



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ**  
**ΤΜΗΜΑ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ**

**Αλγόριθμοι Μηχανικής Μάθησης για την ανάλυση και επεξεργασία  
δεδομένων εκπομπής ρύπων από τα θαλάσσια μέσα μεταφοράς**

**Machine Learning Algorithms for the Analysis and Processing of Emission Data in  
Maritime Transportation**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Ανδρέας Γιαννακόπουλος**  
1072266

**4<sup>η</sup> Κατεύθυνση Εμβάθυνσης**  
**«Συστήματα Βιώσιμων Μεταφορών και Διαχείρισης Έργων»**

**Επιβλέπων:**  
**ΟΙΚΟΝΟΜΟΥ ΠΟΛΥΧΡΟΝΗΣ, Αναπληρωτής Καθηγητής**

**ΠΑΤΡΑ, ΟΚΤΩΒΡΙΟΣ 2025**

Πανεπιστήμιο Πατρών, Τμήμα Πολιτικών Μηχανικών  
Ανδρέας Γιαννακόπουλος  
© 2025 – Με την επιφύλαξη παντός δικαιώματος

## ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία εξετάζει το κρίσιμο ζήτημα της περιβαλλοντικής επιβάρυνσης από τις θαλάσσιες μεταφορές, αναγνωρίζοντας τους περιορισμούς των υφιστάμενων, μονοδιάστατων δεικτών αξιολόγησης. Σκοπός της εργασίας είναι η ανάπτυξη ενός ολοκληρωμένου μεθοδολογικού πλαισίου, το οποίο, μέσω της εφαρμογής αλγορίθμων Μηχανικής Μάθησης, επιτρέπει μια πολυδιάστατη ανάλυση και κατηγοριοποίηση της περιβαλλοντικής απόδοσης των πλοίων.

Η μεθοδολογία βασίστηκε στην ανάλυση δημοσίως διαθέσιμων δεδομένων από τον μηχανισμό THETIS-MRV της Ευρωπαϊκής Ένωσης. Αρχικά, μέσω της τεχνικής της Ανάλυσης Κύριων Συνιστωσών (PCA), τα πολυάριθμα και αλληλοσυσχετιζόμενα δεδομένα εκπομπών και λειτουργίας συμπυκνώθηκαν σε τρεις νέους, ουσιαστικούς και ανεξάρτητους δείκτες: τη «Συνολική Περιβαλλοντική Επίδραση», τη «Λειτουργική Αποδοτικότητα» και το «Περιφερειακό Προφίλ Λειτουργίας». Στη συνέχεια, με την εφαρμογή του αλγορίθμου Ανάλυσης Συστάδων (K-Means) πάνω σε αυτούς τους δείκτες, τα πλοία κατηγοριοποιήθηκαν σε διακριτές ομάδες με ομοιογενή χαρακτηριστικά, αποκαλύπτοντας κρυμμένα πρότυπα συμπεριφοράς που δεν είναι εμφανή με τις παραδοσιακές μεθόδους.

Το τελικό συμπέρασμα που προέκυψε είναι ότι η συνδυαστική προσέγγιση PCA και ανάλυσης συστάδων υπερβαίνει τις αδυναμίες των απλών δεικτών, προσφέροντας ένα ολιστικό και πρακτικό εργαλείο αξιολόγησης. Το μοντέλο αυτό επιτρέπει την αναγνώριση ετερογενών «προφίλ» πλοίων (π.χ. «υψηλού ρυπαντικού φορτίου αλλά αποδοτικά» έναντι «χαμηλού φορτίου αλλά αναποτελεσματικά»), παρέχοντας τη βάση για τη χάραξη στοχευμένων περιβαλλοντικών πολιτικών και τη λήψη τεκμηριωμένων αποφάσεων από τις ναυτιλιακές εταιρείες.

## ABSTRACT

This diploma thesis addresses the critical issue of environmental impact from maritime transport, acknowledging the limitations of existing, one-dimensional assessment indicators. The objective of this work is to develop a comprehensive methodological framework that, through the application of Machine Learning algorithms, enables a multi-dimensional analysis and categorization of the environmental performance of ships.

The methodology was based on the analysis of publicly available data from the European Union's THETIS-MRV mechanism. Initially, using the Principal Component Analysis (PCA) technique, the numerous and inter-correlated emissions and operational data were condensed into three new, meaningful, and independent indices: "Overall Environmental Impact," "Operational Efficiency," and "Regional Operating Profile." Subsequently, by applying the Cluster Analysis algorithm (K-Means) to these indices, ships were categorized into distinct groups with homogeneous characteristics, revealing hidden behavioral patterns that are not apparent with traditional methods.

The final conclusion drawn is that the combined approach of PCA and cluster analysis overcomes the weaknesses of simple indicators, offering a holistic and practical assessment tool. This model allows for the identification of heterogeneous ship "profiles" (e.g., "high-polluting but efficient" versus "low-impact but inefficient"), providing a foundation for designing targeted environmental policies and enabling informed decision-making by shipping companies.

**Λέξεις κλειδιά**

Ανάλυση Δεδομένων, Ανάλυση Κύριων Συνιστωσών (PCA), Ανάλυση Συστάδων, Εκπομπές Ρύπων, Θαλάσσιες Μεταφορές, Κατηγοριοποίηση Πλοίων, Μηχανική Μάθηση, Ναυτιλία, Περιβαλλοντική Απόδοση, THETIS-MRV.

**Keywords**

Cluster Analysis, Data Analysis, Environmental Performance, K-Means, Machine Learning, Maritime Transport, Pollutant Emissions, Principal Component Analysis (PCA), Ship Categorization, Shipping, THETIS-MRV.

## ΠΡΟΛΟΓΟΣ

[Στον Πρόλογο θα αναφέρεται η ακαδημαϊκή μονάδα, όπου εκπονήθηκε η εργασία, και θα αναγνωρίζονται οι συντελεστές που βοήθησαν ουσιαστικά στη διάθεση και συλλογή στοιχείων ή στη χρήση μεθόδων κλπ., όπως επιβλέπων διπλωματικής, εργαστήριο, άλλοι φορείς και προσωπικό αυτών κλπ. Το συνολικό μέγεθος του Προλόγου δεν θα πρέπει να υπερβαίνει τη μια σελίδα.]

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	III
ABSTRACT .....	IV
ΠΡΟΛΟΓΟΣ.....	VI
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ .....	VII
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ ΚΑΙ ΕΙΚΟΝΩΝ .....	X
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ .....	XII
ΣΥΜΒΟΛΙΣΜΟΙ .....	XIII
1. ΕΙΣΑΓΩΓΗ .....	1
1.1 ΓΕΝΙΚΟ ΠΛΑΙΣΙΟ ΚΑΙ ΣΗΜΑΣΙΑ ΤΟΥ ΘΕΜΑΤΟΣ.....	1
1.2 ΠΡΟΒΛΗΜΑ ΚΑΙ ΕΡΕΥΝΗΤΙΚΕΣ ΠΡΟΚΛΗΣΕΙΣ.....	5
1.2.1 ΠΕΡΙΟΡΙΣΜΟΙ ΤΩΝ ΥΠΑΡΧΟΝΤΩΝ ΡΥΘΜΙΣΤΙΚΩΝ ΠΛΑΙΣΙΩΝ..	5
1.2.2 ΠΡΟΒΛΗΜΑΤΑ ΤΥΠΟΠΟΙΗΣΗΣ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΚΕΣ	
ΑΣΥΝΕΠΕΙΕΣ.....	6
1.2.3 ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΚΑΙ ΠΟΙΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ .....	6
1.2.4 ΑΝΑΓΚΗ ΓΙΑ ΕΝΙΑΙΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΑΞΙΟΛΟΓΗΣΗΣ .....	6
1.2.5 ΕΠΙΠΤΩΣΕΙΣ ΣΤΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ.....	7
1.3 ΣΤΟΧΟΙ ΚΑΙ ΕΡΕΥΝΗΤΙΚΕΣ ΕΡΩΤΗΣΕΙΣ .....	7
1.3.1 ΕΙΔΙΚΟΙ ΣΤΟΧΟΙ.....	8
1.3.2 ΕΡΕΥΝΗΤΙΚΕΣ ΕΡΩΤΗΣΕΙΣ .....	8
1.3.3 ΑΝΑΜΕΝΟΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΝΕΙΣΦΟΡΑ .....	9
1.4 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ.....	10
1.5 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ.....	11
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	13
2.1 ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ ΕΚΠΕΜΠΟΜΕΝΟΥΣ ΡΥΠΟΥΣ ΤΩΝ ΠΛΟΙΩΝ	13
2.1.1 ΤΥΠΟΙ ΕΚΠΟΜΠΩΝ ΚΑΙ ΠΗΓΕΣ ΠΡΟΕΛΕΥΣΗΣ.....	13
2.1.2 ΜΕΤΡΗΣΕΙΣ ΚΑΙ ΠΡΟΤΥΠΑ ΕΚΠΟΜΠΩΝ .....	15
2.1.3 ΓΕΩΓΡΑΦΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΕΚΠΟΜΠΩΝ .....	16
2.2 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΑΙΣΘΗΤΗΡΩΝ .....	17

2.2.1	ΣΥΓΧΡΟΝΑ ΣΥΣΤΗΜΑΤΑ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ.....	18
2.2.2	ΤΕΧΝΟΛΟΓΙΕΣ ΙΟΤ ΣΤΗ ΝΑΥΤΗΛΙΑ.....	19
2.2.3	ΠΟΙΟΤΗΤΑ ΚΑΙ ΑΞΙΟΠΙΣΤΙΑ ΔΕΔΟΜΕΝΩΝ .....	20
2.3	ΥΠΑΡΧΟΥΣΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗΣ ΑΠΟΔΟΣΗΣ .....	21
2.3.1	ΡΥΘΜΙΣΤΙΚΑ ΜΕΤΡΑ ΤΟΥ ΙΜΟ.....	22
2.3.2	ΕΝΑΛΛΑΚΤΙΚΟΙ ΔΕΙΚΤΕΣ ΚΑΙ ΠΡΩΤΟΒΟΥΛΙΕΣ .....	23
2.3.3	ΑΝΑΛΥΣΗ ΚΥΚΛΟΥ ΖΩΗΣ ΣΤΗ ΝΑΥΤΙΛΙΑ.....	24
2.4	ΕΦΑΡΜΟΓΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΗ ΝΑΥΤΙΛΙΑ.....	26
2.4.1	ΠΡΟΒΛΕΨΗ ΚΑΤΑΝΑΛΩΣΗΣ ΚΑΥΣΙΜΟΥ .....	26
2.4.2	ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΔΙΑΔΡΟΜΩΝ ΚΑΙ ΛΕΙΤΟΥΡΓΙΩΝ.....	28
2.4.3	ΠΡΟΛΗΨΗ ΒΛΑΒΩΝ ΚΑΙ ΠΡΟΓΝΩΣΤΙΚΗ ΣΥΝΤΗΡΗΣΗ .....	29
2.4.4	ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ.....	31
2.5	ΚΕΝΑ ΣΤΗ ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΑΙ ΕΡΕΥΝΗΤΙΚΕΣ ΕΥΚΑΙΡΙΕΣ .....	33
3.	ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ .....	36
3.1	ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA ANALYSIS) .....	36
3.1.1	ΜΑΘΗΜΑΤΙΚΟ ΥΠΟΒΑΘΡΟ ΤΗΣ PCA.....	37
3.1.2	ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ .....	40
3.1.3	ΠΛΕΟΝΕΚΤΗΜΑΤΑ & ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ PCA.....	41
3.1.4	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	44
3.2	ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΑΣ (CLUSTER ANALYSIS).....	44
3.2.1	ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ.....	45
3.2.2	ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΥΣΤΑΔΑΣ.....	46
3.2.3	ΚΥΡΙΕΣ ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ.....	48
3.2.4	ΠΛΕΟΝΕΚΤΗΜΑΤΑ & ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΥΣΤΑΔΑΣ .....	52
3.2.5	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	54
4.	ΑΝΑΛΥΣΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ .....	56
4.1	ΠΗΓΗ ΑΝΤΛΗΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΗΣ ΕΡΕΥΝΑΣ.....	56
4.2	ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΔΙΑΜΟΡΦΩΣΗ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ .....	57



4.3	ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ .....	61
4.4	ΔΙΕΡΕΥΝΗΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ .....	62
4.5	ΕΦΑΡΜΟΓΗ ΤΗΣ PCA .....	65
4.6	ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΩΝ ΒΑΣΙΣΜΕΝΗ ΣΤΟΥΣ ΤΡΕΙΣ ΔΕΙΚΤΕΣ.....	69
4.6.1	ΕΠΙΛΟΓΗ ΤΟΥ ΒΕΛΤΙΣΤΟΥ ΑΡΙΘΜΟΥ ΣΥΣΤΑΔΩΝ .....	69
4.6.2	ΑΛΓΟΡΙΘΜΟΣ K-MEANS .....	71
4.6.3	ΕΡΜΗΝΕΙΑ ΤΩΝ ΣΥΣΤΑΔΩΝ .....	73
4.6.4	ΑΝΑΛΥΣΗ ΣΥΝΘΕΣΗΣ ΣΥΣΤΑΔΩΝ ΚΑΙ ΤΥΠΩΝ ΠΛΟΙΩΝ .....	75
4.7	ΕΠΙΚΥΡΩΣΗ ΤΩΝ ΠΡΟΤΕΙΝΟΜΕΝΩΝ ΔΕΙΚΤΩΝ ΜΕ PCA.....	76
5.	ΣΥΖΗΤΗΣΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ.....	80
5.1	ΣΥΝΟΨΗ ΤΗΣ ΕΡΕΥΝΑΣ .....	80
5.2	ΚΥΡΙΑ ΕΥΡΗΜΑΤΑ ΚΑΙ ΕΡΜΗΝΕΙΑ.....	81
5.3	ΣΥΝΕΙΣΦΟΡΑ ΤΗΣ ΕΡΓΑΣΙΑΣ .....	82
5.4	ΠΕΡΙΟΡΙΣΜΟΙ ΤΗΣ ΕΡΕΥΝΑΣ .....	84
5.5	ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ.....	85
5.6	ΤΕΛΙΚΕΣ ΣΚΕΨΕΙΣ .....	86
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	88
	ΠΑΡΑΡΤΗΜΑ Α.....	98
A.1	ΕΝΟΤΗΤΑ 1.....	98
A.2	ΕΝΟΤΗΤΑ 2.....	98

**ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ ΚΑΙ ΕΙΚΟΝΩΝ**

	Σελ.
Διάγραμμα 4.1: Διάγραμμα των αποστάσεων Mahalanobis των παρατηρήσεων.	65
Διάγραμμα 4.2: Scree plot για τον προσδιορισμό του αριθμού κύριων συνιστωσών που πρέπει να διατηρηθούν. Η απότομη κάμψη στην τρίτη συνιστώσα υποδεικνύει τη διατήρηση τριών κύριων συνιστωσών.	66
Διάγραμμα 4.3: Biplot των δύο πρώτων κύριων συνιστωσών. Φαίνεται ομαδοποίηση των μεταβλητών κατανάλωσης καυσίμου και εκπομπών στην κατεύθυνση της PC1, ενώ οι δείκτες αποδοτικότητας κατευθύνονται προς την PC2.	68
Διάγραμμα 4.4: Ανάλυση Silhouette για διαφορετικό αριθμό συστάδων (αριστερά) και Gap Statistic για διαφορετικό αριθμό συστάδων (δεξιά).	70
Διάγραμμα 4.5: Απεικόνιση των τεσσάρων συστάδων στο χώρο των κύριων συνιστωσών. (α) PC1–PC2: διαχωρισμός ως προς την περιβαλλοντική επίδραση και την επιχειρησιακή αποδοτικότητα. (β) PC1–PC3: διαφοροποίηση βάσει περιβαλλοντικής επίδρασης και περιφερειακής λειτουργίας. (γ) PC2–PC3: ανάδειξη διαφορών αποδοτικότητας και περιφερειακών προτύπων λειτουργίας. Οι διαφορετικά χρωματισμένες ομάδες αντιπροσωπεύουν τις συστάδες που προέκυψαν από την ανάλυση.	71
Διάγραμμα 4.6: Απεικόνιση των δεκατεσσάρων συστάδων στο χώρο των κύριων συνιστωσών. (α) PC1–PC2: διαφοροποίηση με βάση την περιβαλλοντική επίδραση και την επιχειρησιακή αποδοτικότητα. (β) PC1–PC3: ανάλυση σε σχέση με την περιβαλλοντική επίδραση και την περιφερειακή λειτουργία. (γ) PC2–PC3: εστίαση στην αποδοτικότητα και τα περιφερειακά πρότυπα λειτουργίας. Οι χρωματισμοί απεικονίζουν τις διαφορετικές συστάδες που προέκυψαν από την ανάλυση.	72
Διάγραμμα 4.7: Τιμές των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) στα κεντροειδή των τεσσάρων συστάδων.	73

Διάγραμμα 4.8	Τιμές των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) στα κεντροειδή των δεκατεσσάρων συστάδων.	74
Διάγραμμα 4.9	Σύγκριση των loadings των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) για δύο τύπους πλοίων: Bulk carriers (μπλε) και Container ships (πορτοκαλί). Οι διαφοροποιήσεις στις τιμές των loadings αποτυπώνουν τις διαφορετικές περιβαλλοντικές και λειτουργικές προτεραιότητες των δύο κατηγοριών πλοίων.	77

**ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ**

	Σελ.
Πίνακας 4.1: Παραδείγματα μεταβλητών που αφαιρέθηκαν από το σύνολο των δεδομένων, καθώς και ο λόγος αφαίρεσης της κάθε μιας.	59
Πίνακας 4.2: Περιγραφή συνόλου δεδομένων.	59
Πίνακας 4.3: Πίνακας συσχέτισης των κανονικοποιημένων μεταβλητών.	63
Πίνακας 4.4: Loadings των πρώτων τριών κύριων συνιστωσών (PC1–PC3) για το αρχικό και το μειωμένο σύνολο δεδομένων, καθώς και οι διαφορές μεταξύ τους. Ο πίνακας επιβεβαιώνει τη σταθερότητα της δομής των συνιστωσών μετά την αφαίρεση ακραίων τιμών.	67
Πίνακας 4.5: Loadings των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) για τα υποσύνολα δεδομένων Bulk carriers και Container ships, καθώς και οι διαφορές μεταξύ των δύο φορτίων.	78

## ΣΥΜΒΟΛΙΣΜΟΙ

[Οι Συμβολισμοί είναι προαιρετικοί. Καταγράφονται σε μορφή πίνακα όλα τα σύμβολα που χρησιμοποιούνται στην εργασία καθώς και η σημασία τους. Οι Συμβολισμοί κατατάσσονται αλφαβητικά ως εξής: πεζά Ελληνικά, κεφαλαία Ελληνικά, πεζά Αγγλικά και κεφαλαία Αγγλικά.]

[παράδειγμα:  $k$  (N/m) - σταθερά ελατηρίου]

# **1. ΕΙΣΑΓΩΓΗ**

Οι θαλάσσιες μεταφορές αποτελούν πολύ σημαντικό παράγοντα για την εξέλιξη της παγκόσμιας οικονομίας, εξυπηρετώντας την συντριπτική πλειοψηφία του παγκόσμιου εμπορίου αγαθών και προϊόντων. Πιο συγκεκριμένα, παραπάνω από το 70% του όγκου του παγκόσμιου εμπορίου των αγαθών μεταφέρεται μέσω θαλάσσης (Corbett & Winebrake, 2008) ενώ η αξία αυτών ανέρχεται στο 70% της συνολικής αξίας του παγκόσμιου εμπορίου (Pivetta et al., 2024), νούμερα τα οποία μαρτυρούν την άμεση εξάρτηση της οικονομίας από την ναυτιλία. Η δυνατότητα μαζικής μεταφοράς ανθρώπων, καθώς και υλικών αγαθών, συμβάλλει καθοριστικά στην κοινωνική και πολιτισμική εξέλιξη των κοινωνιών ενθαρρύνοντας την αλληλεπίδραση μεταξύ των λαών, την ανταλλαγή απόψεων, ιδεών και νέων τεχνολογικών ανακαλύψεων. Επιπλέον, η εύρυθμη λειτουργία των θαλάσσιων συγκοινωνιών αποκτά ζωτικής σημασίας ρόλο για την κάλυψη των βασικών αναγκών των νησιωτικών περιοχών και την ανάπτυξη τους. Οι συγκεκριμένοι παράγοντες, σε συνδυασμό με το φαινόμενο της παγκοσμιοποίησης, έχουν επιφέρει ραγδαία εξέλιξη στον τομέα της ναυτιλίας τις τελευταίες δεκαετίες, με αποτέλεσμα ο αριθμός των πλοίων, και ιδίως αυτών που μεταφέρουν εμπορευματοκιβώτια, να έχει αυξηθεί σημαντικά (Kaplan & Solomon, 2016). Βέβαια, παρά τα πολλαπλά θετικά οφέλη στην παγκόσμια οικονομία και την κοινωνία, η αύξηση των πλοίων και των διαδρομών, οδηγεί, αναπόφευκτα, στην αύξηση του όγκου των εκπεμπόμενων ρύπων, προκαλώντας σημαντικές περιβαλλοντικές επιβαρύνσεις (Endresen et al., 2003), συμβάλλοντας στην διόγκωση του προβλήματος της κλιματικής αλλαγής (Eyring et al., 2010).

## **1.1 ΓΕΝΙΚΟ ΠΛΑΙΣΙΟ ΚΑΙ ΣΗΜΑΣΙΑ ΤΟΥ ΘΕΜΑΤΟΣ**

Τα πλοία διοχετεύουν την πλειοψηφία των ρύπων στο περιβάλλον κατά τη διάρκεια εκτέλεσης των διαδρομών τους. Οι περισσότεροι κινητήρες των θαλάσσιων μέσων μεταφοράς δουλεύουν με την καύση μαζούτ (Fridell, 2019), μέσω της οποίας παράγονται μεγάλες ποσότητες διοξειδίου του άνθρακα (CO<sub>2</sub>) (X.-T. Wang et al., 2021), στοιχείο βασικό για την επέκταση του φαινομένου του θερμοκηπίου (Collier et al., 2019). Επιπλέον, οι υψηλές θερμοκρασίες που αναπτύσσονται κατά την καύση των καυσίμων απελευθερώνουν στην ατμόσφαιρα οξείδια του αζώτου (NO<sub>x</sub>), (Winnes & Fridell, 2010) τα οποία εντείνουν την δημιουργία όζοντος στην

τροπόσφαιρα (Grewe et al., 2012). Συγχρόνως, μέσω της καύσης, κατά την διάρκεια λειτουργίας της μηχανής, το θείο που περιέχεται στα καύσιμα μετατρέπεται σε οξείδιο του θείου (SOx) (Williams, 2010), το οποίο, σε συνδυασμό με τα οξείδια του αζώτου, φέρουν ευθύνη για την κλιμάκωση του φαινομένου της όξινης βροχής (Shammas et al., 2020). Ιδιαίτερα επιβλαβή για την ανθρώπινη υγεία και την ποιότητα του αέρα έχουν χαρακτηριστεί και τα αιωρούμενα σωματίδια (PM), που εκπέμπονται από τα πλοία (Williams, 2010) και κατηγορούνται κυρίως για την πρόκληση αναπνευστικών προβλημάτων καθώς και τον περιορισμό της ορατότητας σε παράκτιες περιοχές (Anderson et al., 2012). Βέβαια, η περιβαλλοντική επιβάρυνση που προέρχεται από την λειτουργία των θαλάσσιων μέσων μεταφοράς δεν περιορίζεται στην εκπομπή ρύπων από την καύση των καυσίμων. Τα θαλάσσια οικοσυστήματα προσβάλλονται από την ηχορύπανση που προκαλείται κατά την λειτουργία ενός πλοίου, (Chebanov & Osipova, 2016), ενώ σε αρκετές περιπτώσεις έχουν παρατηρηθεί διαρροές πετρελαίου ή και άλλων απόβλητων προκαλώντας ρύπανση των υδάτων. Τέλος, θα πρέπει να υπογραμμιστεί το γεγονός ότι η εκπομπή ρύπων από τα πλοία περιορίζεται, αλλά δεν μηδενίζεται όσο το πλοίο είναι αγκυροβολημένο ή σε κατάσταση ελλιμενισμού, εφόσον για τον προγραμματισμό ορισμένων λειτουργιών τους κατά την περίοδο που βρίσκονται στο λιμάνι απαιτείται η λειτουργία κύριων και βοηθητικών μηχανών (Styhre & Winnes, 2019).

Η παγκόσμια οικονομία εξαρτάται σε αρκετά υψηλό βαθμό από την συνεχή και ομαλή λειτουργία χιλιάδων θαλάσσιων μέσων μεταφοράς καθημερινά (Garciano, 2023). Την ίδια στιγμή, η εξέλιξη και η διόγκωση της ναυτιλιακής δραστηριότητας καθίσταται κομβική για την κοινωνική και την πολιτισμική ανάπτυξη των πληθυσμών ανά τον κόσμο (Henderson, 2019). Γίνεται αντιληπτό το γεγονός ότι η ζήτηση και οι ανάγκες του παγκόσμιου πληθυσμού, απαιτούν μεγάλο αριθμό θαλάσσιων διελεύσεων, ο οποίος, αναπόφευκτα οδηγεί σε αύξηση των εκπεμπόμενων ρύπων και της ρύπανσης των θαλάσσιων οικοσυστημάτων. Η έντονη περιβαλλοντική επιβάρυνση που προκαλείται, σε συνδυασμό με την κρισιμότητα των θαλάσσιων μέσων μεταφοράς για την εξέλιξη της ανθρωπότητας, καθιστά επιτακτική την ανάγκη μείωσης των εκπεμπόμενων από τη λειτουργία των πλοίων ρύπων.

Τις τελευταίες δεκαετίες παρατηρείται έντονη μετάβαση της ναυτιλιακής βιομηχανίας σε μια πιο ψηφιακή εποχή. Πλέον, η ενσωμάτωση νέων τεχνολογιών καθίσταται απαραίτητη προϋπόθεση για τις ναυτιλιακές εταιρείες, ώστε να ανταποκρίνονται στις απαιτήσεις της αγοράς και να παραμένουν ανταγωνιστικές (Melnik et al., 2023). Η έντονη συσχέτιση της ναυτιλιακής

δραστηριότητας με την τεχνολογία προσφέρει τη δυνατότητα συλλογής μεγάλου όγκου δεδομένων, τόσο για την ενέργεια που καταναλώνουν τα πλοία, όσο και για τους ρύπους τους οποίους εκπέμπουν κατά τη λειτουργία τους. Η αξιοποίηση των δεδομένων με τα κατάλληλα τεχνολογικά μέσα, είναι ικανή να οδηγήσει σε πολύ χρήσιμα και αρκετά ασφαλή συμπεράσματα για την περιβαλλοντική επιβάρυνση που προκαλείται από τα θαλάσσια μέσα μεταφοράς,

Η εξέλιξη της τεχνολογίας και πιο συγκεκριμένα της Τεχνητής Νοημοσύνης, τα τελευταία χρόνια, έχουν προσφέρει στην επιστημονική κοινότητα τους αλγόριθμους μηχανικής μάθησης. Η μηχανική μάθηση αποτελεί μια υποκατηγορία της τεχνητής νοημοσύνης (AI), η οποία εστιάζει στην εκμάθηση και τη βελτίωση της λειτουργίας των υπολογιστών μέσω της λήψης δεδομένων και όχι στον ρητό προγραμματισμό τους για συγκεκριμένες μόνο εργασίες (Peterson et al., 2021). Οι αλγόριθμοι μηχανικής μάθησης είναι εκπαιδευμένοι στο να λαμβάνουν και να αναλύουν δεδομένα από μεγάλα σύνολα και να επιλέγουν την βέλτιστη λύση με βάση την ανάλυση που έχει προηγηθεί. Η βελτίωση των αλγορίθμων αυτών είναι ανάλογη του όγκου των δεδομένων που έχουν ληφθεί και της συχνότητας με την οποία χρησιμοποιούνται (H. Chen et al., 2021). Επιπροσθέτως, η μηχανική μάθηση διαχωρίζεται σε 4 υποκατηγορίες αναφορικά με το είδος των δεδομένων που χρησιμοποιούνται και το ζητούμενο αποτέλεσμα, την επιβλεπόμενη μάθηση, την ημιεπιβλεπόμενη μάθηση, την μη επιβλεπόμενη μάθηση και την ενισχυτική μάθηση. Για κάθε πρόβλημα προς επίλυση στο χώρο της Μηχανικής Μάθησης υπάρχει ένας κατάλληλος τρόπος μάθησης και για κάθε τρόπο μάθησης υπάρχει τουλάχιστον ένας κατάλληλος αλγόριθμος που μπορεί να χρησιμοποιηθεί (Peterson et al., 2021).

Τα τελευταία χρόνια, διάφορα μοντέλα τεχνικής μάθησης έχουν αρχίσει να χρησιμοποιούνται στον τομέα της ναυτιλίας, συμβάλλοντας, κατά κύριο λόγο, στην βελτίωση στον τομέα της ασφάλειας και της αποδοτικότητας. Πιο συγκεκριμένα, μοντέλα μηχανικής μάθησης έχουν αξιοποιηθεί για την πρόβλεψη κινδύνων και την αποφυγή ατυχημάτων σε περιοχές με αυξημένη κυκλοφορία (Feng et al., 2024), την αυτοματοποίηση ορισμένων εργασιών εντός των λιμανιών (Mekkaoui et al., 2022), ενώ, αναλύοντας δεδομένα από τους αισθητήρες των πλοίων προσφέρονται χρήσιμες πληροφορίες για ενδεχόμενη συντήρηση του εκάστοτε πλοίου (Han et al., 2023). Η συλλογή και επεξεργασία των δεδομένων που επικρατούν στην αγορά, σχετικά με την ζήτηση και την προσφορά ορισμένων αγαθών, προσφέρει στις ναυτιλιακές εταιρείες την δυνατότητα να λάβουν πιο συμφέρουσες οικονομικά αποφάσεις για τη λειτουργία των πλοίων τους. Τέλος, η χρήση αλγορίθμων μηχανικής μάθησης στον τομέα των θαλάσσιων μεταφορών



καθιστά δυνατή την βελτιστοποίηση των διαδρομών των πλοίων, αποσκοπώντας στην ελάχιστη δυνατή κατανάλωση καυσίμου. Η βελτιστοποίηση των διαδρομών, που συνεπάγεται μείωση των καταναλωμένων καυσίμων, εκτός από το οικονομικό όφελος για τις ναυτιλιακές εταιρείες, συνεισφέρει στην μείωση των εκπεμπόμενων ρύπων, εφόσον η πλειοψηφία αυτών παράγονται κατά την διαδικασία της καύσης (Moradi et al., 2022). Η μείωση αυτή, κρίνεται σημαντική, αλλά δεν επαρκεί για τον περιορισμό της περιβαλλοντικής επιβάρυνσης που προκαλούν τα πλοία, στα επιθυμητά επίπεδα. Η αντιμετώπιση του προβλήματος προϋποθέτει πολύπλευρη προσέγγιση και πιο ενδελεχή ανάλυση ώστε να κριθεί πραγματικά αποδοτική.

Η συγκεκριμένη μελέτη προτείνει ένα ολοκληρωμένο μοντέλο ανάλυσης δεδομένων με χρήση τεχνικών μηχανικής μάθησης, αποσκοπώντας στην αξιολόγηση και κατηγοριοποίηση της περιβαλλοντικής απόδοσης των θαλάσσιων μέσων μεταφοράς. Σε πρώτο στάδιο, συλλέγονται ιστορικά δεδομένα από τον μηχανισμό Παρακολούθησης, Αναφοράς και Επαλήθευσης (MRV) της Ευρωπαϊκής Ένωσης που περιλαμβάνει αναφορές εκπομπών CO<sub>2</sub> και άλλα αέρια του θερμοκηπίου για πλοία άνω των 5.000 κόντων. Τα δεδομένα αφορούν την ετήσια κατανάλωση καυσίμου, τις εκπομπές διοξειδίου του άνθρακα σε διαφορετικές γεωγραφικές περιοχές, τον χρόνο παραμονής στη θάλασσα, την απόσταση που διανύθηκε, καθώς και τα τεχνικά χαρακτηριστικά των πλοίων όπως διαστάσεις, βάρος φορτίου και ισχύς μηχανών. Εν συνεχεία, μέσω αλγορίθμων μηχανικής μάθησης, αναλύονται και επεξεργάζονται τα ληφθέντα δεδομένα με σκοπό να παρατηρηθούν οι συνθήκες κατά τις οποίες πραγματοποιείται υψηλή εκπομπή ρύπων και η συχνότητα εμφάνισης τους. Η ανάλυση των δεδομένων και η προσέγγιση της βέλτιστης λύσης για το κάθε πλοίο ξεχωριστά καθίσταται ασύμφορη οικονομικά και πρακτικά αδύνατη διαδικασία. Για τον λόγο αυτό, γίνεται χρήση αλγορίθμων μηχανικής μάθησης, επιδιώκοντας την ομαδοποίηση και την κατηγοριοποίηση των δεδομένων.

Στη συνέχεια, πραγματοποιείται διαχωρισμός και κατάταξη των θαλάσσιων μέσων σε υποσύνολα, τέτοια ώστε το εκάστοτε υποσύνολο να εκφράζεται από μια σχέση συσχέτισης, η οποία θα έχει προσδιοριστεί μέσω αλγορίθμων μηχανικής μάθησης, συνδέοντας τα συλλεχθέντα δεδομένα για τα χαρακτηριστικά των πλοίων με τις εκπομπές ρύπων. Με τον τρόπο αυτό, δημιουργούνται ευνοϊκές προϋποθέσεις για την διαχείριση, την ανάλυση και την αξιοποίηση του μεγάλου όγκου των διαθέσιμων δεδομένων. Εν συνεχεία, οι τεχνικές μηχανικής μάθησης προσφέρουν, για το κάθε υποσύνολο, την δυνατότητα εκτίμησης της εκπομπής ρύπων, ανιχνεύουν ενδεχόμενους κινδύνους και τυχόν προβλήματα στις μηχανές των πλοίων, που θα μπορούσαν να

προκαλέσουν αύξηση των εκπεμπόμενων ρύπων, και συγχρόνως μπορούν να προειδοποιήσουν τις ναυτιλιακές εταιρείες σε περίπτωση ανάγκης για συντήρηση. Τέλος, καθίσταται εφικτή η ανάπτυξη μοντέλων φιλικών προς το περιβάλλον τα οποία θα λαμβάνουν υπόψη την ταχύτητα του πλοίου, το βάρος του φορτίου το οποίο μεταφέρει και τις καιρικές συνθήκες που επικρατούν, αποσκοπώντας στην καθοδήγηση των καπετάνιων, ώστε να λαμβάνουν τις απαραίτητες αποφάσεις και μέτρα.

## 1.2 ΠΡΟΒΛΗΜΑ ΚΑΙ ΕΡΕΥΝΗΤΙΚΕΣ ΠΡΟΚΛΗΣΕΙΣ

Παρά τη σημαντική συνεισφορά των θαλάσσιων μεταφορών στην παγκόσμια οικονομία, η ταχεία ανάπτυξη της ναυτιλιακής δραστηριότητας έχει επιφέρει σοβαρές περιβαλλοντικές επιπτώσεις που απαιτούν άμεση αντιμετώπιση. Οι συνολικές εκπομπές αερίων θερμοκηπίου από τη ναυτιλιακή βιομηχανία αυξήθηκαν κατά 9,6% από το 2012 έως το 2018, ενώ προβλέπεται ότι έως το 2050 οι εκπομπές θα αυξηθούν κατά 90-130% σε σχέση με τα επίπεδα του 2008 (Faber et al., 2020).

### 1.2.1 ΠΕΡΙΟΡΙΣΜΟΙ ΤΩΝ ΥΠΑΡΧΟΝΤΩΝ ΡΥΘΜΙΣΤΙΚΩΝ ΠΛΑΙΣΙΩΝ

Τα τρέχοντα ρυθμιστικά μέτρα όπως ο Energy Efficiency Design Index (EEDI) και ο Energy Efficiency Operational Indicator (EEOI) αποτελούν αρχικά σημεία αναφοράς, αλλά αδυνατούν να συνυπολογίσουν τις λειτουργικές και περιβαλλοντικές αβεβαιότητες, όπως οι μεταβαλλόμενες καιρικές συνθήκες και οι συνθήκες ταξιδιού. Αυτό μπορεί να οδηγήσει σε ανακριβείς αξιολογήσεις της πραγματικής απόδοσης των πλοίων και της έντασης άνθρακα, καθιστώντας δύσκολο για τους χειριστές να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με τη συμμόρφωση και τις επενδύσεις σε νέες τεχνολογίες (Vasilikis et al., 2023).

Επιπλέον, πολλοί ρύποι και περιβαλλοντικές επιπτώσεις δεν παρακολουθούνται τακτικά ή δεν περιλαμβάνονται σε αυτά τα πλαίσια, αφήνοντας σημαντικά κενά στη ρυθμιστική κάλυψη (Moldanová et al., 2021). Η αδυναμία ακριβούς πρόβλεψης της κατανάλωσης καυσίμου και των εκπομπών υπό πραγματικές συνθήκες μπορεί να οδηγήσει σε υπο-βέλτιστες επιλογές μεταξύ μηχανικών και ηλεκτρικών συστημάτων προώθησης, καθώς τα αποτελέσματα εξαρτώνται σε μεγάλο βαθμό από συγκεκριμένους λειτουργικούς και περιβαλλοντικούς παράγοντες (Y. Wang et al., 2024).

### 1.2.2 ΠΡΟΒΛΗΜΑΤΑ ΤΥΠΟΠΟΙΗΣΗΣ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΚΕΣ ΑΣΥΝΕΠΕΙΕΣ

Υπάρχει έλλειψη συνέπειας στις μεθοδολογίες που χρησιμοποιούνται για την αξιολόγηση της περιβαλλοντικής απόδοσης, ιδιαίτερα στις μελέτες ανάλυσης κύκλου ζωής (LCA). Βασικά μεθοδολογικά ζητήματα περιλαμβάνουν ασυνεπείς ορισμούς λειτουργικών μονάδων, επιλογή ορίων και πηγών δεδομένων, γεγονός που καθιστά δύσκολη τη σύγκριση αποτελεσμάτων μεταξύ μελετών ή την καθιέρωση βαθμολογιών σε επίπεδο βιομηχανίας.

Χαρακτηριστικά, το 90% των μελετών LCA που εξετάστηκαν είχαν ελλειπείς λειτουργικές μονάδες, και δεν υπήρχε συναίνεση για τα όρια συστήματος ή τις μεθόδους απογραφής, υπονομεύοντας την αξιοπιστία και τη συγκρισιμότητα των ευρημάτων (Jacquet et al., 2024). Αυτή η έλλειψη τυποποίησης περιπλέκει την ανάπτυξη ενιαίων εργαλείων αξιολόγησης και εμποδίζει την εναρμόνιση των κανονισμών.

### 1.2.3 ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΚΑΙ ΠΟΙΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ

Ο ναυτιλιακός τομέας παράγει τεράστιες ποσότητες λειτουργικών και περιβαλλοντικών δεδομένων, αλλά η ποιότητα και η πληρότητα των δεδομένων παραμένουν σημαντικά εμπόδια. Τα ελλιπή ή ασυνεπή δεδομένα μπορούν να μεροληπτούν στα αποτελέσματα και να οδηγήσουν σε παραπλανητικά συμπεράσματα, επηρεάζοντας την εκτίμηση εκπομπών, τον σχεδιασμό πλοίων και τις λειτουργικές αποφάσεις.

Προηγμένες μέθοδοι εκχώρησης και προσεγγίσεις μηχανικής μάθησης αναπτύσσονται για την αντιμετώπιση αυτών των κενών, αλλά οι πρακτικοί περιορισμοί και η ανάγκη για ισχυρά, υψηλής ποιότητας σύνολα δεδομένων επιμένουν (Skarlatos et al., 2023; Sun et al., 2025). Επιπλέον, η πολυπλοκότητα της ενσωμάτωσης διαφορετικών τύπων δεδομένων (π.χ. λειτουργικά, σχεδιαστικά, περιβαλλοντικά) προσθέτει στην πρόκληση (Wang et al., 2024).

### 1.2.4 ΑΝΑΓΚΗ ΓΙΑ ΕΝΙΑΙΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΑΞΙΟΛΟΓΗΣΗΣ

Οι περισσότερες τρέχουσες μέθοδοι αξιολόγησης εστιάζουν σε συγκεκριμένες φάσεις (π.χ. λειτουργική) ή απομονωμένους δείκτες, αντί να υιοθετούν μια ολιστική προοπτική κύκλου ζωής. Αυτή η κατακερματισμένη προσέγγιση περιορίζει την ικανότητα να συλληφθεί το πλήρες

περιβαλλοντικό αντίκτυπο των ναυτιλιακών δραστηριοτήτων και εμποδίζει την ανάπτυξη ολοκληρωμένων στρατηγικών βιωσιμότητας (Jacquet et al., 2024; Mondello et al., 2021).

Οι εκκλήσεις για τυποποιημένες, ολιστικές μεθοδολογίες LCA και ενσωματωμένα εργαλεία αναφοράς αυξάνονται, με συστάσεις για ευθυγράμμιση μελλοντικών μελετών με διεθνή πρότυπα και ανάπτυξη δημόσιων βάσεων δεδομένων προσαρμοσμένων σε ναυτιλιακά πλαίσια (C. Wang et al., 2025).

### 1.2.5 ΕΠΙΠΤΩΣΕΙΣ ΣΤΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ

Αυτές οι προκλήσεις οδηγούν σε αβεβαιότητα και ασυνέπεια στις αξιολογήσεις περιβαλλοντικής απόδοσης, καθιστώντας δύσκολο για τις ναυτιλιακές εταιρείες να δώσουν προτεραιότητα στις επενδύσεις, να επιλέξουν βέλτιστες τεχνολογίες ή να συμμορφωθούν με τους εξελισσόμενους κανονισμούς. Για παράδειγμα, η αδυναμία ακριβούς πρόβλεψης της κατανάλωσης καυσίμου και των εκπομπών υπό πραγματικές συνθήκες μπορεί να οδηγήσει σε υπο-βέλτιστες επιλογές μεταξύ μηχανικών και ηλεκτρικών συστημάτων προώθησης (Vasilikis et al., 2023; Wang et al., 2024).

Η έλλειψη τυποποιημένων, αξιόπιστων εργαλείων αξιολόγησης επίσης περιπλέκει την ανάπτυξη πολιτικής και τη βαθμολόγηση της βιομηχανίας, επιβραδύνοντας την πρόοδο προς τους στόχους μείωσης εκπομπών άνθρακα και βιωσιμότητας (Wang et al., 2025; Jacquet et al., 2024; Moldanová et al., 2021).

## 1.3 **ΣΤΟΧΟΙ ΚΑΙ ΕΡΕΥΝΗΤΙΚΕΣ ΕΡΩΤΗΣΕΙΣ**

Η παρούσα διπλωματική εργασία στοχεύει στην ανάπτυξη ενός ολοκληρωμένου μοντέλου ανάλυσης δεδομένων με χρήση τεχνικών μηχανικής μάθησης για την αξιολόγηση και κατηγοριοποίηση της περιβαλλοντικής απόδοσης των θαλάσσιων μέσων μεταφοράς. Ο κεντρικός στόχος είναι η δημιουργία νέων, αντικειμενικών και κατανοητών δεικτών περιβαλλοντικής απόδοσης που θα αντιμετωπίζουν τις υπάρχουσες προκλήσεις και περιορισμούς των τρεχόντων μεθόδων αξιολόγησης.

### 1.3.1 ΕΙΔΙΚΟΙ ΣΤΟΧΟΙ

Η έρευνα επιδιώκει τη δημιουργία σύνθετων δεικτών ενεργειακής απόδοσης που θα:

- Συνδυάζουν πολλαπλές παραμέτρους περιβαλλοντικής απόδοσης σε ενιαίους, κατανοητούς δείκτες
- Ελαχιστοποιούν τον αριθμό των παραμέτρων που χαρακτηρίζουν την περιβαλλοντική απόδοση ενός πλοίου
- Προσφέρουν πρακτικές ερμηνείες που μπορούν να χρησιμοποιηθούν από τους ενδιαφερόμενους φορείς

Μέσω της εφαρμογής αλγορίθμων συσταδοποίησης, η έρευνα στοχεύει στη:

- Δημιουργία ομάδων πλοίων με παρόμοια περιβαλλοντική απόδοση
- Αναγνώριση προτύπων και συμπεριφορών που συνδέονται με τις εκπομπές ρύπων
- Ανάπτυξη εξειδικευμένων στρατηγικών για κάθε κατηγορία πλοίων

Η έρευνα επιδιώκει την ανάπτυξη:

- Ενός γραφικού συστήματος ετικετοποίησης που θα απεικονίζει την περιβαλλοντική απόδοση
- Εργαλείων οπτικοποίησης που θα διευκολύνουν την κατανόηση και σύγκριση της απόδοσης
- Συστημάτων υποστήριξης αποφάσεων για τη ναυτιλιακή βιομηχανία

### 1.3.2 ΕΡΕΥΝΗΤΙΚΕΣ ΕΡΩΤΗΣΕΙΣ

Η παρούσα έρευνα επιχειρεί να απαντήσει στις ακόλουθες θεμελιώδεις ερωτήσεις:

#### 1. Ερώτηση: Μείωση Διάστασης

**"Πώς μπορούν οι τεχνικές μηχανικής μάθησης, και συγκεκριμένα η Ανάλυση Κύριων Συνιστωσών (PCA), να μειώσουν τη διαστατικότητα των δεδομένων περιβαλλοντικής απόδοσης των πλοίων διατηρώντας παράλληλα τη μέγιστη δυνατή πληροφορία;"**

Η παραπάνω ερώτηση εξετάζει τη δυνατότητα δημιουργίας λιγότερων, αλλά πιο ουσιαστικών δεικτών από τα πολλαπλά χαρακτηριστικά που περιγράφουν την περιβαλλοντική απόδοση ενός πλοίου.

#### 2. Ερώτηση: Φυσικές ομαδοποιήσεις

**"Υπάρχουν φυσικές ομαδοποιήσεις των θαλάσσιων μέσων μεταφοράς βάσει των περιβαλλοντικών τους χαρακτηριστικών και πώς μπορούν αυτές να εντοπιστούν μέσω αλγορίθμων συσταδοποίησης;"**

Η ερώτηση αυτή διερευνά την ύπαρξη κρυμμένων προτύπων στα δεδομένα που θα μπορούσαν να οδηγήσουν σε πιο στοχευμένες πολιτικές περιβαλλοντικής διαχείρισης.

### **3. Ερώτηση: Ερμηνεία αποτελεσμάτων**

**"Πώς μπορούν τα αποτελέσματα των αλγορίθμων μηχανικής μάθησης να μετατραπούν σε κατανοητούς και πρακτικά χρήσιμους δείκτες για τη ναυτιλιακή βιομηχανία;"**

Αυτή η ερώτηση αντιμετωπίζει την πρόκληση της ερμηνευσιμότητας που συχνά συνοδεύει τις προηγμένες τεχνικές ανάλυσης δεδομένων.

### **4. Ερώτηση: Πρακτική εφαρμογή**

**"Μπορεί ένα ενιαίο γραφικό σύστημα να αντικατοπτρίζει αποτελεσματικά τη συνολική περιβαλλοντική απόδοση ενός πλοίου λαμβάνοντας υπόψη το μέγεθος, την περιοχή λειτουργίας και την ενεργειακή απόδοσή του;"**

Η ερώτηση εξετάζει τη δυνατότητα δημιουργίας ενός απλού αλλά περιεκτικού εργαλείου αξιολόγησης.

## **1.3.3 ANAMENOMENA APIOTEΛESMATATA KAI SYNEISΦOPA**

Η παρούσα έρευνα αναμένεται να συνεισφέρει σημαντικά τόσο σε θεωρητικό όσο και σε πρακτικό επίπεδο στον τομέα της περιβαλλοντικής αξιολόγησης των θαλάσσιων μεταφορών. Από θεωρητική άποψη, η εργασία προτείνει την ανάπτυξη νέων μεθοδολογιών συνδυασμού PCA και ανάλυση συστάδων για την ανάλυση περιβαλλοντικών δεδομένων ναυτιλίας. Η δημιουργία νέων σύνθετων δεικτών που συνδυάζουν φυσικά χαρακτηριστικά πλοίων, λειτουργικά δεδομένα και εκπομπές ρύπων αναμένεται να προσφέρει μια πιο ολοκληρωμένη εικόνα της περιβαλλοντικής απόδοσης. Επιπλέον, η πρόταση ενός ολοκληρωμένου πλαισίου αξιολόγησης που αντιμετωπίζει τα υπάρχοντα μεθοδολογικά κενά και προβλήματα τυποποίησης αποτελεί σημαντική θεωρητική συνεισφορά στην επιστημονική κοινότητα.

Σε πρακτικό επίπεδο, η έρευνα αναμένεται να προσφέρει άμεσα εφαρμόσιμα εργαλεία και λύσεις σε διάφορους ενδιαφερόμενους φορείς. Για τη ναυτιλιακή βιομηχανία, τα αποτελέσματα

θα παρέχουν εργαλεία για την αξιολόγηση και βελτίωση της περιβαλλοντικής απόδοσης των στόλων τους, διευκολύνοντας τη λήψη τεκμηριωμένων αποφάσεων για επενδύσεις σε πράσινες τεχνολογίες και βελτιστοποίηση λειτουργιών. Οι ρυθμιστικοί φορείς θα έχουν στη διάθεσή τους νέους, αξιόπιστους δείκτες που μπορούν να ενσωματωθούν σε μελλοντικές ρυθμίσεις και πολιτικές περιβαλλοντικής προστασίας, ενώ παράλληλα θα διευκολυνθεί η διαδικασία παρακολούθησης και ελέγχου της συμμόρφωσης.

Οι περιβαλλοντικοί οργανισμοί θα αποκτήσουν αποτελεσματικά μέσα παρακολούθησης και αξιολόγησης της προόδου προς τους στόχους μείωσης των εκπομπών, ενώ η ακαδημαϊκή κοινότητα θα έχει πρόσβαση σε νέες μεθοδολογίες που μπορούν να προσαρμοστούν και να εφαρμοστούν σε άλλους τομείς περιβαλλοντικής ανάλυσης. Το προτεινόμενο γραφικό σύστημα ετικετοποίησης αναμένεται να γίνει ένα εύκολα κατανοητό εργαλείο που θα μπορεί να χρησιμοποιηθεί από μη ειδικούς για την άμεση αξιολόγηση της περιβαλλοντικής απόδοσης των πλοίων, συμβάλλοντας έτσι στην ευαισθητοποίηση και τη διαφάνεια στον τομέα της ναυτιλίας.

#### **1.4 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ**

Η μεθοδολογική προσέγγιση της παρούσας έρευνας βασίζεται στον συνδυασμό προηγμένων τεχνικών μηχανικής μάθησης για την ανάλυση πραγματικών δεδομένων περιβαλλοντικής απόδοσης θαλάσσιων μέσων μεταφοράς. Η μεθοδολογία ακολουθεί μια συστηματική προσέγγιση που ξεκινά από τη συλλογή και προεπεξεργασία δεδομένων και καταλήγει στη δημιουργία νέων δεικτών και εργαλείων αξιολόγησης.

Τα δεδομένα που συλλέγονται περιλαμβάνουν ετήσια κατανάλωση καυσίμου, εκπομπές διοξειδίου του άνθρακα σε διαφορετικές γεωγραφικές περιοχές, χρόνο παραμονής στη θάλασσα, απόσταση που διανύθηκε.

Η μεθοδολογία ακολουθεί μια διαδικασία πέντε βημάτων που ξεκινά με την προεπεξεργασία των δεδομένων, η οποία περιλαμβάνει τον καθαρισμό, την τυποποίηση και την αντιμετώπιση ελλειπών τιμών. Στη συνέχεια, πραγματοποιείται διερευνητική ανάλυση για την κατανόηση της δομής των δεδομένων, τον εντοπισμό πιθανών ακραίων τιμών και την εξέταση των συσχετίσεων μεταξύ των μεταβλητών. Το τρίτο βήμα αφορά την εφαρμογή της PCA, μιας τεχνικής μη επιβλεπόμενης μάθησης που επιτρέπει τη μείωση της διαστατικότητας των δεδομένων διατηρώντας παράλληλα τη μέγιστη δυνατή πληροφορία.

Η PCA χρησιμοποιείται για τη δημιουργία νέων, ασυσχέτιστων μεταξύ τους μεταβλητών που ονομάζονται κύριες συνιστώσες, οι οποίες αντιπροσωπεύουν γραμμικούς συνδυασμούς των αρχικών μεταβλητών και συγκεντρώνουν τη μέγιστη διακύμανση των δεδομένων. Αυτή η διαδικασία επιτρέπει την απλοποίηση του πολυδιάστατου προβλήματος αξιολόγησης της περιβαλλοντικής απόδοσης σε λιγότερες, αλλά πιο ουσιαστικές διαστάσεις που μπορούν να ερμηνευθούν και να χρησιμοποιηθούν πρακτικά.

Το τέταρτο βήμα περιλαμβάνει την εφαρμογή αλγορίθμων ανάλυσης συστάδας στις κύριες συνιστώσες που προέκυψαν από την PCA. Η ανάλυση συστάδας είναι μια τεχνική μη επιβλεπόμενης μάθησης που στοχεύει στην ομαδοποίηση των πλοίων σε διακριτές κατηγορίες με βάση την ομοιότητα των περιβαλλοντικών τους χαρακτηριστικών. Για την επιλογή του βέλτιστου αριθμού συστάδων χρησιμοποιούνται μέθοδοι όπως η ανάλυση silhouette και το κριτήριο του αγκώνα, ενώ η ποιότητα της συσταδοποίησης αξιολογείται μέσω κατάλληλων μετρικών.

Το πέμπτο βήμα αφορά την επικύρωση των αποτελεσμάτων μέσω στατιστικών τεστ και της εξέτασης της σταθερότητας των συστάδων που προέκυψαν. Η επικύρωση περιλαμβάνει επίσης τη χρήση ενός ξεχωριστού συνόλου δεδομένων από πλοία εμπορευματοκιβωτίων για την αξιολόγηση της γενικευσιμότητας των προτεινόμενων δεικτών.

Η προτεινόμενη μεθοδολογία διακρίνεται για την ολιστική της προσέγγιση, καθώς συνδυάζει πολλαπλές πηγές δεδομένων και εξετάζει ταυτόχρονα τεχνικά χαρακτηριστικά, λειτουργικά δεδομένα και περιβαλλοντικές επιδόσεις. Η χρήση τεχνικών μη επιβλεπόμενης μάθησης επιτρέπει την ανακάλυψη κρυμμένων προτύπων στα δεδομένα χωρίς προκαταλήψεις ή προκαθορισμένες κατηγορίες, ενώ η έμφαση στην ερμηνευσιμότητα των αποτελεσμάτων διασφαλίζει ότι τα νέα εργαλεία που αναπτύσσονται θα είναι πρακτικά χρήσιμα για τη ναυτιλιακή βιομηχανία.

## **1.5 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ**

Η παρούσα διπλωματική εργασία διαρθρώνεται σε πέντε κεφάλαια, τα οποία οδηγούν τον αναγνώστη συστηματικά από το θεωρητικό υπόβαθρο του προβλήματος στην ανάπτυξη της μεθοδολογίας, την ανάλυση των αποτελεσμάτων και την παρουσίαση των τελικών συμπερασμάτων.



Το πρώτο κεφάλαιο αποτελεί την εισαγωγή, όπου παρουσιάζεται το γενικό πλαίσιο και η σημασία του θέματος. Αναλύονται το ερευνητικό πρόβλημα και οι σύγχρονες προκλήσεις, τίθενται οι κεντρικοί στόχοι και οι ερευνητικές ερωτήσεις, και σκιαγραφείται η μεθοδολογική προσέγγιση που ακολουθείται στην έρευνα.

Στο δεύτερο κεφάλαιο πραγματοποιείται εκτενής βιβλιογραφική ανασκόπηση. Εξετάζονται οι κύριοι τύποι ρύπων από τα πλοία, οι σύγχρονες τεχνολογίες συλλογής δεδομένων, οι υφιστάμενες μέθοδοι περιβαλλοντικής αξιολόγησης και οι τρέχουσες εφαρμογές της μηχανικής μάθησης στον ναυτιλιακό τομέα. Το κεφάλαιο ολοκληρώνεται με τον εντοπισμό των κενών στη διεθνή βιβλιογραφία, θεμελιώνοντας την αναγκαιότητα της παρούσας έρευνας.

Το τρίτο κεφάλαιο εστιάζει στην Περιγραφή των Μεθόδων Μηχανικής Μάθησης. Αναλύεται σε βάθος το μαθηματικό υπόβαθρο και η μεθοδολογία των δύο κεντρικών τεχνικών που αξιοποιούνται: της Ανάλυσης Κύριων Συνιστωσών (PCA) για τη μείωση της διαστατικότητας των δεδομένων και της ανάλυσης συστάδας (cluster analysis) για την ομαδοποίηση των πλοίων σε κατηγορίες με ομοιογενή χαρακτηριστικά.

Το τέταρτο κεφάλαιο είναι αφιερωμένο στην Ανάλυση και τα Αποτελέσματα. Περιγράφεται αναλυτικά η πηγή άντλησης των δεδομένων, η διαδικασία προεπεξεργασίας τους, και παρουσιάζονται τα αποτελέσματα από την εφαρμογή της PCA και της ανάλυσης συστάδας. Μέσα από διαγράμματα και πίνακες, οπτικοποιούνται και ερμηνεύονται οι νέοι συνθετικοί δείκτες και οι ομάδες (συστάδες) πλοίων που προέκυψαν.

Η εργασία ολοκληρώνεται με το πέμπτο κεφάλαιο, το οποίο περιλαμβάνει τη Συζήτηση και τα Συμπεράσματα. Σε αυτό το κεφάλαιο, τα ευρήματα συντίθενται και ερμηνεύονται στο πλαίσιο των αρχικών ερευνητικών ερωτημάτων. Επισημαίνεται η συνεισφορά της έρευνας, αναγνωρίζονται οι περιορισμοί της και προτείνονται κατευθύνσεις για μελλοντική έρευνα στον τομέα.

Τέλος, παρατίθεται η Βιβλιογραφία που χρησιμοποιήθηκε καθώς και τα Παραρτήματα που συμπληρώνουν την εργασία.

## 2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Η παρούσα βιβλιογραφική ανασκόπηση εξετάζει συστηματικά τη διεθνή βιβλιογραφία σχετικά με την περιβαλλοντική αξιολόγηση των θαλάσσιων μεταφορών και τις εφαρμογές της μηχανικής μάθησης στον τομέα αυτό. Η ανάλυση οργανώνεται σε πέντε κύριες θεματικές ενότητες που καλύπτουν από τη φύση και τις πηγές των εκπομπών ρύπων έως τις σύγχρονες τεχνολογίες συλλογής δεδομένων και τις καινοτόμες μεθόδους αξιολόγησης. Η βιβλιογραφική έρευνα στοχεύει στον εντοπισμό των τρεχόντων προκλήσεων, των υπαρχόντων λύσεων και των ερευνητικών κενών που υπάρχουν στον τομέα της περιβαλλοντικής αξιολόγησης των πλοίων, παρέχοντας το θεωρητικό πλαίσιο για την ανάπτυξη νέων μεθοδολογιών που συνδυάζουν τεχνικές μηχανικής μάθησης.

### 2.1 ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΥΣ ΕΚΠΕΜΠΟΜΕΝΟΥΣ ΡΥΠΟΥΣ ΤΩΝ ΠΛΟΙΩΝ

Για την κατανόηση της πολυπλοκότητας του προβλήματος της περιβαλλοντικής επιβάρυνσης από τη ναυτιλία, είναι απαραίτητη η λεπτομερής εξέταση των τύπων και της προέλευσης των εκπομπών ρύπων από τα θαλάσσια μέσα μεταφοράς.

#### 2.1.1 ΤΥΠΟΙ ΕΚΠΟΜΠΩΝ ΚΑΙ ΠΗΓΕΣ ΠΡΟΕΛΕΥΣΗΣ

Οι εκπομπές από τα πλοία αποτελούν σημαντική πηγή ατμοσφαιρικής ρύπανσης, ιδιαίτερα κοντά σε λιμάνια και σε περιοχές με έντονη ναυτιλιακή κίνηση. Οι κύριοι ρύποι που εκπέμπονται από τα θαλάσσια μέσα μεταφοράς είναι το διοξείδιο του άνθρακα ( $\text{CO}_2$ ), τα οξείδια του αζώτου ( $\text{NO}_x$ ), τα οξείδια του θείου ( $\text{SO}_x$ ) και τα αιωρούμενα σωματίδια (PM), με τη μεγαλύτερη συνεισφορά να προέρχεται από τις κύριες μηχανές, τις βοηθητικές μηχανές και τους λέβητες. Οι κύριες μηχανές που χρησιμοποιούνται για την προώθηση αποτελούν την κυρίαρχη πηγή των περισσότερων ρύπων, ιδιαίτερα κατά τη διάρκεια της κρουαζιέρας και των χαμηλών ταχυτήτων, ενώ οι βοηθητικές μηχανές και οι λέβητες συνεισφέρουν περισσότερο όταν τα πλοία βρίσκονται στο λιμάνι ή κατά τους ελιγμούς (Aakko-Saksa et al., 2023; Wan et al., 2020).

Το διοξείδιο του άνθρακα αποτελεί τον κύριο ρύπο που συνδέεται με την κλιματική αλλαγή και προέρχεται κυρίως από την καύση καυσίμων στις κύριες μηχανές, με συντελεστές εκπομπής

που κυμαίνονται από 600 έως 700 g/kWh. Τα οξείδια του αζώτου, με συντελεστές εκπομπής από 10 έως 18 g/kWh, εκπέμπονται τόσο από τις κύριες όσο και από τις βοηθητικές μηχανές και συνδέονται άμεσα με την ποιότητα του αέρα, την αναπνευστική υγεία και τη δημιουργία όξινης βροχής (Deng et al., 2020; Ni et al., 2020). Παράλληλα, τα οξείδια του θείου, με συντελεστές εκπομπής που κυμαίνονται από 1 έως 10 g/kWh ανάλογα με την περιεκτικότητα σε θείο του καυσίμου, προέρχονται κυρίως από τις κύριες μηχανές και τους λέβητες και συνδέονται με την όξινη βροχή, τη δημιουργία αιωρούμενων σωματιδίων και σοβαρούς κινδύνους για την υγεία (Ni et al., 2020; X. Wang et al., 2021).

Τα αιωρούμενα σωματίδια ( $PM_{10}$ ,  $PM_{2.5}$ ) εκπέμπονται από όλες τις μηχανές, ιδιαίτερα σε χαμηλές ταχύτητες, με συντελεστές εκπομπής από 0,5 έως 2 g/kWh, και συνδέονται με αναπνευστικά και καρδιαγγειακά προβλήματα υγείας, καθώς και με μείωση της ορατότητας (Ni et al., 2020; Deng et al., 2020). Ιδιαίτερα ανησυχητικό είναι το γεγονός ότι τα αιωρούμενα σωματίδια από τα πλοία περιέχουν βαρέα μέταλλα και πολυαρωματικούς υδρογονάνθρακες, αυξάνοντας σημαντικά τους κινδύνους για την υγεία (Aakko-Saksa et al., 2023; Zhao et al., 2021). Επιπλέον, τα  $NO_x$  και  $SO_x$  αποτελούν πρόδρομες ουσίες για τη δημιουργία δευτερογενών αιωρούμενων σωματιδίων και όζοντος, επιδεινώνοντας περαιτέρω την ποιότητα του αέρα και επηρεάζοντας αρνητικά την υγεία (Aakko-Saksa et al., 2023; Mueller et al., 2022).

Η γεωγραφική κατανομή των εκπομπών δείχνει ότι αυτές είναι υψηλότερες κοντά σε λιμάνια, ναυτιλιακές διαδρομές και κατά τη διάρκεια χαμηλών ταχυτήτων ή ελιγμών, καθιστώντας αυτές τις περιοχές ιδιαίτερα ευάλωτες στην ατμοσφαιρική ρύπανση από τη ναυτιλία (Weng et al., 2020). Τα πλοία εμπορευματοκιβωτίων και τα bulk carriers αποτελούν τους μεγαλύτερους συνεισφέροντες στις συνολικές εκπομπές, με τις εκπομπές να κορυφώνονται κατά τη διάρκεια της κρουαζιέρας και σε μεγάλα λιμάνια (Wan et al., 2020; Wang et al., 2021). Παρά τις πρόσφατες πολιτικές, όπως οι απαιτήσεις για καύσιμα χαμηλής περιεκτικότητας σε θείο που έχουν μειώσει τις εκπομπές  $SO_x$  και PM κατά 30-40%, τα  $NO_x$  και το  $CO_2$  παραμένουν σημαντικές προκλήσεις που απαιτούν περαιτέρω δράση για την αντιμετώπισή τους (Wang et al., 2021; Aakko-Saksa et al., 2023).

### 2.1.2 ΜΕΤΡΗΣΕΙΣ ΚΑΙ ΠΡΟΤΥΠΑ ΕΚΠΟΜΠΩΝ

Η μέτρηση των εκπομπών από τα πλοία σε πραγματικές συνθήκες λειτουργίας είναι απαραίτητη για τη ρυθμιστική συμμόρφωση και την αξιολόγηση των περιβαλλοντικών επιπτώσεων. Οι τρέχουσες μέθοδοι βασίζονται τόσο σε άμεσες επί πλοίου μετρήσεις όσο και σε τυποποιημένα πλαίσια υπολογισμού, αλλά υπάρχουν σημαντικές αβεβαιότητες και διαφορές μεταξύ των εργαστηριακών και των πραγματικών λειτουργικών μετρήσεων (Balcombe et al., 2022; Barreiro et al., 2022).

Οι άμεσες μετρήσεις πραγματοποιούνται με τη χρήση αισθητήρων επί πλοίου και φορητών συστημάτων μέτρησης εκπομπών που καταγράφουν τις εκπομπές από τις καπνοδόχους, τους αεραγωγούς και τις διάχυτες πηγές κατά τη διάρκεια πραγματικών ταξιδιών. Αυτή η προσέγγιση καταγράφει τις πραγματικές λειτουργικές συνθήκες, συμπεριλαμβανομένων των διακυμάνσεων στο φορτίο της μηχανής και τον τύπο καυσίμου (Balcombe et al., 2022). Παράλληλα, ο Διεθνής Ναυτιλιακός Οργανισμός (IMO) έχει καθιερώσει πρότυπα όπως ο Energy Efficiency Existing Ship Index (EEXI) και ο Carbon Intensity Indicator (CII), τα οποία απαιτούν από τα πλοία να παρακολουθούν και να αναφέρουν εκπομπές βάσει της κατανάλωσης καυσίμου, των χαρακτηριστικών της μηχανής και των λειτουργικών δεδομένων (Barreiro et al., 2022; Bayraktar & Yuksel, 2023).

Επιπλέον, τα λιμάνια και τα πλοία δημιουργούν απογραφές εκπομπών για την παρακολούθηση και διαχείριση των εκπομπών, χρησιμοποιώντας συχνά έναν συνδυασμό άμεσων μετρήσεων και τυποποιημένων μεθόδων (Hoang et al., 2022; B. Wang et al., 2023). Ωστόσο, οι πραγματικές μετρήσεις επηρεάζονται από παράγοντες όπως το φορτίο της μηχανής, η ποιότητα του καυσίμου και οι λειτουργικές πρακτικές, γεγονός που δημιουργεί σημαντικές αβεβαιότητες. Χαρακτηριστικό παράδειγμα αποτελούν οι μηχανές LNG, όπου τα ποσοστά διαφυγής μεθανίου είναι υψηλότερα σε χαμηλά φορτία, οδηγώντας σε υποεκτίμηση εάν δεν ληφθούν σωστά υπόψη (Balcombe et al., 2022).

Οι συντελεστές μετατροπής εκπομπών, όπως τα γραμμάρια CO<sub>2</sub> ανά kg καυσίμου, χρησιμοποιούνται για τη μετατροπή των δεδομένων κατανάλωσης καυσίμου σε εκτιμήσεις εκπομπών. Αυτοί οι συντελεστές μπορεί να ποικίλλουν ανάλογα με τον τύπο μηχανής, το καύσιμο και τις λειτουργικές συνθήκες, εισάγοντας περαιτέρω αβεβαιότητα στις εκτιμήσεις (Bayraktar & Yuksel, 2023; Balcombe et al., 2022). Η διαφορά μεταξύ εργαστηριακών και πραγματικών

μετρήσεων είναι ιδιαίτερα σημαντική, καθώς οι εργαστηριακές δοκιμές διενεργούνται σε ελεγχόμενες συνθήκες με τυποποιημένα καύσιμα και σταθερά φορτία μηχανής, παράγοντας αποτελέσματα που μπορούν να αναπαραχθούν. Ωστόσο, τέτοια αποτελέσματα μπορεί να μην αντιπροσωπεύουν την πραγματική μεταβλητότητα (Barreiro et al., 2022).

Αντίθετα, οι λειτουργικές μετρήσεις αντικατοπτρίζουν την πραγματική συμπεριφορά του πλοίου, συμπεριλαμβανομένων μεταβλητών φορτίων, καιρικών συνθηκών και συντήρησης, και συχνά αποκαλύπτουν υψηλότερες ή πιο μεταβλητές εκπομπές από τις εργαστηριακές εκτιμήσεις, ιδιαίτερα για ρύπους όπως το μεθάνιο (Balcombe et al., 2022; Barreiro et al., 2022). Αυτές οι διαφορές υπογραμμίζουν την ανάγκη για πιο ολοκληρωμένες και τυποποιημένες προσεγγίσεις μέτρησης που θα μπορούν να αντιμετωπίσουν τη λειτουργική μεταβλητότητα και τους περιορισμούς των συντελεστών μετατροπής.

### 2.1.3 ΓΕΩΓΡΑΦΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΕΚΠΟΜΠΩΝ

Οι εκπομπές από τα πλοία παρουσιάζουν σημαντικές διακυμάνσεις ανά περιοχή, λειτουργικό πλαίσιο και ρυθμιστικό περιβάλλον. Οι Περιοχές Ελέγχου Εκπομπών (Emission Control Areas - ECAs) και οι λιμενικοί κανονισμοί έχουν οδηγήσει σε σημαντικές μειώσεις των επιβλαβών ρύπων κοντά στις ακτές και στα λιμάνια, ενώ στην ανοιχτή θάλασσα και τα διεθνή ύδατα παρατηρούνται γενικά υψηλότερες εκπομπές λόγω λιγότερο αυστηρών ελέγχων (Mueller et al., 2022; Aakko-Saksa et al., 2023).

Οι υψηλότερες εκπομπές εντοπίζονται κατά μήκος των κύριων ναυτιλιακών διαδρομών και σε πολυσύχναστα λιμάνια, ιδιαίτερα κοντά σε πυκνοκατοικημένες παράκτιες περιοχές, όπου παρατηρούνται οι ισχυρότερες επιπτώσεις στην υγεία από την ατμοσφαιρική ρύπανση που προέρχεται από τα πλοία (Mueller et al., 2022). Οι ECA, όπως αυτές σε μέρη της Ευρώπης, της Βόρειας Αμερικής και της Ασίας, επιβάλλουν αυστηρότερα όρια για το θείο (SOx) και τα οξείδια του αζώτου (NOx), οδηγώντας σε σημαντικές μειώσεις αυτών των ρύπων εντός των ορίων τους σε σύγκριση με τις μη ελεγχόμενες περιοχές. Εκτός των ECA, τα πλοία συχνά χρησιμοποιούν καύσιμα υψηλότερης περιεκτικότητας σε θείο, με αποτέλεσμα μεγαλύτερες εκπομπές SOx και αιωρούμενων σωματιδίων (Mueller et al., 2022; Aakko-Saksa et al., 2023).

Η διάκριση μεταξύ εκπομπών σε λιμάνια και στην ανοιχτή θάλασσα αποκαλύπτει σημαντικές διαφορές στον χαρακτήρα και τις επιπτώσεις της ρύπανσης. Στα λιμάνια, οι εκπομπές

είναι συγκεντρωμένες και μπορούν να επηρεάσουν σημαντικά την τοπική ποιότητα του αέρα και την υγεία, ιδιαίτερα σε αστικές παράκτιες περιοχές, όπου οι λιμενικές δραστηριότητες όπως η φόρτωση/εκφόρτωση και οι βοηθητικές μηχανές συνεισφέρουν σε υψηλή τοπική ρύπανση (Mueller et al., 2022; Hoang et al., 2022). Αντίθετα, στην ανοιχτή θάλασσα οι εκπομπές είναι πιο διασκορπισμένες αλλά συνεισφέρουν στην παγκόσμια ρύπανση και την κλιματική αλλαγή, ενώ οι κανονισμοί είναι γενικά λιγότερο αυστηροί, εκτός από εκείνους εντός των ECAs (Mueller et al., 2022; Aakko-Saksa et al., 2023).

Η διαφοροποίηση μεταξύ διεθνών και εσωτερικών υδάτων επηρεάζει επίσης σημαντικά το επίπεδο των εκπομπών. Τα διεθνή ύδατα τυπικά έχουν λιγότερο αυστηρούς ελέγχους εκπομπών, οδηγώντας σε υψηλότερες εκπομπές ανά πλοίο, ενώ τα εσωτερικά ύδατα συχνά υπόκεινται σε εθνικούς κανονισμούς που μπορεί να είναι αυστηρότεροι, ιδιαίτερα κοντά σε πυκνοκατοικημένες ακτογραμμές ή εντός ECAs (Mueller et al., 2022; Aakko-Saksa et al., 2023). Οι περιορισμοί στη χρήση καυσίμων με υψηλή περιεκτικότητα σε θείο και οι τεχνολογίες μείωσης NOx στις ECAs και σε ορισμένα λιμάνια έχουν αποδειχθεί αποτελεσματικοί στη μείωση των εκπομπών και των συναφών επιβαρύνσεων για την υγεία (Mueller et al., 2022; Aakko-Saksa et al., 2023).

Παρά τις επιτυχίες των ρυθμιστικών μέτρων σε συγκεκριμένες περιοχές, οι εκπομπές παραμένουν υψηλότερες κατά μήκος των κύριων διαδρομών, στα λιμάνια και εκτός των ρυθμιζόμενων περιοχών. Οι τεχνολογικές παρεμβάσεις, όπως τα καθαρότερα καύσιμα και οι συστήματα μεταεπεξεργασίας καυσαερίων, είναι κλειδί για την επίτευξη περαιτέρω μειώσεων (Aakko-Saksa et al., 2023), ενώ η συνεχής επέκταση των ECAs και η υιοθέτηση καθαρότερων τεχνολογιών είναι κρίσιμες για περαιτέρω βελτιώσεις στη γεωγραφική κατανομή των εκπομπών από τη ναυτιλία.

## **2.2 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΑΙΣΘΗΤΗΡΩΝ**

Η κατανόηση των τύπων, των πηγών και της γεωγραφικής κατανομής των εκπομπών από τα πλοία αποτελεί τη βάση για την ανάπτυξη αποτελεσματικών στρατηγικών περιβαλλοντικής διαχείρισης. Ωστόσο, η ακριβής παρακολούθηση και καταγραφή αυτών των εκπομπών απαιτεί προηγμένα τεχνολογικά συστήματα που μπορούν να λειτουργούν αξιόπιστα στις απαιτητικές συνθήκες του θαλάσσιου περιβάλλοντος.

### 2.2.1 ΣΥΓΧΡΟΝΑ ΣΥΣΤΗΜΑΤΑ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ

Η παρακολούθηση των εκπομπών και της κατανάλωσης καυσίμου των πλοίων είναι κρίσιμη για τη ρυθμιστική συμμόρφωση και την περιβαλλοντική βιωσιμότητα στις θαλάσσιες μεταφορές. Τα σύγχρονα συστήματα τελευταίας τεχνολογίας συνδυάζουν αισθητήρες επί πλοίου, παρακολούθηση απόδοσης μηχανών και αυτοματοποιημένη καταγραφή δεδομένων, συχνά ενσωματωμένα με προηγμένες αναλυτικές μεθόδους και ρυθμιστικά πλαίσια (X. Chen et al., 2024; M. Zhang et al., 2024).

Οι αισθητήρες επί πλοίου αποτελούν τον πυρήνα των σύγχρονων συστημάτων παρακολούθησης, μετρώντας σε πραγματικό χρόνο την κατανάλωση καυσίμου, το φορτίο της μηχανής και τις εκπομπές βασικών ρύπων όπως CO<sub>2</sub>, NO<sub>x</sub>, SO<sub>2</sub> και αιωρούμενα σωματίδια (PM). Αυτά τα συστήματα χαρακτηρίζονται από συλλογή δεδομένων υψηλής συχνότητας και υποστηρίζουν άμεσους υπολογισμούς για βελτιωμένη ακρίβεια, με τις άμεσες μεθόδους υπολογισμού που χρησιμοποιούν αισθητήρες επί πλοίου να παράγουν έως και 30% πιο ακριβείς συντελεστές εκπομπών από τα μοντέλα εκτίμησης (Fan et al., 2023; M. Zhang et al., 2024).

Η παρακολούθηση απόδοσης μηχανών συμπληρώνει τα συστήματα αισθητήρων μέσω της παρακολούθησης κρίσιμων παραμέτρων όπως το φορτίο, η ταχύτητα, η ισχύς και η απόδοση της μηχανής. Αυτή η λειτουργικότητα επιτρέπει τη βελτιστοποίηση της λειτουργίας της μηχανής και της χρήσης καυσίμου, συμβάλλοντας στη μείωση των εκπομπών μέσω καλύτερου ελέγχου των λειτουργικών συνθηκών. Παράλληλα, τα αυτοματοποιημένα συστήματα καταγραφής δεδομένων συλλέγουν και καταγράφουν λειτουργικά και περιβαλλοντικά δεδομένα όπως ταχύτητα, καιρικές συνθήκες, διαφορά βάρους και άλλες παραμέτρους, υποστηρίζοντας αναλυτικές μεθόδους μεγάλων δεδομένων και προγνωστική μοντελοποίηση (Chen et al., 2024; Zhang et al., 2024).

Τα σύγχρονα συστήματα ενσωματώνουν επίσης προηγμένες τεχνολογίες τεχνητής νοημοσύνης και βαθιάς μάθησης για την πρόβλεψη κατανάλωσης καυσίμου και εκπομπών υπό πραγματικές συνθήκες. Ιδιαίτερα αποτελεσματικές έχουν αποδειχθεί οι αρχιτεκτονικές Bi-LSTM με μηχανισμούς προσοχής, οι οποίες ξεπερνούν τα παραδοσιακά μοντέλα σε ακρίβεια πρόβλεψης (Zhang et al., 2024). Επιπλέον, οι ρυθμιστικοί δείκτες όπως EEOI, FEEMI, EEXI και CII είναι κεντρικοί στην παρακολούθηση και αναφορά, ποσοτικοποιώντας και δημιουργώντας σημεία αναφοράς για την ενεργειακή απόδοση και την ένταση άνθρακα, καθοδηγώντας λειτουργικές προσαρμογές και αναβαθμίσεις.

Η ολοκλήρωση αυτών των τεχνολογιών επιτρέπει τη δημιουργία ολοκληρωμένων ροών δεδομένων που συνδυάζουν δεδομένα αισθητήρων, αναφορές ταξιδιών και περιβαλλοντικές εισόδους για την παροχή μιας ολιστικής εικόνας της απόδοσης και των εκπομπών του πλοίου (Chen et al., 2024; Zhang et al., 2024). Παρά τις σημαντικές προόδους, παραμένουν προκλήσεις όπως η λειτουργία των πλοίων σε μερικά φορτία μηχανής, που μπορεί να αυξήσει τις εκπομπές ανά μονάδα εξόδου, υπογραμμίζοντας την ανάγκη για συνεχή παρακολούθηση και βελτιστοποίηση (Fan et al., 2023).

### 2.2.2 ΤΕΧΝΟΛΟΓΙΕΣ IOT ΣΤΗ ΝΑΥΤΗΛΙΑ

Οι τεχνολογίες Internet of Things (IoT) και τα ασύρματα δίκτυα αισθητήρων (Wireless Sensor Networks - WSNs) μετασχηματίζουν τη ναυτιλιακή βιομηχανία επιτρέποντας την παρακολούθηση του περιβάλλοντος σε πραγματικό χρόνο, από απόσταση και με ασφάλεια. Αυτά τα συστήματα χρησιμοποιούν έναν συνδυασμό αισθητήρων επί πλοίου και δορυφορικών επικοινωνιών για την αποτελεσματική συλλογή, επεξεργασία και μετάδοση περιβαλλοντικών δεδομένων, ακόμη και σε απομονωμένες περιοχές του ωκεανού (Chaudhry et al., 2023; Narayana et al., 2024; C. Wang et al., 2023).

Τα πλοία και οι θαλάσσιοι τερματικοί σταθμοί εξοπλίζονται με ποικίλους αισθητήρες που μετρούν την ποιότητα του αέρα, τις παραμέτρους του νερού, τις καιρικές συνθήκες και άλλες περιβαλλοντικές μεταβλητές, σχηματίζοντας WSNs για συνεχή περιβαλλοντική παρακολούθηση. Ιδιαίτερα σημαντική είναι η επέκταση της παρακολούθησης σε υποθαλάσσια περιβάλλοντα μέσω των υποβρύχιων ασύρματων δικτύων αισθητήρων, τα οποία υποστηρίζουν εφαρμογές όπως η παρακολούθηση της ρύπανσης και η πρόβλεψη καταστροφών (Chaudhary et al., 2023; Kaidarova et al., 2023).

Οι δορυφορικές επικοινωνίες αποτελούν απαραίτητο στοιχείο για τη μετάδοση δεδομένων σε απομακρυσμένες θαλάσσιες περιοχές όπου η επίγεια κάλυψη δεν είναι διαθέσιμη, εξασφαλίζοντας παγκόσμια συνδεσιμότητα για τις συσκευές IoT και τα WSNs. Αυτά τα δίκτυα επιτρέπουν σε απομονωμένα σκάφη να μεταδίδουν με ασφάλεια περιβαλλοντικά και λειτουργικά δεδομένα σε επίγεια κέντρα για ανάλυση, υποστηρίζοντας έτσι την αποτελεσματική παρακολούθηση ακόμη και στις πιο απομακρυσμένες θαλάσσιες περιοχές (Chaudhry et al., 2023; Wang et al., 2023).



Η ασφάλεια και η ακεραιότητα των δεδομένων αποτελούν κρίσιμες πτυχές των IoT συστημάτων στη ναυτιλία, λόγω της ευαισθησίας των περιβαλλοντικών και λειτουργικών δεδομένων. Προηγμένα σχήματα πιστοποίησης, κρυπτογράφησης και blockchain προστατεύουν τα ευαίσθητα περιβαλλοντικά και λειτουργικά δεδομένα από κυβερνοαπειλές. Επιπλέον, οι έλεγχοι ακεραιότητας δεδομένων και οι μηχανισμοί ανάκτησης εξασφαλίζουν αξιόπιστη μακροπρόθεσμη αποθήκευση και ανάλυση, διασφαλίζοντας ότι τα συλλεχθέντα δεδομένα παραμένουν ακριβή και χρήσιμα για μελλοντικές αναλύσεις (D. Liu et al., 2023; P. Zhang et al., 2023).

Η εφαρμογή των IoT τεχνολογιών και των WSNs, σε συνδυασμό με τις δορυφορικές επικοινωνίες και την υπολογιστική στο edge, επιτρέπει την ισχυρή, σε πραγματικό χρόνο περιβαλλοντική παρακολούθηση στη ναυτιλία. Αυτές οι τεχνολογίες παρέχουν παγκόσμια κάλυψη, ταχεία επεξεργασία δεδομένων και ισχυρή ασφάλεια, υποστηρίζοντας ασφαλέστερες και πιο βιώσιμες ναυτιλιακές λειτουργίες που μπορούν να ανταποκριθούν αποτελεσματικά στις σύγχρονες περιβαλλοντικές προκλήσεις.

### 2.2.3 ΠΟΙΟΤΗΤΑ ΚΑΙ ΑΞΙΟΠΙΣΤΙΑ ΔΕΔΟΜΕΝΩΝ

Η εξασφάλιση υψηλής ποιότητας και αξιοπιστίας δεδομένων στα ναυτιλιακά συστήματα αισθητήρων είναι κρίσιμη για ασφαλείς και αποτελεσματικές ναυτιλιακές λειτουργίες. Οι κύριες προκλήσεις περιλαμβάνουν τη βαθμονόμηση αισθητήρων, τη διαχείριση ελλিপών ή εσφαλμένων δεδομένων και την καθιέρωση ισχυρών προτύπων επικύρωσης δεδομένων. Τα βασικά ζητήματα είναι η ακεραιότητα των δεδομένων, οι βλάβες αισθητήρων, η απώλεια δεδομένων και η έλλειψη ενιαίων πρωτοκόλλων επικύρωσης (J. Liu et al., 2023; Y. Wang et al., 2023).

Τα δεδομένα από τους ναυτιλιακούς αισθητήρες είναι ευάλωτα σε αλλοίωση, απώλεια και κακόβουλες επιθέσεις, ιδιαίτερα σε συστήματα με δυνατότητες IoT. Ευέλικτοι μηχανισμοί ελέγχου ακεραιότητας δεδομένων και ανάκτησης, όπως η κωδικοποίηση διαγραφής (erasure coding), είναι απαραίτητοι για να εξασφαλιστεί ότι τα αλλοιωμένα ή διαγραμμένα δεδομένα μπορούν να αποκατασταθούν και να επαληθευτούν αποτελεσματικά πριν από τη χρήση τους (Liu et al., 2023; Wang et al., 2023). Επιπλέον, η ασφαλής συγκέντρωση και κρυπτογράφηση των δεδομένων αισθητήρων είναι απαραίτητη για τη διατήρηση της αξιοπιστίας, ιδιαίτερα για απομονωμένους ναυτιλιακούς τερματικούς που βασίζονται σε δορυφορικές συνδέσεις (Wang et al., 2023).

Οι βλάβες αισθητήρων και τα σφάλματα βαθμονόμησης μπορούν να επηρεάσουν σημαντικά την αξιοπιστία των δεδομένων. Προσαρμοστικές στρατηγικές ελέγχου και αλγόριθμοι ανοχής βλαβών αισθητήρων είναι απαραίτητοι για τη διατήρηση της σταθερότητας του συστήματος και της ακρίβειας παρακολούθησης, ακόμη και παρουσία βλαβών αισθητήρων ή αβεβαιοτήτων μοντέλου (Li et al., 2024). Ιδιαίτερα προβληματικά είναι τα σφάλματα εντοπισμού κόμβων σε ασύρματα δίκτυα αισθητήρων, τα οποία μπορούν να καταστήσουν τα δεδομένα άχρηστα. Προηγμένοι αλγόριθμοι, όπως αυτοί που βασίζονται σε νοημοσύνη σμήνους (swarm intelligence), βοηθούν στη μείωση των σφαλμάτων εντοπισμού και στη βελτίωση της ακρίβειας των δεδομένων (Rani et al., 2023).

Η απώλεια δεδομένων λόγω κενών επικοινωνίας, όπως σε υποβρύχιες ή απομακρυσμένες περιοχές, και οι βλάβες αισθητήρων είναι συνηθισμένες προκλήσεις. Τεχνικές ανάκτησης δεδομένων και ισχυρή συγχώνευση δεδομένων, που συνδυάζουν δεδομένα AIS (Automatic Identification System) και οπτικά δεδομένα, βοηθούν στον μετριασμό των επιπτώσεων των ελλειπών ή εσφαλμένων δεδομένων (J. Liu et al., 2023; Qu et al., 2023). Οι data-driven μέθοδοι για την ανίχνευση ανώμαλης συμπεριφοράς πλοίων πρέπει να αντιμετωπίσουν προκλήσεις που σχετίζονται με τη συγχώνευση δεδομένων από πολλές πηγές, τα ελλιπή δεδομένα και την ετερογένεια των δεδομένων (Wang et al., 2023).

Παρά τις προόδους στον έλεγχο ακεραιότητας δεδομένων, την ανοχή βλαβών αισθητήρων, την ισχυρή συγχώνευση δεδομένων και την ασφαλή πιστοποίηση, τα ναυτιλιακά συστήματα αισθητήρων εξακολουθούν να αντιμετωπίζουν επίμονες προκλήσεις στην ποιότητα και αξιοπιστία των δεδομένων. Τα πρότυπα σε επίπεδο βιομηχανίας για την επικύρωση δεδομένων παραμένουν υπό ανάπτυξη, καθιστώντας απαραίτητη τη συνεχή ερευνητική προσπάθεια για τη δημιουργία πιο αξιόπιστων και τυποποιημένων συστημάτων διαχείρισης δεδομένων στη ναυτιλία.

## **2.3 ΥΠΑΡΧΟΥΣΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗΣ ΑΠΟΔΟΣΗΣ**

Η διασφάλιση υψηλής ποιότητας και αξιοπιστίας των δεδομένων από τα συστήματα αισθητήρων αποτελεί προϋπόθεση για την αποτελεσματική αξιολόγηση της περιβαλλοντικής απόδοσης των πλοίων, η οποία πραγματοποιείται μέσω διαφόρων μεθόδων και δεικτών που έχουν αναπτυχθεί τόσο από διεθνείς οργανισμούς όσο και από ιδιωτικές πρωτοβουλίες.

### 2.3.1 ΡΥΘΜΙΣΤΙΚΑ ΜΕΤΡΑ ΤΟΥ IMO

Ο IMO έχει εισαγάγει διάφορα ρυθμιστικά μέτρα – EEDI, EEOI, SEEMP και CII – για την αξιολόγηση και βελτίωση της περιβαλλοντικής απόδοσης των πλοίων, με στόχο τη μείωση των εκπομπών αερίων θερμοκηπίου από τον ναυτιλιακό τομέα. Αυτά τα μέτρα χρησιμοποιούν τόσο δείκτες σχεδίασης όσο και λειτουργικούς δείκτες, αλλά η αποτελεσματικότητά τους και η εφαρμογή τους αντιμετωπίζουν τεχνικές, λειτουργικές και πολιτικές προκλήσεις (Bayraktar & Yuksel, 2023; Lee, 2024).

Ο Δείκτης Σχεδιαστικής Ενεργειακής Απόδοσης (EEDI) ορίζει το ελάχιστο αποδεκτό επίπεδο ενεργειακής αποδοτικότητας για τα νεότευκτα πλοία και εκφράζεται ως εκπομπές CO<sub>2</sub> ανά μονάδα μεταφορικού έργου (g CO<sub>2</sub>/τόνο-μίλι), βάσει συγκεκριμένων παραμέτρων σχεδίασης (Εξ. 2.1) (IMO MARPOL Annex VI, Regulation 20, MEPC.245(66)). Ο EEDI εφαρμόζεται μόνο σε νέα πλοία και δεν αντιμετωπίζει τις λειτουργικές πρακτικές, αλλά επιτυγχάνει 30-40% μείωση CO<sub>2</sub> για νέα πλοία έως το 2030 (Lee, 2024; Bayraktar & Yuksel, 2023).

$$EEDI = \frac{\sum_i (P_{ME(i)} \cdot C_{F(i)} \cdot SFC_{ME(i)}) + \sum_j (P_{AE(j)} \cdot C_{F(j)} \cdot SFC_{AE(j)}) - P_{efficiency}}{Capacity \cdot V_{ref}} \quad (2.1)$$

Παράλληλα, ο Energy Efficiency Existing Ship Index (EEXI) εφαρμόζει πρότυπα παρόμοια με τον EEDI σε υπάρχοντα πλοία, υπολογιζόμενος με βάση τη μέγιστη ισχύ της μηχανής και την ταχύτητα. Ωστόσο, οι υπολογισμοί μπορεί να υπερεκτιμούν την απόδοση λόγω της χρήσης μέγιστης ταχύτητας, ενώ υπάρχει κίνδυνος οι ιδιοκτήτες να αυξήσουν την ταχύτητα στα επιτρεπόμενα όρια, με αποτέλεσμα ο EEXI να παρέχει 17-25% μείωση CO<sub>2</sub>, λιγότερη από τον EEDI (Lee, 2024; Bayraktar & Yuksel, 2023).

Το Ship Energy Efficiency Management Plan (SEEMP) απαιτεί από όλα τα πλοία να έχουν ένα σχέδιο διαχείρισης για τη βελτίωση της λειτουργικής απόδοσης, συμπεριλαμβανομένης της παρακολούθησης και αναφοράς. Παρά τη σημασία του, το SEEMP αυξάνει το φόρτο εργασίας του πληρώματος, τη γραφειοκρατία και την τεχνική πολυπλοκότητα, ενώ προσφέρει περιορισμένα κίνητρα για τους ναυτικούς. Η αποτελεσματικότητά του υποστηρίζει τη συμμόρφωση, αλλά εξαρτάται από τη συμμετοχή του πληρώματος και τη διαχείριση (Dewan & Godina, 2023, 2024).

Ο Carbon Intensity Indicator (CII) αξιολογεί τα πλοία ετησίως (A-E) με βάση τις λειτουργικές εκπομπές CO<sub>2</sub> ανά εργασία μεταφοράς (βλ. Εξ. 2.2) (MARPOL Annex VI, Regulation 28). Το 1/3 των πλοίων δεν συμμορφώνεται επί του παρόντος, ενώ απαιτείται μείωση

ταχύτητας ή τεχνικές αναβαθμίσεις, καθώς οι αξιολογήσεις μπορεί να μην αντικατοπτρίζουν την πραγματική περιβαλλοντική επίδραση. Ωστόσο, η μείωση ταχύτητας και η βελτιστοποίηση διαδρομών μπορούν να επιτύχουν συμμόρφωση και έως 50% μείωση εκπομπών (Yuan et al., 2023; Zincir, 2023).

$$CII = \frac{\sum_i (FC_i \cdot CF_i)}{Capacity \cdot Distance} \quad (2.2)$$

Οι προκλήσεις εφαρμογής αυτών των μέτρων περιλαμβάνουν τεχνικές και λειτουργικές δυσκολίες, όπως η αναμόρφωση παλαιότερων πλοίων, η βελτιστοποίηση ταχύτητας και η μετάβαση σε εναλλακτικά καύσιμα που είναι πολύπλοκες και δαπανηρές. Το πλήρωμα αντιμετωπίζει αυξημένο φόρτο εργασίας και χρειάζεται περισσότερη εκπαίδευση και κίνητρα για να εξασφαλίσει τη συμμόρφωση (Bayraktar & Yuksel, 2023; Dewan & Godina, 2023). Επιπλέον, ο IMO αντιμετωπίζει δυσκολίες στη ρύθμιση αναδυόμενων τεχνολογιών, έλλειψη πολιτικής συναίνεσης και περιορισμένη ικανότητα για ολοκληρωμένη επιβολή (Bach & Hansen, 2023), ενώ ορισμένοι δείκτες όπως ο EEXI μπορεί να μην αντικατοπτρίζουν ακριβώς τις πραγματικές λειτουργίες, και οι αξιολογήσεις CII μπορούν να συγκαλύπτουν διαφορές στις πραγματικές εκπομπές μεταξύ πλοίων με την ίδια αξιολόγηση (Lee, 2024; Zincir, 2023).

### 2.3.2 ΕΝΑΛΛΑΚΤΙΚΟΙ ΔΕΙΚΤΕΣ ΚΑΙ ΠΡΩΤΟΒΟΥΛΙΕΣ

Πέραν των ρυθμιστικών μέτρων του IMO, υπάρχουν διάφοροι εναλλακτικοί δείκτες περιβαλλοντικής απόδοσης και πρωτοβουλίες, καθένας με διακριτές μεθοδολογίες και πεδία εφαρμογής. Αυτά τα εργαλεία έχουν σχεδιαστεί για να αξιολογούν, να παρακολουθούν και να βελτιώνουν την περιβαλλοντική απόδοση στη ναυτιλία και σε άλλους τομείς. Αρχικά, το CSI αποτελεί έναν ανεξάρτητο, εθελοντικό και αυστηρά πιστοποιημένο δείκτη, που βαθμολογεί την απόδοση πλοίων σε πέντε κύριες κατηγορίες (CO<sub>2</sub>, NO<sub>x</sub>, SO<sub>x</sub>/PM, χρήση χημικών, διαχείριση νερών και αποβλήτων), με μέγιστο αριθμό 150 βαθμών και ενιαία βαθμολογία 30 ανά κατηγορία (όπου για κάθε τομέα δίνεται βαθμός μόνο εφόσον οι επιδόσεις υπερβαίνουν τη νομοθετική συμμόρφωση) (Clean Shipping Index, 2024). Στη συνέχεια, ο εφοπλιστής υποβάλλει αναλυτικά δεδομένα σε εξωτερική κριτική αρχή (classification society), όπου περιλαμβάνεται και ο υπολογισμός του EEOI κατά τις κατευθυντήριες οδηγίες IMO (MEPC.1/Circ.684), ώστε να διασφαλιστεί η εγκυρότητα της βαθμολογίας.

Επιπλέον, ο ESI (υπό την αιγίδα της IAPH σε συνεργασία με την Green Award Foundation) βαθμολογεί πλοία με βάση τις εκπομπές NO<sub>x</sub>, SO<sub>x</sub> και CO<sub>2</sub>, παρέχοντας μερικές φορές bonus 10 μονάδων για εγκατάσταση On-Shore Power Supply (OPS) (Environmental Ship Index (ESI), 2020). Ο συνολικός υπολογισμός γίνεται ως:

$$ESI\_Score = (2 \times NO_{x\_sub} + SO_{x\_sub} + CO_{2\_sub} + OPS\_bonus), \text{ με ανώτατο } 100 \quad (2.3)$$

Ως αποτέλεσμα, πάνω από 6.700 πλοία και περίπου 60 λιμένες χρησιμοποιούν τον ESI για να χορηγούν εκπτώσεις ή άλλα εμπορικά κίνητρα σε πλοία με χαμηλότερο περιβαλλοντικό αποτύπωμα (International Association of Ports and Harbors (IAPH), 2025).

Καθένας από τους δύο δείκτες εξυπηρετεί διαφορετικό ρόλο: η CSI προσφέρει ολιστική εικόνα πολλαπλών ρύπων και προορίζεται για εμπορική διαφοροποίηση, ενώ το ESI δίνει απευθείας οικονομικά κίνητρα από λιμένες για μειώσεις NO<sub>x</sub>/SO<sub>x</sub>/CO<sub>2</sub>. Παρά την εθελοντική βάση τους, οι δείκτες αυτοί συμπληρώνουν και ενισχύουν τους ρύθμους IMO, αξιοποιώντας στοιχεία που παράγονται μέσω της υποβολής δεδομένων όπως το IMO-DCS (για CO<sub>2</sub> στο ESI και CSI). Κάθε δείκτης δίνει έμφαση είτε στη βελτίωση απόδοσης πέραν της νομοθεσίας (CSI), είτε στη βασισμένη σε υλοποιημένα οφέλη προσέγγιση (ESI), είτε στην θεωρητική αξιολόγηση του σχεδιαστικού δυναμικού (EVDI/GHG Rating). Μαζί παρέχουν πλήρη κάλυψη από την επιλογή σχεδίου μέχρι την εμπορική διαχείριση και την πρακτική λειτουργία.

### 2.3.3 ΑΝΑΛΥΣΗ ΚΥΚΛΟΥ ΖΩΗΣ ΣΤΗ ΝΑΥΤΙΛΙΑ

Η ανάλυση κύκλου ζωής (Life Cycle Assessment - LCA) χρησιμοποιείται ευρέως για την αξιολόγηση των περιβαλλοντικών επιπτώσεων της ναυτιλίας, αλλά διάφορα μεθοδολογικά ζητήματα περιορίζουν την αξιοπιστία και τη συγκρισιμότητά της. Οι βασικές προκλήσεις περιλαμβάνουν ασυνεπή όρια συστήματος, ασαφείς ορισμούς λειτουργικών μονάδων και έλλειψη τυποποιημένων προσεγγίσεων, όλα τα οποία μπορούν να επηρεάσουν σημαντικά τα αποτελέσματα και τη λήψη αποφάσεων (Ha et al., 2023).

Ένα από τα σημαντικότερα μεθοδολογικά ζητήματα αφορά τον καθορισμό των ορίων συστήματος στις μελέτες LCA. Οι μελέτες LCA συχνά διαφέρουν στον τρόπο που ορίζουν τα όρια του συστήματος, όπως το αν θα περιλαμβάνουν μόνο τις άμεσες εκπομπές από τη ναυτιλία ή επίσης τις ανάντη και κατάντη διαδικασίες, όπως την παραγωγή καυσίμων και τις υποδομές (Ha et al.,

2023). Η επιλογή των ορίων, όπως η προσέγγιση Well-to-Tank έναντι της Well-to-Wake, μπορεί να οδηγήσει σε μεγάλες διακυμάνσεις στις αναφερόμενες εκπομπές, ιδιαίτερα όταν λαμβάνονται υπόψη οι διεθνείς εφοδιαστικές αλυσίδες και οι περιφερειακές διαφορές στις αποστάσεις μεταφοράς (Ha et al., 2023). Αυτή η ασυνέπεια στον καθορισμό των ορίων συστήματος μπορεί να οδηγήσει σε εντελώς διαφορετικά συμπεράσματα σχετικά με την περιβαλλοντική απόδοση των ίδιων ναυτιλιακών δραστηριοτήτων.

Οι ορισμοί των λειτουργικών μονάδων αποτελούν έναν άλλο κρίσιμο τομέα προβλημάτων στις ναυτιλιακές μελέτες LCA. Υπάρχει ασυνέπεια στον ορισμό της λειτουργικής μονάδας, όπως ανά τόνο-χιλιόμετρο, ανά MJ καυσίμου ή ανά ταξίδι, καθιστώντας δύσκολη τη σύγκριση αποτελεσμάτων μεταξύ μελετών (Ha et al., 2023). Η λειτουργική μονάδα θα πρέπει να αντικατοπτρίζει τον σκοπό της αξιολόγησης και να είναι σαφώς αιτιολογημένη, αλλά αυτό συχνά δεν συμβαίνει, οδηγώντας σε ασάφεια στην ερμηνεία των αποτελεσμάτων (Ha et al., 2023). Η επιλογή της κατάλληλης λειτουργικής μονάδας είναι ιδιαίτερα σημαντική στη ναυτιλία, όπου διαφορετικοί τύποι πλοίων και φορτίων έχουν εντελώς διαφορετικά χαρακτηριστικά απόδοσης.

Η ποιότητα των δεδομένων και η απογραφή αποτελούν επίσης σημαντικές προκλήσεις στις ναυτιλιακές μελέτες LCA. Πολλές μελέτες βασίζονται σε γενικά ή ελλιπή δεδομένα, τα οποία υπονομεύουν την αξιοπιστία των αποτελεσμάτων. Δεδομένα απογραφής υψηλής ποιότητας και ειδικά για το πλαίσιο είναι απαραίτητα για ακριβείς αξιολογήσεις (Ha et al., 2023). Η μεταβλητότητα στα συστήματα προώθησης, τους τύπους καυσίμων και τις διαδρομές περιπλέκει περαιτέρω τη συλλογή δεδομένων και τη συγκρισιμότητα (Ha et al., 2023). Αυτό το πρόβλημα επιδεινώνεται από το γεγονός ότι η ναυτιλιακή βιομηχανία παρουσιάζει τεράστια ποικιλομορφία σε όρους τεχνολογιών, λειτουργικών πρακτικών και γεωγραφικών περιοχών λειτουργίας.

Η έλλειψη τυποποίησης στις μεθοδολογίες LCA για τη ναυτιλία δημιουργεί σημαντικά προβλήματα συγκρισιμότητας και διαφάνειας. Η απουσία τυποποιημένων οδηγιών για τη ναυτιλιακή LCA οδηγεί σε υποκειμενικές μεθοδολογικές επιλογές, μειώνοντας τη διαφάνεια και τη συγκρισιμότητα (Ha et al., 2023). Υπάρχει έντονη ανάγκη για εναρμονισμένες προσεγγίσεις που να λαμβάνουν υπόψη τις περιφερειακές διαφορές και τις ιδιαιτερότητες των εφοδιαστικών αλυσίδων (Ha et al., 2023). Η έλλειψη αυτή καθιστά δύσκολη την ανάπτυξη συνεκτικών πολιτικών και την αξιόπιστη σύγκριση διαφορετικών ναυτιλιακών τεχνολογιών και πρακτικών.

Συνοψίζοντας, οι μελέτες LCA στη ναυτιλία αντιμετωπίζουν σημαντικές μεθοδολογικές προκλήσεις, ιδιαίτερα όσον αφορά τα όρια συστήματος, τις λειτουργικές μονάδες και την

τυποποίηση. Η αντιμετώπιση αυτών των ζητημάτων είναι κρίσιμη για την παραγωγή αξιόπιστων, συγκρίσιμων και σχετικών με την πολιτική αποτελεσμάτων που θα μπορούν να καθοδηγήσουν αποτελεσματικά την περιβαλλοντική βελτίωση της ναυτιλιακής βιομηχανίας. Χωρίς τη λύση αυτών των θεμελιωδών μεθοδολογικών προβλημάτων, οι μελέτες LCA κινδυνεύουν να παρέχουν παραπλανητικές πληροφορίες που μπορεί να οδηγήσουν σε λανθασμένες αποφάσεις πολιτικής και επενδύσεων.

## **2.4 ΕΦΑΡΜΟΓΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΗ ΝΑΥΤΙΛΙΑ**

Παρά τους περιορισμούς και τις προκλήσεις που αντιμετωπίζουν οι παραδοσιακές μέθοδοι αξιολόγησης περιβαλλοντικής απόδοσης, η ανάπτυξη νέων τεχνολογικών εργαλείων και μεθοδολογιών προσφέρει σημαντικές ευκαιρίες για τη βελτίωση της ακρίβειας και της αποτελεσματικότητας των αξιολογήσεων στον ναυτιλιακό τομέα.

### **2.4.1 ΠΡΟΒΛΕΨΗ ΚΑΤΑΝΑΛΩΣΗΣ ΚΑΥΣΙΜΟΥ**

Οι τεχνικές μηχανικής μάθησης, συμπεριλαμβανομένων των τεχνητών νευρωνικών δικτύων (Artificial Neural Networks - ANN), της παλινδρόμησης Gaussian Process (Gaussian Process Regression - GPR) και των Support Vector Machines (SVM), χρησιμοποιούνται όλο και περισσότερο για την πρόβλεψη της κατανάλωσης καυσίμου και της ισχύος άξονα των πλοίων. Αυτές οι μέθοδοι αξιοποιούν μεγάλα σύνολα δεδομένων από αισθητήρες και λειτουργικά αρχεία για να συλλάβουν πολύπλοκες, πραγματικές επιρροές στην απόδοση των πλοίων. Τα μοντέλα μηχανικής μάθησης, ιδιαίτερα οι προσεγγίσεις βαθιάς μάθησης, γενικά ξεπερνούν τα παραδοσιακά μοντέλα που βασίζονται στην αντίσταση όσον αφορά την ακρίβεια και την προσαρμοστικότητα σε μεταβαλλόμενες συνθήκες (Zhang et al., 2024).

Οι προηγμένες μέθοδοι βαθιάς μάθησης, όπως τα Bi-directional Long Short-Term Memory (Bi-LSTM) δίκτυα με μηχανισμούς προσοχής, μπορούν να επεξεργαστούν δεδομένα υψηλής συχνότητας και πολλαπλών μεταβλητών, όπως ταχύτητα, βύθισμα, καιρικές συνθήκες και κατάσταση θάλασσας, για να προβλέψουν την κατανάλωση καυσίμου με υψηλή ακρίβεια. Αυτά τα μοντέλα μπορούν να μάθουν περίπλοκες σχέσεις και να προσαρμοστούν σε πραγματικές λειτουργικές συνθήκες, καθιστώντας τα κατάλληλα για την υποστήριξη αποφάσεων σε βιώσιμες

ναυτιλιακές λειτουργίες (Zhang et al., 2024). Η ικανότητά τους να ενσωματώνουν χρονικές εξαρτήσεις και να εντοπίζουν μη γραμμικά πρότυπα τα καθιστά ιδιαίτερα αποτελεσματικά για την πρόβλεψη κατανάλωσης καυσίμου σε πραγματικές, δυναμικές συνθήκες λειτουργίας.

Οι παραδοσιακές τεχνικές μηχανικής μάθησης, όπως ANN, GPR και SVM, χρησιμοποιούνται επίσης για εργασίες πρόβλεψης, προσφέροντας καλή απόδοση μέσω της μοντελοποίησης μη γραμμικών σχέσεων μεταξύ μεταβλητών εισόδου και εξόδου. Είναι λιγότερο πολύπλοκα από τα μοντέλα βαθιάς μάθησης, αλλά εξακολουθούν να ξεπερνούν τα απλά φυσικά μοντέλα όταν είναι διαθέσιμα επαρκή δεδομένα (Zhang et al., 2024). Αυτές οι μέθοδοι προσφέρουν ένα καλό συμβιβασμό μεταξύ πολυπλοκότητας και απόδοσης, καθιστώντας τις κατάλληλες για εφαρμογές όπου οι απαιτήσεις υπολογιστικής ισχύος είναι περιορισμένες.

Η σύγκριση με τα παραδοσιακά μοντέλα που βασίζονται στην αντίσταση αποκαλύπτει σημαντικές διαφορές στην απόδοση. Τα παραδοσιακά μοντέλα αντίστασης βασίζονται σε απλοποιημένη φυσική και είναι λιγότερο ακριβή υπό μεταβαλλόμενες λειτουργικές και περιβαλλοντικές συνθήκες. Είναι ευκολότερα στην εφαρμογή αλλά στερούνται προσαρμοστικότητας (Zhang et al., 2024). Αντίθετα, τα μοντέλα μηχανικής μάθησης μπορούν να ενσωματώσουν ένα ευρύ φάσμα μεταβλητών και να προσαρμοστούν σε νέα δεδομένα, παρέχοντας πιο ακριβείς και ισχυρές προβλέψεις, ιδιαίτερα όταν οι λειτουργικές συνθήκες αλλάζουν (Zhang et al., 2024).

Όσον αφορά την ακρίβεια σε πραγματικές συνθήκες, τα μοντέλα που βασίζονται στην αντίσταση επιτυγχάνουν μέτρια απόδοση με χαμηλή προσαρμοστικότητα και χαμηλές απαιτήσεις δεδομένων. Τα παραδοσιακά μοντέλα μηχανικής μάθησης (ANN, GPR, SVM) επιτυγχάνουν υψηλή ακρίβεια με μέτρια προσαρμοστικότητα και μέτριες απαιτήσεις δεδομένων, ενώ τα μοντέλα βαθιάς μάθησης (Bi-LSTM) παρέχουν πολύ υψηλή ακρίβεια με υψηλή προσαρμοστικότητα, αλλά απαιτούν υψηλή ποσότητα δεδομένων (Zhang et al., 2024). Αυτή η ιεραρχία απόδοσης καθιστά τη μηχανική μάθηση, και ιδιαίτερα τη βαθιά μάθηση, σημαντικά βελτιωμένη εναλλακτική έναντι των παραδοσιακών προσεγγίσεων για την πρόβλεψη κατανάλωσης καυσίμου και ισχύος άξονα στη ναυτιλία.



#### 2.4.2 ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΔΙΑΔΡΟΜΩΝ ΚΑΙ ΛΕΙΤΟΥΡΓΙΩΝ

Η μηχανική μάθηση χρησιμοποιείται όλο και περισσότερο στη ναυτιλία για τη βελτιστοποίηση διαδρομών, τη βελτίωση της λειτουργικής απόδοσης και την υποστήριξη στρατηγικών όπως η βελτιστοποίηση ταχύτητας, η δρομολόγηση βάσει καιρικών συνθηκών και οι στρατηγικές αργής πλεύσης. Οι πιο αποτελεσματικές εφαρμογές χρησιμοποιούν προηγμένα προγνωστικά μοντέλα και αναλυτικές μεθόδους μεγάλων δεδομένων για τη μείωση της κατανάλωσης καυσίμου, την ελαχιστοποίηση των κινδύνων και την ενίσχυση της λήψης αποφάσεων (Li et al., 2023; Zhang et al., 2023; Zhang et al., 2024).

Στον τομέα της βελτιστοποίησης διαδρομών και τροχιάς, τα μοντέλα μηχανικής μάθησης και βαθιάς μάθησης, όπως το Kalman Filter, Random Forest, LSTM και Transformer, προβλέπουν τις τροχιές πλοίων χρησιμοποιώντας δεδομένα AIS και περιβαλλοντικά δεδομένα, βοηθώντας στην αποφυγή συγκρούσεων και τη βελτιστοποίηση διαδρομών τόσο για επανδρωμένα όσο και για μη επανδρωμένα πλοία (H. Li et al., 2023; M. Zhang et al., 2023). Παράλληλα, μέθοδοι συσταδοποίησης και παλινδρόμησης, όπως K-means, DB-SCAN και Gaussian Process Regression, αναλύουν την κίνηση των πλοίων και τους περιβαλλοντικούς παράγοντες για να προβλέψουν αποκλίσεις, παρασύρσεις και κινδύνους προσάραξης, υποστηρίζοντας ασφαλέστερη και πιο αποτελεσματική δρομολόγηση (Zhang et al., 2023).

Η βελτιστοποίηση ταχύτητας και οι στρατηγικές αργής πλεύσης αποτελούν κρίσιμους τομείς εφαρμογής της μηχανικής μάθησης. Τα μοντέλα βαθιάς μάθησης, ιδιαίτερα τα Bi-LSTM με μηχανισμούς προσοχής, χρησιμοποιούν δεδομένα αισθητήρων και καιρικά δεδομένα σε πραγματικό χρόνο για να προβλέψουν την κατανάλωση καυσίμου βάσει ταχύτητας, κατεύθυνσης και θαλάσσιων συνθηκών. Αυτά τα μοντέλα επιτρέπουν δυναμικές προσαρμογές ταχύτητας, συμπεριλαμβανομένης της αργής πλεύσης, για την ελαχιστοποίηση της χρήσης καυσίμου και των εκπομπών διατηρώντας παράλληλα τους λειτουργικούς στόχους (Zhang et al., 2024). Επιπλέον, τα συστήματα υποστήριξης αποφάσεων που βασίζονται στη μηχανική μάθηση ενσωματώνουν δεδομένα ταχύτητας, καιρού και λειτουργίας για να συστήσουν βέλτιστες ταχύτητες και διαδρομές για αποδοτικότητα και βιωσιμότητα (Zhang et al., 2024).

Η δρομολόγηση βάσει καιρικών συνθηκών αποτελεί άλλη σημαντική εφαρμογή όπου τα μοντέλα μηχανικής μάθησης ενσωματώνουν υδρομετεωρολογικά δεδομένα για την προσαρμογή διαδρομών και ταχυτήτων ως απάντηση στις μεταβαλλόμενες καιρικές συνθήκες, μειώνοντας τους

κινδύνους και βελτιώνοντας την αποδοτικότητα (Zhang et al., 2023; Zhang et al., 2024). Αυτή η ικανότητα είναι ιδιαίτερα σημαντική για τη ναυτιλία, όπου οι καιρικές συνθήκες μπορούν να επηρεάσουν δραματικά τόσο την ασφάλεια όσο και την αποδοτικότητα των ταξιδιών.

Η συγκριτική ανάλυση των διαφόρων εφαρμογών αποκαλύπτει τα ιδιαίτερα πλεονεκτήματα κάθε προσέγγισης. Η πρόβλεψη τροχιάς χρησιμοποιεί Kalman Filter, LSTM, Transformer και GPR για την αποφυγή συγκρούσεων και τη βελτιστοποίηση διαδρομής (Li et al., 2023; Zhang et al., 2023). Η βελτιστοποίηση ταχύτητας/καυσίμου εφαρμόζει Bi-LSTM με προσοχή και Decision Trees για χαμηλότερη χρήση καυσίμου και υποστήριξη αργής πλεύσης (Zhang et al., 2024). Η δρομολόγηση βάσει καιρού χρησιμοποιεί συσταδοποίηση, παλινδρόμηση και βαθιά μάθηση για ασφαλέστερη και πιο αποτελεσματική δρομολόγηση σε μεταβλητές καιρικές συνθήκες (Zhang et al., 2023; Zhang et al., 2024).

Η μηχανική μάθηση μετασχηματίζει τη ναυτιλία επιτρέποντας εξυπνότερο σχεδιασμό διαδρομών, διαχείριση ταχύτητας και λειτουργική αποδοτικότητα. Τα προγνωστικά μοντέλα και οι αναλυτικές μέθοδοι μεγάλων δεδομένων υποστηρίζουν ασφαλέστερες, πιο φιλικές προς το περιβάλλον και πιο οικονομικά αποδοτικές ναυτιλιακές λειτουργίες, προσφέροντας σημαντικές βελτιώσεις σε σχέση με τις παραδοσιακές μεθόδους βελτιστοποίησης που βασίζονται σε στατικούς αλγορίθμους και περιορισμένη ενσωμάτωση δεδομένων πραγματικού χρόνου.

#### 2.4.3 ΠΡΟΛΗΨΗ ΒΛΑΒΩΝ ΚΑΙ ΠΡΟΓΝΩΣΤΙΚΗ ΣΥΝΤΗΡΗΣΗ

Η μηχανική μάθηση μετασχηματίζει την προγνωστική συντήρηση και τον εντοπισμό βλαβών στις ναυτιλιακές λειτουργίες επιτρέποντας τον έγκαιρο εντοπισμό προβλημάτων εξοπλισμού, τη βελτιστοποίηση των προγραμμάτων συντήρησης και τη μείωση του λειτουργικού χρόνου διακοπής. Η ενσωμάτωση προηγμένων μοντέλων μηχανικής μάθησης με την ανάλυση δεδομένων αισθητήρων οδηγεί σε πιο ακριβείς προβλέψεις αστοχιών εξοπλισμού και εξυπνότερο προγραμματισμό συντήρησης.

Στον τομέα της ανάλυσης δεδομένων αισθητήρων και του εντοπισμού βλαβών, τα μοντέλα μηχανικής μάθησης, ιδιαίτερα η βαθιά μάθηση όπως τα υβρίδια CNN-LSTM, αναλύουν δεδομένα χρονοσειρών από αισθητήρες, όπως δονήσεις, θερμοκρασία και πίεση, για να εντοπίσουν ανωμαλίες και να προβλέψουν βλάβες σε πραγματικό χρόνο, βελτιώνοντας την αξιοπιστία κρίσιμου ναυτιλιακού εξοπλισμού (Borré et al., 2023; Surucu et al., 2023; H. Wang et al., 2023).

Οι μέθοδοι συνδυαστικής μάθησης (ensemble learning), όπως bagging, boosting και stacking, συνδυάζουν πολλαπλά μοντέλα για την ενίσχυση της ακρίβειας εντοπισμού βλαβών και την αντιμετώπιση μη ισορροπημένων ή θορυβωδών δεδομένων αισθητήρων, που είναι συνηθισμένα στα ναυτιλιακά περιβάλλοντα (Mian et al., 2024; Surucu et al., 2023).

Η πρόβλεψη αστοχιών εξοπλισμού αποτελεί κεντρικό στοιχείο της προγνωστικής συντήρησης, όπου τα προγνωστικά μοντέλα χρησιμοποιούν ιστορικά και δεδομένα αισθητήρων πραγματικού χρόνου για να προβλέψουν πιθανές αστοχίες εξοπλισμού, επιτρέποντας προληπτικές παρεμβάσεις πριν συμβούν βλάβες. Η τεχνολογία ψηφιακών διδύμων, που υποστηρίζεται από μηχανική μάθηση, δημιουργεί εικονικά αντίγραφα ναυτιλιακών περιουσιακών στοιχείων, επιτρέποντας τη συνεχή παρακολούθηση και προσομοίωση της υγείας του εξοπλισμού για πιο ακριβείς προβλέψεις σφαλμάτων (C. Chen et al., 2023; H. Wang et al., 2023). Αυτή η προσέγγιση παρέχει τη δυνατότητα εικονικής δοκιμής σεναρίων και πρώιμης προειδοποίησης, επιτρέποντας στους χειριστές να προετοιμαστούν για πιθανά προβλήματα πριν αυτά επηρεάσουν τις λειτουργίες.

Η βελτιστοποίηση προγραμμάτων συντήρησης αποτελεί έναν άλλο κρίσιμο τομέα όπου η μηχανική μάθηση προσφέρει σημαντικά οφέλη. Η προγνωστική συντήρηση που βασίζεται στη μηχανική μάθηση μετατοπίζει τη συντήρηση από αντιδραστική σε προληπτική, βελτιστοποιώντας τα χρονοδιαγράμματα για την ελαχιστοποίηση του χρόνου διακοπής και του κόστους ενώ μεγιστοποιεί τη διάρκεια ζωής του εξοπλισμού (Chen et al., 2023; Borré et al., 2023; Surucu et al., 2023). Η quantile regression και η μοντελοποίηση αβεβαιότητας σε νευρωνικά δίκτυα βοηθούν στη βελτίωση του χρονισμού συντήρησης λαμβάνοντας υπόψη τις αβεβαιότητες πρόβλεψης (Borré et al., 2023), επιτρέποντας πιο εξελιγμένο προγραμματισμό που ισορροπεί τους κινδύνους με το κόστος.

Η συγκριτική ανάλυση των διαφόρων εφαρμογών αποκαλύπτει τα ειδικά πλεονεκτήματα κάθε προσέγγισης. Η ανάλυση δεδομένων αισθητήρων χρησιμοποιεί βαθιά μάθηση (CNN-LSTM, attention) για εντοπισμό ανωμαλιών σε πραγματικό χρόνο και βελτιωμένη ακρίβεια (Borré et al., 2023; Surucu et al., 2023; Wang et al., 2023). Ο εντοπισμός βλαβών εφαρμόζει συνδυαστικής μάθησης και ημι-επιβλεπόμενα μοντέλα για ανθεκτικότητα στο θόρυβο και καλύτερη γενίκευση. Η πρόβλεψη αστοχιών χρησιμοποιεί ψηφιακούς διδύμους και προσομοίωση με μηχανική μάθηση για πρώιμη προειδοποίηση και εικονική δοκιμή σεναρίων (Chen et al., 2023; Wang et al., 2023). Η βελτιστοποίηση συντήρησης εφαρμόζει προγνωστικά μοντέλα και quantile regression για

μειωμένο χρόνο διακοπής και εξοικονόμηση κόστους (Chen et al., 2023; Borré et al., 2023; Surucu et al., 2023).

Η μηχανική μάθηση επιτρέπει μια στροφή προς την προληπτική, βασισμένη σε δεδομένα συντήρηση στις ναυτιλιακές λειτουργίες. Αξιοποιώντας δεδομένα αισθητήρων, προηγμένα μοντέλα και ψηφιακούς διδύμους, οι ναυτιλιακοί χειριστές μπορούν να εντοπίζουν βλάβες νωρίτερα, να προβλέπουν αστοχίες με μεγαλύτερη ακρίβεια και να βελτιστοποιούν τα προγράμματα συντήρησης για μεγαλύτερη αποδοτικότητα και ασφάλεια, μειώνοντας σημαντικά τα λειτουργικά κόστη και βελτιώνοντας την αξιοπιστία των ναυτιλιακών λειτουργιών.

#### 2.4.4 ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ

Η μηχανική μάθηση χρησιμοποιείται όλο και περισσότερο στη ναυτιλία για την αξιολόγηση της περιβαλλοντικής απόδοσης, με εφαρμογές που εκτείνονται από την πρόβλεψη εκπομπών, την κατηγοριοποίηση πλοίων έως τα συστήματα παρακολούθησης πραγματικού χρόνου. Η βαθιά μάθηση και οι αναλυτικές μέθοδοι μεγάλων δεδομένων επιτρέπουν πιο ακριβείς, σε πραγματικό χρόνο και λειτουργικά σχετικές αξιολογήσεις των εκπομπών και της απόδοσης των πλοίων (Zhang et al., 2024; Zhang et al., 2023).

Στον τομέα των μοντέλων πρόβλεψης εκπομπών, προηγμένα μοντέλα βαθιάς μάθησης, όπως τα Bi-LSTM δίκτυα με μηχανισμούς προσοχής, χρησιμοποιούνται για την πρόβλεψη της κατανάλωσης καυσίμου των πλοίων και των σχετικών εκπομπών. Αυτά τα μοντέλα ενσωματώνουν ένα ευρύ φάσμα λειτουργικών και περιβαλλοντικών μεταβλητών, όπως ταχύτητα, καιρικές συνθήκες και θαλάσσιες συνθήκες, για να παρέχουν ακριβείς προβλέψεις πραγματικού χρόνου, υποστηρίζοντας τη λήψη αποφάσεων για βιώσιμες λειτουργίες (Zhang et al., 2024). Επιπλέον, τα μοντέλα Decision Tree βοηθούν στον εντοπισμό των μεταβλητών που επηρεάζουν περισσότερο την κατανάλωση καυσίμου και τις εκπομπές, βελτιώνοντας την ερμηνευσιμότητα του μοντέλου και τις λειτουργικές γνώσεις (Zhang et al., 2024). Αυτή η προσέγγιση επιτρέπει στους χειριστές να κατανοήσουν ποιοι παράγοντες έχουν τη μεγαλύτερη επίδραση στην περιβαλλοντική απόδοση και να εστιάσουν τις προσπάθειές τους σε αυτούς.

Η κατηγοριοποίηση πλοίων και η ανάλυση τροχιάς αποτελούν άλλη σημαντική εφαρμογή της μηχανικής μάθησης στην περιβαλλοντική αξιολόγηση. Μέθοδοι συσταδοποίησης μηχανικής μάθησης, όπως K-means και DB-SCAN, και η PCA χρησιμοποιούνται για την ομαδοποίηση

πλοίων και περιβαλλοντικών παραγόντων, επιτρέποντας την κατηγοριοποίηση βάσει λειτουργικών προφίλ και παραγόντων κινδύνου (Zhang et al., 2023). Η GPR και το Dynamic Time Warping (DTW) εφαρμόζονται για την πρόβλεψη και οπτικοποίηση των τροχιών κίνησης πλοίων, οι οποίες μπορούν να συνδεθούν με την περιβαλλοντική απόδοση και τον κίνδυνο προσάραξης (Zhang et al., 2023). Αυτές οι τεχνικές επιτρέπουν την ανάπτυξη πιο εξειδικευμένων στρατηγικών περιβαλλοντικής διαχείρισης για διαφορετικές κατηγορίες πλοίων.

Τα συστήματα παρακολούθησης πραγματικού χρόνου αντιπροσωπεύουν μια από τις πιο πρακτικά χρήσιμες εφαρμογές της μηχανικής μάθησης στην περιβαλλοντική αξιολόγηση. Δεδομένα αισθητήρων υψηλής συχνότητας, αναφορές ταξιδιών και υδρομετεωρολογικά δεδομένα ενσωματώνονται σε μοντέλα μηχανικής μάθησης για την παρακολούθηση σε πραγματικό χρόνο των ναυτιλιακών λειτουργιών και των περιβαλλοντικών επιπτώσεων (Zhang et al., 2024; Zhang et al., 2023). Αυτά τα συστήματα μπορούν να παρέχουν συνεχή ανατροφοδότηση και υποστήριξη για λειτουργικές προσαρμογές για την ελαχιστοποίηση των εκπομπών και τη βελτίωση της ασφάλειας (Zhang et al., 2024; Zhang et al., 2023). Η ικανότητα για συνεχή παρακολούθηση και άμεση ανταπόκριση καθιστά αυτές τις εφαρμογές ιδιαίτερα πολύτιμες για την επίτευξη βραχυπρόθεσμων και μακροπρόθεσμων περιβαλλοντικών στόχων.

Η συνολική αξιολόγηση των εφαρμογών της μηχανικής μάθησης στην περιβαλλοντική αξιολόγηση ναυτιλίας αποκαλύπτει τα ειδικά πλεονεκτήματα κάθε προσέγγισης. Η πρόβλεψη εκπομπών χρησιμοποιεί Bi-LSTM και Decision Trees για ακριβείς εκτιμήσεις καυσίμου/εκπομπών σε πραγματικό χρόνο (Zhang et al., 2024). Η κατηγοριοποίηση πλοίων εφαρμόζει K-means, DB-SCAN και PCA για ομαδοποίηση βάσει λειτουργικών/περιβαλλοντικών προφίλ (Zhang et al., 2023). Η παρακολούθηση πραγματικού χρόνου ενσωματώνει δεδομένα αισθητήρων, GPR και DTW για συνεχή αξιολόγηση και λειτουργική ανατροφοδότηση (Zhang et al., 2024; Zhang et al., 2023).

Η μηχανική μάθηση μετασχηματίζει την περιβαλλοντική αξιολόγηση απόδοσης στη ναυτιλία επιτρέποντας ακριβή πρόβλεψη εκπομπών, αποτελεσματική κατηγοριοποίηση πλοίων και παρακολούθηση πραγματικού χρόνου. Αυτές οι εξελίξεις υποστηρίζουν πιο βιώσιμες και ασφαλέστερες ναυτιλιακές λειτουργίες, προσφέροντας τη δυνατότητα για πιο στοχευμένες και αποτελεσματικές στρατηγικές περιβαλλοντικής διαχείρισης που μπορούν να προσαρμοστούν δυναμικά στις μεταβαλλόμενες συνθήκες και απαιτήσεις.

## 2.5 ΚΕΝΑ ΣΤΗ ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΑΙ ΕΡΕΥΝΗΤΙΚΕΣ ΕΥΚΑΙΡΙΕΣ

Η συστηματική ανάλυση της διεθνούς βιβλιογραφίας αποκαλύπτει σημαντικά κενά και ερευνητικές ευκαιρίες στον τομέα της περιβαλλοντικής αξιολόγησης των θαλάσσιων μεταφορών. Παρά τις σημαντικές προόδους που έχουν επιτευχθεί στην ανάπτυξη ρυθμιστικών μέτρων, τεχνολογιών παρακολούθησης και εφαρμογών μηχανικής μάθησης, εξακολουθούν να υπάρχουν σημαντικά προβλήματα που περιορίζουν την αποτελεσματικότητα των τρεχόντων προσεγγίσεων και την ανάπτυξη πιο ολοκληρωμένων λύσεων.

Ένα από τα σημαντικότερα κενά αφορά την ανάγκη για ολοκληρωμένους δείκτες περιβαλλοντικής απόδοσης. Οι υπάρχουσες μέθοδοι αξιολόγησης, όπως οι δείκτες του ΙΜΟ (EEDI, EEXI, CII, SEEMP), εστιάζουν κυρίως σε μεμονωμένες πτυχές της περιβαλλοντικής απόδοσης και δεν παρέχουν μια ολιστική εικόνα που να συνδυάζει πολλαπλές διαστάσεις όπως εκπομπές διαφόρων ρύπων, ενεργειακή απόδοση, λειτουργικά χαρακτηριστικά και περιβαλλοντικές συνθήκες. Επιπλέον, η έλλειψη κατανοητών και πρακτικά χρήσιμων δεικτών που να μπορούν να χρησιμοποιηθούν από μη ειδικούς περιορίζει τη διαφάνεια και την ευρεία υιοθέτηση βέλτιστων πρακτικών. Η απουσία ενιαίων μεθοδολογιών αξιολόγησης και η δυσκολία σύγκρισης διαφορετικών τύπων πλοίων δημιουργούν επιπλέον εμπόδια στην ανάπτυξη συνεκτικών πολιτικών περιβαλλοντικής διαχείρισης.

Η περιορισμένη χρήση τεχνικών μη επιβλεπόμενης μάθησης αποτελεί άλλη σημαντική ερευνητική ευκαιρία. Παρά την ευρεία εφαρμογή της μηχανικής μάθησης στη ναυτιλία, οι εφαρμογές της PCA στην περιβαλλοντική ανάλυση ναυτιλίας είναι ελάχιστες, ενώ η χρήση της cluster analysis για την κατηγοριοποίηση πλοίων βάσει περιβαλλοντικών χαρακτηριστικών παραμένει περιορισμένη. Η εξερευνητική ανάλυση δεδομένων και η ανακάλυψη κρυμμένων προτύπων στα περιβαλλοντικά δεδομένα ναυτιλίας μπορεί να αποκαλύψει νέες γνώσεις και σχέσεις που δεν είναι εμφανείς με τις παραδοσιακές μεθόδους ανάλυσης. Αυτές οι τεχνικές έχουν τη δυνατότητα να συμβάλουν στην ανάπτυξη πιο εξειδικευμένων και αποτελεσματικών στρατηγικών περιβαλλοντικής διαχείρισης για διαφορετικές κατηγορίες πλοίων και λειτουργικών συνθηκών.

Η έλλειψη γραφικών εργαλείων αξιολόγησης αποτελεί επίσης σημαντικό περιορισμό στην τρέχουσα πρακτική. Η απουσία απλών οπτικών εργαλείων που να μπορούν να χρησιμοποιηθούν από μη ειδικούς περιορίζει τη διαφάνεια και την προσβασιμότητα των περιβαλλοντικών αξιολογήσεων. Επιπλέον, η ανάγκη για εργαλεία υποστήριξης αποφάσεων που να μπορούν να

ενσωματώσουν πολύπλοκες πληροφορίες σε κατανοητές μορφές είναι εμφανής. Η περιορισμένη διαφάνεια στις μεθόδους αξιολόγησης και οι προκλήσεις στην ερμηνευσιμότητα πολύπλοκων μοντέλων δημιουργούν εμπόδια στην ευρεία υιοθέτηση και εφαρμογή νέων τεχνολογιών από τη ναυτιλιακή βιομηχανία.

Οι προκλήσεις στην ενσωμάτωση δεδομένων από διαφορετικές πηγές αποτελούν ένα άλλο κρίσιμο κενό που χρήζει αντιμετώπισης. Οι δυσκολίες στον συνδυασμό δεδομένων από διαφορετικές πηγές, όπως δεδομένα AIS, μετεωρολογικά δεδομένα, δεδομένα αισθητήρων και λειτουργικά αρχεία, περιορίζουν τη δυνατότητα ανάπτυξης ολοκληρωμένων συστημάτων αξιολόγησης. Η ανάγκη για ανθεκτικές μεθόδους διαχείρισης ελλιπών δεδομένων και οι δυσκολίες στην εναρμόνιση διαφορετικών μορφών τους απαιτούν την ανάπτυξη νέων μεθόδων και προτύπων. Επιπλέον, η έλλειψη προτύπων ποιότητας και διαλειτουργικότητας δεδομένων στη ναυτιλιακή βιομηχανία δημιουργεί εμπόδια στην αποτελεσματική αξιοποίηση των διαθέσιμων πληροφοριών.

Τα μεθοδολογικά κενά στις μελέτες LCA αποτελούν επίσης σημαντικό τομέα για μελλοντική έρευνα. Η έλλειψη τυποποιημένων μεθοδολογιών, οι ασυνέπειες στον καθορισμό ορίων συστήματος και λειτουργικών μονάδων, καθώς και οι αβεβαιότητες σε λειτουργικές και περιβαλλοντικές συνθήκες περιορίζουν την αξιοπιστία και συγκρισιμότητα των αποτελεσμάτων. Αυτά τα προβλήματα επιδεινώνονται από την απουσία συναίνεσης σε μεθόδους απογραφής και την ανάγκη για πιο ολοκληρωμένες προσεγγίσεις που να λαμβάνουν υπόψη τη μεταβλητότητα των ναυτιλιακών λειτουργιών.

Παρά τους περιορισμούς, αυτά τα κενά δημιουργούν σημαντικές ερευνητικές ευκαιρίες. Η ανάπτυξη νέων μεθοδολογιών που συνδυάζουν τεχνικές μη επιβλεπόμενης μάθησης για την περιβαλλοντική ανάλυση ναυτιλίας μπορεί να οδηγήσει σε πρωτοποριακές γνώσεις και εργαλεία. Η δημιουργία ολοκληρωμένων δεικτών που να συνδυάζουν πολλαπλές διαστάσεις περιβαλλοντικής απόδοσης σε κατανοητές και πρακτικά χρήσιμες μορφές αποτελεί επίσης σημαντική ευκαιρία. Επιπλέον, η ανάπτυξη γραφικών συστημάτων αξιολόγησης και εργαλείων υποστήριξης αποφάσεων που να ενσωματώνουν προηγμένες τεχνολογίες σε προσβάσιμες διεπαφές μπορεί να επιταχύνει την υιοθέτηση βέλτιστων πρακτικών από τη ναυτιλιακή βιομηχανία.

Συνολικά, τα εντοπισμένα κενά υπογραμμίζουν την ανάγκη για διεπιστημονικές προσεγγίσεις που να συνδυάζουν τεχνικές γνώσεις, μηχανική μάθηση, περιβαλλοντική επιστήμη

και σχεδιασμό διεπαφών χρήστη. Η αντιμετώπιση αυτών των προκλήσεων απαιτεί συνεργασία μεταξύ ακαδημαϊκών ιδρυμάτων, βιομηχανίας και ρυθμιστικών φορέων για την ανάπτυξη λύσεων που να είναι ταυτόχρονα επιστημονικά ακριβείς, πρακτικά εφαρμόσιμες και κατανοητές από τους τελικούς χρήστες.



### 3. ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Το παρόν κεφάλαιο παρουσιάζει τη μεθοδολογική προσέγγιση που υιοθετήθηκε για την ανάπτυξη νέων δεικτών περιβαλλοντικής απόδοσης των θαλάσσιων μέσων μεταφοράς μέσω της εφαρμογής τεχνικών μηχανικής μάθησης. Η μεθοδολογία βασίζεται στον συνδυασμό της PCA και της ανάλυσης συστάδας για τη μείωση της διαστατικότητας των δεδομένων και την κατηγοριοποίηση των πλοίων αντίστοιχα. Η προσέγγιση ακολουθεί μια συστηματική διαδικασία που ξεκινά από τη συλλογή από τον μηχανισμό MRV της Ευρωπαϊκής Ένωσης και καταλήγει στη δημιουργία ενός γραφικού συστήματος ετικετοποίησης που αντικατοπτρίζει την περιβαλλοντική απόδοση των πλοίων. Το κεφάλαιο περιγράφει λεπτομερώς τις πηγές δεδομένων, τις μεθόδους προεπεξεργασίας, τις τεχνικές ανάλυσης που εφαρμόστηκαν, καθώς και τη διαδικασία επικύρωσης των αποτελεσμάτων. Η μεθοδολογία έχει σχεδιαστεί ώστε να αντιμετωπίσει τα κύρια κενά που εντοπίστηκαν στη βιβλιογραφική ανασκόπηση, εστιάζοντας στην ανάπτυξη κατανοητών, πρακτικά χρήσιμων και επιστημονικά τεκμηριωμένων εργαλείων αξιολόγησης για τη ναυτιλιακή βιομηχανία.

#### 3.1 ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA ANALYSIS)

Η Ανάλυση των Κύριων Συνιστωσών αποτελεί μια τεχνική στατιστικής ανάλυσης, ευρέως χρησιμοποιούμενη στον τομέα της Μηχανικής Μάθησης, η οποία χρησιμοποιείται για την λήψη σημαντικών πληροφοριών από μεγάλα σύνολα δεδομένων. Πιο συγκεκριμένα, κύριο στόχο της PCA analysis αποτελεί η μείωση της διάστασης των δεδομένων που έχουν συλλεχθεί, διατηρώντας όσο το δυνατόν περισσότερες από τις αρχικές πληροφορίες. Από μεγάλα σύνολα δεδομένων  $p$  διαστάσεων λαμβάνονται αλληλοσυσχετιζόμενες μεταβλητές και μέσω της μεθόδου αυτής προκύπτουν νέα, μικρότερα σύνολα  $k$  διαστάσεων, όπου  $k < p$ , μη συσχετιζόμενων μεταβλητών οι οποίες ονομάζονται κύριες συνιστώσες. Οι νέες μεταβλητές που δημιουργούνται (Κύριες Συνιστώσες) απεικονίζουν την μεγαλύτερη διακύμανση των αρχικών δεδομένων. Μάλιστα, η πρώτη κύρια συνιστώσα που προκύπτει από τη μέθοδο συγκεντρώνει το μεγαλύτερο μέρος της μεταβλητότητας των δεδομένων. Η δεύτερη κύρια συνιστώσα αιχμαλωτίζει τη μέγιστη διακύμανση από το υπόλοιπο της πληροφορίας που δεν εξηγήθηκε από την πρώτη κύρια

συνιστώσα, ενώ κάθε επόμενη συνιστώσα ακολουθεί την ίδια αρχή, μεγιστοποιώντας τη διακύμανση του εναπομένοντος τμήματος των δεδομένων (Karamizadeh et al., 2013).

Με τον τρόπο αυτό, και καθώς προχωρά η ανάλυση, κάθε νέα κύρια συνιστώσα η οποία προκύπτει εξηγεί όλο και μικρότερα ποσά διακύμανσης, γεγονός το οποίο επιτρέπει να διατηρηθούν και να ερμηνευθούν μόνο οι πρώτες κύριες συνιστώσες, καθιστώντας την ανάλυση πιο απλή, ενώ ταυτόχρονα διατηρείται το μεγαλύτερο μέρος της αρχικής πληροφορίας. Βασικό χαρακτηριστικό των κύριων συνιστωσών είναι το ότι δεν παρουσιάζουν καμιά σχέση εξάρτησης μεταξύ τους και η συσχέτιση τους είναι μηδενική (Smith, 2002).

### 3.1.1 ΜΑΘΗΜΑΤΙΚΟ ΥΠΟΒΑΘΡΟ ΤΗΣ PCA

Η Ανάλυση Κύριων Συνιστωσών βασίζεται σε θεμελιώδεις στατιστικές έννοιες που αφορούν τη μέτρηση της κεντρικής τάσης και της διασποράς των δεδομένων. Για ένα σύνολο δεδομένων  $X = [X_1, X_2, X_3, \dots, X_p]$ , η μέση τιμή κάθε μεταβλητής αποτελεί το σημείο αναφοράς για την αξιολόγηση της μεταβλητότητας των παρατηρήσεων.

Η τυπική απόκλιση αποτελεί βασικό μέτρο διασποράς των δεδομένων και προσδιορίζει την μέση απόσταση κάθε παρατήρησης από την μέση τιμή (Εξ. 3.1). Η διακύμανση, ως το τετράγωνο της τυπικής απόκλισης, εκφράζει την μεταβλητότητα των δεδομένων και αποτελεί κεντρική έννοια στην PCA (Εξ. 3.2).

$$\bar{X}_i = \frac{\sum_{j=1}^n X_{ij}}{n} \quad (3.1)$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (3.2)$$

Η τυποποίηση των μεταβλητών αποτελεί προαπαιτούμενο για την εφαρμογή της PCA, καθώς εξασφαλίζει ότι όλες οι μεταβλητές έχουν ίσο βάρος στην ανάλυση, ανεξάρτητα από τις μονάδες μέτρησης και την κλίμακά τους.

Για την ανάλυση δεδομένων από διαφορετικές διαστάσεις (X,Y) απαιτείται η εξέταση των σχέσεων μεταξύ των μεταβλητών. Η συνδιακύμανση αποτελεί το βασικό εργαλείο για τη μέτρηση της γραμμικής σχέσης μεταξύ δύο μεταβλητών (Εξ. 3.3). Θετική συνδιακύμανση υποδεικνύει ότι οι μεταβλητές τείνουν να αυξάνονται ή να μειώνονται ταυτόχρονα, ενώ αρνητική συνδιακύμανση δηλώνει αντίστροφη συμπεριφορά.

$$\text{Cov}(X, Y) = \frac{1}{n-1} \times \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (3.3)$$

Ωστόσο, η συνδιακύμανση εξαρτάται από τις μονάδες μέτρησης των μεταβλητών, γεγονός που δημιουργεί προβλήματα στη σύγκριση μεταβλητών με διαφορετικές κλίμακες. Για την επίλυση αυτού του προβλήματος χρησιμοποιείται ο συντελεστής συσχέτισης, ο οποίος κανονικοποιεί τη συνδιακύμανση διαιρώντας την με το γινόμενο των τυπικών αποκλίσεων των δύο μεταβλητών (Εξ. 3.4).

$$r_{x,y} = \frac{\text{Cov}(X,Y)}{S_x \times S_y} \quad (3.4)$$

Ο συντελεστής συσχέτισης λαμβάνει τιμές στο εύρος  $[-1,1]$  και παρέχει μια κανονικοποιημένη μέτρηση της γραμμικής σχέσης, ανεξάρτητη από τις μονάδες μέτρησης και την κλίμακα των μεταβλητών.

Για την ολοκληρωμένη ανάλυση ενός συνόλου πολυδιάστατων δεδομένων απαιτείται η συγκρότηση του πίνακα συνδιακύμανσης. Πρόκειται για έναν συμμετρικό τετραγωνικό πίνακα διαστάσεων  $n \times n$ , όπου  $n$  ο αριθμός των μεταβλητών. Κάθε στοιχείο του πίνακα εκφράζει τη συνδιακύμανση μεταξύ δύο μεταβλητών.

Τα διαγώνια στοιχεία του πίνακα αντιστοιχούν στις διακυμάνσεις των επιμέρους μεταβλητών, ενώ τα εκτός διαγωνίου στοιχεία στις συνδιακυμάνσεις μεταξύ διαφορετικών μεταβλητών. Η συμμετρία του πίνακα προκύπτει από την ιδιότητα της συνδιακύμανσης ότι  $\text{Cov}(X,Y) = \text{Cov}(Y,X)$  (Rodionova et al., 2021).

Με την τυποποίηση των αρχικών μεταβλητών, ο πίνακας συνδιακύμανσης μετατρέπεται σε πίνακα συσχετίσεων, όπου τα διαγώνια στοιχεία γίνονται ίσα με τη μονάδα και τα εκτός διαγωνίου στοιχεία αντιστοιχούν στους συντελεστές συσχέτισης.

Με μαθηματικούς όρους, για  $p$  αρχικές μεταβλητές  $X_1, \dots, X_p$  δημιουργούνται  $p$  κύριες συνιστώσες  $Y_1, \dots, Y_p$  που ορίζονται ως γραμμικοί συνδυασμοί των αρχικών μεταβλητών. Η τεχνική των κύριων συνιστωσών έχει ως στόχο τη δημιουργία  $p$  νέων ασυσχέτιστων μεταξύ τους μεταβλητών, οι οποίες αναφέρθηκαν προηγουμένως ως κύριες συνιστώσες (Εξ. 3.5a-p).

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (1) \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (2) \\ &\vdots \end{aligned} \quad (3.5)$$

$$Y_p = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \quad (p)$$

Υπό μορφή πινάκων, ο μετασχηματισμός εκφράζεται συμπαγώς ως  $Y = AX$ , όπου  $Y$  είναι το διάνυσμα των κύριων συνιστωσών,  $X$  είναι το διάνυσμα των αρχικών μεταβλητών και  $A$  είναι ο πίνακας μετασχηματισμού διαστάσεων  $p \times p$ .

Οι συντελεστές στάθμισης  $a_{ij}$  ορίζουν τη συνεισφορά κάθε αρχικής μεταβλητής στη δημιουργία κάθε κύριας συνιστώσας. Η εκτίμηση αυτών των συντελεστών αποτελεί τον πυρήνα της μεθοδολογίας της PCA και βασίζεται στον υπολογισμό των ιδιοδιανυσμάτων του πίνακα συνδιακύμανσης.

Η εξαγωγή των κύριων συνιστωσών βασίζεται στις έννοιες των ιδιοτιμών και ιδιοδιανυσμάτων του πίνακα συνδιακύμανσης. Για έναν τετραγωνικό πίνακα  $\Sigma$ , οι ιδιοτιμές  $\lambda$  και τα ιδιοδιανύσματα  $a$  αποτελούν λύσεις της θεμελιώδους εξίσωσης της γραμμικής άλγεβρας (Εξ. 3.6). Η εξίσωση αυτή εκφράζει το γεγονός ότι όταν ο πίνακας  $\Sigma$  εφαρμοστεί σε ένα ιδιοδιάνυσμα  $a$ , το αποτέλεσμα είναι το ίδιο διάνυσμα πολλαπλασιασμένο με μια σταθερά  $\lambda$ .

$$\Sigma a = \lambda a \quad (3.6)$$

Τα ιδιοδιανύσματα που προκύπτουν από τον πίνακα συνδιακύμανσης παρουσιάζουν τη σημαντική ιδιότητα ότι είναι κάθετα μεταξύ τους, γεγονός που εξασφαλίζει ότι οι κύριες συνιστώσες θα είναι ασυσχέτιστες. Η ισότητα των διανυσμάτων  $a_i$  προς τη μονάδα επιτυγχάνεται μέσω κανονικοποίησης, ενώ η απαίτηση ότι το άθροισμα των τετραγώνων των συντελεστών να ισούται με τη μονάδα επιτρέπει τη μέγιστη δυνατή διακύμανση για κάθε συνιστώσα.

Στο πλαίσιο της Ανάλυσης Κύριων Συνιστωσών, τα ιδιοδιανύσματα του πίνακα συνδιακύμανσης αποκτούν ιδιαίτερη σημασία καθώς ορίζουν τις κατευθύνσεις μέγιστης διακύμανσης στον χώρο των δεδομένων. Κάθε ιδιοδιάνυσμα αντιστοιχεί σε μία κύρια συνιστώσα και καθορίζει τον τρόπο με τον οποίο οι αρχικές μεταβλητές συνδυάζονται για τη δημιουργία της νέας μεταβλητής.

Οι ιδιοτιμές  $\lambda_i$  εκφράζουν τη διακύμανση που εξηγείται από κάθε κύρια συνιστώσα και αποτελούν το κριτήριο για την κατάταξη των συνιστωσών κατά σειρά σπουδαιότητας. Η ταξινόμηση των ιδιοτιμών σε φθίνουσα σειρά ( $\lambda_1 > \lambda_2 > \dots > \lambda_p$ ) καθορίζει τη σειρά προτεραιότητας των κύριων συνιστωσών (Jolliffe & Jolliffe, 1986).

Η δυνατότητα επιλογής των σημαντικότερων κύριων συνιστωσών, με βάση το μέγεθος των αντίστοιχων ιδιοτιμών, επιτρέπει την επίτευξη της μείωσης διαστατικότητας με ελάχιστη απώλεια πληροφορίας. Συγκεκριμένα, η διατήρηση των  $k$  μεγαλύτερων ιδιοτιμών εξασφαλίζει ότι το μεγαλύτερο μέρος της αρχικής διακύμανσης των δεδομένων θα παραμείνει στο νέο, μειωμένων διαστάσεων σύστημα συντεταγμένων.

### 3.1.2 ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ

Η εφαρμογή της Ανάλυσης Κύριων Συνιστωσών (PCA) ακολουθεί μια συγκεκριμένη μεθοδολογική διαδικασία που εξασφαλίζει την αξιόπιστη εξαγωγή των κύριων συνιστωσών από πολυδιάστατα δεδομένα. Η διαδικασία ξεκινά με την προετοιμασία των δεδομένων, συνεχίζει με υπολογιστικές διαδικασίες και ολοκληρώνεται με την ερμηνεία των αποτελεσμάτων.

Αρχικά, απαιτείται η τυποποίηση (standardization) των δεδομένων, καθώς η PCA είναι ιδιαίτερα ευαίσθητη στις διαφορές κλίμακας μεταξύ των μεταβλητών. Η τυποποίηση εξασφαλίζει ότι όλες οι μεταβλητές συνεισφέρουν εξίσου στην ανάλυση, ανεξάρτητα από τις μονάδες μέτρησης τους. Χωρίς τυποποίηση, μεταβλητές με μεγαλύτερες κλίμακες τιμών θα κυριαρχούν έναντι αυτών με μικρότερες κλίμακες, οδηγώντας σε μεροληπτικά αποτελέσματα. Η διαδικασία περιλαμβάνει την αφαίρεση του μέσου όρου και τη διαίρεση με την τυπική απόκλιση για κάθε μεταβλητή, μετασχηματίζοντας τα δεδομένα ώστε να έχουν μέσο όρο μηδέν και τυπική απόκλιση ίση με τη μονάδα.

Στη συνέχεια, υπολογίζεται ο πίνακας συνδιακύμανσης (covariance matrix) των τυποποιημένων δεδομένων. Αυτός ο πίνακας συνοψίζει τις συνδιακυμάνσεις που σχετίζονται με όλους τους συνδυασμούς ζευγών των αρχικών μεταβλητών στο σύνολο δεδομένων. Ο υπολογισμός του πίνακα συνδιακύμανσης βοηθά στον εντοπισμό των σχέσεων μεταξύ των μεταβλητών, δηλαδή του τρόπου με τον οποίο οι μεταβλητές μεταβάλλονται σε σχέση με τον μέσο όρο. Για τυποποιημένα δεδομένα, ο πίνακας συνδιακύμανσης ταυτίζεται με τον πίνακα συσχετίσεων.

Το επόμενο κρίσιμο στάδιο αφορά την εξαγωγή των ιδιοτιμών (eigenvalues) και ιδιοδιανυσμάτων (eigenvectors) από τον πίνακα συνδιακύμανσης. Τα ιδιοδιανύσματα ορίζουν τις κατευθύνσεις μέγιστης διακύμανσης στα δεδομένα, ενώ οι αντίστοιχες ιδιοτιμές μετρούν τη σημασία αυτών των κατευθύνσεων. Κάθε ιδιοδιάνυσμα αντιστοιχεί σε μία κύρια συνιστώσα και

καθορίζει τον τρόπο με τον οποίο οι αρχικές μεταβλητές συνδυάζονται για τη δημιουργία της νέας μεταβλητής.

Ακολουθεί η κατάταξη των ιδιοτιμών σε φθίνουσα σειρά, καθώς αυτή καθορίζει τη σειρά προτεραιότητας των κύριων συνιστωσών. Η πρώτη κύρια συνιστώσα (PC1) αντιστοιχεί στη μεγαλύτερη ιδιοτιμή και καταγράφει τη μέγιστη διακύμανση, ενώ η δεύτερη κύρια συνιστώσα (PC2) είναι κάθετη προς την πρώτη και καταγράφει τη μέγιστη υπολειπόμενη διακύμανση. Αυτή η διαδικασία συνεχίζεται μέχρι να εξαχθούν όλες οι κύριες συνιστώσες, οι οποίες ισούνται σε αριθμό με τις αρχικές μεταβλητές.

Στη φάση της επιλογής των σημαντικών συνιστωσών, εφαρμόζονται συγκεκριμένα κριτήρια για τον προσδιορισμό του αριθμού των κύριων συνιστωσών που θα διατηρηθούν. Συνήθως διατηρούνται οι συνιστώσες που εξηγούν ένα προκαθορισμένο ποσοστό της συνολικής διακύμανσης, όπως 80 ή 85%. Εναλλακτικά, μπορεί να εφαρμοστεί το κριτήριο Kaiser, όπου διατηρούνται οι συνιστώσες με ιδιοτιμές μεγαλύτερες της μονάδας, ή το κριτήριο scree plot που εντοπίζει την απότομη μείωση των ιδιοτιμών.

Τέλος, πραγματοποιείται ο μετασχηματισμός των αρχικών δεδομένων στον νέο χώρο των κύριων συνιστωσών. Αυτός ο μετασχηματισμός επιτυγχάνεται με τον πολλαπλασιασμό του αρχικού συνόλου δεδομένων με τα επιλεγμένα ιδιοδιανύσματα. Η διαδικασία αυτή αναπροσανατολίζει τα δεδομένα από τους αρχικούς άξονες σε αυτούς που αντιπροσωπεύουν οι κύριες συνιστώσες, παρέχοντας μια νέα οπτική για την καλύτερη αναπαράσταση των δεδομένων (Paul et al., 2013).

Η ολοκληρωμένη εφαρμογή αυτής της μεθοδολογίας διασφαλίζει την αξιόπιστη μείωση της διαστατικότητας των δεδομένων, διατηρώντας παράλληλα τη μέγιστη δυνατή πληροφορία από το αρχικό σύνολο μεταβλητών. Η PCA αποτελεί έναν ευέλικτο εργαλείο που επιτρέπει την ανάλυση συνόλων δεδομένων που μπορεί να περιέχουν πολυσυγγραμμικότητα, ελλειπείς τιμές και θόρυβο.

### 3.1.3 ΠΛΕΟΝΕΚΤΗΜΑΤΑ & ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ PCA

Η Ανάλυση Κύριων Συνιστωσών παρουσιάζει σημαντικά πλεονεκτήματα που την καθιστούν έναν από τους πιο δημοφιλείς αλγορίθμους μείωσης διάστασης, ωστόσο συνοδεύεται και από ορισμένους περιορισμούς που πρέπει να λαμβάνονται υπόψη κατά την εφαρμογή της.

### 3.1.3.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ

Η PCA αποτελεί ένα εξαιρετικά αποτελεσματικό εργαλείο για τη μείωση της διαστατικότητας μεγάλων συνόλων δεδομένων, διατηρώντας παράλληλα την πιο σημαντική πληροφορία, όπως τις κύριες τάσεις και μοτίβα. Αυτή η ικανότητα επιτρέπει την απλοποίηση πολύπλοκων δεδομένων χωρίς σημαντική απώλεια πληροφοριών, καθιστώντας την ιδανική για εφαρμογές που περιλαμβάνουν υψηλών διαστάσεων δεδομένα.

Ένα από τα σημαντικότερα πλεονεκτήματα της PCA είναι η βελτίωση της υπολογιστικής αποδοτικότητας. Οι αλγόριθμοι μηχανικής μάθησης συγκλίνουν, συνήθως, ταχύτερα όταν εκπαιδεύονται σε κύριες συνιστώσες αντί για το αρχικό σύνολο δεδομένων (Jolliffe, 2002). Η επεξεργασία μειωμένων πινάκων επιτρέπει ταχύτερους υπολογισμούς και μικρότερο αποθηκευτικό χώρο, γεγονός που αποτελεί κρίσιμο παράγοντα για εφαρμογές μεγάλης κλίμακας.

Η PCA μειώνει σημαντικά τον χρόνο που απαιτείται για την εκπαίδευση μοντέλων μηχανικής μάθησης και βοηθά στην αποφυγή του υπερπροσαρμογής (overfitting). Η μείωση του αριθμού των χαρακτηριστικών μειώνει την πολυπλοκότητα του μοντέλου, καθώς η προσθήκη κάθε νέου χαρακτηριστικού επηρεάζει αρνητικά την απόδοση του μοντέλου, φαινόμενο που είναι γνωστό ως "κατάρτα της διαστατικότητας".

Η οπτικοποίηση αποτελεί άλλο σημαντικό πλεονέκτημα της PCA. Η PCA είναι ιδιαίτερα αποτελεσματική για την οπτικοποίηση και εξερεύνηση υψηλών διαστάσεων συνόλων δεδομένων, καθώς μπορεί εύκολα να εντοπίσει τάσεις, μοτίβα ή ακραίες τιμές. Η μείωση των χαρακτηριστικών σε 2 ή 3 διαστάσεις επιτρέπει τη δημιουργία διαγραμμάτων διασποράς και άλλων οπτικοποιήσεων, καθιστώντας δυνατό τον εντοπισμό μοτίβων και ομάδων που διαφορετικά θα παρέμεναν κρυμμένα σε χώρους υψηλών διαστάσεων.

Η PCA εξαλείφει την πολυσυγγραμμικότητα μεταξύ των μεταβλητών, καθώς παράγει ένα σύνολο νέων ορθογώνιων αξόνων για την αναπαράσταση των δεδομένων, οι οποίοι είναι ασυσχέτιστοι μεταξύ τους. Αυτό αποτελεί σημαντικό πλεονέκτημα για στατιστικές αναλύσεις και μοντέλα που απαιτούν ανεξάρτητες μεταβλητές (Abdi & Williams, 2010).

Επιπλέον, η PCA συμβάλλει στη μείωση του θορύβου στα δεδομένα. Εστιάζοντας στις συνιστώσες που αντιπροσωπεύουν τη μεγαλύτερη διακύμανση, η PCA φιλτράρει φυσικά τον θόρυβο από το σύνολο δεδομένων. Οι συνιστώσες με χαμηλή διακύμανση, που συχνά σχετίζονται με θόρυβο, απορρίπτονται, βελτιώνοντας έτσι την απόδοση των μοντέλων μηχανικής μάθησης (Libório et al., 2022).

### 3.1.3.2 MEIONEKTHMATA

Παρά τα σημαντικά πλεονεκτήματά της, η PCA παρουσιάζει και ορισμένους περιορισμούς που πρέπει να λαμβάνονται υπόψη. Ο κυριότερος περιορισμός είναι η χαμηλή ερμηνευσιμότητα των κύριων συνιστωσών, καθώς αυτές αποτελούν γραμμικούς συνδυασμούς των χαρακτηριστικών από τα αρχικά δεδομένα, αλλά δεν είναι πάντα εύκολο να ερμηνευτούν. Αυτό καθιστά δύσκολο τον προσδιορισμό των σημαντικότερων χαρακτηριστικών στο σύνολο δεδομένων μετά τον υπολογισμό των κύριων συνιστωσών.

Υπάρχει αναπόφευκτος συμβιβασμός μεταξύ απώλειας πληροφοριών και μείωσης διαστατικότητας. Παρόλο που η μείωση της διαστατικότητας είναι χρήσιμη, συνοδεύεται από κόστος. Η απώλεια πληροφοριών αποτελεί αναγκαίο μέρος της PCA, και η εξισορρόπηση αυτού του συμβιβασμού αποτελεί αναγκαίο στοιχείο που πρέπει να λαμβάνεται υπόψη.

Η PCA είναι ευαίσθητη στην κλίμακα των χαρακτηριστικών και απαιτεί κατάλληλη κανονικοποίηση των δεδομένων πριν από την εφαρμογή, διαφορετικά τα αποτελέσματα μπορεί να είναι παραπλανητικά. Χαρακτηριστικά με τιμές που βρίσκονται σε διαφορετικές τάξεις μεγέθους εμποδίζουν την PCA να υπολογίσει τις βέλτιστες κύριες συνιστώσες.

Η PCA βασίζεται σε γραμμικό μοντέλο και μπορεί να καταγράψει μόνο γραμμικές συσχετίσεις μεταξύ των χαρακτηριστικών, αποτυγχάνοντας όταν παραβιάζεται αυτή η προϋπόθεση. Εάν ένα σύνολο δεδομένων έχει ένα κρυμμένο μοτίβο που είναι μη γραμμικό, τότε η PCA μπορεί πραγματικά να κατευθύνει την ανάλυση στην εντελώς αντίθετη κατεύθυνση από την πρόοδο (Schölkopf et al., 1998)

Η PCA δεν είναι ανθεκτική έναντι ακραίων τιμών, παρόμοια με τα προηγούμενα σημεία, καθώς ο αλγόριθμος θα είναι μεροληπτικός σε σύνολα δεδομένων με ισχυρές ακραίες τιμές. Οι ακραίες τιμές μπορούν δυσανάλογα να επηρεάσουν την κατεύθυνση των κύριων συνιστωσών, οδηγώντας σε παραπλανητικά αποτελέσματα (Hubert et al., 2009).

Τέλος, για πολύ μεγάλα σύνολα δεδομένων, η PCA μπορεί να είναι αργή και να απαιτεί πολλούς πόρους, ενώ υπάρχει κίνδυνος υπερπροσαρμογής όταν χρησιμοποιούνται πάρα πολλές συνιστώσες ή όταν εργάζεται κανείς με μικρό σύνολο δεδομένων. Αυτό μπορεί να οδηγήσει σε μοντέλα που δεν γενικεύουν καλά σε νέα δεδομένα.

Παρά αυτούς τους περιορισμούς, η PCA παραμένει ένα ισχυρό και ευρέως χρησιμοποιούμενο εργαλείο για την ανάλυση δεδομένων, ιδιαίτερα όταν εφαρμόζεται κατάλληλα και λαμβάνοντας υπόψη τους περιορισμούς της (Lippitt et al., 2024).



### 3.1.4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Η Ανάλυση Κύριων Συνιστωσών αποτελεί μια από τις πιο ευρέως διαδεδομένες τεχνικές μείωσης των διαστάσεων. Η μετατροπή των αρχικών δεδομένων σε ένα σύνολο μη συσχετιζόμενων και κάθετων, μεταξύ τους, αξόνων, όπου συμβολίζουν τις κύριες συνιστώσες, συνεισφέρει στην μείωση της πολυπλοκότητας των υπολογισμών και αναδεικνύει τα πιο σημαντικά μοτίβα στο σύνολο των δεδομένων. Επιπλέον, προσφέρει τη δυνατότητα οπτικοποίησης ορισμένων πολυδιάστατων δεδομένων σε δύο ή τρεις διαστάσεις, αφαιρώντας θόρυβο από τις παρατηρήσεις και βελτιώνοντας τη λειτουργία των αλγόριθμων μηχανικής μάθησης.

Από την άλλη πλευρά, η εφαρμογή της ανάλυσης των κυρίων συνιστωσών, οδηγεί σε απώλεια μερικής πληροφορίας, συνήθως της λιγότερο σημαντικής, όταν επιχειρείται μείωση των διαστάσεων των αρχικών δεδομένων, ενώ πρόβλημα μπορεί να αποτελέσει και η ερμηνευσιμότητα των νέων αξόνων που προκύπτουν, των οποίων η πλήρης σημασία δεν είναι εύκολο να αποδοθεί. Εν κατακλείδι, η μέθοδος της ανάλυσης των κυρίων συνιστωσών, παρά τους όποιους περιορισμούς, μπορεί, εάν χρησιμοποιηθεί σωστά, να αποτελέσει ένα ιδιαίτερα χρήσιμο και αποτελεσματικό εργαλείο για την ανάλυση και την κατανόηση μεγάλων και σύνθετων βάσεων δεδομένων.

## 3.2 ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΑΣ (CLUSTER ANALYSIS)

Μετά τη μείωση της διαστατικότητας των δεδομένων μέσω της PCA, το επόμενο βήμα είναι η ομαδοποίηση των θαλάσσιων μέσων σε διακριτές κατηγορίες με ομοιογενή χαρακτηριστικά. Η Ανάλυση Συστάδας είναι μια μέθοδος μη επιβλεπόμενης μάθησης που στοχεύει στον εντοπισμό φυσικών ομαδοποιήσεων (συστάδων) σε ένα σύνολο δεδομένων, όπου τα αντικείμενα εντός της ίδιας συστάδας είναι πιο όμοια μεταξύ τους από ό,τι με τα αντικείμενα σε άλλες συστάδες. Σκοπός της εφαρμογής της στην παρούσα εργασία είναι η κατηγοριοποίηση των πλοίων βάσει των προφίλ εκπομπών ρύπων και των λειτουργικών τους δεδομένων, ώστε να εντοπιστούν πρότυπα και να εξαχθούν εξειδικευμένα συμπεράσματα για κάθε κατηγορία.

### 3.2.1 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Η ανάλυση συστάδας (cluster analysis) είναι μια θεμελιώδης τεχνική της μηχανικής μάθησης και της εξόρυξης δεδομένων, η οποία διαιρεί ένα σύνολο δεδομένων σε ομάδες (συστάδες) που είναι είτε εννοιολογικά σημαντικές, είτε χρήσιμες, είτε και τα δύο (Tan et al., 2016, Chapter 8). Αποτελεί μια μέθοδο μη επιβλεπόμενης μάθησης (unsupervised learning), καθώς εξάγει τις ετικέτες των συστάδων αποκλειστικά από τη δομή των ίδιων των δεδομένων, χωρίς να απαιτείται καμία a priori γνώση ή ταξινόμηση (Tan et al., 2016; Πετρίδης, p. 158). Κεντρικός στόχος της μεθόδου είναι η ομαδοποίηση των αντικειμένων με τέτοιο τρόπο ώστε τα αντικείμενα εντός της ίδιας συστάδας να παρουσιάζουν υψηλή ομοιότητα (homogeneity), ενώ ταυτόχρονα να είναι όσο το δυνατόν πιο ανόμοια με τα αντικείμενα που ανήκουν σε διαφορετικές συστάδες (Tan et al., 2016, Chapter 8). Με άλλα λόγια, η ανάλυση συστάδων επιδιώκει να μεγιστοποιήσει την ενδο-ομαδική ομοιότητα και να ελαχιστοποιήσει τη δια-ομαδική ομοιότητα.

Η ίδια η έννοια της «συστάδας» δεν έχει έναν μοναδικό και αυστηρό ορισμό, καθώς αυτός εξαρτάται από τη φύση των δεδομένων και τον επιδιωκόμενο στόχο της ανάλυσης (Tan et al., 2016, Chapter 8). Ωστόσο, οι συστάδες μπορούν να ταξινομηθούν σε κατηγορίες βάσει των ιδιοτήτων τους (Tan et al., 2016, Chapter 8):

- **Καλά διαχωρισμένες (well-separated):** Κάθε αντικείμενο μιας συστάδας είναι πιο κοντά σε όλα τα άλλα αντικείμενα της ίδιας συστάδας παρά σε οποιοδήποτε αντικείμενο άλλης συστάδας.
- **Βασισμένες σε κάποιο πρωτότυπο (prototype-based):** Κάθε αντικείμενο μιας συστάδας είναι πιο κοντά σε ένα πρωτότυπο (π.χ. κεντροειδής) της δικής του συστάδας από ό,τι στο πρωτότυπο οποιασδήποτε άλλης.
- **Βασισμένες στην πυκνότητα (density-based):** Μια συστάδα ορίζεται ως μια πυκνή περιοχή σημείων, η οποία διαχωρίζεται από άλλες πυκνές περιοχές μέσω μιας περιοχής χαμηλής πυκνότητας (θόρυβος).
- **Βασισμένες στη συνεκτικότητα (contiguity-based):** Μια συστάδα ορίζεται ως ένα σύνολο σημείων όπου κάθε σημείο είναι πιο κοντά σε τουλάχιστον ένα άλλο σημείο της ίδιας συστάδας από ό,τι σε οποιοδήποτε σημείο άλλης συστάδας.

Πυρήνας κάθε αλγορίθμου συσταδοποίησης είναι η μέτρηση της ομοιότητας ή της απόστασης μεταξύ των αντικειμένων. Η επιλογή του κατάλληλου μέτρου είναι κρίσιμη για το

αποτέλεσμα. Για ποσοτικά δεδομένα, η πιο διαδεδομένη μετρική είναι η ευκλείδεια απόσταση (Εξ. 3.7), η οποία υπολογίζεται ως η τετραγωνική ρίζα του αθροίσματος των τετραγώνων των διαφορών των συντεταγμένων (Πετρίδης, p. 162):

$$d_{j,k} = \sqrt{\sum_i (X_{ij} - X_{ik})^2} \quad (3.7)$$

Όπου  $d_{j,k}$  είναι η απόσταση μεταξύ των αντικειμένων  $j$  και  $k$ , και  $X_{ij}$  είναι η τιμή της μεταβλητής  $i$  για το αντικείμενο  $j$ . Όταν επιδιώκεται να δοθεί μεγαλύτερο βάρος σε αντικείμενα που απέχουν πολύ μεταξύ τους, χρησιμοποιείται η τετραγωνική ευκλείδεια απόσταση (Εξ. 3.8) (Πετρίδης, p. 162).

$$d_{j,k} = \sum_i (X_{ij} - X_{ik})^2 \quad (3.8)$$

Μια άλλη σημαντική μετρική είναι η απόσταση Manhattan (Εξ. 3.9), η οποία υπολογίζει το άθροισμα των απόλυτων διαφορών των συντεταγμένων και είναι λιγότερο ευαίσθητη σε ακραίες τιμές (outliers) (Πετρίδης, p. 162).

$$d_{j,k} = \sum_i |X_{ij} - X_{ik}| \quad (3.9)$$

Είναι επιτακτική ανάγκη, πριν την εφαρμογή των μετρικών απόστασης, τα δεδομένα να έχουν τυποποιηθεί (standardized), ώστε οι μεταβλητές με μεγάλο εύρος τιμών να μην κυριαρχούν στον υπολογισμό της απόστασης και να στρεβλώνουν το αποτέλεσμα (Πετρίδης, p. 179).

### 3.2.2 ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΥΣΤΑΔΑΣ

Η εφαρμογή της ανάλυσης συστάδας δεν αποτελεί την εκτέλεση ενός μεμονωμένου αλγορίθμου, αλλά μια επαναληπτική και διαδραστική διαδικασία που περιλαμβάνει συγκεκριμένα στάδια, από την προετοιμασία των δεδομένων έως την τελική ερμηνεία και επικύρωση των αποτελεσμάτων (Πετρίδης, p. 158). Η διαδικασία αυτή είναι εγγενώς διερευνητική, καθώς, ως μέθοδος μη επιβλεπόμενης μάθησης, δεν απαιτεί προκαθορισμένες «ετικέτες» ή κατηγορίες, αλλά αποσκοπεί στην αποκάλυψη της εγγενούς δομής που υπάρχει στα ίδια τα δεδομένα (Tan et al., 2016, Chapter 8). Η μεθοδολογία που ακολουθείται μπορεί να συνοψιστεί στα παρακάτω βήματα.

Αρχικά, κρίνεται απαραίτητη η προεπεξεργασία και κλιμάκωση των δεδομένων. Οι αλγόριθμοι συσταδοποίησης βασίζονται σε μετρικές απόστασης για να ποσοτικοποιήσουν την ομοιότητα μεταξύ των παρατηρήσεων. Εάν οι μεταβλητές (χαρακτηριστικά) του συνόλου δεδομένων βρίσκονται σε διαφορετικές κλίμακες (π.χ. ισχύς μηχανής σε kW και κατανάλωση καυσίμου σε τόνους/ημέρα), οι μεταβλητές με μεγαλύτερο εύρος τιμών θα επηρεάσουν δυσανάλογα τον υπολογισμό της απόστασης. Για τον λόγο αυτό, τα δεδομένα συνήθως κλιμακοποιούνται, μέσω της τυποποίησης (standardization) ή της κανονικοποίησης (normalization), ώστε κάθε μεταβλητή να έχει ισότιμη συνεισφορά στη διαδικασία.

Το επόμενο βήμα είναι η επιλογή μιας κατάλληλης μετρικής απόστασης ή ομοιότητας. Η επιλογή αυτή είναι θεμελιώδης, καθώς καθορίζει πώς ακριβώς ορίζεται η «εγγύτητα» μεταξύ δύο αντικειμένων (πλοίων). Ανάλογα με τη φύση των δεδομένων (ποσοτικά ή κατηγορικά), μπορούν να χρησιμοποιηθούν διάφοροι συντελεστές. Η επιλογή της μετρικής επηρεάζει άμεσα το σχήμα και τη δομή των συστάδων που θα προκύψουν.

Ακολούθως, πραγματοποιείται η επιλογή του αλγορίθμου συσταδοποίησης. Υπάρχουν πολλές κατηγορίες αλγορίθμων, με τις τρεις κυριότερες να είναι η διαμεριστική (partitional), η ιεραρχική (hierarchical) και η βασισμένη στην πυκνότητα (density-based) (Tan et al., 2016, Chapter 8). Η ιεραρχική προσέγγιση δημιουργεί ένα δενδρόγραμμα που απεικονίζει τις συγχωνεύσεις των συστάδων σε διαφορετικά επίπεδα ομοιότητας, ενώ η διαμεριστική (όπως ο K-Means) διαχωρίζει τα δεδομένα σε έναν προκαθορισμένο αριθμό (k) μη αλληλεπικαλυπτόμενων συστάδων (Πετρίδης, p. 160). Η επιλογή εξαρτάται από τους στόχους της έρευνας και τα χαρακτηριστικά των δεδομένων.

Ένα από τα κρίσιμότερα βήματα, ιδίως για τους διαμεριστικούς αλγορίθμους, είναι ο προσδιορισμός του βέλτιστου αριθμού συστάδων (k). Δεδομένου ότι πρόκειται για μη επιβλεπόμενη μάθηση, ο ιδανικός αριθμός συστάδων δεν είναι εκ των προτέρων γνωστός. Για τον προσδιορισμό του χρησιμοποιούνται διάφορες ευρετικές μέθοδοι, όπως:

- **Η μέθοδος του «αγκώνα» (elbow method):** Η μέθοδος αυτή βασίζεται στον υπολογισμό ενός μέτρου ποιότητας, όπως το άθροισμα των τετραγωνικών σφαλμάτων (Sum of Squared Errors - SSE) (Εξ. 3.9), για διαφορετικές τιμές του k. Το SSE ορίζεται ως:

$$SSE = \sum_{i=1}^k \sum_{x \in c_i} \text{dist}(x, c_i)^2 \quad (3.9)$$

Όπου  $C_i$  είναι η  $i$ -οστή συστάδα και  $c_i$  το κέντρο της. Με αυτόν τον τρόπο, σχεδιάζεται το γράφημα του SSE ως προς το  $k$  και επιλέγεται η τιμή του  $k$  στο σημείο όπου το γράφημα σχηματίζει έναν «αγκώνα», υποδηλώνοντας ότι η περαιτέρω αύξηση του  $k$  δεν προσφέρει σημαντική μείωση του σφάλματος (Tan et al., 2016, Chapter 8).

- **Η ανάλυση σιλουέτας (silhouette analysis):** Ο συντελεστής σιλουέτας μετρά το πόσο καλά ταιριάζει ένα αντικείμενο στη συστάδα του σε σύγκριση με τις γειτονικές συστάδες. Υπολογίζεται για κάθε σημείο και ο μέσος συντελεστής για όλες τις παρατηρήσεις μπορεί να χρησιμοποιηθεί για την επιλογή του  $k$  που μεγιστοποιεί αυτή την τιμή (Tan et al., 2016, Chapter 8). Ο συντελεστής αυτός υπολογίζεται ως εξής:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3.10)$$

Όπου  $a(i)$  η μέση απόσταση του σημείου  $i$  από όλα τα άλλα σημεία της ίδιας συστάδας, ενώ  $b(i)$  η ελάχιστη μέση απόσταση του σημείου  $i$  από τα σημεία των άλλων συστάδων.

Τέλος, η διαδικασία ολοκληρώνεται με την επικύρωση και ερμηνεία των αποτελεσμάτων.

Η επικύρωση (cluster validity) εξετάζει την ποιότητα των παραγόμενων συστάδων, αξιολογώντας την εσωτερική συνοχή (cohesion) και τον εξωτερικό διαχωρισμό (separation) τους (Tan et al., 2016, Chapter 8). Η ερμηνεία περιλαμβάνει την ανάλυση των χαρακτηριστικών κάθε συστάδας (π.χ. υπολογισμός των μέσων τιμών των μεταβλητών ανά συστάδα) με σκοπό την απόδοση ενός ουσιαστικού νοήματος σε κάθε ομάδα πλοίων που σχηματίστηκε, συνδέοντας τα στατιστικά αποτελέσματα με το πραγματικό αντικείμενο της έρευνας.

### 3.2.3 ΚΥΡΙΕΣ ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ

Η ανάλυση συστάδας δεν αποτελεί έναν μεμονωμένο αλγόριθμο, αλλά ένα σύνολο τεχνικών που μπορούν να ταξινομηθούν σε ευρύτερες κατηγορίες, καθεμία από τις οποίες είναι κατάλληλη για διαφορετικά είδη δεδομένων και ερευνητικών στόχων.

### 3.2.3.1 ΔΙΑΜΕΡΙΣΤΙΚΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗ (PARTITIONING CLUSTERING)

Η διαμεριστική συσταδοποίηση στοχεύει στη διαίρεση ενός συνόλου δεδομένων σε έναν προκαθορισμένο αριθμό μη επικαλυπτόμενων υποσυνόλων (συστάδων), όπου κάθε αντικείμενο δεδομένων ανήκει σε μία ακριβώς συστάδα (Tan et al., 2016, p. 492).

Η πιο γνωστή και ευρέως χρησιμοποιούμενη τεχνική αυτής της κατηγορίας βασίζεται σε πρωτότυπα (prototypes) και είναι ο αλγόριθμος K-Means. Ο αλγόριθμος K-Means είναι μια τεχνική που επιχειρεί να βρει έναν προκαθορισμένο από τον χρήστη αριθμό συστάδων ( $k$ ), οι οποίες αναπαρίστανται από το κεντροειδές τους (centroid) (Tan et al., 2016, p. 495). Το κεντροειδές ορίζεται συνήθως ως ο μέσος όρος των σημείων που ανήκουν σε μια συστάδα. Η λειτουργία του αλγορίθμου είναι επαναληπτική και περιγράφεται από τα ακόλουθα βήματα (Tan et al., 2016, p. 497):

1. **Αρχικοποίηση:** Επιλέγονται  $K$  αρχικά σημεία ως κεντροειδή των συστάδων.
2. **Επανάληψη:**
  - **Βήμα ανάθεσης:** Κάθε σημείο δεδομένων ανατίθεται στη συστάδα με το πλησιέστερο κεντροειδές.
  - **Βήμα ενημέρωσης:** Το κεντροειδές κάθε συστάδας επανυπολογίζεται ως ο μέσος όρος των σημείων που έχουν ανατεθεί σε αυτήν.
3. **Τερματισμός:** Η διαδικασία επαναλαμβάνεται μέχρι τα κεντροειδή να σταθεροποιηθούν, δηλαδή να μην αλλάζουν πλέον θέση, ή μέχρι να μην αλλάζει η ανάθεση των σημείων στις συστάδες.

Ο στόχος του αλγορίθμου είναι η βελτιστοποίηση μιας αντικειμενικής συνάρτησης, η οποία συνήθως μετρά την ποιότητα της συσταδοποίησης. Για δεδομένα σε Ευκλείδειο χώρο, η πιο κοινή συνάρτηση είναι το SSE (Εξ. 3.9), το οποίο ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει. Το κεντροειδές  $c_i$  που ελαχιστοποιεί το SSE για τη συστάδα  $C_i$  αποδεικνύεται μαθηματικά ότι είναι ο μέσος όρος των σημείων της (Tan et al., 2016, p. 500):

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x \quad (3.11)$$

Όπου  $m_i$  είναι ο αριθμός των σημείων στη συστάδα  $C_i$ .

Παρά την απλότητα και την υπολογιστική του αποδοτικότητα, ο K-Means έχει περιορισμούς. Είναι ευαίσθητος στην αρχική επιλογή των κεντροειδών και μπορεί να συγκλίνει σε

ένα τοπικό, αντί για ολικό, ελάχιστο. Επιπλέον, δυσκολεύεται να εντοπίσει συστάδες που δεν έχουν σφαιρικό σχήμα (non-globular) ή που έχουν σημαντικά διαφορετικά μεγέθη και πυκνότητες (Tan et al., 2016, p. 510-513).

### 3.2.3.2 ΙΕΡΑΡΧΙΚΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗ (HIERARCHICAL CLUSTERING)

Η ιεραρχική συσταδοποίηση παράγει ένα σύνολο ένθετων (nested) συστάδων που οργανώνονται σε μια δενδρική δομή, γνωστή ως δενδρογράμμα (dendrogram) (Tan et al., 2016, p. 492, 515). Σε αντίθεση με τη διαμεριστική, δεν απαιτεί τον προκαθορισμό του αριθμού των συστάδων. Υπάρχουν δύο βασικές στρατηγικές (Πετρίδης, p. 160; Tan et al., 2016, p. 515):

1. **Συσσωρευτική (agglomerative):** Ξεκινά θεωρώντας κάθε σημείο δεδομένων ως μια ξεχωριστή συστάδα. Σε κάθε βήμα, το ζεύγος των πιο «κοντινών» συστάδων συγχωνεύεται, μέχρι τελικά να παραμείνει μία μόνο συστάδα που περιέχει όλα τα δεδομένα. Αυτή είναι η πιο διαδεδομένη προσέγγιση.
2. **Διαιρετική (divisive):** Ξεκινά με μία μόνο συστάδα που περιέχει όλα τα δεδομένα. Σε κάθε βήμα, μία συστάδα διαιρείται σε μικρότερες, μέχρι κάθε σημείο να αποτελεί από μόνο του μια συστάδα.

Η λειτουργία των συσσωρευτικών αλγορίθμων βασίζεται στον ορισμό της απόστασης (proximity) μεταξύ δύο συστάδων. Η επιλογή της μετρικής απόστασης διαφοροποιεί τις τεχνικές μεταξύ τους (Πετρίδης, p. 168; Tan et al., 2016, p. 517):

- **MIN (single linkage ή απλή διασύνδεση):** Η απόσταση μεταξύ δύο συστάδων ορίζεται ως η ελάχιστη απόσταση μεταξύ οποιωνδήποτε δύο σημείων που ανήκουν στις δύο διαφορετικές συστάδες. Είναι καλή στο χειρισμό μη ελλειπτικών σχημάτων, αλλά ευαίσθητη στο θόρυβο.
- **MAX (complete linkage ή πλήρης διασύνδεση):** Η απόσταση ορίζεται ως η μέγιστη απόσταση μεταξύ οποιωνδήποτε δύο σημείων στις δύο συστάδες. Είναι λιγότερο ευαίσθητη στο θόρυβο, αλλά τείνει να «σπάει» μεγάλες συστάδες και ευνοεί τα σφαιρικά σχήματα.
- **Group Average:** Η απόσταση ορίζεται ως η μέση απόσταση μεταξύ όλων των ζευγών σημείων στις δύο διαφορετικές συστάδες, αποτελώντας μια ενδιάμεση προσέγγιση μεταξύ MIN και MAX.

Το κύριο μειονέκτημα των ιεραρχικών μεθόδων είναι η υπολογιστική τους πολυπλοκότητα, η οποία είναι τουλάχιστον  $O(m^2)$ , καθιστώντας τις μη πρακτικές για πολύ μεγάλα σύνολα δεδομένων (Tan et al., 2016, p. 518). Επίσης, οι αποφάσεις συγχώνευσης είναι τελικές και δεν μπορούν να αναιρεθούν σε μεταγενέστερο στάδιο (Tan et al., 2016, p. 525).

### 3.2.3.3 ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΒΑΣΕΙ ΠΥΚΝΟΤΗΤΑΣ (DENSITY-BASED CLUSTERING)

Σε αυτή την προσέγγιση, μια συστάδα ορίζεται ως μια πυκνή περιοχή σημείων, η οποία διαχωρίζεται από άλλες πυκνές περιοχές από αραιές περιοχές (Tan et al., 2016, p. 494). Αυτή η μέθοδος είναι ιδιαίτερα αποτελεσματική για την εύρεση συστάδων με μη σφαιρικό σχήμα και για τον χειρισμό του θορύβου (noise) στα δεδομένα. Ο πιο αντιπροσωπευτικός αλγόριθμος είναι ο DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

Ο DBSCAN βασίζεται σε έναν ορισμό της πυκνότητας που εξετάζει την τοπική γειτονιά κάθε σημείου, ταξινομώντας τα σημεία σε τρεις κατηγορίες (Tan et al., 2016, p. 527):

- **Σημείο πυρήνα (core point):** Ένα σημείο θεωρείται σημείο πυρήνα εάν ο αριθμός των γειτόνων του εντός μιας καθορισμένης ακτίνας (Eps) υπερβαίνει έναν ελάχιστο αριθμό σημείων (MinPts).
- **Σημείο συνόρου (border point):** Ένα σημείο που δεν είναι σημείο πυρήνα, αλλά βρίσκεται εντός της γειτονιάς (Eps) ενός σημείου πυρήνα.
- **Σημείο θορύβου (noise point):** Ένα σημείο που δεν είναι ούτε πυρήνας ούτε σύνορο.

Ο αλγόριθμος λειτουργεί ως εξής: εντοπίζει τα σημεία πυρήνα και τα συνδέει εάν βρίσκονται σε απόσταση μικρότερη ή ίση του Eps. Κάθε συνδεδεμένη ομάδα σημείων πυρήνα σχηματίζει μια συστάδα. Στη συνέχεια, κάθε σημείο συνόρου ανατίθεται στη συστάδα του κοντινότερου σημείου πυρήνα. Τα σημεία θορύβου απορρίπτονται και δεν ανήκουν σε καμία συστάδα (Tan et al., 2016, p. 528).

Το βασικό πλεονέκτημα του DBSCAN είναι η ικανότητά του να βρίσκει συστάδες αυθαίρετου σχήματος και η ανθεκτικότητά του στο θόρυβο. Ωστόσο, δυσκολεύεται όταν οι συστάδες έχουν πολύ διαφορετικές πυκνότητες, καθώς η επιλογή ενός μοναδικού ζεύγους παραμέτρων (Eps, MinPts) για ολόκληρο το σύνολο δεδομένων μπορεί να μην είναι κατάλληλη (Tan et al., 2016, p. 529).



### 3.2.4 ΠΛΕΟΝΕΚΤΗΜΑΤΑ & ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΥΣΤΑΔΑΣ

Η ανάλυση συστάδας αποτελεί ένα εξαιρετικά ισχυρό εργαλείο της μη επιβλεπόμενης μάθησης, ικανό να αποκαλύψει εγγενείς δομές και πρότυπα σε σύνολα δεδομένων. Ωστόσο, όπως κάθε στατιστική τεχνική, παρουσιάζει τόσο σημαντικά πλεονεκτήματα όσο και συγκεκριμένους περιορισμούς που οφείλουν να λαμβάνονται υπόψη κατά την εφαρμογή και την ερμηνεία των αποτελεσμάτων της.

#### 3.2.4.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ

Το θεμελιώδες πλεονέκτημα της ανάλυσης συστάδας είναι ο διερευνητικός της χαρακτήρας. Ως μέθοδος μη επιβλεπόμενης μάθησης, δεν απαιτεί προκαθορισμένες υποθέσεις για τις σχέσεις μεταξύ των δεδομένων ή την ύπαρξη συγκεκριμένων κλάσεων. Αυτό την καθιστά ιδανική για την αρχική φάση ανάλυσης, όπου ο στόχος είναι η ανακάλυψη «φυσικών» ομαδοποιήσεων και η κατανόηση της δομής των δεδομένων.

Η ικανότητά της να εντοπίζει κρυμμένα πρότυπα βρίσκει εφαρμογή σε ένα ευρύ φάσμα επιστημονικών πεδίων, όπως η βιολογία (ταξινόμηση γονιδίων), η ανάκτηση πληροφοριών (ομαδοποίηση αποτελεσμάτων αναζήτησης), οι επιχειρήσεις (τμηματοποίηση πελατών) και η ιατρική (αναγνώριση υποκατηγοριών ασθενειών) (Tan et al., 2016, p. 488).

Ένα δεύτερο σημαντικό πλεονέκτημα είναι η ευελιξία και η ποικιλία των διαθέσιμων αλγορίθμων. Υπάρχει πληθώρα τεχνικών που έχουν αναπτυχθεί για να ανταποκρίνονται σε διαφορετικούς τύπους δεδομένων και διαφορετικούς ορισμούς της έννοιας «συστάδα». Για παράδειγμα, οι διαμεριστικοί αλγόριθμοι, όπως ο K-Means, είναι εξαιρετικά αποδοτικοί για την ανίχνευση σφαιρικών συστάδων που ορίζονται γύρω από ένα κεντροειδές (prototype-based clusters) (Tan et al., 2016, p. 494).

Οι ιεραρχικοί αλγόριθμοι δημιουργούν μια δενδρική δομή εμφωλευμένων συστάδων (δενδρόγραμμα), κάτι που είναι ιδιαίτερα χρήσιμο σε εφαρμογές που απαιτούν ταξινομική οργάνωση, όπως στη βιολογία (Tan et al., 2016, p. 492). Οι αλγόριθμοι βασισμένοι στην πυκνότητα, όπως ο DBSCAN, μπορούν να εντοπίσουν συστάδες αυθαίρετου σχήματος και είναι ανθεκτικοί στην παρουσία θορύβου και ακραίων τιμών (Tan et al., 2016, p. 494).

Τέλος, η ανάλυση συστάδας προσφέρει πρακτικά οφέλη στη διαχείριση μεγάλων συνόλων δεδομένων. Μπορεί να χρησιμοποιηθεί για τη σύνοψη των δεδομένων, όπου αντί να εφαρμοστεί

ένας υπολογιστικά ακριβός αλγόριθμος (π.χ. PCA) σε ολόκληρο το σύνολο δεδομένων, εφαρμόζεται στα πρωτότυπα (centroids) των συστάδων, μειώνοντας δραστικά τον απαιτούμενο χρόνο. Επιπλέον, χρησιμοποιείται για τη συμπίεση δεδομένων (compression), κυρίως σε εικόνες και βίντεο, μέσω της τεχνικής της διανυσματικής κβάντισης (vector quantization) (Tan et al., 2016, p. 489).

#### 3.2.4.2 ΜΕΙΟΝΕΚΤΗΜΑΤΑ

Το κυριότερο μειονέκτημα της μεθόδου είναι η υποκειμενικότητα και η απουσία ενός καθολικά αποδεκτού ορισμού της "συστάδας". Όπως επισημαίνεται, η έννοια της συστάδας δεν είναι καλά ορισμένη και η βέλτιστη λύση εξαρτάται από τη φύση των δεδομένων και τον επιθυμητό στόχο της ανάλυσης (Tan et al., 2016, p. 490). Ένα επικίνδυνο επακόλουθο είναι ότι σχεδόν όλοι οι αλγόριθμοι συσταδοποίησης θα εντοπίσουν συστάδες ακόμη και σε ένα σύνολο δεδομένων που δεν έχει καμία φυσική δομή (π.χ. δεδομένα από ομοιόμορφη τυχαία κατανομή) (Tan et al., 2016, p. 532). Αυτό καθιστά την αξιολόγηση των αποτελεσμάτων (cluster validation) ένα κρίσιμο, αλλά συχνά παραμελημένο, βήμα.

Ένας δεύτερος σημαντικός περιορισμός είναι η ευαισθησία των αλγορίθμων στις αρχικές παραμέτρους. Η τελική μορφή των συστάδων εξαρτάται σε μεγάλο βαθμό από τις επιλογές του αναλυτή, καθιστώντας τη διαδικασία μια «επαναληπτική διαδραστική διεργασία βελτιστοποίησης» παρά μια αυτόματη λύση. Κρίσιμες παράμετροι περιλαμβάνουν την επιλογή της μετρικής απόστασης ή ομοιότητας (π.χ. Ευκλείδεια, Manhattan, Pearson), τον προκαθορισμό του αριθμού των συστάδων ( $k$ ), που απαιτείται από αλγορίθμους όπως ο K-Means, και την επιλογή του ίδιου του αλγορίθμου, καθώς διαφορετικοί αλγόριθμοι έχουν διαφορετικά biases.

Η επιλογή του βέλτιστου  $k$  αποτελεί μια από τις μεγαλύτερες προκλήσεις, αν και τεχνικές όπως η μέθοδος του αγκώνα ή ο συντελεστής silhouette μπορούν να βοηθήσουν (Tan et al., 2016, p. 547).

Επιπλέον, πολλοί διαδεδομένοι αλγόριθμοι έχουν εγγενείς αδυναμίες. Ο K-Means, για παράδειγμα, δυσκολεύεται να εντοπίσει συστάδες που δεν είναι σφαιρικές, που έχουν διαφορετικά μεγέθη ή πυκνότητες, και είναι ευαίσθητος σε ακραίες τιμές (Tan et al., 2016, p. 510, 513). Οι ιεραρχικοί αλγόριθμοι, από την άλλη πλευρά, λαμβάνουν αποφάσεις τοπικά σε κάθε βήμα (π.χ. συγχωνεύοντας τις δύο πλησιέστερες ομάδες), οι οποίες είναι οριστικές και δεν μπορούν να

αναιρεθούν αργότερα. Αυτή η «άπληστη» (greedy) προσέγγιση δεν εγγυάται την εύρεση μιας καθολικά βέλτιστης λύσης (Tan et al., 2016, p. 525).

Επιπλέον, η υπολογιστική τους πολυπλοκότητα, που συχνά είναι τάξης  $O(m^2)$ , τους καθιστά ακατάλληλους για πολύ μεγάλα σύνολα δεδομένων (Πολυμελής, 2018, σελ. 160). Η βελτιστοποίηση ενός καθολικού κριτηρίου, όπως το άθροισμα των τετραγωνικών σφαλμάτων, είναι υπολογιστικά ανέφικτη για μεγάλα δεδομένα. Οι αλγόριθμοι προσπαθούν να ελαχιστοποιήσουν τοπικά τέτοιες συναρτήσεις, αλλά συχνά παγιδεύονται σε τοπικά ελάχιστα (local minima), οδηγώντας σε υπο-βέλτιστες συσταδοποιήσεις (Tan et al., 2016, p. 500, 525).

### 3.2.5 ΣΥΜΠΕΡΑΣΜΑΤΑ

Η ανάλυση συστάδας αποτελεί μια ισχυρή και ευέλικτη τεχνική μη επιβλεπόμενης μάθησης που επιτρέπει την ανακάλυψη φυσικών ομαδοποιήσεων σε πολυδιάστατα σύνολα δεδομένων. Στο πλαίσιο της παρούσας εργασίας, η μέθοδος προσφέρει τη δυνατότητα κατηγοριοποίησης των θαλάσσιων μέσων μεταφοράς βάσει των χαρακτηριστικών εκπομπών τους και των λειτουργικών παραμέτρων τους, συμβάλλοντας στην ανάπτυξη εξειδικευμένων στρατηγικών περιβαλλοντικής διαχείρισης για κάθε κατηγορία πλοίων.

Η ποικιλία των διαθέσιμων αλγορίθμων (διαμεριστικοί, ιεραρχικοί, βασισμένοι στην πυκνότητα) επιτρέπει την προσαρμογή της μεθόδου στις ιδιαιτερότητες των δεδομένων εκπομπών. Οι διαμεριστικοί αλγόριθμοι, όπως ο K-Means, είναι ιδανικοί για την ταχεία κατηγοριοποίηση μεγάλου αριθμού πλοίων σε προκαθορισμένες ομάδες, ενώ οι ιεραρχικοί αλγόριθμοι παρέχουν πολύτιμη πληροφορία για τις σχέσεις μεταξύ των διαφορετικών τύπων πλοίων μέσω της δενδρικής δομής που δημιουργούν.

Ωστόσο, η επιτυχής εφαρμογή της μεθόδου απαιτεί προσεκτική αντιμετώπιση των εγγενών περιορισμών της. Η υποκειμενικότητα στην επιλογή παραμέτρων, η ευαισθησία σε ακραίες τιμές και η δυσκολία προσδιορισμού του βέλτιστου αριθμού συστάδων αποτελούν προκλήσεις που πρέπει να αντιμετωπιστούν μέσω της χρήσης κατάλληλων τεχνικών επικύρωσης και της εφαρμογής πολλαπλών αλγορίθμων για τη σύγκριση αποτελεσμάτων.

Εν κατακλείδι, η ανάλυση συστάδας, σε συνδυασμό με την προηγούμενη εφαρμογή της PCA, δημιουργεί ένα ολοκληρωμένο πλαίσιο ανάλυσης που μειώνει την πολυπλοκότητα των δεδομένων εκπομπών και αποκαλύπτει τα κρυμμένα πρότυπα συμπεριφοράς των θαλάσσιων

μέσων μεταφοράς. Αυτή η προσέγγιση καθιστά δυνατή την ανάπτυξη στοχευμένων πολιτικών μείωσης των περιβαλλοντικών επιπτώσεων, λαμβάνοντας υπόψη τις ιδιαιτερότητες κάθε κατηγορίας πλοίων.

#### 4. ΑΝΑΛΥΣΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

Αφού στο προηγούμενο κεφάλαιο τέθηκαν οι θεωρητικές βάσεις των μεθόδων της μηχανικής μάθησης που θα αξιοποιηθούν, το παρόν κεφάλαιο επικεντρώνεται στην πρακτική εφαρμογή τους, με σκοπό τη μετατροπή του ακατέργαστου συνόλου δεδομένων σε ουσιαστικά συμπεράσματα σχετικά με την περιβαλλοντική απόδοση των θαλάσσιων μέσων μεταφοράς. Η διαδικασία ξεκινά με την περιγραφή της πηγής άντλησης των δεδομένων από την πλατφόρμα THETIS-MRV και την αυστηρή διαδικασία προεπεξεργασίας που ακολουθήθηκε για την εξασφάλιση της ποιότητας και της αξιοπιστίας της ανάλυσης. Στη συνέχεια, παρουσιάζονται τα αποτελέσματα της PCA, μέσω της οποίας οι πολυάριθμες αρχικές μεταβλητές συμπυκνώθηκαν σε τρεις νέους, ερμηνεύσιμους δείκτες. Αυτοί οι δείκτες αποτελούν τη βάση για την ανάλυση συστάδων, όπου αποκαλύπτονται οι φυσικές ομαδοποιήσεις των πλοίων με βάση τα κοινά περιβαλλοντικά και λειτουργικά τους χαρακτηριστικά. Μέσα από διαγράμματα, πίνακες και τη στατιστική ερμηνεία, το κεφάλαιο αυτό αποτυπώνει την πορεία από τα δεδομένα στη γνώση, θέτοντας τις βάσεις για τη συζήτηση και τα τελικά συμπεράσματα που θα ακολουθήσουν.

##### 4.1 ΠΗΓΗ ΑΝΤΛΗΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΗΣ ΕΡΕΥΝΑΣ

Η πλατφόρμα THETIS-MRV, η οποία λειτουργεί υπό την εποπτεία της Ευρωπαϊκής Υπηρεσίας για την Ασφάλεια στη Θάλασσα (EMSA), παρέχει δημόσια πρόσβαση σε αναλυτικά δεδομένα για τις εκπομπές αερίων θερμοκηπίου από τη ναυτιλία που δραστηριοποιείται εντός της Ευρωπαϊκής Ένωσης. Τα δεδομένα αφορούν πλοία άνω των 5.000 GT που πραγματοποιούν πλόες προς ή από λιμάνια της ΕΕ και περιλαμβάνουν μεταξύ άλλων στοιχεία για την κατανάλωση καυσίμου, τις εκπομπές CO<sub>2</sub>, τη διανυθείσα απόσταση, το χρόνο εν πλώ και δείκτες ενεργειακής απόδοσης. Από το 2024, προστίθενται επίσης δεδομένα για τις εκπομπές μεθανίου (CH<sub>4</sub>) και υποξειδίου του αζώτου (N<sub>2</sub>O), ενώ από το 2025 επεκτείνεται η υποχρέωση αναφοράς και σε πλοία 400–5.000 GT ορισμένων τύπων (European Maritime Safety Agency (EMSA), 2025).

Κάθε πλοίο υποβάλλει ετήσια έκθεση εκπομπών (Emissions Report) μέσω της πλατφόρμας, η οποία πρέπει να επαληθεύεται από ανεξάρτητο οργανισμό και να συνοδεύεται από το σχετικό Πιστοποιητικό Συμμόρφωσης (Document of Compliance). Τα δεδομένα που καταχωρούνται είναι διαθέσιμα στο κοινό και επιτρέπουν την παρακολούθηση της προόδου του

τομέα προς την κατεύθυνση της πράσινης μετάβασης. Επιπλέον, αν υπάρξει αλλαγή στη διαχειρίστρια εταιρεία εντός του ίδιου έτους, πρέπει να υποβάλλεται μερική έκθεση (Partial Emission Report) εντός τριών μηνών από την αλλαγή.

Τα δεδομένα του THETIS-MRV χρησιμοποιούνται ευρέως από τις ευρωπαϊκές αρχές για την εκπόνηση πολιτικών, την υποστήριξη των στόχων του European Green Deal και τη σταδιακή ενσωμάτωση της ναυτιλίας στο Ευρωπαϊκό Σύστημα Εμπορίας Δικαιωμάτων Εκπομπών (EU ETS). Η διαθεσιμότητα αυτών των στοιχείων συμβάλλει στη διαφάνεια της ναυτιλιακής δραστηριότητας εντός της ΕΕ και επιτρέπει τη συγκριτική αξιολόγηση της ενεργειακής απόδοσης διαφορετικών πλοίων ή εταιρειών. Συνολικά, το σύστημα MRV (Monitoring, Reporting and Verification) μέσω της πλατφόρμας THETIS-MRV αποτελεί ένα ισχυρό εργαλείο για την παρακολούθηση και αξιολόγηση των περιβαλλοντικών επιδόσεων της ναυτιλίας, ενισχύοντας τον ρόλο της Ευρωπαϊκής Ένωσης στην παγκόσμια προσπάθεια για μείωση των εκπομπών στον ναυτιλιακό τομέα.

## **4.2 ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΔΙΑΜΟΡΦΩΣΗ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ**

Η έκθεση εκπομπών του EU MRV για το 2024 περιλαμβάνει 113 στήλες που καλύπτουν ένα ευρύ φάσμα λειτουργικών, περιβαλλοντικών και τεχνικών παραμέτρων των πλοίων. Ωστόσο, δεν είναι όλες οι μεταβλητές εξίσου σημαντικές για την ανάπτυξη δεικτών περιβαλλοντικής απόδοσης μέσω τεχνικών μηχανικής μάθησης. Η παρούσα ενότητα περιγράφει τη συστηματική προσέγγιση που χρησιμοποιήθηκε για τη μείωση του συνόλου δεδομένων από 113 μεταβλητές σε 16 βασικές μεταβλητές που αιχμαλωτίζουν τις ουσιώδεις πτυχές της περιβαλλοντικής απόδοσης των πλοίων.

Το σύνολο δεδομένων του 2024 αντιπροσωπεύει μια σημαντική επέκταση σε σχέση με τα προηγούμενα έτη, ενσωματώνοντας νέες κανονιστικές απαιτήσεις που εισήχθησαν στο πλαίσιο του ενισχυμένου κλιματικού πλαισίου πολιτικής της ΕΕ. Από την 1η Ιανουαρίου 2024, οι εταιρείες πρέπει να παρακολουθούν και να αναφέρουν τις εκπομπές μεθανίου (CH<sub>4</sub>) και υποξειδίου του αζώτου (N<sub>2</sub>O), επιπλέον των εκπομπών CO<sub>2</sub>, αντικατοπτρίζοντας τη δέσμευση της ΕΕ για ολοκληρωμένη παρακολούθηση των αερίων του θερμοκηπίου στη ναυτιλία.

Η διαδικασία επιλογής μεταβλητών καθοδηγήθηκε από πολλά βασικά κριτήρια που προέκυψαν από το μεθοδολογικό πλαίσιο που καθιέρωσαν οι Skarlatos et al. (2023) και

προσαρμόστηκε για το ενισχυμένο κανονιστικό περιβάλλον του 2024. Πρώτον, δόθηκε προτεραιότητα σε μεταβλητές που σχετίζονται άμεσα με τις εκπομπές αερίων του θερμοκηπίου και την περιβαλλοντική απόδοση, ιδιαίτερα σε αυτές που αιχμαλωτίζουν το διευρυμένο πεδίο παρακολούθησης αερίων θερμοκηπίου που εισήχθη το 2024. Δεύτερον, επιλέχθηκαν μεταβλητές που αντικατοπτρίζουν τη λειτουργική αποδοτικότητα και μπορούν να χρησιμεύσουν ως δείκτες περιβαλλοντικής απόδοσης εν απουσία λεπτομερών τεχνικών προδιαγραφών πλοίων. Τέλος, προτιμήθηκαν μεταβλητές με υψηλή διαθεσιμότητα και αξιοπιστία δεδομένων σε όλο το σύνολο δεδομένων για να διασφαλιστεί ισχυρή στατιστική ανάλυση.

Με βάση αυτά τα κριτήρια, επιλέχθηκαν οι ακόλουθες 16 μεταβλητές για ανάλυση. Οι μεταβλητές αναγνώρισης πλοίου περιλαμβάνουν τον αριθμό IMO ως μοναδικό αναγνωριστικό πλοίου, το όνομα του πλοίου για αναφορά και επαλήθευση, και τον τύπο πλοίου ως κατηγορική μεταβλητή για ταξινόμηση και συγκριτική ανάλυση. Οι βασικές μεταβλητές περιβαλλοντικής απόδοσης περιλαμβάνουν τη συνολική κατανάλωση καυσίμου σε μετρικούς τόνους ως πρωτεύον μέτρο περιβαλλοντικού αντίκτυπου, τις συνολικές εκπομπές CO<sub>2</sub> ως κύριο δείκτη εκπομπών αερίων θερμοκηπίου, τις συνολικές εκπομπές CH<sub>4</sub> και N<sub>2</sub>O που αποτελούν νέα απαίτηση από το 2024, και τις συνολικές εκπομπές CO<sub>2eq</sub> ως ολοκληρωμένο μέτρο αντίκτυπου αερίων θερμοκηπίου. Οι μεταβλητές λειτουργικής αποδοτικότητας περιλαμβάνουν τον χρόνο που διανύθηκε στη θάλασσα σε ώρες ως μέτρο λειτουργικής έντασης, την κατανάλωση καυσίμου ανά απόσταση και τις εκπομπές CO<sub>2</sub> ανά απόσταση ως άμεσους δείκτες αποδοτικότητας, καθώς και την κατανάλωση καυσίμου και τις εκπομπές CO<sub>2</sub> ανά μονάδα μεταφορικού έργου προσαρμοσμένες στη χωρητικότητα φορτίου.

Επιπλέον, επιλέχθηκαν τρεις μεταβλητές που αφορούν στα περιφερειακά πρότυπα λειτουργίας: οι εκπομπές CO<sub>2</sub> από όλα τα ταξίδια μεταξύ λιμανιών υπό τη δικαιοδοσία κράτους μέλους, οι εκπομπές CO<sub>2</sub> από όλα τα ταξίδια που αναχώρησαν από λιμάνια υπό τη δικαιοδοσία κράτους μέλους, και οι εκπομπές CO<sub>2</sub> από όλα τα ταξίδια προς λιμάνια υπό τη δικαιοδοσία κράτους μέλους. Αυτές οι περιφερειακές μεταβλητές αιχμαλωτίζουν τη γεωγραφική κατανομή των λειτουργιών πλοίων εντός των δικαιοδοσιών των κρατών μελών της ΕΕ, παρέχοντας γνώσεις για τα λειτουργικά πρότυπα και την κανονιστική κάλυψη.

Πολλές κατηγορίες μεταβλητών εξαιρέθηκαν από την ανάλυση. Σε αντίθεση με την αρχική μελέτη των Skarlatos et al. (2023), λεπτομερείς τεχνικές προδιαγραφές όπως το νεκρό βάρος, οι διαστάσεις και η ισχύς μηχανών δεν ήταν διαθέσιμες στο δημόσιο σύνολο δεδομένων MRV,

καθιστώντας αναγκαία την εστίαση σε λειτουργικά παρά σε μέτρα αποδοτικότητας βασισμένα στο σχεδιασμό. Επιπλέον, μεταβλητές που παρείχαν πολύ υψηλό βαθμό ανάλυσης ή ήταν ειδικές για συγκεκριμένα λειτουργικά σενάρια εξαιρέθηκαν, ώστε να διατηρηθεί η εστίαση στα συνολικά πρότυπα απόδοσης. Πολλαπλοί παρόμοιοι δείκτες αποδοτικότητας ενοποιήθηκαν για να αποφευχθούν ζητήματα πολυγραμμικότητας στη μετέπειτα στατιστική ανάλυση, ενώ μεταβλητές με σημαντικά ελλιπή δεδομένα σε όλο το σύνολο δεδομένων εξαιρέθηκαν για να διασφαλιστεί ισχυρή στατιστική ανάλυση.

**Πίνακας 4.1:** Παραδείγματα μεταβλητών που αφαιρέθηκαν από το σύνολο των δεδομένων, καθώς και ο λόγος αφαίρεσης της κάθε μιας.

Μεταβλητές που εξαιρέθηκαν	Αιτία αφαίρεσης
Κατανάλωση καυσίμου ανά μεταφορικό έργο (μάζα) σε ταξίδια με φορτίο [g / m τόνους · n μίλια]	Παρατηρήθηκαν 13280 κενές τιμές στη στήλη από το σύνολο των 13280 εγγραφών του dataset.
Εκπομπές CO <sub>2</sub> eq ανά μεταφορικό έργο (επιβάτης) [g CO <sub>2</sub> eq / επιβάτης · n μίλια]	Αρκετές κενές τιμές και πολύ εξειδικευμένο λειτουργικό σενάριο.
Εκπομπές CH <sub>4</sub> από όλα τα ταξίδια μεταξύ λιμένων που υπάγονται στη δικαιοδοσία ενός κράτους μέλους [εκατόν τόνοι]	Μεταβλητή που αναφέρεται σε έναν πολύ εξειδικευμένο υπολογισμό, εκτός του βασικού σκοπού της ανάλυσης.
Μεταβλητές-γνωρίσματα που σχετίζονται με την εταιρεία στην οποία υπάγεται το πλοίο ή η ελεγκτική αρχή που έκανε τις μετρήσεις, όπως Company Name ή Verifier Name.	Κρίνονται μη σχετικές με τον σκοπό της ανάλυσης μου.

Αυτή η συστηματική διαδικασία επιλογής μεταβλητών διασφαλίζει ότι το τελικό σύνολο δεδομένων αιχμαλωτίζει τις ουσιώδεις διαστάσεις της περιβαλλοντικής απόδοσης των πλοίων, διατηρώντας παράλληλα την αναλυτική διαχειρισσιμότητα και την ευθυγράμμιση με τα τρέχοντα κανονιστικά πλαίσια και τους στόχους βιωσιμότητας στη ναυτιλία.

**Πίνακας 4.2:** Περιγραφή συνόλου δεδομένων.

Μεταβλητή	Περιγραφή
-----------	-----------



Αναγνωριστικά Πλοίου	
X1	Μοναδικός αναγνωριστικός κωδικός πλοίου για παρακολούθηση και επικύρωση
X2	Όνομα πλοίου για αναφορά και επαλήθευση
X3	Κατηγορική μεταβλητή για την ταξινόμηση των πλοίων και τη συγκριτική ανάλυση
Βασικές μεταβλητές περιβαλλοντικής απόδοσης	
X4	Συνολική κατανάλωση καυσίμου (μετρικοί τόνοι)
X5	Συνολικές εκπομπές CO <sub>2</sub> (μετρικοί τόνοι)
X6	Συνολικές εκπομπές CH <sub>4</sub> (μετρικοί τόνοι)
X7	Συνολικές εκπομπές N <sub>2</sub> O (μετρικοί τόνοι)
X8	Συνολικές εκπομπές CO <sub>2</sub> eq (μετρικοί τόνοι)
Μεταβλητές λειτουργικής αποδοτικότητας	
X9	Συνολικός χρόνος στη θάλασσα (ώρες)
X10	Κατανάλωση καυσίμου ανά απόσταση (kg / ναυτικό μίλι)
X11	Εκπομπές CO <sub>2</sub> ανά απόσταση (kg CO <sub>2</sub> / ναυτικό μίλι)
X12	Κατανάλωση καυσίμου ανά μεταφορικό έργο (μάζα) (g / μετρικούς τόνους · ναυτικά μίλια)
X13	Εκπομπές CO <sub>2</sub> ανά μεταφορικό έργο (μάζα) (g CO <sub>2</sub> / μετρικούς τόνους · ναυτικά μίλια)
Περιφερειακές μεταβλητές λειτουργίας	
X14	Εκπομπές CO <sub>2</sub> από όλα τα ταξίδια μεταξύ λιμανιών υπό τη δικαιοδοσία κράτους μέλους (μετρικοί τόνοι)
X15	Εκπομπές CO <sub>2</sub> από όλα τα ταξίδια που αναχώρησαν από λιμάνια υπό τη δικαιοδοσία κράτους μέλους (μετρικοί τόνοι)
X16	Εκπομπές CO <sub>2</sub> από όλα τα ταξίδια προς λιμάνια υπό τη δικαιοδοσία κράτους μέλους (μετρικοί τόνοι)

### 4.3 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Πριν από την διερευνητική ανάλυση των δεδομένων (Exploratory Data Analysis – EDA), οι παρατηρήσεις σε κάθε μεταβλητή  $X_i$  τυποποιήθηκαν χρησιμοποιώντας την ακόλουθη φόρμουλα (Εξ. 4.1):

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \quad (4.1)$$

όπου  $\mu_i$  είναι ο μέσος όρος της  $i$ -οστής μεταβλητής και  $\sigma_i$  είναι η τυπική απόκλιση της  $i$ -οστής μεταβλητής. Δεδομένου ότι στις περισσότερες περιπτώσεις οι πραγματικοί μέσοι όροι και διακυμάνσεις είναι άγνωστοι, χρησιμοποιήθηκαν οι αμερόληπτοι δειγματικοί εκτιμητές τους.

Το αρχικό σύνολο δεδομένων περιλάμβανε 13.280 παρατηρήσεις με 16 μεταβλητές. Κατά τη διαδικασία καθαρισμού των δεδομένων, εντοπίστηκαν και αντιμετωπίστηκαν δύο κύριες κατηγορίες προβλημάτων ποιότητας δεδομένων. Πρώτον, εντοπίστηκαν 2.554 τιμές NaN στις μεταβλητές που αφορούν την κατανάλωση καυσίμου και τις εκπομπές CO<sub>2</sub> ανά μεταφορικό έργο (μάζα), υποδεικνύοντας ελλιπή δεδομένα για συγκεκριμένες παρατηρήσεις. Δεύτερον, εντοπίστηκαν τιμές "Division by zero!" σε διάφορες μεταβλητές, με 934 τέτοιες τιμές στις μεταβλητές κατανάλωσης καυσίμου και εκπομπών CO<sub>2</sub> ανά απόσταση, και 1.436 τιμές στις μεταβλητές μεταφορικού έργου. Αυτές οι τιμές προέκυψαν από μαθηματικούς υπολογισμούς όπου ο παρονομαστής ήταν μηδέν, υποδεικνύοντας προβλήματα στην καταγραφή ή τον υπολογισμό συγκεκριμένων λειτουργικών παραμέτρων.

Για την αντιμετώπιση αυτών των προβλημάτων ποιότητας δεδομένων, εφαρμόστηκε μια συστηματική προσέγγιση καθαρισμού. Αρχικά, όλες οι τιμές "Division by zero!" μετατράπηκαν σε τιμές NaN για ομοιόμορφη διαχείριση των ελλিপών δεδομένων. Στη συνέχεια, όλες οι παρατηρήσεις που περιείχαν τουλάχιστον μία τιμή NaN αφαιρέθηκαν από το σύνολο δεδομένων, ακολουθώντας την προσέγγιση που εφαρμόστηκε στην αρχική μελέτη των Skarlatos et al. (2023). Αυτή η διαδικασία οδήγησε στην αφαίρεση 3.990 παρατηρήσεων, με το τελικό καθαρισμένο σύνολο δεδομένων να περιλαμβάνει 9.290 παρατηρήσεις.

Μετά τον καθαρισμό των δεδομένων, εφαρμόστηκε η διαδικασία τυποποίησης στις 13 αριθμητικές μεταβλητές, εξαιρώντας τις μεταβλητές αναγνώρισης (αριθμός IMO, όνομα πλοίου, τύπος πλοίου). Η τυποποίηση πραγματοποιήθηκε χρησιμοποιώντας τον StandardScaler της βιβλιοθήκης scikit-learn, παράγοντας τυποποιημένες μεταβλητές Z1 έως Z13 που αντιστοιχούν στις αρχικές μεταβλητές X1 έως X13. Η επαλήθευση της διαδικασίας τυποποίησης επιβεβαίωσε

ότι όλες οι τυποποιημένες μεταβλητές έχουν μέσο όρο πρακτικά ίσο με το μηδέν ( $\approx 0$ ) και τυπική απόκλιση ίση με τη μονάδα ( $\approx 1$ ), όπως απαιτείται από τη θεωρία.

Η στατιστική ανάλυση των τυποποιημένων δεδομένων αποκάλυψε ενδιαφέρουσα μεταβλητότητα μεταξύ των διαφόρων μεταβλητών. Οι μεταβλητές Z1 έως Z5, που αντιπροσωπεύουν τη συνολική κατανάλωση καυσίμου και τις εκπομπές αερίων θερμοκηπίου, εμφάνισαν εκτεταμένη κατανομή με μέγιστες τιμές που κυμαίνονται από 7,64 έως 53,61, υποδεικνύοντας σημαντική ποικιλομορφία στην περιβαλλοντική απόδοση των πλοίων. Αντίθετα, οι μεταβλητές Z7 έως Z10, που αφορούν σε δείκτες αποδοτικότητας ανά απόσταση και μεταφορικό έργο, εμφάνισαν πιο συμπυκνωμένη κατανομή με ελάχιστες αρνητικές τιμές, γεγονός που υποδεικνύει σχετικά ομοιογενή επίπεδα λειτουργικής αποδοτικότητας στο μεγαλύτερο μέρος του στόλου.

Το τελικό προεπεξεργασμένο σύνολο δεδομένων περιλαμβάνει 9.290 παρατηρήσεις με 13 τυποποιημένες αριθμητικές μεταβλητές, αντιπροσωπεύοντας το 70% του αρχικού συνόλου δεδομένων. Αυτή η μείωση, αν και σημαντική, εξασφαλίζει την ποιότητα και την αξιοπιστία των δεδομένων για τη μετέπειτα στατιστική ανάλυση, εξαλείφοντας παρατηρήσεις με ελλειπείς ή αναξιόπιστες τιμές που θα μπορούσαν να επηρεάσουν αρνητικά τα αποτελέσματα της Ανάλυσης Κύριων Συνιστωσών και της ομαδοποίησης.

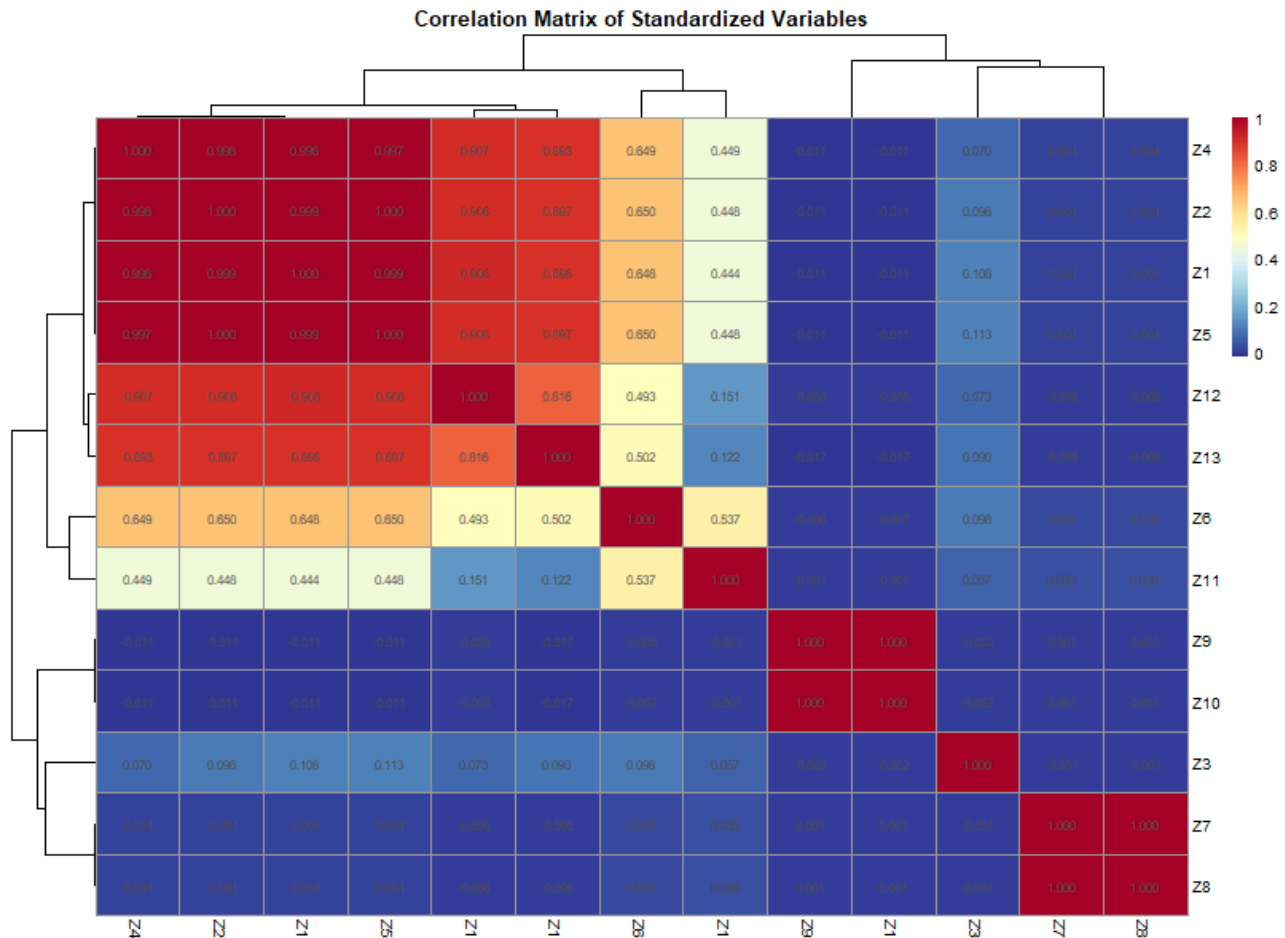
#### 4.4 ΔΙΕΡΕΥΝΗΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Μετά την προεπεξεργασία των δεδομένων, πραγματοποιήθηκε βασική διερευνητική ανάλυση για τον προσδιορισμό αρχικά της συσχέτισης μεταξύ των μεταβλητών και τον εντοπισμό πιθανών ακραίων τιμών. Και οι δύο αυτές αναλύσεις είναι κρίσιμες για την PCA καθώς όχι μόνο δικαιολογούν την εφαρμογή της λόγω της παρουσίας μεταβλητών με υψηλή συσχέτιση, αλλά επίσης οδηγούν σε πιο συγκεκριμένες κύριες συνιστώσες αφαιρώντας οποιαδήποτε ακραία τιμή που ενσωματώνει θόρυβο στο πολυμεταβλητό σύνολο δεδομένων.

Ένα από τα πιο αξιοσημείωτα σημεία από την ανάλυση συσχέτισης ήταν η θετική συσχέτιση μεταξύ της κατανάλωσης καυσίμου (Z1) με τις μεταβλητές που σχετίζονται με περιβαλλοντικά βλαβερές εκπομπές (Z2, Z4, Z5). Συγκεκριμένα, η συσχέτιση μεταξύ της συνολικής κατανάλωσης καυσίμου και των συνολικών εκπομπών CO<sub>2</sub> (Z2) έφτασε το εξαιρετικά

υψηλό επίπεδο του 0,999, ενώ παρόμοια υψηλές συσχετίσεις παρατηρήθηκαν με τις εκπομπές N<sub>2</sub>O (Z4: 0,996) και τις συνολικές εκπομπές CO<sub>2</sub>eq (Z5: 0,999). Αυτές οι σχεδόν τέλειες συσχετίσεις επιβεβαιώνουν τη θεωρητικά αναμενόμενη άμεση σχέση μεταξύ της κατανάλωσης καυσίμου και των εκπομπών αερίων θερμοκηπίου στη ναυτιλία.

**Πίνακας 4.3:** Πίνακας συσχέτισης των κανονικοποιημένων μεταβλητών.



Επιπλέον, υπήρχε ισχυρότερη θετική συσχέτιση μεταξύ των μεταβλητών που σχετίζονται με τα περιφερειακά πρότυπα λειτουργίας των πλοίων. Η συνολική κατανάλωση καυσίμου (Z1) παρουσίασε υψηλή συσχέτιση με τις εκπομπές CO<sub>2</sub> από ταξίδια που αναχώρησαν από λιμάνια κρατών μελών (Z12: 0,908) και προς λιμάνια κρατών μελών (Z13: 0,898). Παρόμοια, η συσχέτιση μεταξύ των εκπομπών από ταξίδια που αναχώρησαν και προς λιμάνια κρατών μελών (Z12-Z13) ήταν 0,816, υποδεικνύοντας συνεπή λειτουργικά πρότυπα εντός της δικαιοδοσίας της ΕΕ. Αξίζει επίσης να σημειωθεί ότι η συνολική κατανάλωση καυσίμου (Z1) συσχετίστηκε θετικά με τον

συνολικό χρόνο στη θάλασσα (Z6: 0,648), γεγονός που σχετίζεται με το συνολικό χρόνο που διανύθηκε στη θάλασσα κατά τη διάρκεια ενός ταξιδιού.

Από την άλλη πλευρά, παρατηρήθηκαν τέλειες συσχετίσεις μεταξύ συγκεκριμένων ζευγών δεικτών αποδοτικότητας. Η κατανάλωση καυσίμου ανά απόσταση (Z7) και οι εκπομπές CO<sub>2</sub> ανά απόσταση (Z8) παρουσίασαν συσχέτιση 1,000, όπως επίσης και η κατανάλωση καυσίμου ανά μεταφορικό έργο (Z9) με τις εκπομπές CO<sub>2</sub> ανά μεταφορικό έργο (Z10). Αυτές οι τέλειες συσχετίσεις αποτελούν αναμενόμενο αποτέλεσμα καθώς οι μεταβλητές αυτές υπολογίζονται με βάση τους ίδιους παρονομαστές και παρόμοιους αριθμητές.

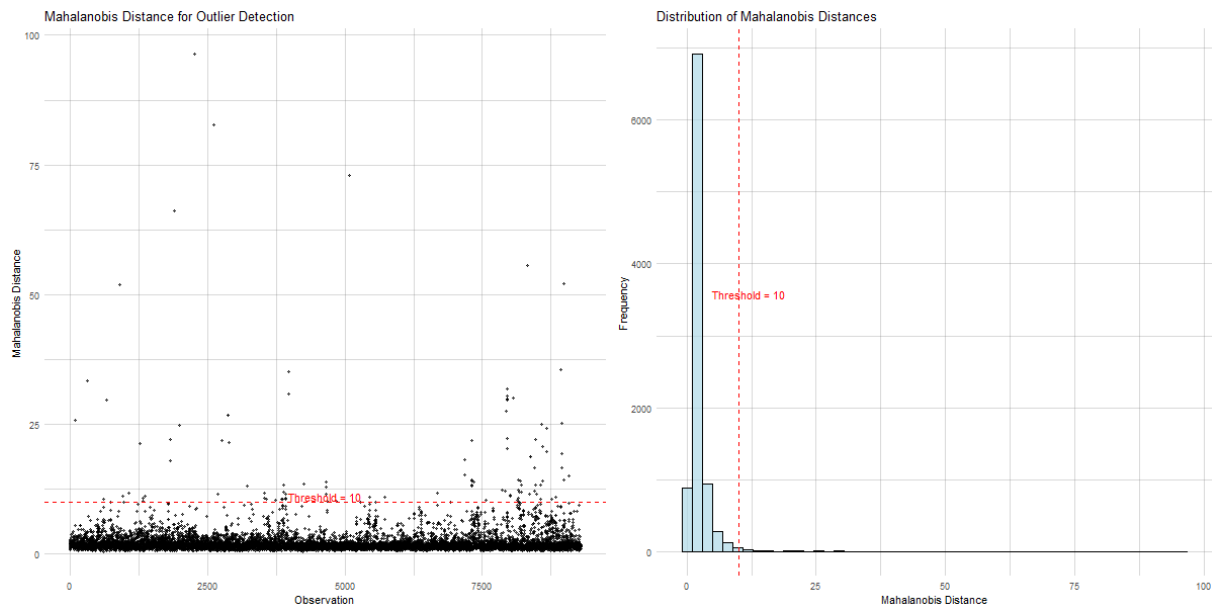
Για τον εντοπισμό πιθανών ακραίων τιμών στο σύνολο δεδομένων, οι μεταβλητές Z1-Z13 εξετάστηκαν ακολουθώντας δύο διαφορετικές προσεγγίσεις: μια μονομεταβλητή γραφική εξέταση και μια πολυμεταβλητή προσέγγιση βασισμένη στις αποστάσεις Mahalanobis (MD) των παρατηρήσεων από το κέντρο (μέσο όρο) της κοινής τους κατανομής. Η απόσταση Mahalanobis (Εξ. 4.2) υπολογίζεται ως:

$$MD = \sqrt{(u - \mu)^T V^{-1} (u - \mu)} \quad (4.2)$$

όπου  $u$  είναι ένας διάνυσμα των παρατηρημένων τιμών,  $\mu$  είναι ο διάνυσμα με τον δειγματικό μέσο όρο των μεταβλητών, και  $V$  είναι ο πίνακας διακύμανσης-συνδιακύμανσης.

Όπως παρατηρείται από το Διάγραμμα 4.1, τα περισσότερα δεδομένα μας δίνουν τιμή MD μικρότερη από 10, ενώ ένας μικρός αριθμός πλοίων δίνει τιμές MD μεγαλύτερες από 10. Από τη γραφική αναπαράσταση, τέθηκε ένα διαισθητικό όριο για τις εντός ορίων τιμές ίσο με 10 και όλες οι παρατηρήσεις με τιμές MD μεγαλύτερες από 10 αφαιρέθηκαν από την περαιτέρω ανάλυση, λόγω πιθανώς εσφαλμένων αναφερόμενων τιμών.

Συγκεκριμένα, εντοπίστηκαν 101 παρατηρήσεις (1,09% του συνόλου) με απόσταση Mahalanobis μεγαλύτερη από 10. Στις πέντε πιο ακραίες περιπτώσεις περιλαμβάνονται το πλοίο ANDROMEDA J (IMO: 9355422), ένα container ship με απόσταση Mahalanobis 96,37 και κατανάλωση καυσίμου 4.860,4 τόνους, το CIUDAD DE CADIZ (IMO: 9383560), ένα Ro-ro ship με απόσταση 82,66, το BEAM (IMO: 9591741), ένα Bulk carrier με απόσταση 73,05, το MORNING CORNET (IMO: 9329473), ένα Vehicle carrier με απόσταση 66,23, και το ADAMASTOS (IMO: 9879698), ένα Gas carrier με την εξαιρετικά υψηλή κατανάλωση καυσίμου των 23.462,4 τόνων και εκπομπές CO<sub>2</sub> 63.630,1 τόνων, που αντιστοιχεί σε απόσταση Mahalanobis 55,64.



**Διάγραμμα 4.1:** Διάγραμμα των αποστάσεων Mahalanobis των παρατηρήσεων.

Η αφαίρεση των ακραίων τιμών είχε σημαντικό αντίκτυπο στη δομή των συσχετίσεων του συνόλου δεδομένων, με τη μέγιστη αλλαγή συσχέτισης να φτάνει το 0,6653. Αυτή η σημαντική μεταβολή υποδεικνύει ότι οι ακραίες τιμές επηρέαζαν σημαντικά τις στατιστικές σχέσεις μεταξύ των μεταβλητών. Το τελικό μειωμένο σύνολο δεδομένων περιλαμβάνει 9.189 παρατηρήσεις με 13 μεταβλητές, αντιπροσωπεύοντας το 98,91% του καθαρισμένου συνόλου δεδομένων.

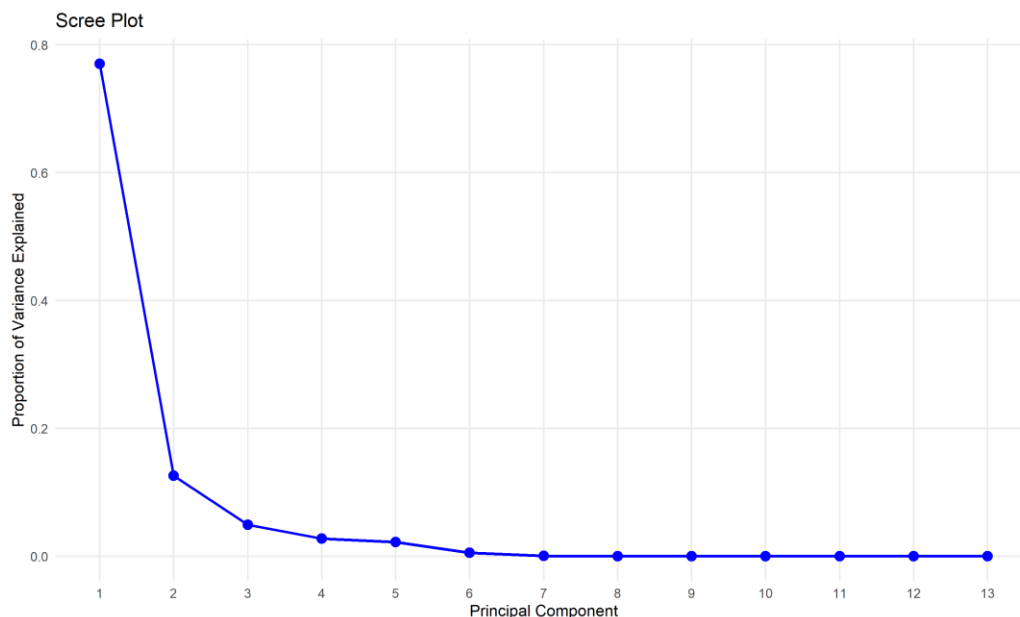
Η ανάλυση της κατανομής των τύπων πλοίων στο τελικό σύνολο δεδομένων αποκαλύπτει μια ποικιλόμορφη σύνθεση στόλου. Τα Bulk carriers αποτελούν τη μεγαλύτερη κατηγορία με 2.919 παρατηρήσεις (31,8%), ακολουθούμενα από τα Container ships με 1.909 παρατηρήσεις (20,8%) και τα Oil tankers με 1.725 παρατηρήσεις (18,8%). Τα Chemical tankers αντιπροσωπεύουν το 13,1% του δείγματος με 1.206 παρατηρήσεις, ενώ οι υπόλοιπες κατηγορίες πλοίων (Vehicle carriers, General cargo ships, Gas carriers, Ro-ro ships και άλλα) αντιπροσωπεύουν μικρότερα ποσοστά του συνολικού δείγματος.

#### 4.5 ΕΦΑΡΜΟΓΗ ΤΗΣ PCA

Η PCA εφαρμόστηκε διαδοχικά σε (α) όλα τα διαθέσιμα δεδομένα και (β) στα δεδομένα μετά την αφαίρεση παρατηρήσεων (πλοίων) με τιμή MD μεγαλύτερη από 10 (μειωμένο σύνολο δεδομένων). Ο αριθμός των κύριων συνιστωσών που πρέπει να διατηρηθούν σε μια PCA

καθορίζεται συνήθως με τη βοήθεια του λεγόμενου scree plot. Από το scree plot που δημιουργήθηκε με βάση όλα τα δεδομένα, μπορεί κανείς να συμπεράνει ότι τρεις κύριες συνιστώσες θα πρέπει να διατηρηθούν καθώς η καμπύλη μειώνεται απότομα και στη συνέχεια κάμπτεται όταν ο αριθμός των PC ισούται με 3, γεγονός που χρησιμεύει ως ένδειξη σημείου αποκοπής.

Από την ανάλυση του scree plot προκύπτει ότι η πρώτη κύρια συνιστώσα (PC1) εξηγεί το 77% της συνολικής διακύμανσης των δεδομένων, η δεύτερη κύρια συνιστώσα (PC2) εξηγεί περίπου το 13% της διακύμανσης, ενώ η τρίτη κύρια συνιστώσα (PC3) συνεισφέρει περίπου το 3% της συνολικής διακύμανσης. Οι τρεις πρώτες κύριες συνιστώσες εξηγούν συνολικά περισσότερο από το 90% της παρατηρούμενης διακύμανσης στις αρχικές μεταβλητές, γεγονός που επιβεβαιώνει την καταλληλότητα της επιλογής διατήρησης τριών συνιστωσών.



**Διάγραμμα 4.2:** Scree plot για τον προσδιορισμό του αριθμού κύριων συνιστωσών που πρέπει να διατηρηθούν. Η απότομη κάμψη στην τρίτη συνιστώσα υποδεικνύει τη διατήρηση τριών κύριων συνιστωσών.

Η σύγκριση των loadings μεταξύ του πλήρους και του μειωμένου συνόλου δεδομένων αποκαλύπτει σημαντική συνοχή στη δομή των κύριων συνιστωσών. Τα loadings των πρώτων 3 PC και στις δύο περιπτώσεις είναι παρόμοια, υποδεικνύοντας ότι η αφαίρεση των ακραίων τιμών δεν επηρέασε σημαντικά την υποκείμενη δομή των δεδομένων. Αυτή η σταθερότητα των αποτελεσμάτων δικαιολογεί τη χρήση του μειωμένου συνόλου δεδομένων για την τελική ανάλυση.

Από τα loadings των κύριων συνιστωσών προκύπτει ότι η πρώτη κύρια συνιστώσα (PC1) χαρακτηρίζεται από υψηλά θετικά loadings σε όλες σχεδόν τις μεταβλητές που αφορούν στην

κατανάλωση καυσίμου και τις εκπομπές αερίων θερμοκηπίου (Z1, Z2, Z3, Z4, Z5). Συγκεκριμένα, οι μεταβλητές της συνολικής κατανάλωσης καυσίμου, των συνολικών εκπομπών CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O και CO<sub>2</sub>eq εμφανίζουν τα υψηλότερα loadings στην PC1, υποδεικνύοντας ότι αυτή η συνιστώσα αντιπροσωπεύει την "Συνολική Περιβαλλοντική Επίδραση" των πλοίων. Επιπλέον, η PC1 παρουσιάζει θετική συσχέτιση με τον χρόνο στη θάλασσα (Z6) και τις περιφερειακές εκπομπές (Z12, Z13).

**Πίνακας 4.4:** Loadings των πρώτων τριών κύριων συνιστωσών (PC1–PC3) για το αρχικό και το μειωμένο σύνολο δεδομένων, καθώς και οι διαφορές μεταξύ τους. Ο πίνακας επιβεβαιώνει τη σταθερότητα της δομής των συνιστωσών μετά την αφαίρεση ακραίων τιμών.

Μεταβλητή	Αρχικό dataset μετά από κανονικοποίηση			Μειωμένο dataset			Ζεύγη διαφορών μεταξύ φορτίων		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Z1	0,394	-0,038	-0,11	0,397	0,001	-0,005	-0,003	-0,039	-0,105
Z2	0,396	-0,036	-0,11	0,397	0,001	-0,005	-0,001	-0,038	-0,105
Z3	0,017	0,032	0,06	0,047	0,001	0,001	-0,03	0,031	0,06
Z4	0,4	-0,04	-0,115	0,396	0,001	-0,005	0,003	-0,041	-0,111
Z5	0,396	-0,036	-0,109	0,397	0,001	-0,005	-0,001	-0,037	-0,104
Z6	0,285	0,5	0,812	0,281	-0,021	0,007	0,004	0,522	0,805
Z7	0	0	0	0,003	-0,611	0,354	-0,003	0,611	-0,354
Z8	0	0	0	0,003	-0,611	0,354	-0,003	0,611	-0,354
Z9	-0,002	0,011	0,003	-0,006	-0,354	-0,612	0,005	0,365	0,615
Z10	-0,002	0,011	0,003	-0,007	-0,354	-0,612	0,005	0,365	0,615
Z11	0,173	0,752	-0,525	0,182	-0,041	0,016	-0,009	0,794	-0,541
Z12	0,362	-0,289	-0,013	0,359	0,012	-0,014	0,003	-0,301	0,001
Z13	0,358	-0,305	0,109	0,356	0,017	-0,006	0,002	-0,323	0,115
Διακύμανση που εξηγείται	0,77	0,126	0,049	0,484	0,154	0,154	0,286	-0,028	-0,105



Η δεύτερα κύρια συνιστώσα (PC2) εμφανίζει διαφορετικό πρότυπο loadings, με υψηλότερα loadings στις μεταβλητές που αφορούν σε δείκτες αποδοτικότητας ανά απόσταση και μεταφορικό έργο (Z7, Z8, Z9, Z10). Αυτή η συνιστώσα φαίνεται να αντιπροσωπεύει την "Λειτουργική Αποδοτικότητα" των πλοίων, καθώς εστιάζει στην αποδοτικότητα κατανάλωσης καυσίμου και εκπομπών σε σχέση με την απόσταση και το μεταφορικό έργο που εκτελείται.

Η τρίτη κύρια συνιστώσα (PC3) παρουσιάζει μεικτό πρότυπο loadings με έμφαση στις περιφερειακές μεταβλητές εκπομπών (Z11, Z12, Z13), υποδεικνύοντας ότι αυτή η συνιστώσα σχετίζεται με τα "Περιφερειακά Πρότυπα Λειτουργίας" των πλοίων εντός και εκτός της δικαιοδοσίας των κρατών μελών της ΕΕ.

Το biplot των πρώτων δύο κύριων συνιστωσών επιβεβαιώνει τις παραπάνω ερμηνείες, δείχνοντας ξεκάθαρο διαχωρισμό μεταξύ των διαφορετικών τύπων μεταβλητών. Οι μεταβλητές συνολικών εκπομπών και κατανάλωσης καυσίμου (Z1-Z5) εμφανίζονται ομαδοποιημένες στην κατεύθυνση της PC1, ενώ οι μεταβλητές αποδοτικότητας (Z7-Z10) τείνουν προς διαφορετική κατεύθυνση. Η οπτικοποίηση των πλοίων με βάση τον τύπο τους στο biplot αποκαλύπτει ενδιαφέρουσα διασπορά, με ορισμένους τύπους πλοίων να εμφανίζουν πιο συμπυκνωμένη κατανομή ενώ άλλοι παρουσιάζουν μεγαλύτερη ποικιλομορφία στις δύο πρώτες κύριες συνιστώσες.



**Διάγραμμα 4.3:** Biplot των δύο πρώτων κύριων συνιστωσών. Φαίνεται ομαδοποίηση των μεταβλητών κατανάλωσης καυσίμου και εκπομπών στην κατεύθυνση της PC1, ενώ οι δείκτες αποδοτικότητας κατευθύνονται προς την PC2.

Οι τρεις κύριες συνιστώσες μπορούν να θεωρηθούν ως τρεις ανεξάρτητοι δείκτες που χαρακτηρίζουν τη συνολική περιβαλλοντική επίδραση, τη λειτουργική αποδοτικότητα, και τα περιφερειακά πρότυπα λειτουργίας των πλοίων αντίστοιχα. Αυτή η ερμηνεία παρέχει μια ολοκληρωμένη προσέγγιση για την αξιολόγηση της περιβαλλοντικής απόδοσης των πλοίων, λαμβάνοντας υπόψη τόσο την απόλυτη περιβαλλοντική επίδραση όσο και την αποδοτικότητα των λειτουργιών τους.

#### **4.6 ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΩΝ ΒΑΣΙΣΜΕΝΗ ΣΤΟΥΣ ΤΡΕΙΣ ΔΕΙΚΤΕΣ**

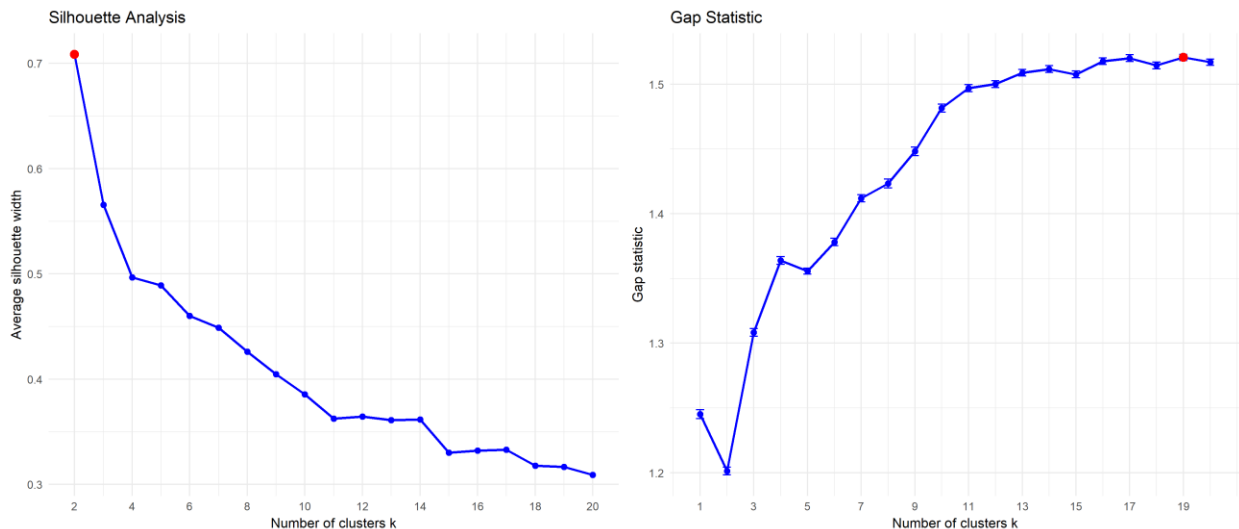
Μετά την εξαγωγή των τριών δεικτών που σχετίζονται με την περιβαλλοντική επίδραση των πλοίων, οι οποίοι προέκυψαν από τον συνδυασμό χρόνου λειτουργίας, εκπομπών και λειτουργικών χαρακτηριστικών, εφαρμόστηκε ο αλγόριθμος K-Means για την περαιτέρω διερεύνηση των δεδομένων. Συγκεκριμένα, ο αλγόριθμος υλοποιήθηκε με βάση τους τρεις δείκτες που προέκυψαν από την PCA, με στόχο την ανίχνευση και ομαδοποίηση πλοίων σε συστάδες με παρόμοια χαρακτηριστικά ως προς την περιβαλλοντική επίδραση, τη λειτουργική αποδοτικότητα και τα περιφερειακά πρότυπα λειτουργίας.

##### **4.6.1 ΕΠΙΛΟΓΗ ΤΟΥ ΒΕΛΤΙΣΤΟΥ ΑΡΙΘΜΟΥ ΣΥΣΤΑΔΩΝ**

Ο καθορισμός του αριθμού  $k$  των συστάδων αποτελεί μία από τις πιο κρίσιμες διαδικασίες στην ανάλυση συστάδων. Υπάρχουν πολλές μέθοδοι για αυτόν τον σκοπό, καθεμία εκ των οποίων εξετάζει διαφορετικές πτυχές των δεδομένων, με αποτέλεσμα να μην οδηγούν πάντα στον ίδιο αριθμό συστάδων. Σε τέτοιες περιπτώσεις, η ανάλυση πρέπει να πραγματοποιείται για διαφορετικά σενάρια και τα αποτελέσματα να συνδυάζονται με την εμπειρική γνώση ενός ειδικού, ώστε να προκύπτει όχι μόνο μια στατιστικά έγκυρη αλλά και μια ουσιαστικά ρεαλιστική επιλογή.

Στην παρούσα μελέτη χρησιμοποιήθηκαν δύο από τις πιο διαδεδομένες μεθόδους: το σκορ silhouette και η στατιστική Gap, προκειμένου να καθοριστεί ο βέλτιστος αριθμός συστάδων  $k$  στο σύνολο δεδομένων των κύριων συνιστωσών. Και οι δύο στατιστικές υπολογίστηκαν για μια σειρά τιμών του  $k$ . Υψηλές τιμές ή τοπικές κορυφώσεις στις αντίστοιχες γραφικές παραστάσεις υποδεικνύουν καλύτερη συνοχή εντός των συστάδων και σαφέστερο διαχωρισμό από γειτονικές συστάδες.

Η ανάλυση silhouette έδειξε ότι η βέλτιστη τιμή εμφανίζεται για  $k = 2$  με σκορ περίπου 0,71, στοιχείο που υποδηλώνει εξαιρετικά καλή συνοχή και διαχωρισμό. Το σκορ μειώνεται σταδιακά όσο αυξάνεται ο αριθμός συστάδων, με σχετικά ικανοποιητικές τιμές να παρατηρούνται για  $k = 4$  ( $\approx 0,49$ ) και  $k = 6$  ( $\approx 0,45$ ). Μετά το  $k = 12$ , οι τιμές σταθεροποιούνται σε χαμηλότερα επίπεδα ( $\approx 0,31$ – $0,36$ ), γεγονός που δείχνει ότι η αύξηση των συστάδων δεν προσφέρει ουσιαστική βελτίωση.



**Διάγραμμα 4.4:** Ανάλυση Silhouette για διαφορετικό αριθμό συστάδων (αριστερά) και Gap Statistic για διαφορετικό αριθμό συστάδων (δεξιά).

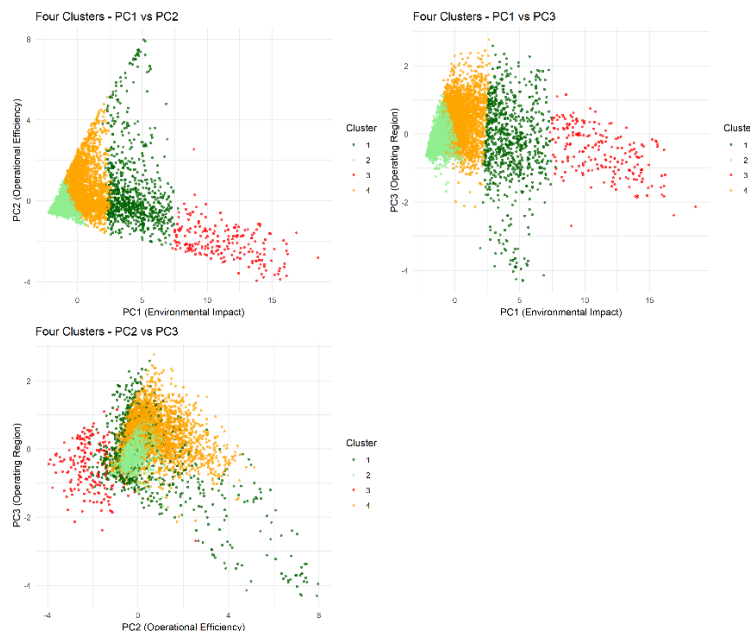
Αντίθετα, η στατιστική Gap ακολουθεί διαφορετική πορεία, με τη μέγιστη τιμή να παρατηρείται στο  $k = 18$  ( $\approx 1,56$ ), ακολουθούμενη από το  $k = 17$  ( $\approx 1,55$ ). Η αύξηση της Gap είναι σταδιακή από  $k = 3$  έως  $k = 10$ , όπου σημειώνεται σημαντική άνοδος, ενώ στη συνέχεια συνεχίζει ανοδικά με πιο ήπιο ρυθμό μέχρι το μέγιστο. Είναι αξιοσημείωτη η χαμηλή τιμή για  $k = 2$ , ένδειξη ότι η μέθοδος αυτή δεν υποστηρίζει τη λύση των δύο συστάδων.

Η απόκλιση μεταξύ των δύο μεθόδων καταδεικνύει ότι η βέλτιστη επιλογή του αριθμού συστάδων εξαρτάται από τα κριτήρια αξιολόγησης και τους στόχους της ανάλυσης. Το silhouette προκρίνει λιγότερες συστάδες με ισχυρή εσωτερική συνοχή, ενώ η Gap τείνει να αναδεικνύει περισσότερες συστάδες που καταγράφουν λεπτότερες διαφοροποιήσεις στα δεδομένα. Για τους σκοπούς της παρούσας ανάλυσης εξετάστηκαν δύο ενδιάμεσες λύσεις, με  $k = 4$  και  $k = 14$  συστάδες, οι οποίες προσφέρουν ισορροπία ανάμεσα στην απλότητα ερμηνείας και την ικανότητα ανίχνευσης πιο σύνθετων προτύπων.

#### 4.6.2 ΑΛΓΟΡΙΘΜΟΣ K-MEANS

Ο αλγόριθμος k-means εφαρμόστηκε στο σύνολο δεδομένων ορίζοντας τον αριθμό των συστάδων ίσο με 4 και 14. Η ανάλυση με τέσσερις συστάδες αποκάλυψε σημαντικές διαφορές στο μέγεθος των ομάδων, με τη δεύτερη συστάδα να περιλαμβάνει το μεγαλύτερο αριθμό πλοίων (5.616 παρατηρήσεις, 61% του συνόλου), ακολουθούμενη από την τέταρτη συστάδα (2.596 παρατηρήσεις, 28%), την πρώτη συστάδα (776 παρατηρήσεις, 8%), και την τρίτη συστάδα (201 παρατηρήσεις, 2%). Το συνολικό άθροισμα τετραγώνων εντός των συστάδων για  $k=4$  ήταν 14.757,32, υποδεικνύοντας την ποιότητα της ομαδοποίησης.

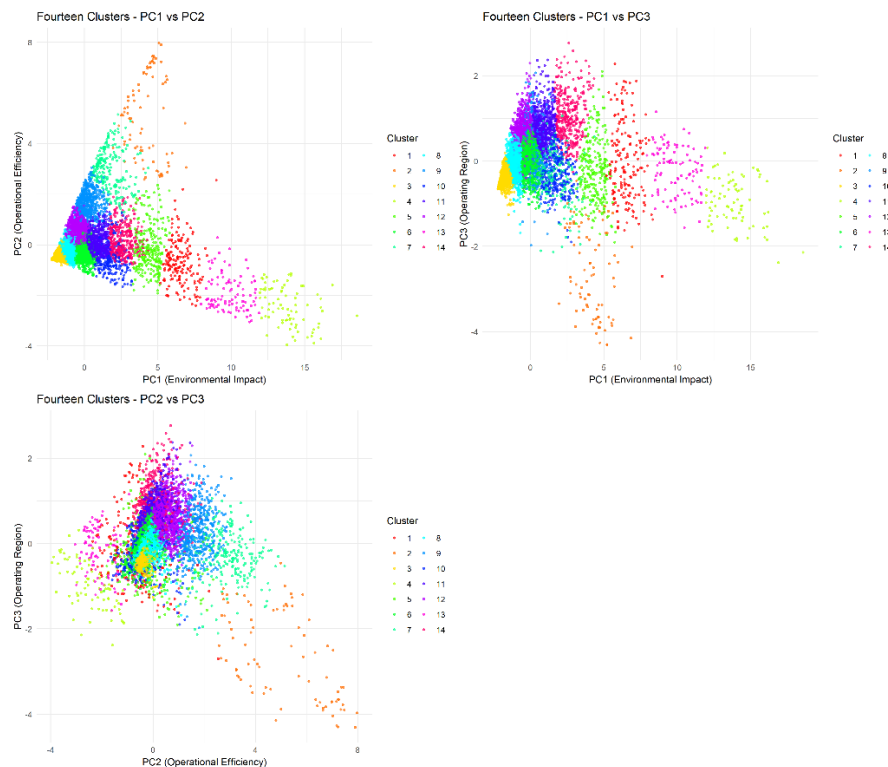
Η ανάλυση με δεκατέσσερις συστάδες παρουσίασε μεγαλύτερη ποικιλομορφία στο μέγεθος των συστάδων, με τη μεγαλύτερη ομάδα (συστάδα 3) να περιλαμβάνει 2.532 πλοία και τη μικρότερη (συστάδα 2) μόλις 82 πλοία. Άλλες σημαντικές συστάδες περιλαμβάνουν τη συστάδα 8 με 2.245 πλοία και τη συστάδα 6 με 943 πλοία. Το συνολικό άθροισμα τετραγώνων εντός των συστάδων για  $k=14$  μειώθηκε σημαντικά σε 4.363,72, επιβεβαιώνοντας ότι η αύξηση του αριθμού των συστάδων βελτιώνει την εσωτερική ομοιογένεια των ομάδων.



**Διάγραμμα 4.5:** Απεικόνιση των τεσσάρων συστάδων στο χώρο των κύριων συνιστωσών. (α) PC1–PC2: διαχωρισμός ως προς την περιβαλλοντική επίδραση και την επιχειρησιακή αποδοτικότητα. (β) PC1–PC3: διαφοροποίηση βάσει περιβαλλοντικής επίδρασης και περιφερειακής λειτουργίας. (γ) PC2–PC3: ανάδειξη διαφορών αποδοτικότητας και περιφερειακών προτύπων λειτουργίας. Οι διαφορετικά χρωματισμένες ομάδες αντιπροσωπεύουν τις συστάδες που προέκυψαν από την ανάλυση.

Από τις γραφικές παραστάσεις προκύπτει ότι ο αλγόριθμος ομαδοποίησης κατατάσσει το σύνολο δεδομένων σε διακριτές ενότητες ως προς τις τρεις κύριες συνιστώσες. Στην ανάλυση με τέσσερις συστάδες, οι συστάδες στο επίπεδο που ορίζεται από την πρώτη και δεύτερη κύρια συνιστώσα παρουσιάζουν καλό διαχωρισμό, με τη συστάδα 1 (πράσινο) να κατανέμεται κυρίως στις υψηλές τιμές της PC1, τη συστάδα 2 (ανοιχτό πράσινο) να καλύπτει το κεντρικό τμήμα, τη συστάδα 3 (κόκκινο) να εμφανίζεται στις πολύ υψηλές τιμές της PC1 με χαμηλές τιμές PC2, και τη συστάδα 4 (πορτοκαλί) να εντοπίζεται σε χαμηλές τιμές PC1 με θετικές τιμές PC2.

Η ανάλυση με δεκατέσσερις συστάδες αποκαλύπτει πιο λεπτομερή πρότυπα διαχωρισμού, με κάθε συστάδα να καταλαμβάνει πιο συγκεκριμένες περιοχές στον τρισδιάστατο χώρο των κύριων συνιστωσών. Η περίπλοκη κατανομή των χρωμάτων υποδεικνύει ότι ο αλγόριθμος εντοπίζει πιο εξειδικευμένα υπο-πρότυπα εντός του συνόλου δεδομένων, που μπορεί να αντιστοιχούν σε διαφορετικούς συνδυασμούς περιβαλλοντικής επίδρασης, λειτουργικής αποδοτικότητας και περιφερειακών προτύπων λειτουργίας.

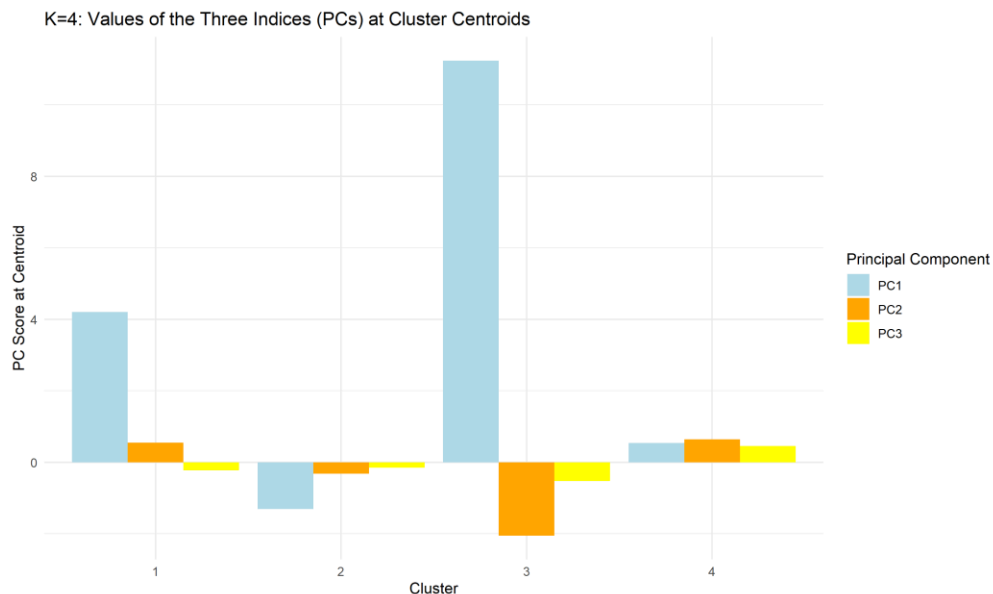


**Διάγραμμα 4.6:** Απεικόνιση των δεκατεσσάρων συστάδων στο χώρο των κύριων συνιστωσών. (α) PC1–PC2: διαφοροποίηση με βάση την περιβαλλοντική επίδραση και την επιχειρησιακή αποδοτικότητα. (β) PC1–PC3: ανάλυση σε σχέση με την περιβαλλοντική επίδραση και την περιφερειακή λειτουργία. (γ) PC2–PC3: εστίαση στην αποδοτικότητα και τα περιφερειακά πρότυπα λειτουργίας. Οι χρωματισμοί απεικονίζουν τις διαφορετικές συστάδες που προέκυψαν από την ανάλυση.

#### 4.6.3 ΕΡΜΗΝΕΙΑ ΤΩΝ ΣΥΣΤΑΔΩΝ

Οι διαφορές μεταξύ των τεσσάρων συστάδων αναδεικνύονται καθαρά από τη γραφική παράσταση των τιμών των τριών δεικτών (PC) στα κεντροειδή των τεσσάρων συστάδων που προσδιορίστηκαν από τον αλγόριθμο k-means. Η ανάλυση των κεντροειδών αποκαλύπτει διακριτά προφίλ περιβαλλοντικής απόδοσης για κάθε ομάδα πλοίων.

Η 1<sup>η</sup> συστάδα χαρακτηρίζεται από υψηλές τιμές της πρώτης κύριας συνιστώσας ( $PC1 = 4,205$ ), υποδεικνύοντας σημαντική περιβαλλοντική επίδραση, συνδυασμένη με θετικές τιμές της δεύτερης κύριας συνιστώσας ( $PC2 = 0,545$ ), που αντιπροσωπεύουν σχετικά καλή λειτουργική αποδοτικότητα. Η τρίτη κύρια συνιστώσα εμφανίζει ελαφρώς αρνητική τιμή ( $PC3 = -0,229$ ), υποδηλώνοντας κυρίως λειτουργίες εκτός της δικαιοδοσίας κρατών μελών.



**Διάγραμμα 4.7:** Τιμές των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) στα κεντροειδή των τεσσάρων συστάδων.

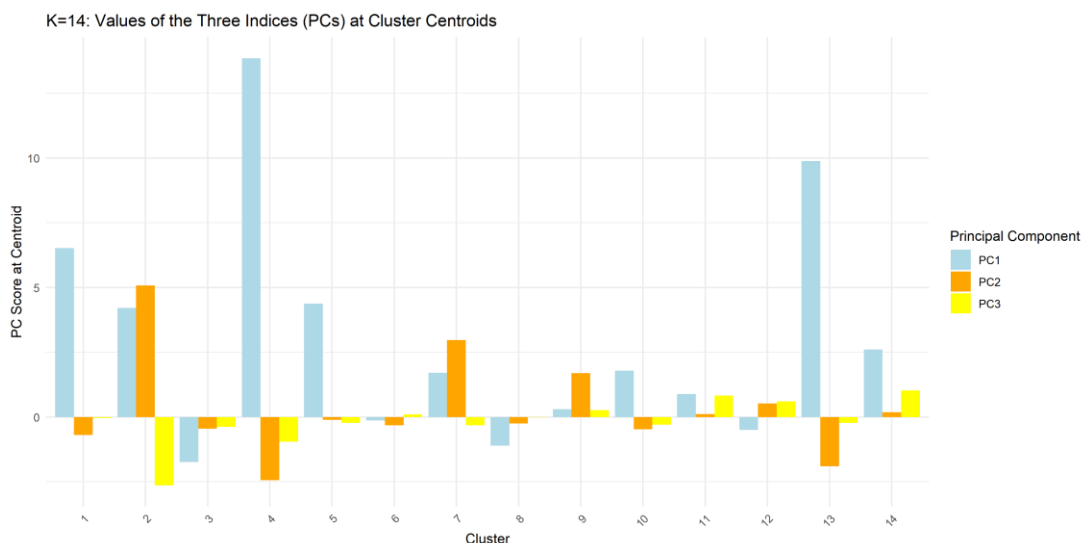
Η 2<sup>η</sup> συστάδα, που αποτελεί τη μεγαλύτερη ομάδα με 5.616 πλοία, παρουσιάζει αρνητικές τιμές τόσο στην PC1 (-1,301) όσο και στην PC2 (-0,318), χαρακτηρίζοντας πλοία με χαμηλή περιβαλλοντική επίδραση αλλά επίσης περιορισμένη λειτουργική αποδοτικότητα. Αυτή η συστάδα πιθανότατα περιλαμβάνει μικρότερα ή λιγότερο ενεργά πλοία που λειτουργούν με χαμηλότερους ρυθμούς κατανάλωσης καυσίμου και εκπομπών.

Η 3<sup>η</sup> συστάδα εμφανίζει την πιο ακραία τιμή στην PC1 (11,235), αντιπροσωπεύοντας πλοία με εξαιρετικά υψηλή περιβαλλοντική επίδραση, ενώ ταυτόχρονα παρουσιάζει σημαντικά αρνητική

τιμή στην PC2 (-2,050), υποδεικνύοντας χαμηλή λειτουργική αποδοτικότητα. Αυτή η συστάδα, αν και μικρή (201 πλοία), αντιπροσωπεύει τα πλοία με τη χειρότερη συνολική περιβαλλοντική απόδοση.

Η 4<sup>η</sup> συστάδα χαρακτηρίζεται από μέτριες τιμές PC1 (0,536) και θετικές τιμές PC2 (0,645), υποδεικνύοντας πλοία με σχετικά χαμηλή περιβαλλοντική επίδραση αλλά καλή λειτουργική αποδοτικότητα. Η υψηλότερη τιμή PC3 (0,458) από όλες τις συστάδες υποδηλώνει ότι αυτά τα πλοία λειτουργούν κυρίως εντός της δικαιοδοσίας κρατών μελών της ΕΕ.

Η ανάλυση με δεκατέσσερις συστάδες παρέχει πιο λεπτομερή διάκριση των προτύπων. Ορισμένες συστάδες, όπως η 4<sup>η</sup> συστάδα (PC1 = 13,848), αναδεικνύουν ακόμη πιο ακραίες περιπτώσεις περιβαλλοντικής επίδρασης, ενώ η Συστάδα 2 (PC2 = 5,077) εμφανίζει εξαιρετικά υψηλή λειτουργική αποδοτικότητα. Αυτή η λεπτομερής διάκριση επιτρέπει τον εντοπισμό υποομάδων που πιθανόν αντιστοιχούν σε συγκεκριμένους τύπους πλοίων ή λειτουργικά σενάρια.



**Διάγραμμα 4.8:** Τιμές των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) στα κεντροειδή των δεκατεσσάρων συστάδων.

Με βάση την ανάλυση των κεντροειδών και την κατανομή των τιμών των κύριων συνιστωσών, οι τέσσερις κύριες συστάδες μπορούν να χαρακτηριστούν ως εξής:

- 1<sup>η</sup> συστάδα: "Πλοία υψηλής περιβαλλοντικής επίδρασης με μέτρια αποδοτικότητα"
- 2<sup>η</sup> συστάδα: "Πλοία χαμηλής περιβαλλοντικής επίδρασης με περιορισμένη αποδοτικότητα"
- 3<sup>η</sup> συστάδα: "Πλοία εξαιρετικά υψηλής περιβαλλοντικής επίδρασης με χαμηλή αποδοτικότητα"

- 4<sup>η</sup> συστάδα: "Πλοία μέτριας περιβαλλοντικής επίδρασης με υψηλή λειτουργική αποδοτικότητα"

Αυτό το πλαίσιο ταξινόμησης αποτελεί ένα χρήσιμο εργαλείο αξιολόγησης της περιβαλλοντικής απόδοσης και μπορεί να στηρίζει τον σχεδιασμό στοχευμένων πολιτικών βελτίωσης για κάθε κατηγορία πλοίων.

#### 4.6.4 ΑΝΑΛΥΣΗ ΣΥΝΘΕΣΗΣ ΣΥΣΤΑΔΩΝ ΚΑΙ ΤΥΠΩΝ ΠΛΟΙΩΝ

Η ανάλυση της κατανομής των τύπων πλοίων μέσα στις συστάδες αναδεικνύει σαφή πρότυπα που ενισχύουν και εμπλουτίζουν την ερμηνεία των τριών βασικών δεικτών περιβαλλοντικής απόδοσης.

Η 1<sup>η</sup> συστάδα (776 πλοία) παρουσιάζει εξαιρετικά συγκεκριμένη σύνθεση με 549 Container ships (70,7%) και 97 Oil tankers (12,5%) να αποτελούν τη συντριπτική πλειοψηφία. Επιπλέον, περιλαμβάνει 109 Ro-ro ships (14%) και μικρότερους αριθμούς άλλων τύπων. Η υψηλή περιβαλλοντική επίδραση ( $PC1 = 4,205$ ) και η σχετικά καλή λειτουργική αποδοτικότητα ( $PC2 = 0,545$ ) χαρακτηρίζουν αυτήν την ομάδα ως πλοία υψηλής δραστηριότητας και μεγάλου μεγέθους.

Η 2<sup>η</sup> συστάδα αποτελεί τη μεγαλύτερη ομάδα (5.616 πλοία, 61% του συνόλου) με ποικιλόμορφη σύνθεση: 2.659 Bulk carriers (47,4%), 942 Oil tankers (16,8%), 719 Chemical tankers (12,8%) και 493 Container ships (8,8%). Αυτή η ετερογένεια αντικατοπτρίζεται στις χαμηλές τιμές τόσο της περιβαλλοντικής επίδρασης ( $PC1 = -1,301$ ) όσο και της λειτουργικής αποδοτικότητας ( $PC2 = -0,318$ ), υποδεικνύοντας πλοία με μέτρια λειτουργική ένταση.

Η 3<sup>η</sup> συστάδα εμφανίζει απόλυτη ομοιογένεια, περιλαμβάνοντας αποκλειστικά 201 Container ships. Αυτή η συστάδα παρουσιάζει την υψηλότερη περιβαλλοντική επίδραση ( $PC1 = 11,235$ ) και τη χειρότερη λειτουργική αποδοτικότητα ( $PC2 = -2,050$ ), υποδεικνύοντας πλοία εμπορευματοκιβωτίων με εξαιρετικά υψηλή κατανάλωση καυσίμου και εκπομπές.

Η 4<sup>η</sup> συστάδα (2.596 πλοία) παρουσιάζει επίσης ποικιλόμορφη σύνθεση με 686 Oil tankers (26,4%), 666 Container ships (25,7%), 487 Chemical tankers (18,8%) και 258 Bulk carriers (9,9%). Η καλή λειτουργική αποδοτικότητα ( $PC2 = 0,645$ ) και η υψηλότερη τιμή  $PC3$  (0,458) υποδηλώνουν πλοία που λειτουργούν αποδοτικά κυρίως εντός της δικαιοδοσίας κρατών μελών της ΕΕ.

Η ανάλυση των 14 συστάδων αποκαλύπτει πιο εξειδικευμένα πρότυπα. Ορισμένες συστάδες εμφανίζουν εξαιρετικά υψηλή ειδίκευση, όπως η 4<sup>η</sup> συστάδα με αποκλειστικά 80



Container ships και την ακραία περιβαλλοντική επίδραση ( $PC1 = 13,848$ ), ή η 13<sup>η</sup> συστάδα με 99 Container ships και  $PC1 = 9,884$ .

Αντίθετα, άλλες συστάδες παρουσιάζουν μεγαλύτερη ποικιλομορφία. Η 3<sup>η</sup> συστάδα (2.532 πλοία) περιλαμβάνει κυρίως Bulk carriers (1.348, 53,2%) και Oil tankers (352, 13,9%), ενώ παρουσιάζει τη χαμηλότερη περιβαλλοντική επίδραση ( $PC1 = -1,738$ ).

Ιδιαίτερο ενδιαφέρον παρουσιάζει η 2<sup>η</sup> συστάδα με μόλις 82 πλοία αλλά εξαιρετικά υψηλή λειτουργική αποδοτικότητα ( $PC2 = 5,077$ ), αποτελούμενη από 41 Container ships και 39 Ro-ro ships. Αυτή η συστάδα πιθανότατα αντιπροσωπεύει πλοία υψηλής τεχνολογίας ή ειδικών λειτουργιών.

Η ανάλυση αποκαλύπτει ότι τα Container ships κατανέμονται σε πολλαπλές συστάδες με διαφορετικά χαρακτηριστικά περιβαλλοντικής απόδοσης, υποδεικνύοντας σημαντική ποικιλομορφία εντός αυτής της κατηγορίας. Αντίθετα, τα Bulk carriers τείνουν να συγκεντρώνονται σε συστάδες με χαμηλότερη περιβαλλοντική επίδραση, ενώ ορισμένοι τύποι όπως τα Chemical tankers εμφανίζουν πιο ομοιόμορφη κατανομή.

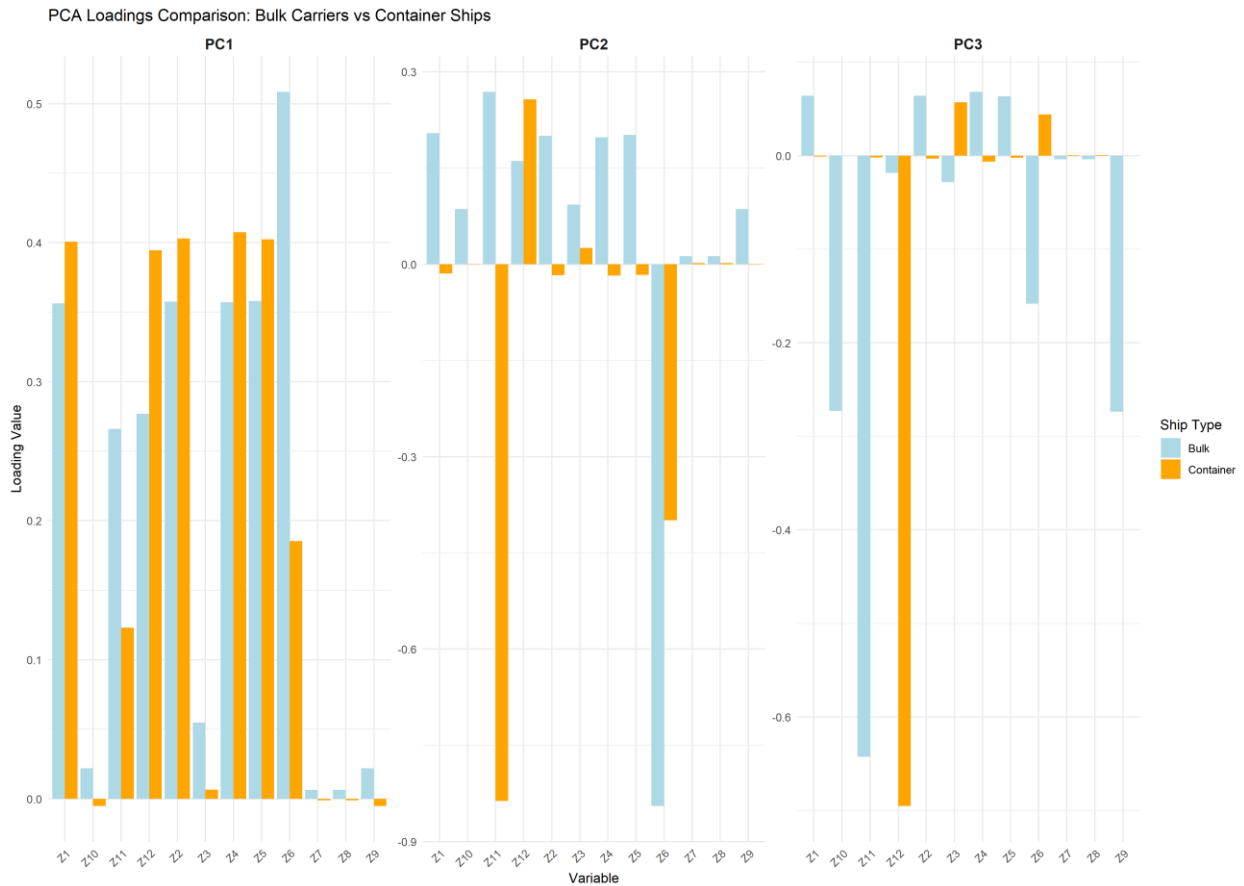
Αυτή η λεπτομερής ανάλυση της σύνθεσης των συστάδων παρέχει πολύτιμες γνώσεις για την ανάπτυξη στοχευμένων πολιτικών περιβαλλοντικής βελτίωσης, καθώς επιτρέπει τον εντοπισμό συγκεκριμένων τύπων πλοίων που χρήζουν προτεραιότητας ή ειδικών παρεμβάσεων για τη βελτίωση της περιβαλλοντικής τους απόδοσης.

#### 4.7 ΕΠΙΚΥΡΩΣΗ ΤΩΝ ΠΡΟΤΕΙΝΟΜΕΝΩΝ ΔΕΙΚΤΩΝ ΜΕ PCA

Για την επικύρωση της ποιότητας, αξιοπιστίας και συνέπειας των αποτελεσμάτων της PCA και των τριών δεικτών που δημιουργήθηκαν, χρησιμοποιήθηκαν ως ανεξάρτητα σύνολα τα 2.919 bulk carriers και τα 1.909 container ships του dataset. Η βασική διαφορά τους είναι ότι τα bulk carriers εξειδικεύονται στη μεταφορά χύδην φορτίων μεγάλου όγκου και βάρους, ενώ τα container ships στη μεταφορά τυποποιημένων εμπορευματοκιβωτίων. Έτσι, οι δύο κατηγορίες μπορούν να χρησιμεύσουν ως ισχυρά σύνολα επικύρωσης για την αξιολόγηση της σταθερότητας και της γενικευσιμότητας των δεικτών σε διαφορετικούς τύπους πλοίων.

Η εφαρμογή της PCA στους bulk carriers έδειξε ότι οι τρεις πρώτες κύριες συνιστώσες εξηγούν το 91,4% της συνολικής διακύμανσης (77,1% η πρώτη, 11,8% η δεύτερη και 2,5% η τρίτη). Αντίστοιχα, για τα container ships το ποσοστό ανήλθε στο 92,3% (79,2%, 10,4% και 2,7%

αντίστοιχα). Η παρόμοια κατανομή της εξηγούμενης διακύμανσης καταδεικνύει τη συνέπεια στη δομή των δεδομένων.



**Διάγραμμα 4.9:** Σύγκριση των loadings των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) για δύο τύπους πλοίων: Bulk carriers (μπλε) και Container ships (πορτοκαλί). Οι διαφοροποιήσεις στις τιμές των loadings αποτυπώνουν τις διαφορετικές περιβαλλοντικές και λειτουργικές προτεραιότητες των δύο κατηγοριών πλοίων.

Η σύγκριση των loadings ανέδειξε σημαντική σταθερότητα στη δομή των κύριων συνιστωσών. Για την PC1, οι μεταβλητές Z1–Z5 (συνολική κατανάλωση καυσίμου και εκπομπές αερίων θερμοκηπίου) παρουσιάζουν παρόμοια υψηλά loadings και στις δύο κατηγορίες, με μέγιστη απόκλιση 0,184 (Z8) και μέσο όρο διαφορών 0,098. Στην PC2 εντοπίζονται μεγαλύτερες διαφορές, ιδίως στις μεταβλητές αποδοτικότητας Z7 και Z8, όπου τα container ships εμφανίζουν υψηλότερα loadings (μέγιστη διαφορά 0,321, μέσος όρος 0,139). Η PC3 παρουσιάζει τη μεγαλύτερη μεταβλητότητα (μέγιστη διαφορά 0,392, μέσος όρος 0,158), αν και οι βασικές μεταβλητές (περιφερειακές εκπομπές Z11–Z13) διατηρούν παρόμοια πρότυπα.

**Πίνακας 4.5:** Loadings των τριών πρώτων κύριων συνιστωσών (PC1, PC2, PC3) για τα υποσύνολα δεδομένων Bulk carriers και Container ships, καθώς και οι διαφορές μεταξύ των δύο φορτίων.

Μεταβλητή	Bulk carriers dataset			Container ships dataset			Ζεύγη διαφορών μεταξύ φορτίων		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Z1	0,356	0,204	0,064	0,401	-0,015	-0,001	0,044	0,219	0,065
Z2	0,358	0,2	0,064	0,403	-0,017	-0,003	0,045	0,217	0,067
Z3	0,055	0,093	-0,028	0,007	0,025	0,057	0,048	0,068	0,085
Z4	0,357	0,198	0,068	0,407	-0,018	-0,006	0,05	0,216	0,075
Z5	0,358	0,201	0,063	0,402	-0,017	-0,002	0,045	0,218	0,066
Z6	0,509	-0,844	-0,158	0,185	-0,399	0,044	0,323	0,445	0,202
Z7	0,006	0,012	-0,004	-0,001	0,002	0,001	0,007	0,01	0,005
Z8	0,006	0,012	-0,004	-0,001	0,002	0,001	0,007	0,01	0,005
Z9	0,022	0,086	-0,274	-0,005	0	0	0,027	0,086	0,274
Z10	0,022	0,086	-0,273	-0,005	0	0	0,027	0,085	0,273
Z11	0,266	0,269	-0,642	0,123	-0,837	-0,002	0,143	1,105	0,641
Z12	0,277	0,161	-0,018	0,395	0,257	-0,695	0,118	0,096	0,677
Z13	0,282	0,062	0,628	0,379	0,27	0,715	0,097	0,208	0,087
Διακύμανση που εξηγείται	0,894	0,05	0,022	0,859	0,085	0,025	0,035	0,035	0,003

Η ανάλυση συνέπειας έδειξε ότι οι ίδιες μεταβλητές παρουσιάζουν υψηλά loadings (>0,3) και στις δύο αναλύσεις. Στην PC1, σταθερά κυρίαρχες είναι οι Z1, Z2, Z4, Z5, Z12 και Z13, ενώ στην PC2 οι Z6, Z7, Z8, Z11 και Z12. Παρά τις μικρές διαφοροποιήσεις, η γενική δομή και ερμηνεία των τριών συνιστωσών ως «Περιβαλλοντική Επίδραση», «Λειτουργική Αποδοτικότητα» και «Περιφερειακά Πρότυπα Λειτουργίας» παραμένει συνεπής.

Οι παρατηρηθείσες διαφορές έχουν σημαντικές πρακτικές επιπτώσεις για τη χρήση των προτεινόμενων δεικτών. Πρώτον, η σταθερότητα της πρώτης κύριας συνιστώσας (PC1) επιβεβαιώνει ότι ο δείκτης "Περιβαλλοντική Επίδραση" είναι αξιόπιστος για τη σύγκριση της

συνολικής περιβαλλοντικής απόδοσης μεταξύ διαφορετικών τύπων πλοίων. Αυτό σημαίνει ότι οι κανονιστικές αρχές μπορούν να χρησιμοποιήσουν τον ίδιο δείκτη για τη θέσπιση στόχων εκπομπών τόσο για Bulk carriers όσο και για Container ships.

Δεύτερον, οι σημαντικότερες διαφορές στη δεύτερη κύρια συνιστώσα (PC2) αποκαλύπτουν ότι ο δείκτης "Λειτουργική Αποδοτικότητα" πρέπει να ερμηνεύεται διαφορετικά για κάθε τύπο πλοίου. Συγκεκριμένα, τα Container ships εμφανίζουν υψηλότερη ευαισθησία στις μεταβλητές αποδοτικότητας ανά απόσταση (Z7, Z8), γεγονός που αντικατοπτρίζει τη φύση των τακτικών δρομολογίων τους και την έμφαση στην ταχύτητα παράδοσης. Αντίθετα, τα Bulk carriers δείχνουν μεγαλύτερη ευαισθησία σε μεταβλητές που σχετίζονται με τον χρόνο στη θάλασσα, αντικατοπτρίζοντας την πιο ευέλικτη φύση των δρομολογίων τους.

Τρίτον, η μεταβλητότητα στην τρίτη κύρια συνιστώσα (PC3) υποδεικνύει ότι ο δείκτης "Περιφερειακά Πρότυπα Λειτουργίας" μπορεί να αντικατοπτρίζει πραγματικές διαφορές στις επιχειρησιακές στρατηγικές των δύο τύπων πλοίων. Τα Container ships, λόγω των τακτικών γραμμών τους, ενδέχεται να έχουν πιο σταθερά γεωγραφικά πρότυπα λειτουργίας εντός της ΕΕ, ενώ τα Bulk carriers μπορεί να παρουσιάζουν μεγαλύτερη ποικιλομορφία ανάλογα με τη διαθεσιμότητα φορτίων.

Συνολικά, η επικύρωση με ανεξάρτητα σύνολα δεδομένων δείχνει ότι οι προτεινόμενοι δείκτες είναι σταθεροί, αξιόπιστοι και γενικεύσιμοι σε διαφορετικούς τύπους πλοίων, επιβεβαιώνοντας την καταλληλότητά τους για εφαρμογή στον ναυτιλιακό κλάδο.

## 5. ΣΥΖΗΤΗΣΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

Φτάνοντας στο τελευταίο κεφάλαιο αυτής της διπλωματικής εργασίας, έχω πλέον ολοκληρώσει ένα ταξίδι που ξεκίνησε από τη θεωρητική ανασκόπηση, πέρασε από τη συλλογή και την επεξεργασία των δεδομένων, και κορυφώθηκε με την εφαρμογή προηγμένων αλγορίθμων μηχανικής μάθησης. Τα αριθμητικά αποτελέσματα και τα διαγράμματα που παρουσιάστηκαν στο προηγούμενο κεφάλαιο, αν και αποκαλυπτικά, αποτελούν μόνο τη μία όψη του νομίσματος. Σκοπός αυτού του κεφαλαίου είναι να συνθέσει αυτά τα ευρήματα, να συνδέσει τις τελείες και να ερμηνεύσει τι πραγματικά σημαίνουν. Εδώ, θα αναλύσουμε τη σημασία των αποτελεσμάτων σε σχέση με τους αρχικούς μας στόχους, θα αναδείξουμε τη συνεισφορά της έρευνας στον τομέα της ναυτιλίας και θα συζητήσουμε τόσο τις δυνατότητες όσο και τους περιορισμούς της προσέγγισής μας, καταλήγοντας σε συγκεκριμένα συμπεράσματα και προτάσεις για το μέλλον.

### 5.1 ΣΥΝΟΨΗ ΤΗΣ ΕΡΕΥΝΑΣ

Το σημείο εκκίνησης για αυτή την εργασία ήταν η αναγνώριση ενός σημαντικού κενού: παρόλο που η περιβαλλοντική απόδοση της ναυτιλίας βρίσκεται στο επίκεντρο του ενδιαφέροντος, τα υπάρχοντα εργαλεία αξιολόγησης είναι συχνά μονοδιάστατα και αδυνατούν να συλλάβουν την πολύπλοκη αλληλεπίδραση μεταξύ των εκπομπών, της λειτουργικής αποδοτικότητας και των τεχνικών χαρακτηριστικών των πλοίων. Οι τρέχοντες δείκτες, αν και χρήσιμοι, συχνά απλοποιούν υπερβολικά την πραγματικότητα, οδηγώντας σε μια εικόνα που δεν είναι πάντα πλήρης. Η έρευνα αυτή, λοιπόν, γεννήθηκε από την ανάγκη να δημιουργηθεί ένα πιο ολοκληρωμένο και αντικειμενικό πλαίσιο για την κατανόηση και την κατηγοριοποίηση της περιβαλλοντικής συμπεριφοράς των θαλάσσιων μέσων μεταφοράς.

Για να αντιμετωπιστεί αυτή η πρόκληση, αναπτύχθηκε ένα καινοτόμο μεθοδολογικό πλαίσιο που αξιοποιεί τη δύναμη της μηχανικής μάθησης και τον πλούτο των δεδομένων που παρέχει ο μηχανισμός THETIS-MRV. Η προσέγγιση ήταν διττή. Αρχικά, εφαρμόστηκε η PCA με σκοπό να αποστάξει τον τεράστιο όγκο των αλληλοσυσχετιζόμενων μεταβλητών σε έναν μικρό αριθμό ουσιαστικών και ανεξάρτητων δεικτών. Στη συνέχεια, αυτοί οι νέοι, συνθετικοί δείκτες χρησιμοποιήθηκαν ως βάση για την εφαρμογή της ανάλυσης συστάδων (K-Means), με στόχο την αποκάλυψη «φυσικών» ομαδοποιήσεων πλοίων με παρόμοια περιβαλλοντικά και λειτουργικά

προφίλ. Ο απώτερος σκοπός δεν ήταν απλώς να κατατάξουμε τα πλοία από το «καλύτερο» στο «χειρότερο», αλλά να κατανοήσουμε τα κρυμμένα πρότυπα που διέπουν τη συμπεριφορά τους. Το προηγούμενο κεφάλαιο παρουσίασε αναλυτικά τα ποσοτικά ευρήματα αυτής της διαδικασίας, αποκαλύπτοντας τη δομή των νέων δεικτών και τις συστάδες που προέκυψαν. Το παρόν κεφάλαιο θα προχωρήσει ένα βήμα παραπέρα, ερμηνεύοντας αυτά τα ευρήματα και συζητώντας τη σημασία τους για τον ναυτιλιακό κλάδο, τους ρυθμιστικούς φορείς και τη μελλοντική έρευνα.

## 5.2 ΚΥΡΙΑ ΕΥΡΗΜΑΤΑ ΚΑΙ ΕΡΜΗΝΕΙΑ

Η εφαρμογή των αλγορίθμων μηχανικής μάθησης δεν ήταν απλώς μια τεχνική άσκηση, αλλά ένα ερευνητικό εργαλείο που μας επέτρεψε να «ακούσουμε» την ιστορία που είχαν να πουν τα ίδια τα δεδομένα. Τα ευρήματα που προέκυψαν δεν περιορίζονται σε απλούς αριθμούς, αλλά αποκαλύπτουν μια βαθύτερη, πολυεπίπεδη εικόνα της περιβαλλοντικής απόδοσης του ναυτιλιακού στόλου. Η ερμηνεία αυτών των ευρημάτων μπορεί να διακριθεί σε δύο βασικούς άξονες: πρώτον, στη δημιουργία ενός νέου, πιο ολοκληρωμένου συστήματος δεικτών, και δεύτερον, στην αποκάλυψη διακριτών προφίλ συμπεριφοράς των πλοίων.

Το πρώτο και θεμελιώδες εύρημα της εργασίας είναι η επιτυχής απόσταξη της πολυπλοκότητας των αρχικών δεδομένων σε τρεις νέους, συνθετικούς δείκτες μέσω της PCA. Αυτοί οι δείκτες δεν είναι απλώς μαθηματικές αφαιρέσεις, αλλά φέρουν συγκεκριμένο και ερμηνεύσιμο νόημα. Ο πρώτος δείκτης (PC1) αναδείχθηκε ως ο κυρίαρχος, εξηγώντας ένα συντριπτικό ποσοστό της συνολικής διακύμανσης. Επειδή επηρεάζεται έντονα από μεταβλητές όπως η συνολική κατανάλωση καυσίμου και οι συνολικές εκπομπές CO<sub>2</sub> και άλλων αερίων, τον ερμηνεύσαμε ως τον δείκτη της «Συνολικής Περιβαλλοντικής Επίδρασης». Ουσιαστικά, αντιπροσωπεύει το απόλυτο περιβαλλοντικό αποτύπωμα ενός πλοίου, ανεξάρτητα από την αποδοτικότητά του. Ο δεύτερος δείκτης (PC2), από την άλλη, συσχετίζεται έντονα με τις μεταβλητές που μετρούν την κατανάλωση και τις εκπομπές ανά μονάδα απόστασης ή μεταφορικού έργου. Γι' αυτό, τον ονομάσαμε δείκτη «Λειτουργικής Αποδοτικότητας». Αυτός ο δείκτης απαντά στο ερώτημα: «Πόσο αποδοτικά εκτελεί ένα πλοίο το έργο του;». Τέλος, ο τρίτος δείκτης (PC3) σχετίζεται με τα γεωγραφικά πρότυπα λειτουργίας, δηλαδή με το αν ένα πλοίο δραστηριοποιείται κυρίως εντός ή εκτός της δικαιοδοσίας των κρατών-μελών της ΕΕ, αποτυπώνοντας τα «Περιφερειακά Πρότυπα Λειτουργίας». Η συνδυασμένη χρήση αυτών των τριών δεικτών

προσφέρει μια ολιστική και ισορροπημένη αξιολόγηση, ικανή να διακρίνει ένα μεγάλο, ρυπογόνο αλλά αποδοτικό πλοίο από ένα μικρότερο, λιγότερο ρυπογόνο αλλά αναποτελεσματικό.

Η πραγματική αξία αυτών των δεικτών, ωστόσο, αποκαλύφθηκε πλήρως στο δεύτερο στάδιο της ανάλυσης, με την ομαδοποίηση των πλοίων. Η ανάλυση συστάδων δεν κατέταξε απλώς τα πλοία σε μια γραμμική κλίμακα, αλλά αποκάλυψε διακριτά και ετερογενή «προφίλ» ή «προσωπικότητες» πλοίων. Η ανάλυση με τέσσερις συστάδες, για παράδειγμα, ήταν ιδιαίτερα διαφωτιστική. Εντοπίστηκε μια ομάδα πλοίων (3<sup>η</sup> συστάδα) με ακραία υψηλή περιβαλλοντική επίδραση και ταυτόχρονα πολύ χαμηλή λειτουργική αποδοτικότητα – οι «προβληματικοί ρυπαντές» του στόλου. Μια άλλη ομάδα (1<sup>η</sup> συστάδα) περιλάμβανε πλοία με επίσης υψηλή συνολική επίδραση, αλλά με μέτρια προς καλή αποδοτικότητα, πιθανότατα μεγάλα πλοία σε γραμμές υψηλής ταχύτητας και συχνότητας. Αντίθετα, η μεγαλύτερη συστάδα (2<sup>η</sup> συστάδα) περιλάμβανε πλοία με χαμηλή συνολική επίδραση και χαμηλή αποδοτικότητα – οι «αφανείς εργάτες» του στόλου, που πιθανόν λειτουργούν λιγότερο εντατικά. Αυτή η κατηγοριοποίηση αποδεικνύει ότι η «κακή περιβαλλοντική απόδοση» δεν είναι μια ενιαία έννοια. Ένα πλοίο μπορεί να είναι προβληματικό λόγω του τεράστιου όγκου καυσίμου που καταναλώνει, ενώ ένα άλλο μπορεί να είναι προβληματικό επειδή είναι εξαιρετικά αναποτελεσματικό, ακόμα κι αν οι συνολικές του εκπομπές είναι χαμηλότερες.

Συνοψίζοντας, τα ευρήματα της εργασίας μετατοπίζουν τη συζήτηση από τη μονοδιάστατη μέτρηση στη πολυδιάστατη κατανόηση. Η συνδυαστική προσέγγιση PCA και Clustering δεν παρέχει απλώς μια κατάταξη, αλλά έναν χάρτη της περιβαλλοντικής συμπεριφοράς του στόλου. Αποκαλύπτει ότι ο στόλος δεν είναι ένα ομοιογενές σύνολο, αλλά ένα οικοσύστημα από διαφορετικές ομάδες πλοίων, καθεμία με τα δικά της μοναδικά χαρακτηριστικά, που απαιτούν διαφορετική προσέγγιση και, ενδεχομένως, διαφορετικές πολιτικές για τη βελτίωσή τους.

### **5.3 ΣΥΝΕΙΣΦΟΡΑ ΤΗΣ ΕΡΓΑΣΙΑΣ**

Η αξία μιας ερευνητικής εργασίας δεν κρίνεται μόνο από τα αποτελέσματά της αυτά καθαυτά, αλλά και από τον τρόπο με τον οποίο εντάσσεται στον υπάρχοντα επιστημονικό διάλογο, απαντώντας σε αναπάντητα ερωτήματα και ανοίγοντας νέους δρόμους. Η παρούσα διπλωματική εργασία τοποθετείται συνειδητά σε αυτό το πλαίσιο, προσφέροντας τόσο μια θεωρητική

καινοτομία όσο και πρακτικά εφαρμόσιμες λύσεις σε προβλήματα που αναδείχθηκαν μέσα από τη βιβλιογραφική ανασκόπηση.

Η ανάλυση της διεθνούς βιβλιογραφίας, όπως παρουσιάστηκε στο Κεφάλαιο 2, κατέδειξε σαφώς ορισμένα κενά στην τρέχουσα προσέγγιση για την περιβαλλοντική αξιολόγηση των πλοίων. Το σημαντικότερο από αυτά ήταν η εξάρτηση από μεμονωμένους, συχνά μονοδιάστατους δείκτες (όπως οι EEDI, EEXI, CII), οι οποίοι, αν και σημαντικοί, αδυνατούν να συλλάβουν την ολιστική εικόνα της απόδοσης. Η παρούσα εργασία απαντά άμεσα σε αυτό το κενό, προτείνοντας ένα μεθοδολογικό πλαίσιο που συνθέτει πολλαπλές παραμέτρους σε τρεις ουσιαστικούς και ορθογώνιους δείκτες. Αντί να εξετάζει μεμονωμένα τις συνολικές εκπομπές ή την αποδοτικότητα, το μοντέλο μας τα συνδυάζει, προσφέροντας μια πιο ισορροπημένη και δίκαιη αξιολόγηση. Επιπλέον, η βιβλιογραφία ανέδειξε την περιορισμένη χρήση τεχνικών μη επιβλεπόμενης μάθησης, όπως η Ανάλυση Συστάδων, για την ανακάλυψη κρυμμένων προτύπων. Η εργασία αυτή καλύπτει και αυτό το κενό, αποδεικνύοντας εμπράκτως πώς η ομαδοποίηση μπορεί να αποκαλύψει ετερογενή προφίλ πλοίων, τα οποία θα παρέμεναν αόρατα με τις παραδοσιακές μεθόδους κατάταξης.

Από θεωρητική σκοπιά, η κύρια συνεισφορά της εργασίας έγκειται στην πρόταση και την επιτυχή εφαρμογή ενός νέου, συνδυαστικού πλαισίου ανάλυσης. Η διαδοχική χρήση της PCA για τη μείωση της διαστατικότητας και τη δημιουργία συνθετικών δεικτών, και ακολούθως της Ανάλυσης Συστάδων για την κατηγοριοποίηση, αποτελεί μια καινοτόμο προσέγγιση στον τομέα της ναυτιλιακής περιβαλλοντικής ανάλυσης. Δεν προτείνουμε απλώς έναν νέο δείκτη, αλλά μια ολιστική, data-driven μεθοδολογία που μπορεί να προσαρμοστεί και να επεκταθεί, προσφέροντας ένα ισχυρό αναλυτικό εργαλείο για την επιστημονική κοινότητα.

Πέρα από τη θεωρητική της αξία, η εργασία έχει και σημαντικές πρακτικές προεκτάσεις. Για τις ρυθμιστικές αρχές, τα ευρήματα αυτά είναι εξαιρετικά πολύτιμα. Η αναγνώριση διακριτών «προφίλ» πλοίων (π.χ. «ρυπαντές», «αποδοτικοί γίγαντες», «αφανείς εργάτες») μπορεί να οδηγήσει στη χάραξη πιο έξυπνων και στοχευμένων πολιτικών. Αντί για έναν οριζόντιο κανόνα που εφαρμόζεται σε όλα τα πλοία ανεξαιρέτως, θα μπορούσαν να σχεδιαστούν κίνητρα ή περιορισμοί προσαρμοσμένοι στις ιδιαιτερότητες κάθε ομάδας. Για παράδειγμα, τα μέτρα για τη βελτίωση της απόδοσης των «ρυπαντών» θα μπορούσαν να είναι διαφορετικά από τα μέτρα για τη μείωση του απόλυτου αποτυπώματος των «γιγάντων». Παράλληλα, οι ναυτιλιακές εταιρείες αποκτούν ένα ισχυρό εργαλείο για εσωτερική αξιολόγηση (benchmarking). Μπορούν να



εντοπίσουν σε ποια συστάδα ανήκει κάθε πλοίο του στόλου τους, να συγκρίνουν την απόδοσή του με άλλα παρόμοια πλοία και να λάβουν τεκμηριωμένες αποφάσεις για επενδύσεις σε νέες τεχνολογίες ή για τη βελτιστοποίηση των λειτουργιών τους. Με αυτόν τον τρόπο, η ανάλυση μετατρέπεται από μια ακαδημαϊκή άσκηση σε ένα πρακτικό εργαλείο υποστήριξης αποφάσεων.

#### 5.4 ΠΕΡΙΟΡΙΣΜΟΙ ΤΗΣ ΕΡΕΥΝΑΣ

Η επιστημονική πρόοδος απαιτεί όχι μόνο την ανάδειξη των νέων γνώσεων, αλλά και την ειλικρινή αναγνώριση των ορίων κάθε έρευνας. Η κατανόηση των περιορισμών της παρούσας εργασίας είναι απαραίτητη τόσο για τη σωστή ερμηνεία των συμπερασμάτων της, όσο και για τον προσανατολισμό της μελλοντικής έρευνας. Οι περιορισμοί αυτοί εντοπίζονται κυρίως στη φύση των δεδομένων που χρησιμοποιήθηκαν και στις εγγενείς υποθέσεις των μεθοδολογιών που εφαρμόστηκαν.

Ο πρώτος και βασικότερος περιορισμός πηγάζει από την ίδια τη βάση δεδομένων THETIS-MRV. Παρόλο που αποτελεί μια εξαιρετικά πλούσια και δημόσια διαθέσιμη πηγή λειτουργικών δεδομένων, δεν περιλαμβάνει λεπτομερή τεχνικά χαρακτηριστικά των πλοίων. Στοιχεία όπως η ηλικία του πλοίου, ο τύπος της κύριας μηχανής, η χρήση τεχνολογιών εξοικονόμησης ενέργειας ή η ποιότητα του καυσίμου, δεν ήταν διαθέσιμα για ανάλυση. Αυτό σημαίνει ότι το μοντέλο μας βασίστηκε αποκλειστικά στα λειτουργικά αποτυπώματα των πλοίων, χωρίς να μπορεί να συνδέσει άμεσα την απόδοση με συγκεκριμένες τεχνικές προδιαγραφές. Κατά συνέπεια, ενώ μπορούμε να εντοπίσουμε ένα πλοίο ως «αναποτελεσματικό», δεν μπορούμε να προσδιορίσουμε με βεβαιότητα αν αυτό οφείλεται σε παλιά τεχνολογία, κακή συντήρηση ή απλώς σε μη βέλτιστες λειτουργικές πρακτικές.

Ένας δεύτερος περιορισμός αφορά τις στατιστικές μεθόδους που επιλέχθηκαν. Η PCA είναι μια ισχυρή τεχνική, αλλά βασίζεται στην υπόθεση ότι οι σχέσεις μεταξύ των μεταβλητών είναι κατά βάση γραμμικές. Ενώ αυτή η υπόθεση φαίνεται να ισχύει σε μεγάλο βαθμό για τα δεδομένα μας, όπως έδειξαν οι υψηλές γραμμικές συσχετίσεις, δεν μπορεί να αποκλειστεί η ύπαρξη πιο σύνθετων, μη γραμμικών σχέσεων που η PCA δεν μπορεί να συλλάβει. Επιπλέον, η ανάλυση συστάδων με τον αλγόριθμο K-Means, αν και ευρέως διαδεδομένη, παρουσιάζει ορισμένες ευαισθησίες. Όπως αναλύθηκε στο Κεφάλαιο 4, η επιλογή του βέλτιστου αριθμού συστάδων ( $k$ ) δεν είναι πάντα μονοσήμαντη, καθώς διαφορετικές μετρικές (silhouette vs. Gap

statistic) μπορεί να υποδεικνύουν διαφορετικές λύσεις. Η τελική επιλογή βασίστηκε σε μια ισορροπία μεταξύ στατιστικής εγκυρότητας και πρακτικής ερμηνευσιμότητας, αλλά ενέχει έναν βαθμό υποκειμενικότητας.

Τέλος, θα πρέπει να επισημανθεί ο περιορισμός της γενικευσιμότητας των ευρημάτων. Η ανάλυση εστιάστηκε σε πλοία που δραστηριοποιούνται εντός της Ευρωπαϊκής Ένωσης, όπως αυτά καταγράφονται από τον μηχανισμό MRV. Αν και αυτό το δείγμα είναι μεγάλο και ποικιλόμορφο, τα λειτουργικά πρότυπα και οι ρυθμιστικές πιέσεις εντός της ΕΕ μπορεί να διαφέρουν από αυτές σε άλλες περιοχές του κόσμου, όπως η Ασία ή η Αμερική. Επομένως, η άμεση γενίκευση των συγκεκριμένων προφίλ των συστάδων σε παγκόσμιο επίπεδο θα πρέπει να γίνεται με προσοχή, αν και το ίδιο το μεθοδολογικό πλαίσιο παραμένει απόλυτα έγκυρο και εφαρμόσιμο σε οποιοδήποτε αντίστοιχο σύνολο δεδομένων.

## 5.5 ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Κάθε περιορισμός της παρούσας εργασίας μετατρέπεται σε μια ευκαιρία και μια πρόταση για περαιτέρω διερεύνηση. Με βάση τα ευρήματα και τα όρια που συζητήθηκαν, διαμορφώνονται ορισμένες σαφείς κατευθύνσεις για τη μελλοντική έρευνα, οι οποίες θα μπορούσαν να εμβαθύνουν και να επεκτείνουν τα συμπεράσματα αυτής της διπλωματικής.

Σε μεθοδολογικό επίπεδο, υπάρχει σημαντικό περιθώριο για εξέλιξη. Θα ήταν εξαιρετικά ενδιαφέρουσα η διερεύνηση εναλλακτικών και πιο προηγμένων αλγορίθμων. Για παράδειγμα, θα μπορούσαν να εφαρμοστούν τεχνικές μη γραμμικής μείωσης διαστάσεων (όπως t-SNE ή UMAP) για να εξεταστεί αν υπάρχουν μη γραμμικές δομές στα δεδομένα που η PCA δεν κατάφερε να αποκαλύψει. Παράλληλα, η χρήση διαφορετικών αλγορίθμων ομαδοποίησης, όπως ο DBSCAN, ο οποίος είναι ικανός να εντοπίζει συστάδες αυθαίρετου σχήματος και να διαχειρίζεται καλύτερα τον «θόρυβο» (outliers), θα μπορούσε να προσφέρει μια διαφορετική οπτική στην κατηγοριοποίηση των πλοίων, ενδεχομένως αποκαλύπτοντας πιο σύνθετες ή εξειδικευμένες ομάδες.

Μια δεύτερη κατεύθυνση, με έντονο πρακτικό προσανατολισμό, είναι η ανάπτυξη ενός εργαλείου υποστήριξης αποφάσεων. Τα αποτελέσματα της παρούσας ανάλυσης θα μπορούσαν να αποτελέσουν τη βάση για τη δημιουργία ενός διαδραστικού πίνακα ελέγχου. Ένα τέτοιο εργαλείο θα επέτρεπε σε χρήστες, όπως στελέχη ρυθμιστικών αρχών ή ναυτιλιακών εταιρειών, να

οπτικοποιούν τη θέση κάθε πλοίου στον τρισδιάστατο χώρο των τριών δεικτών, να φιλτράρουν τα δεδομένα ανά τύπο πλοίου ή εταιρεία και να εξάγουν άμεσα και κατανοητά συμπεράσματα. Αυτό θα μετέτρεπε τη στατική ανάλυση σε ένα δυναμικό εργαλείο παρακολούθησης και στρατηγικού σχεδιασμού.

Τέλος, η παρούσα διερευνητική ανάλυση θέτει τις βάσεις για τη μετάβαση σε προγνωστική μοντελοποίηση. Οι συστάδες που δημιουργήθηκαν μπορούν να χρησιμοποιηθούν ως «ετικέτες» (labels) για την εκπαίδευση μοντέλων επιβλεπόμενης μάθησης (supervised learning). Ένα τέτοιο μοντέλο θα μπορούσε, για παράδειγμα, να εκπαιδευτεί να προβλέπει σε ποιο περιβαλλοντικό προφίλ (συστάδα) είναι πιθανότερο να ανήκει ένα νέο πλοίο, με βάση τα αρχικά τεχνικά και λειτουργικά του χαρακτηριστικά. Αυτό θα μπορούσε να αποδειχθεί εξαιρετικά χρήσιμο για την εκ των προτέρων αξιολόγηση κινδύνου και τη διαμόρφωση προληπτικών πολιτικών.

## 5.6 ΤΕΛΙΚΕΣ ΣΚΕΨΕΙΣ

Η πορεία προς έναν πιο βιώσιμο ναυτιλιακό τομέα δεν είναι ένας απλός, ευθύγραμμος δρόμος, αλλά μια σύνθετη διαδρομή που απαιτεί βαθιά κατανόηση, τεκμηριωμένες αποφάσεις και καινοτόμα εργαλεία. Η παρούσα διπλωματική εργασία επιχείρησε να συμβάλει σε αυτή την προσπάθεια, όχι προτείνοντας μια μαγική λύση, αλλά αναπτύσσοντας ένα πιο έξυπνο και διεισδυτικό πρίσμα για να εξετάσουμε το πρόβλημα.

Αποδείξαμε ότι είναι εφικτό να μετατρέψουμε έναν τεράστιο και πολύπλοκο όγκο δεδομένων εκπομπών σε ένα απλό, κατανοητό και πολυδιάστατο σύστημα αξιολόγησης, το οποίο αποτυπώνει ταυτόχρονα το απόλυτο περιβαλλοντικό αποτύπωμα, τη λειτουργική αποδοτικότητα και τα γεωγραφικά πρότυπα λειτουργίας κάθε πλοίου. Αυτό το μοντέλο δεν παρέχει απλώς μια κατάταξη, αλλά αποκαλύπτει τα κρυμμένα πρότυπα συμπεριφοράς και τις «προσωπικότητες» των πλοίων, ομαδοποιώντας τα σε διακριτές κατηγορίες με κοινά χαρακτηριστικά. Αυτή η γνώση είναι κρίσιμη, διότι μας διδάσκει ότι η βελτίωση της περιβαλλοντικής απόδοσης δεν μπορεί να επιτευχθεί με μια οριζόντια πολιτική που απευθύνεται σε όλους. Αντίθετα, απαιτεί στοχευμένες στρατηγικές, προσαρμοσμένες στις ιδιαιτερότητες κάθε ομάδας.

Σε τελική ανάλυση, η ναυτιλία, ως η ραχοκοκαλιά του παγκόσμιου εμπορίου, βρίσκεται αντιμέτωπη με την ιστορική πρόκληση της πράσινης μετάβασης. Η επιτυχία αυτής της μετάβασης θα εξαρτηθεί σε μεγάλο βαθμό από την ικανότητά μας να αξιοποιούμε τα δεδομένα, όχι απλώς για

να μετράμε το παρελθόν, αλλά για να κατανοούμε το παρόν και να σχεδιάζουμε ένα πιο βιώσιμο μέλλον. Η μεθοδολογία που αναπτύχθηκε σε αυτή την εργασία αποτελεί ένα βήμα προς αυτή την κατεύθυνση: ένα εργαλείο που προσφέρει διαφάνεια, αντικειμενικότητα και, πάνω απ' όλα, μια βαθύτερη κατανόηση, απαραίτητη για τη διαμόρφωση των πολιτικών και των πρακτικών που θα οδηγήσουν τον κλάδο στην επόμενη ημέρα.

## BIBΛΙΟΓΡΑΦΙΑ

- Aakko-Saksa, P., Lehtoranta, K., Kuittinen, N., Järvinen, A., Jalkanen, J., Johnson, K., Jung, H., Ntziachristos, L., Gagné, S., Takahashi, C., Karjalainen, P., Rönkkö, T., & Timonen, H. (2023). Reduction in greenhouse gas and other emissions from ship engines: Current trends and future options. *Progress in Energy and Combustion Science*. <https://doi.org/10.1016/j.pecs.2022.101055>
- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.
- Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health. *Journal of Medical Toxicology*, 8(2), 166–175. <https://doi.org/10.1007/s13181-011-0203-1>
- Bach, H., & Hansen, T. (2023). IMO off course for decarbonisation of shipping? Three challenges for stricter policy. *Marine Policy*. <https://doi.org/10.1016/j.marpol.2022.105379>
- Balcombe, P., Heggo, D., & Harrison, M. (2022). Total Methane and CO<sub>2</sub> Emissions from Liquefied Natural Gas Carrier Ships: The First Primary Measurements. *Environmental Science & Technology*, 56, 9632–9640. <https://doi.org/10.1021/acs.est.2c01383>
- Barreiro, J., Zaragoza, S., & Díaz-Casás, V. (2022). Review of ship energy efficiency. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2022.111594>
- Bayraktar, M., & Yuksel, O. (2023). A scenario-based assessment of the energy efficiency existing ship index (EEXI) and carbon intensity indicator (CII) regulations. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2023.114295>
- Borré, A., Seman, L., Camponogara, E., Stefenon, S., Mariani, V., & Coelho, L. (2023). Machine Fault Detection Using a Hybrid CNN-LSTM Attention-Based Model. *Sensors (Basel, Switzerland)*, 23. <https://doi.org/10.3390/s23094512>
- Chaudhary, M., Goyal, N., Benslimane, A., Awasthi, L., Alwadain, A., & Singh, A. (2023). Underwater Wireless Sensor Networks: Enabling Technologies for Node Deployment and Data Collection Challenges. *IEEE Internet of Things Journal*, 10, 3500–3524. <https://doi.org/10.1109/JIOT.2022.3218766>
- Chaudhry, S. A., Irshad, A., Khan, M. A., Khan, S. A., Nosheen, S., Alzubi, A., & Zikria, Y. (2023). A Lightweight Authentication Scheme for 6G-IoT Enabled Maritime Transport System. *IEEE*

- Transactions on Intelligent Transportation Systems*, 24, 2401–2410.  
<https://doi.org/10.1109/TITS.2021.3134643>
- Chebanov, S., & Osipova, M. (2016). THE PROBLEM OF NOISE POLLUTION FROM SHIPS ON MARINE LIFE. *Достижения и Перспективы Инноваций и Технологий*, 102–105.
- Chen, C., Fu, H., Zheng, Y., Tao, F., & Liu, Y. (2023). The advance of digital twin for predictive maintenance: The role and function of machine learning. *Journal of Manufacturing Systems*.  
<https://doi.org/10.1016/j.jmsy.2023.10.010>
- Chen, H., Chen, J., & Ding, J. (2021). Data evaluation and enhancement for quality improvement of machine learning. *IEEE Transactions on Reliability*, 70(2), 831–847.
- Chen, X., Lv, S., Shang, W.-L., Wu, H., Xian, J., & Song, C. (2024). Ship energy consumption analysis and carbon emission exploitation via spatial-temporal maritime data. *Applied Energy*.  
<https://doi.org/10.1016/j.apenergy.2024.122886>
- Clean Shipping Index. (2024). *Methodology and Reporting Guidelines*.  
<https://cleanshippingindex.com/download/18.15d4621318c8ad3b40f16300/1703674589475/Methodology-and-Reporting-Guidelines-2024.pdf>
- Collier, Sonya, Ruehl, Chris, Yoon, Seungju, Boriboonsomsin, Kanok, Durbin, Thomas D, Scora, George, Johnson, Kent, & Herner, Jorn. (2019). Impact of Heavy-Duty Diesel Truck Activity on Fuel Consumption and Its Implication for the Reduction of Greenhouse Gas Emissions. *Transportation Research Record*, 2673(3), 125–135.  
<https://doi.org/10.1177/0361198119834548>
- Corbett, J. J., & Winebrake, J. J. (2008). International trade and global shipping. In *Handbook on Trade and the Environment*. Edward Elgar Publishing.
- Deng, J., Wang, X., Wei, Z., Wang, L., Wang, C., & Chen, Z. (2020). A review of NO<sub>x</sub> and SO<sub>x</sub> emission reduction technologies for marine diesel engines and the potential evaluation of liquefied natural gas fuelled vessels. *The Science of the Total Environment*, 766, 144319.  
<https://doi.org/10.1016/j.scitotenv.2020.144319>
- Dewan, M. H., & Godina, R. (2023). Roles and challenges of seafarers for implementation of energy efficiency operational measures onboard ships. *Marine Policy*.  
<https://doi.org/10.1016/j.marpol.2023.105746>

- Dewan, M. H., & Godina, R. (2024). Unveiling seafarers' awareness and knowledge on energy-efficient and low-carbon shipping: A decade of IMO regulation enforcement. *Marine Policy*. <https://doi.org/10.1016/j.marpol.2024.106037>
- Endresen, Ø., Sørsgård, E., Sundet, J. K., Dalsøren, S. B., Isaksen, I. S. A., Berglen, T. F., & Gravir, G. (2003). Emission from international sea transportation and environmental impact. *Journal of Geophysical Research: Atmospheres*, 108(D17). <https://doi.org/https://doi.org/10.1029/2002JD002898>
- Environmental Ship Index (ESI). (2020). *ESI Website Manual Version 1.1.1*. [https://assets.ctfassets.net/wnob2adoewm2/2IHIn4SSBbT5aDq9iFSw10/2d0715fc2d623608e5179863d94f526e/ESI\\_website\\_manual\\_2020\\_v1.1.1.pdf](https://assets.ctfassets.net/wnob2adoewm2/2IHIn4SSBbT5aDq9iFSw10/2d0715fc2d623608e5179863d94f526e/ESI_website_manual_2020_v1.1.1.pdf)
- European Maritime Safety Agency (EMSA). (2025). *THETIS-MRV Emissions Reporting System*. European Maritime Safety Agency (EMSA). <https://mrv.emsa.europa.eu/#public/emission-report>
- Eyring, V., Isaksen, I. S. A., Berntsen, T., Collins, W. J., Corbett, J. J., Endresen, O., Grainger, R. G., Moldanova, J., Schlager, H., & Stevenson, D. S. (2010). Transport impacts on atmosphere and climate: Shipping. *Atmospheric Environment*, 44(37), 4735–4771. <https://doi.org/https://doi.org/10.1016/j.atmosenv.2009.04.059>
- Faber, J., Hanayama, S., Zhang, S., Pereda, P., Comer, B., Hauerhof, E., van der Loeff, W. S., Smith, T., Zhang, Y., & Kosaka, H. (2020). Fourth imo ghg study. *London, UK*.
- Fan, A., Yan, J., Xiong, Y., Shu, Y., Fan, X., Wang, Y., He, Y., & Chen, J. (2023). Characteristics of real-world ship energy consumption and emissions based on onboard testing. *Marine Pollution Bulletin*, 194 Pt A, 115411. <https://doi.org/10.1016/j.marpolbul.2023.115411>
- Feng, Y., Wang, H., Xia, G., Cao, W., Li, T., Wang, X., & Liu, Z. (2024). A machine learning-based data-driven method for risk analysis of marine accidents. *Journal of Marine Engineering & Technology*. <https://doi.org/10.1080/20464177.2024.2368914>
- Fridell, E. (2019). Chapter 2 - Emissions and Fuel Use in the Shipping Sector. In R. Bergqvist & J. Monios (Eds.), *Green Ports* (pp. 19–33). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-814054-3.00002-5>
- Garciano, J. R. (2023). Understanding the impact of the maritime shipping industry to a sustainable economic development. *Sciences*, 1(6), 557–563.

- Grewe, V., Dahlmann, K., Matthes, S., & Steinbrecht, W. (2012). Attributing ozone to NO<sub>x</sub> emissions: Implications for climate mitigation measures. *Atmospheric Environment*, 59, 102–107. <https://doi.org/https://doi.org/10.1016/j.atmosenv.2012.05.002>
- Ha, S.-M., Jeong, B., Jang, H., Park, C., & Ku, B. (2023). A framework for determining the life cycle GHG emissions of fossil marine fuels in countries reliant on imported energy through maritime transportation: A case study of South Korea. *The Science of the Total Environment*, 165366. <https://doi.org/10.1016/j.scitotenv.2023.165366>
- Han, Y., Sun, Y., & Wu, J. (2023). A novel solar-driven waste heat recovery system in solar-fuel hybrid power plants. *Energy*, 285, 129396. <https://doi.org/https://doi.org/10.1016/j.energy.2023.129396>
- Henderson, J. (2019). Oceans without History? Marine Cultural Heritage and the Sustainable Development Agenda. *Sustainability*, 11(18). <https://doi.org/10.3390/su11185080>
- Hoang, A., Foley, A., Nižetić, S., Huang, Z.-H., Ong, H. C., Ölçer, A., Pham, V., & Nguyen, X. (2022). Energy-related approach for reduction of CO<sub>2</sub> emissions: A strategic review on the port-to-ship pathway. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2022.131772>
- Hubert, M., Rousseeuw, P., & Verdonck, T. (2009). Robust PCA for skewed data and its outlier map. *Computational Statistics & Data Analysis*, 53(6), 2264–2274.
- International Association of Ports and Harbors (IAPH). (2025). *Environmental Ship Index | IAPH*. International Association of Ports and Harbors (IAPH). <https://www.iaphworldports.org/environmental-ship-index-esi/>
- Jacquet, L., Duigou, L., & Kerbrat, O. (2024). A systematic literature review on holistic lifecycle assessments as a basis to create a standard in maritime industry. *The International Journal of Life Cycle Assessment*. <https://doi.org/10.1007/s11367-023-02269-4>
- Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.
- Jolliffe, I. T., & Jolliffe, I. T. (1986). *Mathematical and statistical properties of sample principal components*. Springer.
- Kaidarova, A., Geraldi, N., Wilson, R., Kosel, J., Meekan, M., Eguíluz, V., Hussain, M., Shamim, A., Liao, H., Srivastava, M., Saha, S. S., Strano, M., Zhang, X., Ooi, B., Holton, M., Hopkins, L., Jin, X., Gong, X., Quintana, F., ... Duarte, C. (2023). Wearable sensors for monitoring



- marine environments and their inhabitants. *Nature Biotechnology*, 41, 1208–1220. <https://doi.org/10.1038/s41587-023-01827-3>
- Kaplan, M. B., & Solomon, S. (2016). A coming boom in commercial shipping? The potential for rapid growth of noise from commercial ships by 2030. *Marine Policy*, 73, 119–121. <https://doi.org/https://doi.org/10.1016/j.marpol.2016.07.024>
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An overview of principal component analysis. *Journal of Signal and Information Processing*, 4(3), 173–175.
- Lee, S.-S. (2024). Analysis of the effects of EEDI and EEXI implementation on CO2 emissions reduction in ships. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2024.116877>
- Li, H., Jiao, H., & Yang, Z. (2023). Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis. *Eng. Appl. Artif. Intell.*, 126, 107062. <https://doi.org/10.1016/j.engappai.2023.107062>
- Li, J., Zhang, G., Zhang, X., & Zhang, W. (2024). Integrating Dynamic Event-Triggered and Sensor-Tolerant Control: Application to USV-UAVs Cooperative Formation System for Maritime Parallel Search. *IEEE Transactions on Intelligent Transportation Systems*, 25, 3986–3998. <https://doi.org/10.1109/TITS.2023.3326271>
- Libório, M. P., Da Silva Martinuci, O., Machado, A. M. C., Machado-Coelho, T. M., Laudares, S., & Bernardes, P. (2022). Principal component analysis applied to multidimensional social indicators longitudinal studies: limitations and possibilities. *GeoJournal*, 87(3), 1453–1468.
- Lippitt, W., Carlson, N. E., Arbet, J., Fingerlin, T. E., Maier, L. A., & Kechris, K. (2024). Limitations of clustering with PCA and correlated noise. *Journal of Statistical Computation and Simulation*, 94(10), 2291–2319.
- Liu, D., Zhang, Y., Wang, W., Dev, K., & Khowaja, S. (2023). Flexible Data Integrity Checking With Original Data Recovery in IoT-Enabled Maritime Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 24, 2618–2629. <https://doi.org/10.1109/TITS.2021.3125070>
- Liu, J., Li, C., Bai, J., Luo, Y., Lv, H., & Lv, Z. (2023). Security in IoT-Enabled Digital Twins of Maritime Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 24, 2359–2367. <https://doi.org/10.1109/TITS.2021.3122566>

- Mekkaoui, S. El, Benabbou, L., & Berrado, A. (2022). Machine Learning Models for Efficient Port Terminal Operations: Case of Vessels' Arrival Times Prediction. *IFAC-PapersOnLine*, 55(10), 3172–3177. <https://doi.org/https://doi.org/10.1016/j.ifacol.2022.10.217>
- Melnyk, O. M., Shcherbina, O. V, Mykhailova, I. V, Obnyavko, T. S., & Korobko, T. O. (2023). Focused research on technological innovations in shipping industry: Review and prospects. *Transport Development*, 1(16), 164–174.
- Mian, Z., Deng, X., Dong, X., Tian, Y., Cao, T., Chen, K., & Jaber, T. Al. (2024). A literature review of fault diagnosis based on ensemble learning. *Eng. Appl. Artif. Intell.*, 127, 107357. <https://doi.org/10.1016/j.engappai.2023.107357>
- Moldanová, J., Hassellöv, I., Matthias, V., Fridell, E., Jalkanen, J., Ytreberg, E., Quante, M., Tröltzsch, J., Maljutenko, I., Raudsepp, U., & Eriksson, K. (2021). Framework for the environmental impact assessment of operational shipping. *Ambio*, 51, 754–769. <https://doi.org/10.1007/s13280-021-01597-9>
- Mondello, G., Salomone, R., Saija, G., Lanuzza, F., & Gulotta, T. (2021). Life Cycle Assessment and Life Cycle Costing for assessing maritime transport: a comprehensive literature review. *Maritime Policy & Management*, 50, 198–218. <https://doi.org/10.1080/03088839.2021.1972486>
- Moradi, M. H., Brutsche, M., Wenig, M., Wagner, U., & Koch, T. (2022). Marine route optimization using reinforcement learning approach to reduce fuel consumption and consequently minimize CO2 emissions. *Ocean Engineering*, 259, 111882. <https://doi.org/https://doi.org/10.1016/j.oceaneng.2022.111882>
- Mueller, N., Westerby, M., & Nieuwenhuijsen, M. (2022). Health impact assessments of shipping and port-sourced air pollution on a global scale: A scoping literature review. *Environmental Research*, 114460. <https://doi.org/10.1016/j.envres.2022.114460>
- Narayana, T., Venkatesh, C., Kiran, A., J, C., Kumar, A., Khan, S. B., Almusharraf, A., & Quasim, T. (2024). Advances in real time smart monitoring of environmental parameters using IoT and sensors. *Heliyon*, 10. <https://doi.org/10.1016/j.heliyon.2024.e28195>
- Ni, P., Wang, X., & Li, H. (2020). A review on regulations, current status, effects and reduction strategies of emissions for marine diesel engines. *Fuel*, 279, 118477. <https://doi.org/10.1016/j.fuel.2020.118477>

- Paul, L. C., Suman, A. Al, & Sultan, N. (2013). Methodological analysis of principal component analysis (PCA) method. *International Journal of Computational Engineering & Management*, 16(2), 32–38.
- Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., & Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, 372(6547), 1209–1214.
- Pivetta, D., Dall’Armi, C., Sandrin, P., Bogar, M., & Taccani, R. (2024). The role of hydrogen as enabler of industrial port area decarbonization. *Renewable and Sustainable Energy Reviews*, 189, 113912. <https://doi.org/https://doi.org/10.1016/j.rser.2023.113912>
- Qu, J., Liu, R. W., Guo, Y., Lu, Y., Su, J., & Li, P. (2023). Improving maritime traffic surveillance in inland waterways using the robust fusion of AIS and visual data. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2023.114198>
- Rani, S., Babbar, H., Kaur, P., Alshehri, M., & Shah, S. (2023). An Optimized Approach of Dynamic Target Nodes in Wireless Sensor Network Using Bio Inspired Algorithms for Maritime Rescue. *IEEE Transactions on Intelligent Transportation Systems*, 24, 2548–2555. <https://doi.org/10.1109/TITS.2021.3129914>
- Rodionova, O., Kucheryavskiy, S., & Pomerantsev, A. (2021). Efficient tools for principal component analysis of complex data— a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 213, 104304. <https://doi.org/https://doi.org/10.1016/j.chemolab.2021.104304>
- Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299–1319.
- Shammas, N. K., Wang, L. K., & Wang, M.-H. S. (2020). Sources, chemistry and control of acid rain in the environment. In *Handbook of environment and waste management: Acid rain and greenhouse gas pollution control* (pp. 1–26). World Scientific.
- Skarlatos, K., Foustieris, A., Georgakellos, D., Economou, P., & Bersimis, S. (2023). Assessing Ships’ Environmental Performance Using Machine Learning. *Energies*. <https://doi.org/10.3390/en16062544>
- Smith, L. I. (2002). *A tutorial on principal components analysis*.
- Styhre, L., & Winnes, H. (2019). Chapter 6 - Emissions From Ships in Ports. In R. Bergqvist & J. Monios (Eds.), *Green Ports* (pp. 109–124). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-814054-3.00006-2>

- Sun, R., Abouarghoub, W., & Demir, E. (2025). Enhancing data quality in maritime transportation: A practical method for imputing missing ship static data. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2024.119722>
- Surucu, O., Gadsden, A., & Yawney, J. (2023). Condition Monitoring using Machine Learning: Theory and Applications. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2023.119738>
- Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to Data Mining*. Pearson India. <https://books.google.gr/books?id=vIqqDwAAQBAJ>
- Vasilikis, N., Geertsma, R., & Coraddu, A. (2023). A digital twin approach for maritime carbon intensity evaluation accounting for operational and environmental uncertainty. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2023.115927>
- Wan, Z., Ji, S., Liu, Y.-T., Zhang, Q., Chen, J., & Wang, Q. (2020). Shipping emission inventories in China's Bohai Bay, Yangtze River Delta, and Pearl River Delta in 2018. *Marine Pollution Bulletin*, 151, 110882. <https://doi.org/10.1016/j.marpolbul.2019.110882>
- Wang, B., Liu, Q., Wang, L., Chen, Y.-J., & Wang, J. (2023). A review of the port carbon emission sources and related emission reduction technical measures. *Environmental Pollution*, 121000. <https://doi.org/10.1016/j.envpol.2023.121000>
- Wang, C., Quang, B. X., & Nguyen, T. (2025). Benchmarking Efficiency, Sustainability, and Corporate Responsibility in Maritime Logistics: An Entropy-GRA Model with Sensitivity Analysis. *Sustainability*. <https://doi.org/10.3390/su17093813>
- Wang, C., Shen, J., Vijayakumar, P., & Gupta, B. (2023). Attribute-Based Secure Data Aggregation for Isolated IoT-Enabled Maritime Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 24, 2608–2617. <https://doi.org/10.1109/TITS.2021.3127436>
- Wang, H., Zheng, J., & Xiang, J. (2023). Online bearing fault diagnosis using numerical simulation models and machine learning classifications. *Reliab. Eng. Syst. Saf.*, 234, 109142. <https://doi.org/10.1016/j.ress.2023.109142>
- Wang, X., Yi, W., Lv, Z., Deng, F., Zheng, S., Xu, H., Zhao, J., Liu, H., & He, K. (2021). Ship emissions around China under gradually promoted control policies from 2016 to 2019. *Atmospheric Chemistry and Physics*. <https://doi.org/10.5194/acp-21-13835-2021>

- Wang, X.-T., Liu, H., Lv, Z.-F., Deng, F.-Y., Xu, H.-L., Qi, L.-J., Shi, M.-S., Zhao, J.-C., Zheng, S.-X., Man, H.-Y., & He, K.-B. (2021). Trade-linked shipping CO<sub>2</sub> emissions. *Nature Climate Change*, 11(11), 945–951. <https://doi.org/10.1038/s41558-021-01176-6>
- Wang, Y., Liu, J., Liu, R. W., Liu, Y., & Yuan, Z. (2023). Data-driven methods for detection of abnormal ship behavior: Progress and trends. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2023.113673>
- Wang, Y., Wright, L., Boccolini, V., & Ridley, J. (2024). Modelling environmental life cycle performance of alternative marine power configurations with an integrated experimental assessment approach: A case study of an inland passenger barge. *The Science of the Total Environment*, 173661. <https://doi.org/10.1016/j.scitotenv.2024.173661>
- Weng, J., Shi, K., Gan, X., Li, G., & Huang, Z. (2020). Ship emission estimation with high spatial-temporal resolution in the Yangtze River estuary using AIS data. *Journal of Cleaner Production*, 248, 119297. <https://doi.org/10.1016/j.jclepro.2019.119297>
- Williams, P. J. le B. (2010). The natural oceanic carbon and sulfur cycles: implications for SO<sub>2</sub> and CO<sub>2</sub> emissions from marine shipping. *Underwater Technology*, 29(1), 5–19.
- Winnes, H., & Fridell, E. (2010). Emissions of NO<sub>x</sub> and particles from manoeuvring ships. *Transportation Research Part D: Transport and Environment*, 15(4), 204–211. <https://doi.org/https://doi.org/10.1016/j.trd.2010.02.003>
- Yuan, Q., Wang, S., & Peng, J.-X. (2023). Operational efficiency optimization method for ship fleet to comply with the carbon intensity indicator (CII) regulation. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2023.115487>
- Zhang, M., Kujala, P., Musharraf, M., Zhang, J.-F., & Hirdaris, S. (2023). A machine learning method for the prediction of ship motion trajectories in real operational conditions. *Ocean Engineering*. <https://doi.org/10.1016/j.oceaneng.2023.114905>
- Zhang, M., Tsoulakos, N., Kujala, P., & Hirdaris, S. (2024). A deep learning method for the prediction of ship fuel consumption in real operational conditions. *Eng. Appl. Artif. Intell.*, 130, 107425. <https://doi.org/10.1016/j.engappai.2023.107425>
- Zhang, P., Wang, Y., Aujla, G., Jindal, A., & Al-Otaibi, Y. (2023). A Blockchain-Based Authentication Scheme and Secure Architecture for IoT-Enabled Maritime Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 24, 2322–2331. <https://doi.org/10.1109/TITS.2022.3159485>

- Zhao, J., Wei, Q.-F., Wang, S., & Ren, X. (2021). Progress of ship exhaust gas control technology. *The Science of the Total Environment*, 799, 149437. <https://doi.org/10.1016/j.scitotenv.2021.149437>
- Zincir, B. (2023). Slow steaming application for short-sea shipping to comply with the CII regulation. *Brodogradnja*. <https://doi.org/10.21278/brod74202>

## ΠΑΡΑΡΤΗΜΑ Α

[Προσοχή!! Τα παραρτήματα αρχίζουν πάντα σε σελίδα με μονό αριθμό.]

**[Τίτλος παραρτήματος]**

### A.1 ΕΝΟΤΗΤΑ 1

Θέση πρώτης παραγράφου

Θέση δεύτερης παραγράφου

.....

### A.2 ΕΝΟΤΗΤΑ 2

Θέση πρώτης παραγράφου

Θέση δεύτερης παραγράφου

.....