



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΗΣ ΙΣΧΥΟΣ

Πρόβλεψη αιολικής και φωτοβολταϊκής
παραγωγής με χρήση τεχνικών μηχανικής
μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΚΩΝΣΤΑΝΤΙΝΟΥ ΠΑΠΑΘΑΝΑΣΙΟΥ

Επιβλέπων: Σταύρος Παπαθανασίου
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2021



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικής Ισχύος

Πρόβλεψη αιολικής και φωτοβολταϊκής
παραγωγής με χρήση τεχνικών μηχανικής
μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΚΩΝΣΤΑΝΤΙΝΟΥ ΠΑΠΑΘΑΝΑΣΙΟΥ

Επιβλέπων: Σταύρος Παπαθανασίου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 15η Μαρτίου 2021.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Σταύρος Παπαθανασίου
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Κορρές
Καθηγητής Ε.Μ.Π.

.....
Παύλος Γεωργιλάκης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2021

(Υπογραφή)

.....

ΚΩΝΣΤΑΝΤΙΝΟΣ ΠΑΠΑΘΑΝΑΣΙΟΥ

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

© 2021 – All rights reserved



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικής Ισχύος

Copyright ©–All rights reserved Κωνσταντίνος Παπαθανασίου, 2021.

Με επιφύλαξη παντός δικαιώματος.

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

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον κύριο Σταύρο Παπαθανασίου, Καθηγητή της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου και επιβλέποντα της παρούσας διπλωματικής εργασίας, για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου αυτό το θέμα, αλλά και για την συνεχή στήριξη και καθοδήγησή του, καθ' όλη την διάρκεια της εκπόνησής του.

Επίσης, ευχαριστώ ιδιαίτερα τον διδάκτορα της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου κύριο Γεώργιο Ψαρρό για τις πολύτιμες συμβουλές και την προθυμία του, όπως και για την άψογη συνεργασία μας κατά την συγγραφή της εργασίας αυτής.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου, για την υποστήριξη και για την ενθάρρυνση τους σε όλα τα χρόνια των σπουδών μου.

Περίληψη

Η αυξανόμενη διείσδυση των Ανανεώσιμων Πηγών Ενέργειας (ΑΠΕ) στα συστήματα παραγωγής ηλεκτρικής ενέργειας, θέτει προκλήσεις αναφορικά με την αξιόπιστη και οικονομική λειτουργία τους. Η τυχαιότητα και η αβεβαιότητα της αιολικής και της φωτοβολταϊκής παραγωγής, εξαιτίας της στοχαστικής τους φύσης, συχνά ενδέχεται να επηρεάσουν την ποιότητα της παραγόμενης ενέργειας, ενώ καθιστούν τις αποφάσεις ένταξης μονάδων και διατήρησης εφεδρειών ιδιαίτερα πολύπλοκες. Η ύπαρξη, επομένως, προβλέψεων της παραγωγής τους κρίνεται αναγκαία και αυξανόμενης σημασίας, για την μετάβαση στην δημιουργία συστημάτων ηλεκτρικής ενέργειας βασισμένα σε ανανεώσιμες πηγές ενέργειας.

Αντικείμενο της παρούσας διπλωματικής αποτελεί η δημιουργία προβλέψεων για την αιολική και την φωτοβολταϊκή παραγωγή με χρονικό ορίζοντα 24 ωρών, χρησιμοποιώντας τεχνικές μηχανικής μάθησης, και συγκεκριμένα τα τεχνητά νευρωνικά δίκτυα. Εξ αυτών, θα εξεταστούν δυο κατηγορίες, τα Αναδρομικά Νευρωνικά Δίκτυα Μακράς και Βραχείας Μνήμης (Long and Short Term Memory ή LSTM) και τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks ή CNN). Για την μελέτη της αποδοτικότητας τους θα υλοποιηθούν δυο παραλλαγές της κάθε κατηγορίας νευρωνικών δικτύων, τα αναδρομικά νευρωνικά δίκτυα (LSTM) με ή χωρίς την χρήση της αρχιτεκτονικής Κωδικοποιητή-Αποκωδικοποιητή (Encoder-Decoder) και τα συνελικτικά νευρωνικά δίκτυα (TCN) και (WaveNet).

Εκτός των διαφορετικών νευρωνικών δικτύων θα δοκιμαστούν και διαφορετικές προσεγγίσεις στην δημιουργία των μοντέλων πρόβλεψης. Πιο αναλυτικά, για την αιολική παραγωγή θα ελεγχθεί η διαφοροποίηση των σφαλμάτων πρόβλεψης, είτε προβλέποντας την συνολική αιολική παραγωγή για μια περιοχή, είτε προβλέποντας την αιολική παραγωγή για κάθε αιολικό πάρκο ξεχωριστά, ενώ για την φωτοβολταϊκή παραγωγή θα εξεταστεί η πρόβλεψη της στο επίπεδο του Ελληνικού συστήματος χρησιμοποιώντας ιστορικές παρατηρήσεις με ή χωρίς επιπλέον χαρακτηριστικά.

Τέλος θα εκτιμηθεί η απόδοση του κάθε μοντέλου πρόβλεψης, συγκρίνοντας τα σφάλματα για την μηνιαία και την ετήσια πρόβλεψη της αιολικής και φωτοβολταϊκής παραγωγής, ενώ θα εξαχθούν συμπεράσματα για την αποτελεσματικότητα του κάθε μοντέλου.

Λέξεις Κλειδιά

Πρόβλεψη Χρονοσειρών, Ανανεώσιμες Πηγές Ενέργειας, Νευρωνικά Δίκτυα, Αναδρομικά Νευρωνικά Δίκτυα, Αρχιτεκτονική Κωδικοποιητή-Αποκωδικοποιητή, Συνελικτικά Νευρωνικά Δίκτυα.

Abstract

The increasing penetration of Renewable Energy Sources (RES) in the electric power systems, poses challenges regarding their reliable and economic operation. The randomness and uncertainty of wind and solar power generation due to their stochastic nature, often render unit commitment and reserve requirement decisions highly complex. Therefore, power generation predictions are necessary and of increasing importance for the transition to energy power systems based on RES.

The main objective of this diploma thesis is the creation of day-ahead wind and solar power predictions, using machine learning techniques, specifically, artificial neural networks. Two different categories will be examined, Long and Short Term Memory Recurrent Neural Networks and Convolutional Neural Networks. In order to study their effectiveness two sub models of each category will be created, recurrent neural networks (LSTM), with and without the use of the Encoder-Decoder Architecture and convolutional neural networks (TCN) and (WaveNet).

Except for the different neural networks, different approaches regarding the creation of the prediction models will be considered too. In more detail, as far as the wind power is concerned, the differences in error measurements, while predicting the whole wind generation or the generation of each individual wind park will be tested. Whereas in the solar power, the prediction of the whole Hellenic Energy system will be examined, by using historical observations with or without extra features.

Lastly, the effectiveness of each model will be identified, comparing the wind and solar power predictions error measurements monthly and yearly, whilst conclusions will be inferred, as regards each model's efficiency.

Keywords

Time series Forecasting, Renewable Energy Sources, Neural Networks, Recurrent Neural Networks, Encoder-Decoder Architecture, Convolutional Neural Networks.

Περιεχόμενα

Ευχαριστίες	1
Περίληψη	3
Abstract	5
Περιεχόμενα	8
Κατάλογος Σχημάτων	11
Κατάλογος Πινάκων	13
1 Εισαγωγή	15
1.1 Αντικείμενο της διπλωματικής εργασίας	15
1.2 Επισκόπηση της σύγχρονης βιβλιογραφίας	15
1.3 Δομή της εργασίας	18
2 Ανανεώσιμες πηγές ενέργειας	21
2.1 Γενικά για τις ανανεώσιμες πηγές ενέργειας	21
2.2 Αξία πρόβλεψης αιολικής παραγωγής	22
2.3 Αξία πρόβλεψης φωτοβολταϊκής παραγωγής	23
2.4 Αξία πρόβλεψης φορτίου	24
3 Θεωρητικό υπόβαθρο	27
3.1 Εισαγωγή στις χρονοσειρές	27
3.1.1 Ποιοτικά χαρακτηριστικά χρονοσειρών	27
3.1.2 Στατιστική ανάλυση χρονοσειρών	30
3.1.3 Στασιμότητα (Stationarity)	32
3.2 Εισαγωγή στα νευρωνικά δίκτυα	36
3.2.1 Διαδικασίες μηχανικής μάθησης	36
3.2.2 Βιολογικά νευρωνικά δίκτυα	37
3.2.3 Τεχνητά νευρωνικά δίκτυα	38
3.2.4 Συναρτήσεις ενεργοποίησης	40

3.2.5	Συνάρτηση κόστους (Cost function)	43
3.2.6	Αλγόριθμος οπισθοδιάδοσης σφάλματος (Backpropagation)	44
3.2.7	Πρόληψη υπερπροσαρμογής	45
3.2.8	Αλγόριθμοι βελτιστοποίησης	49
3.2.9	Αναδρομικά νευρωνικά δίκτυα (RNN)	54
3.2.10	Νευρωνικό δίκτυο μακράς-βραχείας μνήμης (LSTM)	56
3.2.11	Νευρωνικό δίκτυο μακράς-βραχείας μνήμης διπλής κατεύθυνσης (Bidirectional-LSTM)	59
3.2.12	Αρχιτεκτονική Encoder-Decoder	59
3.2.13	Συνελικτικά νευρωνικά δίκτυα (CNN)	60
3.2.14	Temporal Convolutional Networks (TCN)	65
3.2.15	WaveNet	69
3.2.16	Μετρικές σφαλμάτων	70
4	Ανάλυση και προεπεξεργασία δεδομένων	75
4.1	Δεδομένα αιολικής παραγωγής	75
4.2	Δεδομένα φωτοβολταϊκής παραγωγής	83
5	Εφαρμογή μεθόδων πρόβλεψης και αποτελέσματα	87
5.1	Πρόβλεψη αιολικής παραγωγής	87
5.1.1	Εφαρμογή LSTM νευρωνικού δικτύου	87
5.1.2	Εφαρμογή Encoder-Decoder LSTM νευρωνικού δικτύου	90
5.1.3	Εφαρμογή TCN νευρωνικού δικτύου	92
5.1.4	Εφαρμογή WaveNet νευρωνικού δικτύου	94
5.1.5	Εφαρμογή ξεχωριστής πρόβλεψης για το κάθε αιολικό πάρκο	95
5.1.6	Σύγκριση αποτελεσμάτων	97
5.1.7	Ετήσια πρόβλεψη αιολικής παραγωγής	99
5.2	Πρόβλεψη φωτοβολταϊκής παραγωγής	105
5.2.1	Εφαρμογή μοντέλου Παραμένουσας Τιμής (Persistence Model)	106
5.2.2	Εφαρμογή LSTM νευρωνικού δικτύου	107
5.2.3	Εφαρμογή WaveNet νευρωνικού δικτύου	109
5.2.4	Εφαρμογή WaveNet νευρωνικού δικτύου με προσθήκη τεχνητών δεδομένων	111
5.2.5	Σύγκριση αποτελεσμάτων	114
5.2.6	Ετήσια πρόβλεψη φωτοβολταϊκής παραγωγής	117
6	Επίλογος	121
6.1	Σύνοψη και συμπεράσματα	121
6.2	Μελλοντικές επεκτάσεις	122
	Βιβλιογραφία	124

Κατάλογος Σχημάτων

3.1	Πληθυσμός Η.Π.Α (Πηγή: [50])	28
3.2	Πωλήσεις κρασιού στην Αυστραλία (Πηγή: [50])	28
3.3	Μέσος όρος μηνιαίας θερμοκρασίας στο Nottingham (Πηγή: [14])	29
3.4	Θερμοκρασία κάστορα (Πηγή: [14])	29
3.5	Απεικόνιση στάσιμης και μη χρονοσειράς (Πηγή: [68])	33
3.6	Βιολογικό νευρωνικό δίκτυο (Πηγή: [41])	38
3.7	Τεχνητό νευρωνικό δίκτυο (Πηγή: [28])	38
3.8	Πολυεπίπεδο τεχνητό νευρωνικό δίκτυο (Πηγή: [71])	40
3.9	Συνάρτηση βήματος	41
3.10	Σιγμοειδής συνάρτηση	42
3.11	Συνάρτηση υπερβολικής εφαπτομένης	42
3.12	Συνάρτηση ReLU	43
3.13	Δυνατές προσαρμογές μοντέλων μηχανικής μάθησης (Πηγή: [70])	46
3.14	Εφαρμογή επιπέδου εγκατάλειψης (Dropout) (Πηγή: [64])	48
3.15	Βέλτιστο σημείο διακοπής μάθησης (Πηγή: [18])	49
3.16	Βελτιστοποίηση συνάρτησης μέσω αλγορίθμου κατάβασης κλίσης (Πηγή: [2]) .	50
3.17	Συνάρτηση κόστους αλγορίθμου στοχαστικής κατάβασης κλίσης (Πηγή: [75])	51
3.18	Αλγόριθμος κατάβασης κλίσης με και χωρίς την χρήση ορμής (Πηγή: [76]) . .	52
3.19	Αναδρομικό νευρωνικό δίκτυο (RNN) (Πηγή: [8])	55
3.20	Κρυφό κύτταρο LSTM (Πηγή: [8])	56
3.21	Κυτταρική κατάσταση LSTM (Πηγή: [8])	57
3.22	Πύλη εισόδου (Πηγή: [8])	58
3.23	Νέα κυτταρική κατάσταση (Πηγή: [8])	58
3.24	Έξοδος κυττάρου LSTM (Πηγή: [8])	59
3.25	Bidirectional LSTM (Πηγή: [7])	59
3.26	Αρχιτεκτονική Encoder Decoder (Πηγή: [58])	60
3.27	Σύγκριση νευρώνων συνελικτικού και απλού νευρωνικού δικτύου (Πηγή: [65])	61
3.28	Αρχιτεκτονική συνελικτικού νευρωνικού δικτύου (Πηγή: [43])	62
3.29	Συνέλιξη τένσορα εισόδου με φίλτρο (Πηγή: [78])	63
3.30	Εφαρμογή του zero padding (Πηγή: [78])	63
3.31	Συγκεντρωτικό επίπεδο (Πηγή: [65])	64

3.32 Causal Convolutions (Πηγή: [48])	66
3.33 Διεσταλμένες συνελίξεις (Dilated Convolution) (Πηγή: [48])	67
3.34 Υπολειμματική συνδέση (Πηγή: [21])	68
3.35 TCN Residual Block (Πηγή: [21])	68
3.36 Αρχιτεκτονική WaveNet νευρωνικού δικτύου (Πηγή: [48])	69
4.1 Καμπύλες ισχύος αιολικών πάρκων	77
4.2 Καμπύλες ισχύος αιολικών πάρκων, χωρίς θόρυβο	77
4.3 Συνάρτηση αυτοσυσχέτισης αιολικών πάρκων	78
4.4 Συνάρτηση μερικής αυτοσυσχέτισης αιολικών πάρκων	78
4.5 Συνάρτηση αυτοσυσχέτισης και μερικής αυτοσυσχέτισης αιολικών πάρκων	79
4.6 Χρονοσειρά συνολικής παραγωγής αιολικής ενέργειας	80
4.7 Χωρισμός χρονοσειράς σε ζεύγη εισόδου-εξόδου (Πηγή : [35])	82
4.8 Συνάρτηση αυτοσυσχέτισης και μερικής αυτοσυσχέτισης παραγωγής φωτοβολταϊκής ενέργειας	84
4.9 Χρονοσειρά παραγωγής φωτοβολταϊκής ενέργειας	84
5.1 Αποτελέσματα πρόβλεψης αιολικής παραγωγής, LSTM	89
5.2 Ημερήσιο σφάλμα πρόβλεψης LSTM	89
5.3 Αποτελέσματα πρόβλεψης αιολικής παραγωγής, Encoder-Decoder LSTM	91
5.4 Ημερήσιο σφάλμα πρόβλεψης Encoder-Decoder LSTM	91
5.5 Αποτελέσματα πρόβλεψης αιολικής παραγωγής, TCN	93
5.6 Ημερήσιο σφάλμα πρόβλεψης TCN	93
5.7 Αποτελέσματα πρόβλεψης αιολικής παραγωγής, WaveNet	95
5.8 Ημερήσιο σφάλμα πρόβλεψης WaveNet	95
5.9 Αποτελέσματα πρόβλεψης αιολικής παραγωγής	96
5.10 Ημερήσιο σφάλμα πρόβλεψης	97
5.11 Σύγκριση μέσου απόλυτου σφάλματος	98
5.12 Σύγκριση ρίζας μέσου τετραγωνικού σφάλματος	98
5.13 Σύγκριση συντελεστή προσδιορισμού	98
5.14 Σύγκριση κατανομής σφάλματος αιολικής παραγωγής	101
5.15 Σύγκριση τιμών ταχύτητας ανέμου με και χωρίς θόρυβο	103
5.16 Σύγκριση κατανομής σφάλματος αιολικής παραγωγής με προσθήκη θορύβου	104
5.17 Σύγκριση κατανομής σφάλματος αιολικής παραγωγής (με και χωρίς την προσθήκη θορύβου)	105
5.18 Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, Persistence Model	107
5.19 Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, LSTM	108
5.20 Ημερήσιο σφάλμα πρόβλεψης LSTM	109
5.21 Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, Wavenet	110
5.22 Ημερήσιο σφάλμα πρόβλεψης WaveNet	110
5.23 Ιστογράμμα φωτοβολταϊκής παραγωγής	112

5.24	Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, με επιπλέον χαρακτηριστικά	113
5.25	Ημερήσιο σφάλμα πρόβλεψης, με επιπλέον χαρακτηριστικά	113
5.26	Σύγκριση μέσου απόλυτου σφάλματος	114
5.27	Σύγκριση ρίζας μέσου τετραγωνικού σφάλματος	115
5.28	Σύγκριση συντελεστή προσδιορισμού	115
5.29	Κατανομή σφάλματος φωτοβολταϊκής παραγωγής	118
5.30	Κατανομή σφάλματος φωτοβολταϊκής παραγωγής, με επιπλέον χαρακτηριστικά	119
5.31	Σύγκριση κατανομής σφάλματος πρόβλεψης φωτοβολταϊκής παραγωγής	120

Κατάλογος Πινάκων

4.1	Augmented Dickey-Fuller test δεδομένων αιολικής παραγωγής	78
4.2	Augmented Dickey-Fuller test συνολικών δεδομένων αιολικής παραγωγής . .	79
4.3	Ασαφοποίηση ταχύτητας ανέμου	81
4.4	Augmented Dickey-Fuller test δεδομένων φωτοβολταϊκής παραγωγής	83
5.1	Υπερπαραμέτροι LSTM	88
5.2	Συνολικό σφάλμα LSTM	90
5.3	Υπερπαραμέτροι Encoder-Decoder LSTM	91
5.4	Συνολικό σφάλμα Encoder-Decoder LSTM	92
5.5	Υπερπαραμέτροι TCN	92
5.6	Συνολικό σφάλμα TCN	93
5.7	Υπερπαραμέτροι WaveNet	94
5.8	Συνολικό σφάλμα WaveNet	95
5.9	Συνολικό σφάλμα ξεχωριστής πρόβλεψης αιολικών πάρκων	97
5.10	Ετήσιο σφάλμα πρόβλεψης	99
5.11	Ετήσιο σφάλμα ξεχωριστής πρόβλεψης	99
5.12	Ετήσιο σφάλμα πρόβλεψης με προσθήκη θορύβου	103
5.13	Ετήσιο σφάλμα ξεχωριστής πρόβλεψης με προσθήκη θορύβου	103
5.14	Συνολικό σφάλμα Persistence Model	107
5.15	Υπερπαραμέτροι LSTM	108
5.16	Συνολικό σφάλμα LSTM	109
5.17	Υπερπαραμέτροι WaveNet	110
5.18	Συνολικό σφάλμα WaveNet	111
5.19	Ασαφοποίηση φωτοβολταϊκής παραγωγής	112
5.20	Συνολικό σφάλμα	114
5.21	Σφάλματα ετήσιας πρόβλεψης	117
5.22	Σφάλματα ετήσιας πρόβλεψης, με επιπλέον χαρακτηριστικά	119

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η επεξεργασία ενεργειακών δεδομένων σε πανελλαδικό και τοπικό επίπεδο, με στόχο την δημιουργία προβλέψεων των μεγεθών αυτών με χρονικό ορίζοντα 24 ωρών στο μέλλον. Συγκεκριμένα, αρχικά θα αναλυθούν τα βασικά χαρακτηριστικά των χρονοσειρών αιολικής και φωτοβολταϊκής παραγωγής, για την κατανόηση των ιδιαιτεροτήτων που παρουσιάζει το κάθε μέγεθος. Ακόμα, θα δοθεί το μαθηματικό υπόβαθρο για την επεξεργασία χρονοσειρών, όπως και για την δημιουργία προβλέψεων, με την χρήση νευρωνικών δικτύων και βαθιάς μάθησης. Οι διαφορετικές τεχνικές μηχανικής μάθησης θα αναλυθούν εκτενώς παρουσιάζοντας τις αρχιτεκτονικές τους, αλλά και τους αλγορίθμους μάθησης με τους οποίους παράγουν τις προβλέψεις. Για την ορθή αξιοποίηση των μοντέλων αυτών θα δοκιμαστούν διάφορες παραλλαγές της αρχιτεκτονικής τους, όπως και των παραμέτρων εκπαίδευσης τους για την παραγωγή των βέλτιστων αποτελεσμάτων με γνώμονα τα χαρακτηριστικά του κάθε μεγέθους. Τέλος, θα γίνει προσπάθεια βελτιστοποίησης των παραγόμενων αποτελεσμάτων με την προσθήκη επιπλέον δεδομένων (καιρικών) στα μοντέλα πρόβλεψης και θα εξαχθούν συμπεράσματα σχετικά με την αποτελεσματικότητα των εξεταζόμενων μοντέλων, αλλά και προτάσεις για την πιθανή βελτίωση των μοντέλων πρόβλεψης.

1.2 Επισκόπηση της σύγχρονης βιβλιογραφίας

Η πρόβλεψη αιολικής παραγωγής είναι ένα δύσκολο εγχείρημα και χωρίζεται σε διαφορετικές κατηγορίες ανάλογα με τον χρονικό ορίζοντα πρόβλεψης και με την μέθοδο που επιλέγεται για την πραγματοποίησή της. Ανάλογα με τον ορίζοντα πρόβλεψης υπάρχουν προβλέψεις [62]:

- **πολύ μικρού ορίζοντα** προβλέψεις (διάρκειας μερικων δευτερολέπτων έως και 30 λεπτών), οι οποίες χρησιμεύουν κυρίως για ρυθμιστικές ενέργειες.
- **μικρού ορίζοντα** προβλέψεις (διάρκειας 30 λεπτών έως και μερικών ωρών), μέσω των οποίων γίνεται ο προγραμματισμός της οικονομικής φόρτισης των μονάδων και λαμβάνονται οι αποφάσεις μείωσης και αύξησης του φορτίου.

- **μεσαίου ορίζοντα** προβλέψεις (διάρκειας 6 ωρών έως και 1 ημέρας), με τις οποίες αποφασίζεται η ένταξη και η απένταξη των γεννητριών και εξασφαλίζεται η λειτουργική ασφάλεια στην αγορά ηλεκτρικής ενέργειας της επόμενης ημέρας.
- **μεγάλου ορίζοντα** προβλέψεις (διάρκειας 1 μέρας έως και 1 εβδομάδας), οι οποίες χρησιμεύουν στον προγραμματισμό ένταξης των μονάδων, στην εξασφάλιση των απαραίτητων εφεδρειών, όπως και στον προγραμματισμό της συντήρησης για την μείωση του κόστους λειτουργίας.

Για την δημιουργία των προβλέψεων στους διάφορους χρονικούς ορίζοντες, έχουν υιοθετηθεί διάφορες μεθοδολογίες με τις πιο σημαντικές να είναι η φυσική και η στατιστική προσέγγιση. Πιο συγκεκριμένα, οι πρώτες προσπαθούν να εκτιμήσουν την ροή ανέμου στην περιοχή του αιολικού πάρκου, χρησιμοποιώντας φυσικούς νόμους για την μοντελοποίηση της ατμοσφαιρικής συμπεριφοράς, ώστε να εκτιμηθεί η ταχύτητα ανέμου και στην συνέχεια μέσω της καμπύλης ισχύος η αιολική παραγωγή. Για να το επιτύχουν χρησιμοποιούν αριθμητικές προβλέψεις καιρού (Numerical Weather Prediction (NWP)) λύνοντας πολύπλοκα μαθηματικά μοντέλα δεδομένων καιρού. Αντιθέτως, στις στατιστικές μεθόδους μελετάται η σχέση μεταξύ των μετεωρολογικών παραμέτρων και της αντίστοιχης αιολικής παραγωγής με χρήση ιστορικών δεδομένων, ώστε στην συνέχεια να εκτιμηθεί η μελλοντική αιολική παραγωγή μέσω των στατιστικών μοντέλων χωρίς να συνυπολογίζονται τα φυσικά φαινόμενα. Χωρίζοντας τα στατιστικά μοντέλα πρόβλεψης βάσει των παραμέτρων εισόδου τους, πάλι έχουμε δυο κατηγορίες: αυτές που κάνουν χρήση αριθμητικών προγνώσεων καιρού (NWP), και αυτές που χρησιμοποιούν μόνο ιστορικά δεδομένα παραγωγής για την πραγματοποίηση των προβλέψεων. Ωστόσο, λόγω των έντονων μεταβολών που παρατηρούνται στην χρονοσειρά της αιολικής παραγωγής, η χρήση (NWP) θεωρείται απαραίτητη για ορίζοντες πρόβλεψης μεγαλύτερους των 6 ωρών [79]. Τα στατιστικά μοντέλα είναι αρκετά δημοφιλή, καθώς είναι εύκολα στην μοντελοποίηση, δεν χρησιμοποιούν κάποιο εκ των προτέρων ορισμένο μαθηματικό μοντέλο και βασίζονται κυρίως στην μοντελοποίηση μοτίβων. Η υλοποίηση των στατιστικών προσεγγίσεων μπορεί να πραγματοποιηθεί είτε με την χρήση στατιστικών μοντέλων, όπως των αυτοπαλινδρομικών μοντέλων κινητού μέσου όρου (ARMA) και των ολοκληρωμένων αυτοπαλινδρομικών μοντέλων κινητού μέσου όρου (εποχιακών η όχι) s-ARIMA, ARIMA, είτε μέσω νευρωνικών δικτύων τα οποία γνωρίζουν μεγάλη ανάπτυξη τα τελευταία χρόνια. Λόγω της σημασίας των προβλέψεων αιολικής παραγωγής, έχουν ακόμα εμφανιστεί καινούργιες τεχνικές, οι οποίες έχουν ως στόχο την βελτίωση των προβλέψεων, κάνοντας χρήση της ασαφούς λογικής (Fuzzy Logic) ή του μετασχηματισμού Wavelet. Έτσι, μπορούν να δημιουργηθούν υβριδικά μοντέλα, που χρησιμοποιούν συνδυασμό διαφορετικών προσεγγίσεων για την επίτευξη καλύτερων αποτελεσμάτων.

Ανάλογα με τον χρονικό ορίζοντα πρόβλεψης, λίγη έρευνα έχει πραγματοποιηθεί για την πρόβλεψη σε πολύ μικρούς χρονικούς ορίζοντες. Στο [77] χρησιμοποιούνται, η ανάλυση κύριων συνιστωσών (PCA), καιρικές προβλέψεις (NWP) και τα νευρωνικά δίκτυα LSTM για την πρόβλεψη δεδομένων αιολικής παραγωγής με διακριτότητα 5 λεπτών για 24 ώρες. Αντίστοιχα, στο [80] χρησιμοποιείται η βήτα κατανομή (Beta Distribution) σε συνδυασμό με τα νευρωνικά

δίκτυα LSTM και την τεχνική βελτιστοποίησης Particle Swarm Optimization (PSO), για την πρόβλεψη των διαστημάτων (prediction intervals) της αιολικής παραγωγής σε δεδομένα διακριτότητας 10 λεπτών.

Ο χρονικός ορίζοντας ωστόσο με την περισσότερη βιβλιογραφία, είναι αυτός των 30 λεπτών έως και μερικών ωρών. Στο [53] εξετάζονται τα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου, για την δημιουργία προβλέψεων χρονικού ορίζοντα 30 λεπτών, μιας και τριών ωρών, για ένα αιολικό πάρκο, όπως και για την ευρύτερη περιοχή. Συμπεραίνεται, πως η πρόβλεψη για το ένα αιολικό πάρκο έχει μεγαλύτερη διακύμανση σε σχέση με την αντίστοιχη για την ευρύτερη περιοχή, αλλά και ότι η ακρίβεια των προβλέψεων μειώνεται με την αύξηση του ορίζοντα πρόβλεψης. Στο [44] προβλέπεται η ταχύτητα ανέμου όπως και η αιολική παραγωγή για χρονικό ορίζοντα 1 ώρας με την χρήση τεχνητών νευρωνικών δικτύων, τα οποία εκπαιδεύονται μέσω των αλγορίθμων Particle Swarm Optimization (PSO) και Levenberg-Marquardt (LM). Η πρόβλεψη της ταχύτητας του ανέμου μελετάται στο [38], δημιουργώντας προβλέψεις πολλαπλών-βημάτων (multi-step) με βήμα 15 λεπτών. Τα δεδομένα των ταχυτήτων ανέμου αναλύονται σε διαφορετικές συνιστώσες μέσω της τεχνικής SSA (Singular Spectrum Analysis) και η κύρια συνιστώσα προβλέπεται με ένα νευρωνικό δίκτυο που συνδυάζει CNN και GRU νευρωνικά δίκτυα, ενώ οι υπόλοιπες συνιστώσες προσδιορίζονται μέσω των μοντέλων SVR (Support Vector Regression). Για το αντίστοιχο μικρό διάστημα της μιας ώρας, η ταχύτητα του ανέμου προβλέπεται στο [37] με 3 διαφορετικά νευρωνικά δίκτυα, των RBF, ADALINE και των νευρωνικών δικτύων πρόσθιας τροφοδότησης εκπαιδευμένα με τον αλγόριθμο back propagation, δείχνοντας ότι κανένα από τα νευρωνικά δίκτυα δεν υπερτερεί των άλλων σε όλες τις μετρικές σφάλματος που ελέγχθηκαν. Στο [19] εξετάζεται η πρόβλεψη της αιολικής παραγωγής μέσω της βελτίωσης του νευρωνικού δικτύου LSTM. Πιο αναλυτικά, η χρονοσειρά αιολικής παραγωγής αναλύεται σε 3 συνιστώσες (2 κύριες συνιστώσες και 1 συνιστώσα θορύβου), μέσω της μεθόδου Variational Mode Decomposition (VMD) και στην συνέχεια το κελί του νευρωνικού δικτύου LSTM ρυθμίζεται, ώστε στο cell state η τρίτη συνιστώσα να φτάνει μειωμένη κατά έναν παράγοντα a , βελτιώνοντας την ακρίβεια των προβλέψεων, οι οποίες πραγματοποιούνται για χρονικούς ορίζοντες 1 έως και 4 ωρών στο μέλλον. Η μέθοδος αποσύνθεσης VMD, χρησιμοποιείται ακόμα και στο [66], σε συνδυασμό με την μέθοδο συσταδοποίησης K-Means αλλά και τα νευρωνικά δίκτυα LSTM για την δημιουργία προβλέψεων αιολικής παραγωγής για χρονικούς ορίζοντες 15 λεπτών μιας και δύο ωρών. Πιο σύγχρονες προσεγγίσεις της πρόβλεψης αιολικής παραγωγής [45], χρησιμοποιούν GRU νευρωνικά δίκτυα βελτιωμένα μέσω του μηχανισμού προσοχής για την επίτευξη αποτελεσμάτων μεγαλύτερης ακρίβειας. Χρησιμοποιούνται ιστορικά δεδομένα αιολικής παραγωγής, σε συνδυασμό με προβλέψεις καιρικών δεδομένων για την πρόβλεψη της αιολικής παραγωγής για 1-3 ώρες στο μέλλον, δείχνοντας ότι σφάλματα είναι μειωμένα συγκριτικά με άλλα νευρωνικά δίκτυα.

Εκτός όμως των ντετερμινιστικών προσεγγίσεων, πολλοί έχουν αναπτύξει πιθανοτικά μοντέλα για την πρόβλεψη της αιολικής παραγωγής. Μια από τις προσεγγίσεις παρουσιάζεται στο [72], όπου χρησιμοποιούνται νευρωνικά δίκτυα extreme learning machine σε συνδυασμό με την τεχνική Bootstrap, για την εύρεση των διαστημάτων πρόβλεψης (prediction intervals) για την πρόβλεψη της παραγωγής αιολικής ενέργειας, σε χρονικό ορίζοντα 1 ώρας και για δια-

φορετικά επίπεδα εμπιστοσύνης. Αντίστοιχα, στο [73] η χρονοσειρά της παραγωγής αιολικής ενέργειας αποσυντίθεται, μέσω του διακριτού μετασχηματισμού Wavelet και στην συνέχεια μέσω συνελικτικών δικτύων πραγματοποιείται η πρόβλεψη σημείου για ορίζοντα μιας ώρας. Χρησιμοποιώντας την τεχνική ensemble (συνδυασμός προβλέψεων), η αβεβαιότητα του μοντέλου αλλά και των δεδομένων μοντελοποιούνται μέσω Gaussian κατανομών, ώστε τελικά να δημιουργηθούν τα διαστήματα πρόβλεψης. Στο [82], χρησιμοποιείται και πάλι η μέθοδος συσταδοποίησης K-Means, όπως και τα νευρωνικά δίκτυα LSTM, για την εύρεση της πρόβλεψης σημείου για την επόμενη ώρα, όμως για την παραγωγή του πιθανοτικού μοντέλου χρησιμοποιείται το μη παραμετρικό μοντέλο nonparametric kernel density estimation (KDE).

Σχετικά με μεγαλύτερου χρονικού ορίζοντα προβλέψεις, στο [30] εξετάζεται η δημιουργία του στατιστικού μοντέλου f-Arima, οι παράμετροι του οποίου υπολογίζονται μέσω της μεθόδου exact maximum likelihood. Το f-Arima χρησιμοποιείται για την πρόβλεψη της ταχύτητας ανέμου για χρονικούς ορίζοντες 24 και 48 ωρών, ενώ επίσης δημιουργούνται και προβλέψεις αιολικής παραγωγής μέσω της αποτύπωσης της ταχύτητας ανέμου στην καμπύλη ισχύος του αιολικού πάρκου. Στο [61] εξετάζεται η πρόβλεψη αιολικής παραγωγής για τους χρονικούς ορίζοντες 1 και 24 ώρες στο μέλλον, χρησιμοποιώντας δεδομένα αιολικής παραγωγής και προβλέψεις ταχύτητας ανέμου διαθέσιμα από τον διαγωνισμό Global Energy Forecasting Competition (GEFCom2012). Για την πρόβλεψη, εξετάζεται η χρήση συνελικτικών νευρωνικών δικτύων με φίλτρα μιας και δύο διαστάσεων, με τα τελευταία να υπερτερούν των πρώτων και στους δυο χρονικούς ορίζοντες. Συνελικτικά νευρωνικά δίκτυα χρησιμοποιούνται και στο [24], τα οποία συνδυάζονται με RBF νευρωνικά δίκτυα για την δημιουργία προβλέψεων χρονικού ορίζοντα 24 ωρών με δεδομένα τις ιστορικές παρατηρήσεις της αιολικής παραγωγής. Στο [22], αξιοποιούνται τα TCN νευρωνικά δίκτυα για την πρόβλεψη μεγεθών από μετεωρολογικό σταθμό, συμπεριλαμβανομένου της ταχύτητας ανέμου για χρονικούς ορίζοντες 1 έως και 24 ωρών στο μέλλον. Η πρόβλεψη αιολικής παραγωγής προσεγγίζεται στο [81], μέσω της χρήσης νευρωνικών δικτύων πρόσθιας τροφοδότησης, τα οποία εκπαιδεύονται μέσω του αλγορίθμου Levenberg-Marquardt (LM), ενώ ως είσοδοι χρησιμοποιούνται αριθμητικές προβλέψεις καιρού, έπειτα από την εισαγωγή τους σε ένα φίλτρο Kalman για την παραγωγή βελτιωμένων δεδομένων και την πρόβλεψη της αιολικής παραγωγής για χρονικό ορίζοντα μιας ημέρας. Προσπάθεια για την πρόβλεψη της ταχύτητας ανέμου σε ακόμα μεγαλύτερους χρονικούς ορίζοντες από αυτόν της μιας ημέρας γίνεται στο [25]. Οι συγγραφείς, χρησιμοποιούν συνελικτικά νευρωνικά δίκτυα με είσοδο την ταχύτητα ανέμου για χρονικό διάστημα 7 (168 παρατηρήσεις) ημερών, ώστε να προβλεφθεί η τιμή της για τις επόμενες 3 μέρες (72 παρατηρήσεις).

1.3 Δομή της εργασίας

Στο πρώτο κεφάλαιο, γίνεται η εισαγωγή στο αντικείμενο της διπλωματικής εργασίας, ενώ παρουσιάζεται επίσης η επισκόπηση της σύγχρονης βιβλιογραφίας, σχετικά με την παραγωγή προβλέψεων αιολικής παραγωγής.

Στο δεύτερο κεφάλαιο, αναλύεται η σημασία των ανανεώσιμων πηγών ενέργειας για τα

σύγχρονα συστήματα παραγωγής ηλεκτρικής ενέργειας, ενώ εξηγείται η αξία της πρόβλεψης της αιολικής και της φωτοβολταϊκής παραγωγής σε συστήματα ηλεκτρικής ενέργειας με αυξανόμενη διείσδυση ανανεώσιμων πηγών ενέργειας.

Στο τρίτο κεφάλαιο, παρουσιάζεται το θεωρητικό υπόβαθρο για την δημιουργία προβλέψεων ενεργειακών μεγεθών. Αρχικά γίνεται μια εισαγωγή στις χρονοσειρές εξετάζοντας τα βασικά τους χαρακτηριστικά, ενώ στην συνέχεια γίνεται εκτεταμένη αναφορά στα τεχνητά νευρωνικά δίκτυα μέσω της θεωρητικής αλλά και μαθηματικής περιγραφής τους.

Το τέταρτο κεφάλαιο, σχετίζεται με την ανάλυση και την προεπεξεργασία των δεδομένων που θα χρησιμοποιηθούν, για την παραγωγή των προβλέψεων αιολικής και φωτοβολταϊκής παραγωγής.

Στο πέμπτο κεφάλαιο, εξετάζεται η δημιουργία προβλέψεων για την αιολική και φωτοβολταϊκή παραγωγή μέσω βαθιάς μηχανικής μάθησης. Ειδότερα, δημιουργούνται μοντέλα μηχανικής μάθησης βασισμένα στα αναδρομικά και τα συνελικτικά νευρωνικά δίκτυα, ώστε να εξακριβωθεί το πλέον κατάλληλο εξ αυτών για την πρόβλεψη του κάθε μεγέθους, συγκρίνοντάς τα βάσει των παραγόμενων σφαλμάτων.

Στο έκτο και τελευταίο κεφάλαιο, συνοψίζεται το βασικό πλαίσιο της διπλωματικής εργασίας, εξάγοντας συμπεράσματα για την καταλληλότητα των μοντέλων πρόβλεψης, ενώ αναφέρονται επίσης και πιθανές μελλοντικές επεκτάσεις του προβλήματος.

Κεφάλαιο 2

Ανανεώσιμες πηγές ενέργειας

2.1 Γενικά για τις ανανεώσιμες πηγές ενέργειας

Γεγονός αποτελεί, πως η κλιματική αλλαγή είναι ένα από τα σημαντικότερα προβλήματα που έχουν να αντιμετωπίσουν οι σύγχρονες κοινωνίες. Τελευταίες έρευνες, δείχνουν πως οι εκπομπές διοξειδίου του άνθρακα (CO_2) πρέπει να μειωθούν μέχρι το 2030 τουλάχιστον κατά 45 % από τα αντίστοιχα επίπεδα τους το 2010, ώστε να είναι εφικτός ο περιορισμός της αύξησης της θερμοκρασίας κατά $1.5^{\circ}C$ [40]. Ένας από τους κυριότερους παράγοντες για την επίτευξη του στόχου αυτού είναι η απεξάρτηση του τομέα παραγωγής ηλεκτρικής ενέργειας από τα ορυκτά καύσιμα. Στην Ευρωπαϊκή Ένωση (ΕΕ) η παραγωγή διοξειδίου του άνθρακα βιώνει μια σταθερή πτωτική πορεία από το 1990 και μετά, όμως περαιτέρω ρυθμίσεις κρίνονται αναγκαίες για την αντιμετώπιση του προβλήματος. Από το 1997 στην ΕΕ έχουν θεσπιστεί κατάλληλα πλαίσια για την άμβλυνση της κλιματικής αλλαγής, απαιτώντας από κάθε χώρα μέλος την ένταξη ανανεώσιμων πηγών ενέργειας για την εξυπηρέτηση των ηλεκτρικών της αναγκών. Με τον όρο ανανεώσιμες πηγές ενέργειας αναφερόμαστε σε πηγές ενέργειας που προέρχονται από φυσικές διαδικασίες. Πιο συγκεκριμένα, σύμφωνα με το Ευρωπαϊκό Κοινοβούλιο [47] ως ανανεώσιμες πηγές ενέργειας θεωρούνται η αιολική, η ηλιακή, η αεροθερμική, η γεωθερμική, η ενέργεια των ωκεανών, η υδροηλεκτρική, η ενέργεια από βιομάζα, η ενέργεια που εκλύεται από τα αέρια σε χώρους υγειονομικής ταφής, η ενέργεια από αέρια μονάδων επεξεργασίας λυμάτων και τέλος η ενέργεια από βιοαέρια.

Τα βιοκαύσιμα, όπως και η υδροηλεκτρική ενέργεια ήταν από παλιά οι βασικές ανανεώσιμες πηγές που συνέβαλαν στην παραγωγή ηλεκτρικής ενέργειας με ποσά 157.262 (GWh) και 371.167 (GWh) αντίστοιχα κατά το έτος 2015 στην Ευρώπη [40]. Παρόλα αυτά, τα τελευταία χρόνια παρατηρείται ραγδαία άνοδος της αιολικής και φωτοβολταϊκής ενέργειας. Γεγονός αποτελεί, πως η χρήση της αιολικής ενέργειας τετραπλασιάστηκε κατά τις χρονολογίες 2004-2015 από 58.942 (GWh) σε 301.876 (GWh), συνεισφέροντας πλέον κατά το ένα τρίτο της παραγόμενης ηλεκτρικής ενέργειας στην Ευρωπαϊκή Ένωση, ενώ και η φωτοβολταϊκή ενέργεια αντίστοιχα σημειώνει εξίσου ανοδική πορεία με 107.989 (GWh) κατά το έτος 2015 [16]. Σύμφωνα με τις αποφάσεις για την προώθηση των ανανεώσιμων πηγών ενέργειας, στην Ευρώπη ο πρώτος στόχος τέθηκε το 1997 με την απόφαση η κατανάλωση ενέργειας από ανανεώσιμες

πηγές ενέργειας κάθε κράτους μέλους να φτάνει το 12 (%) μέχρι το 2010 [31]. Επίσης, το 2001 ορίστηκε πως το 22.1 (%) της ηλεκτρικής παραγωγής πρέπει να καλύπτεται από ανανεώσιμες πηγές ενέργειας, ενώ το 2009 θεσπίστηκαν συγκεκριμένοι στόχοι για κάθε χώρα της Ευρωπαϊκής Ένωσης μέχρι το 2020, ζητώντας από αυτές να υποβάλλουν το σχέδιο δράσης τους για την ανάπτυξη και εισαγωγή επιπλέον ανανεώσιμων πηγών ενέργειας. Προσθετικά, αποφασίστηκε ότι η συνολική ενεργειακή κατανάλωση μέσω ανανεώσιμων πηγών ενέργειας μέχρι το 2020 θα έπρεπε να φτάνει στο 20 (%) [31], στόχος που επιτεύχθηκε μιας και το 2019 το συνολικό μερίδιο των ανανεώσιμων πηγών ενέργειας ήταν 19.7(%), βιώνοντας μια μεγάλη άνοδο σε σχέση με το μερίδιό τους το 2004 που κυμάνθηκε στο 9.6 (%). Η Ευρωπαϊκή Ένωση, ωστόσο, συνεχίζει να θέτει φιλόδοξους στόχους με τον επόμενο να αποτελεί την αύξηση του ποσοστού αυτού στο 32 (%) μέχρι το 2030, έχοντας σαν απώτερο σκοπό να γίνει η πρώτη ήπειρος χωρίς παραγωγές διοξειδίου του άνθρακα μέχρι το 2050.

2.2 Αξία πρόβλεψης αιολικής παραγωγής

Με την συνεχώς αυξανόμενη διείσδυση της αιολικής ενέργειας στα διασυνδεδεμένα συστήματα ηλεκτρικής ενέργειας οι χειριστές του συστήματος μεταφοράς (TSO) έρχονται αντιμέτωποι όλο και πιο συχνά με αυξανόμενα επίπεδα μεταβλητότητας και αβεβαιότητας. Αυτό μπορεί να εξηγηθεί αν κάποιος σκεφτεί πως η ισχύς που παράγεται από τα αιολικά πάρκα είναι συνάρτηση της ταχύτητας του ανέμου, κάτι που έχει ως αποτέλεσμα την μεταβολή της αιολικής παραγωγής σε ωριαία αλλά και ημερήσια βάση. Ακόμα, δεδομένου ότι η ταχύτητα του ανέμου αποτελεί συνάρτηση των καιρικών συνθηκών, η ταχύτητα του ανέμου που περιμένει ο χειριστής του δικτύου εξαρτάται άμεσα από την αβεβαιότητα των καιρικών προβλέψεων για την επόμενη ημέρα.

Σήμερα οι χειριστές του δικτύου χρησιμοποιούν προβλέψεις φορτίου για την επόμενη ημέρα, ώστε να υπολογίσουν το ποσό της ενέργειας που απαιτείται για την κάθε ώρα της. Αυτή η πρόβλεψη, επιτρέπει την έγκαιρη διαθεσιμότητα των πηγών παραγωγής ηλεκτρικής ενέργειας, κάποιες εκ των οποίων χρειάζεται να ειδοποιηθούν αρκετές ώρες νωρίτερα, για να μπορούν να παράξουν την απαιτούμενη ηλεκτρική ενέργεια. Τα ηλεκτρικά συστήματα με μεγάλη διείσδυση αιολικής παραγωγής χρησιμοποιούν προβλέψεις ταχύτητας ανέμου για την επόμενη ημέρα, για να υπολογίσουν πόση αιολική ενέργεια θα είναι είναι διαθέσιμη για την κάθε ώρα της. Συνδυάζοντας έτσι τις προβλέψεις της αιολικής παραγωγής και του φορτίου, οι χειριστές έχουν την δυνατότητα να αποφασίσουν τον βέλτιστο συνδυασμό πηγών ενέργειας για να εξασφαλίσουν την οικονομική και ασφαλή εξυπηρέτηση του φορτίου της [36].

Παρόλα αυτά οι προβλέψεις δεν είναι τέλειες. Η πρόβλεψη του φορτίου είναι πλέον μια ώριμη επιστήμη, αφού οι χειριστές των ηλεκτρικών δικτύων χρησιμοποιούν τέτοιες προβλέψεις με χρονικό ορίζοντα μιας μέρας για δεκαετίες, με αποτέλεσμα πλέον τα σφάλματά τους να κυμαίνονται σε πολύ χαμηλά ποσοστά. Αντιθέτως, η πρόβλεψη της αιολικής παραγωγής εμφανίζει πολύ μεγαλύτερα σφάλματα που κυμαίνονται σε ποσοστά της τάξης του 15-20 (%) για την περίπτωση της μελέτης ενός αιολικού πάρκου. Λόγω των υψηλών αυτών σφαλμάτων και της αβεβαιότητας της αιολικής παραγωγής, πολλοί μελετούν τις επιπτώσεις της ενσωμάτωσής

της στα συστήματα ηλεκτρικής ενέργειας. Οι οικονομικές αυτές επιπτώσεις, έχουν την μορφή διατήρησης μονάδων εφεδρείας για την αντιμετώπιση της αβεβαιότητας και της μεταβλητότητας του ανέμου, ώστε να διατηρηθεί το σύστημα αξιόπιστο. Σε πολλές από αυτές τις μελέτες, η αβεβαιότητα ή με άλλα λόγια τα μεγάλα σφάλματα στην πρόβλεψη της αιολικής παραγωγής διακατέχουν ένα σημαντικό ρόλο στην αύξηση του κόστους [9].

Ακόμα όμως και με ατελείς προβλέψεις, μελέτες διείσδυσης μεγάλης κλιμακας αιολικών παραγωγών στα συστήματα παραγωγής ηλεκτρικής ενέργειας δείχνουν ότι η χρήση των προβλέψεων (με χρονικό ορίζοντα 24 ωρών) για την κατανομή των μονάδων μπορεί να βελτιώσει σημαντικά τα κόστη λειτουργίας και την αχρησιμοποίητη ενέργεια, διατηρώντας παράλληλα υψηλά επίπεδα αξιοπιστίας του συστήματος. Μπορούμε έτσι να συμπεράνουμε, ότι η βελτίωση των μεθόδων πρόβλεψης της αιολικής παραγωγής θα συνδράμει κατά μεγάλο βαθμό στην οικονομικότερη λειτουργία των συστημάτων παραγωγής ηλεκτρικής ενέργειας.

Όταν οι προβλέψεις της αιολικής παραγωγής είναι χαμηλότερες από την πραγματική παραγωγή των αιολικών πάρκων χρησιμοποιούνται στο ηλεκτρικό σύστημα περισσότερες συμβατικές μονάδες από όσες τελικά χρειάζονται. Αυτό σημαίνει, ότι οι συμβατικές μονάδες θα πρέπει να λειτουργήσουν σε χαμηλότερη ισχύ, κάτι το οποίο δεν είναι αποδοτικό για την συνολική λειτουργία του συστήματος. Ακόμα, αν τα σφάλματα των προβλέψεων είναι αρκετά σημαντικά, ίσως χρειαστεί ορισμένες από τις περιττές αιολικές μονάδες να τεθούν εκτός λειτουργίας, αφήνοντας τα παραγόμενα ποσά ενέργειας αναξιοποίητα. Στην αντίθετη περίπτωση, όπου οι προβλέψεις αιολικής παραγωγής είναι υψηλότερες από την πραγματική παραγωγή των αιολικών πάρκων, χρησιμοποιούνται λιγότερες μονάδες στην κατανομή από όσες πραγματικά χρειάζονται στην μέρα της λειτουργίας. Με την πρόσθεση όμως συμβατικών μονάδων, οι οποίες έχουν την δυνατότητα να συμβάλλουν άμεσα στην παραγωγή διευθετείται το πρόβλημα έλλειψης ηλεκτρικής ενέργειας, ωστόσο αυξάνεται δραματικά το κόστος λειτουργίας του συστήματος.

2.3 Αξία πρόβλεψης φωτοβολταϊκής παραγωγής

Σε αντιστοιχία με την αιολική παραγωγή, λόγω της μείωσης του κόστους στην παραγωγή φωτοβολταϊκών panel, τα τελευταία χρόνια παρατηρείται μεγάλη αύξηση στις επενδύσεις φωτοβολταϊκών εγκαταστάσεων. Η αυξανόμενη διείσδυση τους στο σύστημα ηλεκτρικής ενέργειας έχει επιφέρει προκλήσεις στους χειριστές του συστήματος μεταφοράς, εξαιτίας της μεταβλητότητας και της αβεβαιότητας της παραγωγής [42]. Εκτός αυτού, το μηδενικό οριακό κόστος της ηλιακής ενέργειας επηρεάζει το ηλεκτρικό δίκτυο όσον αφορά την αγορά ενέργειας. Λόγω των παραπάνω, υπάρχει ένα αυξανόμενο ενδιαφέρον για την πρόβλεψη της φωτοβολταϊκής παραγωγής ειδικά σε συστήματα ηλεκτρικής ενέργειας με μεγάλη διείσδυση φωτοβολταϊκών εγκαταστάσεων. Γεγονός αποτελεί, ότι για αυτό το λόγο, πολλές ρυθμιστικές αρχές χρησιμοποιούν εδώ και καιρό συστήματα πρόβλεψης φωτοβολταϊκής παραγωγής. Για παράδειγμα, η California Independent System Operator (CAISO) ξεκίνησε να εντάσσει προβλέψεις φωτοβολταϊκής παραγωγής το 2004 και οι Northwestern Energy, Xcel Energy το 2005 [52].

Η ενσωμάτωση της ηλιακής ενέργειας παραγόμενης από φωτοβολταϊκά έχει φέρει διάφορα προβλήματα που δεν ήταν αξιοσημείωτα, όταν η διείσδυσή της κυμαίνονταν σε χαμηλότερα επίπεδα. Παραδείγματος χάρη, ένα από αυτά είναι η καμπύλη σχήματος πάπιας (duck curve), δηλαδή η μεταβολή της καμπύλης της καθαρής ζήτησης ηλεκτρικής ισχύος, λόγω της μεγάλης φωτοβολταϊκής παραγωγής κατά τις μεσημεριανές ώρες. Η παραγωγή αυτή δημιουργεί προβλήματα, εξαιτίας των απότομων ρυθμών αύξησης ή μείωσης της ηλεκτρικής ισχύος από τις συμβατικές τεχνολογίες ηλεκτροπαραγωγής, αλλά και της δυσκολίας ελέγχου της υπερπαραγωγής ηλεκτρικής ισχύος, κάτι το οποίο μπορεί να οδηγήσει σε ανεπιθύμητες περικοπές της φωτοβολταϊκής παραγωγής αν δεν υπάρχει σωστός προγραμματισμός. Γίνεται, επομένως, κατανοητό ότι λόγω της υψηλής διείσδυσης της φωτοβολταϊκής παραγωγής οι συμβατικές μονάδες παραγωγής ηλεκτρικής ενέργειας αντιμετωπίζουν πρόβλημα στην διαχείριση των καθοδικών και ανοδικών τάσεων στην ζήτηση ηλεκτρικής ισχύος, αλλά και στο γεγονός ότι ενδεχομένως πρέπει να σταματήσουν να λειτουργούν κατά τις μεσημεριανές ώρες μιας και η φωτοβολταϊκή παραγωγή μπορεί να καλύψει της ανάγκες της ζήτησης ηλεκτρικής ενέργειας. Αν κάποιο από τα παραπάνω προβλήματα δεν μπορεί να διευθετηθεί, η φωτοβολταϊκή παραγωγή θα πρέπει να περικοπεί, διαφορετικά το σύστημα ηλεκτρικής ενέργειας θα βρεθεί σε κίνδυνο. Οι ακριβείς προβλέψεις της φωτοβολταϊκής παραγωγής, λοιπόν, μπορούν να συνδράμουν στην μείωση της αβεβαιότητας των προηγούμενων προκλήσεων, καθιστώντας εφικτή την κατανομή των μονάδων με αξιόπιστο τρόπο επιτρέποντας την σταθερότερη και πιο οικονομική λειτουργία του ηλεκτρικού συστήματος.

Ο προγραμματισμός της κατανομής των μονάδων διαφέρει ανάμεσα σε διαφορετικού τύπου ενεργειακές μονάδες. Οι διαφορετικού τύπου μονάδες ανταποκρίνονται σε διαφορετικούς χρόνους εκκίνησης και τερματισμού. Κάποιες μονάδες θα πρέπει να ενταχθούν από την προηγούμενη ημέρα (πυρηνικά, βιομάζα, λιγνίτης), κάποιες άλλες λίγες ώρες πριν (συνδυασμένου κύκλου αερίου, ατμοστρόβιλοι) ενώ άλλες την προηγούμενη ώρα ή και νωρίτερα (μηχανές εσωτερικής καύσης). Εάν στο σύστημα υπάρχει σημαντικός αριθμός φωτοβολταϊκών μονάδων (αντίστοιχα και αιολικών), η όσο το δυνατόν ακριβέστερη πρόβλεψη της παραγωγής τους μαζί με την πρόβλεψη του φορτίου θα βοηθήσει την αρτιότερη κατανομή των μονάδων αυτών.

Προσθετικά και οι αποφάσεις της οικονομικότερης φόρτισης των μονάδων (economic dispatch) επηρεάζονται άμεσα από τις προβλέψεις φωτοβολταϊκής παραγωγής μεσαίου και μικρού χρονικού ορίζοντα, ως αποτέλεσμα της κατανομής μονάδων με διαφορετικούς χρόνους εκκίνησης. Επιπλέον, προβλέψεις φωτοβολταϊκής παραγωγής μικρότερου χρονικού ορίζοντα μπορούν να επιτρέψουν την οικονομικότερη και πιο αξιόπιστη αποστολή ενέργειας, μειώνοντας την αβεβαιότητα πρόβλεψης παραγωγής ηλιακής ενέργειας (από φωτοβολταϊκά συστήματα), μιας και για την οικονομική αποστολή ενέργειας υπάρχουν ήδη οι περιορισμοί της αβεβαιότητας της ζήτησης ηλεκτρικής ενέργειας, της μεταφοράς της, της παραγωγής της και του φορτίου.

2.4 Αξία πρόβλεψης φορτίου

Το φορτίο που καταναλώνεται από τους πολίτες υπόκεινται σε αλλαγές λόγω των ανθρώπινων δραστηριοτήτων, ακολουθώντας ημερήσια, εβδομαδιαία και μηνιαία μοτίβα. Έτσι,

εμφανίζεται αυξημένο κατά την διάρκεια της ημέρας και των πρώτων απογευματινών ωρών, αφού οι βιομηχανικοί εξοπλισμοί είναι σε λειτουργία, όπως και οι οικιακές συσκευές, ενώ εμφανίζεται μειωμένο κατά τις βραδινές ώρες, όπου το μεγαλύτερο μέρος του πληθυσμού δεν είναι ξύπνιο. Η πρόβλεψη φορτίου αποτελεί ένα σημαντικό παράγοντα για τον σχεδιασμό του δικτύου διανομής της ηλεκτρικής ενέργειας [56]. Αυτό αληθεύει, αφού το προβλεπόμενο φορτίο, καθορίζει τις δυνατότητες του συστήματος διανομής που θα πρέπει να πληρούνται, μιας και αν δεν υπολογιστούν ορθά το σύστημα μπορεί να βρεθεί σε κίνδυνο. Γενικότερα, η πρόβλεψη φορτίου είναι καθοριστικής σημασίας για τον σχεδιασμό, την λειτουργία αλλά και τον έλεγχο των ηλεκτρικών συστημάτων, επιτρέποντας την οικονομική και αξιόπιστη λειτουργία τους. Όμως, η συνεχώς αυξανόμενη διείσδυση ανανεώσιμων πηγών ενέργεια στα ηλεκτρικά δίκτυα έχει επηρεάσει κατά πολύ την καταναλωτική συμπεριφορά των πολιτών όπως και την πρόβλεψη του φορτίου [33], κρίνοντάς την εξαιρετικά σημαντική για τον βέλτιστο συνδυασμό των μονάδων, την οικονομική αποστολή, την αναγνώριση των βέλτιστων τάσεων όπως και τις συναλλαγές του εμπορίου ενέργειας [1].

Σύμφωνα με τους διαφορετικούς χρονικούς ορίζοντες, μπορούμε να κατηγοριοποιήσουμε την πρόβλεψη του φορτίου σε μικρού, μεσαίου και μεγάλου χρονικού ορίζοντα. Μικρού χρονικού ορίζοντα θεωρείται η πρόβλεψη του φορτίου μιας ώρας έως και μίας εβδομάδας και χρησιμεύει για την ένταξη των μονάδων, την οικονομική αποστολή, τον προγραμματισμό μεταφοράς ενέργειας, όπως και τους ελέγχους σε πραγματικό χρόνο. Οι μεσαίου χρονικού ορίζοντα προβλέψεις κυμαίνονται από 1 έως 5 χρόνια και χρησιμοποιούνται για την αγορά καυσίμων και τον υπολογισμό ενεργειακών τιμολογίων, ενώ τέλος οι μεγάλου χρονικού ορίζοντα προβλέψεις κυμαίνονται σε διαστήματα μεγαλύτερων των 5 ετών και ενδιαφέρουν μηχανικούς και οικονομολόγους που σχεδιάζουν τις μελλοντικές επεκτάσεις των συστημάτων, δηλαδή τον τύπο και το μέγεθος των μονάδων παραγωγής και των γραμμών μεταφοράς.

Η μικρού χρονικού ορίζοντα πρόβλεψη της ζήτησης του φορτίου έχει μεγάλη οικονομική αξία. Είναι ένα πρόβλημα που πρέπει να αντιμετωπιστεί σε καθημερινή βάση και χρησιμοποιούνται εργαλεία πρόβλεψης που επιτρέπουν τις ενεργειακές εταιρείες να προετοιμαστούν για την μέγιστη ζήτηση και να διανέμουν με αποδοτικό τρόπο τους πόρους, ώστε να μην δημιουργηθούν επιπλοκές στο σύστημα ηλεκτρικής ενέργειας. Ακόμα, επιτρέπει στον χειριστή του συστήματος τον αποτελεσματικό προγραμματισμό των στρεφόμενων εφεδρειών. Εκτός όμως των οικονομικών λόγων, η πρόβλεψη του φορτίου είναι μεγάλης σημασίας για τον έλεγχο σε πραγματικό χρόνο και την διασφάλιση της ασφάλειας των συστημάτων διαχείρισης ενέργειας. Λόγω των παραπάνω, το ποσό των κερδών από τα λειτουργικά κόστη και η αξιοπιστία του συστήματος εξαρτώνται από την ακρίβεια της πρόβλεψης του φορτίου. Σε περίπτωση λοιπόν υποκεκτίμησης, το σύστημα δεν θα μπορεί να παρέχει υψηλής ποιότητας υπηρεσίες και μπορεί να προκύψουν διακοπές ρεύματος, ενώ σε περίπτωση υπερεκτίμησης θα υπάρχει μη αναγκαία διαθεσιμότητα μονάδων αυξάνοντας, το κόστος λειτουργίας του συστήματος.

Για την πρόβλεψη του φορτίου, πολλοί είναι οι παράγοντες που θα καθορίσουν την ακρίβεια της, όπως τα δεδομένα καιρού, ή τα διαφορετικά προφίλ των πελατών. Στις μεσαίου και μεγάλου χρονικού ορίζοντα προβλέψεις, για παράδειγμα, λαμβάνονται υπ όψιν ιστορικά δεδομένα φορτίου, καιρικά δεδομένα, ο αριθμός των πελατών στις διαφορετικές κατηγορίες,

όπως και χαρακτηριστικά της κάθε περιοχής, δηλαδή οικονομικοί παράγοντες και δημογραφικά στοιχεία [56]. Παρόλα αυτά, τρεις είναι εκείνοι οι παράγοντες που επηρεάζουν άμεσα το μοτίβο του φορτίου. Αυτοί είναι η εποχικότητα, οι ημερήσιοι και εβδομαδιαίοι κύκλοι-μοτίβα, όπως και οι ημέρες των διακοπών. Οι εποχικές διακυμάνσεις εξαρτώνται από τις κλιματικές αλλαγές (θερμοκρασία, μήκος της ημέρας), αλλά και τις ανθρώπινες δραστηριότητες (διακοπές, εργασία). Οι εβδομαδιαίες διακυμάνσεις τώρα, βασίζονται στην ημέρα της εβδομάδας μιας και το μοτίβο της κατανάλωσης ενέργειας ενέργειας μεταβάλλεται από το αν η ημέρα είναι καθημερινή ή σαββατοκύριακο. Η ύπαρξη λοιπόν των σαββατοκύριακων στα οποία θα μπορούσαμε να συμπεριλάβουμε και τις ημέρες των διακοπών ή αργιών, επηρεάζουν σημαντικά την κατανάλωση ενέργειας, η οποία εμφανίζεται αρκετά χαμηλότερη σε σχέση με μια καθημερινή ημέρα της εβδομάδας. Πρέπει ακόμα να σημειώσουμε, ότι πολλές υπηρεσίες μεταβάλλουν τις καταναλώσεις τους ανάλογα με τις καιρικές συνθήκες (χρήση κλιματιστικών). Έτσι, βλέπουμε πως οι καιρικές συνθήκες όπως η θερμοκρασία, η ταχύτητα του ανέμου και τα επίπεδα υγρασίας μπορούν να μεταβάλλουν τα μοτίβα του φορτίου. Τέλος, και ο τύπος των πελατών όπου αναφέρθηκε προηγουμένως, επηρεάζει την ενέργεια που καταναλώνεται αφού διαφορετικοί πελάτες (αστικοί, εμπορικοί, βιομηχανικοί) χρησιμοποιούν ενέργεια με διαφορετικά μοτίβα εμφανίζοντας όμως σημαντικές ομοιότητες μεταξύ τους, όταν ανήκουν στην ίδια κατηγορία.

Κεφάλαιο 3

Θεωρητικό υπόβαθρο

3.1 Εισαγωγή στις χρονοσειρές

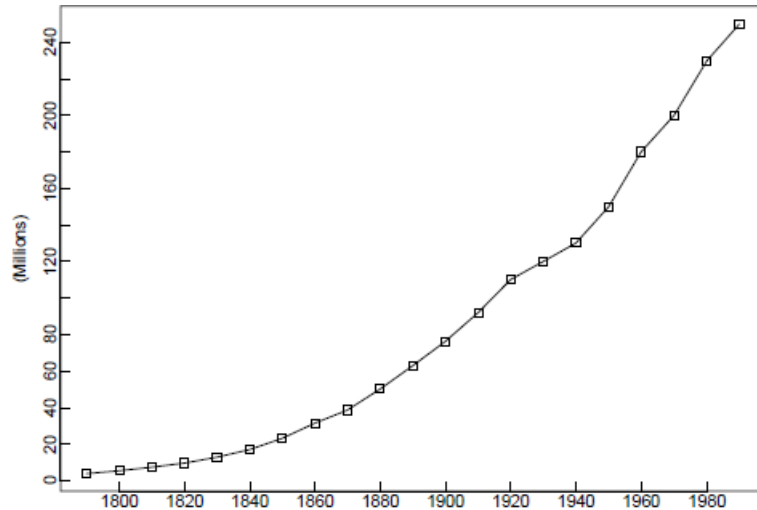
Με τον όρο χρονοσειρά ορίζουμε το σύνολο των διαδοχικών παρατηρήσεων Y_t των τιμών κάποιου φυσικού μεγέθους, οι οποίες έχουν καταγραφεί σε μια χρονική περίοδο. Οι παρατηρήσεις αυτές μπορεί να είναι για ένα ολόκληρο χρονικό διάστημα, ένα τυχαίο δείγμα του διαστήματος αυτού ή από συγκεκριμένα χρονικά σημεία. Οι καταγραφές μπορεί ακόμα να είναι συνεχείς (καταγραφή θερμοκρασίας) ή μια σειρά διακριτών παρατηρήσεων (τιμή κλεισίματος μετοχής). Στην παρούσα εργασία, θα ασχοληθούμε με διακριτές παρατηρήσεις, οι οποίες έγιναν σε σταθερά χρονικά διαστήματα μήκους μιας ώρας. Μαθηματικά μπορούμε να ορίσουμε ως χρονοσειρά την σειρά Y_t , που διατρέχει καθ' όλη τη διάρκεια του χρόνου, δηλαδή $(Y_t)_{t=0, \pm 1, \pm 2, \dots}$, αλλά έχει παρατηρηθεί μόνο στις χρονικές στιγμές $t = 1, \dots, n$ [13]. Έτσι λοιπόν παρατηρούμε τα (Y_1, \dots, Y_n) ή $(Y_t)_{t \in \mathbb{Z}}$, ($\mathbb{Z} = \{\dots, 0, 1, 2, \dots\}$). Μπορούμε να αναφερθούμε στον δείκτη t του Y_t , ως τον χρόνο και να θεωρήσουμε το Y_t , ως την κατάσταση ή έξοδος ενός στοχαστικού συστήματος στην χρονική στιγμή t . Η εξήγηση του δείκτη t ως τον χρόνο που συνέβει μια παρατήρηση δεν έχει μεγάλη σημασία για την μαθηματική επεξεργασία της χρονοσειράς, την οποία απασχολεί κυρίως η κοινή κατανομή των μεταβλητών, όμως η διάταξη αυτή των μεταβλητών στο χρόνο είναι συνήθως πολύ σημαντική [12].

3.1.1 Ποιοτικά χαρακτηριστικά χρονοσειρών

Η πραγματοποίηση της ορθής μελέτης μιας χρονοσειράς γίνεται μέσω της επισκόπησής της στο πεδίο του χρόνου. Για την κατανόηση των βασικών χαρακτηριστικών των χρονοσειρών, μπορούμε να αναλύσουμε την χρονοσειρά στα βασικά της στοιχεία τα οποία είναι : η τάση, η κυκλικότητα, η εποχικότητα, όπως και οι ασυνέχειες (οι οποίες αποτελούν υποσύνολο των μη κανονικών διακυμάνσεων) [69].

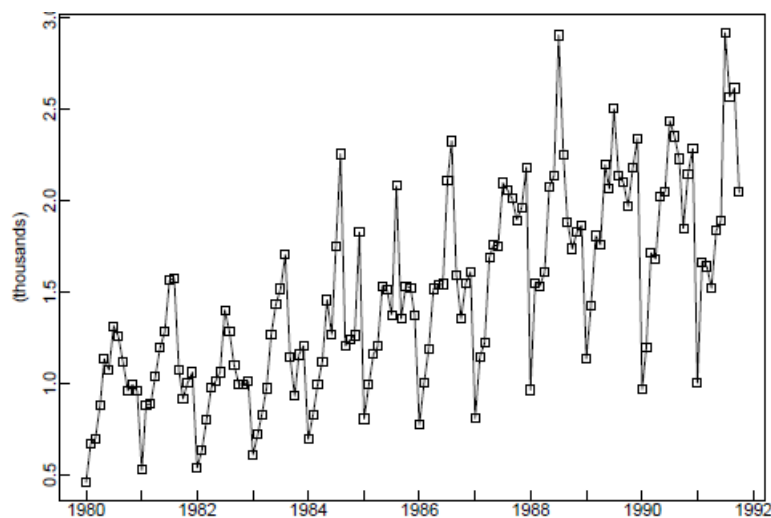
- **Η τάση** (trend), ορίζεται ως η μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών μιας χρονοσειράς. Δίνοντας ιδιαίτερη έμφαση στον όρο μακροπρόθεσμη μεταβολή, για την επιτυχή αναγνώριση της τάσης θα πρέπει να υπάρχει επαρκής αριθμός παρατηρήσεων, όπως και να έχει οριστεί κατάλληλα το μήκος της περιόδου μέσα στην οποία

θα αναζητηθεί η τάση. Η τάση μπορεί να είναι ανοδική, πτωτική ή μηδενική, ενώ σε κάθε περίπτωση αντιπροσωπεύει μια γενική απεικόνιση της χρονοσειράς. Στο σχήμα 3.1, παρουσιάζεται ο πληθυσμός των Η.Π.Α με διακριτότητα 10 ετών από το 1790 έως το 1990. Όπως μπορεί να διαπιστωθεί, η χρονοσειρά παρουσιάζει έντονη εκθετική τάση.



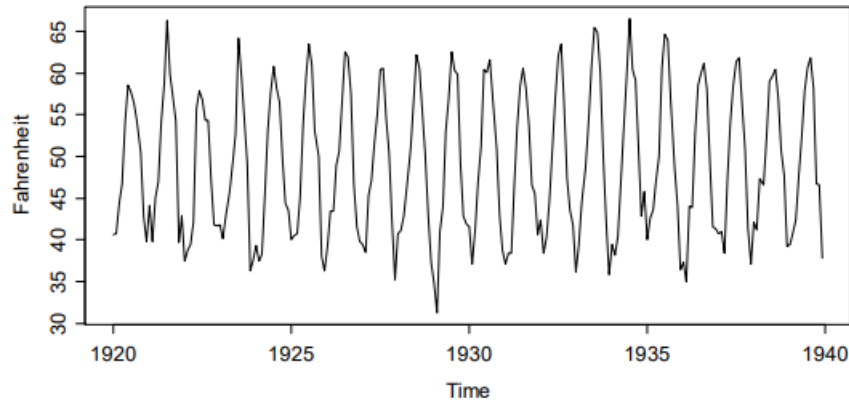
Σχήμα 3.1: Πληθυσμός Η.Π.Α (Πηγή: [50])

- **Η κυκλικότητα** (cyclic), θεωρείται ως μια κυματοειδής μεταβολή (ανοδική ή πτωτική), που εμφανίζεται σε περιόδους και οφείλεται σε εξωτερικούς παράγοντες. Αν και το μήκος της δεν είναι σταθερό τείνει να είναι μεγαλύτερο του ενός έτους. Χρονοσειρές που εμφανίζουν κυκλικότητα είναι παραδείγματος χάρη αυτές των οικονομικών μεγεθών, όπως του Ακαθάριστου Εθνικού Προϊόντος, η οποία οφείλεται σε μεταβολές των γενικότερων οικονομικών συνθηκών. Στο σχήμα 3.2, απεικονίζονται οι πωλήσεις κρασιού στην Αυστραλία την περίοδο 1980 με 1991 και μπορεί να παρατηρηθεί ένα κυκλικό μοτίβο που επαναλαμβάνεται ανά έτος.



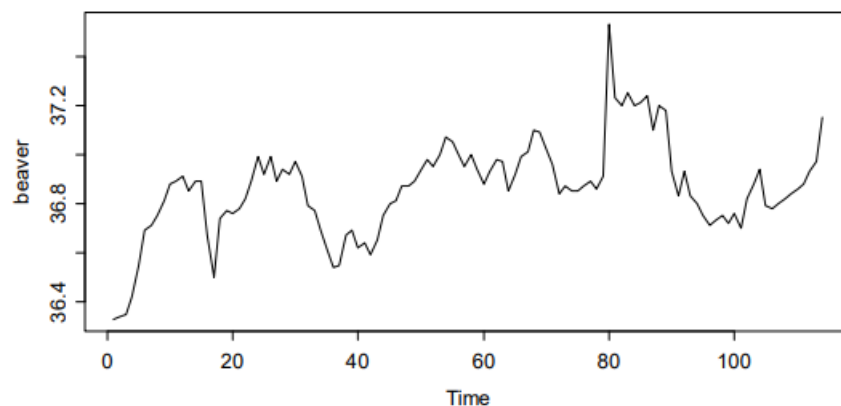
Σχήμα 3.2: Πωλήσεις κρασιού στην Αυστραλία (Πηγή: [50])

- **Η εποχικότητα** (seasonal), μπορεί να οριστεί ως μια περιοδική διακύμανση που έχει σταθερό μήκος μικρότερο του ενός έτους. Το σταθερό μήκος που την διακρίνει είναι και το χαρακτηριστικό που την διαφοροποιεί από την κυκλικότητα, αφού η διακυμάνσή της είναι σταθερή και μπορεί να εμφανίζεται με διακριτότητα ημέρας, εβδομάδας ή μηνιαία. Η συγκεκριμένη διακύμανση είναι άμεσα κατανοητή και προβλέψιμη, αφού επαναλαμβάνεται με το ίδιο μοτίβο σε σχέση με το χρόνο, γεγονός που την κάνει εύκολα εξηγήσιμη. Στο σχήμα 3.3, απεικονίζεται ο μέσος όρος της μηνιαίας θερμοκρασίας στην περιοχή του Nottingham της Αγγλίας, όπου γίνεται φανερό η έντονη εποχικότητα που χαρακτηρίζει την χρονοσειρά.



Σχήμα 3.3: Μέσος όρος μηνιαίας θερμοκρασίας στο Nottingham (Πηγή: [14])

- **Οι ασυνέχειες**, αντιπροσωπεύουν απότομες αλλαγές στο γράφημα κάποιας χρονοσειράς και μπορεί να έχουν παροδικό ή και μόνιμο χαρακτήρα. Οι παροδικές ασυνέχειες (outliers) έχουν μικρή χρονική διάρκεια και συχνά αποτελούν μια ασυνήθιστη παρατήρηση, η οποία οφείλεται σε κάποιο εξαιρετικό γεγονός. Αντιθέτως, οι ασυνέχειες που έχουν μόνιμο χαρακτήρα (level-shifts) έχουν την μορφή μιας απότομης αλλαγής στο μέσο επίπεδο των τιμών μιας χρονοσειράς. Στο σχήμα 3.4, φαίνεται η θερμοκρασία ενός κάστορα, όπου μπορεί να παρατηρηθεί κατά την ώρα 80 ότι η τιμή της θερμοκρασίας αντιπροσωπεύει μια ασυνέχεια σε σχέση με τις υπόλοιπες.



Σχήμα 3.4: Θερμοκρασία κάστορα (Πηγή: [14])

- **Οι μη κανονικές διακυμάνσεις**, είναι οι τιμές που υπολείπονται, έπειτα από την απομόνωση των υπόλοιπων χαρακτηριστικών μιας χρονοσειράς (τάση, κυκλικότητα, εποχιακότητα). Μπορεί να εκφράζουν μια τυχαία μεταβλητή, δηλαδή τον τυχαίο παράγοντα μιας στοχαστικής διαδικασίας, ή μια ασυνέχεια λόγω κάποιου εξαιρετικού γεγονότος.

Για να μπορέσουμε να εξάγουμε, τα παραπάνω χαρακτηριστικά τα οποία θα οδηγήσουν στην καλύτερη κατανόηση και μελέτη της χρονοσειράς χρησιμοποιούμε μαθηματικές σχέσεις για την απομόνωσή τους. Μαθηματικά, μια χρονοσειρά διατυπώνεται ως μια συνάρτηση των βασικών αυτών χαρακτηριστικών ως :

$$Y_t = f(S_t, T_t, C_t, R_t) \quad (3.1)$$

όπου:

Y_t = η παρατήρηση κατά την χρονική περίοδο t

S_t = η συνιστώσα της εποχικότητας κατά την χρονική περίοδο t

T_t = η συνιστώσα της τάσης κατά την χρονική περίοδο t

C_t = η συνιστώσα του κύκλου κατά την χρονική περίοδο t

R_t = η συνιστώσα της τυχαιότητας κατά την χρονική περίοδο t

$t = 0, 1, 2, \dots, n$

Οι πιο συνηθισμένες συναρτησιακές μορφές για την αντιστοίχιση των χαρακτηριστικών της χρονοσειράς σε μια χρονική περίοδο t , με την παρατήρηση στην αντίστοιχη περίοδο είναι η προσθετική, ώστε η συνάρτηση να παίρνει την μορφή :

$$Y_t = S_t + T_t + C_t + R_t \quad (3.2)$$

και η πολλαπλασιαστική :

$$Y_t = S_t \cdot T_t \cdot C_t \cdot R_t \quad (3.3)$$

3.1.2 Στατιστική ανάλυση χρονοσειρών

Μέσω της στατιστικής ανάλυσης των χρονοσειρών μπορούν να υπολογιστούν οι βασικοί στατιστικοί δείκτες σε μια δεδομένη χρονοσειρά Y με μέγεθος n παρατηρήσεων, κάτι το οποίο οδηγεί στην συνέχεια στην αποτελεσματική επιλογή μεθοδολογιών πρόβλεψης. Οι βασικοί στατιστικοί δείκτες είναι [69] [68]:

- **Μέση τιμή (Average)**, η οποία αποτελεί τον γραμμικό μέσο όρο των τιμών των παρατηρήσεων και δείχνει το επίπεδο στο οποίο κυμαίνονται οι πραγματικές τιμές της χρονοσειράς. Μαθηματικά υπολογίζεται από τον τύπο :

$$\mu = \frac{\sum_{i=1}^n Y_i}{n} \quad (3.4)$$

- **Τυπική απόκλιση (Standard Deviation)**, η οποία ονομάζεται και ως διασπορά μιας και είναι ένας δείκτης του βαθμού διασποράς των παρατηρήσεων γύρω από την μέση

τιμή. Μαθηματικά υπολογίζεται από τον τύπο :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (Y_i - \mu)^2}{n}} \quad (3.5)$$

- **Διακύμανση (Variance)**, η οποία εκφράζεται ως το τετράγωνο της τυπικής απόκλισης.
- **Συνδιακύμανση (Covariance)**, η οποία για δύο διακριτές τυχαίες μεταβλητές X και Y , δίνει ένα μέτρο συσχέτισης μεταξύ δύο περιοχών δεδομένων, με αποτέλεσμα να ορίζει το αν οι μεταβλητές μεταβάλλονται ανάλογα (θετική συνδιακύμανση), αντιστρόφως ανάλογα (αρνητική συνδιακύμανση), ή αν οι μεταβλητές είναι ασυσχέτιστες μεταξύ τους (μηδενική συνδιακύμανση). Μαθηματικά υπολογίζεται από τον τύπο :

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \mu_x) \cdot (Y_i - \mu_y)}{n} \quad (3.6)$$

- **Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient)**, ο οποίος αποτελεί ένδειξη της συσχέτισης μεταξύ των παρατηρήσεων της ίδιας μεταβλητής με χρονική υστέρηση k περιόδων. Το εύρος των τιμών του συντελεστή αυτοσυσχέτισης είναι μεταξύ του διαστήματος $[-1, 1]$. Έτσι, εάν η τιμή του είναι κοντά στο μηδέν, δηλώνει την μηδενική συσχέτιση των παρατηρήσεων με χρονική υστέρηση k περιόδων, αν η τιμή του είναι κοντά στο ένα δηλώνει θετική συσχέτιση μεταξύ των παρατηρήσεων αυτών, ενώ αν είναι κοντά στο μείον ένα, την αρνητική συσχέτιση μεταξύ των παρατηρήσεων. Η τιμή του συντελεστή αυτοσυσχέτισης αποδεικνύεται εξαιρετικά χρήσιμη στον προσδιορισμό της εποχιακής συμπεριφοράς μιας σειράς παρατηρήσεων. Μαθηματικά υπολογίζεται από τον τύπο :

$$\rho_k = \frac{\sum_{i=1+k}^n (Y_i - \mu_y) \cdot (Y_{i-k} - \mu_y)}{\sum_{i=1}^n (Y_i - \mu_y)^2} \quad (3.7)$$

- **Συντελεστής μερικής αυτοσυσχέτισης (Partial Autocorrelation Coefficient)**, ο οποίος αποτελεί ένδειξη της συσχέτισης της τιμής μιας χρονοσειράς με την τιμή της παρατήρησης k περιόδων στο παρελθόν, χωρίς όμως να λαμβάνεται υπόψη η επίδραση των ενδιάμεσων τιμών. Για τιμή του $k = 1$, ταυτίζεται με τον συντελεστή αυτοσυσχέτισης και παίρνει και αυτός τιμές στο διάστημα $[-1, 1]$. Υπολογίζεται από τις ακόλουθες

σχέσεις :

$$\varphi_{11} = \rho_1 \quad (3.8\alpha')$$

$$\varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (3.8\beta')$$

$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \varphi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} \rho_j} \text{ για } k = 3 \dots \quad (3.8\gamma')$$

$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk} \varphi_{k-1,k-j} \text{ για } k = 2 \dots j = 1, 2, \dots, k-1 \quad (3.8\delta')$$

3.1.3 Στασιμότητα (Stationarity)

Όπως αναφέρθηκε και στην εισαγωγή, μπορούμε να θεωρήσουμε ότι οι τιμές μιας χρονοσειράς αποτελούν τις τυχαίες μεταβλητές Y_1, \dots, Y_T που με την σειρά τους είναι μέρος μιας μεγαλύτερης στοχαστικής διαδικασίας $\{Y_t : t \in \mathbb{Z}\}$. Τότε, η μέση τιμή και η διακύμανση των τυχαίων μεταβλητών στην ανάλυση της χρονοσειράς είναι ανάλογες της συνάρτησης μέσης τιμής (mean function) και της συνάρτησης αυτόματης διακύμανσης (autocovariance function). Η συνάρτηση mean σε μια χρονοσειρά ορίζεται ως $\mu(t) = EY_t$, ενώ η συνάρτηση (autocovariance) ως $\gamma(s, t) = cov(Y_s, Y_t)$ [15]. Τώρα λοιπόν, μπορούμε να ορίσουμε ότι μια χρονοσειρά είναι αυστηρώς στάσιμη αν για κάθε $k > 0$ και σε κάθε $t_1, \dots, t_k \in \mathbb{Z}$ η κατανομή των

$$(Y_{t_1}, \dots, Y_{t_k})$$

είναι η ίδια με την

$$(Y_{t_1+u}, \dots, Y_{t_k+u})$$

για κάθε τιμή του u . Αν ισχύει το παραπάνω, σημαίνει πως η στοχαστική συμπεριφορά της διαδικασίας δεν αλλάζει στο χρόνο. Ακόμα αν ισχύει ότι η Y_t είναι στάσιμη ισχύει πως :

$$\mu(t) = \mu(0)$$

και

$$\gamma(s, t) = \gamma(s - t, 0)$$

Όπως παρατηρούμε λοιπόν, για τις στάσιμες χρονοσειρές η συνάρτηση μέσης τιμής είναι σταθερή και η συνάρτηση αυτόματης διακύμανσης εξαρτάται μόνο από την χρονική υστέρηση (time-lag) μεταξύ των δυο μεταβλητών για τις οποίες υπολογίστηκε. Εκτός όμως από την αυστηρή στασιμότητα υπάρχει και μια άλλη μορφή της, λιγότερο περιοριστική που ονομάζεται weak stationary. Μια χρονοσειρά λέμε ότι είναι weak stationary όταν $E|Y_t|^2 < \infty$, $\mu(t) = \mu$ και $\gamma(t+u, t) = \gamma(u, 0)$ για κάθε t και u . Όταν η χρονοσειρά είναι στάσιμη μπορούμε να απλουστεύσουμε τις παραμέτρους των συναρτήσεων mean και autocovariance.

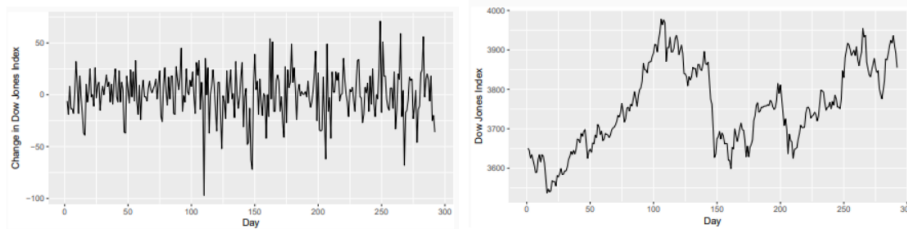
Πιο συγκεκριμένα μπορούν να οριστούν ως $\mu = E(Y_t)$ και $\gamma(u) = cov(Y_{t+u}, Y_t)$. Με αυτή την ορολογία μπορούμε ακόμα να ορίσουμε και μια επιπλέον συνάρτηση που ονομάζεται συνάρτηση αυτοσυσχέτισης (autocorrelation function) και έχει την μορφή:

$$\rho(u) = \frac{\gamma(u)}{\gamma(0)} = corr(Y_{t+u}, Y_t) \quad (3.9)$$

Η συνάρτηση αυτοσυσχέτισης είναι πολύ σημαντική για την επεξεργασία χρονοσειρών, αφού δείχνει το βαθμό συσχέτισης μεταξύ δυο χρονικών στιγμών της χρονοσειράς. Αν τώρα οι τυχαίες μεταβλητές που απαρτίζουν την χρονοσειρά $\{Y_t\}$ είναι ασυσχέτιστες και έχουν μηδενική μέση τιμή και διακύμανση σ^2 , η χρονοσειρά είναι στάσιμη και ονομάζεται λευκός θόρυβος με συνάρτηση αυτοσυσχέτισης :

$$\gamma(u) = \begin{cases} \sigma^2 & \text{εάν } u = 0 \\ 0 & \text{αλλιώς} \end{cases} \quad (3.10)$$

Η διαφορά μια στάσιμης χρονοσειράς από μια μη στάσιμη απεικονίζεται στο σχήμα 3.5



(α') Στάσιμη χρονοσειρά

(β') Μη στάσιμη χρονοσειρά

Σχήμα 3.5: Απεικόνιση στάσιμης και μη χρονοσειράς (Πηγή: [68])

Μέσω των παραπάνω, γίνεται κατανοητό πως για την ορθή μελέτη μιας χρονοσειράς θέλουμε να είναι στάσιμη, ώστε να διατηρεί τα χαρακτηριστικά της στο χρόνο και να μπορεί να μελετηθεί στοχαστικά. Παρόλα αυτά, η απαίτηση της στασιμότητας σπανίως ικανοποιείται εξ αρχής καθώς οι περισσότερες χρονοσειρές χαρακτηρίζονται από τάση, εποχικότητα και ασυνέχειες [68]. Μπορούμε όμως να ικανοποιήσουμε την απαίτηση αυτή, μέσα από κάποιες τεχνικές επεξεργασίας δεδομένων, με τις οποίες επιτυγχάνεται η εξομάλυνση της χρονοσειράς και ο περιορισμός της τυχαιότητας και των ακραίων τιμών. Η πιο απλή από αυτές τις μεθόδους είναι η **λογαρίθμιση**, η οποία έχει το μικρότερο υπολογιστικό κόστος, χωρίς όμως να οδηγεί σε πολλές περιπτώσεις στην μετατροπή μιας χρονοσειράς σε στάσιμη. Παρόμοιας φιλοσοφίας είναι και οι μετασχηματισμοί δυνάμεων, όπως οι μετασχηματισμοί **Box-Cox** οι οποίοι αντίστοιχα με την λογαρίθμιση εφαρμόζονται μόνο σε θετικά δεδομένα και δίνονται από τον τύπο :

$$Y^{(\lambda)} = \begin{cases} \frac{Y^{(\lambda)} - 1}{\lambda} & \text{εάν } \lambda \neq 0 \\ \ln(Y) & \text{εάν } \lambda = 0 \end{cases} \quad (3.11)$$

Σε αυτό το σημείο θα πρέπει να τονίσουμε επίσης, ότι η εφαρμογή μετασχηματισμών επηρεάζει την κλίμακα των δεδομένων, οπότε μετά την ολοκλήρωση της διαδικασίας πρόβλεψης, απαιτείται η εφαρμογή του αντίστροφου μετασχηματισμού για την επαναφορά της αρχικής κλίμακας

των δεδομένων. Ενώ οι προηγούμενοι μετασχηματισμοί περιορίζουν τις τυχαίες διακυμάνσεις, οι διακυμάνσεις που εμφανίζονται συστηματικά και οφείλονται στην ύπαρξη τάσης και εποχιακότητας μένουν ανεπηρεάστες. Για την αποτελεσματική εξάλειψη τους προτείνεται η χρήση της **διαφόρισης (differencing)**. Μέσω της εφαρμογής της για μια χρονοσειρά n παρατηρήσεων δημιουργείται μια νέα, με στοιχεία της, τις διαφορές των παρατηρήσεων της πρώτης [68]. Ανάλογα λοιπόν με την τάξη της διαφόρισης που θέλουμε να επιτύχουμε έχουμε :

$$1^{\text{η}} \text{ τάξη} : Y'_t = Y_t - Y_{t-1}$$

$$2^{\text{η}} \text{ τάξη} : Y''_t = Y'_t - Y'_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}, \text{ κ.ο.κ.}$$

Οι χρονοσειρές που λαμβάνουμε από την διαφόριση $1^{\text{η}}$ και $2^{\text{η}}$ τάξης έχουν $n - 1$ και $n - 2$ παρατηρήσεις αντίστοιχα. Μπορούμε να διαφορίσουμε την χρονοσειρά μέχρι $n - 1$ τάξεις, όμως στην πράξη χρησιμοποιείται η διαφόριση μέχρι $2^{\text{η}}$ τάξης. Ακόμα, μέσω της διαφόρισης μπορούμε να αναγνωρίσουμε τις χρονοσειρές του λευκού θορύβου (white noise) και του τυχαίου περιπάτου (random walk), που χρησιμοποιούνται συχνά για την αξιολόγηση πολυπλοκότερων μοντέλων [68] και περιγράφονται από τους τύπους :

$$\text{white noise} : Y_t = e_t$$

$$\text{random walk} : Y_t = Y_{t-1} + c + e_t$$

Σε αντιστοιχία με την απλή διαφόριση μπορούμε να εφαρμόσουμε και την εποχική διαφόριση, αν οι χρονοσειρές που μελετάμε εμφανίζουν έντονη εποχικότητα. Σε αυτή τη περίπτωση, στις διαφορές των παρατηρήσεων λαμβάνουν μέρος οι προηγούμενες εποχικές περίοδοι (m) δηλαδή :

$$1^{\text{η}} \text{ τάξη} : Y'_t = Y_t - Y_{t-m}$$

$$2^{\text{η}} \text{ τάξη} : Y''_t = Y'_t - Y'_{t-m} = Y_t - 2Y_{t-m} + Y_{t-2m}, \text{ κ.ο.κ.}$$

Για την ευκολότερη περιγραφή της διαφόρισης, μπορούμε ακόμα να ορίσουμε ως B τον τελεστή ολίσθησης, άρα να ισχύει ότι $BY_t = Y_{t-1}$ και $B(BY_t) = B^2Y_t = Y_{t-2}$, ώστε να αναπαραστήσουμε την διαφόριση τάξης n ως $(1 - B)^n Y_t$ και την εποχική διαφόριση, με εποχική περίοδο μήκους m ως $(1 - B^m)^n Y_t$. Τέλος, θα πρέπει να αναφέρουμε ακόμα ότι η διαφόριση μειώνει σημαντικά την αυτοσυσχέτιση των χρονοσειρών, ενώ ταυτόχρονα αυξάνει την τυχαιότητα του μοντέλου, όποτε θα πρέπει να προσέχουμε την τάξη της διαφόρισης, αλλιώς η χρονοσειρά μας ενδέχεται να απαιτεί πολύ πολύπλοκα μοντέλα για την περιγραφή της. Έτσι λοιπόν, συνηθίζεται να μην χρησιμοποιείται η τεχνική της διαφόρισης για αυτοσυσχέτιση μικρότερη του 0.5.

Ενώ αναφερθήκαμε στην σημασία της στασιμότητας για μια χρονοσειρά, αλλά και τις μεθόδους μέσω των οποίων μπορεί μια χρονοσειρά να γίνει στάσιμη, δεν έχουμε αναφερθεί στους τρόπους με τους οποίους ελέγχουμε αν μια χρονοσειρά είναι στάσιμη. Αυτό μπορεί να διαπιστωθεί, είτε από την μελέτη των γραφικών παραστάσεων, είτε μέσω στατιστικών τεστ. Από την απεικόνιση της γραφικής παράστασης λοιπόν, μπορεί κανείς να διακρίνει αν τα χαρακτηριστικά της παραμένουν σταθερά στο χρόνο ή αν εμφανίζουν μεταβολές, όπως τάση ή εποχικότητα με αποτέλεσμα η χρονοσειρά να κριθεί ως μη στάσιμη. Επειδή όμως, η συγκεκριμένη μέθοδος μπορεί να οδηγήσει σε εσφαλμένα συμπεράσματα ένας πιο αποδοτικός τρόπος είναι αυτός της μελέτης της γραφικής παράστασης της αυτοσυσχέτισης. Παρατηρώντας

επομένως το διάγραμμα αυτοσυσχέτισης μιας στασιμης χρονοσειράς, αυτό θα φθίνει σχετικά γρήγορα ενώ μιας μη στασιμης θα μειώνεται αργά [26]. Ακόμα, για μη στάσιμα δεδομένα η τιμή του ρ_1 είναι συνήθως μεγάλη και θετική. Για να κατανοήσουμε αν οι τιμές των διαγραμμάτων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης είναι στατιστικά σημαντικές μπορούμε να τις συγκρίνουμε με αυτές ενός λευκού θορύβου, με τον ίδιο αριθμό παρατηρήσεων [68]. Έτσι, για διάστημα εμπιστοσύνης 95 % οι αυτοσυσχετίσεις ενός λευκού θορύβου $N(0, \frac{1}{n})$, θα λαμβάνουν τιμές στο διάστημα $[-1.96\sqrt{n}, 1.96\sqrt{n}]$. Αν λοιπόν, η χρονοσειρά μας έχει παρόμοια συμπεριφορά τότε δεν είναι στατιστικά διάφορη του λευκού θορύβου άρα είναι και στάσιμη, σε αντίθετη όμως περίπτωση η σειρά μας διαφέρει από τον λευκό θόρυβο και δεν είναι στάσιμη. Ωστόσο, σε αυτό το σημείο πρέπει να σημειώσουμε ότι αν δεν υπάρχει καμία αυτοσυσχέτιση εκτός του ορίου του 95 %, αυτό υποδηλώνει ότι καμία από αυτές δεν είναι στατιστικά σημαντική, δηλαδή κάθε μεταβολή γίνεται τυχαία κατά ένα ποσό το οποίο είναι ασυσχέτιστο με τις προηγούμενες παρατηρήσεις [26]. Οι πιο συνεπείς μέθοδοι για την αξιολόγηση του αν μια χρονοσειρά είναι στάσιμη ή όχι είναι οι στατιστικοί έλεγχοι, οι οποίοι κάνουν μια αρχική υπόθεση (null hypothesis), ότι η χρονοσειρά είναι (ή δεν είναι) στάσιμη και εφαρμόζουν μια σειρά από κριτήρια, ώστε να απορρίψουν ή να δεχτούν την υπόθεση αυτή. Συγκεκριμένα για να αποφανθούν υπολογίζεται ο δείκτης στατιστικής σημαντικότητας (p-value), όπως και ο στατιστικός δείκτης (t-Statistic). Οι πιο γνωστές είναι ο έλεγχος Dickey-Fuller (Dickey-Fuller test) (DF) [11], ο επαυξημένος έλεγχος Dickey-Fuller (Augmented Dickey-Fuller test) (ADF) [20] και ο έλεγχος Kwiatkowski-Phillips-Schmidt-Shin (KPSS) [34]. Για την περιγραφή λοιπόν του Dickey-Fuller test θέλουμε να εκτιμήσουμε την παρακάτω παλινδρόμηση και να την ελέγξουμε για $\rho = 1$:

$$y_t = \rho y_{t-1} + e_t$$

όπου το e_t είναι λευκός θόρυβος, άρα έχει μηδενική μέση τιμή και σταθερή διακύμανση. Η παραπάνω εξίσωση για $\rho = 1$ παριστάνει έναν τυχαίο περίπατο, αλλά επειδή δεν μπορεί να εφαρμοστεί το στατιστικό τεστ (t-test), μιας και η κατανομή της δεν είναι κανονική, αφαιρούμε από κάθε μέλος της τον όρο y_{t-1} , οπότε τελικά παίρνει την μορφή:

$$\Delta y_t = \delta y_{t-1} + e_t \quad (3.12)$$

Έτσι λοιπόν, η μηδενική υπόθεση θα έχει μορφή $H_0 : \delta = 0$, ενώ η εναλλακτική της $H_1 : \delta < 0$. Αν απορρίψουμε την H_0 , η χρονοσειρά μας δεν θα έχει μοναδιαία ρίζα (unit root), ενώ αν όχι η χρονοσειρά μας θα έχει μοναδιαία ρίζα οπότε δεν θα είναι στάσιμη. Μέσω λοιπόν του Dickey-Fuller test, βρίσκουμε την πιθανότητα (p-value) για να ισχύει το παραπάνω σενάριο, θεωρώντας όπως αναφέραμε ως αρχική υπόθεση, ότι η χρονοσειρά έχει μοναδιαία ρίζα και άρα δεν είναι στάσιμη. Εμπειρικά για τιμές της (p-value) μεγαλύτερες του 0.05 (5 %) μπορούμε να υποστηρίξουμε την αρχική υπόθεση. Ακόμα, μέσω του Dickey-Fuller test υπολογίζεται ο t-Statistic και συγκρίνεται με διαφορετικά επίπεδα σημαντικότητας και τις αντίστοιχες κρίσιμες τιμές τους, ώστε να οδηγηθούμε στο τελικό συμπέρασμα. Το παραπάνω τεστ ωστόσο ισχύει όταν το e_t έχει την μορφή λευκού θορύβου. Ενδέχεται όμως να μην είναι ανεξάρτητος αν υπάρχουν συσχετίσεις μεταξύ των εξαρτημένων μεταβλητών της παλινδρόμησης. Για αυτό

το λόγο χρησιμοποιείται το Augmented Dickey-Fuller test, που επαυξάνει το αρχικό τεστ προσθέτοντας j υστερήσεις της εξαρτημένης μεταβλητής :

$$\Delta y_t = \alpha + \lambda t + \delta y_{t-1} + \sum_{j=1}^m \gamma_j \Delta y_{t-1} + e_t \quad (3.13)$$

Τέλος ο έλεγχος KPSS έχει μηδενική υπόθεση ότι η χρονοσειρά είναι στάσιμη γύρω από μια τάση. Η εναλλακτική είναι πως η χρονοσειρά δεν είναι στάσιμη, λόγω της ύπαρξης μιας μοναδιαίας ρίζας. Στο συγκεκριμένο έλεγχο, η χρονοσειρά αναπαρίσταται ως άθροισμα τριών ξεχωριστών συνιστωσών, μια ντετερμινιστική τάση, έναν τυχαίο περίπατο και ένα στάσιμο σφάλμα :

$$y_t = \xi t + r_t + e_t \quad (3.14)$$

όπου το r_t είναι ο τυχαίος περίπατος και περιγράφεται από την εξίσωση:

$$r_t = r_{t-1} + u_t$$

3.2 Εισαγωγή στα νευρωνικά δίκτυα

3.2.1 Διαδικασίες μηχανικής μάθησης

Με τον όρο μηχανική μάθηση αναφερόμαστε στο πεδίο της επιστήμης των υπολογιστών, το οποίο ασχολείται με την κατασκευή αλγορίθμων που μαθαίνουν από δεδομένα και μπορούν να κάνουν προβλέψεις σχετικά με αυτά [74]. Όπως οι ανθρώπινοι οργανισμοί έχουν αναπτύξει διαφορετικούς τρόπους για να μαθαίνουν από το περιβάλλον τους, έτσι και στην μηχανική μάθηση αλγόριθμοι όπως αυτοί των νευρωνικών δικτύων (τα οποία θα αναλυθούν εκτενώς στην συνέχεια της εργασίας) χρησιμοποιούνται με διαφορετικούς τρόπους για την επίτευξη μάθησης. Οι βασικότερες κατηγορίες στις οποίες μπορεί να χωριστούν οι διαδικασίες μάθησης των νευρωνικών δικτύων είναι, η μάθηση με εκπαιδευτή (επιβλεπόμενη μάθηση) και η μάθηση χωρίς εκπαιδευτή, με την δεύτερη να χωρίζεται σε δυο επιμέρους κατηγορίες την μη επιβλεπόμενη μάθηση και την ενισχυτική μάθηση. Συνοψίζοντας τα χαρακτηριστικά της κάθε κατηγορίας μπορούμε να πούμε περιληπτικά [60]:

- **Επιβλεπόμενη μάθηση (Supervised learning).** Η μάθηση αυτού του είδους περιλαμβάνει έναν εκπαιδευτή που έχει γνώση του περιβάλλοντος, το οποίο αντιπροσωπεύεται μέσω παραδειγμάτων εισόδου-εξόδου. Το περιβάλλον ωστόσο είναι άγνωστο για το νευρωνικό δίκτυο. Κατά την εκπαίδευση, ο εκπαιδευτής δίνει στο νευρωνικό δίκτυο ως είσοδο ένα διάνυσμα εκπαίδευσης, ενώ λόγω της γνώσης του περιβάλλοντος δίνει στο νευρωνικό δίκτυο και την επιθυμητή απόκριση για το διάνυσμα εκπαίδευσης. Το νευρωνικό δίκτυο λοιπόν, προσαρμόζει τις παραμέτρους του βάσει της επιρροής του διανύσματος εκπαίδευσης και του σήματος σφάλματος που ορίζεται ως η διαφορά της πραγματικής με την επιθυμητή απόκριση του δικτύου. Με την επανάληψη της παραπάνω διαδικασίας στόχος είναι το νευρωνικό δίκτυο να μπορεί να προσομοιώσει την συμπεριφορά του εκπαιδευτή με αυξανόμενη ακρίβεια μειώνοντας το σήμα σφάλματος. Όταν

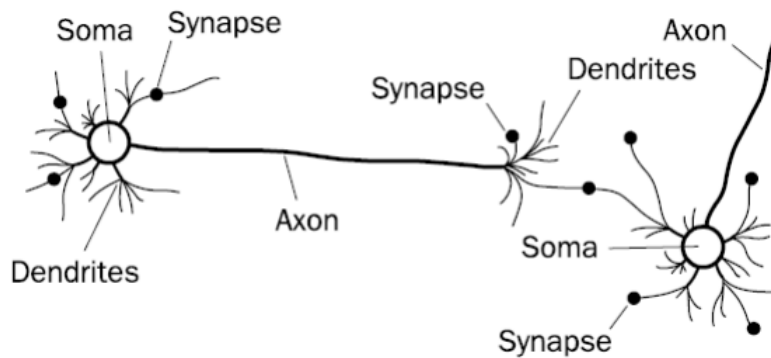
επιτευχθεί ο παραπάνω στόχος, η γνώση του περιβάλλοντος που προσφέρεται από τον εκπαιδευτή θα έχει μεταφερθεί στα συναπτικά βάρη του νευρωνικού δικτύου, οπότε και το δίκτυο θα μπορεί να απαλλαγεί από τον εκπαιδευτή και να προσομοιώσει μόνο του το περιβάλλον.

- **Μη επιβλεπόμενη μάθηση (Unsupervised learning)**. Σε αντίθεση με την επιβλεπόμενη μάθηση, στην μη επιβλεπόμενη δεν υπάρχει κάποιος εξωτερικός εκπαιδευτής ο οποίος δίνει γνώση του περιβάλλοντος στο νευρωνικό δίκτυο και επιβλέπει την εκπαίδευση. Αντιθέτως, ορίζεται ένα ανεξάρτητο μέτρο της ποιότητας της αναπαράστασης, που το νευρωνικό δίκτυο καλείται να μάθει με βάση το οποίο βελτιστοποιούνται οι ελεύθεροι παράμετροι του. Έπειτα από τον συντονισμό του ανεξάρτητου αυτού μέτρου στα δεδομένα εισόδου, το δίκτυο αποκτά την δυνατότητα σχηματισμού εσωτερικών αναπαραστάσεων για την κωδικοποίηση των χαρακτηριστικών εισόδου, δημιουργώντας αυτόματα νέες κλάσεις.
- **Ενισχυτική μάθηση (Reinforcement learning)**. Στόχος της ενισχυτικής μάθησης είναι η μάθηση μέσω συνεχούς αλληλεπίδρασης με το εξωτερικό περιβάλλον, μιας αντιστοιχίας εισόδου-εξόδου με σκοπό την ελαχιστοποίηση ενός βαθμωτού δείκτη απόδοσης. Η ενισχυτική μάθηση διαφοροποιείται από την επιβλεπόμενη και την μη επιβλεπόμενη μάθηση, μιας και ενώ υπάρχει μιας μορφής επίβλεψη αυτή δεν είναι η συσχέτιση κάθε επιθυμητής εξόδου με την αντίστοιχη είσοδο. Αντιθέτως, οι αλγόριθμοι ενισχυτικής μάθησης λαμβάνουν πληροφορίες από το περιβάλλον αφότου έχουν επιλέξει την έξοδο (η οποία ονομάζεται δράση) για τα δεδομένα μιας συγκεκριμένης εισόδου. Οι πληροφορίες του εξωτερικού περιβάλλοντος δείχνουν σε πιο βαθμό η έξοδος ικανοποιεί τις απαιτήσεις του περιβάλλοντος [59].

3.2.2 Βιολογικά νευρωνικά δίκτυα

Νευρώνας ή αλλιώς ένα νευρικό κύτταρο ονομάζεται το βιολογικό κύτταρο που έχει την δυνατότητα επεξεργασίας πληροφορίας. Αποτελείται από το κυτταρικό σώμα, τον άξονα και τους δενδρίτες. Το κυτταρικό σώμα με την σειρά του αποτελείται από τον πυρήνα, στον οποίο περιέχονται πληροφορίες σχετικά με τα κληρονομικά χαρακτηριστικά και το πλάσμα το οποίο συγκρατεί τον μοριακό εξοπλισμό για την παραγωγή των απαιτούμενων συστατικών που χρειάζεται ο νευρώνας [27]. Η βασική λειτουργία του νευρώνα είναι η λήψη σημάτων από άλλους νευρώνες μέσω των δενδριτών τους, αλλά και η μετάδοση σημάτων τα οποία παράγονται από το κυτταρικό του σώμα κατά μήκος του άξονα, ο οποίος διακλαδίζεται σε απολήξεις. Στο τέλος των απολήξεων υπάρχουν οι συνάψεις, οι οποίες είναι οι λειτουργικές μονάδες αλληλεπίδρασης μεταξύ δύο νευρώνων. Όταν ένας παλμός φτάσει στο τέλος της σύναψης απελευθερώνεται μια χημική ουσία που ονομάζεται νευροδιαβιβαστής. Οι νευροδιαβιβαστές απελευθερώνονται στο συναπτικό κενό και βελτιώνουν ή καταστέλλουν (ανάλογα με τον τύπο της σύναψης) την δυνατότητα του νευρώνα να εκπέμψει ηλεκτρικούς παλμούς [27]. Η αποτελεσματικότητα της σύναψης έχει την δυνατότητα να μεταβάλλεται ανάλογα με τα σήματα που την διαπερνούν ώστε να μαθαίνει από τις δραστηριότητες που λαμβάνει μέρος. Ενδιαφέρον είναι, ακόμα,

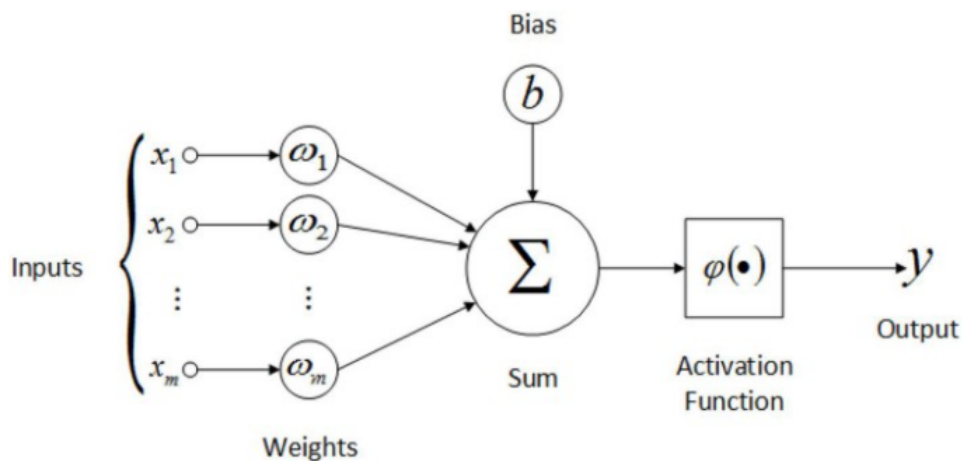
το γεγονός, ότι σύμφωνα με υπολογισμούς οι νευρώνες του εγκεφάλου είναι πέντε με έξι τάξεις μεγέθους πιο αργοί σε σχέση με τις λογικές πύλες (υλοποιημένες από πυρίτιο) που χρησιμοποιούνται στα σημερινά ολοκληρωμένα κυκλώματα. Το παραπάνω φαινόμενο ωστόσο αντισταθμίζεται λόγω του τεράστιου αριθμού νευρώνων στον ανθρώπινο εγκέφαλο, μιας και υπολογίζεται ότι στον ανθρώπινο εγκεφαλικό φλοιό υπάρχουν 10 δισεκατομμύρια νευρώνες οι οποίοι συνδέονται μεταξύ τους με 60 τρισεκατομμύρια συνάψεις [60]. Η αναπαράσταση ενός νευρώνα αποτυπώνεται σχήμα 3.6 :



Σχήμα 3.6: Βιολογικό νευρωνικό δίκτυο (Πηγή: [41])

3.2.3 Τεχνητά νευρωνικά δίκτυα

Σε αντιστοιχία με τους νευρώνες των βιολογικών νευρωνικών δικτύων μπορούν να δημιουργηθούν τεχνητοί νευρώνες, δηλαδή μονάδες επεξεργασίας πληροφορίας για τον σχηματισμό τεχνητών νευρωνικών δικτύων. Ένα τέτοιο μοντέλο μπορεί να απεικονιστεί όπως στο σχήμα 3.7 :



Σχήμα 3.7: Τεχνητό νευρωνικό δίκτυο (Πηγή: [28])

Για την ορθή περιγραφή ενός τεχνητού νευρωνικού δικτύου απαιτούνται τρία βασικά στοιχεία [60]:

1. Αρχικά ένα σύνολο συνάψεων (διασυνδέσεων), κάθε μια από τις οποίες χαρακτηρίζεται

από ένα δικό της ξεχωριστό βάρος. Παραδείγματος χάρη, ένα σήμα x_j στην είσοδο μιας σύναψης j που συνδέεται με το νευρώνα k θα πρέπει να πολλαπλασιαστεί με το συναπτικό βάρος w_{kj} . Μέσω αυτής της περιγραφής ο πρώτος δείκτης του βάρους w_{kj} αναφέρεται στον νευρώνα που θέλουμε να περιγράψουμε, ενώ ο δεύτερος στο άκρο εισόδου της σύναψης στην οποία αντιστοιχεί το βάρος. Μια χαρακτηριστική διαφορά μεταξύ των συναπτικών βαρών του βιολογικού και του τεχνητού νευρωνικού δικτύου είναι πως τα πρώτα μπορούν να πάρουν μόνο θετικές τιμές, σε αντίθεση με τα δεύτερα τα οποία έχουν την δυνατότητα να πάρουν και αρνητικές.

2. Έναν αθροιστή (adder), που χρησιμεύει στην άθροιση των σημάτων εισόδου, τα οποία έχουν πρώτα σταθμιστεί ανάλογα με τα συναπτικά βάρη του κάθε νευρώνα.
3. Μια συνάρτηση ενεργοποίησης (activation function), μέσω της οποίας περιορίζεται το πλάτος του σήματος ενός νευρώνα. Η συνάρτηση ενεργοποίησης ονομάζεται και συνάρτηση περιορισμού (squashing function), λόγω του ότι ενεργεί ως περιοριστής του επιτρεπτού εύρους πλάτους του σήματος εξόδου σε κάποια πεπερασμένη τιμή. Συνηθισμένα εύρη πλάτους τιμών της εξόδου ενός νευρώνα είναι διαστήματα όπως $[0, 1]$, $[-1, 1]$.

Τελευταίο χαρακτηριστικό του νευρωνικού δικτύου είναι η εξωτερικά εφαρμοζόμενη πόλωση b_k , που έχει ως στόχο την αύξηση ή μείωση της δικτυακής διέγερσης της συνάρτησης ενεργοποίησης ανάλογα με το αν είναι θετική ή αρνητική.

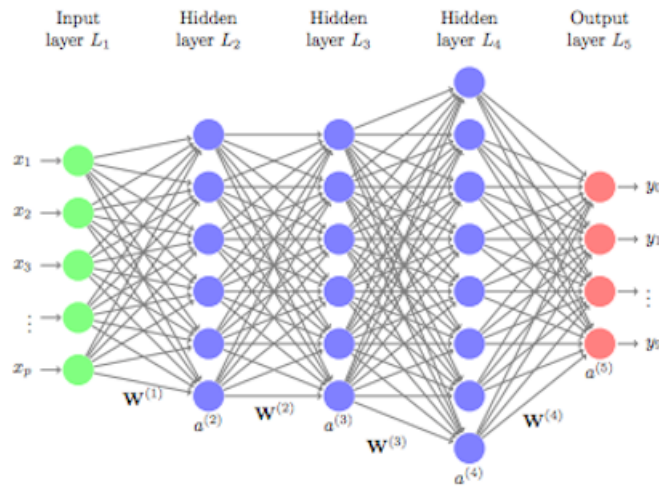
Μοντελοποιώντας μαθηματικά τα παραπάνω, ο νευρώνας k μπορεί να περιγραφεί ως :

$$u_k = \sum_{j=1}^m (w_{kj}x_j) \quad (3.15)$$

$$y_k = \phi(u_k + b_k) \quad (3.16)$$

όπου x_1, x_2, \dots, x_m και $w_{k1}, w_{k2}, \dots, w_{km}$ είναι τα σήματα εισόδου και τα αντίστοιχα συναπτικά βάρη του νευρώνα k . Παρατηρούμε επομένως, πως η έξοδος y_k του νευρωνικού δικτύου είναι το άθροισμα των γινομένων των σημάτων εισόδου επί τα αντίστοιχα βάρη τους και της πόλωσης, αφού πρώτα εισέλθουν από την συνάρτηση ενεργοποίησης.

Οι νευρώνες των νευρωνικών δικτύων ταξινομούνται σε επίπεδα, έτσι ο νευρώνας του σχήματος 3.7 μπορεί να θεωρεί ως ένα νευρωνικό δίκτυο ενός επιπέδου. Συνδυάζοντας πολλαπλά επίπεδα, καθένα από τα οποία μπορεί έχει πολλαπλό αριθμό νευρώνων μπορούμε να δημιουργήσουμε ένα πολυεπίπεδο νευρωνικό δίκτυο το οποίο ονομάζεται Multilayer Perceptron (MLP). Ένα τέτοιο δίκτυο απεικονίζεται στο σχήμα 3.8.



Σχήμα 3.8: Πολυεπίπεδο τεχνητό νευρωνικό δίκτυο (Πηγή: [71])

Ένα πολυεπίπεδο νευρωνικό δίκτυο, αποτελείται από τουλάχιστον τρία επίπεδα με το πρώτο να είναι το επίπεδο εισόδου (Input layer), στο οποίο εισάγουμε ένα διάνυσμα εκπαίδευσης, ένα ή περισσότερα κρυφά επίπεδα (Hidden layer), τα οποία συνδέουν το επίπεδο εισόδου με το επίπεδο εξόδου και ένα επίπεδο εξόδου (Output layer), από το οποίο λαμβάνεται η απόκριση του νευρωνικού δικτύου για το διάνυσμα εκπαίδευσης. Προσθετικά, κάθε νευρώνας με εξαίρεση αυτούς του επιπέδου εισόδου λαμβάνει και μια συνάρτηση ενεργοποίησης, όπως ορίστηκε και στο μοντέλο του απλού νευρώνα. Από το σχήμα 3.8 μπορεί να παρατηρηθεί ακόμα πως κάθε νευρώνας σε οποιοδήποτε επίπεδο συνδέεται με όλους τους άλλους του προηγούμενου επιπέδου, ώστε το νευρωνικό δίκτυο να χαρακτηρίζεται ως πλήρως συνδεδεμένο (fully connected), όπως και ότι η ροή του σήματος έχει κατεύθυνση από τα αριστερά προς τα δεξιά. Λόγω της τελευταίας αυτής ιδιότητάς του νευρωνικού δικτύου ονομάζεται και νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης (Feed-forward neural network).

3.2.4 Συναρτήσεις ενεργοποίησης

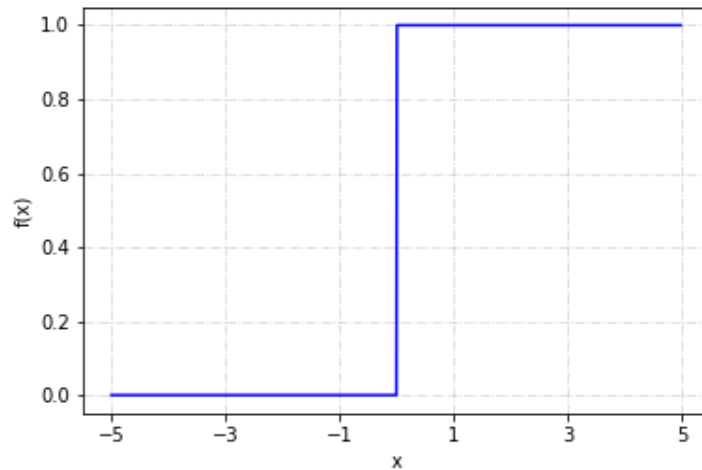
Όπως έχει αναφερθεί, ένας από τους σημαντικότερους παράγοντες κατά την υλοποίηση διαφόρων αρχιτεκτονικών νευρωνικών δικτύων είναι η επιλογή της συνάρτησης ενεργοποίησης. Αντίστοιχα με το μοντέλο του απλού νευρώνα και στα πολυεπίπεδα νευρωνικά δίκτυα αλλά και σε άλλες αρχιτεκτονικές νευρωνικών δικτύων (οι οποίες θα αναλυθούν εκτενώς στην συνέχεια), πρέπει να ορίσουμε την συνάρτηση ενεργοποίησης του κάθε επιπέδου (με εξαίρεση του επιπέδου εισόδου), η οποία θα καθορίζει το πλάτος εξόδου του σήματος από το κάθε επίπεδο του δικτύου. Οι συχνότερα χρησιμοποιούμενες συναρτήσεις ενεργοποιήσεως συνοψίζονται παρακάτω [57]:

Συνάρτηση βήματος (Binary Step Function)

Η συνάρτηση βήματος απεικονίζεται στο σχήμα 3.9 και αποτελεί την πιο απλή συνάρτηση ενεργοποίησης. Ουσιαστικά, είναι μια συνάρτηση κατωφλίου threshold function, αφού αν η

τιμή εισόδου της είναι πάνω από το κατώφλι που έχει οριστεί η έξοδος παίρνει την τιμή 1, ενεργοποιώντας το νευρώνα, προωθώντας το σήμα στο επόμενο επίπεδο, ενώ σε αντίθετη περίπτωση η έξοδος παίρνει την τιμή 0, αφήνοντας τον νευρώνα απενεργοποιημένο. Λόγω της δυνατότητας της να λαμβάνει μόνο δύο τιμές μπορεί να χρησιμοποιηθεί μόνο για δυαδική ταξινόμηση μη έχοντας την δυνατότητα να ταξινομή περισσότερες κατηγορίες. Τέλος, επειδή η κλίση (gradient) της συνάρτησης είναι 0 δημιουργεί προβλήματα κατά την εκπαίδευση μέσω του αλγορίθμου Back propagation που θα αναλυθεί στην συνέχεια. Μαθηματικά η συνάρτηση ορίζεται ως :

$$f(x) = \begin{cases} 1 & \text{εάν } x \geq 0 \\ 0 & \text{εάν } x < 0 \end{cases} \quad (3.17)$$



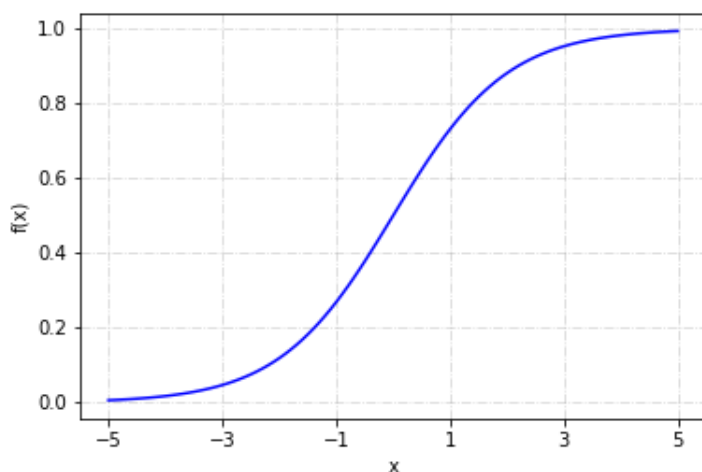
Σχήμα 3.9: Συνάρτηση βήματος

Σιγμοειδής συνάρτηση (Sigmoid Function)

Η σιγμοειδής συνάρτηση είναι από τις πιο γνωστές συναρτήσεις ενεργοποίησης, μιας και είναι μια μη γραμμική συνάρτηση. Μετατρέπει το σήμα εισόδου σε μία τιμή με εύρος από το 0 έως το 1 και μπορεί να οριστεί ως :

$$f(x) = \frac{e^x}{e^x + 1} \quad (3.18)$$

Η συγκεκριμένη συνάρτηση είναι μια συνεχής και διαφορίσιμη συνάρτηση, όμως δεν είναι συμμετρική γύρω από το μηδέν, κάτι το οποίο ωστόσο μπορεί να βελτιωθεί μέσω παραμετροποίησης της. Η σιγμοειδής συνάρτηση απεικονίζεται στο σχήμα 3.10.



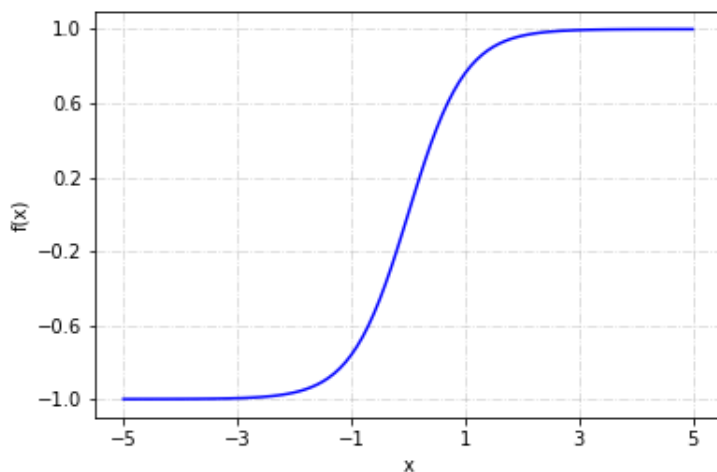
Σχήμα 3.10: Σιγμοειδής συνάρτηση

Συνάρτηση υπερβολικής εφαπτομένης (Tanh Function)

Η συνάρτηση υπερβολικής εφαπτομένης είναι παρόμοια με την σιγμοειδή συνάρτηση με διαφορά ότι είναι συμμετρική και έχει κέντρο της το μηδέν. Ορίζεται ως :

$$f(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (3.19)$$

Η συνάρτηση της υπερβολικής εφαπτομένης είναι συνεχής διαφορίσιμη και οι τιμές της κυμαίνονται από το -1 έως το 1. Προτιμάται σε σχέση με την σιγμοειδή συνάρτηση, μιας και η κλίση της (gradient) είναι πιο απότομη και χρησιμοποιείται για εισόδους με έντονες αρνητικές και θετικές τιμές. Η γραφική της παράσταση φαίνεται στο σχήμα 3.11.



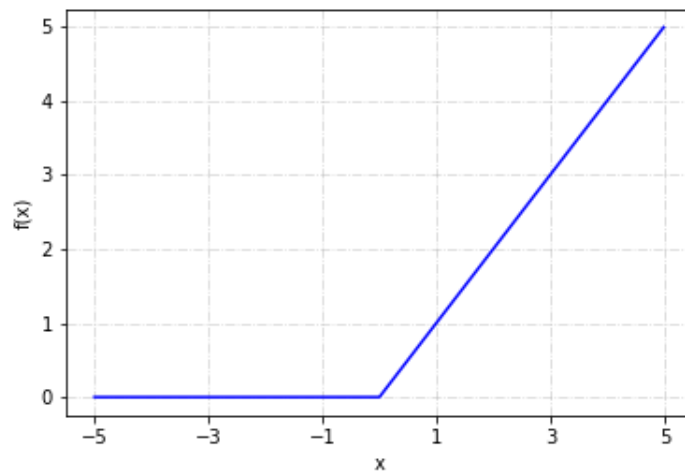
Σχήμα 3.11: Συνάρτηση υπερβολικής εφαπτομένης

Συνάρτηση ReLU (Rectifier Linear Unit)

Η συνάρτηση ReLU είναι μια από τις πιο συχνά χρησιμοποιούμενες, μη γραμμική συνάρτηση ενεργοποίησης. Ορίζεται ως :

$$f(x) = \max(x, 0) \quad (3.20)$$

Η ευρεία χρήση της οφείλεται στο ότι είναι πιο αποδοτική από τις υπόλοιπες συναρτήσεις, μιας και δεν ενεργοποιεί ταυτόχρονα όλους τους νευρώνες του δικτύου. Ακόμα, ένας νευρώνας απενεργοποιείται μόνο όταν η έξοδος από την συνάρτηση είναι μικρότερη από το 0. Αυτό δίνει την δυνατότητα στις τιμές της κλίσης της (gradient) να παίρνουν μεγαλύτερες τιμές, αφού όταν είναι μικρότερες από το μηδέν δεν ανανεώνονται, κάτι το οποίο βοηθάει στο να μη παρουσιάζει κορεσμό, σε αντίθεση με την σιγμοειδή συνάρτηση και τη συνάρτηση υπερβολικής εφαπτομένης. Η συνάρτηση ReLU απεικονίζεται στο σχήμα 3.12.



Σχήμα 3.12: Συνάρτηση ReLU

3.2.5 Συνάρτηση κόστους (Cost function)

Για την αξιολόγηση ενός αλγορίθμου μηχανικής μάθησης θα πρέπει να οριστεί ένα μέτρο της αποτελεσματικότητας του. Για την μέτρηση λοιπόν της επίδοσής του, χρησιμοποιούνται συναρτήσεις που πρέπει να βελτιστοποιήσουμε (να ελαχιστοποιήσουμε ή μεγιστοποιήσουμε), οι οποίες ονομάζονται αντικειμενικές συναρτήσεις (objective functions). Έτσι, η αντικειμενική συνάρτηση σε ένα πρόβλημα ενισχυτικής μάθησης θα είναι η μεγιστοποίηση της συνολικής ανταμοιβής (reward), ενώ σε ένα πρόβλημα παλινδρόμησης η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος. Επειδή η αντικειμενική συνάρτηση αποτελεί μια γενική έννοια, αναφερόμαστε πιο ειδικά σε αυτή ως συνάρτηση κόστους (cost function) ή συνάρτηση απωλειών (σφάλματος) (loss function (error function)). Η διαφορά μεταξύ των τελευταίων δύο είναι ότι μέσω της συνάρτησης απωλειών αναφερόμαστε σε σφάλμα ενός μόνο παραδείγματος εκπαίδευσης, ενώ μέσω της συνάρτησης κόστους σε όλα τα δεδομένα εκπαίδευσης. Μέσω της συνάρτησης σφάλματος επομένως μετράμε το πόσο καλά αποδίδει ο αλγόριθμός μας στα δεδομένα, αφού το σφάλμα μας βοηθάει να κατανοήσουμε πόσο διαφέρουν οι προβλεπόμενες

τιμές από τις κανονικές. Αν οι προβλέψεις μας διαφέρουν πολύ από την πραγματικότητα η συνάρτηση σφάλματος μεγαλώνει, ενώ σε αντίθετη περίπτωση διατηρεί μια μικρή τιμή. Παραμετροποιώντας το μοντέλο μας, η συνάρτηση απωλειών μας βοηθά να κατανοήσουμε αν ο αλγόριθμος μας βελτιώνεται ή όχι. Λόγω των διαφορετικών ειδών προβλημάτων οι συναρτήσεις σφάλματος διαφέρουν μεταξύ τους και χωρίζονται σε αυτές για προβλήματα παλινδρόμησης (regression), όπου προβλέπουμε μια συνεχή τιμή στην έξοδο, δυαδικής ταξινόμησης (Binary Classification), όπου η έξοδος μας μπορεί να πάρει δύο τιμές (0 και 1) και ταξινόμησης πολλών κλάσεων (Multi-Class Classification), όπου μπορούμε να έχουμε πολλές διαφορετικές κατηγορίες που θέλουμε να προβλέψουμε. Παίρνοντας ουσιαστικά το μέσο όρο όλων αυτών των συναρτήσεων, για κάθε δεδομένο εκπαιδευσης δημιουργούμε την συνάρτηση κόστους. Η συνάρτηση κόστους είναι το μέτρο της απόδοσης των αλγορίθμων μηχανικής μάθησης, μιας και υπολογίζει το σφάλμα όλων των προβλεπόμενων τιμών σε σχέση με τις πραγματικές και το ποσοτικοποιεί σε έναν πραγματικό αριθμό. Ανάλογα πάλι με το πρόβλημα που έχουμε να αντιμετωπίσουμε, η συνάρτηση κόστους μπορεί να χρίζει ελαχιστοποίησης ή βελτιστοποίησης. Μια από τις πιο συχνά χρησιμοποιούμενες συναρτήσεις κόστους για προβλήματα παλινδρόμησης είναι αυτή του μέσου τετραγωνικού σφάλματος (Mean Squared Error), (η οποία πρέπει να ελαχιστοποιηθεί) και μετράει όπως προδίδει και το όνομά της το τετραγωνικό σφάλμα της προβλεπόμενης τιμής $f(x_i|\theta)$ σε σχέση με την αναμενόμενη y_i για όλα τα διαθέσιμα δεδομένα (μεγέθους N) x_i και τις παραμέτρους τους θ . Έτσι υπολογίζεται ως :

$$MSE(\theta) = \frac{1}{N} \sum_{i=1}^N (f(x_i|\theta) - y_i)^2 \quad (3.21)$$

3.2.6 Αλγόριθμος οπισθοδιάδοσης σφάλματος (Backpropagation)

Ο αλγόριθμος Backpropagation ανακαλυφθηκε το 1986 [55] και είναι ο πιο συχνά χρησιμοποιούμενος αλγόριθμος για την εκπαίδευση νευρωνικών δικτύων πρόσθιας τροφοδότησης. Η λειτουργία του αλγορίθμου αυτού είναι ο υπολογισμός του σφάλματος πρόβλεψης του μοντέλου και η διάδοσή του (propagation), με ταυτόχρονη προσαρμογή των βαρών του νευρωνικού δικτύου, από το τελευταίο επίπεδο του δικτύου προς το πρώτο. Πιο συγκεκριμένα, μέσω του αλγορίθμου Backpropagation υπολογίζεται η κλίση (gradient) της συνάρτησης κόστους του νευρωνικού δικτύου, ως προς το κάθε βάρος του δικτύου κάνοντας χρήση του κανόνα της αλυσίδας (chain rule). Λόγω της μεγάλης του αποτελεσματικότητας, δίνει την δυνατότητα σε μεθόδους κλίσης (gradient methods) να χρησιμοποιηθούν για την εκπαίδευση πολυεπίδων νευρωνικών δικτύων μέσω της ανανέωσης των βαρών του δικτύου για την ελαχιστοποίηση του σφάλματος της συνάρτησης κόστους. Θεωρούμε λοιπόν ότι $y_j^{(l)}$ είναι η έξοδος του νευρώνα j στο επίπεδο l , m είναι ο αριθμός των νευρώνων στο επίπεδο $l + 1$, σ_l η συνάρτηση ενεργοποίησης του επιπέδου l και $z_j^{(l)} = \sum_{i=0}^K w_{j,i}^{(l)} \cdot y_i^{(l-1)}$ το άθροισμα των εισόδων του νευρώνα j στο επίπεδο l . Για την ανανέωση του κάθε βάρους του νευρωνικού δικτύου, θα πρέπει να υπολογίσουμε την κλίση της συνάρτησης σε σχέση με το βάρος όπου μέσω του αλγορίθμου

Backpropagation και κάνοντας χρήση του κανόνα της αλυσίδας ισούται με:

$$\frac{\partial J(\theta)}{\partial w_{j,k}^{(l)}} = \delta_j^{(l)} \cdot y_k^{(l-1)} \quad (3.22)$$

όπου αναλόγως αν το επίπεδο l είναι επίπεδο εξόδου ή κρυφό επίπεδο το $\delta_j^{(l)}$ υπολογίζεται ως :

$$\delta_j^{(l)} = \begin{cases} \frac{\partial J(\theta)}{\partial y_j^{(l)}} \cdot \sigma'_l(z_j^{(l)}) & \text{εάν το } l \text{ είναι επίπεδο εξόδου} \\ (\sum_{i=1}^m \delta_j^{(l+1)} \cdot w_{i,j}^{(l+1)}) \cdot \sigma'_l(z_j^{(l)}) & \text{εάν το } l \text{ είναι κρυφό εξόδου} \end{cases} \quad (3.23)$$

Όπως παρατηρούμε από τις εξισώσεις, πριν την εκκίνηση του αλγορίθμου Backpropagation πρέπει να προηγηθεί το πρόσθιο πέρασμα ώστε να υπολογιστούν οι τιμές των $z_j^{(l)}$ και της εξόδου $y_j^{(l)}$. Στην συνέχεια, υπολογίζονται οι κλίσεις μέσω του αλγορίθμου Backpropagation, ώστε έπειτα μέσω ενός αλγόριθμου βελτιστοποίησης να γίνει η ανανέωση των βαρών. Έτσι, αν ο αλγόριθμος αυτός είναι ο αλγόριθμος κατάβασης κλίσης που είναι ο πιο συχνά χρησιμοποιούμενος (ο οποίος θα αναλυθεί στην συνέχεια) η ανανέωση των βαρών θα έχει την μορφή :

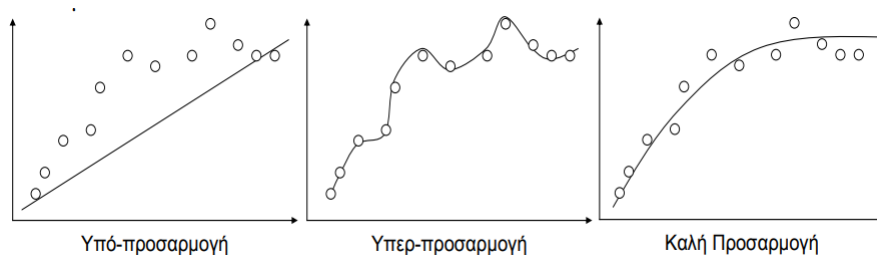
$$w_{j,k}^{(l)} = w_{j,k}^{(l)} - \eta \cdot \frac{\partial J(\theta)}{\partial w_{j,k}^{(l)}} \quad (3.24)$$

Αφού ολοκληρωθεί και η ανανέωση των βαρών, μπορεί να ξεκινήσει η επόμενη επανάληψη του αλγορίθμου μέχρι να ικανοποιηθεί ένα κριτήριο τερματισμού, το οποίο θέτουμε και μπορεί να είναι παραδείγματος χάρη ένας μέγιστος αριθμός επαναλήψεων του αλγορίθμου, ή η μείωση της συνάρτησης σφάλματος σε μια αποδεκτή τιμή.

3.2.7 Πρόληψη υπερπροσαρμογής

Κατά την σχεδίαση στατιστικών μοντέλων και μοντέλων μηχανικής μάθησης βασικό μας μέλημα αποτελεί η υλοποίηση μοντέλων που έχουν την δυνατότητα **γενίκευσης (Generalisation)**, δηλαδή έχουν την δυνατότητα να κάνουν ορθές προβλέψεις για δεδομένα που δεν έχουν δει στο παρελθόν. Τα μοντέλα μηχανικής μάθησης εκπαιδεύονται μέσω δεδομένων εκπαίδευσης και η απόδοσή τους ελέγχεται σε δεδομένα που δεν έχουν ξαναδεί, τα οποία αποτελούν τα δεδομένα ελέγχου. Ένα καλο μοντέλο μηχανικής μάθησης παρουσιάζει παρόμοια συμπεριφορά προβλέποντας με υψηλή ακρίβεια και τα δεδομένα εκπαίδευσης, αλλά και τα άγνωστα για αυτό δεδομένα ελέγχου. Σε αυτή την περίπτωση, θεωρούμε ότι το μοντέλο έχει την δυνατότητα γενίκευσης της πληροφορίας που έμαθε και μπορεί να την εφαρμόσει σε άγνωστα για αυτό δεδομένα. Παρόλα αυτά, η απόκριση των μοντέλων μηχανικής μάθησης δεν είναι πάντα η επιθυμητή. Μια από τις πιθανές καταστάσεις είναι η **υποπροσαρμογή (underfitting)** του μοντέλου μας. Αυτό σημαίνει, πως το μοντέλο μας παρουσιάζει χαμηλή ακρίβεια κατά την δημιουργία προβλέψεων και στα δεδομένα εκπαίδευσης αλλά και στα δεδομένα ελέγχου. Κάτι τέτοιο συμβαίνει όταν οι αλγόριθμοι που χρησιμοποιούνται για την δημιουργία του μοντέλου είναι απλοί, με αποτέλεσμα να μην μπορεί να μάθει σύνθετες αναπαραστάσεις. Ένα

παράδειγμα είναι η χρήση γραμμικών μοντέλων για την πρόβλεψη μη γραμμικών δεδομένων, αφού σε αυτή την περίπτωση το μοντέλο υπεραπλουστεύει τα δεδομένα κάνοντας λανθασμένες υποθέσεις για αυτά, αφού δεν μπορεί να διαχειριστεί τις μη γραμμικές σχέσεις μεταξύ των δεδομένων εισόδου και εξόδου, με αποτέλεσμα να παρουσιάζει μεγάλα σφάλματα. Πρέπει ωστόσο να επισημάνουμε σε αυτό το σημείο, ότι με την ανάπτυξη των αλγορίθμων μηχανικής μάθησης και λόγω της βαθιάς μάθησης μέσω νευρωνικών δικτύων πιο συχνό πρόβλημα αποτελεί η **υπερπροσαρμογή (overfitting)** των μοντέλων. Όταν ένα μοντέλο παρουσιάζει υπερπροσαρμογή σημαίνει πως εμφανίζει πολύ μεγάλη ακρίβεια στην πρόβλεψη των δεδομένων εισόδου, αλλά πολύ μικρότερη στα δεδομένα ελέγχου. Αυτό έχει ως αποτέλεσμα το μοντέλο να χάνει την δυνατότητα γενίκευσης, αφού δεν μπορεί να γενικεύσει τις πληροφορίες που έμαθε κατά την εκπαίδευση στα δεδομένα ελέγχου. Η υπερπροσαρμογή παρουσιάζεται όταν οι αλγόριθμοι που χρησιμοποιούνται είναι πολύ πολύπλοκοι και επηρεάζονται έντονα από τα δεδομένα εκπαίδευσης μαθαίνοντας λεπτομέρειες των μοτίβων που τους παρουσιάζονται, οι οποίες περιέχουν θόρυβο μη μπορώντας να τις γενικεύσουν. Έτσι λοιπόν, τα μοντέλα καταλήγουν να μαθαίνουν τον τυχαίο θόρυβο των δεδομένων εκπαίδευσης αντί να εστιάζουν στην ουσιώδη και γενικευμένη πληροφορία. Όσα περιγράψαμε μέχρι στιγμής μπορούν να παρατηρηθούν και οπτικά μέσω του σχήματος 3.13.



Σχήμα 3.13: Δυνατές προσαρμογές μοντέλων μηχανικής μάθησης (Πηγή: [70])

Λόγω της ραγδαίας αύξησης στην χρήση μοντέλων βαθιάς μάθησης, τα φαινόμενα υπερπροσαρμογής έχουν γίνει όλο και πιο συχνά. Έτσι λοιπόν, έχουν αναπτυχθεί στρατηγικές οι οποίες χρησιμοποιούνται για την μείωση του σφάλματος ελέγχου και την εξάλειψη των φαινομένων υπερπροσαρμογής, τις οποίες θα περιγράψουμε στην συνέχεια :

Ομαλοποίηση (Regularization)

Μέσω της ομαλοποίησης, προστίθεται ένας όρος ποινής στην συνάρτηση κόστους κατά την εκπαίδευση του νευρωνικού δικτύου όσο αυξάνεται η πολυπλοκότητα του μοντέλου, ώστε να εμποδίζεται η υπερπροσαρμογή και να ευνοείται η γενίκευση. Πιο συγκεκριμένα, δεν έχουν όλοι οι συντελεστές του μοντέλου την ίδια συνεισφορά, έτσι μέσω του όρου ποινής (ο οποίος επεμβαίνει σε μεγάλες τιμές των συντελεστών) μειώνουμε όσες από αυτές δεν έχουν μεγάλη συνεισφορά, ώστε να κρατήσουν υψηλές τιμές όσες συνεισφέρουν πραγματικά στην μάθηση. Υπάρχουν δυο διαφορετικές τεχνικές ομαλοποίησης η L_1 και η L_2 . Η ομαλοποίηση L_1

ονομάζεται Lasso Regression και μαθηματικά περιγράφεται από την εξίσωση:

$$J = \sum_{i=1}^n (y_i - \sum_{j=1}^m x_{ij}w_j)^2 + \lambda \sum_{j=1}^m (|w_j|) \quad (3.25)$$

Όπως παρατηρούμε λοιπόν, στην συνάρτηση κόστους προσθέτουμε την απόλυτη τιμή των βαρών του δικτύου, έτσι αυξάνοντας τις τιμές της παραμέτρου λ τα μικρά βάρη θα εκμηδενιστούν. Με αυτό το τρόπο, θα κρατήσουμε μόνο όσα χαρακτηριστικά είναι σημαντικά, δημιουργώντας πίνακες με αραιά χαρακτηριστικά, κάνοντας την μέθοδο αυτή χρήσιμη όταν έχουμε να επιλέξουμε ανάμεσα από πολλά χαρακτηριστικά.

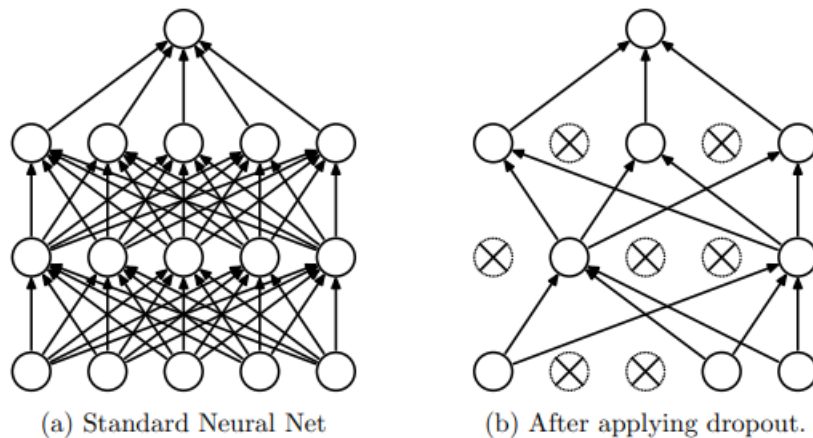
Η ομαλοποίηση L_2 ονομάζεται Ridge Regression και μαθηματικά περιγράφεται από την εξίσωση:

$$J = \sum_{i=1}^n (y_i - \sum_{j=1}^m x_{ij}w_j)^2 + \lambda \sum_{j=1}^m (w_j^2) \quad (3.26)$$

Σε αυτή την περίπτωση στην συνάρτηση κόστους προσθέτουμε το άθροισμα των τετραγώνων των συντελεστών του μοντέλου, κάτι το οποίο έχει παρόμοια λειτουργία με την ομαλοποίηση L_1 , ώστε σε υψηλές τιμές του λ να οδηγεί τους συντελεστές (ή κλίσεις) των μεταβλητών προς το μηδέν χωρίς όμως να τους αφήνει να πάρουν την τιμή αυτή.

Εγκατάλειψη (Dropout)

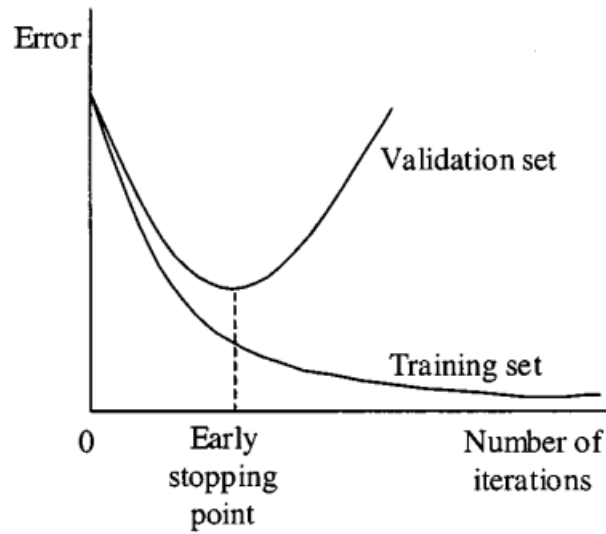
Η εγκατάλειψη (Dropout) [64], είναι μια από τις πιο συχνά χρησιμοποιούμενες τεχνικές ομαλοποίησης. Μέσω αυτής της τεχνικής, κατά την φάση της εκπαίδευσης επιλέγονται με τυχαίο τρόπο κάποιες μονάδες (νευρώνες) από το σύνολό τους, οι οποίες αγνοούνται. Έτσι, κατά τα το πρόσθιο και το οπίσθιο πέρασμα του νευρωνικού αυτές οι μονάδες δε λαμβάνονται υπ όψιν. Αναλυτικότερα, σε κάθε στάδιο εκπαίδευσης ένα ποσοστό p μεμονωμένων κόμβων επιλέγεται να διατηρηθεί κατά την εκπαίδευση, ενώ το υπόλοιπο $1-p$ αφαιρείται μειώνοντας έτσι τον αριθμό των παραμέτρων του δικτύου. Κατά αυτόν τον τρόπο κάθε επανάληψη αποτελείται από ένα διαφορετικό σετ κόμβων, γεγονός που βοηθάει στην μείωση της υπερπροσαρμογής. Η τυχαία εγκατάλειψη κόμβων φαίνεται στο σχήμα 3.14.



Σχήμα 3.14: Εφαρμογή επιπέδου εγκατάλειψης (Dropout) (Πηγή: [64])

Early Stopping

Κατά την εκπαίδευση των μοντέλων βαθιάς μάθησης, συνηθίζεται να παρατηρούμε τις καμπύλες που δημιουργούνται από την πορεία της συνάρτησης σφάλματος σε κάθε στάδιο της εκπαίδευσης. Για τον έλεγχο αυτό υπολογίζουμε το σφάλμα στα δεδομένα εκπαίδευσης, αλλά και στα δεδομένα επαλήθευσης (τα οποία διαφέρουν από αυτά του ελέγχου) που επιλέγονται ως ένα ποσοστό των δεδομένα εκπαίδευσης και δεν λαμβάνουν μέρος στην διαδικασία της εκπαίδευσης, αλλά χρησιμοποιούνται ως μια πιο αντικειμενική ένδειξη του σφάλματος έπειτα από κάθε στάδιο εκπαίδευσης. Ένα μοντέλο θεωρούμε ότι υπερπροσαρμόζεται στα δεδομένα εκπαίδευσης, όταν ενώ το σφάλμα σε αυτά μειώνεται σε κάθε στάδιο της εκπαίδευσης, το σφάλμα στα δεδομένα επαλήθευσης τα οποία είναι ουσιαστικά άγνωστα για το μοντέλο μειώνεται μέχρι ένα σημείο και στην συνέχεια αυξάνεται. Για την αποφυγή της υπερπροσαρμογής, η εκπαίδευση πρέπει να τερματίζεται όταν το σφάλμα στα δεδομένα επαλήθευσης αρχίζει να αυξάνεται, κάτι το οποίο επιτυγχάνεται με το Early Stopping, το οποίο αποθηκεύει τις βέλτιστες παραμέτρους με το μικρότερο σφάλμα επαλήθευσης. Το Early Stopping έχει ακόμα ως παράμετρο την *patience*, μέσω της οποίας ορίζουμε τον αριθμό των εποχών που θα περιμένει το Early Stopping χωρίς να βλέπει περαιτέρω βελτίωση στα δεδομένα επαλήθευσης, μιας και στην πραγματικότητα η καμπύλη αυτών δεν είναι ομαλή αλλά έχει διακυμάνσεις. Τα γραφήματα των σφαλμάτων αλλά και το βέλτιστο σημείο τερματισμού της εκπαίδευσης φαίνονται στο σχήμα 3.15:



Σχήμα 3.15: Βέλτιστο σημείο διακοπής μάθησης (Πηγή: [18])

3.2.8 Αλγόριθμοι βελτιστοποίησης

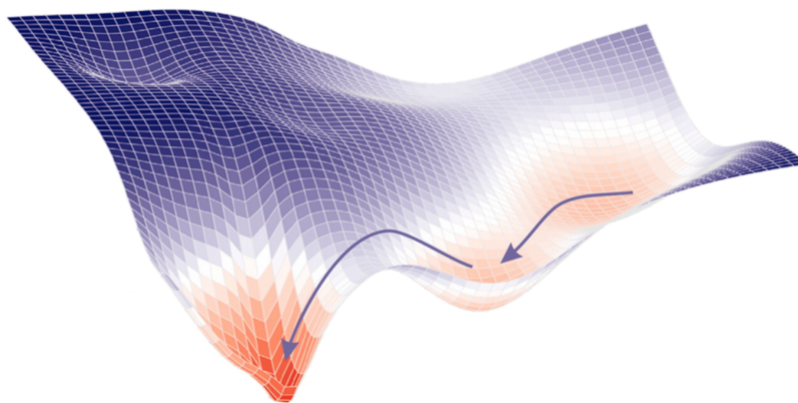
Οι αλγόριθμοι βελτιστοποίησης αποτελούν ένα σημαντικό κομμάτι κατά την δημιουργία συστημάτων μηχανικής μάθησης. Χρησιμοποιούνται για την αλλαγή των βαρών του νευρωνικού δικτύου και ρόλος τους είναι η ελαχιστοποίηση (μεγιστοποίηση) της συνάρτησης στόχου που έχει οριστεί για το σύστημα. Ένας από τους πιο συχνά χρησιμοποιούμενους αλγόριθμους βελτιστοποίησης είναι ο αλγόριθμος κατάβασης κλίσης, ο οποίος χρησιμοποιείται για την ελαχιστοποίηση της συνάρτησης σφάλματος. Παρακάτω θα περιγραφούν οι πιο βασικοί αλγόριθμοι βελτιστοποίησης :

Κατάβασης κλίσης (Gradient Descent)

Ο αλγόριθμος κατάβασης κλίσης χρησιμοποιείται για την ελαχιστοποίηση μιας συνάρτησης $J(\theta)$ όπου το $\theta \in \mathbb{R}^d$ αντιπροσωπεύει το διάνυσμα των παραμέτρων του δικτύου. Ανήκει στην οικογένεια των αλγορίθμων βελτιστοποίησης πρώτης τάξης (first-order optimization algorithms), δηλαδή κάνει χρήση μόνο των πρώτων παραγώγων για την ανανέωση των παραμέτρων. Ο τρόπος με τον οποίο ελαχιστοποιεί την συνάρτηση κόστους, είναι μέσω ανανέωσης των παραμέτρων προς την αντίθετη κατεύθυνση από αυτή της κλίσης (gradient) της συνάρτησης κόστους $\nabla_{\theta} J(\theta)$. Η ανανέωση των παραμέτρων μπορεί να περιγραφεί μέσω της εξίσωσης :

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta) \quad (3.27)$$

Στην παραπάνω εξίσωση το η δηλώνει το ρυθμό μάθησης (learning rate), δηλαδή το μέγεθος του βήματος που κάνει ο αλγόριθμος για να φτάσει στο ελάχιστο. Πιο συγκεκριμένα, ο αλγόριθμος κατάβασης κλίσης ακολουθεί την κατεύθυνση της καμπύλης που δημιουργείται από την αντικειμενική συνάρτηση με κατηγορικό τρόπο έως ότου φτάσει στο ελάχιστο.



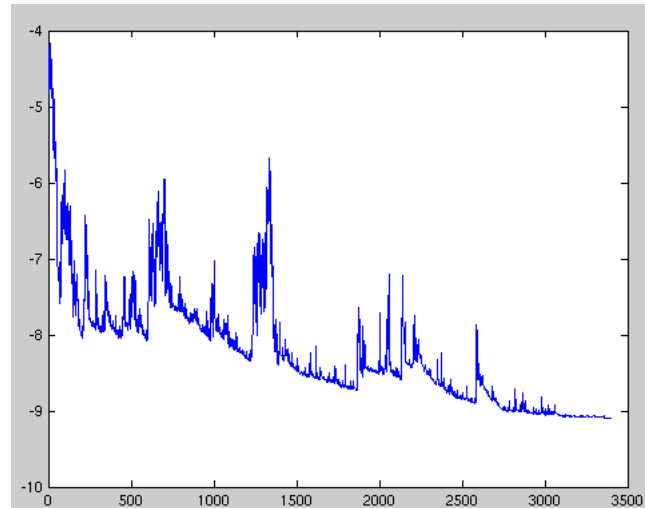
Σχήμα 3.16: Βελτιστοποίηση συνάρτησης μέσω αλγορίθμου κατάβασης κλίσης (Πηγή: [2])

Βάσει του μεγέθους των δεδομένων που χρησιμοποιεί ο αλγόριθμος έχουμε τρεις διαφορετικές παραλλαγές τους. Αρχικά ο Batch gradient descent, υπολογίζει τις κλίσεις της συνάρτησης σφάλματος σε σχέση με όλες τις παραμέτρους θ . Λόγω της ανάγκης του λοιπόν να υπολογίζει όλες τις παραγωγούς των παραμέτρων για κάθε ανανέωση, ο Batch gradient descent είναι αρκετά αργός και μπορεί να δημιουργήσει προβλήματα για ομάδες δεδομένων που δε χωράνε στην μνήμη. Ακόμα ο Batch gradient descent εγγυάται ότι θα συγκλίνει στο ολικό ελάχιστο για κυρτές επιφάνειες συναρτήσεων σφάλματος και στο τοπικό ελάχιστο για μη κυρτές επιφάνειες σφάλματος [54].

Ο Stochastic gradient descent (SGD) σε αντίθεση τώρα πραγματοποιεί ανανέωση των παραμέτρων για κάθε δείγμα εκπαίδευσης $x^{(i)}$ και ετικέτα $y^{(i)}$:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)}) \quad (3.28)$$

Η βασική διαφορά του Batch gradient descent με τον SGD, είναι πως ο πρώτος κάνει περιττούς υπολογισμούς για μεγάλα σετ δεδομένων, καθώς επανα-υπολογίζει τις παραγωγούς για παρόμοια παραδείγματα για κάθε ανανέωση αυτών. Ο SGD παραλείπει τους υπολογισμούς αυτούς, πραγματοποιώντας έναν κάθε φορά. Το γεγονός αυτό, τον καθιστά πολύ γρηγορότερο, βέβαια λόγω του υψηλού ρυθμού ανανέωσης προκαλεί μεγάλες αποκλίσεις στην συνάρτηση κόστους όπως μπορεί να διαπιστωθεί από το σχήμα 3.17.



Σχήμα 3.17: Συνάρτηση κόστους αλγορίθμου στοχαστικής κατάβασης κλίσης (Πηγή: [75])

Ακόμα ενώ ο Batch gradient descent συγκλίνει σε ένα ελάχιστο, ο SGD λόγω των ταλαντώσεων του επιτρέπει την μεταπήδηση σε ένα άλλο τοπικό ελάχιστο ενδεχομένως καλύτερο του προηγούμενου. Έτσι, ενώ αυτό περιπλέκει την σύγκλιση στο ελάχιστο, αποδεικνύεται ότι όταν ο ρυθμός μάθησης του SGD μειώνεται με μικρό βήμα, παρουσιάζει την σύγκλιση του Batch gradient descent και είναι σχεδόν βέβαιο ότι θα συγκλίνει στο τοπικό ή ολικό ελάχιστο, για μη κυρτές και κυρτές βελτιστοποιήσεις αντίστοιχα.

Τέλος ο Mini-batch gradient descent, συνδυάζει τις δυο παραπάνω προσεγγίσεις και ανανεώνεται για κάθε mini-batch κάποιον n δεδομένων εκπαίδευσης :

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i:i+n)}; y^{(i:i+n)}) \quad (3.29)$$

Με αυτόν τον τρόπο, πρώτον μειώνει τις αποκλίσεις κατά την ανανέωση των παραμέτρων, γεγονός που οδηγεί σε σταθερότερη σύγκλιση, ενώ παράλληλα μπορεί να κάνει χρήση πινάκων βελτιστοποίησης, που χρησιμοποιούνται από βιβλιοθήκες βαθιάς μηχανικής μάθησης για τον υπολογισμό παραγώγων σε ομάδες mini-batches με μεγάλη αποτελεσματικότητα [54].

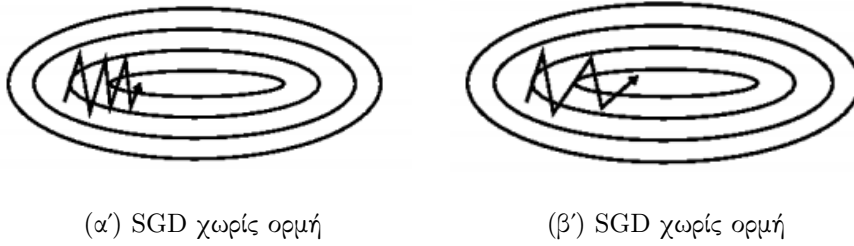
Ωστόσο παρόλες τις τροποποιήσεις, ο αλγόριθμος Mini-batch gradient descent μπορεί να είναι προβληματικός μιας και α) η επιλογή του κατάλληλου ρυθμού μάθησης μπορεί να είναι δύσκολη, αφού αν είναι πολύ μικρός ενδεχομένως να έχουμε αργή σύγκλιση, αντιθέτως αν είναι αρκετά μεγάλος η σύγκλιση μπορεί να εμποδιστεί και να προκληθεί ταλάντωση γύρω από το ελάχιστο της συνάρτησης σφάλματος β) ο ρυθμός μάθησης είναι ο ίδιος σε όλες τις παραμέτρους, κάτι το οποίο δεν είναι αποτελεσματικό σε αραιά δεδομένα όπου θα θέλαμε να μην ανανεώνονται όλες οι παράμετροι με την ίδια συχνότητα γ) η ελαχιστοποίηση μιας μη κυρτής συνάρτησης σφάλματος παγιδεύεται συχνά σε τοπικά ελάχιστα, τα οποία δεν είναι βέλτιστα.

Λόγω των παραπάνω δυσκολιών, προτείνεται η προσθήκη της ορμής (Momentum) στον αλγόριθμο SGD. Ο SGD αντιμετωπίζει πρόβλημα σε επιφάνειες καμπυλών οι οποίες είναι πιο απότομες σε κάποιες διαστάσεις από ότι σε άλλες. Έτσι, καταλήγει να ταλαντώνεται στις τοπικές κλίσεις, κάνοντας διστακτικά βήματα προς το τοπικό ελάχιστο. Μέσω λοιπόν της

ορμής, μπορεί να επιταχυνθεί ο αλγόριθμος SGD προς την κατεύθυνση του ελαχίστου, με την προσθήκη μιας παραμέτρου γ (μια από τις συνηθισμένες τιμές της οποίας είναι το 0.9) που επιτρέπει την ανανέωση μόνο ενός μέρους των παραμέτρων της προηγούμενης χρονικής στιγμής. Η βελτίωση αυτή απεικονίζεται στο σχήμα 3.18 και μαθηματικά περιγράφεται ως :

$$u_t = \gamma u_{t-1} + \eta \cdot \nabla_{\theta} J(\theta) \quad (3.30\alpha')$$

$$\theta = \theta - u_t \quad (3.30\beta')$$



Σχήμα 3.18: Αλγόριθμος κατάβασης κλίσης με και χωρίς την χρήση ορμής (Πηγή: [76])

Adagrad

Ο Adagrad είναι ακόμα ένας αλγόριθμος βασισμένος στην κλίση (gradient-based algorithm). Χαρακτηριστικό του αποτελεί, ότι προσαρμόζει το ρυθμό μάθησης στις παραμέτρους του, με αποτέλεσμα να εκτελεί μεγαλύτερες ανανεώσεις για μη συχνές παραμέτρους και μικρότερες για πιο συχνές παραμέτρους. Αυτή του η ιδιότητα, τον κάνει ιδανικό για διαχείριση αραιών δεδομένων. Σε αντίθεση λοιπόν με τον αλγόριθμο κατάβασης κλίσης, που πραγματοποιεί την ανανέωση των παραμέτρων του θ με τον ίδιο ρυθμό η , στον αλγόριθμο Adagrad, χρησιμοποιείται διαφορετικός ρυθμός μάθησης για κάθε παράμετρο θ_i σε κάθε χρονική στιγμή t . Χρησιμοποιώντας λοιπόν τον συμβολισμό g_t , δηλώνοντας την κλίση κατά την χρονική στιγμή t , με το $g_{t,i}$ δηλώνουμε την μερική παράγωγο της αντικειμενικής συνάρτησης την χρονική στιγμή t ως προς τις παραμέτρους της θ_i :

$$g_{t,i} = \nabla_{\theta} J(\theta_{t,i}) \quad (3.31)$$

Τότε η ανανέωση των παραμέτρων θ σύμφωνα με τον αλγόριθμο SGD θα έχει την μορφή :

$$\theta_{t+1,i} = \theta_{t,i} - \eta \cdot g_{t,i} \quad (3.32)$$

Κατά την ανανέωση όμως του αλγορίθμου Adagrad, ο ρυθμός μάθησης η προσαρμόζεται σε κάθε χρονική στιγμή t για κάθε παράμετρο θ_i βάσει των προηγούμενων κλίσεων που έχουν υπολογιστεί για την συγκεκριμένη παράμετρο :

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} \cdot g_{t,i} \quad (3.33)$$

Όπου το $G_t \in \mathbb{R}^{d \times d}$ αποτελεί έναν διαγώνιο πίνακα με κάθε διαγώνιο στοιχείο του i, i να αντιπροσωπεύει το άθροισμα των τετραγώνων των κλίσεων των παραμέτρων θ_i μέχρι την

χρονική στιγμή t , ενώ το ϵ είναι μια παράμετρος εξομάλυνσης, ώστε να αποφεύγεται η διαίρεση με το μηδέν. Τέλος, η ρίζα χρησιμοποιείται καθώς αποδεικνύεται στην πράξη ότι χωρίς την χρήση της ο αλγόριθμος αποδίδει σημαντικά χειρότερα. Μια από τις αδυναμίες του αλγορίθμου αυτού αποτελεί η συσσώρευση των τετραγώνων των κλίσεων στον παρονομαστή, αφού οι όροι που προστίθενται είναι πάντα θετικοί, ώστε ο ρυθμός μάθησης να μικραίνει και να φτάνει σε σημείο να μη μπορεί να μάθει περαιτέρω πληροφορίες.

Adadelta

Ο αλγόριθμος Adadelta είναι μια προέκταση του αλγορίθμου Adagrad, με στόχο να βελτίωση αυτή την μόνιμη μείωση του ρυθμού μάθησης. Η διαφορά τους έγκειται στο ότι ο Adadelta χρησιμοποιεί ένα παράθυρο περιορισμού μήκους w για την συσσώρευση των παρελθοντικών κλίσεων. Για το παράθυρό αυτό επομένως, ο αλγόριθμος υπολογίζει αναδρομικά το άθροισμα των κλίσεων ως μέσο όρο των προηγούμενων κλίσεων, με αποτέλεσμα ο μέσος όρος $E[g^2]_t$ σε μια χρονική στιγμή t να εξαρτάται από το προηγούμενο μέσο όρο και την κλίση σε αυτή τη χρονική στιγμή :

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1 - \gamma)g_t^2 \quad (3.34)$$

Έτσι η εξίσωση ενημέρωσης των παραμέτρων γίνεται :

$$\Delta\theta_t = -\frac{\eta}{\sqrt{E[g^2]_t + \epsilon}}g_t \quad (3.35)$$

Παρατηρώντας όμως των παρονομαστή της εξίσωσης βλέπουμε ότι αποτελεί την ρίζα του μέσου τετραγώνου των κλίσεων με το οποίο και το αντικαθιστούμε

$$\Delta\theta_t = -\frac{\eta}{RMS[g]_t}g_t \quad (3.36)$$

Έπειτα όμως από παρατήρηση των συντακτών, αναφορικά με τις παραμέτρους (ότι η ανανέωση πρέπει να έχει τις ίδιες υποθετικές μονάδες με αυτές των παραμέτρων), η εξίσωση 3.37 τροποποιήθηκε, ώστε να μην περιλαμβάνει τα τετράγωνα των κλίσεων αλλά τετράγωνα των ανανεώσεων των παραμέτρων :

$$E[\Delta\theta^2]_t = \gamma E[\Delta\theta^2]_{t-1} + (1 - \gamma)\Delta\theta_t^2 \quad (3.37)$$

Έτσι η ρίζα των μέσων τετραγώνων παίρνει την μορφή:

$$RMS[\Delta\theta]_t = \sqrt{E[\Delta\theta^2]_t + \epsilon} \quad (3.38)$$

Πλέον όμως δεδομένου ότι ο όρος $RMS[\Delta\theta]_t$ είναι άγνωστος, τον υπολογίζουμε προσεγγιστικά μέσω του RMS των παραμέτρων που έχουν ανανεωθεί μέχρι την προηγούμενη χρονική στιγμή. Αντικαθιστώντας τον ρυθμό μάθησης η στον προηγούμενο κανόνα ανανέωσης με το $RMS[\Delta\theta]_{t-1}$, προκύπτει η τελική μορφή του αλγορίθμου Adadelta:

$$\Delta\theta_t = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t}g_t \quad (3.39\alpha')$$

$$\theta_{t+1} = \theta_t + \Delta\theta_t \quad (3.39\beta')$$

RMSprop

Ο αλγόριθμος RMSprop είναι ένας αλγόριθμος προσαρμοστικής μάθησης που προτάθηκε από τον Geoff Hinton. Και αυτός αναπτύχθηκε ώστε να προσφέρει λύση στο πρόβλημα του αλγορίθμου Adagrad, σχετικά με την εκμηδένιση του ρυθμού μάθησης. Μαθηματικά είναι ίσος με το διάνυσμα της πρώτης ανανέωσης του Adadelta δηλαδή :

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2 \quad (3.40\alpha')$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \quad (3.40\beta')$$

Ο RMSprop επίσης διαιρεί τον ρυθμό μάθησης με τον εκθετικά μειούμενο μέσο όρο των τετραγώνων των κλίσεων. Ο Geoff Hinton προτείνει η τιμή του γ να είναι 0.9, ενώ μια καλή αρχική τιμή του ρυθμού μάθησης είναι 0.001.

Adam

Ο αλγόριθμος Adaptive Moment Estimation (Adam), είναι ένας ακόμα αλγόριθμος ο οποίος χρησιμοποιεί προσαρμοσμένους ρυθμούς μάθησης για κάθε παράμετρό του. Εκτός όμως από την αποθήκευση του εκθετικά μειούμενου μέσου όρου των παρελθοντικών τετραγώνων των κλίσεων, αποθηκεύει και τον εκθετικά μειούμενο μέσο όρο των παρελθοντικών κλίσεων m_t αντίστοιχα με την προσθήκη της ορμής momentum που έχει προηγουμένως αναλυθεί :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (3.41\alpha')$$

$$u_t = \beta_2 u_{t-1} + (1 - \beta_2) g_t^2 \quad (3.41\beta')$$

όπου τα m_t και u_t είναι εκτιμήσεις της πρώτης ορμής (μέσου) και της δεύτερης ορμής (μη κεντραρισμένη απόκλιση) των κλίσεων. Λόγω του ότι τα m_t και u_t αρχικοποιούνται ως πίνακες μηδενικών, παρατηρήθηκε ότι είναι μεροληπτικά γύρω από αυτή την περιοχή, κυρίως κατά τις αρχικές χρονικές στιγμές. Για την εξάλειψη αυτής της μεροληψίας εκτιμητές του πρώτου και του δεύτερου όρου :

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.42\alpha')$$

$$\hat{u}_t = \frac{u_t}{1 - \beta_2^t} \quad (3.42\beta')$$

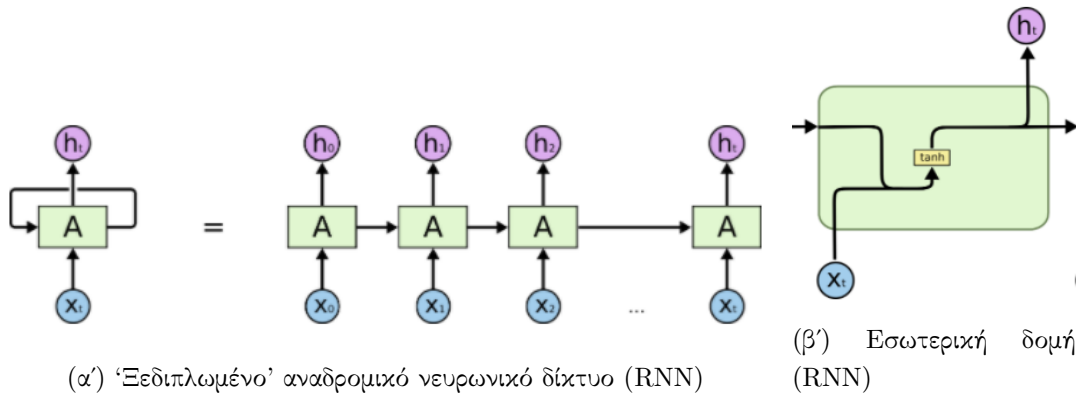
Στην συνέχεια χρησιμοποιούμε τα παραπάνω για την ανανέωση των παραμέτρων σε αντιστοιχία με τις μεθόδους Adadelta και RMSprop :

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{u}_t + \epsilon}} \hat{m}_t \quad (3.43)$$

3.2.9 Αναδρομικά νευρωνικά δίκτυα (RNN)

Τα αναδρομικά νευρωνικά δίκτυα χαρακτηρίζονται από εσωτερικές συνδέσεις, οι οποίες επιτρέπουν την επεξεργασία χρονικών συσχετίσεων. Όταν τα αναδρομικά νευρωνικά δίκτυα επεξεργάζονται σειριακές πληροφορίες, πραγματοποιούν τις ίδιες μετατροπές και διαδικασίες σε κάθε στοιχείο της σειριακής εισόδου. Η έξοδος των RNN σε κάθε χρονικό βήμα εξαρτάται από την προηγούμενη είσοδο και τους προηγούμενους υπολογισμούς του νευρωνικού

δικτύου. Ως συνέπεια, το RNN ενσωματώνει παρελθοντικές πληροφορίες μαζί με πληροφορίες της κάθε εισόδου ώστε να προβλέψει μελλοντικές τιμές. Αυτή η διαδικασία επιτρέπει στο νευρωνικό δίκτυο να σχηματίζει μνήμη των προηγούμενων δεδομένων, την οποία κωδικοποιεί στις μεταβλητές του κρυμμένου επιπέδου (hidden state) του. Λόγω της σειριακής επεξεργασίας δεδομένων τα αναδρομικά νευρωνικά δίκτυα θεωρούνται κατάλληλα για την επεξεργασία χρονοσειρών και άλλων ακολουθιακών δεδομένων [6]. Η απεικόνιση ενός ‘ξεδιπλωμένου’ αναδρομικού δικτύου παρουσιάζεται στο σχήμα 3.19



Σχήμα 3.19: Αναδρομικό νευρωνικό δίκτυο (RNN) (Πηγή: [8])

Παρά την κατάλληλη αρχιτεκτονική τους, για τον χειρισμό ακολουθιακών δεδομένων, παρατηρείται ότι τα RNN νευρωνικά δίκτυα εμφανίζουν προβλήματα στην διαχείριση μεγάλου μήκους ακολουθιακών εισόδων. Αιτία του προβλήματος αποτελεί η εξαφάνιση κλίσεων (vanishing gradients), δηλαδή η μείωση των κλίσεων σε πολύ μικρές τιμές που έχουν ως αποτέλεσμα την πρακτικά μηδενική ανανέωση των βαρών. Κάτι τέτοιο μπορεί να εξηγηθεί λόγω του ότι τα αναδρομικά νευρωνικά δίκτυα για τον υπολογισμό των κλίσεων τους δεν βασίζονται στον αλγόριθμο Backpropagation, αλλά σε μια παραλλαγή του που ονομάζεται Backpropagation through time (BPTT). Θεωρώντας λοιπόν ότι το κρυφό επίπεδο του RNN νευρωνικού δικτύου περιγράφεται από την εξίσωση $h_t = \tanh(W_I x_t + W_R h_{t-1})$ και η έξοδός του από την $y_t = W_O h_t$ για τον αλγόριθμο BPTT, όπως και για τον Backpropagation θα πρέπει να υπολογίσουμε κλίση του σφάλματος E κατά την στιγμή t ως προς το W_R . Εφαρμόζοντας επομένως τον κανόνα της αλυσίδας η μερική αυτή παράγωγος θα πάρει την μορφή :

$$\frac{\partial E_t}{\partial W_R} = \sum_{i=0}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_i} \frac{\partial h_i}{\partial W_R} \quad (3.44)$$

Οι παραπάνω όροι λοιπόν είναι εύκολο να υπολογιστούν με εξαίρεση ίσως το κλάσμα $\frac{\partial h_t}{\partial h_i}$ το οποία απαιτεί την χρήση του κανόνα της αλυσίδας :

$$\frac{\partial h_t}{\partial h_i} = \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdots \frac{\partial h_{i+1}}{\partial h_i} = \prod_{k=i}^{t-1} \frac{\partial h_{k+1}}{\partial h_k} \quad (3.45)$$

Αναλύοντας τώρα ένα από τους όρους του γινομένου αυτού η μερική παράγωγος θα έχει την

μορφή :

$$\frac{\partial h_{k+1}}{\partial h_k} = \text{diag}(f'(W_I x_i + W_R h_{i-1})) W_R \quad (3.46)$$

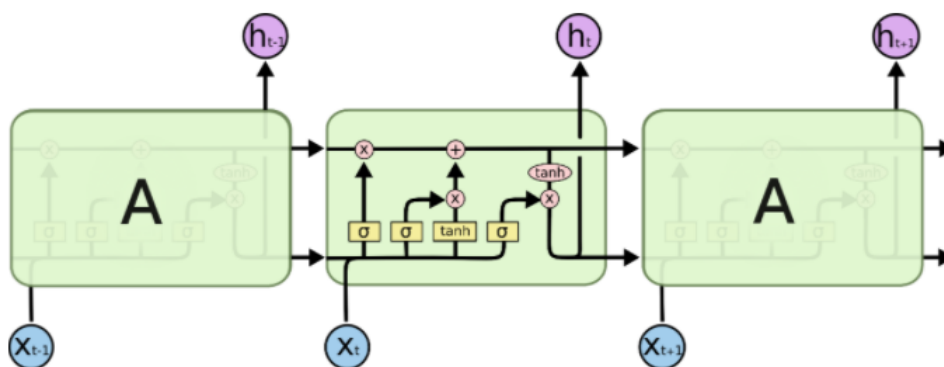
Έτσι λοιπόν με την διάδοση του για k χρονικές στιγμές οι συνολική παράγωγος θα γίνει :

$$\frac{\partial h_k}{\partial h_1} = \prod_i^k \text{diag}(f'(W_I x_i + W_R h_{i-1})) W_R \quad (3.47)$$

Σύμφωνα λοιπόν με τις παραπάνω εξισώσεις [46] αλλά και με τα συμπεράσματα του [49], παρατηρούμε ότι λόγω του ότι η τιμή της παραγώγου $f'(x)$ (μιας και η $f(x) = \tanh(x)$) είναι μικρότερη του 1, οδηγεί το γινόμενο εξαιτίας των πολλαπλών τιμών σε πολύ μικρές τιμές. Αυτό σε συνδυασμό με το ότι το W_R μπορεί να πάρει μικρές τιμές οδηγεί την τιμή της παραγώγου στο 0 έχοντας εξαφάνιση της κλίσης (vanishing gradients), χωρίς να ανανεώνονται τα βάρη. Σε αντίθετη περίπτωση αν η τιμές του W_R μεγαλώσουν πολύ θα οδηγήσουν το γινόμενο σε εκθετική αύξηση υπερνικώντας τις μικρές τιμές των παραγώγων. Αν και στην πράξη η εκθετική αύξηση δεν είναι τόσο συχνή η εξαφάνιση, διαδιδόμενη μέσα σε πολλαπλές χρονικές στιγμές επηρεάζει τα αρχικά κρυφά επίπεδα κάνοντας την επίδραση τους στα επόμενα ασήμαντη και επομένως χάνεται η δυνατότητα συγκράτησης και μάθησης πληροφορίας σε μεγάλου μήκους ακολουθιακών δεδομένων.

3.2.10 Νευρωνικό δίκτυο μακράς-βραχείας μνήμης (LSTM)

Για την επίλυση των παραπάνω προβλημάτων, αναπτύχθηκε το νευρωνικό δίκτυο μακράς-βραχείας μνήμης LSTM το 1997 από τους Hochreiter και Schmidhuber [23]. Η βασική ιδέα πίσω από την δημιουργία του ήταν η ανάγκη αποφυγής των προβλημάτων στην επεξεργασία μεγάλου μήκους εξαρτήσεων και την απόκτηση μνήμης για μεγάλες περιόδους. Ενώ η επαναλαμβανόμενη μονάδα του LSTM ακολουθεί την αναδρομική αρχιτεκτονική του RNN, η εσωτερική της δομή που φαίνεται στο σχήμα 3.20 είναι αρκετά πιο περίπλοκη.

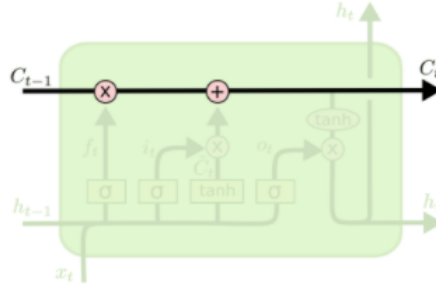


Σχήμα 3.20: Κρυφό κύτταρο LSTM (Πηγή: [8])

Λόγω της σημαντικότητας του μοντέλου αλλά και της πολυπλοκότητάς του, θα αναλυθεί ακριβώς η αρχιτεκτονική του κρυφού του επιπέδου, το οποίο αποτελείται από το κύτταρο cell

και τρεις πύλες (gates), οι οποίες ελέγχουν την ροή της πληροφορίας στην μονάδα του LSTM και ονομάζονται πύλη εισόδου (input gate), πύλη εξόδου (output gate), πύλη λυσμόνησης (forget gate).

Για να ξεκινήσουμε την πιο λεπτομερή περιγραφή, πρέπει να επισημανθεί πως ένα κύτταρο LSTM διατρέχεται κατά μήκος του από την δομή της κυτταρικής κατάστασης (cell state) που φαίνεται στο σχήμα 3.21.



Σχήμα 3.21: Κυτταρική κατάσταση LSTM (Πηγή: [8])

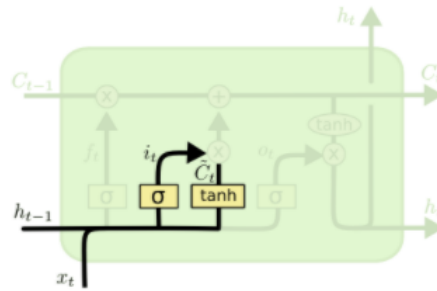
Το cell state, είναι υπεύθυνο για την μεταφορά της πληροφορίας από τον έναν κόμβο της αλυσίδας στο επόμενο. Το κύτταρο του LSTM έχει την δυνατότητα να προσθέτει και να αφαιρεί πληροφορίες στην κυτταρική κατάσταση μέσω των πυλών που αναφέρθηκαν ονομαστικά. Έτσι, κατά το πρώτο βήμα της επεξεργασίας η πύλη λυσμόνησης κοιτάει την προηγούμενη κρυφή κατάσταση h_{t-1} και την είσοδο x_t την χρονική στιγμή t και δίνει σαν έξοδο έναν αριθμό ανάμεσα στην 0 και το 1 για κάθε αριθμό στην κυτταρική κατάσταση C_{t-1} . Ο αριθμός αυτός, δηλώνει το ποσοστό της πληροφορίας που πρέπει να περάσει στο επόμενο στάδιο, με το 0 να δηλώνει το να μην περάσει καθόλου η προηγούμενη πληροφορία και το 1 να σημαίνει να περάσει χωρίς καμία αλλαγή. Η έξοδος του πρώτου σταδίου είναι περιορισμένη στις τιμές 0 και 1 μιας και το αποτέλεσμα των h_{t-1} και x_t της πύλης λυσμόνησης περνάει από μια σιγμοειδή συνάρτηση για τον περιορισμό του. Η συνάρτηση που περιγράφει το πρώτο στάδιο της επεξεργασίας είναι:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.48)$$

Το επόμενο βήμα σχετίζεται με την επιλογή της ποιός καινούργιας πληροφορίας θα αποθηκεύσουμε στην κυτταρική κατάσταση. Το βήμα αυτό χωρίζεται σε δυο επιμέρους καταστάσεις, όπου αρχικά ένα σιγμοειδές επίπεδο το οποίο ονομάζεται επίπεδο εισόδου, αποφασίζει ποιες τιμές θα ανανεώσουμε. Στην συνέχεια ένα επίπεδο υπερβολικής εφαπτομένης σχηματίζει έναν πίνακα των καινούργιων υποψηφίων τιμών \tilde{C}_t , που θα μπορούσαν να προστεθούν στην κυτταρική κατάσταση. Έπειτα, συνδυάζονται οι δυο παραπάνω διαδικασίες, ώστε να υλοποιήσουμε την ανανέωση της κυτταρικής κατάστασης. Το δεύτερο στάδιο απεικονίζεται στο σχήμα 3.22 και περιγράφεται από τις εξισώσεις :

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.49\alpha')$$

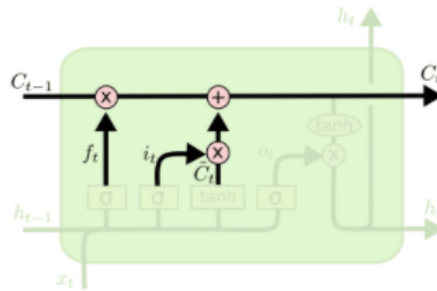
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.49\beta')$$



Σχήμα 3.22: Πύλη εισόδου (Πηγή: [8])

Πλέον, πρέπει να ανανεώσουμε την παλιά κυτταρική κατάσταση C_{t-1} στην καινούργια C_t . Έτσι, πολλαπλασιάζουμε την παλιά κατάσταση με την πύλη λησμόνησης f_t , ώστε να ξεχάσει η παλιά κατάσταση τις πληροφορίες που έχουμε επιλέξει, ενώ ακόμα προσθέτουμε στην παλιά κατάσταση την υποψήφια καινούργια $i_t * \tilde{C}_t$, προσαρμοσμένη κατάλληλα μέσω του όρου i_t . Η απεικόνιση της νέας κυτταρικής κατάστασης φαίνεται στο σχήμα 3.23 και περιγράφεται ως :

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3.50)$$

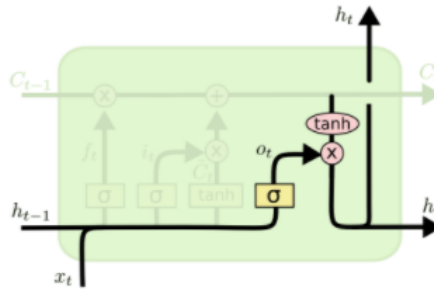


Σχήμα 3.23: Νέα κυτταρική κατάσταση (Πηγή: [8])

Τελευταίο βήμα στην επεξεργασία των πληροφοριών στο κύτταρο του LSTM, αποτελεί η απόφαση της εξόδου του. Η έξοδος θα βασιστεί σε μια φιλτραρισμένη εκδοχή της κυτταρικής κατάστασης. Πιο συγκεκριμένα, αρχικά χρησιμοποιούμε ένα σιγμοειδές επίπεδο το οποίο αποφασίζει από ποια κομμάτια θα αποτελείται η έξοδος. Στην συνέχεια, περνάμε την κυτταρική κατάσταση από ένα επίπεδο υπερβολικής επαφτομένης, ώστε να περιορίσουμε τις τιμές ανάμεσα στο -1 και το 1 και την πολλαπλασιάζουμε με το σιγμοειδές επίπεδο, ώστε τελικά να αφήσουμε στην έξοδο μόνο το κομμάτι το οποίο επιθυμούμε. Η τελική αυτή διαδικασία φαίνεται στο σχήμα 3.24 και μαθηματικά περιγράφεται ως :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.51\alpha')$$

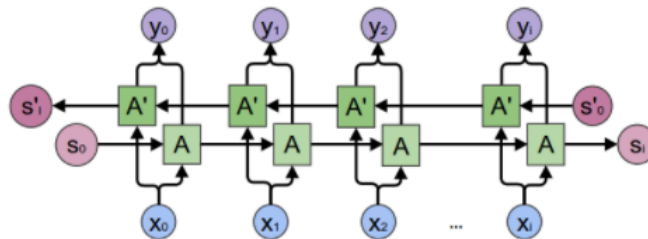
$$h_t = o_t * \tanh(C_t) \quad (3.51\beta')$$



Σχήμα 3.24: Έξοδος κυττάρου LSTM (Πηγή: [8])

3.2.11 Νευρωνικό δίκτυο μακράς-βραχείας μνήμης διπλής κατεύθυνσης (Bidirectional-LSTM)

Το Bidirectional-LSTM νευρωνικό δίκτυο, αποτελεί μια πιο εμπλουτισμένη εκδοχή του απλού LSTM δικτύου. Η αρχιτεκτονική του αποτελείται ουσιαστικά από τον διπλασιασμό του απλού LSTM, όπως φαίνεται και από το σχήμα 3.25. Παρατηρούμε ότι το διάνυσμα εισόδου παρουσιάζεται μία φορά με την κανονική χρονική ακολουθία και μια με την ανάποδη. Μετά την επεξεργασία του κάθε επιπέδου, οι έξοδοι ενώνονται για κάθε χρονική στιγμή. Λόγω της συγκεκριμένης αρχιτεκτονικής λοιπόν, επιτρέπεται στο δίκτυο να επεξεργαστεί τα ακολουθιακά δεδομένα προς δυο κατευθύνσεις κάτι το οποίο αποδεικνύεται ότι σε ορισμένες περιπτώσεις βελτιώνει σημαντικά τα αποτελέσματα σε σχέση με το παραδοσιακό LSTM.

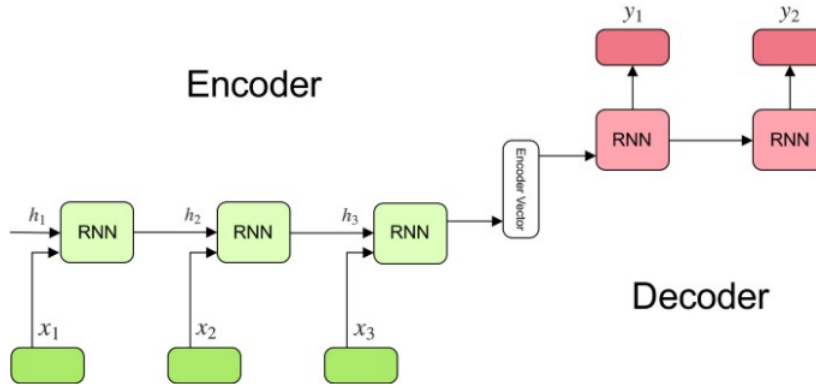


Σχήμα 3.25: Bidirectional LSTM (Πηγή: [7])

3.2.12 Αρχιτεκτονική Encoder-Decoder

Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (encoder-decoder) είναι μια ευρέως χρησιμοποιούμενη τεχνική για την συσχέτιση ακολουθιών διαφορετικού μήκους. Παρουσιάστηκε πρώτη φορά το 2014 από τον Cho [5], για την μετάφραση κειμένων και έχει βρει μεγάλη πληθώρα εφαρμογών έκτοτε. Η λογική της για τον αποτελεσματικό χειρισμό ακολουθιακών δεδομένων εισόδου-εξόδου διαφορετικού μήκους, μπορεί να συνοψιστεί σε τρία διαφορετικά στάδια. Αρχικά, ο κωδικοποιητής (encoder) παίρνει μια μεταβλητού μήκους ακολουθία και την μετατρέπει σε ένα διάνυσμα συγκεκριμένου μήκους, το κωδικοποιημένο διάνυσμα (encoded vector). Το διάνυσμα αυτό είναι το τελευταίο κρυφό επίπεδο που παράχθηκε μέσω του κωδικοποιητή και στόχος του είναι να συμπεριλάβει σημαντικές πληροφορίες από όλα τα

στοιχεία εισόδου. Το διάνυσμα αυτό στην συνέχεια παραλαμβάνεται από τον αποκωδικοποιητή (decoder) και αντιστοιχίζεται σε μια μεταβλητού μήκους έξοδο, παράγοντας μια έξοδο για κάθε χρονικό βήμα. Η αρχιτεκτονική του κωδικοποιητή-αποκωδικοποιητή περιγράφεται στο σχήμα 3.26.



Σχήμα 3.26: Αρχιτεκτονική Encoder Decoder (Πηγή: [58])

Πιο αναλυτικά, ο αποκωδικοποιητής λαμβάνει μια ακολουθιακή είσοδο που αποτελείται από ακολουθιακούς πίνακες $x = (x_1, \dots, x_{T_x})$ και τους μετατρέπει σε ένα πίνακα c . Στην πιο απλή της μορφή, η αρχιτεκτονική αποτελείται από RNN νευρωνικά δίκτυα (ωστόσο συχνά χρησιμοποιείται με μονάδες LSTM ή GRU) και περιγράφεται από τις εξισώσεις :

$$h_t = f(x_t, h_{t-1}) \quad (3.52\alpha')$$

$$c = q(h_1, \dots, h_{T_x}) \quad (3.52\beta')$$

όπου το $h_t \in \mathbb{R}^n$ είναι η κρυφή κατάσταση την χρονική στιγμή t και c είναι ο πίνακας που υπολογίζεται από την ακολουθία των κρυμμένων καταστάσεων. Οι συναρτήσεις f και q είναι μη γραμμικές συναρτήσεις. Ο αποκωδικοποιητής εκπαιδεύεται ώστε να προβλέπει την επόμενη έξοδο y'_t λαμβάνοντας το κωδικοποιημένο διάνυσμα c και όλες τις προηγούμενες εξόδους που έχουν προβλεφθεί y_1, \dots, y_{t-1} . Με άλλα λόγια ο αποκωδικοποιητής εκφράζει την πιθανότητα της εξόδου y :

$$p(y) = \prod_{t=1}^T p(y_t | y_1, \dots, y_{t-1}, c) \quad (3.53)$$

όπου το $y = y_1, \dots, y_{T_y}$. Χρησιμοποιώντας RNN νευρωνικά δίκτυα κάθε μια από τις παραπάνω πιθανότητες του γινομένου μπορεί να υπολογιστεί ως :

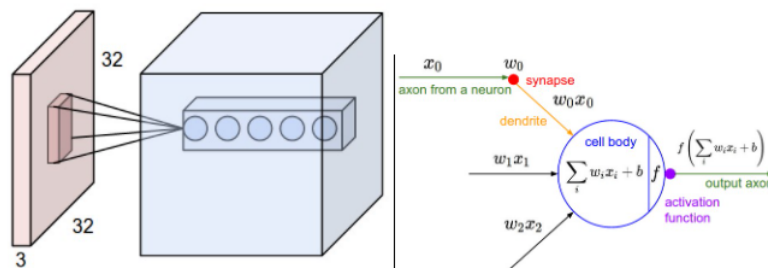
$$p(y_t | y_1, \dots, y_{t-1}, c) = g(y_t - 1, s_t, c) \quad (3.54)$$

όπου g είναι μια μη γραμμική συνάρτηση που στην έξοδό της δίνει την πιθανότητα του y_t και s_t είναι η κρυφή κατάσταση του RNN.

3.2.13 Συνελικτικά νευρωνικά δίκτυα (CNN)

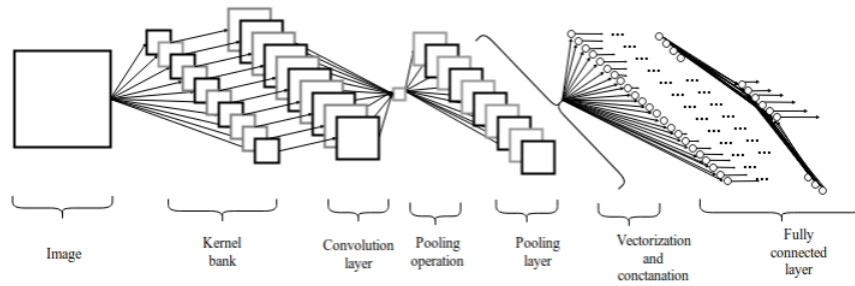
Τα συνελικτικά νευρωνικά δίκτυα (CNN), αποτελούν μια κατηγορία νευρωνικών δικτύων τα οποία έχουν γίνει ιδιαίτερα δημοφιλή τα τελευταία χρόνια. Βασικό τους πλεονέκτημα, απο-

τελεί το ότι έχουν την δυνατότητα να μειώσουν τις παραμέτρους που χρησιμοποιούν συγκριτικά με τα απλά νευρωνικά δίκτυα, κάτι το οποίο έδωσε την δυνατότητα στους επιστήμονες να χρησιμοποιήσουν μεγαλύτερα μοντέλα για την επίλυση σύνθετων υπολογιστικών προβλημάτων. Χρησιμοποιούνται κυρίως για την επεξεργασία δεδομένων που έχουν πλεγματοειδή τοπολογία, μιας και έχει παρατηρηθεί ότι αποδίδουν εξαιρετικά στην μοντελοποίηση δεδομένων με χωρική και χρονική εξάρτηση. Παραδείγματα αυτών, αποτελούν οι χρονοσειρές που έχουν την μορφή πλέγματος μιας διάστασης, παίρνοντας τα δείγματα σε σταθερές χρονικές στιγμές, αλλά και οι εικόνες οι οποίες αποτελούν πλέγματα από πίξελ δύο διαστάσεων. Το όνομα τους λαμβάνεται από την βασική μαθηματική πράξη που τα χαρακτηρίζει δηλαδή την συνέλιξη (convolution). Η συνέλιξη είναι μια απλή γραμμική διαδικασία, μέσω της οποίας μπορούμε να απλοποιήσουμε κατά πολύ την πολυπλοκότητα των νευρωνικών δικτύων, αφού παίρνει την θέση του πολλαπλασιασμού πινάκων που χρησιμοποιείται στα πολυεπίπεδα νευρωνικά δίκτυα. Έτσι, σε αντίθεση με τα απλά νευρωνικά δίκτυα στα επίπεδα των συνελικτικών νευρωνικών δικτύων υπάρχουν νευρώνες ταξινομημένοι σε τρεις διαστάσεις, πλάτος, ύψος και βάθος, με την τελευταία διάσταση να αναφέρεται στην τρίτη διάσταση της εισόδου του δικτύου. Έτσι, μια έγχρωμη εικόνα θα έχει διαστάσεις π.χ $32 \times 32 \times 3$ με την τρίτη διάσταση να αναφέρεται στα 3 διαφορετικά κανάλια χρώματος της έγχρωμης εικόνας ((RGB), Red, Green, Blue). Προσθετικά όπως παρατηρούμε στην εικόνα 3.27, οι νευρώνες ενός συνελικτικού επιπέδου ενός συνελικτικού δικτύου συνδέονται σε τοπικές περιοχές και σε όλο το βάθος, σε αντίθεση με τους νευρώνες των απλών νευρωνικών δικτύων που συνδέονται πλήρως.



Σχήμα 3.27: Σύγκριση νευρώνων συνελικτικού και απλού νευρωνικού δικτύου (Πηγή: [65])

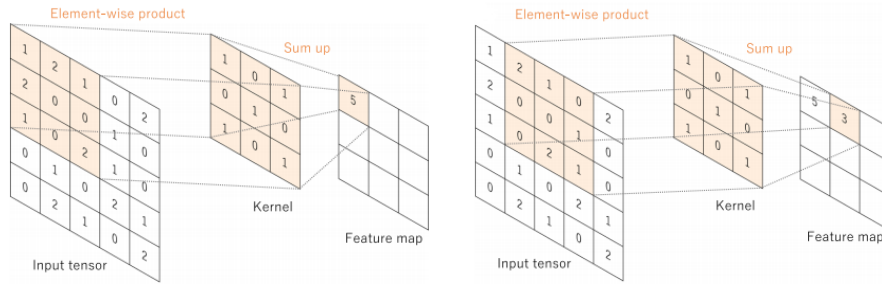
Παρακάτω θα αναλύσουμε τα βασικά επίπεδα που απαρτίζουν ένα συνελικτικό δίκτυο, του οποίου η αρχιτεκτονική περιγράφεται στο σχήμα 3.28:



Σχήμα 3.28: Αρχιτεκτονική συνελικτικού νευρωνικού δικτύου (Πηγή: [43])

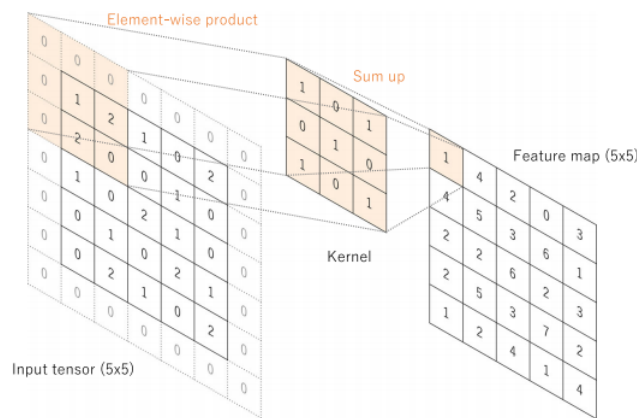
Συνελικτικό επίπεδο (Convolutional Layer)

Το συνελικτικό επίπεδο είναι το πιο σημαντικό επίπεδο των συνελικτικών δικτύων, καθώς μέσω αυτού το συνελικτικό δίκτυο μπορεί να εξάγει χρήσιμα χαρακτηριστικά για την εισόδο του. Η συνέλιξη λοιπόν, είναι μια ειδικού τύπου γραμμική πράξη για την εξαγωγή χαρακτηριστικών, όπου ένας μικρών διαστάσεων πίνακας που ονομάζεται πυρήνας (kernel), εφαρμόζεται στην εισόδο που αποτελείται από ένα πίνακα (τενσορα) (tensor). Υπολογίζεται έτσι, ο ένα-προς-ένα πολλαπλασιασμός κάθε στοιχείου του πυρήνα και της εισόδου για κάθε σημείο της εισόδου και αθροίζεται για να πάρουμε την έξοδο στο αντίστοιχο σημείο, το οποίο ονομάζεται χάρτης χαρακτηριστικών (feature map). Η διαδικασία αυτή, επαναλαμβάνεται για πολλαπλούς πυρήνες για να πάρουμε ένα αυθαίρετα ορισμένο αριθμό χαρτών χαρακτηριστικών. Οι υπερπαράμετροι που ορίζουν την πράξη της συνέλιξης, είναι το μέγεθος και ο αριθμός των φίλτρων. Η συνέλιξη που περιγράψαμε όμως, δεν επιτρέπει το φίλτρο να υπερβεί τα πιο απομακρυσμένα στοιχεία της εισόδου, κάτι το οποίο έχει ως αποτέλεσμα την μείωση των διαστάσεων του χάρτη χαρακτηριστικών σε σχέση με την εισόδο. Αυτή η παράμετρος της συνέλιξης ονομάζεται padding και συχνά χρησιμοποιείται το zero padding, το οποίο αποτελεί μέθοδο επίλυσης του θέματος της διαστατικότητας. Μέσω αυτού, προστίθενται μηδενικές γραμμές και στήλες σε κάθε πλευρά του τένσορα εισόδου, ώστε να μπορεί το φίλτρο να ολισθήσει πάνω τους και να αποτρέψει την αλλαγή των διαστάσεων. Η παραπάνω τεχνική, χρησιμοποιείται εκτεταμένα στα βαθιά νευρωνικά δίκτυα για την διατήρηση των διαστάσεων και την προσθήκη περισσότερων επιπέδων. Χωρίς την χρήση της, κάθε επόμενος χάρτης χαρακτηριστικών θα είχε μικρότερες διαστάσεις μετά από την συνέλιξη. Τέλος, ένα άλλο χαρακτηριστικό των συνελικτικών δικτύων που πρέπει να οριστεί είναι η απόσταση μεταξύ δυο διαδοχικών ολισθήσεων του φίλτρου που ονομάζεται stride. Συνήθως το stride παίρνει την τιμή ένα, ωστόσο πολλές φορές μπορεί να παίρνει και μεγαλύτερες τιμές αν επιθυμούμε την μείωση της διαστατικότητας του χάρτη χαρακτηριστικών. Στο σχήμα 3.29, φαίνεται η πράξη της συνέλιξης μιας εισόδου με ένα φίλτρο μεγέθους 3×3 και stride ένα, ενώ στο σχήμα 3.30 αναδεικνύεται πως το zero padding επιτρέπει την διατήρηση των διαστάσεων της εισόδου στον χάρτη χαρακτηριστικών.



(α') Πρώτο βήμα συνέλιξης εισόδου με (β') Δεύτερο βήμα συνέλιξης εισόδου με φίλτρο 3×3

Σχήμα 3.29: Συνέλιξη τένσορα εισόδου με φίλτρο (Πηγή: [78])



Σχήμα 3.30: Εφαρμογή του zero padding (Πηγή: [78])

Συνοψίζοντας την παραπάνω λειτουργία, για μια είσοδο του συνελικτικού νευρωνικού δικτύου η οποία έχει διαστάσεις $H \times W \times C$ τα συνελικτικά επίπεδα είναι σει από παράλληλους χάρτες χαρακτηριστικών (feature maps), οι οποίοι σχηματίζονται από διαφορετικά φίλτρα (filters) ή πυρήνες (kernels) τα οποία ολισθαίνουν στην εικόνα για την ανίχνευση χαρακτηριστικών πραγματοποιώντας ένα-προς-ένα πολλαπλασιασμούς πινάκων. Για φίλτρο λοιπόν διαστάσεων $k_1 \times k_2 \times C$ ορίζουμε ως K_D το μέγεθος των επιπέδων εξόδου που θα υπολογιστούν μέσω των συνέλιξεων. Ακόμα, θεωρώντας ότι το stride είναι Z_s το padding Z_p είναι zero, για μια είσοδο στο επίπεδο συνέλιξης $H_1 \times W_1 \times D_1$ θα έχουμε μια έξοδο με διαστάσεις:

$$(H_1 \times W_1 \times D_1) = \left(\frac{H + 2Z_p - k_1}{Z_s + 1}, \frac{W + 2Z_p - k_2}{Z_s + 1}, K_D \right) \quad (3.55)$$

Η έξοδος που υπολογίστηκε στο συνελικτικό επίπεδο στην συνέχεια περνάει μέσω της συνάρτησης ενεργοποίησης, η οποία είναι μια μη γραμμική συνάρτηση όπως και στα απλά νευρωνικά δίκτυα.

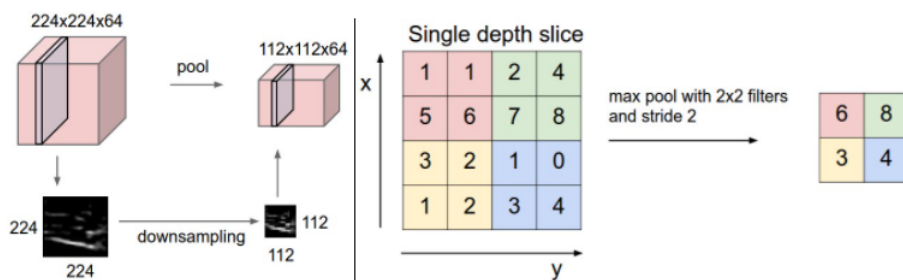
Συγκεντρωτικό επίπεδο (Pooling Layer)

Το συγκεντρωτικό επίπεδο (Pooling layer), συνήθως ακολουθεί το συνελικτικό και στόχος του είναι η μείωση των διαστάσεων των προηγούμενων επιπέδων. Έτσι, μειώνει το πλάτος και

το ύψος των πινάκων που λαμβάνει, αφήνοντας όμως το βάθος των πινάκων-χαρακτηριστικών ανεπηρέαστο. Μέσω αυτής της διαδικασίας, μειώνεται ο αριθμός των παραμέτρων ως συνέπεια της μείωσης των διαστάσεων, επιταχύνοντας την διαδικασία εκπαίδευσης του δικτύου, ενώ παράλληλα αντιμετωπίζει πιθανά προβλήματα υπερπροσαρμογής των δεδομένων. Για την πραγματοποίηση αυτού του επιπέδου, ένα παράθυρο ολισθαίνει πάνω από τα δεδομένα και τα συνδυάζει σε μία μόνο τιμή. Υποθέτοντας ένα παράθυρο (τετραγωνικού σχήματος) με μέγεθος k , D_n αριθμό φίλτρων και Z_s stride η έξοδος των συγκεντρωτικών επιπέδων για μια είσοδο $H_1 \times W_1 \times D_1$ θα είναι :

$$H_2 \times W_2 \times D_2 = \left(\frac{H - k}{Z_s + 1}, \frac{W - k}{Z_s + 1}, D_n \right) \quad (3.56)$$

Εδώ, πρέπει να επισημάνουμε ότι μπορούμε να υπολογίσουμε την τιμή στο συγκεντρωτικό επίπεδο μέσω πολλών συναρτήσεων με τις πιο συνηθισμένες να είναι το average pooling, δηλαδή ως μέσο όρο των στοιχείων που περιέχονται στο παράθυρο, max pooling, δηλαδή ως την μέγιστη τιμή που περιέχεται στο παράθυρο και τέλος, sum pooling δηλαδή ως το άθροισμα των στοιχείων του παραθύρου. Η απεικόνιση του συγκεντρωτικού επιπέδου φαίνεται στο σχήμα 3.31.



Σχήμα 3.31: Συγκεντρωτικό επίπεδο (Πηγή: [65])

Όπως παρατηρούμε η μείωση των διαστάσεων γίνεται μόνο χωρικά, χωρίς να επηρεάζεται το βάθος των δεδομένων εισόδου. Έτσι, στην αριστερή πλευρά του σχήματος παρατηρούμε ότι τα δεδομένα έχουν διαστάσεις $224 \times 224 \times 64$ και μέσω του συγκεντρωτικού επιπέδου, με φίλτρα μεγέθους 2 και stride(μέγεθος βήματος ολίσθησης φίλτρου) μεγέθους 2, οι διαστάσεις μειώνονται σε $112 \times 112 \times 64$. Στην δεξιά πλευρά του σχήματος, βλέπουμε την μείωση των διαστάσεων μέσω του max pooling, δηλαδή κρατώντας μόνο την μέγιστη τιμή του παραθύρου και ολισθαίνοντας με βήμα 2 (δηλαδή το μέγεθος του stride είναι 2) .

Πλήρως συνδεδεμένο επίπεδο (Fully Connected Layer)

Έπειτα από μια σειρά συνελικτικών και συγκεντρωτικών επιπέδων, τα πίκσελ από τις εξόδους τους δίνονται ως είσοδος σε μια συστοιχία από fully connected νευρωνικά δίκτυα, για την λήψη του τελικού αποτελέσματος. Το πλήρως συνδεδεμένο δίκτυο, όπως έχουμε περιγράψει, είναι το δίκτυο όπου κάθε νευρώνας ενός επιπέδου συνδέεται πλήρως με όλους τους υπόλοιπους του προηγούμενου. Το πλήρως συνδεδεμένο νευρωνικό δίκτυο παίρνει ως είσοδο ένα διάνυσμα στήλη, οπότε για το επιτυχές ταίριασμα των απότελεσμάτων των συνελικτικών

και συγκεντρωτικών επιπέδων που δίνουν ως αποτέλεσμα ένα πίνακα τριών διαστάσεων, χρησιμοποιείται ένας μετασχηματισμός που ονομάζεται *flattening*, ο οποίος μετατρέπει τα δεδομένα από πίνακα τριών διαστάσεων σε πίνακα στήλη μιας διάστασης, χωρίς να υπάρχει απώλεια πληροφορίας. Το πλήρως συνδεδεμένο δίκτυο, αξιοποιεί την πληροφορία από τους προηγούμενους υπολογισμούς και σε προβλήματα ταξινόμησης επιλέγει την σωστή κλάση στην οποία ανήκει το διάνυσμα εισόδου, ενώ σε προβλήματα παλινδρόμησης επιλέγει την επόμενη πραγματική τιμή με βάση τις προηγούμενες που έχει λάβει ως είσοδο. Το συγκεκριμένο κομμάτι του δικτύου απεικονίζεται στο τελευταίο βήμα του σχήματος 3.28.

3.2.14 Temporal Convolutional Networks (TCN)

Τα TCN είναι μία κατηγορία συνελικτικών νευρωνικών δικτύων, τα οποία είναι κατάλληλα για την επεξεργασία χρονοσειρών. Τα TCN έχουν δυο βασικά χαρακτηριστικά, η έξοδος τους έχει το ίδιο μήκος με αυτό που έχει η είσοδος (σε αντιστοιχία με τα LSTM), ενώ αποτρέπουν την διαρροή πληροφορίας από το μέλλον στο παρελθόν λόγω των *causal* συνελίξεων [3]. Ένα ακόμα σημαντικό χαρακτηριστικό των TCN, είναι πως έχουν την δυνατότητα να δέχονται μεγάλου μήκους ακολουθιακά δεδομένα σε αντίθεση με τα LSTM, ώστε να αξιοποιούν καλύτερα τις παρελθοντικές παρατηρήσεις για την εξαγωγή συμπερασμάτων για το μέλλον. Για την κατανόηση της λειτουργίας των TCN, αλλά και της καταλληλότητάς τους για την επεξεργασία ακολουθιακών δεδομένων, θα αναφέρουμε τα βασικά τους χαρακτηριστικά, τα οποία τα διαφοροποιούν σε σχέση με τα απλά συνελικτικά δίκτυα δηλαδή, οι *causal* συνελίξεις, οι διεσταλμένες συνελίξεις και οι υπολειμματικές συνδέσεις.

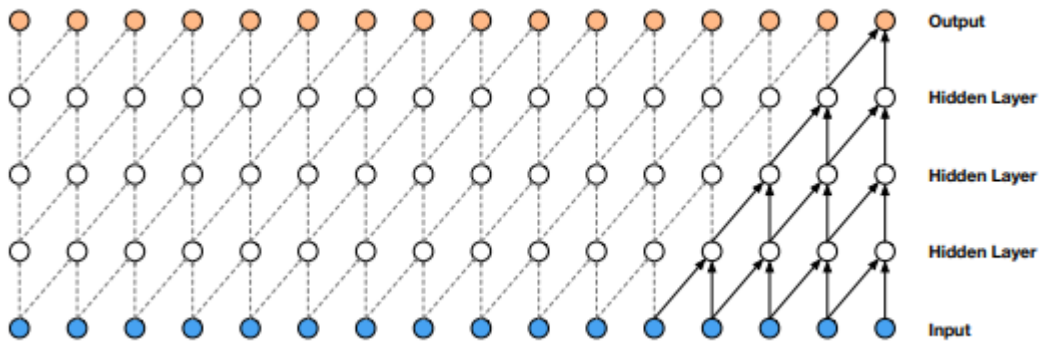
Causal Convolutions

Για την εξήγηση της ανάγκης χρήσης των Causal Convolutions, ώστε να μην υπάρχει πρόβλημα διαρροής πληροφορίας από το μέλλον στο παρελθόν θα πρέπει πρώτα να ορίσουμε την έννοια της ακολουθιακής μοντελοποίησης (*sequence modeling*). Έτσι λοιπόν, δεδομένης μιας ακολουθίας εισόδου x_0, \dots, x_T στο πρόβλημά μας θέλουμε να προβλέψουμε μια ακολουθία εξόδου y_0, \dots, y_T . Ο βασικός περιορισμός μας, είναι πως για την πρόβλεψη της εξόδου y_t για κάποια χρονική στιγμή t πρέπει να χρησιμοποιήσουμε μόνο πληροφορίες που έχουν πρότινος παρατηρηθεί x_0, \dots, x_t . Έτσι, ένα δίκτυο ακολουθιακής μοντελοποίησης είναι κάθε συνάρτηση $f : \mathcal{X}^{T+1} \rightarrow \mathcal{Y}^{T+1}$ η οποία αντιστοιχίζει :

$$\hat{y}_0, \dots, \hat{y}_T = f(x_0, \dots, x_T) \quad (3.57)$$

εάν ικανοποιεί τον *causal* περιορισμό, δηλαδή το y_t να εξαρτάται μόνο από τα x_0, \dots, x_t και όχι από τις μελλοντικές παρατηρήσεις x_{t+1}, \dots, x_T . Ο σκοπός κατά την εκπαίδευση τέτοιων δικτύων, είναι η εύρεση του δικτύου f το οποίο ελαχιστοποιεί το σφάλμα ανάμεσα στις πραγματικές και τις προβλεπόμενες τιμές $L(y_0, \dots, y_T, f(x_0, \dots, x_T))$. Τον παραπάνω περιορισμό στα συνελικτικά δίκτυα μπορούμε να τον επιτύχουμε κάνοντας τις συνελίξεις μας *causal*, ώστε κατά την χρονική στιγμή t να επιτρέπεται η συνέλιξη μιας εξόδου μόνο με στοιχεία από την χρονική αυτή στιγμή και νωρίτερα, αντί για την χρήση των παραδοσιακών συνελίξεων, οι οποίες επιτρέπουν την ολίσθηση του φίλτρου σε μελλοντικές παρατηρήσεις για την λήψη μιας

εξόδου σε προηγούμενες χρονικές στιγμές. Η αποτύπωση των causal συνδέσεων φαίνεται στο σχήμα 3.32.



Σχήμα 3.32: Causal Convolutions (Πηγή: [48])

Εκτός όμως της διαρροής πληροφορίας, αναφέραμε ότι τα TCN παράγουν έξοδο ίσου μήκους με αυτόν της εισόδου, κάτι το οποίο επιτυγχάνουν μέσω συνελίξεων μιας διάστασης (1D) 1×1 (fully-convolutional network (FCN)), όπου κάθε κρυφό επίπεδο έχει ίδιο μήκος με την είσοδο χρησιμοποιώντας zero padding μεγέθους (kernel size - 1).

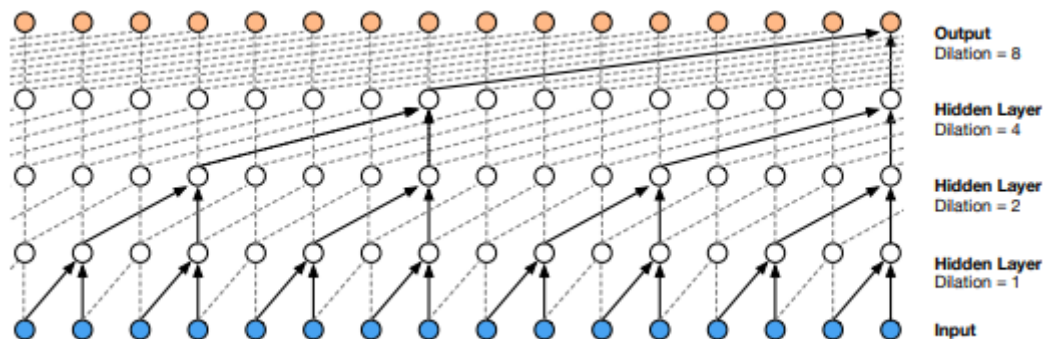
Διεσταλμένες συνελίξεις (Dilated Convolutions)

Μια απλή causal συνέλιξη, είναι δυνατό να κοιτάξει τις παρελθοντικές παρατηρήσεις με γραμμικό μέγεθος σε σχέση με το βάθος του δικτύου της. Κάτι τέτοιο όμως, δημιουργεί προβλήματα στον χειρισμό ακολουθιών, οι οποίες απαιτούν μεγάλο εύρος ιστορικών παρατηρήσεων. Το πρόβλημα αυτό, μπορεί να αντιμετωπιστεί μέσω της χρήσης διεσταλμένων συνελίξεων, οι οποίες επιτρέπουν την εκθετική αύξηση του δεκτικού τους πεδίου (receptive field). Για μια ακολουθία λοιπόν μιας διάστασης $x \in \mathbb{R}^n$ και ένα φίλτρο $f : 0, \dots, k - 1 \rightarrow \mathbb{R}$ η διεσταλμένη συνέλιξη F στο στοιχείο s της ακολουθίας ορίζεται ως :

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (3.58)$$

όπου το d είναι ο συντελεστής διαστολής και το k το μέγεθος του φίλτρου. Έτσι λοιπόν, η διαστολή (dilation) αντιστοιχεί στην εισαγωγή ενός σταθερού βήματος ανάμεσα στα διαδοχικά στοιχεία του πυρήνα του συνελικτικού επιπέδου. Όταν το $d = 1$, η διεσταλμένη συνέλιξη είναι ίδια με την κανονική. Με την αύξηση όμως της παραμέτρου d επιτρέπει στην έξοδο να επηρεάζεται από μεγαλύτερο αριθμό στοιχείων εισόδου, μεγαλώνοντας το δεκτικό της πεδίο, χωρίς όμως να αυξάνει την πολυπλοκότητα του μοντέλου. Τα TCN επομένως, έχουν δύο τρόπους για την αύξηση του δεκτικού τους πεδίου, επιλέγοντας μεγαλύτερο μέγεθος φίλτρου k ή μεγαλώνοντας την παράμετρο d , όπου το δεκτικό πεδίο που θα επηρεάζεται από την έξοδο είναι το $(k - 1)d$. Συχνά κατά την χρήση διεσταλμένων συνελίξεων αυξάνουμε εκθετικά την παράμετρο d σε σχέση με το βάθος του δικτύου, κάτι το οποίο μας διασφαλίζει ότι κάποιο φίλτρο θα επηρεάσει όσα στοιχεία βρίσκονται εντός του δεκτικού πεδίου, ενώ παράλληλα

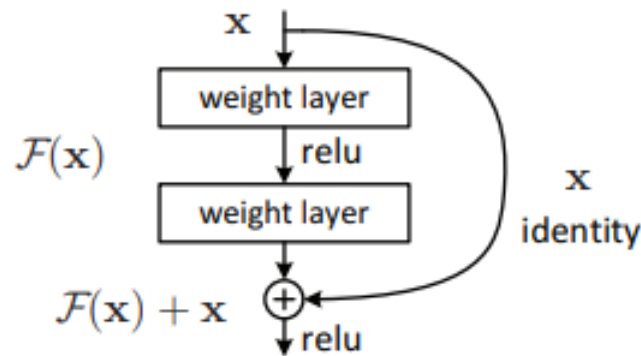
επιτρέπει την δημιουργία πολύ μεγάλων δεχτικών πεδίων[3]. Η απεικόνιση της διεσταλμένης συνελίξης φαίνεται στο σχήμα 3.33.



Σχήμα 3.33: Διεσταλμένες συνελίξεις (Dilated Convolution) (Πηγή: [48])

Υπολειμματικές συνδέσεις (Residual Connections)

Τα βαθιά νευρωνικά δίκτυα, εμφανίζουν δυσκολίες κατά την εκπαίδευσή τους λόγω της πολυπλοκότητας της αρχιτεκτονικής τους. Μελέτες [21], έχουν οδηγήσει στο συμπέρασμα ότι το βάθος του δικτύου διακατέχει καθοριστικό ρόλο στα αποτελέσματα αλλά και στην εκπαίδευση του δικτύου. Παρόλα αυτά έχει παρατηρηθεί ότι ενώ με αύξηση του βάθους του νευρωνικού δικτύου τα αποτελέσματα βελτιώνονται, έπειτα από ένα σημείο το δίκτυο αντιμετωπίζει πρόβλημα εκφυλισμού (degradation problem), δηλαδή ενώ αυξάνουμε το βάθος του η ακρίβεια του έρχεται σε κορεσμό. Ακόμα έχει παρατηρηθεί ότι το πρόβλημα αυτό δεν οφείλεται στην υπερπροσαρμογή του μοντέλου, αλλά προσθέτοντας επιπλέον επίπεδα παρατηρείται αύξηση του σφάλματος εκπαίδευσης. Για την επίλυση του προβλήματος προτείνεται από την βιβλιογραφία η χρήση υπολειμματικών συνδέσεων, οι οποίες έχει παρατηρηθεί ότι βοηθούν το δίκτυο να αποφύγει το πρόβλημα του εκφυλισμού και να συνεχίζει να βελτιώνει την ακρίβεια του με την προσθήκη επιπλέον επιπέδων. Έτσι, αν η επιθυμητή έξοδος ενός κομματιού του δικτύου μας είναι η $\mathcal{H}(x)$ θεωρούμε την αντιστοίχιση $\mathcal{F}(x) = \mathcal{H}(x) - x$. Κατά αυτόν τον τρόπο, η αρχική μας αναπαράσταση μπορεί να αποτυπωθεί ως $\mathcal{F}(x) + x$, το οποίο αποδεικνύεται ότι είναι πιο εύκολο να βελτιστοποιηθεί και να αναπαρασταθεί από το δίκτυο μας σε σχέση με το $\mathcal{H}(x)$. Η παραπάνω αποτύπωση μπορεί να δημιουργηθεί προσθέτοντας στο δίκτυο μέσω συνδέσεων παράκαμψης (shortcut connections) (δηλαδή συνδέσεων που παραλείπουν κάποια κομμάτια του δικτύου) την αρχική είσοδο. Οι shortcut connections απλά προσθέτουν την αρχική είσοδο στο τέλος κάποιων επιπέδων, βοηθώντας στη μάθηση χωρίς να προσθέτουν κάποια επιπλέον παράμετρο ή να αυξάνουν την πολυπλοκότητα του δικτύου. Ένα παράδειγμα υπολειμματικής σύνδεσης φαίνεται στο σχήμα 3.34.

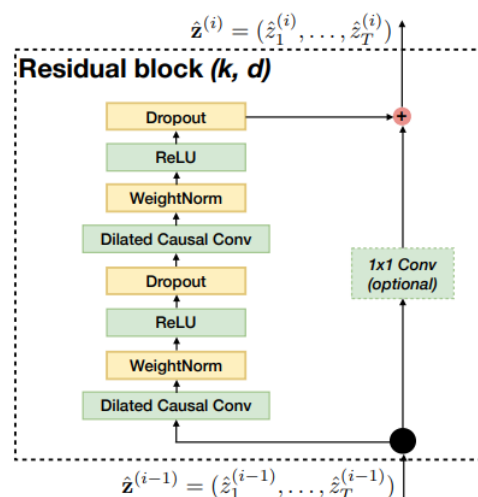


Σχήμα 3.34: Υπολειμματική συνδέση (Πηγή: [21])

Μέσω της παραπάνω μετατροπής λοιπόν, δίνουμε την δυνατότητα στο κομμάτι του δικτύου να μοντελοποιήσει την συνάρτηση $\mathcal{F}(x) + x$, η οποία μπορεί να είναι πιο εύκολη σε σύγκριση με την $\mathcal{H}(x)$, αφού αποτελεί την απόκλιση της x κατά τον όρο $\mathcal{F}(x)$. Όπως έχει παρατηρηθεί λοιπόν, η παραπάνω συσχέτιση βοηθάει την μάθηση, αφού θέτωντας $\mathcal{F}(x) = 0$ λαμβάνουμε την είσοδο στην έξοδο, οπότε μπορεί να συνεχιστεί η εκπαίδευση χωρίς να δημιουργούνται προβλήματα λόγω της μη δυνατότητας συσχέτισης της εισόδου με την έξοδο $\mathcal{H}(x)$.

TCN Residual Block

Συνδυάζοντας τις παραπάνω τεχνικές μπορούμε να δημιουργήσουμε ένα TCN Residual Block, το οποίο και αποτελεί την βασική μονάδα του TCN νευρωνικού δικτύου και φαίνεται στο σχήμα 3.35



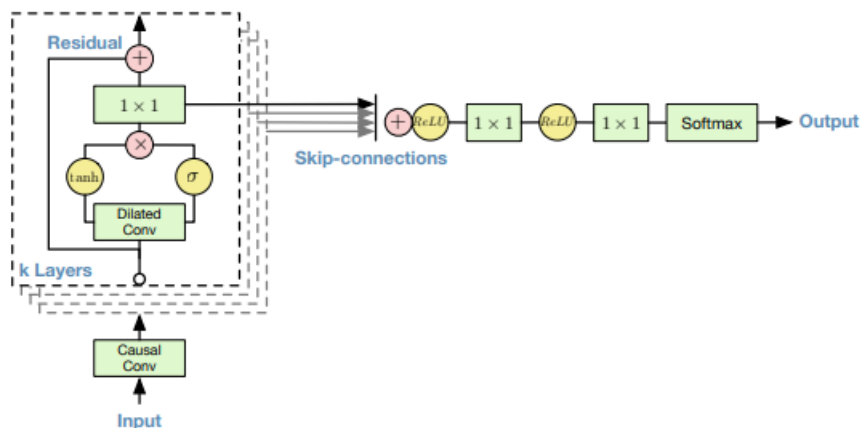
Σχήμα 3.35: TCN Residual Block (Πηγή: [21])

Όπως παρατηρούμε, το TCN block αποτελείται από διεσταλμένες causal συνελίξεις και χρησιμοποιεί για συνάρτηση ενεργοποίησης ReLU, ενώ χρησιμοποιούνται ακόμα και οι τεχνικές του Dropout και της κανονικοποίησης των βαρών για την πρόληψη της υπερπροσαρμογής και την βελτίωση της σύγκλισης των αλγορίθμων κατάβασης κλίσης αντίστοιχα. Βλέπουμε ακόμα, ότι

στην υπολειμματική σύνδεση μπορούμε να χρησιμοποιήσουμε 1×1 συνελίξεις, οι οποίες επιδρούν στο βάθος δηλαδή την τρίτη διάσταση του πίνακα εισόδου διασφαλίζοντας, ότι η είσοδος και η έξοδος θα έχουν το ίδιο μήκος. Τέλος, όπως αναφέραμε και προηγουμένως, η έξοδος επηρεάζεται από το δεκτικό πεδίο που έχει πρόσβαση ανάλογα με τις παραμέτρους που ορίζουμε κατά την σχεδίαση του δικτύου. Για την διασφάλιση επομένως μιας ορθής αρχιτεκτονικής θα πρέπει να ορίζουμε τα μήκη των δεδομένων γνωρίζοντας ότι το δεκτικό πεδίο της εισόδου υπολογίζεται ως ($receptive\ field = no.\ stacked\ blocks \times kernel\ size \times last\ dilation\ factor$), δηλαδή ως το γινόμενο του αριθμού των στοιβαγμένων μπλοκ επί το μέγεθος του πυρήνα (φίλτρο) επί τον τελευταίο συντελεστή διαστολής.

3.2.15 WaveNet

Το WaveNet είναι ένα βαθύ νευρωνικό δίκτυο, το οποίο προτάθηκε από την DeepMind της Google για την παραγωγή ηχητικών κυμάτων και την χρήση του σε text-to-speech εφαρμογές. Βασίζεται και αυτό στα συνελικτικά νευρωνικά δίκτυα και η αρχιτεκτονική του υιοθετεί στοιχεία από αυτή του TCN με την προσθήκη κάποιων επιπλέον χαρακτηριστικών για την βελτίωση των αποτελεσμάτων του. Η αρχιτεκτονική ενός WaveNet νευρωνικού δικτύου απεικονίζεται στο σχήμα 3.36.



Σχήμα 3.36: Αρχιτεκτονική WaveNet νευρωνικού δικτύου (Πηγή: [48])

Από το σχήμα βλέπουμε ότι τα WaveNet νευρωνικά δίκτυα υιοθετούν επίσης την χρήση των causal διεσταλμένων συνελίξεων, όπως και την χρήση υπολειμματικών συνδέσεων για την αποτελεσματικότερη μάθηση και την δυνατότητα της περαιτέρω αύξησης του βάθους τους. Εκτός όμως των παραπάνω χαρακτηριστικών, βλέπουμε ότι η έξοδος των διεσταλμένων συνελίξεων χωρίζεται σε δύο κομμάτια τα οποία στην συνέχεια συνδέονται μέσω του ένα-επί-ένα πολλαπλασιασμού τους. Η σύνδεση αυτή ονομάζεται **gated activation** και αποτελείται από μία *tanh* συνάρτηση ενεργοποίησης, η οποία μπορεί να θεωρηθεί ως το προς μάθηση φίλτρο σε συνδυασμό με μία *sigmoid* συνάρτηση ενεργοποίησης, η οποία ρυθμίζει την ροή πληροφορίας από το φίλτρο. Μαθηματικά μπορούμε να πούμε ότι αντιστοιχίζουμε μια συνελικτική είσοδο

x σε μια έξοδο z μέσω της εξίσωσης :

$$z = \tanh(W_f * x) \odot \sigma(W_g * x) \quad (3.59)$$

Οι ερευνητές της DeepMind χρησιμοποιούν την συγκεκριμένη συνάρτηση ενεργοποίησης αντί για μια από τις πιο συχνά χρησιμοποιούμενες μιας και παρατήρησαν ότι δίνει καλύτερα αποτελέσματα στην μοντελοποίηση των ακουστικών σημάτων, κάτι που ίσως να μπορεί να επεκταθεί και γενικά για τις χρονοσειρές. Ένα επιπλέον χαρακτηριστικό των WaveNet νευρωνικών δικτύων είναι οι συνδέσεις τους που ονομάζονται **skip connentions**. Η ιδέα πίσω από την χρήση τους είναι πως δεδομένα από την επεξεργασία της εισόδου στα αρχικά επίπεδα ενός βαθιού νευρωνικού δικτύου, μπορεί να είναι πολύ χρήσιμα για την λήψη του τελικού αποτελέσματος. Έτσι μέσω των skip connentions εκτός από το να προωθούμε την έξοδο κάθε συνελικτικού επιπέδου στο επόμενο ανά κάποια επίπεδα την αποθηκεύουμε. Στο τέλος των επιπέδων συνενώνουμε τις πληροφορίες αυτές παρέχοντας μια μίξη χαρακτηριστικών καθόλη την επεξεργασία της εισόδου η οποία βοηθάει στην καλύτερη μοντελοποίηση της.

3.2.16 Μετρικές σφαλμάτων

Δεδομένου ότι το πρόβλημα που θα κληθούμε να αντιμετωπίσουμε είναι η πρόβλεψη ενεργειακών μεγεθών, πρέπει να ορίσουμε κατάλληλους τρόπους αξιολόγησης των προβλέψεων που παράγονται από τα μοντέλα που θα υλοποιήσουμε. Αρχικά, πρέπει να επισημάνουμε ότι για την αμερόληπτη αξιολόγηση των μοντέλων θα πρέπει να χωρίσουμε τα δεδομένα μας σε δυο κατηγορίες, τα δεδομένα εκπαίδευσης και τα δεδομένα ελέγχου. Τα δεδομένα εκπαίδευσης χρησιμοποιούνται για την εκπαίδευση και την λήψη των βέλτιστων παραμέτρων των δικτύων μας, ενώ τα δεδομένα ελέγχου για την μέτρηση της αποδοτικότητάς του. Επειδή τα δεδομένα ελέγχου είναι άγνωστα για το μοντέλο μας θα δώσουν μια αξιόπιστη τιμή σφάλματος, προσομοιάζοντας μια πραγματική πρόβλεψη. Ορίζουμε λοιπόν ως σφάλμα της πρόβλεψης την διαφορά μεταξύ της πραγματικής και της προβλεπόμενης τιμής :

$$e_T = y_T - \hat{y}_T \quad (3.60)$$

όπου τα στοιχεία y_1, \dots, y_T αποτελούν τα δεδομένα εκπαίδευσης, ενώ τα δεδομένα y_{T+1}, y_{T+2}, \dots τα δεδομένα ελέγχου [26]. Σε αυτό το σημείο, θα πρέπει να επισημάνουμε ότι υπάρχουν δυο κατηγορίες που μπορούμε να κατατάξουμε τις μετρικές των σφαλμάτων που θα περιγράψουμε στην συνέχεια. Η μία κατηγορία είναι οι μετρικές που υπολογίζονται βάσει του σφάλματος e_T . Σε αυτές τις μετρικές το σφάλμα εκφράζεται στην κλίμακα των παρατηρήσεων. Έτσι, μπορούμε να συγκρίνουμε παρατηρήσεις οι οποίες δίνονται στην ίδια κλίμακα. Στην δεύτερη κατηγορία ανήκουν οι μετρικές οι οποίες εκφράζουν το σφάλμα ποσοστιαία σε εκατοστιαία κλίμακα. Για να το πετύχουν αυτό, διαιρούν το σφάλμα e_T με την πραγματική τιμή της κάθε χρονικής στιγμής y_T . Ενώ αυτές οι μέθοδοι απαλλάσσουν το σφάλμα από την εξάρτηση του από την κλίμακα των δεδομένων, επιτρέποντας συγκρίσεις μεταξύ διαφορετικών μονάδων, εμφανίζουν κάποια σημαντικά μειονεκτήματα. Πιο συγκεκριμένα λόγω της διαίρεσης του σφάλματος με την πραγματική τιμή των παρατηρήσεων το σφάλμα δεν μπορεί να οριστεί αν στην

χρονική στιγμή t έχει την τιμή $y_T = 0$ ή παίρνει ακραίες τιμές ή αν έχει τιμή κοντά στο 0. Παρακάτω θα αναφερθούμε πιο αναλυτικά στις βασικές μετρικές που χρησιμοποιούνται για την αξιολόγηση της πρόβλεψης χρονοσειρών [69], χρησιμοποιώντας τις μεταβλητές του σφάλματος και των παρατηρήσεων που ορίσαμε προηγούμενος θεωρώντας ότι το μέγεθος των δεδομένων είναι N :

- **Μέσο Σφάλμα (Mean Error)**. Υπολογίζεται ως ο προσημασμένος μέσος όρος και αποτελεί ένα μέτρο της συστηματικότητας του σφάλματος. Όταν το $ME \rightarrow 0$ τα σφάλματα είναι τυχαία και όχι συστηματικά. Ακόμα το θετικό ME δηλώνει απαισιοδοξία στις προβλέψεις μας, αφού θα είναι κατά μέσο όρο μικρότερες των πραγματικών, ενώ το αρνητικό απαισιοδοξία αντίστοιχα. Ένα αρνητικό του ME , είναι πως λόγω του ότι τα σφάλματα είναι προσημασμένα, ενδέχεται να αλληλοαναιρούνται και να δίνουν τιμή κοντά στο μηδέν χωρίς να έχουμε ακριβείς προβλέψεις. Υπολογίζεται μέσω του τύπου :

$$ME = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t) = \frac{1}{N} \sum_{t=1}^N e_t \quad (3.61)$$

- **Μέσο Απόλυτο Σφάλμα (Mean Absolute Error)**. Εκφράζει ένα μέτρο της ακρίβειας της πρόβλεψης σε σχέση με τις πραγματικές τιμές, διατηρώντας τις μονάδες μέτρησης της αρχικής χρονοσειράς όπως αναφέραμε και προηγουμένως. Πρακτικά, δηλώνει ένα μέσο μέτρο τις αστοχίας της πρόβλεψης, χωρίς να δίνεται έμφαση στην κατεύθυνση της πρόβλεψης όπως στο Μέσο Σφάλμα. Δείχνει όμως όλο το μέγεθος του σφάλματος, επομένως, ότι όσο μεγαλύτερη είναι η τιμή του δείκτη τόσο μικρότερη προκύπτει η ακρίβεια της μεθόδου που εφαρμόσαμε. Υπολογίζεται ως :

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| = \frac{1}{N} \sum_{t=1}^N |e_t| \quad (3.62)$$

Στις προβλέψεις ενεργειακών χρονοσειρών ωστόσο συνηθίζεται να χρησιμοποιείται μια παραλλαγή του MAE το **κανονικοποιημένο μέσο απόλυτο σφάλμα $NMAE$** [4] [39]. Το $NMAE$ ορίζεται ως :

$$NMAE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_{nom}} \right| = \frac{1}{N} \sum_{t=1}^N \left| \frac{e_t}{y_{nom}} \right| \quad (3.63)$$

Όπως βλέπουμε το $NMAE$ είναι το κανονικοποιημένο MAE ως προς την μέγιστη παρατήρηση ή ως προς την ονομαστική ισχύ ενός ενεργειακού πάγκου ή περιοχής y_{nom}

- **Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error)**. Αποτελεί ένα μέτρο ακρίβειας της πρόβλεψης, το οποίο δίνει μεγαλύτερο βάρος στα μεγάλα σφάλματα εξαιτίας του τετραγωνισμού των σφαλμάτων σε σχέση με τα μικρά σφάλματα. Και αυτό εξαρτάται από τις μονάδες της χρονοσειράς, δίνοντας όμως μια πλήρη εικόνα του σφάλματος όπως και το MAE λόγω του τετραγώνου, αφού δεν λαμβάνει υπόψιν την κατεύθυνση του σφάλματος. Υπολογίζεται από τον παρακάτω τύπο:

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (3.64)$$

• **Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error).**

Υπολογίζεται άμεσα, εφαρμόζοντας την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος. Έχει τις ίδιες ιδιότητες με το MSE , αλλά είναι εκφρασμένο στις μονάδες μέτρησης της αρχικής χρονοσειράς. Επιλέγεται συχνά έναντι του MSE , γιατί είναι απαλλαγμένο από το τετράγωνο της μονάδας μέτρησης:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} = \sqrt{\frac{1}{N} \sum_{t=1}^N e_t^2} \quad (3.65)$$

Αντίστοιχα τώρα με το MAE , συχνά στην πρόβλεψη ενεργειακών μεγεθών χρησιμοποιείται η κανονικοποιημένη του μορφή $NRMSE$, η οποία αποτελεί μια αποτελεσματική μετρική της ακρίβειας του μοντέλου. Όσο μικρότερη είναι η τιμή του, τόσο καλύτερη είναι και η πρόβλεψη του μοντέλου μας. Το $NRMSE$ διατυπώνεται μαθηματικά ως :

$$NRMSE = \frac{1}{y_{nom}} \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} = \frac{1}{y_{nom}} \sqrt{\frac{1}{N} \sum_{t=1}^N e_t^2} \quad (3.66)$$

• **Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Root Mean Absolute Percentage Error).**

Όπως αναφέραμε και στην αρχή της παραγράφου, ορισμένες φορές χρειάζεται ο υπολογισμός των σφαλμάτων πρόβλεψης σε ποσοστιαία μορφή. Κάτι τέτοιο θα ήταν χρήσιμο για την σύγκριση της ακρίβειας μιας μεθόδου πρόβλεψης που έχει εφαρμοστεί σε παραπάνω από μια χρονοσειρές, καθεμία από τις οποίες έχει διαφορετικό επίπεδο μέσης τιμής. Ακόμα, χρησιμεύει για την αξιολόγηση ιδιαίτερα υψηλών πραγματικών τιμών. Είναι απαλλαγμένο από τις μονάδες μέτρησης των χρονοσειρών και δεν επηρεάζεται από μετασχηματισμούς των δεδομένων. Επίσης λόγω της απόλυτης τιμής δεν εξουδετερώνει ετερόσημα σφάλματα. Εκφράζεται επί της εκατό και λαμβάνει τιμές μεγαλύτερες ή ίσες του μηδενός, με τις μικρότερες τιμές να υποδηλώνουν καλύτερη απόδοση του μοντέλου πρόβλεψης. Δίνεται από την παρακάτω εξίσωση:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100(\%) \quad (3.67)$$

• **Συντελεστής προσδιορισμού R^2 (Coefficient of determination)** Ο συντελεστής R^2 ορίζεται ως :

$$R^2 = \frac{\sum_{t=1}^N (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad (3.68)$$

Μπορούμε να θεωρήσουμε τον συντελεστή R^2 , ως το λόγο της διακύμανσης των τιμών \hat{y} ως προς της διακύμανσης των τιμών y . Πιο συγκεκριμένα, για κάθε τιμή y_t υπάρχει μια συνολική διακύμανση $(y_t - \bar{y})$, η οποία αντιπροσωπεύει το πόσο απέχει η τιμή y_t από τον μέσο όρο των τιμών y . Η συνολική διακύμανση επομένως, μπορεί να διαχωριστεί στην μη ερμηνευθείσα διακύμανση $(y_t - \hat{y}_t)$, όπου αντιπροσωπεύει πόσο απέχει η τιμή \hat{y}_t από την πραγματική τιμή y_t και την ερμηνευθείσα διακύμανση $(\hat{y}_t - \bar{y})$, η οποία αντιπροσωπεύει το πόσο απέχει η τιμή \hat{y}_t από τον μέσο όρο των τιμών y . Ο συντελεστής R^2 παίρνει τιμές από στο εύρος $[0,1]$.

Κεφάλαιο 4

Ανάλυση και προεπεξεργασία δεδομένων

Στο παρόν κεφάλαιο, θα ασχοληθούμε εκτενώς με την ανάλυση των δεδομένων που θα χρησιμοποιηθούν για την δημιουργία των μοντέλων μηχανικής μάθησης. Πιο συγκεκριμένα, θα αναφερθούμε στα χαρακτηριστικά των δεδομένων, τον τρόπο συλλογής τους, τις ιδιομορφίες τους αλλά και τον τρόπο επεξεργασίας τους, για την μετατροπή τους σε μορφή κατάλληλη για την εφαρμογή μεθόδων μηχανικής μάθησης. Τα δεδομένα που ελέγχθηκαν είναι δεδομένα αιολικής και φωτοβολταϊκής παραγωγής, που όπως αναφέραμε στην εισαγωγή είναι καίριας σημασίας για τον προγραμματισμό των μονάδων. Για την επεξεργασία των δεδομένων, χρησιμοποιήθηκε η γλώσσα προγραμματισμού (Python) και συγκεκριμένα η έκδοση 3.6.9 και για την σχεδίαση των μοντέλων βαθιάς μάθησης έγινε χρήση του API Keras της Tensorflow, μιας από τις πιο διαδεδομένες βιβλιοθήκες για εφαρμογές μηχανικής μάθησης.

4.1 Δεδομένα αιολικής παραγωγής

Αδιαμφισβήτητα, ο σημαντικότερος παράγοντας κατά την εκπαίδευση μοντέλων επιβλεπόμενης μάθησης (supervised learning) είναι η συγκέντρωση των κατάλληλων δεδομένων σε συνδυασμό με την σωστή προεπεξεργασία τους. Στην παρούσα διπλωματική, αρχικός στόχος ήταν πρόβλεψη της αιολικής παραγωγής σε επίπεδο του Ελληνικού συστήματος για 24 ώρες στο μέλλον. Για την αντιμετώπιση όμως του χρονικού ορίζοντα των 24 ωρών, συνήθης πρακτική είναι η χρήση μετεωρολογικών προβλέψεων [79] [63], αφού μόνο οι ιστορικές παρατηρήσεις της αιολικής παραγωγής δεν αρκούν, εξαιτίας των γρήγορων και απρόβλεπτων μεταβολών της ταχύτητας του ανέμου. Ενώ λοιπόν είχαμε στην διάθεσή μας ιστορικά δεδομένα αιολικής παραγωγής, μέσω του European Network of Transmission System Operators for Electricity (ENTSO-E), ο οποίος συλλέγει δεδομένα για τις Ευρωπαϊκές χώρες που αποστέλλουν οι διαχειριστές του συστήματος μεταφοράς (TSO), η εύρεση περαιτέρω μετεωρολογικών προβλέψεων για διάφορες περιοχές της Ελλάδας δεν ήταν εφικτή. Έτσι, ο αρχικός στόχος της πρόβλεψη της αιολικής παραγωγής σε επίπεδο συστήματος, μεταβλήθηκε στην πρόβλεψη της αιολικής παραγωγής για μια περιοχή της Ελλάδας, στην οποία βρίσκονται τρία αιολικά πάρκα,

όπου μας παραχωρήθηκαν τα δεδομένα της αιολικής παραγωγής και της ταχύτητας ανέμου για τις αντίστοιχες χρονικές στιγμές. Τα δεδομένα έχουν διακριτότητα 10 λεπτών και αποτελούνται από 210384 παρατηρήσεις.

Κατά την προεπεξεργασία των δεδομένων, παρατηρήθηκε αρχικά ότι από το σετ των δεδομένων, έλειπαν κάποιες τιμές πιθανόν λόγω σφαλμάτων καταγραφής, οι οποίες όμως θεωρήθηκαν ως μη στατιστικά σημαντικές, αφού αποτελούσαν το 0.03(%) των δεδομένων, οπότε επιλέχθηκε να αντικατασταθούν από την προηγούμενη διαθέσιμη παρατήρηση. Στην συνέχεια, έχοντας ως στόχο την δημιουργία προβλέψεων με χρονικό ορίζοντα 24 ωρών, υπολογίστηκε η μέση τιμή κάθε έξι παρατηρήσεων, ώστε να δημιουργηθούν ωριαίες παρατηρήσεις, με αποτέλεσμα τα δεδομένα μας να αποτελούνται τελικά από 3 σετ των 35064 παρατηρήσεων το κάθε ένα. Κάθε ένα από τα παραπάνω σετ δεδομένων περιέχει δύο στήλες, όπου καταγράφονται η παραγωγή της αιολικής ενέργειας και η αντίστοιχη ταχύτητα ανέμου σε κάθε χρονική στιγμή. Η παραγωγή της αιολικής ενέργειας δίνεται σε τιμές ανά μονάδα, ανάλογα με την ονομαστική ισχύ του κάθε πάρκου, ενώ η ταχύτητα ανέμου σε μέτρα ανά δευτερόλεπτο (m/s). Για την αποδοτική όμως λειτουργία των αλγορίθμων μηχανικής μάθησης, έχει παρατηρηθεί ότι τα δεδομένα εισόδου πρέπει να δίνονται στο διάστημα [0, 1]. Η μετατροπή των δεδομένων σε αυτό το διάστημα, όχι μόνο βοηθάει την ταχύτητα της σύγκλισης των αλγορίθμων, αλλά και προσαρμόζει όλα τα δεδομένα στην ίδια κλίμακα. Αυτό είναι σημαντικό καθώς δεδομένα με διαφορετικά χαρακτηριστικά, εκφρασμένα σε διαφορετικές κλίμακες, μπορούν να επηρεάσουν αρνητικά τους αλγορίθμους. Πιο συγκεκριμένα, δεδομένα με μεγαλύτερα διαστήματα τιμών θα έχουν μεγαλύτερη επίδραση στους αλγορίθμους συγκριτικά με άλλα με μικρότερα διαστήματα τιμών. Για την προσαρμογή των δεδομένων σε κατάλληλες τιμές ακολουθείται μια διαδικασία που ονομάζεται κανονικοποίηση (normalization), με την οποία τα δεδομένα εισόδου μετατρέπονται σε κατάλληλη κλίμακα. Οι επικρατέστερες μέθοδοι κανονικοποίησης είναι :

- **Κανονικοποίηση ελαχίστου-μεγίστου (Min-Max Scaler)** Χρησιμοποιώντας αυτόν το μετασχηματισμό, επιτυγχάνεται η μετατροπή των δεδομένων σε μια συγκεκριμένη κλίμακα που εμείς επιθυμούμε. Συνήθως αυτή είναι το [0, 1]. Για την εφαρμογή του σε ένα σετ δεδομένων, χρησιμοποιείται η μέγιστη και η ελάχιστη τιμή που παρουσιάζονται στα δεδομένα, οπότε και η καινούρια μεταβλητή υπολογίζεται ως:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- **Κανονικοποίηση z-score** Μέσω αυτής της μεθόδου, τα δεδομένα μετασχηματίζονται ώστε να έχουν τις ιδιότητες της κανονικής κατανομής, δηλαδή μηδενική μέση τιμή και τυπική απόκλιση ίση με ένα. Για μια μεταβλητή x με μέση τιμή μ και τυπική απόκλιση σ η κανονικοποίηση της μέσω του (z-score) θα είναι :

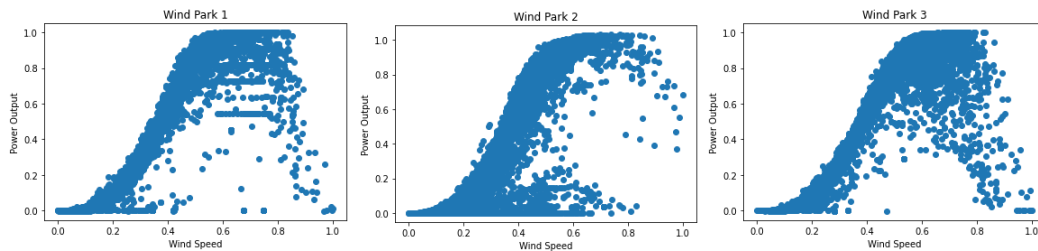
$$x_{new} = \frac{x - \mu}{\sigma}$$

Ο συγκεκριμένος μετασχηματισμός είναι χρήσιμος σε περιπτώσεις όπου τα δεδομένα μας περιέχουν ακραίες τιμές (outliers), πρέπει ωστόσο να τονίσουμε ότι μετά την εφαρμογή

του οι τιμές των δεδομένων δεν μετατρέπονται στην ίδια κλίμακα σε αντίθεση με τον μετασχηματισμό ελαχίστου-μεγίστου.

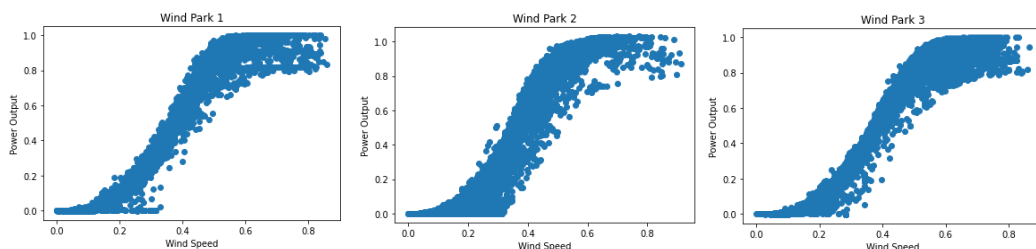
Δεδομένου, ότι οι τιμές της παραγόμενης αιολικής ενέργειας ήταν εκφρασμένες ανά μονάδα, αλλά και αυτές όπως και οι τιμές της ταχύτητας ανέμου δεν παρουσίαζαν ακραίες τιμές, επιλέχθηκε η κανονικοποίηση μόνο των τελευταίων μέσω του (Min-Max Scaler) στο διάστημα $[0, 1]$.

Βασικό χαρακτηριστικό των αιολικών πάρκων είναι η καμπύλη ισχύος τους. Για την καλύτερη κατανόηση επομένως των δεδομένων σχεδιάστηκε η καμπύλη ισχύος του κάθε πάρκου όπως φαίνεται στο σχήμα 4.1.



Σχήμα 4.1: Καμπύλες ισχύος αιολικών πάρκων

Όπως γίνεται αντιληπτό τα δεδομένα περιέχουν θόρυβο, κάτι το οποίο ενδέχεται να επηρεάσει αρνητικά τους αλγορίθμους μηχανικής μάθησης. Για την εξάλειψή του, δεδομένου του σχήματος της καμπύλης ισχύος, επιλέχθηκε η ενσωμάτωση μιας καμπύλης λογιστικής συνάρτησης $f(x) = \frac{L}{1+e^{-k(x-x_0)}}$ στην καμπύλη ισχύος μέσω της συνάρτησης της python `curve_fit`. Η `curve_fit` χρησιμοποιεί την μέθοδο των μη-γραμμικών ελαχίστων τετραγώνων για την ενσωμάτωση μιας συνάρτησης στα ζητούμενα δεδομένα. Μετά την ενσωμάτωση της καμπύλης της λογιστικής συνάρτησης στα δεδομένα, διαγράφονται οι τιμές οι οποίες έχουν μεγάλη απόσταση από αυτή, ώστε να προκύψει η καμπύλη ισχύος των αιολικών πάρκων με μειωμένο θόρυβο. Αφαιρώντας τα ζεύγη ταχύτητας ανέμου - παραγωγής ισχύος που κρίνονται ως ακραίες τιμές (outliers) οι χρονοσειρές εμφανίζονται επίσης με μειωμένο πλήθος παρατηρήσεων. Οι καμπύλες ισχύος μετά την εξάλειψη του θορύβου φαίνονται στο σχήμα 4.2.



Σχήμα 4.2: Καμπύλες ισχύος αιολικών πάρκων, χωρίς θόρυβο

Έχοντας ως στόχο την παράλληλη χρήση των δεδομένων από όλα τα αιολικά πάρκα, κρατάμε από κάθε σετ δεδομένων παρατηρήσεις ίσες με το μήκος του μικρότερου σετ δεδομένων, με αποτέλεσμα τα δεδομένα των πάρκων να αποτελούνται τελικά από 34087 παρατηρήσεις.

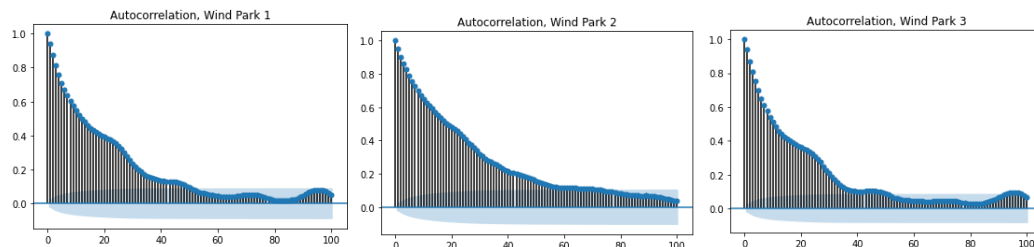
Ελέγχοντας στην συνέχεια τις χρονοσειρές της αιολικής παραγωγής βάσει των στατιστικών τους χαρακτηριστικών, πραγματοποιήθηκε αρχικά το Augmented Dickey-Fuller test για το πρώτες 1000 παρατηρήσεις τα αποτελέσματά του οποίου φαίνονται στον πίνακα 4.1.

ADF Test	Wind Park 1	Wind Park 2	Wind Park 3
Test Statistics	-5.436616	-3.916036	-6.188388e+00
p-value	0.000003	0.001920	6.208457e-08
Critical Value (1%)	-3.436979	-3.437061	-3.436966e+00
Critical Value (5%)	-2.864467	-2.864503	-2.864461e+00
Critical Value (10%)	-2.568328	-2.568348	-2.568325e+00

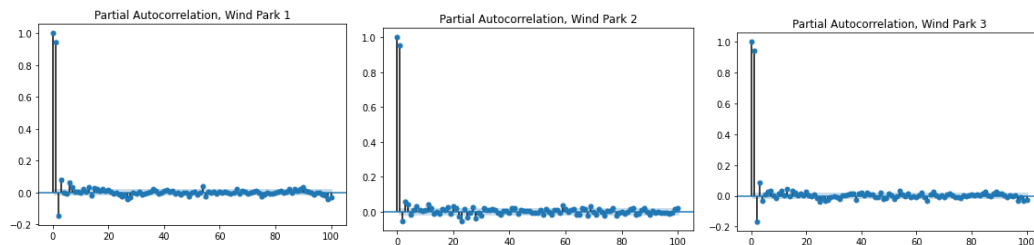
Πίνακας 4.1: Augmented Dickey-Fuller test δεδομένων αιολικής παραγωγής

Παρατηρούμε ότι το p-value των χρονοσειρών των αιολικών πάρκων είναι σε κάθε περίπτωση μικρότερο του 0.05, όπως και τα Test Statistics των κρίσιμων τιμών στα διάφορα επίπεδα εμπιστοσύνης. Οδηγούμαστε λοιπόν στο συμπέρασμα της απόρριψης της μηδικής υπόθεσης (null hypothesis), επομένως οι χρονοσειρές μπορούν να θεωρηθούν στάσιμες.

Για την πλήρη ανάλυση των δεδομένων σχεδιάζονται τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης, ώστε να δούμε κατά πόσο η τιμή σε μια χρονική στιγμή σχετίζεται με τις τιμές σε προηγούμενες παρελθοντικές στιγμές. Για τον υπολογισμό τους χρησιμοποιήθηκαν τα δεδομένα με χρονική περίοδο ενός έτους με υστέρηση μέχρι 100 παρατηρήσεις.



Σχήμα 4.3: Συνάρτηση αυτοσυσχέτισης αιολικών πάρκων



Σχήμα 4.4: Συνάρτηση μερικής αυτοσυσχέτισης αιολικών πάρκων

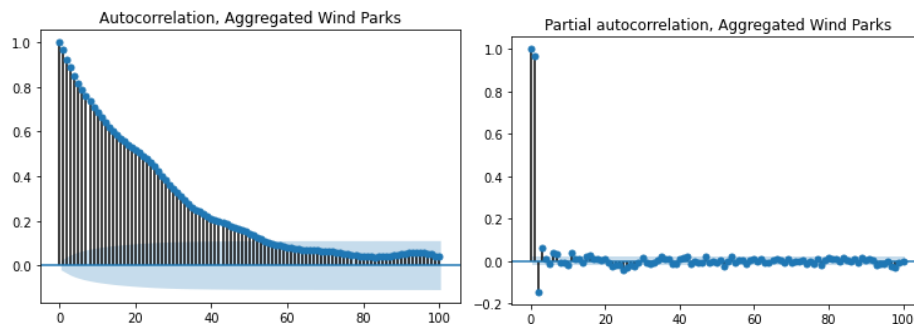
Από τα διαγράμματα, διακρίνουμε τις στατιστικά σημαντικές παρελθοντικές τιμές (τιμές εκτός των σχεδιασμένων ορίων). Βλέπουμε έτσι, πως στις χρονοσειρές της αιολικής παραγωγής

οι 40 έως 50 προηγούμενες τιμές είναι στατιστικά σημαντικές. Η παρατήρηση αυτή, μπορεί να χρησιμοποιηθεί ως βάση για την επιλογή του παρελθοντικού χρονικού ορίζοντα κατά τον χωρισμό των δεδομένων αιολικής παραγωγής, για την εκπαίδευση των μοντέλων μηχανικής μάθησης.

Δεδομένου ότι αυτή η διπλωματική δεν έχει ως στόχο την δημιουργία προβλέψεων μόνο για το επίπεδο ενός πάρκου, αλλά για την συνολική περιοχή που βρίσκονται τα τρία αιολικά πάρκα, τα δεδομένα που αναλύθηκαν θα χρειαστούν περαιτέρω επεξεργασία. Αρχικά, για την εύρεση της παραγωγής αιολικής ενέργειας στην περιοχή των πάρκων, υπολογίστηκε ο σταθμισμένος μέσος όρος των αιολικών πάρκων σε κάθε χρονική στιγμή με βάση την ονομαστική τους ισχύ. Η ίδια τεχνική εφαρμόστηκε και για την ταχύτητα ανέμου, όπου θεωρήθηκε ότι για την πρόβλεψη της παραγωγής αιολικής ενέργειας στην περιοχή, η ταχύτητα ανέμου σε κάθε πάρκο θα επηρεάζει το σύνολο, αναλογικά με την ισχύ που προσφέρει το κάθε αιολικό πάρκο. Σε αντιστοιχία με τις χρονοσειρές των αιολικών πάρκων, θα πραγματοποιηθεί ξανά ο στατιστικός έλεγχος Augmented Dickey-Fuller Test, και θα απεικονιστούν διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης, αναμένοντας παρόμοια συμπεριφορά με τις ξεχωριστές χρονοσειρές των αιολικών πάρκων. Η παραπάνω υπόθεση επαυθευτεται από τον πίνακα 4.2 και το σχήμα 4.5.

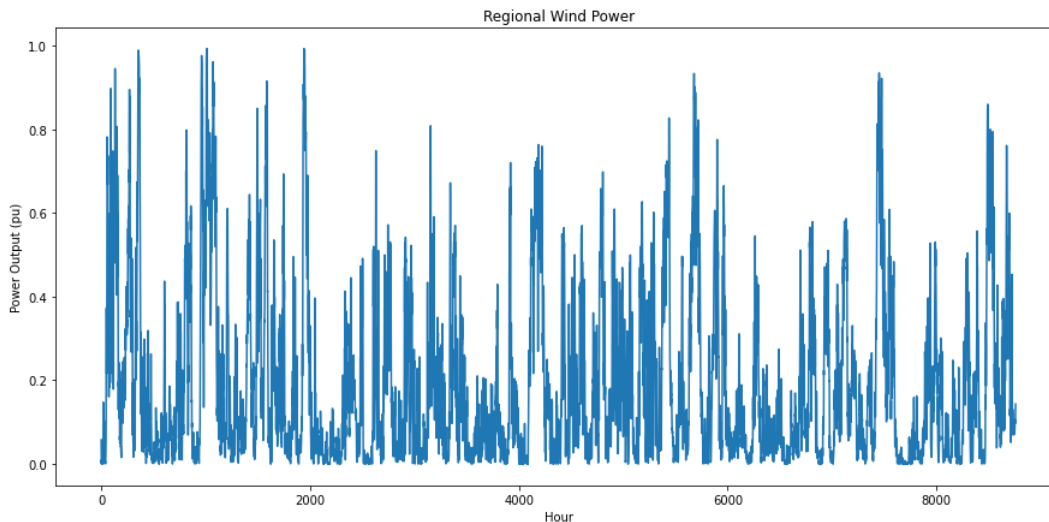
ADF Test	Aggregated Wind Parks
Test Statistics	-4.443562
p-value	0.000248
Critical Value (1%)	-3.436999
Critical Value (5%)	-2.864476
Critical Value (10%)	-2.568333

Πίνακας 4.2: Augmented Dickey-Fuller test συνολικών δεδομένων αιολικής παραγωγής



Σχήμα 4.5: Συνάρτηση αυτοσυσχέτισης και μερικής αυτοσυσχέτισης αιολικών πάρκων

Ακόμα, η συνολική παραγωγή αιολικής ενέργειας για διάστημα ενός χρόνου απεικονίζεται στο σχήμα 4.6.



Σχήμα 4.6: Χρονοσειρά συνολικής παραγωγής αιολικής ενέργειας

Φαίνεται λοιπόν, πως η χρονοσειρά της συνολικής αιολικής παραγωγής παρουσιάζει συχνές και απότομες μεταβολές καθώς και μεγάλη τυχαιότητα, γεγονός που την καθιστά δύσκολα προβλέψιμη. Ελέγχοντας όμως την συσχέτιση της ταχύτητας του ανέμου με την παραγωγή αιολικής ενέργειας, βρίσκουμε πως είναι 0.965. Όπως ήταν αναμενόμενο, η παραγωγή αιολικής ενέργειας εξαρτάται άμεσα από την ταχύτητα ανέμου. Επομένως, μπορούμε να συμπεράνουμε πως για την ακριβέστερη πρόβλεψη της αιολικής παραγωγής, ειδικά για μεγάλους χρονικούς ορίζοντες (πρόβλεψη 24 ωρών στο μέλλον) η ύπαρξη πρόβλεψων ταχύτητας ανέμου αποτελεί καθοριστικό παράγοντα. Η εύρεση όμως τέτοιου είδους προβλέψεων είναι μια δύσκολη διαδικασία και ενώ έγινε προσπάθεια για την πρόσβαση σε τέτοιου είδους δεδομένα, δεν ήταν εφικτή η απόκτηση τους. Για αυτό το λόγο, επιλέχθηκε η χρήση της ταχύτητας του ανέμου από τα αιολικά πάρκα ως πρόβλεψη. Πιο συγκεκριμένα, οι μετεωρολογικές υπηρεσίες δημοσιεύουν προβλέψεις ταχύτητας ανέμου με διακριτότητα 3 ωρών. Σε αντιστοιχία λοιπόν, επιλέχθηκε ως είσοδος των αλγορίθμων μηχανικής μάθησης η ταχύτητα ανέμου της επόμενης ημέρας (δηλαδή αυτής που επιθυμούμε την πρόβλεψη) με διακριτότητα 3 ωρών. Έτσι, οι αλγόριθμοι θα λαμβάνουν ως είσοδο τις ιστορικές παρατηρήσεις της παραγωγής αιολικής ενέργειας (το μήκος των οποίων θα καθοριστεί στην συνέχεια) μαζί με ένα διάγραμμα 8 τιμών, που θα περιέχει την πρόβλεψη της ταχύτητας ανέμου της ημέρας που επιθυμούμε την πρόβλεψη της παραγωγής αιολικής ενέργειας. Ωστόσο, η 'πρόβλεψη' που έχουμε στην διάθεσή μας, αν και μεγαλύτερης διακριτότητας από αυτή της μιας ώρας, θα είναι πολύ πιο ακριβής σε σχέση με μια πρόβλεψη που παράχθηκε από κάποια μετεωρολογική υπηρεσία. Αυτό συμβαίνει διότι, στην πραγματικότητα αποτελεί την καταγραφή της ταχύτητας ανέμου στο κάθε αιολικό πάρκο, για αυτό και επιλέχθηκε η ασαφοποίηση της για την παραγωγή πιο ρεαλιστικών αποτελεσμάτων, προσομοιώνοντας ουσιαστικά μια πρόβλεψη από κάποια μετεωρολογική υπηρεσία. Η προαναφερθείσα ασαφοποίηση, έγινε βάσει των καμπυλών ισχύος του σχήματος 4.1. Πιο αναλυτικά, παρατηρώντας τις καμπύλες ισχύος φαίνεται πως όλα τα αιολικά πάρκα για ταχύτητες ανέμου μεγαλύτερες του 60(%) της μέγιστης τιμής τους παράγουν την μέγιστη δυνατή ισχύ τους.

Αντιθέτως, για ταχύτητες μικρότερες του 20(%) παράγουν σχεδόν μηδενική ισχύ. Έτσι, για την ασαφοποίηση δόθηκε μεγαλύτερη βαρύτητα στις ενδιάμεσες τιμές της ταχύτητας ανέμου σε σχέση με τις μέγιστες και τις ελάχιστες. Οι τιμές που χρησιμοποιήθηκαν φαίνονται στον πίνακα 4.3.

Πραγματική ταχύτητα ανέμου (α.μ)	Ασαφής ταχύτητα ανέμου (α.μ)
≤ 0.2	0
(0.2, 0.3]	0.25
(0.3, 0.4]	0.5
(0.4, 0.5]	0.75
(0.5, 1]	1

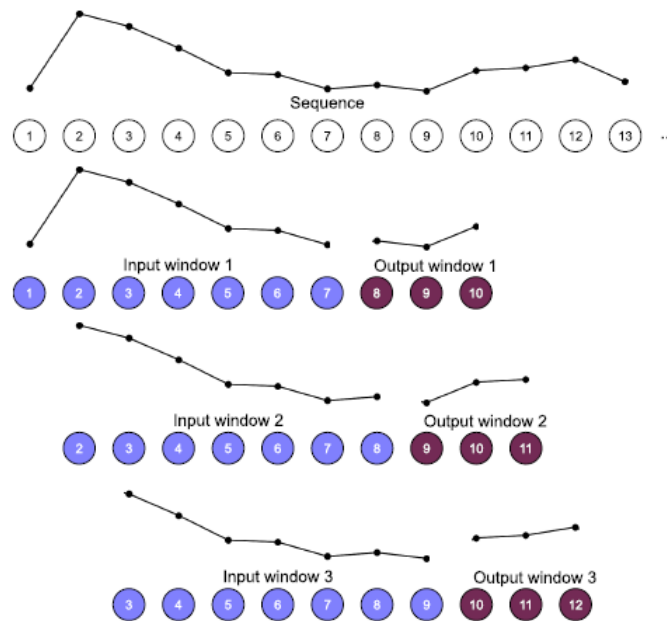
Πίνακας 4.3: Ασαφοποίηση ταχύτητας ανέμου

Στην συνέχεια, αντικαθιστούμε τις τιμές του πίνακα 4.3, στις τιμές της συνολικής ταχύτητας ανέμου για την περιοχή που υπολογίσαμε. Με αυτόν τον τρόπο οι ταχύτητες ανέμου δεν παρουσιάζονται με την πλήρη ακρίβεια με την οποία καταγράφηκαν, αλλά μέσω των τιμών 0, 0.25, 0.5, 0.75, 1, οι οποίες αντιστοιχούν ουσιαστικά σε μια πρόβλεψη, ότι ο άνεμος σε μια περιοχή θα έχει μηδενική, λιγη, μέτρια, μεγάλη ή και την μέγιστη ισχύ του, που θα μπορεί να ληφθεί σε πραγματικές συνθήκες μέσω μετεωρολογικών προβλέψεων.

Το τελευταίο στάδιο της ανάλυσης των δεδομένων, αναφέρεται στον χωρισμό των δεδομένων με κατάλληλο τρόπο για την εφαρμογή των αλγορίθμων μηχανικής μάθησης. Ο χωρισμός των δεδομένων για την πρόβλεψη πολλαπλών χρονικών στιγμών στο μέλλον είναι ένα πρόβλημα που μπορεί να αντιμετωπιστεί με διαφορετικές προσεγγίσεις [67]. Μία από αυτές είναι η αναδρομική, στην οποία το μοντέλο προβλέπει μια τιμή στο μέλλον την οποία στην συνέχεια χρησιμοποιεί ως είσοδο για την πρόβλεψη της επόμενης. Άλλη επιλογή είναι η δημιουργία ενός μοντέλου για κάθε ξεχωριστό χρονικό ορίζοντα πρόβλεψης. Μια από τις πιο αποδοτικές στρατηγικές όμως, είναι αυτή των πολλαπλών εισόδων-πολλαπλών εξόδων (Multi-Input Multi-Output (MIMO)), όπου η έξοδος αποτελεί την συνολική πρόβλεψη όλου του ορίζοντα χρησιμοποιώντας ένα μοντέλο. Η συγκεκριμένη μέθοδος φέρει αρκετά πλεονεκτήματα, όπως αυτό της μοντελοποίησης των εξαρτήσεων μεταξύ των προβλεπόμενων τιμών, αφού στην έξοδο προβλέπεται το σύνολο του μελλοντικού χρονικού ορίζοντα. Ακόμα, αυτή η μέθοδος είναι απαλλαγμένη από την συσσώρευση των σφαλμάτων πρόβλεψης που θα εμφανιζόταν στην αναδρομική στρατηγική [35].

Ακολουθώντας την παραπάνω στρατηγική, πρέπει να δημιουργηθεί ένα κινούμενο παράθυρο το οποίο θα ολισθαίνει στις παρατηρήσεις της χρονοσειράς, δημιουργώντας ζεύγη εισόδου-εξόδου που θα παρουσιαστούν στους αλγορίθμους μας, μετατρέποντας το πρόβλημα της πρόβλεψης σε ένα πρόβλημα επιβλεπόμενης μάθησης. Όλα τα μοντέλα βαθιάς μάθησης που θα παρουσιαστούν δέχονται ως είσοδο ένα συγκεκριμένου μήκους διάνυσμα και έχουν ως έξοδο ένα dense layer, που αποτελείται από 24 νευρώνες όσες δηλαδή και οι μελλοντικές παρατηρήσεις που επιθυμούμε να προβλέψει το νευρωνικό δίκτυο. Η διαδικασία χωρισμού των

παρατηρήσεων μιας χρονοσειράς σε ζεύγη εισόδου-εξόδου φαίνεται στο σχήμα 4.7.



Σχήμα 4.7: Χωρισμός χρονοσειράς σε ζεύγη εισόδου-εξόδου (Πηγή : [35])

Όπως παρατηρείται, ένα παράθυρο σταθερού μήκους ολισθαίνει κατά μήκος της χρονοσειράς, δημιουργώντας τα ζεύγη εισόδου-εξόδου. Εδώ, πρέπει να τονίσουμε ότι για την δημιουργία των δεδομένων εκπαίδευσης, το παράθυρο που φαίνεται και στο σχήμα 4.7 ολισθαίνει με βήμα ένα, για την δημιουργία περισσότερων ζευγαριών εισόδου-εξόδου, μιας και τα βαθιά νευρωνικά δίκτυα απαιτούν πληθώρα δεδομένων για την εκπαίδευσή τους. Ακόμα, η χρονοσειρά αιολικής παραγωγής δεν παρουσιάζει κάποιο ημερήσιο ή εβδομαδιαίο μοτίβο (σε αντίθεση με άλλες χρονοσειρές όπως της παραγόμενης ενέργειας από φωτοβολταϊκά), οπότε επιτρέπει αυτόν τον χωρισμό. Αντιθέτως, για την δημιουργία των ζευγών εισόδου-εξόδου για τα δεδομένα ελέγχου, το παράθυρο ολισθαίνει με βήμα 24, ώστε κάθε φορά να προβλέπεται η παραγωγή ενέργειας για την επόμενη ημέρα. Ακόμα, πρέπει να διευκρινιστεί πως ενώ το μήκος της εξόδου είναι σταθερό, το μήκος των δεδομένων εισόδου είναι μια παράμετρος που πρέπει να αποφασιστεί. Ο καθορισμός του εξαρτάται από πολλούς παράγοντες, όπως η φύση των δεδομένων, το μοντέλο που θα χρησιμοποιηθεί και ο ορίζοντας πρόβλεψης. Στην παρούσα εργασία για την επιλογή του ορίζοντα των δεδομένων εισόδου δοκιμάστηκαν διάφορες τιμές, με γνώμονα το διάγραμμα αυτοσυσχέτισης, αλλά και το ότι το LSTM νευρωνικό δίκτυο έχει περιορισμένη ικανότητα χειρισμού μεγάλου μήκους δεδομένων, ενώ αντιθέτως τα TCN και WaveNet νευρωνικά δίκτυα έχουν δημιουργηθεί για το χειρισμό μεγάλων ακολουθιών όπως θα περιγραφεί και πιο αναλυτικά στην συνέχεια.

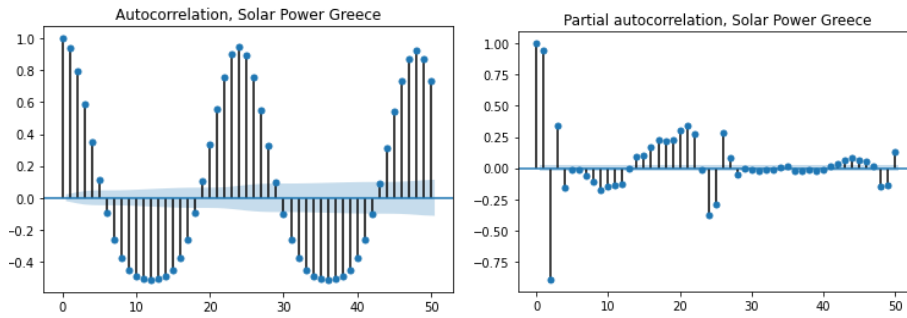
4.2 Δεδομένα φωτοβολταϊκής παραγωγής

Εκτός από την αιολική παραγωγή και η φωτοβολταϊκή παρουσιάζει μεγάλο ενδιαφέρον, με την πρόβλεψή της να είναι σημαντικής αξίας. Σε αντιστοιχία με την αιολική παραγωγή, στόχος είναι η πρόβλεψή της σε επίπεδο συστήματος της Ελλάδας για 24 ώρες στο μέλλον. Η πρόβλεψη της φωτοβολταϊκής παραγωγής κρίνεται ευκολότερη σε σύγκριση με την αιολική, αφού η χρονοσειρά της παρουσιάζει ημερήσια μοτιβά, τα οποία επηρεάζονται κυρίως από την ηλιακή ακτινοβολία. Έτσι, είναι εφικτή η δημιουργία μοντέλων πρόβλεψης της, με χρήση μόνο ιστορικών δεδομένων, αν και η χρήση επιπλέον δεδομένων καιρού θα βοηθούσε σημαντικά στην βελτίωση της ακρίβειας της. Τα δεδομένα που έχουμε στην διάθεσή μας, είναι ιστορικές παρατηρήσεις διάρκειας 5 ετών (2015-2019) για την παραγωγή ηλιακής ενέργειας από φωτοβολταϊκά σε επίπεδο συστήματος της Ελλάδας, τα οποία ήταν διαθέσιμα μέσω του (ENTSO-E). Ενώ οι τιμές ήταν καταγεγραμμένες με την πραγματική παραγωγή ενέργειας της κάθε ώρας σε μονάδες ισχύος, τα δεδομένα κάθε χρόνου κανονικοποιήθηκαν, διαιρώντας τα με την εγκατεστημένη ισχύ φωτοβολταϊκών της Ελλάδας την αντίστοιχη χρονιά. Για την αρτιότερη κατανόηση των δεδομένων φωτοβολταϊκής παραγωγής θα πραγματοποιήσουμε αρχικά τον Augmented Dickey-Fuller έλεγχο, ώστε να διαπιστώσουμε αν οι παρατηρήσεις της φωτοβολταϊκής παραγωγής διατηρούν τα στατιστικά τους χαρακτηριστικά κατά την διάρκεια του χρόνου. Για τον συγκεκριμένο έλεγχο, έγινε χρήση των 1000 πρώτων παρατηρήσεων και τα αποτελέσματά του παρουσιάζονται στο πίνακα 4.4.

ADF Test	Solar Power Greece
Test Statistics	-2.986580
p-value	0.036167
Critical Value (1%)	-3.437054
Critical Value (5%)	-2.864500
Critical Value (10%)	-2.568346

Πίνακας 4.4: Augmented Dickey-Fuller test δεδομένων φωτοβολταϊκής παραγωγής

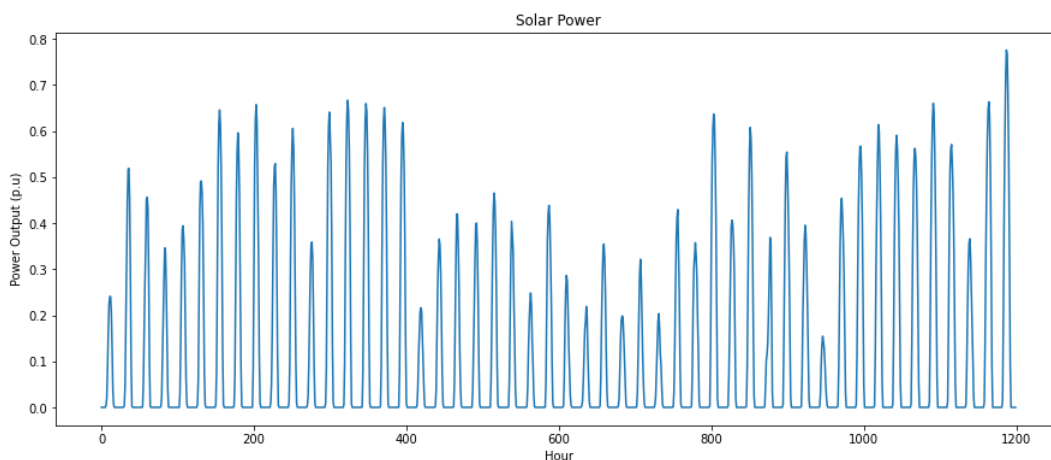
Σύμφωνα με τον έλεγχο Augmented Dickey-Fuller, παρατηρούμε ότι η χρονοσειρά της φωτοβολταϊκής παραγωγής διατηρεί τα χαρακτηριστικά της στο χρόνο και μπορεί να θεωρηθεί στάσιμη, βάσει του (p-value < 0.05) αλλά και του test statistic, σε σύγκριση με τις κρίσιμες τιμές για τα διάφορα διαστήματα εμπιστοσύνης. Στην συνέχεια, παρουσιάζονται και τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης (σχήμα 4.8), ώστε να διαπιστωθεί η σημαντικότητα των παρελθοντικών χρονικών παρατηρήσεων για την δημιουργία των προβλέψεων. Για την σχεδιάσή τους, χρησιμοποιήθηκαν τα δεδομένα φωτοβολταϊκής παραγωγής ενός έτους με υστέρηση 50 παρατηρήσεων.



(α') Αυτοσυσχέτιση χρονοσειράς παραγωγής φωτοβολταϊκής ενέργειας (β') Μερική αυτοσυσχέτιση χρονοσειράς παραγωγής φωτοβολταϊκής ενέργειας

Σχήμα 4.8: Συνάρτηση αυτοσυσχέτισης και μερικής αυτοσυσχέτισης παραγωγής φωτοβολταϊκής ενέργειας

Τα διαγράμματα φανερώνουν ότι οι τιμές της χρονοσειράς φωτοβολταϊκής παραγωγής εμφανίζουν σημαντική αυτοσυσχέτιση με τις αντίστοιχες προηγούμενες 24, κάτι το οποίο ήταν αναμενόμενο εξαιτίας της μορφής της. Έτσι, για την επιλογή του παρελθοντικού χρονικού ορίζοντα, θα χρησιμοποιηθούν οι παρελθοντικές 24 παρατηρήσεις της παραγωγής ενέργειας όπως και πολλαπλάσια τους. Σε αυτό το σημείο θα πρέπει να διευκρινίσουμε ακόμα, ότι λόγω της σημαντικότητας του ημερήσιου χωρισμού των δεδομένων, χρησιμοποιήθηκαν δύο τρόποι για την δημιουργία των ζευγών εισόδου-εξόδου για την εκπαίδευση των νευρωνικών δικτύων. Αρχικά, σε αντίθεση με το σχήμα 4.7, η ολίσθηση του κινούμενου παραθύρου ορίστηκε στις 24 παρατηρήσεις για τα δεδομένα εκπαίδευσης αλλά και ελέγχου (βασισμένοι στην σημασία του μοτίβου που εμφανίζεται ανά 24 ώρες). Η μέθοδος αυτή όμως, τελικά απέδωσε χειρότερα αποτελέσματα από τον ορισμό της ολίσθησης του παραθύρου στην μια παρατήρηση, γεγονός που θεωρείται ότι οφείλεται στην σημασία της ύπαρξης πληθώρας δεδομένων εκπαίδευσης για την άρτια εκπαίδευση νευρωνικών δικτύων βαθιάς μάθησης. Στο σχήμα 4.9 απεικονίζεται η μορφή της χρονοσειράς φωτοβολταϊκής παραγωγής για χρονικό ορίζοντα 50 ημερών της χειμερινής περιόδου.



Σχήμα 4.9: Χρονοσειρά παραγωγής φωτοβολταϊκής ενέργειας

Μπορεί να διακρίνει κανείς πως η χρονοσειρά εμφανίζει παρόμοια ημερήσια συμπεριφορά, με κύρια διαφοροποίηση το σημείο της μέγιστης παραγωγής ενέργειας, κάτι το οποίο εξαρτάται από τα καιρικά φαινόμενα (και συγκεκριμένα κυρίως από την ηλιοφάνεια) που επικρατούν στο σύνολο της Ελληνικής επικράτειας. Για την ορθότητα των προβλέψεων λοιπόν, θεωρείται χρήσιμη η ύπαρξη καιρικών προβλέψεων, δεδομένου ότι ημέρες μικρής παραγωγής μετά από μια σειρά ημερών με μεγάλη παραγωγή ηλιακής ενέργειας η αντίστροφα, δεν είναι εφικτό να προβλεφθούν ή να μοντελοποιηθούν, με μόνο δεδομένο τις ιστορικές παρατηρήσεις της φωτοβολταϊκής παραγωγής.

Κεφάλαιο 5

Εφαρμογή μεθόδων πρόβλεψης και αποτελέσματα

Σε αυτό το κεφάλαιο θα παρουσιαστούν τα αποτελέσματα των αλγορίθμων μηχανικής μάθησης για την πρόβλεψη της αιολικής και φωτοβολταϊκής παραγωγής. Πιο συγκεκριμένα, θα χρησιμοποιηθούν το LSTM νευρωνικό δίκτυο, η αρχιτεκτονική Encoder-Decoder LSTM και τα TCN και WaveNet νευρωνικά δίκτυα. Ο χρονικός ορίζοντας που εξετάστηκε είναι 24 ώρες στο μέλλον. Ωστόσο, οι προβλέψεις για την επόμενη ημέρα (24 ώρες), επαναλήφθηκαν για χρονικό διάστημα ενός μήνα (30 μέρες), ώστε να ελέγξουμε αν τα νευρωνικά δίκτυα έχουν την δυνατότητα προσαρμογής σε διαφορετικές εισόδους, αλλά και να έχουμε ένα αξιόπιστο πλήθος δεδομένων για τον υπολογισμό των μετρικών σφάλματος. Ακόμα σε αυτό το σημείο, πρέπει να επισημάνουμε ότι στην πραγματικότητα οι προβλέψεις της επόμενης ημέρας πρέπει να είναι διαθέσιμες στους χειριστές του συστήματος μεταφοράς (TSO) στις 12 μ.μ της προηγούμενης μέρας. Για αυτό, τα μοντέλα μας πρέπει να έχουν την δυνατότητα πρόβλεψης της παραγωγής ηλεκτρικής ενέργειας (από τα αιολικά η φωτοβολταϊκά πάρκα) με δεδομένα έως εκείνη την ώρα. Επειδή όμως κάτι τέτοιο αυξάνει σημαντικά την δυσκολία πρόβλεψης (μιας και ουσιαστικά ο χρονικός ορίζοντας πρόβλεψης μετατρέπεται σε 36 ώρες από τις οποίες μας ενδιαφέρουν οι ώρες 12 έως και 36), σε συνδυασμό με την μη δυνατότητα πρόσβασης μας σε επιπλέον καιρικά δεδομένα, αποφασίστηκε η δοκιμή των μοντέλων μας, κάνοντας πλήρως γνωστές τις 24 τελευταίες ώρες της προηγούμενης ημέρας στους αλγορίθμους.

5.1 Πρόβλεψη αιολικής παραγωγής

5.1.1 Εφαρμογή LSTM νευρωνικού δικτύου

Η πρώτη μέθοδος πρόβλεψης που εφαρμόστηκε ήταν μέσω των νευρωνικών δικτύων LSTM. Για την εκπαίδευσή τους, τα δεδομένα αιολικής παραγωγής χωρίστηκαν στις εξής τρεις υποομάδες, τα δεδομένα εκπαίδευσης (training data), τα δεδομένα επαλήθευσης (validation data) και τα δεδομένα ελέγχου (test data). Ως δεδομένα ελέγχου, ορίσαμε όπως έχει αναφερθεί τις τελευταίες 30 ημέρες του dataset (720 παρατηρήσεις), ενώ τα υπόλοιπα αποτελούν

τα δεδομένα εκπαίδευσης και επαλήθευσης. Τα τελευταία υπολογίζονται ως το 10(%) των δεδομένων εκπαίδευσης, τα οποία δεν παίρνουν μέρος στην διαδικασία εκπαίδευσης αλλά χρησιμεύουν στο έλεγχο του αν το νευρωνικό μας δίκτυο έχει υπερεκπαιδευθεί. Το επόμενο βήμα για την εκπαίδευση των μοντέλων, μετά των χωρισμό των δεδομένων, είναι η δημιουργία ζευγών εισόδου-εξόδου. Ενώ λοιπόν η έξοδος μας ορίζεται ως οι 24 επόμενες ώρες από την τελευταία ημέρα των δεδομένων εισόδου, το διάνυσμα εισόδου έχει μεταβλητό μήκος. Αρχικά, πρέπει να τονίσουμε ότι στα μοντέλα που δημιουργήθηκαν η είσοδος αποτελείται από το διάνυσμα των ιστορικών παρατηρήσεων της παραγωγής αιολικής ενέργειας, το οποίο και έχει μεταβλητό μήκος, αλλά και από το διάνυσμα πρόβλεψης της ταχύτητας ανέμου για την ημέρα που θέλουμε να προβλέψουμε, το οποίο αποτελείται από 8 παρατηρήσεις, μια πρόβλεψη δηλαδή για κάθε 3 ώρες της επόμενης ημέρας. Σχετικά με το μεταβλητό μήκος του διανύσματος των παρατηρήσεων θεωρείται, ότι πρέπει να είναι τουλάχιστον ίσο με το μήκος των δεδομένων που θέλουμε να προβλέψουμε, αλλά η τιμή του εξαρτάται και από το κάθε μοντέλο πρόβλεψης. Συγκεκριμένα, για το νευρωνικό δίκτυο LSTM δοκιμάστηκαν οι ορίζοντες 24, 48 και 72 ωρών με την καταλληλότερη επιλογή να είναι ο ορίζοντας των 48 ωρών.

Το νευρωνικό δίκτυο LSTM, έχει αρκετές παραμέτρους που πρέπει να προσδιοριστούν πριν γίνει η εκπαίδευση του. Κάποιες από αυτές είναι, το πόσα επίπεδα LSTM θα χρησιμοποιήσουμε, πόσες μονάδες θα έχει το κάθε επίπεδο, η συνάρτηση ενεργοποίησης του, αλλά και το αν το νευρωνικό δίκτυο θα επιστρέφει μια τιμή για κάθε χρονική στιγμή (return_sequence = True) ή αν θα επιστρέφει μόνο την τελική τιμή της τελευταίας χρονικής στιγμής (return_sequence = False). Ακόμα, για την εκπαίδευση του μοντέλου θα πρέπει να οριστεί ο αλγόριθμος βελτιστοποίησης αλλά και το batch_size, δηλαδή μετά από κάθε πόσα πακέτα δεδομένων μορφής (batch size, timesteps, features) θα ανανεώνονται οι κλίσεις (gradients). Για τον καθορισμό των υπερπαραμέτρων του μοντέλου έγιναν δοκιμές σε ένα τμήμα της χρονοσειράς (συγκεκριμένα το πρώτο έτος, 8760 παρατηρήσεις) για την επιλογή των βέλτιστων. Οι παράμετροι που ελέγχθηκαν φαίνονται στον πίνακα 5.1.

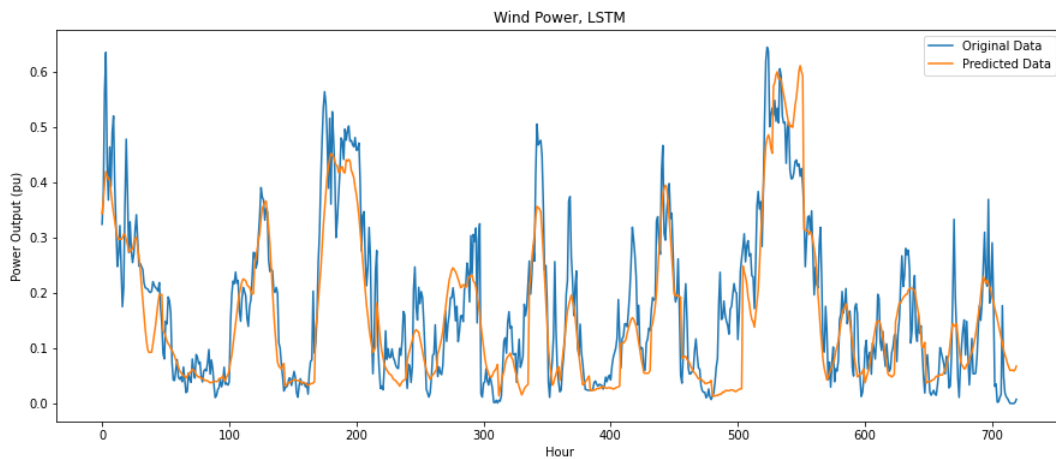
Υπερπαραμέτρος	Εναλλακτικές	Βέλτιστη
Αριθμός επιπέδων LSTM	[1,2]	2
LSTM Units 1 ^{ου} επιπέδου	[100, 150, 200]	200
LSTM Units 2 ^{ου} επιπέδου	[50, 100, 150]	100
Συνάρτηση ενεργοποίησης	[tanh, ReLU]	ReLU
Αλγόριθμος βελτιστοποίησης	[RMSprop, Adam]	Adam
Batch size	[64, 128, 256]	256

Πίνακας 5.1: Υπερπαραμέτροι LSTM

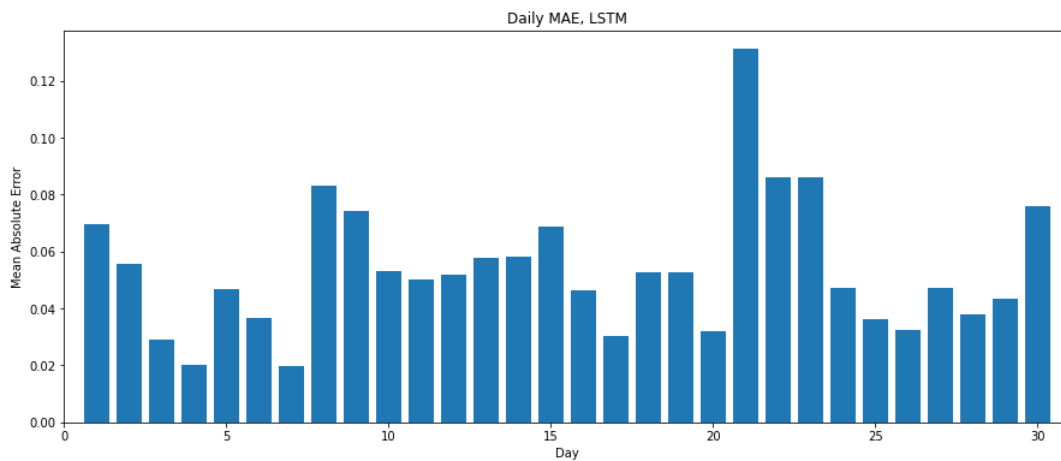
Σύμφωνα με τον πίνακα 5.1, καθορίζουμε την τελική αρχιτεκτονική του LSTM νευρωνικού δικτύου, η οποία αποτελείται από δυο Bidirectional LSTM νευρωνικά δίκτυα, δύο Dropout επίπεδα (με πιθανότητα 0.2) για την πρόληψη της υπερπροσαρμογής ενδιάμεσα τους και τέλος ένα πλήρες συνδεδεμένο επίπεδο με 24 νευρώνες.

Για την εκπαίδευση του μοντέλου ορίζουμε ακόμα τις εποχές εκπαίδευσης σε 50, θεωρώντας

ότι είναι αρκετές για την σύγκλιση του αλγορίθμου, προσθέτοντας επίσης και την μέθοδο ομαλοποίησης (regularization) Early stopping, ώστε να διακοπεί η εκπαίδευση αν δεν υπάρχει μείωση του σφάλματος για παραπάνω από 5 εποχές εκπαίδευσης. Για την εκπαίδευση ακόμα ορίζουμε ως συνάρτηση σφάλματος την MAE(η οποία θα χρησιμοποιηθεί και σε όλα τα μοντέλα που θα παρουσιαστούν), η οποία θεωρείται αξιόπιστη μετρική και χρησιμοποιείται ευρέως για την σύγκριση σφαλμάτων σε προβλέψεις χρονοσειρών ενεργειακής παραγωγής. Μέσω της χρήσης του Early stopping το μοντέλο τελικά εκπαιδεύτηκε για 11 εποχές. Τα αποτελέσματα των προβλέψεων με την χρήση του Bidirectional LSTM για χρονικό ορίζοντα 30 ημερών φαίνονται στο σχήμα 5.1.



Σχήμα 5.1: Αποτελέσματα πρόβλεψης αιολικής παραγωγής, LSTM



Σχήμα 5.2: Ημερήσιο σφάλμα πρόβλεψης LSTM

Βλέπουμε, ότι το LSTM νευρωνικό δίκτυο είναι ικανό να μοντελοποιήσει σε ένα ικανοποιητικό επίπεδο τα ιστορικά δεδομένα και τις προβλέψεις τις ταχύτητας του ανέμου. Λόγω της ασάφειας των δεδομένων καιρού, αλλά και της τυχαιότητας που εμφανίζει η χρονοσειρά αιολικής παραγωγής, δεν μπορούν να προβλεφθούν οι συχνές και απότομες μεταβολές, αλλά το νευρωνικό δίκτυο έχει την δυνατότητα να αντληθεί το επίπεδο που θα κινηθεί η παραγω-

γή της επόμενης ημέρας. Στο σχήμα 5.2, μπορεί να παρατηρηθεί το μέσο απόλυτο σφάλμα υπολογισμένο με τις ανά μονάδα τιμές ισχύος για την κάθε μέρα του μήνα. Φαίνεται πως τις περισσότερες μέρες οι προβλέψεις έχουν αρκετά μικρό σφάλμα, ενώ ημέρες κυρίως μεγαλύτερης παραγωγής εμφανίζουν μεγαλύτερο σφάλμα, όπως ήταν αναμενόμενο. Προσθετικά στον πίνακα 5.2 υπολογίζεται το μέσο απόλυτο σφάλμα όπως και η ρίζα του μέσου τετραγωνικού σφάλματος, για όλη την περίοδο που προβλέφθηκε (720 ώρες), αλλά και η μετρική R^2 , που μας δείχνει το ποσοστό της διακύμανσης της παραγωγής αιολικής ενέργειας που μπορεί να προβλεφθεί από το μοντέλο μας.

Μοντέλο \ Μετρική	MAE	RMSE	R^2
LSTM	0.0538	0.0713	0.7518

Πίνακας 5.2: Συνολικό σφάλμα LSTM

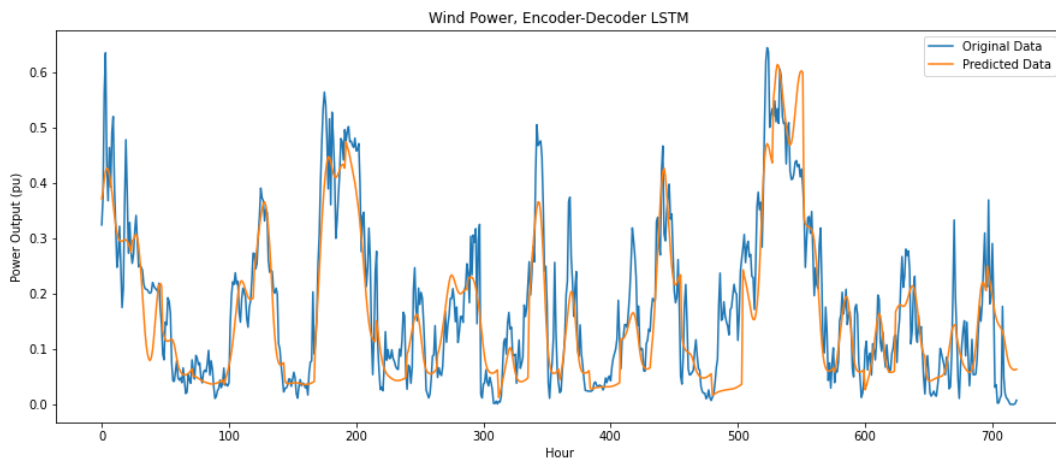
5.1.2 Εφαρμογή Encoder-Decoder LSTM νευρωνικού δικτύου

Η επόμενη μέθοδος που δοκιμάστηκε είναι αυτή του κωδικοποιητή-αποκωδικοποιητή. Παρόλο που, και σε αυτό το μοντέλο χρησιμοποιούνται LSTM νευρωνικά δίκτυα η διαδικασία παραγωγής των προβλέψεων διαφέρει αρκετά. Πιο συγκεκριμένα, για την υλοποίηση αυτής της αρχιτεκτονικής, χρησιμοποιείται ένα (ή περισσότερα) LSTM επίπεδα τα οποία λειτουργούν ως κωδικοποιητής. Έτσι, επεξεργάζονται τα δεδομένα εισόδου και επιστρέφουν την εσωτερική τους κατάσταση, η οποία λειτουργεί ως context vector συγκεντρώνοντας όλες τις σημαντικές πληροφορίες του διανύσματος εισόδου. Από το κομμάτι του κωδικοποιητή, ο context vector είναι το μόνο δεδομένο που μας χρειάζεται, για αυτό και οι έξοδοι των LSTM στο επίπεδο του κωδικοποιητή διαγράφονται. Στην συνέχεια, ένα (ή περισσότερα) LSTM επίπεδα λειτουργούν ως αποκωδικοποιητής της προηγούμενης πληροφορίας. Έτσι, τα LSTM επίπεδα του αποκωδικοποιητή αρχικοποιούνται χρησιμοποιώντας τον context vector, ώστε να δοθεί η πληροφορία του τι πρέπει να παραχθεί. Έπειτα, τα LSTM επίπεδα παράγουν το διάνυσμα εξόδου, παράγοντας μια έξοδο σε κάθε χρονική στιγμή σε αντίθεση με την προηγούμενη αρχιτεκτονική, όπου χρησιμοποιούνταν μόνο η έξοδος της τελευταίας χρονικής στιγμής των LSTM. Και σε αυτή την περίπτωση για την εύρεση της βέλτιστης αρχιτεκτονικής δοκιμάστηκαν παραλλαγές του μοντέλου για να καταλήξουμε στην βέλτιστη όπως φαίνεται στον πίνακα 5.3. Ακόμα, δοκιμάστηκαν διαφορετικοί χρονικοί ορίζοντες για τον χωρισμό των ιστορικών παρατηρήσεων και συγκεκριμένα οι 24, 48 και 72, με τα καλύτερα αποτελέσματα να λαμβάνονται χρησιμοποιώντας 48 ιστορικές τιμές τις χρονοσειράς σε κάθε διάνυσμα εκπαίδευσης. Οι εποχές εκπαίδευσης ορίστηκαν σε 50, όμως μέσω του Early stopping περιορίστηκαν σε 21, ενώ η υλοποίηση του νευρωνικού δικτύου Encoder-Decoder βασίστηκε στο [17].

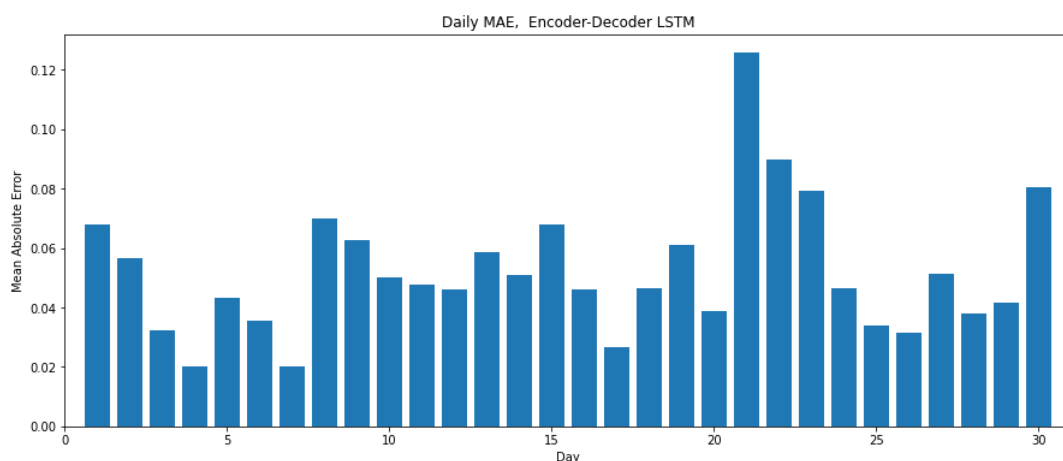
Υπερπαράμετρος	Εναλλακτικές	Βέλτιστη
Αριθμός επιπέδων LSTM	[1,2]	2
LSTM Units 1 ^{ου} επιπέδου	[50, 100, 150, 200]	200
LSTM Units 2 ^{ου} επιπέδου	[50, 100, 150, 200]	200
Συνάρτηση ενεργοποίησης	[tanh, ReLU]	ReLU
Αλγόριθμος βελτιστοποίησης	[RMSprop, Adam]	Adam
Batch size	[64, 128, 256]	256

Πίνακας 5.3: Υπερπαράμετροι Encoder-Decoder LSTM

Στο σχήμα 5.3 φαίνονται οι προβλέψεις χρονικού ορίζοντα 24 ωρών του Encoder-Decoder LSTM μοντέλου για τα δεδομένα ελέγχου διάρκειας ενός μήνα, ενώ στο σχήμα 5.4 το μέσο απόλυτο σφάλμα για την κάθε ημέρα πρόβλεψης.



Σχήμα 5.3: Αποτελέσματα πρόβλεψης αιολικής παραγωγής, Encoder-Decoder LSTM



Σχήμα 5.4: Ημερήσιο σφάλμα πρόβλεψης Encoder-Decoder LSTM

Διαπιστώνουμε ότι μέσω της τεχνικής κωδικοποιητή-αποκωδικοποιητή παράγονται αρκετά ι-

κανοποιητικές προβλέψεις για την παραγωγή της αιολικής ενέργειας. Επίσης, βλέπουμε στο σχήμα 5.4, ότι τα σφάλματα είναι ελαφρώς μικρότερα σε σχέση με την πρώτη μεθοδολογία, γεγονός που επαληθεύεται και από τον πίνακα 5.4, όπου αποτυπώνονται οι διαφορετικές μετρικές σφάλματος.

Μοντέλο \ Μετρική	MAE	RMSE	R ²
Encoder-Decoder LSTM	0.0523	0.0689	0.7687

Πίνακας 5.4: Συνολικό σφάλμα Encoder-Decoder LSTM

5.1.3 Εφαρμογή TCN νευρωνικού δικτύου

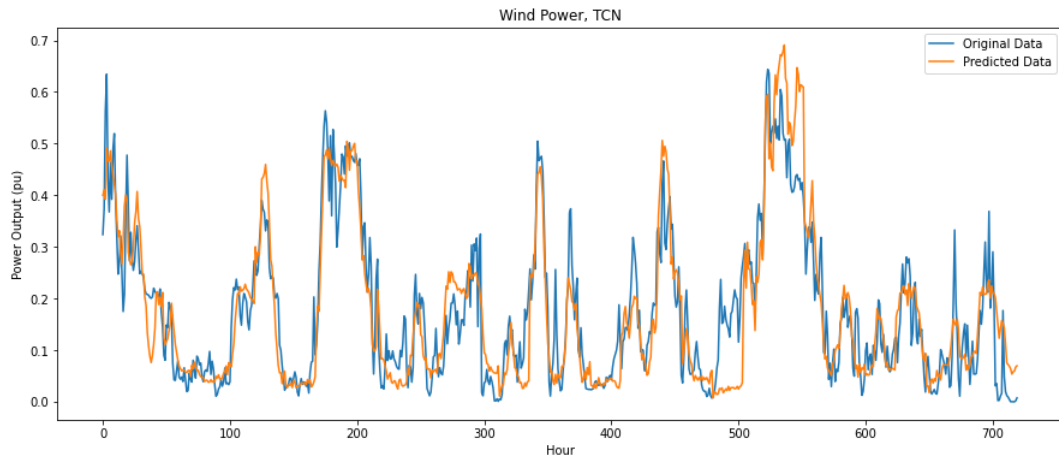
Ενώ τα συνελικτικά δίκτυα είχαν ως αρχικό στόχο εφαρμογές όρασης υπολογιστών, τελευταία χρησιμοποιούνται ευρέως για την πρόβλεψη χρονοσειρών, έχοντας την δυνατότητα να εξάγουν υψηλού επιπέδου χαρακτηριστικά από τα δεδομένα. Για την δημιουργία του μοντέλου, πρέπει να καθορίσουμε διάφορες παραμέτρους, ώστε να τα TCN να λειτουργούν βέλτιστα. Μια από τις πιο βασικές είναι το μήκος των ιστορικών παρατηρήσεων, αφού βάσει αυτού πρέπει να προσαρμόσουμε το μέγεθος των φίλτρων, αλλά και τον αριθμό των διεσταλμένων συνελίξεων όπως και το των αριθμών των TCN νευρωνικών δικτύων, ώστε το δεκτικό πεδίο να επαληθεύει την εξίσωση $(receptive\ field = no.\ stacked\ blocks \times kernel\ size \times last\ dilation\ factor)$. Έτσι, εξασφαλίζεται ότι οι τιμές της εξόδου θα επηρεάζονται από όλες τις τιμές που περιλαμβάνονται στο διάνυσμα εισόδου. Το γεγονός ότι τα TCN χρησιμοποιούν διεσταλμένες συνελίξεις τους δίνει την δυνατότητα, όπως εξηγήσαμε και στην θεωρία, να μπορούν να χειριστούν διανύσματα εισόδου μεγαλύτερου μήκους. Για αυτόν το λόγο, εκτός των μικρών ιστορικών οριζόντων των 24 και 48 τιμών δοκιμάστηκαν και μεγαλύτεροι όπως αυτοί των 96 και 192 τιμών με τα βέλτιστα αποτελέσματα να επιτυγχάνονται με την χρήση 96 τιμών ιστορικών παρατηρήσεων. Το σύνολο των παραμέτρων που δοκιμάστηκαν περιγράφεται στον πίνακα 5.5. Οι εποχές εκπαίδευσης ορίστηκαν σε 50 και μέσω του Early stopping χρειάστηκαν τελικά 22 για την σύγκλιση του, ενώ η υλοποίηση του νευρωνικού δικτύου TCN βασίστηκε στο [51].

Υπερπαραμέτρος	Εναλλακτικές	Βέλτιστη
Αριθμός στοιβαγμένων TCN	[3,4]	3
Μήκος φίλτρου	[2, 3]	2
Αριθμός φίλτρων	[32, 64, 128, 256]	64
Διεσταλμένες συνελίξεις	[1, 2, 4, 8, 16, 32]	[1, 2, 4, 8, 16]
Συνάρτηση ενεργοποίησης	[tanh, ReLU]	ReLU
Αλγόριθμος βελτιστοποίησης	[RMSprop, Adam]	RMSprop
Batch size	[64, 128, 256]	256

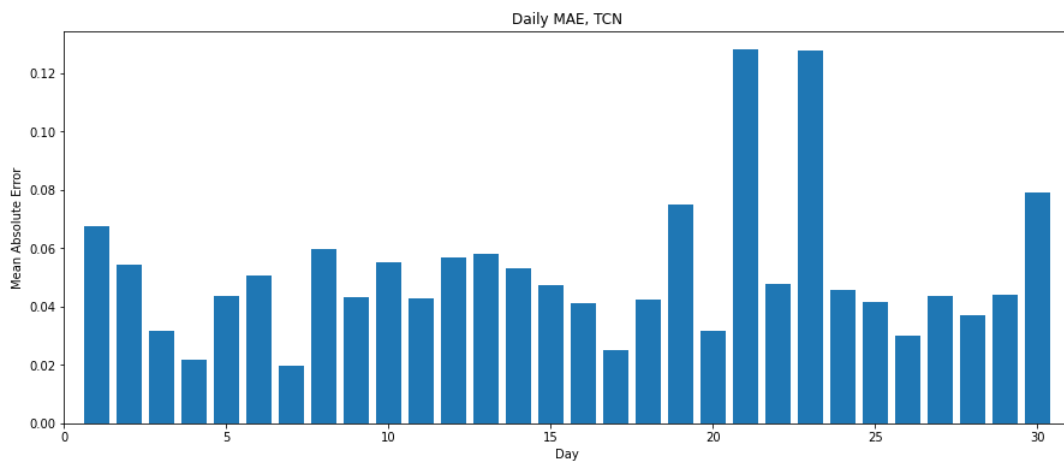
Πίνακας 5.5: Υπερπαραμέτροι TCN

Ορίζοντας τις παραμέτρους του μοντέλου, τελικά το TCN νευρωνικό δίκτυο αποτελείται από

3 Residual Blocks (Σχήμα 3.35) με διασταλμένες συνελίξεις ([1, 2, 4, 8, 16]), οι οποίες καταλήγουν σε 24 νευρώνες για την δημιουργία των προβλέψεων. Στο σχήμα 5.5 φαίνονται οι προβλέψεις του νευρωνικού δικτύου για χρονικό ορίζοντα 24 ωρών και διάρκεια 30 ημερών.



Σχήμα 5.5: Αποτελέσματα πρόβλεψης αιολικής παραγωγής, TCN



Σχήμα 5.6: Ημερήσιο σφάλμα πρόβλεψης TCN

Διαπιστώνεται ότι τα TCN έχουν την δυνατότητα να παράγουν πρόβλεψεις με μεγάλη ακρίβεια και συγχρόνος να εμφανίσουν μικρό ημερήσιο μέσο απόλυτο σφάλμα κατά την διάρκεια των 30 ημερών. Ακόμα, όπως φαίνεται και από τον πίνακα 5.6, οι συνολικές μετρικές σφάλματος είναι χαμηλές, ενώ το R^2 δείχνει ότι το μοντέλο μπορεί να προβλέψει αρκετά καλά τις διακυμάνσεις της παραγωγής αιολικής ενέργειας.

Μοντέλο \ Μετρική	MAE	RMSE	R^2
TCN	0.0515	0.0680	0.7743

Πίνακας 5.6: Συνολικό σφάλμα TCN

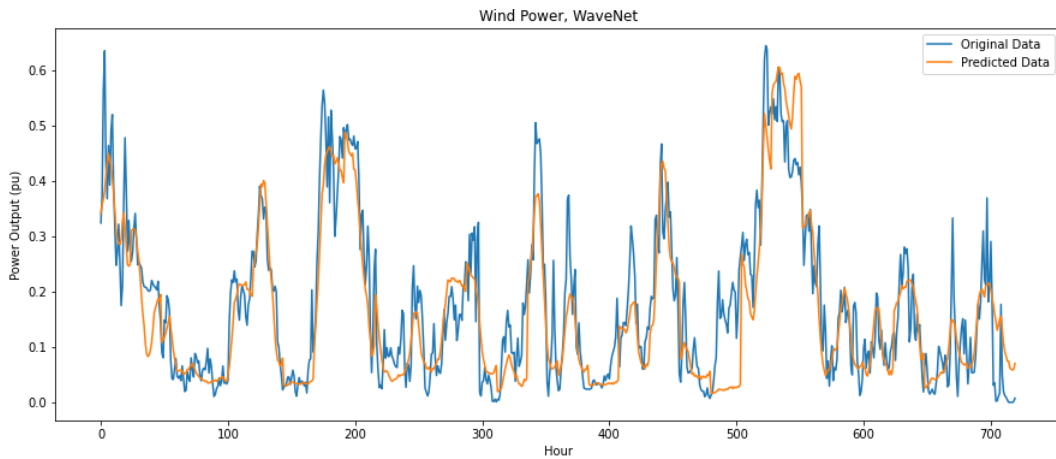
5.1.4 Εφαρμογή WaveNet νευρωνικού δικτύου

Η τελευταία μέθοδος, η οποία δοκιμάστηκε ήταν μέσω των νευρωνικών δικτύων WaveNet. Όπως έχει ήδη αναφέρει, το συγκεκριμένο μοντέλο έχει ως βάση του τα χαρακτηριστικά των TCN, όπως οι causal δισταλμένες συνελίξεις. Εκτός αυτών των χαρακτηριστικών, βασικές προσθήκες είναι οι skip connections, αλλά και οι gated ενεργοποιήσεις οι οποίες ακολουθούν την αρχιτεκτονική των LSTM νευρωνικών δικτύων. Και σε αυτή την περίπτωση για την εύρεση των βέλτιστων παραμέτρων θα πρέπει να διερευνηθούν διαφορετικές αρχιτεκτονικές, όπως τα μήκη ιστορικών παρατηρήσεων, έχοντας ως γνώμονα ότι και τα WaveNet νευρωνικά δίκτυα έχουν υλοποιηθεί για την χρήση μεγάλου μήκους ιστορικών παρατηρήσεων. Θα πρέπει να αναφέρουμε επιπλέον, ότι στις δοκιμές μας πειραματιστήκαμε με την gated ενεργοποίηση αφαιρώντας την, απλουστεύοντας το μοντέλο, κάνοντας χρήση των κλασικών συναρτήσεων ενεργοποίησης tanh και ReLU. Σύμφωνα με τα αποτελέσματά μας, ενώ οι gated ενεργοποίηση συμπεριφέρεται εξαιρετικά σε δεδομένα ήχου [48], στις χρονοσειρές παραγωγής αιολικής ενέργειας οι κλασικές συναρτήσεις ενεργοποίησης εμφάνισαν μικρότερα σφάλματα. Σε αντίστοιχία με τα TCN δοκιμάστηκαν διάφορες παραλλαγές του μοντέλου για την εύρεση του πιο αποδοτικού, με τις τελικές παραμέτρους να φαίνονται στον πίνακα 5.7. Από τους διαφορετικούς ιστορικούς χρονικούς ορίζοντες που δοκιμάστηκαν (24, 48, 96, 192) τα καλύτερα αποτελέσματα εμφάνισε ο χρονικός ορίζοντας των 96 παρελθοντικών παρατηρήσεων. Οι εποχές εκπαίδευσης ορίστηκαν σε 50 όμως μέσω του Early stopping περιορίστηκαν σε 13 για την πλήρη σύγκλιση του, ενώ η υλοποίηση του νευρωνικού δικτύου WaveNet βασίστηκε στο [29].

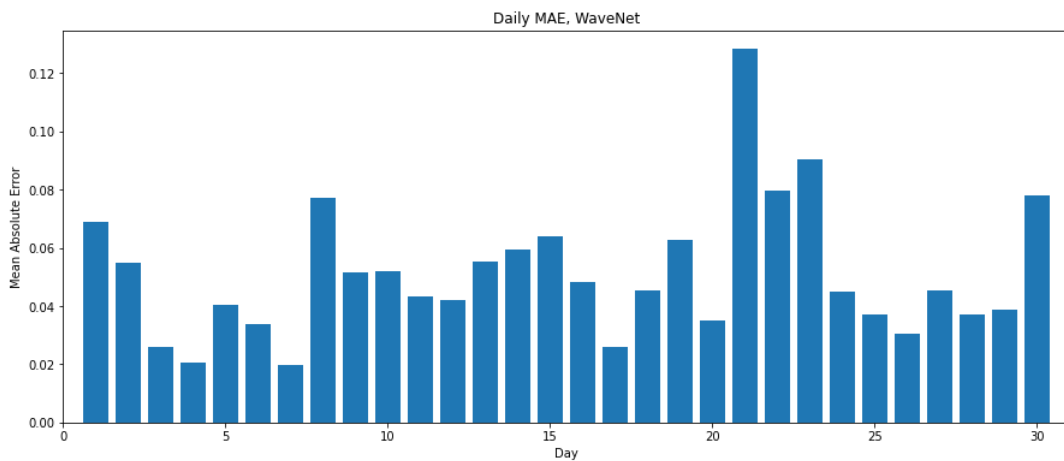
Υπερπαραμέτρος	Εναλλακτικές	Βέλτιστη
Αριθμός φίλτρων	[32, 64, 128, 256]	256
Δισταλμένες συνελίξεις	[1, 2, 4, 8, 16, 32]	[1, 2, 4, 8, 16]
Συνάρτηση ενεργοποίησης	[gated, tanh, ReLU]	ReLU
Αλγόριθμος βελτιστοποίησης	[RMSprop, Adam]	Adam
Batch size	[64, 128, 256, 512]	512

Πίνακας 5.7: Υπερπαραμέτροι WaveNet

Η λογική του νευρωνικού δικτύου WaveNet ακολουθεί αυτή του σχήματος 3.36, με διαφορά, ότι το επίπεδο της gated activation αντικαταστάθηκε από την συνάρτηση ReLU. Ακόμα, οι συνδέσεις από τα Residual Blocks του δικτύου ενώνονται μέσω των skip connections, σε έναν ενιαίο πίνακα, τον οποίο διατρέχει μια 1×1 συνέλιξη 128 φίλτρων η οποία καταλήγει σε 24 νευρώνες για την λήψη του τελικού αποτελέσματος. Ωστόσο στην περίπτωση του δικτύου μας οι 1×1 συνελίξεις δεν ακολουθούνται από την συνάρτηση ενεργοποίησης softmax, αλλά συνδέονται στους 24 νευρώνες οι οποίοι έχουν γραμμική συνάρτηση ενεργοποίησης (linear) για την λήψη των αποτελεσμάτων της παλινδρόμησης. Τα αποτελέσματα του βέλτιστου μοντέλου για το διάστημα ελέγχου φαίνονται στο σχήμα 5.7.



Σχήμα 5.7: Αποτελέσματα πρόβλεψης αιολικής παραγωγής, WaveNet



Σχήμα 5.8: Ημερήσιο σφάλμα πρόβλεψης WaveNet

Αποδεικνύεται, ότι και το WaveNet έχει την δυνατότητα παραγωγής αρκετά καλών αποτελεσμάτων με μικρά ημερήσια σφάλματα. Βλέποντας στη συνέχεια τις μετρικές του σφάλματος για όλη την περίοδο ελέγχου (πίνακας 5.8), διαπιστώνεται ότι μέσω του WaveNet επιτυγχάνονται τα μικρότερα σφάλματα (σε σχέση με τα αναδρομικά νευρωνικά δίκτυα), ενώ παράλληλα προβλέπονται με την μεγαλύτερη ακρίβεια οι διακυμάνσεις της παραγωγής αιολικής ενέργειας.

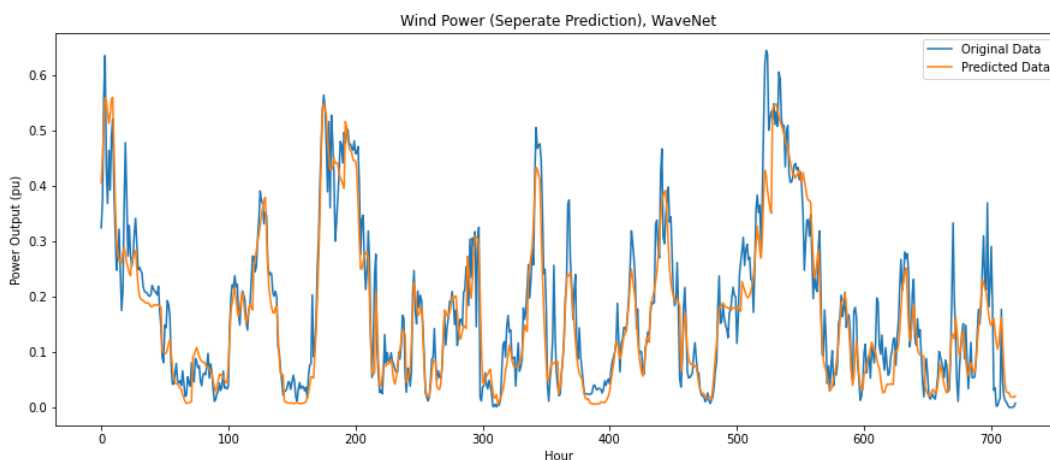
Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0512	0.0680	0.7745

Πίνακας 5.8: Συνολικό σφάλμα WaveNet

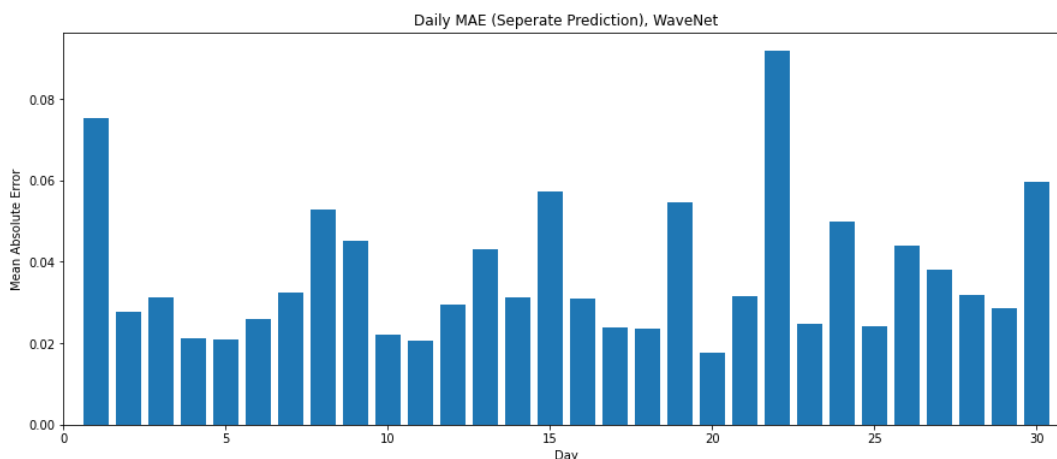
5.1.5 Εφαρμογή ξεχωριστής πρόβλεψης για το κάθε αιολικό πάρκο

Σύμφωνα με τις προβλέψεις των μοντέλων μηχανικής μάθησης μέχρι στιγμής, συμπεραίνουμε πως είναι εφικτή η πρόβλεψη της αιολικής παραγωγής για χρονικό ορίζοντα 24 ωρών

με σημαντική ακρίβεια, στην περίπτωση που υπάρχουν διαθέσιμες προβλέψεις ταχύτητας ανέμου, ακόμα και αν αυτές είναι σε ασαφή μορφή και όχι για κάθε ώρα της επόμενης ημέρας. Μέχρι τώρα όμως, για την πρόβλεψη χρησιμοποιήσαμε το σταθμισμένο μέσο όρο των παραγωγών αιολικής ενέργειας των πάρκων. Κάτι τέτοιο ωστόσο, ενδέχεται να οδηγήσει σε πρόσθετα σφάλματα. Για το λόγο αυτό, ελέγχθηκε και η δημιουργία ενός ξεχωριστού μοντέλου πρόβλεψης για το κάθε αιολικό πάρκο. Με αυτό το τρόπο, κάθε νευρωνικό δίκτυο έχει την δυνατότητα να προσαρμοστεί καλύτερα στις ιδιαίτερες συνθήκες του κάθε αιολικού πάρκου και να παράξει ακριβέστερες προβλέψεις. Η μεθοδολογία που ακολουθείται στην συνέχεια, είναι η ταυτόχρονη εκπαίδευση τριών ξεχωριστών μοντέλων (ένα για κάθε πάρκο) και η στάθμιση των προβλέψεών τους για την λήψη της τελικής παραγωγής αιολικής ενέργειας. Για την δημιουργία του μοντέλου αυτού γίνεται η χρήση του Functional API του Keras, το οποίο χρησιμοποιήθηκε και για την δημιουργία των προηγούμενων μοντέλων, καθώς μας προσφέρει δυνατότητες παραμετροποίησης του νευρωνικού δικτύου ανάλογα με τις ανάγκες της πρόβλεψης. Ως νευρωνικό δίκτυο για την δημιουργία του μοντέλου, επιλέχτηκε το WaveNet δεδομένου ότι τα αποτελέσματα του είχαν μεγάλη ακρίβεια. Οι προβλέψεις για την περίοδο ελέγχου μέσω αυτής της αρχιτεκτονικής απεικονίζονται στο σχήμα 5.9. Οι εποχές εκπαίδευσης ορίστηκαν σε 50, χωρίς να περιοριστούν μέσω του Early stopping.



Σχήμα 5.9: Αποτελέσματα πρόβλεψης αιολικής παραγωγής



Σχήμα 5.10: Ημερήσιο σφάλμα πρόβλεψης

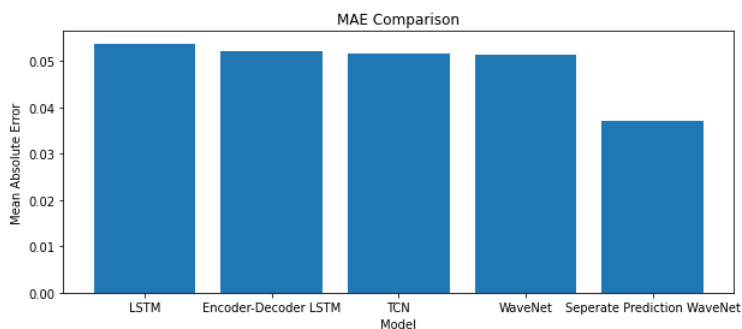
Διαπιστώνεται ότι προβλέποντας ξεχωριστά την παραγωγή αιολικής ενέργειας για το κάθε πάρκο, τα αποτελέσματα έχουν την μεγαλύτερη ακρίβεια. Επίσης, είναι δυνατό να προβλεφθούν απότομες αλλαγές της αιολικής παραγωγής που μέσω των προηγούμενων μεθόδων δεν ήταν εφικτό. Αυτό, επαληθεύεται και μέσω του ημερήσιου διαγράμματος του μέσου απόλυτου σφάλματος του σχήματος 5.10, αλλά και από τον πίνακα 5.9, όπου τα συνολικά σφάλματα είναι σημαντικά μειωμένα, ενώ η ακρίβεια πρόβλεψης των διακυμάνσεων αρκετά υψηλότερη σε σχέση με τις προηγούμενες προσεγγίσεις.

Μοντέλο \ Μετρική	Μετρική	MAE	RMSE	R ²
	WaveNet		0.0371	0.0521

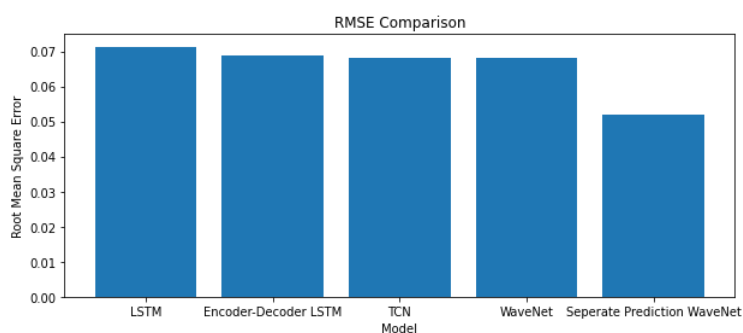
Πίνακας 5.9: Συνολικό σφάλμα ξεχωριστής πρόβλεψης αιολικών πάρκων

5.1.6 Σύγκριση αποτελεσμάτων

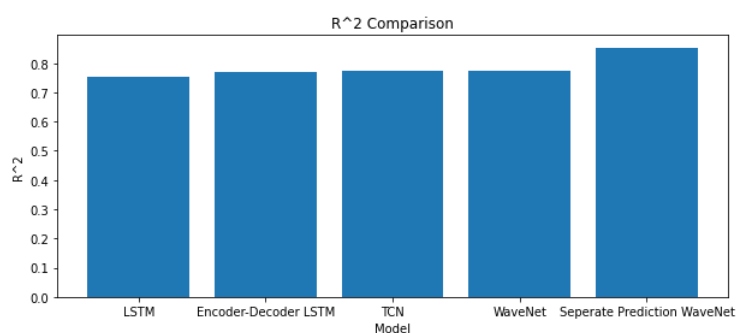
Αναλύοντας, όλα τα διαφορετικά μοντέλα, χρήσιμη θεωρείται η πραγματοποίηση της σύγκρισής τους, ώστε να διαπιστωθεί και να σχολιαστεί η αποδοτικότητα του κάθε μοντέλου σε σχέση με τα υπόλοιπα. Στα σχήματα 5.11, 5.12, 5.13, απεικονίζεται η σύγκριση του μέσου απόλυτου σφάλματος, της ρίζας του μέσου τετραγωνικού σφάλματος όπως και του συντελεστή προσδιορισμού για την περίοδο του ενός μήνα που εξετάστηκε.



Σχήμα 5.11: Σύγκριση μέσου απόλυτου σφάλματος



Σχήμα 5.12: Σύγκριση ρίζας μέσου τετραγωνικού σφάλματος



Σχήμα 5.13: Σύγκριση συντελεστή προσδιορισμού

Από τα διαγράμματα, φαίνεται αρχικά, πως όλα τα μοντέλα των νευρωνικών δικτύων έχουν την ικανότητα να προβλέψουν με ικανοποιητική ακρίβεια την παραγωγή αιολικής ενέργειας για διάστημα 24 ωρών. Επίσης, παρατηρείται ότι τα συνελκτικά δίκτυα αν και δεν έχουν κατασκευαστεί για την χρήση τους σε εφαρμογές πρόβλεψης (σε αντίθεση με τα αναδρομικά δίκτυα), έχουν την ικανότητα να παράγουν προβλέψεις με χαμηλότερα σφάλματα σε σχέση με τα αναδρομικά νευρωνικά δίκτυα. Ωστόσο, οι διαφορές δεν είναι μεγάλου εύρους αφού και πιο απλές αρχιτεκτονικές νευρωνικών δικτύων, όπως τα αναδρομικά νευρωνικά δίκτυα, συσχετίζουν τις παρελθοντικές παρατηρήσεις παραγωγής αιολικής ενέργειας και τις προβλέψεις της ταχύτητας σε ικανοποιητικό βαθμό.

Η διαφορά των προβλέψεων ωστόσο, είναι αισθητή μέσω της αλλαγής στην προσέγγιση του

μοντέλου. Πιο συγκεκριμένα, όπως φαίνεται και από τις γραφικές παραστάσεις των σφαλμάτων, η χρήση ενός ξεχωριστού μοντέλου για κάθε πάρκο και ο μετέπειτα συνδυασμός τους, μπορεί να παράξει προβλέψεις με σημαντικά καλύτερη ακρίβεια. Κάτι τέτοιο ήταν αναμενόμενο, αφού η εκπαίδευση ενός μοντέλου για κάθε πάρκο δίνει την δυνατότητα της προσαρμογής στα χαρακτηριστικά του κάθε πάρκου, ώστε η μετέπειτα στάθμιση των αποτελεσμάτων να οδηγεί σε αποτελέσματα αυξημένης ακρίβειας. Αντιθέτως, η στάθμιση των παραγωγών αιολικής ενέργειας, όπως και των ταχυτήτων κάθε πάρκου εξαλείφει τις ιδιαιτερότητες του κάθε πάρκου, και οδηγώντας σε αυξημένα σφάλματα.

5.1.7 Ετήσια πρόβλεψη αιολικής παραγωγής

Μέσω της επαναλήψης των εικοσιτετράωρων προβλέψεων για χρονικό ορίζοντα 30 ημερών, εξάγονται χρήσιμες πληροφορίες για την αποδοτικότητα των μοντέλων, όπως και για την μεταξύ τους σύγκριση. Ωστόσο, η αύξηση των δεδομένων ελέγχου σε επίπεδο έτους (προβλέποντας 8760 ώρες) μπορεί να δώσει πληροφορίες σχετικά με την δυνατότητα γενίκευσης των μοντέλων, στις μεταβολές των καιρικών συνθηκών σε ετήσια βάση, ενώ παράλληλα προσφέρει ένα ικανοποιητικό μέγεθος δεδομένων ελέγχου για την μελέτη της κατανομής του σφάλματος που παρουσιάζουν τα διαφορετικά μοντέλα.

Στην συνέχεια παρουσιάζονται οι μετρικές σφάλματος του νευρωνικού δικτύου WaveNet για την δημιουργία εικοσιτετράωρων προβλέψεων αιολικής παραγωγής για χρονικό διάστημα ενός έτους (8760 ώρες). Ως μοντέλο πρόβλεψης, χρησιμοποιήθηκε το νευρωνικό δίκτυο WaveNet, με χρήση των ιστορικών παρατηρήσεων του συνόλου της περιοχής για την οποία επιθυμούμε να παράγουμε τις προβλέψεις (όπου για την δημιουργία του εφαρμόστηκε στάθμιση των τιμών των τριών αιολικών πάρκων ανάλογα με την ονομαστική τους ισχύ) αλλά και μέσω της εκπαίδευσης ενός μοντέλου για το κάθε πάρκο (οι προβλέψεις των οποίων σταθμίζονται στην συνέχεια για την λήψη της πρόβλεψης της παραγωγής για το σύνολο της περιοχής). Οι μετρικές σφάλματος για την πρόβλεψη της ετήσιας αιολικής παραγωγής για τα δύο μοντέλα απεικονίζονται στους πίνακες 5.10 και 5.11.

Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0515	0.0684	0.8342

Πίνακας 5.10: Ετήσιο σφάλμα πρόβλεψης

Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0358	0.0504	0.9069

Πίνακας 5.11: Ετήσιο σφάλμα ξεχωριστής πρόβλεψης

Βλέπουμε ότι τα δύο μοντέλα έχουν την δυνατότητα να προβλέψουν με μεγάλη ακρίβεια τις

μεταβολές που λαμβάνουν χώρα κατά την διάρκεια του έτους, παρουσιάζοντας μετρικές σφάλματος παρόμοιες ή και χαμηλότερες σε σχέση με αυτές της μηνιαίας πρόβλεψης. Ωστόσο, σε πραγματικές συνθήκες θα αναμέναμε την αύξηση του σφάλματος μέσω της αύξησης του ορίζοντα πρόβλεψης μιας και η χρονοσειρά της αιολική παραγωγής παρουσιάζει αυξημένη τυχαιότητα. Παρόλα αυτά στα συγκεκριμένα μοντέλα, η εμφάνιση αντίστοιχων σφαλμάτων για την ετήσια και την μηνιαία πρόβλεψη δικαιολογείται καθώς οι προβλέψεις δεν έχουν προέρθει από πραγματικά μετεωρολογικά δεδομένα (των οποίων η ακρίβεια επηρεάζεται από τις τυχαίες και απότομες μεταβολές του καιρού), εμφανίζοντας πάντα πολύ μεγάλη συσχέτιση με τις πραγματικές τιμές της ταχύτητας του ανέμου.

Για την σύγκριση των προβλέψεων των δύο μοντέλων παρουσιάζεται η κατανομή των σφαλμάτων τους. Για την αξιολόγησή της θα υπολογιστεί η **μέση τιμή** των σφαλμάτων (μ), η **τυπική απόκλιση** τους (σ), η **ασυμμετρία (skewness)** (γ) και η **κύρτωση (kurtosis)** (k). Δεδομένου ότι τα μεγέθη της μέσης τιμής και της τυπικής απόκλισης έχουν ήδη περιγραφεί, θα αναλύσουμε τις έννοιες της ασυμμετρίας και της κύρτωσης. Για μια τυχαία μεταβλητή X η ασυμμετρία ορίζεται ως :

$$\gamma = \mathbf{E}\left[\left(\frac{\mathbf{X} - \mu}{\sigma}\right)^3\right]$$

όπου το μ είναι η μέση τιμή, το σ η τυπική απόκλιση και το E η αναμενόμενη τιμή της τυχαίας μεταβλητής X .

Η **ασυμμετρία**, όπως ορίζει και το όνομά της, είναι ένα μέτρο της ασυμμετρίας που παρατηρείται σε ένα δείγμα ή μια κατανομή. Στην περίπτωση επομένως που η τιμή της είναι αρνητική θεωρείται ότι τα δεδομένα είναι αριστερά ασύμμετρα, δηλαδή η αριστερή ουρά της κατανομής είναι μεγαλύτερη από την δεξιά. Σε αντίθετη περίπτωση, αν η τιμή της είναι θετική, τα δεδομένα είναι δεξιά ασύμμετρα και η δεξιά ουρά της κατανομής είναι μεγαλύτερη από την αριστερή. Η ασυμμετρία της κανονικής κατανομής είναι μηδενική, αφού τα δεδομένα είναι κατανομημένα συμμετρικά.

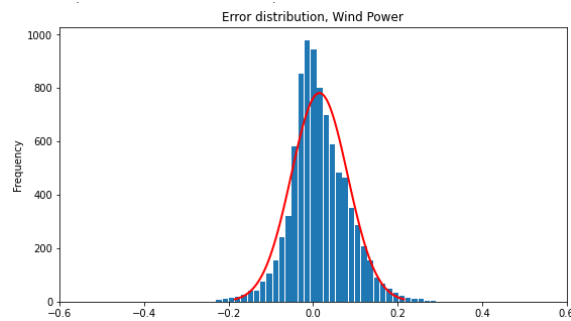
Για την ίδια τυχαία μεταβλητή X μπορεί να οριστεί επίσης και η κύρτωση ως:

$$k = \mathbf{E}\left[\left(\frac{\mathbf{X} - \mu}{\sigma}\right)^4\right]$$

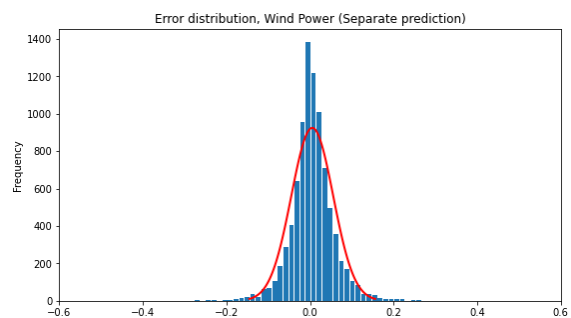
Η **κύρτωση** περιγράφει το μέγεθος της κορυφής της κατανομής. Ωστόσο, μπορεί να θεωρηθεί και ως μέτρηση του πάχους της ουράς μιας κατανομής, συγκριτικά με την κανονική κατανομή. Η κύρτωση αποτελεί ένα σημαντικό μέγεθος, καθώς η υψηλή τιμή της οδηγεί στο σχηματισμό ουράς μεγαλύτερου πάχους, το οποίο και αποτελεί ένδειξη ύπαρξης ακραίων τιμών (outliers). Επίσης σε αυτή την περίπτωση η κατανομή ονομάζεται λεπτόκυρτη. Αντιθέτως, κατανομές δεδομένων με μικρή τιμή κύρτωσης οδηγούν στο σχηματισμό πιο λεπτής ουράς, με αποτέλεσμα την έλλειψη ακραίων τιμών, ενώ σε αυτή την περίπτωση η κατανομή ονομάζεται πλατύκυρτη. Η κύρτωση της κανονικής κατανομής έχει τιμή 3, ωστόσο συνηθίζεται στους υπολογισμούς η κύρτωση της κανονικής κατανομής να υπολογίζεται αφαιρώντας την τιμή 3, ώστε να παίρνει την τιμή 0 για την ευκολότερη σύγκριση της με άλλες κατανομές (το παραπάνω ισχύει και στην βιβλιοθήκη `scipy.stats` της Python, όπου και θα χρησιμοποιήσουμε). Έχοντας ορίσει επαρκώς τα μεγέθη με τα οποία θα αξιολογηθούν οι κατανομές των σφαλμάτων,

μαζί με το ιστόγραμμα τους επιλέχθηκε και η απεικόνιση της συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής με μέση τιμή και τυπική απόκλιση, την μέση τιμή και την τυπική απόκλιση του δείγματος των δεδομένων μας. Πρέπει ωστόσο να επισημανθεί, ότι επειδή στο ιστόγραμμα επιθυμούσαμε την παρουσίαση της συχνότητα εμφάνισης της κάθε στήλης του δείγματος στον άξονα των τεταγμένων, επιλέχθηκε ο πολλαπλασιασμός της συνάρτησης πυκνότητας πιθανότητας με το εμβαδόν του ιστογράμματος. Διαφορετικά, θα μπορούσε να κανονικοποιηθεί η συχνότητα εμφάνισης του κάθε σφάλματος, με αποτέλεσμα την παρουσίαση της πραγματικής συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής με μοναδιαίο εμβαδόν, τεχνική που ωστόσο δεν υιοθετήθηκε.

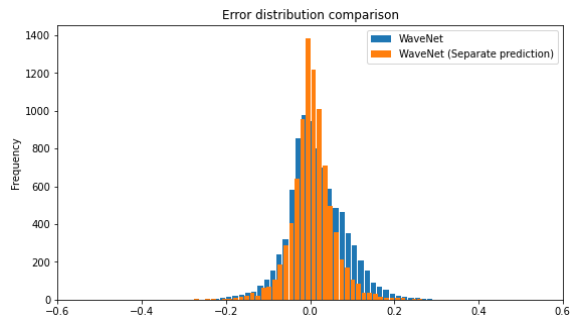
Στο σχήμα 5.14 φαίνεται η κατανομή σφάλματος των δύο μοντέλων ξεχωριστά, αλλά και αποτυπωμένη σε ένα ενιαίο διάγραμμα για την ορθότερη σύγκριση των αποτελεσμάτων.



(α') Κατανομή σφάλματος ($\mu = 0.0149$, $\sigma = 0.0668$, $\gamma = 0.2556$, $\kappa = 1.0654$)



(β') Κατανομή σφάλματος ξεχωριστής πρόβλεψης ($\mu = 0.0054$, $\sigma = 0.0501$, $\gamma = 0.1228$, $\kappa = 3.2601$)



Σχήμα 5.14: Σύγκριση κατανομής σφάλματος αιολικής παραγωγής

Γίνεται αντιληπτό, πως και οι δύο κατανομές παρουσιάζουν δεξιά ασυμμετρία γ , ενώ μπορο-

ύν να χαρακτηριστούν ως λεπτόκυρτες. Επίσης, φαίνεται ότι στην απευθείας πρόβλεψη, το μοντέλο τείνει να υπερεκτιμά σε μεγαλύτερο βαθμό την πρόβλεψη της αιολική παραγωγής, παρουσιάζοντας μεγαλύτερη συγκέντρωση σφαλμάτων αριστερά του μηδενός. Αντιθέτως, στην ξεχωριστή πρόβλεψη τα σφάλματα είναι πιο συμμετρικά κατανομημένα γύρω από το μηδέν. Προσθετικά, το σημαντικότερο χαρακτηριστικό των διαγραμμάτων είναι οι αποκλίσεις των σφαλμάτων από την μέση τιμή του δείγματος. Βλέπουμε πως η απευθείας πρόβλεψη παρουσιάζει μεγαλύτερη τυπική απόκλιση σε σχέση με την ξεχωριστή πρόβλεψη, κάτι το οποίο φαίνεται τόσο μέσω της αριθμητικής τιμής της, όσο και από το συγκριτικό διάγραμμα, όπου απεικονίζονται ταυτόχρονα οι δύο κατανομές.

Προσθήκη θορύβου στις προβλέψεις της ταχύτητας ανέμου

Κατά την διάρκεια της δημιουργίας των μοντέλων μηχανικής μάθησης για την πρόβλεψη της αιολικής παραγωγής χρησιμοποιήθηκαν ως είσοδοι οι πραγματικές ταχύτητες ανέμου αφού πρώτα ασαφοποιήθηκαν. Η τεχνική αυτή αν και προσέγγιζε μια πρόβλεψη καιρικών δεδομένων και ενώ οι τιμές της προέρχονταν μέσω της ασαφοποίησης τους, είχε ως αποτέλεσμα την παραγωγή αποτελεσμάτων με πολύ χαμηλές μετρικές σφάλματος. Κάτι τέτοιο ωστόσο, ενδέχεται να μην ισχύει μέσω της τροφοδότησης του μοντέλου με πραγματικές προβλέψεις ταχύτητας ανέμου για την περιοχή, αφού θα έχουν μεγαλύτερες αποκλίσεις από την πραγματικότητα.

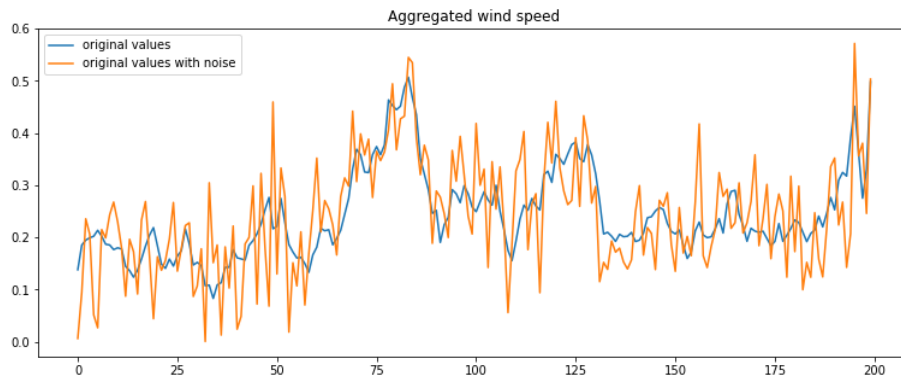
Για αυτό τον λόγο, επιλέξαμε τον έλεγχο των μετρικών σφάλματος για την παραγωγή ετήσιων προβλέψεων, προσθέτοντας περαιτέρω θόρυβο στα δεδομένα ταχύτητας ανέμου, ώστε να ελέγξουμε την διαφοροποίηση των αποτελεσμάτων. Γενικά η ταχύτητα ανέμου μπορεί να οριστεί ως [32] :

$$v^{forecast} = v^{actual} + \epsilon_{wind} \quad (5.1)$$

$$\epsilon_{wind} \sim N(0, \sigma^2) \quad (5.2)$$

όπου το $v^{forecast}$ αναφέρεται στην πρόβλεψη της ταχύτητας ανέμου, το v^{actual} στην πραγματική τιμή της ταχύτητας ανέμου και το ϵ_{wind} αποτελεί το σφάλμα πρόβλεψης της ταχύτητας ανέμου. Φαίνεται από τις παραπάνω εξισώσεις, ότι η πρόβλεψη της ταχύτητας ανέμου από τις αριθμητικές προβλέψεις καιρού (NWP) προσεγγίζεται από τις πραγματικές τιμές της, προσθέτοντας τις τιμές που έχουν προέλθει από μια κανονική κατανομή με μηδενική μέση τιμή και τυπική απόκλιση σ . Ακολουθώντας την παραπάνω λογική, θα προβλέψουμε την ετήσια παραγωγή αιολικής ενέργειας με τα δύο προηγούμενα μοντέλα νευρωνικών δικτύων. Σε αυτή την περίπτωση όμως, ενώ η εκπαίδευση των μοντέλων θα γίνει με τις ασαφοποιημένες παρατηρήσεις της ταχύτητας ανέμου, η λήψη των προβλέψεων θα πραγματοποιηθεί, αλλοιώνοντας τα δεδομένα εισόδου της ταχύτητας ανέμου με την προσθήκη των τιμών που έχουν προέλθει από την κανονική κατανομή, ώστε να επιτευχθεί η μοντελοποίηση των αριθμητικών προβλέψεων καιρού. Η διαφοροποίηση των τιμών της ταχύτητας ανέμου με την προσθήκη των τιμών της κανονικής κατανομής για τις 200 πρώτες ώρες ελέγχου του αιθροίσματος της σταθμισμένης

ταχύτητας των αιολικών πάρκων απεικονίζεται στο σχήμα 5.15.



Σχήμα 5.15: Σύγκριση τιμών ταχύτητας ανέμου με και χωρίς θόρυβο

Φαίνεται ότι η προσθήκη των τυχαίων τιμών που προήλθαν από την κανονική κατανομή, πρόσθεσε αρκετό θόρυβο στις παρατηρήσεις της ταχύτητας του ανέμου, ώστε να θεωρείται ότι προσεγγίζουν τα σφάλματα μιας πραγματικής πρόβλεψης της. Στην συνέχεια χρησιμοποιούμε τα δεδομένα αυτά αφού πρώτα τα ασαφοποιήσουμε, όπως και στην αρχική προσέγγιση, για την λήψη των προβλέψεων της αιολικής παραγωγής για το χρονικό διάστημα του ενός έτους. Οι μετρικές σφάλματος του νευρωνικού δικτύου για εκπαίδευσή του στην αθροιστική χρονοσειρά και στην χρονοσειρά του κάθε πάρκου ξεχωριστά φαίνονται στους πίνακες 5.12 και 5.13.

Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0803	0.1054	0.6069

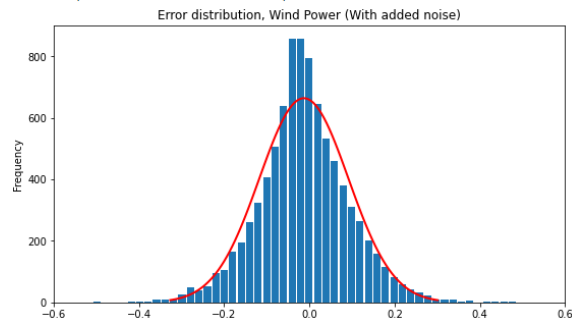
Πίνακας 5.12: Ετήσιο σφάλμα πρόβλεψης με προσθήκη θορύβου

Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0565	0.0755	0.7859

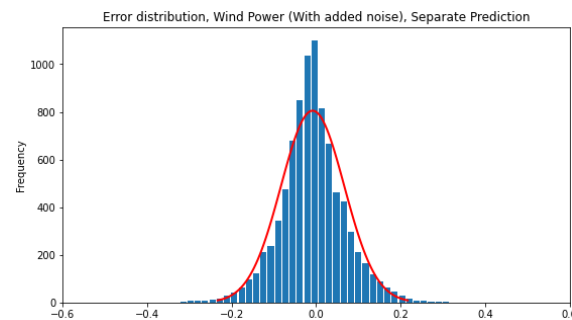
Πίνακας 5.13: Ετήσιο σφάλμα ξεχωριστής πρόβλεψης με προσθήκη θορύβου

Παρατηρείται ότι με την προσθήκη του θορύβου οι μετρικές σφάλματος είναι αυξημένες, ενώ ο συντελεστής προσδιορισμού μειωμένος. Ωστόσο, η πρόβλεψη του κάθε αιολικού πάρκου ξεχωριστά επιφέρει μικρότερη αύξηση των σφαλμάτων σε σχέση με την απευθείας πρόβλεψη του συνόλου της αιολικής παραγωγής της περιοχής. Αυτό, ήταν αναμενόμενο καθώς η ξεχωριστή πρόβλεψη του κάθε πάρκου οδηγεί στην εξάλειψη κάποιων σφαλμάτων κατά την διάρκεια της πρόβλεψης, όπως είχε ήδη παρατηρηθεί, οδηγώντας σε ικανοποιητικά αποτελέσματα, ακόμα και με μεγάλη αλλοίωση των δεδομένων ταχύτητας ανέμου για την επόμενη μέρα.

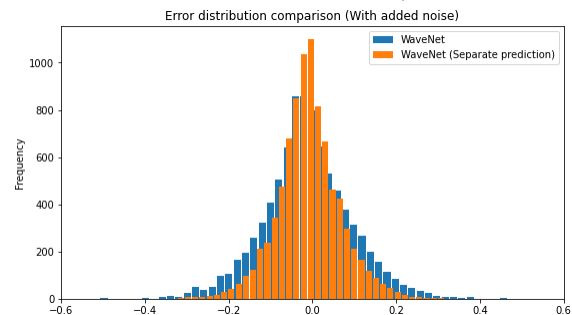
Στο σχήμα 5.16 απεικονίζονται και οι κατανομές σφάλματος των προβλέψεων των δύο μοντέλων.



(α') Κατανομή σφάλματος WaveNet με προσθήκη θορύβου ($\mu = -0.0130$, $\sigma = 0.1046$, $\gamma = 0.0723$, $\kappa = 0.9148$)



(β') Κατανομή σφάλματος WaveNet ξεχωριστής πρόβλεψης με προσθήκη θορύβου ($\mu = -0.0082$, $\sigma = 0.0750$, $\gamma = 0.1271$, $\kappa = 1.4238$)

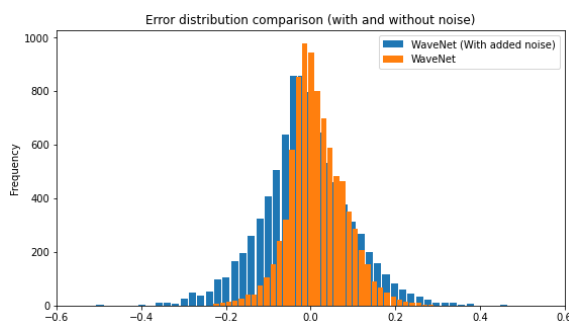


Σχήμα 5.16: Σύγκριση κατανομής σφάλματος αιολικής παραγωγής με προσθήκη θορύβου

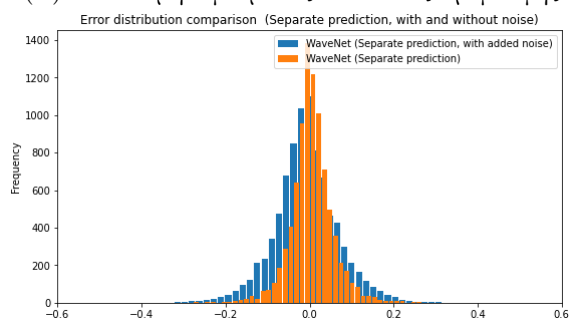
Αναλύοντας τις κατανομές του σφάλματος πρόβλεψης, παρατηρείται πως παρουσιάζουν μεγαλύτερες αποκλίσεις αναφορικά με τις προβλέψεις χωρίς την προσθήκη θορύβου, κάτι το οποίο και αναμέναμε. Επίσης, και σε αυτή την περίπτωση η τυπική απόκλιση της ξεχωριστής πρόβλεψης είναι μικρότερη συγκριτικά με την απευθείας, έχοντας μεγαλύτερη διαφοροποίηση από αυτήν συγκριτικά με την μη προσθήκη θορύβου, γεγονός που μας οδηγεί στο συμπέρασμα ότι η ξεχωριστή πρόβλεψη μπορεί να χειριστεί αποτελεσματικότερα την ύπαρξη σφαλμάτων στα δεδομένα καιρού. Επιπροσθέτως, και πάλι οι δύο κατανομές μπορούν να χαρακτηριστούν ως δεξιά ασύμμετρες, σε μικρότερο βαθμό ωστόσο, αφού όπως παρατηρείται οι δύο κατανομές έχουν αρκετά συμμετρικά κατανεμημένα σφάλματα.

Τέλος στο σχήμα 5.17 αποτυπώνεται η σύγκριση των κατανομών σφάλματος, με και χωρίς την

προσθήκη επιπλέον θορύβου στα δεδομένα της ταχύτητας ανέμου για τα δύο μοντέλα. Συγκριτικά όπως έχει ήδη σχολιαστεί, παρατηρείται πως η χρήση ενός μοντέλου πρόβλεψης για το κάθε αιολικό πάρκο έχει μεγαλύτερη ανοχή στην προσθήκη θορύβου στις παρατηρήσεις. Το παραπάνω συμπέρασμα γίνεται εμφανές, καθώς η κατανομή σφάλματος της ξεχωριστής πρόβλεψης με την προσθήκη θορύβου προσεγγίζει σε μεγαλύτερο βαθμό την κατανομή σφάλματος του μοντέλου χωρίς την προσθήκη θορύβου, έχοντας σημαντικά μικρότερες αποκλίσεις από αυτή, σε σχέση με την απευθείας πρόβλεψη του συνόλου της αιολικής παραγωγής.



(α) Κατανομή σφάλματος απευθείας πρόβλεψης



(β) Κατανομή σφάλματος ξεχωριστής πρόβλεψης

Σχήμα 5.17: Σύγκριση κατανομής σφάλματος αιολικής παραγωγής (με και χωρίς την προσθήκη θορύβου)

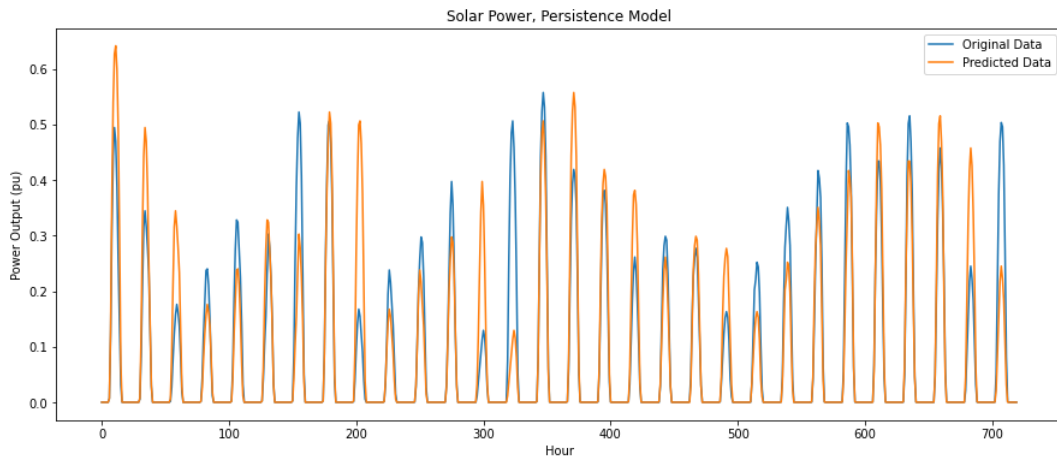
5.2 Πρόβλεψη φωτοβολταϊκής παραγωγής

Αντίστοιχα με την πρόβλεψη της αιολικής παραγωγής, θα δοκιμαστεί η χρήση μοντέλων μηχανικής μάθησης για την πρόβλεψη της φωτοβολταϊκής παραγωγής, με ορίζοντα πρόβλεψης 24 ωρών στο μέλλον. Όπως έχει αναφερθεί, τα δεδομένα μας αποτελούνται από 5 έτη ιστορικών παρατηρήσεων με ωριαία διακριτότητα για την φωτοβολταϊκή παραγωγή σε όλη την Ελλάδα. Για την επεξεργασία τους και την ανάπτυξη των μοντέλων μηχανικής μάθησης θα χωριστούν σε 3 υποομάδες (αντίστοιχα με τα δεδομένα της αιολικής παραγωγής), τα δεδομένα εκπαίδευσης, τα δεδομένα επαλήθευσης και τα δεδομένα ελέγχου. Επιλέχθηκε και σε αυτή των περίπτωση ως δεδομένα επαλήθευσης να χρησιμοποιηθούν τα τελευταία τριάντα εικοσιτετράωρα, ώστε να εξεταστεί η αποδοτικότητα των μοντέλων σε διαφορετικές ημέρες για την μεγαλύτερη αξιοπιστία των μεθόδων. Ακόμα, δεδομένου ότι τα τελευταία τριάντα ει-

κοσιτετράωρα αντιστοιχούν στην εποχή του χειμώνα, θα εξετάσουμε τα μοντέλα μας στην πιο δύσκολη περίπτωση της φωτοβολταϊκής παραγωγής, αφού κατά τους χειμερινούς μήνες υπάρχουν πολλές μεταβολές της ηλιοφάνειας με αποτέλεσμα να είναι δύσκολη η ακριβής πρόβλεψη της παραγωγής για την επόμενη ημέρα. Για την δημιουργία των προβλέψεων θα εξετάσουμε τα δυο από τα τέσσερα μοντέλα που παρουσιάστηκαν στην ανάλυση της αιολικής παραγωγής και συγκεκριμένα τα νευρωνικά δίκτυα Bidirectional-LSTM και WaveNet. Η επιλογή αυτή έγινε καθώς από τα αποτελέσματα της αιολικής παραγωγής μπορεί να εξαχθεί το συμπέρασμα ότι, ενώ τα συνελικτικά νευρωνικά δίκτυα υπερτερούν σε σχέση με τα αναδρομικά στην δημιουργία προβλέψεων και τα τέσσερα μοντέλα παρουσίασαν ικανοποιητικά αποτελέσματα με μικρές διαφορές, οπότε και επιλέχθηκε να δοκιμαστεί στην πρόβλεψη της φωτοβολταϊκής παραγωγής το απλούστερο και το πιο σύνθετο μοντέλο.

5.2.1 Εφαρμογή μοντέλου Παραμένουσας Τιμής (Persistence Model)

Για την σύγκριση των μοντέλων και της δυνατότητας τους να παράξουν προβλέψεις με μικρά σφάλματα, τα αποτελέσματά τους θα συγκριθούν με αυτά που παράγονται με το μοντέλο παραμένουσας τιμής (Persistence Model). Πολλές φορές για να επαληθεύσουμε τις ικανότητες των μοντέλων πρόβλεψης, θεωρείται καλή πρακτική η σύγκρισή τους με πιο απλά, όπως το μοντέλο παραμένουσας τιμής, αφού μόνο τότε είναι σκόπιμη η χρήση τους. Το μοντέλο παραμένουσας τιμής, στην γενική του μορφή υποθέτει ότι η παρατήρηση κατά την χρονική στιγμή x_{t+1} , είναι ίση με την τιμή της παρατήρησης την προηγούμενη χρονική στιγμή x_t [10]. Το μοντέλο παραμένουσας τιμής είναι γνωστό και ως *naïve* και είναι αρκετά χρήσιμο για την σύγκριση διαφορετικών μοντέλων για προβλέψεις ενός χρονικού βήματος. Αντιθέτως, όσο μεγαλώνει ο χρονικός ορίζοντας πρόβλεψης, η ακρίβεια του μειώνεται σε μεγάλο βαθμό καθώς θα συνεχίσει να προβλέπει την προηγούμενη τιμή σχηματίζοντας πρακτικά μια ευθεία γραμμή. Ωστόσο, λόγω του επαναλαμβανόμενου μοτίβου που παρατηρείται στην χρονοσειρά της φωτοβολταϊκής παραγωγής, μπορούμε να εφαρμόσουμε μια παραλλαγή του μοντέλου παραμένουσας τιμής, θεωρώντας ότι η τιμή της παραγωγής σε μια χρονική στιγμή x_t , θα είναι ίση με την αντίστοιχη τιμή παραγωγής της προηγούμενης ημέρας, δηλαδή πριν από 24 ώρες (x_{t-24}). Η εφαρμογή αυτής της μεθόδου, μας δίνει μια *naïve* πρόβλεψη, με αρκετά μικρό σφάλμα, την οποία μπορούμε να χρησιμοποιήσουμε για την σύγκριση των μοντέλων μας. Οι προβλέψεις του μοντέλου παραμένουσας τιμής για τα δεδομένα ελέγχουν φαίνονται στο σχήμα 5.18.



Σχήμα 5.18: Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, Persistence Model

Φαίνεται ότι η πρόβλεψη του μοντέλου παραμένουσας τιμής αποδεικνύεται ακριβής, ειδικά σε ημέρες που έχουν την περίπου ίδια ηλιοφάνεια με την προηγούμενη, αφού τότε η πρόβλεψη θα είναι αρκετά κοντά στην πραγματικότητα. Αντιθέτως, σε μέρες που παρατηρείται απότομη πτώση ή αύξηση της ηλιοφάνειας, το μοντέλο πραγματοποιεί πρόβλεψη αρκετά διαφορετική από την πραγματική. Η αξιολόγηση του μοντέλου της παραμένουσας τιμής, μέσω του μέσου απόλυτου σφάλματος, της ρίζας του μέσου τετραγωνικού σφάλματος όπως και του συντελεστή R^2 απεικονίζεται στον πίνακα 5.14. Αξίζει να σημειωθεί πως η πρώτη γραμμή του πίνακα 5.14 υπολογίστηκε βάσει όλων των τιμών των προβλέψεων της φωτοβολταϊκής παραγωγής. Αντιθέτως η δεύτερη, εξαιρώντας τις τιμές των προβλέψεων όπου η πραγματική τιμή της φωτοβολταϊκής παραγωγής ήταν μηδενική. Ο υπολογισμός των μετρικών με τους δυο αυτούς τρόπους θεωρήθηκε απαραίτητος, εφόσον οι ουσιαστικές προβλέψεις της φωτοβολταϊκής παραγωγής είναι αυτές κατά την διάρκεια της ημέρας, ενώ οι εύκολα προβλέψιμες μηδενικές τιμές της φωτοβολταϊκής παραγωγής κατά τις βραδινές ώρες, μπορούν να οδηγήσουν σε υπερεκτίμηση των δυνατοτήτων των μοντέλων, παρουσιάζοντας χαμηλότερα σφάλματα, όπως διαπιστώνεται και από τον πίνακα 5.14.

Μοντέλο \ Μετρική	MAE	RMSE	R^2
Persistence	0.0301	0.0692	0.7584
Persistence (Non zero values)	0.0697	0.1053	0.5342

Πίνακας 5.14: Συνολικό σφάλμα Persistence Model

5.2.2 Εφαρμογή LSTM νευρωνικού δικτύου

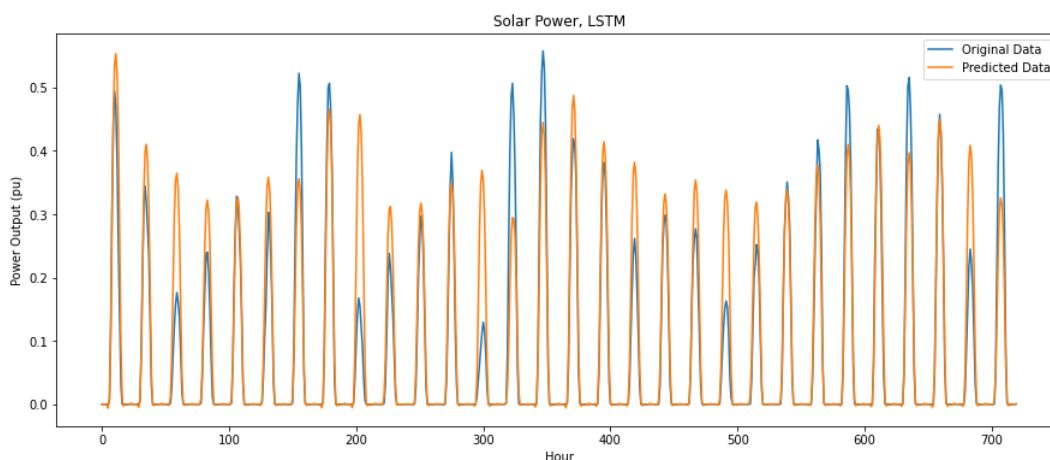
Για την δημιουργία των προβλέψεων μέσω των LSTM νευρωνικών δικτύων θα χρησιμοποιήσουμε και πάλι τα Bidirectional LSTM νευρωνικά δίκτυα. Για την εύρεση της αποδοτικότερης αρχιτεκτονικής, ελέγξαμε διάφορες παραλλαγές του μοντέλου στο ίδιο μέρος των δεδομένων εισόδου και ελέγχου ώστε να βρούμε την βέλτιστη. Ως δεδομένα για την εύρεση των βέλτι-

στων παραμέτρων, θα χρησιμοποιηθούν τα ιστορικά δεδομένα φωτοβολταϊκής παραγωγής για το πρώτο έτος (2015) για την επιτάχυνση της εκπαίδευσης. Οι παράμετροι που δοκιμάστηκαν φαίνονται στο πίνακα 5.15. Επιπλέον επισημάνεται, ότι δοκιμάστηκε η χρήση διαφορετικών χρονικών οριζόντων των παρελθοντικών δεδομένων και συγκεκριμένα οι 24, 48, 72, οι οποίοι ήταν όλοι πολλαπλάσιοι των 24 ωρών, εξαιτίας της μορφής των δεδομένων της φωτοβολταϊκής παραγωγής, με τα καλύτερα αποτελέσματα να λαμβάνονται για χρήση 24 παρελθοντικών τιμών στο διάλυμα εισόδου του νευρωνικού δικτύου.

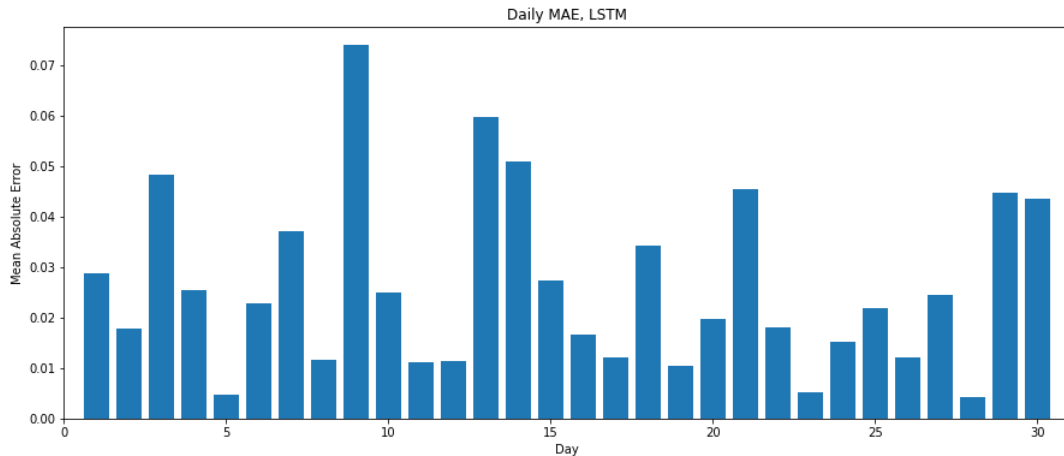
Υπερπαράμετρος	Εναλλακτικές	Βέλτιστη
Αριθμός επιπέδων LSTM	[1,2]	2
LSTM Units 1 ^{ου} επιπέδου	[100, 150, 200]	100
LSTM Units 2 ^{ου} επιπέδου	[50, 100, 150]	100
Συνάρτηση ενεργοποίησης	[tanh, ReLU]	ReLU
Αλγόριθμος βελτιστοποίησης	[RMSprop, Adam]	Adam
Batch size	[64, 128, 256]	128

Πίνακας 5.15: Υπερπαράμετροι LSTM

Ορίζοντας τις βέλτιστες παραμέτρους, εκπαιδεύουμε το νευρωνικό δίκτυο για 50 εποχές (οι οποίες μέσω του Early stopping περιορίστηκαν σε 35) και στην συνέχεια προβλέπουμε τα αντίστοιχα διάλυμα δεδομένων που προβλέφθηκε και με το Persistence μοντέλο. Τα αποτελέσματα της πρόβλεψης για το χρονικό διάστημα 30 ημερών, όπως και το ημερήσιο σφάλμα φαίνονται στα σχήματα 5.19, 5.20.



Σχήμα 5.19: Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, LSTM



Σχήμα 5.20: Ημερήσιο σφάλμα πρόβλεψης LSTM

Από το σχήμα 5.19, παρατηρείται πως η πρόβλεψη χρησιμοποιώντας τα LSTM νευρωνικά δίκτυα είναι ακριβέστερη συγκριτικά με το μοντέλο Persistence, λαμβάνοντας υπόψιν μόνο ιστορικές παρατηρήσεις. Το συμπέρασμα αυτό επαληθεύεται και από τον πίνακα 5.16, όπου απεικονίζονται οι μετρικές σφαλμάτων και ο συντελεστής R^2 .

Μοντέλο \ Μετρική	Μετρική		
	MAE	RMSE	R^2
LSTM	0.0261	0.0565	0.8317
LSTM (Non zero values)	0.0594	0.0859	0.6416

Πίνακας 5.16: Συνολικό σφάλμα LSTM

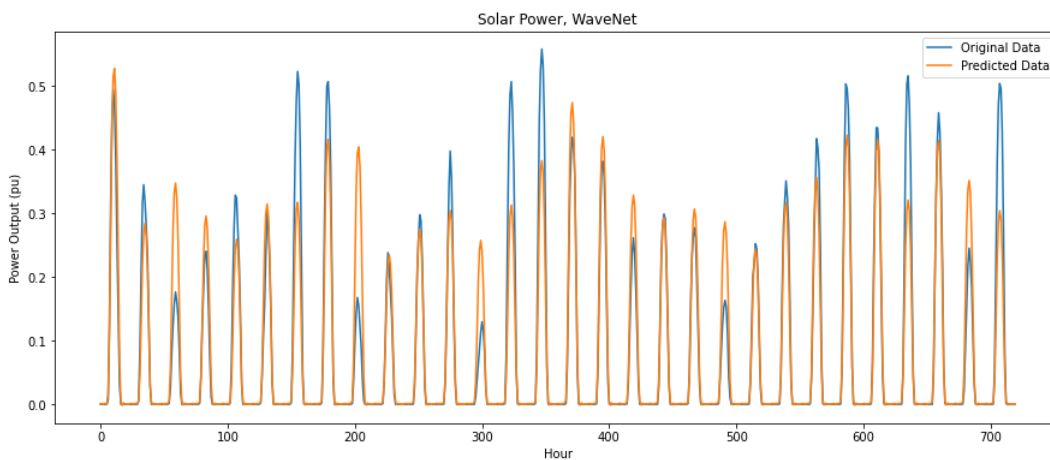
5.2.3 Εφαρμογή WaveNet νευρωνικού δικτύου

Για την σύγκριση των LSTM νευρωνικών δικτύων επιλέχθηκε η δημιουργία ενός δεύτερου μοντέλου, βασισμένου στα νευρωνικά δίκτυα WaveNet. Για την εύρεση των βέλτιστων παραμέτρων του νευρωνικού δικτύου ελέγχθηκαν διαφορετικές παραλλαγές του, όπως φαίνεται στον πίνακα 5.17. Αντίστοιχα με τα LSTM, χρησιμοποιήθηκε το ίδιο μέρος των δεδομένων για να βρεθούν οι βέλτιστες υπερπάρμετροι και διαφορετικά μήκη ιστορικών παρατηρήσεων, τα οποία ήταν τα 24, 48, 96, 192. Σύμφωνα με τα αποτελέσματα των δοκιμών στην χρονοσειρά φωτοβολταϊκής παραγωγής ενός έτους, το βέλτιστο παρελθοντικό μήκος των ιστορικών παρατηρήσεων ορίστηκε σε 96 τιμές. Οι εποχές εκπαίδευσης ορίστηκαν σε 50 όμως μέσω του Early stopping περιορίστηκαν σε 46 για να επιτευχθεί η σύγκλιση του μοντέλου.

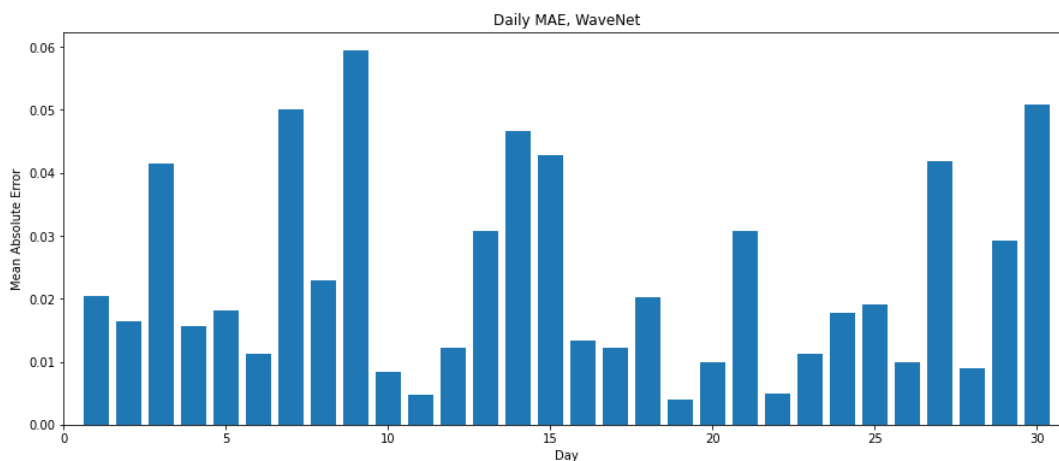
Υπερπαράμετρος	Εναλλακτικές	Βέλτιστη
Αριθμός φίλτρων	[32, 64, 128, 256]	256
Διεσταλμένες συνελίξεις	[1, 2, 4, 8, 16, 32]	[1, 2, 4, 8, 16]
Συνάρτηση ενεργοποίησης	[gated, tanh, ReLU]	gated
Αλγόριθμος βελτιστοποίησης	[RMSprop, Adam]	Adam
batch size	[64, 128, 256]	64

Πίνακας 5.17: Υπερπαράμετροι WaveNet

Εκπαιδεύοντας το μοντέλο με τις βέλτιστες παραμέτρους τα αποτελέσματα των προβλέψεων, όπως και το ημερήσιο μέσο απόλυτο σφάλμα του για τις παρατηρήσεις διάρκειας ενός μήνα φαίνονται στα σχήματα 5.21, 5.22.



Σχήμα 5.21: Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, Wavenet



Σχήμα 5.22: Ημερήσιο σφάλμα πρόβλεψης WaveNet

Ακόμα, τα συνολικά σφάλματα για το σύνολο των παρατηρήσεων φαίνονται στο πίνακα 5.18, όπου παρατηρείται ότι είναι μικρότερα σε σχέση με αυτά της πρόβλεψης μέσω των LSTM

νευρωνικών δικτύων, ενώ ο συντελεστής R^2 μεγαλύτερος, δείχνοντας ότι μπορεί να εντοπιστεί ορθότερα μεγαλύτερο μέρος της διακύμανσης.

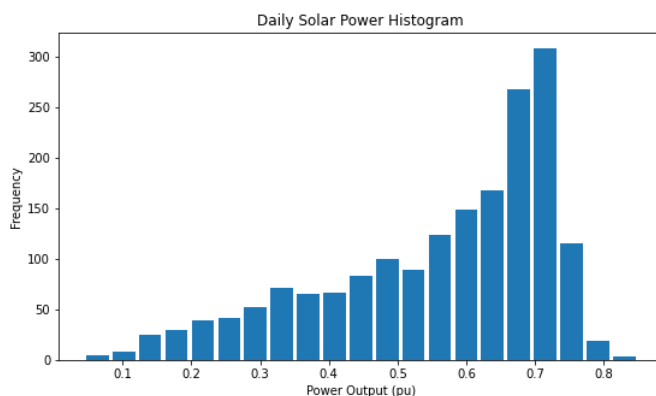
Μοντέλο \ Μετρική	MAE	RMSE	R^2
WaveNet	0.0228	0.0506	0.8652
WaveNet (Non zero values)	0.0517	0.0769	0.6606

Πίνακας 5.18: Συνολικό σφάλμα WaveNet

5.2.4 Εφαρμογή WaveNet νευρωνικού δικτύου με προσθήκη τεχνητών δεδομένων

Συνειδητοποιώντας την αξία εισαγωγής προβλέψεων καιρού για την βελτίωση της πρόβλεψης της παραγωγής αιολικής ενέργειας σε αυτό το σημείο θα προσπαθήσουμε να προσομοιώσουμε την αντίστοιχη τεχνική και στην πρόβλεψη της φωτοβολταϊκής παραγωγής. Πιο συγκεκριμένα, σε πραγματικές συνθήκες μια μέθοδος που θα προήγαγε την ακρίβεια των αποτελεσμάτων, θα ήταν η λήψη προβλέψεων για την ηλιοφάνεια της επόμενης ημέρας (δηλαδή της ημέρας που ζητάμε την πρόβλεψη της φωτοβολταϊκής παραγωγής) από τις περιοχές με την μεγαλύτερη συνεισφορά στην φωτοβολταϊκή παραγωγή με κριτήριο την ονομαστική τους ισχύ. Στην συνέχεια οι προβλέψεις αυτές θα μπορούσαν να σταθμιστούν, ώστε να δημιουργηθεί μια πρόβλεψη για το σύνολο της ηλιοφάνειας, την οποία θα αξιοποιούνταν στα υπάρχοντα μοντέλα. Επειδή ωστόσο, η πρόσβαση σε αυτά τα δεδομένα για την εκπαίδευση των νευρωνικών δικτύων αλλά και τον έλεγχό τους, δεν ήταν εφικτή στα πλαίσια της συγκεκριμένης διπλωματικής εργασίας, θα γίνει η προσπάθεια προσομοίωσης αυτής της πληροφορίας μέσω της πραγματικής παραγωγής της φωτοβολταϊκής παραγωγής. Πιο αναλυτικά, σε αντιστοιχία με την μέθοδο της αιολικής παραγωγής θα ασαφοποιήσουμε την φωτοβολταϊκή παραγωγή της επόμενης ημέρας και συγκεκριμένα την μέγιστη τιμή της, την οποία στην συνέχεια θα δώσουμε ως είσοδο στο νευρωνικό δίκτυο μαζί με τις ιστορικές παρατηρήσεις. Η παραπάνω τεχνική εφαρμόστηκε θεωρώντας ότι μια ποιοτική πρόβλεψη της ηλιακής ακτινοβολίας στο επίπεδο του Ελληνικού συστήματος θα είχε ισχυρή αυτοσυσχέτιση με την τελική παραγωγή, γεγονός που διαπιστώθηκε και από την εξέταση της αιολικής παραγωγής και την αυτοσυσχέτισή της με την ταχύτητα του ανέμου. Έχοντας αυτή την πληροφορία σε μια ασαφή μορφή (καθόλου, λίγη, μέτρια, πολύ ή μέγιστη ηλιοφάνεια), θα μπορούσαμε να βοηθήσουμε τα μοντέλα να προσαρμοστούν και σε απότομες μεταβολές της ηλιοφάνειας κάτι το οποίο μέχρι τώρα δεν ήταν εφικτό.

Για την δημιουργία της ασαφούς πληροφορίας, σχεδιάστηκε το ιστόγραμμα των μέγιστων ημερήσιων παραγωγών ηλιακής ενέργειας για το σύνολο των δεδομένων που έχουμε στην διάθεσή μας, το οποίο απεικονίζεται στο σχήμα 5.23.



Σχήμα 5.23: Ιστογράμμο φωτοβολταϊκής παραγωγής

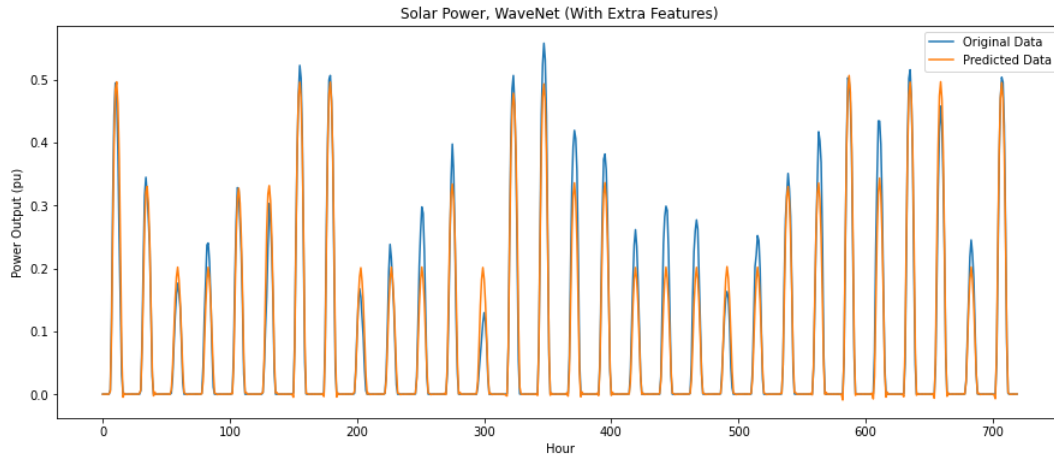
Φαίνεται ότι η φωτοβολταϊκή παραγωγή κυμαίνεται κυρίως σε μέτρια και υψηλά επίπεδα, ενώ οι ημέρες με μικρές παραγωγές είναι αισθητά λιγότερες. Για αυτόν το λόγο, για την μετατροπή των δεδομένων θα εστιάσουμε περισσότερο στις μεγάλες παραγωγές σε σχέση με τις μικρότερες. Μετά από δοκιμές στα διαθέσιμα δεδομένα επιλέξαμε για την ασαφοποίηση των δεδομένων να χρησιμοποιήσουμε τις μετατροπές που φαίνονται στον πίνακα 5.19.

Πραγματική φωτοβολταϊκής παραγωγής (α.μ)	Ασαφής φωτοβολταϊκή παραγωγή (α.μ)
≤ 0.3	0
$(0.3, 0.45]$	0.25
$(0.45, 0.6]$	0.5
$(0.6, 0.7]$	0.75
$(0.7, 1]$	1

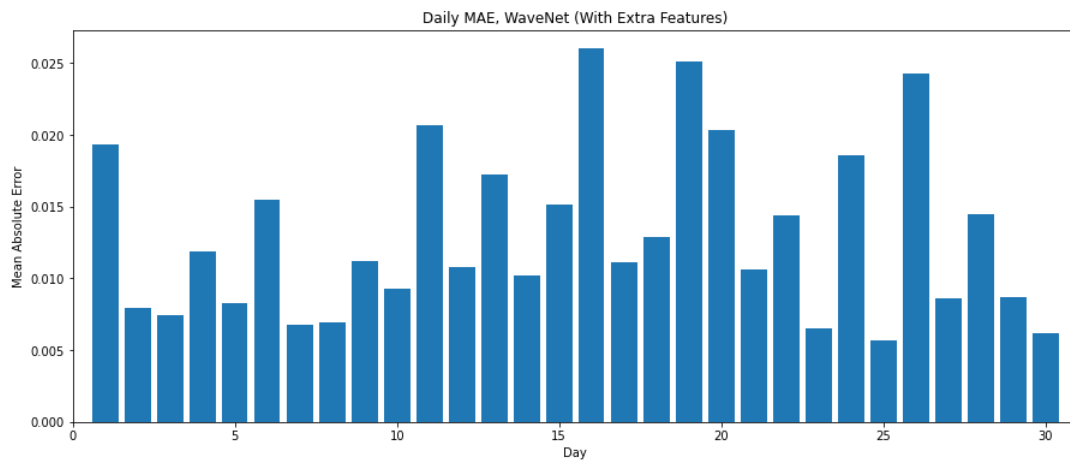
Πίνακας 5.19: Ασαφοποίηση φωτοβολταϊκής παραγωγής

Για την δημιουργία των προβλέψεων, θα χρησιμοποιήσουμε το νευρωνικό δίκτυο WaveNet αφού έδωσε τα καλύτερα αποτελέσματα, με εισόδους του την χρονοσειρά των ιστορικών παρατηρήσεων όπως και στην προηγούμενη περίπτωση, χρησιμοποιώντας επιπλέον και την ασαφή τιμή που θα προκύψει από την μέγιστη τιμή της επόμενης ημέρας αντιστοιχίζοντας την στον πίνακα 5.19. Για την εκπαίδευση θα χρησιμοποιηθούν οι ίδιες παραμέτροι με το αρχικό νευρωνικό δίκτυο WaveNet, ωστόσο μια βασική διαφορά είναι ο χωρισμός των δεδομένων εκπαίδευσης. Αναλυτικότερα ενώ κρατήθηκε το ίδιο σύνολο δεδομένων για τον έλεγχο του μοντέλου, όπως φαίνεται στο σχήμα 5.24, τα ζεύγη των δεδομένων εκπαίδευσης δημιουργήθηκαν ολισθαίνοντας το κινούμενο παράθυρο του σχήματος 4.7 με βήμα 24. Η λογική του διαφορετικού χωρισμού έγκειται στο ότι στόχος του συγκεκριμένου δικτύου είναι η σωστή αντιστοίχιση των παρελθοντικών ιστορικών δεδομένων φωτοβολταϊκής παραγωγής και της τιμής πρόβλεψης της επόμενης ημέρας, με την καμπύλη της φωτοβολταϊκής παραγωγής την επόμενη ημέρα. Έτσι, ως ζεύγη εισόδου-εξόδου επιθυμούμε τις παρελθοντικές ιστορικές τιμές μαζί με την τιμή της πρόβλεψης σε άμεση αντιστοιχία με τις 24 ώρες της επόμενης ημέρας εξαιτίας του επαναλαμβανόμενου μοτίβου που παρατηρείται στην χρονοσειρά. Αντίθετα, στα προηγούμενα μοντέλα

που δεν συμπεριλαμβανόταν η τιμή κάποιας πρόβλεψης, αποδείχθηκε ότι η ολίσθηση του παραθύρου με βήμα ένα είναι πιο αποδοτική, αφού διαθέτουμε στο νευρωνικό δίκτυο περισσότερα δεδομένα εκπαίδευσης. Η απόδοση της τεχνικής, θα πραγματοποιηθεί στα δεδομένα ελέγχου διάρκειας ενός μήνα, όπως φαίνεται στο σχήμα 5.24.



Σχήμα 5.24: Αποτελέσματα πρόβλεψης φωτοβολταϊκής παραγωγής, με επιπλέον χαρακτηριστικά



Σχήμα 5.25: Ημερήσιο σφάλμα πρόβλεψης, με επιπλέον χαρακτηριστικά

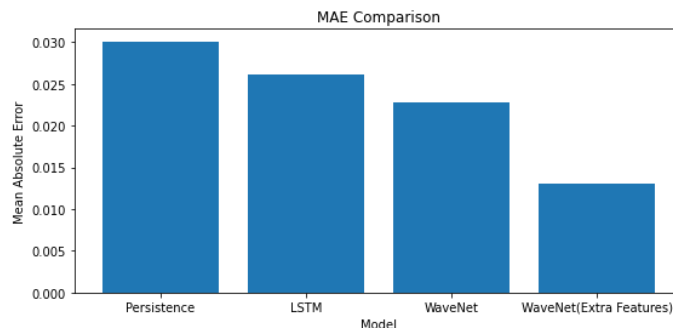
Παρατηρείται όπως ήταν αναμενόμενο, ότι τα σφάλματα έχουν μειωθεί αισθητά σε σχέση με τις προηγούμενες προβλέψεις, ενώ το μοντέλο μας έχει την δυνατότητα προσαρμογής σε διαφορετικές συνθήκες, ανάλογα με τις μεταβολές του καιρού. Έτσι, έχει την δυνατότητα να προβλέψει με μεγάλη ακρίβεια ημέρες χαμηλής παραγωγής έπειτα από ημέρες υψηλής παραγωγής και το αντίθετο. Το παραπάνω επαληθεύεται από τον πίνακα 5.20, όπου φαίνονται οι μετρικές σφάλματος για το σύνολο των δεδομένων που εξετάστηκαν.

Μοντέλο	Μετρική		
	MAE	RMSE	R ²
WaveNet (Extra Features)	0.0131	0.0267	0.9622
WaveNet (Extra Features) (Non zero values)	0.0297	0.0407	0.9216

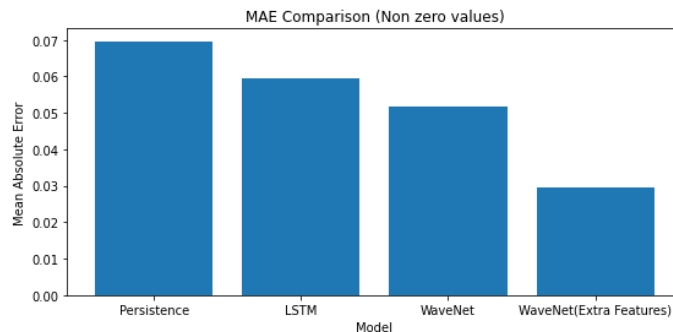
Πίνακας 5.20: Συνολικό σφάλμα

5.2.5 Σύγκριση αποτελεσμάτων

Αφού υλοποιήθηκαν τα διαφορετικά μοντέλα πρόβλεψης της φωτοβολταϊκής παραγωγής, θα γίνει σύγκριση των διαφόρων μετρικών σφάλματος, ώστε να διαπιστωθεί η αποτελεσματικότητα του κάθε μοντέλου. Στα σχήματα 5.26, 5.27, 5.28 φαίνονται οι διαφορετικές μετρικές σφάλματος που χρησιμοποιήθηκαν για το κάθε μοντέλο πρόβλεψης. Στη πρώτη εικόνα απεικονίζονται οι μετρικές σφάλματος κάνοντας χρήση όλων των δεδομένων ελέγχου για τον υπολογισμό τους, ενώ στην δεύτερη χρησιμοποιώντας τις μη μηδενικές τιμές των δεδομένων ελέγχου για την πιο αξιόπιστη αναπαράσταση των σφαλμάτων.

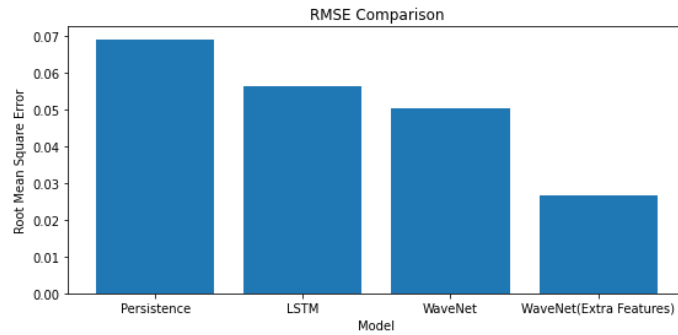


(α) Με χρήση όλων των δεδομένων ελέγχου

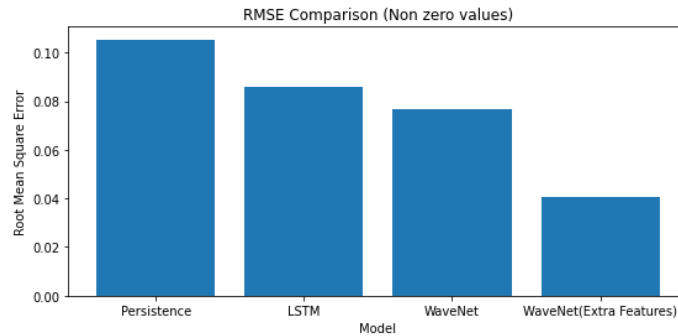


(β) Με χρήση των μη μηδενικών δεδομένων ελέγχου

Σχήμα 5.26: Σύγκριση μέσου απόλυτου σφάλματος

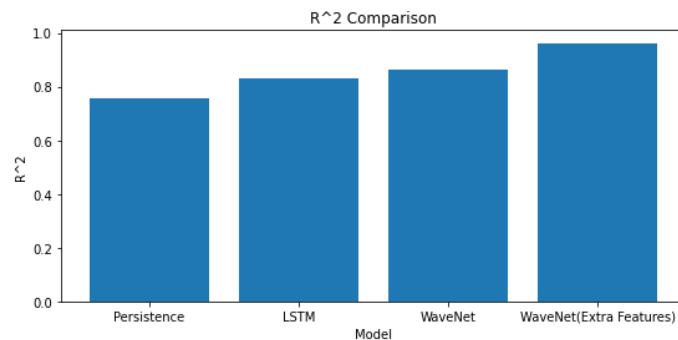


(α') Με χρήση όλων των δεδομένων ελέγχου

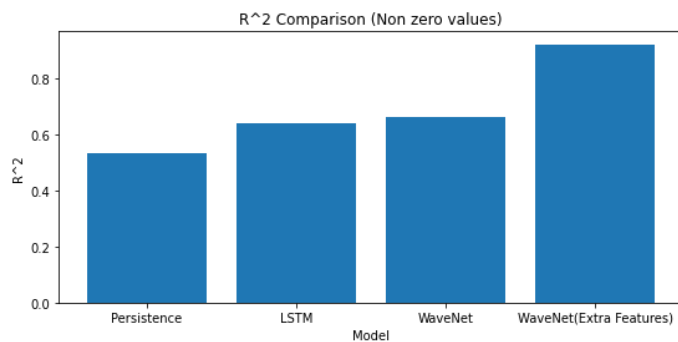


(β') Με χρήση των μη μηδενικών δεδομένων ελέγχου

Σχήμα 5.27: Σύγκριση ρίζας μέσου τετραγωνικού σφάλματος



(α') Με χρήση όλων των δεδομένων ελέγχου



(β') Με χρήση των μη μηδενικών δεδομένων ελέγχου

Σχήμα 5.28: Σύγκριση συντελεστή προσδιορισμού

Η πρώτη παρατήρηση που μπορούμε να κάνουμε, είναι πως όλα μοντέλα παρουσιάζουν μικρότερο μέσο απόλυτο σφάλμα και ρίζα μέσου τετραγωνικού σφάλματος και υψηλότερο συντελεστή R^2 συγκριτικά με το μοντέλο Persistence. Επομένως, έχει νόημα η χρήση τους για την δημιουργία προβλέψεων φωτοβολταϊκής παραγωγής, με χρονικό ορίζοντα 24 ωρών στο μέλλον. Επιπροσθέτως, βλέπουμε ότι το WaveNet, παράγει καλύτερες προβλέψεις σε σχέση με το LSTM νευρωνικό δίκτυο, γεγονός που οφείλεται στην δυνατότητα του να επεξεργάζεται εισόδους μεγαλύτερου μήκους. Με αυτόν το τρόπο, παράγει προβλέψεις λαμβάνοντας υπόψιν τις 4 τελευταίες ημέρες φωτοβολταϊκής παραγωγής, έχοντας την δυνατότητα να προσαρμοστεί καλύτερα σε σχέση με το LSTM νευρωνικό δίκτυο, του οποίου τα καλύτερα αποτελέσματα προέκυψαν με χρήση μόνο της τελευταίας ημέρας ως είσοδο.

Κρίνεται σκόπιμο ωστόσο να αναφερθεί, πως ενώ τα μοντέλα των νευρωνικών δικτύων καταφέρνουν να παράξουν ικανοποιητικά αποτελέσματα για την φωτοβολταϊκή παραγωγή, κανένα από τα δυο μοντέλα δεν έχει την δυνατότητα να προσαρμοστεί σε απότομες μεταβολές της ηλιοφάνειας, όπως ήταν αναμενόμενο. Όπως αποδεικνύουν και τα σχήματα των ημερήσιων σφαλμάτων, τα σφάλματα είναι αισθητά μεγαλύτερα σε τέτοιες περιπτώσεις και η ακρίβεια του μοντέλου μικρότερη. Για την επίλυση του παραπάνω προβλήματος, όπως διαπιστώθηκε, δεν βοηθάει η χρήση κάποιου πιο πολύπλοκου μοντέλου, αλλά πρέπει να γίνει χρήση δεδομένων καιρού και συγκεκριμένα προβλέψεων της ηλιακής ακτινοβολίας, ώστε τα νευρωνικά δίκτυα να μπορούν να προσαρμοστούν και σε τέτοιες δυσμενείς συνθήκες. Με αυτόν το τρόπο, όπως φανερώνουν τα διαγράμματα, τα νευρωνικά δίκτυα ενισχύονται με χρήσιμα χαρακτηριστικά που τους βοηθάνε να βελτιώσουν αισθητά τα αποτελέσματά τους. Συγκεκριμένα, βλέπουμε την σημαντική βελτίωση που προκύπτει με την εισαγωγή μιας ασαφής τιμής στο σύνολο των δεδομένων, καθώς όχι μόνο καταφέρνουμε να μειώσουμε τα σφάλματα, αλλά το νευρωνικό δίκτυο αποκτά την δυνατότητα να συνδυάζει βέλτιστα τις παρελθοντικές ιστορικές τιμές και να προβλέπει με μεγάλη ακρίβεια απότομες αλλαγές της παραγωγής φωτοβολταϊκής παραγωγής.

Τέλος, αξίζει να αναφερθούμε και στην διαφοροποίηση του σφάλματος, ανάλογα με τα δεδομένα που χρησιμοποιήθηκαν. Πιο συγκεκριμένα, φαίνεται πως σε όλες τις προσεγγίσεις οι μετρικές σφάλματος είναι σημαντικά αυξημένες όταν από τα δεδομένα ελέγχου εξαιρούνται οι τιμές που η φωτοβολταϊκή παραγωγή είναι μηδενική. Το παραπάνω ήταν αναμενόμενο, καθώς τα μοντέλα πρόβλεψης μπορούν να διακρίνουν πως κατά τις βραδινές ώρες η παραγωγή ενέργειας είναι πάντα μηδενική και να το προβλέψουν με επιτυχία. Ωστόσο, όταν οι τιμές αυτές συμπεριλαμβάνονται στον υπολογισμό των σφαλμάτων, τα οδηγούν σε χαμηλότερες τιμές αλλοιώνοντας την αξιοπιστία τους. Αυτό συμβαίνει, διότι η αξία της πρόβλεψης της φωτοβολταϊκής παραγωγής έγκειται στην ορθή προσαρμογή των μοντέλων κατά τις πρωινές, μεσημεριανές και απογευματινές ώρες, όπου και η παραγωγή ενέργειας από τις φωτοβολταϊκές εγκαταστάσεις είναι αυξημένη. Παρόλα αυτά, με την εξαίρεση των μηδενικών τιμών, αν και τα σφάλματα είναι αυξημένα, η κατάταξη των μοντέλων είναι η ίδια, με τα WaveNet νευρωνικά δίκτυα να υπερτερούν συγκριτικά με τα LSTM. Επίσης, η εισαγωγή επιπλέον χαρακτηριστικών βλέπουμε ότι βοηθάει σημαντικά στην αποτελεσματικότητα της πρόβλεψης, η οποία έχει αρκετά μικρότερη απόκλιση μεταξύ των δυο προσεγγίσεων (με και χωρίς τη χρήση των μηδενικών τιμών), επιβεβαιώνοντας την υπεροχή της.

5.2.6 Ετήσια πρόβλεψη φωτοβολταϊκής παραγωγής

Αντίστοιχα με την αιολική παραγωγή, εκτός από την επανάληψη της εικοσιτετράωρης πρόβλεψης για μικρούς σχετικά χρονικούς ορίζοντες, όπως αυτός των 30 ημερών, ιδιαίτερης αξίας είναι και η επάκτασή τους. Έτσι, έχοντας διαπιστώσει την υπεροχή του WaveNet νευρωνικού δικτύου, θα χρησιμοποιήσουμε την προηγούμενη αρχιτεκτονική του για την δημιουργία προβλέψεων με χρονικό ορίζοντα 24 ωρών για 365 ημέρες. Με αυτόν τον τρόπο, δεν θα μπορέσουμε μόνο να αξιολογήσουμε τα μοντέλα πρόβλεψης πιο αξιολογικά (λόγω της αύξησης του μεγέθους των δεδομένων ελέγχου) σε όλη την διάρκεια του έτους και των μεταβολών της ηλιοφάνειας, αλλά θα μας δοθεί επίσης η δυνατότητα μελέτης της κατανομής των σφαλμάτων τους, η οποία μπορεί να προσφέρει επιπλέον πληροφορίες για την αποδοτικότητα τους.

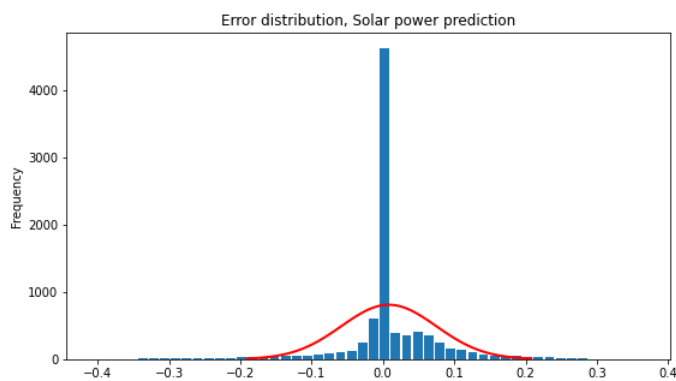
Για την δημιουργία των προβλέψεων θα εκπαιδεύσουμε τα μοντέλα των νευρωνικών δικτύων WaveNet με και χωρίς την χρήση της επιπλέον τιμής ‘πρόβλεψης’. Η εκπαίδευση θα υλοποιηθεί με τις παραμέτρους των νευρωνικών δικτύων που χρησιμοποιήθηκαν για την πρόβλεψη των δεδομένων ελέγχου ενός μήνα, ενώ ο χρονικός ορίζοντας πρόβλεψης θα τεθεί ίσος με 365 προβλέψεις 24 ωρών, δηλαδή 8760 ωρών. Οι μετρικές σφάλματος για πρόβλεψη μέσω του νευρωνικού δικτύου WaveNet με χρήση ως δεδομένων εισόδου μόνο των ιστορικών παρατηρήσεων παρουσιάζεται στον πίνακα 5.21.

Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0348	0.0671	0.9176
WaveNet (Non zero values)	0.0590	0.0878	0.8689

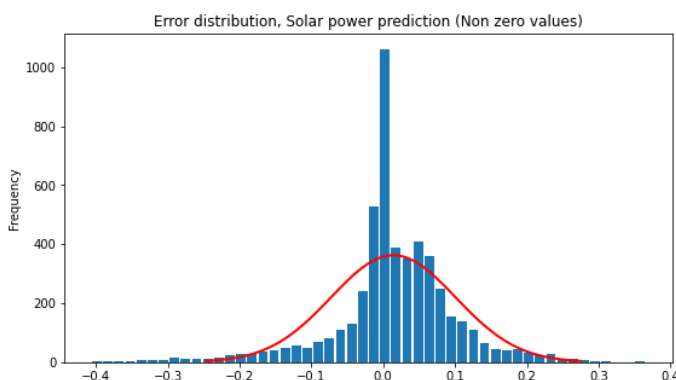
Πίνακας 5.21: Σφάλματα ετήσιας πρόβλεψης

Σε αντιστοιχία με την πρόβλεψη του ενός μήνα, για την φωτοβολταϊκή παραγωγή υπολογίστηκαν οι μετρικές σφάλματος στα πραγματικά δεδομένα, όπως και στα πραγματικά δεδομένα εξαιρώντας τις μηδενικές τιμές. Παρατηρείται, πως τα σφάλματα αν και αυξημένα σε σχέση με την πρόβλεψη του ενός μήνα είναι σε ικανοποιητικό επίπεδο και σε μεγαλύτερους χρονικούς ορίζοντες. Επίσης ο συντελεστής R² είναι αρκετά υψηλός ακόμα και εξαιρώντας τις μηδενικές τιμές, γεγονός που αποδεικνύει πως τα νευρωνικά δίκτυα WaveNet μπορούν να προβλέψουν με αρκετή ακρίβεια τις διακυμάνσεις της φωτοβολταϊκής παραγωγής.

Για τον έλεγχο των αποκλίσεων των προβλέψεων, επιλέχθηκε να σχεδιαστούν και οι κατανομές των σφαλμάτων της πρόβλεψης. Έτσι για N δεδομένα ελέγχου υπολογίζεται ο πίνακας των σφαλμάτων ως $errormatrix = \sum_{i=1}^N y - \hat{y}$ και αποτυπώνεται σε ένα ιστόγραμμα συχνοτήτων. Το ιστόγραμμα συχνοτήτων για τα σφάλματα της ετήσιας πρόβλεψης φαίνεται στο σχήμα 5.29.



(α') Με χρήση όλων των δεδομένων ελέγχου ($\mu = 0.0083$, $\sigma = 0.0666$, $\gamma = -0.6373$, $\kappa = 7.2906$)



(β') Με χρήση των μη μηδενικών δεδομένων ελέγχου ($\mu = 0.0142$, $\sigma = 0.0867$, $\gamma = -0.6967$, $\kappa = 3.2961$)

Σχήμα 5.29: Κατανομή σφάλματος φωτοβολταϊκής παραγωγής

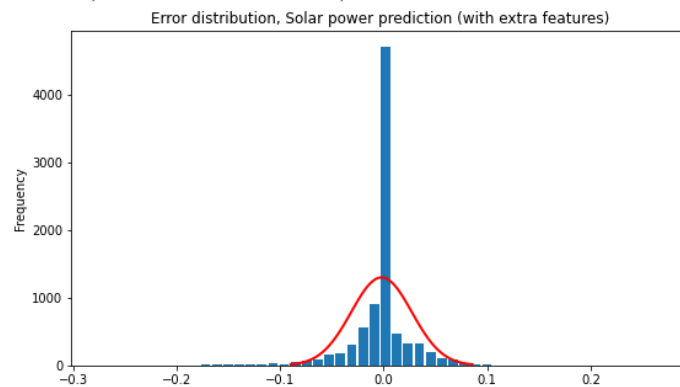
Παρατηρώντας τα δύο διαγράμματα διαπιστώνουμε πως στην κατανομή του σφάλματος που υπολογίστηκε με την χρήση όλων των δεδομένων κυριαρχεί η μηδενική τιμή, κάτι το οποίο ήταν αναμενόμενο λόγω της ορθής πρόβλεψης της μηδενικής παραγωγής ενέργειας των φωτοβολταϊκών κατά τις βραδινές ώρες. Εστιάζοντας στην κατανομή του σφάλματος χωρίς τις μηδενικές τιμές, βλέπουμε ότι εμφανίζει σημαντική αριστερή ασυμμετρία. Το γεγονός αυτό, προκαλεί την συσσώρευση του πλήθους των σφαλμάτων δεξιά του μηδενός, ώστε το μοντέλο να υποεκτιμάει τις προβλέψεις της φωτοβολταϊκής παραγωγής, προβλέποντας συστηματικά τιμές μικρότερες των πραγματικών. Επιπροσθέτως, φαίνεται πως και η τυπική απόκλιση των σφαλμάτων είναι αξιοσημείωτη (παρουσιάζοντας αρκετά αυξημένη τιμή) προσεγγίζοντας την αντίστοιχη των σφαλμάτων της αιολικής παραγωγής. Για την μείωσή της, θα βοηθούσε ενδεχομένως η υιοθέτηση κάποιας άλλης τεχνικής πρόβλεψης, προβλέποντας τα δεδομένα κάθε εποχής ξεχωριστά, για την βελτιστοποίηση της προσαρμογής του νευρωνικού δικτύου σε αυτά.

Συνεχίζοντας ακολουθούμε την ίδια διαδικασία και για την δημιουργία των ετήσιων προβλέψεων, μέσω του μοντέλου WaveNet με την χρήση ως δεδομένων εισόδου, των ιστορικών παρατηρήσεων της φωτοβολταϊκής παραγωγής σε συνδυασμό με την ασαφή τιμή της πρόβλεψη της παραγωγής για την επόμενη ημέρα. Τα αποτελέσματα των μετρικών σφάλματος, όπως και

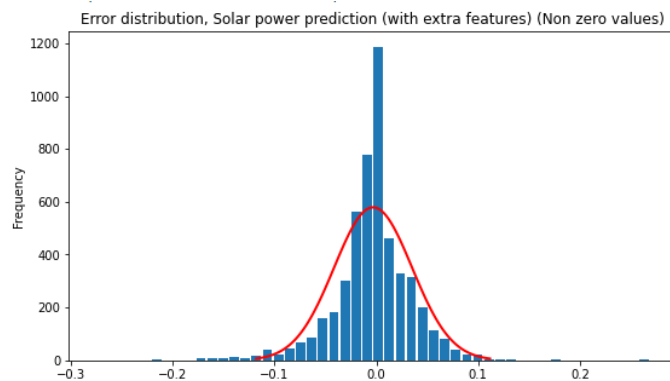
το ιστόγραμμα της κατανομής του σφάλματος του μοντέλου απεικονίζονται στον πίνακα 5.22 και 5.30 αντίστοιχα.

Μοντέλο \ Μετρική	MAE	RMSE	R ²
WaveNet	0.0153	0.0293	0.9843
WaveNet (Non zero values)	0.0258	0.0384	0.9750

Πίνακας 5.22: Σφάλματα ετήσιας πρόβλεψης, με επιπλέον χαρακτηριστικά



(α') Με χρήση όλων των δεδομένων ελέγχου ($\mu = -0.0020$, $\sigma = 0.0293$, $\gamma = -0.7148$, $\kappa = 10.9495$)



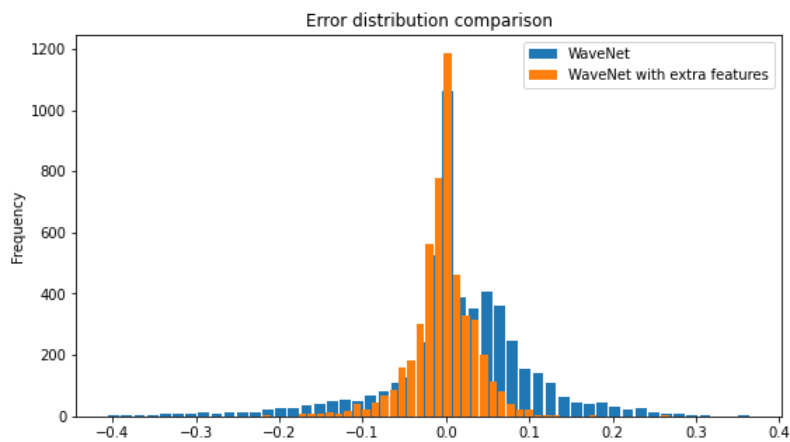
(β') Με χρήση των μη μηδενικών δεδομένων ελέγχου ($\mu = -0.0031$, $\sigma = 0.0383$, $\gamma = -0.4590$, $\kappa = 5.1260$)

Σχήμα 5.30: Κατανομή σφάλματος φωτοβολταϊκής παραγωγής, με επιπλέον χαρακτηριστικά

Διαπιστώνεται ότι η εισαγωγή των επιπλέον χαρακτηριστικών βελτιώνει σε σημαντικό βαθμό την δημιουργία των προβλέψεων, οι οποίες εμφανίζουν μείωση των μετρικών σφαλμάτων τους κατά ποσοστό μεγαλύτερο του 50 % (για τα MAE και RMSE). Επίσης, η κατανομή του σφάλματος προσεγγίζει περισσότερο την κανονική κατανομή, αν και σε αυτή την περίπτωση εμφανίζει αριστερή ασυμμετρία, ενώ επιπλέον μπορεί να χαρακτηριστεί ως λεπτόκυρτη. Αναφορικά με την τυπική απόκλισή της, φαίνεται ότι είναι σημαντικά μειωμένη σε σχέση με αυτή της καμπύλης των προβλέψεων χωρίς την χρήση των επιπλέον δεδομένων εκπαίδευσης.

Το παραπάνω μας οδηγεί στο συμπέρασμα πως για την ορθή αντιμετώπιση της πρόβλεψης της φωτοβολταϊκής παραγωγής ειδικά σε μεγάλους χρονικούς ορίζοντες η ύπαρξη δεδομένων μετεωρολογικών προβλέψεων είναι καίριας σημασίας.

Τέλος στο σχήμα 5.31 παρουσιάζεται η σύγκριση των κατανομών σφάλματος των δύο μοντέλων. Όπως ήδη διαπιστώθηκε, βλέπουμε πως η κατανομή του μοντέλου με την χρήση των επιπλέον δεδομένων ακολουθεί σε μεγαλύτερο βαθμό την κανονική κατανομή, ενώ η τυπική της απόκλιση είναι περιορισμένη σε αξιοσημείωτο βαθμό, σε αντίθεση με αυτή του μοντέλου με την χρήση μόνο ιστορικών παρατηρήσεων, της οποίας οι τιμές εμφανίζουν μεγάλες αποκλίσεις, αρνητικές και θετικές. Το φαινόμενο αυτό μπορεί να εξηγηθεί, καθώς όπως έχει ήδη σχολιαστεί, ημέρες με χαμηλή φωτοβολταϊκή παραγωγή έπειτα από συνεχόμενες ημέρες υψηλής παραγωγής και το αντίθετο δε μπορούν να προβλεφθούν από τα μοντέλα μηχανικής μάθησης, χωρίς την χρήση περαιτέρω δεδομένων καιρού.



Σχήμα 5.31: Σύγκριση κατανομής σφάλματος πρόβλεψης φωτοβολταϊκής παραγωγής

Κεφάλαιο 6

Επίλογος

6.1 Σύνοψη και συμπεράσματα

Στόχος της παρούσας διπλωματικής εργασίας αποτελεί η πρόβλεψη της αιολικής και της φωτοβολταϊκής παραγωγής για την επόμενη ημέρα (24 ώρες).

Για την υλοποίηση των μοντέλων μηχανικής μάθησης και για την πρόβλεψη της αιολικής παραγωγής χρησιμοποιήθηκαν δεδομένα αιολικής παραγωγής και ταχύτητας ανέμου από 3 αιολικά πάρκα μιας ευρύτερης περιοχής, τα οποία αποτελούνταν 210384 παρατηρήσεις με διακριτότητα 10 λεπτών. Αρχικά, επεξεργαστήκαμε τα δεδομένα μετατρέποντάς τα σε ωριαίες παρατηρήσεις και εξαλείφοντας τις θορυβώδεις παρατηρήσεις, βασιζόμενοι στην καμπύλη ισχύος των αιολικών πάρκων καταλήγοντας σε 34084 παρατηρήσεις για κάθε αιολικό πάρκο. Στην συνέχεια, ασαφοποίησαμε τις παρατηρήσεις της ταχύτητας ανέμου, ώστε να μπορούν να χρησιμοποιηθούν σαν πρόβλεψη για την αναμενόμενη ταχύτητα της επόμενης ημέρας. Έπειτα, έχοντας ως σκοπό την πρόβλεψη της αιολικής παραγωγής για το σύνολο της περιοχής, υιοθετήθηκαν οι εξής δυο προσεγγίσεις : η πρόβλεψη του συνόλου της αιολικής παραγωγής της περιοχής (εκπαιδεύοντας τα νευρωνικά δίκτυα με την χρονοσειρά της συνολικής σταθμισμένης παραγωγής αιολικής ενέργειας των πάρκων) και η δημιουργία ενός ξεχωριστού μοντέλου πρόβλεψης για το κάθε πάρκο (με την μετέπειτα στάθμιση τους, για την λήψη του τελικού αποτελέσματος). Η παραγωγή των προβλέψεων πραγματοποιήθηκε με την χρήση βαθιάς μηχανική μάθησης, υλοποιώντας τέσσερα μοντέλα νευρωνικών δικτύων. Δοκιμάστηκαν τα νευρωνικά δίκτυα LSTM, τα νευρωνικά δίκτυα LSTM μέσω της αρχιτεκτονικής Κωδικοποιητή-Αποκωδικοποιητή (Encoder-Decoder), τα συνελικτικά νευρωνικά δίκτυα TCN και τα συνελικτικά νευρωνικά δίκτυα WaveNet. Όλα τους είχαν την δυνατότητα παραγωγής προβλέψεων υψηλής ακρίβειας με το πιο αποδοτικό να είναι το νευρωνικό δίκτυο WaveNet, ωστόσο χωρίς να διαφοροποιείται σημαντικά από τα υπόλοιπα μοντέλα. Ευδιάκριτη διαφοροποίηση παρατηρήθηκε όμως με την αλλαγή της μοντελοποίησης του προβλήματος. Πιο αναλυτικά, σύμφωνα με τα αποτελέσματα η πρόβλεψη του κάθε αιολικού πάρκου ξεχωριστά βελτίωσε σε μεγάλο βαθμό την συνολική ακρίβεια της πρόβλεψης, δίνοντας την δυνατότητα στο μοντέλο να προσαρμοστεί σε ακραίες διακυμάνσεις της παραγωγής. Η αξιοπιστία των μοντέλων δοκιμάστηκε προβλέποντας την παραγωγή αιολική ενέργειας επαναλαμβάνοντας την πρόβλεψη 24 ωρών στο μέλλον, για ένα μήνα και ένα

χρόνο. Συγκρίνοντας τα αποτελέσματα, παρατηρήθηκε πως τα σφάλματα ήταν αντίστοιχα, γεγονός που οφείλεται ενδεχομένως στην μη ύπαρξη πραγματικών αριθμητικών προβλέψεων καιρού (NWP). Για αυτό το λόγο, εξετάστηκε και η προσθήκη επιπλέον θορύβου στις παρατηρήσεις της ταχύτητας ανέμου, προσομοιώνοντας την πρόβλεψη της αιολικής παραγωγής σε αντίστοιχες συνθήκες με τις πραγματικές. Παρατηρήθηκε, ότι αν και τα σφάλματα ήταν αυξημένα όπως και αναμέναμε, τα νευρωνικά δίκτυα είχαν την δυνατότητα να παράξουν προβλέψεις με αρκετή ακρίβεια, ενώ η ξεχωριστή πρόβλεψη του κάθε αιολικού πάρκου εμφάνισε μεγαλύτερη ανοχή στην προσθήκη θορύβου στις προβλέψεις. Αυτός ήταν και ο λόγος που μας οδήγησε στο συμπέρασμα πως είναι και η καταλληλότερη προσέγγιση του προβλήματος της πρόβλεψης αιολική παραγωγής σε μια ευρύτερη περιοχή.

Όσον αφορά την πρόβλεψη της φωτοβολταϊκής παραγωγής επιλέχθηκε η μελέτη της συνολικής παραγωγής της, σε επίπεδο Ελληνικού συστήματος. Στην διάθεσή μας είχαμε τα ιστορικά δεδομένα φωτοβολταϊκής παραγωγής για 5 έτη με ωριαία διακριτότητα, ενώ και σε αυτή την περίπτωση στόχος μας ήταν η δημιουργία προβλέψεων 24 ωρών στο μέλλον. Χρησιμοποιήθηκαν τα νευρωνικά δίκτυα LSTM και WaveNet με εισόδους τις ιστορικές παρατηρήσεις της φωτοβολταϊκής παραγωγής, με τα τελευταία να έχουν την ικανότητα να προβλέψουν με μεγαλύτερη ακρίβεια τις μεταβολές της φωτοβολταϊκής παραγωγής. Επειδή ωστόσο, η ύπαρξη μετεωρολογικών προβλέψεων θα βοηθούσε σημαντικά την αποτελεσματικότητα των μοντέλων, δοκιμάστηκε και η παροχή μιας ένδειξης της μέγιστης ενδεχόμενης φωτοβολταϊκής παραγωγής για την επόμενη μέρα, με τα νευρωνικά δίκτυα να ανταποκρίνονται με εξαιρετική ακρίβεια, ελαχιστοποιώντας τα σφάλματα των προβλέψεων τους. Επιπροσθέτως, οι προβλέψεις για την παραγωγή της επόμενης ημέρας επαναλήφθηκαν για χρονικό ορίζοντα ενός μήνα και ενός έτους, ενώ παρατηρήθηκε, ότι οι προβλέψεις του διαστήματος του ενός έτους παρουσίαζαν μεγαλύτερα σφάλματα, το οποίο αποδίδεται στα διαφορετικά μοτίβα που παρουσιάζει η ηλιοφάνεια κατά την διάρκεια του έτους.

6.2 Μελλοντικές επεκτάσεις

Αναμφίβολα τα τεχνητά νευρωνικά δίκτυα αποτελούν ένα τομέα που βιώνει συνεχή εξέλιξη, με αποτέλεσμα η χρήση τους για την δημιουργία προβλέψεων να προσφέρει την δυνατότητα μελλοντικών επεκτάσεων. Ωστόσο, η βασικότερη επέκταση της παρούσας διπλωματικής θα ήταν ο έλεγχος των αποτελεσμάτων με χρήση πραγματικών μετεωρολογικών προβλέψεων τόσο για την αιολική, αλλά και για την φωτοβολταϊκή παραγωγή.

Αναφορικά με την αιολική παραγωγή, αρχικά χρήσιμη θα ήταν η εξέταση της πρόβλεψης της αιολικής παραγωγής μιας περιοχής με πραγματικές αριθμητικές προβλέψεις καιρού (NWP), ώστε να διαπιστωθεί αν τα μοντέλα που υλοποιήθηκαν μπορούν να αποδώσουν εξίσου μικρά σφάλματα. Στην συνέχεια, ενδιαφέρον θα παρουσίαζε η επέκταση της πρόβλεψης σε μεγαλύτερες περιοχές ή γεωγραφικά διαμερίσματα συγκρίνοντας τις διαφορές που θα παρουσίαζαν οι διαφορετικές προσεγγίσεις πρόβλεψης (προβλέποντας ξεχωριστά ή συνολικά την αιολική παραγωγή). Τέλος ιδιαίτερη αξία θα είχε και η προσπάθεια επέκτασης της πρόβλεψης της αιολικής παραγωγής σε επίπεδο Ελληνικού συστήματος μιας και τέτοιου είδους προβλέψεις

διακατέχουν σημαντικό ρόλο στην ένταξη των μονάδων και χρησιμοποιούνται καθημερινά από τους χειριστές του συστήματος μεταφοράς (TSO).

Σχετικά με την φωτοβολταϊκή παραγωγή, σημαντικός θα ήταν επίσης ο έλεγχος της παραγωγής προβλέψεων, χρησιμοποιώντας πραγματικές καιρικές προβλέψεις. Πιο συγκεκριμένα, θα μπορούσαν να εκμεταλλευτούν προβλέψεις της ηλιοφάνειας από περιοχές της επικράτειας με μεγάλη συνεισφορά στην συνολική φωτοβολταϊκή παραγωγή, οι οποίες στην συνέχεια θα σταθμίζονταν, ώστε να μπορέσουν να αποδώσουν μια συνολική πρόβλεψη της ηλιοφάνειας που ενδέχεται να επικρατήσει στην Ελλάδα για την επόμενη ημέρα.

Κλείνοντας αναφορικά με τα νευρωνικά δίκτυα, διάφορες είναι οι παραλλαγές που θα μπορούσαν να εξεταστούν. Μια από αυτές με ιδιαίτερο ενδιαφέρον θα αποτελούσε η προσθήκη μηχανισμών προσοχής στην αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή των νευρωνικών δικτύων LSTM. Μέσω αυτών, το νευρωνικό δίκτυο θα δεχόταν έναν βάρος σημαντικότητας της κάθε παρατήρησης εισόδου, μαθαίνοντας να εστιάζει την 'προσοχή' του στις παρατηρήσεις που επηρεάζουν περισσότερο την πρόβλεψη. Μια τέτοια προσέγγιση, θα βελτιώνε περαιτέρω την ακρίβεια των προβλέψεων, οδηγώντας στην παραγωγή ακριβέστερων προβλέψεων, λαμβάνοντας υπόψιν τα χαρακτηριστικά των χρονοσειρών που πραγματικά επηρεάζουν τις διακυμάνσεις και τις μελλοντικές μεταβολές τους.

Βιβλιογραφία

- [1] N. B. Adam, M. Elahee, and M. Dauhoo. Forecasting of peak electricity demand in mauritius using the non-homogeneous gompertz diffusion process. *Energy*, 36(12):6763–6769, 2011.
- [2] Article. Gradient descent optimization and challenges. http://primo.ai/index.php?title=Gradient_Descent_Optimization_%26_Challenges.
- [3] S. Bai, J. Z. Kolter, and V. Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.
- [4] N. Chen, Z. Qian, I. T. Nabney, and X. Meng. Wind power forecasts using gaussian processes and numerical weather prediction. *IEEE Transactions on Power Systems*, 29(2):656–665, 2014.
- [5] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [6] Choi, Changkyu. Time series forecasting with recurrent neural networks in presence of missing data. Master’s thesis, UiT Norges arktiske universitet, 2018.
- [7] Christopher Olah. Neural networks, types, and functional programming. <http://colah.github.io/posts/2015-09-NN-Types-FP/>.
- [8] Christopher Olah. Understanding lstm networks. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [9] E. A. DeMeo, G. A. Jordan, C. Kalich, J. King, M. R. Milligan, C. Murley, B. Oakleaf, and M. J. Schuerger. Accommodating wind’s natural behavior. *IEEE power and energy magazine*, 5(6):59–67, 2007.
- [10] H. M. Diagne, P. Lauret, and M. David. Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids. In *WREF 2012-World Renewable Energy Forum*, 2012.

- [11] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.
- [12] Dr. A.W. van der Vaart. Time series notes. <https://www.math.leidenuniv.nl/~avdvaart/timeseries/dictaat.pdf>, 1995-2010.
- [13] Dr. Gesine Reinert. Time series. <http://www.stats.ox.ac.uk/~reinert/time/notesht10short.pdf>, 2010.
- [14] Dr. Marcel Dettling. Applied time series analysis. https://stat.ethz.ch/education/semesters/ss2015/atsa/ATSA_Scriptum_v1_SS15.pdf, 2015.
- [15] Dr. Ross Ihaka . Time series analysis. <https://www.stat.auckland.ac.nz/~ihaka/726/notes.pdf>, 2005.
- [16] "European Commission". *Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Renewable Energy Progress Report, European Commission, Brussels.* European Commission, 2017.
- [17] Francois Chollet. A ten minute introduction to sequence to sequence learning in keras. <https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html>, 2017.
- [18] R. Gençay and M. Qi. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12(4):726–734, 2001.
- [19] L. Han, H. Jing, R. Zhang, and Z. Gao. Wind power forecast based on improved long short term memory network. *Energy*, 189:116300, 2019.
- [20] R. I. Harris. Testing for unit roots using the augmented dickey-fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics letters*, 38(4):381–386, 1992.
- [21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [22] P. Hewage, A. Behera, M. Trovati, E. Pereira, M. Ghahremani, F. Palmieri, and Y. Liu. Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing*, pages 1–30, 2020.

-
- [23] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [24] Y.-Y. Hong and C. L. P. P. Rioflorido. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. *Applied Energy*, 250:530–539, 2019.
- [25] C.-J. Huang and P.-H. Kuo. A short-term wind speed forecasting model by using artificial neural networks with stochastic optimization for renewable energy systems. *Energies*, 11(10):2777, 2018.
- [26] A. Hyndman, R.J. *Forecasting: principles and practice, 2nd edition*. OTexts: Melbourne, Australia, 2018.
- [27] A. K. Jain, J. Mao, and K. M. Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3):31–44, 1996.
- [28] Jayesh Bapu Ahire. Artificial neural model. <https://medium.com/@jayeshbahire/the-artificial-neural-networks-handbook-part-4-d2087d1f583e>.
- [29] Joseph Eddy. High-dimensional time series forecasting with convolutional neural networks: Full-fledged wavenet. https://github.com/JEddy92/TimeSeries_Seq2Seq/blob/master/notebooks/TS_Seq2Seq_Conv_Full.ipynb, 2019.
- [30] R. G. Kavasseri and K. Seetharaman. Day-ahead wind speed forecasting using f-arima models. *Renewable Energy*, 34(5):1388–1393, 2009.
- [31] C. Klessmann, A. Held, M. Rathmann, and M. Ragwitz. Status and perspectives of renewable energy policy and deployment in the european union—what is needed to reach the 2020 targets? *Energy policy*, 39(12):7637–7657, 2011.
- [32] W. Ko, D. Hur, and J.-K. Park. Correction of wind power forecasting by considering wind speed forecast error. *Journal of International Council on Electrical Engineering*, 5(1):47–50, 2015.
- [33] X. Kong, C. Li, C. Wang, Y. Zhang, and J. Zhang. Short-term electrical load forecasting based on error correction using dynamic mode decomposition. *Applied Energy*, 261:114368, 2020.
- [34] D. Kwiatkowski, P. C. Phillips, P. Schmidt, Y. Shin, et al. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics*, 54(1-3):159–178, 1992.
- [35] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme. Temporal convolutional networks applied to energy-related time series forecasting. *Applied Sciences*, 10(7):2322, 2020.
- [36] D. Lew, M. Milligan, G. Jordan, and R. Piwko. Value of wind power forecasting. *Conference*, 4 2011.

- [37] G. Li and J. Shi. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, 87(7):2313–2320, 2010.
- [38] H. Liu, X. Mi, Y. Li, Z. Duan, and Y. Xu. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional gated recurrent unit network and support vector regression. *Renewable Energy*, 143:842–854, 2019.
- [39] M. G. Lobo and I. Sanchez. Regional wind power forecasting based on smoothing techniques, with application to the spanish peninsular system. *IEEE Transactions on Power Systems*, 27(4):1990–1997, 2012.
- [40] C. Mac Domhnaill and L. Ryan. Towards renewable electricity in europe: Revisiting the determinants of renewable electricity in the european union. *Renewable Energy*, 2020.
- [41] F. Makinde, C. Ako, O. Orodu, and I. Asuquo. Prediction of crude oil viscosity using feed-forward back-propagation neural network (ffbpnn). *Petroleum & Coal*, 54(2):120–131, 2012.
- [42] C. B. Martinez-Anido, B. Botor, A. R. Florita, C. Draxl, S. Lu, H. F. Hamann, and B.-M. Hodge. The value of day-ahead solar power forecasting improvement. *Solar Energy*, 129:192–203, 2016.
- [43] P. Murugan. Feed forward and backward run in deep convolution neural network. *arXiv preprint arXiv:1711.03278*, 2017.
- [44] G. Nazaré, R. Castro, and L. R. Gabriel Filho. Wind power forecast using neural networks: Tuning with optimization techniques and error analysis. *Wind Energy*, 23(3):810–824, 2020.
- [45] Z. Niu, Z. Yu, W. Tang, Q. Wu, and M. Reformat. Wind power forecasting using attention-based gated recurrent unit network. *Energy*, 196:117081, 2020.
- [46] Noah Weber. Why lstms stop your gradients from vanishing : A view from the backwards pass. <https://weberna.github.io/blog/2017/11/15/LSTM-Vanishing-Gradients.html>.
- [47] OECD. *OECD Factbook 2015-2016*. OECD Publishing, 2016.
- [48] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
- [49] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318, 2013.

-
- [50] R. A. D. Peter J. Brockwell. *Introduction to Time Series and Forecasting, Second Edition*. Springer, 2002.
- [51] Philippe Remy. Temporal convolutional networks for keras. <https://github.com/philipperemy/keras-tcn>, 2020.
- [52] K. Porter and J. Rogers. Survey of variable generation forecasting in the west: august 2011-june 2012. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2012.
- [53] S. Rajagopalan and S. Santoso. Wind power forecasting and error analysis using the autoregressive moving average modeling. In *2009 IEEE Power & Energy Society General Meeting*, pages 1–6. IEEE, 2009.
- [54] S. Ruder. An overview of gradient descent optimization algorithms. arxiv 2016. *arXiv preprint arXiv:1609.04747*, 2020.
- [55] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [56] Sallam, Abdelhay and Malik, O.P. *Load Forecasting*. WILEY, 10 2018.
- [57] S. Sharma. Activation functions in neural networks. *Towards Data Science*, 6, 2017.
- [58] Simeon Kostadinov. Understanding encoder decoder sequence to sequence model. <https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346>.
- [59] O. Simeone. A very brief introduction to machine learning with applications to communication systems. *IEEE Transactions on Cognitive Communications and Networking*, 4(4):648–664, 2018.
- [60] Simon Haykin. *Νευρωνικά δίκτυα και μηχανική μάθηση*. Εκδόσεις Παπασωτηρίου, 2010.
- [61] M. Solas, N. Cepeda, and J. L. Viegas. Convolutional neural network for short-term wind power forecasting. In *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pages 1–5. IEEE, 2019.
- [62] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium 2010*, pages 1–8. IEEE, 2010.
- [63] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium 2010*, pages 1–8. IEEE, 2010.

- [64] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [65] Stanford University. Convolutional neural networks for visual recognition. <https://cs231n.github.io/convolutional-networks/>.
- [66] Z. Sun, S. Zhao, and J. Zhang. Short-term wind power forecasting on multiple scales using vmd decomposition, k-means clustering and lstm principal computing. *IEEE Access*, 7:166917–166929, 2019.
- [67] S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert systems with applications*, 39(8):7067–7083, 2012.
- [68] Ευάγγελος Σπηλιώτης. Ολοκληρωμένα Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου Όρου (arima). Εθνικό Μετσόβιο Πολυτεχνείο. Σημειώσεις για το μάθημα Τεχνικές Προβλέψεων.
- [69] Φώτιος Περτόπουλος, Βασίλειος Ασημακόπουλος. *Επιχειρησιακές Προβλέψεις*. Εκδόσεις συμμετρία, 2013.
- [70] Ιωάννης Βλαχάβας. Τεχνητά νευρωνικά δίκτυα, 2013. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.
- [71] University of Cincinnati. Feedforward deep learning models. http://uc-r.github.io/feedforward_DNN.
- [72] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 29(3):1033–1044, 2013.
- [73] H.-z. Wang, G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, and Y.-t. Liu. Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied energy*, 188:56–70, 2017.
- [74] Wikipedia. Machine learning. https://en.wikipedia.org/wiki/Machine_learning.
- [75] Wikipedia. Stochastic gradient descent. https://en.wikipedia.org/wiki/Stochastic_gradient_descent.
- [76] Wilamette University. Momentum and learning rate adaption. <https://www.willamette.edu/~gorr/classes/cs449/momrate.html>.
- [77] Q. Xiaoyun, K. Xiaoning, Z. Chao, J. Shuai, and M. Xiuda. Short-term prediction of wind power based on deep long short-term memory. In *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1148–1152. IEEE, 2016.

-
- [78] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.
- [79] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang. Lstm-efg for wind power forecasting based on sequential correlation features. *Future Generation Computer Systems*, 93:33–42, 2019.
- [80] X. Yuan, C. Chen, M. Jiang, and Y. Yuan. Prediction interval of wind power using parameter optimized beta distribution based lstm model. *Applied Soft Computing*, 82:105550, 2019.
- [81] P. Zhao, J. Wang, J. Xia, Y. Dai, Y. Sheng, and J. Yue. Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in china. *Renewable Energy*, 43:234–241, 2012.
- [82] B. Zhou, X. Ma, Y. Luo, and D. Yang. Wind power prediction based on lstm networks and nonparametric kernel density estimation. *IEEE Access*, 7:165279–165292, 2019.