ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

Βραχυπρόθεσμη πρόβλεψη ζήτησης ηλεκτρικής ενέργειας στην Ελλάδα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ιωάννης Εγγονόπουλος

Επιβλέπων: Μαμάσης Νικόλαος
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2020

ΠΕΡΙΛΗΨΗ

Η πρόβλεψη της ζήτησης φορτίου ηλεκτρικής ενέργειας αποτελεί έναν τομέα της επιστήμης των προβλέψεων που παρουσιάζει ιδιαίτερο ενδιαφέρον μετά την απελευθέρωση των αγορών ενέργειας και το ανταγωνιστικό περιβάλλον που έχει δημιουργηθεί. Η εκ των προτέρων γνώση της συμπεριφοράς φορτίου είναι ζωτικής σημασίας για τον σχεδιασμό, την ανάλυση και τη λειτουργία των συστημάτων ισχύος, ώστε να εξασφαλιστεί μια αδιάκοπη, αξιόπιστη, ασφαλής και οικονομική παροχή ηλεκτρικής ενέργειας.

Διάφορες τεχνικές έχουν αναπτυχθεί για τη βραχυπρόθεσμη πρόβλεψη της ηλεκτρικής κατανάλωσης (STLF). Οι τεχνικές αυτές διακρίνονται κυρίως σε στατιστικές μεθόδους πρόβλεψης και μοντέλα τεχνητής νοημοσύνης (AI). Το φορτίο ενός συστήματος αποτελεί μια στοχαστική, μη στάσιμη διαδικασία, ενώ η συμπεριφορά του επηρεάζεται από πολλές μεταβλητές όπως: οι καιρικές συνθήκες, οι κοινωνικοοικονομικές συνθήκες, το είδος του τελικού χρήστη, οι εποχιακοί παράγοντες και τα τυχαία γεγονότα. Λόγω των μη γραμμικών σχέσεων που εμφανίζονται μεταξύ της ηλεκτρικής κατανάλωσης και των επιμέρους παραμέτρων, τα τελευταία χρόνια επιλέγονται μέθοδοι μηχανικής μάθησης (Machine Learning) που βασίζονται στα διαθέσιμα δεδομένα και εκπαιδεύονται μέσα από αυτά. Το γεγονός αυτό τα καθιστά κατάλληλα για την επίλυση τέτοιων προβλημάτων.

Αντικείμενο της διπλωματικής εργασίας είναι η χρονική διερεύνηση της ζήτησης ηλεκτρικής ενέργειας στην Ελλάδα και η ανάπτυξη μιας μεθοδολογίας για τη βραχυπρόθεσμη πρόβλεψη της ενεργειακής ζήτησης. Το πρόβλημα της εξοικονόμησης ενέργειας για οικονομικούς και περιβαλλοντικούς λόγους, καθιστά την ύπαρξη μιας τέτοιας μεθοδολογίας εξαιρετικά σημαντική. Σκοπός επομένως της παρούσας εργασίας, είναι η παράθεση διαφόρων μοντέλων με σκοπό τη βέλτιστη προσέγγιση των πραγματικών δεδομένων. Το χρονικό διάστημα στο οποίο γίνεται η μελέτη είναι από το 2006 έως και το 2016 και αφορά τα δεδομένα της Ελλάδας. Αρχικά, επιχειρείται η εύρεση συσχετίσεων μεταξύ της ηλεκτρικής κατανάλωσης και ορισμένων βασικών μεταβλητών που την επηρεάζουν (θερμοκρασία, εποχή, είδος ημέρας κ.α.). Στη συνέχεια, πραγματοποιείται η βραχυπρόθεσμη πρόβλεψη της ενεργειακής ζήτησης μέσω αλγορίθμων και πακέτων με χρήση της γλώσσας R (2017), που αποτελεί δημοφιλή γλώσσα μηχανικής μάθησης. Συγκεκριμένα, αφού εξετάσουμε ένα μοντέλο αναφοράς με την απλοϊκή μέθοδο (naïve), θα δημιουργήσουμε μοντέλα γραμμικής παλινδρόμησης (linear regression) και μοντέλα random forest τα οποία θα εξετάσουμε σε βάθος. Ένας από τους βασικούς στόχους της μελέτης είναι να διερευνήσουμε πως η απόδοση των μοντέλων σχετίζεται με την επιλογή μεταβλητών σε πρόβλεψη ενός βήματος σύντομων χρονοσειρών (one-step forecasting of short time series). Τέλος, προτείνεται το καταλληλότερο μοντέλο και γίνεται αξιολόγηση των αποτελεσμάτων που προκύπτουν από την ανάλυση. Τα δεδομένα και ο κώδικας από τον οποίον προέκυψαν τα αποτελέσματα της εργασίας, διατίθενται στον παρακάτω σύνδεσμο: **https://doi.org/10.6084/m9.figshare.12630503.v2**

Λέξεις-κλειδιά: διερευνητική ανάλυση δεδομένων, βραχυπρόθεσμη πρόβλεψη, ενεργειακή ζήτηση, μηχανική μάθηση, γραμμική παλινδρόμηση, random forests

ABSTRACT

Future energy demand forecasting is a sector of forecasting science that is of particular interest, especially after the deregulation of energy markets and the competitive environment that has been created. Knowing the load behavior in advance is crucial in planning, analysis and operation of power systems so as to assure an uninterrupted, reliable, secure and economic supply of electricity.

Several methods have been developed for the short-term load forecasting (STLF). These methods are mainly divided in statistical forecasting methods and artificial intelligence (AI) methods. The load of a system is a stochastic, non-stationary process, whose behavior is affected by a number of factors. The most important among these are the weather conditions, the type of the end user, the seasonal factors and random events. Due to the non-linear relations between electric consumption and its components, particular emphasis has been given in the last years on machine learning methods to solve these problems. The machine learning methods are based on the available data and trained through these.

The objective of the diploma thesis is the temporal energy analysis in Greece and the development of a methodology for short-term load forecasting. The problem of energy saving for economic and environmental reasons makes the existence of such a methodology extremely important. The purpose of this project is therefore to create various models in order to optimally approach real data. The time period in which the study takes place is from 2006 to 2016 and concerns the data of Greece. Initially, an attempt is made to analyze the correlations between demand load and certain key parameters that affect it (temperature, season, type of day, etc.). This is followed by a short-term load forecast through algorithms and packages using the R programming language (2017), which is a popular language for machine learning. Specifically, after examining a naïve model (benchmark), we will create linear regression models and random forest models which we will examine in depth. The primary aim of the study is to investigate how the performance of our models is related to the variable selection in one step forecasting of short time series. Finally, the most appropriate model is proposed and the results of the analysis are evaluated. The data and the code, from which the results of the work emerged, are available at the following link: **https://doi.org/10.6084/m9.figshare.12630503.v2**

Keywords: exploratory data analysis, short-term forecasting, energy demand, machine learning, linear regression, random forests

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2019-2020 στον τομέα Υδατικών Πόρων και Περιβάλλοντος της Σχολής Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Υπεύθυνος κατά την εκπόνηση της διπλωματικής ήταν ο Καθηγητής κ. Μαμάσης Νικόλαος στον οποίο οφείλω ιδιαίτερες ευχαριστίες για την ανάθεση αυτής και τη δυνατότητα που μου δόθηκε να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα. Θα ήθελα επίσης να ευχαριστήσω θερμά τον επιβλέποντα της διπλωματικής Τύραλη Χρήστο για την υποστήριξη και την καθοδήγηση που μου παρείχε κατά τη συγγραφή της εργασίας.

Σε αυτό το σημείο θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένειά μου για την αμέριστη στήριξή τους όλα αυτά τα χρόνια ώστε να προσβλέπω δυναμικά στην υλοποίηση των επιδιωκόμενων στόχων μου. Τέλος, θα ήταν παράλειψη να μην ευχαριστήσω όλους τους δικούς μου ανθρώπους για τη διάθεση και την κατανόηση που επέδειξαν σε όλη μου την πορεία.

Ιωάννης Εγγονόπουλος

Φεβρουάριος 2020

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ i

ABSTRACT ii

ΠΡΟΛΟΓΟΣ iii

ΠΕΡΙΕΧΟΜΕΝΑ ΣΧΗΜΑΤΩΝ vi

ΠΕΡΙΕΧΟΜΕΝΑ ΠΙΝΑΚΩΝ ix

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ 1

1.1 Αντικείμενο και Σκοπός 3

1.2 Δομή διπλωματικής εργασίας 7

ΚΕΦΑΛΑΙΟ 2: ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ 9

2.1 Εξόρυξη Δεδομένων 11

2.1.1 Εισαγωγή στην Εξόρυξη Δεδομένων 11

2.1.2 Σημασία της Εξόρυξης Δεδομένων 13

2.1.3 Μοντέλα και Μέθοδοι Εξόρυξης Δεδομένων 14

2.1.3.1 Ταξινόμηση 15

2.1.3.2 Παλινδρόμηση 16

2.1.3.3 Ανάλυση Χρονοσειρών 16

2.1.3.4 Πρόβλεψη 16

2.1.3.5 Συσταδοποίηση 16

2.1.3.6 Περίληψη 17

2.1.3.7 Κανόνες Συσχετίσεων 17

2.1.3.8 Ανακάλυψη Ακολουθιών 17

2.2 Εξαγωγή Δεδομένων 18

2.2.1 Ζήτηση Ενέργειας 18

2.2.2 Ακαθάριστο Εθνικό Προϊόν 18

2.2.3 Θερμοκρασία 19

2.3 Επεξεργασία Δεδομένων 20

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ 22

3.1 Οπτική Διερεύνηση για διάφορες χρονικές κλίμακες 24

3.2 Ζήτηση Ενέργειας και ΑΕΠ 31

3.3 Ζήτηση Ενέργειας και Θερμοκρασία 32

ΚΕΦΑΛΑΙΟ 4: ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ 35

4.1 Στατιστικές Μέθοδοι Πρόβλεψης 37

4.1.1 Απλοϊκή Μέθοδος 39

4.1.2 Μοντέλα Παλινδρόμησης 40

4.1.2.1 Μοντέλα Απλής Γραμμικής Παλινδρόμησης 40

4.1.2.2 Μοντέλα Πολλαπλής Γραμμικής Παλινδρόμησης 43

4.2 Μοντέλα Τεχνητής Νοημοσύνης.....53

4.2.1 Μοντέλα Random Forest.....53

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ 63

5.1 Συμπεράσματα 65

5.2 Προοπτικές 66

ΒΙΒΛΙΟΓΡΑΦΙΑ ..... .67

ΠΕΡΙΕΧΟΜΕΝΑ ΣΧΗΜΑΤΩΝ

Σχήμα 2.1: Διαδικασία εξόρυξης δεδομένων 13

Σχήμα 2.2: Τα βασικότερα βήματα της διαδικασίας Εξόρυξης Γνώσης 14

Σχήμα 2.3: Ταξινόμηση των Τεχνικών Εξόρυξης Δεδομένων 15

Σχήμα 2.4: Τα στάδια της διαδικασίας Συσταδοποίησης 17

Σχήμα 2.5: Πρωτογενή δεδομένα ωριαίας ζήτησης ενέργειας για την Ελλάδα..... 18

Σχήμα 2.6: Πρωτογενή δεδομένα ετήσιου ΑΕΠ για την Ελλάδα 19

Σχήμα 2.7: Θέση του σταθμού της Ηλιούπολης 20

Σχήμα 2.8: Ωριαία ζήτηση ενέργειας..... 21

Σχήμα 2.9: Ιστόγραμμα ωριαίας ζήτησης ενέργειας 21

Σχήμα 3.1: Μέση ωριαία ζήτηση ενέργειας ανά μήνα 25

Σχήμα 3.2: Κατανομή στην ημέρα μέσης ωριαίας ζήτησης ενέργειας ανά μήνα 25

Σχήμα 3.3: Μέση ημερήσια ζήτηση ενέργειας ανά μήνα..... ..26

Σχήμα 3.4: Κατανομή μέσης ημερήσιας ζήτησης ενέργειας ανά μήνα..... ..26

Σχήμα 3.5: Μέση μηνιαία ζήτηση ενέργειας ανά υδρολογικό έτος..... ..27

Σχήμα 3.6: Κατανομή μέσης μηνιαίας ζήτησης ενέργειας στο υδρολογικό έτος 27

Σχήμα 3.7: Μέση μηνιαία ζήτηση ενέργειας ανά υδρολογικό έτος, για τα υδρολογικά έτη 2006, 2009, 2012 και 2015 28

Σχήμα 3.8: Κατανομή μέσης μηνιαίας ζήτησης ενέργειας στο υδρολογικό έτος, για τα υδρολογικά έτη 2006, 2009, 2012 και 2015 28

Σχήμα 3.9: Boxplot με τα ποσοστημόρια 0.25 και 0.75 για τις μέσες μηνιαίες τιμές ζήτησης ενέργειας..... ..29

Σχήμα 3.10: ACF (αριστερά) και PACF (δεξιά) για τη ζήτηση ενέργειας σε ωριαία κλίμακα (άνω), σε ημερήσια κλίμακα (μέση) και μηνιαία κλίμακα (κάτω) 30

Σχήμα 3.11: Μηνιαία μεταβολή ημερήσιας ζήτησης ενέργειας 30

Σχήμα 3.12: Μεταβολή μηνιαίας ζήτησης ενέργειας ανά έτος (άνω) και μηνιαίας διακύμανσης της ίδιας μεταβολής ανά μήνα (κάτω) 31

Σχήμα 3.13: Ακαθάριστο Εθνικό Προϊόν (άνω) και μέση ετήσια ζήτηση ενέργειας (κάτω) ανά υδρολογικό έτος..... ..32

Σχήμα 3.14: Μέση ημερήσια θερμοκρασία στο σταθμό Ηλιούπολης 33

Σχήμα 3.15: Μέση μηνιαία θερμοκρασία ανά υδρολογικό έτος στο σταθμό Ηλιούπολης....33

Σχήμα 3.16: Μέση μηνιαία ζήτηση ενέργειας προς Ακαθάριστο Εθνικό Προϊόν ανά υδρ. έτος..... ..33

Σχήμα 3.17: Διάγραμμα μέσης ημερήσιας θερμοκρασίας στο σταθμό Ηλιούπολης και ημερήσιας ζήτησης ενέργειας..... ..34

Σχήμα 4.1: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (naïve) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 39

Σχήμα 4.2: Διάγραμμα μέσης ημερήσιας θερμοκρασίας και μέσης ημερήσιας ζήτησης ενέργειας ανά ΑΕΠ 42

Σχήμα 4.3: Διάγραμμα μοντέλου GAM προσαρμοσμένο στη μέση ημερήσια θερμοκρασία με τη μέση ημερήσια ζήτηση ενέργειας ανά ΑΕΠ..... ..42

Σχήμα 4.4: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_1) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 43

Σχήμα 4.5: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_2) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 45

Σχήμα 4.6: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_3) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 45

Σχήμα 4.7: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_4) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 46

Σχήμα 4.8: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_5) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 46

Σχήμα 4.9: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_6) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 47

Σχήμα 4.10: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_7) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 47

Σχήμα 4.11: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_8) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 48

Σχήμα 4.12: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_9) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 48

Σχήμα 4.13: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_10) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 49

Σχήμα 4.14: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_11) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 50

Σχήμα 4.15: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_12) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 50

Σχήμα 4.16: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_13) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 51

Σχήμα 4.17: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_14) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 51

Σχήμα 4.18: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_1) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 54

Σχήμα 4.19: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_2) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 55

Σχήμα 4.20: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_3) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 55

Σχήμα 4.21: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_4) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 56

Σχήμα 4.22: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_5) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 56

Σχήμα 4.23: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_6) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 57

Σχήμα 4.24: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_7) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 57

Σχήμα 4.25: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_8) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 58

Σχήμα 4.26: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_9) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 58

Σχήμα 4.27: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_10) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 59

Σχήμα 4.28: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_11) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 59

Σχήμα 4.29: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_12) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 60

Σχήμα 4.30: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_13) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 60

Σχήμα 4.31: Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_14) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ 61

ΠΕΡΙΕΧΟΜΕΝΑ ΠΙΝΑΚΩΝ

Πίνακας 2.1: Ελληνικό Ετήσιο Ακαθάριστο Εθνικό Προϊόν (Πηγή: The World Bank 2016) .. 19

Πίνακας 4.1: Στατιστικά μεγέθη για τα γραμμικά μοντέλα..... 52

Πίνακας 4.2: Στατιστικά μεγέθη για τα μοντέλα random forest 61

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Αντικείμενο και Σκοπός

Η ακριβής πρόβλεψη της μελλοντικής ενεργειακής ζήτησης μετά την απελευθέρωση των αγορών ηλεκτρικής ενέργειας κατέχει πολύ σημαντικό ρόλο όσον αφορά τους τομείς των επενδύσεων σε συστήματα διανομής, τον προγραμματισμό του φορτίου ηλεκτρικής ενέργειας και τις στρατηγικές διαχείρισης σε περιφερειακά και εθνικά συστήματα. Το σφάλμα της πρόβλεψης του ηλεκτρικού φορτίου μπορεί να οδηγήσει στην αύξηση του λειτουργικού κόστους. Σε περίπτωση υποεκτίμησης των μελλοντικών φορτίων η αδυναμία παροχής επαρκών αποθεμάτων συνεπάγεται υψηλό κόστος ανά μονάδα αιχμής. Αντίθετα, υπερεκτίμηση της ζήτησης μπορεί να οδηγήσει σε αχρησιμοποίητη δυναμικότητα που σημαίνει ότι γίνεται σπατάλη πόρων, κυρίως οικονομικών. Ως εκ τούτου, οι μελέτες πρόβλεψης ενεργειακής ζήτησης αποτελούν ζωτικό μέρος της ενεργειακής πολιτικής των χωρών, ειδικά για τις αναπτυσσόμενες χώρες των οποίων η ενεργειακή ζήτηση αυξάνεται σχετικά γρήγορα, όπως στην περίπτωση της Ελλάδας. Επιπλέον, η ελληνική οικονομία είναι εξαιρετικά ευαίσθητη στην εσωτερική και εξωτερική πολιτική της, αλλά και την ανάπτυξη της αγοράς. Λόγω αυτών των αρνητικών ή θετικών επιπτώσεων, το μοτίβο κατανάλωσης ενέργειας παρουσιάζει σημαντικές αυξομειώσεις ακόμη και σε σύντομο χρονικό διάστημα. Παράδειγμα αποτελεί η οικονομική κρίση του 2008 η οποία απέφερε και τις ανάλογες μειώσεις στη ζήτηση ενέργειας, που επισημαίνεται από τη Λιόση (2018).

Το φορτίο ενός συστήματος αποτελεί μια στοχαστική, μη στάσιμη διαδικασία η οποία απαρτίζεται από ένα σύνολο επιμέρους στοιχείων. Επομένως, η συμπεριφορά του φορτίου ενός συστήματος επηρεάζεται από ένα πλήθος παραγόντων που μπορούν να ταξινομηθούν με βάση τη βιβλιογραφία στις ακόλουθες βασικές κατηγορίες: καιρικές συνθήκες, κοινωνικοοικονομικοί παράγοντες, ημερολογιακοί παράγοντες, είδος τελικού χρήστη και τυχαία γεγονότα. Στο πλαίσιο της παρούσας διπλωματικής οι παράμετροι που συγκεντρώθηκαν για τη μοντελοποίηση της ενεργειακής ζήτησης είναι: η θερμοκρασία των προηγούμενων ημερών (έως και μία εβδομάδα πριν), η εποχή, η ημέρα, ο μήνας και το είδος της ημέρας (αργία ή μη).

Οι παράγοντες αυτοί καθιστούν τη διαδικασία πρόβλεψης του ηλεκτρικού φορτίου μια πολύπλοκη διαδικασία. Για την επιλογή της σωστής μεθόδου είναι σημαντική η κατηγοριοποίηση των προβλέψεων με βάση το χρονικό τους ορίζοντα. Έτσι διακρίνονται οι περιπτώσεις της πολύ βραχυπρόθεσμης (very-short), βραχυπρόθεσμης (short), μεσοπρόθεσμης (medium) και μακροπρόθεσμης (long) πρόβλεψης. Η πολύ βραχυπρόθεσμη κατηγορία αφορά προβλέψεις σε χρονικό ορίζοντα από λίγα λεπτά μέχρι μία ώρα. Οι βραχυπρόθεσμες προβλέψεις συνήθως κυμαίνονται από μία ώρα έως μία εβδομάδα, οι μεσοπρόθεσμες προβλέψεις συνήθως κυμαίνονται από μία εβδομάδα έως ένα έτος και οι μακροπρόθεσμες προβλέψεις συνήθως εφαρμόζονται σε χρονικά διαστήματα μεγαλύτερα από ένα έτος, αν και η απουσία ενός προτύπου ενδέχεται να δημιουργεί διαφορές στα χρονικά διαστήματα, κατά τους Hong and Fan (2016). Οι Filik et al. (2011) προσέγγισαν τη μοντελοποίηση της ωριαίας ζήτησης ενέργειας με σκοπό τη μακροπρόθεσμη πρόβλεψη.

Η βραχυπρόθεσμη διακύμανση της ζήτησης φαίνεται να επηρεάζεται από αυξομειώσεις της θερμοκρασίας και να εξαρτάται σε μεγάλο βαθμό από κλιματικούς παράγοντες, γεγονός που υποστηρίζουν οι Zachariadis et al. (2007). Από την άλλη πλευρά, όσον αφορά τη μακροπρόθεσμη πρόβλεψη, υπάρχει ισχυρή συσχέτιση μεταξύ της ζήτησης και των κοινωνικοοικονομικών μεταβλητών, με σημαντικό αριθμό μελετών που εξετάζουν τις πιθανές αιτιώδεις σχέσεις τους, όπως ο Payne (2010) και οι Polemis et al. (2013). Συγκεκριμένα, φαίνεται να επηρεάζεται σε μεγάλο βαθμό από τα οικονομικά και δημογραφικά δεδομένα, αλλά και τους διαφορετικούς τύπους καταναλωτών (οικιακοί, εμπορικοί και βιομηχανικοί). Ως εκ τούτου, η ανάλυση της ζήτησης σε διάφορες χρονικές κλίμακες και η σχέση της με τις καιρικές, κλιματικές και κοινωνικοοικονομικές μεταβλητές είναι απαραίτητη για αυτήν την πρόβλεψη. Σύμφωνα με τους Farzan et al. (2015), η μοντελοποίηση και η πρόβλεψη της ζήτησης είναι σημαντικές για τη διαμόρφωση μίας βιώσιμης ενεργειακής πολιτικής.

Η πρόβλεψη της ζήτησης μπορεί να εξυπηρετήσει βραχυπρόθεσμους, μεσοπρόθεσμους και μακροπρόθεσμους στόχους. Βραχυπρόθεσμα, η πρόβλεψη είναι απαραίτητη για την καθημερινή λειτουργία (ενεργειακές συναλλαγές, ανάλυση ασφάλειας και ενεργειακά αποθέματα) συστημάτων παραγωγής και διανομής, κάτι το οποίο υποστηρίζουν ο Tukey (1977) και οι Filik et al. (2011). Μεσοπρόθεσμα, η πρόβλεψη χρησιμοποιείται από παραγωγούς και μεταπωλητές ηλεκτρικής ενέργειας για προγραμματισμό συντήρησης δικτύων και έρευνα αγοράς. Μακροπρόθεσμα, είναι πολύ κρίσιμη για το σχεδιασμό και την επέκταση των συστημάτων ηλεκτρικής ενέργειας. Διαδραματίζει πολύ ουσιαστικό και σημαντικό ρόλο στο σχεδιασμό της κατασκευής νέων εγκαταστάσεων παραγωγής, την ανάπτυξη του ενεργειακού χάρτη μιας χώρας και την επέκταση των γραμμών μεταφοράς. Στην παρούσα διπλωματική βέβαια, το ενδιαφέρον επικεντρώνεται στη βραχυπρόθεσμη πρόβλεψη της ενεργειακής ζήτησης.

Η σωστή διαχείριση και πρόβλεψη σε βραχυπρόθεσμη κλίμακα μπορεί να βοηθήσει στην εκτίμηση των ροών του φόρτου και στη λήψη αποφάσεων που θα προλάβουν πιθανή υπερφόρτωση. Έγκαιρη υλοποίηση αυτών των αποφάσεων θα οδηγήσει σε αύξηση της αξιοπιστίας του συστήματος και παράλληλα σε μείωση της συχνότητας εμφάνισης αποτυχίας του εξοπλισμού, αλλά και των διακοπών ρεύματος. Μια άλλη πτυχή της σημαντικότητας της πρόβλεψης του φόρτου είναι για αξιολογήσεις συμβάσεων και διάφορων εξελιγμένων χρηματοπιστωτικών προϊόντων σχετικά με την τιμολόγηση της ενέργειας που προσφέρει η αγορά. Ανάλογες μελέτες πραγματοποίησαν οι Pérez-Rodríguez et al. (2005), οι Khashei et al. (2011), ο Yan (2012), οι Babu et al. (2014) και οι Lin et al. (2017). Επίσης, η πρόβλεψη της κατανάλωσης ενέργειας είναι απαραίτητη για περιβαλλοντικά ζητήματα, όπως η υπερθέρμανση του πλανήτη που οφείλονται κυρίως σε εκπομπές αερίων του θερμοκηπίου, αφού τα ορυκτά καύσιμα κατέχουν στις μέρες μας την πιο κοινή πηγή για την παραγωγή ηλεκτρικής ενέργειας. Έτσι, η παραγωγή ηλεκτρισμού εξαντλεί τους φυσικούς πόρους και συμβάλλει στη ρύπανση του περιβάλλοντος. Οι Zou et al. (2004) και οι Papacharalampous et al. (2017) μελέτησαν την πρόβλεψη για τους παραπάνω λόγους.

Η επιστημονική βιβλιογραφία περιλαμβάνει πολλά άρθρα σχετικά με τη ζήτηση ενέργειας στην Ελλάδα. Οι Psiloglou et al. (2009) έχουν διερευνήσει την ενεργειακή ζήτηση της

Αθήνας τη χρονική περίοδο 1997-2001, προσπαθώντας να βρουν τη συσχέτισή της με ένα οικονομικό δείκτη, δηλαδή το Ακαθάριστο Εθνικό Προϊόν (ΑΕΠ) και έναν κλιματικό δείκτη δηλαδή τη θερμοκρασία. Οι Mirasgedis et al. (2006) παρουσιάζουν μια παρόμοια ανάλυση για τη ζήτηση ενέργειας στην Ελλάδα, αλλά και ο Tyralis (2016) για τη χρονική περίοδο 1993-2002 και 2002-2014 αντίστοιχα. Ενδιαφέρον παρουσιάζει επίσης η ανάλυση ζήτησης ενέργειας για ένα νησί εκτός δικτύου, την Αστυπάλαια, με σκοπό την δημιουργία ενός υβριδικού συστήματος ανανεώσιμων πηγών ενέργειας που έγινε από τους Mavroyeoryos et al. (2017). Οι Pappas et al. (2010) υπολογίζουν στατιστικά της χρονοσειράς ζήτησης ενέργειας, για τη χρονική περίοδο 2004-2005, ώστε να προσαρμόσουν ένα στοχαστικό μοντέλο, το οποίο χρησιμοποιούν για πρόβλεψη. Στην παρούσα μελέτη παρουσιάζονται παρόμοια αποτελέσματα, μετά από μια λεπτομερέστερη ανάλυση.

Σήμερα τα διαθέσιμα μοντέλα για την πρόβλεψη χρονοσειρών είναι πολυάριθμα και μπορούν να ταξινομηθούν σύμφωνα με τους De Gooijer και Hyndman (2006) σε οκτώ κατηγορίες: (α) μοντέλα εκθετικής εξομάλυνσης (exponential smoothing models), (β) ARIMA, (γ) εποχιακά μοντέλα, (δ) state space and structural models and the Kalman filter, (ε) μη γραμμικά μοντέλα, (στ) long-range dependence models, π.χ. η οικογένεια μοντέλων Autoregressive Fractionally Integrated Moving Average (ARFIMA), (ζ) Autoregressive Conditional Heteroscedastic / Generalized Autoregressive Conditional Heteroscedastic (ARCH / GARCH) και (η) count data forecasting.

Οι θεωρητικές ιδιότητες των μοντέλων των κατηγοριών (α) - (δ), (στ), (ζ) (εφεξής, θα αναφέρονται ως «στοχαστικά») έχουν διερευνηθεί σε μεγάλο βαθμό, σε αντίθεση με εκείνα των μη γραμμικών μοντέλων και ιδίως των αλγορίθμων Machine Learning (ML). Αυτές οι δύο κύριες κατηγορίες μοντέλων είναι γνωστό ότι αντιπροσωπεύουν δύο διαφορετικές κουλτούρες στη στατιστική μοντελοποίηση και την αλγοριθμική κουλτούρα μοντελοποίησης Breiman (2001). Η πρώτη υποθέτει ότι πίσω από τη δημιουργία των δεδομένων βρίσκεται ένα στοχαστικό μοντέλο αναλυτικά διατυπωμένο, ενώ η δεύτερη ότι πίσω από αυτήν τη διαδικασία βρίσκεται κάτι πολύπλοκο και άγνωστο, το οποίο όμως δεν χρειάζεται να διατυπωθεί αναλυτικά, αρκεί να υπάρχει ένα καθαρά αλγοριθμικό μοντέλο το οποίο να προσφέρει υψηλή ακρίβεια πρόβλεψης. Με άλλα λόγια, η βαθιά κατανόηση και σωστή μοντελοποίηση της (μελλοντικής) συμπεριφοράς μιας διαδικασίας συνδέεται στενά με την κουλτούρα μοντελοποίησης δεδομένων. Η διάκριση μεταξύ αιτιώδους εξήγησης, πρόβλεψης και περιγραφής αναγνωρίζεται και αποσαφηνίζεται σε όρους μοντελοποίησης στο Shmueli (2010).

Όσον αφορά τη βραχυπρόθεσμη πρόβλεψη φορτίου (STLF – short term load forecasting), αρκετές μέθοδοι συμπεριλαμβανομένων των παραδοσιακών μοντέλων και των μεθόδων τεχνητής νοημοσύνης έχουν προταθεί τις τελευταίες δεκαετίες. Τα στατιστικά μοντέλα χρονοσειρών μπορούν να οδηγήσουν σε αριθμητική αστάθεια και έλλειψη ακρίβειας, εάν αγνοηθούν όπως αναφέραμε παραπάνω, παράγοντες υψηλής συσχέτισης όπως οι καιρικές μεταβλητές. Τα παραδοσιακά μοντέλα συνήθως αντικατοπτρίζουν σταθερές γραμμικές σχέσεις μεταξύ των φορτίων και των καιρικών μεταβλητών. Ωστόσο, η πολύπλοκη και μη γραμμική σχέση μεταξύ του φορτίου και των καιρικών συνθηκών μπορεί να τα καταστήσει μη πρακτική επιλογή. Τα τελευταία χρόνια έχει δοθεί μεγάλη προσοχή σε μεθόδους μηχανικής μάθησης, δηλαδή μοντέλα που βασίζονται στα διαθέσιμα δεδομένα και

εκπαιδεύονται μέσα από αυτά, όπως για παράδειγμα τα τεχνητά νευρωνικά δίκτυα (ANN – artificial neural networks), οι μηχανές διανυσμάτων υποστήριξης (SVM – support vector machines) και τα δέντρα αποφάσεων (Decision Trees). Αντίθετα με την ανάλυση της ζήτησης βασισμένης στα αναλυτικά μοντέλα, οι αλγόριθμοι εξόρυξης δεδομένων προσφέρουν ισχυρά εργαλεία για την ανάδειξη των προτύπων από μεγάλα σύνολα δεδομένων. Ένα ακόμα πλεονέκτημα των μοντέλων που στηρίζονται στα δεδομένα είναι ότι μπορούν εύκολα να ανανεώσουν τη βάση τους. Αυτό το στοιχείο είναι σημαντικό, καθώς η σχέση του φορτίου με τις καιρικές συνθήκες δεν είναι στατική.

Ένα από τα βασικότερα προβλήματα που δημιουργούνται κατά την προσπάθεια πρόβλεψης είναι ο «καθαρισμός» δεδομένων. Η ποιότητα των δεδομένων συχνά δημιουργεί προβλήματα λόγω διακοπών λειτουργίας, μεταφορών φορτίου (load transfer) και άλλων σφαλμάτων των δεδομένων. Επίσης, σημαντικό πρόβλημα αποτελεί η ύπαρξη ζωνών που έχουν διαφορετική συμπεριφορά ενεργειακής κατανάλωσης. Λόγου χάρη μία ζώνη που αντιπροσωπεύει μια βιομηχανική περιοχή δεν είναι ευαίσθητη σε καιρικές συνθήκες. Το ευρύτερο αντικείμενο της διπλωματικής εργασίας είναι η πρόβλεψη της ενεργειακής ζήτησης για βραχυπρόθεσμο χρονικό ορίζοντα. Σε πρώτο στάδιο γίνεται η συλλογή δεδομένων από τους αντίστοιχους φορείς. Τα δεδομένα ηλεκτρικής ενέργειας είναι σε ωριαία κλίμακα, και προέρχονται από τον Ανεξάρτητο Διαχειριστή Μεταφοράς Ηλεκτρικής Ενέργειας (ΑΔΜΗΕ, Independent Power Transmission Operator, IPTO), ενώ η θερμοκρασία είναι σε ημερήσια κλίμακα και οι τιμές τις λαμβάνονται από σταθμό της Ηλιούπολης. Τέλος, το Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ) της Ελλάδας λαμβάνεται από τα δεδομένα που παρέχει η Παγκόσμια Τράπεζα. Πριν από τη μοντελοποίηση της ζήτησης, είναι σημαντικό να κατανοήσουμε τη συμπεριφορά της. Για να κατανοήσουμε τα δεδομένα των χρονοσειρών, μια συχνή προσέγγιση που υιοθετήθηκε από τους επιστήμονες δεδομένων είναι η διερευνητική ανάλυση δεδομένων που δίνει έμφαση στη γραφική αναπαράσταση των δεδομένων. Με βάση το κλασικό βιβλίο του Tukey (1977), οι Behrens and Yu (2003) αναφέρουν ότι στο πλαίσιο της ανάλυσης, οι αναλυτές δεδομένων και οι επιστήμονες εργάζονται διαδραστικά σε μια κυκλική διαδικασία εξαγωγής προτύπων και ερμηνείας προτύπων. Ο στόχος είναι να συμπληρωθεί η δημιουργία μοντέλου με βάση τα ευρήματα της ανάλυσης, κάτι το οποίο αναφέρει και ο Behrens (1997). Η ανάλυση αυτή στοχεύει επίσης να εντοπίσει παραπλανητικά μοτίβα. Επομένως, αφού γίνει ο «καθαρισμός» των δεδομένων, θα πραγματοποιηθεί οπτικοποίησή τους σε διάφορες κλίμακες για την αποκρυπτογράφηση μοτίβων επανάληψης και εποχικότητας. Έπειτα, θα δημιουργηθούν τα ορισμένα από το χρόνο πλαίσια δεδομένων (data frames) στο R για κάθε μία από τις μεταβλητές και θα γίνει συγχώνευση αυτών μεταξύ τους με κοινά χρονικά πλαίσια. Με τη βοήθεια ορισμένων πακέτων του R, θα ορίσουμε στη συνέχεια ένα ευρύ φάσμα μοντέλων για πρόβλεψη τα οποία θα εξετάσουμε σε βάθος. Τέλος, από τα αποτελέσματα που θα προκύψουν θα αξιολογήσουμε την αξιοπιστία των μοντέλων και θα σχολιάσουμε τα ευρήματα της εργασίας.

1.2 Δομή διπλωματικής εργασίας

Αρχικά παρατίθεται η περίληψη της διπλωματικής εργασίας στην ελληνική και αγγλική γλώσσα, όπου παρουσιάζονται συνοπτικά τα κύρια σημεία της. Στη συνέχεια ακολουθεί ο πρόλογος και ο αναλυτικός πίνακας περιεχομένων. Τέλος ακολουθεί το κύριο μέρος της εργασίας, το οποίο αποτελείται από 5 κεφάλαια.

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Πρόκειται για το παρόν κεφάλαιο, στο οποίο παρουσιάζεται συνοπτικά το αντικείμενο και ο σκοπός της εργασίας καθώς και το παρόν υποκεφάλαιο, στο οποίο παρουσιάζεται η δομή που υιοθετήθηκε.

ΚΕΦΑΛΑΙΟ 2: ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

Στο κεφάλαιο αυτό εισάγεται αρχικά η έννοια της εξόρυξης γνώσης από ένα σύνολο δεδομένων. Περιγράφονται τα στάδια υλοποίησης που απαιτούνται για την εξαγωγή συσχετίσεων και προτύπων από μια βάση δεδομένων. Επίσης, παρουσιάζονται οι βασικοί αλγόριθμοι εξόρυξης δεδομένων, οι οποίοι μπορούν να ομαδοποιηθούν σε δύο ευρύτερες κατηγορίες, την προβλεπτική και την περιγραφική. Στη συνέχεια, γίνεται εξαγωγή των δεδομένων από τους αντίστοιχους φορείς και η εισαγωγή αυτών στο R. Σε αυτό το σημείο, αφού γίνει ο «καθαρισμός» δεδομένων, θα εξηγηθεί η επιλογή των πακέτων που επιλέχθηκαν από το R τόσο για την ανάλυση όσο και για την πρόβλεψη των δεδομένων.

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Το κεφάλαιο αυτό ξεκινά με την επεξεργασία δεδομένων. Όπως αναφέρθηκε προηγουμένως, για την κατανόηση των ακολουθιών που κρύβουν τα δεδομένα, θα γίνει οπτικοποίηση των δεδομένων σε διάφορες χρονικές κλίμακες. Για το σκοπό αυτό, επιλέγουμε δεδομένα από 2006-09-01 έως 2016-08-31 (10 υδρολογικά έτη) για το φορτίο ζήτησης και τη θερμοκρασία. Επομένως, εκτελούμε ορισμένες λειτουργίες για να λάβουμε σε ημερήσια, εβδομαδιαία και μηνιαία κλίμακα τα νέα δεδομένα. Είναι απαραίτητο να βρεθεί η συσχέτιση της ζήτησης με τη θερμοκρασία, την εποχή, την ημέρα, το μήνα, το είδος της ημέρας (αργία ή μη) και την κατανάλωση των προηγούμενων ημερών. Αυτές θα είναι οι μεταβλητές που θα επηρεάζουν τα μοντέλα που θα χρησιμοποιηθούν. Τέλος, θα γίνει συγχώνευση των χρονοσειρών για κάθε μεταβλητή (φορτίο ζήτησης, θερμοκρασία και ΑΕΠ) με κοινό χρονικό πλαίσιο σε ημερήσια κλίμακα, η οποία θα χρησιμοποιηθεί για πρόβλεψη.

ΚΕΦΑΛΑΙΟ 4: ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ

Το κεφάλαιο αυτό ασχολείται με τη βραχυπρόθεσμη πρόβλεψη που αποτελεί τον άξονα της διπλωματικής. Γίνεται αρχικά μια βιβλιογραφική ανασκόπηση των στατιστικών μεθόδων, που στηρίζονται στη μαθηματική αναπαράσταση της καμπύλης του φορτίου. Στη συνέχεια, θα εξηγηθεί η μεθοδολογία και ο τρόπος με τον οποίο επιλέγεται κάθε μοντέλο. Ως δεδομένα εκπαίδευσης (training data) επιλέγουμε τα πρώτα 5 χρόνια (από 2007-01-01 έως 2011-12-31) και ως δεδομένα δοκιμών (testing data) επιλέγουμε τα ακόλουθα 5 χρόνια

(από 2012-01-01 έως 2016-12-31) σε ημερήσια κλίμακα. Σε αυτή τη φάση, θα ορίσουμε ένα μοντέλο αναφοράς (naïve), μοντέλα γραμμικής παλινδρόμησης και μοντέλα random forest. Το κάθε μοντέλο θα εξαρτάται από τις μεταβλητές που αναφέραμε προηγουμένως και θα εξετάζεται ως προς την καταλληλότητά του με τις στατιστικές παραμέτρους του, αλλά και τη γραφική απεικόνισή του σε σχέση με τα πραγματικά δεδομένα.

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Το κεφάλαιο αυτό αποτελεί τον επίλογο της διπλωματικής εργασίας όπου παρουσιάζονται αναλυτικά τα αποτελέσματα των προβλέψεων με σκοπό την ανάδειξη της βέλτιστης μεθόδου που προσέφερε την υψηλότερη ακρίβεια πρόβλεψης. Επιπλέον, αναδεικνύονται τα κυριότερα συμπεράσματα που προέκυψαν από τα προηγούμενα κεφάλαια της παρούσας μελέτης και προτείνονται σημεία για περαιτέρω ανάλυση.

ΚΕΦΑΛΑΙΟ 2

ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 2: ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

2.1 Εξόρυξη Δεδομένων

2.1.1 Εισαγωγή στην Εξόρυξη Δεδομένων

Η εξόρυξη δεδομένων (data mining), που αναφέρεται και ως εξόρυξη γνώσης (knowledge discovery) σε βάσεις δεδομένων, είναι ένα διεπιστημονικό πεδίο της επιστήμης των υπολογιστών και των στατιστικών με συνολικό στόχο την εξαγωγή πληροφοριών (με έξυπνες μεθόδους) από ένα σύνολο δεδομένων και τη μετατροπή των πληροφοριών σε κατανοητή δομή για περαιτέρω χρήση. Με την εξόρυξη δεδομένων έχουν ασχοληθεί εκτενώς οι Piatetsky-Shapiro and Frawley (1991), ο Χαραλάμπους (2009) και ο Λαδάς (2015). Οι βασικές τεχνικές data mining περιλαμβάνουν αλγόριθμους για κατηγοριοποίηση, παλινδρόμηση, πρόβλεψη συμβάντων, συσταδοποίηση, ανακάλυψη ακολουθιακών προτύπων και εύρεση συσχετίσεων.

Η εξόρυξη δεδομένων αποτελείται από μια σειρά βημάτων υλοποίησης. Τα πιο σημαντικά βήματα παρουσιάζονται συνοπτικά παρακάτω:

- Βήμα 1: Επιλογή Δεδομένων (Data Selection)

Το πρώτο βήμα της διαδικασίας είναι η συλλογή και η αποθήκευση των δεδομένων. Σε αυτό το βήμα γίνεται η ανάκτηση των δεδομένων που σχετίζονται με το αντικείμενο μελέτης.

- Βήμα 2: Καθαρισμός δεδομένων και Προεπεξεργασία δεδομένων (Data Cleaning and Preprocessing)

Ο καθαρισμός δεδομένων είναι μια εργασία που εντοπίζει εσφαλμένα, ελλιπή, ανακριβή ή άσχετα δεδομένα, διορθώνει τυχόν προβλήματα και διασφαλίζει ότι όλα αυτά τα ζητήματα θα διορθωθούν αυτόματα στο μέλλον. Ακολουθούν ορισμένα από τα πιο συνηθισμένα βήματα και μέθοδοι καθαρισμού δεδομένων:

- Αντιμετώπιση δεδομένων που λείπουν
- Τυποποίηση της διαδικασίας
- Επικύρωση ακρίβειας των δεδομένων
- Κατάργηση διπλών δεδομένων
- Χειρισμός δομικών σφαλμάτων
- Απαλλαγή από ανεπιθύμητες παρατηρήσεις

Αντιμετώπιση δεδομένων που λείπουν – το να αγνοήσουμε τις τιμές που λείπουν σε ένα σύνολο δεδομένων είναι ένα τεράστιο λάθος, καθώς οι περισσότεροι αλγόριθμοι απλώς δεν τις αποδέχονται. Μία κοινή πρακτική για την αντιμετώπιση αυτού του προβλήματος είναι να ορίσουμε τις τιμές που λείπουν βάσει άλλων παρατηρήσεων. Έτσι, στη συγκεκριμένη περίπτωση βρίσκουμε τις μέρες που η ζήτηση ενέργειας είναι είτε 0 είτε NA (Not Available) και τις αντιμετωπίζουμε με παρόμοιο τρόπο, ο οποίος θα αναλυθεί

λεπτομερώς στη συνέχεια.

Δομικά σφάλματα - πρόκειται για λάθη που προκύπτουν κατά τη μέτρηση, τη μεταφορά δεδομένων και άλλα ζητήματα που προκύπτουν λόγω κακής διαχείρισης δεδομένων. Τα ασυνεπή σημεία στίξης, τυπογραφικά λάθη και εσφαλμένες ετικέτες είναι τα πιο συνηθισμένα προβλήματα εδώ. Τέτοια σφάλματα δείχνουν τη σημασία καθαρισμού δεδομένων.

Ανεπιθύμητες παρατηρήσεις - συχνά συναντώνται ανεπιθύμητες παρατηρήσεις στα σύνολα δεδομένων. Αυτές μπορεί να είναι διπλές παρατηρήσεις ή αυτές που δεν σχετίζονται με το συγκεκριμένο πρόβλημα που προσπαθούν να λύσουν. Γι' αυτό το λόγο γίνεται έλεγχος των παρατηρήσεων για τον εξορθολογισμό της διαδικασίας των τεχνικών χαρακτηριστικών.

- Βήμα 3: Μετασχηματισμός Δεδομένων (Transformation)

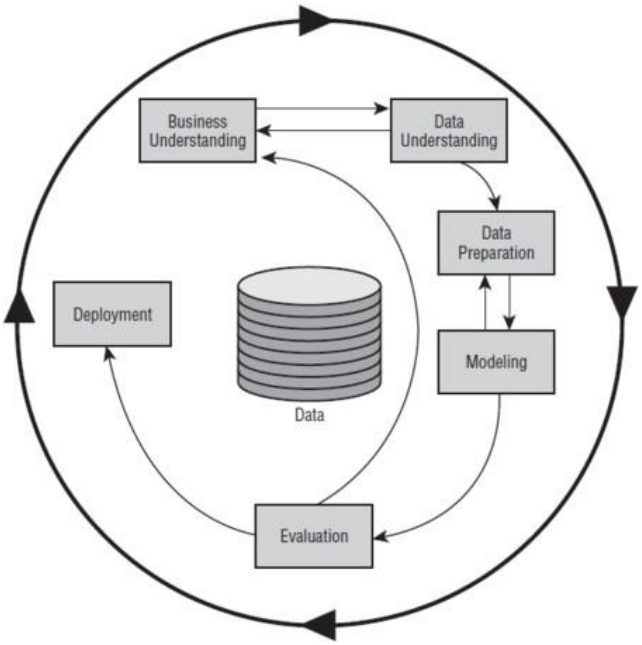
Ουσιαστικά, πρόκειται για τη μετατροπή των δεδομένων κάτω από ένα κοινό πλαίσιο, για επεξεργασία. Χρησιμοποιείται κυρίως για την εξομάλυνση των δεδομένων και απομάκρυνση θορύβου, για τη συνάθροιση των δεδομένων, δηλαδή για την παραγωγή σύνοψής τους, για την κανονικοποίησή τους, δηλαδή την κλιμάκωση των χαρακτηριστικών του συνόλου δεδομένων σε ένα συγκεκριμένο και περιορισμένο εύρος τιμών, ή τέλος για τη δημιουργία νέων χαρακτηριστικών από τα ήδη υπάρχοντα. Η βασική ιδέα είναι η μείωση των πολυάριθμων ποσοτήτων δεδομένων στα ουσιαστικά μέρη.

- Βήμα 4: Εξόρυξη Δεδομένων (Data Mining)

Στο βήμα αυτό τα δεδομένα υφίστανται επεξεργασία με την εφαρμογή τεχνικών εξόρυξης δεδομένων. Στόχος είναι να αναδειχθούν χρήσιμα συμπεράσματα όπως συσχετίσεις, μοντέλα συμπεριφοράς, κανόνες γνώσης κ.α.

- Βήμα 5: Διερμηνεία και Αξιολόγηση (Interpretation & Evaluation)

Τέλος, γίνεται η διερμηνεία και η αξιολόγηση των αποτελεσμάτων που παρήχθησαν από την όλη διαδικασία, με τη βοήθεια ενός πίνακα ή μιας γραφικής παράστασης. Την αξιολόγηση αποτελεσμάτων και πρόβλεψη ενεργειακής ζήτησης έχουν μελετήσει οι Devi and Manonmani (2012), οι Tsekouras et al. (2003) και ο Figueiredo (2005). Στο Σχήμα 2.1, παρουσιάζεται η διαδικασία και τα βήματα υλοποίησης για την εξόρυξη δεδομένων, τα οποία αναφέρει και ο Shearer (2000).

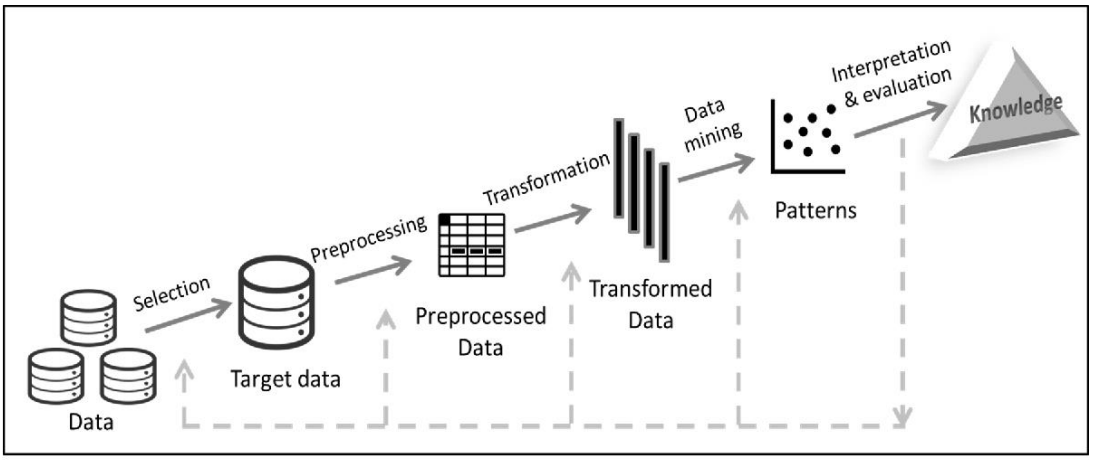


Σχήμα 2.1: Διαδικασία εξόρυξης δεδομένων

2.1.2 Σημασία της Εξόρυξης Δεδομένων

Η εξόρυξη δεδομένων είναι η εξαγωγή «κρυμμένων» πληροφοριών μέσα από μεγάλες βάσεις δεδομένων. Με την ανάκτηση αυτής της γνώσης είναι δυνατό να εμφανιστούν σημαντικές περιοδικότητες και υψηλού επιπέδου πληροφορίες από τα εξεταζόμενα σύνολα δεδομένων. Η εξόρυξη δεδομένων, ωστόσο, είναι μια κρίσιμη διαδικασία και απαιτεί πολύ χρόνο και προσοχή στη συλλογή των επιθυμητών δεδομένων λόγω της πολυπλοκότητας των βάσεων δεδομένων. Οι τεχνικές εξόρυξης δεδομένων θα βοηθήσουν στην επιλογή ενός τμήματος δεδομένων χρησιμοποιώντας τα κατάλληλα εργαλεία για το φιλτράρισμα των ακραίων τιμών και των ανωμαλιών στο σύνολο των δεδομένων. Η εξόρυξη γνώσης αποτελεί μια ισχυρή νέα τεχνολογία με πολλαπλές χρήσεις στον πραγματικό κόσμο σε οικονομικές, ιατρικές, επαγγελματικές εφαρμογές, εφαρμογές πρόβλεψης κ.α. Αξίζει να σημειώσουμε ότι συνδυάζει διάφορα επιστημονικά πεδία, όπως η στατιστική, η τεχνολογία βάσεων δεδομένων, η τεχνητή νοημοσύνη, τα νευρωνικά δίκτυα και η μηχανική μάθηση (Machine Learning).

Η σημασία της εξόρυξης δεδομένων προσδιορίζεται από την εφαρμογή που βρίσκουν τα αποτελέσματά της. Πιο συγκεκριμένα και στη δική μας μελέτη, η γνώση που αποκτούμε σχετικά με τα δεδομένα μας δίνει τη δυνατότητα δημιουργίας ενός προφίλ κατανάλωσης. Σε αυτήν την περίπτωση είναι δυνατή η εξαγωγή πληροφοριών σχετικά με τη συμπεριφορά ενός καταναλωτή και τις συσχετίσεις της κατανάλωσης με τις κλιματολογικές συνθήκες. Όλα τα παραπάνω οδηγούν στη βέλτιστη μοντελοποίηση του προφίλ κατανάλωσης και την ακριβέστερη πρόβλεψη μελλοντικών καταναλώσεων. Αξίζει να αναφέρουμε ότι τα τελευταία χρόνια έχουν κάνει την εμφάνισή τους εφαρμογές που προσφέρουν ενεργειακές υπηρεσίες μέσω διαδικτύου και στηρίζονται στην επεξεργασία ενεργειακών δεδομένων πραγματικού χρόνου (real-time data), όπως η εφαρμογή του Nikolopoulos (2011). Στο Σχήμα 2.2, παρουσιάζονται τα βασικότερα βήματα της διαδικασίας της εξόρυξης γνώσης, όπως παρουσιάζονται από τους Fayyad et al. (1996).



Σχήμα 2.2: Τα βασικότερα βήματα της διαδικασίας Εξόρυξης Γνώσης

2.1.3 Μοντέλα και Μέθοδοι Εξόρυξης Δεδομένων

Σύμφωνα με τους Fayyad et al. (1996) και Παναγιωτάκος (2012), τα συστατικά στοιχεία που απαρτίζουν τους αλγόριθμους εξόρυξη δεδομένων είναι τα ακόλουθα:

1. Μέθοδος Αναζήτησης

Η αναζήτηση παραμέτρων ψάχνει για ελεύθερες παραμέτρους που βελτιστοποιούν την απόδοση ενός τελικού μοντέλου. Αντίστοιχα η αναζήτηση μοντέλου ψάχνει για το κατάλληλο μοντέλο ή την οικογένεια μοντέλων και για κάθε μια τέτοια δομή που βρίσκει, εφαρμόζει τις κατάλληλες παραμέτρους.

2. Παράσταση Μοντέλου

Είναι η γλώσσα που περιγράφει τις μεθόδους που έχουν ανακαλυφθεί. Για να παραχθεί ένα ικανοποιητικό μοντέλο είναι απαραίτητη η εκτενής αναπαράσταση σε συνδυασμό με παραδείγματα και επαρκή χρόνο εκπαίδευσης.

3. Αξιολόγηση Μοντέλου

Εδώ γίνεται η αξιολόγηση του μοντέλου με τις επιλεγμένες παραμέτρους. Είναι αναγκαίο η μέθοδος να πληροί τις προδιαγραφές της διαδικασίας KDD (Knowledge Discovery in Data).

Οι τεχνικές Εξόρυξης Δεδομένων (Data Mining) που αναφέραμε στην παράγραφο 2.1.1 μπορούν να ομαδοποιηθούν σε δύο ευρύτερα μοντέλα. Έτσι έχουμε:

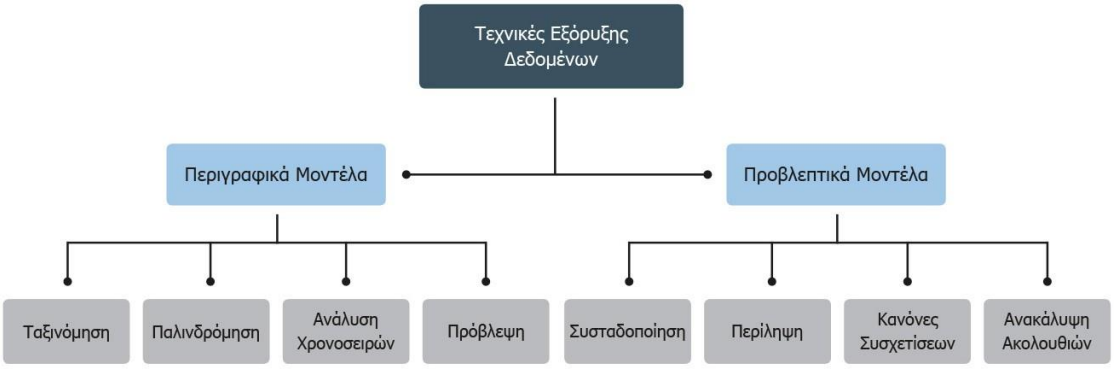
Το προβλεπτικό μοντέλο (predictive model) που έχει στόχο την πρόβλεψη των μελλοντικών δεδομένων, στηριζόμενο σε γνωστά αποτελέσματα που έχει βρει από άλλα δεδομένα. Σε αυτήν την περίπτωση, η χρήση ιστορικών δεδομένων κρίνεται απαραίτητη κατά τη μοντελοποίηση της πρόβλεψης. Το προβλεπτικό μοντέλο απαρτίζεται από τις ακόλουθες μεθόδους data mining:

- Ταξινόμηση (Classification)
- Παλινδρόμηση (Regression)
- Ανάλυση Χρονοσειρών (Time Series Analysis)
- Πρόβλεψη (Prediction)

Το περιγραφικό μοντέλο (descriptive model) που έχει στόχο την αναγνώριση προτύπων και συσχετίσεων μεταξύ των δεδομένων. Με αυτόν τον τρόπο επιτυγχάνεται η διερεύνηση των ιδιοτήτων που φέρουν τα υπό εξέταση δεδομένα. Το περιγραφικό μοντέλο απαρτίζεται από τις ακόλουθες μεθόδους data mining:

- Συσταδοποίηση (Clustering)
- Περίληψη (Summarization)
- Κανόνες Συσχετίσεων (Association Rules)
- Ανακάλυψη Ακολουθιών (Sequential Pattern Discovery)

Στο Σχήμα 2.3 παρουσιάζεται ένα διάγραμμα με τα δύο ευρύτερα μοντέλα και τις μεθόδους που τα απαρτίζουν, όπως τα ταξινομεί ο Dunham (2004):



Σχήμα 2.3: Ταξινόμηση των Τεχνικών Εξόρυξης Δεδομένων

2.1.3.1 Ταξινόμηση

Η ταξινόμηση (classification) είναι μια τεχνική data mining που κατατάσσει τα δεδομένα σε προκαθορισμένες κλάσεις. Είναι γνωστή και ως εποπτευόμενη μάθηση (Supervised Learning) από τους Ripley (1994) και Haykin (1999), καθώς σε αυτήν την περίπτωση οι κλάσεις είναι καθορισμένες πριν την ανάλυση των δεδομένων. Γνωστές τεχνικές ταξινόμησης είναι:

- Naive Bayes μοντέλα
- Δένδρα αποφάσεων (Decision Trees)
- Κοντινότεροι γείτονες (K-Nearest Neighbor)
- Νευρωνικά Δίκτυα (Neural Networks)
- Μηχανές Διανυσμάτων Υποστήριξης (SVM)

2.1.3.2 Παλινδρόμηση

Η παλινδρόμηση (regression) εμφανίζει ομοιότητες με την ταξινόμηση με τη διαφορά ότι η μεταβλητή είναι συνεχής αριθμός. Στη στατιστική μοντελοποίηση, η ανάλυση παλινδρόμησης είναι ένα σύνολο στατιστικών διεργασιών για την εκτίμηση των σχέσεων μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Στόχος της είναι η εκμάθηση μιας συνάρτησης $y = f(x)$ που να απεικονίζει ένα δεδομένο x σε μια μεταβλητή πρόβλεψης y. Μοντέλα παλινδρόμησης βρίσκουμε σε αρκετές μελέτες, όπως αυτές των Bruhns et al. (2005), Liu et al (1996) και Makridakis and Hibon (2000). Προϋπόθεση για τη μοντελοποίηση των δεδομένων είναι να ταιριάζουν σε κάποιο γνωστό είδος συνάρτησης (γραμμική, πολυωνυμική κ.α.). Γνωστές μέθοδοι παλινδρόμησης είναι:

- Γραμμική παλινδρόμηση (Linear)
- Λογιστική παλινδρόμηση (Logistic)
- Δένδρα παλινδρόμησης (Regression Trees)
- Νευρωνικά Δίκτυα (Neural Networks)

2.1.3.3 Ανάλυση Χρονοσειρών

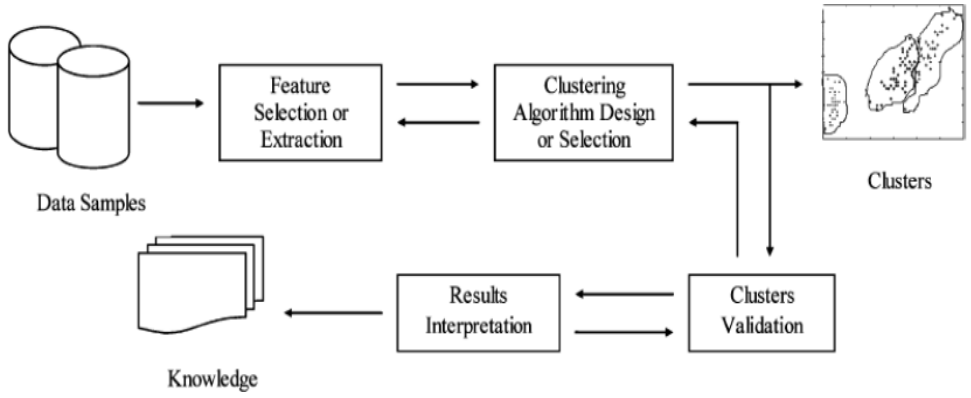
Στόχος της ανάλυσης χρονοσειρών (time series analysis) είναι η μελέτη της μεταβολής μίας μεταβλητής με το πέρασμα του χρόνου. Την παραπάνω ανάλυση έχουν πραγματοποιήσει με διαφορετικές μεταβλητές οι Dunham (2004), οι Han et al. (2006) και οι Belkin and Ord (2000). Οι βασικές λειτουργίες αυτής της τεχνικής data mining είναι: η εξέταση της δομής μιας χρονοσειράς, η εύρεση ομοιοτήτων μεταξύ χρονοσειρών και τέλος η χρήση διαγραμμάτων χρονοσειρών με στόχο την πρόβλεψη μελλοντικών τιμών.

2.1.3.4 Πρόβλεψη

Η πρόβλεψη είναι μια τεχνική που χρησιμοποιεί ιστορικά δεδομένα ως δεδομένα εισόδου για να κάνει τεκμηριωμένες εκτιμήσεις που είναι προγνωστικές για τον καθορισμό της κατεύθυνσης των μελλοντικών τάσεων. Οποιεσδήποτε από τις τεχνικές που χρησιμοποιούνται για την ταξινόμηση μπορούν να προσαρμοστούν και στην πρόβλεψη.

2.1.3.5 Συσταδοποίηση

Η συσταδοποίηση (clustering) είναι μια περιγραφική μέθοδος. Έχοντας ένα σύνολο δεδομένων, στόχος της συσταδοποίησης είναι η δημιουργία συστάδων (clusters), δηλαδή ομάδων, οι οποίες θα περιέχουν όμοια ή παρεμφερή δείγματα. Αναφέρονται ενδεικτικά μελέτες που έχουν ασχοληθεί με τη συσταδοποίηση: Dunham (2004), Han et al. (2006), Zaiane (2005) και Xu and Wunsch (2005). Ουσιαστικά αναζητείται ένα πεπερασμένο σύνολο κατηγοριών ή συστάδων, για να περιγράψει τα δεδομένα. Οι κατηγορίες μπορεί να είναι αμοιβαία αποκλειόμενες και εξαντλητικές ή να έχουν μία πιο σύνθετη αναπαράσταση, όπως για παράδειγμα ιεραρχικές και επικαλυπτόμενες. Στο Σχήμα 2.4, παρατηρούμε τα στάδια της διαδικασίας, όπως περιγράφονται από τους Kaur et al. (2014).



Σχήμα 2.4: Τα στάδια της διαδικασίας συσταδοποίησης

2.1.3.6 Περίληψη

Η περίληψη (summarization) περιλαμβάνει μεθόδους με στόχο την εύρεση συμπαγούς περιγραφής υποσυνόλου των δεδομένων. Πιο αναλυτικά, χαρακτηρίζει τα δεδομένα παράγοντας αντιπροσωπευτικές πληροφορίες, γεγονός που συμβάλλει στην ανάδειξη και κατανόηση μερικών γνωρισμάτων τους. Έννοιες της στατιστικής που εξυπηρετούν αυτόν το σκοπό είναι:

- Μέσος
- Διακύμανση
- Τυπική Απόκλιση
- Ιστόγραμμα
- Διάγραμμα Διασποράς

2.1.3.7 Κανόνες Συσχετίσεων

Η εξαγωγή κανόνων συσχέτισης (Mining Association Rules) θεωρείται μια από τις σημαντικότερες διεργασίες εξόρυξης δεδομένων. Έχει προσελκύσει ιδιαίτερο ενδιαφέρον, καθώς οι κανόνες συσχέτισης παρέχουν έναν συνοπτικό τρόπο για να εκφραστούν οι ενδεχομένως χρήσιμες πληροφορίες που γίνονται εύκολα κατανοητές από τους τελικούς χρήστες. Οι κανόνες συσχέτισης ανακαλύπτουν κρυμμένες «συσχετίσεις» μεταξύ των γνωρισμάτων ενός συνόλου των δεδομένων. Αξίζει να επισημάνουμε ότι η κατανόηση και ανάλυση του καλαθιού αγοράς (market basket analysis) ήταν αυτή που ανέδειξε σε σημαντικό βαθμό τους κανόνες συσχέτισης.

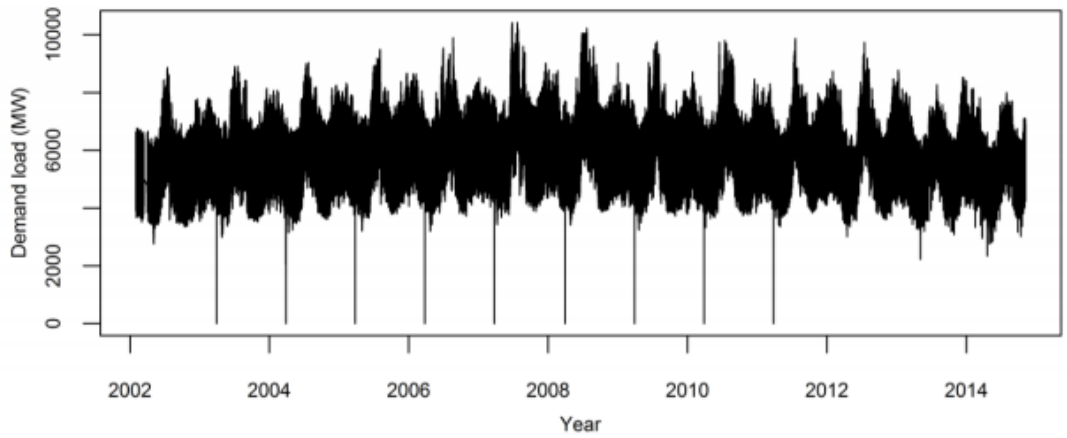
2.1.3.8 Ανακάλυψη Ακολουθιών

Στόχος της ακολουθιακής αυτής ανάλυσης (sequential pattern discovery) είναι ο καθορισμός προτύπων στα δεδομένα. Οι χρονοσειρές και οι ακολουθίες βασίζονται σε χρονικά συνεχόμενα δεδομένα, τα οποία με τη σειρά τους βασίζονται σε εξαρτημένες μεταξύ τους παρατηρήσεις. Το σημαντικό σε αυτήν την περίπτωση είναι ότι η συσχέτιση των προτύπων βασίζεται στο χρόνο.

2.2 Εξαγωγή Δεδομένων

2.2.1 Ζήτηση Ενέργειας

Η ΑΔΜΗΕ (2016) παρέχει πρωτογενή δεδομένα ωριαίας ζήτησης ενέργειας. Τα δεδομένα που λαμβάνονται είναι από 2006-01-01 έως 2016-12-31, δηλαδή 11 χρόνια δεδομένων. Η ποιότητα των δεδομένων είναι καλή με κάποιες τιμές βέβαια να λείπουν, οι οποίες θα αντικατασταθούν στη συνέχεια με χρήση παρεμβολής. Τα πρωτογενή δεδομένα ζήτησης ενέργειας απεικονίζονται στο Σχήμα 2.5. Με μια πρώτη ματιά φαίνεται να υπάρχει κάποια περιοδικότητα στη ζήτηση η οποία οφείλεται στην εποχικότητα.



Σχήμα 2.5. Πρωτογενή δεδομένα ωριαίας ζήτησης ενέργειας για την Ελλάδα

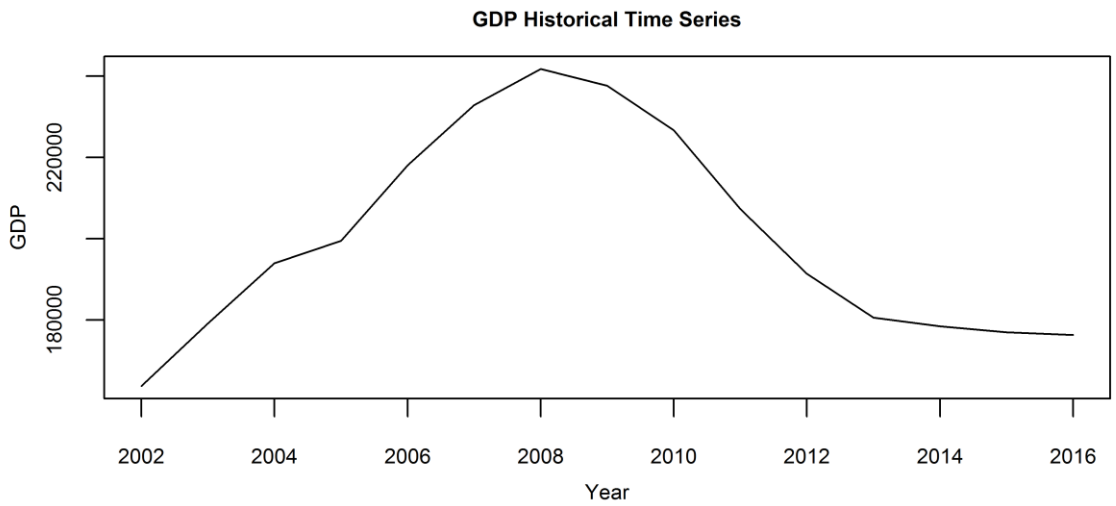
2.2.2 Ακαθάριστο Εθνικό Προϊόν (ΑΕΠ)

Η The World Bank (2016) παρέχει πρωτογενή δεδομένα του ΑΕΠ σε ετήσια βάση, τα οποία παρουσιάζει ο Πίνακας 2.1. Στην παρούσα εργασία, για λόγους απλοποίησης της ανάλυσης, ορίζουμε ως υδρολογικό έτος το έτος το οποίο αρχίζει το Σεπτέμβριο της μιας χρονιάς και λήγει τον Αύγουστο της επόμενης. Υπολογίζουμε το ΑΕΠ του υδρολογικού έτους σύμφωνα με την εξ. (2.1). Στο Σχήμα 2.6 απεικονίζεται το ΑΕΠ της χώρας για την περίοδο 2002 έως 2016.

$$\text{ΑΕΠ}_{\text{υδρ},i} = (1/3)\ \text{ΑΕΠ}_i + (2/3)\ \text{ΑΕΠ}_{i+1},\ i = 2002, 2003, \ldots \qquad (2.1)$$

Πίνακας 2.1. Ελληνικό Ετήσιο Ακαθάριστο Εθνικό Προϊόν (Πηγή: The World Bank 2016).

Έτος	Ετήσιο Ακαθάριστο Εθνικό Προϊόν (10^6 €)	Υδρολογικό έτος	Ακαθάριστο Εθνικό Προϊόν υδρολογικού έτους (10^6 €)
2002	163703.7	2002	173982.8
2003	179122.4	2003	188990.1
2004	193923.9	2004	197633.0
2005	199487.5	2005	211848.9
2006	218029.6	2006	227930.7
2007	232881.2	2007	238795.9
2008	241753.3	2008	239023.0
2009	237657.9	2009	230351.0
2010	226697.6	2010	213755.6
2011	207284.6	2011	196674.9
2012	191370.1	2012	184157.2
2013	180550.7	2013	179136.2
2014	178429.0	2014	177458.3
2015	176973.0	2015	176532.6
2016	176312.4	2016	

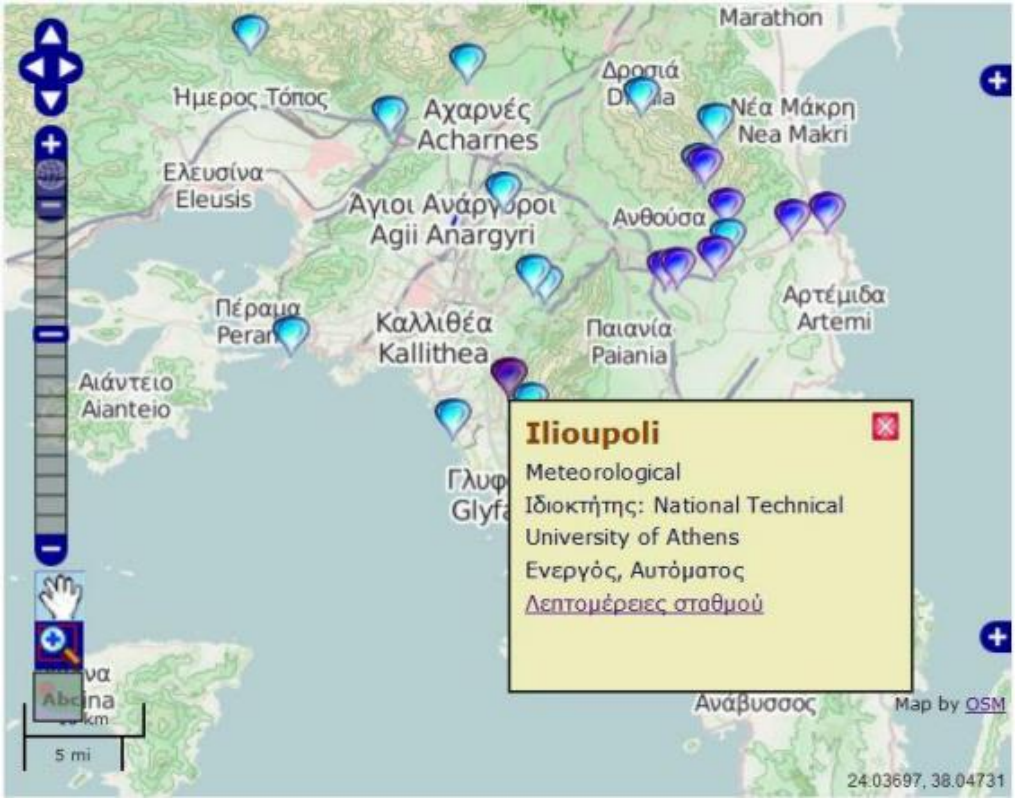


Σχήμα 2.6. Πρωτογενή δεδομένα ετήσιου ΑΕΠ για την Ελλάδα

2.2.3 Θερμοκρασία

Χρησιμοποιήθηκαν δεδομένα θερμοκρασίας του σταθμού της Ηλιούπολης στην Αττική (The Hydrological Observatory of Athens 2016). Ο σταθμός βρίσκεται κοντά στη θάλασσα (Σχήμα 2.7). Θεωρούμε ότι για αυτή την ανάλυση η θερμοκρασία του σταθμού είναι αντιπροσωπευτική για τα δεδομένα της Ελλάδας, διότι στην Αττική κατοικεί το 35% του πληθυσμού της Ελλάδας, είναι συγκεντρωμένη η κύρια βιομηχανική δραστηριότητα, και επίσης λόγω του ήπιου κλίματος στην Αττική και της θέσης του σταθμού είναι λιγότερο ευαίσθητος σε τοπικά ακραία καιρικά φαινόμενα. Επομένως, η θερμοκρασία του σταθμού είναι αντιπροσωπευτική της θερμοκρασίας στην Ελλάδα, τουλάχιστον για τους σκοπούς της

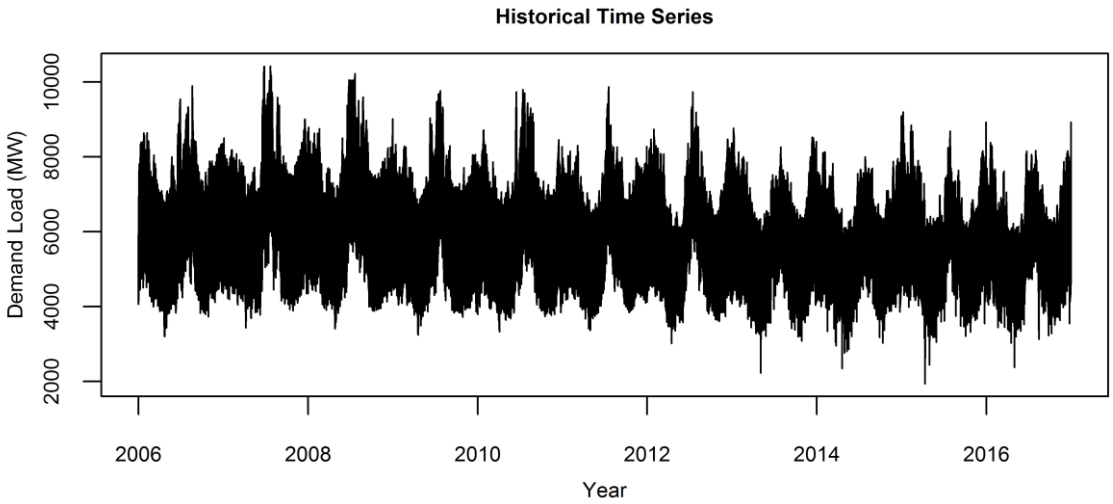
ανάλυσης, παρά τις υψηλές θερμοκρασίες της Ελλάδας λόγω του ορεινού εδάφους της. Επιπλέον, η επιλογή ενός μόνο σταθμού για την ανάλυση δικαιολογείται από την επιλογή μόνο δύο σταθμών από τους Hyndman and Fan (2010) στο προηγμένο μοντέλο πρόβλεψης ζήτησης που χρησιμοποίησαν.



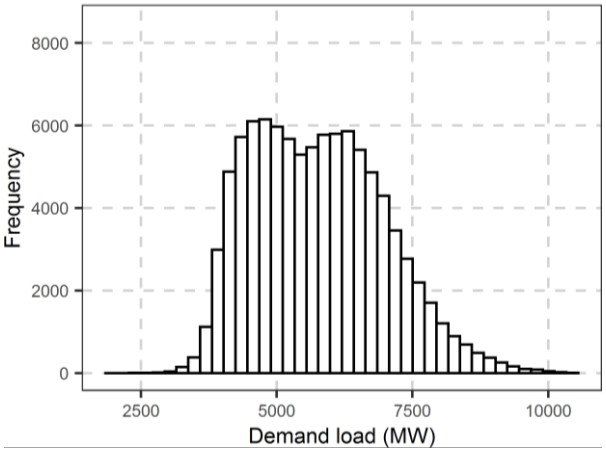
Σχήμα 2.7. Θέση του σταθμού της Ηλιούπολης

2.3 Επεξεργασία Δεδομένων

Πριν αναλύσουμε τα δεδομένα της ενέργειας, προχωρήσαμε σε μια διαδικασία επεξεργασίας και καθαρισμού τους. Όπως φαίνεται από το Σχήμα 2.5 παρατηρούμε κάποιες μηδενικές τιμές, ενώ σε κάποιες περιπτώσεις έλειπαν ορισμένες τιμές (NA). Μετά από διερεύνηση, φαίνεται ότι το φορτίο ζήτησης έχει οριστεί σε NA μερικές Κυριακές τον Μάρτιο. Αυτό συμβαίνει λόγω της θερινής ώρας (DST). Πρώτον, αντικαθιστούμε όλες τις τιμές NA με μηδέν και μετά αντικαθιστούμε όλα τα κελιά που το φορτίο ζήτησης είναι μηδέν με το μέσο όρο των γειτονικών τους κελιών χρησιμοποιώντας την αντίστοιχη εντολή του R (na.approx). Ανάλογη μεθοδολογία ακολουθείται και με τη χρονοσειρά της θερμοκρασίας με τη διαφορά ότι η τιμή 0 είναι σε ορισμένες περιπτώσεις αποδεκτή. Μετά την επεξεργασία των δεδομένων απεικονίζεται η ζήτηση ενέργειας στο Σχήμα 2.8 και το ιστόγραμμα των ωριαίων τιμών στο Σχήμα 2.9. Όπως φαίνεται στο Σχήμα 2.9, το διάγραμμα έχει δύο κορυφές περίπου στα 5.000 MW και τα 6.400 MW για τη ζήτηση ενέργειας.



Σχήμα 2.8. Ωριαία ζήτηση ενέργειας



Σχήμα 2.9. Ιστόγραμμα ωριαίας ζήτησης ενέργειας

ΚΕΦΑΛΑΙΟ 3

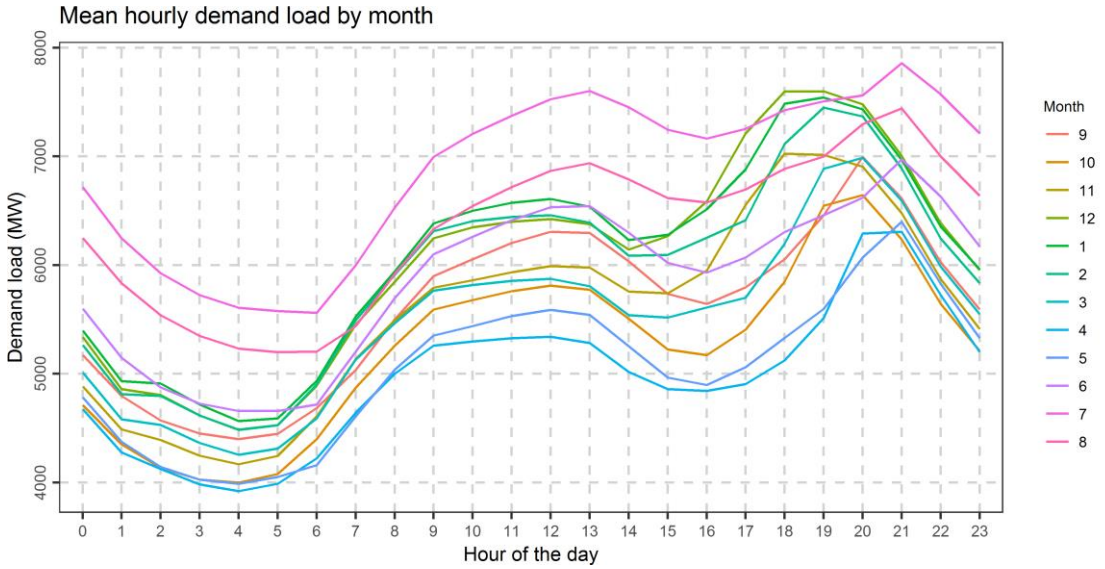
ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

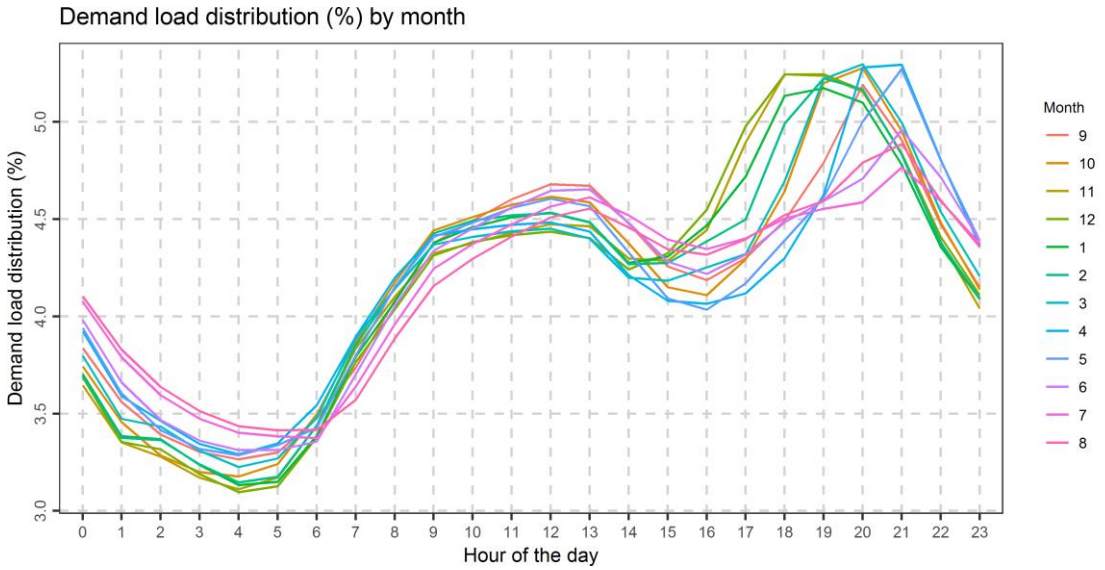
3.1 Οπτική Διερεύνηση για διάφορες χρονικές κλίμακες

Στην Ενότητα 3.1 πραγματοποιείται οπτική διερεύνηση της ζήτησης ενέργειας για διάφορες χρονικές κλίμακες. Η οπτικοποίηση των δεδομένων συχνά βοηθάει στην καλύτερη κατανόηση όχι μόνο των ίδιων των δεδομένων, αλλά και των συσχετίσεων που μπορεί να υπάρχουν μεταξύ τους. Τα δεδομένα τα οποία εξετάζονται αφορούν την Ελλάδα και τη χρονική περίοδο από 2006-09-01 έως 2016-08-31 (10 υδρολογικά έτη). Προκειμένου όλες οι χρονοσειρές να έχουν κοινά χρονικά πλαίσια επιλέγουμε μόνο τα δεδομένα που αφορούν τις παραπάνω ημερομηνίες, «κόβουμε» δηλαδή τις αρχικές χρονοσειρές. Αναζητούμε τη σχέση μεταξύ της ενεργειακής ζήτησης με το ΑΕΠ και τη θερμοκρασία. Από την ανάλυση προκύπτει ότι, ενώ σε γενικές γραμμές τα συμπεράσματα των Psiloglou et al. (2009) εξακολουθούν να ισχύουν, ωστόσο από το 2010 όταν το ΑΕΠ άρχισε να μειώνεται, το σχήμα της ενεργειακής ζήτησης άλλαξε κατά τη χειμερινή περίοδο. Τα αποτελέσματα της ανάλυσης είναι χρήσιμα για την πρόβλεψη ζήτησης ενέργειας και για την παραγωγή συνθετικών χρονοσειρών. Τα δεδομένα και ο κώδικας από τον οποίον προέκυψαν τα αποτελέσματα της εργασίας, διατίθενται ως υποστηρικτικό υλικό.

Οι χρονικές κλίμακες που εξετάζουμε είναι η ωριαία, η ημερήσια, η μηνιαία και η ετήσια ξεκινώντας από τη μικρότερη στη μεγαλύτερη. Στο Σχήμα 3.1 παρουσιάζεται η μέση ωριαία ζήτηση ενέργειας για κάθε μήνα, ενώ στο Σχήμα 3.2 η κατανομή της στη διάρκεια της ημέρας. Οι ελάχιστες τιμές ζήτησης παρουσιάζονται περίπου στις 03:00 – 05:00 καθημερινά, ενώ παρατηρούνται δύο τοπικά μέγιστα. Το πρώτο τοπικό μέγιστο παρατηρείται στις 13:00 ανεξαρτήτως του μήνα, ενώ το δεύτερο τοπικό μέγιστο παρατηρείται στις 18:00 για τους χειμερινούς μήνες και στις 21:00 για τους θερινούς μήνες. Το πρώτο οφείλεται στην αυξημένη ζήτηση ηλεκτρικής ενέργειας για βιομηχανίες, υπηρεσίες και οικιακές δραστηριότητες, ενώ το δεύτερο στη θέρμανση και ψύξη των νοικοκυριών. Επιπρόσθετα, το δεύτερο τοπικό μέγιστο είναι σημαντικά υψηλότερο σε σύγκριση με το προηγούμενο κατά τους χειμερινούς μήνες πιθανώς λόγω της αυξημένης ανάγκης θέρμανσης τη νύχτα, ενώ και τα δύο μέγιστα είναι περίπου ίσα τους θερινούς μήνες, όπου η ψύξη είναι κυρίως απαραίτητη κατά τις απογευματινές ώρες. Η αυξανόμενη τάση, που παρατηρείται από τις 06:00 έως τις 07:00 το καλοκαίρι, κινείται μια ώρα νωρίτερα κατά τους υπόλοιπους μήνες. Οι τιμές ζήτησης ηλεκτρικής ενέργειας κυμαίνονται μεταξύ αυτών των ελάχιστων ή μέγιστων τιμών την άνοιξη και το φθινόπωρο.

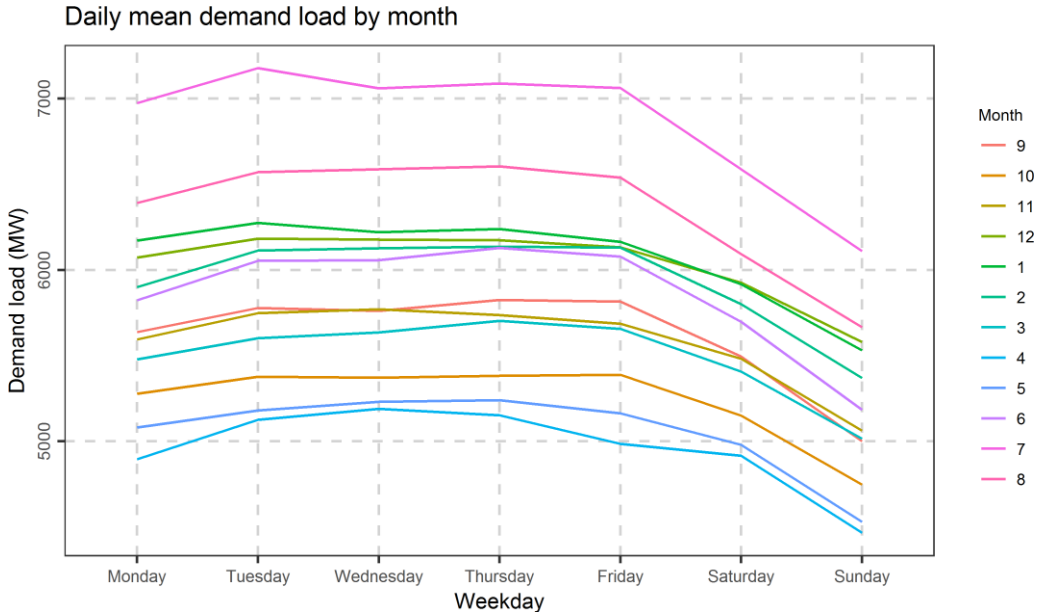


Σχήμα 3.1. Μέση ωριαία ζήτηση ενέργειας ανά μήνα

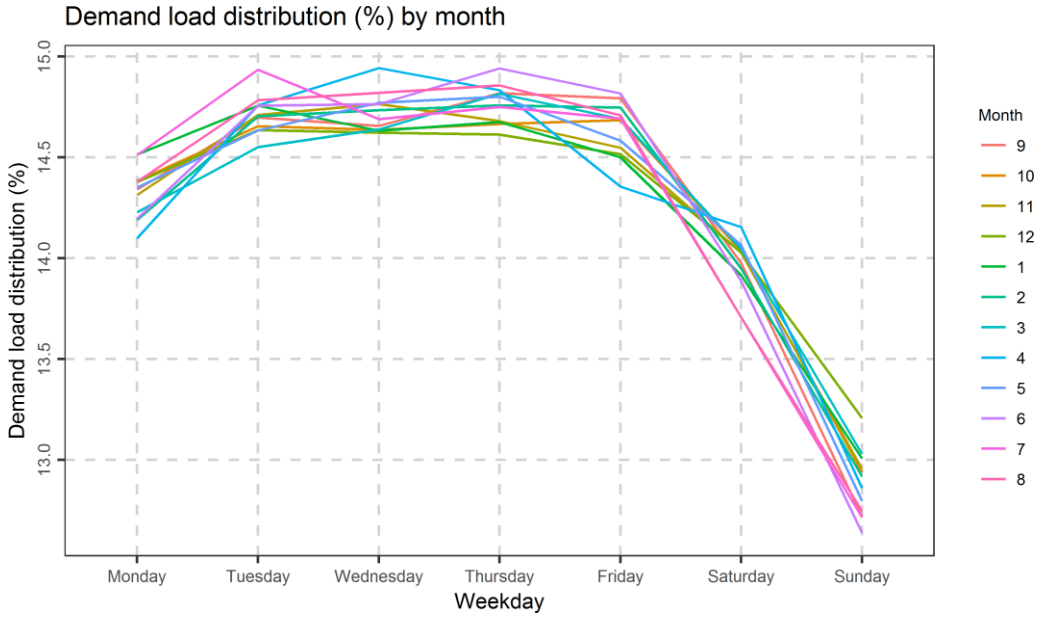


Σχήμα 3.2. Κατανομή στην ημέρα μέσης ωριαίας ζήτησης ενέργειας ανά μήνα

Στο Σχήμα 3.3 παρουσιάζεται η μέση ημερήσια ζήτηση ενέργειας για κάθε μήνα, ενώ στο Σχήμα 3.4 παρουσιάζεται η κατανομή της στη διάρκεια της βδομάδας. Για κάθε μήνα η κατανάλωση ενέργειας κατανέμεται περίπου όμοια στη διάρκεια της εβδομάδας, με τη μέγιστη διακύμανση να παρατηρείται την Κυριακή όπου η μέγιστη από την ελάχιστη τιμή να διαφέρουν περίπου 0.60%. Ο μόνος μήνας που φαίνεται να διαφέρει σημαντικά, ως προς το σχήμα ζήτησης είναι ο Μάρτιος με μια μεγάλη μείωση μεταξύ Πέμπτης και Παρασκευής. Τα μέγιστα της μέσης ημερήσιας ζήτησης ενέργειας παρατηρούνται κάθε Τετάρτη ή Πέμπτη, ενώ τα ελάχιστά τις Κυριακές (και τα Σάββατα). Τα ελάχιστα οφείλονται στη μειωμένη οικονομική δραστηριότητα τα σαββατοκύριακα, κάτι το οποίο επισημαίνεται από τους Psiloglou et al. (2009). Η μέση ημερήσια ζήτηση είναι ίση τη Δευτέρα και την Παρασκευή με τη ζήτηση να αυξάνεται τις άλλες εργάσιμες ημέρες της εβδομάδας.



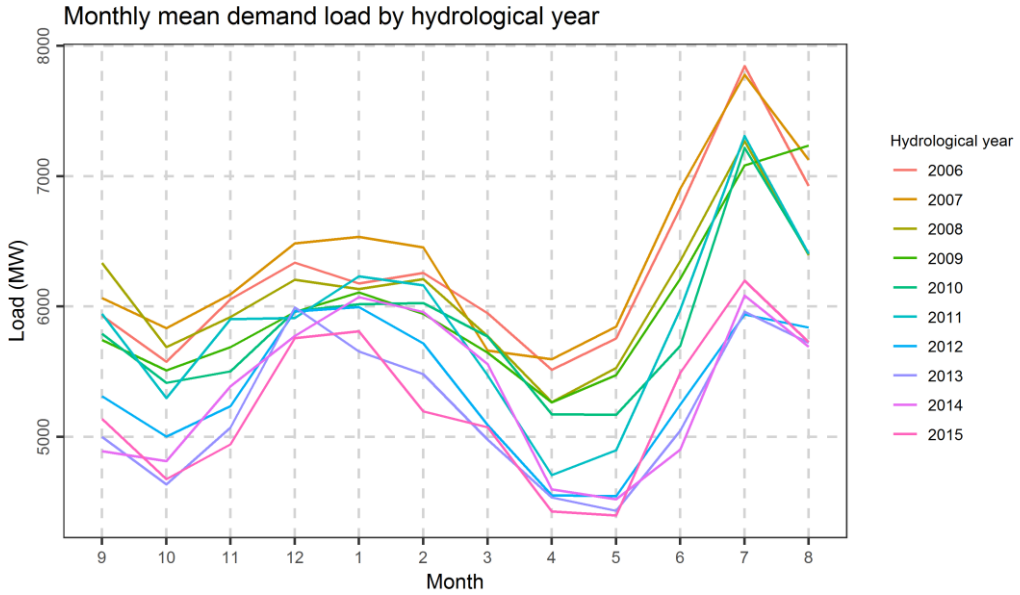
Σχήμα 3.3. Μέση ημερήσια ζήτηση ενέργειας ανά μήνα



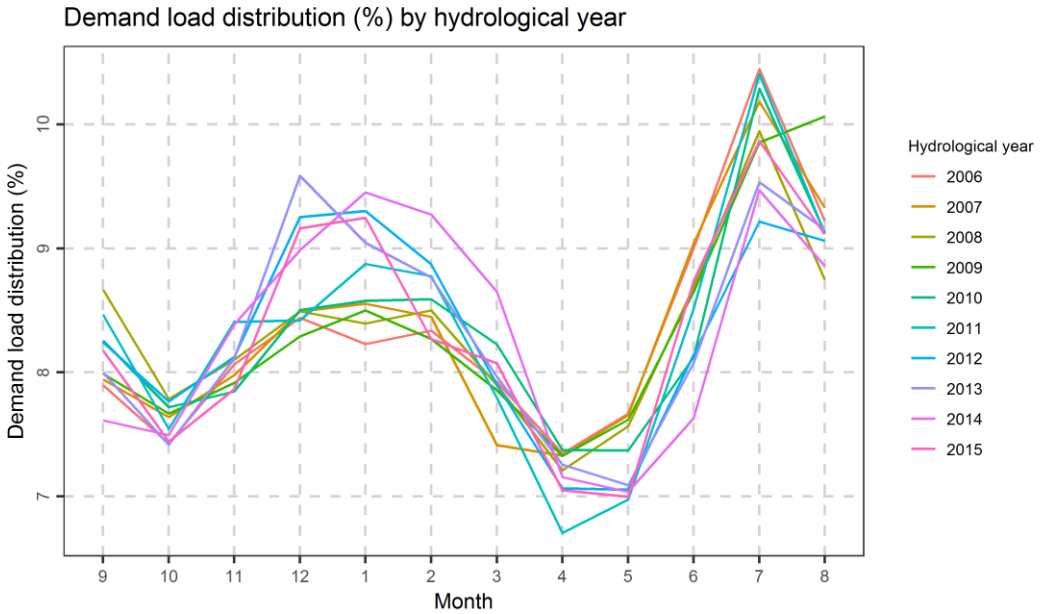
Σχήμα 3.4. Κατανομή μέσης ημερήσιας ζήτησης ενέργειας ανά μήνα

Στο Σχήμα 3.5 παρουσιάζεται η μέση μηνιαία ζήτηση ενέργειας για κάθε υδρολογικό έτος. Παρατηρούμε ότι από το υδρολογικό έτος 2007 και έπειτα παρατηρείται μια πτώση της ζήτησης. Αυτό είναι περισσότερο φανερό στο Σχήμα 3.7, όπου φαίνεται η μεγάλη μείωση ζήτησης ενέργειας, η οποία παρατηρήθηκε το υδρολογικό έτος 2013. Στο Σχήμα 3.6 παρουσιάζεται η κατανομή της μέσης μηνιαίας ζήτησης στη διάρκεια του υδρολογικού έτους. Είναι αξιοσημείωτη η αλλαγή στο σχήμα ζήτησης τους μήνες Δεκέμβριο και Ιανουάριο για τα υδρολογικά έτη 2012 και 2013, όπου παρατηρείται μεγάλη αύξηση. Αυτό είναι περισσότερο εμφανές στο Σχήμα 3.8 για το υδρολογικό έτος 2013, όπου παραλείφθηκαν τα ενδιάμεσα υδρολογικά έτη. Στο Σχήμα 3.9 παρουσιάζεται και η διακύμανση της μέσης μηνιαίας ζήτησης για το σύνολο των υδρολογικών ετών. Παρατηρείται μικρή διακύμανση για τους μήνες Δεκέμβριο – Απρίλιο, ωστόσο υπάρχουν κάποιες ακραίες τιμές, οι οποίες προκύπτουν από την αλλαγή του σχήματος ζήτησης κατά

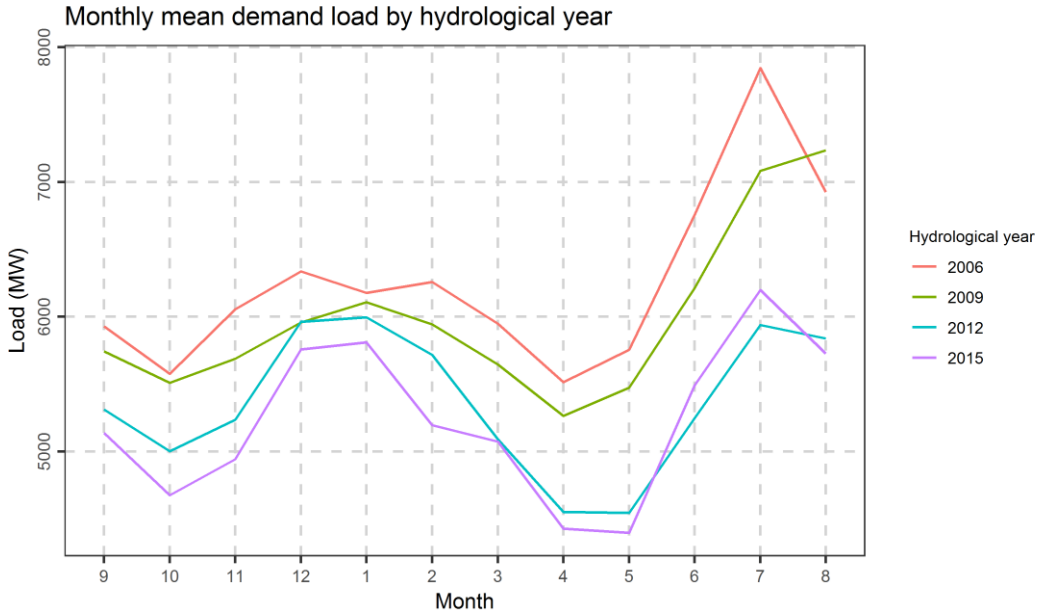
τα υδρολογικά έτη 2012 και 2013. Στα Σχήματα 3.5-3.8, παρατηρείται ότι το μοτίβο της μέσης μηνιαίας ζήτησης Δεκεμβρίου και Ιανουαρίου έχει αλλάξει σημαντικά από το υδρολογικό έτος 2012. Αυτό παρατηρήθηκε αρχικά από τους Dagoumas and Kitsios (2014). Η μελέτη τους αναφέρεται συγκεκριμένα στο υδρολογικό έτος 2012, όπου αποδίδουν αυτήν την αλλαγή στη σημαντική αύξηση του φόρου για το πετρέλαιο θέρμανσης, γεγονός που οδήγησε σε αύξηση της χρήσης της ηλεκτρικής ενέργειας για θέρμανση. Η αύξηση του φόρου επί του πετρελαίου θέρμανσης ήταν αποτέλεσμα της οικονομικής πολιτικής που υιοθέτησε η ελληνική κυβέρνηση λόγω της μείωσης του ΑΕΠ και της οικονομικής κρίσης, η οποία ξεκίνησε το έτος 2010.



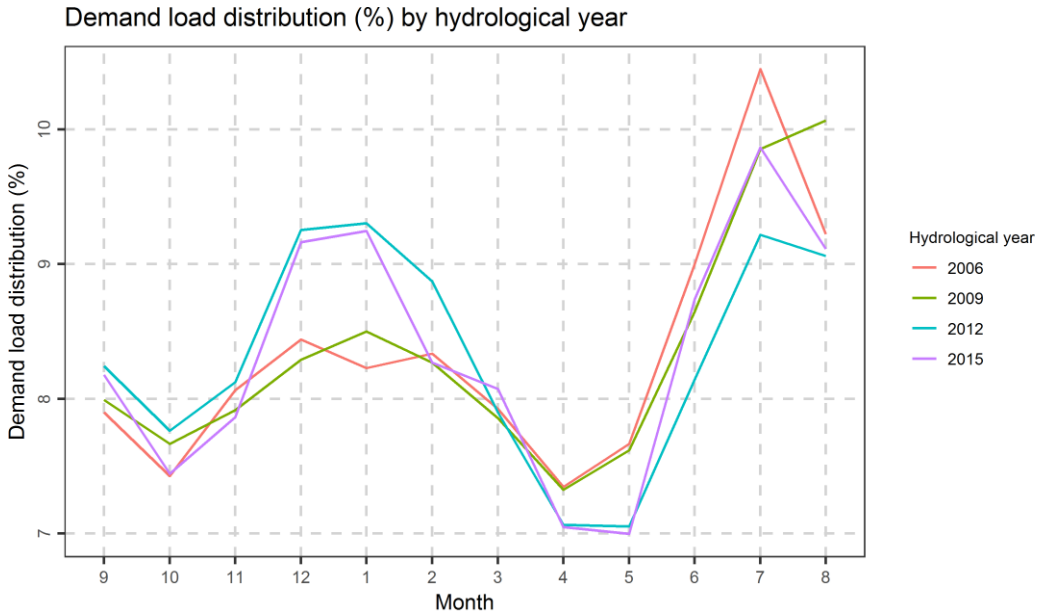
Σχήμα 3.5. Μέση μηνιαία ζήτηση ενέργειας ανά υδρολογικό έτος



Σχήμα 3.6. Κατανομή μέσης μηνιαίας ζήτησης ενέργειας στο υδρολογικό έτος

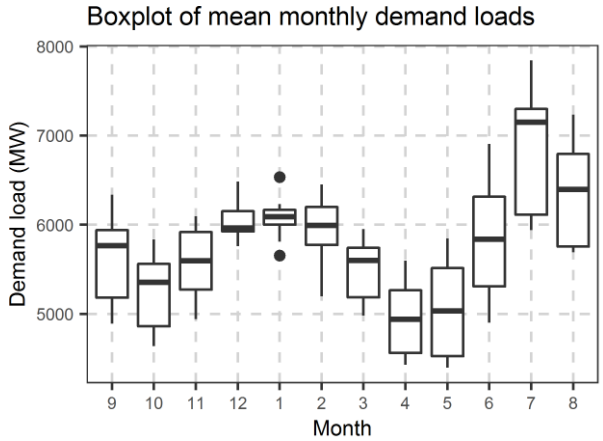


Σχήμα 3.7. Μέση μηνιαία ζήτηση ενέργειας ανά υδρολογικό έτος, για τα υδρολογικά έτη 2006, 2009, 2012 και 2015



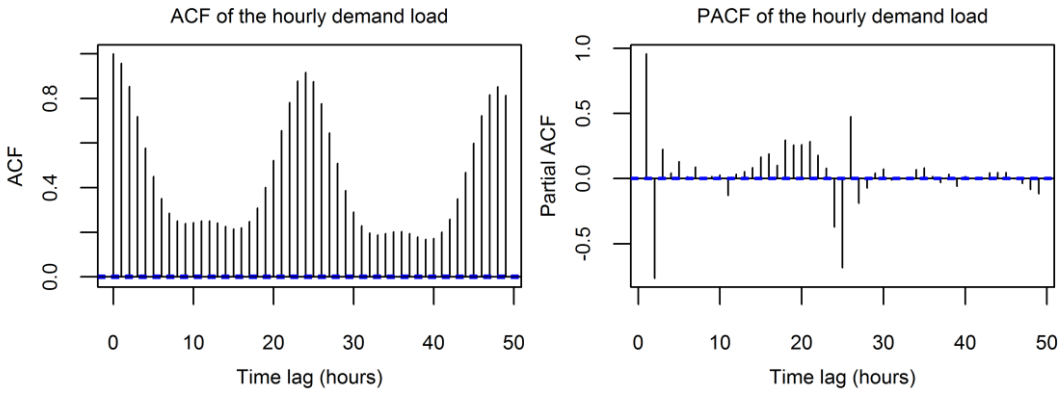
Σχήμα 3.8. Κατανομή μέσης μηνιαίας ζήτησης ενέργειας στο υδρολογικό έτος, για τα υδρολογικά έτη 2006, 2009, 2012 και 2015

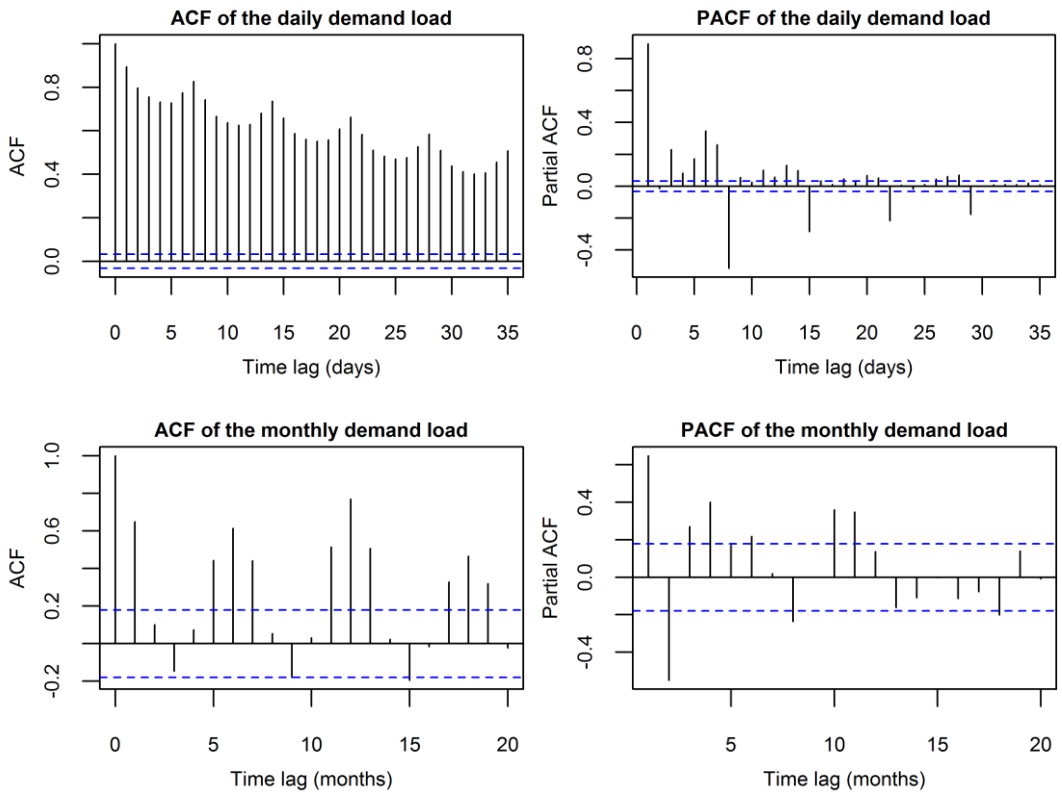
Το Σχήμα 3.9 παρουσιάζει τη διακύμανση της μέσης μηνιαίας ζήτησης. Οι υψηλότερες αποκλίσεις από το μέσο όρο εμφανίζονται τους καλοκαιρινούς μήνες Ιούλιο και Αύγουστο.



Σχήμα 3.9. Boxplot με τα ποσοστημόρια 0.25 και 0.75 για τις μέσες μηνιαίες τιμές ζήτησης ενέργειας

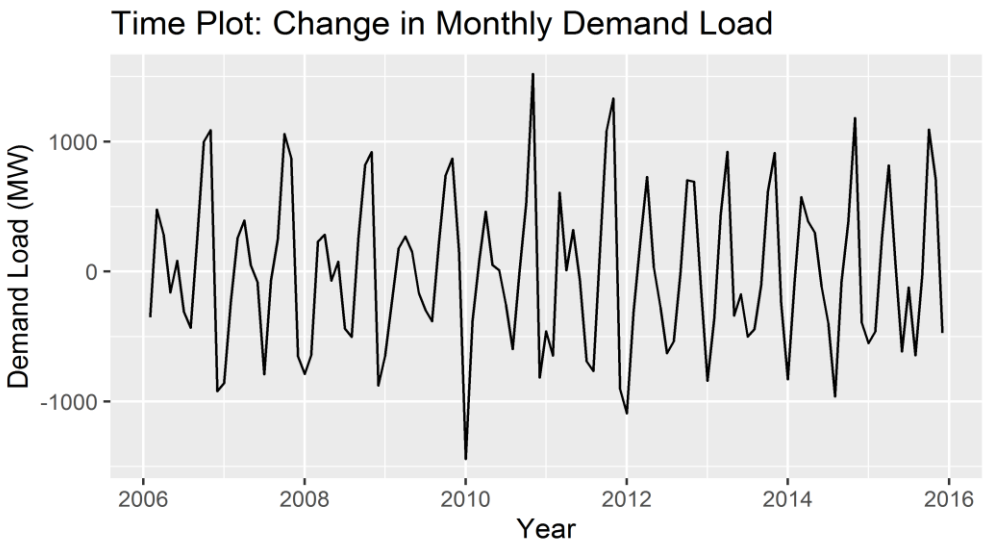
Το Σχήμα 3.10 αντιπροσωπεύει τις συναρτήσεις αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) για τη ζήτηση ενέργειας. Στην ανάλυση χρονοσειρών , η ACF είναι μια συνάρτηση που μας δίνει τιμές αυτοσυσχέτισης οποιασδήποτε σειράς με τις προηγούμενες τιμές της. Με απλούς όρους, περιγράφει πόσο καλά η παρούσα αξία της σειράς σχετίζεται με τις προηγούμενες τιμές της. Μια χρονοσειρά μπορεί να έχει στοιχεία όπως τάση, εποχικότητα, κυκλοσταστιμότητα. Η ACF εξετάζει την ύπαρξη όλων αυτών των στοιχείων και υποδεικνύει την ύπαρξή τους. Η PACF δίνει τη μερική συσχέτιση μιας στάσιμης χρονοσειράς με τις δικές τις τιμές σε προηγούμενα χρονικά βήματα με τις σχέσεις των παρεμβαλλόμενων παρατηρήσεων να αφαιρούνται. Παρατηρούμε πως, η συνάρτηση αυτοσυσχέτισης σε ωριαία κλίμακα μεγιστοποιείται κάθε 24 ώρες και πολλαπλασιάζεται, δηλαδή υπάρχει περιοδικότητα 24 ωρών. Στα διαγράμματα της ημερήσιας ζήτησης, υποδεικνύεται η ύπαρξη ενός περιοδικού μοτίβου 7 ημερών. Η συσχέτιση φαίνεται να είναι ισχυρή ακόμη και για χρονικές καθυστερήσεις ίσες με 35 ημέρες. Τέλος, στο παρακάτω σχήμα, παρουσιάζονται τόσο το διάγραμμα του ACF όσο και του PACF της μέσης μηνιαίας ζήτησης ενέργειας. Παρατηρείται ένα περιοδικό μοτίβο 6 μηνών, ωστόσο η αυτοσυσχέτιση είναι ισχυρότερη κάθε 12 μήνες.



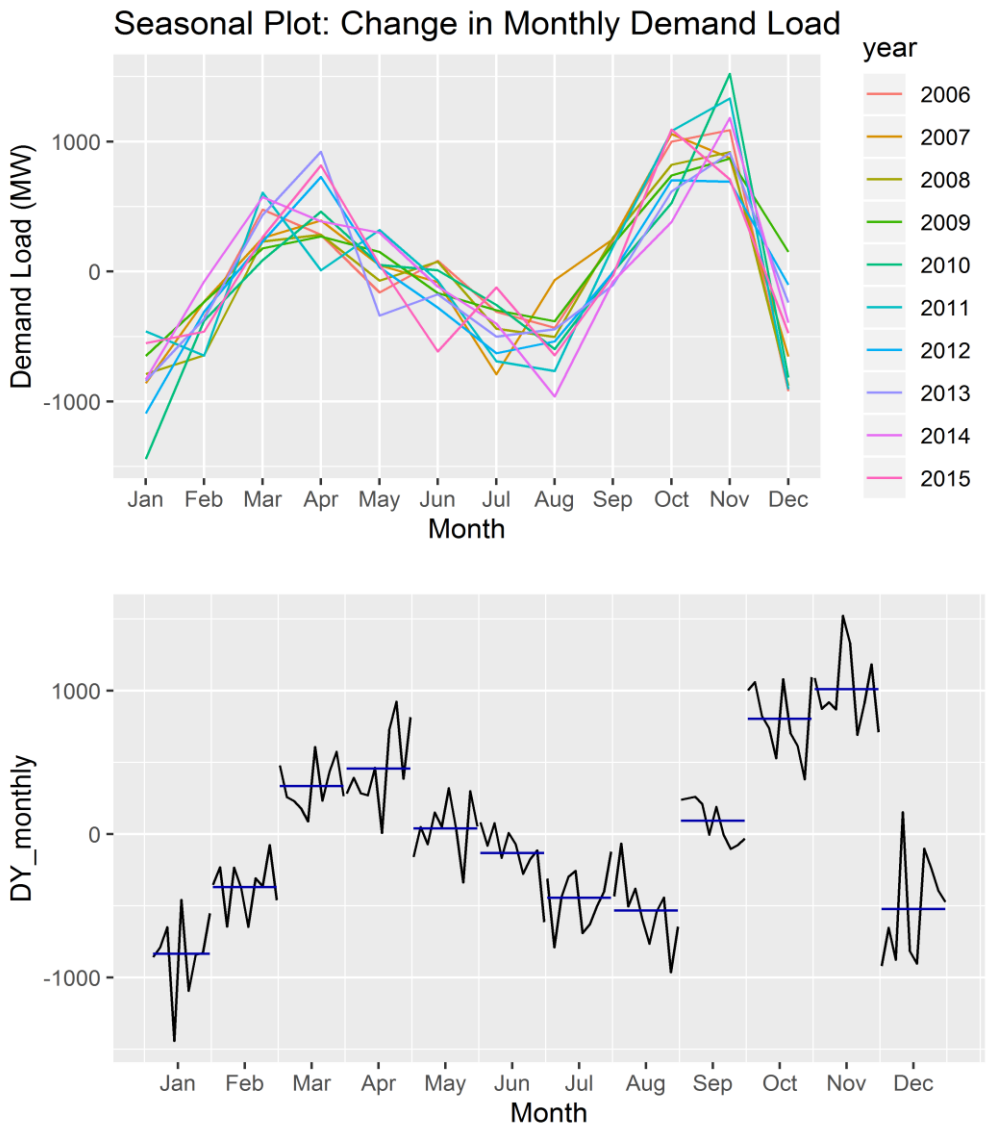


Σχήμα 3.10. ACF (αριστερά) και PACF (δεξιά) για τη ζήτηση ενέργειας σε ωριαία κλίμακα (άνω), σε ημερήσια κλίμακα (μέση) και μηνιαία κλίμακα (κάτω)

Το Σχήμα 3.11 παρουσιάζει τις αλλαγές της μηνιαίας ζήτησης για το χρονικό διάστημα από το 2006 έως το 2016. Είναι αξιοσημείωτη η μεταβολή της ζήτησης του φορτίου κατά το έτος 2010 (μείωση πάνω από 15%) και επαληθεύεται η ύπαρξη εποχικότητας, αλλά και της 12-μηνης περιοδικότητας. Στο Σχήμα 3.12 παρουσιάζονται οι μηνιαίες μεταβολές της ζήτησης ενέργειας για κάθε έτος, με τις πιο σημαντικές αυξομειώσεις να παρατηρούνται κατά τον Ιανουάριο και το Νοέμβριο αλλά και η διακύμανση της ζήτησης του κάθε μήνα με τις μέσες τιμές μεταβολής του.



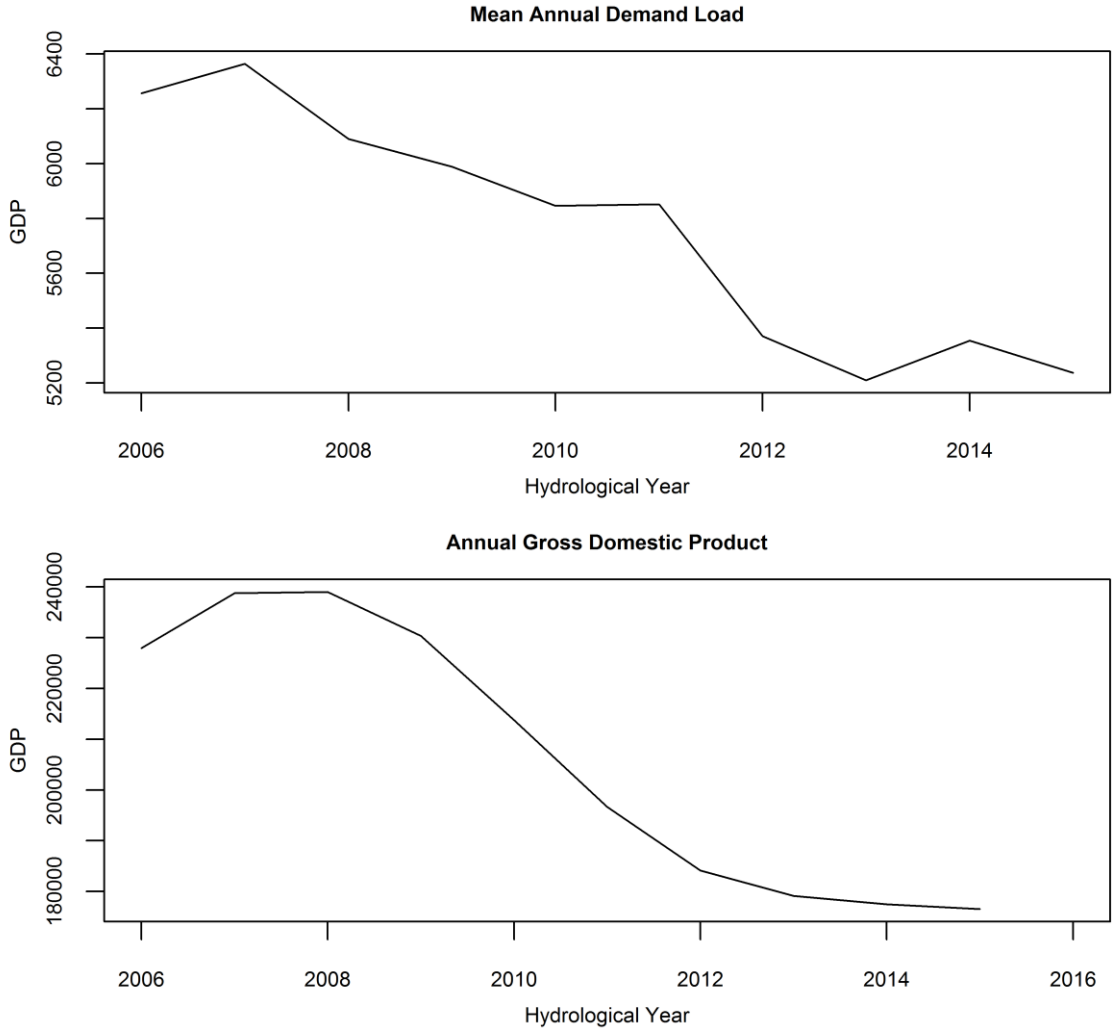
Σχήμα 3.11. Μηνιαία μεταβολή ημερήσιας ζήτησης ενέργειας



Σχήμα 3.12. Μεταβολή μηνιαίας ζήτησης ενέργειας ανά έτος (άνω) και μηνιαίας διακύμανσης της ίδιας μεταβολής ανά μήνα (κάτω)

3.2 Ζήτηση Ενέργειας και ΑΕΠ

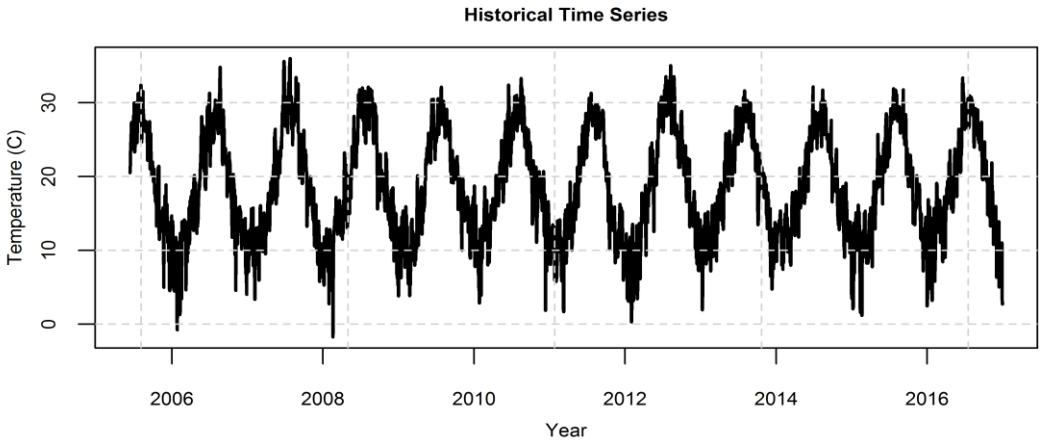
Στο Σχήμα 3.13 παρουσιάζεται η σχέση του ΑΕΠ και της μέσης ετήσιας ζήτησης ενέργειας ανά υδρολογικό έτος η οποία μοιάζει να είναι γραμμική. Παρατηρούμε ότι το υδρολογικό έτος 2008, παρά την αύξηση του ΑΕΠ, υπάρχει μεγάλη μείωση της μέσης ετήσιας ζήτησης ενέργειας, το οποίο κατά την άποψη μας ίσως εξηγείται από το γεγονός ότι το ΑΕΠ δεν είναι από μόνο του αντιπροσωπευτικό της οικονομικής κατάστασης.



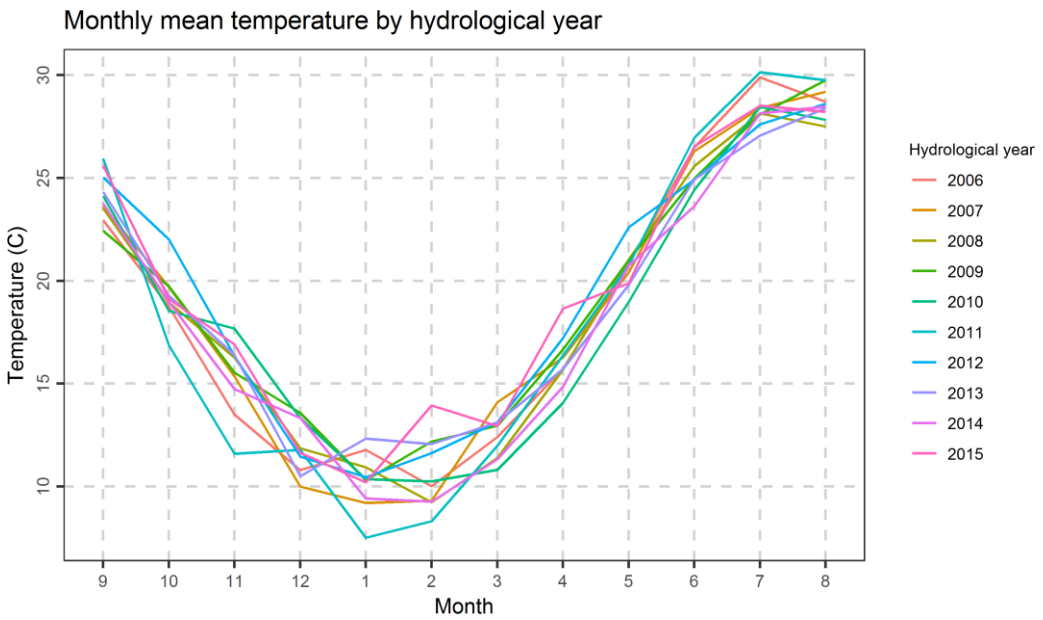
Σχήμα 3.13. Ακαθάριστο Εθνικό Προϊόν (άνω) και μέση ετήσια ζήτηση ενέργειας (κάτω) ανά υδρολογικό έτος

3.3 Ζήτηση Ενέργειας και Θερμοκρασία

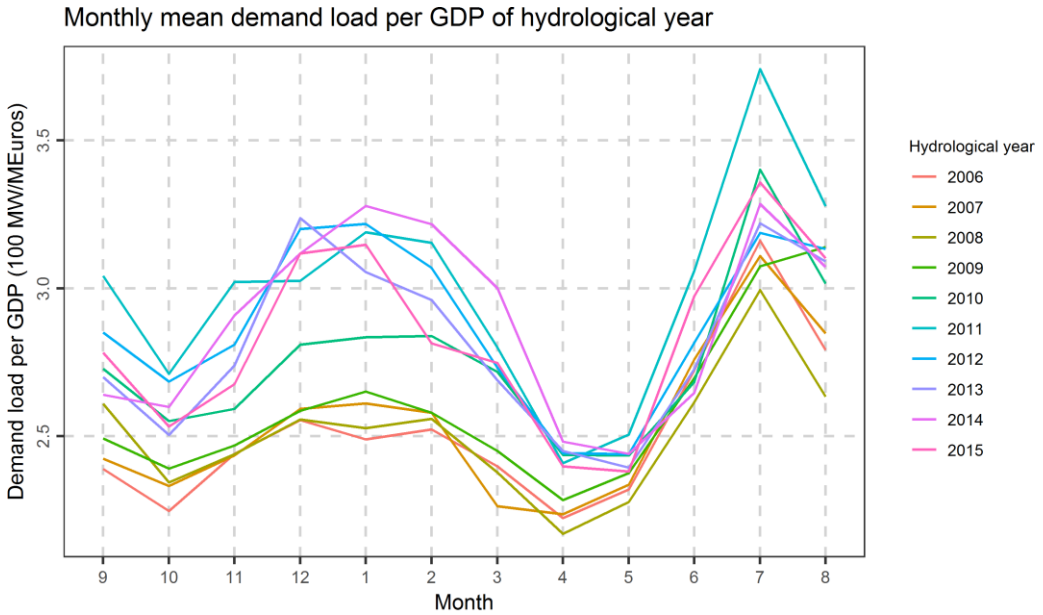
Στο Σχήμα 3.14 παρουσιάζεται η μέση ημερήσια θερμοκρασία του σταθμού της Ηλιούπολης όπου και παρατηρείται μια περιοδικότητα 12 μηνών. Στα Σχήματα 3.15 και 3.16 παρατηρούμε ότι τα τοπικά μέγιστα στη ζήτηση ενέργειας παρουσιάζονται τη χειμερινή περίοδο, που παρατηρούνται οι χαμηλότερες θερμοκρασίες και τη θερινή περίοδο που παρατηρούνται οι υψηλότερες θερμοκρασίες. Αυτό επιβεβαιώνεται και από το Σχήμα 3.17, όπου παρατηρούμε ένα κυρτό σχήμα στο διάγραμμα θερμοκρασίας και ζήτησης ενέργειας. Μάλιστα, το Σχήμα 3.17 είναι παρόμοιο με αυτό των Psiloglou et al. (2009) για την Αθήνα, το οποίο αναφέρεται στη χρονική περίοδο 1997-2001, των Mirasgedis et al. (2006) και Tyralis (2016) για την Ελλάδα, τα οποία αναφέρονται στη χρονική περίοδο 1993-2002 και 2002-2014 αντίστοιχα αλλά και με αυτό των Moral-Carcedo and Vicens-Otero (2005) για την Ισπανία με παρόμοιο κλίμα. Ωστόσο, διαφέρει από αυτό των Hor et al. (2005) για το Ηνωμένο Βασίλειο.



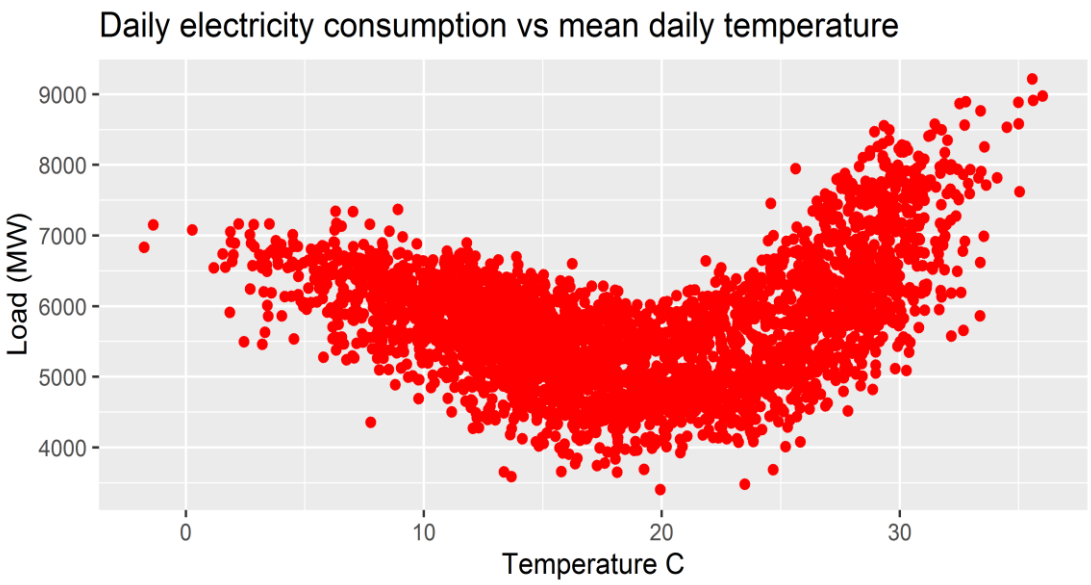
Σχήμα 3.14. Μέση ημερήσια θερμοκρασία στο σταθμό Ηλιούπολης



Σχήμα 3.15. Μέση μηνιαία θερμοκρασία ανά υδρολογικό έτος στο σταθμό Ηλιούπολης



Σχήμα 3.16. Μέση μηνιαία ζήτηση ενέργειας προς Ακαθάριστο Εθνικό Προϊόν ανά υδρ. έτος



Σχήμα 3.17. Διάγραμμα μέσης ημερήσιας θερμοκρασίας στο σταθμό Ηλιούπολης και ημερήσιας ζήτησης ενέργειας

Επισημαίνεται για μία ακόμη φορά πως η διακύμανση της θερμοκρασίας με τη ζήτηση ηλεκτρικής ενέργειας είναι μη γραμμική, και μάλιστα η αύξηση ή μείωση της πρώτης οδηγεί πάντα σε αύξηση της τελευταίας (γεγονός που οφείλεται κυρίως στη χρήση ηλεκτρικών συσκευών θέρμανσης το χειμώνα και κλιματιστικών το καλοκαίρι). Για παράδειγμα, κατά τη διάρκεια του καλοκαιριού, μια αύξηση της μέσης ημερήσιας θερμοκρασίας 1 °C (ας πούμε, από 24 σε 25 °C) θα είχε ως αποτέλεσμα την αύξηση της κατανάλωσης της ηλεκτρικής ενέργειας περίπου 2,6%. Το σύστημα ισχύος αναφοράς φαίνεται να είναι λιγότερο ευαίσθητο στις διακυμάνσεις της θερμοκρασίας το χειμώνα, καθώς η πτώση της μέσης ημερήσιας θερμοκρασίας 1 °C (ας πούμε, από 14 σε 13 °C) θα είχε ως αποτέλεσμα αύξηση της κατανάλωσης ηλεκτρικής ενέργειας περίπου 1,1%. Αυτό οφείλεται κυρίως στο γεγονός ότι οι τελικοί καταναλωτές μπορούν να χρησιμοποιούν μια ποικιλία πηγών ενέργειας για θέρμανση (π.χ. πετρέλαιο, φυσικό αέριο, ηλεκτρικό ρεύμα κ.λπ.) και πρακτικά μόνο ηλεκτρική ενέργεια για ψύξη. Από την άλλη πλευρά, στους 18,5 °C περίπου, η επίδραση της θερμοκρασίας ελαχιστοποιείται και η ζήτηση ηλεκτρικής ενέργειας είναι ανελαστική στις αλλαγές θερμοκρασίας, κάτι το οποίο επισημαίνεται από τους Lalas et al. (2003).

ΚΕΦΑΛΑΙΟ 4

ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ

ΚΕΦΑΛΑΙΟ 4: ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ

4.1 Στατιστικές Μέθοδοι Πρόβλεψης

Οι στατιστικές προσεγγίσεις βασίζονται στη μαθηματική αναπαράσταση της καμπύλης του ηλεκτρικού φορτίου. Το ηλεκτρικό φορτίο μελετάται ως συνάρτηση διάφορων παραγόντων, όπως οι καιρικές συνθήκες, ο τύπος της ημέρας, το ΑΕΠ κ.α. Προκειμένου να γίνει μια σωστή πρόβλεψη προηγήθηκε ανάλυση των δεδομένων μας. Τα αποτελέσματα της μελέτης μας επιβεβαιώνουν και συνοψίζουν τα συμπεράσματα προηγούμενων σχετικών μελετών στην Ελλάδα, καθεμία από τις οποίες όμως αντιμετωπίζει ένα μόνο θέμα και καλύπτει μικρότερες και παλαιότερες χρονικές περιόδους. Επίσης, κάποια απροσδόκητα μοτίβα αποκαλύφθηκαν. Τα αποτελέσματα της ανάλυσης θα χρησιμοποιηθούν στη συνέχεια για την ανάπτυξη ενός συστήματος βραχυπρόθεσμης πρόβλεψης ηλεκτρικής ενέργειας.

Η πρόβλεψη που θα πραγματοποιηθεί αφορά τη συνολική ενεργειακή ζήτηση στην Ελλάδα και γίνεται σε ημερήσια κλίμακα. Για την προσομοίωση, λαμβάνουμε υπόψη την ανάλυση του 3ου Κεφαλαίου, σχετικά με τα οικονομικά και κλιματικά δεδομένα και τις περιοδικότητες των χρονοσειρών. Η μεταβλητή που χρησιμοποιείται για πρόβλεψη είναι η ημερήσια ζήτηση ενέργειας ομαλοποιημένη ως προς το ΑΕΠ της εκάστοτε χρονιάς (εργαζόμαστε πλέον με τα ημερολογιακά έτη και όχι τα υδρολογικά). Επομένως, έχοντας δεδομένα για τη ζήτηση, τη θερμοκρασία και το ΑΕΠ για την περίοδο 2007-2016, δημιουργούμε στο R ένα data frame (χρονικό πλαίσιο) με όλες τις μεταβλητές που αναφέραμε παραπάνω σε κοινή χρονική βάση. Προκειμένου να γίνει η πρόβλεψη θα χρησιμοποιήσουμε τα 5 πρώτα χρόνια (2007-2011) ως δεδομένα βάσης (training data) και τα 5 επόμενα (2012-2016) ως άγνωστα δεδομένα τα οποία θα προσπαθήσουμε να προβλέψουμε (testing data). Ο διαχωρισμός αυτός, στο ήμισυ, γίνεται λόγω του μεγάλου όγκου δεδομένων που έχουμε στη διάθεσή μας και δικαιολογείται σύμφωνα με τους Afendras et al. (2015). Τα μοντέλα που επιλέγουμε είναι αρχικά η απλοϊκή μέθοδος (naïve) και έπειτα μοντέλα γραμμικής παλινδρόμησης και μοντέλα random forest. Οι μεταβλητές που θα επηρεάζουν τα μοντέλα αυτά είναι: η θερμοκρασία των προηγούμενων ημερών (έως και 7 ημέρες), η ημέρα της εβδομάδας, το είδος της ημέρας (αργία ή μη, καθημερινή ή μη) και ο μήνας.

Προκειμένου να αξιολογήσουμε την ποιότητα και την ακρίβεια της πρόβλεψης θα εξετάσουμε ορισμένα μεγέθη. Αρχικά, θα γίνει η οπτικοποίηση των προβλεπόμενων τιμών με τις πραγματικές τιμές των testing data σε ένα κοινό διάγραμμα με κοινή χρονική βάση. Στη συνέχεια, θα λάβουμε υπόψη ορισμένα στατιστικά μεγέθη τα οποία θα μας δώσουν μια καλύτερη εικόνα όσον αφορά την ακρίβειας της πρόβλεψης. Συγκεκριμένα, το πρώτο μέγεθος είναι η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) και δίνεται σύμφωνα με την εξ. 4.1:

$$\mathrm{RMSE} := \left(\left(\sum_{i=1}^{n} (f_i - x_i)^2 \right) / n \right)^{1/2} \tag{4.1}$$

Το RMSE δείχνει την απόλυτη προσαρμογή του μοντέλου στα δεδομένα, δηλαδή πόσο κοντά είναι τα παρατηρούμενα σημεία δεδομένων στις προβλεπόμενες τιμές. Το RMSE

μπορεί να ερμηνευθεί ως η τυπική απόκλιση της ανεξήγητης διακύμανσης και έχει τη χρήσιμη ιδιότητα να βρίσκεται στις ίδιες μονάδες με τη μεταβλητή απόκρισης. Οι χαμηλότερες τιμές του RMSE υποδεικνύουν καλύτερη εφαρμογή. Το RMSE είναι ένα καλό μέτρο για την ακρίβεια με την οποία το μοντέλο προβλέπει την απόκριση και είναι το πιο σημαντικό κριτήριο για την εφαρμογή κατά την πρόβλεψη μοντέλων.

Επόμενο μέγεθος που θα λάβουμε υπόψη είναι το μέσο απόλυτο σφάλμα (MAE). Το MAE είναι απλά, όπως υποδηλώνει το όνομα, ο μέσος όρος των απόλυτων λαθών. Το απόλυτο σφάλμα είναι η απόλυτη τιμή της διαφοράς μεταξύ της προβλεπόμενης τιμής και της πραγματικής τιμής. Το MAE μας δείχνει την απόκλιση κατά μέσο όρο της προβλεπόμενης τιμής από την πραγματική και δίνεται από την εξ. 4.2.

$$MAE = \frac{\sum_{i=1}^{n} |y_{pred,i} - y_i|}{n} \tag{4.2}$$

Ένα πρόβλημα που δημιουργείται με το MAE είναι ότι το σχετικό μέγεθος του σφάλματος δεν είναι πάντα προφανές. Μερικές φορές είναι δύσκολο να διακρίνουμε ένα μεγάλο λάθος από ένα μικρό λάθος. Για να αντιμετωπίσουμε αυτό το πρόβλημα, επιλέγουμε να υπολογίσουμε το μέσο απόλυτο ποσοστιαίο σφάλμα. Το μέσο απόλυτο ποσοστό σφάλματος (MAPE) μας επιτρέπει να συγκρίνουμε προβλέψεις διαφορετικών σειρών σε διαφορετικές κλίμακες και δίνεται από την εξ 4.3.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N} \frac{ABS(Actual_t - Forecast_t)}{Actual_t} * 100\% \tag{4.3}$$

Τέλος, θα υπολογίσουμε το συντελεστή προσδιορισμού (R^2). Το R^2 είναι ένα στατιστικό μέτρο που αντιπροσωπεύει το ποσοστό της διακύμανσης για μια εξαρτημένη μεταβλητή που οφείλεται σε μια ανεξάρτητη μεταβλητή ή μεταβλητές σε ένα μοντέλο παλινδρόμησης. Ενώ η συσχέτιση εξηγεί την ισχύ της σχέσης μεταξύ ανεξάρτητης και εξαρτημένης μεταβλητής, το R^2 εξηγεί σε ποιο βαθμό η διακύμανση μιας μεταβλητής εξηγεί τη διακύμανση της δεύτερης μεταβλητής. Έτσι, εάν το R^2 ενός μοντέλου ισούται με 0,50, τότε περίπου το ήμισυ της παρατηρούμενης παραλλαγής μπορεί να εξηγηθεί από τα δεδομένα εισόδου του μοντέλου, ενώ αν ισούται με 0 τότε το μοντέλο δεν εξηγεί καμία από τις μεταβλητές των δεδομένων απόκρισης γύρω από το μέσο όρο του. Δίνεται από την εξ. 4.4.

$$R^2 = 1 - \left(\sum_{i=1}^{n} (f_i - x_i)^2\right) \Big/ \sum_{i=1}^{n} \left(x_i - \left(\left(\sum_{i=1}^{n} x_i\right)/n\right)\right)^2 \tag{4.4}$$

Στη συνέχεια του κεφαλαίου παρουσιάζονται οι βασικότερες μέθοδοι στατιστικής πρόβλεψης, τις οποίες χρησιμοποιούμε για τη δημιουργία μοντέλων.

4.1.1. Απλοϊκή Μέθοδος

Η απλοϊκή μέθοδος (naïve) είναι η πιο απλή στατιστική μέθοδος. Οι προβλέψεις της στις περισσότερες περιπτώσεις δεν είναι πολύ ακριβείς, ωστόσο χρησιμοποιούνται ως σημείο αναφοράς (benchmark) συγκριτικά με άλλα πιο εξελιγμένα μοντέλα. Με βάση τη naïve μέθοδο η προβλεπόμενη τιμή του εξεταζόμενου μεγέθους είναι ίση με την παρατήρηση της ακριβώς προηγούμενης χρονικής περιόδου. Δηλαδή η πρόβλεψη περιγράφεται από μια σχέση της μορφής:

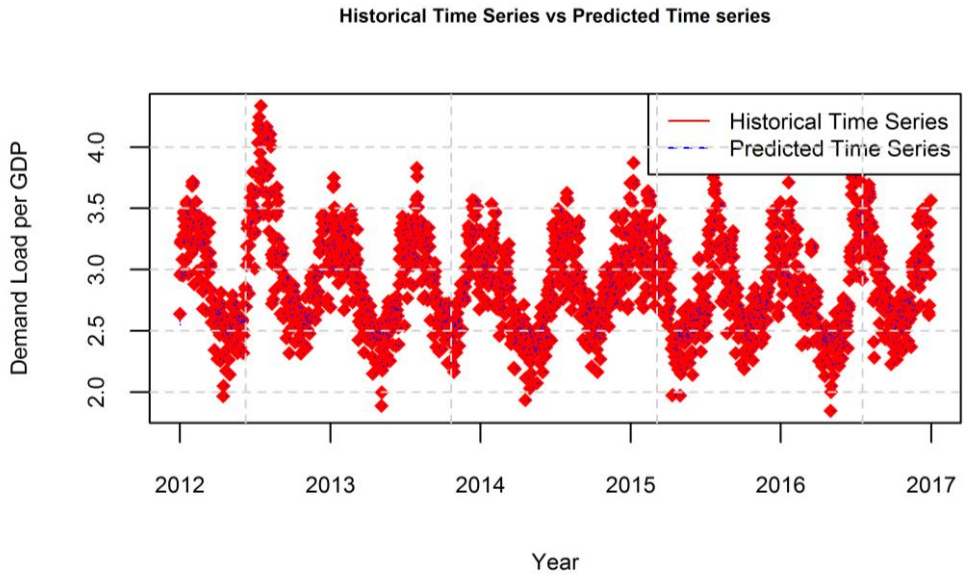
$$F_t = Y_{t-1} \tag{4.5}$$

όπου

F_t: η πρόβλεψη για την περίοδο t

Y_{t-1}: η παρατήρηση για την περίοδο $t-1$

Στη συγκεκριμένη περίπτωση, για την εφαρμογή του μοντέλου θα θεωρήσουμε πως το φορτίο ζήτησης κάθε ημέρας είναι ίσο με το φορτίο της προηγούμενης εβδομάδας. Σε περίπτωση αργίας όμως, το φορτίο ζήτησης εκείνης της ημέρας θα ισούται με τη μέγιστη τιμή των ακόλουθων: είτε το φορτίο ζήτησης της περασμένης Κυριακής είτε το φορτίο ζήτησης της ίδιας ημέρας πριν από μία εβδομάδα. Παρακάτω, στο Σχήμα 4.1 γίνεται απεικόνιση της πρόβλεψης με το μοντέλο naïve. Η χρονοσειρά πρόβλεψης (με μπλε) μοιάζει να ακολουθεί την ιστορική χρονοσειρά πρόβλεψης (με κόκκινο) γραφικά. Αυτό συμβαίνει διότι η χρονοσειρά πρόβλεψης είναι στην περίπτωση μας η ιστορική χρονοσειρά με καθυστέρηση επτά ημερών, εκτός ορισμένων περιπτώσεων που έχουμε αργία. Γίνεται εύκολα αντιληπτό ότι για να θεωρηθεί μια μέθοδος αποτελεσματική πρέπει να δίνει πιο ακριβή αποτελέσματα από την απλοϊκή μέθοδο. Η naïve μέθοδος είναι πολύ απλή στην εφαρμογή της και δίνει καλά αποτελέσματα σε σειρές που παρουσιάζουν σχετικά μικρή τάση ή εποχικότητα, κάτι το οποίο δε συμβαίνει στη δική μας περίπτωση.



Σχήμα 4.1. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (naïve) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

4.1.2 Μοντέλα Παλινδρόμησης

Η κύρια χρήση των μοντέλων παλινδρόμησης (Regression Models) αφορά στην εύρεση συσχετίσεων ανάμεσα σε μια εξαρτημένη μεταβλητή (εξεταζόμενο μέγεθος) και μια ή περισσότερες ανεξάρτητες μεταβλητές που την επηρεάζουν. Ανάλογα με τον αριθμό των ανεξάρτητων μεταβλητών προκύπτουν τα μοντέλα απλής ή πολλαπλής γραμμικής παλινδρόμησης. Σχετικά με τις προβλέψεις, αν θεωρήσουμε δεδομένες τις τιμές των ανεξάρτητων μεταβλητών, μπορούμε με τη βοήθεια των μοντέλων παλινδρόμησης να κατανοήσουμε την εξέλιξη της εξαρτημένης μεταβλητής μέσω της εξίσωσης παλινδρόμησης. Ένα σημαντικό πλεονέκτημα αυτών των μεθόδων έγκειται στη δυνατότητα να προσεγγίσουμε την αλλαγή της εξαρτημένης μεταβλητής για οποιαδήποτε αλλαγή μιας από τις ανεξάρτητες μεταβλητές, εφόσον οι υπόλοιπες παραμένουν σταθερές. Επιπλέον σημαντικό ρόλο παίζει η διακύμανση της εξαρτημένης μεταβλητής γύρω από την εξίσωση παλινδρόμησης, η οποία μπορεί να περιγραφεί μέσω πιθανοτικής κατανομής. Η διαδικασία πρόβλεψης με τη βοήθεια μοντέλων παλινδρόμησης προϋποθέτει τη γνώση ή πρόβλεψη των ανεξάρτητων μεταβλητών, στοιχείο που συχνά καθιστά τη χρήση των μοντέλων δύσκολη.

Η παλινδρόμηση είναι η μία από τις πιο ευρέως χρησιμοποιούμενες στατιστικές τεχνικές. Μοντέλα της χρησιμοποιούνται πολύ συχνά στην πρόβλεψη ηλεκτρικού φορτίου για τη διαμόρφωση της σχέσης μεταξύ της κατανάλωσης φορτίου και εξωτερικών παραγόντων όπως οι καιρικές συνθήκες, ο τύπος της ημέρας, η κατηγορία του πελάτη κ.α.. Εφαρμογές μοντέλων παλινδρόμησης για πρόβλεψη συναντάμε και από τους Feinberg et al. (2005) και Kyriakides et al. (2007). Παρατηρούμε ότι τα μοντέλα αυτά ενσωματώνουν επιρροές από εξωγενείς παράγοντες όπως ο καιρός, ντετερμινιστικές επιδράσεις όπως οι περίοδοι διακοπών (holidays) και στοχαστικές επιδράσεις όπως η μέση τιμή των φορτίων. Πρόσθετες εφαρμογές των μοντέλων παλινδρόμησης σε προβλήματα πρόβλεψης φορτίου παρουσιάζονται από τους Charytoniuk et al. (1998), Haida et al. (1994), Hyde et al. (1997) και Ruzic et al. (2003). Οι μέθοδοι παλινδρόμησης προσφέρουν αρκετά πλεονεκτήματα, όπως επισημαίνουν οι Hahn et al. (2009). Βοηθούν στην κατανόηση της σχέσης μεταξύ της εισόδου και των μεταβλητών εξόδου. Επιπλέον επιτρέπουν σχετικά εύκολα την αξιολόγηση της απόδοσης, κάτι το οποίο υποστηρίζεται από τους Bruhns et al. (2005). Ωστόσο σύμφωνα με τους Kyriakides et al. (2007), μπορεί να υπάρχουν εγγενή προβλήματα στον εντοπισμό του σωστού μοντέλου, γεγονός που αιτιολογείται από τη μη γραμμική σχέση μεταξύ του φορτίου και των παραγόντων που το επηρεάζουν.

Οι δυο βασικές κατηγορίες των μοντέλων παλινδρόμησης που θα αναλυθούν παρακάτω είναι η απλή και η πολλαπλή γραμμική παλινδρόμηση.

4.1.2.1 Μοντέλα Απλής Γραμμικής Παλινδρόμησης

Στην περίπτωση της απλής γραμμικής παλινδρόμησης (Linear Regression) εξετάζουμε τη σχέση του μεγέθους που θέλουμε να προβλέψουμε (εξαρτημένη μεταβλητή), δηλαδή τη ζήτηση ανά ΑΕΠ και μιας μόνο μεταβλητής που το επηρεάζει (ανεξάρτητη μεταβλητή). Υποθέτουμε ότι η σχέση των δύο μεταβλητών είναι γραμμική αν και στην πραγματικότητα αυτό δε συμβαίνει συχνά. Όσον αφορά την πρόβλεψη χρονοσειρών, συνήθως θεωρούμε το

χρόνο ως ανεξάρτητη μεταβλητή. Οι μαθηματικές σχέσεις που περιγράφουν αυτό το μοντέλο πρόβλεψης είναι οι ακόλουθες:

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

$$b = \frac{\sum_{i=1}^{n}[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (4.6)$$

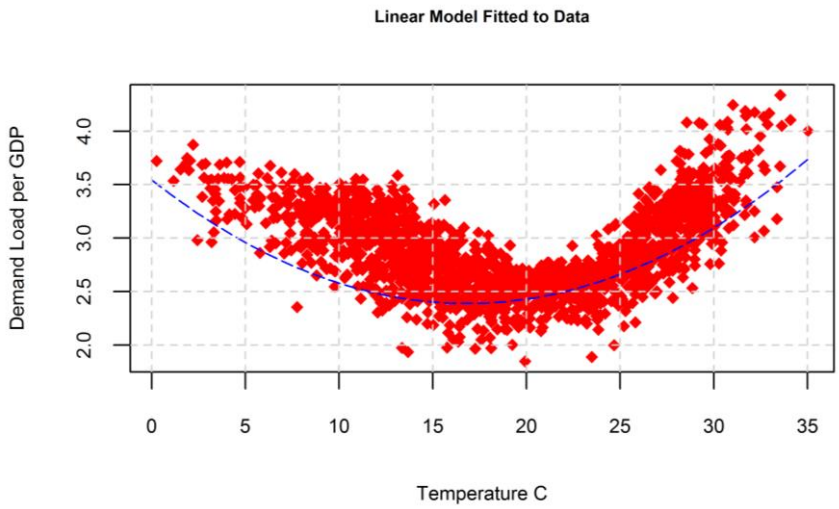
$$\alpha = \bar{Y} - b \cdot \bar{X}$$

όπου
X_i: η ανεξάρτητη μεταβλητή
Y_i: η εξαρτημένη μεταβλητή
$\hat{Y}_i$: η εκτιμώμενη τιμή της εξαρτημένης μεταβλητής
α: η τεταγμένη του σημείου τομής με τον άξονα των εξαρτημένων μεταβλητών
b: η κλίση της ευθείας
$\bar{X}$: η μέση τιμή των ανεξάρτητων μεταβλητών
$\bar{Y}$: η μέση τιμή των εξαρτημένων μεταβλητών
n: ο αριθμός των διαθέσιμων παρατηρήσεων

Η πρώτη εξίσωση περιγράφει την ευθεία γραμμή που συνδέει τις δύο μεταβλητές. Για τον υπολογισμό των συντελεστών α και b γίνεται χρήση της μεθόδου ελαχίστων τετραγώνων. Δηλαδή επιλέγονται εκείνες οι τιμές των συντελεστών που ελαχιστοποιούν το άθροισμα των τετραγώνων των διαφορών ανάμεσα στις πραγματικές και τις προβλεπόμενες τιμές. Το R μας επιτρέπει μέσω των συναρτήσεων lm και nls την εύρεση των παραπάνω συντελεστών με απλό τρόπο.

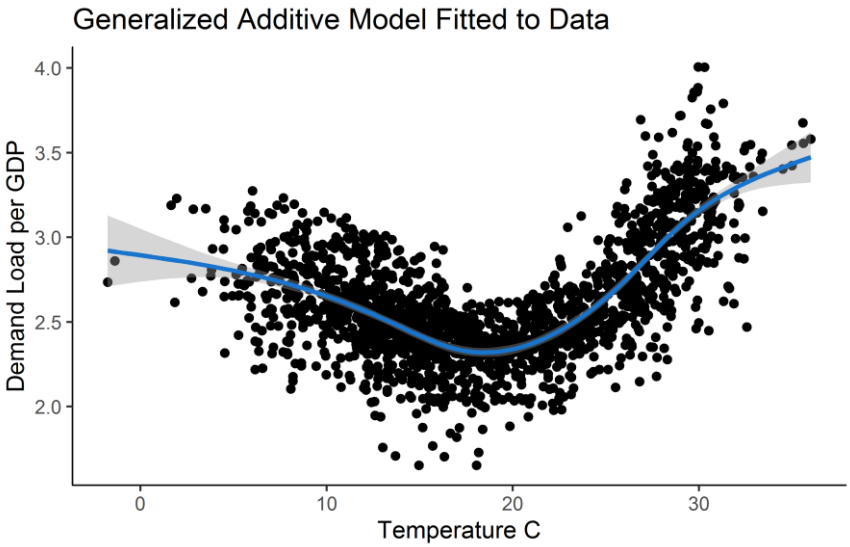
Τα αποτελέσματα που προκύπτουν από την παραπάνω μέθοδο συνήθως μας δίνουν μια καλή εικόνα για τη συμπεριφορά της χρονοσειράς. Όπως αναφέραμε παραπάνω, η σχέση μεταξύ εξαρτημένης και ανεξάρτητης μεταβλητής τις περισσότερες φορές δεν είναι γραμμική. Στη συγκεκριμένη περίπτωση, θα προσπαθήσουμε να βρούμε τη σχέση μεταξύ ζήτησης και θερμοκρασίας. Στο Σχήμα 4.2 παρατηρούμε πως η σχέση των δύο αυτών μεταβλητών είναι κυρτή και θα προσεγγιστεί επομένως από μια εξίσωση της μορφής:

$Y = b_0 + b_1 \cdot X + b_2 \cdot X^2$, όπου το Y αναφέρεται στην ημερήσια ζήτηση ανά ΑΕΠ και το X στην ημερήσια θερμοκρασία.



Σχήμα 4.2. Διάγραμμα μέσης ημερήσιας θερμοκρασίας και μέσης ημερήσιας ζήτησης ενέργειας ανά ΑΕΠ

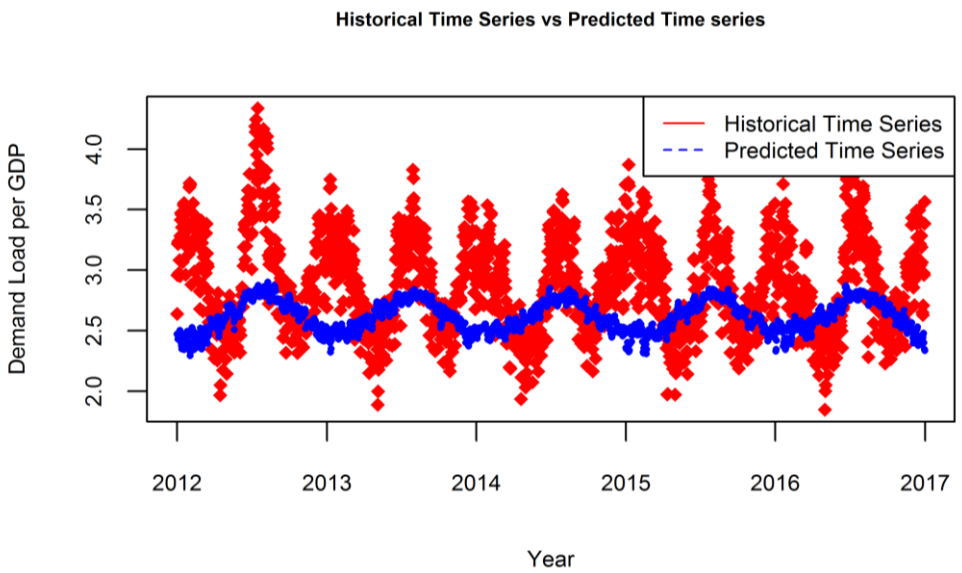
Ύστερα από διερεύνηση προκύπτει ότι η εξίσωση που αντιπροσωπεύει καλύτερα τα δεδομένα μας είναι η: $Y = 3.5441172 - 0.1373002 \cdot X + 0.0040789 \cdot X^2$ και βρίσκεται μέσω της μεθόδου ελάχιστων τετραγώνων. Η παραπάνω εξίσωση αναπαρίσταται στο Σχήμα 4.2 (κυρτή μπλε γραμμή). Οι Mirasgedis et al. (2006) που ανέλυσαν τη σχέση θερμοκρασίας-κανονικοποιημένης ζήτησης (προκύπτει διαιρώντας την πραγματική παρατηρούμενη ζήτηση ηλεκτρικής ενέργειας με τη ζήτηση που αποδίδεται ονομαστικά στην οικονομική ανάπτυξη), κατέληξαν σε εξίσωση κυβικής παλινδρόμησης της μορφής: $Y = \mathrm{a} + \mathrm{b} \cdot X + \mathrm{c} \cdot X^2 + \mathrm{d} \cdot X^3$. Στο Σχήμα 4.3 βρίσκουμε την καμπύλη η οποία αποτελεί τη βέλτιστη για να περιγράψει τα δεδομένα ζήτησης-θερμοκρασίας με ένα γενικευμένο πρόσθετο μοντέλο (Generalized Additive Model). Πρόκειται για ένα μοντέλο, στο οποίο ο γραμμικός προγνωστικός παράγοντας εξαρτάται γραμμικά από άγνωστες ομαλές λειτουργίες ορισμένων μεταβλητών προγνωστικού και το ενδιαφέρον επικεντρώνεται στην εξαγωγή αυτών των ομαλών λειτουργιών. Το μοντέλο συσχετίζει μια μεταβλητή απόκρισης Y, σε ορισμένες μεταβλητές προβλέψεων, X_i.



Σχήμα 4.3. Διάγραμμα μοντέλου GAM προσαρμοσμένο στη μέση ημερήσια θερμοκρασία με τη μέση ημερήσια ζήτηση ενέργειας ανά ΑΕΠ

Στη συνέχεια, με τη χρήση του συντελεστή γραμμικής συσχέτισης των δύο παραπάνω μεταβλητών μπορούμε να αποφανθούμε σχετικά με το βαθμό εξάρτησης αυτών. Ο συντελεστής μπορεί να λάβει τιμές εντός του διαστήματος $[-1,1]$. Πιο συγκεκριμένα για τιμή ίση με το μηδέν προκύπτει ότι οι δυο μεταβλητές είναι ανεξάρτητες μεταξύ τους, ενώ για τιμές ± 1 υπάρχει απόλυτη συσχέτιση. Γενικά για θετικές τιμές του συντελεστή συσχέτισης οι αυξομειώσεις των μεταβλητών έχουν την ίδια κατεύθυνση. Το αντίθετο συμβαίνει για αρνητικές τιμές του συντελεστή. Στη συγκεκριμένη περίπτωση προκύπτει πως ο συντελεστής συσχέτισης ζήτησης θερμοκρασίας ισούται με 0.36, κάτι το οποίο υποδεικνύει μια ήπια συσχέτιση μεταξύ των δύο αυτών μεταβλητών.

Το μοντέλο που θα εξετασθεί για την περίπτωση απλής γραμμικής παλινδρόμησης θα έχει ως εξαρτημένη μεταβλητή την ημερήσια ζήτηση ανά ΑΕΠ και ως ανεξάρτητη μεταβλητή την ημερήσια θερμοκρασία της προηγούμενης ημέρας. Έτσι, με τη βοήθεια της συνάρτησης lm του R λαμβάνουμε δεδομένα εισόδου από τα δεδομένα βάσης (01-01-2007 έως 31-12-2011) και προβλέπουμε τα επόμενα 5 χρόνια. Στο Σχήμα 4.4 φαίνονται οι δύο χρονοσειρές, ιστορική και συνθετική, σε κοινό διάγραμμα που προκύπτουν από την πρόβλεψη. Το μοντέλο αυτό αποτελεί το πρώτο γραμμικό μοντέλο και το ονομάζουμε lm_1.



Σχήμα 4.4. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_1) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

4.1.2.2 Μοντέλα Πολλαπλής Γραμμικής Παλινδρόμησης

Η μέθοδος της πολλαπλής γραμμικής παλινδρόμησης αποτελεί γενίκευση της παραπάνω μεθόδου και περιλαμβάνει περισσότερες από μία ανεξάρτητες μεταβλητές. Σκοπός είναι η πρόβλεψη μιας εξαρτημένης μεταβλητής λαμβάνοντας υπόψη όλες τις μεταβλητές που μπορεί να το επηρεάζουν. Ο όρος γραμμικότητα αναφέρεται στις παραμέτρους και όχι στις ανεξάρτητες μεταβλητές. Έτσι, υπάρχουν και κάποια μοντέλα που είναι γραμμικά ως προς τις παραμέτρους όχι όμως ως προς τις μεταβλητές X. Η μαθηματική σχέση που περιγράφει το μοντέλο πολλαπλής γραμμικής παλινδρόμησης είναι η ακόλουθη:

$$Y = b0 + b1 \cdot X1,i + b2 \cdot X2,i + \cdots + bk \cdot Xk,i + ei \qquad (4.7)$$

όπου
Y: η εξαρτημένη μεταβλητή
X: οι ανεξάρτητες μεταβλητές
bi: συντελεστές (σταθερές παράμετροι)
ei: απόκλιση από τις πραγματικές τιμές

Απαραίτητη προϋπόθεση για την παραγωγή του μοντέλου πολλαπλής παλινδρόμησης είναι η απουσία συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών. Υποθέτουμε ότι η απόκλιση της εκτίμησης από την πραγματική τιμή είναι τυχαία. Η συνάρτηση που περιγράψαμε είναι γραμμική ως προς τους συντελεστές. Στόχος είναι ο υπολογισμός των σταθερών παραμέτρων bi, ο οποίος επιτυγχάνεται με τη μέθοδο των ελαχίστων τετραγώνων. Η διαδικασία εύρεσης της μορφής για την παραπάνω συνάρτηση δεν είναι εύκολη, έτσι ασχολούμαστε αρχικά με τον υπολογισμό των συντελεστών bi και του τυχαίου παράγοντα ei. Πιο αναλυτικά για κάθε συντελεστή υπολογίζονται οι μερικές παράγωγοι του αθροίσματος του τετραγώνου των σφαλμάτων ei, στη συνέχεια τίθενται ίσες με το μηδέν και προκύπτει ένα σύστημα εξισώσεων. Η επίλυση αυτού του γραμμικού συστήματος οδηγεί στην εύρεση των ζητούμενων συντελεστών.

Στη συνέχεια, θα εξετασθούν αρκετά μοντέλα γραμμικής παλινδρόμησης με σκοπό την εύρεση αυτού που θα ταιριάζει απόλυτα στα δεδομένα μας. Θα εξετασθούν οι εξής μεταβλητές σε διάφορους συνδυασμούς:

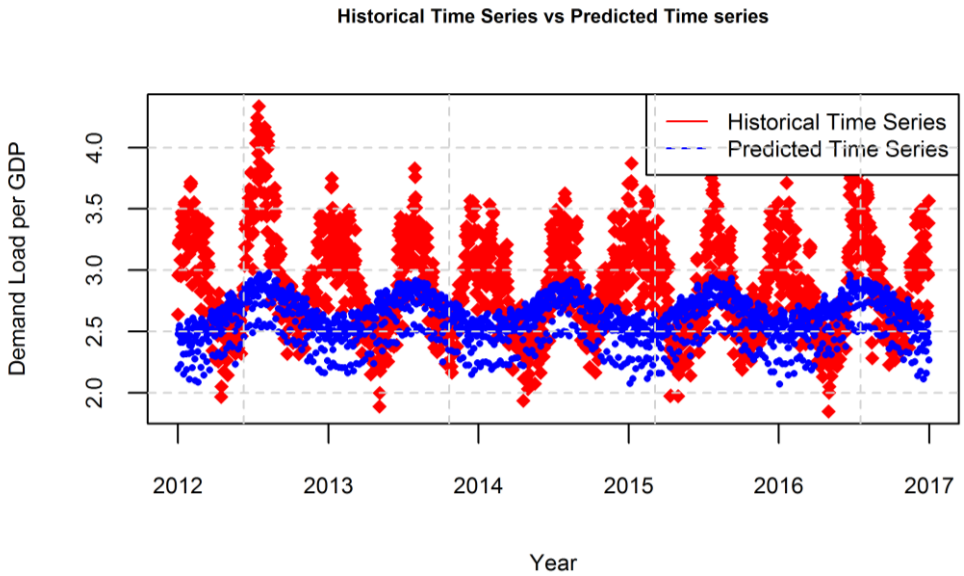
- Θερμοκρασία προηγούμενων ημερών (έως και 7 ημέρες πριν). Χάριν συντομίας θα ονομάσουμε τη μεταβλητή αυτή temp_i, όπου το i λαμβάνει τιμές από το 1 έως το 7, με temp_1 να είναι η θερμοκρασία μίας ημέρας πριν. Η μεταβλητή αυτή έχει αριθμητική αξία.
- Είδος ημέρας (weekday). Η μεταβλητή αυτή ορίζεται ως παράγοντας (factor) στα δεδομένα εισόδου.
- Μήνας (month). Η μεταβλητή αυτή ορίζεται ως παράγοντας (factor) στα δεδομένα εισόδου.

Η μορφή της χρονοσειράς από την οποία λαμβάνουμε τις μεταβλητές μας έχει μήκος 1826 παρατηρήσεις (5 χρόνια) και έχει την εξής μορφή (ενδεικτικά φαίνονται οι πρώτες 6 σειρές):

year	date	month	day	weekday	season	Holiday	Load	temperature	GDP	load_real_100_per_GDP
2012	2012-01-01	1	1	Sunday	Dec_Feb	1	5023.333	10.37	190395	2.638375
2012	2012-01-02	1	2	Monday	Dec_Feb	0	5635.250	8.05	190395	2.959768
2012	2012-01-03	1	3	Tuesday	Dec_Feb	0	6124.333	8.59	190395	3.216646
2012	2012-01-04	1	4	Wednesday	Dec_Feb	0	6150.083	9.86	190395	3.230171
2012	2012-01-05	1	5	Thursday	Dec_Feb	0	6170.833	10.54	190395	3.241069
2012	2012-01-06	1	6	Friday	Dec_Feb	1	5667.583	10.93	190395	2.976750

Ονομάσαμε το πρώτο μοντέλο γραμμικής παλινδρόμησης lm_1. Το αμέσως επόμενο μοντέλο που θα εξετάσουμε θα εξαρτάται από τη θερμοκρασία της προηγούμενης ημέρας

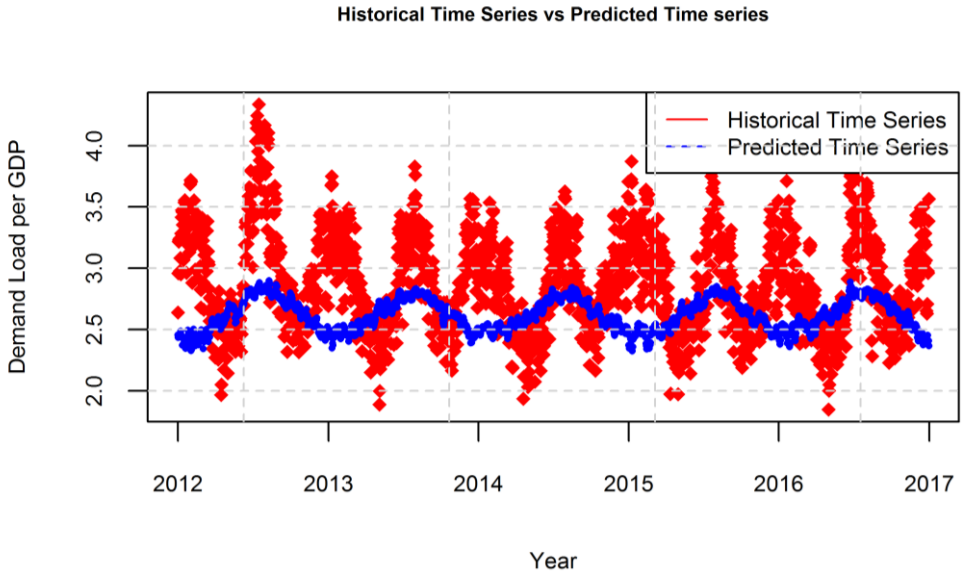
και το είδος της ημέρας. Επομένως το lm_2, θα είναι της μορφής: f(temp_1,weekday). Παρακάτω, στο Σχήμα 4.5, παρουσιάζεται η απεικόνιση της ιστορικής χρονοσειράς με τη συνθετική και υπολογίζονται οι στατιστικές παράμετροι που αναφέρουμε παραπάνω (RMSE, MAE, MAPE, R^2), οι οποίες θα σχολιασθούν στη συνέχεια.



Σχήμα 4.5. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_2) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

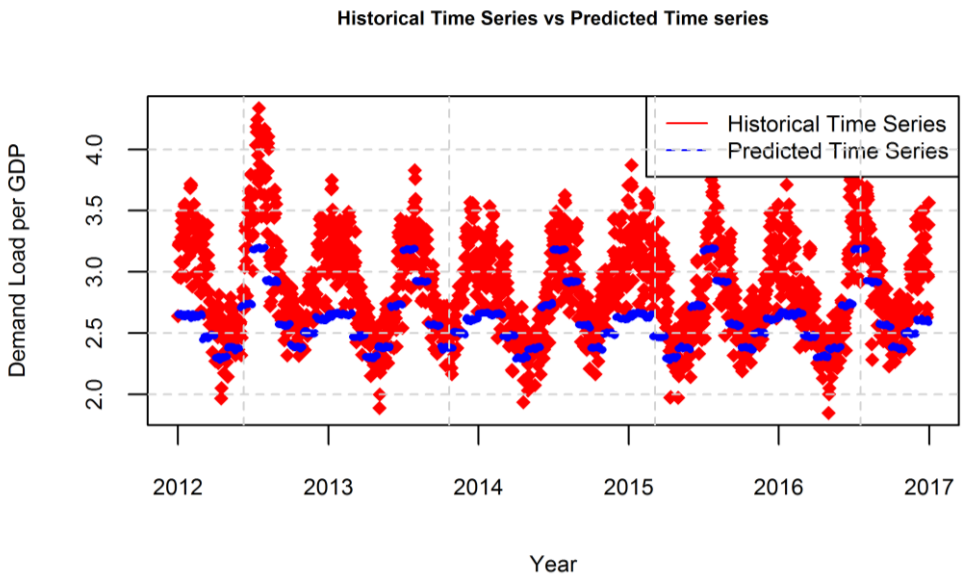
Παρακάτω θα δημιουργήσουμε τα επόμενα μοντέλα και θα τα απεικονίσουμε με τον παραπάνω τρόπο. Συγκεκριμένα τα μοντέλα αυτά είναι:

- Linear Model 3 (lm_3): f(temp_1, temp_2, temp_3), εδώ το μοντέλο επηρεάζεται από τις θερμοκρασίες των τριών προηγούμενων ημερών (Σχήμα 4.6).



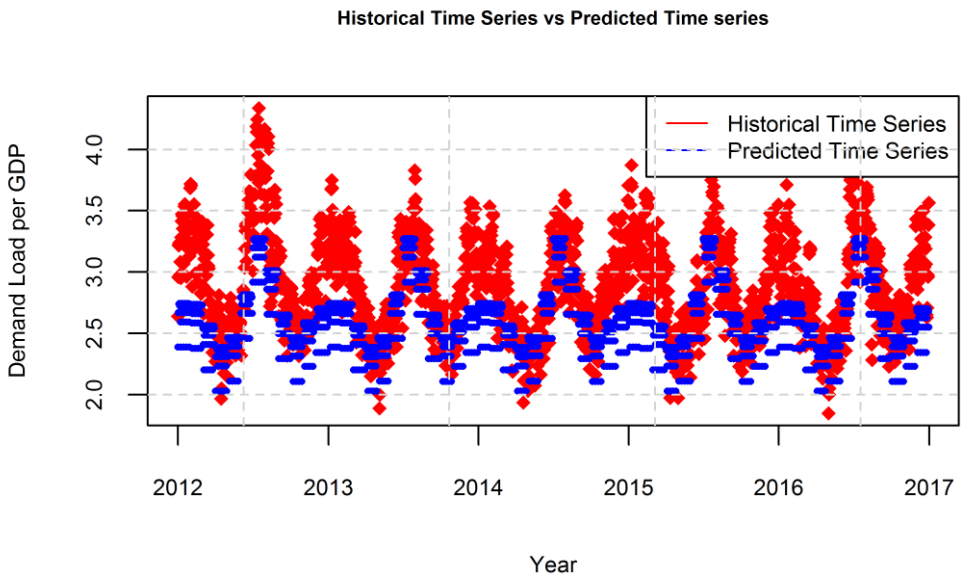
Σχήμα 4.6. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_3) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 4 (lm_4): f(temp_1, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας και το μήνα (Σχήμα 4.7).



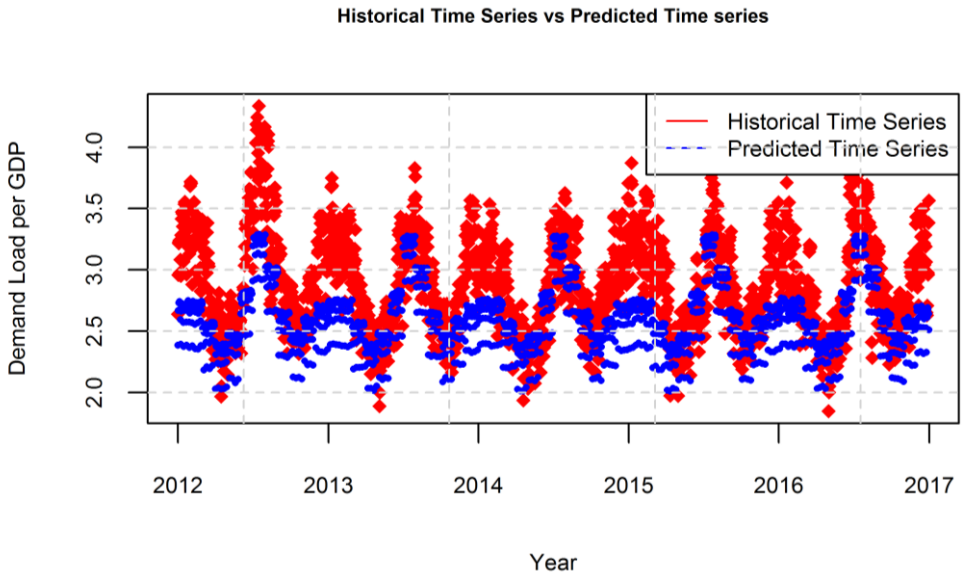
Σχήμα 4.7. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_4) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 5 (lm_5): f(weekday, month), εδώ το μοντέλο επηρεάζεται από την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.8).



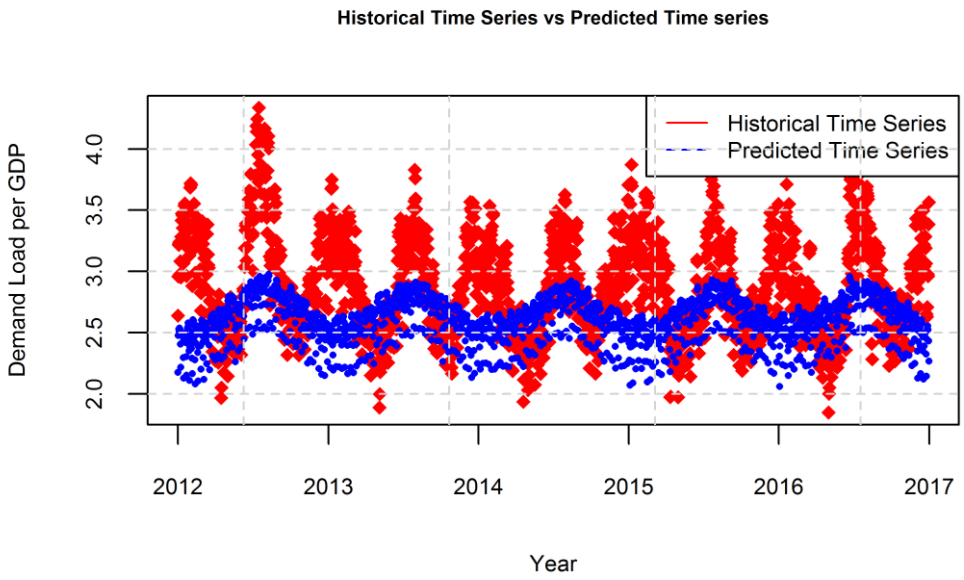
Σχήμα 4.8. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_5) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 6 (lm_6): f(temp_1, weekday, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας, την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.9).



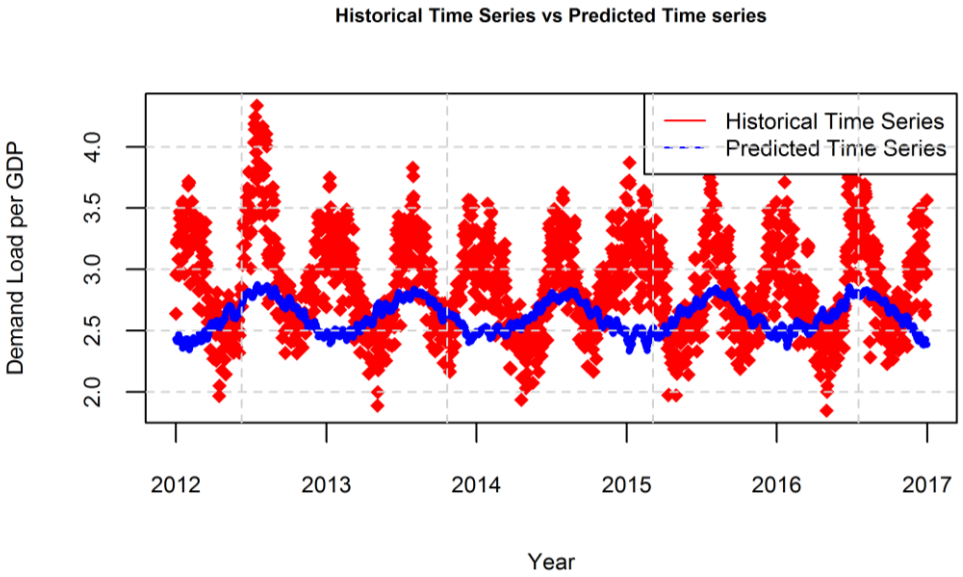
Σχήμα 4.9. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_6) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 7 (lm_7): f(temp_1, temp_2, weekday), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των δύο προηγούμενων ημερών και την ημέρα της εβδομάδος (Σχήμα 4.10).



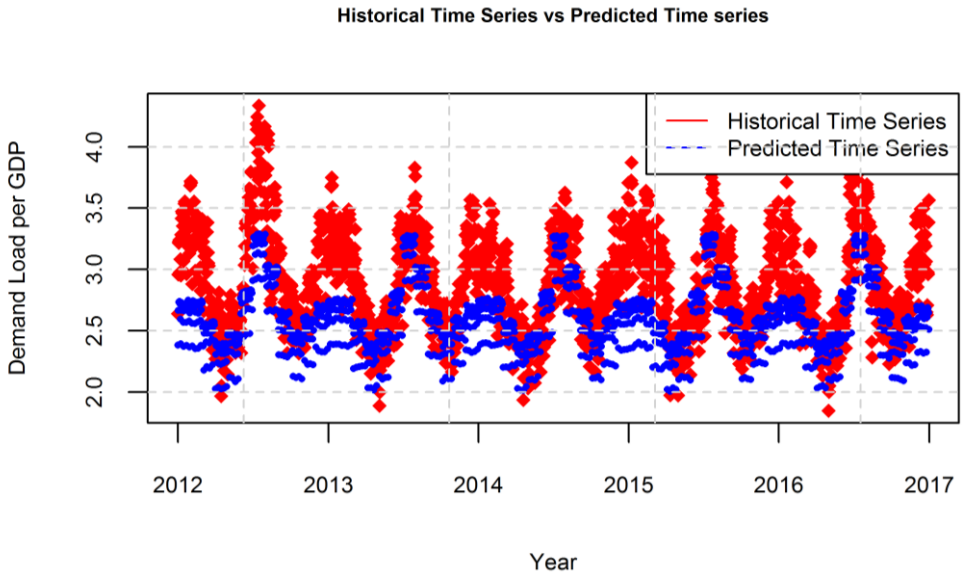
Σχήμα 4.10. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_7) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 8 (lm_8): f(temp_1, temp_2, temp_3, temp_4, temp_5, temp_6, temp_7), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία και των επτά προηγούμενων ημερών (Σχήμα 4.11).



Σχήμα 4.11. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_8) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 9 (lm_9): f(temp_1, temp_2, temp_3, weekday, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των τριών προηγούμενων ημερών, την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.12).



Σχήμα 4.12. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_9) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

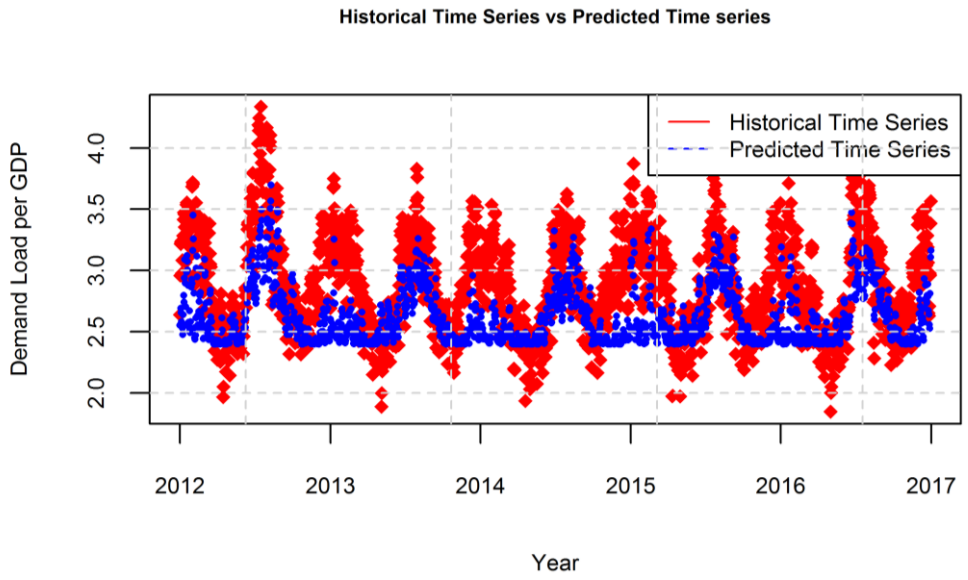
Παρατηρούμε ότι μέχρι στιγμής τα μοντέλα μας, τουλάχιστον οπτικά, δεν προσαρμόζονται αρκετά καλά στα πραγματικά μας δεδομένα. Γι' αυτό το λόγο θα αυξήσουμε την πολυπλοκότητα του μοντέλου. Έτσι, θα εισάγουμε στη συνέχεια την έννοια της πολυωνυμικής γραμμικής παλινδρόμησης. Στη στατιστική, η πολυωνυμική παλινδρόμηση είναι μια μορφή ανάλυσης παλινδρόμησης στην οποία η σχέση μεταξύ της ανεξάρτητης μεταβλητής x και της εξαρτώμενης μεταβλητής y διαμορφώνεται ως πολυώνυμο ν-οστου βαθμού στο x. Η πολυωνυμική παλινδρόμηση ταιριάζει σε μια μη γραμμική σχέση μεταξύ της τιμής του x και του αντίστοιχου υπό όρους μέσου όρου του y, που υποδηλώνεται E (y | x). Αν και η πολυωνυμική παλινδρόμηση ταιριάζει με ένα μη γραμμικό μοντέλο στα δεδομένα, ως πρόβλημα στατιστικής εκτίμησης είναι γραμμικό, με την έννοια ότι η συνάρτηση παλινδρόμησης E (y | x) είναι γραμμική στις άγνωστες παραμέτρους που εκτιμώνται από τα δεδομένα. Για το λόγο αυτό, η πολυωνυμική παλινδρόμηση θεωρείται ειδική περίπτωση πολλαπλής γραμμικής παλινδρόμησης. Η εξίσωση 4.7 πλέον μετασχηματίζεται στην παρακάτω εξίσωση:

$$Y = b0 + b1 \cdot X1,i + b2 \cdot X^2,i + \cdots + bk \cdot Xk,i + ei \quad (4.8)$$

όπου
Y: η εξαρτημένη μεταβλητή
X: οι ανεξάρτητες μεταβλητές
bi: συντελεστές (σταθερές παράμετροι)
ei: απόκλιση από τις πραγματικές τιμές

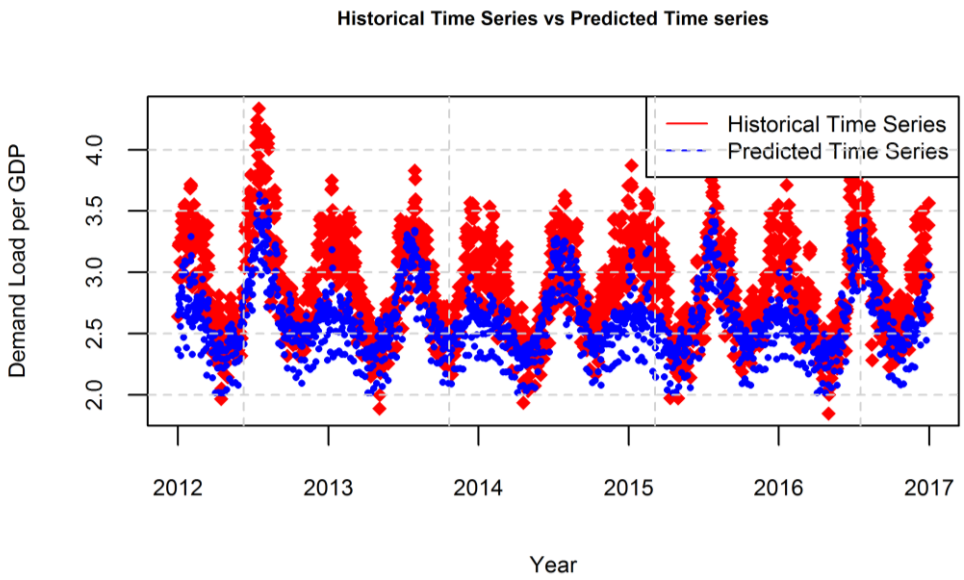
Συνεχίζουμε την παράθεση μοντέλων ορίζοντας πλέον τη σχέση θερμοκρασίας-ζήτησης ενέργειας σε κυρτή (όπως δείξαμε και στην ανάλυση σχέσης μεταξύ τους στο Κεφάλαιο 3).

- Linear Model 10 (lm_10): f((temp_1)^2, temp_1), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας αποκλειστικά (Σχήμα 4.13).



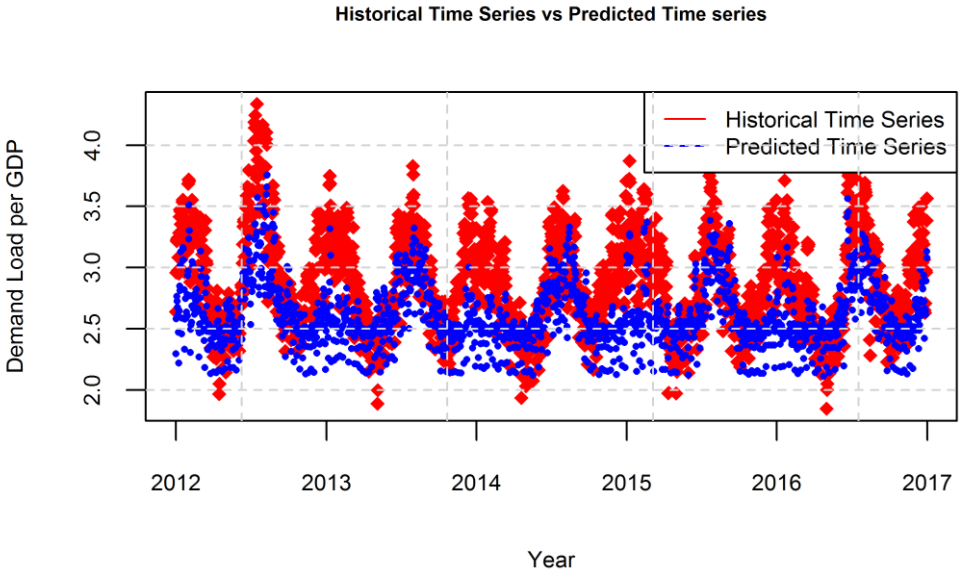
Σχήμα 4.13. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_10) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 11 (lm_11): f((temp_1)^2, temp_1, weekday, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας, την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.14).



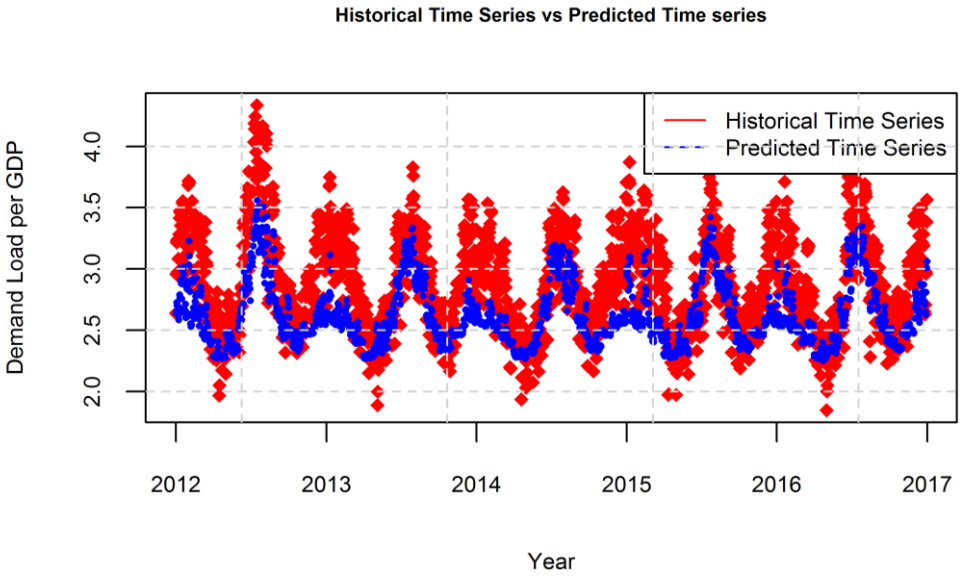
Σχήμα 4.14. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_11) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 12 (lm_12): f((temp_1)^2, temp_1, temp_2, temp_3, weekday), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των τριών προηγούμενων ημερών και την ημέρα της εβδομάδος (Σχήμα 4.15).



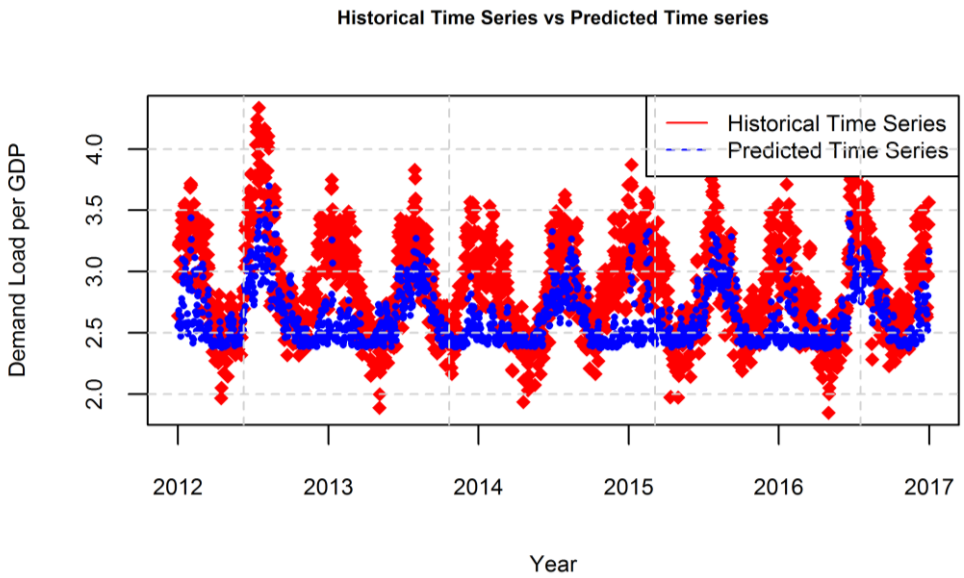
Σχήμα 4.15. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_12) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 13 (lm_13): f((temp_1)^2, temp_1, temp_2, temp_3, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των τριών προηγούμενων ημερών και το μήνα (Σχήμα 4.16).



Σχήμα 4.16. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_13) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Linear Model 14 (lm_14): f((temp_1)^2, temp_1, temp_2, temp_3, temp_4, temp_5, temp_6, temp_7), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των επτά προηγούμενων ημερών (Σχήμα 4.17).



Σχήμα 4.17. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_14) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

Σε αυτό το σημείο και αφού έχουμε απεικονίσει όλα τα γραμμικά μας μοντέλα θα αναλύσουμε την ακρίβεια πρόβλεψης των μοντέλων αυτών. Πέραν από τη γραφική συμπεριφορά του κάθε μοντέλου, είναι απαραίτητο να ληφθούν υπόψη τα στατιστικά μεγέθη κάθε μοντέλου τα οποία αναγράφονται στον Πίνακα 4.1.

Πίνακας 4.1. Στατιστικά μεγέθη για τα γραμμικά μοντέλα

	RMSE	MAE	MAPE	R^2
Naive Model	0.26579313	0.20238420	0.06943970	0.55359382
Linear Model 1	0.49435417	0.37937801	0.12312933	-0.54425531
Linear Model 2	0.48046937	0.36910258	0.11956048	-0.45872733
Linear Model 3	0.49414421	0.37986518	0.12334812	-0.54294387
Linear Model 4	0.40939687	0.32444304	0.10611104	-0.54425531
Linear Model 5	0.39201498	0.30826143	0.10086914	0.02893563
Linear Model 6	0.39212527	0.30593001	0.09994208	0.02838914
Linear Model 7	0.48024246	0.36905142	0.11955304	-0.45734982
Linear Model 8	0.49403065	0.37971640	0.12331308	-0.54223479
Linear Model 9	0.39205849	0.30581943	0.09989810	0.02872003
Linear Model 10	0.41126149	0.32570461	0.10673611	-0.06875655
Linear Model 11	0.36371080	0.29080511	0.09574744	0.16409851
Linear Model 12	0.39408107	0.30695930	0.10028467	0.01867277
Linear Model 13	0.38215110	0.31141193	0.10267076	0.07718865
Linear Model 14	0.41118408	0.32568585	0.10674320	-0.06835426

Προκειμένου να αξιολογήσουμε την ακρίβεια κάθε μοντέλου, είναι απαραίτητο να λάβουμε συνδυαστικά τις παραπάνω πληροφορίες. Είναι προφανές ότι, για τα μεγέθη RMSE, MAE, MAPE ιδανικά θέλουμε η τιμής τους να ισούται με το 0, ενώ για το R^2 η τιμή 1 υποδεικνύει την απόλυτη προσαρμογή της πρόβλεψης στα δεδομένα μας. Τα δεδομένα ζήτησης ανά ΑΕΠ κυμαίνονται από 1.6 έως 4.4. Είναι απαραίτητο να αναφερθεί το εύρος τιμών της μεταβλητής μας προκειμένου να αξιολογηθούν τα μεγέθη RMSE και MAE τα οποία δεν είναι ποσοστιαία. Σε γενικές γραμμές παρατηρούμε πως κανένα από τα παραπάνω μοντέλα δεν προσαρμόζεται πολύ καλά στα ιστορικά δεδομένα για τα έτη 2012-2016 (testing data). Ειδικά τα μοντέλα στα οποία δεν έχει γίνει πολυωνυμική παλινδρόμηση μοιάζουν άστοχα. Συγκεκριμένα, τα μοντέλα lm_3 και lm_8 (με κόκκινο στον πίνακα) στα οποία έχει χρησιμοποιηθεί ως μοναδική μεταβλητή η θερμοκρασία των προηγούμενων ημερών και μάλιστα έχει γραμμική σχέση με τη ζήτηση ενέργειας, παρουσιάζουν μεγάλες αποκλίσεις τόσο γραφικά όσο και στατιστικά. Τα δύο αυτά μοντέλα εμφανίζουν τα μεγαλύτερα σφάλματα (RMSE, MAE, MAPE) και σίγουρα δεν αποτελούν καλή πρόβλεψη. Αντίθετα, τα μοντέλα lm_11 και lm_13 (με πράσινο στον πίνακα) προσέγγισαν πολύ καλύτερα την ιστορική χρονοσειρά και οι στατιστικές τιμές τους είναι αποδεκτές δεδομένου του ορίζοντα πρόβλεψης, της πολύπλοκης συμπεριφοράς και του όγκου των δεδομένων.

4.2 Μοντέλα Τεχνητής Νοημοσύνης

Με τον όρο τεχνητή νοημοσύνη (Artificial Intelligence) ορίζουμε γενικά μια αναλυτική διαδικασία η οποία βασίζεται σε υπολογιστικές μηχανές και εμφανίζει συμπεριφορά και ενέργειες που θεωρούνται "έξυπνες" από ανθρώπινους παρατηρητές. Προσπαθεί δηλαδή να μιμηθεί την ανθρώπινη πορεία σκέψης, συμπεριλαμβανομένων και στοιχείων λογικής και βελτιστοποίησης. Βασικός στόχος των μεθόδων τεχνητής νοημοσύνης είναι να βοηθήσουν στην οργάνωση και την παροχή πληροφοριών που απαιτούνται κατά τη διαδικασία λήψης αποφάσεων με τρόπο που να βελτιώνουν τη συνολική απόδοση. Στη συνέχεια, θα ασχοληθούμε με μοντέλα random forest τα οποία θα χρησιμοποιηθούν για βραχυπρόθεσμη πρόβλεψη.

4.2.1 Μοντέλα Random Forest

Το Random Forest είναι ένας πολύ ισχυρός αλγόριθμος μηχανικής εκμάθησης (Machine Learning) που λειτουργεί με τη δημιουργία πολλαπλών δέντρων αποφάσεων (decision trees) και στη συνέχεια με τον συνδυασμό της παραγωγής που παράγεται από κάθε ένα από τα δέντρα αποφάσεων. Το decision tree είναι ένα μοντέλο ταξινόμησης που λειτουργεί με την έννοια της απόκτησης πληροφοριών σε κάθε κόμβο. Για όλα τα σημεία δεδομένων, το decision tree θα προσπαθήσει να ταξινομήσει τα σημεία δεδομένων σε κάθε έναν από τους κόμβους και να ελέγξει την απόκτηση πληροφοριών σε κάθε κόμβο. Στη συνέχεια, θα ταξινομηθεί στον κόμβο όπου λαμβάνει τις μέγιστες πληροφορίες. Θα ακολουθήσει αυτήν τη διαδικασία έως ότου εξαντληθούν όλοι οι κόμβοι ή δεν υπάρξει περαιτέρω λήψη πληροφοριών. Γενικά, τα decision trees είναι πολύ απλά και κατανοητά μοντέλα.

Το Random Forest συνδυάζει την παραγωγή πολλαπλών decision trees και στη συνέχεια επιτυγχάνει τη δική του έξοδο. Το Random Forest λειτουργεί με την ίδια αρχή με το decision tree. Ωστόσο, δεν επιλέγει όλα τα σημεία δεδομένων και τις μεταβλητές σε κάθε δέντρο. Δείχνει τυχαία σημεία δεδομένων και μεταβλητές σε κάθε δέντρο που δημιουργεί και στη συνέχεια συνδυάζει την πληροφορία στο τέλος. Καταργεί την προκατάληψη που μπορεί να εισαγάγει ένα decision tree στο σύστημα και βελτιώνει σημαντικά την προγνωστική ισχύ σε σχέση με αυτή ενός μεμονωμένου decision tree.

Κατασκευάζουμε ένα πλήθος δέντρων αποφάσεων στο χρόνο εκπαίδευσης και εξάγουμε την τάξη που είναι η μέση πρόβλεψη λειτουργίας (παλινδρόμηση) χρησιμοποιώντας το πακέτο του R που ονομάζεται randomForest.

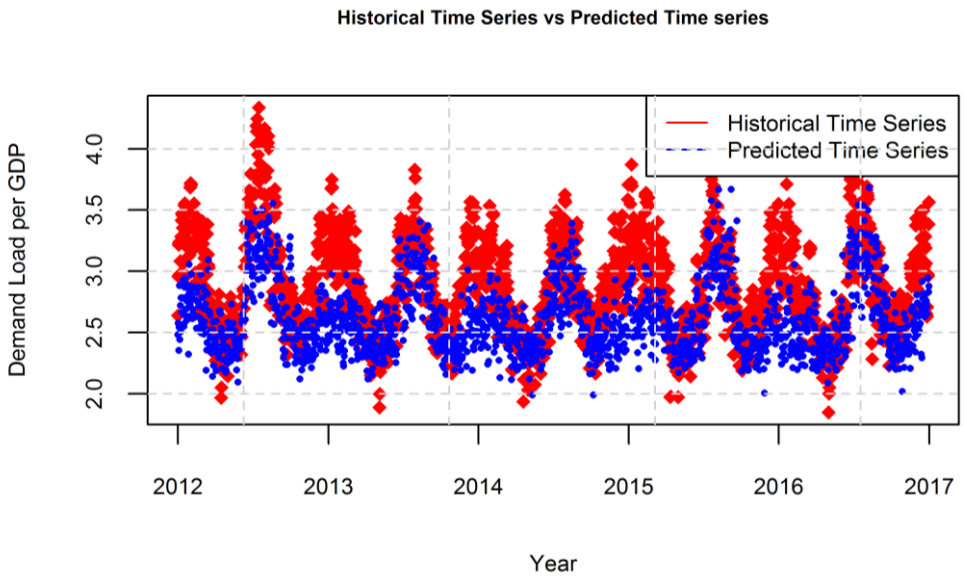
Η σύνταξη για το Random Forest είναι:
randomForest(formula, ntree=n, mtry=FALSE, maxnodes = NULL)
Μεταβλητές:

- formula: Τύπος του προσαρμοσμένου μοντέλου.
- ntree: Αριθμός δέντρων στο δάσος.
- mtry: Ο αριθμός των υποψηφίων που αντλούν για τροφοδοσία του αλγορίθμου. Από προεπιλογή, ισούται με το τετράγωνο του αριθμού των στηλών.
- maxnodes: Η μέγιστη ποσότητα τερματικών κόμβων στο δάσος.

- importance=TRUE: Εάν πρέπει να αξιολογηθεί η σημασία των ανεξάρτητων μεταβλητών στο τυχαίο δάσος.

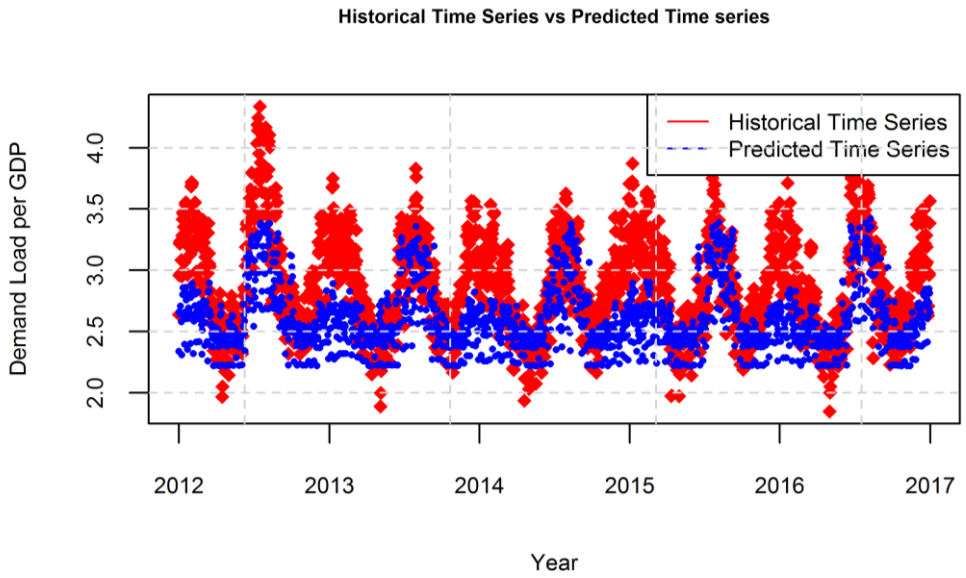
Θα εξετάσουμε διάφορα μοντέλα με τον ίδιο τρόπο όπως και με τα γραμμικά μοντέλα. Από προεπιλογή, ο αριθμός των δέντρων αποφάσεων στο δάσος είναι 500 και ο αριθμός των χαρακτηριστικών που χρησιμοποιούνται ως δυνητικοί υποψήφιοι για κάθε διαίρεση είναι 3. Το μοντέλο θα επιχειρήσει αυτόματα να ταξινομήσει καθένα από τα δείγματα στο σύνολο δεδομένων και να εμφανίσει έναν πίνακα με τα αποτελέσματα. Στη συνέχεια παρατίθενται τα μοντέλα random forest:

- Random forest Model 1 (rf_1): f(temp_1), εδώ το μοντέλο επηρεάζεται μόνο από τη θερμοκρασία της προηγούμενης ημέρας (Σχήμα 4.18).



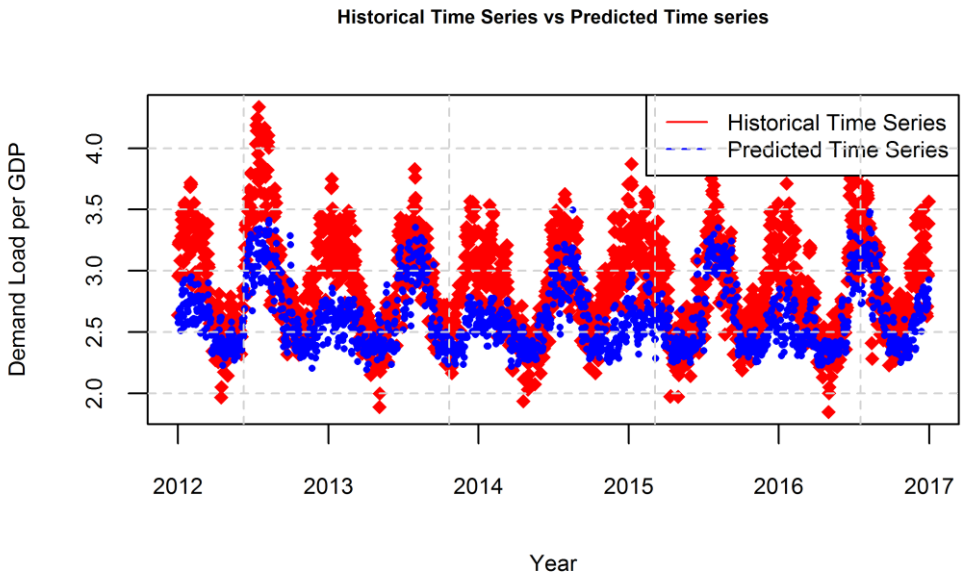
Σχήμα 4.18. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_1) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 2 (rf_2): f(temp_1, weekday), εδώ το μοντέλο επηρεάζεται από τις θερμοκρασίες των τριών προηγούμενων ημερών (Σχήμα 4.19).



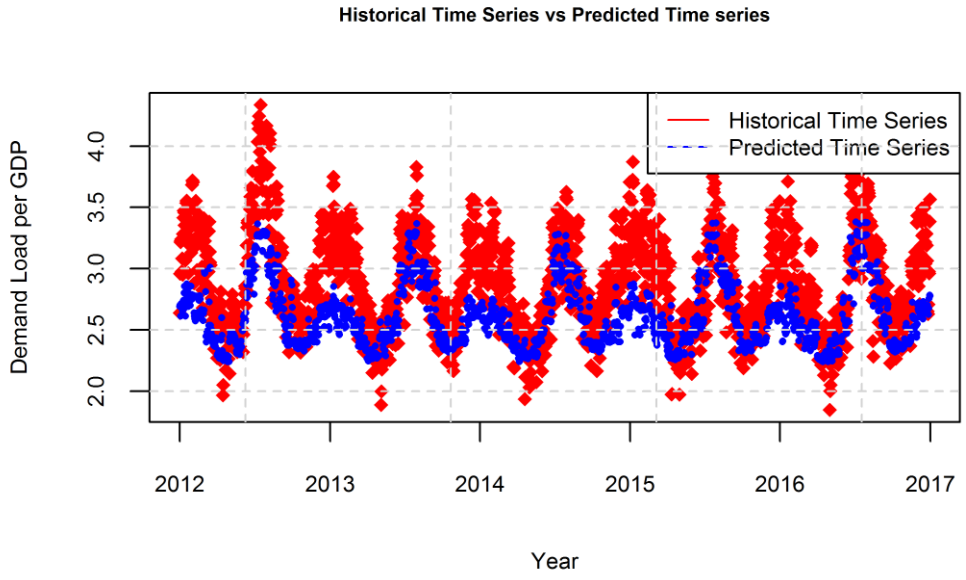
Σχήμα 4.19. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_2) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 3 (rf_3): f(temp_1, temp_2, temp_3), εδώ το μοντέλο επηρεάζεται από τις θερμοκρασίες των τριών προηγούμενων ημερών (Σχήμα 4.20).



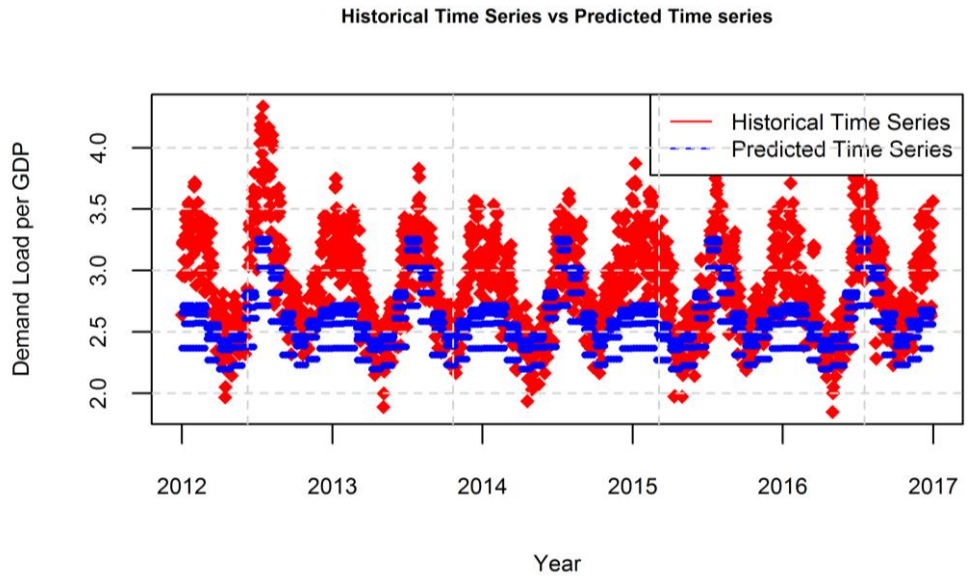
Σχήμα 4.20. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_3) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 4 (rf_4): f(temp_1, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας και το μήνα (Σχήμα 4.21).



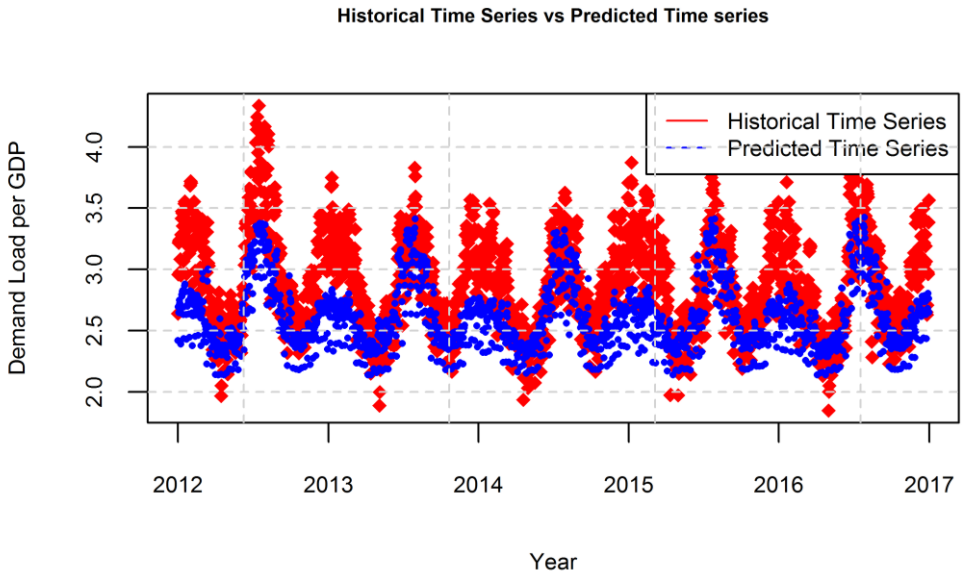
Σχήμα 4.21. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_4) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 5 (rf_5): f(weekday, month), εδώ το μοντέλο επηρεάζεται από την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.22).



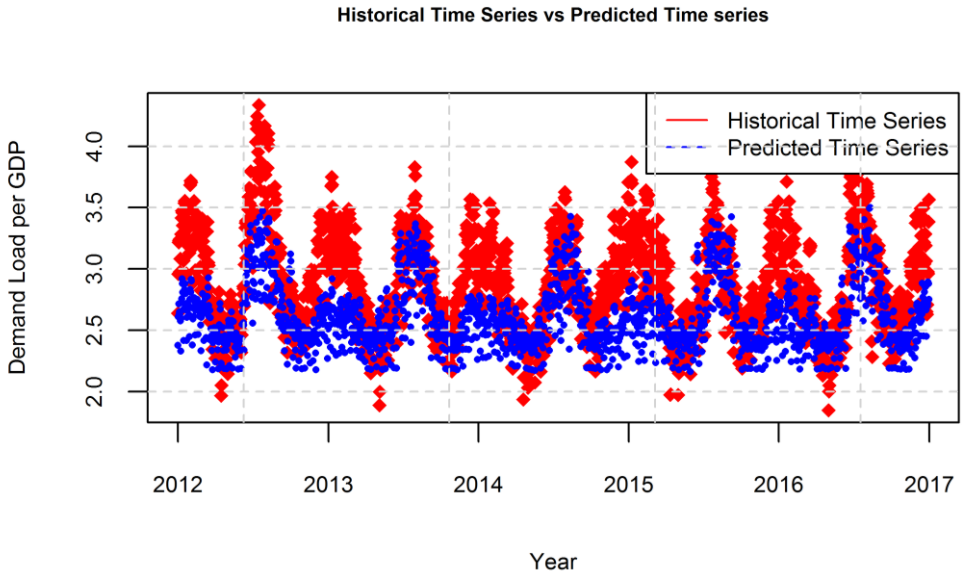
Σχήμα 4.22. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_5) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 6 (rf_6): f(temp_1, weekday, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας, την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.23).



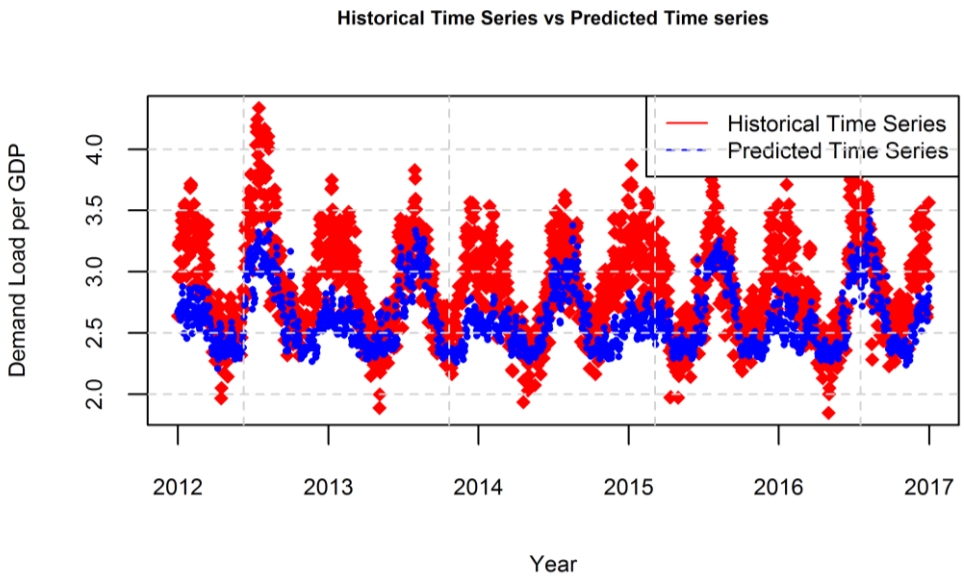
Σχήμα 4.23. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_6) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 7 (rf_7): f(temp_1, temp_2, weekday), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των δύο προηγούμενων ημερών και την ημέρα της εβδομάδος (Σχήμα 4.24).



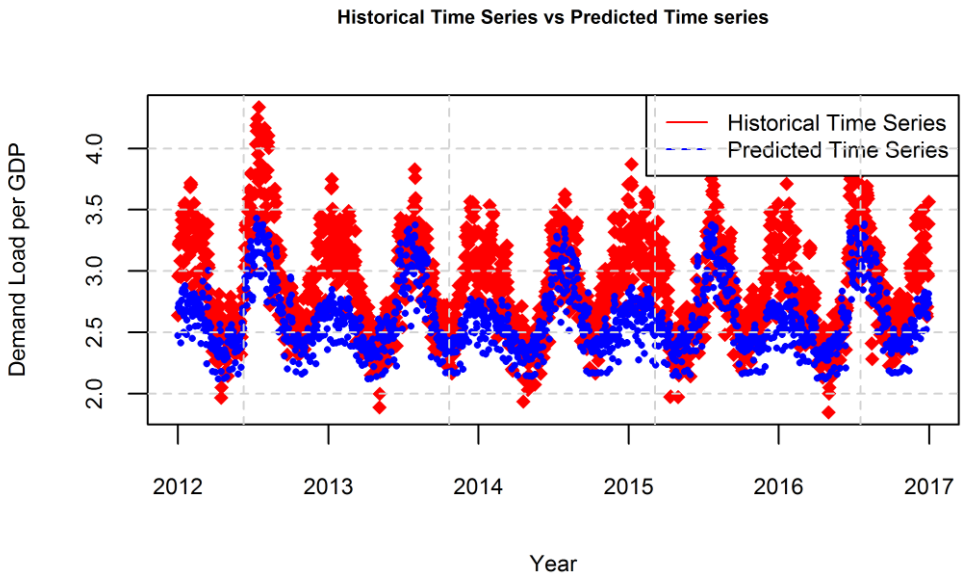
Σχήμα 4.24. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_7) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 8 (rf_8): f(temp_1, temp_2, temp_3, temp_4, temp_5, temp_6, temp_7), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία και των επτά προηγούμενων ημερών (Σχήμα 4.25).



Σχήμα 4.25. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_8) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

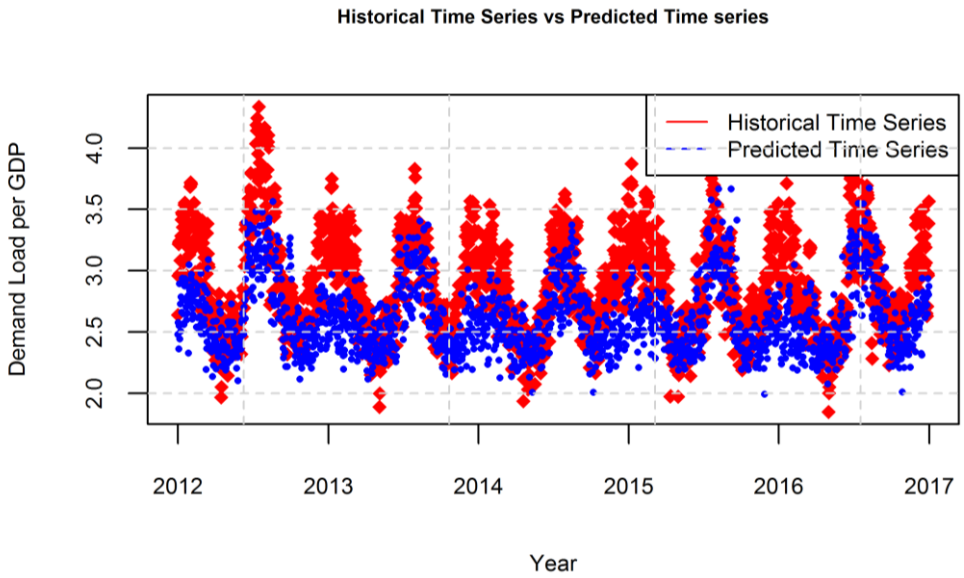
- Random forest Model 9 (rf_9): f(temp_1, temp_2, temp_3, weekday, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των τριών προηγούμενων ημερών, την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.26).



Σχήμα 4.26. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_9) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

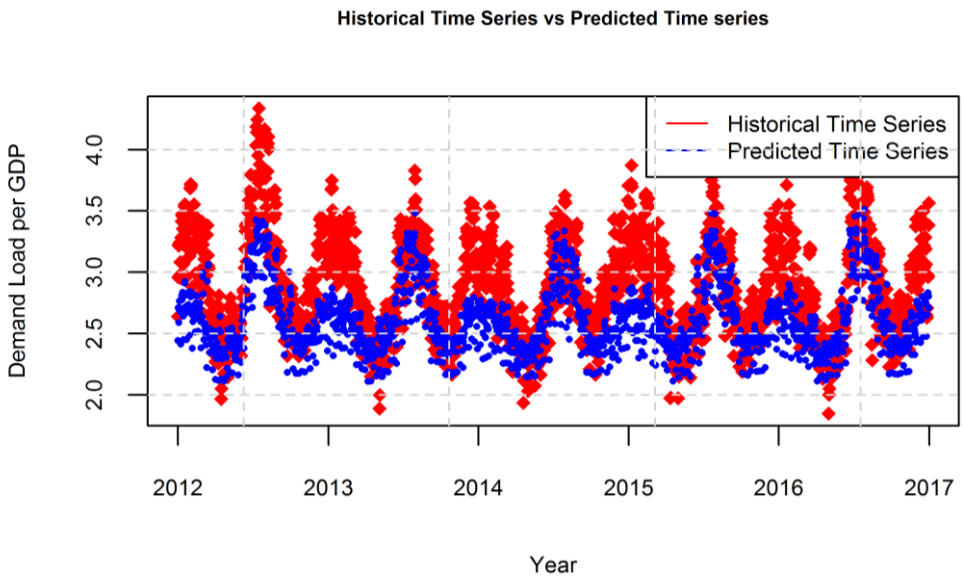
Παρατηρούμε ότι μέχρι στιγμής τα μοντέλα μας, τουλάχιστον οπτικά, πλησιάζουν περισσότερο απ' ότι τα αντίστοιχα γραμμικά τα πραγματικά μας δεδομένα. Βέβαια, σε αυτό το σημείο προκειμένου να βελτιώσουμε την ακρίβεια της πρόβλεψής μας θα αυξήσουμε την πολυπλοκότητα του μοντέλου. Συνεχίζουμε την παράθεση μοντέλων ορίζοντας πλέον τη σχέση θερμοκρασίας-ζήτησης ενέργειας σε κυρτή (όπως δείξαμε και στην ανάλυση σχέσης μεταξύ τους στο Κεφάλαιο 3).

- Random forest Model 10 (rf_10): f((temp_1)^2, temp_1), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας αποκλειστικά (Σχήμα 4.27).



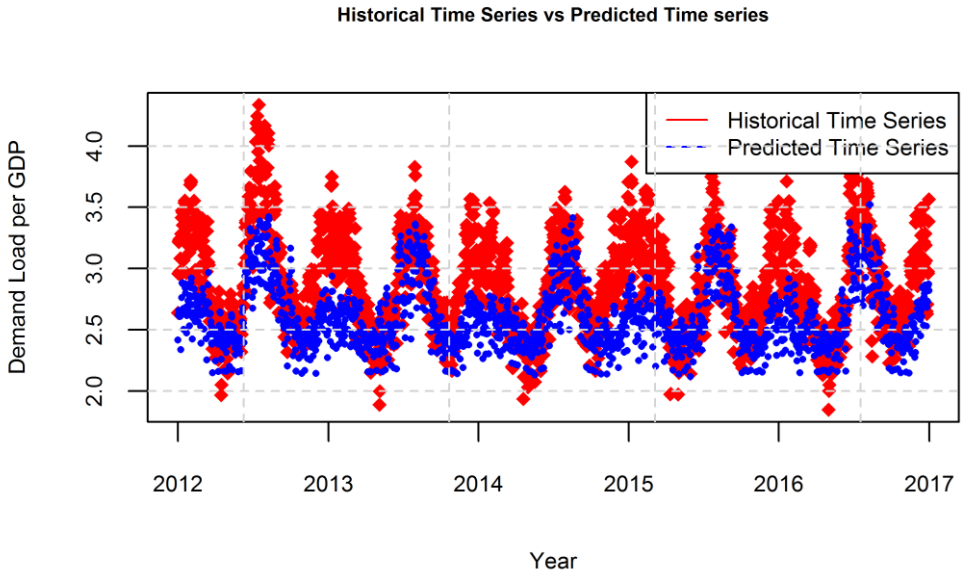
Σχήμα 4.27. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_10) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 11 (rf_11): f((temp_1)^2, temp_1, weekday, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία της προηγούμενης ημέρας, την ημέρα της εβδομάδος και το μήνα (Σχήμα 4.28).



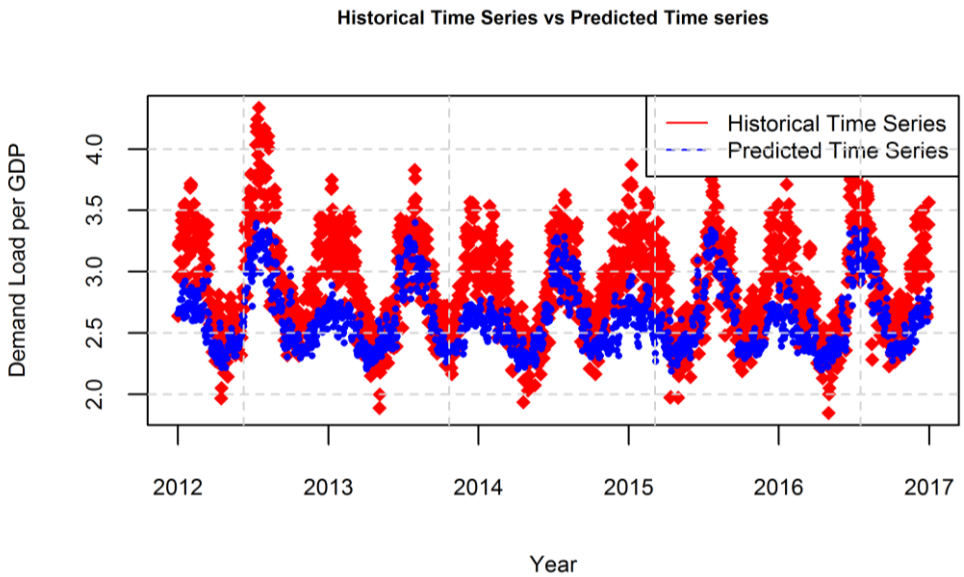
Σχήμα 4.28. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (lm_11) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 12 (rf_12): f((temp_1)^2, temp_1, temp_2, temp_3, weekday), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των τριών προηγούμενων ημερών και την ημέρα της εβδομάδος (Σχήμα 4.29).



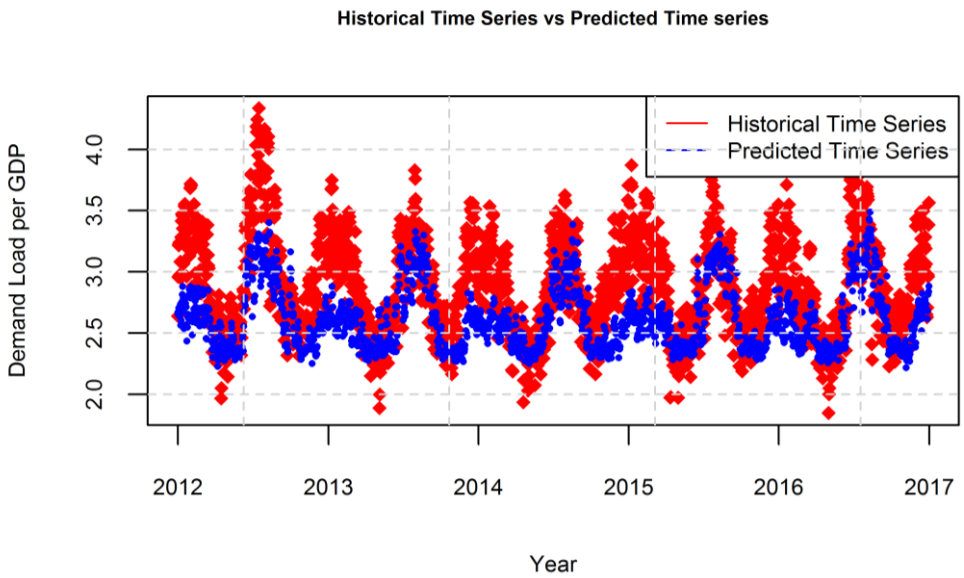
Σχήμα 4.29. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_12) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 13 (rf_13): f((temp_1)^2, temp_1, temp_2, temp_3, month), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των τριών προηγούμενων ημερών και το μήνα (Σχήμα 4.30).



Σχήμα 4.30. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_13) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

- Random forest Model 14 (rf_14): f((temp_1)^2, temp_1, temp_2, temp_3, temp_4, temp_5, temp_6, temp_7), εδώ το μοντέλο επηρεάζεται από τη θερμοκρασία των επτά προηγούμενων ημερών (Σχήμα 4.31).



Σχήμα 4.31. Δεδομένα ιστορικής και συνθετικής χρονοσειράς (rf_14) ζήτησης ηλεκτρικής ενέργειας ανά ΑΕΠ

Σε αυτό το σημείο και αφού έχουμε απεικονίσει όλα τα μοντέλα random forest θα αναλύσουμε την ακρίβεια πρόβλεψης των μοντέλων αυτών και θα τα συγκρίνουμε με τα γραμμικά μοντέλα που προηγήθηκαν. Πέραν από τη γραφική συμπεριφορά του κάθε μοντέλου, είναι απαραίτητο να ληφθούν υπόψη τα στατιστικά μεγέθη κάθε μοντέλου τα οποία αναγράφονται στον Πίνακα 4.2.

Πίνακας 4.2. Στατιστικά μεγέθη για τα μοντέλα random forest

	RMSE	MAE	MAPE	R^2
Random Forest 1	0.423353650	0.339419683	0.112426059	-0.132528974
Random Forest 2	0.390097838	0.306177511	0.099881444	0.038410323
Random Forest 3	0.406006002	0.328966131	0.108567904	-0.041615933
Random Forest 4	0.387174384	0.315493221	0.104108589	0.052768921
Random Forest 5	0.397185909	0.306393612	0.099326864	0.003148722
Random Forest 6	0.370866121	0.295439152	0.096858597	0.130885433
Random Forest 7	0.386008701	0.305089765	0.099879389	0.058464078
Random Forest 8	0.408220543	0.328716734	0.108113945	-0.053009813
Random Forest 9	0.369562778	0.295393931	0.096996834	0.136983394
Random Forest 10	0.423231795	0.339340069	0.112396765	-0.131877112
Random Forest 11	0.369679911	0.296002967	0.097248484	0.136436242
Random Forest 12	0.384630979	0.306054002	0.100320295	0.065173046
Random Forest 13	0.386743635	0.315433352	0.104208882	0.054875425
Random Forest 14	0.406539945	0.328185131	0.108039124	-0.044357414

Παρατηρώντας τις γραφικές απεικονίσεις, αλλά και τα στατιστικά μεγέθη του παραπάνω πίνακα καταλήγουμε στο συμπέρασμα πως η συμπεριφορά των μοντέλων random forest πλησιάζει περισσότερο τα πραγματικά δεδομένα απ' ότι αυτή των γραμμικών. Είναι γεγονός ότι τα μοντέλα γραμμικής παλινδρόμησης έχουν πολύ λίγες παραμέτρους, σε σχέση με τα random forest τα οποία είναι αρκετά πιο πολύπλοκα. Αυτό σημαίνει ότι τα random forest θα προσαρμόζονται πιο εύκολα από τα γραμμικά. Βέβαια, ακόμα και τα μοντέλα αυτά δεν κατάφεραν να προβλέψουν με μεγάλη ακρίβεια τα δεδομένα μας για τα έτη 2012-2016. Συγκεκριμένα, κάποια από αυτά (φαίνονται με κόκκινο στον παραπάνω πίνακα) παρουσιάζουν υψηλά σφάλματα κατά την πρόβλεψη. Αντιθέτως, υπάρχουν μοντέλα τα οποία κατάφεραν να διατηρήσουν τις τάσεις και την εποχικότητα της ιστορικής χρονοσειράς και αποτελούν καλή προσαρμογή των δεδομένων. Τα μοντέλα αυτά φαίνονται με πράσινο χρώμα στον Πίνακα 4.2.

ΚΕΦΑΛΑΙΟ 5

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

5.1 Συμπεράσματα

Όσον αφορά τη συνεισφορά της παρούσας διπλωματικής, αξίζει να αναφερθεί πως δύο παρόμοιες μελέτες έχουν γίνει για τα δεδομένα ηλεκτρικής ενέργειας της Ελλάδας και αναφέρονται στις χρονικές περιόδους 1993-2002 των Mirasgedis et al. (2006) και 2002-2014 του Tyralis (2016). Στην παρούσα εργασία καλύπτεται διαφορετικό χρονικό διάστημα (2006-2016) και χρησιμοποιούνται πιο πρόσφατα δεδομένα. Επιπλέον, δεν βρέθηκε άλλη μελέτη που παρουσιάζει τόσο λεπτομερώς τη ζήτηση ηλεκτρικής ενέργειας στον τομέα του χρόνου και τη σχέση της με μεγάλο αριθμό καιρικών, κλιματικών και κοινωνικοοικονομικών μεταβλητών. Τα αποτελέσματα της παρούσας μελέτης επιβεβαιώνουν και συνοψίζουν τα συμπεράσματα προηγούμενων σχετικών μελετών στην Ελλάδα, με εξαίρεση τις μελέτες που προέβλεπαν τη μέση ετήσια ζήτηση ηλεκτρικής ενέργειας. Αυτές οι μελέτες απέτυχαν να προβλέψουν τη μείωση του ακαθάριστου εγχώριου προϊόντος, που παρατηρήθηκε τα τελευταία χρόνια, και στη συνέχεια απέτυχαν να προβλέψουν μείωση της ζήτησης ηλεκτρικής ενέργειας. Επιπλέον, παρατηρήθηκαν ορισμένες απροσδόκητες συμπεριφορές, που σχετίζονται με την αλλαγή στην ετήσια κατανομή της ζήτησης ηλεκτρικής ενέργειας λόγω της αύξησης των τιμών του πετρελαίου και της μη σταθερότητας σχετικά με τη διακύμανση του μέσου όρου της μηνιαίας ζήτησης ηλεκτρικής ενέργειας, η οποία αν δεν είχε μοντελοποιηθεί σωστά θα μπορούσε να οδηγήσει σε αμφίβολες προβλέψεις. Αυτά τα αποτελέσματα υπογραμμίζουν τη σημασία της διεξαγωγής μιας λεπτομερούς διερευνητικής ανάλυσης δεδομένων πριν από τη μοντελοποίηση της ζήτησης ηλεκτρικής ενέργειας.

Από την άλλη πλευρά, θα πρέπει να δοθεί ιδιαίτερη προσοχή στο βραχυπρόθεσμο και μακροπρόθεσμο σχεδιασμό λόγω της αβεβαιότητας στην πρόβλεψη των κοινωνικοοικονομικών μεταβλητών και της αλληλεπίδρασης που έχουν αυτές με τη ζήτηση ηλεκτρικής ενέργειας σε μικρές χρονικές κλίμακες. Πράγματι, οι κοινωνικοοικονομικές μεταβλητές υπόκεινται σε απροσδόκητες αλλαγές λόγω της απρόβλεπτης ανθρώπινης συμπεριφοράς με αποτέλεσμα να δημιουργούν δυσκολίες κατά την μοντελοποίηση και καθιστούν έτσι απαραίτητη μία διεπιστημονική προσέγγιση. Η διερευνητική ανάλυση δεδομένων της ζήτησης ηλεκτρικής ενέργειας είναι απαραίτητη για τη διερεύνηση της χρησιμότητας ενός μοντέλου ή ενός λογισμικού στη μοντελοποίηση της ζήτησης ηλεκτρικής ενέργειας, προτού χρησιμοποιηθεί. Το πλαίσιο που εισήχθη στην παρούσα μελέτη στην οποία η ζήτηση ηλεκτρικής ενέργειας, το ακαθάριστο εγχώριο προϊόν και η θερμοκρασία απεικονίστηκαν σε διάφορες χρονικές κλίμακες για να επισημανθεί η αλληλεπίδρασή τους, μπορεί να χρησιμεύσει ως σχεδιάγραμμα για την πραγματοποίηση μιας διερευνητικής ανάλυσης δεδομένων. Επίσης, τα αποτελέσματα της μελέτης αυτής είναι χρήσιμα για τη δημιουργία ενός μοντέλου για βραχυπρόθεσμες, μεσοπρόθεσμες και μακροπρόθεσμες προβλέψεις ζήτησης ηλεκτρικής ενέργειας. Αυτό το μοντέλο θα μπορούσε να εφαρμοστεί στη διαμόρφωση μιας πολιτικής διαχείρισης για την καθημερινή λειτουργία και τον προγραμματισμό συντήρησης, αλλά και για την προσομοίωση των τιμολογίων σε διάφορα σενάρια, σχετικά με την εγκατάσταση νέων ενεργειακών συστημάτων, βελτιώνοντας τη στρατηγική σχεδιασμού.

Στο δεύτερο μέρος της μελέτης αναπτύσσεται και εφαρμόζεται μια προσέγγιση για τη βραχυπρόθεσμη πρόβλεψη ενεργειακής ζήτησης με τη χρήση γραμμικών μοντέλων αλλά και μοντέλων random forest. Γίνεται αρχικά μια γραφική απεικόνιση κάθε συνθετικής χρονοσειράς επάνω στα πραγματικά δεδομένα με σκοπό την αξιολόγηση προσαρμογής του μοντέλου. Στη συνέχεια, υπολογίζονται και σχολιάζονται τα στατιστικά μεγέθη των προβλέψεων (RMSE, MAE, MAPE, R^2). Παρατηρείται πως η αύξηση της πολυπλοκότητας ενός μοντέλου δεν οδηγεί απαραίτητα στην καλύτερη προσαρμογή αυτού. Επομένως, από τα μοντέλα που δημιουργήσαμε επιλέγουμε το καταλληλότερο και ταυτόχρονα απορρίπτουμε αυτά τα οποία δεν κατάφεραν με κάποιο τρόπο να προσεγγίσουν τα ιστορικά δεδομένα μας για την περίοδο 2012-2016.

5.2 Προοπτικές

Όσον αφορά την ανάλυση της ζήτησης ηλεκτρικής ενέργειας στον χρόνο, αυτή έγινε για την Ελλάδα χωρίς προηγουμένως να έχουν αφαιρεθεί χρήσεις της ηλεκτρικής ενέργειας, όπως η βιομηχανική και η γεωργική. Το είδος του τελικού χρήστη επηρεάζει σε μεγάλο βαθμό τη συμπεριφορά της ζήτησης. Σύμφωνα με τον Tyralis (2016), διαφορετικές περιοδικότητες εμφανίζει το μοτίβο της ζήτησης για οικιακή και βιομηχανική χρήση. Η οπτικοποίηση της ζήτησης ηλεκτρικής ενέργειας που έγινε αφορά τη χρονική περίοδο 2006-2016, η οποία είναι σχετικά μικρή. Επιπρόσθετα, στη μελέτη αυτή έχουν ληφθεί δεδομένα θερμοκρασίας από έναν αντιπροσωπευτικό σταθμό για την Ελλάδα. Για περαιτέρω ανάλυση, προτείνεται η συλλογή δεδομένων από περισσότερους σταθμούς και για πιο πρόσφατα δεδομένα.

Σε ότι αφορά την προσομοίωση, το ζητούμενο ήταν η παραγωγή μιας συνθετικής χρονοσειράς βραχυπρόθεσμα. Για τη συγκεκριμένη εφαρμογή λήφθηκαν υπόψη παράγοντες θερμοκρασίας, ΑΕΠ, είδος ημέρας, κ.α. και βρέθηκε η συσχέτιση αυτών των μεταβλητών με τη ζήτηση της ηλεκτρικής ενέργειας. Τα μοντέλα που εξετάσθηκαν ήταν μοντέλα γραμμικής παλινδρόμησης και random forest. Λόγω της πολυπλοκότητας των δεδομένων, είναι θεμιτό να ληφθούν υπόψη και άλλα μοντέλα πιο σύνθετα για τη βελτίωση της ακρίβειας της πρόβλεψης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Λαδάς, Π., Διπλωματική Εργασία με Θέμα: *Βραχυπρόθεσμη πρόβλεψη ενεργειακής ζήτησης: Προσεγγίσεις βασισμένες στη Μηχανική Μάθηση*, 2015.

Λιόση, Ε., Διπλωματική Εργασία με Θέμα: *Διερεύνηση της χωροχρονικής ανάλυσης της κατανάλωσης ενέργειας στην Ελλάδα σε επίπεδο NUTS2 και συσχέτισή της με την οικονομική κρίση*, 2018.

Παναγιωτάκος, Θ., και Σ.Τ. Πονής, *Τεχνικές Εξόρυξης Δεδομένων (Data Mining) : Μελέτη Εφαρμογής σε Ελληνική Εταιρεία του κλάδου της Ένδυσης με τη Χρήση του Λογισμικού Weka,* 11ο Φ.Σ. Διοικητικής Επιστήμης και Τεχνολογίας, Αθήνα, Ελλάδα, 13 Μαΐου 2014.

Χαραλάμπους, Κ., Διπλωματική Εργασία με Θέμα: *Οικονομικές Επιπτώσεις Ολυμπιακών Αγώνων και Τεχνικές Εξόρυξης Πληροφορίας: Απόπειρες Εξαγωγής Αποτελεσμάτων και Συμπεράσματα*, Αθήνα, Ιούλιος 2009.

Afendras, G., M. Markatou, *Optimality of Training/Test Size and Resampling Effectiveness of Cross-Validation Estimators of the Generalization Error*, 2015.

Babu, C.N., and B.E. Reddy, *A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data,* Appl. Soft Comput. 23, 27–38, 2014.

Balkin, S., and J. Ord, *Automatic Neural Network Modeling for Univariate Times Series*, International Journal of Forecasting, 16, 509-515, 2000.

Behrens, J.T., *Principles and procedures of exploratory data analysis*. Psychol Methods, 2(2):131e60, 1997.

Behrens, J.T., and CH. Yu, *Exploratory data analysis*, In: Schinka JA, Velicer WF, editors. Handbook of psychology, vol. 2. John Wiley & Sons, Inc; p. 33e64, 2003.

Bishop, C.M., *Neural Network for Pattern Recognition*, Oxford University Press Inc., New York, (1995).

Breiman, L., *Random Forests*, Machine Learning 45, 5–32, 2001.

Bruhns, A., G. Deurveilher, and J.S. Roy, *A non-linear regression model for mid- term load forecasting and improvements in seasonality*, In: 15th Power Systems Computation Conference, 2005.

Charytoniuk, W., M.S. Chen, and P. Van Olinda, *Nonparametric Regression Based Short-Term Load Forecasting*. IEEE Transactions on Power Systems, 13:725–730, 1998.

Dagoumas, A., and F. Kitsios, *Assessing the impact of the economic crisis on energy poverty in Greece*, Sustain Cities Soc; 13:267e78, 2014.

De Gooijer, J.G., and R.J. Hyndman, *25 years of time series forecasting*, Int J Forecast 22(3):443–473, 2006.

Devi, M. R., and R. Manonmani, *Electricity forecasting using data mining techniques in Tamil Nadu and other countries: a survey*, International Journal of Emerging Trends in Engineering and Development, 2012.

Dunham, M.H., *Data Mining introductory and advanced topics*, Prentice Hall, 2004.

Farzan, F., M.A. Jafari, J. Gong, F. Farzan, and A. Stryker, *A multi-scale adaptive model of residential energy demand*, Appl. Energy; 150:258e73, 2015.

Fayyad, U., G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, MIT Press, Cambridge, 1996.

Feinberg, E.A., and D. Genethliou, *Load forecasting*, in Applied mathematics for restructured electric power systems. Springer, pp. 269–285, 2005.

Figueiredo, V., F. Rodrigues, and Z. Vale, *An Electric Energy Consumer Characterization Framework Based On Data Mining Techniques*, IEEE Transactions On Power Systems, Vol. 20, No. 2, May 2005.

Filik, Ü.B., ÖN Gerek, and M. Kurban, *A novel modeling approach for hourly forecasting of long-term electric energy demand*, Energ Convers Manage;52(1):199e211, 2011.

Haykin, S., *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, NJ: Prentice Hall, 2nd edition, 1999.

Hahn, H., S. Meyer-Nieberg, and S. Pickl, *Electric load forecasting methods: Tools for decision making*, European Journal of Operational Research, vol. 199, no. 3, 902–907, 2009.

Han, J., M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, MorganKaufmann Publishers, 2006.

Haida, T., and S. Muto, *Regression Based Peak Load Forecasting using a Transformation Technique,* IEEE Transactions on Power Systems, 9:1788–1794, 1994.

Hong, T., and S. Fan, *Probabilistic electric load forecasting: a tutorial review*. Int J Forecast; 32(3):914e38, 2016.

Hor, C.L., S.J. Watson, and S. Majithia S. Analyzing the impact of weather variables on monthly electricity demand. IEEE Transactions on Power Systems, 20(4):2078e85, 2005.

Hyde, O., and P.F. Hodnett, *An Adaptable Automated Procedure for Short-Term Electricity Load Forecasting*, IEEE Transactions on Power Systems, 12:84–93, 1997.

Hyndman, R.J., and S. Fan, *Density forecasting for long-term peak electricity demand*, IEEE Transactions on Power Systems, 25(2):1142e53, 2010.

Kaur et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(7), pp. 1040-1044, July, 2014.

Khashei, M., and M. Bijari, *A novel hybridization of artificial neural networks and ARIMA models for time series forecasting*, Appl. Soft Comput., 11, 2664–2675, 2011.

Kyriakides, E., and M. Polycarpou, *Short term electric load forecasting: A tutorial*. In: Chen, K., Wang, L. (Eds.), Trends in Neural Computation, Studies in Computational Intelligence, vol. 35. Springer, pp. 391–418 (Chapter 16), 2007.

Lalas, D.P., S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, and A. Adamopoulou, *Impacts of environmental restrictions on the structure of the electricity sector*.,Technical report prepared by National Observatory of Athens for the Greek Public Power Corporation, Lofos Nimfon, Thission, 11810 Athens, 2003.

Lin, L., F. Wang, X. Xie, and S. Zhong, *Random forests-based extreme learning machine ensemble for multi-regime time series prediction*, Expert Syst. Appl. 2017, 85,

164–176, 2017.

Liu, K., S. Subbarayan, R.R. Shoults, and M.T., Manry, *Comparison of very short-term load forecasting techniques*, Autom. & Robotics Res. Inst., Texas Univ., Arlington, TX, USA, 1996.

Makridakis, S., and M. Hibon, *The M3-competition: Results, conclusions and implications*, International Journal of Forecasting 16, 451–476, 2000.

Mavroyeoryos, K., I. Engonopoulos, H. Tyralis, P. Dimitriadis, and D. Koutsoyiannis, *Simulation of electricity demand in a remote island for optimal planning of a hybrid renewable energy system*, Energy Procedia, 125, 435–442, 2017.

Mirasgedis, S., Y. Sarafidis, E. Georgopoulou, D.P. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou, *Models for mid-term electricity demand forecasting incorporating weather influences*, Energy 31(2-3):208-227, 2006.

Moral-Carcedo, J., and J. Vicens-Otero, *Modelling the non-linear response of Spanish electricity demand to temperature variations*, Energ Econ;27(3):477e94, 2005.

Nikolopoulos, V., G. Mpardis, I. Giannoukos, I. Lykourentzou, and V. Loumos, *Web- based decision-support system methodology for smart provision of adaptive digital energy services over cloud technologies*, Software, IET, vol. 5, no. 5, pp. 454–465, 2011.

Papacharalampous, G.A., H. Tyralis, and D. Koutsoyiannis, *Forecasting of geophysical processes using stochastic and machine learning algorithms*, In Proceedings of the 10th World Congress of EWRA on Water Resources and Environment "Panta Rhei", Athens, Greece, 5–9 July 2017.

Pappas, S.S., L. Ekonomou, P. Karampelas, D.C. Karamousantas, S.K. Katsikas, G.E. Chatzarakis, and P.D. Skafidas, *Electricity demand load forecasting of the Hellenic power system using an ARMA model*, Electric Power Systems Research 80(3):256-264, 2010.

Payne, J.E., *A survey of the electricity consumption-growth literature*. Appl Energy; 87(3):723e31, 2010.

Pérez-Rodríguez, J.V., S. Torra, and J. Andrada-Félix, *STAR and ANN models: Forecasting performance on the Spanish "Ibex-35" stock index*, J. Empir. Finance, 12, 490–509, 2005.

Piatetsky-Shapiro, G., and W. J. Frawley, *Knowledge Discovery in Databases*, AAAI (American Association for Artificial Intelligence)/MIT Press, 1991.

Polemis, M.L., and A.S. Dagoumas, *The electricity consumption and economic growth nexus: evidence from Greece*. Energ Policy; 62:798e808, 2013.

Psiloglou, B.E., C. Giannakopoulos, S. Majithia, and M. Petrakis, *Factors affecting electricity demand in Athens, Greece and London, UK: A comparative assessment*, Energy 34(11):1855–1863, 2009.

Ripley, B. D., *Neural Networks and Related Methods for Classification*, Journal of the Royal Statistics Society, 56(3), 409-437, 1994.

Ruzic, S., A. Vuckovic, and N. Nikolic, *Weather Sensitive Method for Short-Term Load Forecasting in Electric Power Utility of Serbia*, IEEE Transactions on Power Systems, 18:1581 1586, 2003.

Shearer, C., *The CRISP-DM model: the new blueprint for data mining*, J Data Warehousing, 5:13—22, 2000.

Shmueli, G., *To Explain or to Predict?*, Statist. Sci. 25, no. 3, 289—310, 2010.

The R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing: Vienna, Austria, 2017.

Tsekouras, G.J., Ch. N. Elias, S. Kavatza, and G.C. Contaxis, *A Hybrid Non-Linear Regression Midterm Energy Forecasting Method Using Data Mining*, 2003 IEEE Bologna Power Tech Conference, Bologna, Italy, June 23th-26th 2003.

Tukey, J.W., *Exploratory data analysis*, Pearson Addison Wesley, 1977.

Tyralis, H., *Spatial and temporal analysis of electricity demand in Greece*, MSc Thesis, 95 pages, Department of Water Resources and Environmental Engineering - National Technical University of Athens, Athens, 2016.

Xu, R., and C.D. II Wunsch, *Survey of Clustering Algorithms*, IEEE Transactions on Neural Networks, vol. 16, no. 3, pp 645-678, 2005.

Yan, W., *Toward automatic time-series forecasting using neural networks*. IEEE Trans. Neural Netw. Lear. Stat., 23, 1028–1039, 2012.

Zachariadis, T., and N. Pashourtidou, *An empirical analysis of electricity consumption in Cyprus*, Energ Econ. 29(2):183e98, 2007.

Zaiane, O., *Machine Learning and Data Mining*, Dunham University, Lecture notes, 2005.

Zou, H. and Y. Yang, *Combining time series models for forecasting,* Int. J. Forecast, 20, 69–84, 2004.