάλαιο. Επιλέχθηκε για τρεις κυρίως λόγους, οι οποίοι αφορούν τα δεδομένα που έχουμε:

- Δεν επηρεάζεται η εποχικότητα που χαρακτηρίζει τα δεδομένα,
- Έχουμε *fast moving data*, δηλαδή γρήγορα μεταβαλλόμενα και όπως φαίνεται και από την παραπάνω εικόνα
- Δεν χαρακτηρίζονται από τάση.

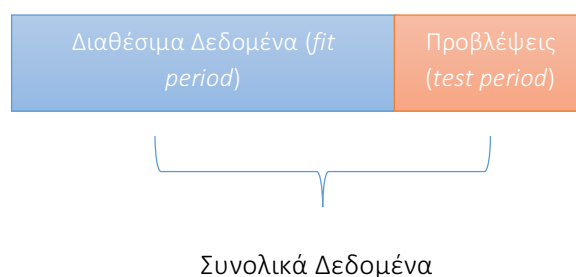
Επομένως αν ανατρέξουμε στον ορισμό και στον τρόπο λειτουργίας της μεθόδου παρατηρούμε ότι η *SES* είναι κατάλληλη για τα δεδομένα που έχουμε στη διάθεση μας.

5.1.4.1 ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ ΜΕ ΜΕΤΑΒΛΗΤΟ ΟΡΙΖΟΝΤΑ ΠΡΟΒΛΕΨΗΣ ΚΑΙ ΜΕΤΑΒΛΗΤΟ ΠΛΗΘΟΣ ΔΕΔΟΜΕΝΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

Στην περίπτωση αυτή του προγράμματος μεταβάλλουμε τα διαθέσιμα δεδομένα παράλληλα με τον ορίζοντα πρόβλεψης, έτσι ώστε σε κάθε επανάληψη ο αριθμός των δεδομένων μαζί με τον ορίζοντα πρόβλεψης να είναι ίσα με τις συνολικές παρατηρήσεις. Αν θεωρήσουμε δηλαδή ως ορίζοντες πρόβλεψης $ppy = c(50, 100, 150, 200)$ τότε τα διαθέσιμα δεδομένα θα είναι *total data-ppy*. Οπότε το πρόγραμμα θα επαναληφθεί και για τους 4 ορίζοντες πρόβλεψης και θα γεμίζουμε κάθε φορά την αντίστοιχη στήλη των πινάκων. Εδώ φαίνεται και αυτό που είπαμε σε παραπάνω ενότητα στην μετατροπή των δεδομένων σε χρονοσειρές. Στο σημείο αυτό είναι που θα κρίνουμε ποια θα είναι τα διαθέσιμα δεδομένα για την αξιολόγηση των μεθόδων. Μετά την μετατροπή σε χρονοσειρές θα θεωρούμε ως διαθέσιμο αριθμό δεδομένων κάθε φορά τα *total data-ppy*.

5.1.4.2 ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ ΜΕ ΜΕΤΑΒΛΗΤΟ ΟΡΙΖΟΝΤΑ ΠΡΟΒΛΕΨΗΣ ΚΑΙ ΣΤΑΘΕΡΟ ΠΛΗΘΟΣ ΔΕΔΟΜΕΝΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ (*Rolling Forecast*)

Στην δεύτερη περίπτωση κρατάμε σταθερό το πλήθος των διαθέσιμων δεδομένων και μεταβάλλουμε τον ορίζοντα πρόβλεψης κάθε φορά κατά μια παρατήρηση. Αρχικά πρέπει να σκεφτούμε πόσες παρατηρήσεις θέλουμε συνολικά να κάνουμε, δηλαδή πόσες φορές θα τρέξουμε τον κώδικα. Στη συνέχεια θα θεωρήσουμε ως διαθέσιμα δεδομένα τα συνολικά πλην το πλήθος των προβλέψεων που θέλουμε να κάνουμε. Επομένως στην περίπτωση αυτή χωρίζουμε τα συνολικά δεδομένα σε δύο μέρη, τα διαθέσιμα δεδομένα (*fit period*) και στα δεδομένα προς πρόβλεψη (*test period*). Το τελευταίο σημείο των διαθέσιμων δεδομένων θα είναι η αρχή των προβλέψεων (*forecast origin*). Όπως έχουμε αναφέρει και παραπάνω το πόσο μπροστά θα προβλέπουμε κάθε φορά ονομάζεται ορίζοντας πρόβλεψης (*forecast horizon*) και το μέγιστο του είναι ίσο με τον αριθμό των υπο πρόβλεψη δεδομένων (*test period*).



Η επιλογή της μεθόδου αυτής μας επιτρέπει να αξιολογήσουμε την ακρίβεια της πρόβλεψης για κάθε χρονοσειρά και για κάθε ορίζοντα πρόβλεψης. Για την μελέτη που πραγματοποιήσαμε θελήσαμε να παράγουμε προβλέψεις σε ένα εύρος τριών συχνοτήτων, δηλαδή 504 προβλέψεις. Επομένως θα θεωρήσουμε ως δεδομένες παρατηρήσεις τις πρώτες 1108 και τις υπόλοιπες 504 θα τις προβλέψουμε.

Ο ορίζοντας πρόβλεψης θα είναι η μονάδα, καθώς είναι ο ορίζοντας που χρησιμοποιείται για την βελτίωση των μεθόδων (*one step-ahead errors*), και σε κάθε επανάληψη θα αυξάνεται κατά 1. Κάθε 24 επαναλήψεις όπου συμπληρώνεται η μια ημέρα διαιρώ την τιμή μου με τον ορίζοντα πρόβλεψης και συμπληρώνω τον πίνακα. Το σφάλμα κάθε επανάληψης προστίθενται

στον πίνακα και μετά την διαίρεση με τον ορίζοντα πρόβλεψης προκύπτει ο μέσος όρος για κάθε ημέρα. Θεωρούμε δηλαδή ορίζοντα πρόβλεψης $rry = rpy + 1$ και διαθέσιμα δεδομένα $1612-3*168$. Επιλέγουμε τυχαία τον αριθμό των διαθέσιμων δεδομένων, ενώ η αρχική τιμή του rry ορίζεται η μονάδα. Για να συμπληρωθεί ο πίνακας σφαλμάτων προσθέτουμε κάθε φορά το εκάστοτε σφάλμα και στο τέλος διαιρούμε με τον αριθμό των επαναλήψεων για να προκύψουν τα μέσα σφάλματα.

Δείκτες Σφαλμάτων
sumerrors[1:24]/24
.
.
.
sumerrors[1:504]/504

Ο τρόπος συμπλήρωσης του πίνακα που περιγράψαμε παραπάνω παρουσιάζεται στον διπλανό πίνακα. Κάθε κελί είναι το πηλίκο της διαίρεσης τους αθροίσματος των σφαλμάτων με τον ορίζοντα πρόβλεψης.

5.1.5 ΠΑΡΑΓΩΓΗ ΤΩΝ ΤΕΛΙΚΩΝ ΠΡΟΒΛΕΨΕΩΝ

Αφού έχει γίνει όλη η παραπάνω ανάλυση είμαστε πλέον έτοιμοι να πραγματοποιήσουμε τις προβλέψεις μας ανάλογα με τις επιλογές που έχουμε κάνει από την παραπάνω ανάλυση. Εδώ πρέπει να προσέξουμε ότι, σε περίπτωση που έχουμε κάνει αποσύνθεση των δεδομένων μας, πρέπει να επαναφέρουμε οποιοδήποτε ποιοτικό χαρακτηριστικό έχουμε αφαιρέσει από την χρονοσειρά. Αν για παράδειγμα έχουμε αφαιρέσει την εποχικότητα πρέπει στη συνέχεια να πολλαπλασιάσουμε τις προβλέψεις με τον δείκτη εποχικότητας για να την επαναφέρουμε. Εδώ πρέπει να προσέξουμε ποια δεδομένα πολλαπλασιάζονται με ποιον δείκτη καθώς δεν χρησιμοποιούμε όλα τα δεδομένα στο σύνολό τους.

Μπορούμε πλέον να παράγουμε τις τελικές προβλέψεις κάνοντας τον συνδυασμό με τις μεθόδους ιεραρχικής συνάθροισης *top-down*, *bottom-up* και *optimal*. Για να ολοκληρώσουμε τις προβλέψεις με βάση την μέθοδο της *bottom-up* ανεβαίνουμε από το κάτω επίπεδο προς τα επάνω, προσθέτουμε δηλαδή τις προβλέψεις από το χαμηλότερο επίπεδο για να φτάσουμε στο υψηλότερο. Στα δικά μας δεδομένα έχουμε $k=0,1,2$ επίπεδα, επομένως θα πάρουμε τις προβλέψεις τους κατώτερου επιπέδου $k=2$ και προσθέτοντάς τες θα φτάσουμε στο επίπεδο $k=0$.

Όσον αφορά την μέθοδο *top-down* πρέπει να πάρουμε την πρόβλεψη που έχει προκύψει στα συνολικά δεδομένα στο ανώτερο επίπεδο $k=0$ και πολλαπλασιάζοντας με κατάλληλους συντελεστές να βρούμε τις προβλέψεις των κατώτερων επιπέδων $k=1,2$. Η επιλογή των συντελεστών μπορεί να διαφέρει από μελέτη σε μελέτη ωστόσο η επιλογή εδώ έγινε διαιρώντας το άθροισμα των δεδομένων που θέλω να προβλέψω με το άθροισμα των συνολικών δεδομένων.

Για την μέθοδο *optimal* πρέπει να δημιουργήσουμε ένα πίνακα συντελεστών με το οποίο θα πολλαπλασιάζουμε τις αρχικές προβλέψεις για να προκύψουν οι τελικές

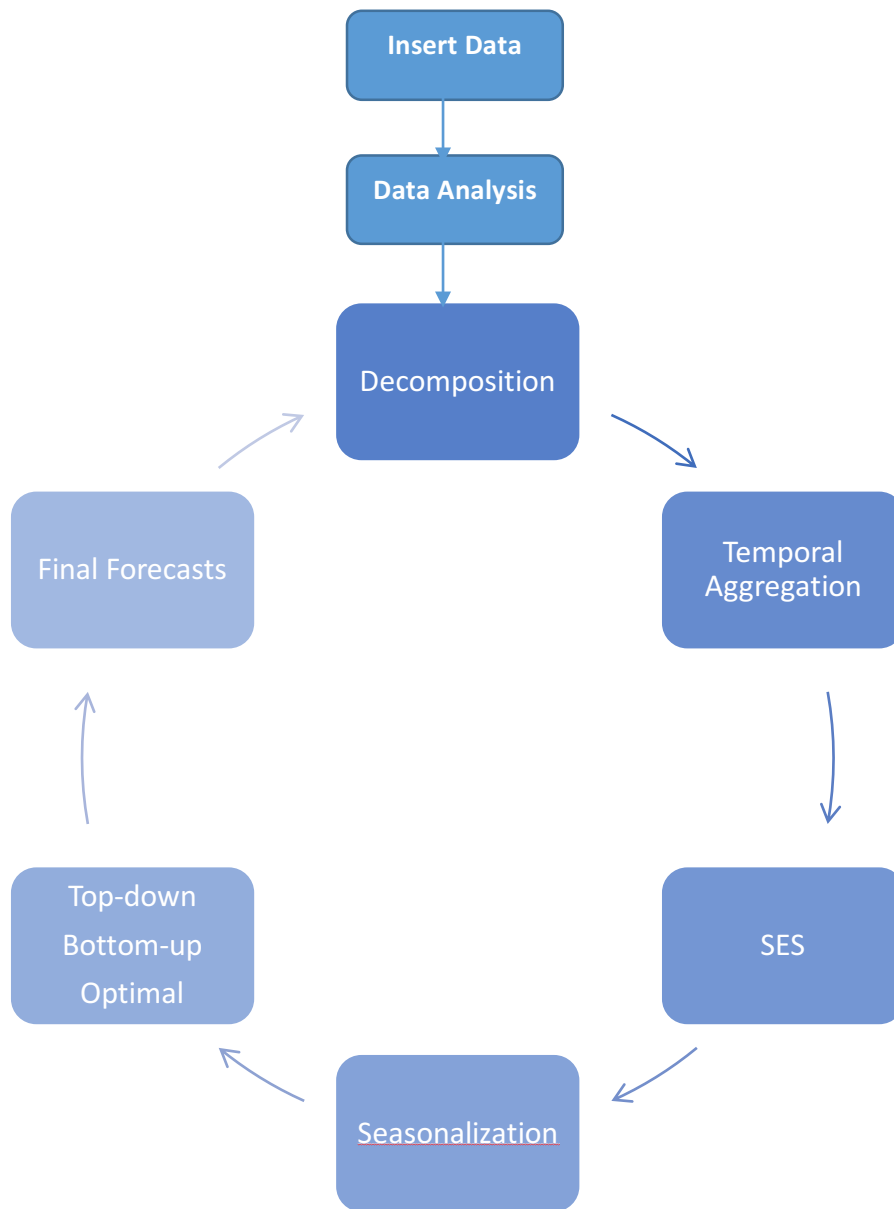
Optimal-S*MAPA

Ο πίνακας αυτός των συντελεστών προκύπτει μετά από πράξεις με βάση την μελέτη που έχει αναφερθεί και σε παραπάνω κεφάλαιο και πραγματοποιήθηκε το 2011 από τους R.J. Hyndman και R.A. Ahmed [18]. Για περισσότερη ανάλυση για το πως προέκυψε ο πίνακας συντελεστών μπορεί κανείς να αντέξει στην παραπάνω μελέτη.

5.1.6 ΣΥΜΠΛΗΡΩΣΗ ΠΙΝΑΚΩΝ

Αφού υπολογίσουμε τις τελικές προβλέψεις με όλους τους παραπάνω τρόπους μπορούμε να προχωρήσουμε στο τελικό στάδιο της μελέτης, στην συμπλήρωση δηλαδή των πινάκων σφαλμάτων που δημιουργήσαμε στην αρχή του προγράμματος. Όπως αναφέρθηκε και παραπάνω οι δείκτες σφαλμάτων υπολογίζονται με την εντολή *accuracy* εκτός από τον δείκτη *sMAPE* που υπολογίζουμε ξεχωριστά μόνοι μας. Η όλη διαδικασία που αναλύθηκε θα τρέξει για όσους ορίζοντες πρόβλεψης ορίσω, οπότε οι πίνακες θα συμπληρώνονται καθόλη τη διάρκεια της εκτέλεσης του προγράμματος στήλη στήλη ή γραμμή γραμμή ανάλογα με το πως έχω ορίσει τους πίνακες.

5.2 ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΜΕΘΟΔΟΛΟΓΙΚΗΣ ΠΡΟΣΕΓΓΙΣΗΣ



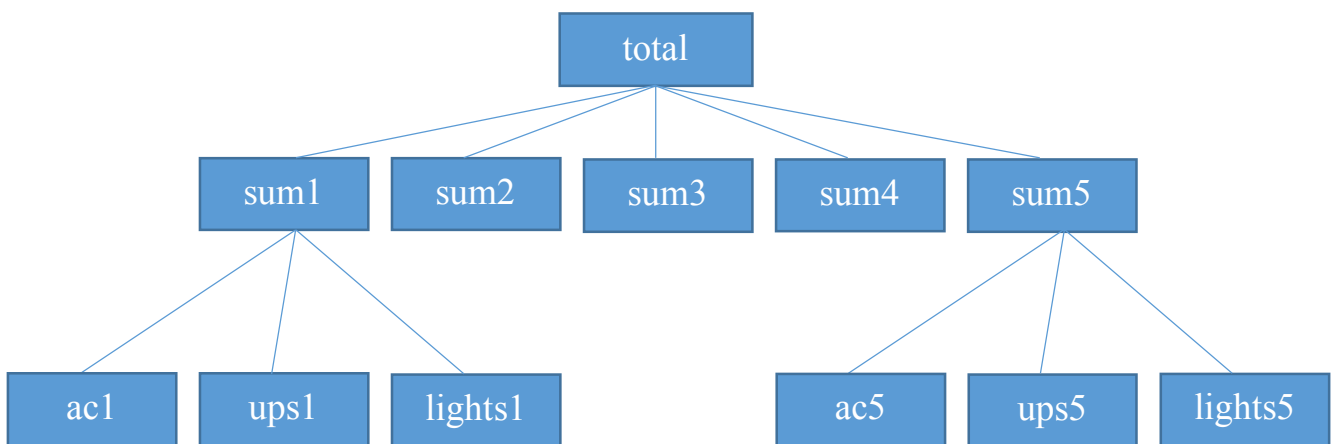
Εικόνα 8 Μεθοδολογική Προσέγγιση

Το παραπάνω σχηματικό διάγραμμα παρουσιάζει συνοπτικά την διαδικασία που αναλύσαμε παραπάνω. Το κυκλικό διάγραμμα επαναλαμβάνεται ανάλογα με τους ορίζοντες πρόβλεψης που χρησιμοποιώ, δηλαδή επαναλαμβάνεται όσο τρέχει το πρόγραμμα.

ΚΕΦΑΛΑΙΟ 6: CASE-STUDY ΣΕ ΣΥΜΠΛΕΓΜΑ ΚΤΗΡΙΩΝ ΤΡΑΠΕΖΑΣ

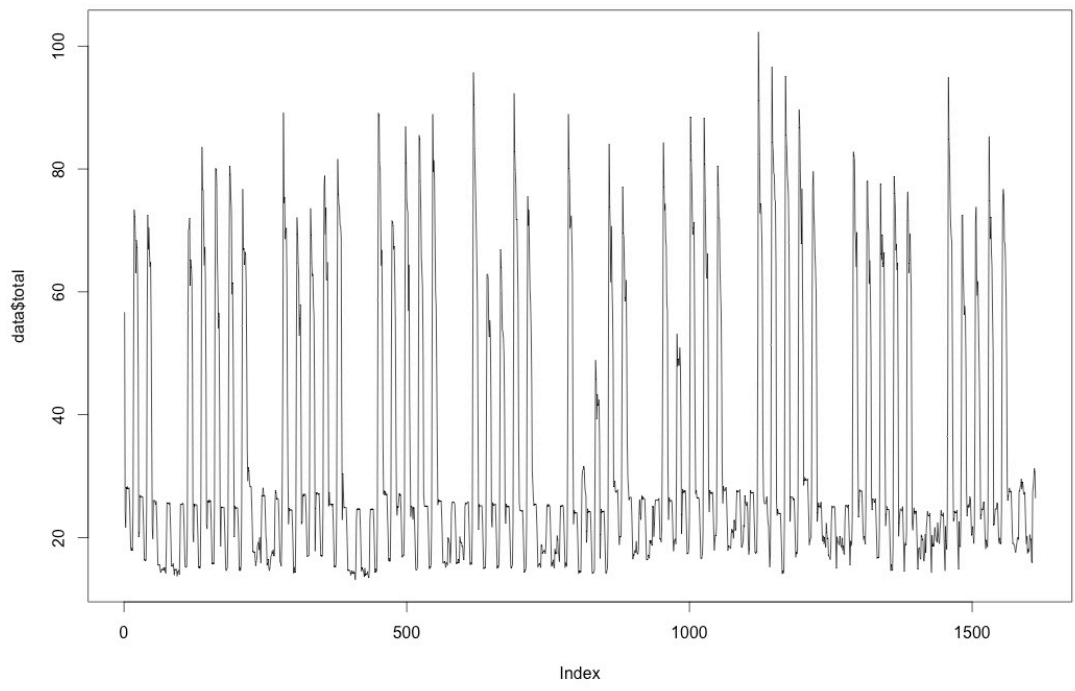
6.1 ΣΥΛΛΟΓΗ ΚΑΙ ΑΝΑΛΥΣΗ ΣΤΟΙΧΕΙΩΝ

Με βάση τα όσα περιγράψαμε παραπάνω ξεκινάμε την μελέτη για τα δικά μας δεδομένα. Για το συγκεκριμένο case-study επιλέχθηκαν 5 καταστήματα της τράπεζας Πειραιώς με δεδομένα, τα οποία συλλέχθηκαν από την ίδια την τράπεζα. Η διάρθρωση των δεδομένων παρουσιάζεται στο παρακάτω διάγραμμα και αποτελείται από 3 επίπεδα.

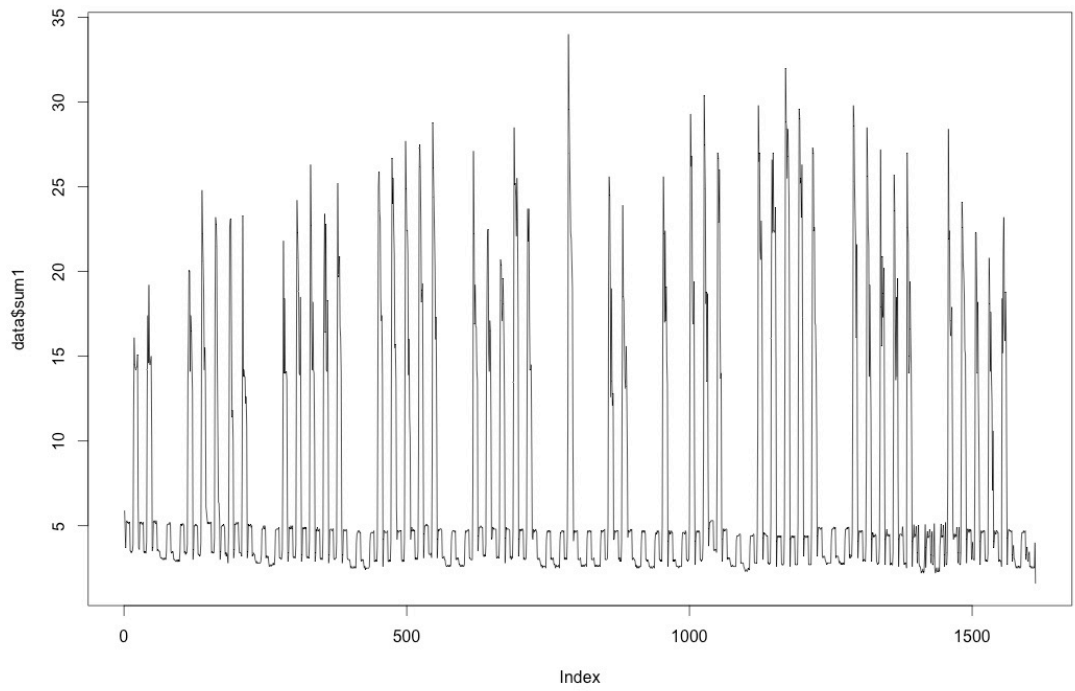


Εικόνα 9 Διάρθρωση Δεδομένων

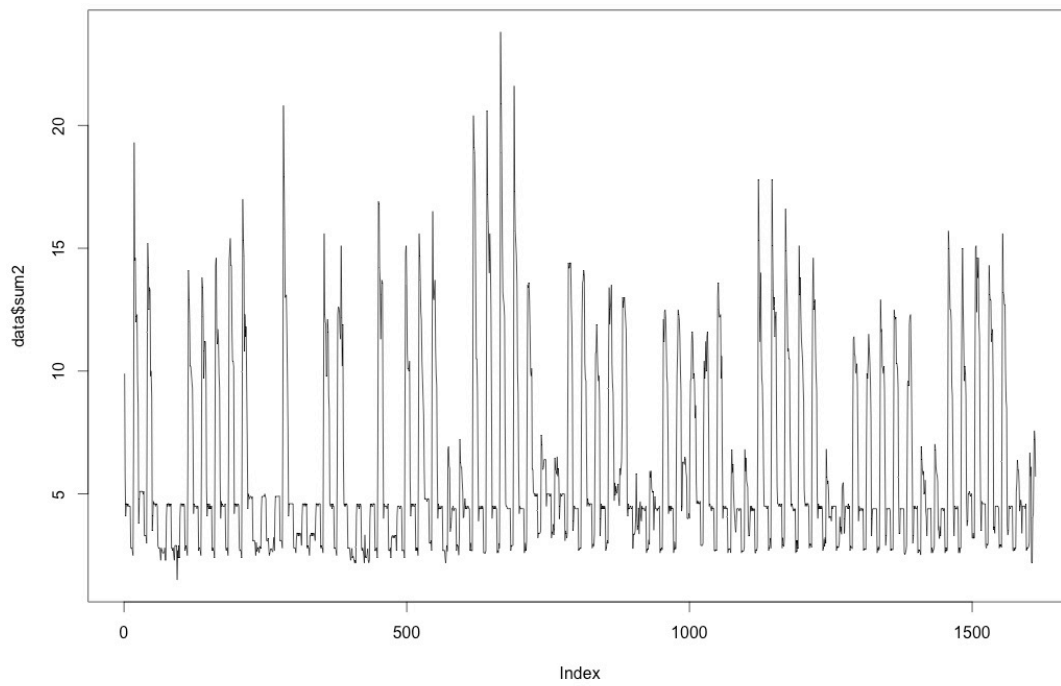
Στο επίπεδο $k=0$ έχω τη συνολική κατανάλωση και των 5 καταστημάτων της τράπεζας (*total*). Συνολικά στην διάθεση μας έχουμε 1612 παρατηρήσεις. Στο επίπεδο $k=1$ έχω την συνολική ενεργειακή κατανάλωση κάθε καταστήματος ξεχωριστά (*sum1*, ..., *sum5*) και στο τελευταίο επίπεδο $k=2$ έχω την κατανάλωση των επιμέρους τμημάτων κάθε καταστήματος (*ac1*, *ups1*, *lights1*, ..., *lights5*). Η κατανάλωση κάθε καστήματος χωρίζεται σε τρία μέρη, τα φώτα (*lights*), τα *ups* (αδιάκοπη παροχή ενέργειας) και το *ac unit* (*air condition*). Η συνολική κατανάλωση είναι το άθροισμα της κατανάλωσης των πέντε καταστημάτων και η συνολική κατανάλωση κάθε καταστήματος είναι το άθροισμα των τριών παραπάνω. Παρακάτω παρουσιάζονται τα διαγράμματα των δεδομένων όπως προκύπτουν από το *R-Studio* και ακολουθεί ανάλυση αυτών. Ακολουθούν 9 διαγράμματα, ένα διάγραμμα για την συνολική κατανάλωση των καταστημάτων (*total*), ένα για κάθε κατάστημα ξεχωριστά (*sum1*, ..., *sum5*) και 3 διαγράμματα για τα επιμέρους τμήματα του πρώτου καταστήματος ξεχωριστά (*ac1*, *ups1*, *lights1*).



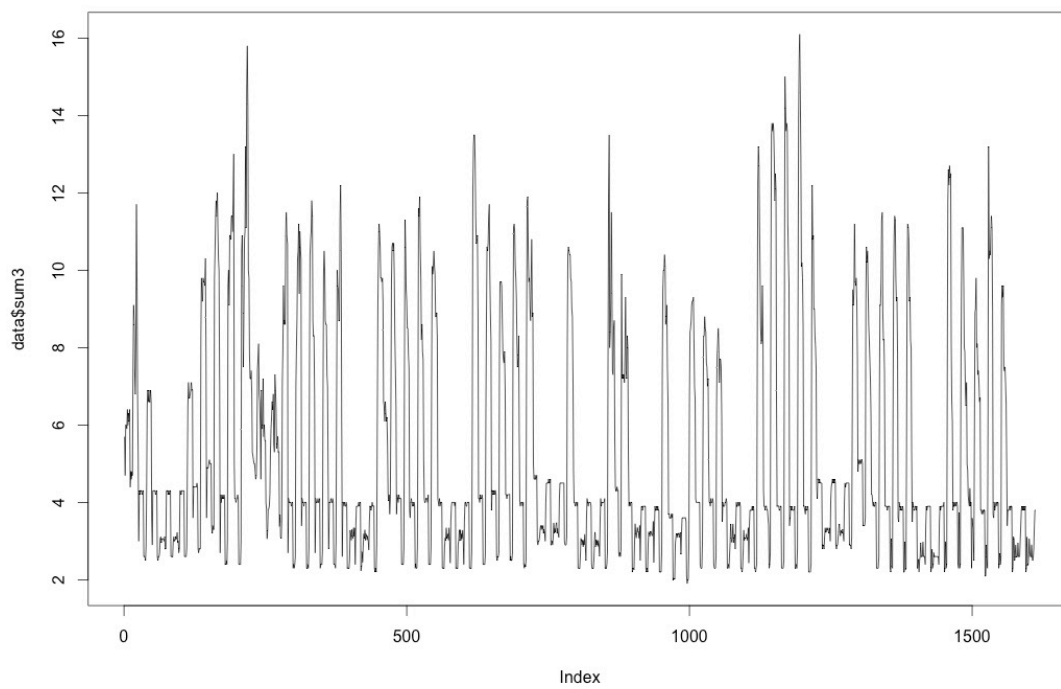
Εικόνα 10 Συνολική Κατανάλωση Κτιρίων (Total)



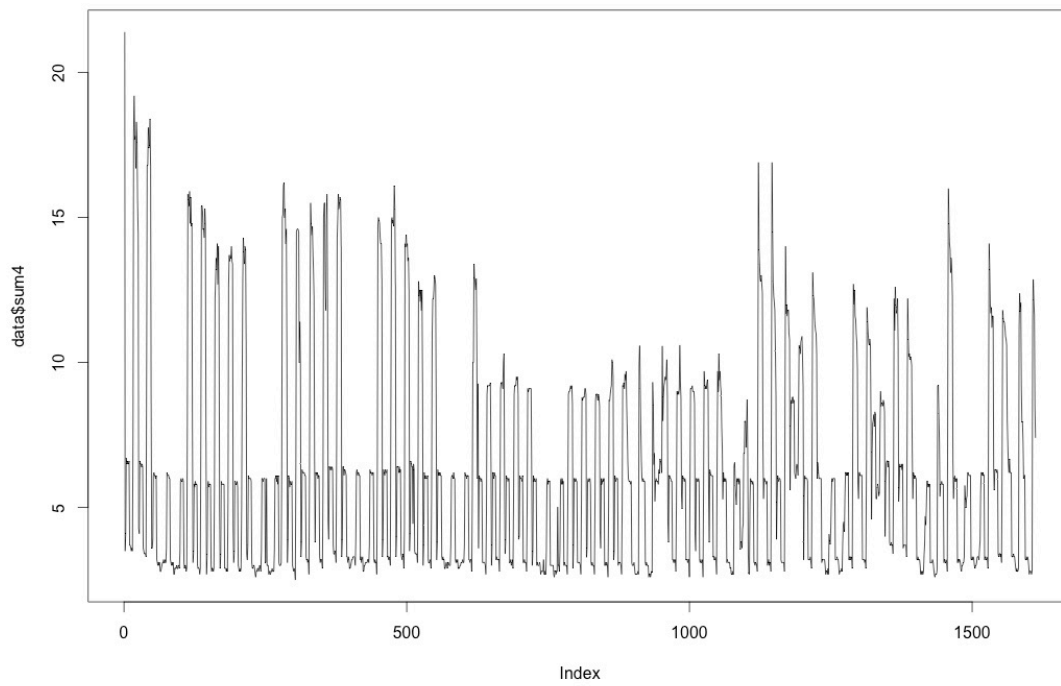
Εικόνα 11 Συνολική Κατανάλωση 1ου Κτιρίου (Sum1)



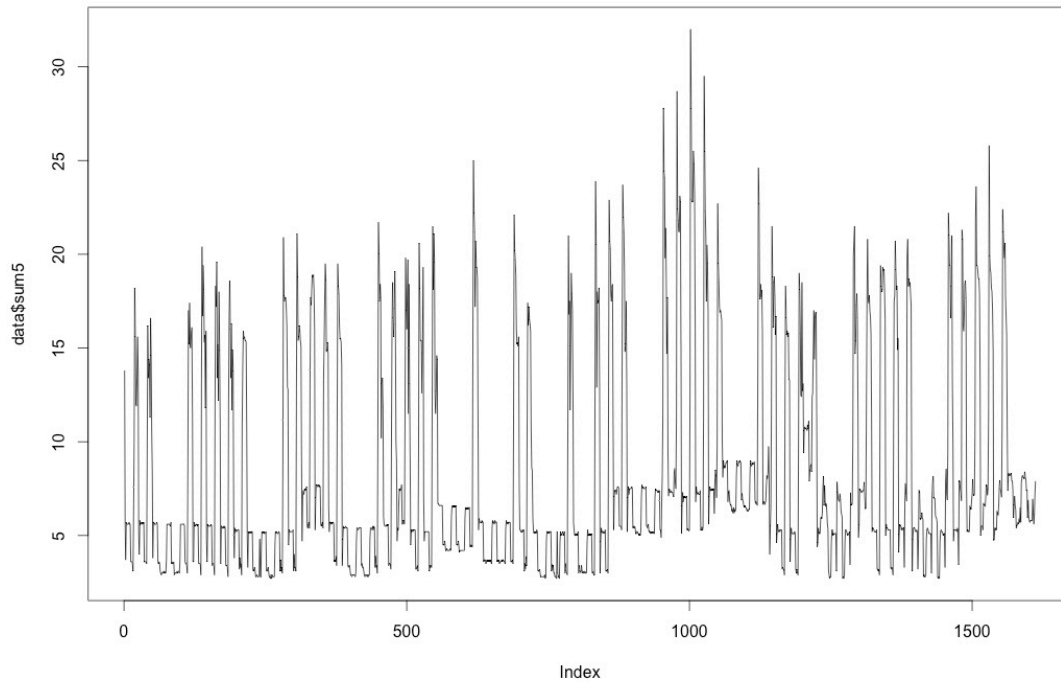
Εικόνα 12 Συνολική Κατανάλωση 2ου Κτιρίου (Sum2)



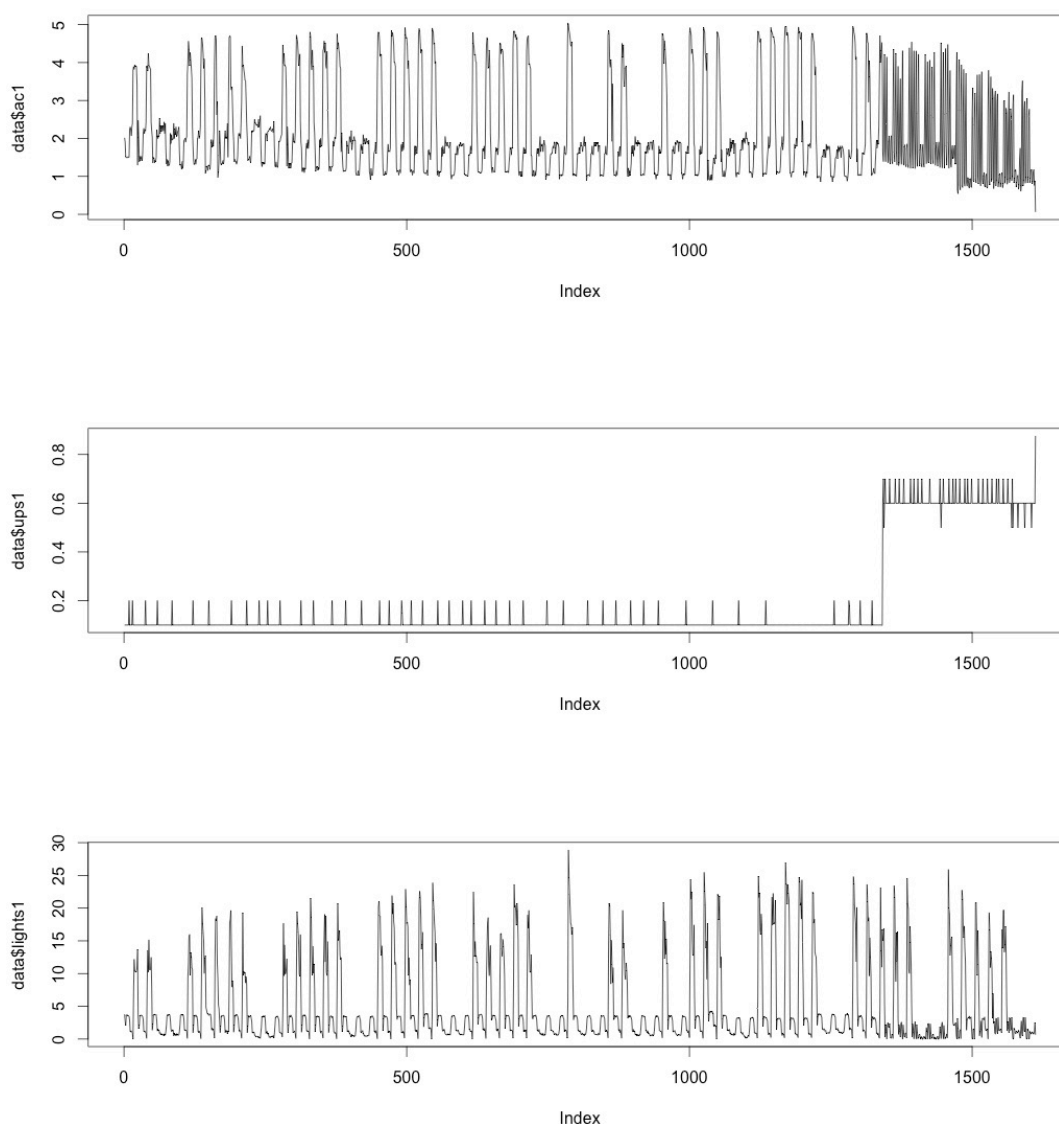
Εικόνα 13 Συνολική Κατανάλωση 3ου Κτιρίου (Sum3)



Εικόνα 14 Συνολική Κατανάλωση 4ου Κτιρίου (Sum4)



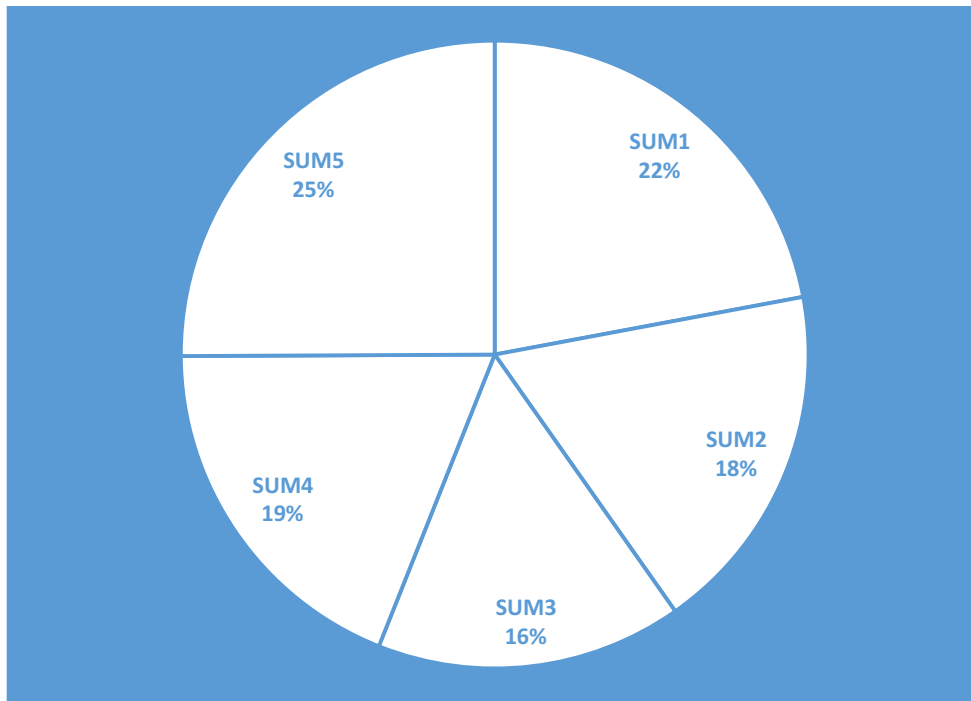
Εικόνα 15 Συνολική Κατανάλωση 5ου Κτιρίου (Sum5)



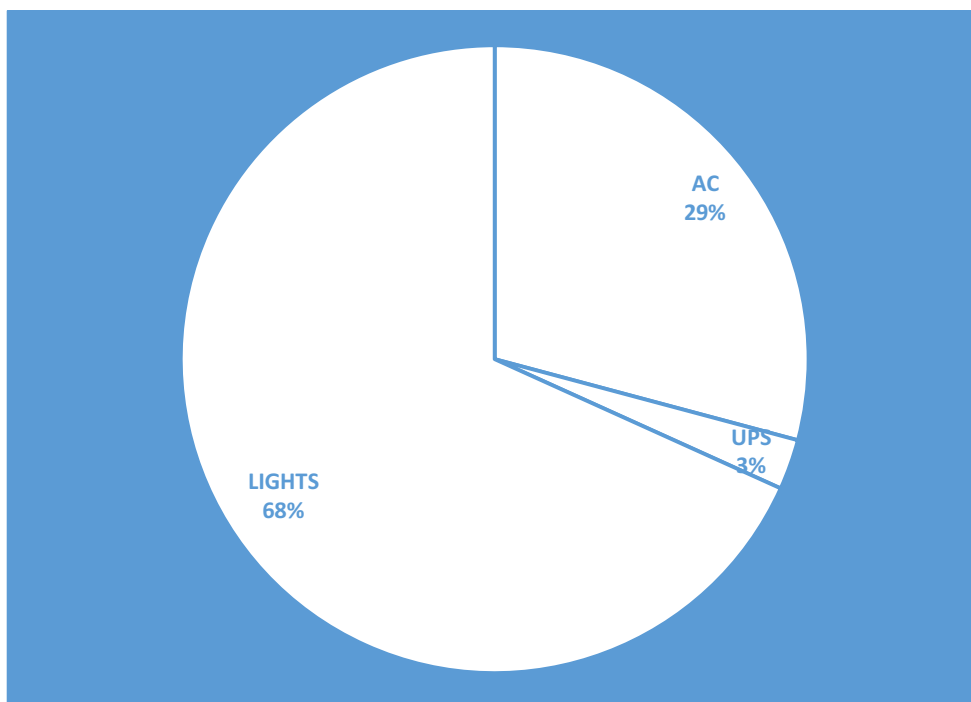
Εικόνα 16 Κατανάλωση ανά τμήμα του 1ου κτιρίου

Η συχνότητα των δεδομένων μας, όπως φαίνεται και από τα διαγράμματα, είναι $f=168$. Πρόκειται δηλαδή για ωριαία δεδομένα εβδομαδιαίας εποχικότητας. Αυτό είναι λογικό αν σκεφτούμε πως πρόκειται για καταστήματα τράπεζας. Δηλαδή αν παρατηρήσουμε τα διαγράμματα θα διαπιστώσουμε πως η ελάχιστη κατανάλωση παρατηρείται το Σάββατο και την Κυριακή που είναι κλειστά, ενώ τις καθημερινές η κατανάλωση εκτοξεύεται στα μέγιστα επίπεδα της. Τα αποτελέσματα αυτά επαναλαμβάνονται κάθε 168 ώρες, δηλαδή κάθε εβδομάδα. Επίσης οι μέγιστες τιμές κατανάλωσης παρατηρούνται τις πρωινές ώρες της ημέρας και αυτό γιατί τότε είναι ανοικτά τα καταστήματα και με βάση τις τιμές του ac καταλαβαίνουμε ότι πρόκειται για θερινή περίοδο. Με βάση τη συχνότητα αυτή επιλέχθηκαν και οι οριζόντες πρόβλεψης. Για την πρώτη περίπτωση επιλέχθηκε ως μέγιστος οριζοντας πρόβλεψης η συχνότητα $f=168$, από μια ημέρα $fh=24$ έως και μια εβδομάδα $fh=168$. Στην δεύτερη περίπτωση του *Rolling Forecasting* επειδή θέλουμε να ελέξουμε πως αντιδρά η

μέθοδος μακροπρόθεσμα επιλέγουμε ως μέγιστο ορίζοντα πρόβλεψης τριπλάσιο της συχνότητας. Παρακάτω παρουσιάζουμε δυο διαγράμματα όπου φαίνεται η ποσοστιαία κατανάλωση κάθε καταστήματος σε σχέση με την συνολική κατανάλωση και η ποσοστιαία κατανάλωση κάθε τμήματος σε σχέση με τη συνολική.



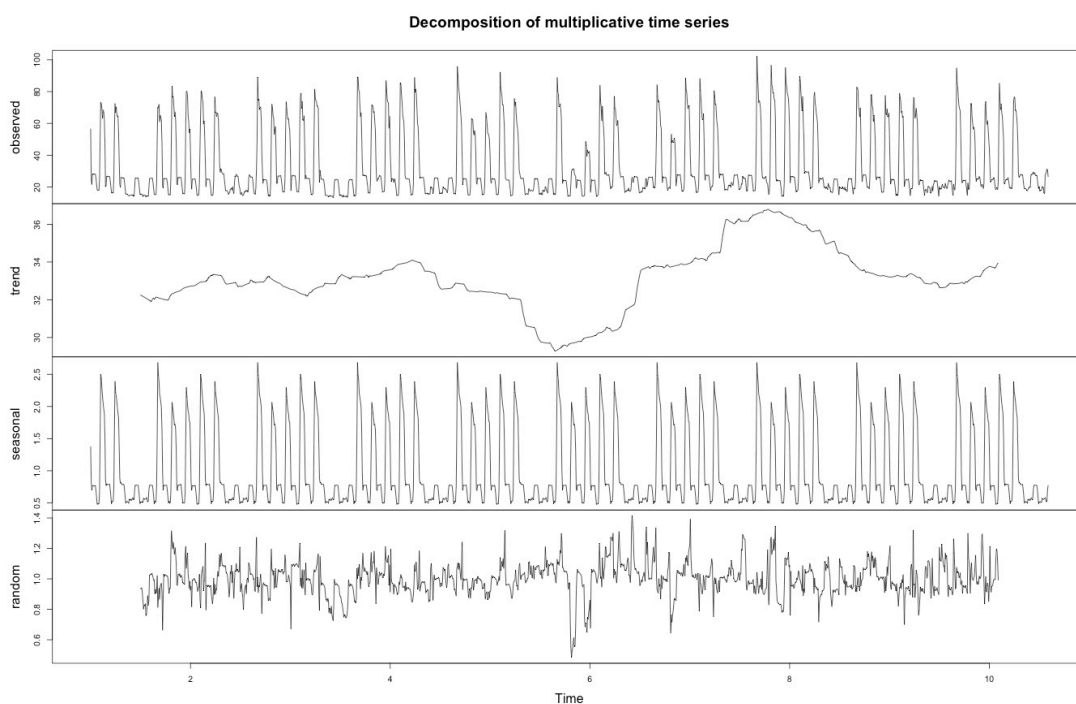
Εικόνα 23 Κατανομή Συνολικής Κατανάλωσης ανά Κτίριο



Εικόνα 24 Κατανομή Συνολικής Κατανάλωσης ανά Τμήμα

Στη συνέχεια δημιουργούμε τους πίνακες σφαλμάτων όπως είπαμε και παραπάνω. Οι πίνακες για τον πρώτο τρόπο θα έχουν διαστάσεις 7×7 , διότι έχουμε 7 δείκτες σφαλμάτων και 7 διαφορετικούς ορίζοντες πρόβλεψης $c=(24,48,72,96,120,144,168)$ ενώ για τον δεύτερο τρόπο θα έχουμε τριπλάσιο αριθμό στηλών. Εδώ πρέπει να σημειώσουμε ότι λόγω των τύπων που χρησιμοποιεί η συνάρτηση *accuracy* για να υπολογίζει τα σφάλματα πρέπει να κάνουμε έναν έλεγχο σε περίπτωση αρνητικού πρόσημου.

Προτού ξεκινήσουμε να κάνουμε τις προβλέψεις πρέπει να αποεποχικοποιήσουμε τις χρονοσειρές που έχουμε με την συνάρτηση *decompose*.



Εικόνα 25 Χαρακτηριστικά Χρονοσειράς

Στην παραπάνω εικόνα παρουσιάζεται η αρχική χρονοσειρά που έχουμε και από κάτω τα ποιοτικά χαρακτηριστικά της. Παίρνοντας το κομμάτι της εποχικότητας που θέλουμε *seasonal* μπορούμε να φτιάξουμε τον παρακάτω πίνακα, ο οποίος παρουσιάζει τους δείκτες εποχικότητας ανά ημέρα και ανά ώρα.

ΔΕΙΚΤΕΣ ΕΠΟΧΙΚΟΤΗΤΑΣ ΑΝΑ ΩΡΑ ΚΑΙ ΗΜΕΡΑ						
ΔΕΥΤΕΡΑ	ΤΡΙΤΗ	ΤΕΤΑΡΤΗ	ΠΕΜΠΤΗ	ΠΑΡΑΣΚΕΥΗ	ΣΑΒΒΑΤΟ	ΚΥΡΙΑΚΗ
1,56	1,36	1,49	1,58	1,60	0,49	0,51
2,68	2,07	2,30	2,50	2,39	0,56	0,58
2,49	1,98	2,16	2,43	2,22	0,55	0,56

2,39	1,91	2,02	2,28	2,16	0,54	0,56
2,20	1,82	1,93	2,15	2,08	0,56	0,55
2,14	1,71	1,88	2,05	1,98	0,54	0,56
2,07	1,72	1,77	1,96	1,89	0,56	0,59
1,99	1,64	1,76	1,89	1,82	0,58	0,55
1,65	1,32	1,38	1,54	1,39	0,55	0,52
0,98	0,71	0,82	0,87	0,88	0,55	0,53
0,70	0,69	0,70	0,77	0,81	0,70	0,70
0,77	0,78	0,77	0,80	0,82	0,77	0,77
0,77	0,78	0,77	0,79	0,79	0,78	0,78
0,77	0,78	0,77	0,79	0,79	0,78	0,77
0,77	0,78	0,76	0,79	0,78	0,78	0,77
0,77	0,79	0,77	0,78	0,79	0,77	0,77
0,78	0,79	0,77	0,79	0,79	0,77	0,77
0,69	0,70	0,68	0,70	0,70	0,69	0,70
0,59	0,60	0,58	0,60	0,60	0,59	0,60
0,50	0,49	0,48	0,49	0,50	0,49	0,48
0,53	0,50	0,48	0,52	0,52	0,52	0,53
0,50	0,50	0,48	0,52	0,52	0,51	0,51
0,50	0,50	0,49	0,53	0,52	0,52	0,53
0,68	0,67	0,67	0,72	0,51	0,51	0,75

Ανάλογα με τον παραπάνω πίνακα θα πολλαπλασιάσουμε αργότερα τις προβλέψεις για να επαναφέρουμε την εποχικότητα στις χρονοσειρές. Χρειάζεται προσοχή ώστε οι προβλέψεις να πολλαπλασιαστούν με τους αντίστοιχους δείκτες και όχι με άλλους διαφορετικής χρονικής περιόδου διότι υπάρχει ο κίνδυνος αλλαγής της εποχικότητας.

6.2 ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΨΕΩΝ

Μπορούμε πλέον να προχωρήσουμε στην παραγωγή των προβλέψεων και στην συμπλήρωση των πινάκων με τα σφάλματα. Παρακάτω παρουσιάζονται κάποιοι από τους πίνακες σφαλμάτων και σημαντικά διαγράμματα από τα σφάλματα που προέκυψαν προκειμένου να προκύψουν τα συμπεράσματα της ανάλυσης. Θα ακολουθήσει ανάλυση των διαγραμμάτων αυτών καθώς και τα συμπεράσματα που προκύπτουν.

Σφάλμα sMAPE με την χρήση MAPA								
BOTTOM UP								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	11,92	12,77	10,09	9,59	9,53	9,30	8,96	10,31
LEVEL 1	24,37	20,77	20,53	18,32	18,55	19,88	19,67	20,30
LEVEL 2	45,05	42,43	41,39	42,78	45,21	45,02	41,83	43,39
AVERAGE	20,59	14,20	13,49	14,09	14,98	15,16	14,42	19,03
TOP DOWN								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	12,28	13,52	10,80	9,74	9,52	9,32	8,88	10,05
LEVEL 1	21,43	21,76	20,18	18,88	20,45	21,85	21,44	20,86
LEVEL 2	53,78	53,56	56,68	56,60	57,05	57,30	57,82	56,11
AVERAGE	29,16	29,61	29,22	28,41	29,01	29,49	29,38	29,18
OPTIMAL								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	12,72	12,33	9,70	9,30	9,34	9,20	8,78	10,20
LEVEL 1	16,60	15,45	16,56	14,80	16,37	17,32	16,96	16,29
LEVEL 2	60,30	56,89	55,50	49,98	50,43	51,14	47,59	53,12
AVERAGE	29,87	28,23	27,25	24,70	25,38	25,88	24,45	26,54

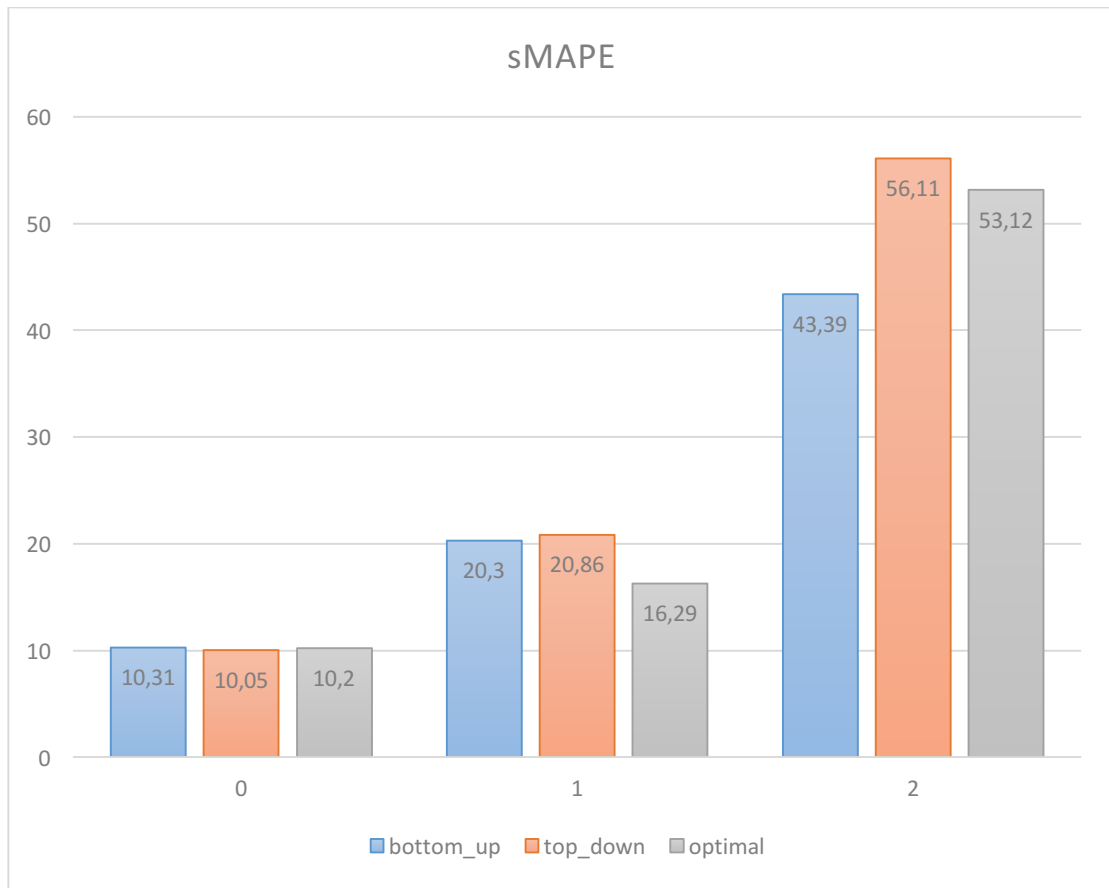
Σφάλμα sMAPE με την απλή SES								
BOTTOM UP								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	22,75	20,26	27,11	30,68	31,53	33,67	36,05	28,86
LEVEL 1	34,77	-0,70	0,99	1,48	1,48	1,97	2,29	6,04
LEVEL 2	54,90	61,85	56,62	56,15	72,11	64,92	70,23	62,40
AVERAGE	28,36	20,85	21,93	23,07	27,53	26,64	28,89	32,43

TOP DOWN								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	20,50	19,50	27,86	31,58	33,16	34,37	36,05	29,00
LEVEL 1	27,05	26,15	35,08	38,87	40,02	40,96	42,96	35,87
LEVEL 2	51,80	50,93	57,29	59,92	61,08	61,34	62,71	57,87
AVERAGE	33,12	32,19	40,08	43,46	44,75	45,56	47,24	40,91
OPTIMAL								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	20,49	19,50	27,68	31,34	32,60	34,21	36,04	28,84
LEVEL 1	32,77	30,86	39,19	41,30	44,93	44,27	45,11	39,78
LEVEL 2	70,09	67,74	58,23	56,90	77,21	65,36	70,32	66,55
AVERAGE	41,12	39,37	41,70	43,18	51,58	47,95	50,49	45,05

Σφάλμα ΜΡΕ με την χρήση ΜΑΡΑ								
BOTTOM UP								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	1,69	9,82	5,79	3,36	3,63	3,68	2,50	4,36
LEVEL 1	-20,50	-14,23	-13,75	-5,98	-5,38	-5,29	-9,75	-10,70
LEVEL 2	-247,15	-437,57	-403,72	-347,95	-471,09	-524,86	-475,85	-415,46
AVERAGE	-66,24	-109,99	-102,17	-86,64	-116,96	-130,12	-119,03	-140,60
TOP DOWN								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	0,80	10,57	7,30	4,62	4,01	4,26	2,96	4,93
LEVEL 1	-6,22	4,59	0,49	-1,75	-4,96	-5,63	-6,38	-2,84
LEVEL 2	-753,95	-711,65	-805,56	-742,37	-791,40	-798,12	-785,28	-769,76
AVERAGE	-253,12	-232,17	-265,92	-246,50	-264,12	-266,50	-262,90	-255,89
OPTIMAL								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	-0,10	9,19	5,86	4,00	3,83	4,11	2,63	4,22
LEVEL 1	-3,94	5,09	2,78	-0,39	-3,45	-4,48	-5,06	-1,35
LEVEL 2	289,48	-42,96	-90,38	-208,63	-393,07	-467,41	-408,59	-188,79
AVERAGE	95,15	-9,56	-27,25	-68,34	-130,90	-155,93	-137,01	-61,98

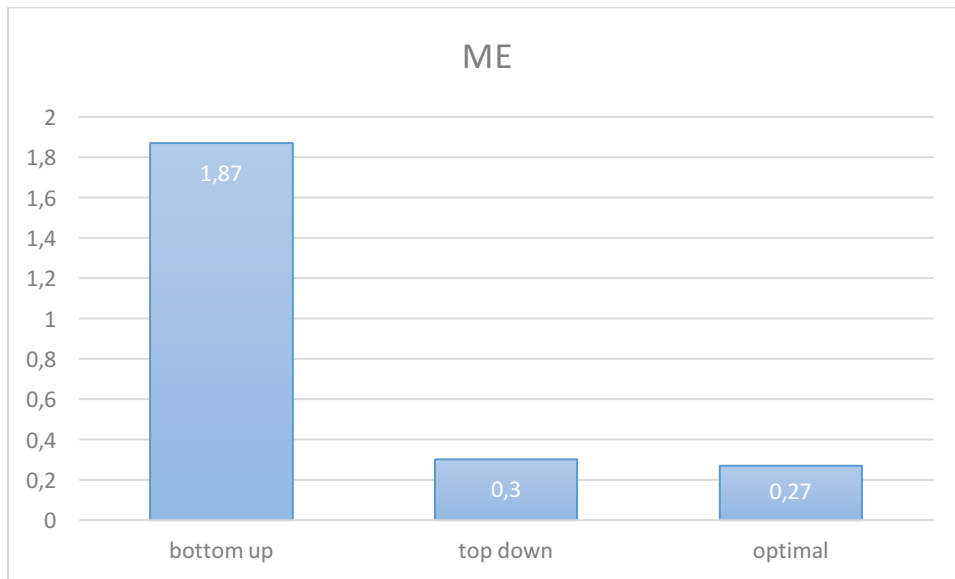
Σφάλμα MPE με την απλή SES								
BOTTOM UP								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	-25,53	-19,76	3,69	7,16	4,69	13,51	16,75	0,07
LEVEL 1	-36,59	-29,25	-7,98	-2,76	-7,09	2,83	6,84	-10,57
LEVEL 2	-747,57	-784,06	-869,15	-582,86	-227,32	-651,34	-543,43	-629,39
AVERAGE	-202,17	-207,77	-217,61	-143,61	-56,18	-157,25	-128,21	-213,30
TOP DOWN								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	-0,66	-0,06	8,59	12,25	14,77	15,77	16,75	9,63
LEVEL 1	-8,07	-7,11	1,40	5,65	7,65	8,35	9,69	2,51
LEVEL 2	-841,26	-778,48	-863,57	-780,03	-808,98	-811,72	-797,57	-811,66
AVERAGE	-283,33	-261,88	-284,53	-254,04	-262,19	-262,53	-257,04	-266,51
OPTIMAL								
	1	2	3	4	5	6	7	AVERAGE
LEVEL 0	-5,93	-3,59	25,02	22,29	23,01	20,31	22,66	14,83
LEVEL 1	-13,08	-9,84	18,53	17,54	8,59	8,95	18,38	7,01
LEVEL 2	-359,63	-342,74	-224,56	-106,05	-165,32	-730,40	-146,21	-296,42
AVERAGE	-126,21	-118,72	-60,34	-22,07	-44,57	-233,71	-35,06	-91,53

Οι παραπάνω πίνακες παρουσιάζουν τους δείκτες σφαλμάτων *sMAPE* και *MPE* για την συνδυασμένη *SES* με την *MAPA* και για την απλή *SES* αντίστοιχα.



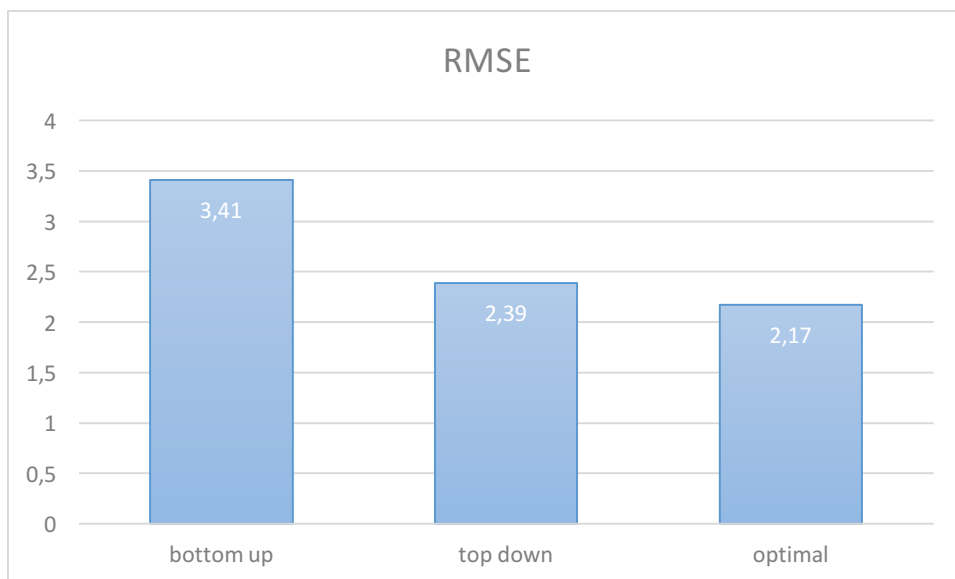
Εικόνα 17 Μέσο Σφάλμα sMAPE, για την SES σε συνδυασμό με την MAPA, για 7 ορίζοντες και για κάθε επίπεδο ξεχωριστά

Το παραπάνω διάγραμμα παρουσιάζει τη σχέση μεταξύ των σφαλμάτων για τις μεθόδους *bottom-up*, *top-down* και *optimal*. Οι τιμές που παρουσιάζονται στο διάγραμμα αποτελούν τα μέσα σφάλματα, αντίστοιχα για τον κάθε δείκτη, από τους 7 ορίζοντες πρόβλεψης για κάθε επίπεδο όπως αναγράφεται στον τίτλο. Αυτό που παρατηρούμε είναι ότι παρουσιάζεται μια συγκεκριμένη ιεραρχία ως προς τις μεθόδους *bottom-up*, *top-down* και *optimal*. Παρατηρούμε ότι για το επίπεδο $k=0$ η μέθοδος *top-down* παρουσιάζει το μικρότερο μέσο σφάλμα για τους 7 ορίζοντες της εκτέλεσης του προγράμματος. Για το επίπεδο $k=1$ η μέθοδος *optimal* παρουσιάζει το μικρότερο μέσο σφάλμα σε σχέση με τις *top-down* και *bottom-up*, ενώ για το επίπεδο $k=2$ η μέθοδος *bottom-up* έχει τα καλύτερα αποτελέσματα. Το συμπέρασμα αυτό είναι ήδη γνωστό στην βιβλιογραφία και επιβεβαιώνεται και στην δική μας μελέτη που περιλαμβάνει δεδομένα ενεργειακής κατανάλωσης κατανεμημένα σε 3 επίπεδα. Επίσης επιβεβαιώνεται και η υπεροχή της μεθόδου *optimal* όπως έχει ήδη αναφερθεί και στο 3^ο Κεφάλαιο. Στα παρακάτω διαγράμματα παρουσιάζεται το μέσο σφάλμα από όλα τα επίπεδα k και για διάφορους δείκτες σφαλμάτων. Η μέθοδος *optimal* υπερτερεί των άλλων δύο και παρουσιάζει τα μικρότερα σφάλματα. Τα αποτελέσματα αυτά δεν φαίνονται σε όλες τις προσομοιώσεις και σε όλους τους δείκτες σφαλμάτων, ωστόσο παρουσιάζεται στην πλειονότητα τους και προκύπτει ως κεντρικό συμπέρασμα.

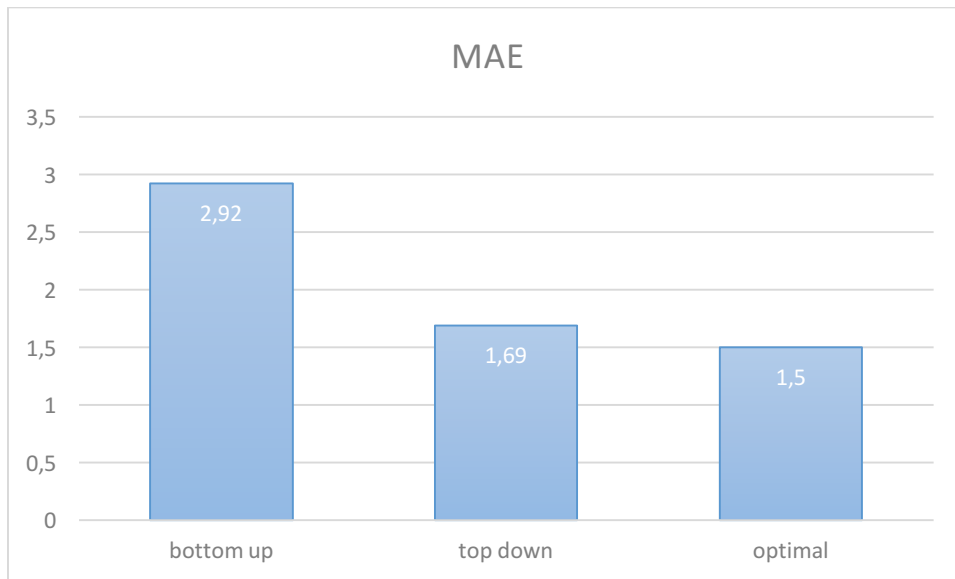


Εικόνα 18 Μέσο Σφάλμα ME για όλους τους ορίζοντες ($fh = 24, \dots, 168$) και για όλα τα επίπεδα ($k=0,1,2$)

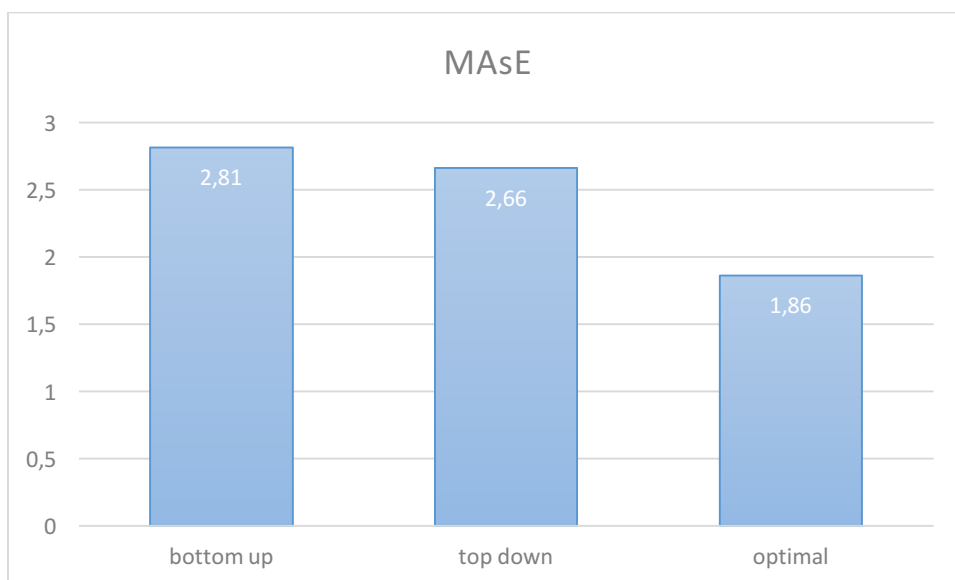
Τα συμπεράσματα αυτά προκύπτουν και εδώ από την εκτέλεση και της απλής *SES* και της *SES* συνδυασμένη με τον αλγόριθμο *MAPA*. Η διαφορά μεταξύ των *bottom-up*, *top-down* και *optimal* διαφέρει σε βαθμό από δείκτη σε δείκτη ωστόσο η *optimal* παραμένει πιο αποτελεσματική από τις άλλες δύο.



Εικόνα 19 Μέσο Σφάλμα RMSE για όλους τους ορίζοντες ($fh = 24, \dots, 168$) και για όλα τα επίπεδα ($k=0,1,2$)



Εικόνα 20 Μέσο Σφάλμα MAE για όλους τους ορίζοντες ($fh = 24, \dots, 168$) και για όλα τα επίπεδα ($k=0,1,2$)

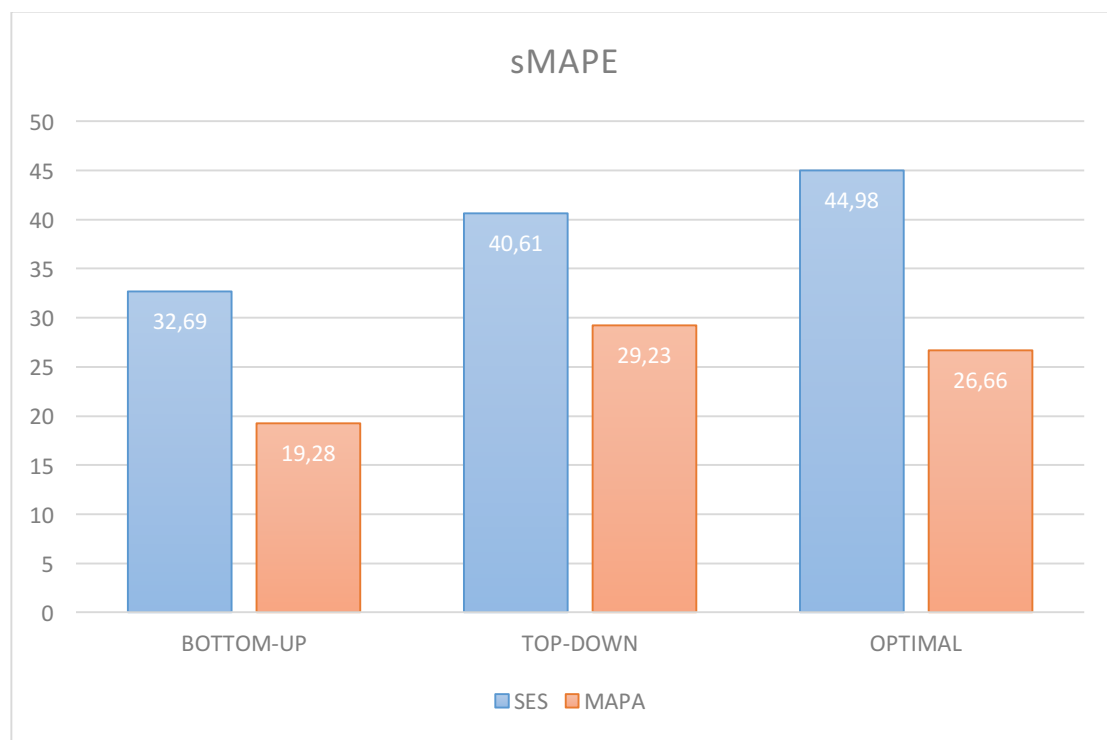


Εικόνα 21 Μέσο Σφάλμα MASe για όλους τους ορίζοντες ($fh = 24, \dots, 168$) και για όλα τα επίπεδα ($k=0,1,2$)

Στα επόμενα διαγράμματα παρουσιάζεται το κεντρικό κομμάτι της διπλωματικής εργασίας και ο σκοπός που πραγματοποιείται. Παρουσιάζεται δηλαδή η σχέση μεταξύ των σφαλμάτων από την παραγωγή των προβλέψεων μέσω της απλής SES και της συνδυασμένης με τον αλγόριθμο MAPA. Στα διαγράμματα επικεντρωνόμαστε σε δυο δείκτες σφαλμάτων, τον δείκτη MPE για την προκατάληψη των προβλέψεων και τον δείκτη sMAPE για την ακρίβεια τους. Η σύγκριση στα διαγράμματα μπορεί να γίνει ως προς τρεις παράγοντες, τους ορίζοντες πρόβλεψης, τα επίπεδα k και τις μεθόδους *bottom-up*, *top-down* και *optimal*.

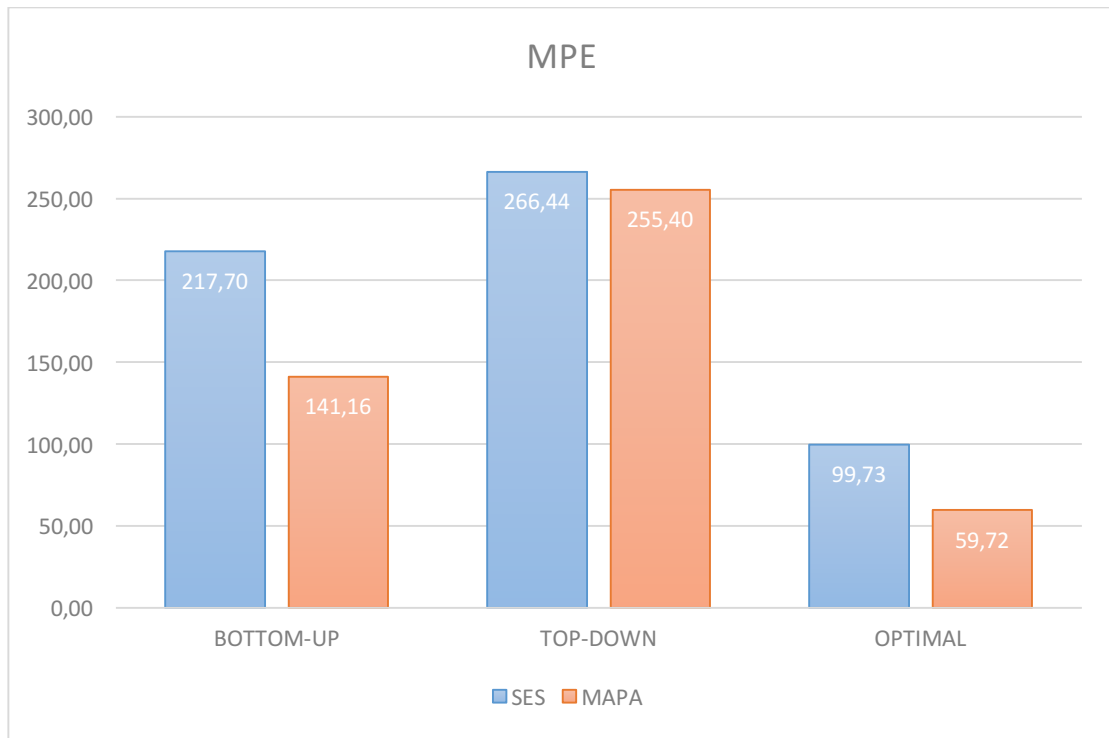
Στα επόμενο διαγράμμα εξετάζουμε τη σχέση μεταξύ των μεθόδων *bottom-up*, *top-down* και *optimal* για την απλή SES και για την συνδυασμένη για όλα τα επίπεδα και για όλους τους

ορίζοντες πρόβλεψης. Για κάθε μια μέθοδο παίρνουμε το μέσο όρο των σφαλμάτων από όλα τα επίπεδα $k=0, 1, 2$ και για τους 7 ορίζοντες.



Εικόνα 22 sMAPE για SES και MAPA για όλα τα επίπεδα και για όλους τους ορίζοντες πρόβλεψης

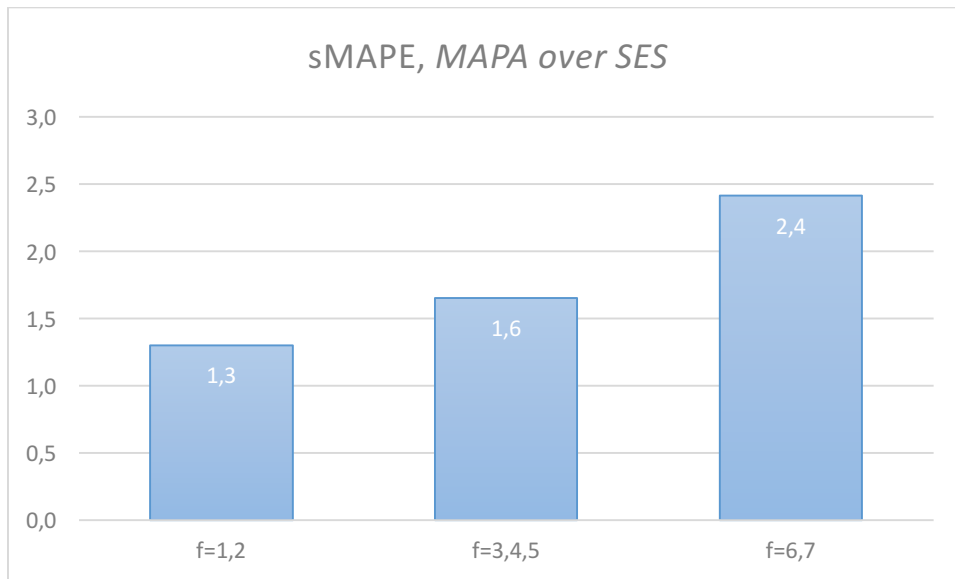
Παρατηρούμε ότι τα σφάλματα της απλής SES είναι μέχρι και 1,7 φορές τα σφάλματα της συνδυασμένης. Πετύχαμε δηλαδή σχεδόν 2 φορές καλύτερη ακρίβεια από την απλή SES. Υπενθυμίζουμε ότι τα διαγράμματα περιλαμβάνουν τα μέσα σφάλματα για όλα τα επίπεδα k και για όλους τους παράγοντες πρόβλεψης.



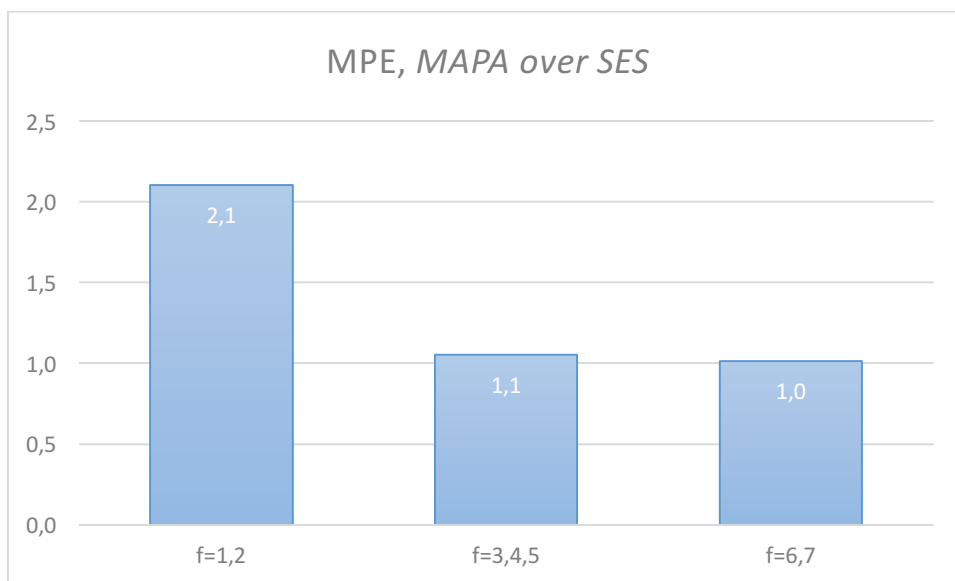
Εικόνα 23 sMAPE, MAPA over SES (bottom-up, top-down)

Αντίστοιχα μπορούμε να εργαστούμε και για τον δείκτη σφάλματος *MPE* όπως φαίνεται και παραπάνω. Παρατηρούμε ότι και στον δείκτη αυτό έχουμε μεγαλύτερα σφάλματα από πλευράς της απλής *SES*. Στην εφαρμογή δηλαδή της απλής *SES* έχουμε μεγαλύτερη προκατάληψη σφαλμάτων όσον αφορά τις *bottom-up* και *optimal* γιατί όπως φαίνεται στο διάγραμμα η *top-down* προέκυψε ίδια (με στρογγυλοποίηση).

Στα επόμενα δυο διαγράμματα βλέπουμε τη σχέση των σφαλμάτων των δύο μεθόδων για τους διάφορους οριζόντες πρόβλεψης. Δεν μας ενδιαφέρει τώρα το επίπεδο των δεδομένων ούτε οι μέθοδοι *bottom-up*, *top-down* και *optimal*. Βρίσκουμε τα μέσα σφάλματα για τις τιμές των οριζόντων $f=1$ και 2 , $f=3, 4$ και 5 , και $f=6$ και 7 για όλα τα επίπεδα k και για τις τρεις παραπάνω μεθόδους. Διαιρούμε πάλι τις τιμές των σφαλμάτων τις απλής *SES* και της συνδυασμένης για να βρούμε τη σχέση μεταξύ τους. Ακολουθούν τα διαγράμματα των *sMAPE* και *MPE* που δείχνουν την παραπάνω διαδικασία.



Εικόνα 24 sMAPE, MAPA over SES σε σχέση με τους ορίζοντες πρόβλεψης

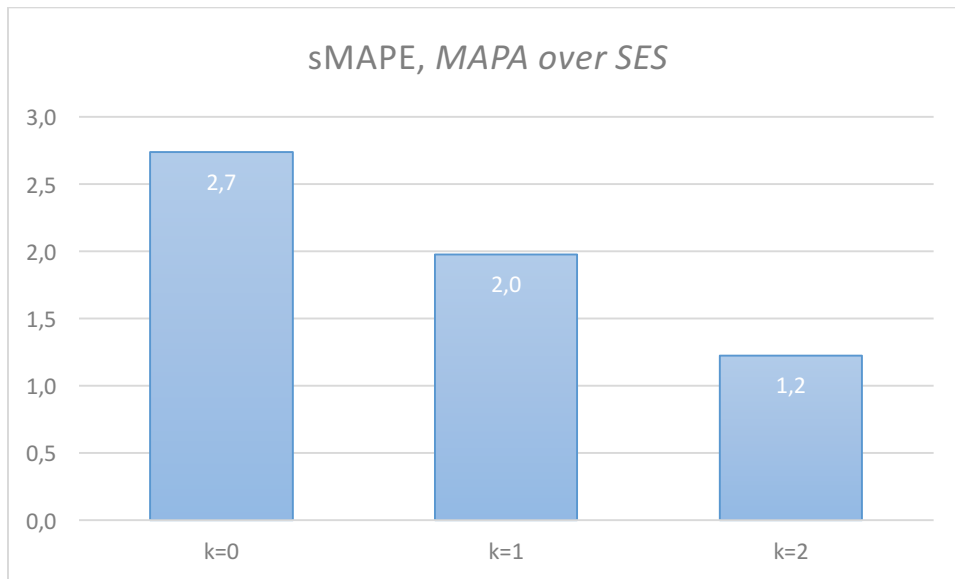


Εικόνα 25 MPE, MAPA over SES σε σχέση με τους ορίζοντες πρόβλεψης

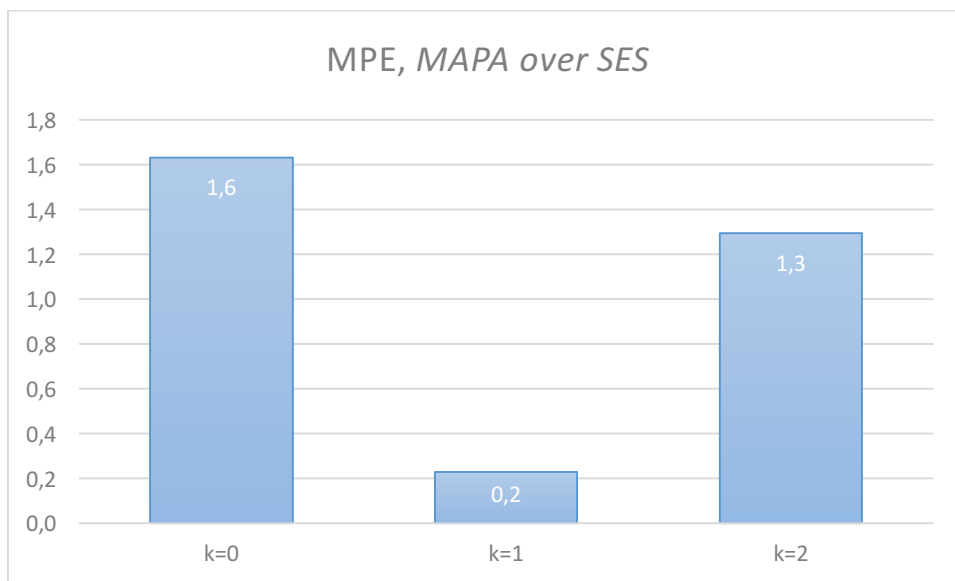
Διαπιστώνουμε από τα δύο παραπάνω διαγράμματα ότι και ως προς τους ορίζοντες πρόβλεψης η συνδυασμένη SES έχει καλύτερα αποτελέσματα και ως προς την ακρίβεια και ως προς την προκατάληψη των προβλέψεων. Πιο έντονα παρατηρείται στον δείκτη sMAPE αλλά και στον δείκτη MPE στους πρώτους δύο ορίζοντες $f=1,2$ η διαφορά των σφαλμάτων είναι αρκετά μεγάλη.

Αντίστοιχα αποτελέσματα βλέπουμε και στα διαγράμματα ως προς τον τελευταίο παράγοντα των δεδομένων μας, τα επίπεδα διάρθρωσης. Στα επόμενα δύο διαγράμματα παίρνω τον μέσο

όρο των σφαλμάτων για όλους τους ορίζοντες πρόβλεψης και για τις τρεις μεθόδους αλλά για κάθε επίπεδο ξεχωριστά.



Εικόνα 26 sMAPE, MAPA over SES (k)

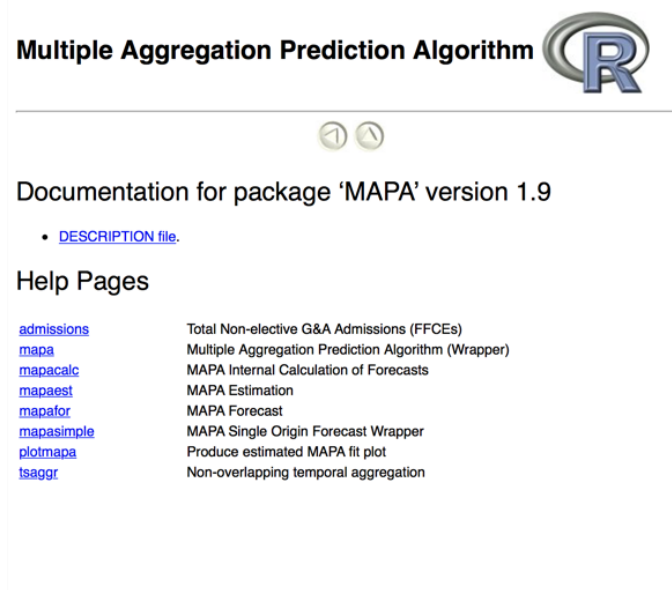



Εικόνα 27 MPE, MAPA over SES (k)

Στο δεύτερο διάγραμμα στο επίπεδο $k=1$ φαίνεται ότι η απλή SES υπερτερεί της συνδυασμένης ωστόσο κατά γενικό κανόνα η συνδυασμένη SES με τον αλγόριθμο MAPA αποδίδει καλύτερα. Στο πρώτο διάγραμμα με τον δείκτη sMAPE βλέπουμε ότι η απλή SES έχει 2,7 φορές μεγαλύτερο σφάλμα από την συνδυασμένη.

6.3 ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ (*RStudio*)

Το *RStudio* (www.rstudio.com) είναι ένα δωρεάν προγραμματιστικό περιβάλλον για στατιστική ανάλυση δεδομένων χρησιμοποιώντας την γλώσσα *R*. Δημιουργήθηκε το 2011 από τον JJ Allaire, με την χρήση της C++ και του Qt για το γραφικό περιβάλλον, και από τότε έχουν κυκλοφορήσει διάφορες εκδόσεις τόσο για περιβάλλον Windows όσο και για Mac OS. Η λειτουργία του στηρίζεται στην χρήση έτοιμων βιβλιοθηκών οι οποίες περιλαμβάνουν συναρτήσεις σχετικές με το αντικείμενο της εκάστοτε βιβλιοθήκης. Ανάλογα με τα δεδομένα που θέλω να αναλύσω και τον τομέα στον οποίο δραστηριοποιούμαι χρησιμοποιώ και τις αντίστοιχες βιβλιοθήκες. Εδώ χρησιμοποιήσαμε την βιβλιοθήκη *forecast* (<https://github.com/robjhyndman/forecast>) που περιλαμβάνει τις μεθόδους προβλέψεων (*Naïve*, *SES*) που θέλω να χρησιμοποιήσω αλλά και τους δείκτες σφαλμάτων (*accuracy*) για την αξιολόγηση αυτών. Επίσης χρησιμοποιήσαμε την βιβλιοθήκη *MAPA* (<http://kourentzes.com/forecasting/2014/04/19/multiple-aggregation-prediction-algorithm-mapa/>) που περιλαμβάνει όλες τις συναρτήσεις σχετικά με τον αλγόριθμο αυτόν. Παρακάτω παρουσιάζονται κάποιες εικόνες από το περιβάλλον του και από τις βιβλιοθήκες που χρησιμοποιήθηκαν.



Multiple Aggregation Prediction Algorithm 

Documentation for package 'MAPA' version 1.9

- [DESCRIPTION file.](#)

Help Pages

admissions	Total Non-elective G&A Admissions (FFCEs)
mapa	Multiple Aggregation Prediction Algorithm (Wrapper)
mapacalc	MAPA Internal Calculation of Forecasts
mapaes	MAPA Estimation
mapafor	MAPA Forecast
mapasimple	MAPA Single Origin Forecast Wrapper
plotmapa	Produce estimated MAPA fit plot
tsaggr	Non-overlapping temporal aggregation

Εικόνα 28 Βιβλιοθήκη *MAPA*

Από τις συναρτήσεις που περιλαμβάνονται στην βιβλιοθήκη *MAPA* χρησιμοποιήθηκε κυρίως η συνάρτηση *mapasimple* που φαίνεται στην παρακάτω εικόνα. Οι κύριες παράμετροι που δέχεται η συνάρτηση είναι η χρονοσειρά που θέλουμε να προβλέψουμε, ο ορίζοντας πρόβλεψης που θα χρησιμοποιήσουμε, η συχνότητα της χρονοσειράς μας και το μοντέλο πρόβλεψης που θα χρησιμοποιήσουμε. Υπάρχουν και άλλες σημαντικές παράμετροι που χρειάζεται η συνάρτηση, όπως οι συντελεστές βαρύτητας που χρησιμοποιεί ο αλγόριθμος, ωστόσο δεν θα ασχοληθούμε με αυτούς στην παρούσα μελέτη.

MAPA Single Origin Forecast Wrapper

Description

The following function estimates MAPA and produces out-of-sample forecasts. It is not suggested, unless only simple single origin forecasts are required.

Usage

```
mapasimple(y, ppy, fh = ppy, minimumAL = 1, maximumAL = ppy, comb = "mean",
           output = "forecast", paral = 0, display = 0, outplot = 1,
           hybrid = TRUE, model = "ZZZ")
```

Εικόνα 29 Συνάρτηση *mapasimple*

Forecasting Functions for Time Series and

Linear Models



Documentation for package 'forecast' version 6.2

- [DESCRIPTION file](#).

Help Pages

accuracy	Accuracy measures for forecast model
Acf	(Partial) Autocorrelation Function Estimation
arfima	Fit a fractionally differenced ARFIMA model
Arima	Fit ARIMA model to univariate time series
arima.errors	ARIMA errors

Εικόνα 30 Βιβλιοθήκη *forecast*

Η βιβλιοθήκη *forecast* περιλαμβάνει όλες τις συναρτήσεις σχετικά με τις προβλέψεις τιμών, τα μοντέλα και τους δείκτες σφαλμάτων. Μια από τις συναρτήσεις που θα χρησιμοποιηθεί στην μελέτη μας είναι η συνάρτηση *accuracy* η οποία περιλαμβάνει τους δείκτες σφαλμάτων που θα χρησιμοποιήσουμε για την αξιολόγηση των προβλέψεων. Στις επόμενες δύο εικόνες παρουσιάζεται η συνάρτηση *accuracy* και οι δείκτες σφαλμάτων που υπολογίζει όπως φαίνονται στο *R Studio*. Ξεχωριστά υπολογίζουμε τον δείκτη *sMAPE*, ο οποίος δεν υπολογίζεται άμεσα από την συνάρτηση αυτή.

Accuracy measures for forecast model

Description

Returns range of summary measures of the forecast accuracy. If x is provided, the function measures out-of-sample (test set) forecast accuracy based on x - f . If x is not provided, the function only produces in-sample (training set) accuracy measures of the forecasts based on $f["x"]$ -fitted(f). All measures are defined and discussed in Hyndman and Koehler (2006).

Usage

```
accuracy(f, x, test=NULL, d=NULL, D=NULL)
```

Εικόνα 31 Συνάρτηση accuracy

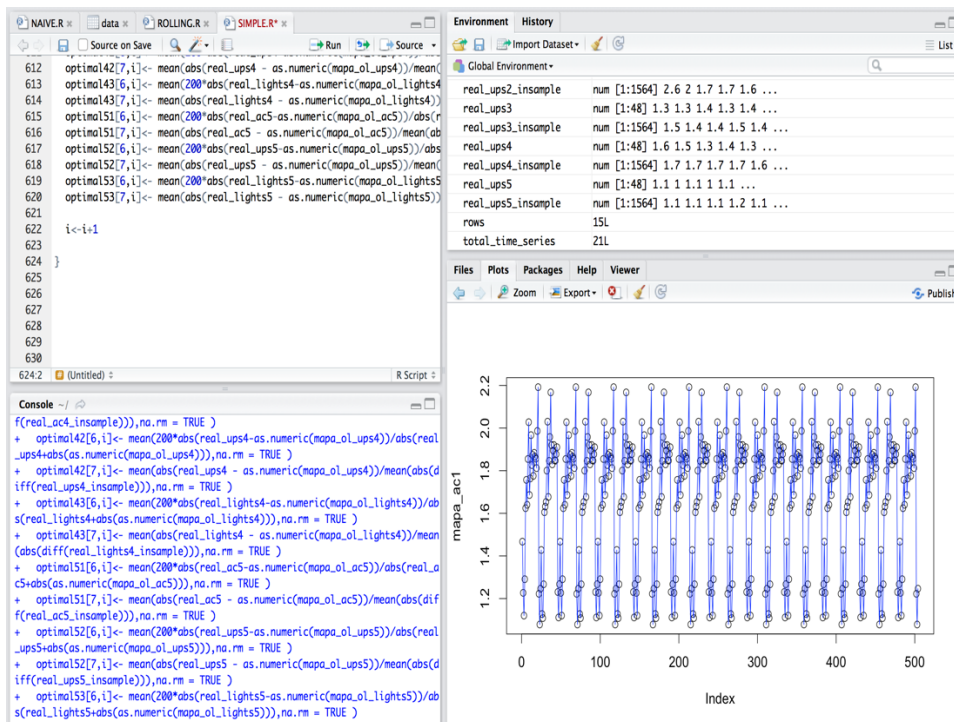
Details

The measures calculated are:

- ME: Mean Error
- RMSE: Root Mean Squared Error
- MAE: Mean Absolute Error
- MPE: Mean Percentage Error
- MAPE: Mean Absolute Percentage Error
- MASE: Mean Absolute Scaled Error
- ACF1: Autocorrelation of errors at lag 1.

Εικόνα 32 Δείκτες Σφαλμάτων

Παρακάτω φαίνεται το γραφικό περιβάλλον του *R Studio* το οποίο χωρίζεται σε τέσσερα μέρη. Στο πάνω αριστερό μέρος έχουμε τον κώδικα που έχουμε γράψει και τα δεδομένα που έχουμε εισάγει. Στο κάτω αριστερό μέρος έχουμε το κομμάτι του κώδικα το οποίο τρέχει στη δεδομένη στιγμή. Στην δεξιά πλευρά της εικόνας έχουμε στο επάνω μέρος τις μεταβλητές που έχουν δημιουργηθεί και αποθηκευτεί και στο κάτω μέρος έχουμε τις γραφικές παραστάσεις από τον κώδικα που τρέξαμε. Στο σημείο αυτό μπορούμε αντί των γραφικών παραστάσεων να αναζητήσουμε πληροφορίες για κάποια συνάρτηση ή βιβλιοθήκη μέσω του *Help*, ενώ στο επάνω μέρος μπορούμε να δούμε το ιστορικό των εντολών που έχουμε τρέξει στο πρόγραμμα.



Εικόνα 33 Περιβάλλον R-Studio

Το περιβάλλον του *R Studio* είναι αρκετά απλό και εύχρηστο και θυμίζει το περιβάλλον του *MatLab*. Αποτελέσε βασικό εργαλείο για την μελέτη αυτή όπως και κάθε μελέτη που αφορά τις τεχνικές προβλέψεων καθώς αποτελεί το πιο σημαντικό εργαλείο ανάλυσης δεδομένων. Με βάση τη λειτουργία του προγράμματος εκτελείται η μεθοδολογία που αναλύθηκε σε παραπάνω κεφάλαιο. Εκτός από το *R Studio* χρησιμοποιήθηκε και το *Microsoft Excel* μέσω του οποίου δημιουργήθηκαν τα διαγράμματα που παρουσιάστηκαν στο παρόν κεφάλαιο σε παραπάνω ενότητα με σκοπό την σύγκριση και την αξιολόγηση των μεθόδων.

ΚΕΦΑΛΑΙΟ 7: ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΠΡΟΕΚΤΑΣΕΙΣ

7.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΕΡΓΑΣΙΑΣ

Στην παρούσα εργασία έγινε πρόβλεψη ηλεκτρικής κατανάλωσης κτιρίων με τον συνδυασμό μεθόδων χρονικής συνάθροισης (*temporal aggregation*), της μετάπτωσης δηλαδή των αρχικών δεδομένων σε πολλαπλές χαμηλότερες συχνότητες και την παραγωγή προβλέψεων σε κάθε μία εξ αυτών ξεχωριστά, και ιεραρχικών μεθόδων πρόβλεψης (*cross sectional aggregation*). Σκοπός ήταν να εξετάσουμε κατά πόσο ο συνδυασμός αυτός μπορεί να έχει θετικές επιδράσεις στην αξιοπιστία των προβλέψεων και στην μείωση των σφαλμάτων. Για το λόγο αυτό εξετάστηκαν οι συχνότερα εμφανιζόμενες στην βιβλιογραφία μέθοδοι *top-down*, *bottom-up* και *optimal* σε συνδυασμό με τον αλγόριθμο *MAPA*. Ως μέθοδο πρόβλεψης επιλέξαμε την *SES* γιατί τα δεδομένα μας δεν χαρακτηρίζονται από τάση, είναι γρήγορα μεταβαλλόμενα και γιατί δεν επηρεάζεται η εποχικότητα.

Παρατηρούμε από την πραγματοποίηση της μελέτης ότι τα συμπεράσματα που προκύπτουν είναι αυτά που περιμέναμε. Αρχικά από τα διαγράμματα που παρουσιάζονται στο παραπάνω κεφάλαιο παρατηρούμε ότι οι ιεραρχικές μέθοδοι είναι κατάλληλες για χρήση σε συγκεκριμένα επίπεδα δεδομένων. Παρατηρούμε δηλαδή ότι στο επίπεδο $k=0$ η μέθοδος *top-down* έχει καλύτερα αποτελέσματα από τις άλλες δυο. Αντίστοιχα στο επίπεδο $k=1$ καλύτερα αποτελέσματα παρουσιάζει η *optimal* και στο επίπεδο $k=2$ η *bottom-up* και είναι λογικό αν σκεφτούμε τον τρόπο με τον οποίο γίνονται οι προβλέψεις στην κάθε μέθοδο.

Όσον αφορά τον αλγόριθμο *MAPA* δύο είναι τα κύρια συμπεράσματα που προκύπτουν από την μελέτη. Αποδεικνύεται και μέσω των αποτελεσμάτων το γεγονός ότι ο αλγόριθμος αυτός δίνει βάρος στα ποιοτικά χαρακτηριστικά της χρονοσειράς και όχι στα ίδια τα δεδομένα. Βλέπουμε ότι όσο αυξάνουμε τον ορίζοντα πρόβλεψης τόσο μειώνονται τα σχετικά σφάλματα και αυτό συμβαίνει γιατί όσο αυξάνουμε τον ορίζοντα τόσο περισσότερο διακρίνονται τα ποιοτικά χαρακτηριστικά των χρονοσειρών. Επίσης παρατηρούμε από τα παραπάνω διαγράμματα ότι η μείωση των σφαλμάτων από την χρήση του αλγορίθμου σε σχέση με την απλή *SES* ανέρχεται έως και 50%.

Τέλος το γενικό συμπέρασμα που προκύπτει από τη μελέτη αυτή είναι ότι ο συνδυασμός της χρονικής με την ιεραρχική συνάθροιση επιφέρει μεγάλη βελτίωση τόσο στην ακρίβεια όσο και στην προκατάληψη των προβλέψεων. Τα ποσοστά βελτίωσης μεταβάλλονται ανάλογα με τον ορίζοντα πρόβλεψης και τις μεθόδους που χρησιμοποιούμε ωστόσο το γενικό συμπέρασμα παραμένει, η βελτίωση δηλαδή της ακρίβειας και της προκατάληψης των προβλέψεων.

7.2 ΠΡΟΕΚΤΑΣΕΙΣ

Όπως αναφέραμε και παραπάνω σκοπός της μελέτης ήταν ο συνδυασμός της χρονικής με την ιεραρχική συνάθροιση. Τρία στοιχεία τα οποία χαρακτηρίζουν και επηρεάζουν τα αποτελέσματα της μελέτης είναι οι μέθοδοι, οι οριζόντες πρόβλεψης και οι συντελεστές βαρύτητας που χρησιμοποιήθηκαν. Και για τα τρία αυτά στοιχεία υπάρχουν πολλές επιλογές που μπορεί κάποιος να κάνει ανάλογα με το τι επιδιώκει να πετύχει ή τι επιλέγει να μελετήσει. Και μόνο η αλλαγή ενός από τα στοιχεία μπορεί να αποτελέσει μια ξεχωριστή μελέτη με διαφορετικά αποτελέσματα.

Επιλέξαμε συγκεκριμένες μεθόδους και αλγόριθμους από την καθεμιά και συνδυάσαμε για να δούμε το κατά πόσο επηρεάζονται οι προβλέψεις μας θετικά ή αρνητικά. Η συγκεκριμένη μελέτη θα μπορούσε να ξαναγίνει με τον ίδιο σκοπό αλλά συνδυάζοντας διαφορετικές μεθόδους ιεραρχικής και χρονικής συνάθροισης. Εδώ χρησιμοποιήθηκαν μέθοδοι πιο γνωστές στην βιβλιογραφία γεγονός το οποίο δεν δεσμεύει τον όποιο ενδιαφερόμενο θελήσει να πραγματοποιήσει την ίδια μελέτη χρησιμοποιώντας διαφορετικές μεθόδους. Οι μέθοδοι *top-down*, *bottom-up* και *optimal* καθώς και ο αλγόριθμος *MAPA* είναι πολύ γνωστά στην βιβλιογραφία αλλά δεν είναι τα μοναδικά.

Επίσης για την πραγματοποίηση της μελέτης θεωρήσαμε δεδομένους τους συντελεστές βαρύτητας που χρησιμοποιεί ο αλγόριθμος *MAPA* και δεν ασχοληθήκαμε με την αξιολόγηση τους. Σε κάποιο αντίστοιχο πείραμα θα μπορούσε να υπάρχει ολόκληρο κεφάλαιο το οποίο θα ασχολείται με την επιλογή κατάλληλων συντελεστών βαρύτητας που χρησιμοποιεί ο αλγόριθμος ανάλογα με τον οριζόντα πρόβλεψης που επιλέγεται ώστε η επίδραση των διαθέσιμων δεδομένων να είναι πιο σωστά κατανεμημένη ανάλογα με την χρονική απόκλιση που έχουν.

Τέλος θα μπορούσαμε να κάνουμε αξιολόγηση της μεθόδου χρησιμοποιώντας πολύ μικρότερους οριζόντες πρόβλεψης. Να κάναμε δηλαδή προσομοίωση ενός *real time* συστήματος το οποίο δέχεται αυτομάτως το *feedback* και αναπροσαρμόζεται.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Hai-xiang Zhao, Frederic Magoules 2011: A review on the prediction of building energy consumption, Applied Mathematics and Systems Laboratory, France
2. Al-Homoud MS. Computer-aided building energy analysis techniques. Building and Environment 2001;36(4):421–33
3. Kimbara A, Kurosu S, Endo R, Kamimura K, Matsuba T, Yamada A. On-line prediction for load profile of an air-conditioning system. ASHRAE Transactions 1995; 101(2):198–207.
4. Hoffman AJ. Peak demand control in commercial buildings with target peak adjustment based on load forecasting. In: Proceedings of the 1998 IEEE international conference on control applications, vol. 2. 1998. p. 1292–6.
5. Aydinalp-Koksal M, Ugursal VI. Comparison of neural network, conditional demand analysis and engineering approaches for modeling end-use energy consumption in the residential sector. Applied Energy 2008; 85(4): 271–96.
6. Kalogirou SA, Bojic M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy 2000;25(5):479–91
7. Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings 2005; 37(5):545–53.
8. Wang X, Chen Z, Yang C, Chen Y. Gray predicting theory and application of energy consumption of building heat-moisture system. Building and Environment 1999; 34(4):417–20.
9. Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: a decomposition approach to forecasting. International Journal of Forecasting 16 (4), 521 {530.
10. Box, George; Jenkins, Gwilym (1970). Time series analysis: Forecasting and control, San Francisco: Holden-Day
11. Commandeur J.J.F.; Koopman S.J. (2007). Introduction to State Space Time Series Analysis (Oxford University Press)
12. Zotteri, G., Kalchschmidt, M., Caniato, F., 2005. The impact of aggregation level on forecasting performance. International Journal of Production Economics 93 94, 479 {491.
13. Petropoulos, F., Kourentzes, N., 2014. Improving forecasting via multiple temporal aggregation. The International Journal of Applied Forecasting In press.
14. Athanasopoulos, G., Ahmed, R. A., Hyndman, R. J., 2009. Hierarchical forecasts for Australian domestic tourism. International Journal of Forecasting 25 (1), 146 {166.
15. Dunn, D., Williams, W., DeChaine, T., 1976. Aggregate versus subaggregate models in local area forecasting. Journal of the American Statistical Association 71 (353), 68 {71.
16. D'Attilio, D. F., 1989. Practical applications of trend analysis in business forecasting. The Journal of Business Forecasting Methods & Systems 8, 9 {11.
17. Schwarzkopf, A. B., Tersine, R. J., Morris, J. S., 1988. Top-down versus bottom-up forecasting strategies. International Journal of Production Research 26 (11), 1833 {1843.
18. Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., Shang, H. L., 2011. Optimal combination forecasts for hierarchical time series. Computational Statistics & Data Analysis 55 (9), 2579 {2589.
19. Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., Assimakopoulos, V., 2011. An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. Journal of the Operational Research Society 62 (3), 544 {554.

20. Abraham, B., 1982. Temporal aggregation and time series. *International Statistical Review* 50, 285 {291.
21. Babai, M. Z., Ali, M. M., Nikolopoulos, K., 2012. Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega* 40 (6), 713 {721.
22. Luiz, K.H., Pedro, A.M., Pedro, L.V.P., 1992. The Effect of Overlapping Aggregation on Time Series Models: An Application to the Unemployment Rate in Brazil. *Brazilian Review of Econometrics* 12, 223-241.
23. Brannas, K., Hellstrom, J., Nordstrom, J., 2002. A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting* 18, 19-30.
24. Mohammadipour, M., Boylan, J.E., 2012. Forecast horizon aggregation in integer autoregressive moving average (INARMA) models. *OMEGA* 40, 703-712.
25. 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.
26. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1982. The accuracy of extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting* 1 (2), 111{153.
27. Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 16 (4), 451 {476.
28. Athanasopoulos, G., Hyndman, R. J., Song, H., Wu, D. C., 2011. The tourism forecasting competition. *International Journal of Forecasting* 27 (3), 822 {844.