



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΥΠΗΡΕΣΙΩΝ**

Πειραματική μελέτη, αξιολόγηση και σύγκριση μεθόδων πρόβλεψης
της Οριακής Τιμής Συστήματος της ηλεκτρικής ενέργειας στην
Ελληνική Αγορά Ενέργειας.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σπυρίδων Ι. Αθανασόπουλος

Επιβλέπων: Φιλιππάκης Μιχαήλ

Καθηγητής ΠΑ.ΠΕΙ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7^η Ιουλίου 2021.

.....
Φιλιππάκης Μιχαήλ

Αναπληρωτής Καθηγητής

ΠΑ.ΠΕΙ

.....
Μαγκλογιάννης Ηλίας

Καθηγητής ΠΑ.ΠΕΙ

Αθήνα, Ιούλιος 2021

.....
Πρέντζα Ανδριάνα

Καθηγήτρια ΠΑ.ΠΕΙ

Copyright © Σπυρίδων Ι. Αθανασόπουλος, 2021

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Ευχαριστίες

Ευχαριστώ πολύ τους καθηγητές και τους συναδέλφους μου, μεταπτυχιακούς φοιτητές στο πρόγραμμα που με τις συζητήσεις και τις παραινέσεις τους γεννήθηκε η ιδέα μελέτης της συγκεκριμένης θεματικής ενότητας στα πλαίσια της διπλωματικής εργασίας μου για την περάτωση του Μεταπτυχιακού Προγράμματος Πληροφοριακά Συστήματα και Υπηρεσίες.

Ευχαριστώ ειδικότερα τον επιβλέποντα καθηγητή κ.Μιχαήλ Φιλιππάκη που με τις υποδείξεις και κατευθύνσεις που μου έδωσε με οδήγησε στην κατά το δυνατόν πληρέστερη κάλυψη του αντικειμένου καθώς και στην αρτιότερη παρουσίαση του παρόντος εγγράφου.

Η παρούσα διπλωματική είναι αφιερωμένη στους κοντινούς μου ανθρώπους και κυρίως στους γονείς μου οι οποίοι μου παρείχαν υλικά και ψυχικά ότι ήταν απαραίτητο καθ' όλη τη διάρκεια των σπουδών μου.

Περίληψη

Σκοπός της παρούσας Διπλωματικής, είναι η εφαρμογή και αξιολόγηση διάφορων μοντέλων πρόβλεψης της Οριακής Τιμής Συστήματος της ηλεκτρικής ενέργειας στην Ελληνική Αγορά Ενέργειας. Αρχικά παρουσιάζεται η φυσική ερμηνεία της Οριακής Τιμής Συστήματος Ενέργειας και η επίδραση της ως δείκτης στη σύγχρονη αγορά Ενέργειας. Ακολούθως, γίνεται προσπάθεια καταγραφής των διάφορων μεθόδων πρόβλεψης της ΟΤΣ και ο διαχωρισμός τους σε Αιτιοκρατικά μοντέλα και μοντέλα Χρονοσειρών. Σε επόμενο στάδιο έγινε μια θεωρητική παρουσίαση των μετρικών σφαλμάτων τα οποία θα υπολογιστούν στη συνέχεια για κάθε μοντέλο προκειμένου στο τέλος να αποτελέσουν τους βασικούς δείκτες για την επιλογή του κατάλληλου μοντέλου για την πρόβλεψη σε μελλοντικά δεδομένα.

Έχοντας καλύψει επαρκώς το θεωρητικό υπόβαθρο του υπό εξέταση εγχειρήματος, σειρά έχει η ενασχόληση με το πρακτικό κομμάτι. Η ανάλυση πραγματοποιήθηκε με δεδομένα από το αρχείο της ΡΑΕ, συγκεκριμένα ωριαία δεδομένα ΟΤΣ από 01/01/2018 έως και 01/11/2020. Με τους κατάλληλους μετασχηματισμούς, τα δεδομένα φορτώθηκαν σε αντίστοιχες δομές δεδομένων και μελετήθηκαν σαν χρονοσειρά ως προς τη στασιμότητα, την εποχικότητα και τις ελλειπούσες τιμές οι οποίες έχουν θεωρητική εξήγηση. Έγινε προσπάθεια εκτίμησης της πολλαπλής εποχικότητας της χρονοσειράς μέσω του Ταχέως Μετασχηματισμού Fourier και οι εποχικότητες αυτές χρησιμοποιήθηκαν παρακάτω. Χρησιμοποιήθηκε μια σειρά από μεθόδους και κρίθηκε αναγκαίο να διαχωριστεί η διαδικασία πρόβλεψης σε μακροπρόθεσμη και βραχυπρόθεσμη, καθώς οι πειραματικές συνθήκες, η φύση των δεδομένων και κυρίως ο όγκος των δεδομένων επιτάσσουν διαφορετική αντιμετώπιση ανά περίπτωση. Ενδεικτικά αναφέρονται για τη βραχυπρόθεσμη πρόβλεψη ότι εκπαιδεύτηκαν στα δεδομένα των 2 πρώτων ετών και δοκιμάστηκαν σ' αυτά του τρίτου έτους η βιβλιοθήκη TBATS, μοντέλα ARMA από διαφορετικές βιβλιοθήκες, υπέρθεση μοντέλων ARIMA με συνιστώσες Fourier και η βιβλιοθήκη FBProphet, με τις τελευταίες 2 μεθόδους να χρησιμοποιούνται και στη μακροπρόθεσμη πρόβλεψη υπό διαφορετικές φυσικά συνθήκες.

Τέλος, παρουσιάστηκαν οι αντίστοιχες μετρικές για κάθε μοντέλο και με βάση αυτές έγινε επιλογή του κατάλληλου μοντέλου πρόβλεψης με σκοπό στο τελευταίο κεφάλαιο να εκπαιδευτεί με το σύνολο των υπάρχοντων δεδομένων και να αποπειραθούν σε μελλοντική πρόβλεψη της ΟΤΣ σε βραχυχρόνια και μακροχρόνια διάρκεια, και να εξαχθεί γνώση από αυτό.

Λέξεις Κλειδιά: Οριακή Τιμή Συστήματος, Ελληνική Αγορά Ηλεκτρικής Ενέργειας, μέθοδοι προβλέψεων, κυλιόμενη πρόβλεψη.

Abstract

The purpose of this Diploma Thesis is the implementation and evaluation of various models for forecasting the System Marginal Price of electricity in the Greek Energy Market. Firstly, the physical interpretation of the Energy System Marginal Price and its effect as an indicator on the modern Energy market is presented. Then, an attempt is made to record the various methods of forecasting the OTS and their separation into causal models and time series models. In the next stage, a theoretical presentation of the metric errors was made, which will then be calculated for each model in order to indicate the appropriate model for the prediction in future data.

Having covered the theoretical background of this project, it is time to deal with the practical part. The analysis was carried out with data from the RAE archive, to be more specific with hourly OTS data from 01/01/2018 until 01/11/2020. After applying appropriate transformations, these data were loaded into corresponding data structures and examined as a time series model in terms of stationarity, seasonality and missing values which have a theoretical explanation. Another attempt was made, this time, to estimate multiple seasonalities of the time series object through the Fast Fourier Transform and these seasonalities were used below. A number of methods were used and it was deemed necessary to separate the forecasting process into long-term and short-term, since the experimental conditions, data nature and especially data volume result to different handling on a case-by-case basis. Regarding short-term prediction TBATS library was tested, ARMA models from different libraries, ARIMA model superposition with Fourier components and the FBProphet library. All the above models were trained in first 2 years data and tested in the third year, using the last 2 methods in long-term forecast under different physical conditions.

Finally, the corresponding metrics for each model were presented and based on them the appropriate forecasting model was selected. As a result, in the last chapter the best model trained with all existing data and future forecasting of OTS was attempted both in short and long term, trying to extract knowledge and business value.

Key words: System's Marginal Price, Market's Clearing Price, Greek Electricity Market, forecasting methods, rolling forecasts.

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	2
ABSTRACT	3
ΚΕΦΑΛΑΙΟ 1	6
1.1 ΕΙΣΑΓΩΓΗ	6
1.2 ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	8
1.2.1 Αγορά Μακροχρόνιας Διαθεσιμότητας Ισχύος	13
1.2.2 Ημερήσιος Ενεργειακός Προγραμματισμός (Αγορά Ενέργειας και Επικουρικών Υπηρεσιών ή Χονδρεμπορική Αγορά)	13
1.3 ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΕΠΙΛΥΣΗ ΤΟΥ ΗΕΠ	14
1.3.1 Συμμετοχή στον ΗΕΠ.....	15
1.4 ΟΡΙΑΚΗ ΤΙΜΗ ΣΥΣΤΗΜΑΤΟΣ (ΟΤΣ)	16
ΚΕΦΑΛΑΙΟ 2	19
2.1 ΥΠΟΔΟΜΗ	19
2.2 ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΤΕΧΝΟΛΟΓΙΩΝ	19
ΚΕΦΑΛΑΙΟ 3	23
3.1 ΠΡΟΒΛΕΨΗ ΟΡΙΑΚΗΣ ΤΙΜΗΣ ΣΥΣΤΗΜΑΤΟΣ (ΟΤΣ)	23
3.1.1 Βραχυπρόθεσμη Πρόβλεψη	23
3.1.2 Μακροπρόθεσμη Πρόβλεψη.....	23
3.2 ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΟΤΣ.....	24
3.3 ΒΑΣΙΚΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΠΡΟΒΛΕΨΗΣ	25
3.3.1 Μοντέλα Χρονοσειρών.....	25
3.3.2 ΑΙΤΙΟΚΡΑΤΙΚΑ ΜΟΝΤΕΛΑ.....	31
3.4 ΜΕΘΟΔΟΙ ΓΙΑ ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ	33
3.5 ΜΕΘΟΔΟΙ ΓΙΑ ΜΑΚΡΟΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ.....	34
3.6 ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ	35
3.6.1 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error)	35
3.6.2 Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error)	35
3.6.3 Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error).....	36
3.6.4 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error).....	36
3.6.5 Συμμετρικό Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (symmetric Mean Absolute Percentage Error).....	36
3.6.6 Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα Εύρους (Mean Absolute Range Normalized Error).....	37
ΚΕΦΑΛΑΙΟ 4	39
4.1 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	39
4.2 ΜΕΛΕΤΗ ΧΡΟΝΟΣΕΙΡΑΣ.....	40
4.2.1 Ποιοτικά Χαρακτηριστικά των Χρονοσειρών.....	40
4.2.2 Ποσοτικά Χαρακτηριστικά των Χρονοσειρών	41
4.2.3 Στασιμότητα χρονοσειράς	43
4.2.4 Μελέτη αυτοσυσχέτισης.....	47
4.3 ΑΠΟΣΥΝΘΕΣΗ ΧΡΟΝΟΣΕΙΡΑΣ (DECOMPOSING)	48
4.4 ΜΕΛΕΤΗ ΕΠΟΧΙΚΟΤΗΤΑΣ ΜΕ FOURIER	51
ΚΕΦΑΛΑΙΟ 5	55
5.1 ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ	55
5.2 ΒΡΑΧΥΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ	55
5.2.1 Μοντέλο TBATS.....	55
5.2.2 Μοντέλο Arima (βιβλιοθήκη pyramid-arima).....	57
5.2.3 Μοντέλο κινητών μέσων (βιβλιοθήκη statsmodels.tsa.statespace.SARIMAX)	59
5.2.4 Μοντέλο SARIMAX με συνιστώσες Fourier	61
5.2.5 Μοντέλο FBProphet	62
5.3 ΜΑΚΡΟΠΡΟΘΕΣΜΗ ΠΡΟΒΛΕΨΗ	67
ΚΕΦΑΛΑΙΟ 6	73

6.1 ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΜΕΤΡΙΚΩΝ.....	73
6.2 ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	74
ΚΕΦΑΛΑΙΟ 7	79
7.1 ΜΕΛΛΟΝΤΙΚΗ ΠΡΟΒΛΕΨΗ	79
7.1.1 Αποτελέσματα Βραχυπρόθεσμης Πρόβλεψης	79
7.1.2 Αποτελέσματα Μακροπρόθεσμης Πρόβλεψης	83
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	85
ΠΑΡΑΡΤΗΜΑ ΡΥΘΜΩΝ	86

Κεφάλαιο 1

1.1 Εισαγωγή

Το έτος 1889 έφτασε ο «ηλεκτρισμός» στην Ελλάδα. Σύμφωνα με τα ιστορικά στοιχεία της ΔΕΗ Α.Ε., η «Γενική Εταιρεία Εργοληψιών» κατασκεύασε στην Αθήνα, στην οδό Αριστείδου, την πρώτη μονάδα παραγωγής ηλεκτρικής ενέργειας. Το πρώτο κτίριο που φωτίζεται είναι τα Ανάκτορα και πολύ σύντομα ο ηλεκτροφωτισμός επεκτείνεται στο σημερινό ιστορικό κέντρο της πόλης. Τον ίδιο χρόνο ηλεκτροδοτείται επίσης η Θεσσαλονίκη, η οποία ανήκει ακόμα στην Οθωμανική Αυτοκρατορία.

Η «Βελγική Εταιρία» αναλαμβάνει απ' τις τουρκικές αρχές το φωτισμό και την τροχοδρόμηση της πόλης με την κατασκευή εργοστασίου παραγωγής ηλεκτρικής ενέργειας. Δέκα χρόνια αργότερα κάνουν την εμφάνισή τους στην Ελλάδα οι πολυεθνικές εταιρίες ηλεκτρισμού. Η αμερικανική εταιρία Thomson - Houston με τη συμμετοχή της Εθνικής Τράπεζας ιδρύουν την «Ελληνική Ηλεκτρική Εταιρία» που αναλαμβάνει την ηλεκτροδότηση μεγάλων ελληνικών πόλεων. Μέχρι το 1929 θα έχουν ηλεκτροδοτηθεί 250 πόλεις με πληθυσμό άνω των 5.000 κατοίκων.

Στις πιο απόμακρες περιοχές – στις οποίες ήταν οικονομικά ασύμφορο για τις μεγάλες εταιρίες να κατασκευάσουν μονάδες παραγωγής ηλεκτρικής ενέργειας – την ηλεκτροδότηση αναλαμβάνουν ιδιώτες ή δημοτικές και κοινοτικές αρχές κατασκευάζοντας μικρά εργοστάσια. Το έτος 1950 υπήρχαν στη Ελλάδα περίπου 400 εταιρίες παραγωγής ηλεκτρικής ενέργειας. Ως πρωτογενή καύσιμα χρησιμοποιούσαν το πετρέλαιο και το γαιάνθρακα, αμφότερα εισαγόμενα από το εξωτερικό.

Η κατάτμηση αυτή της παραγωγής σε πολλές μονάδες, σε συνδυασμό με τα εισαγόμενα καύσιμα, εξωθούσε την τιμή της ηλεκτρικής ενέργειας στα ύψη, φτάνοντας στο τριπλάσιο μέχρι και πενταπλάσιο των τιμών που ίσχυαν στις άλλες ευρωπαϊκές χώρες. Η ηλεκτρική ενέργεια ήταν λοιπόν ένα αγαθό πολυτελείας, αν και τις περισσότερες φορές παρεχόταν με ωράριο και οι ξαφνικές διακοπές ήταν σύνηθες φαινόμενο.

Όσον αφορά στην παραγωγή ηλεκτρικής ενέργειας, αυτή επιτυγχάνεται με την εκμετάλλευση διαφόρων πρωτογενών πηγών ενέργειας και παρουσιάζει μεγάλες διαφοροποιήσεις από χώρα σε χώρα, ανάλογα με τους διαθέσιμους εγχώριους Ενεργειακούς Πόρους, την Ενεργειακή Πολιτική της χώρας, τις γεωλογικές, γεωφυσικές και κλιματολογικές ιδιαιτερότητες αυτής.

Οι πηγές παραγωγής ενέργειας διακρίνονται στις συμβατικές που βασίζονται σε ορυκτά στερεά, υγρά ή αέρια καύσιμα, όπως το πετρέλαιο, ο άνθρακας (λιθάνθρακας και λιγνίτης), το φυσικό αέριο, στην πυρηνική ενέργεια και στις ανανεώσιμες πηγές ενέργειας που χρησιμοποιούν διαχρονικές πηγές (άνεμος, ήλιος, νερό κλπ) και δεν εξαντλούν τα περιορισμένα ενεργειακά αποθέματα (ΡΑΕ, 2012).

Τον Αύγουστο του 1950 ιδρύθηκε η ΔΕΗ και ως εκ τούτου, οι δραστηριότητες παραγωγής, μεταφοράς και διανομής της ηλεκτρικής ενέργειας συγκεντρώθηκαν σε ένα δημόσιο φορέα. Η ΔΕΗ αμέσως στρέφεται προς την αξιοποίηση των εγχώριων πηγών ενέργειας, ενώ ξεκινά και η ενοποίηση των δικτύων μεταφοράς ηλεκτρικής ενέργειας σε ένα εθνικό διασυνδεδεμένο Σύστημα.

Τα πλούσια λιγνιτικά κοιτάσματα του ελληνικού υπεδάφους που είχαν νωρίτερα εντοπισθεί, άρχισαν να εξορύσσονται και να χρησιμοποιούνται ως καύσιμη ύλη στις λιγνιτικές μονάδες ηλεκτροπαραγωγής που δημιουργούσε η ΔΕΗ. Παράλληλα, η Επιχείρηση ξεκίνησε την αξιοποίηση της δύναμης των υδάτων με την κατασκευή υδροηλεκτρικών σταθμών στα μεγάλα ποτάμια της χώρας. Από 1.1.2001 η ΔΕΗ Α.Ε. λειτουργεί ως ανώνυμη εταιρεία, ενώ από 12.12.2001 έχει εισαχθεί στα Χρηματιστήρια Αξιών Αθηνών και Λονδίνου.

Η Ρυθμιστική Αρχή Ενέργειας (ΡΑΕ) είναι υπεύθυνη για τη ρύθμιση της αγοράς ενέργειας στη χώρα μας. Στην αγορά αυτή συμμετέχουν οι Παραγωγοί ηλεκτρικής ενέργειας που παράγουν ηλεκτρική ενέργεια από διαφορετικές πηγές, ο Διαχειριστής του Συστήματος Μεταφοράς Ηλεκτρικής Ενέργειας (ΔΕΣΜΗΕ), ο Διαχειριστής του Δικτύου Διανομής Ηλεκτρικής Ενέργειας (ΔΕΣΔΗΕ), οι Προμηθευτές ηλεκτρικής ενέργειας (όπως η ΔΕΗ Α.Ε.) και οι καταναλωτές.

Ο ΔΕΣΜΗΕ συγκροτήθηκε το 2001 και φροντίζει να υπάρχει ανά πάσα στιγμή ισορροπία παραγωγής και κατανάλωσης έτσι ώστε να καλύπτεται η ζήτηση αφού η παραγόμενη ηλεκτρική ενέργεια δεν αποθηκεύεται σε μεγάλες ποσότητες με οικονομικό τρόπο. Επιπλέον, ο ΔΕΣΜΗΕ μαζί με τον Διαχειριστή του Δικτύου Διανομής (ΔΕΣΔΗΕ), καθήκοντα του οποίου επί του παρόντος εκτελεί τμήμα της ΔΕΗ Α.Ε., είναι υπεύθυνοι να διασφαλίζουν ότι η ηλεκτρική ενέργεια μεταφέρεται με αξιοπιστία και ασφάλεια τόσο μέσω του Συστήματος (Δίκτυο υψηλής τάσης) όσο και μέσω του Δικτύου Διανομής (δίκτυο μέσης και χαμηλής τάσης) από τους ηλεκτροπαραγωγούς σταθμούς στους τελικούς καταναλωτές. Οι Προμηθευτές είναι υπεύθυνοι για την πώληση της ηλεκτρικής ενέργειας στους τελικούς καταναλωτές.

Η ενέργεια που παράγεται στους σταθμούς παραγωγής, αρχικά μεταφέρεται σε κοντινούς μετασχηματιστές, οι οποίοι μετατρέπουν την χαμηλή τάση της ηλεκτρικής ενέργειας σε υψηλή. Με αυτό τον τρόπο, η ηλεκτρική ενέργεια μεταφέρεται με τις γραμμές μεταφοράς σε πολύ μεγάλες αποστάσεις με λιγότερες απώλειες, καθώς οι σταθμοί παραγωγής είναι συνήθως μακριά από μεγάλα αστικά κέντρα. Στο ακόλουθο σχήμα, αναπαρίσταται ο μηχανισμός λειτουργίας της Ελληνική Αγοράς Ηλεκτρικής Ενέργειας:

Σχ. 1.1 Αναπαράσταση Ελληνικής Αγοράς Ηλεκτρικής Ενέργειας



Το δίκτυο μεταφοράς, οδηγεί την ηλεκτρική ενέργεια στους υποσταθμούς μέσης και χαμηλής τάσης, στους οποίους μετατρέπεται η τάση της ηλεκτρικής ενέργειας από υψηλή σε μέση και χαμηλή τάση, προκειμένου με τη βοήθεια εναέριων γραμμών να διανεμηθεί σε βιομηχανίες που χρησιμοποιούν μέση τάση και σε σπίτια που χρησιμοποιούν χαμηλή τάση.

Τα συστατικά στοιχεία των γραμμών μεταφοράς είναι τα ακόλουθα:

- Πυλώνες ή πύργοι, στους οποίους στηρίζονται οι αγωγοί των εναέριων γραμμών
- Μονωτήρες, μέσω των οποίων αναρτώνται στους πυλώνες οι αγωγοί γραμμών
- Αγωγοί, κυρίως από χαλκό και αλουμίνιο.

Το δίκτυο Διανομής, περιλαμβάνει:

- Το δίκτυο διανομής μέσης τάσης (20kV) που μεταφέρει την ηλεκτρική ισχύ από τους υποσταθμούς μεταφοράς στους υποσταθμούς διανομής,
- Το δίκτυο διανομής χαμηλής τάσης (220/380V) που μεταφέρει την ηλεκτρική ισχύ από τους υποσταθμούς διανομής στους καταναλωτές.

(<http://www.rae.gr/>, n.d.)

1.2 Περιγραφή και βασικά χαρακτηριστικά

Η Αγορά Ενέργειας μπορεί να διακριθεί σε δύο βασικές κατηγορίες:

- Απελευθερωμένη αγορά (ολιγοπώλιο)
- Μη απελευθερωμένη αγορά (μονοπώλιο)

Πιο συγκεκριμένα, κατά την απελευθερωμένη αγορά σημαίνει ότι οι ποσότητες και οι τιμές καθορίζονται από τη δράση μεταξύ των παραγωγών και των καταναλωτών, και στις περισσότερες περιπτώσεις καταλήγει σε ολιγοπώλιο (*Vesprucci et al., 2009*).

Στη μη απελευθερωμένη ή μονοπωλιακή αγορά, το αγαθό προσφέρεται από μία μόνο επιχείρηση, δεν είναι άμεσα υποκαταστάσιμο από άλλο παρεμφερές αγαθό, και παράλληλα οι καταναλωτές είναι πολυάριθμοι. Η τιμή και η ποσότητα προσφοράς του αγαθού καθορίζονται μόνο από τις αποφάσεις που παίρνει η μονοπωλιακή επιχείρηση (*Κάπρος, 2003*).

Ήδη πολλοί Ευρωπαίοι καταναλωτές απολαμβάνουν στην καθημερινότητά τους το δικαίωμα επιλογής Προμηθευτή. Όμως στις χώρες της Ευρώπης, στις οποίες δεν έχουν ενεργοποιηθεί ιδιαίτερα άλλοι Προμηθευτές ενέργειας πέραν του αρχικού, όπως στην Ελλάδα, το δικαίωμα αυτό δεν έχει ασκηθεί για τη πλειοψηφία των καταναλωτών.

Το πρώτο βήμα προς το χτίσιμο των απαραίτητων προϋποθέσεων για τη λειτουργία της απελευθερωμένης αγοράς ήταν η λήψη μέτρων σε θεσμικό επίπεδο, προκειμένου να εξαλειφθούν οι στρεβλώσεις που επικράτησαν κατά την εικοσαετία 1980-2000, όπου προτεραιότητα δόθηκε στην ασφάλεια του ενεργειακού εφοδιασμού και στη διατήρηση χαμηλών τιμολογίων.

Ωστόσο είναι γεγονός, ότι η διαδικασία απελευθέρωσης της ηλεκτρικής ενέργειας στην Ελλάδα η οποία ξεκίνησε το έτος 2001 και βρίσκεται ακόμη σε εξέλιξη προχωράει με αργά αλλά σταθερά βήματα και για το λόγο αυτό η Ελληνική Αγορά Ενέργειας μπορεί να χαρακτηριστεί ως αγορά υπό απελευθέρωση.

Η δημιουργία κοινών κανόνων για την αγορά ηλεκτρισμού των χωρών της Ν.Α. Ευρώπης και οι σύγχρονες τάσεις για την ενοποίηση της Ευρωπαϊκής αγοράς ηλεκτρικής ενέργειας αποτελούν τις σημερινές προκλήσεις.

Διαχειριστής της Ελληνικής Αγοράς Ηλεκτρικής Ενέργειας είναι ο ΔΕΣΜΗΕ (Διαχειριστής Ελληνικού Συστήματος Μεταφοράς Ηλεκτρικής Ενέργειας) ο οποίος ιδρύθηκε με το Προεδρικό Διάταγμα 328/12-12-2000 ως ανώνυμη εταιρία, αντικείμενο της οποίας είναι η λειτουργία, η εκμετάλλευση, η διασφάλιση της συντήρησης και η μέριμνα για την ανάπτυξη του Συστήματος Μεταφοράς ηλεκτρικής

Ενέργειας σε ολόκληρη τη χώρα, καθώς και των διασυνδέσεων του με τα άλλα δίκτυα με σκοπό να διασφαλίζεται ο εφοδιασμός της χώρας με ηλεκτρική ενέργεια, κατά τρόπο επαρκή, ασφαλή, οικονομικά αποδοτικό και αξιόπιστο.

Η παρακολούθηση και η τήρηση των κανόνων (Κώδικας) που διέπουν την Ελληνική Αγορά Ενέργειας υπάγονται στις υποχρεώσεις της Ρυθμιστικής Αρχής Ενέργειας (ΡΑΕ). Πιο αναλυτικά, η ΡΑΕ συγκροτήθηκε τον Ιούλιο του 2000 και πρόκειται για μία ανεξάρτητη διοικητική αρχή, η οποία έχει ως βασικό αντικείμενο την παρακολούθηση της ανάπτυξης της Αγοράς Ενέργειας, τόσο στην Ελληνική Αγορά, αλλά και στον τρόπο με τον οποίο αυτή λειτουργεί και αναπτύσσεται σε σχέση με τις ξένες αγορές και ιδίως με εκείνες τις οποίες διασυνδέεται.

Ο Κώδικας εφαρμόστηκε προοδευτικά κατά τη διάρκεια μιας μεταβατικής περιόδου που εκτείνεται από το 2005 έως τα τέλη του 2010 και περιλαμβάνει 5 «ημέρες αναφοράς» οι οποίες χαρακτηρίζουν την πρόοδο που έγινε κατά την προαναφερθείσα περίοδο και τους νέους στόχους που τέθηκαν για το μέλλον.

Αξίζει να αναφερθεί πως η ΡΑΕ παίζει κατά κάποιο τρόπο το ρόλο του συμβούλου του Ελληνικού Υπουργείου Ανάπτυξης, το οποίο είναι αυτό που παίρνει τις τελικές αποφάσεις και καθορίζει τον τρόπο λειτουργίας της Ελληνικής Αγοράς Ενέργειας.

Ιδιαίτερα σημαντικό ρόλο στην Ελληνική Αγορά Ενέργειας έχει η ΔΕΗ Α.Ε.

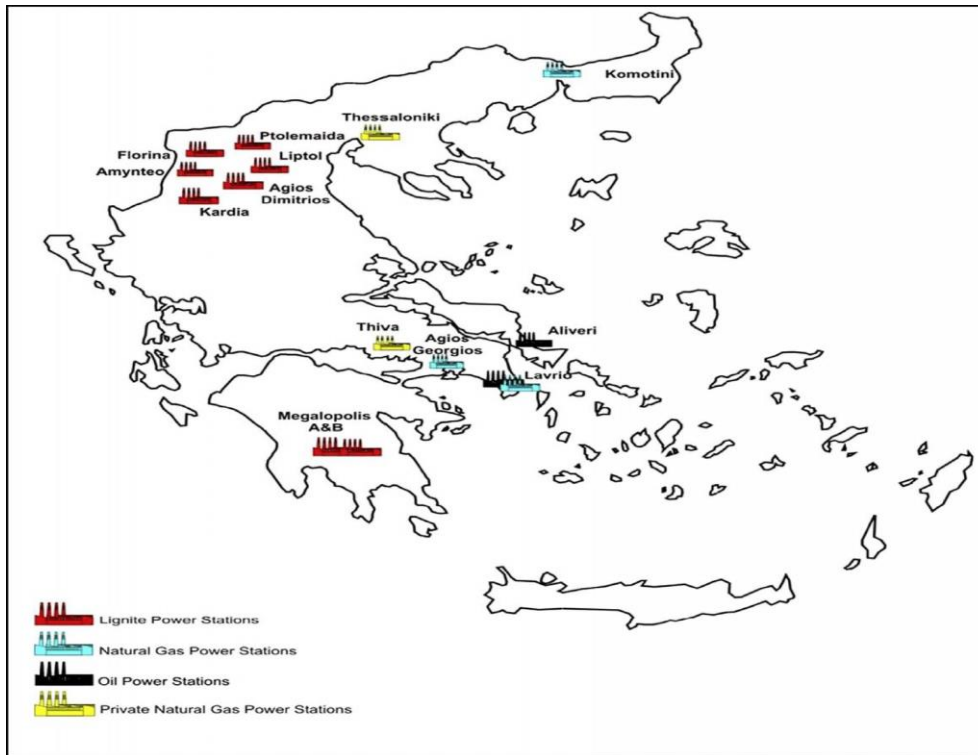
Δραστηριοποιείται ως Παραγωγός και είναι ο κύριος Προμηθευτής ηλεκτρικής ενέργειας, όπου σύμφωνα με τα στοιχεία του 2009, κατέχει περίπου το 89% της εγκατεστημένης ηλεκτρικής ισχύος στην Ελλάδα (12.695 MW) η οποία προέρχεται από λιγνιτικές, πετρελαϊκές, υδροηλεκτρικές μονάδες, μονάδες φυσικού αερίου καθώς επίσης και από αιολικά και ηλιακά πάρκα. Παράγοντας σχεδόν το 55% της ηλεκτρικής της παραγωγής από λιγνίτη, τον καθιστά ως τον 2^ο μεγαλύτερο παραγωγό ηλεκτρικής ενέργειας από λιγνίτη στην Ευρωπαϊκή Ένωση.

Επιπλέον, η ΔΕΗ Α.Ε. ως Κύριος του Συστήματος μεταφοράς υποχρεούται στη συντήρηση του Συστήματος στο πλαίσιο του προγραμματισμού που εκπονεί η ίδια και υπό τις οδηγίες του ΔΕΣΜΗΕ, ενώ παράλληλα εκτελεί χρέη Διαχειριστή του Δικτύου Διανομής μέχρι την ίδρυση του ΔΕΣΔΗΕ, με σημαντικές αρμοδιότητες για τη σύνδεση των νέων χρηστών, την καθημερινή λειτουργία και συντήρηση του Δικτύου, καθώς και για την ανάπτυξή του.

Βασικό χαρακτηριστικό του Ελληνικού συστήματος είναι η γεωγραφική ανισοκατανομή μεταξύ παραγωγής και φορτίων λόγω της μεγάλης συγκέντρωσης σταθμών παραγωγής και της ύπαρξης σημαντικών διεθνών διασυνδέσεων στο Βόρειο τμήμα της χώρας, ενώ το κέντρο κατανάλωσης είναι στο Νότιο τμήμα και ιδιαίτερα στην περιοχή της Αττικής.

Επίσης, η Πελοπόννησος δε συνδέεται στο βασικό κορμό διακίνησης ενέργειας στην ηπειρωτική χώρα (400 kV) (ΤΕΕ, 2010). Η γεωγραφική κατανομή των κυριότερων σταθμών παραγωγής ηλεκτρικής ενέργειας παρουσιάζονται στο ακόλουθο σχήμα (ΡΑΕ,2012):

Σχ. 1.2 - Σταθμοί Παραγωγής Ηλεκτρικής Ενέργειας



Όσον αφορά στις συνδέσεις της Ελληνικής Αγοράς με το εξωτερικό, αυτές βρίσκονται στο βόρειο τμήμα της χώρας. Πιο αναλυτικά, η Ελλάδα συνδέεται στα βόρεια μέσω εναέριων γραμμών μεταφοράς (AC) με την Αλβανία, το FYROM και τη Βουλγαρία, ενώ στα δυτικά συνδέεται με την Ιταλία μέσω καλωδίου (HVDC). Από το Σεπτέμβριο του 2010 εγκαινιάστηκε νέα εναέρια γραμμή μεταφοράς με την Τουρκία. Ιδιαιτερότητες παρουσιάζουν τα Ελληνικά νησιά του Αιγαίου, τα οποία στην πλειοψηφία τους δεν είναι συνδεδεμένα με την κυρίως χώρα, αλλά διαθέτουν αυτόνομους σταθμούς παραγωγής βασισμένους στο πετρέλαιο, σε συνδυασμό με ανεμογεννήτριες και άλλων μορφών ΑΠΕ (*Andrianesis, Biskas & Liberopoulos, 2011*).

Ο χάρτης του ελληνικού διασυνδεδεμένου συστήματος μεταφοράς ηλεκτρικής ενέργειας παρουσιάζεται στο ακόλουθο σχήμα:

Σχ. 1.3 - Χάρτης Διασυνδεδεμένου Συστήματος Μεταφοράς Η.Ε.

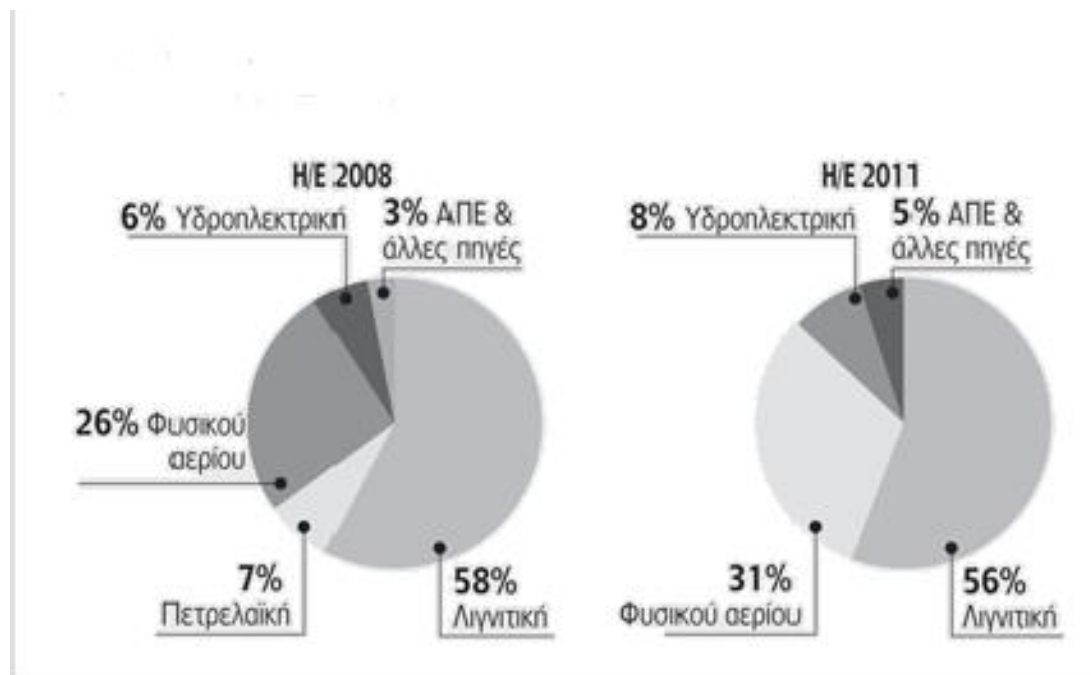


Πηγή: <http://www.investingreece.gov.gr>

Το καύσιμο που χρησιμοποιείται ευρέως στην Ελληνική Αγορά Ενέργειας είναι ο λιγνίτης. Πέρα από τις μονάδες που είναι βασισμένες στο λιγνίτη, συναντάμε και μονάδες αργού πετρελαίου, μονάδες συνδυασμένου κύκλου και ανοιχτού κύκλου με φυσικό αέριο, καθώς και μονάδες ΑΠΕ, όπως μεγάλες υδροηλεκτρικές μονάδες, αιολικά πάρκα και φωτοβολταϊκές εγκαταστάσεις, μικρές υδροηλεκτρικές μονάδες, μονάδες βιομάζας και γεωθερμικές μονάδες.

Η διαμόρφωση του ενεργειακού μίγματος της χώρας (ισοζύγιο παραγωγής Ηλεκτρικής ενέργειας) κατά την περίοδο 2008-2011 παρουσιάζεται στο γράφημα του ακόλουθου σχήματος:

Σχ. 1.4 - Ενεργειακό Μίγμα Ελληνικής Επικράτειας



Πηγή: <http://www.statbank.gr>

Κυρίαρχο ζητούμενο για την εύρυθμη λειτουργία της αγοράς ηλεκτρισμού στην Ελληνική επικράτεια είναι η επίτευξη των στόχων της ασφάλειας εφοδιασμού και της οικονομικής βελτιστοποίησης της αγοράς.

Αυτό επιτυγχάνεται μέσω κατάλληλου συνδυασμού μακροχρόνιων (εγκατάσταση και διαθεσιμότητα ισχύος) καθώς και βραχυχρόνιων αποφάσεων για την ορθή διαχείριση και κατανομή των πόρων στον Ημερήσιο Ενεργειακό Προγραμματισμό.

Σύμφωνα με το παραπάνω πλαίσιο (ΔΕΣΜΗΕ – Εγχειρίδιο Αγοράς, 2010) η Ελληνική Αγορά Ηλεκτρικής Ενέργειας διακρίνεται σε δύο θεμελιώδεις κατηγορίες οι οποίες είναι οι ακόλουθες:

- Η μακροχρόνια αγορά διαθεσιμότητας ισχύος (Capacity Market)
- τη Βραχυχρόνια (Χονδρεμπορική) Αγορά Ενέργειας και επικουρικών υπηρεσιών, που προγραμματίζεται βάσει του Ημερησίου Ενεργειακού Προγραμματισμού (ΗΕΠ) (Energy and Ancillary Services Market).

1.2.1 Αγορά Μακροχρόνιας Διαθεσιμότητας Ισχύος

Η λειτουργία αυτής της αγοράς έχει ως στόχο την όσο το δυνατόν μεγαλύτερη μείωση του επιχειρηματικού κινδύνου του παραγωγού, ο οποίος λαμβάνει αμοιβή έναντι μέρους του κόστους κεφαλαίου επένδυσης, καθώς και του προμηθευτή που επιζητά την αποφυγή ιδιαίτερα υψηλών τιμών στην Ημερήσια Αγορά Ενέργειας και Επικουρικών Υπηρεσιών (βραχυχρόνια αγορά), για το λόγο ότι μειώνεται ο βραχυχρόνιος κίνδυνος του παραγωγού.

Βασικός λόγος που οδήγησε στη δημιουργία της συγκεκριμένης αγοράς είναι η ανάγκη για εξασφάλιση της επάρκειας καθώς και της ποιότητας της ηλεκτρικής ενέργειας σε μακροχρόνια βάση, αμείβοντας παράλληλα την αξιοπιστία κάθε μονάδας.

Η υλοποίηση της Αγοράς Μακροχρόνιας Διαθεσιμότητας Ισχύος επιτυγχάνεται με την έκδοση από κάθε παραγωγό που συμμετέχει σε αυτήν Αποδεικτικών Διαθεσιμότητας Ισχύος (ΑΔΙ) που αντιστοιχούν στην πραγματική διαθεσιμότητα ισχύος κάθε μονάδας του, η οποία προσδιορίζεται από το Διαχειριστή του Συστήματος. Αν για παράδειγμα μία μονάδα εγκατεστημένης ισχύος έχει ισχύ 300 MW, είναι δυνατό μετά τον προσδιορισμό της διαθεσιμότητας της από τον Διαχειριστή του Συστήματος να εκδώσει ΑΔΙ για 250MW.

Όσον αφορά στους προμηθευτές, κάθε ένας από αυτούς συνάπτει με τους παραγωγούς Συμβάσεις Διαθεσιμότητας Ισχύος (ΣΔΙ), με οικονομικούς όρους που συμφωνούνται και από τις δύο πλευρές έτσι ώστε να καλύψουν την Υποχρέωση Επάρκειας Ισχύος που τους αναλογεί.

Η αγορά μπορεί να χαρακτηριστεί ως εξισορροπημένη (Balanced) (ΔΕΣΜΗΕ – Εγχειρίδιο Αγοράς, 2010) όταν ισχύει η ακόλουθη σχέση:

$$\text{ΑΔΙ (MW)} = \text{ΣΔΙ (MW)} + \text{Απαιτούμενη μακροχρόνια εφεδρεία}$$

1.2.2 Ημερήσιος Ενεργειακός Προγραμματισμός (Αγορά Ενέργειας και Επικουρικών Υπηρεσιών ή Χονδρεμπορική Αγορά)

Ο Ημερήσιος Ενεργειακός Προγραμματισμός (ΗΕΠ) αποτελεί τη χονδρεμπορική αγορά και έχει ως κυρίαρχο στόχο το βέλτιστο προγραμματισμό της λειτουργίας των θερμικών και υδροηλεκτρικών μονάδων παραγωγής του Συστήματος, όπως και των μονάδων ΑΠΕ και της διαθέσιμης ενέργειας που προέρχεται από εισαγωγές, ώστε να καλύπτεται σε ημερήσια βάση η ζήτηση ηλεκτρικής ενέργειας από καταναλωτές, η ζήτηση για εξαγωγές ενέργειας και οι απαραίτητες Επικουρικές Υπηρεσίες. Οι μονάδες παραγωγής που συμμετέχουν στον ΗΕΠ οφείλουν να προσφέρουν το σύνολο της διαθεσιμότητάς τους, τόσο σε ενέργεια όσο και σε επικουρικές υπηρεσίες. Ουσιαστικά πρόκειται για ένα μοντέλο αγοράς Υποχρεωτικής Κοινοπραξίας (Mandatory Pool).

Για να γίνει κατανοητός ο όρος «Υποχρεωτική Κοινοπραξία» απαιτείται η παράθεση του ορισμού του:

«Υποχρεωτική Κοινοπραξία» είναι το μοντέλο της Ελληνικής Αγοράς στο οποίο συναλλάσσεται το σύνολο της ηλεκτρικής ενέργειας και των συμπληρωματικών προϊόντων αυτής που θα παραχθούν, θα καταναλωθούν και θα διακινηθούν την επόμενη ημέρα στην αγορά. Όλοι οι συμμετέχοντες στην Ελληνική Αγορά υποχρεούνται να συμμετέχουν στην Υποχρεωτική Κοινοπραξία. Δεν επιτρέπονται φυσικές διμερές συναλλαγές (physical bilateral transactions) μεταξύ των συμμετεχόντων της αγοράς (ΔΕΣΜΗΕ – Εγχειρίδιο Αγοράς έκδοση 2.0,2010).

Στα αυστηρά πλαίσια του παραπάνω μοντέλου κάθε διαμετακόμιση (transit) ηλεκτρικής ενέργειας από μία περιοχή εκτός Ελλάδος σε μία άλλη, μέσω της Ελληνικής επικράτειας, μπορεί να υλοποιηθεί μόνο μέσω εισαγωγής (πώλησης) της εν λόγω ενέργειας στην Υποχρεωτική Κοινοπραξία και εξαγωγής (αγοράς) από αυτήν.

Σημαντικές επιμέρους αγορές-μηχανισμοί οι οποίες υπάγονται στον ΗΕΠ είναι οι ακόλουθες:

- Αγορά Ενέργειας: Ικανοποιούνται οι ποσοτικές ανάγκες των καταναλωτών ηλεκτρικής ενέργειας κατά τις ώρες που αυτή είναι απαραίτητη. Προσφέρουν αλλά και αμείβονται οι εγχώριοι παραγωγοί (θερμικών – υδροηλεκτρικών σταθμών και ΑΠΕ), καθώς και οι εισαγωγείς ενέργειας ενώ αγοράζουν οι εκπρόσωποι του εγχώριου φορτίου (προμηθευτές και επιλεγμένοι πελάτες) και οι εξαγωγείς (προμηθευτές και παραγωγοί).
- Αγορά Επικουρικών Υπηρεσιών (Ancillary Services): Ικανοποιεί τις ανάγκες των καταναλωτών ηλεκτρικής ενέργειας για διασφάλιση της ποιότητας και αξιοπιστίας της τροφοδότησής τους. Παράλληλα προσφέρουν και αμείβονται οι εγχώριοι παραγωγοί (θερμικών – υδροηλεκτρικών σταθμών) και αγοράζουν οι εκπρόσωποι του εγχώριου φορτίου (προμηθευτές και επιλεγμένοι πελάτες), καθώς και οι εξαγωγείς (προμηθευτές και παραγωγοί).
- Μηχανισμός αγοράς για τη χωροθέτηση της παραγωγής κοντά στα κέντρα κατανάλωσης: Σχετίζεται με την παροχή κινήτρων χωροθέτησης των νέων μονάδων όσο τον δυνατόν πιο κοντά στην κατανάλωση. Με δεδομένη την παρούσα συγκέντρωση της ηλεκτρικής παραγωγής στο βορρά, είναι ασφαλές να πούμε ότι αναμένεται αυξημένη συμμετοχή στην ετήσια χρέωση χρήσεως συστήματος των παραγωγών του βορρά, όπως και διαφοροποιημένη αμοιβή παραγωγών που βρίσκονται στο νότιο τμήμα της χώρας, κυρίως σε περιπτώσεις που εμφανίζεται συμφόρηση στη μεταφορά ηλεκτρικής ενέργειας από το βορρά στο νότο κατά τη διάρκεια του ημερήσιου προγραμματισμού.

1.3 Περιγραφή και Επίλυση του ΗΕΠ

Η επίλυση του ΗΕΠ αφορά κυρίως τον προσδιορισμό του τρόπου λειτουργίας κάθε μονάδας για κάθε ώρα της επόμενης ημέρας, με απώτερο στόχο τη μεγιστοποίηση του κοινωνικού οφέλους που προκύπτει από την ικανοποίηση του ενεργειακού ισοζυγίου της αγοράς αλλά και των αναγκών των επικουρικών υπηρεσιών την επόμενη μέρα λαμβάνοντας πάντα υπόψη τους περιορισμούς του συστήματος μεταφοράς.

Ενσωματώνοντας στον ΗΕΠ τις απαραίτητες Επικουρικές Υπηρεσίες και τους τεχνικούς περιορισμούς του συστήματος, οι οποίοι πιθανότατα περιορίζουν την ποσότητα ενέργειας που είναι δυνατόν να διακινείται από το βορρά στο νότο, προκύπτει η πραγματική συνολική αξία της καταναλισκόμενης ηλεκτρικής ενέργειας, έχοντας λάβει υπόψη τα ποσοτικά (όγκος και χρόνος κατανάλωσης) αλλά και τα ποιοτικά χαρακτηριστικά της (συχνότητα, τάση, απαιτούμενη αξιοπιστία παροχής σε κάθε καταναλωτή).

Από την επίλυση του ΗΕΠ προκύπτει η **Οριακή Τιμή Συστήματος (ΟΤΣ)**, η οποία αποτελεί το αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας και θα αναλυθεί εκτενέστερα στη συνέχεια.

Όπως έχει αναφερθεί και προηγουμένως, ο Ημερήσιος Ενεργειακός Προγραμματισμός (ΗΕΠ) είναι η ημερήσια διαδικασία επίλυσης της αγοράς ενέργειας η οποία πραγματοποιείται την προηγούμενη ημέρα της Ημέρας Κατανομής. Ημέρα Κατανομής θεωρείται μία ημερολογιακή ημέρα σε ώρες Ελλάδος. Αποτελείται από 24 ωριαίες περιόδους κατανομής, εκτός φυσικά από τις «μικρές» και «μεγάλες» ημέρες που παρουσιάζονται το φθινόπωρο και την άνοιξη αντίστοιχα και οφείλονται στην αλλαγή της

τοπικής ώρας. Επομένως η Μικρή Ημέρα Κατανομής αποτελείται από 23 περιόδους κατανομής (απουσιάζει η ώρα από τις 3:00 ως τις 4:00), και η Μεγάλη Ημέρα Κατανομής αποτελείται από 25 περιόδους Κατανομής (εμφανίζεται 2 φορές η ώρα από τις 3:00 ως τις 4:00).

1.3.1 Συμμετοχή στον ΗΕΠ

Όλοι όσοι συμμετέχουν στον ΗΕΠ, αλλά και στις άλλες αγορές που διαχειρίζεται ο Διαχειριστής του Συστήματος, είναι απαραίτητο να εγγραφούν στο Μητρώο Συμμετεχόντων του Διαχειριστή. Μέσω αυτής της εγγραφής οι συμμετέχοντες συνάπτουν με το Διαχειριστή του Συστήματος «Σύμβαση Συναλλαγών Ηλεκτρικής Ενέργειας» η οποία διέπεται από τις διατάξεις του ΚΔΣ & ΣΗΕ (Κώδικας Διαχείρισης Συστήματος και Συναλλαγών Ηλεκτρικής Ενέργειας).

Η Σύμβαση Συναλλαγών Ηλεκτρικής Ενέργειας αφορά τις συναλλαγές που πραγματοποιούνται στα πλαίσια του συστήματος Συναλλαγών Ηλεκτρικής Ενέργειας, το οποίο περιλαμβάνει τον ΗΕΠ, τη Διαδικασία Κατανομής, τις Εντολές Κατανομής, την Εκκαθάριση Αποκλίσεων και τον Μηχανισμό Διασφάλισης Επαρκούς Ισχύος. Επιπλέον, οι συμμετέχοντες έχουν το δικαίωμα να προβαίνουν σε κάθε σχετική νόμιμη ενέργεια συμμετοχής στο Σύστημα Συναλλαγών Ηλεκτρικής Ενέργειας, τηρώντας πάντα τις διατάξεις του Κώδικα, και παράλληλα μπορούν να λαμβάνουν και να εξοφλούν τις πληρωμές και τις χρεώσεις που τους αναλογούν. Τα δικαιώματα και οι υποχρεώσεις του Διαχειριστή του Συστήματος ορίζονται από τις διατάξεις του Κώδικα για το Σύστημα Συναλλαγών Ηλεκτρικής Ενέργειας.

Πιο αναλυτικά, η συμμετοχή στον ΗΕΠ περιλαμβάνει τις ακόλουθες διαδικασίες (*ΔΕΣΜΗΕ – Εγχειρίδιο Αγοράς έκδοση 2.0, 2010*):

- Υποβολή μη τιμολογούμενων Δηλώσεων Φορτίου για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από τους Εκπροσώπους Φορτίου για τους Πελάτες που καταναλώνουν ενέργεια εντός της Ελληνικής Επικράτειας.
- Υποβολή τιμολογούμενων Δηλώσεων Φορτίου για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από τους Συμμετέχοντες, όταν αυτοί είναι είτε κάτοχοι Φυσικών Δικαιωμάτων Μεταφοράς (ΦΔΜ) είτε αντισυμβαλλόμενοι κατόχων ΦΔΜ, για εξαγωγές ενέργειας.
- Υποβολή τιμολογούμενων Δηλώσεων Φορτίου για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από Παραγωγούς για αντλητικές μονάδες σε λειτουργία άντλησης.
- Υποβολή μη τιμολογούμενων Δηλώσεων Φορτίου για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από τους Παραγωγούς και τους Προμηθευτές για τα βοηθητικά φορτία των μονάδων όταν αυτά δεν καλύπτονται από την παραγωγή των μονάδων.
- Υποβολή τιμολογούμενων Προσφορών Έγχυσης για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από Παραγωγούς για Κατανεμημένες Μονάδες.
- Υποβολή τιμολογούμενων Προσφορών Έγχυσης για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από το Διαχειριστή του Συστήματος για τις Συμβεβλημένες Μονάδες. Οι τιμολογούμενες προσφορές ενέργειας για τις Συμβεβλημένες Μονάδες πρέπει να είναι σύμφωνες με τις τιμές της σχετικής σύμβασης.
- Υποβολή τιμολογούμενων Προσφορών Έγχυσης για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από τους συμμετέχοντες, όταν αυτοί είναι είτε κάτοχοι Φυσικών Δικαιωμάτων Μεταφοράς (ΦΔΜ) είτε αντισυμβαλλόμενοι κατόχων ΦΔΜ, για εισαγωγές ενέργειας.

- Υποβολή μη τιμολογούμενων Προσφορών Έγχυσης για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από το Διαχειριστή του Συστήματος για τις μονάδες του Άρθρου 35 του Νόμου 2773/1999[2].
- Υποβολή Δηλώσεων Διαχείρισης Υδάτινων Πόρων για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από τους Παραγωγούς για τις Υδροηλεκτρικές Μονάδες, συμπεριλαμβανομένων των Υδροηλεκτρικών Μονάδων Αντλησης.
- Υποβολή τιμολογούμενων Προσφορών Εφεδρειών για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από Παραγωγούς για τις Κατανεμημένες Μονάδες.
- Υποβολή τιμολογούμενων Προσφορών Εφεδρειών για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από το Διαχειριστή τους Συστήματος για τις Συμβεβλημένες Μονάδες. Οι τιμολογούμενες προσφορές εφεδρειών για τις Συμβεβλημένες Μονάδες πρέπει να είναι σύμφωνες με τις τιμές της σχετικής σύμβασης.
- Υποβολή Δηλώσεων Ολική ή Μερικής Μη Διαθεσιμότητας για κάθε Περίοδο Κατανομής της Ημέρας Κατανομής από τους Παραγωγούς για κάθε μονάδα.
- Υποβολή δηλώσεων Τεχνοοικονομικών Στοιχείων από τους Παραγωγούς για τις μονάδες.

1.4 Οριακή Τιμή Συστήματος (ΟΤΣ)

Η Οριακή Τιμή του Συστήματος (ΟΤΣ) είναι η τιμή στην οποία εκκαθαρίζεται η αγορά ηλεκτρικής ενέργειας και είναι η τιμή που εισπράττουν όλοι όσοι εγχέουν ενέργεια στο Σύστημα και πληρώνουν όλοι όσοι ζητούν ενέργεια από το Σύστημα. Συγκεκριμένα, η Οριακή Τιμή του Συστήματος διαμορφώνεται από τον συνδυασμό των προσφορών τιμών και ποσοτήτων που υποβάλλουν κάθε μέρα οι διαθέσιμες μονάδες παραγωγής ηλεκτρικής ενέργειας, και του ωριαίου φορτίου ζήτησης ηλεκτρικής ενέργειας, που διαμορφώνεται σε καθημερινή βάση από τους καταναλωτές.

Επιχειρώντας μια απλή περιγραφή του τρόπου υπολογισμού της Οριακής Τιμής του Συστήματος, σύμφωνα με τις βασικές αρχές της μικροοικονομικής θεωρίας, μπορεί να αναφερθεί ότι οι μονάδες παραγωγής κατατάσσονται αναλόγως των προσφορών τους σε αύξουσα σειρά, ξεκινώντας από την χαμηλότερη προσφερόμενη τιμή για ορισμένη ποσότητα ενέργειας και καταλήγοντας στην υψηλότερη προσφερόμενη τιμή.

Στο σημείο όπου οι προσφερόμενες ποσότητες ενέργειας εξυπηρετούν το ζητούμενο φορτίο, καθορίζεται και η Οριακή Τιμή του Συστήματος. Στην ουσία, η Οριακή τιμή του Συστήματος συμπίπτει με την προσφορά της τελευταίας μονάδας που πρέπει να λειτουργήσει για να καλυφθεί η ζήτηση. Για λόγους προστασίας των καταναλωτών και διαμόρφωσης συνθηκών υγιούς ανταγωνισμού τίθεται διοικητικά ανώτερο όριο ως προς την προσφερόμενη τιμή, το οποίο έχει τεθεί ίσο με 150€/MWh καθώς και κατώτερο επίπεδο προσφορών, το οποίο είναι το μεταβλητό κόστος της μονάδας, ώστε στις περισσότερες περιπτώσεις οι παραγωγοί να πληρώνονται το κόστος καυσίμου τους (ΠΑΕ, 2010).

Η Οριακή Τιμή Συστήματος, πέρα από το γεγονός ότι συνιστά την τιμή εκκαθάρισης των συναλλαγών σε ημερήσια βάση, επιτρέπει την ανάπτυξη καμπυλών αναφοράς οι οποίες αποτελούν σημείο αναφοράς ως προς τη χρηματοοικονομική διαχείριση της αγοράς ηλεκτρικής ενέργειας και την υποκείμενη αξία για επιπλέον προϊόντα, όπως είναι τα προϊόντα αντιστάθμισης, τα προθεσμιακά προϊόντα, τα παράγωγα κ.α., που μπορούν να εξυπηρετούν συγκεκριμένες ανάγκες των συμμετεχόντων (ΔΕΣΜΗΕ – Εγχειρίδιο Αγοράς, 2010).

Η ΟΤΣ υπολογίζεται από τη ΔΕΣΜΗΕ με τη βοήθεια της Οριακής τιμής Παραγωγής (ΟΤΠ). Η ΟΤΠ κάθε λειτουργικής ζώνης ισούται με το σκιαδές κόστος (πολλαπλασιαστής Lagrange) του περιορισμού του ισοζυγίου ισχύος της ζώνης. Στην περίπτωση που οι διαζωνικοί περιορισμοί ροής δεν είναι δεσμευτικοί, οι δύο δυικές τιμές ταυτίζονται και η ΟΤΣ ισούται με αυτή την κοινή τιμή. Στην περίπτωση όμως που οι διαζωνικοί περιορισμοί είναι δεσμευτικοί, οι δύο δυικές τιμές μπορεί να είναι

διαφορετικές και τότε η ΟΤΣ υπολογίζεται ως ο σταθμισμένος (με την παραγωγή των ζωνών) μέσος όρος των οριακών τιμών παραγωγής των ζωνών. Σε κάθε περίπτωση η ΟΤΠ κάθε λειτουργικής ζώνης καθορίζεται από μία ή περισσότερες οντότητες εντός της εκάστοτε λειτουργικής ζώνης, οι οποίες μπορεί να είναι μονάδες παραγωγής, κατανεμημένα φορτία, εισαγωγές ή εξαγωγές.

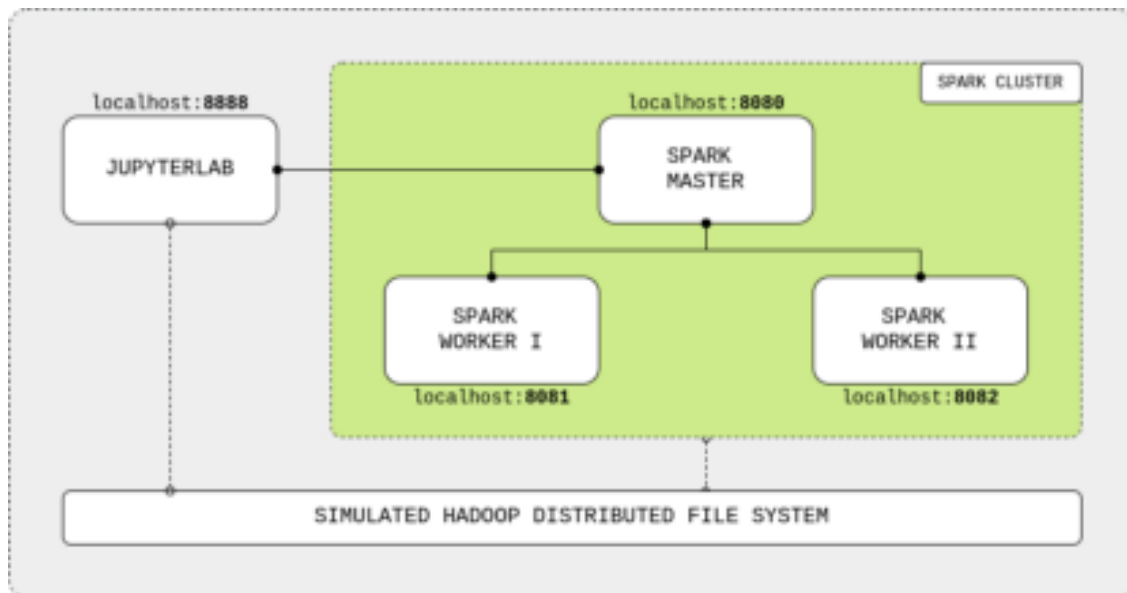
Σε αυτό το σημείο είναι σημαντικό να αναφερθεί πως η επίλυση του ΗΕΠ από τη ΔΕΣΜΗΕ λαμβάνει χώρα από τις 12:30 έως τις 14:00 της προηγούμενης της Ημέρας Κατανομής, και η δημοσίευση των αποτελεσμάτων του ΗΕΠ, συνεπώς και της ΟΤΣ, γίνεται στις 14:00 της προηγούμενης της Ημέρας Κατανομής.

Κεφάλαιο 2

2.1 Υποδομή

Για τη υλοποίηση της παρούσας εργασίας, έγινε προσπάθεια να αναλυθούν τα δεδομένα της ΡΑΕ σε ψευδο-κατανεμημένο περιβάλλον. Με τον όρο «ψευδο-κατανεμημένο», λογίζεται η αρχιτεκτονική όπου στήνεται μία τοπική συστάδα (cluster) από 2 workers που θα φορτώνουν τα δεδομένα κατανεμημένα στη μνήμη τους και έναν master πυρήνα που αναλαμβάνει την ενορχήστρωση των παραπάνω. Επιπρόσθετα, και οι εργασίες πάνω σε αυτά δεδομένα θα γίνονται με τρόπο κατανεμημένο προκειμένου να επιτευχθεί η βέλτιστη απόδοση. Ο πίνακας που ακολουθεί προσφέρει μια σχηματική απεικόνιση της διάταξης :

Σχ. 2.1



Εκτιμήθηκε επίσης για λόγους ευχρηστίας, μία λύση plug-n-play όπου όλα τα παραπάνω θα τρέχουν σε Dockerized περιβάλλον (ένας container ανά worker, ένας για το master και ένας για το περιβάλλον υλοποίησης που είναι το JupyterLab). Αξίζει να σημειωθεί ότι, ο κοινός χώρος αποθήκευσης (SIMULATED HADOOP DISTRIBUTED FILESYSTEM) είναι συνδεδεμένος με το master Node και με το JupyterLab. Πρόκειται για ένα κοινό workspace που προσομοιώνει τη λειτουργικότητα του HDFS storage ενός πραγματικά κατανεμημένου Hadoop cluster και παρέχει τη δυνατότητα πραγματικά παράλληλης επεξεργασίας δεδομένων από το cores της μηχανής υλοποίησης της εφαρμογής.

2.2 Σύντομη Περιγραφή των τεχνολογιών

Όπως προαναφέρθηκε, η εφαρμογή είναι υλοποιημένη σε Dockerized περιβάλλον και χρησιμοποιεί κάποια Docker images. Βασικό image αποτελεί το Spark (έκδοση 3.0.0), μια ενιαία μηχανή ανάλυσης ανοικτού κώδικα για επεξεργασία δεδομένων μεγάλης κλίμακας. Το Spark παρέχει μια διεπαφή για τον προγραμματισμό ολόκληρων ομάδων με έμμεσο παραλληλισμό δεδομένων και ανοχή σφαλμάτων. Για το image του Spark, θα ρυθμιστεί σε standalone mode, την απλούστερη διαμόρφωση ανάπτυξης. Σε αυτήν τη λειτουργία, θα χρησιμοποιηθεί ο διαχειριστής πόρων για να ρυθμίσουμε τα containers ώστε να εκτελούνται είτε ως master είτε ως worker. Αντίθετα, οι διαχειριστές πόρων όπως το Apache YARN εκχωρούν δυναμικά containers ως κόμβους master ή worker ανάλογα με το φόρτο εργασίας του χρήστη.

Επιπλέον, η συγκεκριμένη έκδοση του Spark υποστηρίζεται με Apache Hadoop (έκδοση 2.7) για να επιτραπεί στο cluster να προσομοιώσει τον HDFS χρησιμοποιώντας τον κοινόχρηστο χώρο αποθήκευσης. Σημειώνεται ότι χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python (έκδοση 3.6) για την υλοποίηση της ανάλυσης με πληθώρα βιβλιοθηκών που εξυπηρετούν ανάλυση, επεξεργασία αλλά και πρόβλεψη δεδομένων όπως οι Pyspark, statsmodels.api, scipy, sklearn, pandas, numpy και άλλες.

Επόμενο βήμα αποτελεί η παρουσίαση ενός δείγματος δεδομένων, που χρησιμοποιήθηκαν. Αρχικά, τα δεδομένα ήταν σε μορφή excel file στον ιστότοπο <https://www.enexgroup.gr/en/market/electricity-day-ahead-market/results/> και αφορούσαν την περίοδο 01/01/2018 – 01/11/2020. Η δομή του κάθε αρχείου είναι:

Σχ.2.2


Solved with Load Declarations																											
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL	STDV P	
2	4.140	3.890	3.770	3.726	3.711	3.814	4.197	4.739	5.189	5.541	5.733	5.831	5.900	5.877	5.698	5.446	5.293	5.284	5.512	5.965	5.995	5.667	5.117	4.720	120.754		
4	1.669	1.590	1.511	1.497	1.475	1.477	1.542	1.588	1.813	1.961	2.027	2.107	2.221	2.213	2.185	2.091	2.047	1.997	2.060	2.192	2.275	2.129	1.963	1.815	45.444		
5	2.432	2.270	2.230	2.202	2.209	2.308	2.623	3.092	3.318	3.524	3.657	3.690	3.650	3.635	3.488	3.329	3.208	3.241	3.402	3.697	3.641	3.476	3.112	2.863	74.295		
6	38	30	29	28	28	28	32	58	56	48	34	29	29	25	27	38	46	50	76	80	61	43	42	1.016			
7	39,747	41,047	41,046	41,046	41,046	46,847	67,140	67,423	105,010	100,030	74,000	53,080	51,150	56,500	60,000	65,000	66,500	60,000	56,000	60,000	67,544	68,000	69,570	47,739	47,783	58,885	16,4989
8	39,747	41,047	41,046	41,046	41,046	46,847	67,140	67,423	105,010	100,030	74,000	53,080	51,150	56,500	60,000	65,000	66,500	60,000	56,000	60,000	67,544	68,000	69,570	47,739	47,783	58,885	
9	39,747	41,047	41,046	41,046	41,046	46,847	67,140	67,423	105,010	100,030	74,000	53,080	51,150	56,500	60,000	65,000	66,500	60,000	56,000	60,000	67,544	68,000	69,570	47,739	47,783	58,885	
10	20	84	70	20	0	0	30	38	68	108	108	103	103	83	35	0	0	0	228	843	963	392	90	50	3.436		
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	353,55	339,13	323,13	311,7	299,35	290	287,53	296,83	519,77	996,89	1465,8	1794,3	1978,4	2050,2	2003,7	1877,1	1640,1	1318,5	1091,2	1091,8	1169,9	1254,3	1343,1	1422	25.518		
13	170,81	173,37	173,98	175,51	169,55	163,46	161,19	161,73	179,41	223,94	275,47	343,96	393,04	443,47	475,22	489,91	508,44	531,79	546,09	581,29	615,6	645,33	696,6	695,4	8.995		
14	524,36	512,5	497,11	487,21	468,9	453,46	448,72	458,56	699,18	1220,8	1741,3	2138,3	2371,5	2493,7	2478,9	2367	2148,6	1850,3	1637,3	1673,1	1785,5	1899,6	2039,7	2118	34.513		
15	-1495	-1068	-1082	-1069	-1044	-1006	-1090	-1075	-880	-742	-682,7	-646,6	-616,6	-588	-478,5	-518,4	-421,9	-540,8	-469,9	-248	19,108	-205,1	-429	-497	-16.874		
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	-1495	-1068	-1082	-1069	-1044	-1006	-1090	-1075	-880	-742	-682,7	-646,6	-616,6	-588	-478,5	-518,4	-421,9	-540,8	-469,9	-248	19,108	-205,1	-429	-497	-16.874		
18	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL		
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	150	150	150	150	150	150	150	150	223	223	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150	3.746	
21	155	155	155	155	155	155	155	155	241	241	155	155	155	155	155	155	155	155	155	155	155	155	155	155	155	3.892	
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
32	151	151	151	151	151	151	164,08	213,29	249	249	249	151	151	151	151	151	151	151	151	151	191,66	249	151	151	151	4.132	
33	182	410	410	414	417	417	414	410	405	400	395	395	392	392	392	392	392	395	395	400	400	405	410	410	410	9.429	

Από την παραπάνω πληροφορία στα πλαίσια της παρούσας ανάλυσης ενδιαφέρει η γραμμή 7 που αναπαριστά τις ωριαίες Οριακές Τιμές Συστήματος (η φυσική σημασία του μεγέθους θα αναλυθεί εκτενώς σε ακόλουθο κεφάλαιο) για κάθε μέρα 24 στήλες και τη μέση τιμή τους.

Συνεπώς γίνεται εξόρυξη αυτής πληροφορίας από κάθε αρχείο και αποθηκεύεται σε ένα Spark Dataframe. Πιο συγκεκριμένα, δημιουργούνται 3 Spark Dataframe, ένα ανά έτος, όπου αποθηκεύεται σαν πληροφορία οι ωριαίες τιμές, οι μέσες και το αντίστοιχο timestamp.

Παράδειγμα δεδομένων, φαίνεται στον ακόλουθο πίνακα :

Σχ. 2.3



SMPs	Daily_Average_SMP	date
57.883	52.85216666666667	2018-01-01 00:00:00
58.088	52.85216666666667	2018-01-01 01:00:00
57.959	52.85216666666667	2018-01-01 02:00:00
54.575	52.85216666666667	2018-01-01 03:00:00
54.47	52.85216666666667	2018-01-01 04:00:00
54.848	52.85216666666667	2018-01-01 05:00:00
54.852	52.85216666666667	2018-01-01 06:00:00
53.783	52.85216666666667	2018-01-01 07:00:00
52.57	52.85216666666667	2018-01-01 08:00:00
42.8	52.85216666666667	2018-01-01 09:00:00
50.178	52.85216666666667	2018-01-01 10:00:00
50.646	52.85216666666667	2018-01-01 11:00:00
50.642	52.85216666666667	2018-01-01 12:00:00
52.834	52.85216666666667	2018-01-01 13:00:00
51.169	52.85216666666667	2018-01-01 14:00:00
52.877	52.85216666666667	2018-01-01 15:00:00
51.871	52.85216666666667	2018-01-01 16:00:00
52.932	52.85216666666667	2018-01-01 17:00:00
51.265	52.85216666666667	2018-01-01 18:00:00
51.289	52.85216666666667	2018-01-01 19:00:00

Κεφάλαιο 3

3.1 Πρόβλεψη Οριακής Τιμής Συστήματος (ΟΤΣ)

Οι δύο πιο σημαντικοί παράγοντες της Αγοράς Ηλεκτρικής Ενέργειας είναι οι ακόλουθοι:

- ΟΤΣ (Οριακή Τιμή Συστήματος) ή αλλιώς SMP (System Marginal Price) ή MCP (Market Clearing Price)
- ΟΤΦ (Οριακή Τιμή Φορτίου) ή MCQ (Market Clearing Quantity). Στα πλαίσια της παρούσας διπλωματικής θα επιχειρήσουμε να ασχοληθούμε μόνο με την πρόβλεψη της ΟΤΣ.

Σε γενικές γραμμές, η όσο το δυνατόν ακριβέστερη πρόβλεψη της ΟΤΣ αποτελεί ένα αρκετά δύσκολο εγχείρημα με το οποίο έχουν ασχοληθεί αρκετοί ερευνητές τα τελευταία χρόνια. Πρόκειται για ένα μέγεθος ιδιαίτερα ευμετάβλητο και ασταθές, γεγονός που καθιστά τη διαδικασία πρόβλεψής της εξαιρετικά δύσκολη.

Εύλογα προκύπτει η απορία,για ποιο λόγο είναι ένα μέγεθος με τόσο ενδιαφέρον. Για να γίνει κατανοητή η χρησιμότητα – αναγκαιότητα της πρόβλεψης είναι απαραίτητο να διαχωριστεί πρόβλεψη σε δύο μεγάλες κατηγορίες με βάση τον χρονικό ορίζοντα:

- την βραχυπρόθεσμη (ημέρα – εβδομάδα), και
- τη μακροπρόθεσμη (μήνας – έτος).

Ενίοτε θα μπορούσε να συμπεριληφθεί και η μεσοπρόθεσμη (μήνας), αλλά χωρίς βλάβη της γενικότητας, εντάσσεται στην μακροπρόθεσμη.

3.1.1 Βραχυπρόθεσμη Πρόβλεψη

Οι κυρίαρχοι λόγοι που είναι χρήσιμη η βραχυπρόθεσμη πρόβλεψη (ημέρα έως εβδομάδα) της ΟΤΣ είναι οι ακόλουθοι:

- Βοηθάει στη διαδικασία αγοραπωλησίας ενέργειας και στη υιοθέτηση κατάλληλης στρατηγικής από μία εταιρεία που συμμετέχει στην αγορά ενέργειας με στόχο την ελαχιστοποίηση του κόστους
- Πραγματοποιούνται ευμενέστερες διμερείς συμφωνίες.
- Προγραμματίζεται δέσμευση των παραγωγικών μονάδων με βάση τις προβλεπόμενες τιμές και προσφορά της αναμενόμενης παραγόμενης ενέργειας.

3.1.2 Μακροπρόθεσμη Πρόβλεψη

Η μακροπρόθεσμη πρόβλεψη (μήνας έως έτος) της ΟΤΣ είναι χρήσιμη για τους ακόλουθους λόγους:

- Προγραμματισμός επενδύσεων στη δυναμικότητα ηλεκτροπαραγωγής (generation capacity)
- Ανάλυση ρίσκου αγοράς στο εγγύς η απώτερο μέλλον
- Προγραμματισμός παραγωγής και συντήρησης
- Σύναψη συμβολαίων για καύσιμα
- Τιμολόγηση Παράγωγων.

Περισσότερο κρίσιμη για τους συμμετέχοντες στην αγορά ενέργειας μπορεί να χαρακτηριστεί η βραχυπρόθεσμη πρόβλεψη, καθώς είναι αυτή που μπορεί να τους επιφέρει άμεσα κέρδη ή ζημίες και να καθορίσει την οργάνωση της παραγωγής, στην περίπτωση των παραγωγών, ή αντίστοιχα την κατανάλωση ενέργειας στην περίπτωση των καταναλωτών, στο άμεσο μέλλον.

3.2 Βασικά Χαρακτηριστικά ΟΤΣ

Το μεγαλύτερο πρόβλημα στην πρόβλεψη της ΟΤΣ είναι η μεταβλητότητα, δηλαδή το μέτρο της μεταβολής της τιμής της ηλεκτρικής ενέργειας σε ένα δεδομένο χρονικό διάστημα (μη ομογενής καμπύλη σε σύγκριση με το φορτίο) (Deepak & Swarup, 2011). Υπάρχουν αρκετοί παράγοντες που επηρεάζουν την μεταβλητότητα της ΟΤΣ. Οι βασικότεροι από αυτούς είναι:

- Συμφόρηση κατά τη μεταφορά
- Βασικός λόγος για τα spikes στην τιμή είναι ότι η ζήτηση και η προσφορά πρέπει να συνδέονται με μια δευτερόλεπτο προς δευτερόλεπτο σχέση, κάτι το οποίο δεν συμβαίνει
- Μεταβλητότητα στην τιμή των καυσίμων
- Αβεβαιότητα φορτίου
- Διακυμάνσεις στην υδροηλεκτρική παραγωγή
- Αβεβαιότητα παραγωγής (διακοπές)
- Συμπεριφορά των συμμετεχόντων στην αγορά ενέργειας (βασισμένη στην αναμενόμενη τιμή)
- Καθοδήγηση της αγοράς.

Εξετάζοντας τώρα τα ιδιαίτερα χαρακτηριστικά της ΟΤΣ, είναι πολύ εύκολο να διακρίνουμε το λόγο που το συγκεκριμένο μέγεθος παρουσιάζει υψηλή μεταβλητότητα. Ανάμεσα στις ιδιαιτερότητες της ΟΤΣ μπορούμε να ξεχωρίσουμε τα ακόλουθα 5 στοιχεία (Sanjeev et al., 2009):

- Υψηλή συχνότητα
- Μη σταθερή μέση τιμή και διασπορά
- Πολλαπλή εποχικότητα
- Επηρεάζεται από το ημερολόγιο (calendar effect)
- Υψηλό επίπεδο αστάθειας και υψηλό ποσοστό ασυνήθιστων τιμών (Spikes).

Αναζητώντας του λόγους για τους οποίους η ΟΤΣ παρουσιάζει τα παραπάνω χαρακτηριστικά, μπορούμε να αναφέρουμε ότι αυτά προκύπτουν από το γεγονός ότι:

- Η ενέργεια δεν αποθηκεύεται
- Απαιτείται να υπάρχει ισορροπία ζήτησης – προσφοράς

- Η ζήτηση είναι ανελαστική μέσα σε μικρό χρονικό διάστημα
- Η αγορά λειτουργεί υπό συνθήκες ολιγοπωλίου
- Η ζήτηση και η προσφορά επηρεάζουν την ισορροπία της αγοράς

3.3 Βασικές Μεθοδολογίες Πρόβλεψης

Σύμφωνα με τους Sanjeev, Lalit και Ashwani (2009), υπάρχουν 3 βασικές κατηγορίες μοντέλων πρόβλεψης που χρησιμοποιούνται για την παραγωγή προβλέψεων της ΟΤΣ. Οι κατηγορίες αυτές είναι οι ακόλουθες:

- Game Theory Models (Nash equilibrium, Cournot, Bertrand)
- Simulation Models (MAPS, UPLAN)
- Time Series Models - Stochastic models (ARIMA, AR, MA, GARCH), AI models, Regression/causal models.

Στα πλαίσια της παρούσας διπλωματικής θα εξεταστούν κυρίως τα μοντέλα χρονοσειρών (TimeSeries Models), τα οποία απαιτούν την έκφραση των υπό μελέτη μεγεθών με τη μορφή χρονοσειρών. Η κατηγορία αυτή μπορεί να διαχωριστεί περαιτέρω με τον ακόλουθο τρόπο (Πετρόπουλος, Ασημακόπουλος, 2011):

- Μοντέλα Χρονοσειρών: εξετάζουν τη χρονοσειρά από αριθμητική σκοπιά ως σύνολο τιμών δίνοντας μικρή έμφαση στο τι εκφράζει. Περιλαμβάνουν την αποσύνθεση (decomposition), την εξομάλυνση (smoothing) και τις αυτοπαλινδρομικές μεθόδους κινητού μέσου όρου (Autoregressive Moving Average ή ARMA).
- Αιτιοκρατικά Μοντέλα: εστιάζουν στις σχέσεις αίτιου-αποτελέσματος με συνέπεια να έχει ιδιαίτερη σημασία το τι εκφράζει η εκάστοτε χρονοσειρά. Περιλαμβάνουν τις γραμμικές ή μη γραμμικές παλινδρομήσεις (πολλαπλές ή μη).

Στα αιτιοκρατικά μοντέλα, και συγκεκριμένα στις μη γραμμικές παλινδρομήσεις μπορούν να συμπεριληφθούν και τα μοντέλα νευρωνικών δικτύων.

3.3.1 Μοντέλα Χρονοσειρών

Τα μοντέλα χρονοσειρών, αποτελούν αρκετά δημοφιλή κατηγορία για την πρόβλεψη της ΟΤΣ, κυρίως ως μοντέλα αναφοράς (benchmark models) και έχουν χρησιμοποιηθεί ευρέως σε πάρα πολλές σχετικές έρευνες, είτε μεμονωμένα είτε σε συνδυασμό μεταξύ τους (συνδυαστικές – υβριδικές μέθοδοι). Η διαδεδομένη χρήση τους οφείλεται κυρίως στην ευκολία τους ως προς τη μαθηματική τους διατύπωση σε αντίθεση με την κατηγορία των αιτιοκρατικών. Απαραίτητη προϋπόθεση για την εφαρμογή των συγκεκριμένων μοντέλων είναι η ύπαρξη ιστορικών δεδομένων για την τιμή του μεγέθους το οποίο θέλουμε να προβλέψουμε, σε σταθερές χρονικές περιόδους. Κυρίαρχη υπόθεση είναι ότι η τιμή του μεγέθους μεταβάλλεται ακολουθώντας ένα συγκεκριμένο λανθάνον πρότυπο, χρονικά επαναλαμβανόμενο και σταθερό. Η παραγωγή των προβλέψεων επιτυγχάνεται με την αναγνώριση του ακολουθούμενου προτύπου και την προέκταση του στο μέλλον, με βάση τον χρονικό ορίζοντα που επιθυμούμε. Αξίζει να αναφερθεί πως η σχέση που περιγράφει τέτοιου είδους μοντέλα είναι γραμμική. Θα παρατεθούν ενδεικτικά κάποια μοντέλα παρακάτω.

Μέθοδος Αναφοράς – Naive

Πρόκειται για την απλούστερη στατιστική μέθοδο, η οποία ενώ δεν παράγει ακριβείς προβλέψεις, χρησιμοποιείται από τους ερευνητές ως μέθοδος αναφοράς (benchmark), κάτι το οποίο θα εφαρμόσουμε κι εμείς σε αυτή τη διπλωματική. Η πρόβλεψη που προκύπτει από τη μέθοδο Naive για μια χρονική στιγμή t είναι ίση με την πραγματική παρατήρηση της ακριβώς προηγούμενης χρονικής περιόδου ($t - 1$) (Πετρόπουλος & Ασημακόπουλος, 2011):

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

Στην περίπτωση που θέλουμε να παράγουμε περισσότερες από μία σημειακές προβλέψεις (ορίζοντας $m > 1$), η πρόβλεψη για κάθε χρονική στιγμή στο μέλλον ισούται με την τελευταία διαθέσιμη ιστορική παρατήρηση και η αντίστοιχη εξίσωση που περιγράφει τη μέθοδο είναι:

$$F_{t+m} = Y_t$$

Μέθοδοι Εκθετικής Εξομάλυνσης

Η συγκεκριμένη κατηγορία μεθόδων πρόβλεψης θεωρείται από τις πιο διαδεδομένες στους επιχειρηματικούς κύκλους, γεγονός το οποίο οφείλεται στην μαθηματική ευκολία τους, στην ελάχιστη απαίτηση σε υπολογιστικό χρόνο και στην ανάγκη για σχετικά μικρό αριθμό παρατηρήσεων ούτως ώστε να παράγουν προβλέψεις. Οι μέθοδοι εξομάλυνσης θεωρούνται καταλληλότερες για βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις μεγάλου όγκου χρονοσειρών και παρουσιάζουν καλύτερη απόδοση σε δεδομένα με στασιμότητα ή μικρό ρυθμό ανάπτυξης.

Εμπειρικές μελέτες έχουν δείξει ότι οι μέθοδοι εκθετικής εξομάλυνσης αποδίδουν καλύτερα όσον αφορά τα ποσοστά ακρίβειας σε σύγκριση με πιο περίπλοκες μεθόδους πρόβλεψης, κάτι το οποίο οφείλεται στο ότι δεν επηρεάζονται από τις ιδιομορφίες των προτύπων των δεδομένων ή από περιστασιακά εμφανιζόμενες ακραίες τιμές (spikes) που εμφανίζονται συχνά σε επιχειρησιακά δεδομένα. Στην παρούσα διπλωματική θα χρησιμοποιηθούν 3 από τις πιο βασικές μεθόδους εξομάλυνσης οι οποίες είναι οι ακόλουθες:

- Απλή Εκθετική Εξομάλυνση σταθερού επιπέδου (Simple Exponential Smoothing).
- Εκθετική Εξομάλυνση Γραμμικής Τάσης (Holt Exponential Smoothing).
- Εκθετική Εξομάλυνση μη γραμμικής ή φθίνουσας τάσης (Damped Exponential Smoothing).

Απλή Εκθετική Εξομάλυνση σταθερού επιπέδου (SES)

Το συγκεκριμένο μοντέλο σταθερού επιπέδου, το οποίο είναι γνωστό και ως απλή εκθετική εξομάλυνση (Simple Exponential Smoothing), περιγράφεται από τις ακόλουθες εξισώσεις:

$$e_t = Y_t - F_t$$

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

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

όπου:

- e_t δηλώνει το σφάλμα, δηλαδή την απόκλιση της πραγματικής τιμής από την πρόβλεψη,

- S δηλώνει το επίπεδο,
- F δηλώνει την πρόβλεψη,
- t δηλώνει τη χρονική περίοδο,
- α είναι ο συντελεστής εξομάλυνσης της μεθόδου, ο οποίος λαμβάνει τιμές στο διάστημα $[0,1]$.

Από τις παραπάνω σχέσεις που περιγράφουν την απλή εκθετική εξομάλυνση, καθίσταται σαφές ότι είναι απαραίτητο να οριστεί ένα αρχικό επίπεδο (S0) ώστε να μπορέσει να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης αλλά και των ζητούμενων σημειακών προβλέψεων δοθέντος κάποιου χρονικού ορίζοντα. Υπάρχουν 4 τρόποι για να υπολογίσουμε το αρχικό επίπεδο, η επιλογή των οποίων εξαρτάται από τα ποιοτικά χαρακτηριστικά της χρονοσειράς που εξετάζουμε. Ως σταθερό επίπεδο μπορούμε να θεωρήσουμε ένα από τα ακόλουθα:

1. Το μέσο όρο όλων των παρατηρήσεων.
2. Το μέσο όρο των n πρώτων παρατηρήσεων.
3. Την πρώτη παρατήρηση.
4. Το σταθερό επίπεδο από το μοντέλο απλής γραμμικής παλινδρόμησης (LRL).

Η ορθή επιλογή του σταθερού επιπέδου είναι σημαντική, διότι αφ' ενός επηρεάζει την επιλογή του συντελεστή εξομάλυνσης, αφ' ετέρου αν το σταθερό επίπεδο που θα επιλέξουμε δεν είναι αντιπροσωπευτικό των δεδομένων μας, θα προκαλέσει αρνητική επίδραση στην παραγωγή των προβλέψεων. Αξίζει να σημειωθεί πως στο αρχικό επίπεδο αποτελεί ουσιαστικά και την αρχική πρόβλεψη (F1).

Ιδιαίτερα σημαντικό ρόλο στο συγκεκριμένο μοντέλο πρόβλεψης παίζει ο συντελεστής εξομάλυνσης α . Η επιλογή του βέλτιστου συντελεστή εξομάλυνσης καθορίζεται από δύο αλληλεξαρτώμενους παράγοντες, το ποσοστό του θορύβου στη χρονοσειρά και τη σταθερότητα του μέσου όρου της χρονοσειράς. Όσον αφορά τον πρώτο παράγοντα, όσο περισσότερος είναι ο θόρυβος που εμφανίζεται στα δεδομένα τόσο μικρότερη πρέπει να είναι η τιμή του συντελεστή εξομάλυνσης, προκειμένου να αποφευχθεί υπερβολική αντίδραση στο θόρυβο.

Σχετικά με το δεύτερο παράγοντα, αν ο μέσος όρος παρουσιάζει μεταβολές, ο συντελεστής εξομάλυνσης πρέπει να λάβει μεγάλες τιμές ώστε να μπορούν οι προβλέψεις να ακολουθούν τις μεταβολές των δεδομένων. Στην περίπτωση που ο μέσος όρος είναι σχετικά σταθερός, λαμβάνεται μικρότερη τιμή του α .

Η βελτιστοποίηση της απλής εκθετικής εξομάλυνσης βασίζεται στην εύρεση του βέλτιστου συντελεστή εξομάλυνσης. Η συνηθέστερη διαδικασία που ακολουθείται είναι η γραμμική αναζήτηση η οποία ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (MSE), θέτοντας από την αρχή την ακρίβεια με την οποία θα γίνει η αναζήτηση καθώς και τα επιθυμητά όρια μέσα στα οποία θα κινηθεί ο συντελεστής α .

Πέρα της γραμμικής αναζήτησης, μπορεί να χρησιμοποιηθεί και η μη γραμμική ή δυαδική αναζήτηση για τη βελτιστοποίηση της μεθόδου. Πρόκειται για έναν αναδρομικό αλγόριθμο αναζήτησης ο οποίος χρησιμοποιεί την τεχνική διαίρει και βασίλευε. Η συγκεκριμένη μέθοδος εφαρμόζεται μόνο σε ταξινομημένους πίνακες, σε αντίθεση με τη γραμμική αναζήτηση, γεγονός που μειώνει σημαντικά την πολυπλοκότητα των υπολογισμών.

Το ζητούμενο στη δυαδική αναζήτηση είναι να εντοπιστεί μία συγκεκριμένη τιμή που ονομάζεται κλειδί, στην περίπτωση φυσικά που αυτό είναι δυνατό, δηλαδή αυτή η τιμή υπάρχει μέσα στον πίνακα που εξετάζουμε.

Χρησιμοποιώντας την δυαδική αναζήτηση, ζητούμενο είναι ξανά η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE). Αρχικά η παράμετρος (a) εκφράζεται μέσω ενός πίνακα διαστάσεων 1×2 . Ως ιδανικό a , θεωρείται το πρώτο στοιχείο αυτού του πίνακα, και ως ιδανικό MSE, το αντίστοιχο που προκύπτει από τη συγκεκριμένη τιμή του a .

Ο αλγόριθμος ξεκινάει, υπολογίζοντας για κάθε τιμή του a το αντίστοιχο MSE, και εν συνεχεία το συγκρίνει με το ιδανικό MSE, το οποίο είναι και η τιμή κλειδί. Σε περίπτωση που προκύψει μικρότερο MSE, τότε αυτό παίρνει τη θέση του κλειδιού, και κρατείται το a από το οποίο προέκυψε αυτό το κλειδί.

Στη συνέχεια, οι τιμές του αρχικού πίνακα μεταβάλλονται κατά τη διαφορά του πρώτου στοιχείου διαιρούμενο με τον αριθμό 2. Συγκεκριμένα, πλέον το πρώτο στοιχείο του πίνακα, δηλαδή το $a(1)$, ισούται με την προηγούμενη τιμή του $a(1)$ μείον τη διαφορά που περιγράφηκε, ενώ το νέο στοιχείο $a(2)$ ισούται με το παλιό συν τη διαφορά. Επιπλέον, η νέα διαφορά γίνεται μισή της προηγούμενης, και ο αλγόριθμος συνεχίζει από την αρχή υπολογίζοντας και συγκρίνοντας ξανά τα MSE, σύμφωνα με τις νέες τιμές που προέκυψαν για την παράμετρο a .

Όπως αναφέρθηκε, η δυαδική αναζήτηση παρουσιάζει μικρότερη υπολογιστική πολυπλοκότητα από τη γραμμική, γεγονός το οποίο την καθιστά πολύ πιο γρήγορη. Αξίζει να αναφερθεί πως όταν ο συντελεστής εξομάλυνσης λάβει την ακραία τιμή 1, τότε το μοντέλο ταυτίζεται με την μέθοδο Naïve, ενώ αν είναι $a=0$, κάθε πρόβλεψη είναι ίση με το αρχικό επίπεδο.

Επιπλέον, αν ζητείται ο υπολογισμός περισσότερων από μίας σημειακής πρόβλεψης, τότε όλες οι επόμενες θέτονται ίσες με την τελευταία υπολογισμένη πρόβλεψη (flat forecast) καθώς το μοντέλο θεωρεί την ύπαρξη ενός σταθερού επιπέδου (Πετρόπουλος & Ασημακόπουλος, 2011).

Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing)

Το μοντέλο αυτό αποτελεί μία επέκταση του μοντέλου της απλής εκθετικής εξομάλυνσης, η οποία έχει τη επιπλέον δυνατότητα να διαχειριστεί τη συνιστώσα της τάσης, μέγεθος το οποίο παρατηρείται συχνά στην πλειοψηφία των επιχειρησιακών δεδομένων. Η μέθοδος πήρε το όνομα της το οποίο διατηρεί μέχρι και σήμερα από τον Holt, ο οποίος πρότεινε τη συγκεκριμένη επέκταση το 1957. Η μέθοδος Holt περιγράφεται από τις ακόλουθες εξισώσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + a * e_t$$

$$T_t = T_{t-1} + b * e_t$$

$$F_{t+1} = S_t + m * T_t$$

όπου:

- e_t δηλώνει το σφάλμα, δηλαδή την απόκλιση της πραγματικής τιμής από την πρόβλεψη,
- S δηλώνει το επίπεδο,
- F δηλώνει την πρόβλεψη,

- t δηλώνει τη χρονική περίοδο,
- T δηλώνει την τάση,
- m δηλώνει το χρονικό ορίζοντα της πρόβλεψης,
- α είναι ο συντελεστής εξομάλυνσης της μεθόδου για το επίπεδο, ο οποίος λαμβάνει τιμές στο διάστημα $[0,1]$,
- β είναι ο συντελεστής εξομάλυνσης της μεθόδου για την τάση, ο οποίος λαμβάνει τιμές στο διάστημα $[0,1]$.

Όπως και στην απλή εκθετική εξομάλυνση, έτσι και στην μέθοδο Holt απαιτείται ο προσδιορισμός ενός αρχικού επιπέδου (S_0) αλλά και μίας αρχικής τάσης (T_0), έτσι ώστε να είναι δυνατή η εκκίνηση της διαδικασίας υπολογισμού του μοντέλου πρόβλεψης αλλά και η παραγωγή των ζητούμενων σημειακών προβλέψεων δοθέντος κάποιου χρονικού ορίζοντα.

Το αρχικό επίπεδο προσδιορίζεται όπως και στην μέθοδο SES που αναφέρθηκε προωύτερα. Η αρχική τάση υπολογίζεται ως:

1. Η διαφορά δεύτερης και πρώτης παρατήρησης ($Y_2 - Y_1$)
2. Η διαφορά n -οστής και πρώτης παρατήρησης διαιρεμένης με $n-1$ (πχ. $(Y_8 - Y_1)/7$)
3. Η σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης (LRL).

Και στην μέθοδο Holt, είναι σημαντική η σωστή επιλογή του αρχικού επιπέδου αλλά και της αρχικής τάσης, για την ορθή λειτουργία του μοντέλου πρόβλεψης και την παραγωγή ακριβέστερων σημειακών προβλέψεων. Η επιλογή των δύο αυτών μεταβλητών γίνεται με γνώμονα τα ποιοτικά χαρακτηριστικά των διαθέσιμων ιστορικών δεδομένων για τη χρονοσειρά που εξετάζουμε. Στην περίπτωση μας επιλέγουμε την μέθοδο της απλής γραμμικής παλινδρόμησης λόγω του όγκου των δεδομένων.

Η μέθοδος Holt, όπως και η μέθοδος SES, επιδέχεται βελτιστοποίησης, η οποία βασίζεται στην επιλογή του βέλτιστου συνδυασμού των παραμέτρων α και β . Η γραμμική αναζήτηση αποτελεί και πάλι την πιο διαδεδομένη λύση στο πρόβλημά μας, όμως για τη βελτιστοποίηση της συγκεκριμένης μεθόδου είναι ασφαλέστερη η επιλογή της μη γραμμικής (δυναδικής) αναζήτησης με κριτήριο και πάλι τη μείωση του μέσου τετραγωνικού σφάλματος (MSE), καθώς διαπιστώθηκε με τη βοήθεια πειραμάτων ότι παρουσιάζει καλύτερα αποτελέσματα ενώ παράλληλα μειώνει την πολυπλοκότητα της διαδικασίας υπολογισμού των παραμέτρων α και β .

Σε αυτό το σημείο είναι σημαντικό να αναφερθεί πως στις περισσότερες περιπτώσεις η βέλτιστη τιμή του συντελεστή β για την τάση είναι μικρότερη από την τιμή του συντελεστή α για το επίπεδο, λόγω του ότι η τιμή της τάσης για κάθε περίοδο είναι σημαντικά μικρότερη από την τιμή του αντίστοιχου επιπέδου.

Στην περίπτωση που προσθέσουμε το ίδιο ποσοστό σφάλματος και στο επίπεδο αλλά και στην τάση, υπάρχει το ενδεχόμενο να προκύψουν προβλέψεις που θα χαρακτηρίζονται από μεγάλη αστάθεια.

Μοντέλο μη Γραμμικής Τάσης

Το μοντέλο μη γραμμικής τάσης αποτελεί ουσιαστικά μία τροποποίηση του μοντέλου γραμμικής τάσης. Η ανάγκη τροποποίησης του μοντέλου Holt έγκειται στο γεγονός ότι παρουσιάζει μία τάση για υπερεκτίμηση, δηλαδή οι προβλέψεις που παράγει χαρακτηρίζονται από θετική προκατάληψη, ιδιαίτερα σε περίπτωση μεσοπρόθεσμων ή μακροπρόθεσμων προβλέψεων. Η τροποποίηση αυτή επιτυγχάνεται με τη προσθήκη μίας επιπλέον παραμέτρου η οποία ελέγχει το ρυθμό αύξησης των τιμών. Η παράμετρος αυτή καλείται παράμετρος διόρθωσης της τάσης (trend-modification parameter) και συμβολίζεται φ.

Οι εξισώσεις που περιγράφουν το συγκεκριμένο μοντέλο με τον τρόπο που αυτό παρουσιάστηκε από τους Gardner και McKenzie το 1985 είναι οι ακόλουθες:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \varphi * T_{t-1} + a * e_t$$

$$T_t = \varphi * T_{t-1} + b * e_t$$

$$F_{t+m} = S_t + \sum_{i=1}^m \varphi^i * T_t$$

όπου:

- e_t δηλώνει το σφάλμα, δηλαδή την απόκλιση της πραγματικής τιμής από την πρόβλεψη,
- S δηλώνει το επίπεδο,
- F δηλώνει την πρόβλεψη,
- t δηλώνει τη χρονική περίοδο,
- T δηλώνει την τάση,
- m δηλώνει το χρονικό ορίζοντα της πρόβλεψης,
- a είναι ο συντελεστής εξομάλυνσης της μεθόδου για το επίπεδο, ο οποίος λαμβάνει τιμές στο διάστημα $[0,1]$,
- b είναι ο συντελεστής εξομάλυνσης της μεθόδου για την τάση, ο οποίος λαμβάνει τιμές στο διάστημα $[0,1]$.
- φ είναι η παράμετρος διόρθωσης της τάσης.

Από τις παραπάνω σχέσεις είναι σαφές ότι το μοντέλο είναι σχεδόν πανομοιότυπο με αυτό της γραμμικής τάσης με εξαίρεση την παρουσία της παραμέτρου φ , όπου συμβάλει στην παραγωγή προβλέψεων με μη γραμμικό υπολογισμό της τάσης μέσω του συντελεστή m , σε αντίθεση με πριν (Holt) όπου ο υπολογισμός ήταν γραμμικός.

Σε αντίθεση με τις παραμέτρους a και b , η παράμετρος φ μπορεί να λάβει οποιαδήποτε μεγαλύτερη του μηδενός τιμή, όμως είναι αναγκαία η επιβολή άνω και κάτω ορίων για ορισμένες χαρακτηριστικές τιμές. Επομένως, μπορούμε να χαρακτηρίσουμε το μοντέλο μη γραμμικής τάσης ανάλογα με την τιμή που παίρνει κάθε φορά η παράμετρος φ , ως εξής (Πετρόπουλος & Ασημακόπουλος, 2011):

- Αν $\varphi=0$, τότε προκύπτει το μοντέλο της απλής εκθετικής εξομάλυνσης

(SES), διότι η τάση δεν έχει πλέον καμία επίδραση στον καθορισμό των στατιστικών σημειακών προβλέψεων.

- Αν $0 < \phi < 1$, τότε προκύπτει το μοντέλο φθίνουσας τάσης (damped exponential smoothing), το οποίο θεωρείται καταλληλότερο όταν μας ενδιαφέρει η παραγωγή μεσοπρόθεσμων προβλέψεων, καθώς χαρακτηρίζεται από έλλειψη τάσης για υπεραισιοδοξία.
- Αν $\phi = 1$, τότε προκύπτει το μοντέλο της γραμμικής τάσης, αφού στην τελευταία εξίσωση η οποία παράγει και την πρόβλεψη, τη θέση του αθροίσματος παίρνει το γινόμενο της μεταβλητής χρονικού ορίζοντα m και τα προηγούμενης τάσης T_t .
- Αν $\phi > 1$, τότε προκύπτει το μοντέλο της εκθετικής τάσης, το οποίο χαρακτηρίζεται από μεγάλη θετική προκατάληψη και χρησιμοποιείται μόνο σε περιπτώσεις όπου το ζητούμενο είναι η πρόβλεψη τιμής στην αρχή του κύκλου ζωής ενός προϊόντος ή μιας υπηρεσίας.

Το μεγάλο μειονέκτημα του συγκεκριμένου μοντέλου, είναι η παραγωγή ιδιαίτερα προκατειλημμένων προβλέψεων, κάτι το οποίο ενδέχεται να προκύψει από λανθασμένη επιλογή του συντελεστή ϕ . Γι αυτό το λόγο, συνίσταται ο περιορισμός της παραμέτρου ϕ στο διάστημα $[0,1]$.

Αξίζει να σημειωθεί πως η ακρίβεια των προβλέψεων που παράγονται από το μοντέλο μη γραμμικής τάσης είναι σημαντικά μεγαλύτερη σε σχέση με αυτές του μοντέλου γραμμικής τάσης. Γενικά, όσο μεγαλύτερος είναι ο απαιτούμενος ορίζοντας της πρόβλεψης, τόσο μεγαλύτερο πλεονέκτημα παρουσιάζει το συγκεκριμένο μοντέλο έναντι των υπόλοιπων μοντέλων εκθετικής εξομάλυνσης.

Όσον αφορά τη βέλτιστη επιλογή των παραμέτρων α , β και ϕ , αυτή επιτυγχάνεται με μη γραμμική (δυαδική) αναζήτηση, με κριτήριο πάντα την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE), ενώ η επιλογή αρχικού επιπέδου και αρχικής τάσης πραγματοποιείται με εφαρμογή της απλής γραμμικής παλινδρόμησης, όπως ακριβώς και στην μέθοδο Holt.

3.3.2 Αιτιοκρατικά Μοντέλα

Τα αιτιοκρατικά μοντέλα θεωρούνται ως τα πιο διαδεδομένα μοντέλα για την πρόβλεψη της τιμής της ηλεκτρικής ενέργειας, καθότι έχουν τη δυνατότητα να ενσωματώνουν διάφορες παραμέτρους που επηρεάζουν είτε σε μικρό είτε σε μεγάλο βαθμό της τιμή της ΟΤΣ. Τα μοντέλα αυτά λειτουργούν υπό την θεμελιώδη υπόθεση ότι το υπό πρόβλεψη μέγεθος (εξαρτημένη μεταβλητή) σχετίζεται μέσω κάποιας σταθερής σχέσης με κάποιες παραμέτρους (ανεξάρτητες μεταβλητές), οι οποίες και το επηρεάζουν.

3.3.2.1 Απλή Γραμμική Παλινδρόμηση

Στη μέθοδο της απλής γραμμικής παλινδρόμησης Linear Regression Line (LRL), θεωρούμε ότι υπάρχει γραμμική σχέση ανάμεσα στην μεταβλητή την οποία θέλουμε να προβλέψουμε (εξαρτημένη) και σε κάποια άλλη μεταβλητή (ανεξάρτητη) η οποία συνήθως είναι ο χρόνος στην περίπτωση των χρονοσειρών.

Λόγω της υπόθεσης ότι η σχέση μεταξύ των δύο μεταβλητών είναι γραμμική, η συγκεκριμένη μέθοδος έχει πολύ περιορισμένο πεδίο εφαρμογής. Στόχος της απλής γραμμικής παλινδρόμησης είναι να εκφράσει τη σχέση που αναφέραμε ανάμεσα στις μεταβλητές X και Y με την εξίσωση μίας ευθείας γραμμής, δηλαδή μια σχέση της μορφής:

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

όπου :

- a δηλώνει την τεταγμένη του σημείου τομής της ευθείας με τον άξονα της εξαρτημένης μεταβλητής, δηλαδή όταν θεωρηθεί μηδενική τιμή για την ανεξάρτητη μεταβλητή,
- b δηλώνει την κλίση της ευθείας, δηλαδή το ρυθμό μεταβολής του Y ανά μοναδιαία αύξηση του X.

Με δεδομένες τις πραγματικές τιμές Y της εξαρτημένης μεταβλητής, ο υπολογισμός των συντελεστών a και b γίνεται με γνώμονα την ελαχιστοποίηση των τετραγώνων των διαφορών των πραγματικών τιμών Y από τις τιμές που παράγει η εξίσωση παλινδρόμησης και περιγράφεται από την ακόλουθη σχέση:

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

Η λογική της παραπάνω μεθόδου βασίζεται στο ότι πρέπει να ελαχιστοποιηθεί η απόσταση των πραγματικών παρατηρήσεων Y από τη βέλτιστη γραμμή παλινδρόμησης. Η μέθοδος αυτή καλείται μέθοδος ελαχίστων τετραγώνων και με βάση αυτή οι σχέσεις που δίνουν τις παραμέτρους a και b είναι οι εξής:

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

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

όπου:

- \bar{X} και \bar{Y} είναι οι μέσες τιμές των διανυσμάτων X και Y,
- n είναι ο αριθμός των παρατηρήσεων (ιστορικών δεδομένων του υπό πρόβλεψη μεγέθους) με βάση τις οποίες υπολογίζεται η ευθεία της παλινδρόμησης.

3.3.2.2 Πολλαπλή Γραμμική Παλινδρόμηση

Η πολλαπλή γραμμική παλινδρόμηση αποτελεί στην ουσία μία γενίκευση της απλής γραμμικής παλινδρόμησης, και χρησιμοποιείται στην περίπτωση που θέλουμε να εξετάσουμε την επίδραση περισσότερων της μίας ανεξάρτητων μεταβλητών στην εξαρτημένη μεταβλητή. Η γενική σχέση που εκφράζει τη συγκεκριμένη μέθοδο είναι η εξής:

$$\hat{Y} = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k + e$$

όπου

- Y εκφράζει την εξαρτημένη μεταβλητή,
- X_1, X_2 έως X_k εκφράζουν τις ανεξάρτητες μεταβλητές,
- Οι συντελεστές b_0, b_1 έως b_k είναι σταθερές παράμετροι,
- e εκφράζει τον τυχαίο παράγοντα

- κατανεμημένος γύρω από το μηδέν.

Η παραπάνω εξίσωση είναι γραμμική ως προς τους συντελεστές. Για να επιτύχουμε τη γραμμικότητα της σχέσης θεωρούμε ότι κάθε συντελεστής b είναι υψωμένος σε εκθέτη που ισούται με τη μονάδα, και μπορούμε να υπολογίσουμε τις τιμές των συντελεστών με την χρήση της μεθόδου των ελαχίστων τετραγώνων.

Στην περίπτωση της πολλαπλής παλινδρόμησης, το γράφημα της συνάρτησης που συνδέει την μεταβλητή Y με τις διάφορες μεταβλητές X , δεν είναι πλέον μία ευθεία γραμμή όπως στην περίπτωση της απλής γραμμικής παλινδρόμησης και επομένως είναι δύσκολο να περιγραφεί. Κυρίαρχο ζητούμενο στη διαδικασία της πολλαπλής παλινδρόμησης είναι ο προσδιορισμός των συντελεστών b αλλά και της διακύμανσης του τυχαίου παράγοντα, έχοντας πάντα ένα συγκεκριμένο σύνολο δεδομένων στο οποίο εφαρμόζουμε τη μέθοδο των ελαχίστων τετραγώνων.

Πιο αναλυτικά, ο υπολογισμός των συντελεστών b της πολλαπλής παλινδρόμησης, για παράδειγμα στην περίπτωση που έχουμε δύο ανεξάρτητες μεταβλητές, βασίζεται στην εφαρμογή της μεθόδου των ελαχίστων τετραγώνων ώστε να υπολογιστεί το ελάχιστο άθροισμα των τετραγώνων των σφαλμάτων, ως εξής:

$$(b_0, b_1, b_2) \mid \min[\sum_{i=1}^n e_i^2] ,$$

όπου:

$$e_i = Y_i - \bar{Y}_i$$

Όμως ισχύει η ακόλουθη σχέση για τα σφάλματα:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \bar{Y}_i)^2 = \sum_{i=1}^n (Y_i - b_0 - b_1 * X_{1,i} - b_2 * X_{2,i})^2$$

Άρα, για να υπολογιστούν οι άγνωστοι συντελεστές b_0, b_1, b_2 , οι οποίοι ελαχιστοποιούν την παραπάνω σχέση, πρέπει να υπολογιστούν οι μερικές παράγωγοι αυτής για κάθε έναν από τους συντελεστές, στη συνέχεια να θέσουμε τις προκύπτουσες παραγώγους ίσες με το μηδέν και τέλος να λυθεί ένα γραμμικό σύστημα τριών αγνώστων.

Η διαδικασία που περιγράφηκε, η οποία για λόγους κατανόησης αναφέρεται μόνο σε δύο ανεξάρτητες μεταβλητές, μπορεί να γενικευθεί σε οποιοδήποτε μοντέλο πολλαπλής γραμμικής παλινδρόμησης, με περισσότερες των δύο ανεξάρτητες μεταβλητές.

3.4 Μέθοδοι για βραχυπρόθεσμη πρόβλεψη

Όπως αναφέρθηκε και προηγουμένως η βραχυπρόθεσμη πρόβλεψη της ΟΤΣ αποτελεί το αντικείμενο της πλειοψηφίας των περισσότερων ερευνών σχετικά με το συγκεκριμένο ζήτημα. Όπως διαπιστώνουμε και από τον παραπάνω πίνακα, οι ερευνητές χρησιμοποιούν μία μεγάλη γκάμα από μεθόδους προβλέψεων, ίσως δίνοντας μεγαλύτερη βαρύτητα σε συνδυαστικές μεθόδους, όπως είναι για παράδειγμα η Wavelet-ARIMA, ή συνδυαστικές μέθοδοι παλινδρόμησης, ή ακόμα και συνδυαστικές μέθοδοι που περιέχουν Νευρωνικά Δίκτυα (RBFN, ERBFN, PSO-ANFIS). Τέτοιες μέθοδοι προτιμώνται στην πράξη, καθώς μπορούν να απομονώσουν περισσότερα από τα ιδιαίτερα χαρακτηριστικά του μεγέθους της ΟΤΣ, όπως για παράδειγμα την πολλαπλή εποχικότητα ή τις ασυνήθιστες τιμές (spikes).

Το θέμα της πολλαπλής εποχικότητας, απασχόλησε κατά καιρούς πάρα πολλούς ερευνητές, καθώς αποτελεί ίσως το πιο περίπλοκο χαρακτηριστικό της ΟΤΣ, μαζί με τις ξαφνικές τις μεταβολές των τιμών (spikes).

Βασικές τεχνικές που ακολουθούνται για την εξάλειψη της πολλαπλής εποχικότητας είναι οι μέθοδοι εκθετικής εξομάλυνσης, όπως η μέθοδος διπλής εποχικότητας (Double Seasonal Exponential Smoothing) του Taylor (2003) που αφορά χρονοσειρές που παρουσιάζουν δύο εποχιακούς κύκλους, έναν μικρό που επαναλαμβάνεται με μεγάλη συχνότητα μέσα σε ένα μεγαλύτερο, και η μέθοδος Holt – Winters η οποία αφορά τόσο πολλαπλασιαστική όσο και αθροιστική πολλαπλή εποχικότητα (Phillip G. Gould et al., 2008).

Εναλλακτική αντιμετώπιση θα μπορούσε να αποτελέσει η προσομοίωση της μακράς εποχικότητας με μετασχηματισμό Fourier συγκεκριμένης συχνότητας.

Θα αναλυθεί εκτενέστερα στο θέμα της πολλαπλής εποχικότητας αλλά και στον τρόπο αντιμετώπισής του στο πρακτικό μέρος της διπλωματικής.

Στα πλαίσια της παρούσας εργασίας, ελήφθησαν υπόψιν ωριαία δεδομένα ΟΤΣ από τη ΡΑΕ των τελευταίων τριών ετών και ύστερα από την σχετική προ επεξεργασία των δεδομένων χρησιμοποιήθηκαν:

- Η βιβλιοθήκη TBATS
- Μοντέλα ARMA μέσω της βιβλιοθήκης *statsmodels.api*
- Μοντέλα ARMA μέσω της βιβλιοθήκης *pyramid*
- Μοντέλα SARIMAX Με συνιστώσες Fourier
- Βιβλιοθήκη FB prophet

3.5 Μέθοδοι για μακροπρόθεσμη πρόβλεψη

Περιορισμένη ερευνητική εμπειρία παρουσιάζεται στις μεθόδους μεσοπρόθεσμης και μακροπρόθεσμης πρόβλεψης του δείκτη ΟΤΣ λόγω χαρακτηριστικών, όπως είναι η υψηλή μεταβλητότητα, τα οποία δυσκολεύουν σε μεγάλο βαθμό την πρόβλεψη του στο απώτερο μέλλον, καθώς και έλλειψη πολλών ιστορικών δεδομένων λόγω της βραχύβιας πορείας των ελεύθερων Αγορών Ηλεκτρικής Ενέργειας.

Επιπλέον, όπως αναφέρθηκε και σε προηγούμενη ενότητα, η βραχυπρόθεσμη πρόβλεψη είναι εκείνη που ενδιαφέρει περισσότερο τους συμμετέχοντες στην Αγορά Ενέργειας, επομένως η ανάγκη για μεσοπρόθεσμες προβλέψεις είναι περισσότερο περιορισμένη αλλά όχι και δίχως χρησιμότητα.

Όπως χαρακτηριστικά αναφέρει στην έρευνά του ο Torghaban (2010), η μεσοπρόθεσμη πρόβλεψη της τιμής της ηλεκτρικής ενέργειας είναι απαραίτητη για διάφορες εφαρμογές στην Ηλεκτρική Αγορά Ενέργειας, όπως τιμολόγηση των παραγώγων, ο προγραμματισμός της συντήρησης για τις εταιρείες παραγωγής και ο καθορισμός του προϋπολογισμού και οι συμφωνίες για καύσιμα.

Στην παρούσα εργασία χρησιμοποιήθηκαν οι ακόλουθες μέθοδοι που θα αναλυθούν παρακάτω εκτενώς:

- SARIMA μοντέλο της βιβλιοθήκης *pmddarima* με μόνο συνιστώσες κινούμενων μέσων και μοντελοποίηση μακράς εποχικότητας με μετασχηματισμό Fourier
- FB Prophet

Ως μέτρο σύγκρισης επιλέγονται το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error), το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error), η Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error) καθώς και άλλες μετρικές όπως MAPE, sMAPE και MARNE υπολογίστηκαν για βοηθητικούς σκοπούς.

3.6 Μετρικές αξιολόγησης χρονοσειρών

Όπως αναφέρθηκε και στην αρχή αυτού του κεφαλαίου, η ακρίβεια μιας πρόβλεψης είναι αυτή που καθορίζει το κατά πόσο αυτή είναι αποδεκτή και μας παρέχει όσο το δυνατόν ασφαλέστερα αποτελέσματα. Κυρίαρχο ρόλο στην ανάλυση της ακρίβειας των προβλέψεων παίζουν τα σφάλματα.

Ως σφάλμα ορίζεται η διαφορά μεταξύ της πραγματικής και της προβλεπόμενης τιμής για μία περίοδο. Η μαθηματική διατύπωση του παραπάνω ορισμού είναι η ακόλουθη:

$$e_i = Y_i - F_i$$

όπου:

- i είναι η εξεταζόμενη περίοδος
- e_i είναι το σφάλμα
- Y_i είναι η πραγματική τιμή του μεγέθους
- F_i είναι η πρόβλεψη της τιμής του μεγέθους.

Όμως η παραπάνω απλή σχέση δεν είναι σε θέση να κρίνει την αποδοτικότητα των μεθόδων πρόβλεψης, γι' αυτό και συνίσταται η χρήση πιο ολοκληρωμένων στατιστικών δεικτών σφάλματος, τους οποίους και παρουσιάζουμε στη συνέχεια.

3.6.1 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error)

Ο συγκεκριμένος δείκτης εκφράζει ένα μέτρο της ακρίβειας της πρόβλεψης έναντι των πραγματικών τιμών διατηρώντας τις αρχικές μονάδες της χρονοσειράς. Μεγάλη τιμή του δείκτη σημαίνει μικρή ακρίβεια της μεθόδου πρόβλεψης που εφαρμόστηκε. Η μαθηματική διατύπωση αυτού του στατιστικού δείκτη είναι η εξής:

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

3.6.2 Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error)

Το μέσο τετραγωνικό σφάλμα είναι ένας στατιστικός δείκτης που λειτουργεί ως μέτρο της ακρίβειας της πρόβλεψης, δίνοντας όμως περισσότερο βάρος στα μεγάλα σφάλματα και λιγότερο στα μικρά (λογικό αφού ο τύπος είναι υψωμένος στο τετράγωνο). Ο συγκεκριμένος δείκτης χρησιμοποιείται κυρίως ως κριτήριο επιλογής των κατάλληλων παραμέτρων εξομάλυνσης και υπολογίζεται από τον ακόλουθο τύπο:

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

3.6.3 Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error)

Έχει όμοιες ιδιότητες με τον προηγούμενο δείκτη, μόνο που είναι εκφρασμένος σε μονάδες της αρχικής χρονοσειράς. Ο τύπος που περιγράφει αυτό το στατιστικό δείκτη είναι ο παρακάτω:

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

Από του δείκτες που παρουσιάστηκαν σε αυτό το κεφάλαιο, σημαντικότερος για τον πρόβλημα πρόβλεψης της ΟΤΣ, μπορούν να θεωρηθεί ο δείκτης της Ρίζας Μέσου Τετραγωνικού Σφάλματος.

Η άποψη αυτή στηρίζεται στο γεγονός πως η πλειοψηφία των ερευνητών που ασχολήθηκαν με το συγκεκριμένο θέμα, έθεσαν ως μέτρα αξιολόγησης αυτούς τους δύο στατιστικούς δείκτες, γεγονός που μας οδηγεί να κινηθούμε και εμείς προς αυτήν την κατεύθυνση. Η αξιολόγηση των σφαλμάτων που θα χρησιμοποιηθούν στην παρούσα διπλωματική, θα αναλυθεί εκτενώς στο πρακτικό μέρος της εργασίας.

3.6.4 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error)

Πρόκειται σίγουρα για ένα αμφιλεγόμενο δείκτη, που εκφράζει το μέσο όρο των ποσοστιαίων σφαλμάτων. Συγκεκριμένα, είναι το άθροισμα των μεμονωμένων απόλυτων σφαλμάτων διαιρούμενο με τη ζήτηση (κάθε περίοδο ξεχωριστά) και ο τύπος φαίνεται παρακάτω:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|A_t - F_t|}{A_t}$$

Στα πλεονεκτήματα που προσφέρει είναι ότι εκφράζεται ως ποσοστό, το οποίο είναι ανεξάρτητο από την κλίμακα και μπορεί να χρησιμοποιηθεί για τη σύγκριση προβλέψεων σε διαφορετικές κλίμακες. Ωστόσο, λαμβάνει τιμές που ενδέχεται να υπερβαίνουν το 100%. Επίσης, παίρνει απροσδιόριστες τιμές όταν υπάρχουν μηδενικές πραγματικές τιμές. Επιπλέον, παίρνει ακραίες τιμές όταν οι πραγματικές τιμές είναι πολύ κοντά στο μηδέν.

Το MAPE είναι ασύμμετρο και επιβάλλει βαρύτερη ποινή σε αρνητικά σφάλματα (όταν οι προβλέψεις είναι υψηλότερες από τις πραγματικές) από ότι σε θετικά σφάλματα. Αυτό οφείλεται στο γεγονός ότι το ποσοστό σφάλματος δεν μπορεί να υπερβαίνει το 100% για προβλέψεις που είναι πολύ χαμηλές. Ενώ δεν υπάρχει ανώτερο όριο για τις προβλέψεις που είναι πολύ υψηλές. Συμπερασματικά, το MAPE θα ευνοήσει μοντέλα που κάνουν υπό-πρόβλεψη (under-forecast) παρά αυτά που κάνουν υπέρ-πρόβλεψη (over-forecast).

Επιπλέον, η συνάρτηση υπολογισμού του MAPE δεν είναι παντού παραγωγίσιμη, γεγονός που μπορεί να οδηγήσει σε προβλήματα κατά τη χρήση του ως κριτηρίου βελτιστοποίησης.

3.6.5 Συμμετρικό Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (symmetric Mean Absolute Percentage Error)

Προκειμένου να αντισταθμιστούν πολλά από τα παραπάνω μειονεκτήματα του MAPE, στην παρούσα διπλωματική υπολογίστηκε σε κάθε μοντέλο το sMAPE (symmetric MAPE). Ο τύπος είναι:

$$sMAPE = \frac{100}{n} \cdot \sum_{t=1}^n \frac{|F_t - A_t|}{(|F_t| + |A_t|)/2}$$

Όπως φαίνεται και από την ονομασία, εισάχθηκε για να ξεπεράσει την ασυμμετρία που αναφέρθηκε παραπάνω και συγκεκριμένα την έλλειψη ορίου των προβλέψεων που είναι μεγαλύτερες από τις πραγματικές τιμές.

Με αυτό τον τρόπο επιτυγχάνεται να τεθούν όρια στο ποσοστό του MAPE από 0%-200%. Δεδομένου επίσης ότι το σύνολο δεδομένων δεν περιέχει μηδενικές τιμές όπως και οι προβλέψεις, αποφεύγονται και τυχόν ασταθείς όροι του σφάλματος που μπορεί να οδηγήσουν σε πολύ μεγάλες τιμές του σφάλματος.

3.6.6 Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα Εύρους (Mean Absolute Range Normalized Error)

Το σφάλμα (MARNE) είναι η απόλυτη διαφορά μεταξύ του πραγματικού και πρόβλεψη SMP, κανονικοποιημένο στη μέγιστη πραγματική τιμή όπως φαίνεται και από τον τύπο :

$$MARNE = \frac{100}{n} \cdot \sum_{i=1}^n \frac{|F_t - A_t|}{\max A_t}$$

Το MARNE είναι ένας ποσοστιαίος δείκτης που αποτελεί αξιόπιστη μετρική σε περιπτώσεις χρονοσειρών με τιμές μηδέν ή σχεδόν μηδενικές. Ο δείκτης MAPE χρησιμοποιείται ευρέως σε μελέτες πρόβλεψης φορτίου. Ωστόσο, για χρονοσειρές με σχεδόν μηδενικές τιμές όπως το SMP της ελληνικής αγοράς, το MAPE λαμβάνει εξαιρετικά μεγάλες τιμές, γεγονός που καθιστά λιγότερο πρακτική την εξαγωγή συμπερασμάτων σχετικά με την ανθεκτικότητα ενός μοντέλου.

Κεφάλαιο 4

4.1 Προεπεξεργασία δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν αφορούν ωριαίες τιμές ΟΤΣ για το χρονικό διάστημα 01/01/2018 – 01/11/2020. Βασικό μέλημα πριν προχωρήσουμε στην ανάλυση είναι η ανίχνευση της φύσης των δεδομένων και η υπάρξη NaN τιμών.

Το τελικό Spark Dataframe, έχει το ακόλουθο schema:

Σχ. 4.1

```
final_df.count()

24888

final_df.printSchema()

root
 |-- SMPS: double (nullable = true)
 |-- Daily_Average_SMP: double (nullable = true)
 |-- date: timestamp (nullable = true)
```

και όπως φαίνεται παραπάνω, ένα σύνολο 24888 εγγραφών.

Στην προσπάθεια να ανιχνευτούν NaN τιμές στην στήλη SMPS (System Marginal Prices) που περιέχονται οι ωριαίες τιμές ΟΤΣ για το διάστημα που αναφέρθηκε, διαπιστώθηκε ότι υπάρχουν και θεωρήθηκε ασφαλές να αντικατασταθούν με τις αντίστοιχες ημερήσιες μέσες τιμές.

Παρακάτω φαίνονται οι 26 ελλείπουσες τιμές:

Σχ. 4.2

```
[20]: final_df.where(isnan(col("SMPS"))).show(50)

+-----+-----+-----+
|SMPS|Daily_Average_SMP|date|
+-----+-----+-----+
|NaN|36.55230434782609|2018-03-25 23:00:00|
|NaN|53.08413043478261|2019-03-31 23:00:00|
|NaN|38.19165217391305|2020-03-29 23:00:00|
|NaN|92.223|2020-11-01 01:00:00|
|NaN|92.223|2020-11-01 02:00:00|
|NaN|92.223|2020-11-01 03:00:00|
|NaN|92.223|2020-11-01 04:00:00|
|NaN|92.223|2020-11-01 05:00:00|
|NaN|92.223|2020-11-01 06:00:00|
|NaN|92.223|2020-11-01 07:00:00|
|NaN|92.223|2020-11-01 08:00:00|
|NaN|92.223|2020-11-01 09:00:00|
|NaN|92.223|2020-11-01 10:00:00|
|NaN|92.223|2020-11-01 11:00:00|
|NaN|92.223|2020-11-01 12:00:00|
|NaN|92.223|2020-11-01 13:00:00|
|NaN|92.223|2020-11-01 14:00:00|
|NaN|92.223|2020-11-01 15:00:00|
|NaN|92.223|2020-11-01 16:00:00|
|NaN|92.223|2020-11-01 17:00:00|
|NaN|92.223|2020-11-01 18:00:00|
|NaN|92.223|2020-11-01 19:00:00|
|NaN|92.223|2020-11-01 20:00:00|
|NaN|92.223|2020-11-01 21:00:00|
|NaN|92.223|2020-11-01 22:00:00|
|NaN|92.223|2020-11-01 23:00:00|
```

Η παρεμβολή (interpolation) υλοποιείται μέσω της εντολής `coalesce`, μία `non-aggregate` κανονική συνάρτηση στο `Spark SQL`. Η συνένωση δίδεται με την πρώτη μη μηδενική τιμή μεταξύ των δεδομένων στηλών ή `null` εάν όλες οι στήλες είναι `null`.

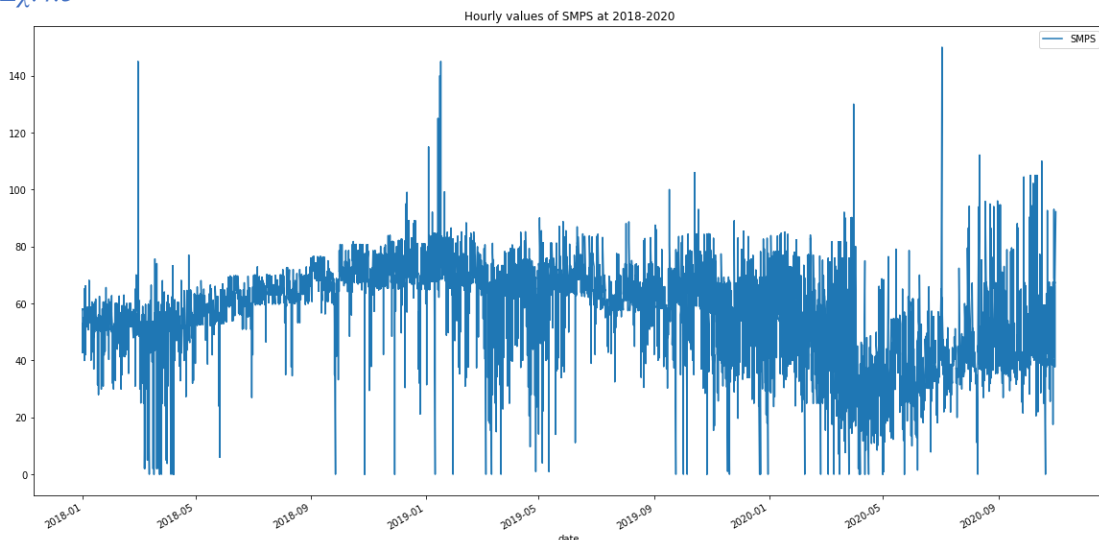
Στο σημείο αυτό, λόγω των παρακάτω συνθηκών της ανάλυσης:

- Βασικό αντικείμενο μελέτης δεν είναι `processing` Μεγάλων Δεδομένων αλλά πρόβλεψη χρονοσειρών
- Το μέγεθος των δεδομένων (24888 εγγραφές) επιτρέπει την `standalone` ανάλυση των δεδομένων σε ένα πυρήνα

γίνεται η παραδοχή ότι η συνέχεια της ανάλυσης δεν θα γίνει σε ψευδο-καταναμημένο περιβάλλον και συνεπώς θα χρησιμοποιηθούν και αντίστοιχες δομές δεδομένων (`numpy arrays` και `pandas dataframes`).

Υιοθετώντας την παραπάνω παραδοχή, το πρώτο που θα παρουσιάσουμε είναι η απεικόνιση της συνολικής χρονοσειράς.

Σχ.4.3



Εποπτικά και διαισθητικά, φαίνεται ότι η σειρά παρουσιάζεται με μια στασιμότητα όμως θα μελετηθεί σε λίγο κατά πόσον ισχύει.

4.2 Μελέτη χρονοσειράς

4.2.1 Ποιοτικά Χαρακτηριστικά των Χρονοσειρών

Απαραίτητη προϋπόθεση για την παραγωγή όσο το δυνατόν ακριβέστερων προβλέψεων, είναι η κατανόηση και ανάλυση των ιδιαίτερων χαρακτηριστικών κάθε χρονοσειράς, αλλά και η επιλογή των κατάλληλων μέτρων σύγκρισης, προκειμένου να αποφανθούμε για το κατά πόσο η πρόβλεψη που παράγεται είναι αποδεκτή. Βασικό εργαλείο για αυτό το σκοπό είναι τα διάφορα σφάλματα τα οποία και θα εξεταστούν στο συγκεκριμένο κεφάλαιο, ενώ θα γίνει αναφορά και στα ποιοτικά αλλά και ποσοτικά χαρακτηριστικά των χρονοσειρών.

Τα βασικά χαρακτηριστικά που προκύπτουν από την αποσύνθεση μιας χρονοσειράς είναι τα εξής:

- Τάση
- Κυκλικότητα
- Εποχικότητα
- Ασυνέχειες - Μη κανονικές Διακυμάνσεις

Στη συνέχεια ακολουθεί μια πιο αναλυτική περιγραφή σχετικά με τη σημασία κάθε ποιοτικού χαρακτηριστικού και πώς αυτό επηρεάζει την διαδικασία πρόβλεψης μιας χρονοσειράς.

Τάση (Trend)

Το μέγεθος της τάσης ορίζεται ως μία «μακροπρόθεσμη» μεταβολή του μέσου επιπέδου των τιμών μιας χρονοσειράς. Όμως θα πρέπει να εξετάσουμε πότε μπορεί να θεωρηθεί μία μεταβολή ως μακροπρόθεσμη. Για να προκύψει ένα ασφαλές συμπέρασμα σχετικά με την τάση μιας χρονοσειράς, είναι απαραίτητος κατάλληλος αριθμός δεδομένων και παράλληλα κατάλληλο μήκος για την περίοδο μέσα στην οποία θα αναζητήσουμε την ύπαρξη τάσης.

Κοκλικότητα (Cycle)

Το συγκεκριμένο μέγεθος εκφράζεται με μια κυματοειδή μεταβολή που συνήθως οφείλεται σε ειδικές εξωγενείς συνθήκες και εμφανίζεται κατά περιόδους. Οι περίοδοι δεν είναι πάντοτε σταθερές και τις περισσότερες φορές έχουν μήκος μεγαλύτερο του έτους.

Εποχικότητα (Seasonality)

Ως εποχικότητα ορίζεται η παρουσία μίας περιοδικής διακύμανσης, συνήθως κατανοητής και προβλέψιμης, με σταθερό και μικρότερο του έτους, μήκος. Ουσιαστικά πρόκειται για αλλαγές που επαναλαμβάνονται με τον ίδιο ακριβώς τρόπο μέσα στο χρόνο. Για παράδειγμα, η κατανάλωση ενέργειας είναι αυξημένη τους καλοκαιρινούς μήνες λόγω της χρήσης κλιματιστικών. Επομένως η απομόνωση του εποχιακού παράγοντα αποτελεί τις περισσότερες φορές εύκολο εγχείρημα, με εξαίρεση ίσως τις χρονοσειρές που παρουσιάζουν πολλαπλή εποχικότητα, όπως για παράδειγμα στην περίπτωση μας η ΟΤΣ. Η χρονοσειρά που προκύπτει απαλείφοντας τον εποχιακό παράγοντα καλείται από εποχικοποιημένη χρονοσειρά.

Ασυνέχειες – Μη κανονικές Διακυμάνσεις

Ασυνέχειες καλούνται ορισμένες απομονωμένες παρατηρήσεις που παρουσιάζονται στο γράφημα μίας χρονοσειράς και ξεφεύγουν από το γενικό πρότυπο συμπεριφοράς της.

Τέτοιου είδους τιμές δεν είναι δυνατό να προβλεφθούν με βάση τα ιστορικά δεδομένα και μπορούν να έχουν είτε μόνιμο είτε παροδικό χαρακτήρα.

Εάν έχουν παροδικό χαρακτήρα ονομάζονται outliers ή special events ή spikes και επιδρούν στη χρονοσειρά για μικρό χρονικό διάστημα. Αυτού του είδους οι ασυνέχειες είναι δυνατό να εμφανιστούν λόγω κάποιου απρόβλεπτου εξωτερικού γεγονότος. Στην περίπτωση που έχουν μόνιμο χαρακτήρα καλούνται level - shifts, αφού εμφανίζονται ως απότομες αλλαγές στο επίπεδο της χρονοσειράς.

Οι μη κανονικές διακυμάνσεις, είναι εκείνες οι οποίες απομένουν μετά την απομόνωση όλων των λοιπών χαρακτηριστικών μιας χρονοσειράς. Οι διακυμάνσεις αυτές μπορεί να εκφράζονται είτε με τον τυχαίο παράγοντα μιας στοχαστικής διαδικασίας, είτε με κάποια ασυνέχεια που οφείλεται σε κάποιο ιδιαίτερο γεγονός.

4.2.2 Ποσοτικά Χαρακτηριστικά των Χρονοσειρών

Πρόκειται για ορισμένους βασικούς στατιστικούς δείκτες, οι οποίοι μπορούν να υπολογιστούν εύκολα για οποιαδήποτε δεδομένη χρονοσειρά με συγκεκριμένο πλήθος παρατηρήσεων, και είναι οι ακόλουθοι:

Μέση τιμή (Average)

Είναι ο απλός γραμμικός μέσος όρος των τιμών των παρατηρήσεων μιας χρονοσειράς και υποδεικνύει το επίπεδο γύρω από το οποίο κινούνται οι πραγματικές τιμές της χρονοσειράς. Ο υπολογισμός του

συγκεκριμένου δείκτη μπορεί να γίνει με χρήση του ακόλουθου τύπου:

$$\bar{Y} = \frac{1}{n} \cdot \sum_{j=1}^n Y_j$$

Όπου,

- Y_j είναι η j-οστή παρατήρηση
- n το πλήθος των παρατηρήσεων

Μέγιστη και Ελάχιστη Τιμή (Maximum and Minimum)

Πρόκειται για τις ακραίες τιμές μιας χρονοσειράς και μπορούν να αποτελούν μία εκτίμηση της διακύμανσης των δεδομένων καθώς και της τυχαιότητάς τους. Συνήθως, όταν η εξεταζόμενη χρονοσειρά αποτελείται από μεγάλο πλήθος δεδομένων, η μέγιστη και ελάχιστη τιμή της υπολογίζονται μέσω υπολογιστικού φύλλου με τις συναρτήσεις max και min αντίστοιχα.

Τοπική Απόκλιση (Standard Deviation)

Ο συγκεκριμένος στατιστικός δείκτης αναφέρεται και ως διασπορά και αντιπροσωπεύει το κατά πόσο οι παρατηρήσεις είναι διεσπαρμένες γύρω από τη μέση τιμή. Αυτός ο δείκτης μπορεί να υπολογιστεί με χρήση των παρακάτω σχέσεων για δείγμα ή πληθυσμό αντίστοιχα:

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

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

Συντελεστής Γραμμικής Συσχέτισης (Linear Correlation Coefficient)

Είναι γνωστός και ως συντελεστής του Pearson και εκφράζει το κατά πόσο είναι συγκεντρωμένες οι τιμές ενός διαγράμματος διασποράς γύρω από την ευθεία παλινδρόμησης δίνοντας παράλληλα ένα μέτρο της γραμμικής συσχέτισης μεταξύ των δύο υπό εξέταση μεταβλητών. Ο εν λόγω δείκτης μπορεί να πάρει τιμές στο διάστημα [-1, 1]. Όταν ο συντελεστής παίρνει τις ακραίες τιμές 1 υπάρχει τέλεια γραμμική συσχέτιση μεταξύ των δύο μεταβλητών. Όσο η τιμή του συντελεστή λαμβάνει τιμές κοντά στις ακραίες, τόσο μεγαλύτερη είναι η γραμμική συσχέτιση, ενώ στην αντίθετη περίπτωση η γραμμική συσχέτιση θεωρείται ασθενής. Ο υπολογισμός του συντελεστή γραμμικής συσχέτισης μπορεί να γίνει με χρήση του ακόλουθου τύπου, θεωρώντας ως μία μεταβλητή το χρόνο:

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

Συντελεστής Αυτοσυσχέτισης (Autocorrelation Coefficient)

Υποδεικνύει τη συσχέτιση μεταξύ παρατηρήσεων της ίδιας μεταβλητής με χρονική υστέρηση k περιόδους.

Ο συγκεκριμένος συντελεστής παίρνει τιμές στο διάστημα $[0, 1]$, και αν λάβει τιμή κοντά στο μηδέν, θεωρείται πως η συσχέτιση των παρατηρήσεων χρονικής υστέρησης k είναι μηδενική, ενώ αν λάβει τιμή κοντά στη μονάδα, τα δεδομένα μας παρουσιάζουν υψηλή συσχέτιση. Πρόκειται για ένα πολύ χρήσιμο στατιστικό δείκτη, ο οποίος χρησιμοποιείται σε περιπτώσεις που απαιτείται αλγεβρικός εντοπισμός της εποχιακής συμπεριφοράς μιας χρονοσειράς. Μπορεί να υπολογιστεί με χρήση του παρακάτω τύπου:

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

Κριτήριο AIC (Akaike Information Criterion)

Το κριτήριο πληροφοριών Akaike (AIC) είναι εκτιμητής του σφάλματος πρόβλεψης και ως εκ τούτου σχετικής ποιότητας στατιστικών μοντέλων για ένα δεδομένο σύνολο δεδομένων. Λαμβάνοντας υπόψη μια συλλογή μοντέλων για τα δεδομένα, το AIC εκτιμά την ποιότητα κάθε μοντέλου, σε σχέση με κάθε ένα από τα άλλα μοντέλα. Έτσι, το AIC παρέχει ένα μέσο για την επιλογή μοντέλου.

Το AIC βασίζεται στη θεωρία της πληροφορίας. Όταν ένα στατιστικό μοντέλο χρησιμοποιείται για την αναπαράσταση της διαδικασίας που δημιούργησε τα δεδομένα, η αναπαράσταση σχεδόν ποτέ δεν θα είναι ακριβής. οπότε ορισμένες πληροφορίες θα χαθούν χρησιμοποιώντας το μοντέλο για να αντιπροσωπεύσει τη διαδικασία. Το AIC εκτιμά τη σχετική ποσότητα πληροφοριών που χάθηκαν από ένα δεδομένο μοντέλο: όσο λιγότερες πληροφορίες χάνει ένα μοντέλο, τόσο υψηλότερη είναι η ποιότητα αυτού του μοντέλου.

Κατά την εκτίμηση του όγκου των πληροφοριών που χάθηκαν από ένα μοντέλο, η AIC ασχολείται με την ανταλλαγή μεταξύ της καλής εφαρμογής του μοντέλου και της απλότητας του μοντέλου. Με άλλα λόγια, η AIC ασχολείται τόσο με τον κίνδυνο υπερ-εφαρμογής (overfitting) όσο και με τον κίνδυνο υπό-εφαρμογής (underfitting).

Υποθέτοντας ένα στατιστικό μοντέλο ορισμένων δεδομένων, έστω k ο αριθμός των εκτιμώμενων παραμέτρων στο μοντέλο και \hat{L} να είναι η μέγιστη τιμή της συνάρτησης πιθανότητας για το μοντέλο. Η τιμή AIC του μοντέλου είναι η ακόλουθη:

$$AIC = 2k - 2 \ln(\hat{L})$$

Δεδομένου ενός συνόλου υποψήφιων μοντέλων για τα δεδομένα, το προτιμώμενο μοντέλο είναι αυτό με την ελάχιστη τιμή AIC. Έτσι, το AIC ανταμείβει την καλή εφαρμογή (όπως εκτιμάται από τη συνάρτηση πιθανότητας), αλλά περιλαμβάνει επίσης μια ποινή που είναι μια αυξανόμενη συνάρτηση του αριθμού των εκτιμώμενων παραμέτρων. Η ποινή αποθαρρύνεται από το overfitting, κάτι που είναι επιθυμητό επειδή η αύξηση του αριθμού των παραμέτρων στο μοντέλο βελτιώνει σχεδόν πάντα το fit των δεδομένων της εφαρμογής.

4.2.3 Στασιμότητα χρονοσειράς

Προτού εφαρμοστεί κάποιο μοντέλο στα δεδομένα, θα πρέπει να εξακριβωθεί αν η χρονοσειρά τμηματικά και συνολικά είναι στάσιμη. Με τον όρο στασιμότητα λογίζεται η ιδιότητα εκείνη της χρονοσειράς κατά την οποία οι διακυμάνσεις των τιμών της δε διαφοροποιούνται με το χρόνο. Η στασιμότητα διακρίνεται σε αυστηρή και ασθενή και συγκεκριμένα ορίζονται ως εξής:

Ορισμός Αυστηρής στασιμότητας

Μια στοχαστική διαδικασία είναι αυστηρώς στάσιμη (strictly stationary) όταν οι ιδιότητές της δεν επηρεάζονται από μια αλλαγή στην αρχή μέτρησης του χρόνου. Αυτό σημαίνει ότι η συνδυασμένη συνάρτηση πιθανότητας με αρχή το χρονικό σημείο t δηλαδή η $f(\psi_t, \psi_{t+1}, \dots, \psi_{t+h})$ είναι ακριβώς η ίδια με τη συνδυασμένη συνάρτηση πιθανότητας με αρχή το χρονικό σημείο $t+S$, την $f(\psi_{t+S}, \psi_{t+1+S}, \dots, \psi_{t+S+T})$. Το s παριστάνει μια αυθαίρετη μετακίνηση κατά μήκος του άξονα του χρόνου είτε προς τα εμπρός είτε προς τα πίσω, δηλαδή μπορεί να είναι είτε θετικό είτε αρνητικό.

Οπότε, από τη στιγμή που δεν μεταβάλλεται η συνάρτηση πιθανότητας με το χρόνο, δεν θα μεταβάλλεται ούτε η περιθωριακή συνάρτηση πιθανότητας και το ίδιο θα ισχύει και για όλες τις διμεταβλητές συναρτήσεις πιθανότητας. Όλα αυτά συνεπάγονται ότι ο μέσος και η διακύμανση του ψ_t δεν μεταβάλλονται με μια αλλαγή του χρόνου, ενώ οι συνδιακυμάνσεις θα είναι συναρτήσεις μόνο της υστέρησης s .

Οπότε, θα δίνονται από τους τύπους:

$$E(\psi_1) = E(\psi_2) = \dots = E(\psi_T) = E(\psi_t) = \mu$$

$$V(\psi_1) = V(\psi_2) = \dots = V(\psi_T) = V(\psi_t) = \sigma^2$$

$$\text{cov}(\psi_1, \psi_{1+s}) = \text{cov}(\psi_2, \psi_{2+s}) = \dots = \text{cov}(\psi_T, \psi_{T+s}) = \gamma_s$$

Ορισμός Ασθενούς στασιμότητας

Ο αυστηρός ορισμός της στασιμότητας αναφέρεται σε όλες τις ιδιότητες μιας στοχαστικής διαδικασίας, γι' αυτό όταν ικανοποιούνται μόνο οι παραπάνω συνθήκες, η στοχαστική διαδικασία χαρακτηρίζεται ασθενώς στάσιμη (weakly stationary). Για την περαιτέρω ανάλυσή μας θα είναι αρκετό μια χρονολογική σειρά να είναι ασθενώς στάσιμη. Δηλαδή, αρκεί να ισχύουν τα εξής:

$$E(\psi_t) = \mu, \text{ ανεξάρτητη από το } t$$

$$V(\psi_t) = \sigma^2, \text{ ανεξάρτητη από το } t$$

$$\text{cov}(\psi_1, \psi_{1+s}) = \text{cov}(\psi_t, \psi_{t+m}) = \gamma, \text{ ανεξάρτητη από το } t$$

Για την εξακρίβωση της στασιμότητας της συνολικής, αρχικά, χρονοσειράς θα χρησιμοποιήσουμε το Επαυξημένο κριτήριο των Dickey-Fuller ή κριτήριο μοναδιαίας ρίζας.

Έλεγχος των Dickey-Fuller (DF)

Οι Dickey - Fuller (1981) μέσω των πειραμάτων Monte - Carlo βρήκαν μια κατάλληλη ασύμμετρη κατανομή που χρησιμοποίησαν για τον έλεγχο της υπόθεσης $H_0: \rho = 1$. Την κατανομή αυτή μπορούμε να χρησιμοποιήσουμε για να ξεχωρίσουμε ένα AR(1) υπόδειγμα από μια ολοκληρωμένη σειρά, δηλαδή την ύπαρξη μοναδιαίας ρίζας I(1). Ο έλεγχος των Dickey - Fuller (DF) βασίζεται στην εμπειρική τιμή t - statistics από μια απλή παλινδρόμηση, όμως η σύγκριση για την αποδοχή ή όχι της H_0 δεν γίνεται με τιμές από την t -κατανομή αλλά με τιμές που έχουν προσδιοριστεί εμπειρικά από τον MacKinnon(1991). Αφαιρώντας το Y_{t-1} από τα δύο μέλη της εξίσωσης:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t$$

$$Y_t - Y_{t-1} = (\rho - 1)Y_{t-1} + u_t$$

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

όπου $\delta = \rho - 1$.

Επομένως, ο έλεγχος για την ύπαρξη μοναδιαίας ρίζας ($\rho = 1$ ή $|\rho| < 1$) μετατρέπεται σε έλεγχο της παραμέτρου δ . Ειδικότερα :

H_0 : $\delta = 0$ η διαδικασία Y_t είναι μη στάσιμη. (υπάρχει μοναδιαία ρίζα)

H_1 : $\delta < 0$ η διαδικασία H_0 είναι στάσιμη. (δεν υπάρχει μοναδιαία ρίζα)

Η μηδενική υπόθεση απορρίπτεται όταν η εμπειρική τιμή της t-statistics για την παράμετρο δ (t_δ) έχει μικρότερη από την κριτική τιμή των πινάκων του το στατιστικό t - student του συντελεστή δ_2 είναι μικρότερο ($t_\delta < \tau_1$) από την κρίσιμη τιμή τ_1 των πινάκων του MacKinnon (1991).

Σε πολλές περιπτώσεις είναι πιθανόν η χρονική σειρά που εξετάζουμε να έχει και κάποιο σταθερό όρο, δηλαδή να συμπεριφέρεται σαν ένα υπόδειγμα τυχαίου περιπάτου με τάση (drift):

$$\Delta Y_t = \alpha + \delta Y_{t-1} + u_t$$

Επίσης, υπάρχουν περιπτώσεις που στο υπόδειγμα εισάγεται μια προσδιοριστική τάση, δηλαδή X_t :

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + u_t$$

και η μηδενική υπόθεση εδώ είναι ότι η μη - στασιμότητα προέρχεται τόσο από την ύπαρξη στοχαστικής τάσης (unit root) όσο και από την ύπαρξη προσδιοριστικής τάσης. Αν με τον έλεγχο στασιμότητας διαπιστωθεί ότι μια σειρά δεν είναι στάσιμη, τότε σχηματίζουμε τις πρώτες διαφορές και επαναλαμβάνουμε τον έλεγχο. Αν και η νέα σειρά δεν είναι στάσιμη, παίρνουμε τις δεύτερες διαφορές και επανελέγχουμε.

Την ίδια διαδικασία την επαναλαμβάνουμε μέχρι να προκύψει στάσιμη σειρά έστω μετά από d φορά.

Τότε λέμε έχουμε ολοκληρωμένη σειρά I(d).

Η διαδικασία αυτή για να επιτευχθεί στασιμότητα, την ονομάζουμε στασιμότητα διαφορών (difference stationarity).

Επαυξημένος Έλεγχος των DICKKEY-FULLER (DF)

Ο προηγούμενος έλεγχος για την ύπαρξη μοναδιαίας ρίζας σε ένα Αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως μπορεί να εφαρμοστεί και στη γενική περίπτωση μιας AR(p) διαδικασίας.

Η γενική μορφή ενός AR(p) υποδείγματος μπορεί να διατυπωθεί ως εξής:

$$Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \dots + \rho_p Y_{t-p} + u_t$$

ή μετροποποιημένη μορφή :

$$\Delta Y_t = \delta Y_{t-1} + \rho_1 \Delta Y_{t-1} + \rho_2 \Delta Y_{t-2} + \dots + \rho_p \Delta Y_{t-p} + u_t$$

Όπου $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$, $\Delta Y_{t-2} = Y_{t-2} - Y_{t-3}$,

Και $\delta = (\rho_1 + \rho_2 + \dots + \rho_p - 1)$ και οι συντελεστές δ^* αποτελούν γενικές συναρτήσεις για τους συντελεστές ρ . Ο έλεγχος για την ύπαρξη μοναδιαίας ρίζας, ότι δηλαδή η σειρά δεν είναι στάσιμη, ισοδυναμεί με έλεγχο της μηδενικής υπόθεσης $H_0: \delta = 0$ έναντι της εναλλακτικής $H_1: \delta < 0$.

Ο έλεγχος της παραπάνω υπόθεσης εκτιμάται με την μέθοδο των ελαχίστων τετραγώνων. Για να γίνει ο παραπάνω έλεγχος, πρέπει να είναι γνωστή η τάξη (p) της αυτοπαλίνδρομης διαδικασίας που βεβαίως δεν είναι γνωστή εκ των προτέρων.

Στην πράξη αυτό που γίνεται είναι να προστίθενται τόσες υστερήσεις της Y_t ώστε τα κατάλοιπα που προκύπτουν από την μέθοδο των ελαχίστων τετραγώνων να μην αυτοσυσχετίζονται. Αντίστοιχα όπως και στον απλό έλεγχο μπορούμε να προσθέσουμε και σταθερό όρο και μεταβλητή χρονικής τάσης. Για την επιλογή του κατάλληλου αριθμού υστερήσεων χρησιμοποιούμε τα κριτήρια Akaike (AIC), Schwartz (SCH) ή (LM). Οι κριτικές τιμές είναι ίδιες με αυτές που χρησιμοποιούνται στον απλό (μη επαυξημένο) έλεγχο Dickey – Fuller.

Στη συγκεκριμένη περίπτωση, ο παραπάνω έλεγχος πραγματοποιήθηκε μέσω της βιβλιοθήκης statsmodels.tsa.stattools με βάση το κριτήριο Akaike (AIC). Τα συνολικά δεδομένα φορτώθηκαν σαν numpy array και το στατιστικό στοιχείο δοκιμής βρέθηκε χαμηλότερο από την κρίσιμη τιμή (p-value) που εμφανίζεται, οπότε απορρίφθηκε μηδενική υπόθεση και συνάγεται ότι η χρονοσειρά είναι στάσιμη.

Το αποτέλεσμα από τον έλεγχο για τα συνολικά δεδομένα φαίνεται παρακάτω:

Σχ.4.4

```
ADF Statistic: -5.869317
p-value: 0.000000
Critical Values:
 1%: -3.431
 5%: -2.862
10%: -2.567
Reject Ho - Time Series is Stationary
```

Θεωρήθηκε σκόπιμο να γίνει ο ίδιος έλεγχος και στο χρονικό διάστημα των δύο πρώτων ετών καθώς είναι το διάστημα πάνω στο οποίο θα δοκιμαστούν τα μοντέλα πρόβλεψης παρακάτω. Το αποτέλεσμα που προέκυψε εκ νέου είναι ότι και αυτή η χρονοσειρά είναι στάσιμη, στοιχείο ιδιαίτερα χρήσιμο για τη μετέπειτα ανάλυση.

Ακολουθούν τα αποτελέσματα:

Σχ.4.5

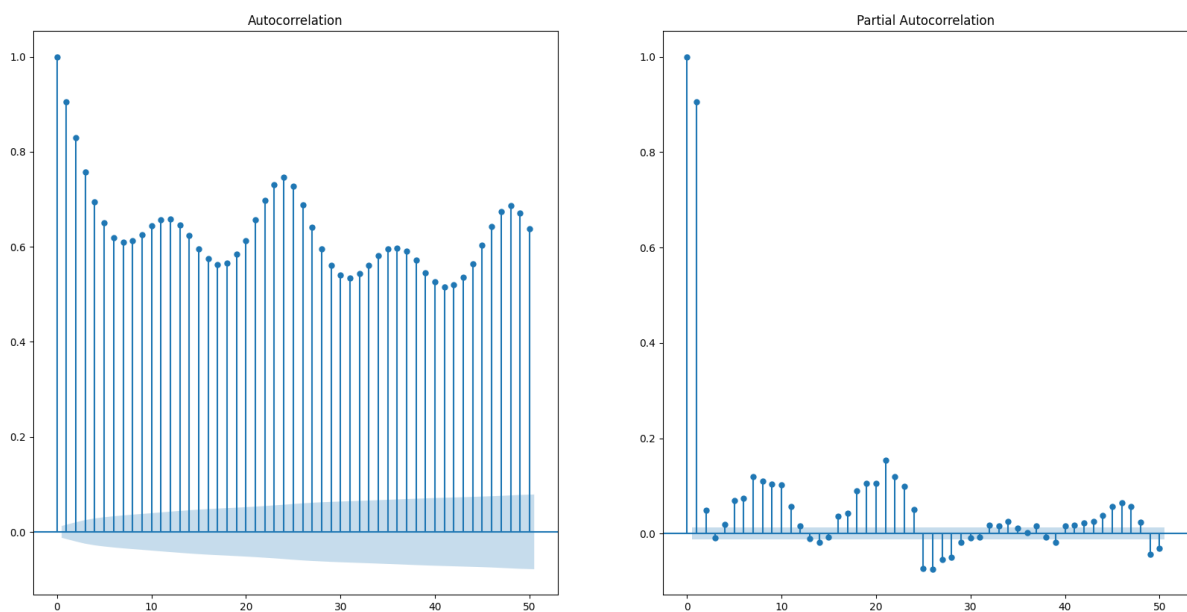
```
ADF Statistic: -7.719552
p-value: 0.000000
Critical Values:
 1%: -3.431
 5%: -2.862
10%: -2.567
Reject Ho - Time Series is Stationary
```

4.2.4 Μελέτη αυτοσυσχέτισης

Από τα βασικά στάδια ελέγχου μιας χρονοσειράς αποτελεί η μελέτη των υπολοίπων (residuals) και η επίδραση τους στην εφαρμογή ενός μοντέλου πρόβλεψης. Πιο συγκεκριμένα, η μελέτη αυτοσυσχέτισης των δεδομένων της χρονοσειράς είναι χρήσιμο εργαλείο στην ανίχνευση patterns και συνεπώς της επιλογής της κατάλληλης μεθόδου πρόβλεψης. Με τον όρο patterns, εννοείται ότι μπορεί να υπάρξει μία εικόνα για την τάση και την εποχικότητα της χρονοσειράς στοιχεία που θα βοηθήσουν στην επιλογή μοντέλου πρόβλεψης στη συνέχεια.

Για τον υπολογισμό και την εικονική αναπαράσταση της αυτοσυσχέτισης των δεδομένων, χρησιμοποιήθηκε η βιβλιοθήκη statsmodels (tsa.stattools και graphics.tsaplots) και τα διαγράμματα ολικής και μερικής αυτοσυσχέτισης φαίνονται παρακάτω:

Σχ. 4.6



Όπως φαίνεται στο πρώτο διάγραμμα της αυτοσυσχέτισης, υπάρχει ένας γραμμοσκιασμένος κώνος, ο οποίος εκφράζει τα όρια της περιοχής εμπιστοσύνης που εν προκειμένω είναι 95%. Η ερμηνεία που μπορεί να δοθεί στο για τις τιμές αυτές που ξεπερνούν κατά πολύ την επιφάνεια αυτή, είναι ότι υπάρχει 95% πιθανότητα να έχουν σοβαρή επίδραση στην τιμή των επόμενων τιμών της χρονοσειράς.

Σημειώνεται εδώ ότι έχουμε χρησιμοποιήσει τις πρώτες 50 τιμές (lags) για λόγους ευχρηστίας. Για να ερμηνευθεί λίγο καλύτερα το ποιοτικό συμπέρασμα, επισημαίνονται τα εξής:

- Το πρώτο σημείο (με δείκτη 0) έχει ύψος 1. Στην πραγματικότητα το πρώτο είναι πάντα 1 πράγμα λογικό, καθώς η τρέχουσα τιμή εξηγεί πάντα την τρέχουσα τιμή.
- Το δεύτερο σημείο είναι περίπου 0,9 που σημαίνει ότι το ακόλουθο σημείο (ακριβώς δίπλα) περιγράφεται στο 90% από την προηγούμενη τιμή.
- Το δέκατο σημείο έχει ύψος 0,65. Αυτό σημαίνει ότι τα τρέχοντα δεδομένα θα επηρεάσουν τα δεδομένα σε 10 ώρες κατά 65% (η ανάλυση δεδομένων είναι ωριαία).

Κάτι άλλο που παρατηρούμε εδώ είναι μια σημαντική αύξηση στο 12ο σημείο (13ο καθώς ξεκινά στο μηδέν). Αυτό υποδηλώνει έναν κύκλο μήκους 12. Δεν αποτελεί έκπληξη και ίσως προϋδεάζει για ένα μοτίβο περιοδικότητας που θα εξακριβωθεί αργότερα. Επίσης, όπως λογικά εξάγεται από το διάγραμμα και αποδεικνύεται και μαθηματικά, υπάρχει η τάση η αυτοσυσχέτιση να τείνει προς το μηδέν με την πάροδο του χρόνου γεγονός το οποίο ερμηνεύεται με το ότι όσο παλιότερα δεδομένα, τόσο μικρότερο αντίκτυπο έχουν.

Το διάγραμμα PACF είναι μια γραφική παράσταση των μερικών συντελεστών συσχέτισης μεταξύ της σειράς και των καθυστερήσεών της. Γενικά, η «μερική» συσχέτιση μεταξύ δύο μεταβλητών είναι το ποσοστό συσχέτισης μεταξύ τους, το οποίο δεν εξηγείται από τους αμοιβαίους συσχετισμούς τους με ένα καθορισμένο σύνολο άλλων μεταβλητών. Με άλλα λόγια, η μερική αυτοσυσχέτιση στο lag k είναι η συσχέτιση που προκύπτει μετά την αφαίρεση της επίδρασης οποιωνδήποτε συσχετισμών λόγω των όρων σε μικρότερες καθυστερήσεις.

4.3 Αποσύνθεση χρονοσειράς (decomposing)

Βασική διαδικασία κατά την ανάλυση χρονοσειρών, αποτελεί αυτή της αποσύνθεσης (decomposing) που στόχο έχει την απομόνωση της εποχικότητας και της τυχαιότητας. Με αυτό τον τρόπο, προκύπτει η σειρά τάσης - κύκλου κάθε αρχικής χρονοσειράς, στην οποία εφαρμόζονται εν συνεχεία οι επιλεγμένες μέθοδοι πρόβλεψης και οι τιμές που προκύπτουν από την εκάστοτε πρόβλεψη εποχικοποιούνται ξανά, ώστε να ληφθεί το τελικό αποτέλεσμα.

Με τη διαδικασία αυτή επιτυγχάνεται αύξηση της αξιοπιστίας των αποτελεσμάτων της πρόβλεψης και δίνεται στο χρήστη η δυνατότητα να εισάγει και τη δική του κρίση, βάσει της εμπειρίας και της διαίσθησής του. Υπάρχουν διάφορες μέθοδοι που δοκιμάστηκαν και φαίνονται παρακάτω.

Οι δύο απλούστερες διατυπώσεις της αποσύνθεσης είναι η προσθετική και η πολλαπλασιαστική, με τη μορφή που παρουσιάζεται στις επόμενες δύο σχέσεις:

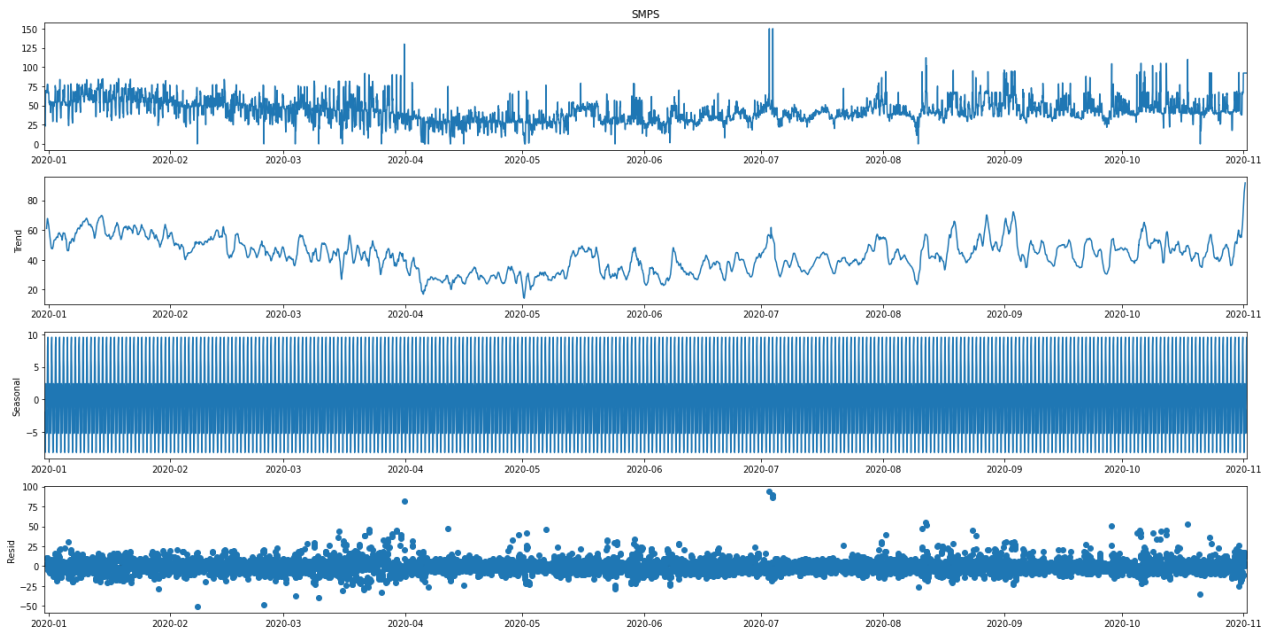
$$Y_t = S_t + T_t + C_t + R_t, \text{ (προσθετικό μοντέλο)}$$

$$Y_t = S_t * T_t * C_t * R_t, \text{ (πολλαπλασιαστικό μοντέλο)}$$

Όπου S_t η περιοδικότητα, T_t η τάση, C_t η κυκλικότητα και R_t η τυχαιότητα. Παρακάτω φαίνονται παραστατικά οι διατυπώσεις της αποσύνθεσης που για λόγους ευκρίνειας και ευχρηστίας αποτυπώνουν δεδομένα ενός έτους (του 2020).

Προσθετικό μοντέλο (*additive naive decomposition*)

Σχ. 4.7



Από το παραπάνω διάγραμμα, δεν μπορούν να βγουν και πολύ χρήσιμα συμπεράσματα για την εποχικότητα των δεδομένων.

Πολλαπλασιαστικό μοντέλο (*multiplicative naive decomposition*)

Δεδομένου, ότι το σύνολο δεδομένων περιέχει μηδενικές τιμές, είναι αδύνατον να αποτυπωθεί πολλαπλασιαστικό μοντέλο. Παρόλα αυτά είναι ρεαλιστικό γεγονός να έχει Οριακή Τιμή Συστήματος μηδενική τιμή και χαρακτηριστικό παράδειγμα είναι η 27^η Σεπτεμβρίου 2018 στις 12 το ξημέρωμα, όπου σύμφωνα με το άρθρο του kiefer εκείνη την ώρα η αιχμή της ζήτησης βρισκόταν στα 4454 MW ενώ στο σύστημα ήταν ενταγμένες 12 θερμικές μονάδες, εκ των οποίων οι 8 λιγνιτικές και οι 4 φυσικού αερίου.

Παράλληλα την επίμαχη ώρα (όπως και σε όλη τη διάρκεια της ημέρας) είχαμε μεγάλη παραγωγή αιολικών στα 1721 MW (στη διάρκεια της ημέρας έφτασε μέχρι και τα 2820MW όταν στο σύστημα μπήκαν και φωτοβολταϊκά). Επίσης σύμφωνα με τα στοιχεία της προ ημερησίας είχαμε εξαγωγές (6,757 GWh), οι οποίες μάλιστα όρισαν για αρκετές ώρες την οριακή τιμή.

STL αποσύνθεση

Η μέθοδος STL είναι μια ευέλικτη μέθοδος για την αποσύνθεση χρονολογικών σειρών, αποτελώντας ακρόνυμο “Seasonal and Trend decomposition using Loess” όπου Loess είναι μια μέθοδος για την εκτίμηση μη γραμμικών σχέσεων. Η μέθοδος STL αναπτύχθηκε από τους R. B. Cleveland, Cleveland, McRae, & Terpenning (1990). Το STL έχει πολλά πλεονεκτήματα έναντι των κλασικών μεθόδων αποσύνθεσης:

- Μπορεί να χειριστεί κάθε είδους εποχικότητα, όχι μόνο μηνιαία και τριμηνιαία δεδομένα.
- Το εποχιακό στοιχείο επιτρέπεται να αλλάζει με την πάροδο του χρόνου και ο ρυθμός αλλαγής μπορεί να ελέγχεται από τον χρήστη.
- Η ομαλότητα του κύκλου τάσης μπορεί επίσης να ελεγχθεί από τον χρήστη.

- Μπορεί να είναι αποτελεσματικό για τα ακραία σημεία (δηλαδή, ο χρήστης μπορεί να καθορίσει μια ισχυρή αποσύνθεση), έτσι ώστε περιστασιακές ασυνήθιστες παρατηρήσεις να μην επηρεάζουν τις εκτιμήσεις του κύκλου τάσης και των εποχιακών στοιχείων.

Από την άλλη πλευρά, το STL έχει κάποια μειονεκτήματα. Συγκεκριμένα, δεν χειρίζεται αυτόματα τις παραλλαγές ημέρας ή ημερολογίου συναλλαγών και παρέχει μόνο διευκολύνσεις για additive αποσυνθέσεις. Εν προκειμένω, φαίνεται η αποσύνθεση στα δεδομένα του 2020:



Hodrick-Prescott αποσύνθεση

Το φίλτρο Hodrick-Prescott διαχωρίζει μια χρονοσειρά Y_t σε ένα στοιχείο τάσης T_t και ένα κυκλικό στοιχείο C_t .

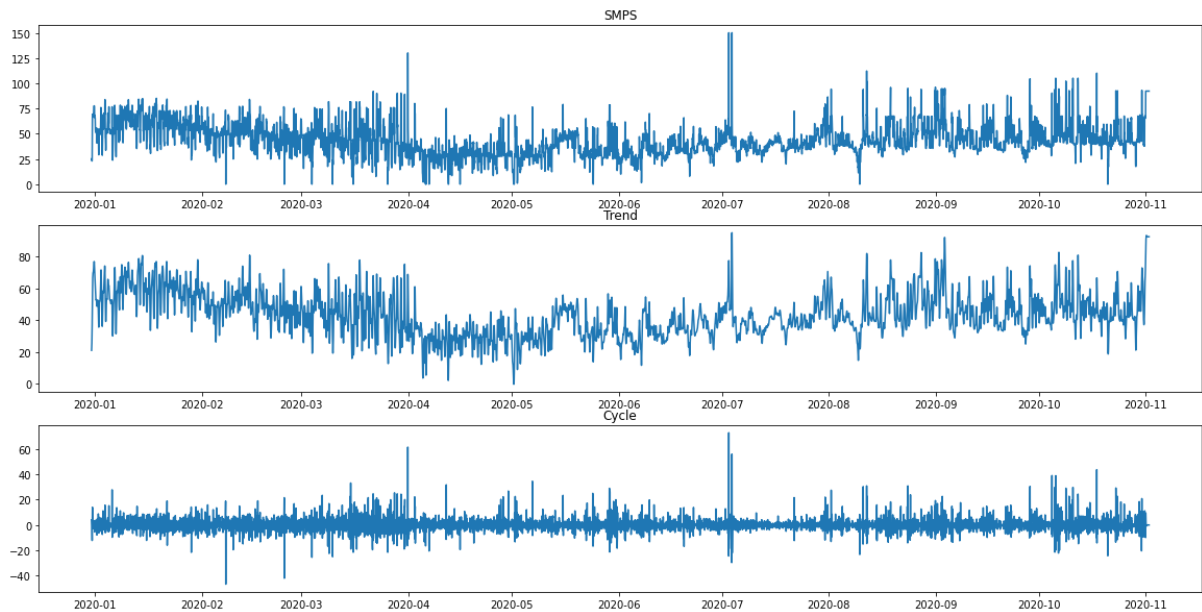
$$Y_t = T_t + C_t$$

Τα στοιχεία προσδιορίζονται ελαχιστοποιώντας την ακόλουθη συνάρτηση τετραγωνικής απώλειας, όπου λ είναι μια παράμετρος εξομάλυνσης:

$$\min_{\tau_t} \sum_{t=1}^T c_t^2 + \lambda \sum_{t=1}^T [(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})]^2$$

Αυτή η τιμή λ χειρίζεται τη διακύμανση του ρυθμού ανάπτυξης της συνιστώσας της τάσης. Προτείνεται τιμή 1600 για τριμηνιαία δεδομένα. Οι Ravn και Uhlig προτείνουν τη χρήση μιας τιμής 6,25 για τα ετήσια δεδομένα και 129600 για τα μηνιαία δεδομένα. Δεν υπάρχει πρόβλεψη για ωριαία δεδομένα και συνεπώς το ακόλουθο διάγραμμα αποσύνθεσης των δεδομένων δεν δείχνει να 'ναι κατάλληλο για συμπεράσματα:

Σχ. 4.9



4.4 Μελέτη εποχικότητας με Fourier

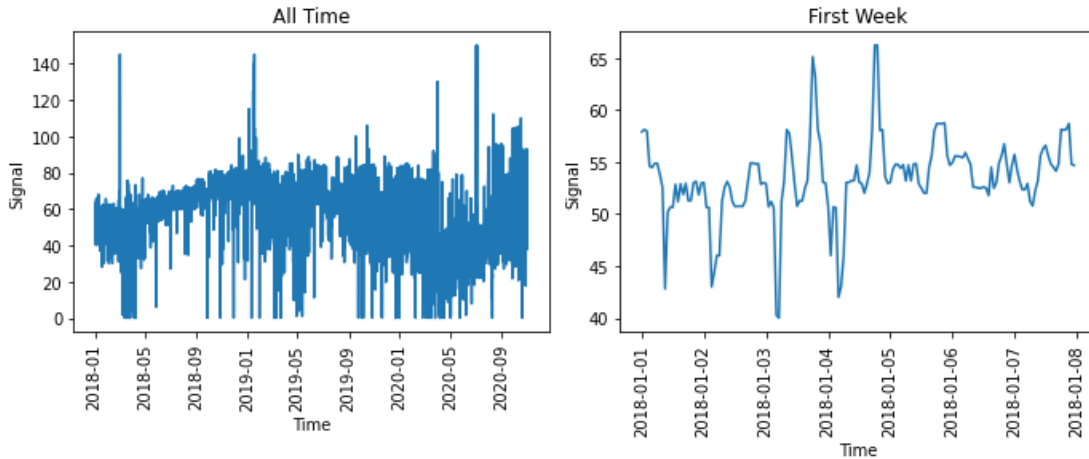
Με βάση τα μέχρι τώρα δεδομένα, θεωρείται πιθανόν τα δεδομένα που εξετάζονται να παρουσιάζουν πολλαπλές εποχικότητες μακροπρόθεσμες ή βραχυπρόθεσμες. Έγινε μια προσπάθεια να αποτυπωθεί με ακρίβεια η περιοδικότητα των δεδομένων μέσω του Ταχέως Μετασχηματισμού Fourier. Ο Ταχύς Μετασχηματισμός Fourier (FFT) είναι ένας αλγόριθμος που υπολογίζει τον διακριτό μετασχηματισμό Fourier (DFT) μιας ακολουθίας ή το αντίστροφο (IDFT). Η ανάλυση Fourier μετατρέπει ένα σήμα από τον αρχικό του τομέα (συχνά χρόνο ή χώρο) σε αναπαράσταση στον τομέα συχνοτήτων και αντιστρόφως. Για να τεθεί απλούστερα, ο μετασχηματισμός Fourier παίρνει δεδομένα βάσει χρόνου, μετρά κάθε πιθανό κύκλο και επιστρέφει τη συνολική «συνταγή κύκλου» (το εύρος, την μετατόπιση και την ταχύτητα περιστροφής για κάθε κύκλο που βρέθηκε). Το DFT λαμβάνεται αποσυνθέτοντας μια ακολουθία τιμών σε συστατικά διαφορετικών συχνοτήτων. Ο ορισμός είναι ο ακόλουθος:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, \dots, N-1$$

όπου x_0, \dots, x_{N-1} μιγαδικοί αριθμοί.

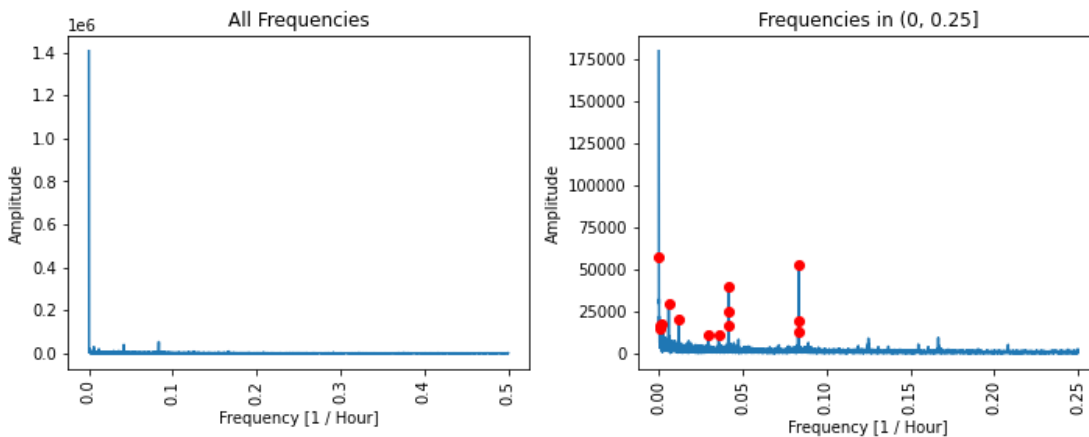
Παρακάτω φαίνεται το διάγραμμα της συνολικής χρονοσειράς των δεδομένων καθώς και τα δεδομένα της πρώτης βδομάδας:

Σχ.4.10



Μέσω της βιβλιοθήκης `scipy`, υλοποιήθηκε ο FFT του συνόλου των δεδομένων και παρακάτω φαίνονται τα αντίστοιχα διαγράμματα στο πεδίο των συχνοτήτων:

Σχ.4.11



Ο ακόλουθος πίνακας περιλαμβάνει τα `reaks`, τις συχνότητες, τα πλάτη των συνιστωσών, οι περίοδοι και οι αντίστοιχες μιγαδικές τιμές του FFT.

Σχ.4.12

[31]:	index	freq (1/hour)	amplitude	period (days)	fft
	0	3	57245.260496	345.666667	52988.894911+21660.952548j
	12	2074	0.083333	0.500000	-6347.640861+52752.359942j
	9	1037	0.041667	1.000000	6197.473962+39615.781176j
	4	148	0.005947	7.006757	-21464.802562-20637.422182j
	8	1034	0.041546	1.002901	-18785.146522+16794.985966j
	5	296	0.011893	3.503378	-5363.447578-19809.882547j
	11	2071	0.083213	0.500724	-19718.6702503-715.3709660j
	3	32	0.001286	32.406250	16188.346040+6135.694158j
	1	18	0.000723	57.611111	14427.783159-8495.235551j
	10	1040	0.041787	0.997115	-8741.810011+13601.525692j
	2	28	0.001125	37.035714	-10639.653357-10645.923361j
	13	2077	0.083454	0.499278	3926.460909+11957.584094j
	7	889	0.035720	1.166479	4244.931878-10496.748375j
	6	741	0.029773	1.399460	-5063.7332830-9973.4550670j

Ο πίνακας αυτός αποτελεί χρήσιμο εργαλείο για να εξαχθούν σοβαρά συμπεράσματα για την εποχικότητα της χρονοσειράς. Πιο συγκεκριμένα, με βάση το διάγραμμα πλάτους - συχνοτήτων καθώς τη στήλη πλάτους των συνιστωσών η οποία είναι ταξινομημένη με φθίνουσα σειρά, θεωρείται ότι οι αντίστοιχες περίοδοι των τριών μεγαλύτερων πλάτων αντικατοπτρίζουν και τις επικρατούσες περιόδους εποχικότητας. Αυτό ισχύει αφού είναι γνωστό από θεωρία ότι οι συχνότητες με το υψηλότερο πλάτος είναι ενδεικτικές των εποχιακών μοτίβων ενώ αυτές με χαμηλό πλάτος είναι θόρυβος.

Συνεπώς, στα δεδομένα του προβλήματος φαίνεται ότι η κύρια μακροχρόνια συνιστώσα εποχικότητας είναι της τάξεως των 346 ημερών και η αμέσως επόμενη είναι βραχυχρόνια της τάξεως των 12 ωρών. Υπάρχει και μια τρίτη ισχυρή συνιστώσα της τάξεως της μίας μέρας (24 ωρών).

Κεφάλαιο 5

5.1 Μοντέλα Πρόβλεψης

Στα πλαίσια της παρούσας διπλωματικής εξετάστηκαν διάφορα μοντέλα πρόβλεψης. Έγινε προσπάθεια να μελετηθούν μοντέλα βραχυπρόθεσμης και μακροπρόθεσμης πρόβλεψης, χρησιμοποιώντας σαν χρονικά διαστήματα αναφοράς από μία έως 2 βδομάδες για τη βραχυπρόθεσμη και έως 11 μήνες για τη μακροπρόθεσμη. Αξίζει επίσης να σημειωθεί ότι τα παραπάνω χρονικά διαστήματα μελετώνται ανηγμένα σε ώρες καθώς τα δεδομένα ΟΤΣ που χρησιμοποιήθηκαν είναι ωριαία.

5.2 Βραχυπρόθεσμη πρόβλεψη

5.2.1 Μοντέλο TBATS

Οι μέθοδοι πρόβλεψης χρονοσειρών που ονομάζονται BATS και TBATS που μπορούν να μοντελοποιήσουν χρονοσειρές με πολλαπλές εποχικότητες.

Τα ονόματα είναι ακρωνύμια των βασικών χαρακτηριστικών των μοντέλων: Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components.

Το μοντέλο TBATS βασίζεται σε εκθετικές μεθόδους εξομάλυνσης και μπορεί να περιγραφεί με τις ακόλουθες εξισώσεις:

Μοντέλο:

$$y_t^{(\lambda)} = I_{t-1} + \varphi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

$$I_t = I_{t-1} + \varphi b_{t-1} + a d_t$$

$$b_t = \varphi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t$$

όπου $y_t^{(\lambda)}$ – η τιμή της χρονοσειράς τη χρονική στιγμή t

$s_t^{(i)}$ – η i -οστή εποχική συνιστώσα

I_t – τοπικό επίπεδο

b_t – τάση με απόσβεση

d_t – ARMA διαδικασία με υπόλοιπα

e_t – Gaussian θόρυβος

Εποχικό κομμάτι:

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos(\omega_i) + s_{j,t-1}^{*(i)} \sin(\omega_i) + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin(\omega_i) + s_{j,t-1}^{*(i)} \cos(\omega_i) + \gamma_2^{(i)} d_t$$

$$\omega_i = 2\pi j/m_i$$

όπου T – το πλήθος των εποχικότητων

m_i – μήκος της i -οστής εποχικότητας

k_i – πλήθος αρμονικών για την i -οστή εποχικότητα

λ – Μετασχηματισμός Box-Cox

α, β – συντελεστές εξομάλυνσης

ϕ – τάση με απόσβεση

φ_i, θ_i – συνιστώσες εποχικότητας ARMA(p,q)

$\gamma_1^{(i)}, \gamma_2^{(i)}$ – εποχικές εξομαλύνσεις (2 για κάθε περίοδο)

Κάθε εποχικότητα διαμορφώνεται από μια τριγωνομετρική αναπαράσταση βασισμένη στη σειρά Fourier. Ένα σημαντικό πλεονέκτημα αυτής της προσέγγισης είναι ότι απαιτεί μόνο 2 seed states ανεξάρτητα από τη διάρκεια της περιόδου. Σε αλγόριθμους που περιλαμβάνουν γεννήτριες τυχαίων αριθμών ή διανύσματα αυτών, τα seeds αποτελούν σημεία εκκίνησης για την παραγωγή μιας ακολουθίας αριθμών. Στην ουσία αποτελεί εργαλείο reproducibility.

Ένα άλλο πλεονέκτημα είναι η ικανότητα μοντελοποίησης εποχιακών συνιστωσών μη ακέραιου μήκους. Για παράδειγμα, δεδομένης μιας σειράς καθημερινών παρατηρήσεων, μπορεί κανείς να μοντελοποιήσει τα άλματα του χρόνου με μια εποχή 365,25.

Στο βάση εκτέλεσης της, η μέθοδος TBATS θα εξετάσει διάφορες εναλλακτικές λύσεις και μοντέλα. Θα εξετάσει μοντέλα:

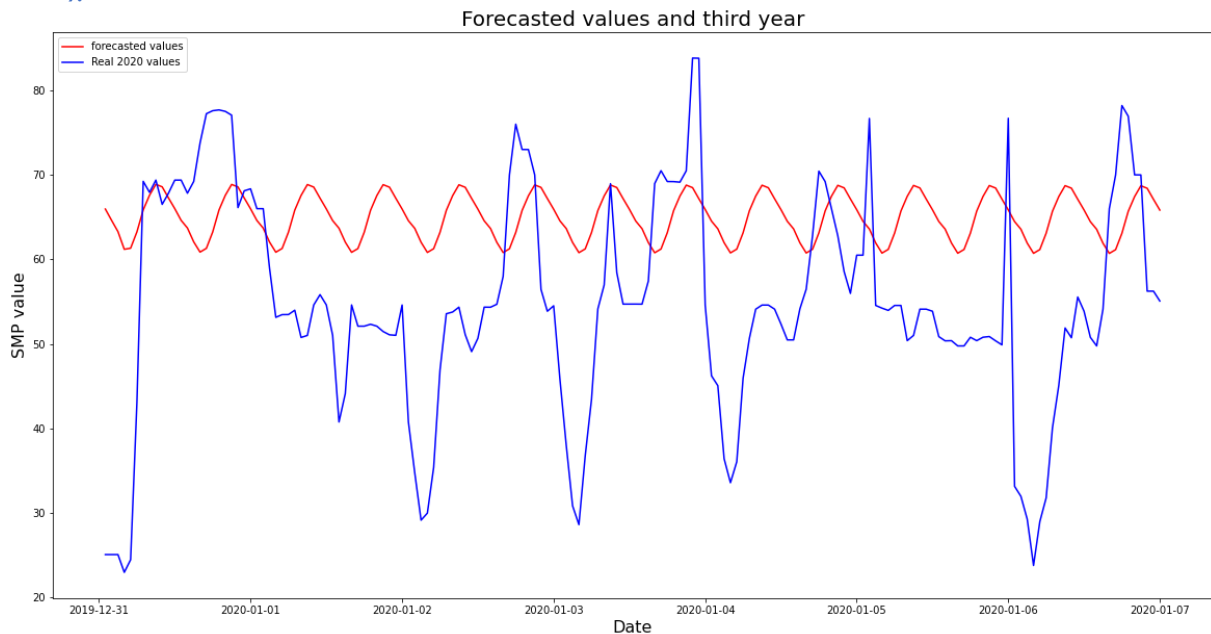
- Με μετασχηματισμό BOX-COX και χωρίς
- Με και χωρίς Τάση
- Με και χωρίς απόσβεση τάσης
- Με και χωρίς μοντέλο ARMA(p,q) για να μοντελοποιήσει τα υπόλοιπα residuals
- Μη εποχικά μοντέλα
- Διάφορα μοντέλα αρμονικών για να μοντελοποιήσει τις εποχικότητες

Το τελικό-βέλτιστο μοντέλο θα καθοριστεί με βάση το Akaike information criterion (AIC) και η εφαρμογή του φαίνεται στο Παράρτημα Python.

Εύκολα παρατηρείται, ότι ορίσματα είναι οι δύο βασικές συνιστώσες εποχικότητας όπως προέκυψαν από τον Ταχύ Μετασχηματισμό Fourier 346 μέρες και 12 ώρες αντίστοιχα.

Κατά την πρώτη απόπειρα πρόβλεψης, το μοντέλο έκανε fit ωριαία δεδομένα της διαιτησίας 2018-2019 με στόχο να προβλεφθούν οι πρώτες 168 ώρες του 2020. Στο παρακάτω διάγραμμα φαίνονται οι τιμές τις πρόβλεψης σε αντιπαραβολή με τις αντίστοιχες πραγματικές τιμές του 2020:

Σχ.5.1



Ήδη εποπτικά, φαίνεται ότι το μοντέλο δεν προσεγγίζει τα δεδομένα επαρκώς. Οι μετρικές που υπολογίστηκαν φαίνονται στον παρακάτω πίνακα:

Πίνακας 5.1

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685

Λεπτομερέστερη αξιολόγηση των μοντέλων ακολουθεί παρακάτω.

5.2.2 Μοντέλο Arima (βιβλιοθήκη pyramid-arima)

Η βιβλιοθήκη Pyramid είναι μια στατιστική βιβλιοθήκη με έναν μοναδικό στόχο: να εφαρμόσει τη λειτουργικότητα της `auto.arima` της R στην Python. Η Pyramid λειτουργεί δένοντας τις `statsmodels.tsa.ARIMA` και `statsmodels.tsa.statespace.SARIMAX` σε μία κατηγορία εκτιμήσεων και δημιουργώντας μια πιο φιλική προς το χρήστη διεπαφή εκτιμητή για προγραμματιστές εξοικειωμένους με το `scikit-learn` βιβλιοθήκη.

Στην παρούσα υλοποίηση θα χρησιμοποιηθεί SARIMA μοντέλο μέσω της μεθόδου `auto_arima` του `pyrdarima` πακέτου, λαμβάνοντας όμως υπόψιν μονάχα τη δεύτερη σημαντικότερη συνιστώσα εποχικότητας (12 ώρες). Αυτό συμβαίνει επειδή, όπως είναι γνωστό από τη θεωρία, έχει δύο σημαντικά μειονεκτήματα:

- μπορεί κανείς να μοντελοποιήσει μόνο ένα εποχιακό μοτίβο
- το εύρος της πρόβλεψης (`forecasted steps`) δεν πρέπει να είναι πολύ μεγάλο.

Η εφαρμογή της μεθόδου `auto_arima`, φαίνεται στο παράρτημα Python.

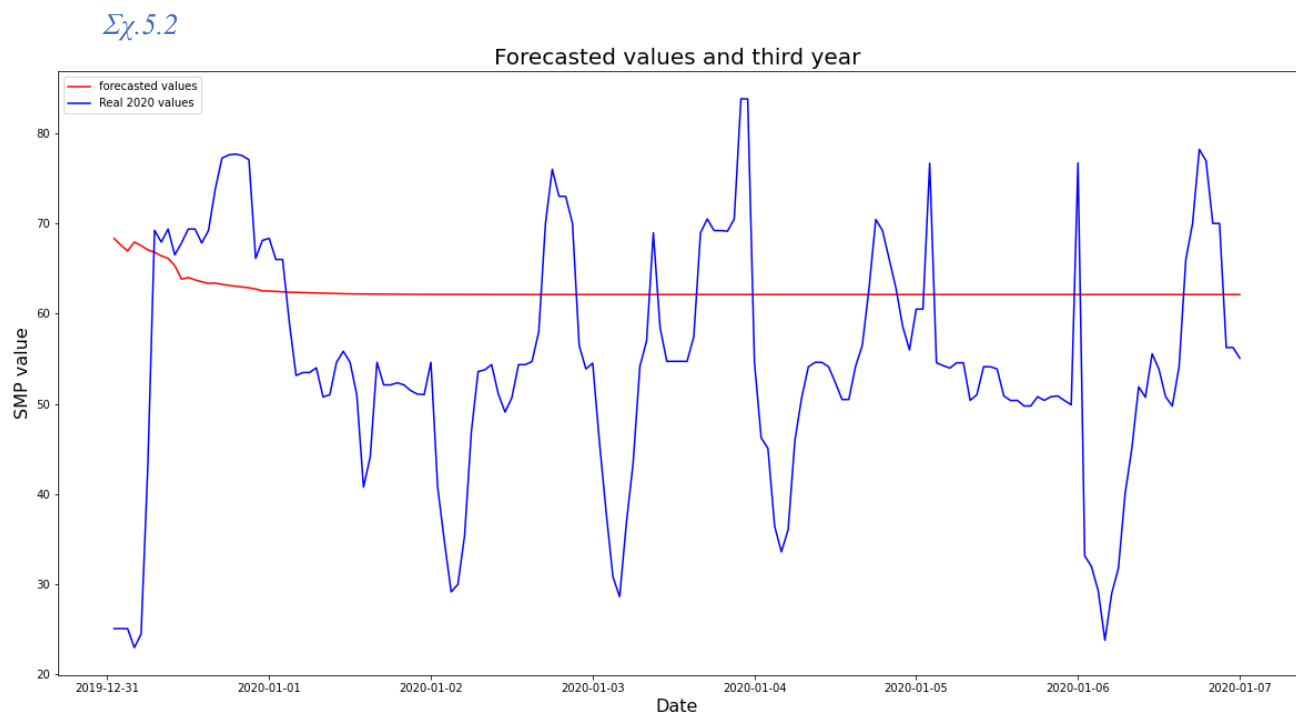
Όπως εύλογα, διακρίνει κάποιος παρατηρητής, αρχικοποιήθηκαν μια σειρά από ιδιότητες του μοντέλου με σκοπό να υπολογιστεί το βέλτιστο μοντέλο με βάση το κριτήριο AIC.

Συγκεκριμένα:

- η τριάδα παραμέτρων (p,d,q) που εκφράζουν την autoregressive συνιστώσα, τη συνιστώσα των απαιτούμενων διαφορών που είναι 0 αφού η χρονοσειρά είναι στάσιμη και η συνιστώσα των κινούμενων μέσων πρόβλεψης σφαλμάτων
- Την αντίστοιχη τριάδα (P,D,Q) για την εποχική συνιστώσα
- Την περίοδο m=12, που προκύπτει με βάση το άρθρο <https://robjhyndman.com/hyndsight/seasonal-periods/> για ωριαία δεδομένα.

Οι υπόλοιπες παράμετροι έχουν τις default τιμές τους δεδομένης της τιμής της παραμέτρου `stepwise=True`, που σημαίνει ότι χρησιμοποιείται ο αλγόριθμος `stepwise` που περιγράφεται από τους Hyndman και Khandakar (2008) για τον προσδιορισμό των βέλτιστων παραμέτρων του μοντέλου. Ο σταδιακός αλγόριθμος μπορεί να είναι σημαντικά ταχύτερος από το να ταιριάζει σε όλους (ή ένα τυχαίο υποσύνολο) συνδυασμών υπερ-παραμέτρων και είναι λιγότερο πιθανό να κάνει `overfitting` στο μοντέλο.

Όπως φαίνεται παραπάνω, το βέλτιστο μοντέλο για τις δεδομένες παραμέτρους είναι το `ARIMA(1,0,0)(1,0,0)[12]`. Συνεπώς, δοκιμάζοντας μια βραχυπρόθεσμη πρόβλεψη με τις ίδιες συνθήκες με το μοντέλο `TBATS`, έχουμε το ακόλουθο διάγραμμα:



Είναι φανερό ότι τιμές πρόβλεψης είναι προβληματικές κάτι που αποδεικνύεται εύκολα χρησιμοποιώντας το συντελεστή διακύμανσης (CV), δηλαδή το λόγο τυπικής απόκλισης προς το μέσο, που είναι $CV = 0.019092214655812508 \ll 1$ που σημαίνει ότι τα δεδομένα έχουν πολύ χαμηλή μεταβλητότητα.

Τυπικά, παρατίθενται και οι μετρικές του μοντέλου παρακάτω:

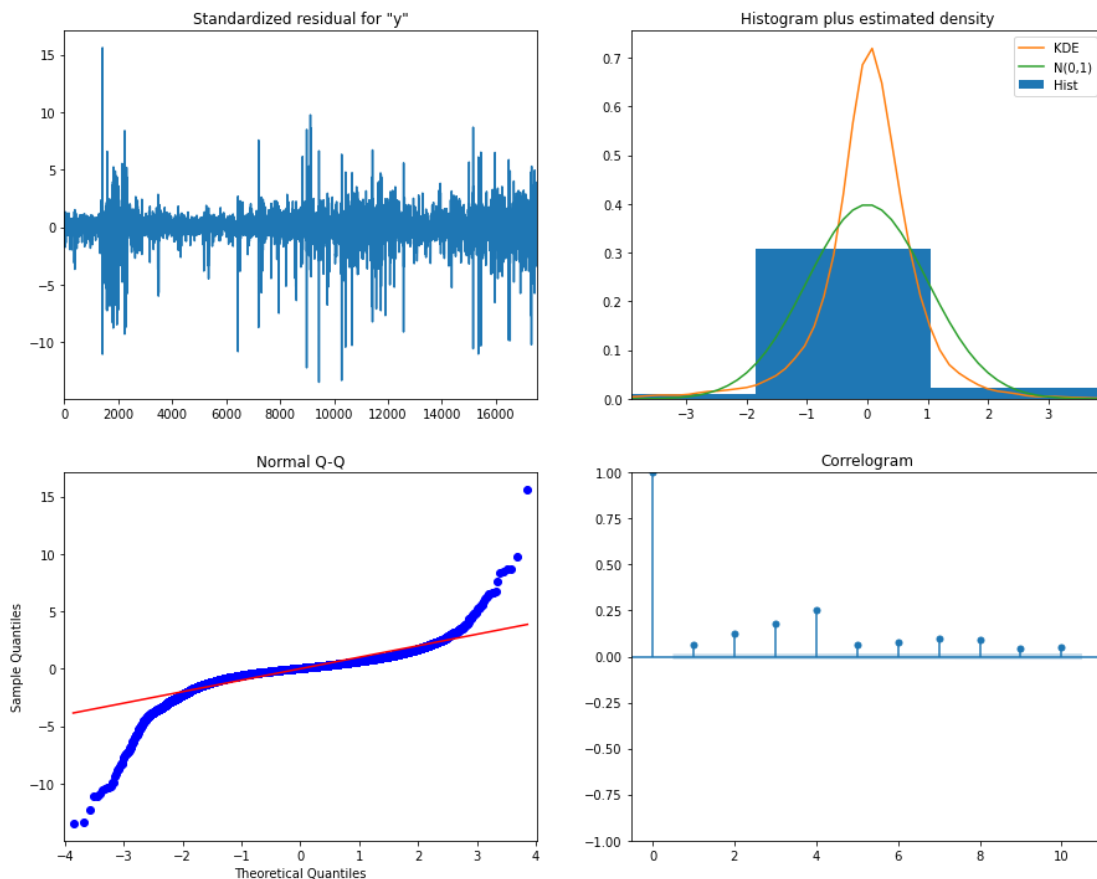
Πίνακας 5.2

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685
PMDARIMA_168	218,1032	14,7683	28,0034	22,43595	12,1290	14,4703

5.2.3 Μοντέλο κινητών μέσων (βιβλιοθήκη statsmodels.tsa.statespace.SARIMAX)

Ύστερα από την υλοποίηση της προηγούμενης ενότητας και κατόπιν πειραματισμού, θεωρήθηκε χρήσιμο να δοκιμαστεί και ένα μοντέλο MA(0,0,q). Το μοντέλο κινητού μέσου όρου είναι ένα μοντέλο χρονοσειρών το οποίο αντιπροσωπεύει πολύ βραχυπρόθεσμη αυτοσυσχέτιση. Συγκεκριμένα, αναλόγως την τιμή της συνιστώσας q, προκύπτει ότι κάθε νέα παρατήρηση είναι ο μέσος των q προηγούμενων παρατηρήσεων. Η τιμή q συνήθως προκύπτει από το διάγραμμα ACF. Στη συγκεκριμένη περίπτωση, με γνώμονα πάντα το AIC κριτήριο δοκιμάστηκε το μοντέλο ARIMA(0,0,9)(1,1,1,12). Παρακάτω φαίνεται το diagnostics διάγραμμά 2*2 του μοντέλου:

Σχ.5.3



Κύριο μέλημά της ανάλυσης είναι να διασφαλιστεί ότι τα υπόλοιπα του μοντέλου δεν συσχετίζονται και κανονικά κατανέμονται με μηδενικό μέσο όρο. Εάν η εποχική συνιστώσα ARIMA δεν ικανοποιεί αυτές τις ιδιότητες, είναι μια καλή ένδειξη ότι μπορεί να βελτιωθεί περαιτέρω. Σε αυτήν την περίπτωση, τα διαγνωστικά μοντέλων μας υποδηλώνουν ότι τα υπόλοιπα του μοντέλου δεν κατανέμονται κανονικά με βάση τα εξής:

- Στην επάνω δεξιά γραφική παράσταση, θα έπρεπε ότι η πορτοκαλί γραμμή KDE να ακολουθεί τη γραμμή N (0,1) (όπου N (0,1)) είναι η κανονική κατανομή με μέση τιμή 0 και τυπική

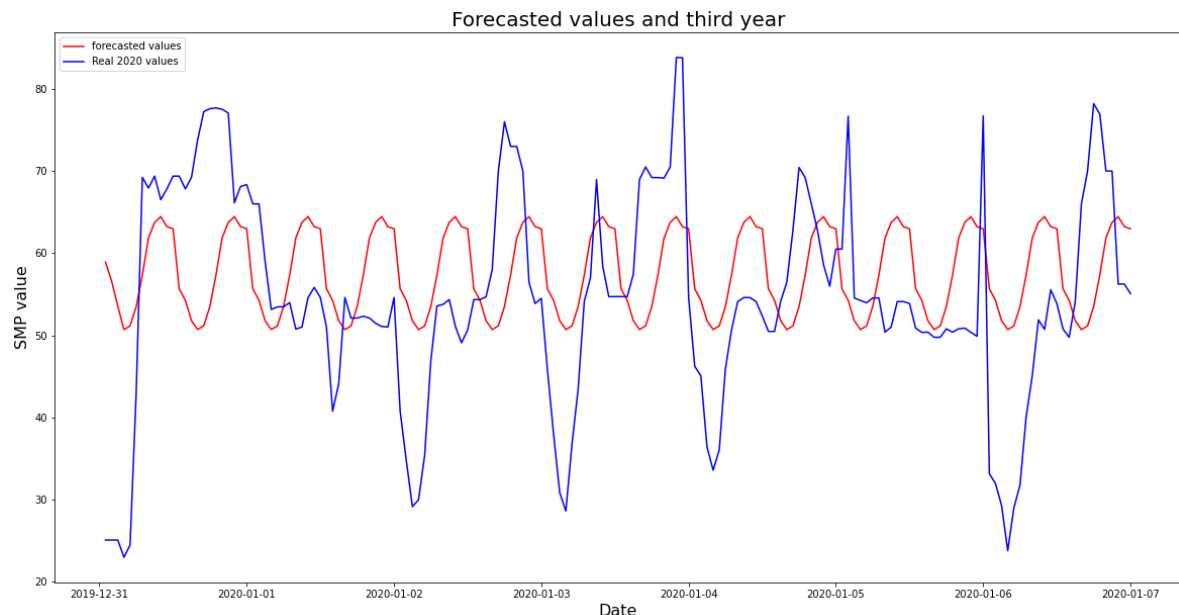
απόκλιση 1). Αυτή είναι μια καλή ένδειξη ότι τα υπόλοιπα δεν κατανέμονται κανονικά.

- Το qq-plot κάτω αριστερά δείχνει ότι η ταξινομημένη κατανομή των υπολειμμάτων (μπλε κουκκίδες) ακολουθεί σε μεγάλο βαθμό τη γραμμική τάση των δειγμάτων που λαμβάνονται από μια τυπική κανονική κατανομή με $N(0, 1)$. Και πάλι όμως, η ένδειξη αυτή δεν πιστοποιεί ότι τα υπολείμματα κατανέμονται κανονικά.
- Τα υπόλοιπα με την πάροδο του χρόνου (πάνω αριστερά διάγραμμα) εμφανίζουν υποψία εποχικότητας και δε φαίνεται να είναι λευκός θόρυβος. Αυτό επιβεβαιώνεται από το διάγραμμα αυτοσυσχέτισης κάτω δεξιά, η οποία δείχνει ότι τα υπόλοιπα των χρονοσειρών έχουν συσχέτιση με καθυστερημένες εκδόσεις του.

Αυτές οι παρατηρήσεις οδηγούν στο συμπέρασμα ότι το μοντέλο δεν παράγει μια ικανοποιητική εφαρμογή και πρέπει αν γίνει περαιτέρω διερεύνηση για την κατανόηση της χρονοσειράς και στην πρόβλεψη μελλοντικών τιμών.

Ακολουθεί το διάγραμμα προβλέψεων (168 ώρες μετά) μαζί με τις αντίστοιχες τιμές από τα δεδομένα του 2020:

Σχ.5.4



Ενδιαφέρον επίσης παρουσιάζουν οι τιμές των σφαλμάτων καθώς φαίνεται ότι είναι χαμηλότερες από αυτές των μέχρι τώρα μοντέλων:

Πίνακας 5.3

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685
PMDARIMA_168	218,1032	14,7683	28,0034	22,43595	12,1290	14,4703
SARIMAX_ARMA_168	145,4428	12,0600	21,8655	19,24457	10,1571	12,1178

5.2.4 Μοντέλο SARIMAX με συνιστώσες Fourier

Παραπάνω έγινε προσπάθεια να γίνει πρόβλεψη λαμβάνοντας υπόψιν 2 εποχικότητες μέσω της βιβλιοθήκης TBATS. Δυστυχώς, οι δυνατότητες BATS και TBATS δεν παρέχονται δωρεάν. Η μέθοδος είναι πολύ generic και καθώς εκτελείται στο περιθώριο, δημιουργεί και αξιολογεί πολλά υποψήφια μοντέλα. Αυτό έχει ως αποτέλεσμα τη βραδύτητα του υπολογισμού. Μία άλλη και εξίσου αποτελεσματική προσέγγιση αποτελεί ο συνδυασμός μοντέλων SARIMAX με σειρές Fourier που οι τελευταίες χειρίζονται τις πολλαπλές εποχικότητες. Θα χρειαστεί να γίνει κάποια προεργασία για να υλοποιηθεί η παραπάνω ιδέα.

Όπως είναι γνωστό από τη Θεωρία, ο μετασχηματισμός Fourier είναι το σημαντικότερο εργαλείο μετάβασης ενός σήματος από το πεδίο του χρόνου στο πεδίο της συχνότητας. Στην περίπτωση μιας περιοδικής συνάρτησης ο μετασχηματισμός Fourier μπορεί να απλοποιηθεί με τον υπολογισμό ενός διακριτού συνόλου σύνθετου πλάτους.

Στην περίπτωση των δεδομένων που εξετάζονται, υλοποιήθηκε μια διαδικασία που προσθέτει κυκλικές σειρές Fourier για περιόδους ώρας, έτους και εβδομάδας. Είναι δυνατόν να προστεθούν πολλαπλές σειρές Fourier με διαφορετικούς όρους k στο $(2 * k * \pi)$ όπου $k = 1, 2, 3 \dots$ κ.λπ. Για να ενσωματωθεί η πολλαπλή εποχικότητα, αυτοί οι πρόσθετοι όροι Fourier προστίθενται στο μοντέλο ARIMA, όπως φαίνεται από την εξίσωση παρακάτω:

$$y_t = \alpha + \sum_{l=1}^M \sum_{K=1}^{K_l} \left[\alpha \sin\left(\frac{2\pi kt}{p_i}\right) + \beta \cos\left(\frac{2\pi kt}{p_i}\right) \right] + N_t$$

Για να γενικευτεί το πρόβλημα, θα μπορούσε να έχει επιλεγεί μια βέλτιστη τιμή k για κάθε σεζόν δοκιμάζοντας μερικές τιμές k και επιλέγοντας τις τιμές που δίνουν το χαμηλότερο AIC. Προέκυψε ύστερα από δοκιμές $k=2$ και συνεπώς μοντελοποιείται η κύρια εποχικότητα των 346 ημερών ως εξής:

- συνιστώσες ημιτόνου και 2 συνημιτόνου για την ετήσια εποχικότητα
- συνιστώσες ημιτόνου και 2 συνημιτόνου για την εβδομαδιαία εποχικότητα
- συνιστώσες ημιτόνου και 2 συνημιτόνου για την ωριαία εποχικότητα

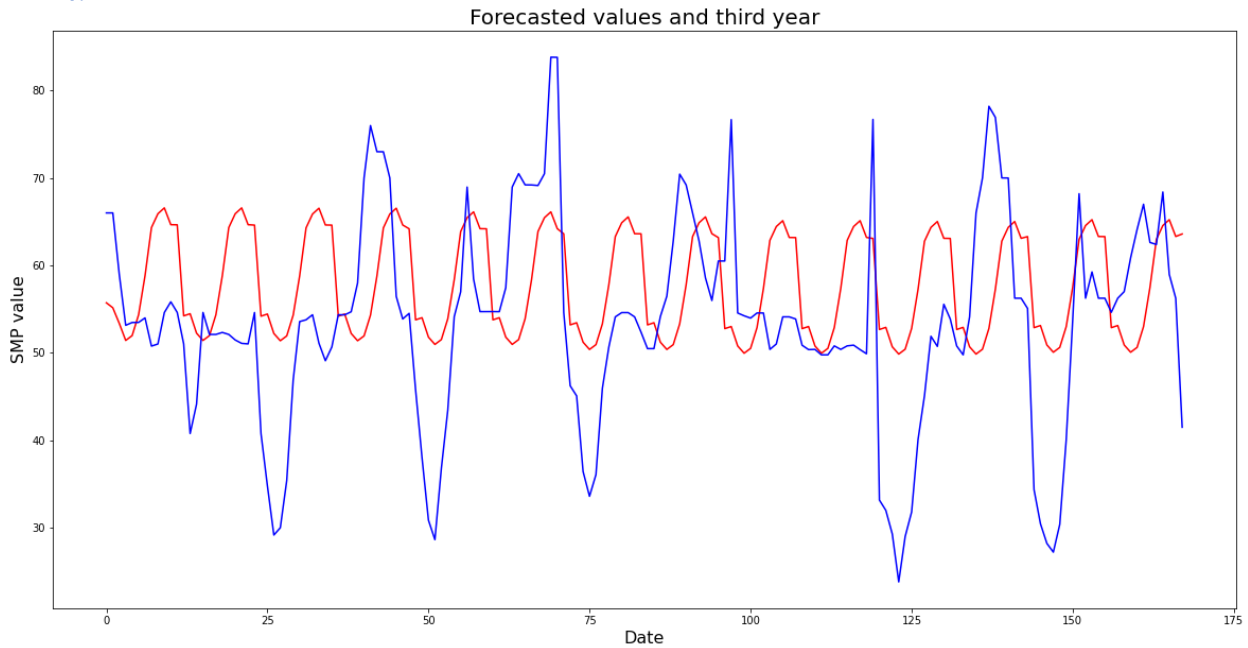
Το αποτέλεσμα της παραπάνω διεργασίας έχει προσθέσει κολώνες στο αρχικό σύνολο δεδομένων όπως φαίνεται παρακάτω:

Σχ.5.5

```
Index (['SMPS', 'Daily_Average_SMP', 'year_sin1', 'year_cos1', 'year_sin2',  
       'Year_cos2', 'week_sin1', 'week_cos1', 'week_sin2', 'week_cos2',  
       'hour_sin1', 'hour_cos1', 'hour_sin2', 'hour_cos2'],  
      dtype='object')
```

Στη συνέχεια, δημιουργούνται τα σύνολα των εξωγενών μεταβλητών εκπαίδευσης και επαλήθευσης προκειμένου να εκπαιδευτεί το μοντέλο. Στη περίπτωση αυτή, θα χρησιμοποιηθεί μοντέλο SARIMAX (0,0, 3)(0,1,1,12). Το διάγραμμα πρόβλεψης για τις πρώτες 168 ώρες του 2020 μαζί με τις αντίστοιχες πραγματικές φαίνεται παρακάτω:

Σχ. 5.6



και οι μετρικές των σφαλμάτων φαίνονται στον παρακάτω πίνακα:

Πίνακας 5.4

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685
PMDARIMA_168	218,1032	14,7683	28,0034	22,43595	12,1290	14,4703
SARIMAX_ARMA_168	145,4428	12,0600	21,8655	19,24457	10,1571	12,1178
SARIMAX_WITH_FOURIER_SHORT	146,5302	12,1050	21,8957	19,30293	10,2055	12,1755

Από την παραπάνω διεργασία, εξήχθησαν τα ακόλουθα συμπεράσματα:

- Το RMSE είναι υψηλό και ίσως χρήζει περαιτέρω διερεύνησης
- Τα μοντέλα SARIMA δεν καταγράφουν καλά πολλές εποχικότητες και είναι επίσης πολύ χρονοβόρα. Επομένως, δεν θα είναι η πρώτη επιλογή αν απαιτείται γρήγορη και ακριβής πρόβλεψη.

5.2.5 Μοντέλο FBProphet

Οι παραδοσιακές προσεγγίσεις όπως τα μοντέλα SARIMA απαιτούν συχνά χειροκίνητα βήματα προεπεξεργασίας δεδομένων (π.χ. διαφοροποίηση για να σταματήσουν τα δεδομένα) και είναι επίσης δύσκολο να εξηγηθεί γιατί αυτά τα μοντέλα παράγουν αποτελέσματα πρόβλεψης σε άτομα χωρίς εμπειρία στο αντικείμενο. Επιπλέον, σε αυτά τα μοντέλα δεν επιτρέπεται η προσθήκη πρόσθετων γνώσεων τομέα για τη βελτίωση της ακρίβειας. Για την επίλυση αυτών των προβλημάτων, οι ερευνητές του Facebook κυκλοφόρησαν πρόσφατα το FBProphet, ένα εργαλείο πρόβλεψης χρονοσειρών που υποστηρίζει τόσο τον Python όσο και τον R.

Το FBProphet παρέχει ένα μοντέλο παλινδρόμησης αποσύνθεσης που είναι επεκτάσιμο και διαμορφώσιμο με ερμηνεύσιμες παραμέτρους. Πλαισιώνει το πρόβλημα πρόβλεψης ως μια άσκηση

προσαρμογής καμπύλης παρά να εξετάζει ρητά την κάθε εξαρτώμενη από το χρόνο παρατήρηση μέσα σε μια χρονοσειρά. Παρόμοια με το SARIMAX, μπορούμε να προσθέσουμε επιπλέον όρους παλινδρόμησης στο μοντέλο.

Στον πυρήνα του, το Prophet είναι ένα προσθετικό μοντέλο με τις ακόλουθες συνιστώσες:

$$\psi(t) = g(t) + s(t) + h(t)$$

Όπου:

- $g(t)$ μοντελοποιεί την τάση, η οποία περιγράφει τη μακροπρόθεσμη αύξηση ή μείωση των δεδομένων. Το Prophet ενσωματώνει δύο μοντέλα τάσης, ένα κορεσμένο μοντέλο ανάπτυξης και ένα γραμμικό μοντέλο, ανάλογα με τον τύπο του προβλήματος πρόβλεψης.
- $s(t)$ μοντελοποιεί την εποχικότητα με σειρές Fourier, η οποία περιγράφει τον τρόπο με τον οποίο επηρεάζονται τα δεδομένα από εποχιακούς παράγοντες.
- $h(t)$ μοντελοποιεί τα αποτελέσματα των διακοπών ή μεγάλων εκδηλώσεων που επηρεάζουν τις επιχειρηματικές σειρές (π.χ. κυκλοφορία νέων προϊόντων, Black Friday, Superbowl κ.λπ.).

Το συνάρτηση Prophet() λαμβάνει ορίσματα για να διαμορφώσει τον επιθυμητό τύπο του μοντέλου, όπως τον τύπο ανάπτυξης, τον τύπο της εποχικότητας και πολλά άλλα. Από προεπιλογή, το μοντέλο θα δουλέψει σκληρά για να “καταλάβει” σχεδόν τα πάντα αυτόματα.

Η συνάρτηση fit() λαμβάνει δεδομένα σε μορφή Dataframe των χρονοσειρών. Το Dataframe πρέπει να έχει συγκεκριμένη μορφή. Πρέπει να διαθέτει μία στήλη “ds” που περιέχει τις ημερομηνίες-ώρες. Η δεύτερη στήλη πρέπει να έχει το όνομα “y” και να περιέχει τις παρατηρήσεις. Οπότε, αφού γίνουν οι σχετικές μετατροπές στο σύνολο δεδομένων των δύο πρώτων ετών, ένα δείγμα δεδομένων του Dataframe θα είναι το ακόλουθο:

Σχ. 5.7

```
[21]: train_df = init_Training_df_for_prophet(first_two_years)
      train_df

[21]:
```

	ds	y	Daily_Average_SMP
0	2018-01-01 00:00:00	57.883	52.852167
1	2018-01-01 01:00:00	58.088	52.852167
2	2018-01-01 02:00:00	57.959	52.852167
3	2018-01-01 03:00:00	54.575	52.852167
4	2018-01-01 04:00:00	54.470	52.852167
...
17539	2019-12-31 19:00:00	77.689	57.407417
17540	2019-12-31 20:00:00	77.515	57.407417
17541	2019-12-31 21:00:00	77.064	57.407417
17542	2019-12-31 22:00:00	66.134	57.407417
17543	2019-12-31 23:00:00	68.123	57.407417

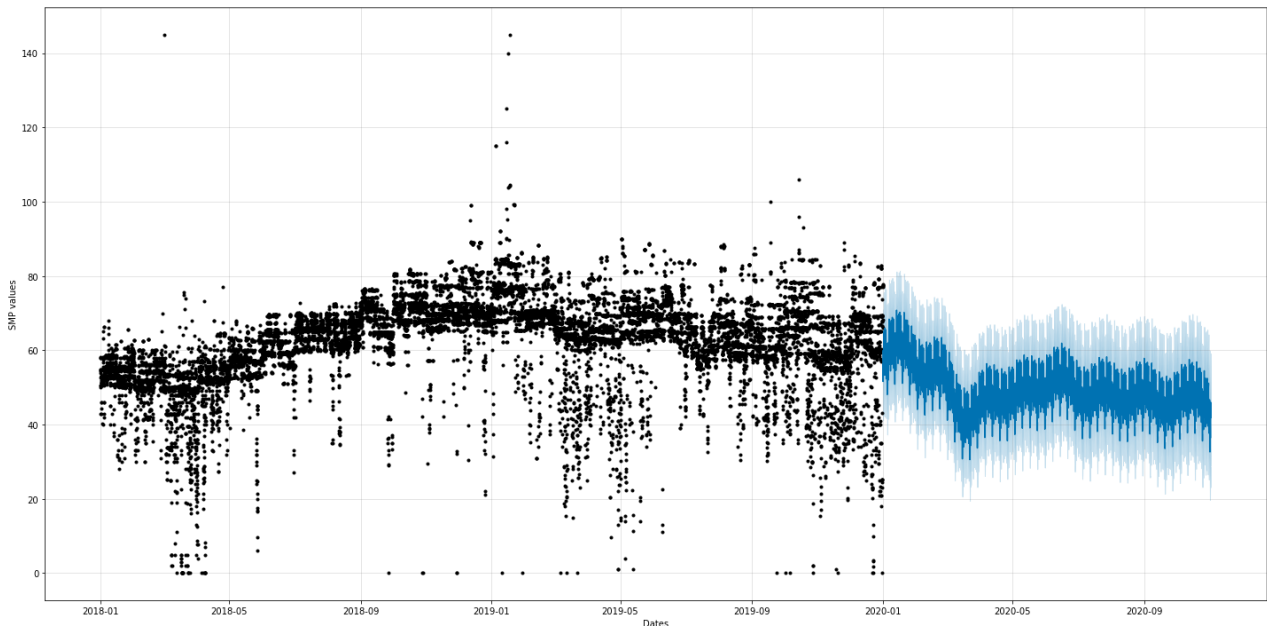
17544 rows x 3 columns

Είναι πλέον τα δεδομένα έτοιμα να προσαρμοστούν σένα μοντέλο Prophet. Για αρχή θεωρήθηκε σάφρον να μην δοθούν ορίσματα διαφορετικά από τα προεπιλεγμένα εκτός από την ενεργοποίηση της ετήσιας εποχικότητας και αυτό με τη λογική, ότι η βασική συνιστώσα εποχικότητας όπως προέκυψε από τον Ταχύ Μετασχηματισμό Fourierσε προηγούμενο εδάφιο, είναι 346 μέρες ~0.95 του έτους.

Υπό αυτές τις συνθήκες το μοντέλο θα ψάξει και θα βρει την κατά το δυνατόν βέλτιστη εποχικότητα στα δεδομένα.

Σαν δεδομένα επαλήθευσης, χρησιμοποιήθηκαν τα δεδομένα ΟΤΣ του τρίτου έτους και παράχθηκε το παρακάτω διάγραμμα δεδομένων με τα αποτελέσματα πρόβλεψης και μάλιστα με τη μορφή ακραίων τιμών (\hat{y}_{lower} - \hat{y}_{upper}).

Σχ.5.8



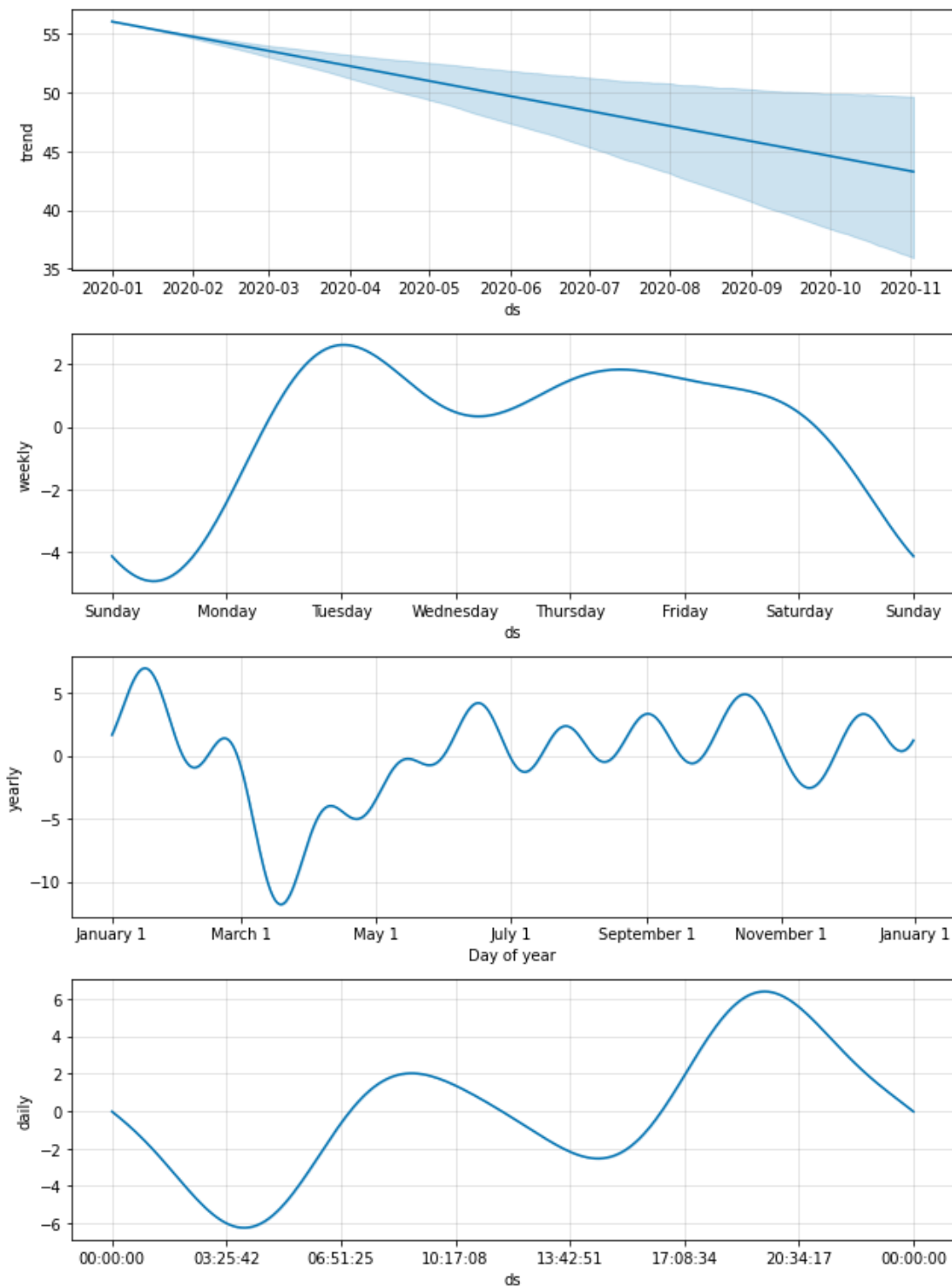
Για να γίνει κατανοητό σε μεγαλύτερο βαθμό το παραπάνω διάγραμμα, σημειώνεται ότι κατά την παραγωγή των προβλέψεων από το μοντέλο Prophet τα αποτελέσματα έχουν τις ακόλουθες κολώνες:

Σχ. 5.9

```
Index (['ds', 'trend', 'yhat_lower', 'yhat_upper', 'trend_lower', 'trend_upper',  
       'additive_terms', 'additive_terms_lower', 'additive_terms_upper',  
       'daily', 'daily_lower', 'daily_upper', 'weekly', 'weekly_lower',  
       'weekly_upper', 'yearly', 'yearly_lower', 'yearly_upper',  
       'multiplicative_terms', 'multiplicative_terms_lower',  
       'multiplicative_terms_upper', 'yhat'],  
      dtype='object')
```

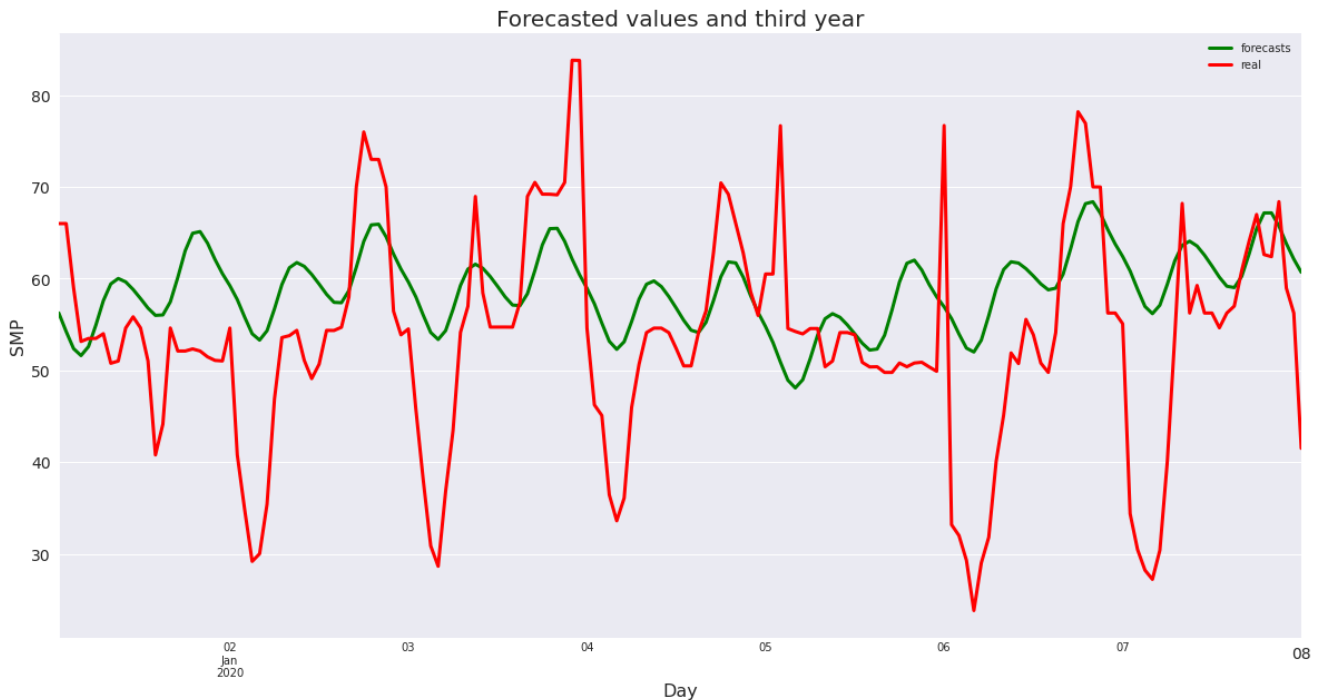
Η πληροφορία αυτή αντιπροσωπεύει και αποτελεί υπέρθεση με τελικό παραγόμενο τη μεταβλητή 'yhat'. Παρακάτω φαίνονται οι γραφικές παραστάσεις των επιμέρους συνιστωσών της τάσης και των επιμέρους εποχικότητων όπως της αποτύπωσε το μοντέλο:

Σχ.5.10



Το διάγραμμα πρόβλεψης προϋποθέτει θετικά για την αποτελεσματικότητα του μοντέλου. Προσπαθώντας να ποσοτικοποιηθεί αυτή η αίσθηση, έγινε βραχεία πρόβλεψη όπως και στα υπόλοιπα μοντέλα που δοκιμάστηκαν δηλαδή τις πρώτες 168 ώρες του 2020. Το αντίστοιχο διάγραμμα φαίνεται παρακάτω:

Σχ.5.11



Παρατηρείται μία αξιοπρεπή προσέγγιση των δεδομένων του 2020 από το μοντέλο, γεγονός που επαληθεύεται και από τις μετρικές των σφαλμάτων που φαίνονται συγκεντρωτικά και σε αντιπαραβολή με τις αντίστοιχες των προηγούμενων μοντέλων:

Πίνακας 5.5

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685
PMDARIMA_168	218,1032	14,7683	28,0034	22,43595	12,1290	14,4703
SARIMAX_ARMA_168	145,4428	12,0600	21,8655	19,24457	10,1571	12,1178
SARIMAX_WITH_FOURIER_SHORT	146,5302	12,1050	21,8957	19,30293	10,2055	12,1755
prophet_auto_short	133,4475	11,5519	23,8796	3319,794	9,0420	2102,9968

Είναι φανερό ότι το Prophet υπερέχει σε αποτελεσματικότητα των άλλων μοντέλων.

Είναι δυνατόν να προστεθούν εποχικότητες ξεχωριστά στο μοντέλο Prophet αντικαθιστώντας τη λειτουργία "auto" των εποχιακών παραπάνω με "FALSE" και προσθέτοντας τις ετήσιες, εβδομαδιαίες, ημερήσιες, μηνιαίες, τριμηνιαίες, κ.λπ. εποχικότητες, καθορίζοντας την περίοδο σε ημέρες για μια εποχικότητα μαζί με τους όρους τεσσάρων που πρέπει να χρησιμοποιηθούν. Μία ακόμα παράμετρος, η ρύθμιση της οποίας έχει ιδιαίτερη σημασία είναι το fourier_order. Αντιπροσωπεύει τον αριθμό των συνιστωσών Fourier που συνθέτουν την κάθε εποχικότητα ξεχωριστά και προκύπτει από ένα εύρος τιμών 10-25 κατόπιν δοκιμών.

Το μοντέλο λοιπόν εκπαιδεύτηκε με 15 συνιστώσες Fourier, χρησιμοποιώντας την εποχικότητα των 12 ωρών καθώς επίσης και τη σημαντικότερη των 346 ημερών ανηγμένη σε εβδομαδιαία (~49 εβδομάδες).

Σημειώνεται εδώ, ότι καθώς δεν υπάρχει ακόμα υποστήριξη του Prophet για ωριαία δεδομένα, οι περίοδοι έχουν αναχθεί σε ώρες. Παρακάτω λοιπόν φαίνονται και οι μετρικές σφαλμάτων:

Πίνακας 5.6

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685
PMDARIMA_168	218,1032	14,7683	28,0034	22,43595	12,1290	14,4703
SARIMAX_ARMA_168	145,4428	12,0600	21,8655	19,24457	10,1571	12,1178
SARIMAX_WITH_FOURIER_SHORT	146,5302	12,1050	21,8957	19,30293	10,2055	12,1755
prophet_manual_short	137,9451	11,7450	21,8168	3069,1065	8,8452	1916,3651
prophet_auto_short	133,4475	11,5519	23,8796	3319,794	9,0420	2102,9968

Οι τιμές που προέκυψαν είναι παραπλήσιες με του μοντέλου με τις default παραμέτρους, πράγμα που σημαίνει ότι το manual μοντέλο προσέγγισε σωστά τα δεδομένα.

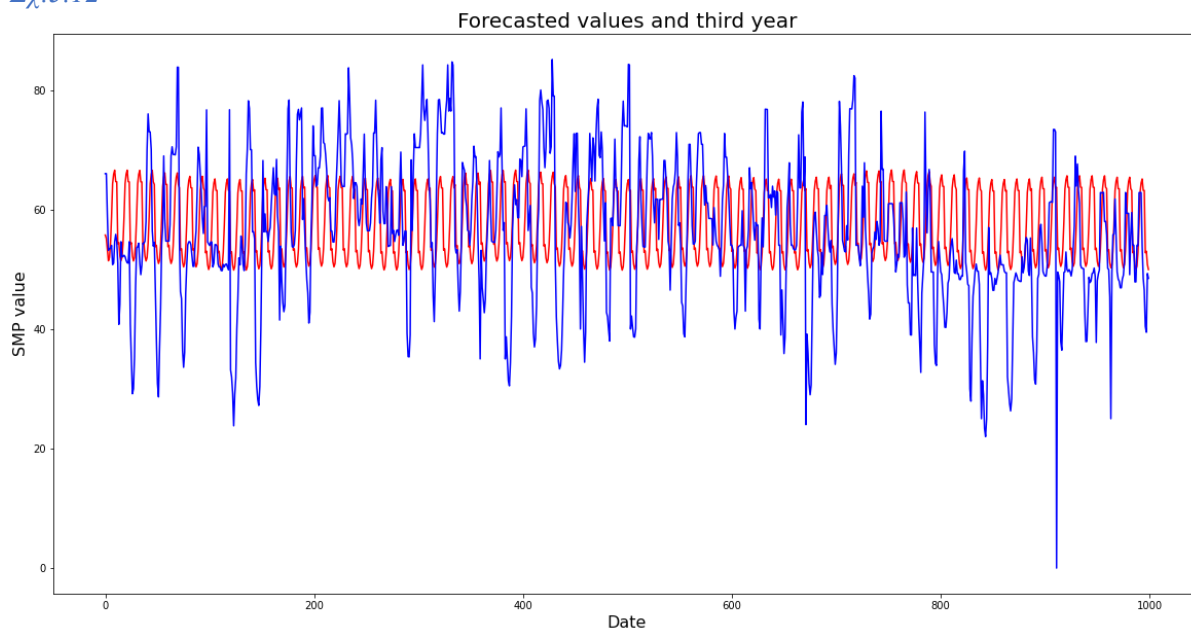
5.3 Μακροπρόθεσμη πρόβλεψη

Στα πλαίσια της παρούσας εργασίας, έγινε επίσης απόπειρα υλοποίησης μακροπρόθεσμης πρόβλεψης της Οριακής Τιμής Συστήματος Ενέργειας. Όπως έχει αναφερθεί παραπάνω, το εγχείρημα αυτό συναντά πολλά εμπόδια όπως η έλλειψη δεδομένων και η μεγάλη μεταβλητότητα των υπαρχόντων. Όμως, αξίζει να δοκιμαστούν τα μοντέλα που έχουν τη δυνατότητα να μοντελοποιήσουν πολλαπλές εποχικότητες προκειμένου να αποκρυσταλλωθεί η συμπεριφορά τους σε μακροπρόθεσμη πρόβλεψη.

Στο σημείο αυτό, θα καθοριστεί το εύρος του διαστήματος πρόβλεψης. Για τη βραχυπρόθεσμη πρόβλεψη έχει οριστεί το διάστημα των 168 ωρών. Στη συγκεκριμένη περίπτωση εξετάζονται 2 περιπτώσεις: 1000 ώρες και 7343 ώρες δηλαδή το σύνολο των δεδομένων του έτους 2020.

Το πρώτο μοντέλο που θα εξεταστεί είναι το SARIMAX με συνιστώσες Fourier. Έχοντας εκπαιδεύσει το μοντέλο με τα ωριαία δεδομένα των δύο πρώτων ετών, από πριν για τη βραχυπρόθεσμη πρόβλεψη, επιχειρήθηκε να γίνει πρόβλεψη για τις πρώτες 1000 ώρες του 2020. Παρακάτω φαίνεται το αντίστοιχο διάγραμμα προβλέψεων με τις αληθινές τιμές:

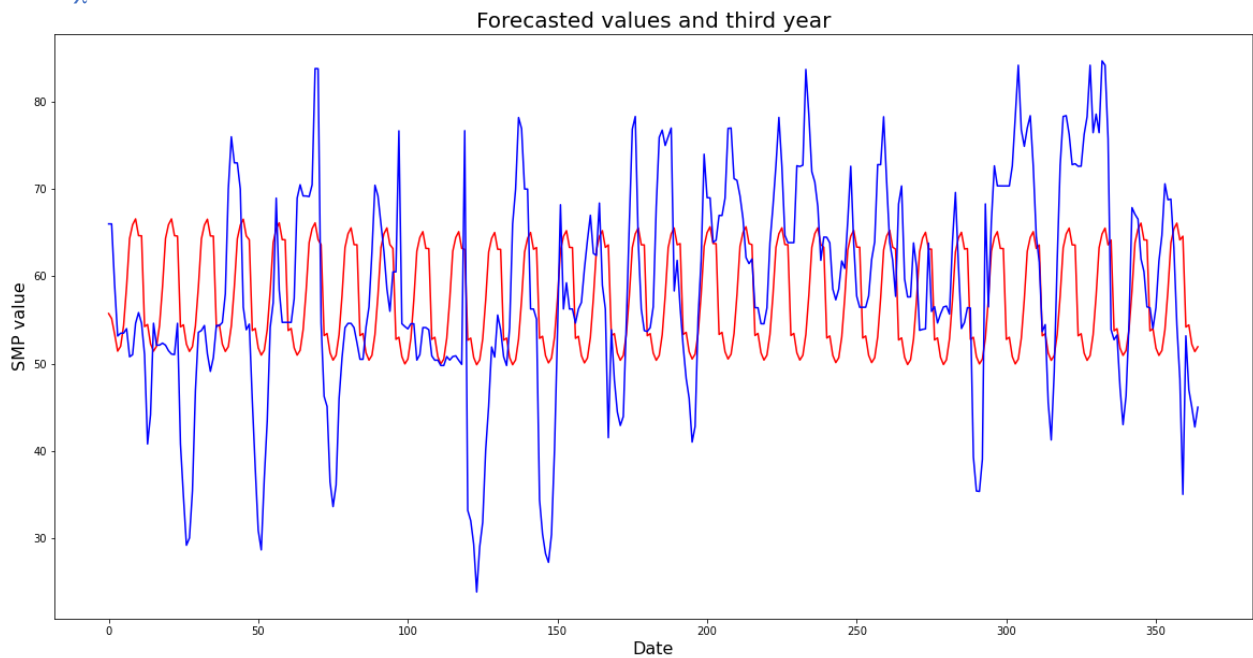
Σχ.5.12



Ήδη εποπτικά φαίνεται ότι η καμπύλη των προβλέψεων αδυνατεί να προσεγγίσει αυτή των πραγματικών δεδομένων.

Αντίστοιχο διάγραμμα φαίνεται να παράγεται και κατά την πρόβλεψη του συνόλου των δεδομένων του 2020 (7343 ωρών):

Σχ.5.13



Ο πίνακας σφαλμάτων πρόβλεψης φαίνεται παρακάτω:

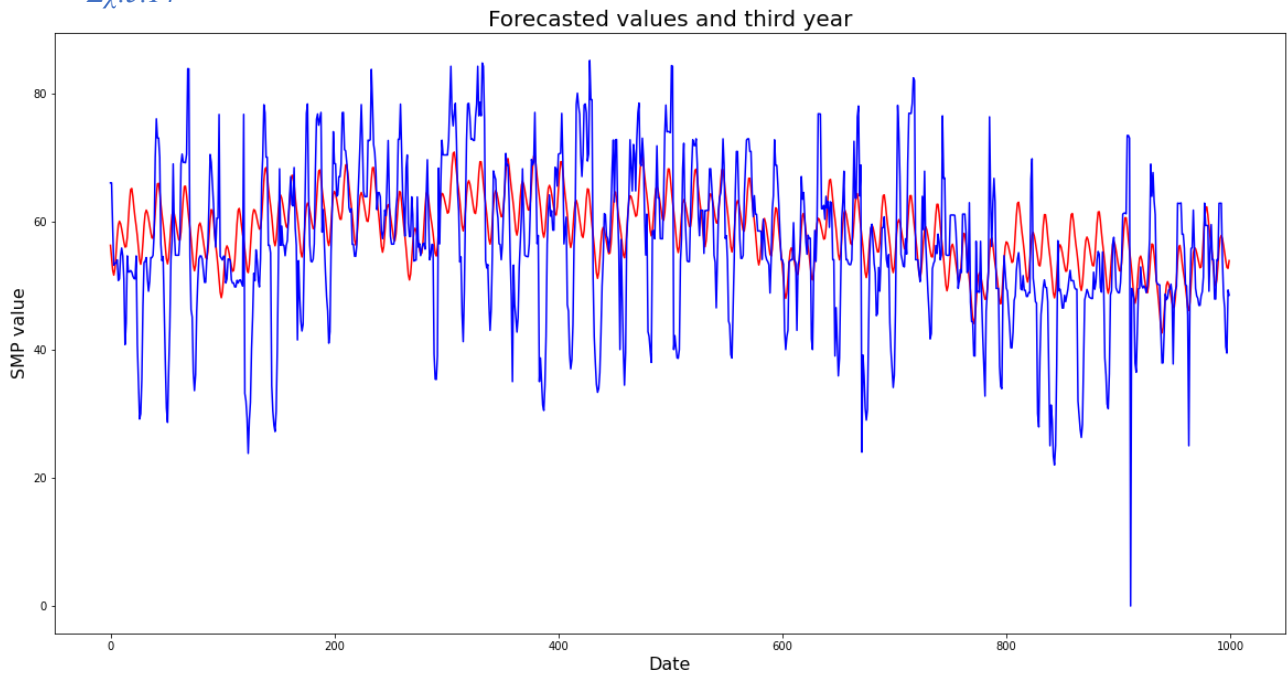
Πίνακας 5.7

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
SARIMAX_WITH_FOURIER_OPT_LONG_1000	138,4398	11,7660	inf	17,3067	9,4695	11,1263
SARIMAX_WITH_FOURIER_OPT_LONG_7343	432,7810	20,8034	inf	38,2966	17,7398	11,8226

Παρατηρείται ότι οι μετρικές σφάλματος των προβλέψεων για τις πρώτες 1000 ώρες του 2020 είναι αξιόλογες και θα αναλυθούν περαιτέρω σε επόμενο κεφάλαιο.

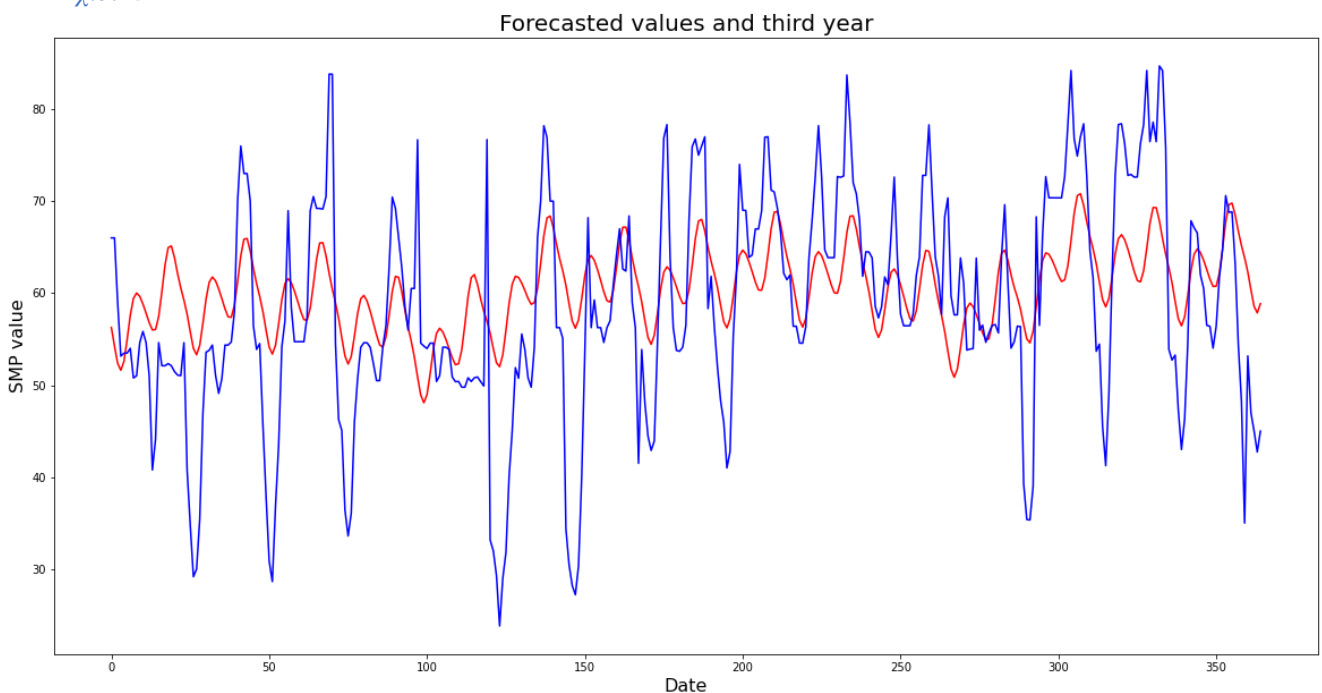
Σειρά έχει η μακροχρόνια πρόβλεψη μέσω του Prophet υπό τις ίδιες συνθήκες. Το διάγραμμά πρόβλεψης για τις πρώτες 1000 ώρες του 2020 φαίνεται παρακάτω:

Σχ.5.14



Είναι φανερό ότι τα δεδομένα προσεγγίζονται από τις προβλέψεις καλύτερα απ' ό τι σε προηγούμενα μοντέλα. Για να φανεί λίγο καλύτερα, παρουσιάζεται το ίδιο διάγραμμα αλλά για τις πρώτες 365 τιμές:

Σχ.5.15



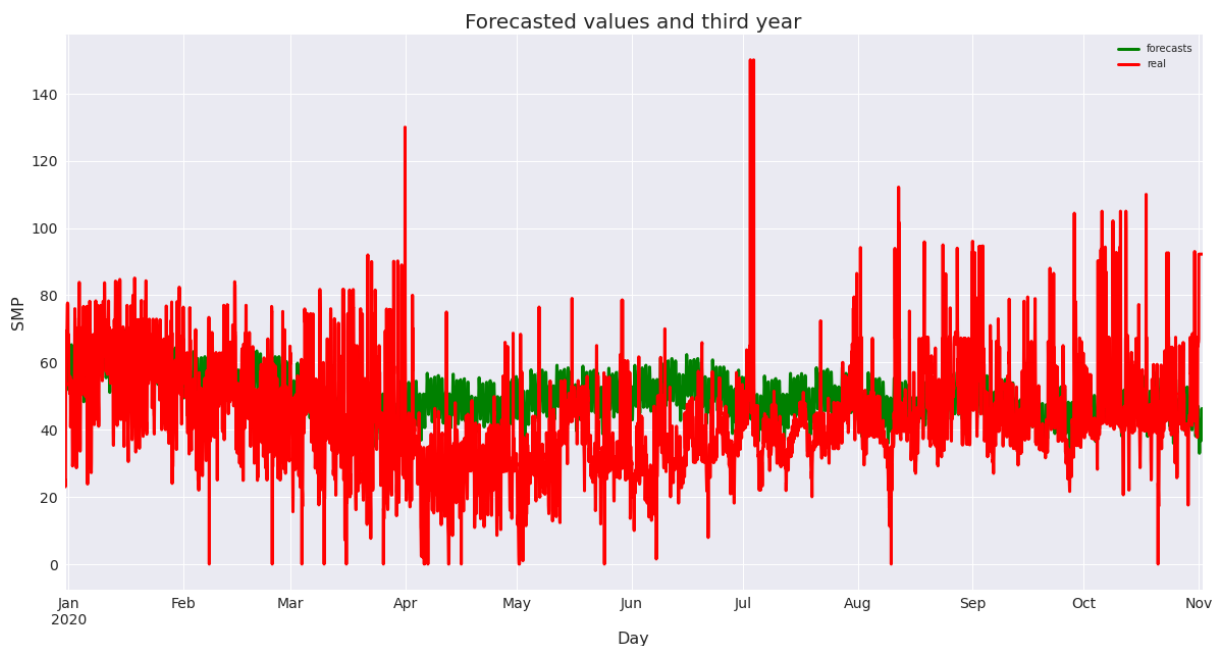
Όπως αναμένεται και οι μετρικές του προβλέψεων θα είναι βελτιωμένες. Πράγματι, όπως φαίνεται από τον πίνακα παρακάτω:

Πίνακας 5.8

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
SARIMAX_FOURIER_OPT_LONG_1000	138,4398	11,7660	inf	17,3067	9,4695	11,1263
SARIMAX_FOURIER_OPT_LONG_7343	432,7810	20,8034	inf	38,2966	17,7398	11,8226
prophet_auto_1000	98,8443	9,9420	inf	19145,0585	7,6178	12387,2271

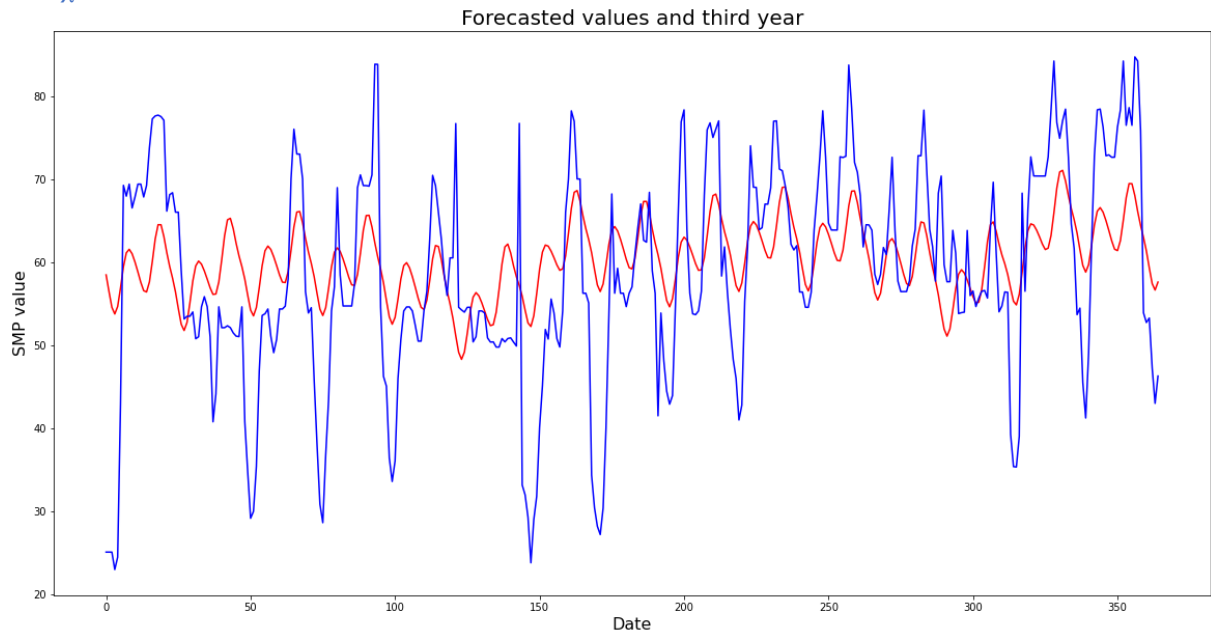
Όμοια με τα SARIMAX μοντέλα, γίνεται πρόβλεψη μέσω Prophet στο σύνολο των δεδομένων του 2020. Ακολουθούν τα διαγράμματα προβλέψεων-παρατηρήσεων και ο συγκεντρωτικός πίνακας με τις μετρικές σφάλματος για τις προβλέψεις στην περίπτωση της μακροπρόθεσμης πρόβλεψης που εξετάστηκε σε αυτή την ενότητα.

Σχ.5.16



Και πάλι παρουσιάζεται σε σμίκρυνση το παραπάνω και συγκεκριμένα για τις πρώτες 365 ώρες προκειμένου να γίνει περισσότερο ευδιάκριτο:

Σχ.5.17



Και πάλι παρατηρείται μία ικανοποιητική προσέγγιση που αποτυπώνεται και στον πίνακα μετρικών :

Πίνακας 5.9

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
SARIMAX_FOURIER_OPT_LONG_1000	138,4398	11,7660	inf	17,3067	9,4695	11,1263
SARIMAX_FOURIER_OPT_LONG_7343	432,7810	20,8034	inf	38,2966	17,7398	11,8226
prophet_auto	224,0473	14,9682	inf	234989,6518	11,6680	68550,8179
prophet_auto_1000	98,8443	9,9420	inf	19145,0585	7,6178	12387,2271

Κεφάλαιο 6

6.1 Παρουσίαση των μετρικών

Στο κεφάλαιο αυτό θα παρατεθούν τα αποτελέσματα πρόβλεψης και οι αντίστοιχες μετρικές σφάλματος για τα μοντέλα που εξετάστηκαν στο κεφάλαιο 5 και θα αξιολογηθούν από πλευράς ακρίβειας πρόβλεψης. Θα πρέπει να σημειωθεί στο σημείο αυτό, ότι αντικειμενικός σκοπός της παρούσας Διπλωματικής εργασίας είναι η εύρεση ενός ικανού μοντέλου που θα χρησιμοποιηθεί στη συνέχεια για την μελλοντική πρόβλεψη τιμών ΟΤΣ, με το αντίστοιχο business value.

Χρησιμοποιούνται περισσότερα του ενός κριτήρια αξιολόγησης των μεθόδων, με βασικότερα την ελαχιστοποίηση του RMSE και του MAE. Έπειτα από την παρουσίαση των σφαλμάτων για όλες τις μεθόδους, παρουσιάζεται και η κατάταξη αυτών βάσει των σφαλμάτων.

Υπενθυμίζεται σε αυτό το σημείο, ότι οι πειραματικές δοκιμές ομαδοποιήθηκαν και μελετήθηκαν ξεχωριστά με κριτήριο αρχικά το εύρος της πρόβλεψης σε βραχυπρόθεσμη (168 ώρες) και μακροπρόθεσμη (1000 και 7343 ώρες).

Ξεκινώντας από τον συγκεντρωτικό πίνακα με τις μετρικές σφαλμάτων για τη βραχυπρόθεσμη πρόβλεψη:

Πίνακας 6.1

	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
TBATS	227,9363	15,0976	29,1541	23,47158	12,7981	15,2685
PMDARIMA_168	218,1032	14,7683	28,0034	22,43595	12,1290	14,4703
SARIMAX_ARMA_168	145,4428	12,0600	21,8655	19,24457	10,1571	12,1178
SARIMAX_WITH_FOURIER_SHORT	146,5302	12,1050	21,8957	19,30293	10,2055	12,1755
prophet_manual_short	137,9451	11,7450	21,8168	3069,1065	8,8452	1916,3651
prophet_auto_short	133,4475	11,5519	23,8796	3319,794	9,0420	2102,9968

Ακολουθεί ο αντίστοιχος πίνακας για τη μακροπρόθεσμη πρόβλεψη:

Πίνακας 6.2

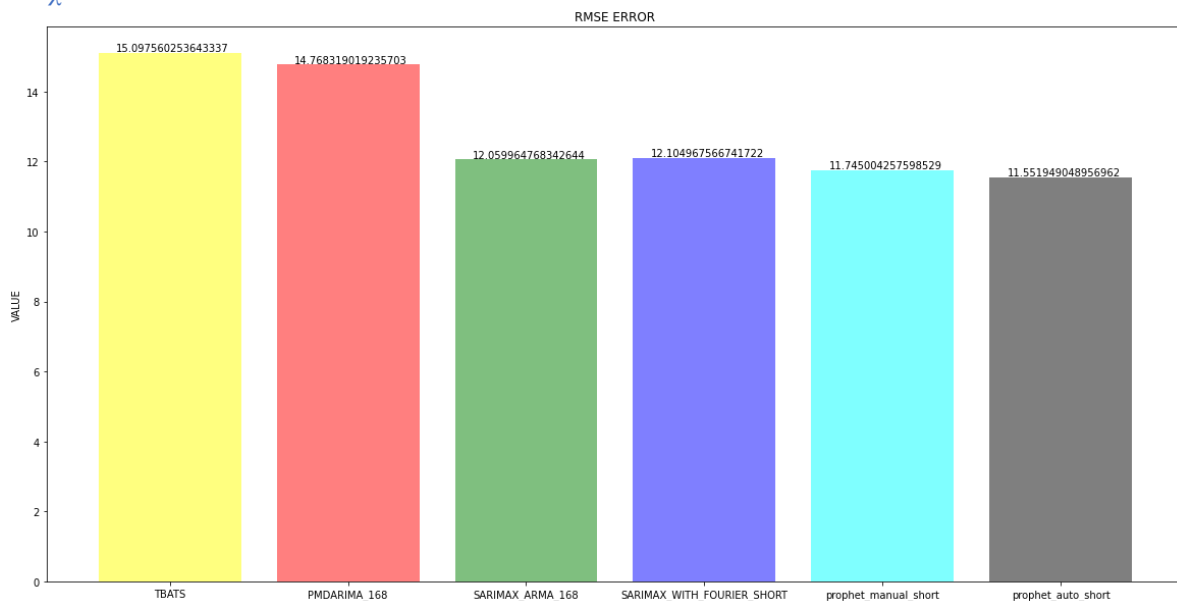
	MSE	RMSE	MAPE	sMAPE	MAE	MARNE
SARIMAX_FOURIER_OPT_LONG_1000	138,4398	11,7660	inf	17,3067	9,4695	11,1263
SARIMAX_FOURIER_OPT_LONG_7343	432,7810	20,8034	inf	38,2966	17,7398	11,8226
prophet_auto	224,0473	14,9682	inf	234989,6518	11,6680	68550,8179
prophet_auto_1000	98,8443	9,9420	inf	19145,0585	7,6178	12387,2271

6.2 Ανάλυση αποτελεσμάτων

Η πρώτη μετρική που ενδιαφέρει και εξετάζεται είναι το RMSE. Καθώς αποτελεί την τετραγωνική ρίζα του Μέσου Τετραγωνικού Σφάλματος, στην ουσία αναπαριστά τη μέση διαφορά προβλεπόμενων τιμών με τις παρατηρήσεις. Αυτό σημαίνει ότι αναμένεται να παρουσιάζει ενδιαφέρον σε μεγάλες διαφορές πρόβλεψης - παρατήρησης και συνεπώς και μία ευαισθησία στις ακραίες τιμές που μεγιστοποιούν τις διαφορές. Λαμβάνοντας υπόψιν και το γεγονός ότι αποτελεί άθροισμα τετραγώνων διαφορών, εξάγεται το συμπέρασμα ότι είναι απαίτηση στο μοντέλο που θα προαχθεί να είναι χαμηλό το RMSE. Από την άλλη μεριά, είναι γνωστό από τη θεωρία ότι η βελτιστοποίηση του οδηγεί σε προβλέψεις σωστές και ακριβείς στο μέσο (on average) που συνεπάγεται και αμερόληπτες (unbiased).

Από τον παραπάνω πίνακα, φαίνεται ότι αυτό συμβαίνει στο μοντέλο FBProphet με την αυτόματη ρύθμιση εποχικότητας.

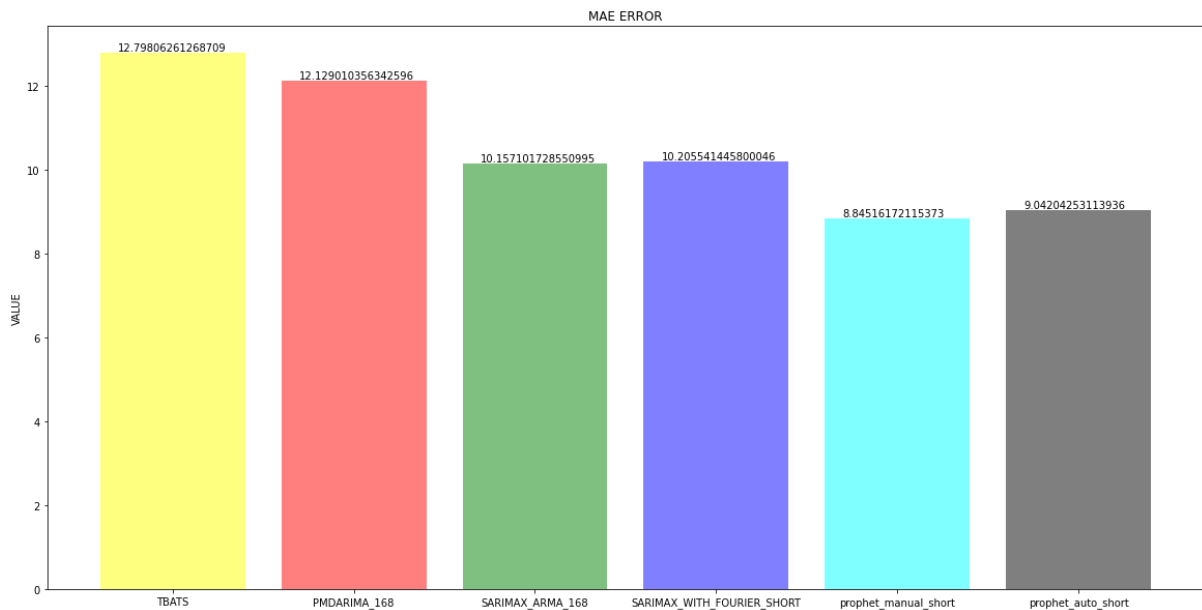
Σχ.6.1



Προχωρώντας, ο επόμενος δείκτης που εξετάζεται είναι το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error). Για τον υπολογισμό του βέλτιστου MAE, αναζητείται μια τιμή που χωρίζει το σύνολο δεδομένων σε δύο ίσα μέρη. Με άλλα λόγια, η βελτιστοποίηση του οδηγεί σε προβλέψεις ακριβείς στο διάμεσο (on median). Θα μπορούσε κανείς να ισχυριστεί, ότι αποτελεί το αντίπαλο δέος του RMSE και αυτό σημαίνει ότι μια πρόβλεψη που ελαχιστοποιεί το MAE θα έχει ως αποτέλεσμα μεροληψία (Bias). Συγκριτικά, μια πρόβλεψη ελαχιστοποίησης του RMSE δεν θα οδηγήσει σε Bias (καθώς στοχεύει στο μέσο όρο). Αυτή είναι σίγουρα η κύρια αδυναμία του MAE.

Παρόλα αυτά, στη γενική περίπτωση η πρόβλεψη με στόχευση το διάμεσο προφανώς είναι περισσότερο ανεκτική σε ακραίες τιμές. Με βάση τα παραπάνω και τον πίνακα αποτελεσμάτων, φαίνεται ότι το βέλτιστο MAE φαίνεται στο FBProphet με τη manual ρύθμιση εποχικότητας.

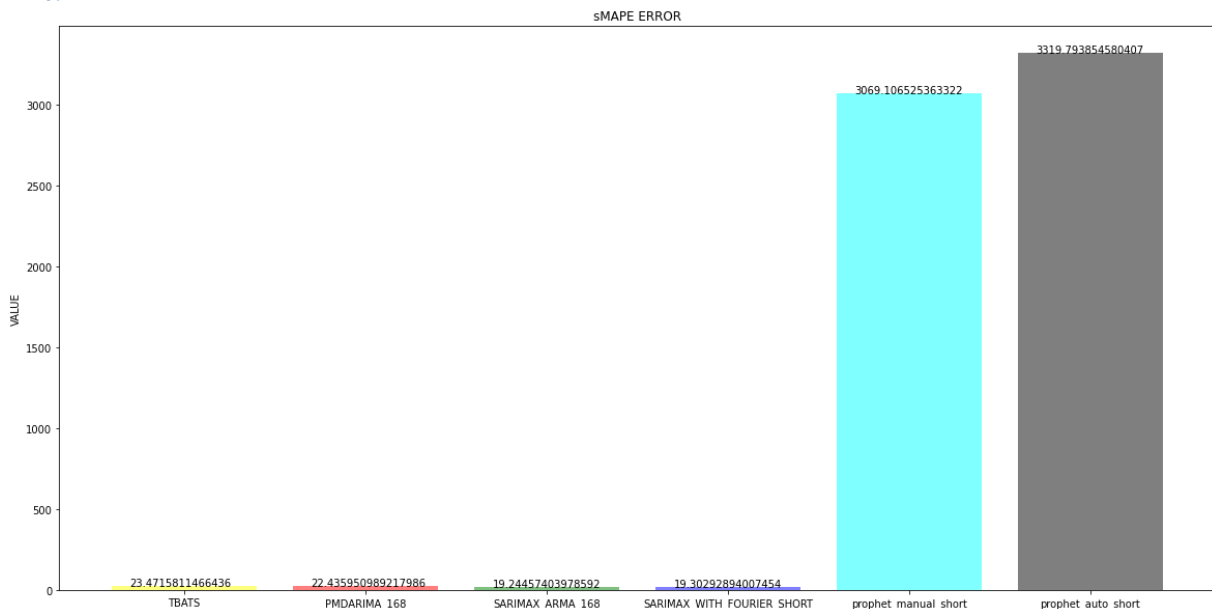
Σχ.6.2



Στο σημείο αυτό γίνεται αντιληπτό ότι πρέπει ληφθούν και άλλες παράμετροι υπόψιν προκειμένου να επιλεγεί το κατάλληλο μοντέλο.

Παρακάτω φαίνεται το αντίστοιχο ιστόγραμμα για το *sMAPE*:

Σχ.6.3



Παρά την συμμετρικοποίηση του MAPE, παρατηρείται ότι και πάλι λαμβάνει πολύ ακραίες τιμές στην περίπτωση των FBProphet μοντέλων και συνεπώς δεν μπορεί να αποτελέσει αξιόπιστο δείκτη.

Στο σημείο αυτό, μένει να αποφασιστεί με βάση ποια μετρική θα επιλεγεί το τελικό μοντέλο. Φαίνεται ότι από τη σκοπιά του RMSE το βέλτιστο μοντέλο είναι το FBProphet με αυτόματη ρύθμιση στην εποχικότητα ενώ από τη σκοπιά του MAE προκύπτει ότι το χαμηλότερο σφάλμα κατέχει το μοντέλο FBProphet με την manual ρύθμιση της εποχικότητας. Όπως αναφέρθηκε παραπάνω, ανάμεσα σε αυτές τις δύο μετρικές υπάρχει μία ακόμα μαθηματική έννοια αυτή της μεροληψίας (bias).

Συγκεκριμένα, θα υπολογιστεί το Μέσο Σφάλμα Μεροληψίας (MeanBiasError) από την παρακάτω σχέση:

$$MBE = \frac{1}{N} \sum_{i=1}^N (f_t - a_t)$$

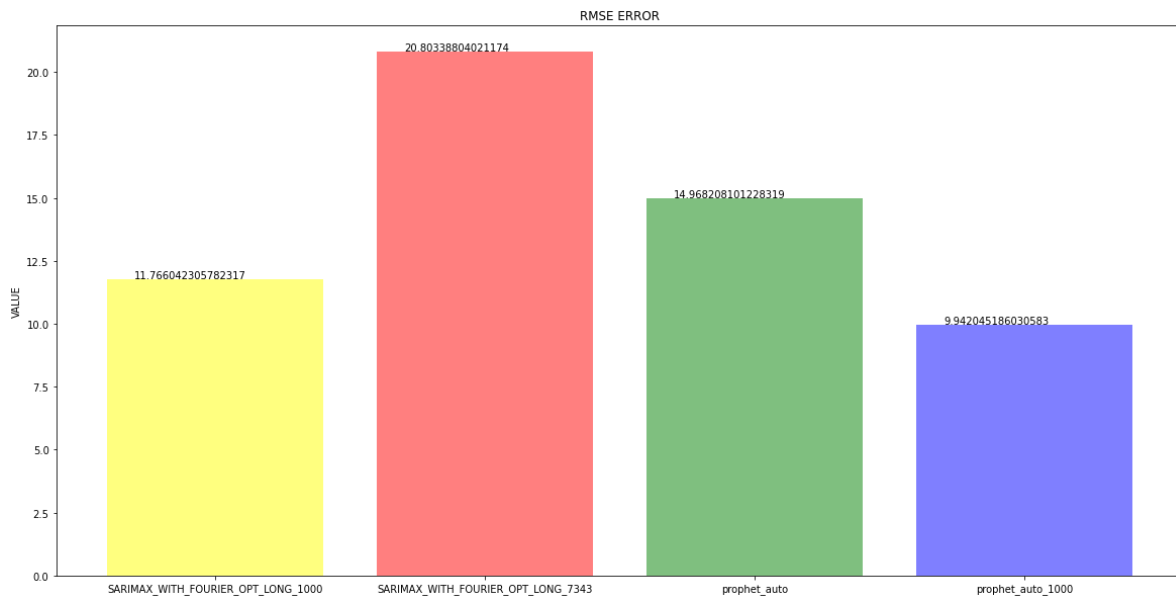
όπου f_t η πρόβλεψη τη χρονική στιγμή t και a_t η πραγματική τιμή την ίδια στιγμή.

Το μέγεθος αυτό αντιπροσωπεύει το συστηματικό σφάλμα ενός μοντέλου πρόβλεψης να κάνει υπο-πρόβλεψη (under-forecasting) ή υπέρ - πρόβλεψη (over-forecasting). Με βάση τον παραπάνω τύπο προκύπτει ότι $MBE_{manual} = 679.4936303862695$ ενώ το $MBE_{auto} = 873.9574248843056$.

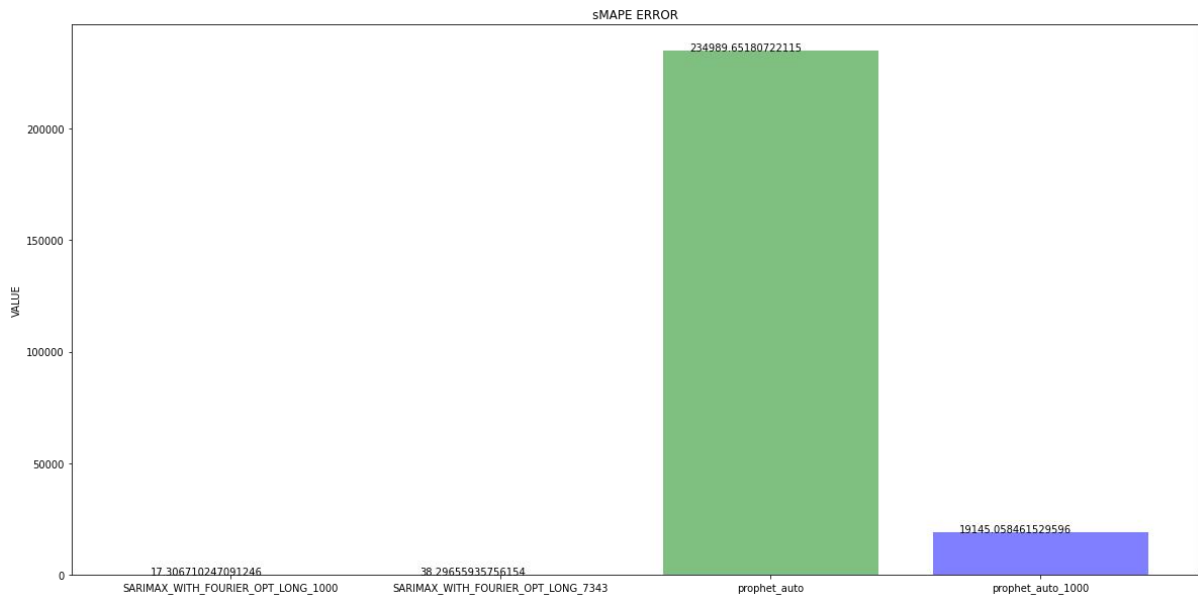
Εύκολα αντιλαμβάνεται κανείς, ότι το μοντέλο με την αυτόματη ρύθμιση εποχικότητας κάνει περισσότερο over-forecasting και δεδομένου ότι το άλλο μοντέλο έχει μικρότερο MAE και μικρότερο MBE με ελάχιστο μεγαλύτερο RMSE, θεωρήθηκε ότι το MAE είναι πιο αξιόπιστος δείκτης και θα είναι το μοντέλο που θα επιλεγεί τελικά για την πραγματική μελλοντική βραχυπρόθεσμη πρόβλεψη στο επόμενο κεφάλαιο.

Στην περίπτωση της μακροπρόθεσμης πρόβλεψης, η εύρεση του κατάλληλου μοντέλου ήταν πιο ξεκάθαρη. Ακολουθούν τα αντίστοιχα ιστογράμματα:

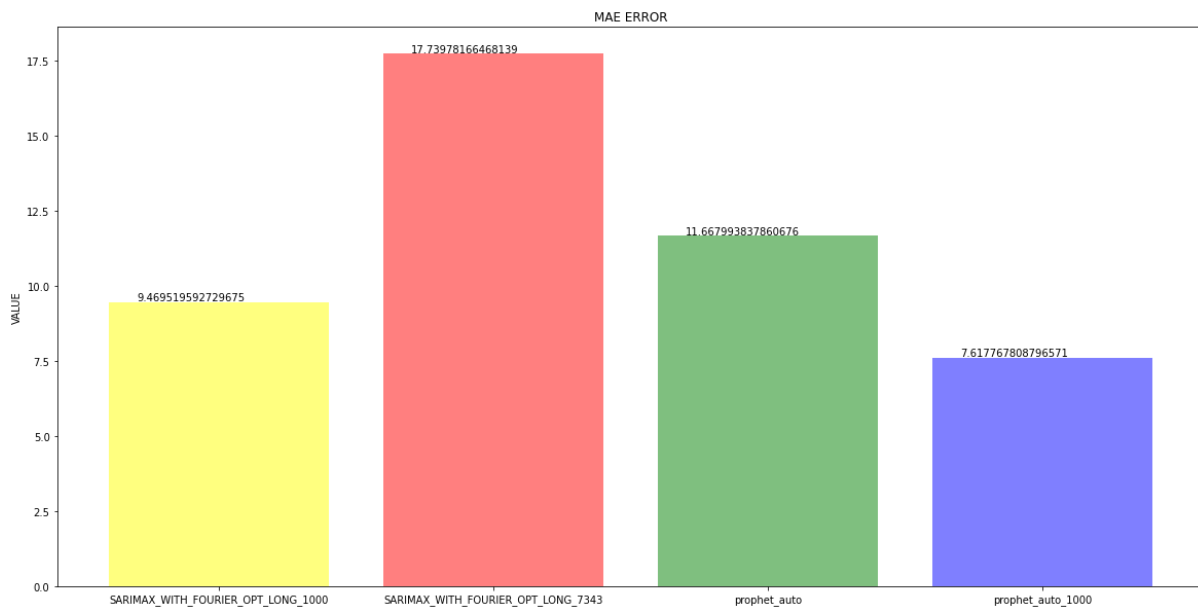
Σχ.6.4



Σχ. 6.5



Σχ. 6.6



Παρατηρείται, ότι ανάμεσα στο μοντέλα που δοκιμάστηκαν, το FBProphet με την αυτόματη ρύθμιση εποχικότητας παρουσίασε τις ελάχιστες τιμές σφαλμάτων RMSE και MAE. Συνεπώς στο επόμενο κεφάλαιο, θα γίνει πρόβλεψη με αυτό.

Κεφάλαιο 7

7.1 Μελλοντική Πρόβλεψη

Στο προηγούμενο κεφάλαιο πραγματοποιήθηκε ανάλυση των αποτελεσμάτων και συγκεκριμένα των μετρικών σφάλματος για τα διάφορα μοντέλα που δοκιμάστηκαν για βραχυπρόθεσμη και μακροπρόθεσμη πρόβλεψη. Από θεωρητικής πλευράς έγινε προσπάθεια να υπάρξει επαρκής τεκμηρίωση και συμπίεση με τη σύγχρονη γνώση γύρω από την αξιολόγηση μοντέλων πρόβλεψης χρονοσειρών.

Η επιλογή τη βιβλιοθήκης FBProphet δεν έγινε τυχαία και φάνηκε ότι το μοντέλο που προσέγγισαν τη χρονοσειρά με βάση τη συγκεκριμένη βιβλιοθήκη μοντελοποίησαν και προσέγγισαν τα δεδομένα με σχετική αποτελεσματικότητα. Υπενθυμίζεται ότι προάχθηκε για τη βραχυπρόθεσμη πρόβλεψη μοντέλο FBProphet με customρύθμιση εποχικότη των σύμφωνα πάντα με τα αποτελέσματα του Ταχέως Μετασηματισμού Fourierπου αφορούν τις εποχικότητες. Το συγκεκριμένο μοντέλο ρυθμίστηκε με εποχικότητα 49 εβδομάδων για να προσεγγίσει την εποχικότητα των 346 ημερών και την εποχικότητα των 12 ωρών και θα χρησιμοποιηθεί για πρόβλεψη μίας μέρα μελλοντικά (one -day-ahead-forecast) και μία εβδομάδα μετά.

Αντίστοιχα, για τη μακροπρόθεσμη πρόβλεψη προάχθηκε το μοντέλο FBProphet με αυτόματη ρύθμιση και πρόβλεψη σε 1000 ώρες και 7343 ώρες αντίστοιχα.

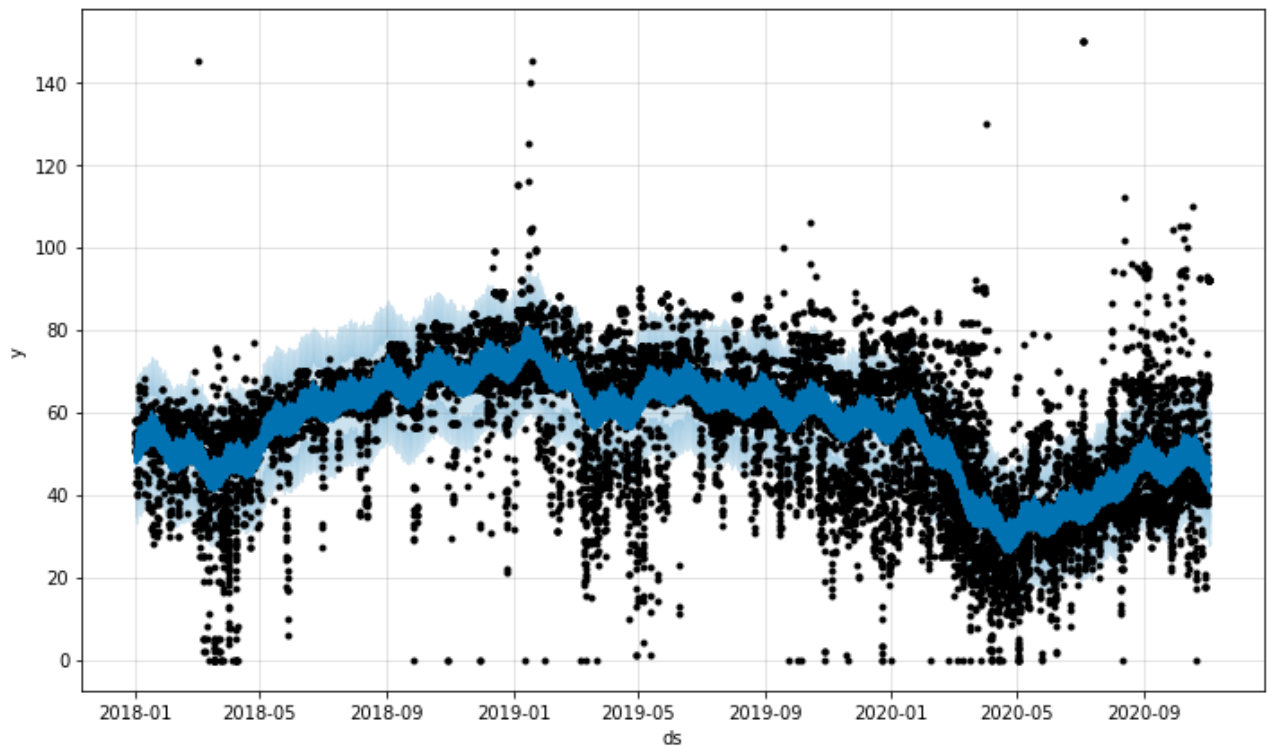
7.1.1 Αποτελέσματα Βραχυπρόθεσμης Πρόβλεψης

Για την παραγωγή μελλοντικών προβλέψεων χρησιμοποιήθηκαν όλα τα διαθέσιμα εξ αρχής δεδομένα (01/01/2018-01/11/2020) σαν δεδομένα εκπαίδευσης. Η πρώτη απόπειρα μελλοντικής πρόβλεψης θα γίνει 24 ώρες μετά και είναι αποτελεί σπουδαίο δείκτη και εργαλείο στην αγορά ενέργειας. Η σημασία της έγκειται στο γεγονός ότι χρησιμοποιείται στην Αγορά επόμενης ημέρας. Η αγορά αυτή επιτρέπει στους συμμετέχοντες να υποβάλλουν εντολές συναλλαγών η/ε με υποχρέωση φυσικής παράδοσης την επόμενη ημέρα. Στην αγορά επόμενης ημέρας δηλώνονται επίσης και οι ποσότητες ενέργειας που έχουν δεσμευτεί μέσω διενέργειας συναλλαγών επί προθεσμιακών προϊόντων, που έχουν πραγματοποιηθεί είτε μέσω της χονδρικής αγοράς προθεσμιακών προϊόντων, είτε εκτός αυτής. Παράλληλα, θα πραγματοποιείται έμμεση κατανομή (implicit allocation) της μεταφορικής ικανότητας στις διασυνδέσεις, μέσω σύζευξης των αγορών επόμενης ημέρας των Ευρωπαϊκών χωρών. Συνεπώς, ότι τιμές πρόβλεψης που θα υπολογισθούν διαθέτουν business value.

Η Αγορά επόμενης ημέρας αποτελεί μία από τις 4 αγορές που ισχύουν με βάση την υπάρχουσα νομοθεσία, μαζί με την Χονδρική αγορά προθεσμιακών προϊόντων, την Ενδοημερήσια αγορά και την Αγορά Εξισορρόπησης.

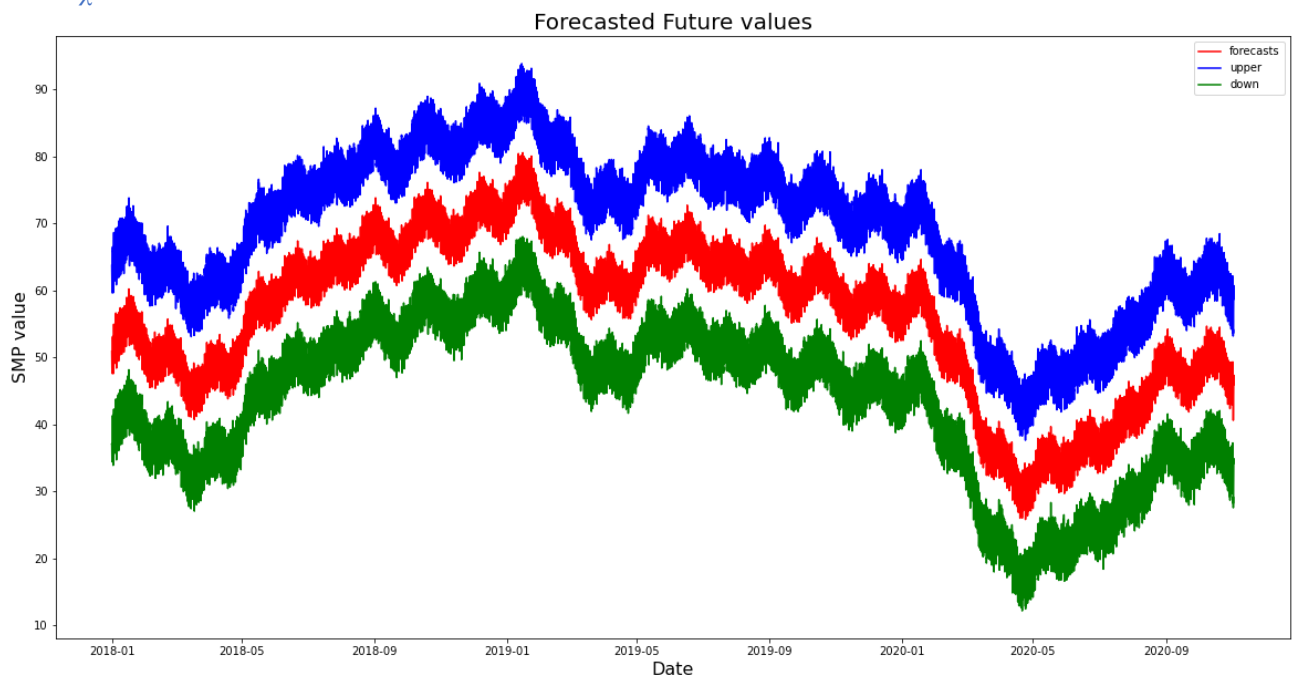
Παρακάτω φαίνεται το διάγραμμα πρόβλεψης με τα πάνω και κάτω όρια πρόβλεψης σε όλο το μελλοντικό σύνολο δεδομένων (future Dataframe). Πρόκειται για το όλο το σύνολο δεδομένων προσαυξημένο με το μελλοντικό διάστημα σε ώρες που είναι επιθυμητή η πρόβλεψη.

Σχ. 7.1



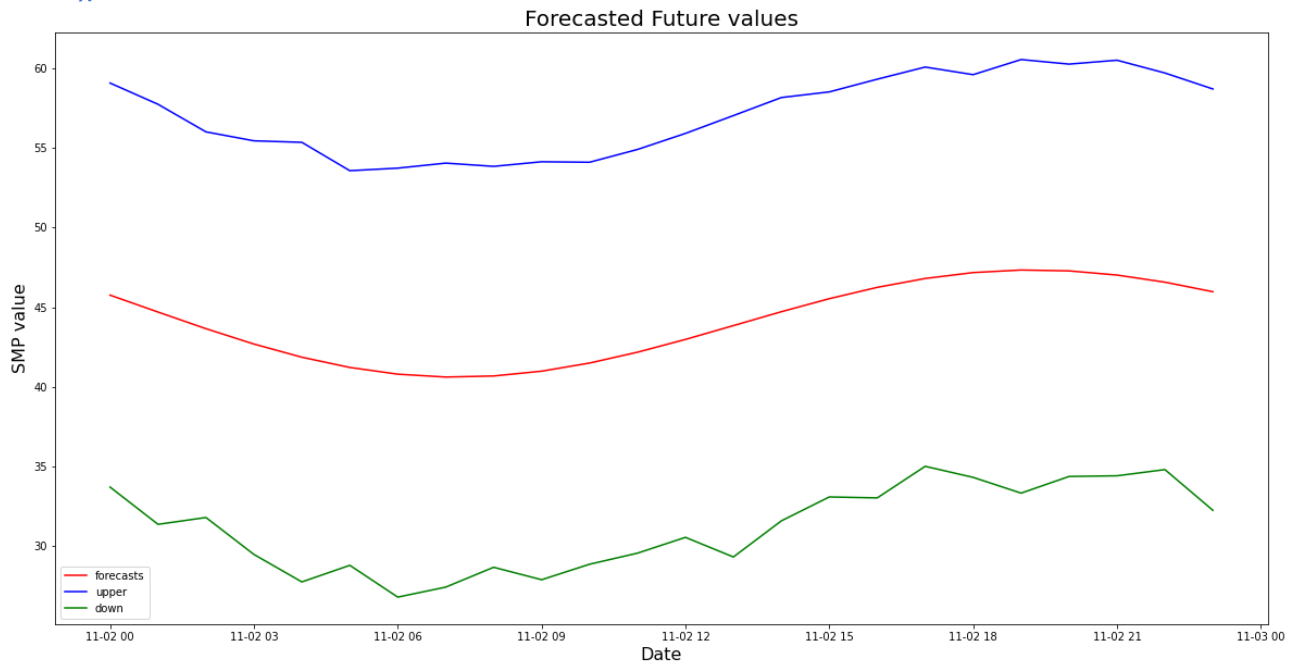
Σειρά έχουν οι προβλέψεις (yhat) μαζί με τα τιμές που αντιστοιχούν στο διάστημα εμπιστοσύνης 95% (yhat_lower, yhat_upper) για το συνολικό dataset και 24 ώρες μετά:

Σχ.7.2



Απομονώνοντας τις προβλέψεις 24 ώρες μετά σε ένα διάγραμμα μαζί με τις αντίστοιχες τιμές που για το διάστημα εμπιστοσύνης 95%, φαίνονται οι ακόλουθες κυματομορφές:

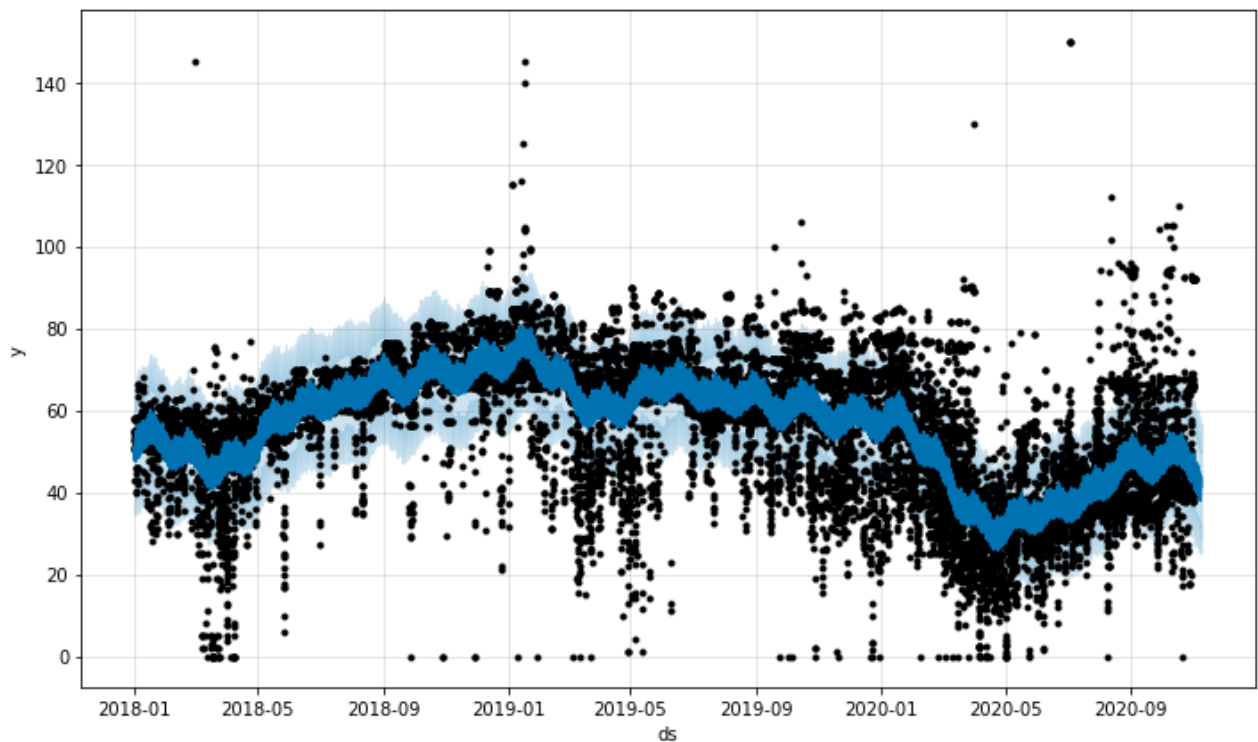
Σχ.7.3



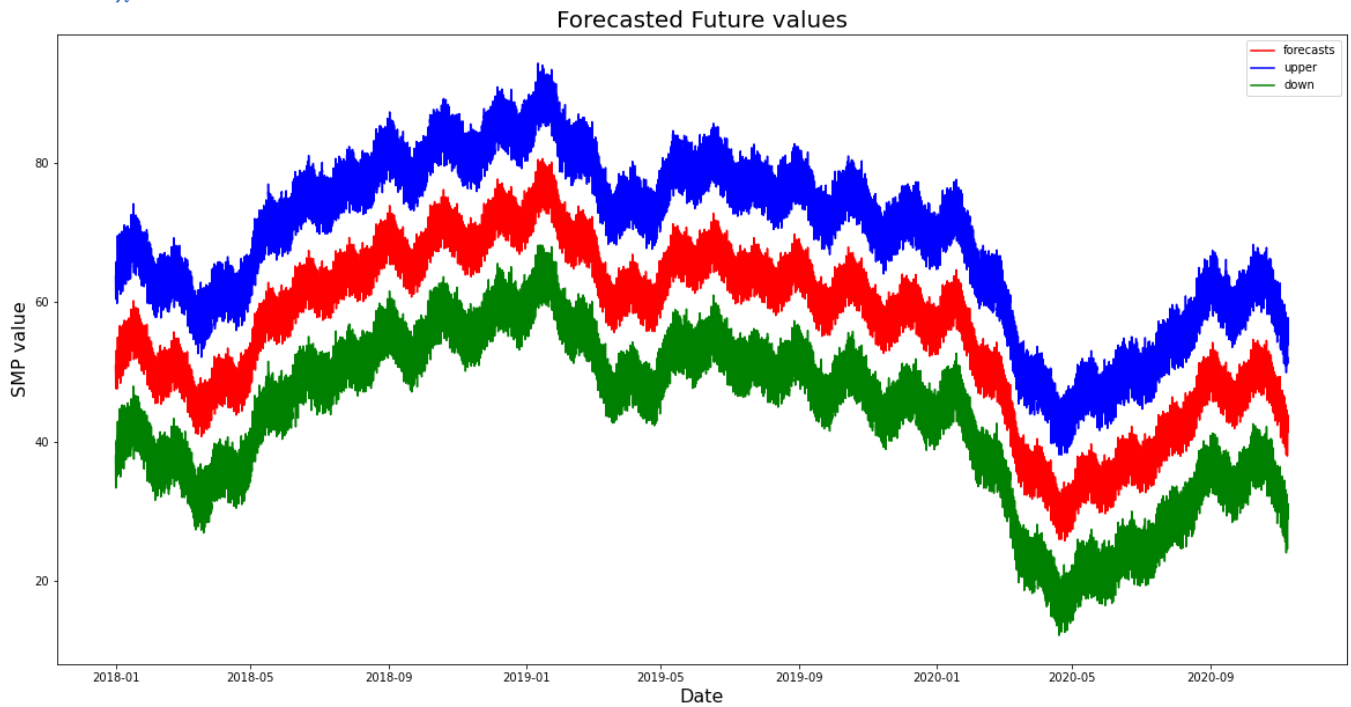
Οι τιμές αυτές παρουσιάζουν μέση τιμή 44.08293968445048, τυπική απόκλιση 2.3962739428334623, μεταβλητότητα 5.991786583411437 και κυμαίνονται σε ένα εύρος [40.60635218082987, 47.335485430090884].

Προχωρώντας στην εβδομαδιαία πρόβλεψη, προέκυψαν τα ακόλουθα διαγράμματα :

Σχ.7.4

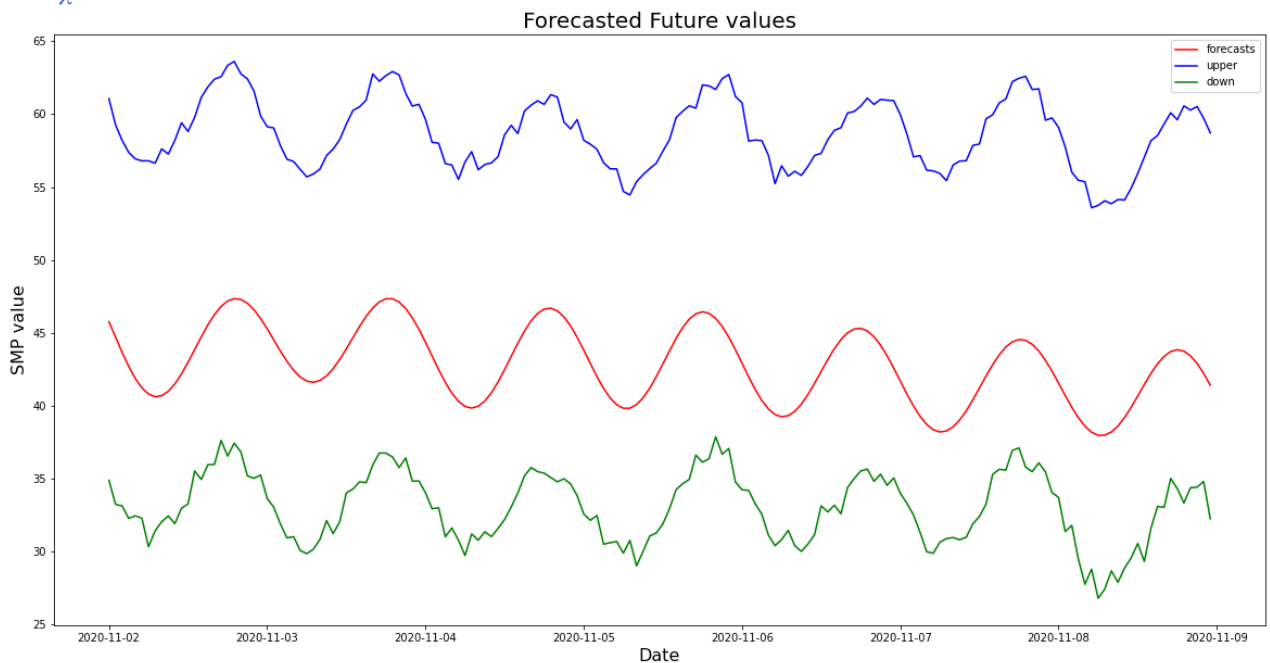


Σχ.7.5



Απομονώνοντας τις προβλέψεις μίας βδομάδας μετά σε ένα διάγραμμα, προκύπτει το ακόλουθο:

Σχ.7.6



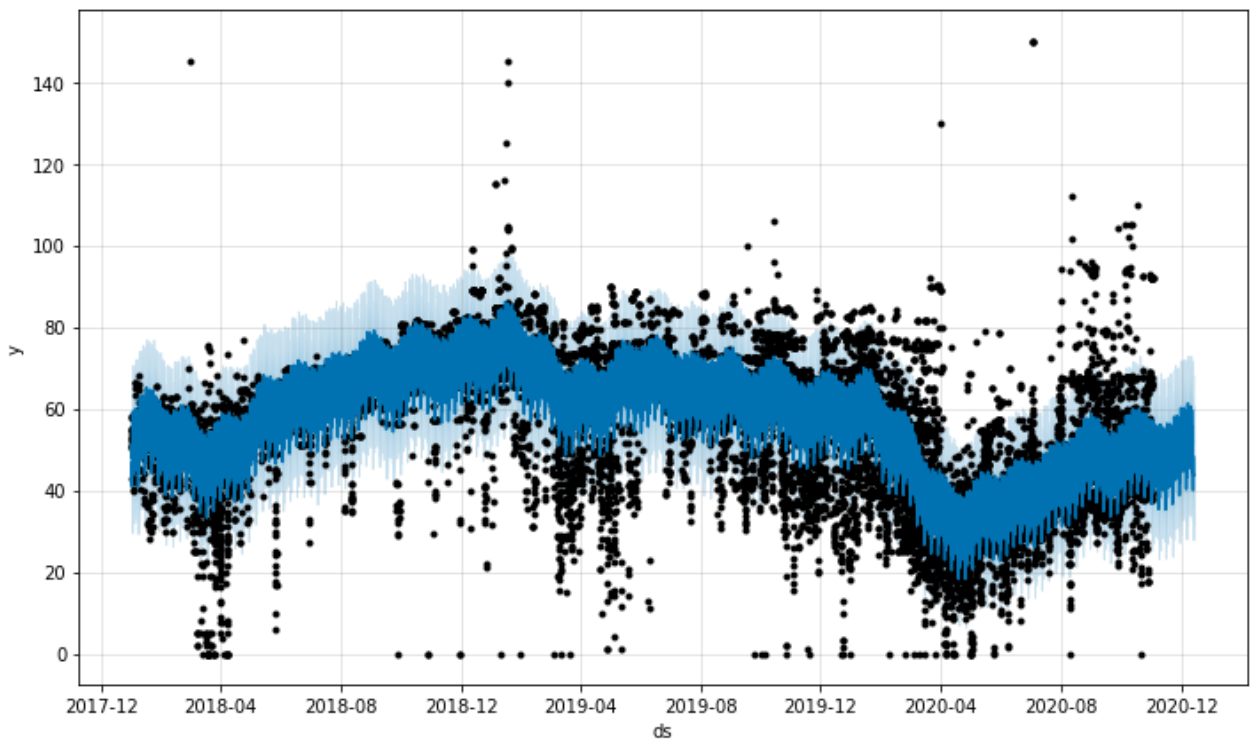
Οι τιμές αυτές παρουσιάζουν μέση τιμή 42.79302806169264, τυπική απόκλιση 2.5411194762358558, μεταβλητότητα 6.495954588867495 και κυμαίνονται σε ένα εύρος [37.956172387970796, 47.335485430090884].

7.1.2 Αποτελέσματα Μακροπρόθεσμης Πρόβλεψης

Η μακροπρόθεσμη πρόβλεψη ενός τόσο σημαντικού δείκτη για την Αγορά Ενέργειας όπως η Οριακή Τιμή Συστήματος, αποτελεί δύσκολο εγχείρημα και εξαρτάται από πολλούς παράγοντες που στο σύνολό τους αναφέρθηκαν σε προηγούμενα κεφάλαια. Το πιο σημαντικό στοιχείο που αποτέλεσε τροχοπέδη στην παρούσα εργασία ήταν η έλλειψη ιστορικών δεδομένων και γι' αυτό χρησιμοποιήθηκαν ωριαία δεδομένα στην ανάλυση. Παρ' όλα αυτά, ύστερα από δοκιμές προέκυψε ότι η βιβλιοθήκη FBPprophet μπορεί να προσεγγίσει ικανοποιητικώς μια μελλοντική πρόβλεψη και θεωρήθηκε σκόπιμο να εξεταστεί σε διάστημα 1000 ωρών μελλοντικά.

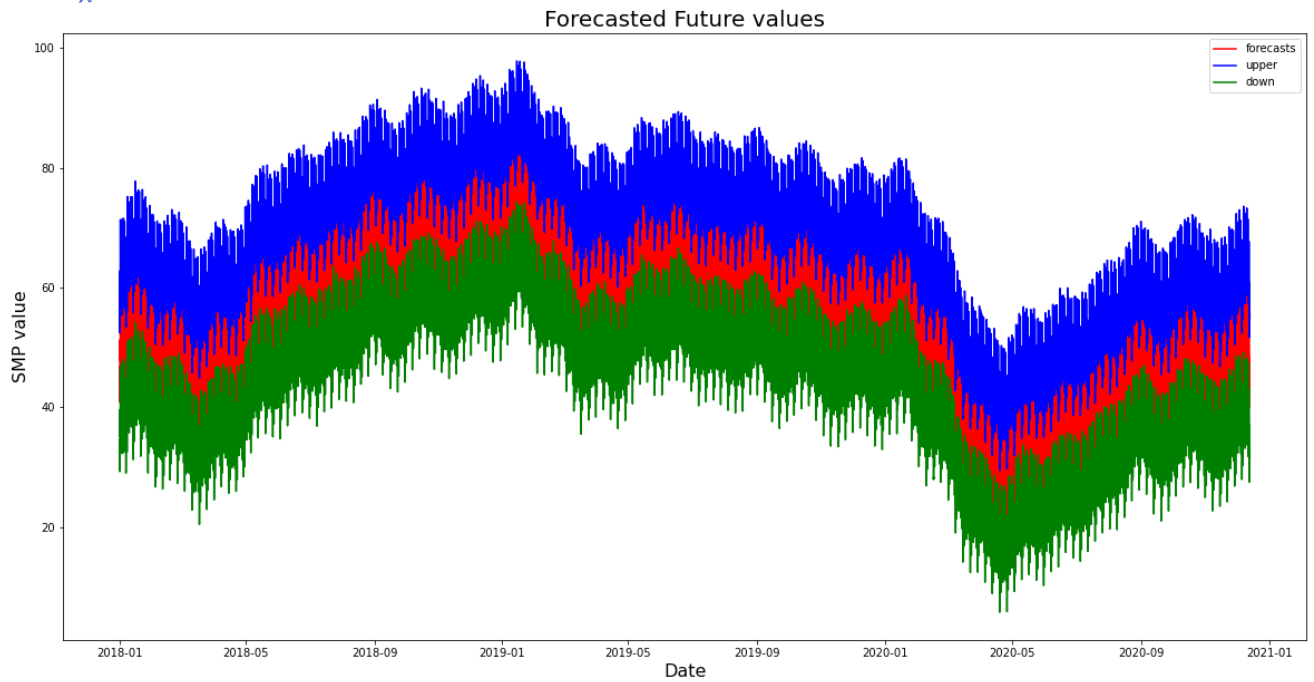
Παρακάτω φαίνεται το διάγραμμα με τα σημεία και τα πάνω και κάτω όρια των προβλέψεων σε μελλοντικό σύνολο εκπαίδευσης 1000 ωρών:

Σχ.7.7



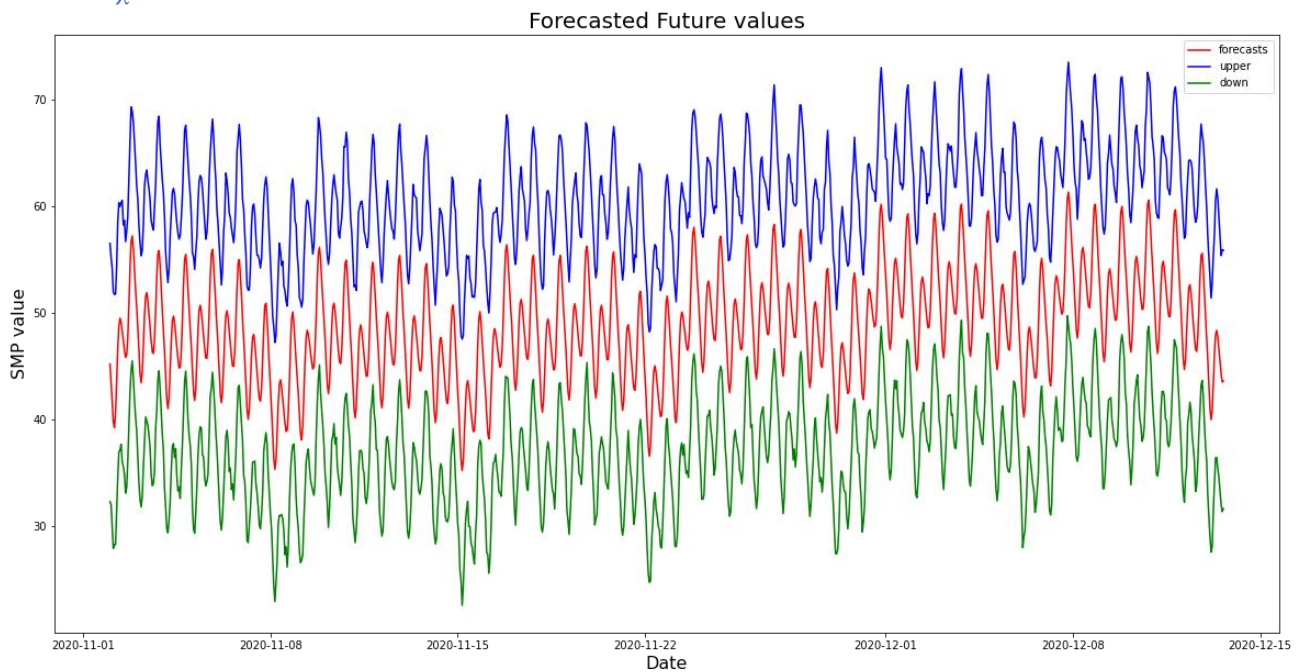
Το ίδιο διάγραμμα αλλά χωρίς τα σημεία φαίνεται παρακάτω με μεγαλύτερη ευκρίνεια και με τις τιμές που αντιστοιχούν στο 95% διάστημα εμπιστοσύνης:

Σχ.7.8



Απομονώνοντας κατά την ίδια πρακτική τις τελευταίες 1000 προβλέψεις, προκύπτει το παρακάτω διάγραμμα:

Σχ.7.9



Αξίζει επίσης να σημειωθεί ότι η λίστα με τις προβλέψεις για 1000 ώρες μετά έχουν μέση τιμή 48.7328778736595, τυπική απόκλιση 4.865830185935885, μεταβλητότητα 23.70000340176662 και κυμαίνονται σε ένα εύρος [35.2264944036852, 61.29598036725487].

Βιβλιογραφία

- Ioannis P. Panapakidis, Marios N. Moschakis (2019), 'Comparison of Machine Learning Models for the Prediction of System Marginal Price of Greek Energy Market'
- Andrianesis P., Biskas P. & Liberopoulos G. (2011) 'An overview of Greece's wholesale electricity market with emphasis on ancillary services', Electric Power Systems Research, Vol. 81, pp. 1631-1642.
- Box G.E.P., G.M. Jenkins (1970) 'Time series analysis: Forecasting and control', San Francisco: Holden-Day
- Assimakopoulos V., Nikolopoulos N. (2000) 'The theta model: a decomposition approach to forecasting', International Journal of Forecasting, Vol. 16, No. 4, pp. 521-530
- Gould P.G., Koehler A.B., Ord J.K., Snyder R.D., Hyndman R.J. & Farshid Vahid-Araghi (2008) 'Forecasting time series with multiple seasonal patterns', European Journal of Operational Research, Vol. 191, pp. 207-222.
- De Livera A.M., Hyndman R.J. (2009) 'Forecasting time series with complex seasonal patterns using exponential smoothing', Working paper 15/09, Department of Econometrics and Business Statistics, Monash University
- Torghaban S.S. (2010) 'Medium-Term Electricity Market Price Forecasting: A Data-driven Approach', Thesis for the Degree of Master of Science, Chalmers University of Technology, Goteborg, Sweden.
- Πετρόπουλος Φ. & Ασημακόπουλος Β. (2011), Επιχειρησιακές Προβλέψεις, Αθήνα: Εκδόσεις Συμμετρία.
- Egcheiridio_KSIE__Ekdosi_2.2__2016.10.27

Για τη συγγραφή της παρούσας Διπλωματικής, πληροφορίες αντλήθηκαν και από τους παρακάτω διαδικτυακούς τόπους:

- <https://www.rae.gr/>
- <https://www.deddie.gr/>
- <https://spark.apache.org/>
- <https://en.wikipedia.org/>
- <http://www.investingreece.gov.gr>
- <http://www.statbank.gr>
- <https://www.enexgroup.gr>

Παράρτημα Python

1. Δείκτες Σφάλματος

```
def calculate_metrics (actual, predicted):

    actual = actual if isinstance(actual, (np.ndarray)) else actual[:len(predicted)].to_numpy()
    pred = predicted if isinstance(predicted, (np.ndarray)) else predicted.to_numpy()
    metrics= []
    mse = mean_squared_error(actual, pred)
    print("Test MSE: %.5f" % mse)
    metrics.append(mse)
    rmse = np.sqrt(mean_squared_error(actual, pred))
    print("Test RMSE: %.5f" % rmse)
    metrics.append(rmse)
    mape= np.mean(np.abs((actual - pred) / actual)) * 100
    print("Test MAPE: %.5f" % mape)
    metrics.append(mape)
    smape = (100/len(actual)) * np.sum(2 * np.abs(pred - actual) / (np.abs(actual) + np.abs(pred)))
    print("Test sMAPE: %.5f" % smape)
    metrics.append(smape)
    mae= mean_absolute_error(actual, pred)
    print("Test MAE: %.5f" % mae)
    metrics.append(mae)
    marne = (100/len(actual)) * np.sum(np.abs(pred - actual) / (np.max(actual)))
    print("Test MARNE: %.5f" % marne)
    metrics.append(marne)
    return metrics
```

2. Έλεγχος στασιμότητας (Dickey_Fuller_Test)

```
def Dickey_Fuller_Test(data):
    t = data
    timeseries = adfuller(t)
    print('ADF Statistic: %f' % timeseries[0])
    print('p-value: %f' % timeseries[1])
    print('Critical Values:')
    for key, value in timeseries[4].items():
        print("\t%s: %.3f" % (key, value))
    if timeseries[0] > timeseries[4]["5%"]:
        print ("Failed to Reject Ho - Time Series is Non-Stationary")
    else:
        print("Reject Ho - Time Series is Stationary")
```

3. Ταχύς μετασχηματισμός Fourier (FFT)

```
fft_output = fft.fft(series)
power = np.abs(fft_output)
freq = fft.fftfreq(len(series))

mask = freq >= 0
freq = freq[mask]
```

```
power = power[mask]
```

4. Μοντέλο TBATS

```
estimator = TBATS(seasonal_periods=(12, 345.67*24))
short_tbats_model = estimator.fit(first_two_years_series)
y_forecast = short_tbats_model.forecast(steps=168)
```

5. SARIMA με Pyramid

```
arima_model = auto_arima(first_two_years_series,
start_p=0, d=0, start_q=0,
max_p=1, max_q=1,
max_P=1,max_Q=1, max_D=1,
max_order=5, m=12,
                        seasonal=True, stationary=True,
information_criterion='aic',
                        alpha=0.05,
                        trace=True,
error_action='warn',
suppress_warnings=True,
                        stepwise=True,
n_jobs=1,
maxiter=100)
y_arima_forecast_168 = arima_model.predict(n_periods=168)
```

6. GridSearch για το βέλτιστο Sarimax μοντέλο για βραχυπρόθεσμη πρόβλεψη

```
p = d = q = range(0, 2)

# Generate all different combinations of p, q and q triplets
pdq = list(itertools.product(p, d, q))
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]
test= range(0, 10)
for param in test:

    for param_seasonal in seasonal_pdq:

        mod = sm.tsa.statespace.SARIMAX(first_two_years_series,
                                        order=(0,0,param),
seasonal_order=param_seasonal,
enforce_stationarity=False,
enforce_invertibility=False)

        results = mod.fit()

        print('ARIMA order{ } x{ } - AIC: {}'.format(param, param_seasonal, results.aic))
```

7. Sarimax μοντέλο για βραχυπρόθεσμη πρόβλεψη

```
arima_model_2 = sm.tsa.statespace.SARIMAX(first_two_years_series,
order=(0, 0, 9),seasonal_order=(1,1,1,12))
```

```
result = arima_model_2.fit()
arimax_forecast= result.forecast(steps=168)
```

8. Προσθήκη συνιστωσών Fourier σε Sarimax μοντέλο

```
def add_fourier_terms(df, year_k, week_k, day_k):
    """
    df: dataframe to add the fourier terms to
    year_k: the number of Fourier terms the year period should have. Thus the model will be fit on
    2*year_k terms (1 term for
    sine and 1 for cosine)
    week_k: same as year_k but for weekly periods
    day_k: same as year_k but for daily periods
    """

    for k in range(1, year_k+1):
        # year has a period of 365.25 including the leap year
        df['year_sin'+str(k)] = np.sin(2 *k* np.pi * df.index.dayofyear/365.25)
        df['year_cos'+str(k)] = np.cos(2 *k* np.pi * df.index.dayofyear/365.25)

    for k in range(1, week_k+1):

        # week has a period of 7
        df['week_sin'+str(k)] = np.sin(2 *k* np.pi * df.index.dayofweek/7)
        df['week_cos'+str(k)] = np.cos(2 *k* np.pi * df.index.dayofweek/7)

    for k in range(1, day_k+1):

        # day has period of 24
        df['hour_sin'+str(k)] = np.sin(2 *k* np.pi * df.index.hour/24)
        df['hour_cos'+str(k)] = np.cos(2 *k* np.pi * df.index.hour/24)
```

9. Μοντέλο Sarimax με συνιστώσες Fourier

```
model_opt = sm.tsa.statespace.SARIMAX(first_two_years.SMPS, order=(0,0,3),
    seasonal_order=(0, 1, 1, 12), exog = exog_to_train)
results = model_opt.fit()
pred = results.get_forecast(steps = 24*7, exog = exog_to_test.iloc[:24*7, :])
pred_ci_2 = pred.conf_int()
```

10. Δημιουργία dataframes για το FBProphet

```
def init_Training_df_for_prophet(dataframe):
    train_df = dataframe.copy()
    train_df.reset_index(inplace=True)
    train_df.rename(columns={'date':'ds', 'SMPS':'y'},inplace=True)
    train_df['ds']=to_datetime(train_df['ds'])
    return train_df

def init_Test_df_for_prophet(dataframe):
    test_df = dataframe.copy()
    test_df.reset_index(inplace=True)
```

```
test_df.rename(columns={'date':'ds', 'SMPS':'y'},inplace=True)
test_df['ds']=to_datetime(test_df['ds'])
df_predict=test_df[['ds']]
return test_df,df_predict
```

11. Μοντέλο FBProphet με αυτόματη και manual ρύθμιση εποχικότητας

```
auto_model = Prophet(yearly_seasonality=True)
auto_model.fit(train_df)
auto_forecast = auto_model.predict(df_predict)
```

```
manual_model = Prophet()
manual_model.add_seasonality(name='daily', period = 12, fourier_order = 15, prior_scale=20)
manual_model.add_seasonality(name='weekly', period = 49*168, fourier_order = 15, prior_scale=20)
manual_model.fit(train_df)
manual_forecast = manual_model.predict(df_predict)
```

12. Υπολογισμός bias για την επιλογή του κατάλληλου μοντέλου

```
def mbe(predicted,actual):
    actual = actual if isinstance(actual,(np.ndarray)) else actual[:len(predicted)].to_numpy()
    predicted = predicted if isinstance(predicted,(np.ndarray)) else predicted.to_numpy()
    return 1/len(actual) * np.sum(predicted - actual)
```