
Στατιστική Ανάλυση και Μοντελοποίηση Χρονοσειρών Αιολικού Δυναμικού

Διπλωματική Εργασία

Ιωάννα Ζαχαρίου



Πολυτεχνείο Κρήτης

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών

Υπολογιστών

Χανιά, Κρήτη, Ελλάδα

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

Τριμελής Επιτροπή:

Διονύσιος Χριστόπουλος, Καθηγητής

Σταυρακάκης Γεώργιος, Καθηγητής, Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Κουτρούλης Ευύχιος, Καθηγητής, Σχολή Ηλεκτρολόγων Μηχανικών και
Μηχανικών Υπολογιστών

Αφιέρωση

Η παρούσα διπλωματική εργασία είναι αφιερωμένη στους γονείς μου!

Ευχαριστίες

Με την παρούσα διπλωματική εργασία θα ήθελα να ευχαριστήσω τον κύριο Χριστόπουλο Διονύσιο, Καθηγητή του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών και επιβλέποντα της διπλωματικής μου εργασίας, τον ο ποίο είχα την τιμή να συνεργαστώ τόσο στα πλαίσια της διπλωματικής μου, όσο και στα πλαίσια του μαθήματος Στοχαστικές Διαδικασίες και Ανάλυση Χρονοσειρών. Επίσης, θα ήθελα να τον ευχαριστήσω για την πολύτιμη καθοδήγηση που μου προσέφερε σε θέματα σύνταξης και δομής της διπλωματικής μου εργασίας καθώς επίσης και για τη γνώση που μου παρείχε. Επιπλέον θα ήθελα να ευχαριστήσω την κ.Σοφία Τσακιρίδου για την πολυτιμή βοήθεια και την στήριξη που μου έδωσε στο τέλος της διπλωματικής μου εργασίας. Στο σημείο αυτό θα ήθελα να ευχαριστήσω την οικογένεια μου, τον πατέρα μου Γιάννη, την μητέρα μου Ναυσικά και τα αδέλφια μου Μιχάλη, Σταύρο και Στάλω. Τους ευχαριστώ για τη στήριξη που μου παρείχαν κατά τη διάρκεια των σπουδών μου, δίνοντας μου και τη δυνατότητα να σπουδάσω αυτό που ήθελα και που πάντα ήταν δίπλα στο πλευρό μου. Τέλος, θα ήθελα να ευχαριστήσω το αγόρι μου Νίκο και τους φίλους μου, για τη συνεχή ενθάρρυνση, κατανόηση και στήριξη σε όλη τη διαδρομή των σπουδών μου.

Περίληψη

Η αιολική ενέργεια είναι μια πηγή βιώσιμης ενέργειας που αναπτύσσεται πολύ γρήγορα σε όλο τον κόσμο. Η πρόβλεψη της ταχύτητας του ανέμου αποτελεί παγκόσμιο πρόβλημα και κρίσιμο ζήτημα για τα συστήματα παραγωγής αιολικής ενέργειας, καθώς επηρεάζει σε μεγάλο βαθμό τον προγραμματισμό της παραγωγής και τον δυναμικό έλεγχο των ανεμογεννητριών. Οι μέθοδοι ανάλυσης χρονοσειρών χρησιμοποιούνται συχνά για την ανάλυση της μεταβλητότητας και την πρόβλεψη της ταχύτητας του ανέμου. Οι στατιστικές μέθοδοι που χρησιμοποιούνται στην παρούσα εργασία είναι βασισμένες στην θεωρία των χρονοσειρών, και στοχεύουν στην πρόβλεψη μελλοντικών τιμών ταχύτητας ανέμου.

Πιο συγκεκριμένα, η παρούσα διπλωματική εργασία ασχολείται με την εφαρμογή του μοντέλου ARIMA (Autoregressive Integrated Moving Average) στην βραχυπρόθεσμη πρόβλεψη της ταχύτητας του ανέμου. Η βραχυπρόθεσμη πρόβλεψη ορίζεται για χρονικό ορίζοντα από λίγα λεπτά έως μερικές ώρες. Τα δεδομένα που αναλύονται στην διπλωματική εργασία είναι ταχύτητες ανέμου για τις περιοχές του Τροόδους και του Αγρού στην Κύπρο, για την χρονική περίοδο μεταξύ 1ης Ιανουαρίου και 31ης Δεκεμβρίου 2020. Οι χρονοσειρές ταχύτητας ανέμου που χρησιμοποιήθηκαν είναι δεδομένα επανανάλυσης (reanalysis) που προέρχονται από την βάση δεδομένων SoDa. Η βάση αυτή παρέχει πρόσβαση σε ιστορικά και πραγματικού χρόνου μετεωρολογικά δεδομένα καθώς και σε προβλέψεις.

Η πρόβλεψη πραγματοποιήθηκε με το μοντέλο ARIMA βάσει χρονοσειρών με ωριαίο βήμα από δύο περιοχές της Κύπρου. Η ανάλυση

έγινε για τα ετήσια δεδομένα καθώς και για τα δεδομένα χειμερινής και καλοκαιρινής περιόδου ξεχωριστά. Αρχικά, με την βοήθεια των διαγραμμάτων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης όσο και των κριτηρίων επιλογής AIC και BIC επιλέχθηκε το καταλληλότερο μοντέλο ARIMA ανά περιοχή και ανά χρονική περίοδο. Η ανάλυση της επίδοσης των μοντέλων σχετικά με την πρόβλεψη βασίζεται στην διαδικασία της διασταυρωτικής επικύρωσης (cross validation) με την βοήθεια των στατιστικών μέτρων της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Square Error - RMSE) και του μέσου απόλυτου σφάλματος (Mean Absolute Error - MAE). Από την ανάλυση των προβλέψεων της ταχύτητας ανέμου για τις τρεις χρονοσειρές (ετήσια, καλοκαιρινής περιόδου, χειμερινής περιόδου) καταλήγουμε στο συμπέρασμα ότι η ακρίβεια της πρόβλεψης μειώνεται όσο αυξάνεται ο χρονικός ορίζοντας (βήμα) πρόβλεψης. Συγκεκριμένα, η πρόβλεψη και στις τρεις περιπτώσεις για χρονικό ορίζοντα των δύο ωρών ήταν αρκετά αξιόπιστη γιατί ο λόγος του RMSE προς την μέση τιμή δεν υπερβαίνει το 10%. Επιπλέον, η περίοδος του καλοκαιριού έδωσε την δυνατότητα καλύτερης εκτίμησης από την χειμερινή περίοδο. Αυτό αποδίδεται στο γεγονός ότι κατά τους χειμερινούς μήνες ο καιρός εμφανίζει μεγαλύτερη μεταβλητότητα σε σχέση με την καλοκαιρινή περίοδο κατά την οποία υπάρχει μεγαλύτερη σταθερότητα. Αυτές οι παρατηρήσεις συμφωνούν με αποτελέσματα από άλλες μελέτες τα οποία δείχνουν ότι το μοντέλο ARIMA είναι ικανοποιητικό για βραχυπρόθεσμη πρόβλεψη, με χρονικό ορίζοντα από 2 έως 6 ώρες.

Λέξεις κλειδιά

Ανάλυση χρονοσειρών, πρόβλεψη χρονοσειρών, μοντέλα χρονοσειρών, ARIMA, ταχύτητα ανέμου.

Abstract

Wind energy is a source of sustainable energy that is growing very fast around the world. Wind speed prediction is a global problem and a critical issue for wind power generation systems, as it greatly affects production scheduling and the dynamic control of wind turbines. Time series analysis methods are often used to analyze the variability and forecast wind speed. The statistical methods used in this thesis are based on time series theory and aim to predict future wind speed values.

This thesis addresses a short-term wind speed forecasting ARIMA (Autoregressive Integrated Moving Average) model. The short-term forecast is defined for a time horizon of a few minutes to a few hours. The data analysed in the thesis are wind speed data for the Troodos and Agros regions in Cyprus, for the time period between January 1st and December 31st, 2020. The wind speed time series used are reanalysis data derived from the SoDa database. This database provides access to historical and real-time meteorological data as well as forecasts.

The prediction was carried out with the ARIMA model based on time series with hourly step from two areas of Cyprus. The analysis was performed for the annual data as well as for the winter and summer period data separately. Initially, with the help of the autocorrelation and partial autocorrelation plots as well as of the AIC and BIC selection criteria, the most appropriate ARIMA model per region and per time period was selected. The performance analysis of the prediction models is based on cross-validation by use of the Root Mean Square Error (RMSE) and of the Mean Absolute Error (MAE) statistical measures. From the analysis of

the wind speed forecasts for the three time series (annual, summer period, winter period) we conclude that the accuracy of the forecast decreases as the forecast time horizon (step) increases. Specifically, the prediction in all three cases was quite reliable for a time horizon of two hours because the RMSE to the mean ratio does not exceed 10%. In addition, the summer period allowed for a better prediction than the winter period. This can be interpreted as the weather is more variable in the winter months compared to the summer period. These observations are in agreement with results from other studies which show that the ARIMA model is satisfactory for short-term forecasting, with a time horizon of 2 to 6 hours.

Keywords: Time series analysis, time series forecasting, time series models, ARIMA, wind speed.

Περιεχόμενα

Περίληψη	iii
Abstract	v
Εισαγωγή	1
1 Βασικές Έννοιες Χρονοσειρών	5
1.1 Ορισμός χρονοσειράς	5
1.2 Βασικά χαρακτηριστικά χρονοσειρών	6
1.2.1 Στασιμότητα	7
1.2.2 Ακραίες τιμές (Outliers)	9
1.2.3 Τάση και περιοδικότητα	10
1.3 Στατιστικά μεγέθη χρονοσειράς	15
1.3.1 Μέση τιμή	16
1.3.2 Αυτοσυνδιακύμανση	16
1.3.3 Αυτοσυσχέτιση	16
2 Κατανομές Χρονοσειρών Ταχύτητας Ανέμου	19
2.1 Κανονική κατανομή	19
2.2 Γενικευμένη κατανομή ακραίων τιμών	21
2.3 Κατανομή Weibull	23

2.4	Λογαριθμο-λογιστική κατανομή (Log-logistic)	24
2.5	Κατανομή Γάμμα	27
2.6	Λογαριθμοκανονική κατανομή (Lognormal)	27
3	Μοντέλα Χρονοσειρών	29
3.1	Γραμμικά μοντέλα πρόβλεψης χρονοσειρών	29
3.2	Αυτοπαλινδρούμενο μοντέλο (AR)	30
3.2.1	Αυτοπαλινδρούμενη διαδικασία πρώτης τάξης	31
3.3	Μοντέλα κινούμενου μέσου MA	32
3.3.1	Στοχαστική διαδικασία πρώτης τάξης MA(1)	32
3.4	Αυτοπαλινδρούμενη διαδικασία κινούμενου μέσου (ARMA)	34
3.5	Ολοκληρωμένο αυτοπαλινδρούμενο μοντέλο κινούμενου μέ- σου (ARIMA)	35
3.6	Ολοκληρωμένο αυτοπαλινδρομούμενο μοντέλο κινούμενου μέσου με εποχικότητα (SARIMA)	36
4	Πρόβλεψη Χρονοσειρών	39
4.1	Βασικά στάδια στην διαδικασία πρόβλεψης	39
4.2	Πρόβλεψη χρονοσειρών με την χρήση του μοντέλου ARIMA	41
4.2.1	Πρόβλεψη χρονοσειρών με την βοήθεια της διαδικα- σίας Box-Jenkins	41
4.2.2	Πρόβλεψη με αυτοπαλινδρούμενα μοντέλα AR(p)	44
4.2.3	Πρόβλεψη με μοντέλο κινούμενου μέσου MA(q)	45
4.2.4	Πρόβλεψη με αυτοπαλινδρούμενα μοντέλα κινούμε- νου μέσου ARMA	46
4.2.5	Πρόβλεψη μη στάσιμων χρονοσειρών με γραμμικά μοντέλα	46

4.2.6	Ανάλυση ακρίβειας μοντέλου για την εκτίμηση χρο- νοσειρών	48
5	Ανάλυση Χρονοσειρών Ταχύτητας Ανέμου	51
5.1	Περιοχή μελέτης και περιγραφή των δεδομένων	51
5.2	Στατιστική ανάλυση δεδομένων	53
5.2.1	Περιοχή του Τροόδους	53
5.2.2	Περιοχή του Αγρού	59
5.2.3	Χρονοσειρές με ωριαίο βήμα	65
5.3	Προσδιορισμός βέλτιστου πιθανοτικού μοντέλου	79
5.3.1	Αποτελέσματα	79
5.4	Πρόβλεψη με την βοήθεια του μοντέλου ARIMA	80
5.4.1	Χρονοσειρά με ωριαίο βήμα για το έτος 2020	81
5.4.2	Χρονοσειρά με ωριαίο βήμα για την καλοκαιρινή περίοδο για την περιοχή Τροόδος	86
5.4.3	Χρονοσειρά με ωριαίο βήμα για την χειμερινή πε- ρίοδο για την περιοχή Τροόδος	90
6	Συμπεράσματα	95
	Βιβλιογραφία	98

Κατάλογος Σχημάτων

1.1	Γραφική παράσταση χρονοσειράς που παρουσιάζει μια ακραία τιμή	9
1.2	Γραφική παράσταση μιας χρονοσειράς που παρουσιάζει γραμμική ανοδική τάση	11
2.1	Τυποποιημένη κανονική κατανομή	20
2.2	Γενικευμένη κατανομή ακραίων τιμών για διάφορες τιμές του ξ	22
2.3	Συνάρτηση πυκνότητας πιθανότητας της Λογαριθμο-λογιστικής κατανομής για παράμετρο κλίμακας $\alpha = 1$	26
2.4	Συνάρτηση διακινδύνευσης της Λογαριθμο-λογιστικής κατανομής, για παράμετρο κλίμακας $\alpha = 1$ και για διάφορες τιμές της παραμέτρου β	26
5.1	Ανάλυση δεδομένων (ιστογράμματα) μηνιαίων χρονοσειρών με δεκάλεπτο βήμα για την περιοχή του Τροόδους για το έτος 2020.	54
5.2	Ανάλυση κανονικής κατανομής (Ιστόγραμμα, διάγραμμα κανονικής πιθανότητας) ετήσιας χρονοσειράς ανά 10 λεπτά για την περιοχή του Τροόδους.	55

5.3	Ανάλυση δεδομένων (ιστογράμματα, διαγράμματα κανονικής πιθανότητας) ετήσιας χρονοσειράς με δεκάλεπτο βήμα για την καλοκαιρινή και χειμερινή περίοδο για την περιοχή του Τροόδους.	58
5.4	Ανάλυση δεδομένων (ιστογράμματα) μηνιαίων χρονοσειρών με δεκάλεπτο βήμα για την περιοχή του Αγρού για το έτος 2020.	60
5.5	Ανάλυση κανονικής κατανομής (Ιστόγραμμα, Διάγραμμα Κανονικής Πιθανότητας) ετήσιας χρονοσειράς ανά 10 λεπτά για την περιοχή του Αγρού.	61
5.6	Ανάλυση κανονικής κατανομής (Ιστογράμματα, Διαγράμματα κανονικής πιθανότητας) χρονοσειράς ανά 10 λεπτά για την καλοκαιρινή και χειμερινή περίοδο για περιοχή του Αγρού.	63
5.7	Ανάλυση Κανονικής Κατανομής (Ιστογράμματα) για κάθε μήνα με βήμα ανά ώρα για την περιοχή του Τροόδους για το έτος 2020.	66
5.8	Ανάλυση κανονικής κατανομής (Ιστόγραμμα, διάγραμμα κανονικής πιθανότητας) ετήσιας χρονοσειράς ανά ώρα για την περιοχή του Τροόδους.	68
5.9	Ανάλυση κανονικής κατανομής (Ιστογράμματα, Διαγράμματα Κανονικής Πιθανότητας) χρονοσειράς ανά ώρα για την καλοκαιρινή και χειμερινή περίοδο για περιοχή του Τροόδους.	70
5.10	Ανάλυση Κανονικής Κατανομής (Ιστογράμματα) για κάθε μήνα με βήμα ανά ώρα για την περιοχή του Αγρού για το έτος 2020.	73
5.11	Ανάλυση κανονικής κατανομής (Ιστόγραμμα, διάγραμμα κανονικής πιθανότητας) ετήσιας χρονοσειράς ταχύτητας ανέμου με χρονικό βήμα ανά ώρα για την περιοχή του Αγρού.	74

5.12 Ανάλυση κανονικής κατανομής (Ιστογράμματα, διαγράμματα κανονικής πιθανότητας) χρονοσειράς ταχύτητας ανέμου με χρονικό βήμα ανά ώρα για την καλοκαιρινή και χειμερινή περίοδο για την περιοχή του Αγρού.	76
5.13 Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για τα ωριαία δεδομένα ταχύτητας ανέμου της περιοχής του Τροόδους για το έτος 2020.	81
5.14 Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για τα διαφορισμένα δεδομένα ταχύτητας ανέμου.	82
5.15 Κατάλοιπα μοντέλου.	84
5.16 Ιστόγραμμα καταλοίπων μοντέλου ARIMA(3,1,12).	84
5.17 Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου και πρόβλεψης με το μοντέλο ARIMA(3,1,12).	86
5.18 Διάγραμμα χρονοσειράς με ωριαίο βήμα για την καλοκαιρινή περίοδο της περιοχής Τροόδους.	87
5.19 Διάγραμμα αυτοσυσχέτισης για την καλοκαιρινή περίοδο.	88
5.20 Διάγραμμα αυτοσυσχέτισης για ωριαία χρονοσειρά για την καλοκαιρινή περίοδο μετά την απαλοιφή των αυτοσυσχετίσεων.	88
5.21 Διάγραμμα μερικής αυτοσυσχέτισης για ωριαία χρονοσειρά για την καλοκαιρινή περίοδο μετά την απαλοιφή των αυτοσυσχετίσεων.	89
5.22 Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου και πρόβλεψης για την καλοκαιρινή περίοδο με το μοντέλο ARIMA(3,1,8).	90
5.23 Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου για την χειμερινή περίοδο για την περιοχή Τροόδους.	91
5.24 Διάγραμμα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για την χειμερινή περίοδο.	92

- 5.25 Διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης
των υπολοίπων από το βέλτιστο μοντέλο ARIMA. 93
- 5.26 Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου και πρό-
βλεψης για την χειμερινή περίοδο με το μοντέλο ARIMA(3,1,10). 94

Κατάλογος Πινάκων

3.1	Εκτίμηση της τάξης των AR και MA με την βοήθεια των συναρτήσεων ACF και PACF	34
5.1	Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά ταχύτητας ανέμου με βήμα δειγματοληψίας 10 λεπτά για την περιοχή του Τροόδους.	56
5.2	Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Τροόδους	57
5.3	Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την χειμερινή περίοδο για την περιοχή του Τροόδους	59
5.4	Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά με δεκάλεπτο βήμα για την περιοχή του Αγρού. .	62
5.5	Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Αγρού.	64
5.6	Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την χειμερινή περίοδο για την περιοχή του Αγρού.	65
5.7	Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά με ωριαίο βήμα για την περιοχή του Τροόδους. .	69

5.8 Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Τροόδους.	71
5.9 Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την χειμερινή περίοδο για την περιοχή του Τροόδους.	72
5.10 Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά ταχύτητας ανέμου με ωριαίο βήμα για την περιοχή του Αγρού.	75
5.11 Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Αγρού.	77
5.12 Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την χειμερινή περίοδο για την περιοχή του Αγρού.	77
5.13 Πίνακας με τις τρεις επικρατέστερες κατανομές για την ετήσια χρονοσειρά με βήμα ανά μία ώρα για την περιοχή του Αγρού.	80
5.14 Πίνακας με τις τρεις επικρατέστερες κατανομές για την ετήσια χρονοσειρά βήμα ανά μία ώρα για την περιοχή του Τροόδους.	80
5.15 Τιμές κριτηρίων AIC και BIC για διαφορετικά μοντέλα ARIMA.	83
5.16 Πίνακας με τις τιμές των μέτρων RMSE και MAE για τα ωριαία βήματα 2,6,12,24.	85
5.17 Πίνακας με τις τιμές των μέτρων RMSE και MAE για τα ωριαία βήματα 2,6,12,24 για την καλοκαιρινή περίοδο. . .	89
5.18 Πίνακας με τις τιμές των μέτρων RMSE και MAE για τα ωριαία βήματα 2,6,12,24 για την χειμερινή περίοδο. . . .	94

Εισαγωγή

Η τιμή της ορυκτής ενέργειας συνεχίζει να αυξάνεται με τη μαζική κατανάλωση ορυκτής ενέργειας και την εντατικοποίηση της παγκόσμιας ενεργειακής κρίσης, οδηγώντας σε μεγαλύτερη προβολή των περιβαλλοντικών προβλημάτων [5]. Η ανάπτυξη και η αξιοποίηση ανανεώσιμων πηγών ενέργειας έχουν καταστεί το κλειδί για την επίλυση αυτού του προβλήματος. Ως αποτέλεσμα, οι χώρες σε όλο τον κόσμο έχουν επενδύσει σημαντικούς ανθρώπινους και υλικούς πόρους στην έρευνα για την παραγωγή ανανεώσιμης ενέργειας. Οι ανανεώσιμες πηγές ενέργειας αναφέρονται στην ανάπτυξη και αξιοποίηση μη συμβατικών πηγών ενέργειας που βασίζονται σε νέες τεχνολογίες. Οι ανανεώσιμες πηγές ενέργειας περιλαμβάνουν την αιολική, ηλιακή, παλιρροϊκή, και ωκεάνια ενέργεια, την ενέργεια από βιομάζα καθώς και άλλες ανανεώσιμες πηγές για την παραγωγή ηλεκτρικής ενέργειας και την κάλυψη της αυξανόμενης ζήτησής της.

Η αιολική ενέργεια έχει σημαντικό ρόλο στις περισσότερες χώρες στον κλάδο των ανανεώσιμων πηγών ενέργειας λόγω της τεράστιας εμπορικής της αξίας και των περιβαλλοντικών της πλεονεκτημάτων που την καθιστούν την ταχύτερα αναπτυσσόμενη και πιο ώριμη μορφή ενέργειας μεταξύ των ανανεώσιμων πηγών. Εκτιμάται ότι η εκμεταλλεύσιμη αποθηκευτική ικανότητα της αιολικής ενέργειας στη γη μπορεί να φθάσει τα 2×10^7 MW και το 1% της αποθηκευτικής ικανότητας της αιολικής ενέργειας μπορεί να αξιοποιηθεί πλήρως για την κάλυψη της τρέχουσας ενεργειακής ζήτησης [5, 10].

Οι προβλέψεις της ταχύτητας του ανέμου θα μειώσουν το λειτουργικό

κόστος, θα αυξήσουν την ισχύ του συστήματος ηλεκτρικής ενέργειας και θα ενισχύσουν επίσης την ικανότητα παραγωγής ηλεκτρικής ενέργειας και την αποδοτικότητα των ανεμογεννητριών. Τα περιβαλλοντικά χαρακτηριστικά του ανέμου, ωστόσο, καταστούν δύσκολη την πρόβλεψη της ταχύτητας του ανέμου με μικρό σφάλμα. Σε σύγκριση με τον άνεμο, η συμβατική ισχύς είναι ευκολότερο να προβλεφθεί. Η παραγωγή αιολικής ενέργειας είναι απρόβλεπτη, καθώς ρυθμίζεται από τον καιρό, την κατεύθυνση της ταχύτητας του ανέμου, τις διακυμάνσεις της εποχικής θερμοκρασίας και της υγρασίας του αέρα. Οι απρόβλεπτες ταχύτητες του ανέμου επηρεάζουν την ποσότητα της παραγόμενης ηλεκτρικής ενέργειας και συνεπώς δημιουργούν προκλήσεις για την εξισορρόπηση της ισχύος μεταξύ προσφοράς και ζήτησης.

Μια από τις λύσεις για να αντιμετωπισθεί η ανεπάρκεια της αιολικής ενέργειας είναι η πρόβλεψη της ταχύτητας του ανέμου. Υπάρχουν διάφορες προσεγγίσεις πρόβλεψης της ταχύτητας του ανέμου που περιλαμβάνουν φυσικές, μαθηματικές και προσεγγίσεις τεχνητής νοημοσύνης (Artificial Intelligence, AI). Τα μοντέλα πρόβλεψης προσφέρουν διαφορετικά αποτελέσματα και διαφορετικό βαθμό ακρίβειας. Οι στατιστικές μέθοδοι είναι κοινές τεχνικές ανάλυσης των δεδομένων σε χρονοσειρές.

Οι πρώτες μελέτες σχετικά με τα προβλήματα πρόβλεψης της ταχύτητας του ανέμου για τις επόμενες ώρες και ημέρες εμφανίστηκαν στις αρχές της δεκαετίας του 1980. Οι μέθοδοι αυτές παρουσιάστηκαν από τους συγγραφείς στο [3], οι οποίοι πρότειναν να δοθούν πληροφορίες σχετικά με την αβεβαιότητα των προβλέψεων. Σήμερα, η βιβλιογραφία που σχετίζεται με την πρόβλεψη του ανέμου και την παραγωγή αιολικής ενέργειας επεκτείνεται πολύ γρήγορα. Ένα μεγάλο μέρος των εξελίξεων στις μεθόδους πρόβλεψης είναι αφιερωμένο στις μεθόδους μοντελοποίησης χρονοσειρών, όπως Autoregression (AR), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) και άλλες. Οι συγγραφείς στο [8] χρησιμοποίησαν ένα μοντέλο ARMA για την πρόβλεψη της ωριαίας μέσης ταχύτητας του ανέμου και συνέκριναν τα αποτελέσματα με το μοντέλο εμμονής (persistence model), για να καταλήξουν στο συμπέρασμα ότι τα σφάλματα του μοντέλου ARMA ήταν

μικρότερα από αυτά του μοντέλου εμμονής. Στο [4], οι συγγραφείς συνέκριναν τις μεθόδους ARIMA και Artificial Neural Network (ANN) που εφαρμόστηκαν σε μια χρονοσειρά μετρήσεων της ταχύτητας του ανέμου. Διαπίστωσαν ότι τα μοντέλα ARIMA παρουσίασαν καλύτερη ευαισθησία στην προσαρμογή και την πρόβλεψη της ταχύτητας του ανέμου. Με βάση το ARIMA και τον βελτιωμένο αλγόριθμο φίλτρου Kalman, οι συγγραφείς πρότειναν ένα νέο μοντέλο πρόβλεψης της ταχύτητας του ανέμου [9].

Συγκεκριμένα, η παρούσα διπλωματική εργασία ασχολείται με την εφαρμογή του μοντέλου ARIMA στην βραχυπρόθεσμη πρόβλεψη της ταχύτητας του ανέμου. Η βραχυπρόθεσμη πρόβλεψη ορίζεται για χρονικό ορίζοντα από λίγα λεπτά έως μερικές ώρες. Τα δεδομένα που αναλύονται στην διπλωματική εργασία είναι ταχύτητες ανέμου για τις περιοχές του Τροόδους και του Αγρού στην Κύπρο, για την χρονική περίοδο μεταξύ 1ης Ιανουαρίου και 31ης Δεκεμβρίου 2020. Οι χρονοσειρές ταχύτητας ανέμου που χρησιμοποιήθηκαν είναι δεδομένα επανανάλυσης (reanalysis) που προέρχονται από την βάση δεδομένων SODA-PRO. Η βάση αυτή παρέχει πρόσβαση σε ιστορικά και πραγματικού χρόνου μετεωρολογικά δεδομένα καθώς και σε προβλέψεις.

Στο πρώτο κεφάλαιο της εργασίας περιγράφονται βασικά στατιστικά μέτρα των χρονοσειρών. Στο δεύτερο κεφάλαιο αναλύεται η προσαρμογή διαφόρων κατανομών πιθανότητας στα δεδομένα. Το τρίτο κεφάλαιο περιγράφει το ολοκληρωμένο αυτοπαλινδρούμενο μοντέλο κινούμενου μέσου (ARIMA), που χρησιμοποιήθηκε για την ανάλυση των δεδομένων. Στο τέταρτο κεφάλαιο παρουσιάζονται τα βασικά στάδια της διαδικασίας Box-Jenkins για την μοντελοποίηση και πρόβλεψη χρονοσειρών. Τέλος, στο πέμπτο κεφάλαιο γίνεται η στατιστική ανάλυση των δεδομένων και η πρόβλεψη της ταχύτητας του ανέμου. Η στατιστική ανάλυση εφαρμόστηκε σε δύο χρονοσειρές που αποτελούνται από 52704 τιμές ταχύτητας ανέμου ανά δεκάλεπτο από την 1η Ιανουαρίου του 2020 έως την 31η Δεκεμβρίου του 2020 στις περιοχές Τροόδους και Αγρού. Επίσης χρησιμοποιήθηκαν χρονοσειρές για τις ίδιες περιοχές και το ίδιο χρονικό διάστημα με ωριαίο βήμα. Οι χρονοσειρές αυτές περιέχουν 8784 τιμές έ-

καστη. Στην συνέχεια έγινε διαχωρισμός όλων των ανωτέρω χρονοσειρών σε δύο χρονικές περιόδους, την καλοκαιρινή περίοδο από 10 Μαΐου έως 10 Οκτωβρίου 2020, και την χειμερινή περίοδο από 1 Δεκεμβρίου 2019 έως 29 Φεβρουαρίου 2020, και πραγματοποιήθηκε αντίστοιχη στατιστική ανάλυση.

Κεφάλαιο 1

Βασικές Έννοιες Χρονοσειρών

1.1 Ορισμός χρονοσειράς

Ως χρονοσειρά ή χρονολογική σειρά (time series) ορίζεται ένα σύνολο δεδομένων, τα οποία συλλέγονται διαχρονικά και εκφράζουν την εξέλιξη των τιμών μιας μεταβλητής κατά τη διάρκεια διαδοχικών χρονικών περιόδων. Ειδικότερα, η χρονοσειρά αποτελείται από ένα σύνολο παρατηρήσεων μιας μεταβλητής, οι τιμές της οποίας είναι ιεραρχημένες με βάση τη χρονική περίοδο στην οποία αναφέρονται, όπως για παράδειγμα ανά έτος, ανά μήνα, ανά εβδομάδα ή ανά ημέρα. Βασική προϋπόθεση για την μελέτη χρονοσειρών αποτελεί ο τρόπος μέτρησης των τιμών της χρονοσειράς ώστε να δημιουργείται η δυνατότητα σύγκρισης των δεδομένων. Εννοώντας ότι, αν η χρονοσειρά αποτελείται από λόγου χάρη ημερήσια δεδομένα, τότε η μέτρηση της κάθε τιμής θα πρέπει να γίνεται την ίδια χρονική στιγμή μέσα στην μέρα και στην ίδια τοποθεσία, αν αυτό επηρεάζει την ίδια την τιμή.

Ένα πολύ σημαντικό χαρακτηριστικό των χρονοσειρών είναι ότι συχνά αναφέρονται μόνο στο παρελθόν. Επομένως είναι κατάλληλες για εξαγωγή συμπερασμάτων από μια μεταβλητή αλλά κυρίως είναι κατάλληλες μέσω της πληροφορίας από το παρελθόν για προβλέψεις που

σχετίζονται με μελλοντικές χρονικές στιγμές. Αναλυτικά, μια χρονοσειρά ορίζεται από τις τιμές Y_1, Y_2, \dots, Y_n κάποιας μεταβλητής Y κατά τις χρονικές στιγμές t_1, t_2, \dots, t_n . Επομένως το Y είναι μια συνάρτηση του t , και συμβολίζεται ως $Y = F(t)$. Η γραφική παράσταση της συνάρτησης $Y = F(t)$ παρουσιάζει την εξέλιξη της μεταβλητής Y στο χρόνο. Ενδιαφέρον παρουσιάζει η γραφική παράσταση μίας χρονοσειράς που περιγράφεται σαν την κίνηση ενός σημείου καθώς κυλάει ο χρόνος. Η βιβλιογραφική εμπειρία αποκαλύπτει κάποιες μορφές χαρακτηριστικών κινήσεων (characteristic movements), που εμφανίζονται συνηθέστερα σε κάποιο βαθμό. Η μετέπειτα ανάλυση και μελέτη αυτών των κινήσεων είναι η κινητήριος δύναμη που σε πολλές περιπτώσεις μας ωθεί στην ορθή πρόβλεψη (forecasting) των μελλοντικών τιμών της χρονοσειράς.

Οι χρονοσειρές μπορούν να διακριθούν ανάλογα με τις τιμές που λαμβάνει ο δείκτης t σε χρονοσειρές συνεχούς χρόνου και σε χρονοσειρές διακριτού χρόνου. Χρονοσειρές **συνεχούς χρόνου** (continuous time series) είναι αυτές που θεωρητικά η μέτρηση των τιμών τους μπορεί να γίνει σε κάθε χρονική στιγμή. Χαρακτηριστικό παράδειγμα αυτής της κατηγορίας είναι η θερμοκρασία που μπορεί να καταγραφεί οποιαδήποτε χρονική στιγμή κατά τη διάρκεια της ημέρας ή και της νύχτας. Χρονοσειρές **διακριτού χρόνου** (discrete time series) είναι αυτές που οι παρατηρήσεις τους λαμβάνονται σε προκαθορισμένα, διακριτά ίσα χρονικά διαστήματα όπως το έτος, το τρίμηνο, ο μήνας και η ημέρα. Στην κατηγορία αυτή ανήκουν οι περισσότερες χρονοσειρές από το χώρο της οικονομίας. Οι χρονοσειρές διακριτού χρόνου μπορούν να προσεγγίσουν ικανοποιητικά και τις χρονοσειρές συνεχούς χρόνου εφόσον επιλεχτεί το κατάλληλο διάστημα δειγματοληψίας.

1.2 Βασικά χαρακτηριστικά χρονοσειρών

Η ανάλυση των χρονοσειρών μέσω της απαραίτητης διερεύνησης μας δίνει την δυνατότητα χρησιμοποίησης αυτής σε πολλαπλές εφαρμογές σε διαφορετικούς επιστημονικούς τομείς. Σκοπός είναι η δημιουργία μοντέλων που θα μπορούν να προβλέψουν τις μελλοντικές τιμές ενός υπό

εξέταση μεγέθους. Η ανάλυση αυτή πραγματοποιείται μέσω γραφικών παραστάσεων, ιστογραμμάτων και διαγραμμάτων. Οι βασικές έννοιες που χαρακτηρίζουν μια χρονοσειρά είναι η στασιμότητα, η τάση, η περιοδικότητα ή εποχικότητα και οι ιδιόμορφες τιμές (outliers).

1.2.1 Στασιμότητα

Αν θέλαμε να δώσουμε έναν διαισθητικό ορισμό για την έννοια της στασιμότητας σε μια χρονοσειρά θα λέγαμε ότι μια χρονοσειρά είναι στάσιμη αν δεν υπάρχει συστηματική αλλαγή του μέσου όρου και της διασποράς της κατά τη διάρκεια του χρόνου. Αυτό σημαίνει ότι σε περίπτωση που υπάρχει τάση για την οποία γίνεται αναφορά παρακάτω 1.2.3, τότε η χρονοσειρά δεν μπορεί να χαρακτηριστεί από στασιμότητα. Η στασιμότητα εκφράζεται με δύο μορφές, την αυστηρή στασιμότητα και την ασθενή στασιμότητα. Η ασθενής στασιμότητα είναι σημαντική όταν μοντελοποιείται ο μέσος όρος μιας διαδικασίας, αν και η αυστηρή στασιμότητα είναι χρήσιμη σε πιο περίπλοκες ρυθμίσεις, όπως τα μη γραμμικά μοντέλα.

Αυστηρή στασιμότητα

Μια στοχαστική διαδικασία Y_t είναι αυστηρά στάσιμη αν η κοινή κατανομή των $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ είναι αμετάβλητη ως προς την ολίσθηση στον χρόνο για κάθε t_1, t_2, \dots, t_n και κάθε τιμή του n . Η αυστηρή στασιμότητα απαιτεί ότι η κοινή κατανομή μιας στοχαστικής διαδικασίας δεν εξαρτάται από το χρόνο και έτσι ο μόνος παράγοντας που επηρεάζει τη σχέση μεταξύ δύο παρατηρήσεων είναι το διάστημα μεταξύ τους. Με άλλα λόγια, μετατοπίζοντας την αρχή του χρόνου κατά ένα χρονικό διάστημα, έστω τ , δεν μεταβάλλεται η από κοινού κατανομή. Αυτή πρέπει να εξαρτάται μόνο από τα διαστήματα μεταξύ των t_1, t_2, \dots, t_n . Η παραπάνω ιδιότητα ισχύει για οποιαδήποτε τιμή του n . Ειδικότερα, για $n = 1$, συνεπάγεται ότι η κατανομή της Y_t πρέπει να είναι ίδια για όλα τα t , και επομένως: $\mu(t) = \mathbb{E}[Y_t] = \mu$ και $\sigma^2(t) = \text{Var}[Y_t] = \sigma^2$. Δηλαδή η μέση τιμή και η διασπορά είναι σταθερές και δεν εξαρτώνται από

την τιμή του t . Από την άλλη πλευρά, η ασθενής στασιμότητα επιβάλλει περιορισμούς μόνο στα στατιστικά μεγέθη πρώτης και δεύτερης τάξης μιας στοχαστικής διαδικασίας. Για τον ορισμό της ασθενούς στασιμότητας μιας στοχαστικής διαδικασίας ή μιας χρονοσειράς χρησιμοποιούνται οι ροπές.

Ασθενής στασιμότητα

Μια στοχαστική διαδικασία Y_t είναι ασθενώς στάσιμη εάν ισχύει

$$\mathbb{E}[Y_t] = \mu, \quad \forall t, \quad (1.1)$$

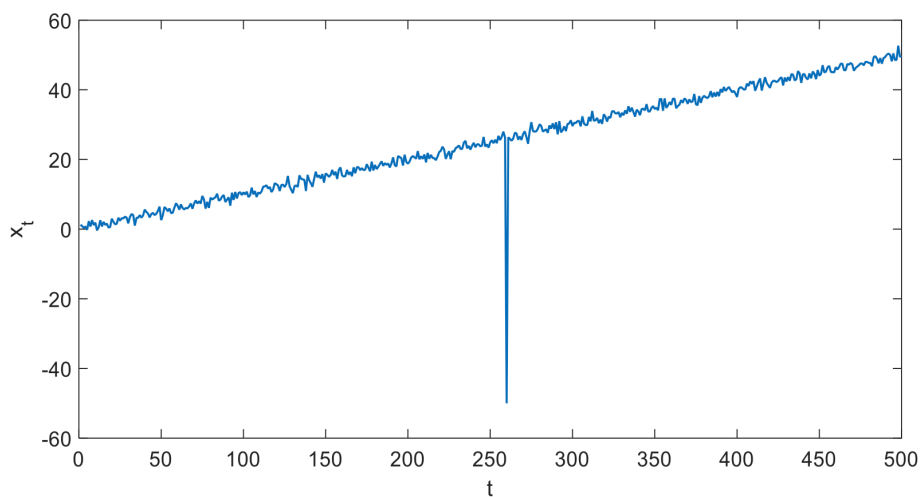
$$\text{Var}[Y_t] = \sigma^2 < \infty, \quad \forall t, \quad (1.2)$$

$$\mathbb{E}[(Y_{t_1} - \mu)(Y_{t_2} - \mu)] = \mathbb{E}[(Y_{t_1+\tau} - \mu)(Y_{t_2+\tau} - \mu)], \quad \forall t_1, t_2, \forall \tau. \quad (1.3)$$

Η ασθενής στασιμότητα απαιτεί ότι τόσο η άνευ όρων μέση τιμή όσο και η άνευ όρων διακύμανση να είναι πεπερασμένες και να μην μεταβάλλονται με το χρόνο. Υπογραμμίζοντας ότι η ασθενής στασιμότητα ισχύει μόνο για τις άνευ όρων ροπές και όχι για τις υπό συνθήκη ροπές, και έτσι μια διαδικασία συνδιακύμανσης μπορεί να έχει μεταβαλλόμενη υπό συνθήκη μέση τιμή (δηλαδή να είναι προβλέψιμη). Αυτοί οι δύο τύποι στασιμότητας σχετίζονται μεταξύ τους, αν και κανένας από τους δύο δεν αποτελεί μέρος του άλλου. Εάν μια διαδικασία είναι αυστηρά στάσιμη και έχει πεπερασμένες δεύτερες ροπές, τότε είναι και ασθενώς στάσιμη. Εάν μια διαδικασία είναι ασθενώς στάσιμη και η κοινή κατανομή των μαθηματικών καταλοίπων δεν εξαρτάται από το χρόνο, τότε η διαδικασία είναι αυστηρά στάσιμη. Ωστόσο, ο ένας τύπος μπορεί να εμφανιστεί χωρίς τον άλλο, μπορεί να εμφανιστούν και οι δύο ή μπορεί να μην ισχύει κανένας από τα δύο σε μια συγκεκριμένη χρονοσειρά. Για παράδειγμα, εάν μια διαδικασία έχει ροπές ανώτερης τάξης που εξαρτώνται από το χρόνο (π.χ. χρονικά μεταβαλλόμενη κύρτωση), μπορεί να είναι ασθενώς στάσιμη αλλά όχι αυστηρά στάσιμη.

1.2.2 Ακραίες τιμές (Outliers)

Οι ακραίες τιμές (outliers) είναι οι απομονωμένες παρατηρήσεις που εμφανίζονται στο γράφημα κάποιας χρονοσειράς ως απότομες αλλαγές στο πρότυπο συμπεριφοράς της. Οι ακραίες τιμές είναι μη προβλέψιμες και η επίδρασή τους στην χρονοσειρά έχει μικρή χρονική διάρκεια. Η ερμηνεία τέτοιων παρατηρήσεων χρειάζεται ιδιαίτερη προσοχή, διότι απαιτείται θεωρητική γνώση, κριτική ικανότητα και κοινή λογική. Ένα ακραίο σημείο μπορεί να αντιπροσωπεύει μια ασυνήθιστη παρατήρηση που οφείλεται σε κάποιο απρόβλεπτο γεγονός. Για παράδειγμα, μια απεργία μπορεί να προκαλέσει μεγάλη πτώση στην παραγωγή μιας βιομηχανικής μονάδας μια μέρα του χρόνου. Ένα χαρακτηριστικό παράδειγμα ακραίας τιμής φαίνεται στο ακόλουθο γράφημα (Σχήμα 1.1).



Σχήμα 1.1: Γραφική παράσταση χρονοσειράς που παρουσιάζει μια ακραία τιμή

1.2.3 Τάση και περιοδικότητα

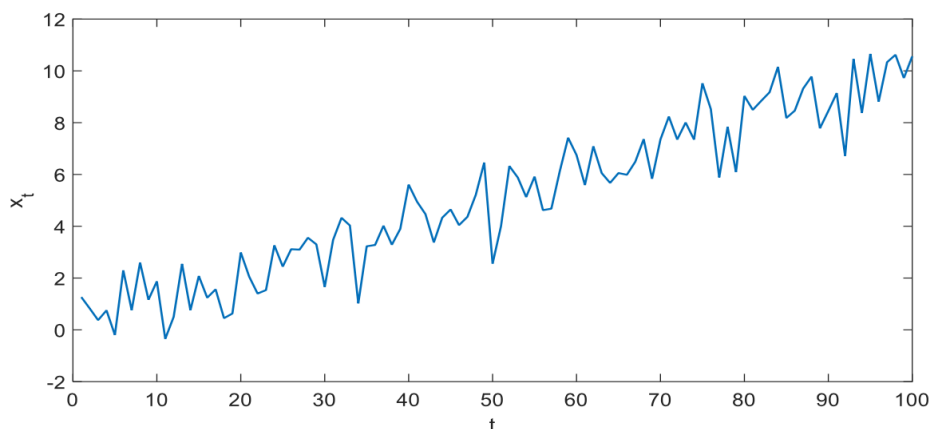
Η τάση (trend) στις χρονοσειρές είναι μια μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών μιας χρονοσειράς. Η τάση των τιμών μιας μεταβλητής μπορεί να είναι ανοδική, πτωτική ή σταθερή σε ένα συγκεκριμένο χρονικό διάστημα. Συχνά, μπορεί να εκτιμηθεί από διάφορες οικογένειες καμπυλών, όπως μια ευθεία γραμμή ή μια εκθετική καμπύλη. Για να είναι ασφαλή τα συμπεράσματα που θα εξαχθούν για το αν μια σειρά παρουσιάζει τάση ή όχι θα πρέπει να έχουμε έναν ικανό αριθμό παρατηρήσεων και να εκτιμηθεί σε ένα κατάλληλο χρονικό διάστημα. Όταν η τάση σε μια χρονοσειρά μπορεί να περιγραφεί από κάποια γνωστή ή εκτιμώμενη συνάρτηση του χρόνου, $\mu_t = f(t)$, ονομάζεται καθοριστική τάση (deterministic trend).

Η περιοδικότητα είναι ένα επαναλαμβανόμενο μοτίβο υψηλών ή χαμηλών τιμών ενός χαρακτηριστικού, το οποίο μεταβάλλεται ανά τακτά χρονικά διαστήματα. Η μέση τιμή των χρονοσειρών συχνά παρουσιάζει μια αυξητική ή φθίνουσα τάση ή εμφανίζει μια κυκλική επαναλαμβανόμενη συνέχεια σε χρονικά διαστήματα ή εποχές. Πριν την απαλοιφή των δυο αυτών φαινομένων πρέπει να γίνει περαιτέρω επεξεργασία των δεδομένων της χρονοσειράς. Μια χρονοσειρά Y_t μπορεί σε κάθε χρονική στιγμή t να αναλυθεί στις συνιστώσες τάσης και περιοδικότητας. Αυτό γίνεται με χρήση του μοντέλου

$$Y_t = \mu_t + s_t + X_t, \quad (1.4)$$

όπου η συνιστώσα μ_t εκφράζει την τάση ως συνάρτηση του χρόνου, η συνιστώσα s_t είναι μια περιοδική συνάρτηση η οποία εκφράζει την εποχικότητα και το X_t αποτελεί τον θόρυβο του συστήματος. Επισημαίνεται πως η απαλοιφή της τάσης και της περιοδικότητας γίνεται όταν δεν μας ενδιαφέρουν οι δυο αυτές έννοιες γιατί θεωρούμε ότι δημιουργούνται από άλλους παράγοντες και αιτιοκρατικά φαινόμενα. Σε περίπτωση πρόβλεψης, είτε συμπεριλαμβάνεται η τάση και η περιοδικότητα στο μοντέλο πρόβλεψης, είτε πραγματοποιείται εκτίμηση του μοντέλου της χρονοσειράς αφαιρώντας την τάση και την περιοδικότητα και στις προβλέψεις που προκύπτουν προστίθενται τα μεγέθη αυτά για να προ-

κύψει η πρόβλεψη του παρατηρούμενου μεγέθους. Στο παρακάτω γράφημα (Σχήμα 1.2) φαίνεται η γραφική παράσταση μιας χρονοσειράς που παρουσιάζει γραμμική ανοδική τάση. Μπορεί όμως η τάση σε μια χρονοσειρά να μην είναι δυνατόν να περιγραφεί από μια γνωστή συνάρτηση του χρόνου, οπότε και ονομάζεται στοχαστική τάση.



Σχήμα 1.2: Γραφική παράσταση μιας χρονοσειράς που παρουσιάζει γραμμική ανοδική τάση

Απαλοιφή τάσης

Αρχικά υποθέτουμε πως η μη στασιμότητα της χρονοσειράς οφείλεται αποκλειστικά στην ύπαρξη τάσης δηλαδή θεωρούμε ότι η Y_t δίνεται ως

$$Y_t = \mu_t + X_t,$$

όπου Y_t είναι η χρονοσειρά και X_t αποτελεί τον θόρυβο του συστήματος. Γενικά, αν η τάση στη χρονοσειρά είναι καθοριστική, μπορούμε να την εκτιμήσουμε και στην συνέχεια να την απαλείψουμε (detrend) με κάποια παραμετρική συνάρτηση $f(t)$ του χρόνου, όπως με πολυώνυμο κάποιου βαθμού p :

$$\mu_t = f(t) = a_0 + a_1 t + \dots + a_p t^p.$$

Όταν η τάση όμως δε φαίνεται να είναι κάποια γνωστή συνάρτηση του χρόνου, η προσαρμογή μιας συνάρτησης του χρόνου σε όλη τη χρονοσειρά δεν είναι κατάλληλη. Σε αυτήν την περίπτωση θα πάρουμε καλύτερα

αποτελέσματα αν προσαρμόσουμε τη συνάρτηση $f(t)$ τοπικά σε συνεχόμενα τμήματα της χρονοσειράς. Όταν η τάση είναι στοχαστική μπορεί να απαλειφθεί παίρνοντας τις **πρώτες διαφορές**. Ο μετασχηματισμός αυτός είναι

$$X_t = \nabla Y_t = Y_t - Y_{t-1} = (1-B)Y_t, \quad (1.5)$$

όπου ∇ δηλώνει τον τελεστή της διαφοράς πρώτης τάξης και B είναι ο τελεστής υστέρησης (log operator). Αν η χρονοσειρά δε γίνει στάσιμη, μπορούμε να πάρουμε πάλι πρώτες διαφορές και αυτός είναι ο **μετασχηματισμός διαφορών δεύτερης τάξης** (second differences)

$$X_t = \nabla^2 Y_t = \nabla(\nabla Y_t) = (1-B)(1-B)Y_t = (1-2B+B^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}.$$

Γενικά, στην πράξη χρησιμοποιούμε συνήθως διαφορές πρώτης τάξης (πρώτες διαφορές) γιατί στις περισσότερες περιπτώσεις οι τοπικές τάσεις μπορεί να προσεγγιστούν ικανοποιητικά από γραμμικά πολυώνυμα. Πράγματι, αν η τάση είναι τοπικά γραμμική, δηλαδή αν $\mu_t = a_0 + a_1 t$, απαλείφεται με τις πρώτες διαφορές. Ο μετασχηματισμός των πρώτων διαφορών δίνει

$$\nabla Y_t = Y_t - Y_{t-1} = \mu_t - \mu_{t-1} + X_t - X_{t-1},$$

όπου η τάση στη νέα χρονοσειρά είναι

$$\mu_t - \mu_{t-1} = a_0 + a_1 t - a_0 - a_1(t-1) = a_1,$$

δηλαδή σταθερή ως προς το χρόνο και άρα έχει απαλειφθεί. Γενικά αν η τάση εκφράζεται τοπικά με πολυώνυμο βαθμού p , απαλείφεται με χρήση διαφορών τάξης p , $\nabla^p Y_t$. Τρίτος τρόπος απαλοιφής της τάσης είναι με τη χρήση φίλτρου κινούμενου μέσου όρου (Moving Average, MA) τάξης $2q+1$. Για κάθε χρονική στιγμή t , $q < t \leq n-q$, το στοιχείο της τάσης μ_t της χρονοσειράς Y_1, Y_2, \dots, Y_n εκτιμάται από τον τοπικό μέσο των παρατηρήσεων στο διάστημα $[t-q, t+q]$, δηλαδή

$$\hat{\mu}_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}. \quad (1.6)$$

Οι τιμές $\hat{\mu}_{q+1}, \hat{\mu}_{q+2}, \dots, \hat{\mu}_{n-q+1}$ αφαιρούνται από τις αντίστοιχες αρχικές παρατηρήσεις και η χρονοσειρά $X_{q+1}, X_{q+2}, \dots, X_{n-q+1}$ που προκύπτει είναι απαλλαγμένη από τάσεις. Οι πρώτες και οι τελευταίες παρατηρήσεις παραλείπονται ή θέτονται ίσες με τις αρχικές. Γενικά μπορεί ο κινούμενος μέσος να είναι σταθμισμένος,

$$\hat{\mu}_t = \sum_{j=-q}^q a_j Y_{t-j}, \quad \sum_{j=-q}^q a_j = 1.$$

Είναι φανερό πως ο απλός κινούμενος μέσος προκύπτει από την παραπάνω έκφραση για συντελεστές στάθμισης

$$a_j = \frac{1}{2q+1}, \quad j = -q, \dots, q.$$

Αν η τάξη του κινούμενου μέσου είναι άρτιος αριθμός $d = 2q$, μπορεί να ορισθεί όπως στην Εξ. (1.6) αλλά με στάθμιση 0.5 στις ακραίες τιμές για χρόνο $t - q$ και $t + q$, δηλαδή

$$\hat{\mu}_t = \frac{1}{d} (0.5Y_{t-d/2} + Y_{t-d/2+1} + \dots + Y_{t+d/2-1} + 0.5Y_{t+d/2}). \quad (1.7)$$

Η επιλογή του q παίζει σημαντικό ρόλο στην εξομάλυνση της σειράς και στη συνέχεια στην απαλοιφή της τάσης. Αν θέλουμε να απαλείψουμε μόνο πολύ αργές μεταβολές (τάσεις) θα πρέπει να χρησιμοποιήσουμε μεγάλη τάξη, ενώ για την απαλοιφή μεταβολών σε μικρότερη χρονική κλίμακα η τάξη πρέπει να είναι αντίστοιχα μικρή. Τελειώνοντας, θα πρέπει να σημειωθεί ότι αν δεν έχουμε κάποια πληροφορία για τη μορφή της τάσης που θέλουμε να απαλείψουμε (σε ποια χρονική κλίμακα πιστεύουμε ότι οι αργές μεταβολές είναι άσχετες με το υπό μελέτη σύστημα και ορίζονται από εξωγενείς παράγοντες που δε μας ενδιαφέρει να τους εμπλέξουμε στην ανάλυση μας), τότε δεν είναι ξεκάθαρο ποια μέθοδος ή παράμετρος μεθόδου απαλοιφής τάσης είναι η πιο κατάλληλη.

Απαλοιφή περιοδικότητας

Υποθέτοντας ότι η χρονοσειρά που μελετήθηκε παραπάνω έχει μόνο περιοδικότητα τότε η Y_t εκφράζεται ως

$$Y_t = s_t + X_t,$$

όπου η περιοδική συνάρτηση s_t έχει γνωστή περίοδο d . Αν, για παράδειγμα, η περιοδικότητα αντιστοιχεί σε κάποια γνωστή περίοδο, όπως 24 ώρες ή 12 μήνες, αναφέρεται ως εποχικότητα. Γενικότερα για γνωστή περίοδο d , ένας απλός τρόπος εκτίμησης της s_t είναι από τους **μέσους όρους των στοιχείων της περιοδικής συνάρτησης** s_i όπου $i = 1, \dots, d$. Αν $k = \lfloor n/d \rfloor$ είναι ο αριθμός των περιόδων στη χρονοσειρά Y_1, Y_2, \dots, Y_n , τότε κάθε στοιχείο της περιοδικής συνάρτησης s_i μπορεί να εκτιμηθεί ως

$$\hat{s}_i = \frac{1}{k} \sum_{j=1}^k Y_{i+jd}. \quad (1.8)$$

Δεύτερος τρόπος εκτίμησης της περιοδικής συνάρτησης είναι με τον **κινούμενο μέσο όρο θέτοντας την τάξη του ίση με την περίοδο d** . Η εξομάλυνση με το φίλτρο αυτό έχει ως αποτέλεσμα να καταστραφεί η περιοδικότητα, δηλαδή παίρνοντας τον κινούμενο μέσο όρο τάξης d σε μεγάλο βαθμό απαλείφουμε το περιοδικό στοιχείο περιόδου d . Για να εκτιμήσουμε το s_t με ακρίβεια (και στη συνέχεια να το απαλείψουμε) θα πρέπει να πάρουμε πρώτα τη διαφορά της αρχικής χρονοσειράς $Y_{q+1}, Y_{q+2}, \dots, Y_{n-q+1}$ και του κινούμενου μέσου $\hat{\mu}_{q+1}, \hat{\mu}_{q+2}, \dots, \hat{\mu}_{n-q+1}$, έστω $w_t = Y_t - \hat{\mu}_t$ (εδώ έχουμε υποθέσει πως $d = 2q+1$, αν το d είναι άρτιος χρησιμοποιούμε τη διόρθωση της Εξ. (1.7)). Στη συνέχεια παίρνουμε το μέσο όρο των $w_t = w_{i+jd}$ ως προς κάθε στοιχείο i για $i = 1, \dots, d$, έστω \bar{w}_i . Αν τα \bar{w}_i για $i = 1, \dots, d$ δεν αθροίζονται στο 0, αφαιρούμε τη μέση τιμή τους και η εκτίμηση της περιοδικής συνάρτησης είναι

$$\hat{s}_i = \bar{w}_i - \frac{1}{d} \sum_{j=1}^d w_j, \quad i = 1, \dots, d \quad (1.9)$$

όπου $\hat{s}_{t-d} = \hat{s}_t$. Αν μας ενδιαφέρει απλά να απαλείψουμε το περιοδικό στοιχείο περιόδου d , δηλαδή αν δεν μας ενδιαφέρει να το εκτιμήσουμε, μπορούμε να πάρουμε της διαφορές υστέρησης d :

$$\nabla^d Y_t = Y_t - Y_{t-d} = (1 - B^d) Y_t \quad t = d+1, \dots, n. \quad (1.10)$$

Απαλοιφή τάσης και περιοδικότητας

Γενικά μια χρονοσειρά μπορούμε να τη χωρίσουμε σε τρεις συνιστώσες ως εξής

$$Y_t = \mu_t + s_t + X_t,$$

όπου μ_t είναι η συνιστώσα της τάσης, s_t η συνιστώσα της περιοδικότητας για κάποια περίοδο d ($s_{t-d} = s_t$) και X_t είναι η χρονοσειρά των υπολοίπων αν αφαιρέσουμε από την παρατηρούμενη χρονοσειρά την τάση και την περιοδικότητα. Η τάση και η περιοδικότητα είναι και οι δύο συναρτήσεις του χρόνου και δεν περιέχουν πληροφορία για τη δυναμική του συστήματος, δηλαδή την εξάρτηση της παρατήρησης Y_t από τις προηγούμενες παρατηρήσεις. Σε κάποια προβλήματα όλη η πληροφορία που μας ενδιαφέρει μπορεί να εντοπίζεται στη συνιστώσα της τάσης και της περιοδικότητας, οπότε το πρόβλημα περιορίζεται στην εκτίμηση των μ_t και s_t . Η εκτίμηση των δυο αυτών συνιστωσών γίνεται μέσω επαναληπτικών διαδικασιών, με σκοπό τον διαδοχικό προσδιορισμό των παραμέτρων ή με μια διαδικασία ταυτόχρονου προσδιορισμού και των δύο. Γενικά υπάρχουν αρκετές μέθοδοι για τον προσδιορισμό των συνιστωσών αυτών όπως η μέθοδος χαμηλής τάσης, η μέθοδος κινούμενου μέσου και η μέθοδος διαφορών με υστέρηση d .

1.3 Στατιστικά μεγέθη χρονοσειράς

Μία χρονοσειρά, όπως έχει αναφερθεί παραπάνω, είναι μία στοχαστική διαδικασία, δηλαδή κάθε παρατήρησή της αποτελεί μία τυχαία μεταβλητή. Τα βασικότερα χαρακτηριστικά μίας τυχαίας μεταβλητής Y είναι η μέση τιμή $\mu = \mathbb{E}[Y]$ και η διασπορά $\sigma^2 = \text{Var}[Y]$. Χαρακτηριστικό μέγεθος για ένα ζεύγος τυχαίων μεταβλητών, έστω X και Y , είναι η μικτή ροπή 2ης τάξης, δηλαδή η συνδιακύμανση $\sigma_{XY} = \text{Cov}(X, Y)$. Παρακάτω παρουσιάζονται τα κυριότερα στατιστικά μεγέθη μίας στοχαστικής διαδικασίας.

1.3.1 Μέση τιμή

Η μέση τιμή μιας χρονοσειράς Y_t ορίζεται ως:

$$\mu_t = \mathbb{E}(Y_t) = \int_{-\infty}^{+\infty} y f_Y(y; t) dy.$$

Η μέση τιμή μ_t σχετίζεται άμεσα με την έννοια της τάσης της χρονοσειράς, εφόσον είναι εκφρασμένη συναρτήσει της χρονικής στιγμής t κάθε παρατήρησης Y_t . Συγκεκριμένα, αν σε μια χρονοσειρά παρουσιάζεται αυξητική ή πτωτική τάση σε ένα χρονικό διάστημα, αυτό θα αποτυπώνεται και στη μέση τιμή.

1.3.2 Αυτοσυνδιακύμανση

Η συνάρτηση συνδιακύμανσης είναι μέτρο του βαθμού συσχέτισεως δύο μεταβλητών και ορίζεται ως

$$\gamma_{s,t} = \text{Cov}(Y_s, Y_t) = \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)],$$

και υπολογίζει τη γραμμική εξάρτηση μεταξύ δύο τιμών της χρονοσειράς που παρατηρούνται σε διαφορετικούς χρόνους. Η αυτοσυνδιακύμανση j -οστής τάξης γ_{jt} για την χρονοσειρά Y_t δίνεται από την σχέση

$$\gamma_{jt} = \mathbb{E}[(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})] = \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} \dots \int_{-\infty}^{-\infty} (y_t - \mu_t)(y_{t-j} - \mu_{t-j}) f_Y(y_1, y_2, \dots, y_j; t, t-1, \dots, t-j) dy_1, dy_2 \dots dy_j.$$

1.3.3 Αυτοσυσχέτιση

Ο συντελεστής αυτοσυσχέτισης είναι ένας από τους πιο σημαντικούς στατιστικούς δείκτες ο οποίος χρησιμοποιείται ευρέως στην ανάλυση χρονοσειρών για την διαπίστωση ύπαρξης τυχαιότητας ή μη στην χρονοσειρά. Η αυτοσυσχέτιση (autocorrelation) j -οστής τάξης ρ_{jt} της τυχαίας μεταβλητής Y_t με μια προηγούμενη χρονική εκδοχή της Y_{t-j} ορίζεται για κάθε t ως εξής:

$$\rho_{jt} = \frac{\gamma_{jt}}{\gamma_{0t}} = \frac{\mathbb{E}[(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})]}{\mathbb{E}[(Y_t - \mu_t)^2]}.$$

Το σύνολο τιμών της αυτοσυσχέτισης ρ_{jt} είναι το $[-1, 1]$. Η αυτοσυσχέτιση είναι ένας σημαντικός στατιστικός δείκτης στη μελέτη χρονοσειρών, διότι έτσι δίνεται ένα μέτρο για τον βαθμό της σχέσης μεταξύ των δύο μεταβλητών. Ειδικότερα, φτιάχνοντας το γράφημα της αυτοσυσχέτισης συναρτήσει της καθυστέρησης j , το οποίο ονομάζεται συνάρτηση αυτοσυσχέτισης (Autocorrelation Function, ACF) δύναται να αποσαφηνιστούν στοιχεία για τα χαρακτηριστικά της χρονοσειράς. Αν η τιμή του ρ_{jt} ισούται περίπου με $+1$ ή -1 , τότε οι παρατηρήσεις Y_t και Y_{t-j} είναι ισχυρά συσχετισμένες. Ενώ, αν ρ_{jt} ισούται με 0 , τότε εύκολα προκύπτει ότι οι παρατηρήσεις Y_t και Y_{t-j} είναι ασυσχέτιστες. Συνεπώς, με γνώμονα την αυτοσυσχέτιση μπορούν να μελετηθούν τα ποιοτικά χαρακτηριστικά των χρονοσειρών, όπως η εποχικότητα και η στασιμότητα. Όπως και η συνάρτηση αυτοσυσχέτισης, η συνάρτηση μερικής αυτοσυσχέτισης (Partial Autocorrelation Function, PACF) αποτελεί πολύτιμη πηγή πληροφόρησης σχετικά με τα χαρακτηριστικά της αλληλεξάρτησης που δημιουργεί μια στοχαστική διαδικασία. Οι συντελεστές μερικής αυτοσυσχέτισης μετρούν το βαθμό της σχέσης μεταξύ των Y_t και Y_{t-k} όταν οι επιδράσεις όλων των άλλων χρονικών υστερήσεων $1, 2, 3, \dots, k-1$ έχουν αφαιρεθεί. Ο συντελεστής μερικής αυτοσυσχέτισης τάξης k συμβολίζεται με α_k και μπορεί να υπολογισθεί εφαρμόζοντας τη μέθοδο της πολλαπλής γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή την Y_t και ανεξάρτητες μεταβλητές τις Y_{t-1}, \dots, Y_{t-k} :

$$Y_t = b_0 + b_1 Y_{t-1} + \dots + b_k Y_{t-k}.$$

Ο συντελεστής α_k ισούται με τον συντελεστή b_k . Ο πρώτος συντελεστής μερικής αυτοσυσχέτισης α_1 είναι πάντα ίσος με τον πρώτο συντελεστή αυτοσυσχέτισης r_1 .

Κεφάλαιο 2

Κατανομές Χρονοσειρών Ταχύτητας Ανέμου

2.1 Κανονική κατανομή

Η κανονική κατανομή αποτελεί την πιο σημαντική κατανομή της στατιστικής μεθοδολογίας για τους εξής λόγους:

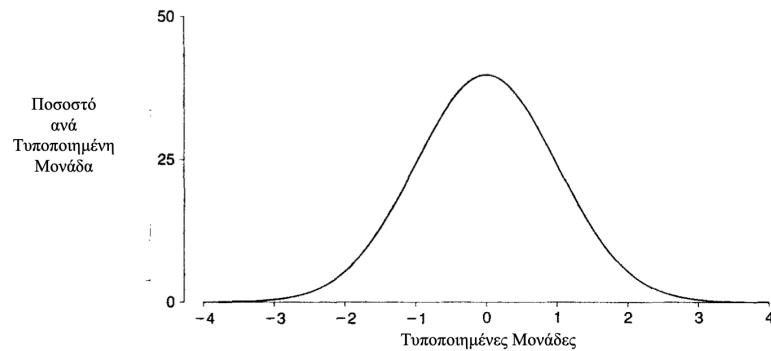
- Πολλές ασυνεχείς κατανομές πιθανοτήτων μπορούν να προσεγγιστούν μέσω της κανονικής κατανομής.
- Η κανονική κατανομή μπορεί να χρησιμοποιηθεί σαν προσέγγιση πολλών άλλων κατανομών.
- Αποτελεί την βάση για πολλές τεχνικές που χρησιμοποιούνται στην στατιστική συμπερασματολογία.
- Τυχαία σφάλματα που εμφανίζονται σε διάφορες μετρήσεις έχουν κανονική κατανομή. Γι' αυτό το λόγο η κανονική κατανομή αναφέρεται πολλές φορές και ως κατανομή σφαλμάτων.

Η κανονική κατανομή είναι συμμετρική γύρω από την μέση τιμή μ και έχει σχήμα καμπάνας με επικρατούσα τιμή την μέση τιμή μ . Η συνεχής κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

καλείται κανονική κατανομή με παραμέτρους μ, σ^2 , όπου $e \equiv 2,71828$, π η γνωστή μαθηματική σταθερά ($\pi \equiv 3,14159$), μ ο μέσος του πληθυσμού, σ η τυπική απόκλιση του πληθυσμού και x μια τιμή της συνεχούς τυχαίας μεταβλητής στο διάστημα $-\infty$ έως ∞ . Η απλούστερη μορφή της κατανομής αυτής, η οποία συνήθως χρησιμοποιείται σε πρακτικές εφαρμογές με μετασχηματισμό, είναι εκείνη που αναφέρεται στην περίπτωση όπου $\mu = 0$ και $\sigma^2 = 1$. Η κατανομή αυτή ονομάζεται τυποποιημένη κανονική κατανομή και έχει προφανώς την παρακάτω μορφή

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$



Σχήμα 2.1: Τυποποιημένη κανονική κατανομή

Όπως γίνεται αντιληπτό τόσο από το γράφημα της κατανομής, αλλά και από την μαθηματική της μορφή, η κατανομή αυτή είναι συμμετρική γύρω από την μέση τιμή της. Προφανώς, όπως σε όλες τις κατανομές, το εμβαδόν της επιφάνειας κάτω από την καμπύλη είναι 1.

2.2 Γενικευμένη κατανομή ακραίων τιμών

Η γενικευμένη κατανομή ακραίων τιμών (Generalized Extreme Value distribution) με παράμετρο θέσης μ , παράμετρο κλίμακας $\sigma > 0$ και παράμετρο σχήματος ξ , έχει αθροιστική συνάρτηση κατανομής που δίνεται από τον παρακάτω τύπο

$$G_\xi(x) = \begin{cases} e^{-[1+\xi\frac{x-\mu}{\sigma}]^{-\frac{1}{\xi}}}, & \text{αν } \xi \neq 0, \\ e^{-e^{[-\frac{x-\mu}{\sigma}]}}, & \text{αν } \xi = 0. \end{cases} \quad (2.1)$$

Η γενικευμένη κατανομή ακραίων τιμών συμβολίζεται ως $GEV(x; \mu, \sigma, \xi)$, $G_\xi(x)$ ή απλά $G(x)$. Η παράμετρος ξ είναι αυτή που καθορίζει τον τύπο των οριακών κατανομών. Για παράδειγμα η $GEV(1, 1, \frac{1}{\alpha})$ για $\alpha > 0$ θα είναι ίδιου τύπου με την $\Phi_\alpha(x)$ ενώ η $GEV(-1, -1, -\frac{1}{\alpha})$ θα είναι ίδιου τύπου με την $\Psi_\alpha(x)$. Παρατηρούμε ότι λαμβάνοντας το όριο στο μηδέν του πρώτου κλάδου της σχέσης (2.1) δηλαδή,

$$\lim_{\xi \rightarrow 0} e^{-[1+\xi\frac{x-\mu}{\sigma}]^{-\frac{1}{\xi}}} \quad (2.2)$$

καταλήγουμε στον δεύτερο κλάδο της σχέσης (2.1). Η παράμετρος σχήματος ξ αναφέρεται συνήθως και ως δείκτης ουράς (tail index) της κατανομής. Αυτό διότι κατέχει σημαντικό ρόλο στον προσδιορισμό της συμπεριφοράς της ουράς της κατανομής. Στο Σχήμα 2.2 παρακάτω φαίνεται το διάγραμμα της γενικευμένης κατανομής ακραίων τιμών για τιμές του ξ που αντιστοιχούν στις τρεις οριακές κατανομές από όπου φαίνεται ότι για μεγάλες τιμές του ξ έχουμε βαριά ουρά [11].

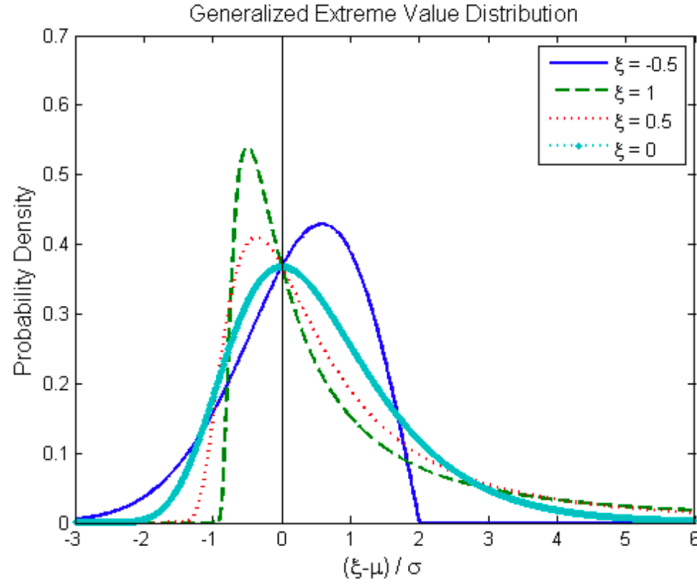
Συγκεκριμένα ισχύει,

$$\Lambda(x) = G_0(x), \quad (2.3)$$

$$\Phi_\alpha(x) = G_{1/\alpha}(\alpha(1-x)), \quad (2.4)$$

$$\Psi_\alpha(x) = G_{-1/\alpha}(\alpha(1-x)). \quad (2.5)$$

Η συνάρτηση πυκνότητας πιθανότητας της γενικευμένης κατανομής ακραίων τιμών προκύπτει παραγωγίζοντας την Εξ. (2.1).



Σχήμα 2.2: Γενικευμένη κατανομή ακραίων τιμών για διάφορες τιμές του ξ

$$g(x) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \frac{x-\mu}{\sigma} \right]^{-\frac{1}{\xi}-1} e^{-[1+\xi \frac{x-\mu}{\sigma}]^{-\frac{1}{\xi}}}, & \text{αν } \xi \neq 0, \\ \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)} e^{-e[-\frac{x-\mu}{\sigma}]}, & \text{αν } \xi = 0. \end{cases} \quad (2.6)$$

Η αντίστροφη συνάρτηση της αθροιστικής συνάρτησης κατανομής της σχέσης (2.1) δίνεται από τον τύπο

$$Q_{G_\xi(x)}(y) = \begin{cases} \mu + \frac{\sigma}{\xi} [-1 + (-\ln y)^{-\xi}], & \text{αν } \xi \neq 0, \\ \mu - \sigma (\ln(-\ln y)), & \text{αν } \xi = 0, \end{cases} \quad (2.7)$$

διότι για $\xi \neq 0$ έχουμε,

$$\begin{aligned} G_{\xi \neq 0}(x) = y &\Leftrightarrow \ln y = - \left(1 + \xi \left(\frac{x-\mu}{\sigma} \right)^{-\frac{1}{\xi}} \right) \\ &\Leftrightarrow x = \mu + \frac{\sigma}{\xi} [(-\ln y)^{-\xi} - 1] \end{aligned}$$

και για $\xi = 0$ έχουμε

$$\begin{aligned} G_{\xi=0}(x) = y &\Leftrightarrow e^{-e^{-\frac{x-\mu}{\sigma}}} = y \\ &\Leftrightarrow \frac{x-\mu}{\sigma} = -\ln(-\ln y) \\ &\Leftrightarrow x = \mu + \sigma(-\ln(\ln y)). \end{aligned}$$

Η μέση τιμή και η διακύμανση της κατανομής δίνεται από τους παρακάτω τύπους:

$$\mathbb{E}(X) = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - \Gamma(1 - \xi)), & \text{αν } \xi < 1, \xi \neq 0, \\ \mu + \sigma\gamma, & \text{αν } \xi = 0, \end{cases} \quad (2.8)$$

$$\mathbb{V}\text{ar}(X) = \begin{cases} \frac{\sigma^2}{\xi^2}(\Gamma(1 - 2\xi) - \Gamma^2(1 - \xi)), & \text{αν } \xi < \frac{1}{2}, \xi \neq 0, \\ \frac{\pi}{6}\sigma^2, & \text{αν } \xi = 0, \end{cases} \quad (2.9)$$

όπου $\gamma = 0.5772$ είναι η σταθερά Euler. Τέλος, ο συντελεστής ασυμμετρίας της γενικευμένης κατανομής ορίζεται ως

$$skew_{G_\xi} := \frac{\mathbb{E}[(X - \mathbb{E}(X))^3]}{\mathbb{V}\text{ar}(X)^{\frac{3}{2}}} \quad (2.10)$$

και είναι ίσος με 0 στο $\xi_0 \simeq -2.8$. Επιπλέον $skew_{G_\xi} > 0$ για $\xi > \xi_0$ και $skew_{G_\xi} < 0$ για $\xi < \xi_0$. Για την Gumbel έχουμε $skew_{G_0} \simeq 1.14$.

2.3 Κατανομή Weibull

Η κατανομή Weibull είναι μια συνεχής κατανομή πιθανότητας, η οποία πήρε το όνομά της από τον Waloddi Weibull, ο οποίος την περιέγραψε λεπτομερώς το 1951. Η συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής που ακολουθεί κατανομή Weibull είναι:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (2.11)$$

όπου $k > 0$ η παράμετρος μορφής και $\lambda > 0$ η παράμετρος κλίμακας. Η συμπληρωματική αθροιστική συνάρτηση κατανομής της είναι μια τεντωμένη εκθετική συνάρτηση.

Η κατανομή Weibull σχετίζεται με μια σειρά από άλλες κατανομές πιθανότητας. Ειδικότερα, παρεμβάλλει μεταξύ της εκθετικής κατανομής ($k = 1$) και της κατανομής Rayleigh ($k = 2$). Εάν η ποσότητα X είναι έτοιμη να "αποτύχει", η κατανομή Weibull δίνει μια κατανομή για την οποία το ποσοστό αποτυχίας είναι ανάλογο με μια δύναμη του χρόνου. Η παράμετρος σχήματος k είναι αυτή η δύναμη συν ένα, και έτσι αυτή η παράμετρος μπορεί να ερμηνευθεί ως εξής:

- Μια τιμή του $k < 1$ δείχνει ότι το ποσοστό αποτυχίας μειώνεται με την πάροδο του χρόνου. Αυτό συμβαίνει όταν υπάρχει σημαντική «θνησιμότητα», ή ελαττωματικά στοιχεία που παραλείπονται νωρίς και το ποσοστό αποτυχίας μειώνεται με την πάροδο του χρόνου, καθώς τα ελαττωματικά στοιχεία εξαλείφονται από τον πληθυσμό.
- Η τιμή του $k = 1$ δείχνει ότι το ποσοστό αποτυχίας είναι σταθερό με την πάροδο του χρόνου. Αυτό μπορεί να υποδηλώνει ότι τυχαία εξωτερικά γεγονότα προκαλούν θνησιμότητα ή αποτυχία.
- Μια τιμή του $k > 1$ δείχνει ότι οι ποσοστό αποτυχίας αυξάνεται με το χρόνο. Αυτό συμβαίνει όταν υπάρχει μια διαδικασία «γήρανσης», ή τμήματα που είναι πιο πιθανό να αποτύχουν, με το πέρασμα του χρόνου.

2.4 Λογαριθμο-λογιστική κατανομή (Log-logistic)

Ορίζουμε την λογαριθμο-λογιστική κατανομή με την βοήθεια της λογιστικής κατανομής. Εάν μία τυχαία μεταβλητή T ακολουθεί λογαριθμο-λογιστική κατανομή, τότε ο λογάριθμος της ακολουθεί λογιστική κατανομή:

$$T \sim \text{log-logistic} \Leftrightarrow Y = \ln T \sim \text{logistic}.$$

Μέσω της συνάρτησης πυκνότητας πιθανότητας της λογιστικής κατανομής,

$$f(t) = \frac{\tau^{-1} \exp\{(t - v)/\tau\}}{[1 + \exp\{(t - v)/\tau\}]^2},$$

όπου $v \in (-\infty, \infty)$ η παράμετρος θέσης και $\tau > 0$ η παράμετρος κλίμακας, προσδιορίζεται η σ.π.π. της λογαριθμο-λογιστικής κατανομής η οποία δίνεται από τη σχέση που φαίνεται παρακάτω:

$$f(t) = \frac{\alpha \beta t^{\beta-1}}{(1 + \alpha t^\beta)^2}, \quad t \geq 0,$$

όπου $\beta = \tau^{-1}$ και $\alpha = e^{-\frac{v}{\tau}}$. Η συνάρτηση αξιοπιστίας εκφράζεται ως:

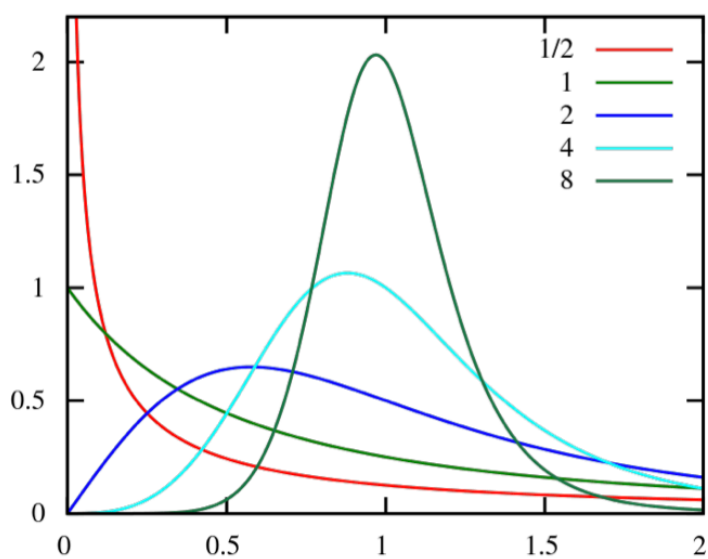
$$S(t) = (1 + \alpha t^\beta)^{-1}$$

και η συνάρτηση διακινδύνευσης ως:

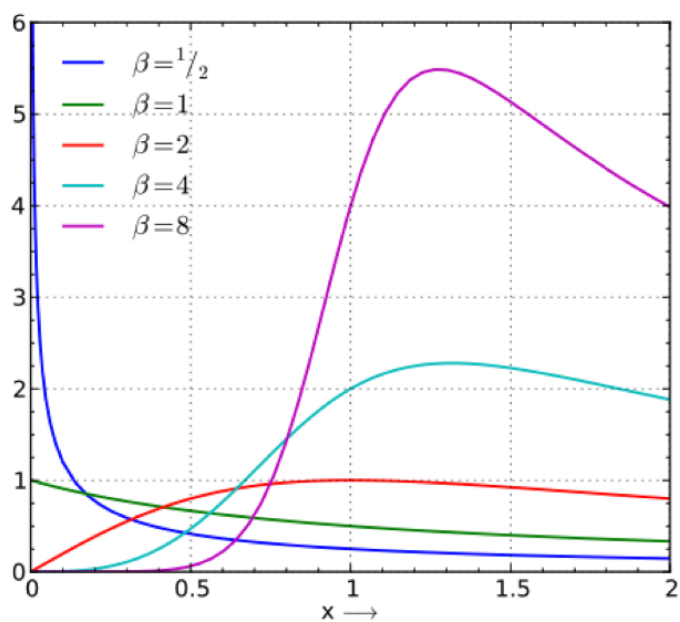
$$h(t) = \frac{\alpha \beta t^{\beta-1}}{1 + \alpha t^\beta}.$$

Στα Σχήματα 2.3 και 2.4 που ακολουθούν φαίνονται οι γραφικές παραστάσεις τόσο της συνάρτησης πυκνότητας πιθανότητας της λογαριθμο-λογιστικής κατανομής, όσο και της συνάρτησης διακινδύνευσης όπου η παράμετρος θέσης $\alpha = 1$ και είναι σταθερή σε όλες τις κατανομές, ενώ η παράμετρος β παίρνει τις τιμές που αναγράφονται στα διαγράμματα. Αποδεικνύεται εύκολα ότι η συνάρτηση διακινδύνευσης της λογαριθμο-λογιστικής κατανομής είναι:

- γνησίως φθίνουσα με $\lim_{t \rightarrow 0^+} h(t) = +\infty$, όταν $\beta < 1$
- γνησίως φθίνουσα με $\lim_{t \rightarrow 0^+} h(t) = +\alpha$, όταν $\beta = 1$



Σχήμα 2.3: Συνάρτηση πυκνότητας πιθανότητας της Λογαριθμο-λογιστικής κατανομής για παράμετρο κλίμακας $\alpha = 1$



Σχήμα 2.4: Συνάρτηση διακινδύνευσης της Λογαριθμο-λογιστικής κατανομής, για παράμετρο κλίμακας $\alpha = 1$ και για διάφορες τιμές της παραμέτρου β

2.5 Κατανομή Γάμμα

Μία τυχαία μεταβλητή T που ακολουθεί κατανομή Γάμμα με παράμετρο κλίμακας λ και παράμετρο σχήματος k έχει συνάρτηση πυκνότητας πιθανότητας:

$$f_T(t) = f(t) = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}, \quad t > 0.$$

Η συνάρτηση αξιοπιστίας ή επιβίωσης (survival) $S(t)$ της κατανομής Γάμμα δεν έχει κλειστή μορφή

$$S(t) = P\{T > t\} = \frac{\Gamma(k, \lambda t)}{\Gamma(k)}.$$

Η συνάρτηση διακινδύνευσης της κατανομής Γάμμα είναι $h(t) = f(t)/S(t)$, άρα,

$$h(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{\Gamma(k, \lambda t)}. \quad (2.12)$$

Ανάλογα με την τιμή της παραμέτρου k ,

$$h(t) \text{ είναι } \begin{cases} \text{γνησίως αύξουσα,} & \text{αν } k > 1, \\ \text{σταθερή,} & \text{αν } k = 1, \\ \text{γνησίως φθίνουσα,} & \text{αν } k < 1. \end{cases}$$

Από τις συναρτήσεις πυκνότητας πιθανότητας και διακινδύνευσης, είναι προφανές ότι για παράμετρο σχήματος $k = 1$, η κατανομή Γάμμα ταυτίζεται με την Εκθετική κατανομή $Exp(\lambda)$. Η αναμενόμενη τιμή και η διασπορά της τυχαίας μεταβλητής δίνεται από τους παρακάτω τύπους:

$$\mathbb{E}(T) = \frac{\Gamma(k+1)}{\lambda \Gamma(k)} = \frac{k}{\lambda}$$

και

$$\mathbb{V}\text{ar}(T) = \frac{k(k+1)}{\lambda^2} - \frac{k^2}{\lambda^2} = \frac{k}{\lambda^2}.$$

2.6 Λογαριθμοκανονική κατανομή (Lognormal)

Η λογαριθμοκανονική κατανομή είναι μια συνεχής κατανομή πιθανότητας μιας τυχαίας μεταβλητής της οποίας ο λογάριθμος ακολουθεί

κανονική κατανομή. Έτσι, εάν μια τυχαία μεταβλητή X είναι λογαριθμοκανονικά κατανεμημένη, τότε η συνάρτηση $Y = \ln X$ ακολουθεί κανονική κατανομή. Ομοίως, εάν η τυχαία μεταβλητή Y ακολουθεί κανονική κατανομή, τότε η συνάρτηση $X = \exp(Y)$ ακολουθεί λογαριθμοκανονική κατανομή. Μια τυχαία μεταβλητή η οποία είναι λογαριθμοκανονικά κατανεμημένη παίρνει μόνο θετικές πραγματικές τιμές. Σε μια λογαριθμοκανονική κατανομή X , οι παράμετροι που συμβολίζονται με μ και σ είναι, αντίστοιχα, ο μέσος όρος και η τυπική απόκλιση του φυσικού λογαρίθμου της μεταβλητής, πράγμα που σημαίνει:

$$X = e^{\mu + \sigma Z}$$

όπου Z είναι μια τυπική κανονική μεταβλητή.

Κεφάλαιο 3

Μοντέλα Χρονοσειρών

3.1 Γραμμικά μοντέλα πρόβλεψης χρονοσειρών

Για την πρόβλεψη χρονοσειρών χρησιμοποιούνται ευρέως δύο μοντέλα, το αυτοπαλινδρούμενο (Autoregressive, AR) μοντέλο και το μοντέλο κινούμενου μέσου (Moving Average, MA) [1]. Συνδυάζοντας αυτά τα δύο προκύπτει το μοντέλο Autoregressive Moving Average (ARMA) και το Autoregressive Integrated Moving Average (ARIMA). Για την πρόβλεψη χρονοσειρών που παρουσιάζουν εποχικότητα χρησιμοποιείται μια παραλλαγή του ARIMA, το Seasonal Autoregressive Integrated Moving Average (SARIMA). Ο γραμμικός χαρακτήρας των παραπάνω μοντέλων περιορίζει την πρόβλεψη μη γραμμικών προβλημάτων, τα οποία αποτελούν και την πλειοψηφία των χρονοσειρών. Έχοντας μια χρονοσειρά X_t και εφαρμόζοντας σε αυτήν την μέθοδο των πρώτων διαφορών προκύπτει μια στάσιμη χρονοσειρά καθώς έχουν απαλειφθεί οι τάσεις. Θα θεωρηθεί ότι η X_t είναι μια στοχαστική διαδικασία, το οποίο σημαίνει ότι το σύστημα οδηγείται σε θόρυβο. Υπάρχουν πολλές μέθοδοι για την δημιουργία μίας εκτίμησης ή πρόβλεψης όταν έχουμε να αντιμετωπίσουμε τέτοιου είδους χρονοσειρές. Στην συνέχεια παρατίθενται κάποιες από τις κυριότερες διαδικασίες για την εκτέλεση τέτοιων αναλύσεων.

3.2 Αυτοπαλινδρούμενο μοντέλο (AR)

Το αυτοπαλινδρούμενο μοντέλο [1] βασίζεται στην ιδέα ότι η τρέχουσα τιμή της σειράς, x_t , μπορεί να εξηγηθεί ως συνάρτηση p παρελθοντικών τιμών, $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, όπου το p καθορίζει τον αριθμό των βημάτων στο παρελθόν που απαιτούνται για την πρόβλεψη της τρέχουσας τιμής.

Αυτοπαλινδρούμενη διαδικασία τάξης p

Μια χρονοσειρά λέγεται αυτοπαλινδρούμενη (autoregressive time series) τάξης p όταν κάθε παρατήρηση x_t εκφράζεται ως ένα σταθμισμένο άθροισμα μιας σταθεράς α , μιας χρονοσειράς λευκού θορύβου και p καθυστερημένων εκδοχών της χρονοσειράς x . Ο όρος αυτοπαλινδρούμενη έχει να κάνει με το γεγονός ότι η εξαρτημένη μεταβλητή x_t παλινδρομείται σε παρελθοντικές τιμές της ίδιας μεταβλητής. Το p αναφέρεται στο μήκος της καθυστέρησης. Ο μαθηματικός ορισμός του μοντέλου είναι:

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t, \quad (3.1)$$

όπου $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$. Στην παραπάνω εξίσωση η w_t είναι στάσιμη χρονοσειρά λευκού θορύβου, η παράμετρος α σχετίζεται με τη μέση τιμή της χρονοσειράς και $(\phi_1, \phi_2, \dots, \phi_p)$ είναι οι παράμετροι της αυτοπαλινδρόμησης. Κάνοντας χρήση του τελεστή υστέρησης, όπου υστερήσεις μπορούν να χαρακτηριστούν οι τελεστές της προς τα πίσω μετατόπισης κατά την δημιουργία προβλέψεων, η εξίσωση (3.1) μπορεί να πάρει την παρακάτω συνοπτική μορφή

$$\Phi(B)x_t = w_t \quad (3.2)$$

όπου το B εκφράζει τον τελεστή υστέρησης και το $\Phi(B)$ αποτελεί το πολυώνυμο υστέρησης της αυτοπαλινδρούμενης διαδικασίας.

3.2.1 Αυτοπαλινδρούμενη διαδικασία πρώτης τάξης

Η αυτοπαλινδρούμενη διαδικασία πρώτης τάξης είναι η πιο απλή διαδικασία παλινδρόμησης, συμβολίζεται ως AR(1) και δίνεται από την σχέση $x_t = \phi x_{t-1} + w_t$. Επαναλαμβάνοντας προς τα πίσω k μετατοπίσεις έχουμε:

$$\begin{aligned} x_t &= \phi x_{t-1} + w_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t \\ &= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\ &= \vdots \\ &= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}. \end{aligned}$$

Η μέθοδος αυτή υποδηλώνει ότι, συνεχίζοντας να επαναλαμβάνουμε μετατοπίσεις προς τα πίσω, και υπό την προϋπόθεση ότι $|\phi| < 1$ και ότι το x_t είναι στάσιμο, μπορούμε να αναπαραστήσουμε ένα μοντέλο AR(1) ως μια γραμμική διαδικασία που δίνεται από την εξίσωση

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}. \quad (3.3)$$

Η διαδικασία AR(1) που ορίζεται από την εξίσωση (3.3) είναι στάσιμη με μέση τιμή

$$\mathbb{E}[x_t] = \sum_{j=0}^{\infty} \phi^j \mathbb{E}[w_{t-j}] = 0$$

και συνάρτηση αυτοσυνδιακύμανσης

$$\begin{aligned} \gamma(h) &= \text{Cov}(x_{t+h}, x_t) = \mathbb{E} \left[\left(\sum_{j=0}^{\infty} \phi^j w_{t+h-j} \right) \left(\sum_{k=0}^{\infty} \phi^k w_{t-k} \right) \right] \\ &= \mathbb{E} \left[\left(w_{t+h} + \dots + \phi^h w_t + \phi^{h+1} w_{t-1} + \dots \right) (w_t + \phi w_{t-1} + \dots) \right] \\ &= \sigma_w^2 \sum_{j=0}^{\infty} \phi^{h+j} \phi^j = \sigma_w^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_w^2 \phi^h}{1 - \phi^2}, \quad h \geq 0. \end{aligned} \quad (3.4)$$

Επισημαίνεται ότι $\gamma(h) = \gamma(-h)$ οπότε από την εξίσωση (3.4) η συνάρτηση αυτοσυσχέτισης της διαδικασίας AR(1) είναι

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \quad h \geq 0. \quad (3.5)$$

3.3 Μοντέλα κινούμενου μέσου MA

Μια από τις βασικές στοχαστικές διαδικασίες είναι αυτή του κινούμενου μέσου τάξης q , MA(q), που μαζί με την αυτοπαλινδρούμενη διαδικασία έχουν καθοριστικό ρόλο στην ανάλυση των χρονοσειρών. Το μοντέλο αυτό προκύπτει από την ακόλουθη σχέση:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (3.6)$$

όπου w_t είναι μια σειρά λευκού θορύβου Gauss με μέση τιμή μηδέν και διακύμανση σ_w^2 και τα $(\theta_1, \theta_2, \dots, \theta_q)$ είναι οι παράμετροι του μοντέλου MA. Ένα από τα βασικά χαρακτηριστικά του μοντέλου MA(q) είναι η διατήρηση της στασιμότητας εφόσον δίνεται από το άθροισμα των όρων του λευκού θορύβου. Μία από τις κυριότερες διαφορές του μοντέλου κινούμενου μέσου με την αυτοπαλινδρούμενη διαδικασία είναι ότι στην δεύτερη το αιτιοκρατικό μέρος αντικαθίσταται από το στοχαστικό, και επομένως η μόνη πληροφορία που δίνεται για την x_t είναι από τις διαταράξεις στους $q+1$ πιο πρόσφατους χρόνους. Ο τελεστής της διαδικασίας του κινητού μέσου παρουσιάζεται στην παρακάτω σχέση:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q.$$

Όπως και η αυτοπαλινδρούμενη διαδικασία, η διαδικασία του κινούμενου μέσου είναι στάσιμη για οποιεσδήποτε τιμές των παραμέτρων $\theta_1, \dots, \theta_q$; λεπτομέρειες αυτού του αποτελέσματος δίνονται στην Ενότητα 3.3.1.

3.3.1 Στοχαστική διαδικασία πρώτης τάξης MA(1)

Το μοντέλο της σχέσης (3.6) για $q = 1$, MA(1), εκφράζεται ως

$$x_t = w_t + \theta w_{t-1}.$$

Επομένως για την διαδικασία αυτή ισχύει $\mathbb{E}[x_t] = 0$,

$$\gamma(h) = \begin{cases} (1 + \theta^2)\sigma_w^2, & h = 0, \\ \theta\sigma_w^2, & h = 1, \\ 0, & h > 1, \end{cases} \quad (3.7)$$

και συνάρτηση αυτοσυσχέτισης

$$\rho(h) = \begin{cases} \frac{\theta}{1+\theta^2}, & h = 1, \\ 0, & h > 1, \end{cases} \quad (3.8)$$

όπου $|\rho(1)| \leq 1/2$ για όλες τις τιμές του θ . Για κάποια υστέρηση h πολλαπλασιάζουμε τον τύπο της MA(1) με x_{t-h} και προκύπτει η παραπάνω συνάρτηση αυτοσυσχέτισης.

Παρατηρώντας την παραπάνω σχέση συμπεραίνουμε πως η μέγιστη τιμή είναι $\rho_1 = 0.5$ για $\theta = 1$ και η ελάχιστη $\rho_1 = -0.5$ για $\theta = -1$. Τέλος αξίζει να σημειωθεί ότι η σχέση που συνδέει την ρ_1 και το θ είναι γραμμική.

Παρακάτω παρατίθεται ένας πίνακας που παρουσιάζει την εκτίμηση της τάξης των AR και MA μοντέλων με την βοήθεια των ACF και PACF [14].

Διαδικασία	ACF	PACF
AR(1)	Εκθετική μείωση: θετικές τιμές αν $\phi > 0$, εναλλαγή πρόσημου ξεκινώντας από αρνητική τιμή, αν $\phi < 0$	Απότομος μηδενισμός της περιόδου. Η τιμή την περίοδο 1 είναι: θετική αν $\phi > 0$ ή αρνητική αν $\phi < 0$
AR(p)	Οι τιμές των συντελεστών του ACF φθίνουν προς το μηδέν ακολουθώντας εκθετική ή ημιτονοειδή πορεία	Μη μηδενικές τιμές για τις πρώτες p περιόδους και στην συνέχεια απότομος μηδενισμός
MA(1)	Απότομος μηδενισμός μετά την περίοδο 1. Η τιμή της περιόδου 1 είναι : θετική αν $\theta < 0$ ή αρνητική αν $\theta > 0$	Εκθετική μείωση: εναλλαγή προσήμου ξεκινώντας από θετική τιμή αν $\theta > 0$ και αρνητικές τιμές αν $\theta < 0$
MA(q)	Μη μηδενικές τιμές για τις πρώτες x περιόδους και στην συνέχεια απότομος μηδενισμός	Εκθετική μείωση ή φθίνουσα ημιτονοειδής συνάρτηση. Το ακριβές μέγεθος εξαρτάται από το πρόσημο και το μέγεθος του θ

Πίνακας 3.1: Εκτίμηση της τάξης των AR και MA με την βοήθεια των συναρτήσεων ACF και PACF

3.4 Αυτοπαλινδρούμενη διαδικασία κινούμενου μέσου (ARMA)

Όπως έγινε αντιληπτό και στις παραπάνω υποενότητες, μια χρονοσειρά μπορεί να περιγραφεί ικανοποιητικά με ένα AR μοντέλο ή ένα MA μοντέλο αν επιτρέψουμε οσοδήποτε μεγάλη τάξη. Πρακτικά θα θέλαμε να αποφύγουμε ένα μοντέλο AR ή MA με μεγάλη τάξη. Γι αυτό η λύση είναι να συνδυάσουμε τους δύο τύπους μοντέλων, βασιζόμενοι στο ότι μια χρονοσειρά μπορεί να αναλυθεί σε δύο μέρη, στο αιτιοκρατικό (το AR μέρος) και στο στοχαστικό (το MA μέρος) [13]. Φυσική επέκταση λοιπόν των μοντέλων τύπου AR και MA είναι ο συνδυασμός τους, τα αυτοπαλινδρομούμενα μοντέλα κινούμενου μέσου (Autoregressive Moving

Average, ARMA). Η γενική μορφή μίας διαδικασίας ARMA είναι

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q}, \quad (3.9)$$

3.5 Ολοκληρωμένο αυτοπαλινδρούμενο μοντέλο κινούμενου μέσου (ARIMA)

Τα αυτοπαλινδρούμενα μοντέλα κινούμενου μέσου ARMA(p, q), αν και εφαρμόζονται μόνο σε στάσιμες χρονοσειρές, μπορούν να χρησιμοποιηθούν και σε μη στάσιμες χρονοσειρές οι οποίες μετατρέπονται σε στάσιμες μέσω της μεθόδου διαφορίσης. Συνδυάζοντας την μέθοδο διαφορίσης για d επαναλήψεις μίας μη στάσιμης χρονοσειράς με το μοντέλο ARMA(p, q), προκύπτει το ολοκληρωμένο μικτό μοντέλο ή ολοκληρωμένο αυτοπαλινδρούμενο μοντέλο κινούμενου μέσου (Autoregressive Integrated Moving Average, ARIMA) που συμβολίζεται με ARIMA(p, d, q). Μια μη στάσιμη χρονοσειρά που εμφανίζει τάση, έστω $\{X_t\}_{t=1}^{\infty}$, διαφορίζεται d φορές μέχρι να προκύψει μία στάσιμη χρονοσειρά. Η στοχαστική διαδικασία που προκύπτει από τις πρώτες διαφορές είναι $X'_t = X_t - X_{t-1}$ για $d = 1$, $X'_t = X_t - 2X_{t-1} + X_{t-2}$ για $d = 2$ και ακολούθως για τις επόμενες τάξεις μέχρι να προκύψει στάσιμη χρονοσειρά. Η διαδικασία αυτή ορίζεται ως εξής:

$$X'_t = \nabla^d X_t$$

Αφού εξασφαλιστεί η στασιμότητα με τις απαιτούμενες διαφορές, ακολουθεί η ανάλυση με το μοντέλο ARMA(p, q) για την τροποποιημένη χρονοσειρά:

$$X'_t = \phi_0 + \phi_1 X'_{t-1} + \dots + \phi_p X'_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

όπου p η τάξη του αυτοπαλινδρουμού μέρους, d η τάξη διαφορίσης για την εξασφάλιση της στασιμότητας και q η τάξη του κινούμενου μέσου. Οι ιδιότητες της στασιμότητας και της αντιστρεψιμότητας όπως ισχύουν για τα μοντέλα ARMA, ισχύουν και για τα ARIMA μοντέλα. Τα ARIMA μοντέλα είναι η πιο γενική περίπτωση από τις στοχαστικές διαδικασίες που έχουν αναλυθεί. Όλα τα μοντέλα μπορούν να αναπαρασταθούν ως

ειδικές περιπτώσεις ARIMA μοντέλων με την επιλογή των κατάλληλων p , d και q . Χαρακτηριστικές περιπτώσεις είναι ο λευκός θόρυβος που αναπαρίσταται ως ARIMA(0, 0, 0), ο τυχαίος περίπατος που αναπαρίσταται ως ARIMA(0, 1, 0) αλλά και τα αυτοπαλινδρούμενα μοντέλα τάξης p ως ARIMA(p , 0, 0) και τα μοντέλα κινούμενου μέσου ως ARIMA(0, 0, q). Μία χρονοσειρά χαρακτηρίζεται ως μη στάσιμη όταν η τάση της δεν είναι σταθερή αλλά και όταν η διακύμανση δεν είναι σταθερή. Δηλαδή, ακόμα κι αν μία χρονοσειρά έχει σταθερή τάση δε σημαίνει πως είναι και στάσιμη. Θα πρέπει να γίνεται έλεγχος και ως προς τη διακύμανση. Στην περίπτωση που η χρονοσειρά χαρακτηρίζεται από αυξημένη διακύμανση θα πρέπει να μετατραπεί σε στάσιμη μέσω κατάλληλων μετασχηματισμών.

3.6 Ολοκληρωμένο αυτοπαλινδρομούμενο μοντέλο κινούμενου μέσου με εποχικότητα (SARIMA)

Το αυτοπαλινδρούμενο μοντέλο κινούμενου μέσου ARMA είναι μια από τις πιο συχνά χρησιμοποιημένες μεθόδους πρόβλεψης χρονοσειρών. Το μεγάλο πρόβλημα που παρουσιάζει το μοντέλο οφείλεται στο γεγονός ότι μπορεί να εφαρμοστεί μόνο σε χρονοσειρές οι οποίες έχουν τάση. Όταν όμως υπάρχει κάποια χρονοσειρά η οποία περιγράφει και τάση και περιοδικότητα τότε το ARMA μοντέλο δεν είναι αποδοτικό και δεν μπορεί να εφαρμοστεί. Η μη στασιμότητα μιας χρονοσειράς, πέρα από την τάση μπορεί να οφείλεται και σε εποχικότητα (ή αλλιώς περιοδικότητα με γνωστή περίοδο) δηλαδή να διαθέτει μια εποχική συνιστώσα που επαναλαμβάνεται κάθε s παρατηρήσεις (εποχικός κύκλος). Μάλιστα είναι δυνατόν να υπάρχει τάση όχι μόνο μεταξύ των συνεχόμενων παρατηρήσεων της χρονοσειράς αλλά και μεταξύ αντίστοιχων παρατηρήσεων σε συνεχόμενους εποχικούς κύκλους, η οποία ονομάζεται εποχική τάση [13]. Επίσης είναι πιθανό να υπάρχουν και περισσότερες από μία εποχικότητες στην ίδια χρονοσειρά. Το κατάλληλο μοντέλο για την αντιμετώπιση της εποχικότητας είναι το γενικό εποχικό πολλαπλασιαστικό μοντέλο (general seasonal multiplicative model) ή αλλιώς SARIMA(p , d , q)x(P , D ,

$Q)s$ (Seasonal Autoregressive Integrated Moving Average, SARIMA).

Για την κατασκευή ενός εποχικού μοντέλου πρέπει αρχικά να εντοπίσουμε την εποχικότητα s , γι' αυτό κατασκευάζουμε το γράφημα της αυτοσυσχέτισης η οποία θα φθίνει εκθετικά ή με κυματοειδή μορφή. Εάν κάποια υστέρηση $\tau = s$ εμφανίζει σημαντική αυτοσυσχέτιση σε σχέση με τις γειτονικές υστερήσεις, τότε θεωρούμε την εποχικότητα s για το μοντέλο μας. Σε περίπτωση που υπάρχει εποχικότητα, πρέπει να ελέγξουμε εάν χρειάζεται να πάρουμε εποχικές διαφορές D . Για το λόγο αυτό κοιτάμε το γράφημα της αυτοσυσχέτισης για όλους τους δυνατούς συνδυασμούς 0 ή 1 των d και D και κρατάμε τις τιμές που μας δίνουν στάσιμη χρονοσειρά, δηλαδή η αυτοσυσχέτιση των υστερήσεων που είναι πολλαπλάσια της εποχικότητας να μειώνεται εκθετικά. Σύμφωνα με το [7] δεν συνιστάται το D να είναι μεγαλύτερο του 1 ή το άθροισμα των d και D να είναι μεγαλύτερο του 2. Στη συνέχεια κάνουμε και τα γραφήματα της αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υστερήσεων $s, 2s, 3s, \dots$ και θεωρούμε ως Q τον αριθμό των σημαντικών αυτοσυσχετίσεων και ως P τον αριθμό των σημαντικών μερικών αυτοσυσχετίσεων. Το εποχικό μοντέλο $ARIMA(P, D, Q)s$ που βρήκαμε το προσαρμόζουμε στη χρονοσειρά μας και παίρνουμε τα υπόλοιπα a_t τα οποία θεωρούμε ως χρονοσειρά.

Κεφάλαιο 4

Πρόβλεψη Χρονοσειρών

4.1 Βασικά στάδια στην διαδικασία πρόβλεψης

Η πρόβλεψη μίας χρονοσειράς είναι από τις σημαντικότερες διεργασίες που εκτελούνται στον τομέα της στατιστικής ανάλυσης. Έχει αποκτήσει ιδιαίτερο ενδιαφέρον το τελευταίο διάστημα αφού ο μεγάλος όγκος των δεδομένων κάνει πραγματοποιήσιμη την επεξεργασία αυτών για την εξαγωγή αποτελεσμάτων και συμπερασμάτων για μελλοντικές χρονικές στιγμές. Η πρόβλεψη είναι ένα σημαντικό πρόβλημα που εκτείνεται σε πολλά πεδία, συμπεριλαμβανομένων των επιχειρήσεων και της βιομηχανίας, της οικονομίας, των περιβαλλοντικών επιστημών, της ιατρικής και των κοινωνικών επιστημών. Τα προβλήματα της πρόβλεψης ταξινομούνται ως βραχυπρόθεσμα, μεσοπρόθεσμα και μακροπρόθεσμα. Τα βραχυπρόθεσμα προβλήματα πρόβλεψης περιλαμβάνουν πρόβλεψη των γεγονότων μόνο μικρών χρονικών περιόδων (ημέρες, εβδομάδες, μήνες) στο μέλλον. Τα μεσοπρόθεσμα προβλήματα εκτείνονται από ένα έως δύο χρόνια στο μέλλον και τα μακροπρόθεσμα μπορούν να επεκταθούν σε περισσότερα από δύο χρόνια [2]. Τα βασικά στάδια σε μια διαδικασία πρόβλεψης είναι τα εξής [12]:

1ο στάδιο: Καθορισμός προβλήματος Αυτό το στάδιο μπορεί να θεωρηθεί το πιο δύσκολο και ταυτόχρονα το πιο σημαντικό μέρος στη

διαδικασία της πρόβλεψης. Αυτό συμβαίνει διότι θα πρέπει να καταστεί σαφής και κατανοητός ο τρόπος με τον οποίο θα χρησιμοποιηθούν οι προβλέψεις και από ποιους.

2ο στάδιο: Συγκέντρωση πληροφοριών Σε αυτό το βήμα απαιτούνται δύο είδη πληροφοριών. Το πρώτο είναι τα αριθμητικά δεδομένα και το δεύτερο η κρίση, η πείρα και η εμπειρία του προσωπικού που ασχολούνται με αυτή τη συλλογή για αυτό το χρονικό διάστημα. Οι παραπάνω πληροφορίες πρέπει να συλλεχθούν πριν ξεκινήσει η διαδικασία της πρόβλεψης.

3ο στάδιο: Προκαταρκτική ανάλυση Σε αυτό το στάδιο μας απασχολεί το είδος της πληροφορίας που αποκομίζεται από τα ακατέργαστα ιστορικά δεδομένα. Αρχικά, αναπαρίστανται γραφικά τα δεδομένα και στη συνέχεια υπολογίζονται οι βασικοί στατιστικοί δείκτες, όπως μέση τιμή, τυπική απόκλιση, ελάχιστη τιμή, μέγιστη τιμή και γραμμική τάση. Με αυτό τον τρόπο δημιουργείται μια αίσθηση των δεδομένων, δίνοντας απαντήσεις σε ερωτήματα όπως αν υπάρχουν πρότυπα κυκλικότητας, αν υπάρχει σημαντική τάση ή εποχικότητα και τέλος αν υπάρχουν ακραίες τιμές. Η ανάλυση αυτή μας οδηγεί στην οικογένεια μοντέλων πρόβλεψης που λογικά αναμένεται να μας προσδώσει ικανοποιητικές προβλέψεις.

4ο στάδιο: Επιλογή και προσαρμογή μοντέλου Εδώ γίνεται η επιλογή και ο καθορισμός των παραμέτρων διαφόρων ποσοτικών μοντέλων πρόβλεψης που έχουν επιλεγεί στο προηγούμενο στάδιο.

5ο στάδιο: Χρήση και αποτίμηση του μοντέλου πρόβλεψης Στο 5ο και τελευταίο στάδιο, εφόσον ένα μοντέλο έχει επιλεγεί υποκειμενικά και οι παράμετροι του έχουν προηγουμένως καθοριστεί, χρησιμοποιείται ώστε να πραγματοποιηθούν οι προβλέψεις. Κατά την εξέλιξη της διαδικασίας, γίνεται συνεχώς αποτίμηση των πλεονεκτημάτων και μειονεκτημάτων του μοντέλου και, εφόσον κριθεί απαραίτητο, επαναλαμβάνονται

ορισμένα βήματα στη διαδικασία.

4.2 Πρόβλεψη χρονοσειρών με την χρήση του μοντέλου ARIMA

Γενικότερα η επιλογή του μοντέλου είναι πολύ σημαντική αφού η καταλληλότητα αυτού είναι που θα κρίνει και θα εγκρίνει την εγκυρότητα της πρόβλεψης που θα πραγματοποιηθεί. Στο προηγούμενο κεφάλαιο (Κεφάλαιο 3), έγινε η ανάλυση του βασικού μοντέλου ARIMA το οποίο θα χρησιμοποιήσουμε ως μοντέλο για την πραγματοποίηση πρόβλεψης.

4.2.1 Πρόβλεψη χρονοσειρών με την βοήθεια της διαδικασίας Box-Jenkins

Η προσέγγιση των Box-Jenkins στην ανάλυση χρονοσειρών είναι μία μέθοδος εύρεσης ενός στατιστικού υποδείγματος ARIMA που να παριστάνει ικανοποιητικά τη στοχαστική διαδικασία από την οποία προήλθαν τα δεδομένα. Η μέθοδος αυτή εφαρμόζεται σε μη στάσιμες χρονοσειρές και απαρτίζεται από τέσσερα στάδια, την ταυτοποίηση, την εκτίμηση, το διαγνωστικό έλεγχο και την πρόβλεψη. Συνοψίζοντας, η μεθοδολογία Box-Jenkins είναι η ανάπτυξη και η κατασκευή υποδειγμάτων ARIMA, τα οποία είναι γραμμικά μοντέλα που μπορούν να περιγράψουν ικανοποιητικά τις συνιστώσες μίας χρονοσειράς.

Ταυτοποίηση

Στο στάδιο όπου πρέπει να γίνει ταυτοποίηση του υποδείγματος πρέπει αρχικά να προσδιοριστεί η τάξη της μη στασιμότητας και η τάξη των AR πολυωνύμων. Για την επίτευξη αυτού του βήματος πρέπει να συγκρίνουμε τη μορφή των δειγματικών συναρτήσεων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης με τη μορφή των θεωρητικών συναρτήσεων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης που αντιστοιχούν σε διαδικασίες με άπειρο πλήθος όρων. Σε αυτό το στάδιο γίνεται η εξειδίκευση

ενός ARIMA υποδείγματος με βάση τις πληροφορίες που λαμβάνονται από το δείγμα. Αυτό σημαίνει ότι καθορίζονται οι τιμές των d , p και q . Επομένως, καθορίζεται ο αριθμός d των διαφορών που απαιτούνται για να μετατραπεί η σειρά σε στάσιμη και στη συνέχεια καθορίζεται η τάξη p της αυτοπαλινδρούμενης διαδικασίας και η τάξη q της διαδικασίας κινούμενου μέσου. Για να διαπιστωθεί η στασιμότητα της σειράς, εξετάζεται η συμπεριφορά της δειγματικής συνάρτησης αυτοσυσχετίσης. Στην περίπτωση που οι αυτοσυσχετίσεις συγκλίνουν με γρήγορο ρυθμό προς το μηδέν, σημαίνει ότι η σειρά είναι μάλλον στάσιμη. Αντίθετα, αν οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό, είναι ένδειξη ότι η σειρά είναι μη στάσιμη και πρέπει να μετατραπεί σε στάσιμη. Η μετατροπή σε αυτή την περίπτωση γίνεται παίρνοντας τις πρώτες ή τις δεύτερες διαφορές. Μετά την μετατροπή προσδιορίζεται η τάξη του υποδείγματος ARIMA, δηλαδή προσδιορίζονται οι τιμές των p και q .

Εκτίμηση

Το 2ο στάδιο στη μεθοδολογία Box-Jenkins είναι αυτό της εκτίμησης των παραμέτρων $\alpha_1, \alpha_2, \dots, \alpha_p$ του αυτοπαλινδρούμενου μέρους καθώς και των παραμέτρων $\theta_1, \theta_2, \dots, \theta_q$ του κινούμενου μέσου μέρους του ARIMA(p, d, q) υποδείγματος που ταυτοποιήθηκε στο προηγούμενο στάδιο. Η εκτίμηση ενός υποδείγματος ARIMA(p, d, q) προϋποθέτει την ύπαρξη ενός μεγάλου δείγματος παρατηρήσεων Y_1, Y_2, \dots, Y_N της χρονοσειράς Y_t . Αυτό το υπόδειγμα ισοδυναμεί με ένα στάσιμο ARMA(p, q) στις διαφορές d τάξεως της Y_t . Για λόγους απλούστευσης μπορεί να θεωρηθεί ότι η χρονοσειρά Y_t είναι στάσιμη αναλύοντας τις μεθόδους εκτίμησης των παραμέτρων ARMA(p, q). Η δυσκολία εκτίμησης των παραμέτρων ARMA(p, q) υποδείγματος εξαρτάται από την τάξη των p και q . Αν πρόκειται για αυτοπαλινδρούμενη σειρά, οι παράμετροι της μπορούν να εκτιμηθούν με τη μέθοδο των ροπών. Όσον αφορά τα υποδείγματα MA ή τις μικτές μορφές, εφαρμόζονται πιο σύνθετες στατιστικές μέθοδοι λόγω της παρουσίας λαθών προηγούμενων περιόδων τα οποία δεν είναι γνωστά. Οι πιο γνωστές μέθοδοι εκτίμησης στην κατηγορία των γραμμικών υποδειγμάτων παλινδρόμησης είναι η μέθοδος ελάχιστων

τετραγώνων (Least Squares, LS) και η μέθοδος μέγιστης πιθανοφάνειας (Maximum Likelihood, ML).

Διαγνωστικός Έλεγχος

Στο στάδιο του διαγνωστικού ελέγχου προσπαθούμε να ελέγξουμε αν το υπόδειγμα που εκτιμήθηκε είναι το καλύτερο. Αν δηλαδή δεν υπάρχει άλλο υπόδειγμα διαφορετικής μορφής, βαθμού ή παραμέτρων που να υπερτερεί του αρχικού υποδείγματος που εκτιμήθηκε σε ότι αφορά τη συμπεριφορά των σφαλμάτων ή την προβλεπτική του ικανότητα. Σε ένα αποτελεσματικό υπόδειγμα τα κατάλοιπα χαρακτηρίζονται καθαρά (λευκός θόρυβος) δηλαδή κατανέμονται ομοιόμορφα με μηδενική μέση τιμή και διακύμανση και επομένως δεν περιέχουν χρήσιμες πληροφορίες.

Μεταδιάγνωση

Στη διαδικασία της προσαρμογής γίνεται έλεγχος της καταλληλότητας του υποδείγματος που έχει εκτιμηθεί συγκρίνοντας το με ένα υπόδειγμα μεγαλύτερης τάξης. Το υπόδειγμα $ARIMA(p, q)$ που εκτιμήθηκε συγκρίνεται με τα υποδείγματα $ARIMA(p + 1, d, q)$ και $ARIMA(p, d, q + 1)$. Αν το υπόδειγμα που εκτιμήθηκε περιγράφει τη διαδικασία από την οποία παράχθηκαν τα δεδομένα μας και στα μεγαλύτερα υποδείγματα οι επιπλέον συντελεστές είναι μηδέν, τότε το υπόδειγμα αυτό είναι το καταλληλότερο για τα δεδομένα μας. Αν οι συντελεστές δεν είναι μηδέν, τότε υπάρχει άλλο, καταλληλότερο υπόδειγμα για τα δεδομένα μας από αυτό που έχει εκτιμηθεί. Παρακάτω αναφέρονται δύο κριτήρια που βοηθούν στην επιλογή του κατάλληλου υποδείγματος, τα οποία χρησιμοποιούνται ευρέως στην ανάλυση των χρονοσειρών. Τα κριτήρια αυτά είναι το κριτήριο πληροφοριών **Akaike (Akaike Information Criterion, AIC)** και το κριτήριο πληροφοριών του **Bayes (Bayesian Information Criterion, BIC)** και ορίζονται ως εξής:

$$AIC = \log(\hat{\sigma}_k^2) + \frac{n + 2k}{n},$$
$$BIC = \log(\hat{\sigma}_k^2) + \frac{k \log n}{n}$$

όπου σ_z^2 είναι η εκτίμηση της διακύμανσης των καταλοίπων, n ο αριθμός εκτιμώμενων παραμέτρων του υποδείγματος $(p + q + 1)$ όπου η μονάδα αντιστοιχεί στην σταθερά. Αν αυξηθεί η τάξη του υποδείγματος προσθέτοντας υστερήσεις είτε για το αυτοπαλινδρούμενο μέρος AR, είτε για το κινούμενου μέσου μέρος MA, μειώνεται το άθροισμα των τετραγώνων των καταλοίπων και ταυτόχρονα μειώνονται οι βαθμοί ελευθερίας λόγω της εκτίμησης περισσότερων παραμέτρων. Με την προσθήκη μιας επιπλέον μεταβλητής στο υπόδειγμα μειώνεται το άθροισμα των τετραγώνων των καταλοίπων, οπότε μειώνεται και η διακύμανση, ταυτόχρονα όμως αυξάνεται η τιμή του n στα κριτήρια AIC και BIC. Η επιλογή των υποδειγμάτων γίνεται με βάση την μικρότερη τιμή των κριτηρίων. Η εφαρμογή αυτών των κριτηρίων χρησιμεύει στην επιλογή του καταλληλότερου μοντέλου ARIMA από πλευράς αριθμού υστερήσεων που θα πρέπει να περιληφθούν.

4.2.2 Πρόβλεψη με αυτοπαλινδρούμενα μοντέλα AR(p)

Εάν η χρονοσειρά x_1, x_2, \dots, x_n αναλύεται με τον καλύτερο δυνατό τρόπο μέσω της διαδικασίας AR τότε το αντίστοιχο μοντέλο για x_{n+1} θα είναι

$$x_{n+1} = \phi_1 x_n + \dots + \phi_p x_{n-p+1} + z_{n+1} \quad (4.1)$$

και η καλύτερη δυνατή πρόβλεψη για ένα χρονικό βήμα μπροστά δίνεται από την παρακάτω εξίσωση

$$x_n(1) = \phi_1 x_n + \dots + \phi_p x_{n-p+1}. \quad (4.2)$$

Επομένως για k χρονικά βήματα μπροστά η πρόβλεψη θα μπορεί να προσδιοριστεί από την σχέση

$$x_n(k) = \phi_1 x_n(k-1) + \dots + \phi_p x_n(k-p) \quad (4.3)$$

όπου κάθε τιμή $x_n(k)$ είναι γνωστή από προηγούμενη πρόβλεψη ή είναι γνωστή καθώς δίνεται από την χρονοσειρά. Η πρόβλεψη που κάνουμε συνιστάται μέσω του αιτιοκρατικού μέρους του μοντέλου, όπου οι παρατηρήσεις οι οποίες δεν μας είναι γνωστές αντικαθιστώνται από τις προβλέψεις μας [13].

Κατά την διαδικασία πρόβλεψης υπάρχει ένα σφάλμα για τα ίδια χρονικά βήματα που εκφράζεται ως ο γραμμικός συνδυασμός των στοιχείων του λευκού θορύβου αλλά μόνο στις χρονικές στιγμές $n+1, \dots, n+k$ [2]:

$$e_n(k) = \sum_{j=0}^{k-1} b_j z_{n+k-j}. \quad (4.4)$$

Έτσι το μοντέλο θα έχει διασπορά

$$\text{Var}[e_n(k)] = \sigma_z^2 \sum_{j=0}^{k-1} b_j^2. \quad (4.5)$$

4.2.3 Πρόβλεψη με μοντέλο κινούμενου μέσου MA(q)

Για να είναι εφικτή η πρόβλεψη με την βοήθεια του μοντέλου κινούμενου μέσου πρέπει αρχικά να προσδιορίσουμε το κατάλληλο μοντέλο για την χρονοσειρά x_1, x_2, \dots, x_n [13]. Έτσι η επόμενη χρονικά παρατήρηση θα δίνεται ως

$$x_{n+1} = z_{n+1} + \theta_1 z_n + \dots + \theta_q z_{n-q+1} \quad (4.6)$$

και, επομένως, η καλύτερη δυνατή πρόβλεψη για ένα χρονικό βήμα μετά θα είναι

$$x_n(1) = \theta_1 z_n + \dots + \theta_q z_{n-q+1}. \quad (4.7)$$

Γενικά, για τυχαία k χρονικά βήματα η εκτίμηση δίνεται από την ακόλουθη σχέση

$$x_n(k) = \theta_k z_n + \theta_{k+1} z_{n-1} + \dots + \theta_q z_{n-q+k}, k \leq 0 \quad (4.8)$$

με

$$x_n(k) = 0$$

Τα σφάλματα z_n, z_{n-1}, z_{n-2} που μπορεί να προκύψουν μπορούν να υπολογιστούν με την βοήθεια των παρατηρήσεων x_n, x_{n-1}, x_{n-2} με την προϋπόθεση ότι οι αρχικές τους τιμές z_1, z_2, \dots, z_q είναι μηδέν. Πιο συγκεκριμένα ο υπολογισμός των $z_{q+1}, z_{q+2}, \dots, z_q$ γίνεται μέσω της λύσης της εξίσωσης του MA(q), όπου θέτουμε $t = q$ και κάνουμε το ίδιο για τους χρόνους $t = q+1, \dots, n-1$.

4.2.4 Πρόβλεψη με αυτοπαλινδρούμενα μοντέλα κινούμενου μέσου ARMA

Το αυτοπαλινδρούμενο μοντέλο κινούμενου μέσου συμβολίζεται ως $\text{ARMA}(p, q)$. Θεωρώντας έτσι μια χρονοσειρά x_1, x_2, \dots, x_n , τότε το αμέσως επόμενο χρονικό βήμα θα είναι

$$x_{n+1} = \phi_1 x_n + \dots + \phi_p x_{n-p+1} + \theta_1 z_n + \dots + \theta_q z_{n-q+1}. \quad (4.9)$$

Έτσι για k χρονικά βήματα μπροστά ισχύει ότι εάν $k \leq q$, τότε

$$x_n(k) = \phi_1 x_n(k-1) + \dots + \phi_p x_n(k-p) + \theta_k z_n + \dots + \theta_q z_{n-q+1}, \quad (4.10)$$

ενώ αν $k > q$, τότε

$$x_n(k) = \phi_1 x_n(k-1) + \dots + \phi_p x_n(k-p). \quad (4.11)$$

Έτσι προκύπτει πως η πρόβλεψη με την βοήθεια του μοντέλου ARMA αποτελεί έναν συνδυασμό των προβλέψεων με τα μοντέλα AR και MA [1].

4.2.5 Πρόβλεψη μη στάσιμων χρονοσειρών με γραμμικά μοντέλα

Οι προβλέψεις στη χρονοσειρά που κάναμε στάσιμη από μια μη στάσιμη χρονοσειρά θα πρέπει να μετασχηματιστούν κατάλληλα για να αναφέρονται στην αρχική χρονοσειρά. Όταν λοιπόν η χρονοσειρά δεν είναι στάσιμη για να εφαρμόσουμε την πρόβλεψη με τα μοντέλα της προηγούμενης παραγράφου πρέπει να κάνουμε τα εξής βήματα:

1. να μετασχηματίσουμε τη χρονοσειρά σε στάσιμη: μη στάσιμη $x_t \rightarrow y_t$ στάσιμη,
2. να κάνουμε την πρόβλεψη του y_{n+k} με κάποιο μοντέλο, π.χ. AR, έστω $y_n(k)$,
3. να μετασχηματίσουμε την πρόβλεψη $y_n(k)$ για την στάσιμη χρονοσειρά στην πρόβλεψη $x_n(k)$ για την αρχική μη στάσιμη χρονοσειρά.

Υποθέτουμε πως η χρονοσειρά x_1, x_2, \dots, x_n δεν είναι στάσιμη. Το κλασικό μοντέλο για το x_t είναι

$$x_t = \mu_t + s_t + y_t, \quad (4.12)$$

όπου μ_t είναι η συνάρτηση τάσης, s_t είναι η περιοδική ή εποχική συνάρτηση και y_t είναι μια στάσιμη χρονοσειρά απαλλαγμένη από τάσεις και περιοδικότητες. Σκοπός είναι να βρούμε την πρόβλεψη για ένα χρονικό βήμα ή γενικά για k χρονικά βήματα μπροστά, δηλαδή να προβλέψουμε το x_{n+k} που ορίζεται ως

$$x_{n+k} = \mu_{n+k} + s_{n+k} + y_{n+k}. \quad (4.13)$$

Αν διαλέξουμε να εκτιμήσουμε τα μ_t και s_t ως συναρτήσεις του χρόνου t , τότε μπορούμε να επεκτείνουμε τις εκτιμήσεις στο χρόνο $n+k$ για να βρούμε τα μ_{n+k} και s_{n+k} . Αφαιρώντας από τις τιμές x_1, x_2, \dots, x_n , τις εκτιμήσεις της τάσης και περιοδικότητας προκύπτουν οι τιμές y_1, y_2, \dots, y_n , και έτσι έχουμε εφαρμόσει το πρώτο βήμα. Για την υλοποίηση του δεύτερου βήματος προβλέπουμε το y_{n+k} χρησιμοποιώντας κάποιο μοντέλο τύπου AR, MA ή ARMA. Η πρόβλεψη $x_n(k)$ προκύπτει απευθείας από την πρόβλεψη $y_n(k)$ και τις επεκτάσεις τάσης και περιοδικότητας μ_{n+k} και s_{n+k} (τρίτο βήμα)

$$x_n(k) = \mu_{n+k} + s_{n+k} + y_n(k). \quad (4.14)$$

Αν διαλέξουμε να απαλείψουμε τα μ_t και s_t χρησιμοποιώντας διαφορές τότε στην ουσία αυτή η πρόβλεψη με τα παραπάνω τρία βήματα είναι η πρόβλεψη με μοντέλα ARIMA ή SARIMA. Οι γενικοί τύποι για τις προβλέψεις με αυτά τα μοντέλα είναι πολύπλοκοι αλλά κάποιος μπορεί να καταλάβει πως γίνεται η πρόβλεψη με ARIMA μοντέλο θεωρώντας το ARIMA($p, 1, q$). Παίρνοντας τις πρώτες διαφορές με υστέρηση ένα, από την αρχική χρονοσειρά x_1, x_2, \dots, x_n προκύπτει η χρονοσειρά y_2, y_3, \dots, y_n , όπου $y_t = x_t - x_{t-1}$, σχηματίζοντας έτσι το πρώτο βήμα της διαδικασίας πρόβλεψης μη στάσιμων χρονοσειρών. Εφαρμόζοντας το μοντέλο ARMA(p, q) στην y_2, y_3, \dots, y_n βρίσκουμε την πρόβλεψη για ένα

χρονικό βήμα $y_n(1)$ (δεύτερο βήμα) και η πρόβλεψη για την αρχική χρονοσειρά (τρίτο βήμα) είναι

$$x_n(1) = x_n + y_n(1). \quad (4.15)$$

Το σφάλμα πρόβλεψης του x_{n+1} , $e_n(1)$, είναι το ίδιο με το σφάλμα πρόβλεψης του y_{n+1} . Γενικά η πρόβλεψη για k χρονικά βήματα είναι

$$x_n(k) = x_n(k-1) + y_n(k), \quad (4.16)$$

όπου $y_n(k)$ είναι η πρόβλεψη του y_{n+k} με το μοντέλο $\text{ARMA}(p, q)$ και το $x_n(k-1)$ είναι γνωστό από την πρόβλεψη του x_{n+k-1} . Για $\text{ARIMA}(p, d, q)$ ή $\text{SARIMA}(p, d, q)(P, D, Q)$ s η διαδικασία της πρόβλεψης του x_{n+1} είναι παρόμοια, δηλαδή βρίσκουμε την πρόβλεψη $y_n(1)$ με μοντέλο τύπου ARMA και την προσθέτουμε στην κατάλληλη έκφραση των τελευταίων παρατηρήσεων x_n, x_{n-1}, \dots σύμφωνα με τις τιμές των d και D . Η πρόβλεψη για k χρονικά βήματα μπορεί να υπολογισθεί αναδρομικά.

4.2.6 Ανάλυση ακρίβειας μοντέλου για την εκτίμηση χρονοσειρών

Η ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error, RMSE), το μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error, MAPE) και το μέσο απόλυτο σφάλμα (Mean Absolute Error, MAE) υιοθετούνται για την αξιολόγηση της ακρίβειας των προσεγγίσεων. Τα μέτρα επιβεβαίωσης αποτελούν ένα σημαντικό ικανό μέσο αξιολόγησης των αποτελεσμάτων που έχουν προκύψει από κάποια πρόβλεψη που έχει υλοποιηθεί.

Το μέσο απόλυτο σφάλμα MAE είναι ένα κοινό μέτρο του σφάλματος πρόβλεψης στην ανάλυση χρονοσειρών, το οποίο μετρά το μέσο μέγεθος των σφαλμάτων σε ένα σύνολο προβλέψεων

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - y_t|, \quad (4.17)$$

όπου n είναι ο αριθμός των παρατηρήσεων στο σύνολο των περιόδων αξιολόγησης, x_t είναι η τιμή της παρατήρησης τη χρονική στιγμή t και

y_t είναι η τιμή πρόβλεψης. Η Εξίσωση (4.17) δείχνει ότι το μέσο τετραγωνικό σφάλμα MAE είναι ο μέσος όρος επί των απόλυτων τιμών των αποκλίσεων μεταξύ της πρόβλεψης και της αντίστοιχης παρατήρησης. Το MAPE υπολογίζεται ως το μέσο απόλυτο ποσοστιαίο σφάλμα:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - y_t}{x_t} \right|, \quad (4.18)$$

Όπως φαίνεται στην Εξίσωση (4.18), ο κύριος σκοπός του μέσου α-πολύτου ποσοστιαίου σφάλματος MAPE είναι να δείξει αν τα δεδομένα είναι σταθερά (η διακύμανση είναι μικρή). Αυτός είναι ο λόγος για τον οποίο το μέσο ποσοστιαίο σφάλμα MAPE είναι σημαντικό στην πρόβλεψη της αιολικής ενέργειας.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - y_t)^2} \quad (4.19)$$

Η παραπάνω Εξίσωση(4.19) δείχνει ότι η ρίζα του μέσου τετραγωνικού σφάλματος RMSE είναι ένας τετραγωνικός κανόνας βαθμολόγησης, ο οποίος μετρά το μέσο μέγεθος του σφάλματος. Η διαφορά μεταξύ των προβλέψεων και των αντίστοιχων παρατηρούμενων τιμών τετραγωνίζεται, αθροίζεται και στη συνέχεια υπολογίζεται ο μέσος όρος για τον αριθμό των δειγμάτων. Τέλος, λαμβάνεται η τετραγωνική ρίζα του μέσου όρου. Δεδομένου ότι τα σφάλματα τετραγωνίζονται πριν από τον υπολογισμό του μέσου όρου, η ρίζα του μέσου τετραγωνικού σφάλματος RMSE δίνει σχετικά υψηλή βαρύτητα στα μεγάλα σφάλματα. Αυτό σημαίνει ότι το RMSE είναι πιο χρήσιμο όταν τα μεγάλα σφάλματα είναι ανεπιθύμητα.

Κεφάλαιο 5

Ανάλυση Χρονοσειρών Ταχύτητας Ανέμου

Σκοπός της διπλωματικής εργασίας είναι η πρόβλεψη σε ταχύτητες ανέμου για ένα χρόνο, για μια ημέρα και σε δεδομένα που αντιστοιχούν σε λεπτά ή ώρες της ημέρας με την βοήθεια των μοντέλων ARIMA. Τα δεδομένα προέρχονται από την ιστοσελίδα SODA-PRO.

5.1 Περιοχή μελέτης και περιγραφή των δεδομένων

Οι περιοχές μελέτης στις οποίες απευθύνονται όλες οι αναλύσεις που πραγματοποιούνται στην διπλωματική εργασία αυτή είναι στο νησί της Κύπρου. Η Κύπρος είναι το τρίτο μεγαλύτερο νησί της Μεσογείου, μετά τη Σικελία και τη Σαρδηνία. Βρίσκεται ανατολικά της Ρόδου, 215 μίλια, και της Κρήτης, νότια των ακτών της Τουρκίας, 37 μίλια, δυτικά των Συριακών ακτών, 56 μίλια, και βόρεια των ακτών της Αιγύπτου, 190 μίλια. Από μορφολογικής άποψης η Κύπρος υποδιαιρείται στις ακόλουθες μορφολογικές περιφέρειες: στο ορεινό σύμπλεγμα Τρόοδος που δεσπόζει στο κεντρο-νοτιοδυτικό τμήμα του νησιού, με ψηλότερη κορυφή τον Όλυμπο (1.953 μ.), στη βόρεια οροσειρά του Πενταδάκτυλου, με ψηλότερη κορυφή τον Κυπαρισσόβουνο (1.024 μ.), στην κεντρική πεδιάδα της Μεσαορίας, που εκτείνεται μεταξύ των δύο προηγούμενων και την ο-

ποία διασχίζουν δυο ποταμοί-χείμαρροι, ο Πεδιαίος και ο Γιαλιάς, στην λοφώδη περιοχή γύρω από το σύμπλεγμα του Τροόδους και τέλος στις παράκτιες πεδιάδες και ακτές. Συγκεκριμένα η μελέτη για την ανάλυση των δεδομένων γίνεται για τις περιοχές του Τροόδους και την περιοχή του Αγρού. Το ορεινό σύμπλεγμα Τρόδος είναι η κύρια οροσειρά της Κύπρου με κορυφές που φτάνουν σε υψόμετρο τα 1954 μέτρα. Ο ορεινός όγκος του Τροόδους καταλαμβάνει το κεντρικό και νοτιοδυτικό μέρος του νησιού, με κορυφογραμμή που εκτείνεται σε κατεύθυνση από τα δυτικά-βορειοδυτικά προς τα ανατολικά-νοτιοανατολικά. Η περιοχή Αγρός, ένα χωριό αμφιθεατρικά κτισμένο στις πλαγιές του όρους Τρόδος, βρίσκεται στην περιοχή της Πιτσιλιάς και σε υψόμετρο 1100 μέτρων.

Τα δεδομένα αποκτήθηκαν από την ιστοσελίδα SODA PRO. Η Soda παρέχει υπηρεσίες πρόσβασης σε ιστορικά μετεωρολογικά δεδομένα από το 1980 μέχρι σήμερα και υπηρεσίες πρόβλεψης μετεωρολογικών δεδομένων. Το σύνολο των δεδομένων που αποκτήθηκαν από τη βάση αποτελείται από 52704 μετρήσεις από την 1η Ιανουαρίου του 2020 έως τις 31 Δεκεμβρίου του 2020. Τα δεδομένα αυτά είναι διαχωρισμένα ανά 10 λεπτά της ώρας. Αρχικά χρησιμοποιείται η χρονοσειρά με τα δεδομένα για ολόκληρο τον χρόνο με βήμα 10 λεπτών. Στην συνέχεια γίνεται διαχωρισμός της χρονοσειράς αυτής σε δύο χρονικές περιόδους, την καλοκαιρινή περίοδο από 10 Μαΐου έως 10 Οκτωβρίου 2020 και την χειμερινή περίοδο από 1 Δεκεμβρίου 2019 έως 29 Φεβρουαρίου 2020. Επιπλέον, για την σύγκριση των δεδομένων και των αποτελεσμάτων έγιναν αναλύσεις και προβλέψεις για την χρονοσειρά ταχύτητας ανέμου με βήμα ανά ώρα για τις ίδιες περιοχές μειώνοντας τα δεδομένα από 52704 σε 8784 ακολουθώντας την ίδια διαδικασία. Ο διαχωρισμός της χρονοσειράς γίνεται για να αντιμετωπιστεί ο μεγάλος όγκος των δεδομένων των χρονοσειρών με χρονικό βήμα 10 λεπτά της ώρας και με χρονικό βήμα μία ώρα άλλα και για να γίνει εφικτή η σύγκριση των αποτελεσμάτων της πρόβλεψης στις δύο διαφορετικές περιόδους. Προφανώς τα αποτελέσματα προσδοκούνται ότι θα είναι διαφορετικά για τις δύο περιόδους λόγω των διαφορετικών ταχυτήτων ανέμου που επικρατούν.

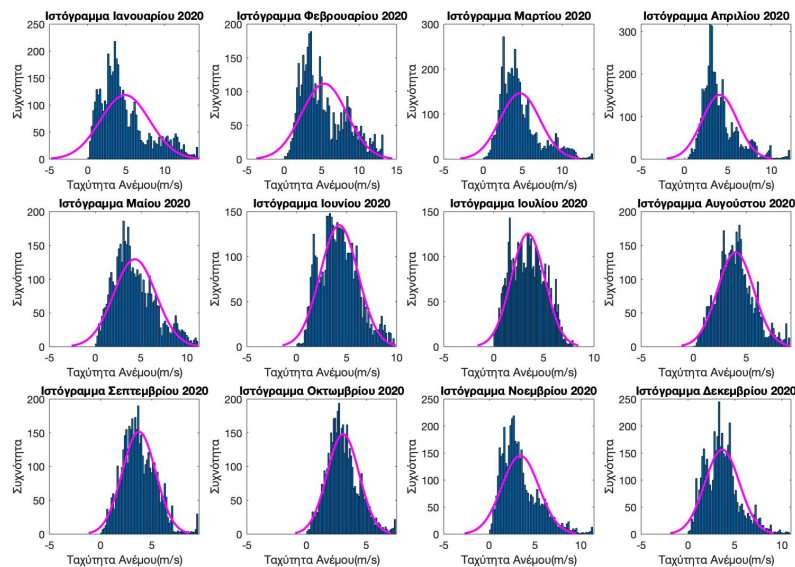
5.2 Στατιστική ανάλυση δεδομένων

Για την ανάλυση των δεδομένων και την εφαρμογή των μεθόδων πρόβλεψης απαιτείται η στατιστική ανάλυση των δεδομένων που έχουμε στην διάθεσή μας. Η ανάλυση αυτή γίνεται με την βοήθεια ιστογραμμάτων, βασικών στατιστικών μέτρων, διαγραμμάτων πιθανοτήτων κ.α. Η στατιστική ανάλυση παρουσιάζεται αρχικά για την χρονοσειρά με δεκάλεπτο βήμα της περιοχής του Τροόδους και στη συνέχεια για την αντίστοιχη της περιοχής του Αγρού. Επιπλέον, παρουσιάζεται η ανάλυση των χρονοσειρών με ωριαίο βήμα για τις δύο υπό μελέτη περιοχές.

5.2.1 Περιοχή του Τροόδους

Χρονοσειρές με δεκάλεπτο βήμα

Από την σύγκριση των ιστογραμμάτων των δεδομένων με τις καμπύλες της Συνάρτησης Πυκνότητας Πιθανότητας (ΣΠΠ) και της Αθροιστικής Συνάρτησης Κατανομής (ΑΣΚ) της κανονικής κατανομής μπορούμε να καταλήξουμε σε χρήσιμα συμπεράσματα σχετικά με την στοχαστική συμπεριφορά των δεδομένων. Στο Σχήμα 5.1 που ακολουθεί παρουσιάζονται τα ιστογράμματα της χρονοσειράς με δεκάλεπτο βήμα για κάθε μήνα του έτους 2020 ξεχωριστά. Επιπλέον, για την σύγκριση, δίνεται η καμπύλη της ΣΠΠ της κανονικής κατανομής που έχει σχήμα "καμπάνας" και η κατανομή αυτή είναι συμμετρική γύρω από την μέση τιμή της.

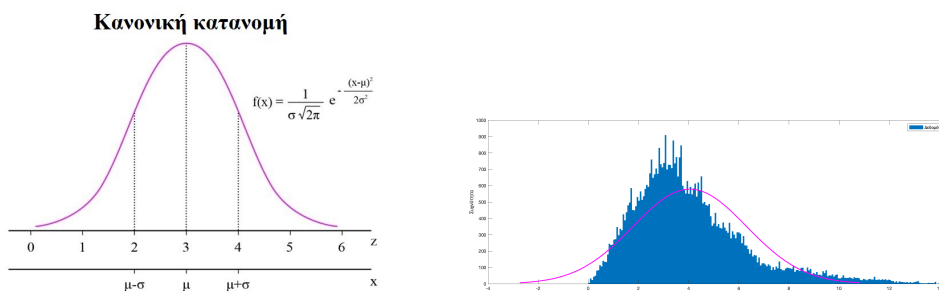


Σχήμα 5.1: Ανάλυση δεδομένων (ιστογράμματα) μηνιαίων χρονοσειρών με δεκάλεπτο βήμα για την περιοχή του Τροόδους για το έτος 2020.

Από τα ιστογραμμάτα των μηνιαίων χρονοσειρών με δεκάλεπτο βήμα καταλήγουμε στο συμπέρασμα ότι οι περισσότερες χρονοσειρές αποκλίνουν από την κανονική κατανομή. Συγκεκριμένα, μεγαλύτερη απόκλιση παρατηρείται τους μήνες Ιανουάριο, Φεβρουάριο, Μάρτιο, Απρίλιο, Μάιο, Νοέμβριο και Δεκέμβριο.

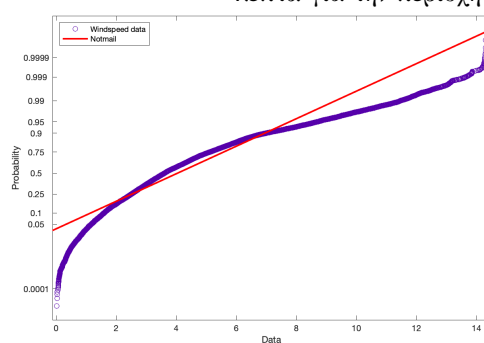
Το ιστόγραμμα των δεδομένων της ετήσιας χρονοσειράς με δεκάλεπτο βήμα για την περιοχή δίνεται στο Σχήμα 5.2. Η γραφική παράσταση της ΣΠΠ της κανονικής κατανομής δίνεται στο Σχήμα 5.2α'. Το διάγραμμα της κανονικής πιθανότητας που φαίνεται πιο κάτω δείχνει την κατανομή των δεδομένων σε σύγκριση με την κανονική κατανομή για την χρονοσειρά με δεκάλεπτο βήμα για την περιοχή του Τροόδους για το έτος 2020.

Παρατηρώντας τα πιο πάνω διαγράμματα γίνεται αντιληπτό πως οι διακυμάνσεις που παρουσιάζει το ιστόγραμμα της χρονοσειράς με βήμα 10 λεπτών μας οδηγεί στο συμπέρασμα ότι η χρονοσειρά αποκλίνει από την κανονική κατανομή. Όσον αφορά το διάγραμμα κανονικής κατανο-



(α) Διάγραμμα κανονικής κατανομής.

(β) Ιστόγραμμα ετήσιας χρονοσειράς ανά 10 λεπτά για την περιοχή του Τροόδους.



(γ) Διάγραμμα κανονικής κατανομής πιθανότητας για την περιοχή Τροόδους ανά 10 λεπτά για το έτος 2020.

Σχήμα 5.2: Ανάλυση κανονικής κατανομής (Ιστόγραμμα, διάγραμμα κανονικής πιθανότητας) ετήσιας χρονοσειράς ανά 10 λεπτά για την περιοχή του Τροόδους.

μής πιθανότητας (Σχήμα 5.2γ) είναι προφανής η μη γραμμικότητα των δεδομένων και κατ' επέκταση η απόκλιση που παρουσιάζουν αυτά από την κανονική κατανομή.

Τέλος, για να συμπληρωθεί η στατιστική ανάλυση της χρονοσειράς με δεκάλεπτο χρονικό βήμα για το έτος 2020 της περιοχής του Τροόδους παραθέτονται κάποια βασικά στατιστικά χαρακτηριστικά στον Πίνακα 5.1 που ακολουθεί.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.5
Τυπική απόκλιση (m/s)	2.26
Διάμεση τιμή (m/s)	3.61
Κύρτωση	4.78
Λοξότητα	1.19
Μέγιστη τιμή (m/s)	14.3
Ελάχιστη τιμή (m/s)	0.02

Πίνακας 5.1: Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά ταχύτητας ανέμου με βήμα δειγματοληψίας 10 λεπτά για την περιοχή του Τροόδους.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.1 συμπεραίνουμε ότι η τιμή της κύρτωσης και η τιμή της λοξότητας αποκλίνουν αρκετά από τις τιμές που αντιστοιχούν στην κανονική κατανομή. Εάν η χρονοσειρά μας δεν απέκλινε από την κανονική κατανομή, η τιμή της κύρτωσης θα έπρεπε να πλησιάζει το 3 και η τιμή της λοξότητας θα έπρεπε να ήταν κοντά στο 0.

Καλοκαιρινή και χειμερινή περίοδος

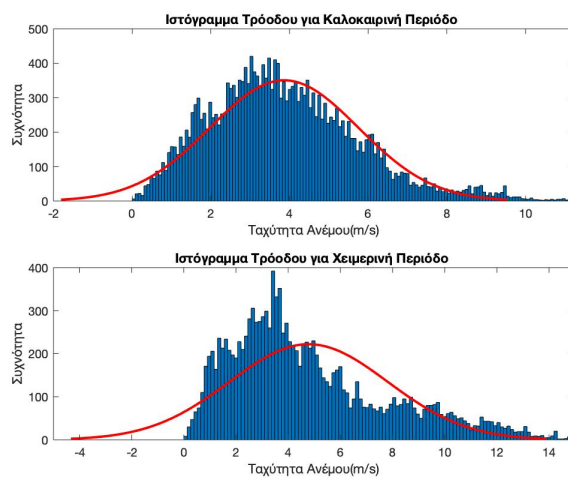
Για την περαιτέρω ανάλυση των δεδομένων των οποίων η κατανομή αποκλίνει από την κανονική, αλλά και για να μειωθεί το μεγάλο μέγεθος της χρονοσειράς, έγινε διαχωρισμός της σε καλοκαιρινή περίοδο (10 Μαΐου - 10 Οκτωβρίου 2020) και χειμερινή περίοδο (1 Δεκεμβρίου 2019 - 29 Φεβρουαρίου 2020). Έτσι τα 52704 δείγματα μειώθηκαν σε 22176 για την καλοκαιρινή και 13104 για την χειμερινή περίοδο. Στο παρακάτω Σχήμα 5.3 φαίνονται τα ιστογράμματα για την καλοκαιρινή και την χειμερινή περίοδο για την χρονοσειρά με δεκάλεπτο βήμα για την περιοχή του Τροόδους, όπως επίσης και τα διαγράμματα κανονικής πιθανότητας.

Όπως είναι αντιληπτό τόσο από τα ιστογράμματα της καλοκαιρινής και χειμερινής περιόδου όσο και από το διάγραμμα κανονικής πιθανότητας στο Σχήμα 5.3, η κατανομή των δεδομένων πλησιάζει περισσότερο την κανονική κατανομή, ειδικότερα κατά την καλοκαιρινή περίοδο, σε σύγκριση με την κατανομή του συνόλου των δεδομένων της ετήσιας χρονοσειράς ταχύτητας ανέμου με το ίδιο δεκάλεπτο βήμα. Αυτό το συμπεραίνουμε λόγω των μικρότερων διακυμάνσεων των τιμών και της καλύτερης εφαρμογής των δεδομένων στην καμπύλη της κανονικής κατανομής στα ιστογράμματα αλλά και από την μελέτη του διαγράμματος κανονικής πιθανότητας. Η ανάλυση αυτή μας οδηγεί στο συμπέρασμα της ομαλοποίησης των διακυμάνσεων των δεδομένων σε σχέση με τον χρόνο και της ύπαρξης κανονικής πλέον κατανομής. Όπως έχει αναφερθεί η παρουσία κανονικής κατανομής είναι απαραίτητο στοιχείο για την ορθή πρόβλεψη μελλοντικών χρονικών στιγμών.

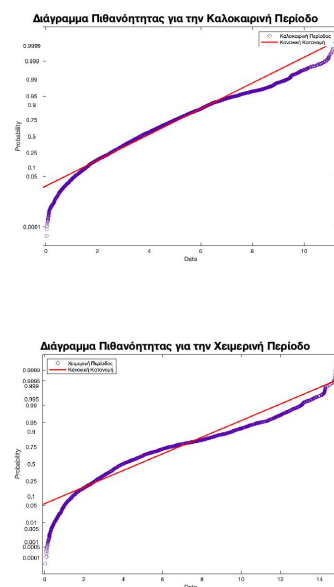
Για το τελευταίο κομμάτι της στατιστικής ανάλυσης, στους Πίνακες 5.2 και 5.3 παρακάτω παρουσιάζονται τα βασικά στατιστικά μέτρα για τα δεδομένα κάθε περιόδου ξεχωριστά.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	3.86
Τυπική απόκλιση (m/s)	1.89
Διάμεση τιμή (m/s)	3.66
Κύρτωση	3.42
Λοξότητα	0.65
Μέγιστη τιμή (m/s)	11.16
Ελάχιστη τιμή (m/s)	0.03

Πίνακας 5.2: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Τροόδους



(α') Ιστογράμματα για την καλοκαιρινή και χειμερινή περίοδο
χρονοσειράς με δεκάλεπτο βήμα για την περιοχή του Τροόδους.



(β') Διαγράμματα κανονικής κατανομής πιθανότητας για την καλοκαι-
ρινή και χειμερινή περίοδο με δεκάλεπτο βήμα για την περιοχή του
Τροόδους.

Σχήμα 5.3: Ανάλυση δεδομένων (ιστογράμματα, διαγράμματα κανονικής πι-
θανότητας) ετήσιας χρονοσειράς με δεκάλεπτο βήμα για την καλοκαιρινή και
χειμερινή περίοδο για την περιοχή του Τροόδους.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.78
Τυπική απόκλιση (m/s)	3.04
Διάμεση τιμή (m/s)	3.97
Κύρτωση	3.17
Λοξότητα	0.91
Μέγιστη τιμή (m/s)	14.82
Ελάχιστη τιμή (m/s)	0.03

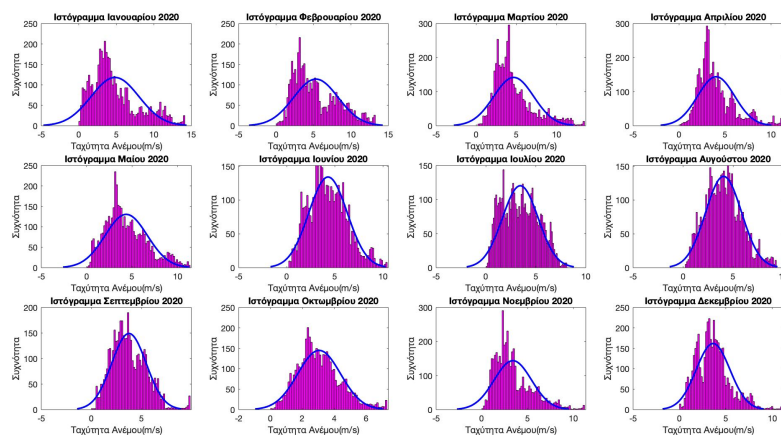
Πίνακας 5.3: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την χειμερινή περίοδο για την περιοχή του Τροόδους

Συμπεραίνουμε από τα αποτελέσματα των τιμών των δύο παραπάνω πινάκων ότι οι τιμές της κύρτωσης και στις δυο περιόδους είναι πολύ πιο κοντά στην τιμή κύρτωσης της κανονικής κατανομής. Επιπλέον, και για τις τιμές της λοξότητας που τείνουν στο 0 και στις δύο περιόδους αυτό δηλώνει πως το σύνολο των δεδομένων πλησιάζει περισσότερο την κανονική κατανομή.

5.2.2 Περιοχή του Αγρού

Χρονοσειρές με δεκάλεπτο βήμα

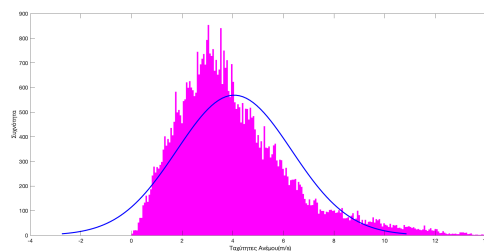
Στην υποενότητα αυτή, η ανάλυση που πραγματοποιήθηκε για τα δεδομένα της περιοχής του Τροόδους επαναλαμβάνεται στην περίπτωση της περιοχής του Αγρού. Και εδώ η χρονοσειρά των δεδομένων περιλαμβάνει δείγματα που έχουν ληφθεί με χρονικό βήμα 10 λεπτών για το χρονικό διάστημα του έτους 2020. Τα ιστογράμματα των μηνιαίων χρονοσειρών παρουσιάζονται στο Σχήμα 5.4.



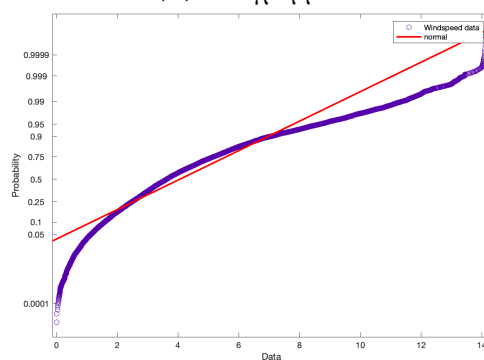
Σχήμα 5.4: Ανάλυση δεδομένων (ιστογράμματα) μηνιαίων χρονοσειρών με δεκάλεπτο βήμα για την περιοχή του Αγρού για το έτος 2020.

Όπως παρατηρούμε από τα ιστογράμματα ταχύτητας ανέμου ξεχωριστά ανά μήνα ενός έτους, τα δεδομένα των περισσότερων μηνιαίων χρονοσειρών αποκλίνουν από την κανονική κατανομή.

Με την βοήθεια του ιστογράμματος και του διαγράμματος κανονικής πιθανότητας που φαίνεται στο Σχήμα 5.5 για την χρονοσειρά με χρονικό βήμα ανά 10 λεπτά για το έτος 2020 μπορούμε να καταλήξουμε σε χρήσιμα συμπεράσματα σχετικά με την κατανομή των δεδομένων. Το διάγραμμα της κανονικής πιθανότητας που φαίνεται πιο κάτω δείχνει την κατανομή των δεδομένων σε σύγκριση με την κανονική κατανομή για την χρονοσειρά με χρονικό βήμα 10 λεπτών της ώρα για την περιοχή του Αγρού για το έτος 2020.



(α') Ιστόγραμμα



(β') Διάγραμμα κανονικής κατανομής πιθανότητας.

Σχήμα 5.5: Ανάλυση κανονικής κατανομής (Ιστόγραμμα, Διάγραμμα Κανονικής Πιθανότητας) ετήσιας χρονοσειράς ανά 10 λεπτά για την περιοχή του Αγρού.

Μέσω των διαγραμμάτων που φαίνονται παραπάνω στο σχήμα γίνεται κατανοητό πως οι διακιμ που παρουσιάζει το ιστόγραμμα της ετήσιας χρονοσειράς για την περιοχή του Αγρού με βήμα ανά 10 λεπτά μας οδηγούν στο συμπέρασμα της απόκλισης της χρονοσειράς από την κανονική κατανομή.

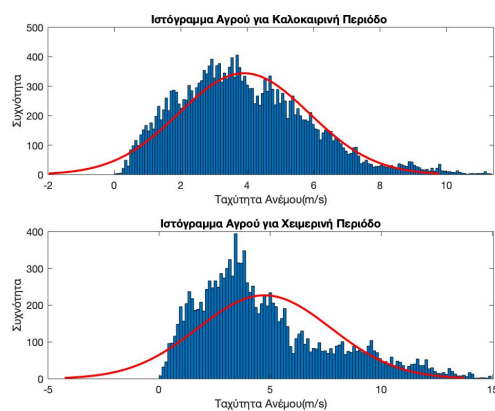
Για το τελευταίο κομμάτι της στατιστικής ανάλυσης της χρονοσειράς με χρονικό βήμα 10 λεπτών για την περιοχή του Αγρού δίνονται κάποια βασικά χαρακτηριστικά στατιστικά μέτρα στον Πίνακα 5.4. Παρατηρούμε όπως και στον Πίνακα 5.1 οι τιμές των στατιστικών μέτρων είναι πολύ κοντά. Επομένως, η τιμή της κύρτωσης από την παρατήρηση του πίνακα αυτού δηλώνει την απόκλιση του συνόλου των δεδομένων από την κανονική κατανομή.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.06
Τυπική απόκλιση (m/s)	2.27
Διάμεση τιμή (m/s)	3.63
Κύρτωση	4.56
Λοξότητα	1.13
Μέγιστη τιμή (m/s)	14.12
Ελάχιστη τιμή (m/s)	0.01

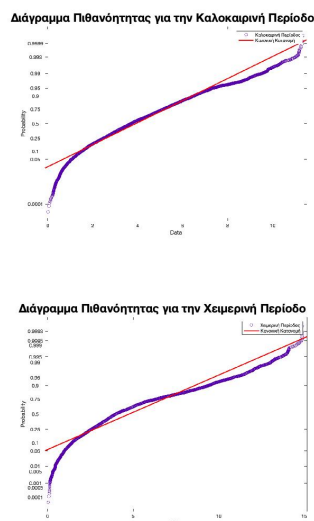
Πίνακας 5.4: Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά με δεκάλεπτο βήμα για την περιοχή του Αγρού.

Καλοκαιρινή και χειμερινή περίοδος

Λόγω της απόκλισης της χρονοσειράς από την κανονική κατανομή αλλά και για να μειωθεί το μεγάλο μέγεθος της χρονοσειράς έχει γίνει διάσπαση της ετήσιας χρονοσειράς με βήμα ανά 10 λεπτά σε καλοκαιρινή και χειμερινή περίοδο. Με αποτέλεσμα τα δεδομένα από 52704 μειώθηκαν σε 22176 για την καλοκαιρινή περίοδο και σε 13104 δεδομένα για την χειμερινή περίοδο. Στο πιο κάτω Σχήμα 5.6 φαίνονται τα ιστογράμματα της καλοκαιρινής και της χειμερινής περιόδου για τις χρονοσειρές με βήμα ανά 10 λεπτά για την περιοχή του Αγρού, όπως και το διάγραμμα κανονικής πιθανότητας.



(α') Ιστογράμματα



(β') Διαγράμματα κανονικής κατανομής πιθανότητας.

Σχήμα 5.6: Ανάλυση κανονικής κατανομής (Ιστογράμματα, Διαγράμματα κανονικής πιθανότητας) χρονοσειράς ανά 10 λεπτά για την καλοκαιρινή και χειμερινή περίοδο για περιοχή του Αγρού.

Με βάση το παραπάνω σχήμα γίνεται αντιληπτό ότι τόσο στα ιστογράμματα της καλοκαιρινής και της χειμερινής περιόδου όσο και στα διαγράμματα κανονικής πιθανότητας ότι τα δεδομένα εφαρμόζουν πολύ καλύτερα στην καμπύλη της κανονικής κατανομής σε σύγκριση με το σύνολο των δεδομένων της ετήσιας χρονοσειράς με βήμα ανά 10 λε-

πτά. Αυτό το συμπεραίνουμε λόγω των μικρότερων διακυμάνσεων των τιμών και της καλύτερης εφαρμογής των δεδομένων στην καμπύλη της κανονικής κατανομής στα ιστογράμματα αλλά και από την μελέτη του διαγράμματος κανονικής πιθανότητας.

Τελειώνοντας την στατιστική ανάλυση για την καλοκαιρινή και την χειμερινή περίοδο με βήμα 10 λεπτά για την περιοχή του Αγρού, οι τιμές βασικών στατιστικών μέτρων παρουσιάζονται στους πιο κάτω πίνακες.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	3.90
Τυπική απόκλιση (m/s)	1.96
Διάμεση τιμή (m/s)	3.68
Κύρτωση	3.29
Λοξότητα	0.63
Μέγιστη τιμή (m/s)	11.34
Ελάχιστη τιμή (m/s)	0.01

Πίνακας 5.5: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Αγρού.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.74
Τυπική απόκλιση (m/s)	2.99
Διάμεση τιμή (m/s)	3.97
Κύρτωση	3.23
Λοξότητα	0.92
Μέγιστη τιμή (m/s)	14.87
Ελάχιστη τιμή (m/s)	0.04

Πίνακας 5.6: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με δεκάλεπτο βήμα για την χειμερινή περίοδο για την περιοχή του Αγρού.

Παρατηρώντας τις τιμές των πιο πάνω πινάκων συμπεραίνουμε ότι οι τιμές της κύρτωσης και στις δυο περιόδους είναι πολύ κοντά στις τιμές της κανονικής κατανομής, όπως επίσης και οι τιμές της λοξότητας που τείνουν στο 0 και στις δύο περιπτώσεις που αυτό υποδηλώνει πως το σύνολο των δεδομένων πλησιάζει περισσότερο την κανονική κατανομή. Επίσης, ρίχνοντας μια ματιά στους Πίνακες 5.2 και 5.3 που αφορούν την περιοχή του Τροόδους για τις ίδιες χρονικές περιόδους με βήμα 10 λεπτών παρατηρούμε ότι οι τιμές είναι σχεδόν ίδιες με τις τιμές των Πινάκων 5.5 και 5.6. Αυτό μας κάνει να αναλογιστούμε ότι παρά την διαφορά υψομέτρου ανάμεσα στις δύο περιοχές, οι τιμές των χαρακτηριστικών στατιστικών μέτρων δεν απέχουν πολύ.

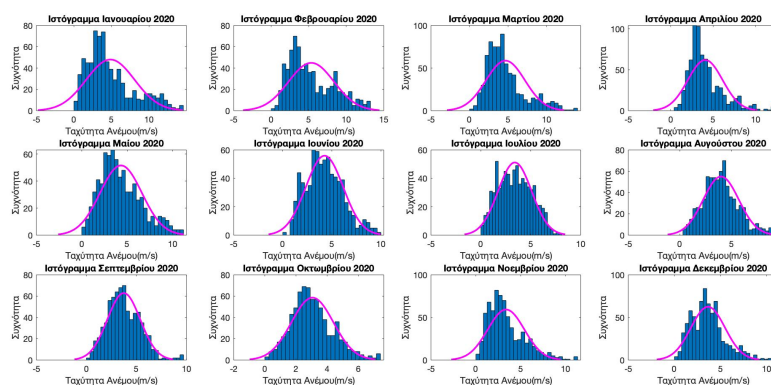
5.2.3 Χρονοσειρές με ωριαίο βήμα

Κάνοντας την στατιστική ανάλυση των χρονοσειρών με βήμα ανά ώρα για τις περιοχές Τροόδους και Αγρού μειώνουμε το μεγάλο μέγεθος της χρονοσειράς που έχει γίνει προηγουμένως με βήμα ανά 10 λεπτά για τις ίδιες περιοχές. Με αποτέλεσμα τα δεδομένα από 52704 μειώθηκαν σε 8784 για το έτος 2020. Αυτή είναι μια τεχνική εξομάλυνσης που βοηθά στην μείωση της μεταβλητότητας των βραχυχρόνιων (υψη-

λής συχνότητας) κυμάνσεων σε μια χρονοσειρά. Στην πραγματικότητα αυτό που επιτυγχάνεται με μια τέτοια εξομάλυνση είναι να απαλειφθεί η άρρυθμη συνιστώσα από τη χρονοσειρά. Κάτι τέτοιο είναι χρήσιμο, καθώς με τον τρόπο αυτό καθίσταται ευκολότερο να διακρίνουμε εποχικές κυμάνσεις και το πόσο κοντά μπορούν τα δεδομένα να φτάσουν την κανονική κατανομή.

Περιοχή του Τροόδους

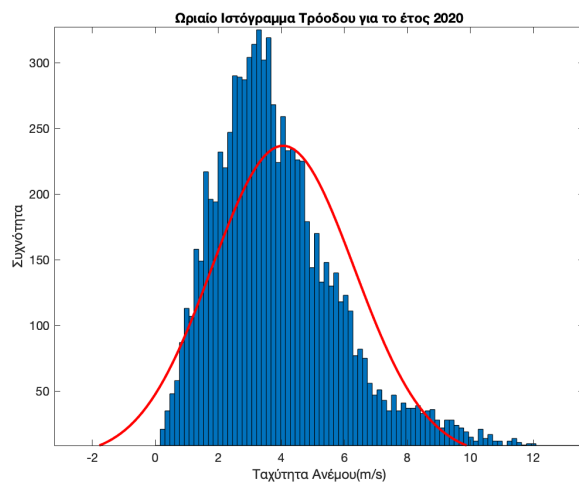
Με την βοήθεια του ιστογράμματος και της καμπύλης της κανονικής κατανομής (normal distribution) για τις χρονοσειρές του κάθε μήνα με χρονικό βήμα ανά ώρα για την περιοχή του Τροόδους μπορούμε να καταλήξουμε σε χρήσιμα συμπεράσματα σχετικά με την κατανομή των στοιχείων. Τα ιστογράμματα παρουσιάζονται στο Σχήμα 5.7.



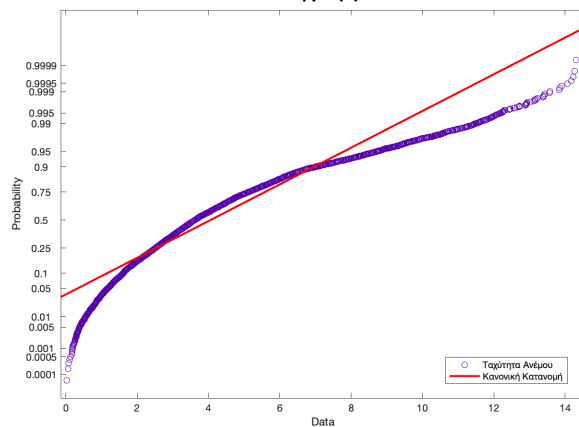
Σχήμα 5.7: Ανάλυση Κανονικής Κατανομής (Ιστογράμματα) για κάθε μήνα με βήμα ανά ώρα για την περιοχή του Τροόδους για το έτος 2020.

Όπως παρατηρούμε από τα διαγράμματα των διαφορετικών ιστογραμμάτων των μηνιαίων χρονοσειρών με βήμα ανά ώρα καταλήγουμε στο συμπέρασμα της απόκλισης ορισμένων χρονοσειρών από την κανονική κατανομή. Συγκεκριμένα τους μήνες του Ιανουαρίου, του Φεβρουαρίου, του Μαρτίου, του Απριλίου, του Νοεμβρίου και τέλος του Δεκεμβρίου η απόκλιση είναι μεγάλη.

Φτιάχνοντας το ιστόγραμμα και το διάγραμμα κανονικής πιθανότητας που φαίνεται στο Σχήμα 5.8 για την χρονοσειρά με χρονικό βήμα ανά ώρα για το έτος 2020 μπορούμε να καταλήξουμε σε χρήσιμα συμπεράσματα σχετικά με την κατανομή των δεδομένων. Μέσω των διαγραμμάτων που φαίνονται παρακάτω στο Σχήμα 5.8 γίνεται κατανοητό πως οι διακυμάνσεις που παρουσιάζει το ιστόγραμμα της ετήσιας χρονοσειράς για την περιοχή του Τροόδους με βήμα ανά ώρα μας οδηγούν στο συμπέρασμα της απόκλισης της χρονοσειράς από την κανονική κατανομή.



(α') Ιστόγραμμα



(β') Διάγραμμα κανονικής κατανομής πιθανότητας.

Σχήμα 5.8: Ανάλυση κανονικής κατανομής (Ιστόγραμμα, διάγραμμα κανονικής πιθανότητας) ετήσιας χρονοσειράς ανά ώρα για την περιοχή του Τροόδους.

Τελειώνοντας το κομμάτι της στατιστικής ανάλυσης της χρονοσειράς με χρονικό βήμα ανά ώρα για την περιοχή του Τροόδους δίνονται κάποια βασικά χαρακτηριστικά στατιστικά μέτρα στον Πίνακα 5.7.

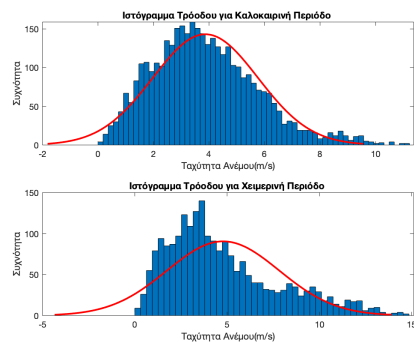
Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.05
Τυπική απόκλιση (m/s)	2.26
Διάμεση τιμή (m/s)	3.61
Κύρτωση	4.78
Λοξότητα	1.19
Μέγιστη τιμή (m/s)	14.31
Ελάχιστη τιμή (m/s)	0.02

Πίνακας 5.7: Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά με ωριαίο βήμα για την περιοχή του Τροόδους.

Παρατηρώντας τα αποτελέσματα του Πίνακα 5.7 συμπεραίνουμε ότι η τιμή της κύρτωσης και η τιμή της λοξότητας αποκλίνουν αρκετά από την κανονική κατανομή που υπό κανονικές συνθήκες αν και εφόσον η χρονοσειρά μας με βήμα ανά ώρα δεν απέκλινε από την κανονική κατανομή οι δύο αυτές τιμές θα έπρεπε να ήταν κοντά στο 3 και στο 0 αντίστοιχα.

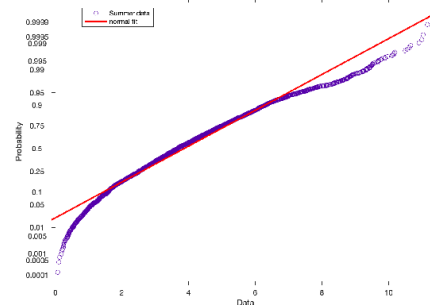
Καλοκαιρινή και χειμερινή περίοδος

Για την αντιμετώπιση της κατανομής που αποκλίνει από την κανονική αλλά και για να μειώσουμε το μεγάλο μέγεθος της χρονοσειράς έγινε όπως και πριν διαχωρισμός σε καλοκαιρινή και χειμερινή περίοδο. Έτσι τα 8784 δεδομένα μειώνονται σε 3696 για την καλοκαιρινή περίοδο και σε 2184 δεδομένα για την χειμερινή περίοδο. Πιο κάτω στο Σχήμα 5.9 φαίνονται τόσο τα ιστογράμματα της καλοκαιρινής περιόδου και της χειμερινής περιόδου με βήμα ανά ώρα για την περιοχή του Τροόδους όσο και τα διαγράμματα κανονικής πιθανότητας.

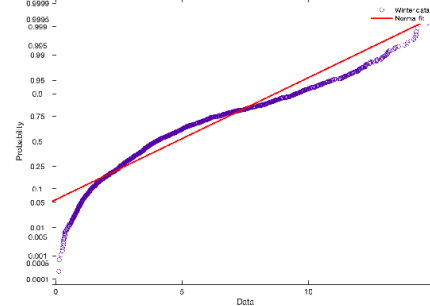


(α') Ιστογράμματα

Διάγραμμα Πιθανότητας για την Καλοκαιρινή Περίοδο



Διάγραμμα Πιθανότητας για την Χειμερινή Περίοδο



(β') Διαγράμματα κανονικής κατανομής πιθανότητας.

Σχήμα 5.9: Ανάλυση κανονικής κατανομής (Ιστογράμματα, Διαγράμματα Κανονικής Πιθανότητας) χρονοσειράς ανά ώρα για την καλοκαιρινή και χειμερινή περίοδο για περιοχή του Τροόδους.

Όπως γίνεται αντιληπτό τόσο στα ιστογράμματα όσο και στα διαγράμματα κανονικής πιθανότητας για τις ωριαίες χρονοσειρές ταχύτη-

τας ανέμου που αφορούν την καλοκαιρινή και την χειμερινή περίοδο για την περιοχή του Τροόδους τα δεδομένα πλησιάζουν περισσότερο την κανονική κατανομή ειδικά στην περίπτωση της καλοκαιρινή περιόδου. Αυτό προκύπτει λόγω των μικρότερων διακυμάνσεων των τιμών στο ι-στόγραμμα αλλά και στο διάγραμμα κανονικής πιθανότητας όπου τα δεδομένα εφαρμόζουν καλύτερα την γραμμή της κανονική κατανομής. Η ανάλυση αυτού μας οδηγεί στο συμπέρασμα της ομαλοποίησης των διακυμάνσεων των δεδομένων σε σχέση με τον χρόνο και της ύπαρξης πλέον κανονικής κατανομής.

Τέλος, στους Πίνακες 5.8 και 5.9 που ακολουθούν παρουσιάζονται βασικά στατιστικά μέτρα για τις ωριαίες χρονοσειρές που αφορούν την καλοκαιρινή και χειμερινή περίοδο για την περιοχή του Τροόδους.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	3.87
Τυπική απόκλιση (m/s)	1.89
Διάμεση τιμή (m/s)	3.66
Κύρτωση	3.42
Λοξότητα	0.65
Μέγιστη τιμή (m/s)	11.17
Ελάχιστη τιμή (m/s)	0.07

Πίνακας 5.8: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Τροόδους.

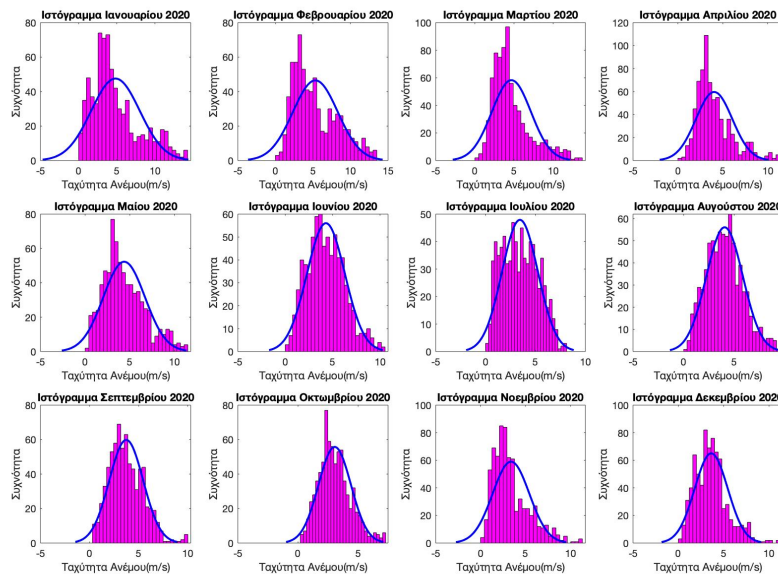
Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.78
Τυπική απόκλιση (m/s)	3.04
Διάμεση τιμή (m/s)	3.98
Κύρτωση	3.18
Λοξότητα	0.91
Μέγιστη τιμή (m/s)	14.83
Ελάχιστη τιμή (m/s)	0.13

Πίνακας 5.9: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την χειμερινή περίοδο για την περιοχή του Τροόδους.

Παρατηρώντας τις τιμές των πιο πάνω πινάκων συμπεραίνουμε ότι οι τιμές της κύρτωσης και στις δυο περιόδους είναι πολύ κοντά στις τιμές της κανονικής κατανομής, όπως επίσης και οι τιμές της λοξότητας που τείνουν στο 0 και στις δύο περιπτώσεις που αυτό υποδηλώνει πως το σύνολο των δεδομένων πλησιάζει περισσότερο την κανονική κατανομή.

Περιοχή Αγρός

Όπως και για την περιοχή του Τροόδους έγινε στατιστική ανάλυση για τα δεδομένα που πάρθηκαν αντίστοιχη στατιστική ανάλυση γίνεται και για την περιοχή του Αγρού με βήμα ανά ώρα για το έτος 2020. Πιο κάτω στο Σχήμα 5.10 φαίνονται τα ιστογράμματα κάθε μηνός ξεχωριστά για το έτος 2020.

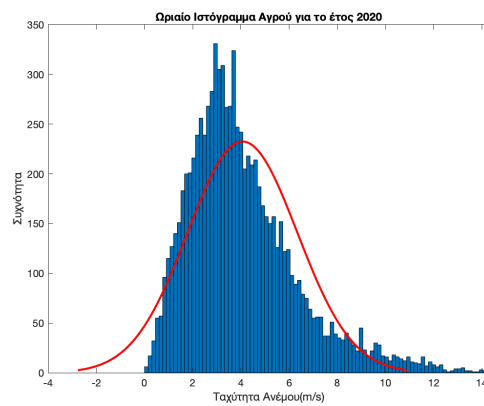


Σχήμα 5.10: Ανάλυση Κανονικής Κατανομής (Ιστογράμματα) για κάθε μήνα με βήμα ανά ώρα για την περιοχή του Αγρού για το έτος 2020.

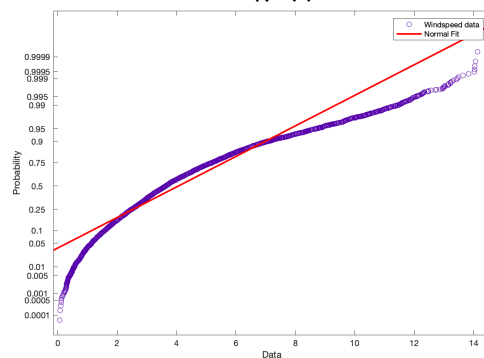
Μελετώντας τα ιστογράμματα του κάθε μηνός ξεχωριστά παρατηρείται ότι στους μήνες Ιούνιο, Ιούλιο, Αύγουστο, Σεπτέμβριο και Οκτώβριο η καμπύλη της κανονικής κατανομής εφαρμόζει καλύτερα στα δεδομένα. Αντίθετα στους υπόλοιπους μήνες παρατηρούμε διακυμάνσεις στα δεδομένα με αποτέλεσμα να απομακρύνονται από την κανονική κατανομή.

Από το ιστόγραμμα και το διάγραμμα κανονικής πιθανότητας που φαίνεται στο Σχήμα 5.11 για την χρονοσειρά ταχύτητας ανέμου με χρονικό βήμα ανά ώρα για το έτος 2020 μπορούμε να καταλήξουμε σε χρήσιμα συμπεράσματα σχετικά με την κατανομή των δεδομένων. Μέσω των διαγραμμάτων που φαίνονται παρακάτω στο Σχήμα 5.11 γίνεται κατανοητό πως οι διακυμάνσεις που παρουσιάζει το ιστόγραμμα της ετήσιας χρονοσειράς ταχύτητας ανέμου για την περιοχή του Αγρού με βήμα ανά ώρα μας οδηγούν στο συμπέρασμα της απόκλισης της χρονοσειράς από την κανονική κατανομή.

Για το τελευταίο κομμάτι της στατιστικής ανάλυσης της χρονοσειράς ταχύτητας ανέμου με χρονικό βήμα ανά ώρα για την περιοχή του



(α') Ιστόγραμμα



(β') Διάγραμμα κανονικής κατανομής πιθανότητας.

Σχήμα 5.11: Ανάλυση κανονικής κατανομής (Ιστόγραμμα, διάγραμμα κανονικής πιθανότητας) ετήσιας χρονοσειράς ταχύτητας ανέμου με χρονικό βήμα ανά ώρα για την περιοχή του Αγρού.

Αγρού δίνονται κάποια βασικά χαρακτηριστικά στατιστικά μέτρα στον Πίνακα 5.10.

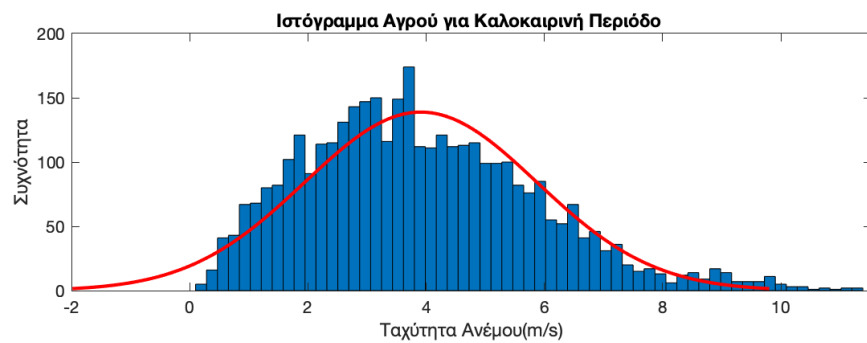
Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.06
Τυπική απόκλιση ((m/s)	2.28
Διάμεση τιμή (m/s)	3.64
Κύρτωση	4.56
Λοξότητα	1.13
Μέγιστη τιμή (m/s)	14.13
Ελάχιστη τιμή (m/s)	0.06

Πίνακας 5.10: Πίνακας βασικών στατιστικών μέτρων για την ετήσια χρονοσειρά ταχύτητας ανέμου με ωριαίο βήμα για την περιοχή του Αγρού.

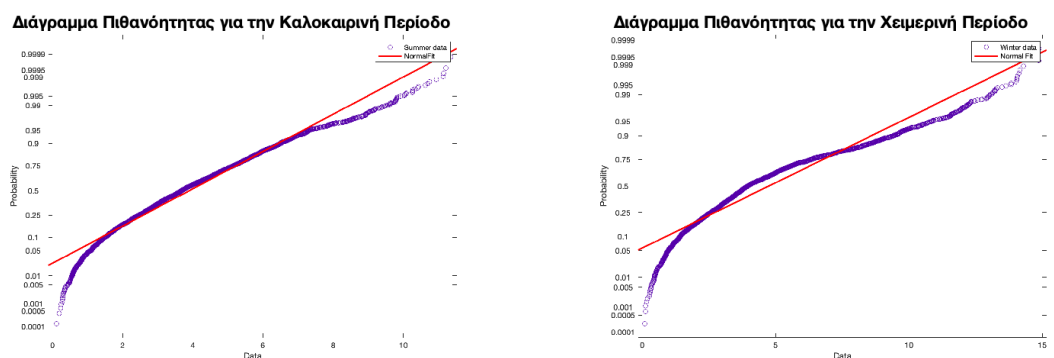
Παρατηρώντας τα αποτελέσματα του Πίνακα 5.10 συμπεραίνουμε ότι η τιμή της κύρτωσης και η τιμή της λοξότητας αποκλίνουν αρκετά από αυτές της κανονικής κατανομής.

Καλοκαιρινή και χειμερινή περίοδος

Όπως και για την περιοχή του Τροόδους αντίστοιχα και για την χρονοσειρά ταχύτητας ανέμου της περιοχής του Αγρού έγινε διάσπασή της σε καλοκαιρινή και χειμερινή περίοδο. Στο πιο κάτω Σχήμα 5.12 φαίνονται τα ιστογράμματα της καλοκαιρινής και της χειμερινής περιόδου για τις χρονοσειρές με βήμα ανά 10 λεπτά για την περιοχή του Αγρού, όπως και το διάγραμμα κανονικής πιθανότητας.



(α) Ιστογράμματα



(β) Διαγράμματα κανονικής κατανομής πιθανότητας

Σχήμα 5.12: Ανάλυση κανονικής κατανομής (Ιστογράμματα, διαγράμματα κανονικής πιθανότητας) χρονοσειράς ταχύτητας ανέμου με χρονικό βήμα ανά ώρα για την καλοκαιρινή και χειμερινή περίοδο για την περιοχή του Αγρού.

Με βάση το παραπάνω Σχήμα 5.12 γίνεται αντιληπτό ότι τόσο στα ιστογράμματα της καλοκαιρινής και της χειμερινής περιόδου όσο και στα διαγράμματα κανονικής πιθανότητας ότι τα δεδομένα εφαρμόζουν πολύ καλύτερα στην καμπύλη της κανονικής κατανομής σε σύγκριση με το σύνολο των δεδομένων της ετήσιας χρονοσειράς ταχύτητας ανέμου με βήμα ανά ώρα.

Τελειώνοντας την στατιστική ανάλυση για την καλοκαιρινή και την χειμερινή περίοδο με βήμα ανά ώρα για την περιοχή του Αγρού παρουσιάζονται στους πιο κάτω πίνακες βασικά στατιστικά μέτρα.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	3.90
Τυπική απόκλιση (m/s)	1.96
Διάμεση τιμή (m/s)	3.68
Κύρτωση	3.28
Λοξότητα	0.63
Μέγιστη τιμή (m/s)	11.35
Ελάχιστη τιμή (m/s)	0.12

Πίνακας 5.11: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την καλοκαιρινή περίοδο για την περιοχή του Αγρού.

Βασικά Στατιστικά Μέτρα	Τιμές
Μέση τιμή (m/s)	4.75
Τυπική απόκλιση (m/s)	3.00
Διάμεση τιμή (m/s)	3.97
Κύρτωση	3.24
Λοξότητα	0.92
Μέγιστη τιμή (m/s)	14.87
Ελάχιστη τιμή (m/s)	0.01

Πίνακας 5.12: Πίνακας βασικών στατιστικών μέτρων της χρονοσειράς με ωριαίο βήμα για την χειμερινή περίοδο για την περιοχή του Αγρού.

Παρατηρώντας τις τιμές των πιο πάνω πινάκων συμπεραίνουμε ότι οι τιμές της κύρτωσης και στις δυο περιόδους είναι πολύ κοντά στις τιμές της κανονικής κατανομής, όπως επίσης και οι τιμές της λοξότητας που τείνουν στο 0 και στις δύο περιπτώσεις που αυτό υποδηλώνει πως το σύνολο των δεδομένων πλησιάζει περισσότερο την κανονική κατανομή. Επίσης αυτό που παρατηρούμε ότι παρά την διαφορά υψομέτρου δεν

υπάρχουν μεγάλες διαφορές στα στατιστικά μέτρα. Επιπλέον κάνοντας και την στατιστική ανάλυση με χρονικό βήμα 10 λεπτών όπως επίσης και με χρονικό βήμα μιας ώρας για τις ίδιες περιοχές, τα αποτελέσματα που παίρνουμε δεν έχουν μεγάλη διαφορά μεταξύ τους. Μεγάλη διαφορά παρατηρείται όταν γίνεται διάσπαση των χρονοσειρών ταχυτήτων ανέμου για την καλοκαιρινή και χειμερινή περίοδο αντίστοιχα και για τις δύο χρονικές περιόδους.

5.3 Προσδιορισμός βέλτιστου πιθανοτικού μοντέλου

Μελετώντας τα αποτελέσματα της στατιστικής ανάλυσης που έγινε πιο πάνω παρατηρούμε ότι, στις περισσότερες περιπτώσεις ανάλυσης των χρονοσειρών ταχυτήτων ανέμου, η κατανομή των δεδομένων αποκλίνει από την κανονική κατανομή. Για αυτό το λόγο χρησιμοποιώντας τις μεθόδους προσαρμογής κατανομής που προσφέρει η εργαλειοθήκη Statistics and Machine Learning Toolbox του MATLAB έγινε σύγκριση στις διαφορετικές κατανομές για ένα σύνολο δεδομένων. Αυτό υλοποιήθηκε με τη συνάρτηση `fitmethis` [6] του MATLAB. Περιοριστήκαμε σε παραμετρικές κατανομές που έχουν το πολύ δύο παραμέτρους, δηλαδή χρησιμοποιήθηκε η `fitmethis` για να γίνει σύγκριση των τιμών λογαριθμο-πιθανοφάνειας (log-likelihood) των ακόλουθων κατανομών: κανονική, εκθετική, γάμμα (gamma), λογιστική, ομοιόμορφη, ακραίας τιμής, Rayleigh, beta, Nakagami, Rician, inverse Gaussian, Birnbaum-Saunders, log-logistic, log-normal και Weibull. Εκτός των προαναφερθέντων κατανομών, ελέγχθηκε και η κατανομή γενικευμένης ακραίας τιμής (Generalized extreme value, Gev) η οποία χαρακτηρίζεται από τρεις παραμέτρους.

5.3.1 Αποτελέσματα

Στους Πίνακες 5.13 και 5.14 δίνονται τα αποτελέσματα της συνάρτησης `fitmethis` για τις ετήσιες χρονοσειρές με χρονικό βήμα ανά μία ώρα στις περιοχές του Αγρού και του Τροόδου αντίστοιχα. Στην πρώτη στήλη δίνεται η κατανομή πιθανότητας, στις στήλες 2 έως 4 δίνονται οι τιμές των παραμέτρων της κατανομής, στην στήλη 5 η τιμή της λογαριθμο-πιθανοφάνειας (LogL) και στην τελευταία στήλη η τιμή του κριτηρίου Akaike.

Περιοχή Αγρός, χρονικό βήμα 10 λεπτών					
Κατανομή	Παρ. 1	Παρ. 2	Παρ. 3	LogL	AIC
Gamma	3.17	1.28	-	-1.87×10^4	3.74×10^4
GEV	0.43	1.69	3.00	-1.87×10^4	3.74×10^4
Weibull	4.59	1.88	-	-1.88×10^4	3.77×10^5

Πίνακας 5.13: Πίνακας με τις τρεις επικρατέστερες κατανομές για την ετήσια χρονοσειρά με βήμα ανά μία ώρα για την περιοχή του Αγρού.

Περιοχή Τροόδους, χρονικό βήμα 10 λεπτών					
Κατανομή	Παρ. 1	Παρ. 2	Παρ. 3	LogL	AIC
GEV	0.45	1.67	3.01	-1.82×10^4	3.72×10^4
Gamma	3.22	1.26	-	-1.87×10^4	3.73×10^4
LogLogistic	1.27	0.33	-	-1.88×10^4	3.75×10^4

Πίνακας 5.14: Πίνακας με τις τρεις επικρατέστερες κατανομές για την ετήσια χρονοσειρά βήμα ανά μία ώρα για την περιοχή του Τροόδους.

Έγινε αντίστοιχη προσαρμογή της συνάρτησης fitmethis για τις ετήσιες χρονοσειρές με χρονικό βήμα ανά 10 λεπτά για τις περιοχές του Αγρού και του Τροόδους και τα αποτελέσματα ήταν ίδια. Από τους παραπάνω πίνακες παρατηρούμε ότι και στις δύο περιοχές αντίστοιχα οι επικρατέστερες κατανομές είναι σχεδόν ίδιες με διαφορά την τρίτη σε σειρά κατανομή που διαφέρει.

5.4 Πρόβλεψη με την βοήθεια του μοντέλου ARIMA

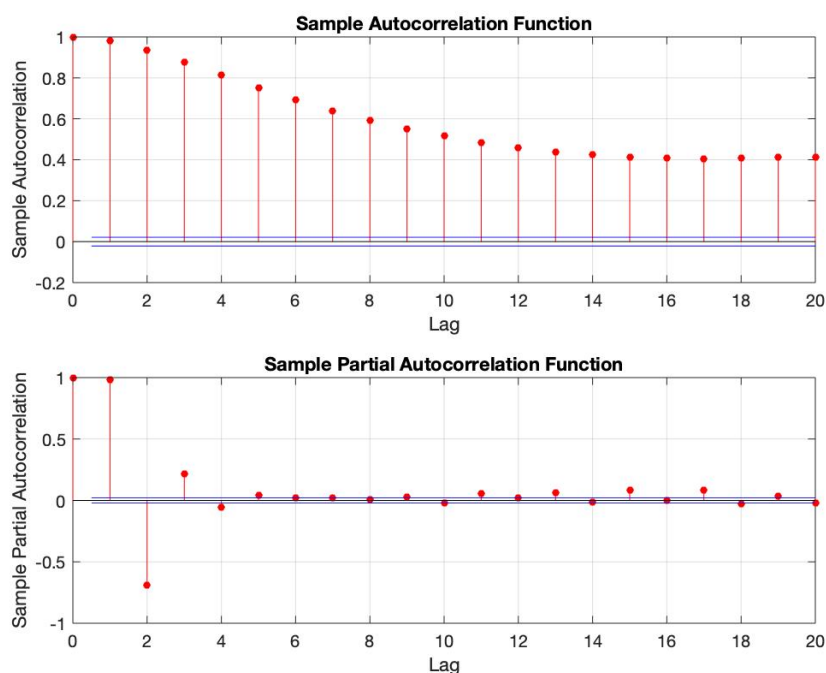
Δεδομένα για την πρόβλεψη χρονοσειρών ταχύτητας ανέμου λήφθηκαν από τις περιοχές Αγρός και Τρόοδος στην Κύπρο, όπως έχει προαναφερθεί. Η διαδικασία για την πρόβλεψη των χρονοσειρών αποτελείται από δύο μέρη: στο πρώτο μέρος γίνεται η εκτίμηση των παραμέτρων του μοντέλου ARIMA και στο δεύτερο μέρος γίνεται η πρόβλεψη με το κατάλληλο μοντέλο. Για την δημιουργία των καλύτερων παραμέτρων του μοντέλου για την πρόβλεψη της ταχύτητας ανέμου αναλύθηκαν τρία

σύνολα δεδομένων:

- 1η Ιανουαρίου του 2020 έως τις 31 Δεκεμβρίου του 2020
- 10 Μαΐου του 2020 έως 10 Οκτωβρίου του 2020
- 1 Δεκεμβρίου του 2019 έως 29 Φεβρουαρίου του 2020.

5.4.1 Χρονοσειρά με ωριαίο βήμα για το έτος 2020

Για τον προσδιορισμό των τάξεων, p και q , εξετάστηκαν τα διαγράμματα των συναρτήσεων αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF). Στα γραφήματα 1 και 2 του Σχήματος 5.13, παρουσιάζονται τα διαγράμματα ACF και PACF για τα ωριαία δεδομένα ταχύτητας ανέμου στην περιοχή του Τροόδους.



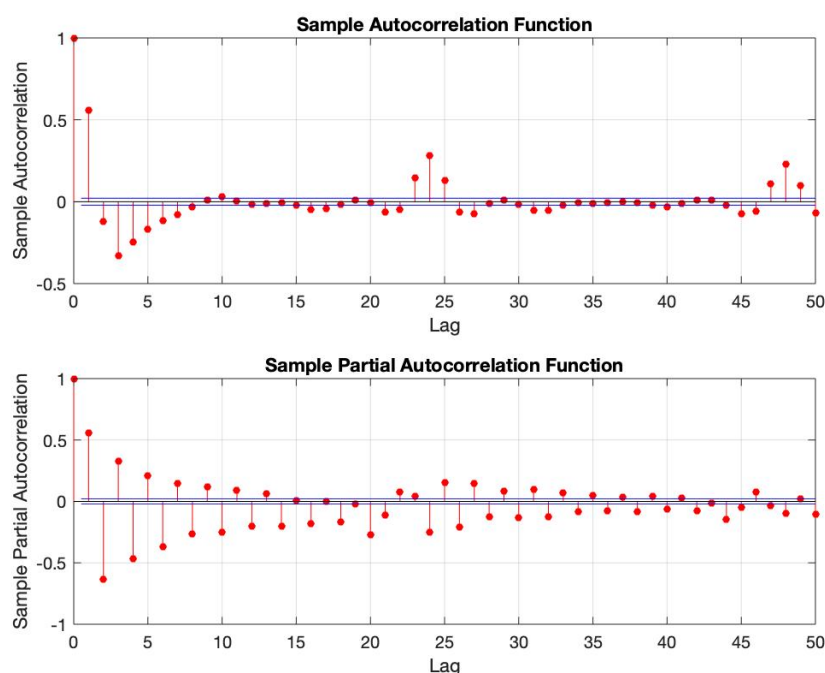
Σχήμα 5.13: Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για τα ωριαία δεδομένα ταχύτητας ανέμου της περιοχή του Τροόδους για το έτος 2020.

Παρατηρώντας το διάγραμμα της συνάρτησης αυτοσυσχέτισης (ACF) γίνεται αντιληπτό ότι τα σημεία ξεπερνούν τα όρια, αυτό σημαίνει ότι

τα δεδομένα της χρονοσειράς δεν είναι στάσιμα και παρουσιάζουν τάση και περιοδικότητα 1.2.3. Με την χρήση της MATLAB έγινε απαλοιφή της τάσης και της περιοδικότητας από τα δεδομένα της χρονοσειράς κάνοντας διαφόριση (differencing). Για την εξάλειψη της τάσης και της περιοδικότητας πήραμε “δεύτερες διαφορές” (second differences) με την βοήθεια του πολυωνύμου

$$y_t = \nabla x^{(2)}t = x^{(1)}(t) - x^{(1)}(t - 1),$$

όπου ο πάνω δείκτης υποδηλώνει την σειρά της διαφοράς. Στο Σχήμα 5.14 παρακάτω φαίνονται οι συναρτήσεις αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF), επομένως είναι φανερό ότι παρά την αποτελεσματική αντιμετώπιση της τάσης και της περιοδικότητας, η εμφάνιση των ακραίων τιμών υπόκειται στην εμφάνιση αυτοσυσχετίσεων μεταξύ των τιμών.



Σχήμα 5.14: Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για τα διαφορισμένα δεδομένα ταχύτητας ανέμου.

Όπως παρατηρούμε και από τα παραπάνω διαγράμματα οι συναρ-

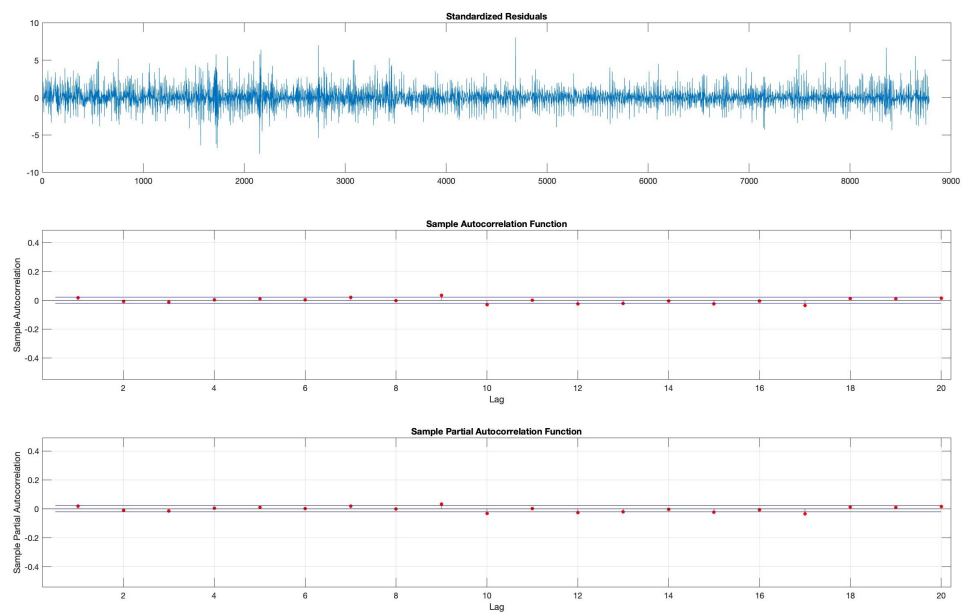
	MONTEAO	AIC Value	BIC Value
1	ARIMA(2,1,9)	4.8532e+034	4.9453e+03
2	ARIMA(3,1,8)	4.8554e+03	4.8554e+03
3	ARIMA(2,1,7)	4.8854e+03	4.9633e+03
4	ARIMA(2,1,11)	4.7837e+03	4.8899e+03
5	ARIMA(3,1,12)	4.5645e+03	4.6849e+03

Πίνακας 5.15: Τιμές κριτηρίων AIC και BIC για διαφορετικά μοντέλα ARIMA.

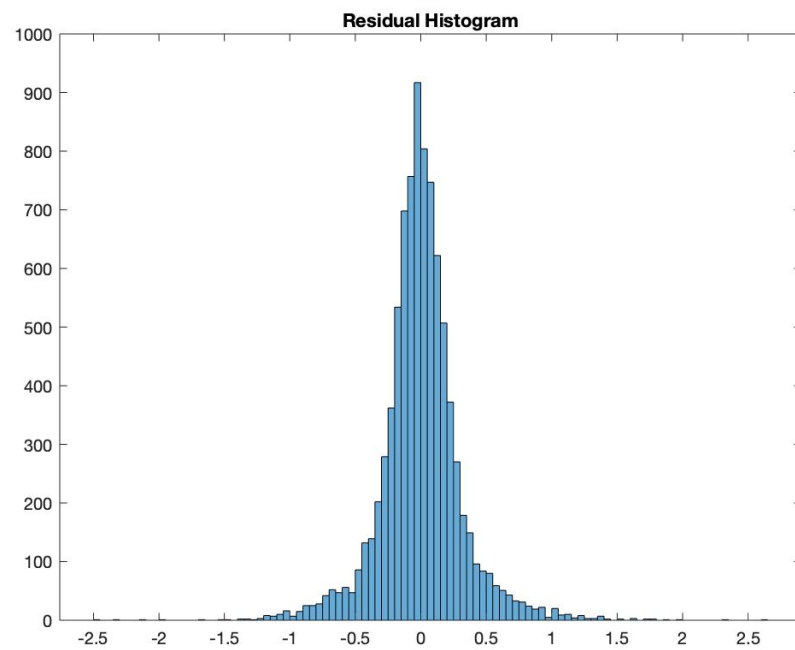
τήσεις μερικής αυτοσυσχέτισης και αυτοσυσχέτισης δεν έδειξαν τις κατάλληλες τιμές των παραμέτρων p και q . Χρησιμοποιώντας τα κριτήρια AIC και BIC για διάφορα μοντέλα γίνεται η επιλογή του κατάλληλου μοντέλου.

Παρατηρώντας τον Πίνακα 5.15, το κατάλληλο μοντέλο είναι το ARIMA(3,1,12) γιατί έχει την μικρότερη τιμή για τα κριτήρια AIC και BIC. Στην συνέχεια γίνεται ο διαγνωστικός έλεγχος για τα κατάλοιπα (residuals) του μοντέλου. Τα κατάλοιπα είναι χρήσιμα για τον έλεγχο του μοντέλου ώστε να διαφανεί κατά πόσο έχουν συλλάβει τις πληροφορίες στα δεδομένα. Επιπλέον, θα πρέπει να μην έχουν συσχετίσεις μεταξύ τους αλλά και να έχουν μηδενικό μέσο όρο. Εκτός από αυτό, θα πρέπει τα κατάλοιπα να ακολουθούν την κανονική κατανομή..

Στο παρακάτω Σχήμα 5.15 παρατηρείται ότι ο μέσος όρος των καταλοίπων (residuals) που φαίνεται στα διαγράμματα των συναρτήσεων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης είναι μέσα στα όρια (δηλαδή κοντά στο μηδέν). Η χρονοσειρά των καταλοίπων δείχνει ότι η διακύμανση παραμένει σταθερή. Αυτό επιβεβαιώνεται και από το ιστόγραμμα στο Σχήμα 5.16, όπου τα κατάλοιπα του μοντέλου ακολουθούν την κανονική κατανομή.



Σχήμα 5.15: Κατάλοιπα μοντέλου.



Σχήμα 5.16: Ιστόγραμμα καταλοίπων μοντέλου ARIMA(3,1,12).

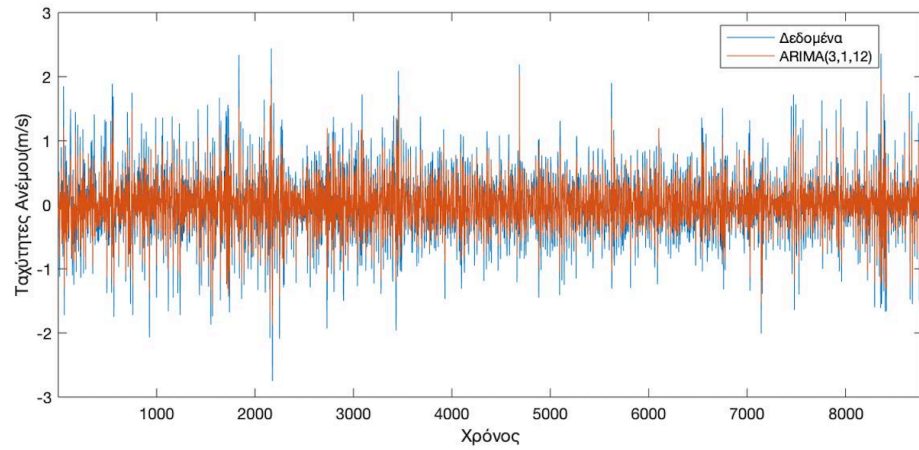
Μετά την απαλοιφή της τάσης και της εποχικότητας και του στοχαστικού μέρους της χρονοσειράς, είναι πλέον εφικτή η δημιουργία πρόβλεψης και αξιολόγησης των αποτελεσμάτων. Παρακάτω φαίνονται τα αποτελέσματα που προέκυψαν από την πρόβλεψη που έγινε με την βοήθεια του μοντέλου ARIMA(3,1,12) μέχρι και 24 ώρες μπροστά. Τα ωριαία βήματα της πρόβλεψης είναι 2,6,12,24. Η σύγκριση των αποτελεσμάτων των πραγματικών τιμών και των προβλεπόμενων τιμών έγινε με την βοήθεια μέτρων επιβεβαίωσης και συγκεκριμένα των RMSE (Root Mean Square Error) και MAE (Mean Absolute Error).

Ώρα (h)	RMSE (m/s)	MAE (m/s)
2	0.3083	0.2612
6	1.5284	1.2546
12	2.9399	2.5454
24	4.4045	4.0075

Πίνακας 5.16: Πίνακας με τις τιμές των μέτρων RMSE και MAE για τα ωριαία βήματα 2,6,12,24.

Παρατηρώντας τον Πίνακα 5.16 είναι προφανές ότι με την αύξηση του ωριαίου βήματος προκαλούνται μεγαλύτερα σφάλματα. Οι τιμές και για τα δύο σφάλματα, RMSE και MAE, αυξάνονται όσο μεγαλώνει το βήμα πρόβλεψης.

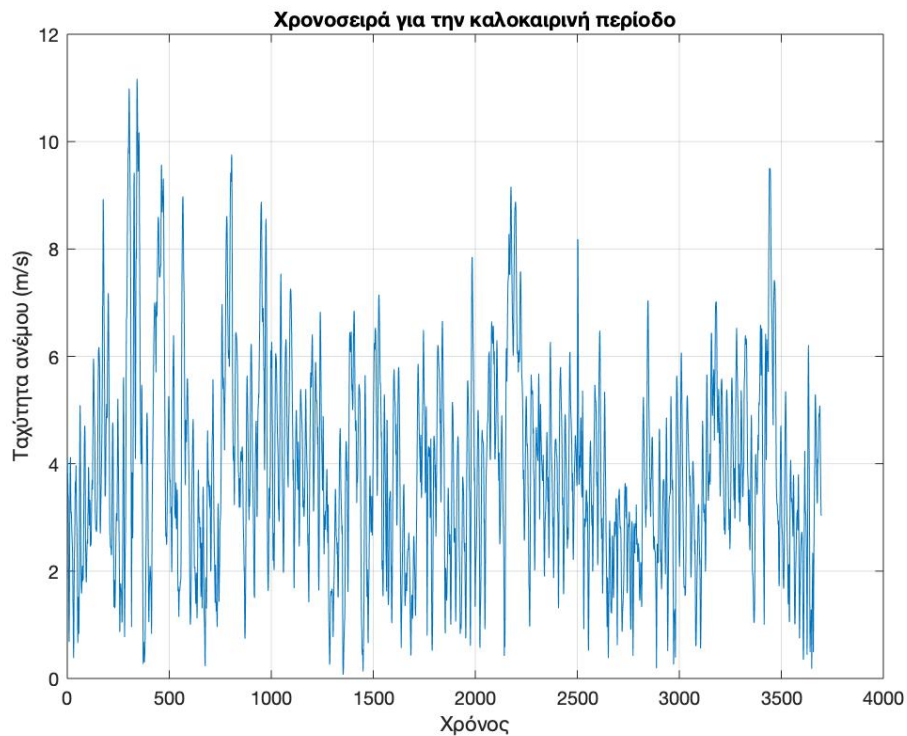
Στο Σχήμα 5.17 φαίνεται στο διάγραμμα με μπλε χρώμα οι πραγματικές τιμές της ταχύτητας ανέμου και με πορτοκαλί χρώμα τα δεδομένα με την χρήση του μοντέλου ARIMA(3,1,12). Αυτό που παρατηρούμε είναι ότι τα δεδομένα με την χρήση του μοντέλου ARIMA(3,1,12) προσαρμόζονται αρκετά καλά στην χρονοσειρά των πραγματικών τιμών της ταχύτητας ανέμου.



Σχήμα 5.17: Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου και πρόβλεψης με το μοντέλο ARIMA(3,1,12).

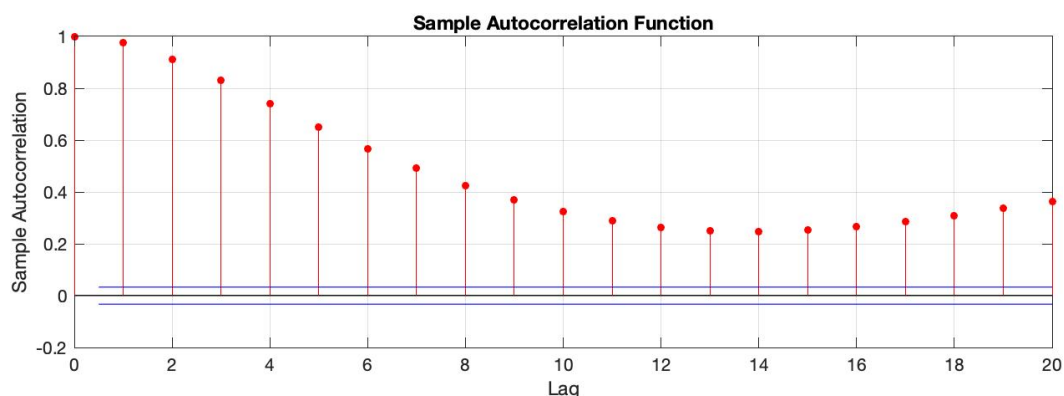
5.4.2 Χρονοσειρά με ωριαίο βήμα για την καλοκαιρινή περίοδο για την περιοχή Τροόδος

Για την καλοκαιρινή περίοδο έγινε επιλογή των δεδομένων από τις 10 Μαΐου έως 10 Οκτωβρίου για την περιοχή Τρόδος. Τα δεδομένα μειώθηκαν από 8784 σε 3696. Η γραφική απεικόνιση της χρονοσειράς για το συγκεκριμένο διάστημα παρουσιάζεται στο Σχήμα 5.18.



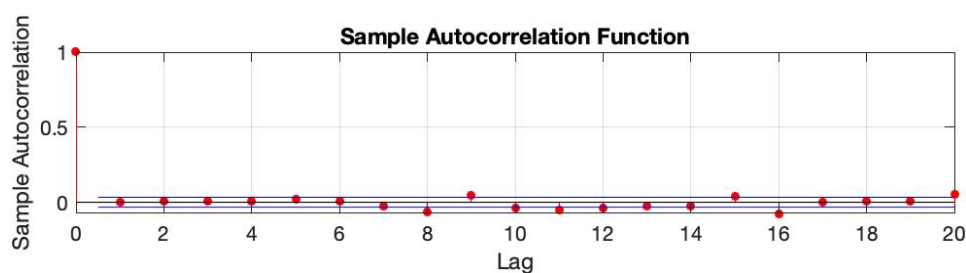
Σχήμα 5.18: Διάγραμμα χρονοσειράς με ωριαίο βήμα για την καλοκαιρινή περίοδο της περιοχής Τροόδος.

Από το παραπάνω διάγραμμα συμπεραίνουμε την παρουσία τάσης και περιοδικότητας αφού μπορούμε να παρατηρήσουμε τις διακυμάνσεις που παρουσιάζει η χρονοσειρά σε σχέση με τον χρόνο. Αυτό θα γίνει πιο ευδιάκριτο με την βοήθεια του διαγράμματος αυτοσυσχέτισης που φαίνεται παρακάτω, μιας και αποτελεί δείγμα για την αναγνώριση τόσο της τάσης όσο και της περιοδικότητας.

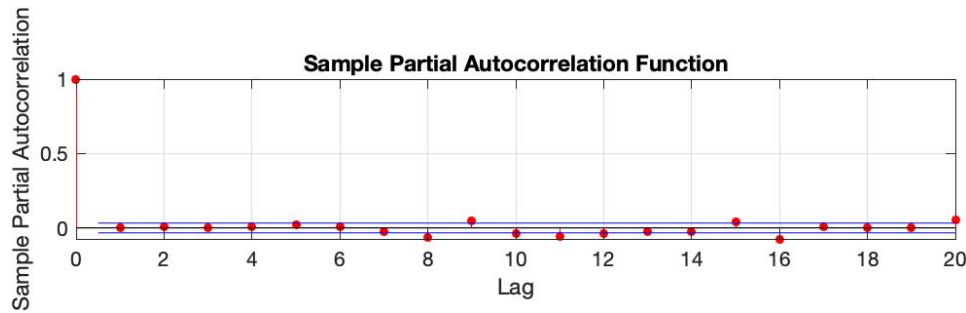


Σχήμα 5.19: Διάγραμμα αυτοσυσχέτισης για την καλοκαιρινή περίοδο.

Όπως μπορούμε να διακρίνουμε τα στοιχεία της χρονοσειράς είναι εκτός των ορίων εμπιστοσύνης (confidence bounds) και επομένως η χρονοσειρά παρουσιάζει τάση και περιοδικότητα την δεδομένη στιγμή. Για να απαλλαγούμε από την παρουσία αυτοσυσχετίσεων, θα εφαρμόσουμε το μοντέλο ARIMA(3,1,8). Η επιλογή του μοντέλου ARIMA(3,1,8) έγινε όπως προηγουμένως με την βοήθεια των κριτηρίων AIC και BIC, επιλέγοντας τον καλύτερο συντελεστή. Στην συνέχεια τα δύο διαγράμματα που ακολουθούν δείχνουν την μερική απαλοιφή των αυτοσυσχετίσεων μεταξύ των τιμών αφού ιδιαίτερα μεταξύ στην περίπτωση του διαγράμματος αυτοσυσχέτισης οι τιμές της χρονοσειρά βρίσκονται σχεδόν μέσα στα όρια εμπιστοσύνης.



Σχήμα 5.20: Διάγραμμα αυτοσυσχέτισης για ωριαία χρονοσειρά για την καλοκαιρινή περίοδο μετά την απαλοιφή των αυτοσυσχετίσεων.

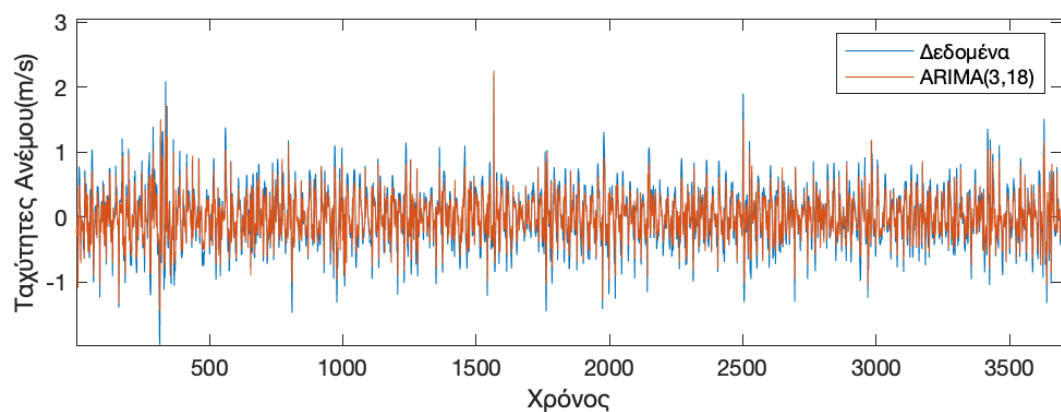


Σχήμα 5.21: Διάγραμμα μερικής αυτοσυσχέτισης για ωριαία χρονοσειρά για την καλοκαιρινή περίοδο μετά την απαλοιφή των αυτοσυσχετίσεων.

Στην συνέχεια αφού έγινε η αφαίρεση του στοχαστικού και αιτιοκρατικού μέρους γίνεται η δημιουργία πρόβλεψης της χρονοσειράς. Για να μπορεί να γίνει η πρόβλεψη με το μοντέλο ARIMA εκπαιδεύουμε το μοντέλο με την χρονοσειρά των $L = 3697$ δεδομένων με το στοιχείο που αντιστοιχεί στην χρονική στιγμή t_1 έως το στοιχείο την χρονική στιγμή t_{L+i} . Γίνεται η εκτίμηση της ταχύτητας ανέμου για την χρονική στιγμή t_{L+2} για 2 ώρες μετά, για 6 ώρες μετά t_{L+6} , για 12 ώρες μετά t_{L+12} και για 24 ώρες μετά t_{L+24} . Οι συγκρίσεις των αποτελεσμάτων μεταξύ των προβλεπόμενων και των πραγματικών τιμών παρουσιάζονται στον παρακάτω πίνακα με την βοήθεια των μέτρων επιβεβαίωσης.

Ώρα (h)	RMSE (m/s)	MAE (m/s)
2	0.2458	0.2661
6	1.2694	1.0418
12	2.2842	1.9984
24	3.0455	2.8241

Πίνακας 5.17: Πίνακας με τις τιμές των μέτρων RMSE και MAE για τα ωριαία βήματα 2,6,12,24 για την καλοκαιρινή περίοδο.



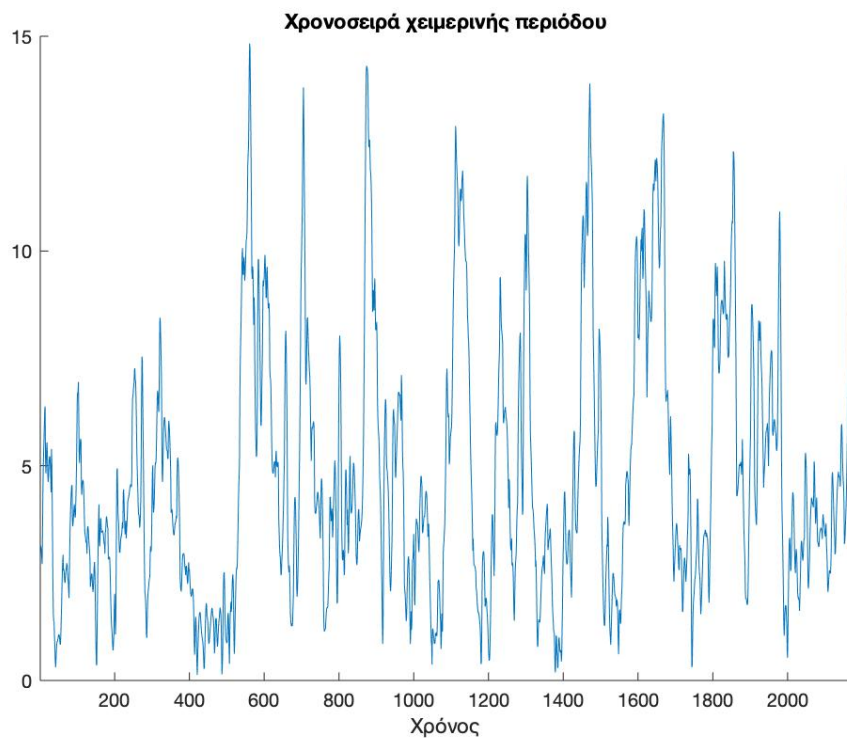
Σχήμα 5.22: Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου και πρόβλεψης για την καλοκαιρινή περίοδο με το μοντέλο ARIMA(3,1,8).

Μπορεί να δηλωθεί με βεβαιότητα ότι μια μεγαλύτερη περίοδος πρόβλεψης της ταχύτητας του ανέμου προκαλεί μεγαλύτερα σφάλματα. Επιπλέον, η πρόβλεψη της ταχύτητας του ανέμου γίνεται όλο και πιο ανακριβής μετά από μια μεγάλη χρονική περίοδο. Όσον αφορά το Σχήμα 5.22 παρατηρούμε είναι ότι τα δεδομένα με την χρήση του μοντέλου ARIMA(3,1,12) προσαρμόζονται αρκετά καλά στην χρονοσειρά των πραγματικών τιμών της ταχύτητας ανέμου

5.4.3 Χρονοσειρά με ωριαίο βήμα για την χειμερινή περίοδο για την περιοχή Τροόδος

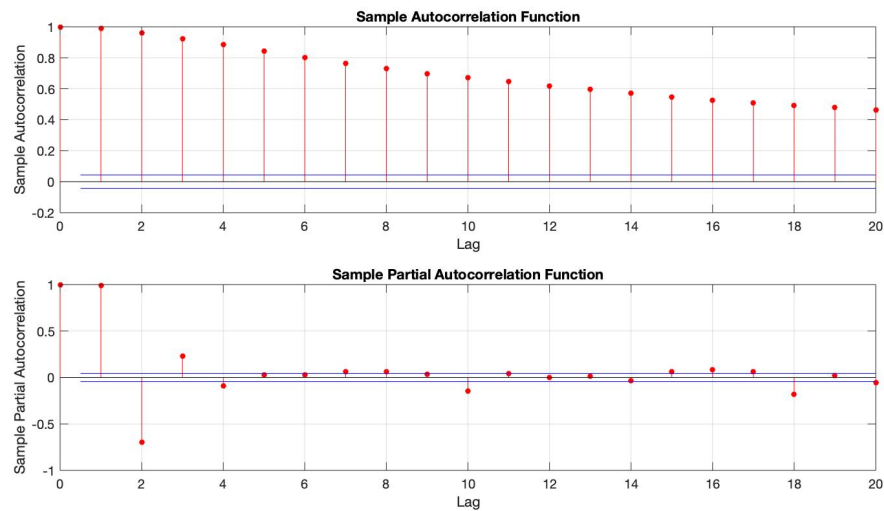
Στο τελευταίο κομμάτι της πρόβλεψης με το αυτοπαλινδρούμενο μοντέλο, πήραμε ένα ακόμα κομμάτι από το σύνολο της χρονοσειράς, το οποίο αναφέρεται στην χειμερινή περίοδο μεταξύ 1ης Δεκεμβρίου 2019 και 29ης Φεβρουαρίου 2020 για την περιοχή Τροόδος.

Το διάγραμμα της χρονοσειράς για την περίοδο αυτή παρουσιάζεται στο παρακάτω Σχήμα 5.23.



Σχήμα 5.23: Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου για την χειμερινή περίοδο για την περιοχή Τροόδος.

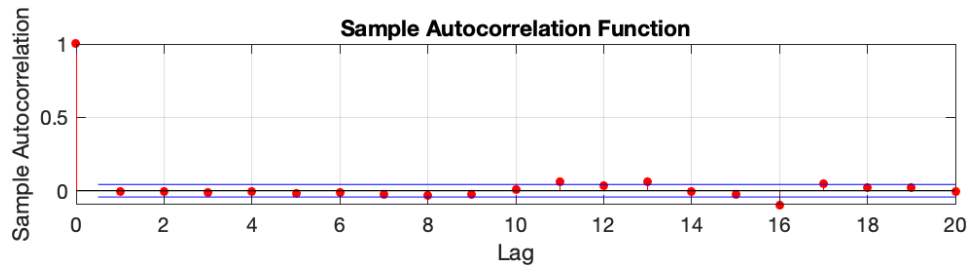
Το διάγραμμα όπως παρατηρούμε παρουσιάζει τάση και περιοδικότητα, αφού διακρίνονται οι μεγάλες διακυμάνσεις που παρουσιάζει η χρονοσειρά σε σχέση με τον χρόνο. Στο Σχήμα 5.24 παρουσιάζεται το διάγραμμα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης της χρονοσειράς τα οποία αποτελούν καλύτερα δείγματα για την αναγνώριση τόσο της τάσης όσο και της περιοδικότητας.



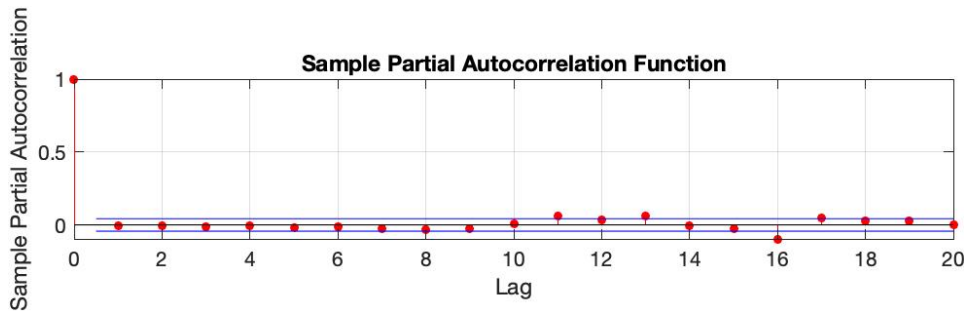
Σχήμα 5.24: Διάγραμμα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για την χειμερινή περίοδο.

Όπως παρατηρούμε οι τιμές της αυτοσυσχέτισης και της μερικής αυτοσυσχέτισης είναι εκτός των ορίων εμπιστοσύνης (confidence bounds), που ορίζουν οι γαλάζιες γραμμές στο Σχήμα 5.24 και έτσι η χρονοσειρά δεν μπορεί να χαρακτηριστεί ως λευκός θόρυβος. Η πολύ αργή μείωση των συσχετίσεων υποδεικνύει την απουσία στασιμότητας.

Όπως και στην περίπτωση της καλοκαιρινής περιόδου έτσι και τώρα για να εκμεταλλευτούμε τις αυτοσυσχετίσεις θα εφαρμόσουμε το μοντέλο $ARIMA(3,1,10)$. Η επιλογή του μοντέλου $ARIMA(3,1,10)$ έγινε όπως προηγουμένως με την βοήθεια των κριτηρίων AIC και BIC, επιλέγοντας τον καλύτερο συντελεστή. Τα δύο διαγράμματα του Σχήματος 5.25 που ακολουθεί και αναφέρονται στα κατάλοιπα (residuals) του μοντέλου δείχνουν την μερική απαλοιφή των αυτοσυσχετίσεων. Αυτή είναι ιδιαίτερα εμφανής στην περίπτωση του διαγράμματος αυτοσυσχέτισης οι τιμές της οποίας βρίσκονται σχεδόν μέσα στα όρια εμπιστοσύνης.



(α') Διάγραμμα αυτοσυσχέτισης



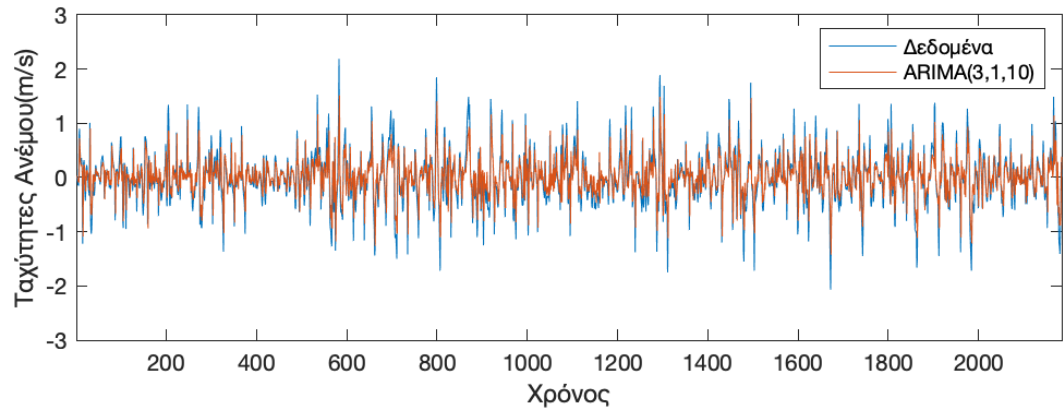
(β') Διάγραμμα μερικής αυτοσυσχέτισης

Σχήμα 5.25: Διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υπολοίπων από το βέλτιστο μοντέλο ARIMA.

Στην συνέχεια για να μπορέσει να γίνει η πρόβλεψη με το μοντέλο ARIMA εκπαιδεύουμε την χρονοσειρά των $L = 2184$ δεδομένων με το στοιχείο που αντιστοιχεί στην χρονική στιγμή t_1 έως το στοιχείο την χρονική στιγμή t_{L+i} . Γίνεται η εκτίμηση της ταχύτητας ανέμου για την χρονική στιγμή t_{L+2} για 2 ώρες μετά, για 6 ώρες μετά t_{L+6} , για 12 ώρες μετά t_{L+12} και για 24 ώρες μετά t_{L+24} . Οι συγκρίσεις των αποτελεσμάτων μεταξύ των προβλεπόμενων και των πραγματικών τιμών γίνονται τόσο με την βοήθεια των μέτρων επιβεβαίωσης που παρουσιάζονται στον Πίνακα 5.18 όσο και με την βοήθεια του διαγράμματος του Σχήματος 5.26.

Ώρα (h)	RMSE (m/s)	MAE (m/s)
2	0.3132	0.2644
6	1.7343	1.4008
12	3.8317	3.2296
24	6.1045	5.4816

Πίνακας 5.18: Πίνακας με τις τιμές των μέτρων RMSE και MAE για τα ωριαία βήματα 2,6,12,24 για την χειμερινή περίοδο.



Σχήμα 5.26: Διάγραμμα πραγματικών τιμών ταχύτητας ανέμου και πρόβλεψης για την χειμερινή περίοδο με το μοντέλο ARIMA(3,1,10).

Παρατηρώντας τον Πίνακα 5.18 αλλά και το Σχήμα 5.26 γίνεται αντιληπτό πως όσο μεγαλώνει το βήμα πρόβλεψης τόσο μεγαλύτερα είναι τα σφάλματα του μοντέλου ARIMA, όπως έχει παρατηρηθεί και για τις προηγούμενες περιπτώσεις. Επομένως οι προβλέψεις που πραγματοποιήθηκαν δεν μπορούν να χαρακτηριστούν αξιόπιστες. Αυτό ίσως να οφείλεται στην απόκλιση της χρονοσειράς από την κανονική κατανομή όπως αναφέρθηκε και στην στατιστική ανάλυση που έγινε. Επίσης οφείλεται στο γεγονός ότι οι συσχετίσεις μεταξύ των τιμών της χρονοσειράς ταχύτητας ανέμου μειώνονται όσο αυξάνεται η απόσταση.

Κεφάλαιο 6

Συμπεράσματα

Στην παρούσα διπλωματική εργασία πραγματοποιήθηκε βραχυπρόθεσμη πρόβλεψη της ταχύτητας ανέμου χρησιμοποιώντας μοντέλα ARI-MA (Autoregressive Integrated Moving Average). Ο στόχος της μελέτης αυτής είναι η δημιουργία αξιόπιστων προβλέψεων της ταχύτητας ανέμου με χρονικό βήμα μιας ώρας για χρονικούς ορίζοντες από λίγες ώρες έως μια ημέρα.

Τα δεδομένα της εργασίας είναι δεδομένα επανανάλυσης (reanalysis) που αποκτήθηκαν από την βάση δεδομένων SoDa. Η βάση δεδομένων SoDa αποτελεί πρόσβαση σε πολλές ηλιακές, μετεωρολογικές, υψομετρικές, αστρονομικές και ατμοσφαιρικές υπηρεσίες. Επίσης, παρέχει υπηρεσίες δεδομένων ηλιακής ακτινοβολίας και μετεωρολογικών δεδομένων για την αναζήτηση θέσεων εγκατάστασης ηλιακών φωτοβολταϊκών και θερμικών εγκαταστάσεων και την παρακολούθηση και πρόβλεψη της παραγωγής ηλεκτρικής ενέργειας. Τα δεδομένα της εργασίας περιλαμβάνουν 52704 μετρήσεις από την 1η Ιανουαρίου 2020 έως την 31η Δεκεμβρίου 2020 με χρονικό βήμα 10 λεπτά της ώρας.

Αρχικά πραγματοποιήθηκε διερευνητική στατιστική ανάλυση για τις χρονοσειρές με βήμα 10 λεπτών. Διαπιστώνεται ότι αυτές παρουσιάζουν σημαντικές αποκλίσεις από την κανονική κατανομή. Η αντιμετώπιση του προβλήματος αυτού έγινε με τον μετασχηματισμό της χρονοσειράς σε ωριαίες τιμές, κάτι το οποίο οδηγεί σε μείωση του πλήθους των δεδομένων

από 52704 σε 8784. Επειδή υπάρχουν σημαντικές διαφοροποιήσεις ανάμεσα στην καλοκαιρινή και χειμερινή περίοδο, η χρονοσειρά διαχωρίζεται σε δύο χρονικές περιόδους, την καλοκαιρινή περίοδο από 10 Μαΐου έως 10 Οκτωβρίου 2020 και την χειμερινή περίοδο από 1 Δεκεμβρίου 2019 έως 29 Φεβρουαρίου 2020. Οι ωριαίες χρονοσειρές περιέχουν αντίστοιχα 3696 τιμές για την καλοκαιρινή περίοδο και 2184 τιμές για την χειμερινή περίοδο.

Με την βοήθεια των μοντέλων ARIMA πραγματοποιήθηκε η διαδικασία πρόβλεψης για τις χρονοσειρές με ωριαίο βήμα. Για την ετήσια χρονοσειρά της περιοχής Τροόδους μέσω των διαγραμμάτων αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) παρατηρήθηκε η ύπαρξη τάσης. Η απαλοιφή της τάσης έγινε με διαφόριση μέσω του μοντέλου ARIMA. Χρησιμοποιώντας τα κριτήρια AIC και BIC γίνεται η επιλογή του καταλληλότερου μοντέλου ARIMA. Για την συγκεκριμένη χρονοσειρά αυτό είναι το μοντέλο ARIMA(3,1,12). Στην συνέχεια γίνεται διαγνωστικός έλεγχος για τα κατάλοιπα του μοντέλου (residuals). Όπως προκύπτει από τον έλεγχο αυτό, το μοντέλο ARIMA προσαρμόζεται αρκετά καλά στις ωριαίες χρονοσειρές.

Η αξιολόγηση των αποτελεσμάτων της πρόβλεψης του μοντέλου ARIMA πραγματοποιείται για χρονικό ορίζοντα δύο, έξι, δώδεκα και 24 ωρών (1 ημέρα). Τα αποτελέσματα τόσο στην τοποθεσία Αγρού όσο και στην τοποθεσία Τροόδους ήταν παραπλήσια επομένως αναφερόμαστε μόνο στην μία περιοχή (Τρόοδος). Η αξιολόγηση των αποτελεσμάτων της πρόβλεψης βασίζεται στην μεθοδολογία της διασταυρωτικής επιβεβαίωσης (επικύρωσης) και περιλαμβάνει τα στατιστικά μέτρα RMSE (Root Mean Square Error) και MAE (Mean Absolute Error). Σημειώνεται ότι δεν χρησιμοποιήθηκε το κριτήριο MAPE (Mean Absolute Percentage Error) γιατί δεν ενδείκνυται για την ανάλυση της ταχύτητας του ανέμου, η οποία περιλαμβάνει και τιμές ίσες ή κοντά στην μηδενική ταχύτητα. Τα αποτελέσματα του μοντέλου ARIMA για την ετήσια χρονοσειρά έδειξαν ότι το RMSE της πρόβλεψης 2 ωρών ήταν 0.31 m/s (7,6%), για τις επόμενες 6 ώρες ήταν 1.53 m/s (37.7%), για τις 12 ώρες 2.94 m/s (72,5%) και για τις 24 ώρες 4.40 m/s (95.7%).

Όσον αφορά τις προβλέψεις που πραγματοποιήθηκαν για την καλοκαιρινή και χειμερινή περίοδο, ακολουθήθηκε η ίδια διαδικασία που έγινε και για την ετήσια χρονοσειρά. Για την καλοκαιρινή περίοδο χρησιμοποιήθηκε το μοντέλο ARIMA(3,1,8). Όπως και προηγουμένως έγινε η πρόβλεψη για 2, 6, 12 και 24 ώρες. Τα συνοπτικά αποτελέσματα έδειξαν ότι μικρότερο RSME επιτεύχθηκε για τον ορίζοντα των δύο ωρών με αποτέλεσμα 0.25 m/s (6.46%) και η υψηλότερη τιμή RMSE παρατηρήθηκε για τον ορίζοντα των 24 ωρών με τιμή 3.05 m/s (78.82%). Για την χειμερινή περίοδο χρησιμοποιήθηκε το μοντέλο ARIMA(3,1,10) και για τον ίδιο χρονικό ορίζοντα πρόβλεψης τα αποτελέσματα κυμαίνονται από 0.31 m/s (6.48%) έως 6.10 m/s (127.5%). Έτσι καταλήγουμε στο συμπέρασμα πως η πρόβλεψη για την καλοκαιρινή περίοδο είναι καλύτερη από την αντίστοιχη για την χειμερινή περίοδο.

Από την διαδικασία των προβλέψεων της ταχύτητας ανέμου για τις τρεις χρονοσειρές καταλήγουμε στο συμπέρασμα ότι η ακρίβεια της πρόβλεψης μειώνεται όσο αυξάνεται το βήμα πρόβλεψης. Συγκεκριμένα, η πρόβλεψη και στις τρεις περιόδους για τον χρονικό ορίζοντα των δύο ωρών ήταν αρκετά αξιόπιστη γιατί το ποσοστό σφάλματος ήταν κάτω από το 10%. Επιπλέον, η περίοδος του καλοκαιριού έδωσε την δυνατότητα καλύτερης εκτίμησης από την χειμερινή περίοδο. Αυτό είναι αναμενόμενο αν αναλογιστούμε ότι τους χειμερινούς μήνες ο καιρός είναι άστατος σε σύγκριση με τους καλοκαιρινούς μήνες όπου οι καιρικές συνθήκες είναι καλύτερες. Παρόλα αυτά και για τις τρεις περιόδους (ολόκληρο το έτος, καλοκαιρινή και χειμερινή περίοδος) οι προβλέψεις έδωσαν σημαντικά ποσοστά σφάλματος.

Για να μειωθούν τα σφάλματα που προκύπτουν, θα μπορούσε σε μια μελλοντική έρευνα να εξεταστεί η χρήση ενός υβριδικού μοντέλου (ARIMA σε συνδυασμό με άλλες τεχνικές). Επίσης θα ήταν ενδιαφέρουσα μια σύγκριση διαφόρων μεθόδων, όπως αυτή των νευρωνικών δικτύων σε συνδυασμό με την χρήση υβριδικών μοντέλων. Επιπλέον θα είχε ενδιαφέρον να γίνει η ανάλυση των χρονοσειρών με χρονικό βήμα μεγαλύτερο της μίας ώρας. Επίσης θα μπορούσαν να αξιοποιηθούν δεδομένα για μεγαλύτερο χρονικό διάστημα. Με αυτό τον τρόπο θα μπορέσουμε να

μελετήσουμε καλύτερα αν υπάρχουν τάσεις κατά την διάρκεια πολλών ετών και πιθανές περιοδικότητες.

Bibliography

- [1] Donald W. K. Andrews and Werner Ploberger. Testing for serial correlation against an ARMA(1, 1) process. *Journal of the American Statistical Association*, 91(435):1331–1342, 1996.
- [2] P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer International Publishing, 2016.
- [3] Barbara G. Brown, Richard W. Katz, and Allan H. Murphy. Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, (8):1184–1195, 1984.
- [4] Erasmo Cadenas and Wilfrido Rivera. Wind speed forecasting in the South Coast of Oaxaca, México. *Renewable Energy*, pages 2116–2128, 2007.
- [5] Global Wind Energy Council. Annual global wind reports. Global Wind Energy Council.
- [6] Francisco de Castro. Function “fitmethis.m”. MATLAB Central File Exchange. Retrieved November 4, 2021.
- [7] Robert Nau. General seasonal ARIMA models. Fuqua School of Business, Duke University.
- [8] J.L. Torres, A. García, M. De Blas, and A. De Francisco. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, pages 65–77, 2005.

-
- [9] Tyass, Ilham, Bellat, Abdelouahad, Raihani, Abdelhadi, Mansouri, Khalifa, and Khalili, Tajeddine. Wind speed prediction based on seasonal ARIMA model. *E3S Web Conf.*, 2022.
- [10] Jing Zhao, Jianzhou Wang, and Zhongyue Su. Power generation and renewable potential in China. *Renewable and Sustainable Energy Reviews*, 40:727–740, 2014.
- [11] Μαριάννα Δ.Τάτση. Η Γενικευμένη κατανομή ακραίων τιμών και εφαρμογές της στον αναλογισμό. Technical report, Πανεπιστήμιο Πειραιώς, 2015.
- [12] Γεωργία Κολομβόκη. Μέθοδοι πρόβλεψης χρονοσειρών: χρονοσειρές στην ελληνική οικονομία. Technical report, Πανεπιστήμιο Πατρών, 2017.
- [13] Δημήτρης Κουγιουμτζής. Ανάλυση Χρονοσειρών. Σημειώσεις του μεταπτυχιακού μαθήματος “Στατιστική και Μοντελοποίηση” του Τμήματος Μαθηματικών ΑΠΘ.
- [14] Θανάσης Μανώλης. Στατιστική ανάλυση δεδομένων παραγωγής ηλεκτρικής ενέργειας, 2019. Διπλωματική εργασία, Πολυτεχνείο Κρήτης.