



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

Παραγωγή ακραίων δεδομένων σε συστήματα
ηλεκτρικής ενέργειας με χρήση μεγάλων
γλωσσικών μοντέλων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΑΝΤΩΝΙΟΣ ΔΗΜΗΤΡΙΟΣ ΑΛΙΚΑΡΗΣ

Επιβλέπων: Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

Παραγωγή ακραίων δεδομένων σε συστήματα
ηλεκτρικής ενέργειας με χρήση μεγάλων
γλωσσικών μοντέλων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΑΝΤΩΝΙΟΣ ΔΗΜΗΤΡΙΟΣ ΑΛΙΚΑΡΗΣ

Επιβλέπων: Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17^η Οκτωβρίου 2025.

(Υπογραφή)

Ιωάννης Ψαρράς
Ομότιμος Καθηγητής Ε.Μ.Π.

(Υπογραφή)

Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

Ευάγγελος Μαρινάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025

Copyright © Αντώνιος Δημήτριος Αλικάρης, 2025.
Με επιφύλαξη παντός δικαιώματος. All rights reserved

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

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

.....
(Υπογραφή)

Αντώνιος Δημήτριος Αλικάρης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Τα δίκτυα ηλεκτρικής ενέργειας αποτελούν θεμελιώδη λίθο για την εύρυθμη λειτουργία και την ικανοποίηση των αναγκών της σύγχρονης κοινωνίας. Είτε για την παραγωγή ηλεκτρικής ενέργειας, με γεννήτριες ή με ανανεώσιμες πηγές ενέργειας, όπως τα φωτοβολταϊκά, είτε για την μεταφορά του ρεύματος προς τους αστικούς ιστούς, έχουν αποδείξει την σημασία τους και την χρηστικότητά τους. Όμως, τα συστήματα αυτά είναι ευάλωτα στα καιρικά φαινόμενα και ειδικά με την εμφάνιση της κλιματικής αλλαγής αυτά τα φαινόμενα γίνονται όλο και πιο ταχτικά, απρόβλεπτα και έντονα. Επομένως, είναι σημαντικό να μπορέσουν να δημιουργηθούν σενάρια ακραίων καιρικών φαινομένων, έτσι ώστε να υπάρχει καλύτερη εσοπτεία της επίδρασής τους στα δίκτυα ηλεκτρικής ενέργειας. Στην παρούσα εργασία, γίνεται ανάπτυξη ενός μοντέλου παραγωγής συνθετικών δεδομένων με σκοπό την παραγωγή σεναρίων που θα προσομοιάζουν δύο συνθήκες καύσωνα, ήπιο και έντονο. Το μοντέλο αυτό αναπτύσσεται μετεκπαιδευόντας ένα μεγάλο γλωσσικό μοντέλο, το Text-To-Text Transfer Transformer (T5) με δεδομένα ενός δικτύου ηλεκτρικής ενέργειας της εταιρίας ASM, στο Τέρενι της Ιταλίας. Για την μετεκπαίδευσή του, τα δεδομένα μετατρέπονται σε κατάλληλη μορφή κειμένου και χρησιμοποιώντας πληροφορίες καιρού, δηλαδή την θερμοκρασία, πληροφορίες μέρας και πληροφορίες έντασης καύσωνα, το μοντέλο καταλαβαίνει τις απαραίτητες πληροφορίες για την παραγωγή της εξόδου. Επίσης, εκπαιδεύτηκαν και άλλα δύο μοντέλα σύγκρισης, το TimeGAN, το οποίο βασίζεται στα Παραγωγικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks - GANs) με εξειδίκευση σε χρονοσειρές, και το GaussianCopula της Synthetic Data vault (SDV), το οποίο υπολογίζει τις εξόδους με χρήση στατιστικών κατανομών. Στην συνέχεια, τα τρία παραπάνω μοντέλα χρησιμοποιούνται για την παραγωγή σεναρίων καύσωνα. Αυτά τα σενάρια αξιολογούνται οπτικά βάσει τις ιστορικές και πραγματικές τιμές και στην συνέχεια συγκρίνονται τα τρία μοντέλα μεταξύ τους με την χρήση των μετρικών αξιολόγησης Dynamic Time Wrapping (DTW), Maximum Mean Discrepancy (MMD) και Pearson Correlation (PC). Τέλος, γίνεται συζήτηση για την πρακτική εφαρμογή αυτού του μοντέλου σε περιβάλλον προσομοίωσης του δικτύου ηλεκτρικής ενέργειας, με σκοπό τον υπολογισμό των τιμών τάσεων, με χρήση ενός προγράμματος υπολογισμού ροής ισχύος (power flow), για την αξιολόγηση της ανθεκτικότητάς του.

Λέξεις κλειδιά

Μεγάλα Γλωσσικά Μοντέλα (LLMs), Text-To-Text Transfer Transformer (T5), Παραγωγικά Αντιπαραθετικά Δίκτυα (GANs), TimeGAN, GaussianCopula, Μετεκπαίδευση Μοντέλων (fine-tuning), Παραγωγή Συνθετικών Δεδομένων Χρονοσειρών, Προσομοίωση Ροής Ισχύος, Στρατηγικές Ανθεκτικότητας Δικτύων

Abstract

Electrical power grids constitute a fundamental cornerstone for the smooth operation and fulfillment of the needs of modern society. Whether for the generation of electricity, through generators or renewable energy sources such as photovoltaics, or for the transmission of electricity to urban areas, they have proven their importance and usefulness. However, these systems are vulnerable to weather events, and especially with the emergence of climate change, such events are becoming increasingly frequent, unpredictable, and intense. Therefore, it is important to be able to construct scenarios of extreme weather events, in order to better monitor their impact on power grids. In the present work, a synthetic data generation model is developed with the purpose of producing scenarios that simulate two heatwave conditions: mild and severe. This model is built by fine-tuning a large language model, the Text-To-Text Transfer Transformer (T5), with data from the power grid of the company ASM in Terni, Italy. For its fine-tuning, the data are transformed into an appropriate textual format, and by incorporating weather information, mainly temperature, day information, and heatwave intensity, the model learns the necessary context for generating its outputs. In addition, two other comparison models were trained: TimeGAN, which is based on Generative Adversarial Networks (GANs) and specialized in time series, and GaussianCopula from the Synthetic Data Vault (SDV), which generates outputs using statistical distributions. Subsequently, the three models are employed to generate scenarios simulating the heatwave conditions. These scenarios are evaluated visually against historical and actual values, and the three models are then compared with each other using the evaluation metrics Dynamic Time Warping (DTW), Maximum Mean Discrepancy (MMD), and Pearson Correlation (PC). Finally, the practical application of this model is discussed within a simulation environment of the power grid, with the aim of calculating voltage values through the use of a power flow simulation, in order to assess the resilience of the grid.

Key words

Large Language Models (LLMs), Text-To-Text Transfer Transformer (T5), Generative Adversarial Networks (GANs), TimeGAN, Gaussian Copula, Model Fine-Tuning, Synthetic Time Series Data Generation, Power Flow Simulation, Grid Resilience Strategies

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Δημήτρη Ασκούνη για την δυνατότητα να ασχοληθώ με το αντικείμενο της παραγωγής ακραίων δεδομένων σε δίκτυα ηλεκτρικής ενέργειας, και την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου αυτό το θέμα. Επίσης, θα ήθελα να ευχαριστήσω θερμά τους υπεύθυνούς μου κ. Σωτήρη Πελέκη και, ειδικά, τον κ. Θεοδόση Πουντρίδη για την συνεχή καθοδήγηση και υποστήριξη σε κάθε βήμα της εργασίας. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και όλους τους φίλους μου, εντός και εκτός σχολής, για την διαρκή υποστήριξή τους όλα αυτά τα χρόνια, καθώς χωρίς αυτούς η εμπειρία της σχολής θα ήταν εντελώς διαφορετική.

Στον παππού μου Αντώνη

Περιεχόμενα

Λίστα Εικόνων	1
Λίστα Πινάκων	3
1 Εισαγωγή	5
1.1 Υπόβαθρο και Κίνητρο	5
1.2 Σκοπός και Συνεισφορά	6
1.3 Διάρθρωση Τόμου	7
2 Θεωρητικό Υπόβαθρο και Σχετική Έρευνα	9
2.1 Σύντομη εισαγωγή στα Συστήματα Ηλεκτρικής Ενέργειας	9
2.1.1 Τι είναι τα Συστήματα Ηλεκτρικής ενέργειας	9
2.1.1.1 Περιοχές λειτουργίας των συστημάτων ισχύος	9
2.1.2 Ανάλυση Ροής Ενέργειας (Power Flow Analysis)	10
2.1.2.1 Ο Αλγόριθμος Newton Raphson	11
2.1.3 Εκτίμηση Κατάστασης (State Estimation)	12
2.2 Ακραία καιρικά φαινόμενα και το αντίκτυπό τους στα Ηλεκτρικά Συστήματα	13
2.3 Αξιολόγηση της Ανθεκτικότητας και Προβλεπτική Συντήρηση	15
2.3.1 Αξιολόγηση της Ανθεκτικότητας (Resilience Assessment)	15
2.3.1.1 Ποιοτική Μέθοδος	16
2.3.1.2 Ποσοτική μέθοδος	16
2.3.1.3 Στρατηγικές Ανθεκτικότητας	16
2.3.2 Προβλεπτική Συντήρηση (Predictive Maintenance)	17
2.4 Παραγωγή συνθετικών δεδομένων	17
2.5 Μοντέλα που χρησιμοποιήθηκαν στην παρούσα εργασία	19
2.5.1 Μεγάλα Γλωσσικά Μοντέλα (LLMs)	19
2.5.1.1 Το Μοντέλο Text-to-Text Transfer Transformer (T5)	20
2.5.1.1.1 Προεκπαίδευση	21
2.5.1.1.2 Ενοποιημένη μεταφορά μάθησης	22
2.5.1.2 Μοντέλα GAN	22
2.5.1.2.1 Παραλλαγές GAN	23
2.5.1.2.2 Conditional GANs (cGANs)	23
2.5.1.2.3 TimeGAN	24
2.5.2 Γκαουσιανά Copulas (Gaussian Copulas)	25
2.6 Σχετική Έρευνα	26
2.6.1 Παραγωγή Σύνθετων Χρονοσειρών	26
2.6.2 Σενάρια ακραίων γεγονότων στα Συστήματα Ηλεκτρικής Ενέργειας	27
3 Μεθοδολογία και πειραματική διαδικασία	29
3.1 Εισαγωγή	29
3.2 Ανάλυση και προεπεξεργασία συνόλου δεδομένων	31
3.2.1 Δεδομένα ενέργειας	31
3.2.1.1 Ανάλυση δεδομένων	31
3.2.1.2 Προεπεξεργασία δεδομένων	34
3.2.1.2.1 Ο τρόπος διαχωρισμού στα υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής	35

3.2.1.2.2	Σύνολο Παραγωγής Δεδομένων	36
3.2.2	Καιρικά δεδομένα	37
3.3	Επιλογή Μοντέλων	38
3.4	Εκπαίδευση Μοντέλων	38
3.4.1	Μεγάλο Γλωσσικό Μοντέλο - LLM	38
3.4.1.1	Μετατροπή σε μορφή κειμένου	39
3.4.1.2	Εκπαίδευση Μοντέλου	40
3.4.2	TimeGAN Μοντέλο	42
3.4.3	Στατιστικό Μοντέλο Gaussian	43
3.5	Εξαγωγή (Inference) Αποτελεσμάτων και Αξιολόγηση Μοντέλων	43
3.5.1	Μετρικές αξιολόγησης	44
3.5.1.1	Dynamic Time Warping	44
3.5.1.2	Maximum Mean Discrepancy	44
3.5.1.3	Pearson Correlation	45
3.5.2	Εξαγωγή (Inference) Αποτελεσμάτων	45
3.5.2.1	Μεγάλο Γλωσσικό Μοντέλο	45
3.5.2.2	Μοντέλο TimeGAN	46
3.5.2.3	Μοντέλο Gaussian	46
3.5.3	Αξιολόγηση Μοντέλων	46
4	Αποτελέσματα Πειραμάτων	49
4.1	Αποτελέσματα του LLM	49
4.2	Αποτελέσματα του TimeGAN μοντέλου	52
4.3	Αποτελέσματα του Gaussian μοντέλου	55
4.4	Αποτελέσματα της αξιολόγησης των μοντέλων	57
5	Συζήτηση	63
5.1	Μεθοδολογία	63
5.2	Αποτελέσματα	64
6	Συμπεράσματα και Μελλοντικές Προεκτάσεις	69
6.1	Συμπεράσματα και Επίλογος	69
6.2	Μελλοντικές προεκτάσεις	70
	Βιβλιογραφία	71
	Παραρτήματα	79
A	Αποτελέσματα προσομοίωσης DPsim όλων των περιπτώσεων	79
B	Αποτελέσματα προσομοίωσης DPsim όλων των περιπτώσεων στην δο- κιμή διπλασιασμού του πληθυσμού	84

Λίστα Εικόνων

2.1	Generative Adversarial Networks (GANs)	23
2.2	Conditional GANs (cGANs)	24
2.3	TimeGAN Architecture	25
3.1	Διάγραμμα κύκλου μηχανικής μάθησης για το LLM Μοντέλο	30
3.2	Τοπολογία Δικτύου	31
3.3	Ενεργός Ισχύς του μετρητή 4	32
3.4	Άεργος Ισχύς του μετρητή 4	32
3.5	Ενεργός Ισχύς του μετρητή w4 μετά την αντικατάσταση άγνωστων τιμών	34
3.6	Άεργος Ισχύς του μετρητή w4 μετά την αντικατάσταση άγνωστων τιμών	34
3.7	Σύγκριση των ειδών schedulers του ρυθμού μάθησης	41
3.8	Σύγκριση DTW με Ευκλίδεια Απόσταση	44
4.1	Αποτελέσματα LLM Inference Ενεργού Ισχύος	50
4.2	Αποτελέσματα LLM Inference Άεργου Ισχύος	51
4.3	Αποτελέσματα TimeGAN Inference Ενεργού Ισχύος	53
4.4	Αποτελέσματα TimeGAN Inference Άεργου Ισχύος	54
4.5	Αποτελέσματα Gaussian μοντέλου Ενεργού Ισχύος	55
4.6	Αποτελέσματα Gaussian μοντέλου Άεργου Ισχύος	56
4.7	Τιμές DTW μετρικής Ενεργού Ισχύος	57
4.8	Τιμές DTW μετρικής Άεργου Ισχύος	58
4.9	Τιμές MMD μετρικής Ενεργού Ισχύος	58
4.10	Τιμές MMD μετρικής Άεργου Ισχύος	59
4.11	Τιμές PC μετρικής Ενεργού Ισχύος	59
4.12	Τιμές PC μετρικής Άεργου Ισχύος	60
4.13	Όλες οι μετρικές Ενεργού Ισχύος σε μορφή Heatmap	61
4.14	Όλες οι μετρικές Άεργου Ισχύος σε μορφή Heatmap	61
5.1	High Level Διάγραμμα της συνολικής διαδικασίας	63
5.2	Αποτελέσματα προσομοίωσης περίπτωσης Moderate Weekday	65
5.3	Αποτελέσματα προσομοίωσης Moderate Weekday σε περίπτωση διπλασιασμού πληθυσμού	67
A.1	Αποτελέσματα προσομοίωσης περίπτωσης Moderate Saturday	79
A.2	Αποτελέσματα προσομοίωσης περίπτωσης Moderate Sunday	80
A.3	Αποτελέσματα προσομοίωσης περίπτωσης Severe Weekday	81
A.4	Αποτελέσματα προσομοίωσης περίπτωσης Severe Saturday	82
A.5	Αποτελέσματα προσομοίωσης περίπτωσης Severe Sunday	83
B.1	Αποτελέσματα προσομοίωσης περίπτωσης Moderate Saturday	84
B.2	Αποτελέσματα προσομοίωσης περίπτωσης Moderate Sunday	85
B.3	Αποτελέσματα προσομοίωσης περίπτωσης Severe Weekday	86
B.4	Αποτελέσματα προσομοίωσης περίπτωσης Severe Saturday	87
B.5	Αποτελέσματα προσομοίωσης περίπτωσης Severe Sunday	88

Λίστα Πινάκων

3.1	Κατανομή δεδομένων ανά συνθήκη και ημέρα της εβδομάδας.	35
3.2	Κατανομή δεδομένων στα υποσύνολα εκπαίδευσης, επικύρωσης και δο- κιμής.	36
3.3	Κατανομή δεδομένων καύσωνα ανά υποσύνολο, ημέρα και ένταση. . . .	36
3.4	Κατανομή τελικών σεναρίων του συνόλου παραγωγής δεδομένων. . . .	37
4.1	Αποτελέσματα συνολικών μετρικών Ενεργού Ισχύος.	60
4.2	Αποτελέσματα συνολικών μετρικών Άεργου Ισχύος.	60

Κεφάλαιο 1

Εισαγωγή

Η κλιματική αλλαγή έχει προκαλέσει πολλές επιπτώσεις στο κλίμα της Γης. Μία από αυτές είναι η αύξηση του πλήθους και της έντασης των ακραίων καιρικών φαινομένων, όπως καύσωνες και θύελλες. Αυτές οι έντονες αλλαγές προκαλούν πολλά προβλήματα στις τοπικές κοινωνίες, στις βιομηχανίες και σε πολλούς άλλους τομείς, ένας από αυτούς είναι τα δίκτυα ηλεκτρικής ενέργειας, αφού έρχονται αντιμέτωπα σε όλο και περισσότερες και πιο έντονες ακραίες καιρικές συνθήκες. Με την προσομοίωση τέτοιων συνθηκών θα υπάρχει καλύτερη εποπτεία στο δίκτυο και μεγαλύτερη βοήθεια στην αποφυγή των βλαβών. Στην εν λόγω εργασία θα ασχοληθούμε με την παραγωγή συνθετικών δεδομένων που θα προσομοιάζουν χρονοσειρές ενός δικτύου ηλεκτρικής ενέργειας υπό ακραίες καιρικές συνθήκες, και συγκεκριμένα συνθήκες καύσωνα.

1.1 Υπόβραθρο και Κίνητρο

Η παραγωγή συνθετικών χρονοσειρών αποτελεί κρίσιμο εργαλείο για την ανάλυση και την αξιολόγηση συστημάτων, ιδίως σε περιπτώσεις όπου η συλλογή πραγματικών δεδομένων είναι πολυδάπανη, δύσκολη ή περιορισμένη, όπως συμβαίνει συχνά στα δίκτυα ηλεκτρικής ενέργειας. Η σχετική βιβλιογραφία έχει εστιάσει κυρίως στην δημιουργία χρονοσειρών σε περιπτώσεις επαύξησης ενός συνόλου δεδομένων (data augmentation) σε προβλήματα ταξινόμησης ή, πολύ συχνότερα, στην ανάπτυξη μοντέλων πρόβλεψης σε τομείς όπως η υγεία, ο καιρός και τα οικονομικά.

Τα περισσότερα μοντέλα που χρησιμοποιούνται για αυτόν τον σκοπό είναι είτε τα στατιστικά, όπου η παραγωγή των νέων χρονοσειρών βασίζεται στα ιστορικά χαρακτηριστικά και τις περιοδικότητες των χρονοσειρών, είτε τα πιο σύγχρονα παραγωγικά μοντέλα όπως τα GANs (π.χ. TimeGAN) ή οι μεταβλητοί αυτοκωδικοποιητές (VAEs).

Παράλληλα, τα Μεγάλα Γλωσσικά Μοντέλα (LLMs) έχουν δείξει σημαντικές δυνατότητες στη μοντελοποίηση ακολουθιών και στη δημιουργία νέων δεδομένων σε πλήθος εφαρμογών. Ωστόσο, η αξιοποίησή τους για τη δημιουργία συνθετικών χρονοσειρών παραμένει περιορισμένη.

Ταυτόχρονα, τα τελευταία χρόνια έχει σημειωθεί σημαντική πρόοδος στην ανάπτυξη και ενσωμάτωση των Ανανεώσιμων Πηγών Ενέργειας (ΑΠΕ) στα δίκτυα ηλεκτρικής ενέργειας, μέσω εγκαταστάσεων όπως τα φωτοβολταϊκά και τα αιολικά πάρκα. Ωστόσο, και αυτές οι μορφές παραγωγής επηρεάζονται από τις ακραίες καιρικές συνθήκες, γεγονός που μπορεί να δημιουργήσει προκλήσεις για τη σταθερότητα και την αξιοπιστία του δικτύου. Για παράδειγμα, τα θαλάσσια αιολικά πάρκα είναι εκτεθειμένα σε κινδύνους που σχετίζονται με το θαλάσσιο περιβάλλον, ενώ η παραγωγή από φωτοβολταϊκά συστήματα μειώνεται σημαντικά σε περιόδους περιορισμένης ηλιοφάνειας ή δυσμενών καιρικών συνθηκών.

Η αυξανόμενη συχνότητα τέτοιων φαινομένων έχει ήδη προκαλέσει σοβαρά περιστατικά αστάθειας ή διακοπών (blackouts) στα δίκτυα ηλεκτρικής ενέργειας, καθώς μέχρι και το 33% των συνολικών διακοπών λειτουργίας οφείλεται σε κλιματικούς κινδύνους σύμφωνα με την Ευρωπαϊκή Επιτροπή [1]. Αυτά τα γεγονότα δείχνουν την ανάγκη για εργαλεία που θα επιτρέπουν την προσομοίωση και ανάλυση της συμπεριφοράς

των δικτύων υπό διαφορετικές καιρικές συνθήκες, ώστε να κατανοηθούν καλύτερα οι μηχανισμοί ευπάθειας και ανθεκτικότητας των συστημάτων.

Έτσι, η ανάπτυξη μεθόδων που μπορούν να προσομοιώνουν τη λειτουργία των ηλεκτρικών δικτύων υπό ακραίες συνθήκες είναι πλέον κρίσιμη. Μέσω τέτοιων προσεγγίσεων μπορεί να επιτευχθεί μια πιο ολοκληρωμένη κατανόηση της λειτουργίας τους, διευκολύνοντας την έγκαιρη πρόληψη και τον περιορισμό επιζήμιων επιπτώσεων. Παράλληλα, μπορούν να υποστηρίξουν τη λήψη αποφάσεων για την εφαρμογή κατάλληλων στρατηγικών ενίσχυσης (grid hardening), εκεί όπου εντοπίζονται αδυναμίες ή πιθανοί κίνδυνοι.

Στο πλαίσιο αυτό, η παρούσα εργασία επιχειρεί να συμβάλει στην κατεύθυνση αυτή, προτείνοντας την ανάπτυξη ενός νέου μοντέλου παραγωγής δεδομένων χρονοσειρών προσομοιάζοντας διαφορετικά σενάρια ακραίων καιρικών συνθηκών

1.2 Σκοπός και Συνεισφορά

Ο σκοπός της παρούσας εργασίας είναι η ανάπτυξη ενός μοντέλου, μετεκπαιδευμένου από ένα γλωσσικό μοντέλο, το οποίο θα μπορεί να δημιουργήσει ρεαλιστικά σενάρια χρονοσειρών, ενός δικτύου ηλεκτρικής ενέργειας, που αντικατοπτρίζουν τη συμπεριφορά των δικτύων υπό διαφορετικές καιρικές συνθήκες, έτσι ώστε να συμβάλει στην καλύτερη κατανόηση και βελτίωση των δικτύων ηλεκτρικής ενέργειας.

Το μοντέλο στοχεύει να διαθέτει τα εξής χαρακτηριστικά:

- Να δέχεται στην είσοδο τα προηγούμενα δεδομένα ενέργειας (ισχύος) και καιρού, έτσι ώστε οι παραγόμενες χρονοσειρές να βασίζονται σε ρεαλιστικά μοτίβα και συσχετίσεις.
- Να λαμβάνει υπόψη τις ημέρες της εβδομάδας που αντιστοιχούν στα ιστορικά δεδομένα, ώστε να αναγνωρίζει τις περιοδικότητες και τις διαφοροποιήσεις στη συμπεριφορά του φορτίου μεταξύ εργάσιμων ημερών, Σαββάτου και Κυριακής.
- Να δέχεται ως είσοδο την ένταση των καιρικών συνθηκών ως μελλοντική συμμεταβλητή (future covariate), επιτρέποντας έτσι τη στοχευμένη παραγωγή χρονοσειρών που ανταποκρίνονται σε συγκεκριμένα σενάρια καιρικής έντασης.

Με αυτόν τον τρόπο, το μοντέλο θα είναι σε θέση να παράγει συνθετικές χρονοσειρές που αντικατοπτρίζουν ρεαλιστικά την κατανάλωση ή την παραγωγή ενέργειας υπό διαφορετικές συνθήκες λειτουργίας του δικτύου.

Επίσης, η εργασία στοχεύει να συμβάλει στην ερευνητική κοινότητα με την ανάπτυξη ενός υπό συνθήκη (conditional) μοντέλου παραγωγής χρονοσειρών που βασίζεται σε Μεγάλα Γλωσσικά Μοντέλα (LLMs), μια προσέγγιση που παραμένει ελάχιστα διερευνημένη.

Τα παραγόμενα σενάρια μπορούν να αξιοποιηθούν για την επαύξηση των συνόλων δεδομένων (data augmentation), προσφέροντας περισσότερα δείγματα σπάνιων ή ακραίων καταστάσεων. Έτσι, θα μπορούν να βελτιώσουν τα αποτελέσματα μελετών και ερευνών που απαιτούν πληρέστερα και πιο αντιπροσωπευτικά σύνολα δεδομένων με ποικιλία σεναρίων ακραίων καιρικών συνθηκών. Παράλληλα, θα μπορέσουν να συμβάλλουν στα μοντέλα πρόβλεψης (forecasting) παρέχοντας περισσότερα δεδομένα ασυνήθιστων συνθηκών, οι οποίες διαφορετικά θα είχαν περιορισμένη επίδραση λόγω της μικρής συχνότητάς τους.

Τέλος, μπορούν να χρησιμοποιηθούν σε διαδικασίες αξιολόγησης της ανθεκτικότητας του δικτύου (resilience assessment) και ελέγχου της ποιότητας των υπηρεσιών (quality of service), παρέχοντας συνθετικά σενάρια λειτουργίας υπό διαφορετικές συνθήκες, βοηθώντας τους ειδικούς να εξετάσουν πως ανταποκρίνεται το δίκτυο σε συνθήκες, όπως υψηλή ζήτηση ή μειωμένη παραγωγή, και να δουν σε ποια σημεία του δικτύου χρειάζεται να εστιάσουν για την αποφυγή βλάβης.

1.3 Διάρθρωση Τόμου

Στο Κεφάλαιο 1 γίνεται εισαγωγή στην παραγωγή σύνθετων χρονοσειρών, καθώς και γίνεται παρουσίαση του προβλήματος και του σκοπού της παρούσας έρευνας.

Στο Κεφάλαιο 2 αναλύεται το θεωρητικό υπόβαθρο των δικτύων ηλεκτρικής ενέργειας, των στρατηγικών ανθεκτικότητας ενός δικτύου καθώς και των προβλημάτων που έχουν υποστεί διάφορα δίκτυα λόγω των ακραίων καιρικών φαινομένων. Επιπλέον, γίνεται θεωρητική εισαγωγή στα μοντέλα που χρησιμοποιήθηκαν και στην σχετική έρευνα.

Στο Κεφάλαιο 3 παρουσιάζεται αναλυτικά η πειραματική διαδικασία που ακολουθήθηκε στην παρούσα εργασία, δηλαδή η μεθοδολογία εκπαίδευσης και μετεκπαίδευσης όλων των μοντέλων που χρησιμοποιήθηκαν, καθώς και ο τρόπος παραγωγής των αποτελεσμάτων.

Στο Κεφάλαιο 4 παρουσιάζονται τα αποτελέσματα των πειραμάτων και συγκρίνονται τα μοντέλα με τρεις μετρικές αξιολόγησης.

Στο κεφάλαιο 5 γίνεται συζήτηση για μια πρακτική εφαρμογή του μοντέλου παραγωγής συνθετικών δεδομένων πάνω σε ένα συγκεκριμένο δίκτυο ηλεκτρικής ενέργειας, με σκοπό την εποπτεία του και παρατήρηση πιθανών βλαβών.

Στο Κεφάλαιο 6 αναλύονται τα συμπεράσματα και οι μελλοντικές προεκτάσεις της εργασίας.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο και Σχετική Έρευνα

Σε αυτό το κεφάλαιο γίνεται ανάλυση των εννοιών που χρειάστηκαν στην παρούσα εργασία. Συγκεκριμένα, γίνεται αναφορά στα συστήματα ηλεκτρικής ενέργειας, στις βλάβες που έχουν υποστεί από ακραία καιρικά φαινόμενα και στους τρόπους που μπορούν να γίνουν πιο ανθεκτικά σε αυτά. Επίσης, γίνεται αναφορά στην παραγωγή δεδομένων, στους τρόπους και τα μοντέλα που έχουν χρησιμοποιηθεί γενικά, αλλά και αναλύονται τα μοντέλα που χρησιμοποιήθηκαν στην παρούσα εργασία. Τέλος, αναφέρονται οι σχετικές έρευνες.

2.1 Σύντομη εισαγωγή στα Συστήματα Ηλεκτρικής Ενέργειας

Στην παρούσα ενότητα γίνεται μία σύντομη αναφορά στα συστήματα ηλεκτρικής ενέργειας και σε μεθόδους που σχετίζονται με την βελτίωση και αποκατάστασή τους.

2.1.1 Τι είναι τα Συστήματα Ηλεκτρικής ενέργειας

Τα ηλεκτρικά συστήματα αποτελούνται από τα συστήματα μεταφοράς, υπο-μεταφοράς, διανομής και παραγωγής. Τα συστήματα μεταφοράς μπορεί να περιλαμβάνουν μεγάλο αριθμό υποσταθμών, οι οποίοι διασυνδέονται μέσω γραμμών μεταφοράς, μετασχηματιστών και άλλων διατάξεων για έλεγχο και προστασία του συστήματος. Η ηλεκτρική ισχύς εισέρχεται στο σύστημα από τις γεννήτριες ή απορροφάται από αυτό μέσω των φορτίων στους εν λόγω υποσταθμούς. Οι τάσεις εξόδου των γεννητριών συνήθως δεν υπερβαίνουν τα 30kV. Συνεπώς, χρησιμοποιούνται μετασχηματιστές για την αύξηση της τάσης σε επίπεδα που κυμαίνονται από 69kV έως και 765kV στους ακροδέκτες των γεννητριών. Η χρήση υψηλής τάσης στο δίκτυο μεταφοράς είναι προτιμητέα για διάφορους λόγους, ένας εκ των οποίων είναι η ελαχιστοποίηση των απωλειών χαλκού, οι οποίες είναι ανάλογες του ρεύματος που διαρρέει τις γραμμές. Στο άλλο άκρο (άκρο υποδοχής), τα συστήματα μεταφοράς συνδέονται με τα συστήματα υπο-μεταφοράς ή διανομής, τα οποία λειτουργούν σε χαμηλότερα επίπεδα τάσης, από 115kV έως 4,16kV. Τα συστήματα διανομής διαμορφώνονται συνήθως σε ακτινική διάταξη (radial configuration), όπου οι γραμμές τροφοδοσίας (feeders) εκτείνονται από τους υποσταθμούς διανομής και σχηματίζουν μια δενδρική δομή, με τη ρίζα στον υποσταθμό και τα κλαδιά να εκτείνονται στην περιοχή διανομής [2].

2.1.1.1 Περιοχές λειτουργίας των συστημάτων ισχύος

Οι συνθήκες λειτουργίας ενός συστήματος ηλεκτρικής ενέργειας σε δεδομένη χρονική στιγμή μπορούν να προσδιοριστούν εφόσον είναι γνωστά το μοντέλο του δικτύου και οι σύνθετοι φασικοί τελεστές τάσης (phasors) σε κάθε ζυγό του συστήματος. Δεδομένου ότι το σύνολο των φασικών τελεστών τάσης περιγράφει πλήρως την κατάσταση του συστήματος, αυτό αναφέρεται ως η στατική κατάσταση του συστήματος. Το σύστημα μπορεί να βρεθεί σε μία από τις τρεις βασικές καταστάσεις: φυσιολογική, έκτακτης ανάγκης και αποκατάστασης [2].

Ένα σύστημα θεωρείται ότι βρίσκεται σε φυσιολογική κατάσταση όταν σε όλα τα φορτία του μπορεί να ασκηθεί ηλεκτρική ισχύς από τις γεννήτριες χωρίς να παραβιαστούν οι περιορισμοί λειτουργίας. Οι περιορισμοί αυτοί περιλαμβάνουν, μεταξύ άλλων, τα όρια ροής ισχύος στις γραμμές μεταφοράς, καθώς και τα ανώτατα και κατώτατα όρια των τιμών τάσης στους ζυγούς. Μια φυσιολογική κατάσταση χαρακτηρίζεται ως ασφαλής, όταν το σύστημα είναι σε θέση να παραμείνει σε αυτή την κατάσταση μετά την εμφάνιση οποιασδήποτε ενδεχόμενης διαταραχής από έναν προκαθορισμένο κατάλογο κρίσιμων περιστατικών. Συνήθη περιστατικά περιλαμβάνουν βλάβες σε γραμμές μεταφοράς ή στις μονάδες παραγωγής λόγω απρόβλεπτων βλαβών στον εξοπλισμό ή λόγω φυσικών φαινομένων, όπως καταιγίδες. Αντίθετα, η φυσιολογική κατάσταση χαρακτηρίζεται ως μη ασφαλής, όπου οι ισορροπίες ισχύος και οι περιορισμοί λειτουργίας ικανοποιούνται σε όλους στους ζυγούς, αλλά το σύστημα παραμένει ευάλωτο σε ορισμένα από τα κρίσιμα περιστατικά. Τότε, απαιτείται η λήψη προληπτικών μέτρων ώστε να αποφευχθεί η μετάβαση σε κατάσταση έκτακτης ανάγκης. Τα μέτρα αυτά μπορούν συνήθως να καθοριστούν μέσω ενός προγράμματος βέλτιστης κατανομής ισχύος με περιορισμούς ασφάλειας (security-constrained optimal power flow program), το οποίο λαμβάνει υπόψη τον κατάλογο των κρίσιμων περιστατικών [2].

Οι συνθήκες λειτουργίας του συστήματος ενδέχεται να αλλάξουν δραστικά λόγω ενός απρόβλεπτου γεγονότος που μπορεί να προκαλέσει την παραβίαση κάποιων περιορισμών λειτουργίας, ενώ το σύστημα θα εξακολουθεί να παρέχει ηλεκτρική ισχύ σε όλα τα φορτία. Σε αυτή την περίπτωση, το σύστημα θεωρείται ότι λειτουργεί σε κατάσταση έκτακτης ανάγκης. Η κατάσταση αυτή απαιτεί άμεση λήψη διορθωτικών ενεργειών από τον χειριστή, ώστε να επανέλθει το σύστημα σε φυσιολογική κατάσταση.

Κατά τη διάρκεια της κατάστασης έκτακτης ανάγκης, είναι δυνατόν με κατάλληλα διορθωτικά μέτρα ελέγχου να αποφευχθεί η κατάρρευση του συστήματος, ακόμη και αν απαιτηθεί η αποσύνδεση φορτίων, γραμμών, μετασχηματιστών ή άλλου εξοπλισμού. Ως αποτέλεσμα, μπορεί να εξαλειφθούν οι παραβιάσεις των ορίων λειτουργίας και το σύστημα να ανακτήσει τη σταθερότητά του με μειωμένο φορτίο και αναδιαμορφωμένη τοπολογία. Στη συνέχεια, απαιτείται η αποκατάσταση της ισορροπίας μεταξύ φορτίου και παραγωγής, ώστε να αρχίσει η τροφοδότηση όλων των φορτίων. Η κατάσταση αυτή ονομάζεται κατάσταση αποκατάστασης και οι ενέργειες που αποσκοπούν στην επαναφορά του συστήματος σε φυσιολογική κατάσταση χαρακτηρίζονται ως ενέργειες αποκατάστασης [2].

2.1.2 Ανάλυση Ροής Ενέργειας (Power Flow Analysis)

Ο υπολογισμός ροής ενέργειας (Power Flow – PF) αποτελεί τη βάση για την ανάλυση της σταθερής μόνιμης κατάστασης ενός συστήματος ηλεκτρικής ενέργειας και είναι ο θεμέλιος λίθος για πολλές εφαρμογές στο πλαίσιο της λειτουργίας και του σχεδιασμού των συστημάτων ισχύος. Οι υπολογισμένες τιμές τάσης και ισχύος μέσω της ανάλυσης ροής ενέργειας είναι απαραίτητες για τον προσδιορισμό της κατάστασης του συστήματος και τη διασφάλιση της ασφάλειας και της ευστάθειας του δικτύου [3].

Η ανάλυση ροής ισχύος αποσκοπεί στον προσδιορισμό των μεγεθών τάσης και των γωνιών φάσης στους ζυγούς φορτίου, της ενεργού και άεργης ισχύος που ρέει μέσω των γραμμών μεταφοράς, καθώς και των γωνιών τάσης και της έγχυσης άεργης ισχύος στους ζυγούς παραγωγής. Οι πληροφορίες αυτές είναι απαραίτητες για την αποτύπωση της σταθερής μόνιμης κατάστασης του δικτύου και τη διασφάλιση της

ασφάλειας και της ευστάθειας του συστήματος. Ωστόσο, η ανάλυση ροής ισχύος είναι ένα μη γραμμικό και υπολογιστικά απαιτητικό πρόβλημα, του οποίου η λύση έχει νόημα μόνο για περιορισμένο χρονικό διάστημα, καθώς η κατάσταση του συστήματος ισχύος μεταβάλλεται συνεχώς [3].

2.1.2.1 Ο Αλγόριθμος Newton Raphson

Ο αλγόριθμος Newton Raphson είναι ένα συνηθισμένο και πολύ βοηθητικό εργαλείο στην επίλυση των προβλημάτων ροής ενέργειας και αυτό γιατί μπορεί να λύσει επαναληπτικά τις μη γραμμικές εξισώσεις των συστημάτων ισχύος και να βρει ένα αποτέλεσμα με όση ακρίβεια θελήσουμε. Παρακάτω φαίνεται με βήματα πως λειτουργεί σε ένα σύστημα ενέργειας.

Το πρόβλημα ροής ενέργειας αποτελεί πρόβλημα εύρεσης μηδενικών για τον προσδιορισμό των τάσεων που ικανοποιούν ένα σύνολο μη γραμμικών εξισώσεων ανισοροπίας ισχύος. Εναλλακτικά, μπορεί να χρησιμοποιηθεί ένα σύνολο μη γραμμικών εξισώσεων ανισοροπίας ρευμάτων ως συναρτήσεις των μεταβλητών κατάστασης.

Έστω ένα σύστημα n ζυγών, όπου ο ζυγός 1 θεωρείται ζυγός αναφοράς (slack bus), με σταθερό μέτρο τάσης και μηδενική γωνία φάσης. Για τον k -οστό ζυγό, η εξίσωση ισοροπίας ρεύματος δίνεται ως εξής [4]:

$$(I_{gen,k} - I_{dem,k}) - \sum_{i=1}^n Y_{ki} V_i = 0 \quad (2.1)$$

όπου $I_{gen,k}$ είναι το ρεύμα γεννήτριας στον ζυγό k , $I_{dem,k}$ είναι το ρεύμα φορτίου, V_k είναι η φασική τάση του ζυγού, και Y_{ki} είναι το στοιχείο της k -οστής γραμμής και i -οστής στήλης του πίνακα αγωγιμότητας.

Δεδομένου ότι τα φορτία εκφράζονται συνήθως ως ισχύς, η εξίσωση 2.1 μπορεί να επαναδιατυπωθεί ως εξής:

$$F_k = \left(\frac{S_{gen,k} - S_{dem,k}}{V_k} \right)^* - \sum_{i=1}^n Y_{ki} V_i = 0 \quad (2.2)$$

Ορίζουμε: $F_k = G_k + jH_k$ ως την ανισοροπία ρεύματος στον ζυγό k ,

$V_k = |V_k| \angle \delta_k$,

$Y_{ki} = |Y_{ki}| \angle \theta_{ki}$,

και $S_{sch,k} = |S_{sch,k}| \angle \phi_k = S_{gen,k} - S_{dem,k}$.

Αντικαθιστώντας στην εξίσωση 2.2 προκύπτει:

$$F_k = \left| \frac{S_{sch,k}}{V_k} \right| \angle (-\phi_k + \delta_k) - \sum_{i=1}^n |Y_{ki} V_i| \angle (\theta_{ki} + \delta_i) = 0 \quad (2.3)$$

Από την οποία έχουμε το πραγματικό και φανταστικό μέρος, αντίστοιχα:

$$P_k = \left| \frac{S_{sch,k}}{V_k} \right| \cos(-\phi_k + \delta_k) - \sum_{i=1}^n |Y_{ki} V_i| \cos(\theta_{ki} + \delta_i) \quad (2.4)$$

$$Q_k = \left| \frac{S_{sch,k}}{V_k} \right| \sin(-\phi_k + \delta_k) - \sum_{i=1}^n |Y_{ki} V_i| \sin(\theta_{ki} + \delta_i) \quad (2.5)$$

Η ανισοροπία μηδενίζεται όταν όλες οι άγνωστες τάσεις στους ζυγούς προσδιοριστούν σωστά. Για την εύρεση των λύσεων μέσω της μεθόδου Newton-Raphson, οι εξισώσεις αναπτύσσονται σε σειρά Taylor:

$$P_k = \sum_{\substack{i=1 \\ i \neq s}}^n \frac{\partial P_k}{\partial \delta_i} \Delta \delta_i + \sum_{\substack{i=1 \\ i \neq s}}^n \frac{\partial P_k}{\partial |V_i|} \Delta |V_i| \quad (2.6)$$

$$Q_k = \sum_{\substack{i=1 \\ i \neq s}}^n \frac{\partial Q_k}{\partial \delta_i} \Delta \delta_i + \sum_{\substack{i=1 \\ i \neq s}}^n \frac{\partial Q_k}{\partial |V_i|} \Delta |V_i| \quad (2.7)$$

όπου s είναι ο ζυγός αναφοράς (slack bus).

Με $n - 1$ άγνωστες μεταβλητές και $n - 1$ εξισώσεις ανισορροπίας, η συμπαγής μορφή πίνακα για την ενημέρωση της λύσης της τάσης δίνεται από:

$$\begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{P}}{\partial \delta} & \frac{\partial \mathbf{P}}{\partial |\mathbf{V}|} \\ \frac{\partial \mathbf{Q}}{\partial \delta} & \frac{\partial \mathbf{Q}}{\partial |\mathbf{V}|} \end{bmatrix} \begin{bmatrix} \Delta \delta \\ \Delta |\mathbf{V}| \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{J}_2 \\ \mathbf{J}_3 & \mathbf{J}_4 \end{bmatrix} \begin{bmatrix} \Delta \delta \\ \Delta |\mathbf{V}| \end{bmatrix} \quad (2.8)$$

Στη μέθοδο Newton–Raphson, η εξίσωση 2.8 επιλύεται επαναληπτικά για τα $\Delta \delta$ και $\Delta |V|$. Ο αλγόριθμος ξεκινάει με μία αρχική πρόβλεψη της τιμής και αν η καθορισμένη νόρμα των ανισορροπιών $\mathbf{P}$ και $\mathbf{Q}$ είναι μικρότερη από το μέγιστο επιτρεπτό όριο ανισορροπίας (κατώφλι ϵ), τότε η λύση των τάσεων θεωρείται ότι έχει επιτευχθεί επιτυχώς. Διαφορετικά, η τρέχουσα λύση τάσης κατά την επανάληψη h πρέπει να ενημερωθεί για την επόμενη επανάληψη $h + 1$ όπως φαίνεται στην επόμενη εξίσωση.

$$\begin{bmatrix} \delta \\ |V| \end{bmatrix}^{(h+1)} = \begin{bmatrix} \delta \\ |V| \end{bmatrix}^{(h)} + \begin{bmatrix} \Delta \delta \\ \Delta |V| \end{bmatrix}^{(h)} \quad (2.9)$$

2.1.3 Εκτίμηση Κατάστασης (State Estimation)

Η εκτίμηση κατάστασης αποτελεί σημαντικό μέρος για την ασφάλεια των συστημάτων ηλεκτρικής ενέργειας. Η λειτουργία του συστήματος ελέγχεται από το Κέντρο Ελέγχου Φορτίου (Load Dispatch Center) με την προϋπόθεση ότι είναι διαθέσιμες διάφορες πληροφορίες σχετικά με την τρέχουσα κατάσταση του συστήματος. Αυτές οι πληροφορίες περιλαμβάνουν μετρήσεις από όργανα μέτρησης, τη θέση των μετασχηματιστών, τη θέση των διακοπών ισχύος και την τοπολογία του δικτύου. Ωστόσο, η μετάδοση αυτών των πληροφοριών προς το κέντρο SCADA δεν είναι πάντα αξιόπιστη. Συγκεκριμένα, τα σφάλματα μπορεί να προκληθούν από λανθασμένη σύνδεση του μετατροπέα, από απώλεια δεδομένων κατά τη μετάδοση ή από την παρουσία ελαττωματικών οργάνων μέτρησης. Επομένως, εάν αυτά τα λανθασμένα δεδομένα χρησιμοποιηθούν για την Ανάλυση Ενδεχομένων (Contingency Analysis) στο σύστημα, μπορεί να προκληθούν εσφαλμένοι συναγερμοί. Για τον λόγο αυτό, είναι απαραίτητο να ελέγχεται η ακρίβεια των μετρήσεων και να εξάγονται οι σωστές τιμές ακόμη και σε περιπτώσεις εσφαλμένων ή ελλιπών δεδομένων, μέσω της εκτίμησης κατάστασης. Η εκτίμηση κατάστασης του συστήματος ηλεκτρικής ενέργειας, στη βασική της μορφή, βασίζεται αποκλειστικά σε μετρήσεις SCADA. Οι Μονάδες Μέτρησης Φασικών Μεγεθών (Phasor Measurement Units – PMU), οι οποίες μπορούν να μετρήσουν την κατάσταση του συστήματος, μπορούν να αυξήσουν σημαντικά την ακρίβεια της τεχνικής εκτίμησης κατάστασης. Επειδή όμως οι PMU έχουν υψηλό κόστος και δεν μπορούν να εγκατασταθούν πρακτικά σε όλους τους ζυγούς, είναι απαραίτητο να αναπτυχθεί ένας αλγόριθμος που να συνδυάζει τις τεχνικές εκτίμησης κατάστασης βασισμένες τόσο στο SCADA όσο και στο PMU [5].

Οι τεχνικές εκτίμησης κατάστασης αποτελούν έναν κρίσιμο συνδυαστικό κρίκο μεταξύ των συσκευών μέτρησης και των κέντρων ελέγχου, επιτρέποντας την παρακολούθηση και απεικόνιση όλων των ηλεκτρικών μεγεθών στα Ενεργά Δίκτυα Διανομής (Active Distribution Networks – ADNs)[6].

Θα ασχοληθούμε με τα συστήματα διανομής, που είναι τα συστήματα όπου έχουμε μικρές τιμές τάσης και είναι τα συστήματα στα εργοστάσια/βιομηχανίες. Με τη γνώση της τοπολογίας του δικτύου και των φασικών τελεστών τάσης, μπορούν να εκτιμηθούν όλες οι μεταβλητές. Συνεπώς, οι ελάχιστες μεταβλητές που μπορούν να περιγράψουν πλήρως ένα σύστημα ισχύος είναι το μέτρο της τάσης και η γωνία τάσης. Αυτές οι μεταβλητές αναφέρονται ως η στατική κατάσταση του συστήματος.

Ανάλογα με τις συνθήκες λειτουργίας, το σύστημα ισχύος μπορεί να βρίσκεται σε μία από τις πέντε καταστάσεις λειτουργίας [7]: Στην κανονική κατάσταση λειτουργίας ικανοποιούνται όλοι οι περιορισμοί ισότητας και ανισότητας. Όταν μειώνονται τα περιστρεφόμενα αποθέματα (spinning reserves) του συστήματος, αυτό μεταβαίνει σε κατάσταση συναγερμού (alert state). Όταν παραβιάζονται τα όρια των ανισοτήτων, το σύστημα εισέρχεται σε κατάσταση έκτακτης ανάγκης (emergency state). Επιπλέον, όταν παραβιάζονται και οι περιορισμοί ισότητας, τότε το σύστημα βρίσκεται σε κατάσταση ακραίου κινδύνου (in extremis). Η κατάσταση αποκατάστασης (restorative state) επαναφέρει σταδιακά το σύστημα στην κανονική κατάσταση, συγχρονίζοντάς το ξανά [5].

Οι μέθοδοι για την εκτίμηση κατάστασης χωρίζονται στις τεχνικές με βάση την τοπολογία, όπου χρησιμοποιούνται αλγόριθμοι όπως η μέθοδος των σταθμισμένων ελαχίστων τετραγώνων (Weighted Least Squares - WLS) [5] και η μέθοδος της εκτίμησης της μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation) [8], και στις τεχνικές με βάση τα δεδομένα, όπου χρησιμοποιούνται πιο έξυπνοι αλγόριθμοι, με κυρίως χρήση τεχνητής νοημοσύνης, όπως παλινδρόμηση με Γκαουσιανές διεργασίες (Gaussian process regression) [9] και μηχανές διανυσμάτων υποστήριξης (Support Vector Machines) [10]. Επίσης, υπάρχουν και υβριδικές μέθοδοι που συνδυάζουν τις παραπάνω τεχνικές [11].

2.2 Ακραία καιρικά φαινόμενα και το αντίκτυπό τους στα Ηλεκτρικά Συστήματα

Η κλιματική αλλαγή έχει αποτελέσει ένα παγκόσμιο ανησυχητικό φαινόμενο που απασχολεί αρκετά την επιστημονική κοινότητα, και όχι μόνο, τις τελευταίες δεκαετίες. Η αύξηση των ποσοστών διοξειδίου του άνθρακα (CO₂) και αερίων του θερμοκηπίου στην ατμόσφαιρα ([12]) έχει οδηγήσει στην ύπαρξη πολλών αλλαγών όπως αύξηση της μέσης θερμοκρασίας της ατμόσφαιρας της γης και την πιο συχνή εμφάνιση των ακραίων καιρικών φαινομένων, όπως καύσωνες, ξηρασίες, πλημμύρες και κύματα ψύχους [13]. Επιπτώσεις της κλιματικής αλλαγής έχουν παρατηρηθεί και στα ηλεκτρικά δίκτυα. Για παράδειγμα, η μεταφορική δυνατότητα των γραμμών μεταφοράς έχει περιοριστεί λόγω των υψηλών θερμοκρασιών και των έντονων καυσώνων ενώ, επίσης, έχουν αυξηθεί και οι απώλειες ενέργειας και η χαλάρωση των γραμμών μεταφοράς [14].

Οι ανεμοθύελλες και οι τυφώνες προκαλούν και αυτοί προβλήματα στις εναέριες γραμμές, είτε λόγω της κρούσης τους με παρασυρόμενα από τον αέρα αντικείμενα, είτε, σε ορισμένες περιπτώσεις, ακόμα και στην πτώση των πυλώνων λόγω της δύναμης του ανέμου [14]. Επιπλέον, τα κύματα ψύχους μπορούν να προκαλέσουν και αυτά βλάβες, καθώς ο πάγος και το χιόνι μπορούν να μαζευτούν στους μονωτές και

να δημιουργήσουν αγωγίμο πέρασμα που μπορεί να προκαλέσει σπινθηρισμούς [14]. Βραχυκύκλωμα, επίσης, μπορεί να προκληθεί και από κεραυνούς πάνω ή κοντά στους εναέριους αγωγούς, οι οποίοι θα πυροδοτήσουν τις ηλεκτρικές ασφάλειες και θα αποσυνδέσουν τις γραμμές. Σε κάποιες περιπτώσεις αυτή η απότομη αλλαγή τάσης μπορεί να μεταφερθεί μέσω των γραμμών και να προκαλέσει ζημιά στους μετασχηματιστές ισχύος [14]. Οι βροχές και οι πλημμύρες, ενώ δεν προκαλούν πρόβλημα στις εναέριες γραμμές μεταφοράς λόγω της καλής μόνωσής τους, μπορούν σε συνδυασμό με ισχυρούς ανέμους ή κεραυνούς να προκαλέσουν ζημιές στον εξοπλισμό των υποσταθμών όπως τα κυκλώματα διακοπής και τους πίνακες ελέγχου [14]. Η κατάσταση των ηλεκτρικών στοιχείων είναι επίσης παράγοντας που επηρεάζεται από τα ακραία καιρικά φαινόμενα. Στην πραγματικότητα, ο φθαρμένος και παλαιός εξοπλισμός στις γραμμές μεταφοράς ενισχύει την ζημιά που μπορεί να προκληθεί από τα ακραία καιρικά φαινόμενα [15]. Σύμφωνα με τα μηνιαία στατιστικά στοιχεία του ENTSO-E για την περίοδο 2010–2016 (22 κράτη μέλη), οι περισσότερες διακοπές ηλεκτρικής ενέργειας στην Ευρωπαϊκή Ένωση (ΕΕ) οφείλονταν σε αστοχίες εξοπλισμού ή υλικών (40%) καθώς και σε ακραία καιρικά φαινόμενα και φυσικές καταστροφές (33%). Επιπλέον, το 80% αυτών των διακοπών καταγράφηκε ως αποτέλεσμα αστοχιών στο επίπεδο διανομής [16].

Τα παραδείγματα όπου τα ακραία φαινόμενα έχουν προκαλέσει διακοπές λειτουργίας στα ηλεκτρικά δίκτυα είναι πολυάριθμα. Αρχικά, το πρώτο και πιο σημαντικό συμβάν είναι η διακοπή ηλεκτρικού ρεύματος το 2005 στο Μύνστερ της Γερμανίας. Εκεί, λόγω του συσσωρευμένου πάχους, των ισχυρών ανέμων και του παλαιού εξοπλισμού, αποδυναμώθηκε η υποδομή του σταθμού. Οι εναέριες γραμμές απέκτησαν πάγο με αποτέλεσμα 82 πυλώνες μεταφοράς υψηλής τάσης να τεθούν εκτός λειτουργίας. Το 2017 ο τυφώνας Μαρία επηρέασε σημαντικά το ηλεκτρικό δίκτυο του Πουέρτο Ρίκο, με την τιμή αποκατάστασης του ηλεκτρικού δικτύου να φτάνει τα 18 δισεκατομμύρια ευρώ. Μετά από ένα μήνα είχε αποκατασταθεί μόνο το 20% των γραμμών μεταφοράς, ενώ η συνολική διάρκεια αποκατάστασης ξεπέρασε τον έναν χρόνο. Το 2021 οι πλημμύρες στην Ευρώπη επηρέασαν το Γερμανικό ηλεκτρικό δίκτυο, αναδεικνύοντας τους κινδύνους της πλημμύρας στον εξοπλισμό των υποσταθμών. Λόγω της δύσκολης προσβασιμότητας και της επισκευής του δικτύου η διακοπή του ρεύματος διήρκησε μέχρι και έξι με οκτώ εβδομάδες [1]. Στα τέλη Δεκέμβρη του 2022, μία χιονοθύελλα στην Βόρεια Αμερική προκάλεσε συνεχόμενες διακοπές ρεύματος, τοπικές ή μεγαλύτερης εμβέλειας, αφήνοντας περίπου 1,5 εκατομμύρια ανθρώπους χωρίς ρεύμα [17]. Τον Ιούλιο του 2022 στο Λονδίνο, ενώ υπήρχε υψηλή ζήτηση και κατανάλωση ρεύματος για τον κλιματισμό των σπιτιών, αρκετές γραμμές ρεύματος βγήκαν σε δυσλειτουργία από διάφορες φωτιές. Για να αποτραπεί η μεγάλη μεγέθους διακοπή του ρεύματος, οι λειτουργοί του συστήματος διανομής του Ηνωμένου Βασιλείου αναγκάστηκαν να αγοράσουν ενέργεια και ρεύμα από τους σταθμούς του Βελγίου, οδηγώντας στην εκτίναξη των τιμών ενέργειας [17]. Τέλος, τον Απρίλη του 2025 υπήρξε διακοπή ρεύματος στην Ιβηρική Χερσόνησο λόγω έντονης αύξησης τάσης και επακόλουθη διακοπή του δικτύου, που οφείλεται στην αστάθεια και την φθορά του εξοπλισμού [18].

Όσον αφορά την Ελλάδα, λόγω της γεωγραφικής της θέσης και γεωμορφολογία της (ορεινή ηπειρωτική χώρα με έντονα διαμελισμένη ακτογραμμή) είναι αρκετά επιρρεπής χώρα σε κλιματικούς κινδύνους. Αποτελεί μία από τις ευρωπαϊκές χώρες με τα χαμηλότερα ποσοστά υπόγειου δικτύου, καθώς το εναέριο δίκτυο διανομής καλύπτει άνω του 88% της περιοχής εξυπηρέτησης. Ως εκ τούτου, οι υποδομές του δικτύου είναι ιδιαίτερα εκτεθειμένες σε ακραία μετεωρολογικά φαινόμενα, όπως κύματα καύσωνα την περίοδο του καλοκαιριού, ισχυρούς ανέμους συνοδευόμενους από χιονοπτώσεις,

έντονες βροχοπτώσεις με κεραυνούς, πλημμύρες κ.ά.[17]. Πράγματι, οι διακοπές ηλεκτροδότησης λόγω καιρικών συνθηκών έχουν αυξηθεί σημαντικά τα τελευταία χρόνια σε εθνικό επίπεδο, επηρεάζοντας μεγάλο αριθμό καταναλωτών. Πιο συγκεκριμένα, τα ακραία καιρικά φαινόμενα ευθύνονταν για περίπου 18,54% των βλαβών στην περιοχή Μέσης Τάσης (MT) για το έτος 2021, αμέσως μετά από τις αστοχίες εξοπλισμού (περίπου 22,72%), προκαλώντας σοβαρές ζημιές στις υποδομές και σημαντικές βλάβες στον εξοπλισμό, σύμφωνα με στοιχεία που παρέχονται από τον Διαχειριστή του Ελληνικού Δικτύου Διανομής Ηλεκτρικής Ενέργειας (ΔΕΔΔΗΕ Α.Ε.), ο οποίος είναι υπεύθυνος για τη λειτουργία και συντήρηση του εθνικού δικτύου διανομής [17].

Τα ακραία καιρικά φαινόμενα φέρουν επιπτώσεις και στο ιταλικό ηλεκτρικό δίκτυο, επηρεάζοντας τόσο τη ζήτηση όσο και την ικανότητα παραγωγής ηλεκτρικής ενέργειας. Η άνοδος της θερμοκρασίας και η αυξανόμενη συχνότητα καυσώνων αυξάνουν σημαντικά την κατανάλωση ηλεκτρικής ενέργειας, κυρίως λόγω της ευρύτερης χρήσης κλιματιστικών καθώς τα νοικοκυριά και οι επιχειρήσεις προσαρμόζονται στις θερμότερες συνθήκες. Αυτό εντείνει τα θερινά φορτία αιχμής, επιβαρύνοντας τα δίκτυα μεταφοράς και διανομής και αυξάνοντας τον κίνδυνο ελλείψεων στην παροχή. Από την πλευρά της παραγωγής, η Ιταλία αντιμετωπίζει μείωση του υδροηλεκτρικού δυναμικού, καθώς οι παρατεταμένες περίοδοι ξηρασίας και η μειωμένη διαθεσιμότητα νερού περιορίζουν τις εισροές στα ρεζερβουάρ, ενώ οι υψηλότερες θερμοκρασίες του αέρα και του νερού μειώνουν την αποδοτικότητα ψύξης και την ισχύ των θερμικών σταθμών. Οι συνδυασμένες αυτές επιδράσεις εντείνουν την ευπάθεια του συστήματος κατά τις περιόδους υψηλής ζήτησης και χαμηλής διαθεσιμότητας ανανεώσιμων πηγών. Επιπλέον, τα ακραία καιρικά φαινόμενα, όπως καύσωνες, καταιγίδες και ισχυροί άνεμοι, μπορούν να προκαλέσουν φυσικές ζημιές στις υποδομές του δικτύου, να διαταράξουν τη μεταφορά ηλεκτρικής ενέργειας και να οδηγήσουν σε δαπανηρές διακοπές λειτουργίας [19].

2.3 Αξιολόγηση της Ανθεκτικότητας και Προβλεπτική Συντήρηση

2.3.1 Αξιολόγηση της Ανθεκτικότητας (Resilience Assessment)

Η ανθεκτικότητα είναι ικανότητα ενός εργαλείου, συστήματος ή και ανθρώπου, να ανακάμπτει και να επιστρέφει στην φυσιολογική λειτουργία μετά από μία μεγάλη αναστάτωση. Στα συστήματα ισχύος, η αξιολόγηση της ανθεκτικότητας του συστήματος είναι ένα σημαντικό εργαλείο για την προετοιμασία, την διαχείριση και την καλύτερη ανάκαμψη ενός συστήματος μετά από ένα επιβλαβές γεγονός, όπως διακοπή ρεύματος, βλάβες εξοπλισμού, που οφείλονται σε ακραία συμβάντα. Τα γεγονότα αυτά χαρακτηρίζονται ως γεγονότα Υψηλής Επίδρασης - Χαμηλής Πιθανότητας (High Impact Low Propability (HILP) events) όπως οι σεισμοί, τυφώνες, πλημμύρες και η αντίδραση του συστήματος σε αυτά είναι το αντικείμενο των περισσότερων ερευνών και μελετών που έχουν προκύψει στην αξιολόγηση της ανθεκτικότητας [20].

Η εκτίμηση της ανθεκτικότητας ενός συστήματος γίνεται τυπικά με δύο τρόπους, ποιοτικά και ποσοτικά. Στη ποσοτική μέθοδο, επιπλέον, υπάρχουν 3 προσεγγίσεις: η αναλυτική μέθοδος, η στατιστική ανάλυση των ιστορικών δεδομένων, και η μέθοδος βασισμένη σε προσομοιώσεις [21]. Από αυτές τις τρεις τεχνικές ποσοτικής εκτίμησης, αυτή που βασίζεται σε προσομοιώσεις είναι αυτή που ευνοείται περισσότερο από ερευνητές παγκοσμίως [20].

2.3.1.1 Ποιοτική Μέθοδος

Στην ποιοτική αξιολόγηση της ανθεκτικότητας λαμβάνονται υπόψη ταυτοχρόνως και οι διαφορετικές πτυχές του συστήματος (όπως διάφορα υπό-συστήματα π.χ. το πληροφοριακό σύστημα, οι αλυσίδες παροχής καυσίμου) αλλά και οι δυνατότητές του (όπως η ετοιμότητα, απόκριση, αποκατάσταση, ύπαρξη πλάνου κρίσιμων καταστάσεων, εκπαίδευση προσωπικού κ.α.). Οι μέθοδοι αξιολόγησης περιλαμβάνουν έρευνες, ερωτηματολόγια, προσωπικές εκτιμήσεις και τεχνικές όπως η Αναλυτική Ιεραρχική Διαδικασία (AHP) οι οποίες μπορούν να μετατρέψουν τις υποκειμενικές γνώμες σε συγκρίνουσες ποσότητες. Η ποιοτική αξιολόγηση παρέχει μία γενική εικόνα του συστήματος και βοηθάει στον σχεδιασμό μιας μακροπρόθεσμης ενεργειακής πολιτικής [21].

2.3.1.2 Ποσοτική μέθοδος

Η ποσοτική μέθοδος είναι βασισμένη στην ποσοτικοποίηση των επιδόσεων του συστήματος. Γι αυτόν τον λόγο χρησιμοποιεί μετρικές, οι οποίες βοηθούν και στην αναγνώριση της υπάρχουσας επίδοσης, αλλά και στην σύγκριση με άλλα συστήματα. Όπως ειπώθηκε και παραπάνω στην ποσοτική μέθοδο εκτίμησης της ανθεκτικότητας υπάρχουν τρεις προσεγγίσεις.

Η αναλυτική μέθοδος χρησιμοποιεί την πιθανότητα βλάβης του συστήματος σε συγκεκριμένη περίπτωση για προβλεπτική λήψη αποφάσεων πριν κάποιο καιρικό φαινόμενο πλήξει το δίκτυο [22].

Η στατιστική μέθοδος αποσκοπεί στην μοντελοποίηση της ζημιάς του συστήματος, της διάρκειας των διακοπών ρεύματος και στην αποκατάσταση μετά από φυσικές καταστροφές. Μερικά από τα πιο χρησιμοποιημένα μοντέλα είναι τα Γενικευμένα Γραμμικά Μοντέλα, τα Γενικευμένα Προσθετικά Μοντέλα, τα μοντέλα Επιταχυνόμενου Χρόνου Αποτυχίας, τα μοντέλα εξόρυξης δεδομένων βασισμένα σε δένδρα κ.α. [23]. Τα μοντέλα αυτά αντλούν δεδομένα από δύο κατηγορίες: από δεδομένα του συστήματος ισχύος και από περιβαλλοντικά δεδομένα. Οι μετρικές που χρησιμοποιούνται για την ακρίβεια προσαρμογής των μοντέλων είναι μέσω σύγκρισης των αποτελεσμάτων πρόβλεψης με τα παρατηρούμενα δεδομένα. Τέτοιες μετρικές σύγκρισης είναι το Μέσο απόλυτο σφάλμα (MAE), η μέση απόλυτη απόκλιση (MAD), το μέσο τετραγωνικό σφάλμα (MSE) και η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE). [23].

Τέλος, η μέθοδος που βασίζεται σε προσομοιώσεις αποσκοπεί στην δημιουργία μοντέλων προσομοίωσης που παρομοιάζουν την εμφάνιση της καταστροφής και την απόκριση του συστήματος, με σκοπό την καλύτερη κατανόηση των φυσικών μηχανισμών της βλάβης και, σημαντικότερα, να προσδιορίσει τις απαραίτητες προετοιμασίες και τις διαδικασίες σκλήρυνσης του συστήματος [23].

2.3.1.3 Στρατηγικές Ανθεκτικότητας

Για να γίνει, πλέον, ένα σύστημα πιο ανθεκτικό χρησιμοποιούνται 2 στρατηγικές ανθεκτικότητας. Αρχικά, γίνεται η σκλήρυνση του συστήματος, δηλαδή γίνονται φυσικές αλλαγές των υποδομών έτσι ώστε να γίνει το σύστημα λιγότερο επιρρεπές σε ακραία γεγονότα. Μερικές από τις αλλαγές είναι η υπογείωση των γραμμών μεταφοράς, η αναβάθμιση των στύλων με πιο ανθεκτικά υλικά, η ανύψωση υπο-σταθμών και η μετακίνηση των εγκαταστάσεων και, τέλος η κοπή των κοντινών δένδρων και άλλης βλάστησης [21] [23]. Ο δεύτερος τρόπος είναι η χρήση τεχνολογίας έξυπνων

δικτύων (smart grid technology) με την οποία αυξάνεται η ορατότητα του συστήματος και προσδιορίζονται γρηγορότερα οι βλάβες [21]. Έτσι, το σύστημα μπορεί να γίνει πιο ανθεκτικό στα ακραία γεγονότα και βλάβες.

2.3.2 Προβλεπτική Συντήρηση (Predictive Maintenance)

Η συντήρηση των συστημάτων ηλεκτρικής ενέργειας αποτελεί ένα αναπόσπαστο μέρος της σταθερής και ασφαλούς λειτουργίας τους. Σύμφωνα με έρευνες ([24]) η απουσία συντήρησης μπορεί να μειώσει την παραγωγή κατά 5-20% ενώ η ύπαρξη της μπορεί να μειώσει τα κόστη συντήρησης κατά 18-25%. Καθώς οι βιομηχανίες εξελίσσονται, τόσο εξελίζονται και οι τεχνικές συντήρησης, με την πιο πρόσφατη να είναι η προβλεπτική συντήρηση [25].

Η πρώτη τεχνική συντήρησης ήταν η Διορθωτική Συντήρηση (Corrective Maintenance) η οποία λαμβάνει δράση μετά από την εμφάνιση βλαβών και αποσκοπεί στην διόρθωση των βλαβών. Η επόμενη ήταν η Αποτρεπτική Συντήρηση (Preventive Maintenance) η οποία κάνει προγραμματισμένους περιοδικούς ελέγχους και αντικαταστάσεις εξαρτημάτων ανά τακτικά χρονικά διαστήματα, έτσι ώστε να μειώσει όσο το δυνατόν περισσότερο τις βλάβες που εμφανίζονται. Η τελευταία, και πρόσφατη, είναι η Προβλεπτική Συντήρηση η οποία χρησιμοποιεί τις τεχνολογίες και τα εργαλεία της νέας γενιάς βιομηχανιών (Βιομηχανία 4.0) με σκοπό να έχει πλήρη και συνεχόμενη επίβλεψη των συστημάτων καθώς και να προβλέπει την μελλοντική απόδοση [24]. Μερικές από τις τεχνολογίες που χρησιμοποιεί είναι το Ίντερνετ των Πραγμάτων (Internet of Things) και τα ψηφιακά αντίγραφα (Digital Twins) όπου δημιουργούν εικονικά μοντέλα των συστημάτων, διεργασιών και των υπηρεσιών προσομοιώνοντας έτσι την συμπεριφορά του συστήματος. Αυτά τα μοντέλα καταλαβαίνουν τις φυσικές οντότητες, προβλέπουν τις αλλαγές, μέσω παρακολούθησης σε πραγματικό χρόνο, και δίνουν καλύτερη εσοπτεία για την κατάσταση του δικτύου [25].

2.4 Παραγωγή συνθετικών δεδομένων

Η παραγωγή συνθετικών δεδομένων (synthetic data generation) είναι ένα πολύ σημαντικό εργαλείο στην διαχείριση και επεξεργασία προβλημάτων που απαιτούν ένα ικανό και επαρκές σύνολο δεδομένων. Χωρίζεται σε δύο κλάδους, στην προσαύξηση του συνόλου δεδομένων (data augmentation) και στην παραγωγή δεδομένων (data production). Ο πρώτος κλάδος αφορά κυρίως τα προβλήματα ταξινόμησης όπου υπάρχει ανισορροπία των κλάσεων, καθώς και σε περιπτώσεις όπου το διαθέσιμο σύνολο δεδομένων δεν επαρκεί για την λύση ενός προβλήματος ή την εκπαίδευση ενός μοντέλου. Μέσω της παραγωγής συνθετικών δεδομένων της μειοψηφικής κλάσης εξισορροπείται το σύνολο δεδομένων βελτιώνοντας την επίδοσή του [26]. Ο δεύτερος κλάδος, και στον οποίο θα επεκταθούμε περισσότερο στην σχετική έρευνα, αφορά κυρίως προβλήματα παλινδρόμησης (regression), όπως πρόβλεψη ή παραγωγή συνεχόμενων δεδομένων, αλλά και περιπτώσεις όπου το σύνολο δεδομένων δεν είναι διαθέσιμο ή δεν μπορεί να χρησιμοποιηθεί για λόγους ιδιωτικότητας. Συγκεκριμένα στις βιομηχανίες, τα πραγματικά δεδομένα είναι, συνήθως, αρκετά προστατευμένα, απαιτούν αρκετό χρόνο για την απόκτησή τους και δεν μπορούν να χρησιμοποιηθούν για υπολογισμούς υποθετικών σεναρίων [27].

Υπάρχουν δύο θεμελιώδεις κατηγορίες της παραγωγής δεδομένων, οι στατιστικές μέθοδοι και τα μοντέλα βαθιάς μάθησης. Οι στατιστικές τεχνικές βασίζονται στην κατανομή των πραγματικών δεδομένων και στη δημιουργία δειγμάτων που ακολουθούν

το ίδιο στατιστικό πρότυπο. Οι αλγόριθμοι αυτοί περιλαμβάνουν συνήθως δύο βασικά στάδια, την εκτίμηση παραμέτρων (parameter estimation) και τη δειγματοληψία. Στην εκτίμηση παραμέτρων υπολογίζονται οι στατιστικές παράμετροι της κατανομής (όπως ο μέσος όρος, η διακύμανση και η συνδιακύμανση) χρησιμοποιώντας μεθόδους όπως η εκτίμηση της μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation) ή η εκτίμηση πυκνότητας με πυρήνες (Kernel Density Estimation). Αφού έχουν εκτιμηθεί οι παράμετροι, τότε γίνεται η δειγματοληψία με την προσαρμοσμένη κατανομή με χρήση γεννητριών τυχαίων αριθμών. Με αυτόν τον τρόπο εξασφαλίζεται ότι τα παραγόμενα δεδομένα διαθέτουν παρόμοια στατιστικά χαρακτηριστικά με εκείνα των αρχικών, όπως η συνάρτηση πυκνότητας πιθανότητας (Probability Density Function) [28]. Αυτές οι μέθοδοι χρησιμοποιούνται κυρίως, αλλά όχι μόνο, στην προσαύξηση του συνόλου δεδομένων.

Μερικά μοντέλα αυτής της περίπτωσης είναι τα Γκαουσιανά Μοντέλα Μίξης καθώς και οι Μαρκοβιανές Αλυσίδες [29]. Οι Μαρκοβιανές αλυσίδες μπορούν εύκολα να εξασφαλίσουν την στατιστική και πιθανοτική ομοιότητα μεταξύ των αρχικών και των παραγόμενων δεδομένων όπως παρουσιάζεται σε έρευνα σχετικά με την παραγωγή ημερησίου προφίλ φορτίου ενέργειας [30].

Τα μοντέλα βαθιάς μάθησης χρησιμοποιούν τα διαθέσιμα δεδομένα προκειμένου να υπολογίσουν την έξοδο. Ένα μοντέλο που χρησιμοποιείται συχνά είναι οι Μεταβλητοί Αυτοκωδικοποιητές (Variational Autoencoders - VAEs) οι οποίοι λειτουργούν μαθαίνοντας να συμπιέζουν τα δεδομένα εισόδου σε μία μικρότερη αναπαράσταση, γνωστή ως λανθάνων χώρος (latent space), από τον οποίο επιχειρούν στην συνέχεια να ανακατασκευάσουν τα αρχικά δεδομένα [31]. Σε αντίθεση με τους κλασικούς αυτοκωδικοποιητές, οι VAEs δεν απομνημονεύουν απλώς τα δεδομένα, αλλά μαθαίνουν τη βασική δομή της κατανομής τους, γεγονός που τους επιτρέπει να παράγουν νέα δείγματα με παρόμοια χαρακτηριστικά. Λόγω αυτής της ιδιότητας, χρησιμοποιούνται ευρέως σε εφαρμογές όπως η παραγωγή εικόνων, η ανίχνευση ανωμαλιών και η συμπίεση δεδομένων. Οι VAEs έχουν καταστεί βασικά εργαλεία σε τομείς όπως η όραση υπολογιστών (computer vision) και η επεξεργασία φυσικής γλώσσας (NLP), λόγω της ικανότητάς τους να μαθαίνουν και να αναπαριστούν πολύπλοκα πρότυπα δεδομένων [28]. Η χρήση αυτοκωδικοποιητών για παραγωγή συνθετικών δεδομένων είναι ευρέως διαδεδομένη. Στο [32] χρησιμοποιήθηκε multichannel autoencoder (MCAE) για να υποβοηθήσει ταξινομητές, με αποτέλεσμα να βελτιώνεται η αναπαράσταση των χαρακτηριστικών. Στο [33] εφαρμόστηκε variational autoencoder (VAE) για ανισόροπα δεδομένα εικόνων, δείχνοντας ότι μπορεί να παράγει νέα δείγματα και να υπερέχει άλλων μεθόδων σε διάφορα σύνολα δεδομένων και μετρικές. Στο [34] ένας VAE παρήγαγε τεχνητά δεδομένα ατυχημάτων δρόμου για χρήση σε πρόβλημα ταξινόμησης, παρουσιάζοντας καλύτερα αποτελέσματα ως προς την εκπαίδευση τους και τις μετρικές ταξινόμησης. Τέλος, στο [35] γίνεται χρήση ενός VAE, μαζί με LSTM, για την πρόβλεψη τιμών για κατανάλωση ενέργειας έξυπνων δικτύων.

Ένα ακόμα μοντέλο που χρησιμοποιείται συχνά στην παραγωγή συνθετικών δεδομένων είναι τα Γενετικά Αντιπαραθετικά Δίκτυα (Generative Adversarial Networks - GANs) στα οποία θα επεκταθούμε θεωρητικά στην επόμενη ενότητα. Λόγω της αρχιτεκτονικής τους, τα GANs μπορούν να παράγουν εξαιρετικά ρεαλιστικά αποτελέσματα, καθιστώντας τα ιδανικά για παραγωγή εικόνων, επαύξηση δεδομένων και δημιουργία νέου περιεχομένου (όπως τέχνης ή μουσικής). Στην περίπτωση της επαύξησης του συνόλου δεδομένων τα GANs έχουν παρουσιάσει σημαντικά αποτελέσματα. Για παράδειγμα, έχουν χρησιμοποιηθεί για την δημιουργία τεχνητών δεδομένων στον τομέα της Ηλεκτροεγκεφαλογραφίας (EEG) παράγοντας εγκεφαλικά σήματα παρό-

μοια με τα πραγματικά [36]. Σε άλλη εφαρμογή, χρησιμοποιήθηκε ένα Conditional GAN (CGAN) για την επαύξηση ενός συνόλου σημάτων διαμόρφωσης που χρησιμοποιείται για αυτόματη ταξινόμηση [37]. Τα δεδομένα αυτά βελτίωσαν την ακρίβεια ενός ταξινομητή CNN που λειτούργησε ως benchmark και οι συγγραφείς συμπέραναν ότι ο εμπλουτισμός με CGAN ωφελεί σημαντικά την εκπαίδευση με CNN παρέχοντας ταχύτερη σύγκλιση και χαμηλότερη απώλεια εκπαίδευσης [37]. Στο [38] προτάθηκε το One-Shot GAN, το οποίο, με είσοδο μία μόνο εικόνα (ή βίντεο), μπορεί να δημιουργήσει εικόνες (ή βίντεο) σημαντικά διαφορετικές από την αρχική. Βελτιώνει την ποιότητα και την ποικιλία σε περιπτώσεις που τα δεδομένα είναι ελάχιστα, μετριάζοντας την απομνημόνευση και παράγοντας δομικά διαφορετικά δείγματα.

2.5 Μοντέλα που χρησιμοποιήθηκαν στην παρούσα εργασία

Στην παρακάτω ενότητα θα εξηγηθεί λεπτομερώς το θεωρητικό υπόβαθρο των μοντέλων που χρησιμοποιήθηκαν στην παρούσα εργασία. Συγκεκριμένα, τα τρία μοντέλα που χρησιμοποιήθηκαν είναι ένα Μεγάλο Γλωσσικό Μοντέλο (Large Language Model - LLM) ως το βασικό μοντέλο και άλλα δύο ως μοντέλα σύγκρισης. Ένα μοντέλο βαθιάς μάθησης, το TimeGAN, και ένα στατιστικό, το Gaussian Copula.

2.5.1 Μεγάλα Γλωσσικά Μοντέλα (LLMs)

Τα μεγάλα γλωσσικά μοντέλα έχουν εμφανίσει ραγδαία άνοδο τα τελευταία χρόνια. Ανήκουν στην κατηγορία μοντέλων επεξεργασίας φυσικής γλώσσας (Natural Language Processing) και η ανάγκη για αυτά τα μοντέλα πηγάζει από την αυξημένη ζήτηση για τον χειρισμό πολύπλοκων γλωσσικών έργων, όπως είναι η μετάφραση, η περίληψη κειμένων, η ανάκτηση πληροφοριών, οι αλληλεπιδράσεις σε διεπαφές συνομιλίας κ.α. [39].

Ιστορικά, η επεξεργασία φυσικής γλώσσας ξεκίνησε με στατιστική μοντελοποίηση η οποία στην συνέχεια αντικαταστάθηκε από τα νευρωνικά δίκτυα. Όμως, η ανάγκη για επεξεργασία δεδομένων μεγάλης κλίμακας καθυστέρωσε τα ήδη υπάρχοντα μοντέλα. Η δημιουργία των transformers έλυσε αυτό το πρόβλημα και πάνω σε αυτά στηρίζονται τα LLMs όπως τα ξέρουμε τώρα. [40]

Μερικά από τα σημαντικά συστατικά στοιχεία των LLMs αναλύονται παρακάτω:

- **Tokenization:** Είναι η διαδικασία μετατροπής του κειμένου σε διακριτές μονάδες (tokens), οι οποίες μπορεί να είναι χαρακτήρες, λέξεις, υπο-λέξεις ή σύμβολα. Είναι ένα πολύ σημαντικό τμήμα της προ-επεξεργασίας της εκπαίδευσης των LLMs [39].
- **Κωδικοποίηση Θέσης - Position Encoding:** Είναι η διαδικασία, των transformers, της διαφορετικής κωδικοποίησης των λέξεων ανάλογα με την θέση στην οποία βρίσκονται μέσα στο κείμενο. Αυτό γίνεται προσθέτοντας ένα διάνυσμα ενσωμάτωσης θέσης (positional embedding vector) στις μονάδες (tokens) που προέκυψαν μετά από το tokenization. Οι τεχνικές ενσωμάτωσης θέσης μπορεί να χρησιμοποιούν απόλυτες, σχετικές ή πολυμαθείς κωδικοποιήσεις θέσης [39].
- **Attention:** Το attention αναθέτει βάρη στις μονάδες tokens εισόδου με βάση τη βαρύτητά τους έτσι ώστε το μοντέλο να δίνει περισσότερη έμφαση στις πιο σχετικές μονάδες. Το attention ως μέρος των transformers υπολογίζει τις αντιστοιχίες των ερωτημάτων (queries), κλειδιών (keys) και των τιμών στις ακολουθίες

εισόδου, και υπολογίζει ένα σκορ πολλαπλασιάζοντας το ερώτημα με το κλειδί και αυτό χρησιμοποιείται για τον υπολογισμό των βαρών [41]. Δύο από τις σημαντικότερες τεχνικές του attention είναι το self-attention και το cross attention. Η πρώτη υπολογίζει το σκορ χρησιμοποιώντας ερωτήματα, κλειδιά και τιμές ίδιου μπλοκ (κωδικοποιητή (encoder) ή αποκωδικοποιητή (decoder)). Ενώ, στη δεύτερη ο κωδικοποιητής βγάζει στην έξοδο του τα ερωτήματα και ο αποκωδικοποιητής προσφέρει τα ζευγάρια κλειδιών-τιμών. [39] [41]

- Μετεκπαίδευση - Fine Tuning: Η διαδικασία με την οποία τα προ-εκπαιδευμένα LLMs εκπαιδεύονται πάλι σε πιο εξειδικευμένες περιπτώσεις χρησιμοποιώντας ένα ειδικό και συγκεκριμένο σύνολο δεδομένων. Υπάρχουν οι εξής πιο σημαντικές προσεγγίσεις:

- Μεταφορά Μάθησης (Transfer Learning) στην οποία γίνεται η επανεκπαίδευση του μοντέλου με συγκεκριμένα δεδομένα αλλά πάνω στο ίδιο task, άρα χρησιμοποιεί ένα ήδη εκπαιδευμένο μοντέλο για να κάνει μιά συγκεκριμένη εργασία.

- Instruction Tuning, στο οποίο γίνεται η επανεκπαίδευση του μοντέλου με ένα σύνολο δεδομένων σε συγκεκριμένη μορφή (instruction form). Αυτή η μορφή αποτελείται από οδηγίες και από δεδομένα σε μορφή ζευγών εισόδου-εξόδου. Οι οδηγίες κατευθύνουν το μοντέλο να απαντήσει ανάλογα με το prompt και την είσοδο [39]. Η έξοδος ορίζουμε να είναι η επιθυμητή έξοδος που θέλουμε να βγάζει το μοντέλο, οπότε είναι ευέλικτο να χρησιμοποιηθεί σε πολλές εργασίες.

- Low Rank Adaptation (LORA): Είναι μια Parameter-Efficient Fine-Tuning (PEFT) μέθοδος η οποία είναι σχεδιασμένη να προσαρμόσει τα προ-εκπαιδευμένα llms χωρίς να επανεκπαιδεύσει όλα τα βάρη/παραμέτρους. Τα βασικά του πλεονεκτήματα είναι ότι μειώνει αρκετά την μνήμη που χρησιμοποιείται κάνοντάς το πιο γρήγορο, αποφεύγει καταστροφικές αλλαγές βαρών αφού τα αρχικά παραμένουν σταθερά, και είναι αρκετά ευέλικτο αφού μπορούμε να έχουμε πολλούς προσαρμογείς και να τους αλλάζουμε ανάλογα την εργασία. Αντί να ενημερώνει όλο το μοντέλο, "παγώνει" τα αρχικά βάρη (W) και χρησιμοποιεί δύο πολύ μικρότερους πίνακες και όπου αποτελούν την κατά προσέγγιση αλλαγή των αρχικών βαρών. Τυπικά, το αρχικό μοντέλο υπολογίζει την έξοδο ως $Y = WX + b$ και το LoRA το αλλάζει σε $Y = (W + BA)X + b$, όπου το W μένει "παγωμένο" και μόνο τα A και B εκπαιδεύονται [42].

Σε συνέχεια του LoRA υπάρχει το Quantized LoRA (QLoRA) το οποίο επεκτείνει την προηγούμενη μέθοδο αποθηκεύοντας τα αρχικά βάρη σε κβαντισμένη μορφή (συνήθως 4-bit) με αποτέλεσμα μεγαλύτερη μείωση των παραμέτρων και την δυνατότητα fine-tuning μεγάλων μοντέλων σε μηχανές με περιορισμένους πόρους GPU [43].

2.5.1.1 Το Μοντέλο Text-to-Text Transfer Transformer (T5)

Ο Text-to-Text Transfer Transformer (T5) είναι από τα πρώτα μοντέλα (μαζί με τα gpt-2, BERT, BART) όπου έδωσαν τα θεμέλια για την ύπαρξη αυτών που ονομάζουμε τώρα llm [44]. Αναλυτικότερα, παρουσιάστηκε στο [45] και η κεντρική ιδέα του είναι να μετατρέψει όλες τις εργασίες επεξεργασίας φυσικής γλώσσας (NLP) σε μία ενιαία μορφή κειμένου-σε-κείμενο, όπου τόσο οι είσοδοι όσο και οι έξοδοι εκφράζονται ως ακολουθίες κειμένου. Για παράδειγμα, ένα πρόβλημα ανάλυσης συναισθήματος μπορεί να εκφραστεί ως "η ταινία ήταν φανταστική" $\rightarrow$ "θετικό", ενώ μια εργασία μετάφρασης

μπορεί να διατυπωθεί ως “μετάφρασε από Αγγλικά σε Γερμανικά: How are you?” → “Wie geht es dir?”. Αυτός ο επαναπροσδιορισμός επιτρέπει την ύπαρξη μίας ενιαίας αρχιτεκτονικής για ένα ευρύ φάσμα εργασιών.

Αρχιτεκτονική Transformer

Το T5 βασίζεται στην αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή Transformer που προτάθηκε στο [41].

Ας θεωρήσουμε μια ακολουθία εισόδου n διακριτών συμβόλων:

$$X = (x_1, x_2, \dots, x_n).$$

Κάθε σύμβολο προβάλλεται σε ένα διάνυσμα διάστασης d και συνδυάζεται με πληροφορία θέσης. Ο κωδικοποιητής εφαρμόζει στρώσεις πολλαπλής κεφαλής self-attention και πλήρως συνδεδεμένων νευρωνικών δικτύων. Το κλιμακώμενο self-attention ορίζεται ως:

$$\text{Attention}(Q, K, V) = \left(\frac{QK^\top}{\sqrt{d_k}} \right) V,$$

όπου d_k είναι η διάσταση των διανυσμάτων-κλειδιών.

Η πολλαπλή κεφαλή προσοχής ορίζεται ως:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O,$$

με κάθε κεφαλή

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V).$$

Ο κωδικοποιητής παράγει συμφοραζόμενες αναπαραστάσεις

$$H = (h_1, h_2, \dots, h_n).$$

Ο αποκωδικοποιητής περιλαμβάνει δύο είδη προσοχής: (i) masked self-attention, ώστε η θέση t να εξαρτάται μόνο από τις προηγούμενες θέσεις, και (ii) cross-attention, όπου τα ερωτήματα του αποκωδικοποιητή παρατηρούν τις εξόδους του κωδικοποιητή H . Δεδομένων των συμβόλων στόχου $Y = (y_1, \dots, y_m)$, ο αποκωδικοποιητής παράγει καταστάσεις $Z = (z_1, \dots, z_m)$. Τέλος, με γραμμική προβολή και softmax προκύπτει η κατανομή:

$$P(y_t | y_{<t}, X) = \text{softmax}(Wz_t + b).$$

2.5.1.1.1 Προεκπαίδευση

Ένα από τα βασικά στοιχεία του T5 είναι το C4 (Colossal Clean Crawled Corpus). Το C4 περιλαμβάνει περίπου 750 GB αγγλικού κειμένου, προερχόμενο από δισεκατομμύρια ιστοσελίδες, αφού του αφαιρέθηκαν τα μη γλωσσικά στοιχεία (π.χ. HTML tags, κώδικας, διαφημίσεις) και εφαρμόστηκαν φίλτρα για την απομάκρυνση τοξικού ή επαναλαμβανόμενου περιεχομένου. Η διαδικασία καθαρισμού στόχευε στη διατήρηση ενός αντιπροσωπευτικού και ποιοτικού συνόλου δεδομένων, το οποίο θα μπορούσε να λειτουργήσει ως γενική «βάση γνώσης» για την προεκπαίδευση.

Το T5 προεκπαιδεύτηκε στο C4 με χρήση της μεθόδου span corruption, όπου τυχαία συνεχή τμήματα κειμένου αντικαθίστανται από ειδικά σύμβολα (sentinel tokens) και το μοντέλο καλείται να τα ανακατασκευάσει. Για παράδειγμα:

Είσοδος: “The quick brown fox <extra_id_0> the lazy dog.”

Έξοδος: “jumps over”

Τυπικά, δεδομένης της εισόδου X και της «κατεστραμμένης» εισόδου $\tilde{X}$, η συνάρτηση κόστους είναι

$$\mathcal{L} = \sum_{t=1}^T \log P(y_t | y_{<t}, \tilde{X}),$$

όπου Y είναι τα τμήματα που λείπουν.

Η χρήση του C4 σε συνδυασμό με το span corruption αποδείχθηκε κρίσιμη για τη γενίκευση του μοντέλου σε ένα ευρύ φάσμα εργασιών.

2.5.1.1.2 Ενοποιημένη μεταφορά μάθησης

Μία από τις σημαντικότερες συνεισφορές του T5 είναι η ενοποίηση όλων των εργασιών NLP σε ένα ενιαίο πλαίσιο κειμένου-σε-κείμενο. Αντί να χρησιμοποιούνται εξειδικευμένες αρχιτεκτονικές ή στρώματα εξόδου για κάθε διαφορετικό τύπο εργασίας (όπως ταξινόμηση, μετάφραση, περίληψη ή ερωταπαντήσεις), όλες οι εργασίες εκφράζονται ως ακολουθίες κειμένου με κοινή μορφοποίηση εισόδου και εξόδου. Για παράδειγμα:

Είσοδος: “summarize: [έγγραφο]” $\rightarrow$ Έξοδος: [περίληψη],

Αυτό το ενοποιημένο σχήμα καθιστά δυνατή τη χρήση του ίδιου μηχανισμού εκπαίδευσης, της ίδιας αρχιτεκτονικής και της ίδιας διαδικασίας αποκωδικοποίησης για κάθε εργασία. Σε συνδυασμό με την προεκπαίδευση στο C4 και το span corruption, το T5 μπορεί να μεταφέρει αποτελεσματικά τη γνώση που απέκτησε σε εξειδικευμένες εργασίες χωρίς την ανάγκη ανασχεδιασμού του μοντέλου. Επίσης, σύμφωνα με το [45] η αύξηση της κλίμακας του μοντέλου (μέχρι και 11 δισεκατομμύρια παραμέτρους στο T5-11B) οδηγεί σε σημαντικές βελτιώσεις απόδοσης, επιτυγχάνοντας κορυφαία αποτελέσματα σε benchmarks όπως GLUE, SuperGLUE, και σύνολα δεδομένων περίληψης.

Με αυτόν τον τρόπο, το T5 κατέστησε εφικτή μία πραγματικά ενοποιημένη προσέγγιση στη μεταφορά μάθησης, η οποία υπήρξε καθοριστική για την εξέλιξη των σύγχρονων μεγάλων γλωσσικών μοντέλων [45].

2.5.1.2 Μοντέλα GAN

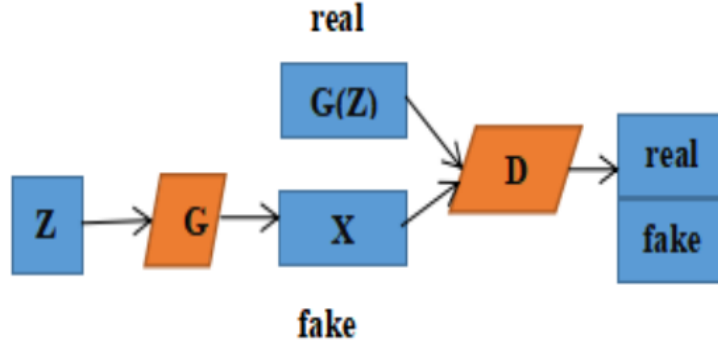
Τα Generative Adversarial Networks (GANs) [46] ανήκουν σε μια κατηγορία μοντέλων βαθιάς μάθησης με σκοπό τη δημιουργία νέων συνθετικών δεδομένων που μιμούνται τα χαρακτηριστικά μιας πραγματικής κατανομής. Η χρησιμότητά τους προκύπτει από την ικανότητά τους να μαθαίνουν πολύπλοκες πιθανότητες κατανομών δεδομένων χωρίς να απαιτούν ρητή παραμετρική περιγραφή. Με αυτόν τον τρόπο, τα GANs έχουν βρει εφαρμογές σε τομείς όπως η σύνθεση εικόνων, η δημιουργία ρεαλιστικών βίντεο, η επεξεργασία ομιλίας, αλλά και σε πιο εξειδικευμένους χώρους, όπως η ιατρική απεικόνιση ή η ανάλυση ενεργειακών δεδομένων.

Ουσιαστικά, η βασική τους λειτουργία είναι να εκπαιδεύουν ένα ζευγάρι νευρωνικών δικτύων τη γεννήτρια (Generator) (G) και τον διακριτή (Discriminator) (D) σε ένα ανταγωνιστικό παιχνίδι. Ο G παίρνει ως είσοδο έναν θόρυβο Z (Σχήμα 2.1) και προσπαθεί να παράγει δείγματα που να μοιάζουν με τα πραγματικά, ενώ ο D προσπαθεί να εντοπίσει αν ένα δείγμα είναι πραγματικό ή παραγόμενο. Αυτή η αντιπαράθεση οδηγεί τον G να παράγει ολοένα και πιο ρεαλιστικά δεδομένα.

Το τυπικό GAN διατυπώνεται ως πρόβλημα ελαχιστοποίησης–μεγιστοποίησης:

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))],$$

όπου p_{data} είναι η πραγματική κατανομή δεδομένων και p_z ένας απλός θόρυβος (π.χ. Γκαουσιανός). Ο D μεγιστοποιεί την ορθή ταξινόμηση, ενώ ο G ελαχιστοποιεί την ικανότητα του D να ανιχνεύει ψευδή δείγματα.



Σχήμα 2.1: Generative Adversarial Networks (GANs)

2.5.1.2.1 Παραλλαγές GAN

Υπάρχουν πολλές παραλλαγές για σταθερότερη εκπαίδευση, καλύτερη ποιότητα και ελεγχόμενη παραγωγή δεδομένων [47]:

- DCGAN: συνελικτικές αρχιτεκτονικές για υψηλότερη ποιότητα εικόνων [48].
- WGAN: χρήση της απόστασης Wasserstein για ομαλότερα gradients και λιγότερη κατάρρευση τρόπων [49].
- PGGAN: προοδευτική αύξηση ανάλυσης κατά την εκπαίδευση [50].
- StyleGAN: style-based σύνθεση με ιεραρχικό έλεγχο χαρακτηριστικών [51].
- InfoGAN: μεγιστοποίηση αμοιβαίας πληροφορίας για αποσυνδεδεμένες (disentangled) αναπαραστάσεις [52].

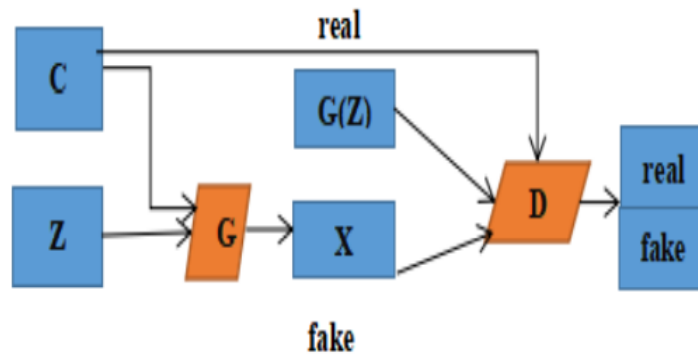
2.5.1.2.2 Conditional GANs (cGANs)

Τα cGANs επεκτείνουν τα απλά GANs εισάγοντας μια επιπλέον συνθήκη y (π.χ. labels, attributes ή άλλα features), την οποία χρησιμοποιούν τόσο ο G όσο και ο D . Έτσι, η παραγωγή γίνεται κατευθυνόμενη προς συγκεκριμένα χαρακτηριστικά [47].

Πιο συγκεκριμένα, για τα cGANs ισχύει:

$$\min_G \max_D \mathbb{E}_{x,y \sim p_{data}(x,y)} [\log D(x | y)] + \mathbb{E}_{z \sim p_z(z), y \sim p_y(y)} [\log(1 - D(G(z | y) | y))].$$

Η συνθήκη y τροφοδοτείται και στα δύο δίκτυα, επιβάλλοντας τα παραγόμενα συνθετικά δεδομένα να είναι όχι μόνο ρεαλιστικά αλλά και να περιέχουν το ζητούμενο χαρακτηριστικό [47].



Σχήμα 2.2: Conditional GANs (cGANs)

Η εισαγωγή της συνθήκης y στο δίκτυο μπορεί να γίνει με διάφορους τρόπους. Μια συνηθισμένη πρακτική είναι η συνένωση (concatenation) του y με το διάνυσμα θορύβου z στην είσοδο της γεννήτριας, έτσι ώστε το παραγόμενο δείγμα να σχετίζεται με τη συνθήκη. Αντίστοιχα, η ίδια πληροφορία μπορεί να ενσωματωθεί στον διακριτή, συνενώνοντας το y με τα πραγματικά ή τα παραγόμενα δείγματα. Σε πιο σύνθετες περιπτώσεις χρησιμοποιούνται τεχνικές όπως ενσωματώσεις (embeddings) ή ειδικές μορφές κανονικοποίησης (π.χ. Conditional Batch Normalization), ώστε η συνθήκη να επηρεάζει τα ενδιαμέσα στρώματα του δικτύου. Με πιο απλά λόγια, η συνθήκη y μπορεί να θεωρηθεί σαν ένα επιπλέον "σήμα καθοδήγησης" που δείχνει στην γεννήτρια τι να φτιάξει και στον διακριτή πώς να το αξιολογήσει [47].

2.5.1.2.3 TimeGAN

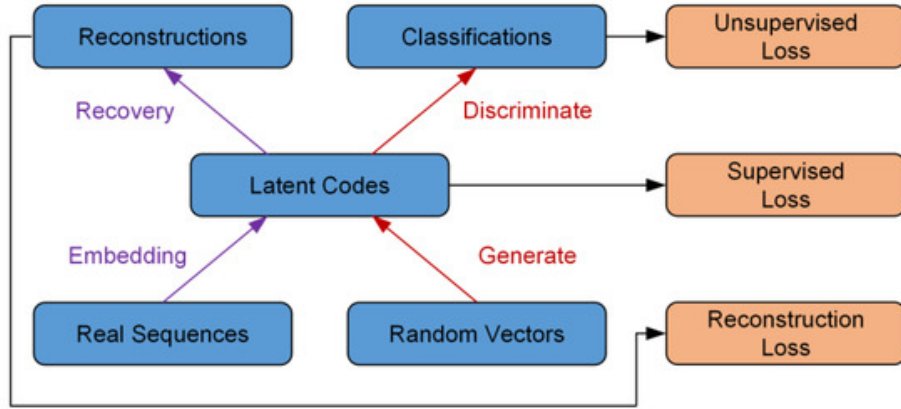
Μία από τις σημαντικότερες επεκτάσεις των GANs σε χρονοσειρές είναι το Time-series Generative Adversarial Network (TimeGAN) [53]. Το TimeGAN αποτελεί ένα state-of-the-art μοντέλο και σχεδιάστηκε ειδικά για να αντιμετωπίσει τη παραγωγή συνθετικών χρονοσειρών που να διατηρούν τόσο τις χρονικές εξαρτήσεις όσο και τις κατανομές των πραγματικών δεδομένων.

Σε αντίθεση με τα κλασικά GANs που εστιάζουν κυρίως στη ρεαλιστική αναπαράσταση μεμονωμένων δεδομένων (π.χ. εικόνων), το TimeGAN συνδυάζει εποπτευόμενη και μη εποπτευόμενη μάθηση ώστε να αποτυπώσει τόσο τη δυναμική όσο και τις χρονικές εξαρτήσεις των χρονοσειρών. Συγκεκριμένα, μέσω της εποπτευόμενης μάθησης επιβάλλεται το κριτήριο ανακατασκευής, δηλαδή τα συνθετικά δεδομένα να μπορούν να αναπαράγουν σωστά τα πραγματικά στο χωροχρονικό λανθάνοντα χώρο. Αντίστοιχα, μέσω της μη εποπτευόμενης μάθησης εφαρμόζεται ο κλασικός ανταγωνιστικός μηχανισμός (adversarial loss), όπου ο διακριτής προσπαθεί να ξεχωρίσει τα πραγματικά από τα συνθετικά δείγματα. Με αυτόν τον συνδυασμό, το TimeGAN εξασφαλίζει όχι μόνο ρεαλισμό στις χρονοσειρές αλλά και διατήρηση της χρονικής τους δομής [53].

Συγκεκριμένα, το TimeGAN αποτελείται από τέσσερα βασικά συστατικά στοιχεία:

- έναν ενσωματωτή (embedder), ο οποίος μετασχηματίζει τις αρχικές χρονοσειρές σε έναν λανθάνοντα χωροχρονικό χώρο,
- έναν ανασυνθέτη (recovery), που ανακατασκευάζει τις χρονοσειρές από τον λανθάνοντα χώρο ώστε να διατηρείται η πιστότητα στα πραγματικά δεδομένα,
- μία γεννήτρια (generator), η οποία παράγει νέες λανθάνουσες ακολουθίες με διατήρηση της χρονικής συνέπειας,

- και έναν διακριτή (discriminator), που καλείται να διαχωρίσει τις πραγματικές από τις συνθετικές ακολουθίες.



Σχήμα 2.3: TimeGAN Architecture

Ο συνολικός στόχος εκπαίδευσης συνδυάζει πολλαπλές συνιστώσες κόστους:

$$\mathcal{L} = \mathcal{L}_{reconstruction} + \mathcal{L}_{supervised} + \mathcal{L}_{adversarial},$$

εξασφαλίζοντας ότι οι συνθετικές χρονοσειρές όχι μόνο μοιάζουν ρεαλιστικές στον διακριτή, αλλά και ακολουθούν τη σωστή χρονική δομή.

Πειραματικά, οι [53] έδειξαν ότι το TimeGAN ξεπερνά άλλες μεθόδους παραγωγής χρονοσειρών σε όρους οπτικής πιστότητας, ικανότητας πρόβλεψης μοντέλων και ποσοτικών μετρικών ομοιότητας. Έτσι, το TimeGAN θεωρείται σήμερα ένα από τα πιο αποδοτικά πλαίσια για συνθετική παραγωγή δεδομένων χρονοσειρών.

2.5.2 Γκαουσιανά Copulas (Gaussian Copulas)

Οι Copulas ως μέλος των πιθανοθεωρητικών συνδέσμων αποτελούν κατάλληλο μέσο για την επίτευξη του συνδυασμού κατανομών πιθανοτήτων. Η βασική θεωρία θεμελιώνεται στο θεώρημα του Sklar, σύμφωνα με το οποίο κάθε πολυδιάστατη συνάρτηση κατανομής F μπορεί να εκφραστεί ως:

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)),$$

όπου F_i είναι οι οριακές κατανομές και C είναι μία συνάρτηση συνεξάρτησης.

Μία ειδική περίπτωση είναι η Gaussian Copula η οποία κατασκευάζεται από μία πολυδιάστατη Γκαουσιανή κατανομή με μηδενικό διάνυσμα μέσου και πίνακα συσχέτισης R [54]. Η δισδιάστατη Γκαουσιανή συνεξάρτηση με παράμετρο συσχέτισης ρ δίνεται από:

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)),$$

όπου Φ^{-1} είναι η αντίστροφη της τυπικής κανονικής συνάρτησης κατανομής και Φ_ρ είναι η δισδιάστατη Γκαουσιανή συνάρτηση κατανομής με συντελεστή συσχέτισης ρ .

Τα Copulas χρησιμοποιούνται περισσότερο σε δεδομένα που βρίσκονται σε μορφή πίνακα. Επεξεργάζεται κάθε στήλη ξεχωριστά και για τον υπολογισμό της συνδιακύμανσης χρησιμοποιεί την μέθοδο των Gaussian Copulas [27].

Ένα παράδειγμα χρήσης του παρουσιάζεται στο [55] όπου οι ερευνητές χρησιμοποιούν ένα περιοδικό Gaussian Copula ώστε να αποτυπώσουν τόσο τη χρονική εξάρτηση όσο και τη χωρική συσχέτιση μεταξύ φωτοβολταϊκών (PV) συστημάτων. Η μέθοδος βασίζεται σε έναν μετασχηματισμό που χαρτογραφεί τις πολυδιάστατες χρονοσειρές παραγωγής ενέργειας σε ανεξάρτητες κανονικές μεταβλητές, επιτρέποντας την κατασκευή ενός πιθανοτικού μοντέλου που είναι ομαλό και επαναλαμβανόμενο στο χρόνο. Από το μοντέλο αυτό μπορούν να αντληθούν νέα δείγματα, τα οποία στη συνέχεια αντιστρέφονται μέσω του μετασχηματισμού ώστε να προκύψουν συνθετικές χρονοσειρές παραγωγής PV που αναπαράγουν τις ημερήσιες συνδιακυμάνσεις και τις συσχετίσεις μεταξύ διαφορετικών τοποθεσιών.

2.6 Σχετική Έρευνα

2.6.1 Παραγωγή Σύνθετων Χρονοσειρών

Η παραγωγή χρονοσειρών είναι ένα αντικείμενο με το οποίο έχει ασχοληθεί ενδελεχώς η επιστημονική κοινότητα τις τελευταίες δεκαετίες. Έχουν προταθεί πολλές μέθοδοι, απλοί οι σύνθετοι, με σκοπό την αποδοτικότερη παραγωγή σύνθετων δεδομένων σε χρονοσειρές ενέργειας [56]. Στο [57] γίνεται πρόταση ενός στοχαστικού limited-ARIMA (LARIMA) μοντέλου το οποίο παράγει χρονοσειρές αιολικής ενέργειας βασισμένες στην παραγωγή ενός αιολικού πάρκου. Συνεχίζοντας με τα αιολικά πάρκα, στο [58] γίνεται χρήση ενός μοντέλου βασισμένο στα Γκαουσιανά μοντέλα μίξης και στις Μαρκοβιανές αλυσίδες, πάλι για τον ίδιο σκοπό.

Η χρήση των GANs στην παραγωγή δεδομένων χρονοσειρών είναι εμφανής και σημαντική. Στα [59] [60] [61] [62] γίνεται χρήση ενός απλού GAN για την παραγωγή σεναρίων χρονοσειρών σε έξυπνα δίκτυα με δεδομένα παραγωγής αιολικής και ηλιακής ενέργειας, και κατανάλωσης ηλεκτρισμού. Στα [63] [64] χρησιμοποιείται το cGAN, στο πρώτο ([63]) για την δημιουργία πρόσθετων χρονοσειρών σε ένα αραιό σύνολο δεδομένων σχετικά με την κατανάλωση ενέργειας ενός συνόλου πολυκατοικιών, και στο δεύτερο ([64]) για να προσομοιάσει τις καταναλώσεις ενέργειας με βάση συγκεκριμένα χαρακτηριστικά. Στα [65] [66] γίνεται πρόταση του μοντέλου Recurrent-GAN (R-GAN) το οποίο χρησιμοποιεί RNNs ως τα δύο βασικά νευρωνικά δίκτυα, τα οποία, σύμφωνα με τους συγγραφείς, βοηθούν στην αντίληψη των χρονικών εξαρτήσεων. Η πρώτη έρευνα αξιοποιεί τα Wasserstein GANs (WGANs) και τα Metropolis-Hastings GANs (MH-GANs) για να βελτιώσει την σταθερότητα της εκπαίδευσης και εξετάστηκε σε δεδομένα κατανάλωσης ισχύος σε δίκτυο ηλεκτρικής ενέργειας, ενώ η δεύτερη χρησιμοποίησε συνθήκες (conditions), άρα RCGAN, για την παραγωγή χρονοσειρών ιατρικών δεδομένων. Τέλος, στο [67] χρησιμοποιείται μία επέκταση του TimeGAN για χρονοσειρές κατανάλωσης ενέργειας. Η διαφορά με το αυθεντικό TimeGAN είναι ότι στο στάδιο recovery, χρησιμοποιείται ένα επιπλέον επίπεδο multi-head self-attention όπου οι συγγραφείς επισημαίνουν ότι βελτιώνει την ικανότητα του μοντέλου να καταλαβαίνει σημαντικές πληροφορίες.

Σημαντική είναι και η χρήση των VAEs στην παραγωγή χρονοσειρών ενέργειας. Στο [68] χρησιμοποιείται ένα απλό VAE για την παραγωγή συνθετικών χρονοσειρών σε δεδομένα έξυπνων δικτύων. Επίσης, στα [69] [70] γίνεται χρήση των Conditional-VAEs για την παραγωγή συνθετικών χρονοσειρών με σκοπό την επαύξηση του συνόλου δεδομένων. Το πρώτο ([69]) χρησιμοποιείται σε σύνολο δεδομένων για τον υπολογισμό κλοπής ηλεκτρικής ενέργειας, ενώ το δεύτερο ([70]) για την πρόβλεψη κατανάλωσης ενέργειας σε κτίρια.

Μέθοδοι με χρήση LLMs

Η χρήση των LLMs στις χρονοσειρές έχει μελετηθεί εκτενώς στην διεθνή βιβλιογραφία. Στην παραγωγή χρονοσειρών οι μελέτες έχουν χωριστεί στις κατηγορίες πρόβλεψης χρονοσειρών (forecast) και παραγωγής χρονοσειρών (generation). Στην πρόβλεψη της πρόβλεψης έχουν προταθεί πολλά νέα μοντέλα όπως το Time-LLM [71] που μπορεί να χρησιμοποιηθεί εύκολα σε δεδομένα ενέργειας, π.χ. πρόβλεψη ισχύος σε φωτοβολταϊκά [72]. Στο [73] χρησιμοποιείται ένα llm για να κάνει forecast το μελλοντικό φορτίο ενέργειας, βάζοντας στο prompt την αυτούσια τιμή κάθε στοιχείου της χρονοσειράς. Τέλος, έχει προταθεί και το FPE-LLM [74] το οποίο έχει εκπαιδευτεί αποκλειστικά στην πρόβλεψη τιμών ενέργειας και έχει μετ-εκπαιδευτεί σε ακραίες καιρικές συνθήκες.

Στην παραγωγή χρονοσειρών οι μελέτες είναι εμφανώς λιγότερες [75]. Στην μέθοδο GenG, η οποία παρουσιάζεται στο [76], οι συγγραφείς χωρίζουν τη διαδικασία σε δύο υπο-διαδικασίες, όπου στην πρώτη γίνεται η μετ-εκπαίδευση ενός LLM για την κατανόηση της εργασίας, και στην δεύτερη χρησιμοποιείται ένα χρησιμοποιείται ένα transformer-based μοντέλο για να παραγάγει τα δεδομένα. Στο [77] χρησιμοποιείται μια παρόμοια μέθοδος με αυτή της παρούσας εργασίας, για την παραγωγή χρονοσειρών σε δεδομένα χρηματοοικονομικά. Οι συγγραφείς εστιάζουν στην διατήρηση της πιστότητας, ιδιωτικότητας και χρηστικότητας των παραγόμενων δεδομένων. Τέλος, στο [78] παρουσιάζεται το SDForger, μία καλά ορισμένη μέθοδος παραγωγής συνθετικών χρονοσειρών η οποία πρώτα μετατρέπει τα δεδομένα σε μορφή πίνακα και στην συνέχεια κάνει μετ-εκπαίδευση και παραγωγή των δεδομένων

2.6.2 Σενάρια ακραίων γεγονότων στα Συστήματα Ηλεκτρικής Ενέργειας

Η ανάλυση και παραγωγή σεναρίων ακραίων καιρικών γεγονότων αποτελεί κρίσιμο πεδίο για τη μελέτη και λειτουργία των συστημάτων ηλεκτρικής ενέργειας. Τα σενάρια αυτά επιτρέπουν την αποτύπωση πιθανών συνθηκών που επηρεάζουν τόσο τη ζήτηση όσο και την παραγωγή, ενώ παρέχουν πολύτιμη πληροφορία για την αξιολόγηση κινδύνων και την ενίσχυση της αξιοπιστίας του δικτύου. Στη βιβλιογραφία συναντώνται διάφορες μεθοδολογίες για τη δημιουργία τέτοιων σεναρίων, βασισμένες είτε σε στατιστικά και πιθανοτικά μοντέλα είτε σε τεχνικές μηχανικής μάθησης, με στόχο την αναπαράσταση των αβεβαιοτήτων και την προσομοίωση της συμπεριφοράς του συστήματος υπό ακραίες συνθήκες.

Μία βασική ερευνητική κατεύθυνση αφορά τη δημιουργία συνθετικών δεδομένων και σεναρίων μέσω μοντέλων μηχανικής μάθησης και στοχαστικών μεθόδων. Για παράδειγμα, στο [79] εφαρμόζεται το TimeGAN (LSTM-based) για την ταυτόχρονη παραγωγή συνθετικών δεδομένων καιρού, φορτίου και παραγωγής φωτοβολταϊκής ισχύος, επιτρέποντας τη διατήρηση των αλληλεξαρτήσεων μεταξύ τους. Στη συνέχεια, τα δεδομένα συσταδοποιούνται σε κατηγορίες που αντανακλούν διαφορετικές λειτουργικές συνθήκες (π.χ. χαμηλό φορτίο με μειωμένη παραγωγή λόγω βροχής), οδηγώντας σε ένα σύνολο αντιπροσωπευτικών σεναρίων. Αντίστοιχα, στο [80] αξιοποιείται το ίδιο μοντέλο (TimeGAN) με στόχο την εξισορρόπηση των δεδομένων καυσώνων και, σε δεύτερο στάδιο, ένα BiLSTM ενσωματώνει τις θερμοκρασιακές πληροφορίες για τη βελτίωση της πρόβλεψης φορτίου κατά τη διάρκεια ακραίων θερμικών γεγονότων. Σχετικά με την παραγωγή συνθετικών δεδομένων ηλιακής ακτινοβολίας, στο [81] προτείνεται ο συνδυασμός ενός ντετερμινιστικού μοντέλου (για την αποτύπωση εποχικών μοτίβων) με ένα στοχαστικό (για τυχαίες διακυμάνσεις όπως η νέφωση), αυξάνοντας σημαντικά το διαθέσιμο σύνολο δεδομένων.

Η δημιουργία σεναρίων έχει επίσης εξεταστεί με στατιστικές και πιθανοτικές μεθόδους. Στο [82], χρησιμοποιούνται συναρτήσεις copula και πυκνότητας πυρήνα ώστε να εξαχθούν οι κατανομές της φωτοβολταϊκής παραγωγής και του φορτίου ηλεκτρικών οχημάτων υπό συνθήκες καύσωνα. Τα αποτελέσματα συσταδοποιούνται σε αντιπροσωπευτικά σενάρια, τα οποία προσομοιάζουν ρεαλιστικά τη συμπεριφορά του συστήματος. Αντίστοιχα, στο [83] αξιοποιείται η μέθοδος IMLE (Implicit Maximum Likelihood Estimation), η οποία με τη χρήση θορύβου και κατάλληλης συνάρτησης απώλειας διασφαλίζει ότι τα συνθετικά δείγματα βρίσκονται κοντά σε πραγματικά δεδομένα, επιτρέποντας τη δημιουργία αξιόπιστων σεναρίων για δίκτυα ΑΠΕ.

Επιπλέον, αρκετές μελέτες επικεντρώνονται στη συνεκτίμηση βλαβών και αντοχών του δικτύου υπό ακραίες συνθήκες. Για παράδειγμα, στο [84] αναλύεται η συμπεριφορά ενός δικτύου με γεννήτριες σε συνθήκες χιονοθύελλας, όπου οι γεννήτριες κατηγοριοποιούνται με βάση τα χαρακτηριστικά τους και υπολογίζεται η πιθανότητα βλάβης καθεμιάς. Στη συνέχεια, σενάρια παράγονται μέσω τυχαίας επιλογής συνδυασμών σοβαρότητας καιρού, αντοχής γεννητριών και επιπέδων βλάβης. Αντίστοιχα, στο [85] εξετάζεται πώς οι βραχυπρόθεσμες μετεωρολογικές προβλέψεις μπορούν να αξιοποιηθούν για τον υπολογισμό πιθανοτήτων βλαβών σε πολλά στοιχεία του δικτύου. Οι πιθανότητες αυτές ενσωματώνονται σε στοχαστικά σενάρια, τα οποία με τη σειρά τους αξιοποιούνται σε ένα πλαίσιο στοχαστικής βελτιστοποίησης, βελτιώνοντας την προετοιμασία και ανθεκτικότητα των συστημάτων έναντι ακραίων γεγονότων.

Συνολικά, οι παραπάνω μελέτες αναδεικνύουν ότι η παραγωγή ρεαλιστικών σεναρίων μέσω συνδυασμού στοχαστικών μοντέλων, πιθανοτικών μεθόδων και αλγορίθμων μηχανικής μάθησης αποτελεί βασικό εργαλείο για την κατανόηση των επιπτώσεων των ακραίων γεγονότων στα συστήματα ισχύος. Τα σενάρια που προκύπτουν χρησιμοποιούνται σε ένα ευρύ φάσμα εφαρμογών, από την πρόβλεψη φορτίου και την ανάλυση παραγωγής ΑΠΕ, έως την εκτίμηση πιθανοτήτων βλαβών και τον σχεδιασμό στρατηγικών ανθεκτικότητας.

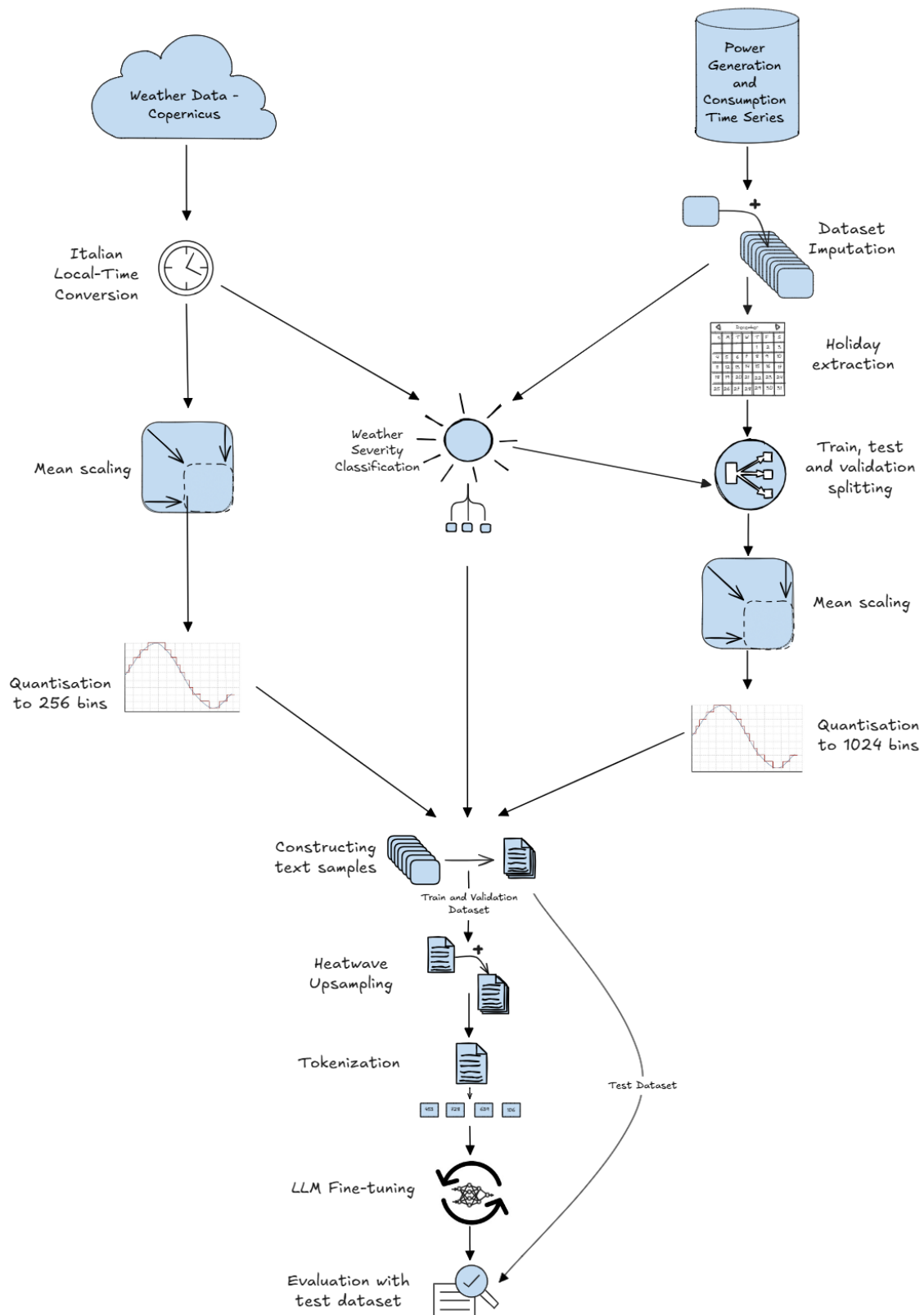
Κεφάλαιο 3

Μεθοδολογία και πειραματική διαδικασία

3.1 Εισαγωγή

Στο παρακάτω κεφάλαιο παρουσιάζεται αναλυτικά η μεθοδολογία που ακολουθήθηκε για την ανάλυση και προεπεξεργασία του συνόλου δεδομένων, την εκπαίδευση των μοντέλων καθώς και τη διαδικασία και αξιολόγηση των πειραμάτων. Πιο συγκεκριμένα, περιγράφεται ο σχεδιασμός του βασικού μοντέλου, η εκπαίδευσή του και η εκπαίδευση των δύο επιπλέον μοντέλων σύγκρισης. Τέλος, αναλύεται ο τρόπος παραγωγής των συνθετικών δεδομένων από όλα τα μοντέλα και ο τρόπος αξιολόγησής τους.

Στο παρακάτω σχήμα παρουσιάζεται η διαδικασία μετ-εκπαίδευσης (fine tuning) του γλωσσικού μοντέλου, ωστόσο κάποια στάδια, όπως η προεπεξεργασία των αρχικών δεδομένων, είναι κοινά και στα άλλα δύο μοντέλα.



Σχήμα 3.1: Διάγραμμα κύκλου μηχανικής μάθησης για το LLM Μοντέλο

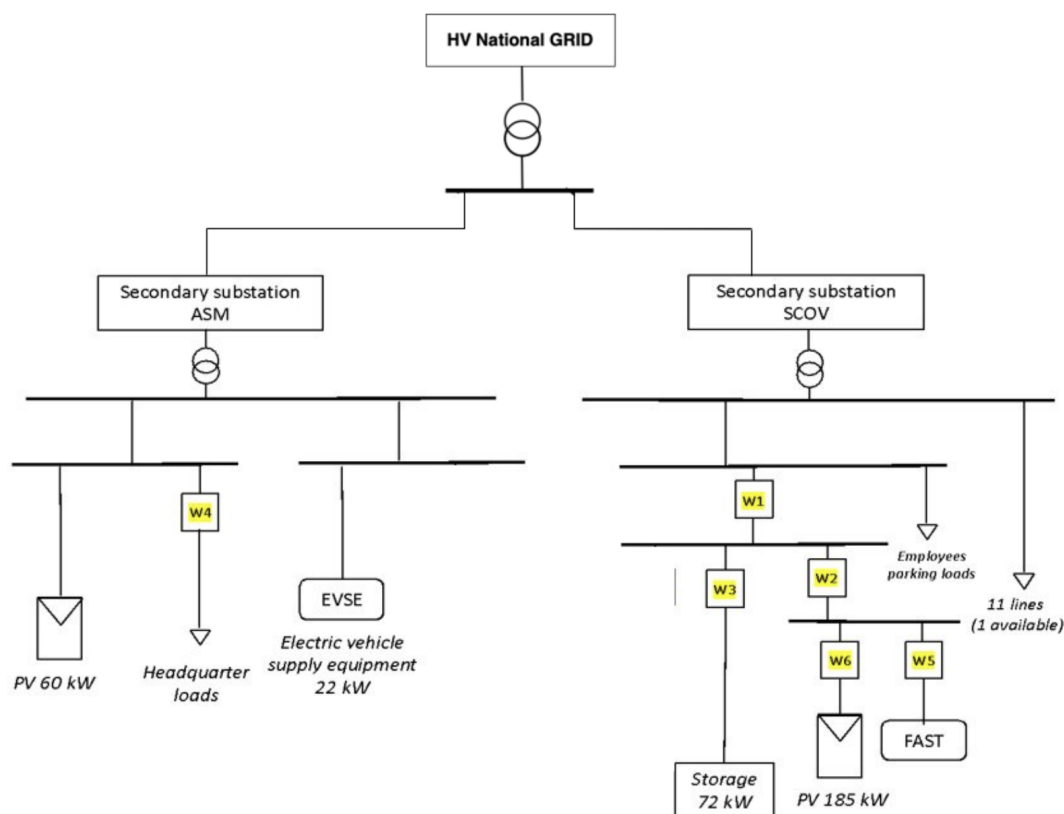
3.2 Ανάλυση και προεπεξεργασία συνόλου δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα εργασία είναι τα δεδομένα ενέργειας που προμήθευσε η εταιρία ASM, καθώς και τα καιρικά δεδομένα τα οποία προέρχονται από το ερευνητικό σύστημα παρατήρησης Γης, τον Κοπέρνικο [86]. Παρακάτω πραγματοποιείται ανάλυση των δεδομένων και περιγράφεται λεπτομερώς η διαδικασία προεπεξεργασίας τους, ώστε να καταστούν κατάλληλα για εισαγωγή και χρήση στα μοντέλα.

3.2.1 Δεδομένα ενέργειας

3.2.1.1 Ανάλυση δεδομένων

Τα δεδομένα που προμηθεύτηκαν από την εταιρία ASM αφορούν την Ενεργό και Άεργο Ισχύ των έξυπνων μετρητών (w_i), σε ανάλυση (resolution) πέντε λεπτών, του παρακάτω δικτύου ηλεκτρικής ενέργειας:



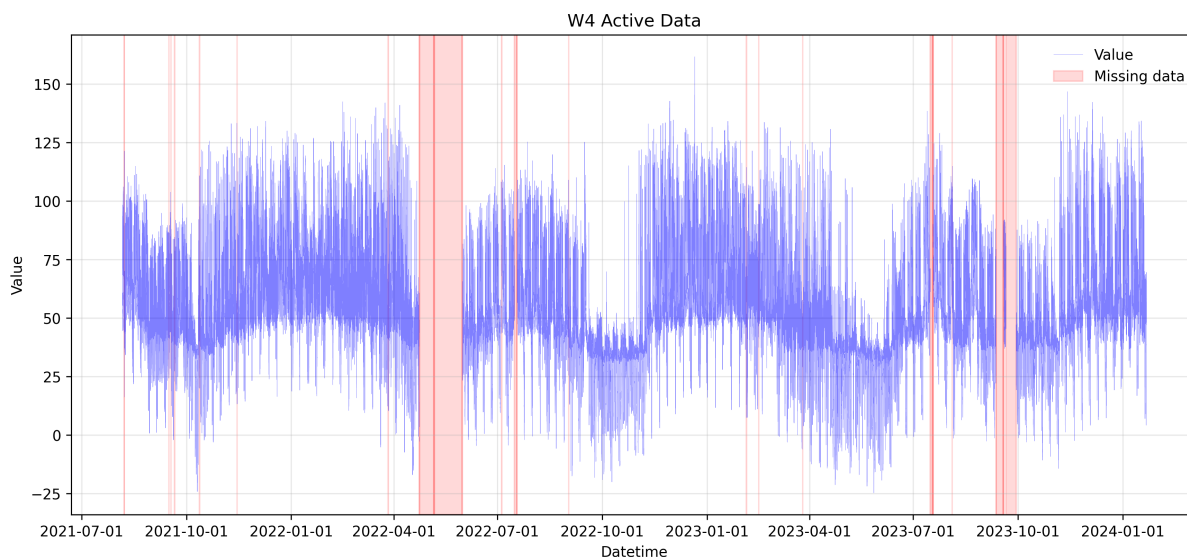
Σχήμα 3.2: Τοπολογία Δικτύου

Το δίκτυο αποτελείται από 2 υποσταθμούς και 6 μετρητές, όπως φαίνονται στην τοπολογία. Ο μετρητής w_3 αναφέρεται στον σταθμό αποθήκευσης, ο w_5 στον FAST φορτιστή και ο w_6 στην παραγωγή της γεννήτριας. Ο μετρητής w_4 αναφέρεται και στην κατανάλωση των κεντρικών γραφείων και στην παραγωγή του φωτοβολταϊκού 60kW (ένα επίπεδο πάνω από ότι βρίσκεται στην παραπάνω εικόνα). Στην παρούσα εργασία δόθηκε έμφαση στους ανθρώπινους παράγοντες και στον τρόπο με τον οποίο

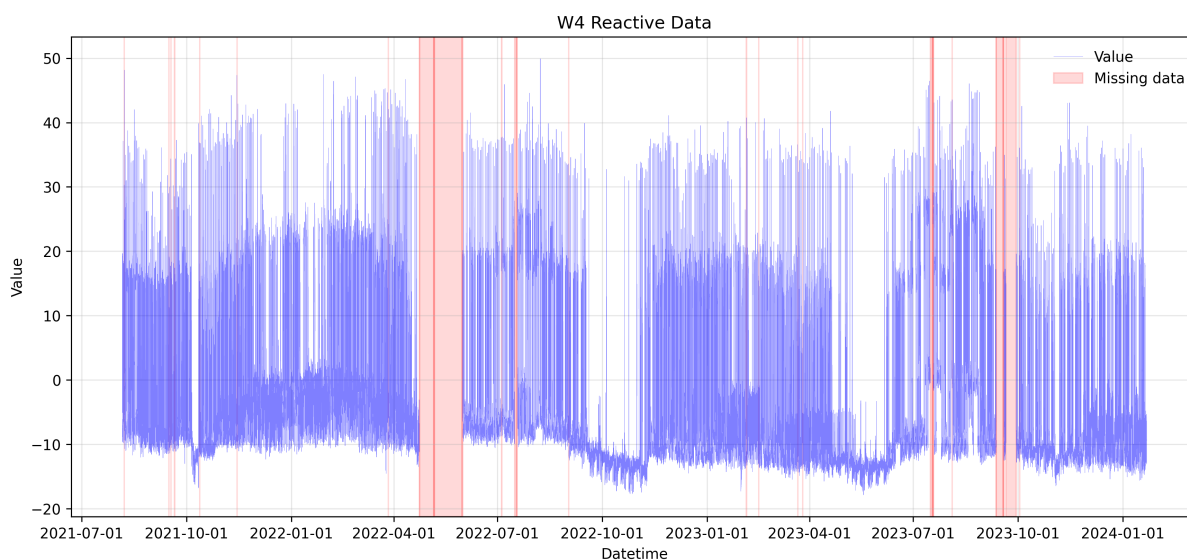
3.2. Ανάλυση και προεπεξεργασία συνόλου δεδομένων

η κατανάλωση επηρεάζεται από τις καιρικές συνθήκες, για τον λόγο αυτό χρησιμοποιήθηκε αποκλειστικά ο μετρητής w4. Επιπλέον, τα δεδομένα των υπόλοιπων μετρητών δεν ήταν αρκετά ποιοτικά καθώς περιείχαν πολλά κενά. Τα διαθέσιμα δεδομένα του w4 εκτείνονται από τον Αύγουστο του 2021 έως και τον Ιανουάριο του 2024.

Στις παρακάτω εικόνες φαίνονται τα δεδομένα του μετρητή w4 πριν από οποιαδήποτε επεξεργασία:



Σχήμα 3.3: Ενεργός Ισχύς του μετρητή 4



Σχήμα 3.4: Άεργος Ισχύς του μετρητή 4

Όπως φαίνεται από το σχήμα, υπάρχουν αρκετά κενά, τόσο συνεχόμενα, όσο και μεμονωμένα. Επομένως, η πρώτη ενέργεια ήταν η αντικατάσταση των άγνωστων τιμών (imputation).

Η αντικατάσταση άγνωστων τιμών είναι μια διαδικασία που συμβαίνει συχνά στην επεξεργασία των δεδομένων των χρονοσειρών, ειδικά όταν τα δεδομένα δεν είναι

επαρκή [87]. Η πιο συνήθης πρακτική στην αντιμετώπιση των άγνωστων τιμών είναι η αγνόησή τους, ωστόσο αυτή η πρακτική μειώνει σημαντικά το διαθέσιμο σύνολο δεδομένων, ειδικά στην συγκεκριμένη περίπτωση που τα δεδομένα των καλοκαιριών, που είναι σημαντικά, είναι ελάχιστα [87].

Μία άλλη, απλή, πρακτική είναι η γραμμική παρεμβολή (linear interpolation - LI). Στην γραμμική παρεμβολή η αντικατάσταση των άγνωστων διαστημάτων γίνεται με την επιλογή των τιμών οι οποίες βρίσκονται στην ευθεία που σχηματίζουν τα άκρα του μικρότερου διαστήματος [88]. Όμως, αυτή η μέθοδος δεν διατηρεί την εποχικότητα της χρονοσειράς, με αποτέλεσμα αυτή να χάνεται σε μεγάλα διαστήματα άγνωστων τιμών. Η μέθοδος που χρησιμοποιήθηκε είναι αυτή που παρουσιάζεται στο [87] η οποία συνδυάζει την γραμμική παρεμβολή με τον ιστορικό μέσο όρο (historical average - HA), δημιουργώντας μια βέλτιστα σταθμισμένη εκτίμηση (optimally weighted average - OWA). Η βασική ιδέα είναι ότι για μικρά διαστήματα άγνωστων τιμών, η γραμμική παρεμβολή παράγει πιο ακριβείς εκτιμήσεις, ενώ για τα μεγάλα διαστήματα, οι τιμές του ιστορικού μέσου όρου, που υπολογίζονται από χρονικά σημεία με παρόμοια εποχικά και ημερήσια χαρακτηριστικά, δίνει καλύτερα αποτελέσματα. Η τελική εκτίμηση προκύπτει ως:

$$\hat{y}_i^{OWA} = w_i \hat{y}_i^{LI} + (1 - w_i) \hat{y}_i^{HA},$$

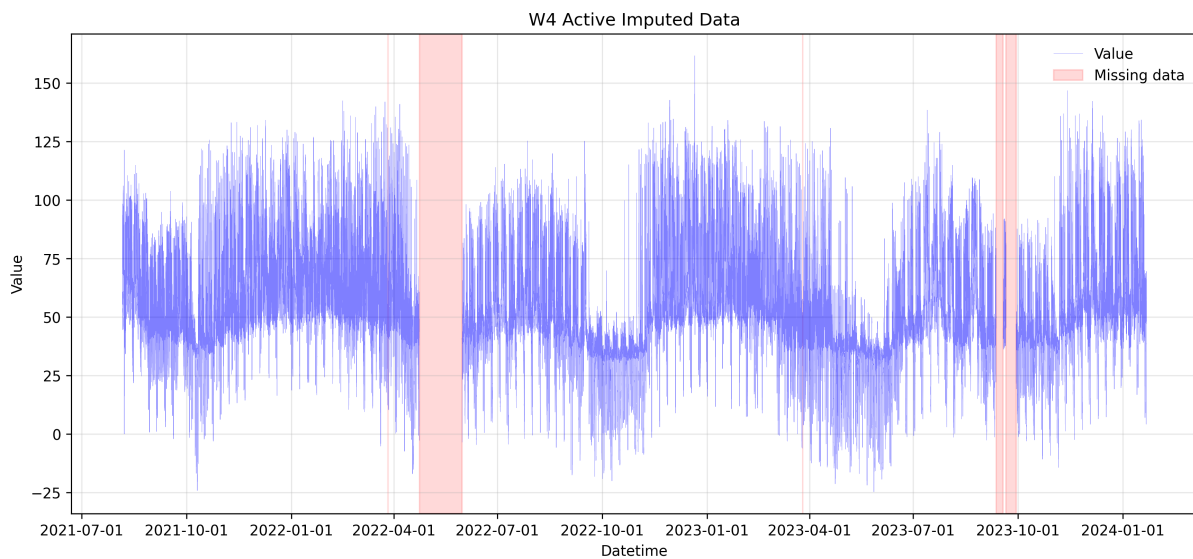
όπου ο συντελεστής βάρους w_i μειώνεται εκθετικά με την απόσταση d_i από το πλησιέστερο διαθέσιμο δείγμα:

$$w_i = e^{-\alpha d_i}, \quad \alpha > 0.$$

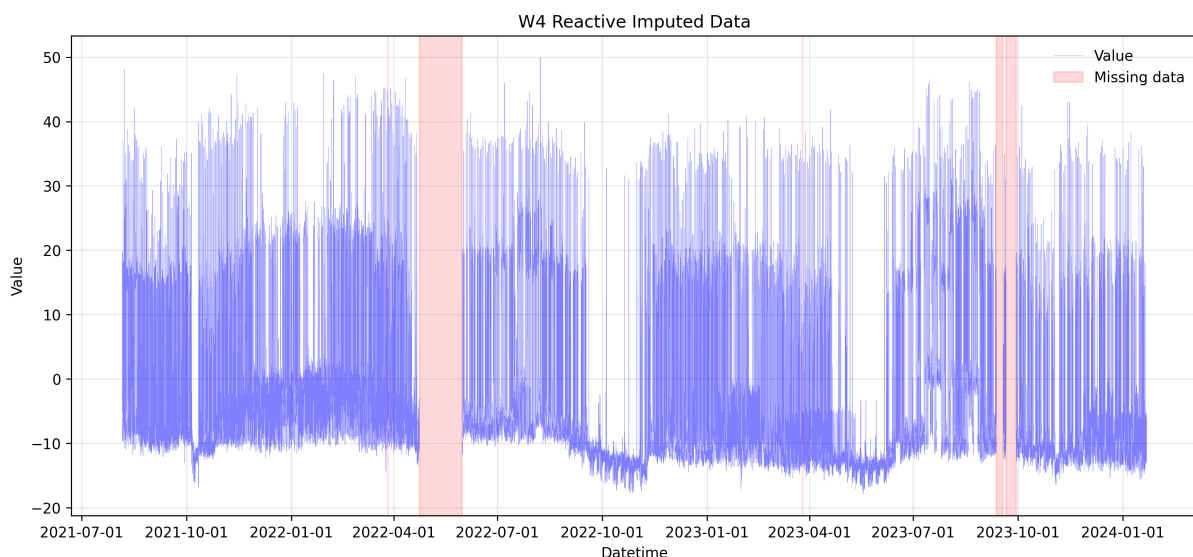
Για μικρές τιμές d_i υπερισχύει η εκτίμηση της LI, ενώ για μεγάλες τιμές d_i υπερισχύει η HA. Η βέλτιστη τιμή του α καθορίζεται από τα χαρακτηριστικά του συνόλου δεδομένων όπως η διακύμανση και τα ιστορικά μοτίβα. Στην παρούσα αντικατάσταση χρησιμοποιήθηκε $\alpha = 0.3$, και όχι $\alpha = 0.1$, όπως χρησιμοποιείται στο [87], διότι, ενώ τα παρόντα δεδομένα είναι ίδιας φύσης με αυτά της δημοσίευσης, δηλαδή μετρήσεις από ένα smart grid, εκείνα έχουν ανάλυση 15 λεπτών ενώ εδώ υπάρχει ανάλυση 5 λεπτών, οπότε προτιμήθηκε μεγαλύτερη τιμή του α για να υπερισχύει η LI για περισσότερα χρονικά βήματα. Τέλος το μέγιστο διάστημα αντικατάστασης, δηλαδή το μέγιστο κενό στο οποίο θα μπορέσει να γίνει το imputation, επιλέξαμε να είναι 24 ώρες.

Τα τελικά δεδομένα του μετρητή w4, μετά από την αντικατάσταση άγνωστων τιμών, είναι:

3.2. Ανάλυση και προεπεξεργασία συνόλου δεδομένων



Σχήμα 3.5: Ενεργός Ισχύς του μετρητή w4 μετά την αντικατάσταση άγνωστων τιμών



Σχήμα 3.6: Άεργος Ισχύς του μετρητή w4 μετά την αντικατάσταση άγνωστων τιμών

Όπου βλέπουμε ότι μειώθηκαν αρκετά οι άγνωστες τιμές και τα συνεχόμενα διαστήματα είναι μεγαλύτερα, ειδικά τους θερινούς μήνες. Στην συνέχεια αναφέρεται η προεπεξεργασία των δεδομένων για την είσοδό τους στα μοντέλα, όπως φαίνεται και στο πάνω μέρος του Σχήματος 3.1.

3.2.1.2 Προεπεξεργασία δεδομένων

Αρχικά, έγινε μετατροπή της 5-λεπτης ανάλυσης των δεδομένων σε ανάλυση μιας ώρας παίρνοντας τον μέσο όρο των δώδεκα 5-λέπτων της κάθε ώρας. Έπειτα, ελέγχοντας τις ημερομηνίες κάθε έτους, βρέθηκαν και αφαιρέθηκαν οι μέρες που είναι εθνικές εορτές στην Ιταλία, καθώς επηρεάζουν αρνητικά την εκπαίδευση του μοντέλου (holiday extraction). Επιπλέον, χρησιμοποιώντας τα καιρικά δεδομένα, έγινε κατηγοριοποίηση των ημερών του συνόλου δεδομένων ανάλογα με την ένταση του καιρού

(weather severity). Στην περίπτωση της παρούσας εργασίας, επειδή το αντικείμενο μελέτης αφορά καύσωνες, οι 3 κατηγορίες είναι: "normal" όταν η θερμοκρασία είναι μικρότερη από 33°C , "moderate heatwave" όταν η θερμοκρασία είναι από 33°C έως 36°C και "severe heatwave" όταν η θερμοκρασία είναι μεγαλύτερη από 36°C . Σημειώνεται, ότι η κάθε ημέρα κατηγοριοποιείται σύμφωνα με την σοβαρότερη συνθήκη καιρού που υπήρξε μέσα στη μέρα.

Παρακάτω παρουσιάζεται η κατανομή των ημερών ανάλογα με την ένταση του καιρού.

Ημέρα	Normal	Moderate Heatwave	Severe Heatwave	Σύνολο
Δευτέρα	103	5	5	113
Τρίτη	106	9	2	117
Τετάρτη	111	7	1	119
Πέμπτη	106	8	3	117
Παρασκευή	110	7	2	119
Σάββατο	110	6	3	119
Κυριακή	102	10	3	115
Σύνολο	748	52	19	819

Πίνακας 3.1: Κατανομή δεδομένων ανά συνθήκη και ημέρα της εβδομάδας.

Συνεχίζοντας, χωρίστηκε το σύνολό στα 3 υποσύνολα: το σύνολο εκπαίδευσης (train set), το σύνολο επικύρωσης (validation set) και το σύνολο δοκιμής (test set). Αυτός ο διαχωρισμός γίνεται σχεδόν πάντα στις εκπαιδεύσεις μοντέλων [89] διότι πρέπει να υπάρχει ένα ξεκάθαρο σύνολο δεδομένων (το σύνολο δοκιμής), τα δεδομένα του οποίου να μην έχουν συμπεριληφθεί στην εκπαίδευση του μοντέλου, έτσι ώστε οι μετρικές απόδοσης να μην είναι πολωμένες (biased). Το σύνολο επικύρωσης υπάρχει έτσι ώστε σε κάθε εποχή της εκπαίδευσης το μοντέλο να υπολογίζει εσωτερικά μετρικές απόδοσης έτσι ώστε να κρατηθούν τα βάρη της εποχής με το καλύτερο σκορ. Ο τρόπος με τον οποίο διαχωρίστηκαν τα δεδομένα παρουσιάζεται αναλυτικά παρακάτω, στο 3.2.1.2.1.

Τέλος, έγινε κανονικοποίηση μέσου όρου, για να μεταφερθούν όλα τα δεδομένα στην ίδια κλίμακα, και κβαντοποίηση των τιμών σε 1024 στάθμες, για να απαλειφθούν τα δεκαδικά. Αυτή η μέθοδος βελτιώνει τις αποδόσεις καθώς μειώνει αρκετά τις αποστάσεις των δεδομένων, με αποτέλεσμα το μοντέλο να μπορεί να καταλάβει καλύτερα τις συσχετίσεις. [90]. Σημειώνεται ότι κβαντοποίηση έγινε μόνο στο LLM μοντέλο επειδή και το TimeGAN και το GaussianCopula διαχειρίζονται καλά τα δεκαδικά ψηφία.

Παρακάτω εξηγείται αναλυτικά ο τρόπος διαχωρισμού του συνόλου στα τρία υποσύνολα.

3.2.1.2.1 Ο τρόπος διαχωρισμού στα υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής

Ο τρόπος διαχωρισμού (splitting) των δεδομένων έγινε με κριτήριο την διατήρηση ομοίων ποσοστών στις διαφορετικές περιπτώσεις που υπάρχουν. Οι διαφορετικές περιπτώσεις των δεδομένων έχουν να κάνουν με την ημέρα της εβδομάδας στην οποία βρίσκονται και με την ένταση του καιρού που επικρατεί. Για λόγους απλότητας, και λόγω ομοιότητας των δεδομένων, ορίστηκαν 3 τύποι ημέρας: η "καθημερινή", το "Σαβ-

3.2. Ανάλυση και προεπεξεργασία συνόλου δεδομένων

βατο” και η ”Κυριακή”. Επιπλέον, οι τύποι έντασης καιρού είναι επίσης τρεις: normal, moderate heatwave, severe heatwave.

Στον διαχωρισμό των υποσυνόλων επιδιώχθηκε τα σύνολα να είναι όσο το δυνατόν πιο συνεχόμενα, αφού τα δεδομένα μας είναι ακολουθιακά, και επιδιώχθηκε όλα τα δεδομένα της ίδιας μέρας (από 00:00 π.μ. μέχρι και 23:00 μ.μ.) να είναι στο ίδιο σύνολο.

Ο τρόπος που έγινε η επιλογή του test set, αρχικά, είναι παίρνοντας ένα κυλιόμενο παράθυρο μήκους 10% των συνολικών δεδομένων και διατρέχοντάς το σε όλο το σύνολο μέχρι να βρεθεί ένα σύνολο που να πληροί συγκεκριμένα κριτήρια. Τα κριτήρια ήταν να υπάρχει τουλάχιστον το 30% των δεδομένων των καυσώνων, να υπάρχει τουλάχιστον μία ημέρα (δηλαδή 24 τιμές που ανήκουν στην ίδια ημερολογιακή μέρα) από τις περιπτώσεις ”normal/moderate/severe sunday”, ”normal/moderate/severe saturday” και τουλάχιστον 4 μέρες της περίπτωσης ”severe weekday”. Προφανώς, είναι απαραίτητοι και οι υπόλοιποι συνδυασμοί (”normal/moderate weekday”) αλλά επειδή υπήρχαν ήδη αρκετά δείγματα από αυτούς, δεν μπήκαν ως κριτήριο στην επιλογή.

Η επιλογή του validation set ήταν παίρνοντας, επίσης, ένα κυλιόμενο παράθυρο μήκους 10% των συνολικών δεδομένων και το μόνο κριτήριο ήταν να περιέχει τουλάχιστον το 10% των δεδομένων των καυσώνων. Λόγω περιορισμένων δεδομένων δεν υπήρχε παράθυρο που να έχει τουλάχιστον μία μέρα από κάθε περίπτωση. Το train set περιέχει τα δεδομένα που έμειναν στο σύνολο δεδομένων μετά από την αφαίρεση των δύο υποσυνόλων.

Παρακάτω παρουσιάζονται πίνακες με τις κατανομές των δεδομένων μετά από τον διαχωρισμό στα τρία διακριτά υποσύνολα.

Υποσύνολο	Ημέρες	% Ημερών	Ημέρες καύσωνα	% Καύσωνα
Εκπαίδευσης (Train)	657	80.2%	36	50.7%
Επικύρωσης (Validation)	81	9.9%	10	14.1%
Δοκιμής (Test)	81	9.9%	25	35.2%
Σύνολο	819	100%	71	100%

Πίνακας 3.2: Κατανομή δεδομένων στα υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής.

Υποσύνολο	Καθημερινή		Σάββατο		Κυριακή		Σύνολο
	Moderate	Severe	Moderate	Severe	Moderate	Severe	
Εκπαίδευσης	20	5	2	2	6	1	36
Επικύρωσης	4	3	1	0	2	0	10
Δοκιμής	12	5	3	1	2	2	25
Σύνολο	36	13	6	3	10	3	71

Πίνακας 3.3: Κατανομή δεδομένων καύσωνα ανά υποσύνολο, ημέρα και ένταση.

3.2.1.2.2 Σύνολο Παραγωγής Δεδομένων

Ιδιαίτερη έμφαση πρέπει να δοθεί στο σύνολο παραγωγής δεδομένων, το οποίο εξαρτάται από το σύνολο δοκιμής (test set), καθώς είναι το υποσύνολό του, το οποίο χρησιμοποιείται για την παραγωγή (generation) των χρονοσειρών από όλα τα μοντέλα. Οι παραγόμενες χρονοσειρές αξιοποιούνται στη συνέχεια τόσο για την αξιολόγηση όσο

και για τον υπολογισμό των μετρικών. Υπενθυμίζεται ότι τα δεδομένα αυτά δεν έχουν χρησιμοποιηθεί στην εκπαίδευση των μοντέλων, γεγονός που βελτιώνει την αντικειμενικότητα της αξιολόγησης. Επομένως, είναι κρίσιμης σημασίας ο προσεκτικός ορισμός των ημερών που θα χρησιμοποιηθούν για την παραγωγή χρονοσειρών.

Αρχικά, από τα δεδομένα του συνόλου δοκιμής διατηρήθηκαν μόνο αυτά των καλοκαιρινών μηνών έτσι ώστε να παραχθούν πιο ρεαλιστικές χρονοσειρές. Το σύνολο δοκιμής περιέχει μέρες καύσωνα, αλλά και συνηθισμένες ("normal") μέρες όπως φαίνεται στον Πίνακα 3.2. Η παραγωγή των δεδομένων θα γίνει για να προσομοιαστούν σενάρια των δύο καυσώνων, επομένως όλα τα σενάρια καυσώνων που υπάρχουν ήδη στο σύνολο δοκιμής θα κρατηθούν στο σύνολο παραγωγής δεδομένων. Τα σενάρια, λοιπόν, διακρίνονται στις έξι κατηγορίες συνδυασμού τύπου ημέρας (καθημερινή, Σάββατο, Κυριακή) και τύπου καύσωνα (moderate, severe).

Επειδή η παραγωγή των σεναρίων θέλουμε να είναι και από ήδη υπάρχοντες καύσωνες, αλλά και από φυσιολογικές μέρες (αφού θέλουμε να προσομοιαστούν σενάρια καυσώνων ανεξάρτητα από την ένταση καύσωνα της εισόδου), και επειδή τα ήδη υπάρχοντα σενάρια καυσώνων στο σύνολο δοκιμής είναι ελάχιστα (βλ. πίνακα 3.3), προστέθηκαν και μέρες φυσιολογικής (normal) έντασης καιρού. Όπως αναφέρθηκε και παραπάνω αυτές οι φυσιολογικές μέρες ανήκουν ήδη σε μέρες καλοκαιριού.

Τα κριτήρια για την προσθήκη των φυσιολογικών ημερών ήταν να προστεθούν όσες φυσιολογικές μέρες του ίδιου τύπου ημέρας υπάρχουν σε κάθε περίπτωση έντασης καύσωνα, έτσι ώστε καμία περίπτωση να μην υπερβεί τις 14 χρονοσειρές. Για παράδειγμα, αν η περίπτωση (moderate heatwave, weekday) είχε 9 χρονοσειρές, τότε από τις υπόλοιπες normal χρονοσειρές που ανήκουν στον τύπο ημέρας weekday θα προστεθούν μέχρι 5 ακόμα (αν υπάρχουν).

Με αυτόν τον τρόπο, υπάρχει ένα ενιαίο σύνολο παραγωγής δεδομένων, το οποίο αποτελείται από διαφορετικά διακριτά σενάρια όπου κάθε ένα από αυτά ανήκει σε κάποια από τις 6 περιπτώσεις και όλα τα μοντέλα θα παράξουν σενάρια με βάση αυτό. Παρακάτω φαίνεται η κατανομή των σεναρίων στις διαφορετικές περιπτώσεις:

Origin	Weekday		Saturday		Sunday		Total
	Moderate	Severe	Moderate	Severe	Moderate	Severe	
Heatwave	11	5	3	1	2	2	24
Normal	3	9	4	4	3	3	26
Total	14	14	7	5	5	5	50

Πίνακας 3.4: Κατανομή τελικών σεναρίων του συνόλου παραγωγής δεδομένων.

3.2.2 Καιρικά δεδομένα

Τα καιρικά δεδομένα πάρθηκαν από τον Κοπέρνικο ([86]) και αφορούν την θερμοκρασία (2μ από την επιφάνεια) της Γης τα έτη 2021-2024. Οι συντεταγμένες των δεδομένων είναι η περιοχή που βρίσκεται ανάμεσα στις 42.41° - 42.56° γεωγραφικό πλάτος και 12.49° - 12.64° γεωγραφικό μήκος και αφορούν την τοποθεσία της πόλης Τέρνι της Ιταλίας, αφού εκεί βρίσκονται οι εγκαταστάσεις της ASM. Τα αρχικά δεδομένα ήταν σε Κέλβιν και έγινε μετατροπή τους σε Κελσίου. Επιπλέον, έγινε μετατροπή των ωρών των δεδομένων από Συντονισμένη Παγκόσμια Ώρα (UTC) σε ώρα Ιταλίας για να υπάρχει απευθείας σύνδεση με τα δεδομένα ενέργειας, τα οποία είναι στην τοπική ώρα Ιταλίας. Τέλος, πραγματοποιήθηκε επίσης κανονικοποίηση μέσου

όρου και κβαντοποίηση, αλλά αυτή τη φορά σε 256 στάθμες γιατί οι αποκλίσεις ήταν μικρότερες.

3.3 Επιλογή Μοντέλων

Τα μοντέλα που επιλέχθηκαν ήταν το T5 μοντέλο, το TimeGAN και το Gaussian Copula της Synthetic Data Vault (SDV), ώστε να καλυφθούν διαφορετικοί τρόποι παραγωγής χρονοσειρών.

Αρχικά, επιλέχθηκε το T5 ως γλωσσικό μοντέλο, λόγω της εύκολης αρχιτεκτονικής του κωδικοποιητή-αποκωδικοποιητή [45]. Είναι ένα ανοιχτού κώδικα μοντέλο με επαρκές documentation και εύκολη χρήση μέσω της πλατφόρμας του HuggingFace [91]. Χρησιμοποιούνται εύκολα prompt-based μέθοδοι, κάνοντάς το ευέλικτο στην παραγωγή των εισόδων. Τέλος, λόγω μεθόδων όπως το LoRA/QLoRA η μετ-εκπαίδευση (fine-tuning) του γίνεται γρήγορη και σε GPUs με περιορισμένους πόρους.

Η επιλογή του TimeGAN έγινε λόγω της μεγάλης ανάπτυξης των GANs στην παραγωγή συνθετικών δεδομένων. Συγκεκριμένα, το TimeGAN είναι ένα state-of-the-art μοντέλο το οποίο εξειδικεύεται πλήρως στην παραγωγή συνθετικών χρονοσειρών [53].

Το Gaussian Copula μοντέλο της Synthetic Data Vault (SDV) επιλέχθηκε με σκοπό να υπάρχει ένα στατιστικό μοντέλο σύγκρισης, ως baseline. Το SDV είναι ένα εύκολο, προς τη χρήση, εργαλείο που δημιουργεί συνθετικά δεδομένα με έτοιμες συναρτήσεις που απαιτούν ελάχιστο χρόνο στην εκπαίδευση. Η Gaussian Copula προτιμήθηκε από τα υπόλοιπα μοντέλα της SDV γιατί αποτυπώνει με απλό και ερμηνεύσιμο τρόπο τις εξαρτήσεις μεταξύ μεταβλητών ενώ διατηρεί τις περιθωριακές κατανομές [27] [55].

3.4 Εκπαίδευση Μοντέλων

Σε αυτήν την ενότητα γίνεται λεπτομερής ανάλυση της εκπαίδευσης όλων των μοντέλων που χρησιμοποιήθηκαν για να παραχθούν τα συνθετικά δεδομένα. Τόσο το βασικό μοντέλο (το μοντέλο LLM) όσο και τα άλλα δύο μοντέλα σύγκρισης χρειάστηκαν κάποια μορφή μετεκπαίδευση έτσι ώστε να παράγουν τις απαραίτητες χρονοσειρές με τα επιθυμητά χαρακτηριστικά.

3.4.1 Μεγάλο Γλωσσικό Μοντέλο - LLM

Το βασικό αντικείμενο της παρούσας εργασίας ήταν η ανάπτυξη ενός γλωσσικού μοντέλου το οποίο θα μπορεί να δημιουργεί συνθετικά δεδομένα εξαρτώμενα από τις καιρικές συνθήκες που θα του ζητηθούν. Η διαδικασία της εκπαίδευσης του μοντέλου φαίνεται στο κάτω μέρος του Σχήματος 3.1. Παρακάτω ακολουθεί λεπτομερής ανάλυση όλων των επιμέρους σταδίων.

Το Μεγάλο Γλωσσικό Μοντέλο (LLM) που χρησιμοποιήθηκε ήταν το Google T5-base model το οποίο έχει 220 εκατομμύρια παραμέτρους. Για να μπορέσει να το χρησιμοποιηθεί για τον σκοπό αυτό χρειάζεται να γίνει μετ-εκπαίδευση (fine-tune). Δηλαδή, να προσαρμοστεί στο συγκεκριμένο πρόβλημα. Ως γλωσσικό μοντέλο, τόσο η είσοδος όσο και η έξοδος του, είναι κείμενο ή λέξεις, επομένως χρειάζεται να μετατραπούν τα δεδομένα στην κατάλληλη μορφή, την μορφή κειμένου (text form).

3.4.1.1 Μετατροπή σε μορφή κειμένου

Έχοντας έτοιμα τα δεδομένα ενέργειας και θερμοκρασίας, το επόμενο βήμα είναι η μετατροπή τους σε μορφή κειμένου ώστε να μπουν στο μοντέλο. Η μετατροπή έγινε παίρνοντας ένα κυλιόμενο παράθυρο σαρανταοκτάωρων συνεχόμενων τιμών και από αυτό πάρθηκαν οι απαραίτητες πληροφορίες που αναφέρονται παρακάτω.

Το κείμενο αποτελείται από δύο μέρη, το κείμενο εισόδου (input text) και το κείμενο στόχος (target text). Και τα δύο κείμενα θα μπουν στην είσοδο, αλλά το κείμενο εισόδου αφορά τις αρχικές πληροφορίες της εισόδου (τιμές, θερμοκρασία, μέρες, τύπος μετρητή), ενώ το κείμενο στόχος είναι το κείμενο που θα πρέπει να βγάλει το μοντέλο τελικά στην έξοδο.

Το κείμενο εισόδου είναι της μορφής:

"Past values: ... Past days: ... Past temperatures: ... Data: Active/Reactive Power. Future Condition: ... Predict next:".

Από όπου έχουμε:

- "Past Values": περιέχει τις 24 συνεχόμενες τιμές ισχύος από το πρώτο μισό του κυλιόμενου παραθύρου (αρχικές τιμές).
- "Past days": περιέχει την ημερολογιακή μέρα (Δευτέρα - Κυριακή) που βρίσκεται η κάθε μία από τις προηγούμενες 24 τιμές αντίστοιχα.
- "Past Temperatures": περιέχει τις θερμοκρασίες που είχαν οι προηγούμενες 24 τιμές ισχύος αντίστοιχα.
- "Data": περιέχει το είδος των δεδομένων (όνομα μετρητή, ενεργός/άεργος ισχύς)
- "Future Condition": αναφέρεται στην κατηγορία έντασης καιρού που είχε η ημέρα στην οποία ανήκει η επόμενη (25η), και μελλοντική, τιμή
- "Predict next": φράση που μπαίνει στο τέλος του κειμένου εισόδου για να προτρέψει το μοντέλο στην αναγνώριση της εργασίας.

Το κείμενο στόχος, αποτελείται από τις 24 τελευταίες τιμές του κυλιόμενου παραθύρου. Περιλαμβάνει τις αμέσως επόμενες τιμές της ημέρας που βρίσκεται στο κείμενο εισόδου. Έτσι, το μοντέλο θα μάθει την σχέση μεταξύ εισόδου και εξόδου, καθώς και θα μάθει τον επιθυμητό τρόπο απάντησης/εξόδου, δηλαδή 24 σκέτες τιμές, χωρίς λέξεις ή κάποιον άλλον χαρακτήρα.

Η επιλογή των παραπάνω παραμέτρων εισόδου έγινε με σκοπό να εκπαιδευτεί ένα μοντέλο που να μπορεί να καταλάβει την σχέση μεταξύ των τιμών ισχύος με την θερμοκρασία που επικρατεί και την ημερολογιακή μέρα που βρίσκεται. Η ύπαρξη της παραμέτρου "Past Days" είναι για να καταλάβει το μοντέλο τόσο την περιοδικότητα μέσα στην ημέρα, όσο και τις διαφορές στις τιμές των καθημερινών με αυτές των Σαββατοκύριακων. Επιπλέον, έχοντας ως έξοδο μόνο τις 24 επόμενες τιμές ισχύος, θα καταλάβει την επιθυμητή έξοδο. Τέλος, πρέπει να σημειωθεί ότι η επιλογή της παραμέτρου "Future Condition" έχει μπει ως μελλοντική μεταβλητή γνωστή εκ των προτέρων (future covariate) και είναι αυτή που οδηγεί το μοντέλο στην παραγωγή τιμών με συνθήκη (conditional generation). Έτσι, το μοντέλο θα μάθει να παράγει δεδομένα τόσο χρησιμοποιώντας ιστορικές τιμές όσο και μελλοντικές συνθήκες.

Έπειτα από αυτή τη διαδικασία, γίνονται δύο επιπλέον βήματα προτού ξεκινήσει η μετεκπαίδευση. Πρώτα, γίνεται αύξηση (augmentation) των δεδομένων του συνόλου

εκπαίδευσης που αναφέρονται σε μελλοντική κατάσταση καύσωνα. Η διαδικασία αύξησης έγινε με απλή αντιγραφή των δεδομένων και προσθήκη 3 φορές του καθενός στο σύνολο. Αυτό έγινε για να εξισορροπηθεί λίγο καλύτερα η αναλογία των κανονικών δεδομένων με τα δεδομένα του καύσωνα, αφού τα δεδομένα καύσωνα είναι αρκετά λιγότερα, όπως φαίνεται στον Πίνακα 3.2. Στην συνέχεια, γίνεται tokenization χρησιμοποιώντας τον T5 Tokenizer, καθώς αυτήν την μορφή αναγνωρίζει το μοντέλο.

Η επεξεργασία των δεδομένων ολοκληρώθηκε και είναι, πλέον, έτοιμο το μοντέλο να εκπαιδευτεί.

3.4.1.2 Εκπαίδευση Μοντέλου

Το μοντέλο που χρησιμοποιήθηκε είναι το Google/t5-base το οποίο αποτελείται από 220 εκατομμύρια παραμέτρους και ανήκει στην οικογένεια των T5 μοντέλων που αναλύθηκε στην Υποενότητα 2.5.1. Χρησιμοποιήθηκε μέσω της βιβλιοθήκης transformers της python, η οποία είναι βασισμένη στα μοντέλα και τις διαδικασίες της πλατφόρμας Hugging Face [91].

Πριν από οποιονδήποτε ορισμό παραμέτρων, τέθηκε μία παγκόσμια τυχαία τιμή (global seed) στην τιμή 420 έτσι ώστε να υπάρχει αναπαραγωγιμότητα της εκπαίδευσης.

Στο μοντέλο εφαρμόστηκε το QLoRA, με κβαντοποίηση 4-bit και τις εξής παραμέτρους:

- `lora_dropout = 0.2`. Τυχαία απόρριψη του 20% των ενεργοποιήσεων του LoRA για βελτίωση της γενίκευσης.
- `r = 32`. Η βαθμίδα των πινάκων χαμηλής τάξης. Καθορίζουν την αποδοτικότητα και χωρητικότητα των παραμέτρων του προσαρμογέα.
- `lora_alpha = 64`. Συντελεστής κλιμάκωσης που ενισχύει τις ενημερώσεις του LoRA πριν ενσωματωθούν ξανά στα βασικά βάρη.
- `target_modules = ["q", "v"]`. Οι προσαρμογείς LoRA εφαρμόζονται μόνο στους πίνακες προβολής query (Q) και value (V).

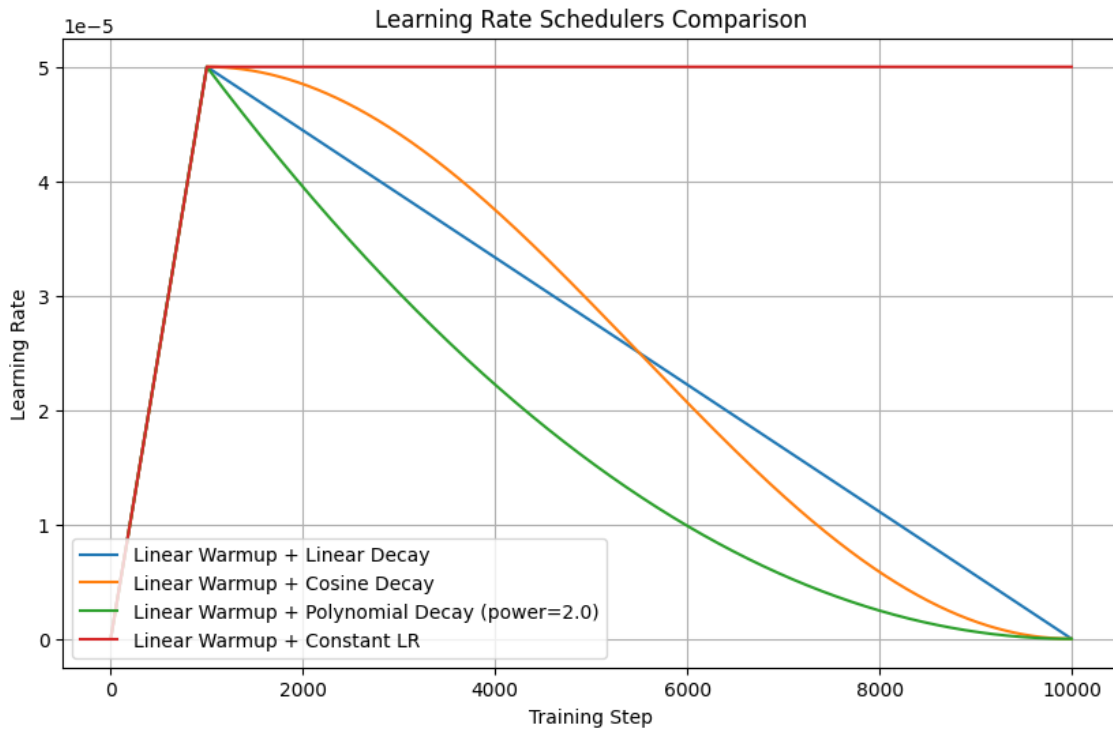
Όπου κάθε παράμετρος παρουσιάζεται και αναλύεται στο [43].

Οι παράμετροι του εκπαιδευτή (Trainer) του μοντέλου εμφανίζονται παρακάτω:

- `number of epochs = 20`. Πόσες εποχές θα τρέξει το μοντέλο αν δεν το σταματήσει το early stopping.
- `learning scheduler type = cosine`. Το είδος της μεταβολής του βαθμού μάθησης. Δες σχήμα 3.7.
- `learning rate =  $5 \times 10^{-5}$` . Ο βαθμός μάθησης
- `warmup steps = 200`. Αρχικά βήματα όπου τα βάρη θα μεταβάλλονται αργά μέχρι να φτάσουν την τελική τιμή. Δες σχήμα 3.7.
- `early stopping patience = 3`. Αριθμός εποχών στις οποίες θα σταματήσει το μοντέλο αν δεν έχει υπάρξει βελτίωση του μοντέλου.
- `random seed = 420`. Η τυχαία τιμή για να εξασφαλιστεί η αναπαραγωγιμότητα.

- `optimizer = adamw_torch`. Χρήση του βελτιστοποιητή AdamW, ο οποίος χρησιμοποιεί τον αλγόριθμο Adam, και προσαρμόζει τους ρυθμούς μάθησης για κάθε παράμετρο και αποτρέπει την υπερβολική αύξηση των βαρών [92].
- `metric_for_best_model="eval_loss"`. Η χρήση της μετρικής `evaluation_loss` ως την μετρική ορισμού καλύτερου μοντέλου. Αναλυτικότερα παρακάτω.
- `evaluation strategy = "epoch"`. Το μοντέλο θα υπολογίζει τις μετρικές (`evaluation loss`, `training loss`) στο τέλος κάθε εποχής.

Ο ρυθμός μάθησης στην παρούσα εργασία δεν ήταν σταθερός, αλλά μεταβαλλόταν. Στο σχήμα 3.7 φαίνεται η ακριβής μεταβολή ενός υποθετικού μοντέλου όπου έχει συνολικά 10000 βήματα και ρυθμό μάθησης 5×10^{-5} . Στο σχήμα παρουσιάζονται τα τέσσερα πιο γνωστά είδη των schedulers καθώς και δείχνεται η εφαρμογή του warmup steps με 1000 βήματα. Στην εργασία χρησιμοποιήθηκε ο cosine καθώς κρίθηκε ότι μειώνεται πιο ομαλά.



Σχήμα 3.7: Σύγκριση των ειδών schedulers του ρυθμού μάθησης

Μετά από κάθε εποχή, το μοντέλο αξιολογείται στο validation dataset, το οποίο δεν χρησιμοποιείται για εκπαίδευση, με σκοπό τον υπολογισμό της μετρικής `validation_loss` (`eval_loss` στις παραμέτρους του Trainer). Η αξιολόγηση πραγματοποιείται με τη μέση συνάρτηση απώλειας διασταυρούμενης εντροπίας (`cross-entropy loss`), η οποία συγκρίνει τις προβλεπόμενες πιθανότητες του μοντέλου σε κάθε token με τα πραγματικά tokens-στόχους [45]. Ο μαθηματικός τύπος που χρησιμοποιείται είναι:

$$\mathcal{L}_{val} = -\frac{1}{N} \sum_{i=1}^N \frac{1}{T_i} \sum_{t=1}^{T_i} \log P_{\theta}(y_{i,t} | y_{i,<t}, x_i)$$

όπου:

- N : ο αριθμός των ακολουθιών στο validation dataset
- T_i : το μήκος (αριθμός tokens) της ακολουθίας-στόχου για το στοιχείο i
- $y_{i,t}$: το σωστό token στη χρονική στιγμή t για το στοιχείο i
- x_i : η είσοδος (input sequence) για το στοιχείο i
- $P_\theta(y_{i,t} | y_{i,<t}, x_i)$: η πιθανότητα που αποδίδει το μοντέλο (με παραμέτρους θ) στο σωστό token, δεδομένων της εισόδου και των προηγούμενων tokens

Η εποχή που έχει το μικρότερο validation loss είναι και αυτή που αποθηκεύει το μοντέλο στο τέλος της εκπαίδευσης. Με την ύπαρξη του Early Stopping, αν περάσει ένας συγκεκριμένος αριθμός εποχών και η μετρική δεν έχει βελτιωθεί, τότε η εκπαίδευση σταματάει και το μοντέλο αποθηκεύει την εποχή με την καλύτερη μέτρηση. Έτσι, ολοκληρώθηκε η εκπαίδευση του LLM μοντέλου

3.4.2 TimeGAN Μοντέλο

Το πρώτο μοντέλο σύγκρισης που εξετάστηκε ήταν το TimeGAN μοντέλο. Παρακάτω παρουσιάζεται αναλυτικά η διαδικασία εκπαίδευσης (training) του μοντέλου.

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση είναι μόνο αυτά του συνόλου εκπαίδευσης. Σε αυτό το μοντέλο δεν γίνεται χρήση του συνόλου επικύρωσης, οπότε αγνοήθηκε ώστε να εκπαιδευτούν όλα τα μοντέλα με το ίδιο σύνολο.

Τα δεδομένα μπήκαν στο TimeGAN ως 3 στήλες, η πρώτη αφορά τα δεδομένα ενεργού ισχύος, η δεύτερη τα δεδομένα άεργου ισχύος και η τρίτη τα δεδομένα θερμοκρασίας. Επίσης, τα δεδομένα μπαίνουν ως ομάδες των 24ων τιμών (διάσταση δεδομένων), δηλαδή ανά μέρα. Για την αξιοποίηση των περιπτώσεων της έντασης καιρού και τύπο ημέρας αποφασίστηκε να χρησιμοποιηθεί η τεχνική των Conditional GANs (CGANs) η οποία προσαρμόστηκε στο TimeGAN. Με αυτήν την τεχνική ο τύπος ημέρας και η ένταση καιρού προστέθηκαν ως εξωτερικές πληροφορίες οι οποίες μπαίνουν στο μοντέλο ως συνθήκη, όπως εξηγήθηκε στο θεωρητικό υπόβαθρο. Ο τύπος μέρας μπήκε με κωδικοποίηση one-hot, δηλαδή τρεις διαστάσεις [1,0,0], [0,1,0], [0,0,1], για τις τρεις διαφορετικές και χωρίς εξάρτηση μέρες, ενώ η ένταση του καιρού μπήκε ως κωδικοποίηση integer, δηλαδή μία διάσταση και ο αριθμός 1 για moderate heatwave και 2 για severe, διότι υπάρχει συσχέτιση (το 2 είναι πιο ισχυρό από το 1).

Στην συνέχεια ξεκίνησε η εκπαίδευση των επιμέρους στοιχείων. Αρχικά, έγινε η εκπαίδευση του Ενσωματωτή (Embedder) και του Επόπτη (Supervisor), το στοιχείο στο latent space που καταλαβαίνει τις χρονικές συσχετίσεις, με τις παραμέτρους:

- $z_dim = 24$. Οι διαστάσεις του τυχαίου διανύσματος θορύβου Z . Ίδιες με την διάσταση των δεδομένων.
- $hidden_dim = 24$. Η διάσταση των κρυμμένων καταστάσεων.
- $batch_size = 32$. Αριθμός των ακολουθιών που επεξεργάζονται
- $train_steps = 4000$. Αριθμός βημάτων.

Τέλος, ακολούθησε η ταυτόχρονη εκπαίδευση των Recovery, Generator και Discriminator με παραμέτρους:

- `train_steps = 17000`. Αριθμός συνολικών βημάτων.
- `warmup_steps = 1000`. Αριθμός πρώτων βημάτων που εκπαιδεύεται μόνο ο Discriminator.

Σε όλες τις παραπάνω εκπαιδεύσεις έχει οριστεί ο τυχαίος αριθμός (random seed) 42 στις αρχικοποιήσεις, για την εξασφάλιση της αναπαραγωγιμότητας. Έτσι, ολοκληρώθηκε και η εκπαίδευση του TimeGAN μοντέλου.

3.4.3 Στατιστικό Μοντέλο Gaussian

Το δεύτερο και τελευταίο μοντέλο σύγκρισης που εξετάστηκε ήταν το Gaussian Copula μοντέλο της SDV Metrics. Το SDV θέλει το αρχείο που θα μπει προς εκπαίδευση να είναι σε μορφή πίνακα με στήλες, επομένως μετατρέψαμε το σύνολο δεδομένων εκπαίδευσης σε τέτοια μορφή, με 24 στήλες για τις 24 ωριαίες τιμές ισχύος, 24 στήλες για τις 24 τιμές θερμοκρασίας, μία στήλη με είδος μέρας (καθημερινή, Σάββατο ή Κυριακή), μία στήλη με την συνθήκη έντασης καιρού και μία στήλη με το είδος ισχύος των τιμών (ενεργός ή άεργος). Άρα δημιουργήθηκε ένας πίνακας 51 στηλών και σειρών όσο είναι τα δεδομένα. Σημειώνεται ότι, όπως και στο TimeGAN, για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε μόνο το train set και όχι και το validation set.

Το επόμενο βήμα της διαδικασίας ήταν το φιλτράρισμα των τιμών που αναφέρονται σε κανονικές συνθήκες θερμοκρασίας. Αυτό έγινε για να είναι τα δεδομένα όσο πιο κοντά γίνεται στα δεδομένα καυσώνων και να μην επηρεαστεί από τιμές διαφορετικών ημερών, αφού είναι στατιστικό. Επιπλέον, έτσι μπορεί να συναγωνιστεί λίγο καλύτερα με τα υπόλοιπα έξυπνα μοντέλα.

Για την εκπαίδευση με το SDV αρχικά χρειάζεται η δημιουργία του αρχείου metadata το οποίο περιέχει πληροφορίες για το είδος των τιμών των στηλών. Στην περίπτωσή μας οι 24 τιμές ισχύος και θερμοκρασίας θεωρήθηκαν numerical ενώ οι άλλες τρεις στήλες με το είδος ημέρας, είδος ισχύος και την συνθήκη καιρού θεωρήθηκαν categorical.

Στην συνέχεια γίνεται η εκπαίδευση του GaussianCopulaSynthesizer με τις εξής παραμέτρους:

- `enforce_rounding=False`. Να μην γίνεται στρογγυλοποίηση.
- `enforce_min_max_values=False`. Να μην οριοθετηθεί από τα μέγιστα και ελάχιστα του συνόλου.
- `default_distribution = gaussian_kde`. Η Gaussian κατανομή

Τέλος, έγινε η εκπαίδευση (fit) του synthesizer και το μοντέλο είναι, πλέον, έτοιμο.

3.5 Εξαγωγή (Inference) Αποτελεσμάτων και Αξιολόγηση Μοντέλων

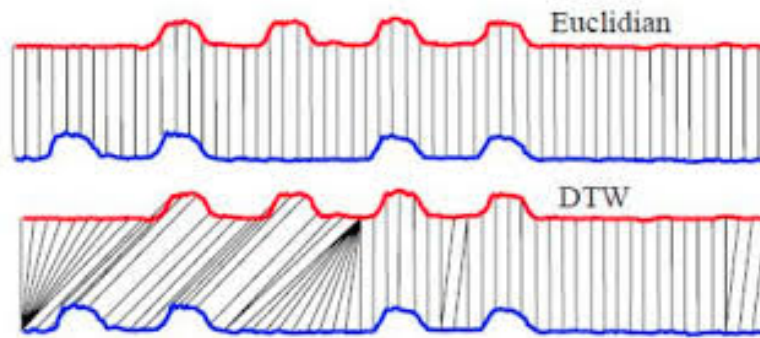
Η αξιολόγηση των μοντέλων έγινε χρησιμοποιώντας το σύνολο δεδομένων δοκιμής (test set) που αναλύθηκε στο 3.2.1.2.2. Αυτό το σύνολο δεν συμμετείχε καθόλου στην διαδικασία της εκπαίδευσης επομένως δεν υπάρχει το πρόβλημα της πόλωσης (bias). Η γενική ιδέα είναι ότι σύμφωνα με το σύνολο αυτό, παράγονται δεδομένα από τα μοντέλα και αυτά συγκρίνονται με τις πραγματικές (ιστορικές) τιμές του συνόλου με

την χρήση κάποιων μετρικών. Παρακάτω παρουσιάζονται οι μετρικές αξιολόγησης και στην συνέχεια ο τρόπος εξαγωγής (inference) αποτελεσμάτων και αξιολόγησης κάθε μοντέλου.

3.5.1 Μετρικές αξιολόγησης

3.5.1.1 Dynamic Time Warping

Το Dynamic Time Warping (DTW) είναι ένας κλασικός αλγόριθμος για τη μέτρηση της ομοιότητας μεταξύ δύο χρονοσειρών [93], επιτρέποντας «ελαστικές» ευθυγραμμίσεις στον άξονα του χρόνου. Χρησιμοποιείται ευρέως σε εφαρμογές χρονοσειρών όπως η ομαδοποίηση, η ταξινόμηση και η ανίχνευση ανωμαλιών. Σε αντίθεση με απλά μέτρα απόστασης (π.χ. Ευκλείδεια απόσταση), το DTW μπορεί να ταιριάζει ακολουθίες που έχουν παρόμοιο σχήμα αλλά παρουσιάζουν μετατόπιση, παραμόρφωση ή διαφορετική ταχύτητα στον χρόνο (Σχήμα 3.8). Λειτουργεί έχοντας έναν πίνακα κόστους από τις αποστάσεις μεταξύ δύο σημείων των ακολουθιών και στη συνέχεια βρίσκει τη βέλτιστη διαδρομή παραμόρφωσης που ελαχιστοποιεί το συνολικό κόστος ευθυγράμμισης [93]. Στην παρούσα εργασία χρησιμοποιείται για να μετρήσει κατά πόσο οι παραγόμενες χρονοσειρές είναι όμοιες με τις αληθινές, έστω και με μετατόπιση η παραμόρφωση στον χρόνο. Στο παρακάτω σχήμα φαίνεται οπτικά ο υπολογισμός της DTW.



Σχήμα 3.8: Σύγκριση DTW με Ευκλείδεια Απόσταση

3.5.1.2 Maximum Mean Discrepancy

Το MMD είναι μία στατιστική μετρική που ποσοτικοποιεί την απόσταση μεταξύ δύο κατανομών πιθανότητας βάσει των δειγμάτων τους. Έστω δείγματα $\{x_i\}_{i=1}^N \sim P$ από τα πραγματικά δεδομένα και $\{y_j\}_{j=1}^M \sim Q$ από τα παραγόμενα δεδομένα, το MMD ορίζεται ως:

$$\text{MMD}^2(P, Q) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(x_i, x_j) - \frac{2}{NM} \sum_{i=1}^N \sum_{j=1}^M k(x_i, y_j) + \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M k(y_i, y_j),$$

όπου $k(\cdot, \cdot)$ είναι μια θετικά ορισμένη συνάρτηση πυρήνα (kernel function), η οποία συνήθως επιλέγεται ως Γκαουσιανή συνάρτηση βάσης (RBF). Αυτή η διατύπωση ενσωματώνει και τις δύο κατανομές σε έναν αναπαραγόμενο χώρο Hilbert (RKHS) και μετρά την τετραγωνική απόσταση μεταξύ των μέσων ενσωματώσεών τους [94]. Στο

πλαίσιο των χρονοσειρών, κάθε ακολουθία μπορεί να θεωρηθεί ως δείγμα, επιτρέποντας στο MMD να συγκρίνει όχι μόνο βασικά στατιστικά μεγέθη όπως ο μέσος όρος και η διακύμανση, αλλά και υψηλότερης τάξης ομοιότητες. Προηγούμενες έρευνες στη παραγωγή δεδομένων χρονοσειρών έχουν δείξει ότι το MMD αποτυπώνει αποτελεσματικά τη σύγκλιση των κατανομών μεταξύ συνθετικών και πραγματικών χρονοσειρών, καθιστώντας το ένα κατάλληλο μέτρο για την αξιολόγηση της πιστότητας των παραγόμενων δεδομένων στην παρούσα έρευνα [66] [62].

3.5.1.3 Pearson Correlation

Ο συντελεστής συσχέτισης Pearson μετρά τη δύναμη και την κατεύθυνση της γραμμικής σχέσης μεταξύ δύο μεταβλητών (ή χρονοσειρών), κανονικοποιώντας τις σε μηδενικό μέσο όρο και μοναδιαία διακύμανση και υπολογίζοντας τη κανονικοποιημένη συνδιακύμανση. Οι τιμές του κυμαίνονται από -1 (τέλεια αρνητική συσχέτιση) έως $+1$ (τέλεια θετική συσχέτιση), ενώ το 0 υποδηλώνει ότι δεν υπάρχει γραμμική σχέση. Στις χρονοσειρές, η συσχέτιση Pearson χρησιμοποιείται ευρέως ως μέτρο ομοιότητας: δύο σειρές θεωρούνται παρόμοιες αν η μία μπορεί να προκύψει από την άλλη μέσω γραμμικής μετατόπισης ή κλιμάκωσης [95]. Στο πλαίσιο της παρούσας εργασίας, ο συντελεστής Pearson χρησιμοποιείται για τη σύγκριση μεταξύ πραγματικών και παραγόμενων χρονοσειρών, παρέχοντας μια ποσοτική εκτίμηση της ομοιότητας τους.

3.5.2 Εξαγωγή (Inference) Αποτελεσμάτων

Αρχικά γίνεται η εξαγωγή (inference) αποτελεσμάτων για την οπτική αναπαράσταση των δεδομένων και για να υπάρχει ποιοτική σύγκριση των αποτελεσμάτων. Αυτή η εξαγωγή γίνεται μόνο μία φορά σε όλα τα μοντέλα.

3.5.2.1 Μεγάλο Γλωσσικό Μοντέλο

Για την εξαγωγή (inference) αποτελεσμάτων του γλωσσικού μοντέλου έγινε μετατροπή του συνόλου δοκιμής (test set) σε μορφή κειμένου, με τον τρόπο που παρουσιάζεται στο 3.4.1.1 και μπήκε το input text των 50 σεναρίων του συνόλου δοκιμής στην είσοδο του γλωσσικού μοντέλου. Οι παράμετροι τις εξαγωγής των αποτελεσμάτων εμφανίζονται παρακάτω:

- seed number = 17. Ο αριθμός τυχαίας αρχικοποίησης των βαρών.
- do_sample = True. Το μοντέλο δεν διαλέγει το πιο πιθανό token, αλλά τυχαία από ένα σύνολο πιθανών tokens.
- top_p = 0.9. Το μοντέλο θα διαλέξει tokens που βρίσκονται στο 90% των πιθανών διαθέσιμων επιλογών. Προσφέρει διαφορετικότητα στις εξόδους.
- temperature = 0.6. Αφού είναι μικρότερη του 1 το μοντέλο γίνεται πιο "συντηρητικό", δηλαδή θα προτιμηθούν τα tokens με την μεγαλύτερη πιθανότητα παρά αυτά με χαμηλή.

Η επιλογή αυτών των παραμέτρων έγινε με σκοπό το μοντέλο να παράγει ρεαλιστικές ισορροπημένες τιμές με λίγη διαφορετικότητα και όχι εντελώς συντηρητικές.

3.5.2.2 Μοντέλο TimeGAN

Για την εξαγωγή (inference) αποτελεσμάτων του μοντέλου TimeGAN χρησιμοποιήθηκε το σύνολο δεδομένων δοκιμής. Το TimeGAN, όπως και όλα τα GANs, δεν παίρνουν είσοδο τις τιμές όπως τα LLMs, γιατί μέσω της εκπαίδευσης έχουν μάθει το είδος της εξόδου που χρειάζεται να παράγουν. Παίρνουν μόνο τον αρχικό θόρυβο Z , ο οποίος λόγω της χρήσης των συνθηκών (conditions), αυτός θα επηρεαστεί από το σύνολο δοκιμής. Το σύνολο αυτό περιέχει 50 σενάρια όπου κάθε ένα από αυτά ανήκει σε μία από τις 6 περιπτώσεις συνδυασμού του τύπου ημέρας και του τύπου έντασης καιρού. Αυτά τα δύο χαρακτηριστικά αποτελούν εξωτερικές συνθήκες του μοντέλου, επομένως για κάθε ένα από τα 50 σενάρια, και με πλήρη αντιστοίχιση, στο TimeGAN μπήκε ο θόρυβος Z επηρεασμένος από τις περιπτώσεις του κάθε σεναρίου και αυτό παρήγαγε έξοδο με βάση αυτές τις συνθήκες. Για παράδειγμα, αν το πρώτο σενάριο αναφερόταν σε περίπτωση καθημερινής με ήπιο καύσωνα, τότε στο TimeGAN ο θόρυβος Z θα περιέχει την συνθήκη καθημερινής $([1,0,0])$ και του moderate heatwave (1) με αποτέλεσμα αυτό να παραγάγει την αντίστοιχη χρονοσειρά. Ο seed που χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων ήταν ο 42.

3.5.2.3 Μοντέλο Gaussian

Για την εξαγωγή (inference) των αποτελεσμάτων του μοντέλου Gaussian έγινε παρόμοια διαδικασία με αυτή του TimeGAN. Το SDV διαθέτει λειτουργία για παραγωγή δεδομένων με συγκεκριμένες σταθερές στήλες (συνθήκες) του πίνακα εκπαίδευσης, χωρίς κάποια συγκεκριμένη ενέργεια στην εκπαίδευση. Έτσι, έγινε η παραγωγή χρονοσειρών με συνθήκες τις 6 διαφορετικές περιπτώσεις του συνδυασμού των διαφορετικών τύπων ημερών και έντασης καιρού. Δηλαδή, για κάθε σενάριο του συνόλου δοκιμής, παράχθηκε ένα αντίστοιχης περίπτωσης σενάριο. Επομένως, παρήχθησαν συνολικά 50 νέα σενάρια με τον αλγόριθμο Gaussian Copula της SDV.

3.5.3 Αξιολόγηση Μοντέλων

Η αξιολόγηση του κάθε μοντέλου έγινε κάνοντας inference σε κάθε μοντέλο, χρησιμοποιώντας το σύνολο δοκιμής, και συγκρίνοντας τα αποτελέσματα με τα ιστορικά δεδομένα του συνόλου. Στο LLM και στο TimeGAN, για να είναι πιο πλήρης η αξιολόγηση, η εξαγωγή των αποτελεσμάτων έγινε για 100 διαφορετικές τυχαίες τιμές αρχικοποίησης (random seeds) που κυμαίνονται από το 1-9999. Για την εξασφάλιση της αναπαραγωγιμότητας η επιλογή των 100 τυχαίων τιμών έγινε με τον seed 42. Το Gaussian μοντέλο δεν χρειάστηκε αριθμό seed επειδή, ενώ βγάζει διαφορετικές εξόδους κάθε φορά που τρέχει δεδομένα ίδιας περίπτωσης, σε κάθε επανάληψη (rerun) του κώδικα βγάζει τις ίδιες, διαφορετικές μεταξύ τους, εξόδους.

Ο υπολογισμός των μετρικών σε όλα τα μοντέλα γίνεται ως εξής. Για κάθε seed, παράγονται 50 σενάρια που ανήκουν σε μία από τις 6 περιπτώσεις συνδυασμού τύπου ημέρας και σοβαρότητας καιρού. Ομαδοποιώντας όλα τα σενάρια των ίδιων περιπτώσεων δημιουργούνται 6 υποσύνολα σεναρίων, ένα για κάθε περίπτωση. Κάθε υποσύνολο περιλαμβάνει τα παραγόμενα σενάρια και τα ιστορικά σενάρια (το κείμενο του target text, δηλαδή η αληθινή (ιστορική) έξοδος). Κάθε ένα παραγόμενο σενάριο συγκρίθηκε, με όλες τις ιστορικές εξόδους του ίδιου υποσυνόλου (cross evaluation), με τις τρεις μετρικές που παρουσιάστηκαν προηγουμένως. Έτσι, υπάρχει μία συνολική μετρική για κάθε υποσύνολο/περίπτωση. Για την ολική μετρική, υπολογίστηκε ο μέσος όρος των μετρικών των 6 υποσυνόλων με βάρη 5/7 για τις περιπτώσεις με "καθημε-

ρινή” ως τύπο ημέρας, $1/7$ για τα ”Σαββατα” και $1/7$ για τις ”Κυριακές”. Αυτό έγινε για να εξισορροπηθεί η επίδραση των καθημερινών (που είναι περισσότερες) με τα Σαββατοκύριακα.

Επιπλέον, στα μοντέλα LLM και TimeGAN που έχουν πολλά seeds, έγινε ο μέσος όρος των μετρικών όλων των seeds. Έχοντας έτσι, μία τελική μετρική για κάθε μοντέλο. Τέλος, υπολογίστηκε ο μέσος όρος των μετρικών που αναφέρονται μόνο σε moderate heatwaves και μόνο σε severe heatwaves, για να δούμε τις μετρικές κάθε τύπου καύσωνα.

Κεφάλαιο 4

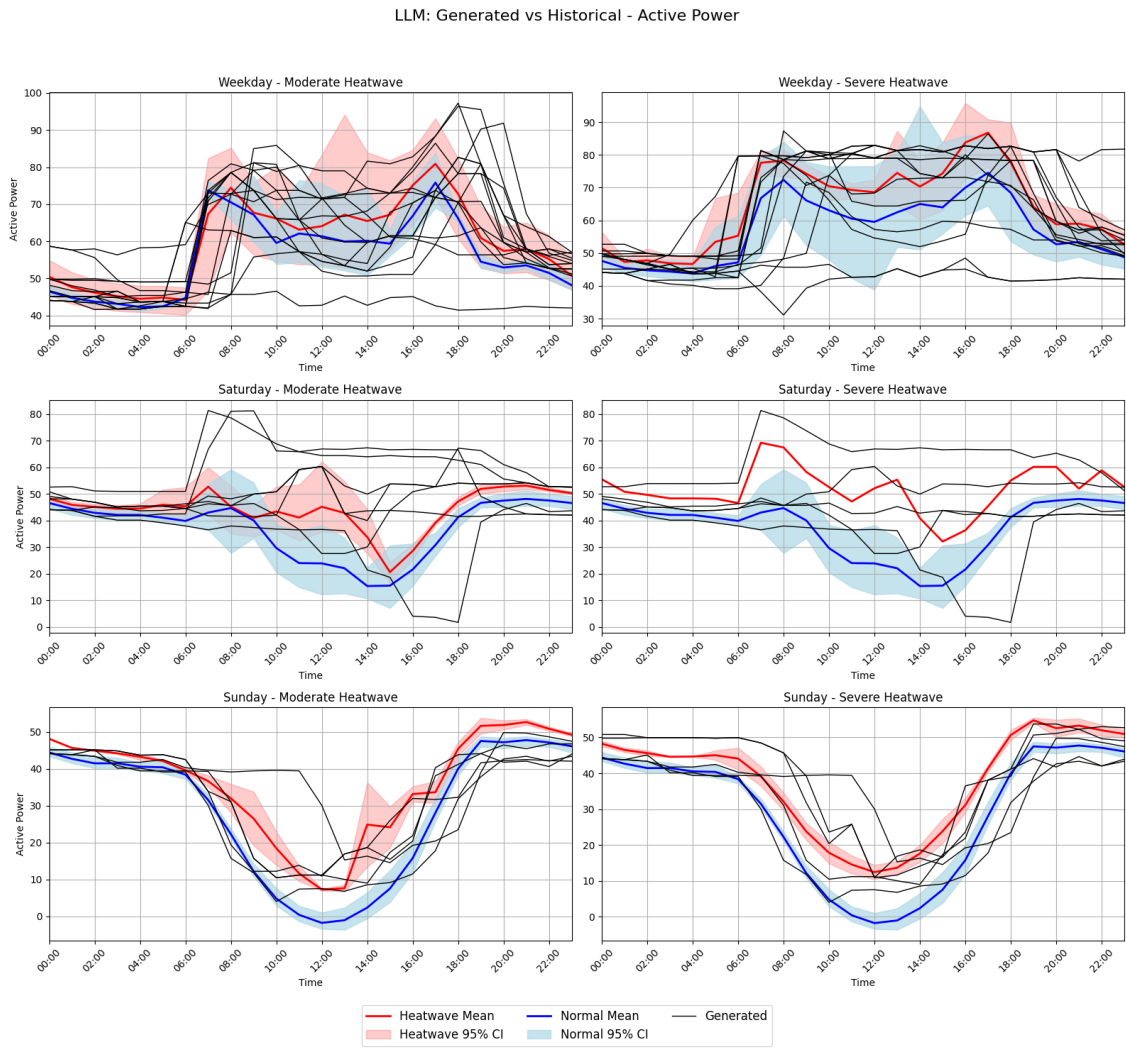
Αποτελέσματα Πειραμάτων

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα των μοντέλων. Αρχικά, παρουσιάζονται τα παραγόμενα δεδομένα των μοντέλων, που προέκυψαν στο στάδιο εξαγωγής (inference) αποτελεσμάτων, σε σύγκριση με τα ιστορικά δεδομένα των 50 σεναρίων του συνόλου δοκιμής, και στην συνέχεια, αναφέρονται, παρουσιάζονται και συγκρίνονται οι μετρικές του κάθε μοντέλου.

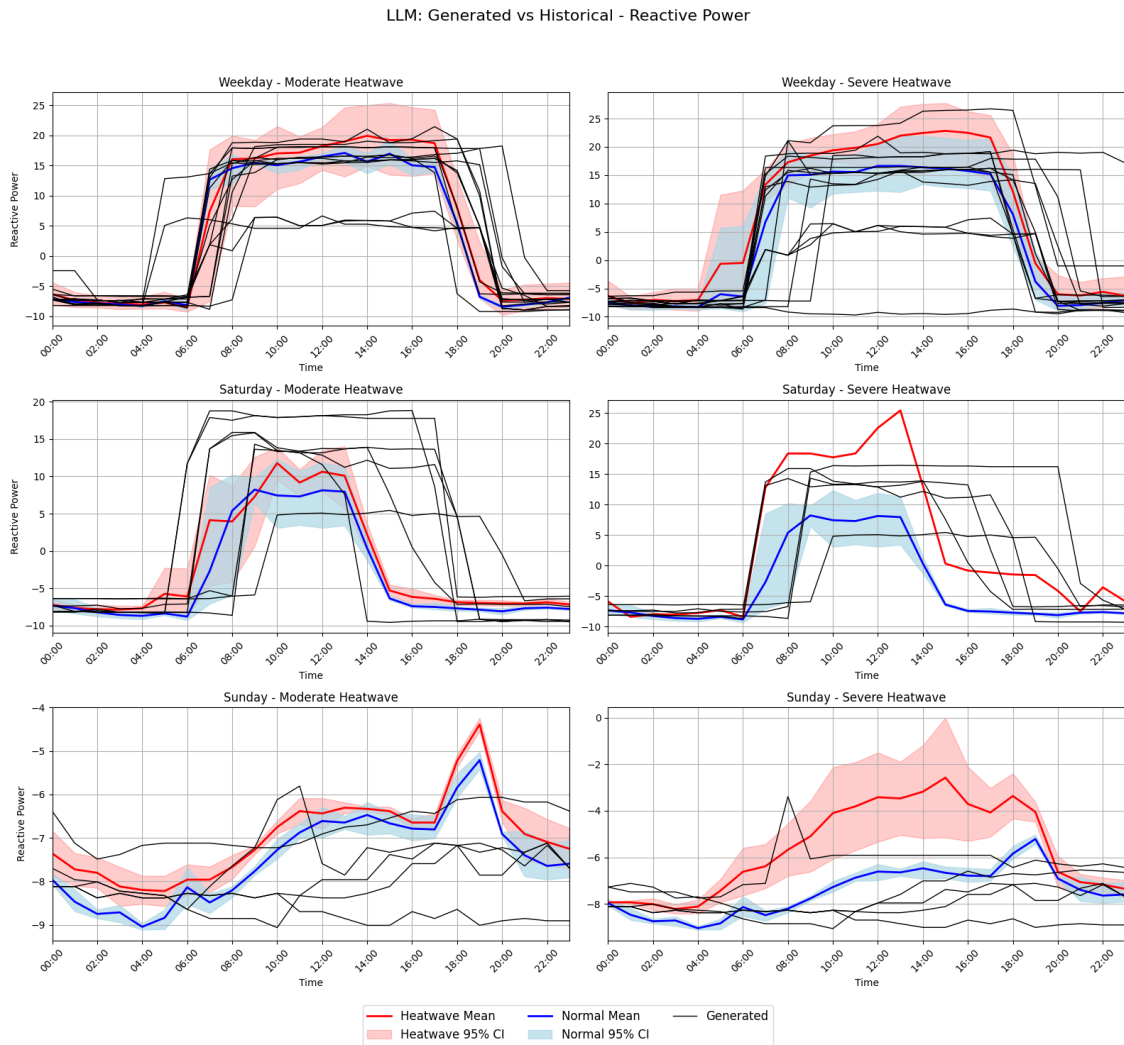
4.1 Αποτελέσματα του LLM

Παρακάτω παρουσιάζονται τα αποτελέσματα του Γλωσσικού Μοντέλου, βάζοντας ως είσοδο κάθε ένα από τα 50 σενάρια του συνόλου παραγωγής δεδομένων. Κατηγοριοποιημένα στις 6 διαφορετικές περιπτώσεις φαίνονται οι παραγόμενες χρονοσειρές ως συνεχόμενες γραμμές και οι αληθινές χρονοσειρές στο παρασκήνιο με την μορφή διαστήματος εμπιστοσύνης (confidence interval). Το κόκκινο διάστημα εμπιστοσύνης αναφέρεται στα ιστορικά δεδομένα που ανήκουν σε δεδομένα καύσωνα, ενώ το μπλε διάστημα αναφέρεται σε αυτά που ήταν φυσιολογικών ημερών αλλά τα προσθέσαμε και λόγω περιορισμένου συνόλου δεδομένων και γιατί θέλουμε να παραχθούμε δεδομένα καυσώνων ανεξάρτητα από την ένταση καιρού της εισόδου.

4.1. Αποτελέσματα του LLM



Σχήμα 4.1: Αποτελέσματα LLM Inference Ενεργού Ισχύος



Σχήμα 4.2: Αποτελέσματα LLM Inference Άεργου Ισχύος

Από τα παραπάνω διαγράμματα γίνεται αντιληπτό ότι το Γλωσσικό Μοντέλο έχει καταλάβει καλά τις εποχικότητες της ημέρας και γενικά συμμορφώνεται πολύ στη μορφή των ιστορικών δεδομένων, ειδικά στα δεδομένα ενεργού ισχύος. Στα δεδομένα άεργου ισχύος υπάρχουν περισσότερες αποκλίσεις, για τις οποίες εκτιμάται ότι ευθύνεται η ποιότητα των δεδομένων άεργου ισχύος. Σημειώνεται ότι στα δεδομένα των Κυριακών ενεργού ισχύος, παρόλο που ήταν ελάχιστα, εμφάνισε θετικά αποτελέσματα, γεγονός που δείχνει πως μπορεί να εκπαιδευτεί και με περιορισμένο σύνολο δεδομένων και λαμβάνει αρκετά υπόψη του τις διαφορετικές περιπτώσεις. Επίσης, από τα ιστορικά διαγράμματα παρατηρείται και η διαφορά μεταξύ της ισχύος των καυσώνων και των ημερών χωρίς καύσιμα. Η ισχύς των καυσώνων στις περιπτώσεις ενεργού ισχύος είναι 5kW (12%) παραπάνω στις καθημερινές moderate heatwave και 10kW (22%) παραπάνω στις καθημερινές severe heatwave, 5kW (14%) παραπάνω στα Σάββατα moderate και 25kW (45%) παραπάνω στα Σάββατα severe, αν και υπάρχει ένα μόνο Σάββατο σε αυτήν την περίπτωση, 4kW (8%) παραπάνω στις Κυριακές moderate και 6kW (11%) παραπάνω στις Κυριακές severe. Στην άεργο ισχύ έχουμε 4kW (13%) παραπάνω στις καθημερινές moderate και 6kW (18%) παραπάνω στις καθημερινές severe, 5kW (22%) παραπάνω στα Σάββατα moderate και 18kW

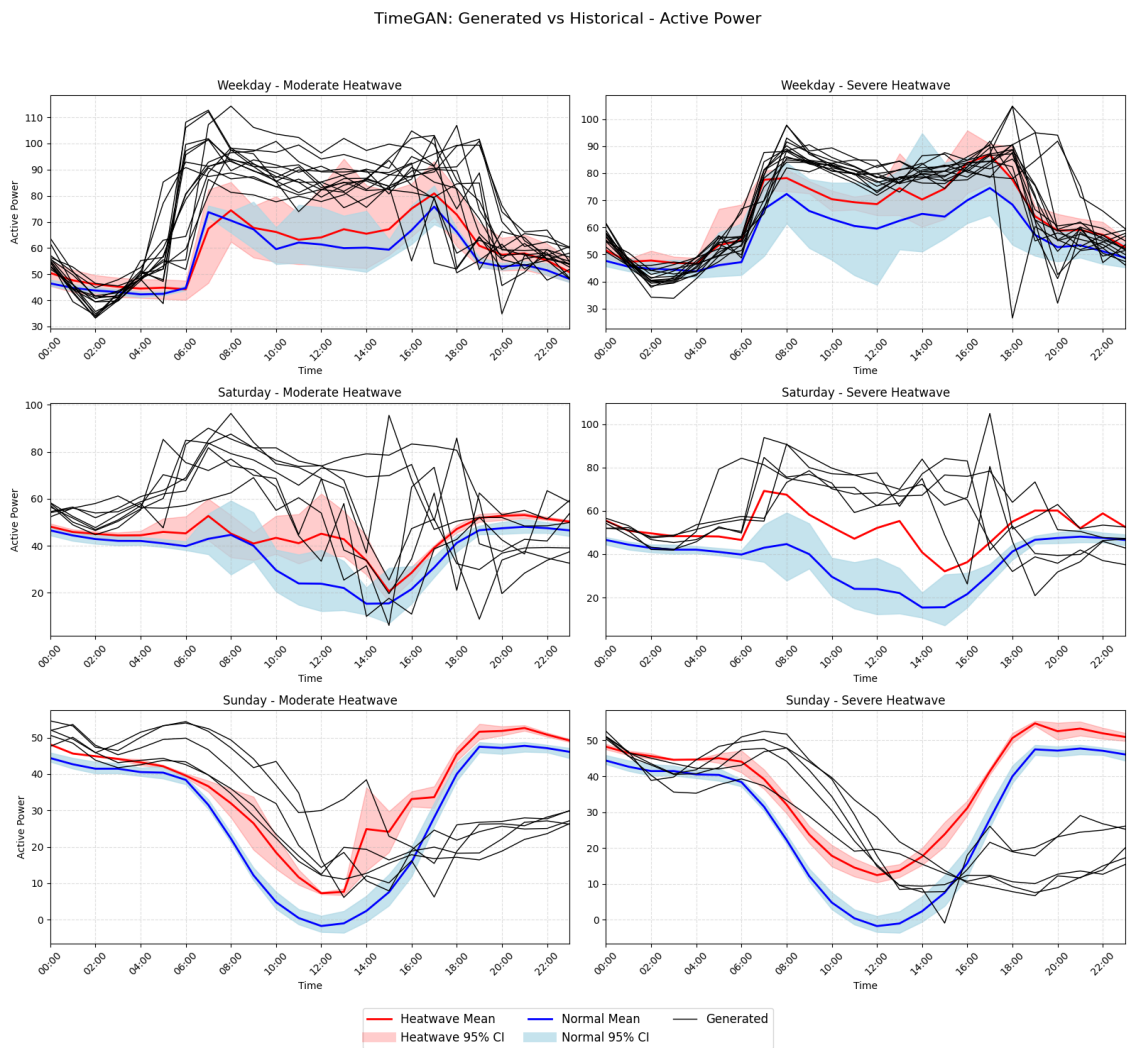
(50%) παραπάνω στα Σάββατα severe, αν και εδώ υπάρχει επίσης μόνο ένα Σάββατο, 1kW(20%) παραπάνω στις Κυριακές moderate και 2.5kW(40%) παραπάνω στις Κυριακές severe. Εν ολίγοις, στις περισσότερες περιπτώσεις ο μέσος όρος των τιμών των καυσώνων βρίσκεται πάνω από τον μέσο όρο των τιμών των φυσιολογικών (normal) ημερών, γεγονός το οποίο οφείλεται στον συνδυασμό μεγαλύτερης παραγωγής ισχύος από το φωτοβολταϊκό και περισσότερη κατανάλωση από τα κεντρικά γραφεία τις ημέρες του καύσωνα.

Ως προς το είδος της έντασης των καυσώνων των ιστορικών δεδομένων παρατηρούμε ότι στις καθημερινές ενεργού ισχύος η μέγιστη τιμή του μέσου severe καύσωνα είναι 5kW (5%) μεγαλύτερη από αυτή του moderate, στα Σάββατα ενεργού ισχύος είναι 20kW (29%) μεγαλύτερη, αν και υπάρχει μόνο μία χρονοσειρά severe καύσωνα, στις Κυριακές ενεργού ισχύος είναι 5kW (9%) μεγαλύτερη, στις καθημερινές άεργου ισχύος είναι 3kW (9%) μεγαλύτερη, στα Σάββατα άεργου ισχύος είναι 15kW (42%) μεγαλύτερη, αν και εδώ έχουμε μόνο μία χρονοσειρά severe καύσωνα, και τέλος στις Κυριακές άεργου ισχύος είναι 1.5kW (20%) μεγαλύτερη. Επίσης, παρατηρούμε ότι οι παραγόμενες χρονοσειρές δεν ακολουθούν πάντα τον προηγούμενο κανόνα, όπως π.χ. στις καθημερινές ενεργού ισχύος όπου η μέγιστη τιμή των moderate είναι 10kW παραπάνω στις 18:00 από αυτές των severe, ωστόσο στην ίδια περίπτωση διαπιστώνουμε ότι περισσότερες χρονοσειρές βρίσκονται πάνω από τον μέσο καύσωνα σε σύγκριση με αυτές των severe. Επομένως, το μοντέλο τείνει να παράγει όντως υψηλότερες χρονοσειρές severe καύσωνα.

Σχετικά με την ποιότητα των παραγόμενων δεδομένων, παρατηρείται ότι τα παραγόμενα δεδομένα ακολουθούν καλά τον μέσο καύσωνα και σε αρκετές περιπτώσεις (όπως καθημερινές moderate καύσωνα) οι τιμές τους είναι υψηλότερες σε φυσιολογικό βαθμό (περίπου 10%). Αυτό το γεγονός αποδεικνύει ότι το μοντέλο είναι ικανό να προσομοιάσει ρεαλιστικούς καύσωνες. Εξαιρέση αποτελούν τα δεδομένα της Κυριακής της άεργου ισχύος τα οποία παρουσιάζουν σταθερές, περίπου, χρονοσειρές χαμηλότερης ισχύος, το οποίο δείχνει την υψηλή πιθανότητα εκπαίδευσης του μοντέλου με κακής ποιότητας δεδομένα της άεργου ισχύος.

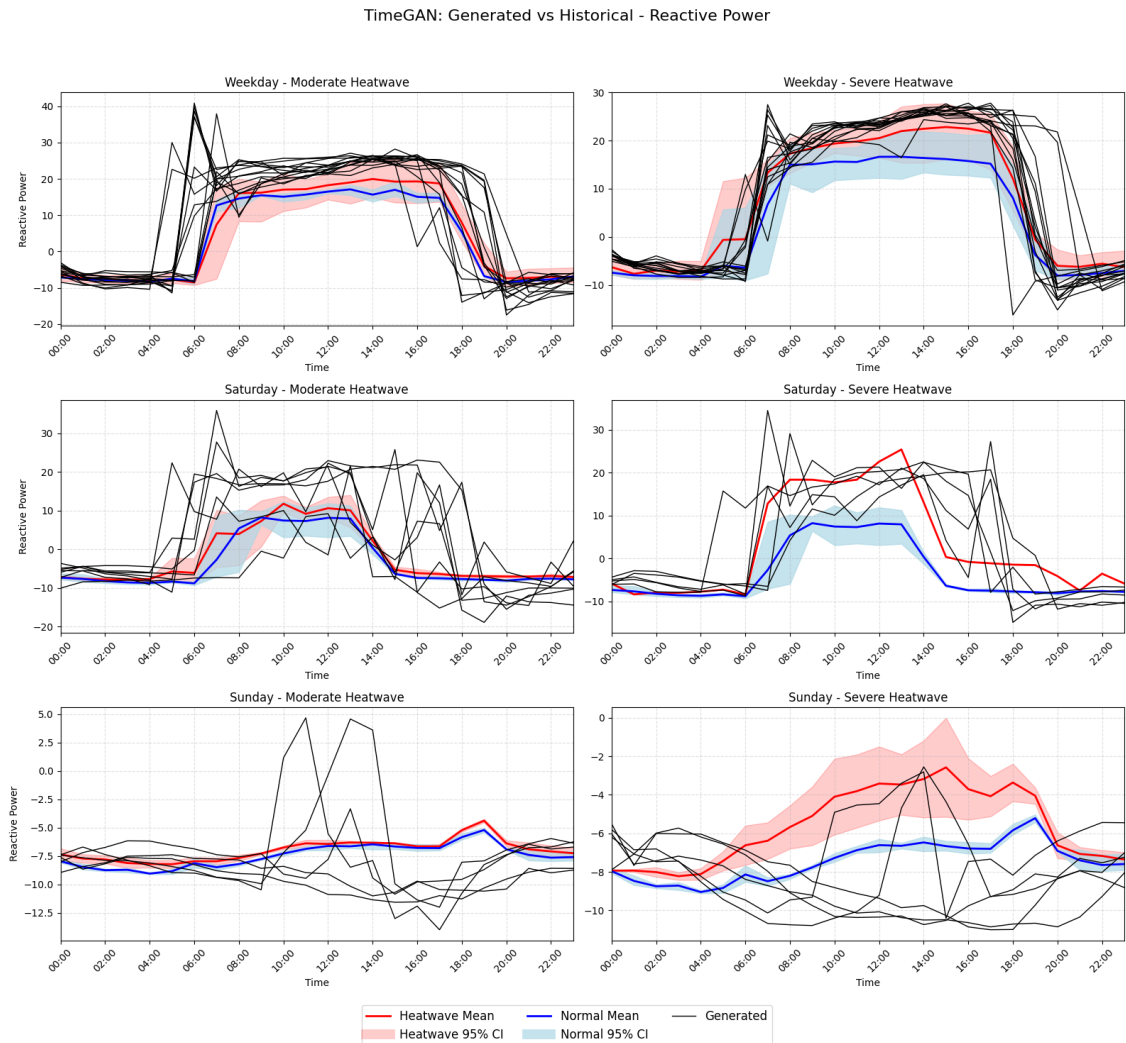
4.2 Αποτελέσματα του TimeGAN μοντέλου

Παρακάτω παρουσιάζονται οι χρονοσειρές που δημιουργήθηκαν μετά από την εξαγωγή αποτελεσμάτων του TimeGAN μοντέλου. Τα δεδομένα εμφανίζονται με τον ίδιο τρόπο που παρουσιάστηκαν στο γλωσσικό μοντέλο.



Σχήμα 4.3: Αποτελέσματα TimeGAN Inference Ενεργού Ισχύος

4.2. Αποτελέσματα του TimeGAN μοντέλου



Σχήμα 4.4: Αποτελέσματα TimeGAN Inference Άεργου Ισχύος

Από τα παραπάνω αποτελέσματα γίνεται αντιληπτό ότι και το TimeGAN έχει καταλάβει τις εποχικότητες της μέρας έως ένα βαθμό. Σε ορισμένα σημεία, όπως στις 18:00 στις καθημερινές severe καύσωνα, στο δεύτερο μισό των Σαββάτων moderate καύσωνα, και στο μεσαίο τμήμα των Κυριακών άεργου ισχύος, οι χρονοσειρές έχουν μεγάλη διακύμανση γεγονός που δείχνει ότι δεν έχει γίνει τόσο κατανοητή η χρονική συσχέτιση της μιας ώρας με την επόμενη. Ειδικά στην άεργο ισχύ οι χρονοσειρές μεγάλη διακύμανση και, όπως και στο γλωσσικό μοντέλο, φαίνεται οπτικά να έχουν χειρότερη απόδοση από αυτές της ενεργού ισχύος. Εξαιρέση έχουν οι καθημερινές οι οποίες φαίνεται να τα πηγαίνουν αρκετά καλά.

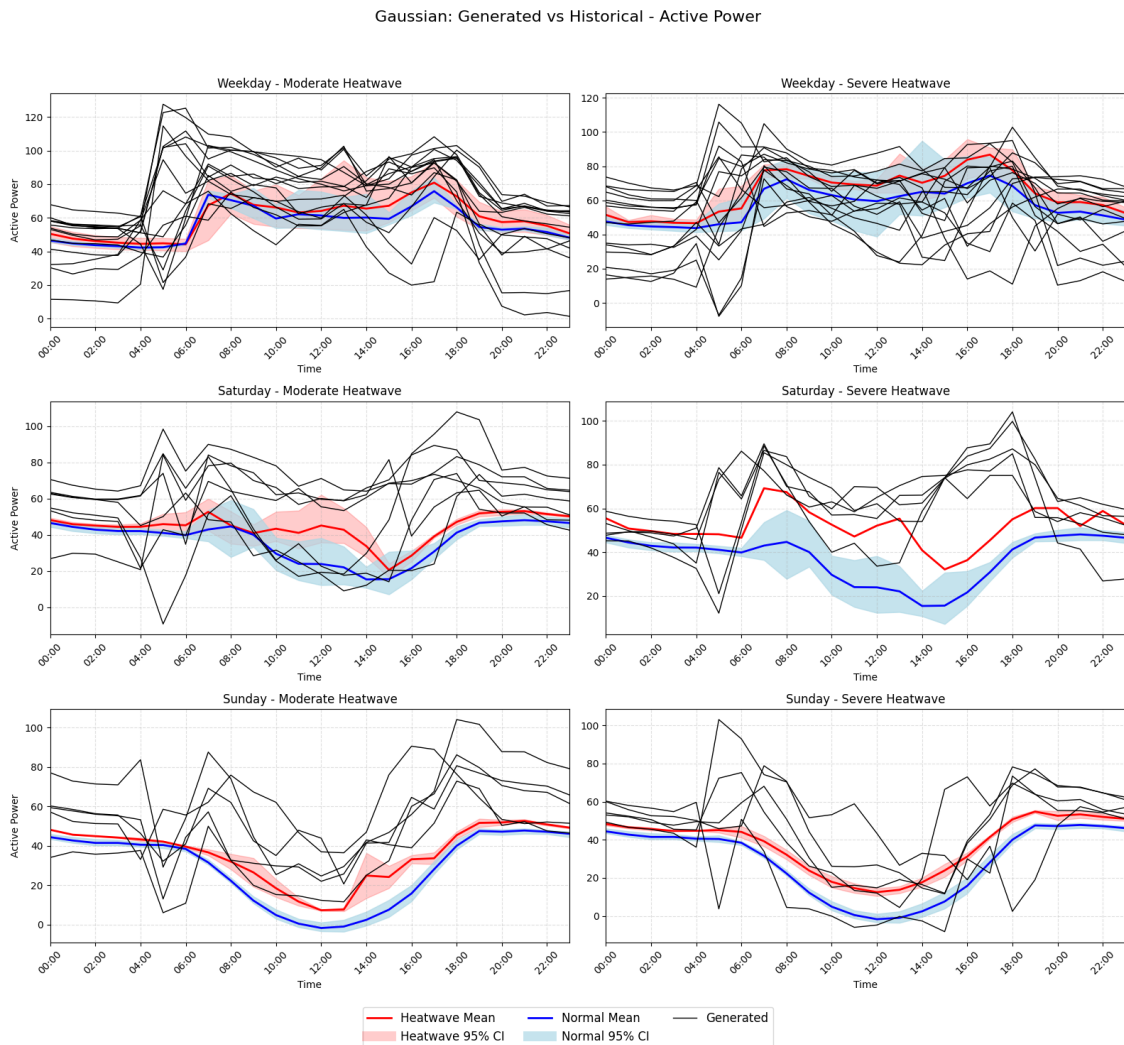
Επιπλέον, παρατηρείται ότι τα δεδομένα του Σαββάτου ακολουθούν κυρίως την συμπεριφορά των καθημερινών δεδομένων, δηλαδή το μοντέλο δύσκολα καταλαβαίνει την διαφορά στον τύπο των περιπτώσεων, ωστόσο είναι και αυτό αποτέλεσμα του περιορισμένου συνόλου δεδομένων. Ενώ το γλωσσικό μοντέλο τα πηγαίνει πολύ καλύτερα στην αναγνώριση των Σαββάτων.

Ως προς την ποιότητα των παραγόμενων δεδομένων, εκ πρώτης όψης, οι χρονοσειρές ακολουθούν αρκετά τον μέσο όρο των καυσώνων, ειδικά στις καθημερινές, ωστόσο, πολλές τιμές τους είναι υπερβολικά υψηλότερες (περίπου 30%) γεγονός που

αμφισβητεί την ρεαλιστικότητά τους. Τέλος, σε όλες τις περιπτώσεις εκτός από την κυριακή άεργου ισχύος, οι χρονοσειρές moderate και severe καύσωνα δεν παρουσιάζουν ουσιαστικές διαφορές, γεγονός που δείχνει ότι δεν μπορούν να καταλάβουν την διαφορά στους τύπους του καύσωνα.

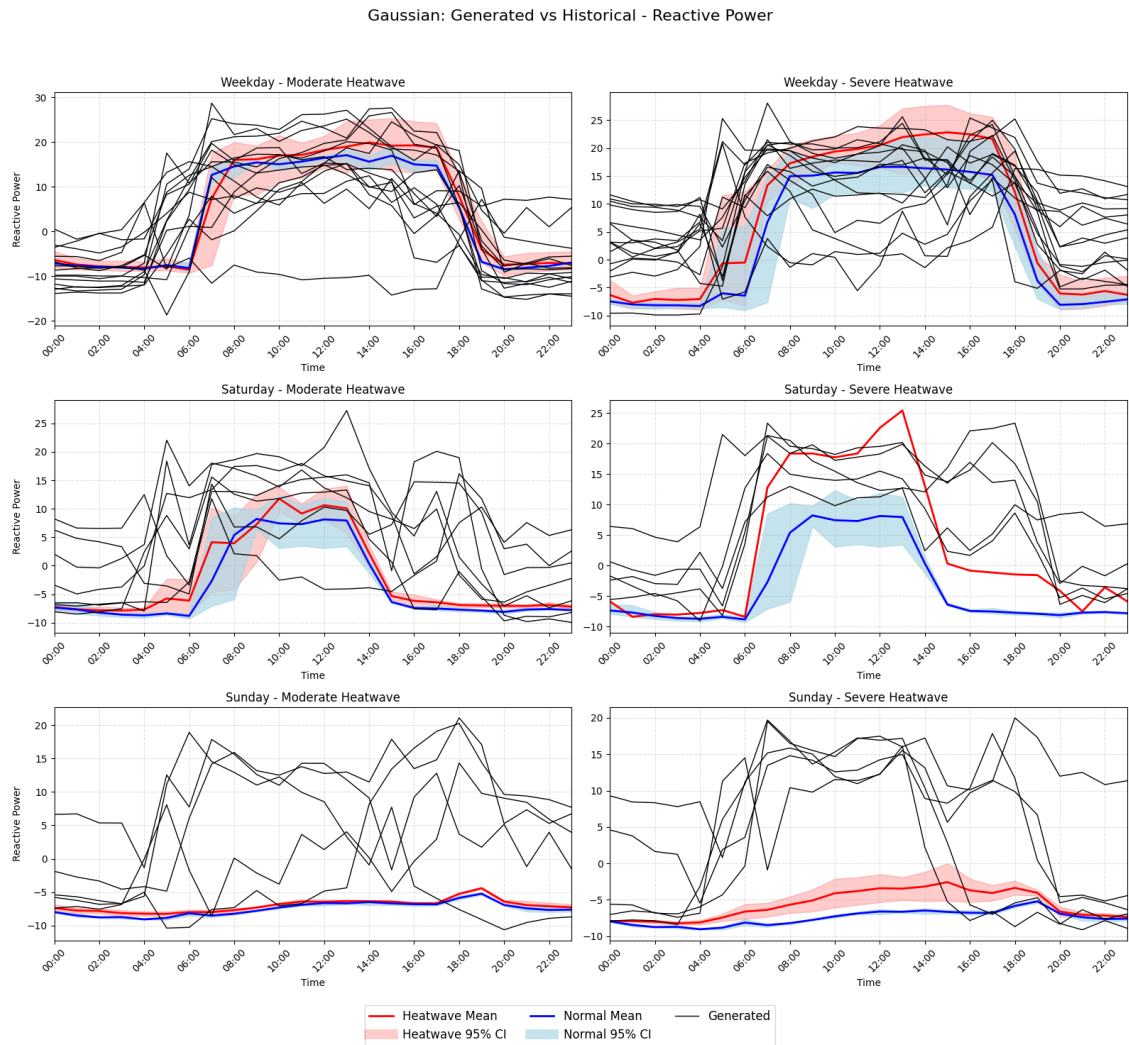
4.3 Αποτελέσματα του Gaussian μοντέλου

Παρακάτω παρουσιάζονται οι χρονοσειρές που δημιουργήθηκαν μετά από την εξαγωγή αποτελεσμάτων του Gaussian μοντέλου. Τα δεδομένα εμφανίζονται με τον ίδιο τρόπο που παρουσιάστηκαν και στα προηγούμενα δύο μοντέλα.



Σχήμα 4.5: Αποτελέσματα Gaussian μοντέλου Ενεργού Ισχύος

4.3. Αποτελέσματα του Gaussian μοντέλου



Σχήμα 4.6: Αποτελέσματα Gaussian μοντέλου Άεργου Ισχύος

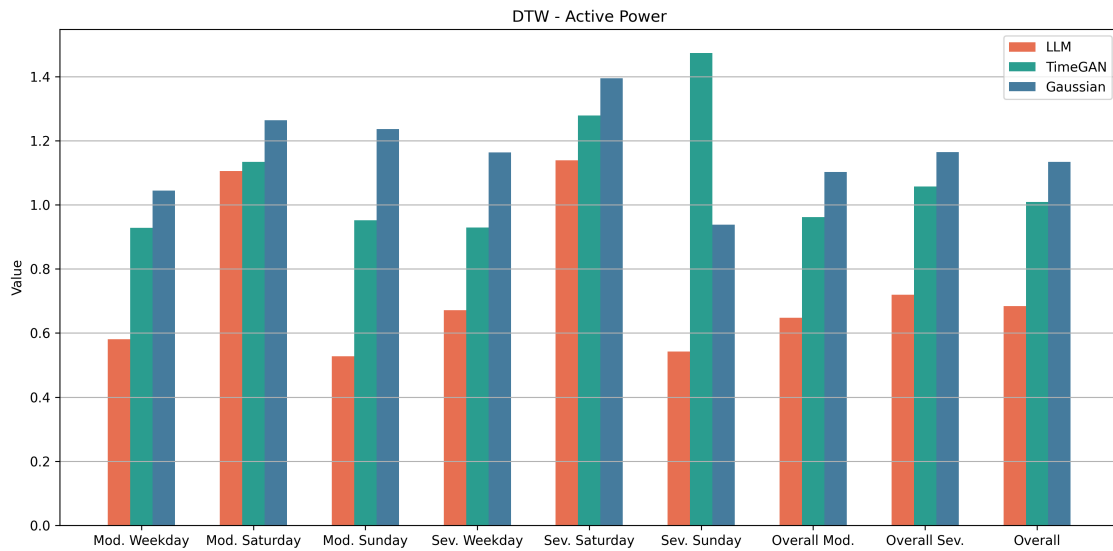
Από τα παραπάνω δεδομένα διαπιστώνεται ότι το Gaussian μοντέλο έχει καταλάβει ελάχιστα την περιοδικότητα της μέρας, και σχεδόν καθόλου τον τύπο ημέρας, εκτός από το μεσαίο τμήμα των Σαββατοκύριακων που πέφτει στα 20kW, όμως στα υπόλοιπα ακολουθεί τις καθημερινές. Επίσης οι χρονοσειρές της άεργου ισχύος έχουν αρκετές διακυμάνσεις και ειδικά στις Κυριακές δεν υπάρχει καθόλου ομοιότητα. Αυτό οφείλεται, πάλι, στον μικρό αριθμό δεδομένων του συνόλου. Ως προς την ποιότητα των δεδομένων, και αυτά ακολουθούν καλά την γραμμή του μέσου καύσωνα, ωστόσο, όπως και στο TimeGAN, έτσι και εδώ το ποσοστό ισχύος ξεπερνάει κατά πολύ αυτά των ιστορικών καυσώνων (περίπου 30%), γεγονός που και εδώ αμφισβητεί την πιστότητα και ρεαλιστικότητα των δεδομένων. Τέλος, ως προς το είδος καύσωνα, ούτε αυτό το μοντέλο έχει παρουσιάσει στοιχειώδεις διαφορές, καθώς οι χρονοσειρές moderate και severe καύσωνα κυμαίνονται στις ίδιες τιμές και μοιάζουν ίδιες.

Συγκρίνοντας οπτικά τα τρία μοντέλα μεταξύ τους παρατηρείται αρχικά η αδυναμία του Gaussian μοντέλου να παραγάγει δεδομένα διαφορετικών περιπτώσεων άεργου ισχύος. Το TimeGAN τα πηγαίνει καλά στις καθημερινές αν και στις υπολοιπες μέρες έχει αρκετές έντονες διακυμάνσεις. Το LLM μοντέλο έχει καταλάβει καλύτερα από όλα τα μοντέλα τις διαφορές στον τύπο ημέρας. Ωστόσο και τα τρία μοντέλα

εμφανίζουν χειρότερα αποτελέσματα στα δεδομένα άεργου ισχύος, με το LLM να παράγει συνεχής γραμμές, το TimeGAN να έχει έντονες διακυμάνσεις και το Gaussian να μπερδεύει και τον τύπο ημέρας και τον τύπο ισχύος. Συμπερασματικά, το LLM μοντέλο δίνει πιο ρεαλιστικά και πολλών περιπτώσεων σενάρια, το TimeGAN δίνει καλά (αλλά λίγο υπερβολικά) αποτελέσματα μόνο στις καθημερινές, ενώ το Gaussian δίνει και αυτό καλά αποτελέσματα μόνο στις καθημερινές ενεργού ισχύος. Ωστόσο, όλα τα μοντέλα έχουν χειρότερα αποτελέσματα, σχεδόν ασυσχέτιστα, στις περιπτώσεις άεργου ισχύος, που δηλώνει την ύπαρξη κάποιας μορφής προβλήματος με την ποιότητα των δεδομένων.

4.4 Αποτελέσματα της αξιολόγησης των μοντέλων

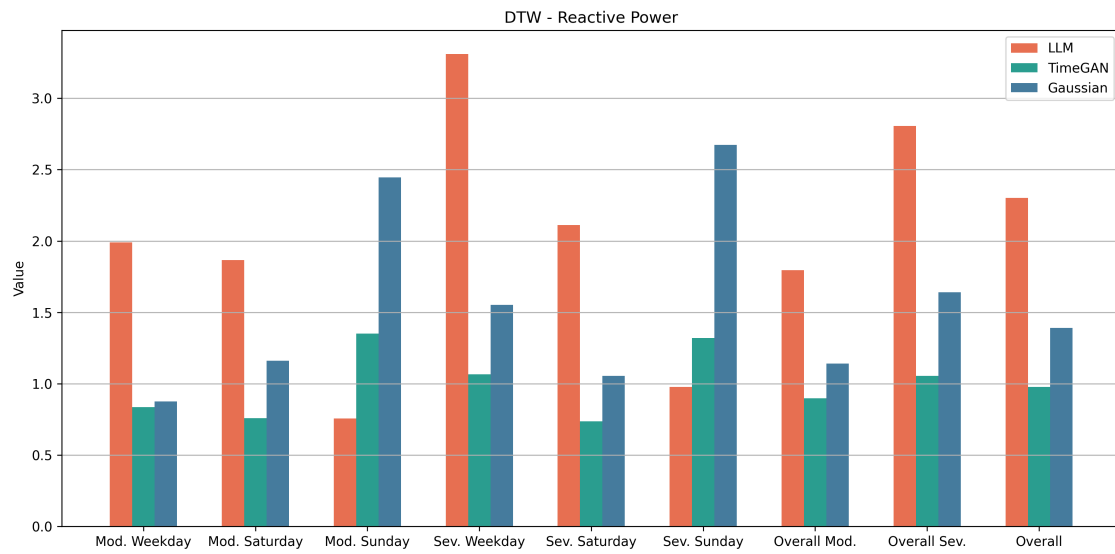
Παρακάτω παρουσιάζονται οι τιμές των μετρικών όλων των μοντέλων, καθώς και γίνεται σύγκριση μεταξύ τους.



Σχήμα 4.7: Τιμές DTW μετρικής Ενεργού Ισχύος

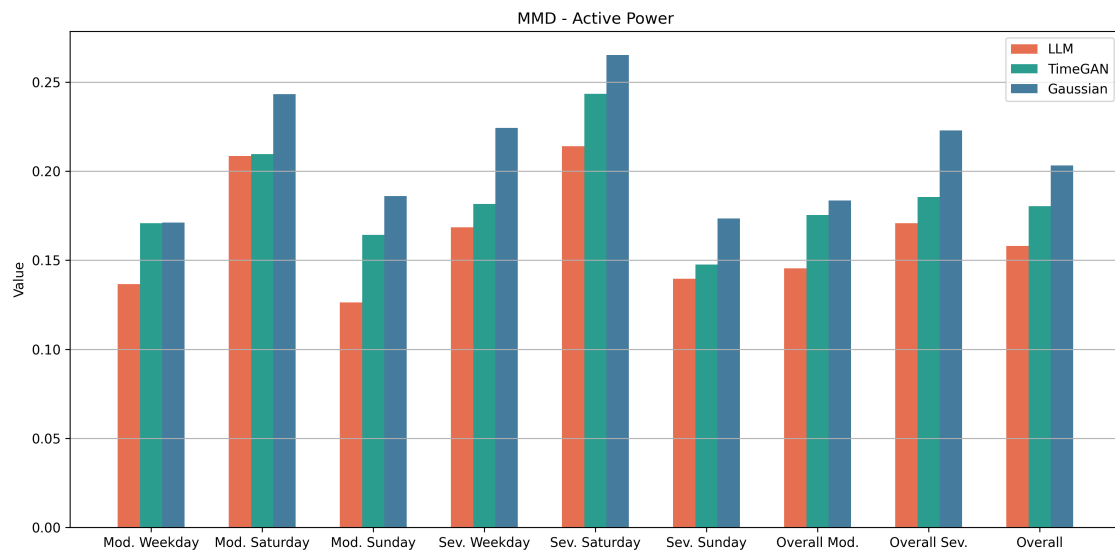
Στην μετρική DTW στα δεδομένα Ενεργού Ισχύος το LLM εμφάνισε αρκετά καλύτερα αποτελέσματα από τα άλλα δύο μοντέλα.

4.4. Αποτελέσματα της αξιολόγησης των μοντέλων



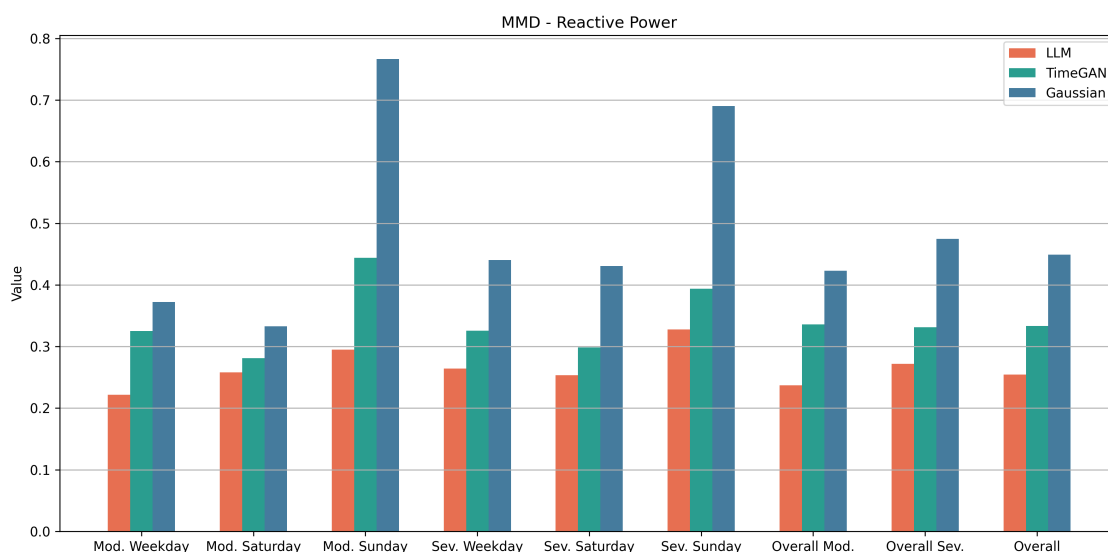
Σχήμα 4.8: Τιμές DTW μετρικής Άεργου Ισχύος

Στην μετρική DTW στα δεδομένα Άεργου Ισχύος το LLM εμφάνισε χειρότερα αποτελέσματα από τα άλλα δύο μοντέλα. Αυτό οφείλεται στις σταθερές τιμές που εμφανίζει στις τιμές άεργου ισχύος, όπως είδαμε στην οπτική απεικόνιση της αξιολόγησης, καθώς με την σταθερή τιμή η DTW "τεντώνεται" αρκετά ώστε να ευθυγραμμιστεί με την επόμενη γραμμή. Τα καλύτερα αποτελέσματα τα παρουσιάζει το TimeGAN μοντέλο.



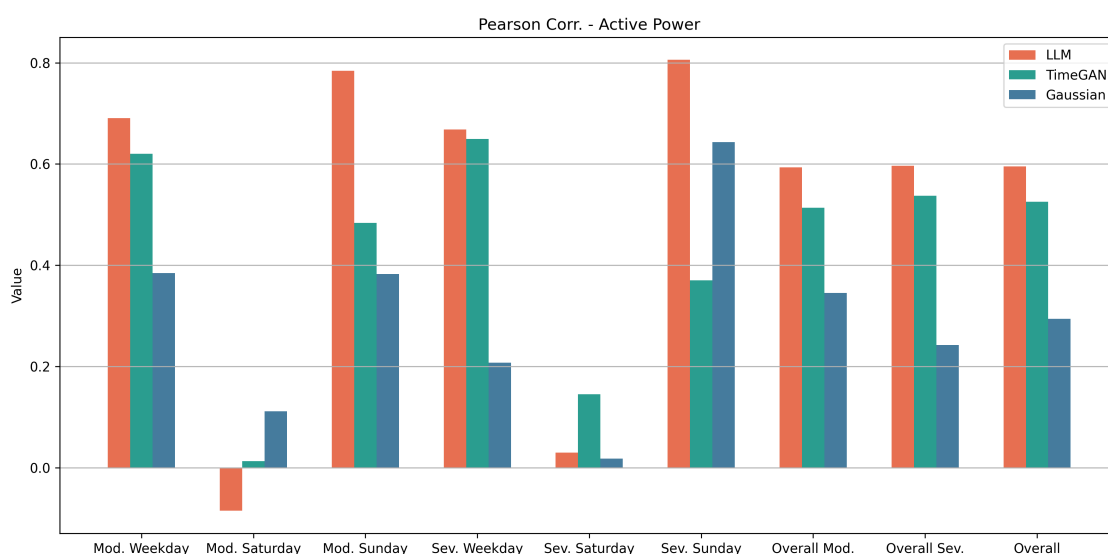
Σχήμα 4.9: Τιμές MMD μετρικής Ενεργού Ισχύος

Στην μετρική MMD στα δεδομένα Ενεργού Ισχύος το LLM εμφάνισε αρκετά καλύτερα αποτελέσματα από τα άλλα δύο μοντέλα.



Σχήμα 4.10: Τιμές MMD μετρικής Άεργου Ισχύος

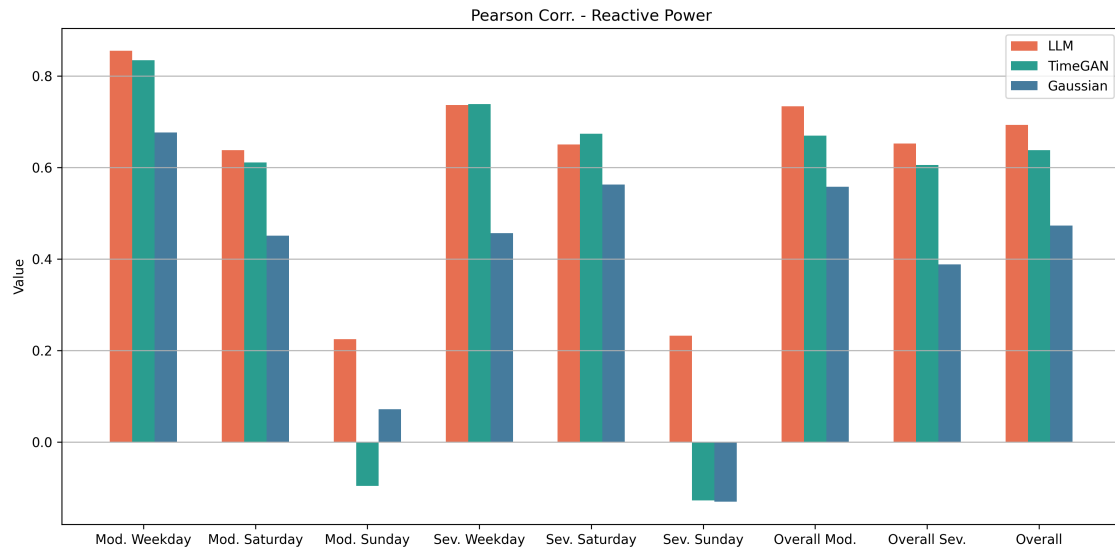
Στην μετρική MMD στα δεδομένα Άεργου Ισχύος το LLM εμφάνισε αρκετά καλύτερα αποτελέσματα από τα άλλα δύο μοντέλα.



Σχήμα 4.11: Τιμές PC μετρικής Ενεργού Ισχύος

Στην μετρική PC στα δεδομένα Ενεργού Ισχύος το LLM εμφάνισε αρκετά καλύτερα αποτελέσματα από τα άλλα δύο μοντέλα, ωστόσο, στις περιπτώσεις Σαββάτου δεν έδειξε καθόλου συσχέτιση των δεδομένων. Όμως ούτε στα άλλα δύο μοντέλα υπήρξε κάποια συσχέτιση γεγονός που μπορεί να οφείλεται στο ότι τα δεδομένα του Σαββάτου ήταν ελάχιστα, και στο ότι τα δεδομένα αυτά έχουν κάποιες ομοιότητες με τα δεδομένα των καθημερινών και το μοντέλο δυσκολεύεται να τα ξεχωρίσει, ενώ στις Κυριαχές, που είναι εντελώς διαφορετικά, δεν δυσκολεύεται.

4.4. Αποτελέσματα της αξιολόγησης των μοντέλων



Σχήμα 4.12: Τιμές PC μετρικής Άεργου Ισχύος

Στην μετρική PC στα δεδομένα Άεργου Ισχύος το LLM εμφάνισε αρκετά καλύτερα αποτελέσματα από τα άλλα δύο μοντέλα.

Παρακάτω εμφανίζονται τα συνολικά αποτελέσματα σε μορφή πίνακα:

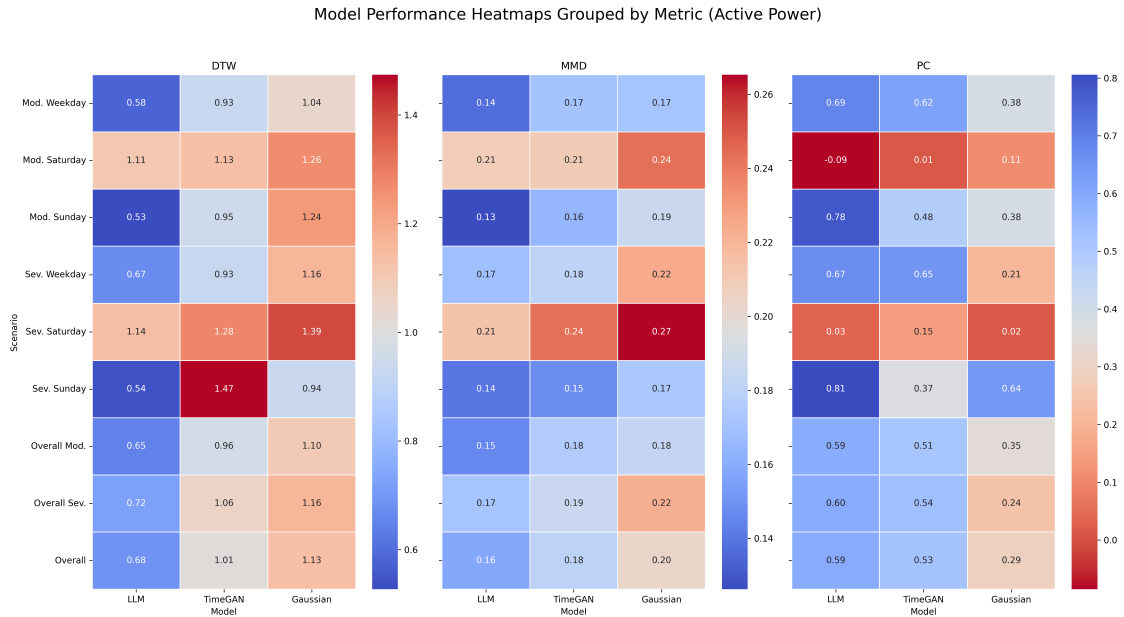
Μετρική	T5	TimeGAN	Gaussian
DTW	0.68	1.01	1.13
MMD	0.16	0.18	0.2
PC	0.59	0.53	0.29

Πίνακας 4.1: Αποτελέσματα συνολικών μετρικών Ενεργού Ισχύος.

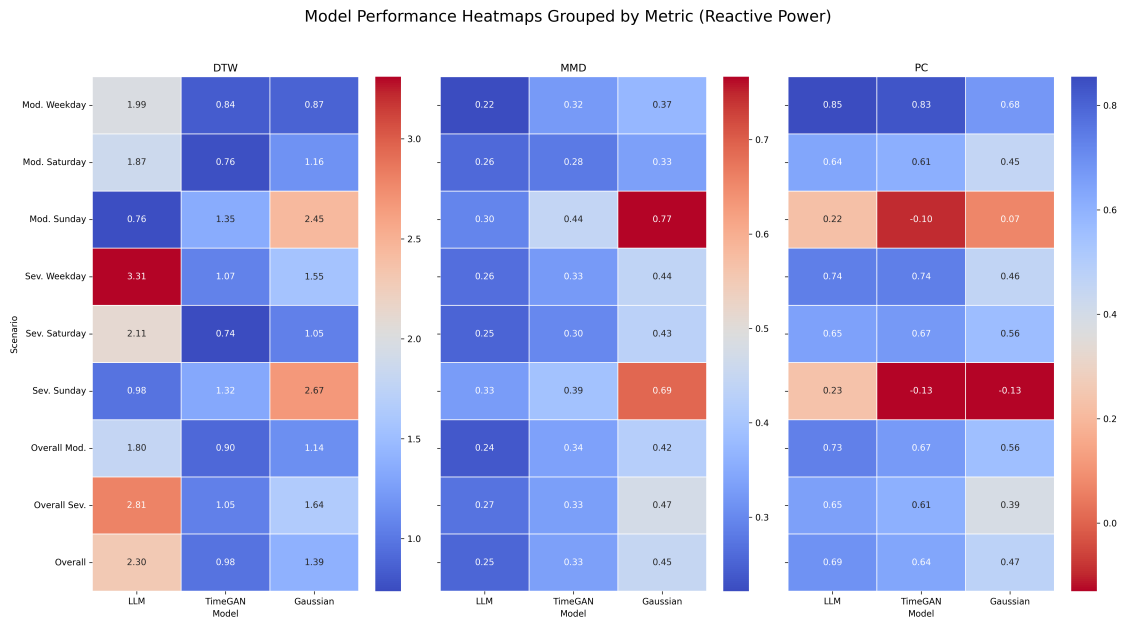
Μετρική	T5	TimeGAN	Gaussian
DTW	2.3	0.98	1.39
MMD	0.25	0.33	0.45
PC	0.69	0.64	0.47

Πίνακας 4.2: Αποτελέσματα συνολικών μετρικών Άεργου Ισχύος.

Παρακάτω παρουσιάζονται συγκεντρωτικά όλα τα αποτελέσματα των μετρικών ανά μοντέλο και περίπτωση με την μορφή χάρτη θερμότητας (Heatmap).



Σχήμα 4.13: Όλες οι μετρικές Ενεργού Ισχύος σε μορφή Heatmap



Σχήμα 4.14: Όλες οι μετρικές Άεργου Ισχύος σε μορφή Heatmap

Συμπερασματικά, σύμφωνα με την μέθοδο αξιολόγησης το LLM παρουσιάζει καλύτερα αποτελέσματα σε πέντε από τις έξι μετρήσεις. Η περίπτωση που υστερεί είναι στα δεδομένα Άεργου Ισχύος με την μετρική DTW, στο οποίο πιστεύουμε ότι οφείλεται η παραγωγή σταθερών τιμών (και απουσία διακύμανσης) ίσως λόγω της επιλογής της παραμέτρου temperature (0.6).

Τα αμέσως καλύτερα αποτελέσματα τα λαμβάνει το TimeGAN μοντέλο λόγω των αρκετά ρεαλιστικών δεδομένων σύμφωνα με την ποιοτική απεικόνιση των δεδομένων, ενώ λίγο πιο πίσω βρίσκεται το Gaussian μοντέλο.

Μπορεί τα αποτελέσματα να μην ήταν τα βέλτιστα σε όλες τις μετρικές, όμως το

4.4. Αποτελέσματα της αξιολόγησης των μοντέλων

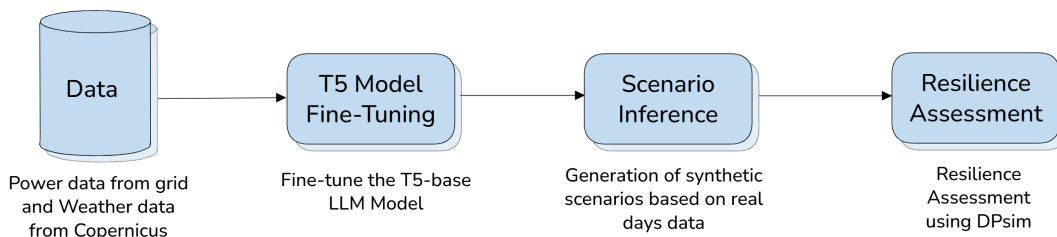
LLM παρέχει τα πιο παρόμοια και ρεαλιστικά αποτελέσματα από όλα τα υπόλοιπα.

Κεφάλαιο 5

Συζήτηση

Στο παρακάτω κεφάλαιο παρουσιάζεται μία πρακτική εφαρμογή της μεθόδου που αναλύθηκε, ως μια ερευνητική επέκταση της εργασίας.

Στα προηγούμενα κεφάλαια έγινε η ανάπτυξη ενός γλωσσικού μοντέλου το οποίο όχι μόνο παράγει συνθετικά δεδομένα, αλλά και λαμβάνει υπόψη του τις καιρικές συνθήκες, δημιουργώντας έτσι πολλά διαφορετικά σενάρια με την ίδια είσοδο δεδομένων ενέργειας και καιρικών δεδομένων. Ως επέκταση της παρούσας έρευνας, αποφασίσαμε να επεξεργαστούμε περαιτέρω τα παραγόμενα σενάρια και να εξετάσουμε αν με τις παραγόμενες χρονοσειρές μπορεί να προκληθεί κάποιο πρόβλημα στο δίκτυο. Παρακάτω ακολουθεί ένα high level διάγραμμα της πρακτικής εφαρμογής ως συνέχεια της ανάπτυξης του μοντέλου:



Σχήμα 5.1: High Level Διάγραμμα της συνολικής διαδικασίας

5.1 Μεθοδολογία

Το μοντέλο που αναπτύχθηκε αφορά κυρίως δεδομένα δικτύου ενέργειας, και ως εκ τούτου εξετάστηκε το ενδεχόμενο χρήσης του στην διαδικασία προβλεπτικής συντήρησής και αξιολόγηση της ανθεκτικότητάς του. Ένα πρώτο βήμα προς αυτή την κατεύθυνση είναι να βρούμε την τάση του συστήματος δεδομένης της τρέχουσας ισχύς, καθώς οι τιμές τάσης παρέχουν πιο σαφή εικόνα για την κατάσταση και τη σταθερότητα του συστήματος. Στην παρούσα εργασία επιλέχθηκε το DPsim, ένα ανοιχτού κώδικα εργαλείο προσομοίωσης ηλεκτρικών συστημάτων. Το DPsim βασίζεται στη μέθοδο των δυναμικών φασμάτων (Dynamic Phasors) και την Τροποποιημένη Ανάλυση Κόμβων (Modified Nodal Analysis – MNA), γεγονός που το καθιστά κατάλληλο για προσομοιώσεις σε πραγματικό χρόνο και για μεγάλης κλίμακας σενάρια [96].

Σε συνεργασία με τους δημιουργούς του DPsim, το πανεπιστήμιο RWTH της Γερμανίας, έγινε μοντελοποίηση του δικτύου της ASM και με τροφοδοσία συγκεκριμένων σεναρίων υπολογίστηκαν οι τιμές τάσης. Τα σενάρια που τροφοδοτήθηκαν στο DPsim ήταν τα 50 παραγόμενα σενάρια της εξαγωγής αποτελεσμάτων του LLM, που παρουσιάστηκαν στην Ενότητα 4.1, καθώς και οι υπόλοιπες φυσιολογικές (normal) μέρες του συνόλου δοκιμής, οι οποίες ήταν 27 στο σύνολό τους, ως baselines. Στο DPsim πρέπει να μπει τιμή για όλους τους μετρητές, όχι μόνο για τον w4, επομένως παρακάτω αναλύεται ο τρόπος υπολογισμού των τιμών των μετρητών που μπήκαν στην

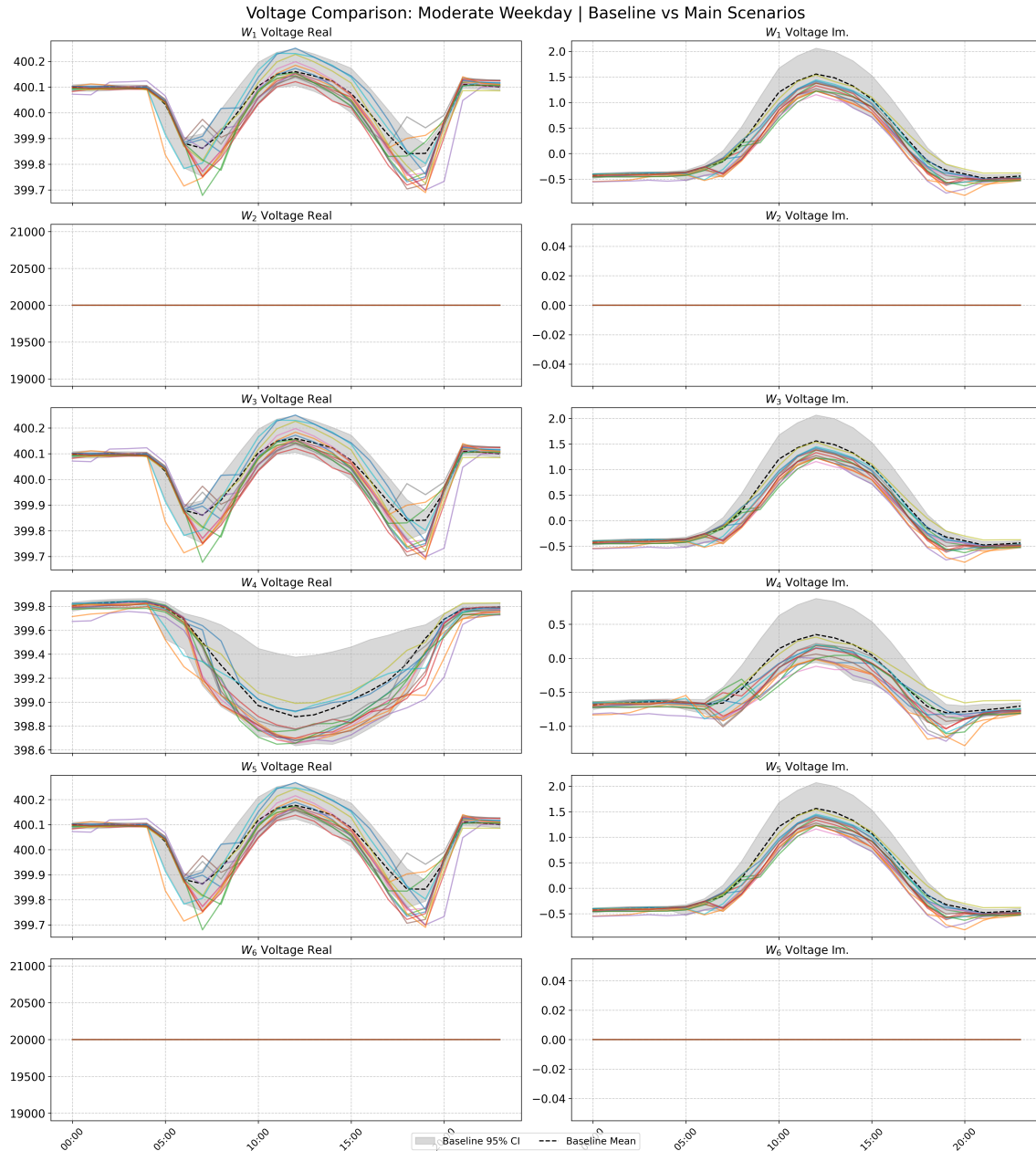
προσομοίωση.

Στα baselines και στους μετρητές w3, w5 και w6 έγινε υπολογισμός των τιμών των μηνών Ιουνίου, Ιουλίου και Αυγούστου και υπολογίστηκε ο μέσος όρος των normal ημερών. Στους μετρητές w1 και w2 δεν προστέθηκαν τιμές, επειδή οι τιμές τους εξαρτώνται από τους άλλους μετρητές, όπως φαίνεται στην τοπολογία 3.2. Στον μετρητή w4 μπήκαν όλες οι διαθέσιμες τιμές των normal ημερών που υπήρχαν στο σύνολο δοκιμής και που ανήκαν σε ημέρες καλοκαιριού (δηλαδή των μηνών Ιούνιο, Ιούλιο και Αύγουστο). Αυτές οι τιμές ήταν 27 στο σύνολό τους και στον μετρητή w4 μπήκαν αυτούσια αυτά τα 27 διαφορετικά σενάρια.

Στα υπόλοιπα 50 βασικά σενάρια, όπου υπενθυμίζεται ότι κάθε σενάριο ανήκει σε περίπτωση mild heatwave ή severe heatwave, στους μετρητές w3, w5 και w6 έγινε υπολογισμός των τιμών των μηνών Ιουνίου, Ιουλίου και Αυγούστου και υπολογίστηκαν δύο τιμές, ο μέσος όρος των ημερών που ανήκουν στις μέρες "mild heatwave" και ο μέσος όρος των ημερών που ανήκουν στις μέρες "severe heatwave". Στους μετρητές w1 και w2 πάλι δεν προστέθηκαν τιμές, όπως προηγουμένως, και στον μετρητή w4 μπήκαν τα 50 σενάρια του συνόλου παραγωγής δεδομένων που αναλύθηκε στο 3.2.1.2.2. Σημειώνεται ότι κάθε σενάριο από τους μετρητές w3, w5 και w6 μπήκε στο αντίστοιχο σενάριο ίδιας έντασης καύσωνα με το σενάριο του w4 που προουπήρχε ήδη στο σύνολο.

5.2 Αποτελέσματα

Η προσομοίωση παρουσίασε τα παρακάτω αποτελέσματα για την περίπτωση έντασης καιρού moderate heatwave και τύπου ημέρας weekday:



Σχήμα 5.2: Αποτελέσματα προσομοίωσης περίπτωσης Moderate Weekday

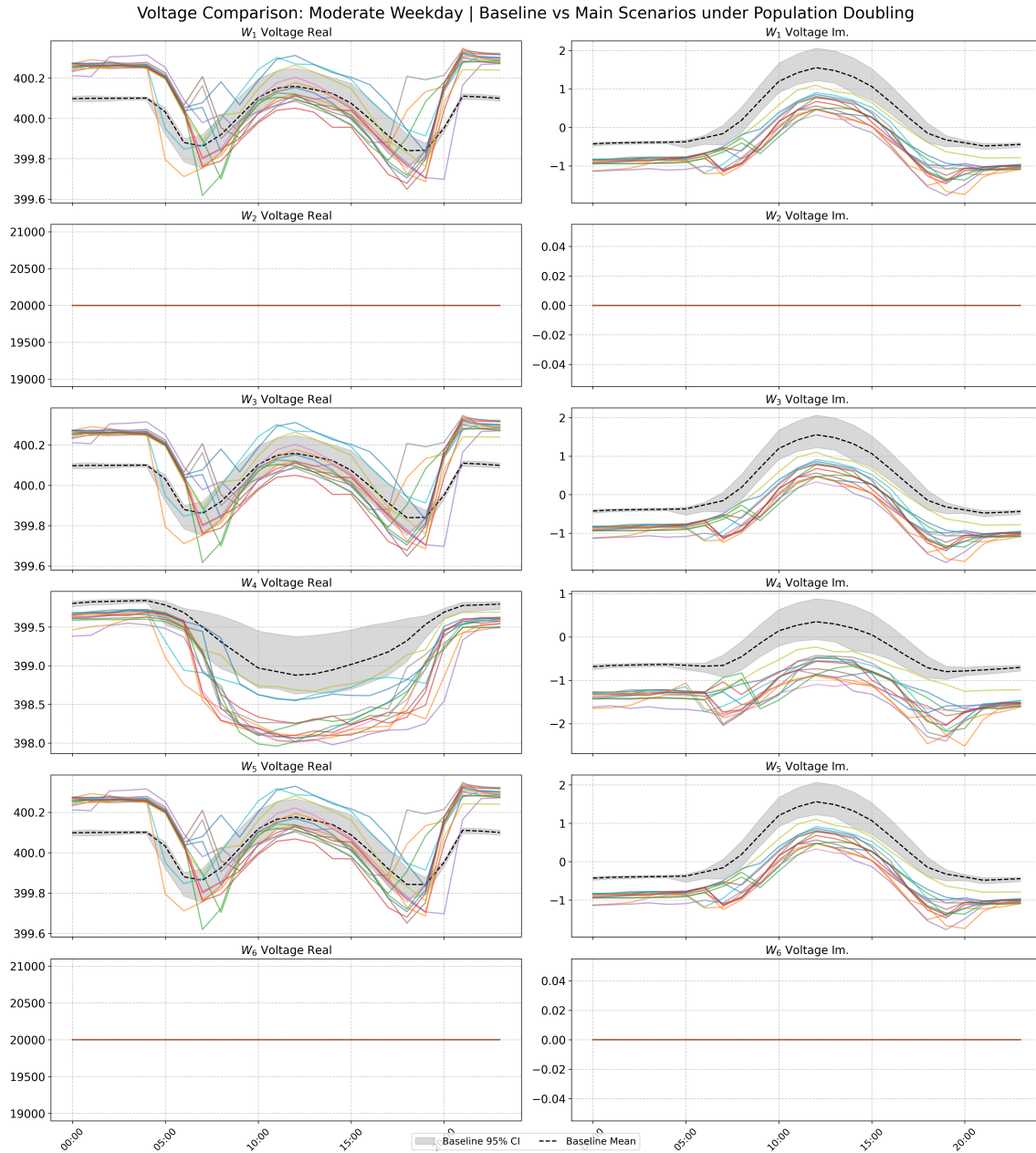
Στο παραπάνω διάγραμμα βλέπουμε την πραγματική και φανταστική τιμή της τάσης όλων των κόμβων της περίπτωσης moderate weekday. Οι χρωματισμένες γραμμές δείχνουν τις τιμές των βασικών σεναρίων, ενώ τα baselines είναι υπό την μορφή διαστήματος εμπιστοσύνης (confidence interval), χρώματος γκρι και ο μέσος όρος τους ως μία γκρι διακεκομμένη γραμμή. Τα διαγράμματα των υπόλοιπων περιπτώσεων βρίσκονται στο παράρτημα Α.

Από την προσομοίωση παρατηρούμε ότι οι παραγόμενες τιμές ισχύος σε συνθήκες moderate καύσινα οδηγούν την τάση σε χαμηλότερες τιμές (398.5V) σε αντίθεση με των μέσο όρο των κανονικών ημερών (398.9V), ωστόσο αυτή η διαφορά δεν είναι σημαντική και δεν προκαλεί πρόβλημα στο δίκτυο.

Γενικά, η προσομοίωση αυτή έδειξε ότι δεν υπάρχει ανησυχία δημιουργίας προβλήματος στο δίκτυο, αφού καμία τιμή της τάσης των μετρητών που βρίσκονται στη

χαμηλή τάση δεν ξεπερνάει τα επιτρεπτά όρια του δικτύου ενέργειας, που συνήθως είναι $\pm 10\%$ της ονομαστικής τάσης (εδώ 400V) [97]. Οι τάσεις του W_2 και W_6 βλέπουμε ότι παραμένουν σταθερές και είναι 20kV διότι βρίσκονται στη Μέση Τάση, ενώ οι υπόλοιπες βρίσκονται στα επιτρεπτά όρια και μέσα στο διάστημα εμπιστοσύνης. Ωστόσο παρατηρούμε σε όλες τις περιπτώσεις ότι στον w4 μετρητή οι τιμές τάσεις των καυσώνων βρίσκονται κάτω από τον μέσο όρων των φυσιολογικών ημερών, αλλά δεν ξεπερνούν το 0,1% της μέσης τιμής της τάσης.

Συνεχίζοντας, δοκιμάσαμε μοντέλα αύξησης πληθυσμού, όπου με την ίδια τοπολογία και το μοντέλο, αυξάνουμε την ισχύ μόνο του μετρητή w4 μέχρι και 2 φορές περισσότερο (διπλάσιος πληθυσμός). Παρακάτω παρουσιάζεται το αποτέλεσμα της προσομοίωσης για την περίπτωση έντασης καιρού moderate heatwave και τύπου ημέρας weekday. Τα διαγράμματα των υπόλοιπων περιπτώσεων βρίσκονται στο Παράρτημα Β.



Σχήμα 5.3: Αποτελέσματα προσομοίωσης Moderate Weekday σε περίπτωση διπλασιασμού πληθυσμού

Στο παραπάνω διάγραμμα βλέπουμε ότι οι περισσότερες τιμές έχουν μετατοπιστεί κατά μία μικρή σταθερά είτε προς τα πάνω (μετρητές w_3 , w_5 , w_6) είτε προς τα κάτω (μετρητής w_4). Συγκεκριμένα ο W_4 έχει μετατοπιστεί κατά 1V (0.25%) προς τα κάτω και πλέον οι τιμές του έχουν απομακρυνθεί από το διάστημα εμπιστοσύνης των baselines, όμως ούτε αυτή η δοκιμή προκαλεί κάποια ανησυχία για βλάβη του συστήματος καθώς η τάση παραμένει στα επιτρεπτά όρια.

Κεφάλαιο 6

Συμπεράσματα και Μελλοντικές Προεκτάσεις

6.1 Συμπεράσματα και Επίλογος

Στην παρούσα εργασία αναπτύχθηκε ένα μεγάλο γλωσσικό μοντέλο το οποίο παράγει συνθετικά δεδομένα χρονοσειρών ενέργειας σε συνθήκες καύσωνα. Τα δεδομένα που χρησιμοποιήθηκαν προμηθεύτηκαν από την εταιρία ASM που βρίσκεται στο Τέρνι, Ιταλίας. Το γλωσσικό μοντέλο που χρησιμοποιήθηκε ήταν το Google-T5-base το οποίο έχει 220 εκατομμύρια παραμέτρους και για το finetuning έγινε χρήση του QLoRA, μία μέθοδος που βελτιώνει και καθιστά πιο γρήγορη την μετ-εκπαίδευση, και την επιτρέπει και σε μηχανήματα με περιορισμένους πόρους GPU.

Το μοντέλο εκπαιδεύτηκε έχοντας ως είσοδο τα δεδομένα ενέργειας, την θερμοκρασία, το είδος των δεδομένων και την επιθυμητή ένταση του καύσωνα, με σκοπό να παράγει ρεαλιστικά υποθετικά σενάρια. Η χρήση της έντασης του καύσωνα λειτουργεί ως μελλοντική συμμεταβλητή (future covariate) κάνοντας το μοντέλο να τείνει να παράγει πιο ρεαλιστικά δεδομένα.

Το μοντέλο συγκρίθηκε με άλλα δύο μοντέλα παραγωγής συνθετικών δεδομένων. Το state-of-the-art μοντέλο TimeGAN, το οποίο πρόκειται για ένα GAN με ένα επιπλέον βήμα όπου για την εκπαίδευσή του τα δεδομένα μεταφέρονται σε έναν λανθάνοντα χώρο (latent space) όπου εκεί γίνεται η εκμάθηση των χρονικών εξαρτήσεων των χρονοσειρών. Το δεύτερο μοντέλο είναι το στατιστικό μοντέλο του Synthetic Data Vault (SDV), το οποίο χρησιμοποιεί την στατιστική κατανομή Gaussian Copula για να υπολογίσει τις τιμές και λειτουργεί ως baseline.

Και τα τρία μοντέλα εξετάστηκαν σε ένα υποσύνολο των συνολικών δεδομένων, το οποίο δεν είχε συμμετάσχει στην εκπαίδευση κανενός μοντέλου. Σύμφωνα με αυτό το σύνολο, έγινε η παραγωγή των χρονοσειρών και από τα τρία μοντέλα και τα αποτελέσματά τους συγκρίθηκαν με τις τρεις μετρικές Dynamic Time Warping (DTW), Maximum Mean Discrepancy (MMD) και Pearson Correlation (PC). Το γλωσσικό μοντέλο προσέφερε αρκετά καλύτερα αποτελέσματα σε 5 από τις 6 μετρήσεις, όπου υστερεί στα δεδομένα άεργου ισχύος στην μετρική DTW, όπου θεωρούμε ότι οφείλεται στο γεγονός ότι οι χρονοσειρές της άεργου ισχύος παρουσιάζουν ευθείες και η DTW εστιάζει στην ομοιότητα των σχημάτων των χρονοσειρών. Ωστόσο, η εμφανώς χειρότερη απόδοση όλων των μοντέλων στις αντίστοιχες περιπτώσεις της άεργου ισχύος, σε σύγκριση με αυτές της ενεργού, υποδηλώνουν ότι η υπάρχει η περίπτωση τα δεδομένα άεργου ισχύος να είναι κακής ποιότητας. Εν ολίγοις, το γλωσσικό μοντέλο είχε καλύτερα αποτελέσματα και πιο ρεαλιστικά από τα δύο υπόλοιπα μοντέλα, γεγονός που επιβεβαιώνει την κυριαρχία των μεγάλων γλωσσικών μοντέλων στα πολλά διαφορετικά tasks κειμένου.

Το TimeGAN μοντέλο ενώ φαίνεται να κατανοεί στις περιπτώσεις τύπου ημέρας, υστερεί στην κατανόηση των περιπτώσεων έντασης και επίσης παρουσιάζει αρκετές διακυμάνσεις, ειδικά στα δεδομένα άεργου ισχύος. Το Gaussian μοντέλο δεν κατάφερε να καταλάβει ούτε τις διαφορετικές περιπτώσεις τύπου ημέρας ούτε τις περιπτώσεις έντασης και επίσης εμφάνισε έντονες διακυμάνσεις στις χρονοσειρές.

Τέλος, ως επέκταση της εργασίας, δοκιμάστηκε η χρήση του μοντέλου ως μέρος της διαδικασίας προβλεπτικής συντήρησης του δικτύου ενέργειας. Χρησιμοποιώντας το

εργαλείο Dpsim, το οποίο προσομοιώνει την τοπολογία του δικτύου και σε πραγματικό χρόνο υπολογίζει τις τιμές τάσεις των μετρητών, δεδομένης της τρέχουσας ισχύος, έγινε δοκιμή των παραγόμενων δεδομένων στο σύστημα προσομοίωσης και οι τιμές τάσης προέκυψαν φυσιολογικές, χωρίς ανησυχία για βλάβη του δικτύου. Στην συνέχεια έγινε και δοκιμή των παραγόμενων δεδομένων στην περίπτωση διπλασιασμού του πληθυσμού, όπου και εκεί μετά την προσομοίωση δεν υπήρξε κάποιο πρόβλημα στο δίκτυο.

Συμπεραίνουμε, λοιπόν, ότι το μοντέλο που αναπτύχθηκε στην παρούσα εργασία παρήγαγε ρεαλιστικά σενάρια, με σεβασμό στην περιοδικότητα και στον τύπο ημέρας. Επίσης, η υπό συνθήκη παραγωγή βοήθησε στην δημιουργία διαφορετικών σεναρίων ανάλογα την ένταση του καύσωνα, γεγονός που ενθαρρύνει την συγκεκριμένη αρχιτεκτονική. Τέλος, με αυτόν τον τρόπο αποδεικνύεται ότι τα μεγάλα γλωσσικά μοντέλα μπορούν να χρησιμοποιηθούν στην δημιουργία δεδομένων χρονοσειρών και είναι ικανά μοντέλα για την παραγωγή χρονοσειρών σε διαφορετικές συνθήκες καύσωνα. Έτσι, ανοίγει το τοπίο στις πολλές εφαρμογές και περιπτώσεις που μπορεί να χρησιμοποιηθεί.

6.2 Μελλοντικές προεκτάσεις

Αρχικά, θα μπορούσαν να δοκιμαστούν περισσότερα μεγάλα γλωσσικά μοντέλα στην διαδικασία παραγωγής συνθετικών δεδομένων, για να εξασφαλιστεί η επιλογή του καλύτερου γλωσσικού μοντέλου. Επιπλέον, θα μπορούσαν να δοκιμαστούν και περισσότερα σενάρια ακραίων καιρικών συνθηκών εκτός από καύσωνα, όπως, ισχυρές καταιγίδες, κύματα ψύχους, ισχυροί άνεμοι. Έτσι, τα σενάρια που θα μπορούσαν να δημιουργηθούν θα ήταν πολλά περισσότερα και θα παρείχαν μία καλύτερη εποπτεία στο σύνολο. Τα παραγόμενα σενάρια μπορούν να αξιοποιηθούν για την επαύξηση των συνόλων δεδομένων (data augmentation) και τη βελτίωση μοντέλων πρόβλεψης (forecasting), παρέχοντας επιπλέον δείγματα ακραίων ή λιγότερο συχνών καταστάσεων. Με αυτόν τον τρόπο, ενισχύουν την εκπαίδευση και τη γενίκευση των μοντέλων, συμβάλλοντας σε πιο ολοκληρωμένα και αντιπροσωπευτικά σύνολα δεδομένων με μεγαλύτερη ποικιλία σεναρίων ακραίων καιρικών φαινομένων.

Επίσης, το μοντέλο θα μπορούσε να γίνει κομμάτι της προβλεπτικής συντήρησης ενός δικτύου ενέργειας, για τον προσδιορισμό των πιθανών σεναρίων με βάση τις καιρικές συνθήκες. Τα συνθετικά σενάρια μπορούν να αξιοποιηθούν σε διαδικασίες αξιολόγησης της ανθεκτικότητας του δικτύου (resilience assessment) Έτσι, οι ειδικοί μπορούν να μελετήσουν πώς ανταποκρίνεται το δίκτυο σε περιπτώσεις υψηλής ζήτησης ή μειωμένης παραγωγής και να εντοπίσουν τα κρίσιμα σημεία που απαιτούν ενίσχυση για την αποφυγή βλαβών ή αστοχιών.

Μία άλλη προσθήκη είναι να εξεταστεί το μοντέλο σε ένα μεγαλύτερο και πιο πολύπλοκο δίκτυο ηλεκτρικής ενέργειας, με περισσότερους μετρητές και δεδομένα, έτσι ώστε να δοκιμαστεί η εποπτεία του και σε αυτό.

Βιβλιογραφία

- [1] G. Karagiannakis, M. Panteli και S. Argyroudis, «Fragility Modeling of Power Grid Infrastructure for Addressing Climate Change Risks and Adaptation», *WIREs Climate Change*, τόμ. 16, 1 Ιαν. 2025, issn: 1757-7780. doi: 10.1002/wcc.930.
- [2] A. Abur και A. G. Expósito, *Power System State Estimation*. CRC Press, Μαρ. 2004, isbn: 9780203913673. doi: 10.1201/9780203913673.
- [3] A. Ahmadi, M. C. Smith, E. R. Collins, V. Dargahi και S. Jin, «Fast Newton-Raphson Power Flow Analysis Based on Sparse Techniques and Parallel Processing», *IEEE Transactions on Power Systems*, τόμ. 37, σσ. 1695–1705, 3 Μάι. 2022, issn: 0885-8950. doi: 10.1109/TPWRS.2021.3116182.
- [4] T. Kulworawanichpong, «Simplified Newton–Raphson power-flow solution method», *International Journal of Electrical Power Energy Systems*, τόμ. 32, σσ. 551–558, 6 Ιούλ. 2010, issn: 01420615. doi: 10.1016/j.ijepes.2009.11.011.
- [5] T. P. Vishnu, V. Viswan και A. M. Vipin, «Power system state estimation and bad data analysis using weighted least square method», στο *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, IEEE, Δεκ. 2015, σσ. 1–5, isbn: 978-1-4673-8072-0. doi: 10.1109/PICC.2015.7455755.
- [6] S. Radhoush, M. Bahramipanah, H. Nehrir και Z. Shahooei, «A Review on State Estimation Techniques in Active Distribution Networks: Existing Practices and Their Challenges», *Sustainability*, τόμ. 14, σ. 2520, 5 Φεβ. 2022, issn: 2071-1050. doi: 10.3390/su14052520.
- [7] A. Wood, B. Wollenberg και G. Sheblé, *Power Generation, Operation, and Control*. Wiley, 2013, isbn: 9780471790556. διεύθν.: <https://books.google.com.na/books?id=TZaIPwAACAAJ>.
- [8] T. Chen, Y. Cao, X. Chen, L. Sun, J. Zhang και G. A. J. Amaratunga, «A Distributed Maximum-Likelihood-Based State Estimation Approach for Power Systems», *IEEE Transactions on Instrumentation and Measurement*, τόμ. 70, σσ. 1–10, 2021. doi: 10.1109/TIM.2020.3024338.
- [9] D. Mukherjee, S. Chakraborty και S. Ghosh, «Power system state forecasting using machine learning techniques», *Electrical Engineering*, τόμ. 104, σσ. 283–305, 1 Φεβ. 2022, issn: 0948-7921. doi: 10.1007/s00202-021-01328-z.
- [10] V. Kecman, «Support Vector Machines – An Introduction», σσ. 1–47, Απρ. 2005, issn: 1860-0808. doi: 10.1007/10984697_1. διεύθν.: https://link.springer.com/chapter/10.1007/10984697_1.
- [11] L. Kamyabi, T. T. Lie, S. Madanian και S. Marshall, «A Comprehensive Review of Hybrid State Estimation in Power Systems: Challenges, Opportunities and Prospects», *Energies*, τόμ. 17, αρθμ. 19, 2024, issn: 1996-1073. doi: 10.3390/en17194806. διεύθν.: <https://www.mdpi.com/1996-1073/17/19/4806>.
- [12] M. I.-u. Rahman, «Climate Change: a Theoretical Review», *Interdisciplinary Description of Complex Systems*, τόμ. 11, σσ. 1–13, 1 2013, issn: 1334-4676. doi: 10.7906/indecs.11.1.1.

- [13] D. M. Konisky, L. Hughes και C. H. Kaylor, «Extreme weather events and climate change concern», *Climatic Change*, τόμ. 134, σσ. 533–547, 4 Φεβ. 2016, issn: 15731480. doi: 10.1007/s10584-015-1555-3.
- [14] M. Panteli και P. Mancarella, «Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies», *Electric Power Systems Research*, τόμ. 127, σσ. 259–270, Οκτ. 2015, issn: 03787796. doi: 10.1016/j.epsr.2015.06.012.
- [15] A. Shafieezadeh, U. Onyewuchi, M. Begovic και R. Desroches, «Age-dependent fragility models of utility wood poles in power distribution networks against extreme wind hazards», *IEEE Transactions on Power Delivery*, τόμ. 29, σσ. 131–139, 1 2014. doi: 10.1109/TPWRD.2013.2281265.
- [16] M. G. Flammini, N. Andreadou, M. Masera, G. Fulli, S. Vitiello και G. Pretico, *Distribution system operators observatory 2018 – Overview of the electricity distribution system in Europe*. Publications Office, 2019. doi: doi/10.2760/104777.
- [17] A. V. Gkika, E. A. Zacharis, D. N. Skikos και E. L. Lekkas, «Battling the extreme: lessons learned from weather-induced disasters on electricity distribution networks and climate change adaptation strategies», *Hydrology Research*, τόμ. 54, σσ. 1196–1226, 10 Οκτ. 2023, issn: 0029-1277. doi: 10.2166/nh.2023.067.
- [18] *Miscalculation by Spanish power grid operator REE contributed to massive blackout, report finds | Reuters*, Πρόσβαση: 2025-10-09. διεύθυν.: <https://www.reuters.com/business/energy/investigation-into-spains-april-28-blackout-shows-no-evidence-cyberattack-2025-06-17/>.
- [19] A. Di Bella και F. P. Colelli, «Mitigation strategies can alleviate power system vulnerability to climate change and extreme weather: a case study on the Italian grid», *Environmental Research: Infrastructure and Sustainability*, τόμ. 5, αρθμ. 1, σ. 015 003, Ιαν. 2025. doi: 10.1088/2634-4505/ada308. διεύθυν.: <https://doi.org/10.1088/2634-4505/ada308>.
- [20] O. S. Omogoye, K. A. Folly και K. O. Awodele, «Review of Proactive Operational Measures for the Distribution Power System Resilience Enhancement Against Hurricane Events», στο *2021 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, IEEE, Ιαν. 2021, σσ. 1–6, isbn: 978-1-6654-0345-0. doi: 10.1109/SAUPEC/RobMech/PRASA52254.2021.9377252.
- [21] Z. Bie, Y. Lin, G. Li και F. Li, «Battling the Extreme: A Study on the Power System Resilience», *Proceedings of the IEEE*, τόμ. 105, σσ. 1253–1266, 7 Ιούλ. 2017, issn: 0018-9219. doi: 10.1109/JPROC.2017.2679040.
- [22] J. C. Whitson και J. E. Ramirez-Marquez, «Resiliency as a component importance measure in network reliability», *Reliability Engineering System Safety*, τόμ. 94, σσ. 1685–1693, 10 Οκτ. 2009, issn: 0951-8320. doi: 10.1016/J.RESS.2009.05.001. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0951832009001069>.
- [23] Y. Wang, C. Chen, J. Wang και R. Baldick, «Research on Resilience of Power Systems Under Natural Disasters-A Review», *IEEE TRANSACTIONS ON POWER SYSTEMS*, τόμ. 31, 2 2016. doi: 10.1109/TPWRS.2015.2429656. διεύθυν.: http://www.ieee.org/publications_standards/publications/rights/index.html.

- [24] M. Molęda, B. Małysiak-Mrozek, W. Ding, V. Sunderam και D. Mrozek, *From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the Power Industry*, Ιούλ. 2023. doi: 10.3390/s23135970.
- [25] S. B. Righetto, M. A. I. Martins, E. G. Carvalho, L. T. Hattori και S. D. Francisci, «Predictive Maintenance 4.0 Applied in Electrical Power Systems», στο *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, IEEE, Φεβ. 2021, σσ. 1–5, ISBN: 978-1-7281-8897-3. doi: 10.1109/ISGT49243.2021.9372230.
- [26] A. Figueira και B. Vaz, «Survey on Synthetic Data Generation, Evaluation Methods and GANs», *Mathematics*, τόμ. 10, αρθμ. 15, 2022, issn: 2227-7390. doi: 10.3390/math10152733. διεύθν.: <https://www.mdpi.com/2227-7390/10/15/2733>.
- [27] M. Endres, A. Mannarapotta Venugopal και T. S. Tran, «Synthetic Data Generation: A Comparative Study», στο *Proceedings of the 26th International Database Engineered Applications Symposium*, σειρά IDEAS '22, Budapest, Hungary: Association for Computing Machinery, 2022, σσ. 94–102, ISBN: 9781450397094. doi: 10.1145/3548785.3548793. διεύθν.: <https://doi.org/10.1145/3548785.3548793>.
- [28] M. Goyal και Q. H. Mahmoud, «A Systematic Review of Synthetic Data Generation Techniques Using Generative AI», *Electronics*, τόμ. 13, αρθμ. 17, 2024, issn: 2079-9292. doi: 10.3390/electronics13173509. διεύθν.: <https://www.mdpi.com/2079-9292/13/17/3509>.
- [29] A. Bauer, S. Trapp, M. Stenger κ.ά., *Comprehensive Exploration of Synthetic Data Generation: A Survey*, 2024. arXiv: 2401.02524 [cs.LG]. διεύθν.: <https://arxiv.org/abs/2401.02524>.
- [30] I. Valova, K. G. Gabrovska-Evstatieva, T. Kaneva και B. I. Evstatiev, «Generation of Realistic Synthetic Load Profile Based on the Markov Chains Theory: Methodology and Case Studies», *Algorithms*, τόμ. 18, αρθμ. 5, 2025, issn: 1999-4893. doi: 10.3390/a18050287. διεύθν.: <https://www.mdpi.com/1999-4893/18/5/287>.
- [31] D. P. Kingma και M. Welling, «Auto-Encoding Variational Bayes», *CoRR*, τόμ. abs/1312.6114, 2013. doi: 10.48550/arXiv.1312.6114. διεύθν.: <https://api.semanticscholar.org/CorpusID:216078090>.
- [32] X. Zhang, Y. Fu, A. Zang, L. Sigal και G. Agam, *Learning Classifiers from Synthetic Data Using a Multichannel Autoencoder*, Αύγ. 2015. doi: 10.13140/RG.2.1.1700.3367.
- [33] Z. Wan, Y. Zhang και H. He, «Variational autoencoder based synthetic data generation for imbalanced learning», Νοέ. 2017, σσ. 1–7. doi: 10.1109/SSCI.2017.8285168.
- [34] Z. Islam, M. Abdel-Aty, Q. Cai και J. Yuan, «Crash data augmentation using variational autoencoder», *Accident Analysis & Prevention*, τόμ. 151, σ. 105950, Μαρ. 2021. doi: 10.1016/j.aap.2020.105950.
- [35] D. Kaur, S. N. Islam και M. A. Mahmud, «A Variational Autoencoder-Based Dimensionality Reduction Technique for Generation Forecasting in Cyber-Physical Smart Grids», στο *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, σσ. 1–6. doi: 10.1109/ICCWorkshops50388.2021.9473748.
- [36] F. Fahimi, Z. Zhang, B. Goh, K. Ang και C. Guan, «Towards EEG Generation Using GANs for BCI Applications», Μάι. 2019, σσ. 1–4. doi: 10.1109/BHI.2019.8834503.

- [37] M. Patel, X. Wang και S. Mao, «Data augmentation with conditional GAN for automatic modulation classification», Ιούλ. 2020, σσ. 31–36. doi: 10.1145/3395352.3402622.
- [38] V. Sushko, J. Gall και A. Khoreva, «One-Shot GAN: Learning to Generate Samples from Single Images and Videos», στο *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021, σσ. 2596–2600. doi: 10.1109/CVPRW53098.2021.00293.
- [39] H. Naveed, A. U. Khan, S. Qiu κ.ά., *A Comprehensive Overview of Large Language Models*, 2024. doi: 10.48550/arXiv.2307.06435.
- [40] A. Chernyavskiy, D. Ilvovsky και P. Nakov, «Transformers: “The End of History” for Natural Language Processing?», στο *Machine Learning and Knowledge Discovery in Databases. Research Track*, N. Oliver, F. Pérez-Cruz, S. Kramer, J. Read και J. A. Lozano, επιμελητές, Cham: Springer International Publishing, 2021, σσ. 677–693, ISBN: 978-3-030-86523-8.
- [41] A. Vaswani, N. Shazeer, N. Parmar κ.ά., «Attention is All you Need», στο *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio κ.ά., επιμελητές, τόμ. 30, Curran Associates, Inc., 2017. doi: 10.48550/arXiv.1706.03762.
- [42] E. J. Hu, Y. Shen, P. Wallis κ.ά., *LoRA: Low-Rank Adaptation of Large Language Models*, 2021. doi: 10.48550/arXiv.2106.09685.
- [43] T. Dettmers, A. Pagnoni, A. Holtzman και L. Zettlemoyer, *QLoRA: Efficient Finetuning of Quantized LLMs*, 2023. doi: 10.48550/arXiv.2305.14314.
- [44] X. Zhang, R. R. Chowdhury, R. Gupta και J. Shang, «Large Language Models for Time Series: A Survey», Αύγ. 2024, σσ. 8335–8343. doi: 10.24963/ijcai.2024/921.
- [45] C. Raffel, N. Shazeer, A. Roberts κ.ά., *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, 2023. doi: 10.48550/arXiv.1910.10683.
- [46] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza κ.ά., *Generative Adversarial Networks*, 2014. doi: 10.48550/arXiv.1406.2661.
- [47] K. S και M. Durgadevi, «Generative Adversarial Network (GAN): a general review on different variants of GAN and applications», στο *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, 2021, σσ. 1–8. doi: 10.1109/ICCES51350.2021.9489160.
- [48] A. Radford, L. Metz και S. Chintala, *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*, 2016. doi: 10.48550/arXiv.1511.06434.
- [49] M. Arjovsky, S. Chintala και L. Bottou, *Wasserstein GAN*, 2017. doi: 10.48550/arXiv.1701.07875.
- [50] T. Karras, T. Aila, S. Laine και J. Lehtinen, *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, 2018. doi: 10.48550/arXiv.1710.10196.
- [51] T. Karras, S. Laine και T. Aila, *A Style-Based Generator Architecture for Generative Adversarial Networks*, 2019. doi: 10.48550/arXiv.1812.04948.
- [52] X. Chen, Y. Duan, R. Houthoofd, J. Schulman, I. Sutskever και P. Abbeel, *InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets*, 2016. doi: 10.48550/arXiv.1606.03657.

- [53] J. Yoon, D. Jarrett και M. van der Schaar, «Time-series Generative Adversarial Networks», στο *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox και R. Garnett, επιμελητές, τόμ. 32, Curran Associates, Inc., 2019. Διεύθυν.: https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.
- [54] N. Patki, R. Wedge και K. Veeramachaneni, «The Synthetic Data Vault», στο *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2016, σσ. 399–410. doi: 10.1109/DSAA.2016.49.
- [55] M. G. Ogut, B. Meyers και S. P. Boyd, *PV Fleet Modeling via Smooth Periodic Gaussian Copula*, 2023. doi: 10.48550/arXiv.2307.00004.
- [56] M. Turowski, B. Heidrich, L. Weingärtner κ.ά., «Generating synthetic energy time series: A review», *Renewable and Sustainable Energy Reviews*, τόμ. 206, σ. 114 842, 2024, issn: 1364-0321. doi: <https://doi.org/10.1016/j.rser.2024.114842>. Διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S1364032124005689>.
- [57] P. Chen, T. Pedersen, B. Bak-Jensen και Z. Chen, «ARIMA-Based Time Series Model of Stochastic Wind Power Generation», *IEEE Transactions on Power Systems*, τόμ. 25, αρθμ. 2, σσ. 667–676, 2010. doi: 10.1109/TPWRS.2009.2033277.
- [58] Y. Li, B. Hu, T. Niu κ.ά., «GMM-HMM-Based Medium- and Long-Term Multi-Wind Farm Correlated Power Output Time Series Generation Method», *IEEE Access*, τόμ. 9, σσ. 90 255–90 267, 2021. doi: 10.1109/ACCESS.2021.3091460.
- [59] W. Dong, X. Chen και Q. Yang, «Data-driven scenario generation of renewable energy production based on controllable generative adversarial networks with interpretability», *Applied Energy*, τόμ. 308, σ. 118 387, 2022, issn: 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2021.118387>. Διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0306261921016251>.
- [60] Y. Chen, Y. Wang, D. Kirschen και B. Zhang, «Model-Free Renewable Scenario Generation Using Generative Adversarial Networks», *IEEE Transactions on Power Systems*, τόμ. 33, σσ. 3265–3275, 3 Μάι. 2018, issn: 0885-8950. doi: 10.1109/TPWRS.2018.2794541.
- [61] R. Tang, P. Leong, J. Dore και A. Vassallo, «Generating Residential PV Production and Electricity Consumption Scenarios via Generative Adversarial Networks», στο *Asia Pacific Solar Research Conference (APSRC)*, Δεκ. 2018.
- [62] C. Zhang, S. R. Kuppannagari, R. Kannan και V. K. Prasanna, «Generative Adversarial Network for Synthetic Time Series Data Generation in Smart Grids», στο *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2018, σσ. 1–6. doi: 10.1109/SmartGridComm.2018.8587464.
- [63] G. Baasch, G. Rousseau και R. Evins, «A Conditional Generative adversarial Network for energy use in multiple buildings using scarce data», *Energy and AI*, τόμ. 5, σ. 100 087, 2021, issn: 2666-5468. doi: <https://doi.org/10.1016/j.egyai.2021.100087>. Διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S2666546821000410>.
- [64] J. Lan, Q. Guo και H. Sun, «Demand Side Data Generating Based on Conditional Generative Adversarial Networks», *Energy Procedia*, τόμ. 152, σσ. 1188–1193, 2018, Cleaner Energy for Cleaner Cities, issn: 1876-6102. doi: <https://doi.org/10.1016/j.egypro.2018.09.157>. Διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S187661021830701X>.

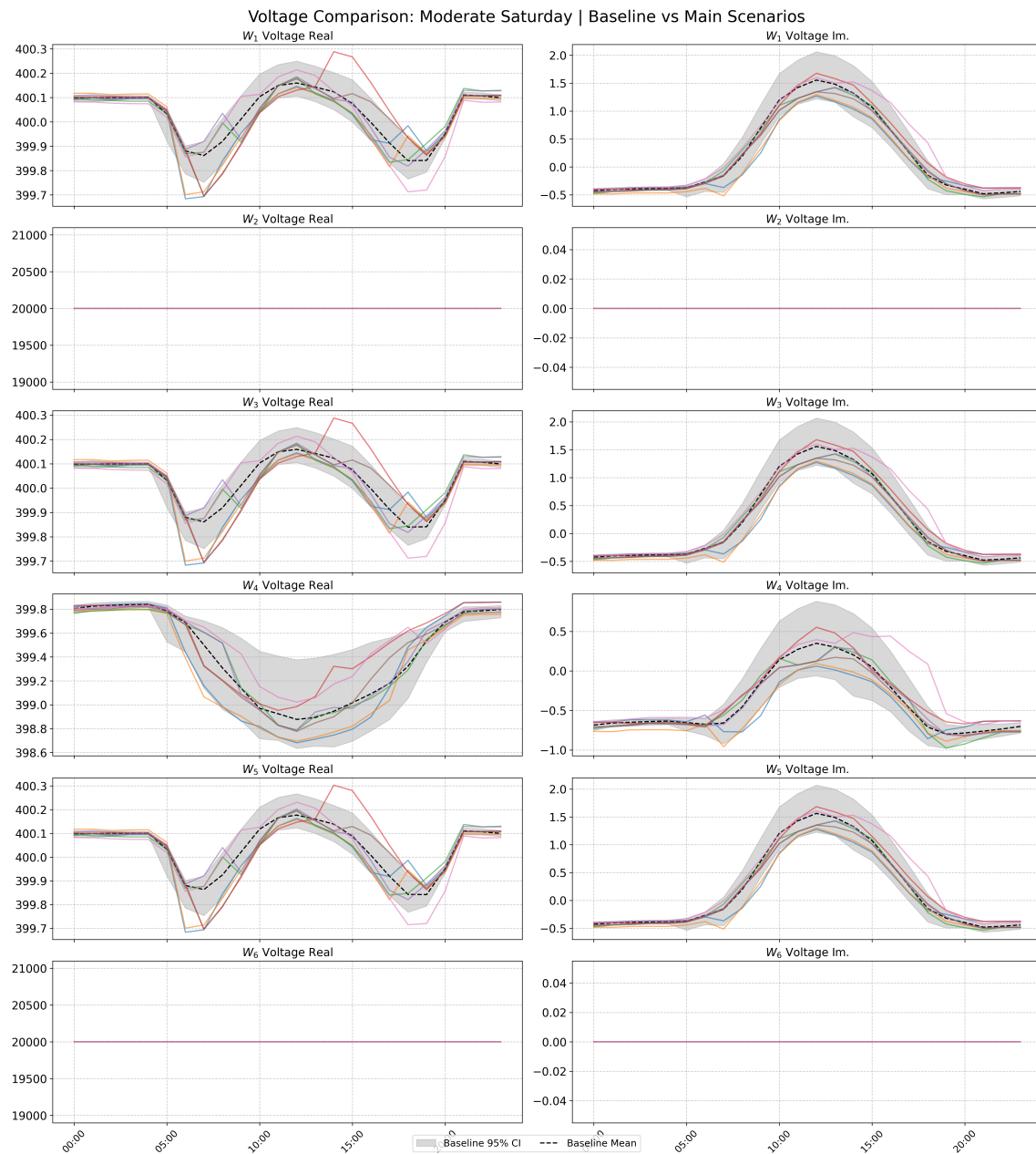
- [65] M. N. Fekri, A. M. Ghosh και K. Grolinger, «Generating Energy Data for Machine Learning with Recurrent Generative Adversarial Networks», *Energies*, τόμ. 13, αρθμ. 1, 2020, issn: 1996-1073. doi: 10.3390/en13010130. διεύθυν.: <https://www.mdpi.com/1996-1073/13/1/130>.
- [66] C. Esteban, S. L. Hyland και G. Rätsch, *Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs*, 2017. doi: 10.48550/arXiv.1706.02633.
- [67] P. Tang, Z. Li, X. Wang, X. Liu και P. Mou, «Time Series Data Augmentation for Energy Consumption Data Based on Improved TimeGAN», *Sensors*, τόμ. 25, αρθμ. 2, 2025, issn: 1424-8220. doi: 10.3390/s25020493. διεύθυν.: <https://www.mdpi.com/1424-8220/25/2/493>.
- [68] R. D. Santos, J. Aguilar και M. D. R-Moreno, «A synthetic Data Generator for Smart Grids based on the Variational-Autoencoder Technique and Linked Data Paradigm», στο *2022 XLVIII Latin American Computer Conference (CLEI)*, 2022, σσ. 1–7. doi: 10.1109/CLEI56649.2022.9959918.
- [69] X. Gong, B. Tang, R. Zhu, W. Liao και L. Song, «Data Augmentation for Electricity Theft Detection Using Conditional Variational Auto-Encoder», *Energies*, τόμ. 13, αρθμ. 17, 2020, issn: 1996-1073. doi: 10.3390/en13174291. διεύθυν.: <https://www.mdpi.com/1996-1073/13/17/4291>.
- [70] C. Fan, M. Chen, R. Tang και J. Wang, «A novel deep generative modeling-based data augmentation strategy for improving short-term building energy predictions», *Building Simulation*, τόμ. 15, σσ. 197–211, 2 Φεβ. 2022, issn: 19968744. doi: 10.1007/S12273-021-0807-6/METRICS. διεύθυν.: <https://link.springer.com/article/10.1007/s12273-021-0807-6>.
- [71] M. Jin, S. Wang, L. Ma κ.ά., *Time-LLM: Time Series Forecasting by Reprogramming Large Language Models*, 2024. doi: 10.48550/arXiv.2310.01728.
- [72] H. Lin και M. Yu, *A Novel Distributed PV Power Forecasting Approach Based on Time-LLM*, 2025. doi: 10.48550/arXiv.2503.06216.
- [73] H. Xue και F. D. Salim, *Utilizing Language Models for Energy Load Forecasting*, 2023. doi: 10.48550/arXiv.2310.17788.
- [74] Z. Qiu, C. Li, Z. Wang κ.ά., *FPE-LLM: Highly Intelligent Time-Series Forecasting and Language Interaction LLM in Energy Systems*, Οκτ. 2024. doi: 10.48550/arXiv.2411.00852.
- [75] S. Abdullahi, K. Usman Danyaro, A. Zakari, I. Abdul Aziz, N. Amila Wan Abdullah Zawawi και S. Adamu, «Time-Series Large Language Models: A Systematic Review of State-of-the-Art», *IEEE Access*, τόμ. 13, σσ. 30 235–30 261, 2025. doi: 10.1109/ACCESS.2025.3535782.
- [76] X. Zhou, Q. Jia, Y. Hu, R. Xie, T. Huang και F. R. Yu, «GenG: An LLM-Based Generic Time Series Data Generation Approach for Edge Intelligence via Cross-Domain Collaboration», στο *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2024, σσ. 1–6. doi: 10.1109/INFOCOMWKSHPS61880.2024.10620716.
- [77] A. Grigoraş και F. Leon, «Synthetic Time Series Generation for Decision Intelligence Using Large Language Models», *Mathematics*, τόμ. 12, αρθμ. 16, 2024, issn: 2227-7390. doi: 10.3390/math12162494. διεύθυν.: <https://www.mdpi.com/2227-7390/12/16/2494>.

- [78] C. Rousseau, T. Boschi, G. Cornacchia, D. Salwala, A. Pascale και J. B. Moreno, *Forging Time Series with Language: A Large Language Model Approach to Synthetic Data Generation*, 2025. doi: 10.48550/arXiv.2505.17103.
- [79] X. Wang, X. Liu, F. Zhong, Z. Li, K. Xuan και Z. Zhao, «A Scenario Generation Method for Typical Operations of Power Systems with PV Integration Considering Weather Factors», *Sustainability*, τόμ. 15, αρθμ. 20, 2023, issn: 2071-1050. doi: 10.3390/su152015007. διεύθυν.: <https://www.mdpi.com/2071-1050/15/20/15007>.
- [80] Y. Xiao, X. Hu, Y. Lin, Y. Lu, R. Jing και Y. Zhao, «Interpretable short-term electricity load forecasting considering small sample heatwaves», *Applied Energy*, τόμ. 398, αρθμ. C, None, NoneNone 2025. doi: 10.1016/j.apenergy.2025.126417. διεύθυν.: <https://ideas.repec.org/a/eee/appene/v398y2025ics030626192501147x.html>.
- [81] A. Grantham, P. Pudney και J. Boland, «Generating synthetic sequences of global horizontal irradiation», *Solar Energy*, τόμ. 162, σσ. 500–509, 2018, issn: 0038-092X. doi: <https://doi.org/10.1016/j.solener.2018.01.044>. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0038092X18300653>.
- [82] X. Li, C. Li και C. Jia, «Electric Vehicle and Photovoltaic Power Scenario Generation under Extreme High-Temperature Weather», *World Electric Vehicle Journal*, τόμ. 15, αρθμ. 1, 2024, issn: 2032-6653. doi: 10.3390/wevj15010011. διεύθυν.: <https://www.mdpi.com/2032-6653/15/1/11>.
- [83] W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, Y. Wang και K. Liu, «Scenario Generations for Renewable Energy Sources and Loads Based on Implicit Maximum Likelihood Estimations», *Journal of Modern Power Systems and Clean Energy*, τόμ. 10, αρθμ. 6, σσ. 1563–1575, 2022. doi: 10.35833/MPCE.2022.000108.
- [84] B. Austgen, M. Garcia, B. Pierre, J. Hasenbein και E. Kutanoglu, «Winter Storm Scenario Generation for Power Grids Based on Historical Generator Outages», στο *2022 IEEE/PES Transmission and Distribution Conference and Exposition (TD)*, 2022, σσ. 1–5. doi: 10.1109/TD43745.2022.9816863.
- [85] M. Sahraei-Ardakani και G. Ou, «Day-ahead preventive scheduling of power systems during natural hazards via stochastic optimization», στο *2017 IEEE Power Energy Society General Meeting*, 2017, σσ. 1–1. doi: 10.1109/PESGM.2017.8274453.
- [86] J. Muñoz Sabater, *ERA5-Land hourly data from 1950 to present*, <https://doi.org/10.24381/cds.e2161bac>, Πρόσβαση 27-02-2025, 2019. doi: 10.24381/cds.e2161bac.
- [87] J. Peppanen, X. Zhang, S. Grijalva και M. J. Reno, «Handling bad or missing smart meter data through advanced data imputation», στο *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, IEEE, Σεπτ. 2016, σσ. 1–5, isbn: 978-1-5090-5167-0. doi: 10.1109/ISGT.2016.7781213.
- [88] *Missing Values Filler - Darts 0.22.0 documentation*. διεύθυν.: https://unit8co.github.io/darts/generated_api/darts.dataprocessing.transformers.missing_values_filler.html?highlight=missing%20value#missing-values-filler (επίσκεψη 02/08/2025).
- [89] J. Tan, J. Yang, S. Wu, G. Chen και J. Zhao, *A critical look at the current train/test split in machine learning*, 2021. doi: 10.48550/arXiv.2106.04525.
- [90] V. Sharma, «A Study on Data Scaling Methods for Machine Learning», *Quarterly International Journal For Global Academic & Scientific Research (IJGASR)*, τόμ. 1, 1 2022. doi: 10.55938/ijgasr.v1i1.4.

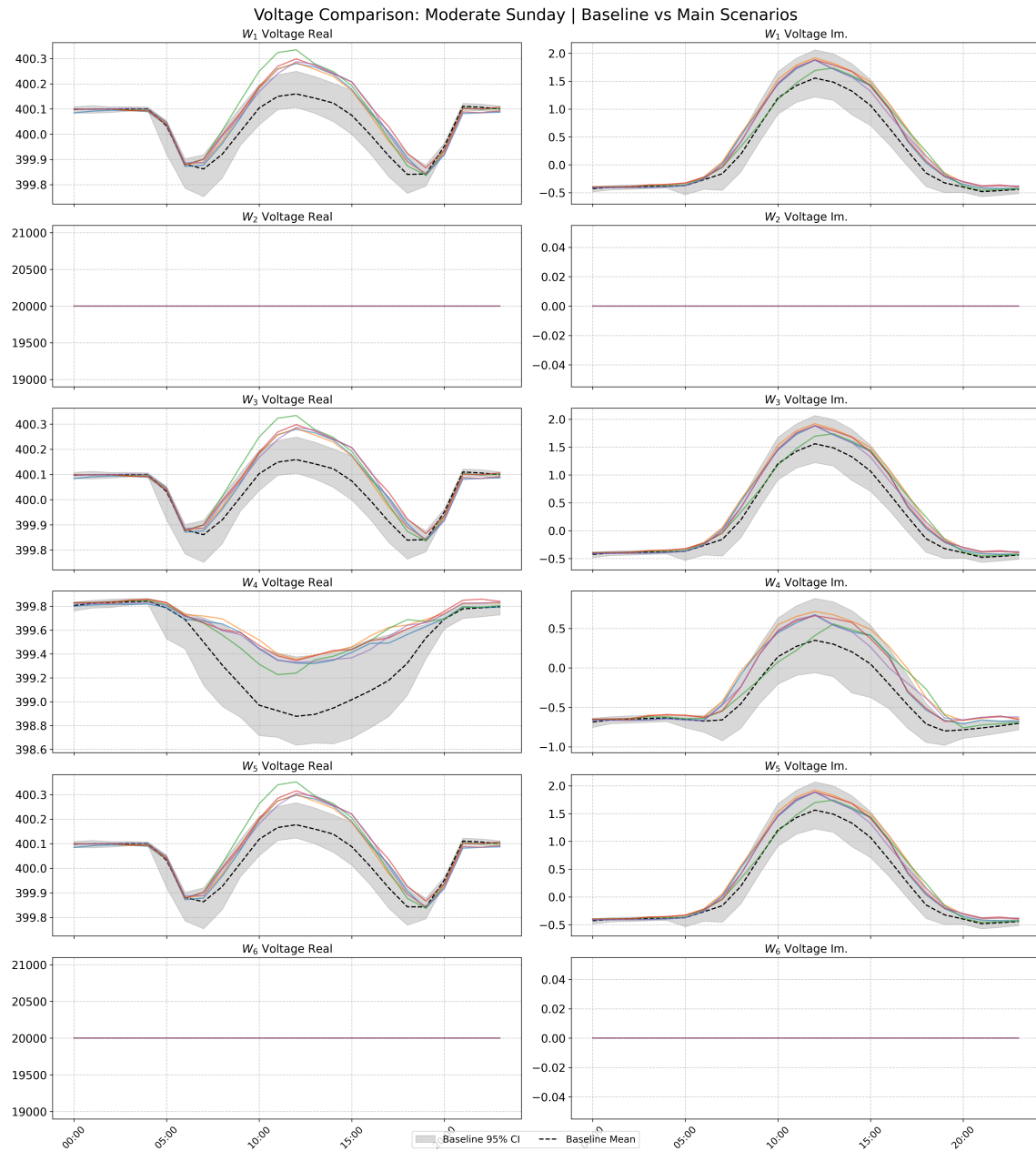
- [91] *google-t5/t5-base* · *Hugging Face*, Πρόσβαση στις 20-02-2025. διεύθυν.: <https://huggingface.co/google-t5/t5-base>.
- [92] P. Zhou, X. Xie, Z. Lin και S. Yan, «Towards Understanding Convergence and Generalization of AdamW», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, τόμ. 46, αρθμ. 9, σσ. 6486–6493, 2024. doi: 10.1109/TPAMI.2024.3382294.
- [93] P. Senin, «Dynamic time warping algorithm review», *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, τόμ. 855, αρθμ. 1-23, σ. 40, 2008.
- [94] Z. Mehraban και A. Pichler, «Quantization Of Probability Measures In Maximum Mean Discrepancy Distance», 2025. doi: 10.48550/arXiv.2503.11868.
- [95] M. R. Berthold και F. Höppner, «On Clustering Time Series Using Euclidean Distance and Pearson Correlation», 2016. doi: 10.48550/arXiv.1601.02213.
- [96] M. Mirz, S. Vogel, G. Reinke και A. Monti, «DPsim—A dynamic phasor real-time simulator for power systems», *SoftwareX*, τόμ. 10, σ. 100 253, 2019, issn: 2352-7110. doi: <https://doi.org/10.1016/j.softx.2019.100253>. διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S2352711018302760>.
- [97] E. C. for Electrotechnical Standardization (CENELEC), *EN 50160: Voltage characteristics of electricity supplied by public distribution systems*, Brussels, Belgium: CENELEC, 2010.

Παραρτήματα

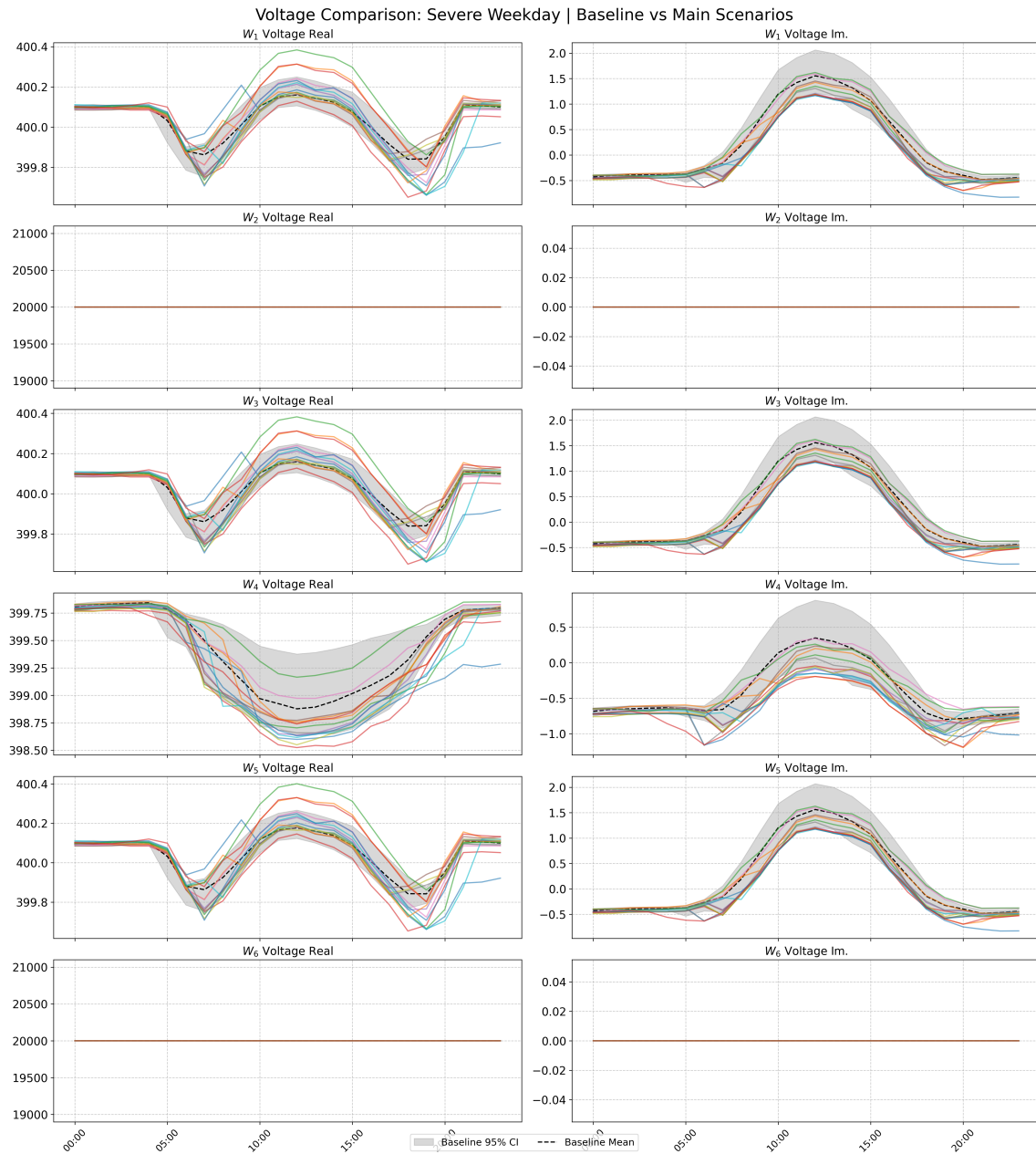
Α Αποτελέσματα προσομοίωσης DPsim όλων των περιπτώσεων



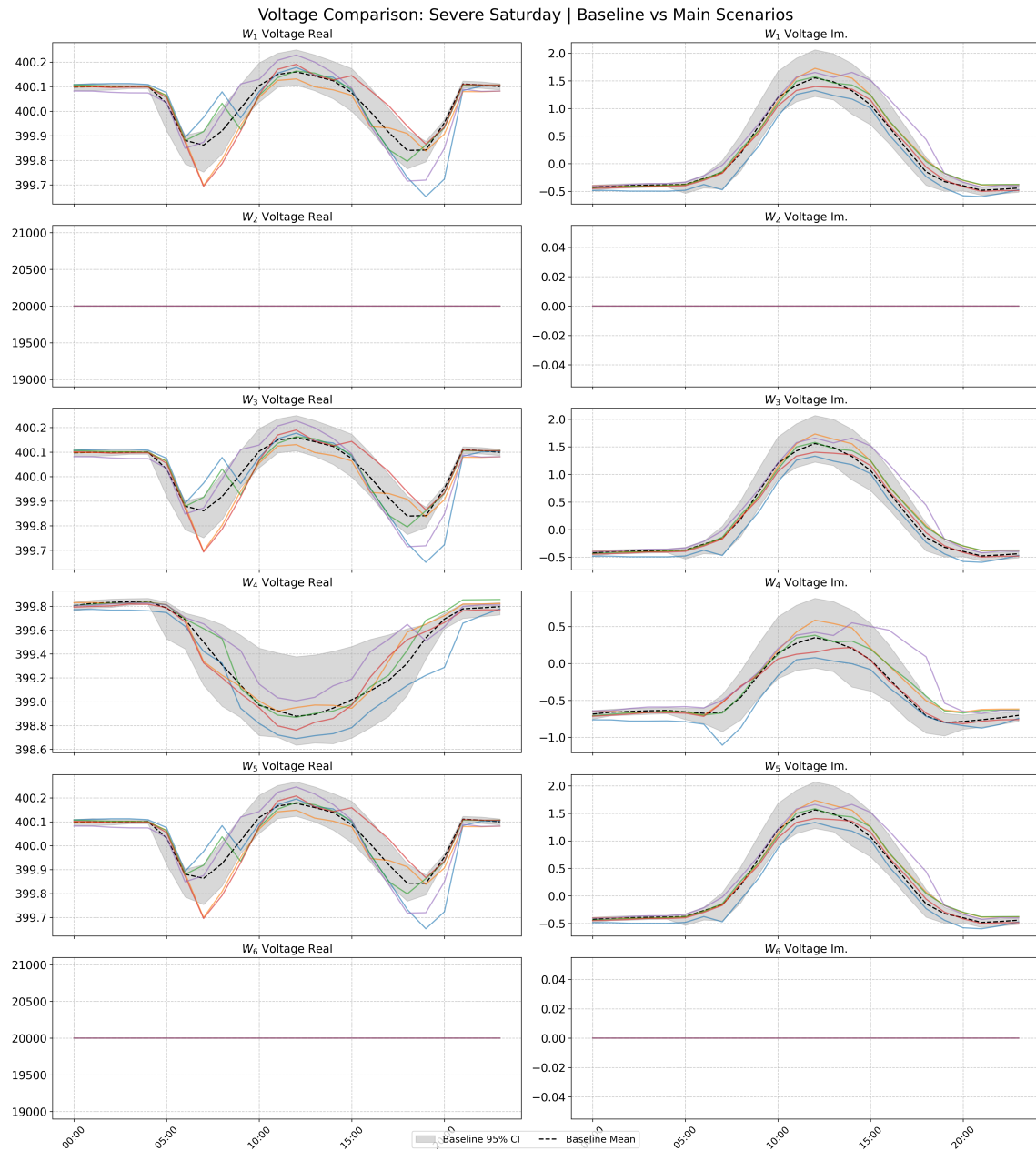
Σχήμα A.1: Αποτελέσματα προσομοίωσης περίπτωσης Moderate Saturday



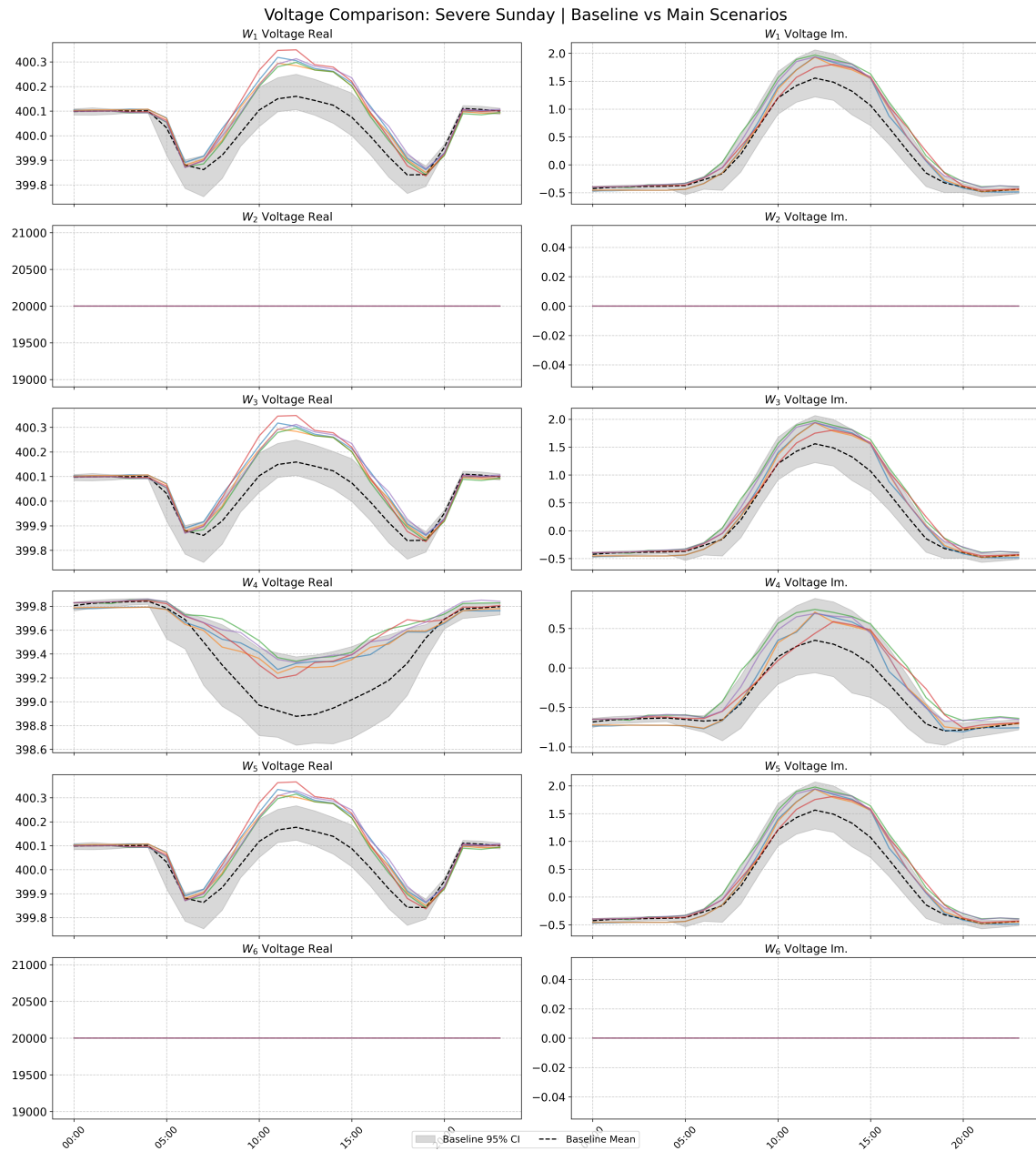
Σχήμα Α.2: Αποτελέσματα προσομοίωσης περίπτωσης Moderate Sunday



Σχήμα Α.3: Αποτελέσματα προσομοίωσης περίπτωσης Severe Weekday

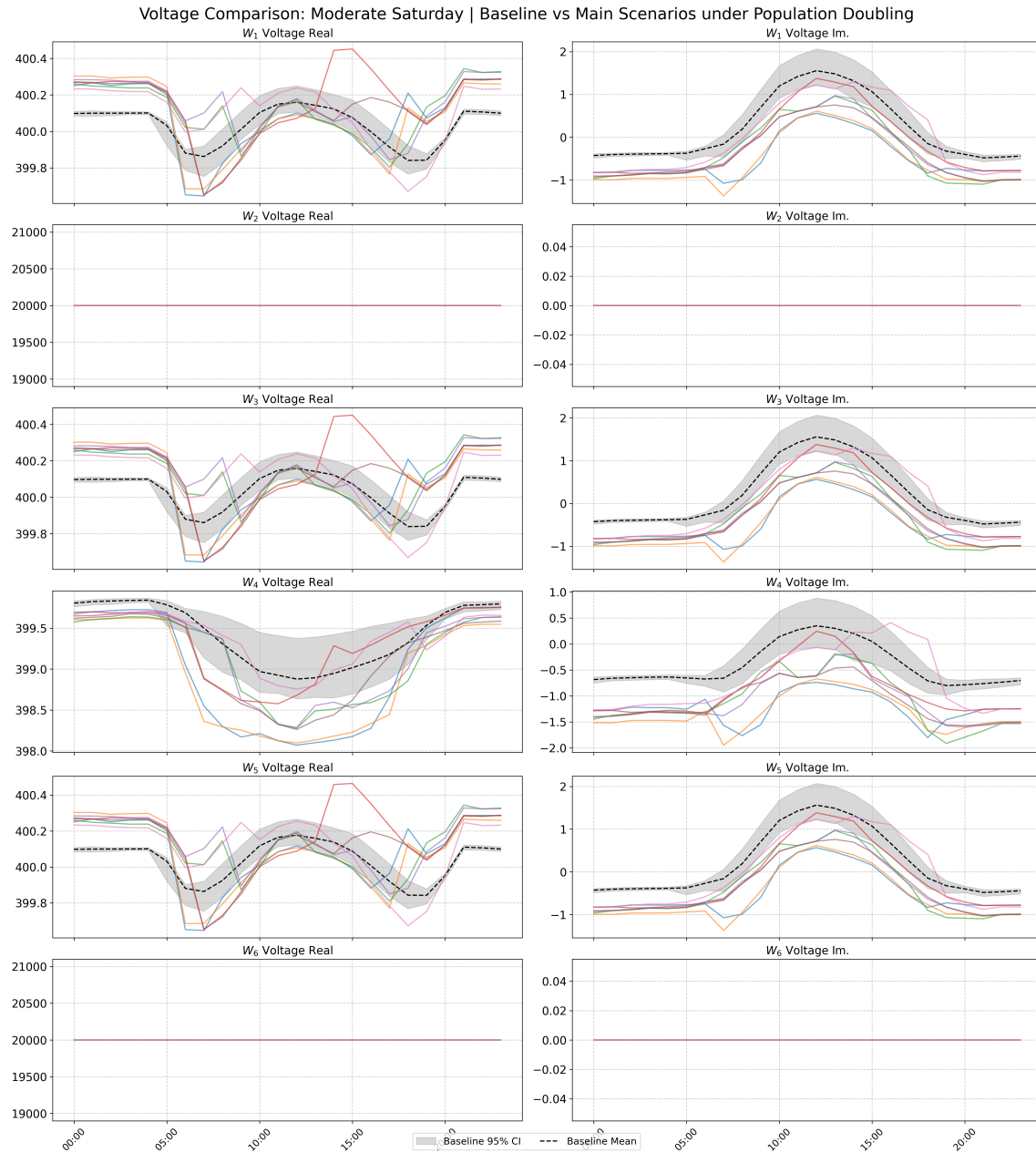


Σχήμα Α.4: Αποτελέσματα προσομοίωσης περίπτωσης Severe Saturday

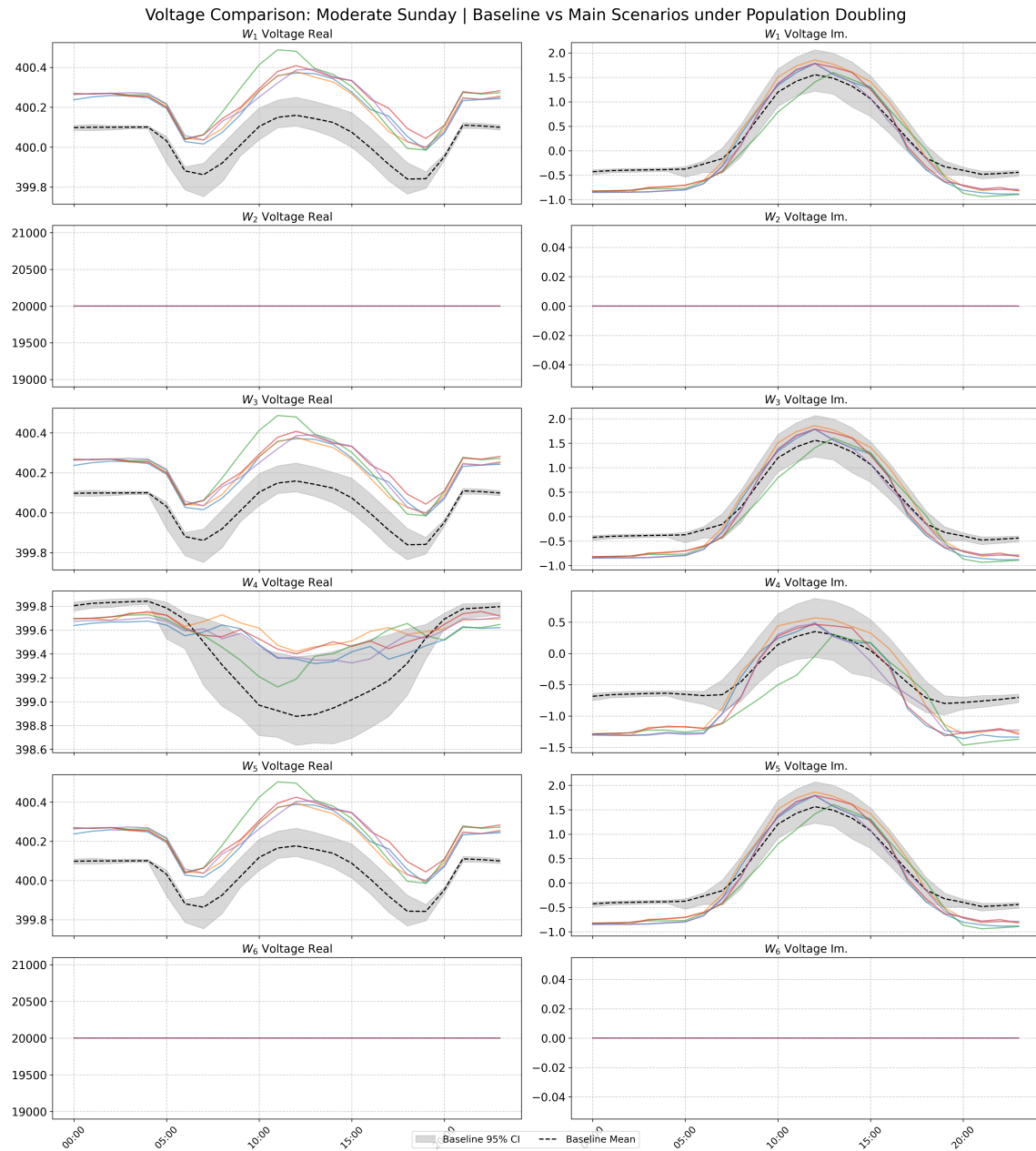


Σχήμα Α.5: Αποτελέσματα προσομοίωσης περίπτωσης Severe Sunday

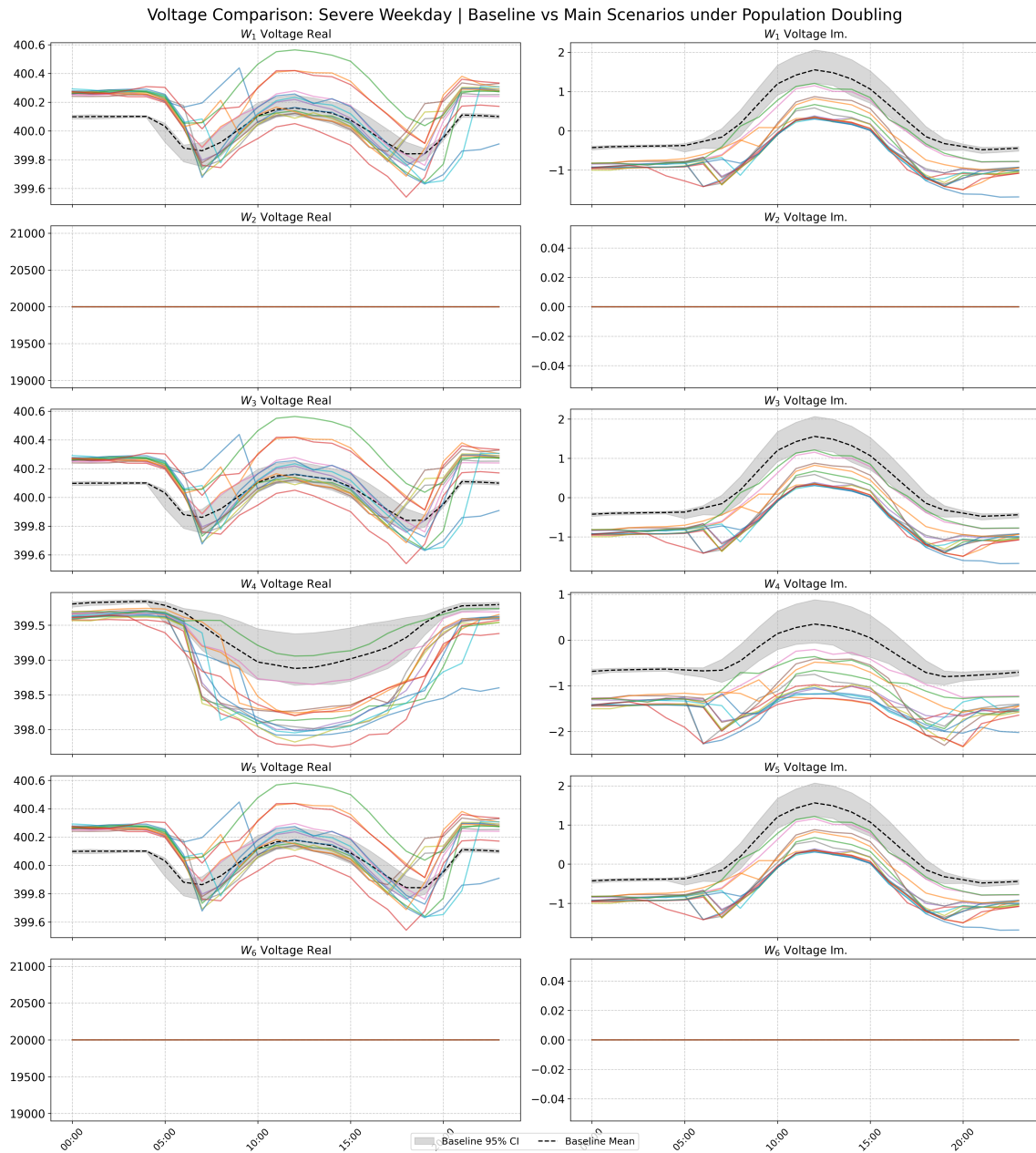
B Αποτελέσματα προσομοίωσης DPsim όλων των περιπτώσεων στην δοκιμή διπλασιασμού του πληθυσμού



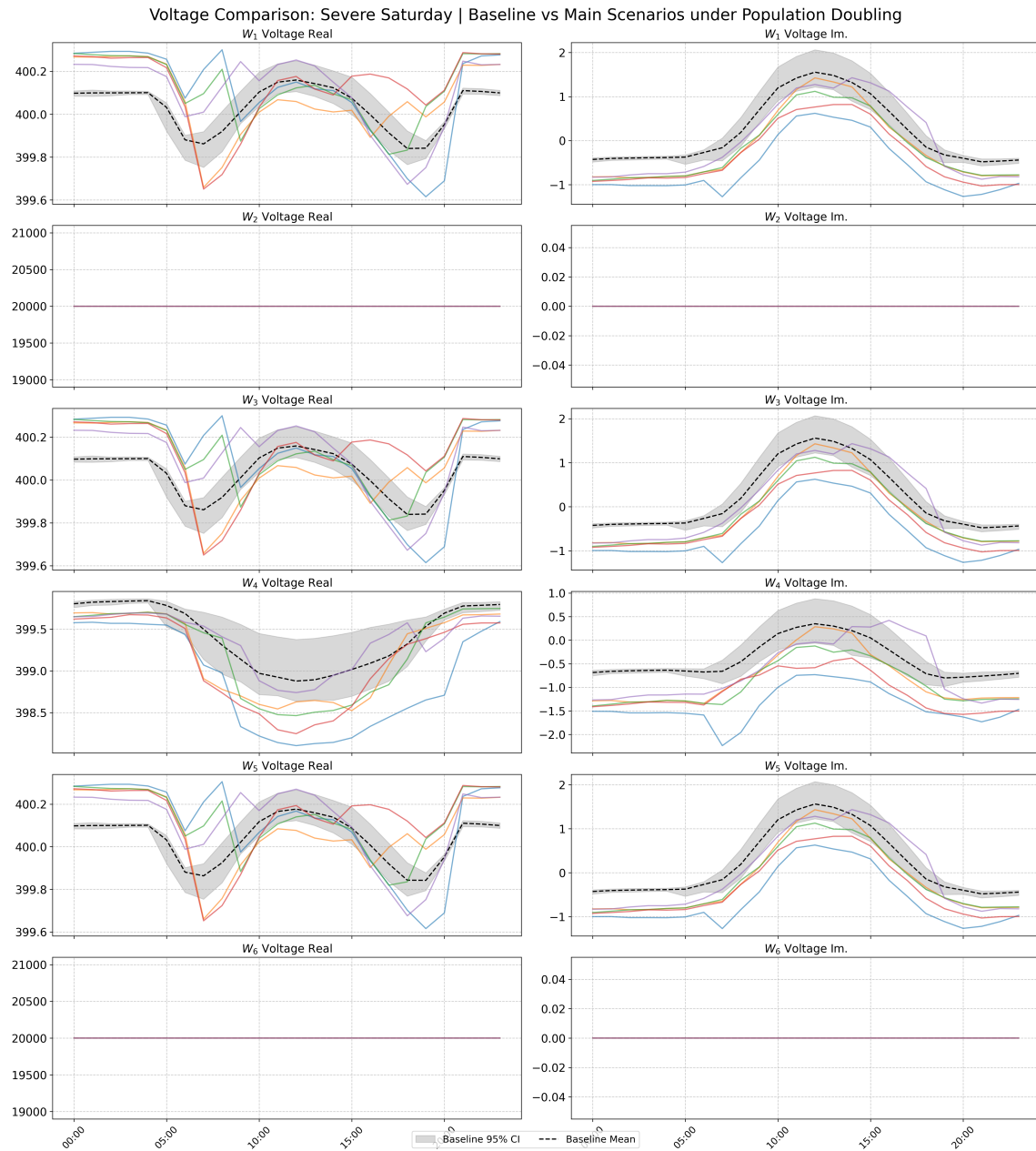
Σχήμα B.1: Αποτελέσματα προσομοίωσης περίπτωσης Moderate Saturday



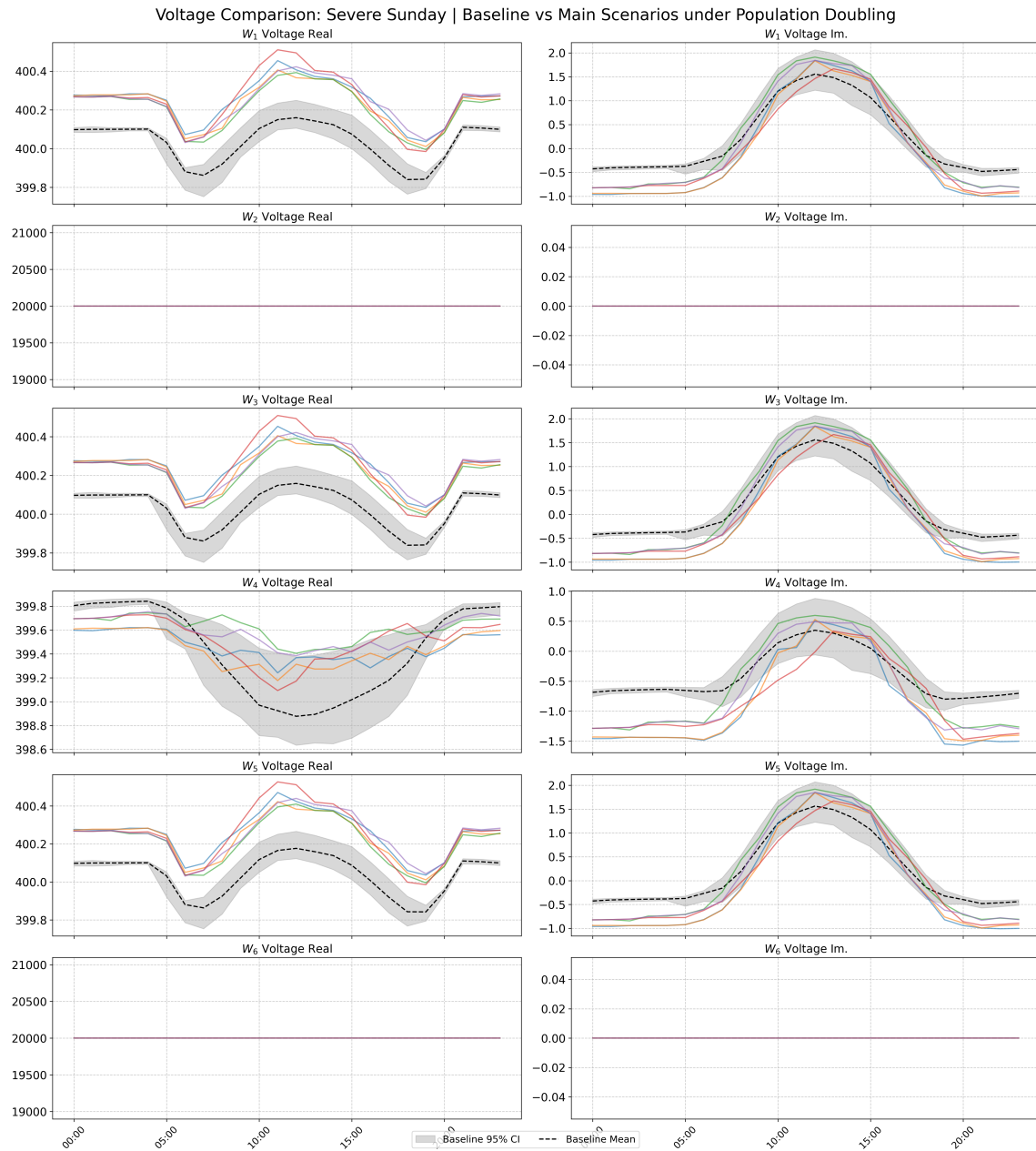
Σχήμα Β.2: Αποτελέσματα προσομοίωσης περίπτωσης Moderate Sunday



Σχήμα Β.3: Αποτελέσματα προσομοίωσης περίπτωσης Severe Weekday



Σχήμα Β.4: Αποτελέσματα προσομοίωσης περίπτωσης Severe Saturday



Σχήμα Β.5: Αποτελέσματα προσομοίωσης περίπτωσης Severe Sunday