
Στατιστική ανάλυση δεδομένων παραγωγής ηλεκτρικής ενέργειας αιολικών πάρκων

Διπλωματική Εργασία
Γεώργιος Ι. Εξαρχάκος



Πολυτεχνείο Κρήτης
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Ακρωτήρι, Τ.Κ. 731 00, Χανιά, Κρήτη

Επιβλέπων: Διονύσιος Χριστόπουλος, Καθηγητής, Σχολή Ηλεκτρολόγων Μηχανικών και
Μηχανικών Υπολογιστών, Πολυτεχνείου Κρήτης

Τριμελής Επιτροπή:

Διονύσιος Χριστόπουλος, Καθηγητής Η.Μ.Μ.Υ

Σταυραράκης Γεώργιος, Καθηγητής Η.Μ.Μ.Υ

Κανέλλος Φώτιος, Καθηγητής Η.Μ.Μ.Υ

Περίληψη

Οι ανθρώπινες κοινωνίες εργαλειοποιούν τον άνεμο από αρχαιοτάτων χρόνων για την κάλυψη διαφόρων αναγκών τους. Στις μέρες μας, μετατρέπουν ένα μέρος της κινητικής του ενέργειας σε ηλεκτρική. Η ενέργεια αυτή είναι πολύτιμη τόσο για την ομαλή λειτουργία της κοινωνίας όσο και για την ανάπτυξή της. Για αυτό το λόγο η επίτευξη σωστού προγραμματισμού των συστημάτων παραγωγής ενέργειας κάθε κοινωνίας φέρει τεράστια σημασία για την βιωσιμότητά της. Για να επιτελεστεί το έργο αυτό, απαιτείται η πρόβλεψη της ηλεκτρικής ενέργειας που παράγεται από τα αιολικά πάρκα και εντάσσονται στο σύστημα παραγωγής να είναι ακριβής. Με βάση αυτήν μπορούν επίσης να παρθούν οι κατάλληλες τεχνοοικονομικές αποφάσεις.

Στη παρούσα εργασία, ερευνήθηκε η δυνατότητα παραγωγής εύστοχων προβλέψεων από την ανάλυση χρονοσειρών μέσων παραγόμενων ποσοτήτων ηλεκτρικής ενέργειας. Ερευνήθηκε, εάν η θεωρία των χρονοσειρών για την μελέτη και ανάπτυξη στατιστικών μοντέλων επαρκεί στη περίπτωση αυτή. Πρόκειται για μια ιδιάζουσα κατάσταση όπου η απρόβλεπτη συμπεριφορά του ανέμου, δημιουργεί χρονοσειρές ευμετάβλητες. Από τη μελέτη αυτών, συλλέχθηκαν πληροφορίες και έγιναν οι απαραίτητες μετατροπές ώστε να εκπαιδευτούν κατάλληλα τα γραμμικά μοντέλα MA, AR, ARMA, ARIMA και SARIMA που θα παράξουν τις προβλέψεις.

Μελετήθηκαν και αναλύθηκαν δύο διαφορετικές χρονοσειρές. Η πρώτη χρονοσειρά δημιουργήθηκε από δεδομένα που αντλήθηκαν από την πλατφόρμα μηχανικής μάθησηςKaggle. Αποτελούνταν από τις μέσες μηνιαίες ποσότητες ηλεκτρικής ενέργειας που παρήχθησαν από 01/1985 έως 01/2018 σε κάποια μονάδα παραγωγής. Η δεύτερη χρονοσειρά συντέθηκε από δεδομένα που παρέχονταν από μια εταιρία διαχείρισης αιολικών πάρκων και εμπεριείχαν τις μέσες ποσότητες ενέργειας που παρήγαγε ένα αιολικό πάρκο στη Θράκη ανά δέκα λεπτά από 01/01/2018 έως 31/12/2020.

Η δεύτερη χρονοσειρά επιλέχθηκε να αναδιαμορφωθεί και να εκφράζει τις μέσες ημερήσιες ποσότητες παραγόμενης ηλεκτρικής ενέργειας για το ημερολογιακό έτος 2020. Ο μεγάλος όγκος των δεδομένων της αρχικής χρονοσειράς καθιστούσε τη μελέτη τους αδύνατη με τη χρήση ενός απλού συμβατικού υπολογιστή. Έτσι, πλέον μελετούνταν δύο χρονοσειρές παρεμφερούς μήκους αλλά διαφορετικού χρονικού βήματος.

Για την κάθε χρονοσειρά πριν τη δημιουργία της γινόταν έλεγχος ποιότητας των δεδομένων που θα την αποτελούσαν. Ακολουθούσε η διερευνητική ανάλυση της φύσης της με βάση το διάγραμμα απεικόνισης, αυτοσυσχέτισης, ανάλυσης συνιστωσών, το ιστόγραμμα της και το ιστόγραμμα υπολοίπων της. Εάν η χρονοσειρά δεν βρισκόταν σε κατάσταση στασιμότητας, εκτελούνταν οι κατάλληλες διεργασίες μετατροπής και επανελεγχόταν η φύση της.

Η εκπαίδευση των βέλτιστων μοντέλων MA, AR, ARMA, ARIMA και SARIMA πραγματοποιούνταν μόνο αφού η χρονοσειρά είχε επέλθει όσο πιο κοντά γινόταν στην στασιμότητα. Για να ορισθούν οι τιμές των παραμέτρων του κάθε μοντέλου αξιοποιούνταν ένας αλγόριθμος εξαντλητικής αναζήτησης. Το μοντέλο με τις καλύτερες παραμέτρους ήταν αυτό που έφερε τη μικρότερη τιμή στο κριτήριο επιλογής AIC. Κάθε πρόβλεψη παραγόταν με βάση την τεχνική της διασταυρωτικής επαλήθευσης. Για την αξιολόγηση των προβλέψεων που παρήγαγαν τα μοντέλα επιστρατεύτηκε μια σειρά κριτηρίων αξιολόγησης.

Από τη μελέτη των αποτελεσμάτων που έφεραν τα κριτήρια αξιολόγησης εξήχθησαν πολύτιμα συμπεράσματα. Στην πρώτη περίπτωση οι προβλέψεις έφεραν πολύ μικρή απόκλιση ενώ στην δεύτερη δεν παρουσίαζαν την ίδια ευστοχία. Η διαφορά έγκειται στην ποιότητα των δεδομένων που αξιοποιούνταν και στο βαθμό στασιμότητας που επιτυγχανόταν για τη κάθε χρονοσειρά. Η δεύτερη χρονοσειρά η οποία παρουσίαζε μεγαλύτερη μεταβλητότητα έφερε λοιπόν χειρότερες προβλέψεις από την πρώτη στην οποία οι διεργασίες μετατροπής σταθεροποίησαν σημαντικά την μέση τιμή και διασπορά της.

Σε μια προσπάθεια παραγωγής πιο εύστοχων προβλέψεων για τη δεύτερη χρονοσειρά αξιοποιήθηκε ο αλγόριθμος μηχανικής μάθησης XGBoost. Αναπτύχθηκαν τρία διαφορετικά μοντέλα και παρήχθησαν τρεις διαφορετικές προβλέψεις. Πριν την εκπαίδευση των μοντέλων αυξήθηκε ο όγκος των δεδομένων εκπαίδευσης συμπεριλαμβάνοντας σε αυτά όσα εξέφραζαν τις μέσες ημερήσιες ποσότητες ενέργειας που παρήχθησαν το ημερολογιακό έτος 2019. Κατόπιν εκπαιδεύτηκε το πρώτο μοντέλο. Αξιοποιούσε τις ίδιες πληροφορίες με τα γραμμικά μοντέλα και οι βέλτιστες τιμές των υπερπαραμέτρων του μοντέλου ανευρίσκονταν μέσω μιας εξαντλητικής αναζήτησης. Το μοντέλο αυτό παρήγαγε τη χειρότερη πρόβλεψη.

Για την εκπαίδευση του δεύτερου μοντέλου, XGB-2, εξήχθησαν περισσότερες πληροφορίες από τις ήδη αξιοποιησίδες. Γνωρίσματα χρόνου ή τιμών υστερημένων κατά κάποια χρονικά βήματα συνέβαλαν στην απόδοση καλύτερων προβλέψεων. Μάλιστα αποτελούσαν τις καλύτερες τόσο σε απόκλιση κατά μέγεθος όσο και κατά κατεύθυνση. Το κύριο μειονέκτημα του XGB-2 αποτελούσε η υπερπροσαρμογή που εμφάνιζε στα δεδομένα. Για να αντιμετωπιστεί το πρόβλημα αυτό αναπτύχθηκε ένα τρίτο μοντέλο, το XGB-3. Χρησιμοποιούσε τα ίδια χαρακτηριστικά γνωρίσματα αλλά οι τιμές στις υπερπαραμέτρους του είχαν μεταβληθεί έτσι ώστε να μειώνουν την υπερεκτίμηση. Πράγματι στις 8 από τις 10 παρατηρήσεις πέτυχε τον σκοπό του. Ωστόσο, το μέγεθος των σφαλμάτων του στις άλλες δύο περιπτώσεις επισκίασαν τη συνολική του ικανότητα.

Λέξεις Κλειδιά

Χρονοσειρές, προεπεξεργασία δεδομένων, διερευνητική ανάλυση χρονοσειρών, μετατροπή χρονοσειρών, εκπαίδευση γραμμικών μοντέλων χρονοσειρών, AR, MA, ARMA, ARIMA, SARIMA, κριτήρια αξιολόγησης προβλέψεων, αλγόριθμος μηχανικής μάθησης XGBoost, δημιουργία χαρακτηριστικών γνωρισμάτων (feature engineering), υπερπροσαρμογή (overfitting), μέση ποσότητα παραγόμενης ηλεκτρικής ενέργειας

Abstract

Human societies have been instrumentalising the wind since ancient times to meet their various needs. Nowadays, they convert part of its kinetic energy into electrical energy. This energy is valuable both for the smooth functioning of society and for its development. That is why achieving proper planning of the power generation systems of any society carries immense importance for its sustainability. In order to accomplish this task, it is required that the forecasting of electricity generated by wind farms and integrated into the generation system be accurate. On this basis, appropriate techno-economic decisions can also be taken.

In this paper, the possibility of generating accurate forecasts by analysing time series of average electricity production has been investigated. It was investigated whether time series theory for the study and development of statistical models is sufficient in this case. This is a peculiar situation where the unpredictable behaviour of the wind generates time series that are highly variable. From the study of these, information was collected and the necessary transformations were made to properly train the linear models MA, AR, ARMA, ARIMA and SARIMA that will produce the forecasts.

Two different time series were studied and analysed. The first time series was created from data extracted from the machine learning platform Kaggle. It consisted of the average monthly quantities of electricity produced from 01/1985 to 01/2018 at a power plant. The second time series was composed from data provided by a wind farm management company and contained the average amounts of energy produced by a wind farm in Thrace every ten minutes from 01/01/2018 to 31/12/2020.

The second time series was chosen to be reconfigured to reflect the average daily quantities of electricity produced for the calendar year 2020. The large volume of data in the original time series made it impossible to study them using a simple conventional computer. Thus, two time series of similar length but different time steps were now studied.

For each time series, a quality check was carried out on the constituent data before it was created. This was followed by an exploratory analysis of its nature based on its plot, autocorrelation, component analysis, its histogram and its residuals histogram. If the time series was not in a stationary state, the appropriate transformation processes were performed and its nature was re-examined.

Training of the optimal models MA, AR, ARMA, ARIMA and SARIMA was only performed after the time series had come as close as possible to stationarity. An exhaustive

search algorithm was utilized to define the parameter values of each model. The model with the best parameters was the one with the lowest value on the selection criterion AIC. Each prediction was generated based on the cross-validation technique. A set of evaluation criteria was employed to evaluate the predictions produced by the models.

Valuable conclusions were drawn from the study of the results produced by the evaluation criteria. In the first case the predictions showed very little deviation while in the second case they did not show the same accuracy. The difference lies in the quality of the data used and the degree of stationarity achieved for each time series. The second time series which exhibited greater variability therefore produced worse predictions than the first in which the conversion processes significantly stabilised the mean and dispersion.

In an effort to produce more accurate predictions for the second time series, the machine learning algorithm XGBoost was utilized. Three different models were developed and three different forecasts were produced. Before training the models, the volume of training data was increased by including in it those expressing the average daily amounts of energy produced in the calendar year 2019. Then the first model was trained. It utilized the same information as the linear models and the optimal values of the model hyperparameters were found through an exhaustive search. This model produced the worst prediction.

To train the second model, XGB-2, more information was extracted than already exploitable. Time traits or values lagged behind by some time steps contributed to yielding better predictions. In fact, they were the best in both magnitude and directional deviation. The main disadvantage of XGB-2 was the overfitting it exhibited to the data. To address this problem a third model, XGB-3, was developed. It used the same features but the values in its hyperparameters were altered to reduce the overestimation. Indeed in 8 out of 10 observations it achieved its purpose. However, the magnitude of his errors in the other two cases overshadowed his overall ability.

Keywords

Time series, data preprocessing, exploratory time series analysis, time series transformation, time, linear time series model training, AR, MA, ARMA, ARIMA, SARIMA, forecast evaluation criteria, XGBoost machine learning algorithm, feature engineering, overfitting, average amount of electricity generated

χαρισμένο στον Αιμίλιο

Ευχαριστίες

Με την επιτυχή ολοκλήρωση και παράδοση της παρούσας διπλωματικής εργασίας, κλείνει άρτια ο κύκλος των προπτυχιακών μου σπουδών στο τμήμα των Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης.

Κρίνω δίκαιο να ευχαριστήσω όλους όσους συνέβαλλαν στην ολοκλήρωση των σπουδών μου. Αρχικά, θα ήθελα να απευθυνθώ στο περιβάλλον των αγαπημένων μου φίλων. Οι φίλοι μου Αλέξανδρος Ζιώγας, Μανώλης Μυλωνάκης, Αλέξανδρος Παπαδάκης και Κωνσταντίνος Σταματόπουλος βρίσκονταν πάντα στο πλάι μου. Είμαι ευγνώμων για την υπομονή που δείξαν στο πρόσωπό μου, την ειλικρίνεια την οποία έφεραν στο λόγο τους, την αγάπη και την ευγένεια με την οποία περιέβαλαν τη σχέση μας και τα γέλια και τα δάκρυα που μοιραστήκαμε όλα αυτά τα χρόνια στα Χανιά. Με ανέπλασαν και ήταν πάντα μια πηγή δύναμης για εμένα.

Ακόμη, οφείλω ένα μεγάλο ευχαριστώ προς όλους τους καθηγητές μου για τις γνώσεις και την εν δύναμή εργασιακή εξέλιξη που μου προσέφεραν. Ιδιαίτερη μνεία θα ήθελα να κάνω για τη βοήθεια που πάντα με ευχαρίστηση μου έδινε ο επιβλέπων εις την παρούσα εργασία καθηγητής μου κύριος Χριστόπουλος Διονύσιος. Τον ευχαριστώ θερμά για την συνεργασία μας, το ενδιαφέρον, τις συμβουλές, την εμπύχωση και την καθοδήγηση του σε όλα τα στάδια της προκείμενης εργασίας. Οι γνώσεις του και το ήθος του συνέβαλλαν καίρια στη συγγραφή της παρούσας μελέτης και στη διεύρυνση των γνώσεων μου στο θεματικό πεδίο της ανάλυσης δεδομένων.

Τολμώ να πω ότι χάριν στους γονείς μου και τον αδερφό μου έφτασα έως εδώ. Η αγάπη, τα διδάγματα και αξίες με τις οποίες με μεγάλωσαν έπαιξαν καθοριστικό ρόλο στην υπέρβαση κάθε εμποδίου που εμφανίστηκε στη ζωή μου. Η προσφορά τους ήταν, είναι και θα είναι πάντα ανεκτίμητη και ανιδιοτελής. Τέλος, θα ήθελα να ευχαριστήσω τη Νάγια που με αντέχει.

Περιεχόμενα

Περίληψη	i
Abstract	iii
Ευχαριστίες	vii
Περιεχόμενα	xi
Κατάλογος Σχημάτων	xvi
Κατάλογος Πινάκων	xviii
1 Εισαγωγή	1
1.1 Οργάνωση Συγγράμματος	2
2 Η Ιστορία της Ενέργειας	5
2.1 Σύνοψη	5
2.2 Αρχαίοι Χρόνοι και Μεσαίωνας	5
2.3 Μετάβαση στη Νεότερη Εποχή με την Αξιοποίηση Ορυκτών Καυσίμων	6
2.3.1 Οφέλη και Πλεονεκτήματα	7
2.3.2 Αρνητικές Συνέπειες για την Φύση και τον Άνθρωπο	9
2.4 Παρόν και Μέλλον με Στόχο την Επίτευξη Ουδέτερου Ισοζυγίου Άνθρακα	13
3 Μελέτη και Ανάλυση του Ανέμου	19
3.1 Σύνοψη	19
3.2 Η Γέννηση του Ανέμου στη Γη	19
3.3 Βασικά Γνωρίσματα του Ανέμου	20
3.4 Κατηγοριοποίηση Ανέμων	22
3.5 Ανεμολογικά Φαινόμενα της Ελλάδας	29
4 Θεμελιώδεις Αρχές Αιολικής Παραγωγής	31
4.1 Σύνοψη	31
4.2 Αιολική Ενέργεια	31
4.3 Ανεμογεννήτριες	33

4.4	Ανεμογεννήτριες Οριζοντίου Άξονα	34
4.4.1	Συστατικά Μέρη Ανεμογεννήτριας	34
4.4.2	Λειτουργία Ανεμογεννήτριας	36
4.4.3	Συντελεστής Ισχύος	37
4.4.4	Καμπύλη Ισχύος Ανεμογεννήτριας	38
4.5	Αιολικά Πάρκα	40
4.5.1	Σχεδίαση και Ανάπτυξη	40
4.5.2	Ενσωμάτωση στο Δίκτυο Ηλεκτρικής Ενέργειας	42
5	Θεωρητικό Υπόβαθρο	45
5.1	Χρονοσειρές	45
5.2	Ανάλυση Στάσιμων Χρονοσειρών	46
5.2.1	Αυτοσυσχέτιση	48
5.3	Ανάλυση Μη Στάσιμων Χρονοσειρών	50
5.4	Μετατροπή μη Στάσιμων Χρονοσειρών σε Στάσιμες	51
5.4.1	Απαλοιφή Τάσης	52
5.4.2	Απαλοιφή Περιοδικότητας	55
6	Πρόβλεψη Χρονοσειρών	57
6.1	Χρονοσειρές Αυτοπαλινδρόμησης AR(p) model	57
6.2	Χρονοσειρές Κινούμενου Μέσου MA(q) model	57
6.3	Αυτοπαλινδρομες Χρονοσειρές Κινούμενου Μέσου ARMA model(p, q)	58
6.4	Ολοκληρωμένες χρονοσειρές τάξης d Integrated process (d)	58
6.5	Αυτοπαλινδρομες Ολοκληρωμένες Κινούμενου Μέσου Χρονοσειρές ARIMA model(p, d, q)	59
6.6	Εποχικές Αυτοπαλινδρομες Ολοκληρωμένες Κινούμενου Μέσου Χρονοσειρές (SARIMA) seasonal ARIMA model	59
6.7	Διαδικασία Box-Jenkins	60
6.7.1	Ταυτοποίηση	60
6.7.2	Εκτίμηση	61
6.7.3	Διαγνωστικός Έλεγχος	62
6.7.4	Διαδικασία Πρόβλεψης	62
6.7.5	Αξιολόγηση προβλέψεων	66
6.8	XGBoost	68
7	Επεξεργασία Ενεργειακών Δεδομένων με Μεθόδους Χρονοσειρών	75
7.1	Εισαγωγή	75
7.2	Ανάλυση Δεδομένων Παραγωγής Ηλεκτρικής Ενέργειας (Kaggle Dataset)	76
7.2.1	Προ επεξεργασία Δεδομένων	76
7.2.2	Διερευνητική Ανάλυση και Έλεγχος Στασιμότητας της Χρονοσειράς	78
7.2.3	Μετατροπή της Χρονοσειράς σε Στάσιμη	85
7.2.4	Παραγωγή Προβλέψεων με Γραμμικά Μοντέλα	91

7.2.5	Αξιολόγηση Προβλέψεων	99
7.2.6	Σύγκριση Προβλέψεων με Αντίστοιχα Μοντέλα	102
7.3	Ανάλυση Δεδομένων Παραγωγής Ηλεκτρικής Ενέργειας από το Αιολικό Πάρκο 2 στην Θράκη	103
7.3.1	Προ επεξεργασία Δεδομένων	104
7.3.2	Διερευνητική Ανάλυση και Έλεγχος Στασιμότητας	105
7.3.3	Μετατροπή της Χρονοσειράς σε Στάσιμη	112
7.3.4	Παραγωγή Προβλέψεων με Γραμμικά Μοντέλα	125
7.3.5	Αξιολόγηση Προβλέψεων	134
7.3.6	Προβλέψεις με Χρήση XGBoost Μοντέλων και Αξιολόγηση	138
8	Συμπεράσματα	153

Κατάλογος Σχημάτων

2.1	Ιστορική εξέλιξη παραγόμενων ποσοτήτων ρύπων CO ₂ ανά τον κόσμο [21]. . .	14
2.2	Μελλοντικές επενδύσεις που απορρέουν από το ελληνικό σχέδιο για την ενέργεια και το κλίμα έως το 2030 [2].	15
3.1	Η γενική κυκλοφορία της ατμόσφαιρας [29].	24
3.2	Στάδια εξέλιξης κυκλώνα.	26
3.3	Οι μηχανισμοί της θαλάσσιας (αριστερά) και απόγειας αύρας (δεξιά) [29]. . . .	27
3.4	Μηχανισμοί αύρων κοιλάδας (πάνω) και βουνού (κάτω) [1].	28
4.1	Περιοχή σάρωσης μια ανεμογεννήτριας [33].	33
4.2	Βασικά μέρη μια ανεμογεννήτριας [18].	36
4.3	Βασικά μέρη μια ανεμογεννήτριας (ανάλυση εσωτερικού ατράχτου) [33].	37
4.4	Καμπύλες συντελεστή ισχύος [19].	38
4.5	Τυπική καμπύλη ισχύος [19].	39
4.6	Τυπική χωροθέτηση ανεμογεννητριών σε αιολικό πάρκο [18].	42
4.7	Η διαδικασία διασύνδεσης της αιολικής παραγωγής στο δίκτυο [18].	44
5.1	Διάγραμμα αυτοσυσχέτισης μιας χρονοσειράς που οι τιμές τις αποτυπώνουν τα ύψη των κυμάτων του νερού σε χιλιοστά δειγματοληπτημένα ανά 0.1 δευτερόλεπτα [10].	49
5.2	Χρονοσειρά του δείκτη Χρηματιστηρίου Αξιών Αθηνών (XAA) κατά το κλεισιμο (στοχαστική τάση) [26].	50
5.3	Χρονοσειρά δείκτη κυκλοφοριακού συμφορισμού (ντετερμινιστική τάση) [10]. .	51
5.4	Εφαρμογή Μετασχηματισμού Box-Cox σε μη στάσιμη χρονοσειρά [26].	55
6.1	Η διαδικασία Box-Jenkins.	60
6.2	Η διαδικασία πρόβλεψης που αναπτύχθηκε και εφαρμόστηκε στη παρούσα διπλωματική εργασία.	63
7.1	Γραμμές του αχατέργαστου αρχείου που περιέχει τα δεδομένα παραγωγής ενέργειας.	77
7.2	Δημιουργία Χρονοσειράς με <code>pandas DataFrame</code>	78
7.3	Απεικόνιση χρονοσειράς	79

7.4	Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας (πρώτο). Kernel Density Estimator (μεσαίο). Συνδυασμός των δύο από πάνω γραφημάτων και σύγκριση των δεδομένων με την κανονικής κατανομής (τελευταίο).	80
7.5	Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στα δεδομένα παραγωγής ηλεκτρικής ενέργειας.	81
7.6	Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας.	83
7.7	Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας.	83
7.8	Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας - ACF plot (150 lags).	84
7.9	Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας - ACF plot.	84
7.10	Αποτέλεσμα του ελέγχου στασιμότητας Augmented-Dickey Fuller Test. . . .	84
7.11	Διενέργεια στατιστικού ελέγχου ADF σε όλων των ειδών τα μετασχηματισμένα δεδομένα.	86
7.12	Χρονοσειρές παραγωγής ηλεκτρικής ενέργειας σε κανονική (πάνω διάγραμμα) και σε λογαριθμική (κάτω διάγραμμα) κλίμακα.	87
7.13	Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στην κανονική κλίμακα (πάνω διάγραμμα) και ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στην λογαριθμική κλίμακα (κάτω διάγραμμα).	87
7.14	Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στη λογαριθμική κλίμακα.	88
7.15	Από πάνω προς τα κάτω: Διάγραμμα απεικόνισης, Ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα αντίστροφης τετραγωνικής ρίζας.	88
7.16	Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα λογαριθμική αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.	89
7.17	Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα αντίστροφης τετραγωνικής ρίζας αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.	90
7.18	Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στη λογαριθμική κλίμακα αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών. . . .	90
7.19	Πρόβλεψη με χρήση του γραμμικού μοντέλου AR.	92
7.20	Πρόβλεψη με χρήση του γραμμικού μοντέλου MA.	93
7.21	Πρόβλεψη με χρήση του γραμμικού μοντέλου ARMA.	93
7.22	Πρόβλεψη με χρήση του γραμμικού μοντέλου ARIMA.	96
7.23	Πρόβλεψη με χρήση του γραμμικού μοντέλου SARIMA.	97
7.24	Απεικόνιση χρονοσειράς.	107
7.25	Απεικόνιση χρονοσειράς παραγωγής ενέργειας με ημερήσιο βήμα σε υποδιαστήματα του έτους 2020.	108

7.26	Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας (πάνω διάγραμμα). Kernel Density Estimator (μεσαίο διάγραμμα). Συνδυασμός των δύο από πάνω γραφημάτων και σύγκριση των δεδομένων με την κανονικής κατανομής (κάτω διάγραμμα).	108
7.27	Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας του Αιολικού πάρκου 2 στη Θράκη.	109
7.28	Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων.	111
7.29	Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη.	111
7.30	Διαγράμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη-ACF plot (40 lags).	113
7.31	Απεικόνιση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα Αιολικού πάρκου 2 στη Θράκη(πάνω διάγραμμα). Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα Αιολικού πάρκου 2 στη Θράκη (μεσαίο διάγραμμα). Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα Αιολικού πάρκου 2 στη Θράκη-ACF plot (κάτω διάγραμμα).	115
7.32	Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα του Αιολικού πάρκου 2 στη Θράκη.	115
7.33	Απεικόνιση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας Αιολικού πάρκου 2 στη Θράκη(πάνω διάγραμμα). Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας Αιολικού πάρκου 2 στη Θράκη (μεσαίο). Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας Αιολικού πάρκου 2 στη Θράκη-ACF plot (κάτω διάγραμμα).	117
7.34	Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας του Αιολικού πάρκου 2 στη Θράκη.	117
7.35	Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε λογαριθμική κλίμακα.	119
7.36	Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη σε λογαριθμική κλίμακα.	119
7.37	Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε κλίμακα τετραγωνικής ρίζας.	120
7.38	Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη σε κλίμακα τετραγωνικής ρίζας.	120

- 7.39 Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα λογαριθμική αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών. 121
- 7.40 Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών. 121
- 7.41 Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στη λογαριθμική κλίμακα αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών. 123
- 7.42 Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών. 123
- 7.43 Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε λογαριθμική κλίμακα. 124
- 7.44 Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε κλίμακα τετραγωνικής ρίζας. 124
- 7.45 Πρόβλεψη με χρήση του γραμμικού μοντέλου AR της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες. 128
- 7.46 Πρόβλεψη με χρήση του γραμμικού μοντέλου MA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες. 128
- 7.47 Πρόβλεψη με χρήση του γραμμικού μοντέλου ARMA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες. 129
- 7.48 Πρόβλεψη με χρήση του γραμμικού μοντέλου ARIMA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες. 131
- 7.49 Πρόβλεψη με χρήση του γραμμικού μοντέλου SARIMA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες. 132
- 7.50 Πρόβλεψη με χρήση του μοντέλου XGB-1 που αξιοποίησε ιστορικά δεδομένα ενός χρόνου για την εκπαίδευση του προκειμένου να προβλέψει τις ποσότητες ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη κάθε μία από τις επόμενες δέκα ημέρες. 140
- 7.51 Πρόβλεψη με χρήση του μοντέλου XGB-2 που αξιοποίησε χρήσιμα χαρακτηριστικά δεδομένα που εξήχθησαν από ιστορικά δεδομένα 2 χρόνων για την εκπαίδευση του χρήσιμες χαρακτηριστικές τιμές προκειμένου να προβλέψει τις ποσότητες ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη κάθε μία από τις επόμενες δέκα ημέρες. 143
- 7.52 Πρόβλεψη με χρήση του μοντέλου XGB-3 που αξιοποίησε χρήσιμα χαρακτηριστικά δεδομένα που εξήχθησαν από ιστορικά δεδομένα 2 χρόνων για την εκπαίδευση του χρήσιμες χαρακτηριστικές τιμές προκειμένου να προβλέψει τις ποσότητες ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη κάθε μία από τις επόμενες δέκα ημέρες. 146

Κατάλογος Πινάκων

2.1	Ετήσιες ποσότητες εκπομπών CO ₂ σε εκατομμύρια τόνους ανά είδος καυσίμου στην Ελλάδα [13].	16
2.2	Ετήσια παραγωγή ηλεκτρικής ενέργειας σε TWh ανά πηγή στην Ελλάδα [14].	16
3.1	Οι άνεμοι ανάλογα την διεύθυνση πνοής τους [29].	21
3.2	Η κλίμακα Μποφόρ [29].	21
3.3	Χωρικές κλίμακες των συστημάτων ανέμου και ένα δείγμα των σχετικών τύπων ανέμου [15].	23
4.1	Περιοχές λειτουργίας ανεμογεννήτριας [18].	40
5.1	Περιπτώσεις συσχέτισης	48
5.2	Μετασχηματισμοί δυναμονομικής (αλγεβρικής) μορφής Box-Cox [26].	53
6.1	Ενδείξεις στα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης οι οποίες βοηθούν στον προσδιορισμό των τάξεων των όρων p, q των γραμμικών μοντέλων.	61
7.1	Χαρακτηριστικά αποτελέσματα αξιολόγησης μοντέλων πρόβλεψης. Χρησιμοποιούνται τα κριτήρια αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος(RMSE), μέσου σφάλματος(ME), συντελεστή συσχέτισης(Corr), μέσου απόλυτου σφάλματος(MAE), ποσοστιαίου μέσου απόλυτου σφάλματος(MAPE), ποσοστιαίου μέσου σφάλματος (MPE) και συντελεστή προσδιορισμού(R^2)	99
7.2	Σύγκριση των προβλέψεων που παρήγαγε με βάση το μέτρο επιβεβαίωσης μέσου τετραγωνισμένου σφάλματοςMSE	103
7.3	Χαρακτηριστικά αποτελέσματα αξιολόγησης μοντέλων πρόβλεψης Αιολικού πάρκου στη Θράκη. Χρησιμοποιούνται τα κριτήρια αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος(RMSE), μέσου σφάλματος(ME), συντελεστή συσχέτισης(Corr), μέσου απόλυτου σφάλματος(MAE), ποσοστιαίου μέσου απόλυτου σφάλματος(MAPE), ποσοστιαίου μέσου σφάλματος (MPE) και συντελεστή προσδιορισμού(R^2)	134
7.4	Τιμές υπερπαραμέτρων των μοντέλων XGB-2 και XGB-3.	145

7.5	Τιμές κριτηρίων αξιολόγησης προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB. Χρησιμοποιούνται τα κριτήρια αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος(RMSE), μέσου σφάλματος(ME) και μέσου απόλυτου σφάλματος(MAE).	148
7.6	Τιμές κριτηρίου αξιολόγησης μέσου απόλυτου σφάλματος MAE των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.	149
7.7	Τιμές κριτηρίου αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος RMSE των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.	149
7.8	Τιμές κριτηρίου αξιολόγησης συντελεστή αυτοσυσχέτισης corr των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.	149
7.9	Τιμές κριτηρίου αξιολόγησης μέσου σφάλματος me των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.	150

Κεφάλαιο 1

Εισαγωγή

Μια κοινωνία δεν κάλυπτε πάντα ένα μέρος της ανάγκης της για ηλεκτρική ενέργεια με την αξιοποίηση ανανεώσιμων πηγών ενέργειας. Κατά το παρελθόν, ωφελούταν της θερμικής ενέργειας που εξέλυαν κατά την καύση τους διάφοροι ορυκτοί πόροι. Αυτή κατόπιν μετατρεπόταν σε ηλεκτρική. Η καύση τους, ωστόσο, έφερε ορισμένα μειονεκτήματα. Οι ποσότητες των επικίνδυνων ρύπων που διαχέονταν στο περιβάλλον κατά καύση τους συντελούσαν στην επιδείνωση της περιβαλλοντικής κατάστασης του πλανήτη. Παράλληλα εντεινόταν η εξάρτηση της ανθρώπινης βιωσιμότητας από αυτούς και τα αποθέματα τους μειώνονταν με ταχύτατους ρυθμούς. Έτσι λοιπόν, τέθηκε ως στόχος ότι η ενεργειακή ανεξάρτηση της κοινωνίας από τους ορυκτούς πόρους θα συντελούνταν σταδιακά με την περαιτέρω εκμετάλλευση των ανανεώσιμων πηγών ενέργειας.

Ανάμεσα στις ανανεώσιμες πηγές ενέργειας που αξιοποιούνται στην παραγωγή ηλεκτρικής ενέργειας, αυτή του ανέμου φαίνεται να ξεχωρίζει. Η κινητική του ενέργεια που μετατρέπεται σε ηλεκτρική, γίνεται να αξιοποιηθεί στη πλειοψηφία των περιοχών του πλανήτη σε αντίθεση με την ηλιακή και να καλύψει τη παγκόσμια ζήτηση. Το πρόβλημα που ανακύπτει όμως από την εκμετάλλευση της προέρχεται από την ίδια της την φύση. Αν και ο άνεμος είναι αέναος, είναι επίσης απρόβλεπτος. Δεν έχει μοντελοποιηθεί ακόμα το σύστημα το οποίο θα προβλέπει με ακρίβεια τις περιόδους πνοής ή νηνεμίας, την διάρκεια και την ακριβή ένταση ή κατεύθυνση του. Η αβεβαιότητα που υπάρχει ακόμα στην εκτίμηση του μεταφέρεται επίσης στο κομμάτι της παραγωγής ηλεκτρικής ενέργειας μέσω της αξιοποίησης του.

Η πραγματοποίηση εύστοχων προβλέψεων μέσω ποσοτήτων ηλεκτρικής ενέργειας που απορρέουν από την αξιοποίηση της αιολικής ενέργειας θα προσφέρουν τον βέλτιστο προγραμματισμό του συστήματος παραγωγής ενέργειας. Θα είναι δυνατό, δηλαδή, τότε να παρίσταται ανάγκη ή όχι λειτουργίας επιπρόσθετων μονάδων παραγωγής για την κάλυψη της ζήτησης. Αυτό θα είχε ως αποτέλεσμα την ελαχιστοποίηση των προβλημάτων στο δίκτυο ηλεκτροδότησης. Ακόμη οι εταιρίες διαχείρισης θα μεγιστοποιούν το κέρδος τους μέσω πιο εύστοχων υποβληθέντων προτάσεων παραγωγής στο Χρηματιστήριο Ενέργειας ενώ επίσης θα μπορούν να πάρουν την σωστή χρονική στιγμή εργασίες οικονομοτεχνικές αποφάσεις όπως τη τέλεση διεργασιών συντήρησης στο αιολικό πάρκο.

Στην παρούσα διπλωματική εργασία έγινε μια προσπάθεια εκτίμησης μελλοντικών μέσων

ποσοτήτων ενέργειας που παρήχθησαν από μια μονάδα παραγωγής και από ένα αιολικό πάρκο. Η προσέγγιση έγινε με τη βοήθεια μαθηματικών και με τη βοήθεια της μηχανικής μάθησης. Κάθε μια παρείχε διαφορετική μεθοδολογία διαχείρισης των δεδομένων και εκπαίδευσης των μοντέλων που θα παρήγαγαν την πρόβλεψη. Είχε μεγάλο ενδιαφέρον να παρατηρηθεί η απόδοση τους επί των ίδιων δεδομένων καθώς επίσης και η πολυπλοκότητα που έφερε η κάθε προσέγγιση.

Στην μαθηματική προσέγγιση εξετάστηκαν δύο διαφορετικά σύνολα παρατηρήσεων. Το μεν πρώτο προερχόταν από την εργασία [36]. Αφού είχε αξιοποιηθεί ήδη από έναν ερευνητή δεδομένων, είχε ενδιαφέρον να παρατηρηθεί εάν συνολικά οι μεθοδολογίες ανάλυσης χρονοσειρών, οι μεθοδολογίες μετατροπής μη στάσιμων χρονοσειρών και οι μεθοδολογίες επιλογής και εκπαίδευσης των βέλτιστων γραμμικών μοντέλων που αναπτύχθηκαν υπό το δικό μας πρίσμα θα απέφεραν καλύτερα αποτελέσματα από τα ήδη υπάρχοντα σε σχέση με το μέγεθος της απόκλισης και το χρόνο που απαιτήθηκε για να παραχθούν.

Το δεύτερο σύνολο δεδομένων αποτελούνταν από παρατηρήσεις δειγματοληπτημένες σε πραγματικό χρόνο σε επίπεδο υποσταθμού. Σε αυτόν καταγράφονταν οι μέσες ποσότητες ενέργειας που προέρχονταν από τη λειτουργία ενός αιολικού πάρκου ορισμένα χιλιόμετρα μακριά. Στη περίπτωση αυτή οι προβλέψεις που θα παράγονταν για το άμεσο μέλλον υλοποιήθηκαν επίσης με τη βοήθεια του αλγορίθμου μηχανικής μάθησης XGBoost. Η σύγκριση των δύο διαφορετικών μεθοδολογιών είχε πολύ μεγάλο ενδιαφέρον. Η κάθε μια εμφανίζει διαφορετικά πλεονεκτήματα ανάλογα τη φύση των δεδομένων που αξιοποιεί. Η σύγκριση των αποτελεσμάτων τους με μέτρα διασταυρωτικής επιβεβαίωσης αποσκοπούσε στην ανάδειξη των προτερήματα της κάθε μιας.

Με την ανάγνωση της εργασίας απαντιούνται πολλά ερωτήματα σε έναν άπειρο ερευνητή δεδομένων. Τόσο τα γραμμικά μοντέλα όσο και τα μοντέλα μηχανικής μάθησης δεν αξιοποιήθηκαν τυχαία. Αποτελούσε στόχο της εργασίας να αναδείξει σε ποια περίπτωση πρέπει να αξιοποιείται το κάθε είδος. Θεωρητικά οι πληροφορίες με βάση τις οποίες θα παίρνεται αυτή η απόφαση βρίσκονται στο στάδιο της διερευνητικής ανάλυσης. Αποτελεί ένα από τα πολλά στάδια της διαδικασίας πρόβλεψης. Για κάθε στάδιο γίνεται μια προσπάθεια εκτίμησης της σημασίας του για τη ποιότητα του τελικού αποτελέσματος. Συλλέγονται πολύτιμες πληροφορίες και χάρη σε αυτές είναι δυνατόν να παρθούν οι κατάλληλες αποφάσεις.

1.1 Οργάνωση Συγγράμματος

Το παρόν σύγγραμμα αποτελείται από επτά κεφάλαια. Καθένα από αυτά σκιαγραφούν τα στάδια που διένυσε η εργασία αυτή για να ολοκληρωθεί. Αρχικά παρατίθεται η θεωρία η οποία στη συνέχεια αξιοποιείται με την εφαρμογή της πάνω σε απτά παραδείγματα. Στη διάρκεια των πειραμάτων, τα αποτελέσματα που προκύπτουν καθώς ερμηνεύονται, είτε ανατροφοδοτούν την έρευνα με στόχο την επίτευξη καλύτερων αποτελεσμάτων είτε οδηγούν στην εξαγωγή πολύτιμων συμπερασμάτων.

Η μελέτη αυτή αρχικά αναφέρεται στον ρόλο της ενέργειας στην ανθρώπινη εξέλιξη. Στο πρώτο κεφάλαιο περιγράφονται οι ενεργειακές ανάγκες και οι τρόποι κάλυψής τους από τον

άνθρωπο ανά τα χρόνια εξέλιξης του. Ουσιαστικά, πραγματοποιείται μια ιστορική αναδρομή στις διάφορες πηγές ενέργειας που εκμεταλλεύτηκε ο άνθρωπος. Αποτυπώνονται τα πλεονεκτήματα αλλά και τα προβλήματα που δημιουργούσαν στη ζωή του. Τα προβλήματα αυτά έπαιξαν καθοριστικό ρόλο στην αναζήτηση τρόπων εκμετάλλευσης νέων πιο ‘καθαρών’ μορφών ενέργειας.

Στα επόμενα δύο κεφάλαια παρέχεται το απαραίτητο γνωστικό υπόβαθρο για την κατανόηση των χαρακτηριστικών του ανέμου και της ενέργειάς του. Σε πρώτη φάση, αναλύονται οι παράγοντες που οδηγούν στη εμφάνισή του και τα κύρια γνωρίσματά του. Σε δεύτερη φάση, λόγος γίνεται για την ενέργειά του και το πώς αυτή αξιοποιείται. Χρήσιμες είναι επίσης οι πληροφορίες που παρέχονται για τα μέσα εκμετάλλευσής της, τις ανεμογεννήτριες. Έπειτα γίνεται αναφορά τόσο στον τρόπο σύστασης και λειτουργίας μιας μονάδας παραγωγής που σχηματίζεται από ένα πλήθος ανεμογεννητριών όσο και στις ανάγκες που δημιουργούνται όταν ένα αιολικό πάρκο ενσωματώνεται στο δίκτυο ηλεκτρικής ενέργειας.

Το τέταρτο κεφάλαιο παρέχει το αναγκαίο θεωρητικό υπόβαθρο για τις χρονοσειρές. Αναλύονται τα στατιστικά χαρακτηριστικά τόσο των στάσιμων όσο και των μη στάσιμων. Τέλος, παραθέτονται μεθοδολογίες των οποίων η εφαρμογή αποφέρει την μετατροπή των μη στάσιμων σε στάσιμες. Στο κεφάλαιο που ακολουθεί, το πέμπτο, περιγράφονται τα διάφορα γραμμικά μοντέλα (MA, AR, ARMA, ARIMA, SARIMA) τα οποία μπορούν να αξιοποιηθούν στη πρόβλεψη χρονοσειρών. Ακόμη γίνεται αναφορά στη διαδικασία Box-Jenkins, η οποία έπαιξε κύριο ρόλο στην διαμόρφωση της μεθοδολογίας που παρουσιάζεται μετέπειτα στην εργασία αυτή και με την οποία μοντελοποιείται μια χρονοσειρά και παράγεται μια πρόβλεψή της. Τέλος, το κεφάλαιο παρέχει αναλυτικές πληροφορίες για τον αλγόριθμο XGBoost που επίσης αξιοποιείται στην παραγωγή προβλέψεων.

Στο έκτο κεφάλαιο παρουσιάζεται η στατιστική ανάλυση των χρονοσειρών που δημιουργήθηκαν από δεδομένα παραγωγής ηλεκτρικής ενέργειας. Σε αυτήν παρουσιάζονται αναλυτικά όλα τα στάδια της από την προεπεξεργασία των δεδομένων έως την τελική πρόβλεψη και αξιολόγηση αυτών. Στο έβδομο και τελευταίο κεφάλαιο γίνεται μια ανασκόπηση των τετελεσμένων στην εργασία αυτή και των συμπερασμάτων που εξήχθησαν από τα αποτελέσματα της πειραματικής διαδικασίας.

Κεφάλαιο 2

Η Ιστορία της Ενέργειας

2.1 Σύνοψη

Στο κεφάλαιο αυτό γίνεται μια προσπάθεια αφήγησης της ιστορίας της ενέργειας. Δίνεται η σημασία που κατείχαν στη ζωή του οι ενεργειακοί πόροι και εξιστορούνται οι τρόποι με τους οποίους κάλυπτε τις ανάγκες του σε ενέργεια ανά περίοδο. Παράλληλα γίνεται αναφορά επί των των οφελών και των αντίκτυπων, που έφεραν οι επιλογές αυτές στη ζωή του. Επιλογές που όχι μόνο άφησαν το στίγμα τους στην ιστορία, αλλά επίσης, προ οικονόμησαν το μέλλον του. Αρχικά δίνονται μερικές πληροφορίες για την πρώιμη περίοδο ανάπτυξης του ανθρώπινου γένους. Μια περίοδος αργής ανάπτυξης, κατά την οποία αξιοποιούνται για πρώτη φορά ανανεώσιμες πηγές ενέργειας, όπως ο άνεμος, για την κάλυψη των αναγκών του. Στη συνέχεια, μεταβαίνουμε σε μια εποχή πολύ σημαντική για την παγκόσμια ιστορία. Είναι το διάστημα κατά το οποίο στη ζωή του επέρχεται αξιοσημείωτη και γοργή ανάπτυξη χάρη στην αξιοποίηση των ορυκτών πόρων. Καταγράφονται τα αίτια, τα οφέλη και τα μειονεκτήματα της απόφασης αυτής για τον άνθρωπο και το περιβάλλον. Με αφορμή τα δεινά που φέρνει η αξιοποίηση των ορυκτών πόρων στη ζωή μας, γίνεται αναφορά στις προσπάθειες που καταβάλλει πλέον η παγκόσμια κοινότητα για να να αποφύγει επικείμενες δυσάρεστες καταστάσεις. Δίνεται, λοιπόν, έμφαση στους πράσινους στόχους που έχουν τεθεί και στους τρόπους με τους οποίους θα επιτευχθούν. Καθώς το μόνο βιώσιμο μέλλον, είναι ένα πράσινο μέλλον, η ανθρωπότητα στρέφεται στις ανανεώσιμες πηγές ενέργειας. Στην προσπάθεια τεκμηρίωσης αυτής της άποψης, δίνονται ορισμένα ενδιαφέροντα στατιστικά στοιχεία που τόσο καταδεικνύουν τόσο την εξάρτηση της Ελλάδας και του κόσμου από τα ορυκτά καύσιμα, όπως και της νέα πράσινης στρατηγικής που έχει χαραχθεί για τα επόμενα χρόνια. Σε αυτήν η αιχμή του δόρατος αποτελεί αξιοποίηση του ανέμου για την κάλυψη των ενεργειακών μας αναγκών σε επίπεδο ηλεκτρικής ενέργειας.

2.2 Αρχαίοι Χρόνοι και Μεσαίωνα

Πριν από περίπου διακόσιες χιλιάδες χρόνια εμφανίζονται τα πρώτα σημάδια ανθρώπινης ύπαρξης στον πλανήτη Γη. Η επιβίωση και η εξέλιξη του πρώτου αυτού ανθρώπου, που

μεταγενέστερα ονομάστηκε Homo Sapiens από τον σημερινό απόγονο του τον Homo Sapiens Sapiens, μελετήθηκε και χωρίστηκε σε διάφορες χρονολογικές περιόδους. Το είδος των ονομάτων, που έλαβαν, καταδεικνύουν ακριβώς τους λόγους για τους οποίους ο άνθρωπος άχμασε ανά εποχή. Κάποιες ονομασίες απορρέουν από τους νέους πόρους που ανακάλυπτε στη φύση και εργαλειοποιούσε (π.χ. Λίθινη Εποχή) ενώ άλλες αποδόθηκαν προς χάριν των πιο σημαντικών τεχνολογικών επιτευγμάτων του (π.χ η Εποχή του Χαλκού ήταν η περίοδος που για πρώτη φορά επεξεργάστηκε τον χαλκό και δημιούργησε χρήσιμα αντικείμενα για την καθημερινότητα του). Συναντούνται σε αυτές τα πρώτα βήματα του ανθρώπινου γένους, που ακόμη και σήμερα προκαλούν το θαυμασμό μας. Σε κάθε ένα επιδεικνύει νέες συμπεριφορές και δημιουργεί νέες ανάγκες που η κάλυψη τους θα το οδηγήσουν σε ένα πιο ευοίωνο μέλλον. Οι διεργασίες αυτές, λοιπόν, συντέλεσαν στη διαμόρφωση μας ανατομικά, πνευματικά και κοινωνικά.

Δεν είναι τυχαίο που η εξελικτική μας πορεία είναι στενά συνδεδεμένη με την ανεύρεση-χρήση νέων πόρων ή τεχνολογιών και την διαχείριση μορφών ενέργειας. Στην αρχή απλές εστίες φωτιάς προσέφεραν θέρμανση και προστασία [24]. Με την εγκατάσταση του σε πιο οργανωμένες κοινωνίες, ο άνθρωπος θα αξιοποιήσει την ηλιακή ενέργεια για την παραγωγή τροφής (αγροτικές καλλιέργειες) και συνεπώς την κάλυψη των επισιτιστικών του αναγκών. Παράλληλα επωφελείται της αιολικής ενέργειας (ιστιοφόρα πλοία) και ανακαλύπτει νέους τόπους, διευκολύνει τις μεταφορές του και αναπτύσσει το εμπόριο. Έρχεται σε επαφή με άλλους πολιτισμούς, αναπτύσσει σχέσεις και ανταλλάσσει ιδέες. Ως όν εφευρετικό και πρωτοπόρο θα ανακαλύψει την αξία της υδραυλικής ενέργειας και θα υδροδοτήσει τους οικισμούς του με χρήση μικρών νερόμυλων. Λίγο πιο μετά θα εφευρεθεί η πρώτη ατμομηχανή. Η ισχύς που απέδιδε η εκμετάλλευση της θερμικής ενέργειας, θα αντικαταστήσει την ισχύ που απέδιδαν πολλά εργατικά χέρια αρκετούς αιώνες αργότερα. Η ανθρωπότητα στο πέρασμα των αιώνων με τη βοήθεια αυτών των μορφών ενέργειας διασφάλισε την επιβίωση της, ένωσε ασφάλεια για το μέλλον της και ανέπτυξε το πνεύμα. Άχμασε ουσιαστικά και οργάνωσε τη κοινωνία στην οποία ζούσε. Έτσι λοιπόν, μέχρι τον 15ο αιώνα, τέλη δηλαδή του ύστερου Μεσαίωνα σχεδόν όλες οι κοινωνίες κάλυπταν επί των πλείστων τις ανάγκες τους και ολοκλήρωναν τις εργασίες τους με την εκμετάλλευση της ανθρώπινης δύναμης και χάρη στην αξιοποίησης της ενέργειας που προσέφεραν ανανεώσιμες πηγές ενέργειας όπως ο ήλιος ή ο άνεμος.

2.3 Μετάβαση στη Νεότερη Εποχή με την Αξιοποίηση Ορυκτών Καυσίμων

Τα χρόνια από τον 15ο αιώνα και ύστερα χαρακτηρίστηκαν από την πρόθεση της ανθρωπότητας να επεκταθεί. Οι χώρες της Ευρώπης άφηναν πίσω αιώνες σκοταδισμού με ταξίδια ανακάλυψης νέων κόσμων ή με την εφεύρεση της τυπογραφίας και τη τύπωση εκατομμυρίων βιβλίων. Πλέον στα βιβλία και στην εκπαίδευση είχαν πρόσβαση άτομα που δεν άνηκαν μήτε στην εκκλησιαστική μήτε στην αστική τάξη. Οι πληροφορίες, οι γνώσεις και οι ιδέες διαδίδονταν πιο εύκολα και πιο γρήγορα σε όλη την Ευρώπη. Οι επιστήμες προάγονται, οι νέες εφευρέσεις αντιμετωπίζονται με ενθουσιασμό από μια κοινωνία που αναπτύσσεται παράλληλα με το πνεύμα.

Μειώνονται οι εμφύλιες και διακρατικές συγκρούσεις και χώρες όπως η Ισπανία και η Αγγλία αναπτύσσουν ιμπεριαλιστικές πολιτικές εκτός Ευρώπης. Με την κατάκτηση νέων εδαφών ή τον οικονομικό έλεγχο άλλων λαών, καθίσταται δυνατή η εξασφάλιση μεγάλων ποσοτήτων νέων ή ήδη αξιοποιήσιμων πόρων και η εγκαθίδρυση νέων κοινωνιών οι οποίες είχαν υποστηρικτικό ρόλο προς τις αναπτυσσόμενες μητροπόλεις τους. Σε αυτές παρατηρούνται έντονες πληθυσμιακές μεταβολές. Κατά τον Μεσαίωνα ορισμένα γεγονότα (λιμοί, πανώλη, διωγμοί, πόλεμοι), είχαν συμβάλει στη δραστική μείωση του πληθυσμού με τα μεγάλα αστικά κέντρα να πληρώνουν το μεγαλύτερο τίμημα. Τώρα πια τα αποδεκατισμένα αστικά κέντρα αναβιώνουν σιγά σιγά χάριν στη ραγδαία αύξηση του πληθυσμού και στις μεγάλες πληθυσμιακές ομάδες που εγκαταλείπουν την ύπαιθρο και μετοικούν προς αυτά. Ήταν λοιπόν ενέργειες που είχαν ως στόχο την εξέλιξη και εδραίωση της ανθρωπότητας στον πλανήτη.

Ο συγκεκριμένος βιολογικός στόχος επιτυγχανόταν μέσω των προαναφερθέντων δραστηριοτήτων, των οποίων η υλοποίηση απαιτούσε τεράστιες ποσότητες ενέργειας και ύλης. Εκείνη την εποχή οι πόροι που μπορούσαν να αξιοποιηθούν ποσοτικά και ποιοτικά όπως επίσης οι διαθέσιμες εφαρμόσιμες τεχνολογίες έκαναν να φαντάζει αδύνατον το συγκεκριμένο εγχείρημα. Έως τότε το ξύλο ήταν το κύριο υλικό που αξιοποιούσε για να καλύψει κάποιες από τις πρωταρχικές του ανάγκες ο άνθρωπος. Πιο συγκεκριμένα, μέσω της καύσης του μπορούσε να θερμάνει την οικία του, να μαγειρέψει και να επεξεργαστεί μεταλλεύματα όπως σίδηρο και ατσάλι. Επίσης το ξύλο ήταν το κύριο δομικό υλικό στις περισσότερες ανθρώπινες κατασκευές και υποδομές και απαραίτητο για την τύπωση βιβλίων. Οι όλο και μεγαλύτερες ποσότητες ξυλείας που έπρεπε να καταναλώνονται για την διατήρηση της ευημερίας και της προόδου των κοινωνιών, οδήγησαν στην αποψίλωση τεράστιων δασικών περιοχών σε περιοχές της Βόρειας Ευρώπης και της Βρετανίας. Γύρω στα τέλη του 17ου αιώνα και ενώ η ζήτηση ξυλείας συνεχίζει να παρουσιάζει αυξητική τάση, τα αποθέματα ξυλείας που διατίθενται προς αξιοποίηση στην αγορά δεν επαρκούν. Πρώτη φορά η ανθρωπότητα αντιμετώπιζε μια ενεργειακή κρίση.

Οι διαθέσιμες ποσότητες ξυλείας σαν ήταν ελλείψεις, έφεραν στο προσκήνιο τον ορυκτό άνθρακα. Ήταν ένα υλικό το οποίο χρησιμοποιούσαν οι υπήκοοι της Βρετανίας στη μεταλλουργία και για τη θέρμανση τους ήδη από το 300 π.Χ. περίπου. Από τότε η εκμετάλλευση του γινόταν σε μικρό βαθμό καθώς η καύση του εμφάνιζε προβλήματα. Ήταν τόσο σοβαρές οι επιπτώσεις που ανέκυπταν στην καθημερινή ζωή από τη χρήση του που ο βασιλιάς της Αγγλίας τον 14ο αιώνα απαγόρευσε την καύση του. Ωστόσο με την ανάπτυξη νέων μεθόδων αξιοποίησης του στην παραγωγή και μετριάζοντας τις αρνητικές συνέπειες που είχε η καύση του, θα αντικαταστήσει σιγά σιγά το ξύλο. Αρχικά η Αγγλία που είχε εύκολη και σε μεγάλες ποσότητες πρόσβαση στον πόρο αυτόν και στη συνέχεια όλος ο κόσμος υιοθέτησαν την πρακτική αυτή και αλλάζουν δυναμική. Έως το 1600 ο άνθρακας εμπορεύεται σχεδόν σε όλο το κόσμο και το υφιστάμενο ενεργειακό πρόβλημα βρίσκει λύση στο όνομα του.

2.3.1 Οφέλη και Πλεονεκτήματα

Με την πάροδο των χρόνων οι αναπτυσσόμενες κοινωνίες της Δύσης οικειοποιούνται τη χρήση και τα οφέλη των ορυκτών καυσίμων. Η θερμότητα που αποδίδει ένα ορυκτό όπως ο

άνθρακας κατά την καύση του καλύπτει ανάγκες που έχει ο άνθρωπος στη θέρμανση κτιρίων, σε διάφορες μεταλλουργικές δραστηριότητες και στην τροφοδότηση των μηχανών για την εύρυθμη λειτουργία τους. Μέχρι τα τέλη του 18ου αιώνα η σχέση-εξάρτηση ανάμεσα στον άνθρωπο και τους ορυκτούς πόρους θα λάβει άλλη διάσταση λόγω μιας σειράς γεγονότων που έλαβαν χώρα. Πιο συγκεκριμένα ο James Watt θα βελτιώσει σημαντικά την υπάρχουσα ατμομηχανή, δίνοντας τη δυνατότητα χρήσης αυτής για την κίνηση μηχανών. Η βιομηχανική επανάσταση η οποία ήδη έχει ξεκινήσει, φουντώνει χάριν στις νέες ατμομηχανές. Η ανέλιξη που θα έφερναν στο βιοτικό επίπεδο του κόσμου ήταν αδιανόητη. Χαρακτηριστικό παράδειγμα αποτελούν οι πρώτες ατμομηχανές των οποίων η ισχύ από καύση άνθρακα ισοδυναμούσε με την ισχύ διακοσίων ανδρών. Βέβαια πιο μεγάλες, πιο ισχυρές και πιο αποδοτικές μηχανές θα ενταχθούν στις μεταφορές, στη γεωργία, στις κατασκευές, στις γραμμές παραγωγής εργοστασίων και μικρών βιομηχανιών. Εισχωρούν και επηρεάζουν κάθε πτυχή της ζωής του ανθρώπου. Ο πληθυσμός πολλαπλασιάζεται, νέες θέσεις εργασίας δημιουργούνται, οι υποδομές και τα μεταφορικά του μέσα αναβαθμίζονται, νέες βλέψεις-όνειρα γεννούνται και το βιοτικό του επίπεδο αλλάζει συνεχώς προς το καλύτερο.

Η ζήτηση για ορυκτά θα εκτιναχθεί. Το γεγονός ότι δίνουν ώθηση στην κοινωνία και ζωή στις μηχανές, τα καθιστούν ζωτικής σημασίας για την εξέλιξη της. Χάρη σε αυτά κάποιες κοινωνίες έρχονται πιο κοντά, συνεργάζονται, προ οδεύουν ενώ άλλες εξαιτίας των αποθεμάτων που διαθέτουν και εμπορεύονται, αποκτούν κύρος που δεν κατείχαν ποτέ πιο πριν στον παγκόσμιο χάρτη. Η αξιοποίηση τους φέρει ανάπτυξη που με τη σειρά της φέρει ευχάριστες αλλαγές στη ζωή του ανθρώπου. Βέβαια οι πιο μεγάλες συναντώνται στην καθημερινότητα του. Στο Λονδίνο για παράδειγμα ήδη από τα πρώτα χρόνια του 19ου αιώνα θα σχεδιαστεί και θα εγκατασταθεί ένα δίκτυο που παρείχε αέριο στα νοικοκυριά. Αυτό προσέφερε αξιόπιστα φως, θέρμανση και ενέργεια. Ένα ακόμη δείγμα γραφής των πλεονεκτημάτων που προσέφερε στη ζωή της ανθρωπότητας η εύρεση και αξιοποίηση νέων καυσίμων αποτελούν οι πρώτες εξορύξεις πετρελαίου που πραγματοποιούνται στα μέσα του 19ου αιώνα. Μεσώ αυτών εξορύχτηκε το ορυκτό το οποίο μπήκε όσο κανένα άλλο στη ζωή του ανθρώπου μέσω των αποσταγμάτων του. Κάθε διύλιση του φέρει κάποια εφαρμογή στη ζωή του. Η βενζίνη κινεί τα οχήματα του, η καύση του προσφέρει ηλεκτρική ενέργεια ή θέρμανση ενώ άλλες μορφές του αποτελούν το κύριο υλικό για πολλά προϊόντα που συναντάμε πλέον στην καθημερινότητα μας όπως είναι οι διαλύτες, τα λιπάσματα και τα απορρυπαντικά. Νέα και πιο αποδοτικά ορυκτά καύσιμα προστίθενται συνέχεια στην διάθεση του ανθρώπου. Είναι τέτοια η συνεισφορά και η δυναμική τους που παράγοντες όπως ο καιρός και οι αποστάσεις δεν επηρεάζουν πλέον σε μεγάλο βαθμό δραστηριότητες του, όπως οι μεταφορές. Μέχρι τα μέσα του 20ου αιώνα το 90 % των ενεργειακών αναγκών μας καλυπτόταν με την καύση ορυκτών πόρων κυρίως άνθρακα. Έως σήμερα τα ορυκτά καύσιμα κατέχουν σχεδόν σε όλες χώρες την πρωτοκαθεδρία στην παραγωγή ενέργειας. Ωστόσο τις τελευταίες δεκαετίες γίνεται μια προσπάθεια στροφής της ανθρωπότητας προς πιο φιλικές μορφές ενέργειας. Υδροηλεκτρικά εργοστάσια, πυρηνικά εργοστάσια, αιολικά πάρκα, μετατροπείς κυματικής ενέργειας και ηλιακοί συλλέκτες εγκαθίστανται και μειώνουν το ποσοστό εξάρτησης του ανθρώπου από τους ορυκτούς πόρους του πλανήτη.

2.3.2 Αρνητικές Συνέπειες για την Φύση και τον Άνθρωπο

Τα τελευταία 200 χρόνια ανέδειξαν την μακροχρόνια συσχέτιση που υπάρχει ανάμεσα στις αναπτυσσόμενες οικονομίες και την αξιοποίηση ορυκτών καυσίμων σε ευρεία κλίμακα από αυτές [25, 8]. Δεν υπάρχει ακμάζουσα οικονομία στον πλανήτη που να μην εξαρτάται από την κατανάλωση τους. Η κατανάλωση τους γίνεται με κύριο σκοπό την τροφοδότηση της βιομηχανίας της οποίας τα παράγωγα (π.χ. μεταφορές, παραγωγή ηλεκτρισμού, παραγωγή προϊόντων, θέρμανση, λειτουργία εργοστασίων) καλύπτουν τις ανάγκες μας και διευκολύνουν τη ζωή μας. Αποτελούν την κινητήρια δύναμη των οικονομιών παγκοσμίως με αποτέλεσμα η διασφάλιση των πόρων αυτών να είναι καίριας σημασίας για αυτές.

Η απορρόφηση των πόρων αυτών από την κοινωνία και η μετουσίωση τους σε ανάπτυξη, αποτελεί μια μακρά διαδικασία. Το πρώτο βήμα στην προσπάθεια αυτή είναι η αναζήτηση τους. Στη συνέχεια πραγματοποιούνται εξορύξεις στα σημεία όπου εντοπίστηκαν ικανοποιητικά κοιτάσματα. Από εκεί οι εκμεταλλεύσιμες ποσότητες που έχουν εξαχθεί, μεταφέρονται σε ειδικές εγκαταστάσεις. Υφίστανται διεργασίες που μεταβάλλουν τη σύνθεση τους έως ότου καταστούν αξιοποιήσιμα για βιομηχανική χρήση. Τέλος μεταφέρονται και αποθηκεύονται σε ασφαλείς δομές από όπου αντλούνται και διατίθενται στις αγορές. Απ' την αρχή ως το τέλος σαν διαδικασία λογίζεται επικίνδυνη τόσο για τον άνθρωπο όσο και για το περιβάλλον.

Γεωπολιτικές Κρίσεις

Η ανθρωπότητα πλέον έχει εξερευνήσει σχεδόν κάθε γωνιά του πλανήτη της. Δεν πραγματοποιεί ευρείας κλίμακας διερευνητικές αποστολές σε αναζήτηση ορυκτών καυσίμων. Οι περισσότερες χώρες έχουν ερευνήσει τα εδάφη τους και έχουν εντοπίσει εάν διαθέτουν ή όχι ορυκτό πλούτο. Για κάποιες τα μεγάλα κοιτάσματα τους αποτελούν το πιο σημαντικό περιουσιακό τους στοιχείο και το εμπορεύονται σε όσους το έχουν ανάγκη. Με τη μείωση των αποθεμάτων και την αύξηση της ζήτησης, η σημασία τους για την παγκόσμια οικονομία αποκτά μεγαλύτερη αξία. Χώρες όμως με πολύ μεγαλύτερη δυναμική συχνά εποφθαλμιούν τους πόρους αυτούς καθώς δεν διαθέτουν τις ποσότητες που επιθυμούν και προσπαθούν να τις ελέγξουν. Πολλές φορές είναι υπεύθυνες των εσωτερικών κρίσεων και των οικονομικών δυσκολιών που ξεσπούν σε χώρες σαν αυτές. Χώρες αποδυναμώνονται ή αποσταθεροποιούνται πολιτικά και οδηγούνται σε πολεμικές συρράξεις. Αθώοι χάνουν τις περιουσίες τους, μεταναστεύουν σε άλλες χώρες ή χάνουν τις ζωές τους. Χώρες που έτυχαν να είναι πλούσιες σε πόρους αλλά όχι σε τεχνογνωσία και μέσα για να τους εκμεταλλευτούν κατά το δοκούν, καταλήγουν έρμαιο της τύχης τους με πολύ άσχημες συνέπειες για τον ίδιο τους τον λαό και την εδαφική τους ακεραιότητα. Βέβαια δεν είναι οι μόνες που επηρεάζονται. Πολλές οικονομίες ανά τον κόσμο διαμορφώνουν πολιτικές ανάλογα τον πόρο που τις κινεί, οπότε και η συνεχή τροφοδότηση τους με αυτόν αποτελεί ζήτημα επιβίωσης. Σε περίπτωση που για λόγους που προαναφέρθηκαν λιγοστεύουν οι πηγές που παρέχουν τον πόρο αυτών ή ακόμη και εκλείψουν η προκείμενη οικονομία υφίσταται γερούς κραδασμούς τους οποίους συχνά δυσκολεύεται να απορροφήσει.

Διατάραξη-Μόλυνση Οικοσυστημάτων

Όλα τα ορυκτά καύσιμα για να περιέλθουν στην κατοχή του ανθρώπου, εξάγονται από την Γή. Οι εξορύξεις που γίνονται στη ξηρά ή οι γεωτρήσεις που πραγματοποιούνται στη θάλασσα, σε βάθος μικρό ή μεγάλο κάτω από την επιφάνεια της γης, είναι διεργασίες που φέρουν σοβαρές επιπτώσεις στο περιβάλλον. Περιοχές φυσικού κάλους όπως δασικές ή θαλάσσιες, μετατρέπονται σε βιομηχανικές καθώς πολύ συχνά αποτελούν τα σημεία εξορύξης. Οι εξορυκτικές δραστηριότητες που αναπτύσσονται είναι τόσο έντονες που διαταράσσουν το ανάγλυφο, το κλίμα, την πανίδα και τη χλωρίδα της περιοχής. Αρχικά εταιρίες αναλαμβάνουν την εκσκαφή, την ανατίναξη και τη διάλυση τεραστίων εκτάσεων γης ώστε να προετοιμάσουν την εξαγωγή των πόρων. Η εδαφική αλλοίωση που υφίσταται μια τέτοια περιοχή συνήθως δεν αποκαθίσταται καθώς είτε υποβαθμίζεται άμεσα η γονιμότητα του εδάφους είτε κατακερματίζονται τεράστιες εκτάσεις και εναποτίθενται σε άλλες περιοχές ώστε να αναπτυχθούν δομές εξορύξης, επεξεργασίας, αποθήκευσης και μεταφοράς. Η άγρια πανίδα με τη σειρά της δεν έχει πλέον στη διάθεση της τα κομμάτια γης που στο παρελθόν χρησιμοποιούσε για να ζει, να αναπαράγεται και να μεταναστεύει, ενώ οι μεγάλες εκτάσεις πρασίνου που αποφιλώνονται ζημιώνουν την χλωρίδα της περιοχής και υποβαθμίζουν την ατμοσφαιρική ποιότητα της. Μεγάλα προβλήματα στο περιβάλλον δημιουργούν επίσης τα απόβλητα που παράγουν οι εξορύξεις ορυκτών καυσίμων ή τα εργοστάσια που τα καταναλώνουν. Ιχνοστοιχεία όπως το κάδμιο, το κοβάλτιο, ο χαλκός κ.α εκ πλένονται από τα απόβλητα των ορυχείων και συγκεντρώνονται στη βλάστηση, στο νερό και το έδαφος. Κάθε χρόνο εκατομμύρια τόνοι ζημιογόνους ύλης που δεν ανακυκλώνεται, παράγεται και διοχετεύεται σε υγειονομικούς τόπους ταφής ή σε χώρους συγκράτησης. Αυτές οι ποσότητες καθώς διαρρέονται από νερό (π.χ βροχής), απελευθερώνουν τοξικά στοιχεία. Κατά αυτόν τον τρόπο μολύνεται το νερό και μετατρέπεται σε όξινο. Σε μια τέτοια περίπτωση τίθεται σε άμεσο κίνδυνο ο υδροφόρος ορίζοντας της περιοχής. Έτσι λοιπόν διαταράσσονται και μολύνονται τα οικοσυστήματα γύρω από το σημείο εξορύξης.

Κίνδυνοι για την Ανθρώπινη Ζωή

Δεινές καταστάσεις αντιμετωπίζει επίσης ο άνθρωπος στη διάρκεια δραστηριοτήτων εξορύξης ή γεώτρησης. Τους προηγούμενους αιώνες όπου τα μέτρα προστασίας των εργατών ήταν ελλιπή, πολλά ατυχήματα είχαν σημειωθεί. Στις μέρες μας αν και έχουν υπάρξει βελτιώσεις τόσο στα μηχανήματα όσο και στα μέτρα που εφαρμόζονται για την προστασία τους, θανατηφόρα ατυχήματα δεν παύουν να εκτυλίσσονται. Από ανατινάξεις, εγκλωβισμούς, εισπνοή δηλητηριωδών αερίων, καταπλακώσεις και καταρρεύσεις έχουν βρει τραγική κατάληξη μέλη του εργατικού δυναμικού εγκαταστάσεων εξορύξης. Άλλοι εργαζόμενοι ταλαιπωρούνται από χρόνια προβλήματα υγείας λόγω των συνθηκών στις οποίες δουλεύουν. Η σκόνη που υπάρχει σε εγκαταστάσεις εξορύξης προκαλεί προβλήματα πνευμονοπάθειας και όρασης ενώ ορισμένα τοξικά ή ραδιενεργά υλικά θέτουν σε υψηλό κίνδυνο τη ζωή αυτών καθημερινά. Στις εγκαταστάσεις γεωτρήσεις καυσίμων όπως το πετρέλαιο ή το φυσικό αέριο, ρύποι εκπέμπονται πολύ πριν καούν τα συγκεκριμένα ορυκτά, είτε σε περίπτωση πυρκαγιάς είτε όταν καταναλώνονται. Η έκθεση σε ρύπους όπως το βενζόλιο έχουν θεωρηθεί υπεύθυνα για την πρόκληση αιμα-

τολογικών διαταραχών και την εμφάνιση λευχαιμίας και κάποιων άλλων ειδών καρκίνου. Η αποστροφή της ανθρωπότητας από πηγές ενέργειας σαν τα ορυκτά καύσιμα θα βελτιώσει τη ζωή πολλών ανθρώπων και ειδικά όσων εργάζονται υπό τόσο δύσκολες συνθήκες.

Περιβαλλοντική Υποβάθμιση - Περιβαλλοντικά Ατυχήματα

Κίνδυνοι συνεχίζουν να υπάρχουν και εκτός των εγκαταστάσεων όπου συλλέγονται. Οι πόροι αυτοί, πριν χρησιμοποιηθούν στην παραγωγή, μεταφέρονται σε ειδικές εγκαταστάσεις. Εκεί, επεξεργάζονται και λαμβάνουν την τελική ωφέλιμη μορφή τους. Η μεταφορά τους απαιτεί πολλά και μεγάλα μέσα (βυτιοφόρα οχήματα, πλοία) που ρυπαίνουν το περιβάλλον γύρω τους καθώς μετακινούνται και διανύουν μεγάλες αποστάσεις. Κατά την μεταφορά τους, υπάρχει κίνδυνος ατυχημάτων όπως διαρροής καυσίμων πτητικών και εύφλεκτων. Η οικολογική καταστροφή που ακολουθεί είναι πολλές φορές τεράστια. Μεγάλες περιοχές και οικοσυστήματα καταστρέφονται. Μολύνονται σε σημείο που κανένας έμβιος οργανισμός δεν μπορεί να επιβιώσει ή να αναπτυχθεί ενώ το διάστημα αποκατάστασης τους υπολογίζεται από χρόνια έως και σε δεκαετίες (π.χ πετρελαιοκηλίδα στον κόλπο του Μεξικού). Παρόμοια καταστροφή παρατηρείται στις περιοχές προ επεξεργασίας καυσίμων (διυλιστήρια) όπου παρατηρείται έντονη μόλυνση του αέρα και των υδάτων γύρω από αυτό. Η πρόκληση ενός ατυχήματος όπως το ξέσπασμα μιας πυρκαγιάς θέτει σε κίνδυνο τη ζωή ανθρώπων, προκαλεί τεράστιες ζημιές και εκλύονται μεγάλες ποσότητες θανατηφόρων αερίων. Πιθανότατα κανένας από αυτούς τους κινδύνους δεν θα υπήρχε εάν εξαρτούμασταν από πράσινες πηγές ενέργειας.

Κλιματική Αλλαγή

Οι πιο ζημιογόνες όμως επιπτώσεις της χρήσης ορυκτών καυσίμων παρατηρούνται κατά την καύση τους, οπότε και απελευθερώνονται στην ατμόσφαιρα οι ζημιογόνοι ρύποι τους [9]. Εκλύονται επικίνδυνα αέρια σε μεγάλες ποσότητες, από τα οποία το κύριο που απειλεί το κλίμα μας, είναι το διοξείδιο του άνθρακα CO_2 . Οι ποσότητες CO_2 που παράγονται από τον άνθρωπο (μεταφορές, εργοστάσια, βιομηχανία) προστίθενται στις ποσότητες που δημιουργούνται με φυσικό τρόπο με αποτέλεσμα το φαινόμενο του θερμοκηπίου να αποκτά αρνητικές διαστάσεις για τον πλανήτη. Υπό κανονικές συνθήκες το φαινόμενο του θερμοκηπίου είναι υπεύθυνο για την επιβίωση μας. Ουσιαστικά από την συνολική ηλιακή ακτινοβολία που δέχεται η Γη ένα μεγάλο μέρος της απορροφάται από την Γη και κάποιο από τα ατμοσφαιρικά αέρια, σωματίδια και νέφη. Το υπόλοιπο ανακλάται από τα σύννεφα, την Γη και τα ατμοσφαιρικά σωματίδια της, με αποτέλεσμα να διαφεύγει στο διάστημα. Τα αέρια αυτά θερμαίνονται και εκπέμπουν υπέρυθη ακτινοβολία της οποίας το μεγαλύτερο ποσοστό επιστρέφει πίσω προς την επιφάνεια του πλανήτη τον οποία τελικά θερμαίνει. Με την αύξηση λοιπόν των ποσοτήτων των αερίων αυτών στην ατμόσφαιρα μας, εντείνεται το φαινόμενο αυτό και θερμαίνεται περισσότερο ο πλανήτης μας.

Η υπερθέρμανση εντείνει τη δύναμη και τη συχνότητα των καιρικών φαινομένων με αποτέλεσμα να βιώνουμε πιο ζεστούς καύσωνες, βαρύτερες βροχοπτώσεις και πιο μακρές ξηρασίες. Καιρικές μεταβολές και συνθήκες που θέτουν σε κίνδυνο ολόκληρα οικοσυστήματα

στον πλανήτη. Η άνοδος της θερμοκρασίας του πλανήτη μεταφράζεται επίσης σε άνοδο της θερμοκρασίας της θάλασσας. Η συγκεκριμένη συνθήκη, επηρεάζει την ένταση των τροπικών καταιγίδων και την βιωσιμότητα των θαλάσσιων οργανισμών. Δημιουργούνται συχνότερα πιο δυνατές τροπικές καταιγίδες, που προκαλούν ζημιές δισεκατομμυρίων κάθε χρόνο και περισσότερους θανάτους. Επίσης η αλλαγή της θερμοκρασίας των νερών, οδηγεί το πλαγκτόν στους πόλους της Γης. Το πλαγκτόν, το οποίο αποτελείται από μικροσκοπικούς φυτικούς και ζωικούς οργανισμούς, επιπλέει ελεύθερο στους ωκεανούς. Βρίσκεται στη βάση της τροφικής αλυσίδας και αποτελεί παράγοντα κλειδί στην απορρόφηση CO₂ και παραγωγή οξυγόνου αφού μπορεί να κάνει φωτοσύνθεση. Η μετεγκατάσταση του θα φέρει δραματικές αλλαγές στα οικοσυστήματα τα οποία επιβίωναν χάρη σε αυτό.

Η μεταβολή της θερμοκρασίας του πλανήτη μας οδηγεί αναπόφευκτα σε σταδιακό λιώσιμο των πάγων και μείωση του όγκου των παγετώνων της. Για παράδειγμα από το 1990 η Ανταρκτική έχει χάσει σχεδόν τέσσερα τρισεκατομμύρια τόνους πάγου. Οι όγκοι νερού που απελευθερώνονται έχουν ως αποτέλεσμα την άνοδο της στάθμης των θαλασσών. Ήδη έχουν βυθιστεί κατοικήσιμες εκτάσεις ενώ στο μέλλον προβλέπεται πως χώρες (μικρά νησιωτικά συμπλέγματα) θα εξαφανιστούν ή τμήματά τους θα καταποντιστούν. Οι πληθυσμοί που θα πληγούν, θα αναγκαστούν να ξεριζωθούν και να μεταναστεύσουν ενώ εκτάσεις χρήσιμες για τον άνθρωπο σε τομείς όπως η καλλιέργεια, η κτηνοτροφία και η δόμηση θα χαθούν. Σε έναν πλανήτη όπου ο πληθυσμός αυξάνεται συνεχώς, η διατήρηση των γόνιμων και κατοικήσιμων εδαφών αποτελεί προτεραιότητα για την ανθρωπότητα.

Τα αναπνευστικά προβλήματα θα είναι πιο έντονα και θα εμφανίζονται σε περισσότερα άτομα όσο δεν μειώνεται η μόλυνση της ατμόσφαιρας από τους ρύπους που εκπέμπουν τα ορυκτά καύσιμα κατά την καύση τους. Πέρα από το διοξείδιο του άνθρακα, αέρια όπως τα νιτρικά και θειικά οξέα που είναι υπεύθυνα για τον σχηματισμό της όξινης βροχής. Το συγκεκριμένο είδος βροχής πέρα από το στοιχείο της μόλυνσης και καταστροφής ολόκληρων καλλιεργήσιμων εκτάσεων, δασικών εκτάσεων και πηγών καταναλώσιμου νερού, φέρει και το στοιχείο της αλλοίωσης και διάβρωσης διαφόρων πετρωμάτων όπως το μάρμαρο ή ο ασβεστόλιθος. Αποτελεί πληγή για πολιτισμικά μνημεία με ιστορία χιλιάδων χρόνων όπως και για την περιουσία πολιτών.

Τα οξείδια του αζώτου συναντώνται σε μεγάλες ποσότητες στην ατμόσφαιρα μεγάλων πόλεων. Απελευθερώνονται κατά την καύση καυσίμων που κινούν τα οχήματα με τα οποία κινείται μέσα σε αυτήν ο άνθρωπος. Στην περίπτωση που η συγκέντρωσή τους παραμένει υψηλή για μεγάλα χρονικά διαστήματα, επιδεινώνεται αθόρυβα η υγεία χιλιάδων ατόμων που τα εισπνέουν. Σοβαρές ασθένειες όπως άσθμα, καρκίνος των πνευμόνων και καρδιαγγειακά προβλήματα διαγιγνώσκονται στους πληθυσμούς των πόλεων των οποίων η ατμόσφαιρα έχει ρυπανθεί από οξείδια του αζώτου. Στην Ευρώπη μόνο υπολογίζονται πώς περίπου τετρακόσιες χιλιάδες ανθρώπινες ζωές χάνονται κάθε χρόνο ελέω των ριπών που εκλύει η κατανάλωση ορυκτών πόρων. Τέλος η Γη διαθέτει συγκεκριμένα αποθέματα πόρων. Οι περισσότεροι πόροι που αντλεί ο άνθρωπος για να καλύψει τις ενεργειακές του ανάγκες, δημιουργούνται έπειτα από φυσικές διεργασίες διάρκειας εκατομμυρίων χρόνων εν αντιθέσει με τους γοργούς ρυθμούς κατανάλωσης τους από τις ανθρώπινες κοινωνίες. Στο εγγύς μέλλον ο προβληματισμός επί

της κάλυψης των ενεργειακών μας αναγκών θα εντείνεται όσο τα αποθέματα θα φτάνουν προς το τέλος τους.

Η ανθρωπότητα δεν είχε την κριτική σκέψη ή ίσως τις γνώσεις να προ οικονομήσει την κατάσταση την οποία θα έφεραν οι επιλογές της. Με την ανάπτυξη όμως που επέφεραν τα ορυκτά καύσιμα στην πορεία, διαθέτει πλέον τα μέσα και τους τρόπους για να ανατρέψει την δυσμενή κατάσταση στην οποία βρίσκεται. Ο στόχος που τέθηκε παγκοσμίως είναι η μείωση των ζημιωγόνων ρύπων. Μια απόφαση που έχει σκοπό την απεξάρτηση εν τέλη των κοινωνιών από τα ορυκτά καύσιμα.

2.4 Παρόν και Μέλλον με Στόχο την Επίτευξη Ουδέτερου Ισοζυγίου Άνθρακα

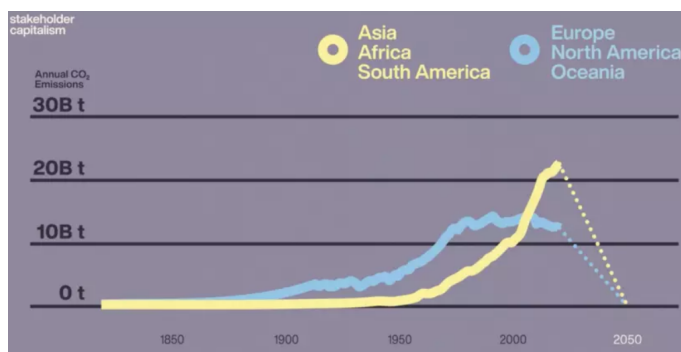
Για τη συγγραφή των παρακάτω ενοτήτων πληροφορίες αντλήθηκαν από τις παρακάτω πηγές [2, 30, 21, 13, 2]. Οι προβληματισμοί που εγείρονται σχετικά με τον τρόπο κάλυψης των ενεργειακών μας αναγκών στο μέλλον και την επερχόμενη κλιματική αλλαγή, αποτελούν αντικείμενο μελέτης, συζήτησης και αντιπαράθεσης ανάμεσα στις επιστημονικές κοινότητες τις τελευταίες δεκαετίες. Στη διάρκεια αυτών αφυπνίστηκε σιγά-σιγά η ανθρωπότητα. Ενημερώθηκε, παρατήρησε τα σημάδια στη φύση γύρω της, πραγματοποίησε διαδηλώσεις, στοχοποίησε ανθρώπινες δραστηριότητες που υπονόμευαν το περιβάλλον, ανέπτυξε συμπεριφορές φιλικές προς αυτό, ξεκίνησε να παράγει “πράσινη” ενέργεια και να προστατεύει συλλογικά σε μεγαλύτερο βαθμό το σπίτι της, τον πλανήτη Γη. Ωστόσο τα τελευταία χρόνια παρατηρείται μια παγκόσμια εγρήγορση γύρω από το θέμα της κλιματικής αλλαγής.

Στις προσπάθειες που καταβάλλονται συνολικά από την ανθρωπότητα για να αντιστραφούν οι συνέπειες της κλιματικής αλλαγής που προκαλεί η συνεχής άνοδος της θερμοκρασίας, χρονιά ορόσημο αποτέλεσε το 2015. Τη χρονιά αυτή επικυρώθηκε η συμφωνία των Παρισίων. Μια συμφωνία που υπογράφηκε ανάμεσα στα 193 μέλη των Ηνωμένων Εθνών. Οι αποφάσεις που ελήφθησαν, αφορούσαν τον αγώνα κατά της κλιματικής αλλαγής. Πιο συγκεκριμένα, για την θερμοκρασία του πλανήτη αποφάνθηκε η διατήρηση της αυξημένης μέσης παγκόσμιας θερμοκρασίας σε επίπεδα πολύ κάτω των 2 βαθμών Κελσίου σε σχέση με τα προ βιομηχανικά επίπεδα 200 χρόνια πριν. Επιθυμητό είναι η αύξηση της να μην ξεπεράσει τους 1,5 βαθμούς στην κλίμακα Celsius. Γι’ αυτό αποφασίστηκε η υιοθέτηση τολμηρών πολιτικών και μέτρων στα οποία δεσμεύθηκαν όλες οι κυβερνήσεις ώστε να μειωθούν οι ρύποι. Αρχικά έως το 2030 οι ρύποι που εκλύονται από καύση ορυκτών καυσίμων θα πρέπει να έχουν μειωθεί κατά 55% τουλάχιστον σε σχέση με όσους παράγονταν το 1990. Η μακροπρόθεσμη, σταδιακή και στρατηγική μείωση των ριπών θα αποφέρει εν τέλη την απεξάρτηση της ευρωπαϊκής οικονομίας από τους πόρους που χάρη σε αυτούς αναπτυσσόταν τα τελευταία 200 χρόνια. Έτσι ως το 2050 κοινωνία και οικονομία θα έχουν επιτύχει τον στόχο της ουδετερότητας τους ως προς τον άνθρακα.

Κάποιες χώρες προβλέπεται πως δεν θα τηρήσουν τη συμφωνία. Από τον 19ο αιώνα έως σήμερα οι χώρες που ανέπτυξαν και στήριξαν τη βιομηχανία τους, παρουσίασαν αξιοσημείωτη πρόοδο. Η εύρυθμη λειτουργία της προϋπόθετε τη καύση ορυκτών καυσίμων. Στον αντίποδα

οι εκπομπές άνθρακα, αυξανόντουσαν χρόνο με τον χρόνο. Οι κοινωνίες στην Ευρώπη, στην Βόρεια Αμερική και στην Ωκεανία ήταν οι πρώτες που επέλεξαν να εξαρτηθούν από τα ορυκτά καύσιμα, για να αναπτυχθούν πιο γρήγορα. Η μεγάλη διαφορά που υπάρχει σήμερα ανάμεσα στις χώρες της Δύσης και στις χώρες, που γεωγραφικά ανήκουν στην Ασία, στην Αφρική και τη Νότια Αμερική, έγκειται στην φάση που βρίσκονται όσον αφορά την παραγωγή ρύπων που σχετίζονται άμεσα με την ανάπτυξη τους. Στις δυτικές κοινωνίες η κορύφωση των εκπομπών τους επήλθε περίπου στις αρχές της δεκαετία του 90 ή λίγο πριν. Μετά το σημείο αυτό και για είκοσι περίπου χρόνια οι αυξομειώσεις, που παρουσίαζαν οι εκπομπές CO₂, ήταν μικρής κλίμακας.

Τα τελευταία χρόνια θετικό σημάδι στην προσπάθεια επίτευξης μηδενικών ριπών μέχρι το 2050, αποτελούν οι ποσότητες ρύπων άνθρακα που μετρούνται κάθε χρόνο στην ατμόσφαιρα, μιας και ακολουθούν καθοδική πορεία. Από την άλλη μεριά, οι αναδυόμενες οικονομίες σε Ασία, Αφρική και Νότια Αμερική δεν έχουν φέρει τις οικονομίες τους στην επιθυμητή θέση, για να ξεκινήσουν να εκλύουν λιγότερο διοξείδιο του άνθρακα, παρά βρίσκονται σε ανοδική πορεία χωρίς να έχουν φτάσει το σημείο κορύφωσης τους ακόμα οι εκπομπές CO₂. Στο Σχήμα 2.1 φαίνεται η πορεία τους. Βρίσκονται, δηλαδή, σε ανοδικό σημείο της περιβαλλοντικής καμπύλης του Kuznets εν αντιθέσει με τις δυτικές ανεπτυγμένες οικονομίες. Σύμφωνα με αυτήν, όσο αναπτύσσεσαι, παράγεις πολλούς ρύπους. Στη συνέχεια της εξέλιξη σου, παράγονται αρκετοί ακόμη έως ότου φτάνεις στο σημείο που έχεις αποκτήσει τη γνώση, τα μέσα και τον τρόπο να έχεις λιγότερες εκπομπές. Από το σημείο εκείνο και μετά οι ρύποι που παράγονται, μειώνονται συνεχώς. Συνεπώς η επίτευξη του στόχου από μεριάς των χωρών αυτών φαντάζει δύσκολος καθώς ο χρόνος πιέζει.

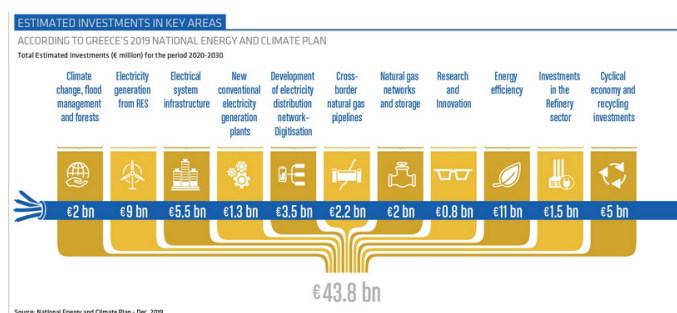


Σχήμα 2.1: Ιστορική εξέλιξη παραγόμενων ποσοτήτων ρύπων CO₂ ανά τον κόσμο [21].

Οι ισχυροί δεσμοί εξάρτησης, που υπάρχουν ανάμεσα στις οικονομικές δραστηριότητες του ανθρώπου και στις ποσότητες ενέργειας, που τις τροφοδοτούν, δεν γίνεται να αλλάξουν. Γίνεται, όμως, να αλλάξουν οι συσχετίσεις ανάμεσα στις εκπομπές CO₂ και στον τρόπο που παράγεται η ενέργεια που χρειάζονται αυτές για να πραγματοποιηθούν. Δύο είδη επενδύσεων στηρίζουν την επιλογή αυτή. Ενέργειες όπως η αποκατάσταση υποβαθμισμένων περιοχών, η προστασία περιοχών φυσικού κάλους, οι αναδασώσεις κλπ. που συμβάλουν στην αναβάθμιση του περιβάλλοντος και στην απορρόφηση με φυσικό τρόπο αερίων του θερμοκηπίου είναι η μια επιλογή. Η άλλη επιλογή αφορά επενδύσεις σε πράσινες υποδομές, σε έργα παραγωγής

ηλεκτρικής ενέργειας από ανανεώσιμες πηγές ενέργειας, στο δίκτυο ηλεκτροδότησης, στις μεταφορές και στην αποδοτικότερη διαχείριση της ενέργειας. Δράσεις που θα βελτιώσουν τους δείκτες βιωσιμότητας, απόδοσης, παραγωγικότητας και ανθεκτικότητας των κοινωνιών σε οικονομικές και κλιματικές αλλαγές. Επενδύσεις που, αν γίνουν, θα συμβάλουν τα μέγιστα στη μείωση των ρύπων άνθρακα και την επίτευξη του μηδενικού ισοζυγίου άνθρακα έως το 2050.

Η χώρα μας, η Ελλάδα, ανήκει στη Δύση. Ακολουθεί παρόμοια πολιτική με τα υπόλοιπα κράτη μέλη της Ευρωπαϊκής Ένωσης. Υποστηρίζει τη συμφωνία των Παρισίων και συνέταξε πλάνο για την επίτευξη των στόχων που ορίζει η συμφωνία ως το 2030. Πριν τη συμφωνία, στη χώρα γίνονταν μικρής κλίμακας “πράσινες” επενδύσεις. Αφορούσαν κυρίως τη παραγωγή ηλεκτρικής ενέργειας από έργα ΑΠΕ όπως την εγκατάσταση μικρών πάρκων ανεμογεννητριών ή μικρών υδροηλεκτρικών σταθμών. Από την συμφωνία και έπειτα οι προσπάθειες της εντάθηκαν καθώς αποτελούσε μια από τις λίγες χώρες στην Ευρώπη που ενεργειακά εξαρτιόταν ακόμα σε μεγάλο ποσοστό από την καύση ορυκτών όπως ο λιγνίτης και το πετρέλαιο. Προσπάθειες που είχαν ως αποτέλεσμα οι ρύποι άνθρακα να ακολουθήσουν καθοδική πορεία και ανοδική οι ποσότητες ενέργειας που οφείλονταν σε ΑΠΕ. Το 2019 ανακοινώθηκαν επενδύσεις ύψους 43,8 δισεκατομμυρίων που αφορούν τους τομείς της ενέργειας και της καταπολέμησης της κλιματικής αλλαγής και έχουν ορίζοντα υλοποίησης το 2030. Στο Σχήμα 2.2 φαίνονται τα ποσά τα οποία θα δαπανηθούν ανά τομέα.



Σχήμα 2.2: Μελλοντικές επενδύσεις που απορρέουν από το ελληνικό σχέδιο για την ενέργεια και το κλίμα έως το 2030 [2].

Ο τομέας ο οποίος θα ενισχυθεί πιο πολύ από όλους είναι ο τομέας της ενεργειακής απόδοσης, του τρόπου δηλαδή με τον οποίο θα παράγεται, θα διανέμεται και θα καταναλώνεται πιο αποδοτικά η αναγκαία ενέργεια. Το δεύτερο μεγαλύτερο ποσό, ύψους 9 δισεκατομμυρίων, έχει δεσμευθεί για την παραγωγή ηλεκτρικής ενέργειας από έργα ΑΠΕ. Αντίθετα ένα εκ των χαμηλότερων, ύψους 2 εκατομμυρίων αφορά την δημιουργία εγκαταστάσεων φυσικού αερίου. Ένα πλάνο που δείχνει την θέληση της ελληνικής κοινωνίας να αποστραφεί της χρήσης ορυκτών πόρων και την μεταστροφή της προς πράσινες πηγές ενέργειας για να επιτύχει τον στόχο μείωσης των ριπών της στον επιθυμητό βαθμό που ορίζει ο χρονικός ορίζοντας του 2030. Πέρα από το πλάνο που έχει συντάξει η πολιτεία, σαφείς ενδείξεις ότι κινούμαστε προς την κατεύθυνση αυτή παρέχουν οι Πίνακες 2.1 και 2.2:

Με μια προσεκτική ανάγνωση των πινάκων αυτών προκύπτουν ενδιαφέροντα συμπεράσμα-

	2007	2014	2020
Άνθρακας	45,19	35,26	9,46
Πετρέλαιο	54,12	33,48	27,83
Φυσικό Αέριο	7,81	5,56	11
Συνολικά	114,59	78,66	54,24

Πίνακας 2.1: Ετήσιες ποσότητες εκπομπών CO₂ σε εκατομμύρια τόνους ανά είδος καυσίμου στην Ελλάδα [13].

	2007	2010	2014	2017	2020
Άνθρακας	34,68	30,8	25,75	17,19	7,39
Πετρέλαιο	9,28	6,2	5,63	5,83	5,83
Φυσικό Αέριο	14,16	9,85	6,79	14,19	17,76
Υδροηλεκτρικά	2,59	7,46	4,48	5,74	3,79
Ανεμογεννήτριες	1,82	2,71	3,69	6,3	8,9
Φωτοβολταϊκά	0,01	0,16	3,79	3,79	4,35
Συνολικά	62,71	57,37	50,34	53,25	48,33

Πίνακας 2.2: Ετήσια παραγωγή ηλεκτρικής ενέργειας σε TWh ανά πηγή στην Ελλάδα [14].

τα. Στον Πίνακα 2.1 φαίνονται οι ετήσιες ποσότητες CO₂ που προήλθαν από καύση άνθρακα, πετρελαίου και φυσικού αερίου. Παρατηρείται ότι σε χρονικό διάστημα σχεδόν δεκαπέντε χρόνων έχει επιτευχθεί μείωση στην παραγωγή αυτών ποσοτικά. Σε αριθμούς αυτό μεταφράζεται σε 70 εκατομμύρια τόνους λιγότερο ετησίως. Το 2020 παρήχθησαν συνολικά περίπου 55 εκατομμύρια τόνοι διοξειδίου του άνθρακα την στιγμή που τις χρονιές 2007 και 2014 παρήχθησαν 115 εκατομμύρια τόνοι και 78 εκατομμύρια τόνοι CO₂ αντίστοιχα από την καύση κάποιου ορυκτού καυσίμου. Οι τιμές ανά καύσιμο αποτυπώνουν μια πτωτική τάση ως προς την παραγόμενη ποσότητα ρύπων. Βέβαια δεν παρατηρείται ίδιου μεγέθους μείωση σε κάθε ένα από αυτά. Στον άνθρακα παρατηρείται η μεγαλύτερη καθώς ο ίδιος και τα παράγωγα του όπως ο λιγνίτης δεν χρησιμοποιούνται πλέον για την παραγωγή ηλεκτρικής ενέργειας παρά στον ελάχιστο βαθμό. Πτωτική πορεία επίσης ακολουθούν οι ρύποι που παράγονται κατά την καύση πετρελαίου. Το πετρέλαιο όμως, καθώς αξιοποιείται για την κάλυψη διαφόρων αναγκών του ανθρώπου όπως η θέρμανση και οι μεταφορές, δεν παρουσιάζει την ίδια μείωση ποσοστιαία με τον άνθρακα. Την στιγμή που οι ρύποι λόγω άνθρακα έχουν μειωθεί κατά 80% οι αντίστοιχοι του πετρελαίου έχουν μειωθεί κατά 50%. Αντίθετα, ανησυχητική θα ήταν η αυξητική τάση που παρουσιάζουν οι ρύποι που προέρχονται από την καύση φυσικού αερίου, εάν αυτό δεν θεωρείτο οικολογικό. Το γεγονός πως είναι απαλλαγμένο από υδρογονάνθρακες το κάνουν πιο φιλικό προς το περιβάλλον. Έτσι λοιπόν σε αντίθεση με τα υπόλοιπα καύσιμα, αυτό επιδεικνύει μια αύξηση της τάξης του 45% αφού πλέον έχει αντικαταστήσει τα πιο επιβλαβή ορυκτά σε πολλές ανθρώπινες δραστηριότητες. Οι Έλληνες πολίτες, οι επιχειρήσεις που δραστηριοποιούνται σε αυτήν και συνολικά η πολιτεία βρίσκονται ξεκάθαρα σε μια προσπάθεια

αποστροφής των ορυκτών καυσίμων που ζημιώνουν το περιβάλλον. Στρέφονται για αυτό το λόγο σε λιγότερο ρυπογόνα ορυκτά όπως το φυσικό αέριο για τη κάλυψη σημαντικών διαχρονικών δραστηριοτήτων τους όπως η θέρμανση, το μαγείρεμα και η παραγωγή ηλεκτρικής ενέργειας.

Παράλληλα όπως υποδεικνύουν τα στοιχεία στον Πίνακα 2.2, σαν χώρα η Ελλάδα μεθοδικά αυξάνει τις ποσότητες ηλεκτρικής ενέργειας που παράγει χάρη σε έργα ΑΠΕ. Αυξάνονται δηλαδή οι ποσότητες παραγωγής καθαρής ενέργειας την στιγμή που μονάδες παραγωγής ηλεκτρικής ενέργειας που χρησιμοποιούν ρυπογόνα καύσιμα όπως πετρέλαιο ή λιγνίτη είτε κλείνουν είτε μειώνουν αισθητά τις ώρες λειτουργίας τους και συνεπώς τη συνεισφορά τους σε ρύπους και ηλεκτρικό ρεύμα. Το 2007 παρήχθησαν περίπου 58 TWh με την καύση ορυκτών καυσίμων, ενώ οι TWh που παρήγαγε χάρη σε πράσινες πηγές ενέργειας δεν ξεπερνούσαν τις 5. Σχεδόν 15 χρόνια πριν η παραγωγή ηλεκτρικού ρεύματος στηριζόταν κατά 90 % στην καύση ορυκτών καυσίμων. Στο πέρασμα των ετών καταγράφονται αυξήσεις στις παραγόμενες ποσότητες ηλεκτρικής ενέργειας από ΑΠΕ και μειώσεις σε αυτές από ορυκτά καύσιμα. Φτάνουμε στο σημείο, όπου το 2020 η χώρα παράγει σχεδόν 17 TWh από έργα ΑΠΕ που ισοδυναμεί με το 35% της παραγωγής της σε αντίθεση με το 7% που λάμβαναν το 2007. Έχει δηλαδή αυξηθεί κατά 500% την ποσότητα ηλεκτρικής ενέργειας που παράγει από καθαρές μορφές ενέργειας.

Από την άλλη μεριά η χρήση λιγνίτη και πετρελαίου για την παραγωγή ηλεκτρισμού μειώνεται δραστικά. Η μείωση των παραγόμενων ρύπων κατά 80% στην περίπτωση του άνθρακα όπως είδαμε από τον Πίνακα 2.1 συνοδεύεται από μείωση κατά 72% της παραγόμενης ποσότητας ηλεκτρικής ενέργειας όπως φανερώνεται από τον Πίνακα 2.2. Στην περίπτωση του πετρελαίου ισχύει το ίδιο. Η μείωση στην ποσότητα ηλεκτρικού ρεύματος κατά 44% οδηγεί στην ταυτόχρονη μείωση των παραγόμενων ρύπων κατά 50%. Αντίθετα αφού αυξάνεται η χρήση του φυσικού αερίου που είναι το ορυκτό το οποίο κατά την καύση του απελευθερώνει τους λιγότερους και τους πιο αβλαβείς ρύπους, αυξάνονται και οι ποσότητες αυτών. Φαίνεται λοιπόν πόσο στενά συνδεδεμένες και αλληλοεξαρτώμενες είναι αυτές οι δύο μεταβλητές. Η σχέση τους είναι κλειδί στην προσπάθεια επίτευξης ουδέτερου ισοζυγίου άνθρακα.

Παρόλα αυτά, είναι φανερό η προσπάθεια που γίνεται ώστε η παραγωγή ηλεκτρισμού να προέρχεται σε όλο και μεγαλύτερα ποσοστά από έργα ΑΠΕ. Από αυτά την πρωτοκαθεδρία στην παραγωγή ρεύματος κατέχουν οι ανεμογεννήτριες και τα αιολικά πάρκα στα οποία εντάσσονται. Η αξιοποίηση του ανέμου και συνεπώς του αιολικού δυναμικού της χώρας προσφέρει πλέον το 20% της συνολικής παραγόμενης ενέργειας και αποτελεί σχεδόν το 50% της συνολικής πράσινης ενέργειας της χώρας μας. Το 2020 παράχθηκαν περίπου 9 TWh, ποσότητα που φτάνει και ξεπερνάει αθροιστικά τη συνολική ποσότητα ηλεκτρικής ενέργειας που παρήγαγαν όλα τα υπόλοιπα έργα ΑΠΕ στην χώρα, την στιγμή που το 2007 παράγονταν λιγότερα από 2 TWh. Μέσα σε 15 χρόνια η ετήσια παραγωγή ηλεκτρικής ενέργειας χάρη στο αιολικό δυναμικό σχεδόν τετραπλασιάστηκε και συνεχίζει να ακολουθεί την αυξητική αυτή τάση. Μια αντίστοιχη παρατήρηση μπορεί να γίνει και για τα πάρκα με ηλιακούς συλλέκτες. Παράξενη παραμένει η αντιμετώπιση της ελληνικής κοινωνίας προς τα υδροηλεκτρικά έργα αφού σε απόδοση παρουσιάζουν την βέλτιστη σε σχέση με όλα τα υπόλοιπα. Αρχικά παρουσίασαν

μεγαλύτερη συμμετοχή στην εγχώρια ετήσια παραγωγή ηλεκτρικής ενέργειας. Έφτασαν σε σημείο να αποτελούν τον πυλώνα της πράσινης ανάπτυξης 10 χρόνια πριν. Τότε κατείχαν το 12% της συνολικής παραγωγής και το 72% από τις πράσινες μορφές ενέργειας. Στην πορεία, όμως, είτε εγκαταλείφθηκαν είτε δεν συντηρήθηκαν με αποτέλεσμα να αχρηστευθούν και να αποτελούν σήμερα μόνο το 7% της συνολικής παραγωγής ηλεκτρικού ρεύματος.

Με την μικρή ανάλυση που έγινε στους δύο στατιστικούς πίνακες, φάνηκε πόσο σημαντική είναι για την χώρα και την υλοποίηση των ενεργειακών στόχων της η αξιοποίηση του υψηλού αιολικού της δυναμικού. Η εκμετάλλευσή του, σε συνδυασμό με τη ραγδαία ανάπτυξη των τεχνολογιών που ενσωματώνεται στις σύγχρονες αποδοτικές ανεμογεννήτριες, συμβάλλει καθοριστικά στη βιώσιμη ανάπτυξη της χώρας, στη μείωση των ρύπων, στην προστασία του περιβάλλοντος και άρα στην αντιμετώπιση της κλιματικής αλλαγής. Ο άνεμος και η αξιοποίηση της ενέργειάς του με τον βέλτιστο δυνατό τρόπο θα ορίζουν την ενεργειακή της ευημερία στο μέλλον.

Κεφάλαιο 3

Μελέτη και Ανάλυση του Ανέμου

3.1 Σύνοψη

Το κεφάλαιο αυτό επικεντρώνεται στην μελέτη και την κατανόηση της υπόστασης και της συμπεριφοράς της μεταβλητής του ανέμου. Αρχικά γίνεται λόγος για την πηγή προέλευσης του, τα κύρια γνωρίσματα του και τη μεταβλητότητα του. Στη συνέχεια, γίνεται μια προσπάθεια κατηγοριοποίησης του με βάση τη διεύθυνση του, την ένταση του και τη χωρική κλίμακα που λαμβάνει κατά την κίνηση του. Αναφέρονται και περιγράφονται συντόμως ανεμολογικά φαινόμενα της κάθε κλίμακας που παρατηρούνται ανά τον κόσμο γενικά. Τέλος, παρέχονται ουσιώδεις πληροφορίες για το αιολικό δυναμικό της Ελλάδας και τα κύρια ανεμολογικά φαινόμενα αυτού.

3.2 Η Γέννηση του Ανέμου στη Γη

Η Γη, καθώς περιστρέφεται γύρω από τον ήλιο, δέχεται αδιάλειπτα ηλιακή ενέργεια [15]. Η ηλιακή ακτινοβολία, που προσπίπτει στη επιφάνεια της, έχει ως αποτέλεσμα, να θερμαίνει την ατμόσφαιρα της. Ωστόσο, από περιοχή σε περιοχή, παρατηρούνται διαφορετικές θερμοκρασίες. Η διαφορά αυτή έγκειται κατά κύριο λόγο στις διαφορετικές γωνίες πρόσπτωσης που σχηματίζει η ηλιακή ακτινοβολία με την επιφάνεια της εκάστοτε περιοχής του πλανήτη ανά εποχή. Ουσιαστικά λόγω της καμπυλότητας της Γης η ηλιακή ακτινοβολία που προσπίπτει στα πιο υψηλά πλάτη εξαπλώνεται σε μεγαλύτερη επιφάνεια την ίδια στιγμή που οι περιοχές κοντά στον Ισημερινό λαμβάνουν πολύ πιο συγκεντρωμένα αυτή τη ποσότητα ηλιακής ακτινοβολίας. Για παράδειγμα, μέρη του κόσμου που βρίσκονται σε υψηλότερα γεωγραφικά πλάτη (πόλοι) θερμαίνονται λιγότερο από αυτές που βρίσκονται στη τροπική ζώνη (0° πλάτος) [18]. Πέρα από την γωνία πρόσπτωσης, σημαντικοί παράγοντες που μπορεί να επηρεάζουν τον ρυθμό με τον οποίο απορροφά θερμοκρασία μια περιοχή είναι είτε η ποσότητα υγρασίας είτε η ανακλαστικότητα του περιβάλλοντος της είτε το πλήθος και τα είδη πετρωμάτων και βλάστησης που συναντώνται σε αυτή. Αυτή η ανομοιομορφία στη θέρμανση της επιφάνειας της Γης, έπειτα από μια φυσική διαδικασία μεταφράζεται σε κίνηση των αέριων μαζών. Η κατεύθυνση αυτής ρυθμίζεται κυρίως χάρη στη περιστροφή της Γης και στην πολυπλοκότητα του αναγλύφου της

(μη ομοιόμορφη κατανομή ξηράς - θάλασσας).

Ανάμεσα σε παρακείμενες περιοχές όπου επικρατούν διαφορετικές θερμοκρασίες είναι πολύ πιθανό να υπάρξει διαφορά στη βαρομετρική πίεση [18]. Στη πιο θερμή περιοχή ο ατμοσφαιρικός αέρας καθώς θερμαίνεται, έχει την ιδιότητα να διαστέλλεται. Γίνεται δηλαδή πιο αραιός, καθώς τόσο η πίεση του όσο και η πυκνότητα του μειώνονται. Ακολουθώντας ανεβαίνει σε υψηλότερο σημείο στην ατμόσφαιρα και εξαπλώνεται γύρω από το σημείο αυτό. Η κίνηση αυτή δημιουργεί ένα κενό, το οποίο αναπληρώνεται από γειτονικές πιο ψυχρές μάζες αέρα. Πιο συγκεκριμένα, αν μια μάζα αέρα ψυχθεί, αυξάνει σε βάρος και πυκνότητα και έχει τη τάση να προβεί σε μια καθοδική κίνηση. Σπρώχνει τις πιο αραιές μάζες και τις αντικαθιστά. Αυτή η κίνηση του αέρα από περιοχές υψηλής πίεσης σε περιοχές χαμηλής πίεσης ονομάζεται άνεμος. Όσο μεγαλύτερη είναι η διαφορά στην πίεση ανάμεσα τους, τόσο πιο γρήγορα ρέει ο αέρας και τόσο δυναμώνει ο άνεμος. Η κίνηση αυτή ενδέχεται να ήταν ευθύγραμμη και απλή ανάμεσα στις επηρεαζόμενες περιοχές εάν η γη δεν περιστρεφόταν. Κατά την περιστροφή της, επιδρά η δύναμη Coriolis στον άνεμο με αποτέλεσμα την εκτροπή της πορείας του και την σχεδόν παράλληλη κίνηση με τις ισοβαρείς γραμμές. Για παράδειγμα στο Βόρειο ημισφαίριο ο αέρας εκτρέπεται κατά τα δεξιά, ενώ στο Νότιο προς την αντίθετη κατεύθυνση. Η δύναμη αυτή επιδρά περισσότερο κοντά στους πόλους, στα πιο υψηλά πλάτη δηλαδή μεγιστοποιείται, ενώ σε μέση κοντά στον ισημερινό, σχεδόν μηδενίζει την επίδραση της [33]. Ακόμη απαραίτητο κρίνεται να αναφερθεί ότι η επίδραση της δεν είναι ίδια σε ανέμους με διαφορετική ένταση. Άνεμοι με μικρή ένταση που η κίνηση τους χωρικά λαμβάνει μικρότερη έκταση, εκτρέπονται λιγότερα εξαιτίας της από ότι άνεμοι οι οποίοι πνέουν πιο δυνατά και έχουν μεγαλύτερη διάσταση.

3.3 Βασικά Γνωρίσματα του Ανέμου

Ο άνεμος είναι ένα από τα βασικά στοιχεία της φύσης γύρω μας. Διαισθητικά ο άνεμος μπορεί να είναι ανόρουβος, αόρατος και απαλός και άλλοτε να είναι δυνατός, θορυβώδης και καταστροφικός. Από τη σκοπιά των μαθηματικών μπορεί να περιγραφεί ως ένα διάνυσμα τριών διαστάσεων. Βέβαια η κατακόρυφη συνιστώσα του έχει τέτοιο μέγεθος σε σύγκριση με το αντίστοιχο της οριζόντιας του που κρίνεται αμελητέα. Έτσι, με την λέξη άνεμος νοείται η οριζόντια κίνηση του και όχι η κατακόρυφη.

Για τον ακριβή προσδιορισμό του είναι ανάγκη να υπάρχει γνώση της έντασης (ταχύτητα) και διεύθυνσης του. Η διεύθυνση του ανέμου ορίζεται από το σημείο του ορίζοντα από όπου πνέει και όχι προς το που πνέει και εκφράζεται σε μοίρες με αρχή τον γήινο μαγνητικό Βορρά [29]. Συνήθως, με βάση την διεύθυνσή του, προσδίδεται σε αυτόν το όνομα, που τον ακολουθεί. Στον Πίνακα 3.1 που ακολουθεί απατώνται οι άνεμοι, όπως επίσημα ονομάζονται και διακρίνονται ανάλογα την διεύθυνση πνοής τους.

Η ισχύς του ανέμου, που συνδέεται άμεσα με την κινητική του ενέργεια, καθορίζεται από την ταχύτητα με την οποία μεταφέρεται στην ατμόσφαιρα. Ο πιο συνήθης τρόπος έκφρασης αυτής της έντασης είναι σε μέτρα ανά δευτερόλεπτο. Βέβαια ανά τόπους και περιστάσεις χρησιμοποιούνται επίσης μετρήσεις αυτής σε χιλιόμετρα ανά ώρα ή μίλια ανά ώρα ή σε κόμβους ή στη κλίμακα Μποφόρ. Η τελευταία ανεμολογική κλίμακα είναι η πιο προσφιλή σε όλους

Διεύθυνση	Επίσημο	Κοινό	Δ.Σ.	Όνομα διεθνές
B (0°)	Βόρειος	Βοριάς	N	North
BA (45°)	Μέσης	Γραίγος	NE	Northeast
A (90°)	Απηλιώτης	Λεβάντες	E	East
NA (135°)	Εύρος	Σιρόκος	SE	Southeast
N (180°)	Νότιος	Νοτιάς	S	South
NΔ (225°)	Λίβας	Γαρμπής	SW	Southwest
Δ (270°)	Ζέφυρος	Πουνέντες	W	West
ΒΔ (315°)	Σκίρων	Μαΐστρος	NW	Northwest

Πίνακας 3.1: Οι άνεμοι ανάλογα την διεύθυνση πνοής τους [29].

αφού αποτελεί μια εμπειρική μέτρηση της ταχύτητας του ανέμου. Ένας απλός παρατηρητής βασιζόμενος στα αποτελέσματα που φέρει ο άνεμος σε στεριά ή θάλασσα μπορεί σύμφωνα με την κλίμακα αυτή, η οποία παρουσιάζεται στον Πίνακα 3.2,

Μποφόρ	Περιγραφή	Παρατηρούμενα Φαινόμενα	Ταχύτητα ανέμου (m/s)
0	Νηνεμία	Ο καπνός υψώνεται κατακόρυφα.	<0.6
1	Υποπνέων	Ο άνεμος μετακινεί τον καπνό αλλά όχι τον ανεμοδείκτη.	0.6-1.7
2-3	Ασθενής	Κινούνται φύλλα και κλαριά δέντρων.	1.8-5.2
4-5	Μέτριος	Κινούνται μικρά κλαδιά και μικρά δέντρα.	5.3-9.8
6	Ισχυρός	Ο αέρας σφυρίζει και μεγάλα κλαδιά κινούνται.	9.9-12.4
7	Σφοδρός	Τα δέντρα κινούνται ολόκληρα και το περπάτημα γίνεται δύσκολο.	12.5-15.2
8-9	Θυελλώδης	Κλαδιά σπάνε, μικρού μεγέθους ζημιές σε σκεπές παρατηρούνται και η όρθια στάση γίνεται δύσκολη.	15.3-21.5
10-11	Καταιγίζων	Δένδρα σπάνε, εκτεταμένες ζημιές σε κτήρια παρατηρούνται και η όρθια στάση γίνεται αδύνατη.	21.6-29
12	Τυφώνας	Μεγάλες και σοβαρές καταστροφές για τον άνθρωπο και την φύση.	>29

Πίνακας 3.2: Η κλίμακα Μποφόρ [29].

να κατηγοριοποιήσει τους ανέμους και να εκτιμήσει την έντασή τους. Αυτή η ένταση στη συνέχεια μπορεί να υπολογισθεί αλγεβρικά μέσω της εμπειρικής σχέσης που περιγράφεται από την Εξ. (3.1)

$$V = 0.836\sqrt{B^3} \text{ [29] ,} \quad (3.1)$$

όπου σαν V και B λογίζονται η ταχύτητα σε m/s και ο εκτιμώμενος αριθμός Μποφόρ αντίστοιχα. Έπειτα καθίσταται δυνατή η μετατροπή της στις επιθυμητές μονάδες μέτρησης ταχύτητας του ανέμου με απλή χρήση των σχέσεων που περιγράφονται παρακάτω στην Εξ. (3.2):

$$1 \text{ m/s} = 3.600 \text{ km/h} = 2.237 \text{ mph} = 1.943 \text{ knots} \text{ [29]}. \quad (3.2)$$

3.4 Κατηγοριοποίηση Ανέμων

Ο άνεμος είναι κατά φύσιν μια απρόβλεπτη μεταβλητή. Θεωρητικά η μεταβλητότητα που του ανέμου μπορεί να διακριθεί σε χωρική και χρονική. Η χωρική υπονοεί ότι ανά χωρική κλίμακα παρατηρούνται άνεμοι με διαφορετικά χαρακτηριστικά. Κάποιοι άνεμοι εξαπλώνονται και επηρεάζουν μεγάλες περιοχές και άλλοι έχουν τοπικό χαρακτήρα αφού έχουν μικρή εμβέλεια. Από την άλλη μεριά με τον όρο χρονική μεταβλητότητα νοείται η διάρκεια εκδήλωσης σε μια περιοχή ενός συγκεκριμένου ανεμολογικού φαινομένου. Υπάρχουν κινήσεις που διαρκούν είτε για λίγες μέρες είτε για λίγες ώρες μέσα στην ημέρα είτε ακόμα για πιο μικρές χρονικές στιγμές. Συνήθως κάθε περιοχή αναπτύσσει ανέμους με ιδιαίτερα γνωρίσματα λόγω των χωρικών ιδιοτήτων της τα οποία παρουσιάζονται ανά χρονικά διαστήματα [5].

Οι παράγοντες που καθορίζουν τα βασικά γνωρίσματα κάθε ανέμου ποικίλουν. Όλοι όμως μπορούν να εξηγηθούν από τη σκοπιά των φυσικών νόμων και διεργασιών. Λόγοι όπως η περιστροφή της Γης και η κλίση του άξονά της, η δύναμη Coriolis, η βαρυτική δύναμη, η γωνία πρόσπτωσης της ηλιακής ακτινοβολίας σε κάθε σημείο της επιφάνειας της Γης, η μορφολογία του εδάφους (βουνά, πεδιάδες, βραχώδεις εκτάσεις, δάση), η πυκνότητα της βλάστησης ή η τραχύτητα του εδάφους που επηρεάζουν το ποσό της ανακλώμενης ακτινοβολίας προς το διάστημα και η ποσότητα υγρασίας οδηγούν στην παρατήρηση διαφορετικών ανεμολογικών φαινομένων ανά τόπο και εποχή. Χάρη σε αυτούς γίνεται δυνατή η λεπτομερής αποσαφήνιση της συμπεριφοράς του κάθε ένα ξεχωριστά. Με τη γνώση και τη κατανόηση της δυναμικής της ατμόσφαιρας μπορούν να βελτιστοποιηθούν ήδη υπάρχουσες εφαρμογές που αφορούν την πρόβλεψη του καιρού για παράδειγμα [31].

Οι διάφορες θερμικές διεργασίες που εμφανίζονται σε διαφορετικές χρονικές και χωρικές κλίμακες σε συνδυασμό με τις δυνάμεις και το ανάγλυφο της φύσης παράγουν πολλά διαφορετικά ανεμολογικά φαινόμενα στον πλανήτη μας. Έτσι λοιπόν στον Πίνακα 3.3 γίνεται μια προσπάθεια κατηγοριοποιήσεως των συστημάτων ανέμου με βάση τη χωρική κλίμακα τους.

Στον πίνακα αυτόν κατηγοριοποιούνται ανάλογα τη χωρική κλίμακα τους οι άνεμοι. Με τον όρο χωρική κλίμακα, όπως προαναφέρθηκε, νοείται το μήκος ή η τυπική διάσταση που καταλαμβάνει κατά την οριζόντια κίνηση της μια αέρια μάζα. Με μια πρώτη ματιά παρατηρούμε τη πλανητική ή παγκόσμια κλίμακα, τη συνοπτική (μεγάλη), τη μέσο-κλίμακα και τη μικρο-κλίμακα. Η κλίμακα τους έχει άμεση σχέση με τη χρονική διάρκεια τους. Όσο μεγαλύτερη είναι τόσο περισσότερο διαρκούν [32]. Στην πλανητική κλίμακα εντάσσονται συστήματα

Χωρική κλίμακα	Τύποι ανέμου	Μήκος οριζόντιας διάστασης
Πλανητική	Παγκόσμια κυκλοφορία	>10000 km
Συνοπτική	Καιρικό σύστημα	1000 km
Μεσοκλίμακα	Τοπικές ορογραφικές ή θερμικά επαγόμενες κυκλοφορίες	10-100 km
Μικροκλίμακα	Τοπική διαμόρφωση της ροής	100 - 1000 m

Πίνακας 3.3: Χωρικές κλίμακες των συστημάτων ανέμου και ένα δείγμα των σχετικών τύπων ανέμου [15].

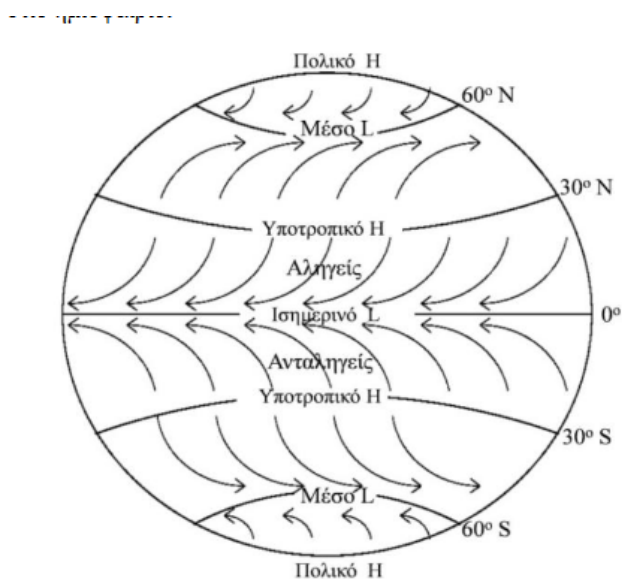
ανέμων τα οποία ως σύνολο μετουσιώνονται στη γενική κυκλοφορία της ατμόσφαιρας. Επηρεάζουν πολύ μεγάλες περιοχές (με έκταση μερικές χιλιάδες χιλιόμετρα) και η εκδήλωσή τους χρονικά διαρκεί από μερικές μέρες έως μερικές εβδομάδες. Στη συνοπτική ή μεγάλη κλίμακα κατατάσσονται φαινόμενα που επηρεάζουν καιρικά συστήματα, όπως τα χαμηλά και υψηλά βαρομετρικά συστήματα ή αλλιώς οι κυκλώνες και αντικυκλώνες των μέσων γεωγραφικών πλατών. Φαινόμενα με ισχύ ικανή να επηρεάσουν αρκετά μεγάλες περιοχές (από εκατοντάδες έως μερικές χιλιάδες χιλιόμετρα) και διάρκεια εξέλιξης από μια ημέρα έως και μερικές μέρες. Έπειτα στη μεσοκλίμακα εντάσσονται άνεμοι που οφείλονται στην τοπική ορογραφία και τοπικές κλιματικές συνθήκες που πνέουν για λιγότερο από 24 ώρες (π.χ αύρες). Τέλος στους μικρο-κλιματικούς ανέμους συναντάμε ανέμους με την εξέλιξη τους να διαρκεί από λίγα λεπτά έως μια ώρα και πνέουν πολύ τοπικά (π.χ ριπές αέρα ή μικροί στρόβιλοι) [15], [5].

Τα ανεμολογικά φαινόμενα που ανήκουν στις κινήσεις πλανητικής κλίμακας χαρακτηρίζονται από την τεράστια οριζόντια σε μήκος διάσταση τους. Είναι τόσο μεγάλη η χωρική κλίμακα τους που ενίοτε συγκρίνεται με τη διάμετρο της Γης (περίπου 10000 km σε μήκος). Τα χαρακτηριστικά αυτών των ανέμων δεν συναντώνται στους ανέμους μέσης ή μικρής κλίμακας. Ακόμη λόγω του ότι αποτελούν τις μεγαλύτερες σε διάσταση κινήσεις, διαρκούν πολύ περισσότερο από όλες τις άλλες. Χαρακτηριστικό παράδειγμα αυτής της κατηγορίας αποτελεί η γενική κυκλοφορία της ατμόσφαιρας, η οποία εξετάζεται συνοπτικά παρακάτω [32].

Η γενική κυκλοφορία της ατμόσφαιρας αποτελεί ένα περίπλοκο σύστημα ανέμων [27], [33]. Απαρτίζεται από διάφορα μέρη τα οποία συνεργάζονται και φέρουν ένα βαθμό συσχέτισης. Χάρη σε αυτήν αποκαθίσταται η ενεργειακή ανισορροπία που προκαλείται από τη ανομοιόμορφη θέρμανση της Γης. Πρόκειται ουσιαστικά για ένα έξυπνο σύστημα κλιματισμού που αποτρέπει τους πόλους της Γης να ψυχθούν περισσότερο και τον ισημερινό να υπερθερμανθεί. Θεωρείται ότι έχει τη μορφή τριών μεγάλων κλειστών ατμοσφαιρικών κυκλοφοριών που συναντώνται τόσο Νότιο όσο και στο Βόρειο ημισφαίριο, όπως φαίνονται στο Σχήμα 3.1.

Αυτές είναι οι:

- Κυκλοφορία Haddley: Αποτελεί το κλειστό σύστημα κυκλοφορίας των μικρών γεωγραφικών πλατών (τροπικών και υποτροπικών περιοχών) του κάθε ημισφαιρίου. Εκτείνεται στα γεωγραφικά πλάτη 0° - 30° νότια και βόρεια του Ισημερινού. Σε αυτές τις περιοχές



Σχήμα 3.1: Η γενική κυκλοφορία της ατμόσφαιρας [29].

θερμαίνεται πολύ ο αέρας, με αποτέλεσμα να γίνεται πιο αραιός και να ανυψώνεται πολύ ψηλά στην ατμόσφαιρα. Από αυτή τη τεράστια μάζα αέρα που έχει συσσωρευθεί στα ανώτερα στρώματα της ατμόσφαιρας και αποτελεί τη υποτροπική ζώνη υψηλών πιέσεων, ένα μέρος της κατέρχεται προς την επιφάνεια της Γης. Στη συνέχεια τείνει να κινηθεί προς τους Πόλους. Ωστόσο, εκτρέπεται αυτής της πορείας χάρη στη δύναμη Coriolis, της οποίας η επίδραση στα σημεία με περίπου 30° γεωγραφικό πλάτος, παύει πλέον να είναι ανεπαίσθητη. Έτσι λοιπόν, την αναγκάζει να επιστρέψει προς τον Ισημερινό όπου και κλείνει τον βρόγχο της. Το ανεμολογικό σύστημα που περιγράφηκε και συναντάται στη κυκλοφορία Haddley, ονομάζεται Αληγείς άνεμοι (Trade Winds) και έχουν κατεύθυνση βορειοανατολική στο Βόρειο ημισφαίριο. Στο Νότιο όπου ονομάζονται Αντί-Αληγείς παρουσιάζεται το ίδιο φαινόμενο απλά με αντίθετη φορά.

- **Κυκλοφορία Ferrel:** Αποτελεί το κλειστό σύστημα κυκλοφορίας των μέσων γεωγραφικών πλατών του κάθε ημισφαιρίου. Εκτείνεται στα γεωγραφικά πλάτη 30° - 60° του νότιου και βόρειου ημισφαιρίου. Αυτός ο βρόγχος εκτελείται ανάμεσα στις περιοχές υψηλής και χαμηλής πίεσης. Η δημιουργία του οφείλεται σε ένα μέρος της αέριας μάζας που κατέρχεται από την υποτροπική ζώνη υψηλών πιέσεων και κινείται προς το Πόλο. Σε γεωγραφικό ύψος περίπου 60° , έρχεται σε σύγκρουση με τους ψυχρούς βαρείς ανέμους που πνέουν στο Βόρειο Πόλο. Σε συνδυασμό με την επίδραση από την περιστροφή της Γης, εκτρέπεται της κανονικά ευθύγραμμης κίνησης και έχουμε σαν αποτέλεσμα την ύπαρξη των Δυτικών ανέμων (Westerlies).
- **Πολική κυκλοφορία:** Αποτελεί το κλειστό σύστημα κυκλοφορίας των υψηλών γεωγραφικών πλατών του Βόρειου ημισφαιρίου. Εκτείνεται στα γεωγραφικά πλάτη 60° - 90° αυτού. Εντοπίζεται ανάμεσα στις υψηλές πιέσεις των πολικών περιοχών και στις χαμηλές των μεγάλων γεωγραφικών πλατών και συντελεί στη δημιουργία των πολικών Ανατολικών

ανέμων (Polar Easterlies). Ουσιαστικά οι ψυχρές μάζες αέρα θα καταποντίζονταν στο Βόρειο Πόλο και θα έρεαν νότια προς τις πιο ζεστές περιοχές, αυτές του Ισημερινού. Ωστόσο, αφού η επίδραση της δύναμης Coriolis είναι τόσο ισχυρή σε τόσο υψηλά γεωγραφικά πλάτη, τις εκτρέπει προς τα νοτιοδυτικά τελικά.

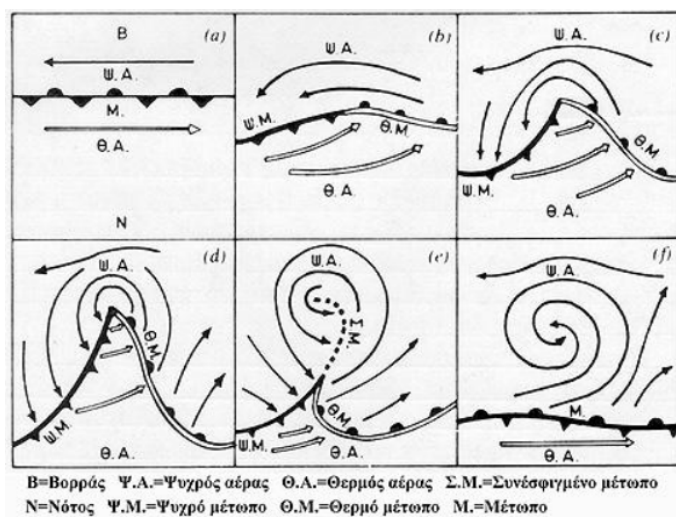
Οι κινήσεις συνοπτικής ή μεγάλης κλίμακας, είναι ουσιαστικά, ανεμολογικά συστήματα τα οποία κατά την οριζόντια διάσταση τους εμφανίζουν μήκος ίσο περίπου με 1000 km. Αυτά τα συστήματα με διάρκεια λίγων ημερών, μικρότερη των πλανητικών αλλά μεγαλύτερη αυτών της μέσης ή μικρής κλίμακας, είναι ικανά να μεταβάλλουν τις καιρικές συνθήκες μεγάλων περιοχών μιας χώρας. Χαρακτηριστικά παραδείγματα αυτής της κατηγορίας είναι κυρίως οι κυκλώνες και οι αντικυκλώνες των μέσων γεωγραφικών πλατών. Οι κυκλώνες που εμφανίζονται στα μέσα πλάτη προκαλούν κακοκαιρίες (βροχοπτώσεις είτε χιονοπτώσεις). Από την άλλη μεριά, ένας αντικυκλώνας φέρνει καλοκαιρία. Σε συνέχεια αυτών, μελετάται με μεγαλύτερη λεπτομέρεια ένα σύστημα μεγάλης κλίμακας, ο κυκλώνας.

Κυκλώνας Μεσαίων Γεωγραφικών Πλατών

Κυκλώνας ή αλλιώς ύφεση ή βαρομετρικό χαμηλό ή απλά χαμηλό είναι όροι, που χρησιμοποιούνται για να περιγραφεί μια περιοχή χαμηλών πιέσεων. Πρόκειται ουσιαστικά για μεγάλες μάζες αέρα οι οποίες κινούνται περιστροφικά γύρω από ένα χαμηλό σε ατμοσφαιρικές πιέσεις, ήρεμο δηλαδή, κέντρο. Όσο κινούμαστε προς την περιφέρεια αυτού οι πιέσεις μεγαλώνουν. Στο Βόρειο ημισφαίριο η φορά σύγκλισης-κίνησης γύρω από το μάτι του κυκλώνα είναι αντίθετη της φοράς του ρολογιού, ενώ σύμφωνη είναι, όταν βρισκόμαστε στο Νότιο ημισφαίριο. Σε κάθε περίπτωση, όμως, το σημείο σύγκλισης είναι το κέντρο. Οι υφέσεις χωρίζονται σε μετωπικές και θερμικές. Κατανέμονται στα μέσα και πιο υψηλά γεωγραφικά πλάτη του Βορείου και Νοτίου ημισφαιρίου τους χειμώνες και τα καλοκαίρια. Η Ευρώπη εμφανίζει τέτοια φαινόμενα σε μεγάλο βαθμό στη Μεσόγειο και στη Βαλτική. Γενικά οι υφέσεις φέρουν έντονες νεφώσεις, βροχές και ανέμους και για αυτό χαρακτηρίζονται ως ο πρόδρομος μιας κακοκαιρίας που θα εξελιχθεί για κάποιες μέρες.

Δυναμικά θα γίνει μια προσπάθεια εξήγησης του τρόπου με τον οποίον γεννάται και εξελίσσεται μια μετωπική ύφεση στα μέσα γεωγραφικά πλάτη. Ένα χαμηλό που εντοπίζεται στα μέσα γεωγραφικά πλάτη (π.χ Ευρωπαϊκός χώρος), σχηματίζεται σταδιακά. Πρώτα πρέπει, να έρθουν σε επαφή δύο μεγάλες μάζες αέρα με κύρια διαφορά τη θερμοκρασία τους. Συνήθως μια τροπική θαλάσσια ή ηπειρωτική μάζα αέρα που πνέει από τα δυτικά επαφίεται έμμεσα μιας πολικής θαλάσσιας ή ηπειρωτικής μάζας που έρχεται από τα ανατολικά. Ανάμεσα τους βρίσκεται ένα μέτωπο (στάσιμο), το οποίο εμποδίζει αυτά τα δύο ρεύματα να αναμιχθούν ελεύθερα, παρά κινούνται παράλληλα με αντίθετη κατεύθυνση το ένα από το άλλο. Μια δύναμη που αναπτύσσεται χάρη στη βαρύτητα και την περιστροφή της Γης, προκαλεί ταλάντωση της μετωπικής επιφάνειας. Τότε ο ψυχρός άνεμος σφηνώνει κάτω από τον θερμό. Η διάβαση του μετώπου αλλάζει απότομα την ένταση και τη διεύθυνση πνοής του αέρα από τα ανατολικά προς τα δυτικά. Πλέον το μέτωπο ταλαντώνεται κατά το οριζόντιο. Εάν η ταλάντωση γίνει αρκετά έντονη και ασταθής, η θερμή μάζα αέρα κυκλώνεται από τη ψυχρή και την οδηγεί να κινηθεί

ανοδικά. Το θερμό ρεύμα αέρα ανέρχεται πάνω από τον ψυχρό και κινείται περιστροφικά με φορά αντίθετη των δεικτών του ρολογιού. Το σύστημα κινείται, ενώ παράλληλα μικραίνει συνεχώς αυτός ο θύλακας θερμού αέρα. Το ψυχρό μέτωπο που ακολουθεί το θερμό, κινείται πιο γρήγορα και τελικά το προ φτάνει όταν πια αυτός ο θύλακας παύει να υφίσταται αφού όλος ο θερμός αέρας έχει μετατοπιστεί πάνω από την επιφάνεια. Ο κυκλώνας λοιπόν αφού ωριμάσει σε διάστημα μιας ημέρας, εξασθενεί και τελικά διαλύεται αφού επέλθει χρονικό διάστημα λίγων ημερών [27]. Το Σχήμα 3.2 υπάρχει για να γίνουν καλύτερα αντιληπτά τα στάδια εξέλιξης του κυκλώνα, όπως περιγράφηκαν παραπάνω.



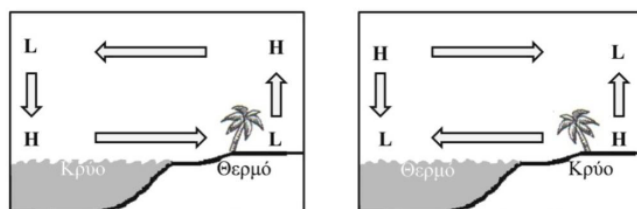
Σχήμα 3.2: Στάδια εξέλιξης κυκλώνα.

Κινήσεις Μεσαίας Κλίμακας Οι μάζες αέρα, οι οποίες κατά την κίνηση τους, λαμβάνουν διαστάσεις από 10 km έως και 1000 km, κατατάσσονται στη μεσοκλίμακα των ατμοσφαιρικών κινήσεων. Αυτοί οι άνεμοι πνέουν ανά τόπους και δείχνουν, να έχουν μια κανονικότητα στην εμφάνισή τους ημερησίως. Αρχικά εκδηλώνονται από αργά το πρωί έως και μέχρι το απόγευμα. Στη συνέχεια κατευνάζονται και πνέουν τις νυχτερινές ώρες σε αντίθετη διεύθυνση σε σχέση με πριν και παύουν νωρίς το πρωί. Αυτός ο κύκλος του ανέμου δικαιολογεί τον ισχυρισμό περί ύπαρξης μιας θερμικής ατμοσφαιρικής κυκλοφορίας. Προκαλείται από τη ανομοιόμορφη θέρμανση του αέρα που παρατηρείται ανά γειτονικές περιοχές και δεν επηρεάζονται από την περιστροφή της Γης. Αρκετά ανεμολογικά φαινόμενα που εντάσσονται στη μέση κλίμακα οφείλονται σε αυτό το φαινόμενο. Τέτοια είναι η θαλάσσια και η απόγεια αύρα όπως επίσης και οι άνεμοι των βουνών και των πεδιάδων που αποτελούν επίσης είδη αύρων ξηράς.

Αύρες Οι αύρες, θαλάσσιες ή απόγειες, εμφανίζονται κατά βάση σε παράκτιες περιοχές των κλιματικά τροπικών και υποτροπικών ζωνών της Γης. Πρόκειται για άνεμους που αναπτύσσονται για κάποιες ώρες μέσα στην ημέρα. Κάθε μια εκδηλώνεται σε διαφορετικό χρονικό διάστημα από την άλλη και αφού ύπαρξη μια παύση ανάμεσά τους. Ακόμη διαφέρουν μεταξύ τους σε ένταση και διεύθυνση.

Η απόγεια αύρα είναι ο άνεμος, που αναπτύσσεται από το βράδυ έως και λίγο μετά την αυγή. Πνέει σταθερά αλλά όχι ιδιαίτερα δυνατά, ενώ η διεύθυνση της είναι από την ξηρά προς την θάλασσα. Στη συνέχεια έπειτα από μια παύση λίγων ωρών, εξελίσσεται η θαλάσσια αύρα. Αρχικά πνέει από την θάλασσα προς την ξηρά, δηλαδή αντίθετα σε σχέση με την απόγεια, αλλά στη συνέχεια πνέει εκατέρωθεν ξηράς και θάλασσας. Ακόμη δεν έχει σταθερή πνοή, αλλά αυξάνει σταδιακά την ένταση της. Τείνει και γίνεται πιο ισχυρή από την απόγεια, όταν προσεγγίζει τη μέγιστη τιμή της το μεσημέρι. Μετά από αυτό το σημείο σταδιακά εξασθενεί, ώσπου παύει περίπου με την δύση του ηλίου. Ακολουθεί μια σύντομη άπνοια την οποία αντικαθιστά η εμφάνιση της απόγειας αύρας. Έτσι, εμφανίζονται καθημερινά αυτοί οι δύο άνεμοι.

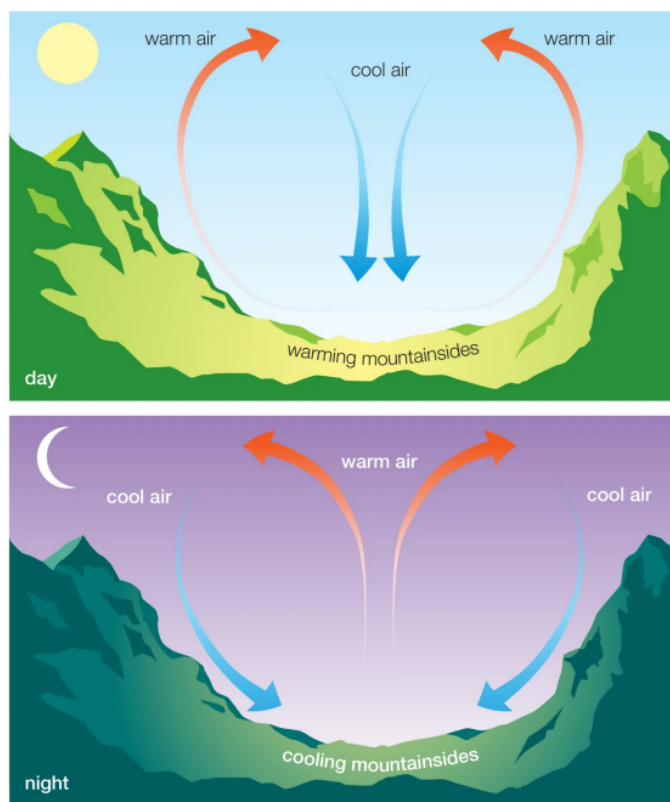
Οι αύρες αποτελούν ανέμους που επηρεάζονται έντονα από μεγαλύτερα συστήματα ανέμων που ανήκουν είτε στη συνοπτική κλίμακα είτε στη πλανητική. Απουσία αυτών και καιρικών καταστάσεων που δεν ευνοούν ηλιοφάνεια (π.χ.νεφελώδης ουρανός) οι αύρες εμφανίζονται πιο έντονες και μπορεί να εξηγηθεί πιο ξεκάθαρα ο τρόπος με τον οποίο αναπτύσσονται. Σαν ξημερώνει σε μια παράκτια περιοχή, ζεσταίνεται με πιο γοργό ρυθμό και πολύ περισσότερο η ξηρά σε σχέση με την επιφάνεια της παρακείμενης και μεγάλης σε όγκο υδάτινης πηγής (λίμνη ή θάλασσα). Οι μάζες αέρα που υφίστανται πάνω από την ξηρά γίνονται πιο αραιές αφού διαστέλλονται και ανεβαίνουν σε μεγαλύτερα ύψη. Παράλληλα στην θάλασσα ο αέρας που βρίσκεται σε χαμηλότερη θερμοκρασία σε σχέση με την ξηρά, συστέλλεται και δημιουργούνται υψηλές πιέσεις. Υφίσταται μια διαφορά πιέσεων η οποία προκαλεί την κίνηση του αέρα από την θάλασσα προς τη στεριά. Ο πιο βαρύς και ψυχρός αέρας αντικαθιστά τον πιο ελαφρύ και πιο θερμό της ξηράς που έχει ανέλθει. Την αντίθετη διαδρομή ακολουθεί στα πιο υψηλά στρώματα ο αέρας με σκοπό την ενεργειακή εξισορρόπηση, με αποτέλεσμα να υπάρχει πνοή ανέμου και προς τις δύο κατευθύνσεις. Τέλος, όσο η μέρα πλησιάζει στη δύση της η ξηρά αποθεμαίνεται πιο γρήγορα σε σύγκριση με τη θάλασσα που έχει μεγάλη θερμοχωρητικότητα. Κατά αυτόν τρόπο εξασθενεί και εν τέλει παύει να υφίσταται η θαλάσσια αύρα. Κατά τη διάρκεια της νύχτας εμφανίζεται η απόγεια αύρα, που ουσιαστικά αποτελεί την αντίστροφη διαδικασία από αυτήν που περιγράφηκε. Βέβαια μιας και το έδαφος όταν ψύχεται δεν προκαλεί τόση έντονη θερμική ανισορροπία όπως όταν θερμαίνεται σε σχέση με τη θάλασσα, έχει ως αποτέλεσμα οι αύρες ξηράς να μην τόσο ισχυρές όσο αυτές της θάλασσας. Στο Σχήμα 3.3, που ακολουθεί, παρίστανται με γραφικό τρόπο οι μηχανισμοί δημιουργίας των αύρων θαλάσσης και ξηράς [31], [32].



Σχήμα 3.3: Οι μηχανισμοί της θαλάσσιας (αριστερά) και απόγειας αύρας (δεξιά) [29].

Αύρες ξηράς δημιουργούνται επίσης χωρίς την παρουσία του υδάτινου στοιχείου. Αρκεί οι

παρακείμενες περιοχές να εμπεριέχουν στο ανάγλυφο τους τα χαρακτηριστικά μιας κοιλάδας και κάποιου ορεινού σχηματισμού. Οφείλονται στην ανομοιόμορφη θέρμανση των πλαγιών του βουνού και της κοιλάδας. Σημειώνεται δε η εξάρτηση τους από την κλίση, τον προσανατολισμό, τη βλάστηση και τη τραχύτητα του εδάφους των πλαγιών. Είναι δηλαδή θερμικά επαγόμενες κυκλοφορίες.



Σχήμα 3.4: Μηχανισμοί αύρων κοιλάδας (πάνω) και βουνού (κάτω) [1].

Με τη βοήθεια του Σχήματος 3.4 είναι δυνατό να γίνει πιο εύκολα κατανοητός ο τρόπος με τον οποίο σχηματίζονται και ο χρόνος κατά τον οποίο εμφανίζονται αυτές οι αύρες. Με τον όρο αύρα κοιλάδας νοείται η κίνηση αέρα που έχει φορά ανηφορική (αναβατικός άνεμος) την ημέρα προς αυτήν. Με τον όρο αύρα βουνών θεωρούμε την αντίστροφη κίνηση του αέρα το βράδυ, την κάθοδο δηλαδή αυτού προς την βάση της κοιλάδας από τις πλαγιές των βουνών (καταβατικός) [29]. Ως αύρες φέρουν μια κανονική μεταβολή στη πνοή τους επί ημερήσιας βάσης. Αφού δύσει ο ήλιος, το έδαφος των πλαγιών ενός βουνού ψύχεται γρηγορότερα από αυτό μιας κοιλάδας λόγω της μικρής θερμοχωρητικότητας του και ψύχει με τη σειρά του τον αέρα κοντά σε αυτόν. Ο αέρας στις πλαγιές των βουνών λοιπόν συστέλλεται, γίνεται πιο βαρύν και καταβαρυνώνεται, παίρνει την κάθοδο αυτών. Κατεβαίνει και οδηγείται μέσω των κοιλάδων σε περιοχές με πιο υψηλή θερμοκρασία (καταβατικός άνεμος). Την ημέρα, οι αέριες μάζες που βρίσκονται στις πλαγιές διαστέλλονται καθώς θερμαίνονται περισσότερο από την ηλιακή ακτινοβολία σε σχέση με τις κοιλάδες. Γίνονται πιο ελαφριές απ' τις παρακείμενες αέριες μάζες και ανέρχονται στις πλαγιές (αναβατικός άνεμος). Ωστόσο καθώς κάθε κοιλάδα

έχει τα δικά της γνωρίσματα όπως διαστάσεις, υψόμετρο, βλάστηση και γεωγραφικό πλάτος, οι αύρες που αναπτύσσονται σε αυτές τείνουν να γίνονται απρόβλεπτες ορισμένες φορές.

3.5 Ανεμολογικά Φαινόμενα της Ελλάδας

Η Ελλάδα είναι μια χώρα προικισμένη με τεράστιο αιολικό δυναμικό. Πολλά διαφορετικά είδη ανέμων πνέουν στην επικράτεια της πρόσκαιρα και ανά εποχές. Η ποικιλομορφία του ελληνικού εδάφους (οροσειρές, κοιλάδες, πεδιάδες), η μεγάλη ακτογραμμή της, η συνεχής εναλλαγή στεριάς με θάλασσα και ο συνδυασμός υψηλών βαρομετρικών πιέσεων σε Σιβηρία και Ατλαντικό Ωκεανό με τις υφέσεις που δημιουργούνται στη Μεσόγειο είναι οι κύριοι λόγοι παρουσίας διάφορων τύπων και εντάσεως ανέμων στην επικράτεια της ανά εποχή.

Το χειμώνα, το φθινόπωρο και την άνοιξη πνέουν πολύπλοκοι και απρόβλεπτοι άνεμοι στον Ελλαδικό χώρο. Το μεγαλύτερο μερίδιο ευθύνης φέρουν τα υψηλά βαρομετρικά συστήματα (κυκλώνες) που αναπτύσσονται στη Σιβηρία και στις Αζόρες του Ατλαντικού. Από τη Σιβηρία, με διεύθυνση νοτιοδυτική, πνέουν οι λεγόμενοι ρωσικοί "Μπουράν". Στην Ελλάδα καταλήγουν ως έντονοι, ψυχροί, βορειοανατολικοί άνεμοι που συνήθως φέρνουν χιόνια μαζί τους. Οι βαρομετρικές πιέσεις στον Ατλαντικό σπρώχνουν μάζες αέρα οι οποίες θερμαίνονται στην Βόρεια Αφρική και έπειτα στρέφονται προς την Ελλάδα. Στο μεταίχμιο απορροφούν υδρατμούς από την Μεσόγειο θάλασσα με αποτέλεσμα να φέρουν βροχές στη χώρα όταν φτάσουν τελικά. Υπάρχει ένας άνεμος όμως ο οποίος παρατηρείται ότι πνέει από Νοέμβρη έως Μάη. Αυτός είναι ο Σιρόκος. Πρόκειται για άνεμο νότιο νοτιοδυτικό που φέρνει μαζί του άμμο από την Αφρική. Εάν διαπεράσει κάποια οροσειρά μεταβάλλεται σε καταβατικό ξερό άνεμο.

Η ποικιλομορφία του ανάγλυφου στην Ελλάδα και η συνεχής εναλλαγή στεριάς θάλασσας είναι ο λόγος που πολλές φορές παρατηρείται σε γειτονικές περιοχές άνεμος με διαφορετική διεύθυνση. Ακόμη τα σχήματα των κόλπων στις ακτές ή των βουνών στη στεριά ή των κοιλάδων αλλάζουν τη διεύθυνση τους και αυξάνουν την ένταση τους. Έτσι κάθε τόπος διαθέτει τις δικές του τοπικές ανεμολογικές συνθήκες.

Κεφάλαιο 4

Θεμελιώδεις Αρχές Αιολικής Παραγωγής

4.1 Σύνοψη

Το κεφάλαιο αυτό επικεντρώνεται στην κατανόηση των βασικών αρχών που διέπουν την αιολική παραγωγή. Αρχικά γίνεται λόγος για την ενέργεια του ανέμου, την αιολική. Στη συνέχεια μελετάται αναλυτικά η μηχανή μετατροπής της αιολικής ενέργειας σε ηλεκτρική. Αφού παρατεθούν διάφορα είδη ανεμογεννητριών που υφίστανται στην παραγωγή, η μελέτη επικεντρώνεται στο είδος που συναντάται πιο συχνά. Για τις μηχανές οριζοντίου άξονα παρατίθενται σημαντικές πληροφορίες για τη δομή και το τρόπο λειτουργίας της. Τέλος, γίνεται αναφορά στα αιολικά πάρκα που δημιουργούνται από συστοιχίες ανεμογεννητριών. Περιγράφονται οι διαδικασίες εγκατάστασης ενός πάρκου σε μια περιοχή και ο τρόπος με τον οποίο ενσωματώνεται η παραγωγή του στο δίκτυο.

4.2 Αιολική Ενέργεια

Ο άνθρωπος από τα πρώιμα ιστορικά χρόνια εκμεταλλεύεται την ενέργεια του ανέμου, την αιολική. Πρόκειται για μια καθαρή μορφή ενέργειας, που η αξιοποίηση της παρουσιάζει χαμηλό περιβαλλοντικό αποτύπωμα. Κατά το παρελθόν, ιστιοφόρα πλοία κινούνταν χάρη σε αυτήν, ανεμόμυλοι θέτονταν σε λειτουργία, ώστε, να αλέθονται τα σιτηρά ή να υδροδοτούνται οικισμοί. Πλέον, η εκμετάλλευση της κινητικής ενέργειας του ανέμου γίνεται για λόγους ηλεκτροπαραγωγής. Υπάρχουν πολλοί λόγοι που εξηγούν την απόφαση αυτή. Αρχικά, το 'καύσιμο' της, ο άνεμος δεν παράγει ρύπους (π.χ. CO_2). Επίσης τα αποθέματα του ανανεώνονται συνεχώς, το κόστος του είναι μηδαμινό, βρίσκεται εύκολα σε περιοχές μακριά από μεγάλα αστικά κέντρα και με την εξέλιξη της τεχνολογίας βελτιώνεται ο βαθμός απόδοσης με την οποία τον εκμεταλλευόμαστε. Ωστόσο μεγάλη πρόκληση στη παραγωγή ηλεκτρικής ενέργειας από την αιολική είναι η υψηλή μεταβλητότητα του αέρα ανά χρονικά διαστήματα. Οι μεταβολές στην ένταση πνοής του μεταφράζονται κυρίως σε αβεβαιότητα, ως προς τη δυνατότητα κάλυψης του απαιτούμενου μέρους της ζήτησης, καθώς η παραγωγή δεν έχει σταθερό ρυθμό.

Πρέπει να σημειωθεί ότι η ισχύς και η ενέργεια του ανέμου αποτελούν έννοιες ετερόνυμες. Η αιολική ενέργεια που αποδίδει ένα ρεύμα ανέμου καθώς πνέει εξαρτάται από τη μάζα και τη ταχύτητά του. Έτσι, μπορούμε να ισχυριστούμε ότι κάθε μάζα αέρα η οποία βρίσκεται σε περιστροφική ή μεταβαλλόμενη κίνηση, έχει κινητική ενέργεια ίση με

$$E_k = \frac{1}{2} m \bar{u}^2, \quad (4.1)$$

όπου m η αέρια μάζα, \bar{u} η μέση ταχύτητα του αέρα στη διάρκεια ενός εύλογου χρονικού διαστήματος (π.χ. δέκα λεπτά ή μια ώρα).

Η παράγωγος της κινητικής ενέργειας ως προς τον χρόνο t δίνει την αιολική ισχύ P_w :

$$P_w = \frac{dE_k}{dt} = \frac{1}{2} \dot{m} \bar{u}^2. \quad (4.2)$$

Ένα ποσοστό αυτής της ισχύος μετατρέπεται σε ηλεκτρική ενέργεια με την χρήση ανεμογεννητριών. Πιο συγκεκριμένα, οι έλικες μιας ανεμογεννήτριας τείνουν να περιστραφούν καθώς τις διαπερνά αέρας. Μια αναπαράσταση της διαδικασίας αυτής παρουσιάζεται στο Σχήμα 4.1. Ανάλογα με το εμβαδόν που σαρώνουν οι έλικες της A , την μέση ταχύτητα ροής του ανέμου \bar{u} και την πυκνότητα του ανέμου ρ μπορούμε να ορίσουμε την ποσότητα ανέμου που γίνεται να αξιοποιηθεί \dot{m} ως

$$\dot{m} = \rho A \bar{u}, \quad (4.3)$$

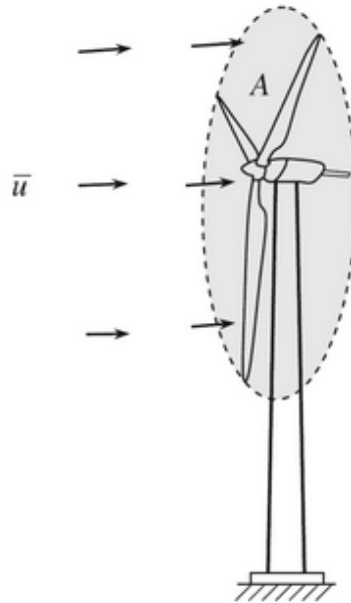
έτσι ώστε να μπορέσουμε να ορίσουμε με ακρίβεια στην Εξ. (4.4) το μέρος της ανεμολογικής ισχύς που τίθεται προς αξιοποίηση P_w τελικά

$$P_w = \frac{1}{2} \rho A \bar{u}^3. \quad (4.4)$$

Μια προσεκτική παρατήρηση της Εξ. (4.4) μας οδηγεί στο συμπέρασμα ότι, περισσότερη αξιοποιήσιμη αιολική ισχύ έχουμε όταν:

- Αυξάνεται το μήκος των ελίκων της ανεμογεννήτριας, ώστε να αυξηθεί το εμβαδόν σάρωσης A αυτών.
- Αυξάνεται η μέση ένταση \bar{u} πνοής της αέριας μάζας που διαπερνά τους έλικες της ανεμογεννήτριας.
- Αυξάνεται η πυκνότητα ρ της αέριας μάζας, γίνεται δηλαδή πιο βαριά (συνήθως αυτό συμβαίνει όταν ψύχεται).

Τέλος, κρίνεται σκόπιμο να αναφερθεί η σχέση που υπάρχει ανάμεσα στη αποδιδόμενη τελική αιολική ισχύ και την μέση ταχύτητα ανέμου [33]. Το γεγονός ότι είναι ανάλογα ποσά και κυρίως το γεγονός ότι η δεύτερη είναι υψωμένη σε τρίτη δύναμη, οδηγεί σε πολύ μεγάλες μεταβολές στο τελικό αποτέλεσμα της πρώτης. Παρεμφερείς ταχύτητες πνοής ανέμου, όταν αξιοποιούνται από μια ανεμογεννήτρια παράγουν τιμές ισχύς με σημαντική διαφορά μεταξύ τους.



Σχήμα 4.1: Περιοχή σάρωσης μια ανεμογεννήτριας [33].

4.3 Ανεμογεννήτριες

Οι πρώτες ανεμογεννήτριες με στρόβιλο εμφανίστηκαν στα τέλη του 20ου αιώνα (1891) στην Δανία χάρη στον Poul la Cour. Ήδη στην Ολλανδία και στις κάτω Χώρες είχαν γίνει προσπάθειες παραγωγής ηλεκτρικής ενέργειας μέσω των ανεμόμυλων αλλά η χαμηλή απόδοση και τα προβλήματα αποθήκευσης της ενέργειας που ανέκυπταν, καθιστούσαν το εγχείρημα αυτό ασύμφορο. Ο προαναφερθέν Δανός εφευρέτης και επιστήμονας, υπερέκρασε το πρόβλημα αυτό με τη βοήθεια της ηλεκτρόλυσης. Με τη παροχή ηλεκτρικού ρεύματος σε μια ποσότητα νερού διασπούσε το νερό σε καθαρό υδρογόνο (και οξυγόνο). Την ενέργεια του υδρογόνου την χρησιμοποιούσε μετέπειτα για τη παραγωγή ηλεκτρικού ρεύματος. Έτσι λοιπόν, ξεκίνησε δειλά δειλά η ανάπτυξη των ανεμογεννητριών με σκοπό την εκμετάλλευση της αιολικής ενέργειας για την παραγωγή ηλεκτρικής ενέργειας.

Μέχρι και στις μέρες μας το κύριο μέσο μετατροπής της αιολικής ισχύς σε μηχανική και στη συνέχεια σε ηλεκτρική είναι η ανεμογεννήτρια. Βέβαια, στην πάροδο των τελευταίων τριάντα χρόνων με την εξέλιξη της τεχνολογίας επήλθαν μεγάλες αλλαγές στην σχεδίαση και κατασκευή διαφόρων τύπων ανεμογεννητριών. Αναβαθμίσεις όπως η σχεδίαση εξελιγμένων αεροδυναμικών μηχανών περιστροφής (ρότορες) ή η εύρεση και τοποθέτηση νέων πιο αποδοτικών εξαρτημάτων παραγωγής, μετάδοσης και ρύθμισης της ενέργειας σε αυτές βελτίωσαν την απόδοση λειτουργίας τους. Σημαντική ήταν επίσης η εξέλιξη που επήλθε στο τρόπο παρατήρησης και πρόβλεψης ανεμολογικών φαινομένων μεσοκλίμακας ή μικρο-κλίμακας. Όλοι αυτοί οι παράγοντες έχουν συμβάλει σε μεγάλο βαθμό στην καθιέρωση των ανεμογεννητριών σε παγκόσμια κλίμακα ως έναν φιλικό προς το περιβάλλον, αποδοτικό και αξιόπιστο μέσο αξιοποίησης της αιολικής ενέργειας για τη παραγωγή ηλεκτρικής ενέργειας [33].

Ανά τον κόσμο υπάρχουν πολλά διαφορετικά είδη εγκατεστημένων ανεμογεννητριών.

Κάποια από αυτά έχουν ευρεία εμπορική χρήση, ενώ άλλα δεν συναντώνται τόσο πολύ στις γραμμές παραγωγής ηλεκτρικής ενέργειας. Έχουν αναπτυχθεί για διαφορετικές εφαρμογές και υπάρχουν πολλοί τρόποι με τους οποίους κατηγοριοποιούνται αυτές. Ο πιο απλός βασίζεται στο τρόπο εκμετάλλευσης του ανέμου από τον άξονα περιστροφής τους και τα είδη ανέμου που δύναται να εκμεταλλευτεί το κάθε είδος. Με βάση αυτό το κριτήριο υπάρχουν οι ανεμογεννήτριες οριζοντίου άξονα και οι ανεμογεννήτριες κατακόρυφου άξονα. Οι δεύτερες είναι οι πιο ευρέως διαδεδομένες. Επιπλέον χωρίζονται σε ανεμογεννήτριες ξηράς ή παράκτιες. Σε αυτή τη περίπτωση το κριτήριο διάκρισης τους είναι ο χώρος εγκατάστασής τους. Περαιτέρω κατηγοριοποίηση μπορεί να γίνει εάν ληφθούν υπόψιν κάποιο χαρακτηριστικό τους, όπως, η εγκατεστημένη ισχύ τους [33].

4.4 Ανεμογεννήτριες Οριζοντίου Άξονα

Στην εργασία αυτή, οι ανεμογεννήτριες των υπό μελέτη αιολικών πάρκων είναι οριζοντίου άξονα. Αποτελούν το είδος ανεμογεννητριών που χρησιμοποιείται κατά κόρον για την παραγωγή ηλεκτρικής ενέργειας. Σε αυτές ο άξονας περιστροφής είναι πάντα παράλληλος με τον αποπνέοντα άνεμο. Προκειμένου να συμβαίνει αυτό διαθέτουν αναλόγως το μέγεθος τους είτε κάποιο ανεμοδείκτη είτε κάποιον αισθητήρα που τις προσανατολίζει προς τη φορά διεύθυνση του ανέμου τον ρότορα. Τοποθετείται στη κορυφή του πύργου μαζί με την ηλεκτρογεννήτρια και το ρότορα. Ο πύργος τους έχει μεγάλο μήκος έτσι ώστε ο ρότορας να βρίσκεται όσο πιο ψηλά γίνεται σε σχέση με το έδαφος, προκειμένου να περιστρέφεται από το βέλτιστο δυνατό άνεμο. Το μεγάλο του πλεονέκτημα έγκειται στο γεγονός ότι τα πτερύγια του μπορούν να ξεκινούν να περιστρέφονται ακόμη και με ανέμους μικρούς σε ένταση (cut-in speed 3 - 4 m/s). Εν γένει αποτελούν συστήματα με υψηλή απόδοση. Δεν παύουν παρόλα αυτά να παρουσιάζουν ορισμένα μειονεκτήματα. Καθώς ο ρότορας και η ηλεκτρογεννήτρια βρίσκονται σε μεγάλο ύψος, περιπλέκεται ο σχεδιασμός της ανεμογεννήτριας, αυξάνεται το κόστος κατασκευής λόγω μεγέθους και είναι δύσκολο να συντελεστούν διεργασίες επιδιόρθωσης σε περίπτωση βλάβης. Το κόστος κατασκευής τους αυξάνεται επίσης και με τον αριθμό των πτερυγίων που έχουν. Συναντώνται ανεμογεννήτριες τέτοιου τύπου με ένα ή δύο ή τρία πτερύγια με τις περισσότερες να αποτελούνται από τρία καθώς παρουσιάζουν καλύτερη αεροδυναμική ισορροπία. Τέλος, κάποιες από αυτές είναι κατασκευασμένες με τον ρότορα τους στραμμένο προς τον άνεμο (up-wind ανεμογεννήτριες), ενώ σε κάποιες άλλες ο άνεμος επαφίεται πρώτα του θαλάμου που περιέχει την ανεμογεννήτρια και έπειτα του ρότορα (down-wind ανεμογεννήτριες) [18].

4.4.1 Συστατικά Μέρη Ανεμογεννήτριας

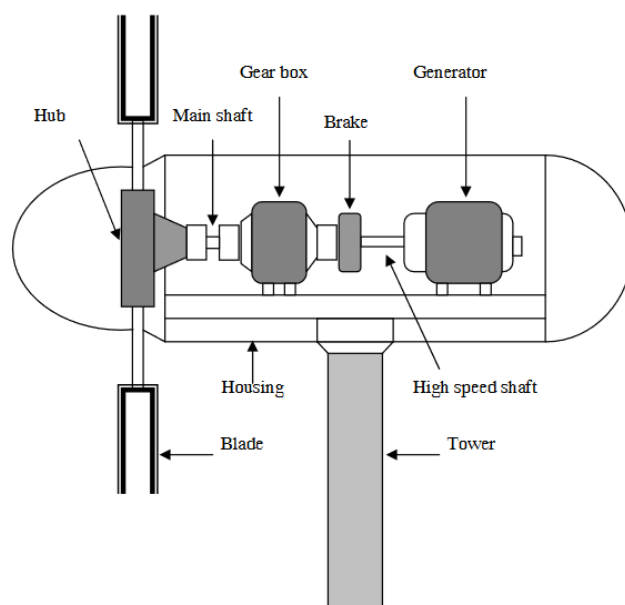
Οι ανεμογεννήτριες οριζοντίου άξονα, όπως προαναφέρθηκε, αποτελούν τον πιο συνήθη τύπο ανεμογεννήτριας στην παραγωγή ηλεκτρικής ενέργειας. Συνήθως χρησιμοποιούνται αυτές, που έχουν ρότορες με τρία πτερύγια. Πέρα από αυτά, άλλα κύρια συστατικά μέρη μιας τέτοιας ανεμογεννήτριας είναι:

- Πύργος (Tower): Ο πύργος της ανεμογεννήτριας είναι ο κορμός αυτής της κατασκευής. Κατασκευάζεται συνήθως από χάλυβα και έχει τη μορφή σωλήνας. Χάρη σε αυτόν στηρίζεται ο ρότορας, η άτρακτος και μεταφέρονται στο έδαφος τα φορτία που αναπτύσσονται σε ύψος πάνω από το έδαφος. Όσο μεγαλύτερο είναι αυτό το ύψος τόσο μεγαλύτερη είναι η ενέργεια που παράγεται αφού ο άνεμος πνέει ισχυρότερος σε πιο μεγάλα υψόμετρα και επίσης δεν παρεμποδίζεται από φυσικά ή τεχνητά εμπόδια (π.χ κτήρια ή βλάστηση) [18].
- Ρότορας (Rotor): Ο ρότορας είναι το πιο σημαντικό μέρος αυτής της μηχανής. Ουσιαστικά μετατρέπει την κινητική ενέργεια του αέρα σε μηχανική δύναμη μέσω της περιστροφής του κύριου άξονα main shaft του. Τα πτερύγια, που είναι τοποθετημένα στη πλήμνη (hub) του, συνήθως κατασκευάζονται από πολυστρωματικό υαλοβάμβακα [18]. Φέρουν κατάλληλη αεροδυναμική σχεδίαση για να αποδίδουν όσο δυνατόν καλύτερα γίνεται και να μην προκαλούν θόρυβο που θα ζημιώνει το οικοσύστημα γύρω του. Τέλος, χαίρουν αντικεραυνικής προστασίας.
- Άτρακτος (Nacelle): Βρίσκεται στο υψηλότερο σημείο του πύργου. Εκεί βρίσκονται τα περισσότερα μηχανικά υποσυστήματα της ανεμογεννήτριας. Μερικά από αυτά είναι η ηλεκτρογεννήτρια (generator), το σύστημα πέδησης (brake system), το κιβώτιο ταχυτήτων gearbox, ο άξονας χαμηλής ταχύτητας (κύριος άξονας ρότορα), ο άξονας υψηλής ταχύτητας (άξονας γεννήτριας), το σύστημα προσανεμισμού (yaw system) όπως επίσης και διάφοροι αισθητήρες και εξαρτήματα που φροντίζουν να λειτουργεί ομαλά και χωρίς κινδύνους βλάβης η ανεμογεννήτρια συνολικά [33], [18]. Πάνω από την άτρακτο (θάλαμο) βρίσκονται συνδεδεμένα ένας ανεμοδείκτης και ένα ανεμόμετρο, ικανά να προσδιορίζουν ανά πάσα στιγμή τη διεύθυνση και τη ταχύτητα του αποπνέοντα ανέμου.
- Κιβώτιο ταχυτήτων (Gearbox): Κατέχει πολύ σημαντικό ρόλο για το συνολικό σύστημα ισχύος της ανεμογεννήτριας. Βρίσκεται ανάμεσα στον άξονα χαμηλών ταχυτήτων και στον άξονα υψηλών ταχυτήτων. Ουσιαστικά αυξάνει την αρχική στροφική ταχύτητα του ρότορα που τελικά αξιοποιεί η ηλεκτρογεννήτρια για την παραγωγή ηλεκτρικού ρεύματος. Για παράδειγμα, όταν ο άξονας του ρότορα περιστρέφεται με μια τυπική ταχύτητα της τάξεως των 30-50 στροφών ανά λεπτό, η ταχύτητα με την οποία περιστρέφεται ο άξονας υψηλών ταχυτήτων, την οποία και εκμεταλλεύεται η ηλεκτρογεννήτρια φθάνει τις 1000-1500 στροφές το λεπτό.
- Γεννήτρια (Generator): Εντάσσεται στα πολύ σημαντικά συστατικά στοιχεία του συστήματος μετατροπής της αιολικής ενέργειας σε ηλεκτρική. Η συγκεκριμένη ηλεκτρογεννήτρια έχει μια μεγάλη διαφορά σε σχέση με άλλες συμβατικές που χρησιμοποιούνται για άλλες εφαρμογές. Λειτουργεί ανάλογα τη συνεχώς μεταβαλλόμενη ταχύτητα του ανέμου. Στις μεγάλες μεγέθους κατασκευές όπου υπάρχει διασύνδεση με το δίκτυο, συναντώνται τριφασικές AC γεννήτριες, είτε ασύγχρονες είτε σύγχρονες. Οι ασύγχρονες γεννήτριες αποδίδουν εντυπωσιακά καλά παρά τις μεταβαλλόμενες συνθήκες λειτουργίας, είναι οικονομικές και δεν υπάρχει υψηλή απαίτηση σε θέματα συντήρησης

και της φροντίδας. Αυτοί είναι μερικοί από τους λόγους προτιμούνται συνήθως έναντι των σύγχρονων γεννητριών [18].

- Σύστημα πέδησης (Safety brakes): Υπάρχουν δύο συστήματα πέδησης σε μια ανεμογεννήτρια, ένα μηχανικό και ένα αεροδυναμικό. Εάν και εφόσον υπάρξουν εξαιρετικά υψηλές ταχύτητες αέρα ενεργοποιείται το σύστημα πέδησης για λόγους ασφαλείας. Το ίδιο σύστημα ενεργοποιείται σε περίπτωση που αποσυνδεθεί η ηλεκτρογεννήτρια ή υπάρξει βλάβη στη γραμμή του ρεύματος. Τέτοιες περιπτώσεις, όπου, μπορεί να υπάρξει απότομη και πολύ μεγάλη επιτάχυνση δεν είναι διαχειρίσιμες και μπορούν να καταστρέψουν το σύστημα. Για αυτό όταν ενεργοποιούνται αυτές οι συσκευές ασφαλείας, το σύστημα παύει να λειτουργεί και εξασφαλίζεται η ασφάλεια του [18].

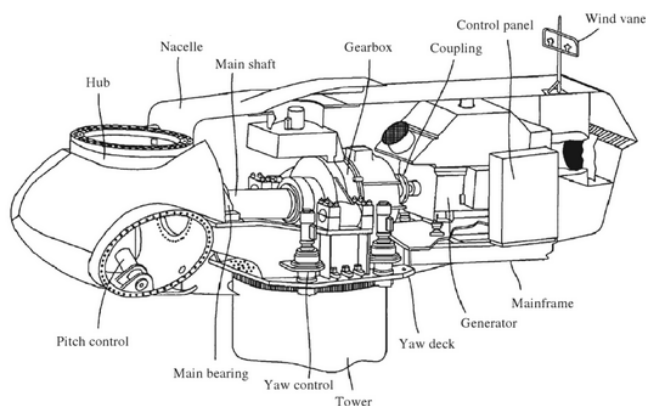
Τα Σχήματα 4.2 και 4.3 παρέχονται για να γίνει ακόμα πιο αντιληπτός ο τρόπος δόμησης και λειτουργίας μιας ανεμογεννήτριας οριζοντίου άξονα.



Σχήμα 4.2: Βασικά μέρη μια ανεμογεννήτριας [18].

4.4.2 Λειτουργία Ανεμογεννήτριας

Η λειτουργία μιας τέτοιας αιολικής μηχανής ξεκινάει με την περιστροφή των πτερυγίων της. Οι μετρήσεις της στιγμιαίας ταχύτητας και της διεύθυνσης του ανέμου δίνουν τις απαραίτητες πληροφορίες στο σύστημα προσανατολισμού πνοής του ανέμου. Κατά αυτόν τον τρόπο τα πτερύγια της ανεμογεννήτριας βρίσκονται πάντα προσανατολισμένα σωστά (στραμμένα προς τον άνεμο), ώστε να αξιοποιήσουν στο έπακρο σε συνδυασμό με το μηχανισμό περιστροφής των πτερυγίων το διαθέσιμο αιολικό δυναμικό. Η διερχόμενος άνεμος ωθεί τα πτερύγια σε περιστροφική κίνηση με κέντρο τον οριζόντιο άξονα της πλήμνης. Μια κίνηση που θέτει σε ανάλογη κίνηση τη πλήμη (hub) του δρομέα αφού σε αυτήν προσδένονται τα



Σχήμα 4.3: Βασικά μέρη μια ανεμογεννήτριας (ανάλυση εσωτερικού ατράκτου) [33].

περιστρεφόμενα πτερύγια. Με αυτόν τον τρόπο γίνεται εκμεταλλεύσιμη η κινητική ενέργεια του ανέμου. Η πλήμνη του δρομέα είναι επίσης συνδεδεμένη με το κιβώτιο ταχυτήτων μέσω ενός στελέχους μέσα στην άτρακτο. Είναι ο άξονας χαμηλών ταχυτήτων της ανεμογεννήτριας. Αυτός συνδέεται με το κιβώτιο ταχυτήτων στο εσωτερικό του θαλάμου. Αυτό συνδέεται με την ηλεκτρογεννήτρια μέσω του άξονα υψηλών ταχυτήτων. Το κιβώτιο ταχυτήτων αναλόγως τη ζήτηση και τις δυνατότητες παραγωγής αυξάνει την αργή περιστροφική κίνηση του δρομέα (άξονας χαμηλών ταχυτήτων) της πλήμνης (ταχύτητα περιστροφής: 10-60 στροφές το λεπτό) και οδηγεί τον δρομέα της ηλεκτρογεννήτριας σε μια πολύ πιο γρήγορη περιστροφική κίνηση (περίπου 1500 στροφές το λεπτό) αφού το στέλεχος που περιστρέφεται είναι συνδεδεμένο με αυτό. Έτσι λοιπόν, η γεννήτρια μετατρέπει την κινητική αυτή ενέργεια σε ηλεκτρική [33].

Συνοπτικά, η αιολική ενέργεια του ανέμου πριν μετατραπεί σε ηλεκτρική, μετασχηματίζεται σε μηχανική. Υπάρχουν, δηλαδή, δύο στάδια τα οποία πρέπει να ολοκληρωθούν πριν παραχθεί ηλεκτρικό ρεύμα μέσα από την αξιοποίηση του αιεθαλή ανέμου. Στη διάρκεια της διαδικασίας αυτής και ανά στάδιο υφίστανται απώλειες, καθώς το σύστημα μετατροπής, η ανεμογεννήτρια, δεν μπορεί να μετατρέψει εξ ολοκλήρου την προσφερόμενη αιολική ενέργεια σε ηλεκτρική. Έπειτα από μελέτη έχει διαπιστωθεί ότι υπάρχουν ορισμένοι παράμετροι που μπορούν να συμβάλουν στην μερική εκτίμηση της αιολικής ισχύς που αξιοποιείται τελικά. Παράμετροι όπως ο συντελεστής ισχύος, οι απώλειες που εμφανίζουν τα μηχανικά μέρη της ανεμογεννήτριας ή η καμπύλη ισχύος της ανεμογεννήτριας μπορούν να προσδώσουν μια αρχική εκτίμηση για την ηλεκτρική ενέργεια που μπορεί να παραχθεί τελικά από μια ανεμογεννήτρια.

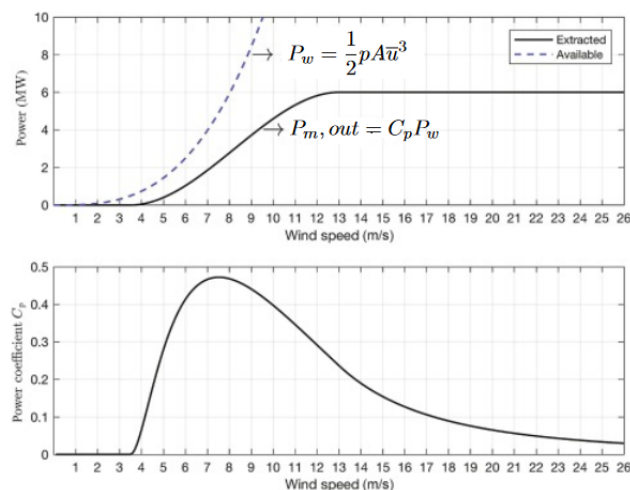
4.4.3 Συντελεστής Ισχύος

Σε πρώτο στάδιο, όπως προαναφέραμε, η κινητική ενέργεια του ανέμου μετατρέπεται σε μηχανική. Ο αποπνέων άνεμος κινεί τα πτερύγια της ανεμογεννήτριας και αυτά στη συνέχεια καθώς περιστρέφονται, τείνουν σε ανάλογη κίνηση το δρομέα μιας ηλεκτρογεννήτριας. Μετατρέπεται έτσι κατά ένα μέρος η κινητική ενέργεια του ανέμου σε μηχανική με την περιστροφή του κυρίου άξονα και εν τέλη αυτού της ηλεκτρογεννήτριας. Για να μετρήσουμε την απόδοση αυτής της μετατροπής, δηλαδή το μέρος της κινητικής ενέργειας του ανέμου που αξιοποιείται

και μετατρέπεται σε μηχανική ισχύ, χρησιμοποιούμε τον συντελεστή ισχύος C_p . Πρόκειται, ουσιαστικά, για τον λόγο που σχηματίζει η πραγματική μηχανική ισχύς που αποδίδουν τα πτερύγια $P_{m,out}$ με την αιολική ισχύ του ανέμου P_w . Από την Εξ. (4.5) περιγράφεται η σχέση αυτή

$$C_p = \frac{P_{m,out}}{P_w} = \frac{P_{m,out}}{\frac{1}{2}\rho A \bar{u}^3}. \quad (4.5)$$

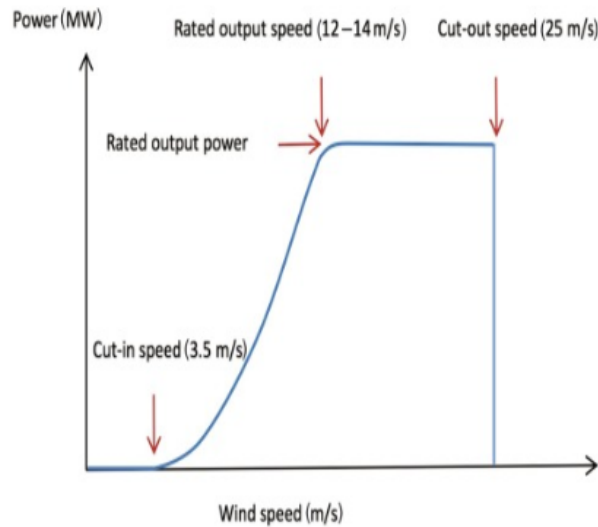
Συνήθως όσο πιο αποτελεσματική είναι η αξιοποίηση του ανέμου από τα πτερύγια της ανεμογεννήτριας, τόσο μεγαλύτερη είναι η μηχανική ισχύς που αποδίδουν. Βέβαια, οι απώλειες που εμφανίζονται λόγω ελαττωματικής αεροδυναμικής σχεδίασης ή λόγω αστοχιών του συστήματος περιστροφής ή ανάκλησης των πτερύγιων, οδηγούν τον συντελεστή ισχύος να κυμαίνεται τελικά γύρω στο 30-45%. Τέλος στο Σχήμα 4.4 φαίνεται η διαφορά που υπάρχει ανά ταχύτητα ανέμου μεταξύ της αιολικής ισχύς που διατίθεται (available) και αυτής που τελικά αξιοποιείται (extracted). Σαν συντελεστής είναι ανάλογος της ταχύτητας του ανέμου, ωστόσο παρουσιάζει ορισμένες ιδιαιτερότητες σε πολύ μικρές ή πολύ μεγάλες τιμές [19]. Παρεμφερή αναφορά σε αυτή τη συμπεριφορά θα γίνει στην επόμενη ενότητα, όπου μελετάται η καμπύλη ισχύος μιας τυπικής ανεμογεννήτριας οριζοντίου άξονα.



Σχήμα 4.4: Καμπύλες συντελεστή ισχύος [19].

4.4.4 Καμπύλη Ισχύος Ανεμογεννήτριας

Η καμπύλη ισχύος μιας τυπικής ανεμογεννήτριας μας παρέχει πολύ σημαντικές πληροφορίες. Αφορούν τον τρόπο λειτουργίας αυτής [19]. Πιο συγκεκριμένα, παρουσιάζει την σχέση που υπάρχει ανάμεσα στην μέση ταχύτητα ανέμου \bar{u} και την πραγματική ηλεκτρική ενέργεια που έχει παραχθεί από την ανεμογεννήτρια καθώς αξιοποιεί τη ταχύτητα αυτή. Με την παρατήρηση της καμπύλης ισχύος μιας ανεμογεννήτριας μπορούμε να εξάγουμε συμπεράσματα για την αεροδυναμική φύση της και την απόδοση που έχει στην μετατροπή και παραγωγή ηλεκτρικής ενέργειας συνολικά.



Σχήμα 4.5: Τυπική καμπύλη ισχύος [19].

Μια τυπική καμπύλη ισχύος δίδεται από το Σχήμα 4.5. Με μια πρώτη ματιά, παρατηρείται ότι, όταν ο αποπνέων άνεμος είναι πολύ ασθενής είτε πολύ δυνατός, η ανεμογεννήτρια δεν λειτουργεί. Επίσης έντονα σημειωμένες είναι κάποιες από τις ταχύτητες πνοής του ανέμου (ταχύτητα εισόδου, ονομαστική ταχύτητα και ταχύτητα αποκοπής). Η λειτουργία της ξεκινάει όταν προσεγγιστεί ή ξεπεραστεί μια σχετικά μικρή ταχύτητα ανέμου. Αυτή η ταχύτητα ονομάζεται ταχύτητα εισόδου (cut-in speed) και ανάλογα το σχεδιασμό της εκάστοτε ανεμογεννήτριας ποικίλει. Συνήθως η τιμή της κυμαίνεται στα 3-5 m/s. Από το σημείο αυτό και έως ένα σημείο παράγεται αξιόλογη ποσότητα ηλεκτρικής ενέργειας η οποία αυξάνεται, όσο αυξάνεται η ένταση του ανέμου. Σαν φθάσει ο άνεμος να πνέει με ταχύτητα ίδια ή μεγαλύτερη της ονομαστικής (12-14 m/s), η τιμή της παραγόμενης ηλεκτρικής ισχύος παραμένει αμετάβλητη και ίδια με αυτήν που παράγεται όταν πνέει άνεμος ταχύτητας ίσης με την ονομαστική. Βρισκόμαστε στο επιθυμητό σημείο από άποψη τεχνική και οικονομική, καθώς παρά την μεταβλητότητα της πνοής του αέρα, παράγεται σταθερή ποσότητα ηλεκτρικής ενέργειας με τη λιγότερη δυνατή φύρα. Τέλος, εάν ο άνεμος γίνει θυελλώδης και ξεπεράσει μια σχετικά μεγάλη ταχύτητα ανέμου, την ταχύτητα αποκοπής (γύρω στα 25 m/s), η ανεμογεννήτρια παύει τη λειτουργία της. Αυτό συμβαίνει για λόγους ασφαλείας, καθώς ο ρότορας ή τα γρανάζια του κιβωτίου ταχύτητας μπορούν να υποστούν μεγάλη ζημιά [18].

Συνοπτικά, υπάρχουν τέσσερις περιοχές λειτουργίας της ανεμογεννήτριας, οι οποίες φαίνονται στον Πίνακα 4.1. Η κάθε μια ορίζεται με βάση τη σχέση που υπάρχει ανάμεσα στη παραγόμενη ηλεκτρική ισχύ και την ταχύτητα πνοής του ανέμου [18]. Σημειώνεται ότι ως V_I λογίζεται η ταχύτητα εισόδου, ως V_R η ονομαστική και ως V_0 η ταχύτητα αποκοπής ή εξόδου.

Περιοχή	Ταχύτητα αέρα	Παραγόμενη ηλεκτρική ισχύς
1	$[0, V_I)$	Μηδενική, καθώς ο άνεμος είναι ανίκανος να περιστρέψει τα πτερύγια.
2	$[V_I, V_R)$	Αυξάνεται, όσο αυξάνεται η ταχύτητα του αποπνέων ανέμου.
3	$[V_R, V_O)$	Παράγεται σταθερά η μεγαλύτερη δυνατή ηλεκτρική ισχύς.
4	$> V_O$	Μηδενική, αφού παύει τη λειτουργία της για λόγους ασφαλείας.

Πίνακας 4.1: Περιοχές λειτουργίας ανεμογεννήτριας [18].

4.5 Αιολικά Πάρκα

Στις μέρες μας η αιολική παραγωγή γνωρίζει μεγάλη ανάπτυξη. Έχουν αναπτυχθεί και χρησιμοποιούνται ανεμογεννήτριες διαφόρων ειδών και μεγεθών. Αναλόγως το ποσό της ενέργειας που απαιτεί μια εφαρμογή, χρησιμοποιούνται αντίστοιχα οι κατάλληλες μηχανές σε μέγεθος και πλήθος. Για τοπικές ανάγκες ή μικρής κλίμακας βιομηχανικές χρήσεις σε περιοχές μακριά από αστικά κέντρα χρησιμοποιούνται μεμονωμένα μικρές ανεμογεννήτριες. Στη περίπτωση, όμως, που στόχος είναι η παραγωγή μεγάλων ποσοτήτων ηλεκτρικής ενέργειας, με σκοπό την τροφοδοσία του τοπικού δικτύου ηλεκτρικής ενέργειας, το τοπίο αλλάζει άρδην. Πιο μεγάλες και πιο πολλές μηχανές τοποθετούνται συγκεντρωτικά σε μια περιοχή και εγκαθίστανται. Αυτές οι ομάδες ανεμογεννητριών αποτελούν τα γνωστά και κατακριτέα από μια μερίδα κόσμου αιολικά πάρκα αν και φέρουν ορισμένα πλεονεκτήματα έναντι των μεμονωμένων εγκατεστημένων αιολικών μηχανών. Αρχικά, δε ρυπαίνουν την ατμόσφαιρα καθώς παράγουν ρεύμα χωρίς την καύση κάποιου ορυκτού καυσίμου. Επιπλέον, η εγκατάσταση, η λειτουργία και η συντήρηση μιας ομάδας ανεμογεννητριών γίνεται με μεγαλύτερη ευκολία. Τέλος, οι απώλειες στη μεταδιδόμενη ισχύ είναι πολύ λιγότερες αφού μπορεί να μετασχηματίζεται το ρεύμα σε πιο υψηλή τάση.

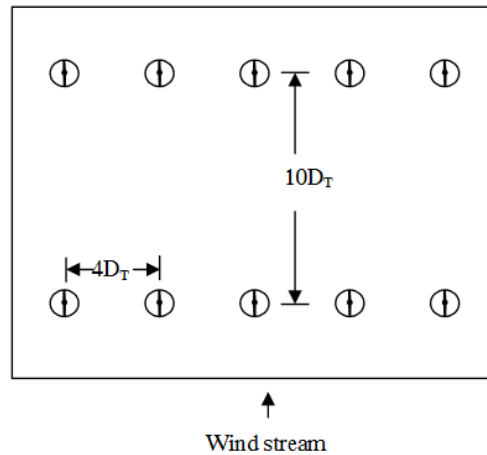
4.5.1 Σχεδίαση και Ανάπτυξη

Η σχεδίαση και ανάπτυξη ενός αιολικού πάρκου απαιτεί την ολοκλήρωση συγκεκριμένων βημάτων που προηγούνται [18]. Πρώτο βήμα προς την κατεύθυνση αυτή είναι η εύρεση του χώρου όπου θα γίνει η εγκατάσταση του πάρκου. Γενικά κατάλληλες θεωρούνται οι περιοχές οι οποίες έχουν τουλάχιστον 5 m/s μέση ετήσια ταχύτητα αέρα. Τα μέσα που χρησιμοποιούνται για να βρεθεί μια τέτοια περιοχή είναι ανεμολογικά δεδομένα που έχουν συλλεχθεί από παραπλήσιους μετεωρολογικούς σταθμούς ή εξαχθεί από τη μελέτη αναιμικών χαρτών. Μόλις βρεθεί μια τέτοια περιοχή, κατάλληλα υπολογιστικά μοντέλα εφαρμόζονται με σκοπό να εκτιμηθεί το αιολικό δυναμικό της ανά διάφορες χρονικές περιόδους. Σε επόμενη φάση διενεργείται μια προσπάθεια λεπτομερής κατανόησης της συμπεριφοράς του ανέμου σε αυτή τη περιοχή. Για αυτό το λόγο γίνονται πιο αναλυτικές μετρήσεις και για μεγαλύτερο χρονικό διάστημα. Για παράδειγμα, μετρήσεις της ταχύτητας του ανέμου στο ύψος της πλήμνης και στη θέση που υποθετικά μπορεί να τοποθετηθεί μια ανεμογεννήτρια πραγματοποιούνται σε διαφορετικές χρονικές περιόδους. Γίνονται ουσιαστικά προσομοιώσεις, ώστε να επαληθευτεί και να εκτιμη-

θεί ο επιθυμητός βαθμός απόδοσης του πάρκου. Σε αυτές λαμβάνονται υπόψιν χαρακτηριστικά διαφόρων τύπων ανεμογεννητριών, ώστε να βρεθεί η βέλτιστη επιλογή για τη συγκεκριμένη τοποθεσία και να εξακριβωθεί το μέγεθος της συνολικής ισχύς που θα εγκατασταθεί.

Βέβαια στη παρούσα φάση δεν εξετάζονται μόνο τεχνικά κριτήρια που αφορούν τη τοποθεσία του πάρκου αλλά και πάσης φύσεως ιδιαιτερότητες. Είτε η μορφολογία του εδάφους είτε η απουσία κατάλληλων δρόμων ή δομών για τη διανομή της παραγόμενης ενέργειας μπορεί να καταστήσουν ασύμφορη τελικά την υλοποίηση ενός τέτοιου έργου. Οπότε, τέτοιοι παράγοντες λαμβάνονται πάντα υπόψιν στην οικονομική μελέτη του έργου όπως επίσης και το κόστος των εξαρτημάτων των μηχανών που θα εγκατασταθούν. Σε επόμενη φάση και αφού έχει εξασφαλιστεί η οικονομική βιωσιμότητα του έργου, σειρά λαμβάνει η αποδοχή του περιβαλλοντικά και κοινωνικά από την κοινωνία στην οποία εντάσσεται γεωγραφικά ο χώρος στον οποίον θα ανεγερθεί το πάρκο αυτό. Συνήθως οι περιοχές, που εκμεταλλεύονται τέτοια έργα ΑΠΕ, είναι περιοχές μακριά από αστικά κέντρα με πλούσια φυσική ομορφιά, που φιλοξενούν άγρια πανίδα και οι ζωές των πολιτών είναι στενά συνδεδεμένες με τη φύση. Έτσι δημιουργούνται θέματα οπτικής και ακουστικής ενόχλησης στους κατοίκους, θέματα ενδημικής ενόχλησης για τη πανίδα της περιοχής και θέματα που αφορούν την επικείμενη οικολογική καταστροφή της περιοχής. Γίνονται λοιπόν εκτενείς συζητήσεις και διαβουλεύσεις με τους αρμόδιους φορείς ώστε να παρθούν οι σχετικές άδειες. Χωρίς την άδεια αυτών και των πολιτών, είναι πολύ πιθανών σε μετέπειτα στάδια να δημιουργηθούν εμπόδια ανυπέρβλητα, ικανά να αφήσουν ανολοκλήρωτο ένα τέτοιο έργο. Έτσι εάν σε κάποια τοποθεσία δεν μπορούν να ξεπεραστούν τα κοινωνικό-περιβαλλοντικά προβλήματα, η τοποθεσία απορρίπτεται.

Η τοποθεσία που επιλέγεται τελικά, πληρεί τα προαναφερόμενα κριτήρια (οικονομικά, τεχνικά, κοινωνικά, περιβαλλοντικά) κατά το βέλτιστον. Το τελικό βήμα που απομένει είναι η επίσημη σύνταξη της μελέτης για το έργο. Εκεί καταγράφονται όλα τα στοιχεία που καθιστούν αξιόλογη τη τοποθεσία για ανάπτυξη αιολικού πάρκου, ο οικονομικός προϋπολογισμός του έργου, τα σχέδια εγκατάστασης και οι υφιστάμενοι περιορισμοί που υπάρχουν για τη κατασκευή του έργου. Στη συνέχεια υποβάλετε στην αρμόδια αρχή και κατόπιν έγκριση της ξεκινά η κατασκευή και χωροθέτηση των ανεμογεννητριών. Συνήθως τοποθετούνται ανά σειρές η μια δίπλα στην άλλη. Κάθε σειρά απέχει $10D_T$, ενώ η απόσταση που υπάρχει ανάμεσα σε διαδοχικές ανεμογεννήτριες στην ίδια σειρά είναι ίση με $4D_T$. Με τον όρο D_T νοείται μια απόσταση ίση με το μήκος της διαμέτρου του ρότορα (ήτοι, δηλαδή, το διπλάσιο από το μήκος του πτερυγίου της Α/Γ). Αυτές οι μεγάλες αποστάσεις ανάμεσα στις σειρές αφήνονται, ώστε ο άνεμος να προλαβαίνει να αναχτά την ένταση του μετά από κάθε σειρά ανεμογεννητριών που διαπερνά. Πιο μικρές αποστάσεις αφήνονται μεταξύ ανεμογεννητριών της ίδια σειράς, έτσι ώστε οι στροβιλισμοί που αναπτύσσονται λόγω της περιστροφής των πτερυγίων να μην επηρεάζουν πολύ την απόδοση των παρακείμενων μηχανών. Στο Σχήμα 4.6 απεικονίζεται ένας τυπικός τρόπος χωροθέτησης μηχανών σε ένα αιολικό πάρκο.



Σχήμα 4.6: Τυπική χωροθέτηση ανεμογεννητριών σε αιολικό πάρκο [18].

4.5.2 Ενσωμάτωση στο Δίκτυο Ηλεκτρικής Ενέργειας

Δεν είναι παρά λίγα τα χρόνια που η συνεισφορά των αιολικών στη παραγωγή ηλεκτρικής ενέργειας έχει πάψει να είναι άνευ σημασίας για το δίκτυο που τροφοδοτεί την κοινωνία με ηλεκτρικό ρεύμα. Ο ρυθμός διεξόδου τέτοιων έργων ΑΠΕ στην αγορά είναι ακόμα αργός. Παρόλα αυτά δείχνει να ακολουθεί σταθερά ανοδική τάση. Πλέον, κατέχουν σημαντικό μερίδιο στη γραμμή παραγωγής όλων των Ευρωπαϊκών χωρών. Υπάρχουν χώρες, όπως η Ελλάδα, όπου το 35% της εγχώριας παραγωγής ηλεκτρικού ρεύματος οφείλεται στην αξιοποίηση του αιολικού δυναμικού, ενώ χώρες όπως η Δανία προσεγγίζουν το 80%. Για αυτό το λόγο η ενσωμάτωση τους στο δίκτυο ηλεκτρικής ενέργειας αποτελεί μια πολύ σημαντική διαδικασία η οποία χρήζει μεγάλης προσοχής και λεπτομέρειας.

Πολύ σοβαρά λαμβάνεται υπόψη η ποιότητα και ο τρόπος με τον οποίο η παραγόμενη από το αιολικό πάρκο ηλεκτρική ενέργεια διοχετεύεται στο δίκτυο. Αρχικά, πρέπει η συχνότητα αυτού να διατηρείται στα επίπεδα των 50 Hz. Επίσης καθώς ένα αιολικό πάρκο δεν παράγει σταθερά ούτε προσφέρει τα ίδια ποσά ηλεκτρικής ενέργειας, το δίκτυο πρέπει να είναι σχεδιασμένο έτσι ώστε να αντέχει αυτές τις μεταβολές. Περιπτώσεις υπερφόρτωσης ή αστάθειας μπορούν να φέρουν βλάβη ή διακοπή στη λειτουργία του πάρκου.

Η γραμμή παραγωγής πλέον αποτελεί ένα υβριδικό σύστημα. Σε αυτό συμμετέχουν αιολικά πάρκα ή υδροηλεκτρικά εργοστάσια ή φωτοβολταϊκά πάρκα, έργα με μεταβλητή παραγωγή μαζί με έργα με σχετικά σταθερή παραγωγή όπως είναι οι μονάδες συμπαραγωγής (μονάδες όπου γίνεται καύση φυσικού αερίου). Κάποιες από αυτές λειτουργούν για να παράξουν το φορτίο (βασικό) που θα καλύψει τη ζήτηση, ενώ οι υπόλοιπες είναι εν αναμονή λειτουργίας σε περίπτωση που χρειασθεί πρόσθετη παραγωγή ηλεκτρικής ενέργειας. Συνηθίζεται να προσφέρουν πρώτα την παραγόμενη ηλεκτρική ενέργεια τους προκειμένου να καλυφθεί η ζήτηση, οι μονάδες παραγωγής που έχουν υψηλό κεφάλαιο επένδυσης και ελάχιστα έξοδα κατόπιν στη λειτουργία τους (π.χ αιολικά πάρκα, υδροηλεκτρικά). Στον αντίποδα μονάδες παραγωγής με μικρό επενδυτικό κεφάλαιο αλλά υψηλά λειτουργικά κόστη καλύπτουν την εάν υπάρξει επιπλέον ζήτηση (π.χ μονάδες λιγνίτη). Αυτό συμβαίνει για ξεκάθαρα περιβαλλοντικούς και

οικονομικούς λόγους.

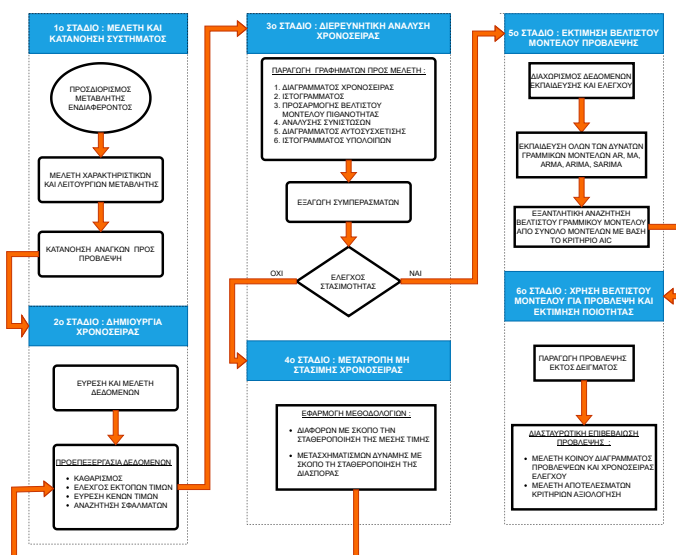
Η μη κάλυψη της ζήτησης από την γραμμή παραγωγής μπορεί να δημιουργήσει τεράστια προβλήματα στην κοινωνία. Για να καλύπτεται πάντα πρέπει να υπάρχει πάντα σωστός συντονισμός και προγραμματισμός των φορτίων μεταξύ των γραμμών παραγωγής. Για να γίνει αυτό χρειάζεται να μπορούν να προβλεφθούν αξιόπιστα τόσο η ζήτηση όσο και τα φορτία που μπορεί να προσφέρει η κάθε γραμμή παραγωγής, για να γίνει ο κατάλληλος προγραμματισμός.

Από την μια πλευρά, υπάρχει η ζήτηση η οποία ανά χρονική στιγμή μέσα στην ημέρα ή ανά εποχές αλλάζει. Τις στιγμές που η κοινωνία είναι πιο παραγωγική ή έχει ανάγκες που η κάλυψη τους απαιτεί τεράστια ποσά ενέργειας, η ζήτηση αυξάνεται. Αντίθετα άλλες χρονικές στιγμές όπως τις νυκτερινές ώρες η ζήτηση για ενέργεια μειώνεται καθώς η κοινωνία είναι πιο αδρανής. Είναι μια μεταβλητή η οποία επηρεάζεται από τις δραστηριότητες του ανθρώπου, το κλίμα (αυξάνεται σε περίπτωση καύσωνα λόγω των κλιματιστικών) ή τις εποχιακές γιορτές. Για την πρόβλεψη της δηλαδή χρησιμοποιούνται δεδομένα καιρού, μοντελοποιούνται τάσεις για κάποιες ημέρες ανά εποχές ή ώρες ανά ημέρα και προσομοιώνονται γραμμικά μοντέλα πρόβλεψης επί ιστορικών δεδομένων της ζήτησης για να γίνει μια ακριβή πρόβλεψη της ζήτησης ανά ώρα κάθε ημέρα. Παρόμοιες τεχνικές χρησιμοποιούνται κατά την προσπάθεια πρόβλεψης της παραγωγής. Δεδομένα καιρού και ιστορικά δεδομένα παραγωγής τροφοδοτούν τα μοντέλα πρόβλεψης που κάνουν τις προβλέψεις.

Η ζήτηση ορίζει τη παραγωγή. Μια παραγωγή η οποία αποτελείται από πολλούς διαφορετικούς παραγωγούς ενέργειας. Αυτοί πρέπει ανά ώρα να μπορούν να γνωρίζουν τι μέρος της ζητούμενης ενέργειας μπορούν να καλύψουν. Οι εταιρίες που έχουν στην κατοχή τους έργα ΑΠΕ έχουν το πρώτο λόγο στο Χρηματιστήριο της Ενέργειας. Υποβάλουν πρώτοι τις προσφορές τους για την κάλυψη της ανά ώρας ζήτησης (day-ahead market). Είναι οι παραγωγοί που μπορούν να προσφέρουν το μέγιστο των δυνατοτήτων τους σε ενέργεια αφού έχουν χαμηλά λειτουργικά κόστη σε σχέση με άλλους παραγωγούς (λιγνίτη). Προσφέρουν σε μηδενική τιμή την παραγόμενη ενέργεια τους. Στη συνέχεια κάποιο μέρος της ζήτησης καλύπτουν παραγωγοί με λίγο πιο υψηλές τιμές και τέλος αποδεκτές γίνονται προσφορές και σε πιο υψηλές τιμές ακόμα έως ότου καλυφθεί και η τελευταία MWh. Η τιμή με την οποία πωλήθηκε η τελευταία ποσότητα που κάλυψε την ζήτηση ονομάζεται οριακή τιμή συστήματος και είναι η τιμή με την οποία πληρώνονται τελικά όλοι οι παραγωγοί. Αυτή είναι η κύρια ημερήσια αγορά ενέργειας, αλλά υπάρχει και η αγορά εξισορρόπησης η οποία λειτουργεί καθ' όλη τη διάρκεια της ημέρας. Αυτή ενεργοποιείται εάν υπάρξει λάθος στην πρόβλεψη της ζήτησης. Τότε ζητάτε περισσότερη ενέργεια την οποία σπεύδουν να καλύψουν παραγωγοί με υψηλά λειτουργικά κόστη.

Είναι πολύ σημαντικό να προβλέπεται επακριβώς η ποσότητα ενέργειας που μπορεί να παράγει ένα έργο ΑΠΕ (αιολικό πάρκο). Εάν η πρόβλεψη είναι αξιόπιστη δεν δημιουργείται οικονομική ζημία, τη στιγμή χιόλας που δεν υπάρχει όριο στην τιμή που μπορεί να προσφέρει το προϊόν του ο παραγωγός. Επίσης δεν δημιουργούνται προβλήματα στη κάλυψη της ζήτησης. Καθώς το καύσιμο των ΑΠΕ έργων (ήλιος, αέρας) είναι μεταβλητό, μπορεί να υπάρξουν προβλήματα στην τελική ποσότητα ενέργειας που παράχθηκε με αυτήν που αρχικά είχε εκτιμηθεί. Σε περίπτωση που χρειασθεί μια μονάδα παραγωγής να επέμβει (μονάδες με μηχανές καύσης

ορυκτών καυσίμων), υπάρχει περίπτωση να μην είναι κατάλληλα προετοιμασμένη για να λειτουργήσει. Συνήθως οι μηχανές καύσης της χρειάζονται κάποιες ώρες για να προετοιμαστούν και να λειτουργήσουν. Σε μια τέτοια περίπτωση το δίκτυο δεν μπορεί να καλύψει τη ζήτηση και δημιουργείται μεγάλο πρόβλημα στην ηλεκτροδότηση σημαντικών εγκαταστάσεων. Στο Σχήμα 4.7 αποτυπώνεται το σύστημα ενσωμάτωσης της αιολικής παραγωγής στο ηλεκτρικό δίκτυο.



Σχήμα 4.7: Η διαδικασία διασύνδεσης της αιολικής παραγωγής στο δίκτυο [18].

Κεφάλαιο 5

Θεωρητικό Υπόβαθρο

Κάθε χρονοσειρά αναλύεται διεξοδικά πριν εφαρμόσουμε σε αυτήν κάποιο μοντέλο που την προσομοιώνει. Κατά την ανάλυση της, μελετώνται τα κύρια γνωρίσματα της. Βασική απαίτηση, πριν γίνει η οποιαδήποτε επιλογή μοντέλου προσομοίωσης, είναι η ύπαρξη στασιμότητας. Για αυτό το λόγο σε μια ενότητα αναλύεται η έννοια της στασιμότητας και της αυτοσυσχέτισης. Υπάρχουν φορές όμως που διαπιστώνεται ότι μια χρονοσειρά δεν είναι στάσιμη. Στην περίπτωση αυτή, αρχικά γίνεται μια προσπάθεια κατανόησης των συνιστωσών που περιέχει μια μη στάσιμη χρονοσειρά. Έπειτα παρατίθενται χρήσιμοι μέθοδοι που εφαρμόζονται, ώστε να απαλειφθούν οι οι συνιστώσες αυτές και η χρονοσειρά να μετατραπεί σε στάσιμη.

5.1 Χρονοσειρές

Στις μέρες μας η συλλογή και αξιοποίηση ιστορικών δεδομένων είναι καθοριστικής σημασίας σε σχεδόν όλους τους τομείς της βιομηχανίας. Γίνεται στατιστική ανάλυση επί αυτών και κατόπιν είτε εκπαιδεύονται γραμμικά μοντέλα είτε εκπαιδεύονται μοντέλα που βασίζονται σε αλγόριθμους μηχανικής μάθησης για την πρόβλεψη μελλοντικών τιμών τους. Αυτές οι ενέργειες πραγματοποιούνται κυρίως για δύο λόγους. Αρχικά για να εξαχθεί πολύτιμη γνώση με τη βοήθεια της οποίας μπορεί να περιγραφεί και να εξηγηθεί η συμπεριφορά της εκάστοτε υπό μελέτης μεταβλητής ή συστήματος και εν συνεχεία για να καθίσταται δυνατή η πρόβλεψη της συμπεριφορά της βραχυπρόθεσμα ή μακροπρόθεσμα. Έτσι λοιπόν, πολύ σημαντικές αποφάσεις μπορούν να παρθούν είτε διοικητικές είτε κοινωνικές και πολύπλοκα συστήματα να λειτουργούν συντονισμένα και αρμονικά μεταξύ τους χωρίς να προκαλούν σύγχυση στη καθημερινότητα μας.

Τα δεδομένα που συλλέγονται και αξιοποιούνται τις περισσότερες φορές έχουν την μορφή χρονοσειρών (time series). Μια χρονοσειρά αποτελεί ένα σύνολο δειγματοληπτημένων ποσοτικών παρατηρήσεων (observations) μιας ενδιαφέρουσας μεταβλητής. Η δειγματοληψία αυτών γίνεται σειριακά στη διάρκεια ενός χρονικού διαστήματος και σε χρόνους που απέχουν ίσα χρονικά. Η συχνότητα με την οποία καταγράφονται οι τιμές αυτής της μεταβλητής ποικίλουν ανάλογα τη χρήση της και τις εφαρμογές της. Για παράδειγμα, η χρονοσειρά που περιέχει τις τιμές του εθνικού ακαθάριστου προϊόντος μιας χώρας έχει ετήσια συχνότητα δειγματο-

ληψίας, ενώ μια χρονοσειρά που περιέχει τις καταγεγραμμένες μέσες ταχύτητες ανέμου μιας περιοχής μπορεί να έχει συχνότητα δειγματοληψίας του δευτερολέπτου ή ακόμη και των δέκα λεπτών [20]. Με την ανάλυση των χρονοσειρών γίνεται δυνατό να μελετηθεί και να περιγραφεί ο τρόπος με τον οποίο εξελίσσεται η υπό μελέτη μεταβλητή ανά ίσα χρονικά διαστήματα στη διάρκεια μιας μικρής ή μεγάλης περιόδου. Με την κατανόηση του τρόπου εξέλιξης του συστήματος, γίνεται τελικά δυνατή η επιλογή ενός κατάλληλου μοντέλου προσομοίωσης και πρόβλεψης.

Η δειγματοληψία των τιμών της μεταβλητής ενδιαφέροντος της χρονοσειράς γίνεται είτε συνεχόμενα είτε σε διακριτές χρονικές στιγμές. Με βάση αυτό το κριτήριο μπορούν να χωριστούν σε χρονοσειρές συνεχούς χρόνου (continuous time) και σε χρονοσειρές διακριτού χρόνου (discrete time) [3]. Εάν αυτές καταγράφονται ανά διακριτά διαστήματα χρόνου κατατάσσονται στις διακριτού χρόνου, ενώ εάν η καταγραφή τους πραγματοποιείται συνεχώς για ένα χρονικό διάστημα εντάσσονται στην κατηγορία των χρονοσειρών συνεχούς χρόνου [4]. Πολύ σημαντικό να αναφερθεί ότι οι περισσότερες μελετημένες χρονοσειρές ανήκουν στη πρώτη κατηγορία. Ακόμη σημαντικό κρίνεται να αναφερθεί ότι μπορεί να υπάρξει παραπάνω κατηγοριοποίηση ανάλογα το πλήθος των μεταβλητών που αναλύονται από την εκάστοτε χρονοσειρά. Στη βάση αυτή, διακρίνονται σε μονοδιάστατες (univariate) και πολυδιάστατες (multivariate). Στην παρούσα εργασία θα ασχοληθούμε μόνο με μονοδιάστατες χρονοσειρές διακριτού χρόνου.

Αυτή η μεταβλητή συστήματος, όπως προείπαμε, δειγματοληπτείται με σταθερό βήμα (χρόνος δειγματοληψίας). Έχουμε, δηλαδή, ανά κάθε χρονική στιγμή t μια πραγματική τιμή x_t του παρατηρούμενου μεγέθους X . Το σύνολο των τιμών αυτών μέσα σε μια χρονική περίοδο n , όπου κάθε μια διακρίνεται από τον χρόνο t στον οποίο δειγματοληπτήθηκε, αποτελεί τη χρονοσειρά της μεταβλητής του συστήματος [28]

$$\{X_t\}_{t=1}^n = \{x_1, x_2, \dots, x_n\}. \quad (5.1)$$

Μια ακολουθία σαν αυτή, καλείται επίσης ως στοχαστική διαδικασία διακριτού χρόνου $\{X_t : t \in \mathcal{T}\}$, όπου \mathcal{T} είναι το σύνολο των χρονικών στιγμών κατά τις οποίες ορίζεται αυτή η ακολουθία τυχαίων μεταβλητών X_t . Οι τιμές της χρονοσειράς αποτελούν τις πραγματοποιήσεις αυτής της τυχαίας διαδικασίας. Από εδώ και πέρα με τον όρο χρονοσειρά θα νοούνται τόσο μια στοχαστικής διαδικασίας $\{X_t : t \in \mathcal{T}\}$ όσο και η πραγματοποίηση αυτής $\{X_t\}$ [10, 28].

5.2 Ανάλυση Στάσιμων Χρονοσειρών

Για να καταστεί δυνατή η μελέτη και η πρόβλεψη μιας χρονοσειράς, είναι απαραίτητη η ύπαρξη στασιμότητας. Με απλά λόγια, στάσιμη θεωρείται μια χρονοσειρά, όταν με την εξέλιξη του χρόνου δεν παρατηρείται ότι επηρεάζονται σημαντικές στατιστικές ιδιότητες της. Αυτό σημαίνει ότι η πιθανολογική συμπεριφορά της δεν αλλάζει με το πέρασμα του χρόνου αλλά παραμένει σταθερή και έτσι μπορεί να καταστεί κατά κάποιον τρόπο προβλέψιμη. Υπάρχουν δύο είδη στασιμότητας τα οποία θα εξετασθούν, η αυστηρή και η ασθενής [10].

Μια χρονοσειρά είναι αυστηρά στάσιμη αν μόνο και μόνο αν η πιθανολογική συμπεριφορά της δεν μεταβάλλεται με τον χρόνο. Αυτό σημαίνει ότι μια συλλογή τιμών $\{x_{t_1}, x_{t_2}, x_{t_3}, \dots, x_{t_n}\}$ κατανέμεται παρόμοια με την μετατοπισμένη χρονικά συλλογή τιμών της $\{x_{t_1+k}, x_{t_2+k}, \dots, x_{t_n+k}\}$. Δηλαδή, η πιθανότητα να εμφανιστεί η τιμή της παρατήρησης x_{t_1+k} είναι ίδια με αυτήν της x_{t_1} , όπως υποδεικνύεται στην Εξ. (5.2) [23]

$$\mathcal{P}\{x_{t_1} \leq l_1, x_{t_2} \leq l_2, \dots, x_{t_n} \leq l_n\} = \mathcal{P}\{x_{t_1+k} \leq l_1, x_{t_2+k} \leq l_2, \dots, x_{t_n+k} \leq l_n\}, \quad (5.2)$$

όπου οι t_1, t_2, \dots, t_n αποτελούν όλες τις χρονικές στιγμές, οι l_1, l_2, \dots, l_n αποτελούν τις σταθερές τιμές πιθανότητας εμφάνισης της κάθε παρατήρησης και $k \in (0, \pm 1, \pm 2, \dots)$ είναι οι χρονικές μετατοπίσεις [26].

Μια χρονοσειρά $\{X_t : t \in \mathcal{T}\}$ είναι ασθενώς στάσιμη εάν οι ροπές πρώτης και δεύτερης τάξης δεν εξαρτώνται από τον χρόνο. Ειδικότερα, πρέπει να ισχύουν ταυτόχρονα τρεις συνθήκες που αφορούν την μέση τιμή, την διασπορά και την αυτοδιασπορά της χρονοσειράς. Η πρώτη συνθήκη αναφέρει ότι η μέση τιμή της πρέπει να είναι σταθερή και ανεξάρτητη του χρόνου t , δηλαδή,

$$m(t) = \mathbb{E}[X_t] = \mu_t = \mu \quad \forall t \in \mathcal{T}. \quad (5.3)$$

Η δεύτερη συνθήκη που πρέπει να ικανοποιείται, απαιτεί η συνάρτηση αυτοδιασποράς $\gamma(t_1, t_2)$ των τιμών μιας χρονοσειράς, να εξαρτάται από την διαφορά $\tau = t_1 - t_2$, που ονομάζεται χρονική υστέρηση (lag) τ των χρονικών στιγμών t_1, t_2 . Πράγματι, σε αυτή τη περίπτωση, εάν $t_1 = t$ και $t_2 = \tau - t$ ισχύει ότι

$$\begin{aligned} \gamma(t_1, t_2) &= \text{Cov}(X_{t_1}, X_{t_2}) = \mathbb{E}[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] \quad \forall t_1, t_2 \in \mathcal{T} \\ \iff \gamma(t_1, \tau - t_1) &= \gamma(t, \tau - t) = \gamma(\tau). \end{aligned} \quad (5.4)$$

Τέλος, υπάρχει μια ακόμη συνθήκη η οποία ικανοποιείται ως απόρροια των δύο προηγούμενων. Η διασπορά $\text{Var}(X_t)$ πρέπει πάντα να είναι πεπερασμένη και να μην υπολογίζεται ως συνάρτηση του χρόνου t παρά της υστέρησης τ . Στην Εξ. (5.5), εξηγείται μαθηματικά αυτή η παράθεση

$$\sigma^2(t) = \text{Var}(X_t) = \text{Cov}(X_t, X_t) = \gamma(0) = \sigma^2. \quad (5.5)$$

Στη πράξη, όταν γίνεται έλεγχος της στασιμότητας μιας χρονοσειράς, συνήθως ελέγχουμε τη μέση τιμή και τη διασπορά μιας χρονοσειράς. Είναι λάθος, καθώς δεν λαμβάνουμε υπόψιν την συνθήκη που υπάρχει για την αυτοδιασπορά της. Ακόμη η αυστηρή στασιμότητα επιτυγχάνεται σε πολύ λίγες περιπτώσεις καθώς στις περισσότερες πραγματικές χρονοσειρές εμφανίζονται οι συνιστώσες της τάσης και της περιοδικότητας. Αποτέλεσμα της ύπαρξης τους είναι η μεταβολή των τιμών της μέσης τιμής, της διασποράς και της αυτοσυσχέτισης που απορρέει από την Εξ. (5.4) της αυτοδιασποράς. Έτσι, από δω και πέρα, μια αναφορά επί της στασιμότητας μιας χρονοσειράς, θα υπονοεί την ύπαρξη ασθενούς στασιμότητας σε αυτήν.

5.2.1 Αυτοσυσχέτιση

Με την συνάρτηση αυτοσυσχέτισης (ACF) μιας στάσιμης χρονοσειράς μπορεί να υπολογιστεί ο βαθμός συσχέτισης που υπάρχει ανάμεσα σε μια παρατήρηση x_{t_1} και μια παρατήρηση x_{t_2} . Η γραμμική αυτοσυσχέτιση των δύο παρατηρήσεων που μεταξύ τους παρεμβάλλεται μια χρονική απόσταση ίση με τ υπολογίζεται όπως φαίνεται στην Εξ. (5.6) [23, 10]

$$\begin{aligned} \text{Cor}(X_{t_1}, X_{t_2}) &= \text{Cor}(X_{t_1}, X_{\tau-t_1}) = \text{Cor}(X_t, X_{\tau-t}) \\ \iff \text{Cor}(X_t, X_{\tau-t}) &= \frac{\text{Cov}(X_t, X_{\tau-t})}{\sqrt{\text{Var}(X_t)\text{Var}(\tau-t)}} \\ \iff \rho(\tau) &= \frac{\gamma(\tau)}{\gamma(0)}. \end{aligned} \quad (5.6)$$

Οι τιμές που μπορεί να προκύψουν κατά τον υπολογισμό της γραμμικής αυτοσυσχέτισης είναι μεταξύ -1 και 1 . Στον Πίνακα 5.1 παρουσιάζονται τα συμπεράσματα που μπορούμε να εξαγάγουμε για τη συσχέτιση δύο στοιχείων μιας χρονοσειράς ανάλογα τη προκύπτουσα τιμή αυτοσυσχέτισης.

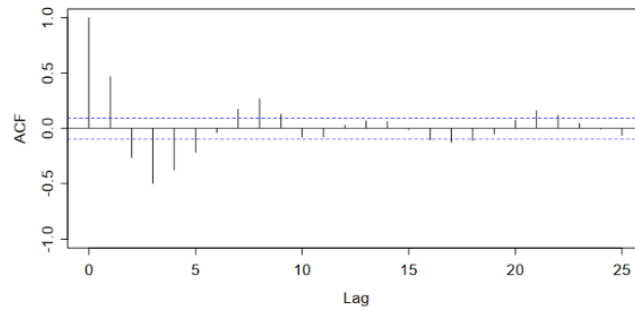
ρ	Τιμή συσχέτισης
(0,+1)	Θετική, η οποία γίνεται πιο ισχυρή όσο η τιμή της πλησιάζει το 1.
0	Απουσία συσχέτισης
(-1,0)	Αρνητική, η οποία γίνεται πιο ισχυρή όσο η τιμή της πλησιάζει το -1 .
+1 ή -1	Ισχυρή συσχέτιση κατά το βέλτιστον

Πίνακας 5.1: Περιπτώσεις συσχέτισης

Η παρατήρηση του Πίνακα 5.1 οδηγεί σε ένα πολύ σημαντικό συμπέρασμα. Το συμπέρασμα αυτό αφορά τη γραμμική συσχέτιση μεταξύ δύο παρατηρήσεων x_{t_1} , x_{t_2} της χρονοσειράς, όπου η x_{t_2} έπεται της x_{t_1} κατά χρονική υστέρηση τ . Όσο η τιμή της γραμμικής συσχέτισης μεταξύ τους πλησιάζει το 1 ή το -1 , τόσο πιο εύστοχη μπορεί να καταστεί μια πρόβλεψη της x_{t_2} από την x_{t_1} . Αντίθετα, όσο πλησιάζει το μηδέν, αυτή καθίσταται όλο και λιγότερο εφικτή. Μια γραμμική σχέση της μορφής $x_{t_2} = \alpha_1 x_{t_1} + \alpha_0$, όπου το α_1 παίρνει θετικές τιμές αν η τιμή της αυτοσυσχέτισης είναι θετική και αρνητικές αν είναι αρνητική, μπορεί να εκφράσει μαθηματικά τη πρόβλεψη αυτή.

Ένα πολύ χρήσιμο εργαλείο εκτίμησης αυτοσυσχέτισης μεταξύ παρατηρήσεων που απέχουν κατά μια χρονική υστέρηση τ (lag) μεταξύ τους είναι το διάγραμμα αυτοσυσχέτισης (correlogram). Το Σχήμα 5.1 που ακολουθεί, είναι ένα κλασσικό παράδειγμα απεικόνισης της αυτοσυσχέτισης που υπάρχει μεταξύ των παρατηρήσεων μιας χρονοσειράς.

Το πρώτο πράγμα που παρατηρείται στο διάγραμμα αυτό είναι οι άξονες του. Στον οριζόντιο βρισκουμε τις υστερήσεις (lags), δηλαδή, την χρονική απόσταση που απέχουν δύο στοιχεία μιας χρονοσειράς. Αυτή η απόσταση είναι ανάλογη του χρόνου δειγματοληψίας της χρονοσειράς. Για παράδειγμα εάν εξετάζονται μηνιαία δεδομένα παραγωγής ενέργειας, τότε αποτυπώνεται η



Σχήμα 5.1: Διάγραμμα αυτοσυσχέτισης μιας χρονοσειράς που οι τιμές τις αποτυπώνουν τα ύψη των κυμάτων του νερού σε χιλιοστά δειγματοληπτημένα ανά 0.1 δευτερόλεπτα [10].

αυτοσυσχέτιση μεταξύ παρατηρήσεων της χρονοσειράς που απέχουν $\tau = 1, 2, 3, \dots, n$ μήνες κάθε φορά. Η τιμή της αυτοσυσχέτισης προσδίδεται από την κορυφή της κάθε αιχμής (spike) και τις τιμές του άξονα y στις οποίες αντιστοιχούν. Όπως προαναφέρθηκε αλλά και φαίνεται, κυμαίνονται από -1 έως 1. Οι αιχμές τοποθετούνται κατά μήκος του οριζοντίου άξονα ανά lag. Το πρώτο lag που έχει τιμή μηδέν, έχει πάντα αυτοσυσχέτιση ίση με 1. Στο παράδειγμα με τα μηνιαία δεδομένα αυτό μεταφράζεται σε αυτοσυσχέτιση μεταξύ πρώτου μήνα και πρώτου μήνα. Αντίστοιχα, γίνεται η ερμηνεία για τα υπόλοιπα.

Πέρα από τις τιμές αυτοσυσχέτισης, στο διάγραμμα διακρίνονται οι ζώνες εμπιστοσύνης. Είναι οι περιοχές που ορίζονται με βάση τις διακεκομμένες γραμμές στο Σχήμα 5.1 και τον οριζόντιο άξονα. Στις περιπτώσεις που μελετώνται στάσιμες χρονοσειρές, θεωρείται ότι όσες τιμές αυτοσυσχέτισης $\rho(\tau)$, όπου τ είναι το εκάστοτε lag, είναι εντός των διαστημάτων εμπιστοσύνης, μπορεί να θεωρηθούν μηδέν ή ότι μόνο κατά τύχη δεν είναι μηδέν. Αντιθέτως, όσες εξέρχουν των διαστημάτων αυτών, είναι σίγουρα διάφορες του μηδενός [10] και φανερώνουν έναν βαθμό συσχέτισης.

Έλεγχος Στασιμότητας

Ένα διάγραμμα αυτοσυσχέτισης, δεν χρησιμοποιείται μόνο για την αποτίμηση της συσχέτισης μεταξύ παρατηρήσεων μιας στάσιμης χρονοσειράς. Πολλές φορές εφαρμόζεται σε μη στάσιμες χρονοσειρές. Όταν συμβαίνει αυτό, δύναται η δυνατότητα εξαγωγής πληροφοριών χρήσιμων για την κατανόηση των ιδιοτήτων της χρονοσειράς. Γίνεται, δηλαδή, να εντοπισθεί η ύπαρξη τάσης ή περιοδικότητας μέσα στη χρονοσειρά. Για παράδειγμα, στο Σχήμα 5.1 ευσταθεί ο ισχυρισμός περί ύπαρξης σχετικής περιοδικότητας, δεδομένου της συμπεριφοράς που επαναλαμβάνεται ανά 8 lags. Σε κάθε διάγραμμα ACF, η εμφάνιση σταθερής ή φθίνουσας περιοδικότητας, είναι ένα ξεκάθαρο σημάδι μη στασιμότητας της χρονοσειράς. Συνεπώς, το διάγραμμα αυτοσυσχέτισης μπορεί να χρησιμοποιηθεί επίσης ως εργαλείο ελέγχου της στασιμότητας μιας χρονοσειράς.

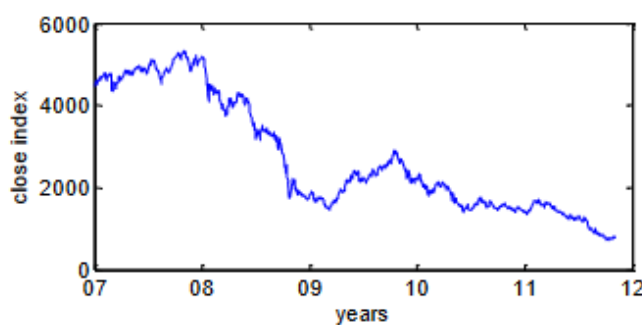
Ένας άλλος τρόπος για να ελεγχθεί η ύπαρξη ή όχι στασιμότητας στη χρονοσειρά, είναι η διενέργεια του ελέγχου Augmented Dickey–Fuller [22]. Πρόκειται για ένα στατιστικό έλεγχο ο οποίος χρησιμοποιείται ευρέως προκειμένου να διαπιστωθεί εάν μια χρονοσειρά είναι στάσιμη.

Ο έλεγχος αυτός αποτελεί έναν έλεγχο στατιστικής σημαντικότητας (statistical significance test) που εφαρμόζεται σε μονοδιάστατες χρονοσειρές. Σαν έλεγχος ανήκει στην οικογένεια των ελέγχων μοναδιαίας ρίζας (unit root test) και αποτελεί μια παραλλαγή του ελέγχου Δικτεψ–Φυλλερ. Ελέγχονται δύο υποθέσεις, η μηδενική υπόθεση H_0 η οποία εκφράζει το ενδεχόμενο να υπάρχει μοναδιαία ρίζα και η συμπληρωματική αυτής H_1 . Κατά τη διενέργεια του εξεταζόμενου ελέγχου, υπολογίζονται ορισμένα σημαντικά στατιστικά στοιχεία και οι κρίσιμες τιμές (critical values). Με βάση τις τιμές αυτών απορρίπτεται ή όχι η μηδενική υπόθεση H_0 . Εάν απορριφθεί, τότε η χρονοσειρά είναι στάσιμη. Σε διαφορετική περίπτωση η ύπαρξη μοναδιαίας ρίζας υποδεικνύει ότι η τάση που υπάρχει στη χρονοσειρά αποσταθεροποιεί την μέση τιμή της χρονοσειράς στη πάροδο του χρόνου και την καθιστά μη στάσιμη.

5.3 Ανάλυση Μη Στάσιμων Χρονοσειρών

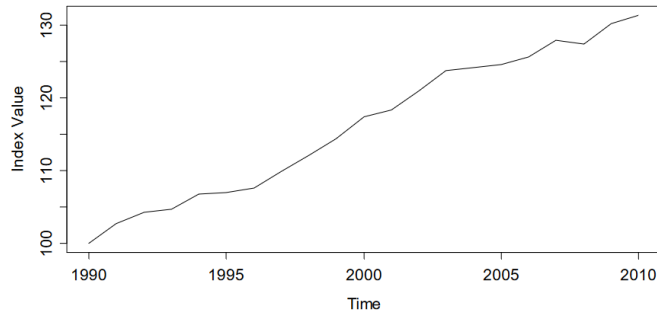
Η μη στασιμότητα είναι το πιο σημαντικό πρόβλημα που εγείρεται στη διαδικασία παραγωγής προβλέψεων μελλοντικών τιμών της. Μια χρονοσειρά είναι μη στάσιμη όταν παρατηρηθεί ότι τα στατιστικά χαρακτηριστικά της μεταβάλλονται με τη πάροδο του χρόνου. Η δημιουργία στατιστικών διαγραμμάτων και γραφικών παραστάσεων μιας τέτοιας χρονοσειράς συνήθως αποκαλύπτει την ύπαρξη κυρίως των παρακάτω συνιστωσών:

- Τάση (Trend): Μια χρονοσειρά μπορεί να εμφανίζει είτε καθοριστική είτε στοχαστική τάση. Μια καθοριστική (deterministic) τάση είναι δυνατό να αναπαρασταθεί από μια γνωστή συνάρτηση του χρόνου και οφείλεται σε εξωτερικούς παράγοντες. Ουσιαστικά, παρατηρείται ότι η μέση τιμή της χρονοσειράς μεταβάλλεται μακροχρόνια είτε με ανοδική είτε με καθοδική πορεία [10, 26]. Στη περίπτωση που δεν γίνεται να περιγραφεί από κάποια ήδη γνωστή συνάρτηση του χρόνου, δηλαδή, δεν επέρχονται αλλαγές με κάποιο συγκεκριμένο τρόπο, τότε λέμε ότι έχουμε στοχαστική τάση. Στα Σχήματα 5.3 και 5.2 δίνονται δύο χαρακτηριστικά παραδείγματα ντερμινιστικής και στοχαστικής τάσης.



Σχήμα 5.2: Χρονοσειρά του δείκτη Χρηματιστηρίου Αξιών Αθηνών (XAA) κατά το κλείσιμο (στοχαστική τάση) [26].

- Εποχικότητα (Seasonality): Εποχικότητα παρουσιάζει μια χρονοσειρά εάν οι τιμές του παρατηρούμενου μεγέθους μεταβάλλονται σταθερά ανά κάποια χρονικά διαστήματα (εποχή του χρόνου ή διάστημα ωρών στη μέρα). Είναι δηλαδή μια συνιστώσα η οποία



Σχήμα 5.3: Χρονοσειρά δείκτη κυκλοφοριακού συμφορισμού (ντετερμινιστική τάση) [10].

εμφανίζεται σταθερά και περιοδικά στη χρονοσειρά και την μεταβάλλει. Χαρακτηριστικό παράδειγμα αποτελούν οι εποχικές συνιστώσες σε δεδομένα αεροπορικών κρατήσεων που έχουν δειγματοληπτηθεί ανά μήνα για μια περίοδο πολλών χρόνων ή σε ωριαία δεδομένα ζήτησης ενέργειας που είναι διαθέσιμα για πολλές ημέρες και εμφανίζουν ημερήσιο μοτίβο [10]. Στα μεν πρώτα είναι πολύ πιθανό σε κάθε χρονιά να εμφανίζονται πρόσκαιρες αυξήσεις στο πλήθος των κρατήσεων τους καλοκαιρινούς μήνες των διακοπών και στα δε δεύτερα τις πρωινές ώρες όταν η κοινωνία δραστηριοποιείται περισσότερο. Φαίνεται να έχει κάθε παρατηρούμενο μέγεθος "εποχιακή" περιοδικότητα στην εξέλιξη του, δηλαδή, μια κυκλικά επαναλαμβανόμενη εξέλιξη των τιμών της χρονοσειράς [28].

Μια χρονοσειρά για να χαρακτηριστεί μη στάσιμη δεν προϋποθέτει την παράλληλη παρουσία των δύο αυτών συνιστωσών. Μπορεί είτε να παρουσιάζει μόνο τάση είτε μόνο εποχικότητα είτε και τα δύο μαζί. Μια χρονοσειρά X_t μπορεί να περιγραφεί ως το άθροισμα συνιστωσών καθοριστικής τάσης m_t , εποχικότητας s_t και του όρου που εκφράζει τα στάσιμα υπολείμματα της χρονοσειράς R_t (θόρυβος) [10, 28]. Το γενικό μοντέλο που περιγράφηκε ονομάζεται αθροιστικό και εκφράζεται ως:

$$X_t = m_t + s_t + R_t. \quad (5.7)$$

Σε κάθε περίπτωση οφείλεται να απαλείφονται οι συνιστώσες της τάσης και της περιοδικότητας προκειμένου να εργαζόμαστε με στάσιμες χρονοσειρές.

5.4 Μετατροπή μη Στάσιμων Χρονοσειρών σε Στάσιμες

Μια χρονοσειρά η οποία παρουσιάζει τάση ή περιοδικότητα αποτελεί μια μη στάσιμη χρονοσειρά. Για να μετατραπεί σε στάσιμη αρκεί να αντιμετωπιστούν αυτές και να γίνουν οι απαραίτητοι μετασχηματισμοί. Στη περίπτωση μιας πρόβλεψης, δεν χρειάζεται να μελετηθούν οι αλλαγές που επέρχονται στην χρονοσειρά λόγω της επικρατούσας τάσης ή περιοδικότητας. Θεωρείται ότι είναι το αποτέλεσμα εξωγενών παραγόντων [26]. Βέβαια υπάρχουν άλλες εφαρμογές, όπου η μελέτη και ο προσδιορισμός της τάσης ή της περιοδικότητας της χρονοσειράς είναι ο στόχος της ανάλυσης. Σε μια τέτοια περίπτωση δεν γίνεται προσπάθεια απαλοιφής παρά

εκτίμησης αυτών των συνιστωσών. Στην περίπτωση που μας ενδιαφέρει η πρόβλεψη, υπάρχουν δύο τρόποι με τους οποίους μπορούμε να εργαστούμε για την αντιμετώπιση αυτών των συνιστωσών. Η μια επιλογή είναι, να απαλειφθούν πρώτα οι δυο συνιστώσες και έπειτα αφού ολοκληρωθεί η πρόβλεψη να προστεθούν ξανά. Η άλλη επιλογή είναι, να γίνει η πρόβλεψη, συνεκτιμώντας την τάση και την περιοδικότητα στην όλη διαδικασία.

Στις παρακάτω υπό ενότητες αναπτύσσονται περιεκτικά μεθοδολογίες αφαίρεσης τάσης και περιοδικότητας που αξιοποιήθηκαν στη ανάλυση της παρούσας εργασίας και σκοπό είχαν την σταθεροποίηση των στατιστικών μέτρων της χρονοσειράς. Για την μαθηματική περιγραφή τους χρησιμοποιείται ως δεδομένο το γενικό μοντέλο που περιγράφει η Εξ. (5.7).

5.4.1 Απαλοιφή Τάσης

Όταν στην υπό μελέτη χρονοσειρά η συνιστώσα της τάσης είναι ο λόγος που αυτή δεν εμφανίζει στασιμότητα και παράλληλα η επίδραση της εποχικής περιοδικότητας στην χρονοσειρά μπορεί να θεωρηθεί αμελητέα, η X_t μπορεί να περιγραφεί όπως υποδεικνύεται από την Εξ. (5.8) [28, 26]

$$X_t = m_t + R_t \text{ όταν } t = 1, 2, \dots, n. \quad (5.8)$$

Εάν αυτή η τάση m_t είναι καθοριστική, δηλαδή, είναι δηλαδή γραμμική ή εκθετική [28], πρώτα εκτιμάται και έπειτα απαλείφεται. Ο προσδιορισμός της μπορεί να γίνει με τη χρήση ενός γραμμικού πολυωνύμου τάξης p [26]

$$m_t = a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p \quad (5.9)$$

και κατόπιν να αφαιρεθεί (detrend).

Εάν η τάση όμως είναι στοχαστική, τότε μόνο απαλείφεται. Η φύση μιας ήδη γνωστής συνάρτησης χρόνου $f(t)$ όπως περιγράφεται από την Εξ. (5.9) δεν μπορεί να λειτουργήσει κατάλληλα σε όλο το μήκος της χρονοσειράς. Στη περίπτωση αυτή εφαρμόζονται μετασχηματισμοί Box-Cox ώστε να σταθεροποιηθεί η διασπορά της X_t και χρησιμοποιούνται μετασχηματισμοί διαφορών πρώτης ή δεύτερης τάξης και διαδικασίες κινούμενου μέσου όρου ώστε να αφαιρεθεί η τάση.

Η μέθοδος των διαφορών πρώτης τάξης οδηγεί στην αφαίρεση της στοχαστικής τάσης από τη χρονοσειρά. Ουσιαστικά το αποτέλεσμα της μεθόδου αυτής είναι η παραγωγή μιας καινούργιας χρονοσειράς Y_t από την X_t , που όμως αυτή έχει τάση ίση του μηδενός [26]. Ανά διαδοχικούς παρατηρήσεις της X_t υπολογίζονται οι διαφορές, έχουμε δηλαδή τον εξής μετασχηματισμό της $\{X_t\}_{t=1}^n$

$$Y_t = \nabla X_t = X_t - X_{t-1} \quad (5.10)$$

για $t = 2, 3, \dots, n$, όπου με ∇ συμβολίζεται ο τελεστής διαφοράς πρώτης τάξης. Σε περίπτωση που δεν επιτευχθεί στασιμότητα με τη μέθοδο των πρώτων διαφορών, τότε εφαρμόζουμε ξανά την ίδια μέθοδο και έτσι ουσιαστικά χρησιμοποιούμε διαφορές δεύτερης τάξης ώστε η τάση να μένει σταθερή στη πάροδο του χρόνου [26]

$$\begin{aligned}
Y_t &= \nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) \\
&= X_t - 2X_{t-1} + X_{t-2} \text{ για } t = 3, 4, \dots, n.
\end{aligned}
\tag{5.11}$$

Βέβαια τις περισσότερες φορές χρησιμοποιούμε διαφορές πρώτης τάξης. Ανά μικρά χρονικά διαστήματα, είναι δυνατό να εκφραστεί η τάση γραμμικά, δηλαδή, τοπικά να είναι της μορφής $m_t = a_0 + a_1 t$. Με συνέπεια, αν σε μια χρονοσειρά X_t τάσης m_t πάρουμε πρώτες διαφορές [26]

$$\begin{aligned}
\nabla Y_t &= Y_t - Y_{t-1} \\
&= X_t + m_t - X_{t-1} - m_{t-1},
\end{aligned}
\tag{5.12}$$

τότε η νέα τάση που διαμορφώνεται από την Εξ. (5.12) ισούται με

$$m_t - m_{t-1} = a_0 + a_1 t - a_0 - a_1(t-1) = +a_1. \tag{5.13}$$

Σύμφωνα με την Εξ. (5.13), καταλήγουμε εν τέλη σε μια τάση η οποία είναι ανεξάρτητη του χρόνου t που τελικά έχει εξαλειφθεί.

Πολλές φορές προτιμούμε να μην εφαρμόσουμε απευθείας τη μέθοδο των διαφορών. Επιλέγουμε να εφαρμόσουμε κάποιον μετασχηματισμό δύναμης στη χρονοσειρά μας. Ένας τέτοιου είδους μετασχηματισμός ονομάζεται Box-Cox και ο ορισμός του δίδεται μέσω της Εξ. (5.14) [26]

$$Y_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln(X_t), & \lambda = 0. \end{cases}
\tag{5.14}$$

Επίσης για συγκεκριμένες τιμές της παραμέτρου λ ο μετασχηματισμός Box-Cox δίνει στη νέα μετασχηματισμένη χρονοσειρά Y_t κάποια από τις γνωστές μαθηματικές μορφές, όπως αυτές φαίνονται στο Πίνακα 5.2.

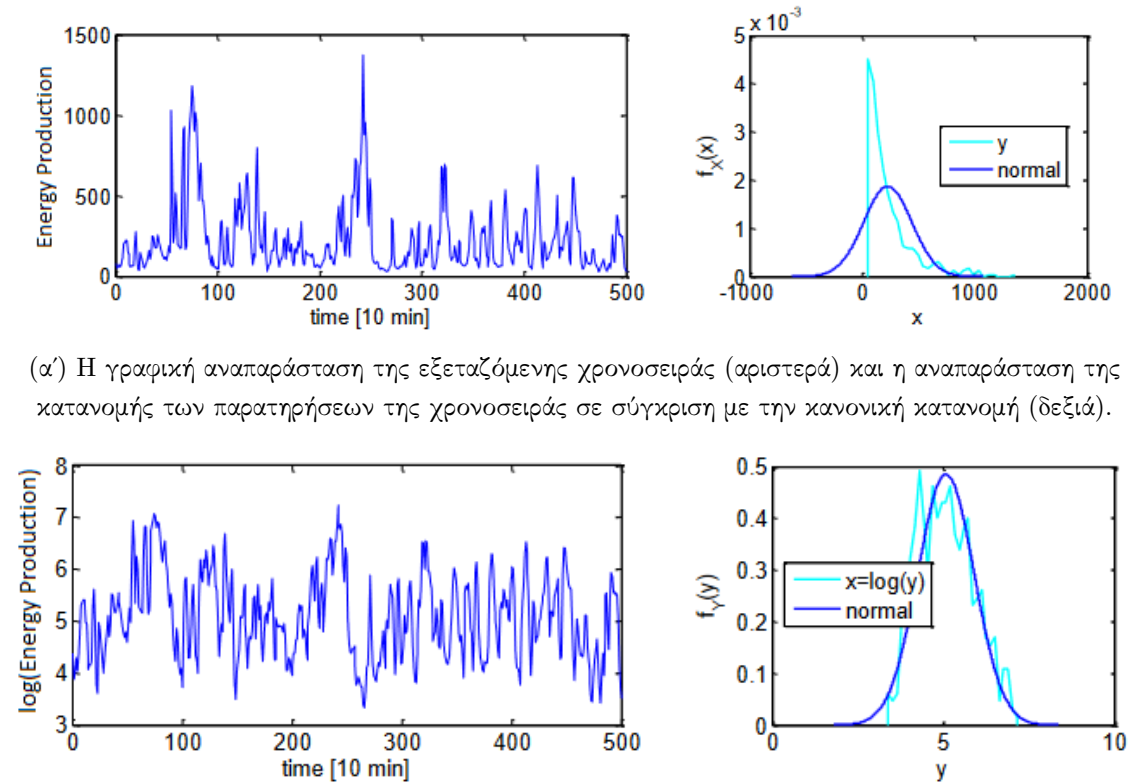
λ	Μετασχηματισμένη χρονοσειρά Y_t
-3	$\frac{1}{X_t^3}$
-2	$\frac{1}{X_t^2}$
-1	$\frac{1}{X_t}$
0.5	$\frac{1}{\sqrt{X_t}}$
0	$\ln(X_t)$
0.5	$\sqrt{X_t}$
2	X_t^2
3	X_t^3

Πίνακας 5.2: Μετασχηματισμοί δυναμονομικής (αλγεβρικής) μορφής Box-Cox [26].

Με την επιλογή του μετασχηματισμού των δεδομένων με τον τρόπο αυτό επιτυγχάνονται δύο πράγματα [10]. Αρχικά, σταθεροποιείται η διασπορά της χρονοσειράς X_t . Είναι ένα σημαντικό πλεονέκτημα που κατέχει σαν μετασχηματισμός, αφού για παράδειγμα, η διαφύση των δεδομένων εξαλείφει την τάση αλλά δεν σταθεροποιεί τη διασπορά. Η τάση η οποία υπάρχει μεταβάλλει τη μέση τιμή της χρονοσειράς όσο αυτή εξελίσσεται, με αποτέλεσμα η διασπορά να μην παραμένει σταθερή. Ο μετασχηματισμός αυτός τελικά οδηγεί στην σταθεροποίηση της, εάν βρεθεί κατάλληλος παράγοντας λ . Πάντα υπολογίζεται με βάση τα εξεταζόμενα δεδομένα. Μια σημαντική παρατήρηση, πρέπει να γίνει σε αυτό το σημείο. Εάν η βέλτιστη τιμή της παραμέτρου λ υπολογιστεί εντός του διαστήματος $[-0.3, 0.3]$, τότε επιλέγεται ο λογαριθμικός μετασχηματισμός (με βάση \exp) έναντι του μετασχηματισμού δύναμης για την τιμή λ που βρέθηκε στο προαναφερθέν διάστημα. Η επιλογή αυτή γίνεται για λόγους πρακτικούς. Ουσιαστικά, το κέρδος είναι πολύ μικρό για να υπερκεράσει τις ιδιότητες και την ευκολία που φέρει στην ερμηνεία της μια λογαριθμική συνάρτηση. Η εφαρμογή αυτού του μετασχηματισμού μπορεί να κάνει τα δεδομένα να προσεγγίζουν καλύτερα τη κανονική κατανομή αν και αυτό δεν επιτυγχάνεται πάντα. Το Σχήμα 5.4 δίνεται ώστε να προβληθούν τα θετικά αποτελέσματα που μπορεί να έχει η εφαρμογή ενός μετασχηματισμού Box Cox σε μια μη στάσιμη χρονοσειρά.

Αρχικά παρατηρείται το αριστερό διάγραμμα του Σχήματος 5.4 α'. Σε αυτό αναπαρίσταται η υπό μελέτη χρονοσειρά. Τα δεδομένα της αρχικής χρονοσειράς έχουν θετικές τιμές και φαίνεται να μην υπάρχει στασιμότητα σε αυτήν. Σε διαφορετικές χρονικές στιγμές παρουσιάζονται διαφορετικού εύρους διακυμάνσεις. Χαρακτηριστικές είναι οι απότομες και μεγάλες αιχμές που εμφανίζονται σε κάποιες χρονικές στιγμές. Η διασπορά της, δηλαδή, φαίνεται να αλλάζει στον χρόνο. Προκειμένου να επιβεβαιωθεί η ύπαρξη μη στασιμότητας, δημιουργείται και απεικονίζεται το δεξιό γράφημα του Σχήματος 5.4 α'. Σε αυτό, είναι δυνατόν, να παρατηρηθεί η οριακή κατανομή της χρονοσειράς σε σχέση με την κανονική κατανομή την οποία έπρεπε να ακολουθεί ή να την πλησιάζει. Εάν την προσέγγιζε, αυτό θα ήταν αρκετό για να υποστηρίξουμε τον ισχυρισμό περί ύπαρξης στασιμότητας αν και δεν θα ήταν βέβαιο. Στο γράφημα αυτό, η οριακή κατανομή των δεδομένων δεν πλησιάζει την κανονική, παρά είναι δεξιά στρεβλωμένη (right skewed) με μια μεγάλη ουρά, να αντικατοπτρίζει τις απότομες μεγάλες διακυμάνσεις. Συνεπώς, μελετάται μια μη στάσιμη χρονοσειρά. Για το λόγο αυτό, εφαρμόζεται μετασχηματισμός (Box-Cox). Πιο συγκεκριμένα, επιλέγεται ο λογαριθμικός ($\lambda = 0$), λόγω των όσων προαναφέρθηκαν (θετικές τιμές, ασταθής διασπορά, απότομες διακυμάνσεις, μη κανονική κατανομή). Από τα διαγράμματα στο Σχήμα 5.4 β', φαίνεται ότι οι εμφανίσεις απότομων διακυμάνσεων μειώθηκαν σε μεγάλο βαθμό, ενώ επίσης η οριακή κατανομή της φαίνεται να πλησιάζει τη κανονική καλύτερα από πριν. Ωστόσο, δεν παύουν να υπάρχουν αποκλίσεις που δεν την καταστούν κανονική τελικά.

Αφού εργαστούμε με τη μετασχηματισμένη χρονοσειρά και παράξουμε κάποιο αποτέλεσμα (π.χ. προβλέψεις) με βάση αυτή, έπειτα θα χρειαστεί να επαναφέρουμε τα μετασχηματισμένα δεδομένα (τιμές των παρατηρήσεων της χρονοσειράς) στην αρχική κλίμακα τους. Στη περίπτωση αυτή πρέπει πάντα να λαμβάνεται υπόψη ότι ένας απλός ανεστραμμένος μετασχηματισμός θα περιέχει μεροληψία στο αποτέλεσμα του. Ουσιαστικά, διατηρεί την διάμεση τιμή αλλά όχι την αναμενόμενη τιμή της κατανομής των στοιχείων που προβλέφθηκαν. Με τη διόρθωσή της,



(α') Η γραφική αναπαράσταση της εξεταζόμενης χρονοσειράς (αριστερά) και η αναπαράσταση της κατανομής των παρατηρήσεων της χρονοσειράς σε σύγκριση με την κανονική κατανομή (δεξιά).

(β') Η γραφική αναπαράσταση της μετασχηματισμένης χρονοσειράς (εφαρμόστηκε Box-Cox μετασχηματισμός με $\lambda=0$ (δηλαδή λογαριθμικός μετασχηματισμός των τιμών» (αριστερά) και η αναπαράσταση της κατανομής των παρατηρήσεων της μετασχηματισμένης χρονοσειράς σε σύγκριση με την κανονική κατανομή (δεξιά).

Σχήμα 5.4: Εφαρμογή Μετασχηματισμού Box-Cox σε μη στάσιμη χρονοσειρά [26].

δίνεται η αναμενόμενη τιμή της πρόβλεψης. Στην περίπτωση του μετασχηματισμού δύναμης (Box-Cox), το έργο αυτό επιτελείται με την εφαρμογή της Εξ. (5.15) [10]

$$(\lambda \hat{x}_t + 1)^{1/\lambda} \cdot \left(1 + \frac{\hat{\sigma}_h^2(1-\lambda)}{2(\lambda \hat{x}_t + 1)^2} \right), \quad (5.15)$$

όπου ως $\hat{\sigma}_h^2$ ορίζεται η διασπορά των προβλέψεων που έχουν γίνει h βήματα μπροστά από την τελευταία αξιοποιήσιμη παρατήρηση της χρονοσειράς x_t . Στη περίπτωση του λογαριθμικού μετασχηματισμού, η Εξ. (5.15) γίνεται:

$$\exp(\hat{x}_t) \cdot \left(1 + \frac{\hat{\sigma}_h^2}{2} \right)$$

5.4.2 Απαλοιφή Περιοδικότητας

Υπάρχει περίπτωση η επίτευξη στασιμότητας σε μια χρονοσειρά να εμποδίζεται από την ύπαρξη περιοδικότητας σε αυτήν. Σε μια τέτοια περίπτωση, όπου επίσης θεωρούμε, ότι, δεν εμφανίζεται τάση, η χρονοσειρά μπορεί να εκφραστεί με βάση τη φόρμουλα της Εξ. (5.16)

$$X_t = s_t + R_t, ; \text{ για } t = 1, 2, \dots, n. \quad (5.16)$$

Στη παραπάνω σχέση ως s_t λογίζεται η συνιστώσα της περιοδικότητας στη χρονοσειρά και ως R_t τα στάσιμα υπόλοιπά της. Για την περιοδική συνάρτηση, παίρνεται ως δεδομένη η γνώση της περιόδου d που την χαρακτηρίζει. Η περιοδικότητα μπορεί, επίσης, να μεταφραστεί σε εποχικότητα, αν η περίοδος της χαρακτηρίζεται από κάποιο σύννηθες χρονικό διάστημα (π.χ έναν χρόνο, ένα τρίμηνο ή μία ημέρα).

Στην πιο απλή περίπτωση, η εκτίμηση της περιοδικότητας μπορεί να καταστεί δυνατή με τη βοήθεια μιας γνωστής συνάρτησης του χρόνου $f(t)$. Κύριο γνώρισμα της μορφής μιας τέτοιας συνάρτησης είναι φυσικά η περίοδος της. Συναρτήσεις όπως αυτές του ημίτονου ή του συνημιτόνου ίσως φανούν χρήσιμες σε μεμονωμένες περιπτώσεις όπου η περιοδικότητα σαν συνάρτηση μπορεί να αποτυπωθεί με συγκεκριμένη μορφή.

Ένας ακόμη απλός τρόπος εκτίμησης της περιοδικής συνάρτησης s_t (με περίοδο d) μιας χρονοσειράς, μπορεί να γίνει μέσω μιας άλλης περιοδικής συνάρτησης s_i , για $i = 1, 2, \dots, d$. Η κατασκευή της, προϋποθέτει τον υπολογισμό των επιμέρων στοιχείων της \hat{s}_i . Ο υπολογισμός τους γίνεται με βάση την Εξ. (5.17) [26]

$$\hat{s}_i = \frac{1}{k} \sum_{j=1}^k x_{i+jd}, \quad (5.17)$$

όπου k είναι το πλήθος των περιόδων της χρονοσειράς με παρατηρήσεις x_1, x_2, \dots, x_n . Είναι προφανές, ότι εάν εξαχθούν οι μέσοι όροι στοιχείων της περιοδικής συνάρτησης, είναι δυνατόν να κατασκευασθεί αρχικά η βοηθητική περιοδική συνάρτηση s_i και στη συνέχεια με χρήση της να αφαιρεθεί η συνιστώσα της περιοδικότητας από τη χρονοσειρά μας.

Υπάρχει περίπτωση, βέβαια, να μην παρουσιάζει ιδιαίτερο ενδιαφέρον η εκτίμηση της όσο η εξάλειψή της. Υπό αυτές τις συνθήκες, συνηθίζεται να εφαρμόζονται διαφορές d τάξης (d - differencing). Με την χρήση της μεθόδου αυτής παράγεται μια νέα χρονοσειρά Y_t από την X_t , που δεν έχει περιοδικότητα. Ανά παρατηρήσεις που απέχουν χρονικά όσο η περίοδος d που χαρακτηρίζει την X_t υπολογίζονται οι διαφορές. Έτσι για χρονική υστέρηση d , η $\{X_t\}_{t=1}^n$ μετασχηματίζεται σε [26]

$$Y_t = \nabla_d X_t = X_t - X_{t-d}. \quad (5.18)$$

Κεφάλαιο 6

Πρόβλεψη Χρονοσειρών

6.1 Χρονοσειρές Αυτοπαλινδρόμησης AR(p) model

Μια χρονοσειρά θεωρείται αυτοπαλινδρομη τάξεως p όταν κάθε παρατήρηση αυτής X_t καθίσταται δυνατό να εκφραστεί ως ένα άθροισμα του οποίου οι όροι σταθμίζονται από σταθερούς συντελεστές a_0, a_1, \dots, a_p (autoregressive coefficients). Σαν όροι αυτού λογίζονται οι παρελθοντικές εκδοχές αυτής και αριθμούν σε πλήθος p , όσο δηλαδή, το μήκος της καθυστέρησης που οδηγεί στην πρόβλεψη της τρέχουσας τιμής της χρονοσειράς. Η τρέχουσα τιμή της χρονοσειράς δηλαδή μπορεί να εκφραστεί συναρτήσει των παρελθοντικών τιμών της υποδεικνύοντας την εξάρτηση που υπάρχει στην διαμόρφωση της από αυτές. Τέλος, στο άθροισμα αυτό δεν παραλείπεται να συνυπολογιστεί η διαδικασία λευκού θορύβου με μέση τιμή μηδέν και διακύμανση χρονικά ανεξάρτητη η οποία υφίσταται [16, 17, 11]. Αυτή η γραμμική συνάρτηση μπορεί να συμβολιστεί ως $AR(p)$ και δίδεται από την εξίσωση (6.1)

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \epsilon_t. \quad (6.1)$$

όπου με ϵ_t νοείται η διαδικασία του λευκού θορύβου. Πρόκειται ουσιαστικά για μια τυχαία απόκλιση η οποία προστίθεται τελικώς μαζί με τη τιμή της κάθε παρατήρησης που καταγράφηκε σε χρονικό διάστημα p πριν από την τρέχουσα και αποτελεί το τυχαίο μέρος της χρονοσειράς. Βέβαια, δεν πρέπει να παραληφθεί το γεγονός ότι κάθε μια από αυτές τις $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ είναι πολλαπλασιασμένη με μια σταθερά, έναν συντελεστή δηλαδή, ο οποίος καταδεικνύει τον βαθμό επίδρασης αυτής στην υπό μελέτη παρατήρηση x_t .

6.2 Χρονοσειρές Κινούμενου Μέσου MA(q) model

Σε μια χρονοσειρά κινούμενου μέσου η τρέχουσα τιμή της, X_t , μπορεί να οριστεί γραμμικά από ένα άθροισμα που εμπεριέχει όρους που προκύπτουν από τις q προηγούμενες παρατηρήσεις της. Πιο συγκεκριμένα, η εξάρτηση αυτή περιγράφεται από ένα άθροισμα που αποτελείται από μια διαδικασία λευκού θορύβου ϵ_t της τρέχουσας παρατήρησης και επιπλέον q τέτοιων διαδικασιών. Αυτές απορρέουν από τις q στο πλήθος προηγμένες παρατηρήσεις που χρησιμοποιούνται

και παράλληλα είναι σταθμισμένες από έναν παράγοντα θ η κάθε μία. Η σχέση αυτή μπορεί να εκφραστεί μαθηματικά από την Εξίσωση (6.2) :

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (6.2)$$

όπου ως θ λογίζονται οι παράγοντες ουσιαστικά με τους οποίους αποτυπώνεται η επίδραση της κάθε τυχαίας απόκλισης που παρατηρήθηκε στο παρελθόν στον υπολογισμό της τιμής της παρατήρησης X_t που μελετάται την παρούσα στιγμή. Ένα τέτοιου είδους γραμμικό μοντέλο αξιοποιείται για την περιγραφή θορύβων μέσα στη χρονοσειρά που δεν οφείλονται σε μοτίβα ή τάσεις που υφίστανται κατά μήκος όλης της χρονοσειράς. Αξιοποιούνται λοιπόν για την εκτίμηση του στοχαστικού μέρους της χρονοσειράς.

6.3 Αυτοπαλίνδρομες Χρονοσειρές Κινούμενου Μέσου ARMA model(p, q)

Υπάρχουν περιπτώσεις που ορισμένες χρονοσειρές ενώ είναι στάσιμες δεν επαρκεί η περιγραφή τους αποκλειστικά από κάποια διαδικασία MA ή AR . Για να μην καταφύγουμε στην λογική μιας μεγάλης τάξεως διαδικασίας είτε MA είτε AR όπου το λογιστικό κόστος μπορεί να αποτελέσει πρόβλημα, οδηγούμαστε στις χρονοσειρές $ARMA(p, q)$ ή αλλιώς στις μικτές διαδικασίες αυτοπαλινδρόμησης και κινούμενου μέσου τάξεως p και q αντίστοιχα. Είναι στην ουσία ένα γραμμικό μοντέλο το οποίο συνδυάζει τα δύο γραμμικά μοντέλα που περιγράφηκαν στις δυο προηγούμενες υπό ενότητες για να προσομοιώσει τόσο το αιτιοκρατικό όσο και το στοχαστικό τμήμα αυτής. Σε ένα τέτοιο μοντέλο μια παρατήρηση X_t μπορεί να περιγραφεί από την Εξ. (6.3),

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}. \quad (6.3)$$

όπου ως ϵ_t συμβολίζεται ο λευκός θόρυβος που στην ουσία υπάρχει για να υπολογίζεται τυχαία αστοχίες. Στην περίπτωση που $p = 0$ το μοντέλο εξαθλιώνεται και μετατρέπεται σε αμιγώς κινούμενου μέσου ενώ στην περίπτωση που $q = 0$ το μοντέλο πλέον μετατρέπεται σε αυτοπαλινδρούμενο.

6.4 Ολοκληρωμένες χρονοσειρές τάξης d Integrated process (d)

Όταν η χρονοσειρά X που εξετάζεται δεν μπορεί να χαρακτηριστεί στάσιμη, τότε με την εφαρμογή κάποιας τάξεως d διαφορών είναι πολύ πιθανό η χρονοσειρά να παρουσιάσει στοιχεία στασιμότητας. Ονομάζονται ολοκληρωμένες χρονοσειρές τάξης d ή πιο σύντομα μπορούν να καταγραφούν ως $I(d)$. Η μαθηματική σχέση που μπορεί να φανερώσει την αλλαγή αυτή αν X_t η υπό μελέτη χρονοσειρά που δεν παρουσιάζει στασιμότητα και K_t η νέα ολοκληρωμένη

χρονοσειρά φαίνεται από τη Εξ. (6.4)

$$K_t = X_t - X_{t-1} + \epsilon_t \quad (6.4)$$

όπου ϵ_t λογίζεται ο λευκός θόρυβος που προσμετράται για να αποτυπωθεί το μέρος των τυχαίων διακυμάνσεων. Η Εξ. (6.4) δεν αποτελεί τον ορισμό αλλά μια εφαρμογή που καταδεικνύει την εφαρμογή μιας διαφορίσης πρώτης τάξεως.

6.5 Αυτοπαλίνδρομες Ολοκληρωμένες Κινούμενου Μέσου Χρονοσειρές ARIMA model(p, d, q)

Στις περισσότερες περιπτώσεις μια χρονοσειρά όταν βρίσκεται σε κατάσταση στασιμότητας μπορεί να περιγραφεί από ένα αυτοπαλίνδρομο μοντέλο κινούμενου μέσου. Παρόλα αυτά υπάρχει πάντα η πιθανότητα, η χρονοσειρά να μην είναι στάσιμη. Τα ολοκληρωμένα αυτοπαλίνδρομα μοντέλα κινούμενου μέσου χρησιμοποιούνται στη περίπτωση αυτή. Η εξεταζόμενη χρονοσειρά που δεν είναι στάσιμη, προκειμένου να γίνει, εφαρμόζεται σε αυτήν αρχικά μια διαδικασία διαφορίσης τάξης d . Στη συνέχεια, τίθεται σε εφαρμογή το κατάλληλο μοντέλο $ARMA$ τάξεως (p, q) ώστε να περιγράψει την πλέον στάσιμη χρονοσειρά. Η σχέση που περιγράφει το ολοκληρωμένο αυτοπαλίνδρομο μοντέλο κινούμενου μέσου $ARIMA$ δίδεται από την Εξ. (6.5) :

$$Z_t = a_1 K_{t-1} + a_2 K_{t-2} + \dots + a_p K_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}. \quad (6.5)$$

όπου K_t , λογίζεται η αρχική μη στάσιμη χρονοσειρά αφού επήλθε σε αυτή η διαφορίση τάξεως d που την κατέστησε στάσιμη. Εάν για παράδειγμα $d = 1$, τότε $K_t = X_t - X_{t-1}$ όπως άλλωστε ορίζει η Εξ. (6.4). Υπενθυμίζεται ότι p, q αποτελούν τους βαθμούς αυτοπαλινδρόμησης και κινούμενου μέσου αντίστοιχα.

Τα ARIMA χρησιμοποιούνται πιο γενικευμένα και δίνουν τη δυνατότητα περιγραφής διαφορών στοχαστικών διαδικασιών. Κυρίως αναπτύσσονται σαν μέσα πρόβλεψης τιμών της χρονοσειράς που μελετάται. Η διαδικασία με την οποία επιτυγχάνεται το εγχείρημα αυτό ονομάζεται διαδικασία Box - Jenkins και θα μελετηθεί πιο αναλυτικά στη συνέχεια.

6.6 Εποχικές Αυτοπαλίνδρομες Ολοκληρωμένες Κινούμενου Μέσου Χρονοσειρές (SARIMA) seasonal ARIMA model

Τα εποχικά μοντέλα ARIMA, γνωστά ως SARIMA, έρχονται και λύνουν τα προβλήματα αποδοτικότητας που έχουν τα υπόλοιπα γραμμικά μοντέλα όταν η χρονοσειρά που εξετάζεται περιέχει περιοδικότητα πέρα από τάση. Όταν μια χρονοσειρά δεν είναι στάσιμη, τότε είτε η συνιστώσα της τάσης είτε αυτής της περιοδικότητας είτε η συνύπαρξη αυτών είναι οι λόγοι που προκαλείται η κατάσταση αυτή. Με τη διαφορίση της χρονοσειράς αντιμετωπίζεται η τάση αλλά

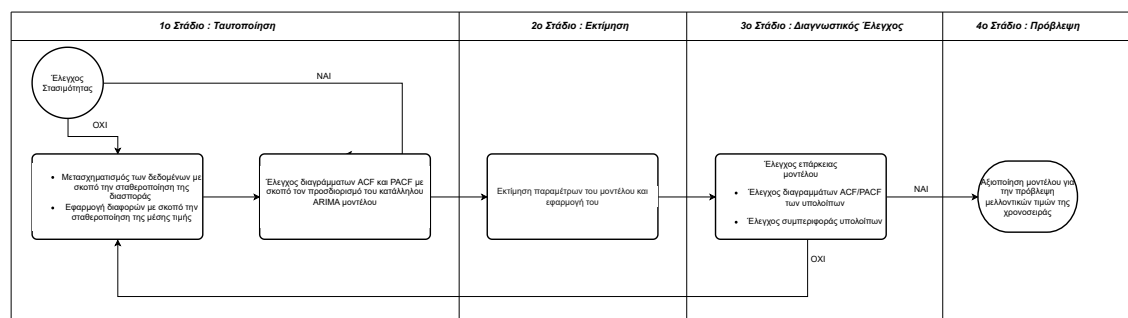
όχι η περιοδικότητα ή αλλιώς η συνιστώσα που επαναλαμβάνεται κάθε k παρατηρήσεις. Για να απαλειφθούν, λοιπόν, χρησιμοποιούνται οι εποχικές διαφορές. Είναι σημαντικό να αναφερθεί ότι ιδιαίτερη προσοχή δίδεται στο πλήθος των διαφορών που εφαρμόζονται τελικά. Αυτές μπορούν να αριθμούν το μέγιστο τις 2. Σε καμία περίπτωση, παρόλα αυτά, η τάξη της εποχικής διαφορίσης δεν θα ξεπεράσει τη πρώτη ($D = 1$). Οι υπόλοιποι εποχικοί όροι SAR και SMA που απαιτούνται για την ολοκληρωμένη εφαρμογή του γραμμικού μοντέλου, προσδιορίζονται χάριν στις τιμές της αυτοσυσχέτισης που αφορούν την εποχική συνιστώσα. Σε περίπτωση που παρουσιάζονται θετικές τιμές, αξιοποιείται ένας όρος εποχικής αυτοπαλινδρόμησης SAR. Σε αντίθετη περίπτωση, γίνεται αξιοποίηση ενός εποχικού όρου κινούμενου μέσου SMA. Πάντως στη γενική του μορφή ένα τέτοιο εποχικό μοντέλο ARIMA μπορεί να εκφραστεί ως :

$$ARIMA(p, d, q)x(P, D, Q)_s \quad (6.6)$$

όπου ως s λογίζεται η εποχικότητα που έχει βρεθεί ότι υπάρχει στη χρονοσειρά, ενώ οι όροι P και Q που υφίστανται προσδιορίζουν το πλήθος των εποχικών όρων SAR και SMA.

6.7 Διαδικασία Box-Jenkins

Η μέθοδος που μελετήθηκε και αξιοποιήθηκε σε επόμενα κεφάλαια όπου εξετάζονται χρονοσειρές παραγωγής ηλεκτρικής ενέργειας ανήκει στους Box-Jenkins. Ο σκοπός που επιτυγχάνεται με την εφαρμογή αυτής είναι ο προσδιορισμός ενός γραμμικού μοντέλου ARIMA το οποίο δύναται να την αναπαραστήσει σε ικανοποιητικό βαθμό. Μάλιστα, ο βαθμός αυτός φανερώνει πολλές φορές και την ικανότητα του μοντέλου αυτού να προβλέψει στη συνέχεια μελλοντικές τιμές της μεταβλητής ενδιαφέροντος της χρονοσειράς. Αποτελείται από τέσσερα στάδια τα οποία απεικονίζονται επιγραμματικά στο Σχήμα 6.1, ενώ η πιο λεπτομερή παρουσίαση τους έπεται στη συνέχεια.



Σχήμα 6.1: Η διαδικασία Box-Jenkins.

6.7.1 Ταυτοποίηση

Πρώτο βήμα στην διαδικασία αυτή αποτελεί η ταυτοποίηση η αλλιώς ο προσδιορισμός του μοντέλου που θα αξιοποιηθεί. Καθορίζονται ουσιαστικά οι τάξεις αυτοπαλινδρόμησης,

διαφόρισης και κινούμενου μέσου του μοντέλου. Πριν, όμως, γίνει αυτό, απαραίτητος κρίνεται ο έλεγχος στασιμότητας της χρονοσειράς. Το κύριο εργαλείο που αξιοποιείται για την ολοκλήρωση της ενέργειας αυτής είναι η ανάγνωση του γραφήματος της συνάρτησης αυτοσυσχέτισης της χρονοσειράς. Στη περίπτωση που φθίνουν γρήγορα προς το μηδέν οι τιμές αυτής, συμπεραίνετε ότι η χρονοσειρά είναι στάσιμη κατά πάσα πιθανότητα. Σε αντίθετη περίπτωση η χρονοσειρά εμφανίζει μη στασιμότητα οπότε και είναι αναγκαίο να μετατραπεί σε στάσιμη. Με την εφαρμογή διαφορών σταθεροποιείται η μέση τιμή της ενώ με την μετατροπή των δεδομένων με κάποιου είδους μετασχηματισμό δύναμης σταθεροποιείται η διακύμανση της.

Όταν η χρονοσειρά είτε είναι στάσιμη είτε επέλθει σε κατάσταση στασιμότητας, γίνεται απλά καθορισμός των τάξεων των όρων p, q . Ο πιο σύννηθες τρόπος για να γίνει αυτό είναι η προσεκτική ανάγνωση των διαγραμμάτων $ACF, PACF$. Στον Πίνακα 6.1 φαίνονται συνοπτικά οι ενδείξεις από τις οποίες εκμαιεύονται αυτές.

Μοντέλο	Ενδείξεις Αυτοσυσχέτισης	Ενδείξεις Μερικής Αυτοσυσχέτισης
Λευκός Θόρυβος	Όταν όλες οι τιμές ισούνται με το μηδέν	Όταν όλες οι τιμές ισούνται με το μηδέν
AR(p)	Όταν οι τιμές προσεγγίζουν το μηδέν με εκθετική μείωση	Μετά από p lags αρχίζουν οι μηδενικές τιμές
MA(q)	Μετά από q lags αρχίζουν οι μηδενικές τιμές	Γεωμετρική μείωση των τιμών μετά την q υστέρηση
ARMA(p, q)	Γεωμετρική μείωση των τιμών της μετά από p υστερήσεις	Μείωση των τιμών της περιοδικά ή γεωμετρικά μετά από q υστερήσεις

Πίνακας 6.1: Ενδείξεις στα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης οι οποίες βοηθούν στον προσδιορισμό των τάξεων των όρων p, q των γραμμικών μοντέλων.

6.7.2 Εκτίμηση

Το στάδιο που ακολουθεί της ταυτοποίησης είναι αυτό της εκτίμησης. Στο δεύτερο στάδιο της διαδικασίας των Box-Jenkins αξιοποιείται το μοντέλο που προσδιορίστηκε ότι μπορεί να προσφέρει αξιολογικά αποτελέσματα προηγουμένως. Για να συμβεί αυτό πρέπει να υπολογιστούν οι συντελεστές των μοντέλων της αυτοπαλινδρόμησης a_1, a_2, \dots, a_p και του κινούμενου μέσου $\theta_1, \theta_2, \dots, \theta_q$. Η εκτίμηση αυτών πραγματοποιείται ανάλογα τη φύση του μοντέλου. Όταν υπό αξιοποίηση υπάρχει ένα αυτοπαλίνδρομο μοντέλο (AR), η εκτίμηση τους γίνεται με τη μέθοδο ελαχίστων τετραγώνων LS. Το ίδιο δε μπορεί να συμβεί στη περίπτωση που το μοντέλο περιέχει διαδικασία κινούμενου μέσου (MA). Στη περίπτωση αυτή είναι ανάγκη να εφαρμοστούν μη γραμμικοί τρόποι εκτίμησης. Συνεπώς, κάθε φορά η μέθοδος που θα εφαρμοστεί ποικίλει, ανάλογα τη φύση του μοντέλου που απαιτεί η χρονοσειρά για να περιγραφεί και τη κατάσταση στασιμότητας της χρονοσειράς. Στις περιπτώσεις που χρησιμοποιείται λογισμικό για την

εκτίμηση αυτών, εφαρμόζεται ο αλγόριθμος που περιλαμβάνει τη διαδικασία βελτιστοποίησης του Marquardt και εκτιμώνται αυτόματα οι παράμετροι αυτοί [17].

6.7.3 Διαγνωστικός Έλεγχος

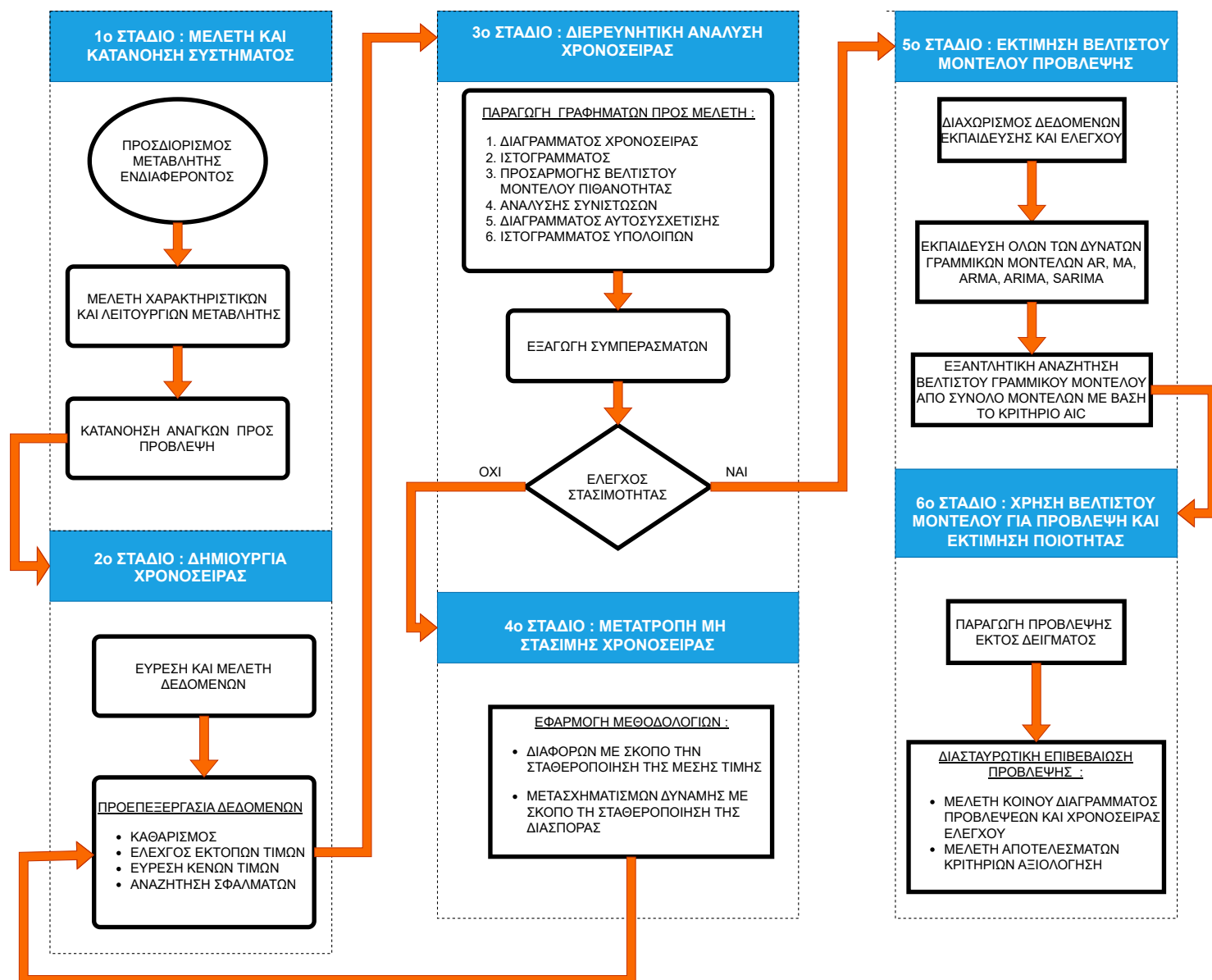
Το στάδιο που έπεται της ταυτοποίησης και της εκτίμησης καλείται στάδιο διαγνωστικού ελέγχου. Σε αυτό το σημείο το μοντέλο το οποίο έχει προσδιοριστεί και εκτιμηθεί ελέγχεται αν μπορεί να εξάγει αξιόλογα αποτελέσματα στη επικείμενη πρόβλεψη που θα πραγματοποιηθεί. Ουσιαστικά, εξετάζεται η προσαρμοστικότητα του στα δεδομένα. Εάν το μοντέλο μπορεί να περιγράψει ποιοτικά την χρονοσειρά, είναι πολύ πιθανό σε αυτή τη περίπτωση τα υπόλοιπα της χρονοσειράς να φέρουν τα χαρακτηριστικά του λευκού θορύβου. Είτε γίνεται έλεγχος των αυτοσυσχετίσεων αυτών μέσω του γραφήματος της εν λόγω συνάρτησης όπου εντοπίζεται ότι έχουν μηδενικές αυτοσυσχετίσεις είτε παράγεται το ιστόγραμμα αυτών όπου παρατηρείται ότι έχουν μέση τιμή μηδέν και σταθερή διακύμανση με την συνάρτηση πυκνότητας πιθανότητας αυτής να προσομοιώνει απόλυτα την Gaussian. Ακόμη δεν παραλείπεται να ελεγχθεί η τάξη του γραμμικού μοντέλου. Για να ελεγχθεί εάν υπάρχει μοντέλο το οποίο ικανοποιεί καλύτερο τον σκοπό μας, πραγματοποιείται σύγκριση με μοντέλα ανώτερης τάξης. Εάν το μοντέλο που έχει επιλεγεί είναι τάξεως p, q , τότε αυτά που χρησιμοποιούνται στη σύγκρισή είναι δύο ακόμη. Το πρώτο φέρει την ίδια τάξη p στον όρο αυτοπαλινδρόμησης αλλά κατά ένα αυξημένη την τάξη του όρου του κινούμενου μέσου. Το δεύτερο έχει την ίδια τάξη q στον όρο του κινούμενου μέσου και κατά ένα αυξημένη την τάξη του όρου της αυτοπαλινδρόμησης. Ωστόσο, πρέπει να λαμβάνεται υπόψη το κόστος που υπάρχει όταν εξετάζεται ένα μοντέλο μεγαλύτερης τάξης πέρα από τα οφέλη που έχει η ενέργεια αυτή. Καθώς γίνεται εκτίμηση παραπάνω παραμέτρων, υφίσταται μείωση των βαθμών ελευθερίας αλλά από την άλλη μειώνεται το άθροισμα των τετραγώνων των υπολοίπων. Για το λόγο αυτό, η επιλογή του μοντέλου γίνεται με βάση ένα κριτήριο πληροφωρίας. Ονομάζεται Akaike Information Criterion AIC και ορίζεται ως :

$$AIC = \ln \frac{\sum \hat{u}^2}{N} + \frac{2\kappa'}{N} \quad (6.7)$$

με το $\sum \hat{u}^2$ να εκφράζει το άθροισμα των τετραγώνων των υπολοίπων, το N το πλήθος παρατηρήσεων της χρονοσειράς και το κ' να ορίζει το πλήθος των παραμέτρων ($p + q + 1$) που υπολογίζονται. Έτσι λοιπόν, το μοντέλο το οποίο στο κριτήριο αυτό λάβει τη πιο μικρή τιμή υποδεικνύεται ως το πλέον κατάλληλο για τις μετέπειτα προβλέψεις.

6.7.4 Διαδικασία Πρόβλεψης

Η παραγωγή μιας πρόβλεψης αξιοποιώντας τις πληροφορίες που παρέχει μια χρονοσειρά καθώς εφαρμόζονται γραμμικά μοντέλα πρόβλεψης δεν αποτελεί μια απλή διαδικασία. Στη παρούσα διπλωματική δεν υλοποιήθηκε μια διαδικασία πρόβλεψης γνωστή εκ των προτέρων. Παρόλα αυτά υιοθετήθηκαν από ήδη γνωστές διαδικασίες χαρακτηριστικά στοιχεία. Η διαδικασία που επηρέασε πιο πολύ από όλες τη διαδικασία που θα περιγραφεί ήταν αυτή των Box - Jenkins. Στο Σχήμα 6.2 φαίνονται όλα τα στάδια αυτής καθώς και τα βήματα που υλοποιούνται σε κάθε ένα από αυτά.



Σχήμα 6.2: Η διαδικασία πρόβλεψης που αναπτύχθηκε και εφαρμόστηκε στη παρούσα διπλωματική εργασία.

Όπως γίνεται εύκολα αντιληπτό, η διαδικασία αναπτύσσεται και ολοκληρώνεται σε 6 στάδια. Αυτά είναι τα ακόλουθα :

1. **Μελέτη και Κατανόηση Συστήματος :** Το πρώτο στάδιο αποτελεί τον αχρονωγωναίο λίθο της διαδικασίας αυτής. Καμία πρόβλεψη δεν μπορεί να πραγματοποιηθεί εάν δεν μελετηθεί και τελικώς κατανοηθεί εις βάθος η μεταβλητή ενδιαφέροντος και το σύστημα στο οποίο εντάσσεται αυτή. Αρχικά, απαιτείται να πραγματοποιηθεί έρευνα και να αξιοποιηθεί βιβλιογραφία με την οποία ο ερευνητής θα καταλάβει ουσιαστικά τα χαρακτηριστικά της μεταβλητής που εξετάζει καθώς και τους μηχανισμούς λειτουργίας αυτής. Θα εξετάσει τις πτυχές αυτής ώστε να μπορεί να διακρίνει και να αντιληφθεί έπειτα κατά την στατιστική ανάλυση αυτής τα όσα παρουσιάζονται. Έτσι, αποκτά χρήσιμες πληροφορίες για τα επόμενα βήματα του. Για παράδειγμα, στην περίπτωση της χρονοσειράς ποσοτήτων παραγόμενης ηλεκτρικής ενέργειας από αιολικά πάρκα, δεν αρκεί να μελετηθεί μόνο ο τρόπος παραγωγής αυτής. Είναι απολύτως απαραίτητο ο ερευνητής να μπορεί να δικαιολογήσει γιατί μεταβάλλεται ανά εποχές η παραγωγικότητα ή ανά ώρες μέσα στη διάρκεια της ίδιας της ημέρας. Γεγονότα τα οποία προκαλούν τις μεταβολές αυτές, στη συγκεκριμένη περίπτωση, αποτελούν φυσικές διεργασίες που επηρεάζουν τη χρονοσειρά αυτή. Ακόμη δεν πρέπει να λησμονηθεί το γεγονός ότι για να πραγματοποιηθεί μια πρόβλεψη, πρέπει να υπάρχει μια ανάγκη η οποία πρέπει να καλυφθεί. Πέρα λοιπόν από την κατανόηση του συστήματος του οποίου τη χρονοσειρά θα μελετήσει ο ερευνητής παράλληλα πρέπει να αντιληφθεί και τις ανάγκες που καλύπτει η πρόβλεψη του. Σύμφωνα με αυτές ορίζει ένα στόχο πρόβλεψης και καθορίζει τα οφέλη αυτής που μπορεί να φύσει οικονομικής, οργανωτικής ή επιστημονικής. Στη περίπτωση μας, η ανάγκη κατάθεσης προτάσεων παραγωγής στο Χρηματιστήριο Ενέργειας από την εταιρία διαχείρισής του αιολικού πάρκου ωφελεί τόσο σύστημα διαχείρισης ενέργειας μιας χώρας όσο και την ίδια την εταιρία.
2. **Προσδιορισμός Χρονοσειράς :** Το δεύτερο στάδιο στην προσπάθειά πρόβλεψης αφορά τον προσδιορισμό της χρονοσειράς που θα αξιοποιηθεί. Ξεκινάει με την αναζήτηση δεδομένων. Χωρίς τον επιθυμητό όγκο δεδομένων για το σκοπό που έχει τεθεί προηγουμένως, καμία πρόβλεψη δεν είναι εφικτή να πραγματοποιηθεί. Αυτά τα δεδομένα ανακτούνται είτε από βάσεις δεδομένων κοινής χρήσης είτε από ιδιωτικές κατά κύριο λόγο. Με την ανάκτηση αυτών, ξεκινάει η διαδικασία μελέτης αυτών. Κατά την μελέτη τους, αποσαφηνίζεται η ποιότητα και η πληρότητα αυτών. Υπάρχει περίπτωση ανάμεσα σε όλα αυτά να υπάρχουν ελαττωματικές μετρήσεις ή να απουσιάζουν ακόμη και τιμές. Σε κάθε περίπτωση πριν την δημιουργία της χρονοσειράς, οφείλεται να καθαρίζονται και να ελέγχεται η ορθότητα των τιμών αυτής. Τελικά δημιουργείται η χρονοσειρά είτε με τον όγκο των δεδομένων που διατίθενται είτε με λιγότερα καθώς επιλέγεται να γίνει μια επαναδειγματοληψία η οποία θα ταιριάζει στον ορίζοντα πρόβλεψης. Ακόμη ίσως απαιτείται να πραγματοποιηθεί κάποια αλγεβρική διαδικασία επί των τιμών αυτής πριν την αξιοποίησή τους.
3. **Διερευνητική Ανάλυση Χρονοσειράς :** Στο στάδιο αυτό αξιοποιούνται κατά

το μέγιστο δυνατό βαθμό οι πληροφορίες που περιέχονται στη χρονοσειρά. Για να γίνει αυτό, παράγονται μια σειρά γραφημάτων προς παρατήρηση. Με την παρατήρηση αυτών είναι δυνατό να μελετηθούν οι συνιστώσες της χρονοσειράς, η συμπεριφορά της στον χρόνο, η στασιμότητα που δύναται να εμφανίζει, η συνάρτηση πυκνότητας πιθανότητας που ακολουθούν και η καταλληλότητα εφαρμογής γραμμικών μοντέλων για τη πρόβλεψη. Μετά την ανάγνωση αυτών για την εξαγωγή χρήσιμων πληροφοριών για τη φύση της χρονοσειράς, παράλληλα γίνεται έλεγχος στασιμότητας. Στη περίπτωση που παρουσιάζονται στοιχεία στασιμότητας, η διαδικασία μεταβαίνει στο 5ο στάδιο και όχι στο 4ο όπου μετατρέπεται η χρονοσειρά σε στάσιμη.

4. **Μετατροπή Μη Στάσιμης Χρονοσειράς σε Στάσιμη :** Το στάδιο αυτό κρίνεται πολύ σημαντικό για την διεκπεραίωσή της διαδικασίας πρόβλεψης. Μια πρόβλεψη για να παραχθεί καθώς βασίζεται στη χρήση γραμμικών μοντέλων πρόβλεψης απαιτεί η χρονοσειρά να είναι στάσιμη. Σε διαφορετική περίπτωση τα αποτελέσματα μετά την εφαρμογή αυτών, είναι πολύ πιθανόν να μη χαίρουν αξιοπιστίας και απόδοσης. Προκειμένου να μετατραπεί μια χρονοσειρά σε στάσιμη οφείλουν να εφαρμοστούν ορισμένες διαδικασίες οι οποίες είτε θα σταθεροποιήσουν την μέση τιμή είτε την διακύμανση είτε και τις δύο. Στη προκείμενη διαδικασία τα εργαλεία που πραγματοποιούν τις ρυθμίσεις αυτές είναι η διαφόριση των τιμών της χρονοσειράς και ο μετασχηματισμός τους με την χρήση κάποιας δύναμης. Με το πέρας της ολοκλήρωσης αυτών η νέα πλέον χρονοσειρά διερευνάται ξανά, υλοποιείται ουσιαστικά το 3ο Στάδιο για την ανάγνωση και κατανόηση των αποτελεσμάτων που είχαν οι μέθοδοι αυτοί στη χρονοσειρά μας. Υπάρχει περίπτωση να παραχθεί μια νέα χρονοσειρά στάσιμη, ικανή να αξιοποιηθεί στη μετέπειτα πορεία αλλά υπάρχει πάντα η πιθανότητα η χρονοσειρά να μην μετατράπηκε οπότε να συνεχιστεί η διαδικασία μετατροπή της έως ότου επέλθει το επιθυμητό αποτέλεσμα.
5. **Εκτίμηση Βέλτιστου Μοντέλου Πρόβλεψης :** Στο σημείο αυτό εκτιμάται το μοντέλο το οποίο θα χρησιμοποιηθεί για την παραγωγή της πρόβλεψης στο επόμενο στάδιο. Παρόλα αυτά στην αρχή του σταδίου αυτού γίνεται μια διχοτόμηση στα δεδομένα. Χωρίζονται σε δεδομένα εκπαίδευσης (κύριος όγκος) και δεδομένα ελέγχου (μικρό δείγμα). Τα δεδομένα εκπαίδευσης θα χρησιμοποιηθούν στο στάδιο αυτό. Με βάση αυτά γίνεται η εκπαίδευση και η εκτίμηση των γραμμικών μοντέλων που μπορούν καλύτερα από όλα να περιγράψουν την χρονοσειρά. Για να βρεθεί είτε το βέλτιστο μοντέλο AR είτε το βέλτιστο μοντέλο MA είτε το βέλτιστο μοντέλο ARMA είτε το βέλτιστο μοντέλο ARIMA είτε το βέλτιστο μοντέλο SARIMA πραγματοποιείται μια εξαντλητική αναζήτηση σε κάθε περίπτωση. Στην αναζήτηση αυτήν τίθεται ένα εύρος τιμών στο οποίο γίνεται αναζήτηση της βέλτιστης τάξης. Αφού γίνει η εκπαίδευση του για την κάθε περίπτωση, υπολογίζεται η τιμή του κριτηρίου AIC. Το μοντέλο με την μικρότερη τιμή στη κάθε κατηγορία μοντέλων αποτελεί το βέλτιστο.
6. **Πρόβλεψη :** Τελευταίο στάδιο στην διαδικασία πρόβλεψης είναι αυτό όπου παράγεται η πρόβλεψη, αξιολογείται και προσδιορίζεται η καλύτερη που μπορεί να προσφερθεί από

τα διαθέσιμα γραμμικά μοντέλα προς αξιοποίηση. Αρχικά παράγονται εκτός δείγματος εκπαιδευσης οι προβλέψεις. Ο αριθμός των παρατηρήσεων που θα προβλεφθούν μετά τη τελευταία παρατήρηση του δείγματος εκπαίδευσης ορίζεται από το πλήθος του δείγματος ελέγχου. Αυτό έχει οριστεί με βάση τον επιθυμητό ορίζοντα πρόβλεψης. Παράγονται λοιπόν οι προβλέψεις με βάση το κάθε μοντέλο και οπτικοποιούνται τόσο αυτές όσο και οι αληθινές τιμές. Με τη παρατήρηση αυτών όσων και των τιμών των κριτηρίων αξιολόγησης που υπολογίζονται, γίνεται η αξιολόγηση και σύγκριση τους. Έτσι, τελικά προσδιορίζεται η βέλτιστη πρόβλεψη και συνάμα ποιο ήταν τελικά το βέλτιστο μοντέλο.

6.7.5 Αξιολόγηση προβλέψεων

Πολύ σημαντικό στην στατιστική ανάλυση δεν είναι μόνο η παραγωγή προβλέψεων παρά επίσης και η αξιολόγηση τους σε ύστερη φάση. Κατά αυτόν τον τρόπο εκτιμάται επίσης ο βαθμός απόδοσης των γραμμικών μοντέλων που παρήγαγαν την πρόβλεψη. Για το λόγο αυτό επιστρατεύονται μια σειρά από μέτρα αξιολόγησης. Στην παρούσα εργασία τα μέτρα αξιολόγησης τα οποία κλήθηκαν να εκτιμήσουν την ποιότητα των προβλέψεων είναι τα ακόλουθα :

- ME - Μέσο Σφάλμα

Το κριτήριο μέσου σφάλματος χρησιμοποιείται κυρίως για να ελεγχθεί η περίπτωση ύπαρξης μεροληψίας επί της πρόβλεψης. Μέσω αυτού διαπιστώνεται εάν έγινε υποεκτίμηση ή υπερεκτίμηση. Στη περίπτωση που το πρόσημο του αποτελέσματος είναι θετικό τότε υπάρχει υπερεκτίμηση ενώ στην αντίθετη περίπτωση υποεκτίμηση. Με την σχέση που εκφράζεται από την Εξ. (6.8) υπολογίζεται η τιμή του κριτηρίου αυτού :

$$ME = \frac{1}{N} \sum_{t=1}^N (f_t - x_t) \quad (6.8)$$

όπου ως N λογίζεται το πλήθος των δεδομένων που συμμετέχουν στη πρόβλεψη εκτός δείγματος, ενώ ως f_t λογίζεται η προβλεπόμενη τιμή και ως x_t η τιμή που πραγματικά έχει παρθεί ως μέτρηση την χρονική στιγμή t .

όπου ως N λογίζεται το πλήθος των παρατηρήσεων που προβλέφθηκαν.

- MPE - Μέσο Ποσοστιαίο Σφάλμα

Το κριτήριο του μέσο ποσοστιαίου σφάλματος είναι ένα ακόμη κριτήριο με το οποίο μετράται η μεροληψία στη πρόβλεψη. Είναι ένας διαφορετικός τρόπος να γίνει αντιληπτή η μεροληψία που υφίσταται στην πρόβλεψη. Η διαφορά τους έγκειται στο γεγονός ότι στο κριτήριο μέσου σφάλματος το αποτέλεσμα εκφράζεται με τις μονάδες μέτρησης της μεταβλητής ενδιαφέροντος ενώ στο κριτήριο αυτό εκφράζεται ποσοστιαία. Η Εξ 6.9 που δίνει τον τρόπο υπολογισμού του ακολουθεί :

$$MPE = \frac{1}{N} \sum_{t=1}^N (f_t - x_t) \cdot 100 \quad (6.9)$$

όπου ως N λογίζεται το πλήθος των παρατηρήσεων που προβλέφθηκαν.

- **RMSE - Ρίζα Μέσου Τετραγωνικού Σφάλματος**

Το κριτήριο της ρίζας μέσου τετραγωνικού σφάλματος αποτελεί το κύριο κριτήριο με βάση το οποίο αξιολογήθηκαν τα γραμμικά μοντέλα που παρήγαγαν προβλέψεις. Αυτό συμβαίνει διότι η χρήση του υιοθετεί την υπόθεση ότι τα λάθη δεν περιέχουν προκατάληψη παρά ακολουθούν την κανονική κατανομή κάτι το οποίο απαιτείται επίσης ως προϋπόθεσή από τα υπόλοιπα της χρονοσειράς προκειμένου να είναι στάσιμα. Το συγκεκριμένο κριτήριο υπολογίζεται αφού υπολογιστεί το τετράγωνο των διαφορών ανάμεσα στις τιμές που προβλέφθηκαν και τις πραγματικές και έπειτα αφού υπολογιστεί ο μέσος όρος αυτών, παρθεί η ρίζα. Συνεπώς οι μικρότερες τιμές στο κριτήριο αυτό προσδίδουν την πληροφορία περί ενός γραμμικού μοντέλου με καλή απόδοση αφού η μέση διακύμανση των λαθών δεν θα είναι πολύ μακριά από την πραγματική και άρα οι προβλέψεις είναι ικανοποιητικά εύστοχες. Στην Εξ 6.10 που ακολουθεί, παρουσιάζεται ο τρόπος υπολογισμού :

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (f_t - x_t)^2} \quad (6.10)$$

όπου ως N λογίζεται το πλήθος των δεδομένων που συμμετέχουν στη πρόβλεψη εκτός δείγματος, ενώ ως f_t λογίζεται η προβλεπόμενη τιμή και ως x_t η τιμή που πραγματικά έχει παρθεί ως μέτρηση την χρονική στιγμή t .

- **MAE - Μέσο Απόλυτο Σφάλμα**

Το μέσο απόλυτο σφάλμα MAE αποτελεί ένα κριτήριο του οποίου η χρήση αποσκοπεί στην εκτίμηση της διάστασης του σφάλματος. Ουσιαστικά εκτιμάται η μέση απόσταση της προβλεπόμενης τιμής από την πραγματική. Σημαντικό πλεονέκτημα του κριτηρίου αυτού αποτελεί το γεγονός ότι η ερμηνεία του είναι εύκολη αφού η τιμή του βρίσκεται στην ίδια μονάδα μέτρησης με την μεταβλητή που εξετάζεται. Στην Εξ. (6.11) παρουσιάζεται η σχέση με την οποία είναι εφικτός ο υπολογισμός του :

$$MAE = \frac{1}{N} \sum_{t=1}^N |f_t - x_t| \quad (6.11)$$

όπου ως N λογίζεται το πλήθος των δεδομένων που συμμετέχουν στη πρόβλεψη εκτός δείγματος, ενώ ως f_t λογίζεται η προβλεπόμενη τιμή και ως x_t η τιμή που πραγματικά έχει παρθεί ως μέτρηση την χρονική στιγμή t .

- **MAPE - Μέσο Απόλυτο Τετραγωνικό Σφάλμα**

Η ποσοστοτικοποίηση του κριτηρίου του μέσου απόλυτου σφάλματος ονομάζεται δημιουργεί το κριτήριο του ποσοστιαίου μέσου απόλυτου σφάλματος (MAPE). Με τη βοήθεια του κριτηρίου αυτού γίνεται πιο εύκολη η αντίληψη της μέση διαφοράς που υπάρχει

ανάμεσα στην προβλεπόμενη και πραγματική τιμή καθώς πλέον είναι εκπεφρασμένη επί της %. Η τιμή του κριτηρίου αυτή όπως είναι λογικό αυξάνεται γραμμικά με το μέγεθος του λάθους. Συνεπώς μοντέλα με μικρές τιμές αποδίδονται σε μοντέλα που έφεραν καλές προβλέψεις. Στην Εξ. (6.12) παρουσιάζεται η σχέση η οποία χρησιμοποιείται για να υπολογιστεί η τιμή του κριτηρίου αυτού :

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{f_t - x_t}{x_t} \right| \cdot 100 \quad (6.12)$$

όπου ως N λογίζεται το πλήθος των δεδομένων που συμμετέχουν στη πρόβλεψη εκτός δείγματος, ενώ ως f_t λογίζεται η προβλεπόμενη τιμή και ως x_t η τιμή που πραγματικά έχει παρθεί ως μέτρηση την χρονική στιγμή t .

- R^2 - Συντελεστής προσδιορισμού

Ένα ακόμη βασικό κριτήριο που χρησιμοποιείται στην αξιολόγηση των προβλέψεων είναι ο συντελεστής προσδιορισμού R^2 . Με το κριτήριο αυτό προσδιορίζεται εν πολλοίς η προσαρμοστικότητα του μοντέλου. Οι τιμές που λαμβάνει βρίσκονται στο διάστημα 0% έως 100%. Ισχύει ότι επί του ίδιου όγκου δεδομένων, όσο μεγαλύτερη η τιμή που λαμβάνει, τόσο μικρότερη είναι η διαφορά που υφίσταται ανάμεσα στις πραγματικές τιμές και στις τιμές που εκτίμησε και απέδωσε το μοντέλο που προσαρμόστηκε στα δεδομένα εκπαίδευσης. Συνήθως τιμές που προσεγγίζουν το 100% απορρέουν από μοντέλα που αποδίδουν άριστες και έμπιστες προβλέψεις. Σε κάθε περίπτωση η τιμή του κριτηρίου εξηγεί το ποσοστό μεταβλητότητας των δεδομένων. Με την Εξ. (6.13) γίνεται δυνατός ο υπολογισμός του :

$$R^2 = 1 - \frac{\sum_{t=1}^N (f_t - x_t)^2}{\sum_{t=1}^N (\bar{x}_t - x_t)^2} \quad (6.13)$$

όπου ως N λογίζεται το πλήθος των δεδομένων που συμμετέχουν στη πρόβλεψη εκτός δείγματος, ενώ ως f_t λογίζεται η προβλεπόμενη τιμή και ως x_t η τιμή που πραγματικά έχει παρθεί ως μέτρηση την χρονική στιγμή t . Τέλος, ως \bar{x}_t συμβολίζεται η μέση τιμή των πραγματικών τιμών των παρατηρήσεων της χρονοσειράς.

6.8 XGBoost

Το XGBoost ((eXtreme Gradient Boosting) αποτελεί έναν από τους πλέον αποτελεσματικούς αλγόριθμους μηχανικής μάθησης (machine learning) σε προβλήματα κατηγοριοποίησης (classification) δεδομένων και πρόβλεψης μέσω παλινδρόμησης (regression). Ο αλγόριθμος αυτός, ο οποίος έχει υλοποιηθεί σε διάφορες γλώσσες προγραμματισμού όπως η python, προκειμένου να γίνει πιο εύκολα αντιληπτός, κρίνεται σημαντικό να εξηγηθούν πρότερα αυτού μερικές από τις τεχνικές που εμπεριέχονται σε αυτόν. Ακολουθούν, λοιπόν, οι τεχνικές και οι αλγόριθμοι οι οποίοι θα μελετηθούν συνοπτικά ώστε να γίνει κατανοητός ο τρόπος λειτουργίας του συνολικά [7, 34, 12, 35].

Δένδρα Απόφασης (Παλινδρόμησης) Decision (Regression) Trees

Ένα δέντρο παλινδρόμησης αποτελεί ένα είδος δέντρου απόφασης. Μέσω της ανάπτυξης ενός τέτοιου δέντρου είναι δυνατό να γίνει με ικανοποιητικό τρόπο η προσαρμογή των δεδομένων εκπαίδευσης. Σε ένα τέτοιο δέντρο, κάθε κόμβος αποτελεί ένα κατώφλι με βάση το οποίο διαχωρίζει βέλτιστα σε περιοχές τιμών τις παρατηρήσεις. Κάθε φύλλο αυτού αποτελεί τη πρόβλεψη και ως τιμή περιέχει τη μέση τιμή των παρατηρήσεων που βρίσκονται στη περιοχή τιμών που όρισε ο κόμβος που το δημιουργήσε. Το κατώφλι που ορίζεται κάθε φορά γίνεται με βάση το κριτήριο του ελαχίστου αθροίσματος τετραγώνων των υπολοίπων. Για να βρεθεί, υπολογίζονται όλα τα δυνατά και επιλέγεται εν τέλη το μικρότερο. Η διαδικασία αυτή επαναλαμβάνεται έως ότου δεν υπάρχουν προς διαχωρισμό ένα πλήθος τιμών το οποίο ορίζουμε εξ αρχής με βάση το πλήθος των δεδομένων που μελετώνται.

Πρόκειται για μια διαδικασία η οποία χρονικά κοστίζει και συνήθως αν και μπορεί να περιγράψει με ικανοποιητικό τρόπο τα δεδομένα εκπαίδευσης οδηγεί τελικά σε ένα μοντέλο με χαμηλή μεροληψία και υψηλή μεταβλητότητα (overfitting) όταν αυτό εφαρμόζεται σε νέα σετ δεδομένων. Για να λυθεί εν μέρη το πρόβλημα αυτό επιλέγεται το κλάδεμα ενός δέντρου (αφαίρεση φύλων) με βάση κάποιο κριτήριο πολυπλοκότητας και η αντικατάστασή τους με τη μέση τιμή μιας μεγαλύτερης ομάδας παρατηρήσεων. Το μειονέκτημα που εμπεριέχεται στη παραπάνω λύση είναι ότι είναι πιθανό τα δένδρα αν κλαδευτούν πολύ να μην αποφέρουν καρπούς αφού η εκπαίδευσή τους είναι πλέον ανεπαρκής.

Bagging

Ένας τρόπος αντιμετώπισης του προβλήματος μεταβλητότητας overfitting αποτελεί η τεχνική Bagging. Βασικός στόχος της μεθόδου αυτή είναι να μειώσει τη μεταβλητότητα αυξάνοντας την μεροληψία (bias), δηλαδή, την μην ικανότητα επαρκούς περιγραφής της σχέσης που υπάρχει ανάμεσα στις μεταβλητές που ορίζουν το σετ δεδομένων μας, ώστε να είναι πιο εύστοχο στις προβλέψεις του το μοντέλο μηχανικής μάθησης. Ο πόρος που αξιοποιείται στο συγκεκριμένο αλγόριθμο είναι ένα σετ δεδομένων εκπαίδευσης από το οποίο θα προκύψουν K διαφορετικά σετ δεδομένων εκπαίδευσης. Αυτά πάντα θα περιέχουν σε πλήθος όσες παρατηρήσεις περιείχε το αρχικό.

Δημιουργούνται δειγματοληπτώντας ομοιόμορφα το αρχικό. Είτε επιλέγονται όλες τις παρατηρήσεις του με τυχαία σειρά είτε επιλέγονται μερικές από αυτές με τυχαία σειρά. Στη δεύτερη περίπτωση κάποια από αυτές δειγματοληπτείται παραπάνω από μια φορά και τόσες ώστε να δημιουργηθεί ένα σετ δεδομένων εκπαίδευσης πλήθους ίδιο με το αρχικό. (Bootstrap). Για το κάθε ένα δέντρο δημιουργείται ένα δέντρο παλινδρόμησης και πραγματοποιείται μια πρόβλεψη. Έτσι για τη κάθε παρατήρηση δημιουργούνται K προβλέψεις όπου η μέση τιμή αυτών αποτελεί την τελική εκτίμηση για την εκάστοτε παρατήρηση (Aggregation). Έτσι λοιπόν το μοντέλο που εκπαιδεύεται, δημιουργείται από τις διεργασίες που προαναφέρθηκαν και οι οποίες γίνονται παράλληλα για το κάθε δέντρο.

Random Forrest

Η μέθοδος Random Forest έρχεται να δώσει λύση στη περίπτωση που όταν αξιοποιείται κάποιο δέντρο εμφανίζεται overfitting. Η μέθοδος αυτή δανείζεται χαρακτηριστικά από την τεχνική Bagging καθώς επίσης και τα προτερήματά της. Η βασική ιδέα στην Random Forest είναι το αποτέλεσμα να εξάγεται με βάση πολλά διαφορετικά δέντρα και όχι αποκλειστικά ένα. Παράγονται έτσι προβλέψεις οι οποίες δεν έχουν αυστηρή συσχέτιση μεταξύ τους με αποτέλεσμα αυτές να είναι τελικά πιο εύστοχες. Για να γίνει βέβαια αυτό πρέπει :

1. Να δημιουργηθεί ένα bootstrapped dataset από σετ δεδομένων εκπαίδευσης. Έχει δηλαδή ίδιο πλήθος παρατηρήσεων με το αρχικό αλλά οι παρατηρήσεις που επιλέγονται ομοιόμορφα από το αρχικό βρίσκονται σε τυχαία σειρά και είναι δυνατό να μην έχει επιλεγεί κάποια και να έχει επιλεγεί κάποια άλλη παραπάνω από μια φορά.
2. Το δέντρο που κατασκευάζεται από το bootstrapped dataset επιλέγεται να εμπεριέχει ένα τυχαίο υποσύνολο a από τις διαθέσιμες παρατηρήσεις b σε κάθε βήμα.
3. Η διαδικασία που ορίζουν τα βήματα 1 και 2 καλείται να επαναληφθεί N φορές, τον οποίο αυτόν αριθμό ορίζει ο εκάστοτε ερευνητής αναλόγως το πρόβλημα.
4. Όπως και στη τεχνική Bagging η τελική πρόβλεψη προκύπτει από ένα μοντέλο το οποίο για να την παράξει υπολογίζει την μέση τιμή από τις προβλέψεις που έδωσαν τα N δέντρα που παρήχθησαν.

Συνήθως επιλέγεται η σχέση $a = \sqrt{b}$ προκειμένου να υπάρχει μια ισορροπία ανάμεσα στην απόδοση του δέντρου και την μη συσχέτισή των προβλέψεων.

Gradient Boosting

Μια τεχνική ακόμα που βασίζεται σε δέντρα απόφασης και κρίνεται αναγκαία να μελετηθεί για την τελική κατανόηση του αλγορίθμου XGB ονομάζεται Gradient Boosting. Έχει αρκετές ομοιότητες με την τεχνική Random Forest και μια κύρια διάφορα. Τα δέντρα που παράγονται και παράγουν την τελική πρόβλεψη για την κάθε παρατήρηση του train dataset δεν είναι ασυσχέτιστα μεταξύ τους παρά δημιουργούνται σε σειρά. Ουσιαστικά παράγεται ένας μοντέλο με ισχυρή ικανότητα πρόβλεψης από μοντέλα με ισχνή ικανότητα. Το πρώτο δέντρο έχει μια ιδιομορφία αποτελείται από ένα μόνο φύλλο.

Το φύλλο αυτό περιέχει σαν τιμή τη μέση τιμή των τιμών όλων των παρατηρήσεων του σετ εκπαίδευσης και αποτελεί την αρχική πρόβλεψη. Με βάση αυτή υπολογίζεται η διαφορά ανάμεσα στην παρατηρήσιμη τιμή κάθε παρατήρησης και την προβλεπόμενη. Είναι το μέγεθος του λάθους πρόβλεψης του αρχικού δέντρου και με βάση αυτή θα αναπτυχθεί το επόμενο δέντρο. Το λάθος αυτό ονομάζεται ψευτό υπόλοιπο pseudo-residual. Τα δέντρα που προστίθενται σε σειρά μετά από το αρχικό (ο αριθμός ορίζεται από τον ερευνητή) δεν λαμβάνουν μόνο υπόψιν τους το κάθε ένα τις πληροφορίες που εξάγονται από τα λάθη που έκανε το προηγούμενο στη διαδικασία πρόβλεψης.

Για να χτιστεί κατάλληλα το δέντρο, αναγκαία προϋπόθεσή είναι να γίνει ο κατάλληλος διαχωρισμός του δέντρου με βάση άλλα γνωρίσματα που μπορούν να εξαχθούν από τις πληροφορίες που προσφέρει μια παρατήρηση. Με βάση αυτές για την κάθε παρατήρηση το ψευτό υπόλοιπο *pseudoresidual* που της αντιστοιχεί θα καταλήξει στα φύλλα του δέντρου. Στα φύλλα του κάθε νέου δέντρου το πλήθος των οποίων ορίζεται εξ αρχής είναι δυνατόν να καταλήξουν παραπάνω από ένα ψευτό υπόλοιπα *pseudo-residual*. Σε αυτήν την περίπτωση αντικαθίστανται με τη μέση τιμή όσων βρίσκονται σε αυτό.

Μόλις καταλήξουν στα φύλλα τα ψευτό υπόλοιπα *pseudoresiduals*, δηλαδή, τα λάθη του προηγούμενου δέντρου, ξεκινά η διαδικασία πρόβλεψης του δέντρου. Για να γίνει η νέα πρόβλεψη για την κάθε παρατήρηση προστίθεται η τιμή που περιέχει το αρχικό φύλλο (αρχική πρόβλεψη) και η τιμή του ψευτό υπολοίπου *pseudo-residual* που της αντιστοιχεί σταθμισμένη. Σταθμίζεται από μια παράμετρο που ονομάζεται ρυθμός εκμάθησης *learning rate* που μετριάζει τη συνεισφορά του δέντρου στη τελική πρόβλεψη. Υπολογίζονται λοιπόν οι νέες προβλέψεις και κατόπιν τα νέα ψευτουπόλοιπα *pseudo-residual*. Αν το μέγεθος αυτών έχει μειωθεί τότε έχει γίνει ένα βήμα προς τη σωστή κατεύθυνση. Η διαδικασία αυτή επαναλαμβάνεται έως ότου δημιουργηθούν όλα τα αρχικώς ορισμένα δέντρα ή έως ότου δεν επιφέρει ιδιαίτερη μείωση στο μέγεθος των ψευτό υπολοίπων *pseudo-residual* κάθε νέο δέντρο που προστίθεται. Η παράμετρος *learning rate* ορίζει τον ρυθμό με τον οποίο μαθαίνει ένα δέντρο και λαμβάνει τιμές στο διάστημα τιμών $[0, 1]$ με τις τιμές στο διάστημα $[0.1, 0.3]$ να δίδουν συνήθως τα καλύτερα αποτελέσματα. Είναι λοιπόν μια διαδικασία η οποία βασίζεται σε δέντρα παλινδρόμησης όπου το κάθε ένα βασίζεται στα λάθη που έκανε το προηγούμενο.

Extreme Gradient Boosting (XGboost)

Ο αλγόριθμος μηχανικής μάθησης XGboost αποτελεί μια προέκταση των αλγορίθμων για δέντρα Gradient Boosting και τα τελευταία χρόνια κυριαρχεί σε διαγωνισμούς επιστήμης δεδομένων. Δανείζεται χαρακτηριστικά από τον αλγόριθμο Gradient Boosting για να δημιουργήσει τα δέντρα του αλλά προσθέτει στοιχεία στη διαδικασία που κάνουν τελικά ξεχωριστό τον τρόπο με τον οποίο αυτά δημιουργούνται. Επίσης αξιοποιεί χαρακτηριστικά από αλγορίθμους κανονικοποίησης για να αποφύγει την πιθανότητα εμφάνισης του φαινομένου υπερπροσαρμογής του μοντέλου στο σετ δεδομένων εκπαίδευσης. Τέλος, χαρακτηριστικά αυτού του αλγόριθμου τα οποία είναι μοναδικά θα παρουσιαστούν κατόπιν μαζί με τα πλεονεκτήματα που δίδουν έναντι των άλλων αλγορίθμων που παρουσιάστηκαν παραπάνω.

Ο αλγόριθμος XGboost εκτελείται επαναληπτικά N φορές έως ότου δημιουργηθούν όσα δέντρα έχουν οριστεί από τον ερευνητή πριν την έναρξη της διαδικασίας. Σε κάθε επανάληψη δημιουργείται ένα δέντρο το οποίο λαμβάνει υπόψιν του τα υπολείμματα που δημιούργησε η πρόβλεψη που πραγματοποίησε το προηγούμενο προκειμένου να επιφέρει μια καλύτερη πρόβλεψη. Όλα ξεκινούν από ένα αρχικό φύλλο. Πριν χτιστεί το πρώτο δέντρο γίνεται μια αρχική πρόβλεψη η οποία εξ ορισμού ορίζεται και με βάση αυτήν υπολογίζονται τα υπόλοιπα (*Residuals*), η απόσταση δηλαδή της πραγματικής τιμής από την πρόβλεψη που έχει γίνει αρχικά. Έτσι, λοιπόν, έχοντας υπόψιν τα υπολείμματα γίνεται μια προσπάθεια να χτιστεί το πρώτο

δέντρο XGBoost.

Για να χτιστεί κάθε από τα N δέντρα παλινδρόμησης XGboost που βασίζεται στα υπόλοιπα της προηγούμενης πρόβλεψης μια ακολουθία βημάτων απαιτείται να εκτελεστεί. Αρχικά κάθε δέντρο όπως και το πρώτο ξεκινάει με ένα φύλλο μόνο. Σε αυτό το τοποθετούνται όλα τα υπόλοιπα και υπολογίζεται ο βαθμός ποιότητας Quality Score ή Similarity Score. Η Εξ. (6.14) παρουσιάζει τη σχέση που αξιοποιείται στη προκείμενη περίπτωση :

$$SS = \frac{\sum_t (x_t - y_t)^2}{R + \lambda} \quad (6.14)$$

όπου με SS νοείται ο βαθμός ποιότητας. Στον αριθμητή τοποθετείται τετραγωνισμένο το άθροισμα των υπολοίπων Residuals που βρίσκονται στο συγκεκριμένο φύλλο, ενώ στον παρονομαστή βρίσκεται ένα άθροισμα το οποίο αποτελείται από το πλήθος R των υπολοίπων Residuals και μιας παραμέτρου κανονικοποίησης λ .

Στη συνέχεια της διαδικασίας το δέντρο πρέπει να διαχωριστεί. Το φύλλο που περιείχε όλα τα υπόλοιπα Residuals θα μετατραπεί σε κατώφλι. Αυτό ορίζεται αρχικά από κάποια χαρακτηριστική πληροφορία των παρατηρήσεων χωρίς να είναι απαραίτητα το βέλτιστο. Για παράδειγμα, αυτό μπορεί να είναι η μέση τιμή των τιμών ενός πλήθους παρατηρήσεων ή κάποιο άλλο χαρακτηριστικό γνώρισμα που εξάγεται από τις παρατηρήσεις. Τα υπόλοιπα Residuals που βρίσκονταν έως τώρα στο αρχικό φύλλο, διαχωρίζονται σε δυο νέες ομάδες με βάση αυτό. Δημιουργούνται δηλαδή δύο νέα φύλλα με διαφορετικά υπόλοιπα και έπειτα υπολογίζεται για το κάθε νέο φύλλο ο νέος βαθμός ποιότητας Quality Score ή Similarity Score του.

Στη φάση αυτή του αλγορίθμου κρίσιμο είναι να υπολογισθεί το κέρδος διάσπασης Gain των υπολοίπων σε δύο νέες ομάδες με βάση αυτό το κατώφλι. Για το λόγο αυτό γίνεται χρήση μιας σχέσης που παρουσιάζεται από την Εξ. (6.15) :

$$Gain = SS_l + SS_r + SS_{node} \quad (6.15)$$

όπου για το αριστερό φύλλο ο βαθμός ποιότητας συμβολίζεται ως SS_l ενώ για το δεξί ως SS_r . Ως SS_{node} λογίζεται ο βαθμός ποιότητας του κατωφλίου από το οποίο προκύπτουν τα δύο φύλλα.

Για να βρεθεί, ωστόσο, το βέλτιστο κατώφλι η διαδικασία αυτή επαναλαμβάνεται για όλα τα διαθέσιμα κατώφλια. Ως βέλτιστο κατώφλι ορίζεται τελικά αυτό που λαμβάνει τη μεγαλύτερη τιμή.

Το δέντρο δεν έχει ολοκληρωθεί ακόμη βέβαια. Για να ολοκληρωθεί, απαιτείται είτε να φτάσει σε βάθος τον προκαθορισμένο αριθμό είτε πλέον στα φύλλα να βρίσκεται μόνο μια τιμή. Εάν δεν βρίσκεται πρέπει να γίνει ξανά διαχωρισμός και άρα να επαναληφθεί η διαδικασία εύρεσης του καταλληλότερου κατωφλίου που αναλύθηκε προηγουμένως. Μόλις, λοιπόν, ολοκληρωθεί το χτίσιμο του δέντρου ακολουθεί το κλάδεμα του από κάτω προς τα πάνω. Σε κάθε κόμβο (κατώφλι) υπολογίζεται η διαφορά ανάμεσα στη τιμή κέρδους διάσπασης (Gain) αυτού και της τιμής μιας παραμέτρου κανονικοποίησης γ η οποία ορίζεται αυθαίρετα από τον ερευνητή. Το αποτέλεσμα της πράξης αυτής καθορίζει το αν θα αφαιρεθεί ο κλάδος αυτός τελικά ή όχι. Εάν είναι θετικό, σε καμιά αλλαγή δεν εμπίπτει το δέντρο. Εάν όμως είναι αρ-

νητικό τότε έπεται κλάδεμα του κλάδου και συνέχεια στη διαδικασία αναζήτησης των κλάδων που θα αφαιρεθούν όσο ανεβαίνουμε προς τα πάνω.

Το δέντρο μετά τη διαδικασία κλαδέματος (pruning) έχει πλέον ολοκληρωθεί. Ακολουθεί ο υπολογισμός της τιμής εξόδου κάθε τελικού φύλλου. Στην Εξ. (6.16) παρουσιάζεται η σχέση που αξιοποιήθηκε για τον υπολογισμό αυτόν :

$$Output = \frac{\sum_t x_t - y_t}{R + \lambda} \quad (6.16)$$

όπου ως αριθμητής χρησιμοποιείται το άθροισμα των υπολοίπων Residuals του φύλλου και ως παρονομαστής το πλήθος αυτών αθροισμένο με μια παράμετρο λ .

Ο υπολογισμός της τιμής αυτής συμπεριλαμβάνεται στον υπολογισμό που γίνεται για την τελική νέα πρόβλεψη. Η νέα πρόβλεψη ολοκληρώνεται μέσω ενός αθροίσματος. Σε αυτό συμμετέχουν η αρχική πρόβλεψη και η σταθμισμένη από τη παράμετρο του ρυθμού μάθησης η (Learning Rate) εξόδιος τιμή του φύλλου που αντιστοιχεί στην υπό έρευνα παρατήρηση. Έτσι λοιπόν, γίνονται οι προβλέψεις για όλες τις παρατηρήσεις του σετ δεδομένων εκπαίδευσης και υπολογίζονται τα νέα υπόλοιπα Residuals. Αν το μέγεθος αυτών μειώθηκε, τότε έγινε ένα βήμα προς τη σωστή κατεύθυνση και με βάση αυτά επαναλαμβάνεται η διαδικασία χτισίματος ενός νέου δέντρου που σκοπό θα έχει να μειώσει ακόμα παραπάνω το μέγεθος τους. Η διαδικασία τερματίζεται εάν πλέον δεν υφίσταται κάποια αξιοπρόσεκτη μείωση ή εάν έχουν δημιουργηθεί και αξιοποιηθεί όλα τα δέντρα που αρχικά είχαν οριστεί. Η συνάρτηση απώλειας που χρησιμοποιείται για να μετρηθεί η προβλεπτική ικανότητα του μοντέλου στην κάθε παρατήρηση δίδεται από την Εξ. (6.17) :

$$L(x_t, y_t) = \frac{(x_t - y_t)^2}{2} \quad (6.17)$$

όπου ως $L(x_t, y_t)$ συμβολίζεται η συνάρτηση απώλειας των παρατηρήσιμων τιμών x_t και αυτών που προβλέφθηκαν y_t .

Ο αλγόριθμος XGBoost δεν χαίρει εκτίμησης μόνο επειδή συνδυάζει τεχνικές Gradient Boosting ή Regularization για να χτίσει τα δέντρα του με μοναδικό τρόπο και να προσαρμόσει το μοντέλο στα δεδομένα εκπαίδευσης. Στον αλγόριθμο αυτόν εμπεριέχονται τεχνικές οι οποίες τον καθιστούν αποδοτικό και γρήγορο όταν καλείται να επεξεργαστεί τεράστια σετ δεδομένων εκπαίδευσης.

Με δύο τεχνικές γνωστές ως Cache-Aware Split Finding και Blocks For Out Of Core Computation δύναται να βελτιστοποιήσει την απόδοση του χρονικά όταν μεταχειρίζεται σχετικά πολύ μεγάλα σετ δεδομένων εκπαίδευσης. Μέσω την εφαρμογή της Cache-Aware Split Finding επιτυγχάνονται οι υπολογισμοί των Similarity Score και Output Value οι οποίοι υπολογιστικά έχουν κόστος. Για να γίνει αυτό αποθηκεύονται στην πιο γρήγορη μνήμη του υπολογιστή την Cache και εκτελούνται εξοικονομώντας πολύ χρόνο από την πρόσβαση σε μια πιο αργή μνήμη όπως ο σκληρός δίσκος του υπολογιστή. Χρόνος επίσης εξοικονομείται με την τεχνική Blocks For Out Of Core Computation. Η διεργασία αυτή πραγματοποιείται όταν τα δεδομένα είναι τόσα πολλά που χρειάζεται μέρος αυτών να αποθηκευτεί στο σκληρό

δίσκο όπου το διάβασμα ή εγγραφή πληροφοριών καθυστερεί πολύ την διαδικασία συνολικά. Μια μερική λύση στον προβληματισμό αυτό είναι η συμπίεση και έπειτα αποσυμπίεση και αξιοποίηση (διάβασμα-εγγραφή) τους στην κύρια μνήμη. Εκεί η εκτέλεση των διεργασιών αυτών κοστίζουν χρονικά πολύ λιγότερο. Παράλληλα εάν υπάρχει η δυνατότητα αξιοποίησης παραπάνω από ενός σκληρού δίσκου, αξιοποιούνται παράλληλα με τον διαχωρισμό του σετ δεδομένων σε αντίστοιχα σε πλήθος ξεχωριστών σετ δεδομένων και αποστολή αυτών σε αυτούς για εγγραφή και ανάγνωση. Αυτές λοιπόν είναι οι δύο κύριες τεχνικές που βελτιστοποιούν την ταχύτητα εκτέλεσης του και δεν υπόκεινται στο επιστημονικό πεδίο της στατιστικής.

Στην περίπτωση που υπάρχουν missing values στο σετ των δεδομένων εκπαίδευσης ο αλγόριθμος XGBoost μπορεί να υπερκεράσει το εμπόδιο αυτό μέσω της τεχνικής Sparsity - Aware Split Finding η οποία εντάσσεται σε αυτόν. Αυτός ο αλγόριθμος καλείται να εκτελεστεί και να δώσει τις προβλέψεις που απαιτούνται χωρίζοντας με ένα συγκεκριμένο τρόπο τα δεδομένα και πιο συγκεκριμένα σε δύο dataset. Στο ένα βρίσκονται όσα δεν εμφανίζουν ελλείψεις και στο άλλο τα 'υγείη' δεδομένα. Η ανάλυση του αλγορίθμου δεν είναι αντικείμενο μελέτης της παρούσας εργασίας, παρά αρκούμαστε στην γνώση ότι ο συγκεκριμένος αλγόριθμος είναι υπεύθυνος για την επίλυση προβλημάτων που προκύπτουν από την αξιοποίηση σετ δεδομένων με missing values.

Στη περίπτωση που ο αλγόριθμος δέχεται ως είσοδο ένα τεράστιο σετ δεδομένων εκτελούνται ακόμη δύο αλγόριθμοι. Αυτοί είναι ο Parallel Learning και ο Weighted Quantile Sketch. Στην ουσία ένα τεράστιο σετ δεδομένων διαχωρίζεται σε μικρότερα και επεξεργάζονται παράλληλα διαφορετικοί κόμβοι από διαφορετικές υπολογιστικές μονάδες (Parallel Learning). Ένας Quantile Sketch αλγόριθμος αναλαμβάνει να συνδυάσει τις τιμές που υπολογίστηκαν σε κάθε subset ώστε να δημιουργήσει ένα προσεγγιστικό ιστόγραμμα. Αυτό το ιστόγραμμα στη συνέχεια χωρίζεται σε κάποια quantiles. Τα συγκεκριμένα είναι ζυγισμένα (weighted) και το βάρος της κάθε παρατήρησης ισούται με τη δεύτερη παράγωγη της συνάρτησης απώλειας. Αυτός είναι ο λόγος που στα προβλήματα παλινδρόμησης όλες οι παρατηρήσεις έχουν το ίδιο βάρος και οπότε δεν λαμβάνονται τελικώς υπόψιν. Αυτά τα quantiles λειτουργούν τελικά ως προσεγγιστικά κατώφλια στο χτίσιμο των δέντρων καθώς η συνήθης διαδικασία θα είχε τεράστιο χρονικό κόστος.

Κεφάλαιο 7

Επεξεργασία Ενεργειακών Δεδομένων με Μεθόδους Χρονοσειρών

7.1 Εισαγωγή

Στα προηγούμενα κεφάλαια μελετήθηκε η θεωρία και τα μοντέλα τα οποία χρησιμοποιούνται για να παραχθεί μια αξιόπιστη πρόβλεψη μιας μονοδιάστατης χρονοσειράς. Στο κεφάλαιο αυτό θα μελετηθούν 2 ξεχωριστές περιπτώσεις πρόβλεψης χρονοσειρών παραγωγής ηλεκτρικής ενέργειας. Η πρώτη περίπτωση αφορά δεδομένα παραγωγής, που παρέχονται από την Kaggle. Η άλλη πραγματοποιείται με δεδομένα παραγωγής, που προέρχονται από αιολικό πάρκο εγκατεστημένο στην περιφέρεια Θράκης όπου και λειτουργεί έως σήμερα. Παρέχονται οι χρονοσειρές με τις ποσότητες παραγόμενης ενέργειας στο επίπεδο σύνδεσης με το δίκτυο. Αυτό σημαίνει ότι οι μετρήσεις προέρχονται από τον υποσταθμό που εξυπηρετεί το κάθε αιολικό πάρκο και οποίος βρίσκεται μερικά χιλιόμετρα μακριά από αυτά.

Αξιοποιούνται τα ιστορικά διαθέσιμα δεδομένα, δημιουργείται η χρονοσειρά η οποία στη συνέχεια μελετάται ώστε να γίνει τελικά μια βραχυπρόθεσμη πρόβλεψη με χρήση γραμμικών μοντέλων *AR*, *MA*, *ARMA*, *ARIMA* και *SARIMA*. Στη συνέχεια γίνεται σύγκριση των μοντέλων, ώστε να εξαχθεί το βέλτιστο, αυτό δηλαδή που απέδωσε την καλύτερη δυνατή πρόβλεψη με βάση κάποια στατιστικά μέτρα. Σε προσπάθεια που έγινε για να βελτιωθεί η ποιότητα της πρόβλεψης που παράχθηκε αξιοποιήθηκαν επίσης μοντέλα μηχανικής μάθησης *XGB*. Δημιουργήθηκαν τρία διαφορετικά μοντέλα και αφού παρήγαγαν προβλέψεις για τα αντίστοιχα χρονικά διαστήματα με τα γραμμικά, αξιολογήθηκαν με παρόμοια στατιστικά μέτρα.

Μια σωστή πρόβλεψη παραγωγής ηλεκτρικής ενέργειας από ένα αιολικό πάρκο, αποφέρει πολλά οφέλη στην εκάστοτε εταιρία διαχείρισης και αξιοποίησης του πάρκου αυτού. Χρησιμοποιεί τη βεβαιότητα που του παρέχει μια αξιόπιστη πρόβλεψη, ώστε να μπορεί να προσφέρει με ασφάλεια τη μέγιστη δυνατή παραγωγή του πάρκου στο Χρηματιστήριο της ενέργειας. Εκεί ο παραγωγός θα επιτύχει το μέγιστο δυνατό κέρδος, καθώς είναι υποχρεωμένος να συμμετέχει και να προσφέρει τη συνολική παραγόμενη ενέργεια από το πάρκο αυτό. Μια ποσότητα,

όμως, η οποία δηλώνεται πως μπορεί να προσφερθεί πριν τη παραγωγή της και όχι εκ των υστέρων. Με τον τρόπο αυτό, επίσης, το δίκτυο μπορεί να προγραμματίσει την ένταξη ή όχι επιπλέον μονάδων παραγωγής ηλεκτρικής ενέργειας προκειμένου να αντεπεξέλθει στην ζήτηση της αγοράς.

7.2 Ανάλυση Δεδομένων Παραγωγής Ηλεκτρικής Ενέργειας (Kaggle Dataset)

Από τη διαδικτυακή πλατφόρμα επιστήμης δεδομένων Kaggle ανακτήθηκαν δεδομένα παραγωγής ηλεκτρικής ενέργειας Electric_Production σε μορφή .csv [36]. Τα δεδομένα αυτά προέρχονται από την εργασία του Sercan Yeşilöz με τίτλο Electricity Production Forecasting (ARIMA). Η χρήση τους έγινε με σκοπό τον έλεγχο ορθότητας και ποιότητας των αλγορίθμων που υλοποιήθηκαν στο πλαίσιο της παρούσας διπλωματικής εργασίας. Είναι αλγόριθμοι οι οποίοι εξετάζουν την φύση των δεδομένων, πραγματοποιούν μια στατιστική ανάλυση, μετατρέπουν την χρονοσειρά σε στάσιμη και τέλος βρίσκουν και εκπαιδεύουν τα γραμμικά μοντέλα που προσαρμόζονται κατά το βέλτιστον δυνατόν στα δεδομένα εκπαίδευσης ώστε να παράξουν ύστερα προβλέψεις επί της μέσης ποσότητας ηλεκτρικής ενέργειας που παράχθηκε σε μηνιαία βάση από τη μονάδα παραγωγής για ένα χρονικό διάστημα 12 μηνών μετά το τέλος των δεδομένων εκπαίδευσης. Δε παραλείπεται να γίνει σχολιασμός επί των αποτελεσμάτων και σύγκριση με τα αποτελέσματα της ανάλυσης του Sercan Yeşilöz. Η σύγκριση αυτή αναφέρεται στις διαφορετικές μεθοδολογίες και στα αποτελέσματα που έφερε η κάθε μια και καταλήγει στο ποια από τις δύο τελικά εκπλήρωσε καλύτερα τον σκοπό της.

Τα δεδομένα παραγωγής ηλεκτρικής ενέργειας αποτελούν μηνιαίες μετρήσεις της μέσης ποσότητας ηλεκτρικής ενέργειας που έχει παραχθεί. Το δυσάρεστο γεγονός είναι ότι δεν αναφέρονται στοιχεία σχετικά με τη γεωγραφική, τη συνολικά καλυπτόμενη περιοχή του αιολικού πάρκου και των μονάδων μέτρησης των ποσοτήτων. Είναι διαθέσιμες προς αξιοποίηση 397 μηνιαίες μετρήσεις που έχουν συλλεχθεί από 01/1985 έως 01/2018. Έτσι, δίνεται η δυνατότητα να δημιουργηθεί μια χρονοσειρά 397 παρατηρήσεων με χρονικό βήμα ενός μήνα η οποία φανερώνει την εξέλιξη της παραγωγής σε διάρκεια χρόνων.

7.2.1 Προ επεξεργασία Δεδομένων

Το πρώτο βήμα στη προσπάθεια πρόβλεψης μιας χρονοσειράς, είναι η δημιουργία της. Για να δημιουργηθεί η χρονοσειρά, πρώτα από όλα, πρέπει να φορτωθούν τα δεδομένα που υπάρχουν στο ακατέργαστο αρχείο της μορφής .csv με όνομα Electric_Production. Στο Σχήμα 7.1 φαίνονται οι πρώτες 10 γραμμές αυτού του αρχείου.

Παρατηρείται ότι κάθε στήλη έχει σαν διαχωριστικό το κόμμα (,) και μια επικεφαλίδα που την προσδιορίζει (DATE, IPG22211A2N). Στη προκείμενη περίπτωση υπάρχουν δύο στήλες, οπότε κάθε γραμμή αποτελεί εν δυνάμει ένα ζεύγος τιμών της χρονοσειράς (χρονική στιγμή, ποσότητα παραγόμενης ενέργειας). Στη πρώτη στήλη υπάρχουν οι μήνες για τους οποίους έχουν πραγματοποιηθεί μετρήσεις με επικεφαλίδα DATE. Οι μετρήσεις που αντιστοιχούν στον

```
DATE, IPG2211A2N
1/1/1985, 72.5052
2/1/1985, 70.672
3/1/1985, 62.4502
4/1/1985, 57.4714
5/1/1985, 55.3151
6/1/1985, 58.0904
7/1/1985, 62.6202
8/1/1985, 63.2485
9/1/1985, 60.5846
10/1/1985, 56.3154
```

Σχήμα 7.1: Γραμμές του ακατέργαστου αρχείου που περιέχει τα δεδομένα παραγωγής ενέργειας.

κάθε μήνα βρίσκονται στη δεύτερη στήλη με όνομα επικεφαλίδας IPG2211A2N. Εδώ, πρέπει να τονιστεί ότι δεν παρέχεται η μονάδα μέτρησης των ποσοτήτων ενέργειας που παρήχθησαν. Αυτή θα μπορούσε να είναι είτε kWh είτε MWh είτε GWh ανάλογα το μέγεθος της υπό μελέτης μονάδας παραγωγής.

Οι αλγόριθμοι με τους οποίους έγινε η προεπεξεργασία των δεδομένων της χρονοσειράς (όπως περιγράφεται παρακάτω) υλοποιήθηκαν στο προγραμματιστικό περιβάλλον της (python). Πιο συγκεκριμένα:

1. Έγινε φόρτωση των απαραίτητων βιβλιοθηκών με τις οποίες θα γίνει η ανάλυση και η πρόβλεψη των χρονοσειρών. Μεταξύ όλων οι πιο σημαντικές ήταν οι **Pandas**, **Matplotlib**, **Numpy**, **Scipy**, **Statsmodels**, **Datetime**, **Fitter** και η **Sklearn**. Κάθε μια από αυτές περιείχε συναρτήσεις και κλάσεις που εξυπηρετούσαν διαφορετικούς σκοπούς στην διαδικασία.
2. Η φόρτωση του αρχείου των δεδομένων πραγματοποιήθηκε με την συνάρτηση `read_csv`. Δημιουργείται μια δομή ετερογενών δεδομένων και δυναμικού μεγέθους **pandas DataFrame**. Στη προκείμενη περίπτωση, είναι ένας δισδιάστατος πίνακας, ο οποίος είναι μεταβλητός και μπορεί να αξιοποιήσει το σύνολο των συναρτήσεων που προσφέρει η βιβλιοθήκη ανάλυσης δεδομένων της **python** η **Pandas**. Αυτός ο πίνακας έχει για κάθε στήλη μια επικεφαλίδα και κάθε γραμμή έναν δείκτη που τις αριθμεί. Η αρχική μορφή του είναι αυτή που παρουσιάζεται στο Σχήμα 7.2 α'.
3. Στη συνέχεια, προκειμένου να δημιουργηθεί η επιθυμητή χρονοσειρά, αλλάζουμε τη μορφή του. Πιο συγκεκριμένα, αρκεί να αντικαταστήσουμε τον δείκτη που αριθμεί τις γραμμές του πίνακα με τα στοιχεία της στήλης **DATE**. Αυτή περιέχει της ημερομηνία δειγματοληψίας της κάθε παρατήρησης της χρονοσειράς (στήλη IPG2211A2N στο Σχήμα 7.1). Για ευκολία στη μεταχείριση τους αργότερα, πρότερα της αλλαγής του δείκτη, οι επικεφαλίδες των δύο στηλών, είχαν μετονομαστεί σε **'Date'** και **'energy'** αντίστοιχα. Έτσι, λαμβάνουμε τελικά τη τελική μορφή του **DataFrame** μας, όπως φαίνεται στο Σχήμα 7.2β'. Αυτή η δομή πλέον, ένα **DataFrame** δηλαδή το οποίο έχει μια στήλη μόνο στην οποία βρίσκονται οι τιμές της μεταβλητής ενδιαφέροντος μας και όπου για την κάθε μια υπάρχει σαν δείκτης η ημερομηνία δειγματοληψίας της (χρήση συνάρτησης `set_index`), αποτελεί τη μονοδιάστατη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας μηνιαίου βήματος που θα μελετήσουμε και θα παράγουμε προβλέψεις.

4. Τελευταίο βήμα στην προεπεξεργασία των δεδομένων ήταν η αναζήτηση τυχών τιμών που μπορεί να έλλειπαν από το `dataset` και έχουν τιμή 'NaN' ή τιμών που έφεραν μηδενικές ή αρνητικές τιμές. Στη περίπτωση αυτή δε βρέθηκε κάποια τέτοια περίπτωση παρά την εφαρμογή των συναρτήσεων `isnull` και `values.any` επί του `DataFrame` που αποτελεί τη δομή που περιγράφει τη χρονοσειρά μας.

Index	DATE	IPG2211A2N
0	1/1/1985	72.5052
1	2/1/1985	70.672
2	3/1/1985	62.4502
3	4/1/1985	57.4714
4	5/1/1985	55.3151
5	6/1/1985	58.0904
6	7/1/1985	62.6202
7	8/1/1985	63.2485
8	9/1/1985	60.5846
9	10/1/1985	56.3154
10	11/1/1985	58.0005

(α') Αρχικός Πίνακας Δεδομένων
(`DataFrame`).

Date	energy
1985-01-01 00:00:00	72.5052
1985-02-01 00:00:00	70.672
1985-03-01 00:00:00	62.4502
1985-04-01 00:00:00	57.4714
1985-05-01 00:00:00	55.3151
1985-06-01 00:00:00	58.0904
1985-07-01 00:00:00	62.6202
1985-08-01 00:00:00	63.2485
1985-09-01 00:00:00	60.5846
1985-10-01 00:00:00	56.3154

(β') Τελικός Πίνακας Δεδομένων
(`DataFrame`). Εμπεριέχει τη Χρονοσειρά
Παραγωγής Ηλεκτρικής Ενέργειας.

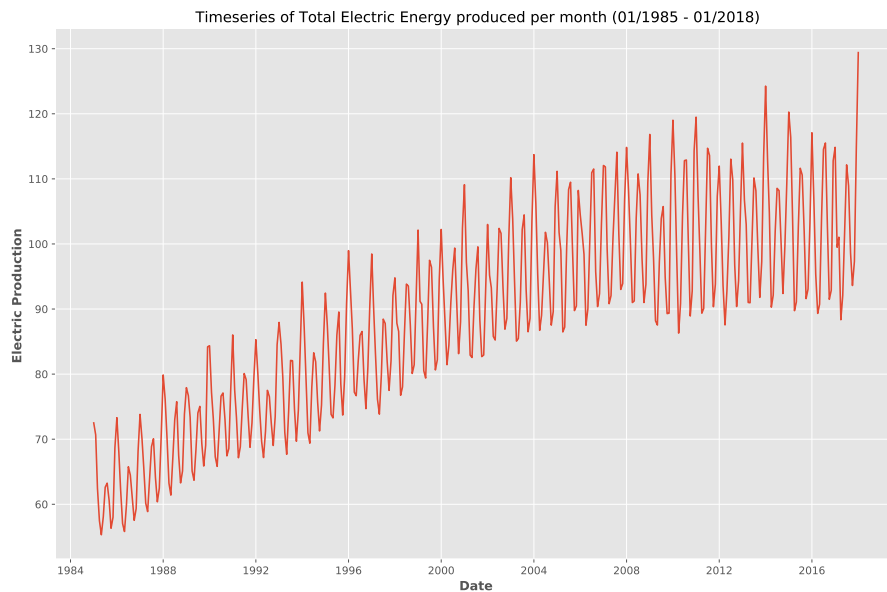
Σχήμα 7.2: Δημιουργία Χρονοσειράς με `pandas DataFrame`

7.2.2 Διερευνητική Ανάλυση και Έλεγχος Στασιμότητας της Χρονοσειράς

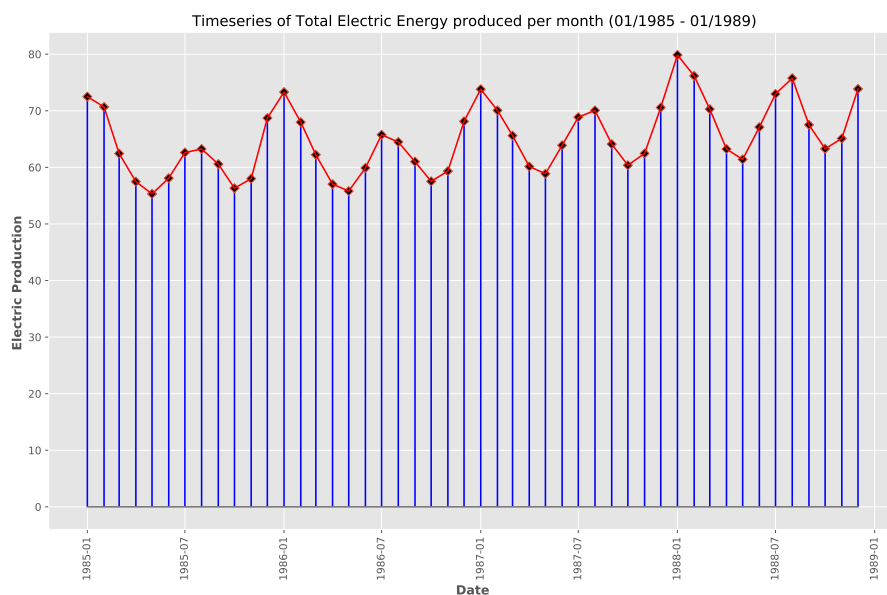
Επόμενο σημαντικό βήμα που προηγείται της πρόβλεψης, είναι η διερευνητική ανάλυση των δεδομένων της χρονοσειράς που προ επεξεργαστήκαμε προηγουμένως. Με την ανάλυση αυτή, μπορούν να εξαχθούν ασφαλή συμπεράσματα περί της συμπεριφοράς και των ιδιοτήτων της χρονοσειράς συναρτήσει του χρόνου ως προς την στασιμότητα της. Είναι μια διαδικασία η οποία υλοποιείται σταδιακά στο προγραμματιστικό περιβάλλον της `python`. Χρησιμοποιούνται συναρτήσεις από τις βιβλιοθήκες `Matplotlib`, `Statsmodels`, `Pandas` και `Fitter` της `python`. Με αυτές καθίσταται δυνατή η επεξεργασία των δεδομένων, η δημιουργία των απαραίτητων γραφημάτων, η διενέργεια ελέγχων στασιμότητας της χρονοσειράς για την εξαγωγή χρήσιμων και σίγουρων αποτελεσμάτων. Οι διεργασίες και τα αποτελέσματα της διαδικασίας αυτής παρουσιάζονται παρακάτω με τη σειρά που εκτελέστηκαν.

Αρχικά παράγεται το γράφημα απεικόνισης της υπό μελέτης χρονοσειράς, όπως φαίνεται στο Σχήμα 7.3α'. Με την οπτικοποίηση των δεδομένων, μελετάται εμπειρικά η χρονοσειρά. Στο Σχήμα 7.3α', παρουσιάζεται ολόκληρη η χρονοσειρά. Στον οριζόντιο άξονα αποτυπώνονται οι χρονικές στιγμές στις οποίες δειγματοληπτήθηκαν οι παρατηρήσεις της χρονοσειράς, των οποίων οι δειγματοληπτημένες τιμές φαίνονται στον κατακόρυφο άξονα. Καθώς τα δεδομένα αριθμούν σε πλήθος τα 397, αποδεικνύεται σωστά η αρχική υπόθεση ότι είναι μια χρονοσειρά μήκους $n=397$. Ωστόσο, λόγω του μεγάλου πλήθους δεν φαίνεται τόσο καθαρά το χρονικό βήμα της χρονοσειράς ή αλλιώς η χρονική υστέρηση τ που υφίσταται ανάμεσα σε δύο διαδοχικές παρατηρήσεις. Για το λόγο αυτό παράγεται το Σχήμα 7.3β'. Σε αυτό παρουσιάζεται ένα τμήμα της χρονοσειράς. Πιο συγκεκριμένα, παρουσιάζονται οι πρώτες 48 παρατηρήσεις. Εάν παρατηρηθεί προσεκτικά, χρονικό βήμα της χρονοσειράς είναι όντως ο ένας μήνας. Επι-

βεβαιώνεται, έτσι η υπόθεση ότι τα δεδομένα που αξιοποιήθηκαν αποτελούν μια χρονοσειρά μηνιαίου βήματος μέσω παραγόμενων ποσοτήτων ηλεκτρικής ενέργειας.



(α') Απεικόνιση της χρονοσειράς παραγωγής ηλεκτρικής ενέργειας.



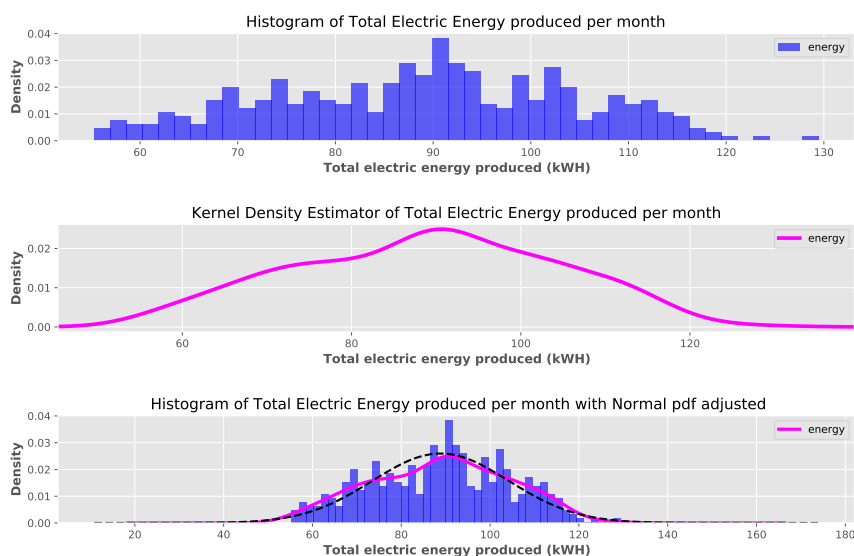
(β') Απεικόνιση των πρώτων 48 παρατηρήσεων της χρονοσειράς προκειμένου να καταστεί εύκολη η εκτίμηση του βήματος της χρονοσειράς.

Σχήμα 7.3: Απεικόνιση χρονοσειράς

Η ανάγνωση του Σχήματος 7.3α' παρέχει επιπλέον σημαντικές πληροφορίες που αφορούν την φύση της χρονοσειράς. Φαίνεται ότι αποτελεί μια μη στάσιμη χρονοσειρά. Πρώτον, είναι εμφανές, ότι οι ποσότητες ηλεκτρικής ενέργειας που παράγονται στη πάροδο του χρόνου αυξάνονται σταθερά. Δεύτερον, παρατηρείται η ύπαρξη ενός κυκλικά επαναλαμβανόμενου μοτίβου. Η μεν πρώτη παρατήρηση μεταφράζεται ως τάση, ενώ η δεύτερη ως εποχικότητα. Αν

παρατηρηθεί το Σχήμα 7.3β', ο ισχυρισμός περί ύπαρξης ετήσιας εποχικότητας στη χρονοσειρά μοιάζει πάρα πολύ πιθανός. Σταθερά κάθε χρόνο τους χειμερινούς και τους καλοκαιρινούς μήνες εμφανίζει τις υψηλότερες τιμές της, ενώ τους ανοιξιάτικους και φθινοπωρινούς τις χαμηλότερες. Με το πέρασμα των χρόνων είναι λογικό να αυξάνεται η παραγωγή καθώς όσο ο πληθυσμός αυξάνεται, τόσο αυξάνονται οι ενεργειακές ανάγκες του. Η αύξησης στη παραγωγή επίσης ανά εποχή δείχνουν επίσης να έχουν λογική. Συνάδουν με περιόδους διακοπών και εποχικών περιόδων όπου η ζήτηση για ενέργεια αυξάνεται (θέρμανση, ψύξη κλπ.). Έτσι, κατανοείται καλύτερα, το σύστημα που μελετάται και η συμπεριφορά του στο χρόνο.

Η ύπαρξη των συνιστωσών τάσης και περιοδικότητας στην υπό μελέτη χρονοσειρά, αποτελούν σαφές ενδείξεις μη στασιμότητας. Βέβαια, παραπάνω έλεγχοι πρέπει να πραγματοποιηθούν, καθώς το παραπάνω συμπέρασμα είναι προϊόν εμπειρικού ελέγχου. Αρχικά μια χρονοσειρά, η οποία ακολουθεί ή προσεγγίζει σε πολύ μεγάλο βαθμό την κανονική κατανομή, θα μπορούσε να αποτελεί μια στάσιμη χρονοσειρά χωρίς βέβαια να την καθιστά στάσιμη με απόλυτη σιγουριά το γεγονός αυτό. Με την παραγωγή του ιστογράμματος μιας χρονοσειράς, είναι δυνατόν να εξεταστεί και να διαπιστωθεί το είδος της περιθώριας κατανομής που ακολουθούν τα δεδομένα που εξετάζονται. Στο Σχήμα 7.4 παρέχεται το ιστόγραμμα της χρονοσειράς που εξετάζεται στη προκειμένη περίπτωση.

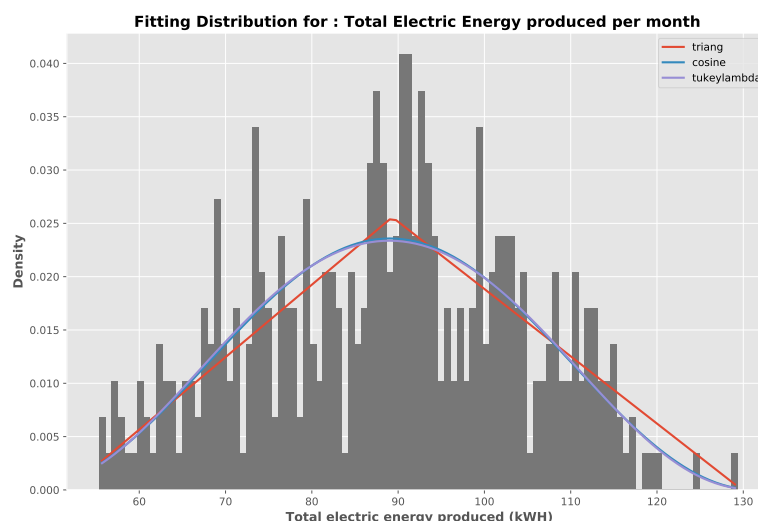


Σχήμα 7.4: Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας (πρώτο). Kernel Density Estimator (μεσαίο). Συνδυασμός των δύο από πάνω γραφημάτων και σύγκριση των δεδομένων με την κανονικής κατανομής (τελευταίο).

Σύμφωνα με αυτό η κατανομή που ακολουθούν τα δεδομένα μας δεν προσεγγίζει σε καμιά περίπτωση τη κανονική. Το γεγονός και μόνο ότι, φαίνεται να σχηματίζονται παραπάνω από μια κορυφές, δικαιολογεί τον ισχυρισμό αυτόν. Προφανώς, επειδή ανά εποχές είναι διαφορετική η ανάγκη για παραγωγή ηλεκτρικής ενέργειας και επειδή με το πέρασμα των χρόνων η παραγωγή αυξανόταν σταθερά διαμορφώθηκαν οι κορυφές αυτές. Προκειμένου να εκτιμηθεί η

κατανομή πυκνότητας πιθανότητας που ακολουθεί, δημιουργείται ακόμη το μεσαίο διάγραμμα του Σχήματος 7.4. Σε αυτό το διάγραμμα, η εκτίμηση που αφορά την συνάρτηση πυκνότητας πιθανότητας της χρονοσειράς, γίνεται μη παραμετρικά. Η εκτιμώμενη συνάρτηση πιθανότητας απέχει από την κανονική όπως φαίνεται στο διάγραμμα αυτό όσο και στο από κάτω του. Το τελευταίο διάγραμμα στο Σχήμα 7.4 παρέχει συγκεντρωτικά τις γνώσεις των παραπάνω διαγραμμάτων και σχιαγραφείται επίσης η κανονική κατανομή που θα έπρεπε να προσεγγίζει η εξεταζόμενη χρονοσειρά. Συμπερασματικά, η κατανομή της χρονοσειράς δεν είναι κανονική κάτι το οποίο υποδεικνύει ότι η εφαρμογή γραμμικών μοντέλων για τη παραγωγή προβλέψεων της χρονοσειράς θα είναι αναποτελεσματική στη παρούσα φάση.

Τα διαγράμματα στο Σχήμα 7.4 μπορεί να δίνουν κάποιες σημαντικές πληροφορίες ως προς την κατανομή που ακολουθούν τα δεδομένα αλλά δεν είναι αρκετές. Γνωρίζουμε ότι δεν ακολουθούν κανονική κατανομή αλλά δεν γνωρίζουμε το είδος της. Για αυτόν τον λόγο χρησιμοποιούμε από τη βιβλιοθήκη *Fitter* τις συναρτήσεις *get_distributions*, *fit*, *get_best* και *summary*. Με αυτές υπολογίσθηκαν και αποτυπώθηκαν τελικά οι τρεις κατανομές πυκνότητας πιθανότητας που προσεγγίζουν καλύτερα από όλες την κατανομή που ακολουθούν τα δεδομένα μας. Το αποτέλεσμα αυτό, φαίνεται στο Σχήμα 7.5.



Σχήμα 7.5: Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στα δεδομένα παραγωγής ηλεκτρικής ενέργειας.

Οι κατανομές που φαίνονται να προσεγγίζουν καλύτερα τον τρόπο που κατανέμονται τα δεδομένα μας είναι οι:

1. Κατανομή τριγωνική
2. Κατανομή συνημιτόνου
3. Κατανομή Tukey Lambda

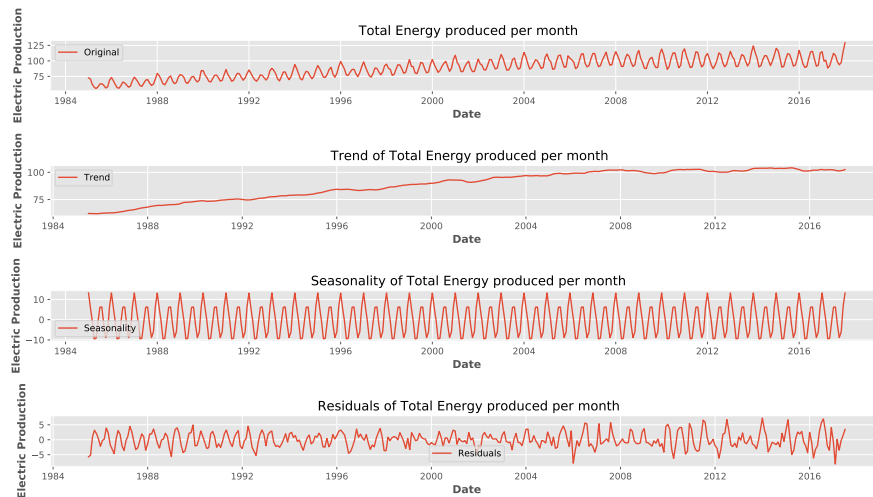
Πρόκειται για κατανομές οι οποίες παρουσιάζουν μια συμμετρία γύρω από τη μέση τιμή της χρονοσειράς. Μια μέση τιμή, όμως, που δεν μένει σταθερή στο χρόνο λόγω της συνιστωσών τάσης και περιοδικότητας που υφίστανται. Πέρα από την συμμετρία που φέρουν, δεν υπάρχει άλλο χαρακτηριστικό που να τις φέρει κοντά στη κανονική κατανομή. Φαίνεται, λοιπόν, πως τα δεδομένα μας, ίσως μπορούν να προσεγγίσουν σε μεγαλύτερο βαθμό, που είναι και το πλέον επιθυμητό, την κανονική κατανομή. Βέβαια για να γίνει αυτό, ορισμένες αλλαγές πρέπει να γίνουν στη χρονοσειρά μας, οι οποίες θα παρουσιαστούν και θα εξηγηθούν στην επόμενη υπό ενότητα.

Πριν μελετηθούν οι μεταβολές, πρέπει να μελετηθούν περαιτέρω οι συνιστώσες που καθιστούν τη χρονοσειρά μας μη στάσιμη. Με γνώμονα το σκοπό αυτό δημιουργούνται δύο διαγράμματα, τα οποία παρουσιάζονται στα Σχήματα 7.6 και 7.9. Το μεν πρώτο σκιαγραφεί την χρονοσειρά καθώς την αποδομεί, ενώ το δεύτερο αποτελεί το διάγραμμα αυτοσυσχέτισης της.

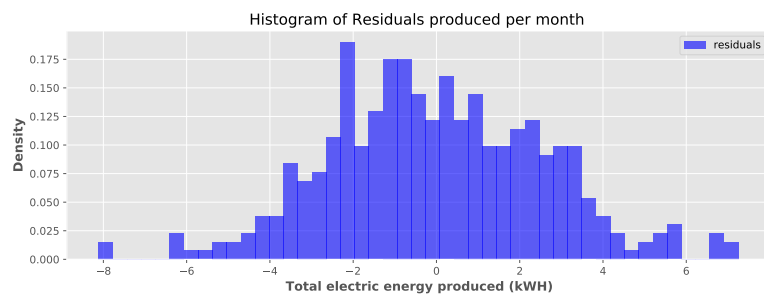
Στο Σχήμα 7.6 παρατηρείται η αποδόμηση της χρονοσειράς παραγωγής ηλεκτρικής ενέργειας. Θεωρείται ότι ακολουθεί το αθροιστικό μοντέλο και έχει άρα τη μορφή $X_t = m_t + s_t + R_t$. Η ντετερμινιστική τάση m_t της χρονοσειράς X_t , όπως παρουσιάζεται στο δεύτερο διάγραμμα του Σχήματος 7.6, φαίνεται, να είναι αυξητική. Αυξάνεται σταθερά στο χρόνο. Ίσως με τη βοήθεια ενός πολυωνύμου κάποιου βαθμού p να είναι δυνατή η εκτίμηση της ή με την εφαρμογή των μεθόδων των διαφορών πρώτης τάξης να είναι δυνατή η απαλοιφή της. Το διάγραμμα κάτω από αυτό της τάσης στο Σχήμα 7.6, παρουσιάζει την εποχικότητα s_t που υπάρχει στη χρονοσειρά. Φαίνεται, να υπάρχει ετήσια περιοδικότητα. Εύκολα παρατηρείται ότι, ανά τέσσερα χρόνια ολοκληρώνονται τέσσερις περίοδοι. Τέλος, στο ίδιο Σχήμα, το τελευταίο διάγραμμα, το οποίο αποτυπώνει τα υπολείμματα R_t της χρονοσειράς δεν μπορεί να μας δώσει πολλές πληροφορίες ακόμα, πέραν του ότι φαίνεται να μην είναι στάσιμα. Πιο συγκεκριμένα, παρατηρείται ότι στη διάρκεια του χρόνου δεν παραμένει σταθερή η διασπορά τους (π.χ. παρατηρούνται απότομες, μεγάλες και άνισες διακυμάνσεις σε διαφορετικά χρονικά διαστήματα). Συνεπώς, η μελέτη του Σχήματος 7.6, δηλαδή του διαγράμματος αποδόμησης της χρονοσειράς, αποφέρει πολύ σημαντικές και ακριβείς πληροφορίες για τις συνιστώσες που καθιστούν μη στάσιμη την χρονοσειρά.

Για να μελετηθούν καλύτερα τα υπόλοιπα της χρονοσειράς, R_t , δηλαδή η πληροφορία που απομένει για τη χρονοσειρά αφού απαλειφθεί η τάση και η περιοδικότητα παράγεται το ιστόγραμμα τους. Αυτό αποτυπώνεται από το Σχήμα 7.7. Σε περίπτωση, που αυτά είναι στάσιμα, το ιστόγραμμα τους δείχνει ότι ακολουθούν μια κατανομή της οποίας τα γνωρίσματα είναι η σταθερή μέση τιμή και διασπορά, όπως για παράδειγμα η κανονική κατανομή. Από το Σχήμα 7.7 γίνεται αντιληπτό ότι πρόκειται για ένα ιστόγραμμα που δείχνει ότι ναι μεν απέχει η κατανομή πιθανότητας από την κανονική αλλά η απόκλιση αυτή δεν μπορεί να θεωρηθεί τεράστια. Συνεπώς, αντιλαμβανόμαστε ότι υπάρχει μεγάλη πιθανότητα με τους κατάλληλους μετασχηματισμούς να φέρουμε τη χρονοσειρά σε μια κατάσταση η οποία θα μπορέσει να την καταστήσει στάσιμη.

Το διάγραμμα αυτοσυσχέτισης είναι ένα πολύ χρήσιμο εργαλείο για τον έλεγχο ύπαρξης τάσης και περιοδικότητας στη χρονοσειρά. Γίνεται, δηλαδή, να προσδιοριστεί εάν και κατά



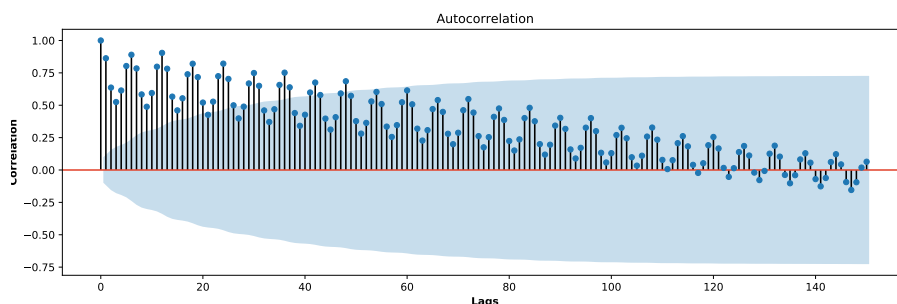
Σχήμα 7.6: Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας.



Σχήμα 7.7: Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας.

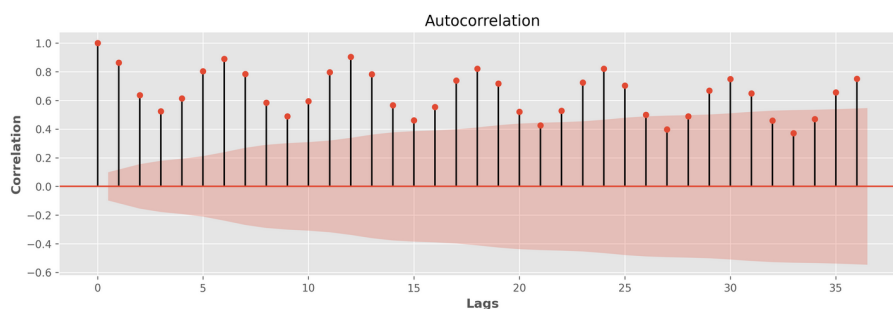
πόσο είναι στάσιμη η υπό μελέτη χρονοσειρά. Στη προσπάθεια αυτή, παρατηρούνται τα Σχήματα 7.8 και 7.9. Τόσο το ένα όσο και το άλλο αποτελούν διαγράμματα αυτοσυσχέτισης της χρονοσειράς που μελετάται εκφρασμένο σε διαφορετικό πλήθος χρονικών υστερήσεων (lags). Παρατηρείται εκθετική πτώση των τιμών της αυτοσυσχέτισης μετά από λίγες χρονικές υστερήσεις lags. Συνήθως μετά τα lags 1 ή 2, αναμένεται να μην υπάρχει τιμή αυτοσυσχέτισης που να βρίσκεται εκτός των διαστημάτων εμπιστοσύνης. Σε αντίθετη περίπτωση, όταν δηλαδή, εμφανίζονται υψηλές τιμές αυτοσυσχέτισης και η μείωση τους μεταξύ χρονικών υστερήσεων είναι αργή, τότε η χρονοσειρά δεν είναι στάσιμη, όπως φαίνεται και στη περίπτωση αυτή. Η ύπαρξη τάσης προκαλεί την αργή αυτή μείωση, ενώ αν ακόμη υπάρχει περιοδικότητα αναμένεται να υπάρχουν ταλαντώσεις. Η περίοδος των ταλαντώσεων αυτών υπολογίζεται σε χρονικές υστερήσεις lags και οι κορυφές της εντοπίζονται σε lags πολλαπλάσια της περιόδου.

Στο Σχήμα 7.8, είναι εύκολο να αναγνωρισθούν ενδείξεις που μας οδηγούν στο συμπέρασμα ότι η προκείμενη χρονοσειρά είναι μη στάσιμη αφού εμφανίζει τάση και περιοδικότητα. Οι τιμές συσχέτισης μεταξύ χρονικών υστερήσεων παρουσιάζουν αργή μείωση, απόδειξη ύπαρξης τάσης. Επιπλέον, τα lags 1,6,12 αποτελούν τις κορυφές μιας ταλάντωσης που έχει περίοδο ίση με 12 lags, δηλαδή 12 μήνες ή αλλιώς ετήσια εποχικότητα, όπως φαίνεται καλύτερα από το Σχήμα 7.9. Τα διαστήματα εμπιστοσύνης σχηματίζουν μια μπλε κωνοειδή περιοχή στο



Σχήμα 7.8: Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας -ACF plot (150 lags).

διάγραμμα αυτοσυσχέτισης. Στις υστερήσεις στις οποίες ο συντελεστής αυτοσυσχέτισης λαμβάνει τιμή η οποία εμπίπτει εντός αυτής της περιοχής, δεν θεωρείται στατιστικά σημαντικός, παρά ότι είναι μηδενικός [6]. Έτσι λοιπόν, με τη βοήθεια του διαγράμματος αυτοσυσχέτισης επιβεβαιώθηκαν και ισχυροποιήθηκαν οι προηγούμενοι ισχυρισμοί μας.



Σχήμα 7.9: Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας - ACF plot.

Τελευταίο βήμα, στη διαδικασία ελέγχου ύπαρξης στασιμότητας σε μια χρονοσειρά είναι η διενέργεια του ελέγχου **Augmented-Dickey Fuller (ADF) Test**. Με βάση τα όσα έχουν εξετασθεί, αναμένεται να γίνει αποδεκτή η υπόθεση H_0 , δηλαδή ότι, η χρονοσειρά μας δεν είναι στάσιμη. Παρακάτω στο Σχήμα 7.10 παρουσιάζεται το αποτέλεσμα της εκτέλεσης του ελέγχου ADF στην python το οποίο είναι και αυτό που περιμέναμε.

```
ADF Statistic: -2.257
p-value: 0.186
#####
Critical Values:
1%: -3.448
5%: -2.869
10%: -2.571
#####
ADF > Critical Values
Failed to reject null hypothesis, time series is non-stationary.
```

Σχήμα 7.10: Αποτέλεσμα του ελέγχου στασιμότητας Augmented-Dickey Fuller Test.

Φαίνεται ότι οι p-value που υπολογίζονται είναι μεγαλύτερες από το όριο του 0.05 και ότι επίσης τα στατιστικά ADF είναι μεγαλύτερα σε τιμή από τις οριακές τιμές (critical values).

Έτσι, η μηδενική υπόθεση H_0 : μη στάσιμη χρονοσειρά, δεν απορρίπτεται και αποδεικνύεται οριστικά, ότι εξετάζεται μια μη στάσιμη χρονοσειρά.

7.2.3 Μετατροπή της Χρονοσειράς σε Στάσιμη

Με την διερευνητική ανάλυση που έγινε προηγουμένως, εξετάσθηκε και εκτιμήθηκε με μεγάλη βεβαιότητα, ότι η χρονοσειρά είναι μη στάσιμη, απόρροια των συνιστώσεων τάσης και περιοδικότητας που συναντώνται σε αυτήν. Εάν η χρονοσειρά παραμείνει ως έχει, η πρόβλεψη μελλοντικών παρατηρήσεων της χρονοσειράς με γραμμικά μοντέλα πρόβλεψης, θα είναι αναποτελεσματική. Η μετατροπή της χρονοσειράς από μη στάσιμη σε στάσιμη, είναι ο επόμενος σημαντικός στόχος στην στατιστική ανάλυση των δεδομένων παραγωγής ηλεκτρικής ενέργειας που δημιουργούν την εξεταζόμενη χρονοσειρά. Ο στόχος αυτός θα επιτευχθεί με την απαλοιφή της τάσης και της περιοδικότητας που περιέχονται επί της χρονοσειράς μας και τη σταθεροποίηση της διακύμανσης της.

Για αρχή, με τη βοήθεια των μετασχηματισμών δύναμης Box-Cox, γίνεται να μετασχηματισθούν τα δεδομένα μας. Σαν μετασχηματισθούν με τη βοήθεια ενός τέτοιου μετασχηματισμού, είναι δυνατόν η νέα χρονοσειρά, που προκύπτει, να είναι πλέον πιο κοντά στη στασιμότητα. Επίσης, μετά από έναν μετασχηματισμό δύναμης είναι πιθανό η κατανομή πυκνότητας πιθανότητας της προκύπτουσας χρονοσειράς, να αποκλίνει λιγότερο από την κανονική. Είναι ανάγκη να ακολουθούν τα υπόλοιπα της χρονοσειράς την κανονική κατανομή, καθώς μόνο υπό αυτήν προϋπόθεση τα γραμμικά μοντέλα πρόβλεψης (AR, MA, ARMA, ARIMA, SARIMA) παράγουν προβλέψεις εύστοχες στις οποίες ο βαθμός εμπιστοσύνης τους είναι μεγάλος. Ανάλογα την τιμή της παραμέτρου λ , πραγματοποιείται ο ανάλογος μετασχηματισμός δύναμης. Στη προκείμενη περίπτωση δοκιμάστηκαν όλοι οι πιθανοί μετασχηματισμοί δύναμης Box-Cox, που αναφέρονται στον Πίνακα 5.2, ώστε να επιλεγεί ο καταλληλότερος.

Η επιλογή του κατάλληλου μετασχηματισμού έγινε, αφού μελετήθηκαν και συγκρίθηκαν τα ιστογράμματα των προκύπτουσών χρονοσειρών και τα γραφήματα απεικόνισης χρονοσειράς, διάσπασης χρονοσειράς και αυτοσυσχέτισης χρονοσειράς. Δε παραλείφθηκε επίσης να γίνει ο απαραίτητος έλεγχος ADF. Οι μετασχηματισμένες χρονοσειρές, οι οποίες φάνηκε να μην είναι στάσιμες από τα παραπάνω διαγράμματα και που δεν πέρασαν τον έλεγχο ADF και φαίνονται στο Σχήμα 7.11 δεν εξετάσθηκαν περαιτέρω.

Από όλους τους μετασχηματισμούς οι μόνοι που έφεραν το επιθυμητό αποτέλεσμα ήταν τελικά αυτοί του αντίστροφου μετασχηματισμού (reciprocal) και του μετασχηματισμού που φέρνει τις τιμές των αρχικών παρατηρήσεων σε κλίμακα αντίστροφης τετραγωνικής ρίζας (reciprocal square root). Βέβαια, επιλέχθηκε να εξετασθεί ακόμη ο λογαριθμικός μετασχηματισμός. Ο λογαριθμικός προτιμήθηκε έναντι των υπολοίπων, που απορρίφθηκαν, επειδή έχει την ιδιότητα να σταθεροποιεί την διασπορά. Την επίδραση αυτή φανερώνει η μελέτη των Σχημάτων 7.12 και 7.13. Στο πρώτο εμφανίζονται τα διαγράμματα απεικόνισης της χρονοσειράς πριν και μετά το λογαριθμικό μετασχηματισμό και στο δεύτερο τα αντίστοιχα ιστογράμματα πριν και μετά την μεταφορά των τιμών των αρχικών παρατηρήσεων σε λογαριθμική κλίμακα.

Αν και δεν προσεγγίζεται απόλυτα η κανονική κατανομή, οι διακυμάνσεις στις τιμές της

```

Log Transformation ADF test:
ADF Statistic: -3.145
p-value: 0.023373401033
#####
Critical Values:
1%, -3.448
Critical Values:
5%, -2.869
Critical Values:
10%, -2.571
#####
ADF > Critical Values
Failed to reject null hypothesis, time series is non-stationary.
-----

Square Root Transformation ADF test::
ADF Statistic: -2.687
p-value: 0.076252340647
#####
Critical Values:
1%, -3.448
Critical Values:
5%, -2.869
Critical Values:
10%, -2.571
#####
ADF > Critical Values
Failed to reject null hypothesis, time series is non-stationary.
-----

Reciprocal Square root Transformation ADF test::
ADF Statistic: -3.617
p-value: 0.005451781722
#####
Critical Values:
1%, -3.448
Critical Values:
5%, -2.869
Critical Values:
10%, -2.571
#####
ADF < Critical Values
p-value < 0.05
Reject null hypothesis, time series is stationary.
-----

```

(α')

```

Reciprocal Transformation ADF test:
ADF Statistic: -4.085
p-value: 0.001026836995
#####
Critical Values:
1%, -3.448
Critical Values:
5%, -2.869
Critical Values:
10%, -2.571
#####
ADF < Critical Values
p-value < 0.05
Reject null hypothesis, time series is stationary.
-----

Cube Root Transformation ADF test:
ADF Statistic: -2.838
p-value: 0.053100852917
#####
Critical Values:
1%, -3.448
Critical Values:
5%, -2.869
Critical Values:
10%, -2.571
#####
ADF > Critical Values
Failed to reject null hypothesis, time series is non-stationary.
Auto Box-Cox Transformation ADF test:
ADF Statistic: -2.196
p-value: 0.207822293277
#####
Critical Values:
1%, -3.448
Critical Values:
5%, -2.869
Critical Values:
10%, -2.571
#####
ADF > Critical Values
Failed to reject null hypothesis, time series is non-stationary.
-----

```

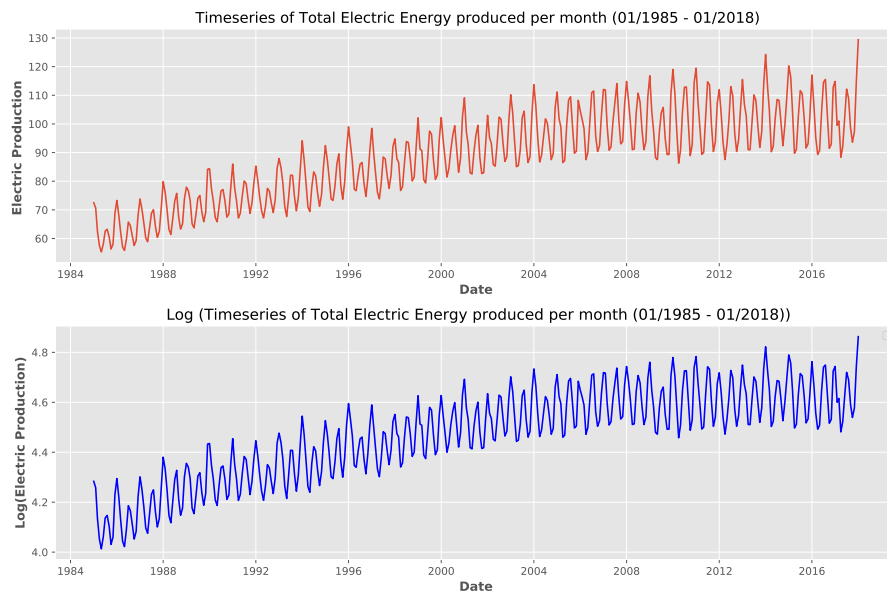
(β')

Σχήμα 7.11: Διενέργεια στατιστικού ελέγχου ADF σε όλων των ειδών τα μετασχηματισμένα δεδομένα.

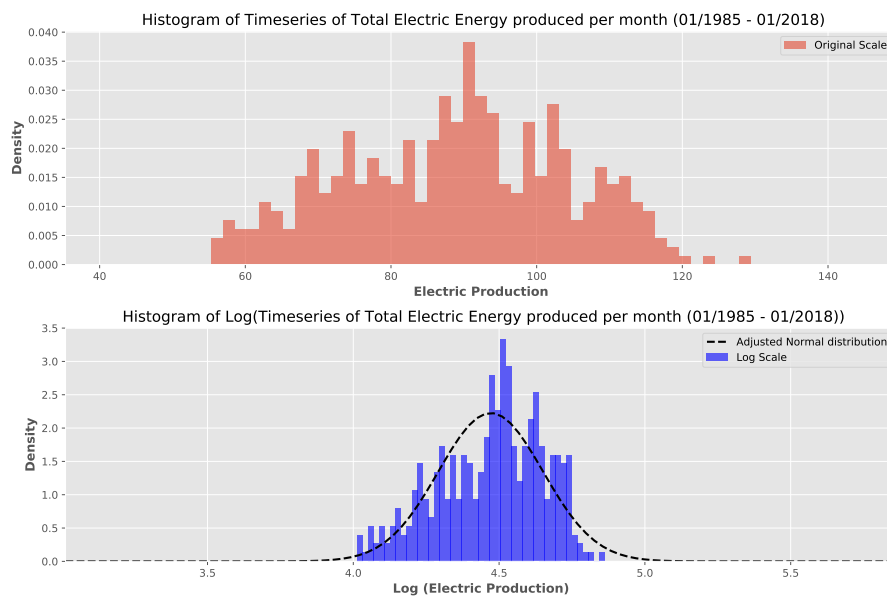
χρονοσειράς παραγωγής φαίνεται να έχουν σταθεροποιηθεί. Μειώθηκε το πλήθος των απότομων διακυμάνσεων ενώ το εύρος τιμών της πλέον είναι μικρότερο. Σημαντικό, δε, να τονιστεί ότι ο συγκεκριμένος μετασχηματισμός, δεν αντιμετώπισε την τάση και την περιοδικότητα που υπήρχε από πριν. Επίσης, όπως φαίνεται από το ιστόγραμμα των υπολοίπων της λογαριθμικά μετασχηματισμένης χρονοσειράς που παράχθηκε και απεικονίζεται στο Σχήμα 7.14, η περιθώρια κατανομή τους αποκλίνει της κανονικής σε μικρότερο βαθμό από πριν.

Ο έλεγχος ADF για τα μετασχηματισμένα δεδομένα μας με χρήση αντίστροφου και ρίζα αντιστρόφου μετασχηματισμού, υπέδειξε ότι σχηματίζουν μια στάσιμη χρονοσειρά. Ωστόσο, το γεγονός ότι δεν έχει αντιμετωπιστεί η περιοδικότητα και κυρίως η τάση, εγείρει προβληματισμούς ως προς την ορθότητα του αποτελέσματος. Γι' αυτό το λόγο κατασκευάζεται μια σειρά γραφημάτων προς μελέτη και συλλογή χρήσιμων πληροφοριών. Πληροφορίες που θα χρηνούν αξιόλογες κατά τη μετατροπή της χρονοσειράς σε στάσιμη στην περίπτωση που δεν είναι ήδη.

Δημιουργούνται τα γραφήματα απεικόνισης της μετασχηματισμένης χρονοσειράς, τα ιστογράμματα τους και τα διαγράμματα αυτοσυσχέτισης, ώστε να ελεγχθεί εκ πρώτης άποψης η ορθότητα του αποτελέσματος αυτού. Επιλέγονται τα δεδομένα που έχουν μετασχηματιστεί με ρίζα αντίστροφου και παράγονται τα προαναφερόμενα γραφήματα, τα οποία παρουσιάζονται στο

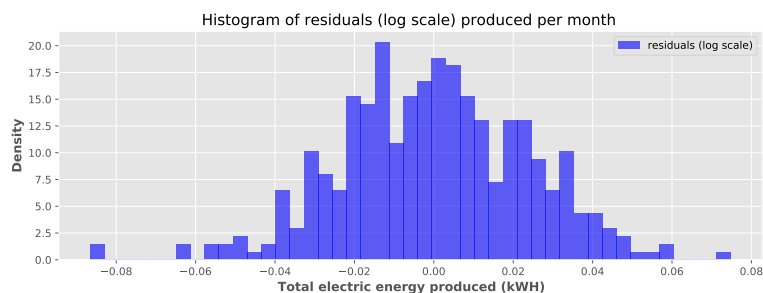


Σχήμα 7.12: Χρονοσειρές παραγωγής ηλεκτρικής ενέργειας σε κανονική (πάνω διάγραμμα) και σε λογαριθμική (κάτω διάγραμμα) κλίμακα.



Σχήμα 7.13: Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στην κανονική κλίμακα (πάνω διάγραμμα) και ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στην λογαριθμική κλίμακα (κάτω διάγραμμα).

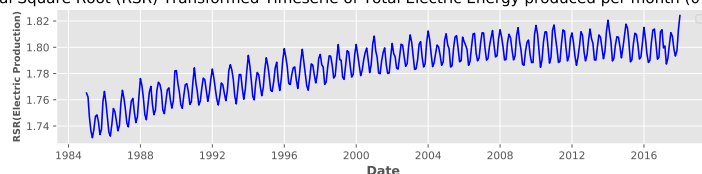
Σχήμα 7.15. Παρατηρείται ότι ο συγκεκριμένος μετασχηματισμός δε μετατρέπει σε στάσιμη τη χρονοσειρά που εκφράζει τη παραγωγή ηλεκτρικής ενέργειας με μηνιαίο βήμα. Είναι ξεκάθαρο από το διάγραμμα αυτοσυσχέτισης ότι η τάση και η περιοδικότητα συνεχίζουν να υφίστανται στη χρονοσειρά. Ακόμη, αφού η διασπορά των τιμών της φαίνεται να σταθεροποιείται μετά από ένα χρονικό διάστημα, αυτό σημαίνει ότι δεν παραμένει σταθερή στο χρόνο.



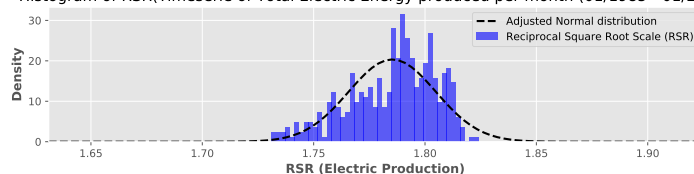
Σχήμα 7.14: Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στη λογαριθμική κλίμακα.

Στο διάγραμμα απεικόνισης παρατηρείται επίσης αύξηση της μέσης τιμής παραγωγής με το πέρασμα του χρόνου. Τέλος, το ιστόγραμμα της φανερώνει μια απόκλιση από την κανονική κατανομή. Συμπερασματικά, μπορεί ο στατιστικός έλεγχος ADF να υπέδειξε ότι ένα είδος μετασχηματισμού μπορεί να επιφέρει στασιμότητα στη χρονοσειρά, αλλά αυτό δεν επαληθεύεται πάντα. Η διενέργεια των ελέγχων στασιμότητας που έγιναν τόσο σε αυτή στη περίπτωση όσο και στη περίπτωση που τα δεδομένα βρίσκονταν στην αρχική κλίμακα τους είναι απαραίτητοι προκειμένου να μην παραπλανηθούμε.

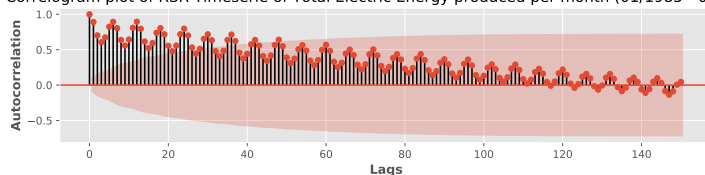
Reciprocal Square Root (RSR) Transformed Timeserie of Total Electric Energy produced per month (01/1985 - 01/2018)



Histogram of RSR(Timeserie of Total Electric Energy produced per month (01/1985 - 01/2018))



Correlogram plot of RSR Timeserie of Total Electric Energy produced per month (01/1985 - 01/2018))

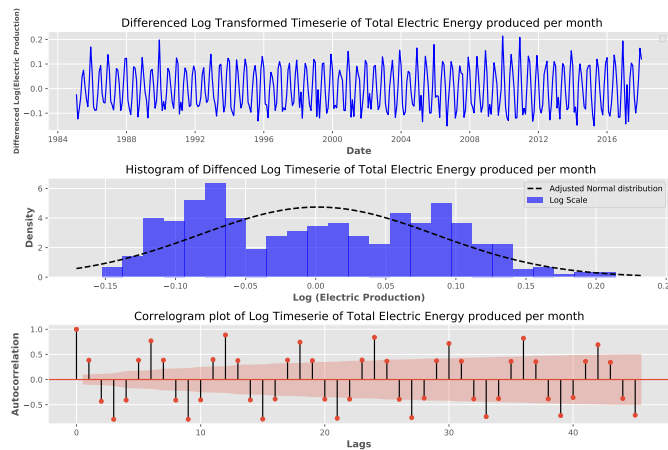


Σχήμα 7.15: Από πάνω προς τα κάτω: Διάγραμμα απεικόνισης, Ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα αντίστροφης τετραγωνικής ρίζας.

Για τις προβλέψεις που θα πραγματοποιηθούν θα χρησιμοποιηθούν τόσο τα δεδομένα στη λογαριθμική κλίμακα όσο και τα δεδομένα που προέκυψαν μετά τον μετασχηματισμό αντίστροφης τετραγωνικής ρίζας. Η χρονοσειρά που προκύπτει από τον αντίστροφο μετασχηματισμό δεν θα μελετηθεί περαιτέρω καθώς δεν οδηγεί σε αποτελέσματα ανάλογης ποιότητας με τους

άλλους δύο που προαναφέρθηκαν. Το συμπέρασμα αυτό προέκυψε μετά την ανάγνωση των βασικών γραφημάτων.

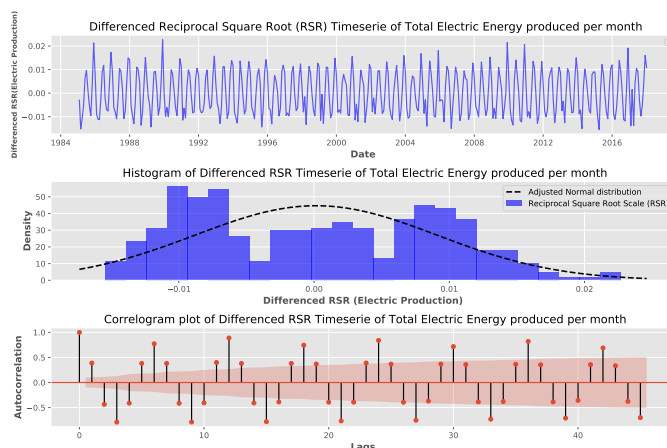
Πριν χρησιμοποιηθούν για τη παραγωγή κάποιας πρόβλεψης, θα αντιμετωπιστεί η τάση που περιέχουν οι χρονοσειρές και δημιουργούν συνθήκες μη στασιμότητας για τις χρονοσειρές. Για να εξαλειφθεί η τάση, εφαρμόζεται η μέθοδος των πρώτων διαφορών. Αφού εφαρμόζεται η μέθοδος αυτή, παράγονται τα κατάλληλα διαγράμματα, ώστε να ελεγχθεί εκ νέου η στασιμότητα της χρονοσειράς. Στο Σχήμα 7.16 παρουσιάζονται το διάγραμμα απεικόνισης, το ιστόγραμμα και το διάγραμμα αυτοσυσχέτισης της χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα, ενώ στο Σχήμα 7.17 παρουσιάζονται τα αντίστοιχα διαγράμματα της χρονοσειράς που δημιουργούν τα δεδομένα που έχουν μετασχηματιστεί χάριν στο μετασχηματισμό αντίστροφης τετραγωνικής ρίζας. Υπενθυμίζεται ότι τόσο στην μια όσο και στην άλλη περίπτωση έχει εφαρμοστεί η μέθοδος των πρώτων διαφορών.



Σχήμα 7.16: Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα λογαριθμική αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.

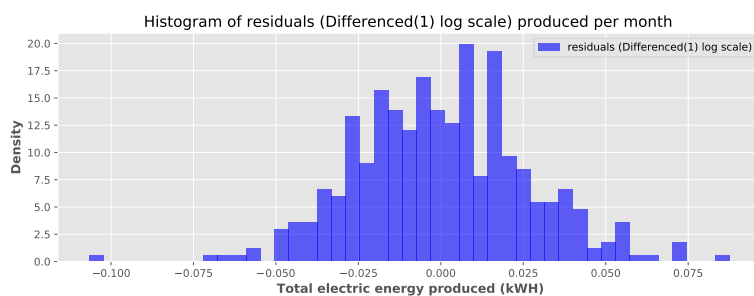
Με μια πρώτη ανάγνωση των Σχημάτων 7.16 και 7.17 γίνεται αντιληπτό ότι η τάση εξαλείφεται. Τα διαγράμματα που απεικονίζουν τις χρονοσειρές αλλά και τα διαγράμματα αυτοσυσχέτισης της κάθε χρονοσειράς δικαιολογούν το γεγονός αυτό. Καθώς μελετώνται τα διαγράμματα απεικόνισης, παρατηρείται, ότι η διασπορά των τιμών φαίνεται να σταθεροποιείται ακόμα παραπάνω. Το ίδιο συμβαίνει στη μέση τιμή της χρονοσειράς, η οποία πλέον δεν αυξάνεται στη πάροδο του χρόνου. Ένα επαναλαμβανόμενο μοτίβο αλλαγών ανάμεσα στις τιμές των παρατηρήσεων υπάρχει, προ δίδοντας την διατήρηση της συνιστώσας της περιοδικότητας επί της χρονοσειράς. Από τα διαγράμματα αυτοσυσχέτισης και για τους δύο μετασχηματισμούς απουσιάζει πλέον η αργή μείωση της αυτοσυσχέτισης μεταξύ χρονικών υστερήσεων μετά την πρώτη σημαντική υστέρηση ($lag = 0$). Πλέον στα διαγράμματα αυτοσυσχέτισης δικαιολογείται μόνο η ύπαρξη περιοδικότητας ανά 12 lags, ήτοι μια περίοδος 12 μηνών.

Στη συνέχεια πραγματοποιείται ο στατιστικός έλεγχος ADF και υποδεικνύει ότι οι χρονοσειρές που εξετάστηκαν είναι στάσιμες. Ένα αποτέλεσμα το οποίο αναμενόταν, καθώς η



Σχήμα 7.17: Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα αντίστροφης τετραγωνικής ρίζας αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.

μέση τιμή και η διασπορά αυτών έχει σταθεροποιηθεί σημαντικά. Τέλος, για τα δεδομένα που είναι στη λογαριθμική κλίμακα παράγεται ξανά το ιστόγραμμα των υπολοίπων, όπως φαίνεται στο Σχήμα 7.18, αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών. Φαίνεται να έχουν μέση τιμή πολύ κοντά στο μηδέν, ενώ επίσης η διασπορά τους υπερβαίνει οριακά το 1. Είναι ενδείξεις ότι η κατανομή πυκνότητας πιθανότητας που προσεγγίζουν τα υπολείμματά R_t , είναι η κανονική. Αποτελεί μια θετική εξέλιξη στη προσπάθεια που έγινε πριν τη παραγωγή πρόβλεψης. Το γεγονός αυτό δεν καθιστά τη χρονοσειρά μας στάσιμη αναγκαστικά, αλλά υποδεικνύει ότι η εφαρμογή γραμμικών μοντέλων (AR, MA, ARMA, ARIMA, SARIMA) για τη παραγωγή των επιθυμητών προβλέψεων πρόκειται να αποφέρουν καλά και έμπιστα αποτελέσματα εάν χρησιμοποιηθούν τα δεδομένα στη παρούσα μορφή αφού τα υπόλοιπα της χρονοσειράς προσεγγίζουν πλέον σε ικανοποιητικό βαθμό την κανονική κατανομή.



Σχήμα 7.18: Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στη λογαριθμική κλίμακα αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.

7.2.4 Παραγωγή Προβλέψεων με Γραμμικά Μοντέλα

Για την παραγωγή των προβλέψεων επί της συγκεκριμένης χρονοσειράς χρησιμοποιήθηκαν μόνο γραμμικά μοντέλα AR, MA, ARMA, ARIMA και SARIMA και μελετήθηκε η απόδοση τους. Η επιλογή τους έγινε με βάση το κριτήριο AIC και η αξιολόγηση τους με βάση το κριτήριο της ρίζας του μέσου τετραγωνικού σφάλματος RMSE καθώς και άλλων στατιστικών μέτρων όπως του μέσου σφάλματος ME ή του μέσου απόλυτου σφάλματος MAE. Τέλος, ιδιαίτερη βάση δόθηκε στην μελέτη του ιστογράμματος των υπολοίπων που αντιστοιχεί σε κάθε μοντέλο. Με τη μελέτη αυτού εξήχθησαν απαραίτητες πληροφορίες, οι οποίες χρησιμοποιήθηκαν επίσης στην αξιολόγηση της επάρκειας μοντέλου που εκπαιδεύτηκε και αξιοποιήθηκε για να παράξει τελικώς μια πρόβλεψη επί της χρονοσειράς που μελετάται.

Πριν παραχθεί η πρόβλεψη, πρέπει, να αποφασηριστεί η χρησιμότητα και ο λόγος πραγματοποίησης της. Στη περίπτωση των δεδομένων παραγωγής ηλεκτρικής ενέργειας που ανακτήθηκαν από τη διαδικτυακή πλατφόρμα επιστήμης δεδομένων Kaggle, εξετάζονται μηνιαία δεδομένα συνολικής παραγωγής τα οποία μπορούν να χρησιμοποιηθούν προκειμένου να παραχθούν προβλέψεις επί της παραγωγής για τους επόμενους μήνες. Επιλέγεται να γίνει πρόβλεψη της παραγωγής για κάθε έναν από τους 12 επόμενους μήνες από το τελευταίο εξεταζόμενο δείγμα. Με αυτόν τον τρόπο μπορεί η εταιρία που παράγει ορισμένες ποσότητες ενέργειας να πληροφορηθεί για τις επικείμενες ποσότητες που πρόκειται να έχει διαθέσιμες προς πώληση στο άμεσο μέλλον. Συνεπώς, με βάση αυτές μπορεί να γνωρίζει τι μέρος της ζήτησης μπορεί να καλύψει στην αγορά, να μεγιστοποιήσει το κέρδος της, να προγραμματίσει εργασίες συντήρησης ή αναβάθμισης, να διαμορφώσει την πολιτική της και να λάβει τις κατάλληλες οικονομοτεχνικές αποφάσεις που θα διασφαλίσουν την βιωσιμότητά της.

Όλα τα γραμμικά μοντέλα, που θα εφαρμοστούν, θα παράγουν μια πρόβλεψη για την παραγωγή, που πρόκειται να υπάρχει για κάθε έναν από τους επόμενους 12 μήνες. Οι αλγόριθμοι οι οποίοι υλοποιούσαν τους υπολογισμούς που ορίζει το κάθε μοντέλο, παρέχονταν από τη βιβλιοθήκη `Statsmodel` της python. Σε αυτή μέσα υπάρχουν οι κλάσεις ARIMA και SARIMA οι οποίες παρέχουν κατάλληλες συναρτήσεις για την εκπαίδευση των μοντέλων που θα παράγουν τις προβλέψεις μας.

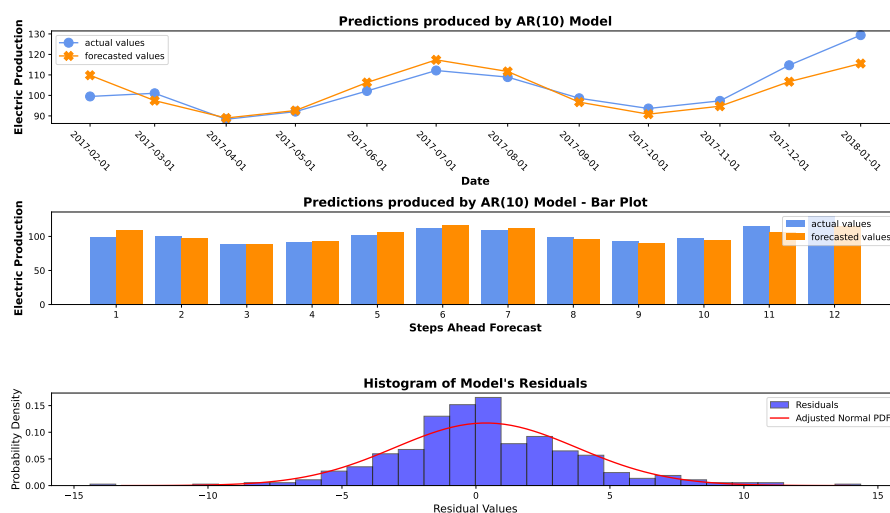
Τα δεδομένα παραγωγής χωρίστηκαν σε δεδομένα εκπαίδευσης (train dataset) και δεδομένα ελέγχου (test dataset). Πιο συγκεκριμένα, οι πρώτες 385 παρατηρήσεις της χρονοσειράς χρησιμοποιούνται, για να εκπαιδευτεί ένα πιθανό γραμμικό μοντέλο πρόβλεψης. Αποτελούν τα δεδομένα εκπαίδευσης (train dataset) με βάση τα οποία θα εκτιμηθεί το μοντέλο που προσαρμόζεται καλύτερα σε αυτά και θα χρησιμοποιηθεί στη συνέχεια για να πραγματοποιηθεί η πρόβλεψη εκτός δείγματος. Σκοπός δεν είναι άλλος παρά η αναζήτηση και εύρεση του μοντέλου που θα αντιλαμβάνεται και θα προσομοιώνει με καλύτερο τρόπο τις σχέσεις και τα μοτίβα που υπάρχουν μεταξύ των παρατηρήσεων. Η εκτίμηση του ολοκληρώνεται αφού υπολογισθούν με βάση τα ιστορικά αυτά δεδομένα οι τάξεις αυτοπαλινδρόμησης p , κινούμενου μέσου q και διαφορών d .

Από το σύνολο των δεδομένων που αποτελείτο από 397 παρατηρήσεις οι εναπομείναντες 12 αποτελούν τα δεδομένα ελέγχου (test dataset). Θα χρησιμοποιηθούν, ώστε να υπολογισθεί

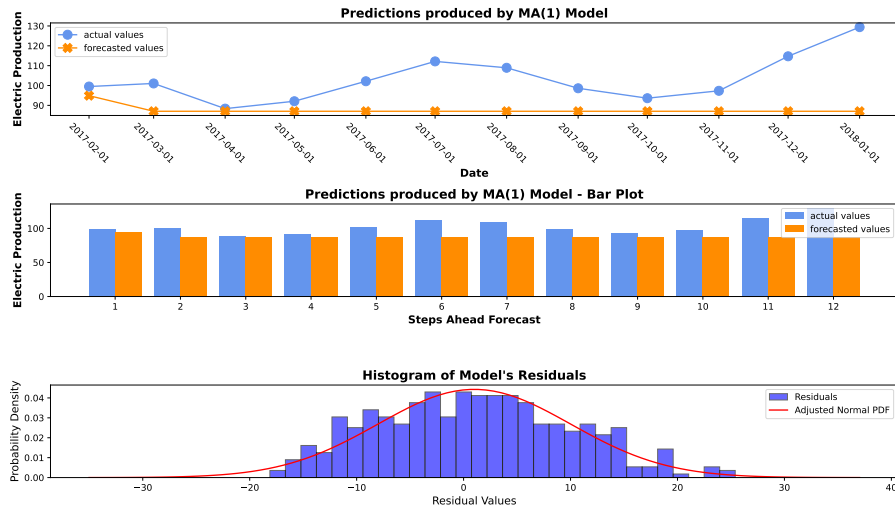
η απόκλιση των τιμών που προβλέφθηκαν από των πραγματικών. Για τη άμεση σύγκριση και την αξιολόγηση των μοντέλων χρησιμοποιείται το στατιστικό μέτρο της ρίζας του μέσου τετραγωνικού σφάλματος RMSE. Για το βέλτιστο μοντέλο υπολογίζονται επιπλέον στατιστικά μέτρα τα οποία είναι ο συντελεστής προσδιορισμού (R^2), ο συντελεστής συσχέτισης ($corr$), το μέσο απόλυτο σφάλμα (MAE), το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE), το μέσο σφάλμα (ME) και το μέσο ποσοστιαίο σφάλμα (MPE).

Για το κάθε μοντέλο γίνεται μια οπτική σύγκριση ανάμεσα στις πραγματικές τιμές και σε αυτές που προβλέφθηκαν από τα γραμμικά μοντέλα. Σε κάθε σχήμα παρέχεται ένα διάγραμμα παράλληλης απεικόνισης των προβλέψεων και της πραγματικής χρονοσειράς, ένα ραβδογράφημα αυτών και ένα ιστόγραμμα υπολοίπων του μοντέλου που εκπαιδεύτηκε και αξιοποιήθηκε για την πρόβλεψη. Έτσι μπορεί να παρατηρηθεί το πόσο καλά ανταπεξέρχεται το μοντέλο μας από πρόβλεψη σε πρόβλεψη και να γίνει μια πιο λεπτομερής σύγκριση ανάμεσα στις μηνιαίες πραγματικές τιμές ποσοτήτων παραγόμενης ενέργειας και αυτές που προβλέφθηκαν τελικά.

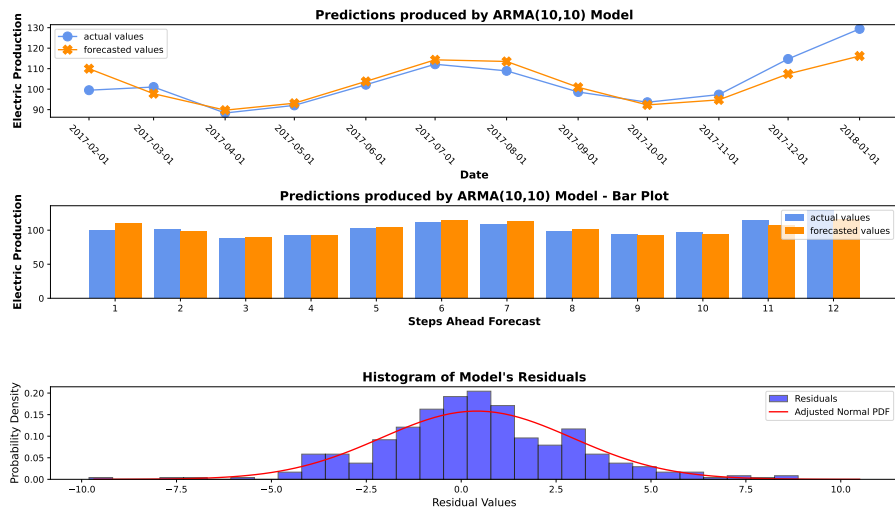
Το ιστόγραμμα των υπολοίπων εξυπηρετεί πολλούς σκοπούς στην ανάλυση χρονοσειρών με γραμμικά μοντέλα πρόβλεψης. Προσφέρει μια αξιολόγηση επί της προσαρμογής του μοντέλου ή αλλιώς της ποιότητας πρόβλεψης του goodness of fit. Ακόμη είναι δυνατόν να γίνει έλεγχος για την ύπαρξη τυχόν ακραίων τιμών ή κανονικότητας. Με βάση αυτό μπορεί να δικαιολογηθεί η επάρκεια ή ανεπάρκεια του μοντέλου. Η σωστή ανάγνωση του είναι δυνατό να δικαιολογήσει την ανάγκη βελτίωσης του μοντέλου που έχει εκπαιδευτεί είτε επειδή εμφανίζονται σημάδια αποσταθεροποίησης της διακύμανσης ή μέσης τιμής είτε επειδή το ιστόγραμμα δε φαίνεται να προσομοιώνει αυτό μιας χρονοσειράς λευκού θορύβου. Τέλος, μπορεί να χρησιμοποιηθεί ως μέτρο σύγκρισης με άλλα γραμμικά μοντέλα καθώς αυτό που φέρεται να βρίσκεται πιο κοντά σε ένα ιστόγραμμα λευκού θορύβου αποτελεί τη καλύτερη επιλογή για τη παραγωγή της πρόβλεψης μας. Θα είναι ουσιαστικά το μοντέλο το οποίο εξηγεί όσο καλύτερα μπορεί τα μοτίβα και τις σχέσεις ανάμεσα στα δεδομένα με αποτέλεσμα το κάθε υπόλειμμα να είναι απλά τυχαίος θόρυβος.



Σχήμα 7.19: Πρόβλεψη με χρήση του γραμμικού μοντέλου AR.



Σχήμα 7.20: Πρόβλεψη με χρήση του γραμμικού μοντέλου MA.



Σχήμα 7.21: Πρόβλεψη με χρήση του γραμμικού μοντέλου ARMA.

Για την παραγωγή της πρόβλεψης, επιλέχθηκε να δοκιμαστούν διάφορα γραμμικά μοντέλα. Πιο συγκεκριμένα, έγινε δοκιμή με μοντέλα $AR(p)$, $MA(q)$, $ARMA(p,q)$, $ARIMA(p,d,q)$ και $SARIMA(p,d,q),(P,D,Q)_{12}$. Στα Σχήματα 7.19, 7.20, 7.21, 7.22, 7.23 παρουσιάζονται οι προβλέψεις που απέδωσε το καλύτερο μοντέλο από το κάθε είδος. Για να βρεθεί η βέλτιστη δυνατή τάξη των παραγόντων (p,q) του κάθε μοντέλου, εφαρμόστηκε ένας αλγόριθμος εξαντλητικής αναζήτησης στο σύνολο τιμών $[0,10]$ τόσο για τον παράγοντα p όσο και για τον παράγοντα q . Για τους εποχικούς συντελεστές (P,Q) ορίστηκε το διάστημα αναζήτησης $[0,2]$, ενώ οι παράγοντες διαφορίσης d,D τέθηκαν ίσοι με το 1 στις περιπτώσεις των $ARIMA$ και $SARIMA$ μοντέλων προκειμένου να αντιμετωπιστεί η τάση ($d = 1$) και η εποχικότητα ($D = 1$) ώστε να απαλειφθεί η τάση και η συνιστώσα της εποχικότητας από την χρονοσειρά. Την σταθεροποίηση της διασποράς είχε αναλάβει ο λογαριθμικός μετασχηματισμός των τιμών των παρατηρήσεων της χρονοσειράς.

Με την ολοκλήρωση της διαδικασίας αναζήτησης είχαν εκπαιδευτεί όλα τα δυνατά μοντέλα με βάση τα δεδομένα εκπαίδευσης. Το μοντέλο που όμως επιλεγόταν τελικά να χρησιμοποιηθεί, ώστε να γίνει η πρόβλεψη, ήταν αυτό που απέδιδε τη μικρότερη τιμή στο κριτήριο AIC κατά την εκπαίδευση του. Το μοντέλο αυτό στη συνέχεια χρησιμοποιούνταν για την επιθυμητή πρόβλεψη εκτός δείγματος (out-of-sample prediction). Τέλος, οι προβλέψεις που μελετώνται και συγκρίνονται παρακάτω είναι βασισμένες στα λογαριθμικά δεδομένα και όχι στα δεδομένα αντίστροφης τετραγωνικής ρίζας. Τα μοντέλα που εκπαιδεύτηκαν με βάση τα λογαριθμημένα δεδομένα είχαν τις μικρότερες τιμές στο κριτήριο AIC οπότε και εικάσθηκε ότι θα επέφεραν πιο ακριβείς προβλέψεις έναντι αυτών που είχαν μετασχηματιστεί σε κλίμακα αντίστροφης τετραγωνικής ρίζας. Έτσι επιλέχθηκαν να παρουσιαστούν μόνο αυτά.

Οι προβλέψεις που αποτυπώνονται στα Σχήματα 7.19, 7.20, 7.21, οφείλονται στην αξιοποίηση AR, MA και ARMA μοντέλων αντίστοιχα. Στην ίδια κλίμακα με αυτήν που βρίσκονταν τα δεδομένα εκπαίδευσης πραγματοποιήθηκαν οι προβλέψεις. Ωστόσο, πριν παρουσιαστούν και βγουν συμπεράσματα για την απόδοση του μοντέλου και την ποιότητα της πρόβλεψης, επαναφέρθηκαν στην αρχική κλίμακα οι τιμές τους. Τα τρία αυτά γραμμικά μοντέλα δεν χρησιμοποιούν τη μέθοδο των διαφορών εν αντιθέσει με τα μοντέλα ARIMA και SARIMA που παρουσιάζονται στα Σχήματα 7.22 και 7.23. Μπορεί η μεταφορά των τιμών στη λογαριθμική κλίμακα να μειώνει τις απότομες διακυμάνσεις και άρα να σταθεροποιεί εν μέρη τη διακύμανση αλλά δεν σταθεροποιείται απόλυτα η μέση τιμή της χρονοσειράς. Παρόλα αυτά επιλέγεται να εφαρμοστούν κανονικά τα συγκεκριμένα μοντέλα καθώς πλέον οι μεταβολές που υπάρχει στη τιμή της μέσης τιμής και της διασποράς στη πάροδο του χρόνου είναι μικρές μετά την εφαρμογή του λογαριθμικού μετασχηματισμού.

Το μοντέλο κινούμενου μέσου MA δίνει την χειρότερη πρόβλεψη. Είναι ένα απλό μοντέλο το οποίο δεν συλλαμβάνει τις ιδιαιτερότητες της συγκεκριμένης χρονοσειράς όπως τη τάση και τη περιοδικότητα. Η ανάλυση που έγινε επί του ιστογράμματος των υπολοίπων που προέκυψε μετά την εκπαίδευση του μοντέλου προσφέρει πολύτιμες πληροφορίες για την απόδοση του μοντέλου αλλά και για την φύση της χρονοσειράς. Όπως φαίνεται από το Σχήμα 7.20, το ιστόγραμμα δεν προσομοιώνει τον λευκό θόρυβο που θα ήταν το ιδανικό. Δεν παρουσιάζει σταθερή διακύμανση τιμών εκατέρωθεν της μέσης τιμής αφού δεν διακυμαίνονται συμμετρικά οι τιμές. Υφίσταται μια μικρή σε βαθμό δεξιά λοξότητα right skewness, η οποία υποδηλώνει ότι πιθανότατα το μοντέλο που αξιολογείται (MA(1)), θα έχει την τάση να υποτιμά κατά την διαδικασία της πρόβλεψης όπως και γίνεται τελικά στη πλειοψηφία των περιπτώσεων που κλήθηκε να κάνει μια πρόβλεψη.

Από την άλλη μεριά, το μοντέλο αυτοπαλινδρόμησης AR έδωσε μια καλή πρόβλεψη. Είναι καλύτερη από αυτήν του κινούμενου μέσου αλλά χειρότερη από αυτή του μοντέλου ARMA, όπου συνδυάζονται και εφαρμόζονται, οι διαδικασίες κινούμενου μέσου και αυτοπαλινδρόμησης. Το μοντέλο αυτό άλλοτε φαίνεται να υπερεκτιμά και άλλοτε να υποτιμάει. Σε γενικές γραμμές πάντως φαίνεται να μπορεί με ακρίβεια να εκτιμήσει τις μελλοντικές ποσότητες ενέργειας που θα παραχθούν για το συγκεκριμένο διάστημα όπου του ανατέθηκε να προβλέψει.

Οι ιδιότητες αυτές δικαιολογούνται από το ιστόγραμμα υπολοίπων του. Παρουσιάζει μια υψηλή κορυφή κοντά στο μηδέν, η οποία σαν ένδειξη έχει θετικό πρόσημο καθώς φανερώνει

την ικανότητα του να προσφέρει ακριβείς προβλέψεις σε πολλές περιπτώσεις. Ωστόσο, σημασία πρέπει να δοθεί επίσης στον τρόπο που εξαπλώνονται οι τιμές γύρω από αυτήν. Φαίνεται, να διακυμαίνονται με διαφορετικό τρόπο και σε διαφορετικό βαθμό εκατέρωθεν αυτής. Δεν υπάρχει συμμετρία αρχικά και παρουσιάζει δυο κορυφές στο αριστερό μέρος του, ένδειξη του ότι υπάρχουν πολλές περιπτώσεις που υπερτιμά με μικρή απόκλιση τις πραγματικές τιμές. Φαίνεται να υπάρχουν περιπτώσεις που υποεκτίμησε ή υπερεκτίμησε σε μεγάλο βαθμό καθώς υπάρχουν ακραίες τιμές εκτός του κυρίου σώματος του ιστογράμματος. Αυτές ίσως είναι οι περιπτώσεις που επειδή δεν έχει εξαλειφθεί περιοδικότητα ή η τάση δεν μπόρεσε να εκτιμήσει σωστά την μεταβολή των τιμών. Πράγματι, στις προβλέψεις το πιο μεγάλο σφάλμα το παρουσιάζει κατά την 11η και 12η πρόβλεψη όπου η περιοδικότητα έχει ισχυρή επίδραση στη μεταβολή των τιμών της χρονοσειράς. Με αλλά λόγια, είναι πιθανό ότι σε κάποιο άλλο διάστημα οι προβλέψεις που θα προσέφερε η αξιοποίηση του μοντέλου AR(10) να μην ήταν αντίστοιχης ποιότητας.

Το μοντέλο αυτοπαλίνδρομου κινούμενου μέσου ARMA που αξιοποιήθηκε έφερε το καλύτερο αποτέλεσμα ανάμεσα σε αυτά που χρησιμοποιήθηκαν χωρίς να σταθεροποιείται η μέση τιμή τους μέσω της εφαρμογής πρώτων διαφορών. Το μοντέλο αυτό σε απόδοση ξεπερνάει κατά πολύ το μοντέλο κινούμενου μέσου και λίγο αυτό της αυτοπαλινδρόμησης. Βέβαια, κατά πάσα περίπτωση προσφέρει πιο εύστοχες προβλέψεις από το μοντέλο αυτοπαλινδρόμησης AR(10). Το συμπέρασμα αυτό προκύπτει τόσο από την παρατήρηση των προβλεπόμενων τιμών όσο και από την παρατήρηση του ιστογράμματος των υπολοίπων του υπό μελέτη μοντέλου.

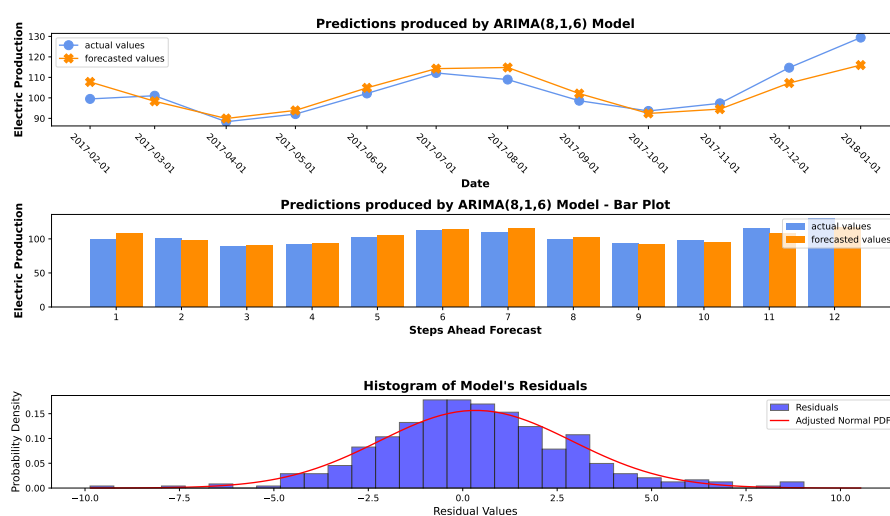
Στο ιστόγραμμα υπολοίπων παρατηρείται υψηλή συγκέντρωση τιμών κοντά στο μηδέν και μια πιο συμμετρική κατανομή των τιμών γύρω από αυτό. Η συγκέντρωση αποτελείται από τρεις ισούψείς κορυφές οι οποίες ξεπερνούν ελάχιστα την κορυφή που θα είχε το ιδανικό ιστόγραμμα λευκού θορύβου. Παρόλα αυτά υπάρχουν ορισμένα αρνητικά που αξίζουν να αναφερθούν αφού δεν προσομοιώνει τον λευκό θόρυβο απόλυτα. Αρχικά υπάρχουν ορισμένες ιδιόμορφες τιμές outliers. Αυτές βρίσκονται εκατέρωθεν του κύριου σώματος του ιστογράμματος. Με την παρατήρηση αυτή, γίνεται αντιληπτό ότι υπάρχουν κάποιες ελάχιστες περιπτώσεις όπου το μοντέλο δεν δύναται να προσομοιώσει με επιτυχία. Επίσης, παρουσιάζει μια μικρού βαθμού δεξιά σκίαση που αποτελεί ένδειξη ότι οι περιπτώσεις υποεκτίμησης είναι περισσότερες από αυτές της υπερεκτίμησης. Συνεπώς, είναι ένα μοντέλο το οποίο αναμένεται να έχει καλή απόδοση παρά τις ελάχιστες αστοχίες που μπορεί να φέρει.

Στην ανάλυση των προβλέψεων που έφεραν τα τρία μοντέλα πέρα από το αποτέλεσμα, άξιος σχολιασμού, είναι επίσης ο χρόνος που δαπανήθηκε κατά την αναζήτηση και εύρεση του κάθε ενός. Πρέπει, δηλαδή, να συνυπολογιστεί η χρονική πολυπλοκότητα των αλγορίθμων αναζήτησης της καλύτερης τάξης των όρων p, q και εκπαίδευσης των μοντέλων επί του πλήθους των δεδομένων. Υπό αυτή την συνθήκη και εκ των υστέρων βλέποντας τα αποτελέσματα που προσέφεραν σαν μοντέλα πρόβλεψης η επιλογή του αυτοπαλίνδρομου μοντέλου AR είναι πιο αποδοτική. Με τη χρήση ενός συμβατικού υπολογιστή, η εκτέλεση και ολοκλήρωση του αλγόριθμου που φέρει στο προσκήνιο το καλύτερο μοντέλο αυτοπαλίνδρομου κινούμενου μέσου ARMA απαιτεί περίπου δέκα φορές παραπάνω χρόνο σε σχέση με το μοντέλο AR (12 λεπτά έναντι 1.5 λεπτών).

Σημαντικό να αναφερθεί ότι ο χρόνος αυτός αυξάνεται σημαντικά όταν εξετάζονται και

εκπαιδεύονται μοντέλα μεγαλύτερης τάξης p και πιο ειδικά τάξης μεγαλύτερη του 6. Υπενθυμίζεται ότι σε όλες τις αναζητήσεις το εύρος αναζήτησης των όρων p, q είναι από 0 έως 10. Συνεπώς, το ARMA μοντέλο να μεν αποδίδει καλύτερη πρόβλεψη με RMSE ίσο με 5.93 μονάδες ποσότητας ηλεκτρικής ενέργειας έναντι 6.09 του AR μοντέλου, αλλά απαιτεί πολύ παραπάνω χρόνο η εύρεση και εκπαίδευση του. Στη συγκεκριμένη περίπτωση ο χρόνος αυτός δεν είναι απαγορευτικός ως προς την αξιοποίηση του γραμμικού μοντέλου ARMA καθώς γίνεται πρόβλεψη με ορίζοντα 12 μηνών μπροστά.

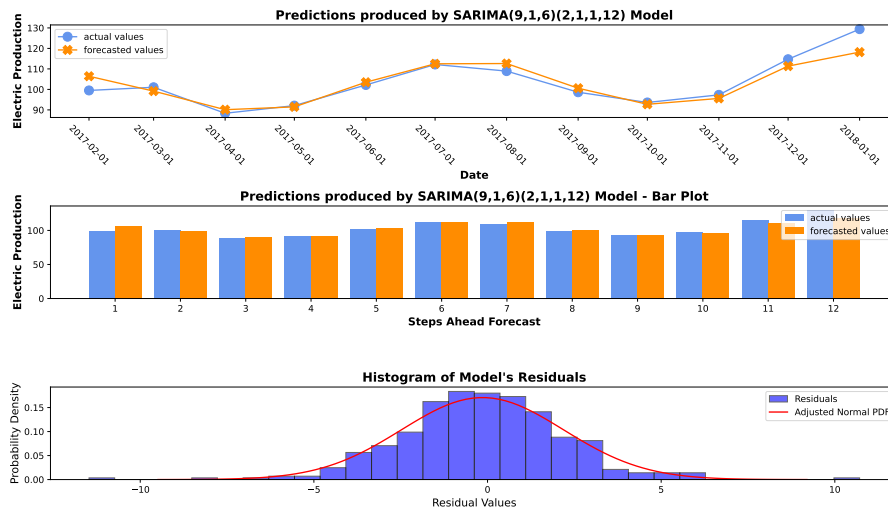
Στη περίπτωση των προβλέψεων ποσοτήτων παραγόμενης ηλεκτρικής ενέργειας ο χρόνος είναι πολύτιμος καθώς υπάρχουν πάντα συγκεκριμένα χρονικά περιθώρια μέσα στα οποία πρέπει να ολοκληρώνεται και να υποβάλλονται οι προτάσεις στο Χρηματιστήριο Ενέργειας. Για να γίνει αυτό, πρώτο και σημαντικό βήμα αποτελεί η παραγωγή και αξιολόγηση μιας πρόβλεψης εντός λογικών χρονικών πλαισίων.



Σχήμα 7.22: Πρόβλεψη με χρήση του γραμμικού μοντέλου ARIMA.

Για τις προβλέψεις, που αποτυπώνονται στα Σχήματα 7.22 και 7.23, χρησιμοποιήθηκαν επίσης τα λογαριθμικά μετασχηματισμένα δεδομένα. Στο πρώτο προαναφερόμενο σχήμα απεικονίζεται η πρόβλεψη που παρήγαγε το βέλτιστο αυτοπαλίνδρομο ολοκληρωμένο μοντέλο κινούμενου μέσου ARIMA, ενώ στο δεύτερο απεικονίζεται η πρόβλεψη που παρήγαγε το βέλτιστο εποχικό αυτοπαλίνδρομο ολοκληρωμένο μοντέλο κινούμενου μέσου SARIMA. Ακολούθησε η ίδια διαδικασία όπως με τα μοντέλα AR, MA, ARMA όσον αφορά τον τρόπο εύρεσης της βέλτιστης τάξης των όρων p, q για το μη εποχικό και $(p, q), (P, Q)$ για το εποχικό. Σε αυτά τα μοντέλα υπήρχε επίσης η δυνατότητα διαφορίσης των δεδομένων που σταθεροποιεί τη μέση τιμή της χρονοσειράς αφού απαλείφει τη τάση και μετατρέπει σε στάσιμη τη χρονοσειρά. Το μοντέλα που έφεραν την μικρότερη τιμή στο κριτήριο επιλογής AIC, δηλαδή τα ARIMA(8, 1, 6) και SARIMA(9, 1, 6)(2, 1, 1)₁₂ επιλέχθηκαν ως τα καλύτερα όσον αφορά την προσαρμογή τους στα δεδομένα εκπαίδευσης.

Πρώτα παρατηρείται η πρόβλεψη που προήλθε από την αξιοποίηση του μοντέλου ARIMA (Σχήμα 7.22). Το μοντέλο αυτό φαίνεται να παρήγαγε ακόμα πιο εύστοχες προβλέψεις. Οι



Σχήμα 7.23: Πρόβλεψη με χρήση του γραμμικού μοντέλου SARIMA.

τιμές που προβλέπει φέρουν μικρές αποκλίσεις από τις πραγματικές. Οι περιπτώσεις που παρουσιάζει υπερεκτίμηση είναι περισσότερες από τις περιπτώσεις που υποτιμά. Κατά πάσα πιθανότητα, το γεγονός αυτό οφείλεται στην πολυπλοκότητα του. Ίσως η επιλογή χαμηλότερων τάξεων αυτοπαλινδρόμησης και κινούμενου μέσου να μετρίαζε την κατάσταση αυτή. Παρόλα αυτά, επειδή, είναι μικρή σε μέγεθος δεν οδηγεί σε άσχημη πρόβλεψη. Ενώς ακόμη παράγοντας που θεωρητικά οδηγεί στην εμφάνιση της υπερεκτίμησης (overestimation) είναι η έλλειψη εποχικού μέρους.

Το ιστόγραμμα των υπολοίπων του συγκεκριμένου μοντέλου μελετάται στη συνέχεια. Το συγκεκριμένο μοντέλο έχει ισχυρή συγκέντρωση τιμών κοντά στο μηδέν με τις περισσότερες αυτών να βρίσκονται στο κομμάτι των θετικών τιμών. Αν και σε αυτό βλέπουμε καλύτερη προσομοίωση του λευκού θορύβου από ότι προηγουμένως δεν έχει εξαλειφθεί η ύπαρξη ελάχιστων ακραίων τιμών, δείγμα του ότι θα υπάρξουν περιπτώσεις που το μοντέλο δεν θα μπορεί να επιδείξει άριστη προβλεπτική ικανότητα. Παρόλα αυτά η ισχυρή συγκέντρωση τιμών και η ικανοποιητικά ομοιόμορφη κατανομή τιμών γύρω από τη μέση τιμή αυτού, που προσεγγίζει το μηδέν, δίνει μια καλή προσομοίωση του λευκού θορύβου και την απόδειξη της καλής απόδοσης του μοντέλου επί προβλέψεων εκτός δείγματος. Το ARIMA μοντέλο επιφέρει καλύτερα αποτελέσματα από τα MA, AR και ARMA αφού σε αντίθεση με αυτά, στο συγκεκριμένο αντιμετωπίζεται η τάση αξιοποιώντας έτσι στην ουσία δεδομένα από μια στάσιμη χρονοσειρά για την εκπαίδευση του πριν την παραγωγή προβλέψεων.

Με την παρατήρηση του Σχήματος 7.23 που παρουσιάζει τις προβλέψεις που παρήγαγε το μοντέλο SARIMA γίνεται εύκολα αντιληπτό ότι το συγκεκριμένο μοντέλο δίδει τις καλύτερες προβλέψεις. Το συγκεκριμένο μοντέλο λαμβάνει επιπλέον υπόψιν του την χρονική περιοδικότητα που υπάρχει ή ετήσια εποχικότητα και την απαλείφει. Διαφορίζει ουσιαστικά ανά 12 υστερήσεις ήτοι μια εποχική περίοδο (ετήσια) τις παρατηρήσεις της χρονοσειράς για να εξαλείψει την εποχική συνιστώσα της χρονοσειράς με το να θέτει ίσο με την τιμή 1 τον εποχικό παράγοντα διαφόρισης D ($D = 1$). Η ποιότητα των προβλέψεων είναι υψηλή. Οι περιπτώσεις

που υπερεκτιμά είναι ανάλογες με τις περιπτώσεις που υποτιμά. Υπάρχει μια ισορροπία, δείγμα της εκπαίδευσης του με τη πλέον στάσιμη χρονοσειρά.

Η εξαντλητική αναζήτηση έφερε ένα μοντέλο το οποίο έχει όσες τάξεις πραγματικά χρειάζονται για να περιγραφεί η χρονοσειρά έτσι ώστε η πολυπλοκότητα του να μην οδηγήσει σε υπερπροσαρμογή των δεδομένων. Το ιστόγραμμα υπολοίπων της χρονοσειράς φανερώνει την καλύτερη προσομοίωση σε λευκό θόρυβο με σχεδόν υποδειγματική κατανομή των τιμών γύρω από μια μέση τιμή κοντά στο μηδέν. Ακόμη παρατηρείται να υπάρχει ελαφρώς αριστερή σκίαση και υπάρχουν ορισμένες τιμές οι οποίες συγκεντρώνονται σε διαστήματα μακριά από το κύριο σώμα του ιστογράμματος και αποτελούν τις περιπτώσεις που το μοντέλο δεν μπόρεσε να προσομοιώσει με το βέλτιστο δυνατό τρόπο κάποιες παρατηρήσεις της χρονοσειράς. Εάν δεν υπήρχαν αυτές ίσως τότε να ευσταθούσε ο ισχυρισμός ότι το ιστόγραμμα προσομοιώνει ιδανικά τον λευκό θόρυβο. Η εικόνα αυτή οφείλεται στο γεγονός ότι το μοντέλο αυτό σταθεροποιώντας την τάση και εξαλείφοντας την εποχική περιοδικότητα μπορεί να προσφέρει προς αξιοποίηση κατά την εκπαίδευση δεδομένα που συνθέτουν μια χρονοσειρά στάσιμη. Το μοντέλο αυτό μπορεί να περιγράψει επαρκώς τη χρονοσειρά και δεν εισάγει μεροληψία κατά την πρόβλεψη τιμών εκτός δείγματος στη πλειοψηφία των περιπτώσεων.

Τα δύο βέλτιστα μοντέλα (ARIMA, SARIMA) φέρουν μεγάλη διαφορά στον χρόνο εκτέλεσης και εκπαίδευσης τους. Η εύρεση και εκπαίδευση του καλύτερου μοντέλου ARIMA χρειάζεται περίπου τον μισό χρόνο σε σχέση με το SARIMA (30 λεπτά έναντι σχεδόν 60 λεπτών). Προκειμένου να κερδηθεί χρόνος στη περίπτωση του SARIMA η αναζήτηση των τιμών της τάξεως των παραγόντων p, q έγινε στο διάστημα $[5, 10]$ καθώς αναμενόταν ότι για μικρότερους όρους δεν θα δινόταν τόσο καλή πρόβλεψη. Οι εποχικοί όροι P, Q είχαν εύρος αναζήτησης $[0, 2]$. Παρατηρήθηκε ξανά ότι ο χρόνος αυξάνεται σημαντικά όταν εξετάζονται και εκπαιδεύονται μοντέλα μεγαλύτερης τάξης p, q και P, Q .

Είναι λογικό να εγείρονται ξανά προβληματισμοί ως προς τον χρόνο απόκρισης των αλγορίθμων σε σχέση με την ποιότητα του αποτελέσματος. Στη προκειμένη περίπτωση, θεωρώ ότι η επιλογή του πιο ποιοτικού μοντέλου έναντι του λιγότερου ποιοτικού αλλά επίσης λιγότερο χρονοβόρου μοντέλου είναι προτιμητέα. Πρόκειται για μια πρόβλεψη η οποία αναφέρεται στη μηνιαία παραγωγή του επόμενου χρόνου. Είναι χρονικά πολύ μικρό το αντίτιμο σε σχέση με τον χρονικό ορίζοντα της πρόβλεψης. Επίσης, αν η μονάδα μέτρησης των ποσοτήτων ενέργειας που παράχθηκαν ήταν γνωστή, η άποψη αυτή θα μπορούσε να τεκμηριωθεί και ποιοτικά. Εάν το βέλτιστο μοντέλο SARIMA αποκλείει κατά 4.28 μονάδες μέτρησης και το βέλτιστο ARIMA κατά 5.48, τότε η μεταξύ τους διαφορά έγκειται στις 1.20 μονάδες μέτρησης της ποσότητας της ενέργειας που παράχθηκε. Αυτή η ποσότητα είναι αμελητέα εάν συζητάμε περί kWh και αξιοσημείωτη εάν μιλάμε για GWh.

Πολύ σημαντικές παρατηρήσεις μπορούν να εξαχθούν με την παρατήρηση των σχημάτων που παρουσιάζουν τις προβλέψεις του εκάστοτε μοντέλου. Παρατηρείται στο Σχήμα 7.23 ότι το μικρότερο σφάλμα μέσης τετραγωνικής ρίζας παρουσιάζει η πρόβλεψη η οποία έγινε 9 χρονικά βήματα μετά το τελευταίο δείγμα που λήφθηκε υπόψιν στην εκπαίδευση του μοντέλου. Μετά το 9ο βήμα αρχίζει να αυξάνει το σφάλμα του μοντέλου. Από το σημείο εκείνο και έπειτα γίνονται πρόβλεψη σε χρονικές στιγμές οι οποίες βρίσκονται πέραν του ορίζοντα εξέτασης της

AR διαδικασίας ($p=9$). Σε μια τέτοια περίπτωση τόσο ένα απλό μοντέλο ARIMA όσο και ένα εποχιακό έχουν την τάση να μετατρέπονται σε AR μοντέλα των οποίων οι οι προβλέψεις που παράγονται εξαρτώνται πλέον από τις τιμές που έχουν ήδη προβλεφθεί και όχι από τις παρελθοντικές.

Είναι προτιμητέο να μην γίνονται προβλέψεις με μοντέλα ARIMA σε βάθος χρόνου παρά σε πιο μικρούς χρονικούς ορίζοντες. Ουσιαστικά, το μοντέλο πέραν του ορίζοντα αυτού αποδίδει κάτω των δυνατοτήτων του (underachieving). Σε αυτήν την περίπτωση είτε μειώνουμε τον ορίζοντα πρόβλεψης και τότε ανατροφοδοτούμε το μοντέλο με τις τιμές που παρατηρήθηκαν στη πραγματικότητα είτε αυξάνουμε την τάξη των συνιστωσών AR, MA και πληρώνουμε σε χρόνο δίχως να υπάρχει εγγύηση ότι τα αποτελέσματα μας θα έχουν ακρίβεια.

7.2.5 Αξιολόγηση Προβλέψεων

Μια ανάγνωση των γραφημάτων που προαναφέρθηκαν μπορεί να προσφέρει μια εμπειρική εκτίμηση της ποιότητας των προβλέψεων που τελικά παράχθηκαν. Ωστόσο, στη περίπτωση που θέλουμε να μελετήσουμε με μεγαλύτερη προσοχή και ασφάλεια τα δεδομένα μας και τα αποτελέσματα μας, πρέπει να υπολογίσουμε και να εξετάσουμε ορισμένα κριτήρια με βάση τα οποία θα αξιολογηθούν οι προβλέψεις.

Οι τιμές της ρίζας μέσου τετραγωνικού σφάλματος (RMSE) με βάση τις οποίες αξιολογήθηκε η απόδοση των προβλέψεων που παρήγαγε το κάθε μοντέλο και έγινε η κατάταξη των γραμμικών μοντέλων παρουσιάζονται με αύξουσα σειρά στον Πίνακα 7.1. Παράλληλα υπολογίζονται και παρουσιάζονται επιπλέον στατιστικά κριτήρια αξιολόγησης για το κάθε μοντέλο πρόβλεψης.

Μοντέλο	AIC	RMSE (kWh)	R^2	Corr	MAE (kWh)	MAPE	ME (kWh)	MPE
SARIMA(9,1,6),(2,1,1,12)	-1691.89	4.28	0.84	0.93	2.98	-0.29	-0.006	$-63 \cdot 10^{-5}$
ARIMA(8,1,6)	-1682.61	5.48	0.75	0.87	4.32	0.04	1.46	0.016
ARMA(10,10)	-1631.71	5.93	0.73	0.85	4.20	0.038	-0.19	$-8 \cdot 10^{-4}$
AR(10)	-1422.61	6.09	0.69	0.83	4.57	0.04	-0.44	0.0015
MA(1)	-638.46	19.21	-2.06	-0.1	15.47	0.14	-15.47	-0.14

Πίνακας 7.1: Χαρακτηριστικά αποτελέσματα αξιολόγησης μοντέλων πρόβλεψης. Χρησιμοποιούνται τα κριτήρια αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος (RMSE), μέσου σφάλματος (ME), συντελεστή συσχέτισης (Corr), μέσου απόλυτου σφάλματος (MAE), ποσοστιαίου μέσου απόλυτου σφάλματος (MAPE), ποσοστιαίου μέσου σφάλματος (MPE) και συντελεστή προσδιορισμού (R^2).

Ένα από τα κριτήρια αξιολόγηση, είναι ο συντελεστής προσδιορισμού R^2 . Οι τιμές που μπορεί να λάβει κυμαίνονται από 0% έως 100%. Συνήθως τιμές που προσεγγίζουν το 100% απορρέουν από μοντέλα που αποδίδουν άριστες και έμπιστες προβλέψεις αφού μπορούν να ερμηνεύσουν άριστα τη σχέση που υπάρχει ανάμεσα στη μεταβλητή που προβλέπεται και την

μεταβλητή από την οποία προβλέπεται. Στη περίπτωση μας, όπως φαίνεται από τον Πίνακα 7.3, το μοντέλο SARIMA έχει τη καλύτερη τιμή, την μεγαλύτερη από όλα τα άλλα. Την ίδια στιγμή εμφανίζει την μικρότερη τιμή μέσου τετραγωνικού σφάλματος ρίζας RMSE. Συνολικά παρατηρείται ότι όσο πιο περίπλοκο είναι το μοντέλο που εκπαιδεύτηκε, τόσο πιο μεγάλες τιμές συντελεστή προσδιορισμού λαμβάνει. Είναι ένα κριτήριο το οποίο επιβεβαιώνει την ευελιξία και τις δυνατότητες αυτών των μοντέλων να μπορούν να συλλαμβάνουν καλύτερα τα μοτίβα με αποτέλεσμα να μπορούν να προσαρμοστούν καλύτερα στα δεδομένα εκπαίδευσης και εν τέλη να παράγουν αξιολογες προβλέψεις. Τα πιο απλά μοντέλα που αξιοποιούν για παράδειγμα μόνο μια διαδικασία κινούμενου μέσου αδυνατούν να αντεπεξέλθουν της ερμηνείας πολύπλοκων σχέσεων που υπάρχουν στα δεδομένα και για αυτό λαμβάνουν αρνητικές τιμές.

Παρατηρείται ότι το μέσο τετραγωνικό σφάλμα ρίζας μειώνεται όσο μεταβαίνουμε από ένα πιο απλό μοντέλο (MA, AR) σε ένα πιο σύνθετο (ARMA, ARIMA, SARIMA). Πιο συγκεκριμένα, ο χειρίστος συντελεστής προσδιορισμού R^2 συνάδει με την μεγαλύτερη τιμή RMSE, τις οποίες και εμφανίζει το μοντέλο MA. Πρόκειται για το μοντέλο στο οποίο οι τιμές που εκτίμησε αποκλίνουν περισσότερο από τις πραγματικές, έκανε δηλαδή τις χειρότερες προβλέψεις, όπως εκτιμήσαμε εμπειρικά από το Σχήμα 7.20.

Το μοντέλο ARIMA, κατέχει τη δεύτερη θέση σε αποτελεσματικότητα, με βάση το συντελεστή προσδιορισμού αλλά και τη ρίζα του μέσου τετραγωνικού σφάλματος. Αν και δεν αντιμετωπίζει την περιοδικότητα, το γεγονός και μόνο ότι με την διαφόριση πρώτη βαθμού απαλείφθηκε η τάση, συντέλεσε στην καλύτερη προσαρμογή των δεδομένων αφού αξιοποιήθηκαν τιμές που καθιστούσαν πλέον τη χρονοσειρά στάσιμη και εν τέλη τη παραγωγή πιο αποτελεσματικών και αξιόπιστων προβλέψεων. Ακολουθεί σε απόδοση το ARMA και έπειτα το AR. Τα μοντέλα αυτά αποκλίνουν κατά 5.93 και 6.09 μονάδες ποσοτήτων παραγόμενης ηλεκτρικής ενέργειας αντίστοιχα με βάση το κριτήριο ρίζας μέσου τετραγωνικού σφάλματος. Το γεγονός ότι δεν αντιμετωπίζεται η τάση ή περιοδικότητα της χρονοσειράς που χρησιμοποιούν για να εκπαιδευτούν είναι ο λόγος που δεν υπάρχει ουσιαστική διαφορά στην απόδοση τους. Το ARMA ουσιαστικά εμφανίζει λίγο καλύτερη απόδοση επειδή εμπεριέχει διαδικασίες κινούμενου μέσου σε αντίθεση με το αυτοπαλίνδρομο μοντέλο.

Ένα ακόμη κριτήριο αξιολόγησης το οποίο λαμβάνουμε σοβαρά υπόψιν μας είναι αυτό του συντελεστή συσχέτισης (*corr*). Σε όλα τα μοντέλα πλην του MA η τιμή του ανέρχεται του 0,8. Η πληροφορία αυτή είναι ιδιαίτερη χρήσιμη για την εξαγωγή συμπεράσματος επί της γραμμικής σχέσης που υπάρχει ανάμεσα στις τιμές που προβλέπονται και τις τιμές που έχουν πραγματικά παρατηρηθεί. Στην περίπτωση αυτή, που λαμβάνει τιμή θετική και κοντά στη μονάδα, αντιλαμβανόμαστε ότι υπάρχει πολύ ισχυρή θετική γραμμική σχέση ανάμεσα τους. Ουσιαστικά, η σχέση αυτή υποδηλώνει ότι ακολουθείτε η ίδια κατεύθυνση στις μεταβολές των δύο μεταβλητών τόσο σε μέγεθος όσο και κατεύθυνση. Στον αντίποδα, για το MA μοντέλο όπου ο συντελεστής προσδιορίζεται σχεδόν ίσος με το μηδέν, συμπεραίνετε ότι δεν υπάρχει κάποια σχέση ανάμεσα στις δύο μεταβλητές. Οι τιμές που λαμβάνουν συνεπώς δικαιολογούν επίσης το ποιόν των προβλέψεων που φέρει το κάθε μοντέλο. Όσο πιο ισχυρή η γραμμική σχέση τόσο καλύτερες έφερε και στη πραγματικότητα το κάθε μοντέλο.

Ένα κριτήριο από μόνο του δεν είναι ικανό να παρέχει μια πλήρη αξιολόγηση της απόδοσης

τους οπότε και προσφεύγουμε στη χρήση επιπλέον μέτρων αξιολόγησης που περιγράφονται παρακάτω. Παρόλα αυτά επιδεικνύει την ικανότητα των μοντέλων να προβλέπουν με ασφάλεια την κατεύθυνση με την οποία μεταβάλλεται η παραγωγή ενέργειας από μήνα σε μήνα. Στις επόμενες στήλες του Πίνακα 7.3 συναντάμε τις τιμές που υπολογίστηκαν για τα κριτήρια αξιολόγησης MAE, MAPE, ME, MPE.

Υπολογίζεται σε δύο βήματα το απόλυτο μέσο σφάλμα MAE. Υπολογίζεται στην αρχή το απόλυτο της διαφοράς ανάμεσα στην κάθε προβλεπόμενη τιμή και την αντίστοιχη πραγματική. Αθροίζονται και τέλος διαιρούνται δια του πλήθους αυτών. Έτσι, έχουμε τον μέσο όρο σφάλματος σε απόλυτη μορφή. Είναι μια εκτίμηση που στοχεύει στο μέγεθος του σφάλματος και όχι στο πρόσημο (κατεύθυνση) του. Με άλλα λόγια, υποδηλώνεται ότι για αυτό το κριτήριο όσο πιο χαμηλές τιμές λάβει το εκάστοτε μοντέλο, τόσο πιο μεγάλη ακρίβεια στις προβλέψεις του παρουσιάζει. Είναι ένα ακόμη κριτήριο το οποίο επισημαίνει τη ποιότητα του μοντέλου που παρήγαγε την εκάστοτε πρόβλεψη. Η ερμηνεία του θεωρείται εύκολη, αφού η τιμή του σφάλματος βρίσκεται στην ίδια μονάδα μέτρησης με την μεταβλητή που εξετάζεται. Αναμένεται να έχει το μοντέλο SARIMA τη μικρότερη τιμή όπως και έχει όπως φαίνεται από τον Πίνακα, ενώ την μεγαλύτερη το μοντέλο με τις χειρότερες προβλέψεις, το MA δηλαδή, όπως και συμβαίνει στη προκειμένη περίπτωση. Ακόμη, παρατηρείται ότι τα μοντέλα ARIMA, ARMA και AR έχουν παραπλήσιες τιμές.

Αν και το μοντέλο ARIMA αναμενόταν να έχει μικρότερη τιμή αφού προσδίδει καλύτερα αποτελέσματα, τελικά αυτό δεν συμβαίνει. Σε αυτή τη περίπτωση φανερώνεται το μειονέκτημα του κριτηρίου αυτού και η παραπλάνηση η οποία μπορεί να δημιουργήσει ορισμένες φορές. Το μειονέκτημα του έγκειται στο γεγονός ότι η προβλεπόμενη τιμή όταν αποκλίνει σημαντικά από τη πραγματική τιμή, διογκώνει το σφάλμα αυτό. Έτσι, αν και το ARIMA έφερε καλά αποτελέσματα, η αστοχία του στη 1η και 12η παρατήρηση εν τέλη το κάνει να φαίνεται με βάση αυτό το κριτήριο, χειρότερη επιλογή σε σχέση με το ARMA, ενώ δεν είναι. Φαίνεται, λοιπόν, ο λόγος που γίνεται εξέταση της απόδοσης των μοντέλων με παραπάνω από ένα κριτήριο.

Το κριτήριο μέσου σφάλματος ME εν αντίθεση μας υποδεικνύει εάν υπερεκτιμάται ή υποτιμάται η τιμή που προβλέπεται σε σχέση με την πραγματική. Εάν για παράδειγμα είναι θετικό στο πρόσημο, υπερεκτιμάται η ποσότητα ενέργειας που επρόκειτο να παραχθεί. Στην αντίθετη περίπτωση έχουμε υποεκτίμηση. Κατά αυτόν τον τρόπο πραγματοποιείται ποσοτικοποίηση της συνολικής μεροληψίας (bias) που υπάρχει στην πρόβλεψη. Στα μοντέλα που εξετάζονται εντοπίζουμε υποεκτίμηση σε όλα εκτός του ARIMA, όπου εκεί παρατηρείται υπερεκτίμηση στις ποσότητες ενέργειας που προβλέπεται ότι θα παραχθούν.

Για το μοντέλο SARIMA εντοπίζουμε μια τιμή αρνητική μεν αλλά ελάχιστα αποκλειόμενη από το μηδέν. Μια παρατήρηση η οποία συνάδει με τα εξαιρετικά αποτελέσματα που έφερε η πρόβλεψη με το μοντέλο αυτό, δηλαδή τις σχεδόν άριστα προβλεπόμενες τιμές και την ελάχιστη διαφορά τους από τις πραγματικές. Αντίστοιχα για το MA η τιμή του δε προκαλεί ταραχή. Αναμενόταν να έχει τη χειρότερη τιμή και μάλιστα αρνητική καθώς όπως φαίνεται και από το Σχήμα 7.20 σχεδόν σε όλες τις περιπτώσεις γίνεται υποεκτίμηση στη ποσότητα ενέργειας που παράγεται ανά μήνα. Το SARIMA μοντέλο έχει τέτοια τιμή επειδή άλλοτε υπερεκτιμά και άλλοτε υποτιμά ελάχιστα. Υπάρχει μια ισορροπία η οποία αποδεικνύει την επάρκεια των

δεδομένων για την εκπαίδευση του μοντέλου και τη καταλληλότητα του για το σκοπό που χρησιμοποιείται. Αντίθετα, όταν δεν υπάρχει όπως στη περίπτωση του ARIMA, όπου επειδή στις περισσότερες περιπτώσεις υπερεκτιμήθηκε η τιμή της ποσότητας ενέργειας που επρόκειτο να παραχθεί, η τιμή του κριτηρίου κατέληξε να είναι θετική.

Το ιστόγραμμα υπολοίπων του κάθε μοντέλου ανάλογα την σκίαση που έχει δείχνει εάν υπάρχει υποτίμηση ή υπερεκτίμηση επί τω πλείστων. Υποδηλώνει την τάση του χωρίς όμως να προδικάζει τη συμπεριφορά του σε ένα διάστημα πρόβλεψης. Σε περιπτώσεις όπου του μοντέλου ARIMA το ιστόγραμμα υπολοίπων στο στάδιο εκπαίδευσης υπέδειξε μια ελαφριά δεξιά σκίαση η οποία μεταφράζεται ως μικρή τάση του μοντέλου να υπερεκτιμά. Παρόλα αυτά στο στάδιο των προβλέψεων η τάση του να υπερεκτιμά ήταν μεγαλύτερη. Βέβαια, στα υπόλοιπα μοντέλα η τάση υποεκτίμησης που υπήρχε στο στάδιο εκπαίδευσης συνεχίστηκε και στο στάδιο πρόβλεψης. Συνεπώς, το ιστόγραμμα υπολοίπων μπορεί να προσφέρει χρήσιμη πληροφορία για τη συμπεριφορά του μοντέλου και το τρόπο προσαρμογής του αλλά η χρήση επιπλέον κριτηρίων αξιολόγησης κρίνεται αναγκαία.

Παράλληλα με το MAE, εξετάζεται το ποσοστιαίο μέσο απόλυτο σφάλμα (MAPE). Η χρήση του και η ερμηνεία του δεν εμπεριέχει ιδιαίτερες δυσκολίες. Εκφράζει τη ποσοστιαία διαφορά ανάμεσα στις πραγματικές τιμές της μεταβλητής που εξετάζεται και τις τιμές που προβλέφθηκαν. Όσο πιο μικρή η τιμή που λαμβάνει, τόσο καλύτερο μπορεί να θεωρηθεί το μοντέλο που παρήγαγε τη πρόβλεψη. Στη περίπτωση του SARIMA μοντέλου αναμέναμε τη μικρότερη τιμή, αλλά συναντάται αρνητική τιμή. Το γεγονός αυτό δικαιολογείται από το γεγονός ότι οι τιμές που προβλέπονται είναι μεγαλύτερες των πραγματικών για τους περισσότερους μήνες, όπως φαίνεται και από το Σχήμα 7.23. Η ποινή η οποία αυξάνει γραμμικά σε αυτό το κριτήριο είναι μεγαλύτερη για τα λάθη με αρνητικό πρόσημο. Αντίθετα, τα άλλα μοντέλα στα οποία οι τιμές που προβλέφθηκαν ήταν μικρότερες των αναμενόμενων έχουν θετικό πρόσημο. Φαίνεται πώς ευνοούνται τα μοντέλα που υποεκτιμούν στις προβλέψεις. Τα μοντέλα AR, ARMA και ARIMA, καθώς υποεκτιμούν συνεχώς από την 9η παρατήρηση και έπειτα σε μεγαλύτερο βαθμό από ότι το SARIMA αποφέρουν τελικά την θετική τιμή τους λίγο πάνω από το 0. Φέρουν παραπλήσιες τιμές καθώς μετά από εκείνο το σημείο σε όλα τα μοντέλα έχει φθίνει η απόδοση τους.

7.2.6 Σύγκριση Προβλέψεων με Αντίστοιχα Μοντέλα

Τα αποτελέσματα της ανάλυσης των δεδομένων που έγινε στην παρούσα εργασία συγκρίθηκαν με τα αποτελέσματα της ανάλυσης του Sercan Yeşilöz [36]. Ο Sercan Yeşilöz δεν ακολούθησε ακριβώς τη διαδικασία που περιγράφηκε παραπάνω. Συγκεκριμένα, οι κύριες διαφορές που υπάρχουν ανάμεσα στις δύο αναλύσεις είναι η διαφορετική αντιμετώπιση της τάσης και της εποχικότητας. ο Sercan Yeşilöz επέλεξε να αντιμετωπίσει την εποχικότητα με έναν αλγόριθμο εκθετικής εξομάλυνσης, ενώ πιο πριν είχε εφαρμόσει ένα φίλτρο κινούμενου μέσου (το μήκος του εξεταζόμενου παραθύρου κάθε φορά αριθμούσε τις 12 παρατηρήσεις, δηλαδή, παρατηρήσεις 12 μηνών) για την αντιμετώπιση της τάσης.

Στον αντίποδα, στην ανάλυση που περιγράφηκε παραπάνω η τάση εξαλείφθηκε με διαφόριση

μιας τάξης, ενώ για την αντιμετώπιση της εποχικότητας στη χρονοσειρά χρησιμοποιήσαμε εποχικά μοντέλα ARIMA, τα λεγόμενα SARIMA. Επίσης, τα γραμμικά μοντέλα τα οποία χρησιμοποιήθηκαν τόσο στην μια ανάλυση όσο και στην άλλη ήταν τα AR, MA, ARIMA. Έτσι έμεινε εκτός σύγκρισης το μοντέλο ARMA το οποίο εξετάστηκε στην παρούσα εργασία εκ μέρους μου. Στη συνέχεια, ακολουθεί οι Πίνακας 7.2, ο οποίος περιέχει τις τιμές του στατιστικού σφάλματος MSE με το οποίο αξιολογήθηκαν οι προβλέψεις που παρήγαγαν τα προαναφερόμενα μοντέλα για τη πρόβλεψη της ποσότητας ενέργειας που θα παραγόταν τους επόμενους 12 μήνες.

Μοντέλο	$RMSE_{Exarchakos}$	$RMSE_{Yesiloz}$
AR	0.054	0.145
MA	0.19	0.12
ARIMA	0.048	0.091
SARIMA	0.037	0.091

Πίνακας 7.2: Σύγκριση των προβλέψεων που παρήγαγε με βάση το μέτρο επιβεβαίωσης μέσου τετραγωνισμένου σφάλματος MSE

Οι τιμές που φαίνονται σε αυτόν δεν αφορούν την αρχική κλίμακα στην οποία βρίσκονταν πριν την επεξεργασία τους τα δεδομένα. Τα δεδομένα είχαν μετασχηματιστεί στη λογαριθμική κλίμακα προκειμένου να επιτευχθεί ο στόχος της στασιμότητας στη χρονοσειρά. Ο Yesiloz επέλεξε να μην επαναφέρει στην αρχική κλίμακα τις τιμές των προβλέψεων και των πραγματικών τιμών. Για το λόγο αυτό αφέθηκαν επίσης στην ίδια κλίμακα, την λογαριθμική, τα δεδομένα μου, ώστε να γίνει η σύγκριση των αποτελεσμάτων σε παρόμοια κλίμακα. Σε όλα τα μοντέλα πλην του MA αποδεικνύεται ότι η μεθοδολογία που αναπτύχθηκε εκ μέρους μου προσφέρει πολύ καλύτερα αποτελέσματα από αυτήν του ερευνητή Yesiloz. Ίσως το αποτέλεσμα αυτό να οφείλεται στο γεγονός ότι εκ μέρους μου δαπανήθηκε λίγος περισσότερος χρόνος για την αναζήτηση και επιλογή των βέλτιστων παραμέτρων στο κάθε μοντέλο και όχι τόσο στον τρόπο με τον οποίον έφερε ο κάθε ένας σε κατάσταση στασιμότητας την ίδια χρονοσειρά.

7.3 Ανάλυση Δεδομένων Παραγωγής Ηλεκτρικής Ενέργειας από το Αιολικό Πάρκο 2 στην Θράκη

Από εταιρία διαχείρισης Αιολικών Πάρκων στη Θράκη αποκτήθηκαν τα δεδομένα που θα αξιοποιηθούν στη στατιστική ανάλυση που ακολουθεί. Οι τοποθεσία εγκατάστασης του πάρκου δεν αποκαλύφθηκε οπότε θα χρησιμοποιηθεί η ονομασία Τοποθεσία 2 για το χωρικό προσδιορισμό του αιολικού πάρκου. Το αιολικό πάρκο στην Τοποθεσία 2 είναι εγκατεστημένης ισχύος 16.1 MW και αποτελείται από 7 ανεμογεννήτριες της ENERCON τύπου E-70 E4 με ονομαστική ισχύ 2.3MW η κάθε μια από αυτές. Το πάρκο βρίσκεται σε λειτουργία από το 2017, ωστόσο τα δεδομένα που ανακτήθηκαν, αφορούν το χρονικό διάστημα 01/01/2019 - 31/12/2020. Πρόκειται για μετρήσεις που λαμβάνονταν ανά δέκα λεπτά από τη ΔΕΔΔΗΕ

στον υποσταθμό που βρισκόταν πλησίον του πάρκου αυτού. Το αρχείο .csv το οποίο χρησιμοποιήθηκε, περιείχε σε κάθε σειρά την χρονική στιγμή της μέτρησης και τη ποσότητα ηλεκτρικής ενέργειας (kWh) που προσφερόταν προς κατανάλωση στο δίκτυο χάρη στη λειτουργία του αιολικού πάρκου. Πρόκειται, δηλαδή, να εξετασθούν δεδομένα παραγωγής ηλεκτρικής ενέργειας από αιολικό πάρκο δειγματοληπτημένα ανά δέκα λεπτά.

7.3.1 Προ επεξεργασία Δεδομένων

Το αρχείο περιείχε δεκάδες χιλιάδες τέτοιες μετρήσεις. Ένας τεράστιος όγκος δεδομένων, ο οποίος με τη χρήση ενός απλού συμβατικού υπολογιστή θα οδηγούσε με μαθηματική ακρίβεια σε άκαρπα αποτελέσματα καθώς η ανάλυση τους θα ήταν εξαιρετικά χρονοβόρα. Μετά από ενδελεχή μελέτη και χρονοβόρο ποιοτικό έλεγχο των δεδομένων αποφασίστηκε να αξιοποιηθούν οι μετρήσεις που αφορούν το ημερολογιακό έτος του 2020, από 01/01/2020 έως και 31/12/2020. Αποτελούν 52560 σε πλήθος ακατέργαστα δεδομένα (*raw data*), τα οποία όμως μπορούν να κριθούν ότι είναι πιο ποιοτικά από αυτά του έτους 2019. Ήταν πολύ σημαντικό να υπάρχουν όλες οι μετρήσεις και να μην υπάρχουν μετρήσεις με τιμές 'Nan' ή μηδέν, όπως και ίσχυε τελικά σε αντίθεση με τα δεδομένα για το έτος 2019.

Οι 50 χιλιάδες και πλέον παρατηρήσεις θα δημιουργήσουν μια πολύ μεγάλη χρονοσειρά. Η εκπαίδευση μοντέλων σε μια χρονοσειρά με τέτοιο μήκος θα είχε ανυπολόγιστο χρονικό κόστος. Θα καθιστούσε την διαδικασία παραγωγή προβλέψεων ανέφικτη και άρα ανώφελη. Επίσης, ο ορίζοντας πρόβλεψης, αυτός των προβλέψεων ανά δέκα λεπτά δεν ήταν ο πλέον ενδιαφέρον. Για αυτούς τους λόγους έγινε μια επεξεργασία στο σύνολο τους. Υπολογίστηκε για την κάθε ημέρα η μέση παραγωγή ποσοτήτων ενέργειας. Για δημιουργία κάθε μίας παρατήρησης, γινόταν απορρόφηση των πληροφοριών που περιείχαν 144 αρχικές παρατηρήσεις. Ο αριθμός αυτός δικαιολογείται από το πλήθος των μετρήσεων που πραγματοποιούνταν κάθε μέρα στον υποσταθμό. Έτσι αν αθροιστούν όλες αυτές οι τιμές και κατόπιν διαιρεθούν δια του πλήθους τους (144) προκύπτει η μέση παραγωγή ηλεκτρικής ενέργειας μετρημένη σε kiloWatthours. Η χρονοσειρά η οποία τελικά θα αξιοποιηθεί για την επικείμενη στατιστική ανάλυση αποτελούνταν πλέον από 366 παρατηρήσεις, οι οποίες παράχθηκαν μέσω της επεξεργασίας που επεξηγήθηκε παραπάνω.

Οι αλγόριθμοι που πραγματοποίησαν την ανωτέρα διαδικασία της προ επεξεργασίας των δεδομένων υλοποιήθηκε στο προγραμματιστικό περιβάλλον της *python*.

Πιο ειδικά:

1. Έγινε φόρτωση των απαραίτητων βιβλιοθηκών με τις οποίες θα γίνει και αυτή η ανάλυση και η πρόβλεψη των τιμών της χρονοσειράς. Αυτές ήταν οι *Pandas*, *Matplotlib*, *Numpy*, *Scipy*, *Statsmodels*, *Datetime*, *Fitter* και η *Sklearn*. Κάθε μια από αυτές περιείχε συναρτήσεις και κλάσεις που υποστήριζαν την επίτευξη των σκοπών αυτής της ανάλυσης.
2. Η φόρτωση του αρχείου των δεδομένων πραγματοποιήθηκε με την συνάρτηση *read_csv*. Δημιουργείται μια δομή ετερογενών δεδομένων και δυναμικού μεγέθους *pandas DataFrame*. Στη προκειμένη περίπτωση, είναι ένας διδιάστατος πίνακας, ο οποίος είναι μεταβλητός

και μπορεί να αξιοποιήσει το σύνολο των συναρτήσεων που προσφέρει η βιβλιοθήκη ανάλυσης δεδομένων της `python` η `Pandas`. Αυτός ο πίνακας αποτελείται από μια στήλη στην οποία συναντάται η μέρα και χρονική στιγμή δειγματοληψίας και μια στήλη στην οποία αναφέρεται η τιμή της ενέργειας που παράχθηκε. Κάθε στήλη έχει την επικεφαλίδα της και κάθε γραμμή έναν δείκτη που τις αριθμεί.

3. Στη συνέχεια, προκειμένου να δημιουργηθεί η επιθυμητή χρονοσειρά, αλλάζουμε τη μορφή του. Πιο συγκεκριμένα, αρκεί να αντικαταστήσουμε τον δείκτη που αριθμεί τις γραμμές του πίνακα με τα στοιχεία της στήλης που περιέχει τη χρονική στιγμή δειγματοληψίας της κάθε παρατήρησης της χρονοσειράς. Για ευκολία στη μεταχείριση τους αργότερα, πρότερα της αλλαγής του δείκτη, οι επικεφαλίδες των δύο στηλών, είχαν μετονομαστεί σε `'Date'` και `'energy'` αντίστοιχα. Στη συνέχεια, λαμβάνει μέρος η αναμόρφωση της χρονοσειράς. Για να δημιουργηθεί η χρονοσειρά μέσης παραγωγής ηλεκτρικής ενέργειας με ημερήσιο βήμα, εφαρμόζουμε δύο συναρτήσεις. Αρχικά την `resample` και έπειτα την `mean`. Η πρώτη μετασχηματίζει το βήμα δειγματοληψίας στο οποίο πλέον θα αναφερόμαστε και η δεύτερη υπολογίζει την μέση παραγωγή εντός αυτού του διαστήματος. Έτσι, λαμβάνουμε τελικά τη τελική μορφή του `DataFrame` μας. Αυτή η δομή πλέον, ένα `DataFrame` δηλαδή, το οποίο έχει μια στήλη μόνο στην οποία βρίσκονται οι τιμές της μεταβλητής ενδιαφέροντος μας και όπου για την κάθε μια υπάρχει σαν δείκτης η στιγμή δειγματοληψίας της (χρήση συνάρτησης `set_index`), αποτελεί την εν δυνάμει μονοδιάστατη χρονοσειρά μέσης ποσότητας παραγόμενης ηλεκτρικής ενέργειας ημερήσιου βήματος που θα μελετήσουμε και θα παράγουμε προβλέψεις.
4. Τελευταίο βήμα στην προεπεξεργασία των δεδομένων ήταν η αναζήτηση τυχών τιμών που μπορεί να έλλειπαν από το `dataset` μας και έχουν τιμή `'NaN'` μετά την προεπεξεργασία των τιμών του. Στη περίπτωση αυτή δε βρέθηκε κάποια τέτοια περίπτωση μετά την εφαρμογή των συναρτήσεων `isnull` και `values.any` επί του `DataFrame` που αποτελεί τη δομή που περιγράφει τη χρονοσειρά μας.

7.3.2 Διερευνητική Ανάλυση και Έλεγχος Στασιμότητας

Ακολουθεί η διερευνητική ανάλυση των δεδομένων που εξετάζουμε και δημιουργούν την υπό παρακολούθηση χρονοσειρά. Αποτελεί μια μακροσκελή διαδικασία, στο πλαίσιο της οποίας παράγονται διάφορων ειδών γραφήματα και γίνεται ανάγνωσή για την εξαγωγή χρήσιμων πληροφοριών που αφορούν τη φύση της. Συμπληρωματικά, πραγματοποιούνται έλεγχοι που αποφέρουν ασφαλή συμπεράσματα για την ύπαρξη ή όχι στασιμότητας στη χρονοσειρά.

Πρώτο βήμα στην διαδικασία είναι η παραγωγή και ανάγνωση του γραφήματος απεικόνισης της χρονοσειράς. Στο Σχήμα 7.24 αποτυπώνεται η υπό μελέτη χρονοσειρά. Οι τιμές των ποσοτήτων ενέργειας (kWh) που παρήχθησαν κατά μέσο όρο ανά ημέρα φανερώνονται στον κατακόρυφο άξονα. Στον οριζόντιο άξονα συναντώνται οι χρονικές στιγμές για την οποία κάθε μια από αυτές αντιστοιχίζεται. Ειδικά, το Σχήμα 7.24β' δημιουργήθηκε για να επαληθευτεί ότι είναι γνωστό από πρότερη επεξεργασία. Διαπιστώνεται στην πράξη ότι το χρονικό βήμα

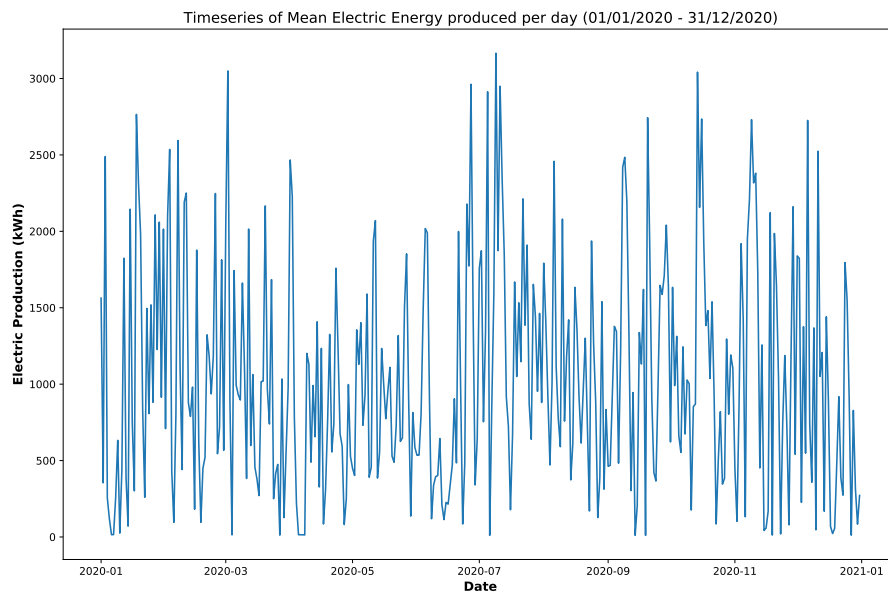
της χρονοσειράς είναι η μια ημέρα, καθώς παρατηρούμε σε αυτό τις πρώτες 48 ημερήσιες παρατηρήσεις της χρονοσειράς.

Με την εμπειρική παρατήρηση του Σχήματος 7.24, θεωρείται εκ πρώτης όψης ότι η χρονοσειρά είναι μη στάσιμη. Η ύπαρξη διακυμάνσεων με διαφορετικό εύρος στις τιμές των μέσων ημερήσιων παραγόμενων ποσοτήτων ενέργειας στο πέρασμα του χρόνου ομολογούν το γεγονός αυτό. Αποτέλεσμα της συμπεριφοράς αυτής είναι ότι η μέση τιμή της αλλάζει επίσης, δεν μένει δηλαδή αναλλοίωτη στο χρόνο. Φαίνεται για παράδειγμα τους ανοιξιάτικους μήνες να λαμβάνει πολύ μικρότερη τιμή από ότι τους καλοκαιρινούς. Επίσης τους καλοκαιρινούς μήνες παρατηρούνται πιο απότομες μεταβολές στη παραγωγή ενέργειας. Πιθανότητα η εποχική ενίσχυση του αιολικού δυναμικού και συνεπώς της παραγωγής του αιολικού πάρκου οφείλεται στην εμφάνιση πρόσκαιρων ανεμολογικών συστημάτων όπως αυτή των μελετεμάτων. Όσον αφορά την τάση της χρονοσειράς, είναι σημαντικό να τονιστεί ότι δε φαίνεται να μπορεί να εξαχθεί ξεκάθαρο συμπέρασμα. Άλλοτε φαίνεται να επικρατεί μια καθοδική πτώση των τιμών και άλλοτε μια ανοδική.

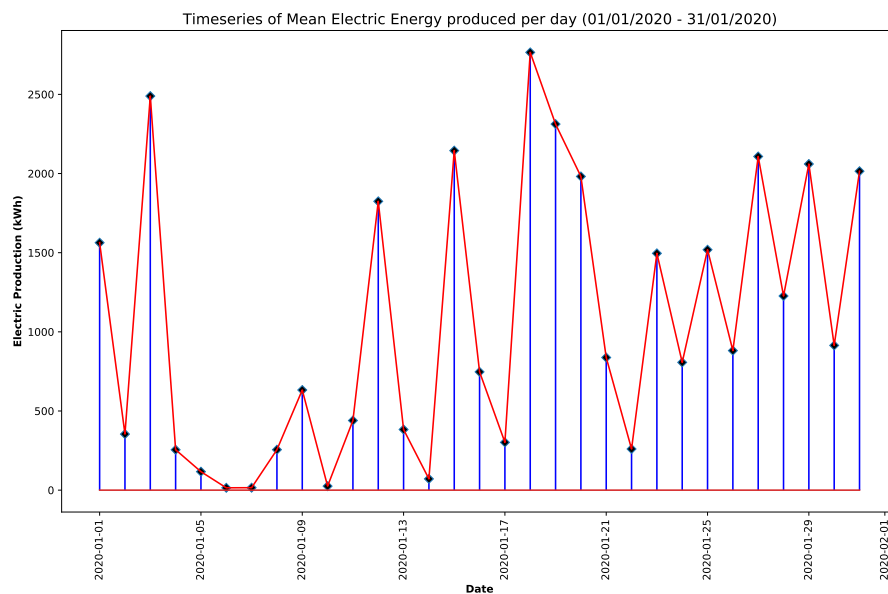
Η συνιστώσα της εποχικότητας της χρονοσειράς ήταν η πλέον δύσκολη να εντοπισθεί. Για το λόγο αυτό παράχθηκε επιπλέον το Σχήμα 7.25. Σε αυτό προβάλλεται η χρονοσειρά σε τέσσερα μικρότερα υποδιαστήματα. Κάθε ένα από αυτά απεικονίζει ένα διάστημα τριών μηνών. Γίνεται μια σύγκριση των μεταβολών ανάμεσα τους και με όση ασφάλεια μπορεί να προσφέρει μια εμπειρική παρατήρηση μπορεί να ειπωθεί ότι υπάρχει εποχικότητα στη χρονοσειρά. Για παράδειγμα, φαίνεται ανάμεσα στα διαστήματα 01/01 - 15/01 και 01/04 - 15/04 να υπάρχει μια συσχέτιση ως προς τον τρόπο που μεταβάλλονται οι τιμές της χρονοσειράς. Άλλο τέτοιο παράδειγμα αποτελούν τα διαστήματα 13/07 - 25/07 και 13/10 - 21/10. Τόσο στο ένα όσο και στο άλλο φαίνεται να μεταβάλλονται κατά παρόμοιο τρόπο οι τιμές των μέσων ποσοτήτων ηλεκτρικής ενέργειας που παράγονταν.

Η εξέταση της χρονοσειράς δεν μένει στον παραπάνω εμπειρικό τρόπο. Για αρχή, δημιουργείται το ιστόγραμμα της χρονοσειράς καθώς επίσης και η εκτιμήτρια της πυκνότητας κατανομής της. Στο Σχήμα 7.26 φαίνονται τα προαναφερόμενα διαγράμματα και επιπλέον ένα τρίτο διάγραμμα. Στο τελευταίο γίνεται ένας συνδυασμός των δύο παραπάνω και παράλληλα διαγράφεται η κανονική κατανομή που θα έπρεπε περίπου να ακολουθεί η χρονοσειρά μας εάν αυτή ήταν στάσιμη. Τα πλέον ενδιαφέροντα διαγράμματα στο Σχήμα 7.26 είναι τα πρώτα δύο. Το άνω γράφημα αποτελεί το ιστόγραμμα, ενώ αυτό που ακολουθεί απεικονίζει την εκτιμήτρια της κατανομής πιθανότητας της χρονοσειράς. Από τα δύο αυτά γραφήματα, φαίνεται ότι η χρονοσειρά μας δεν ακολουθεί μια κανονική κατανομή παρά εμφανίζει μια υψηλή συγκέντρωση στις μικρότερες τιμές της μεταβλητής ενδιαφέροντος και στη συνέχεια εκθετική πτώση των πιθανοτήτων εμφάνισης μεγαλύτερων τιμών αυτής. Υπάρχει, δηλαδή, πολύ μεγαλύτερη πιθανότητα εμφάνισης μικρών σε μέγεθος τιμών παρά πιο μεγάλων. Τέλος, εμφανίζει κάποιες κορυφές με την μεγαλύτερη να εντοπίζεται κοντά στις 1000 kWh. Ίσως αποτελεί κάποιο είδος εκθετικής κατανομής, η οποία φανερώνει την απόκλιση της από την κανονική.

Η μελέτη της κατανομής πιθανότητας που ακολουθεί η χρονοσειρά συνεχίζεται με την εκτίμηση των τριών καλύτερα εφαρμοζόμενων κατανομών επί των δεδομένων της χρονοσειράς. Χρησιμοποιείται η βιβλιοθήκη *Fitter* και των συναρτήσεων της για τον σκοπό αυτό. Έτσι,



(α') Απεικόνιση χρονοσειράς ποσοτήτων ηλεκτρικής ενέργειας που παρήχθησαν κατά μέσο όρο σε καθημερινή βάση στο Αιολικό Πάρκο 2 στη Θράκη καθ' όλη τη διάρκεια του έτους 2020.

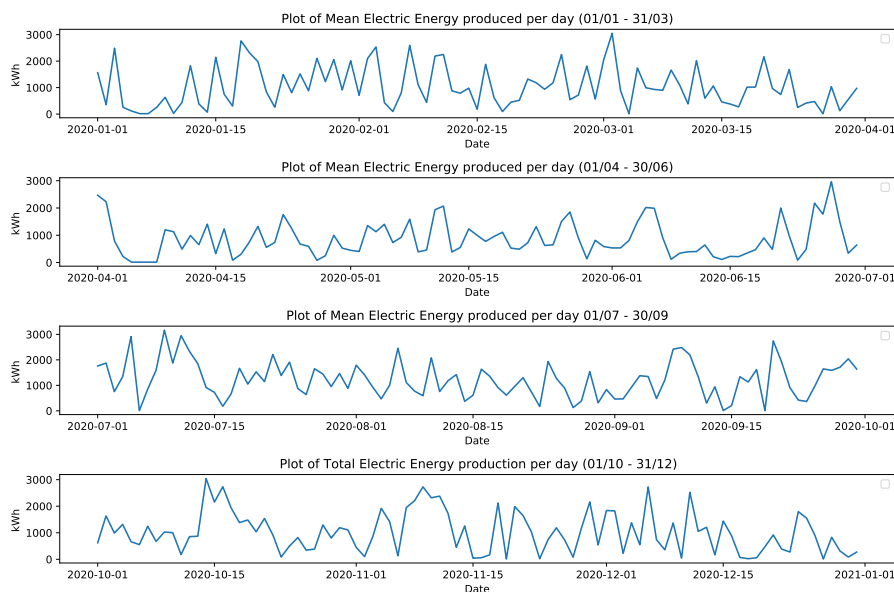


(β') Απεικόνιση των πρώτων 48 παρατηρήσεων της χρονοσειράς προκειμένου να καταστεί εύκολη η εκτίμηση του βήματος της χρονοσειράς.

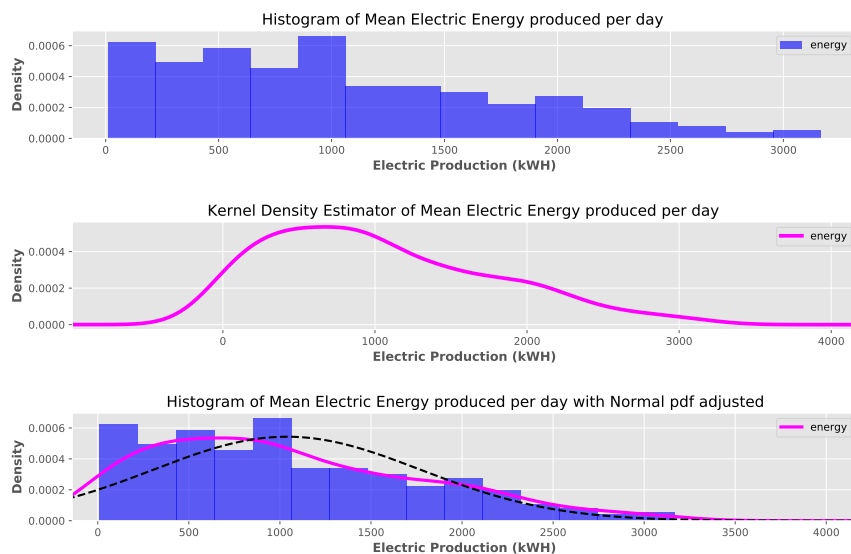
Σχήμα 7.24: Απεικόνιση χρονοσειράς

παράγεται το Σχήμα 7.27.

Σε αυτό φαίνεται το ιστόγραμμα της χρονοσειράς, πάνω στο οποίο αφού έγιναν οι κατάλληλες προσομοιώσεις. Παρουσιάζονται οι τρεις καλύτερα εφαρμοζόμενες κατανομές επί της χρονοσειράς. Είναι οι κατανομές οι οποίες θα μπορούσαν να αναπαραστήσουν την κατανομή την οποία ακολουθούν οι τιμές της μεταβλητής ενδιαφέροντος της χρονοσειράς κατά τον πιο πιστό



Σχήμα 7.25: Απεικόνιση χρονοσειράς παραγωγής ενέργειας με ημερήσιο βήμα σε υποδιαστήματα του έτους 2020.

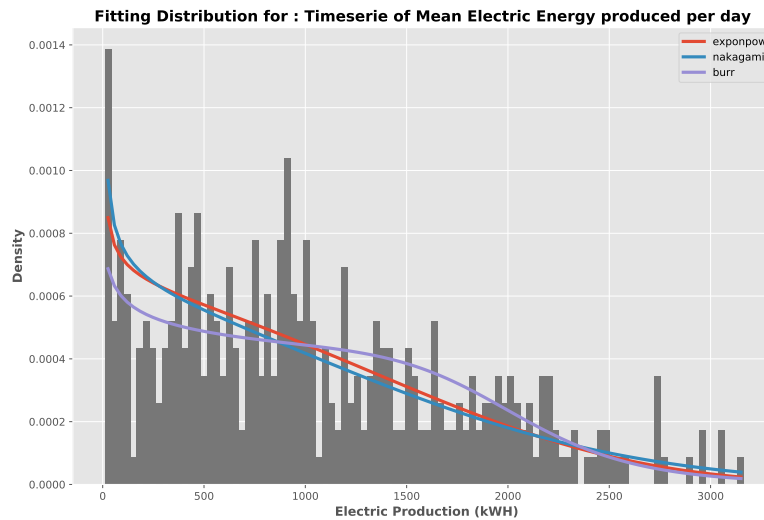


Σχήμα 7.26: Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας (πάνω διάγραμμα). Kernel Density Estimator (μεσαίο διάγραμμα). Συνδυασμός των δύο από πάνω γραφημάτων και σύγκριση των δεδομένων με την κανονικής κατανομής (κάτω διάγραμμα).

τρόπο.

Οι κατανομές που φαίνονται να προσεγγίζουν καλύτερα τον τρόπο που κατανέμονται τα δεδομένα μας είναι οι:

1. Κατανομή εκθετικής δύναμης (exponential power distribution)



Σχήμα 7.27: Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας του Αιολικού πάρκου 2 στη Θράκη.

2. Κατανομή Nakagami

3. Κατανομή Burr

Πρόκειται για κατανομές οι οποίες όπως φαίνεται από το Σχήμα 7.27 κύριο χαρακτηριστικό που παρουσιάζουν είναι μια θετική ή δεξιά σκίαση. Αυτό υποδηλώνει ότι, οι παρατηρήσεις της χρονοσειράς μπορούν να λάβουν τιμές οι οποίες εμπίπτουν στο ένα άκρο τιμών της. Το πρόβλημα είναι ότι είναι πολύ σπάνιο να συμβεί αυτό και ειδικά πολύ πιο σπάνιο από ότι αν έπαιρναν τις τιμές που υπάρχουν στο άλλο άκρο. Καμιά από αυτές δεν προσεγγίζει την κανονική. Τις χαρακτηρίζει η ασυμμετρία. Οι ακραίες τιμές της ουράς επηρεάζουν πολύ την μέση τιμή της χρονοσειράς. Είναι κατανομές οι οποίες έχουν το λεγόμενο σχήμα μπανιέρας (bathtub shaped), αφού φανερώνουν μια μακρυά ουρά στο δεξιό μέρος της κατανομής (δεξιά από την κορυφή).

Εάν η χρονοσειρά μας ακολουθούσε την κανονική κατανομή, τα γραμμικά μοντέλα πρόβλεψης θα απέδιδαν εν τέλη τη καλύτερη δυνατή πρόβλεψη. Για να συμβεί όμως, αυτό, ορισμένες αλλαγές πρέπει, να επέλθουν εις την χρονοσειρά μας. Δεν είναι βέβαιο ωστόσο ότι θα καταστεί επιτυχημένη η προσπάθεια. Οι συγκεκριμένες κατανομές δεν επιτρέπουν εύκολη την επίλυση του προβλήματος σταθεροποίησης της μέσης τιμής και διασποράς στη διάρκεια του χρόνου.

Στη συνέχεια, με σκοπό τη μελέτη των συνιστώσεων που καθιστούν τη χρονοσειρά μας μη στάσιμη δημιουργούνται τα Σχήματα 7.28, 7.30. Είναι τα διαγράμματα αποδόμησης και αυτο-συσχέτισης της χρονοσειράς ποσοτήτων ενέργειας που παρήγαγε το αιολικό πάρκο κατά μέσο όρο σε ημερήσια βάση. Με τη μελέτη αυτών θα επανεξετασθεί και επιβεβαιωθεί με μεγαλύτε-

ρη σιγουριά το ενδεχόμενο ύπαρξης τάσης ή περιοδικότητας ή και των δύο στη χρονοσειρά. Θα αποτελέσει μια πιο ενδελεχή έρευνα επί των συνιστώσεων που προκαλούν συνθήκες μη στασιμότητας στη χρονοσειρά.

Το Σχήμα 7.28, το διάγραμμα αποδόμησης της χρονοσειράς, αποτελείται από τέσσερα υποδιαγράμματα. Αρχικά από πάνω προς τα κάτω απεικονίζεται η χρονοσειρά που εξετάζεται και ακολουθούν τα διαγράμματα απεικόνισης της τάσης, της περιοδικότητας και των υπολοίπων της. Η τάση φαίνεται ότι χαρακτηρίζεται άλλοτε από αυξητική και άλλοτε από καθοδική κατεύθυνση. Δεν είναι σταθερή πορεία της με αποτέλεσμα τη μεταβλητότητα της μέσης τιμής της στη διάρκεια του χρόνου. Για παράδειγμα, παρατηρείται η συμπεριφορά της τάσης τους καλοκαιρινούς μήνες (06/2020 - 08/2020). Γίνεται αντιληπτό ότι η τάση αυξάνεται σταθερά παρά τις μικρές διακυμάνσεις που υφίστανται αρχικά και έπειτα από το μέσον του διαστήματος αυτού ξεκινάει να εμφανίζει μια ανάλογα σταθερή πτώση. Η εξήγηση αυτή ίσως οφείλεται στο ανεμολογικό σύστημα των πρόσκαιρων καλοκαιρινών μελτεμίων. Πρόκειται για ανέμους που εντείνονται και κορυφώνονται τον Ιούλιο και κατόπιν σιγά σιγά εξασθενούν. Στη συνέχεια τους φθινοπωρινούς μήνες, οι μεταβολές είναι πιο έντονες. Παρατηρείται ένα επαναλαμβανόμενο μοτίβο όπου η τάση όσο απότομα αυξάνεται τόσο απότομα μειώνεται. Ίσως το γεγονός ότι οι θερμοκρασίες στο περιβάλλον το φθινόπωρο παρουσιάζουν πιο έντονες μεταβολές καθώς ο καιρός αλλάζει, να δημιουργούν πιο έντονες και πιο απότομες ατμοσφαιρικές πιέσεις και συνεπώς αέρα. Είναι μια προσπάθεια κατανόησης ενός ευμετάβλητου συστήματος, το οποίο ξεκάθαρα αναδεικνύει ότι οι μεταβολές της τάσης της χρονοσειράς επηρεάζονται από την κατά τόπο και χρόνο επικρατούσα ανεμολογική συνθήκη.

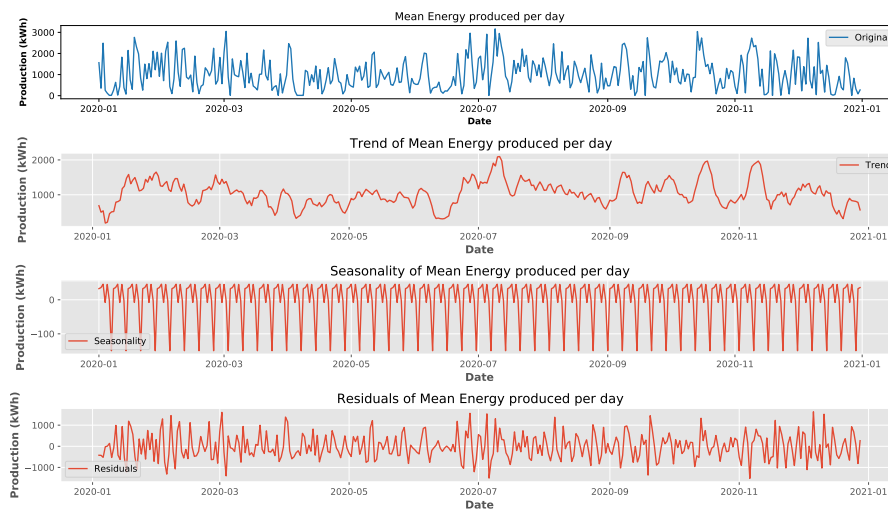
Στο διάγραμμα που ακολουθεί αυτό της τάσης αποτυπώνεται η περιοδικότητα της χρονοσειράς. Μπορούν να καταμετρηθούν 53 περίοδοι σε διάστημα ενός έτους, όσες δηλαδή περίπου οι εβδομάδες ενός ολόκληρου έτους. Συνεπώς, δεν θα ήταν αφελές να ισχυριστούμε ότι η περιοδικότητα που φαίνεται να υπάρχει είναι εβδομαδιαία. Στη διάρκεια ολοκλήρωσης μιας περιόδου φαίνεται ότι αυτή η συνιστώσα προκαλεί πολλές μεταβολές στις παραγόμενες ποσότητες ενέργειας. Αρχικά δείχνει να αυξάνεται έπειτα για λίγο να μειώνεται και μετά να ξανά αυξάνεται έως ότου εμφανιστεί μια τεράστια πρώτη την οποί ακολουθεί μια μεγάλη άνοδος.

Στο τέταρτο και τελευταίο διάγραμμα παρακολουθείται η εξέλιξη των υπολειμμάτων των παρατηρήσεων της χρονοσειράς. Υπενθυμίζεται ότι τα υπόλοιπα της χρονοσειράς προκύπτουν από την αφαίρεση της τάσης και της περιοδικότητας από της υπάρχουσες τιμές. Οι απότομες αλλαγές και οι υψηλές και διαφορετικού μεγέθους διακυμάνσεις ανά διαφορετικά χρονικά διαστήματα δημιουργούν την εικόνα μη στάσιμων υπολοίπων. Η συμπεριφορά που έχουν είναι απρόβλεπτη, άλλοτε οι τιμές τους έχουν μικρές μεταβολές και άλλοτε μεγάλες. Φαίνεται να μην υπάρχει μια σταθερή διακύμανση των τιμών τους στη διάρκεια εξέλιξης της χρονοσειράς.

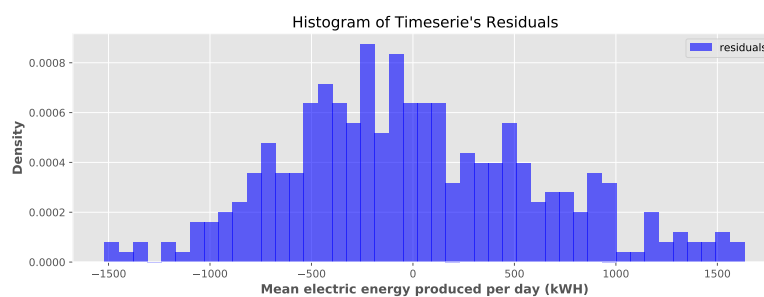
Την πιο ξεκάθαρη εικόνα, όμως, για την φύση των υπολοίπων της χρονοσειράς, δίδει το ιστόγραμμα τους. Στο Σχήμα 7.29 παρουσιάζεται αυτό και μελετάται προκειμένου να εξαχθούν οι πλέον ασφαλείς πληροφορίες που αφορούν την κατανομή πυκνότητας πιθανότητας που ακολουθούν τα υπόλοιπα της χρονοσειράς μας. Η κατανομή πυκνότητας πιθανότητας που φαίνεται να ακολουθούν δεν είναι σίγουρα η κανονική. Αρχικά, φαίνεται να μη παρουσιάζει μόνο μια κορυφή. Η μεγαλύτερη κορυφή από όσες εμφανίζονται βρίσκεται μετατοπισμένη

προς τις αρνητικές τιμές του διαγράμματος. Ακόμη υπάρχει, μια μακρυά ουρά στα δεξιά μαζί με δύο πιο μικρές κορυφές. Είναι μια παρατήρηση την οποία κάναμε επίσης στο ιστόγραμμα της συνολικής χρονοσειράς. Δικαιολογείται κατά αυτόν τον τρόπο ο αρχικός ισχυρισμός μας μετά την ανάγνωση του γραφήματος απεικόνισης των υπολοίπων της χρονοσειράς.

Απουσία της κανονικής κατανομής τα γραμμικά μοντέλα πρόβλεψης δεν πρόκειται να λειτουργήσουν κατά το δοκούν. Στη περίπτωση που διαπιστωθεί ότι ακολουθούν μια κανονική κατανομή με μέση τιμή 0 και σταθερή διασπορά, είναι δηλαδή λευκός θόρυβος, τότε γνωρίζουμε εκ του ασφαλούς ότι τα γραμμικά μοντέλα πρόβλεψης που θα εφαρμοστούν επί των δεδομένων της χρονοσειράς θα προσφέρουν ένα ασφαλές αποτέλεσμα. Ωστόσο, φαίνεται ότι η διασπορά των τιμών γύρω από τη μέση τιμή δεν είναι συμμετρική. Η μέση τιμή της δεν προσεγγίζει το 0 επίσης και συνολικά από την παρατήρηση του εξάγεται το συμπέρασμα ότι δεν προσεγγίζει το λευκό θόρυβο. Συνεπώς, είναι βέβαιο ότι πρέπει να γίνουν αλλαγές επί των δεδομένων πριν εφαρμόσουμε οποιοδήποτε γραμμικό μοντέλο, εάν θέλουμε η προσπάθειά μας να καρποφορήσει.



Σχήμα 7.28: Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων.



Σχήμα 7.29: Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη.

Τελευταίο διάγραμμα που θα εξετασθεί είναι αυτό της αυτοσυσχέτισης της χρονοσειράς. Στο Σχήμα 7.30 αποτυπώνεται το γράφημα αυτό για τις πρώτες 40 χρονικές υστερήσεις lags της χρονοσειράς. Με την παρατήρηση του δεν εξάγονται μόνο πληροφορίες για τις αυτοσυσχετίσεις που υφίστανται μεταξύ παρατηρήσεων που απέχουν k lags, παρά είναι δυνατό να συλλεχθούν πληροφορίες με τις οποίες μπορεί να αποσαφηνιστεί τελικά η ύπαρξη ή όχι στασιμότητας στη χρονοσειρά. Αρχικά παρατηρείται η εκθετική πτώση των τιμών αυτοσυσχέτισης εντός των διαστημάτων εμπιστοσύνης (πρακτικά μηδέν) έπειτα από ένα μόλις lag. Φαίνεται, ότι η συνιστώσα της τάσης την οποία εντοπίσαμε προηγουμένως, δεν μπορεί να ανιχνευθεί από το διάγραμμα αυτό σε αυτή τη περίπτωση, όπως επίσης ούτε αυτή της εβδομαδιαίας περιοδικότητας. Δεν παρατηρείται, δηλαδή, αργή μείωση των τιμών εκτός διαστημάτων εμπιστοσύνης με κάποιο επαναλαμβανόμενο μοτίβο παρά εκθετική με ισχυρή αυτοσυσχέτιση να εμφανίζεται μόνο για $lag = 1$.

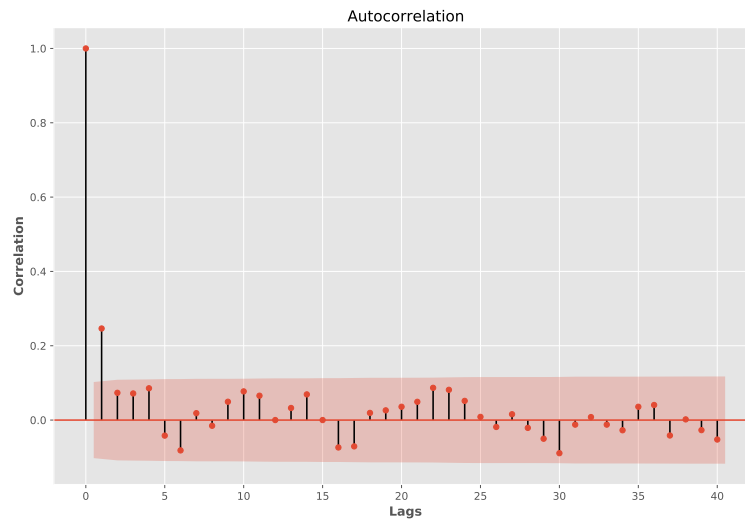
Από την παρατήρηση του διαγράμματος αυτοσυσχέτισης θα συμπεραίναμε ότι η χρονοσειρά μας είναι στάσιμη και ότι πιθανότατα ένα AR μοντέλο τάξης 1 θα ήταν ικανό να προσαρμοστεί κατάλληλα στα δεδομένα και να αποδώσει ωφέλιμες προβλέψεις εκτός δείγματος. Παρόλα αυτά δεν εμπιστευόμαστε τον ισχυρισμό αυτό καθώς έρχεται σε αντίθεση με όσα προσεκτικά ανακαλύψαμε προηγουμένως από το ιστόγραμμα και τα διαγράμματα απεικόνισης και αποδόμησης της χρονοσειράς. Αυτό που πραγματικά συμβαίνει είναι ότι πρόκειται για μια χρονοσειρά ευμετάβλητη και με απρόβλεπτη φύση με την κάθε παρατήρηση να μην έχει εξαρτήσεις από παρελθοντικές τιμές. Εμφανίζει υψηλή τυχαιότητα και οι συνιστώσες τάσης και περιοδικότητας δεν γίνεται να περιγραφούν από κάποια γραμμική σχέση. Δε παραμένουν σταθερές στο διάστημα μελέτης της χρονοσειράς. Φαίνεται ότι η μοντελοποίηση της παρούσας χρονοσειράς από κάποιο γραμμικό σύστημα είναι ένα πολύ δύσκολο εγχείρημα.

Εάν είχαμε διενεργήσει μόνο έναν από τους παραπάνω ελέγχους στασιμότητας ίσως να είχαμε οδηγηθεί σε λάθος συμπέρασμα. Στη στατιστική ανάλυση χρονοσειρών πρέπει να βολιδοσκοπούμε ένα πρόβλημα από πολλές οπτικές, ώστε να μην ξεγελαστούμε και χάσουμε τελικά πολύτιμες πληροφορίες.

Τελευταίο μέτρο επιβεβαίωσης της μη στάσιμης χρονοσειράς υπήρξε η διενέργεια του ελέγχου Augmented Dickey Fuller Test. Η ολοκλήρωση του και η ανάγνωση του υπέδειξε ότι εξετάζεται μια μη στάσιμη χρονοσειρά. Σε συνδυασμό με τα παραπάνω πάρθηκε η απόφαση μετατροπής της μη στάσιμης χρονοσειράς σε στάσιμη πριν την αξιοποίηση γραμμικών μοντέλων για την πρόβλεψη τιμών της ενδιαφερόμενης μεταβλητής μας.

7.3.3 Μετατροπή της Χρονοσειράς σε Στάσιμη

Με τις διερευνητικές ενέργειες, που περιγράφηκαν στις προηγούμενες ενότητες, κατανοήθηκε η φύση και η συμπεριφορά της χρονοσειράς στην διάρκεια του χρόνου στον οποίο μελετάται. Με μεγάλη βεβαιότητα, εξαχθή ένα πολύ σημαντικό συμπέρασμα. Η χρονοσειρά μας δεν βρίσκεται σε τέτοια κατάσταση, ώστε να αποφέρει αποτελεσματικές προβλέψεις. Δεν είναι στάσιμη και δεν δύναται να κατανοηθεί η συμπεριφορά της στην εξέλιξη του χρόνου. Τα γραμμικά μοντέλα δεν γίνεται να προσαρμοστούν στα δεδομένα εκπαίδευσης και να περιγράψουν



Σχήμα 7.30: Διαγράμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη-ACF plot (40 lags).

τις σχέσεις που αναπτύσσονται. Η αξιοποίηση της χρονοσειράς στη παρούσα κατάσταση θα οδηγήσει σε φτωχά αποτελέσματα στη διαδικασία πρόβλεψης αφού η φύση της είναι εξαιρετικά απρόβλεπτη.

Προκειμένου η χρονοσειρά να γίνει στάσιμη ή να έλθει πιο κοντά στην κατάσταση αυτή, ορισμένες αλλαγές πρέπει να γίνουν σε αυτή. Ο μετασχηματισμός των δεδομένων έχει σημαντικό ρόλο στην επίτευξη του στόχο αυτού. Με τη βοήθεια μετασχηματισμών δύναμης Box-Cox δημιουργούνται νέες χρονοσειρές. Δεν είναι βέβαιο ότι θα αποφέρουν χρονοσειρές στάσιμες και για αυτό το λόγο μελετώνται διεξοδικά πριν την αξιοποίησή τους για τις τελικές προβλέψεις. Κάποιες από αυτές φαίνεται να σταθεροποιούν καλύτερα τη διασπορά και τη μέση τιμή στην εξέλιξη του χρόνου. Παράλληλα, για τη κατανομή των υπολοίπων τους εξετάζεται η κατανομή πυκνότητας πιθανότητας. Σε περίπτωση που προσεγγίζουν την κανονική καλύτερα από πριν, τότε ο συγκεκριμένος μετασχηματισμός κερδίζει επιπλέον βαθμούς εμπιστοσύνης.

Στα υπάρχοντα δεδομένα εφαρμόστηκαν οι εξής μετασχηματισμούς δύναμης:

- άναστροφος(reciprocal)
- άναστροφος τετραγωνικής ρίζας (reciprocal square root)
- τετραγωνικής ρίζας (square root)
- τρίτης ρίζας (cube root)
- λογαριθμικός (log).

Η επιλογή της μετασχηματισμένης χρονοσειράς γίνεται μετά την εξαγωγή συμπερασμάτων από την παρατήρηση κάποιων διαγραμμάτων. Προτού γίνει η προαναφερόμενη επιλογή, μελετήθηκαν τα γραφήματα απεικόνισης, αποδόμησης και αυτοσυσχέτισης καθώς επίσης και τα

ιστογράμματα όλων των μετασχηματισμένων χρονοσειρών. Στην συνέχεια, διενεργήθηκε ο έλεγχος ADF και αναγνώστηκαν τα αποτελέσματα του. Οι μετασχηματισμένες χρονοσειρές που φάνηκε να παραμένουν μη στάσιμες και παράλληλα δεν πέρασαν τον έλεγχο ADF απορρίφθηκαν και έπαψε η μελέτη τους. Έτσι οι μετασχηματισμένες χρονοσειρές που απέμειναν και των οποίων η μελέτη και σύγκριση θα παρουσιαστεί είναι αυτές που έφεραν τα δεδομένα σε λογαριθμική κλίμακα και σε κλίμακα τετραγωνικής ρίζας. Γίνεται παράλληλη μελέτη και σύγκριση, έτσι ώστε να αποφασιστεί ποια από τις δύο επιλογές θα επιλεγεί για να επιφέρει καλύτερες προβλέψεις.

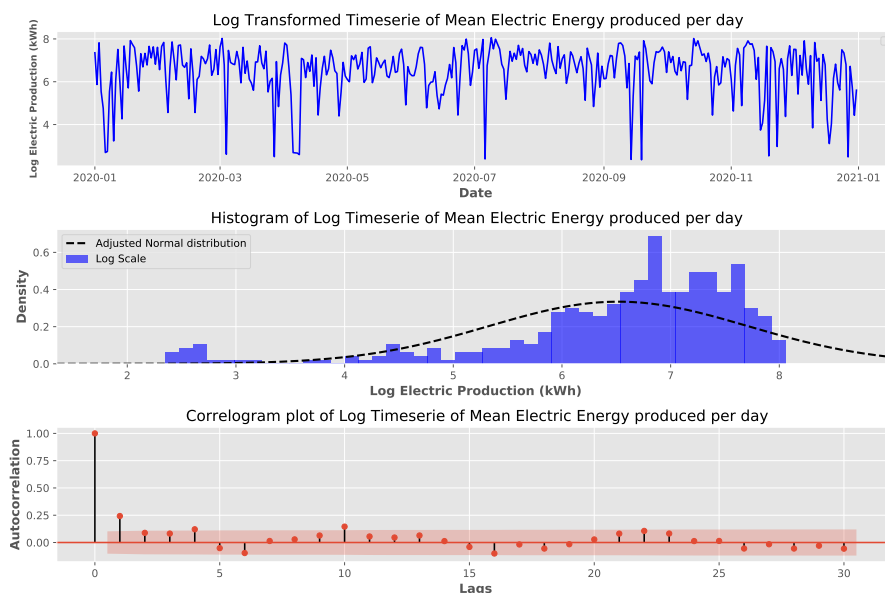
Στη προσπάθεια αποσαφήνισης της χρονοσειράς που θα επιλεγεί για τη παραγωγή των προβλέψεων μελετώνται αρχικά τα Σχήματα 7.31 και 7.33. Στο καθένα από αυτά βρίσκουμε συγκεντρωτικά τρία γραφήματα. Από πάνω προς τα κάτω, παρουσιάζονται το γράφημα απεικόνισης, το ιστόγραμμα και το διάγραμμα αυτοσυσχέτισης της εκάστοτε μετασχηματισμένης χρονοσειράς. Στο Σχήμα 7.31 παρουσιάζονται τα διαγράμματα για τη χρονοσειρά μέσω των ποσοτήτων παραγόμενης ηλεκτρικής ενέργειας από το αιολικό πάρκο σε ημερήσια βάση σε λογαριθμική κλίμακα, ενώ στο Σχήμα 7.33 παρουσιάζονται αυτά που αφορούν τη χρονοσειρά μέσω των ποσοτήτων παραγόμενης ηλεκτρικής ενέργειας από το αιολικό πάρκο σε ημερήσια βάση σε κλίμακα τετραγωνικής ρίζας.

Πρώτα θα μελετηθούν τα διαγράμματα του Σχήματος 7.31 για τη χρονοσειρά σε λογαριθμική κλίμακα. Από το διάγραμμα απεικόνισης της χρονοσειράς συμπεραίνεται ότι ο συγκεκριμένος μετασχηματισμός επέφερε σημαντικές και θετικές αλλαγές. Η διασπορά και η μέση τιμή έχουν σταθεροποιηθεί σε σημαντικό βαθμό σε κάποια διαστήματα αφού με αυτόν τον μετασχηματισμό συρρικνώνεται το εύρος τιμών των παρατηρήσεων της χρονοσειράς. Πλέον, για μεγαλύτερα χρονικά διαστήματα παρατηρούνται ομοίου ύψους διακυμάνσεις στις τιμές των μέσων ποσοτήτων ενέργειας. Ωστόσο, ανάμεσα σε αυτά τα διαστήματα υπάρχουν διαστήματα με ανομοιόμορφες και μεγάλες διακυμάνσεις στις τιμές της μεταβλητής, δείγμα του ότι δεν έχει σταθεροποιηθεί σε όλο το μήκος της χρονοσειράς η διασπορά της. Δε παύει, ωστόσο, να είναι είναι μια θετική εξέλιξη.

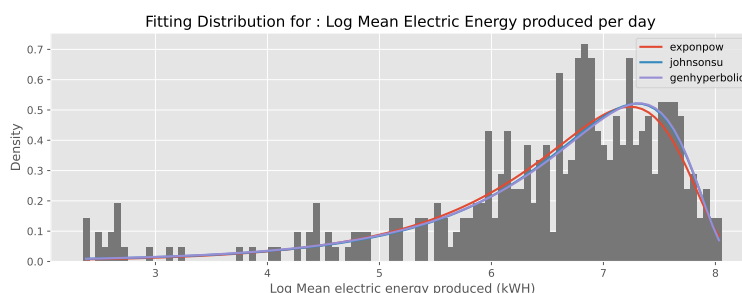
Το επόμενο διάγραμμα που ερευνάται και αποτελεί το ιστόγραμμα της χρονοσειράς. Η κατανομή που φαίνεται να ακολουθεί η χρονοσειρά δεν αποτελεί δείγμα στάσιμης χρονοσειράς όπως και αναμενόταν. Ωστόσο, η ουρά που προϋπήρχε στα δεξιά του ιστογράμματος έχει αντιμετωπιστεί και πλέον η χρονοσειρά μας φαίνεται να είναι σχισμένη αρνητικά αποτέλεσμα της λογαριθμικής μετατροπής των δεδομένων. Με διακεκομμένη μαύρη γραμμή φαίνεται η εφαρμοζόμενη κανονική κατανομή επί των συγκεκριμένων δεδομένων. Φαίνεται και είναι ξεκάθαρη η απόκλιση της χρονοσειράς που μελετάται από την κανονική. Επόμενο και τελευταίο γράφημα στο Σχήμα 7.31 είναι αυτό της αυτοσυσχέτισης της χρονοσειράς. Σε αυτό είναι εύκολο να παρατηρηθεί η εκθετική πτώση των τιμών της εντός διαστημάτων εμπιστοσύνης μετά από τη πρώτη χρονική υστέρηση (lag). Η αλλαγή που υπάρχει σε σχέση με πριν είναι ότι η συνιστώσα της περιοδικότητας είναι πλέον πιο εμφανής. Υπάρχει η τάση μετά από 6 lags η τιμή της αυτοσυσχέτισης να ξαναεμφανίζεται εκτός των διαστημάτων εμπιστοσύνης.

Η χρονοσειρά παραμένει μη στάσιμη προς το παρόν αλλά υπάρχουν ενδείξεις και χαρακτηριστικά που ίσως να μπορούν πλέον να μοντελοποιηθούν ώστε να περιγραφεί με ακρίβεια αυτή.

Γίνεται αντιληπτό, λοιπόν, ότι ο λογαριθμικός μετασχηματισμός δεν εξαλείφει τις συνιστώσες τάσης και περιοδικότητας που δημιουργούν μια μη στάσιμη χρονοσειρά παρά σταθεροποιεί για κάποια χρονικά διαστήματα τη μέση τιμή και διασπορά της, μειώνοντάς κατά αυτόν τον τρόπο το μεγάλο πλήθος απότομων, μεγάλων και απρόβλεπτων διακυμάνσεων των τιμών της μεταβλητής ενδιαφέροντος. Η λογαριθμική χρονοσειρά φαίνεται να προσεγγίζει καλύτερα μια κατανομή εκθετικής δύναμης σύμφωνα με το Σχήμα 7.32.



Σχήμα 7.31: Απεικόνιση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα Αιολικού πάρκου 2 στη Θράκη (πάνω διάγραμμα). Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα Αιολικού πάρκου 2 στη Θράκη (μεσαίο διάγραμμα). Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα Αιολικού πάρκου 2 στη Θράκη-ACF plot (κάτω διάγραμμα).



Σχήμα 7.32: Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα του Αιολικού πάρκου 2 στη Θράκη.

Ακολούθως έγινε η παρατήρηση των διαγραμμάτων του Σχήματος 7.33, που αφορούν

την χρονοσειρά, η οποία φέρει πλέον τις τιμές της σε κλίμακα τετραγωνικής ρίζας. Είναι ένας μετασχηματισμός που είναι τόσο επιθετικός στην συρρίκνωση του εύρους τιμών της χρονοσειράς όσο αυτός που φέρει τη χρονοσειρά σε λογαριθμική κλίμακα. Αρχικά τίθεται υπό διερεύνηση το διάγραμμα απεικόνισης της. Ο συγκεκριμένος μετασχηματισμός δύναμης δεν μίκρυνε τόσο πολύ τις τιμές της μέσης παραγόμενης ενέργειας παρά τις έφερε σε μια κλίμακα η οποία όμως είναι πιο προσφιλή και ευανάγνωστη. Για παράδειγμα, τα διαστήματα στα οποία η τάση εμφανίζει μια πτωτική πορεία διακρίνονται πιο εύκολα όπως ακόμη είναι πιο εύκολα να διακριθεί το γεγονός ότι υπάρχουν επαναλαμβανόμενα μοτίβα στη χρονοσειρά. Υπάρχουν διαστήματα που φαίνεται να έχει σταθεροποιηθεί η διακύμανση των τιμών της χρονοσειράς αλλά συνολικά παρατηρείται ότι η χρονοσειρά δεν έχει σταθερή διασπορά και μέση τιμή. Η διασπορά και η μέση τιμή έχουν σταθεροποιηθεί σε μικρότερο βαθμό σε σχέση με την πρώτη μετασχηματισμένη χρονοσειρά μας και για πολύ μικρότερα χρονικά διαστήματα. Πλέον, για μεγαλύτερα χρονικά διαστήματα παρατηρούνται ανισοϋφείς διακυμάνσεις στις τιμές των μέσων ποσοτήτων παραγόμενης ενέργειας.

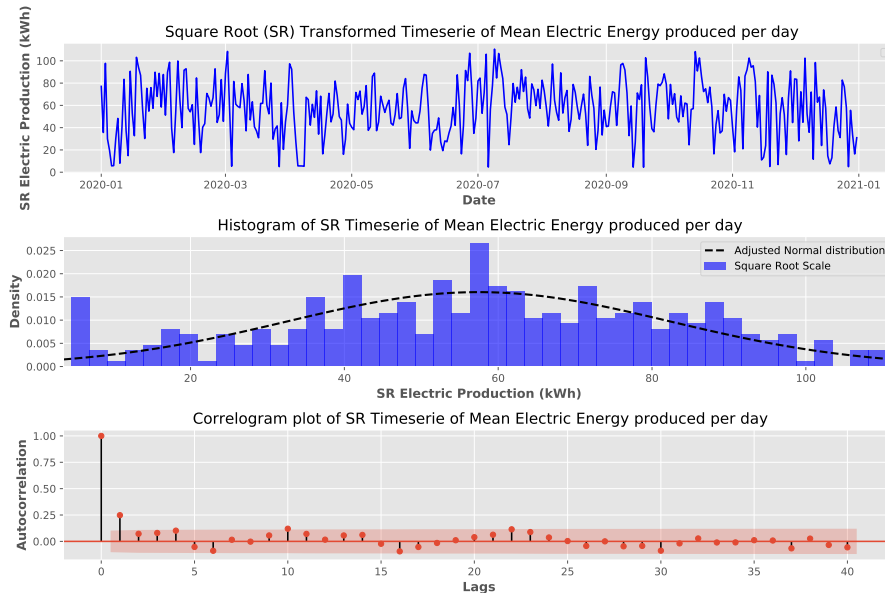
Στη συνέχεια, ερευνάται, το επόμενο διάγραμμα το οποίο και αποτελεί το ιστόγραμμα της χρονοσειράς. Από το συγκεκριμένο διάγραμμα διαπιστώνεται ότι η κατανομή που φαίνεται πλέον να ακολουθεί η χρονοσειρά δεν είναι η κανονική. Πρώτον, παρουσιάζει μια υψηλή συγκέντρωση τιμών κοντά στο μηδέν. Δεύτερον, η δεξιά σκίαση που υπήρχε στην αρχική χρονοσειρά μπορεί να μην υφίσταται πλέον αλλά έχουν δημιουργηθεί αρκετές κορυφές γύρω από την μεγαλύτερη. Παρόλα αυτά, φαίνεται να έχουν κατανεμηθεί στα υπόλοιπα διαστήματα τιμών, οι τιμές καλύτερα από ότι στο ιστόγραμμα της λογαριθμικά μετασχηματισμένης χρονοσειράς όσον αφορά τη προσέγγιση στη κανονική κατανομή.

Στη χρονοσειρά που εξετάζεται γίνεται μια προσπάθεια εύρεσης της κατανομής πυκνότητας πιθανότητας που εφαρμόζεται καλύτερα επί των δεδομένων. Στο Σχήμα 7.34 παρουσιάζεται το αποτέλεσμα της προσπάθειας εύρεσης της κατανομής αυτής. Ως βέλτιστη εμφανίζεται η κατανομή Von Mises. Πρόκειται για την κυκλική κανονική κατανομή, κάτι το οποίο ισχυροποιεί την άποψή μας περί μη προσέγγισης της κανονικής κατανομής. Η συγκεκριμένη κατανομή δεν υποδεικνύει στασιμότητα παρά ότι υπάρχουν κυκλικά επαναλαμβανόμενα μοτίβα γύρω από τη μέση τιμή της.

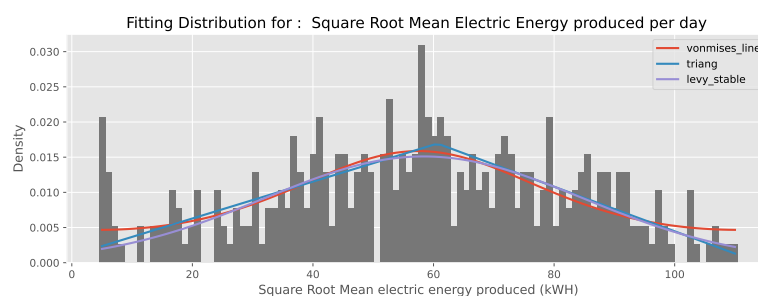
Τελευταίο γράφημα στο Σχήμα 7.33 είναι αυτό της αυτοσυσχέτισης της χρονοσειράς. Είναι ένα γράφημα από το οποίο θα εξάγουμε περισσότερες πληροφορίες για τις συνιστώσες που καθιστούν τη χρονοσειρά μας μη στάσιμη. Δε μπορεί παρά να παρατηρηθεί η εκθετική πτώση των τιμών της αυτοσυσχέτισης εντός διαστημάτων εμπιστοσύνης μετά από τη πρώτη υστέρηση lag. Εάν γινόταν μόνο αυτή η παρατήρηση, θα μπορούσε να θεωρηθεί στάσιμη προς το παρόν. Ωστόσο, ισχυρή φαίνεται να είναι η συνιστώσα της περιοδικότητας. Προσεγγιστικά το μήκος της σε υστερήσεις είναι 6, δηλαδή, 6 μέρες. Πιο απλά, η περιοδικά επαναλαμβανόμενη εξέλιξη στις τιμές της αυτοσυσχέτισης συντελείται σχεδόν ανά εβδομάδα. Συνεπώς, δεν γίνεται παρά να στηριχθεί ο ισχυρισμός περί μελέτης μιας μη στάσιμης χρονοσειράς με ισχυρά επαναλαμβανόμενα μοτίβα ανά 6 ημέρες.

Ο μετασχηματισμός (τετραγωνικής ρίζας) δεν εξαλείφει τις συνιστώσες τάσης και περιοδικότητας. Σταθεροποιεί για μικρά χρονικά διαστήματα τη μέση τιμή και διασπορά της και

φέρει μια κατανομή κοντά στη μορφή της κυκλικής κανονικής. Παρόλα αυτά τόσο σε αυτή τη περίπτωση όσο και στη περίπτωση της χρονοσειράς του λογαριθμικού μετασχηματισμού επιπλέον αλλαγές πρέπει να συντελεστούν ώστε να προκύψει μια στάσιμη χρονοσειρά. Μια χρονοσειρά, η οποία στη συνέχεια θα μπορεί να αποδώσει άξιες και έμπιστες προβλέψεις.



Σχήμα 7.33: Απεικόνιση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας Αιολικού πάρκου 2 στη Θράκη (πάνω διάγραμμα). Ιστόγραμμα χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας Αιολικού πάρκου 2 στη Θράκη (μεσαίο). Διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας Αιολικού πάρκου 2 στη Θράκη-ACF plot (κάτω διάγραμμα).



Σχήμα 7.34: Εκτίμηση των τριών καλύτερων κατανομών πιθανότητας που προσαρμόζονται κατά το βέλτιστον στη χρονοσειρά παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας του Αιολικού πάρκου 2 στη Θράκη.

Τελευταίο βήμα πριν την εφαρμογή των κατάλληλων μεθόδων για την αντιμετώπιση των συνιστώσεων που καθιστούν τη χρονοσειρά μας μη στάσιμη, είναι η περαιτέρω μελέτη των συγκεκριμένων. Για το λόγο αυτό παράγονται τα διαγράμματα αποδόμησης των δύο μετασχη-

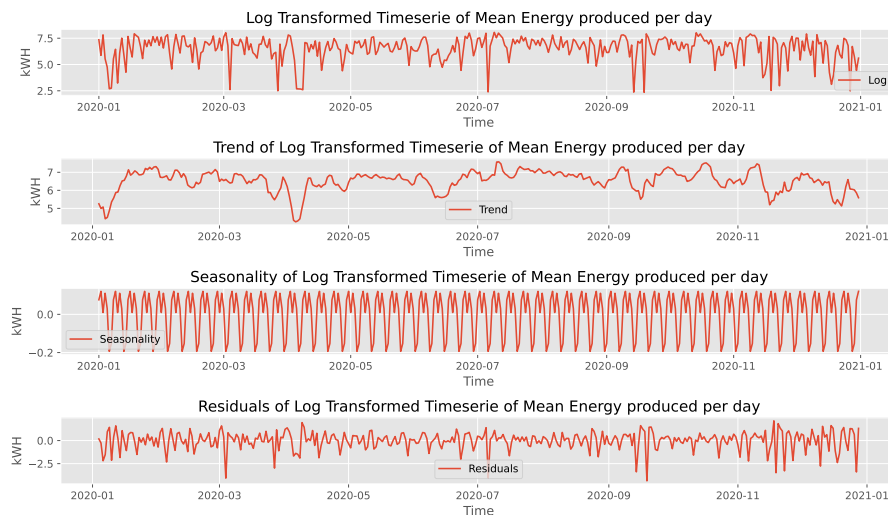
ματισμένων χρονοσειρών. Στο Σχήμα 7.35 αποτυπώνεται αυτό που αφορά την λογαριθμικό μετασχηματισμό ενώ αυτό που φανερώνεται από το Σχήμα 7.37 αφορά το διάγραμμα αποδόμησης της χρονοσειράς που βρίσκεται σε κλίμακα τετραγωνικής ρίζας. Και τα δύο περιέχουν τέσσερα υποδιαγράμματα. Από πάνω προς τα κάτω κατά σειρά τοποθετούνται το διάγραμμα απεικόνισης της χρονοσειράς και ακολουθούν τα διαγράμματα απεικόνισης της τάσης, της περιοδικότητας και των υπολοίπων της χρονοσειράς. Βαρύτητα κυρίως θα δοθεί στα διαγράμματα απεικόνισης τάσης και περιοδικότητας.

Αρχικά εξετάζεται το Σχήμα 7.35 που αποτυπώνει την αποδόμηση της λογαριθμικά μετασχηματισμένης χρονοσειράς. Παρατηρείται ότι η τάση επηρεάζει συναρτήσει της περιοδικότητας έντονα την τελική τιμή της μέσης ημερήσιας ενέργειας. Η τάση δεν φαίνεται να έχει σταθερή κατεύθυνση. Είτε είναι αυξητική είτε είναι πτωτική είτε σταθερή. Σε κάθε περίπτωση επηρεάζει ανάλογα την τιμή της μεταβλητής ενδιαφέροντος. Για παράδειγμα, παρατηρείται η συμπεριφορά της τάσης τους καλοκαιρινούς μήνες (07/2020 - 09/2020). Η τάση τους μήνες αυτούς φανερώνει πτώση στις τιμές της παρά τις μικρές διακυμάνσεις που υπάρχουν. Αρχικά και έπειτα από το μέσον του διαστήματος αυτού ξεκινάει να εμφανίζει μια ανάλογα σταθερή πτώση. Η συμπεριφορά αυτή οφείλεται στο ανεμολογικό σύστημα των πρόσκαιρων καλοκαιρινών μελτεμίων. Πρόκειται για ανέμους που αυξάνουν την ένταση τους έως ότου κορυφωθούν τον Ιούλιο. Μετά από αυτό το σημείο σιγά σιγά εξασθενούν, με αποτέλεσμα να υπάρχει μείωση στη ποσότητα αιολικής ενέργειας που μπορεί να αξιοποιηθεί για να παραχθεί ηλεκτρική ενέργεια.

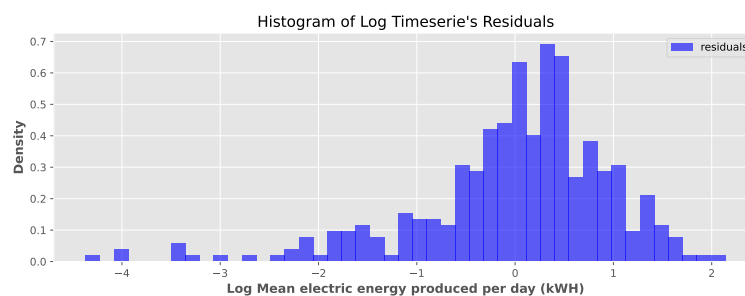
Στο διάγραμμα, που αποτυπώνεται η περιοδικότητα της χρονοσειράς, μπορούν να καταμετρηθούν 53 περίοδοι σε διάστημα ενός έτους. Τόσες είναι περίπου οι εβδομάδες ενός ολόκληρου έτους. Δεν θα ήταν αφελές να ισχυριστούμε ότι η περιοδικότητα που φαίνεται να υπάρχει είναι εβδομαδιαία. Είναι ένας ισχυρισμός ο οποίος δεν έρχεται σε αντιπαράθεση με την εκτίμηση που κάναμε μέσω της παρατήρησης που έγινε προηγουμένως όταν μελετήθηκε το διάγραμμα αυτοσυσχέτισης.

Στο τέταρτο και τελευταίο διάγραμμα παρακολουθείται η εξέλιξη των υπολειμμάτων των παρατηρήσεων της χρονοσειράς. Είναι το διάγραμμα απεικόνισης τους. Τόσο το διάγραμμα αυτό όσο και το ιστόγραμμα αυτών που παρουσιάζεται στο Σχήμα 7.36 μελετώνται προσεκτικά. Από την παρατήρηση τους προκύπτουν χρήσιμες πληροφορίες για την στασιμότητα τους και την κατανομή που ακολουθούν. Η μέση τιμή και η διασπορά τους δεν παραμένουν σταθερές παρά εμφανίζουν κατά διαστήματα αυξομειώσεις. Είναι εν γένη μη στάσιμα. Το ιστόγραμμα που παράχθηκε επιβεβαιώνει τον ισχυρισμό αυτό και παράλληλα υποδεικνύει ότι ακολουθούν μια κατανομή σκιασμένη κατά τα αριστερά. Δεν είναι κανονική και συνεπώς τα γραμμικά μοντέλα πρόβλεψης δεν δύναται ακόμη να λειτουργήσουν κατά τον κατάλληλο τρόπο επί αυτών των δεδομένων για να παράγουν έμπιστες προβλέψεις.

Το Σχήμα 7.37, αποτελεί το διάγραμμα αποδόμησης της χρονοσειράς που έχει σχηματιστεί με βάση τον μετασχηματισμό τετραγωνικής ρίζας. Η ανάγνωση των διαγραμμάτων τάσης και περιοδικότητας, υποδεικνύει την έντονη παρουσία και επιρροή τους στη διαμόρφωση των τιμών της μέσης ημερήσιας παραγόμενης ενέργειας. Αποδεικνύεται ότι είναι οι κύριοι λόγοι που η χρονοσειρά δεν είναι στάσιμη. Η περιοδικότητα υπολογίζεται πως είναι εβδομαδιαία



Σχήμα 7.35: Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε λογαριθμική κλίμακα.

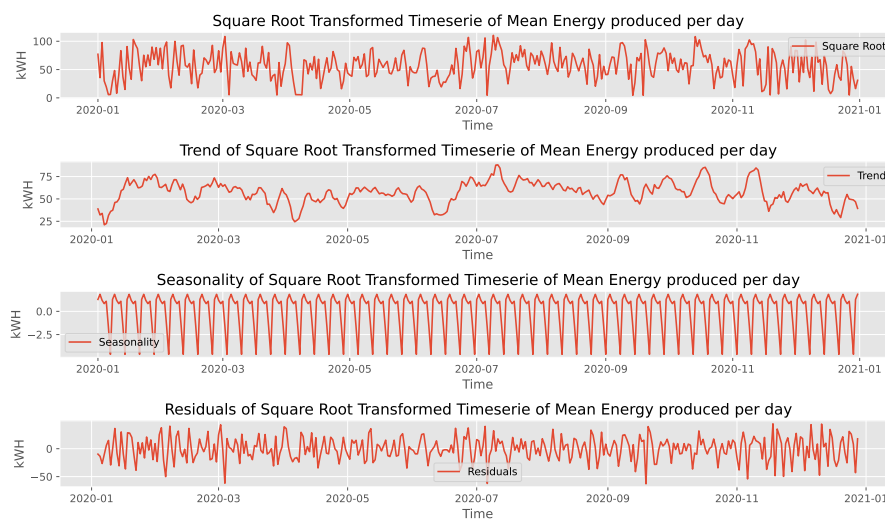


Σχήμα 7.36: Ιστογράμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη σε λογαριθμική κλίμακα.

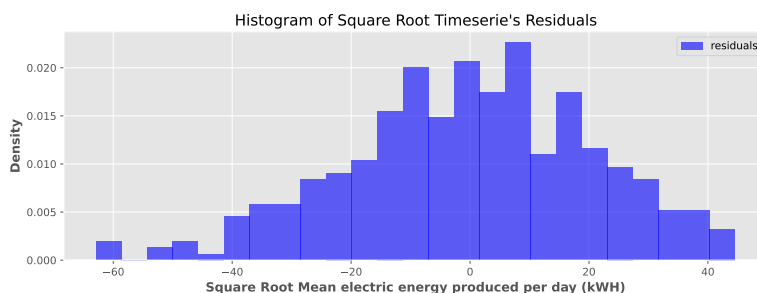
καθώς μπορούν να απαριθμηθούν 53 περίοδοι στη διάρκεια του διαστήματος που μελετάται η χρονοσειρά, δηλαδή, ενός έτους. Από την άλλη μεριά για την τάση δεν γίνεται να βγουν ασφαλή συμπεράσματα για τον τρόπο εξέλιξης της. Άλλοτε είναι ανοδική, άλλοτε σταθερή και άλλοτε πτωτική. Ο μετασχηματισμός αυτός δεν αντιμετωπίζει κατά κάποιο βαθμό την τάση ή την περιοδικότητα και σταθεροποιεί σε μικρότερα χρονικά διαστήματα την διακύμανση των τιμών.

Στο τελευταίο διάγραμμα του Σχήματος 7.37 παρακολουθείται το διάγραμμα απεικόνισης των υπολειμμάτων των παρατηρήσεων της χρονοσειράς. Από τη παράλληλη μελέτη αυτού του διαγράμματος και του ιστογράμματος αυτών που παρουσιάζεται στο Σχήμα 7.38, προκύπτουν χρήσιμα συμπεράσματα για την στασιμότητα τους και την κατανομή που ακολουθούν. Η μέση τιμή και η διασπορά τους δεν παραμένουν σταθερές παρά εμφανίζουν κατά διαστήματα αυξομειώσεις. Επίσης εμφανίζονται απότομες μεγάλες διακυμάνσεις, υποδεικνύοντας τη μη στασιμότητα εν γένει. Έπειτα το ιστογράμμα που παράχθηκε και μελετήθηκε, ισχυροποιεί την άποψη αυτή. Φαίνεται να ακολουθούν μια κατανομή η οποία προσεγγίζει την κανονική αλλά

έχει μεγαλύτερη συγκέντρωση τιμών σε κάποιες μεγάλες και σε κάποιες μικρές. Τέλος, δεν υπάρχει συμμετρία στην κατανομή τους γύρω από κάποιες κορυφές που εμφανίζονται στο μέσο του ιστογράμματος.



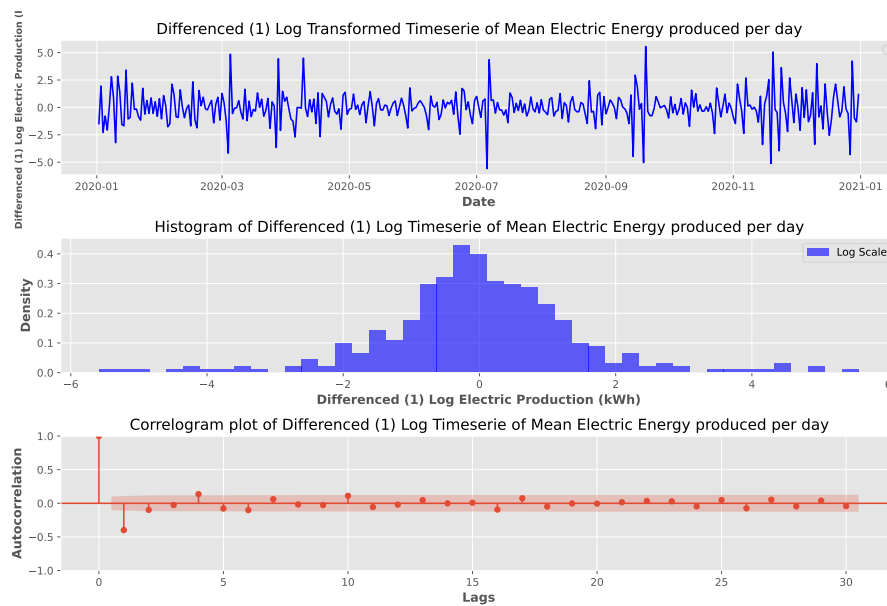
Σχήμα 7.37: Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιodikότητας και υπολοίπων σε κλίμακα τετραγωνικής ρίζας.



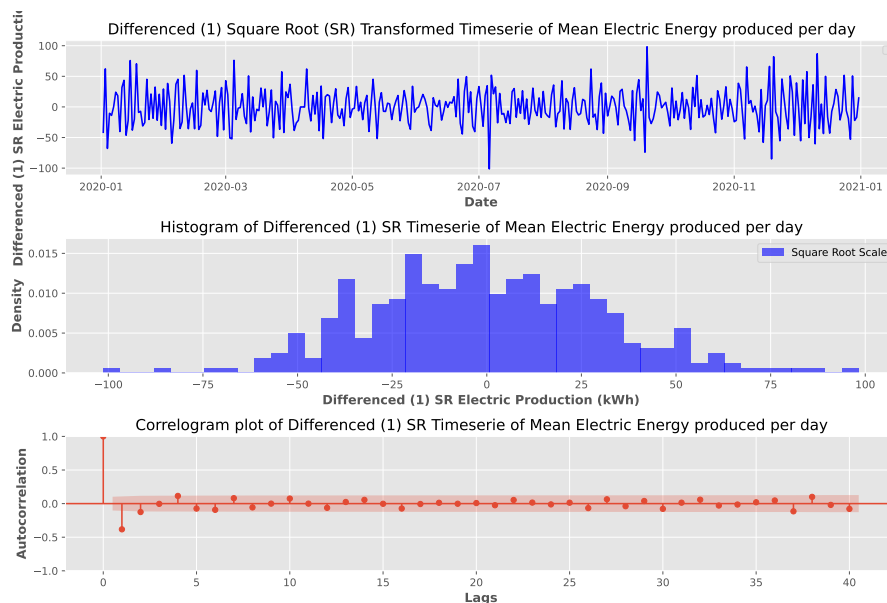
Σχήμα 7.38: Ιστογράμμο υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας Αιολικού πάρκου 2 στη Θράκη σε κλίμακα τετραγωνικής ρίζας.

Ο μετασχηματισμός των χρονοσειρών είτε στην κλίμακα τετραγωνικής ρίζας είτε στην λογαριθμική δεν επέφερε το επιθυμητό αποτέλεσμα. Προκειμένου να μετατραπούν σε πιο στάσιμες χρονοσειρές και στη συνέχεια να γίνει η επιλογή της καταλληλότερης χρονοσειράς για την παραγωγή προβλέψεων εφαρμόζεται μια μαθηματική μέθοδος. Τόσο στη μία όσο και στην άλλη περίπτωση εφαρμόζεται η μέθοδος των πρώτων διαφορών και γίνεται επαναξιολόγηση της κατάστασης. Γίνεται η εφαρμογή αυτής στη προσπάθεια αντιμετώπισης της τάσης και δημιουργούνται τα απαραίτητα διαγράμματα για την άντληση πληροφοριών επί της συμπεριφοράς της χρονοσειράς στο χρόνο.

Στο Σχήμα 7.39 παρουσιάζονται το διάγραμμα απεικόνισης, το ιστογράμμο και το διάγραμμα αυτοσυσχέτισης της χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε λογαριθμική κλίμακα, ενώ στο Σχήμα 7.40 παρουσιάζονται τα αντίστοιχα διαγράμματα της χρονοσειράς που



Σχήμα 7.39: Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα λογαριθμική αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.



Σχήμα 7.40: Από πάνω προς τα κάτω: διάγραμμα απεικόνισης, ιστόγραμμα και διάγραμμα αυτοσυσχέτισης χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.

δημιουργούν τα δεδομένα που έχουν μετασχηματιστεί χάριν στο μετασχηματισμό τετραγωνικής ρίζας. Υπενθυμίζεται ότι τόσο στην μια όσο και στην άλλη περίπτωση έχει εφαρμοστεί η μέθοδος των πρώτων διαφορών πριν παραχθούν.

Η ανάγνωση των Σχημάτων 7.39, 7.40 γίνεται προσεκτικά και παράλληλα ταυτόχρονα.

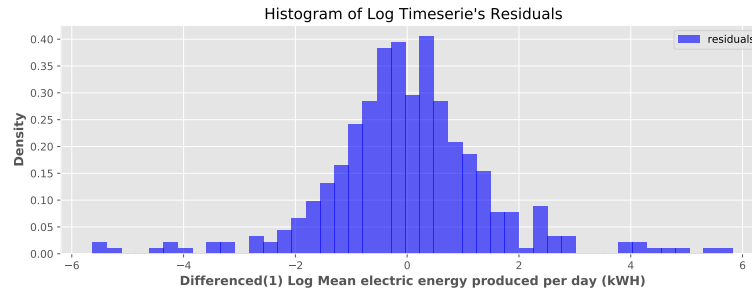
Η ανάλυση βρίσκεται στο σημείο στο οποίο θα πρέπει να γίνει η επιλογή της χρονοσειρά με την οποία στη συνέχεια θα προχωρήσουμε στη παραγωγή των απαιτούμενων προβλέψεων. Η παρατήρηση τους ξεκινάει με τα διαγράμματα απεικόνισης τα οποία βρίσκονται πάνω πάνω στα Σχήματα.

Στη περίπτωση της λογαριθμικής χρονοσειράς φαίνεται ότι η εφαρμογή της μεθόδου εξομάλυνε ακόμα περισσότερο την χρονοσειρά. Αν και δεν εξαλείφθηκαν απόλυτα οι απότομες διακυμάνσεις στις τιμές της χρονοσειράς, το πλήθος αυτών έχει μειωθεί σημαντικά. Πλέον μόνο ελάχιστα μικρά διαστήματα παρουσιάζουν απρόβλεπτες διακυμάνσεις. Είναι ένα σημάδι, αντιμετώπισης της τάσης. Από την άλλη μεριά, με μια πρώτη ματιά στο διάγραμμα απεικόνισης στη χρονοσειρά τετραγωνικής ρίζας παρατηρείται ότι η διαφορίση κατά μια τάξη της χρονοσειράς σταθεροποιεί την μέση τιμή της χρονοσειράς σημαντικά. Παρόλα αυτά φαίνεται ακόμα πως οι απότομες διακυμάνσεις επηρεάζουν πιο έντονα την χρονοσειρά αυτή παρά την λογαριθμικά μετασχηματισμένη. Η συνιστώσα που επηρεάζει πλέον πιο πολύ την χρονοσειρά είναι αυτή της περιοδικότητας και στις δύο περιπτώσεις. Ειδικά στη λογαριθμικά μετασχηματισμένη χρονοσειρά είναι πιο εύκολο να παρατηρηθεί ότι η εξέλιξη των τιμών δείχνει να προκαλείται σύμφωνα με αυτήν.

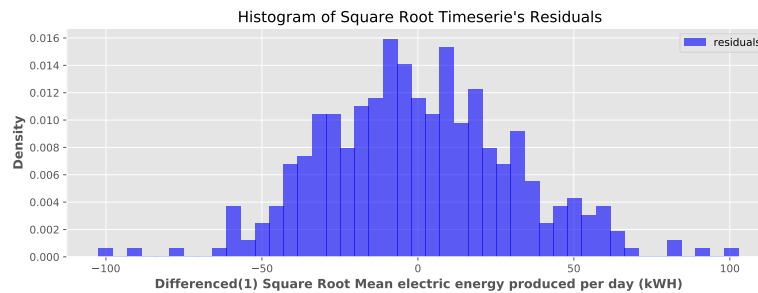
Τα ιστογράμματα που ακολουθούν στην μέση των Σχημάτων δίδουν χρήσιμες πληροφορίες για την κατανομή πυκνότητας πιθανότητας των χρονοσειρών. Τόσο στην μια όσο και στην άλλη περίπτωση, αποδεικνύεται ότι η εφαρμογή της μεθόδου πρώτων διαφορών ήταν ένα σωστό βήμα προς της μετατροπή τους σε στάσιμες. Οι ουρές έχουν εξαλειφθεί τόσο στην μια όσο και στην άλλη περίπτωση. Εμφανίζουν επιτέλους μια κορυφή γύρω από την οποία κατανέμονται οι πιθανότητες εμφάνισης μιας τιμής όπως υποδεικνύει μια κανονική κατανομή. Είναι δυο κατανομές οι οποίες θα μπορούσαν να θεωρηθούν ότι προσεγγίζουν την κανονική.

Το ιστόγραμμα της λογαριθμικά μετασχηματισμένης χρονοσειράς φαίνεται να είναι αυτό το οποίο φέρει τη καλύτερη προσέγγιση. Η κατανομή της προσεγγίζεται από την κανονική κατανομή πολύ καλύτερα σε σχέση πάντα με την περίπτωση της χρονοσειράς τετραγωνικής ρίζας. Είναι ένα προοίμιο ότι τα υπόλοιπα της χρονοσειράς αυτής θα ακολουθούν μια παρόμοια κατανομή, η οποία θα φέρει σε πολύ μεγάλο βαθμό την κανονική. Για το λόγο αυτό παράγουμε τα Σχήματα 7.41 και 7.42, που παρουσιάζουν τα ιστογράμματα των υπολοίπων των χρονοσειρών που μελετώνται. Το πρώτο προέρχεται από την λογαριθμικά μετασχηματισμένη χρονοσειρά και το δεύτερο από την χρονοσειρά που βρίσκεται σε κλίμακα τετραγωνικής ρίζας.

Είναι ξεκάθαρο ότι το ιστόγραμμα των υπολοίπων που φαίνονται στο Σχήμα 7.41 προσεγγίζει πολύ καλύτερα την κανονική κατανομή. Τα υπόλοιπα της λογαριθμικά μετασχηματισμένης χρονοσειράς έχουν κορυφή και μέση τιμή κοντά στο μηδέν ενώ συμμετρικά γύρω από αυτήν απλώνεται οι υπόλοιπες τιμές. Στη περίπτωση της χρονοσειράς που μετασχηματίστηκε με βάση τον μετασχηματισμό τετραγωνικής ρίζας από το Σχήμα 7.42 γίνεται αντιληπτό ότι δεν υπάρχει εξίσου καλή συμμετρία γύρω από την κορυφή του ιστογράμματος και τη μέση τιμή αυτών. Ακόμη συνεχίζουν να υπάρχουν κάποιες κορυφές γύρω από αυτήν όπως και συνεχίζει να υφίσταται μια μικρή δεξιά σκίαση. Είναι ισχυρές αποδείξεις ότι τα γραμμικά μοντέλα πρόβλεψης θα λειτουργήσουν καλύτερα με τα δεδομένα τα οποία έχουν μετατραπεί σε λογαριθμικά και έπειτα διαφοριστεί.



Σχήμα 7.41: Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στη λογαριθμική κλίμακα αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.

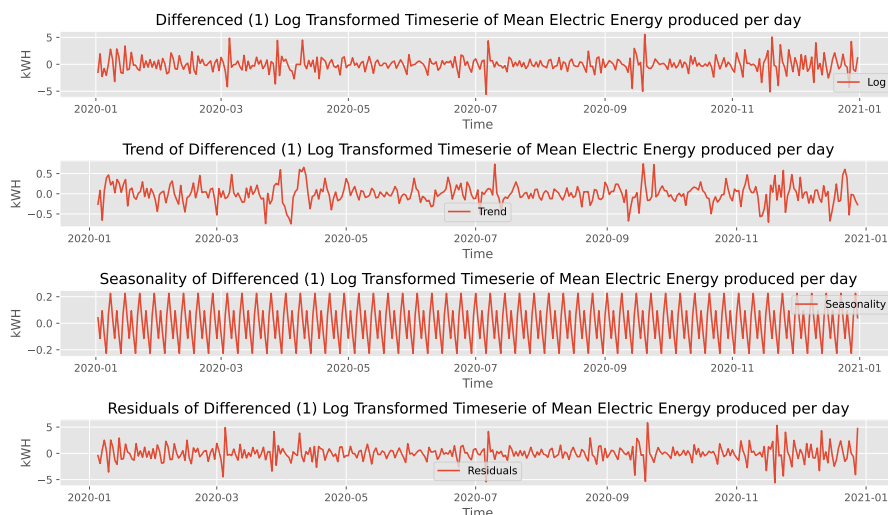


Σχήμα 7.42: Ιστόγραμμα υπολοίπων χρονοσειράς παραγωγής ηλεκτρικής ενέργειας σε κλίμακα τετραγωνικής ρίζας αφού εφαρμόστηκε η μέθοδος των πρώτων διαφορών.

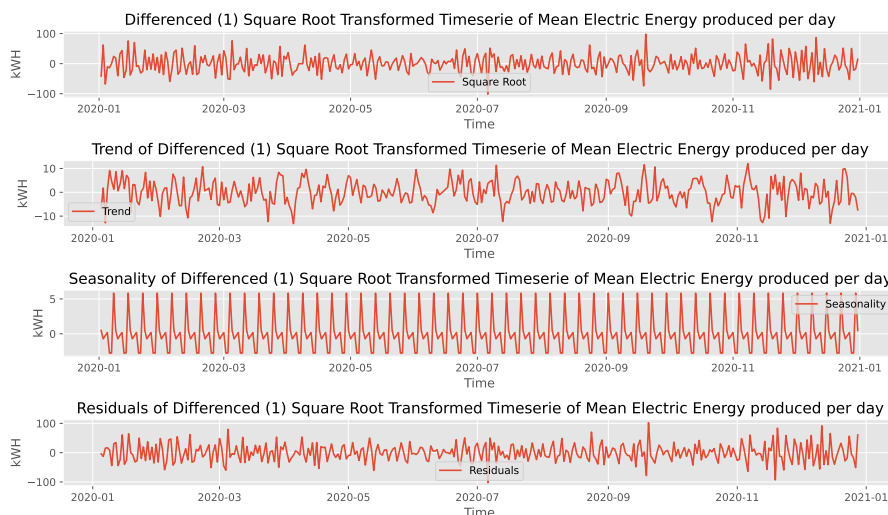
Τελευταία διαγράμματα προς ανάγνωση και σύγκριση στα Σχήματα 7.39 και 7.40 είναι αυτά της αυτοσυσχέτισης και βρίσκονται στο κατώτατη θέση αυτού. Είναι διαγράμματα τα οποία περιέχουν πληροφορίες χρήσιμες για την υπόδειξη ή όχι στασιμότητας στη χρονοσειρά. Και στις δύο περιπτώσεις παρατηρείται εκθετική πτώση των τιμών της αυτοσυσχέτισης εντός των διαστημάτων εμπιστοσύνης μετά τις πρώτες δύο υστερήσεις. Ωστόσο μετά από αυτές υπάρχουν κάποιες οι οποίες μεμονωμένα εμπίπτουν εκτός των διαστημάτων αυτών. Ίσως να μπορούν να προσφέρουν πληροφορία ώστε να εκφραστεί καλύτερα μια παρατήρηση με βάση τις παρελθοντικές τιμές της σε σχέση με την αρχική χρονοσειρά που αυτό δεν ήταν εφικτό.

Όσον αφορά τη περιοδικότητα της χρονοσειράς που δεν έχει αντιμετωπιστεί ακόμα, φαίνεται να μπορεί να οριστεί στις 6 περίπου χρονικές υστερήσεις. Στις πρώτες σημαντικές υστερήσεις είναι πιο ορατή, ενώ όσο προχωράμε σε μεταγενέστερες αυτή φαίνεται να μην καθίσταται δυνατή να εντοπιστεί. Προκειμένου να μελετηθούν ακόμα παραπάνω οι συνιστώσες της χρονοσειράς παράγονται τα διαγράμματα αποδόμησης σε κάθε περίπτωση. Αυτά παρουσιάζονται από τα Σχήματα 7.43 και 7.44.

Το διάγραμμα αποδόμησης της χρονοσειράς που οι τιμές της βρίσκονται σε λογαριθμική κλίμακα και έχει διαφοριστεί κατά μια τάξη αποτυπώνεται από το Σχήμα 7.43, ενώ αυτό που αφορά την χρονοσειρά τετραγωνικής ρίζας βρίσκεται στο Σχήμα 7.44. Από το διάγραμμα της περιοδικότητας διακρίνουμε τόσο στην μια όσο και στην άλλη περίπτωση ότι στο διάστημα που μελετάμε (ενός έτος) ολοκληρώνονται 53 περίοδοι. Εύλογα συμπεραίνουμε ότι υπάρχει περίπου εβδομαδιαία περιοδικότητα ακόμα στις χρονοσειρές. Είναι μια παρατήρησή που αναμέναμε



Σχήμα 7.43: Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε λογαριθμική κλίμακα.



Σχήμα 7.44: Αποδόμηση χρονοσειράς παραγωγής ηλεκτρικής ενέργειας στις συνιστώσες τάσης, περιοδικότητας και υπολοίπων σε κλίμακα τετραγωνικής ρίζας.

καθώς η μέθοδος των πρώτων διαφορών δεν επιλύει το πρόβλημα αυτό.

Η εφαρμογή της μεθόδου πρώτων διαφορών αντιμετωπίζει κυρίως την συνιστώσα της τάσης. Στη περίπτωση της λογαριθμικής χρονοσειράς φαίνεται ότι αντιμετωπίστηκε σε σημαντικό βαθμό. Υπάρχουν διαστήματα που επηρεάζονταν οι τιμές είτε λόγω αυξητικής τάσης είτε λόγω μειωτικής τάσης αλλά πλέον οφείλονται κυρίως στη συνιστώσα της περιοδικότητας. Φαίνεται ότι η τάση όπου δεν εξαλείφθηκε, παρουσιάζεται να επηρεάζει τη χρονοσειρά κατά το τρόπο που το κάνει η συνιστώσα της περιοδικότητας. Για αυτό το λόγο άλλωστε φαίνεται η λογαριθμημένη και ολοκληρωμένη χρονοσειρά να βρίσκεται πιο κοντά από ποτέ στη στασιμότητα.

Στη περίπτωση της χρονοσειράς που βρισκόταν σε κλίμακα τετραγωνικής ρίζας και εφαρμόστηκε η μέθοδος των διαφορών πρώτης τάξης μέσω του διαγράμματος αποδόμησης της χρονοσειράς είναι εύκολο να διαπιστωθεί ότι η μέση τιμή της χρονοσειράς έχει σταθεροποιηθεί περισσότερο αλλά ακόμη υφίσταται. Και σε αυτήν την περίπτωση η τάση εμφανίζεται να επηρεάζει τις τιμές της χρονοσειράς με το τρόπο που το κάνει η περιοδικότητα.

Τέλος τόσο για τη μια όσο και για την άλλη περίπτωση παρατηρήθηκαν τα διαγράμματα απεικόνισης των υπολειμμάτων των χρονοσειρών. Τα υπόλοιπα της ολοκληρωμένης λογαριθμικής χρονοσειράς φαίνεται να είναι πιο στάσιμα από της χρονοσειράς που βρίσκεται στη κλίμακα τετραγωνικής ρίζας. Τα κύρια χαρακτηριστικά τους, μέση τιμή και διασπορά, διατηρούνται εν γένη πιο σταθερά στο χρόνο. Δεν υπάρχουν ιδιαίτερα σημαντικά απρόβλεπτες και μεγάλες διακυμάνσεις στις τιμές τους. Όσες υπάρχουν είναι θεωρητικά μετρήσιμες. Εάν συνυπολογίσουμε την φύση του ανέμου, του μέσου που αξιοποιείται για την παραγωγή ηλεκτρικής ενέργειας, που δεν γίνεται να μοντελοποιηθεί από απολύτως γραμμικές σχέσεις, τότε μπορεί να θεωρηθεί ικανοποιητικός ο βαθμός στασιμότητας της χρονοσειράς που εκφράζει την ημερήσια μέση ποσότητα παραγόμενης ηλεκτρικής ενέργειας από ένα αιολικό πάρκο.

Με βάση τις πληροφορίες που εξήχθησαν από τη διαδικασία μετατροπής της αρχικής χρονοσειράς, ο λογαριθμικός μετασχηματισμός ακολουθούμενος από την εφαρμογή της μεθόδου πρώτων διαφορών προσφέρει μια χρονοσειρά η οποία αν και δεν είναι αυστηρά στάσιμη βρίσκεται πιο κοντά στη κατάσταση στασιμότητας που σε σχέση με την αρχική χρονοσειρά. Επιλέχθηκε αυτή η χρονοσειρά για την εκπαίδευση των γραμμικών μοντέλων και την παράγωγή των προβλέψεων επειδή τα υπόλοιπα της παρατηρήθηκε ότι ήταν πιο στάσιμα αφού είχε σταθεροποιηθεί σε μεγαλύτερα χρονικά διαστήματα η μέση τιμή και διασπορά της χρονοσειράς.

7.3.4 Παραγωγή Προβλέψεων με Γραμμικά Μοντέλα

Για να δημιουργηθούν οι απαιτούμενες προβλέψεις, έγινε χρήση των γραμμικών μοντέλων AR, MA, ARMA, ARIMA και SARIMA. Στη συνέχεια μελετήθηκαν τα αποτελέσματα που προέκυψαν μετά την εφαρμογή τους και αξιολογήθηκε η απόδοσή τους. Σε κάθε περίπτωση, η επιλογή του κατάλληλου μοντέλου βασίστηκε στο κριτήριο AIC, ενώ η αξιολόγηση τους στο κριτήριο της ρίζας του μέσου τετραγωνικού σφάλματος RMSE όπως και άλλων στατιστικών μέτρων.

Πριν την παραγωγή της πρόβλεψης, πρέπει να αποσαφηνιστεί ο σκοπός και ο λόγος πραγματοποίησης της. Στην ουσία, με αυτόν τον τρόπο καθορίζεται ο χρονικός ορίζοντας στον οποίο θα προβλεφθούν οι τιμές της μεταβλητής ενδιαφέροντος. Στη περίπτωση των δεδομένων παραγωγής ηλεκτρικής ενέργειας από το Αιολικό Πάρκο 2 στη Θράκη, εξετάζονται ημερήσια δεδομένα μέσης παραγωγής τα οποία μπορούν να χρησιμοποιηθούν προκειμένου να παραχθούν προβλέψεις επί της παραγωγής για τις επόμενες μέρες. Πιο συγκεκριμένα, η πρόβλεψη που γίνεται, αφορά τις τιμές της μέσης παραγωγής για τις επόμενες δέκα μέρες μετά το τελευταίο εξεταζόμενο δείγμα. Με αυτόν τον τρόπο μπορεί η εταιρία που παράγει την ενέργεια να πληροφορηθεί για τις επικείμενες ποσότητες που πρόκειται να έχει διαθέσιμες προς πώληση στο Χρηματιστήριο Ενέργειας. Συνεπώς, με βάση αυτές μπορεί να γνωρίζει τι μέρος της

ζήτησης μπορεί να καλύψει στην αγορά, να μεγιστοποιήσει το κέρδος της, να αποφασίσει εάν θα διατελέσει ή όχι εργασίες συντήρησης ή αναβάθμισης και να προγραμματίσει το δίκτυο σε περίπτωση που αδυνατούν οι μονάδες παραγωγής αιολικής ενέργειας να προσφέρουν το απαιτούμενο φορτίο προς κατανάλωση άλλες εφεδρικές πηγές ενέργειας να τεθούν προς λειτουργία ώστε να μην υπάρξουν προβλήματα ηλεκτροδότησης.

Οι αλγόριθμοι, οι οποίοι υλοποιούσαν τους υπολογισμούς που ορίζει το κάθε μοντέλο, παρέχονταν από τη βιβλιοθήκη `Statsmodel` της `python`. Σε αυτή μέσα υπάρχουν οι κλάσεις `ARIMA` και `SARIMA` οι οποίες παρέχουν κατάλληλες συναρτήσεις για την παραγωγή των προβλέψεων μας. Αρχικά τα δεδομένα παραγωγής διαχωρίστηκαν σε δεδομένα εκπαίδευσης (`train dataset`) και δεδομένα ελέγχου (`test dataset`). Οι πρώτες 355 παρατηρήσεις της χρονοσειράς αποτελούν τα δεδομένα εκπαίδευσης. Είναι το λεγόμενο (`train dataset`). Χρησιμοποιούνται, για να εκπαιδευτεί το εκάστοτε γραμμικό μοντέλο πρόβλεψης. Οι 10 παρατηρήσεις που απομένουν, αποτελούν τα δεδομένα ελέγχου (`test dataset`). Ουσιαστικά, αφού εκπαιδευτεί το μοντέλο, θα γίνει μια πρόβλεψη εκτός δείγματος h βημάτων μπροστά, όπου στη συγκεκριμένη περίπτωση $h = 10$ (`out of sample prediction`). Θα χρησιμοποιηθούν, ώστε να υπολογισθεί η απόκλιση των τιμών που προβλέφθηκαν από των πραγματικών.

Η τιμή της ποσότητας ενέργειας που παράχθηκε και προβλέπεται για κάθε μία από αυτές τις 10 ημέρες συγκρίνεται με την πραγματική όπως αυτή είχε καταγραφεί στη πραγματικότητα και παρέχεται από τα δεδομένα ελέγχου (`test dataset`). Το στατιστικό μέτρο της ρίζας του μέσου τετραγωνικού σφάλματος `RMSE` θα χρησιμοποιηθεί για να γίνει μια πρώτη σύγκριση των μοντέλων. Ωστόσο, μόλις βρεθεί το βέλτιστο μοντέλο υπολογίζονται επιπλέον στατιστικά μέτρα. Αυτά είναι ο συντελεστής προσδιορισμού (R^2), ο συντελεστής συσχέτισης (`corr`), το μέσο απόλυτο σφάλμα (`MAE`), το μέσο απόλυτο ποσοστιαίο σφάλμα (`MAPE`), το μέσο σφάλμα (`ME`) και το μέσο ποσοστιαίο σφάλμα (`MPE`). Τέλος, δεν παραλείπεται να δημιουργηθούν τα διαγράμματα των προβλέψεων ώστε να γίνει επιπλέον οπτική σύγκριση ανάμεσα στις πραγματικές τιμές και σε αυτές που προβλέφθηκαν από τα γραμμικά μοντέλα. Σε κάθε σχήμα παρέχεται ένα γράφημα απεικόνισης, ένα ραβδογράφημα καθώς επίσης και το ιστόγραμμα υπολοίπων του κάθε μοντέλου που εκπαιδεύτηκε. Με την αξιοποίηση τους δύναται να γίνει μια πρώτη εκτίμηση της προβλεπτικής ικανότητας του κάθε μοντέλου.

Με την χρήση του ιστογράμματος υπολοίπων αποκτούμε μια πρώτη άποψη γενικά ως προς την συμπεριφορά και την απόδοση που πρόκειται να έχει το μοντέλο πρόβλεψης δίχως να πρέπει να βασιστούμε σε στατιστικά μέτρα. Με μια προσεκτική ανάγνωση του γίνεται έλεγχος για την ύπαρξη τυχόν ακραίων τιμών ή προσαρμογής σε διαδικασία λευκού θορύβου και να δικαιολογηθεί η επάρκεια ή ανεπάρκεια του μοντέλου να παράξει μια εύστοχη και έμπιστη πρόβλεψη. Σε περίπτωση που εμφανίζονται σημάδια αποσταθεροποίησης της διακύμανσης ή μέσης τιμής υπάρχει ανάγκη βελτίωσης του μοντέλου που έχει εκπαιδευτεί. Τέλος, μπορεί να χρησιμοποιηθεί ως μέτρο σύγκρισης με άλλα γραμμικά μοντέλα καθώς αυτό που φέρεται να βρίσκεται πιο κοντά σε ένα ιστόγραμμα λευκού θορύβου αποτελεί τη καλύτερη επιλογή για τη παραγωγή της πρόβλεψης μας. Το επιθυμητό δηλαδή είναι η διαφορά ανάμεσα στην πραγματική και τη προβλεπόμενη τιμή, να είναι απλά τυχαίος θόρυβος, δίδοντας με αυτόν τον τρόπο στα υπόλοιπα έναν χαρακτήρα ανεξαρτησίας και μη συσχέτισης με τον χρόνο.

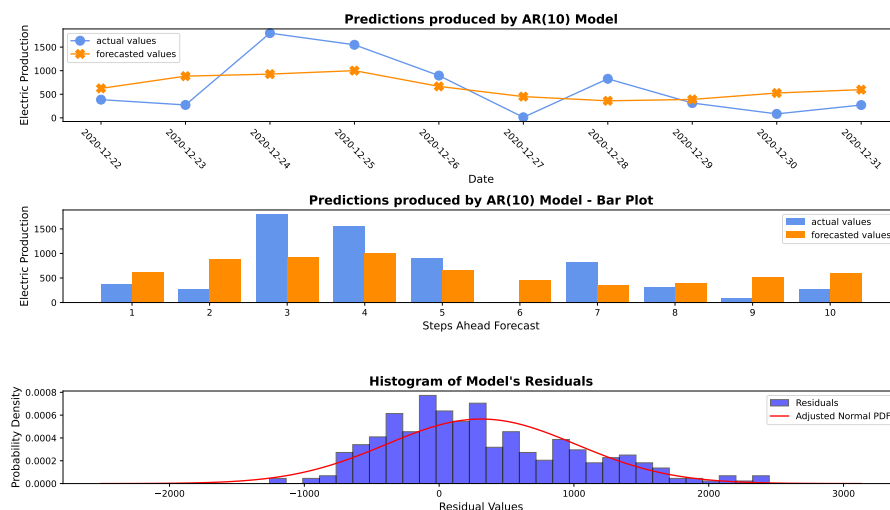
Για την παραγωγή της πρόβλεψης, επιλέχθηκε να δοκιμαστούν μοντέλα $AR(p)$, $MA(q)$, $ARMA(p,q)$, $ARIMA(p,d,q)$ και $SARIMA(p,d,q),(P,D,Q)_s$. Στα Σχήματα 7.45 – 7.49 παραθέτονται οι προβλέψεις που απέδωσε το καλύτερο μοντέλο από το κάθε είδος. Η βέλτιστη δυνατή τάξη των παραγόντων (p,q) του κάθε μοντέλου βρέθηκε χωρίς την βοήθεια των διαγραμμάτων ACF και PACF παρά αναπτύχθηκε ένας αλγόριθμος εξαντλητικής αναζήτησης στο σύνολο τιμών $[0,10]$ τόσο για τον έναν παράγοντα (p) όσο και για τον άλλο παράγοντα (q) . Για τους εποχικούς συντελεστές (P,Q) ορίστηκε το διάστημα αναζήτησης $[0,2]$, ενώ οι παράγοντες διαφορίσης d,D τέθηκαν ίσοι με το 1 στις περιπτώσεις των ARIMA, SARIMA μοντέλων προκειμένου να αντιμετωπιστεί η τάση ($d = 1$) και η εποχικότητα ($D = 1$) με σκοπό την απαλοιφή των συνιστωσών τάσης και εποχικότητας που καθιστούσαν την εξεταζόμενη χρονοσειρά μη στάσιμη.

Η εφαρμογή λογαριθμικού μετασχηματισμού στα δεδομένα είχε ως αποτέλεσμα την σταθεροποίηση της διασποράς. Με την ολοκλήρωση της εκτέλεσης της διαδικασίας αυτής είχαν εκπαιδευτεί όλα τα δυνατά μοντέλα βάση των δεδομένων εκπαίδευσης. Το μοντέλο που όμως επιλεγόταν τελικά να χρησιμοποιηθεί, ώστε να γίνει η πρόβλεψη, ήταν αυτό που απέδιδε τη μικρότερη τιμή στο κριτήριο AIC κατά την εκπαίδευση του. Όποιο μοντέλο προσαρμοζόταν καλύτερα (best fit), έδινε δηλαδή, την καλύτερη προσαρμογή στα δεδομένα εκπαίδευσης, χρησιμοποιόταν έπειτα για την πρόβλεψη εκτός δείγματος (ουτ-οφ-σαμπλε πρεδιστιον). Τα λογαριθμημένα δεδομένα επέφεραν πιο ακριβείς προβλέψεις έναντι αυτών που είχαν μετασχηματιστεί σε κλίμακα αντίστροφης τετραγωνικής ρίζας, οπότε και επιλέχθηκαν να παρουσιαστούν, όντας τα βέλτιστα, καθώς όπως αποδείχθηκε σε πρότερα στάδια επέφεραν μια χρονοσειρά η οποία μετά τη διαφορίσης πλησίαζε περισσότερο τη στασιμότητα.

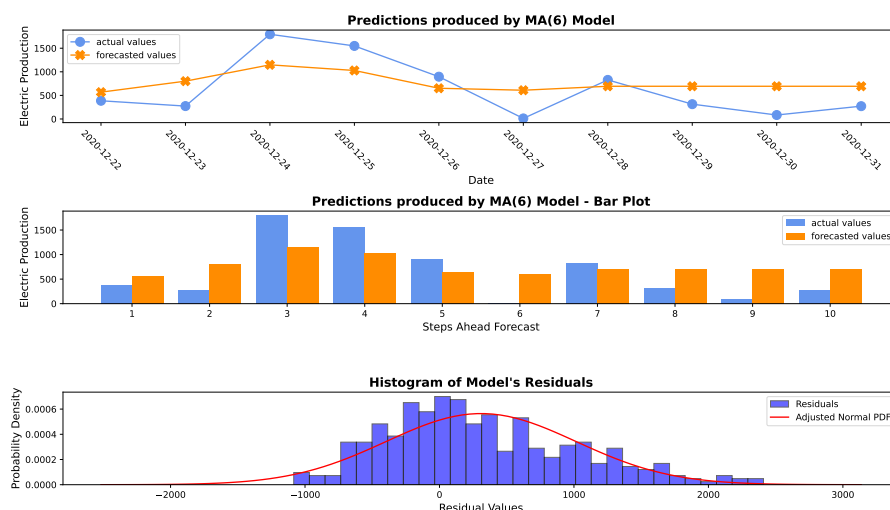
Οι προβλέψεις που αποτυπώνονται στα Σχήματα 7.45, 7.46, 7.47, παράχθηκαν αφού εκπαιδεύτηκαν τα μοντέλα AR, MA και ARMA αντίστοιχα. Η κλίμακα στην οποία θα παρουσιάζονταν οι προβλέψεις θα ήταν η ίδια με αυτήν των δεδομένων που χρησιμοποιήθηκαν, δηλαδή, της λογαριθμικής, όμως πριν παρουσιαστούν επιλέχθηκε να επαναφερθούν στην αρχική κλίμακα.

Τα μοντέλα αυτοπαλινδρόμησης AR και κινούμενου μέσου MA ήταν τα πιο απλά μοντέλα που χρησιμοποιήθηκαν καθώς αξιοποίησαν διαδικασίες αυτοπαλινδρόμησης και κινούμενου μέσου αντίστοιχα. Αντίθετα ο συνδυασμός των δύο αυτών διαδικασιών έφερε ένα πιο σύνθετο μοντέλο το οποίο στη θεωρία δύναται να προσαρμοστεί καλύτερα στις σχέσεις που αναπτύσσονται μεταξύ των τιμών της μεταβλητής ενδιαφέροντος. Είναι σημαντικό να αναφερθεί ότι επειδή δεν έχει γίνει διαφορίση κάποιας τάξης η τάση δεν έχει απαλειφθεί. Παρόλα αυτά επειδή η λογαριθμική μετατροπή σταθεροποίησε εν μέρη τη διακύμανση των τιμών της μεταβλητής ενδιαφέροντος επιλέγεται να εφαρμοστούν κανονικά τα γραμμικά μοντέλα στα συγκεκριμένα δεδομένα. Πλέον οι μεταβολές που υφίστανται στη τιμή της μέσης τιμής και της διασποράς στη πάροδο του χρόνου είναι πιο μικρές. Σίγουρα δεν αναμένεται να εμφανιστεί το επιθυμητό αποτέλεσμα καθώς δεν προσεγγίζεται απόλυτα η στασιμότητα και η κανονική κατανομή από τα υπόλοιπα της χρονοσειράς.

Με την παρατήρηση των τριών σχημάτων φαίνεται ότι το μοντέλο ARMA δεν δίνει καλύτερη πρόβλεψη έναντι των AR και MA μοντέλων όπως αναμενόταν. Μπορεί να συνδυάζεται τόσο διαδικασίες αυτοπαλινδρόμησης όσο και κινούμενου μέσου αλλά δεν καταφέρνει να φέρει

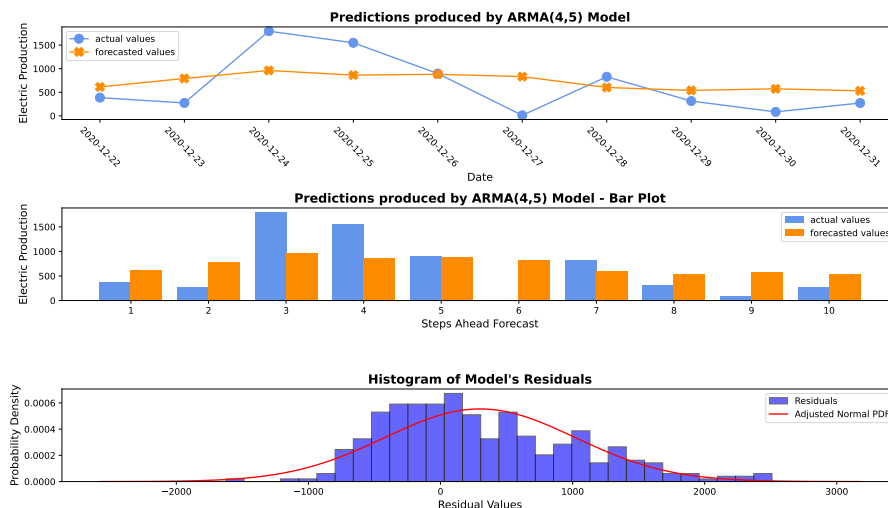


Σχήμα 7.45: Πρόβλεψη με χρήση του γραμμικού μοντέλου AR της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες.



Σχήμα 7.46: Πρόβλεψη με χρήση του γραμμικού μοντέλου MA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες.

ένα καλύτερο αποτέλεσμα. Αν και αποδίδει τη μικρότερη τιμή στο κριτήριο AIC σε σχέση με όλα τα υπόλοιπα, δίνει τη δεύτερη χειρότερη πρόβλεψη με βάση το κριτήριο RMSE το οποίο υπολογίζει μια απόκλιση ίση με 506.45 kWh. Δε μπορούμε να θεωρήσουμε ούτε κακή αλλά ούτε και καλή την προβλεπτική του ικανότητα. Σε κάποιες περιπτώσεις πραγματοποιεί εκτιμήσεις με μικρές αποκλίσεις (για $h = 1, 5, 7, 8, 10$) ενώ στις υπόλοιπες παρατηρείται ότι αποκλίνει σημαντικά των πραγματικών. Σε γενικές γραμμές επίσης φαίνεται να μπορεί να ακολουθήσει σε μέγεθος και κατεύθυνση τις πραγματικές τιμές στο πέρασμα του χρόνου, υπάρχει όμως χώρος για βελτίωση. Ακόμη στο μοντέλο αυτό φαίνεται να έχει γίνει μια σωστή επιλογή τάξεων p, q καθώς είναι μοιρασμένες οι φορές που υπερεκτίμησε ή υποεκτίμησε σε σχέση με τις πραγματικές τιμές.



Σχήμα 7.47: Πρόβλεψη με χρήση του γραμμικού μοντέλου ARMA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες.

Στο Σχήμα 7.47 πολύ χρήσιμες πληροφορίες παρέχει το τελευταίο διάγραμμα, δηλαδή, το ιστόγραμμα των υπολοίπων του μοντέλου εκπαιδεύτηκε και παρήγαγε τη πρόβλεψη που φαίνεται στα από πάνω διαγράμματα. Με μια ανάγνωση του, συμπεραίνεται ότι δεν προσομοιώνει μια κανονική κατανομή και ιδανικά αυτή του λευκού θορύβου. Οι ενδείξεις που οδήγησαν στο συμπέρασμα αυτό είναι κυρίως η ύπαρξη μιας μακριάς ουράς η οποία εκτείνεται στο δεξιό μέρος του ιστογράμματος. Υπάρχει δηλαδή δεξιά σκίαση. Φανερώνει ότι το μοντέλο αυτό έχει την τάση να κάνει υποεκτίμηση των πραγματικών τιμών και μάλιστα σε μεγαλύτερη κλίμακα από όταν υπερεκτιμά. Η παρατήρηση αυτή ευσταθεί εάν παρατηρηθεί η ασυμμετρία στην κατανομή των τιμών του ιστογράμματος γύρω από την κορυφή του ιστογράμματος που αυτή βρίσκεται σε τιμές κοντά στο μηδέν. Στις περιπτώσεις υπερεκτίμησης που τα υπόλοιπα έλαβαν τιμές έως 2500 kWh ενώ σε κάθε περίπτωση που υποτίμησε το λάθος δεν ανήλθε ποτέ το μέγεθος των 1800 kWh. Θετική ένδειξη για την απόδοση του μοντέλου αποτελεί η υψηλή συγκέντρωση τιμών γύρω από το μηδέν που υποδηλώνει ότι κατά τη διάρκεια της εκπαίδευσης υπήρχαν πολλές φορές που προσαρμόστηκε με μεγάλη επιτυχία.

Στον αντίποδα τα μοντέλα MA και AR έδωσαν μια καλύτερη πρόβλεψη αν και στη θεωρία αποτελούσαν μια αμφιβόλου ποιότητας και εμπιστοσύνης επιλογή, καθώς ήταν τα πιο απλά μοντέλα και καλούνταν να αξιοποιήσουν δεδομένα μη στάσιμης χρονοσειράς για την εκπαίδευση τους. Το AR μοντέλο με μέσο τετραγωνικό σφάλμα ρίζας τις 474,78 kWh έρχεται δεύτερο στη συνολική κατάταξη με καλύτερο όλων το MA με μέσο τετραγωνικό σφάλμα ρίζας τις 461 kWh περίπου. Τα δύο αυτά μοντέλα κατάφεραν με την απλότητα τους να φέρουν παραπλήσια καλά αποτελέσματα υπενθυμίζοντας μας ότι κάποιες φορές η πιο απλή λύση στο πρόβλημα μας ίσως να είναι καλύτερη από μια πιο σύνθετη.

Στα δύο μοντέλα φαίνεται ότι η εξαντλητική αναζήτηση των βέλτιστων τάξεων για τις διαδικασίες αυτοπαλινδρόμησης και κινούμενου μέσου έφερε καλά αποτελέσματα αφού ακολουθούνται στις περισσότερες περιπτώσεις οι αλλαγές των πραγματικών τιμών σε μέγεθος

και κατεύθυνση. Ειδικά το μοντέλο MA δείχνει μια ισχυρή ικανότητα στην προσπάθεια αυτή. Μπορεί να μην προβλέπει ακριβώς τις τιμές που θα έπρεπε αλλά έχει αποτυπώσει σχεδόν άριστα την κατεύθυνση των αλλαγών στις τιμές της μέση παραγόμενης ποσότητας ηλεκτρικής ενέργειας από ημέρα σε ημέρα. Φαίνεται ότι προβλέπει τόσο τις περιπτώσεις που υπάρχει αύξηση στη τιμή της παραγόμενης ενέργειας όσο και αυτές που υπάρχει πτώση σε αντίθεση με τα άλλα δύο μοντέλα (AR, ARMA) που δεν προβλέπουν σωστά ορισμένες φορές την κατεύθυνση της αλλαγής.

Η δυσχέρεια που εμφανίζουν τα (AR, ARMA) να εκτιμήσουν σωστά τη μεταβολή στη παραγωγή ηλεκτρικής ενέργειας φαίνεται από τα διαγράμματα απεικόνισης των προβλέψεων τους. Για παράδειγμα το AR μοντέλο από την 1η στη 2η, από την 3η στην 4η, από την 6η στην 7η και από την 7η στην 8η προβλέπει σε ανάποδη κατεύθυνση την μεταβολή των τιμών ενώ το ARMA πράττει ακόμη χειρότερα στο τομέα αυτόν με το να προβλέπει σωστά την κατεύθυνση της μεταβολής μόνο από την 3η στην 4η και από την 7η στην 8η. Πιθανότατα, ο κύριος λόγος που ένα πιο απλό γραμμικό μοντέλο όπως αυτού του κινούμενου μέσου παραδίδει σε εμάς τα καλύτερα αποτελέσματα να οφείλεται στο γεγονός ότι η χρονοσειρά μας δεν χαρακτηρίζεται από ξεκάθαρα patterns καθ' όλη τη διάρκεια εξέλιξής της παρά εμφανίζει διακυμάνσεις οι οποίες δεν χωρούν ντετερμινιστική εξήγηση.

Τα μοντέλα που περιέχουν αυτοπαλίνδρομες διαδικασίες για να περιγράψουν την τιμή μιας παρατήρησης στηρίζονται στις παρελθοντικές τιμές αυτής. Ωστόσο όπως διαπιστώθηκε από τα διαγράμματα αυτοσυσχέτισης δεν γίνεται να στηριχθούμε σε πολλές παρελθοντικές στιγμές πέρα της πρώτης για την περιγραφή μιας παρατήρησης της χρονοσειράς. Έτσι, δεν αποδίδουν καλύτερα από το μοντέλο κινούμενο μέσο το οποίο για να εκφράσει τη τιμή μιας παρατήρησης βασίζεται στο γραμμικό συνδυασμό τιμών και τυχαίων θορύβων παρελθοντικών παρατηρήσεων.

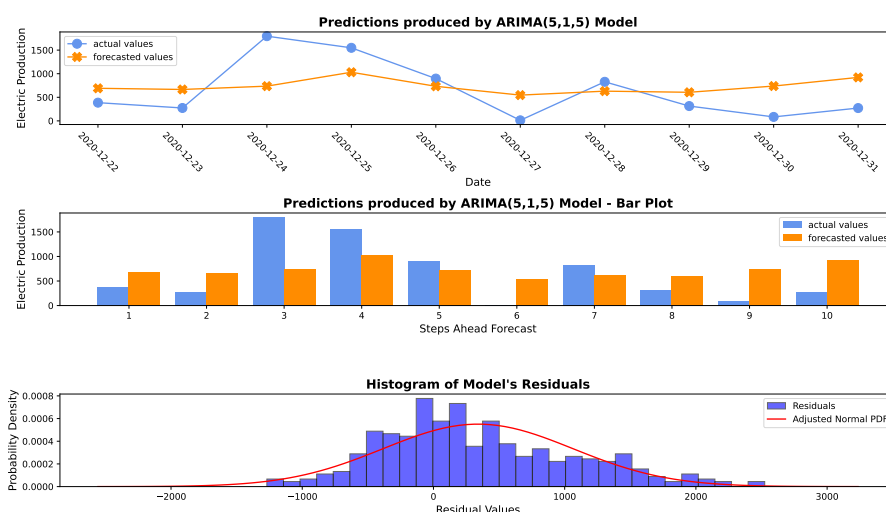
Ιδιαίτερη προσοχή δίδεται επίσης στα ιστογράμματα υπολοίπων των AR και MA μοντέλων. Τόσο στο ένα όσο και στο άλλο παρατηρείται δεξιά λοξότητα. Η μικρή διαφορά που εμφανίζουν στην προβλεπτική τους ικανότητα ίσως δικαιολογείται από το επίπεδο συγκέντρωσης τιμών κοντά στο μηδέν. Σε αυτό παρατηρείται ότι το MA μοντέλο παρουσιάζει ελάχιστα μεγαλύτερη γύρω από αυτό σε σχέση με το AR το οποίο παρουσιάζει κορυφές και μακρυά από το μηδέν. Επίσης το ιστόγραμμα του AR μοντέλου παρουσιάζει ισχυρότερη δεξιά λοξότητα. Με τη παρατήρηση αυτή αντιλαμβανόμαστε ότι το μοντέλο AR κατά την εκπαίδευση του όταν έτεινε να κάνει υπερεκτίμηση των πραγματικών τιμών το έκανε σε μέγεθος μεγαλύτερο από αυτό των υποεκτιμήσεων. Στο MA παρατηρούνται παρόμοιες ενδείξεις σε μικρότερο βαθμό όμως. Αν κάτι είναι κοινό στα δύο ιστογράμματα είναι το μέγεθος μέχρι το οποίο υποεκτιμούν.

Στην αξιολόγηση, πρέπει πάντα να προστίθεται ο παράγοντας της χρονικής πολυπλοκότητας των αλγορίθμων αναζήτησης της καλύτερης τάξης των όρων p, q και εκπαίδευσης των μοντέλων επί του πλήθους των δεδομένων. Με τη χρήση ενός συμβατικού υπολογιστή, η εκτέλεση των δύο βασικών αυτών αλγορίθμων για το μοντέλο ARMA απαιτεί περίπου δέκα φορές παραπάνω χρόνο σε σχέση με το μοντέλα AR ή MA (5 λεπτών έναντι μισού). Σημαντικό να αναφερθεί ότι ο χρόνος αυτός αυξάνεται σημαντικά όταν εξετάζονται και εκπαιδεύονται μοντέλα μεγαλύτερης τάξης p και πιο ειδικά τάξης μεγαλύτερη του 6. Συνολικά από τα 3 μοντέλα το μοντέλο MA(6) έδωσε την καλύτερη πρόβλεψη αν και δεν είχε την μικρότερη τιμή

από όλα στο κριτήριο AIC. Ωστόσο, το MA μοντέλο χρειάζεται τον λιγότερο χρόνο για την διεκπεραίωση της όλης διαδικασίας και προσφέρει την πιο εύστοχη πρόβλεψη. Αυτοί είναι λόγοι ικανοί να δικαιολογήσουν τη προτίμηση μας σε αυτό σε σχέση με τα άλλα.

Στη περίπτωση των προβλέψεων στη παραγωγή ηλεκτρικής ενέργειας ο χρόνος είναι πολύτιμος καθώς υπάρχουν πάντα συγκεκριμένα χρονικά περιθώρια μέσα στα οποία πρέπει να ολοκληρώνεται και να υποβάλλονται οι προτάσεις στο Χρηματιστήριο Ενέργειας. Για να γίνει αυτό, πρώτο και σημαντικό βήμα αποτελεί η παραγωγή και αξιολόγηση μιας πρόβλεψης εντός λογικών χρονικών πλαισίων.

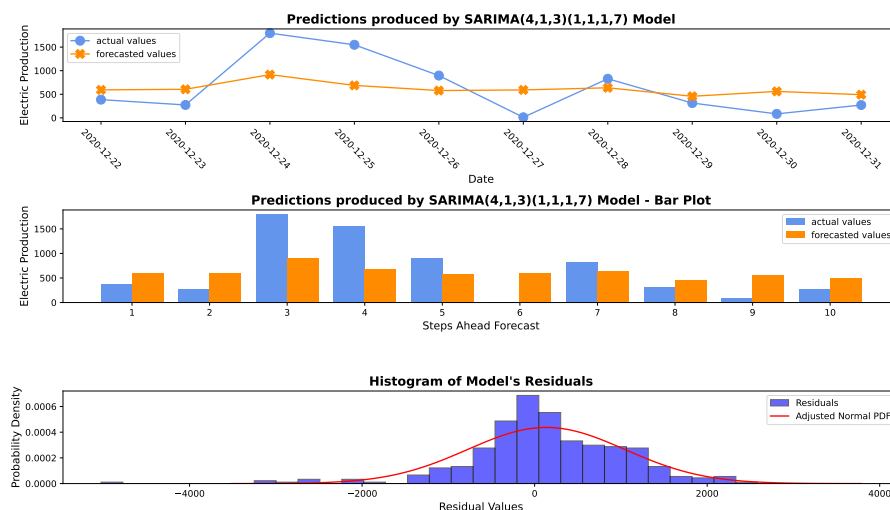
Το MA μοντέλο χρειάζεται τον ελάχιστο χρόνο για την διεκπεραίωση της όλης διαδικασίας. Το γεγονός αυτό σε συνάρτηση με την απόδοση των καλύτερων αποτελεσμάτων ενισχύει την ανωτερότητα του έναντι των άλλων γραμμικών μοντέλων. Οι χρόνοι εκπαίδευσης των μοντέλων δεν διέφεραν από τους χρόνους εκπαίδευσης των αντίστοιχων μοντέλων που εξετάστηκαν στην ανάλυση των δεδομένων της Kaggle λόγω του εύρους τιμών αναζήτησης των τάξεων και του πλήθους των δεδομένων που είναι παρεμφερή με την ανάλυση αυτή.



Σχήμα 7.48: Πρόβλεψη με χρήση του γραμμικού μοντέλου ARIMA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες.

Οι προβλέψεις που έγιναν με τα μοντέλα ARIMA και SARIMA παρουσιάζονται με τα Σχήματα 7.48 και 7.49. Στο μεν πρώτο, απεικονίζεται η πρόβλεψη που παρήγαγε το βέλτιστο μοντέλο ARIMA, ενώ στο δεύτερο απεικονίζεται η πρόβλεψη που παρήγαγε το βέλτιστο εποχικό μοντέλο πρόβλεψης SARIMA. Διεσχάθηκε η ίδια διαδικασία όπως με τα μοντέλα AR, MA, ARMA όσον αφορά τον τρόπο εύρεσης της βέλτιστης τάξης των όρων p, q για το μη εποχικό μοντέλο ARIMA και $(p, q), (P, Q)$ για το εποχικό. Τόσο στο μη εποχικό μοντέλο όσο και στο εποχικό δύναται να διαφοριστούν οι τιμές των παρατηρήσεων της χρονοσειράς. Η διαφορίση των λογαριθμικών δεδομένων κατά μια τάξη ($d = 1$) απαλείφει τη τάση που υπάρχει στη χρονοσειρά και σταθεροποιεί τη μέση τιμή όπως ανακαλύψαμε κατά τη διαδικασία μετατροπής της μη στάσιμης χρονοσειράς μας σε στάσιμη.

Το μοντέλα που έφεραν την μικρότερη τιμή AIC ήταν τα ARIMA(5, 1, 5) και SARIMA(4, 1, 5)(1, 1, 1)₇.



Σχήμα 7.49: Πρόβλεψη με χρήση του γραμμικού μοντέλου SARIMA της ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη για τις επόμενες δέκα ημέρες.

Το SARIMA αξιοποιεί επίσης το εποχικό μέρος του και με διαφόριση ανά 7 παρατηρήσεις ήτοι μια εβδομάδα. Απάλειψε τη συνιστώσα της περιοδικότητας και έδωσε μια χρονοσειρά περισσότερο στάσιμη προς αξιοποίηση στην εκπαίδευση του μοντέλου. Με βάση αυτά παρήχθησαν οι δύο προαναφερόμενες προβλέψεις που φαίνονται στα Σχήματα 7.48 και 7.49. Με την παρατήρηση των δύο αυτών σχημάτων φαίνεται ότι το μοντέλο SARIMA δίνει καλύτερη πρόβλεψη (Σχήμα 7.49) σε σχέση με αυτήν του μη εποχικού μοντέλου ARIMA (Σχήμα 7.48) αλλά και του μοντέλου ARMA (Σχήμα 7.47). Παρόλα αυτά τα πιο απλά σε πολυπλοκότητα μοντέλα AR και MA φέρουν καλύτερες προβλέψεις τόσο από το ARIMA όσο και από το SARIMA.

Το μοντέλο ARIMA, γίνεται άμεσα αντιληπτό ότι έφερε τη χειρότερη πρόβλεψη από όλα. Σε λίγες μελλοντικές παρατηρήσεις προσεγγίζει πάρα πολύ τις πραγματικές τιμές (βλέπε $h = 5, 7$) ενώ στις υπόλοιπες αποκλίνει τόσο ώστε να θεωρηθούν άστοχες. Ακόμη φαίνεται συνήθως να μην ακολουθεί την κάθοδο ή την άνοδο στις τιμές της μεταβλητής ενδιαφέροντος. Ίσως το γεγονός αυτό να προέρχεται από το γεγονός ότι περίπλοκα μοντέλα σαν και αυτό καταφέρνουν να μοντελοποιούν εν μέρη το θόρυβο που πιθανότητα έχουν τα συγκεκριμένα δεδομένα και ο οποίος τελικά οδηγεί σε λάθος εκτιμήσεις το μοντέλο.

Το SARIMA από μεριάς του φέρει προβλεπόμενες τιμές κοντά στις πραγματικές για $h = 1, 2, 5, 7, 8, 10$. Ακόμη σε αντίθεση με το μοντέλο ARIMA προβλέπει σωστά την άνοδο ή πτώση των τιμών ανά παρατήρηση σε σχεδόν όλες τις περιπτώσεις γεγονός που το κατατάσσει τρίτο σε αποτελεσματικότητα με RMSE ίσο με 493,62 kWh. Ωστόσο, το SARIMA ήταν επίσης υπολογιστικά το πιο κοστοβόρο από όλα καθώς η εύρεση του απαιτούσε περίπου 100 φορές τον χρόνο που καταναλώθηκε για το μοντέλο ARMA ή ARIMA. Το μοντέλο SARIMA επειδή πιθανότατα καταφέρνει να λάβει υπόψιν του την περιοδικότητα μπορεί και αποδίδει καλά μετά από την 6η παρατήρηση σε αντίθεση με το MA του οποίου η απόδοση εξασθενούσε μετά από εκείνο το σημείο.

Τα μοντέλα φέρουν μεγάλη διαφορά στον χρόνο εύρεσης και εκπαίδευσης τους. Ενδεικτι-

κά, η εύρεση και εκπαίδευση του καλύτερου μοντέλου ARIMA χρειάζεται περίπου 5 λεπτά με το SARIMA να απαιτεί σχεδόν 50 λεπτά. Προκειμένου να κερδηθεί χρόνος στη περίπτωση του SARIMA η αναζήτηση των τιμών της τάξεως των παραγόντων p, q έγινε στο διάστημα $[4,10]$ καθώς αναμενόταν ότι για μικρότερους όρους δεν θα δινόταν τόσο καλή πρόβλεψη. Οι εποχικοί όροι P, Q είχαν εύρος αναζήτησης $[0,2]$. Παρατηρήθηκε ξανά ότι ο χρόνος αυξάνεται σημαντικά όταν εξετάζονται και εκπαιδεύονται μοντέλα μεγαλύτερης τάξης p, q και P, Q .

Εγείρονται ξανά προβληματισμοί ως προς τον χρόνο απόκρισης των αλγορίθμων σε σχέση με την ποιότητα του αποτελέσματος. Στη προκείμενη περίπτωση, θεωρώ ότι η επιλογή του πιο ποιοτικού μοντέλου έναντι του λιγότερου ποιοτικού είναι προτιμητέα ακόμη και αν το κόστος χρόνου είναι υψηλό. Πρόκειται για μια πρόβλεψη η οποία αναφέρεται στη ημερήσια παραγωγή των επόμενου δεκαημέρου και είναι χρονικά πολύ μικρό το αντίτιμο σε σχέση με τον χρονικό ορίζοντα της πρόβλεψης. Στη περίπτωση μας όμως οι διαφορές ανάμεσα στη ποιότητα των αποτελεσμάτων του κάθε μοντέλου είναι φανερά μεγάλες με αποτέλεσμα να μην μπορούμε στη διαδικασία να μελετήσουμε το κόστους χρόνου σαν αντίβαρο στην απόφαση μας. Προτιμητέο θα κριθεί το μοντέλο που απέκλινε λιγότερο.

Όσον αφορά τα ιστογράμματα υπολοίπων, σε αυτό του SARIMA παρατηρείται καλύτερη προσομοίωση στον λευκό θόρυβο έναντι όλων. Είναι ελαφρώς αρνητικά σκιασμένο με την κορυφή του να προσεγγίζει το μηδέν και μάλιστα να παρουσιάζει πολύ υψηλότερη συγκέντρωση από ότι θα παρουσίαζε στο σημείο αυτό ένα ιδανικό μοντέλο. Αποτελεί μια ένδειξη ότι έχει καταφέρει να προσομοιώσει πολύ καλά πολλές από τις τιμές των δεδομένων εκπαίδευσης αλλά δε παύουν να υπάρχουν συγκεντρώσεις σε τιμές μακριά από αυτό στο αριστερό μέρος του ιστογράμματος. Μάλιστα ορισμένες από αυτές υποδεικνύουν πολύ μεγάλες αποκλίσεις. Παρόλα αυτά δεν εμφανίζει μακριές ουρές ούτε δεξιά ούτε αριστερά της κορυφής του. Είναι πιο συμμετρικά κατανεμημένα γύρω από την κορυφή. Εάν μπορεί να εξαχθεί μια πληροφορία γύρω από τον τρόπο που κατανέμονται αυτά ίσως είναι το γεγονός ότι στις περιπτώσεις υποεκτίμησης η διαφορά ανάμεσα στις πραγματικές και τις προβλεπόμενες τιμές θα είναι μεγαλύτερη από όταν θα κάνει υπερεκτίμηση για κάποιες περιπτώσεις. Μάλιστα είναι το μόνο μοντέλο που παρουσιάζει ότι κατά την εκπαίδευση του προέβη σε τόσο μεγάλο μεγέθους υποτιμήσεις.

Το ARIMA φέρει ένα ιστόγραμμα υπολοίπων από την εκπαίδευση του το οποίο δε προσομοιώνει το λευκό θόρυβο. Έχει υψηλή συγκέντρωση τιμών γύρω από το μηδέν αλλά έχει επίσης μια πολύ μακριά ουρά στο δεξιό μέρος του. Αυτές οι παρατηρήσεις μας οδηγούν στο συμπέρασμα ότι αν και μπορεί να εκτιμήσει με ακρίβεια πολλές φορές υπάρχουν πολλές περιπτώσεις που υποτιμά. Υπερεκτιμά επίσης αλλά αναμένεται το φαινόμενο υποτίμησης να είναι πιο έντονο στις μελλοντικές προβλέψεις του. Το γεγονός αυτό αποδεικνύεται με τη παρατήρηση των προβλέψεων. Μπορεί να υπάρχουν 7 περιπτώσεις υπερεκτίμησης και μόλις 3 περιπτώσεις υποτίμησης αλλά το μέγεθος της απόκλισης στις περιπτώσεις της υποτίμησης είναι μεγαλύτερο από αυτό της υπερεκτίμησης.

7.3.5 Αξιολόγηση Προβλέψεων

Σε πρώτη φάση πραγματοποιήθηκε μια εμπειρική ανάγνωση των γραφημάτων που παρουσιάζουν τις προβλέψεις που παρήγαγε το κάθε γραμμικό μοντέλο. Κατά αυτόν τον τρόπο είναι δυνατό να εκτιμηθεί εμπειρικά η ποιότητα πρόβλεψης του κάθε μοντέλου. Είναι αναγκαίο, όμως, να γίνει μια πιο ενδελεχής μελέτη καθώς η ανάγνωση και μελέτη μόνο του κριτηρίου ρίζας μέσου τετραγωνικού σφάλματος RMSE δεν επαρκεί. Για το λόγο αυτό κάνουμε χρήση μερικών ακόμα κριτηρίων αξιολόγησης και γίνεται σύγκριση των τιμών ανά μοντέλο.

Ιδιαίτερη βαρύτητα, αρχικά, δόθηκε στις τιμές της ρίζας μέσου τετραγωνικού σφάλματος (RMSE) με βάση τις οποίες αξιολογήθηκε η απόδοση των προβλέψεων που παρήγαγε το κάθε μοντέλο και έγινε η κατάταξη των γραμμικών μοντέλων που παρουσιάζονται με φθίνουσα σειρά στον Πίνακα 7.3. Παράλληλα υπολογίζονται και παρουσιάζονται επιπλέον στατιστικά κριτήρια αξιολόγησης για τη πρόβλεψη που παρήγαγε το κάθε μοντέλο.

Μοντέλο	AIC	RMSE (kWh)	R^2	Corr	MAE (kWh)	MAPE	ME (kWh)	MPE
ARIMA(5,1,5)	1109.36	540.74	0.14	0.46	476.55	5.91	88.7	5.64
ARMA(4,5)	1102.98	506.45	0.24	0.63	429.92	7.93	77.68	7.69
SARIMA(4,1,3),(1,1,1,7)	1127.29	493.62	0.28	0.80	420.7	5.85	-29.1	5.53
AR(10)	1109.19	474.78	0.338	0.64	426.66	4.76	2.05	4.43
MA(6)	1106.66	461.65	0.37	0.83	424.44	6.11	117.11	6.1

Πίνακας 7.3: Χαρακτηριστικά αποτελέσματα αξιολόγησης μοντέλων πρόβλεψης Αιολικού πάρκου στη Θράκη. Χρησιμοποιούνται τα κριτήρια αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος(RMSE), μέσου σφάλματος(ME), συντελεστή συσχέτισης(Corr), μέσου απόλυτου σφάλματος(MAE), ποσοστιαίου μέσου απόλυτου σφάλματος(MAPE), ποσοστιαίου μέσου σφάλματος (MPE) και συντελεστή προσδιορισμού(R^2)

Από τα κριτήρια αξιολόγησης που αξιοποιήθηκαν, πρώτα θα μελετηθεί αυτό του συντελεστή προσδιορισμού R^2 με το ίδιο να λαμβάνει τιμές από 0% έως 100% γενικά. Στη περίπτωση των μοντέλων που αξιοποιήθηκαν κανένα από αυτά δεν ξεπέρασε το 40%. Καθώς οι τιμές τους δεν προσεγγίζουν το 100%, εύκολα γίνεται αντιληπτό ότι γιατί τα μοντέλα δεν απέδωσαν άριστες προβλέψεις. Γενικά, ισχύει ότι επί του ίδιου όγκου δεδομένων, όσο μεγαλύτερη είναι η τιμή του, τόσο μικρότερη είναι η διαφορά που υφίσταται ανάμεσα στις τις πραγματικές τιμές και στις τιμές πρόβλεψης που εκτίμησε και παρήγαγε το μοντέλο που προσαρμόστηκε στα δεδομένα εκπαίδευσης.

Στη περίπτωση μας, όπως φαίνεται από τον Πίνακα 7.3, το μοντέλο MA(6) λαμβάνει το υψηλότερο ποσοστό της τάξεως του 37% και ακόμη παρουσιάζει την μικρότερη τιμή RMSE και ίση με 461,65 kWh. Επιβεβαιώνεται, συνεπώς ότι και από αυτά τα δύο κριτήρια αξιολόγησης ότι η εμπειρική εκτίμηση ως προς το μοντέλο MA και την διάκριση του ως το καλύτερο έναντι των άλλων ήταν ορθή. Η τιμή R^2 που λαμβάνει το μοντέλο MA(6), γύρω στο 37% υποδεικνύει ότι αν και γίνεται μια σεβαστή προσπάθεια από το μοντέλο να παράξει μια καλή πρόβλεψη, αλλά

υπάρχει χώρος για βελτίωση στον τρόπο με τον οποίο μπορεί να εξηγηθεί η μεταβλητότητα των τιμών της εξαρτημένης μεταβλητής.

Το πιο σημαντικό συμπέρασμα όμως από την αξιοποίηση του κριτηρίου αυτού προκύπτει εάν παρατηρηθούν συνολικά οι τιμές που λαμβάνουν όλα τα γραμμικά μοντέλα. Μέτα το μοντέλο MA ακολουθεί με ποσοστό 34% αυτό της αυτοπαλινδρόμησης AR και έπονται με 24% και 14% αυτά του αυτοπαλινδρόμου κινούμενου μέσου ARMA και αυτοπαλινδρόμου ολοκληρωμένου μέσου ARIMA αντίστοιχα. Παρατηρείται, δηλαδή, ότι όσο πιο περίπλοκο το μοντέλο που αναπτύχθηκε τόσο πιο μικρή είναι η τιμή που έλαβε στο κριτήριο αυτό. Εξαιρέση στη περίπτωση αυτό αποτέλεσε το SARIMA το οποίο αν και δεν έφερε χειρότερη τιμή από τα AR και MA αλλά καλύτερα από τα υπόλοιπα. Η παρατήρηση αυτή είναι ιδιαίτερα σημαντική για να εξηγηθεί ο λόγος που τα πιο σύνθετα γραμμικά μοντέλα δεν έδωσαν καλύτερες προβλέψεις από τα πιο απλά. Δεν θα ήταν λάθος να υποστηριχθεί ότι τα πιο σύνθετα μοντέλα πιθανώς είχαν ανάγκη μεγαλύτερου όγκου δεδομένων ώστε να υπολογίσουν με ακρίβεια τις παραμέτρους τους σε σχέση με τα πιο απλά.

Ένας άλλος λόγος που τα πιο σύνθετα μοντέλα (ARMA, ARIMA, SARIMA) δεν κατάφεραν να ξεπεράσουν σε απόδοση τα πιο απλά (AR, MA είναι η ύπαρξη θορύβου στα δεδομένα της χρονοσειράς. Θόρυβος ή τυχαίες και απότομες διακυμάνσεις, επηρεάζουν την απόδοση των σύνθετων αυτών μοντέλων καθώς μπορούν να τον προσαρμόζουν κατά την εκπαίδευση τους, κάνοντας τα ευαίσθητα στην παρουσία του. Ακόμη το γεγονός ότι η χρονοσειρά αυτή άλλοτε εμφάνιζε σε ορισμένα τμήματα της τάση και περιοδικότητα ενώ σε άλλα δεν εμφάνιζε ή εμφάνιζε μια από τις δύο συνιστώσες, μπορεί να οδήγησε αυτά τα μοντέλα εσφαλμένα σε παρανόηση ως προς το σε ποια κομμάτια της χρονοσειράς έπρεπε να προσαρμόζεται στις συνιστώσες αυτές και πότε όχι. Το SARIMA πιθανώς προσημείωσε καλύτερα τα δεδομένα επειδή η εξάλειψη της εποχικότητας θα έφερε τη χρονοσειρά πολύ πιο κοντά σε στασιμότητα από ότι θα την έφερνε για παράδειγμα μόνο η σταθεροποίηση της μέσης τιμής της με την εξάλειψη της τάσης. Ένας άλλος πιθανώς παράγοντας είναι ότι χρησιμοποιεί χαμηλότερης τάξης διαδικασίες αυτοπαλινδρόμησης και κινούμενου μέσου. Δεν αυξάνει την πολυπλοκότητα του ιδιαίτερα κατά αυτόν τον τρόπο.

Επόμενο κριτήριο αξιολόγησης το οποίο αξιοποιείται στη διαδικασία αξιολόγησης είναι αυτό της μέσης τετραγωνικής ρίζας $RMSE$. Την μεγαλύτερη τιμή $RMSE$ εμφανίζει το μοντέλο ARIMA, ενώ το μικρότερο μέσο τετραγωνικό σφάλμα ρίζας κατέχει το μοντέλο κινούμενου μέσου MA. Όπως και στο κριτήριο του συντελεστή προσδιορισμού, έτσι και στο κριτήριο αυτό παρατηρείται η τιμή του να αυξάνεται από το πιο απλό στο πιο σύνθετο μοντέλο. Το μοντέλο AR έρχεται δεύτερο με ελάχιστη διαφορά, ενώ το SARIMA τρίτο. Η διαφορά απόκλισης ανάμεσα σε αυτά τρία μοντέλα είναι μικρότερη των 35 kWh. Έχει ενδιαφέρον η παρατήρηση των τιμών σε αυτό το μοντέλο καθώς μετριούνται σε ποσότητες μετρήσιμες και προσφέρουν καλύτερη αντίληψη του σφάλματος πρόβλεψης.

Το μέσο τετραγωνικό σφάλμα ρίζας από μοντέλο σε μοντέλο έχει διαφορά μερικών δεκάδων kWh και συνολικά υπάρχει διαφορά περίπου 80 kWh ανάμεσα στο χειρότερο μοντέλο (ARIMA) και το καλύτερο (MA). Με την παρατήρηση αυτή, γίνεται ακόμη πιο ασφαλές το συμπέρασμα ότι ελάχιστη διαφορά υπάρχει στην προβλεπτική τους ικανότητα. Πιθανώς απο-

δίδουν παρόμοια επειδή οι τιμές των παρατηρήσεων συνθέτουν μια χρονοσειρά η οποία εμπεριέχει χαρακτηριστικά μη γραμμικότητας. Σε αυτή τη περίπτωση μοντέλα σαν αυτά, γραμμικά δηλαδή, αδυνατούν να προσομοιώνουν αυτές τις υποκύπτουσες μη γραμμικές σχέσεις που υπάρχουν. Ειδικά το SARIMA πέρα του ότι έχει μοντελοποιηθεί με βάση τις καλύτερες δυνατές παραμέτρους, έχει σταθεροποιήσει τη μέση τιμή, την διασπορά και εξαλείφει συνιστώσες τάσης και εποχικότητας. Εάν αναπτύσσονταν αμιγώς γραμμικές σχέσεις θα μπορούσε να τις προσομοιώσει καλύτερα από όλα αλλά αντίθετα αυτό συμβαίνει από ένα μοντέλο κινούμενο μέσου. Αυτής της οικογένειας τα γραμμικά μοντέλα θεωρούνται συνήθως καταλληλότερα όταν καλούνται να προσομοιώσουν χρονοσειρές με τυχαίες διακυμάνσεις και βραχυπρόθεσμες εξαρτήσεις (short term dependencies).

Ένα ακόμη κριτήριο αξιολόγησης το οποίο λαμβάνουμε υπόψιν μας είναι αυτό του συντελεστή αυτοσυσχέτισης (*corr*). Σε όλα τα μοντέλα η τιμή του ανέρχεται του 0 και άρα συμπεραίνετε ότι υπάρχει κάποια θετική γραμμική σχέση ανάμεσα στις δύο μεταβλητές. Για τα MA, AR, ARMA και ARIMA συναντώνται οι τιμές 0.83, 0.64, 0.63 και 0.46 αντίστοιχα. Πληροφορίες χρήσιμες επί της γραμμικής σχέσης που υπάρχει ανάμεσα στις τιμές που προβλέπονται από το μοντέλο και τις πραγματικές, μπορούν να εξαχθούν μελετώντας αυτές. Για παράδειγμα, στη περίπτωση του MA μοντέλου που η τιμή του συντελεστή αυτοσυσχέτισης είναι θετική και κοντά στη μονάδα, αντιλαμβανόμαστε ότι υπάρχει ισχυρή θετική γραμμική σχέση ανάμεσα στην ποσότητα ενέργειας που παράχθηκε και αυτή που προβλέφθηκε. Οι τιμές των συντελεστών αυτοσυσχέτισης στις προβλέψεις που προέρχονται από τα AR, ARMA και ARIMA μοντέλα αναδεικνύουν ότι υπάρχει μια θετική γραμμική σχέση μέσης ισχύς ανάμεσα στις δύο μεταβλητές η οποία όμως μειώνεται όσο μεταβαίνουμε σε ένα πιο περίπλοκο μοντέλο. Αντίθετα, τα πιο απλά μοντέλα δείχνουν να έχουν προβλέψεις οι οποίες είναι πιο κοντά σε μέγεθος και κατεύθυνση προς τις πραγματικές από ότι τα πιο περίπλοκα.

Σε ακόμη ένα κριτήριο αποδεικνύεται ότι όσο μεγαλύτερη πολυπλοκότητα εμφανίζει ένα μοντέλο τόσο πιο πολύ έχει δυσκολευτεί να προσαρμοστεί αποτελεσματικά στα συγκεκριμένα δεδομένα εκπαίδευσης. Από την αλλή μεριά τα πιο απλά μοντέλα φαίνεται ξανά ότι έχουν προσαρμοστεί καλύτερα και αποτυπώνουν σε καλύτερο βαθμό τις γραμμικές σχέσεις τάσης που υπάρχουν. Ωστόσο, το κριτήριο αυτό δεν αρκεί για την πλήρη αξιολόγηση της απόδοσης των μοντέλων πρόβλεψης αφού για το SARIMA φαίνεται να υπάρχει παρόμοια με το MA ισχυρή θετική γραμμική σχέση. Φαίνεται ότι παρότι δεν προσφέρει τη καλύτερη δυνατή πρόβλεψη είναι πολύ ικανό στο να προβλέψει την μείωση ή αύξηση της μέσης ημερήσιας παραγόμενης ποσότητας ηλεκτρικής ενέργειας από ημέρα σε μέρα. Είναι μια απόδειξη ότι το SARIMA μας που αναπτύχθηκε για να συλλάβει τα εποχικά μοτίβα της χρονοσειράς, επιτυγχάνει το έργο του αλλά παράλληλα φαίνεται ότι μοντελοποιεί το θόρυβο και για αυτό δεν επιτυγχάνει ακόμα καλύτερες προβλέψεις.

Στις επόμενες στήλες του Πίνακα 7.3 συναντάμε τις τιμές που υπολογίσθηκαν για τα κριτήρια αξιολόγησης MAE, MAPE, ME, MPE. Αρχικά, το απόλυτο μέσο σφάλμα MAE αποτελεί μια εκτίμηση που στοχεύει στο μέγεθος του σφάλματος και όχι στο πρόσημο (κατεύθυνση) του. Με άλλα λόγια αυτό το κριτήριο υποδηλώνει ότι όσο πιο χαμηλές τιμές λάβει, τόσο πιο μεγάλη ακρίβεια στις προβλέψεις του παρουσιάζει το εξεταζόμενο μοντέλο. Με βάση

αυτό το θεωρητικό υπόβαθρο και τα όσα έχουν παρατηρηθεί έως τώρα, το μοντέλο ARIMA είναι αναμένεται να έχει τη μεγαλύτερη τιμή. Πράγματι, αυτό συμβαίνει, όπως φαίνεται από τον Πίνακα 7.3. Αντίθετα την μικρότερη τιμή στο κριτήριο αυτό λαμβάνει το μοντέλο SARIMA δηλαδή, το μοντέλο που όπως υποδεικνύουν τα προηγούμενα κριτήρια δεν είχε κάνει τις καλύτερες προβλέψεις. Σε αυτή τη περίπτωση αναδεικνύεται ο λόγος για τον οποίον αξιοποιούνται παραπάνω από ένα κριτήρια αξιολόγησης.

Το κριτήριο MAE δεν δίνει βάρος σε μεγάλη λάθη όπως γίνεται με το RMSE το οποίο τα τιμωρεί. Αυτός είναι ο λόγος που αυτό κριτήριο αναδεικνύει ως καλύτερο μοντέλο πρόβλεψης το SARIMA και όχι το MA. Είναι ένα ακόμη κριτήριο το οποίο επισημαίνει τη ποιότητα του μοντέλου που παρήγαγε την εκάστοτε πρόβλεψη αφού η ερμηνευση του θεωρείται εύκολη μιας και η τιμή του σφάλματος βρίσκεται στην ίδια κλίμακα με την μεταβλητή που εξετάζεται δηλαδή τη μέση ποσότητα ηλεκτρικής ενέργειας που παράχθηκε σε μια ημέρα μετρημένη σε kWh. Επίσης αν και παρατηρείται ότι όσο αυξάνεται η περιπλοκότητα του μοντέλου, αυξάνεται επίσης το μέσο απόλυτο σφάλμα, η διαφορά που υπάρχει ανάμεσα τους είναι αμελητέα. Πέρα από το ARIMA μοντέλο του οποίου το MAE προσεγγίζει τις 476 kWh, των υπολοίπων η τιμή στο κριτήριο έγκειται στο διάστημα [420,429] kWh. Είναι μια ακόμη παρατήρηση με την οποία φανερώνεται η ομοιότητα των μοντέλων σε βαθμό απόδοσης.

Το κριτήριο μέσου σφάλματος ME εάν ερμηνευθεί σωστά υποδεικνύει εάν υπερεκτιμήθηκε ή έγινε υποεκτίμηση στην προβλεπόμενη τιμή της μεταβλητής ενδιαφέροντος σε σχέση με την πραγματική. Εάν για παράδειγμα είναι θετικό στο πρόσημο, υποτιμάται η ποσότητα ενέργειας που επρόκειτο να παραχθεί, αλλιώς έχουμε υπερεκτίμηση. Κατά αυτόν τον τρόπο γίνεται ποσοτικοποίηση της συνολικής μεροληψίας (bias) και αποτυπώνεται στην πρόβλεψη. Συνολικά, το κριτήριο μέσου σφάλματος είναι χρήσιμο αν είναι επιθυμητό να ελεγχθεί η μεροληψία επί της πρόβλεψης και όχι το μέγεθος του λάθους επί αυτής.

Στα μοντέλα που εξετάζονται εντοπίζουμε υποεκτίμηση σε όλα εκτός του SARIMA, όπου εκεί παρατηρείται γενικά υπερεκτίμηση στις ποσότητες ενέργειας που προβλέπεται ότι θα παραχθούν. Για το μοντέλο AR εντοπίζουμε μια τιμή θετική μεν αλλά ελάχιστα αποκλειόμενη από το μηδέν αφού προσεγγίζει το 2. Μια παρατήρηση η οποία συνάδει με τα καλά και παρπλήσια αποτελέσματα που έφερε επίσης το μοντέλο αυτό και με το γεγονός ότι στις μισές περίπου περιπτώσεις υπερεκτίμησε και στις άλλες μισές υποεκτίμησε με αποτέλεσμα να απαλειφθούν οι διαφορές. Από την άλλη μεριά το MA μοντέλο και η τιμή που φέρει μας προκαλεί έναν προβληματισμό ο οποίος όμως λύνεται μόλις παρατηρηθεί το Σχήμα 7.46. Αναμενόταν να έχει τη χαμηλότερη τιμή αλλά όπως φαίνεται και από το Σχήμα 7.46 μετά την 6η παρατήρηση φθίνει σε απόδοση με αποτέλεσμα να αυξάνει το σφάλμα του. Τυγχάνει στα σημεία που υπερεκτιμά σε σχέση με τις πραγματικές τιμές να υπερκεράσει τις διαφορές που αναπτύσσονται στις περιπτώσεις που αν και υποτιμάει, φέρει προβλέψεις πολύ κοντά στις πραγματικές. Έτσι, δημιουργεί αυτή την αίσθηση ότι υπό εκτιμά σε μεγαλύτερο βαθμό σε σχέση με όλα τα υπόλοιπα.

Παράλληλα με το MAE, εξετάζεται η ποσοστοτικοποίηση του κριτηρίου αυτού, η οποία ονομάζεται ποσοστιαίο μέσο απόλυτο σφάλμα (MAPE). Η χρήση του και η ερμηνεία του είναι σχετικά εύκολη. Εκφράζει τη ποσοστιαία διαφορά ανάμεσα στις πραγματικές τιμές της μετα-

βλητής που εξετάζεται και τις τιμές που προβλέφθηκαν. Όσο πιο μικρή η τιμή που λαμβάνει, τόσο καλύτερο μπορεί να θεωρηθεί το μοντέλο που παρήγαγε τη πρόβλεψη. Στη περίπτωση των μοντέλων αυτών λαμβάνει πολύ μεγάλες τιμές του MA μοντέλου αναμέναμε τη μικρότερη τιμή, αλλά δεν συναντάται αυτή. Αντίθετα, κατέχει τη δεύτερη χειρότερη. Για παράδειγμα, το μοντέλο AR λαμβάνει ποσοστό 476% γιατί υπολογίζεται ότι οι προβλέψεις του είναι κατά μέσο όρος 5 φορές μεγαλύτερες από τη πραγματική. Αντικρίζοντας τα ποσοστά αυτά μεμονωμένα αποδεικνύεται ότι τα μοντέλα αυτά δεν μπορούν να παράξουν έμπιστα αποτελέσματα.

7.3.6 Προβλέψεις με Χρήση XGBoost Μοντέλων και Αξιολόγηση

Η διεργασία που ολοκληρώθηκαν στη στατιστική ανάλυση του Αιολικού Πάρκου 2 της Θράκης, οδήγησαν τελικά στην εκπαίδευση γραμμικών μοντέλων πρόβλεψης για την παραγωγή προβλέψεων. Ωστόσο, αυτά δεν έφεραν τα πλέον ικανοποιητικά αποτελέσματα. Με στόχο την παραγωγή καλύτερων προβλέψεων επιλέχθηκε να αξιοποιηθεί ο αλγόριθμος μηχανικής μάθησης XGBoost. Είναι ένας αλγόριθμος ο οποίος στοχεύει στην συνεχή εκπαίδευση αδύναμων μαθητών. Κάθε ένας από αυτούς έχει ως στόχο τη διόρθωση των λαθών που έκανε ο προηγούμενος με βάση τον οποίο εκπαιδεύεται ώστε τελικά να δημιουργηθεί ένα ισχυρό μοντέλο πρόβλεψης. Με βάση αυτόν τον αλγόριθμο εκπαιδεύτηκαν τρία μοντέλα, το XGB-1, το XGB-2 και τα το XGB-3.

Τα τρία μοντέλα XGBoost που εκπαιδεύτηκαν είχαν μια σημαντική διαφορά με τα γραμμικά. Εκπαιδεύτηκαν με διαφορετικό όγκο δεδομένων. Τα γραμμικά μοντέλα εκπαιδεύτηκαν αξιοποιώντας 356 παρατηρήσεις που κάθε μια από αυτές αποτυπώνει την μέση ποσότητα ηλεκτρικής ενέργειας που παράχθηκε ανά ημέρα καθ όλη τη διάρκεια του 2020. Τα XGBoost μοντέλα κατά την εκπαίδευση τους χρησιμοποίησαν ίδιας φύσεως δεδομένα αλλά διπλάσια σε όγκο. Άντλησαν πληροφορίες από της ημερήσιες παρατηρήσεις δύο ετών (2019-2020), ήτοι 725 παρατηρήσεις, για να εκπαιδευτούν. Η επιλογή αυτή έγινε γιατί τα μοντέλα μηχανικής μάθησης δεν εμφάνιζαν τα μειονεκτήματα που παρουσίαζαν τα γραμμικά μοντέλα στον ίδιο όγκο πληροφοριών. Δεν επηρεάζονται από τυχόν αστοχίες των μετρήσεων (π.χ κενές τιμές) αλλά ούτε αυξάνεται χρονικά το κόστος εύρεσης και εκπαίδευσης του βέλτιστου μοντέλου σε επίπεδα που να θεωρείται ασύμφορο. Ο ορίζοντας πρόβλεψης που τέθηκε ήταν ο ίδιος που τέθηκε και στα γραμμικά.

Μετά τον διαχωρισμό των δεδομένων, σε δεδομένα εκπαίδευσης και επαλήθευσης σειρά είχε η εκπαίδευση των μοντέλων. Καίριο ρόλο στην εκπαίδευση και ανάπτυξη ενός αποδοτικού μοντέλου XGBoost διαδραματίζει η ρύθμιση των υπερπαραμέτρων (Hyperparameter tuning) του. Μια άρτια ρύθμιση αυτών μπορεί είτε να βελτιώσει την απόδοση του μοντέλου είτε να μειώσει την πολυπλοκότητα του ώστε να εμφανίζει λιγότερες περιπτώσεις υπερεκτίμησης είτε να μειώσει το χρόνο που χρειάζεται για να εκπαιδευτεί.

Για την αναζήτηση των βέλτιστων τιμών υπερπαραμέτρων προτείνονται αρκετές μεθοδολογίες όπου ανάμεσα σε αυτές βρίσκεται αυτή της τυχαίας αναζήτησης ή της εξαντλητικής. Στη προκείμενη περίπτωση χρησιμοποιήθηκε η εξαντλητική αναζήτηση. Αφού επιλέχθηκαν οι υπερπαραμέτροι που θα ρυθμιζόνταν βέλτιστα, ορίστηκαν πιθανές τιμές αυτών και αναζητήθηκε

ο συνδυασμός που βελτιστοποιεί την απόδοση του μοντέλου στο στάδιο εκπαίδευσης. Ουσιαστικά, καλύτερες τιμές παραμέτρων κρίθηκαν αυτές με τις οποίες το μοντέλο προσαρμοζόταν καλύτερα στα δεδομένα εκπαίδευσης. Με την ολοκλήρωση της και την εύρεση τους πραγματοποιήθηκε χειροκίνητα μια ακόμη πιο λεπτομερής δοκιμή τιμών σε περιοχές τιμών γύρω από τις βέλτιστες. Προς ρύθμιση επιλέχθηκαν οι ακόλουθες μεταβλητές :

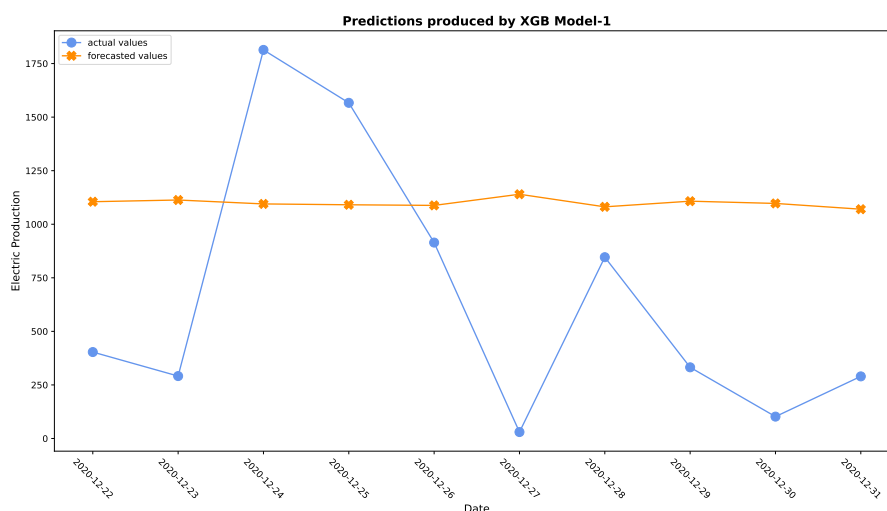
- **learning rate** : Πρόκειται για την υπερπαραμέτρο που ορίζει την τιμή του ρυθμού εκμάθησης των δέντρων που δημιουργούνται, εκπαιδεύονται και προστίθενται στο μοντέλο. Με βάση τον ρυθμό αυτό, σταθμίζεται ο ρυθμός εκπαίδευσης των νέων δέντρων, έτσι ώστε να μην γίνει γρήγορη προσαρμογή στα δεδομένα και να αποφευχθεί η υπερπροσαρμογή (overfitting) του μοντέλου σε αυτά. Εάν επιλεγεί ένας μικρός ρυθμός στάθμισης, στην ουσία επιλέγεται να γίνουν λιγότερες διορθώσεις ανά δέντρο. Σε αντίθετη περίπτωση γίνονται μεγαλύτερου βαθμού διορθώσεις, οπότε το μοντέλο καταλήγει να ελαχιστοποιεί τη συνάρτηση κόστους πολύ γρήγορα. Δοκιμάστηκαν οι τιμές 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, και 0.01.
- **(n_estimators** : Πρόκειται για την υπερπαραμέτρο που ορίζει το πλήθος των δέντρων που θα εκπαιδευτούν. Συνήθως με την αύξηση του πλήθους αυτών, βελτιώνεται η απόδοση του μοντέλου. Βέβαια, η πιθανότητα εμφάνισης υπεραπροσαρμογής αυξάνεται παράλληλα. Δοκιμάστηκαν οι τιμές 300, 250, 200, 150, 100 και 50.
- **max_depth** : Πρόκειται για την υπερπαραμέτρο που ορίζει το μέγιστο βάθος των δέντρων. Με την αύξηση του, αυξάνεται ο βαθμός πολυπλοκότητας του δέντρου. Πιο συγκεκριμένα, το μοντέλο καταφέρνει κατά αυτόν τον τρόπο να προσομοιώσει καλύτερα τα δεδομένα εκπαίδευσης. Ωστόσο, στη περίπτωση αυτή, λόγω υπερπροσαρμογής υπάρχει ο κίνδυνος να μη μπορεί να αποδώσει το ίδιο καλά σε νέα δεδομένα. Δοκιμάστηκαν οι τιμές 2, 3 και 4.
- **gamma** : Πρόκειται για μια υπερπαραμέτρο που η αξιοποίηση της έχει ως σκοπό την αποτροπή της εμφάνισης του φαινομένου της υπερπροσαρμογής. Με βάση την τιμή αυτής ελέγχεται εάν επιτρέπεται να διαιρεθεί περαιτέρω το δέντρο κατά τη κατασκευή του. Εάν το κέρδος διάσπασης (gain) που υπολογίζεται στον εκάστοτε κόμβο είναι μεγαλύτερο από την τιμή της παραμέτρου αυτής, η οποία ουσιαστικά λειτουργεί σαν κατώφλι, τότε το δένδρο διασπάται καθώς υπάρχει περιθώριο μείωσης του μέσου απόλυτου σφάλματος. Ορίζει ουσιαστικά την ελάχιστη μείωση που πρέπει να επιτευχθεί κατά τη διάσπαση του δέντρου. Η αύξηση στη τιμή του οδηγεί σε ένα πιο συντηρητικό μοντέλο, το οποίο γενικεύει την απόδοση του καλύτερα καθώς ελέγχεται η ανάπτυξη του δέντρου και δεν οδηγούμαστε σε μοντέλα τα οποία προσαρμόζονται πολύ καλά στα δεδομένα εκπαίδευσης μόνο. Η ανάπτυξη δέντρων όπου η μεταβλητή αυτή λαμβάνει μικρές τιμές, οδηγεί στη δημιουργία δέντρων πιο περίπλοκων με υψηλή πιθανότητα υπερπροσαρμογής. Δοκιμάστηκαν οι τιμές 0, 5, 10, 25, 50 και 75.
- **objective** : Πρόκειται για την υπερπαραμέτρο με την οποία διευκρινίζεται ποια συνάρτηση απώλειας θα ελαχιστοποιηθεί. Επιλέχθηκε η squarederror που χρησιμοποιείται σε

προβλήματα παλινδρόμησης και στοχεύει στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος ρίζας (RMSE) που ορίστηκε ως μέτρο αξιολόγησης μέσω της μεταβλητής `eval_metric`.

Εκπαίδευση του XGB-1 και παραγωγή προβλέψεων

Με την ολοκλήρωση του αλγορίθμου εξαντλητικής αναζήτησης πραγματοποιήθηκε σε πρώτο φάση η λεπτομερής ρύθμιση (fine tuning) των τιμών των παραμέτρων του πρώτου μοντέλου XGB-1. Σε δεύτερη φάση έγινε αναζήτηση τιμών γύρω από αυτές που έφερε ως βέλτιστες ο αλγόριθμος της εξαντλητικής αναζήτησης. Τελικά επιλέχθηκε να χρησιμοποιηθεί για τη διαδικασία της πρόβλεψης το μοντέλο το οποίο κατάφερε να προσαρμόζεται καλύτερα στα δεδομένα εκπαίδευσης. Με το πέρας της διαδικασίας εκπαίδευσης, αξιολογήθηκε η απόδοση του εκπαιδευμένου μοντέλου με βάση το κριτήριο αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος (RMSE). Ο υπολογισμός αυτού έδωσε ότι κατά μέσο όρο η τιμή που προσάρμοζε το XGB-1 απείχε από την αντίστοιχη αληθινή 751.03 kWh. Πρόκειται για μια ευκαταφρόνητη διαφορά ανάμεσα τους αλλά όχι τέτοια σε μέγεθος που να αποτρέψει την προσπάθεια πρόβλεψης.

Το XGB-1 δεν προσαρμόζεται λεπτομερώς στα δεδομένα εκπαίδευσης και ίσως αυτό να είναι το πλεονέκτημα του έναντι ενός μοντέλου που θα τα προσάρμοζε κατά αυτόν τον τρόπο. Παρόλα αυτά, οι προβλέψεις που έφερε το μοντέλο σε παρατηρήσεις εκτός δείγματος ήταν εξίσου άστοχες. Στο Σχήμα 7.50 αποτυπώνονται αυτές παράλληλα με τις πραγματικά καταγεγραμμένες τιμές. Εμπειρικά παρατηρείται ότι δεν μπορεί να προβλέψει ούτε την κατεύθυνση των μεταβολών της μέσης ημερήσιας παραγωγής ούτε το μέγεθος αυτής. Με βάση τις ενδείξεις αυτές, η προσπάθεια πρόβλεψης από το XGB-1 αποτελεί κατά πάσα περίπτωση την χειρότερη από όσες έχουν γίνεις έως τώρα.



Σχήμα 7.50: Πρόβλεψη με χρήση του μοντέλου XGB-1 που αξιοποίησε ιστορικά δεδομένα ενός χρόνου για την εκπαίδευση του προκειμένου να προβλέψει τις ποσότητες ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη κάθε μία από τις επόμενες δέκα ημέρες.

Για την ολοκλήρωση της αξιολόγησης της απόδοσης του χρησιμοποιήθηκαν ορισμένα κριτήρια αξιολόγησης. Με το κριτήριο ρίζας μέσου τετραγωνικού σφάλματος υπολογίσθηκε ότι υπήρχε κατά μέσο όρο μια διαφορά της τάξεως των 736,81 kWh κατά μέσο όρο ανάμεσα σε πραγματική και προβλεπόμενη τιμή. Αποτελεί μια προσπάθεια πρόβλεψης η οποία είναι η χειρότερη από όσες έχουν υλοποιηθεί και αξιολογηθεί έως τώρα. Αυτή η απόκλιση είναι κατά περίπου 200 kWh μεγαλύτερη από αυτή του γραμμικού μοντέλου ARIMA. Το γραμμικό αυτό μοντέλο υπενθυμίζεται ότι έφερε την χειρότερη πρόβλεψη από όσα γραμμικά αξιοποιήθηκαν με βάση το κριτήριο RMSE το οποίο υποδείκνυε μια απόκλιση της τάξης των 540.74 kWh. Με βάση επίσης το κριτήριο του μέσου σφάλματος το XGB-1 εμφάνιζε ακόμη μεγάλη υπερεκτίμηση τιμών αφού η τιμή που λάμβανε ήταν της τάξεως των 430 περίπου kWh. Σε ένα ακόμη πολύ βασικό κριτήριο εμφάνιζε τη χειρότερη τιμή από όσα είχαν αξιοποιηθεί και αξιολογηθεί.

Από τα φτωχά αποτελέσματα που έφερε η προηγούμενη διαδικασία, εξήχθη ένα πολύ χρήσιμο συμπέρασμα. Έγινε αντιληπτό ότι ένα μοντέλο XGB παρά το γεγονός ότι κατατάσσεται στα πιο ισχυρά εργαλεία πρόβλεψης δεν φέρει απαραίτητα άριστα αποτελέσματα χωρίς τα κατάλληλα δεδομένα προς αξιοποίηση. Δεν αρκεί η σωστή ρύθμιση των παραμέτρων του. Το μοντέλο XGB-1 εικάζεται ότι δεν απέδωσε ικανοποιητικά αφού η διαθέσιμη πληροφορία που είχε προς αξιοποίηση στο στάδιο εκπαίδευσης του δεν ήταν αρκετή για την κατάλληλη διάσπαση των δέντρων. Τα δεδομένα εκπαίδευσης παρείχαν την χρονική στιγμή που δειγματοληπτήθηκε η τιμή της παρατήρησης και την τιμή αυτής ως χαρακτηριστικά προς αξιοποίηση. Με βάση αυτά δεν κατάφερε το μοντέλο να ελαχιστοποιήσει το μέσο τετραγωνικό σφάλμα ρίζας και να προσαρμόσει ικανοποιητικά τα δεδομένα εκπαίδευσης.

Εκπαίδευση του XGB-2 και παραγωγή προβλέψεων

Τα δεδομένα με την μορφή που είχαν έως τώρα δεν γινόταν να αξιοποιηθούν από ένα μοντέλο XGBoost με τρόπο ικανοποιητικό για την σκιαγράφηση γραμμικών ή μη γραμμικών σχέσεων που υπήρχαν ανάμεσα στις παρατηρήσεις. Αποφασίστηκε η επεξεργασία των δεδομένων με τέτοιο τρόπο ώστε να εξαχθούν πιο χαρακτηριστικές πληροφορίες. Η διαδικασία αυτή είναι γνωστή ως feature engineering (εξαγωγή χαρακτηριστικών). Τα υπάρχοντα διαθέσιμα δεδομένα επεξεργάστηκαν ώστε για κάθε παρατήρηση να δημιουργηθεί ένα σύνολο από νέες πληροφορίες προς αξιοποίηση για το μοντέλο εκπαίδευσης. Από τις ημερομηνίες καταγραφής της κάθε παρατήρησης, κρίθηκε κρίσιμο να υπολογιστούν για την κάθε μια τα εξής χαρακτηριστικά :

- η ημέρα της χρονιάς με τιμές στο διάστημα [1,365]
- η ημέρα της εβδομάδας με τιμές στο διάστημα [1, 7]
- ο μήνας της χρονιάς με τιμές στο διάστημα [1, 12]
- ο χρόνος με τιμές στο διάστημα [2019, 2020]
- η εβδομάδα της χρονιάς με τιμές στο διάστημα [1, 52]

Τα χρονικά χαρακτηριστικά δεν ήταν τα μόνο επιπλέον γνωρίσματα που εξήχθησαν από τη χρονοσειρά. Κρίθηκε χρήσιμο να ληφθούν υπόψιν οι δέκα προηγούμενες τιμές της κάθε μιας και να αποτελούν επιπρόσθετο γνώρισμα μιας παρατήρησης. Τα χαρακτηριστικά που εξάγονταν από τις χρονικές στιγμές δειγματοληψίας δημιουργήθηκαν για να βοηθήσουν το μοντέλο να σχηματίζει πιθανές σχέσεις που υπάρχουν ανάμεσα σε παρατηρήσεις ή μοτίβα που επαναλαμβάνονται στη ακολουθία των τιμών της μεταβλητής που μας ενδιαφέρει να προβλέψουμε. Από την άλλη, η ανάπτυξη των χαρακτηριστικών υστέρησης, θα προσφέρει πληροφορία στο μοντέλο περί των συσχετίσεων που μπορεί να υπάρχουν ανάμεσα σε μια παρατήρηση και τις παρελθοντικές του. Τέλος, γίνεται να περιγραφούν οι παροδικές είτε αυξητικές είτε πτωτικές τάσεις που εμφανίζονται στη χρονοσειρά αυτή. Έτσι, το νέο μοντέλο θα αξιοποιεί πληροφορίες με τις οποίες θεωρητικά θα μπορεί να προσαρμόζεται πιο εύστοχα στις τιμές των παρατηρήσεων.

Συνολικά πλέον από την κάθε παρατήρηση έχουν εξαχθεί και αξιοποιούνται στο στάδιο εκπαίδευσης τα εξής γνωρίσματα :

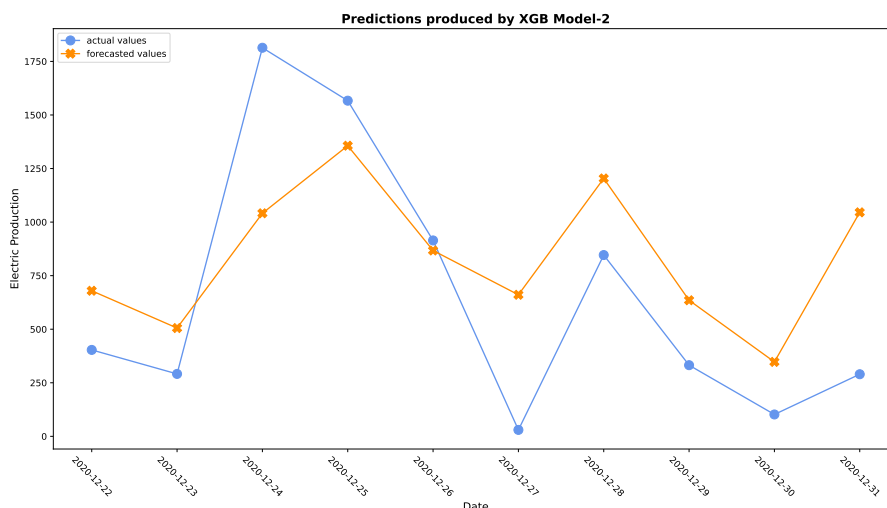
- `day_of_year` : η ημέρα της χρονιάς με τιμές στο διάστημα [1,365]
- `day_of_week` : ημέρα της εβδομάδας με τιμές στο διάστημα [1, 7]
- `month_of_year` : ο μήνας της χρονιάς με τιμές στο διάστημα [1, 12]
- `year` : ο χρόνος με τιμές στο διάστημα [2019, 2020]
- `week_of_year` : η εβδομάδα της χρονιάς με τιμές στο διάστημα [1, 52]
- X_{t-n} : η τιμή της παρατήρησης που υστερεί κατά n χρονική απόσταση από την υπό μελέτη παρατήρηση όπου $n \in \{1, 10\}$ και $n \in \mathbb{Z}$.

Τη διαδικασία δημιουργίας των χαρακτηριστικών ακολούθησε η διαδικασία επιλογής των γνωρισμάτων που θα αξιοποιούνταν τελικώς. Τα 15 διαφορετικά χαρακτηριστικά που δημιουργήθηκαν όταν χρησιμοποιήθηκαν όλα μαζί για την εκπαίδευση ενός μοντέλου έφεραν πολύ άσχημα αποτελέσματα. Όταν επίσης ελέγχθηκε η σημασία που δίνει στο κάθε ένα από αυτά το μοντέλο πολλά από αυτά παραμελούνταν πλήρως. Δε λαμβάνονταν υπόψιν στη διαδικασία εκπαίδευσης. Σε πρώτη φάση αποφασίστηκε να μειωθεί ο αριθμός αυτών που θα συμμετείχαν περίπου στο μισό, έτσι ώστε να μην υπάρξει κάποιο που δεν αξιοποιήθηκε. Η πράξη αυτή δεν οδηγούσε σε αποπροσανατολισμό του μοντέλου εκπαίδευσης ενώ παράλληλα μειωνόταν ο υπολογιστικός χρόνος εκπαίδευσης του κάθε μοντέλου.

Επιλέχθηκαν μετά από πολλές δοκιμές και διαφορετικούς συνδυασμούς χαρακτηριστικών έξι μόνο γνωρίσματα. Τα χαρακτηριστικά της χρονοσειράς τα οποία θα βοηθούσαν στη καλύτερη εκπαίδευση του XGBoost μοντέλου ήταν αυτά των τιμών των παρατηρήσεων που υστερούν από 2 έως 6 χρονικές αποστάσεις (lags) και επίσης το χρονικό χαρακτηριστικό που αποτυπώνει σε ποια ημέρα της εβδομάδας βρίσκεται η κάθε παρατήρηση. Επιλέχθηκε να χρησιμοποιηθεί μόνο αυτό το χρονικό χαρακτηριστικό καθώς ήταν το πλέον κατάλληλο για την περιγραφή της εβδομαδιαίας περιοδικότητας που υπάρχει στη χρονοσειρά. Όσον αφορά τα

χαρακτηριστικά υστερήσεων, στις διάφορες δοκιμές που πραγματοποιήθηκαν, φανερώθηκε ότι το εκάστοτε μοντέλο λάμβανε πιο σοβαρά τις τιμές που υπήρχαν σε χρονικές στιγμές κατά 2 ή 3 ή 4 ή 5 ή 6 χρονικών αποστάσεων (lags) πριν από την μεταβλητή που εκτιμούσε. Το XGB-2 μοντέλο που επιλέχθηκε ανάμεσα σε όσα εκπαιδεύτηκαν ήταν αυτό που κατάφερε να προσαρμοστεί καλύτερα στα δεδομένα εκπαίδευσης.

Η αξιολόγηση της προσαρμογής στα δεδομένα εκπαίδευσης έγινε με βάση το κριτήριο αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος (RMSE). Η απόκλιση ανάμεσα στις τιμές που προσάρμοζε και τις πραγματικές υπολογίστηκε στις 9.98 kWh. Αποδεικνύεται κατά αυτόν τον τρόπο ότι εν τέλει η τακτική πρόσθεσης επιπλέον χαρακτηριστικών απέφερε καρπούς. Η βελτίωση που υπάρχει στην απόδοση του μοντέλου XGB-2 κατά τη διαδικασία εκπαίδευσης είναι εξαιρετική αν αναλογιστούμε ότι το προηγούμενο XGB μοντέλο (XGB-1) στο αντίστοιχο στάδιο προσάρμοζε κατά μέσο όρο 736,81 kWh πέραν της πραγματικής, τις τιμές του. Με δεδομένο ότι το μοντέλο XGB-2 κατάφερε να προσαρμοστεί τόσο καλά στα δεδομένα εκπαίδευσης, γίνεται μια προσπάθεια πρόβλεψης δέκα παρατηρήσεων εκτός δείγματος όπως προηγουμένως. Το αποτέλεσμα της προσπάθειας αυτής φαίνεται στο Σχήμα 7.51.



Σχήμα 7.51: Πρόβλεψη με χρήση του μοντέλου XGB-2 που αξιοποίησε χρήσιμα χαρακτηριστικά δεδομένα που εξήχθησαν από ιστορικά δεδομένα 2 χρόνων για την εκπαίδευση του χρήσιμες χαρακτηριστικές τιμές προκειμένου να προβλέψει τις ποσότητες ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη κάθε μία από τις επόμενες δέκα ημέρες.

Το XGB-2 χρησιμοποίησε διαφορετικό σετ δεδομένων για να εκπαιδευτεί με πολύ πιο στοχευμένες πληροφορίες προς επεξεργασία. Μετά το στάδιο της εκπαίδευσης παρήγαγε μια αξιολογη πρόβλεψη χάριν στα χαρακτηριστικά χρόνου και υστερήσεων που του παρέχονταν πλέον. Τα μεν πρώτα βοήθησαν να σχηματιστεί και να ερμηνευτεί η εβδομαδιαία περιοδικότητα ενώ τα δεύτερα συντέλεσαν στην αποτύπωση των σχέσεων εξάρτησης της κάθε παρατήρησης από βραχείες παρελθοντικές στιγμές. Μια πρώτη ανάγνωση του Σχήματος 7.51 θετική εντύπωση προκαλεί η ικανότητα του XGB-2 να προβλέπει σωστά πότε θα αυξηθεί η ποσότητα

παραγόμενης ενέργειας και τότε θα μειωθεί αυτή. Από παρατήρηση σε παρατήρηση εύστοχα υποδεικνύει εάν θα έχουμε αύξηση ή μείωση στη παραγωγή. Εάν μάλιστα μπορούσε να προβλέψει σωστά σε μέγεθος την ποσότητα που θα παραγόταν από ημέρα σε ημέρα τότε θα αντικρίζαμε την τέλεια πρόβλεψη.

Δεν παρατηρείται μια τέλεια πρόβλεψη γιατί φαίνεται να μην μπορεί να εκτιμήσει σε μέγεθος τις απότομες αυξήσεις ή μειώσεις. Για παράδειγμα, όταν στη πραγματικότητα από την δεύτερη στην τρίτη παρατήρηση υπάρχει μια αύξηση στη μέση παραγωγή της τάξεως των 1500 kWh περίπου το συγκεκριμένο μοντέλο αν και σωστά προβλέπει μια μεγάλη αύξηση, δεν προβλέπει ότι αυτή θα ήταν πέρα των 600 kWh περίπου. Αντίστοιχα από την πέμπτη παρατήρηση στην έκτη όπου φαίνεται να μειώνεται η παραγωγή κατά 900 kWh περίπου το μοντέλο που εκπαιδεύτηκε αν και εκτιμά ότι θα υπάρξει μείωση, αυτή είναι της τάξεως των 150 kWh. Παρά το μέγεθος των δύο αυτών λαθών όπως και της μεγάλης αστοχίας παρεμφερή σε μέγεθος που εμφανίζει στην τελευταία παρατήρηση το μοντέλο κατατάσσεται με βάση το κριτήριο της μέσης τετραγωνικής ρίζας RMSE ως το καλύτερο μοντέλο προβλέψεων συνολικά. Μάλιστα, την ίδια θέση κατέχει και με βάση το κριτήριο μέσου απόλυτου σφάλματος MAE. Στο μεν πρώτο, λαμβάνει την τιμή των 448,78 kWh και στο δεύτερο την τιμή των 381 περίπου.

Η απόκλιση ανάμεσα στα δύο κριτήρια φανερώνει ότι το μοντέλο αυτό πραγματικά μπορεί και φέρνει καλές προβλέψεις αλλά υπάρχουν περιπτώσεις που το σφάλμα του είναι τέτοιο σε μέγεθος που υποθάλπει την προβλεπτική του ικανότητα. Σε όλες τις παρατηρήσεις η κατεύθυνση πρόβλεψης είναι σωστή ενώ σε μέγεθος πέραν των τριών που αναφέρθηκαν δεν αποκλίνει πέρα των 300 kWh η εκτιμώμενη τιμή από την πραγματική. Γεγονός που αποτιμά την απόδοση του μοντέλου ως ικανοποιητική και το ίδιο ως το πιο έμπιστο. Μελανό σημείο αναφοράς στο μοντέλο αυτό αποτελεί το γεγονός ότι υπερεκτιμά στις 7 από τις 10 περιπτώσεις πρόβλεψης. Το γεγονός ότι ενώ το μοντέλο XGB-2 εμφάνισε ένα πολύ μικρό μέσο σφάλμα ρίζας κατά την προσαρμογή του στα δεδομένα εκπαίδευσης δεν έφερε ανάλογο αποτέλεσμα στην πρόβλεψη αποδεικνύει ότι δεν πρέπει να εξειδικεύουμε το μοντέλο σε τόσο μεγάλο βαθμό. Σε μια τέτοια περίπτωση φαίνεται ότι το μοντέλο δεν γενικεύεται το ίδιο αποτελεσματικά σε νέα δεδομένα παρά εμφανίζει το φαινόμενο της υπερπροσαρμογής. Επίσης αυτό το φαινόμενο απορρέει και από το γεγονός ότι υπάρχει θόρυβος στα δεδομένα.

Εκπαίδευση του XGB-3 και παραγωγή προβλέψεων

Σε μια προσπάθεια αντιμετώπισης του φαινομένου υπερπροσαρμογής αποφασίστηκε να γίνουν ορισμένες αλλαγές στον τρόπο με το οποίο θα μοντελοποιηθεί το υποψήφιο XGB μοντέλο. Για να αντιμετωπιστεί ως ένα βαθμό το πρόβλημα αυτό προτείνεται είτε να αυξηθεί ο ρυθμός εκπαίδευσης και να μειωθούν το πλήθος των δέντρων που χτίζονται είτε να μειωθεί το βάθος τους είτε να αυξηθεί ο όγκος πληροφορίας εκπαίδευσης. Το σετ δεδομένων που αξιοποιήθηκε ήταν το ίδιο σε όγκο δεδομένων και πληροφορία. Στη περίπτωση μας δεν διατίθονταν παραπάνω παρελθοντικά δεδομένα πέραν των όσων ήδη αξιωνόντουσαν οπότε το πλήθος τους παρέμεινε το ίδιο. Ακόμη, δεν προστέθηκαν παραπάνω χαρακτηριστικά όπως κάποιο χαρακτηριστικό χρονικό της χρονοσειράς ή κάποιο χαρακτηριστικό υστερήσεων.

Πριν την εκπαίδευση των μοντέλων που μελετώνται είχε εμπειρικά διαπιστωθεί ότι όσα χρησιμοποιήθηκαν την εκπαίδευση του XGB-2 επαρκούσαν. Με δεδομένη την κατάσταση αυτή αποφασίστηκε να γίνει μια αναζήτηση μικρότερων τιμών για το επιθυμητό πλήθος δέντρων και μεγαλύτερων για το ρυθμό εκπαίδευσης, ενώ η τιμή της παραμέτρου που όριζε το βάθος του δέντρου διατηρήθηκε ως είχε. Οι διαφορές στις τιμές των παραμέτρων των δύο μοντέλων γίνονται πιο ξεκάθαρες εάν παρατηρηθεί ο Πίνακας 7.4. Το νέο μοντέλο XGB-3 είχε τον ίδιο όγκο δεδομένων να αξιοποιήσει προς εκπαίδευση αλλά μεγαλύτερο ρυθμό εκπαίδευσης κατά 9.3% από 0.43 σε 0.47 και ίσο με 0.47, ενώ το πλήθος των δέντρων που θα εκπαιδευτούν μειώθηκε κατά 67% από τα 125 στα 41. Με τις τιμές των υπόλοιπων παραμέτρων να παραμένουν αμετάβλητες εκπαιδεύτηκε το μοντέλο XGB-3.

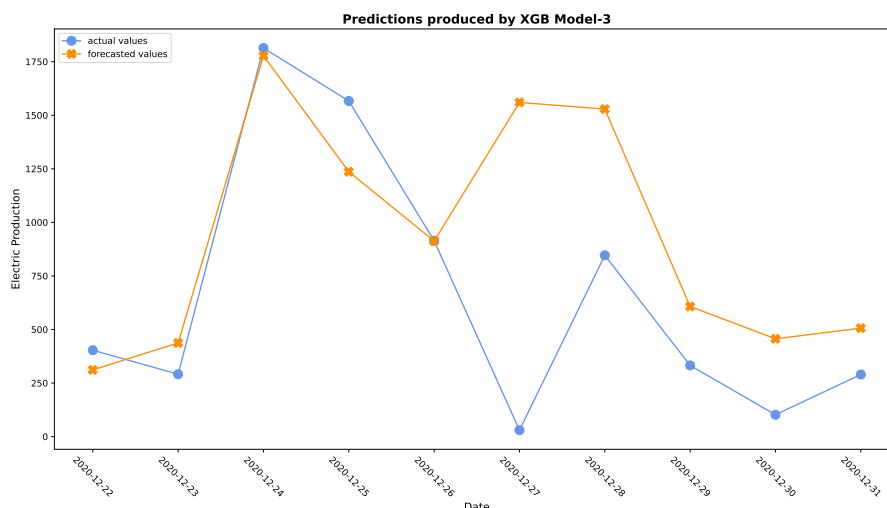
Hyperparameter	XGB-2	XGB-3
Learning rate	0.43	0.47
n_estimators	125	41
max_depth	6	6
gamma	75	75

Πίνακας 7.4: Τιμές υπερπαραμέτρων των μοντέλων XGB-2 και XGB-3.

Στο μοντέλο XGB-3 αφού εκπαιδεύτηκε, έγινε εκτίμηση της προσαρμογής του στα δεδομένα εκπαίδευσης. Με βάση την τιμή των κατά μέσο όρο 70,89 kWh που έλαβε στο κριτήριο ρίζας μέσου τετραγωνικού σφάλματος αποδεικνύεται ότι δεν προσαρμόζεται στον ίδιο βαθμό με το XGB-2 του οποίου οι εκτιμώμενες τιμές απέκλιναν κατά μέσο όρο μόνο κατά 9.98 kWh από τις πραγματικές. Η ένδειξη αυτή πιθανότατα υποδεικνύει ότι δημιουργήθηκε ένα μοντέλο το οποίο θα παρουσιάζει σε μικρότερο βαθμό περιπτώσεις υπερπροσαρμογής. Για να αποτιμήσουμε επίσης την προβλεπτική του ικανότητα παρήχθη πρόβλεψη μέσων ποσοτήτων παραγόμενης ηλεκτρικής ενέργειας. Το διάστημα στο οποίο έγινε η πρόβλεψη ήταν το ίδιο με το διάστημα πρόβλεψης που ορίστηκε προηγουμένως στη περίπτωση του XGB-2. Στο Σχήμα 7.52 παρουσιάζονται οι προβλέψεις του XGB-3 για το διάστημα αυτό μαζί με τις τιμές που πραγματική παράχθηκαν.

Η ανάγνωση του Σχήματος 7.52 όπου φαίνονται οι προβλέψεις του XGB-3 έγινε με πολύ προσοχή και λεπτομέρεια. Το XGB-3 σε σχέση με το XGB-2 δείχνει να μπορεί να προβλέψει με μεγαλύτερη ακρίβεια τις μέσες ποσότητες ηλεκτρικής ενέργειας που επρόκειτο να παραχθούν τις 5 πρώτες μελλοντικές ημέρες. Πιο ειδικά, με βάση το κριτήριο μέσου τετραγωνικού σφάλματος MAE στο διάστημα αυτό, αποκλίνει κατά μέσο όρο 121 kWh περίπου από τις πραγματικές εν αντιθέσει με το XGB-2 το οποίο αποκλίνει με βάση το ίδιο κριτήριο κατά 303 kWh κατά μέσο όρο περίπου.

Η βελτίωση που επιδεικνύει στις προβλέψεις είναι άξια αναφοράς. Οφείλεται στο γεγονός ότι σε σχέση με το προηγούμενο μοντέλο, το XGB-3 μπορεί να προβλέψει πολύ καλύτερα τόσο τις δύο αρχικές τιμές όσο και την μεγάλη αλλαγή στη παραγωγή που υπάρχει από τη δεύτερη στη τρίτη ημέρα που είναι της τάξεως των 1500 kWh. Το XGB-3 πέφτει εκτός πρόβλεψης στη τρίτη παρατήρηση για μόλις 30 kWh την ίδια στιγμή που στην αντίστοιχη περίπτωση το



Σχήμα 7.52: Πρόβλεψη με χρήση του μοντέλου XGB-3 που αξιοποίησε χρήσιμα χαρακτηριστικά δεδομένα που εξήχθησαν από ιστορικά δεδομένα 2 χρόνων για την εκπαίδευση του χρήσιμες χαρακτηριστικές τιμές προκειμένου να προβλέψει τις ποσότητες ενέργειας που θα παράγει το Αιολικό πάρκο 2 στη Θράκη κάθε μία από τις επόμενες δέκα ημέρες.

XGB-2 πέφτει εκτός κατά 780 kWh αφού δεν προβλέπει ότι αυτή η αύξηση θα ήταν πέραν των 600kWh. Και τα δύο μοντέλα στο διάστημα αυτό ακολούθησαν ισχυρά τις ανόδους και πτώσεις (κατευθύνσεις) στις τιμές της μέσης παραγωγής αλλά το XGB-3 υπερτερεί καθώς προβλέπει με ακρίβεια και το μέγεθος αυτών.

Το XGB-3 υπερεκτίμησε λιγότερο σε σχέση με το XGB-2 στις πέντε αυτές παρατηρήσεις. Παρατηρήθηκε μια περίπτωση έναντι δύο του XGB-2, με την φορά που υπερεκτίμησε να είναι για μόλις 145 kWh. Παρατηρούνται επίσης οι περιπτώσεις υποτίμησης. Συγκρίνονται με τις αντίστοιχες περιπτώσεις υποτίμησης του XGB-2 και συμπεραίνεται ότι αυτές επίσης ήταν μικρότερου βαθμού. Εύλογα αποτιμάται ότι η μείωση της πολυπλοκότητας του μοντέλου (αύξηση στο συντελεστή τιμωρίας λαθών και μείωση στο πλήθος των δέντρων) οδήγησε σε ένα μοντέλο το οποίο στις πρώτες 5 παρατηρήσεις δείχνει πιο ισορροπημένο και καταλληλότερα ρυθμισμένο από το XGB-2 αφού υπερτιμά ελάχιστα και όταν υποτιμάει το κάνει σε πολύ μικρό βαθμό. Συνολικά, αποδεικνύει ότι ο σκοπός για τον οποίο σχεδιάστηκε το XGB-3 επιτεύχθηκε στο πρώτο αυτό υποδιάστημα του διαστήματος πρόβλεψης.

Στη συνέχεια το εξεταζόμενο μοντέλο ξεκινάει να υπερτιμά από την έκτη έως τη δέκατη παρατήρηση. Αποτελεί μια κοινή παρατήρηση ανάμεσα στο XGB-3 και XGB-2. Η διαφορά τους έγκειται κυρίως στο μέγεθος και λιγότερο στη κατεύθυνση με την οποία προβλέπουν ότι θα κινηθεί η μέση παραγωγή, αν δηλαδή θα αυξηθεί ή θα μειωθεί από ημέρα σε ημέρα. Το XGB-3 παρουσιάζει πολύ φτωχή απόδοση στο μεσοδιάστημα της 6ης και 7ης παρατήρησης. Δεν προβλέπει ότι από την 5η στην 6η παρατήρηση θα υπάρξει θεαματική μείωση της παραγωγής της τάξεως των 900 kWh περίπου αλλά αύξηση της τάξεως των 650 kWh από τις 900 kWh περίπου που είχαν παραχθεί τη προηγούμενη ημέρα. Στη συνέχεια, προβλέπει ότι η παραγωγή

θα μειωνθεί μερικές δεκάδες kWh ενώ έπρεπε να προβλέψει ότι θα αυξανόταν κατά 800 kWh από τις 30 kWh περίπου. Σε αυτό το υποδιάστημα το μοντέλο εμφανίζει μέσο απόλυτο σφάλμα 1106 kWh περίπου εν αντιθέσει με το XGB-2 που εμφανίζει στο ίδιο υποδιάστημα λίγο πάνω από 500 και πιο συγκεκριμένα 512 kWh. Αυτό συμβαίνει διότι και στις δύο παρατηρήσεις αυτές είχε προβλέψει σωστά την πτώση και άνοδο της παραγωγής. Μάλιστα ενώ στην πτώση η πρόβλεψη της σε μέγεθος ήταν 300 kWh λιγότερες από τη προηγούμενη τιμή αντί για 900, στην άνοδο που προβλέπει ακολούθως σωστά ότι θα υπάρξει, αποτυπώνει ότι το μέγεθος αυτής θα ήταν της τάξεως των 600 kWh αντί για 800 kWh που είναι πραγματικά. Σε αυτό λοιπόν το υποδιάστημα βλέπουμε ότι το XGB-2 αποδίδει πολύ καλύτερα από το XGB-3 του οποίου τα αποτελέσματα ήταν πραγματικά κακά και παραπλανητικά για την επιχείρηση που θα στηριζόταν σε αυτά.

Το τελευταίο υποδιάστημα που εξετάζεται είναι των τριών τελευταίων παρατηρήσεων του χρονικού διαστήματος για το οποίο παράχθηκαν οι προβλέψεις. Το XGB-3 επανέρχεται δυναμικά και δείχνει ξανά ικανό να προβλέψει τόσο την πτώση των τιμών στη παραγωγή τις πρώτες 2 ημέρες του διαστήματος αυτού όσο και την μικρή άνοδο την τελευταία. Επίσης, το XGB-3 καθώς έχει σχεδιαστεί για να υπερεκτιμά λιγότερο από το XGB-2 αποκλίνει μόνο κατά 262 kWh σύμφωνα με το κριτήριο του μέσου απόλυτου σφάλματος εν αντιθέσει με το XGB-2 το οποίο σύμφωνα με το ίδιο κριτήριο λαμβάνει ως τιμή απόκλισης κατά μέσο όρο αυτή των 434 kWh περίπου. Τόσο το ένα όσο και το άλλο δείχνουν ικανά να προβλέψουν τις πτώσεις ή την άνοδο στη παραγωγή με το XGB-3 να επιτυγχάνει το έργο αυτό πιο αποτελεσματικά αφού η σχεδίαση του να υπερεκτιμά λιγότερο ήταν επιτυχής οπότε και έφερε προβλέψεις σε μέγεθος πιο κοντά στις αληθινές αν και πάνω από αυτές.

Συνολικά το μοντέλο XGB-3 προκαλεί μια σύγχυση για την αποτελεσματικότητά του. Ενώ φανέρωσε την ικανότητά του να προβλέπει άριστα στο άμεσο μέλλον, δηλαδή, για τις πρώτες πέντε ημέρες, στη συνέχεια και για δύο μόνο ημέρες απέδωσε πολύ κακά πριν επαναφέρει την αποτελεσματικότητά του σε ικανοποιητικά επίπεδα για τις τελευταίες 3 παρατηρήσεις. Φαινομενικά, εάν δεν απέδιδε τόσο κακές παρατηρήσεις στη μέση του διαστήματος πρόβλεψης θα κατατασσόταν πρώτο σε αποτελεσματικότητα όπως δείχνει άλλοτε η τιμή του κριτηρίου μέσου απόλυτου σφάλματος η οποία συνολικά υπολογίστηκε στις 366 kWh περίπου. Με βάση όμως το κριτήριο ρίζας μέσου τετραγωνικού σφάλματος RMSE το μοντέλο XGB-3 κατατάσσεται προτελευταίο με τιμή αυτή των 565 kWh περίπου, μπροστά μόνο από το XGB-1. Η τιμή που λαμβάνει θα ήταν πολύ μικρότερη εάν δεν υπέκυπτε σε τόσο μεγάλα σφάλματα.

Για την περαιτέρω αξιολόγηση της απόδοσης του XGB-3, επιστρατεύτηκε επίσης συντελεστή συσχέτισης corr. Αυτό δείχνει ότι στα διαστήματα που το μοντέλο XGB-3 έφερε τις καλές προβλέψεις υπήρχε πολύ ισχυρή γραμμική σχέση ανάμεσα στις τιμές τις πραγματικές και αυτές που προβλέφθηκαν με τιμή πολύ κοντά στο 1. Το πρώτο υποδιάστημα καλών προβλέψεων εμπίπτει στις ημερομηνίες 22/12 έως και 26/12 (Τρίτη έως Σάββατο) και το δεύτερο στις ημερομηνίες 29/12 έως και 31/12 (Τρίτη έως Πέμπτη). Φαίνεται λοιπόν ότι σε αυτές τις ημέρες πιθανώς υπάρχουν ισχυρά εβδομαδιαία μοτίβα τα οποία συντελούν στη παραγωγή τελικά καλών προβλέψεων. Αντίθετα, εικάζεται ότι τα μοτίβα που υπάρχουν από Κυριακή σε Κυριακή ή από Δευτέρα σε Δευτέρα, δηλαδή, για το διάστημα 27/12 έως και 28/12 (Κυριακή

με Δευτέρα) δε φαίνεται να είναι ισχυρά ώστε να βοηθήσουν στη σωστή πρόβλεψη των τιμών των συγκεκριμένων ημερών.

Ο συντελεστής συσχέτισης στο διάστημα που έδωσε το μοντέλο πολύ κακές προβλέψεις ήταν -1. Η τιμή του κριτηρίου για το διάστημα των δύο ημερών που προβλέφθηκε με μεγάλη απόκλιση η ποσότητα της ενέργειας που παράχθηκε υποδηλώνει ότι υπάρχει τέλεια γραμμική αρνητική σχέση ανάμεσα στις πραγματικές και τις εκτιμώμενες τιμές. Αυτό σημαίνει ότι όσο μειώνεται η πραγματική τόσο αυξάνεται η εκτιμώμενη. Προβλέφθηκε δηλαδή σωστά κατά μέγεθος η αλλαγή αλλά όχι προς τη σωστή κατεύθυνση. Το XGB-2 από την άλλη μεριά σε όλα τα υποδιαστήματα προβάλλει ότι αναπτύσσονται θετικές γραμμικές ισχυρές σχέσεις ανάμεσα στις πραγματικές τιμές και τις εκτιμώμενες. Πιο συγκεκριμένα, στο πρώτο υποδιάστημα ο συντελεστής αυτοσυσχέτισης υπολογίζεται σε 0.87, στο δεύτερο σε 1 και στο τρίτο σε 0.69. Μέσω της παρατήρησης των τιμών του κριτηρίου αυτού στηρίζεται ακόμα πιο ισχυρά η εμπειρική παρατήρηση που πραγματοποιήθηκε προηγουμένως για την ισχυρή ικανότητα του XGB-2 να προβλέπει εάν από ημέρα σε ημέρα η παραγωγή θα αυξηθεί ή θα μειωθεί και μάλιστα πολύ κοντά σε αναλογία με την πραγματικότητα.

Οι τιμές κριτηρίων που εκμεταλλευτήκαμε για την αξιολόγηση των μοντέλων βρίσκονται στους Πίνακες 7.5, 7.6, 7.7, 7.8 και 7.9 που ακολουθεί. Εκεί βρίσκονται όσες χρησιμοποιήθηκαν παραπάνω για την αξιολόγηση των XGB-2, XGB-3 και αναφέρθηκαν παραπάνω καθώς επίσης και ορισμένες ακόμη των οποίων ο σχολιασμός ακολουθεί.

Μοντέλο XGB	RMSE (kWh)	MAE (kWh)	ME (kWh)
XGB-1	736.04	678.77	439.46
XGB-2	448.78	381.22	175.27
XGB-3	565.4	366.4	274.44

Πίνακας 7.5: Τιμές κριτηρίων αξιολόγησης προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB. Χρησιμοποιούνται τα κριτήρια αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος(RMSE), μέσου σφάλματος(ME) και μέσου απόλυτου σφάλματος(MAE).

Στον Πίνακα 7.9 εμπεριέχονται οι τιμές μέσου σφάλματος των δύο μοντέλων (XGB-2, XGB-3) για τις προβλέψεις που παρήχθησαν για το κάθε ένα από τα τρία υποδιαστήματα. Παράλληλα στον Πίνακα 7.5 είναι διαθέσιμο το συνολικό μέσο σφάλμα για το κάθε μοντέλο. Με διαθέσιμες αυτές τις πληροφορίες θα εξεταστεί περαιτέρω ο βαθμός υπερπροσαρμογής που εμφάνισαν και τα δύο μοντέλα.

Το κριτήριο του μέσου ME σφάλματος είναι σημαντικό να παρατηρηθεί γιατί υποδεικνύει κατά μέσο όρο αν το μοντέλο που αξιοποιήθηκε υποτιμά ή υπερεκτιμά κατά βάση όταν εκτιμά μελλοντικές τιμές της χρονοσειράς. Το XGB-3 το οποίο σχεδιάστηκε με στόχο να μην υπερεκτιμά στο βαθμό που το έκανε το XGB-2 παρουσιάζει συνολικά μεγαλύτερο μέσο θετικό σφάλμα από ότι το XGB-2, 274.44 έναντι 175.27 kWh. Σε πρώτη ανάγνωση φαίνεται ότι η εύρεση των κατάλληλων παραμέτρων, η μετέπειτα εκπαίδευση του και παραγωγή προβλέψε-

Υποδιάστημα παρατηρήσεων	MAE XGB-2 (kWh)	MAE XGB-3 (kWh)
22/12-26/12	303.19	121.48
27/12-28/12	494.71	1106.16
29/12-31/12	434.69	282.03

Πίνακας 7.6: Τιμές κριτηρίου αξιολόγησης μέσου απόλυτου σφάλματος MAE των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.

Υποδιάστημα παρατηρήσεων	RMSE XGB-2 (kWh)	RMSE XGB-3 (kWh)
22/12-26/12	391.12	168.83
27/12-28/12	512.68	1184.79
29/12-31/12	490.35	287.65

Πίνακας 7.7: Τιμές κριτηρίου αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος RMSE των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.

Υποδιάστημα παρατηρήσεων	Corr XGB-2	Corr XGB-3
22/12-26/12	0.87	0.97
27/12-28/12	0.96	-1
29/12-31/12	0.69	0.85

Πίνακας 7.8: Τιμές κριτηρίου αξιολόγησης συντελεστή αυτοσυσχέτισης corr των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.

ων σε διάστημα εκτός του διαστήματος προβλέψεων δεν έφερε το αναμενόμενο αποτέλεσμα αφού το συνολικό μέσο απόλυτο σφάλμα δεν μειώθηκε αλλά αυξήθηκε. Μια πιο προσεκτική ματιά όμως στη τιμή που λαμβάνει το κριτήριο για το κάθε μοντέλο στα υποδιαστήματα της πρόβλεψης είναι χρήσιμη καθώς εξάγονται πολύτιμα συμπεράσματα.

Στο πρώτο υποδιάστημα και στα δύο μοντέλα το μέσο σφάλμα λαμβάνει αρνητική τιμή. Όσο πιο μακριά από το μηδέν αυτή, τόσο μεγαλύτερος ο βαθμός υποτίμησης από το εκάστοτε μοντέλο. Στη προκειμένη περίπτωση το XGB-3 εμφανίζει μεγαλύτερο μέσο σφάλμα, με

Υποδιάστημα παρατηρήσεων	ME XGB-2	ME XGB-3
22/12-26/12	-107.91	-62.81
27/12-28/12	494.13	1106.16
29/12-31/12	434.69	282.03

Πίνακας 7.9: Τιμές κριτηρίου αξιολόγησης μέσου σφάλματος με των προβλέψεων μέσης ημερήσιας ποσότητας ηλεκτρικής ενέργειας από το Αιολικό πάρκο 2 στη Θράκη που απορρέουν από μοντέλα XGB για τα υποδιαστήματα του χρονικού διαστήματος στο οποίο έγινε η πρόβλεψη.

συνέπεια να αποδεικνύεται ότι στο 50% των τιμών των παρατηρήσεων που έπρεπε να προβλεφθούν το μοντέλο αυτό φέρει προβλέψεις οι οποίες κατά μέσο όρο θα είναι κατά 62 kWh μικρότερες από την αναμενόμενη τιμή, ενώ αντίθετα το XGB-2 φέρει μέσο σφάλμα το οποίο υπονοεί ότι οι εκτιμώμενες τιμές κατά μέσο όρο θα είναι 102 kWh μικρότερες από τις πραγματικές. Στο πρώτο αυτό υποδιάστημα το XGB-3 φάνηκε ότι υποτιμάει λιγότερο σε σχέση με το το XGB-2. Είναι μια παρατήρηση η οποία δεν έρχεται σε αντιπαράβολή με τις παρατηρήσεις που εκφράστηκαν μετά την σύγκριση και αξιολόγηση των μοντέλων με προηγούμενα κριτήρια. Το XGB-3 καθώς παράγει καλύτερες προβλέψεις για τις πρώτες 5 ημέρες, δείχνετε ότι υποτιμάει λιγότερο επίσης σε σχέση με το XGB-2.

Στο δεύτερο υποδιάστημα, στο διάστημα των προβλέψεων έκτης και έβδομης ημέρας, το μέσο σφάλμα και στα δύο μοντέλα αυξάνεται πολύ και λαμβάνει θετική τιμή. Στη περίπτωση αυτή, το μεν XGB-3 εκτιμά κατά μέσο όρο 1106 kWh πάνω από τις πραγματικές τιμές, ενώ το XGB-2 κατά μέσο όρο 494 kWh παραπάνω. Το XGB-3 το οποίο σχεδιάστηκε να υπερτιμά λιγότερο από το XGB-2, δεν το καταφέρνει για το συγκεκριμένο υποδιάστημα. Όπως παρατηρήσαμε και προηγουμένως, είναι το υποδιάστημα όπου είναι πάρα πολύ άστοχο καθώς αν και προβλέπει σε μέγεθος την αλλαγή που θα υπάρξει στην παραγωγή δεν προβλέπει σωστά την πορεία αυτής. Αντί να προβλέψει ότι θα μειωθεί προβλέπει ότι θα αυξηθεί. Αντίθετα το XGB-2 φέρει μικρότερο μέσο σφάλμα γιατί αν και δε προβλέπει το μέγεθος αυτής σωστά προβλέπει ότι από την πέμπτη στην έκτη ημέρα θα υπάρξει μείωση στη παραγωγή.

Στο τελευταίο υποδιάστημα, όπου το XGB-3 προβλέπει σωστά από παρατήρηση σε παρατήρηση αν θα υπάρξει μείωση ή αύξηση στη παραγωγή φέρει μικρότερο μέσο σφάλμα από το XGB-2. Στη περίπτωση αυτή αποδεικνύει ότι για ο σκοπός για τον οποίο δημιουργήθηκε, επιτεύχθηκε. Μάλιστα υπερτιμά κατά περίπου 150 kWh λιγότερο σε σχέση με το XGB-2 αφού λαμβάνει την τιμή των 282 kWh αντί των 434 kWh που υπολογίζεται ως μέσο σφάλμα για το XGB-2. Συνολικά η αποτίμηση που μπορεί να γίνει είναι ότι το XGB-3 στις 8 από τις 10 παρατηρήσεις έφερε πολύ καλύτερες προβλέψεις. Η απόδοση του επισκιαζεται από την υψηλή αστοχία που φέρει στο μέσο του διαστήματος πρόβλεψης.

Η επιλογή του καταλληλότερου μοντέλου δεν μπορεί να γίνει αυστηρά με βάση κάποιο κριτήριο. Εάν ενδιαφέρον της εταιρίας είναι να προβλέπει την αύξηση ή μείωση στη παραγωγή σκεπτόμενη ότι οι ακριβές ποσότητες δεν είναι δυνατόν να υπολογισθούν τότε το XGB-2

είναι το πλέον κατάλληλο μοντέλο για τον σκοπό αυτό. Εάν πάλι στόχος της είναι η ακριβής πρόβλεψη των ποσοτήτων αυτών τότε ένα μοντέλο σαν το XGB-3 είναι καταλληλότερο αρκεί να χρησιμοποιείται για προβλέψεις με μικρότερο χρονικό ορίζοντα. Σε κάθε περίπτωση τόσο το ένα όσο και το άλλο μοντέλο φέρουν περιθώρια βελτίωσης είτε αυξάνοντας τον όγκο πληροφορίας τους στο στάδιο εκπαίδευσης είτε προσδίδοντας τους νέες χαρακτηριστικές πληροφορίες που θα τα βοηθήσουν να εκπαιδευτούν ακόμη καλύτερα. Για παράδειγμα επικείμενα μετεωρολογικά φαινόμενα μέσης κλίμακας τα οποία έχουν διάρκεια κάποιων ημερών μπορεί να παύσουν την λειτουργία του αιολικού πάρκου για λόγους ασφάλειας είτε να αυξήσουν άμεσα την παραγωγή ανάλογα τη σφοδρότητα τους και κατόπιν με το πέρας αυτήν να την μειώσουν απότομα.

Κεφάλαιο 8

Συμπεράσματα

Σε αυτήν την διπλωματική εργασία μελετήθηκαν και αναπτύχθηκαν μεθοδολογίες πρόβλεψης της μέσης παραγόμενης ηλεκτρικής ενέργειας από αιολικά πάρκα. Ενδιαφέρον σε τέτοιες προβλέψεις έχουν οι εταιρείες διαχείρισης των αιολικών πάρκων ώστε να γνωρίζουν εκ των προτέρων τις ποσότητες ενέργειας που μπορούν να διαθέσουν στην αγορά ώστε να προβούν στις κατάλληλες τεχνοοικονομικές αποφάσεις. Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση γραμμικών μοντέλων παλινδρόμησης (MA, AR, ARMA, ARIMA, SARIMA) και για το μοντέλο μηχανικής μάθησης XGBoost αποτελούν μονοδιάστατες χρονοσειρές. Τα δεδομένα αυτά αποτελούν μετρήσεις μέσης παραγόμενης ηλεκτρικής ενέργειας κάθε δέκα λεπτά στο χρονικό διάστημα από 01/01/2018 έως 31/12/2020. Οι μετρήσεις αυτές αφορούν σε αιολικό πάρκο στην Θράκη και προέρχονται από υποσταθμό ο οποίος βρίσκεται ορισμένα χιλιόμετρα μακριά από το αιολικό πάρκο.

Στο πλαίσιο της παρούσας εργασίας, αρχικά αναπτύχθηκε στο προγραμματιστικό περιβάλλον της *python* η διαδικασία ανάλυσης και πρόβλεψης χρονοσειρών. Η ποιότητα της αναλυτικής διαδικασίας ελέγχθηκε χρησιμοποιώντας δεδομένα παραγωγής ηλεκτρικής ενέργειας από την εργασία [36] (διαθέσιμη στην διαγωνιστική πλατφόρμα μηχανικής μάθησης Kaggle) και συγκρίνοντας με τα αντίστοιχα αποτελέσματα.

Η διαδικασία της ανάλυσης των δεδομένων συνοψίζεται ως εξής: (1) Διενέργεια ελέγχου της ποιότητας των δεδομένων (έλεγχος για έκτοπες τιμές και κενά δεδομένων). (2) Διερευνητική ανάλυση (διάγραμμα απεικόνισης χρονοσειράς, ιστόγραμμα χρονοσειράς, ανάλυση συνιστωσών, διάγραμμα αυτοσυσχέτισης, ιστόγραμμα υπολοίπων μετά την αφαίρεση τάσης και περιοδικότητας). (3) Έλεγχος στασιμότητας. (4) Εφαρμογή λογαριθμικού μετασχηματισμού σταθεροποίησης. (5) Υπολογισμός διαφορών και επανεκτίμηση στασιμότητας. (6) Διαχωρισμός δεδομένων σε δεδομένα εκπαίδευσης και επαλήθευσης. (7) Αναζήτηση βέλτιστων παραμέτρων των γραμμικών μοντέλων (με τη χρήση αλγορίθμου εξαντλητικής αναζήτησης). (8) Εκπαίδευση των γραμμικών μοντέλων. (9) Αξιολόγηση εκπαίδευσης (ιστόγραμμα υπολοίπων). (10) Εφαρμογή γραμμικών μοντέλων AR, MA, ARMA, ARIMA, SARIMA για παραγωγή προβλέψεων εκτός δείγματος. (11) Αξιολόγηση απόδοσης γραμμικών μοντέλων (παράλληλη οπτικοποίηση εκτιμώμενων και πραγματικών τιμών, υπολογισμός κριτηρίων αξιολόγησης).

Από τη διαδικασία πρόβλεψης της χρονοσειράς που δημιουργήθηκε από δεδομένα παραγω-

γής ηλεκτρικής ενέργειας που παρέχονταν από την εργασία [36] βγήκαν τα εξής συμπεράσματα :

- Ο έλεγχος ποιότητας των δεδομένων έδειξε ότι αξιοποιούνται δεδομένα καλής ποιότητας. Μια τέτοια ένδειξη πάντα ενθαρρύνει την συνέχεια της διαδικασίας.
- Η διερευνητική ανάλυση της χρονοσειράς προσφέρει χρήσιμες πληροφορίες για τη μη στάσιμη φύση της χρονοσειράς. Μελετάται μια χρονοσειρά με σταθερή αυξητική τάση και περιοδικότητα ενός χρόνου.
- Ο λογαριθμικός μετασχηματισμός μπορεί να σταθεροποιήσει σε ικανοποιητικό βαθμό τη διακύμανση.
- Η διαφόριση της χρονοσειράς κατά μια τάξη αφαίρεσε την τάση και σταθεροποίησε τη μέση τιμή της χρονοσειράς.
- Ο έλεγχος στασιμότητας πρέπει να γίνεται κάθε φορά που μεταβάλλεται η χρονοσειρά ώστε να αποτιμώνται οι επιδράσεις της μεταβολής στη φύση της χρονοσειράς και αναλόγως να αναθεωρείται ή όχι η στρατηγική μετατροπής της χρονοσειράς σε στάσιμη.
- Όσο αυξάνεται η πολυπλοκότητα του γραμμικού μοντέλου τόσο αυξάνεται ο χρόνος εκπαίδευσης του. Χαρακτηριστικά αναφέρεται ότι το μοντέλο MA έφερε χρόνο εκπαίδευσης τα 20 δευτερόλεπτα, ενώ τα AR, ARMA, ARIMA και SARIMA έφεραν χρόνο εκπαίδευσης 90, 720, 1800 και 3600 δευτερόλεπτα αντίστοιχα.
- Στο στάδιο εκπαίδευσης το ιστόγραμμα υπολοίπων λειτουργεί ως οριζόντιος άξονας της απόδοσης του μοντέλου στην πρόβλεψη εκτός δείγματος. Το μοντέλο SARIMA του οποίου το ιστόγραμμα προσομοίωνε καλύτερα από όλα τον λευκό θόρυβο έφερε την καλύτερη πρόβλεψη. Αντίθετα το μοντέλο MA που έφερε τη χειρότερη πρόβλεψη, παρουσίαζε στο στάδιο εκπαίδευσης το ιστόγραμμα που προσομοίωνε χειρότερα από όλα τον λευκό θόρυβο.
- Σύμφωνα με το κριτήριο αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος RMSE το μοντέλο SARIMA έκανε την καλύτερη πρόβλεψη και τη χειρότερη το MA. Μάλιστα το SARIMA αποκλίνει κατά μέσο όρο μόνο κατά 4.28 μονάδες σε αντίθεση με το MA που αποκλίνει κατά 19.21. Κάνει, δηλαδή, μια πρόβλεψη κατά 78% πιο εύστοχη.
- Σύμφωνα με το ίδιο κριτήριο το SARIMA κάνει επίσης κατά 20% πιο εύστοχη πρόβλεψη από το δεύτερο σε απόδοση ARIMA. Αυτό συμβαίνει γιατί το SARIMA μπορεί να αντιμετωπίζει επιπλέον την εποχικότητα πέρα από τάση που αντιμετωπίζει τόσο αυτό όσο και το ARIMA.
- Ο συντελεστής συσχέτισης υπολογίστηκε ότι ανερχόταν του 0.8 σε όλα τα μοντέλα πέραν του MA. Υπήρχε δηλαδή ισχυρή θετική σχέση ανάμεσα στις πραγματικές και εκτιμώμενες τιμές. Η ένδειξη αυτή επιβεβαιώνει την ικανότητα τους να προβλέπουν με ακρίβεια εάν θα υπάρξει αύξηση ή μείωση στη παραγωγή από μήνα σε μήνα.

- Τα γραμμικά μοντέλα έφεραν τιμές παραπλήσιες του μηδενός στο κριτήριο μέσου σφάλματος, ένδειξη της επαρκούς εκπαίδευσης τους η οποία δεν οδηγεί σε έντονα φαινόμενα υπερεκτίμησης ή υποτίμησης.
- Σε σύγκριση με τα αποτελέσματα από την εργασία [36] τα μοντέλα που εφαρμόστηκαν στην παρούσα εργασία έφεραν πολύ καλύτερες προβλέψεις. Χαρακτηριστικά αναφέρεται ότι η καλύτερη πρόβλεψη που παράχθηκε σε αυτήν ήταν 59% χειρότερη από την καλύτερη πρόβλεψη της υπό μελέτης διαδικασίας.

Η διαδικασία ανάλυσης και πρόβλεψης χρονοσειρών που παρατέθηκε παραπάνω, στη συνέχεια εξετάστηκε επί μιας χρονοσειράς ποσοτήτων ηλεκτρικής ενέργειας που παρήχθησαν από ένα αιολικό πάρκο. Το βήμα δειγματοληψίας της χρονοσειράς ήταν ημερήσιο και όχι μηνιαίο όπως προηγούμενως. Αποφασίστηκε ότι θα χρησιμοποιηθούν τα δεδομένα του ημερολογιακού έτους 2020 γιατί ήταν ποιοτικά ανώτερα σε σχέση με των ετών 2018 ή 2019. Έτσι, δημιουργήθηκε η χρονοσειρά η οποία είναι πολύ πιθανό να εμπεριέχει εσφαλμένες μετρήσεις που όμως δεν ανιχνεύθηκαν.

Η διαδικασία ελέγχου των δεδομένων ήταν το πρώτο βήμα της ανάλυσης. Στη συνέχεια, εκτελέστηκε ακέραια στο σύνολο της η διαδικασία όπως προηγούμενως. Κατά την εκτέλεση της και από τα αποτελέσματα που παρήχθησαν και αξιολογήθηκαν, εξήχθησαν ορισμένα ενδιαφέροντα συμπεράσματα. Στα πιο βασικά συμπεράσματα συγκαταλέγονται τα εξής :

- Η διερευνητική ανάλυση υπέδειξε τη μελέτη μιας ευμετάβλητης χρονοσειράς. Παρατηρούνται έντονες διακυμάνσεις και απόκλιση από τη Gaussian κατανομή. Μάλιστα, η κατανομή που προσομοιώνει είναι ένας είδος εκθετικής.
- Η συνιστώσα της τάσης άλλοτε έχει πορεία αυξητική και άλλοτε πτωτική. Η περιοδικότητα της είναι εβδομαδιαία αλλά στο διάγραμμα αυτοσυσχέτισης δεν εμφανίζονται σημαντικές τιμές αυτοσυσχέτισης ανάμεσα στη τιμή μιας παρατήρησης και υστερημένων τιμών αυτής κατά κάποια χρονική απόσταση πέραν της πρώτης. Είναι, δηλαδή, ελάχιστη η παρελθοντική πληροφορία στην οποία γίνεται να εξαρτηθεί μια παρατήρηση για να περιγραφεί από αυτήν.
- Ο λογαριθμικός μετασχηματισμός και η διαφόριση της χρονοσειράς σταθεροποίησαν εν μέρει τη διακύμανση και τη μέση τιμή όμως πολλές περιπτώσεις έντονης διακύμανσης παρέμειναν.
- Παρά τη διαδικασία μετατροπής η χρονοσειρά ήταν πιο κοντά στη στασιμότητα αλλά όχι στον επιθυμητό βαθμό.
- Αφού εκπαιδεύτηκαν τα βέλτιστα μοντέλα AR, MA, ARMA, ARIMA SARIMA ερμηνεύτηκε το ιστόγραμμα υπολοίπων τους. Κανένα από τα ιστογράμματα δεν προσομοίωνε σε ικανοποιητικό βαθμό τον λευκό θόρυβο. Πάρα τη την υψηλή συγκέντρωση τιμών κοντά στο μηδέν, οι σκιάσεις που υπήρχαν αποτελούσαν απόδειξη πολλών περιπτώσεων υπερεκτίμησης.

- Τη καλύτερη πρόβλεψη έκανε το μοντέλο MA σύμφωνα με το κριτήριο αξιολόγησης ρίζας μέσου τετραγωνικού σφάλματος RMSE και την χειρότερη το ARIMA. Ήταν πιο εύστοχη για μόλις 15%.
- Γενικά, όλα τα γραμμικά μοντέλα έφεραν παρεμφερείς προβλέψεις οι οποίες έφεραν σφάλμα από 460 περίπου έως 540 kWh. Αποτελούν μέτριες προβλέψεις.
- Το γεγονός ότι ένα MA μοντέλο και όχι ένα SARIMA που αντιμετωπίζει τάση και εποχικότητα έφερε τις καλύτερες προβλέψεις αποδεικνύει ότι η φύση της χρονοσειράς δεν ήταν αμιγώς στάσιμη αλλά εμπεριείχε στοιχεία μη αιτιοκρατικά. Αυτές οι μη γραμμικές σχέσεις δεν μπορούν να μοντελοποιηθούν άριστα από γραμμικά μοντέλα.

Οι προσπάθειες πρόβλεψης με χρήση γραμμικών μοντέλων πρόβλεψης δεν είχαν το ιδανικό αποτέλεσμα. Ο κύριος λόγος είναι το είδος της πρώτης ύλης που δέχτηκαν για να εκπαιδευτούν. Δεν υπήρχαν αμιγώς γραμμικές σχέσεις ανάμεσα στις τιμές της χρονοσειράς. Η συνιστώσα της περιοδικότητας αλλά κυρίως αυτή της τάσης δεν αντιμετωπίστηκαν στο βαθμό που έπρεπε. Μπορεί ο λογαριθμικός μετασχηματισμός να σταθεροποιήσει εν μέρη τη διακύμανση αλλά η τάση η οποία ανά διαστήματα άλλαζε φορά και ρυθμό έκανε την χρονοσειρά να αποσταθεροποιείται. Οι απότομες και ενίοτε ξαφνικές μεταβολές που εμφάνιζε δεν ήταν δυνατό να εξηγηθούν μέσω γραμμικών σχέσεων και άρα ούτε κάποιο από τα διαθέσιμα γραμμικά μοντέλα μπορούσε να τις περιγράψει. Οι εκτιμήσεις τους δεν έφεραν ιδιαίτερες διαφορά σε ποιότητα. Λίγο καλύτερες ήταν αυτές του απλού μοντέλου κινούμενου μέσου και του εποχικού μοντέλου SARIMA.

Μια νέα προσπάθεια πρόβλεψης πραγματοποιήθηκε με μοντέλα μηχανικής μάθησης XG-Boost. Αυτά τα μοντέλα δεν ενδιαφέρονται τόσο για την ύπαρξη τόσο ικανοποιητικών σε ποιότητα δεδομένων, την κανονικότητα που πρέπει να φέρουν στην κατανομή τους και την ανάπτυξη γραμμικών σχέσεων μεταξύ των τιμών της χρονοσειράς. Οι αλγόριθμοι μηχανικής μάθησης με τους οποίους έχουν αναπτυχθεί φροντίζουν να μην αποτελούν αυτές οι συνθήκες κατασταλτικοί παράγοντες σε προβλήματα προβλέψεων. Αναπτύχθηκαν 3 διαφορετικά XG-Boost μοντέλα. Κάθε ένα από αυτά αξιοποιούσε τον διπλάσιο όγκο δεδομένων από ότι ένα γραμμικό μοντέλο. Αξιοποιήθηκαν λοιπόν τα δεδομένα παραγωγής του έτους 2019 παρά τα όσα σφάλματα υπήρχαν σε αυτά.

Η διαδικασία πρόβλεψης περιελάμβανε τα εξής βήματα: (1) Εξαγωγή χαρακτηριστικών γνωρισμάτων από τα δεδομένα. (2) Διαχωρισμός δεδομένων σε δεδομένα εκπαίδευσης και επαλήθευσης. (3) Καθορισμός πιθανών βέλτιστων τιμών υπερπαραμέτρων του μοντέλου (4) Εξαντλητική αναζήτηση του μοντέλου που κατά τη προσαρμογή του στα δεδομένα εκπαίδευσης έφερε το μικρότερο σφάλμα. (5) Αξιολόγηση εκπαίδευσης (κριτήριο ρίζας μέσου τετραγωνικού σφάλματος). (6) Εφαρμογή μοντέλου XGBoost για παραγωγή προβλέψεων εκτός δείγματος. (7) Αξιολόγηση απόδοσης γραμμικών μοντέλων (παράλληλη οπτικοποίηση εκτιμώμενων και πραγματικών τιμών, υπολογισμός κριτηρίων αξιολόγησης).

Το πρώτο εξ αυτών, το μοντέλο XGB-1, εκπαιδεύτηκε λαμβάνοντας ως χαρακτηριστικά γνωρίσματα την ημερομηνία της κάθε παρατήρησης της χρονοσειράς καθώς επίσης και την

τιμή της. Το XGB-2 αναπτύχθηκε για να φέρει καλύτερες προβλέψεις. Εκπαιδεύτηκε με τα ίδια δεδομένα αλλά είχε επιπλέον χαρακτηριστικά γνωρίσματα σε σχέση με το XGB-1. Τέτοια ήταν για παράδειγμα η ημέρα της εβδομάδας που αντιστοιχούσε στην κάθε ημερομηνία δειγματοληψίας ή η τιμή της παρατήρησης πριν από 3 χρονικές υστερήσεις. Είχαν εξαχθεί από τα υπάρχοντα δεδομένα επιπλέον γνωρίσματα χρόνου και υστερήσεων. Μετά από πολλούς συνδυασμούς επιλέχθηκαν όσα παρατηρήθηκαν να οδηγούν το μοντέλο σε μικρότερο απόλυτο μέσο σφάλμα στο στάδιο της εκπαίδευσης. Το τελευταίο μοντέλο, XGB-3 αναπτύχθηκε για να αντιμετωπιστεί η υπερπροσαρμογή που εμφάνιζε το XGB-2. Χρησιμοποιούσε τα ίδια χαρακτηριστικά γνωρίσματα, τον ίδιο όγκο δεδομένων κατά την εκπαίδευση του αλλά υψηλότερο ρυθμό εκπαίδευσης και λιγότερα δέντρα.

Από τη διαδικασία πρόβλεψης της χρονοσειράς με μοντέλα XGBoost βγήκαν τα εξής συμπεράσματα :

- Το μοντέλο XGB-1 έφερε φτωχά αποτελέσματα στο στάδιο εκπαίδευσης όσο και στη διαδικασία πρόβλεψης εκτός δείγματος. Αυτή η απόδοση το κατάρταζε τελευταίο μεταξύ των όσων είχαν χρησιμοποιηθεί έως τώρα. Μάλιστα η πρόβλεψη του ήταν χειρότερη κατά 26% από τη χειρότερη έως τώρα.
- Η διαδικασία επιλογής των κατάλληλων χαρακτηριστικών για το XGB-2 αν και χρονοβόρα απέφερε καρπούς. Το μοντέλο πλέον προσαρμοζόταν στα δεδομένα εκπαίδευσης με πολύ μεγάλη ακρίβεια αφού το σφάλμα του ανερχόταν στις 10 kWh περίπου. Η τροφοδότηση του μοντέλου με πιο χαρακτηριστικές πληροφορίες στο στάδιο εκπαίδευσης βελτιώνει την απόδοση του κατά 86% σε σχέση με πριν.
- Στο στάδιο της πρόβλεψης το XGB-2 έφερε προβλέψεις καλύτερες από όλα τα μοντέλα. Μπορεί σε μέγεθος να μην εκτιμούσε ακριβώς την μέση ποσότητα που θα παραγόταν αλλά εκτιμούσε άριστα εάν από ημέρα σε ημέρα θα υπήρχε αύξηση ή μείωση της παραγωγής καθώς και το μέγεθος αυτής. Το γεγονός ότι το μοντέλο υπερεκτιμούσε σε έναν βαθμό οδηγούσε στην όποια απόκλιση. Η απόκλιση αυτή με βάση το κριτήριο ρίζας μέσου τετραγωνικού σφάλματος ανερχόταν στις 445 kWh περίπου.
- Το XGB-3 που αναπτύχθηκε για να προσφέρει καλύτερες προβλέψεις μειώνοντας τον βαθμό υπερεκτίμησης του XGB-2 πέτυχε το σκοπό του. Στις 8 από τις 10 περιπτώσεις που έπρεπε να προβλέψει υπερεκτιμούσε πολύ λιγότερο από το XGB-2 και απέκλινε ελάχιστα. Χαρακτηριστικά αναφέρεται ότι βελτιώνει την προβλεπτική του ικανότητα κατά 45% σε αυτές τις περιπτώσεις αφού αποκλίνει περίπου για 200kWh όταν το XGB-2 αποκλίνει περίπου για 375 kWh.
- Το XGB-3 συνολικά εμφάνιζε μεγαλύτερο σφάλμα από το XGB-2 με βάση το κριτήριο ρίζας μέσου τετραγωνικού σφάλματος επειδή ήταν πολύ άστοχο στις εναπομείναντες δύο περιπτώσεις. Χαρακτηριστικά αναφέρεται ότι σε αυτές τις δύο η απόκλιση ανέρχεται των 1100 kWh.

Bibliography

- [1]
- [2] Energy sector overview: Green revolution. . Accessed: 2022-05-24.
- [3] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. Springer, 2002.
- [4] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer Science & Business Media, 2009.
- [5] Tony Burton, Nick Jenkins, David Sharpe, and Ervin Bossanyi. Wind energy handbook. 2011.
- [6] Christopher Chatfield. *The analysis of time series: theory and practice*. Springer, 2013.
- [7] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [8] Melissa Denchak. Fossil fuels: the dirty facts. NRDC, 2018. <https://www.nrdc.org/stories/fossil-fuels-dirty-facts>.
- [9] Melissa Denchak and Jeff Turrentine. Global climate change: What you need to know. NRDC, 2021. <https://www.nrdc.org/stories/global-climate-change-what-you-need-know>.
- [10] Marcel Dettling. Applied time series analysis. *ETH Scripts*, 2013.
- [11] Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami, and Abdeslam Lachhab. Forecasting of demand using arima model. *International Journal of Engineering Business Management*, 10:1847979018808673, 2018.
- [12] Andrea Ferrario and Roger Hämmmerli. On boosting: Theory and applications. *Available at SSRN 3402687*, 2019.
- [13] Max Roser Hannah Ritchie and Pablo Rosado. CO₂ and greenhouse gas emissions. *Our World in Data*, 2020.

- [14] Max Roser Hannah Ritchie and Pablo Rosado. Energy. *Our World in Data*, 2020. <https://ourworldindata.org/energy>.
- [15] Alexander Kalmikov. Wind power fundamentals. In *Wind Energy Engineering*, pages 17–24. Elsevier, 2017.
- [16] Helmut Lütkepohl and Markus Krätzig. *Applied time series econometrics*. Cambridge university press, 2004.
- [17] Spyros Makridakis and Michele Hibon. Arma models and the box-jenkins methodology. *Journal of forecasting*, 16(3):147–163, 1997.
- [18] Sathyajith Mathew. *Wind energy: fundamentals, resource analysis and economics*, volume 1. Springer, 2006.
- [19] Simon P Neill and M Reza Hashemi. *Fundamentals of ocean renewable energy: generating electricity from the sea*. Academic Press, 2018.
- [20] Avishek Pal and PKS Prakash. *Practical time series analysis: master time series data processing, visualization, and modeling using python*. Packt Publishing Ltd, 2017.
- [21] Robin Pomeroy. Planet versus profit. World Economic Forum, 2021. <https://www.weforum.org/agenda/2021/12/stakeholder-capitalism-profit-planet-nicholas-stern-risalat-khan/>.
- [22] Selva Prabhakaran. Augmented dickey fuller test (adf test) – must read guide, November 2019. url: <https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test>.
- [23] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer, 2000.
- [24] Vaclav Smil. World history and energy. *Encyclopedia of energy*, 6:549–561, 2004.
- [25] Linqiup Team. What are the disadvantages of fossil fuels?. Linqiup, 2021. <https://www.linqiup.com/blog/disadvantages-of-fossil-fuels>.
- [26] Δημήτρης Κουγιουμτζής. Ανάλυση Χρονοσειρών. Σημειώσεις μεταπτυχιακού μαθήματος του ΑΠΘ, 2014.
- [27] Κατσούλης Δημήτρης. Μαθήματα Μετεωρολογίας και Κλιματολογίας. 2015. <https://olympias.lib.uoi.gr/jspui/handle/123456789/26966>.
- [28] Κοκολάκης Γεώργιος. Σημειώσεις Ανάλυσης Χρονοσειρών. Τομέας Μαθηματικών, ΣΕΜΦΕ, ΕΜΠ, Αθήνα, 2004.
- [29] Μελάς, Δημήτρης Μπάης, Αλκιβιάδης Μπάης, Δημήτρης Μπάλης. Ατμοσφαιρική τεχνολογία. Kallipos, Anoiktes Akadhmaikes Ekdoseis, 2015.

-
- [30] Μυρτώ Μπότσιου. Πλανήτης Γη σε αναζήτηση ενέργειας, μέρος 1ο. 2020. <http://www.eef.edu.gr/el/arthra/planitis-gi-se-anazitisi-energeias-meros-1o/>.
- [31] Στάθης, Δημήτρης. Μαθήματα Δασικής Μετεωρολογίας και κλιματολογίας. Technical report, Aristotle University of Thessaloniki, 2015.
- [32] Χαλδούπης Χρήστος. Δυναμική της Ατμόσφαιρας. 2016. <http://hdl.handle.net/11419/3278>.
- [33] Wei Tong. *Wind power generation and wind turbine design*. WIT press, 2010.
- [34] xgboost developers. Xgboost documentation, 2022. Accessed: 08/07/2023.
- [35] xgboost developers. xgboost release 2.1.0-dev, 2023. Accessed: 31/10/2023.
- [36] Sercan Yeşilöz. Electricity production forecasting (arima), June 2021. Version 34 with url: <https://www.kaggle.com/code/sercanyesiloz/electricity-production-forecasting-arima/notebook>.

