



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**Προσδιορισμός των Επιπέδων Αδράνειας Συστημάτων
Ηλεκτρικής Ενέργειας με Χρήση Μεθόδων Μηχανικής
Μάθησης**

Διπλωματική Εργασία

Θεμιστοκλής Πρόκο

Επιβλέπων: Ελευθέριος Κόντης

Μήνας 2025



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

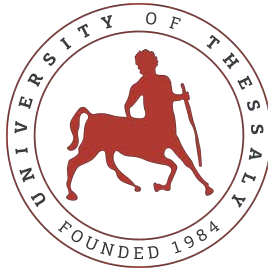
**Προσδιορισμός των Επιπέδων Αδράνειας Συστημάτων
Ηλεκτρικής Ενέργειας με Χρήση Μεθόδων Μηχανικής
Μάθησης**

Διπλωματική Εργασία

Θεμιστοκλής Πρόκο

Επιβλέπων: Ελευθέριος Κόντης

Μήνας 2025



UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Power System Inertia Estimation Using Machine Learning Techniques

Diploma Thesis

Themistoklis Proko

Supervisor: Eleftherios Kontis

Month 2025

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων **Ελευθέριος Κόντης**

Επίκουρος Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος **Φώτιος Πλέσσας**

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος **Ασπασία Δασκαλοπούλου**

Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Ευχαριστίες

Έχοντας πλέον φτάσει στη γραμμή τερματισμού των προπτυχιακών μου σπουδών, θα ήθελα να εκφράσω τη βαθύτατη ευγνωμοσύνη μου σε όσους με στήριξαν σε αυτό το ταξίδι. Πρώτα απ' όλα, ευχαριστώ από τα βάθη της καρδιάς μου την οικογένειά μου, της οποίας η υποστήριξη τόσο οικονομική όσο και ψυχολογική υπήρξε καθοριστική για την ολοκλήρωση των σπουδών μου. Επίσης, θέλω να ευχαριστήσω τους φίλους και συμφοιτητές μου για τις αξέχαστες στιγμές και τις εμπειρίες που μοιραστήκαμε κατά τη διάρκεια αυτών των τεσσάρων ετών στον Βόλο, και για το ότι στάθηκαν δίπλα μου τόσο στις εύκολες όσο και στις δύσκολες στιγμές. Τέλος, οφείλω θερμές ευχαριστίες στον επιβλέποντα καθηγητή μου, κ. Ελευθέριο Κοντή, για την πολύτιμη συνεισφορά του στην εκπόνηση της παρούσας διπλωματικής εργασίας. Η προθυμία του να με καθοδηγεί με σύνεση και υπομονή σε κάθε στάδιο της διαδικασίας αποτέλεσε σημαντικό στήριγμα και έμπνευση για μένα.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

«Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής».

Ο/Η Δηλών/ούσα

Θεμιστοκλής Πρόκο

Διπλωματική Εργασία

Προσδιορισμός των Επιπέδων Αδράνειας Συστημάτων Ηλεκτρικής Ενέργειας με Χρήση Μεθόδων Μηχανικής Μάθησης

Θεμιστοκλής Πρόκο

Περίληψη

Καθώς οι συμβατικές μονάδες παραγωγής ισχύος αποτελούμενες από σύγχρονες γεννήτριες αντικαθίστανται σταδιακά από Ανανεώσιμες Πηγές Ενέργειας (ΑΠΕ) διασυνδεδεμένες μέσω μετατροπέων ισχύος, η αδράνεια του Συστήματος Ηλεκτρικής Ενέργειας (ΣΗΕ) ελαττώνεται σημαντικά, μειώνοντας την ικανότητά του δικτύου να ανταποκρίνεται σε διαταραχές συχνότητας. Στην παρούσα διπλωματική εργασία αναπτύσσονται και αξιολογούνται μοντέλα μηχανικής και βαθιάς μάθησης για την εκτίμηση σε πραγματικό χρόνο των σταθερών αδράνειας ανά περιοχή του διασυνδεδεμένου ΣΗΕ, μέσω συνδυασμού των μετρήσεων συχνότητας ανά περιοχή και ενεργού ισχύος στις διασυνδετικές γραμμές του ΣΗΕ. Για τη δημιουργία δεδομένων εκπαίδευσης και επικύρωσης χρησιμοποιούνται μοντέλα απόκρισης συχνότητας τόσο δύο όσο και τριών περιοχών αντίστοιχα, υλοποιημένα σε MATLAB/Simulink, με μεταβαλλόμενα επίπεδα αδράνειας και μίγματα παραγωγής. Τα μοντέλα πρόβλεψης που αναπτύχθηκαν σε Python, προσδιορίζουν με εξαιρετική ακρίβεια τις τιμές αδρανειών με τα περισσότερα να παρουσιάζουν μέσο απόλυτο ποσοστιαίο σφάλμα κάτω από 4 %. Οι εκτιμήσεις των επιπέδων αδράνειας σε πραγματικό χρόνο επιτρέπουν στους διαχειριστές των συστημάτων μεταφοράς την έγκαιρη εφαρμογή διορθωτικών ενεργειών, όπως για παράδειγμα την άμεση ένταξη σύγχρονων γεννητριών σε περιπτώσεις χαμηλής αδράνειας ώστε να βελτιωθεί η ευστάθεια του συστήματος.

Λέξεις-κλειδιά:

Αδράνεια, απόκριση συχνότητας, μεταβολή ισχύος διασυνδέσεων, περιοχές διασυνδεδεμένου ΣΗΕ, συμβατικές πηγές ενέργειας, ΣΗΕ

Diploma Thesis

Power System Inertia Estimation Using Machine Learning Techniques

Themistoklis Proko

Abstract

As centralized synchronous generators are gradually replaced by converter-based renewable units, the system's rotational inertia has declined, reducing its ability to counter frequency disturbances. This thesis develops and evaluates several machine-learning and deep-learning models capable of real-time estimation of regional inertia constants by optimally fusing wide-area frequency measurements with tie-line active-power flows. Frequency-response models of a multi-area system are built in MATLAB/Simulink to generate training and validation data under varying inertia levels, generation mixes, and region counts. Implemented in Python, the proposed estimators demonstrate excellent performance, achieving mean absolute percentage errors below 4 % across multiple scenarios of generation source mixtures. Real-time inertia estimates enable network operators to dynamically adjust remedial actions to ensure frequency stability of low-inertia power systems.

Keywords:

Conventional power plants, frequency response, interconnected region, power grid, rotational inertia, tie-line active-power flows,

Πίνακας περιεχομένων

Ευχαριστίες	ix
Περίληψη	xii
Abstract	xiii
Πίνακας περιεχομένων	xv
Κατάλογος σχημάτων	xix
Κατάλογος πινάκων	xxiii
1 Εισαγωγή	1
1.1 Αντικείμενο της διπλωματικής	1
1.1.1 Συνεισφορά	2
1.2 Οργάνωση του τόμου	3
2 Η Έννοια της Αδράνειας στα Ηλεκτρικά Συστήματα	5
2.1 Εισαγωγή	5
2.2 Ορισμός αδράνειας	5
2.3 Συνολική αδράνεια περιοχής	7
2.4 Ο ρόλος της αδράνειας σε ένα ΣΗΕ	8
3 Αλγόριθμοι Μηχανικής Μάθησης	13
3.1 Εισαγωγή	13
3.2 Οργάνωση των δεδομένων και προεπεξεργασία	14
3.2.1 Δομή των δεδομένων εισόδου	14
3.2.2 Παράγοντες επίδοσης σε προβλήματα παλινδρόμησης και ταξινόμησης	15

3.2.3	Βήματα προεπεξεργασίας δεδομένων	16
3.2.4	Διαχωρισμός δεδομένων εκπαίδευσης και αξιολόγησης	18
3.2.5	Μετρικές ποσοτικοποίησης σφάλματος	21
3.3	Αλγόριθμοι προσδιορισμού της αδράνειας	23
3.3.1	Πολλαπλή γραμμική παλινδρόμηση	23
3.3.2	Random forest	25
3.3.3	Support vector machines	28
3.3.4	EXtreme gradient boosting	30
3.3.5	Multi-layer perceptron	32
4	Μοντέλα απόκρισης συχνότητας διασυνδεδεμένων ΣΗΕ	39
4.1	Εισαγωγή	39
4.2	Βασικό δομικό διάγραμμα	39
4.2.1	Συνάρτηση Μεταφοράς ΣΗΕ (Power System component)	41
4.2.2	Είσοδος διαταραχής ΔP_d	41
4.2.3	Ελεγκτής (Controller)	45
4.2.4	Υποσυστήματα τύπου Power Plants Area	46
4.2.5	Συνοπτική Περιγραφή Σεναρίων	50
4.2.6	Λοιπές παράμετροι του δικτύου	51
4.2.7	Αριθμητικές Τιμές Παραμέτρων	52
5	Ανάπτυξη Προτεινόμενων Μοντέλων	57
5.1	Εισαγωγή	57
5.2	Κατασκευή δεδομένων για την εκπαίδευση των μοντέλων	57
5.2.1	Κατασκευή δοκιμαστικού σήματος (Probing signal)	59
5.2.2	Εκτίμηση σταθερών αδράνειας	60
5.2.3	Δειγματοληψία και συλλογή δεδομένων	62
5.3	Προ-επεξεργασία δεδομένων (Preprocessing)	64
5.3.1	Εξαγωγή χαρακτηριστικών	65
5.3.2	Απόρριψη Ασταθών σημάτων	66
5.3.3	Επιλογή χαρακτηριστικών (Feature selection)	68
5.3.4	Κλιμάκωση και διαχωρισμός	69

6	Αξιολόγηση των Αποτελεσμάτων	71
6.1	Εισαγωγή	71
6.2	Ενδεικτικά αποτελέσματα - Σενάριο 1	72
6.2.1	Αξιολόγηση αλγορίθμου Random Forest	72
6.2.2	Αξιολόγηση Αλγορίθμου MLP	76
6.2.3	Αξιολόγηση των υπόλοιπων αλγορίθμων μηχανικής μάθησης	79
6.3	Ενδεικτικά αποτελέσματα - Σενάριο 2	81
6.4	Ενδεικτικά αποτελέσματα - Σενάριο 3	83
6.5	Online Προσδιορισμός Αδρανειών	85
7	Επέκταση των μοντέλων σε διασυνδεδεμένο σύστημα 3 περιοχών	87
7.1	Εισαγωγή	87
7.2	Υλοποίηση στο simulink	87
7.2.1	Βασικό δομικό διάγραμμα τριών διασυνδεδεμένων περιοχών. . . .	88
7.2.2	Υποσύστημα intrerconnected lines	89
7.2.3	Υποσύστημα power plants 3	90
7.3	Προετοιμασία δεδομένων για τους αλγορίθμους μηχανικής μάθησης	93
7.3.1	Συλλογή των Δεδομένων	93
7.3.2	Προεπεξεργασία των δεδομένων	94
7.4	Αξιολόγηση των Αλγορίθμων Μηχανικής Μάθησης	97
7.4.1	Αξιολόγηση αλγορίθμου MLR	97
7.4.2	Αξιολόγηση Αλγορίθμου MLP	101
7.5	Αξιολόγηση των Υπόλοιπων Αλγορίθμων	104
8	Συμπεράσματα	107
8.1	Σύνοψη και συμπεράσματα	107
8.2	Μελλοντικές επεκτάσεις	109
	Βιβλιογραφία	111

Κατάλογος σχημάτων

2.1	Η μείωση αδράνειας σε ένα σύστημα ισχύος με υψηλή διείσδυση ηλεκτρονικών ισχύος [1].	6
2.2	Σύνολο γεννητριών που λειτουργούν συγχρονισμένα στο δίκτυο ηλεκτρικής ενεργείας [2].	8
2.3	Απόκριση συχνότητας σε διάφορες χρονικές στιγμές κατά την διάρκεια μιας βηματικής αύξησης της ζητούμενης ισχύος [1].	9
2.4	Αναπαράσταση δυναμικού ισοδυναμίου ενός συστήματος ηλεκτρικής ενεργείας πολλαπλών περιοχών [1].	11
3.1	Στάδια της προεπεξεργασίας των δεδομένων (Data Preprocessing in Machine Learning by PythonGeeks Team)	18
3.2	Διαμέριση του αρχικού συνόλου δεδομένων σε υποπίνακες για εκπαίδευση και επικύρωση How to Split Data into Train and Test Sets in Python with sklearn	20
3.3	Αναπαράσταση ενός δέντρου στην κόσμο των υπολογιστών και της πληροφορικής. [Decision Tree: A Tree-based Algorithm in Machine Learning] . .	26
3.4	Αναπαράσταση λειτουργίας του αλγορίθμου Random Forest. [Random Forest]	28
3.5	Αναπαράσταση λειτουργίας του αλγορίθμου SVR στον διδιάστατο χώρο. [From Theory to Practice: Implementing Support Vector Regression for Predictions in Python]	30
3.6	Αναπαράσταση λειτουργίας του αλγορίθμου XGBoost [3]	31
3.7	Αρχή λειτουργίας ενός νευρώνα του hidden layer	35
3.8	Πλήρως διασυνδεδεμένο νευρωνικό δίκτυο.	36
4.1	Διάγραμμα δυο διασυνδεδεμένων περιοχών έλεγχου	40
4.2	Συνάρτηση μεταφοράς του Power System	41

4.3	Πραγματική μεταβολή της ζητούμενης ισχύος στην περιοχή 1.	42
4.4	Μοντελοποίηση μίτρας τυχαίων διαταραχών εισόδου.	43
4.5	Τελικό σήμα ζητούμενης ισχύος, όπως προκύπτει από τον συνδυασμό της πραγματικής καμπύλης ζήτησης με τη βηματική μεταβολή που προκαλεί η φόρτιση της μπαταρίας.	45
4.6	Υλοποίηση υποσυστήματος ΔP_d	45
4.7	Υλοποίηση υποσυστήματος Controller περιοχής 1	46
4.8	Διάγραμμα βαθμίδων του υποσυστήματος Power Plants Area 1.	47
4.9	Διάγραμμα βαθμίδων του υποσυστήματος Power Plants Area 2.	48
5.1	Παράδειγμα δυναμικών αποκρίσεων για τυχαίο ζεύγος (H_1, H_2)	58
5.2	Κατασκευή δοκιμαστικού σήματος.	60
5.3	Forward mapping (πάνω): $(H_1, H_2) \mapsto (\Delta f_1, \Delta f_2, \Delta P_{tie})$ Inverse mapping (κάτω): $(\Delta f_1, \Delta f_2, \Delta P_{tie}) \mapsto (H_1, H_2)$	61
5.4	Αναπαράσταση τενσορα (3,2,5)	63
5.5	Τρόπος αποθήκευσης σύνολου δεδομένων σε τένσορα	64
5.6	Πινάκας Συσχέτισης (Correlation Matrix)	69
6.1	Απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών των σταθερών H_1, H_2 από το μοντέλο Random Forest.	73
6.2	Θηκογράμματα του απόλυτου σφάλματος για τις προβλέψεις των H_1 και H_2 με Random Forest.	74
6.3	Αθροιστική Συνάρτηση Κατανομής (CDF) του απόλυτου σφάλματος πρόβλεψης των H_1, H_2 με Random Forest.	76
6.4	Απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών των σταθερών H_1, H_2 από το μοντέλο MLP.	77
6.5	Θηκόγραμμα του ζεύγους H_1, H_2 από το μοντέλο MLP.	78
6.6	Συνάρτηση CDF του απόλυτου σφάλματος πρόβλεψης από το μοντέλο MLP για τις σταθερές H_1 και H_2	79
6.7	Σύγκριση επίδοσης των μοντέλων ανά μετρική για το σενάριο 1	81
6.8	Σύγκριση επίδοσης των μοντέλων ανά μετρική για το σενάριο 2	83
6.9	Σύγκριση επίδοσης των μοντέλων ανά μετρική για το σενάριο 3	84
6.10	Online προσδιορισμός ζεύγους αδρανειών	86

7.1	Σενάρια διασύνδεσης τριών περιοχών.	88
7.2	Βασικό block διάγραμμα τριών διασυνδεδεμένων περιοχών	89
7.3	Υλοποίηση υποσυστήματος interconnected lines	90
7.4	Υποσύστημα Power Plants 3.	93
7.5	Πινάκας συσχέτισης (Correlation Matrix) σεναρίου 3 περιοχών.	96
7.6	Απόκλιση πραγματικών από προσδιοριζόμενων τιμών των σταθερών H_1 , H_2, H_3 από το μοντέλο MLR.	99
7.7	Θηκόγραμμα του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLR για τις σταθερές αδράνειας H_1 , H_2 και H_3	100
7.8	Συνάρτηση CDF του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLR για τις σταθερές H_1, H_2 και H_3	100
7.9	Απόκλιση πραγματικών από προσδιοριζόμενων τιμών των σταθερών H_1 , H_2, H_3 από το μοντέλο MLP.	102
7.10	Θηκόγραμμα του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLP για τις σταθερές αδράνειας H_1 , H_2 και H_3	103
7.11	Συνάρτηση CDF του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLP για τις σταθερές H_1, H_2 και H_3	103
7.12	Σύγκριση Επίδοσης των Μοντέλων ανά Μετρική-Σενάριο 3 Περιοχών. . .	106

Κατάλογος πινάκων

2.1	Τυπικές τιμές σταθεράς αδράνειας για διάφορους τύπους σύγχρονων γεννητριών [4]	11
3.1	Δημοφιλέστερες συναρτήσεις ενεργοποίησης [5]	34
4.1	Συνοπτική παρουσίαση παραμέτρων μοντέλων Περιοχής 1 και Περιοχής 2 .	49
4.2	Συνοπτική παρουσίαση των σεναρίων μονάδων παραγωγής στις δύο περιοχές	51
4.3	Πίνακας παραμέτρων σεναρίου 1	53
4.4	Πίνακας παραμέτρων σεναρίου 2	55
4.5	Πίνακας παραμέτρων σεναρίου 3	56
6.1	Αποτελέσματα αξιολόγησης του αλγορίθμου Random Forest για τις εξόδους H_1 και H_2	73
6.2	Αποτελέσματα αξιολόγησης του MLP για τις εξόδους H_1 και H_2	78
6.3	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο 1).	80
6.4	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο 1).	80
6.5	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο 2).	82
6.6	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο 2).	82
6.7	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο 3).	84
6.8	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο 3).	84

7.1	Αριθμητικές τιμές των παραμέτρων της Περιοχής 3 [6].	92
7.2	Αποτελέσματα αξιολόγησης του αλγορίθμου MLR για τις εξόδους H_1 , H_2 και H_3	98
7.3	Αποτελέσματα αξιολόγησης του αλγορίθμου MLP για τις εξόδους H_1 , H_2 και H_3	101
7.4	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο Τριών Περιοχών).	104
7.5	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο Τριών Περιοχών).	104
7.6	Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_3 (Σενάριο Τριών Περιοχών).	105

Κεφάλαιο 1

Εισαγωγή

Αν και η ιδέα της παραγωγής ηλεκτρικής ισχύος από πηγές με ανεξάντλητα αποθέματα καυσίμου και σχεδόν μηδενικό περιβαλλοντικό αποτύπωμα φαντάζει ένα εντυπωσιακό και πολλά υποσχόμενο σενάριο στον τομέα της ενέργειας, η αλήθεια είναι πως η ευρείας κλίμακας διείσδυση των μονάδων ΑΠΕ έχει φέρει στο προσκήνιο μια σειρά από νέες προκλήσεις για την ευστάθεια και τη λειτουργία των συστημάτων ηλεκτρικής ενέργειας (ΣΗΕ). Μία από τις σημαντικότερες επιπτώσεις αυτής της ενεργειακής μετάβασης έγκειται στη μείωση των επιπέδων αδράνειας των σύγχρονων ΣΗΕ. Πράγματι, οι περιστρεφόμενες μάζες, που εντοπίζονται στις σύγχρονες γεννήτριες των συμβατικών μονάδων παραγωγής (όπως για παράδειγμα στις μονάδες ατμοστροβίλων και αεριοστροβίλων), προσδίδουν φυσική αδράνεια στο δίκτυο μέσω της κινητικής τους ενέργειας. Καθώς αυτές οι μονάδες σταδιακά αποσύρονται, τη θέση τους καταλαμβάνουν μονάδες ΑΠΕ, οι οποίες συνδέονται στο δίκτυο μέσω μετατροπών ισχύος και, ως εκ τούτου, δεν συνεισφέρουν άμεσα σε αδράνεια, λόγω της απουσίας μηχανικής σύζευξης με το δίκτυο [1].

1.1 Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εργασία έχει ως στόχο την ανάπτυξη μοντέλων μηχανικής και βαθιάς μάθησης για τον όσο το δυνατόν ακριβέστερο προσδιορισμό της συνολικής αδράνειας ανά περιοχή σε ένα διασυνδεδεμένο ΣΗΕ. Η γνώση της διαθέσιμης αδράνειας σε πραγματικό χρόνο δίνει τη δυνατότητα στον διαχειριστή του δικτύου να λαμβάνει έγκαιρα και τεκμηριωμένα τις κατάλληλες αποφάσεις, ενεργοποιώντας τις απαραίτητες μονάδες παραγωγής, ώστε να μπορεί να εγγυηθεί την ευστάθεια συχνότητας του συστήματος και να αποτρέπει αποτε-

λεσματικά ανεπιθύμητες καταστάσεις, όπως για παράδειγμα εκτεταμένες διακοπές ρεύματος (brown-outs) ή ολικές σβέσεις συστήματος (blackout).

Η συμβολή της εργασίας έγκειται στο γεγονός ότι αξιοποιεί σύγχρονες τεχνικές τεχνητής νοημοσύνης για την πρόβλεψη ενός κρίσιμου αλλά δύσκολα μετρούμενου φυσικού μεγέθους, όπως η αδράνεια. Παρέχει ένα εργαλείο υποστήριξης αποφάσεων για τους διαχειριστές δικτύων, ιδίως σε περιβάλλοντα με υψηλή διείσδυση μονάδων ΑΠΕ, όπου η παραδοσιακή φυσική αδράνεια μειώνεται σημαντικά. Η εργασία επομένως συνεισφέρει τόσο στην ενίσχυση της ευστάθειας των σύγχρονων ενεργειακών συστημάτων όσο και στην ομαλή μετάβαση προς ένα πιο βιώσιμο και «έξυπνο» ενεργειακό μέλλον.

1.1.1 Συνεισφορά

Η συνεισφορά της παρούσας διπλωματικής εργασίας συνοψίζεται στα εξής:

1. Πραγματοποιήθηκε μελέτη διαφόρων συνδυασμών μονάδων παραγωγής σε ένα πρότυπο διασυνδεδεμένο ΣΗΕ δύο περιοχών, υλοποιημένο στο περιβάλλον MATLAB/Simulink.
2. Για κάθε διαφορετικό συνδυασμό τιμών αδράνειας ανά περιοχή (H_1, H_2), προσομοιώθηκε η αντίστοιχη διάταξη του συστήματος στο περιβάλλον Simulink. Σε κάθε προσομοίωση καταγράφηκαν οι δυναμικές αποκρίσεις συχνότητας των δύο περιοχών, καθώς και η δυναμική απόκριση της μεταβολής της ισχύος της διασυνδετικής γραμμής. Συνολικά εκτελέστηκαν 1000 προσομοιώσεις, καθεμία με διαφορετικό ζεύγος τιμών (H_1, H_2), κατά τις οποίες εφαρμόστηκε δοκιμαστικό σήμα (probing signal) μικρής διαταραχής μέσω συστήματος αποθήκευσης μπαταρίας. Το δοκιμαστικό σήμα μοντελοποιείται σε συνδυασμό με την ήδη υπάρχουσα και τυχαία μεταβαλλόμενη ζήτηση ισχύος ανά δευτερόλεπτο. Το 80% των δεδομένων χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων, ενώ το υπόλοιπο 20% αξιοποιήθηκε για επικύρωση των αποτελεσμάτων.
3. Αναπτύχθηκαν και εφαρμόστηκαν συνολικά πέντε αλγόριθμοι μηχανικής μάθησης για την εκτίμηση της συνολικής αδράνειας σε κάθε σενάριο παραγωγής, με στόχο τη συγκριτική αξιολόγηση της ικανότητάς τους να προσεγγίζουν με ακρίβεια τις πραγματικές τιμές αδράνειας του ΣΗΕ. Οι αλγόριθμοι που μελετήθηκαν ήταν οι: Random Forest, Multiple Linear Regression, XGBoost, Support Vector Regression, και Multilayer Perceptron.

4. Η απόδοση κάθε αλγορίθμου αξιολογήθηκε με βάση το σφάλμα πρόβλεψης. Διαπιστώθηκε ότι ο αλγόριθμος MLP παρείχε ιδιαίτερα υψηλή ακρίβεια, με μέσο απόλυτο ποσοστιαίο σφάλμα μικρότερο του 4% στην εκτίμηση των αδρανειών των δύο περιοχών.
5. Η ίδια μεθοδολογία εφαρμόστηκε και σε ένα εκτενέστερο σενάριο με τρεις διασυνδεδεμένες περιοχές και πολλαπλές μονάδες παραγωγής ανά περιοχή, γεγονός που επέφερε αυξημένες απαιτήσεις σε επεξεργασία δεδομένων και έλεγχο γενίκευσης των μοντέλων. Οι εξαιρετικές επιδόσεις των μοντέλων ελαττώθηκαν σημαντικά με μοναδική εξαίρεση το μοντέλο MLP που συνέχισε να προσδιορίζει με εντυπωσιακή ακρίβεια τις τιμές των τριών αυτήν την φορά εξόδων (H_1 , H_2 , H_3) με το σφάλμα να περιορίζεται στο μόλις 4% ανά έξοδο.

1.2 Οργάνωση του τόμου

Στο Κεφάλαιο 2 αναλύεται η έννοια και η σημασία της αδράνειας στα ΣΗΕ. Στο Κεφάλαιο 3, ο αναγνώστης μπορεί να μελετήσει τη λειτουργική αρχή κάθε αλγορίθμου που χρησιμοποιείται για τον προσδιορισμό της αδράνειας. Το Κεφάλαιο 4 εστιάζει στη μοντελοποίηση και την περιγραφή των χρησιμοποιούμενων μοντέλων απόκρισης συχνότητας (System Frequency Response - SFR models), τα οποία υιοθετούνται στα πλαίσια της διπλωματικής εργασίας για την μελέτη απόκρισης συχνότητας διασυνδεδεμένων ΣΗΕ. Στο Κεφάλαιο 5, αναπτύσσεται η προτεινόμενη μεθοδολογία. Έμφαση δίνεται στη συλλογή και προεπεξεργασία των δεδομένων που προέκυψαν από τις προσομοιώσεις των μοντέλων απόκρισης συχνότητας στο Simulink. Στο Κεφάλαιο 6, παρουσιάζεται η αξιολόγηση όλων των μεθόδων εκτίμησης της αδράνειας με τρόπο διαδραστικό και συγκριτικό. Το Κεφάλαιο 7 επεκτείνει την ανάλυση στο πιο σύνθετο σενάριο ενός συστήματος με τρεις διασυνδεδεμένες περιοχές. Τέλος, στο Κεφάλαιο 8 παρουσιάζονται τα συμπεράσματα της μελέτης, ενώ προτείνονται κατευθύνσεις για μελλοντική έρευνα σε συναφή προβλήματα.

Κεφάλαιο 2

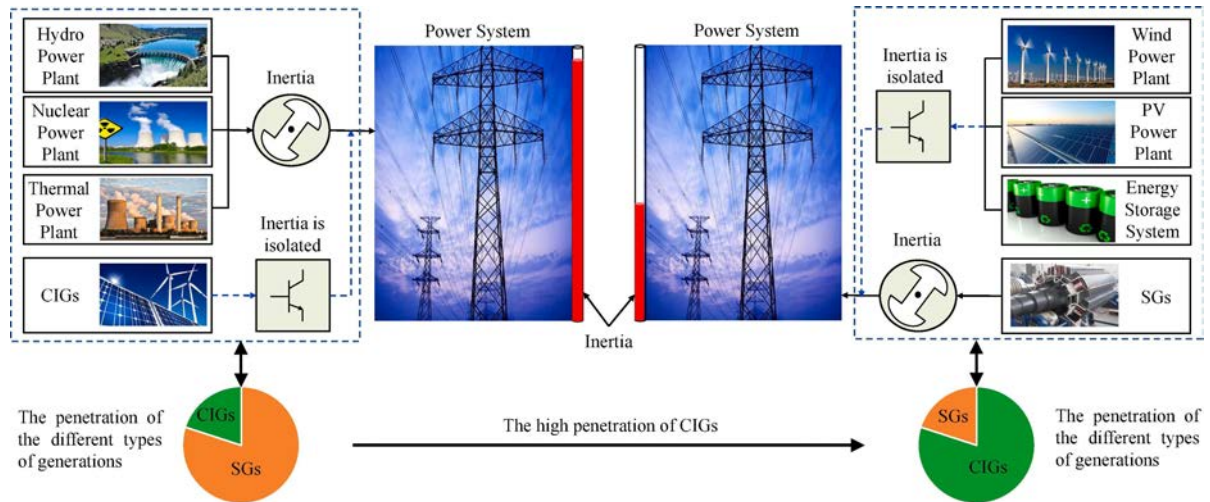
Η Έννοια της Αδράνειας στα Ηλεκτρικά Συστήματα

2.1 Εισαγωγή

Η έννοια της αδράνειας είναι γνωστή από την φυσική ως η τάση των σωμάτων να αντιστέκονται σε οποιαδήποτε μεταβολή της κινητικής τους κατάστασης. Στην περίπτωση των σύγχρονων γεννητριών, η αδράνεια αντικατοπτρίζει την τάση των περιστρεφόμενων μαζών τους να διατηρούν την υπάρχουσα κινητική τους κατάσταση και να συνεχίζουν την περιστροφική τους κίνηση. Η γνώση της αδράνειας σε ένα ΣΗΕ είναι καθοριστικής σημασίας για την ομαλή και αξιόπιστη λειτουργία του. Όσο υψηλότερο το επίπεδο αδράνειας, τόσο μικρότερη η μεταβολή της συχνότητας σε περιπτώσεις διαταραχών του ισοζυγίου παραγωγής και κατανάλωσης [1].

2.2 Ορισμός αδράνειας

Η αδράνεια στα ΣΗΕ απορρέει από την κινητική ενέργεια που είναι αποθηκευμένη στις περιστρεφόμενες μάζες των σύγχρονων γεννητριών, δηλαδή στους στροβίλους και στους δρομείς των μονάδων παραγωγής. Αυτή η φυσική ιδιότητα προσφέρει στο δίκτυο μια άμεση, αυτόματη και παθητική αντίσταση σε απότομες μεταβολές της συχνότητας, όταν συμβαίνουν ξαφνικές ανισορροπίες μεταξύ παραγόμενης και καταναλισκόμενης ισχύος (π.χ. από απώλεια μονάδας ή απότομη μεταβολή φορτίου). Με απλά λόγια, όσο περισσότερη αδράνεια διαθέτει το σύστημα, τόσο πιο αργά αλλά και ομαλά μεταβάλλεται η συχνότητα, δίνοντας



Σχήμα 2.1: Η μείωση αδράνειας σε ένα σύστημα ισχύος με υψηλή διείσδυση ηλεκτρονικών ισχύος [1].

κρίσιμο χρόνο για την ενεργοποίηση των μηχανισμών ελέγχου, όπως η πρωτεύουσα ρύθμιση συχνότητας (primary frequency response). Αντίθετα, σε συστήματα χαμηλής αδράνειας η συχνότητα μπορεί να μεταβληθεί πολύ ταχύτερα (μεγάλο Rate of Change of Frequency - RoCoF), οδηγώντας σε ενεργοποίηση συστημάτων προστασίας, όπως για παράδειγμα από-ζευξη φορτίων λόγω υποσυχνότητας (under frequency load shedding - UFLS) ή ακόμη και σε διακοπές ρεύματος (blackouts) [2]. Καθώς οι παραδοσιακές μονάδες με σύγχρονες γεννήτριες αντικαθίστανται από φωτοβολταϊκά και ανεμογεννήτριες που συνδέονται μέσω μετατροπών ισχύος (και επομένως δεν προσφέρουν φυσική αδράνεια), η συνολική αδράνεια του δικτύου καθίσταται δυναμική και μεταβαλλόμενη (σχήμα 2.1). Αυτό καθιστά απαραίτητη την εκτίμηση της αδράνειας σε πραγματικό χρόνο (online inertia estimation), ώστε να υπάρχει άμεση εικόνα της κατάστασης του συστήματος και να λαμβάνονται έγκαιρα αποφάσεις για την ενίσχυση της ευστάθειας.

Κάθε συμβατική πηγή ενέργειας που διαθέτει περιστρεφόμενη μάζα λειτουργεί συγχρονισμένα με τις υπόλοιπες γεννήτριες του δικτύου και συνεισφέρει στη συνολική τιμή της περιστροφικής αδράνειας κάθε περιοχής, όπως φαίνεται και στο σχήμα 2.2

Η σταθερά αδράνειας κάθε μεμονωμένης σύγχρονης γεννήτριας με περιστρεφόμενη μάζα περιγράφεται από την ακόλουθη σχέση [1]:

$$H_{SG} = \frac{E_k}{S_B} = \frac{\frac{1}{2} \cdot J \cdot \omega_{m0}^2}{S_B} \quad (2.1)$$

Όπου:

- E_k είναι η κινητική ενέργεια του δρομέα (Joule),
- J είναι η ροπή αδράνειας του δρομέα ($\text{kg}\cdot\text{m}^2$),
- ω_{m0} είναι η γωνιακή ταχύτητα περιστροφής (rad/s),
- S_B είναι η ονομαστική ισχύς της γεννήτριας (VA).

2.3 Συνολική αδράνεια περιοχής

Σε ένα διασυνδεδεμένο σύστημα ηλεκτρικής ενέργειας, κάθε περιοχή αποτελείται από ένα σύνολο σύγχρονων μονάδων παραγωγής, οι οποίες διαθέτουν περιστρεφόμενες μάζες και επομένως συνεισφέρουν στην αδράνεια του συστήματος. Η συνολική αδράνεια μιας περιοχής εξαρτάται από τη συνεισφορά όλων των συμβατικών μονάδων που συμμετέχουν στην παράγωγη.

Συγκεκριμένα, όταν n σύγχρονες γεννήτριες λειτουργούν εντός της ίδιας περιοχής, η συνολική αδράνεια H_{total} υπολογίζεται ως η σταθμισμένη μέση τιμή των αδρανειών τους με βάση την ισχύ εξόδου τους [1]:

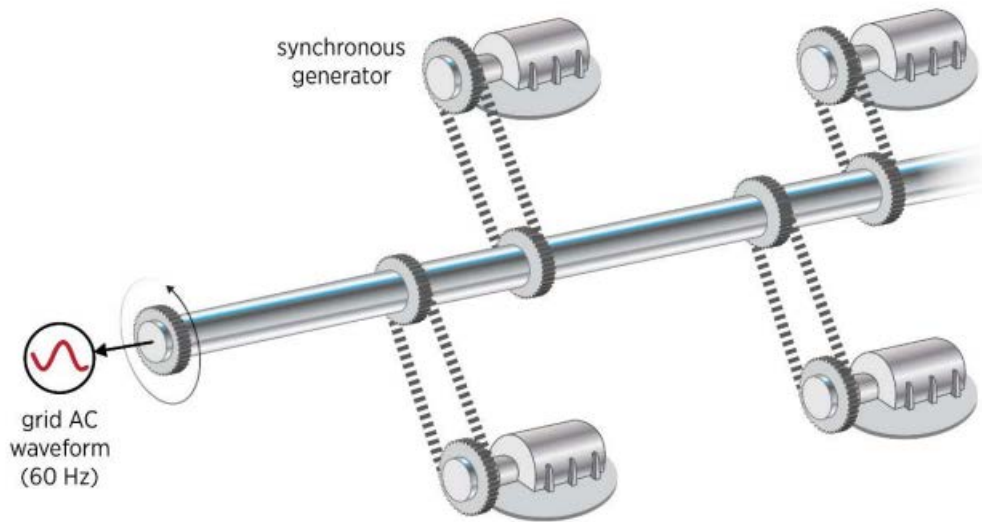
$$H_{\text{total}} = \frac{\sum_{i=1}^n H_i \cdot S_{G,i}}{\sum_{i=1}^n S_{G,i}} \quad (2.2)$$

Όπου:

- H_i είναι η σταθερά αδράνειας της i -οστής γεννήτριας (σε δευτερόλεπτα),
- $S_{G,i}$ η ονομαστική ισχύς της γεννήτριας (σε MVA ή p.u.).

Ο παραπάνω τύπος καταδεικνύει ότι η αδράνεια μιας περιοχής εξαρτάται από τις μονάδες παραγωγής που είναι κάθε χρονική στιγμή ενταγμένες στο ΣΗΕ. Η συνολική αδράνεια μεταβάλλεται λοιπόν δυναμικά ανάλογα με το μείγμα παραγωγής.

Σε περιόδους αυξημένης ζήτησης, η ενεργοποίηση πολλών συμβατικών μονάδων παραγωγής μπορεί να ενισχύσει τα επίπεδα αδράνειας και άρα την ευστάθεια συχνότητας του συστήματος. Αντίθετα, σε ώρες χαμηλού φορτίου η συνολική αδράνεια μπορεί να μειωθεί σημαντικά, ιδίως όταν η κάλυψη του φορτίου βασίζεται σε μονάδες ΑΠΕ, οι οποίες δεν προσφέρουν φυσική αδράνεια. Η αξιολόγηση της αδράνειας ανά περιοχή αποκτά έτσι ιδιαίτερη σημασία στα σύγχρονα ΣΗΕ, καθώς σε πολλές περιοχές κυριαρχούν (ή έχουν σημαντικό



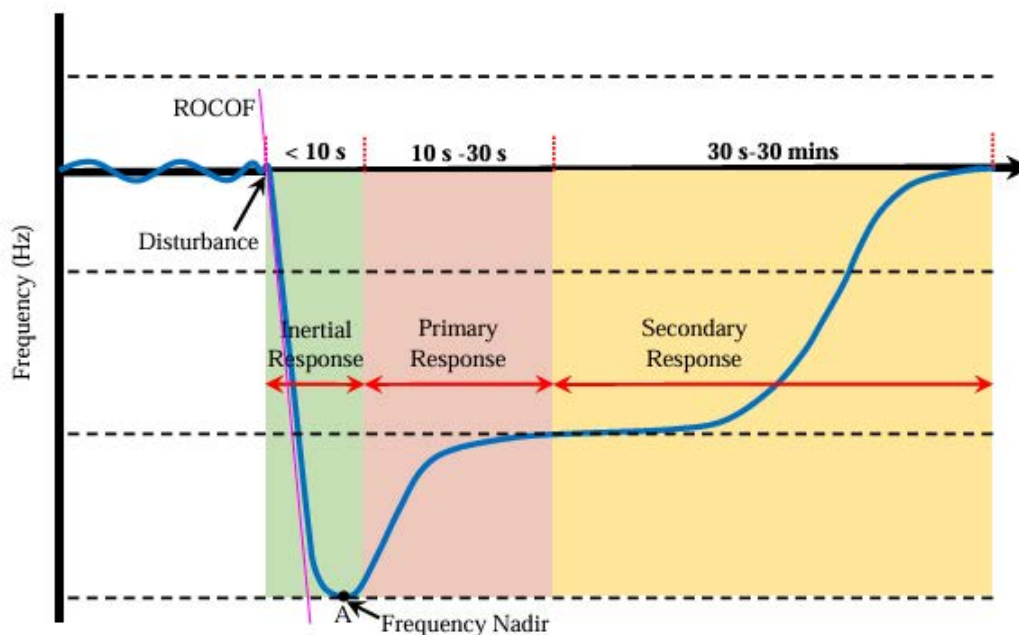
Σχήμα 2.2: Σύνολο γεννητριών που λειτουργούν συγχρονισμένα στο δίκτυο ηλεκτρικής ενέργειας [2].

μερίδιο) σε συγκεκριμένες ώρες της ημέρας μονάδες ΑΠΕ, επηρεάζοντας σημαντικά τη δυναμική συμπεριφορά της συχνότητας του δικτύου.

2.4 Ο ρόλος της αδράνειας σε ένα ΣΗΕ

Τα επίπεδα αδράνειας μίας περιοχής παίζουν καθοριστικό ρόλο στην ασφαλή και αξιόπιστα λειτουργία ενός ΣΗΕ. Κατά τη διάρκεια μιας ανισορροπίας μεταξύ παραγωγής και ζήτησης, διακυμάνσεις στη συχνότητα του συστήματος κάνουν την εμφάνιση τους. Σύμφωνα με τους κανονισμούς ασφαλείας των ευρωπαϊκών διαχειριστών συστημάτων μεταφοράς [7], η συχνότητα πρέπει να διατηρείται εντός του εύρους 49.5 Hz έως 50.5 Hz. Εάν, για οποιονδήποτε λόγο, η συχνότητα υπερβεί αυτά τα όρια, οι επιπτώσεις στο δίκτυο ηλεκτροδότησης μπορεί να είναι σοβαρές ή ακόμη και καταστροφικές [2].

Όταν συμβεί μια αιφνίδια διαταραχή στη ζήτηση ή την παραγωγή ισχύος, ενεργοποιείται μια σειρά μηχανισμών ελέγχου με στόχο τη σταδιακή αποκατάσταση της συχνότητας στην ονομαστική τιμή των 50 Hz, όπως φαίνεται ενδεικτικά στο Σχήμα 2.3, όπου απεικονίζεται ένα σενάριο βηματικής αύξησης της ζητούμενης ισχύος.



Σχήμα 2.3: Απόκριση συχνότητας σε διάφορες χρονικές στιγμές κατά την διάρκεια μιας βηματικής αύξησης της ζητούμενης ισχύος [1].

Όπως φαίνεται Σχήμα 2.3, αμέσως μετά τη διαταραχή, λαμβάνει χώρα η αδρανειακή απόκριση (*inertial response*). Κατά την περίοδο αυτή, μέρος της κινητικής ενέργειας των στρεφόμενων μαζών μετατρέπεται σε ηλεκτρική ισχύ, συμβάλλοντας προσωρινά στην κάλυψη του ελλείμματος ισχύος και στη μείωση της ελάχιστης τιμής συχνότητας (*nadir*), καθώς και του ρυθμού μεταβολής της συχνότητας (*RoCoF* – Rate of Change of Frequency). Αντίστοιχα, στην περίπτωση περίσσειας παραγόμενης ισχύος, η πλεονάζουσα ενέργεια μετατρέπεται σε κινητική ενέργεια, αυξάνοντας τη γωνιακή ταχύτητα των δρομέων των μηχανών.

Μερικά δευτερόλεπτα μετά τη διαταραχή ενεργοποιείται η πρωτεύουσα ρύθμιση συχνότητας, η οποία εκτείνεται τυπικά συνήθως από 10 έως 30 s. Στη φάση αυτή, αναλαμβάνει δράση ο ρυθμιστής στροφών των σύγχρονων γεννητριών, προσαρμόζοντας τη μηχανική ισχύ εισόδου με σκοπό την αντιστάθμιση της διαφοράς μεταξύ παραγωγής και ζήτησης. Συγκεκριμένα, σε περίπτωση ελλείμματος ισχύος, ο ρυθμιστής αυξάνει το άνοιγμα της βαλβίδας εισόδου ατμού (ή καυσίμου), ενώ σε περίπτωση περίσσειας ισχύος τη μειώνει, μεταβάλλοντας ανάλογα τη ροπή του στροβίλου. Με τον τρόπο αυτό, επιτυγχάνεται μια νέα προσωρινή ισορροπία μεταξύ παραγωγής και κατανάλωσης, και η συχνότητα σταθεροποιείται σε μία νέα τιμή, πλησιέστερη στην ονομαστική [4].

Τέλος, σε μεταγενέστερο χρόνο από την ενεργοποίηση της πρωτεύουσας ρύθμισης συχνότητας, ενεργοποιείται και η δευτερεύουσα ρύθμιση συχνότητας, κατά την οποία μέσω της

χρήσης αναλογικών, διαφορικών και ολοκληρωτικών ελεγκτών (PID controllers) αποστέλλονται κατάλληλα σήματα παραγωγής στις μονάδες, ώστε η συχνότητα να επανέλθει στην ονομαστική της τιμή [4].

Η συνεισφορά της αδρανειακής απόκρισης εξαρτάται σε μεγάλο βαθμό από τον αριθμό και το μέγεθος των σύγχρονων γεννητριών που είναι συνδεδεμένες στο ηλεκτρικό δίκτυο κάθε χρονική στιγμή. Όταν λάβει χώρα μια διαταραχή (π.χ. απότομη μεταβολή φορτίου ή απώλεια μονάδας), δημιουργείται μια ανισορροπία ροπής στον δρομέα κάθε γεννήτριας, η οποία οδηγεί σε επιτάχυνση ή επιβράδυνση της περιστροφής της. Ως αποτέλεσμα, μεταβάλλεται η γωνιακή ταχύτητα και επομένως η συχνότητα του συστήματος [2].

Η δυναμική συμπεριφορά κάθε σύγχρονης γεννήτριας, σε ένα σύστημα ηλεκτρικής ενέργειας που αποτελείται από N_{SG} περιστρεφόμενες μηχανές, μπορεί να περιγραφεί από τη λεγόμενη εξίσωση ταλάντωσης (*swing equation* [1]):

$$2H_{SG,i} \cdot \frac{d\Delta\omega_{g,i}}{dt} = \Delta P_{m,i} - \Delta P_{e,i} - D_i \cdot \Delta\omega_{g,i} \quad (2.3)$$

Όπου:

- $H_{SG,i}$ είναι η σταθερά αδράνειας της γεννήτριας i (σε δευτερόλεπτα),
- $\Delta\omega_{g,i}$ είναι η μεταβολή της γωνιακής ταχύτητας της γεννήτριας i (rad/s),
- D_i είναι ο συντελεστής αυτορρύθμισης της γεννήτριας i (σε p.u.),
- $\Delta P_{m,i}$ είναι η μεταβολή της μηχανικής ισχύος εισόδου της γεννήτριας i (σε p.u.),
- $\Delta P_{e,i}$ είναι η μεταβολή της ηλεκτρικής ισχύος εξόδου της γεννήτριας i (σε p.u.).

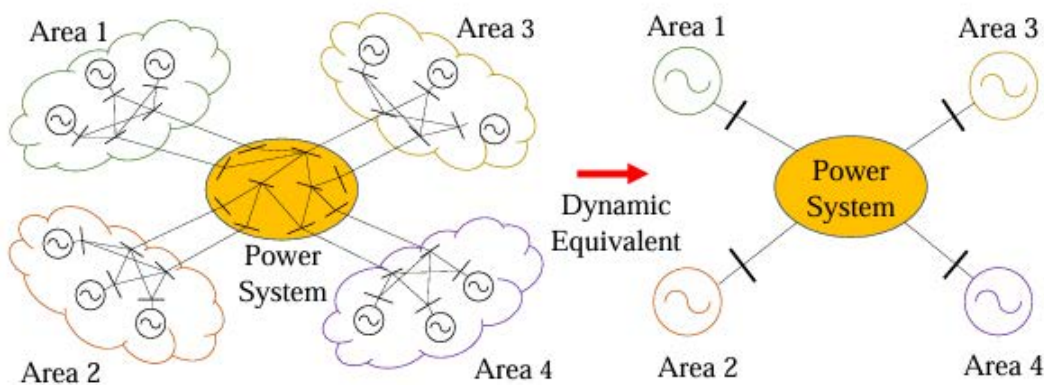
Ανάλογα με τον τύπο της σύγχρονης γεννήτριας και της μονάδας παραγωγής στην οποία ενσωματώνεται, η τιμή της σταθεράς αδράνειας κυμαίνεται μεταξύ 1.75 s και 10 s, όπως φαίνεται στον Πίνακα 2.1

Σε ένα σύστημα ηλεκτρικής ενέργειας που αποτελείται από πολλαπλές σύγχρονες γεννήτριες, μπορεί να οριστεί μία κοινή συχνότητα περιστροφής, γνωστή ως **συχνότητα κέντρου αδράνειας** (ω_{sys}). Η ποσότητα αυτή αντιστοιχεί στη σταθμισμένη μέση γωνιακή ταχύτητα όλων των γεννητριών του συστήματος, όπου το βάρος κάθε γεννήτριας δίνεται από τη σταθερά αδράνειάς της. Με αυτόν τον τρόπο, όλες οι μονάδες μπορούν να αναχθούν σε ένα ισοδύναμο μοντέλο μίας γεννήτριας, διευκολύνοντας τη δυναμική ανάλυση του συστήματος [1].

Πίνακας 2.1: Τυπικές τιμές σταθεράς αδράνειας για διάφορους τύπους σύγχρονων γεννητριών [4]

Τύπος γεννήτριας	Σταθερά αδράνειας (s)
Υδροηλεκτρική γεννήτρια	1.75 – 4.75
Πυρηνική γεννήτρια	4
Θερμική γεννήτρια (ατμοστρόβιλος)	2 – 10

Το συμπέρασμα αυτό έχει κομβική σημασία για την παρούσα εργασία, καθώς πάνω στην συγκεκριμένη θεώρηση θα στηριχθεί η ανάπτυξη των γραμμικών μοντέλων στο Simulink στις επόμενες ενότητες τα οποία θα προσομοιώνουν τη δυναμική συμπεριφορά των γεννητριών. Πιο συγκεκριμένα, το σύνολο όλων των γεννητριών ενός κοινού τύπου ομαδοποιείται και αντιπροσωπεύεται ως ένα ενιαίο μοντέλο. Με αυτήν την προσέγγιση, γίνεται δυνατή η αποτελεσματική διαχείριση της συμπεριφοράς ολόκληρου του δικτύου, αφού αντί να χειριζόμαστε εκατοντάδες μεμονωμένες γεννήτριες, η ανάλυση περιορίζεται σε 3 με 4 ομάδες (οικογένειες) γεννητριών ανά περιοχή (βλέπε σχήμα 2.4).



Σχήμα 2.4: Αναπαράσταση δυναμικού ισοδύναμου ενός συστήματος ηλεκτρικής ενεργείας πολλαπλών περιοχών [1].

Η εξίσωση ταλάντωσης επαναδιατυπώνεται για αυτό το ισοδύναμο σύστημα, λαμβάνοντας υπόψη τη συνολική αδράνεια H_{sys} , την απόκλιση μηχανικής και ηλεκτρικής ισχύος $(\Delta P_m, \Delta P_e)$ και τον συντελεστή απόσβεσης του συστήματος D_{sys} . Η τελική της μορφή φαίνεται στην παρακάτω σχέση [1]:

$$2H_{\text{sys}} \cdot \frac{d\Delta\omega_{\text{sys}}}{dt} = \Delta P_m - \Delta P_e - D_{\text{sys}} \cdot \Delta\omega_{\text{sys}} \quad (2.4)$$

Σε γενικές γραμμές, η συνολική σταθερά αδράνειας ενός συστήματος τείνει να αυξάνεται όσο μεγαλύτερη είναι η ονομαστική ισχύς των γεννητριών που το απαρτίζουν, σύμφωνα με τη σχέση (2.2) πράγμα που δεν ισχύει υποχρεωτικά για μια μεμονωμένη γεννήτρια. Για μια μεμονωμένη μονάδα συμβίωνα με την σχέση (2.1), μια αύξηση της ονομαστικής ισχύος συνεπάγεται συνήθως αύξηση της ροπής αδράνειας J , λόγω των μεγαλύτερων διαστάσεων και μάζας του ρότορα. Παρόλα αυτά, επειδή η αύξηση της ονομαστικής ισχύος (S_n) εμφανίζεται στον παρονομαστή του κλάσματος έχει ως αποτέλεσμα η σχέση μεταξύ της ονομαστικής ισχύος και της σταθεράς αδράνειας μιας μεμονωμένης γεννήτριας να μην είναι ευθέως ανάλογη και να καθορίζεται από το σχέδιο κάθε μηχανής, γεγονός που την καθιστά δύσκολως προβλέψιμη και συγκρίσιμη μεταξύ διαφορετικών μονάδων. Όσο μεγαλύτερη είναι η αδράνεια μιας μονάδας, τόσο μεγαλύτερη είναι και η ικανότητά της να αντιστέκεται σε απότομες μεταβολές ισχύος. Παράλληλα, μπορεί να προσφέρει ή να απορροφήσει μεγαλύτερα ποσά κινητικής ενέργειας για μεγαλύτερη χρονική διάρκεια [2].

Η σταθερά αδράνειας εκφράζεται σε δευτερόλεπτα (s) και υποδηλώνει το χρονικό διάστημα για το οποίο η γεννήτρια είναι ικανή να παρέχει την ονομαστική της ισχύ στο δίκτυο, κάνοντας αποκλειστική χρήση της αποθηκευμένης κινητικής της ενέργειας [2].

Η κινητική ενέργεια του ρότορα λειτουργεί ουσιαστικά ως μια δεξαμενή εφεδρικής ισχύος: αποδίδεται στο σύστημα όταν υπάρχει έλλειμμα και απορροφάται όταν υπάρχει περίσσεια. Όσο μεγαλύτερη είναι η αδράνεια μιας γεννήτριας, τόσο μεγαλύτερη είναι και η "χωρητικότητα" αυτής της δεξαμενής. [8].

Κεφάλαιο 3

Αλγόριθμοι Μηχανικής Μάθησης

3.1 Εισαγωγή

Ο προσδιορισμός των επιπέδων αδράνειας αποτελεί μια ιδιαίτερα πολύπλοκη και χρονοβόρα διαδικασία, ειδικά όταν επιχειρείται ο αναλυτικός υπολογισμός της μέσω των εξισώσεων 2.1 και 2.2. Αυτό οφείλεται στο γεγονός ότι το δίκτυο περιλαμβάνει χιλιάδες σύγχρονες γεννήτριες, οι οποίες ενδέχεται να διαφέρουν σημαντικά μεταξύ τους ως προς τα κατασκευαστικά χαρακτηριστικά τους και την ικανότητα παραγωγής τους. Επιπλέον, η συχνή έλλειψη κρίσιμων τεχνικών στοιχείων και πληροφοριών για τις γεννήτριες του συστήματος καθιστά πρακτικά αδύνατο τον προσδιορισμό της αδράνειας σε πραγματικό χρόνο με τη χρήση αυστηρά μαθηματικών μοντέλων [9].

Για τον λόγο αυτό, έχει επικρατήσει στην πράξη η προσέγγιση του προσδιορισμού της αδράνειας μέσω της ανάλυσης μετρήσεων της συχνότητας και της ισχύος των διασυνδέσεων. Σε αυτό το πλαίσιο, η συμβολή των αλγορίθμων μηχανικής μάθησης και βαθιάς μάθησης είναι καθοριστική, καθώς επιτρέπουν τον προσδιορισμό της ισοδύναμης αδράνειας του δικτύου με υψηλή ακρίβεια [10, 11].

Η διεθνής βιβλιογραφία περιλαμβάνει πληθώρα διαφορετικών αλγορίθμων και προσεγγίσεων για την επίλυση του εν λόγω προβλήματος. Η επιλογή του κατάλληλου αλγορίθμου εξαρτάται από το σύνολο των διαθέσιμων δεδομένων προς επεξεργασία για τον προσδιορισμό της αδράνειας [11, 12, 13, 14, 15].

Στην παρούσα διπλωματική εργασία, προσεγγίσαμε τις τιμές των επιπέδων αδράνειας χρησιμοποιώντας τους εξής αλγορίθμους: Multi-Linear Regression (MLR), Random Forest, XGBoost, Support Vector Regression (SVR), και Multilayer Perceptron (MLP). Στο παρόν

κεφάλαιο αρχικά θα αναλυθεί ο τρόπος οργάνωσης και διαχείρισης των δεδομένων και έπειτα θα περιγραφεί αναλυτικά η λειτουργία και τα βασικά χαρακτηριστικά των προαναφερθέντων αλγορίθμων.

3.2 Οργάνωση των δεδομένων και προεπεξεργασία

Κρίνεται απαραίτητο, πριν προχωρήσουμε σε μια εκτενή περιγραφή πολύπλοκων και συχνά δυσνόητων αλγορίθμων, ο αναγνώστης να έχει κατανοήσει σε βάθος τις θεμελιώδεις έννοιες και τη βασική αρχή λειτουργίας ενός αλγορίθμου μηχανικής μάθησης. Σε γενικές γραμμές, όλες οι τεχνικές της μηχανικής και βαθιάς μάθησης είναι άρρηκτα συνδεδεμένες με βασικούς κλάδους των μαθηματικών, όπως η γραμμική άλγεβρα και η αριθμητική ανάλυση, οι οποίοι εστιάζουν, μεταξύ άλλων, στην επίλυση συστημάτων εξισώσεων μέσω της αξιοποίησης των ιδιοτήτων των πινάκων. Η συμβολή αυτών των μαθηματικών εργαλείων σε επιστημονικά πεδία που σχετίζονται με την ανάλυση δεδομένων, όπως η επεξεργασία φωνής και λόγου, η υπολογιστική όραση και η ανάπτυξη μοντέλων πρόβλεψης, είναι τεράστια. Είναι, συνεπώς, ιδιαίτερος σημαντικό να διαθέτουμε μια σφαιρική εικόνα για τον τρόπο με τον οποίο οργανώνονται τα δεδομένα σε μορφή πινάκων, καθώς και για το είδος της επεξεργασίας που υφίστανται πριν χρησιμοποιηθούν ως είσοδοι σε ένα μοντέλο, το οποίο θα προσδιορίσει τις τιμές εξόδου.

3.2.1 Δομή των δεδομένων εισόδου

Η έξοδος ενός συστήματος προβλέψεων προκύπτει κάθε φορά ως αποτέλεσμα ενός συνόλου παραγόντων, οι οποίοι συμβάλλουν, καθένας με διαφορετικό βαθμό, στον προσδιορισμό της ζητούμενης τελικής τιμής. Οι παράγοντες αυτοί αναφέρονται στη βιβλιογραφία με τον όρο *features* (χαρακτηριστικά), και αποτελούν στην πραγματικότητα την αρχική μορφή της εισόδου που τροφοδοτείται στο μοντέλο. Τα περισσότερα σύνολα δεδομένων (*datasets*) που δημοσιεύονται από εταιρείες, κρατικούς οργανισμούς, αθλητικούς συλλόγους ή φορείς υγείας μπορούν να αναπαρασταθούν ως ένας πίνακας X διαστάσεων $a \times b$. Σε αυτήν τη μορφή, κάθε στήλη ($x_1, x_2, x_3, \dots, x_b$) αντιστοιχεί σε μία συγκεκριμένη χαρακτηριστική ποσότητα (π.χ. δείκτης πληθωρισμού, θερμοκρασία, τυπική απόκλιση, ταχύτητα ανέμου κ.ά.), ενώ κάθε γραμμή ή αλλιώς πλειάδα, όπως συνηθίζεται να λέγεται, περιλαμβάνει μια πλήρη εγγραφή όλων των χαρακτηριστικών που απαιτούνται για τον προσδιορισμό μιας αντίστοιχης

εξόδου. Ο βαθμός επιρροής στη διαμόρφωση της τελικής τιμής εξόδου ποικίλει απο feature σε feature, καθώς ορισμένα χαρακτηριστικά είναι ιδιαίτερα κρίσιμα για την επιτυχία του μοντέλου, ενώ άλλα ενδέχεται να έχουν αμελητέα ή και μηδαμινή συμβολή. Για παράδειγμα, σε ένα μοντέλο ταξινόμησης που αποσκοπεί στην πρόβλεψη του αν μια εταιρεία πρόκειται να χρεοκοπήσει ή όχι, ένα χαρακτηριστικό όπως τα ετήσια έσοδα της εταιρείας ενδέχεται να διαδραματίζει καθοριστικό ρόλο στην πρόβλεψη. Αντιθέτως, άλλα χαρακτηριστικά, όπως ο αριθμός των εργαζομένων ή η ημερομηνία ίδρυσης, ενδέχεται να έχουν μικρότερη σημασία [16].

3.2.2 Παράγοντες επίδοσης σε προβλήματα παλινδρόμησης και ταξινόμησης

Δύο είναι τα βασικά γρανάζια που θα εξασφαλίσουν υψηλή ακρίβεια προβλέψεων σε προβλήματα παλινδρόμησης (regression) και ταξινόμησης (classification): η επιλογή του κατάλληλου μοντέλου και η σωστή προεπεξεργασία των δεδομένων.

Όσον αφορά την επιλογή του κατάλληλου μοντέλου, αυτό θα πρέπει να είναι σε θέση να κατανοήσει τη φύση και τη συμπεριφορά των δεδομένων εισόδου σε σχέση με την έξοδο, ώστε να επιτύχει σωστή γενίκευση του προβλήματος. Στην πλειονότητα των εφαρμογών, είναι σύνηθες να εφαρμόζονται πολλαπλοί αλγόριθμοι μηχανικής μάθησης για την εκτίμηση της ζητούμενης τιμής, καθώς η καταλληλότητα κάθε μεθόδου εξαρτάται σε μεγάλο βαθμό από τη δομή, την κατανομή και την πολυπλοκότητα των δεδομένων. Συνεπώς, δεν είναι εφικτό να γνωρίζουμε εκ των προτέρων ποιος αλγόριθμος θα αποδώσει καλύτερα ή θα κατανοήσει αποτελεσματικότερα τη συμπεριφορά των δεδομένων ([17]).

Τα περισσότερα μοντέλα προβλέψεων είναι προκαθορισμένα και έτοιμα προς χρήση από τον αναλυτή μέσω των διαθέσιμων βιβλιοθηκών της Python, με εξαίρεση κυρίως τα νευρωνικά δίκτυα. Στην περίπτωση αυτή, συχνά απαιτείται εξατομικευμένος σχεδιασμός της αρχιτεκτονικής του δικτύου, όπως η επιλογή του αριθμού και του είδους των επιπέδων (*dense layers*), του πλήθους των νευρώνων, των συναρτήσεων ενεργοποίησης (*activation functions*) και άλλων παραμέτρων.

Το δεύτερο, και ενδεχομένως σημαντικότερο, γρανάζι για την επιτυχία ενός μοντέλου είναι η σωστή προεπεξεργασία των δεδομένων (*preprocessing*). Η προεπεξεργασία αποτελεί μια σειρά από διαδικασίες στις οποίες υποβάλλονται τα αρχικά δεδομένα στη μορφή πίνακα X , πριν αυτά εισαχθούν στα επιλεγμένα μοντέλα. Στόχος αυτών των διαδικασιών είναι η

μετατροπή και ο καθαρισμός του υπάρχοντος dataset, ώστε να καταστεί κατάλληλο για εκπαίδευση και πρόβλεψη.

3.2.3 Βήματα προεπεξεργασίας δεδομένων

Το πρώτο βήμα στην προεπεξεργασία των δεδομένων είναι ο έλεγχος για την ύπαρξη ελλιπών τιμών (*missing values*). Εάν το ποσοστό των ελλιπών τιμών σε μία στήλη είναι εξαιρετικά υψηλό, η συγκεκριμένη στήλη ενδέχεται να αφαιρεθεί πλήρως από τον πίνακα X , καθώς η συνεισφορά της στην πρόβλεψη μπορεί να θεωρηθεί αμελητέα. Σε διαφορετική περίπτωση, οι ελλιπούς τιμές συνήθως αντικαθίστανται με στατιστικές τιμές όπως ο μέσος όρος ή η διάμεσος της αντίστοιχης στήλης, ανάλογα με τη φύση της κατανομής [16].

Ακολουθεί ο εντοπισμός και η διαχείριση των (*outliers*), δηλαδή των τιμών που αποκλίνουν σημαντικά από το σύνολο των υπόλοιπων δεδομένων. Η παρουσία τους μπορεί να επηρεάσει αρνητικά την απόδοση του μοντέλου, ιδιαίτερα σε αλγόριθμους που βασίζονται σε αποστάσεις ή υποθέσεις κανονικότητας. Κοινές μέθοδοι ανίχνευσης περιλαμβάνουν τον Z-score και την ανάλυση IQR (Interquartile Range). Ανάλογα με την περίπτωση, οι outliers είτε τροποποιούνται (*clipping*), είτε απορρίπτονται, είτε αντικαθίστανται με στατιστικά επιτρεπτές τιμές [16].

Μια επίσης αξιοσημείωτη διαδικασία στην οποία υποβαλονται τα δεδομένα κατά την διάρκεια της προεπεξεργασίας είναι η μετατροπή των τιμών των στηλών που περιέχουν categorical data όπως για παράδειγμα η χώρα προέλευσης ενός προϊόντος (Greece, Germany, Italy) σε αριθμητικές τιμές (one-hot, one-cold) αφού η πλειοψηφία των αλγορίθμων μηχανικής μάθησης αποτυγχάνει να διαχειριστεί τέτοιου είδους δεδομένα. Για την αντιμετώπιση αυτού του προβλήματος, χρησιμοποιούνται τεχνικές όπως η one-hot encoding, κατά την οποία δημιουργείται μία νέα δυαδική στήλη για κάθε ξεχωριστή κατηγορία, με τιμές 0 ή 1 για να δηλωθεί η παρουσία ή απουσία της εκάστοτε κατηγορίας. Στο παραπάνω παράδειγμα, η στήλη χώρα θα αντικατασταθεί από τρεις νέες στήλες: χώρα_Greece, χώρα_Germany, χώρα_Italy [16].

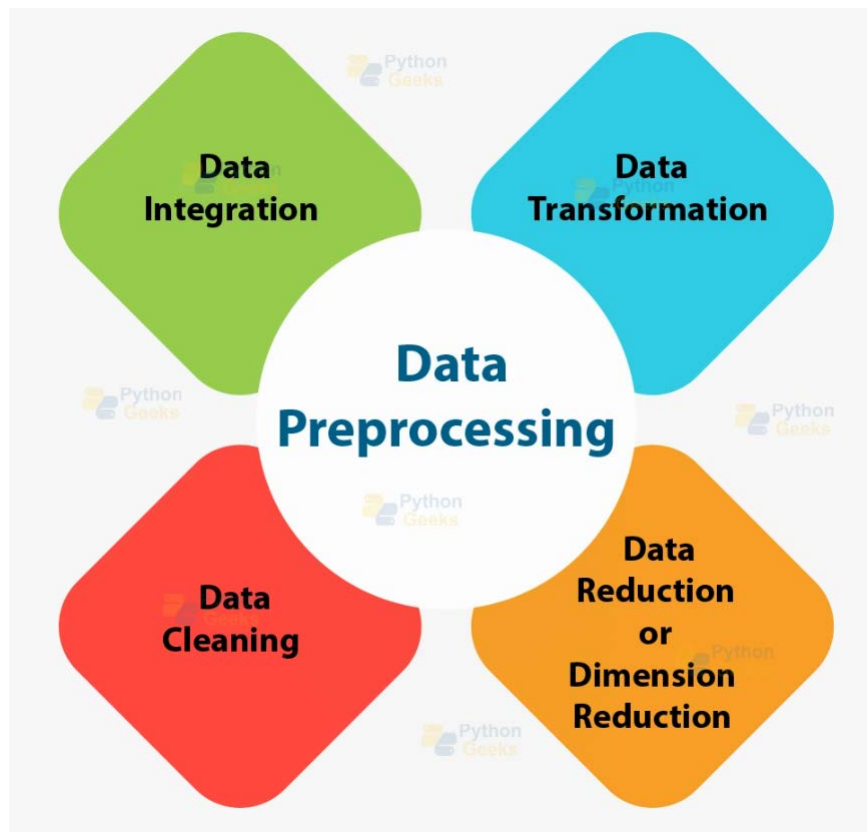
Ένα κρίσιμο στάδιο της προεπεξεργασίας αποτελεί η κλιμάκωση (*scaling*) των αριθμητικών χαρακτηριστικών. Η διαδικασία αυτή εξασφαλίζει ότι τα δεδομένα κάθε στήλης βρίσκονται εντός ενός καθορισμένου εύρους τιμών, συχνά μεταξύ $[-1, 1]$ ή $[0, 1]$, χωρίς να αλλοιώνεται η σχετική διάταξή τους. Η πιο διαδεδομένη τεχνική είναι το *standard scaling*, όπου τα δεδομένα μετασχηματίζονται ώστε να έχουν μέσο όρο μηδέν και τυπική απόκλιση

ιση με ένα. Αυτή η διαδικασία είναι ουσιώδης για τη σωστή λειτουργία των περισσότερων αλγορίθμων μηχανικής και βαθιάς μάθησης [16].

Το τελευταίο στάδιο της προεπεξεργασίας δεν θα μπορούσε να είναι άλλο από το feature selection. Σε αυτό το στάδιο, το ήδη τροποποιημένο σύνολο δεδομένων, όπως έχει διαμορφωθεί από τα προηγούμενα βήματα, μειώνεται ως προς το μέγεθος και την πολυπλοκότητά του, μέσω της απόρριψης χαρακτηριστικών (στηλών) των οποίων η συνεισφορά στην έξοδο του μοντέλου είναι περιορισμένη ή αμελητέα. Η διαδικασία αυτή είναι σημαντική για τη βελτίωση της αποδοτικότητας του μοντέλου, τη μείωση του υπολογιστικού κόστους και την αποφυγή του (*overfitting*). Υπάρχει πλήθος τεχνικών που επιτρέπουν την αξιολόγηση της σημασίας κάθε χαρακτηριστικού ως προς την προβλεπόμενη έξοδο, όπως οι μέθοδοι βάσει στατιστικών μέτρων (*ANOVA*, χ^2), μέθοδοι ενσωματωμένες σε μοντέλα (*feature importances* από *Random Forests*, *Lasso*), ή μεθοδολογίες επαναληπτικής επιλογής (*Recursive Feature Elimination*). Ωστόσο η αναλυτική παρουσίαση των τεχνικών αυτών ξεφεύγει από τους στόχους της παρούσας διπλωματικής εργασίας. Τα βασικότερα σημεία την προεπεξεργασίας δεδομένων απεικονίζονται στο σχήμα 3.1

Ωστόσο, αξίζει να σημειωθεί ότι η σειρά και ο αριθμός των συγκεκριμένων διαδικασιών προεπεξεργασίας ποικίλλουν από πρόβλημα σε πρόβλημα. Σε ορισμένα σύνολα δεδομένων απαιτείται εξαιρετικά ενδελεχής και προσεκτική προεπεξεργασία, ενώ σε άλλα ορισμένα βήματα μπορεί να παραλείπονται εντελώς ή να εφαρμόζονται με διαφορετική σειρά. Στο τέλος της μέρας εξάλλου στον κόσμο των δεδομένων καμιά διαδικασία δεν είναι τυποποιημένη αλλά εξαρτάται πάντα από την φύση των δεδομένων [18].

Υπάρχουν πολυάριθμοι τρόποι υλοποίησης και εκπαίδευσης μοντέλων μηχανικής μάθησης με στόχο τον προσδιορισμό μιας εξαρτημένης μεταβλητής. Οι μέθοδοι αυτοί εξαρτώνται τόσο από τη μορφή και τον όγκο των δεδομένων εισόδου, όσο και από το επίπεδο ακρίβειας που επιδιώκεται σε κάθε εφαρμογή. Σε κρίσιμες περιπτώσεις, όπως η ιατρική διάγνωση ή η πρόβλεψη αστοχίας σε βιομηχανικά συστήματα, απαιτείται εξαιρετικά υψηλή ακρίβεια. Αυτή η απαίτηση συνήθως καλύπτεται με τη χρήση μοντέλων αυξημένης πολυπλοκότητας, όπως τα νευρωνικά δίκτυα, τα οποία όμως απαιτούν σημαντική υπολογιστική ισχύ και μεγαλύτερο χρόνο εκπαίδευσης. Αντιθέτως, όταν το πρόβλημα παρουσιάζει αυξημένη ανοχή στο σφάλμα ή οι πόροι είναι περιορισμένοι, είναι κοινή πρακτική να θυσιάζεται ένα ποσοστό ακρίβειας προς όφελος της ταχύτερης εκπαίδευσης και της χαμηλότερης υπολογιστικής απαίτησης. Σε τέτοιες περιπτώσεις προτιμώνται απλούστερα μοντέλα, όπως η γραμμική πα-



Σχήμα 3.1: Στάδια της προεπεξεργασίας των δεδομένων (Data Preprocessing in Machine Learning by PythonGeeks Team)

λινδρόμηση (linear regression) ή δέντρα απόφασης (decision trees). Ο στόχος σε κάθε περίπτωση παραμένει κοινός: το μοντέλο πρέπει να είναι σε θέση να κατανοήσει τις σχέσεις μεταξύ των μεταβλητών εισόδου (features) και της εξόδου, και να κατασκευάσει κατά την εκπαίδευσή του μια συνάρτηση πρόβλεψης. Η συνάρτηση αυτή οφείλει να αποτυπώνει επαρκώς τη συσχέτιση μεταξύ των ανεξάρτητων και εξαρτημένων μεταβλητών, ώστε το μοντέλο να μπορεί να γενικευθεί σε νέες, άγνωστες τιμές εισόδου που δεν έχει συναντήσει στα δεδομένα εκπαίδευσης [19].

3.2.4 Διαχωρισμός δεδομένων εκπαίδευσης και αξιολόγησης

Το αρχικό σύνολο δεδομένων (*dataset*), αφού έχει προηγουμένως υποβληθεί στη διαδικασία της προεπεξεργασίας, διαιρείται συνήθως σε δύο υποσύνολα: το *training dataset* και το *testing dataset*.

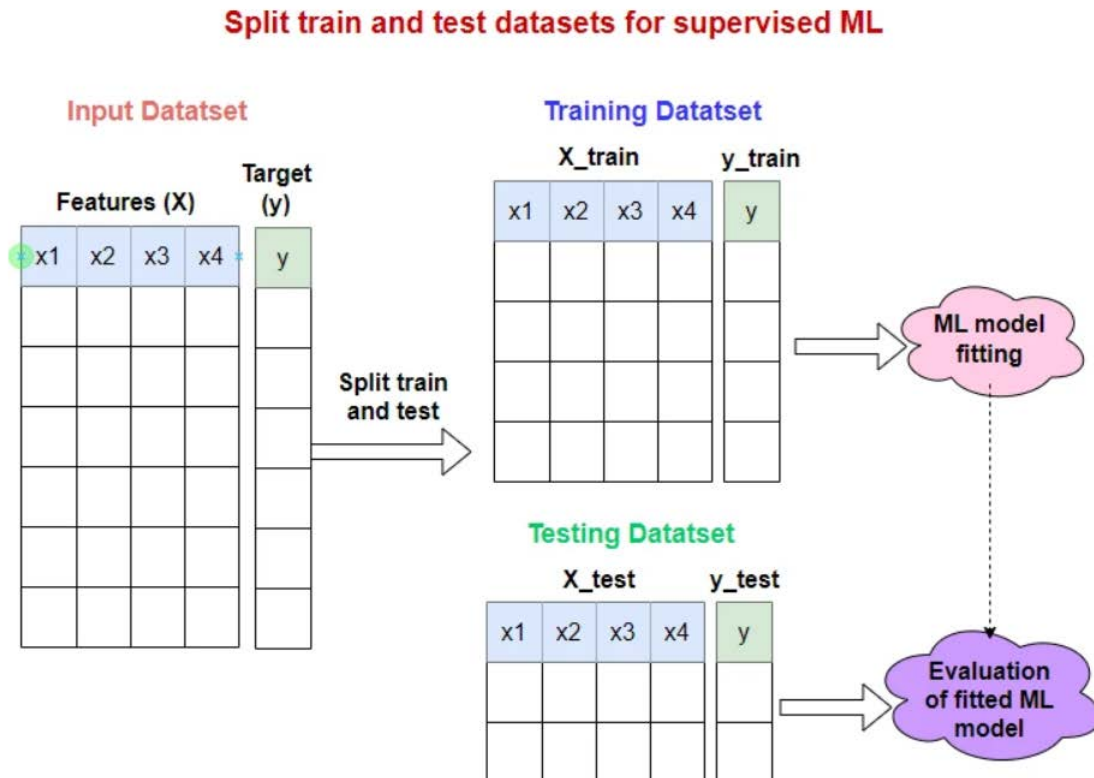
Κατά κανόνα, το *training set* περιλαμβάνει περίπου το 80% του αρχικού συνόλου δεδομένων και χρησιμοποιείται για την εκπαίδευση του μοντέλου, δηλαδή για να μάθει το

μοντέλο να συνδέει τα διάφορα χαρακτηριστικά εισόδου (features) με τις αντίστοιχες τιμές εξόδου. Αντίστοιχα, το *testing set* περιλαμβάνει το υπόλοιπο 20% και χρησιμοποιείται για την αξιολόγηση της ικανότητας του μοντέλου να γενικεύει, δηλαδή, να προβλέπει σωστά τις τιμές εξόδου για δεδομένα εισόδου που δεν έχει «δει» κατά τη διάρκεια της εκπαίδευσης. Είναι ωστόσο σημαντικό να διευκρινιστεί ότι ο διαχωρισμός των δειγμάτων σε *training* και *testing* γίνεται συνήθως τυχαία (random split) από ολόκληρο το σύνολο δεδομένων, και όχι σειριακά. Δηλαδή, δεν ισχύει ότι τα πρώτα 8 στα 10 δείγματα θα περιλαμβάνονται στο *training set* και τα τελευταία 2 στο *testing set* [16]. Κάτι τέτοιο θα μπορούσε να εισάγει χρονική ή δομική μεροληψία. Μία βασική εξαίρεση στον κανόνα αυτό αποτελεί η πρόβλεψη χρονοσειρών (*time series forecasting*), όπου η χρονική σειρά των δεδομένων είναι καθοριστικής σημασίας, και ο προσδιορισμός των πιο πρόσφατων χρονικά τιμών αποτελεί το ζητούμενο. Σε αυτή την περίπτωση, ο διαχωρισμός γίνεται χρονικά διαδοχικά, ώστε να εξασφαλιστεί ότι το μοντέλο προβλέπει το μέλλον βασιζόμενο αποκλειστικά στο παρελθόν.

Αμέσως μετά τον οριζόντιο διαχωρισμό του αρχικού συνόλου δεδομένων σε *training* και *testing* υποσύνολα, ακολουθεί μία δεύτερη, κάθετη αυτή τη φορά, διάσπαση. Στόχος της διαδικασίας είναι ο διαχωρισμός των μεταβλητών εισόδου (ανεξάρτητες μεταβλητές) από τις μεταβλητές εξόδου (εξαρτημένες μεταβλητές), δηλαδή να διακριθούν τα *features* από τα *target values* [16].

Με τον τρόπο αυτό, καταλήγουμε στη δημιουργία τεσσάρων επιμέρους υποπινάκων, οι οποίοι προκύπτουν από τον αρχικό πίνακα δεδομένων. Σύμφωνα με τη βιβλιογραφία, ο πίνακας των μεταβλητών εισόδου συμβολίζεται συνήθως με X , ενώ ο πίνακας των μεταβλητών εξόδου με y . Συνεπώς, τα τέσσερα υποσύνολα που δημιουργούνται είναι τα: X_{train} , y_{train} , X_{test} , y_{test} . Αυτός ο διαχωρισμός επιτρέπει την ανεξάρτητη επεξεργασία των εισόδων και εξόδων και αποτελεί τυπική πρακτική σε κάθε πρόβλημα (*supervised learning*) [16].

Η βασική φιλοσοφία που διέπει την πλειοψηφία των αλγορίθμων μηχανικής και βαθιάς μάθησης βασίζεται στην αρχική υιοθέτηση μιας μαθηματικής συνάρτησης με τυχαία αρχικοποιημένες παραμέτρους (π.χ. βάρη και σταθερές), και στη σταδιακή τροποποίηση αυτών των παραμέτρων μέσα από επαναληπτική διαδικασία (*epochs*), με στόχο τη μείωση του σφάλματος μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Πιο συγκεκριμένα, κατά την πρώτη επανάληψη, το μοντέλο ξεκινά με μια συνάρτηση πρόβλεψης η οποία συνδέει τις μεταβλητές εισόδου (X_{train}) με τις αντίστοιχες τιμές εξόδου (y_{train}), χρησιμοποιώντας παραμέτρους που έχουν επιλεγεί τυχαία. Οι εκτιμήσεις που προκύπτουν καταγράφονται σε έναν



Σχήμα 3.2: Διαμέριση του αρχικού συνόλου δεδομένων σε υποπίνακες για εκπαίδευση και επικύρωση How to Split Data into Train and Test Sets in Python with sklearn

νέο πίνακα προβλέψεων, συνήθως συμβολιζόμενο ως y_{pred} , ο οποίος ανανεώνεται σε κάθε επανάληψη. Λόγω της τυχαίας αρχικοποίησης, οι προβλεπόμενες τιμές στην πρώτη επανάληψη αποκλίνουν σημαντικά από τις πραγματικές τιμές. Μετά το πέρας κάθε επανάληψης, οι παράμετροι του μοντέλου ενημερώνονται μέσω αλγορίθμων βελτιστοποίησης, όπως ο *Gradient Descent*, με σκοπό τη σταδιακή μείωση του σφάλματος. Η διαδικασία αυτή επαναλαμβάνεται για αρκετές επαναλήψεις, έως ότου το σφάλμα μεταξύ των y_{pred} και y_{train} κριθεί μικρό σε ικανοποιητικό βαθμό. Τότε, θεωρείται ότι το μοντέλο έχει κατανοήσει τη σχέση εισόδων-εξόδων και είναι έτοιμο να εφαρμοστεί σε νέα, άγνωστα δεδομένα, των οποίων οι τιμές εξόδου δεν είναι γνωστές εκ των προτέρων [16].

Έχοντας πλέον ολοκληρώσει τη διαδικασία εκπαίδευσης του εκάστοτε μοντέλου, ακολουθεί το στάδιο της αξιολόγησης, κατά το οποίο εξετάζεται κατά πόσο το μοντέλο είναι ικανό να προβλέπει με ακρίβεια τις τιμές εξόδου για δεδομένα εισόδου που δεν έχει «δεν» ποτέ στο παρελθόν. Ο στόχος είναι να επιβεβαιωθεί ότι το μοντέλο δεν έχει απλώς απομνημονεύσει τις τιμές του y_{train} , αλλά αντίθετα έχει μάθει την υποκείμενη σχέση μεταξύ εισόδου και εξόδου και μπορεί πλέον να γενικεύει. Σε αυτό ακριβώς το στάδιο γίνεται ουσιαστική

η χρήση του testing dataset, το οποίο είχε αναφερθεί στην προηγούμενη ενότητα. Εάν η τελική μορφή της μαθηματικής συνάρτησης που διαμορφώθηκε κατά την εκπαίδευση όντως περιγράφει επαρκώς τη σχέση μεταξύ μεταβλητών εισόδου και εξόδου, τότε το μοντέλο αναμένεται να προβλέψει με ακρίβεια τις τιμές εξόδου y_{pred} , έχοντας ως είσοδο τις τιμές του πίνακα X_{test} [16].

Το ερώτημα που τίθεται σε αυτό το στάδιο είναι το εξής: *πώς γνωρίζουμε αν οι τιμές εξόδου που παρήγαγε το μοντέλο είναι όντως σωστές*; Η απάντηση βρίσκεται στον πίνακα y_{test} , ο οποίος περιέχει τις πραγματικές τιμές εξόδου για το testing dataset. Αν και το μοντέλο δεν έχει πρόσβαση στο y_{test} κατά τη διάρκεια της εκπαίδευσης ή της πρόβλεψης, οι τιμές αυτές είναι γνωστές στο χρήστη. Έτσι, συγκρίνοντας τις προβλεπόμενες τιμές y_{pred} , όπως αυτές υπολογίστηκαν από το εκπαιδευμένο μοντέλο βάσει του X_{test} , με τις αντίστοιχες πραγματικές τιμές του y_{test} , μπορούμε να εκτιμήσουμε την ακρίβεια των προβλέψεων. Η διαφορά μεταξύ των δύο αυτών πινάκων, δηλαδή το σφάλμα πρόβλεψης, αποτελεί το βασικό μέγεθος αξιολόγησης της απόδοσης του μοντέλου. Όλα όσα ειπωθήκαν σε αυτήν την παράγραφο συνοψίζονται στην εικόνα 3.2.

3.2.5 Μετρικές ποσοτικοποίησης σφάλματος

Ανάλογα με το είδος του προβλήματος (παλινδρόμηση ή ταξινόμηση), χρησιμοποιούνται κατάλληλες μετρικές για την ποσοτικοποίηση του σφάλματος. Στα περισσότερα προβλήματα παλινδρόμησης, όπως και στο πλαίσιο της παρούσας διπλωματικής εργασίας, γίνεται χρήση τεσσάρων τέτοιων μετρικών. Πιο συγκεκριμένα, χρησιμοποιούνται οι Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) και ο συντελεστής προσδιορισμού R^2 .

Περιγραφή μετρικών παλινδρόμησης

- **Mean Absolute Error (MAE)**

Εκφράζει τη μέση απόλυτη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών, χωρίς να λαμβάνεται υπόψη το πρόσημο της διαφοράς. Αν και πρόκειται για μία από τις απλούστερες μετρικές ως προς την ερμηνεία της, αποτελεί μία από τις πιο ευρέως χρησιμοποιούμενες μεθόδους αξιολόγησης σφάλματος σε προβλήματα

παλινδρόμησης. Η τιμή της μετρικής αυτής δίνεται από την παρακάτω σχέση [20]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

όπου n είναι ο αριθμός των παρατηρήσεων, y_i είναι η πραγματική τιμή και $\hat{y}_i$ είναι η προβλεπόμενη τιμή.

- **Mean Absolute Percentage Error (MAPE)**

Εκφράζει το σφάλμα ως ποσοστό επί της πραγματικής τιμής [21]:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

όπου n είναι ο αριθμός των παρατηρήσεων, y_i είναι η πραγματική τιμή και $\hat{y}_i$ είναι η προβλεπόμενη τιμή. Αν και το MAPE είναι επίσης μια δημοφιλής μετρική, μπορεί να παρουσιάσει προβλήματα όταν οι πραγματικές τιμές y_i είναι κοντά στο μηδέν, προκαλώντας απότομη αύξηση του ποσοστιαίου σφάλματος. Για τον λόγο αυτό, δεν συνιστάται η χρήση του όταν οι τιμές εξόδου είναι μικρές. Όσο μικρότερη είναι η τιμή του MAPE, τόσο πιο ακριβές θεωρείται το μοντέλο.

- **Root Mean Squared Error (RMSE)**

Η μετρική αυτή μετρά τη μέση απόκλιση μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Υπολογίζεται ως η τετραγωνική ρίζα της μέσης τιμής του τετραγώνου των διαφορών ανάμεσα στα αντίστοιχα ζεύγη τιμών. Λόγω της ύψωσης στο τετράγωνο, το RMSE δίνει μεγαλύτερη βαρύτητα στα μεγάλα σφάλματα. Όσο μικρότερη είναι η τιμή του, τόσο καλύτερα το μοντέλο προσεγγίζει τα πραγματικά σημεία εξόδου. Η τιμή του δίνεται από την παρακάτω μαθηματική διατύπωση [22].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

όπου n είναι ο αριθμός των παρατηρήσεων, y_i είναι η πραγματική τιμή και $\hat{y}_i$ είναι η προβλεπόμενη τιμή.

- **Συντελεστής Προσδιορισμού (R^2)**

Ο συντελεστής προσδιορισμού R^2 μετρά το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο. Αν η τιμή του R^2 είναι κοντά στο 1,

σημαίνει ότι το μοντέλο περιγράφει ικανοποιητικά τη διακύμανση της εξόδου. Αντιθέτως, τιμές κοντά στο 0 υποδηλώνουν ότι το μοντέλο αδυνατεί να αποδώσει τη μεταβλητότητα των δεδομένων. Αυτός ο δείκτης υπολογίζεται από τον εξής τύπο [23]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

όπου n είναι ο αριθμός των παρατηρήσεων, y_i είναι η πραγματική τιμή, $\hat{y}_i$ είναι η προβλεπόμενη τιμή και $\bar{y}$ είναι ο μέσος όρος των πραγματικών τιμών.

3.3 Αλγόριθμοι προσδιορισμού της αδράνειας

Όπως αναφέραμε και προηγουμένως χρησιμοποιήθηκαν συνολικά πέντε αλγόριθμοι για τον προσδιορισμό των σταθερών αδράνειας κάθε διασυνδεδεμένης περιοχής. Ο τρόπος λειτουργίας κάθε αλγορίθμου αναλύεται διεξοδικά στις παρακάτω υποτεινόμενες.

3.3.1 Πολλαπλή γραμμική παλινδρόμηση

Η πολλαπλή γραμμική παλινδρόμηση, πιο γνωστή με την αγγλική της ορολογία ως Multi-Linear Regression (MLR), είναι ένας supervised αλγόριθμος μηχανικής μάθησης που προσπαθεί να προσδιορίσει την έξοδο του συστήματος ως έναν γραμμικό συνδυασμό των εισόδων του [16]. Συνεπώς, η εξάρτηση μεταξύ της εξαρτημένης μεταβλητής και των διάφορων ανεξάρτητων μεταβλητών της θεωρείται γραμμική. Η μαθηματική περιγραφή που συνδέει την έξοδο του συστήματος, η οποία στη δική μας περίπτωση αντιστοιχεί στις σταθερές αδράνειας, με τις διάφορες τιμές των χαρακτηριστικών εισόδου (*features*) δίνεται από τη σχέση:

$$y = w_0 + w_1x_1 + w_2x_2 + \dots + w_bx_b \quad (3.1)$$

Όπου οι όροι $x_1, x_2, \dots, x_b$ είναι τα αριθμητικά περιεχόμενα των στηλών των χαρακτηριστικών εισόδου με διαστάσεις $n \times 1$, ενώ οι μεταβλητές $w_1, w_2, \dots, w_b$ αποτελούν τους συντελεστές ή αλλιώς τα βάρη (*weights*) αυτών των εισόδων και στην πράξη δηλώνουν τον βαθμό επιρροής κάθε εισόδου στην τελική έξοδο. Τέλος, ο όρος w_0 αποτελεί μια σταθερά (*intercept*) που στοχεύει στη βελτίωση της τελικής προσέγγισης.

Το ζητούμενο είναι ο ακριβής προσδιορισμός των βαρών ($w_0, w_1, w_2, \dots, w_b$) με σκοπό την ελαχιστοποίηση του αθροίσματος τετραγώνων των σφαλμάτων μεταξύ της πραγματικής

και της προβλεπόμενης από το μοντέλο τιμής Sum of Squared Errors [5]. Το SSE δίνεται από την ακόλουθη σχέση:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2} + \dots + w_b x_{ib})]^2 \quad (3.2)$$

Ο συγκεκριμένος αλγόριθμος μηχανικής μάθησης ακολουθεί την εξής λογική:

Ξεκινάμε με την τυχαία αρχικοποίηση των τιμών των βαρών $w_0, w_1, w_2, \dots, w_b$ και υπολογίζουμε το σύνολο των εξόδων y με διαστάσεις $n \times 1$. Είναι σχεδόν βέβαιο, ωστόσο, ότι οι τιμές του πίνακα y που προέκυψαν θα διαφέρουν σημαντικά από τις πραγματικές τιμές της εξόδου (y_{train}), λόγω της αυθαίρετης επιλογής των αρχικών βαρών. Συνεπώς, μία μόνο επανάληψη της διαδικασίας δεν αρκεί αλλά αντίθετα, απαιτείται ένας μεγάλος αριθμός επαναλαμβανόμενων βημάτων μέχρι να προσδιοριστούν με ικανοποιητική ακρίβεια οι τιμές των εξόδων.

Το ερώτημα, λοιπόν, είναι πώς μπορεί το μοντέλο μας να ξέρει αν οι τιμές που προσδιόρισε είναι έγκυρες ή αποκλίνουν σημαντικά από την πραγματικότητα. Η απάντηση βρίσκεται στη σχέση (3.2), όπου σε κάθε επανάληψη υπολογίζεται μία νέα τιμή για τη μεταβλητή SSE , η οποία λειτουργεί ως δείκτης της απόκλισης μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Όσο μικρότερη είναι η τιμή του SSE , τόσο καλύτερη είναι η προσέγγιση που επιτυγχάνει το μοντέλο.

Υπάρχει ένας πολύ μεγάλος αριθμός μεθόδων με στόχο τον προσδιορισμό των παραμέτρων $w_0, w_1, w_2, \dots, w_b$ έτσι ώστε να ελαχιστοποιηθεί η τιμή της SSE , δηλαδή να επιτευχθεί το μικρότερο δυνατό σφάλμα. Η πιο δημοφιλής από αυτές δεν είναι άλλη από τον Gradient Descent [24].

Σύμφωνα με τον συγκεκριμένο αλγόριθμο βελτιστοποίησης, σε κάθε βήμα οι τιμές των βαρών ενημερώνονται από την ακόλουθη γενική σχέση:

$$w_j \leftarrow w_j - \gamma \frac{\partial SSE(w_0, w_1, \dots, w_b)}{\partial w_j} \quad (3.3)$$

όπου ο όρος $\gamma > 0$ είναι γνωστός ως *learning rate* και δηλώνει τον ρυθμό με τον οποίο ενημερώνονται οι τιμές των βαρών σε κάθε βήμα. Οι τιμές των βαρών $w_0, w_1, \dots, w_b$ ενημερώνονται σε κάθε επανάληψη και χρησιμοποιούνται εκ νέου στον προσδιορισμό των τελικών τιμών εξόδου. Εάν η προσέγγιση της τελικής τιμής είναι ικανοποιητική, δηλαδή μικρότερη από ένα μέγιστο επιτρεπόμενο σφάλμα, ή αν δεν παρατηρείται περαιτέρω βελτίωση έπειτα

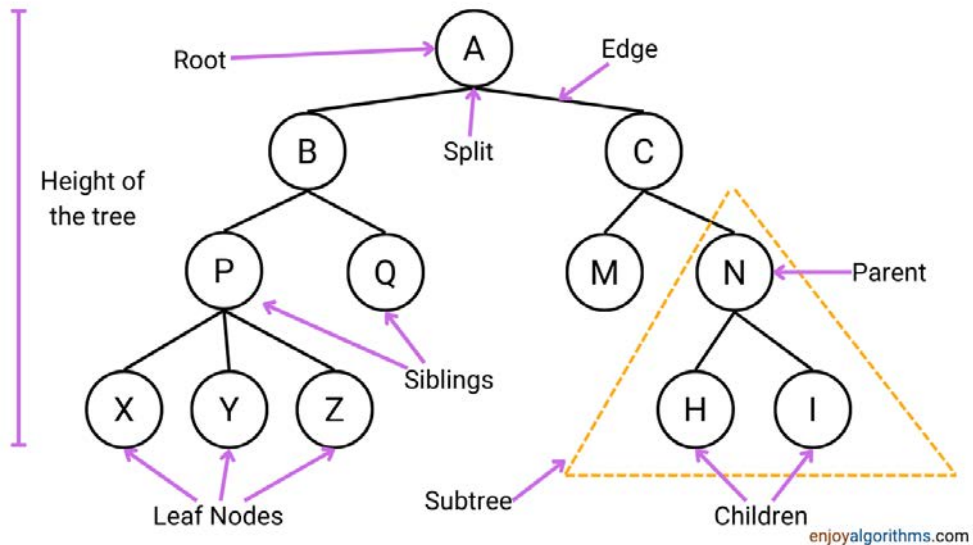
από έναν προκαθορισμένο αριθμό επαναλήψεων, η διαδικασία σταματά και η εξίσωση (3.1) λαμβάνει την τελική της μορφή.

3.3.2 Random forest

Ένα σημαντικό μειονέκτημα του προηγούμενου αλγορίθμου είναι η αδυναμία του να προσεγγίζει με ακρίβεια την έξοδο σε προβλήματα όπου η εξάρτηση μεταξύ των μεταβλητών εισόδου και εξόδου δεν είναι γραμμική. Για τον λόγο αυτό, με την πάροδο του χρόνου αναπτύχθηκε μια σειρά άλλων αλγορίθμων, με στόχο την επίλυση προβλημάτων αυτής της φύσης. Ένας πανίσχυρος και αρκετά ιδιαίτερος αλγόριθμος που χρησιμοποιείται ευρέως σε επίλυση προβλημάτων παλινδρόμησης είναι ο Random Forest, όπου η τελική τιμή εξόδου προκύπτει από τον μέσο όρο όλων των αποτελεσμάτων που λαμβάνουμε από ένα σύνολο δέντρων απόφασης [19].

Για την ερμηνεία του αλγορίθμου *Random Forest* κρίνεται αναγκαίο να αναλυθούν αρχικά οι λειτουργικές αρχές των δομικών λίθων που τον αποτελούν, δηλαδή των δέντρων απόφασης (Decision Trees). Σύμφωνα λοιπόν με την εικόνα 3.3, ένα δέντρο απόφασης αποτελείται από ένα σύνολο κόμβων και ακμών. Ο πρώτος κόμβος, που βρίσκεται στο ανώτερο σημείο του δέντρου, ονομάζεται *ρίζα*, ενώ οι κόμβοι που βρίσκονται στο κατώτερο σημείο ονομάζονται *φύλλα*. Το μήκος του μεγαλύτερου μονοπατιού, δηλαδή ο αριθμός των ακμών που μεσολαβούν από τη ρίζα έως το πιο απομακρυσμένο φύλλο, ορίζεται ως το *ύψος* του δέντρου. Το *βάθος* ενός κόμβου, από την άλλη, ορίζεται ως ο αριθμός των ακμών που μεσολαβούν από αυτόν μέχρι τη ρίζα του δέντρου. Συνεπώς, όταν αναφερόμαστε στο βάθος ολόκληρου του δέντρου, υπολογίζουμε το βάθος του πιο απομακρυσμένου φύλλου του από τη ρίζα. Με αυτή τη λογική, το βάθος και το ύψος ενός δέντρου ταυτίζονται ως έννοιες. Μια ακόμη σημαντική πληροφορία είναι ότι δύο κόμβοι που βρίσκονται σε διαδοχικά επίπεδα του δέντρου και συνδέονται μεταξύ τους έχουν σχέση *γονέα–παιδιού*. Σε προβλήματα παλινδρόμησης και ταξινόμησης, ωστόσο, σε αντίθεση με την εικόνα 3.3, γίνεται αποκλειστική χρήση *πλήρως δυαδικών δέντρων*, δηλαδή δέντρων στα οποία τόσο η ρίζα όσο και οι εσωτερικοί κόμβοι έχουν ακριβώς δύο παιδιά. Ο προσδιορισμός της τελικής τιμής y ακολουθεί την παρακάτω λογική:

Ένα αρχικό σύνολο δεδομένων (*training dataset*) αναμένει προσωρινά στη ρίζα του δέντρου. Κάθε κόμβος του δέντρου απόφασης, εκτός από τα φύλλα, περιέχει μια συνθήκη που διαχωρίζει τα δεδομένα που φτάνουν σε αυτόν, ανάλογα με το αν τα χαρακτηριστικά εισόδου



Σχήμα 3.3: Αναπαράσταση ενός δέντρου στην κόσμο των υπολογιστών και της πληροφορικής. [Decision Tree: A Tree-based Algorithm in Machine Learning]

(ανεξάρτητες μεταβλητές $x_1, x_2, \dots, x_b$) πληρούν την εκάστοτε συνθήκη. Συνήθως η συνθήκη αυτή είναι μια ανισωτική σχέση εάν μιλάμε για την δίκη μας περίπτωση που έχουμε συνεχείς μεταβλητές να επεξεργαστούμε και να προβλέψουμε, όπως για παράδειγμα $x_1 \geq 3$. Έτσι, κάθε κόμβος λειτουργεί σαν ένα σημείο διέλευσης ή, ακόμη καλύτερα, σαν μια διασταύρωση δεδομένων, όπου η πορεία που θα ακολουθήσει κάθε παρατήρηση καθορίζεται από τις τιμές των χαρακτηριστικών της [19]. Αφού το αρχικό *training dataset* διαχωριστεί στη ρίζα του δέντρου, δύο νέα υποσύνολα δεδομένων θα βρεθούν στα παιδιά της, έτοιμα να συναντήσουν νέες ανισωτικές συνθήκες και να διαιρεθούν περαιτέρω. Ο γενικός κανόνας διαχωρισμού είναι ότι, εάν η συνθήκη ικανοποιείται, η εκάστοτε πλειάδα (εγγραφή του *training dataset*) θα οδηγηθεί στο αριστερό παιδί του κόμβου που φέρει τη συνθήκη· διαφορετικά, θα μεταβεί στο δεξί παιδί. Η διαδικασία αυτή συνεχίζεται για όλες τις πλειάδες του συνόλου εκπαίδευσης, έως ότου φτάσουν στα φύλλα, τα οποία αποτελούν και τον τελικό προορισμό τους. Το βάθος κάθε δέντρου απόφασης καθορίζεται από τον εκάστοτε αναλυτή δεδομένων ανάλογα με τις ανάγκες του προβλήματος.

Πώς όμως διαμορφώνονται οι συνθήκες που συναντώνται σε κάθε κόμβο, έτσι ώστε να επιτυγχάνεται ο βέλτιστος διαμερισμός των δεδομένων;

Ένας από τους πολλούς τρόπους με τους οποίους μπορεί να περιγραφεί ένα σύνολο δεδομένων, ανεξαρτήτως μεγέθους, είναι μέσω της *διασποράς* του, η οποία αποτελεί μέτρο αταξίας και τυχαιότητας των δεδομένων που το απαρτίζουν. Κατάλληλη θεωρείται η συν-

θήκη εκείνη η οποία, σε κάθε βήμα διαχωρισμού, επιφέρει τη μεγαλύτερη δυνατή μείωση της διασποράς μεταξύ γονέα και παιδιών, καθώς στόχος είναι η ομαδοποίηση πλειάδων με κοινά χαρακτηριστικά. Η μαθηματική σχέση που μας επιτρέπει να υπολογίσουμε τη διασπορά ενός συνόλου δεδομένων δίνεται από την εξίσωση (3.4):

$$Var = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3.4)$$

Σε κάθε στάδιο διάσπασης, κατά τη διάρκεια της εκπαίδευσης, υπολογίζεται η διασπορά όλων των πιθανών συνόλων δεδομένων που μπορούν να προκύψουν από κάθε πιθανή διάσπαση. Για παράδειγμα, μπορεί να εξετάζονται τα δύο υποσύνολα που θα προκύψουν από τη συνθήκη $x_1 \geq 21$ ή τα δύο υποσύνολα που θα προκύψουν από τη συνθήκη $x_4 < -3$.

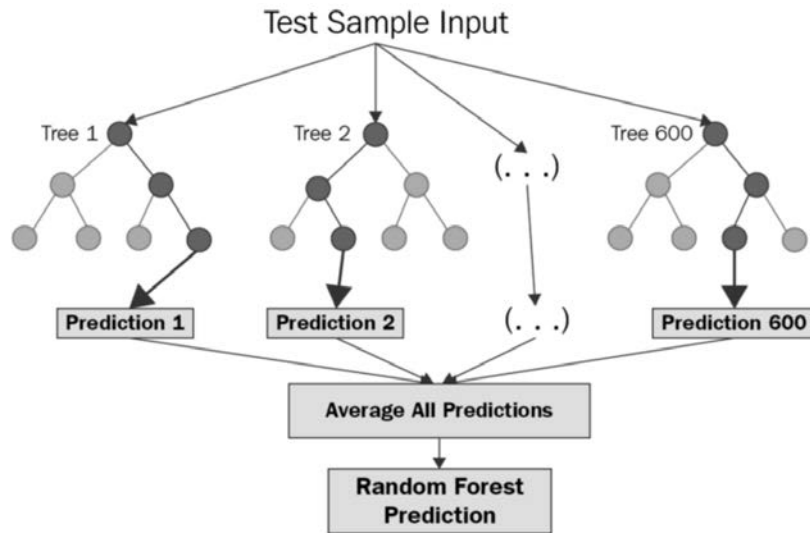
Σε κάθε σενάριο διαχωρισμού λοιπόν υπολογίζεται η διαφορά διασποράς μεταξύ του γονέα και των παιδιών. Η διαφορά αυτή δίνεται από την ακόλουθη εξίσωση:

$$Var_{\text{reduction}} = Var(\text{parent}) - \sum_i w_i Var(\text{child}_i) \quad (3.5)$$

όπου ο όρος w_i δηλώνει το ποσοστό του αρχικού συνόλου δεδομένων που κληρονομήθηκε από κάθε παιδί.

Η διαδικασία αυτή επαναλαμβάνεται για όλα τα πιθανά σενάρια διαχωρισμού σε όλους τους κόμβους του δέντρου, έως ότου φτάσουμε στα φύλλα. Σε αυτό το σημείο, το δέντρο είναι πλέον έτοιμο να υποδεχθεί τα δεδομένα επικύρωσης (*testing dataset*). Τα στοιχεία του *testing dataset*, για τα οποία το μοντέλο δεν γνωρίζει την έξοδο, εισέρχονται διαδοχικά στο δέντρο και ακολουθούν την πορεία που υπαγορεύουν οι συνθήκες κάθε κόμβου, έως ότου καταλήξουν σε κάποιο φύλλο. Εκεί, η τελική τιμή εξόδου προσδιορίζεται ως ο μέσος όρος των εξόδων των πλειάδων εκπαίδευσης που ανήκουν στο συγκεκριμένο φύλλο.

Στην περίπτωση του *Random Forest*, η παραπάνω διαδικασία εκπαίδευσης και πρόβλεψης εφαρμόζεται σε ένα σύνολο πολλών δέντρων απόφασης, τα οποία λειτουργούν παράλληλα. Κάθε δέντρο εκπαιδεύεται σε ένα διαφορετικό δείγμα του ίδιου αρχικού συνόλου εκπαίδευσης, το οποίο προκύπτει μέσω δειγματοληψίας με επανατοποθέτηση (*bootstrapping*). Έτσι, αν και όλα τα δέντρα «βλέπουν» δεδομένα από την ίδια αρχική πηγή, η σύνθεση του συνόλου εκπαίδευσης κάθε δέντρου διαφέρει, γεγονός που αυξάνει τη διαφοροποίηση μεταξύ τους. Μετά την εκπαίδευση, κάθε δέντρο του *Random Forest* δέχεται το *testing dataset* και παράγει τη δική του πρόβλεψη. Στο πρόβλημα παλινδρόμησης, η τελική έξοδος του μοντέλου προκύπτει ως ο μέσος όρος των προβλέψεων όλων των δέντρων όπως φαίνεται και



Σχήμα 3.4: Αναπαράσταση λειτουργίας του αλγορίθμου Random Forest. [Random Forest]

στο σχήμα 3.4

Αν και ο συγκεκριμένος αλγόριθμος πετυχαίνει εντυπωσιακή ακρίβεια στις προβλέψεις του, χαρακτηρίζεται από υψηλό υπολογιστικό κόστος λόγω του μεγάλου όγκου πράξεων και συγκρίσεων που απαιτούνται για την εκπαίδευση και λειτουργία πολυάριθμων δέντρων. Επιπρόσθετα, για να εξασφαλιστούν τα καλύτερα δυνατά αποτελέσματα, είναι απαραίτητος ο κατάλληλος προσδιορισμός παραμέτρων εισόδου, όπως το μέγιστο βάθος των δέντρων και ο συνολικός αριθμός τους. Ωστόσο παρά το υπολογιστικό κόστος, η παρούσα τεχνική παραμένει μία από τις δημοφιλέστερες μεθόδους σε προβλήματα παλινδρομησης. Από μαθηματική σκοπιά η προβλεπόμενη από το μοντέλο έξοδος θα προκύψει από την σχέση (3.6).

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x) \quad (3.6)$$

Όπου T ο αριθμός των δέντρων του δάσους και $f_t(x)$ η πρόβλεψη του κάθε δέντρου.

3.3.3 Support vector machines

Ο αλγόριθμος *Support Vector Regression* (SVR) αποτελεί ακόμη ένα ιδιαίτερα πολύτιμο εργαλείο στη συλλογή μας για τον προσδιορισμό της αδράνειας. Ξεχωρίζει για την ευελικτη και αποδοτική διαχείριση μη γραμμικών συνόλων δεδομένων, αξιοποιώντας τον πυρήνα (*kernel*) ώστε να προβάλει τα δεδομένα σε χώρους μεγαλύτερης διάστασης, όπου οι σχέσεις μεταξύ μεταβλητών γίνονται γραμμικά διαχωρίσιμες [25]. Αν και συχνά εξασφαλίζει εξαιρε-

τικές επιδόσεις τόσο σε προβλήματα παλινδρόμησης όσο και σε προβλήματα ταξινόμησης, καθώς και αποτελεσματική αντιμετώπιση των outliers, στις περισσότερες περιπτώσεις απαιτεί ιδιαίτερα προσεκτικό preprocessing και σωστή επιλογή υπερπαραμέτρων, ώστε να παραχθούν αξιόπιστα και σταθερά αποτελέσματα αυξάνοντας με αυτόν τον τρόπο το απαιτούμενο υπολογιστικό κόστος.

Η βασική ιδέα του SVR είναι η αναζήτηση μιας συνάρτησης $f(\mathbf{x})$ που να προσεγγίζει τις τιμές-στόχους y_i με σφάλμα μικρότερο από μια προκαθορισμένη ανοχή ϵ , ενώ ταυτόχρονα να εξασφαλίζεται η ελάχιστη δυνατή πολυπλοκότητα. Η γενική μορφή του μοντέλου είναι:

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (3.7)$$

όπου $\phi(\mathbf{x})$ είναι ένας μετασχηματισμός των εισόδων σε χώρο μεγαλύτερης διάστασης, $\mathbf{w}$ το διάνυσμα βαρών και b η σταθερά μετατόπισης.

Ο στόχος είναι η ελαχιστοποίηση της ακόλουθης συνάρτησης κόστους [25]:

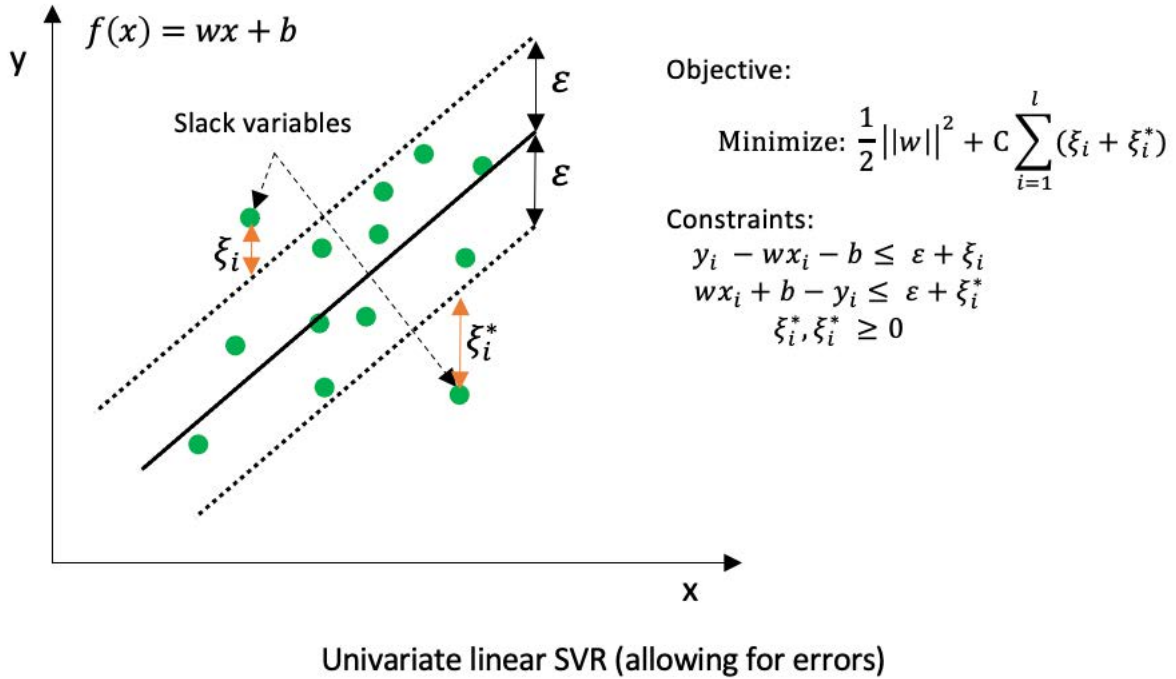
$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3.8)$$

υπό τους περιορισμούς:

$$\begin{cases} y_i - f(\mathbf{x}_i) \leq \epsilon + \xi_i \\ f(\mathbf{x}_i) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3.9)$$

όπου C είναι παράμετρος που ελέγχει την ισορροπία μεταξύ πολυπλοκότητας του μοντέλου και ανεκτών σφαλμάτων, και οι ξ_i, ξ_i^* είναι μεταβλητές χαλάρωσης που επιτρέπουν αποκλίσεις μεγαλύτερες από ϵ . Εικονική αναπαράσταση αυτής της προσέγγισης στο δισδιάστατο χώρο φαίνεται στο σχήμα 3.5

Με απλά λόγια, το SVR προσπαθεί να βρει μια γραμμή ή υπερεπίπεδο στην δική μας περίπτωση λόγω του πλήθους των μεταβλητών εισόδου που περνά “κοντά” στα δεδομένα, αγνοώντας μικρές αποκλίσεις (ζώνη ανεκτικότητας ϵ) και τιμωρώντας τις μεγάλες αποκλίσεις, έτσι ώστε το μοντέλο να έχει καλή γενίκευση σε νέα δεδομένα. Όσο μεγαλώνει ο αριθμός των ανεξάρτητων μεταβλητών ωστόσο τόσο αυξάνεται και το πλήθος των διαστάσεων επομένως από τρεις μεταβλητές και έπειτα μας είναι ολοκληρωτικά αδύνατον να αναπαραστήσουμε σχηματικά τόσο το υπερεπίπεδο όσο και τις τιμές εξόδου των δεδομένων που θα προσεγγίζει.



Σχήμα 3.5: Αναπαράσταση λειτουργίας του αλγορίθμου SVR στον διδιάστατο χώρο. [From Theory to Practice: Implementing Support Vector Regression for Predictions in Python]

3.3.4 EXtreme gradient boosting

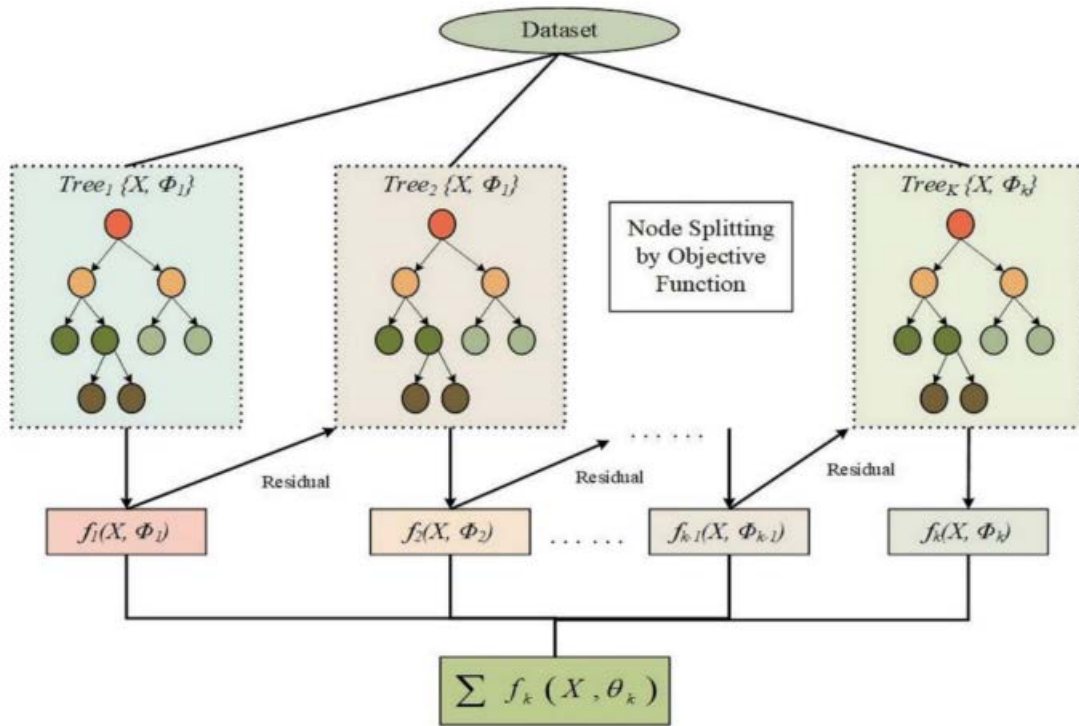
Ο XGBoost (*Extreme Gradient Boosting*) είναι ένας ιδιαίτερα ισχυρός αλγόριθμος μηχανικής μάθησης βασισμένος σε δέντρα απόφασης, όπως ακριβώς και ο *Random Forest*. Η λειτουργία του στηρίζεται στην τεχνική του *gradient boosting* [26], εξασφαλίζοντας με αυτόν τον τρόπο εντυπωσιακή ακρίβεια στις προβλέψεις σε σύντομο χρονικό διάστημα. Η βασική ιδέα στηρίζεται στην κατασκευή διαδοχικών απλών μοντέλων (*weak learners*), συνήθως μικρών δέντρων, τα οποία βελτιώνουν σταδιακά τα λάθη των προηγούμενων. Αντί, λοιπόν, να προσπαθεί να μάθει τη συνάρτηση πρόβλεψης $f(x)$ κατευθείαν, ο XGBoost κατασκευάζει ένα άθροισμα από K απλών συναρτήσεων (δέντρα):

$$\hat{y}_i = \sum_{k=1}^K f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F}$$

όπου $\mathcal{F}$ είναι το σύνολο όλων των πιθανών δέντρων απόφασης.

Όπως προαναφέρθηκε, η εκπαίδευση γίνεται βήμα-βήμα. Όταν βρισκόμαστε στο t -οστό βήμα, έχοντας το μοντέλο $\hat{y}_i^{(t-1)}$, προσθέτουμε ένα νέο δέντρο f_t που μειώνει το σφάλμα:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)$$



Σχήμα 3.6: Αναπαράσταση λειτουργίας του αλγορίθμου XGBoost [3]

Στόχος, όπως και στους περισσότερους αλγορίθμους μηχανικής μάθησης, είναι η ελαχιστοποίηση μιας συνάρτησης κόστους. Στην προκειμένη περίπτωση, ο XGBoost προσπαθεί να ελαχιστοποιήσει την ακόλουθη σχέση [27]:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k)$$

όπου l είναι η συνάρτηση απώλειας (π.χ. τετραγωνικό σφάλμα) και $\Omega(f_k)$ είναι όρος κανονικοποίησης που περιορίζει την πολυπλοκότητα του δέντρου (π.χ. βάθος και αριθμός φύλλων). Για την εύρεση νέων δέντρων f_t , γίνεται ανάπτυξη Taylor 2ης τάξης γύρω από τις προβλέψεις του προηγούμενου μοντέλου:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t)$$

όπου:

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}$$

είναι οι πρώτες και δεύτερες παραγώγοι (gradient και hessian).

Κατά την κατασκευή του δέντρου, για κάθε υποψήφιο κόψιμο (*split*), ο XGBoost υπολο-

γίζει το *gain*:

$$\text{Gain} = \frac{1}{2} \left[\frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_i + \lambda} + \frac{(\sum_{i \in R} g_i)^2}{\sum_{i \in R} h_i + \lambda} - \frac{(\sum_{i \in L \cup R} g_i)^2}{\sum_{i \in L \cup R} h_i + \lambda} \right] - \gamma$$

όπου L και R είναι τα αριστερά και δεξιά παιδιά, η παράμετρος λ ελέγχτει την ομαλότητα των βαρών και το γ είναι το κόστος δημιουργίας νέου φύλλου.

Συνοπτικά, λοιπόν, οι τελικές τιμές εξόδου του πίνακα y προκύπτουν από το συνδυαστικό άθροισμα των προβλέψεων όλων των δέντρων απόφασης. Ξεκινώντας από το πρώτο δέντρο παράγεται μια αρχική πρόβλεψη των τιμών του πίνακα y , η οποία προκύπτει από τη μέση τιμή της στήλης των πραγματικών τιμών εξόδου και, όπως είναι αναμενόμενο, απέχει αρκετά από την πραγματικότητα. Στη συνέχεια, υπολογίζεται η διαφορά μεταξύ πραγματικής και προβλεπόμενης τιμής (*residue*) και ακολουθεί η διοχέτευση της ως είσοδο στο αμέσως επόμενο δέντρο. Έπειτα, το επόμενο δέντρο υπολογίζει τις τιμές ενός νέου πίνακα, ο οποίος θα προστεθεί στον αρχικό, με σκοπό να μειώσει το σφάλμα μεταξύ πραγματικών και προβλεπόμενων τιμών. Η διαδικασία αυτή επαναλαμβάνεται για συγκεκριμένο πλήθος δέντρων. Σε αντίθεση με το *Random Forest*, όπου τα δέντρα λειτουργούν ανεξάρτητα, στον XGBoost κάθε δέντρο κατασκευάζεται ώστε να διορθώνει τα σφάλματα του προηγούμενου, με αποτέλεσμα η τελική πρόβλεψη να αποτελεί το αθροιστικό αποτέλεσμα όλων των δέντρων. Η βασική αρχή λειτουργίας του παρόντος αλγορίθμου απεικονίζεται στην Εικόνα 3.6.

3.3.5 Multi-layer perceptron

Το τελευταίο μοντέλο που εξετάστηκε, και μάλιστα, όπως θα δούμε στη συνέχεια, παρήγαγε και τις ακριβέστερες προβλέψεις για τις τιμές της αδράνειας, είναι ο αλγόριθμος Multi-Layer Perceptron (MLP). Ο συγκεκριμένος αλγόριθμος ανήκει στην οικογένεια των νευρωνικών δικτύων τα οποία αποτελούν μια ιδιαίτερα δημοφιλή οικογένεια αλγορίθμων, γνωστή με τον όρο *deep learning models*, η οποία είναι υποκατηγορία των αλγορίθμων *Machine Learning*. Το φάσμα εφαρμογών όπου χρησιμοποιούνται τα νευρωνικά δίκτυα είναι αρκετά ευρύ και εκτείνεται από απλούς αλγορίθμους προσδιορισμού τιμών έως τη ρομποτική, την επεξεργασία εικόνας και τη φυσική επεξεργασία γλώσσας. Η τεράστια αυτή δημοτικότητα οφείλεται σε μεγάλο βαθμό στην ικανότητα αυτών των αλγορίθμων να μπορούν να διαχειρίζονται αποτελεσματικά μη γραμμικά δεδομένα. Δεδομένα, δηλαδή, των οποίων οι τιμές εξόδου δεν εξαρτώνται γραμμικά από τις τιμές εισόδου [5].

Κοινό χαρακτηριστικό όλων των μοντέλων που ανήκουν στην οικογένεια των νευρωνι-

κών δικτύων είναι οι νευρώνες. Οι τελευταίοι, οργανωμένοι σε επίπεδα (layers), λαμβάνουν ως είσοδο πληροφορία, εφαρμόζουν σε αυτήν έναν μαθηματικό μετασχηματισμό με τη χρήση συναρτήσεων ενεργοποίησης και μεταβιβάζουν το τελικό αποτέλεσμα στους νευρώνες των επόμενων επιπέδων.

Το MLP μοντέλο αποτελεί την απλούστερη μορφή ενός νευρωνικού δικτύου και δομείται από έναν συγκεκριμένο αριθμό επιπέδων (layers), τα οποία συνδέονται μεταξύ τους με ακμές. Πιο συγκεκριμένα, διακρίνουμε το επίπεδο εισόδου (input layer), ένα ή περισσότερα κρυφά επίπεδα (hidden layers) και, τέλος, το επίπεδο εξόδου (output layer), όπου προκύπτει η τελική τιμή της πρόβλεψης.

Το συγκεκριμένο νευρωνικό δίκτυο λαμβάνει ως είσοδο ένα διάνυσμα που αποτελείται από αριθμητικές τιμές, καθεμία από τις οποίες αντιστοιχεί στην τιμή ενός χαρακτηριστικού (feature) μίας πλειάδας του συνόλου δεδομένων. Κάθε επίπεδο αποτελείται από έναν συγκεκριμένο αριθμό νευρώνων, με το input layer να διαθέτει ακριβώς τόσους νευρώνες όσα είναι και τα χαρακτηριστικά εισόδου του δικτύου (δηλαδή οι ανεξάρτητες μεταβλητές), ενώ το output layer να διαθέτει τόσους νευρώνες όσες είναι και οι μεταβλητές εξόδου.

Οι νευρώνες, ή αλλιώς κόμβοι όπως συνηθίζεται να λέγονται, στο συγκεκριμένο μοντέλο είναι πλήρως συνδεδεμένοι μεταξύ τους (fully connected layers) πράγμα που σημαίνει ότι κάθε κόμβος των κρυφών επιπέδων δέχεται ως είσοδο τις εξόδους όλων των νευρώνων του προηγούμενου layer και, στη συνέχεια, μεταβιβάζει το αποτέλεσμα της εξόδου που παράγει σε όλους τους κόμβους του επόμενου layer. Οι νευρώνες δύο γειτονικών επιπέδων, όπως αναφέρθηκε προηγουμένως, συνδέονται μεταξύ τους μέσω ακμών, καθεμία από τις οποίες διαθέτει το δικό της βάρος (weight). Το βάρος αυτό στην πραγματικότητα αποτελεί έναν αριθμητικό συντελεστή που πολλαπλασιάζει την έξοδο ενός νευρώνα πριν αυτή μεταβιβαστεί στον επόμενο. Η τιμή κάθε βάρους ($w_0, w_1, w_2, \dots, w_b$) εκφράζει τον βαθμό επιρροής της εξόδου ενός νευρώνα στον αμέσως επόμενο, ενώ ο τελικός προσδιορισμός των τιμών τους αποτελεί το κλειδί για την κατασκευή ενός λειτουργικού και αποτελεσματικού μοντέλου πρόβλεψης [5].

Η διαδικασία της εκπαίδευσης ξεκινά με την είσοδο στο input layer της πρώτης πλειάδας του training dataset, η οποία έχει τη μορφή πίνακα-στήλης ($b \times 1$) και περιέχει τις αριθμητικές τιμές των χαρακτηριστικών εισόδου για μία εγγραφή. Σε αυτό το επίπεδο, κάθε νευρώνας «συγκρατεί» μία και μόνο τιμή, η οποία αντιστοιχεί σε κάθε χαρακτηριστικό εισόδου. Οι νευρώνες του input layer δεν επιβάλλουν κάποιον μαθηματικό μετασχηματισμό στα δεδομένα,

απλώς τα μεταβιβάζουν στο επόμενο επίπεδο, δηλαδή στα *hidden layers*, όπου εφαρμόζονται πιο πολύπλοκοι μαθηματικοί υπολογισμοί στην εισερχόμενη πληροφορία, κάνοντας χρήση μη γραμμικών συναρτήσεων ενεργοποίησης.

Αφού, λοιπόν, τα δεδομένα μεταβιβαστούν στο πρώτο *hidden layer*, αρχικά κάθε κόμβος αθροίζει όλα τα δεδομένα εισόδου τα οποία προήλθαν από το *input layer* και έχουν πολλαπλασιαστεί με τον αντίστοιχο συντελεστή βάρους της κάθε ακμής. Έπειτα μόλις ολοκληρωθεί αυτή η αθροίση, το αποτέλεσμα διοχετεύεται ως όρισμα σε μία συνάρτηση ενεργοποίησης, η οποία τείνει να διαφέρει ανάλογα με τη φύση του προβλήματος που αντιμετωπίζουμε. Οι πιο γνωστές και ευρέως χρησιμοποιούμενες συναρτήσεις ενεργοποίησης καθώς και η αντίστοιχη μαθηματική περιγραφή τους φαίνεται στον 3.1.

Πίνακας 3.1: Δημοφιλέστερες συναρτήσεις ενεργοποίησης [5]

Συνάρτηση Ενεργοποίησης	Μαθηματική Εξίσωση f
<i>ReLU</i>	$\text{ReLU}(x) = \max(0, x)$
<i>Sigmoid</i>	$\sigma(x) = \frac{1}{1 + e^{-x}}$
<i>tanh</i>	$\tanh(x)$

Αν και είναι πιθανό να χρησιμοποιούνται διαφορετικές συναρτήσεις ενεργοποίησης σε ολόκληρη την έκταση ενός νευρωνικού δικτύου, ο διαχωρισμός του σε επίπεδα επιβάλλει την αποκλειστική χρήση ενός μόνο είδους τέτοιας συνάρτησης σε καθένα από αυτά.

Αφού η συνάρτηση ενεργοποίησης του κάθε νευρώνα παράγει την τιμή εξόδου, αυτή προωθείται στο επόμενο επίπεδο, το οποίο μπορεί να είναι είτε ακόμη ένα κρυφό επίπεδο είτε το επίπεδο εξόδου. Ο αριθμός των κρυφών επιπέδων, σε συνδυασμό με τον αριθμό των νευρώνων που ανήκουν σε αυτά και τη συνάρτηση ενεργοποίησης που χρησιμοποιείται, αποτελούν υπερπαραμέτρους (*hyperparameters*) οι οποίες πρέπει να ληφθούν υπόψη κατά τον σχεδιασμό του μοντέλου.

Στη γλώσσα των μαθηματικών, η παραπάνω διαδικασία εκφράζεται με την εξής πρόταση: Έστω ότι διαθέτουμε τις τιμές εισόδου $x_1, x_2, \dots, x_n$, οι οποίες τροφοδοτούνται σε έναν νευρώνα j που ανήκει στο επίπεδο h . Η έξοδος του νευρώνα αυτού υπολογίζεται ως:

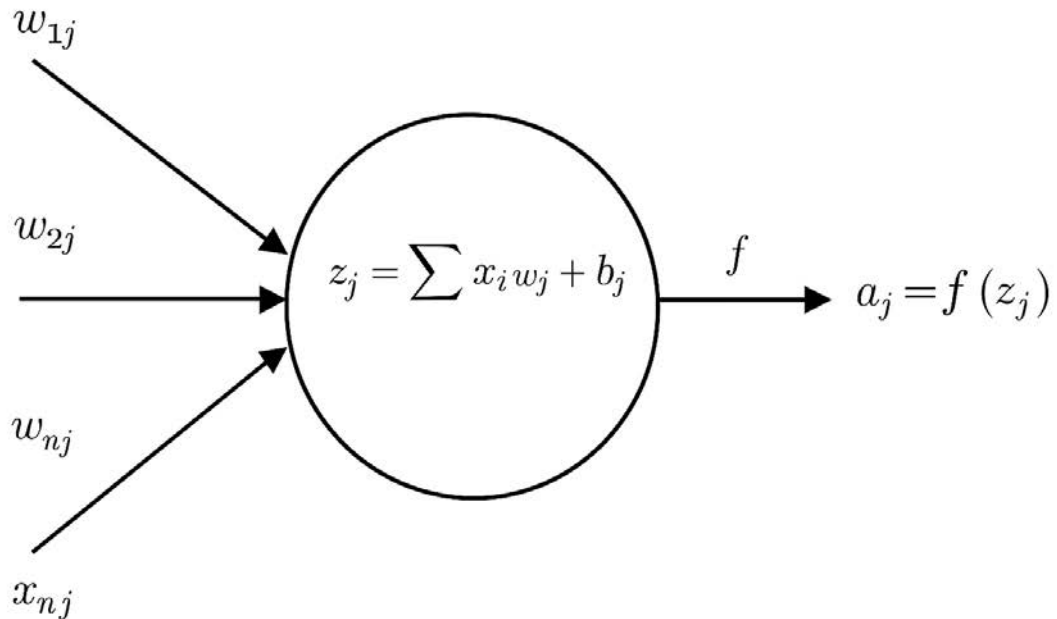
$$z_j^{(h)} = \sum_{i=1}^n w_{ij}^{(h)} x_i + b_j^{(h)} \quad (3.10)$$

όπου $w_{ij}^{(h)}$ είναι το βάρος της σύνδεσης μεταξύ του i -οστού νευρώνα του προηγούμενου επιπέδου και του j -οστού νευρώνα του επιπέδου h , ενώ $b_j^{(h)}$ είναι η προκατάληψη (*bias*) του

νευρώνα. Στη συνέχεια, εφαρμόζεται η συνάρτηση ενεργοποίησης $f(\cdot)$ ώστε να προκύψει η τελική έξοδος:

$$a_j^{(h)} = f\left(z_j^{(h)}\right). \quad (3.11)$$

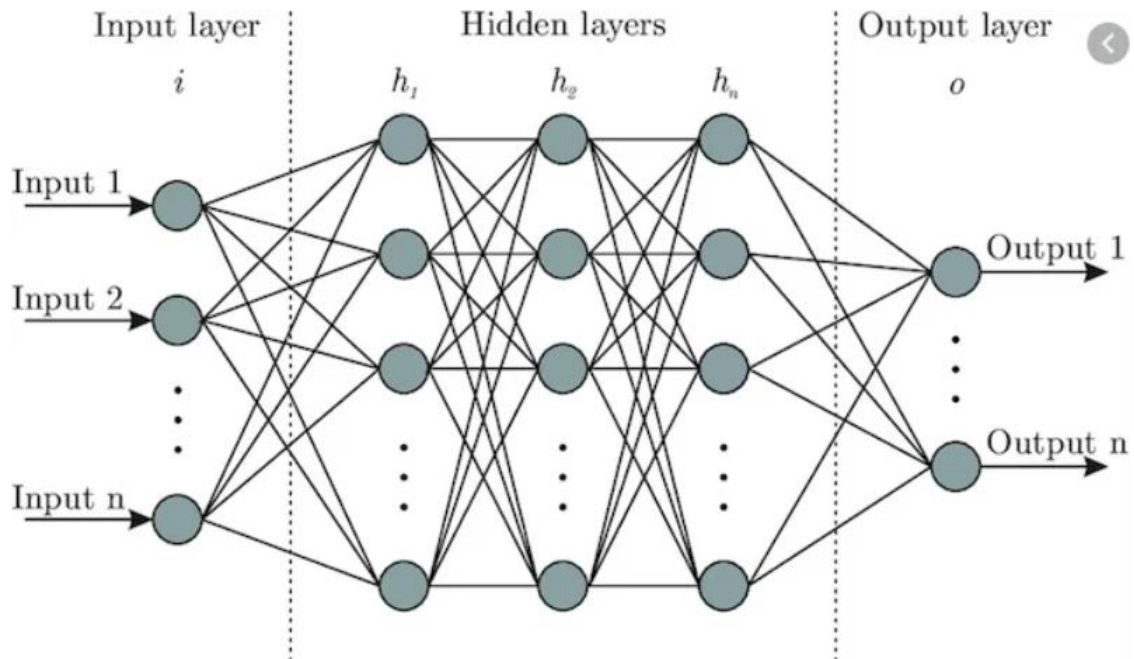
Το σχήμα 3.7 απεικονίζει με ακρίβεια την πρακτική εφαρμογή των εξισώσεων 3.10,3.11



Σχήμα 3.7: Αρχή λειτουργίας ενός νευρώνα του hidden layer

Η ίδια λογική ισχύει για όλους τους κόμβους που ανήκουν στα κρυφά επίπεδα του νευρωνικού δικτύου. Όσον αφορά το επίπεδο εξόδου, τα δεδομένα, αφού έχουν ήδη υποστεί ένα μεγάλο πλήθος μαθηματικών μετασχηματισμών στα *hidden layers*, εισέρχονται για μία τελευταία φορά σε ένα καινούργιο επίπεδο νευρώνων. Εάν πρόκειται για πρόβλημα ταξινόμησης οι κόμβοι θα αθροίσουν τις τιμές των διαφόρων ακμών, θα τις διοχετεύσουν σε μια συνάρτηση ενεργοποίησης και θα προσδιορίσουν τις τελικές τιμές. Σε προβλήματα παλινδρόμησης ωστόσο μιας και στόχος είναι ο προσδιορισμός συνεχών τιμών η χρήση συναρτήσεων ενεργοποίησης στο output layer αποφεύγεται. Μια αναπαράσταση ενός πλήρως διασυνδεδεμένου νευρωνικού δικτύου, παρόμοιου με αυτό που χρησιμοποιήθηκε στο πλαίσιο της παρούσας διπλωματικής, φαίνεται στο Σχήμα 3.8.

Οι νευρώνες στο *output layer* παράγουν τις τελικές τιμές εξόδου του συστήματος. Αφού



Σχήμα 3.8: Πλήρως διασυνδεδεμένο νευρωνικό δίκτυο.

ολοκληρωθεί η διέλευση της πρώτης πλειάδας του *training dataset*, ακολουθεί η δεύτερη πλειάδα και η διαδικασία επαναλαμβάνεται.

Το βασικό ερώτημα σε αυτό το σημείο είναι πώς η διάταξη του Σχήματος 3.8 μπορεί να προσδιορίσει με ακρίβεια τις τιμές της αδράνειας μιας περιοχής ενός ΣΗΕ.

Τα δεδομένα εισόδου, αφού προηγουμένως έχουν υποβληθεί σε *pre-processing* ώστε να είναι ευκολότερα διαχειρίσιμα και συμβατά με το νευρωνικό δίκτυο, τροφοδοτούνται σε αυτό οργανωμένα σε ομάδες, οι οποίες ονομάζονται *batches*. Οι πλειάδες κάθε batch εισέρχονται σειριακά στο νευρωνικό δίκτυο και υποβάλλονται στις διεργασίες που περιγράψαμε νωρίτερα. Αφού παραχθούν οι προβλέψεις του μοντέλου για όλες τις πλειάδες του πρώτου batch, ακολουθεί μια θεμελιώδης διαδικασία για τα νευρωνικά δίκτυα: το *backpropagation*.

Αρχικά, τα βάρη των ακμών αρχικοποιούνται με τυχαίο τρόπο και, συνεπώς, οι προβλεπόμενες τιμές που θα προκύψουν μετά τη διέλευση του πρώτου *batch* ενδέχεται να απέχουν αρκετά από την πραγματικότητα. Αφού παραχθεί το πρώτο σύνολο προβλέψεων, η εγκυρότητά τους αξιολογείται με τη χρήση διαφόρων μετρικών, οι οποίες ελέγχουν συνήθως τη διαφορά μεταξύ της πραγματικής και της προβλεπόμενης τιμής.

Στη συνέχεια, κατά τη διάρκεια του *backpropagation*, οι τιμές των βαρών ενημερώνονται σε όλη την έκταση του νευρωνικού δικτύου με στόχο την ελαχιστοποίηση του σφάλματος. Ο τρόπος με τον οποίο διαμορφώνονται οι νέες τιμές εξαρτάται από τον εκάστοτε *optimizer*

που έχει επιλεγεί (π.χ. Adam, SGD).

Μετά τη διέλευση του πρώτου batch και την τροποποίηση των βαρών, ακολουθεί το πέρας των πλειάδων του δεύτερου batch στο νευρωνικό δίκτυο. Ο αριθμός των εγγραφών κάθε batch είναι συνήθως δύναμη του 2 (π.χ. 32, 64, 128) και αποτελεί επίσης υπερπαράμετρο, η οποία πρέπει να επιλέγεται με προσοχή ώστε να εξασφαλίζεται υψηλή ακρίβεια στις προβλέψεις. Η διαδικασία αυτή επαναλαμβάνεται για όλα τα batches του *training dataset*, μέχρι όλα τα δεδομένα εισόδου να έχουν περάσει τουλάχιστον μία φορά από το νευρωνικό δίκτυο. Τότε λέμε ότι ολοκληρώθηκε μία επανάληψη (*epoch*).

Η επιλογή του αριθμού των epochs γίνεται με τέτοιο τρόπο ώστε να επιτυγχάνεται η μέγιστη δυνατή ακρίβεια. Πρακτικά, εξετάζεται η απόδοση του μοντέλου για διάφορα σενάρια με διαφορετικό αριθμό epochs και επιλέγεται το βέλτιστο. Έπειτα από την ολοκλήρωση του συνολικού αριθμού των επαναλήψεων, τα βάρη έχουν πλέον διαμορφωθεί και το μοντέλο είναι έτοιμο να δοκιμάσει τις δυνατότητές του σε άγνωστα για εκείνο δεδομένα (*testing dataset*). Αν το σφάλμα κυμαίνεται σε ικανοποιητικά πλαίσια, αυτό σημαίνει ότι το μοντέλο έχει καταφέρει να διαμορφώσει την κατάλληλη συνάρτηση που συνδέει τις εισόδους με τις εξόδους και έχει ισχυρή ικανότητα πρόβλεψης.

Κεφάλαιο 4

Μοντέλα απόκρισης συχνότητας διασυνδεδεμένων ΣΗΕ

4.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται αναλυτικά η προσομοίωση της δυναμικής συμπεριφοράς δύο διασυνδεδεμένων περιοχών ελέγχου ενός ΣΗΕ, με χρήση μοντέλων απόκρισης συχνότητας (system frequency response models). Στο πλαίσιο της παρούσας διπλωματικής, αναπτύχθηκαν σε περιβάλλον Matlab-Simulink πολλαπλά σενάρια με διαφορετικούς συνδυασμούς μονάδων παραγωγής, με στόχο τη διερεύνηση της δυναμικής απόκρισης του συστήματος υπό ποικίλες συνθήκες λειτουργίας. Τα σενάρια αυτά καλύπτουν ένα ευρύ φάσμα γραμμικών μοντέλων που προσεγγίζουν τη συμπεριφορά διαφορετικών τύπων μονάδων παραγωγής, όπως υδροηλεκτρικοί σταθμοί, ατμοστρόβιλοι με και χωρίς αναθέρμανση ενώ σε επόμενες ενότητες θα εξεταστούν και αεριοστροβιλικόι σταθμοί και μονάδες βιομάζας.

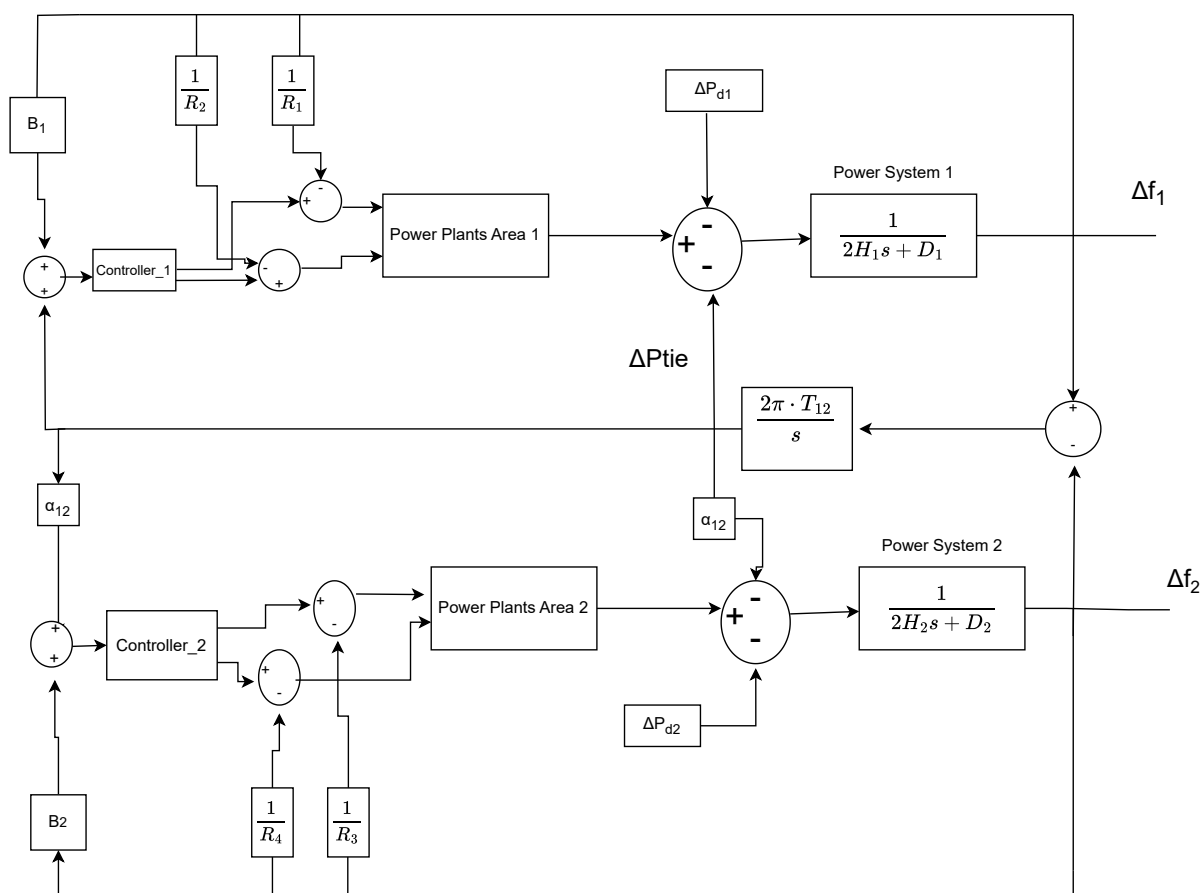
4.2 Βασικό δομικό διάγραμμα

Το διάγραμμα βαθμίδων που απεικονίζει τις δύο διασυνδεδεμένες περιοχές ελέγχου παρουσιάζεται στο Σχήμα 4.1. Παρατηρώντας το διάγραμμα, μπορούν να διακριθούν τέσσερις βασικές συνιστώσες για κάθε περιοχή:

- **Συνάρτηση μεταφοράς ΣΗΕ (Power System Component):** Πρόκειται για μια συνάρτηση μεταφοράς της οποίας η έξοδος αντιστοιχεί στη συνολική απόκριση συχνότητας της περιοχής (Δf), εκφρασμένη σε p.u.

- **Είσοδος Διαταραχής (ΔP_d)** : Εκφράζει τη μεταβολή της ζητούμενης ισχύος.
- **Controller**: Υποσύστημα που μέσω της χρήσης ολοκληρωτικών ελεγκτών, έχει ως στόχο να επαναφέρει τη συχνότητα στην ονομαστική της τιμή. Πρόκειται ουσιαστικά για την υλοποίηση της δευτερεύουσας ρύθμισης συχνότητας κάθε περιοχής.
- **Μονάδες Παραγωγής (Power Plants)**: Υποσύστημα που αποτελείται από ένα μείγμα διαφόρων μονάδων παραγωγής, του οποίου το περιεχόμενο διαφοροποιείται ανάλογα με το σενάριο που εξετάζεται.

Η λειτουργικότητα και η εσωτερική αρχιτεκτονική καθεμιάς από της προαναφερθείσες διατάξεις θα αναλυθεί εκτενώς στα υποκεφάλαια που ακολουθούν.

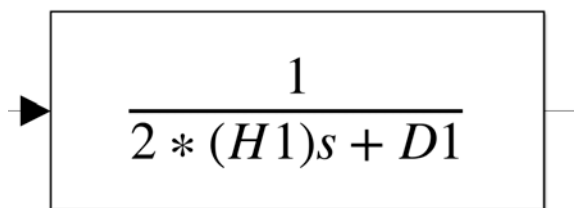


Σχήμα 4.1: Διάγραμμα δυο διασυνδεδεμένων περιοχών έλεγχου .

4.2.1 Συνάρτηση Μεταφοράς ΣΗΕ (Power System component)

Η συγκεκριμένη συνάρτηση μεταφοράς περιλαμβάνει τη συνολική σταθερά αδράνειας της εκάστοτε περιοχής, H (σε s), καθώς και τον συντελεστή αυτορύθμισης φορτίου D (damping factor). Ο αντίστοιχος βρόχος για την Περιοχή 1 παρουσιάζεται με μεγαλύτερη ευκρίνεια στο Σχήμα 4.2.

Γίνεται επομένως φανερό ότι τόσο η τιμή της σταθεράς αδράνειας H_{eq} κάθε περιοχής όσο και εκείνη της σταθεράς απόσβεσης D επηρεάζουν σημαντικά τη δυναμική απόκριση της περιοχής. Η σταθερά αδράνειας έχει μεγαλύτερη επίδραση στην απόκριση της συχνότητας του συστήματος και συνήθως μεταβάλλεται αρκετά κατά τη διάρκεια της ημέρας. Αντίθετα, οι συντελεστές απόσβεσης D μπορούν να θεωρηθούν περίπου σταθεροί και γνωστοί. Έτσι, θεωρώντας γνωστές τις τιμές των D_1 και D_2 , στόχος της παρούσας διπλωματικής εργασίας είναι ο ακριβής προσδιορισμός των τιμών των επιμέρους σταθερών αδράνειας H_1 και H_2 .

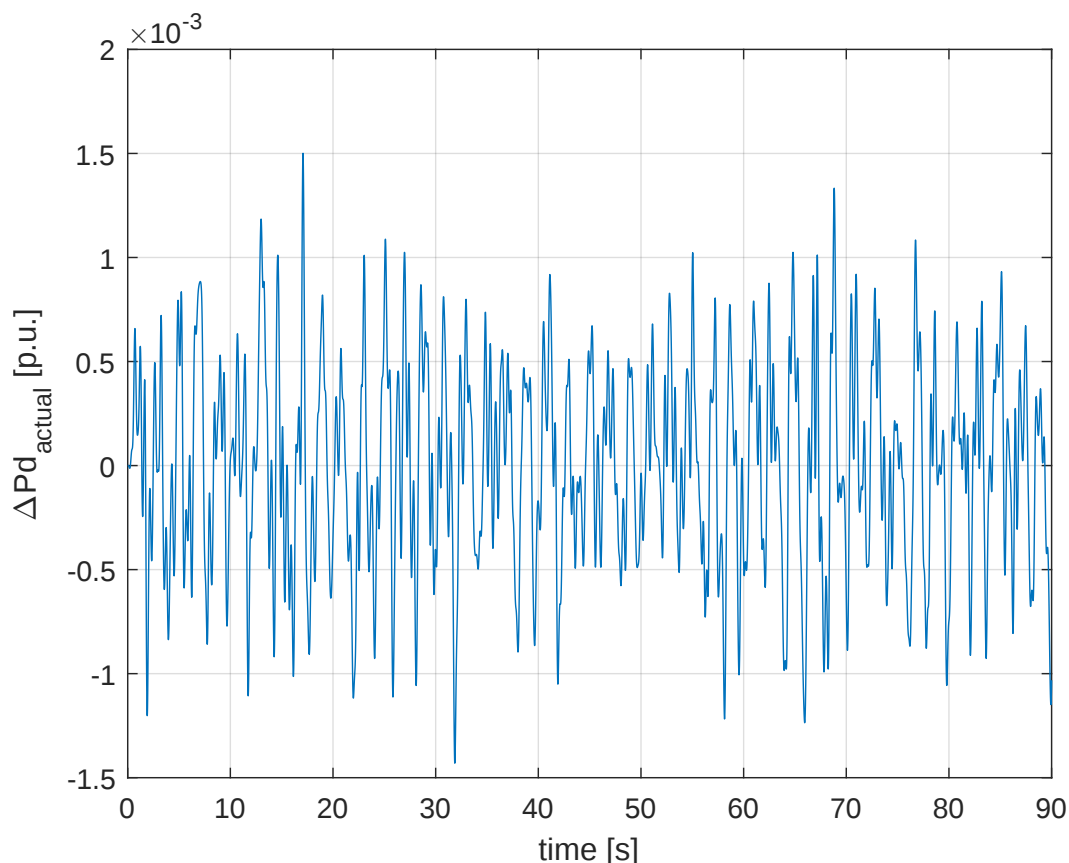


Σχήμα 4.2: Συνάρτηση μεταφοράς του Power System

4.2.2 Είσοδος διαταραχής ΔP_d

Όπως αναφέρθηκε και παραπάνω, το ΔP_d εκφράζει τη μεταβολή της ζητούμενης ισχύος σε p.u. Εάν η τιμή του είναι αρνητική, αυτό σημαίνει ότι υπάρχει περίσσεια ισχύος στο δίκτυο· συνεπώς, η συχνότητα τείνει να αυξηθεί και πρέπει να δοθεί κατάλληλο σήμα στον ρυθμιστή στροφών ώστε να μειώσει την ισχύ εξόδου των μονάδων παραγωγής. Αντιθέτως, όταν η τιμή του ΔP_d είναι θετική, δηλώνεται αύξηση της ζήτησης ηλεκτρικής ενέργειας, με αποτέλεσμα πτώση της συχνότητας, γεγονός που απαιτεί την αύξηση της παραγόμενης ισχύος. Στην πράξη, η ζητούμενη ισχύς δεν παραμένει ποτέ σταθερή, αλλά μεταβάλλεται συνεχώς, παρουσιάζοντας άλλοτε θετικές και άλλοτε αρνητικές αποκλίσεις. Οι διακυμάνσεις αυτές οφείλονται κυρίως στη στοχαστική συμπεριφορά των φορτίων, όπως για παράδειγμα στο άνοιγμα και κλείσιμο οικιακών συσκευών, στη μεταβολή της βιομηχανικής δραστηριότητας ή ακόμη και σε εξωγενείς παράγοντες όπως οι καιρικές συνθήκες.

Στο πλαίσιο της παρούσας διπλωματικής εξετάστηκε ένα χρονικό διάστημα διάρκειας 90 δευτερολέπτων, κατά το οποίο η τιμή της ενεργειακής ζήτησης μεταβάλλεται συνεχώς και με τυχαίο τρόπο, όπως απεικονίζεται στο Σχήμα 4.3. Κατά τη διάρκεια κάθε προσομοίωσης, η χρονοσειρά που αναπαριστά τη μεταβολή της ζήτησης διαφοροποιείται τυχαία, προκειμένου να αναπαρασταθούν ρεαλιστικότερες συνθήκες λειτουργίας του συστήματος.



Σχήμα 4.3: Πραγματική μεταβολή της ζητούμενης ισχύος στην περιοχή 1.

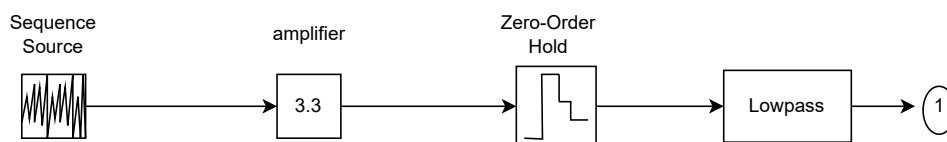
Η μοντελοποίηση του υποσυστήματος που περιέχει την μήτρα παραγωγής των τυχαίων χρονοσειρών διακύμανσης εισόδου φαίνεται στην εικόνα 4.4. Παρατηρώντας κανείς με προσοχή τη συγκεκριμένη εικόνα μπορεί να παρατηρήσει ότι αποτελείται από μια πηγή σήματος, η οποία λαμβάνει τιμές από μια ημιτονοειδή συνάρτηση ορισμένη στο MATLAB. Οι τιμές των συνιστωσών που τη συγκροτούν αλλάζουν για κάθε προσομοίωση διαφορετικού ζευγαριού (H_1, H_2) , ενισχύοντας με αυτόν τον τρόπο τη στοχαστική φύση των διαταραχών του προβλήματος. Το πλάτος των ημιτονοειδών σημάτων έχει οριστεί από τη συνάρτηση του MATLAB να κυμαίνεται σε ένα εύρος τιμών p.u., παρόμοιο με εκείνο μικρών ρεαλιστικών διαταραχών στο δίκτυο $([-1.5, 1.5] \text{ p.u.})$.

Αμέσως μετά το σήμα εισόδου οδηγείται σε έναν ενισχυτή (amplifier), ο οποίος διευρύνει

το συγκεκριμένο σύνολο τιμών κατά 3.3 φορές. Ο σκοπός της συγκεκριμένης διάταξης είναι να εξουδετερώσει την επίδραση των επόμενων blocks, τα οποία έχουν την τάση να ελαττώνουν το πλάτος με την τιμή 3.3 να αποδεικνύεται, ύστερα από μεγάλο αριθμό πειραμάτων, ως το ιδανικό μέγεθος ενίσχυσης ώστε το τελικό αποτέλεσμα εξόδου να κυμαίνεται πάλι στο επιθυμητό εύρος.

Η χρονοσειρά στη συνέχεια εισέρχεται στο block τύπου Zero-Order Hold (ZOH), του οποίου η λειτουργία είναι η μετατροπή ενός σήματος συνεχούς χρόνου σε διακριτό. Η τιμή που λαμβάνεται τη χρονική στιγμή της δειγματοληψίας διατηρείται σταθερή σε όλο το χρονικό διάστημα μεταξύ δύο διαφορετικών δειγμάτων, το οποίο στη δική μας περίπτωση επιλέγεται ίσο με 0.01 s (σύνολο 9001 χρονικά δείγματα). Η συγκεκριμένη διάταξη γεφυρώνει τον αναλογικό με τον ψηφιακό κόσμο και είναι ιδιαίτερα δημοφιλής σε περιπτώσεις που χρειάζεται να συνδυάσουμε διακριτά blocks (π.χ. φίλτρα) με συνεχή σήματα στο ίδιο μοντέλο.

Τέλος, το πλέον διακριτοποιημένο σήμα προωθείται σε ένα χαμηλοπερατό φίλτρο (Low Pass FIR), το οποίο έχει σκοπό να απομακρύνει τις υψηλές συχνότητες από το σήμα, αφήνοντας να περάσουν μόνο οι χαμηλές. Με αυτόν τον τρόπο μειώνεται ο θόρυβος και εξομαλύνεται η κυματομορφή, ενώ διατηρείται η κύρια μορφή του σήματος. Με απλά λόγια, το συγκεκριμένο φίλτρο καθαρίζει το σήμα από γρήγορες, ανεπιθύμητες μεταβολές και κρατάει την αργή βασική μορφή του. Στο συγκεκριμένο φίλτρο επιλέχθηκε συνχोτητα διέλευσης (passband frequency) ίση με 1 Hz, πράγμα που σημαίνει ότι όλες οι συχνότητες μέχρι το 1 Hz περνούν σχεδόν χωρίς εξασθένηση, ενώ η συχνότητα αποκοπής (stopband frequency) ορίστηκε ίση με 5 Hz, υποδεικνύοντας ότι από τα 5 Hz και πάνω το σήμα αποκόπτεται σημαντικά. Το διάστημα ανάμεσα στις δύο τιμές (1–5 Hz) είναι η ζώνη μετάβασης, όπου η εξασθένηση αυξάνεται προοδευτικά.

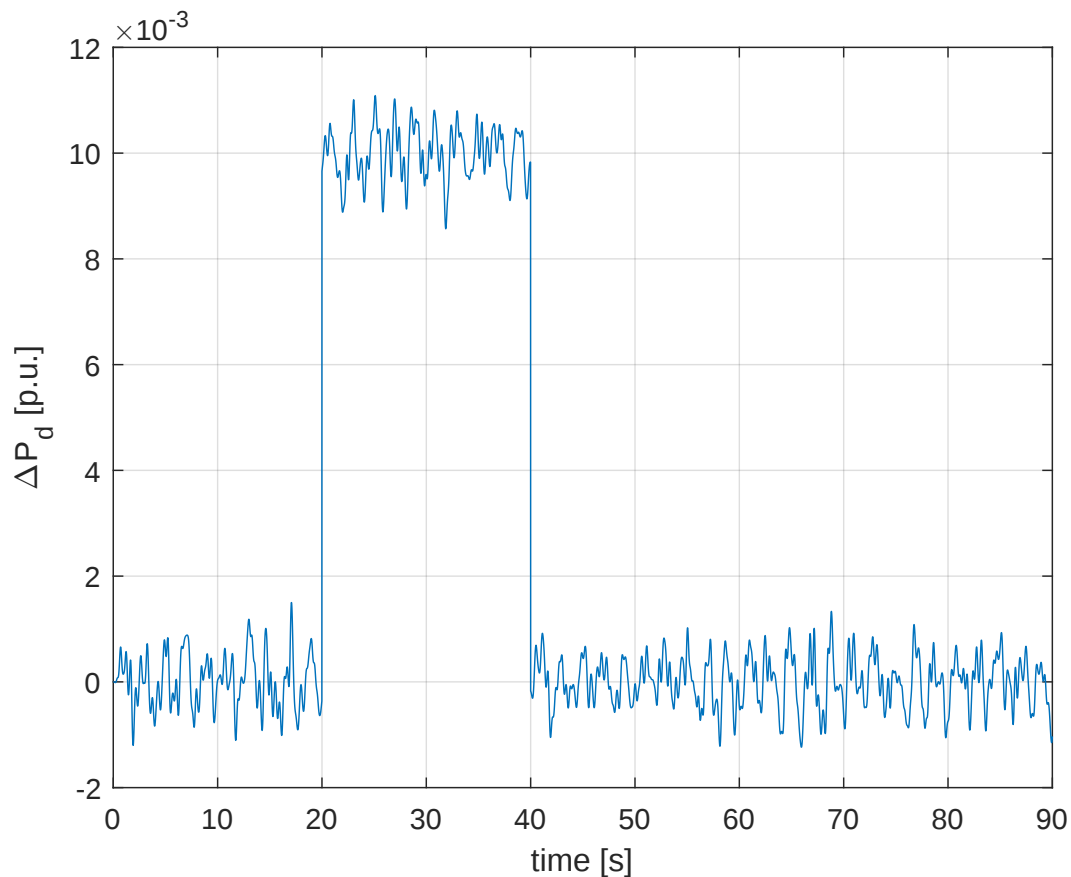


Σχήμα 4.4: Μοντελοποίηση μίτρας τυχαίων διαταραχών εισόδου.

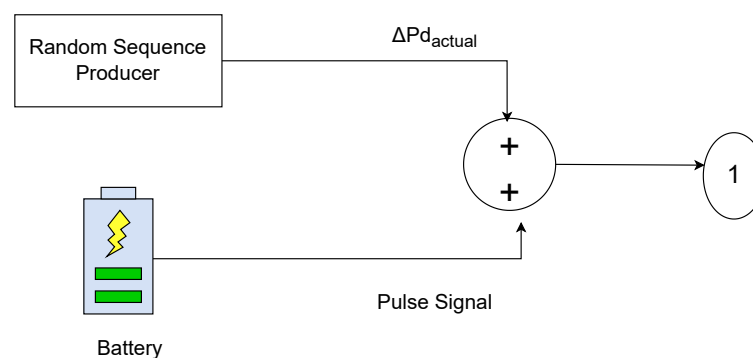
Η στοχαστικότητα της μεταβολής της ζητούμενης ισχύος παρατηρείται, όπως είναι φυσικό, τόσο στην περιοχή 1 όσο και στην περιοχή 2. Λόγω όμως αυτής της στοχαστικής φύσης, όπως παρατηρήθηκε κατά τη διάρκεια των προσομοιώσεων, τα μοντέλα μηχανικής μάθησης αδυνατούν να εξάγουν ασφαλείς και αξιόπιστες εκτιμήσεις για τις σταθερές αδράνειας σε τόσο μικρές και διαρκώς μεταβαλλόμενες διαταραχές. Εάν κάποιος ανατρέξει στη διεθνή βιβλιογραφία, θα διαπιστώσει ότι στην πλειοψηφία των περιπτώσεων η σταθερά αδράνειας προσδιορίζεται με ακρίβεια κατά τη διάρκεια σχετικά μεγάλων και συνήθως βηματικών διαταραχών [13, 14]. Επειδή όμως οι μεγάλες διαταραχές ενδέχεται να προκαλέσουν αποσυντονισμό του συστήματος και να θέσουν το δίκτυο σε κίνδυνο, καταφεύγουμε στη χρήση δοκιμαστικών ελεγχόμενων διαταραχών (probing signals).

Πιο συγκεκριμένα, για να μετριάσουμε τα προβλήματα που προκαλεί η στοχαστικότητα της ζήτησης στην έξοδο του συστήματος, σε κάθε προσομοίωση ενσωματώνουμε στα φορτία της Περιοχής 1 μία μπαταρία, την οποία θέτουμε σε λειτουργία φόρτισης κατά το χρονικό διάστημα 20–40 s. Η ενεργοποίηση της φόρτισης εφαρμόζεται με πολύ μικρό χρόνο ανόδου, ώστε η επαγόμενη μεταβολή στο ισοζύγιο ισχύος να προσεγγίζει ικανοποιητικά μία βηματική διαταραχή. Μετά τη λήξη του διαστήματος αυτού, η μπαταρία τίθεται εκ νέου σε εκφόρτιση. Κατ' αυτόν τον τρόπο, το τελικό σήμα ζήτησης προκύπτει ως συνδυασμός δύο διαδοχικών βηματικών μεταβολών, δηλαδή μίας αύξησης λόγω της φόρτισης και μίας μείωσης λόγω της εκφόρτισης, οι οποίες υπερτίθενται στην πραγματική, τυχαία μεταβαλλόμενη ζήτηση, όπως φαίνεται στο Σχήμα 4.5. Το αντίστοιχο υποσύστημα ΔP_d απεικονίζεται στο Σχήμα 4.6.

Η εφαρμογή του probing signal αρκεί να γίνει σε μία περιοχή του ΣΗΕ, την οποία εμείς επιλέξαμε τυχαία ως Περιοχή 1. Αντίθετα, στη δεύτερη διασυνδεδεμένη περιοχή, η μεταβλητή ΔP_{d2} αντιστοιχεί άμεσα στην πραγματική τιμή της μεταβολής της ζητούμενης ισχύος της Περιοχής 2. Με την υιοθέτηση αυτής της διάταξης, διευκολύνουμε σε σημαντικό βαθμό την ταυτοποίηση των παραμέτρων αδράνειας H , όπως θα αναλυθεί στην επόμενη ενότητα.



Σχήμα 4.5: Τελικό σήμα ζητούμενης ισχύος, όπως προκύπτει από τον συνδυασμό της πραγματικής καμπύλης ζήτησης με τη βηματική μεταβολή που προκαλεί η φόρτιση της μπαταρίας.

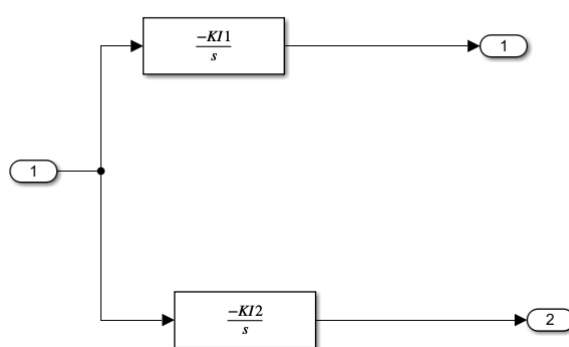


Σχήμα 4.6: Υλοποίηση υποσυστήματος ΔP_d .

4.2.3 Ελεγκτής (Controller)

Κάθε ένα από τα δύο υποσυστήματα που διακρίνονται στο Σχήμα 4.1 αποτελείται από την παράλληλη σύνδεση δύο ολοκληρωτικών ελεγκτών, οι οποίοι πολλαπλασιάζονται με τον όρο -1 . Το αρνητικό πρόσημο είναι απαραίτητο, διότι ένα θετικό σφάλμα συχνότητας πρέ-

πει να οδηγεί σε εντολή μείωσης της ισχύος εξόδου. Κάθε ολοκληρωτικός ελεγκτής πολλαπλασιάζεται με τον όρο K_I που ονομάζεται ολοκληρωτικό κέρδος και επηρεάζει σε μεγάλο βαθμό τον χρόνο αποκατάστασης δηλαδή τον χρόνο που χρειάζεται να μηδενιστεί το σφάλμα. Έπειτα ο καθένας από τους ελεγκτές αντιστοιχίζεται με μια και μόνο μονάδα παράγωγης του υποσυστήματος Power Plants. Στόχος της συγκεκριμένης διάταξης είναι η προσομοίωση της δυναμικής συμπεριφοράς της δευτερεύουσας ρύθμισης συχνότητας. Το διάγραμμα του υποσυστήματος Controller της Περιοχής 1 παρουσιάζεται στο Σχήμα 4.7, ενώ αντίστοιχη είναι και η υλοποίηση για την Περιοχή 2.



Σχήμα 4.7: Υλοποίηση υποσυστήματος Controller περιοχής 1

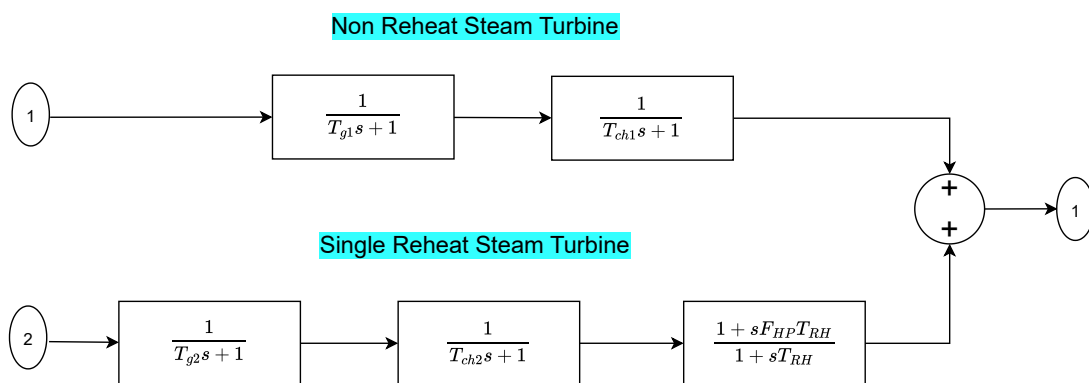
4.2.4 Υποσυστήματα τύπου Power Plants Area

Όπως αναφέρθηκε και στην αρχή της παρούσας ενότητας, έχουν αναπτυχθεί πολλαπλά σενάρια που βασίζονται σε διαφορετικούς συνδυασμούς μονάδων παραγωγής. Η διαφοροποίηση της υλοποίησης κάθε σεναρίου στο Simulink εντοπίζεται αποκλειστικά στο τελευταίο είδος υποσυστήματος που απομένει να αναλυθεί, δηλαδή στο *Power Plants Area*. Δύο τέτοια υποσυστήματα εμφανίζονται στη διάταξη του Σχήματος 4.1: το πρώτο περιλαμβάνει τις μονάδες παραγωγής της Περιοχής 1 και το δεύτερο τις μονάδες παραγωγής της Περιοχής 2.

Στο πλαίσιο της εργασίας εξετάστηκαν συνολικά τρία διακριτά σενάρια, τα οποία προκύπτουν από εναλλακτικούς συνδυασμούς μονάδων παραγωγής για δύο περιοχές του εξεταζόμενου ΣΗΕ. Τα σενάρια αυτά παρουσιάζονται και αναλύονται λεπτομερώς στην συνέχεια.

ΣΕΝΑΡΙΟ 1

Στο παρόν σενάριο, η Περιοχή 1 αποτελείται από την παράλληλη σύνδεση δύο διαφορετικών σύνολων συναρτήσεων μεταφοράς, οι οποίες περιγράφουν τη δυναμική συμπεριφορά δύο τύπων ατμοστροβίλων. Στην πρώτη περίπτωση, η συνάρτηση μεταφοράς αντιστοιχεί σε έναν ατμοστρόβilo μονής αναθερμάνσης (single reheat steam turbine), ενώ στη δεύτερη περιγράφει την λειτουργία ατμοστροβίλου χωρίς αναθερμάνση (non-reheat steam turbine). Το υποσύστημα της Περιοχής 1, το οποίο αποτελείται από αυτήν την παράλληλη σύνδεση των block διαγραμμάτων του κάθε ατμοστροβίλου, φαίνεται στην εικόνα 4.8.



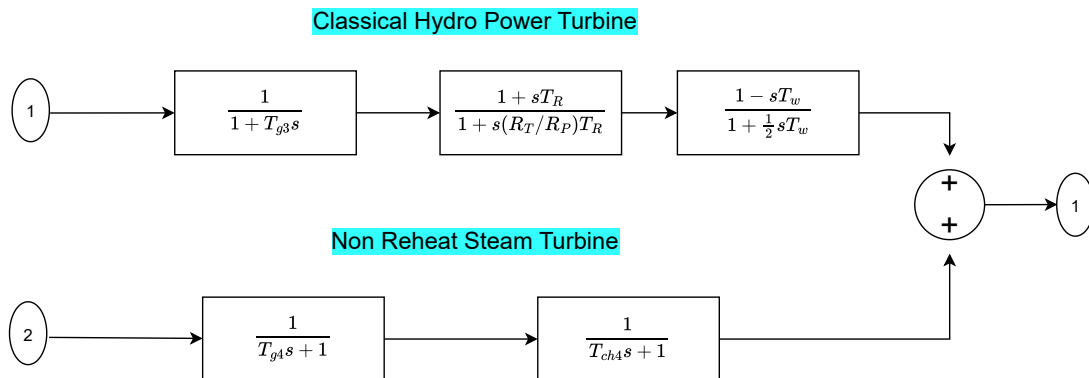
Σχήμα 4.8: Διάγραμμα βαθμίδων του υποσυστήματος Power Plants Area 1.

Σε αυτό το σχήμα παρατηρούμε ότι η είσοδος οδηγείται σε δυο κλάδους. Ο πάνω κλάδος περιγράφει την δυναμική συμπεριφορά ενός ατμοστροβίλου που λειτουργεί χωρίς αναθέρμανση. Η παράμετρος T_{g1} παριστάνει την χρονική σταθερά του ρυθμιστή στροβίλου ενώ η μεταβλητή T_{ch1} παριστάνει χρονική σταθερά του θαλάμου υψηλής πίεσης.

Όσον αφορά τον κάτω κλάδο, περιγράφονται οι συναρτήσεις μεταφοράς που συνιστούν έναν ατμοστρόβilo με λειτουργία μονής αναθέρμανσης. Στον δεύτερο κλάδο λοιπόν, υπάρχει ένα επιπλέον στάδιο αναθέρμανσης μετά τον θάλαμο υψηλής πίεσης, όπου ο ατμός επαναθερμαίνεται προτού συνεχίσει προς τον στρόβilo. Η μεταβλητή T_{RH} παριστάνει τη χρονική σταθερά του αναθερμαντήρα ενώ η F_{HP} παριστάνει το ποσοστό της ισχύος που διέρχεται από το στάδιο υψηλής πίεσης. Οι τιμές T_{g2} , T_{ch2} αντικατοπτρίζουν τις αντίστοιχες χρονικές σταθερές του δεύτερου ατμοστροβίλου. Στο τέλος αυτοί οι δυο κλάδοι αθροίζονται για να δώσουν την τελική έξοδο του υποσυστήματος.

Η Περιοχή 2 με την σειρά της δομείται από την παράλληλη σύνδεση δυο block διαγραμμάτων που προσομοιώνουν την λειτουργία ενός υδροστροβίλου (hydro turbine) και ξανά

ενός ατμοστροβίλου χωρίς αναθέρμανση. Το εσωτερικό του υποσυστήματος των μονάδων παραγωγής της Περιοχής 2 παρουσιάζεται αναλυτικά στο Σχήμα 4.9.



Σχήμα 4.9: Διάγραμμα βαθμίδων του υποσυστήματος Power Plants Area 2.

Όσον αφορά τον ατμοστροβίλο χωρίς αναθέρμανση, η δομή του είναι πανομοιότυπη με εκείνη του αντίστοιχου υποσυστήματος της Περιοχής 1. Η παράμετρος T_{g4} παριστάνει τη χρονική σταθερά του ρυθμιστή στροβίλου, ενώ η παράμετρος T_{ch4} εκφράζει τη χρονική σταθερά του θαλάμου υψηλής πίεσης. Για λόγους διαφοροποίησης των σεναρίων και ρεαλιστικότερης προσομοίωσης, οι αριθμητικές τιμές των παραμέτρων αυτών τέθηκαν ελαφρώς διαφορετικές σε σχέση με τις αντίστοιχες του ατμοστροβίλου της Περιοχής 1.

Ο υδροηλεκτρικός σταθμός, από την άλλη πλευρά, προσομοιώνει τη δυναμική συμπεριφορά ενός τυπικού υδροστροβίλου. Η παράμετρος T_{g3} αντιπροσωπεύει τη χρονική σταθερά του ρυθμιστή, ενώ η μεταβλητή T_R παριστάνει τη χρονική σταθερά του ίδιου του στροβίλου. Οι μεταβλητές R_T και R_P σχετίζονται αντίστοιχα με την υδραυλική αντίσταση και την πίεση του νερού στο σύστημα, ενώ ο λόγος τους (R_T/R_P) επηρεάζει άμεσα τη δυναμική απόκριση του στροβίλου. Τέλος, ο όρος T_w εκφράζει τη χρονική σταθερά του συστήματος νερού (*Water Time Constant*), η οποία σχετίζεται με τη ροή του νερού εντός του αγωγού πίεσης.

Οι μονάδες και η φυσική σημασία όλων των παραμέτρων των διαγραμμάτων που περιγράφουν την δυναμική συμπεριφορά των μονάδων παραγωγής φαίνονται στον πίνακα 4.1

Πίνακας 4.1: Συνοπτική παρουσίαση παραμέτρων μοντέλων Περιοχής 1 και Περιοχής 2

Περιοχή	Σύμβολο	Μονάδα	Φυσική σημασία
Περιοχή 1	T_{g1}	s	Χρονική σταθερά ρυθμιστή στροβίλου (governor) ατμοστροβίλου χωρίς αναθέρμανση.
	T_{ch1}	s	Χρονική σταθερά θαλάμου υψηλής πίεσης (steam chest).
	T_{g2}	s	Χρονική σταθερά ρυθμιστή στροβίλου (governor) ατμοστροβίλου μονής αναθέρμανσης.
	T_{ch2}	s	Χρονική σταθερά θαλάμου υψηλής πίεσης (steam chest).
	T_{RH}	s	Χρονική σταθερά αναθερμαντήρα στον ατμοστροβίλο με αναθέρμανση.
	F_{HP}	-	Ποσοστό ισχύος που διέρχεται από το στάδιο υψηλής πίεσης στον ατμοστροβίλο με αναθέρμανση.
Περιοχή 2	T_{g4}	s	Χρονική σταθερά ρυθμιστή στροβίλου (governor) ατμοστροβίλου χωρίς αναθέρμανση.
	T_{ch4}	s	Χρονική σταθερά θαλάμου υψηλής πίεσης ατμοστροβίλου.
	T_{g3}	s	Χρονική σταθερά ρυθμιστή στροβίλου (governor) υδροστροβίλου.
	T_R	s	Χρονική σταθερά υδροστροβίλου (αδράνεια δρομέα και μηχανισμού ροής).
	R_T, R_P	-	Παράμετροι υδραυλικής αντίστασης και πίεσης νερού στον αγωγό πίεσης· ο λόγος τους επηρεάζει τη δυναμική απόκριση.
	T_w	s	Χρονική σταθερά νερού (Water Time Constant), που εκφράζει τη δυναμική καθυστέρηση ροής στον αγωγό πίεσης.

Σενάριο 2

Το συγκεκριμένο σενάριο χαρακτηρίζεται από αυξημένη διείσδυση υδροηλεκτρικών μονάδων. Στην Περιοχή 1, περιλαμβάνονται δύο διαφορετικά block συναρτήσεων μεταφοράς, οι οποίες περιγράφουν τη λειτουργία δύο διαφορετικών υδροστροβίλων (υδροστροβίλος 1, υδροστροβίλος 2). Οι δύο αυτές μονάδες είναι συνδεδεμένες παράλληλα και η ισχύς εξόδου τους αθροίζεται, ώστε να προκύψει η συνολική παραγόμενη ισχύς της Περιοχής 1. Η Περιοχή 2, από την άλλη πλευρά, αποτελείται από τον παράλληλο συνδυασμό ενός ατμοστροβίλου μονής αναθέρμανσης και ενός χωρίς καθόλου αναθέρμανση.

Οι τιμές των παραμέτρων που χρησιμοποιούνται στην υλοποίηση της διάταξης των δύο διασυνδεδεμένων περιοχών ελέγχου είτε αυτές σχετίζονται με τις μονάδες παραγωγής, είτε με τα μεγέθη της πρωτεύουσας και δευτερεύουσας ρύθμισης συχνότητας ενδέχεται να διαφοροποιούνται από σενάριο σε σενάριο, με στόχο την εξασφάλιση της ευσταθούς λειτουργίας του δικτύου.

Σενάριο 3

Σε αυτό το σενάριο, η Περιοχή 1 οικοδομείται από την παράλληλη λειτουργία ενός γραμμικού μοντέλου που περιγράφει έναν υδροστροβίλο και ενός ατμοστροβίλου μονής αναθέρμανσης. Πανομοιότυπη με την Περιοχή 1 είναι η υλοποίηση της Περιοχής 2, όπου συναντώνται οι ίδιες μονάδες παραγωγής, αλλά με μικρές διακυμάνσεις στις τιμές των παραμέτρων τους. Το σενάριο αυτό αντιπροσωπεύει μια πιο ισορροπημένη κατανομή μονάδων, καθώς και στις δύο περιοχές συνδυάζονται μια υδροηλεκτρική και μια θερμική μονάδα με αναθέρμανση, γεγονός που επιτρέπει τη μελέτη της αλληλεπίδρασης των δύο τεχνολογιών σε συνθήκες διασύνδεσης.

4.2.5 Συνοπτική Περιγραφή Σεναρίων

Οι μονάδες παραγωγής που δομούν τις περιοχές του εξεταζόμενου ΣΗΕ για κάθε σενάριο συνοψίζονται στον Πινάκα 4.2.

Πίνακας 4.2: Συνοπτική παρουσίαση των σεναρίων μονάδων παραγωγής στις δύο περιοχές

Σενάριο	Περιοχή 1	Περιοχή 2
Σενάριο 1	Ατμοστρόβιλος χωρίς αναθέρμανση, Ατμοστρόβιλος μονής αναθέρμανσης	Ατμοστρόβιλος χωρίς αναθέρμανση, Υδροστρόβιλος
Σενάριο 2	Υδροστρόβιλος 1, Υδροστρόβιλος 2	Ατμοστρόβιλος χωρίς αναθέρμανση, Ατμοστρόβιλος μονής αναθέρμανσης
Σενάριο 3	Υδροστρόβιλος, Ατμοστρόβιλος μονής αναθέρμανσης	Υδροστρόβιλος, Ατμοστρόβιλος μονής αναθέρμανσης

4.2.6 Λοιπές παράμετροι του δικτύου

Πέρα από τα τέσσερα βασικά υποσυστήματα που περιλαμβάνονται στη διάταξη του Σχήματος 4.1, υπάρχουν και ορισμένες επιπλέον παράμετροι η οποίες σχετίζονται με τους μηχανισμούς της πρωτεύουσας και δευτερεύουσας ρύθμισης συχνότητας καθώς και την λειτουργικότητα της διασύνδεσης μεταξύ των περιοχών έλεγχου του ΣΗΕ. Αρχικά, παρατηρούμε τις σταθερές R_1, R_2, R_3, R_4 , καθεμία εκ των οποίων αντιστοιχεί στο στατισμό μίας μονάδας παραγωγής. Επιπλέον, οι παράμετροι B_1 και B_2 , συμβολίζουν τους συντελεστές πόλωσης και προσδιορίζουν την απόκριση του δευτερογενούς ελέγχου κάθε περιοχής. Η τιμή τους δίνεται από τη σχέση

$$B_i = \frac{1}{R_{eq,i}} + D_i \quad (\text{σε p.u.}),$$

όπου $R_{eq,i}$ είναι η ισοδύναμη τιμή του στατισμού των μονάδων της i περιοχής και D_i ο συντελεστής αυτορύθμισης φορτίου της αντίστοιχης περιοχής.

Τέλος, παρατηρούμε τη συνάρτηση μεταφοράς

$$\frac{2\pi T_{12}}{s},$$

όπου η σταθερά T_{12} ονομάζεται *σταθερά συγχρονισμού* μεταξύ της Περιοχής 1 και της Περιοχής 2 και μετριέται σε δευτερόλεπτα. Η παράμετρος αυτή εκφράζει τη δυναμική σύζευξη των δύο περιοχών, δηλαδή το πόσο γρήγορα και σε ποιο βαθμό οι μεταβολές ισχύος στη μία περιοχή επηρεάζουν τη συχνότητα της άλλης.

4.2.7 Αριθμητικές Τιμές Παραμέτρων

Όπως αναφέρθηκε και προηγουμένως, κάθε σενάριο διαφοροποιείται από τα υπόλοιπα, αφενός λόγω της χρήσης διαφορετικού μείγματος παραγωγής (διαφορετικού τύπου μονάδων) και αφετέρου λόγω της χρήσης διαφορετικών τιμών στις παραμέτρους των μοντέλων. Για ορισμένες παραμέτρους, όπως για παράδειγμα η κλίση των καμπυλών στατισμού και τα κέρδη ολοκλήρωσης K_I των ελεγκτών, η διαφοροποίηση κρίθηκε απαραίτητη ώστε το δίκτυο να παραμένει ευσταθές. Αντίθετα, άλλες παράμετροι μεταβλήθηκαν ελαφρώς ανά σενάριο για να ανταποκρίνονται σε πιο ρεαλιστικές συνθήκες.

Ενδέχεται το σημείο στο οποίο εμφανίζεται μια παράμετρος να διαφέρει από σενάριο σε σενάριο, γεγονός που σημαίνει ότι η ίδια ονομασία μπορεί να αντιστοιχεί σε διαφορετικό φυσικό μέγεθος. Για παράδειγμα, η παράμετρος T_{g1} στο Σενάριο 1 δηλώνει τη χρονική σταθερά του ρυθμιστή στροβίλου ενός ατμοστροβίλου με λειτουργία μονής αναθέρμανσης, ενώ στο Σενάριο 3 δηλώνει τη χρονική σταθερά του ρυθμιστή ενός υδροστροβίλου. Παρακάτω συνοψίζονται σε πινάκες οι τιμές των παραμέτρων για κάθε σενάριο καθώς αναλύεται παράλληλα και η φυσική τους σημασία σε κάθε περίπτωση.

Σενάριο 1

Όσον αφορά το σενάριο 1 η φυσική σημασία των παραμέτρων αναλύθηκε προηγουμένως στον πίνακα 4.1 ενώ στον πίνακα 4.3 παρουσιάζονται συγκεντρωτικά οι αριθμητικές τιμές του σύνολου των παραμέτρων που συναντώνται στο συγκεκριμένο σενάριο. Τρεις χρωματιστές γραμμές στον πίνακα υποδηλώνουν το διαχωρισμό μεταξύ των δύο περιοχών και του ενδιάμεσου δικτύου, καθιστώντας σαφές αν η παράμετρος ανήκει σε μονάδα παραγωγής που λειτουργεί στην Περιοχή 1, στην Περιοχή 2 ή αν συναντάται ανάμεσα τους.

Πίνακας 4.3: Πίνακας παραμέτρων σεναρίου 1

Παράμετρος	Αριθμητική Τιμή
D_1	0.9
T_{g1}	0.23
T_{ch1}	0.39
T_{g2}	0.14
T_{ch2}	0.5
T_{rh}	7
F_{hp}	0.05
R_1	0.05
R_2	0.05
K_{I1}	0.25
K_{I2}	0.25
B_1	40.9
D_2	1.5
T_{g3}	0.4
T_w	2.25
T_r	5
R_t	0.38
R_p	0.05
T_{g4}	0.3
T_{ch4}	0.45
R_3	0.05
R_4	0.05
K_{I3}	0.25
K_{I4}	0.25
B_2	41.5
T_{12}	2
a_{12}	-1

Σενάριο 2

Στο Σενάριο 2, η Περιοχή 1 αποτελείται, όπως ήδη αναφέρθηκε, από την παράλληλη σύνδεση δύο blocks που αναπαριστούν δύο διαφορετικούς υδροστρόβιλους. Στο παρόν σενάριο,

οι παράμετροι T_{g3} και T_{g4} αντιστοιχούν στις χρονικές σταθερές των ρυθμιστών των αντίστοιχων υδροστροβίλων, ενώ οι όροι T_R , R_T , R_P και T_w , οι οποίοι παραμένουν κοινοί και στα δύο μοντέλα, αντιστοιχούν στα φυσικά μεγέθη της χρονικής σταθεράς του υδροστροβίλου, της υδραυλικής αντίστασης, της πίεσης του νερού και της χρονικής σταθεράς του νερού.

Όσον αφορά την Περιοχή 2, αναφέραμε ότι συγκροτείται από δυο ατμοστροβίλους εκ των οποίων ο ένας δεν λειτουργεί με αναθέρμανση ενώ ο δεύτερος αναθερμαίνει μια φορά τον ατμό του. Οι παράμετροι T_{g1} και T_{g2} αντιστοιχούν στις χρονικές σταθερές των ρυθμιστών του κάθε στροβίλου, ενώ οι όροι T_{ch1} και T_{ch2} εκφράζουν τη χρονική σταθερά των θαλάμων υψηλής πίεσης του ατμοστροβίλου χωρίς αναθέρμανση και του στροβίλου με μονή αναθέρμανση, αντίστοιχα. Τέλος, οι παράμετροι T_{RH} και F_{HP} αφορούν αποκλειστικά τον ατμοστροβίλο με μονή αναθέρμανση και αντιστοιχούν στη χρονική σταθερά του αναθερμαντήρα και στο ποσοστό ισχύος που διέρχεται από το στάδιο υψηλής πίεσης, αντίστοιχα.

Οι αριθμητικές τιμές όλων των παραμέτρων των μοντέλων που συνθέτουν τις δύο περιοχές του Σενάριου 2 παρουσιάζονται στον Πίνακα 4.4.

Σενάριο 3

Στο σενάριο 3, και οι δύο περιοχές συντίθενται από την παράλληλη σύνδεση δύο γραμμικών μοντέλων, τα οποία αναπαριστούν έναν υδροστροβίλο και έναν ατμοστροβίλο μονής επαναθέρμανσης.

Στα πλαίσια του συγκεκριμένου σεναρίου, οι χρονικές σταθερές T_{g1} και T_{g2} αντιπροσωπεύουν τις σταθερές των ρυθμιστών στροβίλου για τον υδροστροβίλο και τον ατμοστροβίλο της Περιοχής 1, αντίστοιχα. Οι παράμετροι T_R , R_T , R_P και T_w διατηρούν την ίδια σημασία όπως και στα προηγούμενα σενάρια. Επιπλέον, ο όρος T_{ch2} αποτελεί τη χρονική σταθερά του θαλάμου υψηλής πίεσης του ατμοστροβίλου στην Περιοχή 1.

Αντίστοιχα, στην Περιοχή 2, οι παράμετροι T_{g3} και T_{g4} σχετίζονται με τις χρονικές σταθερές των ρυθμιστών στροβίλου για τον υδροστροβίλο και τον ατμοστροβίλο, ενώ ο όρος T_{ch4} αντιστοιχεί στη χρονική σταθερά του θαλάμου υψηλής πίεσης του ατμοστροβίλου. Οι αριθμητικές τιμές όλων των παραμέτρων παρουσιάζονται στον Πίνακα 4.5.

Πίνακας 4.4: Πίνακας παραμέτρων σεναρίου 2

Παράμετρος	Αριθμητική Τιμή
D_1	0.9
T_{g1}	0.23
T_{ch1}	0.39
T_{g2}	0.14
T_{ch2}	0.5
T_{rh}	7
F_{hp}	0.4
R_1	0.25
R_2	0.25
K_{I1}	0.25
K_{I2}	0.25
B_1	8.9
D_2	1.5
T_{g3}	0.4
T_w	2.25
T_r	5
R_t	0.38
R_p	0.05
T_{g4}	0.5
T_{ch4}	—
R_3	0.05
R_4	0.05
K_{I3}	0.25
K_{I4}	0.25
B_2	41.5
T_{12}	2
a_{12}	-1

Πίνακας 4.5: Πίνακας παραμέτρων σεναρίου 3

Παράμετρος	Αριθμητική Τιμή
D_1	0.9
T_{g1}	0.45
T_{g2}	0.14
T_{ch2}	0.5
T_{rh}	7
F_{hp}	0.4
R_1	0.25
R_2	0.05
K_{I1}	0.25
K_{I2}	0.25
B_1	24.9
D_2	1.5
T_{g3}	0.4
T_w	2.25
T_r	5
R_t	0.38
R_p	0.05
T_{g4}	0.16
T_{ch4}	0.55
R_3	0.25
R_4	0.05
K_{I3}	0.25
K_{I4}	0.25
B_2	25.5
T_{12}	2
a_{12}	-1

Κεφάλαιο 5

Ανάπτυξη Προτεινόμενων Μοντέλων

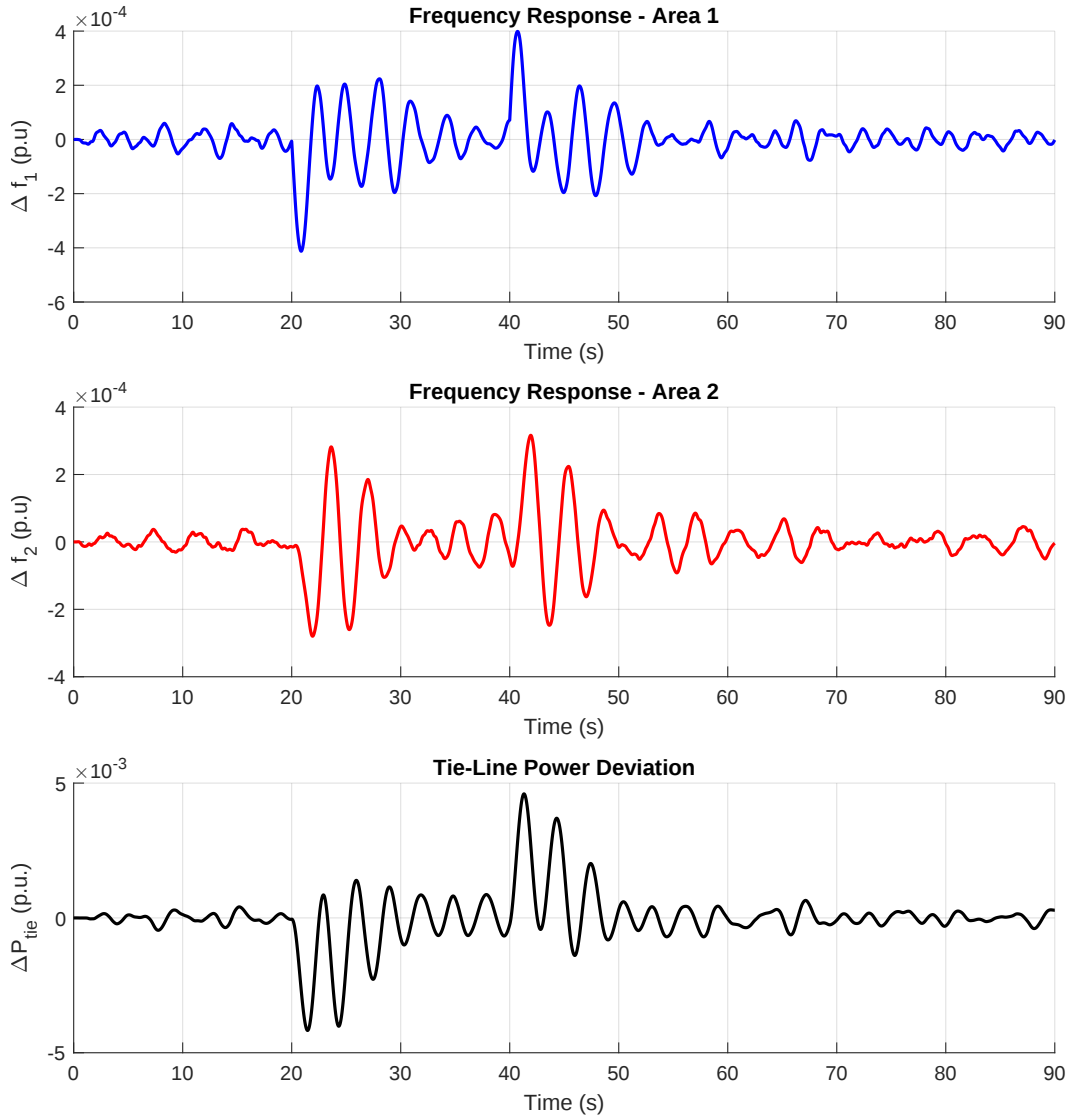
5.1 Εισαγωγή

Στο κεφάλαιο αυτό αναλύεται η προσέγγιση που ακολουθήθηκε για την ανάπτυξη των προτεινόμενων μοντέλων. Συγκεκριμένα, γίνεται αναλυτική αναφορά στον τρόπο συλλογής και επεξεργασίας των δεδομένων που θα χρησιμοποιηθούν για την ανάπτυξη των μεθόδων μηχανικής μάθησης. Επίσης, παρουσιάζεται λεπτομερώς η φιλοσοφία εκπαίδευσης των μοντέλων μηχανικής μάθησης για την εκτίμηση σταθερών αδράνειας από μετρήσεις ΣΗΕ.

5.2 Κατασκευή δεδομένων για την εκπαίδευση των μοντέλων

Πρώτο βήμα για την ανάπτυξη των μοντέλων αποτελεί η συλλογή των δεδομένων εκπαίδευσης. Λόγω του περιορισμένου βαθμού πρόσβασης σε επίσημα δεδομένα πραγματικών μετρήσεων, στην παρούσα διπλωματική αναπτύσσονται ρεαλιστικές αποκρίσεις συχνότητας ΣΗΕ με τη βοήθεια των μοντέλων απόκρισης συχνότητας που αναλύθηκαν στο Κεφάλαιο 4 και υλοποιήθηκαν στο Simulink. Με αυτό τον τρόπο δημιουργήθηκε ένα σύνολο τεχνητών δεδομένων (synthetic dataset). Πιο συγκεκριμένα, τα δεδομένα που δημιουργήθηκαν με τη βοήθεια του Simulink περιλαμβάνουν την απόκριση συχνότητας της Περιοχής 1 (Δf_1), την απόκριση συχνότητας της Περιοχής 2 (Δf_2), και τη μεταβολή της ισχύος στη γραμμή διασύνδεσης (ΔP_{tie}). Σε όλες τις περιπτώσεις που εξετάζονται, τα δεδομένα έχουν χρονική διάρκεια 90 δευτερολέπτων. Ένα ενδεικτικό παράδειγμα δυναμικών αποκρίσεων παρουσιάζεται στο Σχήμα 5.1. Τα μοντέλα μηχανικής μάθησης που θα αναπτυχθούν στα πλαίσια της διπλωμα-

τικής έχουν ως στόχο να προσδιορίσουν/εκτιμήσουν τις σταθερές αδράνειας H_1 και H_2 των δύο περιοχών του ΣΗΕ, χρησιμοποιώντας ως εισόδους τις δυναμικές αποκρίσεις Δf_1 , Δf_2 , και ΔP_{tie} .



Σχήμα 5.1: Παράδειγμα δυναμικών αποκρίσεων για τυχαίο ζεύγος (H_1 , H_2)

Κρατώντας σταθερές όλες τις παραμέτρους που αναλύθηκαν στο προηγούμενο κεφάλαιο, για κάθε σενάριο μονάδων παραγωγής λαμβάνουμε και αποθηκεύουμε τις αποκρίσεις της συχνότητας καθώς και τη μεταβολή της ισχύος στη διασύνδεση θεωρώντας έναν μοναδικό συνδυασμό τιμών για τις σταθερές αδράνειας H_1 και H_2 . Για κάθε συνδυασμό των H_1 και H_2 , το μοντέλο απόκρισης συχνότητας διεγείρεται με τυχαίες μικρές διαταραχές ισχύος και προκύπτουν οι δυναμικές αποκρίσεις Δf_1 , Δf_2 , και ΔP_{tie} , οι οποίες θα χρησιμοποιηθούν σε δεύτερο χρόνο ώστε να εκτιμηθούν οι τιμές H_1 και H_2 . Θα πρέπει να σημειωθεί, ότι εκτός από τις H_1 και H_2 και οι υπόλοιποι παράμετροι του συστήματος επηρεάζουν σε

κάποιο βαθμό τις αποκρίσεις Δf_1 , Δf_2 , και ΔP_{tie} [6, 28] όπως για παράδειγμα ο στατισμός R , και ο συντελεστής αυτορύθμισης φορτίου D . Σε κάθε περίπτωση, ο συγκεκριμένοι συντελεστές των μονάδων μπορούν να θεωρηθούν ότι έχουν γνωστές τιμές. Για το λόγο αυτό στην παρούσα διπλωματική εστιάζουμε στην εκτίμηση, μέσω μεθόδων μηχανικής μάθησης, των σταθερών αδράνειας H_1 και H_2 .

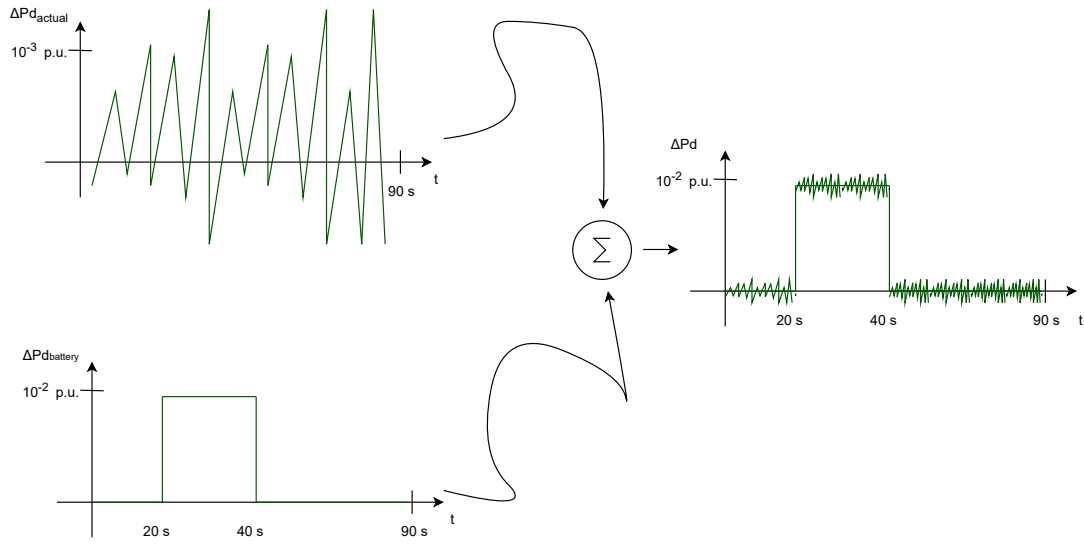
5.2.1 Κατασκευή δοκιμαστικού σήματος (Probing signal)

Θεωρώντας, λοιπόν, το σύνολο των παραμέτρων που διέπουν το διάγραμμα βαθμίδων του σχήματος 4.1 σταθερό σε κάθε προσομοίωση, στόχος μας είναι ο ακριβής προσδιορισμός των τιμών του ζεύγους (H_1, H_2) υπό την επίδραση τυχαίας διαταραχής ισχύος για δεδομένο σύνολο μονάδων παραγωγής. Ωστόσο, η απρόβλεπτη φύση της ζητούμενης ισχύος ανά πάσα στιγμή καθιστά δύσκολη την κατασκευή ενός μοντέλου μηχανικής μάθησης το οποίο λαμβάνοντας ως είσοδο τις αποκρίσεις συχνότητας και της ισχύος στη διασύνδεση να μπορεί να προβλέψει το ζεύγος των σταθερών αδράνειας H_1 και H_2 . Ο λόγος είναι ότι η μορφή των δυναμικών αποκρίσεων Δf_1 , Δf_2 , και ΔP_{tie} επηρεάζεται σε πολύ μεγάλο βαθμό από τη στοχαστική μεταβολή της ζητούμενης ισχύος, η οποία διαφοροποιείται τυχαία σε κάθε προσομοίωση.

Για να αντιμετωπιστεί αυτό το πρόβλημα καταφεύγουμε στη χρήση δοκιμαστικών σημάτων (probing signals). Τα probing signals είναι σήματα τα οποία επιβάλλονται από το διαχειριστή του συστήματος μέσω επιβαλλόμενης μεταβολής ισχύος. Με τον τρόπο αυτό, πραγματοποιείται μια μικρή, πλήρως ελεγχόμενη διαταραχή, η οποία επιτρέπει, όπως θα δούμε και στο κεφάλαιο με τα σχετικά αποτελέσματα, την πολύ ακριβή εκτίμηση των σταθερών αδράνειας H_1 και H_2 . Για τη δημιουργία των probing signals, εκμεταλλευόμαστε τη φύση των μπαταριών, οι οποίες χαρακτηρίζονται από την ικανότητα σχεδόν ακαριαίας ενεργοποίησης σε λειτουργία φόρτισης και εκφόρτισης. Έτσι, δημιουργούμε δοκιμαστικά σήματα που προσεγγίζουν ικανοποιητικά τη μορφή παλμών, με πλάτος μιας τάξης μεγέθους μεγαλύτερο σε σχέση με την τυχαία ζήτηση ισχύος.

Έτσι, το τελικό σήμα ΔP_d σε κάθε προσομοίωση προκύπτει από τον συνδυασμό των δύο διαταραχών, όπως φαίνεται και στο Σχήμα 5.2. Το αποτέλεσμα είναι ότι η μορφή των τελικών αποκρίσεων των Δf_1 , Δf_2 , ΔP_{tie} επηρεάζεται κυρίως από το επιβαλλόμενο δοκιμαστικό σήμα και πολύ λιγότερο από την τυχαία και μη προβλέψιμη μεταβολή του φορτίου.

Κατά συνέπεια, το πρόβλημα της τυχειότητας της εισόδου μετατρέπεται σε ψευδοτυχαίο,



Σχήμα 5.2: Κατασκευή δοκιμαστικού σήματος.

επιτρέποντας την αξιόπιστη συσχέτιση μεταξύ του ζεύγους (H_1, H_2) και των παρατηρούμενων αποκρίσεων. Με άλλα λόγια, οι τιμές των H_1 και H_2 και οι αποκρίσεις συχνότητας είναι πλέον πολύ πιο ισχυρά εξαρτώμενες. Με αυτόν τον τρόπο, διευκολύνεται σημαντικά η εκπαίδευση των μοντέλων μηχανικής μάθησης.

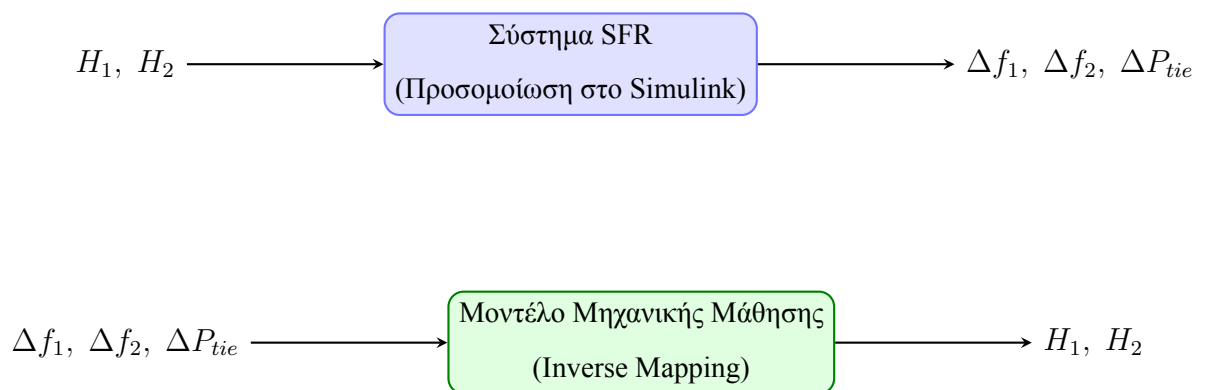
5.2.2 Εκτίμηση σταθερών αδράνειας

Για να κατανοήσει κανείς σε βάθος το πρόβλημα που καλούμαστε να επιλύσουμε, είναι χρήσιμο να ανατρέξει στην αρχική διάταξη των δύο διασυνδεδεμένων περιοχών στο Simulink, η οποία παρουσιάζεται στο Σχήμα 4.1. Το διάγραμμα αυτό, δεχόμενο ως είσοδο τις μεταβολές της ζητούμενης ισχύος στις δύο περιοχές $(\Delta P_{d1}, \Delta P_{d2})$, παράγει τρεις βασικές εξόδους: την απόκριση συχνότητας της Περιοχής 1 (Δf_1) , την απόκριση συχνότητας της Περιοχής 2 (Δf_2) καθώς και τη μεταβολή της ισχύος στη διασύνδεση (ΔP_{tie}) κατά την οποία υπάρχει ανταλλαγή ισχύος μεταξύ των περιοχών.

Αντικαθιστώντας όλες τις παραμέτρους που αναφέρθηκαν στο κεφάλαιο 4 συμπεριλαμβανομένων των αδρανειών H_1, H_2 με αριθμητικές τιμές που αντλούνται από σχετικά πειράματα στη βιβλιογραφία [6, 29], το μοντέλο παράγει για κάθε ζεύγος εισόδων $(\Delta P_{d1}, \Delta P_{d2})$ ένα σύνολο από τρεις χρονοσειρές εξόδου: Δf_1 , Δf_2 και ΔP_{tie} .

Όπως αναφέρθηκε και στην προηγούμενη υποενότητα, κατά τη διάρκεια κάθε προσομοίωσης το σύστημα δέχεται μία ψευδοτυχαία διαταραχή εισόδου, η οποία οφείλεται κυρίως στη λειτουργία φόρτισης-εκφόρτισης της μπαταρίας στην Περιοχή 1, ενώ όλες οι υπόλοιπες παράμετροι διατηρούνται σταθερές, πλην του ζεύγους H_1, H_2 . Με στόχο να εξετάσουμε την ευαισθησία του συστήματος σε διαφορετικές τιμές ισοδύναμων αδρανειών ανά περιοχή οι οποίες αντανακλούν τον βαθμό διείσδυσης των ΑΠΕ στο δίκτυο χρησιμοποιούμε συνολικά 1000 διαφορετικά ζεύγη τιμών (H_1, H_2) , με κάθε τιμή να κυμαίνεται στο εύρος $[3, 8]$ s, το οποίο προσεγγίζει ρεαλιστικά σενάρια λειτουργίας. Επειδή πλέον η μοναδική παράμετρος που μπορεί να επηρεάσει το αποτέλεσμα είναι η αδράνεια προσπαθούμε να κατασκευάσουμε ένα μοντέλο που θα συνδέει την αδράνεια με τις πραγματικές αποκρίσεις του συστήματος διαμορφώνοντας έτσι μια σχέση εισόδου - εξόδου μεταξύ αδρανειών και αποκρίσεων του συστήματος.

Ολόκληρη η συλλογιστική πίσω από την κατασκευή των μοντέλων μηχανικής μάθησης που θα χρησιμοποιηθούν βασίζεται στην απλή ιδέα ότι, από τη στιγμή που ένα μοναδικό ζεύγος αδρανειών H_1, H_2 καθορίζει τη μορφή των αποκρίσεων συχνότητας και της διασυνδεδεμένης ισχύος του συστήματος στο Σχήμα 4.1, τότε λογικά και ο αντίστροφος συλλογισμός θα πρέπει να ισχύει: δηλαδή, ένας μοναδικός συνδυασμός αποκρίσεων συχνότητας $(\Delta f_1, \Delta f_2)$ και διασυνδεδεμένης ισχύος (ΔP_{tie}) αντιστοιχεί αποκλειστικά σε ένα και μόνο ζεύγος τιμών H_1, H_2 . Συνεπώς, η διαδικασία της εκπαίδευσης θα στηρίζεται κάθε φορά στην αντιστοίχιση μιας ομάδας τριών χρονοσειρών με δύο συγκεκριμένες τιμές εξόδου, του ζεύγους αδρανειών.



Σχήμα 5.3: Forward mapping (πάνω): $(H_1, H_2) \mapsto (\Delta f_1, \Delta f_2, \Delta P_{tie})$ Inverse mapping (κάτω): $(\Delta f_1, \Delta f_2, \Delta P_{tie}) \mapsto (H_1, H_2)$

Στον πραγματικό κόσμο, τα δεδομένα που είναι διαθέσιμα για επεξεργασία είναι οι έξοδοι του συστήματος του Σχήματος 4.1 $(\Delta f_1, \Delta f_2, \Delta P_{tie})$ οι οποίες μπορούν εύκολα να λη-

φθούν στην πράξη μέσω μετρήτών ενώ αυτό που δεν μπορούμε να γνωρίζουμε και καλούμαστε να βρούμε είναι το ζεύγος των αδρανειών H_1 και H_2 , το οποίο είναι υπεύθυνο για τη συμπεριφορά των χρονοσειρών εξόδου του συστήματος. Για τον λόγο αυτό, εφαρμόζουμε μια αντιστροφή ως προς το τι θεωρούμε είσοδο και τι έξοδο στο πρόβλημα μας. Παρόλο που στην πραγματικότητα οι τιμές των H_1 και H_2 επηρεάζουν την έξοδο του συστήματος λειτουργώντας ως παράμετροι, εμείς τώρα θεωρούμε ότι η είσοδος του αντίστροφου προβλήματος είναι οι ίδιες οι χρονοσειρές (αποκρίσεις συχνότητας και ισχύος), ενώ η έξοδος είναι το ζεύγος των H_1 και H_2 , το οποίο καθορίζει μοναδικά το σχήμα των αποκρίσεων. Αυτή η διαδικασία ονομάζεται αντίστροφη αντιστοίχιση (inverse mapping) και η λογική λειτουργίας της απεικονίζεται στο Σχήμα 5.3.

Το πρόβλημα λοιπόν που καλούμαστε να επιλύσουμε στην επιστήμη δεδομένων αντιστοιχεί σε ένα sequence-to-value regression πρόβλημα. Σε τέτοιες περιπτώσεις στόχος είναι, δεδομένου ενός συνόλου χρονοσειρών, να μπορέσουμε να εκτιμήσουμε μια συσχετισμένη με αυτές βαθμωτή τιμή εξόδου. Προβλήματα τέτοιας φύσης συναντώνται ευρέως στη βιβλιογραφία, με την πρόβλεψη ζήτησης ενέργειας από ιστορικά δεδομένα φορτίου, την εκτίμηση της υπολειπόμενης διάρκειας ζωής βιομηχανικού εξοπλισμού βάσει μετρήσεων αισθητήρων, καθώς και την ανάλυση βιοϊατρικών σημάτων (π.χ. καρδιογραφήματα - ECG, ηλεκτροεγκεφαλογραφήματα - EEG) να αποτελούν μερικά από τα πιο κλασικά παραδείγματα [5]

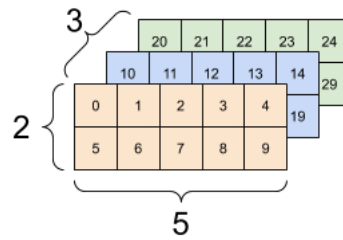
5.2.3 Δειγματοληψία και συλλογή δεδομένων

Οι δυναμικές αποκρίσεις των συχνοτήτων στις δύο περιοχές, καθώς και η δυναμική απόκριση της μεταβολής της ισχύος διασύνδεσης για ένα τυχαίο ζεύγος τιμών H_1 , H_2 (6.5, 5.86) του διαγράμματος 4.1, για το Σενάριο 1, παρουσιάστηκε προηγούμενος στο σχήμα 5.1.

Για να εξασφαλιστεί η αποτελεσματικότερη και οικονομικότερη αποθήκευση των χρονοσειρών, χωρίς απώλεια της πληροφορίας που περιέχουν, αποφασίστηκε να πραγματοποιηθεί δειγματοληψία κάθε χρονοσειράς με βήμα 0.01 s. Συνεπώς, καθεμία από τις χρονοσειρές, από τη χρονική στιγμή 0 έως και τα 90 s, θα αποτελείται συνολικά από 9001 δείγματα, τα οποία θα είναι ικανά να την ανοικοδομήσουν εφόσον χρειαστεί. Ο χρόνος των 90 s κρίθηκε κατάλληλος ως σημείο τερματισμού της καταγραφής, διότι μέσα σε αυτό το διάστημα προλαβαίνουν να ενεργοποιηθούν τόσο η πρωτεύουσα όσο και η δευτερεύουσα ρύθμιση συχνότητας έπειτα από την βηματική μεταβολή λόγω της μπαταρίας στα 20 και στα 40 δευτερόλεπτα.

Για την αποθήκευση, λοιπόν, του δειγματοληπτημένου σήματος για καθεμία από τις ομά-

δες των τριών χρονοσειρών που αντιστοιχούν σε καθένα από τα 1000 διαφορετικά ζεύγη (H_1, H_2) , υιοθετήθηκε η διάταξη του τένσορα (tensor). Ο τένσορας αποτελεί στην πραγματικότητα έναν τρισδιάστατο πίνακα, ο οποίος προκύπτει από την παράλληλη τοποθέτηση ενός συνόλου δισδιάστατων πινάκων, όπως φαίνεται και στο Σχήμα 5.4. Ένας τρισδιάστατος τένσορας περιγράφεται από μία τριάδα αριθμών (w, d, l) . Παρατηρώντας το Σχήμα 5.4, εύκολα αντιλαμβάνεται κανείς ότι η πρώτη συντεταγμένη αντιστοιχεί στο πάχος του, δηλαδή στον αριθμό των στρώσεων (ή, καλύτερα, των παράλληλων πινάκων που τον αποτελούν). Η δεύτερη συντεταγμένη αντιστοιχεί στον αριθμό των γραμμών κάθε πίνακα, ενώ η τρίτη συντεταγμένη αφορά τον αριθμό των στηλών του.



Σχήμα 5.4: Αναπαράσταση τένσορα $(3,2,5)$

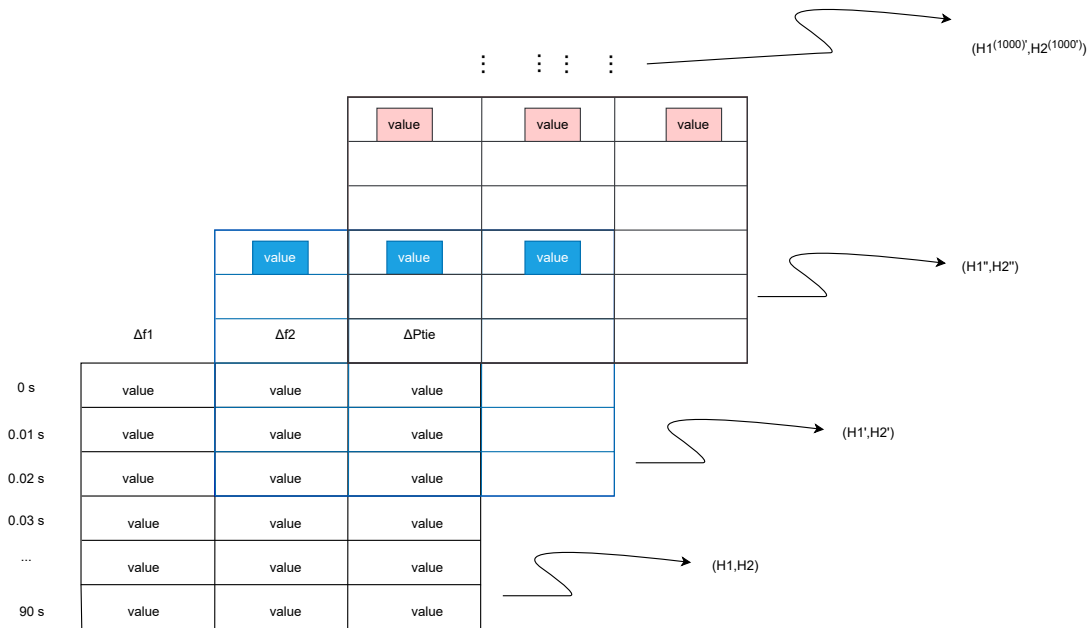
Στη δική μας περίπτωση, λοιπόν, ο αριθμός των στηλών κάθε πίνακα αντιστοιχεί στην τριάδα των χρονοσειρών εισόδου $(\Delta f_1, \Delta f_2, \Delta P_{tie})$, ενώ ο αριθμός των γραμμών κάθε πίνακα αντιστοιχεί στο πλήθος των δειγμάτων που προέκυψαν από τη δειγματοληψία. Τα περιεχόμενα του κάθε πίνακα είναι οι αριθμητικές τιμές των χρονοσειρών για κάθε χρονική στιγμή, ανάλογα με τον ρυθμό δειγματοληψίας, και συνδέονται αποκλειστικά με ένα συγκεκριμένο ζεύγος τιμών (H_1, H_2) . Κάθε πίνακας, συνεπώς, αποθηκεύει ένα σύνολο τριών χρονοσειρών που αντιστοιχούν σε έναν μοναδικό συνδυασμό (H_1, H_2) . Το σύνολο όλων των πινάκων είναι ίσο με το πλήθος των διαφορετικών συνδυασμών (H_1, H_2) , δηλαδή 1000, και αποθηκεύεται παράλληλα, όπως φαίνεται στο Σχήμα 5.5.

Μαθηματικά, το συνολικό σύνολο χρονοσειρών μπορεί να αναπαρασταθεί ως ένας τένσορας

$$X \in \mathbb{R}^{1000 \times N \times 3},$$

όπου $N = 9001$ είναι το πλήθος των δειγμάτων κάθε χρονοσειράς και ο τελευταίος άξονας αντιστοιχεί στις τρεις εισόδους $(\Delta f_1, \Delta f_2, \Delta P_{tie})$. Αντίστοιχα, οι μεταβλητές εξόδου, δηλαδή τα ζεύγη των αδρανειών, αναπαρίστανται ως ένας διδιάστατος πίνακας

$$Y \in \mathbb{R}^{1000 \times 2},$$



Σχήμα 5.5: Τρόπος αποθήκευσης σύνολου δεδομένων σε τένσορα

όπου κάθε γραμμή περιέχει το αντίστοιχο ζεύγος (H_1, H_2) για τις τρεις χρονοσειρές του ίδιου πίνακα στον X δηλαδή θα υπάρχει αντιστοίχιση της πρώτης από τις τρεις συντεταγμένες του τένσορα με την κάθε γραμμή του πίνακα Y .

Η ίδια διαδικασία επαναλαμβάνεται για όλα τα σενάρια των διαφορετικών μειγμάτων μονάδων παραγωγής, καθώς η μορφή των αποκρίσεων συχνότητας και ισχύος επηρεάζεται από το είδος των γραμμικών συναρτήσεων που χρησιμοποιούνται για την προσομοίωση της δυναμικής συμπεριφοράς των διαφόρων μονάδων παραγωγής ενέργειας. Στη συνέχεια του παρόντος κεφαλαίου θα χρησιμοποιούμε ως αναφορά το Σενάριο 1, καθώς οι στρατηγικές σχεδιασμού και διαχείρισης δεδομένων που εφαρμόζονται στα υπόλοιπα δύο σενάρια είναι παρόμοιες.

5.3 Προ-επεξεργασία δεδομένων (Preprocessing)

Αφού πλέον το τεχνητό σύνολο δεδομένων που κατασκευάστηκε, αποθηκεύτηκε με επιτυχία, σειρά έχει η προ-επεξεργασία του πριν αυτό τροφοδοτηθεί σε ένα Machine Learning μοντελο το οποίο θα μας παράγει προβλέψεις.

5.3.1 Εξαγωγή χαρακτηριστικών

Ο πρώτος μετασχηματισμός στον οποίο υποβάλλεται ο τένσορας μας είναι μια διαδικασία που ονομάζεται εξαγωγή χαρακτηριστικών γνωστότερη με την αγγλική της ορολογία ως *feature extraction*. Επειδή η πολυπλοκότητα και ο όγκος των χρονοσειρών δυσχεραίνουν την άμεση επεξεργασία τους, αποφασίστηκε να μελετηθεί κάθε χρονοσειρά ξεχωριστά και να κατασκευαστεί ένα νέο σύνολο δεδομένων, αποτελούμενο από δείκτες που περιγράφουν τη συμπεριφορά της. Από καθεμία χρονοσειρά εισόδου ($\Delta f_1, \Delta f_2, \Delta P_{tie}$) αντλούνται αρχικά οι εξής δείκτες:

- Μέση τιμή (mean):

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

- Τυπική απόκλιση (standard deviation):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

- Μέγιστη τιμή (max value):

$$x_{\max} = \max_{1 \leq i \leq N} (x_i)$$

- Ελάχιστη τιμή ή ναδίρ (min value / nadir):

$$x_{\min} = \min_{1 \leq i \leq N} (x_i)$$

- Ρίζα Μέσης Τετραγωνικής Τιμής (RMS):

$$\text{RMS}(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

- Peak-to-peak (διαφορά μέγιστης από ελάχιστη τιμή):

$$P2P = x_{\max} - x_{\min}$$

- Χρόνος ναδίρ (time of nadir):

Η χρονική στιγμή που παρατηρείται η μεγαλύτερη πτώση της συχνότητας.

$$t_{\min} = \arg \min_{t_i} (x(t_i))$$

Το νέο σύνολο δεδομένων αποθηκεύεται εκ νέου, αλλά αυτή τη φορά σε έναν διδιάστατο πίνακα. Ο πίνακας αυτός αποτελείται από 23 στήλες, εκ των οποίων οι 21 πρώτες αντιστοιχούν στα στατιστικά στοιχεία κάθε χρονοσειράς εισόδου και στην ουσία συνιστούν το σύνολο των ανεξάρτητων μεταβλητών εισόδου. Οι 2 τελευταίες στήλες περιέχουν το ζεύγος των αδρανειών (H_1, H_2) , οι οποίες αποτελούν τις μεταβλητές εξόδου. Σύμφωνα με αυτή τη λογική, ο πίνακας περιλαμβάνει συνολικά 1000 γραμμές, όπου κάθε γραμμή περιγράφει το σύνολο των τριών χρονοσειρών που αντιστοιχούν σε έναν μοναδικό συνδυασμό εξόδου (H_1, H_2) .

Μαθηματικά, το σύνολο δεδομένων μπορεί να αναπαρασταθεί ως εξής:

$$D \in \mathbb{R}^{1000 \times 23}, \quad D = [X \mid Y],$$

όπου

$$X \in \mathbb{R}^{1000 \times 21}$$

αντιπροσωπεύει τις ανεξάρτητες μεταβλητές (features) που προκύπτουν από τα στατιστικά των χρονοσειρών εισόδου, ενώ

$$Y \in \mathbb{R}^{1000 \times 2}$$

περιέχει τις μεταβλητές εξόδου, δηλαδή το ζεύγος των αδρανειών (H_1, H_2) .

5.3.2 Απόρριψη Ασταθών σημάτων

Αφού έχει κατασκευαστεί ο αρχικός διδιάστατος πίνακας, ενώ παράλληλα διατηρούμε πρόσβαση στον τένσορα που περιέχει τις γραφικές παραστάσεις όλων των σημάτων εισόδου $(\Delta f_1, \Delta f_2, \Delta P_{tie})$, το επόμενο βήμα είναι ο καθαρισμός του dataset από άκυρα δεδομένα. Λόγω της τυχαίας καταχώρησης αριθμητικών τιμών στις μεταβλητές H_1, H_2 κατά τη διάρκεια των προσομοιώσεων στο *Simulink*, όπου το εύρος τους κυμαινόταν στο διάστημα $[3, 8]$, ορισμένες από τις αποκρίσεις της διάταξης του Σχήματος 4.1 παρουσίασαν φαινόμενα αστάθειας δηλαδή παρατηρήθηκε εμφάνιση μη αποσβενούμενων ταλαντώσεων, συνεχώς αυξανόμενη απόκλιση της συχνότητας ή της ισχύος, και γενικότερα αδυναμία του συστήματος να επανέλθει σε μια νέα ισορροπία μετά από την διαταραχή [4]. Με άλλα λόγια, υπάρχουν συνδυασμοί τιμών αδρανειών (H_1, H_2) που καθιστούν το σύστημα ασταθές, και συνεπώς πρέπει να απορριφθούν τόσο οι συγκεκριμένες τιμές όσο και οι αποκρίσεις ισχύος και συχνότητας που τις συνοδεύουν, καθώς στο πλαίσιο της παρούσας διπλωματικής εξετάζονται αποκλειστικά σενάρια ευσταθούς λειτουργίας.

Το ερώτημα που τίθεται σε αυτό το σημείο είναι πώς μπορεί ο υπολογιστής να αποφανθεί για την ύπαρξη αστάθειας ή μη.

Ένας κλασικός τρόπος είναι η εφαρμογή του κριτηρίου Routh–Hurwitz, το οποίο είναι ικανό να προσδιορίσει την ύπαρξη πόλων της συνάρτησης μεταφοράς του κλειστού βρόχου στο δεξί ημιεπίπεδο. Ο γενικός κανόνας είναι ότι, εάν εντοπιστεί έστω και μία ρίζα στο πρώτο ή στο τέταρτο τεταρτημόριο, τότε το σύστημα θεωρείται ασταθές. Ωστόσο, η ύπαρξη πολλαπλών μονάδων παραγωγής καθιστά την κατασκευή της συνάρτησης μεταφοράς του κλειστού βρόχου μια ιδιαίτερα δύσκολη και χρονοβόρα διαδικασία.

Για τον λόγο αυτό αποφασίστηκε να εκμεταλλευτούμε την τάση των ασταθών σημάτων να παρουσιάζουν σημαντική αύξηση του πλάτους τους με την πάροδο του χρόνου. Η μέθοδος εντοπισμού της αστάθειας που υιοθετήθηκε βασίζεται στον υπολογισμό της RMS τιμής κάθε σήματος στο τελευταίο 10% του συνολικού χρόνου απόκρισης (δηλαδή από $t = 81$ s έως $t = 90$ s). Στη συνέχεια, υπολογίζεται η RMS τιμή του ίδιου σήματος στο πρώτο 30% του χρόνου απόκρισης (δηλαδή από $t = 0$ s έως $t = 27$ s). Οι δύο αυτές τιμές διαιρούνται, και το τελικό αποτέλεσμα ορίζεται ως δείκτης *RMS_ratio*:

$$\text{RMS_ratio} = \frac{\text{RMS}(x_{81-90s})}{\text{RMS}(x_{0-27s})}.$$

Κάθε ένα από τα τρία σήματα εισόδου στο μοντέλο μας διαθέτει τον δικό του δείκτη, ο οποίος περιγράφει τη σχέση μεταξύ του μεγέθους της διακύμανσης του πλάτους των αποκρίσεων στην αρχή και στο τέλος της περιόδου εξέτασης. Εάν ο λόγος αυτός υπερβαίνει την τιμή 2, αυτό σημαίνει ότι το σήμα στο τέλος έχει τουλάχιστον διπλάσια RMS τιμή σε σχέση με την αρχή, γεγονός που αποτελεί σαφή ένδειξη αστάθειας.

Επομένως, τρία επιπλέον features προστίθενται στον αρχικό δισδιάστατο πίνακα, αυξάνοντας τη διάστασή του σε 1000×26 .

Από τη στιγμή που κάθε πλειάδα του πίνακα διαθέτει πλέον τρεις επιπλέον δείκτες, οι οποίοι υποδεικνύουν εάν τα στατιστικά στοιχεία των χρονοσειρών που φιλοξενούνται αντιστοιχούν σε ασταθή ή όχι σήματα, προχωράμε στη διαγραφή γραμμών. Συγκεκριμένα, εάν έστω και ένα από αυτά τα features έχει τιμή μεγαλύτερη του 2, τότε η γραμμή που περιέχει αυτήν την τιμή διαγράφεται από τον πίνακα.

Αφού λοιπόν τα δεδομένα σε όλη την έκταση του πίνακα υποβλήθηκαν στον παραπάνω έλεγχο, εντοπίστηκαν οι ασταθείς περιπτώσεις που αντιστοιχούν σε μη ρεαλιστικά σενάρια λειτουργίας και εξαιρέθηκαν από την περαιτέρω ανάλυση.

5.3.3 Επιλογή χαρακτηριστικών (Feature selection)

Επόμενο βήμα της προεπεξεργασίας δεδομένων στο παρόν πρόβλημα είναι το feature selection. Όπως αναφέρθηκε στο κεφάλαιο της οργάνωσης των δεδομένων, η συγκεκριμένη διαδικασία αποσκοπεί στην απομάκρυνση χαρακτηριστικών των οποίων ο βαθμός επιρροής στο τελικό αποτέλεσμα είναι αμελητέος. Έχει αποδειχθεί ότι αυτός ο μετασχηματισμός βελτιώνει σημαντικά την ακρίβεια των προβλέψεων, καθώς πολλά χαρακτηριστικά είναι πρακτικά ασυσχέτιστα με την έξοδο και συχνά αποπροσανατολίζουν το μοντέλο από τη σωστή λύση. Επιπλέον, η μείωση του μεγέθους των δεδομένων εξασφαλίζει μικρότερο υπολογιστικό κόστος και ταχύτερη εκπαίδευση των μοντέλων.

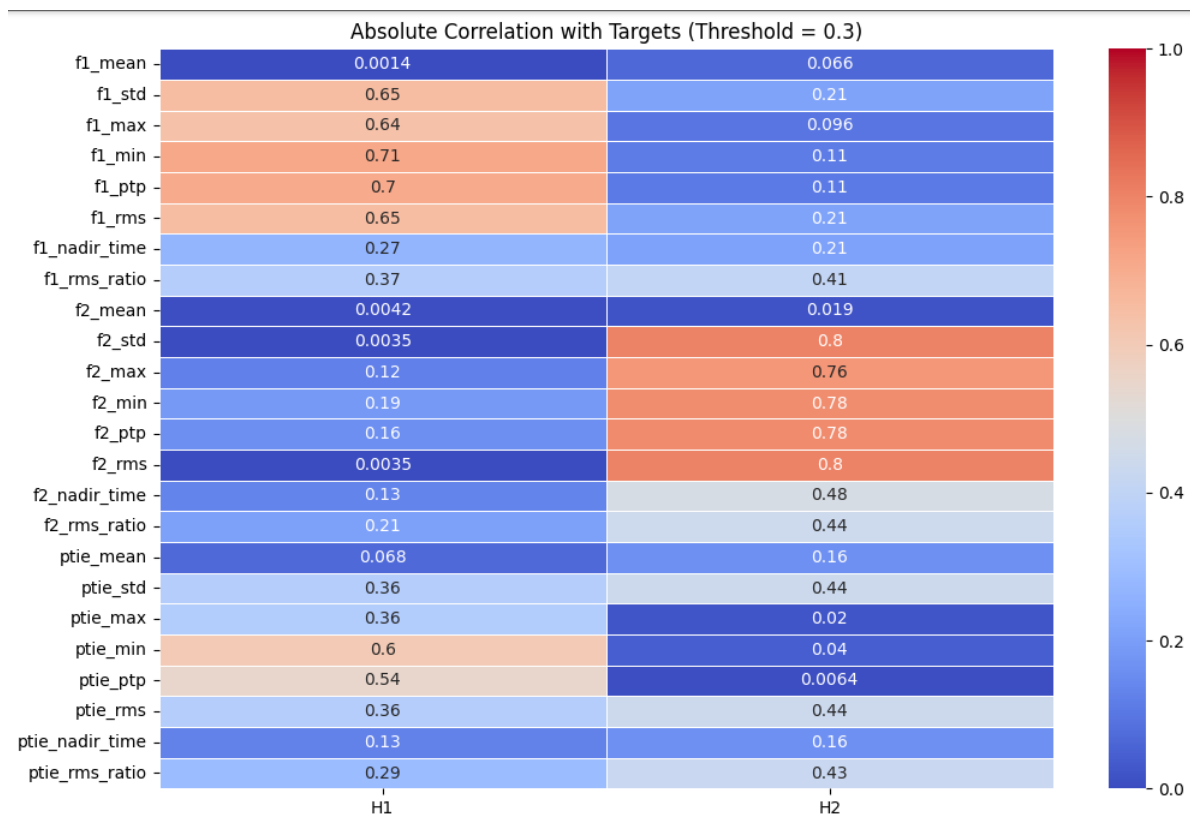
Στα πλαίσια της παρούσας διπλωματικής εργασίας, για την επιλογή των πλέον αντιπροσωπευτικών χαρακτηριστικών χρησιμοποιήθηκε ο πίνακας συσχέτισης (correlation matrix). Ο πίνακας αυτός ποσοτικοποιεί τον βαθμό γραμμικής συσχέτισης μεταξύ των χαρακτηριστικών και της εξόδου, επιτρέποντας την απομάκρυνση εκείνων που εμφανίζουν χαμηλή ή μηδενική συσχέτιση. Με τον τρόπο αυτό, διατηρούνται μόνο τα χαρακτηριστικά με ουσιαστική συμβολή στην πρόβλεψη, γεγονός που οδηγεί σε πιο αποδοτικά και ακριβή μοντέλα.

Ο πιο διαδεδομένος δείκτης που χρησιμοποιείται για τον υπολογισμό της συσχέτισης είναι ο συντελεστής Pearson, ο οποίος για δύο τυχαίες μεταβλητές X και Y ορίζεται ως:

$$r_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

όπου $\text{cov}(X, Y)$ είναι η συνδιακύμανση των δύο μεταβλητών, ενώ σ_X και σ_Y οι τυπικές αποκλίσεις τους. Οι τιμές του συντελεστή κυμαίνονται στο διάστημα $[-1, 1]$, με τιμές κοντά στο $+1$ να υποδεικνύουν ισχυρή θετική συσχέτιση, κοντά στο -1 ισχυρή αρνητική συσχέτιση, ενώ τιμές κοντά στο 0 δηλώνουν έλλειψη γραμμικής σχέσης. Ο πίνακας συσχέτισης ο οποίος δείχνει τον βαθμό επιρροής κάθε χαρακτηριστικού σε κάθε μια από τις δυο εξόδους του μοντέλου φαίνεται στο σχήμα 5.6.

Ως ελάχιστο όριο συντελεστή συσχέτισης ορίστηκε η τιμή 0.3 πράγμα που σημαίνει ότι οποιοδήποτε χαρακτηριστικό παρουσιάζει βαθμό γραμμικής εξάρτησης και με τις δύο εξόδους του συστήματος μικρότερο αυτής της τιμής, απορρίπτεται άμεσα από το σύνολο δεδομένων. Παρατηρώντας τις πληροφορίες που παρέχει ο πίνακας συσχέτισης, μπορεί κανείς να συμπεράνει ότι τα στατιστικά στοιχεία που περιγράφουν την απόκριση συχνότητας μιας συγκεκριμένης περιοχής έχουν μεγαλύτερο βαθμό επιρροής στη διαμόρφωση της ισοδύναμης αδράνειας αυτής της περιοχής, αλλά όχι της άλλης. Τα χρώματα χρησιμοποιούνται για



Σχήμα 5.6: Πινάκας Συσχέτισης (Correlation Matrix)

να απεικονίσουν την εγγύτητα του συντελεστή συσχέτισης προς τη μονάδα: όσο πιο κόκκινο εμφανίζεται ένα κελί, τόσο περισσότερο η τιμή προσεγγίζει το +1, ενώ όσο πιο μπλε, τόσο πιο κοντά βρίσκεται στο μηδέν.

Με βάση λοιπόν τον πίνακα συσχέτισης, απορρίφθηκαν συνολικά 5 χαρακτηριστικά εισόδου που δεν πληρούσαν το κριτήριο του ελάχιστου συντελεστή συσχέτισης. Τα συγκεκριμένα *features* είναι τα εξής: *f1_mean*, *f1_nadir_time*, *f2_mean*, *ptie_mean*, *ptie_nadir_time*. Κατά συνέπεια, ο τελικός πίνακας δεδομένων διαμορφώνεται πλέον σε διαστάσεις 722×21 .

5.3.4 Κλιμάκωση και διαχωρισμός

Τελευταίο στάδιο της προεπεξεργασίας του υπάρχοντος πίνακα δεδομένων είναι η κλιμάκωση (*scaling*). Ανάμεσα στις διάφορες μεθόδους κλιμάκωσης που συναντώνται στον χώρο της επιστήμης δεδομένων, στην παρούσα εργασία επιλέχθηκε η μέθοδος *Min-Max Scaler*. Η επιλογή αυτή δικαιολογείται από το γεγονός ότι οι αριθμητικές τιμές των περισσότερων χαρακτηριστικών εκφράζονται ήδη σε ανά μονάδα (p.u.), με αποτέλεσμα να εμφανίζουν διαφορετικά εύρη τιμών μεταξύ τους. Χωρίς κατάλληλη κλιμάκωση, τα χαρακτηριστικά με με-

γαλύτερα απόλυτα μεγέθη θα μπορούσαν να κυριαρχήσουν στη διαδικασία εκπαίδευσης, οδηγώντας το μοντέλο σε λανθασμένες εκτιμήσεις.

Με την εφαρμογή της μεθόδου *Min-Max*, όλες οι τιμές κανονικοποιούνται στο διάστημα $[0, 1]$, εξασφαλίζοντας ότι κάθε χαρακτηριστικό συμβάλλει ισότιμα στη διαδικασία εκπαίδευσης και βελτιώνοντας τόσο τη σταθερότητα όσο και τη σύγκλιση των αλγορίθμων μηχανικής μάθησης. Ο μαθηματικός ορισμός της κανονικοποίησης *Min-Max* για μια μεταβλητή x δίνεται από τον τύπο:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

όπου $x_{\min}$ και $x_{\max}$ είναι οι ελάχιστες και μέγιστες τιμές της μεταβλητής αντίστοιχα, ενώ x' η κανονικοποιημένη τιμή στο διάστημα $[0, 1]$.

Στη συνέχεια, ακολουθεί η κατασκευή δύο υποπινάκων από τον αρχικό πίνακα. Ο πρώτος θα περιλαμβάνει το 70% των στοιχείων και θα χρησιμοποιηθεί για την εκπαίδευση των μοντέλων (*training dataset*), ενώ ο δεύτερος θα περιλαμβάνει το υπόλοιπο 30% και θα αξιοποιηθεί για την αξιολόγηση κάθε εκπαιδευμένου μοντέλου με δεδομένα που του είναι άγνωστα (βλ. κεφάλαιο 3). Έπειτα, ακολουθεί η διχοτόμηση των δύο νέων υποπινάκων σε εξαρτημένες και ανεξάρτητες μεταβλητές. Ο πίνακας των ανεξάρτητων μεταβλητών διαθέτει 19 στήλες, ενώ ο πίνακας των εξαρτημένων μεταβλητών διαθέτει 2.

Στο σημείο αυτό, τα δεδομένα διοχετεύονται ως είσοδοι στα διάφορα μοντέλα μηχανικής μάθησης που αναλύθηκαν στο Κεφάλαιο 3. Τα αποτελέσματα κάθε μοντέλου, καθώς και η ερμηνεία τους, θα παρουσιαστούν στο επόμενο κεφάλαιο. Το παρόν κεφάλαιο αποτέλεσε σημείο αναφοράς για ολόκληρη τη διπλωματική εργασία, καθώς σε αυτό αναλύθηκε η διαδικασία κατασκευής, αποθήκευσης και προεπεξεργασίας των δεδομένων, τα οποία θα αποτελέσουν βασικά γρανάζια της επίτευξης του πολυπόθητου στόχου μας, δηλαδή τον προσδιορισμό της αδράνειας που θα παρουσιαστεί στην επόμενη ενότητα.

Κεφάλαιο 6

Αξιολόγηση των Αποτελεσμάτων

6.1 Εισαγωγή

Στο παρόν κεφάλαιο πραγματοποιείται πλήρης αξιολόγηση των προτεινόμενων μεθόδων προσδιορισμού της αδράνειας που περιγράφηκαν στα προηγούμενα κεφάλαια. Επειδή η απόδοση ενός αλγορίθμου μηχανικής μάθησης δεν μπορεί να είναι γνωστή εκ των προτέρων, στην πλειονότητα των εφαρμογών είναι κοινή πρακτική να δοκιμάζονται πολλαπλοί αλγόριθμοι για την εκτίμηση της ζητούμενης τιμής. Η καταλληλότητα κάθε μεθόδου εξαρτάται σε μεγάλο βαθμό από τη φύση και τη δομή των δεδομένων. Συνεπώς, δεν είναι δυνατό να προβλεφθεί ποιος αλγόριθμος θα αποδώσει καλύτερα ή θα κατανοήσει αποτελεσματικότερα τη συμπεριφορά του συστήματος

Για τον λόγο αυτό, στο παρόν κεφάλαιο παρουσιάζονται αναλυτικά οι μετρικές επίδοσης κάθε αλγορίθμου σε όλα τα σενάρια μονάδων παραγωγής. Παράλληλα, παρατίθενται διαδραστικά διαγράμματα που αποτυπώνουν το σφάλμα και οπτικοποιούν τα αποτελέσματα από διαφορετικές πτυχές, διευκολύνοντας την κατανόηση των διαφορών μεταξύ των μεθόδων. Με αυτόν τον τρόπο επιτυγχάνεται μια ολοκληρωμένη και συγκριτική αξιολόγηση της αξιοπιστίας και της αποδοτικότητάς τους. Τέλος έχοντας πλέον εκπαιδεύσει τα μοντέλα μας και έχοντας αποφανθεί σχετικά με την ικανότητα του κάθε αλγορίθμου στον προσδιορισμό του ζεύγους αδρανειών παρουσιάζουμε με συνοπτικό τρόπο πως θα μπορούσε η μέθοδος μας να λειτουργήσει σε ένα πραγματικό δίκτυο επιτρέποντας την σε πραγματικό χρόνο (online) παρακολούθηση της αδράνειας του συστήματος.

6.2 Ενδεικτικά αποτελέσματα - Σενάριο 1

Το ζεύγος των αδρανειών (H_1 , H_2) προσδιορίζεται με τη χρήση των πέντε αλγορίθμων μηχανικής μάθησης που αναλύθηκαν στο Κεφάλαιο 3, δηλαδή των MLR, Random Forest, XGBoost, SVR και MLP. Η αρχική φύση του προβλήματος το κατατάσσει στην κατηγορία των sequence-to-value προβλημάτων. Ωστόσο, μέσω της διαδικασίας προεπεξεργασίας που περιγράφηκε στο προηγούμενο κεφάλαιο, κατασκευάστηκε ένα νέο σύνολο δεδομένων προερχόμενο από το αρχικό, με αποτέλεσμα το πρόβλημα να μετατραπεί σε ένα κλασικό πρόβλημα παλινδρόμησης με πολλές μεταβλητές εισόδου και δύο μεταβλητές εξόδου (*multi-output regression*). Ο ταυτόχρονος προσδιορισμός των δύο μεταβλητών εξόδου δεν αποτελεί απλώς μια τεχνική επιλογή, αλλά αντανακλά και τη φυσική πραγματικότητα του συστήματος. Πράγματι, δεδομένου ότι οι σταθερές H_1 και H_2 αναφέρονται σε διασυνδεδεμένες περιοχές, επιδρούν ταυτόχρονα στη δυναμική συμπεριφορά του συστήματος. Η παράλληλη εκτίμησή τους μέσω multi-output παλινδρόμησης επιτρέπει στο μοντέλο να αξιοποιήσει αυτήν την αλληλεξάρτηση, οδηγώντας σε πιο συνεπή και αξιόπιστα αποτελέσματα. Σε αυτό το σημείο αξίζει να σημειωθεί ότι το σφάλμα διαφοροποιείται ανάλογα με τη σταθερά αδράνειας που εκτιμάται σε κάθε μοντέλο, δηλαδή σε ορισμένες περιπτώσεις ο προσδιορισμός της σταθεράς H_1 ενδέχεται να είναι πιο ακριβής, δηλαδή πιο κοντά στην πραγματική της τιμή, σε σύγκριση με τον αντίστοιχο προσδιορισμό της σταθεράς H_2 του ίδιου ζεύγους, και αντιστρόφως.

6.2.1 Αξιολόγηση αλγορίθμου Random Forest

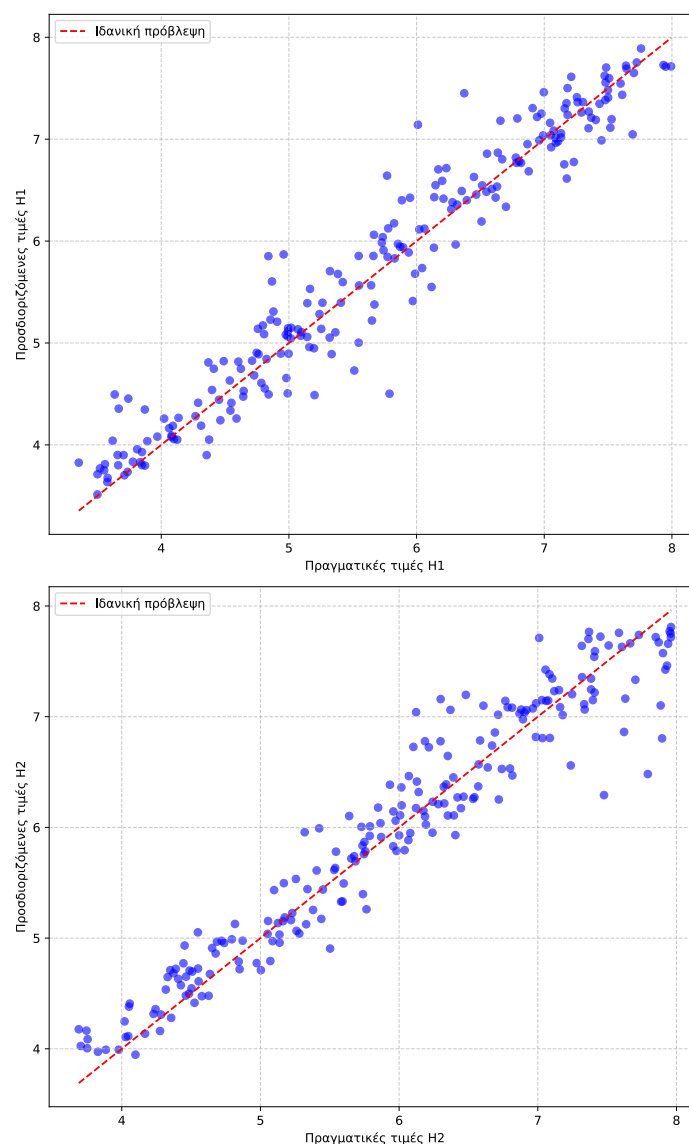
Τα αποτελέσματα αξιολόγησης για καθεμία από τις εξόδους, με τη χρήση του αλγορίθμου Random Forest, συνοψίζονται στον Πίνακα 6.1.

Παρατηρείται ότι ο συγκεκριμένος αλγόριθμος επιτυγχάνει πολύ υψηλές τιμές της μετρικής (R^2), με τα νούμερα να ξεπερνάνε το 0.92 και για τις δύο εξόδους, γεγονός που καταδεικνύει την ικανότητά του να προσαρμόζεται με μεγάλη ακρίβεια στα δεδομένα. Οι τιμές των σφαλμάτων RMSE και MAE παραμένουν σε χαμηλά επίπεδα και είναι συγκρίσιμες μεταξύ των δύο εξόδων, υποδηλώνοντας ότι η απόδοση του μοντέλου είναι ισορροπημένη. Επιπλέον, το MAPE κυμαίνεται γύρω στο 4%, κάτι που σημαίνει ότι η μέση σχετική απόκλιση από τις πραγματικές τιμές είναι περιορισμένη. Συνολικά, τα αποτελέσματα επιβεβαιώνουν ότι ο αλγόριθμος Random Forest προσφέρει μια ιδιαίτερα αξιόπιστη εκτίμηση των σταθερών αδράνειας, με ελαφρώς καλύτερη απόδοση στην εκτίμηση του H_2 σε σχέση με το H_1 .

Πίνακας 6.1: Αποτελέσματα αξιολόγησης του αλγορίθμου Random Forest για τις εξόδους H_1 και H_2 .

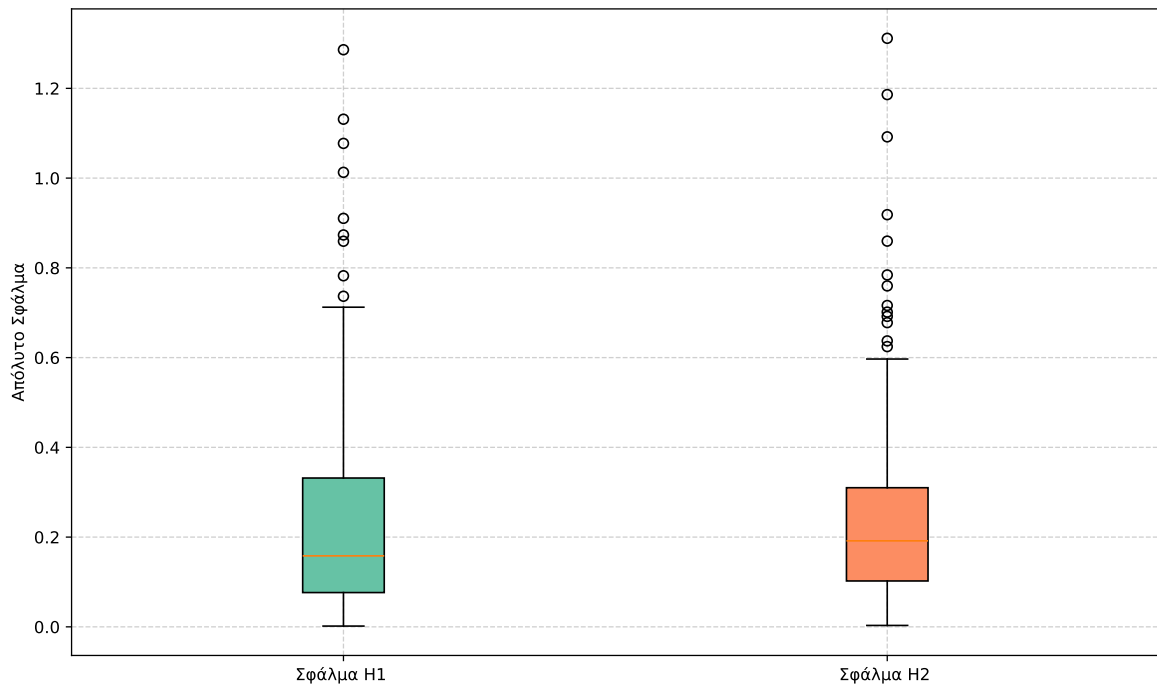
Έξοδος	R^2	RMSE	MAE	MAPE (%)
H_1	0.93537	0.32258	0.23332	4.34
H_2	0.92647	0.31709	0.23889	4.06

Η απόκλιση μεταξύ πραγματικών και προσδιοριζόμενων απο το μοντέλο τιμών για κάθε αδράνεια μπορεί να οπτικοποιηθεί καλύτερα στο σχήμα 6.1



Σχήμα 6.1: Απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών των σταθερών H_1 , H_2 από το μοντέλο Random Forest.

Σύμφωνα με το συγκεκριμένο Σχήμα, για καθεμία από τις αδράνειες H_1 και H_2 μπορούμε να αναπαριστήσουμε το σφάλμα τους κάνοντας χρήση ενός συστήματος δύο αξόνων. Στον οριζόντιο άξονα παριστάνεται η πραγματική τιμή της αδράνειας στο εύρος $[3, 8]$, ενώ στον κατακόρυφο η αντίστοιχη προβλεπόμενη τιμή που προκύπτει από το εκάστοτε μοντέλο. Κάθε μπλε κουκίδα που εμφανίζεται στο επίπεδο περιγράφεται από ένα ζεύγος συντεταγμένων, όπου η τετμημένη αντιστοιχεί στην πραγματική τιμή και η τεταγμένη στην προβλεπόμενη. Η βέλτιστη περίπτωση, δηλαδή ο μηδενισμός του σφάλματος, αντιστοιχεί σε πλήρη ταύτιση των πραγματικών και των προβλεπόμενων τιμών. Γεωμετρικά, αυτό σημαίνει ότι όλες οι κουκίδες θα βρίσκονταν επάνω στην ευθεία $y = x$. Συνεπώς, η απόσταση κάθε σημείου από την ευθεία αυτή αντανakλά άμεσα την ικανότητα του εκάστοτε αλγορίθμου να προσδιορίσει με ακρίβεια την αδράνεια.



Σχήμα 6.2: Θηκογράμματα του απόλυτου σφάλματος για τις προβλέψεις των H_1 και H_2 με Random Forest.

Ένας εναλλακτικός τρόπος οπτικοποίησης των σφαλμάτων είναι μέσω των θηκογραμμάτων (box plots), όπως φαίνεται στο Σχήμα 6.2. Σε αντίθεση με τα προηγούμενα διαγράμματα που εστιάζουν άμεσα στη σύγκριση πραγματικών και προβλεπόμενων τιμών, τα διαγράμματα κουτιών στοχεύουν στην απεικόνιση της κατανομής του σφάλματος.

Πιο συγκεκριμένα, στο Σχήμα 6.2 ο άξονας y αναπαριστά το απόλυτο σφάλμα (Absolute error), δηλαδή το μέτρο της διαφοράς μεταξύ της πραγματικής τιμής κάθε αδράνειας και

αυτής που προσδιόρισε το μοντέλο ($|H_{\text{actual}} - H_{\text{predicted}}|$), ενώ ο άξονας x περιλαμβάνει τις δύο αδράνειες H_1 και H_2 των οποίων τις κατανομές σφάλματος επιχειρούμε να δείξουμε. Για καθεμία από τις αδράνειες κατασκευάζεται ένα ανεξάρτητο θηκόγραμμα. Η διαδικασία ακολουθεί τα εξής βήματα: για κάθε προβλεπόμενη τιμή αδράνειας υπολογίζεται το απόλυτο σφάλμα και τα αποτελέσματα συγκεντρώνονται σε μία λίστα. Στη συνέχεια, η λίστα ταξινομείται σε αύξουσα σειρά και υπολογίζονται τα τεταρτημόρια. Το κάθε κουτί του διαγράμματος (box plot) εκτείνεται από το πρώτο τεταρτημόριο ($Q1$, το 25% των δεδομένων) έως το τρίτο τεταρτημόριο ($Q3$, το 75% των δεδομένων), περιλαμβάνοντας έτσι το μεσαίο 50% της κατανομής. Στο εσωτερικό του κουτιού υπάρχει μια μαύρη γραμμή που αντιστοιχεί στη διάμεσο (median) του σφάλματος, η οποία είναι ανθεκτική σε ακραίες τιμές και αντιπροσωπεύει το «τυπικό» σφάλμα. Από την άλλη πλευρά, τα μουστάκια (whiskers) εκτείνονται μέχρι ένα εύρος $[a, b]$, το οποίο καθορίζεται από τις σχέσεις:

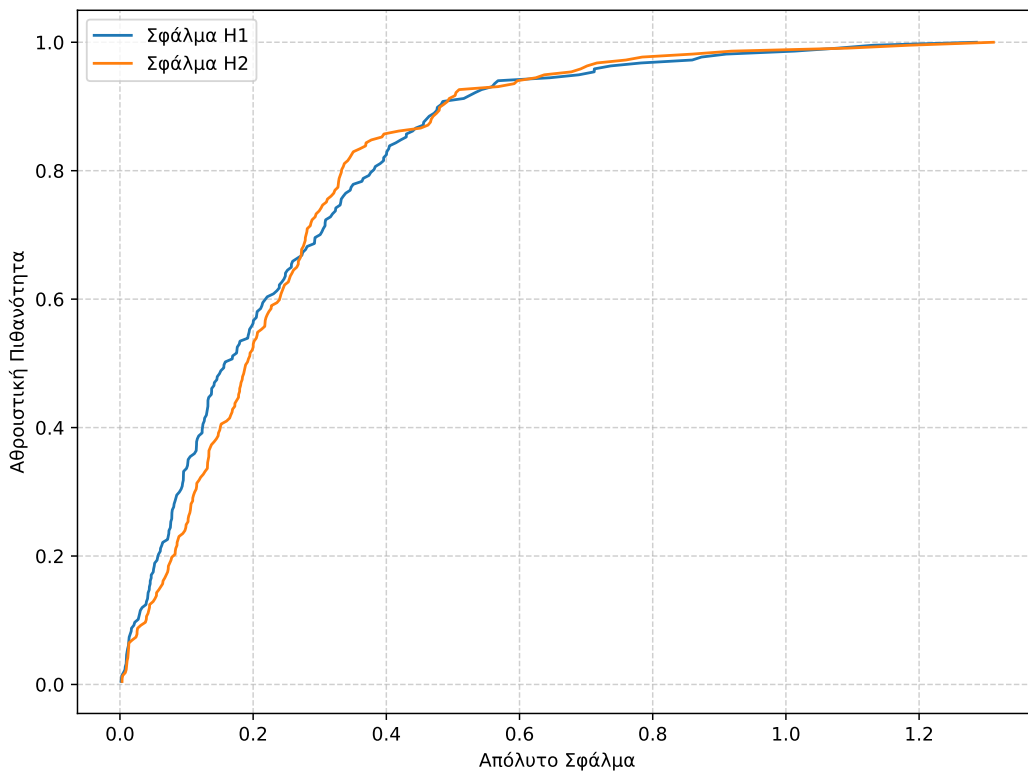
$$a = Q1 - 1.5 \cdot IQR, \quad b = Q3 + 1.5 \cdot IQR, \quad IQR = Q3 - Q1 \quad (6.1)$$

Όπου οι τιμές $Q1$ και $Q3$ αντιστοιχούν στο απόλυτο σφάλμα που βρίσκεται στο 25% και 75% της ταξινομημένης λίστας, αντίστοιχα. Οποιαδήποτε τιμή βρίσκεται εκτός του εύρους $[a, b]$ θεωρείται ακραία τιμή (outlier) και παριστάνεται στο διάγραμμα ως ξεχωριστή κουκκίδα.

Η ερμηνεία των *box plots* είναι ιδιαίτερα χρήσιμη για τη σύγκριση των δύο εξόδων. Ένα μικρότερο εύρος (πλάτος του κουτιού) υποδηλώνει ότι τα σφάλματα είναι πιο συγκεντρωμένα γύρω από τη διάμεσο, άρα το μοντέλο παρουσιάζει μεγαλύτερη σταθερότητα. Αντίθετα, ένα μεγαλύτερο εύρος σημαίνει ότι το μοντέλο εμφανίζει μεγαλύτερη διασπορά στα σφάλματα. Επιπλέον, η ύπαρξη πολλών ακραίων τιμών (outliers) μπορεί να υποδηλώνει ότι το μοντέλο σε ορισμένες περιπτώσεις αποτυγχάνει να προσαρμοστεί σωστά [30]. Έτσι, στην δική μας περίπτωση παρατηρούμε ότι το H_2 πάρα το γεγονός ότι εμφανίζει πολλά outliers εμφανίζει συγχρόνως μικρότερο εύρος σφαλμάτων από το H_1 , συνεπώς μπορούμε να συμπεράνουμε ότι το μοντέλο εκτιμά την αδράνεια της δεύτερης περιοχής με μεγαλύτερη συνέπεια.

Τελευταία μέθοδος οπτικοποίησης του σφάλματος που χρησιμοποιήθηκε στην πτυχιακή είναι η συνάρτηση αθροιστικής κατανομής πιο γνωστή με την αγγλική της ορολογία Cumulative Distribution Function (CDF) του απόλυτου σφάλματος. Η συνάρτηση αυτή λοιπόν δείχνει το ποσοστό των προβλέψεων που παρουσιάζουν σφάλμα μικρότερο ή ίσο από μια δεδομένη τιμή [31]. Όπως φαίνεται στο Σχήμα 6.3, οι καμπύλες για τις δύο εξόδους H_1 και H_2 είναι ιδιαίτερα παραπλήσιες, γεγονός που υποδηλώνει παρόμοια απόδοση του μοντέλου και στις δύο περιπτώσεις. Παρατηρείται ότι περίπου το 80% των προβλέψεων έχει

σφάλμα μικρότερο από 0.4, ενώ για χαμηλότερες τιμές σφάλματος (< 0.2) το H_1 εμφανίζει ελαφρώς μεγαλύτερη συγκέντρωση προβλέψεων με πολύ μικρό σφάλμα μιας και περιβάλλει την καμπύλη του σφάλματος προσδιορισμού της H_2 . Συνολικά, η CDF καταδεικνύει ότι το μοντέλο επιτυγχάνει ικανοποιητική ακρίβεια και συνεπή συμπεριφορά και στις δύο σταθερές αδράνειας.



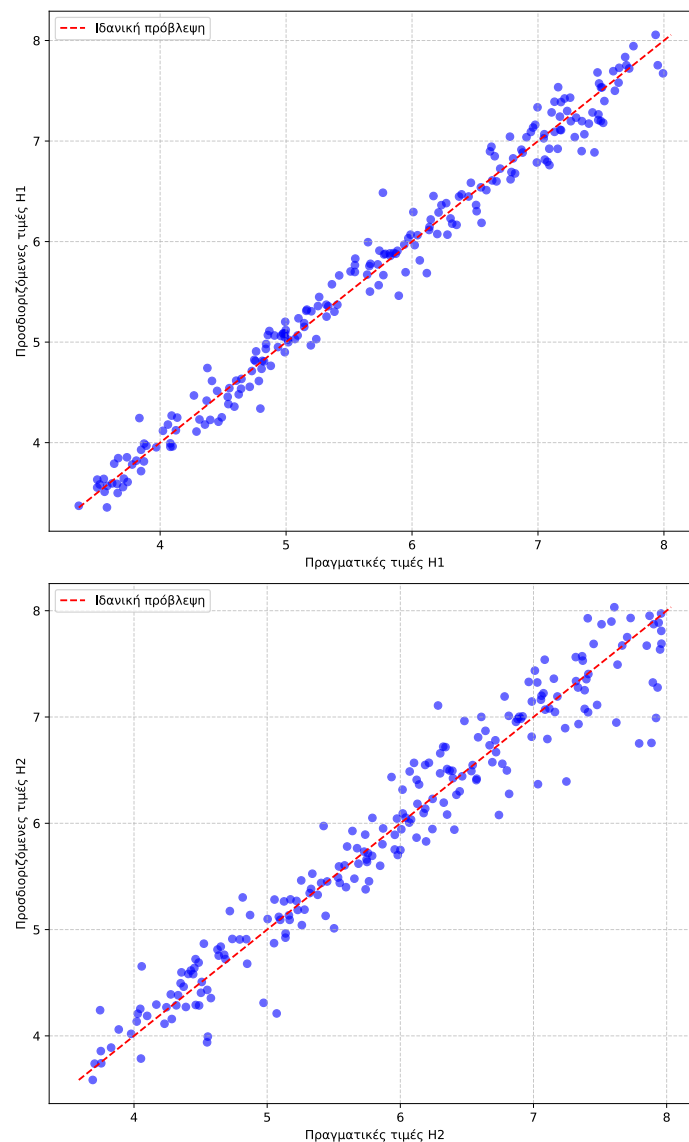
Σχήμα 6.3: Αθροιστική Συνάρτηση Κατανομής (CDF) του απόλυτου σφάλματος πρόβλεψης των H_1 , H_2 με Random Forest.

6.2.2 Αξιολόγηση Αλγορίθμου MLP

Σε αντίθεση με τα υπόλοιπα μοντέλα προσδιορισμού, το νευρωνικό δίκτυο πέτυχε την υψηλότερη ακρίβεια, υπερβαίνοντας σε απόδοση κάθε άλλο μοντέλο.

Πιο συγκεκριμένα λοιπόν συμφώνα με τον πίνακα 6.2 η ακρίβεια που πέτυχε ο συγκεκριμένος αλγόριθμος ήταν εξαιρετική ιδιαιτέρως στη εκτίμηση της σταθεράς H_1 , όπου το R^2 ξεπερνά το 0.98 και τα σφάλματα RMSE και MAE παραμένουν σε χαμηλά επίπεδα. Το MAPE στην περίπτωση του H_1 είναι μόλις 2.42%, γεγονός που καταδεικνύει εξαιρετική ακρίβεια. Όσον αφορά τον προσδιορισμό της αδράνειας H_2 η απόδοση συνεχίζει να παραμένει πολύ καλή με R^2 περίπου 0.94 και MAPE ίσο με 3.39%, ωστόσο είναι εμφανώς χαμηλό-

τερη σε σύγκριση με την περίπτωση της H_1 . Συνολικά, τα αποτελέσματα υποδεικνύουν ότι το νευρωνικό δίκτυο αποτελεί την πιο αποδοτική μέθοδο μεταξύ των εξεταζόμενων αλγορίθμων, παρέχοντας ιδιαίτερα αξιόπιστες προβλέψεις. Παρακάτω παρουσιάζονται τα διαγράμματα οπτικοποίησης που αφορούν το συγκεκριμένο μοντέλο με το σχήμα 6.4 να αφορά την απόκλιση μεταξύ των πραγματικών και των προσδιοριζόμενων τιμών για τις 2 αδράνειες, το σχήμα 6.5 να απεικονίζει τα θηκογράμματα του απολυτού σφάλματος των εξόδων και τέλος το σχήμα 6.6 να παρουσιάζει την συνάρτηση CDF των απολύτων σφαλμάτων των τιμών των αδρανειών H_1, H_2



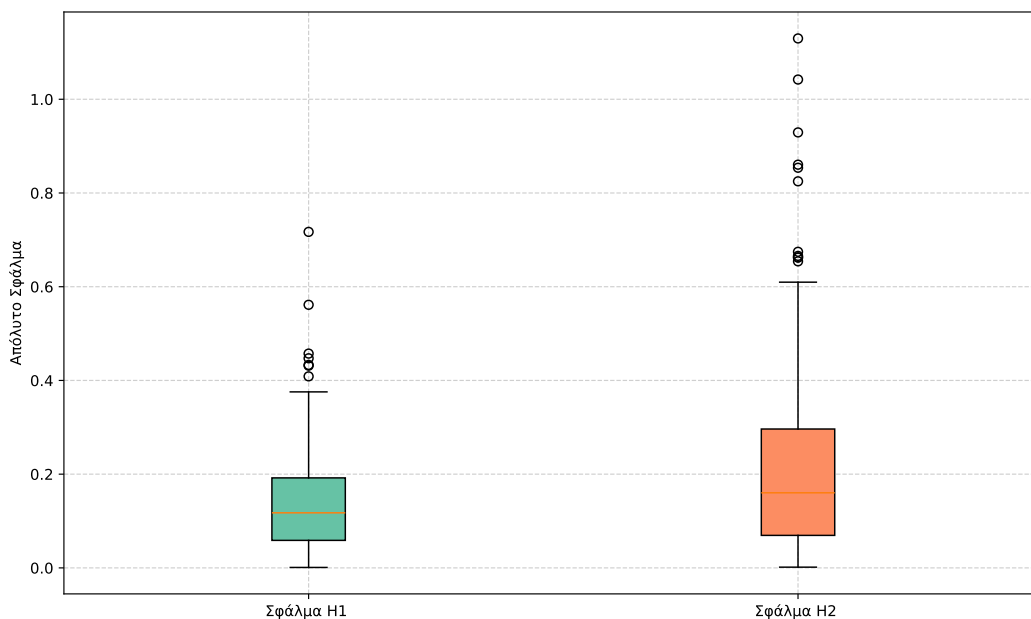
Σχήμα 6.4: Απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών των σταθερών H_1, H_2 από το μοντέλο MLP.

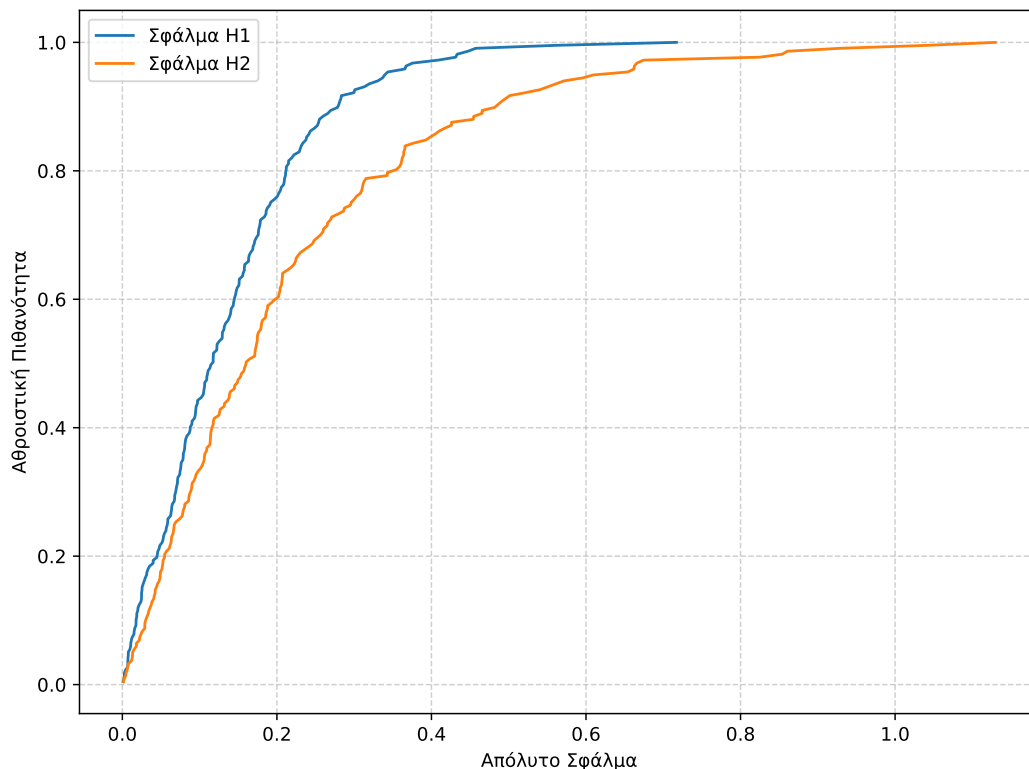
Πίνακας 6.2: Αποτελέσματα αξιολόγησης του MLP για τις εξόδους H_1 και H_2 .

Έξοδος	R^2	RMSE	MAE	MAPE (%)
H_1	0.98178	0.17126	0.13489	2.42
H_2	0.94111	0.28377	0.20352	3.39

Δίνοντας προσοχή στο σχήμα 6.4 παρατηρούμε ότι στην περίπτωση του συγκεκριμένου αλγορίθμου οι μπλε κουκίδες βρίσκονται σε σαφώς μικρότερη απόσταση από την ευθεία $y = x$ σε σύγκριση με τις ίδιες κουκίδες που προέκυψαν από το μοντέλο *Random Forest* γεγονός αυτό υποδεικνύει ισχυρότερη ικανότητα προσδιορισμού των ζητούμενων τιμών αδράνειας από το νευρωνικό δίκτυο.

Σύμφωνα με το *box plot* του Σχήματος 6.5, που αντιστοιχεί στο μοντέλο MLP, παρατηρείται ότι το εύρος των σφαλμάτων για την αδράνεια της πρώτης περιοχής είναι σαφώς μειωμένο σε σύγκριση με εκείνο που προέκυψε από το *Random Forest*. Το γεγονός αυτό υποδηλώνει μεγαλύτερη σταθερότητα του MLP στην εκτίμηση του H_1 . Αντίθετα, το εύρος του κουτιού που αντιστοιχεί στη δεύτερη περιοχή (H_2) δεν παρουσιάζει σημαντική διαφοροποίηση μεταξύ των δύο μοντέλων. Τέλος, αξίζει να σημειωθεί ότι ο αριθμός των ακραίων τιμών (*outliers*), τόσο για την πρώτη όσο και για τη δεύτερη αδράνεια, είναι αισθητά μειωμένος στην περίπτωση του MLP, γεγονός που καταδεικνύει μεγαλύτερη ανθεκτικότητα του μοντέλου σε ακραίες αποκλίσεις.

Σχήμα 6.5: Θηκόγραμμα του ζεύγους H_1, H_2 από το μοντέλο MLP.



Σχήμα 6.6: Συνάρτηση CDF του απόλυτου σφάλματος πρόβλεψης από το μοντέλο MLP για τις σταθερές H_1 και H_2 .

Η αθροιστική συνάρτηση κατανομής (CDF) του απόλυτου σφάλματος για το μοντέλο MLP (βλέπε σχήμα 6.6) δείχνει ότι η εκτίμηση της αδράνειας H_1 είναι πιο ακριβής σε σχέση με εκείνη του H_2 . Συγκεκριμένα, η καμπύλη του H_1 βρίσκεται σταθερά υψηλότερα, γεγονός που σημαίνει ότι μεγαλύτερο ποσοστό των σφαλμάτων συγκεντρώνεται σε χαμηλές τιμές. Ενδεικτικά, περίπου το 90% των σφαλμάτων για το H_1 είναι μικρότερα από 0.4, ενώ το αντίστοιχο ποσοστό για το H_2 είναι χαμηλότερο. Αυτό καταδεικνύει ότι το MLP παρουσιάζει αυξημένη ακρίβεια και σταθερότητα στην πρόβλεψη του H_1 , ενώ η εκτίμηση του H_2 παραμένει ικανοποιητική αλλά με μεγαλύτερη διασπορά σφαλμάτων.

6.2.3 Αξιολόγηση των υπόλοιπων αλγορίθμων μηχανικής μάθησης

Παρόμοια προσέγγιση χρησιμοποιήθηκε και για την οπτικοποίηση των αποτελεσμάτων των υπόλοιπων αλγορίθμων μηχανικής μάθησης, οι οποίοι χρησιμοποιήθηκαν για την εκτίμηση των τιμών των αδρανειών H_1 και H_2 . Η ακρίβεια των υπολογισμών για τα ζεύγη των εξόδων κυμάνθηκε σε παρόμοιο εύρος για όλους τους αλγορίθμους, με μικρές αποκλίσεις ανά περίπτωση.

Οι μετρικές αξιολόγησης κάθε αλγορίθμου παρουσιάζονται αναλυτικά για καθεμία από τις δύο εξόδους H_1 , H_2 αντίστοιχα στους δύο ακόλουθους πίνακες (6.3,6.4)

Πίνακας 6.3: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο 1).

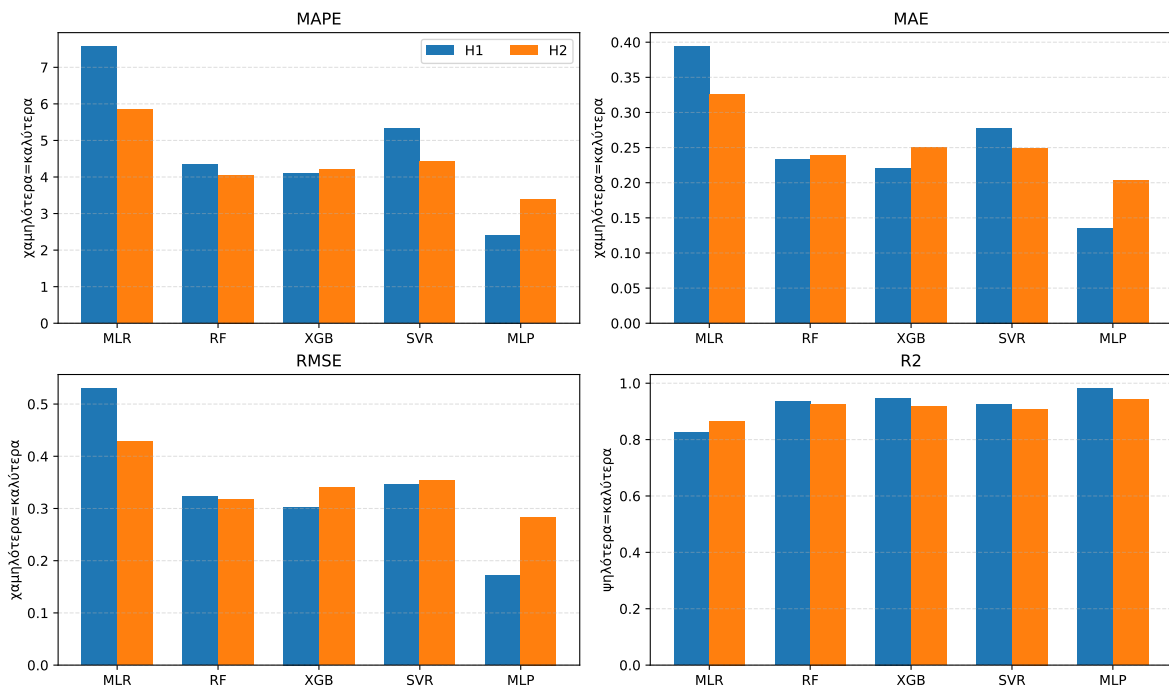
Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.93537	0.32258	0.23332	4.34
MLR	0.82549	0.53008	0.39391	7.57
XGBoost	0.94772	0.30286	0.22047	4.09
SVR	0.92548	0.34640	0.27811	5.33
MLP	0.98178	0.17126	0.13489	2.42

Πίνακας 6.4: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο 1).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.92647	0.31709	0.23889	4.06
MLR	0.86626	0.42765	0.32525	5.84
XGBoost	0.91648	0.34033	0.25008	4.21
SVR	0.90838	0.35395	0.24953	4.42
MLP	0.94111	0.28377	0.20352	3.39

Από τα αποτελέσματα παρατηρείται ότι ο αλγόριθμος MLP παρουσιάζει την καλύτερη επίδοση και για τις δύο εξόδους, με πολύ υψηλές τιμές R^2 και χαμηλά σφάλματα (RMSE, MAE, MAPE). Οι αλγόριθμοι Random Forest και XGBoost ακολουθούν με παρόμοια και σχετικά ικανοποιητικά αποτελέσματα, ενώ το SVR εμφανίζει ελαφρώς χαμηλότερη ακρίβεια. Τέλος, η Γραμμική Παλινδρόμηση καταγράφει τη χαμηλότερη απόδοση συγκριτικά με τα υπόλοιπα μοντέλα, γεγονός που επιβεβαιώνει την περιορισμένη ικανότητά της να αποτυπώσει την πολυπλοκότητα του προβλήματος.

Τα αποτελέσματα αξιολόγησης των αλγορίθμων μπορούν να ομαδοποιηθούν ανά μετρική και να παρουσιαστούν γραφικά στο Σχήμα 6.7. Σε κάθε υποδιάγραμμα εμφανίζεται μία μετρική αξιολόγησης, ενώ για κάθε αλγόριθμο αποτυπώνεται η επίδοσή του τόσο για την έξοδο H_1 όσο και για την έξοδο H_2 , διευκολύνοντας έτσι την άμεση συγκριτική αξιολόγηση.



Σχήμα 6.7: Σύγκριση επίδοσης των μοντέλων ανά μετρική για το σενάριο 1 .

6.3 Ενδεικτικά αποτελέσματα - Σενάριο 2

Παρόμοιες στρατηγικές αναπαράστασης εφαρμόστηκαν και για το Σενάριο 2. Παρατηρώντας όμως τις τιμές των μετρικών αξιολόγησης για τις δύο εξόδους, στους Πίνακες 6.5 και 6.6, διαπιστώνεται μια ασυνήθιστη συμπεριφορά στο σύνολο των αλγορίθμων. Ενώ οι επιδόσεις των μοντέλων για την έξοδο H_1 είναι εξαιρετικές και σαφώς βελτιωμένες σε σύγκριση με το Σενάριο 1, επιδεικνύοντας πολύ υψηλές τιμές R^2 (> 0.99) και εξαιρετικά χαμηλά σφάλματα (π.χ., $\text{MAPE} < 2\%$ για τα MLP, Random Forest, MLR), την ίδια στιγμή αδυνατούν να προσεγγίσουν με παρόμοια ακρίβεια τη σχέση μεταξύ των δεδομένων εισόδου και της αδρανείας της δεύτερης περιοχής (H_2). Η απόδοση για την H_2 καταγράφει, στην πλειοψηφία των περιπτώσεων, σφάλματα που κυμαίνονται από διπλάσιες έως και τετραπλάσιες τιμές συγκριτικά με αυτές της H_1 . Αυτή η ασυμμετρία στην ικανότητα πρόβλεψης μεταξύ των δύο περιοχών υποδηλώνει μια εγγενή πολυπλοκότητα στη δυναμική του συστήματος που σχετίζεται με τη δεύτερη περιοχή, η οποία δεν ανιχνεύεται εξίσου εύκολα από τα μοντέλα.

Παρά τις σημαντικές αυτές αποκλίσεις στην ακρίβεια πρόβλεψης μεταξύ των δύο εξόδων, τα μοντέλα MLP και Random Forest εξακολουθούν να παράγουν, για άλλη μια φορά, τα πιο αξιόπιστα και ισορροπημένα αποτελέσματα και για τις δύο περιοχές, διατηρώντας υψηλό R^2 και τα χαμηλότερα σφάλματα στο σύνολο των αλγορίθμων. Στο σχήμα 6.8 γίνεται ακόμα

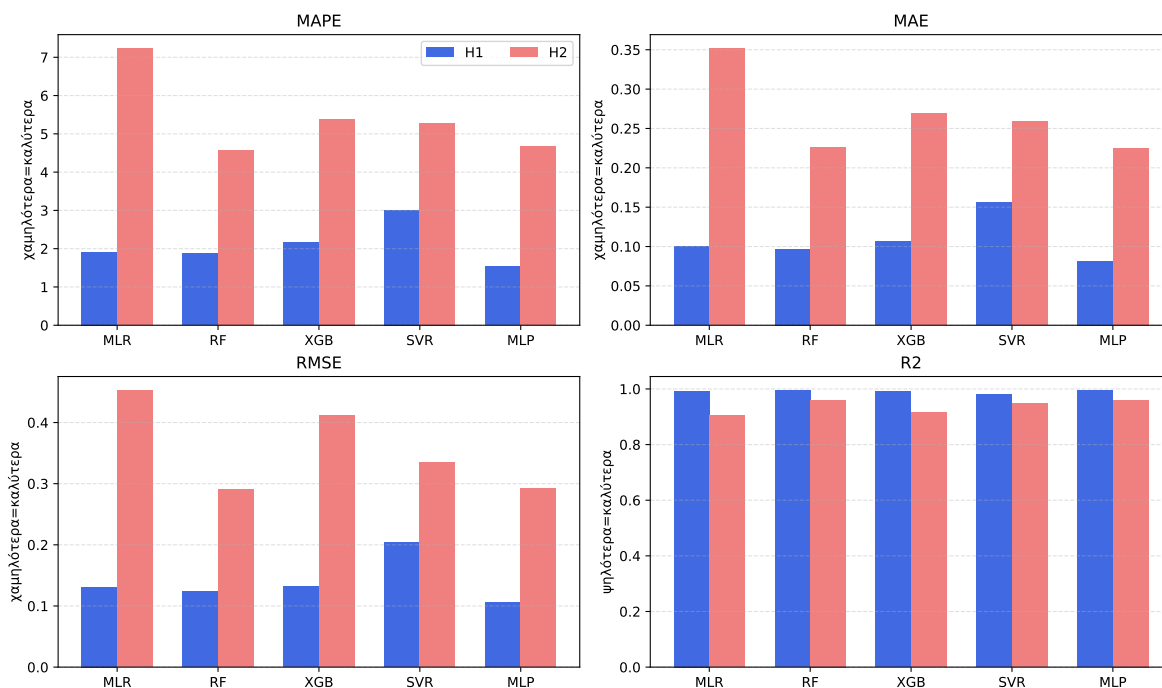
πιο εμφανής η ανισορροπία στην ακρίβεια προσδιορισμού των μοντέλων μεταξύ των εξόδων H_1 , H_2 ανά μετρική.

Πίνακας 6.5: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο 2).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.99311	0.1237	0.09672	1.88
MLR	0.99228	0.13094	0.09973	1.91
XGBoost	0.99182	0.13234	0.10656	2.16
SVR	0.98124	0.20411	0.15570	3.01
MLP	0.99495	0.10593	0.08187	1.53

Πίνακας 6.6: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο 2).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.96067	0.29017	0.22586	4.57
MLR	0.90431	0.45263	0.35163	7.23
XGBoost	0.91417	0.41129	0.26915	5.39
SVR	0.94757	0.33506	0.25895	5.28
MLP	0.96011	0.29223	0.22541	4.67



Σχήμα 6.8: Σύγκριση επίδοσης των μοντέλων ανά μετρική για το σενάριο 2 .

6.4 Ενδεικτικά αποτελέσματα - Σενάριο 3

Τα αποτελέσματα του Σεναρίου 3, όπως φαίνονται στους Πίνακες 6.7 και 6.8, αποκαλύπτουν ότι τα σφάλματα των μοντέλων είναι εμφανώς μικρότερα συγκριτικά με τα προηγούμενα σενάρια ενώ παράλληλα παρατηρείται μια ισορροπία στην απόδοση μεταξύ των δύο εξόδων H_1 και H_2 με τις τιμές των μετρικών για τις δύο περιοχές πλησιάζουν πολύ μεταξύ τους, γεγονός που υποδηλώνει ότι τα μοντέλα πλέον κατανοούν και προβλέπουν με παρόμοια ακρίβεια τη δυναμική και των δύο σταθερών αδράνειας.

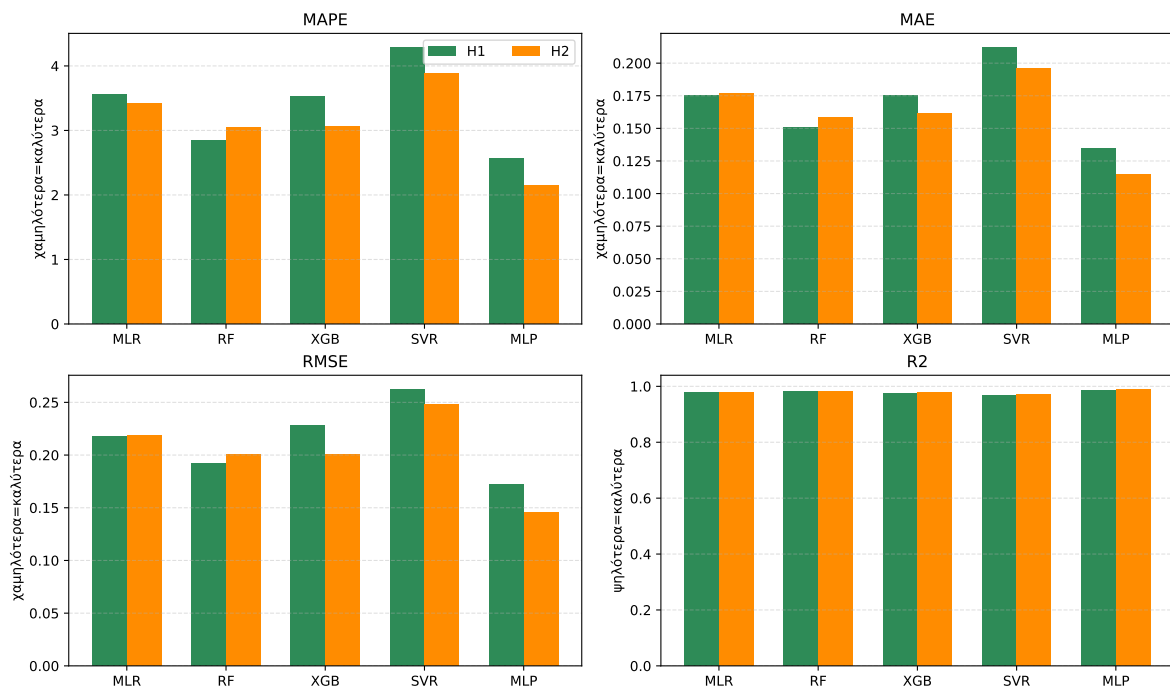
Η συμπεριφορά αυτή συνδέεται άμεσα με το γεγονός της ύπαρξης ισοροποποιημένου μείγματος παραγωγής στις 2 περιοχές αποκαλύπτοντας με αυτόν τον τρόπο ότι το μείγμα των μονάδων παράγωγης διαδραματίζει επίσης σημαντικό ρόλο στην ακρίβεια προσδιορισμού των αδρανειών. Ο MLP διατηρεί ξανά τα σκήπτρα της πρωτιάς ως ο πιο ισχυρός αλγόριθμος, καταγράφοντας τις βέλτιστες τιμές στις μετρικές αξιολόγησης και για τις 2 εξόδους με το MAPE της αδράνειας H_2 να φτάνει ακόμη και στο 2.14%. Η οπτικοποίηση των τιμών των μετρικών για το συγκεκριμένο σενάριο φαίνεται στο Σχήμα 6.8

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.98337	0.19216	0.15073	2.85
MLR	0.97857	0.21816	0.17489	3.56
XGBoost	0.97561	0.22851	0.17525	3.52
SVR	0.96895	0.26258	0.21222	4.29
MLP	0.98668	0.17198	0.13422	2.56

Πίνακας 6.7: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο 3).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.98119	0.20070	0.15827	3.05
Multiple Linear Regression	0.97770	0.21850	0.17698	3.42
XGBoost	0.97957	0.20069	0.16164	3.06
SVR	0.97115	0.24855	0.19567	3.89
MLP	0.99014	0.14531	0.11451	2.14

Πίνακας 6.8: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο 3).



Σχήμα 6.9: Σύγκριση επίδοσης των μοντέλων ανά μετρική για το σενάριο 3.

6.5 Online Προσδιορισμός Αδρανειών

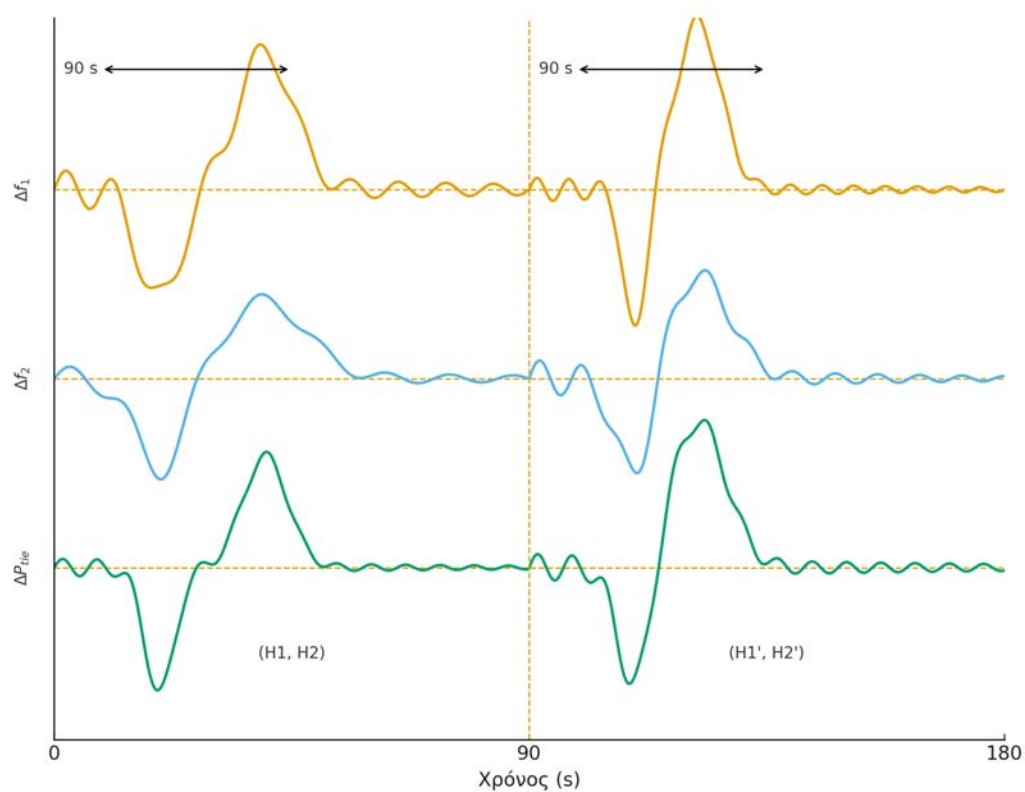
Η ανάπτυξη και η εκπαίδευση του συνόλου των αλγορίθμων που παρουσιάστηκαν στη διπλωματική εργασία επιτρέπει την εκτίμηση της αδράνειας ΣΗΕ σε πραγματικό χρόνο, χρησιμοποιώντας δεδομένα από δυναμικές αποκρίσεις του συστήματος. Συγκεκριμένα λοιπόν, μετά την αξιολόγηση των αλγορίθμων και καταλήγοντας στο συμπέρασμα ότι ο πιο αξιόπιστος αλγόριθμος μηχανικής μάθησης είναι ο MLP, προτείνεται η ανάπτυξη ενός συστήματος εκτίμησης αδράνειας σε πραγματικό χρόνο.

Το προτεινόμενο σύστημα θα βασίζεται σε ένα δίκτυο μετρητών που θα καταγράφουν τις απαιτούμενες δυναμικές αποκρίσεις (Δf_1 , Δf_2 , ΔP_{tie}) κάθε 90 δευτερόλεπτα. Οι μετρητές αυτοί θα επικοινωνούν μέσω ασφαλούς πρωτοκόλλου με το Κέντρο Ελέγχου (ΚΕ).

Στο ΚΕ, ένας κεντρικός server θα φιλοξενεί το εκπαιδευμένο μοντέλο MLP. Ο server θα λαμβάνει τα δεδομένα από ολόκληρο το δίκτυο των μετρητών, θα τα προεπεξεργάζεται και θα τα τροφοδοτεί στο μοντέλο μηχανικής μάθησης, υπολογίζοντας τις αντίστοιχες τιμές αδράνειας για κάθε διάστημα 90 s.

Πρέπει να τονιστεί, ότι αυτή η διαδικασία εκτελείται συνεχώς. Κάθε 90 s, λοιπόν, ο server στο ΚΕ λαμβάνει νέα δεδομένα, ανανεώνοντας την εκτίμηση των αδρανειών σε συνεχή βάση. Αυτό επιτρέπει την ζωντανή παρακολούθηση του ΣΗΕ και την άμεση προσαρμογή των εκτιμήσεων, βασιζόμενες στις τελευταίες διαθέσιμες τιμές. Η μέθοδος αυτή συνιστά έναν online προσδιορισμό αδρανειών, καθώς οι εκτιμήσεις ενημερώνονται αδιάκοπα και παρέχουν μία ακριβή απεικόνιση της κατάστασης του συστήματος σε κάθε δεδομένο χρονικό σημείο, αναλόγως με την εξέλιξη των συνθηκών στο σύστημα παραγωγής και τη συχνότητα του ΣΗΕ (δες σχήμα 6.10).

Με αυτόν τον τρόπο, καθίσταται εφικτή η συνεχής και ακριβής παρακολούθηση των αδρανειών του συστήματος σε πραγματικό χρόνο, γεγονός που αποτελεί ένα κρίσιμο εργαλείο για τη διαχείριση και τον έλεγχο των σύγχρονων ΣΗΕ. Ωστόσο, πρέπει να τονιστεί ότι για να λειτουργήσει το μοντέλο που εκπαιδεύσαμε και να προσδιορίσει με ακρίβεια το ζεύγος των αδρανειών εξόδου, απαιτείται η επαναλαμβανόμενη εφαρμογή ενός ορθογωνίου παλμού σε περιοδικά χρονικά διαστήματα. Ο παλμός αυτός θα τροποποιεί τη ζήτηση ισχύος στην περιοχή 1, όπως αναλύθηκε στο Κεφάλαιο 5 και θα εφαρμόζεται πάντα από το 20 έως το 40 δευτερόλεπτο κάθε περιόδου καταγραφής.



Σχήμα 6.10: Online προσδιορισμός ζεύγους αδρανειών

Κεφάλαιο 7

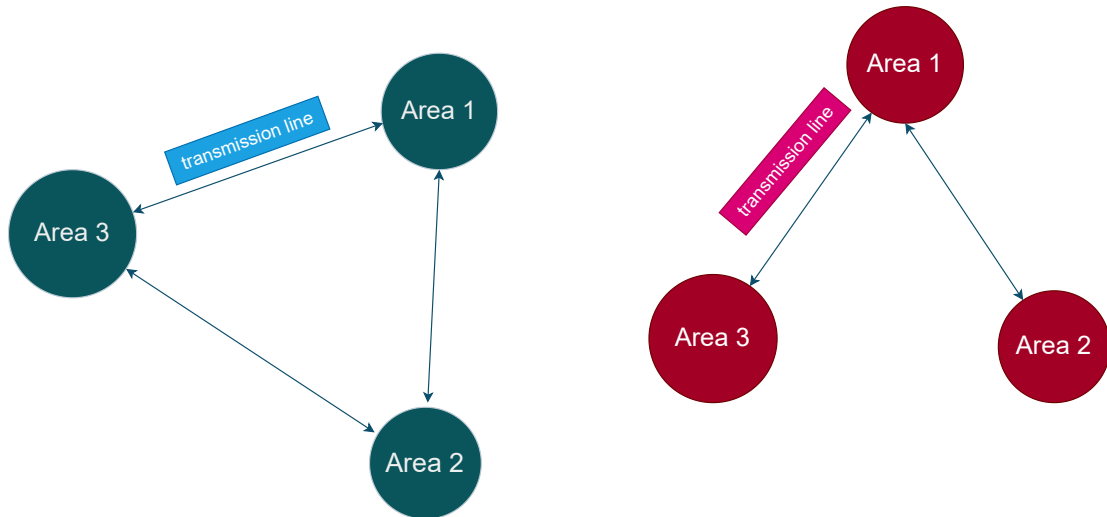
Επέκταση των μοντέλων σε διασυνδεδεμένο σύστημα 3 περιοχών

7.1 Εισαγωγή

Οι βασικές αρχές και στρατηγικές που εφαρμόστηκαν στα Κεφάλαια 4, 5 και 6 για το διασυνδεδεμένο σύστημα δύο περιοχών, επεκτείνονται στο παρόν κεφάλαιο σε ένα πιο εκτεταμένο σενάριο, το οποίο αυτή τη φορά αποτελείται από τρεις περιοχές διασυνδεδεμένες μεταξύ τους. Όπως είναι αναμενόμενο, ο όγκος δεδομένων, οι απαιτήσεις υπολογιστικής ισχύος και η πολυπλοκότητα του προβλήματος αυξάνονται σημαντικά. Παρ' όλα αυτά, η ανάλυση ενός συστήματος τριών περιοχών παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς επιτρέπει την εις βάθος κατανόηση της δυναμικής αλληλεπίδρασης μεταξύ των περιοχών. Μάλιστα, η συγκεκριμένη διάταξη προσεγγίζει με μεγαλύτερη ακρίβεια τις πραγματικές συνθήκες που συναντώνται στα σύγχρονα διασυνδεδεμένα ΣΗΕ, όπου η πολυπλοκότητα και η αλληλεξάρτηση αποτελούν βασικά χαρακτηριστικά (π.χ. Δίκτυο Βόρειας Αμερικής με 4 περιοχές [2]).

7.2 Υλοποίηση στο simulink

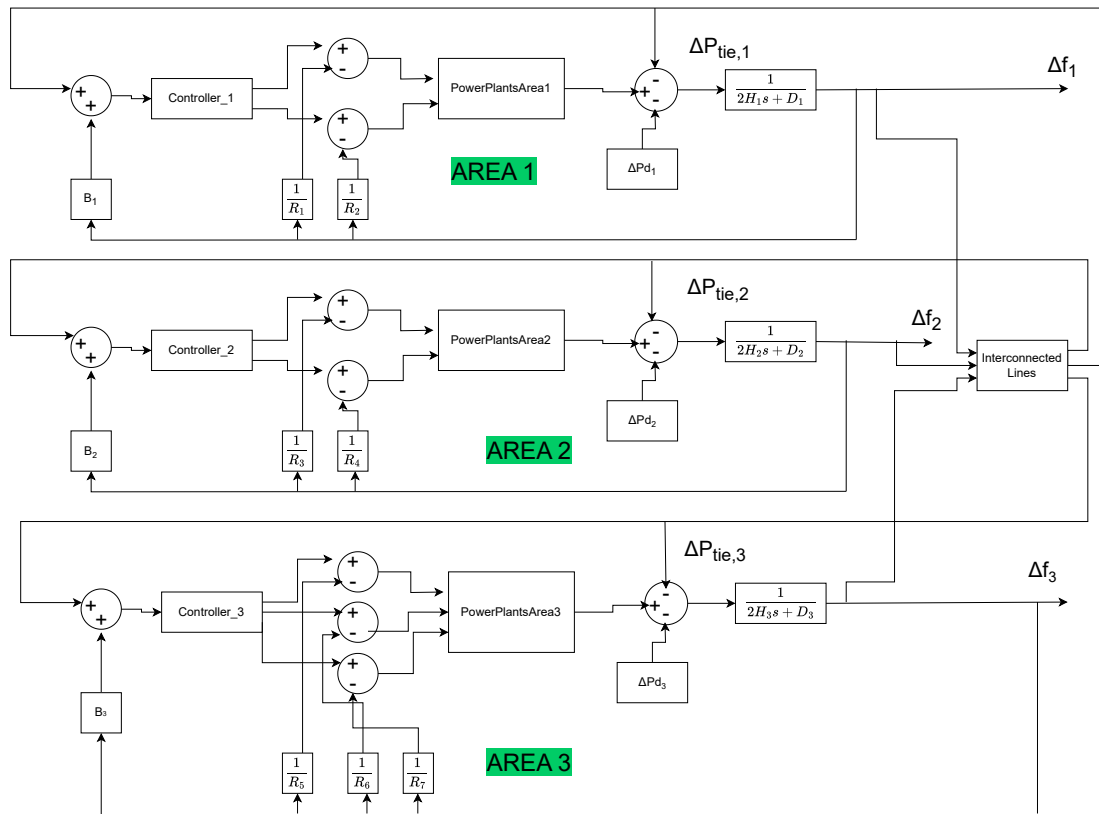
Υπάρχουν δύο δυνατοί τρόποι διασύνδεσης τριών περιοχών, οι οποίοι απεικονίζονται στο Σχήμα 7.1. Στην μια εκδοχή, κάθε περιοχή είναι απευθείας συνδεδεμένη με τις υπόλοιπες, ενώ στην δεύτερη εκδοχή η Περιοχή 1 λειτουργεί ως συνδετικός κρίκος μεταξύ των Περιοχών 2 και 3. Στο πλαίσιο της παρούσας εργασίας εξετάστηκε η πιο σύνθετη περίπτωση, δηλαδή εκείνη της πλήρους διασύνδεσης όλων των περιοχών.



Σχήμα 7.1: Σενάρια διασύνδεσης τριών περιοχών.

7.2.1 Βασικό δομικό διάγραμμα τριών διασυνδεδεμένων περιοχών.

Η μοντελοποίηση των τριών διασυνδεδεμένων περιοχών στο *Simulink* αποτελεί φυσική επέκταση του Σεναρίου 1. Σε αυτή την περίπτωση, αντί για ένα σύνολο από ζεύγη υποσυστημάτων, όπως αυτά που αναλύθηκαν στο Κεφάλαιο 4 και αντιστοιχούσαν το καθένα στην δικιά του περιοχή, έχουμε πλέον ένα σύνολο από τριάδες υποσυστημάτων. Οι μονάδες παραγωγής στις περιοχές 1 και 2 παραμένουν ίδιες με εκείνες του Σεναρίου 1, ενώ στην περιοχή 3 υλοποιούνται τρεις μονάδες παραγωγής. Από αυτές, οι δύο αποτελούν νέους τύπους, που προσομοιώνουν τη δυναμική συμπεριφορά ενός αεριοστρόβιλου και μιας μονάδας βιομάζας, με στόχο την αύξηση της ποικιλομορφίας του μείγματος παραγωγής έτσι ώστε να προσεγγίζει ρεαλιστικότερα ένα σύγχρονο ΣΗΕ, ενώ η τρίτη είναι ένας ατμοστρόβιλος μονής αναθέρμανσης. Η βασική διάταξη του συστήματος απεικονίζεται στο Σχήμα 7.2 [32].



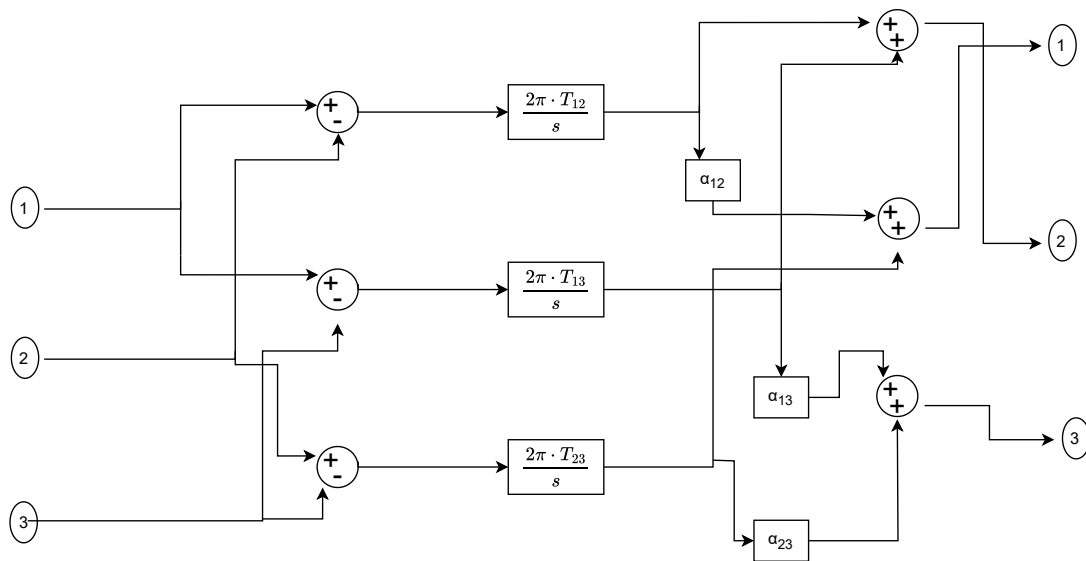
Σχήμα 7.2: Βασικό block διάγραμμα τριών διασυνδεδεμένων περιοχών

Μια επίσης σημαντική διαφορά με το σενάριο 1 στην κατασκευή των μοντέλων στο Simulink και, κατ' επέκταση, στην εκπαίδευση των αλγορίθμων μηχανικής μάθησης είναι η χρήση μπαταριών για τη σύνθεση του probing signal τόσο στην Περιοχή 1 όσο και στην Περιοχή 3. Αυτή η διαφοροποίηση κρίθηκε απαραίτητη, καθώς ένας τεχνητός παλμός στη ζήτηση ισχύος της Περιοχής 1 δεν στάθηκε ικανός να περιορίσει την επιρροή της στοχαστικότητας της πραγματικής μεταβολής της ζήτησης ισχύος. Ως αποτέλεσμα, τα μοντέλα μηχανικής μάθησης αδυνατούσαν να κατανοήσουν τη σχέση μεταξύ εισόδου και εξόδου. Συνεπώς, δοκιμαστικά σήματα επιβάλλονται, με χρήση μπαταριών, τόσο στην Περιοχή 1 όσο και στην Περιοχή 3.

7.2.2 Υποσύστημα intrerconnected lines

Πέρα από τις διαφορές με το Σενάριο 1, όσον αφορά την υιοθέτηση διάταξης μπαταρίας σε δύο διαφορετικές περιοχές και, προφανώς, την προσθήκη μιας επιπλέον περιοχής (Περιοχή 3), παρατηρούμε ότι, εξετάζοντας προσεκτικά τη διάταξη του Σχήματος 7.2, εμφανίζεται ένα νέο υποσύστημα με την ονομασία **Interconnected Lines**. Η συγκεκριμένη διάταξη

περιλαμβάνει όλες τις συνδεσμολογίες των γραμμών μεταφοράς μεταξύ των τριών περιοχών, συνδέοντας στο τέλος την κάθε περιοχή με την αντίστοιχη διασυνδεδετική της γραμμή ($\Delta P_{tie,1}$, $\Delta P_{tie,2}$, $\Delta P_{tie,3}$). Στο σχήμα 7.3 παρατηρούμε το σύνολο των συνδεσμολογιών που φιλοξενεί η συγκεκριμένη διάταξη στο εσωτερικό της. Η χρησιμότητα του ακόλουθου υποσυστήματος έγκειται αποκλειστικά στην απλότητα και την κατανόηση του τελικού σχήματος επιτρέποντας μια πιο κοσμία, ευκρινή και οργανωμένη αναπαράσταση των τριών διασυνδεδεμένων περιοχών.



Σχήμα 7.3: Υλοποίηση υποσυστήματος interconnected lines

7.2.3 Υποσύστημα power plants 3

Το υποσύστημα Power Plants της Περιοχής 3 φαίνεται στο Σχήμα 7.4. Παρατηρώντας το, μπορεί κανείς να διακρίνει την παράλληλη σύνδεση τριών γραμμικών συναρτήσεων μεταφοράς, οι οποίες προσεγγίζουν τη δυναμική λειτουργία των τριών ειδών μονάδων παραγωγής που συγκροτούν τη συγκεκριμένη περιοχή.

Η πρώτη μονάδα αντιστοιχεί στη μοντελοποίηση μιας μονάδας βιομάζας της οποίας η υλοποίηση είναι πανομοιότυπη με εκείνη ενός ατμοστροβίλου μονής αναθέρμανσης. Η διαφορά της ως προς τη δυναμική συμπεριφορά εντοπίζεται αποκλειστικά στα διαφορετικά εύρη τιμών που λαμβάνουν οι παράμετροι που την χαρακτηρίζουν.

Η δεύτερη μονάδα παραγωγής, που επίσης φαίνεται στο Σχήμα 7.4, προσομοιώνει τη δυναμική συμπεριφορά ενός αεριοστροβίλου (Gas Turbine). Ο αεριοστροβίλος χρησιμοποιεί

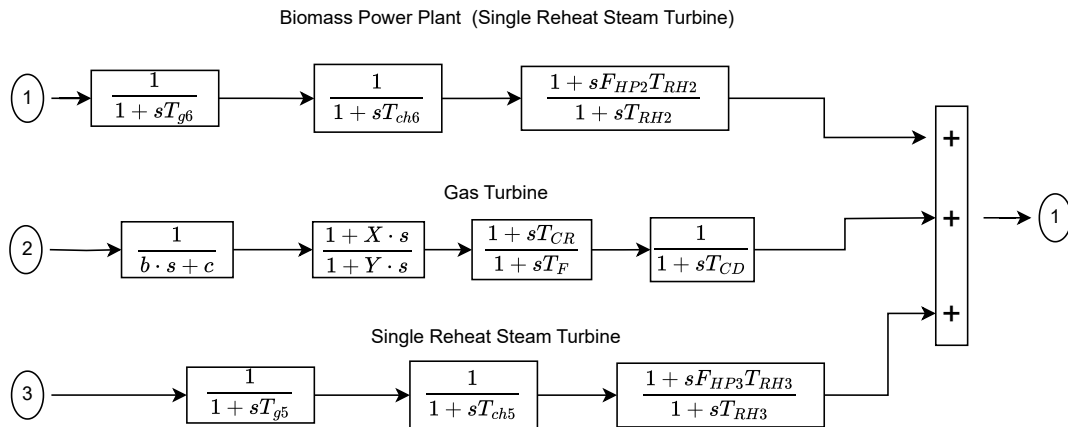
ως μέσο καύσης αέρια (π.χ. ατμοσφαιρικό αέρα, CO_2 , ή He) αντί για το νερό που αξιοποιείται στους ατμοστροβίλους. Ο αέρας εισέρχεται μέσω των διαφόρων υποδοχών εισόδου, συμπιέζεται, αναφλέγεται στον θάλαμο καύσης και μέσω της εκτόνωσης των καυσαερίων παράγεται μηχανικό έργο (Θερμοδυναμικός κύκλος Brayton). Στο γραμμικό μοντέλο που προσομοιώνει τη λειτουργία του αεριοστροβίλου, συναντάμε ένα σύνολο παραμέτρων μεταξύ των οποίων, η σταθερά c αντιπροσωπεύει τη θέση της βαλβίδας ελέγχου (*valve positioner constant*), ενώ οι παράμετροι b , X , Y , T_F , T_{CR} και T_{CD} αντιστοιχούν σε χρονικές σταθερές που χαρακτηρίζουν τη δυναμική απόκριση του συστήματος [6].

Τέλος συναντάμε ξανά μια μονάδα αντιστοιχεί στη μοντελοποίηση ενός ατμοστροβίλου με μονή αναθέρμανση, ενός από τους πιο διαδεδομένους τύπους μονάδων παραγωγής στα σύγχρονα ηλεκτρικά δίκτυα. Η παράμετρος T_{g5} εκφράζει τη χρονική σταθερά του ρυθμιστή στροβίλου, ενώ η παράμετρος T_{ch5} αντιστοιχεί στη χρονική σταθερά του θαλάμου υψηλής πίεσης. Όπως έχει ήδη αναφερθεί στο Κεφάλαιο 4, οι τιμές των παραμέτρων που χρησιμοποιούνται για τη μοντελοποίηση αντίστοιχων μονάδων μπορεί να διαφοροποιούνται, ώστε να προσεγγίζουν ρεαλιστικότερα τις πραγματικές συνθήκες λειτουργίας στο δίκτυο.

Οι αριθμητικές τιμές των παραμέτρων που συνθέτουν την Περιοχή 3 σε συνδυασμό με τις τιμές των σταθερών που συναντώνται στο υποσύστημα interconnected lines συνοψίζονται στον Πίνακα ??.

Πίνακας 7.1: Αριθμητικές τιμές των παραμέτρων της Περιοχής 3 [6].

Παράμετρος	Τιμή	Μονάδα Μέτρησης
T_{g5}	0.1	s
T_{ch5}	0.4	s
F_{HP3}	0.35	-
T_{RH3}	5	s
c	1	-
b	0.05	s
X	0.6	s
Y	1	s
T_F	0.23	s
T_{CR}	0.01	s
T_{CD}	0.2	s
T_{g6}	0.08	s
T_{ch6}	0.3	s
T_{RH2}	10	s
F_{HP2}	0.3	-
T_{12}	2	s
T_{13}	2	s
T_{23}	2	s
α_{12}	-1	-
α_{13}	-1	-
α_{23}	-1	-
K_{I5}	0.25	-
K_{I6}	0.25	-
K_{I7}	0.25	-
R_5	0.1	p.u.
R_6	0.1	p.u.
R_7	0.1	p.u.
B_3	31.3	p.u.
D_3	1.3	p.u.



Σχήμα 7.4: Υποσύστημα Power Plants 3.

7.3 Προετοιμασία δεδομένων για τους αλγορίθμους μηχανικής μάθησης

7.3.1 Συλλογή των Δεδομένων

Παρόμοια βήματα, όπως και στα προηγούμενα κεφάλαια, ακολουθήθηκαν και για τη συλλογή των δεδομένων που αφορούν το ΣΗΕ τριών περιοχών. Σε αυτήν την εκδοχή του προβλήματος, τα δεδομένα που κατασκευάζονται περιλαμβάνουν τις δυναμικές αποκρίσεις συχνότητας των Περιοχών 1, 2 και 3 (Δf_1 , Δf_2 , Δf_3), καθώς και τις μεταβολές της ισχύος στις διασυνδέσεις μεταξύ των τριών περιοχών. Συγκεκριμένα, πρόκειται για τη μεταβολή της ισχύος που ρέει μεταξύ της περιοχής 1 και των υπολοίπων δύο περιοχών ($\Delta P_{tie,1}$), τη μεταβολή της ισχύος που ρέει μεταξύ της περιοχής 2 και των περιοχών 1 και 3, ($\Delta P_{tie,2}$), καθώς και τη μεταβολή ισχύος μεταξύ τρίτης περιοχής και των υπολοίπων δυο ($\Delta P_{tie,3}$). Η χρονική διάρκεια κάθε χρονοσειράς απόκρισης παραμένει στα 90 δευτερόλεπτα και, πλέον, στόχος είναι η εκτίμηση των τριών σταθερών αδράνειας (H_1 , H_2 , H_3). Σε αυτή την περίπτωση, στα μοντέλα μηχανικής μάθησης που αναλύθηκαν στα προηγούμενα κεφάλαια, θα έχουμε ως εισόδους έξι διαφορετικές δυναμικές αποκρίσεις (τρεις δυναμικές αποκρίσεις συχνότητας και τρεις δυναμικές αποκρίσεις από τις ισχύεις των διασυνδέσεων).

Διατηρώντας σταθερές τις τιμές των παραμέτρων του πίνακα 7.1, αποθηκεύουμε αρχικά σε διάταξη τένσορα τις έξι αποκρίσεις του συστήματος που προκύπτουν λόγω της μεταβολής της ζητούμενης ισχύος σε κάθε περιοχή για κάθε μοναδικό συνδυασμό τιμών για τις σταθε-

ρές αδράνειας H_1 , H_2 και H_3 . Το σύστημα των τριών περιοχών ελέγχου, λοιπόν, για τυχαίες τιμές H_1 , H_2 , H_3 οι οποίες ανήκουν όλες στο εύρος [3,8] διεγείρεται, όπως αναφέρθηκε και νωρίτερα, από τον συνδυασμό τυχαίων διαταραχών ισχύος και ελεγχόμενων διαταραχών (δοκιμαστικών σημάτων) από μπαταρίες. Τα δοκιμαστικά σήματα των μπαταριών επιβάλλονται στην Περιοχή 1 και 3.

Αφού οι χρονοσειρές εξόδου του συστήματος δειγματοληπτούνται με ρυθμό 100 Hz (ένα δείγμα κάθε 0.01 δευτερόλεπτα), αποθηκεύονται σε έναν τένσορα του οποίου το μέγεθος έχει διπλασιαστεί συγκριτικά με εκείνον της Εικόνας 5.5. Οι διαστάσεις του νέου τένσορα που περιέχει το σύνολο των χρονοσειρών θα είναι:

$$X \in \mathbb{R}^{1000 \times N \times 6},$$

όπου $N = 9001$ είναι το πλήθος των δειγμάτων κάθε χρονοσειράς και ο τελευταίος άξονας αντιστοιχεί στις έξι εισόδους ($\Delta f_1, \Delta f_2, \Delta f_3, \Delta P_{tie,1}, \Delta P_{tie,2}, \Delta P_{tie,3}$). Το πλήθος των διαφορετικών τριάδων (H_1, H_2, H_3) των οποίων εξετάζεται η επίδραση στην έξοδο του συστήματος αποφασιστήκατε να είναι πάλι 1000. Αντίστοιχα, οι μεταβλητές εξόδου, δηλαδή οι τριάδες των σταθερών αδράνειας, αναπαρίστανται ως ένας δισδιάστατος πίνακας:

$$Y \in \mathbb{R}^{1000 \times 3},$$

όπου κάθε γραμμή περιέχει την αντίστοιχη τριπλέτα (H_1, H_2, H_3), υπεύθυνη για τις έξι χρονοσειρές του ίδιου πίνακα στον X .

7.3.2 Προεπεξεργασία των δεδομένων

Μετά την επιτυχή συλλογή και αποθήκευση των δεδομένων, προχωράμε στην προεπεξεργασία τους. Η διαδικασία δεν διαφέρει καθόλου με εκείνη που περιγράφηκε διεξοδικά στο Κεφάλαιο 5. Αρχικά πραγματοποιείται η εξαγωγή των στατιστικών στοιχείων για κάθε χρονοσειρά (μέση τιμή, τυπική απόκλιση, μέγιστη τιμή, ναδίρ, RMS και χρόνος ναδίρ) και έπειτα ακολουθεί η κατασκευή του νέου διδιάστατου πίνακα που θα φιλοξενήσει αυτά τα στατιστικά στοιχεία για κάθε χρονοσειρά. Συγκεκριμένα, ο πίνακας αυτός θα περιέχει συνολικά 45 στήλες, εκ των οποίων οι 42 θα αντιστοιχούν στους περιγραφικούς δείκτες κάθε χρονοσειράς και οι υπόλοιποι τρεις στην τριπλέτα εξόδων (H_1, H_2, H_3). Ο αριθμός των γραμμών του πίνακα παραμένει ίσος με 1000, όσο ακριβώς είναι και το σύνολο των διαφορετικών συνδυασμών εξόδων. Στη γλώσσα των μαθηματικών, αυτό μεταφράζεται ως εξής:

$$K \in \mathbb{R}^{1000 \times 45}, \quad D = [X \mid Y],$$

όπου

$$X \in \mathbb{R}^{1000 \times 42}$$

αντιπροσωπεύει το σύνολο των στατιστικών δεικτών που περιγράφουν τη συμπεριφορά των χρονοσειρών εισόδου, ενώ

$$Y \in \mathbb{R}^{1000 \times 3}$$

περιέχει τις μεταβλητές εξόδου, δηλαδή το ζεύγος των αδράνειας (H_1, H_2, H_3).

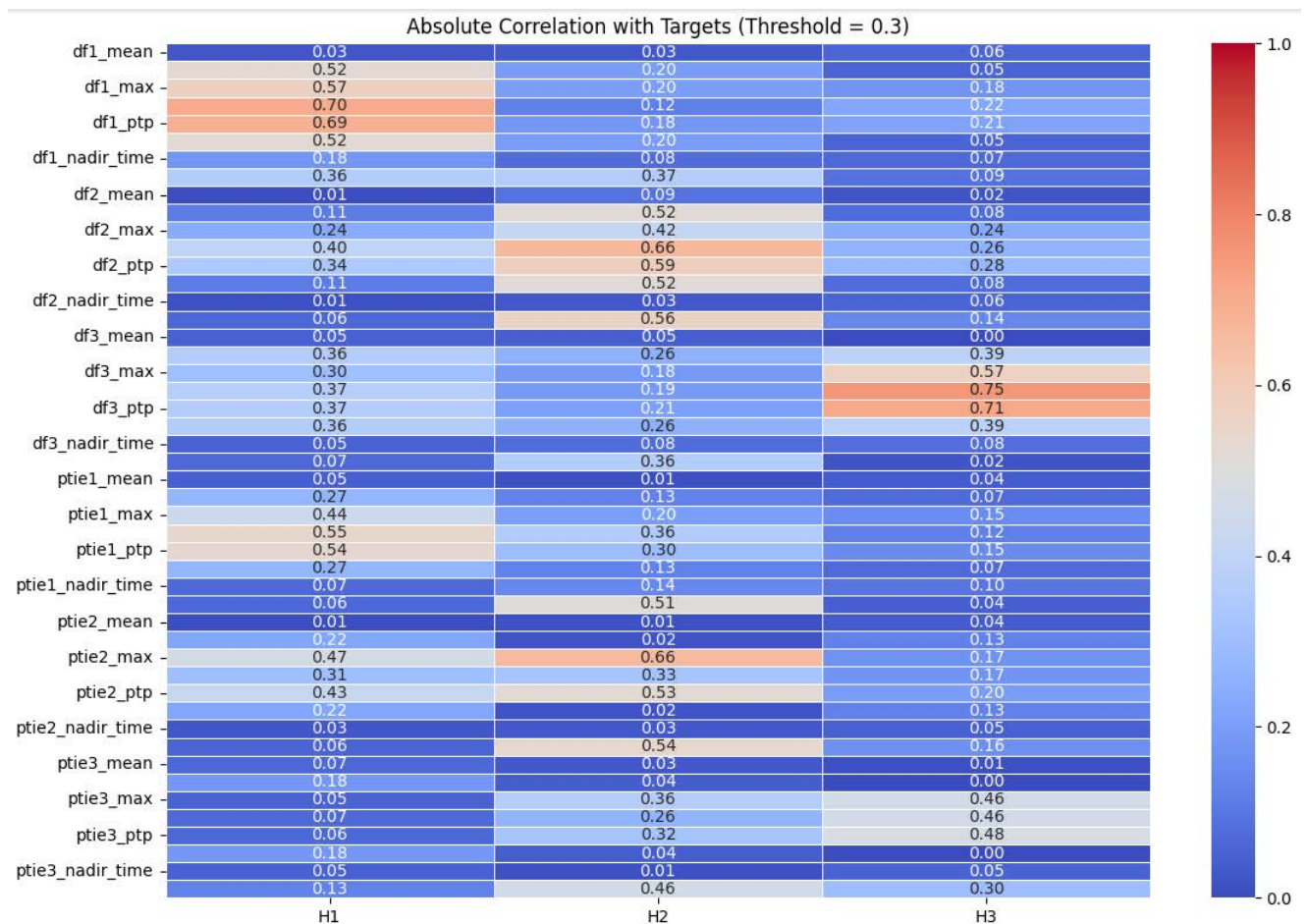
Αμέσως μετά, σειρά έχει ο «καθαρισμός» του συνόλου δεδομένων από ασταθείς χρονοσειρές και τις αντίστοιχες τριάδες εξόδων αδράνειας που προκαλούν αυτή τη συμπεριφορά. Για τον εντοπισμό τους καταφεύγουμε, για άλλη μια φορά, στον υπολογισμό του μεγέθους RMS_{ratio} για κάθε μία από τις 6 χρονοσειρές, του οποίου η τιμή θα τις κατατάξει στην κατηγορία των ασταθών ή ευσταθών σημάτων (threshold 2.5). Συνολικά, απορρίπτονται από τον πίνακα 75 εγγραφές, οι οποίες αφορούν ασταθή σήματα και τις τριάδες αδράνειας που τα προκαλούν ενώ στον αρχικό πίνακα έχουν ήδη προστεθεί οι 6 στήλες που εμπεριέχουν τις τιμές του RMS_{ratio} για το σύνολο των 6 χρονοσειρών.

Στη συνέχεια, ακολουθεί ένα επιπλέον φιλτράρισμα του πίνακα, επιλέγοντας αυτή τη φορά τις στήλες με τη μεγαλύτερη συνεισφορά στη διαμόρφωση της τελικής εξόδου. Ως εργαλείο για τον εντοπισμό του βαθμού επιρροής κάθε χαρακτηριστικού (*feature*) σε κάθε έξοδο χρησιμοποιείται ξανά ο δείκτης Pearson, με ελάχιστη επιτρεπόμενη τιμή 0.3. Εάν ένα χαρακτηριστικό δεν πληρεί αυτή τη συνθήκη για καμία από τις τρεις εξόδους, ολόκληρη η στήλη του απομακρύνεται από τον τελικό πίνακα.

Σύμφωνα με τον πίνακα συσχέτισης του σχήματος 7.5, απομακρύνονται συνολικά 18 στήλες από το dataset, οι οποίες αντιστοιχούσαν στα ακόλουθα χαρακτηριστικά:

$$\begin{aligned} &df1_mean, df1_nadir_time, \\ &df2_mean, df2_nadir_time, \\ &df3_mean, df3_nadir_time, \\ &ptie1_mean, ptie1_std, ptie1_rms, ptie1_nadir_time, \\ &ptie2_mean, ptie2_std, ptie2_rms, ptie2_nadir_time, \\ &ptie3_mean, ptie3_std, ptie3_rms, ptie3_nadir_time. \end{aligned}$$

Όπου $ptie1, ptie2, ptie3$ συμβολίζουν της μεταβολές της διασυνδετικής ισχύος μεταξύ του δικτύου και κάθε μεμονωμένης περιοχής.



Σχήμα 7.5: Πινάκας συσχέτισης (Correlation Matrix) σεναρίου 3 περιοχών.

Σειρά τώρα έχει το βήμα της κλιμάκωσης, όπου επιλέγεται, όπως και στο Σενάριο 1 του συστήματος των δύο διασυνδεδεμένων περιοχών, η μέθοδος *Min-Max Scaler*. Αυτό επιλέχθηκε αφού, μετά από μια σειρά πειραμάτων με διαφορετικές μεθόδους κλιμάκωσης, αποδείχθηκε για ακόμη μια φορά ότι επιδράει αποτελεσματικότερα στον ακριβή προσδιορισμό των εξόδων.

Το τελευταίο βήμα πριν τα δεδομένα εισαχθούν στα μοντέλα μηχανικής μάθησης είναι η διαμοιρασμός του dataset σε σύνολα εκπαίδευσης και επικύρωσης, και στη συνέχεια η οργάνωση σε μεταβλητές εισόδου και μεταβλητές στόχου. Ο πίνακας των μεταβλητών εισόδου (X_{train}) διαθέτει 30 στήλες, ενώ ο πίνακας που περιέχει τις μεταβλητές εξόδου διαθέτει συνολικά 3 στήλες.

7.4 Αξιολόγηση των Αλγορίθμων Μηχανικής Μάθησης

Για τον προσδιορισμό των αδρανειών του ΣΗΕ τριών περιοχών χρησιμοποιήθηκαν ξανά οι πέντε αλγόριθμοι μηχανικής μάθησης που εξετάστηκαν και στα προηγούμενα Κεφάλαια: Random Forest, MLR, SVR, XGBoost και MLP. Σε αντίθεση όμως με τα σενάρια που εξετάστηκαν στα προηγούμενα κεφάλαια, η ακρίβεια στις τιμές των τριών αδρανειών εξόδου δεν υπήρξε ικανοποιητική για την πλειοψηφία των μοντέλων, με μοναδική εξαίρεση το μοντέλο MLP, το οποίο για ακόμη μία φορά πέτυχε εξαιρετική ακρίβεια στον προσδιορισμό των τριών μεταβλητών εξόδου.

Η απόδοση τόσο του MLP όσο και του αλγορίθμου της πολυγραμμικής παλινδρόμησης όπου αποδείχθηκε η πιο αδύναμη μέθοδος προσέγγισης των εξόδων οπτικοποιείται στην συνέχεια αυτής της ενότητας (Σχήματα 7.6, 7.7, 7.8, 7.9, 7.10, 7.11). Στο τέλος της παρούσας ενότητας συνοψίζονται, σε πίνακες και διαγράμματα (7.4, 7.5, 7.6, 7.12), οι μετρικές αξιολόγησης του συνόλου των αλγορίθμων για κάθε μία από τις μεταβλητές στόχου (H_1 , H_2 , H_3).

7.4.1 Αξιολόγηση αλγορίθμου MLR

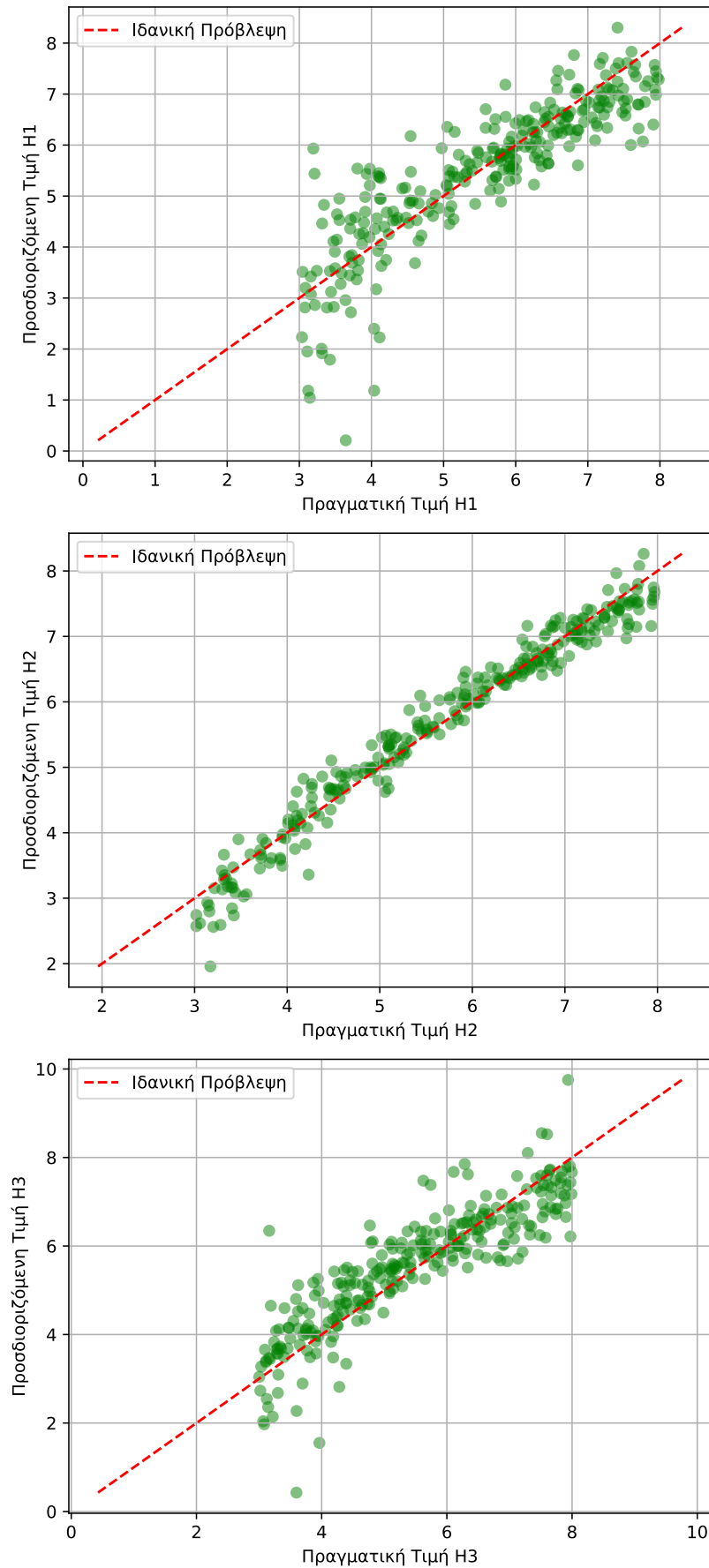
Οι μετρικές αξιολόγησης της ακρίβειας προσδιορισμού για καθεμία από τις τρεις εξόδους του συστήματος, χρησιμοποιώντας τον αλγόριθμο MLR, συνοψίζονται στον Πίνακα 7.2. Παρατηρούμε ότι ο αλγόριθμος MLR επιτυγχάνει αρκετά υψηλή ακρίβεια όσον αφορά τον προσδιορισμό της τιμής της αδράνειας της περιοχής 2, με τη μετρική R^2 να φτάνει σε πολύ υψηλά επίπεδα (πάνω από 0.96), υποδεικνύοντας εξαιρετική ικανότητα κατανόησης της συσχέτισης μεταξύ των μεταβλητών εισόδου και της συγκεκριμένης εξόδου του μοντέλου.

Επιπλέον, τα σφάλματα RMSE και MAE παραμένουν εξαιρετικά χαμηλά, αναδεικνύοντας την ακρίβεια που επιτυγχάνει το μοντέλο στις εκτιμήσεις του. Τέλος, η μετρική MAPE παραμένει κάτω από 5%, ποσοστό σφάλματος το οποίο θεωρείται γενικά αποδεκτό σε εφαρμογές εκτίμησης αδράνειας [33], υποδεικνύοντας ότι η μέθοδος παρουσιάζει υψηλή αξιοπιστία και ακρίβεια στην πρόβλεψη της συγκεκριμένης εξόδου. Ωστόσο, όσον αφορά τις υπόλοιπες δύο περιοχές, οι αποκλίσεις που προκύπτουν είναι σημαντικές ($MAPE > 10\%$), καθιστώντας το μοντέλο συνολικά αδύναμο και ακατάλληλο να διαχειριστεί τη συγκεκριμένη φύση των δεδομένων και να παράγει έγκυρες προβλέψεις για αυτές τις εξόδους.

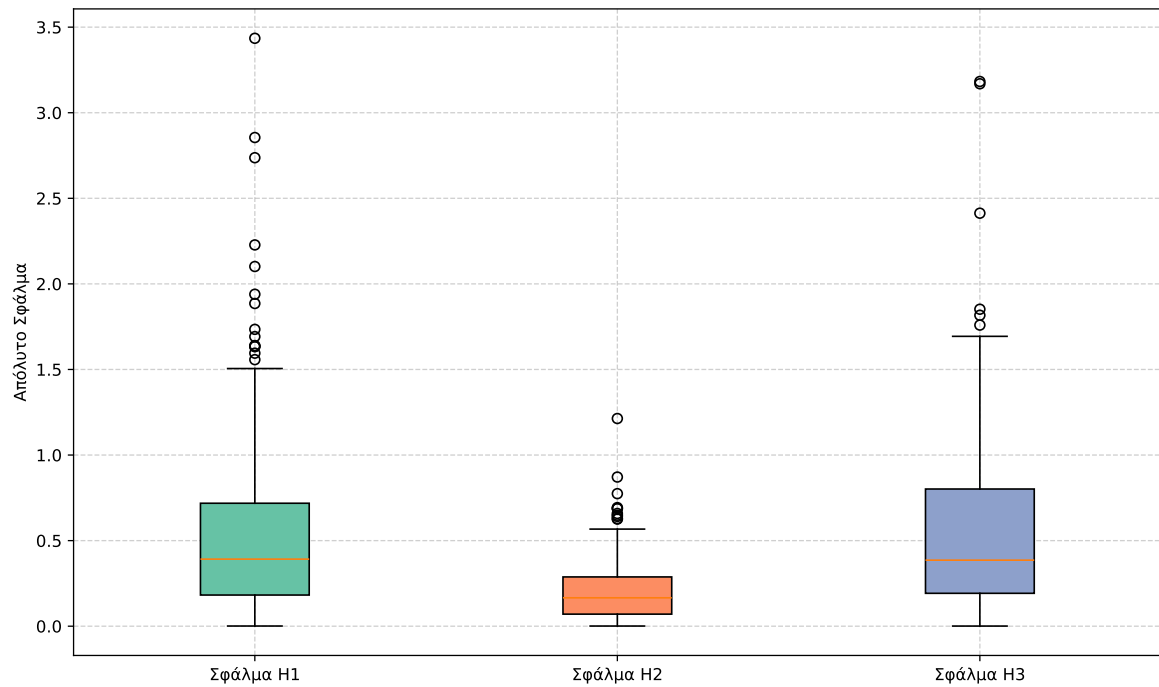
Πίνακας 7.2: Αποτελέσματα αξιολόγησης του αλγορίθμου MLR για τις εξόδους H_1 , H_2 και H_3 .

Έξοδος	R^2	RMSE	MAE	MAPE (%)
H_1	0.72639	0.73961	0.53768	11.23
H_2	0.96251	0.27483	0.20851	4.10
H_3	0.75632	0.72430	0.54063	11.04

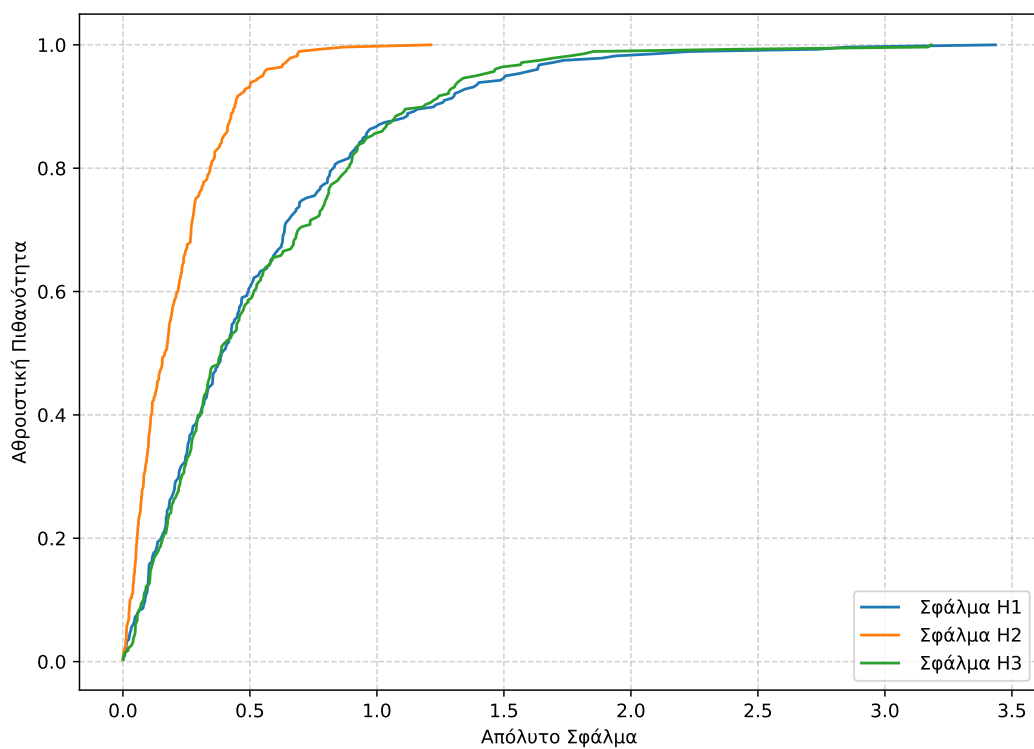
Η Εικόνα 7.6 απεικονίζει την απόκλιση μεταξύ των πραγματικών και των προσδιοριζόμενων τιμών που προκύπτουν από το μοντέλο MLR για καθεμία από τις εξόδους του συστήματος, ενώ το Σχήμα 7.7 παρουσιάζει τα θυκογράμματα (box plots) για κάθε μία από τις αδράνειες των τριών περιοχών (H_1 , H_2 , H_3). Τέλος, η Εικόνα 7.8 αναπαριστά τις γραφικές παραστάσεις των συναρτήσεων αθροιστικής κατανομής (CDF) του απόλυτου σφάλματος για καθεμία από τις τρεις προσδιοριζόμενες τιμές εξόδου.



Σχήμα 7.6: Απόκλιση πραγματικών από προσδιοριζόμενων τιμών των σταθερών H_1 , H_2 , H_3 από το μοντέλο MLR.



Σχήμα 7.7: Θηκόγραμμα του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLR για τις σταθερές αδράνειας H_1 , H_2 και H_3 .



Σχήμα 7.8: Συνάρτηση CDF του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLR για τις σταθερές H_1, H_2 και H_3 .

Οι τρεις αυτές εικόνες επιβεβαιώνουν τα αποτελέσματα του Πίνακα 7.2, με την αδράνεια της περιοχής 2 να αποτελεί τη μοναδική που προσδιορίζεται με υψηλή ακρίβεια από το συγκεκριμένο μοντέλο, ενώ τα σφάλματα για τον προσδιορισμό των αδρανειών H_1 και H_3 να είναι μη αποδεκτά.

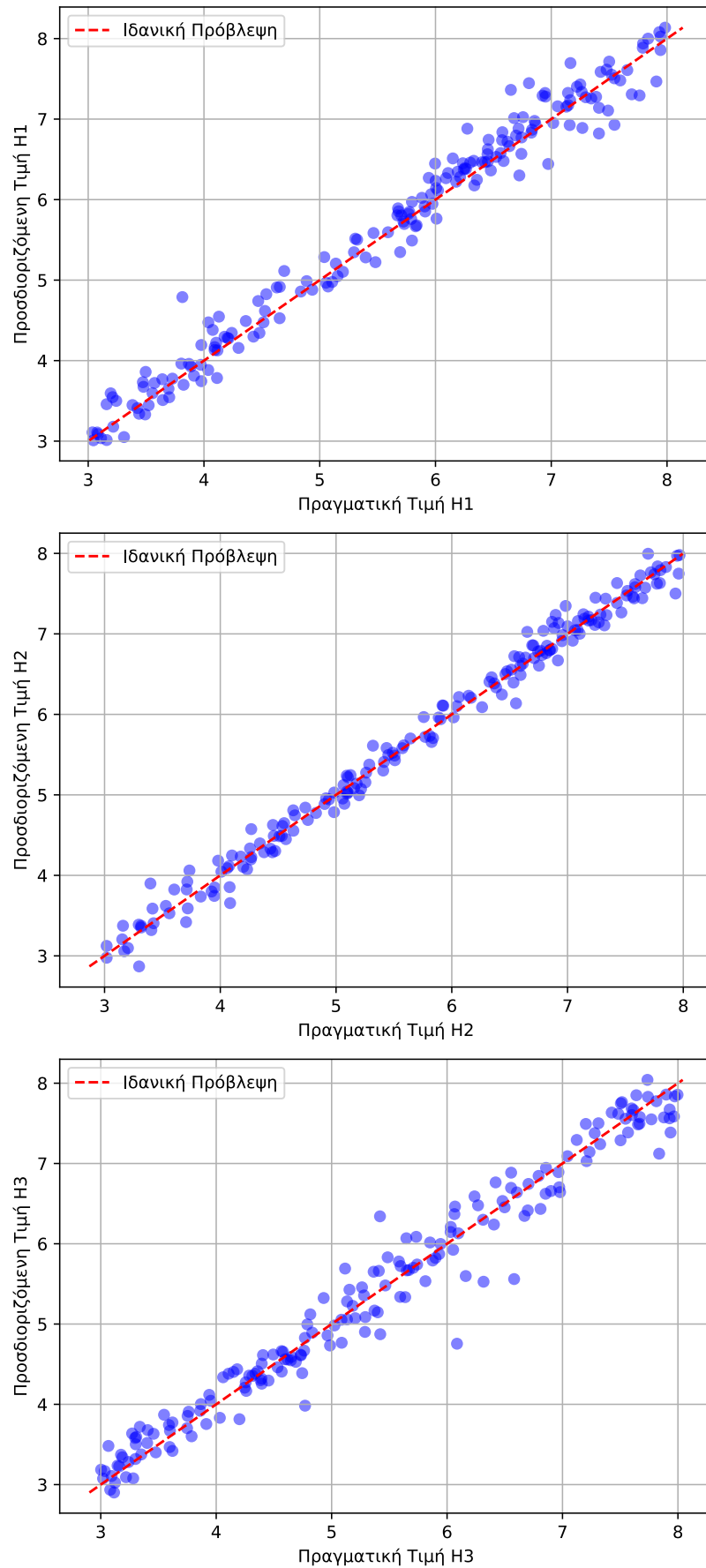
7.4.2 Αξιολόγηση Αλγορίθμου MLP

Σαφώς βελτιωμένες εμφανίζονται οι τιμές των μετρικών του αλγορίθμου MLP, όπως φαίνονται στον Πίνακα 7.3. Παρατηρούμε ότι, σε αντίθεση με την πολυγραμμική παλινδρόμηση, ο αλγόριθμος MLP επιτυγχάνει εξαιρετική ακρίβεια και για τις τρεις περιοχές του συστήματος. Συγκεκριμένα, η μετρική R^2 φτάνει σε πολύ υψηλά επίπεδα (πάνω από 0.96 για όλες τις εξόδους), υποδεικνύοντας την ικανότητα του MLP να συλλαμβάνει με ακρίβεια τη συσχέτιση μεταξύ των μεταβλητών εισόδου και των εξόδων του μοντέλου. Επιπλέον, τα σφάλματα RMSE και MAE παραμένουν πολύ χαμηλά, ενώ το MAPE δεν ξεπερνά το 4%, γεγονός που καταδεικνύει την αξιοπιστία του μοντέλου στις προβλέψεις του.

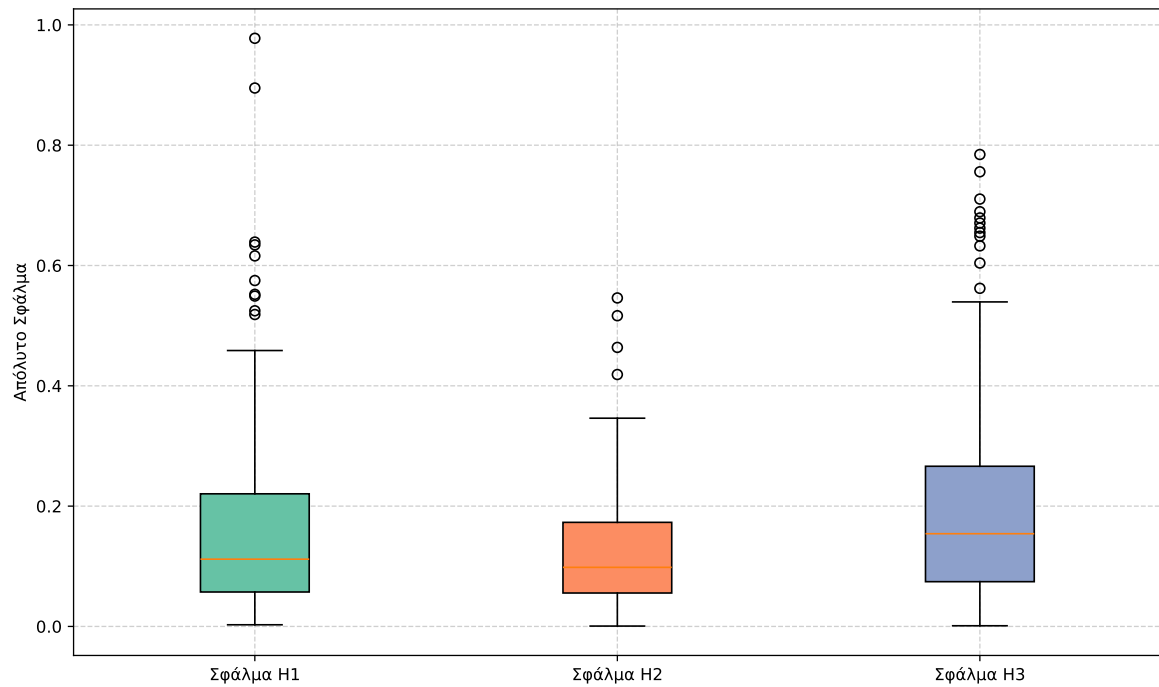
Πίνακας 7.3: Αποτελέσματα αξιολόγησης του αλγορίθμου MLP για τις εξόδους H_1 , H_2 και H_3 .

Έξοδος	R^2	RMSE	MAE	MAPE (%)
H_1	0.96842	0.25532	0.17874	3.22
H_2	0.98268	0.18651	0.14908	2.77
H_3	0.96509	0.27971	0.20856	3.91

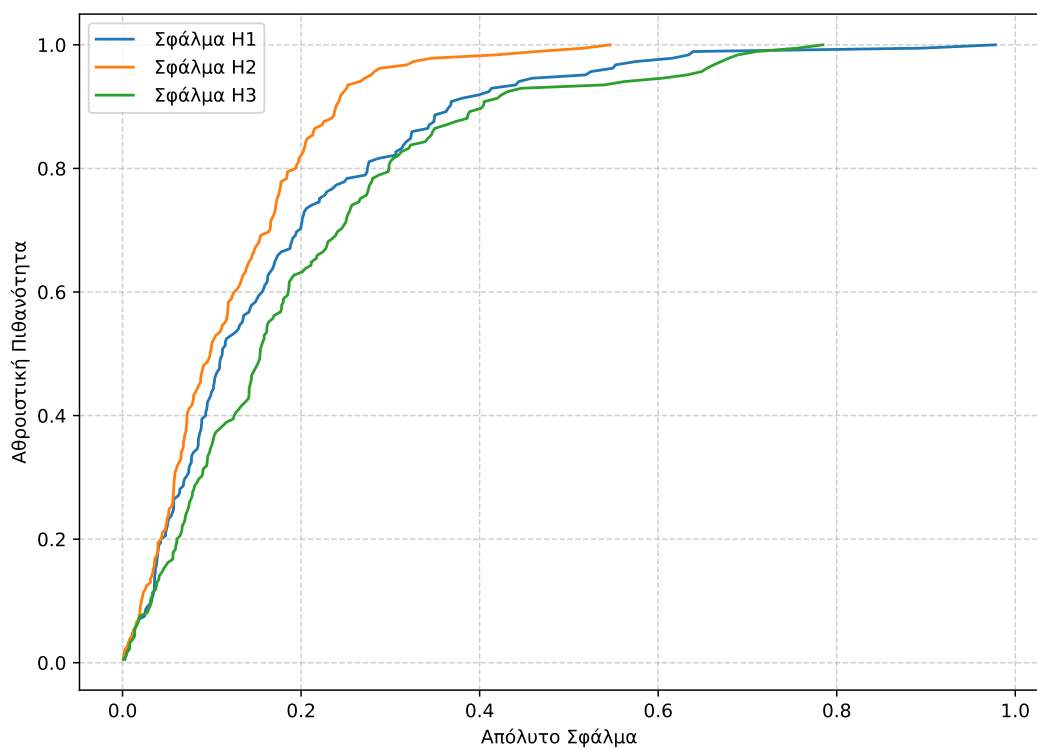
Οι Εικόνες 7.9, 7.10, 7.11 οπτικοποιούν την απόκλιση μεταξύ των πραγματικών και των προβλεπόμενων τιμών της αδράνειας για κάθε περιοχή, τα θηκογράμματα των τριών αδρανειών, καθώς και τις γραφικές παραστάσεις των συναρτήσεων αθροιστικής κατανομής (CDF) του απόλυτου σφάλματος πρόβλεψης για κάθε μία από τις τρεις προσδιοριζόμενες εξόδους, αντίστοιχα.



Σχήμα 7.9: Απόκλιση πραγματικών από προσδιοριζόμενων τιμών των σταθερών H_1 , H_2 , H_3 από το μοντέλο MLP.



Σχήμα 7.10: Θηκόγραμμα του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLP για τις σταθερές αδράνειας H_1 , H_2 και H_3 .



Σχήμα 7.11: Συνάρτηση CDF του απόλυτου σφάλματος πρόβλεψης του μοντέλου MLP για τις σταθερές H_1, H_2 και H_3 .

Η σύγκριση των εικόνων 7.9, 7.10, 7.11 με τις αντίστοιχες εικόνες που αφορούσαν τον αλγόριθμο MLR (7.6, 7.7, 7.8), επιβεβαιώνει για μία ακόμη φορά την υπεροχή του νευρωνικού δικτύου MLP σε σύγκριση με το μοντέλο της πολυγραμμικής παλινδρόμησης, όσον αφορά την εκτίμηση των αδρανειών και στις τρεις περιοχές του συστήματος.

7.5 Αξιολόγηση των Υπόλοιπων Αλγορίθμων

Οι επιδόσεις των υπόλοιπων αλγορίθμων μηχανικής μάθησης όσον αφορά την ακρίβεια προσδιορισμού των σταθερών αδράνειας H_1 , H_2 και H_3 βρίσκονται μεταξύ των δύο ακραίων περιπτώσεων που περιγράψαμε νωρίτερα, χωρίς ωστόσο να αποτελούν αξιόπιστη λύση για το συγκεκριμένο πρόβλημα. Οι μετρικές αξιολόγησης της ακρίβειας κάθε αλγορίθμου στον προσδιορισμό κάθε μίας από τις εξόδους συνοψίζονται στους Πίνακες 7.4, 7.5, 7.6.

Πίνακας 7.4: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_1 (Σενάριο Τριών Περιοχών).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.89401	0.47207	0.37476	7.31
MLR	0.72639	0.73961	0.53768	11.23
XGBoost	0.84580	0.55645	0.39973	7.65
SVR	0.89764	0.45238	0.33046	6.73
MLP	0.96842	0.25532	0.17874	3.22

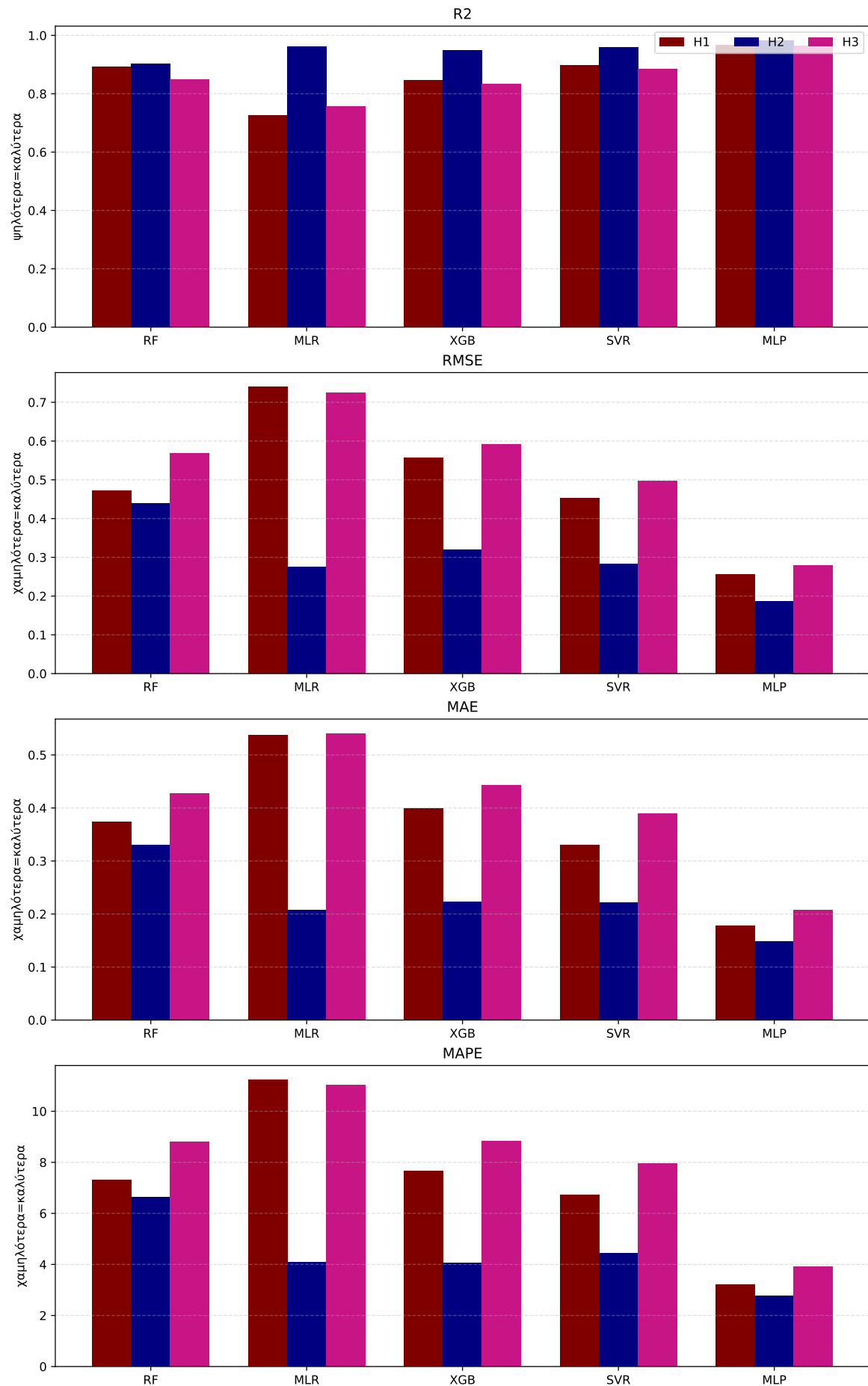
Πίνακας 7.5: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_2 (Σενάριο Τριών Περιοχών).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.90435	0.43897	0.33079	6.64
MLR	0.96251	0.27483	0.20851	4.10
XGBoost	0.94893	0.31987	0.22360	4.07
SVR	0.96029	0.28286	0.22208	4.44
MLP	0.98268	0.18651	0.14908	2.77

Πίνακας 7.6: Αποτελέσματα αξιολόγησης των αλγορίθμων μηχανικής μάθησης για την έξοδο H_3 (Σενάριο Τριών Περιοχών).

Αλγόριθμος	R^2	RMSE	MAE	MAPE (%)
Random Forest	0.84969	0.56886	0.42812	8.80
MLR	0.75632	0.72430	0.54063	11.04
XGBoost	0.83401	0.59078	0.44327	8.83
SVR	0.88514	0.49727	0.38984	7.96
MLP	0.96509	0.27971	0.20856	3.91

Ένα συμπέρασμα που μπορούμε να εξάγουμε παρατηρώντας προσεκτικά τους παραπάνω πίνακες είναι ότι η τιμή της αδράνειας της περιοχής 2 προσδιορίζεται με ικανοποιητική ακρίβεια, ανεξαρτήτως του αλγορίθμου που χρησιμοποιείται. Η εικόνα 7.12 παρουσιάζει με διαδραστικό τρόπο τις διαφορές στις επιδόσεις του κάθε αλγορίθμου στον προσδιορισμό των τελικών τιμών εξόδου ανά μετρική.



Σχήμα 7.12: Σύγκριση Επίδοσης των Μοντέλων ανά Μετρική-Σενάριο 3 Περιοχών.

Κεφάλαιο 8

Συμπεράσματα

8.1 Σύνοψη και συμπεράσματα

Η μεγάλη διείσδυση μονάδων ΑΠΕ στα σύγχρονα ΣΗΕ συνεπάγεται έναν μεγάλο αριθμό προκλήσεων στις οποίες οι ενεργειακοί μηχανικοί καλούνται να βρουν καινοτόμες λύσεις. Μια από τις σημαντικότερες προκλήσεις της διαρκώς αυξανόμενης διείσδυσης των μονάδων ΑΠΕ, και της αντικατάστασης των συμβατικών πηγών ενέργειας που αυτή συνεπάγεται, αφορά στη μείωση της συνολικής αδράνειας των ΣΗΕ. Η συνολική αδράνεια του δικτύου αποτελεί έναν δείκτη της ανθεκτικότητας του ΣΗΕ σε απότομες διαταραχές του ισοζυγίου παραγωγής/κατανάλωσης. Γεννάται, επομένως, η επιτακτική ανάγκη παρακολούθησης της αδράνειας των ΣΗΕ σε πραγματικό χρόνο, με στόχο την έγκαιρη λήψη κατάλληλων διορθωτικών ενεργειών από τους διαχειριστές του δικτύου, όπως για παράδειγμα η ενεργοποίηση εφεδρικών υδροηλεκτρικών εργοστασίων και σύγχρονων πυκνωτών, με στόχο τη διασφάλιση της ευστάθειας συχνότητας του συστήματος. Η γνώση της αδράνειας επιτρέπει στην πραγματικότητα στους διαχειριστές των δικτύων να γνωρίζουν κάθε χρονική στιγμή τα όρια ευστάθειας συχνότητας του ΣΗΕ.

Στα πλαίσια της παρούσας διπλωματικής εργασίας αναπτύχθηκε ένα σύνολο μεθόδων για τον προσδιορισμό των επιπέδων αδράνειας ΣΗΕ, κάνοντας χρήση αλγορίθμων μηχανικής μάθησης. Οι προτεινόμενες μέθοδοι αξιολογήθηκαν αρχικά σε ΣΗΕ δύο περιοχών και στη συνέχεια η επικύρωση τους πραγματοποιήθηκε και σε ΣΗΕ τριών περιοχών. Δεδομένα εκπαίδευσης για τους αλγόριθμους μηχανικής μάθησης αποκτήθηκαν μέσω της προσομοίωσης της δυναμικής συμπεριφοράς ΣΗΕ στο περιβάλλον Simulink.

Αρχικά, η επικύρωση των προτεινόμενων μεθόδων πραγματοποιήθηκε σε ΣΗΕ δύο πε-

ριοχών. Εξετάστηκαν τρία διαφορετικά σενάρια τα οποία διαφοροποιούνταν ως προς το μείγμα παραγωγής, το είδος, δηλαδή, των μονάδων παραγωγής. Σε κάθε σενάριο, εξετάζονται διαφορετικά επίπεδα αδράνειας (H_1 , H_2) και καταγράφονται οι δυναμικές αποκρίσεις του συστήματος (Δf_1 , Δf_2 , ΔP_{tie}). Εν συνεχεία, από τις δυναμικές αποκρίσεις υπολογίζεται ένα σύνολο περιγραφικών δεικτών, έτσι ώστε το εξεταζόμενο πρόβλημά να μετατραπεί σε πρόβλημα παλινδρόμησης. Έπειτα, οι περιγραφικοί δείκτες χρησιμοποιούνται ως είσοδοι για την εκπαίδευση πέντε διαφορετικών αλγορίθμων μηχανικής μάθησης: του Random Forest, της Multiple Linear Regression, του XGBoost, του SVR και τέλος του μοντέλου MLP το οποίο ανήκει στην κατηγορία των νευρωνικών δικτύων.

Όσον αφορά την αξιολόγηση της εγκυρότητας των εκτιμώμενων επιπέδων αδράνειας από τα εξεταζόμενα μοντέλα μηχανικής μάθησης, και στα τρία σενάρια που εξετάστηκαν παρατηρήθηκε ένα κοινό μοτίβο, κατά το οποίο οι αλγόριθμοι MLP και Random Forest παρήγαγαν τις ακριβέστερες προβλέψεις, κατα σειρά, με τους XGBoost και SVR να ακολουθούν. Το μεγαλύτερο σφάλμα συνολικά κατέγραψε το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης, το οποίο παρουσίασε ελαφρώς μεγαλύτερα σφάλματα από τους υπόλοιπους αλγορίθμους στα σενάρια 1 και 2 με τις επιδόσεις, ωστόσο, στο σενάριο 3, να ξεπερνάνε εκείνες του XGBoost και του SVR. Η ακρίβεια που επιτεύχθηκε υπήρξε εντυπωσιακή, με το μέσο ποσοστιαίο σφάλμα (MAPE) σε ορισμένες περιπτώσεις (Random Forest, XGBoost, MLR στο σενάριο 3 και MLP σε όλα τα σενάρια), να παραμένει κάτω από το 4%, απόκλιση που θεωρείται αποδεκτή στις εφαρμογές προσδιορισμού αδράνειας, όπως αναφέρεται σε σχετικές μελέτες [33]

Στη συνέχεια, η μελέτη μας επεκτάθηκε στον προσδιορισμό των αδρανειών ενός διασυνδεδεμένου ΣΗΕ, αποτελούμενου από τρεις περιοχές. Παρά τον διπλασιασμό του όγκου των δεδομένων, η προσέγγιση του προβλήματος παρέμεινε πανομοιότυπη. Τα αποτελέσματα καταδεικνύουν ότι σε αυτή την περίπτωση ο MLP αποτελεί τη μοναδική αξιόπιστη λύση, προσφέροντας εκτιμήσεις για τις αδράνεις όλων των περιοχών του ΣΗΕ με *MAPE* μικρότερο του 3.91%.

Αν και η εκπαίδευση των μοντέλων βασίστηκε στην χρήση συνθετικού συνόλου δεδομένων (synthetic dataset) και όχι πραγματικών μετρήσεων, τα παραγόμενα μοντέλα είναι θεωρητικά ικανά να εφαρμοστούν σε πραγματικές συνθήκες δικτύου. Συγκεκριμένα, λαμβάνοντας ως είσοδο ένα σύνολο δυναμικών αποκρίσεων του ΣΗΕ, μπορούν να προσδιορίζουν με υψηλή ακρίβεια και να ενημερώνουν σε πραγματικό χρόνο τις σταθερές αδράνειας των

περιοχών του ΣΗΕ.

Αξίζει να αναφερθεί ότι τα μοντέλα μηχανικής μάθησης που εξετάστηκαν υστερούν στο να εκτιμήσουν τη συσχέτιση μεταξύ των δυναμικών αποκρίσεων κανονικής λειτουργίας του συστήματος και των σταθερών αδράνειας. Είναι απαραίτητη, λοιπόν, η εφαρμογή ενός δοκιμαστικού σήματος, το οποίο όμως εύκολα μπορεί να εφαρμοστεί με τη βοήθεια μιας μπαταρίας. Συνεπώς, η εφαρμογή των προτεινόμενων μοντέλων μηχανικής μάθησης σε πραγματικές συνθήκες απαιτεί τη χρήση αυτοματισμού που θα θέτει μια μπαταρία σε κύκλους φόρτισης και εκφόρτισης, ώστε να παράγονται τα απαραίτητα δοκιμαστικά σήματα.

Ο μεγαλύτερος περιορισμός του συνόλου των μοντέλων που αναπτύχθηκαν, και γενικότερα ολόκληρης της προσέγγισης που αναπτύχθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας, ήταν η υπόθεση ότι η μορφή των αποκρίσεων συχνότητας και ισχύος εξαρτάται αποκλειστικά από τις τιμές των σταθερών αδράνειας που έχουμε ανά περιοχή. Στην πραγματικότητα, υπάρχουν και άλλες παράμετροι των οποίων οι τιμές επηρεάζουν τη μορφή των δυναμικών αποκρίσεων του ΣΗΕ, όπως για παράδειγμα ο στατισμός του συστήματος (R), η σταθερά αυτορρύθμισης του φορτίου (D) και τα κέρδη των ολοκληρωτικών ελεγκτών. Στα πλαίσια της παρούσας διπλωματικής εργασίας οι συγκεκριμένες παράμετροι θεωρήθηκαν γνωστές και σταθερές κατά τη διάρκεια όλων των προσομοιώσεων.

Τέλος, αξίζει να αναφερθεί ότι το μείγμα παραγωγής διαδραματίζει επίσης σημαντικό ρόλο στην ακρίβεια του προσδιορισμού των σταθερών αδράνειας. Συνεπώς, εάν τα προτεινόμενα μοντέλα εφαρμοστούν σε συνθήκες πραγματικού δικτύου, πρέπει να εκπαιδευτούν σε συγκεκριμένο μείγμα παραγωγής για να παρέχουν αξιόπιστες προβλέψεις.

8.2 Μελλοντικές επεκτάσεις

Η παρούσα διπλωματική εργασία μπορεί να λάβει ένα μεγάλο αριθμό μελλοντικών επεκτάσεων. Οι σημαντικότερες εξ αυτών είναι εκείνες που μπορούν να συνδράμουν ώστε να ξεπεραστούν τα προβλήματα και οι περιορισμοί που αναφέρθηκαν παραπάνω. Πιο συγκεκριμένα, προτείνεται η επέκταση των προτεινόμενων μοντέλων, ώστε αυτά να λαμβάνουν υπόψη τους τους στατισμούς των μονάδων (R), τη σταθερά αυτορρύθμισης του φορτίου (D) και τα κέρδη των ολοκληρωτικών ελεγκτών του βρόγχου δευτερεύουσας ρύθμισης συχνότητας (K_I).

Μια ακόμα σημαντική μελλοντική επέκταση αφορά την ενίσχυση της γενικευσιμότητας

των μοντέλων, ώστε να μπορούν να αποδίδουν αξιόπιστα κάτω από διαφορετικά μείγματα μονάδων παραγωγής. Όπως αναφέρθηκε και προηγουμένως, διαφορετικά μείγματα παραγωγής έχουν ως αποτέλεσμα διαφορετικές δυναμικές αποκρίσεις εξόδου (αποκρίσεις (Δf_1 , Δf_2 , ΔP_{tie}) για τα ίδια επίπεδα αδράνειας. Για να αντιμετωπιστεί αυτό το ζήτημα, μπορούν να εκπαιδευτούν τα μοντέλα μηχανικής μάθησης με ένα μεγαλύτερο σύνολο δεδομένων εκπαίδευσης. Δηλαδή ένα σύνολο δεδομένων που θα συνδυάζει διαφορετικά επίπεδα αδράνειας και διαφορετικά μείγματα παραγωγής. Με αυτόν τον τρόπο, το μοντέλο μαθαίνει να αναγνωρίζει τις βασικές σχέσεις μεταξύ των δυναμικών αποκρίσεων του ΣΗΕ και των επιπέδων αδράνειας, ανεξάρτητα από τις συγκεκριμένες μονάδες που παράγουν ισχύ κάθε χρονική στιγμή. Μια άλλη προέκταση, η οποία θα πρέπει να εξεταστεί, είναι η τεχνική του transfer learning, η οποία επιτρέπει την προσαρμογή ενός ήδη εκπαιδευμένου μοντέλου σε νέα σενάρια παραγωγής χρησιμοποιώντας μικρότερο όγκο δεδομένων εκπαίδευσης, χωρίς να χρειάζεται επανεκπαίδευση των μοντέλων από την αρχή. Οι δύο αυτές προσεγγίσεις/προεκτάσεις εξασφαλίζουν ότι τα αναπτυσσόμενα μοντέλα παραμένουν αξιόπιστα και ακριβή, ακόμη και όταν οι συνθήκες λειτουργίας του δικτύου αποκλίνουν σημαντικά από τις συνθήκες που εκπαιδεύτηκαν.

Τέλος, για την παράκαμψη της προ επεξεργασίας των δυναμικών αποκρίσεων και του υπολογισμού των στατιστικών δεικτών που απαιτούνται για την ανάπτυξη των μοντέλων που εξετάστηκαν, μπορεί να αξιοποιηθεί η χρήση σύγχρονων νευρωνικών δικτύων τα οποία είναι σε θέση να επεξεργάζονται απευθείας τα ακατέργαστα δεδομένα (raw time series). Αρχιτεκτονικές όπως τα Recurrent Neural Networks (RNNs) και οι παραλλαγές τους LSTM/GRU, τα Temporal Convolutional Networks (TCN), καθώς και οι Transformers για χρονοσειρές, ενδέχεται να φανούν ιδιαίτερα αποτελεσματικές τεχνικές. Μάλιστα, η εφαρμογή αυτών των τεχνικών θα μπορούσε να επιτρέψει την ακριβέστερη και πιο ευέλικτη εκτίμηση της δυναμικής συμπεριφοράς ενός ΣΗΕ, ενισχύοντας έτσι τη δυνατότητα παρακολούθησης των επιπέδων αδράνειας σε πραγματικό χρόνο.

Συνοψίζοντας, παρά τις προκλήσεις που υπάρχουν στις μέρες μας, το αντικείμενο του προσδιορισμού της αδράνειας σύγχρονων ΣΗΕ είναι ένα ερευνητικό πεδίο με σημαντικές προοπτικές περαιτέρω εξέλιξης. Η αναζήτηση νέων, γρήγορων και αποτελεσματικών μεθόδων για την παρακολούθηση σε πραγματικό χρόνο των επιπέδων αδράνειας ΣΗΕ αποτελεί αναγκαιότητα, ώστε να διασφαλιστεί η ευστάθεια σύγχρονων ΣΗΕ στα οποία κυριαρχούν μονάδες ΑΠΕ.

Βιβλιογραφία

- [1] Bendong Tan, Junbo Zhao, Marcos Netto, Venkat Krishnan, Vladimir Terzija, and Yingchen Zhang. Power system inertia estimation: Review of methods and the impacts of converter-interfaced generations. *International Journal of Electrical Power Energy Systems*, 134:107362, 2022.
- [2] Wesley Cole and Bethany Frew. Inertia and the power grid: A guide without the spin. Report NREL/TP-5C00-73856, National Renewable Energy Laboratory (NREL), Golden, CO, 2020. Accessed: [2025-01-09].
- [3] Islam Uddin, Hamid Awan, Majdi Khalid, Salman Khan, Shahid Akbar, Mahidur Sarker, Maher Abdolrasol, and Thamer A. H. Alghamdi. A hybrid residue based sequential encoding mechanism with xgboost improved ensemble model for identifying 5-hydroxymethylcytosine modifications. *Scientific Reports*, 14, 09 2024.
- [4] Prabha Kundur. *Power System Stability and Control*. McGraw-Hill, 1994.
- [5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [6] Ana Fernández-Guillamón, Eduard Muljadi, and Angel Molina-García. Frequency control studies: A review of power system, conventional and renewable generation unit modeling. *Electric Power Systems Research*, 211:108191, 2022.
- [7] ENTSO-E. Winter outlook report 2023-2024. Technical report, European Network of Transmission System Operators for Electricity (ENTSO-E), October 2023.
- [8] D.C. Mikulecky and R. Rosen. Why are organisms different from machines. *Intelligent Engineering Systems Through Artificial Neural Networks*, 9:193–198, 1999.

- [9] ENTSO-E. Frequency stability evaluation criteria for the synchronous zone of continental europe. Report, European Network of Transmission System Operators for Electricity, 2020.
- [10] Hadi Alharbi, Ahmed Abu-Siada, Mohammad AlMuhaini, and Yateendra Mishra. Machine learning-based inertia estimation in power systems: A review of methods and challenges. *Electric Power Systems Research*, 236:109765, 2024.
- [11] Mahdi Heidari, Lei Ding, Mostafa Kheshti, Weiyu Bao, Xiaowei Zhao, Marjan Popov, and Vladimir Terzija. A review on application of machine learning-based methods for power system inertia monitoring. *International Journal of Electrical Power Energy Systems*, 162:110279, 2024.
- [12] Zhen Tang, Lihua Hao, Jiapeng Li, Qiuyi Chen, Yujun Li, and Zhao Xu. Power system inertia estimation based on long short term memory network. In *2021 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 2542–2547, 2021.
- [13] Sayanee Das and Paramita Chattopadhyay. Data-driven inertia estimation of power systems using autoencoder-based unsupervised feature learning. In *2023 IEEE 3rd Applied Signal Processing Conference (ASPCON)*, pages 222–226, 2023.
- [14] Vasupalli Sagar and Sachin Jain. System identification-based estimation of power system inertia using pmu data. pages 1–6, 12 2020.
- [15] Lei Zhang, Zhihao Guo, Qianhui Tao, Zhizhi Xiong, and Jing Ye. Xgboost-based short-term prediction method for power system inertia and its interpretability. *Energy Reports*, 9:1458–1469, 2023. Selected papers from 2022 International Conference on Frontiers of Energy and Environment Engineering.
- [16] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, NY, USA, 2006.
- [17] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.
- [18] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, Sebastopol, CA, 2nd edition, 2019.

- [19] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2 edition, 2009.
- [20] Rob J Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2006.
- [21] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting models and prediction intervals for the multiplicative holt-winters method. *International Journal of Forecasting*, 27(4):956–979, 2011.
- [22] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.
- [23] Nico JD Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.
- [24] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
- [25] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
- [26] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [27] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [28] P.M. Anderson and M. Mirheydar. A low-order system frequency response model. *IEEE Transactions on Power Systems*, 5(3):720–729, 1990.
- [29] Banaja Mohanty, Sidhartha Panda, and P.K. Hota. Controller parameters tuning of differential evolution algorithm and its application to load frequency control of multi-source power system. *International Journal of Electrical Power Energy Systems*, 54:77–85, 2014.

- [30] John W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.
- [31] Sheldon M. Ross. *A First Course in Probability*. Pearson, 9 edition, 2014.
- [32] Hêmin Golpîra and H. Bevrani. A framework for economic load frequency control design using modified multi-objective genetic algorithm. *Electric Power Components and Systems*, 42, 06 2014.
- [33] D. Linaro et al. Continuous estimation of power system inertia using convolutional neural networks. *Nature Communications*, 14(1):1–10, 2023.