



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων

**Ανάλυση Χαρτοφυλακίων Αιολικών Πάρκων για τη βέλτιστη
Διαφοροποίηση και Διαχείρισή τους**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΗΜΗΤΡΙΟΣ ΙΩΑΝΝΙΔΗΣ

Υπεύθυνος : Ευάγγελος Σπηλιώτης
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Προβλέψεων και Στρατηγικής

**Ανάλυση Χαρτοφυλακίων Αιολικών Πάρκων για τη βέλτιστη
Διαφοροποίηση και Διαχείρισή τους**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΗΜΗΤΡΙΟΣ ΙΩΑΝΝΙΔΗΣ

Υπεύθυνος: Ευάγγελος Σπηλιώτης
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16η Οκτωβρίου 2025.

.....
Ευάγγελος Σπηλιώτης
Επίκ. Καθηγητής Ε.Μ.Π.

.....
Ευάγγελος Μαρινάκης
Επίκ. Καθηγητής Ε.Μ.Π.

.....
Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025

Copyright © Δημήτριος Ιωαννίδης, 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)

.....

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

Περίληψη

Στη σύγχρονη εποχή, η αγορά των ανανεώσιμων πηγών ενέργειας βρίσκεται σε συνεχή ανάπτυξη και οι φορείς που εισχωρούν σε αυτή γίνονται όλο και πολυπληθέστεροι. Αυτοί, ως παραγωγοί, κάνουν προβλέψεις για τα ποσά ενέργειας που θα παράξουν, όμως ενδεχόμενες αστοχίες στις εκτιμήσεις τους επιφέρουν ρήτρες, γεγονός που πλήττει την αξιοπιστία τους, την κερδοφορία τους, καθώς και την ακρίβεια προγραμματισμού που μπορεί να κάνει ο διαχειριστής του ενεργειακού δικτύου. Η παρούσα διπλωματική εστιάζει στην εύρεση μιας μεθόδου μείωσης των σφαλμάτων αυτών και των χρηματικών ποινών στις οποίες υποβάλλονται, με χρήση βέλτιστα καταστρωμένων χαρτοφυλακίων αιολικών πάρκων.

Άφου έγινε μια σύντομη παρουσίαση της αγοράς ενέργειας στην Ελλάδα και στις σημαντικότερες έννοιες της, αναπτύχθηκαν στατιστικοί δείκτες περιγραφής των ιδιαίτερων χαρακτηριστικών της παραγωγής αιολικών πάρκων, που υπολογίστηκαν από τις διαθέσιμες χρονοσειρές των δοσμένων αιολικών πάρκων, έτσι ώστε να χρησιμοποιηθούν στην συσταδοποίηση των εν λόγω πάρκων. Σε επόμενο βήμα αναπτύχθηκαν εκτενώς οι ποικίλες μέθοδοι συσταδοποίησης και αξιολόγησης ποιότητάς τους, ενώ έγινε αναφορά και σε χρήσιμες τεχνικές επεξεργασίας των δεδομένων. Για τη συσταδοποίηση, δοκιμάστηκαν διάφορες μέθοδοι με στόχο την επιβεβαίωση της στιβαρότητας της κατηγοριοποίησης, και αποδείχθηκε πως οι συστάδες είναι ισχυρά και κατά ουσιαστικό τρόπο χωρισμένες η μια από την άλλη. Στη συνέχεια, εξετάστηκαν τα χαρακτηριστικά της κάθε συστάδας που την καθιστούν διαφορετική από τις άλλες, ενώ επίσης εντοπίστηκαν συστάδες με μεγάλο βαθμό συμπληρωματικότητας.

Το τελευταίο σκέλος της εργασίας αφορά την αξιοποίηση πιθανοτικών μοντέλων προσομοίωσης, ώστε να εντοπιστούν με όσο δυνατόν μεγαλύτερη ακρίβεια βάσει πιθανοτήτων οι συνδυασμοί εκείνοι από συστάδες που ελαχιστοποιούν την διακύμανση της παραγωγής ενέργειας ενός χαρτοφυλακίου, και επομένως ελαχιστοποιούν τον κίνδυνο οι ταχύτητες των ανέμων να παρεκκλίνουν από την αναμενόμενη μέση τιμή τους. Τα αποτελέσματα των προσομοιώσεων επιβεβαίωσαν τις αρχικές εκτιμήσεις της συμπληρωματικής συμπεριφοράς μερικών συνδυασμών συστάδων, και κατέδειξαν την πιθανότητα με βάση ένα αρχικό χαρτοφυλάκιο τυχαίων αιολικών πάρκων ο έμπορος με στοχευμένες προσθήκες πάρκων από προσεκτικά επιλεγμένες συστάδες να μειώσει σε πολύ ουσιαστικό βαθμό τη διακύμανση των πάρκων που κατέχει. Στην κατακλείδα, έγινε αναφορά σε μελλοντικές επεκτάσεις της παρούσας έρευνας, για βελτίωση της μεθοδολογίας που θα συμβαδίζει με την αναμενόμενη περαιτέρω ανάπτυξη της αγοράς ενέργειας.

Λέξεις Κλειδιά:

συστάδες, χαρτοφυλάκια, ανανεώσιμες πηγές ενέργειας, ελάττωση διακύμανσης, διαφοροποίηση, συμπληρωματικότητα, αιολικό πάρκο

Abstract

In the modern era, the renewable energy market is constantly growing and the number of merchants entering it is increasing. As producers, they make predictions about the amounts of energy they will produce, but any inaccuracies in their estimates result in penalties, which undermine their credibility and reliability. This thesis focuses on finding a method to minimize these errors and the financial penalties they incur, using optimally constructed energy portfolios.

After a brief presentation of the Energy Market in Greece and its most important concepts, statistical distribution data, change indicators, normalized average values and seasonality intensities were drawn from the available time series of the given wind farms, so that they could be used in the clustering of these farms, and subsequently expanded upon. In the following step, the various clustering and quality of clustering assessment methods were discussed in detail, while useful data processing tools were also mentioned. For the clustering, various methods were tested to confirm the robustness of the categorization, and it was found that the clusters are strong and essentially separate from each other. Next, the characteristics of each cluster were briefly discussed and clusters with a high degree of complementarity were identified.

The last part of the body of work concerns the use of probabilistic simulation models to identify, as accurately as possible based on probabilities, those combinations of clusters that minimize the variance of a portfolio's time series and therefore minimize the risk of wind speeds deviating from their expected mean value. The results of the simulations confirmed the initial estimates of the complementary behavior of some cluster combinations and demonstrated the possibility, based on an initial portfolio of random wind parks, for a trader with additions of random parks from targeted clusters to significantly reduce the variance of the parks he owns. In conclusion, reference was made to future extensions of the present research to improve the methodology in line with the expected further development of the energy market.

Keywords:

clusters, portfolios, renewable energy sources, variance reduction, diversification, complementarity, wind park

Ευχαριστίες

Η παρούσα εργασία εκπονήθηκε στο πλαίσιο ολοκλήρωσης του προπτυχιακού κύκλου σπουδών της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα αρχικά να ευχαριστήσω όλους τους καθηγητές και του επιβλέποντες της Μονάδας Προβλέψεων και Στρατηγικής για την ευκαιρία που μου έδωσαν και για την επιλογή της παρούσας διπλωματικής. Πιο συγκεκριμένα, θα ήθελα να ευχαριστήσω τον καθηγητή και υπεύθυνο μου Ευάγγελο Σπηλιώτη, για τις στοχευμένες συμβουλές του και την υπομονή που έδειξε καθ' όλη τη διάρκεια εκπόνησης της εργασίας.

Θα ήθελα επίσης να ευχαριστήσω θερμά τους γονείς μου, Βασίλη και Ασημίνα, καθώς και τον αδερφό μου Γιώργο, για όλη τη συμπαράσταση που μου προσέφεραν απλόχερα όλα αυτά τα χρόνια, και ιδιαίτερα τα τελευταία των σπουδών μου. Ήταν πάντα εκεί όταν τους χρειάστηκα για συμβουλές, και κατά κάποιον τρόπο μόνο εκείνοι ξέρουν για τον άθλο μου. Δε θα τα είχα καταφέρει χωρίς αυτούς.

Οι φίλοι μου όλων αυτών των χρόνων, από ή και όχι από τη σχολή, εντός ή και εκτός Ελλάδας, επίσης με στήριξαν πέρα από τα βασικά και μου έδωσαν την απαραίτητη ώθηση να συνεχίσω, ακόμα και όταν η διαδρομή ήταν δύσκολη. Μικρά λόγια ενθάρρυνσης ήταν αυτά που χρειαζόμουν πολλές φορές, και αισθάνομαι τυχερός που είχα ανθρώπους εκεί για μένα για να μου τα δώσουν στην ανάγκη. Δε θα τα είχα καταφέρει ούτε χωρίς αυτούς.

Δημήτρης

Περιεχόμενα

Κεφάλαιο 1 : Εισαγωγή.....	18
1.1 Ανανεώσιμες Πηγές Ενέργειας και Ενεργειακή Μετάβαση.....	18
1.2 Το Ενεργειακό Χρηματιστήριο στην Ελλάδα.....	20
1.3 Έννοια και Σημασία του Ενεργειακού Χαρτοφυλακίου.....	24
1.4 Σύγχρονη Θεωρία Χαρτοφυλακίων.....	25
1.5 Αντικείμενο και Σκοπός της Εργασίας.....	27
Κεφάλαιο 2 : Χρήσιμα Χαρακτηριστικά.....	29
2.1 Δημιουργία και Επιλογή Στατιστικών Δεικτών για την Περιγραφή των Χαρακτηριστικών Παραγωγής Ενέργειας.....	30
2.1.1. Χαρακτηριστικά Διασποράς και Κατανομής.....	30
2.1.2. Δείκτες Μεταβολής της Χρονοσειράς.....	33
2.1.3. Ημερήσια Κανονικοποιημένη Μέση Τιμή Ανέμου.....	35
2.1.4. Ένταση Εποχικότητας Χρονοσειράς.....	35
2.2 Μείωση και Απλοποίηση των Χαρακτηριστικών.....	36
Κεφάλαιο 3: Μέθοδοι Συσταδοποίησης.....	39
3.1 Αλγόριθμοι Συσταδοποίησης.....	39
3.2 Μέτρα Επιλογής Αριθμού Συστάδων και Αξιολόγησης Συσταδοποίησης.....	42
3.3 Μέτρα Ομοιότητας Συστάδων.....	45
Κεφάλαιο 4: Προτεινόμενη Μεθοδολογία.....	46
4.1 Κανονικοποίηση Δεδομένων.....	46
4.2 Μείωση Διαστάσεων.....	47
4.3 Μέτρα Αξιολόγησης Χαρτοφυλακίων.....	50
4.4 Μεθοδολογία Εύρεσης Βέλτιστων Στρατηγικών για το Χτίσιμο Χαρτοφυλακίων.....	50
Κεφάλαιο 5 : Μελέτη Περίπτωσης.....	53
5.1 Παρουσίαση Δεδομένων.....	53
5.2 Επιλογή αριθμού συστάδων.....	56
5.3 Εφαρμογή Αλγορίθμου Συσταδοποίησης k-means και PCA.....	57
5.4 Εφαρμογή Αλγορίθμου Ιεραρχικής Συσταδοποίησης.....	60
5.5 Εφαρμογή Αλγορίθμου DBSCAN.....	65
5.6 Έλεγχος Ομοιότητας των Συστάδων.....	68
5.7 Χαρακτηρισμός και Ανάλυση των Συστάδων.....	69
5.8 Κατασκευή των Χαρτοφυλακίων.....	71
5.8.1 Αφετηρία με μια συστάδα.....	71
5.8.2 Αφετηρία με δύο συστάδες.....	81
Κεφάλαιο 6 : Συμπεράσματα και Μελλοντικές Επεκτάσεις.....	93
6.1 Συμπεράσματα.....	93
6.2 Μελλοντικές Επεκτάσεις.....	94
Βιβλιογραφία.....	97

Ευρετήριο Πινάκων

Πίνακας 1: Υψηλά συσχετιζόμενα στοιχεία πριν την απλοποίηση.....	38
Πίνακας 2: Διακύμανση PCA που καλύπτεται.....	59
Πίνακας 3: Αυτόματη Επιλογή του eps δεδομένου ενός min_samples.....	65
Πίνακας 4 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 0.....	71
Πίνακας 5 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 1.....	72
Πίνακας 6 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 2.....	72
Πίνακας 7 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 3.....	73
Πίνακας 8 : Αποτελέσματα για όλα τα πιθανά χαρτοφυλάκια, για την περίπτωση προσθήκης 4 πάρκων.....	76
Πίνακας 9 : Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 0 - 1.....	81
Πίνακας 10: Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 0 - 2.....	82
Πίνακας 11: Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 0 - 3.....	82
Πίνακας 12: Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 1 - 2.....	82
Πίνακας 13 : Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 1 - 3.....	83
Πίνακας 14 : Αποτελέσματα για αρχικό χαρτοφυλάκιο 2 - 3.....	83
Πίνακας 15 : Αποτελέσματα για όλους τους συνδυασμούς, για προσθήκη 4 πάρκων.....	87

Ευρετήριο Εικόνων

Εικόνα 1: Ζήτηση ηλεκτρισμού που καλύπτεται από κάθε πηγή ενέργειας μετρημένη σε GWh.....	18
Εικόνα 2: Στατιστικά χωρητικότητας αιολικής ενέργειας ανά τα έτη.....	19
Εικόνα 3: Προβλεπόμενες τιμές ηλεκτρισμού στις χώρες της ΕΕ, για τα τελευταία 5 έτη.....	19
Εικόνα 4: Ημερήσια απόκλιση των προβλεπόμενων τιμών ηλεκτρισμού στις χώρες της ΕΕ.....	20
Εικόνα 5: Μοντέλο λειτουργίας ενεργειακής αγοράς.....	22
Εικόνα 6: Διάγραμμα μεταβολής τιμών αποκλίσεων για μια τυχαία εβδομάδα.....	23
Εικόνα 7: Αρνητική και θετική λοξότητα, αντίστοιχα.....	31
Εικόνα 8: Ενδεικτικές μέσες τιμές ανέμου, υπολογισμένες για κάθε πάρκο ξεχωριστά.....	37
Εικόνα 9: Παράδειγμα k-distance plot για εύρεση του κατάλληλου ϵ κατά την εφαρμογή DBSCAN.....	40
Εικόνα 10: Τυχαίο παράδειγμα διαγράμματος για εξαγωγή του βέλτιστου k	43
Εικόνα 11: Παράδειγμα διαγράμματος για υπολογισμό του silhouette score.....	44
Εικόνα 12: Παράδειγμα μετασχηματισμού PCA.....	49
Εικόνα 13: Γεωγραφική τοποθεσία των πάρκων.....	53
Εικόνα 14: Μέσες τιμές ταχύτητας για όλα τα πάρκα.....	55
Εικόνα 15: Elbow Method για επιλογή του βέλτιστου αριθμού συστάδων k	56
Εικόνα 16: Δείκτες Silhouette και Davies-Bouldin για το σύνολο δεδομένων.....	56
Εικόνα 17: Δισδιάστατη αναπαράσταση των πάρκων, με κάθε συστάδα να χαρακτηρίζεται από διαφορετικό χρώμα.....	57
Εικόνα 18: Τρισδιάστατη απεικόνιση των πάρκων.....	58
Εικόνα 19: Πίνακας Βαρών των βασικών συνιστωσών.....	59
Εικόνα 20: Κέντρα των συστάδων στις αρχικές τιμές τους.....	60
Εικόνα 21: Δενδρόγραμμα ομάδων σύμφωνα με την ιεραρχική συσταδοποίηση.....	62
Εικόνα 22: Δείκτης σιλουέτας για διάφορα k κατά την ιεραρχική συσταδοποίηση.....	63
Εικόνα 23: Δείκτης Davies-Bouldin για διάφορα k κατά την ιεραρχική συσταδοποίηση.....	63
Εικόνα 24: Δισδιάστατη αναπαράσταση των πάρκων ιεραρχικής συσταδοποίησης.....	64
Εικόνα 25: Τρισδιάστατη αναπαράσταση των πάρκων.....	64
Εικόνα 26: Αρχικές τιμές κέντρων των συστάδων για ιεραρχική συσταδοποίηση.....	65
Εικόνα 27: Διάγραμμα βοήθειας επιλογής παραμέτρου ϵ	66
Εικόνα 28: Δισδιάστατο διάγραμμα πάρκων κατόπιν DBSCAN.....	66
Εικόνα 29: Τρισδιάστατο διάγραμμα πάρκων κατόπιν DBSCAN.....	67
Εικόνα 30: Συνημιτονοειδής Ομοιότητα για ιεραρχική συσταδοποίηση.....	68
Εικόνα 31: Ευκλείδεια Απόσταση κατά ιεραρχική συσταδοποίηση.....	68
Εικόνα 32: Radar Plot των κέντρων των συστάδων.....	69
Εικόνα 33: Τελικό radar plot των προφίλ των συστάδων.....	70
Εικόνα 34: Μείωση CoV για όλα τα σενάρια αρχικών χαρτοφυλακίων, συναρτήσει του αριθμού προστιθέμενων πάρκων.....	75
Εικόνα 35: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 3.....	77
Εικόνα 36: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 3.....	78
Εικόνα 37: Κατανομή τιμών έπειτα από προσθήκη 1 πάρκου.....	78
Εικόνα 38: Κατανομή τιμών έπειτα από προσθήκη 4 πάρκων.....	79
Εικόνα 39: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 3.....	79
Εικόνα 40: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 3.....	80
Εικόνα 41: Κατανομή τιμών έπειτα από προσθήκη 1 πάρκου.....	80
Εικόνα 42: Κατανομή τιμών έπειτα από προσθήκη 4 πάρκων.....	81
Εικόνα 43: Μείωση CoV για όλα τα σενάρια αρχικών χαρτοφυλακίων, συναρτήσει του αριθμού προστιθέμενων πάρκων.....	84
Εικόνα 44: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 2.....	89
Εικόνα 45: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 2.....	89
Εικόνα 46: Κατανομή τιμών έπειτα από προσθήκη 1 πάρκου.....	90

Εικόνα 47: Κατανομή τιμών έπειτα από προσθήκη 4 πάρκων.....	90
Εικόνα 48: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 3.....	91
Εικόνα 49: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 3.....	91
Εικόνα 50: Κατανομή τιμών για προσθήκη 1 πάρκου της συστάδας 3.....	92
Εικόνα 51: Κατανομή τιμών για προσθήκη 4 πάρκων της συστάδας 3.....	92

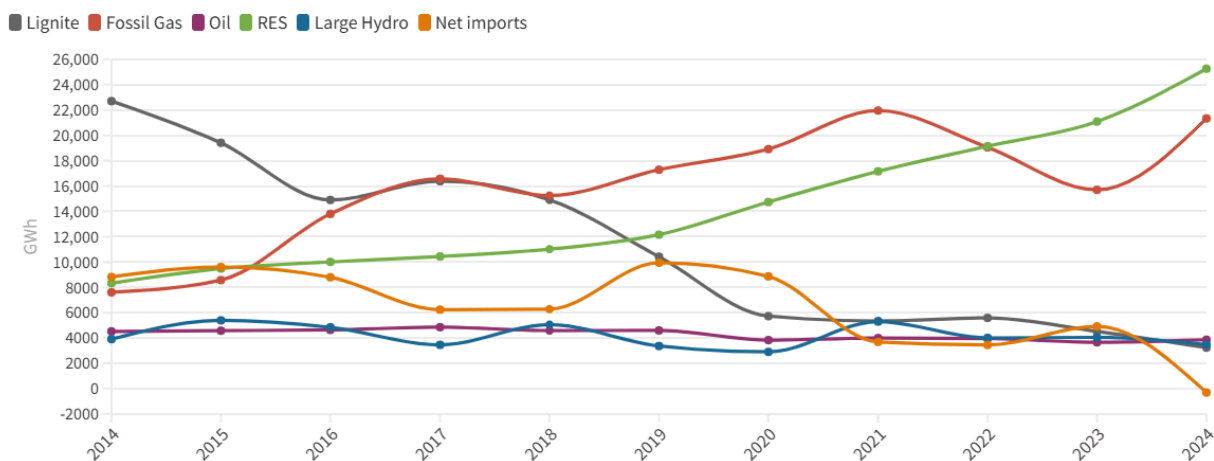
Κεφάλαιο 1 : Εισαγωγή

1.1 Ανανεώσιμες Πηγές Ενέργειας και Ενεργειακή Μετάβαση

Η μετάβαση προς ένα πιο βιώσιμο και καθαρό ενεργειακό σύστημα αποτελεί βασική προτεραιότητα της σύγχρονης κοινωνίας σε παγκόσμιο επίπεδο. Οι ανανεώσιμες πηγές ενέργειας (ΑΠΕ), όπως η ηλιακή και η αιολική ενέργεια, παίζουν καθοριστικό ρόλο σε αυτή τη μετάβαση, αντικαθιστώντας σταδιακά τα ορυκτά καύσιμα. Οι λόγοι που οδηγούν σε αυτή την αλλαγή είναι κυρίως περιβαλλοντικοί. Η σύγχρονη κοινωνία και βιομηχανία καταναλώνει ορυκτά καύσιμα σε έναν ακραίο και πρωτοφανή βαθμό, και ο αντίκτυπός της κατανάλωσης αυτής στο περιβάλλον είναι τεράστιος. Πολλοί ειδικοί θεωρούν πως η στιγμή για την πολυαναμενόμενη και οριστική αλλαγή από τα ορυκτά καύσιμα στις “πράσινες” ενέργειες είναι τώρα, όντας πεπεισμένοι πως οι εκπομπές διοξειδίου του άνθρακα είναι σχεδόν ανεξέλεγκτες πλέον και αποτελούν ένα φλέγον θέμα που δε μπορεί να περιμένει. [1]

Η αυξανόμενη επομένως προσπάθεια διείσδυσης των ΑΠΕ στο ενεργειακό μείγμα των χωρών της Ευρωπαϊκής Ένωσης, και ειδικότερα της Ελλάδας, αποτελεί έναν ιδιαίτερα στρατηγικό στόχο, διότι το πρόβλημα των αυξημένων εκπομπών αερίων του θερμοκηπίου, σε συνδυασμό με την ιδιαίτερη επιβάρυνση της ατμόσφαιρας του πλανήτη, καθιστά τη στροφή προς τις “πράσινες” πηγές ενέργειας απαραίτητη. Τον Ιούλιο του 2025, μάλιστα, η επιτροπή της Ευρωπαϊκής Ένωσης έθεσε ως στόχο τη μείωση κατά το εντυπωσιακό 90% των εκπομπών αερίων του θερμοκηπίου, δεδομένου ότι ο στόχος της Ένωσης για μείωση κατά 55% των εκπομπών μέχρι το έτος 2030, σύμφωνα με όλες τις προβλέψεις, θα επιτευχθεί. [2]

Στην Ελλάδα, η πράσινη ενέργεια έχει αρχίσει ήδη να εγκαθιδρύεται στο μείγμα του ηλεκτρισμού. Πιο συγκεκριμένα, η πράσινη ενέργεια το 2024 (ΑΠΕ + μεγάλα υδροηλεκτρικά πάρκα) κάλυψε 50,5% της ζήτησης· μόνο τα αιολικά και φωτοβολταϊκά κάλυψαν 44,4%. Το φυσικό αέριο έφτασε το 37,5%, το πετρέλαιο 6,8% και ο λιγνίτης 5,7%. [3]

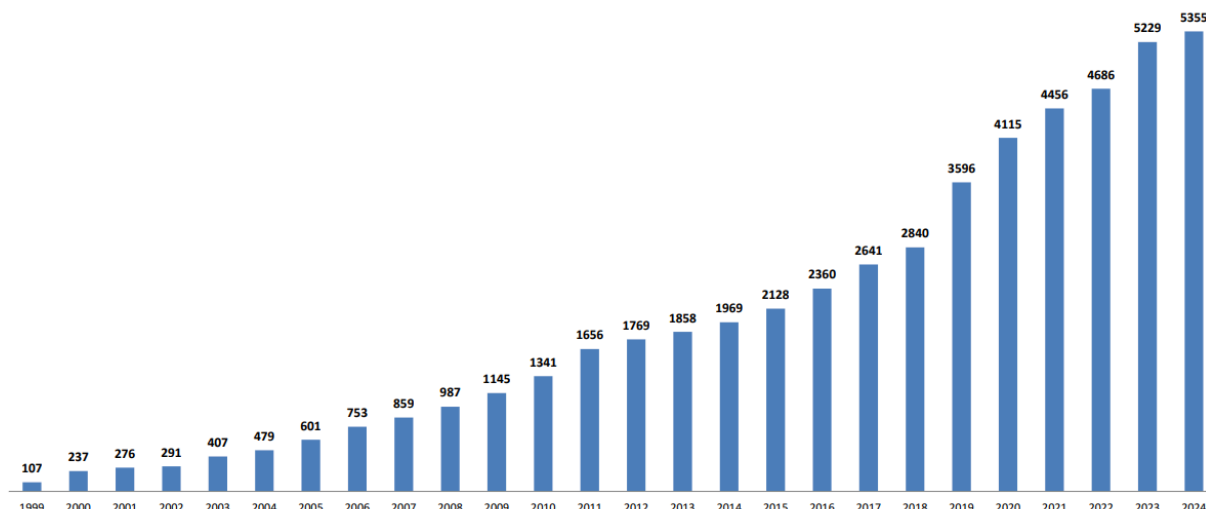


Εικόνα 1: Ζήτηση ηλεκτρισμού που καλύπτεται από κάθε πηγή ενέργειας, μετρημένη σε GWh. Πηγή: The Green Tank

Εστιάζοντας στην αιολική ενέργεια, σύμφωνα με έρευνα της Ελληνικής Επιστημονικής Ένωσης Αιολικής Ενέργειας (ΕΛΕΤΑΕΝ) που διενεργήθηκε στο τέλος του 2023, η παραγόμενη αιολική ισχύς βρίσκεται σε διαρκή και σταθερή αύξηση, και η ίδια έρευνα καταδεικνύει πως η χωρητικότητα του αιολικού δικτύου βρίσκεται αυτή τη στιγμή στα 5,355 GW, καταδεικνύοντας ραγδαίες αυξήσεις σε σχέση με τα προηγούμενα έτη και επιβεβαιώνοντας τις εκτιμήσεις της ΕΕ για την θέση που κατέχει η αιολική ενέργεια στο ενεργειακό μίγμα της χώρας.

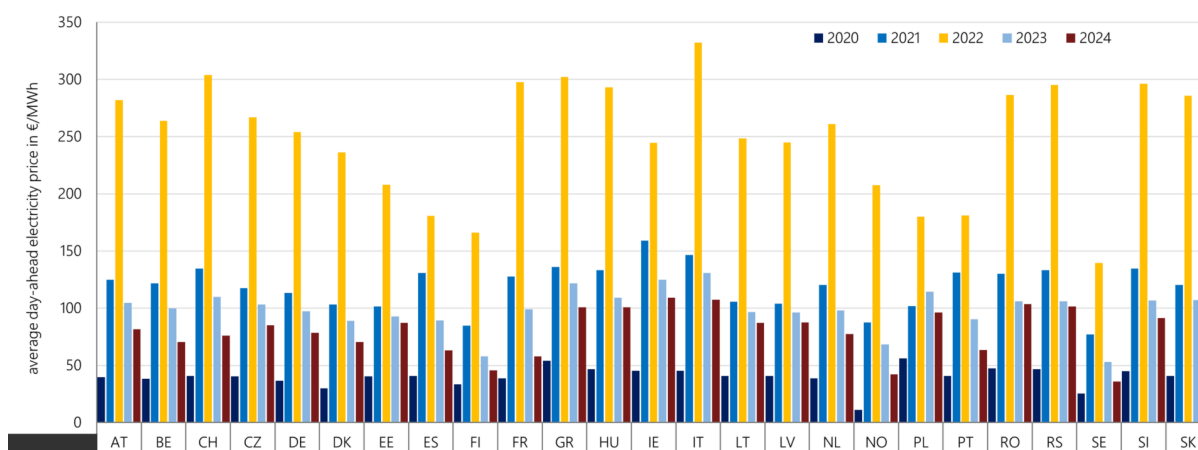


Total capacity to the grid (MW) per year



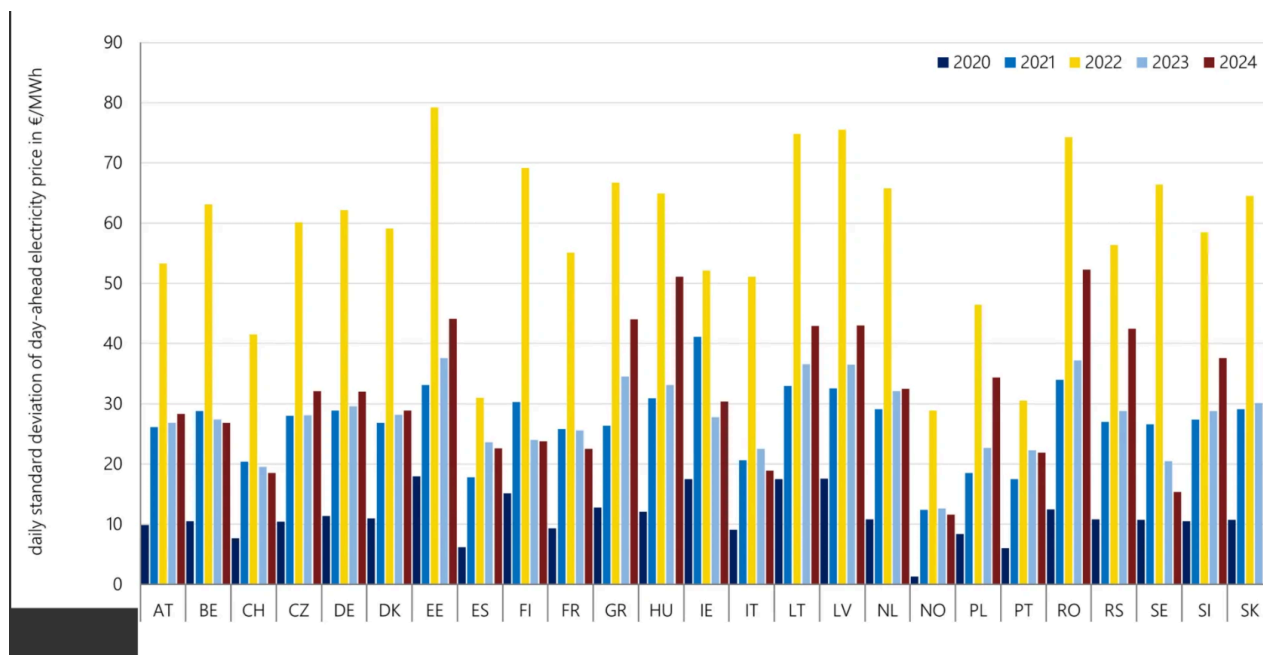
Εικόνα 2: Στατιστικά χωρητικότητας αιολικής ενέργειας ανά τα έτη. Πηγή: ΕΛΕΤΑΕΝ

Αυτή ακριβώς η αυξανόμενη εδραίωση των ΑΠΕ και της πράσινης ενέργειας δεν μπορεί να αφήσει ανεπηρέαστες τις τιμές του ηλεκτρισμού. Όταν παρατηρείται αύξηση της παραγωγής αιολικής και ηλιακής ενέργειας, εκτοπίζεται ακριβότερη θερμική παραγωγή και η χονδρική τιμή του ηλεκτρισμού, κατά συνέπεια, μειώνεται. Μάλιστα για πρώτη φορά μετά από πολλά χρόνια, παρατηρήθηκε πανευρωπαϊκά μείωση των τιμών του ηλεκτρισμού, αναδεικνύοντας μια επερχόμενη περίοδο σταθερότητας.



Εικόνα 3: Προβλεπόμενες τιμές ηλεκτρισμού στις χώρες της ΕΕ, για τα τελευταία 5 έτη. Πηγή: FfE

Παρόλα αυτά, οι αποκλίσεις των προβλεπόμενων τιμών έχουν ανάμεικτα αποτελέσματα ανά τις χώρες, δείχνοντας είτε μείωση είτε αύξηση σε σχέση με το προηγούμενο ακριβώς έτος. Χώρες όπως η Νορβηγία και η Σουηδία είχαν μειωμένη μεταβλητότητα, σε αντίθεση με την Ρουμανία και τη Σερβία. Στην Ελλάδα, η απόκλιση ναι μεν μειώθηκε συγκριτικά με το 2023, αλλά ελάχιστα, και εξακολουθεί να κατέχει έναν από τους υψηλότερους δείκτες μεταβλητότητας σε ολόκληρη την Ευρώπη. Αυτό υπογραμμίζει το ζήτημα των μη βέλτιστων προβλέψεων, που οφείλει να επιλυθεί ή να περιοριστεί, για να αποφευχθούν οι οικονομικές ποινές στους παραγωγούς ενέργειας. [4]



Εικόνα 4: Ημερήσια απόκλιση των προβλεπόμενων τιμών ηλεκτρισμού στις χώρες της ΕΕ. Πηγή: FfE

1.2 Το Ενεργειακό Χρηματιστήριο στην Ελλάδα

Η αιολική και η ηλιακή ενέργεια χαρακτηρίζονται, εκ φύσεως, από υψηλό βαθμό στοχαστικότητας και τυχαιότητας. Η παραγόμενη ισχύς μεταβάλλεται δυναμικά και με διαφορετικούς τρόπους κατά τη διάρκεια της ημέρας και από περιοχή σε περιοχή, εξαρτώμενη από φυσικά φαινόμενα όπως η ταχύτητα του ανέμου και η ηλιακή ακτινοβολία. Αυτή η μεταβλητότητα προσδίδει έναν παράγοντα αβεβαιότητας στις προβλέψεις και δυσκολεύει την αξιόπιστη συμβολή των ΑΠΕ στην αγορά του ηλεκτρισμού.

Η ακριβής πρόβλεψη της παραγωγής των ΑΠΕ αποτελεί, επομένως, κρίσιμο εργαλείο για τους διαχειριστές δικτύου αλλά και για τους εμπόρους που χρησιμοποιούν την αγορά ενέργειας. Οι προβλέψεις αυτές, κυρίως σε ημερήσια βάση, χρησιμοποιούνται, εκτός από τον προγραμματισμό της παραγωγής, και για τη συμμετοχή στη χονδρεμπορική αγορά ηλεκτρικής ενέργειας. Σφάλματα στις προβλέψεις οδηγούν σε αποκλίσεις (imbalances) μεταξύ δηλωμένης και πραγματικής παραγωγής, οι οποίες συνεπάγονται οικονομικές ρήτρες για τους παραγωγούς.

Ένας τύπος για τις αποκλίσεις είναι ο $\Delta E(t) = E_{\text{μετρούμενη}}(t) - \text{Final Position}(t)$,

όπου $E_{\text{μετρούμενη}}$ η ενέργεια που πράγματι παράχθηκε από έναν έμπορο, και Τελική Θέση (Final Position) η ενέργεια αυτή που δεσμεύεται ότι θα παραδώσει ένας έμπορος στην αγορά. Αν η ενέργεια που έχει δεσμευτεί - προβλέψει ο έμπορος ότι θα παραχθεί είναι περισσότερη από αυτή που πράγματι παρήγαγε, τότε η απόκλιση $\Delta E(t)$ θα είναι αρνητική, κάτι που οδηγεί σε έλλειμμα παραγωγής και ποινές για τον έμπορο.

Η αύξηση της παραγωγής ΑΠΕ και η όλο και υψηλότερη σημασία που έχουν αυτές στον σύγχρονο κόσμο έχει καταστήσει την ενεργειακή αγορά ιδιαίτερα ανταγωνιστική. Παράλληλα, είναι αναγκαίο οι έμποροι που εισέρχονται στην αγορά και κάνουν εκτίμηση της παραγωγής που μπορούν να προσφέρουν να είναι υπεύθυνοι των προβλέψεων αυτών και να τους επιβάλλονται οικονομικές χρεώσεις, στην περίπτωση που αστοχήσουν. Όλα αυτά τα γεγονότα καθιστούν απαραίτητη την ίδρυση αρμόδιων φορέων και επιχειρήσεων, ώστε να εξασφαλιστεί η σωστή λειτουργία της αγοράς, η τήρηση συγκεκριμένων προδιαγραφών και η αποφυγή αυθαιρεσιών από την πλευρά των εμπόρων.

Ακολουθούν οι σημαντικότεροι Φορείς και Αρχές της ελληνικής ενεργειακής αγοράς:

Ρυθμιστική Αρχή Αποβλήτων, Ενέργειας και Υδάτων (ΡΑΑΕΥ, πρώην Ρυθμιστική Αρχή Ενέργειας - ΡΑΕ):

Η ΡΑΑΕΥ ιδρύθηκε ως ανεξάρτητη ρυθμιστική αρχή για την αγορά ηλεκτρικής ενέργειας και, αργότερα κατόπιν εμπλουτισμού των Νόμων, για τον ηλεκτρισμό και το φυσικό αέριο. Οι αρμοδιότητές της είναι κυρίως θεσμοφορικού και εισηγητικού τύπου, με τις βασικότερες από τις οποίες να είναι:

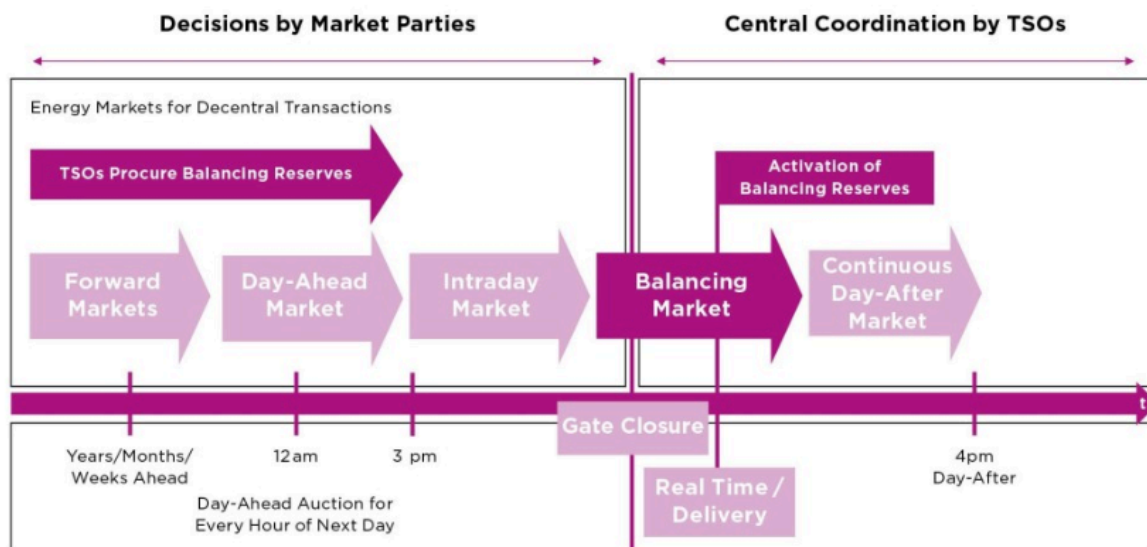
- η παρακολούθηση και η εποπτεία της αγοράς ενέργειας, η μελέτη, σύνταξη και δημοσίευση εκθέσεων με σκοπό την εύρεση ενδεχόμενων αναθεωρήσεων των κανονισμών. Ταυτόχρονα αποφασίζει ή εισηγείται στα αρμόδια όργανα τη λήψη αναγκαίων μέτρων, ιδίως για την τήρηση των κανόνων του ανταγωνισμού και των ρυθμιστικών υποχρεώσεων.
- η προστασία των καταναλωτών, όσον αφορά καταγγελίες υποβληθείσες εναντίον ενεργειακών επιχειρήσεων
- η χορήγηση και, αν καταστεί αναγκαίο, η ανάκληση αδειών για άσκηση ενεργειακών δραστηριοτήτων, σε περιπτώσεις μη τήρησης των προδιαγραφών
- η παρακολούθηση της ασφάλειας του ενεργειακού ανεφοδιασμού της Ελλάδας.
- η παροχή αντικειμενικής πληροφόρησης σε όλους τους ενδιαφερόμενους μέσω στατιστικών στοιχείων, προβλέψεων παραγωγής και κατανάλωσης, στα οποία μπορεί ο καθένας να έχει δωρεάν πρόσβαση. Η πρακτική αυτή αποσκοπεί στην ενίσχυση του ελεύθερου ανταγωνισμού και στην καταπολέμηση ενδεχομένου μονοπωλίου από ιδιοτελείς φορείς ή εμπόρους.

Ανεξάρτητος Διαχειριστής Μεταφοράς Ηλεκτρικής Ενέργειας (ΑΔΜΗΕ):

Ο ΑΔΜΗΕ ασκεί τις αρμοδιότητες του και εκτελεί τα καθήκοντα του ως διαχειριστής του Ελληνικού Συστήματος Μεταφοράς Ηλεκτρικής Ενέργειας (ΕΣΜΗΕ). Τα καθήκοντα του φορέα είναι, ανάμεσα σε άλλα:

- η λειτουργία, ο έλεγχος, η συντήρηση και η ανάπτυξη του ΕΣΜΗΕ, ώστε να διασφαλίζεται η ασφαλής και αξιόπιστη ηλεκτροδότηση της χώρας,
- η εξασφάλιση ισοδύναμης πρόσβασης όλων των χρηστών στο Σύστημα και η εγγύηση ύπαρξης διαφάνειας κατά την παροχή πληροφοριών προς τους φορείς της αγοράς, με στόχο την τελική ενίσχυση του υγιούς και αμερόληπτου ανταγωνισμού,
- η δημοσίευση του καταλόγου των τιμολογίων, βάσει των οποίων οι χρήστες του Συστήματος χρεώνονται,
- ο υπολογισμός της Οριακής Τιμής Αποκλίσεων,
- η βέλτιστη αξιοποίηση των διαθέσιμων πόρων της Ελλάδας, προς το δημόσιο όφελος της χώρας,
- η εκκαθάριση των Αποκλίσεων Παραγωγής και Ζήτησης, ο υπολογισμός των χρεοπιστώσεων και η διευθέτηση των χρηματικών συναλλαγών στο πλαίσιο της διευθέτησης των Αποκλίσεων Παραγωγής και Ζήτησης. [5]

Η ελληνική χονδρεμπορική αγορά ενέργειας λειτουργεί αυτή τη στιγμή σύμφωνα με το εξής μοντέλο:



Εικόνα 5: Μοντέλο λειτουργίας ενεργειακής αγοράς. Πηγή: Amprion

Τα σημαντικότερα σημεία της αγοράς είναι:

α) Η Αγορά Επόμενης Ημέρας (Day-Ahead Market, DAM): Σε αυτή τη φάση υποβάλλονται στο χρηματιστήριο οι προβλεπόμενοι αριθμοί παραγωγής ενός δικτύου ενέργειας και ακολουθούνται από τις δράσεις των επενδυτών, αναλόγως τη ζήτηση που έχει ο καθένας.

β) Η Ενδοημερήσια Αγορά (Intra-day Market, IDM): Αυτό το στάδιο είναι απαραίτητο για την κάλυψη προσωρινών αναγκών σε ενέργεια που δεν είχαν προβλεφθεί εκ των προτέρων από τον αγοραστή, και μπορούν να καλυφθούν σε περιπτώσεις που η παραγωγή του δικτύου ενέργειας ξεπερνά τις προβλεπόμενες τιμές. Τυχόν επιπρόσθετες MWh που θα απαιτηθούν από τον αγοραστή, για παράδειγμα εξαιτίας διακοπής ρεύματος, μπορούν να βρεθούν εδώ. Το τωρινό μοντέλο προσφέρει τρεις ευκαιρίες εντός της ημέρας (IDM1, IDM2, IDM3) σε έναν επενδυτή, ώστε αυτός να καλύψει τις ανάγκες του.

Η HEnEx (Hellenic Energy Exchange) είναι η υπεύθυνη για τον καθορισμό των τιμών του IDM και του DAM, όπως ορίστηκε από τη ΠΑΑΕΥ - ΠΑΕ. Αντίστοιχα η αδελφική εταιρεία EnExClear είναι η υπεύθυνη για την εκκαθάριση και το διακανονισμό των συναλλαγών του DAM και του IDM, καθώς και για την εκκαθάριση των θέσεων της Αγοράς Εξισορρόπησης. Και οι δύο εταιρείες ανήκουν στον όμιλο EnExGroup, που είναι ο όμιλος του Ελληνικού Χρηματιστηρίου Ενέργειας.

Η ημερήσια διαπραγμάτευση ενεργειακών περιουσιακών στοιχείων με βάση τη ζήτηση συμβάλλει στην εκπλήρωση συμβολαίων υποχρεώσεων, στη μείωση του κόστους ανισορροπίας, στη μεγιστοποίηση των κερδών και στην ενίσχυση της σταθερότητας του δικτύου.

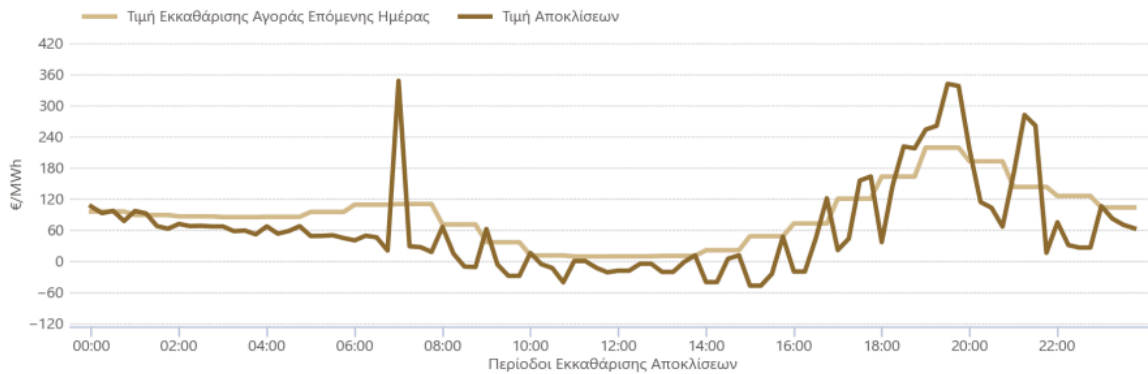
γ) Αγορά Εξισορρόπησης (Balancing Market), όπου εκκαθαρίζονται οι αποκλίσεις τόσο στις προβλέψεις, δηλαδή στην προσφορά, αλλά και στη ζήτηση. Σε αυτή τη φάση, αν υπάρχει συνολικά έλλειμμα στην αγορά για μια συγκεκριμένη χρονική στιγμή, η Αγορά Εξισορρόπησης μπορεί να επιστρατεύσει αποθέματα ενέργειας (energy reserves) για την αντιστάθμιση της ελλειμματικής ποσότητας ενέργειας και την κάλυψη των αναγκών του δικτύου. Φορείς όπως ο ΑΔΜΗΕ εξασφαλίζουν ότι στο δίκτυο υπάρχει, συνολικά, επαρκής διαθέσιμη ισχύς για παροχή υπηρεσιών εξισορρόπησης σύμφωνα με τις εκτιμήσεις του. [5]

Στην Αγορά Εξισορρόπησης λαμβάνει χώρα η πολύ σημαντική Περίοδος Εκκαθάρισης Αποκλίσεων (Imbalance Settlement Period, ISP). Ως ISP ορίζεται η δεκαπεντάλεπτη χρονική περίοδος για την οποία υπολογίζεται η απόκλιση των φορέων από τις προβλεπόμενες τιμές παραγωγής τους. Εντός αυτού του δεκαπενταλεπτού χρονικού διαστήματος, αποτιμάται η Τιμή Ανισορροπίας (Imbalance Price). Σύμφωνα με τον ΑΔΜΗΕ, η Τιμή Ανισορροπίας είναι η τιμή (€/MWh) με την οποία εκκαθαρίζονται οι αποκλίσεις κάθε συμβαλλόμενου με ευθύνη ισοζυγίου, ανά 15λεπτο ISP. Ο μαθηματικός της τύπος ορίζεται ρητά στον Κανονισμό Αγοράς Εξισορρόπησης ως

«η τιμή που υπολογίζεται σύμφωνα με το Άρθρο 19.6, με την οποία χρεοπιστώνονται τα συμβαλλόμενα μέρη με ευθύνη ισοζυγίου». [6]

Η τιμή ανισορροπίας εξαρτάται από πολλούς παράγοντες και αλλάζει διαρκώς τιμή, όπως φαίνεται στο επόμενο διάγραμμα, υπολογισμένο για την πρώτη εβδομάδα του Σεπτεμβρίου 2025. Οι τιμές είναι οι μαθηματικοί μέσοι για κάθε δεκαπεντάλεπτο της ημέρας, για τις 7 μέρες της εβδομάδας αυτής.

Ημερήσιο Προφίλ Μέσου Όρου 15-min Τιμών



Εικόνα 6: Διάγραμμα μεταβολής τιμών αποκλίσεων για μια τυχαία εβδομάδα. Πηγή: ΑΔΜΗΕ.

Η εκάστοτε τιμή, στην οποία βρίσκεται μια χρονική στιγμή η καμπύλη της Τιμής Ανισορροπίας, είναι κρίσιμη για την κοστολόγηση της ρήτρας, καθώς με βάση αυτή χρεώνεται εν τέλει ο παραγωγός.

Ο τύπος, από τον οποίο προκύπτει το τελικό κόστος της ρήτρας, είναι ο:

$$Cost_{imb} = \Delta E \cdot P_{imb} + \Lambda \Pi$$

όπου ΔE οι αποκλίσεις, P_{imb} η Τιμή Ανισορροπίας για μια συγκεκριμένη χρονική στιγμή και $\Lambda \Pi$ οι Λογαριασμοί Προσαυξήσεων (Uplifts).

Οι Λογαριασμοί Προσαυξήσεων αποτελούνται από 3 επιμέρους λογαριασμούς:

- ΛΠ-1, ο Λογαριασμός Προσαυξήσεων Απωλειών του ΕΣΜΗΕ. Χρησιμοποιείται για τον επιμερισμό του κόστους των απωλειών ΕΣΜΗΕ, το οποίο υπολογίζεται από το άθροισμα των ποσών που προκύπτουν από την εκκαθάριση της Αγοράς Επόμενης Ημέρας, την εκκαθάριση της Ενδοημερήσιας Αγοράς και την Εκκαθάριση Αποκλίσεων για τις απώλειες αυτές. Πρακτικά ανακτά το κόστος των απωλειών στο δίκτυο μεταφοράς.
- ΛΠ-2, ο Λογαριασμός Προσαυξήσεων Ισχύος Εξισορρόπησης. Χρησιμοποιείται για τον επιμερισμό του κόστους παροχής Ισχύος Εξισορρόπησης από τους Παρόχους Υπηρεσιών Εξισορρόπησης. Η ισχύς εξισορρόπησης είναι η διαθεσιμότητα εφεδρειών σε MW που δεσμεύει ο ΑΔΜΗΕ από παρόχους ώστε, αν χρειαστεί, να ενεργοποιήσει γρήγορα ενέργεια εξισορρόπησης σε περιπτώσεις που αυτό καταστεί απαραίτητο. Στην πράξη, είναι ένα κόστος που πληρώνει ο έμπορος ώστε να μπορεί ανά πάσα στιγμή που αυτός επιθυμεί να δράσει στην ηλεκτρική αγορά, το σύστημα στο οποίο βασίζεται να είναι διαθέσιμο (stand-by).
- ΛΠ-3, ο Λογαριασμός Προσαυξήσεων Οικονομικής Ουδετερότητας. Η χρηματοοικονομική ουδετερότητα (Financial Neutrality) είναι ο μηχανισμός με τον οποίο ο ΑΔΜΗΕ εξουδετερώνει τα οικονομικά της Αγοράς Εξισορρόπησης που εκκρεμούν έτσι ώστε η αγορά να μένει ουδέτερη (ούτε έλλειμμα ούτε πλεόνασμα). Πρακτικά, οποιοδήποτε υπόλοιπο που απομένει μετά τις χρεώσεις / πιστώσεις από την εκκαθάριση κατανέμεται στα Συμβαλλόμενα Μέρη με Ευθύνη Ισοζυγίου μέσω του ΛΠ-3 ως προσαύξηση σε €/MWh. Αν το υπόλοιπο είναι θετικό, ο ΛΠ-3 θα είναι χρέωση, ενώ αν το υπόλοιπο είναι αρνητικό τότε ο ΛΠ-3 θα είναι πίστωση. [6]

Το κόστος της ρήτηρας, δεδομένου ότι εξαρτάται από την Τιμή Ανισορροπίας που είναι μεταβαλλόμενη μεταβλητή στο χρόνο και μάλιστα κάθε 15 λεπτά, υπολογίζεται επίσης διαφορετικά και θα είναι και διαφορετικό για κάθε συνεχόμενο δεκαπεντάλεπτο. Στο τέλος της ημέρας, τα επιμέρους κόστη προστίθενται ώστε να προκύψει η τελική ημερήσια ρήτρα.

Η κοστολόγηση των αποκλίσεων, και συνεπώς των ρητρών, προκύπτει υπολογιστικά βάσει του Κανονισμού Αγοράς Εξισορρόπησης (KAE-Balancing Market Rulebook) και τις μεθοδολογίες που ισχύουν για την ελληνική αγορά.

Ο ΚΑΕ ορίζει τους κανόνες και τις διαδικασίες λειτουργίας της Αγοράς Εξισορρόπησης. Τον Κανονισμό και τις μεθοδολογίες τους εγκρίνει η ΡΑΑΕΥ, και ο ΑΔΜΗΕ τις εφαρμόζει όπως αυτοί έχουν θεσπιστεί, και συμμορφώνεται με αυτούς. Μεταξύ των αρμοδιοτήτων του ΚΑΕ συγκαταλέγονται επίσης, εκτός του καθορισμού των οικονομικών επιβαρύνσεων λόγω μη εκπληρωμένων δεσμεύσεων,:

- ορισμός των κανόνων και των προϋποθέσεων συμμετοχής στην Αγορά Εξισορρόπησης,
- εγγύηση της ορθής λειτουργίας της Αγοράς Εξισορρόπησης, της Αγοράς Επόμενης Ημέρας και της Ενδοημερήσιας Αγοράς, τυποποιώντας τη διασύνδεση των αγορών αυτών,
- ορισμός της διαδικασίας Εκκαθάρισης της Αγοράς Εξισορρόπησης,
- ορισμός της διαδικασίας ανταλλαγής πληροφοριών με τους υπόλοιπους σχετιζόμενους φορείς,
- φροντίδα για την κάλυψη των υποχρεώσεων παροχής στοιχείων και πληροφοριών και παρακολούθησης της Αγοράς Εξισορρόπησης,
- τυποποίηση των διαδικασιών προς προστασία των εμπορικά ευαίσθητων πληροφοριών.
- καθορισμός κυρώσεων, στην περίπτωση που ένας έμπορος παραμελεί τα καθήκοντά του ή δε συμμορφώνεται με τις εγκεκριμένες προδιαγραφές του Κανονισμού.

1.3 Έννοια και Σημασία του Ενεργειακού Χαρτοφυλακίου

Η έννοια του «χαρτοφυλακίου» (portfolio) έχει τις ρίζες της στον χρηματοοικονομικό τομέα και χρησιμοποιείται για να περιγράψει τη συλλογή επενδυτικών προϊόντων που κατέχει ένας επενδυτής (όπως μετοχές, ομόλογα, ακίνητα κ.λπ.). Ο στόχος του επενδυτή είναι η βέλτιστη ισορροπία μεταξύ απόδοσης και κινδύνου, μέσα από την κατάλληλη διαφοροποίηση των στοιχείων που απαρτίζουν το χαρτοφυλάκιο. Η βασική αρχή στην οποία στηρίζεται αυτή η στρατηγική είναι ότι ένας διαφοροποιημένος συνδυασμός επενδύσεων έχει μικρότερη συνολική μεταβλητότητα και μεγαλύτερη ανθεκτικότητα σε απρόβλεπτες διακυμάνσεις, συγκριτικά με μεμονωμένα και μη διαφοροποιημένα περιουσιακά στοιχεία.

Η ίδια αυτή λογική έχει μεταφερθεί και στον χώρο των ΑΠΕ, όπου ο όρος ενεργειακό χαρτοφυλάκιο χρησιμοποιείται για να περιγράψει το σύνολο των μονάδων παραγωγής ενέργειας που ανήκουν ή διαχειρίζονται από έναν φορέα. Αυτές οι μονάδες μπορεί να είναι αιολικά πάρκα, φωτοβολταϊκά συστήματα ή άλλες μορφές ΑΠΕ, κατανεμημένες γεωγραφικά σε διάφορες περιοχές. Το χαρτοφυλάκιο μπορεί να περιλαμβάνει επίσης και υβριδικά ή αποθηκευτικά συστήματα, εφόσον συνυπάρχουν στην ίδια επιχειρησιακή μονάδα.

Ειδικότερα για τους παραγωγούς που συμμετέχουν στην ενεργειακή αγορά, η σταθερότητα που μπορεί να προσφέρει ένα ισχυρό ενεργειακό χαρτοφυλάκιο αποκτά κρίσιμη σημασία, αφού οι αποκλίσεις μεταξύ δηλωμένης και πραγματικής παραγωγής επιφέρουν ρήτρες ή οικονομικές απώλειες. Μέσω της συγκρότησης και βελτιστοποίησης του χαρτοφυλακίου τους, οι παραγωγοί μπορούν να περιορίσουν την έκθεσή τους σε τέτοια σφάλματα και να βελτιώσουν την απόδοσή τους στις αγορές.

Ο πιο σημαντικός χρήστης - έμπορος της Ελλάδας είναι οι Φορείς Σωρευτικής Εκπροσώπησης (ΦοΣΕ). Αυτοί είναι φορείς που αναλαμβάνουν την εκπροσώπηση μεγάλου αριθμού μικρότερων παραγωγών - εμπόρων ΑΠΕ στην χονδρεμπορική αγορά ηλεκτρικής ενέργειας, όπως στις προαναφερθείσες DAM και IDM. Οι ΦοΣΕ διαχειρίζονται τη συμμετοχή τους στην αγορά, διασφαλίζοντας τη συμμόρφωση με τις εθνικές προδιαγραφές και την έμπιστη εμπορική εκπροσώπηση. Ταυτόχρονα,

- αναλαμβάνουν την ευθύνη της εξισορρόπησης και το κόστος των αποκλίσεων,
- βελτιστοποιούν την πρόβλεψη παραγωγής των μονάδων ΑΠΕ που εκπροσωπούν, αξιοποιώντας και προσαρμόζοντας διαφορετικές τεχνικές προβλέψεων, είτε με βάση γνωστές, είτε ενσωματώνοντας καινοτόμες, μεθοδολογίες,

- κατανέμουν την ενέργεια των μονάδων ΑΠΕ στις κατάλληλες επιμέρους αγορές, σύμφωνα με εκτιμήσεις που γίνονται από έμπειρους ειδήμονες.

Καθώς οι ΦοΣΕ έχουν ανάγκη και από νέες μεθοδολογίες που μπορούν να συνδυάσουν με βέλτιστο τρόπο περιουσιακά στοιχεία ΑΠΕ, με απώτερο στόχο την ελαχιστοποίηση των σφαλμάτων και τη μείωση με αυτό τον τρόπο των οικονομικών ποινών, το αντικείμενο της παρούσας διπλωματικής έχει ιδιαίτερο ενδιαφέρον για τους φορείς αυτούς. Η μεθοδολογία που θα παρουσιαστεί στοχεύει στην εξασφάλιση αποτελεσμάτων με όσο το δυνατόν μικρότερη αβεβαιότητα γύρω τους, το οποίο οδηγεί αναμφίβολα σε εγκυρότερες προβλέψεις και επομένως μείωση των ρητρών. Αναλογιζόμενοι την πολύ ευέλικτη Τιμή Ανισορροπίας, που μπορεί να δείξει μεγάλες αποκλίσεις εντός της ίδιας ημέρας, είναι ευνόητο πως ακόμα και μικρές βελτιώσεις στη μείωση της αβεβαιότητας μιας πρόβλεψης μπορούν συνολικά να αποβούν ευεργετικές για έναν φορέα, ο οποίος επενδύει μεγάλα ποσά και δεσμεύεται για αυτά καθημερινά.

1.4 Σύγχρονη Θεωρία Χαρτοφυλακίων

Η ανάγκη μείωσης των αποκλίσεων καθιστά αναγκαία την ανάπτυξη, αλλά και τη βελτιστοποίηση, μεθόδων που ενισχύουν την προβλεψιμότητα και περιορίζουν όσο το δυνατόν περισσότερο τα σφάλματα. Μία τέτοια προσέγγιση είναι η ανάλυση και η βελτιστοποίηση του ενεργειακού χαρτοφυλακίου ενός φορέα ΑΠΕ, εμπνευσμένη από τη Θεωρία Χαρτοφυλακίου του Markowitz (Portfolio Theory of Markowitz), όπου ο στόχος είναι η ελαχιστοποίηση της διακύμανσης (κίνδυνος) για δεδομένη προσδοκώμενη απόδοση (ενεργειακή παραγωγή), ή αντίστροφα η μεγιστοποίηση της απόδοσης για δεδομένα επίπεδα κινδύνου.

Ο ισχυρισμός του Markowitz βασίστηκε στο γεγονός πως οι περισσότερες επενδύσεις σε περιουσιακά στοιχεία επιφέρουν είτε μικρά κέρδη για μικρό ρίσκο, είτε μεγάλα κέρδη για μεγάλο ρίσκο. Σκοπός ήταν η εύρεση ενός ταιριαστού συνδυασμού περιουσιακών στοιχείων, έτσι ώστε να επιτευχθεί ένα ενιαίο χαρτοφυλάκιο που μειώνει τον κίνδυνο που θα επέφεραν τα πιο επισφαλή (risk-averse) περιουσιακά στοιχεία του, και ταυτόχρονα αυξάνει το ενδεχόμενο κέρδος του, επωφελούμενο από τα ασφαλή (risk-free) εκείνα στοιχεία. Με αφετηρία αυτό, ένας επενδυτής θέτει ως στόχο μια συγκεκριμένη επιθυμητή επιστροφή από την επένδυσή του, και μπορεί να χτίσει ένα χαρτοφυλάκιο που συμπεριλαμβάνει τα διαθέσιμα περιουσιακά στοιχεία του και τα συνδυάζει με κατάλληλο τρόπο, ώστε να επιτυγχάνεται αυτός ο στόχος. Φυσικά δεδομένου ότι τα περισσότερα περιουσιακά στοιχεία είναι ασταθή (volatile) στον χρόνο, η σωστή και ορθολογική τακτική είναι η διαρκής έρευνα και έλεγχος των επενδύσεων, ώστε να εντοπιστούν ήδη κατεχόμενες επενδύσεις που στα νέα δεδομένα του χαρτοφυλακίου περιέχουν μεγάλο ποσοστό κινδύνου ή να προστεθούν νέα περιουσιακά στοιχεία που φαίνεται να είναι ιδιαίτερα ευνοϊκά.

Εξίσου σημαντική είναι και η έννοια της διαφοροποίησης, καθώς σύμφωνα με τη θεωρία του Markowitz, οι επενδυτές οφείλουν να διανέμουν τα κεφάλαιά τους σε διαφορετικά προϊόντα - επενδύσεις, ώστε να επιτευχθεί μείωση της επισφάλειας του χαρτοφυλακίου τους.

Σε αυτό το σημείο, θα πρέπει να γίνει αναφορά σε ορισμένες χρήσιμες έννοιες της Θεωρίας Χαρτοφυλακίων.

Μια τέτοια βασική έννοια είναι η απόδοση (return), δηλαδή η αξία ενός περιουσιακού στοιχείου σε ένα συγκεκριμένο χρονικό σημείο συγκρινόμενη με μια παλαιότερη τιμή του ίδιου περιουσιακού στοιχείου. Εκφράζει την ποσοστιαία αλλαγή της αξίας και ορίζεται ως:

$$R_t = \frac{V_t - V_{t-1}}{V_{t-1}}, \text{ όπου}$$

V_t η τιμή του στοιχείου τη χρονική στιγμή t , και αντίστοιχα V_{t-1} η τιμή του στοιχείου την ακριβώς προηγούμενη στιγμή $t-1$.

Η μέση απόδοση (mean return) υπολογίζει την αναμενόμενη απόδοση ενός στοιχείου και ορίζεται ως:

$$E[R_t] = \sum_{t=1}^n p_t R_t, \text{ όπου}$$

p_t η πιθανότητα να υπάρξει επιστροφή R_t τη στιγμή t και R_t η επιστροφή.

Μπορεί επίσης να ποσοτικοποιηθεί ο κίνδυνος (risk) μιας επένδυσης με χρήση της διακύμανσης ή τυπικής απόκλισης (variance) της απόδοσής του με τον εξής τύπο:

$$Var(R_t) = \sigma_t^2 = \sum p_t (R_t - E[R_t])^2$$

Ως σ_t^2 ορίζεται η διακύμανση της απόδοσης του περιουσιακού στοιχείου, και η τυπική απόκλιση είναι απλά η τετραγωνική ρίζα της διακύμανσης:

$$\sigma_t = \sqrt{Var(R_t)}$$

Η συνδιακύμανση (covariance) μεταξύ δύο στοιχείων υπολογίζει το βαθμό που η απόδοση του ενός επηρεάζει την απόδοση του άλλου. Με άλλα λόγια, μπορεί να αναδείξει τη σχετική κίνηση των αποδόσεων μεταξύ τους. Η ποσότητα αυτή καταδεικνύει τον κίνδυνο που επιφέρει ο συνδυασμός των δύο αυτών περιουσιακών στοιχείων, και αν αυτός είναι αρκετά χαμηλός ώστε να προστεθούν μαζί στο ίδιο χαρτοφυλάκιο. Ο τύπος ορίζεται ως:

$$Cov[R_t, R_w] = \sigma_{tw} = \sum_{i=1}^n p_i (R_{ti} - E[R_t])(R_{wi} - E[R_w]), \text{ όπου}$$

σ_{tw} η συνδιακύμανση των αποδόσεων των στοιχείων και $E[R_t]$, $E[R_w]$ οι μέσες αποδόσεις των στοιχείων. Όσο μικρότερη είναι αυτή η ποσότητα, τόσο πιο ασφαλής είναι η προσθήκη τους στο ίδιο χαρτοφυλάκιο, διότι τα στοιχεία παρουσιάζουν μεγάλες διαφορές στη συμπεριφορά τους, και επομένως είναι κατάλληλες επιλογές για διαφοροποίηση.

Επίσης χρήσιμος είναι ο ορισμός της Αναμενόμενης Απόδοσης Χαρτοφυλακίου (mean portfolio return). Ο επενδυτής μπορεί να χωρίσει τον πλούτο που θα επενδύσει σε ποσοστά, το καθένα από τα οποία πολλαπλασιάζεται με το περιουσιακό στοιχείο στο οποίο επενδύονται. Ο μαθηματικός τύπος αυτής της διαδικασίας είναι ο εξής:

$$E[R_p] = \sum_{t=1}^n w_t E[R_t], \text{ όπου}$$

w_t τα ποσοστά επένδυσης του αρχικού πλούτου στα περιουσιακά στοιχεία t και $E[R_t]$ η μέση απόδοση του στοιχείου t .

Φυσική επέκταση των τελευταίων εννοιών της διακύμανσης και της αναμενόμενης απόδοσης χαρτοφυλακίου είναι η έννοια του κινδύνου του χαρτοφυλακίου, που εξάγεται από τον τύπο:

$$\sigma_p^2 = \sum_{t=1}^n \sum_{w=1}^n w_t w_w Cov[R_t, R_w]$$

Η τυπική απόκλιση του χαρτοφυλακίου, που είναι ουσιαστικά ο κίνδυνος του χαρτοφυλακίου, είναι η τετραγωνική ρίζα της παραπάνω ποσότητας:

$$\sigma_p = \sqrt{\sigma_p^2}$$

Η διακύμανση μπορεί να γραφεί και σε διανυσματική μορφή, για εκείνες τις περιπτώσεις που περιέχουν πολλά περιουσιακά στοιχεία:

$$\sigma_p^2 = w^T \Sigma w$$

Φυσικά ως w ορίζεται το διάνυσμα των ποσοστών που επενδύονται και Σ ορίζεται ο πίνακας των συνδιακυμάνσεων.

Για την επίτευξη ελαχιστοποίησης ρίσκου ενός χαρτοφυλακίου, αρκεί η ελαχιστοποίηση της προαναφερθείσας ποσότητας, δηλαδή

$$\min (w^T \Sigma w), \text{ με δεδομένο ότι } \sum_{t=1}^n w_t = 1 \text{ και } R^T w = \mu, \text{ λύνοντας ως προς } \mu.$$

$R^T w$ ορίζεται εδώ η αναμενόμενη επιστροφή του χαρτοφυλακίου.

Η λογική όμως της διαφοροποίησης στα χαρτοφυλάκια ΑΠΕ δεν περιορίζεται μόνο στη Θεωρία Χαρτοφυλακίων του Markowitz· μεταφράζεται επιχειρησιακά σε τρεις μορφές συμπληρωματικότητας (complementarity) που εξομαλύνουν την παραγωγή και τα έσοδα.

1. Χωρική συμπληρωματικότητα: αιολικά πάρκα σε διαφορετικά περιβάλλοντα (π.χ. ορεινά, νησιωτικά) ή ανεμολογικά καθεστώτα τις περισσότερες φορές δεν έχουν ταυτόχρονα διακυμάνσεις στην παραγωγή τους, άρα οι χαμηλές αποδόσεις σε μια περιοχή συχνά αντισταθμίζονται από υψηλότερες σε μια άλλη.
2. Τεχνολογική συμπληρωματικότητα: καθώς οι ανεμογεννήτριες μπορούν να κατασκευαστούν από διαφορετικές εταιρείες με διαφορετικές προδιαγραφές, η ικανότητα με την οποία μπορούν να παράγουν ενέργεια ενδέχεται να διαφέρει από πάρκο σε πάρκο, λόγω χαμηλότερων επιπέδων ισχύος.
3. Χρονική συμπληρωματικότητα: σταθμοί με διαφορετικές καμπύλες και ρυθμούς μεταβολής (ramps) αλληλοκαλύπτονται, μειώνοντας ακραίες μεταβολές που ενδέχεται να επιβαρύνουν το σύστημα και την αγορά. Επιπλέον, τα φωτοβολταϊκά τείνουν να παράγουν περισσότερη ενέργεια μεσημέρι και καλοκαίρι· οι ανεμογεννήτριες μπορούν να παράγουν ενέργεια και τη νύχτα και το καλοκαίρι, επομένως ο συνδυασμός τους κάνει το ημερήσιο/εποχικό προφίλ πιο πλήρες, με λιγότερα κενά παραγωγής.

Κρίσιμο είναι ότι στη σημερινή αγορά το κόστος προκύπτει όχι μόνο από τη μεταβλητότητα της παραγωγής, αλλά και από τη μεταβλητότητα του σφάλματος πρόβλεψης. Ένα «έξυπνα» διαφοροποιημένο χαρτοφυλάκιο μειώνει ταυτόχρονα τις αστοχίες πρόβλεψης (γιατί τα λάθη διαφορετικών μονάδων δεν συμπίπτουν και επομένως μπορούν να αλληλοαναιρεθούν) και άρα περιορίζει τις χρεώσεις αποκλίσεων στην εξισορρόπηση, και ελαχιστοποιεί το ρίσκο επένδυσης για κάποιον ελπιδοφόρο επενδυτή. [7]

1.5 Αντικείμενο και Σκοπός της Εργασίας

Στην σημερινή αγορά ενέργειας, οι προβλέψεις παίζουν πολύ σημαντικό ρόλο. Η εγκυρότητα τους είναι κρίσιμη και αποτελεί βασικό μέλημα του κάθε εμπόρου, καθώς όχι μόνο θα μπορέσουν δυνητικά να ελαττωθούν οι χρηματικές ποινές που επιβάλλονται σε περιπτώσεις προβλεπτικών αστοχιών, αλλά μέσω αυτής ακριβώς της εγκυρότητας θα μπορέσουν να αποκτήσουν πλεονέκτημα στην αγορά. Πράγματι, χάριν σε μια ικανοποιητική προβλεπτική ικανότητα, ένας φορέας θα θεωρηθεί έμπιστος, ενώ ένας απλός χρήστης της αγοράς μπορεί να γνωρίζει σε ποιους φορείς μπορεί με μεγαλύτερη ασφάλεια να επενδύσει τα χρήματά του.

Είναι λογικό όμως, πως η πραγματοποίηση έγκυρων εκτιμήσεων παραγωγής δεν είναι διόλου εύκολη. Η φύση των ταχυτήτων του ανέμου ως μέγεθος είναι περίπλοκη και η εξάρτησή τους από πολυάριθμους παράγοντες,

αρκετοί από τους οποίους μπορούν να επηρεάσουν σε σύντομο χρονικό διάστημα τα πραγματικά δεδομένα που γίνεται προσπάθεια να προβλεφθούν, κάνει το έργο των εξειδικευμένων ομάδων δύσβατο. Μερικά παραδείγματα φαινομένων που επιβαρύνουν και δυσχεραίνουν την ικανότητα πρόβλεψης είναι ξαφνικές καταιγίδες, απότομες αλλαγές κατεύθυνσης των ανέμων και μελέμια, αν το πάρκο βρίσκεται κοντά σε θαλάσσια περιοχή.

Με αυτά τα δεδομένα, δημιουργείται το εύλογο ερώτημα αν ένας επενδυτής μπορεί, χωρίς την χρήση προβλέψεων και την εξάρτηση από αυτές, να εκτιμήσει εκ των προτέρων ποιοί τύποι πάρκων ταιριάζουν ο ένας με τον άλλο κατάλληλα με βάση το αιολικό προφίλ του καθενός.

Ο κεντρικός σκοπός της παρούσας εργασίας είναι η ανάπτυξη και εφαρμογή μιας μεθοδολογίας, η οποία σε πρώτο στάδιο θα αναλύει αιολικά δεδομένα και θα εξορύσσει στατιστικά στοιχεία από αυτά, και στη συνέχεια θα συσταδοποιεί τα πάρκα με βάση το συνολικό αιολικό προφίλ τους. Τα επόμενα βήματα είναι η κατασκευή, με τη χρήση των συστάδων που θα προκύψουν, ενεργειακών χαρτοφυλακίων, τα οποία έχουν ως κύριο στόχο τη βέλτιστη διαφοροποίηση εντός τους, και κατ'επέκταση τη μείωση των σφαλμάτων στην πρόβλεψη παραγωγής, επομένως και των ρητρών. Η διαφοροποίηση επιτυγχάνεται μέσω της συνδυαστικής αξιολόγησης των προαναφερθέντων στατιστικών χαρακτηριστικών των χρονοσειρών για κάθε πάρκο, ώστε να εντοπιστούν εκείνα τα πάρκα με συμπληρωματικά προφίλ, ώστε να αλληλοαναιρούνται οι επιμέρους διακυμάνσεις στις συμπεριφορές τους. Η μεθοδολογία αυτή θα μπορέσει, με τη σειρά της, να αξιοποιηθεί από εμπόρους στην ενεργειακή αγορά, που επιθυμούν να επιλέξουν ένα σύνολο πάρκων, χωρίς αυτά να είναι συγκεκριμένα, για την ένταξή τους σε ένα ενεργειακό χαρτοφυλάκιο που θα μπορέσει να ανταπεξέλθει στους στόχους τους με όσο το δυνατόν μικρότερο ρίσκο.

Η ανάλυση που θα γίνει θα βασίζεται στους εξής άξονες:

- Επεξεργασία και συσταδοποίηση πάρκων, με σκοπό την εύρεση συστάδων πάρκων που συνδυάζονται ευνοϊκά μεταξύ τους
- Κατασκευή χαρτοφυλακίων βάσει συμπληρωματικότητας των συστάδων και των πάρκων τους
- Αριθμητική επιβεβαίωση των συνδυασμών από συστάδες εκείνων, που με μεγαλύτερη συνέπεια και λιγότερο ρίσκο επιφέρουν ελάχιστο κόστος αποκλίσεων.

Μέσω της παραπάνω ανάλυσης, η εργασία φιλοδοξεί να προσφέρει ένα εργαλείο υποστήριξης αποφάσεων για παραγωγούς, διαχειριστές και φορείς που εμπλέκονται στη λειτουργία και στρατηγική ανάπτυξη αιολικών πάρκων. Η προτεινόμενη μεθοδολογία είναι γενικεύσιμη και μπορεί να εφαρμοστεί και σε άλλα είδη ΑΠΕ ή σε μικτά χαρτοφυλάκια. Με δεδομένο επίσης ότι ο κλάδος της εκμετάλλευσης των ΑΠΕ είναι ιδιαίτερα νέος και σε φάση ανάπτυξης στη χώρα μας, εκτιμάται πως η παρούσα εργασία θα αποτελέσει μια αφετηρία για περαιτέρω μελέτες στο αντικείμενο από μια χρηματιστηριακή πλευρά, που δεν έχει ερευνηθεί σε ιδιαίτερα μεγάλο βάθος στην Ελλάδα.

Κεφάλαιο 2 : Χρήσιμα Χαρακτηριστικά

Η στατιστική ανάλυση μιας μεταβλητής ή ενός συνόλου δεδομένων είναι μια απαραίτητη διαδικασία σε όλους τους επιστημονικούς κλάδους, και ιδιαίτερα στον κλάδο των προβλέψεων και της αξιολόγησης του ρίσκου. Μέσω αυτής, μπορούν να εξορυχθούν σημαντικές πληροφορίες για τη συμπεριφορά μιας μεταβλητής στο πέρασμα του χρόνου, για τυχόν επαναλαμβανόμενα μοτίβα, καθώς και για την κατανομή των τιμών της μεταβλητής. Η σωστή ανάλυση και ερμηνεία των πληροφοριών αυτών, αλλά και η ορθή χρήση τους αναλόγως τον στόχο που έχει η εκάστοτε έρευνα ή εργασία, συμβάλλει σε μεγάλο βαθμό στην εξαγωγή ορθών επιστημονικών συμπερασμάτων και στην επίλυση μιας ευρείας ποικιλίας προβλημάτων, που μοιάζουν δυσεπίλυτα.

Ένας πολύ συχνός τύπος δεδομένων είναι η χρονοσειρά (time series), οι οποίες αποτελούν ένα σύνολο διαχρονικών παρατηρήσεων μιας μεταβλητής. Συνήθως οπτικοποιούνται μέσω Διαγραμμάτων Χρόνου (time plots), τα οποία απλά απεικονίζουν τις τιμές της μεταβλητής για κάθε ένα προκαθορισμένο χρονικό σημείο (π.χ. οι τιμές της ταχύτητας των ανέμων συνηθίζεται να μετρώνται κάθε μία ώρα, ενώ οι απογραφές πληθυσμών συνηθίζεται να μετρώνται κάθε χρόνο). Μπορούν επίσης να οπτικοποιηθούν και με Διαγράμματα Εποχών (seasonal plots), αν στόχος είναι η ανάδειξη της επικράτησης ή μειωμένης επιρροής μιας συγκεκριμένης εποχής.

Μια χρονοσειρά μπορεί να θεωρηθεί ως η επαλληλία 4 χαρακτηριστικών, τα οποία είναι:

- η τάση (trend), η οποία είναι η μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς. Μπορεί να είναι είτε ανοδική είτε πτωτική, αλλά μπορεί ακόμα και να μην υπάρχει (μηδενική). Ανοδική τάση παρατηρείται π.χ. στις τιμές πώλησης σπιτιών στην Αμερική, που λόγω αυξημένης ακρίβειας και πληθωρισμού, οδηγούν την ενοικιαστική αγορά στην αύξηση των τιμών,
- η εποχικότητα (seasonality), περιοδική διακύμανση που παρατηρείται ανά σταθερά χρονικά διαστήματα. Συνήθίζεται να έχει μήκος ένα έτος, και χαρακτηριστικά παραδείγματα είναι ο αριθμός εγγραφών σε ένα γυμναστήριο τους πρώτους μήνες του χρόνου ή η αύξηση των αριθμών επιβατών πτήσεων τις καλοκαιρινές περιόδους,
- η κυκλικότητα (cyclical), επαναλαμβανόμενη μεταβολή που παρατηρείται έπειτα από συγκεκριμένες περιόδους, αλλά όχι ανά σταθερά χρονικά διαστήματα. Ένα χαρακτηριστικό παράδειγμα κυκλικότητας είναι η πτώση των τιμών μιας μετοχής έπειτα από μια περίοδο άνθησης. Η σημαντική διαφορά μεταξύ κυκλικότητας και εποχικότητας είναι η χρονική διάρκεια του φαινομένου, που επίσης δύναται να μην είναι γνωστή. Σε χρονοσειρές με έντονη εποχικότητα, τα διαστήματα μεταξύ κορύφωσης και πτώσης θα είναι σταθερά και θα μπορούν να οριστούν (π.χ. 6 μήνες), ενώ σε χρονοσειρές με έντονη κυκλικότητα, τα διαστήματα μεταξύ επαναλήψεων της ίδιας συμπεριφοράς μπορεί να διαρκούν χρόνια και να μην μπορούν να οριστούν εύκολα,
- ο ασταθής παράγοντας (irregular component), που περιγράφει ασταθείς, απρόβλεπτες μεταβολές της χρονοσειράς. [8]

Συνολικά, μια χρονοσειρά μπορεί να γραφεί ως

$$Y_t = T_t + S_t + C_t + I_t,$$

όπου Y η χρονοσειρά, T η τάση της χρονοσειράς, S η εποχικότητα, C η κυκλικότητα και I ο ασταθής παράγοντας.

Βάσει των Διαγραμμάτων Χρόνου, που αποτελούν τα διαθέσιμα δεδομένα της εργασίας, μπορούν να αντληθούν χρήσιμα μαθηματικά στοιχεία (mathematical features) που καθιστούν την κάθε χρονοσειρά ξεχωριστή μεμονωμένα, αλλά που ενδέχεται να παρουσιάζουν και αρκετές ομοιότητες ως προς τη συμπεριφορά τους. Τα στοιχεία αυτά μπορεί να αφορούν τη μεταβολή των τιμών των παρατηρήσεων, την εξάρτηση μεταγενέστερων τιμών των παρατηρήσεων από παλαιότερες, καθώς και τη διασπορά των παρατηρήσεων όσον αφορά το πλάτος (amplitude) τους. Σε αυτό το πλαίσιο, η ανάλυση διαφόρων στατιστικών στοιχείων κατά μήκος ενός αριθμού ξεχωριστών χρονοσειρών είναι κρίσιμη για την κατηγοριοποίηση των χρονοσειρών σε συγκεκριμένες ομάδες, κάθε μια από τις οποίες θα έχει ένα δικό της χαρακτηριστικό γνώρισμα που την καθιστά ξεχωριστή. Έτσι, χρονοσειρές που παρουσιάζουν μεγάλες ομοιότητες μεταξύ τους σε ορισμένα χαρακτηριστικά θα μπορούν να θεωρηθούν παρόμοιες, και αντιστρόφως. Ακολουθώντας αυτή τη μέθοδο, οι χρονοσειρές μπορούν να μετατραπούν κατά κάποιον τρόπο σε ένα σύνολο από στατιστικά, τα οποία μπορούν να τις περιγράψουν πλήρως χωρίς αυτά τα στατιστικά να είναι φανερά απλά παρατηρώντας τις τιμές των παρατηρήσεων της χρονοσειράς.

Έπειτα, θα μπορέσουν να ομαδοποιηθούν με βάση τις μεταξύ τους ομοιότητες και διαφορές, ώστε αυτές οι ομάδες να αξιοποιηθούν για τις επόμενες φάσεις της μελέτης.

Σε αυτό το κεφάλαιο θα αναπτυχθούν τα στοιχεία που θα συνεισφέρουν στην συσταδοποίηση των διαθέσιμων δεδομένων, αναλύοντας ταυτόχρονα τι μετρά το καθένα και γιατί επιλέχθηκε. Η κάθε υποενότητα θα περιέχει τις κατάλληλες πληροφορίες για τη σφαιρική κατανόηση του σκεπτικού που ακολουθήθηκε για την εργασία.

Η εξόρυξη και ανάλυση των χρήσιμων χαρακτηριστικών των χρονοσειρών, που θα χρησιμεύσουν στο παρόν έργο, έγινε αποκλειστικά στο περιβάλλον της γλώσσας προγραμματισμού Python, και οι κύριες βιβλιοθήκες που χρησιμοποιήθηκαν για την ανάλυση ήταν η SciPy και η NumPy.

Η SciPy έχει ενσωματωμένο μεγάλο βαθμό συναρτήσεων υπολογισμού στατιστικών μιας χρονοσειράς, και αποτελεί θεμελιώδες εργαλείο της προγραμματιστικής ανάλυσης δεδομένων, διότι επιτρέπει τον ταχύ υπολογισμό περίπλοκων στατιστικών στοιχείων, ενώ ταυτόχρονα δεν επιβαρύνει τον χρήστη, καθώς χρησιμοποιεί απλουστευμένο συντακτικό. Η NumPy είναι ιδιαίτερα χρήσιμη στην αποθήκευση και στην ανάλυση πινάκων και διανυσμάτων, ενώ ενσωματώνει περίπλοκες μαθηματικές διανυσματικές πράξεις και πράξεις μεταξύ πινάκων, που ενδέχεται χωρίς τη χρήση της βιβλιοθήκης να αυξήσουν την πολυπλοκότητα και το χρόνο εκτέλεσης του προγράμματος. [9]

2.1 Δημιουργία και Επιλογή Στατιστικών Δεικτών για την Περιγραφή των Χαρακτηριστικών Παραγωγής Ενέργειας

Η ανάλυση και η ερμηνεία της συμπεριφοράς του ανέμου σε διαφορετικά αιολικά πάρκα αποτελεί μια σύνθετη διαδικασία, η οποία απαιτεί τη χρήση ποικίλων στατιστικών και δυναμικών χαρακτηριστικών. Η ταχύτητα του ανέμου παρουσιάζει εγγενή μεταβλητότητα, επηρεαζόμενη τόσο από περιοδικά φαινόμενα (όπως η εναλλαγή ημέρας-νύχτας και οι εποχές) όσο και από πιο χαοτικούς παράγοντες, όπως τυχόν μικροκλιματικές συνθήκες.

Για την πληρέστερη αποτύπωση αυτής της δυναμικής, η παρούσα ενότητα καταστρώνει και αναλύει ένα οργανωμένο σύνολο χαρακτηριστικών από τις χρονοσειρές ταχύτητας ανέμου κάθε πάρκου. Τα χαρακτηριστικά αυτά ομαδοποιούνται σε τέσσερις βασικές κατηγορίες, καθεμία από τις οποίες αποτυπώνει μια διαφορετική πτυχή της αιολικής συμπεριφοράς. Η προσέγγιση αυτή επιτρέπει μια πιο στοχευμένη και ερμηνεύσιμη ανάλυση, ενισχύοντας τη συγκρισιμότητα και τη δυνατότητα συσταδοποίησης μεταξύ των πάρκων.

Στο τέλος κάθε κατηγορίας, θα αναφερθούν επίσης χαρακτηριστικά τα οποία αρχικά υπολογίστηκαν, αλλά στη συνέχεια λόγω υψηλής αυτοσυσχέτισης με άλλα χαρακτηριστικά (αρκετά από αυτά περιέγραφαν όμοια πράγματα, απλώς με διαφορετικό πρόσημο ή κατεύθυνση), “κόπηκαν” από τη λίστα.

Οι κατηγορίες χαρακτηριστικών που θα εξεταστούν είναι οι εξής:

- Στατιστικά Χαρακτηριστικά Κατανομής και Διασποράς: Περιγράφουν τη μορφή της κατανομής των ταχυτήτων ανέμου, όπως η ασυμμετρία, η διασπορά και η μεταβλητότητα.
- Δείκτες Ρυθμού Αλλαγών και Μνήμης της Χρονοσειράς: Αντιπροσωπεύουν τη χρονοσειριακή πολυπλοκότητα, τη συχνότητα των μεταβολών και τη χρονική συσχέτιση των τιμών.
- Ημερήσια Μέση Τιμή Ανέμου: Αποτυπώνουν την μέση κανονικοποιημένη τιμή του ανέμου, υπολογισμένη για κάθε ώρα της ημέρας.
- Εποχικότητα: Εξετάζει την ύπαρξη ανοδικών ή πτωτικών συμπεριφορών μέσα στη διάρκεια ενός έτους.

2.1.1. Χαρακτηριστικά Διασποράς και Κατανομής

- Λοξότητα (skewness): είναι ένα στατιστικό μέτρο που περιγράφει τον βαθμό ασυμμετρίας μιας κατανομής σε σχέση με τον μέσο όρο της.
Ο τύπος της λοξότητας είναι:

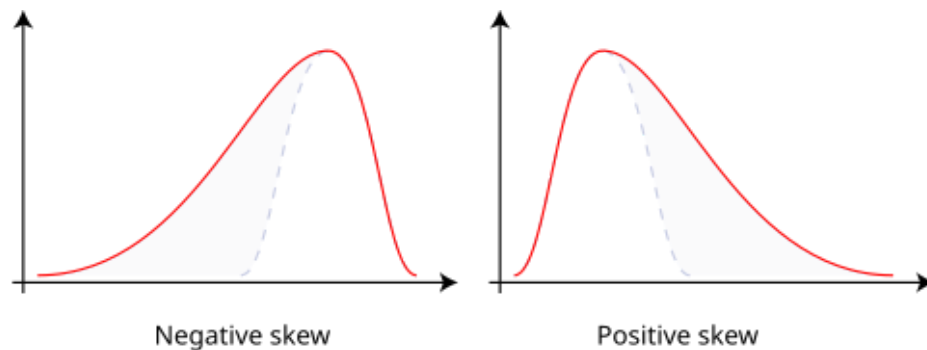
$$\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3 = \frac{\mu_3}{\sigma^3}$$

όπου μ ο μέσος όρος του δείγματος και σ η τυπική απόκλιση

Η τρίτη δύναμη χρησιμοποιείται για να δοθεί βάρος στις ακραίες τιμές της λοξότητας.

Παρατηρείται πως αν η λοξότητα είναι θετική, τα περισσότερα δεδομένα έχουν χαμηλές τιμές και υπάρχουν μερικά με πολύ υψηλές τιμές. Αντίθετα, αν η λοξότητα είναι αρνητική, τα περισσότερα δεδομένα είναι υψηλά, με κάποιες σπάνιες μικρές τιμές.

Οπτικά, η λοξότητα μπορεί να φανεί εύκολα από το σχήμα του ιστογράμματος. Μια θετικά λοξή κατανομή έχει «ουρά» προς τα δεξιά, ενώ μια αρνητικά λοξή έχει ουρά προς τα αριστερά.



Εικόνα 7: Αρνητική και θετική λοξότητα, αντίστοιχα. Πηγή: Wikiversity

Σε περίπτωση μηδενικής λοξότητας, ο μέσος όρος του δείγματος ταυτίζεται με τη μέτρηση του δείγματος, επομένως η κατανομή είναι κανονική ή γκαουσιανή, και το δείγμα έχει πλήρως συμμετρική συμπεριφορά.

- Εντροπία (entropy): είναι ένα μέτρο της αβεβαιότητας, τυχαιότητας ή "απροβλεψιμότητας" μιας κατανομής ή ενός συστήματος. Μετράει πόσο ομοιόμορφη ή πόσο διάχυτη είναι η πληροφορία.

Για διακριτές τιμές μιας κατανομής, ο τύπος της εντροπίας είναι:

$$H = - \sum_i p(i) \log_b p(i)$$

όπου $p(i)$ είναι η πιθανότητα να εμφανιστεί η τιμή i .

Η εντροπία συνδέεται συχνά με τα ιστογράμματα της συνάρτησης πυκνότητας πιθανότητας μιας μεταβλητής. Υψηλή εντροπία συνεπάγεται μεγαλύτερη μεταβλητότητα, καθώς οι διάφορες τιμές που μπορεί να λάβει μια μεταβλητή έχουν παρόμοιες πιθανότητες, ενώ χαμηλές τιμές εντροπίας συνεπάγονται μεγαλύτερη σταθερότητα.

Εύκολα επομένως βγαίνει το συμπέρασμα, πως πάρα με υψηλή εντροπία αναμένεται να είναι τα πιο αλλοπρόσαλλα και δύσκολα στην εκτίμηση, ενώ αυτά με χαμηλή εντροπία θα είναι πιο εύκολα στην πρόβλεψη.

- Εύρος (range): ένα απλό αλλά σημαντικό στατιστικό μέτρο που περιγράφει τη διασπορά των τιμών σε ένα σύνολο δεδομένων. Μετράει την πλήρη έκταση των τιμών που εμφανίζονται στο σύνολο δεδομένων. Δηλαδή, πόσο "απέχουν" η μικρότερη από τη μεγαλύτερη τιμή, και ο τύπος του είναι

$$\text{Range} = \max(x) - \min(x),$$

όπου $\max$ και $\min$ είναι η μέγιστη και ελάχιστη τιμή, αντίστοιχα, που παρατηρείται σε όλο το μήκος του δείγματος.

Το εύρος είναι άλλος ένας δείκτης της διασποράς που παρουσιάζει ένα πάрко, αλλά η απλότητα του ως μέτρο το καθιστά πολύ ευαίσθητο σε ακραίες τιμές (outliers), ενώ δεν δείχνει πόσο συχνά εμφανίζονται οι τιμές κοντά στο ελάχιστο ή το μέγιστο. Για αυτό και το εύρος από μόνο του δεν επαρκεί για πλήρη αξιολόγηση της διακύμανσης και συνδυάζεται πάντα με άλλα μέτρα σχετικά με τη μεταβλητότητα, όπως π.χ. η προαναφερθείσα εντροπία.

- Δείκτης ή Συντελεστής Διακύμανσης (Coefficient of Variation, CoV): ένα στατιστικό μέτρο που χρησιμοποιείται για να ποσοτικοποιήσει τη σχετική διακύμανση της ταχύτητας του ανέμου. Μετράει πόσο "ασταθείς" είναι οι τιμές σε σχέση με τον μέσο όρο τους, ανεξαρτήτως από το απόλυτο μέγεθος των ταχυτήτων, κάνοντας το ένα πολύτιμο εργαλείο στον διαχωρισμό ασταθών και σταθερών πάρκων.

Ο τύπος του δείκτη είναι:

$$CoV = \frac{\sigma}{\mu}$$

όπου σ η τυπική απόκλιση μιας κατανομής, και μ ο μέσος όρος των τιμών της κατανομής αυτής.

Σύμφωνα με τον τύπο βγαίνει το συμπέρασμα πως χαμηλές τιμές του CoV καταδεικνύουν σταθερή μικρή διακύμανση γύρω από το μέσο όρο ταχυτήτων ενός πάρκου, ενώ αντίθετα υψηλές τιμές καταδεικνύουν έντονη διακύμανση των τιμών.

- Πλήθος Κορυφών (Number of Peaks): είναι ένα στοιχείο που μετράει πόσες τοπικές κορυφές (τοπικά μέγιστα) υπάρχουν μέσα σε μια χρονοσειρά. Το πλήθος κορυφών υπολογίζει τα σημεία της χρονοσειράς των οποίων η τιμή είναι μεγαλύτερη από τις γειτονικές της. Καθώς ουσιαστικά μετράει πόσα τοπικά "σκαμπανεβάσματα" έχει η χρονοσειρά, μπορεί να θεωρηθεί ένας απλός στην εφαρμογή και κατανόηση δείκτης πολυπλοκότητας και μεταβλητότητας, δείχνοντας πόσο δυναμική είναι η χρονοσειρά της ταχύτητας του ανέμου. [10]

Για να θεωρηθεί ένα σημείο i ως κορυφή, θα πρέπει να ισχύει:

$$x[i] > x[i-1] \text{ και } x[i] > x[i+1]$$

Είναι προφανές πως υψηλές τιμές του πλήθους κορυφών είναι ενδεικτικές πάρκων που παρουσιάζουν συχνές μεταβολές και ριπές. Φαίνεται επομένως ποια πάρκα παρουσιάζουν ήρεμη ή ταραχώδη συμπεριφορά, και μπορούν να καταταχθούν αναλόγως.

Σε αυτό το σημείο, είναι καλό να γίνει αναφορά σε δύο στοιχεία αυτής της κατηγορίας που είχαν υπολογιστεί και ληφθεί υπόψη κατά την έρευνα, αλλά έπειτα από δοκιμές απορρίφθηκαν κατά το πειραματικό μέρος, για λόγους υψηλής συσχέτισης με άλλα χαρακτηριστικά, όπως θα φανεί σε επόμενη ενότητα:

1. Το πρώτο τέτοιο στοιχείο είναι η κύρτωση (kurtosis). Ως στοιχείο μετρά την "αιχμηρότητα" ή αντίστοιχα την "επιπεδότητα" της κορυφής της κατανομής, αν δηλαδή υπάρχουν πολλές τιμές άκρων.

Ενώ σαν χαρακτηριστικό δείχνει πόσο συχνά συναντώνται ακραίες τιμές, συνήθως είναι υψηλά συσχετισμένο με την λοξότητα. Πράγματι, εκτελώντας κώδικα που υπολογίζει τις συσχετίσεις όλων των επιλεγμένων χαρακτηριστικών, προέκυψε πως τα δύο χαρακτηριστικά έχουν συσχέτιση της τάξης του 0.93, γεγονός που καθιστά λογικό βήμα την απόρριψη ενός εκ των δύο χαρακτηριστικών προς αποφυγή έντονης επιρροής των αποτελεσμάτων, καθώς η περίληψη και των δύο είναι περιττή.

Η κύρτωση απορρίφθηκε επίσης, διότι οι στόχοι της εργασίας εστιάζουν στην πρόβλεψη και μετέπειτα μοντελοποίηση των διαθέσιμων δεδομένων, και η λοξότητα συνδράμει περισσότερο με αυτούς τους στόχους. Η λοξότητα δείχνει αν ένα πάρκο έχει συχνότερα υψηλές ή χαμηλές τιμές, δηλαδή την ύπαρξη τάσεων προς υψηλούς ή ήπιους ανέμους. Αντίθετα, η κύρτωση δείχνει την πυκνότητα των ακραίων τιμών, επομένως είναι και λιγότερο κρίσιμη.

2. Ένα δεύτερο στοιχείο είναι η διασπορά (variance). Ως στοιχείο μετράει την απόσταση των τιμών μιας μεταβλητής από το μέσο όρο της, και ο τύπος της είναι:

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

Καθώς όμως ο CoV και η εντροπία έχουν ήδη συμπεριληφθεί ως στοιχεία, τα οποία αναμένεται να έχουν έντονη σχέση με τη διασπορά, θεωρείται ασφαλές βήμα το να απορριφθεί τη διασπορά προς όφελος της καθαρότητας των αποτελεσμάτων. Πράγματι προέκυψε πως η διασπορά είχε υψηλές τιμές συσχέτισης και με τα δύο παραπάνω στοιχεία (κοντά στο 0.9), επομένως ορθά μπορεί να απορριφθεί στη συγκεκριμένη έρευνα. [11]

2.1.2. Δείκτες Μεταβολής της Χρονοσειράς

- Μέση Απόλυτη Αλλαγή (Mean Absolute Change - MAC): ένα στατιστικό μέτρο που υπολογίζει τη μέση απόλυτη διαφορά μεταξύ διαδοχικών τιμών σε μια χρονοσειρά. Είναι ιδιαίτερα χρήσιμο για τη μέτρηση της ομαλότητας ή αστάθειας στη διακύμανση ενός σήματος — στην προκειμένη περίπτωση, της ταχύτητας του ανέμου. Με άλλα λόγια, το MAC μετράει πόσο "συχνά και έντονα" αλλάζουν διαδοχικά οι τιμές σε μια χρονική σειρά, σε αντίθεση με την τυπική απόκλιση που εξαρτάται από το πόσο μακριά είναι οι τιμές από τον μέσο όρο. Αυτό το καθιστά πιο ευαίσθητο στη ροή της μεταβολής και λιγότερο επηρεασμένο από ακραίες τιμές.

Ο μαθηματικός τύπος του MAC είναι

$$MAC = \frac{1}{n-1} \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

όπου x_i η τιμή της χρονοσειράς στη χρονική στιγμή i και n το πλήθος των τιμών στη χρονοσειρά.

Με βάση τον τύπο βγαίνει το πόρισμα πως ένα πάρκο με υψηλό MAC έχει ασταθές, έντονα μεταβαλλόμενο προφίλ, ενώ ένα πάρκο με χαμηλό MAC έχει σταθερότερη ροή και λιγότερες αιφνίδιες αλλαγές στις ταχύτητες, δημιουργώντας ένα σταθερό προφίλ.

- Αυτοσυσχέτιση (Autocorrelation, AC): είναι ένα στατιστικό μέτρο που ποσοτικοποιεί τη γραμμική εξάρτηση μεταξύ των τιμών X_t και X_{t+h} μιας τυχαίας μεταβλητής και την ομοιότητα μεταξύ των παρατηρήσεων της. Με πιο απλά λόγια, είναι ένας δείκτης της μνήμης που έχει μια χρονοσειρά, και πόσο επηρεάζουν οι προγενέστερες τιμές της σειράς τις μελλοντικές.

Στα πλαίσια της έρευνας, ενδιαφέρον έχει κυρίως η αυτοσυσχέτιση σε ένα χρονικό διάστημα μιας ημέρας, καθώς στόχος είναι η ελαχιστοποίηση των σφαλμάτων στις προβλέψεις.

Ο γενικός τύπος της αυτοσυσχέτισης είναι:

$$AC(k) = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

όπου $\bar{x}$ ο μέσος όρος της χρονοσειράς, n ο αριθμός των μετρήσεων της χρονοσειράς και k η καθυστέρηση που έχει επιλεγεί.

Στην προκειμένη περίπτωση, το k ισούται με 24 και ο τύπος υπολογίζει κατά πόσο η ταχύτητα του ανέμου για μια συγκεκριμένη ώρα της ημέρας ισούται με την ταχύτητα του ανέμου την ίδια ώρα της προηγούμενης ημέρας.

Πολύ θετικές τιμές της αυτοσυσχέτισης υποδηλώνουν επαναληψιμότητα εντός του 24ώρου, επομένως είναι ένα ευνοϊκό στοιχείο για την προβλεψιμότητα και την εύρεση τυχόν περιοδικότητας της χρονοσειράς. Τιμές κοντά στο 0 υποδηλώνουν τυχαία ή ασταθή συμπεριφορά, που προβλέπεται πιο δύσκολα, ενώ αρνητικές τιμές υποδηλώνουν ότι οι υψηλές τιμές ακολουθούνται συστηματικά από χαμηλές (και αντίστροφα), κάτι που δεν ευνοεί σταθερή πρόβλεψη.

- Κυρίαρχη συχνότητα κατά FFT – Fast Fourier Transform (Dominant FFT Frequency): είναι το συχνотικό χαρακτηριστικό που δείχνει σε ποια συχνότητα "συμβαίνει" το ισχυρότερο, ενεργειακά, περιοδικό μοτίβο μέσα σε μια χρονοσειρά, όπως αυτή της ταχύτητας ανέμου. Με άλλα λόγια, υπολογίζει τη συχνότητα με το μέγιστο πλάτος στο φάσμα της συχνότητας Fourier, η οποία συχνότητα είναι και η πιο επαναλαμβανόμενη συνιστώσα της χρονοσειράς. Μονάδα μέτρησης της είναι οι κύκλοι ανά ώρα ή cycles per hour (cph).

Αρχικά υπολογίζεται το φάσμα συχνοτήτων, αφαιρώντας ταυτόχρονα τη μέση τιμή που θα αντιστοιχούσε στη συνιστώσα των 0 Hz, άρα στη DC συνιστώσα. Η συμπερίληψη της DC συνιστώσας δεν είναι επιθυμητή, καθώς θα έχει το μεγαλύτερο φάσμα και θα επηρεάσει τα αποτελέσματα, "καλύπτοντας" τις υπόλοιπες συνιστώσες που σχετίζονται με την περιοδικότητα της χρονοσειράς. Αποτρέπεται επομένως η απώλεια σημαντικής πληροφορίας, εστιάζοντας στις κυκλικές συμπεριφορές που θα παρουσιάζει η χρονοσειρά και όχι στην απόλυτη τιμή. [12]

Έχοντας κάνει τα παραπάνω, καταγράφεται η μέγιστη τιμή του index που έχει δημιουργηθεί για κάθε πάρκα, και η συχνότητα που επιστρέφεται και αποθηκεύεται είναι η "κυρίαρχη".

Τα αποτελέσματα που αναμένονται είναι υψηλές τιμές FFT συχνότητας για πάρκα με μικρή περίοδο, χαμηλές τιμές FFT συχνότητας για πάρκα με μεγάλη περίοδο και τιμές κοντά στο μηδέν για πάρκα με πολύ μεγάλη περίοδο ή πάρκα που δείχνουν οριακή περιοδική και χαοτική συμπεριφορά, και επομένως ελάχιστη επαναληψιμότητα.

- Πολυπλοκότητα Lempel-Ziv (Lempel-Ziv Complexity, LZC): ένα μέτρο που εκτιμά την πολυπλοκότητα ή το πλήθος διαφορετικών μοτίβων μέσα σε μια συμβολοσειρά ή χρονοσειρά. Προέρχεται από τον αλγόριθμο συμπίεσης Lempel-Ziv και χρησιμοποιείται για να μετρήσει πόσο "μη-επαναλαμβανόμενο" ή "μη-προβλέψιμο" είναι ένα σήμα. [13]

Ο αλγόριθμος Lempel-Ziv λειτουργεί ως εξής:

1. Υπολογίζεται η διάμεσος / μέση τιμή του σήματος.
2. Στη συνέχεια, κάθε παρατήρηση (δηλαδή η ταχύτητα του ανέμου κάθε ώρα) μετατρέπεται σε δυαδικό σύστημα, λαμβάνοντας την τιμή 1 αν η ταχύτητα είναι πάνω από τη μέση τιμή ή την τιμή 0 αν η ταχύτητα είναι κάτω από τη μέση τιμή.
3. Το δεύτερο μέρος του αλγορίθμου αναλαμβάνει να μετρήσει πόσες διαφορετικές “υποακολουθίες” υπάρχουν εντός του σήματος, μετρώντας παράλληλα πόσες φορές εμφανίζεται η κάθε υποακολουθία.

Βάσει της λειτουργίας του αλγορίθμου είναι λογικό συμπέρασμα πως μεγαλύτερη τιμή της πολυπλοκότητας σημαίνει περισσότερες υποακολουθίες και συνεπώς μειωμένη προβλεψιμότητα. Τα ευκόλως προβλέψιμα πάρκα είναι εκείνα με χαμηλή πολυπλοκότητα και λιγότερες υποακολουθίες.

2.1.3. Ημερήσια Κανονικοποιημένη Μέση Τιμή Ανέμου

- Κανονικοποιημένη Μέση Ταχύτητα (Standardised Mean Speed - SMS): ένα απλό μέτρο υπολογισμού της μέσης ταχύτητας ανά ώρα της ημέρας σε κάθε πάρκο, κανονικοποιημένο σύμφωνα με την τυπική απόκλιση και τη μέση τιμή κάθε πάρκου ώστε να δοθεί έμφαση στο μοτίβο αύξησης ή μείωσης της ταχύτητας και όχι στην καθαρή απόλυτη τιμή των ταχυτήτων.

Αρχικά υπολογίζεται η μέση ταχύτητα για κάθε ώρα της ημέρας:

$$\mu_i = \frac{1}{N_i} \sum_{t=1}^{N_i} v_{i,t}$$

και η τυπική απόκλιση:

$$\sigma_i = \sqrt{\frac{1}{N_i-1} \sum_{t=1}^{N_i} (v_{i,t} - \mu_i)^2}$$

Έπειτα υπολογίζεται η κανονικοποιημένη (z-τυποποιημένη) μέση ταχύτητα της ώρας h ως:

$$z_{i,h} = \frac{1}{n_{i,h}} \sum_{t \in H_{i,h}} \frac{v_{i,t} - \mu_i}{\sigma_i}$$

Δεδομένου ότι υπήρχε σημαντική διαφορά στις ταχύτητες των ανέμων ανά πάρκο, όπως θα φανεί και σε επόμενο κεφάλαιο, η κανονικοποίηση είναι απαραίτητη για την ήπια επιρροή των ταχυτήτων στην συσταδοποίηση, για να αποφευχθεί πιθανή στρέβλωση (skewing) των αποτελεσμάτων. Άλλωστε, το σημαντικότερο είναι να συμβάλλουν οι ταχύτητες στην συσταδοποίηση με τη συμπεριφορά τους, και όχι με τον καθαρό αριθμό τους. Η συμπεριφορά δύο πάρκων τελείως διαφορετικών ταχυτήτων μπορεί να είναι παρόμοια (π.χ. ίδια ώρα μείωσης της ταχύτητας), οπότε θα μπορούν να ομαδοποιηθούν μαζί.

2.1.4. Ένταση Εποχικότητας Χρονοσειράς

- Η εποχικότητα (seasonality), όπως αναφέρθηκε και στην αρχή του κεφαλαίου, είναι ένα πολύ σημαντικό χαρακτηριστικό που καταδεικνύει την επαναληψιμότητα μιας χρονοσειράς σε διάρκεια περίπου ενός έτους. Μπορεί επομένως να δείξει αν μια χρονοσειρά επηρεάζεται ανα σταθερά χρονικά διαστήματα από εξωγενείς παράγοντες, στην περίπτωση της εργασίας από π.χ. καλοκαιρινούς ανέμους, χειμωνιάτικες

κακοκαιρίες, μεταξύ άλλων. Η Ένταση Εποχικότητας επομένως θα είναι ένα χρήσιμο εργαλείο που θα καταδεικνύει αν ένα πάρκο επηρεάζεται από εποχικές συμπεριφορές στη διάρκεια ενός χρόνου, και μάλιστα σε τι βαθμό θα συμβαίνει αυτό.

Ανακαλώντας τον τύπο στην αρχή του κεφαλαίου για την αποσύνθεση μιας χρονοσειράς:

$$Y_t = T_t + S_t + C_t + I_t$$

η Ένταση Εποχικότητας μπορεί να οριστεί ως:

$$F_s = \max(0, 1 - \frac{Var(I_t)}{Var(I_t + S_t)})$$

υπολογίζοντας τις διακυμάνσεις του ασταθή παράγοντα και του αθροίσματος του ασταθή παράγοντα και του παράγοντα εποχικότητας. Ο τύπος αυτός είναι γνωστός και ως Ένταση Εποχικότητας κατά Loess, και η μέθοδος αποσύνθεσης αυτή ως Εποχική-Τασική Αποσύνθεση κατά Loess (Seasonal-Trend decomposition with Loess, STL), η οποία αναπτύχθηκε το 1990 από τους Cleveland et al. [14]

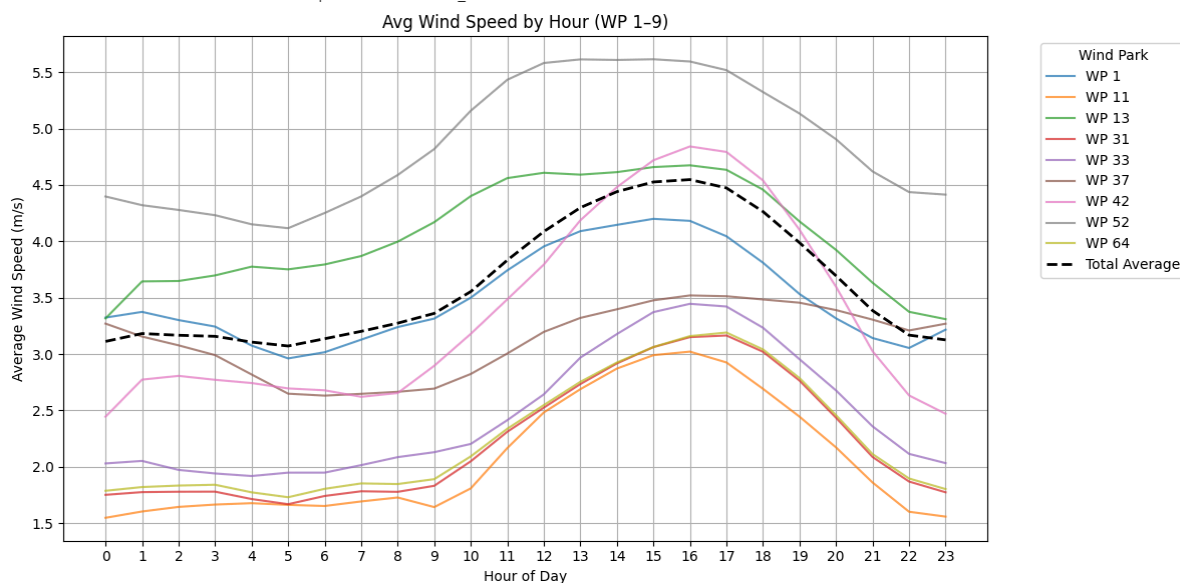
Είναι προφανές πως οι τιμές της Έντασης Εποχικότητας βρίσκονται αποκλειστικά στο διάστημα [0, 1]. Τιμή κοντά στο 1 σημαίνει πως υπάρχει έντονη εποχικότητα για τη χρονοσειρά, επομένως επηρεάζεται σε μεγάλο βαθμό από εποχικές συμπεριφορές, ενώ αντίστροφα μια τιμή κοντά στο 0 σημαίνει πως η χρονοσειρά δεν καταδεικνύει κάποια επαναλαμβανόμενη συμπεριφορά κατά τη διάρκεια του χρόνου, όπως π.χ. μια αύξηση των ταχυτήτων του ανέμου την καλοκαιρινή περίοδο.

2.2 Μείωση και Απλοποίηση των Χαρακτηριστικών

Προσθέτοντας όλα τα προαναφερθέντα στοιχεία, ο αριθμός που θα χρησιμοποιηθεί για την συσταδοποίηση καταλήγει να είναι 35. Ένας τόσο μεγάλος αριθμός όμως κοστίζει πόρους, κάνει περίπλοκη τη διαδικασία της συσταδοποίησης προσθέτοντας πολλές μεταβλητές, ενώ εγκυμονεί πάλι ο κίνδυνος να διαστρεβλωθούν τα αποτελέσματα με την υπερβολική βάση στις μέσες τιμές. Είναι απαραίτητο λοιπόν να μειωθούν, ελέγχοντας ζευγάρια μεταβλητών, των οποίων η ταυτόχρονη συμπερίληψη είναι περιττή επειδή μεταφέρουν ίδια πληροφορία ποιοτικά.

Ένα εύκολο πρώτο μέτρο περιορισμού των μεταβλητών είναι η περικοπή των 24 ωριαίων στατιστικών που έχουν υπολογιστεί και ο χωρισμός της εκάστοτε ημέρας σε τμήματα ωρών. Είναι εφικτό να περιορίσουμε την πληροφορία της μέσης τιμής των ταχυτήτων σε μερικά ωριαία διαστήματα, κατά τα οποία δεν παρατηρούνται ιδιαίτερες διακυμάνσεις, και στην “αλλαγή βάρδιας” τους παρατηρούνται αλλαγές στα μοτίβα των μέσων τιμών, είτε αυτές είναι τάσεις απότομης αύξησης ή μείωσης, είτε κάποια ταλάντωση στη μέση τιμή.

Αν εκτυπωθεί ενδεικτικά ένα διάγραμμα των μέσων ταχυτήτων των ανέμων για 9 πάρκα, λαμβάνεται το εξής:



Εικόνα 8: Ενδεικτικές μέσες τιμές ανέμου, υπολογισμένες για κάθε πάρκο ξεχωριστά.

Παρατηρώντας προσεκτικά, φαίνονται ορισμένες στιγμές της ημέρας όπου η συνάρτηση των ταχυτήτων αλλάζει με συνέπεια είτε κατεύθυνση είτε έχει έντονη αλλαγή κλίσης. Τέτοιες χρονικές στιγμές είναι οι 22:00, 17:00 και 05:00. Με βάση όλα τα παραπάνω, λαμβάνεται η απόφαση η ημέρα να χωριστεί στις εξής περιόδους:

1. 02:00 - 05:00
2. 06:00 - 09:00
3. 10:00 - 13:00
4. 14:00 - 17:00
5. 18:00 - 21:00
6. 22:00 - 01:00

Ο υπολογισμός των νέων αυτών μεταβλητών δεν απαιτεί κάτι παραπάνω από τον υπολογισμό του αριθμητικού μέσου των ταχυτήτων για το εκάστοτε 4ωρο, και την μετέπειτα κανονικοποίηση των αριθμών αυτών, με βάση τον τύπο κανονικοποίησης της ενότητας 2.1.

Έχοντας μειώσει επομένως τα στοιχεία μέσης ταχύτητας από 24 σε 6, και τα συνολικά στοιχεία από 35 σε 17, το λειτουργικό κόστος και η πολυπλοκότητα του αλγορίθμου που πρόκειται να τρέξει έχει ελαττωθεί σημαντικά.

Παρόλα αυτά, περαιτέρω απλοποιήσεις είναι εφικτές και επιθυμητές. Στοιχεία με πολύ υψηλή συσχέτιση, είτε θετική είτε αρνητική, επηρεάζουν με τον ίδιο τρόπο και την ίδια κατεύθυνση τα δεδομένα, επομένως ελλοχεύει ο κίνδυνος η επιπρόσθετη πληροφορία να επηρεάσει αρνητικά τα αποτελέσματα, μεροληπτώντας ουσιαστικά υπερ των στοιχείων αυτών και μειώνοντας την "ίση" συμμετοχή των υπόλοιπων στοιχείων. Για το σκοπό αυτό, ελέγχεται ποια από τα 20 χαρακτηριστικά παρουσιάζουν μεγάλη συσχέτιση με κάποια άλλα, και στη συνέχεια είτε ενώνονται (merge) σε ένα ενιαίο χαρακτηριστικό είτε αφαιρείται (drop) ένα από τα δύο.

Υπολογίζονται επομένως οι συσχετίσεις (correlations) όλων των ζευγών στοιχείων της λίστας που παρουσιάζουν συσχέτιση r άνω του 0.85, ορίζοντας το ως ένα όριο μετά από το οποίο δύο στοιχεία θα επηρεάζουν μαζί σε υπερβολικό βαθμό το τελικό αποτέλεσμα της συσταδοποίησης. Αναλόγως με τα ζευγάρια που θα προκύψουν, λαμβάνεται η κατάλληλη απόφαση για τη διαχείριση της "περίσσειας" πληροφορίας. Τα αποτελέσματα είναι τα εξής:

ΣΤΟΙΧΕΙΟ A	ΣΤΟΙΧΕΙΟ B	r
Skewness	Kurtosis	+0.943
SMS 22-1	SMS 6-9	-0.890
SMS 18-21	SMS 6-9	-0.865

Πίνακας 1: Υψηλά συσχετιζόμενα στοιχεία πριν την απλοποίηση.

Με βάση την επεξήγηση που δόθηκε στο 2.2.1, αφαιρείται η κύρτωση ως χαρακτηριστικό, βασιζόμενοι αποκλειστικά στην λοξότητα. Η αρνητική συσχέτιση που εμφανίζεται και στις μέσες ταχύτητες για τις παραπάνω τετράωρες περιόδους μπορεί επίσης να φανεί χρήσιμη. Αξιοποιώντας το γεγονός πως τόσο η περίοδος από 18:00 με 21:00 όσο και η περίοδος από 22:00 με 01:00 έχουν περίπου την ίδια συσχέτιση με την περίοδο από 06:00 με 09:00, υπολογίζεται ο μέσος όρος των μέσων τιμών από 18:00-01:00 και στη συνέχεια αφαιρείται η μέση τιμή της περιόδου 06:00-09:00. Έτσι, δημιουργείται ένα νέο στοιχείο, που θα φέρει το όνομα Evening-Night vs Morning Contrast, ενώνοντας 3 συσχετιζόμενα στοιχεία σε 1, χωρίς να χάνεται πρακτικά πληροφορία.

Έχει επιτευχθεί επομένως ο περιορισμός των στοιχείων που γίνουν η βάση της συσταδοποίησης από τα πολύ χρονοβόρα και περίπλοκα 35 στα πολύ πιο εύρωστα και διαχειρίσιμα 14, χωρίς σημαντική απώλεια πληροφορίας.

Στο επόμενο κεφάλαιο, θα γίνει αναφορά στις μεθόδους που θα χρησιμοποιήσουν τα θεωρητικά αυτά στοιχεία και πως αυτές λειτουργούν, προτού δοθεί εμβάθυνση σε άλλο μεταγενέστερο κεφάλαιο.

Κεφάλαιο 3: Μέθοδοι Συσταδοποίησης

Έχοντας παρουσιάσει ήδη τις μεταβλητές που θα χρησιμοποιηθούν στην εργασία και το θεωρητικό υπόβαθρο πίσω από αυτές, σε αυτό το κεφάλαιο θα δοθεί το θεωρητικό υπόβαθρο για τους αλγόριθμους και τις μεθόδους που θα χρησιμοποιηθούν για την συσταδοποίηση των δεδομένων και την αξιολόγηση, και ενδεχόμενη βελτιστοποίηση μέσω σύγκρισης των μεθόδων, των αποτελεσμάτων της συσταδοποίησης αυτής.

3.1 Αλγόριθμοι Συσταδοποίησης

Αφού τα στοιχεία που περιέχουν τις σημαντικότερες πληροφορίες σχετικά με τον στόχο της έρευνας επιλέχθηκαν, πρέπει να εφαρμοστεί μια ή και παραπάνω μέθοδοι συσταδοποίησης, ώστε τα διαθέσιμα πάρκα να χωριστούν σε ουσιαστικές και διαφοροποιημένες κατηγορίες, η κάθε μια από αυτές με τις δικές της βασικές χαρακτηριστικές ιδιότητες. Παράλληλα, θα πρέπει να ελεγχθούν τυχόν αποκλίσεις στα αποτελέσματα της κάθε συσταδοποίησης και να εντοπιστεί ένας κορμός πάρκων που ομαδοποιούνται με ίδιο τρόπο για κάθε μεθοδολογία. Ο διαχωρισμός αυτός μπορεί να επιτευχθεί με τη χρήση αλγορίθμων συσταδοποίησης.

Στο παρόν έργο θα γίνει αναφορά σε και θα χρησιμοποιηθούν τρεις από τους πιο διαδεδομένους.

- Αλγόριθμος συσταδοποίησης κατά K-means (K-means clustering)

Σύμφωνα με αυτό τον αλγόριθμο, τα στοιχεία προς συσταδοποίηση πρέπει να διαιρεθούν σε K συστάδες με βάση τη συνοχή και την ομοιότητά τους. Τα σημεία κάθε ομάδας οφείλουν να είναι όσο το δυνατόν πιο κοντά μεταξύ τους και ταυτόχρονα όσο το δυνατόν πιο μακριά από τα σημεία άλλων ομάδων. Η απόσταση μετριέται συνήθως με την Ευκλείδεια απόσταση, μέθοδος που θα χρησιμοποιηθεί και στο παρόν έργο.

Λειτουργία του αλγορίθμου:

1. Ως πρώτο βήμα, επιλέγεται ένας αριθμός συστάδων K με τη βοήθεια των μέτρων αξιολόγησης συσταδοποίησης που θα αναφερθούν στην ενότητα 3.2. Πολύ συχνά, πριν εκτελεστεί ο K-means συνηθίζεται να εκτελείται ο δείκτης σιλουέτας έτσι ώστε να έχουμε μια εμποπτεία του αριθμού συστάδων που είναι βέλτιστος.
2. Επιλέγονται τυχαία K σημεία από τα δεδομένα ως τα αρχικά κέντρα (centroids) των συστάδων.
3. Για κάθε σημείο x_j , υπολογίζεται η απόστασή του από κάθε κέντρο και το σημείο αντιστοιχίζεται στο πλησιέστερο:

$$\text{Cluster}(x_j) = \arg \min_{i \in \{1, \dots, k\}} \|x_j - \mu_i\|$$

4. Αφού ομαδοποιηθούν τα σημεία, υπολογίζεται εκ νέου το κέντρο κάθε ομάδας ως μέσος όρος των σημείων της:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

Παρατηρείται πως “αθροίζοντας” τους δύο τελευταίους τύπους, προκύπτει ο συνολικός τύπος του WCSS.

5. Ως τελικό βήμα, ο αλγόριθμος επαναλαμβάνει τα βήματα 3 και 4 (ο k-means είναι επαναληπτικός - iterative) μέχρις ότου οι αναθέσεις στις συστάδες σταθεροποιηθούν και επιτευχθεί ένας μέγιστος αριθμός επαναλήψεων, ο οποίος επιλέγεται εκ των προτέρων.

Ο αλγόριθμος είναι απλός στην κατανόηση και την υλοποίηση, και ενδείκνυται για μεγάλα σύνολα, παρέχοντας συμπαγείς και διακριτές συστάδες. Παρ'όλα αυτά, η εκ των προτέρων επιλογή του k και ο υπολογισμός σφαιρικών ομάδων (λόγω της χρήσης Ευκλείδειας απόστασης) θέτει κάποιους περιορισμούς στην αποτελεσματικότητά του. Σε τέτοιες περιπτώσεις, μπορεί να είναι προτιμότερο να χρησιμοποιηθούν μέθοδοι όπως DBSCAN ή Γκαουσιανά Μοντέλα Ανάμιξης (Gaussian Mixture Models - GMMs).

- Αλγόριθμος DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Είναι ένας πολύ δημοφιλής αλγόριθμος συσταδοποίησης που βασίζεται στην πυκνότητα των δεδομένων. Σχεδιάστηκε για να αναγνωρίζει συστάδες με αυθαίρετο σχήμα σε μεγάλα σύνολα δεδομένων με θόρυβο και ακραία σημεία (noise/outliers), φέρνοντας τον σε αντίθεση με τον K-Means, που προϋποθέτει ότι οι συστάδες είναι σφαιρικές και ισομεγεθείς. [15]

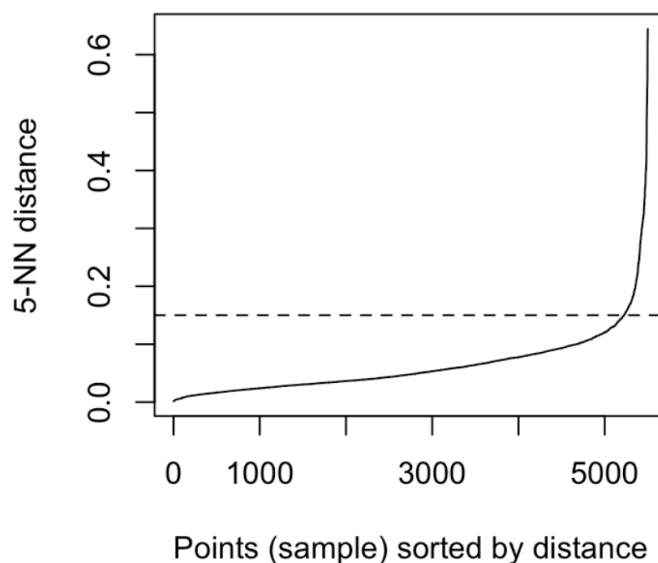
Ο στόχος είναι τα σημεία που ανήκουν στην ίδια συστάδα να είναι αρκετά κοντά μεταξύ τους και να σχηματίζουν περιοχές υψηλής πυκνότητας.

Λειτουργία του αλγορίθμου:

1. Αρχικοποίηση των δύο βασικών παραμέτρων του αλγορίθμου, του ϵ και του min_samples . Το ϵ είναι η μέγιστη επιτρεπτή απόσταση ώστε ένα σημείο να θεωρείται «γείτονας» ενός άλλου, ενώ το min_samples είναι ο ελάχιστος αριθμός γειτονικών σημείων που απαιτούνται ώστε ένα σημείο να χαρακτηριστεί ως πυκνό (core point). Σε αυτό το σημείο είναι σημαντικό να δοθούν μερικοί ορισμοί που θα είναι χρήσιμοι στη συνέχεια.

- Ως πυκνό μπορεί να χαρακτηριστεί ένα σημείο που έχει τουλάχιστον min_samples σημεία (συμπεριλαμβανομένου και του εαυτού του) σε απόσταση μέχρι ϵ από εκείνο.
- Ως οριακό (border point) μπορεί να χαρακτηριστεί ένα σημείο που έχει λιγότερους από min_samples γείτονες, αλλά ανήκει στην εμβέλεια ενός πυκνού σημείου.
- Τέλος, ο θόρυβος ή ακραίο σημείο (outlier) είναι ένα σημείο που δεν είναι ούτε πυκνό, ούτε οριακό, επομένως και εκτός όλων των συστάδων.

Μια συνηθισμένη μέθοδος για την εύρεση του κατάλληλου ϵ είναι η χρήση του διαγράμματος των k -αποστάσεων (k-distance plot), όπου παρατηρείται το “γόνατο” της καμπύλης. Η τιμή min_samples μπορεί να οριστεί αρχικά ως $\ln(n)$, όπου n το πλήθος των παρατηρήσεων.



Εικόνα 9: Παράδειγμα k-distance plot για εύρεση του κατάλληλου ϵ κατά την εφαρμογή του DBSCAN. Πηγή: Sefidian Academy

2. Για κάθε σημείο x , υπολογίζονται όλοι οι γείτονες του σημείου εντός της απόστασης ϵ .
3. Αν ο αριθμός των γειτόνων είναι μεγαλύτερος του $\min_samples$, τότε το σημείο x είναι πυκνό, και είτε δημιουργείται μια νέα συστάδα είτε προστίθεται σε μια ήδη υπάρχουσα συστάδα.
4. Οι γείτονες του σημείου προστίθενται στην ίδια συστάδα, και αν κάποιος από αυτούς είναι πυκνό σημείο, ελέγχονται και οι δικοί του γείτονες. Με αυτό τον τρόπο, μια συστάδα μπορεί να επεκταθεί χωρίς κίνδυνο σφάλματος.
Αντιθέτως αν το σημείο δεν έχει τους απαιτούμενους $\min_samples$ γείτονες, είτε είναι θόρυβος (αν δεν ανήκει σε κάποιο πυκνό σύνολο) είτε οριακό σημείο (αν είναι κοντά σε κάποιο πυκνό σημείο)
5. Ο αλγόριθμος, όπως και ο k-means, είναι επαναληπτικός, άρα εκτελεί τα παραπάνω βήματα για όλα τα σημεία των δεδομένων.

Ο DBSCAN, σε αντίθεση με τον k-means, δεν απαιτεί εκ των προτέρων αριθμό συστάδων, ενώ δεν αντιμετωπίζει προβλήματα στην αναγνώριση συστάδων οποιουδήποτε σχήματος και μπορεί εύστοχα να διακρίνει σημεία που θεωρούνται θόρυβος. Κάποια μειονεκτήματα του είναι η έλλειψη ευρωστίας σε δεδομένα μεγάλου μεγέθους και η εξάρτησή του από τις αρχικές παραμέτρους ϵ και $\min_samples$, που καθορίζουν την εκτέλεσή του.

- Αλγόριθμος Ιεραρχικής Συσταδοποίησης (Hierarchical Clustering)

Είναι μέθοδος συσταδοποίησης που στοχεύει στη δημιουργία μιας ιεραρχικής δομής συστάδων. Σε αντίθεση με μεθόδους όπως ο K-Means, που παράγουν έναν συγκεκριμένο αριθμό συστάδων, ο ιεραρχικός αλγόριθμος δημιουργεί μια δένδρική δομή ή διάγραμμα δένδρου (dendrogram) που δείχνει πολλαπλά επίπεδα συσταδοποίησης, επιτρέποντας μεγαλύτερη ευελιξία και ερμηνευσιμότητα. Κάθε κλαδί του δένδρου είναι και ένα σημείο των δεδομένων, και ο χρήστης μπορεί να επιλέξει το κατάλληλο σημείο και το κατάλληλο K που επιθυμεί για να “χωρίσει” τα δεδομένα. [16]

Λειτουργία του αλγορίθμου:

1. Κάθε σημείο θεωρείται δικό του ξεχωριστή συστάδα.
2. Υπολογίζονται οι αποστάσεις μεταξύ κάθε πιθανού ζευγαριού από συστάδες
3. Εντοπίζονται οι δύο ομάδες που έχουν την μικρότερη απόσταση από όλες, και συγχωνεύονται σε μια ενιαία συστάδα.
4. Μετά από μια συγχώνευση, οι νέες αποστάσεις μεταξύ των συστάδων επανυπολογίζονται, και η διαδικασία συνεχίζεται μέχρις ότου όλα τα σημεία να ανήκουν σε μια συστάδα, επομένως να μην είναι εφικτή κάποια περαιτέρω συγχώνευση.
Για τον επανυπολογισμό των αποστάσεων, χρησιμοποιείται η μέθοδος του Ward (Ward's Linkage Method). Η μέθοδος του Ward βασίζεται σε ένα κριτήριο ελαχιστοποίησης της αύξησης της ενδοσυσταδικής διακύμανσης (Within-Cluster Variance) κατά την ένωση δύο συστάδων. Κάθε φορά που συγχωνεύονται δύο συστάδες, επιλέγεται εκείνος ο συνδυασμός που προκαλεί τη μικρότερη δυνατή αύξηση της ολικής ενδοσυσταδικής διασποράς (WCSS).
Η ποσότητα που οφείλει να ελαχιστοποιηθεί είναι η εξής

$$E(C) = \sum_{x_i \in C} \|x_i - \mu_C\|^2$$

όπου C είναι η νέα συστάδα μετά τη συγχώνευση δύο συστάδων, x_i τα σημεία της συστάδας και μ_C το κέντρο βάρους της νέας συστάδας, το οποίο ορίζεται ως:

$$\mu_C = \frac{1}{|C|} \sum_{x_i \in C} x_i$$

5. Τέλος, κατασκευάζεται το διάγραμμα δένδρου, το οποίο καταγράφει (από κάτω προς τα πάνω) τις διαδικασίες συγχώνευσης. Δείχνει ποιες συγχωνεύσεις έγιναν πρώτες, καθώς οι προγενέστερες συγχωνεύσεις έχουν μικρότερο “κλαδί”.

Η μέθοδος της ιεραρχικής συσταδοποίησης έχει ορισμένα πλεονεκτήματα που δε διαθέτουν οι άλλες δύο μέθοδοι, ήτοι η έλλειψη αρχικών συνθηκών και η πλήρης εικόνα ομοιοτήτων που προσφέρει. Ταυτόχρονα όμως είναι ευαίσθητη σε outliers. Επιπλέον, καθώς οι συγχωνεύσεις είναι μια μη αναστρέψιμη διαδικασία λόγω της απληστίας (greediness) του αλγορίθμου, δεν μπορεί να γίνει επανεξέταση μιας πιθανώς εκ παραδρομής εσφαλμένης συγχώνευσης λόγω ύπαρξης θορύβου ή ακραίων σημείων, το οποίο ενδεχομένως οδηγήσει σε λανθασμένα συμπεράσματα και αποτελέσματα.

Παρόλα αυτά, αν και υπολογιστικά πιο απαιτητική από άλλες μεθόδους, προσφέρει πληρέστερη εικόνα των σχέσεων μεταξύ των παρατηρήσεων και επιτρέπει ευέλικτη ερμηνεία μέσω του δένδρου διαγράμματος.

3.2 Μέτρα Επιλογής Αριθμού Συστάδων και Αξιολόγησης Συσταδοποίησης

Η συσταδοποίηση αποτελεί αναμφίβολα έναν από τους βασικότερους πυλώνες της ανεπιτήρητης μηχανικής μάθησης, με εφαρμογές σε πλήθος τομέων, από την ανάλυση μεγάλων δεδομένων έως την επεξεργασία σημάτων και τη χαρτογράφηση συμπεριφορών. Ωστόσο, σε αντίθεση με τις επιβλεπόμενες μεθόδους, η ορθότητα ή η ποιότητα μιας συσταδοποίησης δεν μπορεί να αξιολογηθεί απευθείας μέσω πραγματικών ετικετών (labels). Συνεπώς, καθίσταται απαραίτητη η χρήση κατάλληλων μέτρων αξιολόγησης ποιότητας, τα οποία επιτρέπουν τη συγκριτική εκτίμηση διαφορετικών μοντέλων συσταδοποίησης και την επιλογή του κατάλληλου αριθμού συστάδων.

Σε αυτό το κεφάλαιο παρουσιάζονται τα κυριότερα μέτρα εσωτερικής αξιολόγησης (internal evaluation metrics), τα οποία βασίζονται αποκλειστικά στα ίδια τα δεδομένα και στη γεωμετρική δομή των συστάδων. Αναλύονται ενδεικτικά η μέθοδος του αγκώνα (Elbow Method), ο δείκτης σιλουέτας (Silhouette Score) και ο δείκτης Davies-Bouldin (DBI), εξηγώντας τόσο τη μαθηματική τους θεμελίωση όσο και τη σημασία τους στην αποτίμηση της συνοχής και του διαχωρισμού μεταξύ των ομάδων.

Η κατανόηση και η σωστή εφαρμογή αυτών των μέτρων είναι καθοριστικής σημασίας για την αξιόπιστη ερμηνεία των αποτελεσμάτων συσταδοποίησης, ιδιαίτερα όταν δεν υπάρχει προκαθορισμένη ταξινόμηση. Στο πλαίσιο αυτής της εργασίας, τα εν λόγω μέτρα αξιοποιούνται για την αξιολόγηση διαφορετικών μοντέλων συσταδοποίησης που εφαρμόστηκαν σε δεδομένα ταχύτητας ανέμου, με στόχο την ανακάλυψη λειτουργικά ομοιογενών ομάδων αιολικών πάρκων.

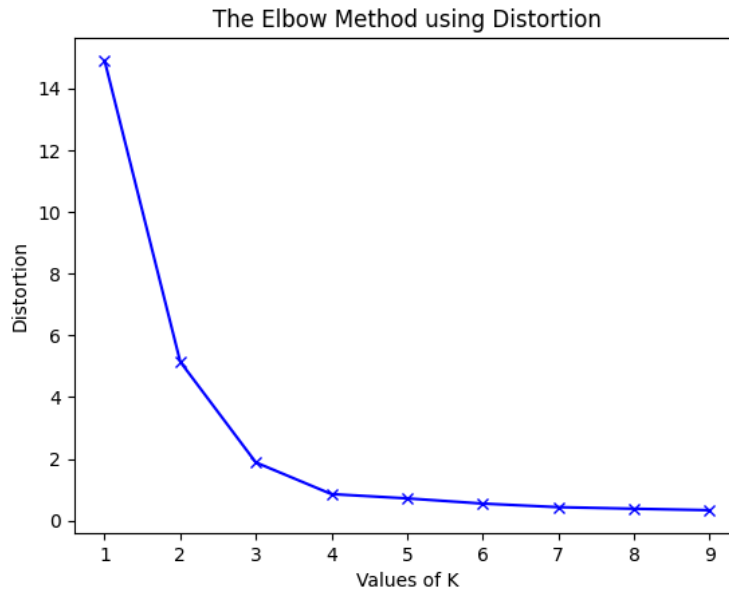
- Μέθοδος του αγκώνα (Elbow Method)

Η μέθοδος του αγκώνα είναι ένα διαγνωστικό εργαλείο για να προτείνει πόσες συστάδες είναι “αρκετές”, χωρίς να γίνεται υπερβολικός διαχωρισμός. Είναι ταυτόχρονα και ένα σημαντικό πρώτο βήμα στην αξιολόγηση της συσταδοποίησης που πρόκειται να γίνει. Βασίζεται στον υπολογισμό του ενδοσυσταδικού σφάλματος για κάθε πιθανό αριθμό συστάδων k . Το σφάλμα αυτό είναι γνωστό ως WCSS (Within-Cluster Sum of Squares), το οποίο είναι το άθροισμα των τετραγωνικών αποστάσεων των σημείων μιας συστάδας από το κέντρο της συστάδας. Ο τύπος του WCSS είναι:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

όπου C_i η συστάδα i , μ_i το κέντρο της συστάδας και x κάθε σημείο που ανήκει στην εν προκειμένω συστάδα.

Λειτουργία της μεθόδου: Αφού επιλεγθεί μια σειρά πιθανών τιμών του αριθμού των συστάδων k , εφαρμόζεται συσταδοποίηση κατά k -means και υπολογίζεται το αντίστοιχο WCSS. Στο γράφημα που θα προκύψει, παρατηρείται πως η τιμή του WCSS πέφτει, αλλά ο ρυθμός πτώσης αλλάζει. Στόχος της εφαρμογής της μεθόδου του αγκώνα είναι η εύρεση, διαισθητικά με την παρατήρηση του διαγράμματος, του σημείου από το οποίο και μετά η πτώση του ενδοσυσταδικού σφάλματος είναι σημαντικά μειωμένη και μπορεί να θεωρηθεί πως η χρήση παραπάνω συστάδων από το σημείο αγκώνα k είναι υπερβολική. [17]



Εικόνα 10: Τυχαίο παράδειγμα διαγράμματος για εξαγωγή του βέλτιστου k για εφαρμογή συσταδοποίησης. Στο σημείο για k = 3, το διάγραμμα έχει καμπή, επομένως τα δεδομένα μάλλον χωρίζονται με φυσικό τρόπο σε τρεις ομάδες. Πηγή: GeeksForGeeks

Καθώς όμως η επιλογή του k είναι απόφαση του χρήστη, υπάρχει ένας βαθμός υποκειμενικότητας. Επιπλέον, είναι συχνό φαινόμενο τα δεδομένα πραγματικού κόσμου (real-world data) να μην αναδεικνύουν με ευκρίνεια τον βέλτιστο αριθμό συστάδων, επομένως το διάγραμμα του αγκώνα σε αυτές τις περιπτώσεις να μη δίνει κάποιο ξεκάθαρο αποτέλεσμα. Για αυτό το λόγο, συνηθίζεται να χρησιμοποιούνται και οι επόμενες μέθοδοι σε συνδυασμό με τη μέθοδο αγκώνα.

Η επιλογή όμως πολύ μεγάλου αριθμού συστάδων σημαίνει πως το WCSS θα συνεχίσει να μειώνεται, γεγονός που οδηγεί σε υπερπροσαρμογή (overfitting). Αυτό θα σημαίνει πως κάθε σημείο σχεδόν θα γίνει δική του συστάδα, επομένως η συσταδοποίηση δεν έχει νόημα εξ αρχής. Η επιλογή, επομένως, του αριθμού k θα πρέπει να συγκριθεί και με τον αριθμό των δεδομένων, ώστε να ελεγχθεί αν, μαθηματικά, οι συστάδες που δημιουργούνται είναι ουσιαστικές ή “επιβλήθηκαν”.

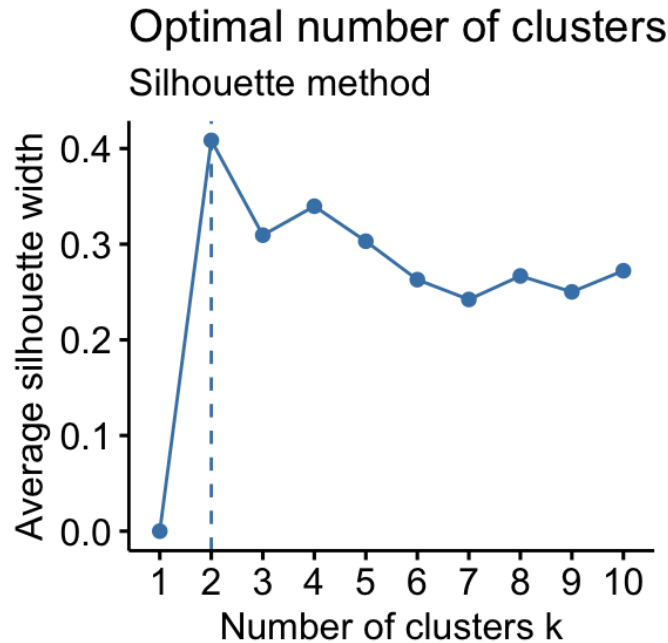
- Δείκτης Σιλουέτας (Silhouette Score): είναι ένα μέτρο αξιολόγησης ποιότητας συσταδοποίησης, το οποίο δείχνει πόσο καλά έχει τοποθετηθεί κάθε πάρκο μέσα σε μια συστάδα, σε σύγκριση με άλλες συστάδες. Εισήχθη το 1990 από τους Kaufman και Rousseeuw. Ο δείκτης σιλουέτας είναι ιδιαίτερα χρήσιμος για να αξιολογηθούν οι συστάδες ως προς το πόσο συμπαγείς και καλά διαχωρισμένες είναι. [17]

Για κάθε παρατήρηση / πάρκο i υπολογίζονται τα εξής δύο ποσά:

- $a(i)$, η μέση απόσταση του σημείου από όλα τα άλλα σημεία της ίδιας συστάδας (δείχνει πόσο “ανήκε” στη συστάδα του) και
- $b(i)$, η ελάχιστη μέση απόσταση του σημείου από τα σημεία άλλων συστάδων (δείχνει πόσο “κοντά” είναι στην πιο κοντινή άλλη συστάδα).

Με βάση αυτά τα δύο ποσά, υπολογίζεται ο δείκτης σιλουέτας ως εξής:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$



Εικόνα 11: Παράδειγμα διαγράμματος για υπολογισμό του silhouette score. Για k = 2 συστάδες ο δείκτης μεγιστοποιείται. Πηγή: datanovia.com

Σε γενικές γραμμές, η τιμή του δείκτη σιλουέτας κυμαίνεται μεταξύ του -1 και του 1.

Επομένως για τα πάρκα που ανήκουν “καλά” στη συστάδα τους, η τιμή του δείκτη θα είναι κοντά στο 1, ενώ για τα πάρκα που βρίσκονται στο μεταίχμιο, η τιμή του δείκτη θα κυμαίνεται κοντά στο 0 ή και στις αρνητικές τιμές, αν το πάρκο φαίνεται να ανήκει σε άλλη τελείως συστάδα.

Οι Kaufman και Rousseeuw αναφέρουν πως τιμές του δείκτη που κυμαίνονται κάτω του 0.25 υποδεικνύουν αδύναμη ή ακόμα και έλλειψη δομής ενός cluster. Η εκτίμηση αυτή όμως χαρακτηρίζεται ως υποκειμενική και μπορεί να αλλάξει κατά βούληση και αναλόγως του συνόλου δεδομένων. Στο παρόν έργο θα επιλεχθεί το πιο συντηρητικό όριο (threshold) της τιμής 0.15, καθώς τα δεδομένα είναι αρκετά ώστε να γίνει έρευνα, αλλά όχι τόσα ώστε να ανακαλυφθούν ισχυρές δομές στα clusters.

Η χαμηλή τιμή του δείκτη σιλουέτας για ορισμένα αιολικά πάρκα μπορεί επίσης να υποδεικνύει υβριδική συμπεριφορά, δηλαδή ότι το πάρκο παρουσιάζει χαρακτηριστικά δύο ή περισσότερων ομάδων. Με χρήση του δείκτη σιλουέτας μπορούν να καταγραφούν ποια πάρκα των δεδομένων δεν ταιριάζουν τόσο καλά στη συστάδα που έχουν ταξινομηθεί, και κατ’ επέκταση να ληφθεί κάποια απόφαση σχετικά με αυτά, είτε εντάσσοντας τα σε ειδικές υποσυστάδες, είτε διαχειριζόμενοι τα με ιδιαίτερο τρόπο, σημειώνοντας τα ως “αταίριαστα” (misfits), πάρκα δηλαδή που δεν ταιριάζουν σε καμία συστάδα από τις διαθέσιμες με σαφή και εμφανή τρόπο.

- Δείκτης Davies-Bouldin (Davies-Bouldin Index- DBI): είναι ένα μέτρο αξιολόγησης συσταδοποίησης, που χρησιμοποιείται για να μετρήσει πόσο καλά έχουν διαχωριστεί οι συστάδες μεταξύ τους, λαμβάνοντας υπόψη και τη συνοχή (συμπαγή διάταξη) κάθε συστάδας. Ταυτόχρονα, συνδυάζει τη διάμεση πυκνότητα κάθε συστάδας (πόσο κοντά είναι τα σημεία στο κέντρο της) και την απόσταση της κάθε συστάδας από τις υπόλοιπες. [18]

Ο δείκτης DBI για k συστάδες μπορεί να υπολογιστεί με χρήση του τύπου:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max(\frac{S_i + S_j}{M_{ij}}), i \neq j$$

όπου S_i η μέση απόσταση των σημείων της συστάδας i από το κέντρο του, και M_{ij} η απόσταση μεταξύ των κέντρων των συστάδων i και j.

Ο DBI λαμβάνει τον μέγιστο λόγο (worst-case) μεταξύ κάθε ζεύγους από συστάδες, υπογραμμίζοντας τις πιο "στριμωγμένες", και στη συνέχεια υπολογίζει τον μέσο όρο. Σαν μέθοδος επομένως είναι πιο συντηρητική και ευαίσθητη σε κακές διαχωριστικές περιπτώσεις.

Φαίνεται ότι ο DBI σαν μέτρο έχει τάση αντίθετη από το σκορ σιλουέτας, καθώς στόχος είναι η ελαχιστοποίηση του. Τιμές κοντά στο 0 σημαίνουν μεγάλη απόσταση μεταξύ των κέντρων των συστάδων και ταυτόχρονα μικρές αποστάσεις των πάρκων των συστάδων από το κέντρο τους, επομένως έχει επιτευχθεί ισχυρή συσταδοποίηση. Σε αντίθετη περίπτωση, η συσταδοποίηση δε μπορεί να χαρακτηριστεί άριστη, καθώς υπάρχει ένας βαθμός αλληλοκάλυψης μεταξύ των διαφορετικών συστάδων.

3.3 Μέτρα Ομοιότητας Συστάδων

Προκειμένου να ελεγχθεί κατά πόσο ο στόχος του διαχωρισμού των πάρκων σε ουσιαστικές ομάδες με διαφορετικά χαρακτηριστικά έχει επιτευχθεί, μπορούν να εφαρμοστούν δύο μέτρα που ποσοτικοποιούν την ομοιότητα των κέντρων των συστάδων.

1. Το πρώτο που θα χρησιμοποιηθεί είναι η Ευκλείδεια Απόσταση (Euclidean Distance), η οποία δεν είναι άλλο από το γεωμετρικό μήκος του ευθύγραμμου τμήματος που ενώνει τα κέντρα των συστάδων στο πολυδιάστατο χώρο των χαρακτηριστικών. Υπολογίζεται από τον τύπο

$$d(A, B) = \|A - B\|_2 = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$$

όπου n ο αριθμός των χαρακτηριστικών που έχουν χρησιμοποιηθεί και A, B τα διανύσματα χαρακτηριστικών των δύο συστάδων (τα κέντρα τους)

Μεγάλη τιμή της Ευκλείδειας Απόστασης σημαίνει μεγάλη διαφορά των κέντρων, επομένως και τα πάρκα των συστάδων θα είναι κι αυτά διαφορετικά μεταξύ τους.

2. Το δεύτερο μέτρο είναι η Συνημιτονοειδής Ομοιότητα (Cosine Similarity) μεταξύ των κέντρων. Μετράει τη γωνιακή ομοιότητα μεταξύ δύο διανυσμάτων, δηλαδή αν "κοιτούν" προς την ίδια κατεύθυνση στο χαρακτηριστικό χώρο. Υπολογίζεται ως:

$$\sin(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

όπου A, B ξανά τα προαναφερθέντα διανύσματα χαρακτηριστικών και $\|A\|, \|B\|$ τα μέτρα τους.

Προφανώς οι τιμές κυμαίνονται μεταξύ -1 και 1. Μια τιμή -1 υποδεικνύει εντελώς διαφορετική κατεύθυνση, επομένως οι συστάδες "δίνουν βάρος" σε αντίστροφα χαρακτηριστικά. Αντίστοιχα μια τιμή +1 υποδεικνύει ίδιο προσανατολισμό, επομένως και ίδια χαρακτηριστικά. Η πλήρως επιθυμητή τιμή είναι η μηδενική, καθώς σημαίνει πως οι συστάδες είναι ορθογώνιες, δηλαδή ασχέτως προσανατολισμένες.

Κεφάλαιο 4: Προτεινόμενη Μεθοδολογία

Έχοντας αναφερθεί ήδη στα στατιστικά στοιχεία που εξάγονται από τις χρονοσειρές και θα αξιοποιηθούν από τους αλγορίθμους συσταδοποίησης που αναπτύχθηκαν στο προηγούμενο κεφάλαιο, το τρέχον κεφάλαιο στις πρώτες ενότητες θα εστιάσει στην παρουσίαση του πειράματος που θα πραγματοποιηθεί, συνάμα με τις απαραίτητες προκαταρκτικές διαδικασίες τροποποίησης των δεδομένων, ώστε τα αποτελέσματα να είναι ευπαρουσίαστα, κατανοητά και αντικειμενικά (unbiased).

Στις επόμενες ενότητες του κεφαλαίου, θα εισαχθούν οι απαραίτητες έννοιες σχετικές με τα χαρτοφυλάκια που θα προσομοιωθούν, καθώς επίσης και η μέθοδος αξιολόγησης της ποιότητας τους. Τέλος, θα αναπτυχθεί σε βάθος η μέθοδος προσομοίωσης των χαρτοφυλακίων και οι σχετικές πιθανοτικές θεωρίες, προς κατανόηση του αναγνώστη.

4.1 Κανονικοποίηση Δεδομένων

Έχοντας αποφασίσει για τα εκάστοτε στοιχεία που θα χρησιμοποιηθούν κατά την συσταδοποίηση, γίνεται γρήγορα κατανοητό ένα πρόβλημα, κι αυτό είναι πως κάποια στοιχεία ενδέχεται να επηρεάσουν την συσταδοποίηση παραπάνω από ότι άλλα λόγω των μέτρων των τιμών τους. Για παράδειγμα, οι ταχύτητες των ανέμων μπορεί να είναι πάρα πολύ υψηλές για έναν μικρό αριθμό πάρκων, επειδή αυτά συναντώνται σε περιοχές κοντά σε θάλασσα. Η ανισορροπία αυτή ενδέχεται να επηρεάσει τα αποτελέσματα σε σημαντικό βαθμό, σε σημείο που τα πάρκα αυτά ενδέχεται να ομαδοποιηθούν ξεχωριστά από όλα τα άλλα, ακόμα και αν οι συμπεριφορές και η κατανομή των παρατηρήσεών τους παρουσιάζει μεγάλες ομοιότητες με αυτές άλλων πάρκων. Η μεθοδολογία που μπορεί να αίρει αυτό το πρόβλημα είναι αυτή της κανονικοποίησης (standardisation).

Η κανονικοποίηση είναι μια τεχνική επεξεργασίας που μετασχηματίζει τις τιμές ενός συνόλου δεδομένων ώστε να βρίσκονται σε μια κοινή κλίμακα – συνήθως μεταξύ 0 και 1 ή εναλλακτικά ώστε να έχουν μέσο όρο 0 και τυπική απόκλιση 1, χωρίς να αλλοιώνεται η μορφή της κατανομής και η σχετική πληροφορία που περιέχεται. Διαφορετικές κλίμακες σε χαρακτηριστικά μπορούν να οδηγήσουν σε ασυμμετρίες στην απόσταση και, συνεπώς, σε μεροληπτική συσταδοποίηση.

Με πιο απλά λόγια, η κανονικοποίηση δίνει έμφαση στα μοτίβα και τη μορφή των κατανομών, αφαιρώντας το βάρος που έχουν εκ φύσεως οι απόλυτες τιμές, οι οποίες μπορούν να επηρεάσουν δυσανάλογα τα αποτελέσματα μιας συσταδοποίησης.

Η κανονικοποίηση συναντά διαδεδομένη χρήση σε διάφορα επιστημονικά πεδία, μεταξύ άλλων αυτά [19]:

1. της μηχανικής μάθησης, καθώς πολλοί αλγόριθμοι ευρείας χρήσεως βασίζονται στην εξομάλυνση των αποστάσεων μεταξύ δεδομένων και την εύρεση μιας κοινής κλίμακας
2. της συσταδοποίησης, που αφορά την παρούσα διπλωματική,
3. της ανάλυσης χρονοσειρών, για την ανίχνευση μοτίβων ανεξαρτήτων από απόλυτες τιμές και κλίμακες με σκοπό την έγκυρη και εύστοχη πρόβλεψη,
4. της βιοιατρικής, καθώς η σύγκριση για παράδειγμα βιοδεικτών ή τιμών αίματος μεταξύ διαφορετικών ατόμων και συνθηκών καθιστούν την κανονικοποίηση απαραίτητη.

Υπάρχουν δύο διαδεδομένες μέθοδοι κανονικοποίησης, η πρώτη εκ των οποίων είναι αυτή που θα χρησιμοποιηθεί στην παρούσα εργασία.

Η τυποποίηση κατά Z (Z-Score Standardisation) φέρνει τα δεδομένα σε μορφή κανονικής / γκαουσιανής κατανομής, δηλαδή μετατρέπει την διαθέσιμη κατανομή σε μια καινούρια, με μέση τιμή μ ίση με 0 και τυπική απόκλιση σ ίση με 1. Ο τύπος είναι:

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

Η τυποποίηση κατά Z είναι κατάλληλη, διότι προσφέρει ανθεκτικότητα ακόμα και σε περιπτώσεις που επικρατούν διαφορετικές κλίμακες στα δεδομένα.

Η δεύτερη μέθοδος είναι η min-max κανονικοποίηση, η οποία φέρνει τα δεδομένα εντός του διαστήματος [0,1]. Η μέθοδος αυτή ταιριάζει περισσότερο όταν το εύρος των τιμών είναι γνωστό και δεν υπάρχουν ακραίες τιμές. Ο τύπος της είναι:

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

4.2 Μείωση Διαστάσεων

Οι αλγόριθμοι συσταδοποίησης από τη φύση τους ως μαθηματικά εργαλεία είναι γενικεύσιμα, που σημαίνει πως μπορούν να λάβουν πληθώρα στοιχείων και να παράγουν ένα αποτέλεσμα. Πολλές φορές όμως τα αποτελέσματα των αλγορίθμων είναι περίπλοκα και απλώνονται σε μεγάλο αριθμό διαστάσεων, οι οποίες δεν μπορούν φυσικά να γίνουν κατανοητές από τον άνθρωπο. Η διάταξη του πειράματος ανήκει σε αυτή την κατηγορία, καθώς ο αλγόριθμος καλείται να ομαδοποιήσει τα πάρα βάσει 14 μεταβλητών, παρά την απλοποίηση, αφαίρεση και απόρριψη των αρχικών στοιχείων. Αυτή η διάταξη δεν επιτρέπει μια καθαρή εικόνα των σχετικών θέσεων των δεδομένων, και επομένως δημιουργείται η επιτακτική ανάγκη χρήσης μιας μεθόδου που θα μειώσει τις διαστάσεις των δεδομένων, χωρίς όμως να γίνει αυτό εις βάρος της περιεχόμενης πληροφορίας.

Μια τέτοια μέθοδος είναι η Ανάλυση Βασικών Συνιστωσών (Principal Component Analysis - PCA), και χρησιμοποιείται ευρέως στην ανάλυση δεδομένων και στην απεικόνιση πολυδιάστατων συνόλων. Δεν είναι αλγόριθμος ταξινόμησης ή συσταδοποίησης, αλλά ένα εργαλείο μετασχηματισμού των δεδομένων, που βοηθά στη διατήρηση της σημαντικότερης πληροφορίας με τη χρήση λιγότερων μεταβλητών και τη συμπίεση των δεδομένων.

Η PCA μετατρέπει το σύνολο πιθανώς συσχετισμένων μεταβλητών, επομένως, σε ένα νέο σύνολο ορθογωνίων μεταβλητών, που ονομάζονται κύριες ή βασικές μεταβλητές (Principal Components - PCs). Κάθε κύρια συνιστώσα εξηγεί λιγότερη διασπορά και διατηρεί ασυσχέτιστες τις νέες μεταβλητές.

Λειτουργία της μεθόδου:

1. Αρχικά, προαπαιτείται η κανονικοποίηση όλων των δεδομένων, για να αποφευχθεί η κυριαρχία των απόλυτων τιμών εις βάρος των μοτίβων των μεταβλητών.
2. Έπειτα, κάθε στήλη των δεδομένων (που περιέχει κάθε στοιχείο που επιλέχθηκε) κεντράρεται, σύμφωνα με τον τύπο

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad Z_{ij} = x_{ij} - \bar{x}_j$$

και παίρνουμε τον τελικό πίνακα Z. Ο πίνακας αυτός αποτελείται από 34 γραμμές, μία για κάθε δεδομένο - πάρκο, και 14 στήλες, για κάθε μεταβλητή που επιλέχθηκε.

3. Για τον κεντραρισμένο πίνακα Z, υπολογίζεται ο πίνακας συνδιακύμανσης ή συσχέτισης, του οποίου τα στοιχεία C_{jk} υποδηλώνουν κατά πόσο τα στοιχεία j, k επηρεάζουν το ένα τη συμπεριφορά του άλλου. Ο πίνακας συνδιακύμανσης, ο οποίος πλέον είναι πίνακας 14 x 14, υπολογίζεται σύμφωνα με τον τύπο

$$C = \frac{1}{n-1} Z^T Z$$

4. Υπολογίζονται οι ιδιοτιμές και τα ιδιοδιανύσματα, λύνοντας την ιδιοτιμική εξίσωση

$$C v = \lambda v, \text{ όπου}$$

λ η ιδιοτιμή (eigenvalue) και v το ιδιοδιάνυσμα (eigenvector). Οι ιδιοτιμές είναι υπεύθυνες για τη διασπορά που “εξηγεί” η εκάστοτε συνιστώσα, ενώ τα ιδιοδιανύσματα είναι η κατεύθυνση στον χώρο των στοιχείων.

5. Επιλέγεται τι ποσοστό της διασποράς θέλουμε να κρατήσουμε, και αντιστοίχως κρατώνται οι ιδιοτιμές των οποίων το άθροισμα είναι μεγαλύτερο ή ίσο του ποσοστού αυτού. Με πιο απλά λόγια, σε περίπτωση που επιθυμούμε να κρατήσουμε π.χ. 80% της πληροφορίας, οπότε διασπορά 0.8, θα υπολογιστούν όσες ιδιοτιμές χρειάζονται ώστε το άθροισμα τους να είναι τουλάχιστον 0.8.
6. Κατασκευάζεται ο πίνακας των βαρών (Loading Chart) $P_{p \times k} = [v_1, \dots, v_k]$. Κάθε στήλη υποδεικνύει πώς συνδυάζονται τα αρχικά χαρακτηριστικά για να σχηματίσουν την αντίστοιχη κύρια συνιστώσα.
7. Τέλος, δε μένει παρά να προβληθούν τα δεδομένα στο νέο χώρο $T_{n \times k} = Z P$, όπου T ο πίνακας των νέων συντεταγμένων του κάθε πάρκου.

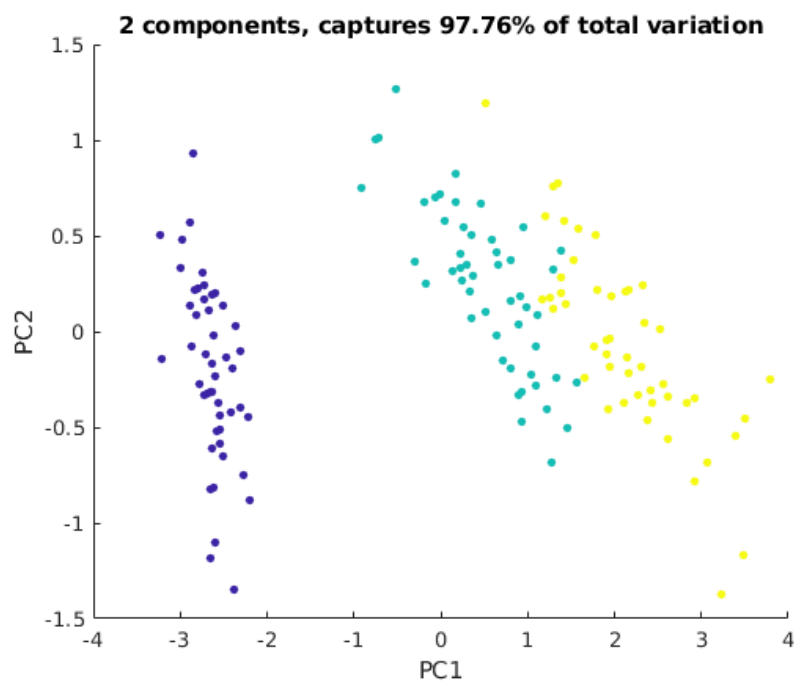
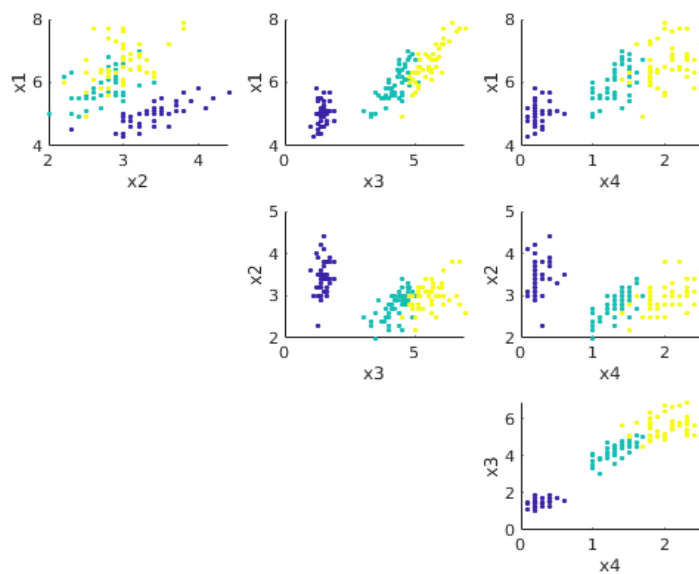
Έχοντας κάνει τους υπολογισμούς, γίνεται να οπτικοποιηθούν τα αποτελέσματα σε δισδιάστατο ή τρισδιάστατο διάγραμμα (αν χρειάζονται τουλάχιστον 3 βασικές μεταβλητές).

Η μέθοδος PCA εφαρμόζεται σε πλήθος επιστημονικών πεδίων, όπως:

- Στη βιοπληροφορική, για ανάλυση γονιδιακών εκφράσεων.
- Στην οικονομία, για σύνολα οικονομικών δεικτών.
- Στη μηχανική μάθηση, για συμπίεση δεδομένων και βελτίωση της υπολογιστικής απόδοσης.
- Στην οπτική αναγνώριση προτύπων, για μείωση χαρακτηριστικών σε εικόνες και σήματα.

Αν και ισχυρή ως τεχνική, το PCA έχει και ορισμένους περιορισμούς, καθώς η ερμηνεία των νέων αξόνων δεν είναι πάντα άμεση και τόσο ξεκάθαρη. Επιπλέον, πρόκειται για μια γραμμική μέθοδο, επομένως μη γραμμικές συσχετίσεις δεν μπορούν να ανιχνευτούν. Τέλος, είναι κρίσιμη η σωστή επιλογή της διασποράς προς κράτηση, ειδάλλως υπάρχει σοβαρός κίνδυνος να χαθεί ζωτική πληροφορία σε περίπτωση επιλογής χαμηλού ποσοστού διασποράς.

Παρακάτω βρίσκονται δύο εικόνες, που δείχνουν ένα τυχαίο σύνολο δεδομένων 4 μεταβλητών, με κάθε συνδυασμό μεταβλητών να έχει σκιαγραφηθεί σε διαγράμματα δύο διαστάσεων. Από κάτω φαίνεται το αποτέλεσμα του PCA, κρατώντας μόνο τις 2 από τις 4 διαστάσεις, καθώς πρακτικά καλύπτουν όλη τη διασπορά και χάνεται ελάχιστη πληροφορία.



Εικόνα 12: Παράδειγμα μετασχηματισμού PCA, με την κάτω εικόνα να δείχνει την απλοποιημένη εικόνα του συνόλου δεδομένων που μπορεί να χρησιμοποιηθεί. Πηγή: University of Maryland, Department of Mathematics

Συμπερασματικά διαφαίνεται πως, παρόλο που το PCA δεν επηρεάζει την συσταδοποίηση που γίνεται, βοηθά τόσο στην οπτική εμφάνιση των αποτελεσμάτων, κάνοντας τα πολύ πιο κατανοητά και “όμορφα”, αλλά και στην ανάδειξη εκείνων των στοιχείων που παίζουν μεγαλύτερο ή και μικρότερο ρόλο στον διαχωρισμό των δεδομένων.

4.3 Μέτρα Αξιολόγησης Χαρτοφυλακίων

Αφού έχουν δημιουργηθεί οι συστάδες και τα πάρκα έχουν διαχωριστεί, μπορεί να αρχίσει η έρευνα για την κατασκευή ενός χαρτοφυλακίου που αποτελείται από τυχαία πάρκα των συστάδων. Όπως έχει ήδη αναφερθεί στην ενότητα 1.4, η Θεωρία Χαρτοφυλακίων του Markowitz προϋποθέτει πως, για να έχει ένα χαρτοφυλάκιο πρακτική σημασία και να μπορεί να θεωρηθεί αξιόλογο και αξιόπιστο, θα πρέπει ο κίνδυνος να ελαχιστοποιείται και το αναμενόμενο κέρδος να αυξάνεται. Αυτό μπορεί να επιτευχθεί στην περίπτωση που τα περιουσιακά στοιχεία που προστίθενται σε αυτό έχουν όσο το δυνατόν χαμηλότερη συσχέτιση μεταξύ τους.

Στο περιβάλλον της παρούσας εργασίας, τα περιουσιακά στοιχεία είναι τα διαθέσιμα πάρκα, και η συσχέτιση που παρουσιάζουν μπορεί να ερμηνευθεί ως η συμπληρωματικότητα των συστάδων και κατ'επέκταση των πάρκων τους.

Μια πολύ σημαντική έννοια για την αξιολόγηση της ποιότητας ενός χαρτοφυλακίου είναι εκείνη του Συντελεστή Μεταβλητότητας (Coefficient of Variation, CoV), στοιχείο που χρησιμοποιήθηκε και κατά την συσταδοποίηση.

Ο CoV μετρά πόσο «σκορπίζουν» οι τιμές ενός δείγματος σε σχέση με τον μέσο όρο του. Είναι μέγεθος χωρίς μονάδες, άρα ιδανικό για να συγκρίνει τη σχετική διακύμανση μεταξύ διαφορετικών μεγεθών ή μονάδων. Ορίζεται ως ο λόγος της τυπικής απόκλισης προς τον μέσο όρο, δηλαδή

$$\text{CoV} = \sigma / \mu$$

Ερμηνευτικά, όσο μικρότερος ο CoV, τόσο πιο «συγκεντρωμένες» είναι οι τιμές γύρω από τον μέσο όρο, και με μεγαλύτερη πιθανότητα μια τιμή της χρονοσειράς θα βρεθεί κοντά στην αναμενόμενη τιμή. Αντίθετα, μεγάλος CoV υποδηλώνει υψηλή σχετική αβεβαιότητα και εξίσου μεγάλη πιθανότητα μια τιμή της χρονοσειράς να βρεθεί αρκετά πιο μακριά από την αναμενόμενη τιμή.

Στον τομέα των χαρτοφυλακίων, ο CoV μετρά τον κίνδυνο ανά μονάδα αναμενόμενης απόδοσης (ή, για ενεργειακά χαρτοφυλάκια, ανά μονάδα μέσης παραγωγής/σφάλματος), άρα βοηθά στην ιεράρχηση επιλογών με βάση την αποδοτικότητα του ρίσκου. Είναι ιδιαίτερα χρήσιμος για:

- συγκρίσεις μεταξύ περιουσιακών στοιχείων ή πάρκων με διαφορετικά επίπεδα μέσης απόδοσης / παραγωγής,
- αξιολόγηση της επίδρασης της διαφοροποίησης—χαμηλότερες συσχετίσεις μειώνουν τη συνολική διασπορά και κατ'επέκταση τον CoV και
- επιλογή πιο «σταθερών» χαρτοφυλακίων που περιορίζουν επιχειρησιακούς κινδύνους (π.χ. πρόστιμα από σφάλματα πρόβλεψης στην ενέργεια). Εν συντομία, ο CoV συμπυκνώνει σε έναν αριθμό τον συμβιβασμό (trade-off) μεταξύ διακύμανσης και μέσου, καθιστώντας τον θεμελιώδες εργαλείο για τεκμηριωμένες αποφάσεις. [20]

Ένα άλλο μέτρο αξιολόγησης απόδοσης των χαρτοφυλακίων που συνήθως χρησιμοποιείται είναι η εντροπία, όπως αυτή έχει ήδη αναπτυχθεί στην ενότητα 2.

Η χρήση του CoV παρέχει μια πολυδιάστατη εικόνα της ποιότητας ενός χαρτοφυλακίου, επιτρέποντας την εκτίμηση τόσο της προβλεψιμότητας του, όσο και του ρίσκου που επιφέρει μια ενδεχόμενη επένδυση ή προσθήκη σε αυτό.

4.4 Μεθοδολογία Εύρεσης Βέλτιστων Στρατηγικών για το Χτίσιμο Χαρτοφυλακίων

Το χτίσιμο ενός βέλτιστου χαρτοφυλακίου είναι μια διαδικασία που απαιτεί εκ των προτέρων επίγνωση των βασικών στοιχείων που καθορίζουν τις συστάδες, και επομένως τα πάρκα που ανήκουν σε αυτές. Στην ιδανική περίπτωση που ο αριθμός των πάρκων είναι περιορισμένος, μια θεωρητική αλλά ανορθόδοξη μεθοδολογία ανεύρεσης του βέλτιστου χαρτοφυλακίου θα ήταν οι συνεχείς δοκιμές των μικρών σε αριθμό συνδυασμών από πάρκα και η μεταξύ τους σύγκριση, ώστε να εξαχθεί ο συνδυασμός εκείνος που αποφέρει τον ελάχιστο CoV, δηλαδή τη μικρότερη δυνατή συνδιακύμανση των πάρκων. Ενώ αποτελεσματική, μια τέτοια μεθοδολογία είναι

εύκολο να κατανοηθεί πως είναι αδύνατον να γενικευθεί, όταν τα διαθέσιμα πάρκα αυξηθούν πολύ σε αριθμό, καθώς η πολυπλοκότητα θα αυξηθεί εκθετικά.

Ακόμα όμως και στην περίπτωση που το πρόβλημα αυτό μπορούσε να αντιμετωπιστεί, ο βασικός άξονας του ζητήματος είναι η εξέταση του τι θα συμβεί στη μέση περίπτωση των διαφόρων συνδυασμών. Με άλλα λόγια, σημασία για έναν επενδυτή και έναν έμπορο δεν έχει η εύρεση του τέλει χαρτοφυλακίου από ολόκληρο το σύνολο δεδομένων, αλλά η κατά δύναμιν μέγιστη βελτίωση της απόδοσης του μέσου τυχαίου χαρτοφυλακίου, που θα αποτελείται από πάρκα διαφορετικών συστάδων. Με άλλα λόγια, δεν έχει σημασία επομένως η ακριβής εύρεση των κατάλληλων πάρκων και ποιά είναι αυτά που ελαχιστοποιούν καθολικώς τον CoV και το ρίσκο, αλλά η όσο το δυνατόν ακριβέστερη εύρεση των κατάλληλων συνδυασμών από τυχαία πάρκα, τα οποία ανταποκρίνονται σε διαφορετικά προφίλ συμπεριφοράς, και που σε μία τυχαία μέση περίπτωση μπορούν να επιτύχουν τον ελάχιστο CoV.

Η μέθοδος που θα χρησιμοποιηθεί στο παρόν έργο είναι οι Προσομοιώσεις Monte Carlo (Monte Carlo Simulations - MCS). Αναπτύχθηκε από τον Πολωνό μαθηματικό Stanislaw Ulam στα τέλη της δεκαετίας του 1940.

Ο θεμελιώδης στόχος των Προσομοιώσεων είναι η εκτίμηση ενός αποτελέσματος, δεδομένης της ύπαρξης πολλών τυχαίων και απρόβλεπτων μεταβλητών που μπορούν να επηρεάσουν το αποτέλεσμα. Τέτοια παραδείγματα είναι η έρευνα για την εύρεση του βέλτιστου φαρμάκου για την ίαση πολλαπλών ασθενειών ή η πρόβλεψη του μέσου όρου ύψους των ανθρώπων σε μια χώρα. Αυτό επιτυγχάνεται με την εκτέλεση μεγάλου αριθμού προσομοιώσεων και μετέπειτα καταγραφής των αποτελεσμάτων. Πάνω στο προαναφερθέν παράδειγμα της πρόβλεψης του μέσου όρου ύψους, επιλέγοντας έναν μικρό αριθμό προσομοιώσεων, με άλλα λόγια κάνοντας μικρή δειγματοληψία ανθρώπων, ελλοχεύει ο κίνδυνος οι τυχαίες επιλογές μας να είναι συμπτωματικά μεροληπτικές (biased). Αυτό στο παράδειγμα μεταφράζεται ως η συμπτωματική επιλογή πολύ παραπάνω ψηλών ανθρώπων από ότι κοντών, που ενδεχομένως θα επηρεάσει έντονα την εκτίμηση. [21]

Οι MCS έχουν μεγάλη ποικιλία εφαρμογών, όπως στους τομείς:

- των χρηματοοικονομικών, για την πρόβλεψη με τη μέγιστη δυνατή ακρίβεια των τιμών μετοχών,
- της φαρμακευτικής, για την εκτίμηση της πιθανότητας καταπολέμησης μιας ασθένειας κατόπιν κατανάλωσης ενός φαρμάκου,
- της φυσικής, για την ασφαλή εκτίμηση του αποτελέσματος ενός πειράματος που δεν μπορεί να επιλυθεί με υπολογιστικούς τρόπους.

Επιστρέφοντας στο περιβάλλον της εργασίας, είναι φανερό η σημασία που θα έχουν οι MCS στην εκτίμηση της βέλτιστης στρατηγικής χτίσιμου ενεργειακών χαρτοφυλακίων. Εκκινώντας από ένα αρχικό χαρτοφυλάκιο με έναν προκαθορισμένο αριθμό πάρκων μιας συγκεκριμένης συστάδας, είναι εφικτό έπειτα από μεγάλο αριθμό προσομοιώσεων να εκτιμηθεί η μέση τιμή CoV που μπορεί να επιτευχθεί κατόπιν προσθήκης ενός επίσης προκαθορισμένου αριθμού πάρκων μιας συστάδας.

Διάταξη του πειράματος της εργασίας:

1. Επιλέγεται μία συστάδα και ένας αριθμός πάρκων αυτής της συστάδας και προστίθενται σε ένα χαρτοφυλάκιο, το οποίο θα ονομαστεί χαρτοφυλάκιο βάσης (baseline portfolio). Τα πάρκα που επιλέγονται είναι τυχαία.
2. Επιλέγεται ένας αριθμός k πάρκων που θα προστεθούν στο χαρτοφυλάκιο βάσης.
3. Στη συνέχεια, οι MCS εκτελούνται για μεγάλο αριθμό επαναλήψεων (iterations). Ο αριθμός των επαναλήψεων ορίζεται σε αυτό το βήμα. Για κάθε μια συστάδα ξεχωριστά, προστίθενται τα k πάρκα στο χαρτοφυλάκιο βάσης και για κάθε έναν πιθανό συνδυασμό πάρκων που προστίθενται, καταγράφεται η επίδραση του συνδυασμού στον CoV του χαρτοφυλακίου βάσης, είτε αυτή είναι θετική είτε αρνητική. Ενδέχεται ο CoV να αυξηθεί, αν προστεθούν πάρκα που δεν επιτυγχάνουν καλή διαφοροποίηση.
4. Για κάθε συστάδα προσθήκης και κάθε συνδυασμό, υπολογίζονται οι εξής ποσότητες:
 - Πιθανότητα Βελτίωσης του CoV (Improvement Probability, `prob_improve` στις μετέπειτα προσομοιώσεις), η οποία πρακτικά είναι το ποσοστό των επαναλήψεων των MCS που επέφεραν μείωση στον CoV του χαρτοφυλακίου βάσης

- Μέση Τιμή CoV κατόπιν προσθηκών (Mean CoV), η οποία είναι ο μέσος όρος όλων των τελικών τιμών του CoV στο νέο εμπλουτισμένο χαρτοφυλάκιο που καταγράφηκαν στις επαναλήψεις των MCS
- Μέση Τιμή Μεταβολών του CoV κατόπιν προσθηκών (Mean CoV Change ή ΔCoV Mean), η οποία είναι όπως λέει και το όνομα ο μέσος όρος όλων των μεταβολών του CoV που καταγράφηκαν στις επαναλήψεις των MCS
- Δείκτες εκατοστημορίων Q_{05} , Q_{95} για τη μείωση του CoV

Τα εκατοστημόρια (percentiles) περιγράφουν σε ποιο σημείο “κόβεται” η κατανομή των τιμών. Αν όλες οι παρατηρήσεις μείωσης ή αύξησης του CoV μετατραπούν σε μια κατανομή της οποίας υπολογίζονται οι πιθανότητες να επιλεχθεί παρατήρηση “κάτω” ή “πάνω” από ένα προεπιλεγμένο όριο, τότε το πεμπτο εκατοστημόριο (Q_{05}) είναι η τιμή κάτω από την οποία βρίσκεται το 5% των παρατηρήσεων· αντίστοιχα το ενενηκοστό πεμπτο (Q_{95}) είναι η τιμή κάτω από την οποία βρίσκεται το 95% των παρατηρήσεων (άρα μόνο 5% είναι μεγαλύτερες). Στην πράξη, η συνάρτηση που περιγράφει τις τυχαίες μεταβολές του CoV, καθώς είναι μια μεταβλητή με πάρα πολλές παρατηρήσεις, ουσιαστικά θα μοιάζει με την Γκαουσιανή Κανονική Κατανομή. Το διάστημα $[Q_{05}, Q_{95}]$ ονομάζεται συχνά «κεντρικό 90%» γιατί περίπου 9 στις 10 εκβάσεις πέφτουν εκεί μέσα. Στην περίπτωση που αυτοί οι δείκτες μετρούν την μείωση του CoV, το πεμπτο εκατοστημόριο δείχνει πως 5% όλων των δοκιμών απέφερε μείωση του CoV μεγαλύτερη από την τιμή του δείκτη. Αντίστροφα, το ενενηκοστό πέμπτο τεταρτημόριο δείχνει πως μόνο 5% των μεταβλητών απέφερε μικρότερη μείωση του CoV από τον δείκτη.

Φαίνεται επομένως πως αν ο Q_{95} έχει αρνητική τιμή, σημαίνει πως κατά 95% πιθανότητα, προσθήκες οποιουδήποτε συνδυασμού πάρκων σε οποιονδήποτε ήδη υπάρχοντα συνδυασμό πάρκων στο αρχικό χαρτοφυλάκιο θα επιφέρει μείωση του CoV. Πιθανοτικά μιλώντας, αυτό σημαίνει πρακτικά εγγυημένη μείωση, που είναι άλλωστε και το επιθυμητό αποτέλεσμα. Αντιθέτως, αν η απόσταση του διαστήματος $[Q_{05}, Q_{95}]$ είναι σχετικά μεγάλη και οι δείκτες έχουν αντίθετα πρόσημα, η προσθήκη πάρκου ή πάρκων από την εν λόγω συστάδα είναι μια ριψοκίνδυνη κίνηση, που έχει μεγάλες πιθανότητες να είναι ασύμφορη.

Στην πρώτη υποενοότητα, λαμβάνεται ως αφετηρία μια από τις τέσσερις συστάδες, εκκινώντας από 4 πάρκα που θα προστεθούν στη βάση του χαρτοφυλακίου. Στο επόμενο βήμα, ελέγχεται ποιες συστάδες επηρεάζουν θετικά ή αρνητικά το αρχικό χαρτοφυλάκιο, επομένως εμφανίζεται το προφίλ πάρκων που είναι πιο ωφέλιμο να προστεθεί σε κάθε αρχικό χαρτοφυλάκιο. Σε κάθε αρχικό χαρτοφυλάκιο, η πρώτη εκδοχή του πειράματος απαιτεί την προσθήκη 1 μόνο πάρκου στο βασικό χαρτοφυλάκιο, ενώ για τη δεύτερη εκδοχή προστίθενται 4. Το πείραμα θα εκτελεστεί 1000 φορές, ελέγχοντας τις μέσες τιμές του CoV κατόπιν των προσθηκών, καθώς και οι ποσοστιαίες μεταβολές τους. Επίσης χρήσιμο είναι να υπολογιστούν τα εκατοστημόρια για κάθε συνδυασμό, ώστε να υπάρχει εποπτεία των τιμών μείωσης του CoV μεταξύ των οποίων κυμαίνεται η συντριπτική πλειοψηφία των επαναλήψεων.

Κεφάλαιο 5 : Μελέτη Περίπτωσης

Για την εκτέλεση του πειράματος, χρησιμοποιήθηκε το προγραμματιστικό περιβάλλον του Kaggle, και η γλώσσα προγραμματισμού που επιλέχθηκε ήταν η Python, λόγω της ευκολίας στη χρήση που παρουσιάζει και της διαδεδομένης χρήσης της στον επιστημονικό και ερευνητικό κλάδο.

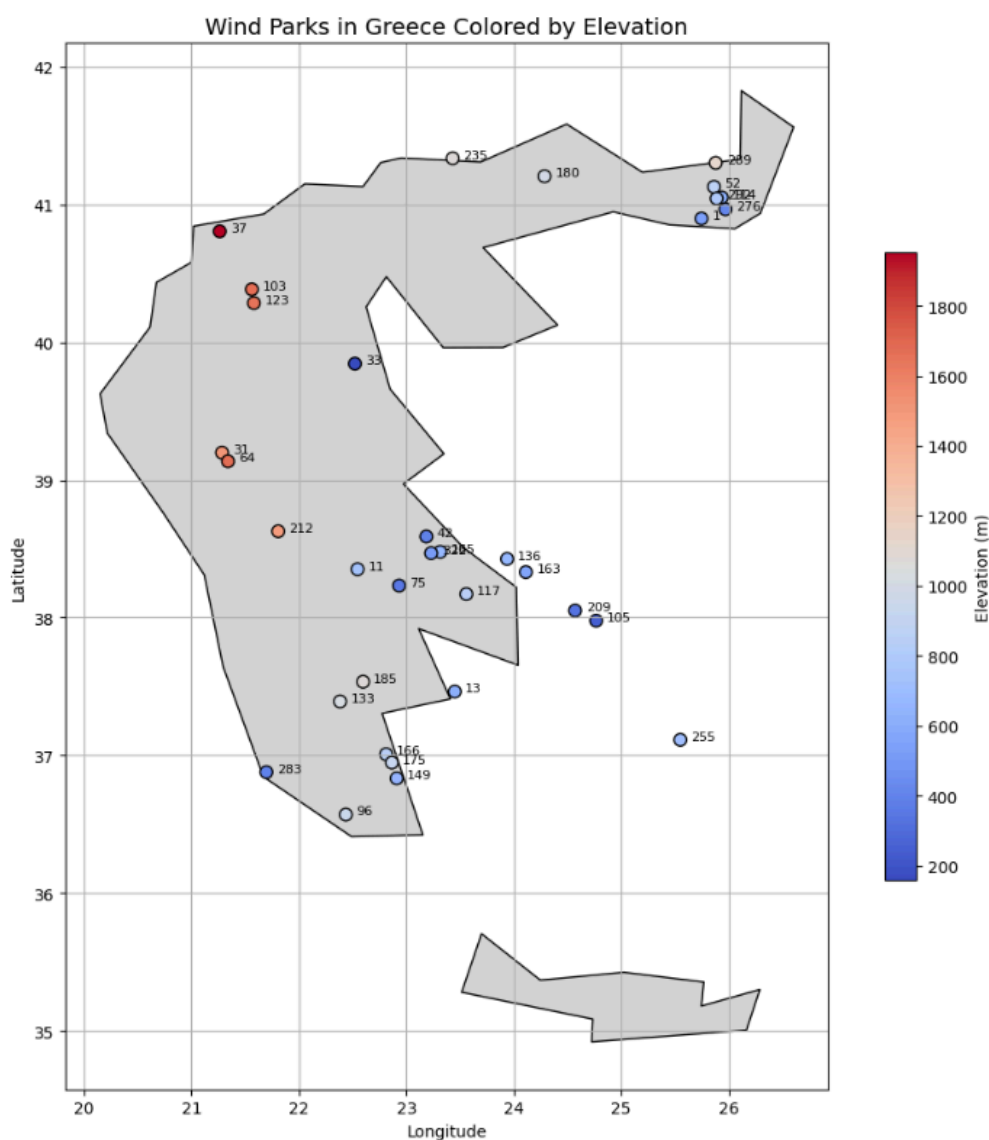
5.1 Παρουσίαση Δεδομένων

Τα διαθέσιμα δεδομένα για την παρούσα διπλωματική είναι τα δεδομένα ταχυτήτων ανέμων σε 34 διαφορετικά αιολικά πάρκα. Επίσης διαθέσιμα είναι τα γεωγραφικά μήκη, γεωγραφικό πλάτη, καθώς και τα υψόμετρα στο οποία βρίσκονται τα εκάστοτε πάρκα

Οι μετρήσεις των ταχυτήτων έγιναν εντός ενός διαστήματος 3 ετών, μεταξύ Φεβρουαρίου 2022 και Μαρτίου 2025, με λίγους μήνες κενών όπου δεν καταγράφηκαν αιολικά δεδομένα.

Κατασκευάζεται αρχικά ένας πρόχειρος χάρτης της Ελλάδας, δείχνοντας τις τοποθεσίες και τα υψόμετρα του κάθε πάρκου, βάσει των δοσμένων αριθμών για τα γεωγραφικά πλάτη, μήκη και υψόμετρα.

Έχουμε λοιπόν την εξής εικόνα:



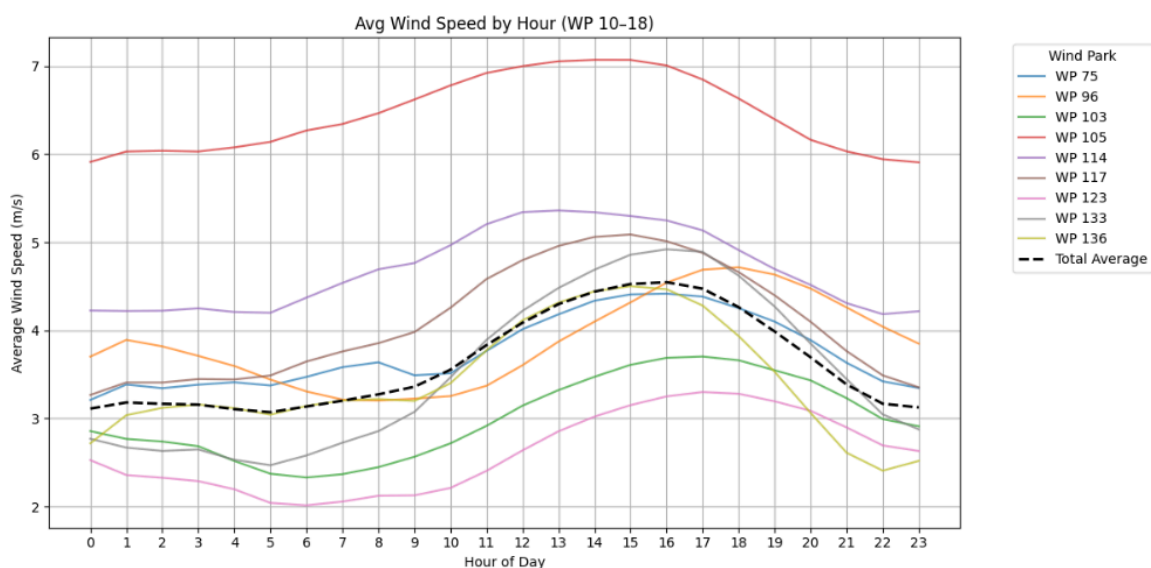
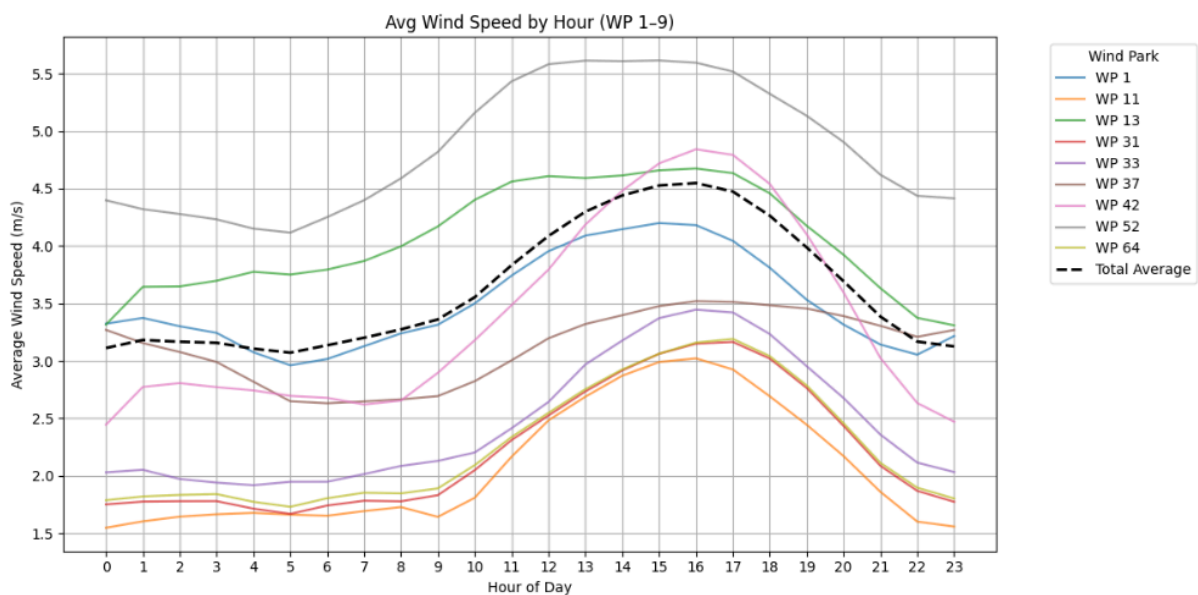
Εικόνα 13: Γεωγραφική τοποθεσία των πάρκων.

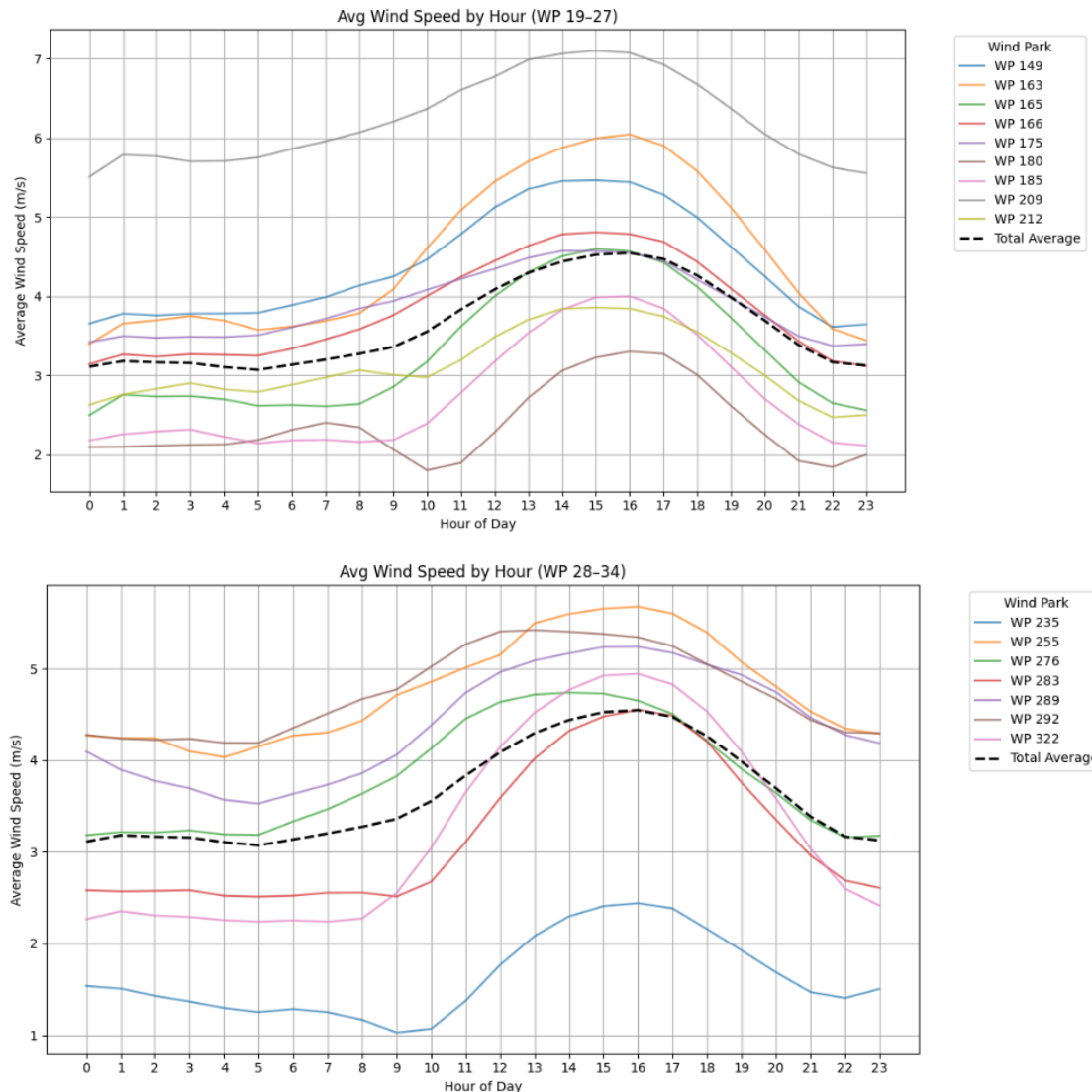
Πολλά από τα 34 πάρκα είναι απομονωμένα σε σχέση με τα υπόλοιπα, αλλά υπάρχουν ορισμένες περιοχές με πολλά αιολικά πάρκα συγκεντρωμένα σε μικρή ακτίνα. Μεγάλος αριθμός παρατηρείται στους Νομούς Βοιωτίας, Φωκίδας και Φθιώτιδας (11, 42, 75, 117, 165, 322), νομοί με έντονη δραστηριότητα στον τομέα της αιολικής ενέργειας. Επίσης 3 από τα πάρκα (136, 163, 209) βρίσκονται στην Εύβοια, που μαζί με τους προαναφερθέντες νομούς, έχουν πολύ μεγάλο ποσοστό των συνολικών εγκατεστημένων ανεμογεννητριών σε ολόκληρη τη χώρα. 4 από τα πάρκα (96, 149, 166, 175) βρίσκονται στο Νομό Λακωνίας, επίσης πρόσφορο έδαφος για εκμετάλλευση αιολικής ενέργειας λόγω του πλήθους οροσειρών.

Πολύ υψηλός αριθμός πάρκων (6, τα 1, 52, 114, 276, 289, 292) βρίσκεται επίσης στον Νομό Έβρου, περιοχή που τα τελευταία χρόνια βλέπει ραγδαίες εξελίξεις στον τομέα της αιολικής ενέργειας. Λόγω της τοποθεσίας κοντά στα σύνορα, η εγκατάσταση πολυπληθών ανεμογεννητριών είναι μια λογική από επενδυτική σκοπιά κίνηση, καθώς με αυτό τον τρόπο διευκολύνεται αρκετά η εξαγωγή ενέργειας σε γειτονικές χώρες.

Παρατηρούνται επίσης 6 πάρκα (31, 37, 64, 103, 123, 212) στην ορεινή ραχοκοκαλιά της Ελλάδας, πολλά από τα οποία θα φανεί στη συνέχεια πως παρουσιάζουν μεγάλες ομοιότητες μεταξύ τους ως προς τη συμπεριφορά.

Έχοντας δει τις τοποθεσίες των πάρκων πάνω στο χάρτη, υπολογίζονται οι μέσες τιμές των ταχυτήτων των ανέμων για κάθε ώρα της ημέρας, επομένως προκύπτουν 24 παρατηρήσεις:





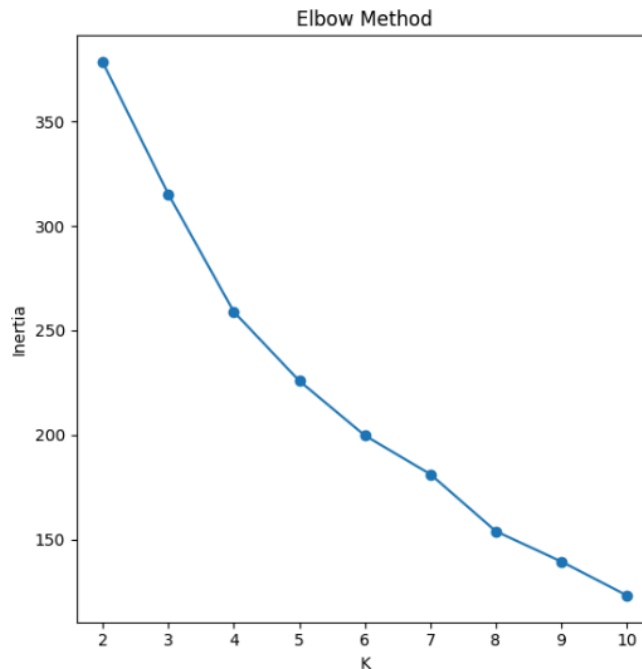
Εικόνα 14: Μέσες τιμές ταχύτητας για όλα τα πάρκα.

Με μια πρώτη ματιά, γίνεται να εντοπιστούν ορισμένα πάρκα με ιδιαιτερότητες:

- Τα πάρκα 105 και 209 έχουν πανομοιότυπη συμπεριφορά σε μηνιαίο και ωριαίο επίπεδο, αλλά οι ταχύτητες κυμαίνονται κατά μέσο όρο σε πολύ υψηλότερα επίπεδα από τα άλλα πάρκα. Αναζητώντας τα στον χάρτη, φαίνεται πως είναι πάρκα που βρίσκονται γεωγραφικά πολύ κοντά, στη Νότια Εύβοια και στην Άνδρο αντίστοιχα. Οι ταχύτητές τους επομένως επηρεάζονται σίγουρα από τις ιδιαίτερες κλιματικές συνθήκες της θάλασσας.
- Επιπλέον, ορισμένα πάρκα έχουν πτώση στη μέση ταχύτητα των ανέμων τους για τις πρωινές ώρες, κυρίως μεταξύ 07:00-11:00. Τέτοια πάρκα είναι τα 37, 96, 103 και 123, και σε μικρότερο βαθμό το 180. Τρία από τα τέσσερα αυτά πάρκα βρίσκονται στη Βόρεια Ελλάδα, στην περιοχή της Κοζάνης και της Φλώρινας, ενώ το τέταρτο βρίσκεται στην Πύλο. Γενικά αυτή η συμπεριφορά παρατηρείται σε μεγάλα υψόμετρα, άνω των 1000 μέτρων. Παρατηρούνται επομένως και εδώ μικροκλιματικές ομοιότητες στα κοντινά πάρκα.
Τα ίδια αυτά πάρκα έχουν επίσης αύξηση των ταχυτήτων τους κατά τις απογευματινές-νυχτερινές ώρες, από τις 18:00 και έπειτα. Αυτή η παρατήρηση θα φανεί χρήσιμη αργότερα.
- Μερικά πάρκα, όπως για παράδειγμα τα 11, 180 και 235, παρουσιάζουν πτώσεις στις ταχύτητες των ανέμων για τις πρωινές ώρες μεταξύ 08:00-10:00 περίπου.
- Άλλα πάρκα, όπως το 165 και το 322, για τις ίδιες ώρες παρουσιάζουν κατά μέσο όρο έντονη αύξηση στις ταχύτητες των ανέμων.

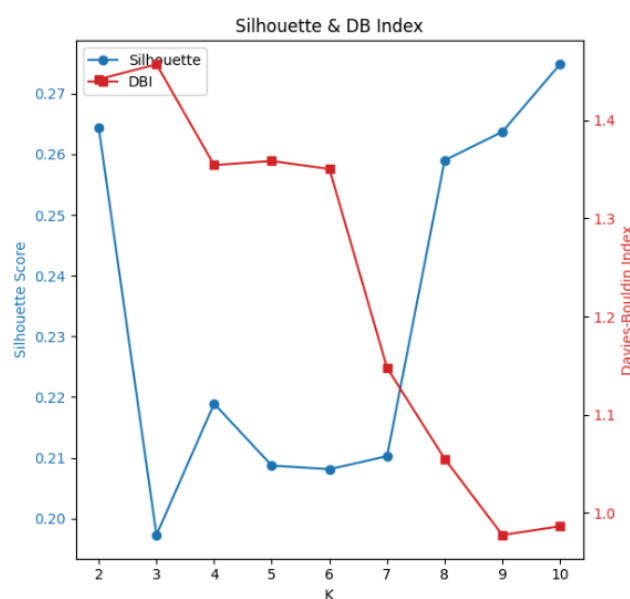
5.2 Επιλογή αριθμού συστάδων

Έχοντας προετοιμάσει όλα τα δεδομένα, αρχικά γίνεται μια προεργασία, σχεδιάζοντας τα διαγράμματα της μεθόδου αγκώνα, του σκορ σιλουέτας και του δείκτη Davies-Bouldin, ώστε να φανεί ο αριθμός k που συμφέρει να επιλεγεί, για να υπάρξει βέλτιστη συσταδοποίηση.



Εικόνα 15: Elbow Method για επιλογή του βέλτιστου αριθμού συστάδων k .

Με μια πρώτη ματιά, φαίνεται πως για $k = 4, 7$ ή 8 υπάρχουν σημεία όπου το WCSS μειώνεται με αυξανόμενο ρυθμό. Παρ'όλα αυτά, δεν είναι πολύ ξεκάθαρα αυτά τα σημεία, επομένως θα χρησιμοποιηθούν και τα σκορ σιλουέτας σε συνδυασμό με τον δείκτη Davies-Bouldin, για να εντοπιστεί το k που συνδυάζει τα κατάλληλα χαρακτηριστικά.



Εικόνα 16: Δείκτες Σιλουέτας και Davies-Bouldin για το σύνολο δεδομένων

Από τις προαναφερθείσες τιμές 4, 7 και 8 παρατηρείται πως οι 4 συστάδες ταιριάζουν αρκετά στα δεδομένα. Οι τιμές 7 και 8 όμως από θεωρητική οπτική γωνία ταιριάζουν καλύτερα, διότι εγγυώνται χαμηλότερες τιμές του DBI. Ειδικά για $k = 8$, το σκορ σιλουέτας είναι μάλιστα υψηλότερο και από αυτό για $k = 4$, ακόμα και αν οι διαφορές είναι της τάξης του 10^{-2} .

Καθώς όμως υπάρχουν μόνο 34 πάρκα, μπορεί να θεωρηθεί υπερβολική αυστηρότητα να χωριστούν σε 8 ομάδες. Σε αυτό το σημείο θα γίνει ένας σιωπηρός συμβιβασμός και θα επιλεγεί ως αριθμός ομάδων το $k = 4$, καθώς εμπειρικά και διαισθητικά φαίνεται να είναι ο καταλληλότερος αριθμός.

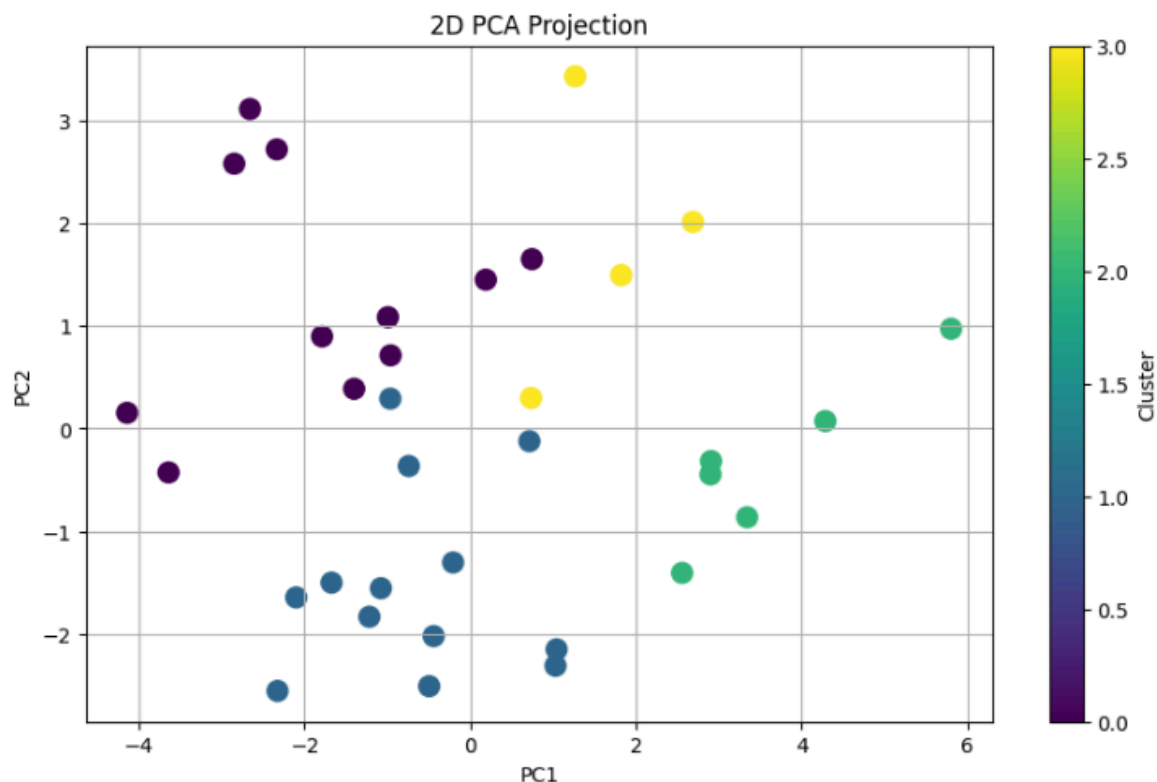
Φυσικά η αύξηση του k θα μειώσει το DBI, αλλά δε θα πρέπει για χάρη της απόλυτης “καθαρότητας” του συνόλου να αυξηθεί κατά πολύ η πρακτική πολυπλοκότητα, δημιουργώντας 8 ομάδες. Επιπρόσθετα, το DBI κυμαίνεται σε σχετικά υψηλές τιμές ακόμα και για μεγάλο αριθμό ομάδων, το οποίο είναι αναμενόμενο για δεδομένα πραγματικού κόσμου. Υπάρχουν σίγουρα επικαλύψεις και γενικότερα ένα “χάος” που περιγράφει απόλυτα το σύνολο δεδομένων.

Για $k = 4$, το DBI υπολογίστηκε ίσο με 1.3544, διακύμανση που μπορεί να θεωρηθεί αποδεκτή.

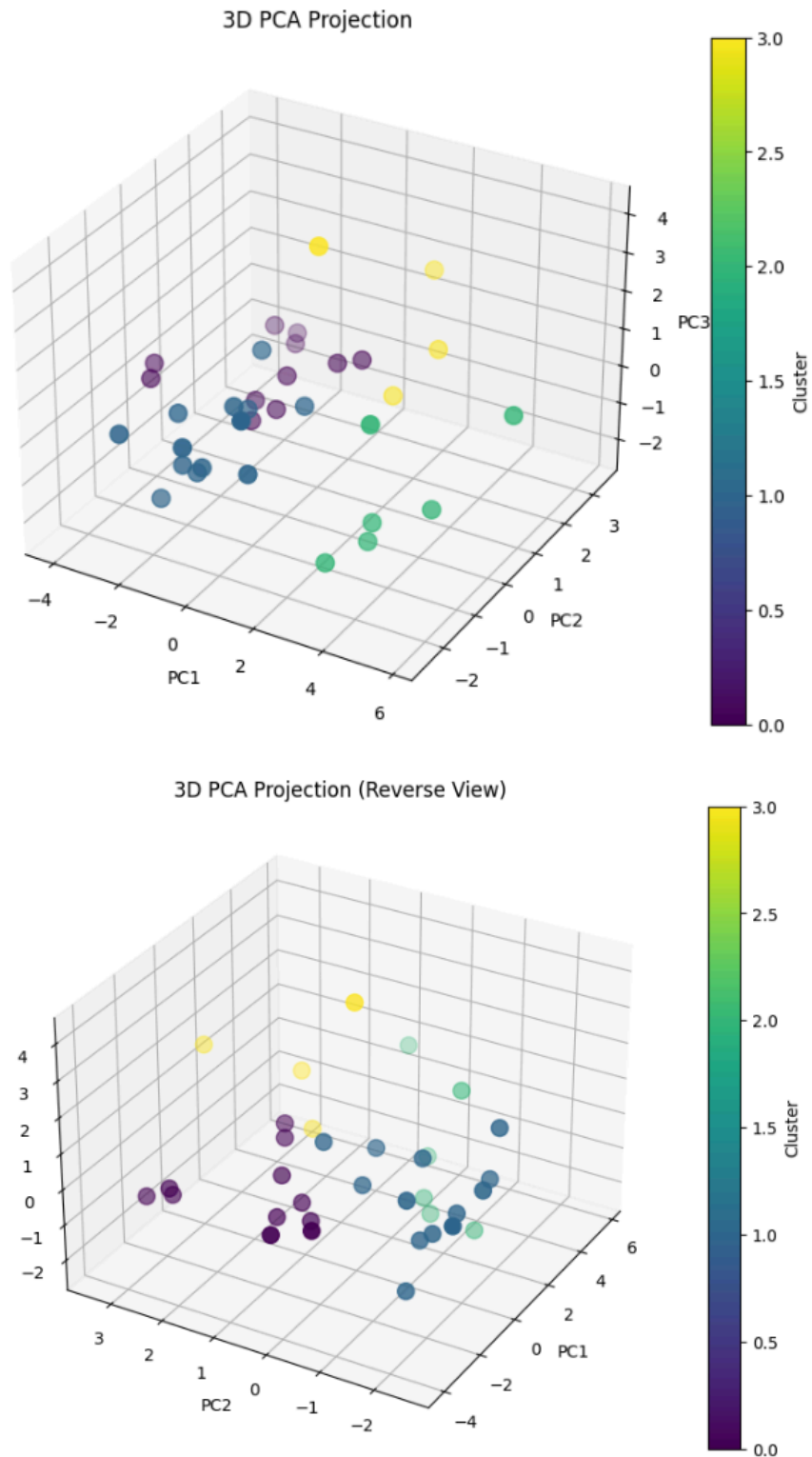
5.3 Εφαρμογή Αλγορίθμου Συσταδοποίησης k-means και PCA

Επόμενο βήμα είναι η εφαρμογή του αλγορίθμου k-means και η μετέπειτα ανάλυση βασικών συνιστωσών, ώστε να γίνει εφικτή η απεικόνιση των συστάδων. Μέσω αυτού προκύπτει ποιο στοιχείο έχει μεγαλύτερο βάρος, επομένως παίζει σημαντικότερο ρόλο, στο διαχωρισμό των πάρκων.

Τα πρώτα διαγράμματα που θα απεικονιστούν είναι των πάρκων στο νέο χώρο κατόπιν εφαρμογής του PCA. Θα απεικονιστούν τα πάρκα τόσο σε διαγράμματα δύο, όσο και τριών, διαστάσεων, με σκοπό να καταδειχθεί ο χωρικός διαχωρισμός των ομάδων.



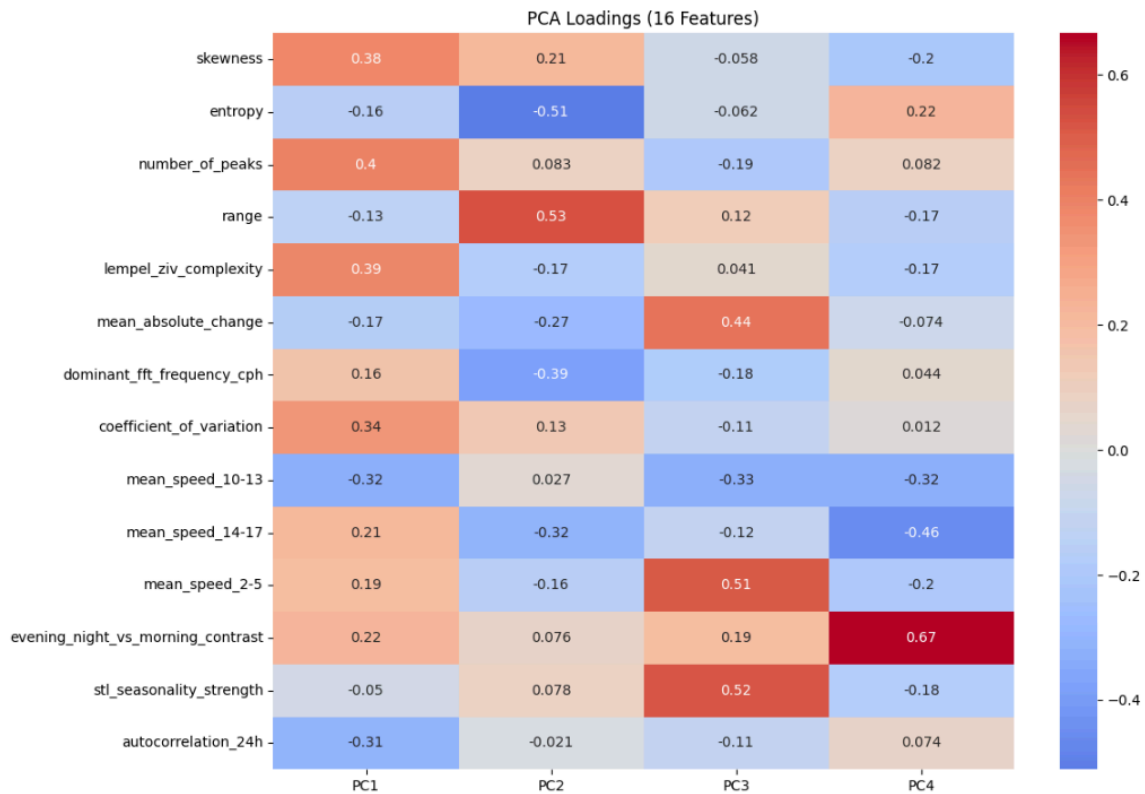
Εικόνα 17: Διοδίστατη αναπαράσταση των πάρκων, με κάθε συστάδα να χαρακτηρίζεται από διαφορετικό χρώμα.



Εικόνα 18: Τρισδιάστατη απεικόνιση των πάρκων.

Η τρισδιάστατη απεικόνιση των συστάδων δείχνει με πιο ξεκάθαρο τρόπο πως πάρκα που στη δισδιάστατη απεικόνιση βρίσκονται φαινομενικά κοντά στα υπόλοιπα, όπως π.χ. τα πάρκα της τρίτης συστάδας, στην πραγματικότητα λόγω των πολλών διαστάσεων, βρίσκονται σε εντελώς άλλο σημείο του χώρου. Ήδη από τις παραπάνω εικόνες φαίνεται πώς οι συστάδες 0 και 1 βρίσκονται κοντά, και υπάρχει επικάλυψη κοντά στα όρια των 2 συστάδων. Αντίθετα, η συστάδα 3 φαίνεται ήδη αρκετά διαφορετική από τις υπόλοιπες, κάτι που θα φανεί

παραπάνω και στη συνέχεια. Επίσης, με προσεκτική παρατήρηση φαίνεται η αρχή των αξόνων να λειτουργεί σαν νοητό σύνορο διαχωρισμού των συστάδων.
Έχουν υπολογιστεί επίσης τα βάρη που έχουν τα features που χρησιμοποιήθηκαν, και φαίνονται στον πίνακα που ακολουθεί, τον πίνακα βαρών (Loading Chart).



Εικόνα 19: Πίνακας Βαρών των βασικών συνιστωσών.

Στην παράγραφο 4.2, αναφέρθηκε το ποσοστό της διασποράς, δηλαδή το ποσοστό της πληροφορίας, που θα διατηρηθεί. Ένα καλό ποσοστό είναι το 80-90% της αρχικής πληροφορίας, με χώρο για ευελιξία αναλόγως φυσικά των δεδομένων. Ένα τέτοιο ποσοστό εγγυάται πως ο επιλεγμένος αριθμός συνιστωσών δίνει πρακτικά την “πλήρη εικόνα”, με μικρές διαφοροποιήσεις, οι οποίες για χάρη της απλοποίησης είναι αποδεκτές.

Με διαδοχικούς ελέγχους προκύπτει πως κρατώντας τις 4 πρώτες συνιστώσες, επιτυγχάνεται μία συνολική διακύμανση 81,09%. Η 5η συνιστώσα προσφέρει μόνο ~5% επιπλέον διακύμανση, και κατόπιν ελέγχων τα διαγράμματα δεν άλλαξαν καθόλου. Επομένως για απλοποίηση, κρατώνται μόνο τις πρώτες τέσσερις, ακόμα και αν ο θεωρητικός στόχος του 80-90% καλύπτεται οριακά. Τα ποσοστά διακύμανσης που καλύπτονται φαίνονται στον ακόλουθο Πίνακα 2:

PCA Variance που καλύπτεται	
PC1	42,49%
PC2	20,92%
PC3	9,13%
PC4	8,55%

Ο πίνακας βαρών δίνει σημαντικές πληροφορίες, δείχνοντας ποιες μεταβλητές είναι πιο επιδραστικές στο διαχωρισμό των πάρκων στις ποικίλες κατηγορίες. Αναζητώνται εκείνες τις μεταβλητές, οι οποίες έχουν μεγάλη απόλυτη τιμή στο βάρος τους, ασχέτως αν ο καθαρός αριθμός είναι θετικός ή αρνητικός.

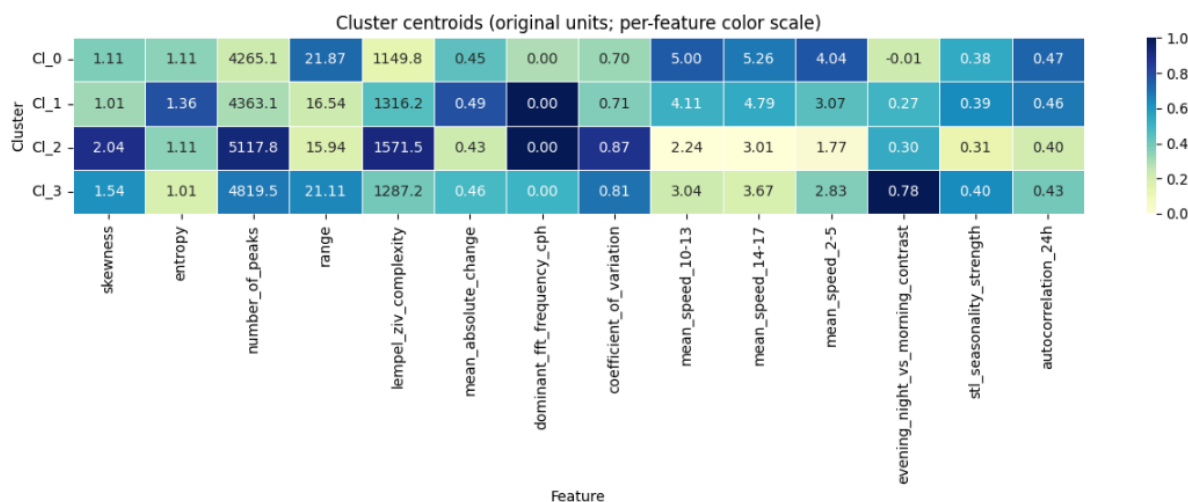
Εστιάζοντας στις πρώτες 3 μεταβλητές, διότι η τέταρτη και τελευταία δεν μπορεί να οπτικοποιηθεί, φαίνεται πως η PC3, που ευθύνεται κυρίως για την συστάδα 3, επηρεάζεται κυρίως από τη μέση κανονικοποιημένη ταχύτητα για τις πρωινές ώρες 02:00-05:00, καθώς και από την ένταση της εποχικότητας. Πράγματι, σε επόμενη ενότητα θα αποδειχθεί πως η συστάδα 3 αποτελείται από εκείνα τα πάρκα, στα οποία παρατηρούνται υψηλότερες απογευματινές-βραδινές ταχύτητες από ότι πρωινές.

Η PC1 επηρεάζεται κυρίως από την πολυπλοκότητα, τον αριθμό των κορυφών και την λοξότητα, ενώ η PC2 επηρεάζεται κατά βάση από το εύρος και την εντροπία, και μάλιστα από χαμηλές τιμές της. Οι 2 συστάδες με υψηλή την PC2, είναι πράγματι αυτές που καταδεικνύουν χαμηλή εντροπία.

Έχοντας ολοκληρώσει την συσταδοποίηση και έχοντας μια διαισθητική εκτίμηση ως προς τα χαρακτηριστικά που θα καθορίζουν την κάθε συστάδα, θα απεικονιστούν τα κέντρα της κάθε συστάδας, δείχνοντας τις τιμές γύρω από τις οποίες κυμαίνονται τα πάρκα της κάθε μίας για όλα τα επιλεγμένα στοιχεία. Άλλωστε το κέντρο υπολογίζεται μετρώντας τον μέσο όρο κάθε στοιχείου, δείχνοντας έτσι το μέσο προφίλ ενός πάρκου που ανήκει στην εκάστοτε συστάδα.

Τα κέντρα θα υπολογιστούν αν απλώς αντιστραφεί η κανονικοποίηση των αριθμών που είχε γίνει αρχικά ώστε να επιστραφούν οι κανονικοποιημένες τιμές στις αρχικές (original units). Αυτό συμβαίνει ώστε να υπάρχει φυσική σημασία στα αποτελέσματα.

Έτσι τα κέντρα για κάθε συστάδα είναι τα εξής:



Εικόνα 20: Κέντρα των συστάδων στις αρχικές τιμές τους.

Ενδιαφέρουσες παρατηρήσεις αποτελούν οι σχεδόν μηδενικές τιμές της κύριας συχνότητας, που υποδηλώνουν αντίστοιχα μεγάλη περίοδο. Μια σημαντική παρατήρηση είναι επίσης η πολύ μεγαλύτερη πολυπλοκότητα της τρίτης συστάδας σε σχέση με τις άλλες, ενώ η χαμηλότερη αυτοσυσχέτιση που παρατηρείται ανήκει στην ίδια συστάδα. Επομένως η τρίτη συστάδα θα είναι η πιο δύσκολη στην πρόβλεψη, γιατί θα είναι η πιο ασταθής και φτωχή μοτίβων. Στον αντίποδα, η πρώτη συστάδα έχει πολύ χαμηλούς δείκτες εντροπίας, πολυπλοκότητας και MAC, που μεταφράζεται ως ήπια συμπεριφορά, χωρίς ιδιαίτερες απότομες αλλαγές.

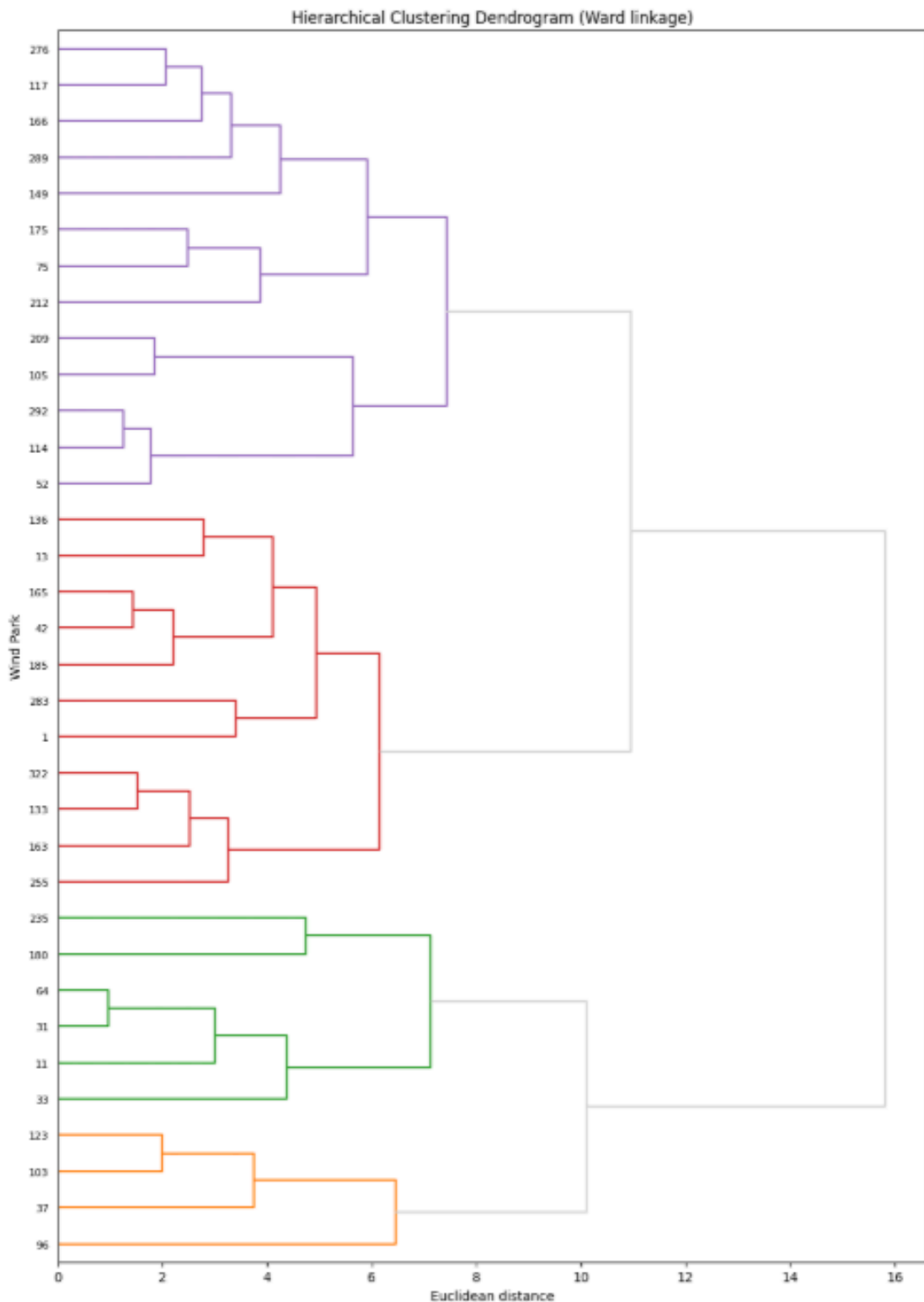
Παρατηρείται πως η συστάδα 2 έχει γενικά πολύ χαμηλές τιμές μέσης ταχύτητας ανέμων, ενώ η συστάδα 3 αντιθέτως έχει πολύ υψηλές, με τις νυχτερινές ταχύτητες μάλιστα να υπερτερούν σημαντικά των πρωινών. Η συστάδα 0 φαίνεται επίσης να αποτελείται από τα πάρκα που παρουσιάζουν την “τυπική” συμπεριφορά όσον αφορά τις ώρες μεγαλύτερης δραστηριότητας, δηλαδή τις πρωινές ώρες.

5.4 Εφαρμογή Αλγορίθμου Ιεραρχικής Συσταδοποίησης

Ο αλγόριθμος k-means είναι ένα πολύ χρήσιμο και αξιόπιστο εργαλείο συσταδοποίησης δεδομένων, και εκτιμάται πως τα αποτελέσματά του είναι αποδεκτά. Παρόλα αυτά, είναι εξίσου σημαντικό τα αποτελέσματα αυτά να συγκριθούν και με άλλους αλγορίθμους συσταδοποίησης, με σκοπό την ανάδειξη κάποιας νέας, βέλτιστης, συσταδοποίησης, που αποδίδει καλύτερα από τον k-means, αν φυσικά κάποια τέτοια υπάρχει. Επίσης η

εφαρμογή εναλλακτικών αλγορίθμων συσταδοποίησης μπορεί να βοηθήσει στην επιβεβαίωση της αρχικής επιλογής των αριθμών συστάδων. Αν παρατηρηθεί πως και οι άλλες μέθοδοι καταλήγουν σε παρόμοια συμπεράσματα για τις τέσσερις συστάδες που εικάζεται ότι είναι κατάλληλες, με μικρές αναμενόμενες παραλλαγές, μπορεί να εξαχθεί με μεγάλη βεβαιότητα μια τελική απόφαση για τον ουσιαστικό αριθμό ομάδων πάρκων με διαφορετικά προφίλ το ένα από το άλλο.

Για αυτό το λόγο, σε δεύτερη φάση εφαρμόζεται ο αλγόριθμος της ιεραρχικής συσταδοποίησης, ώστε να συγκριθούν τα αποτελέσματα. Εφαρμόζοντας τη μεθοδολογία που αναπτύχθηκε πρωτύτερα στην παράγραφο 3.1, προκύπτει το εξής δενδρόγραμμα:

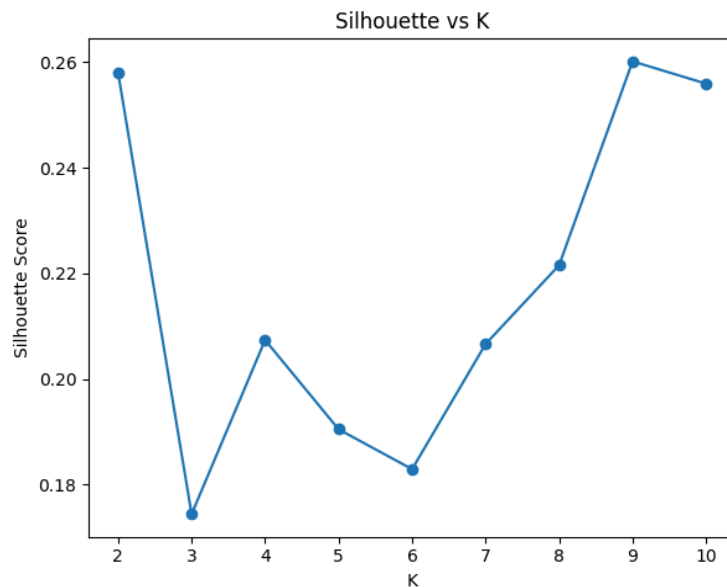


Εικόνα 21: Δενδρόγραμμα ομάδων σύμφωνα με την ιεραρχική συσταδοποίηση.

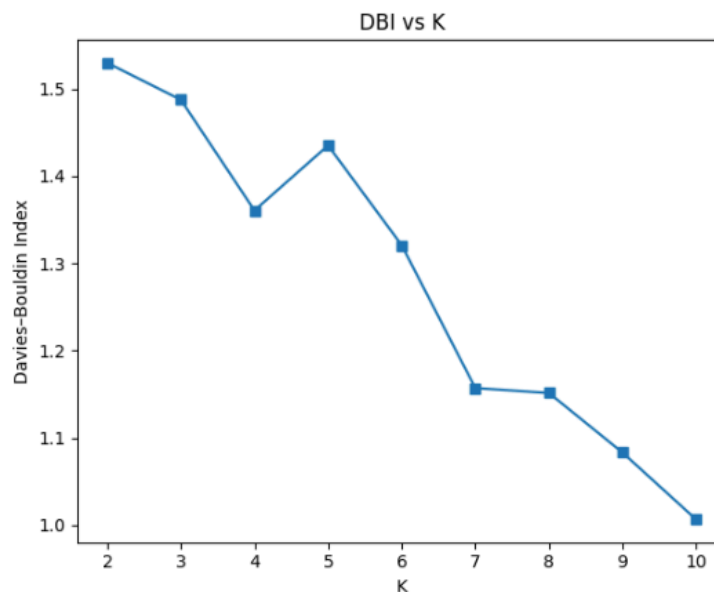
Το δενδρόγραμμα οπτικοποιεί με ξεκάθαρο τρόπο σε τι “ύψος” ενώνονται δύο κλαδιά του δένδρου, επομένως σε ποιο στάδιο δύο υποσυστάδες από όλες τις διαθέσιμες είναι οι πιο όμοιες. Τα πάρκα που αναμένεται να παρουσιάζουν τις μεγαλύτερες ομοιότητες, με μια πρώτη εκτίμηση λόγω της γεωγραφικής τους εγγύτητας, είναι πράγματι συνήθως από τα πρώτα που “ενώνονται”. Τέτοια παραδείγματα είναι τα ζευγάρια πάρκων 31 και 64, καθώς επίσης και τα 42, 165 και 103, 123.

Το δένδρόγραμμα επίσης είναι πάρα πολύ κατατοπιστικό στην προκειμένη περίπτωση, διότι δείχνει πως ένας πολύ κατάλληλος και λογικός αριθμός συστάδων είναι οι τέσσερις, διότι φαίνεται να είναι το πιο “φυσικό” σημείο τομής των δεδομένων. Ξεκινώντας από τα δεξιά, από το σημείο που οι συστάδες γίνονται 4, φαίνεται πως η απόσταση μέχρι τον επόμενο διαχωρισμό είναι αρκετά μεγαλύτερη από την απόσταση π.χ. μεταξύ του διαχωρισμού από 6 σε 7 ομάδες. [22]

Για επιβεβαίωση, υπολογίζονται και οι δείκτες αξιολόγησης συσταδοποίησης:



Εικόνα 22: Δείκτης σιλουέτας για διάφορα k κατά την ιεραρχική συσταδοποίηση.

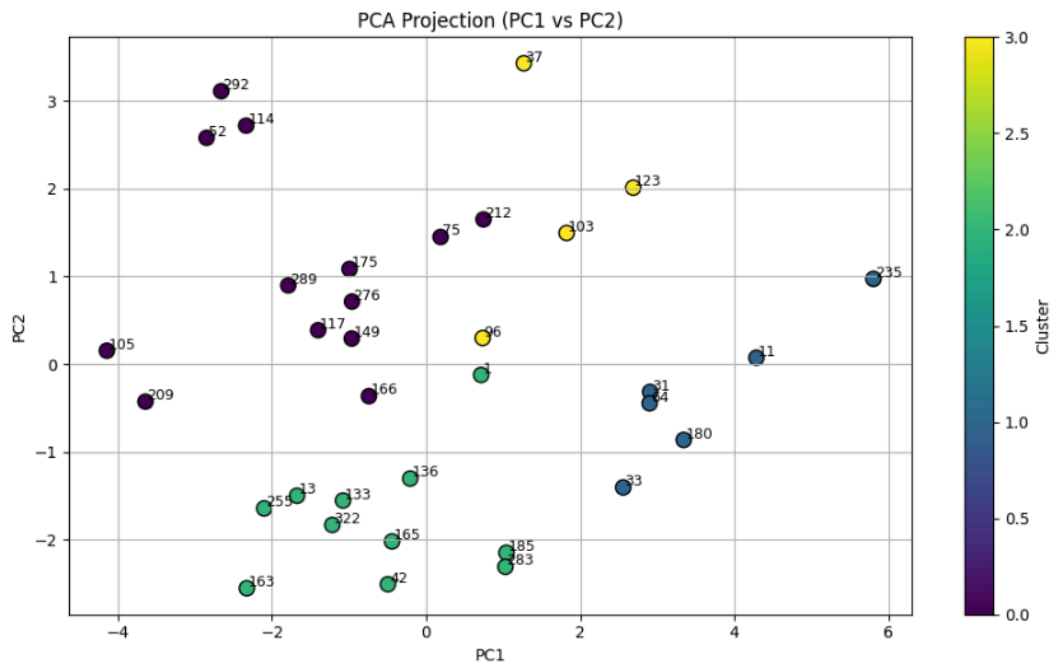


Εικόνα 23: Δείκτης Davies-Bouldin για διάφορα k κατά την ιεραρχική συσταδοποίηση.

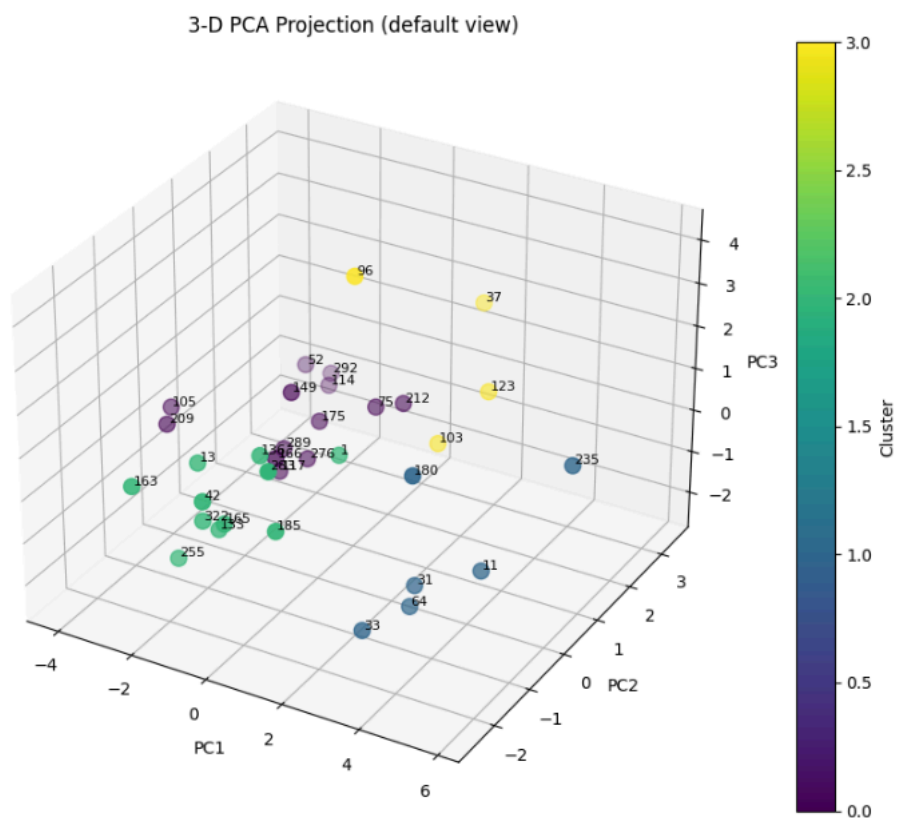
Παρατηρείται πως τοπικά το $k = 4$ παρουσιάζει ελάχιστο για το DBI και μέγιστο για το σκορ σιλουέτας, άρα ο ισχυρισμός ότι οι τέσσερις συστάδες είναι κατάλληλες επιβεβαιώνεται. Μάλιστα, οι επιλογές $k = 7$, $k = 8$ παρουσιάζουν οριακά το ίδιο σκορ σιλουέτας με το $k = 4$.

Έχοντας επιλέξει επομένως τον κατάλληλο αριθμό, ομαδοποιούνται εκ νέου τα πάρκα. Είναι σημαντικό να παρατηρηθούν και να σημειωθούν τυχόν διαφορές σε σχέση με το διάγραμμα PC της συσταδοποίησης k-means, διότι τα πάρκα αυτά θα είναι εκείνα που είναι στο “μεταίχμιο” μεταξύ ομάδων και δεν ανήκουν με ασφάλεια σε μια ξεκάθαρη συστάδα.

Τα διαγράμματα PC σε δύο και τρεις διαστάσεις είναι τα ακόλουθα:



Εικόνα 24: Δισδιάστατη αναπαράσταση των πάρκων ιεραρχικής συσταδοποίησης.

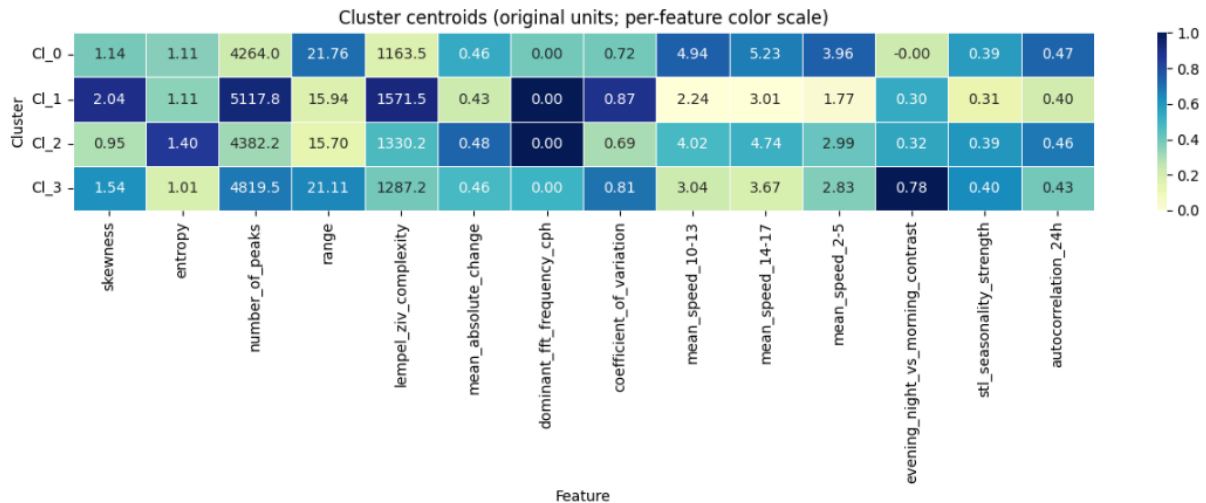


Εικόνα 25: Τρισδιάστατη αναπαράσταση των πάρκων.

Με μια πρώτη ματιά φαίνεται πως οι συστάδες έχουν μείνει σχεδόν ίδιες, επομένως υπάρχει ένας κορμός πάρκων που παρουσιάζουν αρκετές διαφορές μεταξύ τους ώστε να διαχωρίζονται με συνέπεια και χωρίς

παραλλαγές. Οι συστάδες 1 και 3 μάλιστα έχουν μείνει απaráλλαχτες. Με σύγκριση, καταγράφονται μόνο 2 πάρκα που άλλαξαν ομάδα, τα υπ αριθμόν 149, 166. Πράγματι, παρατηρώντας προσεκτικότερα την εικόνα 25 και συγκρίνοντας με την εικόνα 18, τα 2 αυτά πάρκα βρίσκονται κοντά σε άλλες συστάδες είτε σε μια πιθανώς “άσχετη” θέση στον πολυδιάστατο χώρο, που τα καθιστά ευάλωτα σε αναπηδήσεις μεταξύ ομάδων. Το γεγονός αυτό καθιστά την συσταδοποίηση ιδιαίτερα στιβαρή, διότι τα πάρκα που αναπηδούν από συστάδα σε συστάδα είναι μικρό ποσοστό του συνόλου.

Ακριβώς όπως και πριν, εκτυπώνονται τα κέντρα των συστάδων στις αρχικές τους τιμές:



Εικόνα 26: Αρχικές τιμές κέντρων των συστάδων για ιεραρχική συσταδοποίηση.

Παρατηρούνται μικρές διαφορές, οι οποίες φυσικά δικαιολογούνται με την προσθήκη ή αφαίρεση πάρκων από κάποιες συστάδες. Επίσης, ο αλγόριθμος αρίθμησε διαφορετικά τις συστάδες, εξ ου και η ανάποδη σειρά με την οποία εκτυπώνονται. Καθώς όμως οι εναλλαγές πάρκων από μία συστάδα σε κάποια άλλη είναι ελάχιστες, τα εξέχοντα χαρακτηριστικά της καθεμιάς παρέμειναν ως είχαν. Αναμενόμενα, τα κέντρα των συστάδων 2 και 3 μάλιστα δεν άλλαξαν καθόλου, καθώς τα ίδια πάρκα με την συσταδοποίηση κατά k-means είναι μέλη τους και κατά ιεραρχική συσταδοποίηση. Συνεχίζουμε με την εφαρμογή του αλγορίθμου DBSCAN, βλέποντας τι επιπλέον πληροφορίες μπορούν να αντληθούν από τα αποτελέσματά του.

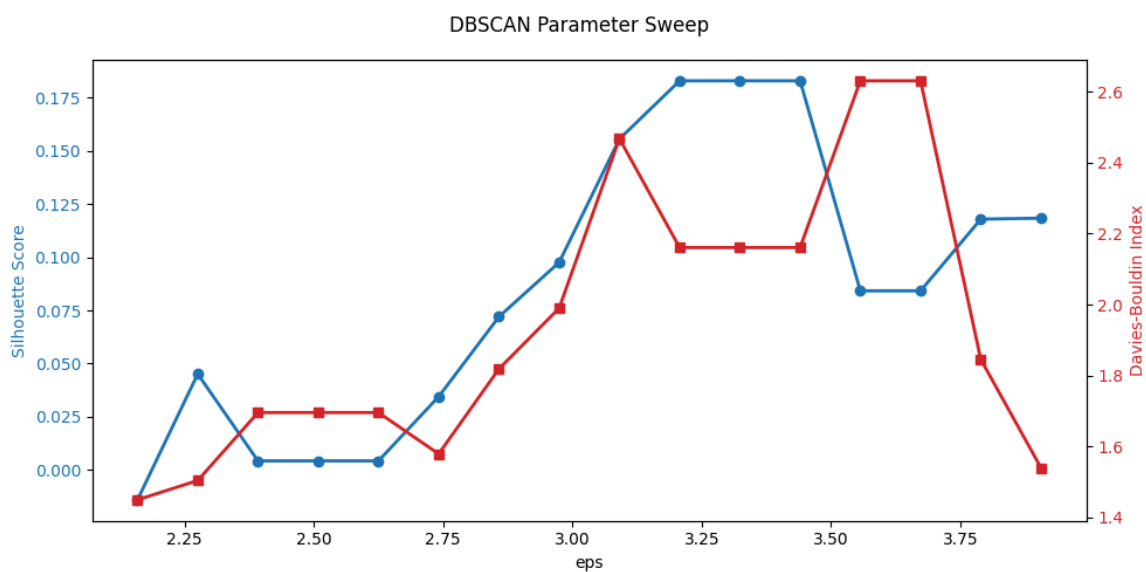
5.5 Εφαρμογή Αλγορίθμου DBSCAN

Ο τελευταίος αλγόριθμος που θα βοηθήσει στην συσταδοποίηση είναι ο DBSCAN, με τη βοήθεια του οποίου εντοπίζονται καλύτερα μη-κυκλικές δομές και πάρκα - “θόρυβος”, που δεν ανήκουν ικανοποιητικά σε κάποια συστάδα. Στόχος είναι η εύρεση παρόμοιων συστάδων με τις προηγούμενες δοκιμές με διαφορετικούς αλγόριθμους, όσο αυτό είναι εφικτό, καθώς και ο εντοπισμός πάρκων που χαρακτηρίζονται ως θόρυβος από τον αλγόριθμο. Αν τα πάρκα αυτά είναι λίγα σε αριθμό σε σχέση με τον συνολικό αριθμό, μπορεί με ασφάλεια να βγει το συμπέρασμα πως η συσταδοποίηση είναι ποιοτική και ανθεκτική.

Εφαρμόζοντας λοιπόν τον αλγόριθμο ώστε το ϵ να επιλέγεται αυτόματα και με βέλτιστο τρόπο, δεδομένου του `min_samples`, επιλέγοντας ως `min_samples` το 3, ώστε να αποφευχθεί η εμφάνιση πολλών συστάδων που αποτελούνται μόνο από ένα ή και δύο πάρκα:

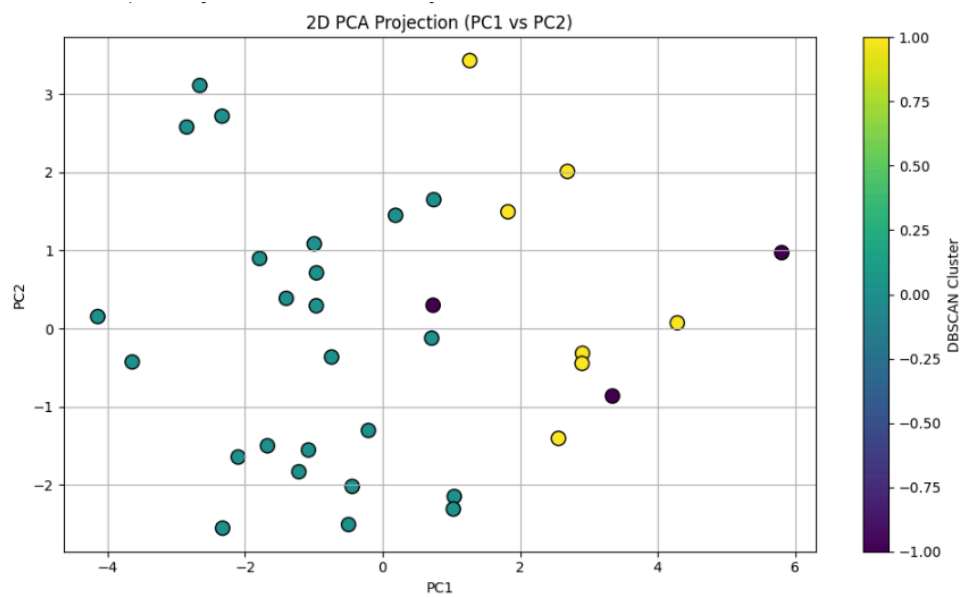
Selected eps value	3.2071 (min_samples = 3)
DBI	2.1607

Πίνακας 3: Αυτόματη Επιλογή του eps δεδομένου ενός `min_samples`.

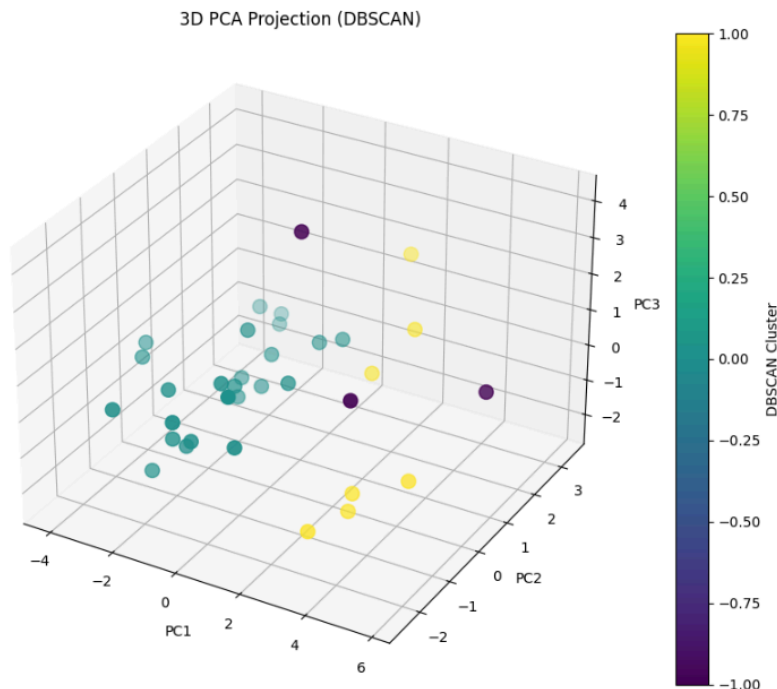


Εικόνα 27: Διάγραμμα βοήθειας επιλογής παραμέτρου ϵ , δεδομένου του ελάχιστου αριθμού πάρκων μιας συστάδας.

Παρακάτω ακολουθούν τα δισδιάστατα και τρισδιάστατα διαγράμματα των πάρκων και των συστάδων στα οποία ανήκουν:



Εικόνα 28: Δισδιάστατο διάγραμμα πάρκων κατόπιν DBSCAN.



Εικόνα 29: Τρισδιάστατο διάγραμμα πάρκων κατόπιν DBSCAN.

Κατά την εφαρμογή του αλγορίθμου DBSCAN, παρατηρήθηκαν διαφορετικά αποτελέσματα, τα οποία όμως σχετίζονται με τη φύση του αλγορίθμου. Παρατηρείται για άλλη μια φορά πως η αρχή των αξόνων, όπως και στις άλλες περιπτώσεις, λειτουργεί σαν διαχωριστικό πέρα από το οποίο δεν υπάρχουν εναλλαγές πάρκων από συστάδα σε συστάδα. Η ένωση των συστάδων από 4 σε 2 είναι δείγμα πως οι συστάδες δεν έχουν τέλειο διαχωρισμό, αλλά αυτό είναι αναμενόμενο λόγω της μικρής σχετικά απόστασης μεταξύ των πάρκων στον πολυδιάστατο χώρο. Ο αλγόριθμος DBSCAN μπορεί εύκολα να προσθέσει πάρκα σε μια κυκλική σχηματικά συστάδα, ακόμα και αν οι αποστάσεις μεταξύ των εξωτερικών πάρκων με των πυκνών σημείων είναι αρκετά μεγάλες. Εφόσον όμως οι προηγούμενες δύο μέθοδοι συσταδοποίησης κατέδειξαν εμφανώς πως ο λογικότερος αριθμός συστάδων θα είναι 4, μπορούμε να καταγράψουμε τα αποτελέσματα του DBSCAN ως ελαφρώς παραπλανητικά.

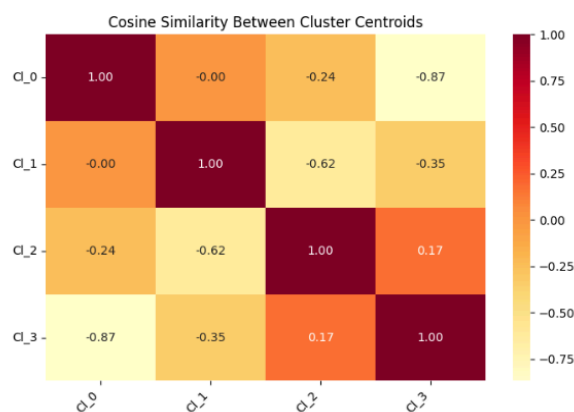
Τα πάρκα 96, 105, 180 και 209 θεωρήθηκαν θόρυβος από τον αλγόριθμο, δηλαδή αποτελούν πάρκα που δε μπορούν με σαφήνεια να καταταχθούν σε μία συστάδα.

Τα περισσότερα φαίνεται να έχουν μια υπολογίσιμη απόσταση από τις συστάδες, αλλά κάποια φαίνονται κοντά σε άλλες συστάδες. Αυτό οφείλεται στις υπόλοιπες συνιστώσες, οι οποίες όμως δε μπορούν να οπτικοποιηθούν, μια απεικόνιση η οποία θα προσέφερε ένα μεγαλύτερο επίπεδο σαφήνειας.

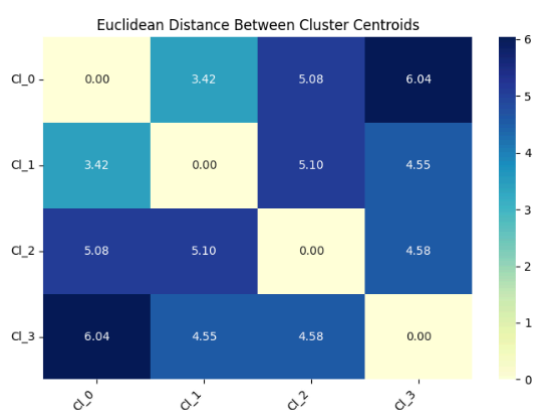
Ως επόμενα βήματα, θα ελεγχθεί πόσο συμπαγείς εσωτερικά είναι οι συστάδες, δηλαδή αν υπάρχουν πάρκα που με συνέπεια δεν ταιριάζουν στις ομάδες τους, και πιο συγκεκριμένα πόσα είναι αυτά. Βρίσκοντας πως μικρό ποσό από τα διαθέσιμα πάρκα δυσκολεύεται να ταξινομηθεί σε μια συγκεκριμένη συστάδα, επιβεβαιώνεται η ανθεκτικότητα της ομαδοποίησης. [23]

5.6 Έλεγχος Ομοιότητας των Συστάδων

Ως τελευταίο έλεγχο για την επιβεβαίωση ότι οι συστάδες είναι διακριτές, και οι αναπόφευκτες αλληλοκαλύψεις είναι όσο το δυνατόν λιγότερες και περιορισμένες, υπολογίζονται οι Ευκλείδειες Αποστάσεις και η Συνημιτονοειδής Ομοιότητα για κάθε ζευγάρι συστάδων:



Εικόνα 30: Συνημιτονοειδής Ομοιότητα για ιεραρχική ομαδοποίηση.



Εικόνα 31: Ευκλείδεια Απόσταση κατά ιεραρχική ομαδοποίηση.

Επιλέχθηκαν οι συστάδες που δημιουργήθηκαν από τον αλγόριθμο Ιεραρχικής Ομαδοποίησης, καθώς ήταν εκείνος που οπτικά έδινε τα καλύτερα αποτελέσματα, και είχε επίσης τις καλύτερες αξιολογήσεις χωρισμού συστάδας, βάσει των σκορ σιλουέτας και του DBI.

Εφόσον οι Ευκλείδειες αποστάσεις είναι σχετικά μεγάλες και ταυτόχρονα η Συνημιτονοειδής Ομοιότητα τείνει στην ξεκάθαρη ανομοιότητα, ειδικά για την πρώτη και τη δεύτερη συστάδα που η Ομοιότητα ισούται με 0, γεγονός που σημαίνει απόλυτη ανομοιότητα. Βγαίνει επομένως το συμπέρασμα πως τα κέντρα των συστάδων είναι καλά διαχωρισμένα στον χώρο και αντιστοιχούν σε γενικώς διαφορετικά χαρακτηριστικά.

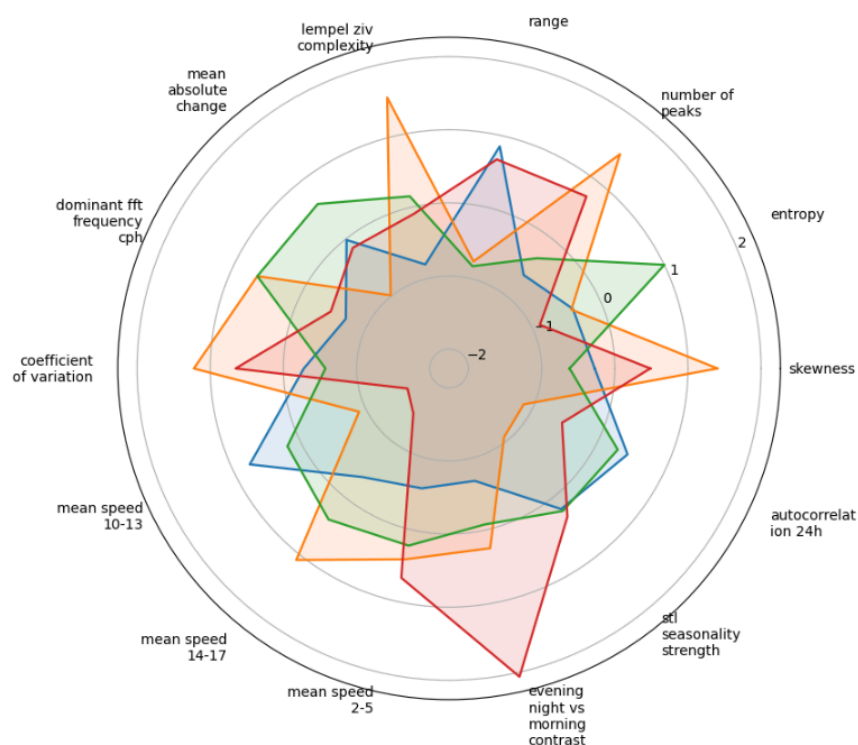
Εν κατακλείδι, ακολουθεί η λίστα με τις συστάδες και τα πάρκα από τα οποία αποτελούνται.

1. Συστάδα 0: 52, 75, 105, 114, 117, 149, 166, 175, 209, 212, 276, 289, 292
2. Συστάδα 1: 11, 31, 33, 64, 180, 235
3. Συστάδα 2: 1, 13, 42, 133, 136, 163, 165, 185, 255, 283, 322
4. Συστάδα 3: 37, 96, 103, 123

Με μια γρήγορη ματιά στον χάρτη της Ελλάδας στην αρχή του κεφαλαίου, πάρκα που βρίσκονται σε κοντινές τοποθεσίες κατανέμονται και στις ίδιες συστάδες. Αυτό είναι ένα σημάδι ότι η συσταδοποίηση ήταν επιτυχημένη και τα συμπεράσματα είναι πολύ εύστοχα και έγκυρα.

5.7 Χαρακτηρισμός και Ανάλυση των Συστάδων

Εφόσον έχουν επιλεγεί οι συστάδες, το επόμενο βήμα είναι να κατονομαστούν, αξιοποιώντας τα βασικότερα χαρακτηριστικά της καθεμίας. Για να οπτικοποιηθούν καλύτερα οι διαφορές, θα χρησιμοποιηθεί ένα εργαλείο που ονομάζεται Διάγραμμα Ραντάρ (Radar Plot), το οποίο μετατρέπει τα κέντρα των συστάδων, που ορίζονται από τη μαθηματική τιμή που έχουν τα στοιχεία για το δεδομένο κέντρο, σε κλειστά πολύγωνα.

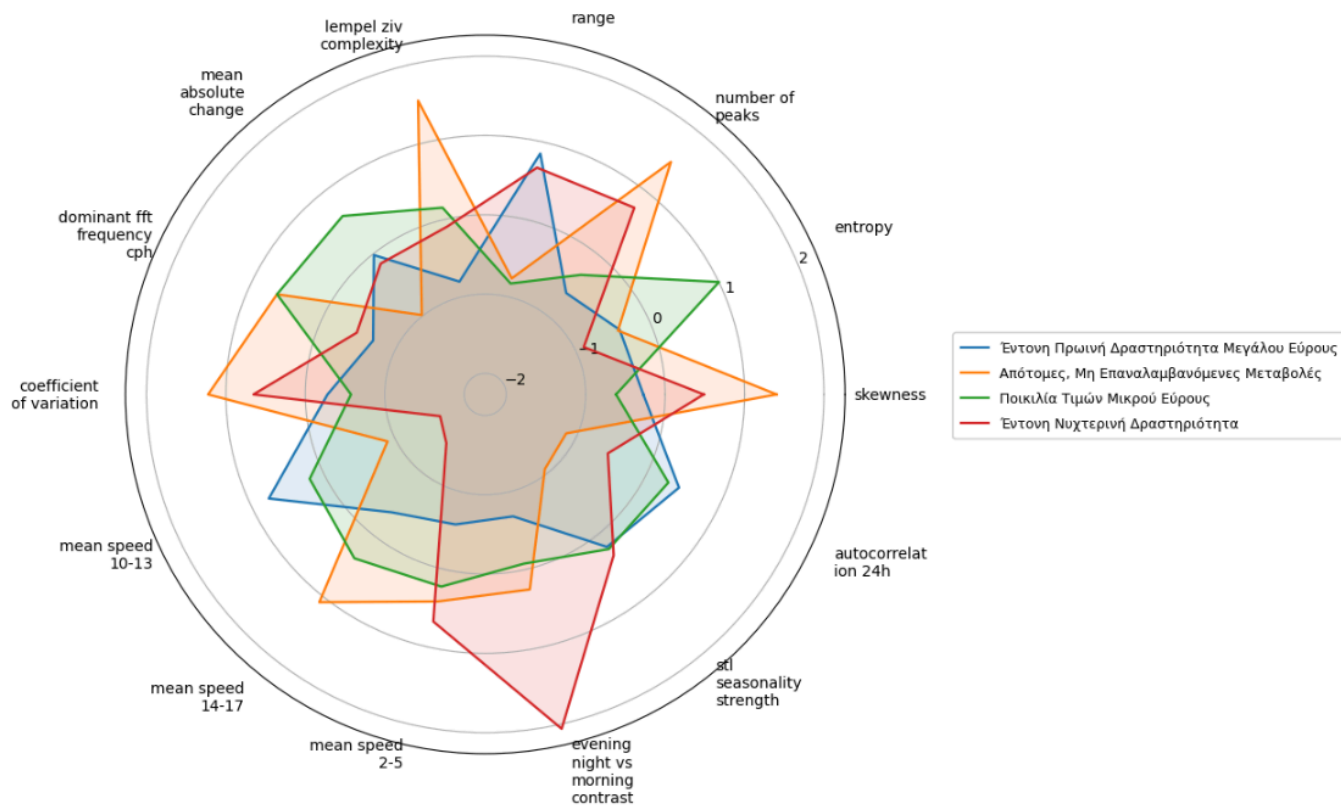


Εικόνα 32: Radar Plot των κέντρων των συστάδων.

Έτσι δημιουργείται καλύτερη εποπτεία των χαρακτηριστικών κάθε συστάδας, ενώ ταυτόχρονα επιβεβαιώνεται πως δεν υπάρχει κάποιο ζευγάρι συστάδων που να παρουσιάζει έντονες ομοιότητες μεταξύ του. Κάτι τέτοιο θα σήμαινε πως κάποια συστάδα θα ήταν περιττή και θα έπρεπε να αναθεωρηθεί ο αριθμός k των συνολικών συστάδων.

1. Η πρώτη συστάδα έχει υψηλό εύρος και μέση κανονικοποιημένη ταχύτητα για την περίοδο 10-13. Ταυτόχρονα έχει πολύ χαμηλή ένταση εποχικότητας, καθώς και ιδιαίτερα χαμηλή πολυπλοκότητα Lempel-Ziv. Τα πάρκα αυτής της συστάδας επομένως χαρακτηρίζονται κυρίως από τις ιδιαίτερα υψηλές και χαμηλές ταχύτητες που μπορούν να έχουν και από την επικράτηση των πρωινών ωρών όσον αφορά τις ταχύτητες των ανέμων. Δευτερευόντως χαρακτηρίζονται από την έλλειψη πολύπλοκων δομών στη χρονοσειρά τους, την επαναλαμβανόμενη συμπεριφορά και την προβλεψιμότητά τους. Επομένως ένα όνομα για τη συστάδα είναι το “Έντονη Πρωινή Δραστηριότητα Χαμηλού Εύρους”.
2. Η δεύτερη συστάδα έχει ιδιαίτερα υψηλή τιμή πολυπλοκότητας Lempel-Ziv, ενώ ο αριθμός κορυφών και η λοξότητα είναι επίσης σε υψηλά επίπεδα. Ταυτόχρονα η αυτοσυσχέτιση είναι χαμηλή. Έχουμε λοιπόν ένα σύνολο από πάρκα που παρουσιάζουν έντονες διακυμάνσεις και χαρακτηρίζονται από ιδιαίτερη πολυπλοκότητα. Επομένως η συστάδα μπορεί να ονομαστεί “Απότομες, Μη Επαναλαμβανόμενες Μεταβολές”.

3. Η τρίτη συστάδα έχει υψηλή εντροπία και MAC, και επίσης πολύ χαμηλό εύρος. Ο αριθμός κορυφών και η λοξότητα είναι επίσης χαμηλά. Τα πάρκα επομένως παρουσιάζουν μοτίβα και χαμηλή διακύμανση, λόγω των χαμηλών αντίστοιχων δεικτών, ενώ η υψηλή εντροπία υποδεικνύει πυκνή δομή όσον αφορά τη διασπορά των τιμών της χρονοσειράς. Άρα σε αυτή τη συστάδα δίνεται το όνομα “Ποικιλία Τιμών Μικρού Εύρους”.
4. Η τέταρτη και τελευταία συστάδα έχει πολύ υψηλή αντίθεση πρωινών-βραδινών ταχυτήτων, ενώ υψηλοί είναι και οι δείκτες διακύμανσης και αριθμού κορυφών. Αντίθετα, είναι πολύ χαμηλές οι ταχύτητες και για τις υπόλοιπες πρωινές ώρες, και παρατηρείται και χαμηλή εντροπία. Τα πάρκα επομένως χαρακτηρίζονται από ασυμμετρίες, αλλά όχι απαραίτητα από απρόβλεπτες συμπεριφορές. Η συστάδα θα ονομαστεί “Έντονη Νυχτερινή Δραστηριότητα”.



Εικόνα 33: Τελικό radar plot των προφίλ των συστάδων.

Είναι σημαντικό σε αυτή τη φάση να εντοπιστούν οι συστάδες που είναι συμπληρωματικές μεταξύ τους.

- Με μια πρώτη ματιά φαίνεται πως η “Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους” και η “Απότομες, Μη Επαναλαμβανόμενες Μεταβολές” αλληλοσυμπληρώνονται σε πολύ μεγάλο βαθμό. Οι τιμές LZC, αριθμού κορυφών και δείκτη διακύμανσης είναι αντίστροφες. Ο συνδυασμός αυτός είναι φυσικός, καθώς η μια συστάδα προσδίδει σταθερότητα λόγω υψηλής προβλεψιμότητας και η άλλη είναι η πιο απρόβλεπτη και ακανόνιστη από τις τέσσερις. Χρησιμοποιώντας αυτόν τον συνδυασμό συστάδων, “αλληλοεξουδετερώνονται” κατά κάποιον τρόπο οι ακραίες τιμές των δύο συστάδων, καταλήγοντας σε ένα σταθερό και ήπιο αποτέλεσμα.

Στο παραπάνω ζευγάρι μπορεί επίσης να προστεθεί και η συστάδα “Έντονη Νυχτερινή Δραστηριότητα”, διότι προσθέτει ύψη στα σημεία που οι άλλες δύο συστάδες έχουν πτώση, όπως στο στοιχείο night vs morning contrast. Με αυτό τον τρόπο, καλύπτονται βραχυπρόθεσμες, υψηλής συχνότητας διακυμάνσεις, που δεν καλύπτονται από καμία από τις άλλες δύο συστάδες. Μπορεί έτσι να επιτευχθεί η δημιουργία ενός χαρτοφυλακίου που είναι πλήρως καλυμμένο σε όλο το φάσμα των στοιχείων από άποψη σταθερότητας, τυχαιότητας και ημερήσιας διακύμανσης.

Η προσθήκη στο παραπάνω ζευγάρι και της συστάδας “Ποικιλία Τιμών Μικρού Εύρους” επίσης θα επιφέρει καλά αποτελέσματα, καθώς και εκείνη συμπληρώνει καλά τα κενά των άλλων 2 συστάδων.

- Ένα ακόμα ζευγάρι που ταιριάζει μαζί είναι το “Έντονη Νυχτερινή Δραστηριότητα” με το “Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους”. Ενώ δεν επιτυγχάνεται η καθολική συμπληρωματικότητα του συνδυασμού που αναφέρθηκε προηγουμένως, το στοιχείο που καλύπτεται σε πολύ μεγάλο βαθμό με αυτόν τον συνδυασμό είναι η ημερήσια παραγωγή ενέργειας, καθώς το “Έντονη Νυχτερινή Δραστηριότητα” ευθύνεται για τους υψηλούς απογευματινούς-βραδινούς ανέμους, τους οποίους δεν παρέχει το “Χαμηλή Πολυπλοκότητα”. Αναμένονται επομένως ικανοποιητικές αποδόσεις από αυτό το ζευγάρι, και θα αποτελέσει ένα χαρτοφυλάκιο, του οποίου ο στόχος είναι η κάλυψη των ενεργειακών αναγκών.
Η προσθήκη της συστάδας “Ποικιλία Τιμών Μικρού Εύρους” θα βοηθήσει εν δυνάμει την κάλυψη όλου του φάσματος των στοιχείων.

5.8 Κατασκευή των Χαρτοφυλακίων

Έχοντας μια ιδέα εξαρχής για τους καλούς συνδυασμούς από συστάδες, θα ελεγχθεί κατά πόσο τα πειράματα αντιστοιχούν στις πρώιμες εκτιμήσεις. Στην πρώτη υποενοότητα, λαμβάνεται ως αφετηρία μια από τις τέσσερις συστάδες, παίρνοντας 4 τυχαία πάρκα που θα προστεθούν στην άδεια βάση του χαρτοφυλακίου, και έπειτα ελέγχεται ποιες συστάδες επηρεάζουν θετικά ή αρνητικά το αρχικό χαρτοφυλάκιο. Μέσω αυτού του πειράματος, θα εμφανιστεί το προφίλ πάρκων που είναι πιο ωφέλιμο να προστεθεί σε κάθε αρχικό χαρτοφυλάκιο, και με πόση βεβαιότητα μπορεί μια συστάδα να λειτουργήσει ευνοϊκά στη βάση του χαρτοφυλακίου.

Στην πρώτη εκδοχή του πειράματος, σε κάθε αρχικό χαρτοφυλάκιο προστίθεται 1 καινούργιο τυχαίο πάρκο από μια τυχαία συστάδα κάθε φορά, με το πείραμα να εκτελείται 1000 φορές, ελέγχοντας τις μέσες τιμές του CoV καθώς και τα ποσοστά μείωσης τους. Η προσθήκη μόνο ενός πάρκου είναι ένα ασφαλές και συνηθισμένο σενάριο στον κλάδο του ενεργειακού χρηματιστηρίου, επομένως έχει ουσία να ελεγχθεί πόσο επηρεάζει και με πόση βεβαιότητα η προσθήκη μονάχα ενός περιουσιακού στοιχείου. Επίσης χρήσιμο είναι να υπολογιστούν τα εκατοστημόρια q_{05} , q_{95} για κάθε συνδυασμό, ώστε να υπάρχει εποπτεία των τιμών CoV μεταξύ των οποίων κυμαίνεται η συντριπτική πλειοψηφία των επαναλήψεων. Με τη βοήθεια του διαστήματος $[q_{05}, q_{95}]$ άλλωστε επιβεβαιώνεται και η μέση τιμή του CoV για το χαρτοφυλάκιο.

Στην δεύτερη εκδοχή του πειράματος, σε κάθε αρχικό χαρτοφυλάκιο θα προστεθούν επιπλέον 4, αντί για 1, πάρκα, και θα γίνει σύγκριση με τα αποτελέσματα της προσθήκης 1 πάρκου. Το πείραμα θα εκτελεστεί ξανά 1000 φορές. Αναμένεται πτώση του CoV, καθώς μεγαλύτερος αριθμός πάρκων συνεπάγεται σίγουρα περισσότερη διαφοροποίηση εντός ενός χαρτοφυλακίου, και θα μελετηθεί πόσο αυξάνεται ή μειώνεται η πιθανότητα κερδών CoV.

5.8.1 Αφετηρία με μια συστάδα

1. Αρχική συστάδα 0 (Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους)

Κατόπιν εκτέλεσης των προσομοιώσεων, προκύπτουν τα εξής αποτελέσματα:

Βάση Χαρτοφυλακίου	Προσθήκ η από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C0:4	C0	0.549	0.6204	0.6128	-1,2	-0.0576	0.025
	C1	0.831		0.6063	-2,2	-0.0413	0.004
	C2	0.807		0.6065	-2,2	-0.06	0.005
	C3	1		0.592	-4,6	-0.047	-0.013

Πίνακας 4 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 0.

Αυτό το σενάριο δείχνει πως τα κέρδη σε CoV δεν κυμαίνονται σε υψηλά επίπεδα, αλλά ο βασικός CoV ήταν ήδη σχετικά χαμηλός (0.6204). Η προσθήκη 1 πάρκου από τη συστάδα 0 είναι κίνηση που δε μπορεί με τίποτα να

συστηθεί, καθώς επιφέρει πολύ ρίσκο (πιθανότητα μείωσης του CoV ελάχιστα ευνοϊκότερη από ρίψη κέρματος, 0.549) και ελάχιστη αξία (high risk - low reward). Αντίθετα, η προσθήκη πάρκου από τη συστάδα 3 προσφέρει εγγυημένο κέρδος, αλλά με μέτριο κέρδος σε CoV (μείωση κατά περίπου 4.6%). Οι προσθήκες από τις άλλες δύο συστάδες ομοίως έχουν πολλές πιθανότητες να βελτιώσουν τον CoV, αλλά για λίγο κέρδος, επομένως κι εκείνες δε φαίνεται να ωφελούν ιδιαίτερα.

2. Αρχική συστάδα 1 (Απότομες, Μη Επαναλαμβανόμενες Μεταβολές)

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C1:4	C0	0.796	0.6387	0.6044	-5,3	-0.1142	0.0185
	C1	0.658		0.626	-2	-0.067	0.0331
	C2	0.956		0.599	-6,2	-0.1053	-0.0014
	C3	0.965		0.6165	-3,4	-0.0533	-0.0015

Πίνακας 5 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 1.

Για αυτή την αρχική κατάσταση, παρατηρούνται καλύτερα και πιο ομοιόμορφα αποτελέσματα. Προσθήκες από τις συστάδες 2 και 3 προσφέρουν σχεδόν εγγυημένη βελτίωση (πιθανότητες μείωσης CoV 0.956 και 0.965, αντίστοιχα), ενώ τα κέρδη τους σε CoV κυμαίνονται στο 6,2% και 3,4%, αντίστοιχα. Μάλιστα, το διάστημα $[q_{05}, q_{95}]$ είναι πολύ πιο λεπτό στην περίπτωση προσθήκης από τη συστάδα 3, κάτι που δείχνει πως η συγκεκριμένη προσθήκη έχει πιο ομοιόμορφα αποτελέσματα σε σχέση με την προσθήκη από τη συστάδα 1, όπου το διάστημα είναι σχεδόν διπλάσιο. Συνολικά, φαίνεται πως η προσθήκη από τη συστάδα 3 είναι μια ισορροπημένη, εγγυημένη επένδυση. Αντιθέτως η προσθήκη από τη συστάδα 2, παρόλο που είναι εγγυημένη η βελτίωση, ενδέχεται να επιφέρει πολύ μικρότερη από την αναμενόμενη.

Η επιπλέον προσθήκη από τη συστάδα 0 έχει και σε αυτή την περίπτωση, όπως και στην προηγούμενη, ανώφελο ρίσκο, αφού τα κέρδη σε CoV είναι μικρά. Τέλος, η προσθήκη από τη συστάδα 1 προσφέρει γενικά μέτρια αποτελέσματα και δεδομένης της ύπαρξης καλύτερων επιλογών, δε μπορεί να συστηθεί ως καλή επένδυση.

Μια επιπλέον ενδιαφέρουσα παρατήρηση είναι πως οι τιμές των εκατοστημορίων q_{05} και q_{95} έχουν μεγάλη απόσταση, σε σχέση με τις άλλες περιπτώσεις. Με πιθανότητα 5% ο CoV μπορεί να μειωθεί κατά ~16,5%, ενώ υπάρχει 5% πιθανότητα να μειωθεί μόνο κατά 0,22%. Αυτή η μεγάλη απόκλιση οφείλεται στο αλλοπρόσαλλο προφίλ που έχουν τα πάρκα της συστάδας 1, τα οποία είναι τα πιο ασταθή και απρόβλεπτα, και συνεπώς επηρεάζουν και την επίδραση που θα έχει το πάρκο από κάποια άλλη συστάδα. Για βελτίωση των αποτελεσμάτων, θα χρειαστούν περαιτέρω προσθήκες πάρκων για να εξισορροπηθεί η αστάθεια που εν γένει φέρνουν τα πάρκα αυτά.

3. Αρχική συστάδα 2 (Ποικιλία Τιμών Μικρού Εύρους)

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C2:4	C0	0.567	0.6035	0.5971	-1,05	-0.045	0.029
	C1	0.748		0.59	-2,1	-0.037	0.008
	C2	0.526		0.5972	-1,05	-0.059	0.017
	C3	1		0.571	-4,67	-0.0472	-0.021

Πίνακας 6 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 2.

Σε αυτή την περίπτωση, τα αποτελέσματα από μαθηματική άποψη είναι μεν καλύτερα συγκριτικά με τις άλλες περιπτώσεις, αλλά αυτό οφείλεται επίσης και στο χαμηλότερο επίπεδο CoV που υπήρχε και πριν την προσθήκη. Η προσθήκη ενός πάρκου από τη συστάδα 3 εγγυάται και εδώ βελτίωση κατά περίπου 4,7%, και η μέση τελική τιμή του CoV είναι η χαμηλότερη από τους τέσσερις πιθανούς συνδυασμούς αρχικών χαρτοφυλακίων (0.571). Προσθήκες από τις άλλες 3 συστάδες επιφέρουν μεγάλο ρίσκο για μικρό κέρδος, μάλιστα για ελάχιστο κέρδος στην περίπτωση των συστάδων 0 και 2, οπότε δε μπορούν να συστηθούν σε κανένα σενάριο, καθώς δεν επικρατούν της συστάδας 3 σε κανέναν τομέα. Συγκεκριμένα για την προσθήκη από τη συστάδα 0, το εκατοστημόριο q_{95} μπορεί να επιφέρει αύξηση του CoV κατά περίπου 0.03, τη στιγμή που η κατά μέσο όρο μείωση του CoV είναι μόλις 0.0064. Είναι προφανές επομένως πως προσθήκη από εκείνη τη συστάδα μπορεί να δράσει καταστροφικά στο χαρτοφυλάκιο.

4. Αρχική συστάδα 3 (Έντονη Νυχτερινή Δραστηριότητα)

Σε αυτή την περίπτωση δε γίνεται να προστεθούν πάρκα από τη συστάδα 3, καθώς αυτή περιέχει μόνο τα 4 πάρκα που βρίσκονται ήδη στο χαρτοφυλάκιο:

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C3:4	C0	1	0.6843	0.6123	-10,51	-0.1522	-0.037
	C1	0.89		0.6515	-4,79	-0.069	0.0014
	C2	1		0.61	-10,85	-0.127	-0.045

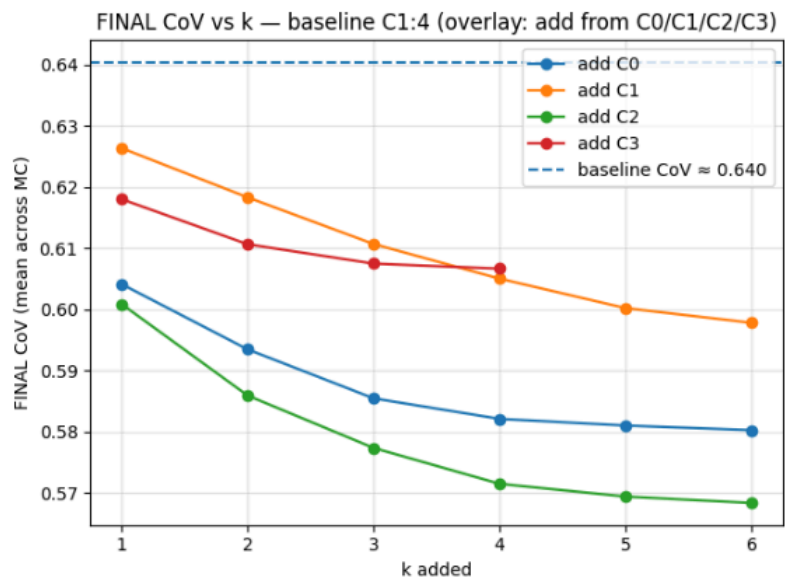
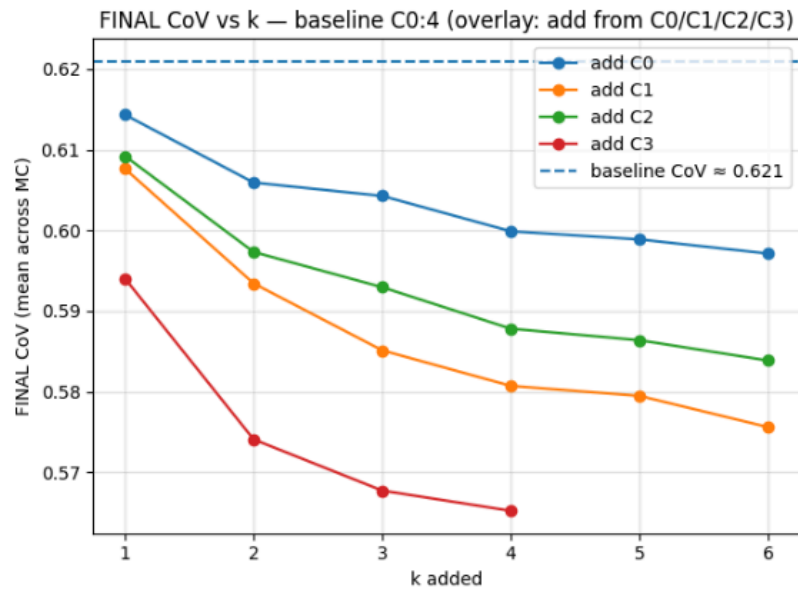
Πίνακας 7 : Αποτελέσματα για αρχικό χαρτοφυλάκιο της συστάδας 3.

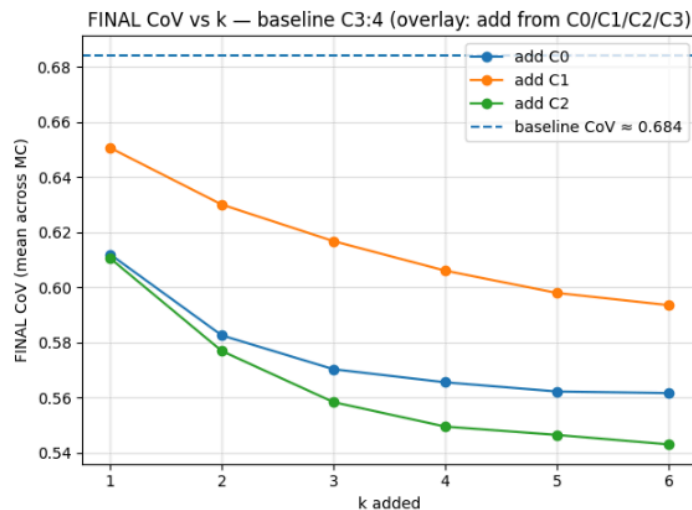
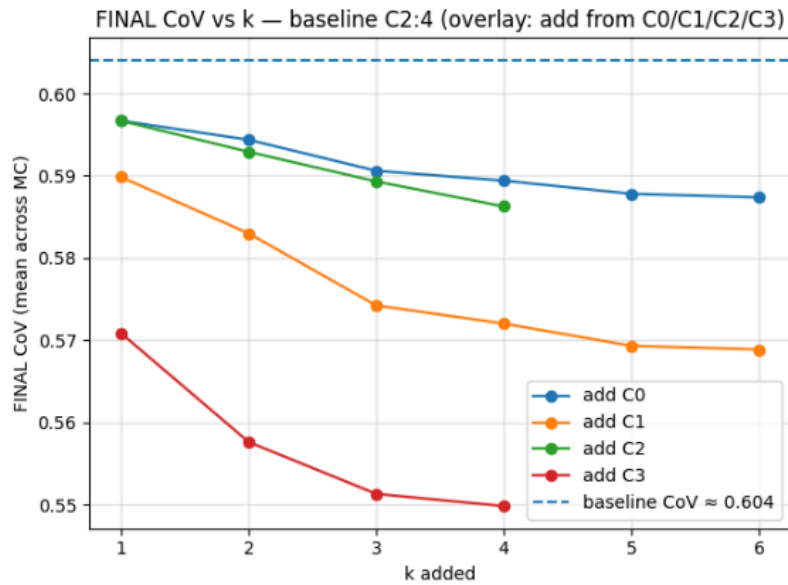
Παρατηρούνται πολύ μεγαλύτερες μειώσεις του CoV, αλλά το αρχικό χαρτοφυλάκιο είχε ιδιαίτερα υψηλή αρχική κατάσταση. Οι μέσες τελικές τιμές του CoV, άλλωστε, είναι υψηλότερες από τις άλλες περιπτώσεις. Παρόλα αυτά, οι προσθήκες από τις συστάδες 0 και 2 προσφέρουν εγγυημένη μείωση της τάξης του 10,9% και 10,5% αντίστοιχα. Όπως και στην περίπτωση αρχικού χαρτοφυλακίου πάρκων της συστάδας 1, και εδώ τα εκατοστημόρια q_{05} , q_{95} εμφανίζουν μεγάλες αποκλίσεις. Και η συστάδα 3 αποτελείται από πάρκα με ιδιόρρυθμο προφίλ, και για μεγαλύτερη ευστάθεια και καλύτερα αποτελέσματα προτείνονται περαιτέρω προσθήκες πάρκων προς διαφοροποίηση του χαρτοφυλακίου.

Η προσθήκη από τη συστάδα 1 δε συμφέρει συγκριτικά με τις άλλες δύο περιπτώσεις, καθώς μειώνει κατά πολύ λιγότερο ποσοστό τον CoV (-4,8%), ενώ δεν προσφέρει εγγύηση βελτίωσης του CoV.

Έχοντας ελέγξει τη βελτίωση του CoV για το σενάριο προσθήκης μόνο ενός πάρκου, μπορεί επίσης να εξεταστεί πόσο επηρεάζει η προσθήκη μεγαλύτερου αριθμού πάρκων από μια συστάδα στο αρχικό χαρτοφυλάκιο. Φυσικά, όσο περισσότερα πάρκα προστίθενται, τόσο θα χαμηλώνει ο CoV και το ρίσκο του χαρτοφυλακίου, καθώς επιτυγχάνεται μεγαλύτερη διαφοροποίηση. Το ουσιαστικό αποτέλεσμα είναι ο εντοπισμός του αριθμού εκείνου των νέων πάρκων, για τον οποίο επιτυγχάνεται η μέγιστη ποσοστιαία μείωση του CoV.

Με βάση τα παραπάνω, προκύπτουν οι εξής πίνακες για κάθε ενδεχόμενη αρχική συστάδα 4 πάρκων:





Εικόνα 34: Μείωση CoV για όλα τα σενάρια αρχικών χαρτοφυλακίων, συναρτήσει του αριθμού προστιθέμενων πάρκων.

Οι πίνακες αυτοί όμως καλύπτουν μόνο το κομμάτι της μείωσης του CoV και όχι τις πιθανότητες της μείωσης του CoV. Οπότε για περαιτέρω έλεγχο, εκτελείται για δεύτερη φορά το πείραμα της παραγράφου, επιλέγοντας αυθαίρετα την προσθήκη 4 πάρκων αντί για 1 στο ήδη υπάρχον χαρτοφυλάκιο. Με αυτό τον τρόπο, επιλέγεται μια ενδιάμεση τιμή προστιθέμενων πάρκων, με σκοπό να φανούν τυχόν αυξήσεις ή μειώσεις της πιθανότητας βελτίωσης του CoV. Έτσι λαμβάνονται τα εξής αποτελέσματα:

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C0:4	C0	0.638	0.6207	0.6008	-3,19	-0.0917	0.0365
	C1	0.964		0.5818	-6,25	-0.08	-0.0037
	C2	0.877		0.589	-5,09	-0.078	0.008
	C3	1		0.5642	-9,085	-0.0942	-0.0274
C1:4	C0	0.878	0.6395	0.5855	-8,43	-0.1397	0.0173
	C1	0.757		0.6049	-5,4	-0.1221	0.0357
	C2	0.958		0.5729	-10,4	-0.1409	-0.0018
	C3	0.895		0.6055	-5,3	-0.0844	0.0038
C2:4	C0	0.693	0.6043	0.5888	-2,55	-0.0628	0.0346
	C1	0.926		0.573	-5,16	-0.0701	0.0033
	C2	0.564		0.5862	-3	-0.0696	0.0264
	C3	1		0.5501	-8,95	-0.0817	-0.0289
C3:4	C0	1	0.6843	0.5653	-17,39	-0.1204	-0.0276
	C1	1		0.607	-11,3	-0.1695	-0.0767
	C2	1		0.5505	-19,54	-0.1628	-0.0983

Πίνακας 8: Αποτελέσματα για όλα τα πιθανά χαρτοφυλάκια, για την περίπτωση προσθήκης 4 πάρκων.

Συνδυάζοντας τον παραπάνω πίνακα με τα προηγούμενα αποτελέσματα, διαφαίνονται τα εξής αναλυτικά συμπεράσματα:

- Η προσθήκη πάρκων από τη συστάδα 3 (Νυχτερινή Δραστηριότητα) οφείλει να είναι μια πάγια τακτική. Οι προσθήκες είναι ταυτόχρονα η πιο ασφαλής επιλογή σχεδόν για όλα τα δυνατά σενάρια και η επιλογή που οδηγεί στις χαμηλότερες τιμές του CoV. Τα πάρκα της συστάδας 3 είναι εκείνα που προσφέρουν τη μεγαλύτερη συμπληρωματικότητα, επομένως το αποτέλεσμα αυτό είναι αναμενόμενο. Για εκκίνηση από πάρκα της συστάδας 1, για την οποία η συστάδα 3 δεν είναι η βέλτιστη επιλογή, παραμένει ασφαλέστερη παρά τη μικρότερη μείωση του CoV.
- Η προσθήκη πάρκων από τη συστάδα 2 (Περιοδικότητα) είναι η δεύτερη καλύτερη επιλογή. Τα κέρδη είναι μικρότερα, αλλά η ασφάλεια με την οποία μπορούν αυτά να εξασφαλιστούν είναι επίσης πολύ υψηλή.
- Προσθήκες από ίδια συστάδα με την αρχική είναι οι χειρίστες επιλογές. Ως κινήσεις, είναι εκείνες με τη μικρότερη πιθανότητα μείωσης του CoV, επιφέροντας ρίσκο, ενώ ταυτόχρονα τα κέρδη είναι κατά μέσο όρο τα χαμηλότερα. Είναι μια στρατηγική η οποία δεν προσφέρει κανένα πλεονέκτημα σε σχέση με τις προσθήκες από άλλες συστάδες. Η έλλειψη ποικιλίας επηρεάζει και έμμεσα την ποιότητα του χαρτοφυλακίου.
Με βάση την παρατήρηση αυτή, συνεπάγεται πως ένα άνω όριο πάρκων στο ίδιο χαρτοφυλάκιο με προέλευση την ίδια συστάδα είναι το 4.
- Οι καλύτερες επιλογές για κάθε αρχική θέση υπερέχουν σε ουσιαστικό βαθμό των δεύτερων καλύτερων επιλογών. Ελέγχοντας τη μείωση του CoV για κάθε μία από τις καλύτερες επιλογές φαίνονται σημαντικές διαφορές, της τάξης του 0.01 και υψηλότερα, ενώ οι υπόλοιπες επιλογές προσφέρουν αρκετά χαμηλότερα κέρδη και τελικές τιμές CoV σημαντικά υψηλότερες από αυτές των βέλτιστων επιλογών. Η βέλτιστη στρατηγική επομένως είναι πολύ ξεκάθαρη, και δεν χρειάζονται κριτήρια διευθέτησης διαφορών (tiebreakers) για τη λήψη της τελικής απόφασης.
- Ο καλύτερος συνδυασμός προκύπτει από τις συστάδες 2 (Περιοδικότητα) και 3 (Νυχτερινή Δραστηριότητα), οι οποίες με απόλυτη βεβαιότητα πετυχαίνουν μείωση του CoV, καθώς και τη

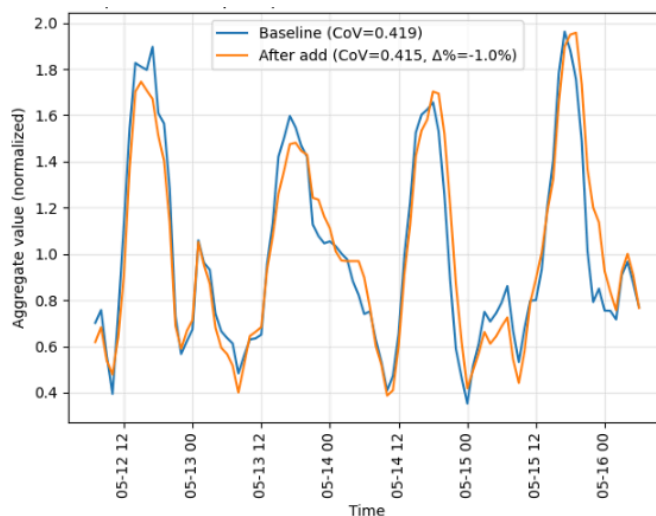
χαμηλότερη τιμή του CoV από όλους τους συνδυασμούς (κυμαινόμενες τιμές γύρω από το 0.55). Το αποτέλεσμα αυτό έχει ιδιαίτερο ενδιαφέρον, καθώς η αύξηση των προστιθέμενων πάρκων επιφέρει ραγδαία μείωση του CoV, παρά την υψηλή τιμή του CoV στο αρχικό χαρτοφυλάκιο. Αν το αρχικό χαρτοφυλάκιο έχει πάρκα της συστάδας 3, μπορεί να επιτευχθεί η μέγιστη ποσοστιαία μείωση του CoV (-19,5% μείωση). Αυτό οφείλεται στις μεγάλες διαφορές των προφίλ των πάρκων που προστίθενται, που εγγυώνται μεγάλη διαφοροποίηση στο χαρτοφυλάκιο.

- Ο δεύτερος καλύτερος συνδυασμός προκύπτει από τον συνδυασμό των συστάδων 0 και 3, καταφέροντας να μειώσουν τον CoV γύρω στο 0.56, επίσης με εγγυημένα αποτελέσματα (πιθανότητα μείωσης ίση με 1).
- Οι συνδυασμοί από συστάδες που προσφέρουν τον χαμηλότερο CoV είναι επίσης εκείνες που είναι οι πιο σίγουρες κινήσεις, με πιθανότητα βελτίωσης του CoV πολύ κοντά στο 1, ακόμα και 1. Αυτό σημαίνει πως στην πράξη, η προσθήκη πάρκων σε δεύτερη φάση, που θα αναλυθεί στην ακριβώς επόμενη υποενότητα, είναι η κρίσιμη κίνηση που θα καθορίσει πόσο χαμηλός κατά μέσο όρο CoV μπορεί να επιτευχθεί με τη μεγαλύτερη δυνατή βεβαιότητα.
- Η αλλαγή από 1 σε 4 προστιθέμενα πάρκα δεν επηρεάζει, αναλογικά, τα αποτελέσματα. Ο CoV μειώνεται και για τα τέσσερα διαφορετικά σενάρια, αντιθέτως το χαρακτηριστικό που αλλάζει, είναι η πιθανότητα να μειωθεί πράγματι ο CoV. Παίρνοντας το παράδειγμα του αρχικού χαρτοφυλακίου 2, η προσθήκη μόνο ενός πάρκου από τη συστάδα 1 έχει πιθανότητα μείωσης του CoV 0.75 στην τιμή του 0.59. Αντιθέτως, αν προστεθούν 4 πάρκα από τη συστάδα 1, ο CoV μπορεί να μειωθεί με πιθανότητα 0.93 στην τιμή του 0.573. Βέβαια δεν αποτελεί τη βέλτιστη επιλογή για τις δεδομένες αρχικές συνθήκες, καθώς η προσθήκη πάρκων από τη συστάδα 3 υπερτερεί.
- Εκκινώντας από τη συστάδα 3, φαίνεται πως προσθέτοντας μεγαλύτερο αριθμό πάρκων επιτυγχάνεται και αύξηση της πιθανότητας βελτίωσης του CoV. Αυτό συμβαίνει, διότι η συστάδα αυτή “Έντονη Νυχτερινή Δραστηριότητα” παρουσιάζει μεγάλες διακυμάνσεις και απρόβλεπτη συμπεριφορά, και είναι λογικό να απαιτούνται περισσότερα πάρκα για την εξομάλυνση της αθροιστικής χρονοσειράς του χαρτοφυλακίου.

Για τους δύο βέλτιστους συνδυασμούς που αναφέρθηκαν προηγουμένως, μπορεί να ελεγχθεί και διαγραμματικά η επιρροή της προσθήκης των πάρκων από τη νέα συστάδα στη χρονοσειρά του χαρτοφυλακίου. Εδώ είναι σημαντικό να αναφερθεί πως μικρές αποκλίσεις σε όλες τις τιμές του CoV των χρονοσειρών σε σχέση με τις τιμές των παραπάνω πινάκων οφείλονται σε διαφορετικούς τρόπους μέτρησης του CoV, και πιο συγκεκριμένα στον διαφορετικό τρόπο διαχείρισης των κενών παρατηρήσεων της χρονοσειράς.

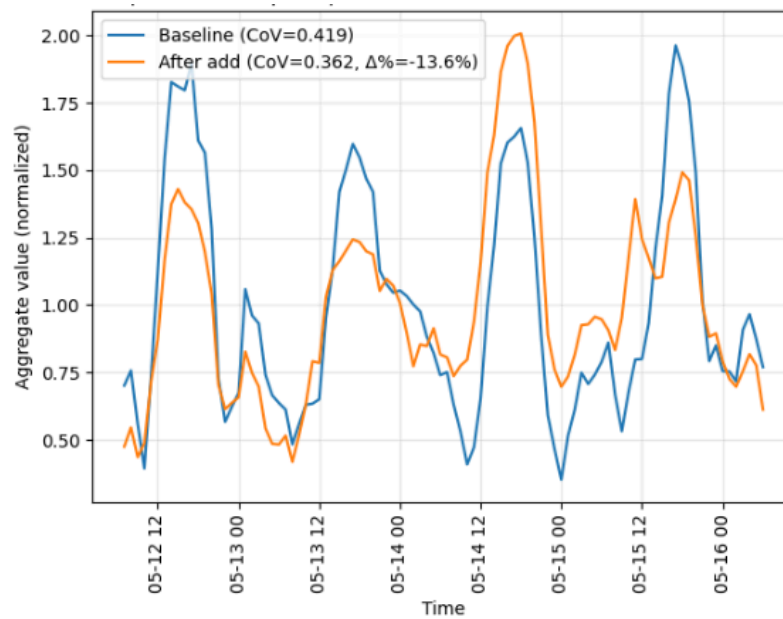
Στην πρώτη υλοποίηση, από την οποία προέκυψαν οι πίνακες την ενότητας αυτής, οι κενές παρατηρήσεις αγνοούνταν και δεν επηρέασαν τη μέση τιμή. Αυτή η στρατηγική επιλέχθηκε για την μείωση της πολυπλοκότητας του κώδικα που εκτελέστηκε, εξαιτίας της χρήσης των προσομοιώσεων Monte Carlo. Αντίθετα, στην δεύτερη υλοποίηση, οι κενές παρατηρήσεις απλά χαρακτηρίζονται ως μηδενικές, με αποτέλεσμα η μέση τιμή να μειωθεί, που επιφέρει ως συνέπεια την αύξηση του CoV (υπενθυμίζεται πως $CoV = \sigma / \mu$).

- Συνδυασμός 2 / 3 (Ποικιλία Τιμών Μικρού Εύρους / Έντονη Νυχτερινή Δραστηριότητα)
Αρχικά δοκιμάζεται η προσθήκη 1 πάρκου:



Εικόνα 35: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 3.

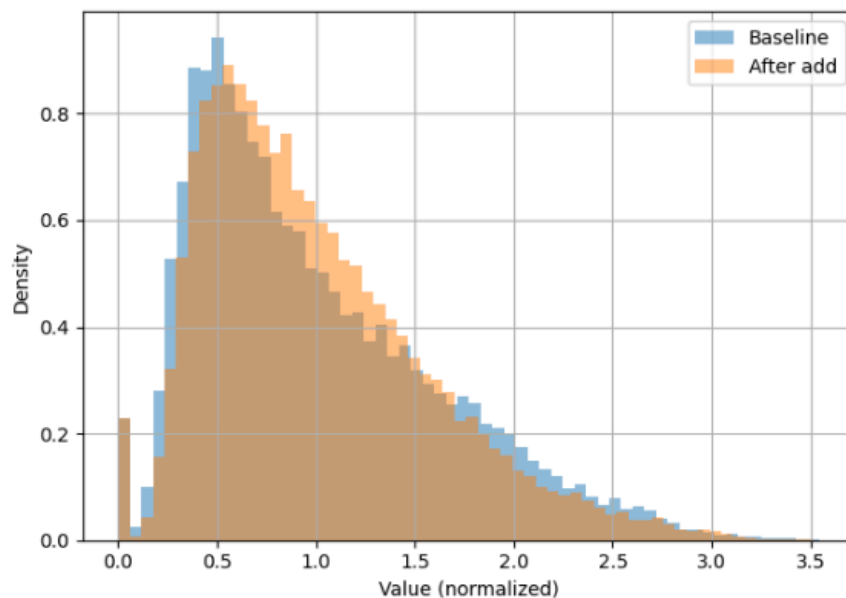
Ακολουθεί η δοκιμή για προσθήκη 4 πάρκων:



Εικόνα 36: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 3.

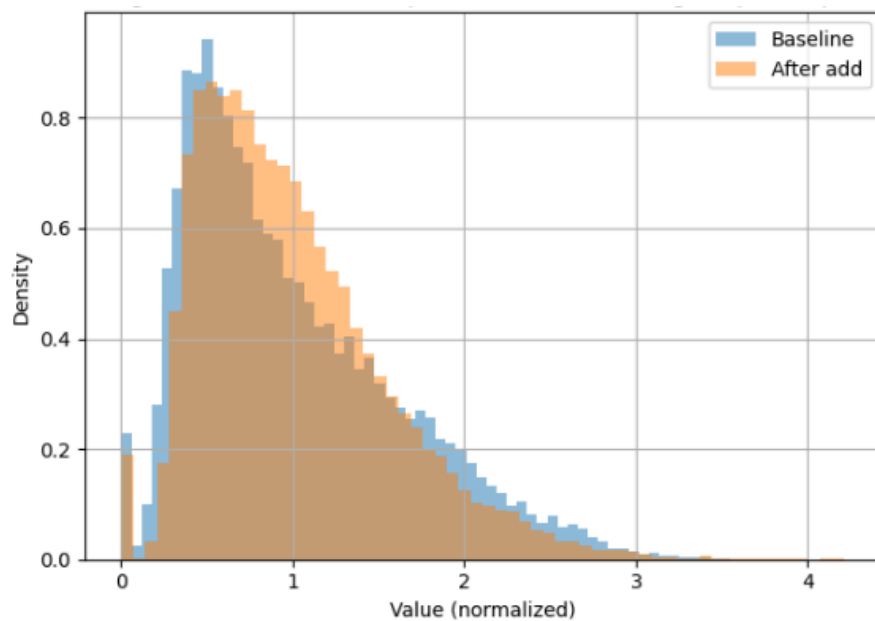
Για την προσθήκη 1 πάρκου, παρατηρείται μια μείωση της διακύμανσης της χρονοσειράς, μολονότι δεν είναι τόσο έντονη. Η προσθήκη 4 πάρκων είναι προφανές πως βοήθησε παραπάνω, μειώνοντας τον CoV κατά σχεδόν 14% στο ενδεικτικό χαρτοφυλάκιο. Φαίνεται επομένως μια ξεκάθαρη βελτίωση της απόδοσης του χαρτοφυλακίου 2 / 3 με την προσθήκη μεγαλύτερου αριθμού πάρκων.

Ακολουθεί τέλος και το ιστόγραμμα πυκνότητας πιθανότητας των τιμών, κανονικοποιημένο γύρω από την τιμή 1, αρχικά για προσθήκη 1 και έπειτα και για προσθήκη 4 πάρκων:



Εικόνα 37: Κατανομή τιμών έπειτα από προσθήκη 1 πάρκου.

Με την προσθήκη 4 πάρκων:

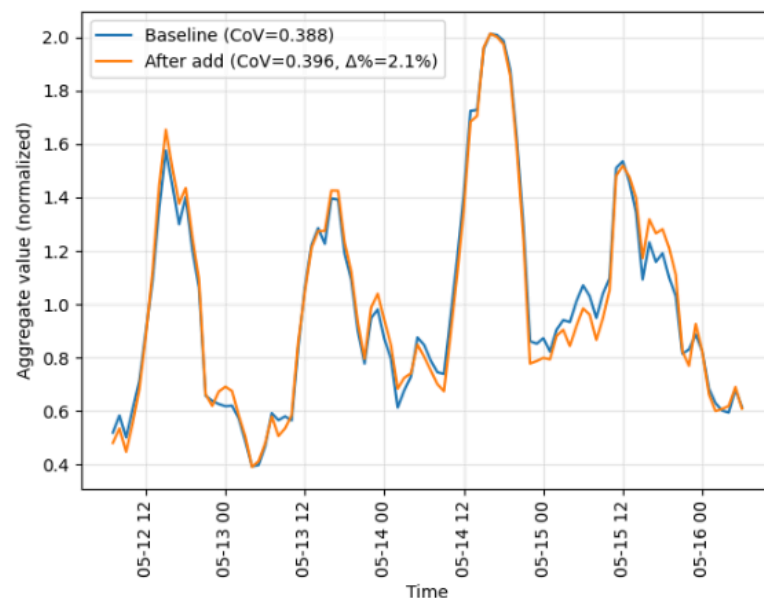


Εικόνα 38: Κατανομή τιμών έπειτα από προσθήκη 4 πάρκων.

Η βελτίωση του CoV είναι πιο ξεκάθαρη εδώ. Ειδικά για την προσθήκη 4 πάρκων, οι “ουρές” του ιστογράμματος έχουν περιοριστεί σε πολύ μεγάλο βαθμό, με τις περισσότερες τιμές ταχυτήτων να κυμαίνονται πιο έντονα γύρω από την κανονικοποιημένη τιμή 1

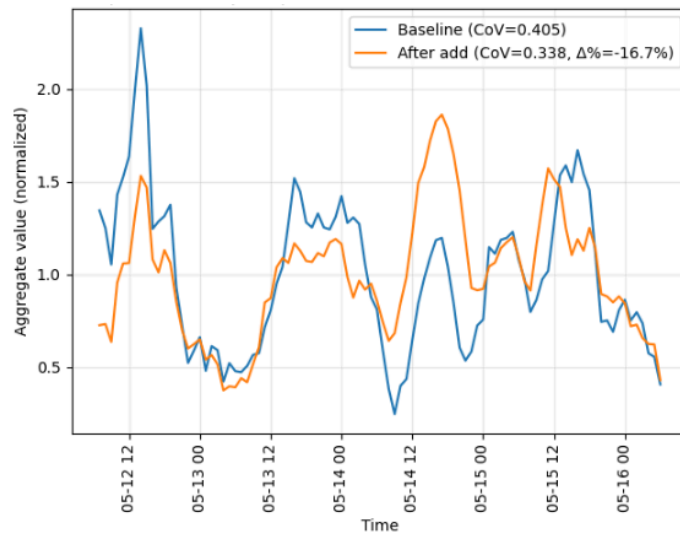
- Συνδυασμός 0 / 3 (Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους / Έντονη Νυχτερινή Δραστηριότητα)

Η προσθήκη ενός πάρκου αποφέρει την εξής ομαλοποιημένη χρονοσειρά:



Εικόνα 39: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 3.

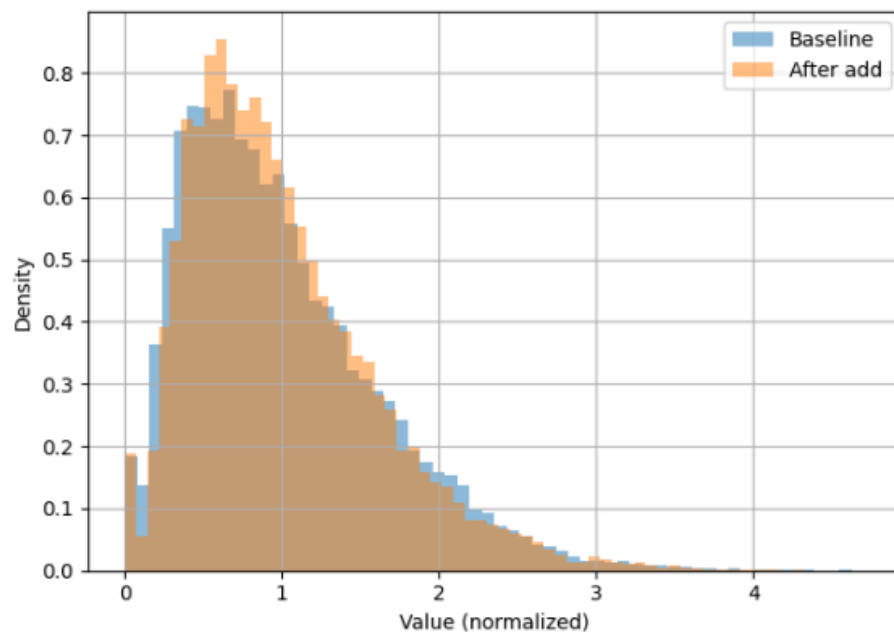
Και η προσθήκη 4 πάρκων:



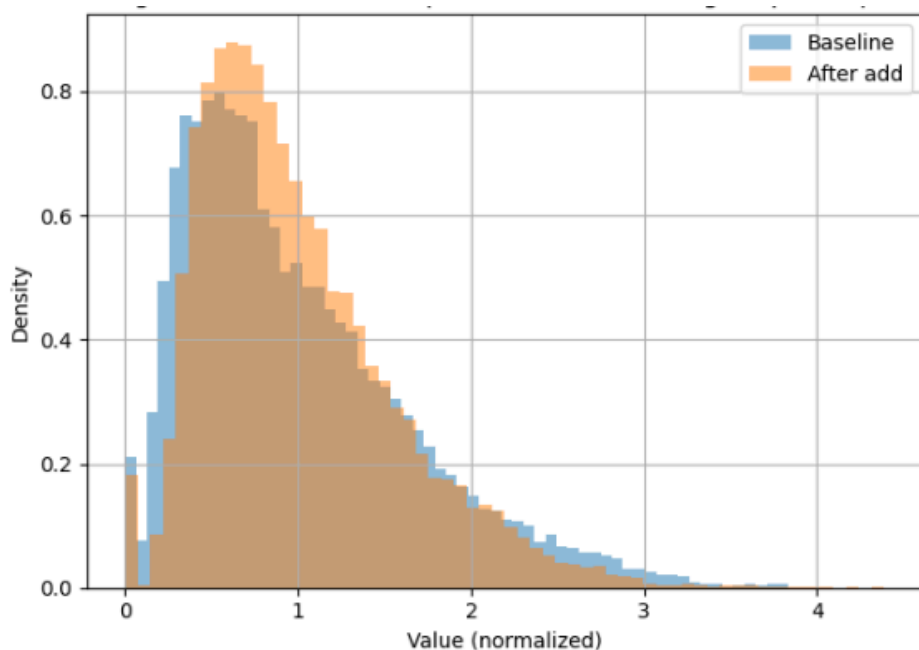
Εικόνα 40: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 3.

Σε αυτή την περίπτωση, παρατηρείται για την προσθήκη 1 πάρκου στο ενδεικτικό χαρτοφυλάκιο μια μικρή αύξηση του CoV, αλλά επειδή το παράδειγμα είναι τυχαίο δεν επηρεάζει φυσικά τα αποτελέσματα. Μεγάλη μείωση του CoV κατά ~17% για τυχαία επιλογή ενός μέσου συνδυασμού 4 πάρκων. Τα επίπεδα του CoV είναι επίσης εξαιρετικά, αλλά δεύτερα καλύτερα σε σχέση με τον συνδυασμό 2 / 3 στη μέση περίπτωση.

Τα ιστογράμματα των συναρτήσεων πυκνότητας πιθανότητας είναι ως εξής:



Εικόνα 41: Κατανομή τιμών έπειτα από προσθήκη 1 πάρκου.



Εικόνα 42: Κατανομή τιμών έπειτα από προσθήκη 4 πάρκων.

Και εδώ παρατηρείται φανερή βελτίωση της ομαλοποίησης του χαρτοφυλακίου, με τις τιμές να είναι πολύ πιο συγκεντρωμένες γύρω από την τιμή 1. Μπορεί όμως να εκτιμηθεί με το μάτι πως οι ακραίες τιμές σε αυτό το χαρτοφυλάκιο είναι περισσότερες και δεν έχουν εξισορροπηθεί τόσο καλά, όσο στην περίπτωση των συνδυασμών 2 / 3.

5.8.2 Αφετηρία με δύο συστάδες

Σε αυτή την υποενότητα, θα ελεγχθεί ποιες προσθήκες πάρκων είναι ευνοϊκές, με αφετηρία 4 + 4 πάρκα από δύο αρχικές συστάδες. Η πρώτη εκδοχή αφορά προσθήκη 1 πάρκου, η δεύτερη προσθήκη 4 πάρκων.

1. Αρχικές συστάδες 0 / 1 (Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους / Απότομες, Μη Επαναλαμβανόμενες Μεταβολές)

Βάση Χαρτοφυλακίου	Προσθήκη η από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C0:4, C1:4	C0	0.453	0.58235	0.5862	+0,31	-0.04	0.0195
	C1	0.66		0.5784	-0,77	-0.0249	0.011
	C2	0.593		0.5774	-0,84	-0.0343	0.0074
	C3	0.996		0.5702	-2,06	-0.0236	-0.0032

Πίνακας 9 : Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 0 - 1.

Εκτός της προσθήκης πάρκου από τη συστάδα 3, οι υπόλοιπες εναλλακτικές είναι ασύμφορες και επιφέρουν μεγάλο ρίσκο για πολύ μικρά κέρδη. Μάλιστα, η προσθήκη απλώς ενός πάρκου από τη συστάδα 0 κατά μέσο όρο αυξάνει τον CoV. Παρόλο που μια προσθήκη πάρκου της συστάδας 3 έχει πρακτικά εγγυημένα αποτελέσματα (πιθανότητα βελτίωσης του CoV ίση με 0.996), τα κέρδη δεν είναι μεγάλα και κυμαίνονται γύρω από το 2,1%.

2. Αρχικές συστάδες 0 / 2 (Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους / Ποικιλία Τιμών Μικρού Εύρους)

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C0:4, C2:4	C0	0.547	0.589	0.5876	-0,24	-0.0238	0.0171
	C1	0.746		0.5821	-1,15	-0.0227	0.0042
	C2	0.496		0.5861	-0,48	-0.03	0.0065
	C3	1		0.5722	-2,84	-0.0256	-0.009

Πίνακας 10: Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 0 - 2.

Και σε αυτή την περίπτωση, η προσθήκη από τη συστάδα 3 είναι η μόνη συμφέρουσα επιλογή, επιφέροντας με εγγύηση μείωση του CoV κατά 3%. Γενικότερα, τα αποτελέσματα είναι αδύναμα, ακόμα και για την προσθήκη από την έμπιστη συστάδα 3, καθώς θα φανεί με τις επόμενες περιπτώσεις πως η τιμή αυτή του CoV δεν είναι ιδανική και μπορεί με άλλες προσθήκες πάρκων να μειωθεί περαιτέρω.

3. Αρχικές συστάδες 0 / 3 (Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους / Έντονη Νυχτερινή Δραστηριότητα)

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C0:4, C3:4	C0	0.729	0.5654	0.5629	-0,42	-0.041	0.0165
	C1	0.67		0.5626	-0,48	-0.022	0.011
	C2	0.457		0.5581	-1,27	-0.03	0.0047

Πίνακας 11: Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 0 - 3.

Σε αυτή την περίπτωση καμία προσθήκη δεν είναι ιδιαίτερα ασφαλής, αλλά για ένα πάρκο από τη συστάδα 2, ο CoV κυμαίνεται σε χαμηλά επίπεδα (0.558) συγκριτικά με τα υπόλοιπα σενάρια. Προσθήκες από τις άλλες δύο συστάδες δεν συμβάλλουν ιδιαίτερα στη μείωση του CoV και με μεγάλο παράγοντα ρίσκου, προτείνεται να αποφευχθούν.

4. Αρχικές συστάδες 1 / 2 (Απότομες, Μη Επαναλαμβανόμενες Μεταβολές / Ποικιλία Τιμών Μικρού Εύρους)

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C1:4, C2:4	C0	0.577	0.5725	0.5705	-0,35	-0.032	0.0223
	C1	0.607		0.5696	-0,5	-0.0231	0.0141
	C2	0.382		0.5699	-0,44	-0.0395	0.0115
	C3	1		0.5598	-2,2	-0.0227	-0.0057

Πίνακας 12: Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 1 - 2.

Σε αυτή την περίπτωση, η προσθήκη πάρκου από τη συστάδα 3 είναι μία εξαιρετική κίνηση. Ο CoV κατεβαίνει σε χαμηλά επίπεδα με απόλυτη βεβαιότητα, παρά τη μικρή ελάττωση (μείωση κατά ~2%). Αντιθέτως, οι προσθήκες από άλλες συστάδες δεν έχουν κάποιο ιδιαίτερο πλεονέκτημα που να αντισταθμίζει το ρίσκο που επιφέρει η προσθήκη τους, ενώ η προσθήκη από τη συστάδα 2 έχει πολύ κακές πιθανότητες μείωσης CoV (0,38).

Η παραπάνω προσθήκη οδηγεί στο τελικό χαρτοφυλάκιο με πάρκα από τις συστάδες 1, 2 και 3, που όπως αναφέρθηκε στην ενότητα 4.8 είναι από τους πιο στιβαρούς συνδυασμούς.

5. Αρχικές συστάδες 1 / 3 (Απότομες, Μη Επαναλαμβανόμενες Μεταβολές / Έντονη Νυχτερινή Δραστηριότητα)

Βάση Χαρτοφυλάκιο u	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C1:4, C3:4	C0	0.955	0.6055	0.5823	-3,82	-0.0782	-0.0008
	C1	0.715		0.5984	-1,17	-0.0313	0.0161
	C2	0.998		0.5795	-4,28	-0.063	-0.007

Πίνακας 13 : Αποτελέσματα για αρχικό χαρτοφυλάκιο συστάδων 1 - 3.

Τα τελικά αποτελέσματα εδώ είναι τα πιο αδύναμα, κάτι που είναι αναμενόμενο αφού οι συστάδες 1 και 3 από μόνες τους δεν συνδυάζονται καλά. Αποτελούνται και οι δύο από τα πάρκα εκείνα που παρουσιάζουν ιδιαίτερο προφίλ και αλλοτπρόσαλλη συμπεριφορά, επομένως ο ταυτόχρονος συνδυασμός τους προσφέρει μεγαλύτερη αβεβαιότητα.

Παρόλα αυτά, η προσθήκη ενός πάρκου από τις συστάδες 0 ή 2 έχει πρακτικά εγγυημένα αποτελέσματα (πιθανότητες βελτίωσης 0,998 και 0,955 αντίστοιχα), επιτυγχάνοντας ελάττωση του CoV κατά 4,3% και 2,6% αντίστοιχα. Ενδεχομένως η προσθήκη παραπάνω πάρκων να ωφεληθεί αυτό τον αρχικό συνδυασμό περισσότερο από άλλους.

6. Αρχικές συστάδες 2 / 3 (Ποικιλία Τιμών Μικρού Εύρους / Έντονη Νυχτερινή Δραστηριότητα)

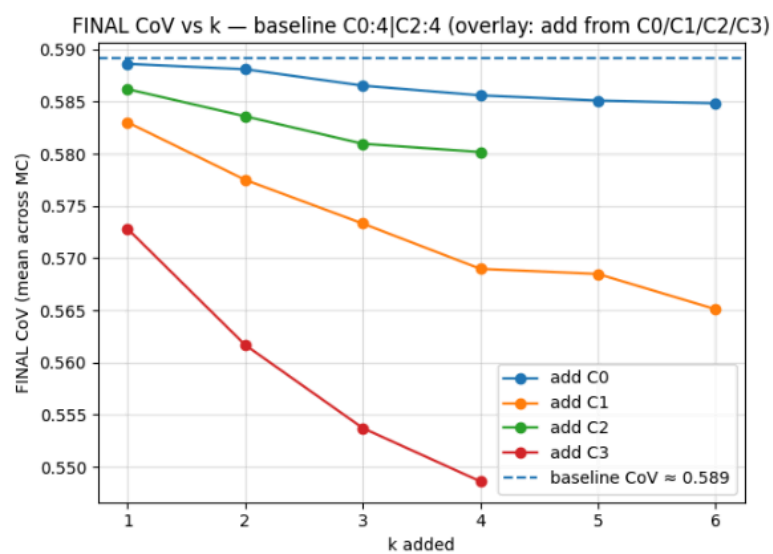
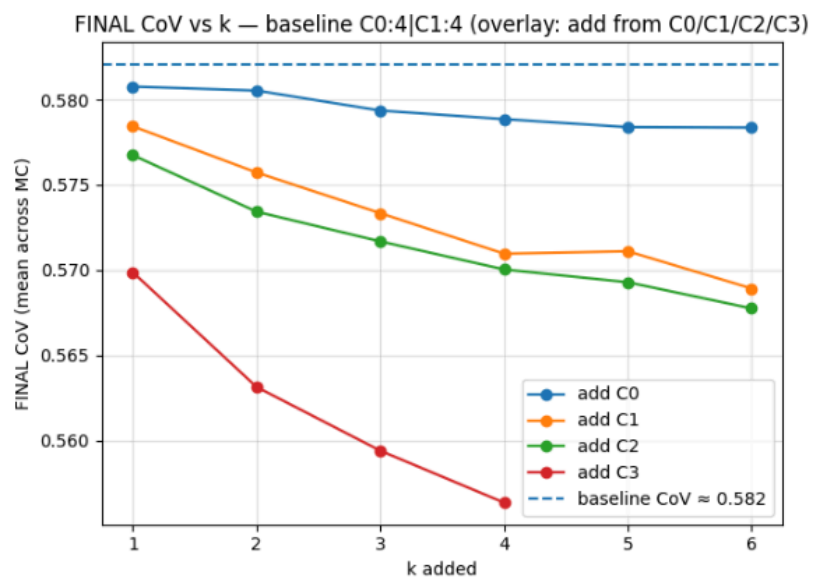
Βάση Χαρτοφυλάκιο u	Προσθήκη από	prob_impr	CoV_base	CoV_mean	dCoV%	dCoV_q05	dCoV_q95
C2:4, C3:4	C0	0.536	0.5491	0.5474	-0,31	-0.0336	0.0076
	C1	0.532		0.5485	-0,11	-0.0167	0.0173
	C2	0.506		0.5444	-0,86	-0.0378	0.0143

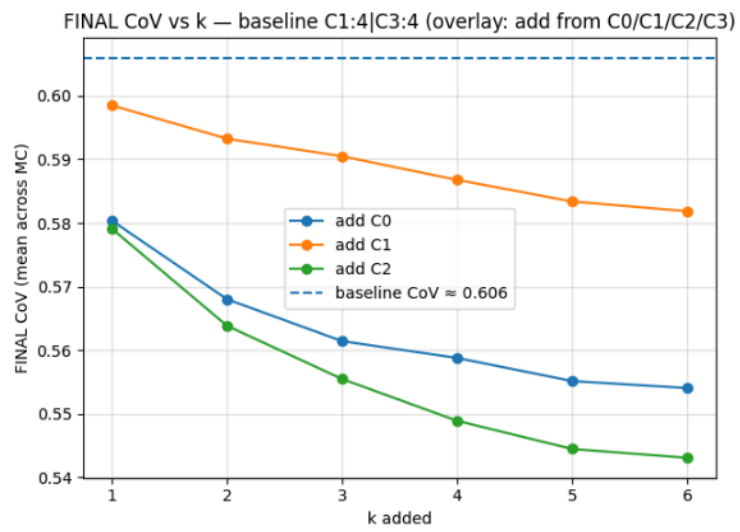
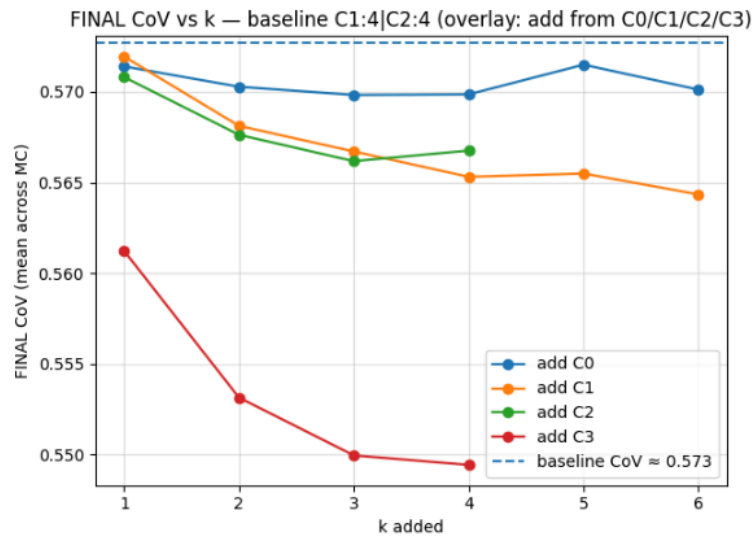
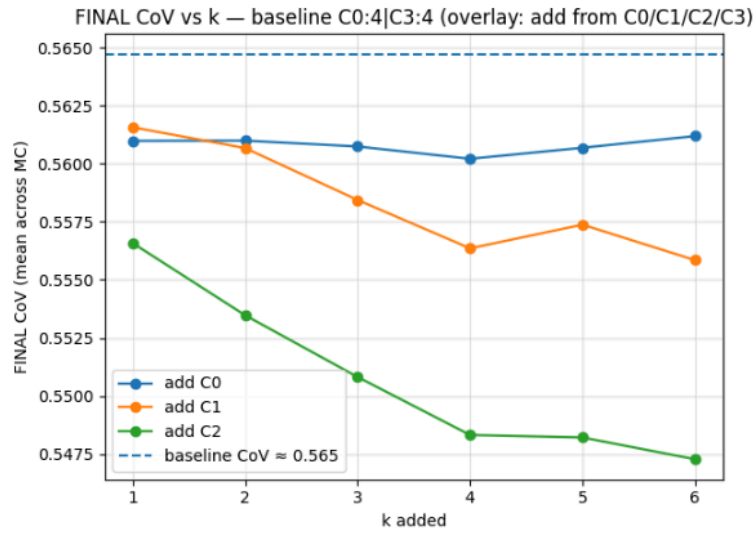
Πίνακας 14 : Αποτελέσματα για αρχικό χαρτοφυλάκιο 2 - 3.

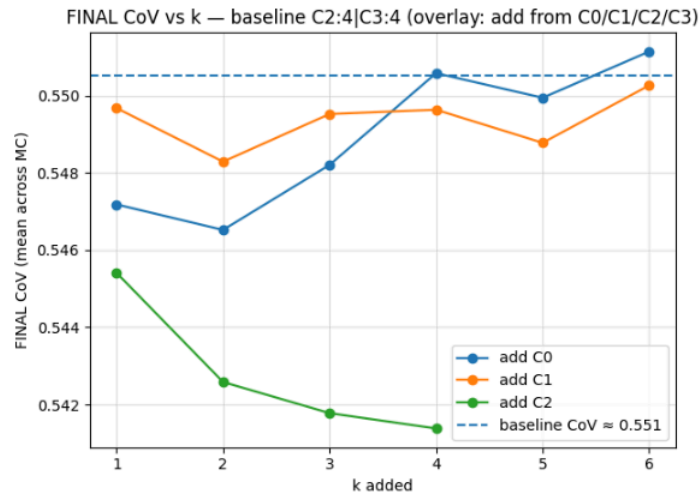
Η περίπτωση αυτή δεν επιτρέπει, πρακτικά, καμία περαιτέρω κίνηση. Ο αρχικός μέσος CoV είναι ήδη σε εξαιρετικά επίπεδα (0,549) προτού ερευνηθεί το ενδεχόμενο περαιτέρω προσθηκών, και πράγματι οποιαδήποτε προσθήκη συνοδεύεται από μεγάλο ρίσκο (πιθανότητες αντίστοιχες με ρίψη νομίσματος για όλες τις συστάδες) και μικρά κέρδη (μέγιστη μέση μείωση του CoV κατά ~0,86%).

Επιβεβαιώνεται η εκτίμηση που είχε γίνει στην ενότητα 4.9.1 για τον συνδυασμό 2 - 3, καθώς και για το απλό σενάριο προσθήκης πάρκου σε βασικό χαρτοφυλάκιο μόνο 4 πάρκων (είτε ξεκινούσαμε από την 2 είτε από την 3) έφερνε έντονη μείωση του CoV με μεγάλη πιθανότητα.

Και σε αυτή την ενότητα, θα ελεγχθεί η μείωση του CoV βάσει του αριθμού των πάρκων που προστίθενται στο βασικό χαρτοφυλάκιο. Προκύπτουν τα εξής διαγράμματα:







Εικόνα 43: Μείωση CoV για όλα τα σενάριο αρχικών χαρτοφυλακίων, συναρτήσει του αριθμού προστιθέμενων πάρκων.

Σε γενικές γραμμές, η αναλογία έχει τηρηθεί και για χαρτοφυλάκια αυξημένου αριθμού πάρκων. Παρόλα αυτά, ενδιαφέρον έχουν οι παρατηρήσεις πως για προσθήκη 3 ή και περισσότερων καινούριων πάρκων, ο CoV ενδέχεται να αυξηθεί, ή ακόμα και να ταλαντωθεί. Αυτό συμβαίνει για παράδειγμα στην περίπτωση αρχικού χαρτοφυλακίου 0 / 3, όπου η προσθήκη 5 πάρκων αυξάνει τον CoV, ενώ η προσθήκη 6 πάρκων αντίθετως μειώνει τον CoV, και μάλιστα σε χαμηλότερη στάθμη από ότι για προσθήκη 4 πάρκων.

Παρόμοια κατάσταση παρατηρείται στην περίπτωση αρχικού χαρτοφυλακίου 2 / 3, όπου 5 πάρκα από τη συστάδα 1 είναι πιο ευνοϊκά από 3 πάρκα της ίδιας συστάδας.

Εννοείται πως αυτά τα αποτελέσματα, όπως και στην ενότητα 4.9.1, για να είναι επιλέξιμα θα πρέπει να ακολουθούνται και από ικανοποιητική πιθανότητα μείωσης του CoV, σε διαφορετική περίπτωση το ρίσκο αυξάνεται πολύ και η επένδυση ενδέχεται να αποβεί ανωφελής ή και καταστροφική. Μπορεί επομένως να επαναληφθεί το πείραμα, με την προσθήκη 4 πάρκων ως ενδιάμεση επιλογή, και να παρατηρηθούν τα πιθανοτικά αποτελέσματα.

Ακολουθεί ο συμπυκνμένος πίνακας κατόπιν των 4 προσθηκών, κατανεμημένος κατά αύξοντα CoV:

Βάση Χαρτοφυλακίου	Προσθήκη από	prob_ impr	CoV_base	CoV_mean	dCoV_%	dCoV_ q05	dCoV_q95
C0:4 - C1:4	C0	0.479	0.582	0.579	-0,53	-0.053	0.035
	C1	0.678		0.572	-1,72	-0.04	0.016
	C2	0.708		0.57	-2,06	-0.046	0.018
	C3	0.997		0.557	-4,23	-0.044	-0.008
C0:4 - C2:4	C0	0.525	0.588	0.585	-0,39	-0.04	0.031
	C1	0.942		0.569	-3,25	-0.04	0.000
	C2	0.603		0.58	-1,39	-0.036	0.015
	C3	1		0.547	-6,8	-0.058	-0.024
C0:4 - C3:4	C0	0.498	0.564	0.56	-0,74	-0.052	0.034
	C1	0.677		0.556	-1,31	-0.032	0.015
	C2	0.756		0.548	-2,75	-0.048	0.012
C1:4 - C2:4	C0	0.541	0.573	0.57	-0,47	-0.042	0.035
	C1	0.632		0.565	-1,38	-0.04	0.02
	C2	0.531		0.567	-1,05	-0.043	0.027
	C3	1		0.549	-4,15	-0.039	-0.009
C1:4 - C3:4	C0	0.976	0.605	0.559	-7,62	-0.092	-0.006
	C1	0.781		0.586	-3,19	-0.063	0.018
	C2	1		0.549	-9,31	-0.09	-0.022
C2:4 - C3:4	C0	0.514	0.55	0.548	-0,31	-0.035	0.019
	C1	0.518		0.549	-0,16	-0.025	0.022
	C2	0.552		0.541	-1,6	-0.044	0.02

Πίνακας 15 : Αποτελέσματα για όλους τους συνδυασμούς, για προσθήκη 4 πάρκων.

Με αφετηρία αυτόν τον πίνακα, εξάγονται τα εξής συμπεράσματα:

- Επιβεβαιώνεται για άλλη μια φορά η τεράστια σημασία που κατέχει η συστάδα 3 (Νυχτερινή Δραστηριότητα) στην ποικιλία, καθώς οποιοδήποτε σενάριο δεν την περιέχει ήδη, οφείλει να την προσθέσει. Προσφέρει με απόλυτη συνέπεια μείωση του CoV και μάλιστα πρακτικά εγγυημένα (πιθανότητες από 0.997 μέχρι 1). Η συστάδα αυτή είναι ο ισχυρότερος «διαφοροποιητής».
- Η προσθήκη πάρκων από τη συστάδα 2 (Περιοδικότητα) είναι πολύ συχνά η δεύτερη καλύτερη επιλογή, όταν δεν υπάρχουν στο αρχικό χαρτοφυλάκιο ήδη πάρκα από αυτή τη συστάδα. Και για αυτή τη συστάδα όμως φαίνεται πως το όριο πάρκων σε ένα χαρτοφυλάκιο είναι 4.
- Η συνύπαρξη παραπάνω από 4 πάρκων με καταγωγή το ίδιο cluster εμπεριέχει μεγάλο ρίσκο αύξησης του CoV και μάλιστα οι τιμές του CoV που μπορεί να προκύψουν, σε περίπτωση μείωσης του, είναι πολύ αδύναμες. Επομένως μπορεί να θεωρηθεί αυτό ως ένα πρακτικό ανώτερο όριο του εκάστοτε χαρτοφυλακίου. Η φαινομενική εκτίμηση επομένως, ότι περισσότερες προσθήκες πάρκων από την ίδια συστάδα μειώνουν τον CoV ισχύει, αλλά δεν δείχνουν ολόκληρη την εικόνα. Προσθήκες περισσότερων από 4 πάρκων δεν προσφέρουν

πολύ καλύτερες πιθανότητες από τη ρίψη ενός νομίσματος, και δε μπορούν να θεωρηθούν ευνοϊκές κινήσεις.

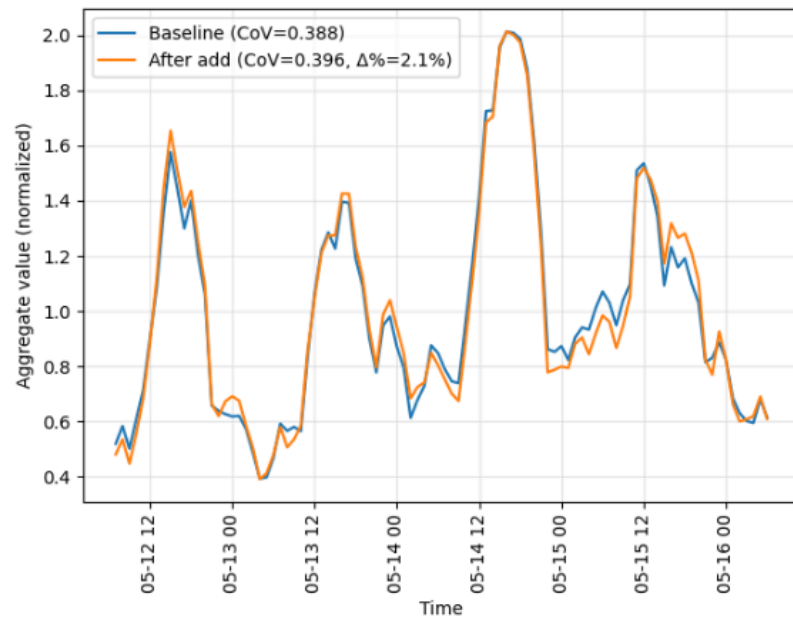
Παρόλα αυτά:

- Ο ελάχιστος μέσος CoV (0.5414) μπορεί να επιτευχθεί με προσθήκη επιπλέον 4 πάρκων από τη συστάδα 2 στο αρχικό χαρτοφυλάκιο 4 πάρκων από τις συστάδες 2 και 3. Η μείωση όμως του CoV στα επίπεδα αυτά είναι ιδιαίτερα επισφαλής κίνηση, καθώς τα ενδεχόμενα μείωσης ή αύξησης του CoV είναι σχεδόν ισοπίθανα (0.55). Αν μια στρατηγική μεγάλου κέρδους αλλά και μεγάλου ρίσκου (high risk - high reward) είναι επιθυμητή, μπορεί να ακολουθηθεί αυτό το μονοπάτι. Διαφορετικά, η συνύπαρξη 4 πάρκων από τις συστάδες 2 και 3 προσφέρει ήδη αρκετά χαμηλό CoV (0.55), οπότε μια επιπλέον στρατηγική είναι να μην προστεθεί κανένα επιπλέον πάρκο.
- Εκτός ελαχίστων εξαιρέσεων, η συντριπτική πλειοψηφία των συνδυασμών που επιτυγχάνουν ικανοποιητικά επίπεδα CoV με μικρό ρίσκο περιορίζεται στα 8 πάρκα, καθώς η περαιτέρω προσθήκη 4 πάρκων επιφέρει μεγάλο ρίσκο όσον αφορά τη μείωση του CoV. Οι ιδιαίτερες βέλτιστες περιπτώσεις, οι οποίες θα αναφερθούν σε επόμενη παρατήρηση, επιτρέπουν και 12 πάρκα. Τα καλύτερα "συμβιβαστικά" μονοπάτια τείνουν να έχουν πάρκα από 2-3 συστάδες, πάντα με την παρουσία της τρίτης· συχνά τα 8 πάρκα είναι αρκετά, ενώ το πέρασμα στα 12 έχει μικρό αλλά υπαρκτό επιπλέον όφελος όταν η τρίτη συστάδα επιλέγεται προσεκτικά.
- Βάσεις με υψηλότερο αρχικό CoV έχουν μεγαλύτερο περιθώριο για βελτίωση. Όσο καλύτερη είναι ήδη η βάση, τόσο μικρότερο το περιθώριο κέρδους και τόσο μεγαλύτερο είναι και το ρίσκο η προσθήκη να δράσει καταστροφικά στον CoV.
- Ο καλύτερος και ασφαλέστερος συνδυασμός είναι 4 πάρκα από καθεμία από τις συστάδες 1 / 2 / 3, για τελικό σύνολο 12 πάρκων. Αρχίζοντας από τη συστάδα 3 μέσου CoV 0.6843, γίνεται με 100% πιθανότητα βελτίωσης να επιτευχθεί μέσος CoV 0.607, προσθέτοντας πάρκα από τη συστάδα 1. Ως δεύτερο βήμα, ξανά με 100% πιθανότητα βελτίωσης, γίνεται να επιτευχθεί τιμή CoV γύρω στο 0.5492, με τις προσθήκες 4 πάρκων από τη συστάδα 2. Έτσι ο αρχικός μέσος CoV υφίσταται μείωση κατά σχεδόν 20%, ένα πολύ μεγάλο ποσοστό.
- Ένας επίσης εξαιρετικός συνδυασμός, αλλά λιγότερο ασφαλής, είναι 4 πάρκα από τις συστάδες 0 / 2 / 3. Ξεκινώντας από 4 πάρκα της συστάδας 0 με μέσο CoV 0.6207, προστίθενται 4 πάρκα από την συστάδα 2, κίνηση η οποία έχει αρκετά μεγάλη πιθανότητα να βελτιώσει τον CoV (0.877) στην τιμή 0.5891. Στη συνέχεια θα προστεθούν τα πάρκα της συστάδας 3, τα οποία έχουν 100% πιθανότητα να βελτιώσουν τον CoV, ώστε να επιτευχθεί η τελική τιμή του 0.5478. Παρατηρείται μια μείωση 12%, ικανοποιητική μεν αλλά υστερεί σε σχέση με την παραπάνω επιλογή.

Και στις περιπτώσεις της ένωσης 3 συστάδων, μπορεί να ελεγχθεί και γραφικά η μείωση της διακύμανσης στις χρονοσειρές των χαρτοφυλακίων.

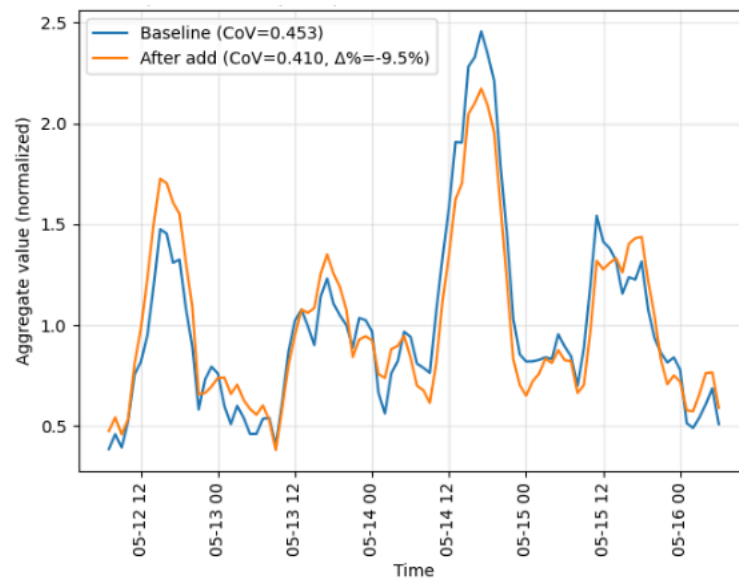
- Συνδυασμός 1 / 2 / 3

Θα ακολουθηθεί η διαδρομή 3 -> 1 -> 2, και για προσθήκη 1 πάρκου από τη συστάδα 2, προκύπτει:



Εικόνα 44: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 2.

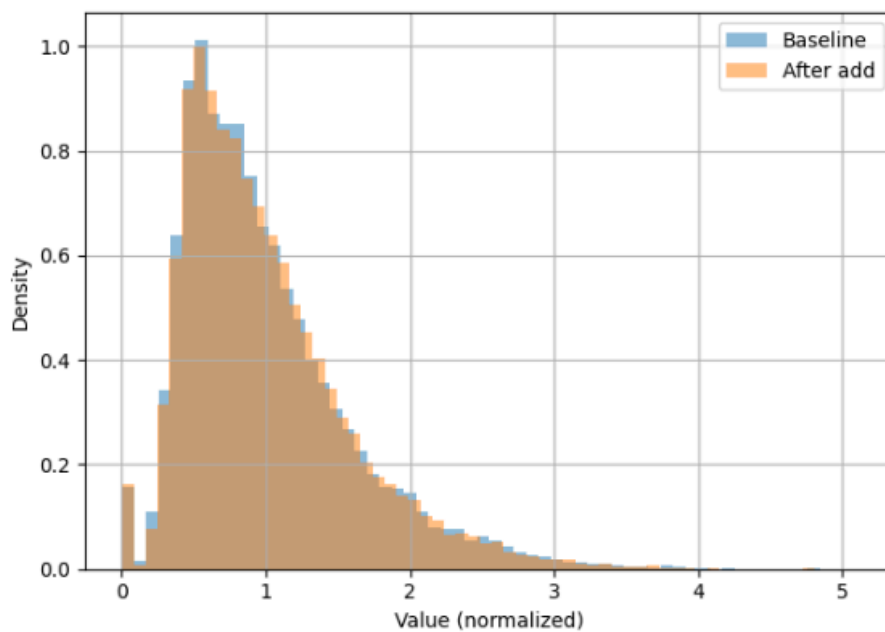
Ακολουθεί και το διάγραμμα της χρονοσειράς για προσθήκη 4 πάρκων:



Εικόνα 45: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 2.

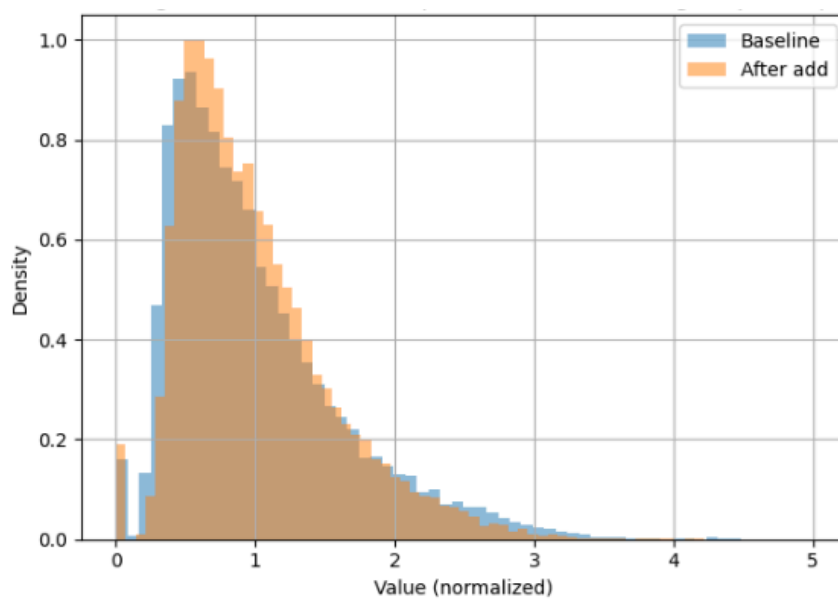
Επιβεβαιώνονται τα αποτελέσματα του πίνακα πως ο συγκεκριμένος συνδυασμός αποτελεί ιδανική επιλογή για ασφαλή και επικερδή επένδυση, καθώς ομαλοποιούν σε καλό βαθμό τη χρονοσειρά του χαρτοφυλακίου.

Ακολουθούν και τα ιστογράμματα της πυκνότητας πιθανότητας της χρονοσειράς:



Εικόνα 46: Κατανομή τιμών έπειτα από προσθήκη 1 πάρκου.

και για προσθήκη 4 πάρκων:

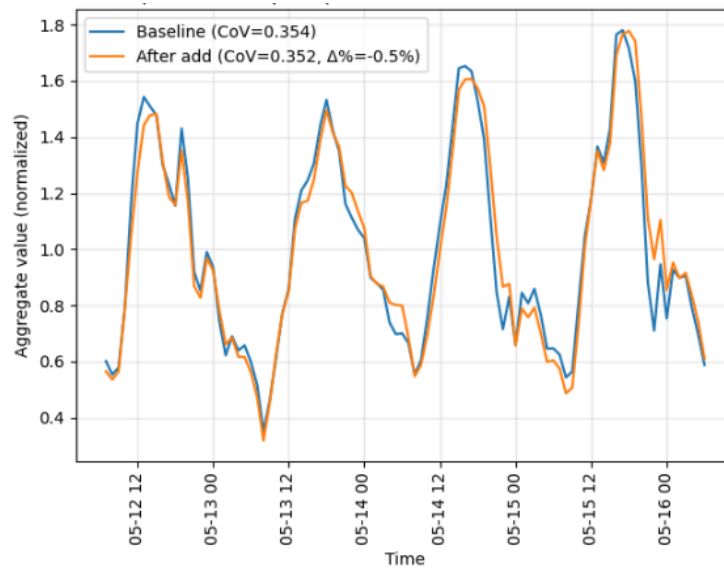


Εικόνα 47: Κατανομή τιμών έπειτα από προσθήκη 4 πάρκων.

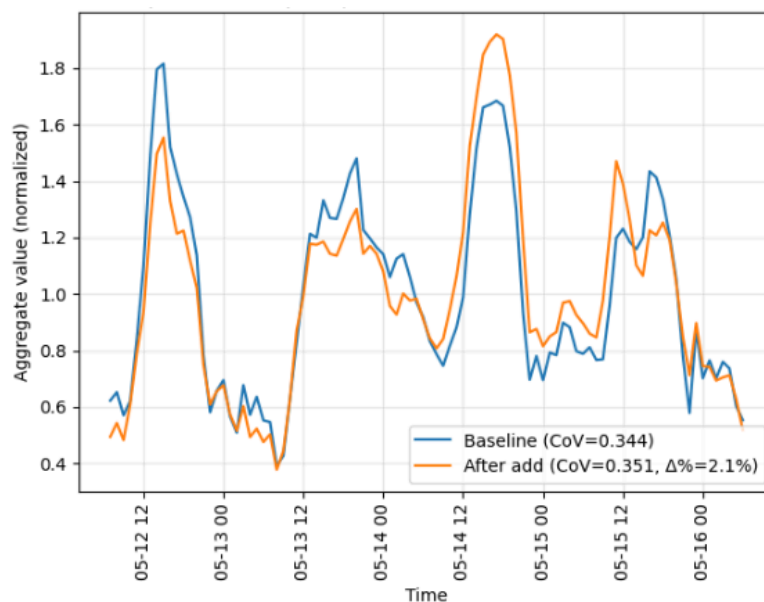
Εμφανής η θετική επίδραση της διαφοροποίησης εντός του χαρτοφυλακίου, με τις τιμές των ταχυτήτων έντονα συγκεντρωμένες γύρω από το 1.

- Συνδυασμός 0 / 2 / 3

Σε αυτή την περίπτωση, θα ακολουθηθεί η διαδρομή 0 -> 2 -> 3, και προκύπτει:



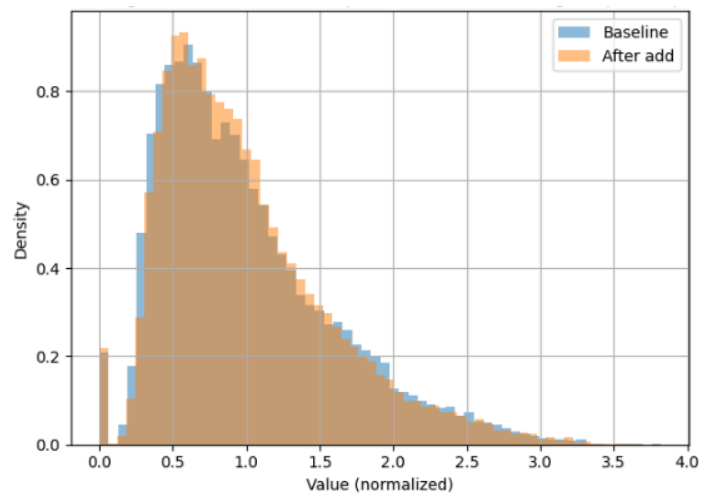
Εικόνα 48: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 1 πάρκου της συστάδας 3.



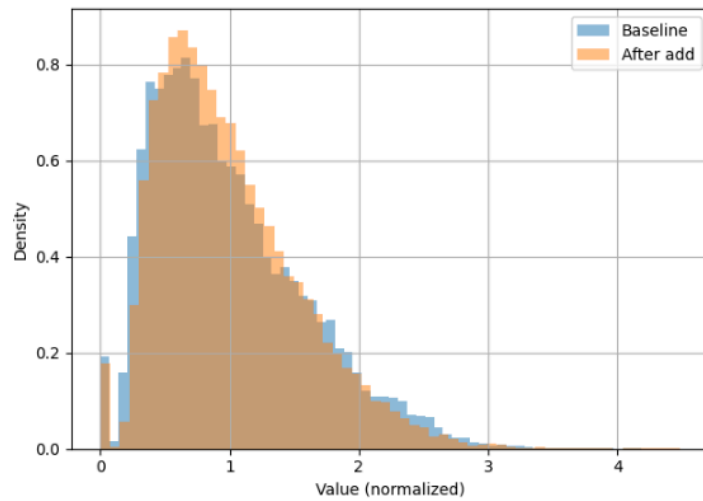
Εικόνα 49: Χρονοσειρά του χαρτοφυλακίου για προσθήκη 4 πάρκων της συστάδας 3.

Η ποσοστιαία μείωση του CoV στο τυχαίο συγκεκριμένο παράδειγμα είναι μικρότερη, καθώς το αρχικό χαρτοφυλάκιο κυμαινόταν ήδη σε πιο χαμηλά επίπεδα. Το συμπέρασμα όμως μένει ίδιο και η τιμή του CoV μειώνεται, ομαλοποιώντας τη χρονοσειρά.

Ακολουθούν τα ιστογράμματα, για περαιτέρω οπτικοποίηση των αποτελεσμάτων:



Εικόνα 50: Κατανομή τιμών για προσθήκη 1 πάρκου της συστάδας 3.



Εικόνα 51: Κατανομή τιμών για προσθήκη 4 πάρκων της συστάδας 3.

Και σε αυτή την περίπτωση, φαίνεται πως έχει επιτευχθεί καλή εξομάλυνση της χρονοσειράς, και η προσθήκη των επιπλέον πάρκων επιφέρει αναμενόμενα καλύτερα αποτελέσματα, παραμένει όμως η δεύτερη καλύτερη επιλογή για διαφοροποίηση χαρτοφυλακίου με χρήση τριών συστάδων.

Κεφάλαιο 6 : Συμπεράσματα και Μελλοντικές Επεκτάσεις

6.1 Συμπεράσματα

Η παρούσα διπλωματική είχε ως σκοπό να εξετάσει τη δυνατότητα της εκ των προτέρων εκτίμησης της απόδοσης ενός χαρτοφυλακίου ανανεώσιμων πηγών ενέργειας με την αποκλειστική χρήση μεθόδων εξόρυξης στατιστικών στοιχείων και συσταδοποίησης. Πιο συγκεκριμένα, διερευνήθηκε η μείωση του CoV για κάθε πιθανό συνδυασμό συστάδων, εφαρμόζοντας τη μέθοδο Προσομοιώσεων Monte Carlo για την εύρεση των χαμηλότερων μέσων τελικών επιπέδων CoV. Διερευνήθηκε επίσης η πιθανότητα του κάθε συνδυασμού συσταδικών προσθηκών να καταλήξει πράγματι στο μέσο επίπεδο CoV που εκτιμήθηκε, ώστε ο χρήστης - έμπορος - φορέας να μπορεί να έχει επίγνωση του ενδεχόμενου ρίσκου που επιφέρει η αγορά ενός τυχαίου πάρκου μιας συγκεκριμένης συστάδας. Εκείνος θα μπορεί είτε να το αποδεχτεί για χάρη της μέγιστης δυνατής μείωσης της αβεβαιότητας για την παραγωγή ενέργειας και την εκτίμηση των μελλοντικών ταχυτήτων του αιολικού πάρκου, είτε να το απορρίψει, επιλέγοντας να προσθέσει πάρκο ή πάρκα από εκείνες τις συστάδες που εγγυώνται καλά αποτελέσματα λόγω της διαφοροποίησης που θα προσφέρουν στο χαρτοφυλάκιο που ήδη διαθέτει ο χρήστης.

Ένα γρήγορο πλεονέκτημα της μεθόδου είναι η πλήρης απουσία χρήσης τεχνικών προβλέψεων και δεικτών λάθους. Παρόλο που το θέμα της διπλωματικής σχετίζεται άμεσα με τον επιστημονικό τομέα των προβλέψεων, η μεθοδολογία που παρουσιάστηκε δεν απαιτεί την πρόβλεψη μελλοντικών τιμών, καθώς η “αβεβαιότητα” που είναι συνώνυμη με τις προβλέψεις ουσιαστικά καλύπτεται από τις Προσομοιώσεις Monte Carlo. Μπορούν με μεγάλη ακρίβεια να αποφανθούν για το ποιες προσθήκες είναι κατά μέσο όρο ευνοϊκές και ποιες κατά μέσο όρο θα δράσουν αρνητικά, ή ακόμα και καταστροφικά σε κάποιες περιπτώσεις που παρατηρήθηκαν, στο ήδη υπάρχον χαρτοφυλάκιο.

Αυτή η μεγαλύτερη βεβαιότητα για το τελικό αποτέλεσμα είναι υψίστης σημασίας για τους ΦοΣΕ, και γενικότερα για μεγάλους και μικρούς εμπόρους στην ηλεκτρική αγορά της Ελλάδας, διότι με τη χρήση της σχετικά απλής αυτής μεθοδολογίας, θα έχουν καλύτερη εποπτεία των προφίλ πάρκων στα οποία θα πρέπει να εστιάσουν και να επενδύσουν. Με αυτό τον τρόπο, θα μπορέσει να επιτευχθεί ο βασικός τους στόχος, που δεν είναι άλλος από την ελάττωση των χρηματικών ποινών και των ρητρών που τους επιβάλλονται στη συνηθισμένη περίπτωση οι προβλέψεις παραγωγής ενέργειάς τους να είναι εσφαλμένες. Στο χρηματιστήριο, οποιαδήποτε μείωση της αβεβαιότητας είναι φυσικά καλοδεχούμενη, επομένως ακόμα και πώση των επιπέδων του CoV κατά π.χ. 3%, ενώ φαντάζει μικρό και ασήμαντο, μπορεί να αποτελεί καταλυτικό παράγοντα για την επίτευξη των ενεργειακών στόχων μιας εταιρείας ή ενός εμπόρου. Δεδομένης επίσης της μεγάλης βεβαιότητας με την οποία μπορούν να επιτευχθούν παρόμοιες μειώσεις, είναι προφανές πως η μεθοδολογία που αναπτύχθηκε μπορεί να είναι ιδιαίτερα χρήσιμη στον κλάδο των ενεργειακών οικονομικών.

Έχοντας ολοκληρώσει τα πειράματα και βλέποντας τα αποτελέσματα, συμπεραίνουμε πως αυτά ανταποκρίθηκαν σε πολύ μεγάλο βαθμό στις αρχικές εκτιμήσεις, οι οποίες έγιναν βάσει της συμπληρωματικότητας των συστάδων των πάρκων. Οι συστάδες εκείνες που έχουν μικρότερο βαθμό επικάλυψης μεταξύ τους σε σχέση με τις άλλες, όπως η “Έντονη Πρωινή Δραστηριότητα Μεγάλου Εύρους” και η “Απότομες, Μη Επαναλαμβανόμενες Μεταβολές”, συνδυάζονται εξαιρετικά και αποφέρουν στις περισσότερες περιπτώσεις τους χαμηλότερους αριθμούς CoV. Όχι μόνο αυτό, αλλά είχαν πολύ μεγάλες πιθανότητες, σε μερικές περιπτώσεις υπήρχε και βεβαιότητα, πως τα επίπεδα του CoV θα κυμαίνονταν τόσο χαμηλά.

Ακολουθούν τα κυριότερα συμπεράσματα, όπως αυτά αναπτύχθηκαν στις ενότητες 5.9.1 και 5.9.2:

- Υπήρξε επιβεβαίωση του προφανούς συμπεράσματος πως σε γενικές γραμμές, οι περισσότερες προσθήκες πάρκων δρουν ευνοϊκά στα χαρτοφυλάκια, καθώς αυξάνεται η διαφοροποίησή του. Οι πιθανότητες επίσης επίτευξης του αναμενόμενου CoV δεν άλλαξαν σε μεγάλο βαθμό με την προσθήκη περισσότερων πάρκων.
- Η σειρά με την οποία προστίθενται τα πάρκα από τις συστάδες δεν επηρεάζει την τελική προσεγγιστική τιμή του CoV, επηρεάζει όμως την πιθανότητα να επιτευχθεί η τελική αυτή τιμή. Θεωρώντας ως παράδειγμα τον τελικό συνδυασμό 1 / 2 / 3 που αναφέρθηκε παραπάνω, μπορούμε να καταλήξουμε στα τελικά 12 πάρκα είτε με το μονοπάτι 3 -> 1 -> 2, το οποίο έχει 100% βεβαιότητα, είτε με το 2 -> 3 -> 1,

του οποίου το τελευταίο βήμα έχει πιθανότητα βελτίωσης του CoV μόλις 0.51. Αυτό σημαίνει ότι επιτυγχάνεται πολύ μεγάλη μείωση του CoV μόνο με ένα βήμα, το οποίο φυσικά περιορίζει το περιθώριο για περαιτέρω κέρδη στο δεύτερο βήμα. Έτσι βγαίνει το συμπέρασμα πως στόχος για σταθερά κέρδη είναι η προσθήκη των 4 πάρκων με συγκεκριμένη σειρά, ώστε κάθε βήμα να εγγυάται σχετικά μέτρια κέρδη, τα οποία όμως δε θα είναι απότομα και δε θα επιτυγχάνονται σε ένα μόνο βήμα. Πρακτικά προστίθενται σε κάθε βήμα πάρκα που ναι μεν βελτιώνουν την διαφοροποίηση, δεν την εξαντλούν όμως, καθώς αυτό θα καταστήσει την προσθήκη πάρκων στο τρίτο βήμα ριψοκίνδυνη.

- Τα πάρκα εκείνα με το πιο ασυνήθιστο προφίλ, όπως αυτά της συστάδας 3 “Έντονη Νυχτερινή Δραστηριότητα”, θα πρέπει να κατέχουν κρίσιμη θέση σε όλα τα χαρτοφυλάκια. Πράγματι, και οι τέσσερις επιλεγμένοι βέλτιστοι συνδυασμοί περιέχουν τα πάρκα της συγκεκριμένης συστάδας. Αυτό συμβαίνει καθώς τα πάρκα αυτά παρουσιάζουν τις μεγαλύτερες αποκλίσεις στη συμπεριφορά τους σε σχέση με τα άλλα πάρκα. Ενώ οι υπόλοιπες συστάδες διαφέρουν κυρίως στην κατανομή των τιμών της χρονοσειράς τους, τα πάρκα της συστάδας 3 λειτουργούν διαφορετικές ώρες. Το γεγονός αυτό τα κάνει πολύτιμα, καθώς μπορούν να εξισορροπήσουν τα ανενεργά βράδια των υπόλοιπων πάρκων του χαρτοφυλακίου με τα εκρηκτικά, δραστήρια δικά τους, εγγυώμενα 24ωρη κάλυψη ενεργειακής παραγωγής.
- Παρά την θετική επίδραση των όσο δυνατόν περισσότερων προστιθέμενων πάρκων από την ίδια συστάδα σε ένα χαρτοφυλάκιο, καθώς η προσθήκη αυτή επιφέρει μεγαλύτερη διαφοροποίηση εντός του χαρτοφυλακίου, φαίνεται πως για 4 πάρκα ίδιας συστάδας η μέθοδος αρχίζει να αποδίδει ολοένα και πιο επισφαλής αποτελέσματα. Με άλλα λόγια, αν επιλεγθεί η προσθήκη παραπάνω από 4 πάρκων στο ίδιο χαρτοφυλάκιο, ελλοχεύει ο κίνδυνος υψηλού ρίσκου, το οποίο σε ορισμένες περιπτώσεις μπορεί να είναι και σε δυσμενείς πιθανότητες (π.χ. για 4 εκάστοτε πάρκα από τις συστάδες 1 και 2, η προσθήκη ακόμα και ενός πάρκου από τη συστάδα 2 έχει πιθανότητα μόλις 0,38 να μειώσει τον CoV).
- Στην εκδοχή του πειράματος με την σύσταση ενός χαρτοφυλακίου έπειτα από δύο διαδοχικές κινήσεις προσθήκης πάρκων, η επίτευξη του ελάχιστου CoV αποτελεί επικίνδυνη κίνηση από πλευράς ρίσκου. Η πιθανότητα επίτευξης αυτών των επιπέδων CoV έχει πιθανότητα περίπου 0.55, οιωνοί ελάχιστα καλύτεροι από τη ρίψη ενός νομίσματος, και σίγουρα δεν είναι ένα ρίσκο που πολλοί φορείς και έμποροι θα είναι πρόθυμοι να πάρουν. Από τα υπόλοιπα αποτελέσματα, είναι φανερά πολύ ευκολότερη η επίτευξη ενός λίγο πιο μέτριου CoV, ο οποίος όμως έχει πιθανότητες από 0.99 έως και απόλυτη βεβαιότητα 1 να επιτευχθεί, κατά μέσο όρο. Από στρατηγική πλευρά, αυτή η κίνηση είναι πολύ πιο λογική και συμφέρουσα, καθώς με την προσθήκη επιπλέον πάρκων ενδέχεται η πιθανότητα επιπλέον βελτίωσης του CoV να παραμένει πολύ υψηλή, ακόμη και να εξακολουθήσει να υπάρχει βεβαιότητα. Στην περίπτωση, για παράδειγμα, των συστάδων 1 “Απότομες, Μη Επαναλαμβανόμενες Μεταβολές”, 2 “Ποικιλία Τιμών Μικρού Εύρους” και 3 “Έντονη Νυχτερινή Δραστηριότητα”, μπορεί να επιτευχθεί μείωση του CoV ακόμα και με την προσθήκη 4 πάρκων της συστάδας 2, με το αρχικό χαρτοφυλάκιο να αποτελείται από 12 πάρκα. Είναι χωρίς αμφιβολία η βέλτιστη επιλογή, χωρίς να επιφέρει τον ελάχιστο δυνατό CoV.

6.2 Μελλοντικές Επεκτάσεις

Η παρούσα εργασία δίνει πρόσφορο έδαφος για μελλοντικές επεκτάσεις και βελτιώσεις της μεθοδολογίας που χρησιμοποιεί, που μπορούν να εξελιχθούν σε διάφορες κατευθύνσεις. Ένα σημείο που ενδείκνυται για βελτίωση είναι η μελέτη ενός μεγαλύτερου συνόλου δεδομένων, καθώς τα αιολικά πάρκα που μελετήθηκαν ήταν 34. Ικανοποιητικός αριθμός για τη μελέτη και την επιβεβαίωση ορθής λειτουργίας του κορμού της μεθόδου, αλλά για εγκυρότερα αποτελέσματα απαιτούνται παραπάνω δεδομένα. Ενδεχομένως η εφαρμογή της μεθοδολογίας σε κάποια άλλη χώρα της Ευρωπαϊκής Ένωσης που διαθέτει ευρύτερο δίκτυο ανεμογεννητριών, όπως για παράδειγμα είναι η Γερμανία, να είναι ένα κατάλληλο περιβάλλον για την εξονυχιστική έρευνα και βελτίωση της εν λόγω μεθόδου.

Ταυτόχρονα, η ανάλυση μεγαλύτερου αριθμού από αιολικά πάρκα μπορεί να αναδείξει πιο ιδιαίτερες συστάδες, με αποτέλεσμα τα ποικίλα προφίλ των πάρκων να χωριστούν ακόμα πιο καλά και ουσιαστικά.

Η εφαρμογή διαφορετικών στατιστικών ή μετεωρολογικών στοιχείων ως βάση για την συσταδοποίηση, όπως είναι π.χ. η υγρασία ή οι μετεωρολογικές προβλέψεις, μπορεί να βοηθήσει την εύρεση ενός καλύτερου συνδυασμού στοιχείων, που δε βασίζεται τόσο στις στατιστικές των χρονοσειρών αλλά στις μικροκλιματικές διαφορές μεταξύ των πάρκων.

Πέραν αυτών, η συμπερίληψη περισσότερων τύπων ΑΠΕ, όπως είναι τα φωτοβολταϊκά και οι υδροηλεκτρικές μονάδες, ενδέχεται να προσφέρουν επιπλέον διαφοροποίηση στο χαρτοφυλάκιο, και η εξέταση της επίδρασης των υβριδικών αυτών ειδών παραγωγής ενέργειας είναι μια επέκταση που παρουσιάζει ιδιαίτερο ενδιαφέρον.

Επεκτάσεις μπορούν να γίνουν επίσης και στη μεθοδολογία που ακολουθείται, χωρίς να αλλάξει ο κύριος στόχος και η ουσία της έρευνας. Η πανταχού παρουσία των τεχνικών μηχανικής μάθησης, όπως οι αλγόριθμοι *deep learning*, μπορούν να προσφέρουν μια άλλη οπτική στο ζήτημα της εκτίμησης αποδοτικότητας των χαρτοφυλακίων, που ενδέχεται να είναι πιο ολοκληρωμένη στο σύνολό της. Ταυτόχρονα, μπορούν να εφαρμοστούν διαφορετικές στοχαστικές μέθοδοι αντί των MCS, όπως είναι η μέθοδος *bootstrapping*, με στόχο την σύγκριση των αποδόσεων και την λήψη ενός συμπεράσματος για την βέλτιστη μεθοδολογία ανάλυσης των προς κατασκευή χαρτοφυλακίων.

Επίσης σημαντικές επεκτάσεις μπορούν να γίνουν και στην ανάλυση δεδομένων και εφαρμογών πραγματικού κόσμου. Η εφαρμογή της μεθόδου που αναπτύχθηκε σε πραγματικά χαρτοφυλάκια ενός ή πολλών ΦοΣΕ και η σύγκριση των αποτελεσμάτων χωρίς και με ένταξη της μεθόδου μπορεί να δείξει έμπρακτα την αξία που έχει το εργαλείο που συντέθηκε στην διπλωματική εργασία.

Βιβλιογραφία

- [1] Rosemary Potter, “Global CO₂ Emissions Surge Out of Control: This Alarming Spike Signals a Dangerous New Phase in the Climate Crisis”. Sustainability Times, 2025
- [2] European Union, EU's Climate Law presents a new way to get to 2040, https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1687
- [3] The Green Tank, Trends in electricity production – December 2024, 2025
- [4] FfE, European day-ahead electricity prices in 2024, 2025
<https://www.ffe.de/en/publications/european-day-ahead-electricity-prices-in-2024/>
- [5] Όμιλος ΑΔΜΗΕ, <https://admieholding.gr/el/omilos-admie/>
- [6] ΑΔΜΗΕ, ΚΕΦΑΛΑΙΟ 19 ΕΚΚΑΘΑΡΙΣΗ ΕΝΕΡΓΕΙΑΣ ΕΞΙΣΟΡΡΟΠΗΣΗΣ ΚΑΙ ΑΠΟΚΛΙΣΕΩΝ, <https://www.admie.gr/agora/rythmistiko-plaisio-agoras/kanonismos-agoras-exisorropisis/arthro-196-ypologismos-timis>
- [7] Nikolaos S. T., Santos, F.J., Pozo, D., & Usaola J . Optimal management of wind and solar energy resources, 2016.
- [8] Shanthababu Pandian, Time Series Analysis: Definition, Components, Methods, and Applications, AnalyticsVidhya, 2025
- [9] SciPy, The main SciPy namespace, https://docs.scipy.org/doc/scipy/reference/main_namespace.html
- [10] Luciano Telesca, Fabian Guignard, Mohamed Laib, Mikhail Kanevski, Analysis of temporal properties of wind extremes, 2018
- [11] Nenad S. Mitić, Saša N. Malkov, Mirjana M. Maljković Ružičić, Aleksandar N. Veljković, Ivan Lj. Čukić, Xin Lin, Minjie Lyu, Vladimir Brusić, Correlation-based feature selection of single cell transcriptomics data from multiple sources, 2025
- [12] Pawel Kuklik, Dominant Frequency, <https://pawelkuklik.com/dominant-frequency/>
- [13] Fernando Rosas, Pedro Mediano, When and how to use Lempel-Ziv complexity, 2019.
- [14] Robert B. Cleveland, William S. Cleveland, Jean E. McRae and Irma Terpenning, STL: A Seasonal-Trend Decomposition Procedure Based on Loess, 1990
- [15] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, 1996
- [16] Pedro Contreras, Fionn Murtagh, Algorithms for hierarchical clustering: an overview, 2011

- [17] Leonard Kaufman, Peter J. Rousseeuw, Finding Groups in Data An Introduction to Cluster Analysis, σ. 86-90, 1990
- [18] David L. Davies; Donald W. Bouldin, A Cluster Separation Measure, 1979.
- [19] Thomas Schultze, Lukas Thürmer, The importance of standardised approaches for experimental research, 2024.
- [20] Richard L. Sweeney, Joseph Wilske, Excess Variance in Decentralized Renewable Energy Investment, 2025
- [21] Samik Raychaudhuri, Introduction to Monte Carlo simulation, 2008
- [22] Iliyas Karim Khan, Hanita Binti Daud, Nooraini Binti Zainuddin, Rajalingam Sokkalingam, Muhammad Farooq, Muzammil Elahi Baig, Gohar Ayub, Mudasar Zafar, Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm, 2024
- [23] John Pavlopoulos, Georgios Vardakas, Aristidis Likas, Revisiting Silhouette Aggregation, 2024.