



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

**ΠΜΣ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ
ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΑΝΑΛΥΤΙΚΗ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**

Διπλωματική εργασία

Ανάλυση Συγκεντρώσεων Διοξειδίου του Θείου (SO₂)



Ηλίας Ν. Περγαντής

[me 2322]

**Επιβλέπων καθηγητής
Μιχαήλ Φιλιππάκης**

Σεπτέμβριος 2025

Πανεπιστήμιο Πειραιώς. Κάτοχος όλων των δικαιωμάτων
University of Piraeus. All rights reserved.
Συγγραφέας /Author: Ηλίας Ν. Περγαντής - ME2322

ΣΕΛΙΔΑ ΕΓΚΥΡΟΤΗΤΑΣ

Όνοματεπώνυμο Φοιτητή: Ηλίας Ν. Περγαντής

Τίτλος Μεταπτυχιακής Διπλωματικής Εργασίας: **Ανάλυση Συγκεντρώσεων Διοξειδίου του Θείου (SO₂)**

Η παρούσα Μεταπτυχιακή Διπλωματική Εργασία υποβάλλεται ως μερική εκπλήρωση των απαιτήσεων του Προγράμματος Μεταπτυχιακών Σπουδών “Πληροφοριακά Συστήματα & Υπηρεσίες” του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς και εγκρίθηκε από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή

Επιβλέπων/ (Τμήμα Ψηφιακών Συστημάτων, Πανεπιστήμιο Πειραιώς) Φιλιππάκης Μιχαήλ, Καθηγητής

Μέλος Εξεταστικής Επιτροπής: Κυριαζής Δημοσθένης, Καθηγητής

Μέλος Εξεταστικής Επιτροπής: Μαρία Χαλκίδη, Καθηγήτρια

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΑΥΘΕΝΤΙΚΟΤΗΤΑΣ

*Εγώ ο Ηλίας Ν. Περγαντής, γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «**Ανάλυση Συγκεντρώσεων Διοξειδίου του Θείου (SO₂)**», αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.*

Επιπλέον δηλώνω υπεύθυνα ότι η συγκεκριμένη Μεταπτυχιακή Διπλωματική Εργασία έχει συγγραφεί από εμένα προσωπικά και δεν έχει υποβληθεί ούτε έχει αξιολογηθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου. Σε κάθε περίπτωση, αναληθούς ή ανακριβούς δηλώσεως, υπόκειμαι στις συνέπειες που προβλέπονται τις διατάξεις που προβλέπει η Ελληνική και Κοινωνική Νομοθεσία περί πνευματικής ιδιοκτησίας.

Ο ΔΗΛΩΝ

Όνοματεπώνυμο: Ηλίας Ν. Περγαντής

Αριθμός Μητρώου: ΜΕ2322

Υπογραφή:

Ευχαριστίες / Acknowledgements

Με την παρούσα διπλωματική εργασία ολοκληρώνεται ένας θαυμαστός κύκλος σπουδών στο Μεταπτυχιακό Πρόγραμμα «Πληροφοριακά Συστήματα & Υπηρεσίες». Αρωγός, σε αυτό το εγχείρημα, στάθηκαν οι καθηγητές του Πανεπιστημίου Πειραιώς, με τις βαθιές και πολύπλευρες γνώσεις τους, την ικανότητα να μεταλαμπαδεύουν τις γνώσεις τους και να προκαλούν διαρκώς το έντονο ενδιαφέρον των συμμετεχόντων.

Τις πιο σημαντικές ευχαριστίες μου οφείλω στη σύζυγό μου, που με στήριξε με κάθε τρόπο σε αυτό το ταξίδι, στην οικογένειά μου, για τη διαρκή και πολύτιμη συνδρομή τους και στον πρόσφατα εκλιπόντα πατέρα μου, που το χαμόγελο και η περηφάνια του, μου έδιναν πάντοτε δύναμη για περεταίρω εξέλιξη.

Contents

ΕΙΣΑΓΩΓΗ	7
ΚΕΦΑΛΑΙΟ 1: ΤΟ ΔΙΟΞΕΙΔΙΟ ΤΟΥ ΘΕΙΟΥ (SO₂)	8
1.1 ΕΙΣΑΓΩΓΗ.....	8
1.2 ΦΥΣΙΚΕΣ ΚΑΙ ΧΗΜΙΚΕΣ ΙΔΙΟΤΗΤΕΣ.....	8
1.3 ΦΥΣΙΚΕΣ ΚΑΙ ΑΝΘΡΩΠΟΓΕΝΕΙΣ ΠΗΓΕΣ.....	9
1.4 ΑΤΜΟΣΦΑΙΡΙΚΟΙ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΙΚΕΣ ΕΠΙΠΤΩΣΕΙΣ.....	10
1.5 ΕΠΙΠΤΩΣΕΙΣ ΣΤΗΝ ΥΓΕΙΑ.....	12
1.6 ΧΡΗΣΕΙΣ ΤΟΥ SO ₂	13
1.7 ΙΣΤΟΡΙΚΟ ΠΛΑΙΣΙΟ ΚΑΙ ΡΥΘΜΙΣΕΙΣ.....	14
ΚΕΦΑΛΑΙΟ 2: ΤΟ ΔΙΟΞΕΙΔΙΟ ΤΟΥ ΘΕΙΟΥ ΩΣ ΑΕΡΙΟ ΤΟΥ ΘΕΡΜΟΚΗΠΙΟΥ ΚΑΙ ΟΙ ΣΥΝΕΠΕΙΕΣ ΤΟΥ	15
2.1 ΤΟ ΦΑΙΝΟΜΕΝΟ ΤΟΥ ΘΕΡΜΟΚΗΠΙΟΥ.....	15
2.2 ΤΟ ΔΙΟΞΕΙΔΙΟ ΤΟΥ ΘΕΙΟΥ ΚΑΙ ΤΑ ΑΕΡΙΑ ΤΟΥ ΘΕΡΜΟΚΗΠΙΟΥ.....	15
2.3 ΚΛΙΜΑΤΙΚΗ ΕΠΙΔΡΑΣΗ - ΕΠΙΠΤΩΣΕΙΣ.....	16
2.4 ΛΟΙΠΕΣ ΕΠΙΣΗΜΑΝΣΕΙΣ.....	17
2.4.1 Διάρκεια ζωής στην ατμόσφαιρα.....	17
2.4.2 Σύνδεση με πολιτικές καθαρού αέρα.....	17
2.4.3 Αλληλεπίδραση με νέφη.....	17
2.4.4 Γεωγραφική διαφοροποίηση.....	18
ΚΕΦΑΛΑΙΟ 3: ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΧΡΟΝΟΣΕΙΡΩΝ	18
3.1 ΟΡΙΣΜΟΣ ΚΑΙ ΕΝΝΟΙΑ ΤΗΣ ΧΡΟΝΟΣΕΙΡΑΣ.....	18
3.1.1 Βασικά συστατικά μιας χρονοσειράς.....	19
3.2 ΑΠΟΣΥΝΘΕΣΗ ΧΡΟΝΟΣΕΙΡΩΝ (TIME SERIES DECOMPOSITION).....	21
3.2.1 Μέθοδοι εξομάλυνσης.....	23
3.3 ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ.....	25
3.3.1 Holt-Winters – ως επέκταση της τριπλής εκθετικής εξομάλυνσης.....	25
3.3.2 Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX).....	26
3.3.3 Long Short-Term Memory (LSTM) νευρωνικά δίκτυα.....	29
ΚΕΦΑΛΑΙΟ 4: ΑΝΑΛΥΣΗ ΣΥΓΚΕΝΤΡΩΣΕΩΝ ΔΙΟΞΕΙΔΙΟΥ ΤΟΥ ΘΕΙΟΥ	33
4.1 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ SO ₂	33
4.2 ΣΥΛΛΟΓΗ ΜΕΤΕΩΡΟΛΟΓΙΚΩΝ ΔΕΔΟΜΕΝΩΝ.....	34
4.3 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ.....	35
4.4 ΤΕΧΝΙΚΗ ΔΙΑΔΙΚΑΣΙΑ ΑΠΟΣΥΝΘΕΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ.....	37
4.5 ΑΝΑΛΥΣΗ ΕΠΟΧΙΚΟΤΗΤΑΣ.....	43
4.5.1 Αξιολόγηση έντασης εποχικών μοτίβων.....	43
4.5.2 Ποιοτική αξιολόγηση των εποχικών μοτίβων.....	45
4.5.3 Οπτική αξιολόγηση εποχικής μεταβλητότητας.....	48
4.6 ΑΝΑΛΥΣΗ ΤΑΣΗΣ.....	50
4.6.1 Έλεγχος ύπαρξης τάσης.....	50
4.6.2 Εκτίμηση ρυθμού μεταβολής της τάσης.....	54
4.6.3 Υπολογισμός αβεβαιότητας του ρυθμού μεταβολής τάσης.....	58
4.6.4 Στατιστική σύγκριση ρυθμού μεταβολής τάσης μεταξύ περιφερειών.....	61
4.6.5 Ανάλυση σταθερότητας τάσης (Rolling trend analysis).....	64
4.6.6 Ανάλυση μεταβολών τάσης (Trend Shift Analysis).....	74
4.7 ΑΝΑΛΥΣΗ ΚΑΤΑΛΟΙΠΩΝ ΑΠΟΣΥΝΘΕΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ.....	80
4.7.1 Υπολογισμός στατιστικών δεικτών καταλοίπων.....	81
4.7.2 Ανάλυση κατανομής και ακραίων τιμών καταλοίπων.....	82
4.8 ΑΝΑΛΥΣΗ ΕΠΙΠΤΩΣΕΩΝ ΒΡΑΧΥΠΡΟΘΕΣΜΩΝ ΣΥΜΒΑΝΤΩΝ.....	86
4.8.1 Ανάλυση συγκεντρώσεων διοξειδίου του θείου τους μήνες Μάρτιο - Απρίλιο - Μάιο.....	87
4.8.2 Ανάλυση επιρροής συγκεντρώσεων διοξειδίου του θείου από εξωγενείς παράγοντες.....	88

4.9 ΣΥΣΧΕΤΙΣΗ ΔΙΟΞΕΙΔΙΟΥ ΤΟΥ ΘΕΙΟΥ ΜΕ ΜΕΤΕΩΡΟΛΟΓΙΚΟΥΣ ΠΑΡΑΓΟΝΤΕΣ ΚΑΙ ΤΟ ΟΖΟΝ (O ₃)	93
4.9.1. Ανάλυση συσχέτισης συγκεντρώσεων διοξειδίου του θείου με το όζον (O ₃).	93
4.9.2 Ανάλυση συσχέτισης συγκεντρώσεων διοξειδίου του θείου με μετεωρολογικούς παράγοντες.	97
ΚΕΦΑΛΑΙΟ 5: ΠΡΟΒΛΕΨΗ ΣΥΓΚΕΝΤΡΩΣΕΩΝ ΔΙΟΞΕΙΔΙΟΥ ΤΟΥ ΘΕΙΟΥ.....	102
5.1 ΕΦΑΡΜΟΓΗ ΜΟΝΤΕΛΟΥ HOLT - WINTERS EXPONENTIAL SMOOTHING.....	102
5.2 ΕΦΑΡΜΟΓΗ ΜΟΝΤΕΛΟΥ SARIMAX.....	106
5.3 ΕΦΑΡΜΟΓΗ ΜΟΝΤΕΛΟΥ LSTM.....	110
ΚΕΦΑΛΑΙΟ 6. ΣΥΝΟΨΗ ΜΕΛΕΤΗΣ ΚΑΙ ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ	114
ΒΙΒΛΙΟΓΡΑΦΙΑ – ΠΑΡΑΠΟΜΠΕΣ	117

Εισαγωγή

Η ατμοσφαιρική ρύπανση αποτελεί έναν από τους σημαντικότερους περιβαλλοντικούς και υγειονομικούς κινδύνους του αιώνα που διανύουμε. Ο Παγκόσμιος Οργανισμός Υγείας εκτιμά ότι εκατομμύρια πρόωροι θάνατοι κάθε χρόνο σχετίζονται με την έκθεση σε ρύπους, υπογραμμίζοντας μάλιστα τις μακροχρόνιες επιπτώσεις στην ανθρώπινη υγεία, την κλιματική ισορροπία, τη γεωργία και την οικονομία.

Ανάμεσα στους σημαντικότερους ατμοσφαιρικούς ρύπους συγκαταλέγεται το διοξείδιο του θείου (SO_2) που θα αποτελέσει και το βασικό στοιχείο της μελέτης μας. Πρόκειται για ένα αέριο που εκπέμπεται στην ατμόσφαιρα τόσο από ανθρωπογενείς πηγές όπως η καύση ορυκτών καυσίμων και η βιομηχανική δραστηριότητα όσο και από φυσικές πηγές όπως τα ηφαίστεια και τα θαλάσσια συστήματα. Στην ατμόσφαιρα, μετασχηματίζεται, μέσω χημικών αντιδράσεων, σε θειικό οξύ (H_2SO_4) και θειικά άλατα, τα οποία αποτελούν βασικά συστατικά των λεπτών αιωρούμενων σωματιδίων ($\text{PM}_{2.5}$). Οι ενώσεις αυτές, έχουν άμεσες και σοβαρές επιπτώσεις στην υγεία, στην εμφάνιση όξινης βροχής, στην οξίνιση εδαφών και υδάτων, στη διάβρωση κατασκευών και στην αλλοίωση του ισοζυγίου ακτινοβολίας της γης.

Η παρούσα διπλωματική εργασία εστιάζει στη μελέτη των συγκεντρώσεων διοξειδίου του θείου, με σκοπό την κατανόηση των εποχικών μοτίβων (seasonality), της μακροχρόνιας τάσης (trend), της βραχυπρόθεσμης και μη συστηματικής μεταβλητότητας (residuals), καθώς και στη διερεύνηση τυχόν συσχετίσεων με εξωγενείς παράγοντες και άλλους ατμοσφαιρικούς ρύπους. Παράλληλα, αναπτύσσονται μοντέλα πρόβλεψης που δύναται να χρησιμοποιηθούν για την πρόβλεψη της πορείας των συγκεντρώσεων διοξειδίου του θείου, παρέχοντας έτσι ένα σημαντικό μηχανισμό εξεύρεσης των κατάλληλων μέτρων για τον περιορισμό του.

Δομή της Διπλωματικής Εργασίας

Η εργασία αναπτύσσεται σε έξι κεφάλαια:

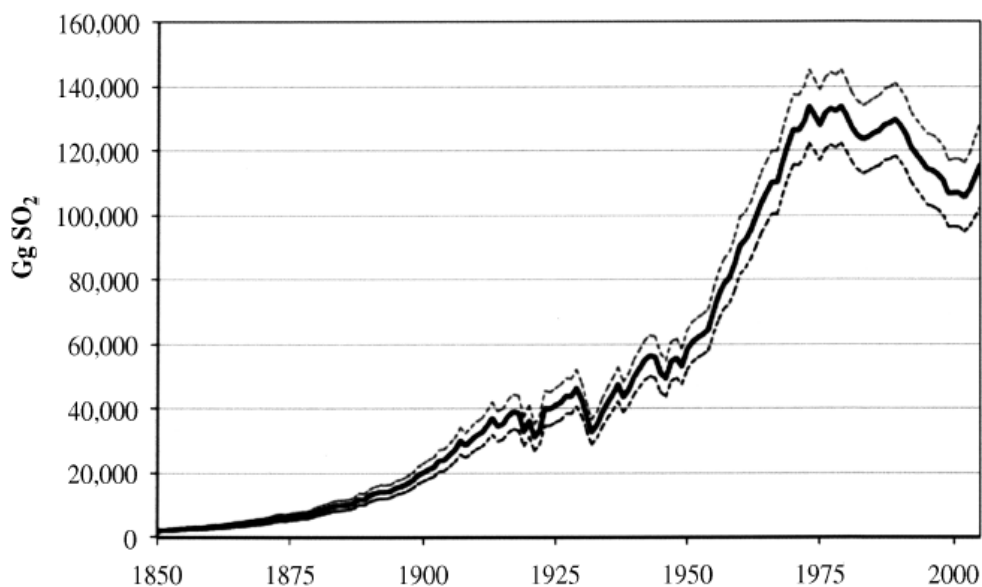
- Κεφάλαιο 1: Εισαγωγή - παρουσίαση του διοξειδίου του θείου, οι χημικές του ιδιότητες, η χρησιμότητά του και οι επιπτώσεις που προκαλεί τόσο στο περιβάλλον όσο και στην υγεία.
- Κεφάλαιο 2: Το διοξείδιο του θείου ως έμμεσο αέριο του θερμοκηπίου, η κλιματική επίδρασή του καθώς και η σχέση του με άλλα αέρια του θερμοκηπίου.
- Κεφάλαιο 3: Παρουσίαση των μεθόδων ανάλυσης και πρόβλεψης χρονοσειρών, όπου παρατίθενται εκτενώς το θεωρητικό υπόβαθρο και η μεθοδολογία που θα ακολουθήσουμε στη μελέτη μας.
- Κεφάλαιο 4: Το βασικό κεφάλαιο της μελέτης μας, με τη σε βάθος ανάλυση της συμπεριφοράς του διοξειδίου του θείου, τα εποχικά μοτίβα, την τάση και τις διαφορές μεταξύ των πόλεων της Ευρώπης και των ΗΠΑ. Πρόσθετα, διερευνάται η σχέση του με μετεωρολογικούς παράγοντες καθώς και με το όζον.
- Κεφάλαιο 5: Παρουσιάζονται τρεις διαφορετικοί μέθοδοι πρόβλεψης της πορείας των συγκεντρώσεων διοξειδίου του θείου, και παρατίθενται συγκριτικά τα αποτελέσματα από την εφαρμογή τους.
- Κεφάλαιο 6: Σύντομη παρουσίαση της μελέτης μας και παράθεση γενικών συμπερασμάτων.

Βασικός μας στόχος, είναι να παραθέσουμε με φυσικό και κατανοητό λόγο τα αποτελέσματα της μελέτης μας, συμπεριλαμβάνοντας όπου χρειάζεται πίνακες με τα στοιχεία της ανάλυσής μας, καθώς και σχετικά διαγράμματα. Το πρωτογενές υλικό, ο κώδικας που αναπτύχθηκε και τα αναλυτικά αποτελέσματα θα είναι διαθέσιμα σε κάθε ενδιαφερόμενο για επισκόπηση και τυχόν αξιολόγηση.

Κεφάλαιο 1: Το Διοξείδιο του Θείου (SO₂)

1.1 Εισαγωγή

Το διοξείδιο του θείου σε κανονικές συνθήκες, είναι ένα άχρωμο τοξικό αέριο με έντονη οσμή και ανήκει στην οικογένεια των οξειδίων του θείου (SO_x). Παρά την απλή μοριακή του δομή του, διαδραματίζει κεντρικό ρόλο στη χημεία της ατμόσφαιρας, στην περιβαλλοντική υποβάθμιση και κατ' επέκταση στη δημόσια υγεία [1]. Παράγεται με φυσικό τρόπο από την ηφαιστειακή δραστηριότητα και τις εκπομπές από τους ωκεανούς, δηλαδή, τη φυσική απελευθέρωση οργανικών ενώσεων θείου από τη θάλασσα προς την ατμόσφαιρα, οι οποίες, στη συνέχεια, μετατρέπονται σε διοξείδιο του θείου και τελικά σε θειικά αερολύματα. Μετά τη Βιομηχανική Επανάσταση, οι ανθρωπογενείς εκπομπές από την καύση ορυκτών καυσίμων κυριάρχησαν, καθιστώντας το έναν από τους βασικούς και πλέον μελετημένους ρυθμιζόμενους ρύπους [2].



Γράφημα 1: Παγκόσμιες εκπομπές SO₂ από καύση ορυκτών καυσίμων και άλλες χημικές διεργασίες [2].

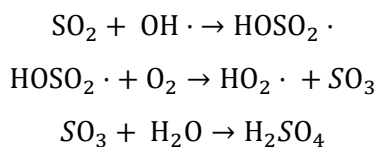
Η περιβαλλοντική υποβάθμιση αναφέρεται στην αλλοίωση του φυσικού περιβάλλοντος και στην περίπτωση του διοξειδίου του θείου, εκδηλώνεται με διάφορους τρόπους, όπως,

- την πρόκληση όξινης βροχής, η οποία οξινίζει τα εδάφη και τα υδάτινα οικοσυστήματα, οδηγώντας σε απώλεια βιοποικιλότητας και σημαντικές βλάβες στη βλάστηση και τις καλλιέργειες,
- την επιτάχυνση της διάβρωσης των κατασκευών και γενικά των δομικών υλικών,
- τη μείωση της ορατότητας λόγω σχηματισμού θεικών αερολυμάτων στην ατμόσφαιρα και
- την αλλοίωση του ισοζυγίου ακτινοβολίας της γης, δηλαδή τη σχέση μεταξύ της ηλιακής ακτινοβολίας που εισέρχεται στην ατμόσφαιρα και της υπέρυθρης ακτινοβολίας που εκπέμπεται προς το διάστημα.

Γενικά, η παρουσία του διοξειδίου του θείου στην ατμόσφαιρα συνδέεται στενά με τη μείωση της ποιότητας ζωής και την υποβάθμιση του περιβάλλοντος, κυρίως μέσω των προϊόντων μετασχηματισμών του, στα οποία θα αναφερθούμε εκτενώς παρακάτω.

1.2 Φυσικές και Χημικές Ιδιότητες

Το διοξείδιο του θείου είναι μια ανόργανη χημική ένωση με μοριακό τύπο SO₂ και μοριακή μάζα 64.06 g/mol. Έχει σημείο βρασμού τους -10°C και υγροποιείται εύκολα υπό πίεση. Το σημείο τήξης του είναι -75.5°C και έχει πυκνότητα περίπου 2.26 φορές αυτή του αέρα. Είναι ιδιαίτερα διαλυτό στο νερό και συμμετέχει ενεργά σε διάφορες ατμοσφαιρικές διεργασίες. Οξειδώνεται δε σε τριοξείδιο του θείου (SO₃), το οποίο με τη σειρά του αντιδρά με υδρατμούς σχηματίζοντας έτσι το θειικό οξύ (H₂SO₄) και συγκεκριμένα μέσω των κάτωθι χημικών αντιδράσεων:



Σημειώνουμε πως το 'OH·' είναι η ρίζα του υδροξυλίου (hydroxyl radical), όπου η τελεία '·' στο τέλος του στοιχείου, δηλώνει ότι το μόριο έχει μονήρες (ασύζευκτο) ηλεκτρόνιο, δηλαδή είναι ελεύθερη ρίζα. Οι ρίζες είναι εξαιρετικά δραστικές, γιατί προσπαθούν, υπό μία έννοια, να συμπληρώσουν το ηλεκτρόνιο τους.

Οι αντιδράσεις αυτές λαμβάνουν χώρα σε χρονικές κλίμακες από ώρες μέχρι ημέρες, ανάλογα βέβαια τις εκάστοτε μετεωρολογικές συνθήκες [3]. Σε περιβάλλοντα νεφοσταγόνων ή ομίχλης, όπου κυριαρχεί η υδατική φάση, η οξειδωση του διοξειδίου του θείου επιταχύνεται σημαντικά, με τα κυριότερα οξειδωτικά στοιχεία να είναι το υπεροξείδιο του υδρογόνου (H₂O₂) και το όζον (O₃), οδηγώντας τελικά στο σχηματισμό θειικού οξέος (H₂SO₄).

Συγκεκριμένα, αυτό επιτυγχάνεται μέσω,

α) οξειδωσης από υπεροξείδιο του υδρογόνου: $\text{SO}_2 (\text{aq}) + \text{H}_2\text{O}_2 \rightarrow \text{H}_2\text{SO}_4$ και

β) οξειδωσης από όζον: $\text{SO}_2 (\text{aq}) + \text{O}_3 \rightarrow \text{H}_2\text{SO}_4$

Οι συγκεντρώσεις του διοξειδίου του θείου στην ατμόσφαιρα εκφράζονται με διαφορετικές μονάδες μέτρησης, ανάλογα με το πλαίσιο εφαρμογής τους. Για την υγεία, χρησιμοποιούνται συνήθως μικρογραμμάρια ανά κυβικό μέτρο (μg/m³), με τον Παγκόσμιο Οργανισμό Υγείας να θέτει ως όριο της ημερήσιας μέσης συγκέντρωσης τα 40 μg/m³ [4]. Στη χημεία της ατμόσφαιρας και σε δίκτυα παρακολούθησης, συχνά χρησιμοποιούνται αναλογίες κατ' όγκο, όπως ppb, ενώ σε δορυφορικές μετρήσεις η συγκέντρωση εκφράζεται σε Dobson Units (DU), που αντιστοιχούν σε 2.69 × 10¹⁶ μόρια/cm². Σε σύγκριση με τις μονάδες συγκέντρωσης μg/m³, 1 DU ισοδυναμεί περίπου με 28.5 μg/m³, όταν το διοξείδιο του θείου κατανέμεται σε στρώμα 1 km πάχους στην κατώτερη ατμόσφαιρα, ενώ η τιμή μειώνεται αντιστρόφως ανάλογα με το ύψος του στρώματος κατανομής. Δηλαδή, εάν είναι κατανεμημένο σε μεγαλύτερο ύψος (πχ. 5 - 10 km), η πραγματική συγκέντρωση ανά κυβικό μέτρο είναι μικρότερη. Έτσι, η αντιστοιχία DU → μg/m³ δεν είναι μοναδική αλλά εξαρτάται πάντοτε από την κατακόρυφη κατανομή του.

Ενδεικτικά,

Πάχος στρώματος κατανομής	Συγκέντρωση (μg/m ³)
1 km	28.5 μg/m ³
2 km	14.3 μg/m ³
5 km	5.7 μg/m ³
10 km	2.9 μg/m ³

1.3 Φυσικές και Ανθρωπογενείς Πηγές

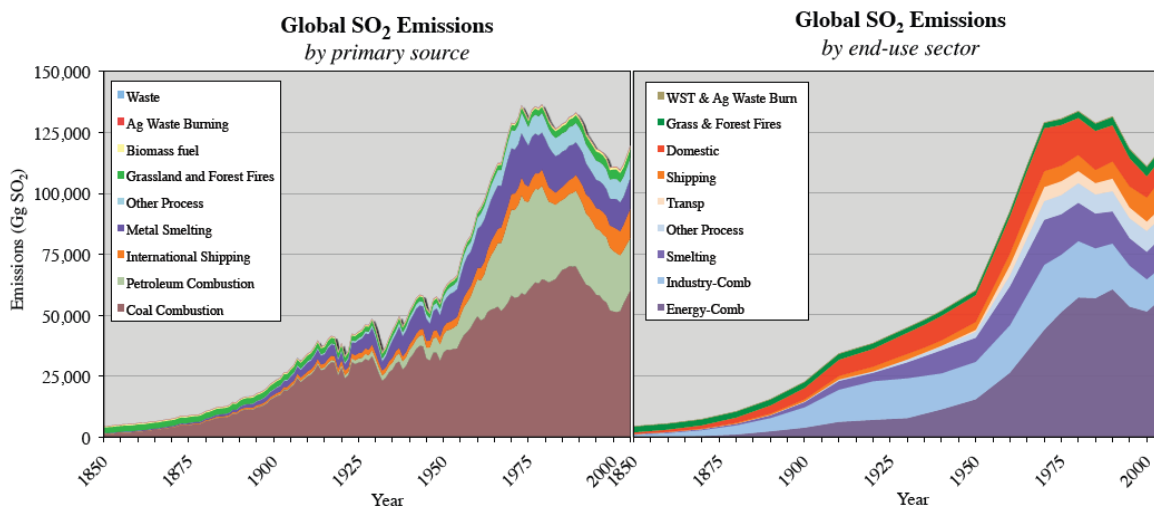
Το διοξείδιο του θείου προέρχεται κυρίως από ανθρωπογενείς δραστηριότητες που συνδέονται με την καύση ορυκτών καυσίμων, όπως ο άνθρακας και το πετρέλαιο, που περιέχουν θείο. Η χρήση αυτών των καυσίμων σε θερμοηλεκτρικούς σταθμούς, βιομηχανικές εγκαταστάσεις και μέσα μεταφοράς αποτελεί τη βασικότερη πηγή εκπομπών του. Σημαντική συμβολή έχουν επίσης οι διεργασίες τήξης μεταλλευμάτων που περιέχουν θειούχες ενώσεις (πχ. χαλκός, μόλυβδος, ψευδάργυρος). Πέρα από τις ανθρωπογενείς πηγές, το διοξείδιο του θείου απελευθερώνεται και από φυσικές διεργασίες, με κυριότερη την ηφαιστειακή δραστηριότητα, η οποία μπορεί να εκλύσει τεράστιες ποσότητες αερίου στην ατμόσφαιρα σε σύντομο χρονικό διάστημα, επηρεάζοντας τόσο την τοπική όσο και την παγκόσμια ποιότητα του αέρα. Σε μικρότερη κλίμακα, το διοξείδιο του θείου μπορεί να παραχθεί και από τη φυσική αποσύνθεση οργανικής ύλης και μικροβιακές διεργασίες στο έδαφος.

Φυσικές Πηγές:

- **Ηφαίστεια:** Εκπέμπουν τεράστιες ποσότητες SO₂ κατά τις εκρήξεις αλλά και μέσω της συνεχούς εκπομπής αερίων [5]. Το μάγμα περιέχει διαλυμένο υδρόθειο (H₂S), SO₂, καθώς και άλλα θειούχα αέρια. Όταν το μάγμα της λάβας ανεβαίνει προς την επιφάνεια, η πίεσή του μειώνεται (αποπίεση) και το υδρόθειο έρχεται σε επαφή με το οξυγόνο (O₂) της ατμόσφαιρας. Έτσι, το υδρόθειο οξειδώνεται, δηλαδή χάνει τα ηλεκτρόνια του και μετατρέπεται σε διοξείδιο του θείου και νερό, μέσω της αντίδρασης $H_2S + 1.5 O_2 \rightarrow SO_2 + H_2O$.
- **Ωκεανοί και υγρότοποι:** Εκλύουν ενώσεις θείου, όπως διμεθυλοσουλφίδιο (DMS), οι οποίες οξειδώνονται στην ατμόσφαιρα σε SO₂ και στη συνέχεια σε θειικά άλατα [6].

Ανθρωπογενείς Πηγές:

- **Καύση ορυκτών καυσίμων:** Ο λιγνίτης, ο άνθρακας και το πετρέλαιο περιέχουν οργανικό και ανόργανο θείο, το οποίο κατά την καύση απελευθερώνεται ως διοξείδιο του θείου. Οι μονάδες ηλεκτροπαραγωγής και τα βιομηχανικά καύσιμα αποτελούν ιστορικά τις κυριότερες πηγές SO₂ [2].
- **Βιομηχανικές διεργασίες:** Η μεταλλουργία μη σιδηρούχων μετάλλων, τα διυλιστήρια πετρελαίου και η παραγωγή τσιμέντου εκπέμπουν σημαντικές ποσότητες διοξειδίου του θείου [7]. Ο πυρίτης, δηλαδή θειούχος σίδηρος (FeS₂) είναι ένα ορυκτό που βρίσκεται σε κάρβουνα, μεταλλεύματα, σε πετρώματα, ακόμη και σε ηφαίστεια. Όταν ο πυρίτης θερμανθεί ή εκτεθεί σε οξυγόνο, σε υψηλές θερμοκρασίες, τότε οξειδώνεται, μέσω της αντίδρασης, $4 FeS_2 + 11 O_2 \rightarrow 2 Fe_2O_3 + 8 SO_2$. Ουσιαστικά, το θείο του πυρίτη οξειδώνεται σε διοξείδιο του θείου (SO₂), ενώ ο σίδηρος (Fe) οξειδώνεται σε οξείδιο του σιδήρου (Fe₂O₃) ή αλλιώς αιματίτη.
- **Ναυτιλία:** Τα πλοία που χρησιμοποιούν για την καύση τους βαριά καύσιμα παραμένουν σημαντική πηγή SO₂, ιδιαίτερα κοντά σε λιμάνια και παράκτιες ζώνες [8].
- **Μέσα μεταφοράς:** Τα παλαιότερα οχήματα με βενζίνη και κυρίως πετρέλαιο παρήγαγαν σημαντικές εκπομπές διοξειδίου του θείου λόγω της υψηλής περιεκτικότητας θείου στα καύσιμα. Αν και η χρήση καυσίμων χαμηλού θείου έχει μειώσει σημαντικά τις εκπομπές στην Ευρώπη και τις ΗΠΑ, σε πολλές αναπτυσσόμενες χώρες τα οδικά μέσα παραμένουν σημαντική πηγή ατμοσφαιρικής ρύπανσης από διοξείδιο του θείου [4], [9].

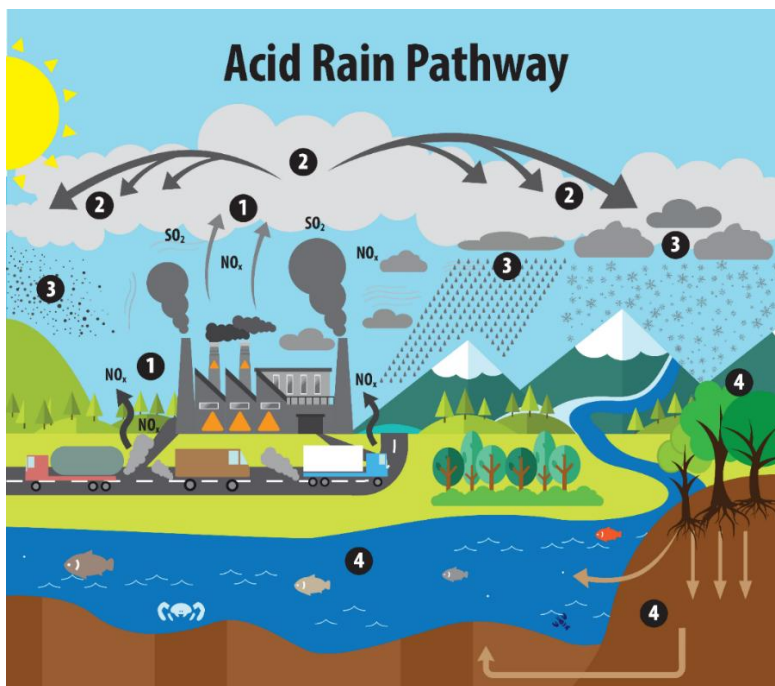


Γράφημα: Πηγές διοξειδίου του θείου και βασικοί τομείς χρήσης (Smith et al, 2011)

1.4 Ατμοσφαιρικοί Μετασχηματισμοί και Περιβαλλοντικές Επιπτώσεις

Η παρουσία διοξειδίου του θείου στην ατμόσφαιρα συμβάλλει σε σημαντικές περιβαλλοντικές διεργασίες, ήτοι, **στη δημιουργία όξινης απόθεσης (Acid Deposition)**, όπου το εκλυόμενο διοξείδιο του θείου όταν έρχεται σε επαφή με υδρατμούς, οξυγόνο και άλλους οξειδωτικούς παράγοντες, μετασχηματίζεται σε θειικό οξύ (H₂SO₄). Ο όρος 'όξινη απόθεση' (acid deposition) περιγράφει τη διαδικασία μέσω της οποίας όξινα συστατικά, όπως το θειικό οξύ και το νιτρικό οξύ (HNO₃) που σχηματίζονται από την οξείδωση του SO₂ και των NO_x αντίστοιχα, αποτίθενται από την ατμόσφαιρα στην επιφάνεια της Γης. Η απόθεση αυτή μπορεί να είναι είτε υγρή (όξινη βροχή, χιόνι, ομίχλη) είτε ξηρή (σωματίδια και αέρια που προσκολλώνται απευθείας σε εδάφη, φυτά και κατασκευές).

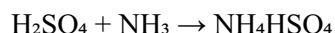
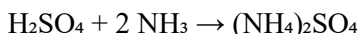
Το φαινόμενο αυτό έχει συνδεθεί με σοβαρές επιπτώσεις στα υδάτινα οικοσυστήματα, τα δάση, τα εδάφη, αλλά και στην υποβάθμιση κτιρίων και μνημείων, καθιστώντας την όξινη απόθεση ένα μείζον διασυννοριακό, μεγάλης κλίμακας, πρόβλημα ατμοσφαιρικής ρύπανσης



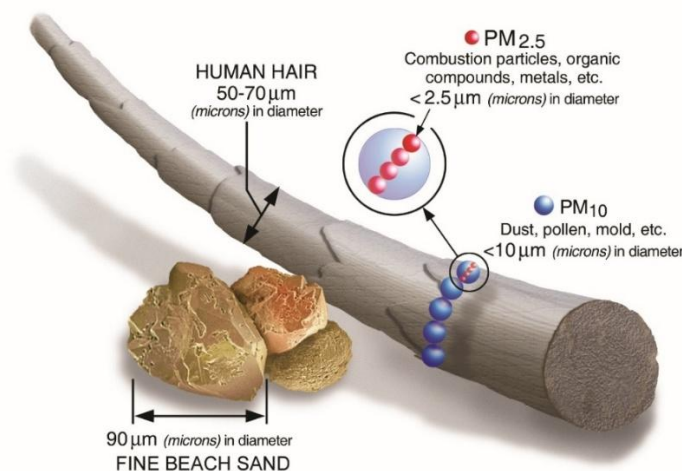
Εικόνα 1: Σχηματική απεικόνιση του μηχανισμού σχηματισμού της όξινης βροχής. Οι εκπομπές SO₂ και NO_x απελευθερώνονται στην ατμόσφαιρα, μετασχηματίζονται σε όξινα σωματίδια τα οποία δύνανται να μεταφερθούν σε μεγάλες αποστάσεις και τελικά επιστρέφουν στην επιφάνεια της Γης, μέσω υγρής ή ξηρής εναπόθεσης, προκαλώντας αρνητικές επιπτώσεις σε εδάφη, δάση, υδάτινα οικοσυστήματα και υποδομές (πηγή: <https://www.epa.gov/acidrain/what-acid-rain>).

και

στη δημιουργία εξαιρετικά λεπτών σωματιδίων (PM_{2.5}), όπου το διοξείδιο του θείου, είτε προέρχεται από φυσικές είτε από ανθρωπογενείς πηγές, απελευθερώνεται στην ατμόσφαιρα και υπόκειται σε διαδοχικές οξειδωτικές διεργασίες, με κυριότερη αυτών τη μετατροπή του σε θειικό οξύ. Το τελευταίο δεν παραμένει σε αέρια μορφή, καθώς είναι ιδιαίτερα υγροσκοπικό, αλλά απορροφά υδατμούς και είτε συμπυκνώνεται πάνω σε ήδη υπάρχοντα ατμοσφαιρικά σωματίδια είτε οδηγεί στη δημιουργία νέων πυρήνων συμπύκνωσης. Στην ατμόσφαιρα, η παρουσία αμμωνίας (NH₃), κυρίως από γεωργικές δραστηριότητες, επιδρά καθοριστικά στη χημεία των προϊόντων της οξείδωσης του SO₂, όπου αντιδρά με το θειικό οξύ σχηματίζοντας άλατα, όπως το θειικό αμμώνιο [(NH₄)₂SO₄] και το όξινο θειικό αμμώνιο (NH₄HSO₄), σύμφωνα με τις αντιδράσεις:



Τα θειικά άλατα αυτά, αποτελούν την κυριότερη ανόργανη συνιστώσα των λεγόμενων δευτερογενών ανόργανων αιωρούμενων σωματιδίων (secondary aerosols), που ταξινομούνται στην κατηγορία των σωματιδίων PM_{2.5} με διάμετρο < 2.5 μm (Particulate Matter < 2.5 μm).



Εικόνα 2: Απεικόνιση διείσδυσης σωματιδίων PM_{2.5} στην τρίχα της κεφαλής.

Στο σημείο αυτό, να διευκρινίσουμε ότι τα PM_{2.5} δεν αποτελούν ένα ομοιογενές υλικό, αλλά μίγμα από θειικά άλατα, νιτρικά άλατα, οργανικά σωματίδια, μαύρο άνθρακα, σκόνη και μέταλλα. Σύμφωνα με μετρήσεις [10], [11], σε αστικά και βιομηχανικά περιβάλλοντα τα θειικά συστατικά μπορεί να αποτελέσουν μέχρι και το 40% του PM_{2.5}. Σε περιοχές με έντονη καύση άνθρακα και πετρελαίου, τα θειικά σωματίδια ξεπερνούν ακόμη και το 50% του PM_{2.5}. Αντίθετα, σε αγροτικές περιοχές, με χαμηλές εκπομπές διοξειδίου του θείου, η περιεκτικότητα του PM_{2.5} σε θειικά σωματίδια είναι σαφώς μικρότερη.

Ο λόγος αναφοράς μας στα σωματίδια αυτά, εξηγείται εκτενώς παρακάτω.

1.5 Επιπτώσεις στην Υγεία

Το διοξείδιο του θείου και τα δευτερογενή προϊόντα του, κυρίως τα θειικά αερολύματα (sulfate aerosols), τα οποία αποτελούν σημαντικό κλάσμα των αιωρούμενων σωματιδίων PM_{2.5}, έχουν αποδεδειγμένα σοβαρές επιπτώσεις στη δημόσια υγεία. Μάλιστα, ο Παγκόσμιος Οργανισμός Υγείας αναγνωρίζει το διοξείδιο του θείου ως έναν από τους πλέον επικίνδυνους ατμοσφαιρικούς ρύπους, καθώς συμβάλει στην πρόκληση τόσο οξείας όσο και χρόνιας επιπτώσεων στην ανθρώπινη υγεία [4].

Αναλυτικότερα,

Οξείες επιπτώσεις (βραχυχρόνια έκθεση): Η βραχυχρόνια αύξηση των συγκεντρώσεων διοξειδίου του θείου, έχει συσχετιστεί με αναπνευστικά προβλήματα, όπως βρογχόσπασμα, ερεθισμό των αεραγωγών και έξαρση των συμπτωμάτων άσθματος. Η ανάλυση 225 μελετών κατέδειξε ότι ακόμα και βραχυπρόθεσμες αυξήσεις μπορούν να επιδεινώσουν περαιτέρω υφιστάμενα προβλήματα υγείας, ειδικά σε ευάλωτους πληθυσμούς όπως παιδιά, ηλικιωμένους και ασθενείς με ΧΑΠ [12], [13].

Μεγάλης κλίμακας διεθνής μελέτη σε 399 πόλεις από 23 χώρες, κατέδειξε ότι η ημερήσια αύξηση του διοξειδίου του θείου σχετίζεται με 0.5% αύξηση στη συνολική θνησιμότητα (all-cause mortality) [14]. Παρόλο που τα επίπεδα εκπομπών έχουν μειωθεί σε πολλές περιοχές τις τελευταίες δεκαετίες, η συσχέτιση αυτή παραμένει στατιστικά σημαντική.

Χρόνιες επιπτώσεις (μακροχρόνια έκθεση): Η παρατεταμένη έκθεση σε διοξείδιο του θείου συνδέεται με τον αυξημένο κίνδυνο χρόνιας αναπνευστικής και καρδιαγγειακής νοσηρότητας. Σε πληθυσμιακές μελέτες έχει παρατηρηθεί αύξηση περιστατικών άσθματος, μείωση της πνευμονικής λειτουργίας και αυξημένες εισαγωγές σε νοσοκομεία εξαιτίας αυτών [15]. Ορισμένες μελέτες μάλιστα υποδεικνύουν πιθανές επιπτώσεις στην εγκυμοσύνη, όπως χαμηλό βάρος γέννησης του βρέφους και αυξημένη βρεφική θνησιμότητα, αν και τα ευρήματα αυτά χρήζουν περαιτέρω διερεύνησης [12].

Έμμεσες επιπτώσεις, μέσω σωματιδίων PM_{2.5}: Ο σχηματισμός δευτερογενών θεικών σωματιδίων (sulfate particles), τα οποία αποτελούν ένα από τα κυρίαρχα συστατικά των αιωρούμενων σωματιδίων με διάμετρο

μικρότερη από 2.5 μm (PM_{2.5}), έχει σημαντικές επιπτώσεις στην υγεία, καθώς τα σωματίδια αυτά είναι ικανά να διεισδύουν βαθιά στους πνεύμονες και να φτάνουν ακόμη και στην κυκλοφορία του αίματος. Σύμφωνα με μελέτες, η έκθεση σε τέτοια σωματίδια σχετίζεται με αυξημένο κίνδυνο εμφάνισης καρδιαγγειακών παθήσεων, εγκεφαλικών επεισοδίων και πρόωρης θνησιμότητας [16], [17].

Ουσιαστικά, ενεργοποιούν το ανοσοποιητικό σύστημα προκαλώντας συστηματική φλεγμονή, ενώ με τη διείσδυση στους πνεύμονες, ενεργοποιούν τα μακροφάγα (κύτταρα του ανοσοποιητικού συστήματος που απομακρύνουν νεκρά κύτταρα, μικρόβια κλπ. μέσω της φαγοκυττάρωσης) και άλλα κύτταρα του ανοσοποιητικού, κάτι που προκαλεί την απελευθέρωση φλεγμονωδών μεσολαβητών (κυτοκίνες ή κυτταροκίνες όπως IL-6, TNF-α). Το αποτέλεσμα αυτής της διαδικασίας είναι η δημιουργία χρόνιας φλεγμονής όχι μόνο στους πνεύμονες αλλά και σε ολόκληρο τον οργανισμό, κάτι που συνδέεται σαφώς και με τα καρδιαγγειακά νοσήματα [16].

Παράλληλα, τα εν λόγω σωματίδια προάγουν την εμφάνιση οξειδωτικού στρες (oxidative stress), μέσω της υπερβολικής παραγωγής δραστικών μορφών οξυγόνου (ROS – Reactive Oxygen Species), όπως το υπεροξειδίο και η ρίζα υδροξυλίου, διαδικασία που οδηγεί σε βλάβες στις κυτταρικές δομές. Και τούτο διότι, περιέχουν μέταλλα και θειικά ιόντα που ευνοούν τη δημιουργία αυτών των δραστικών μορφών οξυγόνου. Μάλιστα, όταν η παραγωγή τους υπερβαίνει την αντιοξειδωτική άμυνα του οργανισμού, προκαλείται βλάβη σε κυτταρικές δομές, όπως στις μεμβράνες, τα μιτοχόνδρια, ακόμη και στο DNA, οδηγώντας έτσι σε κυτταρική δυσλειτουργία [18].

Μία ακόμη σημαντική επίπτωση είναι η διείσδυση των σωματιδίων PM_{2.5} στην κυκλοφορία του αίματος, η οποία συμβάλλει σε ενδοθηλιακή δυσλειτουργία και αυξημένο κίνδυνο θρόμβωσης, αποτελώντας βασικό πρόδρομο ανάπτυξης καρδιαγγειακών νοσημάτων [19]. Παράλληλα, οι διαταραχές του ανοσοποιητικού συστήματος που προκαλούνται από την έκθεση σε PM_{2.5}, σχετίζονται με αυξημένη ευπάθεια σε λοιμώξεις [20]. Νεότερες έρευνες δείχνουν ότι τα σωματίδια με διάμετρο μικρότερη του 0.1 μm, μπορούν να έχουν νευροτοξικές επιδράσεις, καθώς, εξαιτίας του μεγέθους τους, μπορούν να φτάσουν ακόμη και στον εγκέφαλο, συμβάλλοντας έτσι σε γνωστική έκπτωση και σε νευροεκφυλιστικές νόσους, όπως η νόσος Alzheimer και η νόσος Parkinson [4], [21], [22].

1.6 Χρήσεις του SO₂

Παρά τις αρνητικές επιπτώσεις του στην υγεία και το περιβάλλον, το διοξείδιο του θείου έχει διαχρονικά σημαντική βιομηχανική και οικονομική αξία. Οι βασικές του χρήσεις είναι οι εξής:

Παραγωγή θειικού οξέος (H₂SO₄): Η κυριότερη και πιο εκτεταμένη χρήση του είναι ως πρόδρομη ένωση για την παραγωγή θειικού οξέος. Το θειικό οξύ αποτελεί το χημικό με τη μεγαλύτερη παγκόσμια παραγωγή, με πάνω από 200 εκατομμύρια τόνους ετησίως [23]. Χρησιμοποιείται δε στην παραγωγή λιπασμάτων (κυρίως υπερφωσφορικών), παρασκευή απορρυπαντικών και χημικών διεργασιών (πχ. βαφές, φαρμακευτικά), καθώς και στη μεταλλουργία, για την εκχύλιση μετάλλων όπως χαλκός, ψευδάργυρος και νικέλιο.

Βιομηχανία τροφίμων και ποτών: Το διοξείδιο του θείου χρησιμοποιείται εδώ και χρόνια ως συντηρητικό και αντιοξειδωτικό σε αποξηραμένα φρούτα και λαχανικά, όπου αποτρέπει την ανάπτυξη μικροοργανισμών, στο κρασί και στους χυμούς, όπου εμποδίζει την οξείδωση των πολυφαινόλων που οδηγούν σε αλλοίωση της γεύσης και του χρώματος, καθώς και στη ζυθοποιία, όπου χρησιμοποιείται σε ορισμένες περιπτώσεις για τον έλεγχο της ανάπτυξης μικροβίων.

Βέβαια, η χρήση του είναι αυστηρά ρυθμιζόμενη, αφού σε υψηλές συγκεντρώσεις μπορεί να προκαλέσει αλλεργικές αντιδράσεις ή ακόμη και κρίσεις άσθματος σε ευαίσθητα άτομα [24].

Παλαιότερα, χρησιμοποιούνταν και σε ψυκτικά συστήματα (αρχές του 20ού αιώνα), ως ψυκτικό αέριο, όμως αυτή η χρήση του εγκαταλείφθηκε λόγω της τοξικότητάς του και αντικαταστάθηκε από το γνωστό φρέον και πρόσφατα από άλλα πιο ασφαλή ψυκτικά μέσα.

Τα παραπάνω εξηγούν γιατί το διοξείδιο του θείου, παρά τις βλαβερές του επιπτώσεις τόσο στην υγεία όσο και στο περιβάλλον, παραμένει πολύτιμο και δημοφιλές στην παγκόσμια βιομηχανία.

Για να αντιληφθεί κανείς το πραγματικό μέγεθος της βιομηχανικής αξιοποίησης του SO₂, μέσω της εκτενούς παραγωγής H₂SO₄, το έτος 2011-2012, η παγκόσμια παραγωγή θειικού οξέος υπολογίστηκε σε περίπου 231 εκατ. τόνους [25] και σύμφωνα με πιο πρόσφατη ανάλυση της αγοράς, η παγκόσμια παραγωγή H₂SO₄, ανήλθε σε 261

εκ. τόνους [26]. Η αξία αγοράς του το 2022 ανήλθε σε USD 13,39 δισεκατομμύρια, με προβλεπόμενη αύξηση σε 17,94 δισεκατομμύρια έως το 2030 [27], [28].

1.7 Ιστορικό Πλαίσιο και Ρυθμίσεις

Η ατμοσφαιρική ρύπανση από το διοξείδιο του θείου κορυφώθηκε τον 20ό αιώνα στις βιομηχανικές περιοχές της Ευρώπης και της Βόρειας Αμερικής, όταν η μαζική καύση άνθρακα για θέρμανση και βιομηχανία, παρήγαγε υψηλές συγκεντρώσεις αιωρούμενων σωματιδίων και SO₂. Το πιο γνωστό παράδειγμα είναι η ‘Μεγάλη Ομίχλη’ του Λονδίνου το 1952, η οποία μέσα σε λίγες ημέρες προκάλεσε περισσότερους από 4.000 θανάτους, κυρίως λόγω καρδιοαναπνευστικών επιπλοκών, με μεταγενέστερες αναλύσεις να ανεβάζουν τον αριθμό των θανάτων σε 12.000 [29]. Παρόμοια επεισόδια παρατηρήθηκαν και σε άλλες πόλεις, όπως το Donora των ΗΠΑ το 1948, όπου ένα επεισόδιο ‘καπνομίχλης (smog)’ σκότωσε τουλάχιστον 20 άτομα ενώ χιλιάδες ήταν εκείνοι που αρρώστησαν [30].



Εικόνα 3: Η Donora Pennsylvania (29-10-1948) κατά τη διάρκεια του επεισοδίου της ατμοσφαιρικής ρύπανσης (πηγή: Jacobson Z. M., 2002, «Atmospheric pollution : History, Science and Regulation», Cambridge University Press, New York, ISBN-13 978-0-521-81171-2.)



Εικόνα 4: Αιθαλομίχλη στο Λονδίνο (Δεκ. 1952) (πηγή: Ahrens C. D, Henson R., 2016, «Meteorology Today: An Introduction to Weather, Climate, and the Environment», 11th Edition, Cengage Learning, 20 Channel Center Street, Boston, MA 02210, USA, ISBN-13 : 978-1-305-11358-9)

Αυτά τα δυσάρεστα περιστατικά αποτέλεσαν την αφετηρία για μια σειρά νομοθετικών παρεμβάσεων. Στις ΗΠΑ, ο ‘Clean Air Act’ του 1970 και οι μεταγενέστερες τροποποιήσεις του, καθιέρωσαν συγκεκριμένα όρια εκπομπών διοξειδίου του θείου, οδηγώντας στην ανάπτυξη τεχνολογιών αποθείωσης καυσαερίων FGD (Fuel Gas Desulfurization), περιορίζοντας έτσι σημαντικά τις εκπομπές τουλάχιστον από σταθμούς ηλεκτροπαραγωγής.

Στην Ευρώπη, σημαντικό βήμα αποτέλεσε η οδηγία για τις ‘Μεγάλες Εγκαταστάσεις Καύσης’ [31], η οποία επέβαλε αυστηρά όρια στις εκπομπές SO₂, NO_x και άλλων σωματιδίων. Οι πολιτικές αυτές, σε συνδυασμό με τη μετάβαση σε καύσιμα χαμηλής περιεκτικότητας σε θείο (όπως πχ φυσικό αέριο), οδήγησαν σε δραστική μείωση των εκπομπών διοξειδίου του θείου κατά τις τελευταίες δεκαετίες [9]. Πιο συγκεκριμένα, στοιχεία του European Environment Agency, δείχνουν ότι οι εκπομπές διοξειδίου του θείου στην ΕΕ μειώθηκαν κατά περισσότερο από 90% μεταξύ των ετών 1990 και 2020 [9]. Αντίστοιχα, στις ΗΠΑ, οι εκπομπές μειώθηκαν κατά 94% από το 1980 έως το 2020, κυρίως λόγω των προγραμμάτων ελέγχου ρύπανσης και της μείωσης χρήσης άνθρακα [32].

Ωστόσο, σε ταχέως εκβιομηχανιζόμενες χώρες, όπως η Κίνα και η Ινδία, η εκτεταμένη καύση άνθρακα οδήγησε σε εκρηκτική αύξηση των εκπομπών κατά το τέλος του 20ού αιώνα [2]. Μόνο μετά το 2006, με την υιοθέτηση αυστηρότερων περιβαλλοντικών κανονισμών και την εγκατάσταση τεχνολογιών FGD, η Κίνα τουλάχιστον πέτυχε ουσιαστική μείωση των εκπομπών διοξειδίου του θείου [33].

Σε διεθνές επίπεδο, η σύμβαση του ΟΗΕ για τη ‘Διασυνοριακή Ατμοσφαιρική Ρύπανση σε Μεγάλες Αποστάσεις’ [34] αποτέλεσε ορόσημο, εισάγοντας συμφωνίες για τον περιορισμό του διοξειδίου του θείου σε διακρατική βάση, ιδιαίτερα στην Ευρώπη και τη Βόρεια Αμερική, όπου το πρόβλημα ήταν εντονότερο.

Συνολικά, οι ιστορικές εξελίξεις δείχνουν ότι η ρύπανση που προκαλείται από το διοξείδιο του θείου, θεωρείται ελέγξιμη, όταν εφαρμόζονται σωστά οι τεχνολογικές λύσεις και ακολουθούνται οι νομοθετικοί κανόνες. Παρ' όλα αυτά, η πρόκληση παραμένει, ιδιαίτερα σε περιοχές με αυξανόμενη ενεργειακή ζήτηση και χαλαρότερο ελεγκτικό πλαίσιο.

Κεφάλαιο 2: Το διοξείδιο του θείου ως αέριο του θερμοκηπίου και οι συνέπειές του

2.1 Το φαινόμενο του θερμοκηπίου

Το φαινόμενο του θερμοκηπίου είναι η φυσική διεργασία, μέσω της οποίας, η ατμόσφαιρα της γης απορροφά και επανεκπέμπει μέρος της υπέρυθρης ακτινοβολίας που εκλύεται από την επιφάνειά της. Με τον τρόπο αυτό διατηρούνται οι κατάλληλες θερμοκρασίες για την ανάπτυξη ζωής. Χωρίς αυτό, η μέση θερμοκρασία στην επιφάνεια της γης θα ήταν περίπου $-18\text{ }^{\circ}\text{C}$ αντί για $+15\text{ }^{\circ}\text{C}$ που είναι σήμερα [35], [36].

Τα βασικά αέρια του θερμοκηπίου (Greenhouse Gases - GHGs) περιλαμβάνουν το διοξείδιο του άνθρακα (CO_2), το μεθάνιο (CH_4), το υποξείδιο του αζώτου (N_2O) και τους υδρατμούς (H_2O), οι οποίοι αποτελούν και το πιο σημαντικό αέριο του θερμοκηπίου [10].

Αυτά τα αέρια απορροφούν και επανεκπέμπουν υπέρυθρη ακτινοβολία, αυξάνοντας έτσι το καθαρό ενεργειακό φορτίο στην ατμόσφαιρα της γης. Η επιρροή τους μετρείται μέσω της έννοιας Radiative Forcing (RF), ή σε ελεύθερη μετάφραση στα Ελληνικά 'Ακτινοβολιακή εξανάγκαση', δηλαδή η μεταβολή στο ενεργειακό ισοζύγιο του πλανήτη που προκαλεί ένας παράγοντας (θετική RF \rightarrow θερμαντική επίδραση, αρνητική RF \rightarrow ψυκτική επίδραση) [37]. Επιπλέον, η ικανότητα κάθε αερίου να συμβάλλει στη θέρμανση, σε σχέση με το CO_2 , αποτυπώνεται με το Δυναμικό Παγκόσμιας Θέρμανσης (Global Warming Potential - GWP), που εξαρτάται από τη διάρκεια ζωής του αερίου στην ατμόσφαιρα και την ικανότητά του να απορροφά υπέρυθρη ακτινοβολία [10]. Για παράδειγμα, το μεθάνιο έχει GWP περίπου 28 - 34 φορές υψηλότερο από το διοξείδιο του άνθρακα σε ορίζοντα 100 ετών, ενώ το υποξείδιο του αζώτου είναι σχεδόν 300 φορές πιο ισχυρό. Συνεπώς, το φαινόμενο του θερμοκηπίου είναι ένα πολύπλοκο αλλά βασικό στοιχείο διατήρησης της κλιματικής ισορροπίας. Οι αρνητικές επιπτώσεις, όπως η υπερθέρμανση του πλανήτη, προέρχονται όταν αυτή η ισορροπία διαταράσσεται από ανθρωπογενείς παράγοντες.

2.2 Το Διοξείδιο του Θείου και τα Αέρια του Θερμοκηπίου

Όπως παρουσιάστηκε στο Κεφάλαιο 1, το διοξείδιο του θείου, μέσω των ατμοσφαιρικών μετασχηματισμών του, οδηγεί σε σημαντικές περιβαλλοντικές και υγειονομικές συνέπειες, όπως η όξινη βροχή και η δημιουργία δευτερογενών θειικών σωματιδίων, που αποτελούν κύρια συνιστώσα των λεπτών αιωρούμενων σωματιδίων $\text{PM}_{2.5}$ [1]. Σε αντίθεση με τα κλασικά αέρια του θερμοκηπίου (GHGs), που απορροφούν άμεσα την υπέρυθρη ακτινοβολία και συμβάλλουν στη θέρμανση του πλανήτη, το διοξείδιο του θείου δεν δρα ως τέτοιο. Ο ρόλος του είναι έμμεσος και συνδέεται κυρίως με το σχηματισμό θειικών αερολυμάτων [10], [38], [39] που διασκορπίζουν την εισερχόμενη ηλιακή ακτινοβολία, μειώνοντας έτσι την ορατότητα και δημιουργώντας μία χαρακτηριστική 'ομίχλη' [38].

Με τον όρο 'χαρακτηριστική ομίχλη' δεν αναφερόμαστε στη φυσική ομίχλη που αποτελείται από υδρατμούς, αλλά σε ένα φωτοχημικό φαινόμενο που δημιουργείται από τα θειικά αερολύματα στην ατμόσφαιρα. Τα θειικά αερολύματα είναι συστήματα πολύ μικρών σωματιδίων που παραμένουν αιωρούμενα στον αέρα, είναι εξαιρετικά λεπτά κι έχουν μεγάλη ικανότητα σκέδασης (scattering) της ηλιακής ακτινοβολίας. Η σκέδαση, συνιστά το φαινόμενο κατά το οποίο η πορεία ενός κύματος, όπως είναι το φως, μεταβάλλεται όταν συναντά σωματίδια ή άλλες ανωμαλίες στο διάβα του. Ουσιαστικά, αντί το φως να συνεχίσει ευθύγραμμο, διασκορπίζεται σε πολλές κατευθύνσεις.

Από τα ανωτέρω, γίνεται αντιληπτό ότι το διοξείδιο του θείου, από τη μία πλευρά συνεισφέρει σε αρνητική RF, δρώντας ως ψυκτικός παράγοντας στο κλίμα, ενώ από την άλλη, έμμεσα, συνιστά σοβαρή απειλή για τη δημόσια υγεία. Η διπλή αυτή δράση του είναι που το καθιστά μοναδικό μεταξύ των λοιπών ατμοσφαιρικών ρύπων, αφού η μείωση των εκπομπών του, αν και είναι απαραίτητη για την προστασία της ανθρώπινης υγείας και του

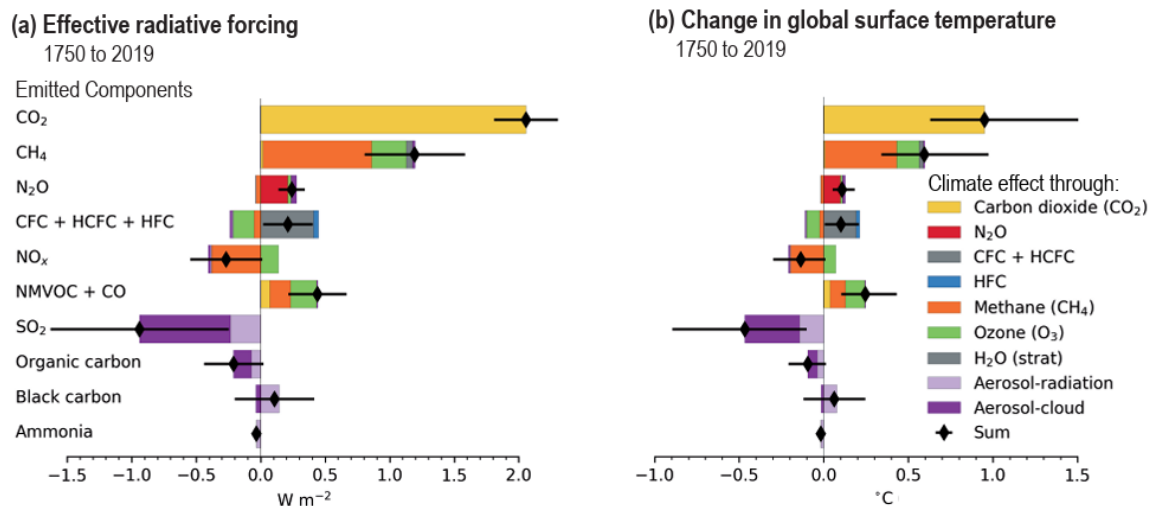
περιβάλλοντος, συνεπάγεται ταυτόχρονα και τον περιορισμό της ψυκτικής ιδιότητας που προσφέρει, μέσω των θεικών αερολυμάτων.

2.3 Κλιματική Επίδραση - Επιπτώσεις

Τα θεικά αερολύματα έχουν δύο βασικές κλιματικές επιδράσεις:

Άμεσο αποτέλεσμα, όπου σκεδάζουν την ηλιακή ακτινοβολία μικρού μήκους κύματος, μειώνοντας έτσι την ποσότητα ενέργειας που καταφθάνει στην επιφάνεια της Γης. Το φαινόμενο αυτό οδηγεί σε μείωση της θερμοκρασίας, γνωστό και ως παγκόσμια συσκότιση (global dimming) [38], [39].

Έμμεσο αποτέλεσμα (cloud albedo effect), λειτουργώντας ως πυρήνες συμπύκνωσης νεφών (CCN - cloud condensation nuclei), όπου σχηματίζουν νέφη με περισσότερα και μικρότερα σταγονίδια, τα οποία είναι πιο ανακλαστικά (μεγαλύτερο albedo) και διαρκούν περισσότερο. Αυτό ενισχύει περαιτέρω το ψυκτικό αποτέλεσμα [40], [41].



Γράφημα 2: Εκτιμώμενη συνεισφορά στην (α) μεταβολή της επιβαλλόμενης ακτινοβολίας (Effective Radiative Forcing - ERF) και (β) μεταβολή της παγκόσμιας θερμοκρασίας επιφάνειας από εκπομπές συστατικών για την περίοδο 1750–2019 (πηγή: IPCC Sixth Assessment Report (2021) - <https://www.ipcc.ch/report/ar6/wg1/>)

Σε παγκόσμια κλίμακα, τα θεικά αερολύματα συμβάλλουν σε αρνητική RF, καθώς διασκορπίζουν και ανακλούν την εισερχόμενη ηλιακή ακτινοβολία πίσω στο διάστημα, μειώνοντας έτσι την ενεργειακή ροή που φτάνει στην επιφάνεια της Γης. Το αποτέλεσμα είναι ένα ψυκτικό φαινόμενο (cooling effect), το οποίο αντισταθμίζει εν μέρει τη θετική RF των αερίων του θερμοκηπίου. Ωστόσο, υπενθυμίζουμε ότι η ευνοϊκή αυτή επίδραση, συνοδεύεται από σοβαρές αρνητικές συνέπειες στη δημόσια υγεία, αφού τα ίδια αερολύματα αποτελούν την κύρια συνιστώσα των λεπτών αιωρούμενων σωματιδίων PM_{2.5} [38]. Σύμφωνα με την Έκθεση IPCC AR6 του 2021 [10], η καθαρή αποτελεσματική RF των αερολυμάτων εκτιμάται σε $-1.1 W/m^2$ (εύρος -1.7 έως $-0.4 W/m^2$), καθιστώντας τα αερολύματα αυτά ως τον ισχυρότερο γνωστό παράγοντα αρνητικής RF.

Αξίζει να σημειωθεί ότι τη δεκαετία του 1970, ορισμένοι ερευνητές είχαν την άποψη ότι τα αερολύματα μπορούν να 'σκεπάσουν' τη θέρμανσή, από τα αέρια του θερμοκηπίου, οδηγώντας ακόμη και σε πιθανή παγκόσμια ψύξη [42]. Η άποψη αυτή στηρίχθηκε σε μετρήσεις που έδειχναν μείωση της θερμοκρασίας στο βόρειο ημισφαίριο, κατά τα μέσα του 20ού αιώνα, περίοδος που συνέπεσε με την έντονη βιομηχανική ρύπανση και τις υψηλές εκπομπές διοξειδίου του θείου. Σήμερα, αν και επιβεβαιώνεται η άποψη ότι μειώνουν τη θέρμανση, δεν είναι σε θέση να την αντισταθμίσουν πλήρως [10], [38]. Δηλαδή, θα μπορούσαμε να πούμε πως λειτουργούν σα φρένο στο φαινόμενο του θερμοκηπίου.

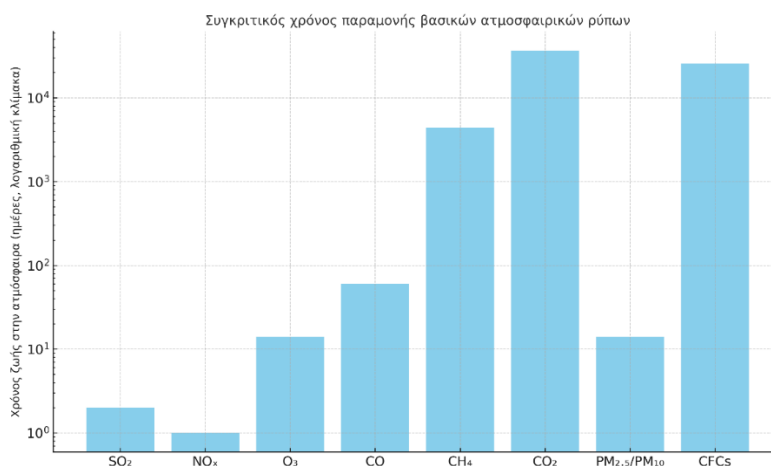
2.4 Λοιπές επισημάνσεις

2.4.1 Διάρκεια ζωής στην ατμόσφαιρα

Μία επίσης βασική διαφορά ανάμεσα στα αέρια του θερμοκηπίου (GHGs) και τα θεϊκά αερολύματα είναι η διάρκεια παραμονής τους στην ατμόσφαιρα. Τα GHGs έχουν μακρά διάρκεια ζωής, κάτι που σημαίνει πως η επίδρασή τους είναι σωρευτική και ανεξάρτητη από τις άμεσες εκπομπές τους.

Συγκεκριμένα, α) το CO₂, που προέρχεται κυρίως από την καύση ορυκτών καυσίμων (άνθρακας, πετρέλαιο και φυσικό αέριο), τη βιομηχανία και τις μεταφορές, έχει διάρκεια ζωής έως και εκατοντάδες χρόνια, β) το CH₄, που εκπέμπεται από γεωργικές δραστηριότητες, χωματερές, καθώς και από την εξόρυξη και μεταφορά πετρελαίου και φυσικού αερίου, έχει διάρκεια ζωής περίπου 12 χρόνια και γ) το N₂O, που εκπέμπεται κυρίως από λιπάσματα, καύση καυσίμων και βιομηχανικές δραστηριότητες, έχει διάρκεια ζωής περίπου 120 χρόνια.

Αντίθετα, τα θεϊκά αερολύματα έχουν διάρκεια ζωής μόλις λίγες ημέρες έως εβδομάδες, διότι απομακρύνονται από την ατμόσφαιρα μέσω της βροχόπτωσης (wet deposition) ή της καθίζησης (dry deposition). Δηλαδή, η ψυκτική τους ιδιότητα, που αναφέραμε παραπάνω, είναι βραχύβια και σε τοπικό επίπεδο. Γι' αυτό και χαρακτηρίζεται ως μία 'μάσκα ψύξης (cooling mask)', η οποία μειώνει προσωρινά τη θέρμανση αλλά δεν μπορεί να συγκριθεί με τα αποτελέσματα της μακροχρόνιας θετικής RF των GHGs.



Γράφημα 3: Συγκριτικός χρόνος παραμονής βασικών ατμοσφαιρικών ρύπων στην ατμόσφαιρα (σε λογαριθμική κλίμακα ημερών) (προσαρμοσμένο από Seinfeld, & Pandis, 2016)

2.4.2 Σύνδεση με πολιτικές καθαρού αέρα

Οι πολιτικές περιορισμού των εκπομπών διοξειδίου του θείου, που εφαρμόστηκαν τόσο στην Ευρώπη όσο και στη Βόρεια Αμερική (πχ. Οδηγία 2001/80/EC· Clean Air Act στις ΗΠΑ, Οδηγία 2008/50/EK του Ευρωπαϊκού Κοινοβουλίου), είχαν άμεσα οφέλη, όπως τη βελτίωση της ποιότητας του αέρα, τη μείωση της όξινης βροχής αλλά και τη δραστική μείωση της νοσηρότητας και της θνησιμότητας από αναπνευστικά και καρδιαγγειακά νοσήματα [4].

Ωστόσο, αυτές οι πολιτικές οδήγησαν σε ένα φαινόμενο γνωστό ως 'climate penalty of clean air policies' [43], [44]. Πιο συγκεκριμένα, εξαιτίας της μείωσης των εκπομπών διοξειδίου του θείου η ψυκτική επίδραση που αναφέραμε παραπάνω εξασθένησε, με αποτέλεσμα να είναι πιο έντονο το πραγματικό θερμαντικό αποτέλεσμα των GHGs. Η επίδραση αυτή καταγράφεται και στα δεδομένα, όπου παρότι οι εκπομπές SO₂ στην Ευρώπη και στις ΗΠΑ μειώθηκαν περισσότερο από 90% από το 1990 [10], [45], η παγκόσμια θέρμανση όχι μόνο δεν επιβραδύνθηκε, αλλά επιταχύνθηκε.

2.4.3 Αλληλεπίδραση με νέφη

Τα θεϊκά αερολύματα δεν επηρεάζουν μόνο το άμεσο ισοζύγιο ακτινοβολίας, αλλά και τις ιδιότητες των νεφών. Αναλυτικότερα,

α) όταν υπάρχουν περισσότερα αερολύματα, τα νέφη σχηματίζονται με αντίστοιχα περισσότερα αλλά μικρότερα σταγονίδια. Αυτά τα νέφη είναι πιο φωτεινά και ανακλούν μεγαλύτερο ποσοστό της ηλιακής ακτινοβολίας πίσω στο διάστημα (Cloud albedo effect) [41] και

β) τα νέφη με μικρότερα σταγονίδια εξατμίζονται πιο δύσκολα και τείνουν να διαρκούν περισσότερο. Αυτό σημαίνει ότι η παρουσία θεικών αερολυμάτων οδηγεί σε αύξηση της διάρκειας ζωής των νεφών, ενισχύοντας ακόμη περισσότερο το ψυκτικό αποτέλεσμα (Cloud lifetime effect) [40].

Αυτές οι αλληλεπιδράσεις, γνωστές και ως έμμεσες επιδράσεις των αερολυμάτων, είναι πιο δύσκολο να ποσοτικοποιηθούν από την άμεση σκέδαση. Παρά ταύτα, θεωρούνται κρίσιμες για την κατανόηση του συνολικού ρόλου των αερολυμάτων στο κλιματικό σύστημα [10].

2.4.4 Γεωγραφική διαφοροποίηση

Ένα ακόμη σημαντικό σημείο είναι ότι το ψυκτικό αποτέλεσμα των θεικών αερολυμάτων δεν είναι ομοιόμορφο παγκοσμίως. Και τούτο διότι, όπως προαναφέραμε, τα αερολύματα έχουν βραχύ χρόνο ζωής και παραμένουν κοντά στις πηγές εκπομπής τους. Έτσι, οι υψηλότερες συγκεντρώσεις και το ισχυρότερο ψυκτικό αποτέλεσμα παρατηρείται στο Βόρειο Ημισφαίριο, κοντά σε βιομηχανικές περιοχές της Ευρώπης, της Βόρειας Αμερικής και της Ανατολικής Ασίας. Αυτό δημιουργεί ανισορροπίες στο ενεργειακό ισοζύγιο μεταξύ βόρειου και νότιου ημισφαιρίου, επηρεάζοντας ακόμη και τα μοτίβα βροχοπτώσεων αλλά και την ένταση των μουσώνων [39].

Η γεωγραφική αυτή διαφοροποίηση, σε συνδυασμό με τον σχετικά περιορισμένο χρόνο ζωής τους στην ατμόσφαιρα είναι σημαντική και εξηγούν το γιατί το ψυκτικό αποτέλεσμα των αερολυμάτων δεν μπορεί να θεωρηθεί ότι είναι σε θέση να αντισταθμίσει σε παγκόσμια κλίμακα το φαινόμενο του θερμοκηπίου, αλλά περισσότερο αποτελεί περιφερειακό τροχοπέδη στο φαινόμενο αυτό.

Κεφάλαιο 3: Μέθοδοι ανάλυσης και πρόβλεψης χρονοσειρών

Στα προηγούμενα κεφάλαια παρουσιάσαμε τη φυσικοχημική ταυτότητα του διοξειδίου του θείου, τις πηγές εκπομπής του, τις χρήσεις του, τον ρόλο του στις κλιματικές διεργασίες, καθώς και τις επιπτώσεις του στο περιβάλλον και τη δημόσια υγεία. Στο παρόν κεφάλαιο θα παραθέσουμε κάποιες βασικές έννοιες και το αντίστοιχο θεωρητικό υπόβαθρο, ώστε να μην αναλωθούμε σε αυτές κατά την παράθεση των αποτελεσμάτων της μελέτης μας για την εν γένει πορεία και συμπεριφορά του.

3.1 Ορισμός και έννοια της χρονοσειράς

Μια χρονοσειρά ορίζεται ως μια ακολουθία από παρατηρήσεις μιας μεταβλητής, οι οποίες καταγράφονται σε διαδοχικές χρονικές στιγμές και συνήθως σε σταθερά χρονικά διαστήματα. Οι παρατηρήσεις αυτές μπορεί να είναι ημερήσιες, μηνιαίες, ετήσιες κλπ., ανάλογα με τη φύση των δεδομένων και τον σκοπό της ανάλυσης.

Η βασική ιδέα των χρονοσειρών είναι η χρονική εξάρτηση των παρατηρήσεων, δηλαδή, οι τιμές που παρατηρούνται στο παρόν συνδέονται με τις τιμές του παρελθόντος και συχνά ακολουθούν συγκεκριμένα μοτίβα ή τάσεις [46]. Η εξάρτηση αυτή εκφράζεται μαθηματικά μέσω της αυτοσυσχέτισης, κάτι που σημαίνει πως η τιμή μιας μεταβλητής τη χρονική στιγμή t (Y_t) σχετίζεται στατιστικά με τις τιμές της ίδιας μεταβλητής σε προηγούμενες χρονικές στιγμές (Y_{t-1}, Y_{t-2}, \dots) [47].

Οι χρονοσειρές διακρίνονται σε δύο βασικές κατηγορίες, τις τακτικές (regular) και τις μη τακτικές (irregular), ανάλογα με το αν οι παρατηρήσεις μίας μεταβλητής καταγράφονται σε σταθερά χρονικά διαστήματα ή όχι [48]. Στις τακτικές χρονοσειρές, οι παρατηρήσεις πραγματοποιούνται με σταθερή συχνότητα, όπως για παράδειγμα κάθε ημέρα, κάθε μήνα κλπ. Αυτή η σταθερότητα του χρονικού βήματος επιτρέπει την εφαρμογή κλασικών στατιστικών, καθώς και σύγχρονων μηχανικών μοντέλων πρόβλεψης, όπως οι μέθοδοι αποσύνθεσης, οι εκθετικές εξομαλύνσεις, τα μοντέλα ARIMA/SARIMA και τα δίκτυα LSTM (Long Short Term Memory).

Ένα παράδειγμα τακτικής χρονοσειράς είναι οι συγκεντρώσεις του διοξειδίου του θείου που παρατηρούνται σε μια πόλη, σε συγκεκριμένα χρονικά διαστήματα, όπως κάθε ημέρα. Η ακολουθία αυτών των παρατηρήσεων μας επιτρέπει να τη μελετήσουμε με τη χρήση διάφορων παραμετρικών και μη παραμετρικών μεθόδων. Αντίθετα, στις μη τακτικές χρονοσειρές, οι παρατηρήσεις δεν καταγράφονται σε ίσα χρονικά διαστήματα, δηλαδή μπορεί να

υπάρχουν μεγάλα κενά μεταξύ των παρατηρήσεων, ή οι μετρήσεις να πραγματοποιούνται μόνο όταν συμβαίνει ένα συγκεκριμένο γεγονός. Ένα χαρακτηριστικό παράδειγμα μίας τέτοιας χρονοσειράς, είναι οι παρατηρήσεις των εκπομπών διοξειδίου του θείου από ηφαιστειακές εκρήξεις, καθώς αυτές δεν συμβαίνουν σε προκαθορισμένα χρονικά διαστήματα αλλά σε άτακτες χρονικές στιγμές.

Γενικά, οι τακτικές χρονοσειρές είναι πιο εύκολα επεξεργάσιμες, αφού η σταθερή δομή τους επιτρέπει την απευθείας εφαρμογή μεθόδων ανάλυσης και πρόβλεψης. Αντιθέτως, οι μη τακτικές σειρές απαιτούν πιο εξειδικευμένες τεχνικές, όπως διαδικασίες που βασίζονται σε γεγονότα (event history analysis) ή μεθόδους ανομοιόμορφης δειγματοληψίας, ώστε να ληφθεί υπόψη η ακανόνιστη φύση των δεδομένων [49].

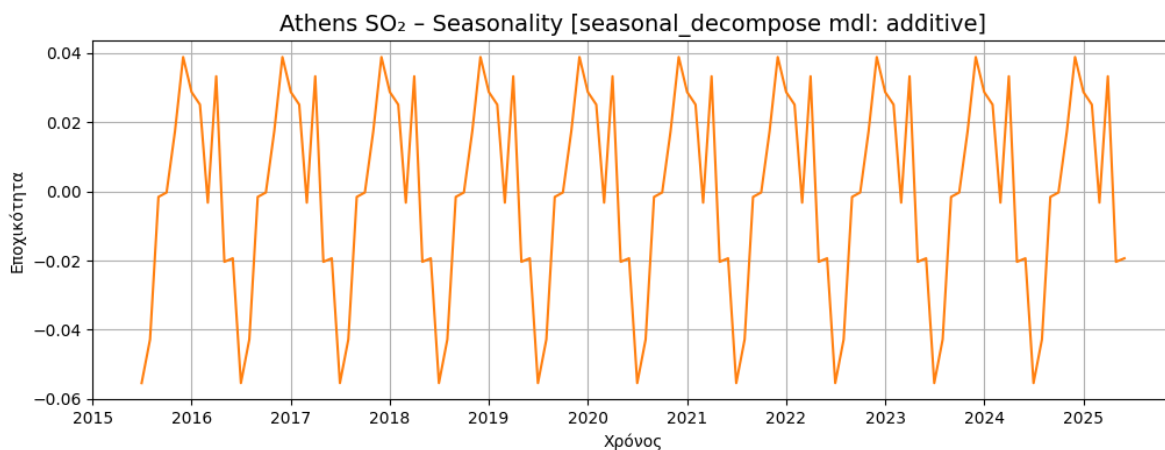
3.1.1 Βασικά συστατικά μιας χρονοσειράς

Μια χρονοσειρά Y_t μπορεί να αναλυθεί σε τρεις κύριες συνιστώσες:

Εποχικότητα (Seasonality, S_t), που αντιπροσωπεύει τα περιοδικά μοτίβα σε επαναλαμβανόμενα χρονικά διαστήματα, όπως για παράδειγμα οι αυξημένες συγκεντρώσεις διοξειδίου του θείου το χειμώνα, λόγω χρήσης θερμαντικών μέσων ή οι μειωμένες τιμές το καλοκαίρι.

Μαθηματικά, η εποχικότητα έχει περιοδικό χαρακτήρα και εκφράζεται από τη σχέση $S_{t+m} = S_t$, όπου m είναι το μήκος του εποχικού κύκλου (πχ. $m = 12$ για μηνιαία δεδομένα με ετήσια εποχικότητα).

Σχηματικά, η εποχικότητα φαίνεται ως επαναλαμβανόμενες κυματώσεις στη διάσταση του χρόνου.



Γράφημα 4: Εποχική διακύμανση, μετά την αποσύνθεση της χρονοσειράς των συγκεντρώσεων διοξειδίου του θείου της πόλης Αθήνα.

Στο παραπάνω στιγμιότυπο αποτυπώνεται η εποχικότητα των συγκεντρώσεων διοξειδίου του θείου στην Αθήνα που προέκυψε από την εφαρμογή της κλασικής αποσύνθεσης (σ.σ. αναλύεται παρακάτω). Η περιοδικότητα εμφανίζεται ιδιαίτερα σταθερή και επαναλαμβανόμενη σε ετήσια βάση, με κορυφώσεις κατά τους χειμερινούς μήνες και χαμηλότερες τιμές το καλοκαίρι, κάτι αναμενόμενο άλλωστε καθώς οι χειμερινοί μήνες χαρακτηρίζονται από αυξημένη χρήση θερμαντικών μέσων και λιγότερο ευνοϊκές μετεωρολογικές συνθήκες διασποράς των ρύπων. Αντίθετα, τους καλοκαιρινούς μήνες, οι υψηλές θερμοκρασίες, η αυξημένη ηλιακή ακτινοβολία και η μεγαλύτερη ανάμιξη της ατμόσφαιρας, συμβάλλουν στη μείωση των συγκεντρώσεων.

Σημείωση: Το καλοκαίρι, ο αέρας κοντά στο έδαφος ζεσταίνεται από την ηλιακή ακτινοβολία, γίνεται πιο ελαφρύς και ανέρχεται στην ατμόσφαιρα. Αυτό, δημιουργεί κάθετη ανάμιξη, όπου ο αέρας κυκλοφορεί και οι ρύποι διασκορπίζονται σε μεγαλύτερο χώρο (όγκο). Αντίθετα, το χειμώνα, ο αέρας κοντά στο έδαφος είναι πιο ψυχρός (πιο βαρύς) και παγιδεύεται από ένα στρώμα θερμότερου αέρα που βρίσκεται σε υψηλότερα επίπεδα στην ατμόσφαιρα. Αυτό ονομάζεται θερμοκρασιακή αναστροφή, εμποδίζοντας έτσι την κάθετη ανάμιξη και έχει ως αποτέλεσμα ο ψυχρός αέρας να μην ανεβαίνει στην ατμόσφαιρα και οι ρύποι να εγκλωβίζονται κοντά στο έδαφος. Έτσι, το χειμώνα παρατηρούμε υψηλότερες συγκεντρώσεις ρύπων στην επιφάνεια της γης.

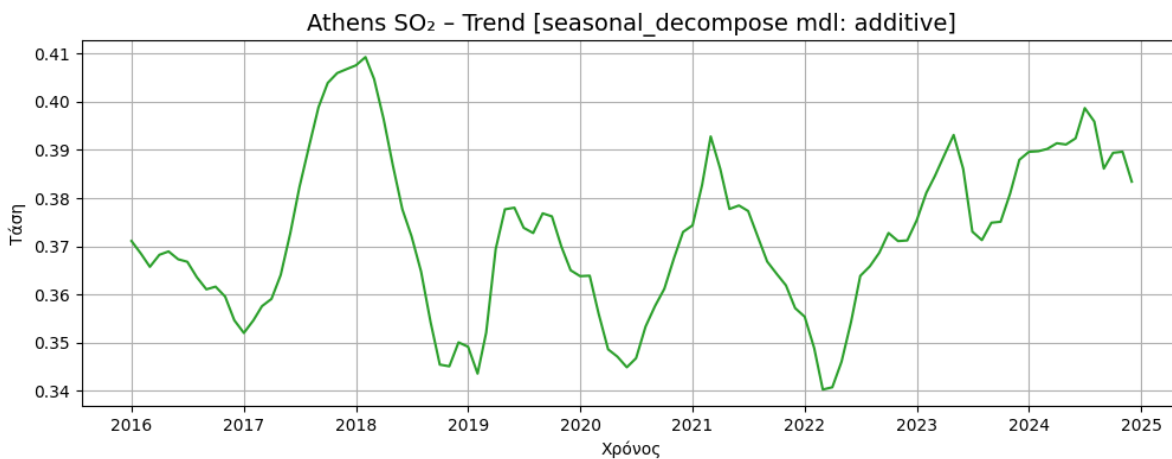
Τάση (Trend, T_t), που εκφράζει τη μακροχρόνια κατεύθυνση που ακολουθεί μια μεταβλητή, ανεξάρτητα από τις βραχυπρόθεσμες διακυμάνσεις και την εποχικότητα. Οι κ.κ. Rob J. Hyndman και George Athanasopoulos, στο

έργο τους 'Forecasting: Principles and Practice'[48], αναφέρουν ρητά ότι τάση υφίσταται όταν παρατηρείται μια μακροχρόνια αύξηση ή μείωση στα δεδομένα. Δεν είναι απαραίτητο η τάση να είναι γραμμική. Ορισμένες φορές αναφερόμαστε σε τάση που 'αλλάζει κατεύθυνση', όταν από αυξητική μετατρέπεται σε μειωτική ή το αντίστροφο.

Στην περίπτωση του διοξειδίου του θείου, η τάση αποτυπώνει τη μακροχρόνια πορεία των συγκεντρώσεων, η οποία μπορεί να εκφράζει είτε καθοδική πορεία (πχ. λόγω εφαρμογής περιβαλλοντικών πολιτικών) είτε ανοδική, εξαιτίας εντατικοποίησης ανθρωπογενών δραστηριοτήτων, όπως η καύση ορυκτών καυσίμων και οι βιομηχανικές διεργασίες ή γενικά να προέρχεται από μεταβολές των μετεωρολογικών συνθηκών κλπ.

Μαθηματικά, η τάση μπορεί να εκφραστεί με γραμμικά ή μη γραμμικά μοντέλα. Στην απλούστερη μορφή της, η γραμμική τάση εκφράζεται ως $T_t = \alpha + \beta t$ όπου α είναι η σταθερά εκκίνησης και β ο ρυθμός μεταβολής (slope).

Σχηματικά, η τάση απεικονίζεται ως μία ιδανικά ομαλή καμπύλη που διατρέχει τη χρονοσειρά.



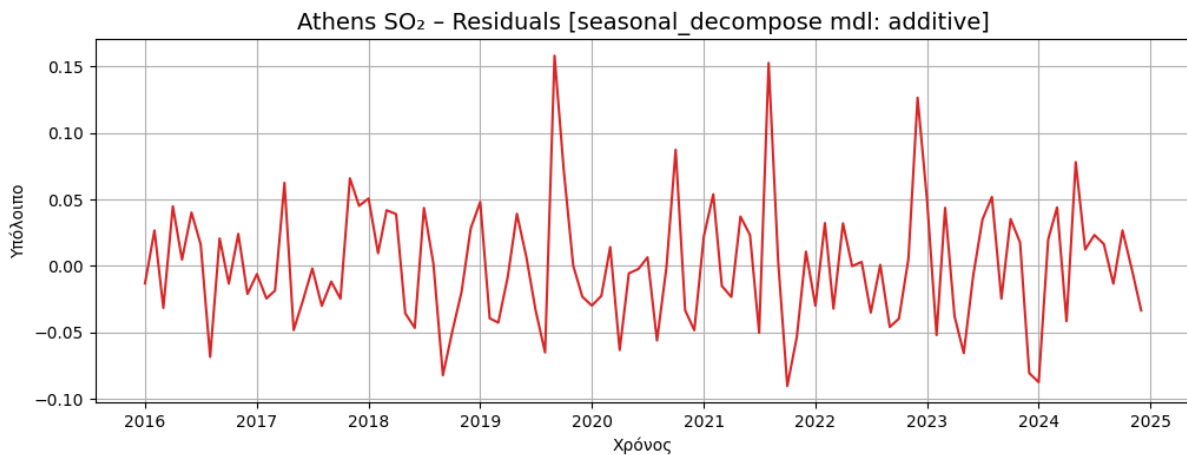
Γράφημα 5: Μακροχρόνια τάση, μετά την αποσύνθεση της χρονοσειράς των συγκεντρώσεων διοξειδίου του θείου της πόλης Αθήνα

Στο παραπάνω στιγμιότυπο παρουσιάζεται η εκτίμηση της μακροχρόνιας τάσης των συγκεντρώσεων διοξειδίου του θείου στην Αθήνα, όπως προέκυψε μετά την εφαρμογή της κλασικής μεθόδου αποσύνθεσης, όπου εμφανίζονται διακυμάνσεις με χαρακτηριστικές αιχμές και απότομες ή σταδιακές μεταβολές.

Κατάλοιπο ή Στοχαστικό στοιχείο (Residual/Irregular component, R_t), που περιλαμβάνει τυχαίες διακυμάνσεις που δεν εξηγούνται από την τάση ή την εποχικότητα. Στην περίπτωση του διοξειδίου του θείου, τέτοιες αποκλίσεις μπορεί να προκύπτουν από ηφαιστειακές εκρήξεις, ακραία καιρικά φαινόμενα ή έκτακτες βιομηχανικές δραστηριότητες.

Μαθηματικά, τα κατάλοιπα υπολογίζονται είτε με το προσθετικό μοντέλο (Additive Model) $R_t = Y_t - T_t - S_t$, όπου η εποχικότητα και οι τυχαίες διακυμάνσεις εκφράζονται σε σταθερές απόλυτες τιμές (πχ. Dobson Units) είτε με το πολλαπλασιαστικό μοντέλο (Multiplicative Model) $R_t = \frac{Y_t}{T_t \times S_t}$, όπου εκφράζονται σε ποσοστό, ανάλογα με τη τιμές της χρονοσειράς.

Σχηματικά, τα κατάλοιπα απεικονίζονται ως ακανόνιστες διακυμάνσεις γύρω από το μηδέν για το προσθετικό μοντέλο και γύρω από το 1 για το πολλαπλασιαστικό. Αν παρατηρούνται επαναλαμβανόμενα ή συστηματικά μοτίβα (πχ. κάθε Αύγουστο αιχμές, ή συστηματική μείωση της τιμής των καταλοίπων με το χρόνο), τότε η διαδικασία της αποσύνθεσης δεν έχει καταφέρει να απομονώσει πλήρως την εποχικότητα ή/και την τάση. Αντίθετα, αν τα κατάλοιπα είναι τυχαία διασκορπισμένα, χωρίς εμφανή μοτίβα και οι τιμές τους είναι πλησίον του μηδενός (ή 1 για το πολλαπλασιαστικό μοντέλο), τότε η διαδικασία της αποσύνθεσης κατάφερε να απομονώσει την εποχικότητα και την τάση. Η εμφάνιση σπάνιων έντονων αιχμών είναι φυσιολογικές και υποδηλώνουν έκτακτα γεγονότα τα οποία δεν αλλοιώνουν τη συνολική δομή της χρονοσειράς.



Γράφημα 6: Κατάλοιπα αποσύνθεσης της χρονοσειράς των συγκεντρώσεων διοξειδίου του θείου της πόλης Αθήνα.

Στο παραπάνω διάγραμμα παρουσιάζονται τα κατάλοιπα της αποσύνθεσης της χρονοσειράς των συγκεντρώσεων διοξειδίου του θείου στην Αθήνα, όπου παρατηρούνται αρκετές έντονες κορυφώσεις, οι οποίες πιθανόν να συνδέονται με βραχυπρόθεσμα επεισόδια αυξημένων συγκεντρώσεων, λόγω τοπικών πηγών ρύπανσης ή έκτακτων μετεωρολογικών συνθηκών. Από την απλή οπτική παρατήρηση, φαίνεται να υπάρχει μία ελαφριά αλλαγή συμπεριφοράς, πριν το 2019 (μικρότερη διακύμανση, χωρίς έντονες κορυφώσεις) και μετά το 2021 (τυχαία διασπορά τιμών με μεγαλύτερη διακύμανση και μερικές έντονες αιχμές). Σε τέτοιες περιπτώσεις απαιτείται περεταίρω ανάλυση, ώστε να διαπιστωθεί αν πράγματι υπάρχουν μοτίβα που δεν κατόρθωσε να εξηγήσει η αποσύνθεση.

3.2 Αποσύνθεση Χρονοσειρών (Time Series Decomposition)

Η διαδικασία με την οποία διασπάται μια χρονοσειρά στα δομικά συστατικά που παραθέσαμε προηγουμένως (εποχικότητα, τάση, κατάλοιπα) ονομάζεται αποσύνθεση (decomposition) [50].

Στη βιβλιογραφία συναντώνται διάφορες μέθοδοι αποσύνθεσης χρονοσειρών. Παρακάτω θα αναφερθούμε συνοπτικά στην κλασική αποσύνθεση (Classical Decomposition) και στην αποσύνθεση μέσω τοπικής παλινδρόμησης (Seasonal-Trend Decomposition using LOESS - STL), την οποία και χρησιμοποιήσαμε στη μελέτη μας.

Η κλασική αποσύνθεση αποτελεί μια από τις παλαιότερες και πιο διαδεδομένες στατιστικές μεθόδους ανάλυσης χρονοσειρών. Η βασική υπόθεση της μεθόδου είναι ότι η χρονοσειρά μπορεί να αναπαρασταθεί είτε ως άθροισμα είτε ως γινόμενο των συνιστωσών εποχικότητας, τάσης και καταλοίπων. Στην προσθετική (additive) μορφή της, η διακύμανση παραμένει σταθερή στο χρόνο και ο εποχικός κύκλος δεν μεταβάλλεται στο πέρασμα των ετών. Δηλαδή μπορεί να περιγραφεί ως,

$$Y_t = T_t + S_t + I_t$$

Στην πολλαπλασιαστική (multiplicative) εκδοχή της, η οποία είναι καταλληλότερη για δεδομένα με εξ αρχής γνωστή μεταβαλλόμενη διακύμανση, η εποχικότητα και τα κατάλοιπα εκφράζονται αναλογικά ως ποσοστιαίες μεταβολές της τάσης:

$$Y_t = T_t \times S_t \times I_t$$

Η εκτίμηση της τάσης (T_t) μπορεί να υπολογιστεί μέσω κινητών μέσων (moving averages), διαδικασία κατά την οποία επιλέγεται ένα συγκεκριμένο παράθυρο (window) παρατηρήσεων για την εξομάλυνση της χρονοσειράς. Το μέγεθος του παραθύρου καθορίζει το βαθμό εξομάλυνσης και επηρεάζει άμεσα την ακρίβεια της εκτιμώμενης τάσης. Ουσιαστικά, κάθε τιμή της χρονοσειράς αντικαθίσταται από το μέσο όρο ενός καθορισμένου πλήθους γειτονικών παρατηρήσεων, δημιουργώντας έτσι μια εξομαλυσμένη εκδοχή της χρονοσειράς, στην οποία έχουν ελαχιστοποιηθεί οι βραχυχρόνιες διακυμάνσεις (θόρυβος) και έτσι γίνεται πιο εμφανής η μακροχρόνια τάση.

Ακολουθώς, η εποχικότητα (S_t) υπολογίζεται ως η μέση απόκλιση από την τάση για κάθε χρονική περίοδο (πχ. για κάθε μήνα), ενώ οι λοιπές ανωμαλίες (Irregularities) που δεν εξηγήθηκαν από την εποχικότητα και την τάση (I_t) προκύπτουν από το λόγο μεταξύ των αρχικών παρατηρήσεων και των δύο συνιστωσών, εποχικότητας και τάσης.

Η μέθοδος αυτή είναι σχετικά απλή, ωστόσο προορίζεται κυρίως για χρονοσειρές με σταθερή εποχικότητα και χωρίς ακραίες τιμές. Για το λόγο αυτό, σε περιβαλλοντικά και μετεωρολογικά δεδομένα συστήνεται η χρήση άλλων μεθόδων που να μπορούν να προσαρμόζονται δυναμικά σε μη γραμμικές τάσεις και σε μεταβαλλόμενα εποχικά μοτίβα [48].

Μία από αυτές τις μεθόδους, είναι STL, η οποία προτάθηκε από τους Robert Cleveland, William Cleveland, Jean McRae και Irma Terpenning, το 1990 [51], ως μια πιο ευέλικτη και ανθεκτική προσέγγιση για την αποσύνθεση χρονοσειρών. Σε αντίθεση με την κλασική αποσύνθεση, η STL επιτρέπει την εκτίμηση μη γραμμικών τάσεων και μεταβαλλόμενης εποχικότητας (non-stationary seasonality), γεγονός που την καθιστά ιδανική για περιβαλλοντικά και ατμοσφαιρικά δεδομένα, όπου η ένταση ή το σχήμα του εποχικού μοτίβου μπορεί να μεταβάλλονται με το χρόνο.

Η βασική αρχή της μεθόδου είναι ότι η παρατηρούμενη χρονοσειρά Y_t μπορεί να αναλυθεί ως άθροισμα τριών επιμέρους συνιστωσών:

$$Y_t = T_t + S_t + R_t$$

όπου,

T_t , είναι η τάση,

S_t , η εποχική συνιστώσα και

R_t , το μη συστηματικό υπόλοιπο ή καλύτερα κατάλοιπο (residual) της χρονοσειράς.

Η καινοτομία της STL είναι στον τρόπο με τον οποίο εκτιμώνται οι ανωτέρω συνιστώσες. Αντί να χρησιμοποιεί απλούς κινητούς μέσους, όπως η κλασική αποσύνθεση, η STL εφαρμόζει τοπική παλινδρόμηση (Locally Estimated Scatterplot Smoothing - Loess, σ.σ. σε κάποιες βιβλιογραφικές αναφορές συναντάται και ως Locally Weighted Scatterplot Smoothing - Lowess). Πρόκειται για μια μη παραμετρική τεχνική εξομάλυνσης, η οποία εφαρμόζει σε κάθε σημείο της χρονοσειράς μια πολυωνυμική παλινδρόμηση, βασισμένη σε ένα τοπικό υποσύνολο παρατηρήσεων που καθορίζεται από ένα παράθυρο εξομάλυνσης (smoothing window). Η διαδικασία αυτή, εφαρμόζεται ξεχωριστά για την εκτίμηση της τάσης και ξεχωριστά για την εποχικότητα, χρησιμοποιώντας δύο διαφορετικά παράθυρα εξομάλυνσης. Το παράθυρο που χρησιμοποιείται για την τάση είναι ευρύτερο, ώστε να συλλαμβάνει τη μακροπρόθεσμη συμπεριφορά της χρονοσειράς, ενώ το παράθυρο που χρησιμοποιείται για την εποχικότητα είναι στενότερο, επιτρέποντας την ανίχνευση επαναλαμβανόμενων εποχικών μοτίβων.

Η ύπαρξη αυτών των δύο ανεξάρτητων παραθύρων εξασφαλίζει ότι η STL μπορεί να απομονώνει αποτελεσματικά τόσο τις βραχυπρόθεσμες περιοδικότητες όσο και τις μακροπρόθεσμες τάσεις της χρονοσειράς, ακόμη και όταν η εποχικότητα μεταβάλλεται στο χρόνο, όπως άλλωστε συμβαίνει σε περιβαλλοντικές χρονοσειρές. Η διαφορά μεταξύ των παρατηρήσεων και του αθροίσματος των συνιστωσών της τάσης και της εποχικότητας παράγουν τα κατάλοιπα, δηλαδή, τις βραχυπρόθεσμες ανωμαλίες που δεν κατάφεραν να εντοπίσουν η εποχικότητας και η τάση.

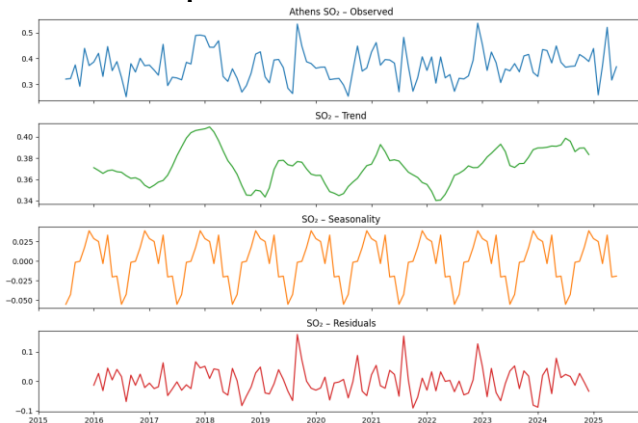
Η επιλογή του μεγέθους του παραθύρου εξομάλυνσης καθορίζει τον βαθμό ευαισθησίας της μεθόδου στις βραχυπρόθεσμες διακυμάνσεις. Μικρότερα παράθυρα καθιστούν τη μέθοδο πιο ευαίσθητη σε απότομες μεταβολές της χρονοσειράς, ενώ μεγαλύτερα παράθυρα εξομαλύνουν περισσότερο τις διακυμάνσεις, αναδεικνύοντας έτσι τη μακροπρόθεσμη τάση.

Επιπλέον, η STL διαθέτει τη 'robust' εκδοχή, όπου μετά την αρχική εκτίμηση των συνιστωσών (εποχικότητα και τάση), επαναλαμβάνεται η διαδικασία εξομάλυνσης, αλλά αυτή τη φορά δίδοντας μικρότερο βάρος στα σημεία που αποκλίνουν πολύ από την υπόλοιπη χρονοσειρά (δηλαδή, στα outliers). Αυτό το χαρακτηριστικό την καθιστά ιδιαίτερα κατάλληλη για περιβαλλοντικές χρονοσειρές, οι οποίες συχνά εμφανίζουν αιφνίδιες αυξήσεις ή μειώσεις στις τιμές των παρατηρήσεων.

Σημειώνουμε πως στην παρούσα μελέτη μας, επιλέξαμε να χρησιμοποιήσουμε τη ‘robust’ εκδοχή της μεθόδου STL, εφαρμόζοντας μάλιστα, προσαρμοστική επιλογή παραθύρων με ταυτόχρονους διαγνωστικούς ελέγχους ποιότητας, ώστε να εξασφαλιστεί κατά το δυνατό η στατιστική εγκυρότητα της αποσύνθεσης.

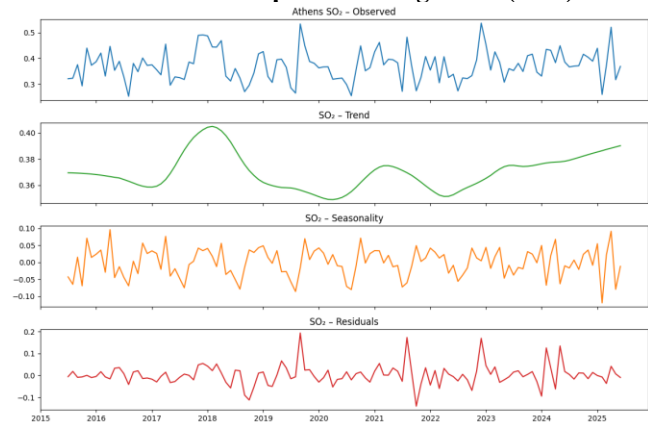
Όπως γίνεται αντιληπτό, η μέθοδος της κλασικής αποσύνθεσης διαφέρει σημαντικά από την STL. Για την καλύτερη κατανόηση της διαφοράς τους, στο παρακάτω σχήμα απεικονίζεται η αποσύνθεση της ίδιας χρονοσειράς, εφαρμόζοντας τις δύο ως άνω μεθόδους που περιγράψαμε.

Classical Decomposition



Γράφημα 7: Αποτέλεσμα εφαρμογής κλασικής αποσύνθεσης της χρονοσειράς συγκεντρώσεων διοξειδίου του θείου της πόλης της Αθήνας.

Seasonal-Trend Decomposition using Loess (STL)



Γράφημα 8: Αποτέλεσμα εφαρμογής της μεθόδου STL στη χρονοσειρά συγκεντρώσεων διοξειδίου του θείου της πόλης της Αθήνας.

Από τις ως άνω εικόνες, γίνεται εμφανής η δυναμική της STL έναντι της κλασικής μεθόδου αποσύνθεσης και απεικονίζεται τόσο στην τάση, η οποία είναι πιο ομαλή, όσο και στο εποχικό μοτίβο που είναι πιο ρεαλιστικό ενθυλακώνοντας εντός του τις μεταβολές των διακυμάνσεων στην εξέλιξη των ετών. Όσον αφορά τα κατάλοιπα, η διακύμανσή τους είναι σαφώς πιο περιορισμένη.

3.2.1 Μέθοδοι εξομάλυνσης

Στις παραπάνω ενότητες χρησιμοποιήσαμε εκτεταμένα την έννοια της εξομάλυνσης και δώσαμε βαρύτητα στη μέθοδο της εξομάλυνσης LOESS που χρησιμοποιεί η STL. Οπότε παρακάτω θα αναφερθούμε στις μεθόδους στις οποίες βασίζεται και η LOESS.

Η εξομάλυνση (smoothing) είναι η διαδικασία που χρησιμοποιούμε για να μετριάσουμε τις βραχυχρόνιες διακυμάνσεις (θόρυβο) μιας χρονοσειράς, ώστε να αναδειχθούν πιο καθαρά τα εποχικά μοτίβα και η μακροχρόνια τάση [46]. Όπως αναφέραμε και προηγουμένως, με την εξομάλυνση δημιουργείται μία νέα πιο ‘καθαρή’ χρονοσειρά, που μας επιτρέπει να εστιάσουμε στη βασική δομή των παρατηρήσεων, χωρίς αυτή να επηρεάζεται από έκτακτες τυχαίες διακυμάνσεις.

Γενικά, υπάρχουν διάφορες μέθοδοι εξομάλυνσης, οι κυριότερες από τις οποίες είναι,

α) Κινητοί μέσοι (Moving Averages), όπου για κάθε χρονικό σημείο υπολογίζουμε μία νέα τιμή βάσει προκαθορισμένου πλήθους των αμέσως προηγούμενων διαδοχικών παρατηρήσεων. Αυτό έχει ως αποτέλεσμα, τη δημιουργία μιας νέας εξομαλυμένης χρονοσειράς, όπου φαίνεται πιο καθαρά η τάση της. Ενδεικτικά, υπολογίζεται ως εξής:

$$MA_t = \frac{Y_{t-k} + Y_{t-k+1} + \dots + Y_{t+k}}{2k + 1}$$

όπου $2k + 1$ είναι το μήκος του παραθύρου (πόσες διαδοχικές παρατηρήσεις).

β) Απλή εκθετική εξομάλυνση (SES – Simple Exponential Smoothing), η οποία ευνοεί περισσότερο τις πιο πρόσφατες παρατηρήσεις και λιγότερο τις παρελθοντικές. Γενικά, χρησιμοποιείται σε χρονοσειρές χωρίς έντονη τάση ή εποχικότητα και προκύπτει από τη σχέση:

$$S_t = \alpha Y_t + (1 - \alpha)S_{t-1}$$

όπου, $0 < \alpha < 1$ είναι ο συντελεστής εξομάλυνσης και S_t είναι η εξομαλυμένη τιμή.

γ) Διπλή εκθετική εξομάλυνση (Holt's Linear Trend Method), η οποία επεκτείνει την απλή εκθετική εξομάλυνση, επιτρέποντας την εκτίμηση και εξομάλυνση της γραμμικής τάσης μιας μη εποχικής χρονοσειράς. Ουσιαστικά, χρησιμοποιεί δύο εξισώσεις, μία για το επίπεδο (σ.σ. τιμή γύρω από την οποία κινούνται οι παρατηρήσεις σε μία δεδομένη χρονική στιγμή) και μία για την τάση [52].

Για να γίνει πιο κατανοητή η λειτουργία της διπλής εκθετικής εξομάλυνσης, αν Y_t είναι η πραγματική τιμή στη χρονική στιγμή t , L_t η εκτίμηση του επιπέδου τη χρονική στιγμή t , T_t η εκτίμηση της τάσης τη χρονική στιγμή t και $\alpha, \beta \in (0,1)$ οι συντελεστές εξομάλυνσης για το επίπεδο και την τάση αντίστοιχα, τότε οι εξισώσεις της διπλής εκθετικής εξομάλυνσης είναι:

$$\text{εξίσωση επιπέδου: } L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}),$$

$$\text{εξίσωση τάσης: } T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

και

$$\text{εξίσωση πρόβλεψης: } \hat{Y}_{t+h} = L_t + hT_t,$$

όπου h είναι το βήμα πρόβλεψης (πχ. αν έχουμε μηνιαία δεδομένα, '1' είναι ο επόμενος μήνας και '12' ο επόμενος χρόνος).

Σχετικά με την αναφορά μας στην εξίσωση πρόβλεψης, πρόκειται για την αναμενόμενη τιμή της γραμμικής τάσης που υπολογίστηκε στο χρονικό σημείο t , μετά από h χρονικές περιόδους. Οπότε, γίνεται αντιληπτό, πως η διπλή εκθετική εξομάλυνση, πέραν του βασικού σκοπού της, μπορεί να χρησιμοποιηθεί και ως μέσο εκτίμησης της τάσης.

δ) Τριπλή εκθετική εξομάλυνση (Holt-Winters), η οποία επεκτείνει περαιτέρω τη διπλή εκθετική εξομάλυνση, λαμβάνοντας υπόψη, πέραν της τάσης και την εποχικότητα. Μαθηματικά, χρησιμοποιούνται δύο εκδοχές, ήτοι, η προσθετική, που επιλέγεται κυρίως για χρονοσειρές με σταθερή εποχικότητα και υπολογίζεται ως:

$$\text{εξίσωση επιπέδου: } L_t = \alpha(Y_t - S_{t-m}) + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$\text{εξίσωση τάσης: } T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1},$$

$$\text{εξίσωση εποχικότητας: } S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-m},$$

$$\text{εξίσωση πρόβλεψης: } \hat{Y}_{t+h} = L_t + hT_t + S_{t-m+h},$$

και η πολλαπλασιαστική, που επιλέγεται κυρίως σε χρονοσειρές όπου οι εποχικές διακυμάνσεις αυξάνονται ή μειώνονται αντίστοιχα με το επίπεδο της χρονοσειράς (μέση τιμή γύρω από την οποία κινούνται οι παρατηρήσεις σε μία δεδομένη χρονική στιγμή) και υπολογίζεται ως:

$$\text{εξίσωση επιπέδου: } L_t = \alpha \frac{Y_t}{S_{t-m}} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$\text{εξίσωση τάσης: } T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1},$$

$$\text{εξίσωση εποχικότητας: } S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-m},$$

$$\text{εξίσωση πρόβλεψης: } \hat{Y}_{t+h} = (L_t + hT_t) \times S_{t-m+h},$$

όπου, α είναι ο συντελεστής εξομάλυνσης του επιπέδου (δηλαδή πόσο βάρος δίδουμε σε νέες τιμές), β ο συντελεστής εξομάλυνσης της τάσης, γ ο συντελεστής εξομάλυνσης της εποχικότητας και m το μήκος του εποχικού κύκλου (πχ. αν έχουμε μηνιαία δεδομένα με ετήσια εποχικότητα, $m = 12$) [48], [52], [53].

Παρότι οι προαναφερθείσες μέθοδοι εξομάλυνσης, προσφέρουν έναν αποτελεσματικό τρόπο εξαγωγής της δομής μιας χρονοσειράς, έχουν κάποιους περιορισμούς. Οι κινητοί μέσοι χρησιμοποιούν σταθερά βάρη και επομένως αδυνατούν να προσαρμοστούν δυναμικά σε τοπικές μεταβολές ή σε μη γραμμικές τάσεις. Οι εκθετικές μέθοδοι

από την άλλη, αν και δίδουν μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις, υποθέτουν ότι η σχέση μεταξύ επιπέδου, τάσης και εποχικότητας είναι γραμμική και ότι οι παράμετροι εξομάλυνσης παραμένουν σταθεροί στο χρόνο.

Η μέθοδος LOESS, στην οποία βασίζεται η αποσύνθεση STL, ξεπερνά αυτούς τους περιορισμούς χρησιμοποιώντας τοπική πολυωνυμική παλινδρόμηση με σταθμισμένα βάρη. Ουσιαστικά και σε αντίθεση με τις παραπάνω μεθόδους, η LOESS δεν εφαρμόζει σταθερό τύπο εξομάλυνσης σε ολόκληρη τη χρονοσειρά, αλλά προσαρμόζει δυναμικά μια πολυωνυμική καμπύλη σε κάθε τμήμα των παρατηρήσεων, αποδίδοντας μεγαλύτερα βάρη στις κοντινές χρονικά παρατηρήσεις και μικρότερα στις πιο απομακρυσμένες. Με αυτόν τον τρόπο επιτυγχάνει ομαλή εξομάλυνση που προσαρμόζεται στις τοπικές μεταβολές της χρονοσειράς.

Κατόπιν αυτών, θα μπορούσαμε να πούμε πως η LOESS αποτελεί γενίκευση των κλασικών μεθόδων εξομάλυνσης, καθώς ενσωματώνει μεν την ιδέα του σταθμισμένου μέσου, όπως κάνουν οι κινητοί μέσοι και η εκθετική εξομάλυνση, αλλά με προσαρμοζόμενα βάρη και τοπική παλινδρόμηση, προσφέροντας έτσι μεγαλύτερη ευελιξία και ακρίβεια στην ανάλυση περιβαλλοντικών χρονοσειρών.

3.3 Μέθοδοι πρόβλεψης

Στην παράγραφο §3.2.1 περιγράψαμε τις μεθόδους εξομάλυνσης και συγκεκριμένα τον τρόπο με τον οποίο η εξομάλυνση μας επιτρέπει να κατανοήσουμε τις παρατηρήσεις μίας χρονοσειράς, αφαιρώντας από το θόρυβο και αναδεικνύοντας την τάση και την εποχικότητα. Το επόμενο στάδιο της μελέτης μας είναι η πρόβλεψη των μελλοντικών τιμών της χρονοσειράς, είτε μέσω στατιστικών μοντέλων που επεκτείνουν τη διαδικασία εξομάλυνσης είτε μέσω μεθόδων μηχανικής μάθησης.

Οι μέθοδοι πρόβλεψης που θα χρησιμοποιήσουμε στην παρούσα εργασία είναι τρεις. Η επέκταση της τριπλής εκθετικής εξομάλυνσης (Holt-Winters) που αναφέραμε προηγουμένως, καθώς και δύο πιο προχωρημένες μέθοδοι που χρησιμοποιούν και εξωγενείς παράγοντες, όπως μετεωρολογικές μετρήσεις, για την αποδοτικότερη παροχή προβλέψεων.

3.3.1 Holt-Winters – ως επέκταση της τριπλής εκθετικής εξομάλυνσης

Η μέθοδος Holt-Winters μπορεί να χρησιμοποιηθεί και ως ένα πλήρες προγνωστικό μοντέλο. Πιο συγκεκριμένα, αφού εκτιμηθούν η τάση και η εποχικότητα σε κάθε χρονική στιγμή της χρονοσειράς, μπορούμε να παρέχουμε προβλέψεις για μελλοντικές περιόδους, με τη βοήθεια της ακόλουθης σχέσης:

$$\hat{Y}_{t+h} = (L_t + h \cdot T_t) \cdot S_{t+h-m(k)}$$

όπου,

\hat{Y}_{t+h} είναι η τιμή πρόβλεψης της χρονοσειράς για h βήματα μετά τη χρονική στιγμή t ,

L_t είναι η τιμή γύρω από την οποία κινείται η χρονοσειρά τη δεδομένη χρονική στιγμή t , λαμβάνοντας υπόψη την τάση, αφ'ότου όμως έχουμε αφαιρέσει την εποχικότητα. Ουσιαστικά, υπολογίζεται δυναμικά από το συνδυασμό της τρέχουσας παρατήρησης, των προηγούμενων εκτιμήσεων, αλλά και των παραμέτρων εξομάλυνσης,

T_t είναι η τάση της χρονοσειράς τη χρονική στιγμή t , δηλαδή, απεικονίζει αν το επίπεδο της χρονοσειράς αυξάνεται, μειώνεται ή μένει σταθερό και σε περίπτωση που μεταβάλλεται το ρυθμό μεταβολής,

$S_{t+h-m(k)}$ είναι ο εποχικός συντελεστής που αντιστοιχεί στη χρονική θέση του βήματος πρόβλεψης μέσα στον εποχικό κύκλο και προκύπτει από την επανάληψη του εκτιμηθέντος εποχικού μοτίβου,

h είναι αριθμός βημάτων πρόβλεψης,

m είναι το μήκος της εποχικότητας (πχ. 12 για μηνιαία δεδομένα με ετήσιο κύκλο),

$k = \lfloor (h - 1)/m \rfloor$ είναι το πλήθος των πλήρων κύκλων εποχικότητας που έχουν μεσολαβήσει από τη χρονική στιγμή t μέχρι τη στιγμή $t + h$. Δηλαδή, πόσες φορές (κρατώντας το ακέραιο μέρος του υπολογισμού) 'χωράει' το μήκος της εποχικότητας m στο διάστημα των $h - 1$ βημάτων.

Έτσι, το μοντέλο Holt-Winters αξιοποιεί τις τιμές που υπολογίζει σε κάθε βήμα της εξομάλυνσης, ώστε να προεκτείνει την τάση και να επαναλάβει το εποχικό μοτίβο στο μέλλον.

Σημειώνουμε ότι η ανωτέρω μέθοδος, δεν έχει τη δυνατότητα να αξιοποιήσει εξωγενείς παράγοντες όπως για παράδειγμα μετεωρολογικά δεδομένα.

3.3.2 Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX)

Μία χρονοσειρά, μπορεί να περιγραφεί με τη χρήση των όρων, αυτοπαλινδρόμησης (AutoRegressive), διαφοροποίησης (Integrated) και κινητού μέσου όρου σφαλμάτων πρόβλεψης (Moving Average). Εξ' ου και η ονομασία του γενικού μοντέλου ARIMA [47], [54] που ορίζεται ως:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t$$

όπου,

B είναι ο τελεστής υστέρησης ($By_t = y_{t-1}$),

$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ είναι το πολυώνυμο των όρων αυτοπαλινδρόμησης,

$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ είναι το πολυώνυμο των όρων κινητού μέσου,

p είναι η τάξη της αυτοπαλινδρόμησης,

q είναι η τάξη του κινητού μέσου σφαλμάτων πρόβλεψης,

d είναι ο βαθμός διαφοροποίησης και

ε_t είναι το σφάλμα της πρόβλεψης (το μοντέλο το θεωρεί ως 'λευκό θόρυβο').

Το SARIMAX γενικεύει το ARIMA, ενσωματώνοντας σε αυτό την εποχικότητα και εξωγενείς παράγοντες [48], [55], που τους χρησιμοποιεί για να πετύχει καλύτερη πρόβλεψη, ενώ ορίζεται ως:

$$\varphi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta(B)\Theta(B^s)\varepsilon_t + \beta X_t$$

εισάγοντας επιπλέον τις παραμέτρους,

$\Phi(B^s)$ το πολυώνυμο της εποχικής αυτοπαλινδρόμησης,

$\Theta(B^s)$ το πολυώνυμο του εποχικού κινητού μέσου σφαλμάτων πρόβλεψης,

X_t είναι οι χρονοσειρές των εξωγενών παραγόντων,

β αποτελεί το διάνυσμα των συντελεστών παλινδρόμησης των εξωγενών παραγόντων,

P είναι η τάξη της εποχικής αυτοπαλινδρόμησης,

Q είναι η τάξη του εποχικού κινητού μέσου σφαλμάτων πρόβλεψης,

D είναι ο βαθμός της εποχικής διαφοροποίησης και

s είναι η εποχική περίοδος.

Ας δούμε αναλυτικά τι σημαίνουν και πως χρησιμοποιούνται οι ως άνω όροι:

α) Αυτοπαλινδρόμηση σημαίνει ότι η τρέχουσα τιμή της χρονοσειράς εξαρτάται γραμμικά από τις τιμές των προηγούμενων παρατηρήσεων. Η δυναμική αυτής της εξάρτησης εκφράζεται από τους συντελεστές αυτοπαλινδρόμησης (Auto Regressive coefficient) ϕ , που είναι διαφορετικοί για κάθε προηγούμενη υστέρηση (lag). Το πλήθος τους δε, που τελικά θα ληφθεί υπόψη από το μοντέλο, εκφράζεται από την τάξη της αυτοπαλινδρόμησης p και για το σκέλος της εποχικότητας P .

Ο τρόπος επιλογής του κατάλληλου p ορίζεται κατά την εφαρμογή του μοντέλου, όπου μπορούμε να χρησιμοποιήσουμε είτε διαγνωστικές μεθόδους όπως ACF (Autocorrelation Function) και PACF (Partial Autocorrelation Function) με συγκεκριμένα κριτήρια (πχ. BIC - Bayesian information Criterion, AIC - Akaike information Criterion) είτε με τη χρήση εξειδικευμένων αλγορίθμων πχ. auto arima (σε Python, μέσω της βιβλιοθήκης pmdarima) [47], [48]. Οι αλγόριθμοι αυτοί εξετάζουν διάφορους συνδυασμούς παραμέτρων σε ένα προκαθορισμένο εύρος τιμών και επιλέγουν το βέλτιστο, βάσει των ίδιων ως άνω κριτηρίων.

β) ο κινητός μέσος όρος σφαλμάτων περιγράφει την τρέχουσα παρατήρηση ως γραμμικό συνδυασμό του τυχαίου σφάλματος της τρέχουσας παρατήρησης και των σφαλμάτων προηγούμενων παρατηρήσεων. Στην περίπτωση ενός μοντέλου που χρησιμοποιεί το σφάλμα μόνο της αμέσως προηγούμενης παρατήρησης, η εξίσωση είναι $Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$, όπου, Y_t η τιμή τη χρονική στιγμή t , μ ο μέσος όρος, ε_t το τυχαίο σφάλμα της τρέχουσας παρατήρησης (σ.σ. θεωρείται ως λευκός θόρυβος από το μοντέλο), ε_{t-1} το σφάλμα της πρόβλεψης της αμέσως

προηγούμενης παρατήρησης και θ_1 ο συντελεστής που καθορίζει το πόσο επηρεάζει η προηγούμενη παρατήρηση την τρέχουσα πρόβλεψη [54]. Να σημειωθεί ότι κατά την πρόβλεψη μελλοντικών περιόδων, τα μελλοντικά σφάλματα δεν είναι γνωστά και λαμβάνουν την τιμή 0, ώστε να υπολογιστεί η πρόβλεψη. Η τάξη του κινητού μέσου σφαλμάτων πρόβλεψης q και Q για το σκέλος της εποχικότητας, εκφράζουν πόσα προηγούμενα σφάλματα λαμβάνονται υπόψη από το μοντέλο και ορίζονται κατά την εφαρμογή του, είτε μέσω εφαρμογής διαγνωστικών μεθόδων όπως ACF είτε μέσω αλγορίθμων σε Python όπως `auto arima`.

γ) ο τελεστής υστέρησης B (Backshift Operator) ή εναλλακτικά Lag, εκφράζει τη μετακίνηση μίας παρατήρησης προς τα πίσω ενώ ο εκθέτης του το συγκεκριμένο βήμα (πχ. B^2 είναι το δεύτερο βήμα και γενικά $B^2 y_t = y_{t-2}$). Η ανώτατη τιμή του εκθέτη ορίζεται από την τάξη της αυτοπαλινδρόμησης που αναφέραμε παραπάνω, ως B^p και γενικά ως $B^p y_t = y_{t-p}$.

δ) για να εφαρμοστεί το μοντέλο, είναι απαραίτητο η χρονοσειρά να είναι στάσιμη. Δηλαδή, ο μέσος όρος των παρατηρήσεων να παραμένει σταθερός στο χρόνο, η διακύμανση (variance) των τιμών των παρατηρήσεων (σ.σ. η διασπορά γύρω από το μέσο) να είναι επίσης σταθερή στο χρόνο και η συνδιακύμανση (covariance ή καλύτερα αυτοσυνδιακύμανση - autocovariance) να εξαρτάται μόνο από τη χρονική απόσταση (lag) μεταξύ των παρατηρήσεων και όχι από τον ίδιο τον χρόνο.

Σημείωση: η σχέση μεταξύ των τιμών των παρατηρήσεων να εξαρτάται μόνο από τη χρονική απόστασή τους lag. Για να γίνει πιο κατανοητό, αν η σχέση μεταξύ τιμών εκπομπών του SO_2 σήμερα και πριν από μία ημέρα είναι η ίδια με τις αντίστοιχες πριν 2 χρόνια τότε η χρονοσειρά έχει σταθερή αυτοσυνδιακύμανση. Εάν όμως, αυτή η σχέση αλλάζει στη διάρκεια του χρόνου, τότε η συνδιακύμανση δεν είναι σταθερή και κατ'επέκταση η χρονοσειρά δεν είναι στάσιμη.

Και τούτο διότι, τα περισσότερα στατιστικά μοντέλα (ARMA, ARIMA κλπ.) βασίζονται στην υπόθεση ότι δεν παρατηρείται σταθερή αυξητική ή μειωτική τάση ή έντονη εποχικότητα. Σε τέτοιες περιπτώσεις, θα πρέπει να προβούμε σε ξεχωριστές τεχνικές για να είναι δυνατή η εφαρμογή των μοντέλων.

Η πιο συνηθισμένη τεχνική είναι η διαφοροποίηση (differencing), μέσω της οποίας αφαιρείται η τάση και σε ορισμένες περιπτώσεις μειώνεται και η εποχικότητα [54], [55]. Οι τύποι διαφοροποίησης είναι οι εξής,

απλή διαφοροποίηση (first order), όπου για κάθε χρονικό σημείο t , αφαιρούμε την τιμή που είχε η χρονοσειρά στο αμέσως προηγούμενο χρονικό σημείο $t - 1$.

Μαθηματικά, υπολογίζετε ως $Y'_t = Y_t - Y_{t-1}$, όπου Y_t είναι η τιμή της χρονοσειράς στο χρόνο t , Y_{t-1} η τιμή της χρονοσειράς στο προηγούμενο χρονικό σημείο (πχ. τον προηγούμενο μήνα) και Y'_t η νέα διαφοροποιημένη πλέον τιμή. Το αποτέλεσμα αυτής της τεχνικής είναι η παραγωγή μίας νέας χρονοσειράς που πλέον εμφανίζει τις μεταβολές μεταξύ διαδοχικών τιμών και όχι τις ίδιες τις τιμές.

διπλή διαφοροποίηση (second order), όπου σε περίπτωση που δε λάβουμε τα επιθυμητά αποτελέσματα από την πρώτη διαφοροποίηση επαναλαμβάνουμε τη διαδικασία στην ήδη 'διαφοροποιημένη' χρονοσειρά ($Y''_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$) και αυτό συνεχίζεται τόσες φορές μέχρι το καθορισμένο d που αποτελεί το βαθμό διαφοροποίησης [54], [55]. Η τεχνική που χρησιμοποιούμε για την εύρεση του βέλτιστου πλήθους διαφοροποιήσεων είναι μέσω στατιστικών ελέγχων στασιμότητας όπως πχ. το ADF (Augmented Dickey-Fuller) [56]. Στο σημείο αυτό να σημειώσουμε ότι, εμπειρικά, εάν και μετά τη διπλή διαφοροποίηση η χρονοσειρά δεν καταστεί στάσιμη, τότε πιθανότατα το μοντέλο που χρησιμοποιείται δεν είναι το κατάλληλο. Και τούτο διότι, παραπάνω διαφοροποιήσεις μετατρέπουν τη χρονοσειρά εξαιρετικά επίπεδη, με αποτέλεσμα να χαθεί πολύτιμη πληροφορία.

εποχική διαφοροποίηση (seasonal differencing), όπου εφαρμόζεται όταν η χρονοσειρά παρουσιάζει επαναλαμβανόμενα μοτίβα σε σταθερά χρονικά διαστήματα, δηλαδή όταν υπάρχει εποχικότητα. Στην περίπτωση αυτή, η νέα τιμή προκύπτει αφαιρώντας από την τρέχουσα παρατήρηση την αντίστοιχη που είχε η χρονοσειρά σε ένα προηγούμενο χρονικό σημείο, το οποίο απέχει ακριβώς ένα πλήρες εποχικό διάστημα (m) από αυτήν. Το εποχικό διάστημα μπορεί να αντιστοιχεί σε οποιαδήποτε περιοδικότητα, όπως, ημέρες, εβδομάδες, μήνες κλπ, ανάλογα πάντα με τη φύση των δεδομένων.

Έτσι, η νέα μετασχηματισμένη πλέον χρονοσειρά γράφεται:

$$y'_t = y_t - y_{t-m}.$$

Με την ανωτέρω διαδικασία, αφαιρείται η περιοδικότητα και η χρονοσειρά γίνεται πιο στάσιμη, διευκολύνοντας έτσι την εφαρμογή του μοντέλου. Το πλήθος των εποχικών διαφοροποιήσεων (D) μπορεί να εντοπιστεί μέσω των ACF και PACF που αναφέραμε παραπάνω ή με τη χρήση αλγορίθμων όπως το `auto_arima`. Συνήθως, μέχρι μια εποχική διαφοροποίηση είναι αρκετή, καθώς πέραν αυτής, ελλοχεύει ο κίνδυνος να χαθεί σημαντική πληροφορία.

ε) ο λευκός θόρυβος, όπως αναφέραμε και στην §3.2.1, είναι μια ακολουθία τυχαίων μεταβλητών με μηδενικό μέσο όρο, σταθερή διακύμανση και απουσία αυτοσυσχέτισης. Δηλαδή, δεν παρουσιάζει κανένα συστηματικό μοτίβο ή δομή που να μπορεί να αξιοποιηθεί προβλεπτικά. Στο πλαίσιο της πρόβλεψης, εξετάζουμε κατά πόσο τα σφάλματα προσαρμογής (residuals), δηλαδή οι διαφορές μεταξύ των παρατηρούμενων τιμών και των τιμών που εκτιμά το μοντέλο εντός του δείγματος, προσεγγίζουν λευκό θόρυβο. Όταν αυτά εμφανίζουν τα χαρακτηριστικά του λευκού θορύβου, υποδηλώνει ότι το μοντέλο έχει αποτυπώσει επαρκώς τη δομή της χρονοσειράς και ότι δεν παραμένει εκμεταλλεύσιμη πληροφορία στα σφάλματα [55].

στ) με τον όρο βX_t , το SARIMAX ενσωματώνει στο μηχανισμό του τη χρήση εξωγενών παραγόντων (exogenous regressors). Δηλαδή, δεν περιορίζεται μόνο στα συστατικά στοιχεία της χρονοσειράς που εξετάζει, αλλά χρησιμοποιεί και άλλες μεταβλητές σε μορφή ξεχωριστών χρονοσειρών, ενισχύοντας έτσι την ικανότητα προβλέψεων [47], [48]. Στην περίπτωση των εκπομπών διοξειδίου του θείου, τέτοιες μεταβλητές είναι για παράδειγμα μετεωρολογικές μετρικές όπως η θερμοκρασία, η υγρασία, ο άνεμος κλπ.

Μαθηματικά, αυτό το πετυχαίνει με την προσθήκη στο μηχανισμό του μίας εξίσωσης παλινδρόμησης όπου η εξαρτημένη μεταβλητή, όπως οι τιμές των παρατηρήσεων του διοξειδίου του θείου, συνδυάζονται με τις ξεχωριστές χρονοσειρές των εξωγενών παραγόντων.

Μετά το θεωρητικό υπόβαθρο παραθέτουμε παρακάτω συνοπτικά τα βήματα εφαρμογής του μοντέλου SARIMAX.

Βήμα 1. Έλεγχος στασιμότητας χρονοσειράς, εάν η χρονοσειρά δεν είναι στάσιμη, εφαρμόζεται κατάλληλη διαφοροποίηση (απλή ή και εποχική), ώστε να καταστεί στάσιμη και χρησιμοποιείται ο έλεγχος ADF (Augmented Dickey–Fuller) για την επαλήθευση της στασιμότητας.

Βήμα 2. Ορισμός των παραμέτρων (p, d, q, P, D, Q, s). Οι παράμετροι p και q (μη εποχικοί όροι AR και MA) και P και Q (εποχικοί όροι AR και MA σε πολλαπλάσια της περιόδου s) εκτιμώνται αρχικά από τα διαγράμματα ACF και PACF της ‘στατικής’ πλέον χρονοσειράς και οι υπονήφιοι συνδυασμοί αυτών, αξιολογούνται βάσει των κριτηρίων AIC και BIC, καθώς και με ταυτόχρονο έλεγχο των σφαλμάτων προσαρμογής. Εναλλακτικά χρησιμοποιείται αυτοματοποιημένη διαδικασία όπως η `auto_arima` σε Python, η οποία επιλέγει τις κατάλληλες παραμέτρους βάσει των κριτηρίων AIC και BIC.

Βήμα 3. Ενσωμάτωση μεταβλητών εξωγενών παραγόντων (X_t), όπως για παράδειγμα μετεωρολογικούς.

Βήμα 4. Εκπαίδευση μοντέλου, με τη μέθοδο μέγιστης πιθανοφάνειας (MLE), όπου εκτιμώνται ταυτόχρονα οι παράμετροι ϕ, θ, Φ και Θ , καθώς και οι συντελεστές παλινδρόμησης των εξωγενών μεταβλητών β .

Βήμα 5. Έλεγχος σφαλμάτων προσαρμογής, με χρήση γραφημάτων ACF/PACF και στατιστικών ελέγχων όπως το Ljung-Box test (σ.σ. σχετική αναφορά γίνεται παρακάτω), ώστε να επιβεβαιωθεί ότι τα σφάλματα προσαρμογής δεν εμφανίζουν στατιστικά σημαντική αυτοσυσχέτιση και επομένως προσεγγίζουν λευκό θόρυβο.

Βήμα 6. Αξιολόγηση μετρικών απόδοσης, βάσει δεικτών, όπως, RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), και SMAPE (Symmetric Mean Absolute Percentage Error).

Βήμα 7. Πρόβλεψη, όπου το τελικό μοντέλο εφαρμόζεται σε νέα δεδομένα.

Τέλος, σημειώνουμε πως η δυνατότητα του SARIMAX να χρησιμοποιεί την εποχικότητα αλλά και εξωγενείς παράγοντες, έχει βρει ευρεία εφαρμογή στα πεδία της οικονομίας, μετεωρολογίας, επιδημιολογίας και φυσικά στην ανάλυση περιβαλλοντικών δεδομένων [47], [55].

3.3.3 Long Short-Term Memory (LSTM) νευρωνικά δίκτυα

Τα LSTM δίκτυα αποτελούν μια ειδική κατηγορία επαναλαμβανόμενων νευρωνικών δικτύων (Recurrent Neural Networks – RNNs), ειδικά σχεδιασμένα για την ανάλυση και πρόβλεψη χρονοσειρών.

Η ιστορία τους ξεκινά στη μακρινή δεκαετία του 1980, από την ανάπτυξη δηλαδή των επαναλαμβανόμενων νευρωνικών δικτύων (RNNs). Τα RNNs σχεδιάστηκαν με την προοπτική να επεξεργάζονται δεδομένα ακολουθιών, όπως κείμενα και χρονοσειρές, προκειμένου να αξιοποιούν την εξάρτηση των δεδομένων στη διάσταση του χρόνου. Ωστόσο, γρήγορα εντοπίστηκε μία σημαντική αδυναμία, ήτοι, η διαδικασία εκπαίδευσης μέσω της μεθόδου οπισθοδιάδοσης στο χρόνο (Back-Propagation Through Time - BPTT) οδηγούσε σε εξαφάνιση ή εκρηκτική αύξηση της βαθμίδας (σ.σ. οι παράγωγοι του σφάλματος ως προς τα βάρη δείχνουν προς ποια κατεύθυνση και πόσο έντονα πρέπει να αλλάξουν τα βάρη ώστε να μειωθεί το σφάλμα πρόβλεψης), γεγονός που καθιστούσε εξαιρετικά δύσκολη την εκμάθηση μακροπρόθεσμων εξαρτήσεων [57].

Για να γίνει κατανοητό, μπορούμε να δούμε πώς λειτουργεί πρακτικά ένα RNN. Σε κάθε χρονικό βήμα t , η ίδια δομή του δικτύου εφαρμόζεται εκ νέου χρησιμοποιώντας τα ίδια βάρη που μοιράζονται όλα τα βήματα της ακολουθίας. Ακολουθώς, παράγονται οι προβλέψεις σε κάθε χρονικό βήμα και υπολογίζεται το συνολικό σφάλμα πρόβλεψης (loss), δηλαδή η διαφορά ανάμεσα στις πραγματικές τιμές και στις προβλέψεις για όλη την ακολουθία [58].

Επειδή τα ίδια βάρη χρησιμοποιούνται σε όλα τα χρονικά βήματα, η οπισθοδιάδοση (BPTT) τα ενημερώνει λαμβάνοντας υπόψη το σφάλμα πρόβλεψης από ολόκληρη την ακολουθία. Έτσι, το δίκτυο ‘μαθαίνει’ τις χρονικές εξαρτήσεις που εκτείνονται σε πολλαπλά χρονικά βήματα της χρονοσειράς και όχι μόνο σε μεμονωμένα. Ωστόσο, η ενημέρωση των βαρών βασίζεται σε παραγώγους που πολλαπλασιάζονται επανειλημμένα σε κάθε χρονικό βήμα.

Συγκεκριμένα, η παράγωγος της συνάρτησης κόστους ως προς τα βάρη, περιλαμβάνει όρους της μορφής:

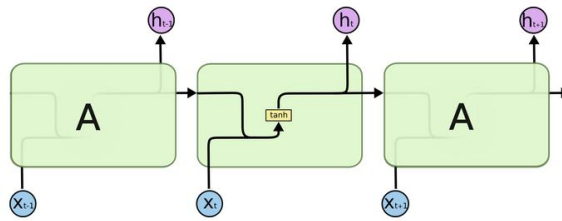
$$\frac{\partial L}{\partial W} \propto \prod_{t=1}^T \frac{\partial h_t}{\partial h_{t-1}}$$

δηλαδή, οι παράγωγοι περιέχουν συνεχόμενους πολλαπλασιασμούς από όρους της μορφής $\frac{\partial h_t}{\partial h_{t-1}}$. Εάν ο πίνακας παραγώγων (Ιακωβιανή) $\frac{\partial h_t}{\partial h_{t-1}}$ έχει ιδιοτιμές μικρότερες του 1 (σ.σ. σε απόλυτο μέγεθος), τότε ο πολλαπλασιασμός επαναλαμβανόμενων μικρών αριθμών οδηγεί σε σταδιακή εξαφάνιση της βαθμίδας (vanishing gradient), αφού οι παράγωγοι τείνουν προς το 0. Αντίθετα, εάν οι ιδιοτιμές είναι μεγαλύτερες του 1, οι παράγωγοι αυξάνονται εκθετικά οδηγώντας έτσι σε εκρηκτική αύξηση της βαθμίδας (exploding gradient).

Στη μεν πρώτη περίπτωση το δίκτυο αδυνατεί να ‘μάθει’ μακροπρόθεσμες εξαρτήσεις, ενώ στη δεύτερη η εκπαίδευση γίνεται ασταθής, καθώς τα βάρη διογκώνονται υπερβολικά και το δίκτυο δε μπορεί να συγκλίνει.

Για να ξεπεραστεί αυτή η αδυναμία, το 1997 οι Hochreiter και Schmidhuber παρουσίασαν την αρχιτεκτονική των Long Short-Term Memory (LSTM) δικτύων. Η καινοτομία τους ήταν η εισαγωγή μιας κυψέλης μνήμης (memory cell) και ενός συστήματος από πύλες (gates) που ουσιαστικά ελέγχουν τη ροή της πληροφορίας, επιτρέποντας έτσι στο δίκτυο να αποθηκεύει αλλά και να ανακαλεί πληροφορία σε πολύ μεγαλύτερους χρονικούς ορίζοντες [59].

Στην αρχιτεκτονική των LSTM, το τμήμα επανάληψης (repeating module) είναι πιο σύνθετο σε σχέση με τα κλασικά RNNs. Σε ένα RNN, το τμήμα επανάληψης περιλαμβάνει μόνο ένα επίπεδο, το οποίο λαμβάνει ως είσοδο την τρέχουσα παρατήρηση x_t και την προηγούμενη έξοδο h_{t-1} , τις συνδυάζει γραμμικά και μέσω μιας μη γραμμικής συνάρτησης ενεργοποίησης (πχ. \tanh) υπολογίζει τη νέα βραχυπρόθεσμη μνήμη h_t , σύμφωνα με τη σχέση $h_t = \tanh(W \cdot [h_{t-1}, x_t] + b)$.



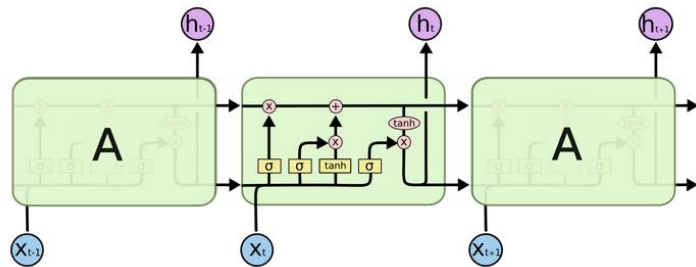
Εικόνα 5: Αρχιτεκτονική ενός απλού RNN αποτελούμενη από ένα μόνο στοιχειώδες layer (πηγή: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

Αντίθετα, σε ένα δίκτυο LSTM, το αντίστοιχο τμήμα αποτελείται από τέσσερα διακριτά επίπεδα (layers) που αλληλοεπιδρούν μεταξύ τους. Τρία από αυτά αντιστοιχούν στις πύλες που ελέγχουν τη ροή της πληροφορίας, ενώ το τέταρτο είναι υπεύθυνο για τον υπολογισμό της υποψήφιας κατάστασης μνήμης (candidate memory state).

Πιο συγκεκριμένα,

- το Forget Gate επίπεδο αποφασίζει ποια πληροφορία θα διατηρηθεί και ποια θα απορριφθεί από τη μακροπρόθεσμη μνήμη C_{t-1} ,
- το Input Gate επίπεδο ρυθμίζει ποια νέα πληροφορία θα εισαχθεί στη μνήμη,
- το επίπεδο υποψήφιας μνήμης υπολογίζει τις νέες τιμές \tilde{C}_t που προτείνονται για εισαγωγή και
- το Output Gate επίπεδο καθορίζει ποιο μέρος της επικαιροποιημένης μνήμης C_t θα αξιοποιηθεί για την παραγωγή της εξόδου h_t .

Έτσι, τα επίπεδα των LSTM ενσωματώνουν τις πύλες ως μηχανισμούς ελέγχου, συνδυάζοντας την αποθήκευση, την ανανέωση και την ανάκληση πληροφορίας σε κάθε χρονικό βήμα t .



Εικόνα 6: Αρχιτεκτονική ενός δικτύου LSTM αποτελούμενο από 4 διασυνδεδεμένα μεταξύ τους layers (πηγή: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

Όσον αφορά την κυψέλη μνήμης,

α) ελέγχεται από τις τρεις πύλες,

- Forget Gate,
- Input Gate και
- Output Gate,

β) διαχειρίζεται δύο βασικές καταστάσεις:

- τη μακροπρόθεσμη μνήμη C_t (Cell state) και
- τη βραχυπρόθεσμη μνήμη h_t (Hidden state), η οποία λειτουργεί και ως έξοδος στο χρονικό βήμα t

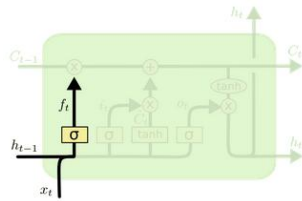
και

γ) σε κάθε χρονικό βήμα t , λαμβάνει:

- την είσοδο x_t (διάνυσμα εισόδου στο βήμα t),
- την προηγούμενη έξοδο h_{t-1} ,
- και τη μακροπρόθεσμη μνήμη από το προηγούμενο χρονικό βήμα C_{t-1} , όπως είχε διαμορφωθεί μέχρι εκείνη τη στιγμή.

Για να κατανοήσουμε καλύτερα τη λειτουργία της κυψέλης μνήμης, θα αναλύσουμε χωριστά τις τρεις πύλες της και το διακριτό ρόλο που επιτελούν [58], [59], [60].

Η Forget Gate, που αποφασίζει ποια πληροφορία από τη μακροπρόθεσμη μνήμη C_{t-1} θα διατηρηθεί ή θα απορριφθεί, περιγράφεται από τη σχέση:



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

Εικόνα 7: Λειτουργία σιγμοειδούς συνάρτησης ενεργοποίησης ενός δικτύου LSTM (πηγή: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

όπου,

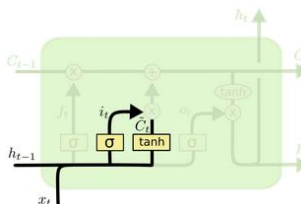
σ είναι η σιγμοειδής συνάρτηση ενεργοποίησης, η οποία παράγει τιμές στο διάστημα από 0 έως 1,

W_f είναι ο πίνακας βαρών που καθορίζει το γραμμικό συνδυασμό της τρέχουσας εισόδου x_t και της προηγούμενης κατάστασης h_{t-1} , αποδίδοντας διαφορετικούς συντελεστές (βάρη) σε κάθε στοιχείο, ώστε να προσδιοριστεί η τελική έξοδος της Forget Gate και

b_f είναι ένας σταθερός συντελεστής που προστίθεται στο γραμμικό συνδυασμό των εισόδων, πριν από τη συνάρτηση ενεργοποίησης και μετατοπίζει το σημείο ενεργοποίησης της πύλης, ανεξάρτητα από τις τρέχουσες τιμές των εισόδων.

Επισημαίνεται ότι τιμές της f_t κοντά στο 0 υποδεικνύουν ότι η αντίστοιχη πληροφορία απορρίπτεται, ενώ τιμές κοντά στο 1 σημαίνουν ότι η πληροφορία διατηρείται.

Η Input Gate, που καθορίζει ποια νέα πληροφορία θα προστεθεί στη μακροπρόθεσμη μνήμη C_t , περιγράφεται από τις σχέσεις:



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Εικόνα 8: Συνδυασμός σιγμοειδούς συνάρτησης ενεργοποίησης και υποψήφιας κατάστασης μνήμης ενός δικτύου LSTM (πηγή: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

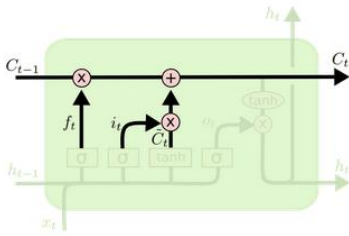
όπου,

i_t είναι η πύλη εισόδου, η οποία, μέσω της σιγμοειδούς συνάρτησης, αποφασίζει το μέγεθος της συνεισφοράς που θα έχει η υποψήφια κατάσταση μνήμης \tilde{C}_t στην τελική μνήμη C_t και

\tilde{C}_t είναι η υποψήφια κατάσταση μνήμης που προκύπτει μέσω της συνάρτησης tanh, δηλαδή είναι ένας ενδιάμεσος υπολογισμός που προτείνει νέες τιμές για τη μακροπρόθεσμη μνήμη C_t . Δεν εισέρχεται απευθείας στη μνήμη, αλλά φιλτράρεται από την πύλη εισόδου i_t . Έτσι, η ενημέρωση της μνήμης γίνεται στοιχείο προς στοιχείο, όπου ορισμένα εξ' αυτών μπορεί να διατηρηθούν, άλλα να αφαιρεθούν και άλλα να εμπλουτιστούν με νέες τιμές.

Με την ίδια λογική που περιγράψαμε και παραπάνω (Forget Gate), τα W_i και b_i χρησιμοποιούνται για τον υπολογισμό της πύλης εισόδου i_t , ενώ τα W_C και b_C χρησιμοποιούνται για τον υπολογισμό της υποψήφιας κατάστασης μνήμης \tilde{C}_t .

Η ενημέρωση της μακροπρόθεσμης μνήμης C_t συνδυάζει τα αποτελέσματα των πυλών Forget Gate και Input Gate και σύμφωνα με τη σχέση:



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

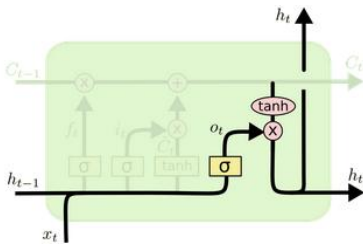
Εικόνα 9: Ενημέρωση μακροπρόθεσμης μνήμης σε ένα δίκτυο LSTM (πηγή: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

όπου,

το πρώτο σκέλος $f_t \odot C_{(t-1)}$ εκφράζει ποια τμήματα της προηγούμενης μνήμης διατηρούνται, ενώ το δεύτερο σκέλος $i_t \odot \tilde{C}_t$ εκφράζει την ενσωμάτωση της νέας πληροφορίας.

Έτσι, στο τέλος κάθε χρονικού βήματος t , το μοντέλο έχει επικαιροποιήσει τόσο τη μακροπρόθεσμη μνήμη C_t όσο και τη βραχυπρόθεσμη μνήμη h_t .

Αφού ενημερωθεί η μακροπρόθεσμη μνήμη C_t , το δίκτυο πρέπει να αποφασίσει ποιο μέρος αυτής της πληροφορίας θα αξιοποιηθεί άμεσα. Αυτή τη λειτουργία αναλαμβάνει η Output Gate, η οποία ρυθμίζει ποιο τμήμα από τη μακροπρόθεσμη μνήμη C_t θα οδηγηθεί προς τη βραχυπρόθεσμη μνήμη h_t , δηλαδή την έξοδο του μοντέλου στο χρονικό βήμα t . Η λειτουργία της περιγράφεται από τις σχέσεις:



$$o_t = \sigma(W_o \cdot [h_{(t-1)}, x_t] + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$

Εικόνα 10: Διαδικασία απόφασης για το τμήμα της μακροπρόθεσμης μνήμης που θα οδηγηθεί προς την έξοδο του μοντέλου (πηγή: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

όπου,

o_t είναι η πύλη εξόδου που αποφασίζει ποιο τμήμα της κατάστασης μνήμης θα περάσει στην έξοδο,

h_t είναι η βραχυπρόθεσμη μνήμη (hidden state) στο χρονικό βήμα t και προκύπτει από την εφαρμογή της συνάρτησης \tanh στη μνήμη C_t , το αποτέλεσμα της οποίας συνδυάζεται με την πύλη εξόδου o_t ,

το σύμβολο \odot δηλώνει τον πολλαπλασιασμό κατά στοιχείο (element-wise multiplication). Για κάθε χρονικό βήμα t , τα διανύσματα h_t, C_t, f_t, i_t, o_t αποτελούνται από n στοιχεία, όπου n είναι ο αριθμός των μονάδων του LSTM. Σημειώνουμε πως ο πολλαπλασιασμός κατά στοιχείο υποδηλώνει ότι για κάθε θέση k , το στοιχείο $h_{t,k}$ προκύπτει από τον πολλαπλασιασμό των αντίστοιχων στοιχείων των διανυσμάτων που συμμετέχουν στη σχέση και με την ίδια λογική που περιγράψαμε παραπάνω, τα W_o και b_o αποτελούν τις παραμέτρους που χρησιμοποιούνται για τον υπολογισμό της πύλης εξόδου o_t .

Η παραπάνω αρχιτεκτονική επιτρέπει στα δίκτυα LSTM να προσαρμόζονται τόσο σε βραχυπρόθεσμες όσο και σε μακροπρόθεσμες εξαρτήσεις μίας χρονοσειράς, κάτι που τα καθιστά ιδανικά για την πρόβλεψη της μελλοντικής πορείας συγκεντρώσεων περιβαλλοντικών ρύπων, όπως το διοξείδιο του θείου, όπου συναντάμε περίπλοκα εποχικά αλλά και μακροχρόνια μοτίβα [61].

Τέλος, αξίζει να αναφέρουμε ότι η σημασία των LSTM αναδείχθηκε ιδιαίτερα στη δεκαετία του 2000, όταν χρησιμοποιήθηκαν με επιτυχία στην αναγνώριση φωνής και στη μηχανική μετάφραση [62]. Σταδιακά, με την αύξηση της υπολογιστικής ισχύος (GPU computing) και την ανάπτυξη μηχανισμών βαθιάς μάθησης (deep

learning), όπως TensorFlow και PyTorch, τα LSTM καθιερώθηκαν ως μια από τις βασικές μεθόδους πρόβλεψης δεδομένων ακολουθιών.

Κεφάλαιο 4: Ανάλυση συγκεντρώσεων διοξειδίου του θείου

Η ανάλυση των συγκεντρώσεων του διοξειδίου του θείου αποτελεί τον πυρήνα της παρούσας διπλωματικής εργασίας και αποσκοπεί στη σε βάθος μελέτη και κατανόηση της συμπεριφοράς του στο χρόνο. Αν και δεν εξαντλήσαμε τις σχετικές βιβλιογραφικές αναφορές, η μεθοδολογία που ακολουθήσαμε περιγράφεται τόσο στα πρόσφατα EMEP Reports (2021 – 2023), που εκδίδονται από το Norwegian Institute for Air Research (NILU), στο πλαίσιο του European Monitoring and Evaluation Programme (EMEP), όσο και στις πρακτικές που χρησιμοποιούν οι κ.κ. Rob J Hyndman και George Athanasopoulos στο έργο τους ‘Forecasting: Principles and Practice (3rd ed) [48].

4.1 Συλλογή δεδομένων παρατηρήσεων SO₂

Βασική πηγή άντλησης δεδομένων αποτέλεσε το Goddard Earth Sciences Data and Information Services Center (GES DISC), ένα από τα επίσημα κέντρα δεδομένων της NASA, το οποίο λειτουργεί στο πλαίσιο του Earth Observing System Data and Information System (EOSDIS). Πρόκειται για τον κεντρικό κόμβο διάθεσης δορυφορικών δεδομένων που σχετίζονται με την ατμόσφαιρα, παρέχοντας ελεύθερη πρόσβαση σε ένα πλήθος μετρικών και γενικά περιβαλλοντικών παρατηρήσεων (GES DISC – <https://disc.gsfc.nasa.gov>).

Στο πλαίσιο της παρούσας διπλωματικής εργασίας χρησιμοποιήσαμε το σύνολο δεδομένων OMI/Aura Sulfur Dioxide (SO₂) Total Column L3 1-Day Best Pixel in 0.25° × 0.25° Version 3 (OMSO2e) [64]. Το συγκεκριμένο σύνολο περιλαμβάνει ημερήσια δορυφορικά δεδομένα μετρήσεων από το δορυφόρο Aura της NASA, ο οποίος είναι εξοπλισμένος με το όργανο Ozone Monitoring Instrument (OMI). Το OMI αποτελεί έναν υπεριώδη φασματογράφο απεικόνισης (UV–VIS imaging spectrometer), σχεδιασμένο για την καταγραφή της ηλιακής ακτινοβολίας που διαχέεται και ανακλάται από την ατμόσφαιρα και την επιφάνεια της Γης, επιτρέποντας την ποσοτική εκτίμηση των συγκεντρώσεων διάφορων ατμοσφαιρικών ρύπων, με υψηλή χωρική και χρονική ανάλυση σε διαφορετικά ύψη της ατμόσφαιρας.

Η μέτρηση των συγκεντρώσεων του διοξειδίου του θείου πραγματοποιείται έμμεσα, δια της μεθόδου Διαφορικής Φασματοσκοπικής Οπτικής Απορρόφησης (Differential Optical Absorption Spectroscopy - DOAS), όπου η απορρόφηση της υπεριώδους ακτινοβολίας σε συγκεκριμένα μήκη κύματος σχετίζεται με την ποσότητα διοξειδίου του θείου στην ατμόσφαιρα. Οι μετρήσεις αυτές υποβάλλονται σε περεταίρω επεξεργασία, μέσω εξειδικευμένων αλγορίθμων και τελικά παράγεται το ημερήσιο σύνολο δεδομένων επιπέδου 3 (Level-3) του OMSO2e. Κάθε κελί πλέγματος (0.25° × 0.25°) περιλαμβάνει μία μόνο τιμή της ολικής ποσότητας διοξειδίου του θείου στο οριακό στρώμα της τροπόσφαιρας (Planetary Boundary Layer – PBL), η οποία προκύπτει από την εφαρμογή ενός βελτιωμένου αλγορίθμου ανάλυσης κύριων συνιστωσών (Principal Component Analysis – PCA) στα δεδομένα του αμέσως προηγούμενου επιπέδου ανάλυσης (Level-2 OMSO2). Από όλες τις διαθέσιμες μετρήσεις του επιπέδου ανάλυσης 2, που εμπίπτουν στο ίδιο κελί πλέγματος και αφορούν στο χρονικό διάστημα 00:00:00–23:59:59 UTC, επιλέγεται εκείνη με την υψηλότερη ποιότητα (best pixel), ώστε να αντιπροσωπεύει τη βέλτιστη ημερήσια εκτίμηση για το συγκεκριμένο γεωγραφικό σημείο.

Εκτός από τη συγκέντρωση του διοξειδίου του θείου, κάθε αρχείο περιλαμβάνει συνοδευτικές παραμέτρους, όπως το ποσοστό νεφοκάλυψης (cloud fraction), το υψόμετρο εδάφους (terrain height), τις γωνίες θέασης ηλίου και δορυφόρου (solar/satellite viewing angles), καθώς και δείκτες ανωμαλιών (row anomaly flags) και ποιότητας (quality flags).

Τα ημερήσια αρχεία δεδομένων που αντλήσαμε από το GES-DISC είναι της μορφής HDF-EOS5 (Hierarchical Data Format, έκδοση 5) και περιλαμβάνουν τις εξής μετρικές παραμέτρους:

Μεταβλητή (Variable)	Περιγραφή (Description)	Μονάδες μέτρησης (Units)
CloudRadianceFraction	Ποσοστό Νεφοκάλυψης	NoUnits
ColumnAmountO3	Στήλη Όζοντος	Dobson Units (DU)
ColumnAmountSO2	Στήλη Διοξειδίου του Θείου	Dobson Units (DU)
Latitude	Γεωγραφικό Πλάτος	degrees_north
LineNumber	Αριθμός Γραμμής	NoUnits
Longitude	Γεωγραφικό Μήκος	degrees_east
OrbitNumber	Αριθμός Τροχιάς	NoUnits
QualityFlags_SO2	Δείκτες Ποιότητας SO ₂	NoUnits
RelativeAzimuthAngle	Σχετική Γωνία Αζιμουθίου	deg (East of North)
SceneNumber	Αριθμός Σκηνής	NoUnits
SolarZenithAngle	Γωνία Ηλιακής Ζενίθιας	deg
TerrainHeight	Υψόμετρο Εδάφους	m
Time	Χρόνος	seconds since 1993-01-01
ViewingZenithAngle	Γωνία Θέασης Δορυφόρου	deg
StructMetadata.0	Μεταδεδομένα Δομής HDFEOS	—

Σημειώνεται πως λόγω του όγκου των υπόψη αρχείων, επιλέξαμε ως χρονική διάρκεια μελέτης, τις μετρήσεις που πραγματοποιήθηκαν στο διάστημα από 1/7/2015 έως 30/6/2025, δηλαδή 10 έτη.

4.2 Συλλογή μετεωρολογικών δεδομένων

Πέραν των δεδομένων των παρατηρήσεων διοξειδίου του θείου, αντλήσαμε και δεδομένα μετεωρολογικών μετρήσεων από το ERA5, το πιο πρόσφατο σύστημα επανάλυσης (reanalysis) του European Centre for Medium-Range Weather Forecasts (ECMWF). Πρόκειται για ένα εκτενές σύνολο δεδομένων [65] που προκύπτει από το συνδυασμό παρατηρήσεων (δορυφορικών, μετεωρολογικών σταθμών, ραδιοβολίσεων κ.ά.) με το Numerical Weather Prediction (NWP) μοντέλο του ECMWF.

Το σύνολο δεδομένων του ERA5 περιλαμβάνει ένα ευρύ φάσμα μετεωρολογικών παραμέτρων, όπως θερμοκρασία, πίεση, άνεμος, σχετική υγρασία, βροχόπτωση κ.α. με χωρική ανάλυση $0.25^\circ \times 0.25^\circ$ (~30 km), δηλαδή ίδια με εκείνη του συνόλου OMSO2e., επιτρέποντάς μας να συνενώσουμε με ευκολία τα δύο αυτά σύνολα.

Και σε αυτήν την περίπτωση, επιλέχθηκαν οι ημερήσιες μετρήσεις που καλύπτουν την περίοδο 1/7/2015 έως 30/6/2025 και η άντληση των σχετικών αρχείων πραγματοποιήθηκε μέσω του συστήματος Copernicus Climate Data Store (CDS), με χρήση της βιβλιοθήκης cdsapi σε περιβάλλον Python.

Οι μετρικές που επιλέξαμε από το σύνολο δεδομένων του ERA5 είναι οι ακόλουθες,

Μεταβλητή	Περιγραφή	Μονάδες μέτρησης
t2m	Θερμοκρασία αέρα στα 2 m	Kelvin (K)
d2m	Θερμοκρασία δρόσου στα 2 m	Kelvin (K)
u10	Ανατολική συνιστώσα ανέμου στα 10 m	m/s
v10	Βόρεια συνιστώσα ανέμου στα 10 m	m/s
blh	Ύψος οριακού στρώματος	m
tcc	Ολική νεφοκάλυψη (κλάσμα)	0–1 (αναλογία)
msl	Ατμοσφαιρική πίεση στη μέση στάθμη θάλασσας	Pa (Pascal)
number	Ensemble member (μέλος συνόλου πρόγνωσης)	Αδιάστατο (integer index)
tp	Ολική κατακρήμνιση	m (ύψος νερού)

4.3 Προεπεξεργασία δεδομένων

Εξαιτίας του μεγάλου όγκου των δεδομένων, όπου περιείχαν ημερήσιες παρατηρήσεις 10 ετών από δύο διαφορετικά σύνολα (OMSO2e και ERA5), σε συνδυασμό με την περιορισμένη ισχύ του οικιακού υπολογιστικού συστήματος που είχαμε στη διάθεσή μας, επιλέξαμε να κρατήσουμε τις μετρήσεις που καταγράφηκαν σε 10 τυχαίες πόλεις από κάθε χώρα της Ευρώπης και κάθε πολιτείας των ΗΠΑ. Δηλαδή, συνολικά 410 πόλεις από 41 χώρες της Ευρώπης και 511 πόλεις από 51 πολιτείες της Αμερικής, καθώς και την πρωτεύουσα των ΗΠΑ (Washington DC). Ακολούθως ταξινομήσαμε τις πόλεις και χώρες/πολιτείες, σε περιφέρειες (Βορράς, Νότος, Ανατολή, Δύση, Κέντρο), λαμβάνοντας υπόψη διάφορους παράγοντες, όπως η επιρροή από θαλάσσια κλίματα και συγκρίσιμο πλήθος πόλεων μεταξύ Ευρώπης και ΗΠΑ και όχι απόλυτα τη γεωγραφική τους θέση.

Όσον αφορά την ενοποίηση των μετρικών από το σύνολο δεδομένων OMSO2e με εκείνες του ERA5, προβήκαμε σε 1:1 αντιστοίχιση των γεωγραφικών συντεταγμένων (latitude, longitude) καταγραφής των παρατηρήσεων. Με τον τρόπο αυτό, κάθε τιμή συγκέντρωσης διοξειδίου του θείου συνδυάστηκε με τις μετεωρολογικές μετρήσεις του ίδιου γεωγραφικού κελιού. Να σημειώσουμε εδώ πως για τα μετεωρολογικά δεδομένα ERA5, οι ωριαίες τιμές ανάλυσης μετατράπηκαν σε ημερήσιες χρονοσειρές και οι ποσοτικές μετρικές (θερμοκρασία, άνεμος, πίεση κλπ) υπολογίστηκαν ως ημερήσιοι μέσοι όροι, ενώ οι αθροιστικές μεταβλητές (πχ. ολική κατακρήμνιση) υπολογίστηκαν ως ημερήσια αθροίσματα.

Για την κατά το δυνατό ταχύτερη και ευκολότερη επεξεργασία των δεδομένων, μετατρέψαμε τα ημερήσια ενοποιημένα αρχεία σε μορφή parquet, μία column oriented μορφή αποθήκευσης δεδομένων, η οποία έχει σχεδιαστεί για μεγάλα σύνολα δεδομένων και βελτιστοποιεί την ανάγνωση και την επεξεργασία τους σε σύγχρονα περιβάλλοντα ανάλυσης (big data frameworks) ([63]Apache Parquet. (2025)).

Μετά την ενοποίηση των δύο βασικών συνόλων δεδομένων διενεργήσαμε ελέγχους πληρότητας και ποιότητας, από όπου προέκυψαν σημαντικά ευρήματα, ήτοι,

α) απουσίαζαν περίπου 30 ημερήσιες εγγραφές από το σύνολο δεδομένων OMSO2e (δεν υπήρχαν στη βάση του GES DISC) και

β) η NASA συνοδεύει κάθε παρατήρηση του συνόλου OMSO2e με ένα δείκτη ποιότητας (Quality Flag), ο οποίος αποτυπώνει την αξιοπιστία της μέτρησης βάσει παραμέτρων όπως, η ζενίθια γωνία ηλιακής ακτινοβολίας, η νεφοκάλυψη κ.α. Μετά την εφαρμογή των κριτηρίων που συστήνονται από τη NASA, για την επιλογή αξιόπιστων μετρήσεων (good data), το 46% των δορυφορικών παρατηρήσεων διοξειδίου του θείου (σ.σ. και όζοντος, που περιλαμβάνεται στο σύνολο OMSO2e), χαρακτηρίστηκαν ως μη αξιόπιστες.

Για την αντιμετώπιση των παραπάνω αδυναμιών προβήκαμε στην άντληση τριών νέων συνόλων δεδομένων (M2TUNXADG [64], M2T1NXAER [65] και M2T1NXCHM [66]), τα οποία έχουν ήδη υποστεί σχετική προεπεξεργασία από το σύστημα reanalysis της NASA (Global Modeling and Assimilation Office – GMAO) MERRA-2. Το σύστημα αυτό βασίζεται στο Numerical Weather Prediction μοντέλο GEOS (Goddard Earth Observing System) και ενσωματώνει δεδομένα παρατηρήσεων από δορυφόρους αλλά και επίγειους σταθμούς. Εστιάζει δε στην αναπαράσταση αερολυμάτων και άλλων χημικών στοιχείων της ατμόσφαιρας, όπως SO₂, O₃, PM_{2.5}, CO κλπ. Αν και παρέχει παγκόσμια κάλυψη, έχει χωρική ανάλυση 0.5° × 0.625°, δηλαδή διαφορετική από εκείνη του αρχικού μας ενοποιημένου συνόλου (OMSO2e και ERA5). Προς τούτο, για την ορθή αντιστοίχιση των γεωγραφικών συντεταγμένων εφαρμόσαμε τον τύπο Haversine [67], ο οποίος λαμβάνει υπόψη τη σφαιρικότητα της Γης. Η μέθοδος αυτή χρησιμοποιεί τριγωνομετρικές συναρτήσεις για να υπολογίσει το μήκος του τόξου κύκλου με ακτίνα R = 6.371 km, που αποτελεί και τη μέση ακτίνα της Γης. Έτσι, προκύπτει μια αξιόπιστη και επαρκώς ακριβή εκτίμηση της απόστασης δύο σημείων, σε αντίθεση με τις απλοποιημένες ευκλείδειες προσεγγίσεις, που δε λαμβάνουν υπόψη τους τις αποστάσεις σε καμπύλη επιφάνεια.

Τα δεδομένα του MERRA-2 για το διοξείδιο του θείου, προέρχονται από το προϊόν M2T1NXAER, το οποίο παρέχει τη μεταβλητή SO2CMASS, δηλαδή το column mass density σε μονάδες kg/m². Για να είναι συμβατά με τις μετρήσεις του προϊόντος OMSO2e, που εκφράζονται σε DU, προβήκαμε στη μετατροπή των μονάδων μέτρησης. Συγκεκριμένα, η μάζα στήλης (kg/m²) μετατράπηκε σε αριθμό μορίων ανά m², μέσω διαίρεσης με τη μοριακή μάζα του διοξειδίου του θείου και πολλαπλασιασμού με τον αριθμό Avogadro και στη συνέχεια

μετατράπηκε σε DU, χρησιμοποιώντας τον τυπικό συντελεστή $1 \text{ DU} = 2.69 \times 10^{20} \text{ molecules/m}^2$. Η διαδικασία αυτή εφαρμόστηκε σε όλα τα ημερήσια αρχεία SO₂, πριν την ενοποίηση τους με τα αντίστοιχα σύνολα O₃ και μετεωρολογικών παραγόντων.

Μετά από αυτή την εξέλιξη, ο όγκος των δεδομένων που κληθήκαμε να επεξεργαστούμε και η απαραίτητη πλέον προς τούτο ισχύς, υπερέβαινε κατά πολύ τις δυνατότητες ενός οικιακού υπολογιστικού συστήματος. Προκειμένου να μη χαθεί πολύτιμος χρόνος, καταφύγαμε καταρχάς σε υπολογιστικές πλατφόρμες νέφους που παρέχουν σημαντικούς πόρους για την επεξεργασία μεγάλου όγκου δεδομένων.

Όσον αφορά τη χρονική κλίμακα των παρατηρήσεων, αυτή μετατράπηκε από ημερήσιες χρονοσειρές σε μηνιαίες, απομειώνοντας σημαντικά τον προς επεξεργασία όγκο δεδομένων. Ωστόσο, ακολουθήσαμε διαφορετική τακτική στον υπολογισμό των μέσων, ανά χωρικό επίπεδο. Συγκεκριμένα, για τις χρονοσειρές των πόλεων υπολογίστηκαν οι μέσες μηνιαίες τιμές που προέκυψαν από τα ημερήσια δεδομένα. Για τις χώρες τις περιφέρειες και το επίπεδο ηπείρου, εφαρμόστηκε η μέθοδος των σταθμισμένων μέσων όρων, προκειμένου οι πόλεις με υψηλότερη μακροπρόθεσμη συγκέντρωση SO₂ (και O₃) να έχουν ελαφρώς μεγαλύτερη βαρύτητα στη διαμόρφωση των συνολικών τιμών.

Αναλυτικότερα, για κάθε πόλη υπολογίστηκε η μακροπρόθεσμη μέση τιμή διοξειδίου του θείου και οι σχετικές τιμές κανονικοποιήθηκαν εντός της χώρας\περιφέρειας\ηπείρου, ώστε να προκύψουν τα αντίστοιχα βάρη:

$$w_{SO_2,i} = \frac{S\bar{O}_{2i}}{\sum_j S\bar{O}_{2j}}$$

Στη συνέχεια, με τα βάρη αυτά υπολογίστηκαν οι μηνιαίοι σταθμισμένοι μέσοι όροι των αντίστοιχων χωρικών επιπέδων, δίδοντας μας μία πιο ρεαλιστική εκτίμηση των μετρήσεων.

Το τελικό σύνολο περιέχει μηνιαίες παρατηρήσεις για τις οντότητες πόλης, χώρας, περιφέρειας και ηπείρου και διαμορφώθηκε με τις ακόλουθες μεταβλητές:

Μεταβλητή	Περιγραφή	Μονάδες Μέτρησης
city	Όνομασία Πόλης	-
country	Όνομασία Χώρας	-
region	Όνομα Περιφέρειας	-
year_month	Έτος - μήνας παρατήρησης (YYYY-MM)	DateTime
SO2_mean	Μέση τιμή SO ₂	Dobson Units (DU)
O3_mean	Μέση τιμή O ₃	Dobson Units (DU)
t2m_mean	Μέση θερμοκρασία αέρα στα 2m	Kelvin (K)
d2m_mean	Μέση θερμοκρασία δρόσου στα 2m	Kelvin (K)
u10_mean	Μέση τιμή οριζόντιας συνιστώσας ανέμου στα 10 m (διεύθυνση ανατολή - δύση)	m/s
v10_mean	Μέση τιμή οριζόντιας συνιστώσας ανέμου στα 10 m (διεύθυνση βορράς - νότος)	m/s
blh_mean	Μέση τιμή ύψους οριακού στρώματος	m
tcc_mean	Μέση τιμή ολικής νεφοκάλυψης	0-1 (αναλογία)
tp_mean	Μέση τιμή ολικής κατακρήμνισης	m (ύψος νερού)
mssl_mean	Μέση τιμή ατμοσφαιρικής πίεσης στη μέση στάθμη της θάλασσας	Pa (Pascal)
cloudfrac_mean	Μέση τιμή νεφοκάλυψης (μέτρηση από OMSO2)	0-1 (αναλογία)
dist_mean	Μέση τιμή απόστασης από συντεταγμένες μέτρησης	Km

Σημείωση: Τα σύνολα δεδομένων που αφορούν στις οντότητες χώρας, περιφέρειας και ηπείρου, φέρουν στις αντίστοιχες ονομασίες των μεταβλητών την ένδειξη 'mean_weighted' αντί 'mean', που υποδεικνύει ότι ως μέσες τιμές λαμβάνονται υπόψη οι μηνιαίοι σταθμισμένοι μέσοι όροι που αναφέραμε παραπάνω.

4.4 Τεχνική διαδικασία αποσύνθεσης χρονοσειρών

Η διαδικασία της αποσύνθεσης που ακολουθήσαμε περιλάμβανε αρχικά την εφαρμογή της μεθόδου STL (με `robust = true`) και εφόσον δεν πληρούνταν συγκεκριμένα κριτήρια στατιστικής επάρκειας, γίνονταν αυτόματη μετάβαση σε προσαρμογή της επιλογής καταλληλότερου χρονικού παραθύρου. Για την εύρεση του χρονικού παραθύρου, ακολουθήσαμε μια ανεξάρτητη διαδικασία φασματικής ανάλυσης της αρχικής χρονοσειράς, μέσω περιοδογράμματος (periodogram), βασισμένου σε μετασχηματισμό Fourier. Η ανάλυση αυτή επιτρέπει τον προσδιορισμό της κυρίαρχης συχνότητας της χρονοσειράς, η οποία μετατρέπεται στην αντίστοιχη περίοδο επανάληψης του εποχικού μοτίβου. Η εκτιμώμενη αυτή περίοδος χρησιμοποιήθηκε αποκλειστικά για την αυτόματη προσαρμογή του εποχικού παραθύρου εξομάλυνσης της STL (adaptive STL), αποφεύγοντας έτσι τον χειροκίνητο ορισμό παραμέτρων [51], [68].

Στην περίπτωση αποτυχίας και αυτής της διαδικασίας εφαρμόστηκε και τρίτη εναλλακτική με χρήση της πολυεποχικής αποσύνθεσης MSTL (Multi-Seasonal STL), η οποία αποτελεί επέκταση της STL και εφαρμόζει διαδοχικά εξομάλυνση LOESS για κάθε προκαθορισμένη περίοδο (σ.σ. στη δική μας περίπτωση χρησιμοποιήσαμε 2 περιόδους, 12 και 6 μηνών), ώστε να απομονώσει ξεχωριστά πολυεπίπεδα εποχικά μοτίβα. Έπειτα αφαιρούνται τα εποχικά σήματα, εξάγεται η τάση και υπολογίζονται τα κατάλοιπα.

Αν αποτύγχανε και αυτή η μέθοδος, τότε επιλέγονταν εκείνη που εμφάνιζε τη μεγαλύτερη συνολική βαθμολογία διαγνωστικών ελέγχων, με την ταυτόχρονη επισήμανση ως 'fall-back'.

Οι διαγνωστικοί έλεγχοι που πραγματοποιήθηκαν τεχνικά, για την επιλογή της καλύτερης μεθόδου αποσύνθεσης, είναι οι ακόλουθοι,

α) Ljung-Box test

Ο έλεγχος Ljung-Box [69] εφαρμόστηκε στα κατάλοιπα της αποσύνθεσης, ώστε να διαπιστωθεί αν αυτά είναι μη συσχετισμένα μεταξύ τους (λευκός θόρυβος). Ο στατιστικός αυτός έλεγχος βασίζεται στη συνάρτηση αυτοσυσχέτισης και δίδεται από τη σχέση,

$$Q = n(n + 2) \sum_{k=1}^h \frac{r_k^2}{n - k}$$

όπου n είναι το μέγεθος του δείγματος, r_k η δειγματική αυτοσυσχέτιση κατά τη χρονική υστέρηση k , και h ο αριθμός των υστερήσεων.

Η μηδενική υπόθεση (H_0) είναι ότι τα κατάλοιπα δεν εμφανίζουν αυτοσυσχέτιση. Οπότε, αν $p > 0.05$, η υπόθεση H_0 δεν απορρίπτεται υποδηλώνοντας ότι η αποσύνθεση κατάφερε ικανοποιητικά να απομονώσει τα κατάλοιπα από την εποχικότητα και την τάση.

β) Έλεγχος ADF (Augmented Dickey-Fuller)

Ο έλεγχος (ADF) [70], όπως ορίζεται και από την ονομασία του, αποτελεί μία 'επαυξημένη' εκδοχή του Dickey Fuller test και χρησιμοποιείται για τη διερεύνηση της στασιμότητας μιας χρονοσειράς, δηλαδή αν οι στατιστικές της ιδιότητες της (μέσος, διακύμανση, αυτοσυσχέτιση) παραμένουν σταθερές στο χρόνο. Μια μη στάσιμη χρονοσειρά χαρακτηρίζεται από την ύπαρξη μοναδιαίας ρίζας (unit root), κάτι που σημαίνει ότι οι τιμές της εξαρτώνται έντονα από τις προηγούμενες και δε συγκλίνουν γύρω από έναν σταθερό μέσο όρο. Η εξίσωση παλινδρόμησης έχει τη μορφή:

$$\Delta y_t = \alpha + \beta t + \gamma y_{(t-1)} + \sum_{i=1}^p \theta_i \Delta y_{t-i} + \varepsilon_t$$

όπου,

Δy_t , εκφράζει τη μεταβολή της χρονοσειράς μεταξύ δύο διαδοχικών περιόδων,
 α , ο τυχόν σταθερός όρος, όπου αν υφίσταται σημαίνει ότι η χρονοσειρά έχει σημείο αναφοράς διαφορετικό από το 0 (σ.σ. μια μέση τιμή γύρω από την οποία κινείται).

βt , η γραμμική τάση (αν υπάρχει γενική ανοδική ή καθοδική πορεία στο χρόνο),
 γ , ο συντελεστής του όρου $y_{(t-1)}$ που ελέγχεται για την ύπαρξη μοναδιαίας ρίζας,
και ε_t είναι το τυχαίο σφάλμα.

Η μηδενική υπόθεση H_0 είναι ότι $\gamma = 0$, δηλαδή ότι η χρονοσειρά διαθέτει μοναδιαία ρίζα και δεν είναι στάσιμη. Αν η p-value του ελέγχου ADF είναι μικρότερη από 0.05, τότε απορρίπτεται η H_0 και η χρονοσειρά θεωρείται στάσιμη.

γ) Durbin-Watson d statistic test

Ο δείκτης DW [71], χρησιμοποιείται για τη διερεύνηση αυτοσυσχέτισης πρώτης τάξης στα κατάλοιπα μιας χρονοσειράς, δηλαδή για να εξεταστεί αν κάθε τιμή σχετίζεται με την αμέσως προηγούμενή της (lag = 1) και υπολογίζεται από τη σχέση,

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

όπου, το d λαμβάνει τιμές από 0 έως 4, e_t είναι τα κατάλοιπα της εξεταζόμενης χρονικής στιγμής και n το πλήθος των παρατηρήσεων.

Τιμές d κοντά στο 2 υποδηλώνουν απουσία αυτοσυσχέτισης ή πιο σωστά τα κατάλοιπα μπορούν να θεωρηθούν ως μη συσχετισμένα. Τιμές d μικρότερες από 1.5 ή μεγαλύτερες από 2.5 θεωρούμε (σ.σ. εμπειρικά) ότι υποδηλώνουν θετική ή αρνητική αυτοσυσχέτιση αντίστοιχα.

Όπως προαναφέραμε, για κάθε χρονοσειρά εφαρμόστηκε αυτόματος αλγόριθμος που αξιολογούσε διαδοχικά τα αποτελέσματα των μεθόδων αποσύνθεσης και επιλέγονταν η πρώτη κατά σειρά μέθοδος όπου πληρούνταν ταυτόχρονα και οι τρεις ως άνω διαγνωστικοί έλεγχοι, δηλαδή, Ljung-Box p-value > 0.05 και ADF p-value < 0.05 και $1.5 < DW \text{ result} < 2.5$.

Από αυτήν τη διαδικασία προέκυψαν τα κάτωθι αποτελέσματα:

α) μέθοδος αποσύνθεσης που επιλέχθηκε αυτόματα:

Europe							
Level	n	STL	STL Fallback	Adaptive STL	Adaptive STL Fallback	MSTL	MSTL Fallback
City	410	340 (82.9%)	19 (4.6%)	28 (6.8%)	15 (3.7%)	-	8 (2%)
Country	41	29 (70.7%)	4 (9.8%)	6 (14.6%)	2 (4.9%)	-	-
Region	5	3 (60%)	-	-	2 (40%)	-	-
Continent	1	1 (100%)	-	-	-	-	-

USA							
Level	n	STL	STL Fallback	Adaptive STL	Adaptive STL Fallback	MSTL	MSTL Fallback
City	511	416 (81.4%)	16 (3.1%)	50 (9.8%)	21 (4.1%)	-	8 (1.6%)
State	52	42 (80.8%)	1 (1.9%)	7 (13.5%)	2 (3.8%)	-	-
Region	5	5 (100%)	-	-	-	-	-
Continent	1	1 (100%)	-	-	-	-	-

β) αποτελέσματα διαγνωστικών ελέγχων κατά τη διαδικασία αποσύνθεσης:

Europe				USA			
Level	n	Succeeded	Failed	Level	n	Succeeded	Failed
City	410	368 (89.7%)	42 (10.3%)	City	511	466 (91.2%)	45 (8.8%)
Country	41	35 (85.4%)	6 (14.6%)	State	52	49 (94.2%)	3 (5.8%)
Region	5	3 (60%)	2 (40%)	Region	5	5 (100%)	-
Continent	1	1 (100%)	-	Continent	1	1 (100%)	-

Σαφώς και δεν περιοριστήκαμε στα ως άνω αποτελέσματα, αλλά μετά την ολοκλήρωση της διαδικασίας αποσύνθεσης, προβήκαμε και σε οπτική εξέταση των διαγραμμάτων των καταλοίπων, με σκοπό να αξιολογηθεί κατά πόσο αυτά προσεγγίζουν τη συμπεριφορά του ‘λευκού θορύβου’ και να εντοπιστούν τυχόν υπολειπόμενα μοτίβα ή ακραίες τιμές.

Η εξέταση αυτή, βασίστηκε σε ένα σύνολο διαγνωστικών διαγραμμάτων, που παρήχθησαν για κάθε οντότητα κάθε χωρικού επιπέδου.

Συγκεκριμένα, χρησιμοποιήσαμε τα διαγράμματα,

- Residuals over time, που απεικονίζει τα κατάλοιπα σε συνάρτηση με το χρόνο, ώστε να ελέγξουμε για την τυχόν ύπαρξη συστηματικών μοτίβων ή μεταβαλλόμενες διακυμάνσεις. Ιδανικά, τα κατάλοιπα θα πρέπει να είναι τυχαία κατανομημένα γύρω από το 0, χωρίς εμφανή τάση ή εποχικότητα.
- Histogram + KDE (Kernel Density Estimation), που παρουσιάζει την κατανομή των καταλοίπων. Ιδανικά, τα κατάλοιπα προσεγγίζουν κανονική κατανομή με μέσο = 0 και συμμετρική διάταξη γύρω του.
- Q-Q Plot (Quantile-Quantile Plot), που συγκρίνει τα κατανομή των καταλοίπων με τη θεωρητική κανονική κατανομή. Ιδανικά, τα σημεία θα πρέπει να βρίσκονται πολύ κοντά στη διαγώνιο, χωρίς συστηματικές αποκλίσεις γύρω από αυτή και χωρίς να εμφανίζουν έντονες καμπύλες στα άκρα (ένδειξη ύπαρξης ακραίων τιμών ή απόκλιση από την κανονικότητα).
- ACF (Auto Correlation Function) και PACF (Partial Auto Correlation Function), που δείχνουν την αυτοσυσχέτιση και μερική αυτοσυσχέτιση των καταλοίπων αντίστοιχα. Ιδανικά οι συντελεστές αυτοσυσχέτισης πρέπει να κυμαίνονται τυχαία γύρω από το 0 και να βρίσκονται εντός των ορίων εμπιστοσύνης (95%), γεγονός που υποδηλώνει πως τα κατάλοιπα προσεγγίζουν το ‘λευκό θόρυβο’.
- Residuals vs Fitted (Trend + Seasonal), που ελέγχει αν τα κατάλοιπα παρουσιάζουν συστηματική σχέση με τις την εποχικότητα και την τάση. Ιδανικά, τα σημεία θα πρέπει να κατανέμονται τυχαία, χωρίς σαφές μοτίβο, ένδειξη ότι η τάση και η εποχικότητα έχουν απομονωθεί επιτυχώς.

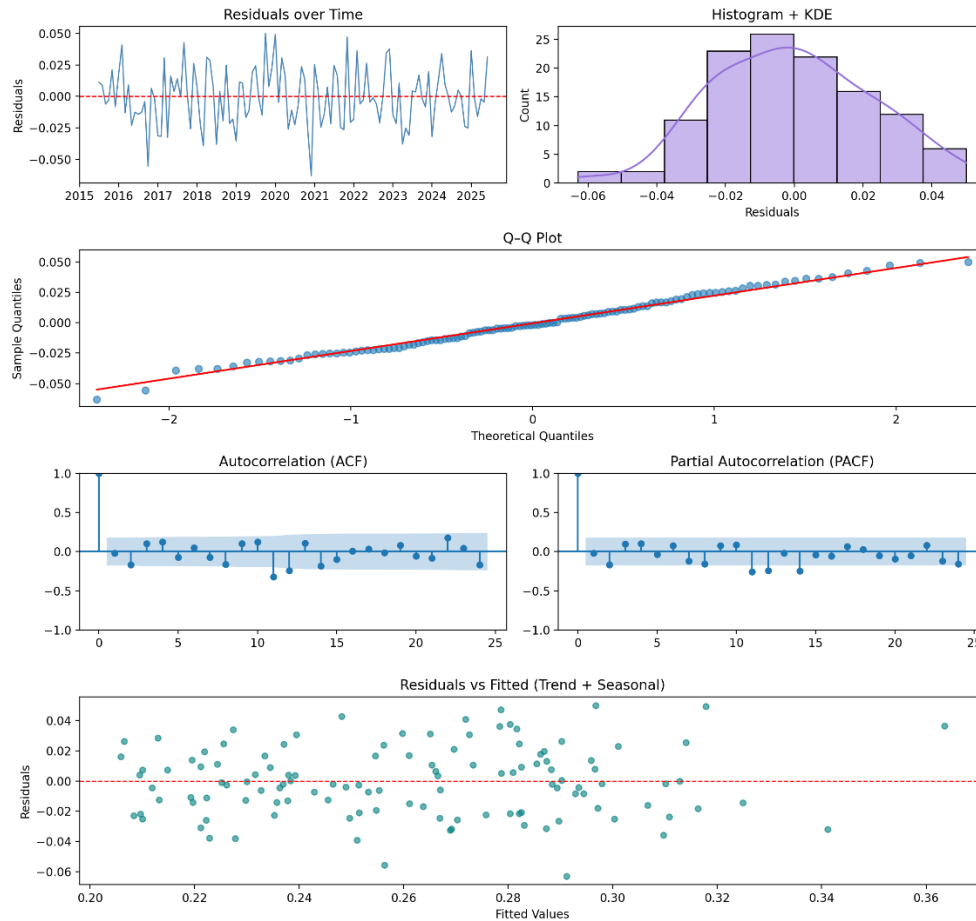
Ας δούμε αναλυτικά 2 παραδείγματα, ένα όπου οι διαγνωστικοί στατιστικοί έλεγχοι δεν ήταν επιτυχείς και ένα όπου ήταν επιτυχείς.

Παράδειγμα 1

Πρόκειται για την περίπτωση της Ανατολικής Ευρώπης, όπου ο έλεγχος Ljung-Box είχε αποτέλεσμα $p = 0.000257$, υποδεικνύοντας στατιστικά σημαντική αυτοσυσχέτιση στα κατάλοιπα. Οπότε και οι διαγνωστικοί έλεγχοι οδηγήθηκαν σε κατάσταση ‘Failed’.

Παρακάτω παρουσιάζονται τα διαγράμματα διάγνωσης:

East - Residual Diagnostics (Region)



Γράφημα 9: Οπτική απεικόνιση καταλοίπων αποσύνθεσης για την Ανατολική περιφέρεια της Ευρώπης.

Από την οπτική εξέταση των ανωτέρω, προκύπτουν τα εξής,

Residuals over time: Τα κατάλοιπα κατανέμονται τυχαία γύρω από το 0, χωρίς εμφανή τάση ή κάποιο επαναλαμβανόμενο εποχικό μοτίβο. Η διακύμανσή τους είναι σχετικά σταθερή σε όλο το χρονικό διάστημα, χωρίς να συσσωρεύονται μεγάλες θετικές ή αρνητικές αποκλίσεις. Δηλαδή, φαίνεται ότι η αποσύνθεση απομόνωσε σωστά την τάση και την εποχικότητα και τα κατάλοιπα μάλλον αποτελούν ‘λευκό θόρυβο’.

Histogram + KDE: Το ιστόγραμμα είναι σχεδόν συμμετρικό γύρω από το 0 και η καμπύλη KDE έχει μορφή που προσεγγίζει την κανονική κατανομή. Άρα, τα κατάλοιπα είναι σχεδόν κανονικά κατανεμημένα, οπότε δε φαίνεται μετά την αποσύνθεση να έχουν απομείνει συστηματικά σφάλματα.

Q-Q Plot: Τα σημεία ευθυγραμμίζονται σε μεγάλο βαθμό με τη διαγώνιο, γεγονός που δείχνει ότι το κεντρικό τμήμα της κατανομής των καταλοίπων προσεγγίζει την κανονική κατανομή. Μικρές αποκλίσεις εμφανίζονται μόνο στα άκρα, υποδηλώνοντας περιορισμένη παρουσία ακραίων τιμών. Η συνολική συμπεριφορά παραμένει ομαλή και συνεπής με την εικόνα του Histogram.

ACF και PACF: Οι μπάρες (κάθετες μπλε σκούρες γραμμές με ένα χρωματισμένο κύκλο στο τέλος τους) των συντελεστών αυτοσυσχέτισης βρίσκονται εντός της σκιασμένης μπλε περιοχής εμπιστοσύνης (95%) και οι τιμές τους κυμαίνονται κοντά στο 0, με ελαφρώς τυχαίες θετικές και αρνητικές διακυμάνσεις. Παρόμοια εικόνα παρατηρούμε και στις μπάρες μερικής αυτοσυσχέτισης, όπου οι τιμές κυμαίνονται γύρω από το 0, χωρίς κάποιο μοτίβο ή επαναλαμβανόμενες αιχμές. Ωστόσο, μεταξύ των lag 11 - 14, φαίνεται να υπάρχει μια ασθενής αυτοσυσχέτιση και μάλιστα με κάποια αλληλουχία (οι δύο πρώτες μπλε μπάρες είναι συνεχόμενες και έχουν το ίδιο μήκος).

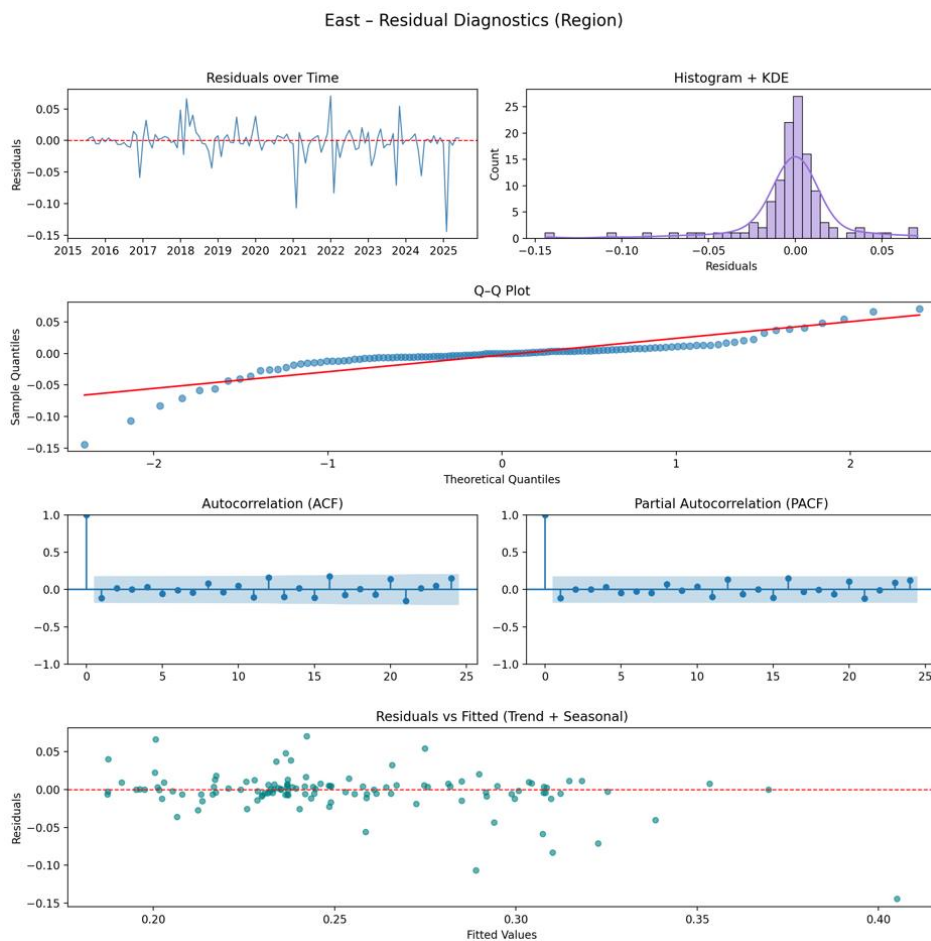
Residuals vs Fitted: Τα σημεία κατανέμονται τυχαία γύρω από τη γραμμή $y = 0$, χωρίς εμφανή μοτίβα ή σχήμα. Δηλαδή, δεν υπάρχει ένδειξη συστηματικής σχέσης μεταξύ των τιμών [Εποχικότητας + Τάσης] και των καταλοίπων.

Φαίνεται λοιπόν, ότι ο έλεγχος Ljung-Box test, αναγνώρισε τη σχεδόν αμελητέα ύπαρξη αυτοσυσχέτισης στα κατάλοιπα, μεταξύ των lag 11 - 14, με αποτέλεσμα η Ανατολική περιφέρεια της Ευρώπης να ταξινομηθεί στην κατηγορία 'Fail'. Και τούτο διότι, ο υπόψη έλεγχος είναι αρκετά συντηρητικός και ακόμη και η ελάχιστη απόκλιση έχει εξαιρετικά σημαντικό αντίκτυπο στο αποτέλεσμά του (σ.σ. προκύπτει από τον τύπο που αναφέραμε παραπάνω, όπου ουσιαστικά τετραγωνίζονται και αθροίζονται οι αυτοσυσχετίσεις, με αποτέλεσμα ακόμη και μία μικρή απόκλιση από το 0, εκτοξεύει το αποτέλεσμα του ελέγχου).

Παράδειγμα 2

Πρόκειται για την περίπτωση της Ανατολικής περιφέρειας των ΗΠΑ, όπου όλοι οι διαγνωστικοί έλεγχοι ήταν επιτυχείς. Μάλιστα ο έλεγχος Ljung-Box είχε αποτέλεσμα $p = 0.710911$, ιδιαίτερα υψηλό, υποδηλώνοντας ότι δεν υπάρχει στατιστικά σημαντική αυτοσυσχέτιση στα κατάλοιπα.

Παρακάτω παρουσιάζονται τα διαγράμματα διάγνωσης:



Γράφημα 10: Οπτική απεικόνιση καταλοίπων αποσύνθεσης για την Ανατολική περιφέρεια των ΗΠΑ.

Από την οπτική εξέταση των παραπάνω διαγραμμάτων προκύπτουν τα εξής:

Residuals over time: Τα κατάλοιπα κατανέμονται γύρω από το 0, χωρίς εμφανή τάση ή κάποιο επαναλαμβανόμενο εποχικό μοτίβο. Ωστόσο, υπάρχουν κάποιες έκτακτες διακυμάνσεις από το 2020 και μετά, αλλά δε φαίνεται να έχουν κάποια συστηματικότητα ή μοτίβο. Παρά ταύτα, τα κατάλοιπα φαίνεται να είναι τυχαία και στατικά κάτι που υποδηλώνει πως η αποσύνθεση απομόνωσε σωστά την τάση και την εποχικότητα.

Histogram + KDE: Οι τιμές του ιστογράμματος αν και κινούνται γύρω από το 0 φαίνεται να υπάρχει μια καθαρή ασυμμετρία στη διακύμανση των τιμών προς τα αριστερά (αρνητικές τιμές). Η καμπύλη KDE αποκλίνει από την κανονική κατανομή, ιδιαίτερα στα αριστερά άκρα.

Q-Q Plot: Η πλειοψηφία των σημείων βρίσκεται πλησίον της διαγωνίου, σχεδόν σε όλο το μήκος της, εκτός από την αριστερή ουρά, όπου φαίνεται οι τιμές να αποκλίνουν έντονα προς τα κάτω. Η εικόνα αυτή συνάδει και με το σχετικό Histogram, υποδεικνύοντας την ύπαρξη έκτακτων γεγονότων, που όμως δε φαίνεται να επηρεάζουν τη συνολική εικόνα της χρονοσειράς.

ACF και PACF: Τόσο οι μπάρες αυτοσυσχέτισης όσο και μερικής αυτοσυσχέτισης βρίσκονται εντός της σκιασμένης μπλε περιοχής εμπιστοσύνης (95%) και οι τιμές κυμαίνονται πολύ κοντά στο 0, χωρίς έκτακτες κορυφές. Ένδειξη δηλαδή, ότι τα κατάλοιπα δεν παρουσιάζουν υπολειμματική συσχέτιση και συνεπώς δε διατηρούν κάποια δομή μετά την αποσύνθεση.

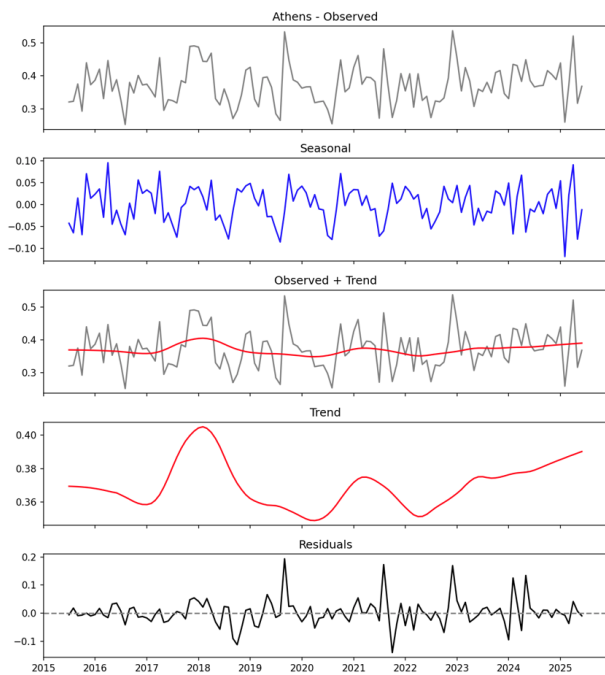
Residuals vs Fitted: Τα σημεία κατανέμονται τυχαία γύρω από τη γραμμή $y = 0$, χωρίς εμφανή μοτίβα ή σχήμα. Κάποιες συγκεντρώσεις τιμών περίπου μεταξύ των τιμών 0.23 - 0.24, δε φαίνεται να έχουν κάποιο συγκεκριμένο μοτίβο ή συστηματική απόκλιση. Συνεπώς, δεν προκύπτει ένδειξη συστηματικής σχέσης μεταξύ των τιμών [Εποχικότητας + Τάσης] και των καταλοίπων.

Από την ερμηνεία των ανωτέρω διαγνωστικών διαγραμμάτων, επιβεβαιώνεται ότι η διαδικασία της αποσύνθεσης κατάφερε να απομονώσει επιτυχώς το εποχικό μοτίβο και την τάση, ενώ ανέδειξε την ύπαρξη κάποιων ακραίων τιμών που ίσως να αφορούν σε έκτακτα βραχυπρόθεσμα γεγονότα.

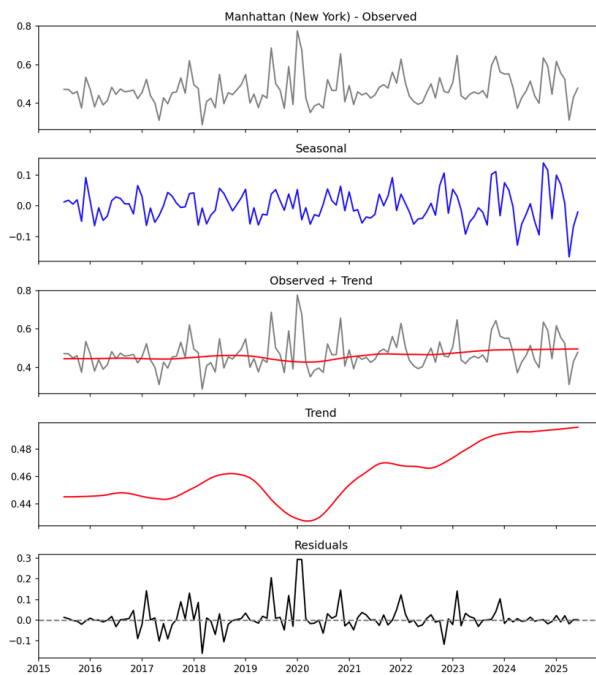
Συνολικά, η εικόνα της Ανατολικής περιφέρειας της Ευρώπης, φαίνεται πιο ‘καθαρή’ από εκείνη των ΗΠΑ, ωστόσο, στη μεν πρώτη περίπτωση, ο τεχνικός διαγνωστικός έλεγχος ήταν ανεπιτυχής (εξαιτίας της τιμής του Ljung-Box test), στη δε δεύτερη, όλοι οι διαγνωστικοί ήταν επιτυχείς. Το γεγονός αυτό, μας δείχνει τη σημαντικότητα στην ερμηνεία των αποτελεσμάτων συνδυαστικά με την οπτική διάγνωση τους, όπου αυτό βέβαια είναι δυνατό. Σαφώς δε θα πρέπει να παραλείψουμε το γεγονός ότι υπάρχουν και άλλες τεχνικές που μπορούν να συνδράμουν στην καλύτερη αξιολόγηση των αποτελεσμάτων της αποσύνθεσης και ειδικότερα στην ανάλυση των καταλοίπων, κάτι για το οποίο θα ασχοληθούμε σε ξεχωριστό κεφάλαιο. Σε κάθε περίπτωση, η παρουσία μεμονωμένων ακραίων τιμών στα κατάλοιπα δεν υποδηλώνει σφάλμα ή αποτυχία της διαδικασίας της αποσύνθεσης, αλλά υποδεικνύει την ύπαρξη έκτακτων βραχυπρόθεσμων επεισοδίων, κάτι αναμενόμενο άλλωστε σε περιβαλλοντικά δεδομένα.

Καταλήγοντας, σημειώνουμε ότι η αποσύνθεση των χρονοσειρών, μέσω των μεθόδων που περιγράψαμε παραπάνω, στη συντριπτική πλειοψηφία των οντοτήτων όλων των χωρικών επιπέδων, μας επιτρέπει να προχωρήσουμε στην ανάλυση των εποχικών μοτίβων και της μακροπρόθεσμης τάσης των συγκεντρώσεων διοξειδίου του θείου.

Οπτικά, το αποτέλεσμα της αποσύνθεσης παρατίθεται παρακάτω, ενδεικτικά για 2 πόλεις, όπου απεικονίζονται η αρχική χρονοσειρά των παρατηρήσεων διοξειδίου του θείου και οι συνιστώσες εποχικότητας, τάσης και καταλοίπων,



Γράφημα 11: Συνιστώσες αποσύνθεσης για τις συγκεντρώσεις διοξειδίου του θείου της Αθήνας.



Γράφημα 12: : Συνιστώσες αποσύνθεσης για τις συγκεντρώσεις διοξειδίου του θείου του Manhattan.

Στις επόμενες ενότητες θα ασχοληθούμε αναλυτικά με κάθε μία εκ των ως άνω συνιστωσών.

4.5 Ανάλυση εποχικότητας

Η ανάλυση των εποχικών μοτίβων στις συγκεντρώσεις διοξειδίου του θείου, έγινε σε όλες τις οντότητες όλων των χωρικών επιπέδων. Έτσι προβήκαμε σε διάφορες μεθόδους αξιολόγησης της εποχικότητας, με σκοπό να αξιολογήσουμε τα αποτελέσματα της διαδικασίας αποσύνθεσης και να μελετήσουμε την εν γένει συμπεριφορά των εποχικών μοτίβων.

4.5.1 Αξιολόγηση έντασης εποχικών μοτίβων

Μετά την αποσύνθεση των χρονοσειρών, προβήκαμε σε ποσοτική αξιολόγηση της έντασης της εποχικότητας για κάθε οντότητα. Για το σκοπό αυτό χρησιμοποιήθηκε ο δείκτης Seasonal Strength (F_s) που προτάθηκε αρχικά από κ.κ. Rob J Hyndman και George Athanasopoulos στο έργο τους ‘Forecasting: Principles and Practice (2018)’ [48], ως μέτρο ποσοτικής αποτίμησης της έντασης της εποχικότητας σε μια χρονοσειρά και ορίζεται μαθηματικά ως εξής:

$$F_s = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right)$$

όπου, R_t , είναι το κατάλοιπο και S_t η εποχική συνιστώσα της αποσύνθεσης. Τιμές του δείκτη F_s κοντά στο 0 υποδηλώνουν απουσία ή πολύ ασθενής εποχικότητα, ενώ τιμές κοντά στο 1 υποδεικνύουν ισχυρά επαναλαμβανόμενα εποχικά μοτίβα. Για ευκολία στην παρουσίαση και ερμηνεία των αποτελεσμάτων, τα ταξινομήσαμε σε 5 διακριτές κατηγορίες ως εξής:

Τιμή F_s	Ερμηνεία
$F_s < 0.2$	Πολύ ασθενής (Very Weak)
$0.2 \leq F_s < 0.4$	Ασθενής (Weak)
$0.4 \leq F_s < 0.6$	Μέτρια (Moderate)
$0.6 \leq F_s < 0.8$	Ισχυρή (Strong)
$F_s \geq 0.8$	Πολύ ισχυρή (Very Strong)

Τα αποτελέσματα από την παραπάνω διαδικασία δίδουν μία πρώτη εικόνα για την ένταση της εποχικότητας, σε όλες τις οντότητες, όλων των χωρικών επιπέδων. Αναλυτικότερα,

City level	Europe		USA	
strength_rank	count	%	count	%
Strong	41	10.00	51	9.98
Moderate	160	39.03	232	45.40
Weak	183	44.63	174	34.05
Very Weak	26	6.34	54	10.57

Στην Ευρώπη, η πλειονότητα των πόλεων (περίπου 84 %) εμφανίζει ασθενή έως μέτρια εποχικότητα στις μηνιαίες συγκεντρώσεις διοξειδίου του θείου, με μόλις ένα μικρό ποσοστό (10 %) να έχει ισχυρή εποχικότητα. Η εικόνα αυτή αντανακλά την επίδραση μικτών πηγών ρύπανσης και σαφώς την επίδραση των μετεωρολογικών παραγόντων που είναι διαφορετικοί ανά γεωγραφική τοποθεσία.

Στις ΗΠΑ παρατηρείται παρόμοια κατανομή, αλλά με ελαφρώς υψηλότερο ποσοστό πόλεων που εμφανίζουν μέτρια εποχικότητα (45.4 %) και περισσότερες περιπτώσεις με ασθενή εποχικότητα (10,6 %), γεγονός που υποδηλώνει πιο ανομοιογενή χωρική διακύμανση και μεγαλύτερη εξάρτηση από τοπικούς παράγοντες εκπομπών διοξειδίου του θείου και κλίματος.

Συνολικά, τα αποτελέσματα δείχνουν ότι η εποχικότητα των συγκεντρώσεων διοξειδίου του θείου σε επίπεδο πόλεων είναι παρούσα αλλά όχι ιδιαίτερα έντονη.

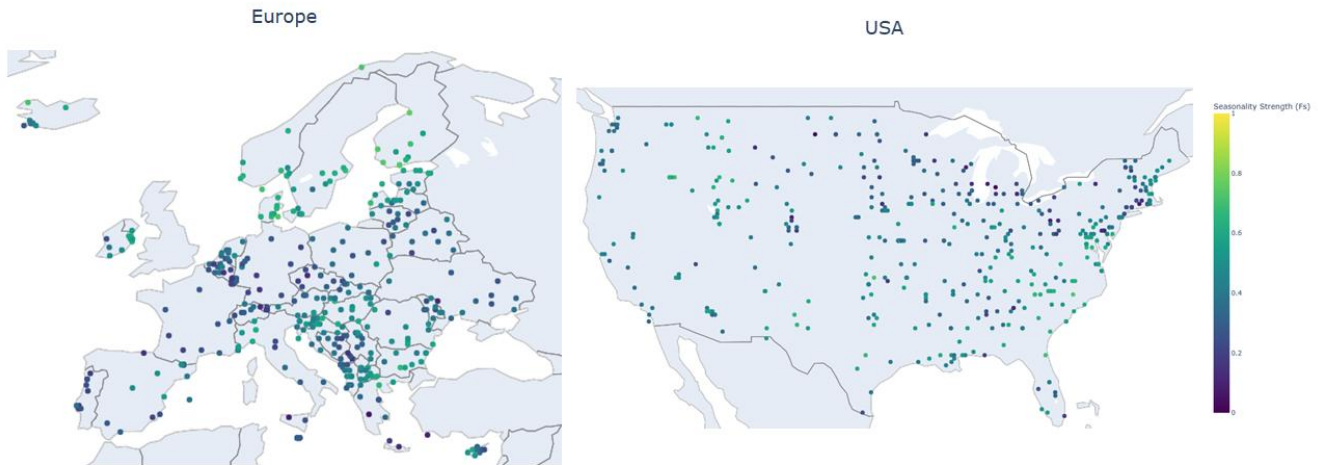
Προκειμένου να επιβεβαιώσουμε ότι η ένταση της εποχικότητας μεταβάλλεται όσο ανεβαίνουμε χωρικό επίπεδο, καταρτίσαμε τους παρακάτω πίνακες που εμφανίζουν τα αποτελέσματα σε επίπεδο χωρών και περιφερειών:

Country level	Europe		USA		Region level	Europe		USA	
strength_rank	count	%	count	%	strength_rank	count	%	count	%
Very Strong	1	2.44	1	1.93	Very Strong	1	20	-	-
Strong	11	26.83	17	32.69	Strong	1	20	2	40
Moderate	15	36.59	21	40.38	Moderate	2	40	2	40
Weak	14	34.14	8	15.38	Weak	1	20	1	20
Very Weak	-	-	5	9.62	Very Weak	-	-	-	-

Όπως άλλωστε ήταν αναμενόμενο, η ένταση της εποχικότητας σε εθνικό επίπεδο, εμφανίζεται εντονότερη σε σχέση με εκείνη των πόλεων. Περίπου δύο στις τρεις χώρες τόσο στην Ευρώπη όσο και στις ΗΠΑ εμφανίζουν μέτρια ή ισχυρή εποχικότητα, ενώ εμφανίζονται και περιπτώσεις (ελάχιστες) με πολύ έντονη εποχικότητα. Η αιτία πίσω από αυτή την εικόνα, είναι πως οι τιμές σε επίπεδο κρατών προέρχονται από τις μέσες τιμές των πόλεων που αντιπροσωπεύουν, οπότε τα τοπικά χαρακτηριστικά των τελευταίων εξομαλύνονται και αναδεικνύεται πιο καθαρά το εποχικό σήμα.

Σε επίπεδο περιφερειών, είναι εμφανές ότι τα εποχικά μοτίβα γίνονται ακόμη πιο καθαρά, λόγω της εξομάλυνσης, ενώ παρατηρούμε και μία σχετική ομοιομορφία μεταξύ της Ευρώπης και των ΗΠΑ. Μάλιστα, αξίζει να σημειωθεί ότι σε αυτό το χωρικό επίπεδο οι πόλεις ομαδοποιούνται γεωγραφικά, κάτι που σημαίνει ότι πόλεις στην ίδια περιφέρεια, τουλάχιστο βάσει των μετεωρολογικών συνθηκών, είναι πιθανό να εμφανίζουν παρόμοια εποχικά μοτίβα.

Η προβολή της έντασης των εποχικών μοτίβων σε χωρικό επίπεδο φαίνεται παρακάτω:



Εικόνα 11: Συγκριτική απεικόνιση έντασης εποχικότητας Ευρώπης και ΗΠΑ.

Στην Ευρώπη παρατηρείται μία γεωγραφική διαφοροποίηση, όπου εμφανίζονται υψηλότερες τιμές στις βόρειες και κεντρικές πόλεις, πιθανότατα λόγω της χαμηλής ανάμιξης της ατμόσφαιρας κατά τους χειμερινούς μήνες και ίσως τις αυξημένες εκπομπές λόγω χρήσης θερμαντικών σωμάτων. Αντίθετα, στις πόλεις της Μεσογείου η εποχικότητα εμφανίζεται ασθενέστερη, πιθανότατα αντανακλώντας το ηπιότερο κλίμα.

Στις ΗΠΑ, ο χωρικός καταμερισμός παρουσιάζει μεγαλύτερη διασπορά σε σχέση με εκείνον της Ευρώπης. Οι περισσότερες πόλεις εμφανίζουν μέτρια έως ασθενή εποχικότητα, χωρίς σαφή γεωγραφική συγκέντρωση. Ελαφρώς εντονότερη εποχικότητα εντοπίζεται στις βορειοδυτικές και μεσο-ανατολικές περιοχές, πιθανότατα λόγω ψυχρότερου κλίματος. Συμπερασματικά, η ένταση της εποχικότητας συγκεντρώσεων διοξειδίου του θείου στις ΗΠΑ έχει περισσότερο τοπικό χαρακτήρα, σε αντίθεση με ότι παρατηρήσαμε στην Ευρώπη.

4.5.2 Ποιοτική αξιολόγηση των εποχικών μοτίβων

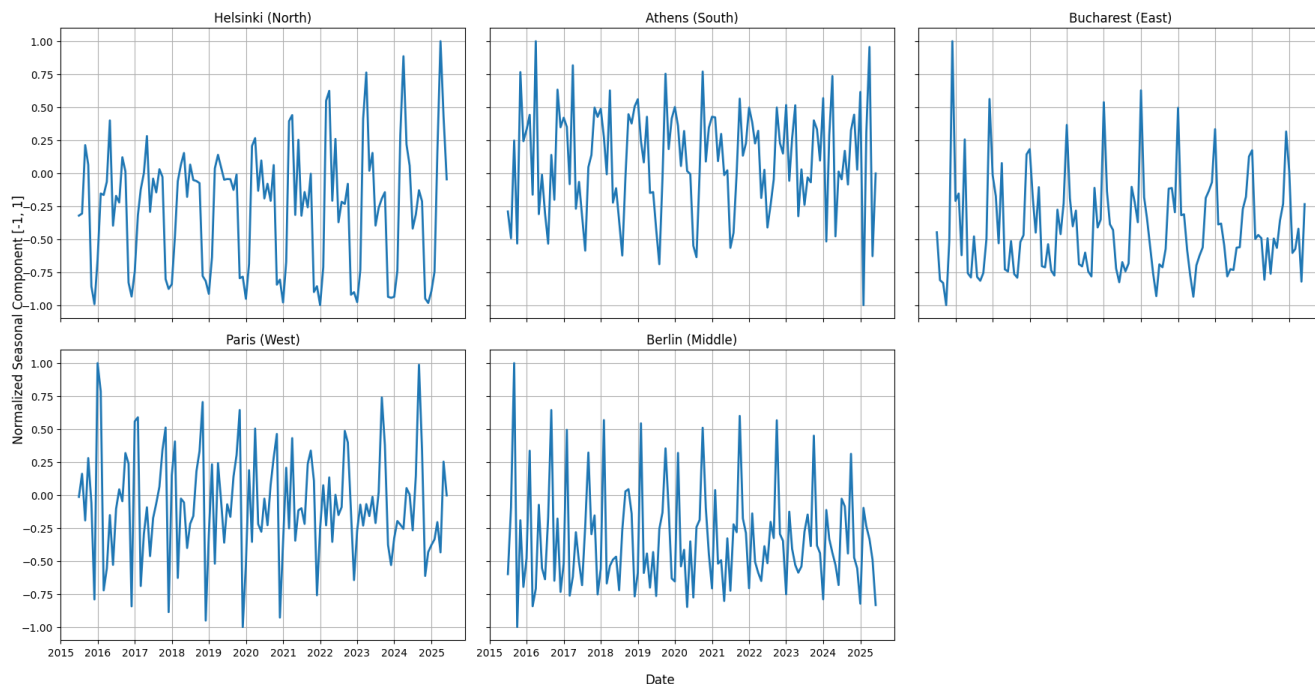
Μετά τη διερεύνηση της έντασης των εποχικών μοτίβων, στην παρούσα υποενότητα θα επικεντρωθούμε στα ποιοτικά χαρακτηριστικά τους, όπως τη μορφή, τη σταθερότητα που παρουσιάζουν αλλά και τη συμπεριφορά τους στο χρόνο.

I. Κανονικοποίηση εποχικής συνιστώσας.

Η διαδικασία της κανονικοποίησης της εποχικής συνιστώσας, μας επιτρέπει τη συγκριτική απεικόνιση των μοτίβων μεταξύ διαφορετικών πόλεων, ανεξάρτητα από τα απόλυτα επίπεδα συγκέντρωσης. Στα παρακάτω διαγράμματα παρουσιάζονται οι κανονικοποιημένες εποχικές συνιστώσες για πέντε αντιπροσωπευτικές πόλεις της Ευρώπης (Helsinki, Athens, Bucharest, Paris, Berlin), μία από κάθε περιφέρεια και ομοίως των Ηνωμένων Πολιτειών (Manhattan, Orlando, Indianapolis, Sacramento, Denver), για την περίοδο 2015 - 2025.

Normalized Seasonal Components of SO₂ (Europe, 2015–2025)

— SO₂ (normalized)

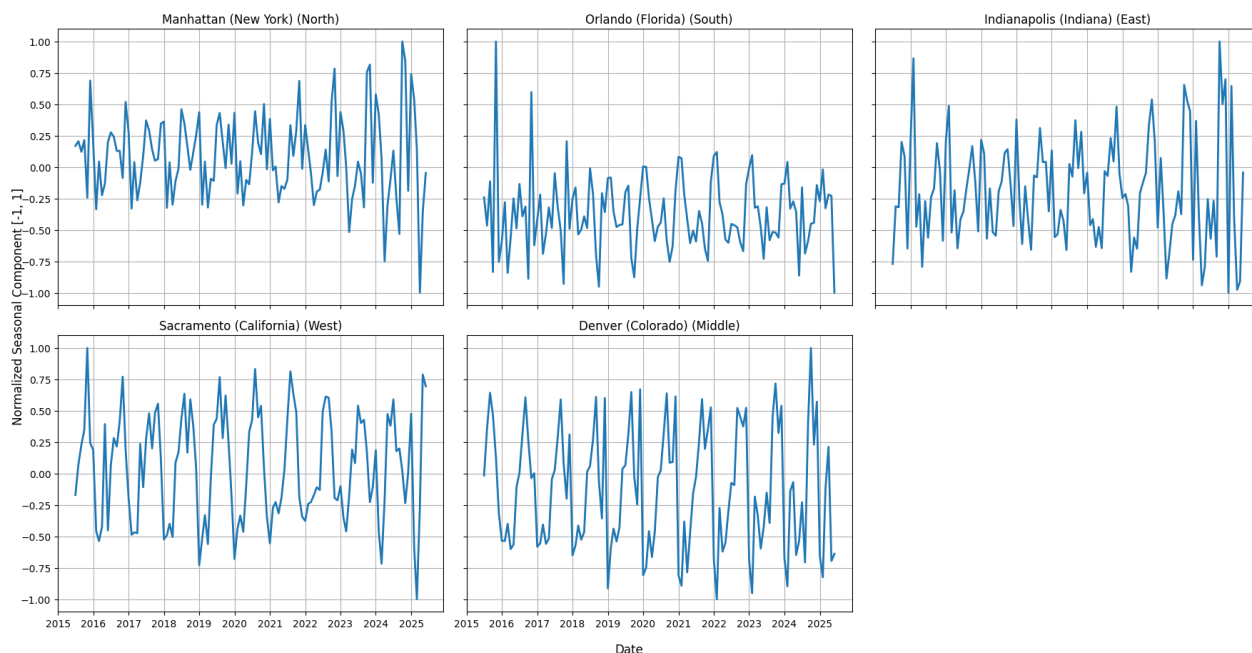


Γράφημα 13: Απεικόνιση εποχικής συνιστώσας επιλεγμένων Ευρωπαϊκών πόλεων.

Στην Ευρώπη, τα εποχικά μοτίβα φαίνεται να έχουν κοινή συμπεριφορά με ξεκάθαρες υψηλές συγκεντρώσεις το χειμώνα και ελάχιστες την άνοιξη και το καλοκαίρι. Το μοτίβο αυτό παραμένει σταθερό στις περισσότερες πόλεις, αν και η ένταση των διακυμάνσεων διαφοροποιείται γεωγραφικά. Στις βόρειες και ανατολικές πόλεις (Helsinki, Bucharest) οι κορυφώσεις είναι εντονότερες και πιο συμμετρικά κατανομημένες, ενώ στις δυτικές και νότιες (Paris, Athens) η εποχικότητα φαίνεται να είναι πιο ήπια και λιγότερο συμμετρική, γεγονός που μπορεί να αποδοθεί σε ηπιότερες θερμοκρασιακές μεταβολές.

Normalized Seasonal Components of SO₂ (USA, 2015–2025)

— SO₂ (normalized)



Γράφημα 14: Απεικόνιση εποχικής συνιστώσας επιλεγμένων πόλεων των ΗΠΑ.

Στις ΗΠΑ, τα εποχικά μοτίβα παρουσιάζουν μικρότερη συνοχή μεταξύ γεωγραφικών περιοχών. Στο Manhattan και Indianapolis εμφανίζονται θετικές εποχικές αιχμές, πιθανότατα λόγω του κλίματος που δεν είναι ιδιαίτερα θερμό. Στο Orlando διαπιστώνουμε ένα συνδυασμό υψηλής έντασης διακυμάνσεις με ακανόνιστες όμως κορυφές ιδιαίτερα μετά το 2020. Στο Orlando το κλίμα είναι σχετικά ήπιο οπότε δεν αποκλείεται αυτή η συμπεριφορά να οφείλεται σε άλλους παράγοντες, πέραν του κλίματος. Στο Sacramento και το Denver εμφανίζονται έντονες και ξεκάθαρους εποχικού μοτίβου διακυμάνσεις με αρνητική μάλιστα συμπεριφορά, κάτι που ίσως να οφείλεται και σε πολιτικές περιορισμού των εκπομπών διοξειδίου του θείου.

Συνολικά, η ανάλυση των κανονικοποιημένων εποχικών συνιστωσών δύναται να χρησιμοποιηθεί για συγκριτικούς σκοπούς, αποκαλύπτοντας σε αρκετές περιπτώσεις διαφορές στη χρονική κατανομή και την ένταση των εποχικών μοτίβων μεταξύ γεωγραφικών περιοχών.

II. Διάγνωση περιοδικών μοτίβων

Η οπτική διερεύνηση των κανονικοποιημένων εποχικών συνιστωσών ανέδειξε σαφείς περιοδικές διακυμάνσεις στις συγκεντρώσεις του διοξειδίου του θείου, με διαφοροποιήσεις ως προς τη χρονική διακύμανση και το εύρος τους, μεταξύ διαφορετικών γεωγραφικών περιοχών. Για να αξιολογηθεί εάν τα εποχικά αυτά μοτίβα παρουσιάζουν σταθερή και περιοδικά επαναλαμβανόμενη συμπεριφορά στο χρόνο, ακολουθεί η φασματική ανάλυση (periodicity diagnostics), μέσω της οποίας μπορούμε να εκτιμήσουμε την κυρίαρχη περίοδο και την ισχύ της εποχικότητας.

Η φασματική ανάλυση αποτελεί βασικό εργαλείο για την ανίχνευση περιοδικών μοτίβων σε χρονοσειρές και στηρίζεται στην παραδοχή ότι κάθε χρονοσειρά μπορεί να θεωρηθεί ως συνδυασμός κυμάτων διαφορετικών συχνοτήτων και εντάσεων (σ.σ. πόσο ισχυρή είναι η συμβολή της αντίστοιχης συχνότητας στη συνολική χρονοσειρά), όπως περιγράφεται από τη θεωρία Fourier. Μέσω αυτής της προσέγγισης εκτιμάται η φασματική πυκνότητα ισχύος (spectral density), η οποία δείχνει τη συνολική διακύμανση της χρονοσειράς στις επιμέρους συχνότητες [49], [68].

Προς τούτο, υπολογίσαμε το φάσμα συχνοτήτων της χρονοσειράς κάθε οντότητας όλων των χωρικών επιπέδων, ώστε να προσδιορίσουμε την περίοδο με τη μέγιστη φασματική ισχύ. Η διαδικασία αυτή πραγματοποιήθηκε με ρυθμό δειγματοληψίας μία παρατήρηση ανά μήνα και εφαρμόστηκε τόσο στις αρχικές παρατηρήσεις των συγκεντρώσεων του διοξειδίου του θείου, όσο και στην εποχική συνιστώσα. Τα αποτελέσματα για τις πόλεις περιγράφονται παρακάτω,

Raw Data	Europe		USA		Seasonal	Europe		USA	
Dominant period (m)	count	%	count	%	Dominant period (m)	count	%	count	%
12	223	54.4	477	93.3	12	225	54.9	481	94.1
6	134	32.7	6	1.2	6	141	34.4	13	2.5
4	17	4.2	5	1.0	4	18	4.4	6	1.2
3	5	1.2	10	2.0	3	9	2.2	6	1.2
2	6	1.5	2	0.4	2	16	3.9	5	1.0
Other	25	6.0	11	2.1	Other	1	0.2	-	-

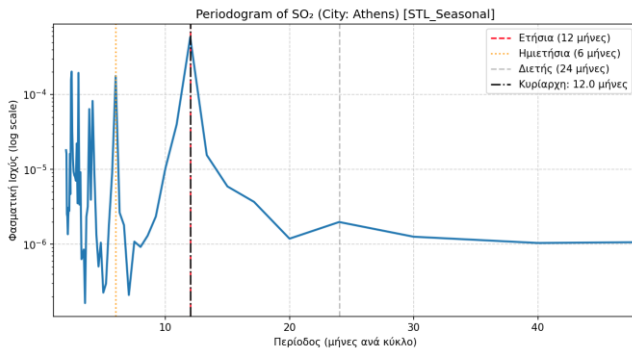
Η κατανομή των κυρίαρχων περιόδων, μέσω της φασματικής ανάλυσης, ανέδειξε την ύπαρξη μιας εντελώς διαφορετικής δομής εποχικότητας, μεταξύ των πόλεων της Ευρώπης και εκείνων των ΗΠΑ.

Συγκεκριμένα, στις ΗΠΑ η συντριπτική πλειοψηφία των πόλεων εμφανίζει μονοσήμαντα ετήσια περίοδο (12 μήνες), τόσο στα αρχικά δεδομένα όσο και στην εποχική συνιστώσα, γεγονός που υποδηλώνει ότι ο ετήσιος κύκλος διοξειδίου του θείου είναι εξαιρετικά σταθερός και με επαναλαμβανόμενη χρονική συνέπεια στο πέρασμα των ετών. Αντίθετα, στην Ευρώπη, αν και η ετήσια περίοδος παραμένει κυρίαρχη σε ποσοστό περίπου στο 55% των πόλεων, παρατηρείται έντονη παρουσία δευτερευόντων περιοδικότητων κυρίως εξαμηνιαίων, δείχνοντας πως το εποχικό σήμα είναι πιο σύνθετο. Ωστόσο, μία τόσο σημαντική διαφορά, σε σύγκριση με τις πόλεις των ΗΠΑ,

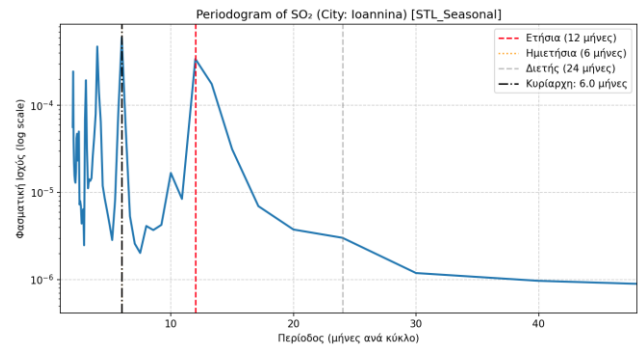
δε μπορεί να αποδοθεί αποκλειστικά στη γεωγραφική θέση ή στις διαφορετικές κλιματολογικές συνθήκες που επικρατούν στην Ευρώπη, αλλά φαίνεται να συνδέεται και με άλλους παράγοντες, όπως η ανομοιογένεια στις ανθρωπογενείς εκπομπές ρύπων και οι διαφοροποιήσεις στο βαθμό εφαρμογής των περιβαλλοντικών πολιτικών.

Οπτικά, η πιο συνηθισμένη μέθοδος είναι το periodogram, στο οποίο ο οριζόντιος άξονας αντιστοιχεί στη συχνότητα (cycles per unit time) και ο κατακόρυφος άξονας στην ένταση της κάθε συχνότητας. Οι κορυφές του διαγράμματος αντιστοιχούν σε κυρίαρχες περιόδους, δηλαδή σε χρονικά διαστήματα όπου η διακύμανση της χρονοσειράς είναι εντονότερα περιοδική.

Παρακάτω παραθέτουμε τα περιοδογράμματα 2 πόλεων, της Αθήνας και των Ιωαννίνων, όπου η μεν πρώτη έχει ξεκάθαρο ετήσιο εποχικό κύκλο ενώ στα Ιωάννινα η κυρίαρχη περίοδος είναι 6μηνιαία.



Γράφημα 15: Περιοδογράμμα εποχικής περιόδου Αθήνας.



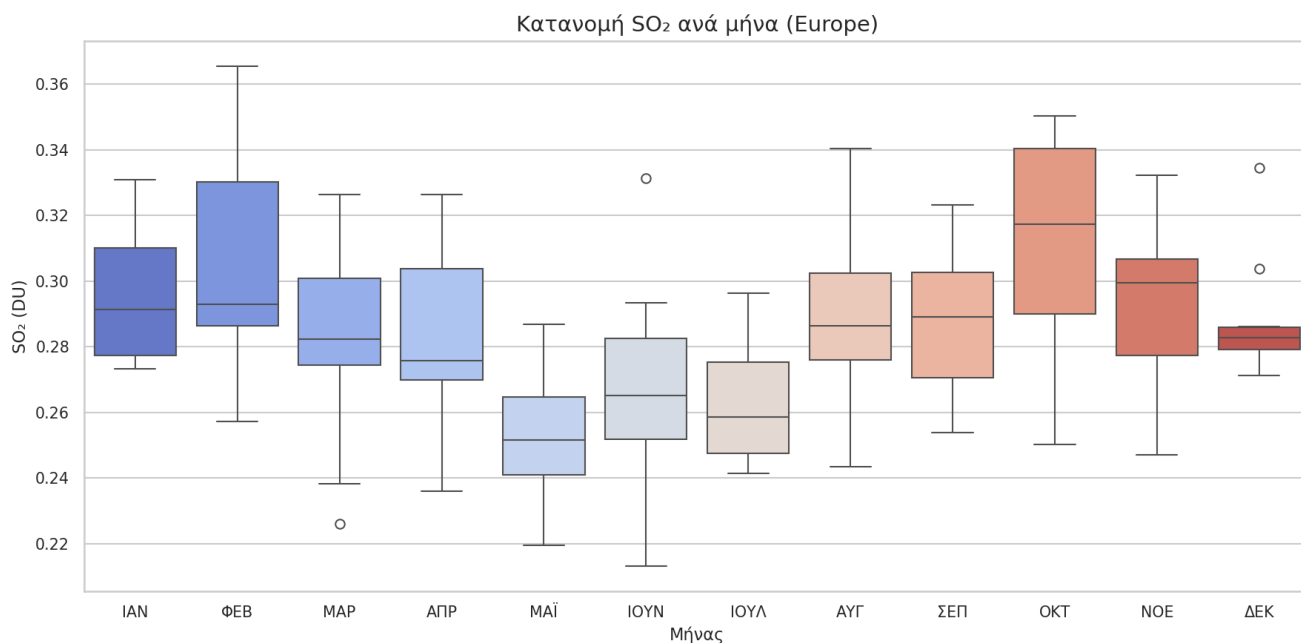
Γράφημα 16: Περιοδογράμμα εποχικής περιόδου Ιωαννίνων.

Η διαφορά μεταξύ των 2 πόλεων είναι εμφανής, όπου στην Αθήνα η μαύρη διακεκομμένη γραμμή (κυρίαρχη περίοδος) τέμνεται απόλυτα με την κόκκινη διακεκομμένη (ετήσια συχνότητα), ενώ στα Ιωάννινα, η μαύρη διακεκομμένη γραμμή έχει τοποθετηθεί στην 6μηνιαία συχνότητα.

4.5.3 Οπτική αξιολόγηση εποχικής μεταβλητότητας

Για την απεικόνιση της εποχικής μεταβλητότητας και τις διαφορές μεταξύ χρονικών περιόδων, επιλέξαμε να χρησιμοποιήσουμε διαγράμματα τύπου Box-Plot, που συνοψίζουν τη μηνιαία κατανομή των συγκεντρώσεων διοξειδίου του θείου, απεικονίζοντας τη διάμεσο, τα τεταρτημόρια, και τις τυχόν ακραίες τιμές. Όπως και κάθε άλλη διαδικασία, πραγματοποιήθηκε για όλες τις οντότητες κάθε χωρικού επιπέδου.

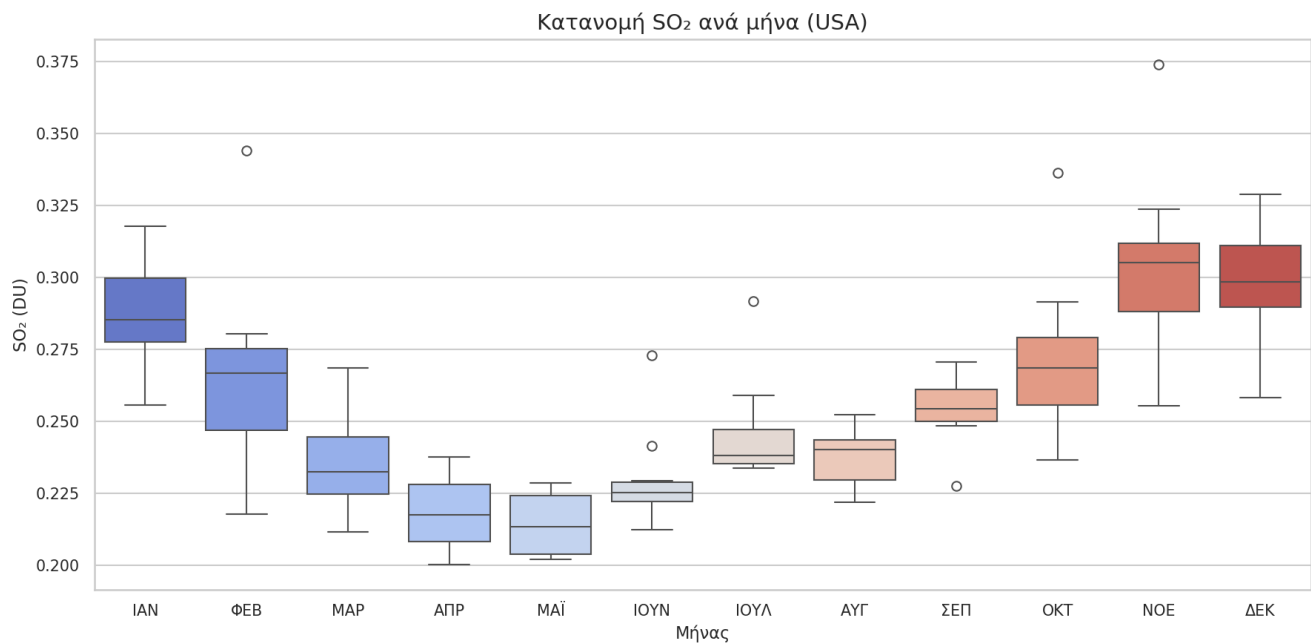
Παρακάτω παραθέτουμε τα σχετικά διαγράμματα για την Ευρώπη και τις ΗΠΑ.



Γράφημα 17: box-and-whisker plot κατανομής εποχικής μεταβλητότητας στην Ευρώπη.

Στην Ευρώπη, οι συγκεντρώσεις διοξειδίου του θείου παρουσιάζουν δύο διακριτές περιόδους αυξημένων τιμών, η μία κατά τους χειμερινούς μήνες (Δεκέμβριος - Φεβρουάριος) και η δεύτερη το φθινόπωρο (Οκτώβριος - Νοέμβριος). Όπως έχουμε ήδη αναφέρει, οι χειμερινοί μήνες χαρακτηρίζονται από αυξημένες εκπομπές ρύπων, λόγω καύσης ορυκτών καυσίμων, καθώς και εξαιτίας του χαμηλότερου ύψους οριακού στρώματος (PBL), στοιχεία που εξηγούν τις υψηλές συγκεντρώσεις διοξειδίου του θείου. Από την άλλη, οι χαμηλότερες τιμές εμφανίζονται την άνοιξη και στις αρχές του καλοκαιριού, κάτι που συνάδει με την εποχική μεταβολή των μετεωρολογικών συνθηκών και τη μειωμένη ενεργειακή δραστηριότητα που σχετίζεται με εκπομπές διοξειδίου του θείου.

Ένα επιπλέον χαρακτηριστικό που προκύπτει από το διάγραμμα, είναι ότι κατά τους θερινούς μήνες (Ιούνιος–Αύγουστος) δε μειώνεται μόνο το επίπεδο των συγκεντρώσεων διοξειδίου του θείου, αλλά και η μεταξύ των πόλεων διακύμανση, καθώς τα boxplots εμφανίζονται στενότερα. Αυτό υποδηλώνει ότι οι Ευρωπαϊκές πόλεις συγκλίνουν σε παρόμοια επίπεδα διοξειδίου του θείου κατά την περίοδο αυτή, σε αντίθεση με το χειμώνα και το φθινόπωρο όπου η μεταβλητότητα είναι μεγαλύτερη.



Γράφημα 18: box-and-whisker plot κατανομής εποχικής μεταβλητότητας στις ΗΠΑ.

Στις ΗΠΑ, οι συγκεντρώσεις διοξειδίου του θείου παρουσιάζουν μία σαφή και συμμετρική ετήσια διακύμανση, με τις υψηλότερες τιμές να καταγράφονται κατά τους χειμερινούς μήνες (Νοέμβριος – Φεβρουάριος) και τις χαμηλότερες κατά τη διάρκεια της άνοιξης και των αρχών του καλοκαιριού (Απρίλιος – Ιούλιος). Το μοτίβο αυτό υποδηλώνει ότι η εποχικότητα του διοξειδίου του θείου είναι πιο καθαρή και σταθερή σε σύγκριση με εκείνη της Ευρώπης, χωρίς εμφανείς δευτερεύουσες κορυφώσεις, στοιχείο που συνάδει απόλυτα με τα αποτελέσματα της φασματικής ανάλυσης. Οι συγκεντρώσεις του διοξειδίου του θείου μειώνονται ομαλά από το χειμώνα προς το καλοκαίρι και επανέρχονται το φθινόπωρο αναδεικνύοντας τη σταθερή περιοδική συμπεριφορά τους.

Πρόσθετα, οι συγκεντρώσεις διοξειδίου του θείου τείνουν να συγκλίνουν σε πολύ παρόμοια επίπεδα τους θερινούς μήνες, περισσότερο από ό,τι στην Ευρώπη. Επίσης, η μείωση των συγκεντρώσεων από το χειμώνα προς τις αρχές του καλοκαιριού, είναι σχεδόν μονοτονική και ομαλή, χωρίς απότομες διακυμάνσεις ή τοπικά μέγιστα, στοιχείο που υποδηλώνει μια πιο καθαρή και σταθερή εποχική συμπεριφορά σε σχέση με εκείνη της Ευρώπης.

4.6 Ανάλυση Τάσης

Μετά την ανάλυση των εποχικών μοτίβων, το επόμενο στάδιο της μελέτης μας εστιάζει στην εκτίμηση της μακροχρόνιας τάσης των συγκεντρώσεων διοξειδίου του θείου. Σκοπός μας είναι να διερευνηθεί εάν οι αποεποχικοποιημένες χρονοσειρές, δηλαδή οι συνιστώσες της τάσης και των καταλοίπων, παρουσιάζουν στατιστικά σημαντική μονοτονική μεταβολή στο χρόνο, και σε θετική περίπτωση, να ποσοτικοποιηθεί ο ρυθμός της μεταβολής της. Ως μονάδα έκφρασης έχουμε ορίσει DU/year (Dobson Units per year), σύμφωνα και με την πρακτική που ακολουθείται διεθνώς κυρίως για δεδομένα που προέρχονται από δορυφορικές μετρήσεις.

Για κάθε οντότητα, όλων των χωρικών επιπέδων, εφαρμόσαμε τους ακόλουθους στατιστικούς υπολογισμούς, δημιουργώντας έτσι μια βάση για την περαιτέρω ερμηνεία των αποτελεσμάτων της μελέτης μας.

4.6.1 Έλεγχος ύπαρξης τάσης

Προκειμένου να εντοπίσουμε εάν οι χρονοσειρές εμφανίζουν μονοτονική μεταβολή στο χρόνο και σε θετική περίπτωση να εκτιμήσουμε την κατεύθυνσή τους (αυξητική ή μειωτική), χρησιμοποιήσαμε το μη παραμετρικό έλεγχο Mann-Kendall.

Ο έλεγχος Mann-Kendall [72], [73], [74] αποτελεί μία από τις ευρέως χρησιμοποιούμενες στατιστικές μεθόδους για τον εντοπισμό μονοτονικής τάσης σε χρονοσειρές περιβαλλοντικών δεδομένων. Ο έλεγχος αυτός δεν

προϋποθέτει συγκεκριμένη κατανομή των δεδομένων (όπως κανονικότητα) και είναι ιδιαίτερα ανθεκτικός στην παρουσία ακραίων τιμών ή ακόμη και σε περιπτώσεις όπου απουσιάζουν παρατηρήσεις από τη χρονοσειρά.

Ο Richard O. Gilbert (Pacific Northwest Laboratory), στο βιβλίο του 'Statistical Methods for Environmental Pollution Monitoring' [75] περιγράφει αναλυτικά τα βήματα που ακολουθεί η μέθοδος Mann-Kendall test. Επιγραμματικά,

Για πλήθος μέχρι 40 παρατηρήσεις, σε κάθε ζεύγος $(x_j - x_k)$ με $j > k$, υπολογίζεται το πρόσημο της διαφοράς $(x_j - x_k)$. Ο δείκτης S ορίζεται ως το άθροισμα όλων αυτών των προσήμων:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k)$$

όπου,

$$\text{sign}(x_j - x_k) = \begin{cases} 1, & \text{αν } x_j - x_k > 0 \\ 0, & \text{αν } x_j - x_k = 0 \\ -1, & \text{αν } x_j - x_k < 0 \end{cases}$$

Σε περίπτωση μη ύπαρξης τάσης, οι θετικές και αρνητικές διαφορές αλληλοαναιρούνται, οπότε η τιμή του S κυμαίνεται γύρω από το μηδέν. Αντίθετα, μια θετική τιμή του S υποδεικνύει αυξητική τάση, ενώ μια αρνητική τιμή δηλώνει μειωτική τάση.

Για δείγματα με μεγάλο πλήθος παρατηρήσεων ($n > 40$, σ.σ. αν και ο Kendall υποστηρίζει πως η κάτωθι μέθοδος μπορεί να χρησιμοποιηθεί και για μικρότερο πλήθος παρατηρήσεων όπως 10), χρησιμοποιείται ο παραπάνω δείκτης S, ώστε να υπολογιστεί ο στατιστικός δείκτης Z ως ακολούθως:

$$Z = \begin{cases} \frac{S - 1}{[\text{Var}(S)]^{1/2}}, & \text{αν } S > 0 \\ 0, & \text{αν } S = 0 \\ \frac{S + 1}{[\text{Var}(S)]^{1/2}}, & \text{αν } S < 0 \end{cases}$$

όπου,

$$\text{VAR}(S) = \frac{1}{18} [n(n-1)(2n+5) - \sum_{p=1}^g t_p(t_p-1)(2t_p+5)]$$

με n , το συνολικό πλήθος των παρατηρήσεων (πχ. αριθμός ετών ή μηνών),
 g , ο αριθμός των ομάδων ισοδύναμων τιμών (δηλαδή 'ties') και
 t_p , το πλήθος των ισοδύναμων τιμών στην p -οστή ομάδα.

Μία θετική τιμή του Z υποδηλώνει ανοδική τάση, ενώ μία αρνητική υποδηλώνει καθοδική. Εάν η μηδενική υπόθεση H_0 , που εκφράζει την απουσία τάσης, είναι αληθής, τότε ο στατιστικός δείκτης Z ακολουθεί την τυπική κανονική κατανομή.

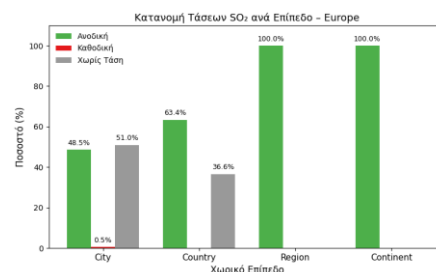
Για έλεγχο ύπαρξης είτε ανοδικής είτε καθοδικής τάσης (διμερής έλεγχος) στο επίπεδο σημαντικότητας α , η H_0 απορρίπτεται όταν η απόλυτη τιμή του Z είναι μεγαλύτερη από το $Z_{1-\alpha/2}$

Αν η εναλλακτική υπόθεση αφορά μόνο ανοδική τάση (μονόπλευρος έλεγχος), η H_0 απορρίπτεται όταν η τιμή του Z είναι μεγαλύτερη από το $Z_{1-\alpha}$. Αντίθετα, απορρίπτουμε τη H_0 υπέρ της εναλλακτικής υπόθεσης ύπαρξης καθοδικής τάσης, όταν το Z είναι αρνητικό και η απόλυτη τιμή του είναι μεγαλύτερη από το $Z_{1-\alpha/2}$.

Ακολουθώντας τα παραπάνω, εφαρμόσαμε το Mann-Kendall test σε όλες τις οντότητες, τηρώντας απαρέγκλιτα ότι τιμές $p\text{-value} < 0.05$ θεωρούνται ως στατιστικά σημαντικές, ενώ μεγαλύτερες τιμές $p\text{-value}$ υποδηλώνουν την απουσία σημαντικής τάσης, οπότε κατηγοριοποιούνται ως 'No Trend'. Η κατεύθυνση της τάσης (Increasing ή

Decreasing) καθορίστηκε αποκλειστικά στις στατιστικά σημαντικές περιπτώσεις. Παρακάτω παραθέτουμε τα αποτελέσματα από αυτή τη διαδικασία:

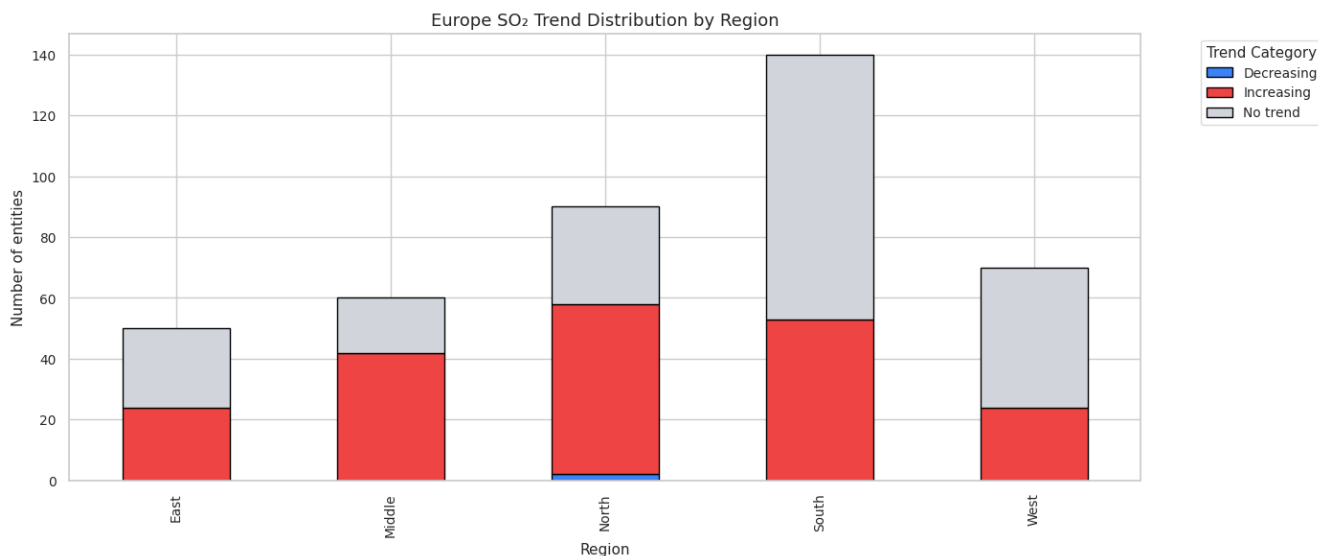
Europe		Increasing		Decreasing		No trend	
Level	n	count	%	count	%	count	%
city	410	199	48,54	2	0,49	209	50,98
country	41	26	63,41	0	0	15	36,59
region	5	5	100	0	0	0	0
continent	1	1	100	0	0	0	0



Γράφημα 19: Κατανομή τάσης ανά χωρικό επίπεδο στην Ευρώπη.

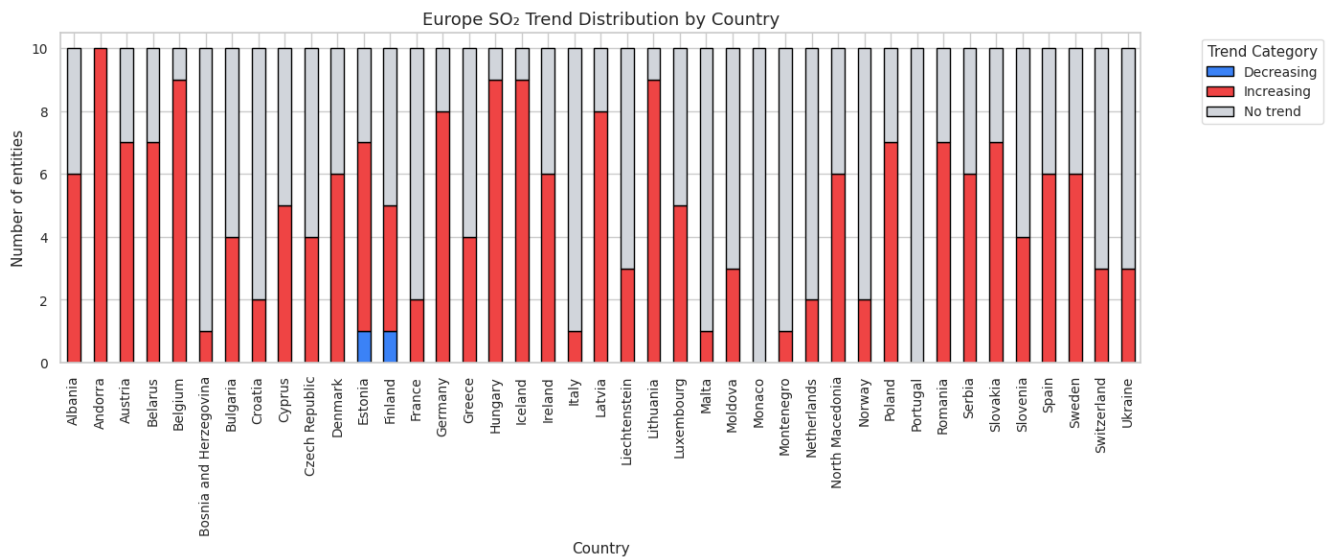
Βάσει των ανωτέρω, για την Ευρώπη φαίνεται ότι σε επίπεδο πόλεων, περίπου οι μισές παρουσιάζουν στατιστικά σημαντική αυξητική τάση στις συγκεντρώσεις του διοξειδίου του θείου, ενώ μόλις το 0,5% εμφανίζει σημαντική μειωτική τάση (η επιλογή του κόκκινου χρώματος στο σχετικό bar plot για την καθοδική τάση έγινε επί σκοπού, ώστε να είναι εμφανής η μειονότητα από το όλον). Το υπόλοιπο 51% των πόλεων δεν παρουσιάζει στατιστικά αξιόπιστη μεταβολή (No trend), γεγονός που μάλλον υποδηλώνει σταθεροποίηση των επιπέδων του διοξειδίου του θείου τα τελευταία χρόνια. Σε επίπεδο κρατών, το ποσοστό των αυξητικών τάσεων είναι ακόμη υψηλότερο (63,4%), ενώ κανένα κράτος δεν εμφανίζει στατιστικά σημαντική μείωση. Στο επίπεδο περιφερειών και ηπείρου, όλες οι χρονοσειρές καταγράφουν στατιστικά σημαντική αύξηση, αντικατοπτρίζοντας τη γενική μακροχρόνια τάση στην Ευρώπη, όπου εξακολουθεί να είναι ανοδική και αντανακλώντας πιθανώς είτε ανεπαρκή μέτρα μείωσης των εκπομπών είτε μετεωρολογικές συνθήκες που δεν ευνοούν την απομάκρυνσή του διοξειδίου του θείου.

Για να δούμε την κατανομή των πόλεων σε ανώτερα χωρικά επίπεδα, καταρτίσαμε τα κάτωθι διάγραμμα:



Γράφημα 20: Ταξινόμηση πόλεων ανά περιφέρεια βάσει της τάσης στην Ευρώπη.

Από το παραπάνω παρατηρούμε ότι σε όλες τις περιφέρειες της Ευρώπης οι αυξητικές τάσεις υπερτερούν σαφώς έναντι των καθοδικών. Η νότια Ευρώπη συγκεντρώνει το μεγαλύτερο πλήθος πόλεων με θετικές ή σταθερές τάσεις, ενώ οι πόλεις της βόρειας και της κεντρικής περιφέρειας εμφανίζουν μία πιο ισορροπημένη αλλά και πάλι κυρίως αυξητική συμπεριφορά. Συνολικά, το διάγραμμα καταδεικνύει ότι, αν και η πλειονότητα των περιοχών παραμένει χωρίς στατιστικά σημαντικές μεταβολές, οι τοπικές αυξήσεις των συγκεντρώσεων διοξειδίου του θείου είναι διάχυτες σε όλη την ήπειρο, με εντονότερη παρουσία στην νότια και κεντρική περιφέρεια.

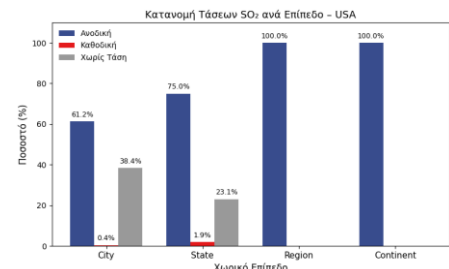


Γράφημα 21: Ταξινόμηση κρατών Ευρώπης βάσει την εκτίμηση της πορείας της μακροχρόνιας τάσης.

Και σε αυτή την προβολή, οι αυξητικές τάσεις κατακλύζουν το διάγραμμα, με τις περισσότερες χώρες να παρουσιάζουν τουλάχιστον στο 50% των πόλεων που αντιπροσωπεύουν, ανοδική συμπεριφορά τάσης, ενώ καθοδική τάση εμφανίζουν μόνο πόλεις της Εσθονίας και της Φινλανδίας.

Τα παραπάνω διαγράμματα μας επιτρέπουν να δούμε πως τα αστικά ή βιομηχανικά κέντρα επηρεάζουν τα ανώτερα χωρικά επίπεδα και συνολικά την Ευρώπη. Αυτό θα γινόταν ακόμη πιο εμφαντικό αν διαθέταμε τα τεχνικά μέσα, να μελετήσουμε τις συγκεντρώσεις διοξειδίου του θείου όχι μόνο 10 πόλεων ανά χώρα, αλλά στο σύνολο των πόλεων.

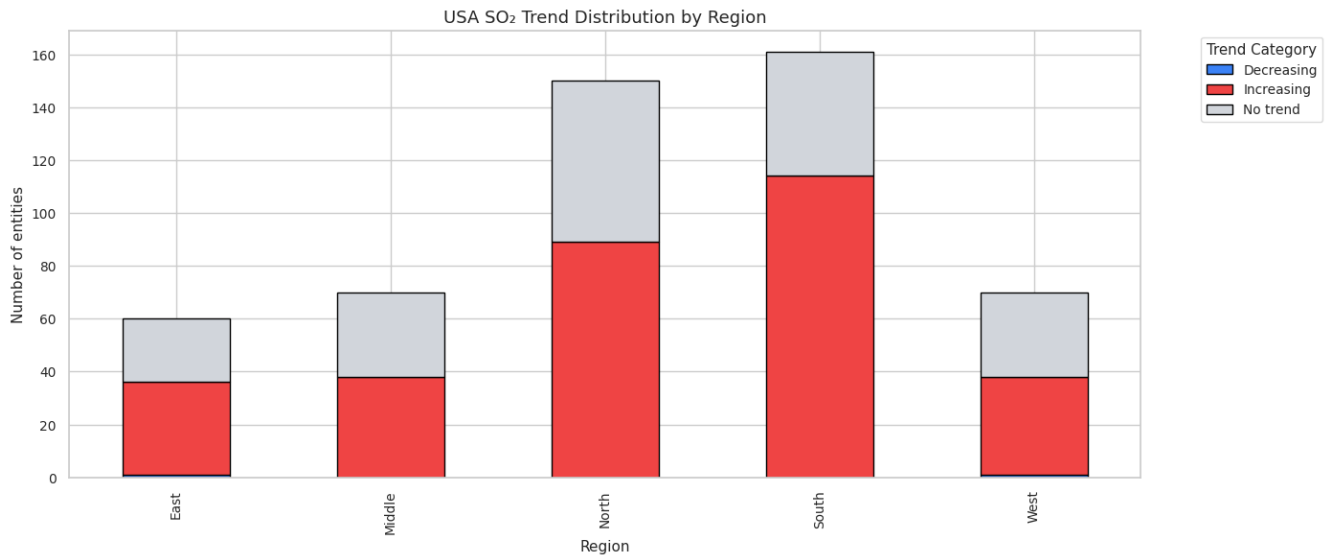
USA		Increasing		Decreasing		No trend	
Level	n	count	%	count	%	count	%
city	511	313	61,25	2	0,39	196	38,36
state	52	39	75	1	1,92	12	23,08
region	5	5	100	0	0	0	0
continent	1	1	100	0	0	0	0



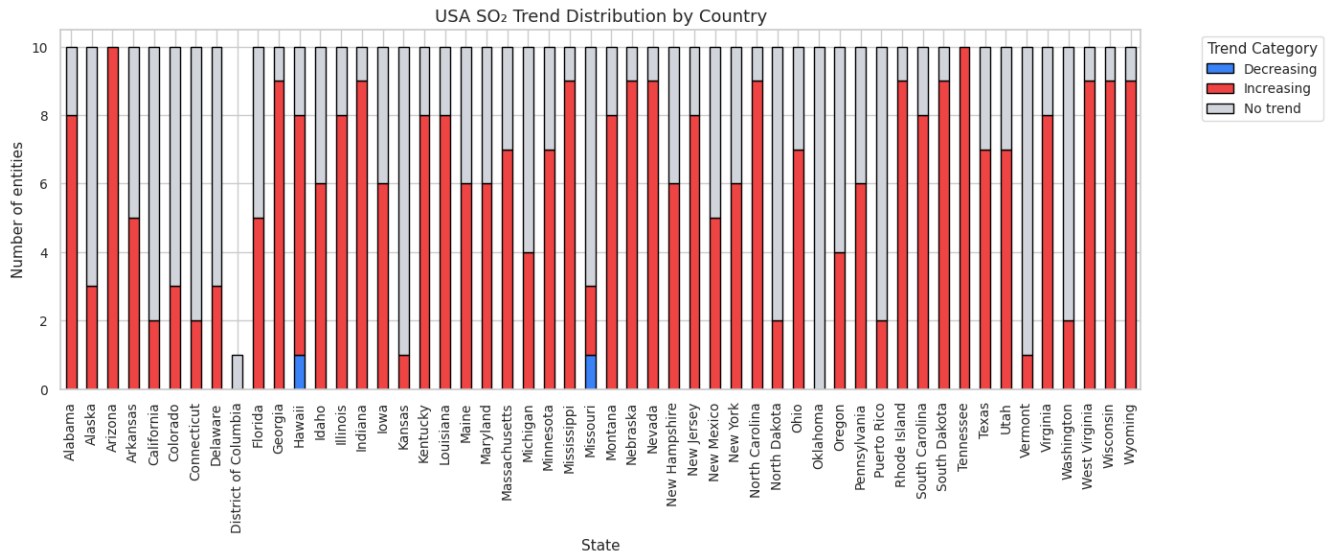
Γράφημα 22: Κατανομή τάσης ανά χωρικό επίπεδο στις ΗΠΑ.

Για τις ΗΠΑ, παρατηρούμε μια ακόμη μεγαλύτερη αυξητική πορεία στις συγκεντρώσεις του διοξειδίου του θείου. Σε επίπεδο πόλεων, το 61,3% παρουσιάζει στατιστικά σημαντική αύξηση, ενώ μόλις το 0,4% εμφανίζει σημαντική μείωση. Το υπόλοιπο περίπου 38% των πόλεων δεν παρουσιάζει στατιστικά αξιόπιστη μεταβολή. Σε επίπεδο πολιτειών, η εικόνα είναι ακόμη πιο έντονη, με το 75% να παρουσιάζει ανοδική τάση και μόλις μία πολιτεία καθοδική. Στις περιφέρειες και σε επίπεδο ηπείρου, οι τάσεις είναι αποκλειστικά αυξητικές, όπως συμβαίνει και στην Ευρώπη. Τα ευρήματα αυτά δείχνουν ότι, σε αντίθεση με την ιστορική μείωση των εκπομπών SO₂ που είχε καταγραφεί τις προηγούμενες δεκαετίες (EPA - Air Pollutant Emissions Trends Data - National Tier 1 CAPS Trends [76]), τα τελευταία χρόνια παρατηρείται μία ήπια αύξηση των συγκεντρώσεων διοξειδίου του θείου, πιθανώς συνδεδεμένη με την αύξηση της ενεργειακής ζήτησης και την εκτεταμένη χρήση ορυκτών καυσίμων.

Προβάλλοντας τις πόλεις των ΗΠΑ σε ανώτερα χωρικά επίπεδα,



Γράφημα 23: Ταξινόμηση πόλεων ανά περιφέρεια βάσει της τάσης στις ΗΠΑ.



Γράφημα 24: Ταξινόμηση πολιτειών των ΗΠΑ, βάσει την εκτίμηση της πορείας της μακροχρόνιας τάσης.

παρατηρούμε μια πιο έντονη αυξητική συμπεριφορά σε σχέση με την Ευρώπη, όπου και εδώ οι τοπικές διαφοροποιήσεις επηρεάζουν συνολικά τις πολιτείες των ΗΠΑ.

4.6.2 Εκτίμηση ρυθμού μεταβολής της τάσης

Για να εκτιμήσουμε το ρυθμό μεταβολής της μακροχρόνιας τάσης των συγκεντρώσεων διοξειδίου του θείου, χρησιμοποιήσαμε τη μη παραμετρική μέθοδο Theil-Sen, όπως αρχικά προτάθηκε από τον Henri Theil το 1950 [77] και επεκτάθηκε αργότερα από τον Pranab Kumar Sen το 1968 [78]. Η μέθοδος αυτή, που περιγράφεται αναλυτικά στο έργο του Gilbert ‘Statistical Methods for Environmental Pollution Monitoring’ [75], αποτελεί μία από τις πιο ευρέως αποδεκτές προσεγγίσεις στην ανάλυση τάσεων περιβαλλοντικών δεδομένων και είναι στενά συνδεδεμένη με τον έλεγχο Mann-Kendall, καθώς βασίζεται στα ίδια διατεταγμένα ζεύγη παρατηρήσεων.

Πιο συγκεκριμένα,

για μια σειρά X_i με n χρονικές παρατηρήσεις (πχ. ετήσιες ή μηνιαίες), υπολογίζονται όλες οι επιμέρους εκτιμήσεις των κλίσεων Q μεταξύ κάθε ζεύγους παρατηρήσεων (i, i') με $i' > i$:

$$Q_{i'i} = \frac{X_{i'} - X_i}{i' - i}$$

Ο συνολικός αριθμός αυτών των εκτιμήσεων είναι $N' = n(n - 1)/2$ εφόσον υπάρχει μία τιμή ανά χρονική περίοδο, ενώ σε περίπτωση ύπαρξης πολλαπλών μετρήσεων ανά περίοδο, $N' < n(n - 1)/2$, καθώς δε μπορούν να υπολογιστούν διαφορές για $i' = i$.

Η διάμεσος όλων των N' τιμών $Q_{i'i}$ αποτελεί την εκτίμηση της πραγματικής κλίσης:

$$\text{Sen's slope} = \begin{cases} Q_{[(N'+1)/2]}, & \text{αν } N' \text{ είναι περιττός} \\ \frac{Q_{[N'/2]} + Q_{[(N'+2)/2]}}{2}, & \text{αν } N' \text{ είναι άρτιος} \end{cases}$$

όπου, $Q_{[k]}$ υποδηλώνει την k -οστή κατά αύξουσα σειρά τιμή των επιμέρους εκτιμήσεων των κλίσεων.

Τα βήματα αυτής της διαδικασίας είναι:

-Επιλογή του επιπέδου εμπιστοσύνης α και εύρεση της τιμής $Z_{1-\alpha/2}$.

-Υπολογισμός $C_\alpha = Z_{1-\alpha/2} [\text{VAR}(S)]^{1/2}$, όπου $\text{VAR}(S)$ (βλ. Mann-Kendall).

-Υπολογισμός $M_1 = (N' - C_\alpha)/2$ και $M_2 = (N' + C_\alpha)/2$.

Τα όρια του διαστήματος εμπιστοσύνης είναι οι M_1^{th} και $(M_2 + 1)^{\text{th}}$ μεγαλύτερες τιμές της ταξινομημένης λίστας των $Q_{i'i}$. Δηλαδή, το διάστημα εμπιστοσύνης της Sen's slope προκύπτει απευθείας από τις ταξινομημένες επιμέρους κλίσεις, χωρίς καμία γραμμική προσέγγιση.

Πριν παραθέσουμε τα αποτελέσματα από εκτίμηση των κλίσεων των τάσεων, σημειώνουμε πως για τον υπολογισμό της κλίσης Theil-Sen σε Dobson Units ανά έτος (DU/year), μετατρέψαμε το δείκτη του χρόνου της χρονοσειράς σε συνεχή μεταβλητή με ακρίβεια μηνών, ως

$$t = \text{year} + \frac{\text{month} - 1}{12}$$

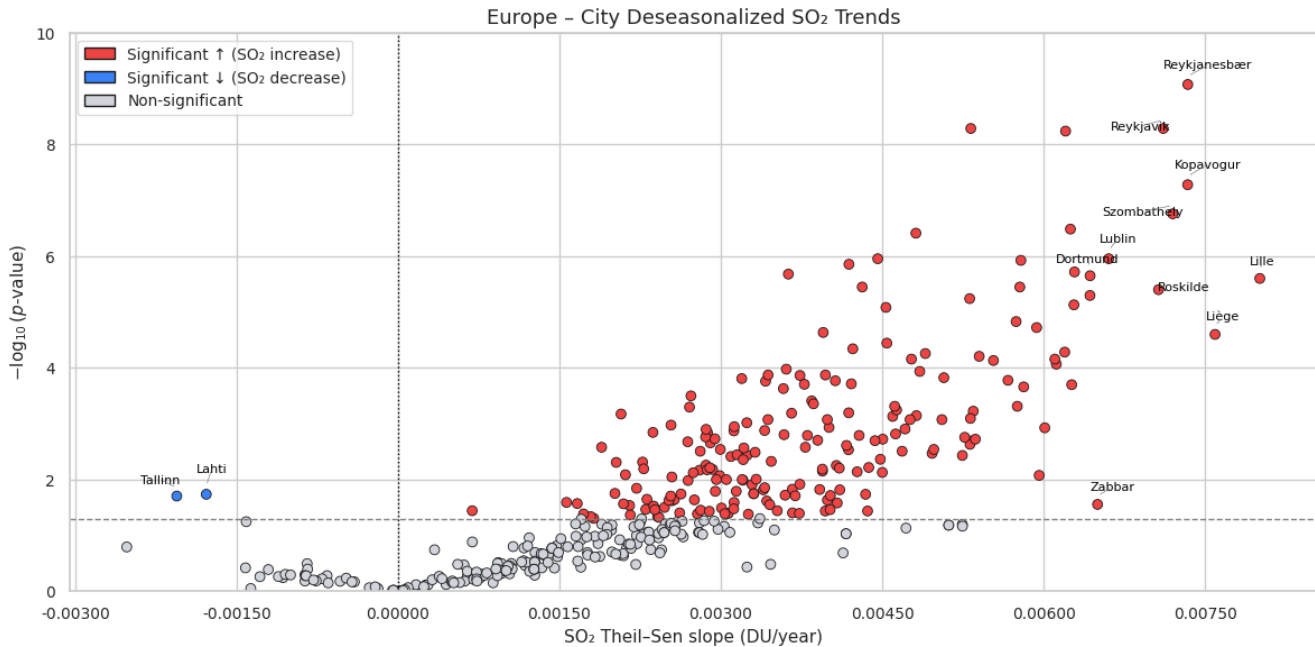
όπου, η αφαίρεση του ελάχιστου έτους t_{\min} (δηλαδή $X = t - t_{\min}$) εξασφαλίζει πως η παλινδρόμηση ξεκινάει από μηδενικό χρονικό σημείο. Έτσι, η κλίση που επιστρέφει η συνάρτηση `theilslopes` από τη βιβλιοθήκη `scipy.stats` [79] εκφράζει τη μέση μεταβολή της συγκέντρωσης διοξειδίου του θείου ανά έτος, δηλαδή το ρυθμό μεταβολής της τάσης σε DU/year.

Παρακάτω παραθέτουμε τα αποτελέσματα υπολογισμού των κλίσεων των τάσεων:

Europe		Increasing				Decreasing			
Level	n	count	%	median slope	IQR	count	%	median slope	IQR
city	410	199	48,54	0,003692	0,001798	2	0,49	-0,001919	0,000137
country	41	26	63,41	0,003206	0,000921	0	0	0	0
region	5	5	100	0,002658	0,000117	0	0	0	0
continent	1	1	100	0,002446	0	0	0	0	0

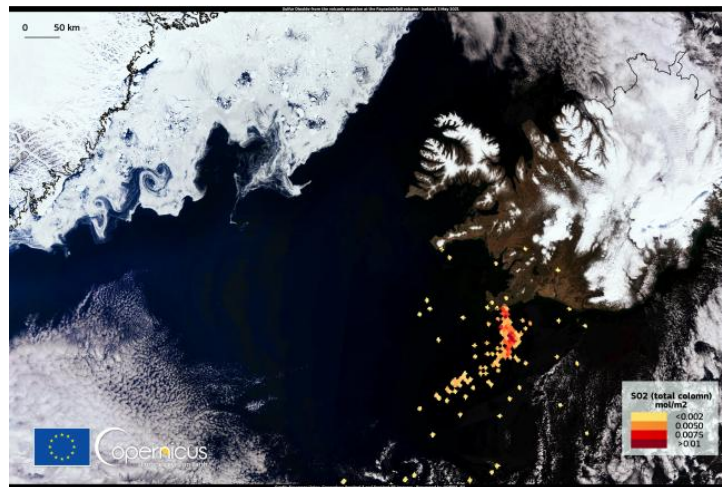
Η ανάλυση των κλίσεων Theil-Sen για την Ευρώπη, σε επίπεδο πόλεων, δείχνει μία ανοδική πορεία της τάσης κατά μέσο όρο 0.0037 DU/year, ενώ για 2 πόλεις που εμφανίζουν καθοδική πορεία, η διάμεση κλίση είναι μόλις 0.002. Αυτά, σε συνδυασμό με το ενδοετησιακό εύρος (0.0018 και 0.00014 αντίστοιχα) υποδηλώνουν μία σχετικά ομοιογενή αυξητική συμπεριφορά, αν και δε θα πρέπει να παραβλέψουμε ότι το 50% των πόλεων δεν εμφανίζει στατιστικά σημαντική ανοδική ή καθοδική τάση. Σε επίπεδο Ευρωπαϊκών κρατών παρατηρούμε μία παρόμοια συμπεριφορά με σχετικά περισσότερες χώρες να εμφανίζουν αυξητική τάση, σε σχέση με εκείνες που δεν εμφανίζουν στατιστικά σημαντική τάση. Σε περιφερειακό επίπεδο, φαίνεται μία ομαλή αύξηση με διάμεση κλίση 0.0027 DU/year, κάτι που αποτυπώνεται και στο ανώτατο χωρικό επίπεδο.

Η οπτική κατανομή των πόλεων ανάλογα με την κλίση τους παρατίθεται παρακάτω,



Γράφημα 25: Κατανομή Ευρωπαϊκών πόλεων ανάλογα τη σημαντικότητα της κλίσης της τάσης.

όπου μας επιτρέπει να εντοπίσουμε τις πόλεις εκείνες που επηρεάζουν σημαντικά θετικά ή αρνητικά τη γενική συμπεριφορά της τάσης σε Ευρωπαϊκό επίπεδο. Για παράδειγμα, η πόλη Reykjanesbaer της Ισλανδίας που φαίνεται επάνω δεξιά με την πιο σημαντική θετική κλίση, βρίσκεται πλησίον του ηφαιστείου Fagradalsfjall (περίπου 18 km), το οποίο ενεργοποιήθηκε το 2021, με αποτέλεσμα να παρατηρούνται αυξημένες εκπομπές διοξειδίου του θείου, όχι μόνο τοπικά αλλά και στην ευρύτερη περιοχή επηρεάζοντας μάλιστα και την Ευρώπη (Copernicus Atmosphere Monitoring Service - CAMS)[80].



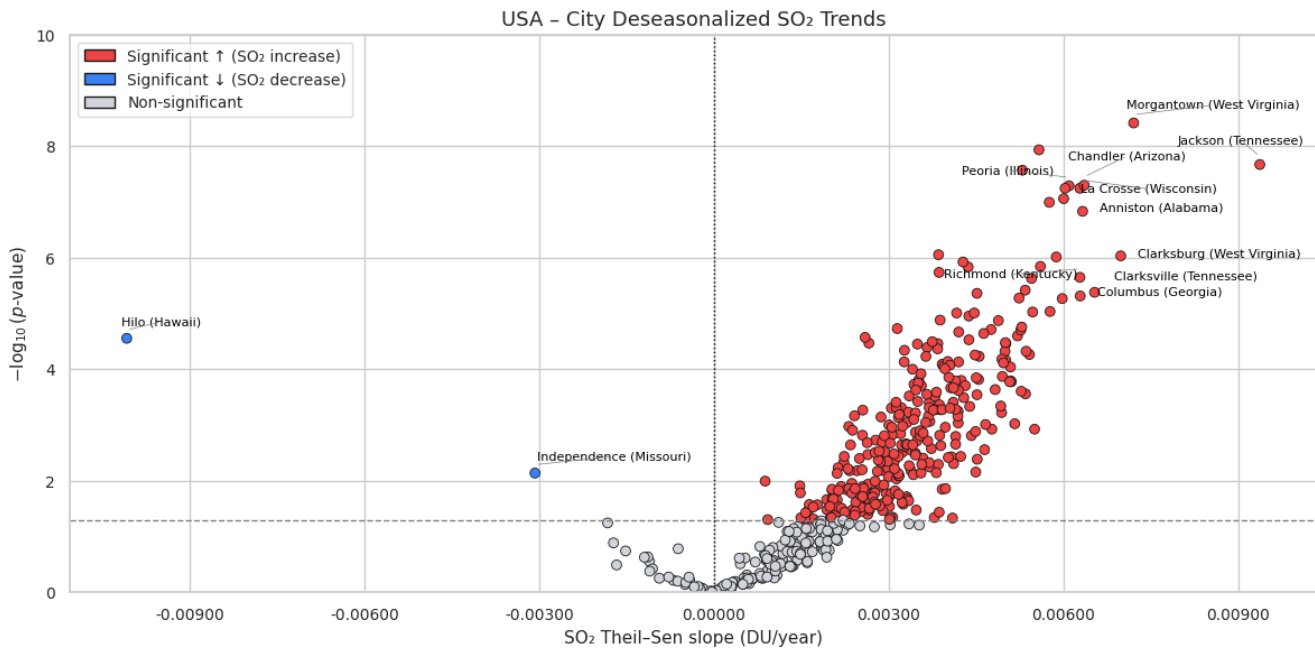
Εικόνα 12: Διασπορά SO₂ από τη δραστηριότητα του ηφαιστείου Fagradalsfjall το Μάιο του 2021

Στο Dortmund, η βιομηχανική δραστηριότητα, ιδιαίτερα στην περιοχή της Ρουρ που περιλαμβάνει παραδοσιακές βιομηχανίες όπως η χαλυβουργία, μπορεί να συμβάλει στην αύξηση των εκπομπών διοξειδίου του θείου, σε βαθμό που να την κατατάσσει στις 10 πόλεις (βάσει των τυχαία επιλεγμένων πόλεων του δείγματός μας) με την πιο σημαντική κλίση τάσης.

USA		Increasing				Decreasing			
Level	n	count	%	median slope	IQR	count	%	median slope	IQR
city	511	313	61,25	0,003386	0,001416	2	0,39	-0,00658	0,003506
state	52	39	75	0,002896	0,000896	1	1,92	-0,00228	0
region	5	5	100	0,00259	0,000456	0	0	0	0
continent	1	1	100	0,00245	0	0	0	0	0

Στην αντίπερα όχθη, αν και οι κλίσεις των τάσεων κυμαίνονται στα ίδια επίπεδα με εκείνα της Ευρώπης, παρατηρούμε ότι η ανοδική πορεία των συγκεντρώσεων του διοξειδίου του θείου καταγράφεται σε μεγαλύτερο πλήθος πόλεων και ακόμη περισσότερο σε επίπεδο πολιτειών, εμφανίζοντας διαφορά σε σχέση με την Ευρώπη πάνω από 10 ποσοστιαίες μονάδες.

Από την απεικόνιση των κλίσεων των τάσεων συνολικά για τις πόλεις των ΗΠΑ,



Γράφημα 26: Κατανομή πόλεων των ΗΠΑ, ανάλογα τη σημαντικότητα της κλίσης της τάσης.

θα σταθούμε στην περίπτωση της πόλης Hilo της Χαβάης, που παρατηρήσαμε μια ιδιαίτερη εικόνα καθοδικής πορείας της τάσης, μοναδική στο σύνολο του δείγματός μας. Από περαιτέρω διερεύνηση διαπιστώσαμε ότι το αποτέλεσμα αυτό δεν οφείλεται σε ανθρωπογενείς πολιτικές μείωσης ρύπων, αλλά σε φυσική διακύμανση που προέκυψε από την απότομη αλλαγή συμπεριφοράς του ηφαιστείου Kīlauea. Πιο συγκεκριμένα, το Kīlauea, πριν το 2018 είχε πολύ υψηλές εκπομπές διοξειδίου του θείου, με τιμές που ξεπερνούσαν τους 50.000 τόνους την ημέρα. Τον Αύγουστο του 2018, η δραστηριότητα του ηφαιστείου σταμάτησε απότομα και ο κρατήρας Halema'ūma'u κατέρρευσε, με αποτέλεσμα οι εκπομπές να μειωθούν κατακόρυφα σε λιγότερο από 200 τόνους την ημέρα.

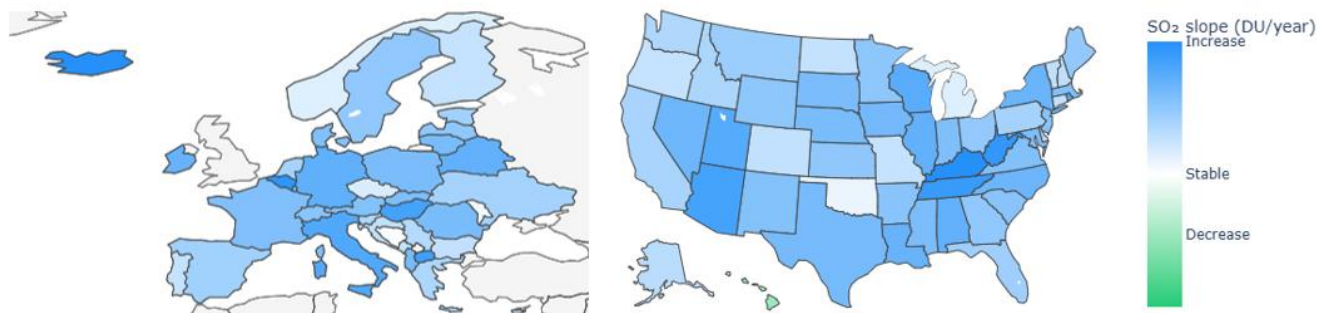


Εικόνα 13: Κατάρρευση κρατήρα Halema'uma'ū του ηφαιστείου Kīlauea (εικόνα πριν την κατάρρευση και μετά, πηγή: <https://www.nps.gov/havo/learn/nature/2018-eruption.htm>)

Έτσι, η χρονοσειρά για την πόλη Hilo δείχνει μια έντονη πτωτική πορεία, γιατί τα προηγούμενα έτη (ακόμη και πριν το δείγμα μας) είχε πολύ υψηλές τιμές συγκεντρώσεων, ενώ τα τελευταία χρόνια οι εκπομπές διοξειδίου του θείου είναι σχεδόν μηδενικές.

Καταλήγοντας, στους κάτωθι χάρτες απεικονίζεται η ένταση των κλίσεων Theil-Sen's στο ανώτατο χωρικό επίπεδο, ηπείρων:

SO₂ Trends (2015–2025)



Εικόνα 14: Χάρτες Ευρώπης και ΗΠΑ, με χρωματική απεικόνιση του ρυθμού μεταβολής της τάσης.

4.6.3 Υπολογισμός αβεβαιότητας του ρυθμού μεταβολής τάσης.

Ακολουθώντας την εκτίμηση της κατεύθυνσης και της στατιστικής σημαντικότητας της τάσης, μέσω του ελέγχου Mann-Kendall, προβήκαμε στην ανάλυση της αβεβαιότητας των εκτιμώμενων κλίσεων Theil-Sen. Όπως έχουμε ήδη αναφέρει τιμές που παρήχθησαν από το Mann-Kendall test με $p < 0.05$ θεωρούνται ως στατιστικά σημαντικές, ενώ μεγαλύτερες τιμές p υποδηλώνουν την απουσία σημαντικής τάσης, οπότε κατηγοριοποιούνται ως 'No trend'. Η κατεύθυνση της τάσης (ανοδική ή καθοδική) καθορίστηκε αποκλειστικά στις περιπτώσεις όπου η στατιστική σημαντικότητα επιβεβαιώθηκε.

Η ανάλυση της αβεβαιότητας δεν αποσκοπεί στην επανεκτίμηση της ύπαρξης τάσης, αλλά στην ποσοτική αξιολόγηση της αξιοπιστίας της εκτιμώμενης κλίσης για τις ήδη στατιστικά σημαντικές περιπτώσεις. Ως επίπεδο σημαντικότητας [81] επιλέξαμε $\alpha = 0.05$ και για κάθε οντότητα υπολογίσαμε το 95 % διάστημα

εμπιστοσύνης (CI) [81] της κλίσης Theil-Sen. Από αυτό προέκυψε το εύρος της αβεβαιότητας και η σχετική αβεβαιότητα (σ.σ. δείκτης ποιότητας εκτίμησης) ως ο λόγος του εύρους προς το απόλυτο μέγεθος της κλίσης,

$$U_{rel} = \frac{|CI_{high} - CI_{low}|}{2 |\beta|}$$

όπου, β είναι η εκτιμώμενη κλίση Theil-Sen που υπολογίσαμε προηγουμένως.

Τα άκρα του διαστήματος εμπιστοσύνης προσδιορίζονται με βάση τη διακύμανση του στατιστικού S που υπολογίσαμε κατά τη διενέργεια του ελέγχου Mann-Kendall. Οπότε, για επίπεδο σημαντικότητας α (0.05), ο συντελεστής C_α που καθορίζει τα όρια εμπιστοσύνης υπολογίζεται ως,

$$C_\alpha = Z_{1-\alpha/2} \sqrt{\text{Var}(S)}$$

και οι αντίστοιχες θέσεις των άκρων του διαστήματος στις ταξινομημένες επιμέρους κλίσεις β_i , δίδονται από τις σχέσεις,

$$M_1 = \frac{N - C_\alpha}{2}, M_2 = \frac{N + C_\alpha}{2}$$

ενώ, τα όρια του 95 % CI είναι,

$$CI_{low} = \beta_{(M_1)}, CI_{high} = \beta_{(M_2)}$$

Με βάση τα παραπάνω, κάθε χρονοσειρά ταξινομήθηκε σε τρεις κατηγορίες αξιοπιστίας, ήτοι,

α) Significant: στατιστικά ισχυρή και αξιόπιστη (το CI δεν περιλαμβάνει το 0 και $U_{rel} \leq 1$),

β) Non-Significant: το CI περιλαμβάνει το 0, δηλαδή υπάρχει περίπτωση να αλλάζει το πρόσημο της τάσης ή ακόμη και να μην υφίσταται τάση και

γ) Unstable: το CI δεν περιλαμβάνει το 0 αλλά η σχετική αβεβαιότητα είναι υψηλή ($U_{rel} > 1$), υποδεικνύοντας ασταθή ή ποσοτικά αβέβαιη εκτίμηση.

Η προσέγγιση αυτή μας επιτρέπει όχι μόνο την αναγνώριση της ύπαρξης στατιστικά σημαντικής τάσης, αλλά και την εκτίμηση της ποσοτικής αξιοπιστίας της κλίσης που προκύπτει από τη μέθοδο Theil-Sen. Πρόσθετα, θα πρέπει να επισημανθούν τα εξής,

α) οι έννοιες Significant και Non-Significant, στο πλαίσιο ανάλυσης της αβεβαιότητας, δε θα πρέπει να συγχέονται με την ύπαρξη ή μη στατιστικά σημαντικής τάσης, όπως αυτή αναλύθηκε στην προηγούμενη ενότητα και

β) η μονάδα '1' που τέθηκε ως κατώφλι στο δείκτη ποιότητας εκτίμησης U_{rel} , χρησιμοποιήθηκε ως τεχνικό κριτήριο, ώστε το μισό πλάτος του διαστήματος εμπιστοσύνης να μην υπερβαίνει το εκτιμώμενο μέγεθος της κλίσης.

Παρακάτω παρατίθενται τα αποτελέσματα αυτής της διαδικασίας.

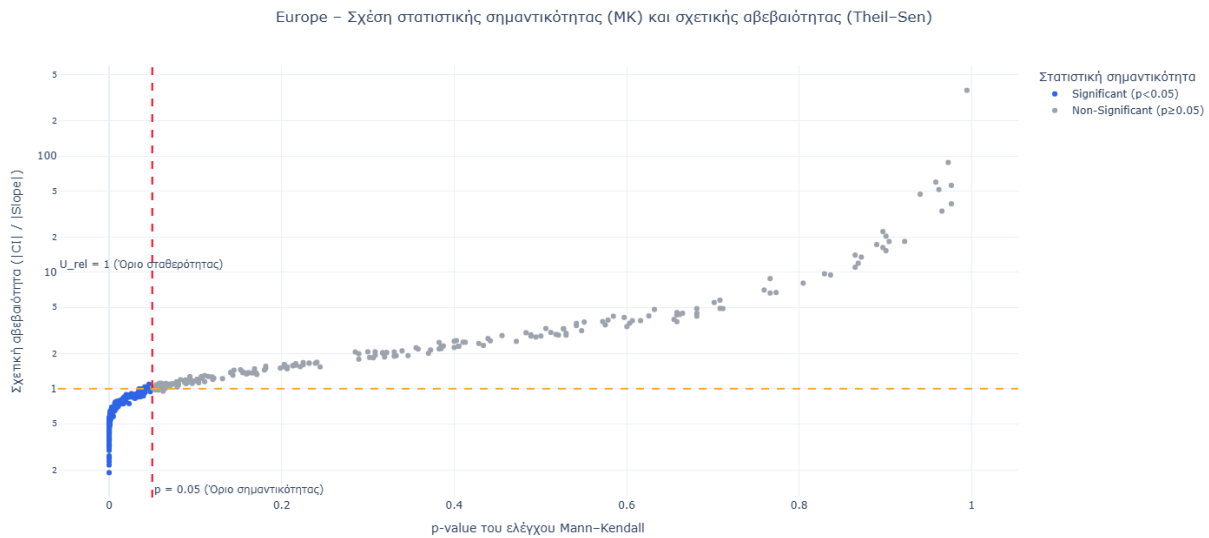
Europe

Level	n	median slope	median uncertainty	mean relative uncertainty	R Non-Significant	R Unstable	R Significant	MK Significant
city	410	0.0025	0.0024	3.609	50.98 %	1.22%	47.80%	49.02%
country	41	0.0028	0.0019	1.596	36.59 %	0	63.41%	63.41%
region	5	0.0027	0.0016	0.633	0	0	100%	100%
continent	1	0.0024	0.0012	0.508	0	0	100%	100%

Η ανάλυση της αβεβαιότητας επιβεβαίωσε πλήρως τα αποτελέσματα που είχαν προκύψει από τον υπολογισμό της στατιστικής σημαντικότητας, μέσω του Mann Kendall test και των κλίσεων Theil-Sen.

Η μόνη διαφοροποίηση αφορά στο επίπεδο των πόλεων, όπου από τις 201 πόλεις (εκ των 410, ήτοι 49,02%) που εμφάνιζαν στατιστικά σημαντική τάση, μόλις 5 κατηγοριοποιήθηκαν ως ‘Unstable’ (1.02% επί του συνόλου), δηλαδή με υψηλή σχετική αβεβαιότητα.

Η σχέση της στατιστικής σημαντικότητας, όπως υπολογίστηκε από την εφαρμογή του Mann-Kendall test και η σχετική αβεβαιότητα των κλίσεων των τάσεων σε επίπεδο πόλεων, απεικονίζονται στο κάτωθι διάγραμμα,



Γράφημα 27: Σχέση στατιστικής σημαντικότητας και σχετικής αβεβαιότητας στην Ευρώπη.

από όπου φαίνεται και οπτικά ότι οι περισσότερες στατιστικά σημαντικές περιπτώσεις ($p < 0.05$) συνοδεύονται από χαμηλή σχετική αβεβαιότητα. Αντίθετα, οι μη στατιστικά σημαντικές περιπτώσεις παρουσιάζουν αυξημένη αβεβαιότητα, γεγονός που, όπως ήταν αναμενόμενο, υποδηλώνει μειωμένη αξιοπιστία στην εκτίμηση της τάσης.

USA

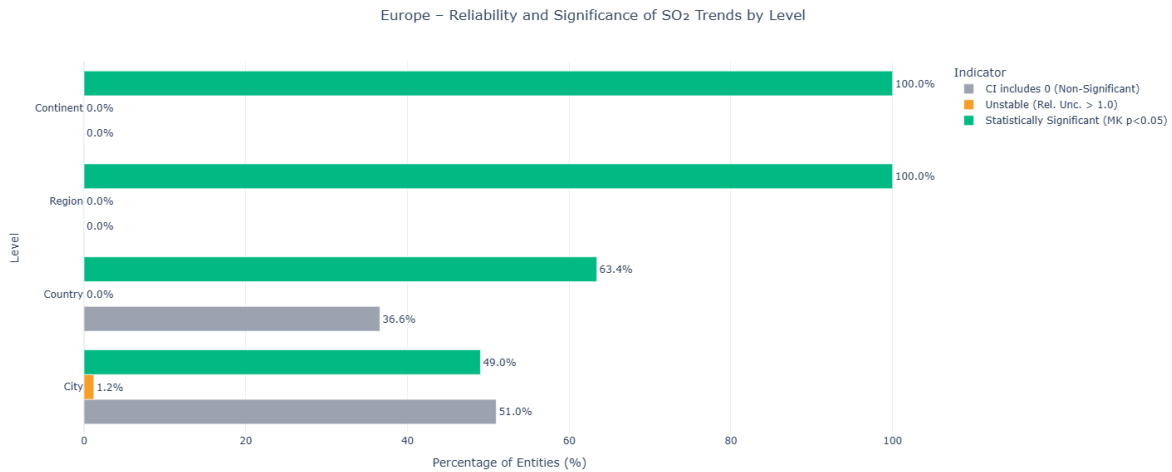
Level	n	median slope	median uncertainty	mean relative uncertainty	R Non-Significant	R Unstable	R Significant	MK Significant
city	511	0.0026	0.0021	2.626	38.36 %	0.98%	60.67%	61.64%
state	52	0.0026	0.0014	0.800	23.08 %	1.92%	75%	76.92%
region	5	0.0026	0.0010	0.399	0	0	100%	100%
continent	1	0.0024	0.0008	0.310	0	0	100%	100%

Και στην περίπτωση των ΗΠΑ τα αποτελέσματα της διαδικασίας εκτίμησης της αβεβαιότητας παρουσίασαν πλήρη συνέπεια με εκείνα του ελέγχου σημαντικότητας Mann-Kendall και των εκτιμώμενων κλίσεων Theil-Sen, επιβεβαιώνοντας και εδώ τη στατιστική αξιοπιστία των τάσεων που υπολογίστηκαν.

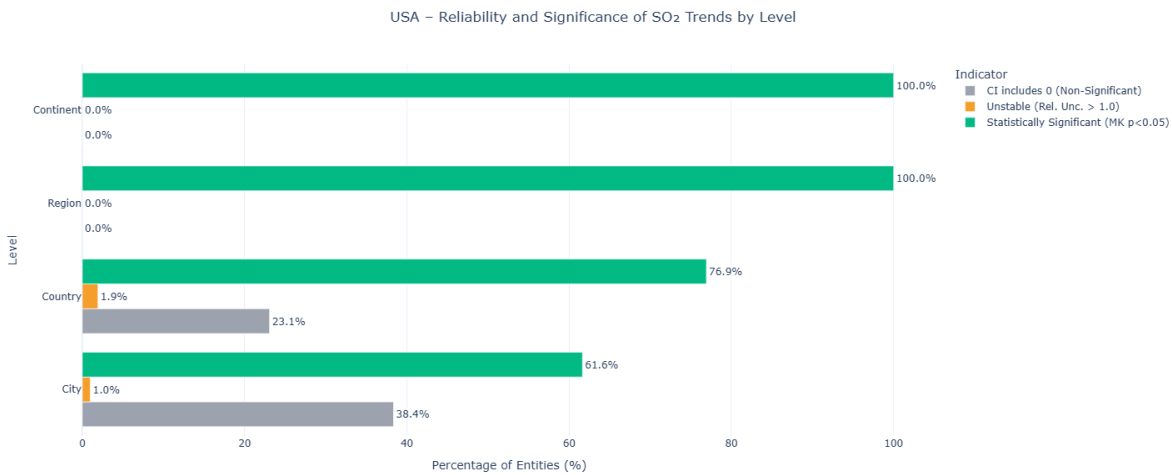
Η μέση σχετική αβεβαιότητα (2.63) είναι σημαντικά χαμηλότερη από εκείνη της Ευρώπης (3.61), γεγονός που υποδηλώνει μεγαλύτερη σταθερότητα των τάσεων στις πόλεις των ΗΠΑ.

Σε επίπεδο πολιτειών, η διάμεση αβεβαιότητα περιορίζεται σε 0.0014 DU/year έναντι της αντίστοιχης τιμής της Ευρώπης (0.0019 DU/year), ενώ η μέση σχετική αβεβαιότητα μειώνεται περαιτέρω σε 0.8, δηλαδή κάτω από το όριο σταθερότητας που θέσαμε (όριο = 1). Αυτό σημαίνει ότι η πλειονότητα των πολιτειών (75 %) παρουσιάζει στατιστικά και ποσοτικά αξιόπιστες αυξητικές τάσεις, σε αντίθεση με την Ευρώπη, όπου το αντίστοιχο ποσοστό είναι σημαντικά μικρότερο (63 %).

Συνοπτικά, τα συγκριτικά αποτελέσματα της διαδικασίας υπολογισμού αβεβαιότητας παρατίθενται στα παρακάτω διαγράμματα:



Γράφημα 28: Απεικόνιση αβεβαιότητας και στατιστικής σημαντικότητας, ανά χωρικό επίπεδο Ευρώπης.



Γράφημα 29: : Απεικόνιση αβεβαιότητας και στατιστικής σημαντικότητας, ανά χωρικό επίπεδο ΗΠΑ.

4.6.4 Στατιστική σύγκριση ρυθμού μεταβολής τάσης μεταξύ περιφερειών

Προκειμένου να εξετάσουμε αν ο ρυθμός μεταβολής του διοξειδίου του θείου διαφέρει σημαντικά μεταξύ γεωγραφικών περιοχών, καταφύγαμε στη χρήση του μη παραμετρικού ελέγχου Kruskal-Wallis [82], ο οποίος αποτελεί γενίκευση του ελέγχου Mann-Whitney σε περισσότερες από δύο ομάδες.

Ο έλεγχος Kruskal-Wallis, αντί να συγκρίνει μέσους όρους, βασίζεται στις τάξεις (ranks) των παρατηρήσεων και ελέγχει αν οι κατανομές των ομάδων προέρχονται από τον ίδιο πληθυσμό.

Αρχικά, όλες οι παρατηρήσεις ταξινομούνται σε μία κοινή σειρά από το 1 έως το n, ανεξάρτητα από την ομάδα στην οποία ανήκουν. Για κάθε ομάδα i, υπολογίζεται το άθροισμα των αντίστοιχων τάξεων R_i, και μαθηματικά ο στατιστικός έλεγχος Kruskal-Wallis υπολογίζεται από τη σχέση:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

όπου,

n_i είναι το πλήθος παρατηρήσεων της ομάδας i, n το συνολικό πλήθος παρατηρήσεων, R_i όπως προαναφέραμε το άθροισμα τάξεων της ομάδας i και k το πλήθος των ομάδων.

Για μεγάλο πλήθος παρατηρήσεων ο στατιστικός έλεγχος Kruskal-Wallis (H) ακολουθεί κατά προσέγγιση την κατανομή χ^2 με $k - 1$ βαθμούς ελευθερίας.

Η μηδενική υπόθεση (H_0) του ελέγχου Kruskal-Wallis είναι ότι όλες οι ομάδες προέρχονται από τον ίδιο πληθυσμό, δηλαδή ότι οι κατανομές των κλίσεων δεν διαφέρουν συστηματικά μεταξύ τους και κατ' επέκταση ότι οι διάμεσοι είναι ίσες.

Το πόσο στατιστικά σημαντικό είναι το αποτέλεσμα ελέγχεται με την τιμή p , όπου, εάν $p < 0.05$, τότε απορρίπτεται η μηδενική υπόθεση H_0 και συνεπώς, τουλάχιστο μία ομάδα έχει διαφορετική κατανομή τιμών από τις υπόλοιπες, είτε επειδή παρουσιάζει διαφορετική διάμεση τιμή είτε διαφορετική θέση κατανομής είτε και τα δύο.

Παράλληλα με τον έλεγχο Kruskal - Wallis, υπολογίστηκε ο δείκτης Kendall's W (Kendall's Coefficient Concordance) [83], [84], ως μη παραμετρικό μέτρο μεγέθους επίδρασης, ώστε να ποσοτικοποιηθεί ο βαθμός ομοιογένειας ή διαφοροποίησης των κατατάξεων των ρυθμών μεταβολής της τάσης μεταξύ των περιφερειών.

Μαθηματικά, ο δείκτης Kendall's W , υπολογίζεται ως εξής:

$$W = \frac{H}{n(k - 1)}$$

όπου, H είναι η στατιστική του ελέγχου Kruskal-Wallis, n είναι το συνολικό πλήθος των παρατηρήσεων και k είναι το πλήθος των ανεξάρτητων ομάδων. Οι τιμές που λαμβάνει κυμαίνονται από 0 έως 1, όπου, τιμές κοντά στο 0 υποδεικνύουν την απουσία κοινού μοτίβου στις ομάδες, ενώ τιμές κοντά στο 1 δείχνουν ότι οι ομάδες έχουν κοινή συμπεριφορά.

Για να εντοπίσουμε ποιες είναι οι ομάδες που τελικά διαφέρουν μεταξύ τους, εφαρμόσαμε τον post-hoc Dunn έλεγχο [85], με διόρθωση Holm-Bonferroni πολλαπλών συγκρίσεων. Με τον τρόπο αυτό, αρχικά υπολογίζεται η διαφορά των μέσων τάξεων κάθε ζεύγους ομάδων. Δηλαδή, όλες οι τιμές των παρατηρήσεων διοξειδίου του θείου όλων των περιφερειών, ταξινομούνται από τη μικρότερη στη μεγαλύτερη και στη συνέχεια υπολογίζεται για κάθε περιφέρεια ο μέσος όρος των θέσεων (τάξεων) που κατέχουν οι τιμές της, εντός αυτής της συνολικής ταξινόμησης.

Οι μέσες τάξεις (mean ranks) συγκρίνονται μεταξύ όλων των ζευγών περιφερειών και για κάθε ζεύγος (i, j) υπολογίζεται η στατιστική τιμή z , η οποία δίδεται από τη σχέση:

$$z_{ij} = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\frac{n(n+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

όπου n είναι το συνολικό πλήθος των παρατηρήσεων, \bar{R}_i και \bar{R}_j οι μέσες τάξεις των ομάδων i και j , και n_i, n_j τα μεγέθη τους.

Ουσιαστικά, η διαφορά των μέσων τάξεων κανονικοποιείται ως προς την τυπική απόκλισή τους, ώστε να εκτιμηθεί πόσο μεγάλη είναι σε σχέση με αυτό που θα αναμενόταν τυχαία. Αν η απόλυτη τιμή του z είναι αρκετά μεγάλη, τότε θεωρείται ότι οι δύο ομάδες διαφέρουν στατιστικά σημαντικά μεταξύ τους (σ.σ. θέτουμε ως επίπεδο σημαντικότητας $\alpha = 0.05$).

Με τον τρόπο αυτό προκύπτει ένας πίνακας που παρουσιάζει για κάθε ζεύγος περιφερειών το αντίστοιχο p -value, από τον οποίο εντοπίζονται εκείνες οι περιφέρειες που εμφανίζουν διαφορετικές στατιστικά κλίσεις των τάσεων των συγκεντρώσεων διοξειδίου του θείου.

Επειδή ο post-hoc Dunn έλεγχος πραγματοποιεί ταυτόχρονα πολλές συγκρίσεις μεταξύ όλων των πιθανών ζευγών περιφερειών, αυξάνεται η πιθανότητα να εντοπιστούν διαφορές που στην πραγματικότητα οφείλονται στην τύχη (σφάλμα τύπου I). Για να περιοριστεί αυτή η πιθανότητα και να διατηρηθεί το συνολικό επίπεδο σημαντικότητας στο 5% ($\alpha = 0.05$), εφαρμόστηκε η διόρθωση Holm-Bonferroni [86], η οποία προβαίνει σε σταδιακή προσαρμογή των τιμών σημαντικότητας, εκκινώντας από τη μικρότερη τιμή p (δηλαδή την πιο σημαντική διαφορά) και

προχωράει προς τις μεγαλύτερες. Έτσι, λαμβάνει υπόψη τόσο το πλήθος όσο και τη σειρά των επιμέρους ελέγχων, μειώνοντας τον κίνδυνο ψευδώς θετικών αποτελεσμάτων, χωρίς όμως να γίνεται υπερβολικά αυστηρή, όπως η κλασική μέθοδος Bonferroni.

Στους παρακάτω πίνακες παραθέτουμε τα αποτελέσματα αυτής της διαδικασίας:

Europe

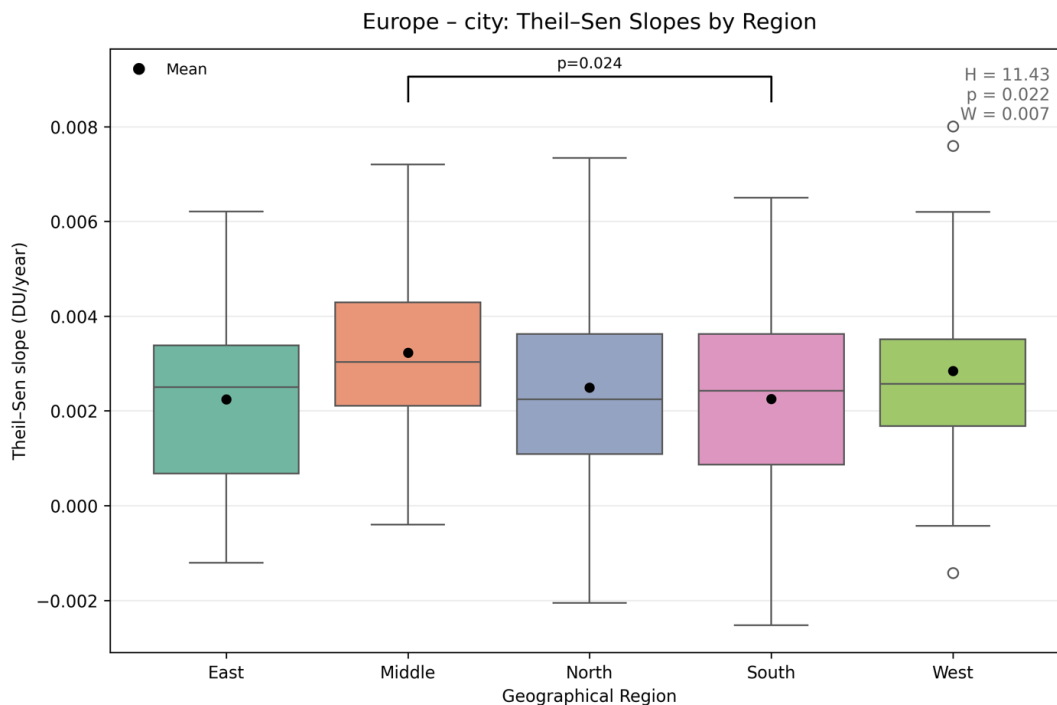
Level	n	H	p-value	Kendall's W
City	410	11.43	0.022	0.007
Country	41	2.06	0.725	0.013

Post hoc Dunn	East	Middle	North	South	West
East	1	0.122	1	1	1
Middle	0.122	1	0.142	0.024	1
North	1	0.142	1	1	1
South	1	0.024	1	1	0.621
West	1	1	1	0.621	1

Η εφαρμογή του μη παραμετρικού ελέγχου Kruskal-Wallis, ανέδειξε ότι οι ρυθμοί μεταβολής της τάσης των συγκεντρώσεων διοξειδίου του θείου, παρουσιάζουν στατιστικά ανιχνεύσιμες διαφοροποιήσεις μεταξύ των 5 περιφερειών ($H = 11.43$, όπου για επίπεδο σημαντικότητας $\alpha = 0.05$, το κατώφλι είναι $\chi^2(0.95, 5 - 1) = 9.49$). Οι διαφορές αυτές είναι μεν στατιστικά σημαντικές ($p = 0.022$), ωστόσο, δε συνοδεύονται από αντίστοιχη συνεπή ή ομοιογενή κατανομή των κατατάξεων των κλίσεων μεταξύ των περιφερειών, καθώς ο δείκτης Kendall's W είναι εξαιρετικά χαμηλός (0.007).

Ο post-hoc Dunn έλεγχος έδειξε ότι η στατιστικά σημαντική διαφορά εντοπίζεται μεταξύ Κεντρικής και Νότιας Ευρώπης.

Το παρακάτω διάγραμμα απεικονίζει την κατανομή των κλίσεων Theil-Sen των πόλεων ανά περιφέρεια, με τα αποτελέσματα του στατιστικού ελέγχου.



Γράφημα 30: Κατανομή κλίσεων Theil-Sen ανά περιφέρεια της Ευρώπης.

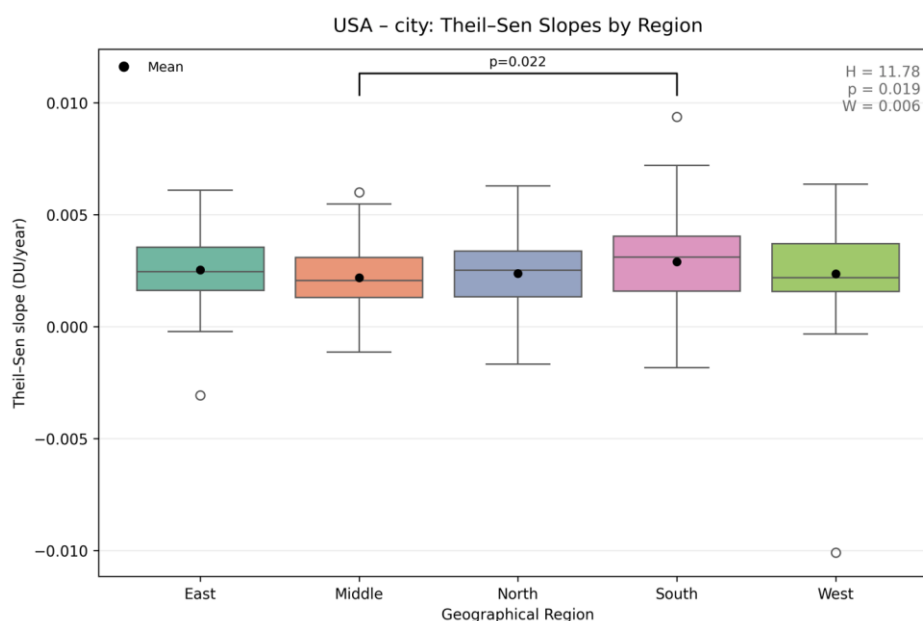
USA

Level	n	H	p-value	Kendall's W
City	410	11,78	0.019	0.006
Country	41	4,70	0.319	0.022

Post hoc Dunn	East	Middle	North	South	West
East	1	1	1	1	1
Middle	1	1	1	0.022	1
North	1	1	1	0.096	1
South	1	0.022	0.096	1	0.660
West	1	1	1	0.660	1

Παρόμοια μοτίβο με εκείνο της Ευρώπης, παρατηρούμε και στις πόλεις των ΗΠΑ, όπου η διαφορά στους ρυθμούς μεταβολής των συγκεντρώσεων διοξειδίου του θείου εντοπίζεται κι εδώ μεταξύ Κεντρικής και Νότιας περιφέρειας, με εξίσου όμως σχεδόν απόλυτη απουσία ομοιογενούς ή συνεπούς κατανομής των κατατάξεων των ρυθμών μεταβολής της τάσης ($W = 0.006$).

Και απεικονιστικά,



Γράφημα 31: Κατανομή κλίσεων Theil-Sen ανά περιφέρεια των ΗΠΑ.

Συμπερασματικά, ο έλεγχός μας ανέδειξε παρόμοια εικόνα σε επίπεδο περιφερειών, μεταξύ Ευρώπης και ΗΠΑ. Και στις δύο ηπείρους, ο post-hoc Dunn έλεγχος έδειξε ότι η στατιστικά σημαντική διαφοροποίηση των ρυθμών μεταβολής της τάσης, εντοπίζεται μεταξύ της Κεντρικής και της Νότιας περιφέρειας. Ωστόσο, ο δείκτης Kendall's W κατέδειξε πως η διαφοροποίηση αυτή δε συνοδεύεται από ομοιογενή ή συνεπή κατανομή των κατατάξεων των κλίσεων των περιφερειών, αντικατοπτρίζοντας πιθανότατα την ανομοιογένεια μεταξύ των πόλεων και κρατών, που συνθέτουν τις περιφέρειες.

4.6.5 Ανάλυση σταθερότητας τάσης (Rolling trend analysis)

Στην προηγούμενη υποενότητα υπολογίσαμε την κλίση της τάσης συνολικά σε ολόκληρη τη χρονοσειρά κάθε οντότητας, αποτυπώνοντας το γενικό ρυθμό μεταβολής των συγκεντρώσεων διοξειδίου του θείου. Στην παρούσα υποενότητα, διερευνούμε τη σταθερότητα της εκτιμώμενης τάσης μέσα από μια πιο λεπτομερή χρονική ανάλυση, εφαρμόζοντας τον εκτιμητή Theil-Sen σε διαδοχικά επικαλυπτόμενα χρονικά παράθυρα, ώστε να καταγράψουμε την εξέλιξη της τοπικής κλίσης σε διαφορετικά τμήματα της χρονοσειράς και να αξιολογήσουμε τη συνέπεια της συμπεριφοράς της στο χρόνο. Προς τούτο, καταστήσαμε 3 δείκτες που εφαρμόζονται στις κυλιόμενες κλίσεις και αποτυπώνουν τη συνέπεια του πρόσημου, δηλαδή αν η τάση παραμένει ανοδική ή καθοδική, την απόκλιση της έντασης από τη συνολική κλίση, καθώς και ένα σύνθετο δείκτη που εκφράζει συνολικά τον βαθμό σταθερότητας της τάσης.

Αναλυτικότερα, υπολογίσαμε τις κλίσεις Theil-Sen σε διαδοχικά επικαλυπτόμενα χρονικά παράθυρα μήκους 36, 48 και 60 μηνών, με βήμα μετατόπισης 3, 6 και 12 μηνών αντίστοιχα και αναπτύξαμε τους εξής δείκτες:

α) Δείκτης συνέπειας πρόσημου (Sign Consistency Index), ο οποίος εκφράζει τη σταθερότητα του προσήμου της κλίσης στα διαδοχικά χρονικά παράθυρα και δίδεται από τη σχέση:

$$SCI = 1 - \frac{N_{\text{changes}}}{N - 1}$$

όπου, N_{changes} είναι το πλήθος των αλλαγών πρόσημου μεταξύ διαδοχικών κλίσεων και N το πλήθος των χρονικών παραθύρων.

Ο υπόψη δείκτης που είναι επίσης κανονικοποιημένος, λαμβάνει τιμές από 0 έως 1, όπου 0 σημαίνει ότι το πρόσημο αλλάζει σε κάθε βήμα και 1 υποδηλώνει πλήρως συνεπές πρόσημο της κλίσης της τάσης, δηλαδή η κλίση είναι πάντα ανοδική ή καθοδική.

β) Δείκτης απόκλισης (Deviation Index), ο οποίος υπολογίζει την απόκλιση της μέσης κλίσης του κυλιόμενου παραθύρου από τη συνολική κλίση της τάσης και δίδεται από τη σχέση:

$$D = \min \left(\frac{|\bar{\beta}_{TS} - \beta_{TS,global}|}{|\beta_{TS,global}|}, 1 \right)$$

όπου, $\bar{\beta}_{TS}$ είναι η μέση κλίση του παραθύρου και $\beta_{TS,global}$ είναι η συνολική κλίση της τάσης της χρονοσειράς. Ο ως άνω δείκτης είναι κανονικοποιημένος και λαμβάνει τιμές από 0 έως 1, όπου 0 σημαίνει πλήρη συμφωνία με τη συνολική κλίση και 1 υποδεικνύει τη μέγιστη απόκλιση. Επισημαίνεται πως στην περίπτωση όπου $\beta_{TS,global} = 0$ τότε ο δείκτης D δεν υπολογίζεται (σ.σ. τεχνικά επιστρέφεται NaN).

γ) Σχετικός δείκτης σταθερότητας τάσης, (Relative Trend Stability Index), ο οποίος συνδυάζει τους δύο προηγούμενους δείκτες, ώστε να προκύψει η μετρική της σταθερότητας, μέσω της σχέσης:

$$RTSI = SCI \times (1 - D)$$

Ομοίως, οι τιμές που λαμβάνει ο δείκτης $RTSI$ κυμαίνονται από 0 έως 1, όπου 0 σημαίνει πλήρη αστάθεια και 1 απόλυτα σταθερή κλίση τάσης.

Επικουρικά των ανωτέρω, χρησιμοποιήσαμε δύο ελέγχους, ήτοι,

α) Levene test [87], ο οποίος χρησιμοποιείται για να εξετάσει αν δύο ή περισσότερες ομάδες δεδομένων παρουσιάζουν ομοιογένεια στη διασπορά των τιμών τους. Στην περίπτωσή μας εφαρμόζεται στις κλίσεις των τάσεων (Theil-Sen slopes) που προκύπτουν από τα διαδοχικά χρονικά παράθυρα. Αρχικά, για κάθε παράθυρο υπολογίζεται η τοπική κλίση της τάσης (δηλαδή ο ρυθμός μεταβολής των συγκεντρώσεων ανά έτος). Στη συνέχεια, το πλήθος των παραγόμενων κλίσεων (πχ. για παράθυρο 36 μηνών με βήμα 3 μηνών, προκύπτουν 29 τοπικές κλίσεις) χωρίζεται σε τρεις ίσες ομάδες, ήτοι, στην αρχική, στη μεσαία και στην τελική περίοδο. Έπειτα, ελέγχεται εάν οι τρεις αυτές περίοδοι παρουσιάζουν παρόμοια διασπορά των τοπικών κλίσεων.

Η μηδενική υπόθεση (H_0) δηλώνει ότι όλες οι ομάδες περιόδων έχουν ίση διασπορά, ενώ η εναλλακτική (H_1) ότι τουλάχιστον μία παρουσιάζει διαφορετική διασπορά. Συνεπώς, αν το p-value είναι < 0.05 , η H_0 απορρίπτεται, υποδεικνύοντας ότι οι διασπορές των κλίσεων διαφέρουν σημαντικά μεταξύ των περιόδων. Δηλαδή η διακύμανση της τοπικής κλίσης δεν παραμένει σταθερή στο χρόνο, γεγονός που αποτελεί ένδειξη πιθανής αστάθειας της συνολικής τάσης.

και

β) μία παραλλαγή του Sequential Mann-Kendall test (ή εναλλακτικά Mann-Kendall-Sneyers test) για την ανίχνευση σημείων μεταβολής της κλίσης της τάσης (change points) που υποδεικνύουν διαστήματα όπου η κατεύθυνση ή η ένταση της τάσης μεταβάλλεται σημαντικά.

Πιο συγκεκριμένα, ο Raymond Sneyers, στο έργο του 'On the Statistical Analysis of Series of Observations' [88], αναφέρει ότι ο μη παραμετρικός έλεγχος Mann-Kendall είναι ιδιαίτερα κατάλληλος για εφαρμογή, όταν, στην

περίπτωση που διαπιστώνεται στατιστικά σημαντική τάση, επιθυμούμε να εντοπίσουμε το χρονικό σημείο έναρξης του φαινομένου.

Για το σκοπό αυτό προτείνει τη χρήση διαδοχικής (sequential) ανάλυσης, δηλαδή την επαναληπτική εφαρμογή του ελέγχου Mann-Kendall κατά μήκος της χρονοσειράς, έτσι ώστε να υπολογίζεται προοδευτικά το στατιστικό $u(t)$ για κάθε χρονικό βήμα. Η εξέλιξη αυτών των τιμών επιτρέπει την ανίχνευση χρονικών σημείων όπου μεταβάλλεται η κατεύθυνση ή η ένταση της τάσης, δηλαδή τα σημεία μεταβολής της κλίσης (change-points).

Στην περίπτωση μας, η εφαρμογή του παραπάνω προοδευτικού ελέγχου οδηγεί κάποιες φορές στον εντοπισμό προσωρινών αλλαγών κατεύθυνσης, οι οποίες δεν αντανακλούν στην ουσιαστική μεταβολή της τάσης αλλά οφείλονται σε βραχυπρόθεσμα γεγονότα, όπως για παράδειγμα η περίοδος COVID-19, όπου για ένα συγκεκριμένο χρονικό διάστημα άλλαξε σημαντικά η καθημερινότητα σε σημαντικό πλήθος περιοχών του πλανήτη, με σαφείς επεκτάσεις και στις συγκεντρώσεις του διοξειδίου του θείου.

Για την αποφυγή τέτοιων περιστατικών προβήκαμε στην εφαρμογή συγκεκριμένων κανόνων που επηρεάζουν (όχι σημαντικά) τα αποτελέσματα του sequential Mann-Kendall test. Αναλυτικότερα, α) όταν το πρόσημο της πρώτης και τελευταίας κλίσης της τάσης είναι το ίδιο, τότε δε θεωρείται πραγματική αντιστροφή, β) αν η διαφορά της μεταβολής είναι μικρότερη από 0.001 DU/year, τότε η μεταβολή θεωρείται αμελητέα και γ) μεταβολές που προκύπτουν μόνο από μία αντιστροφή πρόσημου, χαρακτηρίζονται ως βραχύχρονες και δε λαμβάνονται υπόψη. Με τον τρόπο αυτό εξασφαλίζουμε κατά το δυνατό ότι αναγνωρίζονται μόνο διαρκείς και ουσιαστικές μεταβολές στην κατεύθυνση της τάσης, ενώ αγνοούνται προσωρινές διαταραχές που δεν αλλοιώνουν το γενικό μοτίβο.

Κατόπιν των ανωτέρω, καταστήσαμε την ακόλουθη κατηγοριοποίηση,

Category	Criteria
High stability trend	MK p-value < 0.05 and RTSI \geq 0.70
Stable trend	MK p-value < 0.05 and $0.50 \leq$ RTSI < 0.70
Moderate stability trend	MK p-value < 0.05 and $0.30 \leq$ RTSI < 0.50
Unstable trend	MK p-value < 0.05 and RTSI < 0.30
No trend	MK p-value \geq 0.05

Τα αποτελέσματα από την εφαρμογή της ανωτέρω διαδικασίας παρατίθενται αναλυτικά παρακάτω.

Europe

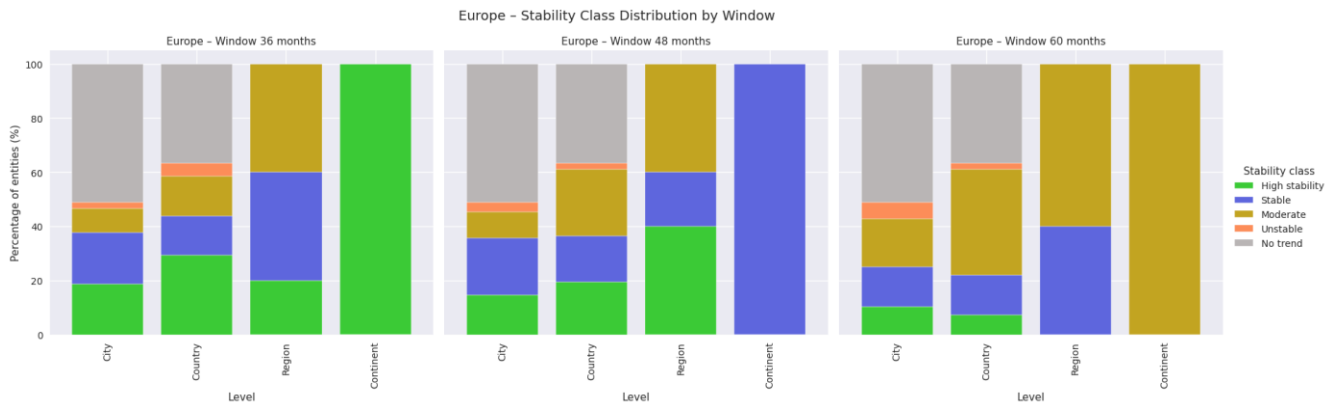
window	step	Level	n	Mean slope	High stability trend	Stable trend	Moderate stability trend	Unstable trend	No trend
36	3	City	410	0,002305	77	78	36	10	209
		Country	41	0,002591	12	6	6	2	15
		Region	5	0,002501	1	2	2	0	0
		Continent	1	0,002240	1	0	0	0	0
48	6	City	410	0,002066	60	87	39	15	209
		Country	41	0,002207	8	7	10	1	15
		Region	5	0,002072	2	1	2	0	0
		Continent	1	0,001899	0	1	0	0	0
60	12	City	410	0,001967	42	61	73	25	209
		Country	41	0,002180	3	6	16	1	15
		Region	5	0,001982	0	2	3	0	0
		Continent	1	0,001771	0	0	1	0	0

Στην Ευρώπη, για παράθυρο 36 μηνών (βήμα 3μ), περίπου το 38% των πόλεων εμφανίζουν σταθερές ή υψηλής σταθερότητας τάσεις (155 από 410), ενώ οι ασταθείς περιπτώσεις είναι ελάχιστες (10 περιπτώσεις, ήτοι περίπου 2%). Το υπόλοιπο 51% των πόλεων δεν παρουσιάζει στατιστικά σημαντική τάση, κάτι που αναφέραμε και στην αρχή της παρούσας ενότητας. Σε ανώτερα χωρικά επίπεδα, η εικόνα είναι σχεδόν ανάλογη με τις περισσότερες χώρες (18 από 41 περιπτώσεις) να εμφανίζουν σταθερή ή υψηλής σταθερότητας τάση και οι περιφέρειες να κατατάσσονται κυρίως στις κατηγορίες σταθερή και μέτριας σταθερότητας τάση. Όσον, αφορά το επίπεδο ηπείρου, παρατηρούμε υψηλή σταθερότητα τάσης με χαμηλό θετικό ρυθμό μεταβολής των συγκεντρώσεων του διοξειδίου του θείου.

Καθώς αυξάνεται το μήκος του παραθύρου, παρατηρείται σταδιακή μετατόπιση προς χαμηλότερες κατηγορίες σταθερότητας, καταλήγοντας στο παράθυρο των 60 μηνών (βήμα 12μ), σε επίπεδο πόλεων, τα ποσοστά να περιορίζονται από 38% σε 25% για τις σταθερές και υψηλής σταθερότητας κατηγορίες και σε επίπεδο ηπείρου η σταθερότητα της τάσης να κατηγοριοποιείται στη βαθμίδα μέσης σταθερότητας.

Βάσει των πιο πάνω αποτελεσμάτων, φαίνεται πως παρότι η αύξηση του μήκους του παραθύρου οδηγεί σε εξομάλυνση της εκτιμώμενης τάσης, η σταθερότητά της μειώνεται, διότι τα μεγαλύτερα παράθυρα ενσωματώνουν περιόδους με διαφορετικούς ρυθμούς μεταβολής των συγκεντρώσεων του διοξειδίου του θείου, καθιστώντας έτσι τη συνολική κλίση λιγότερο συνεπή μεταξύ διαδοχικών παραθύρων.

Οπτικά και συγκριτικά για όλα τα παράθυρα, παραθέτουμε το ακόλουθο στιγμιότυπο,



Γράφημα 32: Ταξινόμηση χωρικών επιπέδων, ανά κατηγορία σταθερότητας τάσης, στην Ευρώπη.

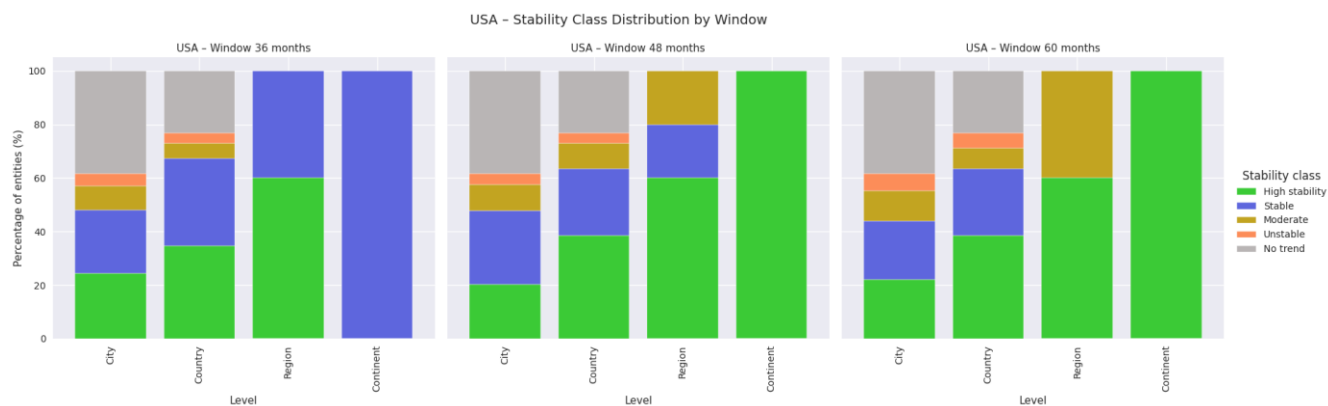
όπου παρατηρούμε πως όσο αυξάνεται το μήκος του παραθύρου, μειώνεται σταδιακά το ποσοστό της υψηλής σταθερότητας τάσης και αυξάνεται εκείνο της μέτριας σταθερότητας.

USA

window	step	Level	n	Mean slope	High stability trend	Stable trend	Moderate stability trend	Unstable trend	No trend
36	3	City	511	0,002314	124	122	46	23	196
		State	52	0,002209	18	17	3	2	12
		Region	5	0,002347	3	2	0	0	0
		Continent	1	0,002130	0	1	0	0	0
48	6	City	511	0,002308	103	141	50	21	196
		State	52	0,002180	20	13	5	2	12
		Region	5	0,002249	3	1	1	0	0
		Continent	1	0,002157	1	0	0	0	0
60	12	City	511	0,002451	113	111	58	33	196
		State	52	0,002357	20	13	4	3	12
		Region	5	0,002325	3	0	2	0	0
		Continent	1	0,002537	1	0	0	0	0

Στην αντίπερα όχθη, παρατηρούμε μια ελαφρώς διαφορετική εικόνα, όπου στο παράθυρο των 36 μηνών (βήμα 3μ) το 48% των πόλεων εμφανίζουν σταθερές ή υψηλής σταθερότητας τάσεις (246 από 511), ενώ οι ασταθείς περιπτώσεις είναι πολύ λίγες (23 περιπτώσεις, ήτοι περίπου 4.5 %). Σε ανώτερα χωρικά επίπεδα, η εικόνα είναι σχεδόν ανάλογη με τις περισσότερες χώρες (35 από 52 περιπτώσεις) να εμφανίζουν σταθερή ή υψηλής σταθερότητας τάση και οι περιφέρειες να κατατάσσονται αποκλειστικά στις κατηγορίες υψηλής σταθερότητας και σταθερής τάσης.

Όσο αυξάνεται το μήκος του παραθύρου, παρατηρείται μια μικρή μείωση των περιπτώσεων υψηλής και σταθερής τάσης και αντίστοιχη μικρή αύξηση των μέτριων ή ασταθών περιπτώσεων. Σε επίπεδο ηπείρου, παρατηρούμε ότι μια λίγο διαφορετική συμπεριφορά, όπου όσο αυξάνεται το παράθυρο, αυξάνεται και η κατηγοριοποίηση της σταθερότητας (από σταθερή σε υψηλής σταθερότητας τάσης). Εκτίμησή μας είναι και θα το δούμε και παρακάτω, ότι η κατηγοριοποίηση ως Stable στο παράθυρο των 36 μηνών οφείλεται σε αποτίμηση οριακά κάτω από το κατώφλι της υψηλής σταθερότητας (κατά μόλις 0.02 μονάδες).



Γράφημα 33: : Ταξινόμηση χωρικών επιπέδων, ανά κατηγορία σταθερότητας τάσης, στις ΗΠΑ.

Πάντως, στο παραπάνω στιγμιότυπο, φαίνεται ξεκάθαρα η σταθερότητα μεταξύ των παραθύρων για όλα τα επίπεδα.

Αξιοσημείωτη είναι η διαφορετική συμπεριφορά μεταξύ Ευρώπης και ΗΠΑ, όπου για την πρώτη παρατηρούμε ότι όσο αυξάνεται το παράθυρο παρατήρησης μειώνεται ο ρυθμός μεταβολής των συγκεντρώσεων διοξειδίου του

θείου (πχ. για τις πόλεις: 0.002305 (36μ) → 0.002066 (48 μ) → 0.001967 (60 μ)), ενώ στις ΗΠΑ δεν εμφανίζεται παρόμοιο μοτίβο (πχ. 0.002314 (36μ) → 0.002308 (48μ) → 0.002451 (60μ)), αλλά αντίθετα παρατηρείται μία ήπια μεταβλητότητα έως σταθερότητα. Το φαινόμενο αυτό πιθανόν να υποδηλώνει ότι η Ευρώπη έχει φτάσει σε ένα σημείο όπου ο ρυθμός μεταβολής των συγκεντρώσεων αρχίζει να σταθεροποιείται, έπειτα από μια μακρά περίοδο μείωσης, ενώ στις ΗΠΑ φαίνεται να υπάρχει διαχρονικά μία πολύ μικρή αλλά σταθερή ανοδική πορεία.

Πέραν αυτών, η ανάλυσή μας επεκτάθηκε τόσο στις τιμές των δεικτών που υπολογίσαμε όσο και στους δύο ελέγχους που εφαρμόσαμε. Παρακάτω παραθέτουμε τα αποτελέσματα ξεχωριστά για τις οντότητες που σύμφωνα με τον αρχικό έλεγχο Mann-Kendall, έχουν στατιστικά σημαντική τάση και για εκείνες που κατατάχθηκαν στην κατηγορία ‘No trend’.

Πιο συγκεκριμένα,

Europe

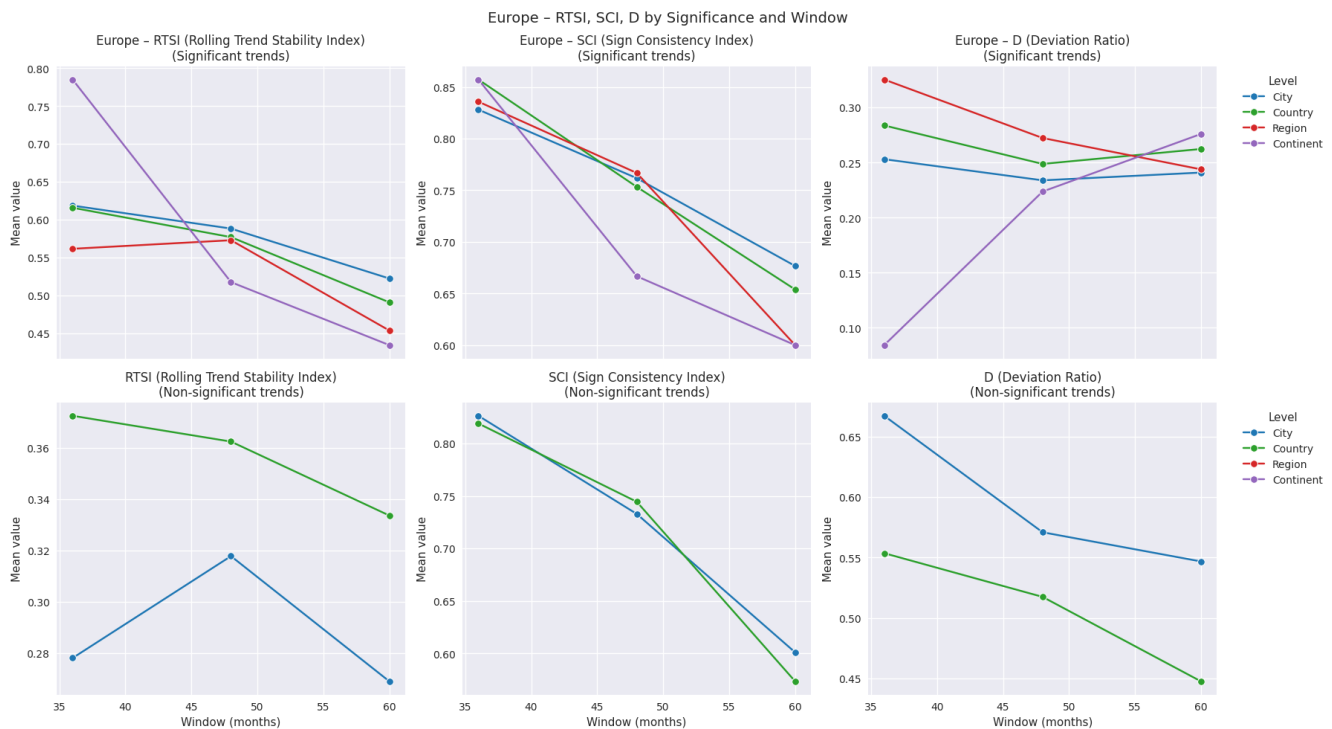
Win (step)	Level	n		Mean SCI		Mean D		Mean RTSI		Mean p _{Levene}		Change point	
		Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
is trend significant:													
36 (3)	City	201	209	0.83	0.83	0.25	0.67	0.62	0.28	0.36	0.37	48	49
	Country	26	15	0.86	0.82	0.28	0.55	0.61	0.37	0.39	0.37	2	2
	Region	5	0	0.84	-	0.33	-	0.56	-	0.22	-	0	-
	Continent	1	0	0.86	-	0.08	-	0.78	-	0.73	-	0	-
48 (6)	City	201	209	0.76	0.73	0.23	0.57	0.59	0.32	0.47	0.45	13	22
	Country	26	15	0.75	0.74	0.25	0.52	0.58	0.36	0.44	0.59	1	0
	Region	5	0	0.77	-	0.27	-	0.57	-	0.54	-	0	-
	Continent	1	0	0.67	-	0.22	-	0.52	-	0.63	-	0	-
60 (12)	City	201	209	0.68	0.60	0.24	0.55	0.52	0.26	0	0	0	0
	Country	26	15	0.65	0.57	0.26	0.45	0.49	0.33	0	0	0	0
	Region	5	0	0.60	-	0.24	-	0.45	-	0	-	0	-
	Continent	1	0	0.60	-	0.28	-	0.43	-	0	-	0	-

Στο επίπεδο των πόλεων, τα αποτελέσματα δείχνουν ότι η συνέπεια του πρόσημου των επιμέρους κλίσεων (Mean SCI) παραμένει σε υψηλά επίπεδα ανεξαρτήτως παραθύρου και από 0.8 για το παράθυρο των 36 μηνών υποχωρεί σταδιακά σε 0.6 για εκείνο των 60 μηνών. Αυτό υποδηλώνει ότι οι τάσεις στις περισσότερες πόλεις διατηρούν την ίδια κατεύθυνση, χωρίς συχνές αντιστροφές πρόσημου.

Ο δείκτης D, ο οποίος εκφράζει την απόκλιση της μέσης κυλιόμενης κλίσης από τη συνολική τάση, παρουσιάζει εντονότερες διαφοροποιήσεις μεταξύ των στατιστικά σημαντικών και μη σημαντικών περιπτώσεων. Για τις οντότητες με στατιστικά σημαντική μακροχρόνια τάση, οι τιμές του παραμένουν ιδιαίτερα χαμηλές σε όλα τα παράθυρα, γεγονός που υποδηλώνει ότι η τοπική συμπεριφορά της τάσης ταυτίζεται με εκείνη της συνολικής εκτίμησης. Αντίθετα, για τις μη σημαντικές τάσεις, οι τιμές του δείκτη είναι συστηματικά υψηλότερες, κάτι που υποδεικνύει μεγαλύτερη μεταβλητότητα και ασυνέπεια της τάσης στο χρόνο.

Ο δείκτης σταθερότητας RTSI, ο οποίος υπενθυμίζουμε πως συνδυάζει τη συνέπεια πρόσημου (SCI) και έντασης (D), παρουσιάζει αντίστοιχο μοτίβο, όπου οι οντότητες με στατιστικά σημαντική τάση, εμφανίζουν τιμές οι οποίες ταξινομούνται στις κατηγορίες σταθερής και υψηλής σταθερότητας τάσης σε όλα τα παράθυρα ανάλυσης.

Συνολικά, τα ευρήματά μας, υποδεικνύουν ότι η κατεύθυνση της τάσης παραμένει κατά κανόνα σταθερή, αλλά η συνέπεια της έντασης της μειώνεται όσο αυξάνεται το παράθυρο της ανάλυσης.



Γράφημα 34: Συνοπτική απεικόνιση δεικτών σταθερότητας ανά χωρικό επίπεδο στην Ευρώπη.

Το παραπάνω στιγμιότυπο επιβεβαιώνει πλήρως τα ευρήματά μας, δείχνοντας με σαφήνεια ότι στις περιπτώσεις στατιστικά σημαντικής τάσης (επάνω σειρά διαγραμμάτων), οι τιμές των δεικτών RTSI και SCI είναι αισθητά υψηλότερες σε σχέση με εκείνες των μη στατιστικά σημαντικών περιπτώσεων και μειώνονται σταδιακά όσο αυξάνεται το μήκος του παραθύρου ανάλυσης. Αντίθετα, ο δείκτης D, στις περιπτώσεις στατιστικά σημαντικής τάσης, διατηρείται συστηματικά σε χαμηλά επίπεδα, υποδηλώνοντας μεγαλύτερη συνοχή μεταξύ των τοπικών και της συνολικής τάσης, σε σύγκριση με τις μη σημαντικές περιπτώσεις.

USA

Win (step)	Level	n		Mean SCI		Mean D		Mean RTSI		Mean p_{Levene}		Change point	
		Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
36 (3)	City	315	196	0.82	0.81	0.24	0.57	0.62	0.34	0.32	0.30	186	120
	State	40	12	0.83	0.80	0.21	0.39	0.66	0.48	0.32	0.30	18	7
	Region	5	0	0.85	-	0.15	-	0.73	-	0.50	-	3	-
	Continent	1	0	0.79	-	0.13	-	0.68	-	0.43	-	1	-
48 (6)	City	315	196	0.79	0.75	0.23	0.56	0.61	0.33	0.42	0.42	111	82
	State	40	12	0.80	0.81	0.19	0.36	0.65	0.53	0.32	0.33	15	5
	Region	5	0	0.85	-	0.16	-	0.71	-	0.30	-	1	-
	Continent	1	0	0.92	-	0.12	-	0.81	-	0.27	-	0	-
60 (12)	City	315	196	0.75	0.64	0.21	0.55	0.60	0.29	0	0	0	0
	State	40	12	0.78	0.68	0.17	0.36	0.65	0.43	0	0	0	0
	Region	5	0	0.80	-	0.18	-	0.66	-	0	-	0	-
	Continent	1	0	0.80	-	0.04	-	0.77	-	0	-	0	-

Στις ΗΠΑ, σε όλα τα επίπεδα, ο δείκτης συνέπειας πρόσημου (SCI) παραμένει σταθερά υψηλός, με τιμές γύρω στο 0.8 για όλα τα παράθυρα ανάλυσης, ακόμη και στο μεγαλύτερο παράθυρο των 60 μηνών, όπου στην Ευρώπη

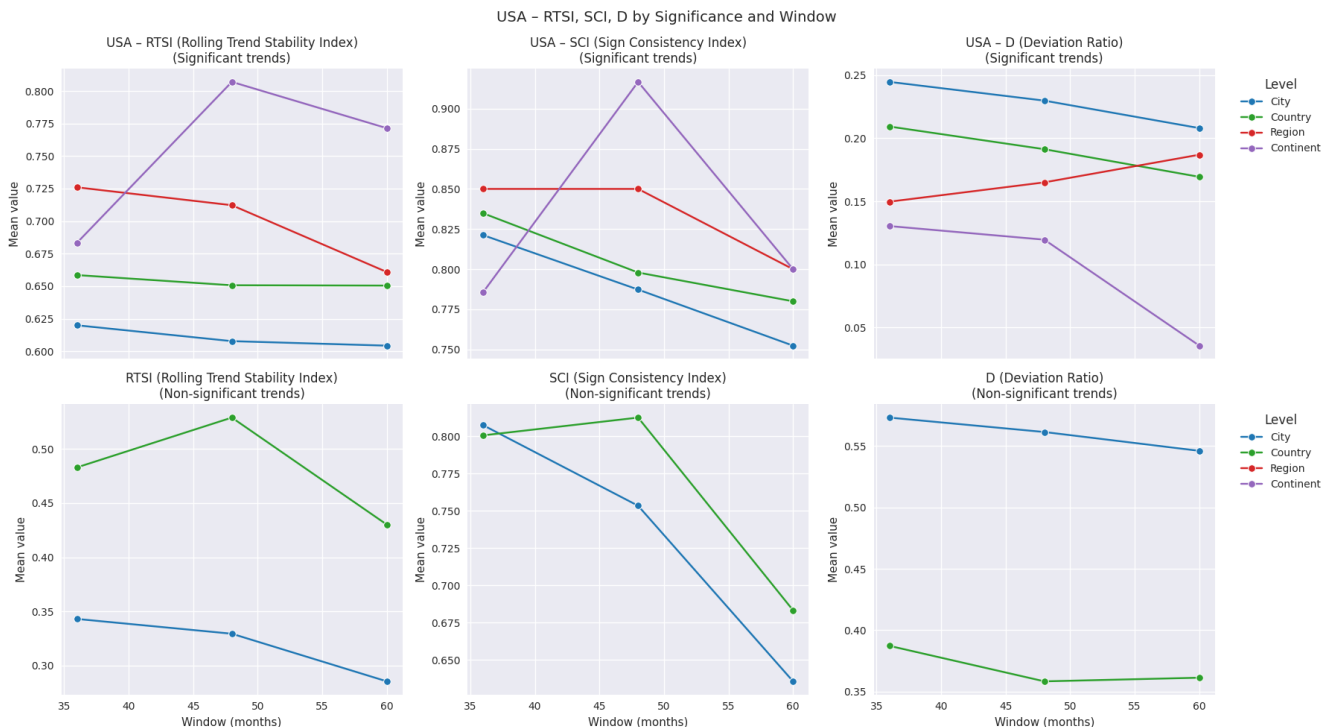
ο υπόψη δείκτης είχε μειωθεί αισθητά (0.6). Αυτό υποδηλώνει ότι ο ρυθμός μεταβολής των συγκεντρώσεων διοξειδίου του θείου χαρακτηρίζεται από μεγαλύτερη συνέπεια κατεύθυνσης μέσα στον χρόνο, με σημαντικά λιγότερες αντιστροφές πρόσημου.

Παρόμοια εικόνα εμφανίζει και ο δείκτης D, όπου οι τιμές του είναι σταθερά χαμηλότερες από ότι στην Ευρώπη, ιδίως για τις οντότητες με στατιστικά σημαντική τάση. Αυτό δείχνει ότι οι τοπικές κυλιόμενες κλίσεις προσεγγίζουν με μεγαλύτερη συνέπεια τη συνολική τάση, κάτι που υποδεικνύει μία πιο σταθερή συμπεριφορά έναντι της Ευρώπης. Αντίθετα, οι μη σημαντικές τάσεις εμφανίζουν ελαφρώς αυξημένες αποκλίσεις, χωρίς όμως να παρατηρούνται τόσο σημαντικές διαφορές όπως στην Ευρώπη.

Αυτή η συμπεριφορά αποτυπώνεται και στο δείκτη RTSI, διατηρώντας σταθερά υψηλές τιμές ανεξαρτήτως παραθύρου. Μάλιστα, κι εδώ παρατηρούμε ότι όσο αυξάνεται το παράθυρο η αντίστοιχη μείωση των τιμών του δείκτη είναι ελάχιστη σε σχέση με το μοτίβο που παρατηρήσαμε στην Ευρώπη, όπου η πτώση ήταν αισθητά πιο έντονη.

Ο δείκτης p_{Levene} επιβεβαιώνει την ομοιογένεια των διακυμάνσεων, όπου οι μέσες τιμές του κυμαίνονται σταθερά γύρω στο 0.3, χωρίς κάποια ένδειξη αστάθειας στη διασπορά. Αξιοσημείωτο είναι ότι ο αριθμός των εντοπισμένων 'change points' στις ΗΠΑ είναι σαφώς μεγαλύτερος από ότι στην Ευρώπη, κάτι που μας οδηγεί στο συμπέρασμα ότι παρά τη σταθερότητα στη συνολική τάση, εμφανίζονται περισσότερες τοπικές μεταβατικές φάσεις, που όμως δεν αλλοιώνουν τη γενική κατεύθυνση. Όπως, παρατηρήθηκε και προηγουμένως, όσο αυξάνεται το παράθυρο ανάλυσης, μειώνονται αισθητά αυτές οι μεταβατικές φάσεις.

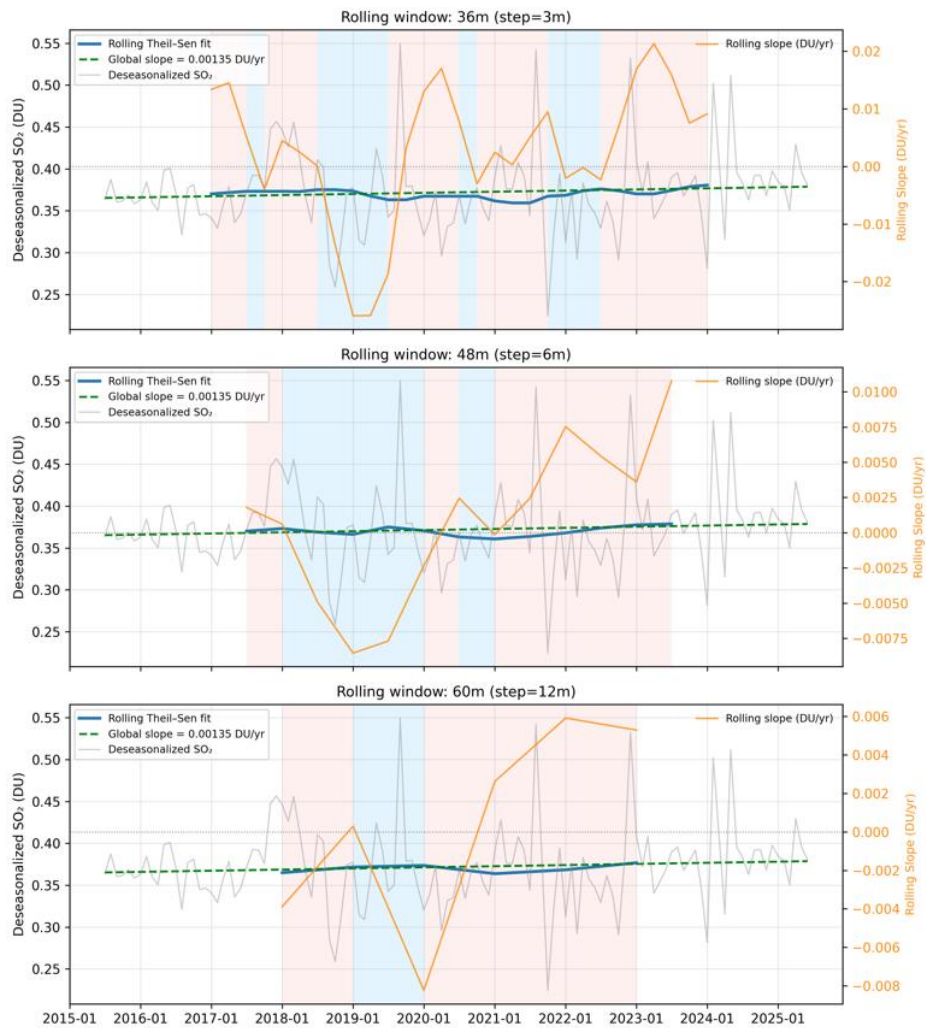
Το παρακάτω διάγραμμα επιβεβαιώνει και εδώ τα ευρήματα της ανάλυσής μας, καθώς στις ΗΠΑ οι δείκτες RTSI και SCI διατηρούνται σε υψηλά επίπεδα ανεξαρτήτως παραθύρου, ενώ ο δείκτης D παραμένει χαμηλός και σταθερός, υποδεικνύοντας υψηλότερη συνοχή και σταθερότητα της τάσης σε σχέση με εκείνη της Ευρώπης.



Γράφημα 35: Συνοπτική απεικόνιση δεικτών σταθερότητας ανά χωρικό επίπεδο στις ΗΠΑ.

Προτού ολοκληρώσουμε την ανάλυσή μας, κρίνουμε σκόπιμο να προβούμε στη συγκριτική αξιολόγηση της δυναμικής και της σταθερότητας της τάσης μεταξύ δύο πόλεων που ανήκουν σε διαφορετικές ηπείρους, ήτοι, της Αθήνας και του Manhattan (New York). Όπως θα περιγράψουμε και παρακάτω, οι δύο αυτές πόλεις παρουσιάζουν μία σαφώς διαφορετική συμπεριφορά ως προς το ρυθμό μεταβολής των συγκεντρώσεων διοξειδίου του θείου.

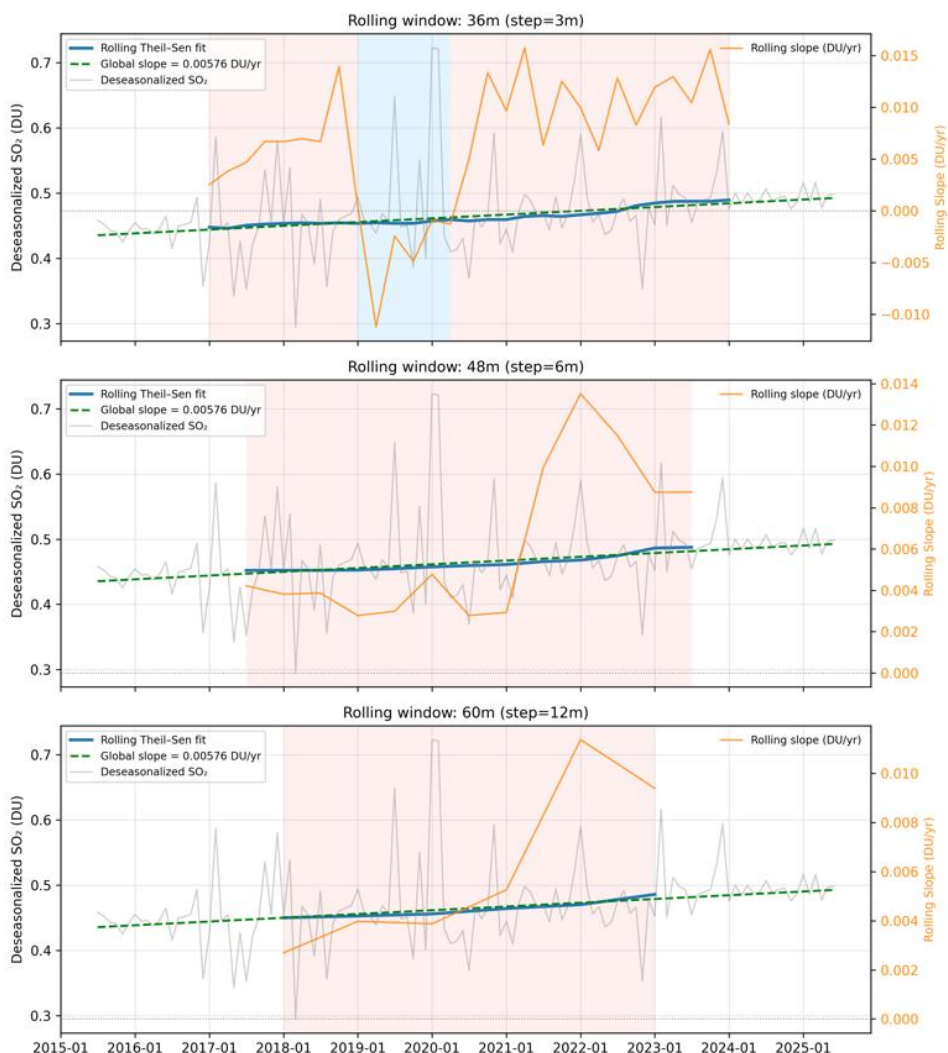
Greece – Athens (City)
Continuous Rolling Theil-Sen Trend Curves (36m, 48m, 60m)



Γράφημα 36: Συγκριτική απεικόνιση διαδικασίας υπολογισμού σταθερότητας τάσης ανά παράθυρο ελέγχου, για την πόλη της Αθήνας.

Στην Αθήνα, η συνολική κλίση είναι θετική αλλά μικρού μεγέθους (0.0013 DU/year), ενώ το αποτέλεσμα του ελέγχου Mann-Kendall δεν είναι στατιστικά σημαντικό ($p = 0.216$). Αυτό σημαίνει ότι η μακροχρόνια ανοδική τάση δεν τεκμηριώνεται με στατιστική βεβαιότητα, γεγονός που αποτυπώνεται και στις κυλιόμενες εκτιμήσεις, όπου οι τοπικές κλίσεις μεταβάλλονται έντονα στο χρόνο, με εναλλαγές προσήμου και περιόδους προσωρινής καθόδου, όπως στην περίοδο 2019-2020.

New York — Manhattan (New York) (City)
Continuous Rolling Theil-Sen Trend Curves (36m, 48m, 60m)



Γράφημα 37: Συγκριτική απεικόνιση διαδικασίας υπολογισμού σταθερότητας τάσης ανά παράθυρο ελέγχου, για την πόλη του Manhattan της Νέας Υόρκης..

Αντίθετα, στο Manhattan της Νέας Υόρκης, η συνολική κλίση είναι σημαντικά υψηλότερη (0.0058 DU/year) και στατιστικά ισχυρή (Mann-Kendall test: $p < 10^{-7}$), γεγονός που υποδηλώνει μια σταθερή και συνεχή αύξηση των συγκεντρώσεων διοξειδίου του θείου σε όλη τη χρονική περίοδο ελέγχου. Οι κυλιόμενες κλίσεις παρουσιάζουν υψηλό βαθμό συνέπειας πρόσημου, παραμένοντας θετικές σε όλα τα παράθυρα (36m, 48m, 60m), με μόνο χαμηλής έντασης και περιορισμένου χρόνου μειώσεις.

Συγκριτικά, το Manhattan εμφανίζει μία πολύ πιο σταθερή ανοδική τάση, σε αντίθεση με την Αθήνα, όπου η γενική κατεύθυνση της τάσης είναι μεν ανοδική αλλά ασταθής, με σημαντική τοπική μεταβλητότητα και μη στατιστική τεκμηρίωση. Το αποτέλεσμα αυτό εναρμονίζεται πλήρως με τα συνολικά μοτίβα σταθερότητας που παρατηρήθηκαν σε ηπειρωτικό επίπεδο.

Συνοψίζοντας όσα προεκτέθηκαν, οι ΗΠΑ εμφανίζουν υψηλό βαθμό χρονικής συνέπειας και ομοιογένειας της τάσης, καθώς οι δείκτες σταθερότητας παραμένουν σε σχετικά σταθερά επίπεδα, ανεξαρτήτως μήκους παραθύρου. Το γεγονός αυτό υποδηλώνει πιθανότατα ότι οι μεταβολές που εξελίχθηκαν κατά τις προηγούμενες δεκαετίες έχουν πλέον σταθεροποιηθεί και οι εσωτερικές διακυμάνσεις στο χρόνο είναι περιορισμένες. Αντίθετα, στην Ευρώπη, αν και παρατηρείται συνολικά υψηλή συνέπεια πρόσημου, οι δείκτες σταθερότητας μειώνονται

σταδιακά όσο αυξάνεται το χρονικό παράθυρο, γεγονός που θα μπορούσε να υποδηλώνει ότι βρίσκεται ακόμη σε εξέλιξη κάποια διαδικασία εξισορρόπησης των συγκεντρώσεων διοξειδίου του θείου.

Σε κάθε περίπτωση, τα ευρήματά μας δείχνουν ότι και στις δύο ηπείρους οι τάσεις παραμένουν σταθερές και συνεπείς στη διάρκεια του χρόνου, με τις ΗΠΑ να παρουσιάζουν πιο εδραιωμένη σταθερότητα, ενώ η Ευρώπη φαίνεται να ακολουθεί μια πιο δυναμική, αλλά ελαφρώς μεταβατική, φάση προσαρμογής.

4.6.6 Ανάλυση μεταβολών τάσης (Trend Shift Analysis)

Η ανάλυση σταθερότητας της τάσης, μέσω των κυλιόμενων κλίσεων (Rolling Theil-Sen) και του Sequential Mann-Kendall ελέγχου, μας βοήθησε να διερευνήσουμε το κατά πόσο η τάση των συγκεντρώσεων του διοξειδίου του θείου παραμένει συνεπής στο χρόνο, ως προς το πρόσημο και την έντασή της. Ωστόσο, με την προσέγγισή μας αυτή, δεν εντοπίζουμε τα χρονικά σημεία εκείνα όπου η τάση μεταβάλλεται σημαντικά, ούτε αν πρόκειται για απότομη αλλαγή σε επίπεδο συγκεντρώσεων διοξειδίου του θείου ή για ουσιαστική αλλαγή στο ρυθμό μεταβολής τους.

Για να απαντήσουμε και σε αυτό το ερώτημα, εφαρμόσαμε τον αλγόριθμο PELT (Pruned Exact Linear Time), με συνάρτηση κόστους βασισμένη σε Radial Basis Function (RBF kernel), τόσο στις τοπικές κλίσεις (Theil-Sen slope), ώστε να εντοπιστούν εκείνες οι χρονικές στιγμές όπου αλλάζει σημαντικά ο ρυθμός μεταβολής των συγκεντρώσεων διοξειδίου του θείου όσο και απευθείας στις αποεποχικοποιημένες χρονοσειρές, προκειμένου να ανιχνευτούν μεταβολές στη συνολική κατανομή των τιμών πριν και μετά από κάθε σημείο αλλαγής. Αξίζει να σημειωθεί ότι η μέθοδος αυτή εντοπίζει δομικές μεταβολές στη συμπεριφορά της χρονοσειράς, δηλαδή γεγονότα που αλλάζουν τη μετέπειτα πορεία της τάσης και όχι μεμονωμένα ή βραχύβια επεισόδια.

Πιο συγκεκριμένα,

Ο αλγόριθμος PELT προτάθηκε από τους Rebecca Killick, Paul Fearnhead και Idris Eckley το 2012, στο άρθρο τους 'Optimal detection of changepoints with a linear computational cost' [89], ως μία βελτιστοποιημένη μέθοδος εντοπισμού σημείων μεταβολής (changepoints) σε μια χρονοσειρά, η οποία ελαχιστοποιεί το χρόνο εκτέλεσης, έναντι άλλων μεθόδων (Segment Neighbourhood και Binary Segmentation) και παράλληλα παράγει πιο ακριβή αποτελέσματα.

Ο τρόπος λειτουργίας του αλγόριθμου, όπως περιγράφεται και στο σχετικό άρθρο, είναι ο ακόλουθος:

Είσοδος:

Ένα σύνολο δεδομένων της μορφής (y_1, y_2, \dots, y_n) , όπου $y_i \in \mathbb{R}$,
η συνάρτηση κόστους $C(\cdot)$, όπου όσο μικρότερη η τιμή της C τόσο καλύτερη η προσαρμογή στα δεδομένα του τμήματος της χρονοσειράς που ελέγχουμε και
η ποινή β , η οποία δεν εξαρτάται από το πλήθος ή τη θέση των σημείων αλλαγής (change points)

Αρχικοποίηση:

n = μήκος των δεδομένων και

$F(0) = -\beta$, $cp(0) = 0$ (κενή λίστα), $pts = \{0\}$ (σύνολο υποψήφιων σημείων προς αξιολόγηση)

Επαναληπτικά για $\tau^* = 1, \dots, n$ (όπου τ^* είναι το τρέχον σημείο επεξεργασίας):

1. Υπολόγισε $F(\tau^*) = \min_{\tau \in (0, pts, \tau^* - 1)} [F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta]$
2. $\tau^1 = \arg \{ \min_{0 \leq \tau < \tau^*} [F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta] \}$
3. Θέσε $cp(\tau^*) = [cp(\tau^1), \tau^1]$
4. (pruning) $pts = \arg_{\tau} \{ F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta + K > F(\tau^*) + \beta \}$

Έξοδος: Τα σημεία αλλαγών (change points), που είναι τα στοιχεία του $cp(n)$ εκτός του n .

Συνοπτικά, ο αλγόριθμος PELT εξετάζει διαδοχικά κάθε σημείο της χρονοσειράς για να διαπιστώσει αν σε εκείνο το σημείο αλλάζει κάτι ουσιαστικό. Για παράδειγμα αν οι τιμές μετά από αυτό το σημείο ακολουθούν διαφορετική συμπεριφορά, από ότι πριν. Εάν εντοπίσει ότι πράγματι από εκείνο το σημείο και μετά η συμπεριφορά είναι διαφορετική, τότε χωρίζει τη χρονοσειρά σε δύο τμήματα, δηλαδή στο τμήμα πριν από το εντοπισμένο σημείο και

μετά από αυτό. Αν δεν προκύψει κάποια διαφορά στη συμπεριφορά μεταξύ των τμημάτων αυτών, τότε ο αλγόριθμος απορρίπτει οριστικά το υπόψη σημείο και προχωράει στον επόμενο έλεγχο.

Αυτή η διαδικασία επαναλαμβάνεται για όλα τα σημεία της χρονοσειράς, με αποτέλεσμα να δημιουργούνται πολλαπλά τμήματα, εντός των οποίων οι τιμές είναι σχετικά σταθερές, αλλά το κάθε τμήμα διαφέρει σε κάποιο βαθμό από το γειτονικό του. Ωστόσο, αυτή η προσέγγιση θα μπορούσε να οδηγήσει σε υπερβολικό διαχωρισμό της χρονοσειράς, δημιουργώντας ακόμη και πολύ μικρά τμήματα ένα για κάθε παρατήρηση. Για το λόγο αυτό ο αλγόριθμος χρησιμοποιεί ένα βαθμό ποινής, ο οποίος καθορίζει το πόσο σημαντική πρέπει να είναι μια διαφορά, μεταξύ τμημάτων για να θεωρηθεί πραγματικό το σημείο που τα χωρίζει. Έτσι, εντοπίζει μόνο τα ουσιαστικά σημεία μεταβολής, περιορίζοντας ταυτόχρονα και το πλήθος των τμημάτων που έχουν διαχωριστεί.

Όσον αφορά τη συνάρτηση κόστους, επιλέξαμε να χρησιμοποιήσουμε την ‘costRbf’, από τη βιβλιοθήκη της python ‘ruptures’ [90], η οποία είναι μη-παραμετρική και μπορεί να ανιχνεύει μη-γραμμικές μεταβολές στη μέση τιμή, τη διακύμανση ή και στη συνολική κατανομή των δεδομένων.

Για τον υπολογισμό της βέλτιστης τιμής ποινής (β), χρησιμοποιήσαμε μια τροποποιημένη εκδοχή του Bayesian Information Criterion (BIC) που προτάθηκε από τους Nancy R. Zhang και David O. Siegmund το 2007, στο άρθρο τους ‘A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data’ [91]. Συγκεκριμένα, χρησιμοποιήσαμε την εκδοχή αυτή ως κριτήριο επιλογής της τιμής της ποινής, μέσω ελαχιστοποίησης της σχέσης,

$$mBIC = Cost + k \log(n) + \gamma k \log(k)$$

όπου,

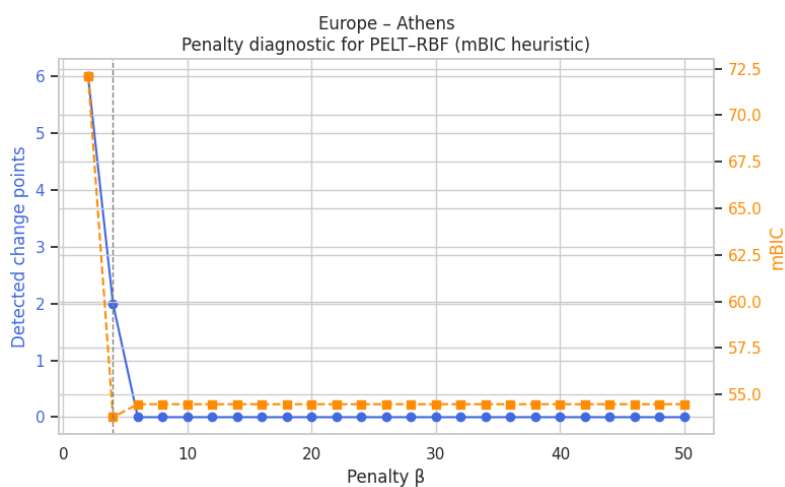
$Cost$, είναι το άθροισμα των ενδοτμηματικών σφαλμάτων,

k , ο αριθμός των σημείων αλλαγών (change points),

n , είναι το μήκος της χρονοσειρά και

γ , είναι η σταθερά που προσδιορίζει το πόσο αυστηρός θα είναι ο αλγόριθμος, ώστε να αποτραπεί η υπερβολική ανίχνευση μεταβολών. Εμπειρικά, επιλέξαμε την τιμή 1.5 που κατά την εκτίμησή μας αποτελεί μία ισορροπημένη επιλογή, αγνοώντας πολύ μικρές διακυμάνσεις, αλλά όχι σε βαθμό που να χάνει τις πραγματικά σημαντικές αλλαγές.

Ενδεικτικά, για την πόλη της Αθήνας, παραθέτουμε το κάτωθι γράφημα όπου απεικονίζεται η διαδικασία εύρεσης της κατάλληλης τιμής ποινής,



Γράφημα 38: Διαδικασία εύρεσης τιμής ποινής για τον αλγόριθμο PELT, για την πόλη της Αθήνας.

Σύμφωνα με το ανωτέρω διάγραμμα, παρατηρούμε πως για πολύ μικρές τιμές ποινής ($\beta < 3$) ο αλγόριθμος ανιχνεύει υπερβολικά μεγάλο αριθμό μεταβολών (έως και 6), γεγονός που υποδηλώνει υπερπροσαρμογή. Αντίθετα, για τιμές β μεγαλύτερες του 6 δεν ανιχνεύεται καμία μεταβολή. Η ελάχιστη τιμή του mBIC εντοπίζεται

περίπου στο 4 (για την ακρίβεια 3.986), υποδεικνύοντας πως αυτό είναι και το βέλτιστο σημείο ισορροπίας μεταξύ ευαισθησίας και σταθερότητας του αλγορίθμου.

Πρόσθετα σημειώνουμε ότι για τον υπολογισμό των κυλιόμενων κλίσεων χρησιμοποιήσαμε χρονικό παράθυρο 24 μηνών, ώστε να εξασφαλίζει επαρκή αριθμό παρατηρήσεων, διατηρώντας παράλληλα υψηλή χρονική ανάλυση στην ανίχνευση στατιστικά σημαντικών μεταβολών της τάσης.

Επικουρικά, για την αξιοπιστία των μεταβολών που εντοπίστηκαν από τον αλγόριθμο PELT, εφαρμόσαμε το στατιστικό έλεγχο Mann-Whitney U [92], ο οποίος συγκρίνει τις κατανομές των τιμών πριν και μετά από κάθε σημείο αλλαγής, χωρίς να προϋποθέτει κανονικότητα των δεδομένων και εξετάζει αν οι δύο ομάδες προέρχονται από τον ίδιο πληθυσμό. Εάν η διαφορά μεταξύ των δύο ομάδων είναι στατιστικά σημαντική ($p < 0.05$), τότε το σημείο αυτό χαρακτηρίζεται ως πραγματική στατιστικά επιβεβαιωμένη μεταβολή και διατηρείται στα αποτελέσματά μας.

Πέραν του ελέγχου σημαντικότητας, κατασκευάσαμε κι έναν στατιστικά σταθμισμένο δείκτη συνέπειας κατεύθυνσης των μεταβολών Consistency Weighted Index (CWI), που βασίζεται στον πιο απλό δείκτη Net Direction Index (NDI) και υπολογίζει τη συνολική συνέπεια και κατεύθυνση των μεταβολών, λαμβάνοντας υπόψη όχι μόνο την κατεύθυνση, δηλαδή αν είναι ανοδική ή καθοδική, αλλά πρόσθετα, την ένταση της μεταβολής και τη στατιστική της σημαντικότητα (μέσω p-value). Εφαρμόζεται δε μόνο στις στατιστικά επιβεβαιωμένες περιπτώσεις (από την εφαρμογή του Mann-Whitney U τεστ) και βασίζεται στη συνάρτηση,

$$CWI = \frac{\sum(\text{sign}(\Delta) \cdot |\Delta| \cdot (1 - p))}{\sum(|\Delta| \cdot (1 - p))}$$

όπου, ως είσοδο έχει μία ομάδα μεταβολών (ανά έτος και ανά επίπεδο) και για κάθε μεταβολή υπολογίζει ένα βάρος $w = |\Delta| \cdot (1 - p)$, με $|\Delta|$ να αντιστοιχεί στην απόλυτη τιμή της έντασης (μεγαλύτερες αλλαγές έχουν και μεγαλύτερο βάρος) και $(1 - p)$ να είναι η σημαντικότητα που έχει προκύψει από τον έλεγχο Mann-Whitney U, όπου όσο μικρότερο είναι το p τόσο μεγαλύτερο γίνεται το $(1 - p)$. Πολλαπλασιάζοντας το βάρος με το πρόσημο της μεταβολής $\text{sign}(\Delta)$, λαμβάνουμε +1 για ανοδική και -1 για καθοδική μεταβολή.

Συνολικά, το άθροισμα των βαρών με την προσθήκη του προσήμου, διαιρούμενο από το άθροισμα των απλών βαρών, μας δίδει ένα αποτέλεσμα εντός των ορίων -1 και +1.

Οπότε, όταν το αποτέλεσμα είναι κοντά στο +1 σημαίνει ότι όλες οι μεταβολές του έτους ανά επίπεδο (πχ. πόλεις) είναι ανοδικές, με μεγάλη ένταση και υψηλή σημαντικότητα. Ομοίως, όταν το αποτέλεσμα είναι κοντά στο -1 σημαίνει ότι όλες οι μεταβολές είναι καθοδικές, επίσης με μεγάλη ένταση και υψηλή σημαντικότητα. Οι τιμές που είναι κοντά στο 0, ερμηνεύονται ως ασυνεπείς, δηλαδή έχουν μικτή κατεύθυνση και υπάρχει σχετική ισορροπία μεταξύ των ανοδικών και καθοδικών μεταβολών.

Ο λόγος που μας οδήγησε στην κατασκευή και χρήση του υπόψη δείκτη, έναντι του NDI, ήταν γιατί αρχικά αντιμετωπίσαμε το εξής πρόβλημα. Αν έχουμε 100 πόλεις, από τις οποίες οι 60 έχουν καθοδικές μεταβολές και οι 40 έχουν ανοδικές, τότε ο δείκτης NDI παράγαγε ως αποτέλεσμα -0.2 $((40 - 60) / 100 = -0.2)$. Ωστόσο, αν οι 60 πόλεις εμφανίζουν χαμηλής έντασης και μη στατιστικά σημαντικές πτωτικές μεταβολές και αντίθετα οι 40 πόλεις μεγάλης έντασης και στατιστικά πολύ σημαντικές ανοδικές μεταβολές, τότε το αποτέλεσμα του NDI είναι ξεκάθαρα παραπλανητικό. Αντίθετα, ο δείκτης CWI θα εμφανίσει θετική και πιθανότατα υψηλή τιμή γιατί λαμβάνει υπόψη του όχι μόνο τα ποσοτικά αλλά και τα ποιοτικά χαρακτηριστικά των μεταβολών.

Τα αποτελέσματα από την εφαρμογή των ανωτέρω παρατίθενται παρακάτω, συνοπτικά για το επίπεδο των πόλεων:

Europe – Slope & Level Shifts (for slope window = 24 months)

Shift Year	n Downward shifts		n Upward shifts		Overall Entities		Consistency Weighted Index	
	slope	level	slope	level	slope	level	slope	level
2017	30	1	29	7	0,144	0,02	-0,095	0,891
2018	22	9	35	30	0,139	0,083	0,29	0,504
2019	43	2	52	1	0,227	0,007	0,276	-0,67
2020	29	6	30	6	0,144	0,029	0,046	-0,173
2021	44	3	49	13	0,222	0,039	-0,039	0,816
2022	28	3	23	2	0,124	0,012	-0,17	-0,371
2023	41	3	18	7	0,144	0,024	-0,531	0,504
2024	35	1	3	43	0,093	0,107	-0,978	0,965

Βάσει των παραπάνω αποτελεσμάτων, οι όποιες μεταβολές εντοπίζονται, αφορούν σε ένα μικρό σχετικά πλήθος πόλεων και δεν εκφράζουν γενικευμένα μοτίβα σε επίπεδο Ευρώπης.

Κατά τα έτη 2017 - 2019 παρατηρούμε μία μικτή εικόνα, με τις περιπτώσεις ανοδικών αλλαγών να κυριαρχούν έναντι των καθοδικών και ταυτόχρονα μία σημαντική διάκριση μεταξύ αλλαγών στο ρυθμό μεταβολής των συγκεντρώσεων του θείου (slope shift) και του ύψος των τιμών του (level shift). Ο δείκτης συνέπειας κατεύθυνσης για τα slope shifts κυμαίνεται σε χαμηλά επίπεδα (-0.095 - 0.276) έναντι των level shifts (0.891 - 0.67) που αποκαλύπτουν σημαντικά επεισόδια απότομων μεταβολών με θετικό μάλιστα πρόσημο.

Το έτος 2020 φαίνεται να αποτελεί μία φάση σταθεροποίησης, τόσο ως προς το πλήθος των μεταβολών, όπου παρατηρούμε μία ισοκατανομή μεταξύ ανοδικών και καθοδικών, όσο και ως προς τη σημαντικότητά τους. Μάλιστα ο δείκτης συνέπειας κατεύθυνσης τόσο σε επίπεδο slope shifts όσο και σε level shifts προσεγγίζει το μηδέν, δείχνοντας μια αλληλεξουδετέρωση μεταξύ των ανοδικών και καθοδικών γεγονότων.

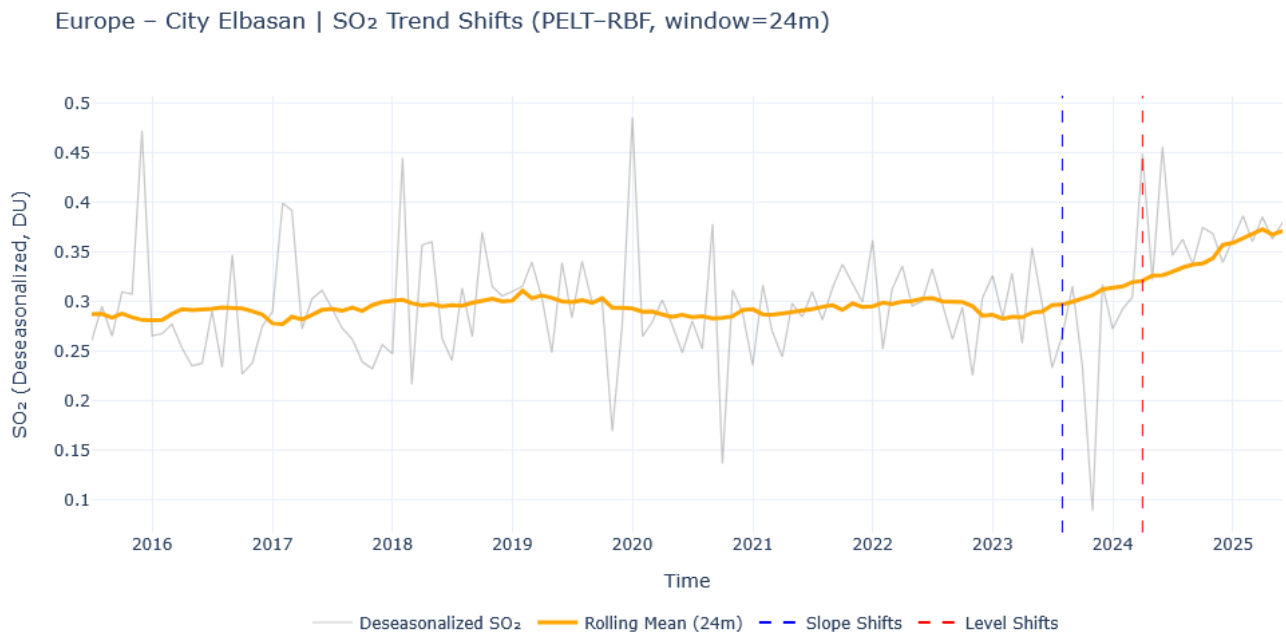
Κατά τα έτη 2021 - 2023 παρατηρούμε μία εκ νέου αύξηση του πλήθους των μεταβολών, όπου από τη μία, σε επίπεδο slope shifts, οι καθοδικές υπερτερούν των ανοδικών και από την άλλη, σε επίπεδο level shifts, οι ανοδικές είναι περισσότερες από τις καθοδικές. Ο δείκτης συνέπειας κατεύθυνσης δείχνει μία σταθερή επιβράδυνση του ρυθμού μεταβολής των συγκεντρώσεων του διοξειδίου του θείου και ταυτόχρονα μία μικτή συμπεριφορά σημαντικών ανοδικών και καθοδικών τιμών, υποδεικνύοντας πως πρόκειται για φαινόμενα τοπικού χαρακτήρα.

Το έτος 2024, είναι και το πιο ενδιαφέρον, καθώς παρατηρούμε μία σημαντική διαφοροποίηση στη συμπεριφορά των μεταβολών μεταξύ slope shifts και level shifts. Σε επίπεδο slope shifts, οι περιπτώσεις των μεταβολών με αρνητικό πρόσημο υπερτερούν σημαντικά εκείνων με θετικό και μάλιστα με πολύ υψηλό αρνητικό δείκτη συνέπειας κατεύθυνσης. Αντίθετα, σε επίπεδο level shifts παρατηρούμε σημαντική αύξηση του πλήθους των ανοδικών επεισοδίων (43 περιπτώσεις), με το δείκτη συνέπειας κατεύθυνσης να επιβεβαιώνει πως οι υπόψη μεταβολές είναι σημαντικές και παρόμοιας έντασης (+0.965). Δηλαδή, από τη μία έχουμε πόλεις με επιβράδυνση του ρυθμού μεταβολής των συγκεντρώσεων και από την άλλη πόλεις με έντονα επεισόδια αύξησης των τιμών των συγκεντρώσεων του διοξειδίου του θείου (σε αποεποχικοποιημένες χρονοσειρές).

Για να εντοπίσουμε αν πράγματι αυτά τα φαινόμενα επηρεάζουν τη συνολική τάση των χρονοσειρών των εκάστοτε πόλεων, θα πρέπει να προβούμε σε περαιτέρω ανάλυση της εν γένει συμπεριφοράς τους.

Ενδεικτικά, θα αναφερθούμε στην πόλη της Elbasan της Αλβανίας, η οποία εμφανίζει τον Απρίλιο του 2024, μία στατιστικά σημαντική (MWUtest, p-value = 0.0022) ανοδική μεταβολή, κατά 0.11 Dobson Units. Να σημειώσουμε πως η μακροχρόνια τάση της πόλης Elbasan είναι ανοδική κατά 0.0063 DU/year και η μέση τιμή των συγκεντρώσεων διοξειδίου του θείου (στην αποεποχικοποιημένη χρονοσειρά) τους 6 προηγούμενους μήνες του επεισοδίου ήταν 0.26 DU και τους 6 επόμενους ήταν 0.37 DU. Δηλαδή, δεν πρόκειται για ένα απλό γεγονός αλλά μάλλον για ένα επεισόδιο που δύναται να αλλάξει σημαντικά τη συμπεριφορά της τάσης.

Προς επίρρωση των ανωτέρω, παραθέτουμε το σχετικό διάγραμμα από όπου φαίνεται πως πράγματι, πρόκειται για ένα σημαντικό γεγονός που σηματοδοτεί την είσοδο σε μία πιο έντονη φάση αύξησης των συγκεντρώσεων διοξειδίου του θείου και μάλιστα δεν είναι το μόνο, καθώς έχει προηγηθεί άλλο επεισόδιο που φαίνεται να επηρεάζει σημαντικά το ρυθμό μεταβολής των συγκεντρώσεων διοξειδίου του θείου.

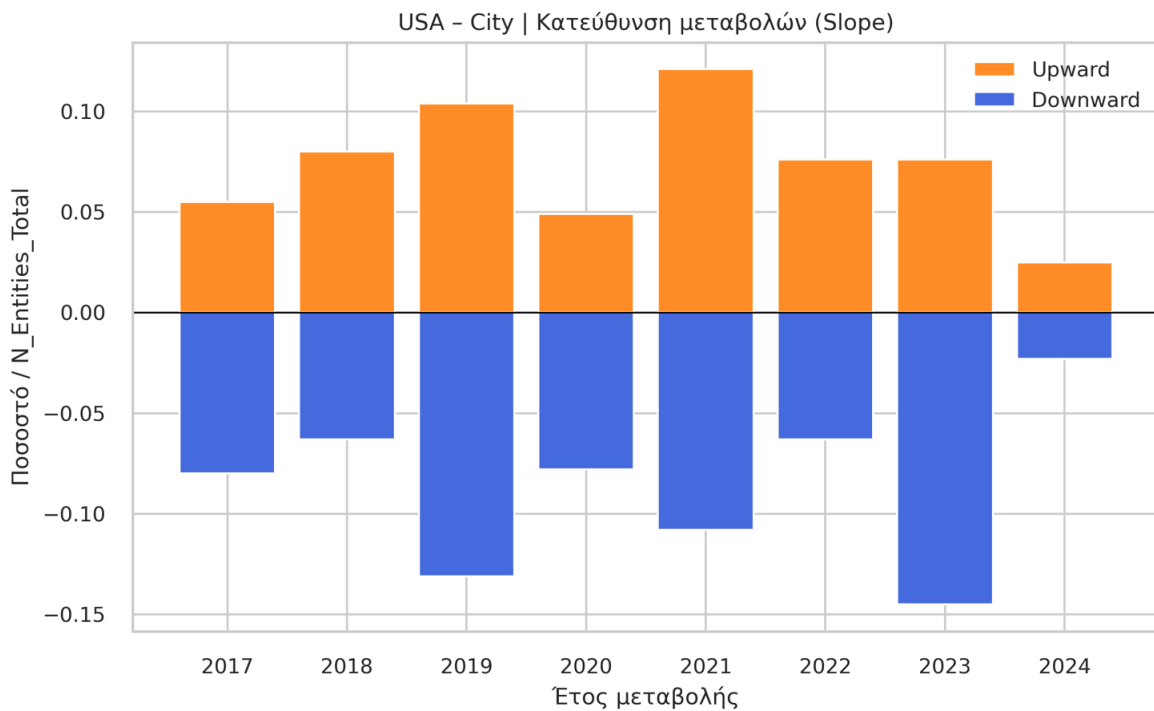


Γράφημα 39: Απεικόνιση σημαντικών γεγονότων που μετέβαλαν την τάση για την πόλη Elbasan.

USA – Slope & Level Shifts (for slope window = 24 months)

Shift Year	n Downward shifts		n Upward shifts		Overall Entities		Consistency Weighted Index	
	slope	level	slope	level	slope	level	slope	level
2017	41	6	28	10	0,135	0,031	-0,208	0,402
2018	32	8	41	10	0,143	0,035	0,325	0,041
2019	67	1	53	15	0,233	0,031	-0,269	0,901
2020	40	4	25	21	0,127	0,049	-0,222	0,72
2021	55	11	62	6	0,227	0,033	0,216	-0,378
2022	32	12	39	11	0,139	0,045	0,233	0,048
2023	74	6	39	21	0,221	0,053	-0,409	0,48
2024	12	4	13	39	0,049	0,084	-0,1	0,868

Στις πόλεις των ΗΠΑ, εντοπίζονται περισσότερες και διάσπαρτες χρονικά μεταβολές, σε σχέση με την Ευρώπη και γενικά παρατηρούμε μία πιο έντονη αλλά και ετερογενή δυναμική μεταβολών, που όμως συμμετέχουν λίγες πόλεις και σίγουρα δεν πρόκειται για κάποιο γενικευμένο μοτίβο. Θα λέγαμε ότι η γενική εικόνα των μεταβολών αλλάζει από χρόνο σε χρόνο με διαφορετικό πρόσημο και δυναμική κάθε φορά, ιδιαίτερα ως προς το πλήθος των περιπτώσεων που αφορούν στο ρυθμό μεταβολής των συγκεντρώσεων διοξειδίου του θείου. Προς τούτο παραθέτουμε σε διαγραμματική μορφή την εξέλιξη του πλήθους των μεταβολών στη διάσταση του χρόνου:



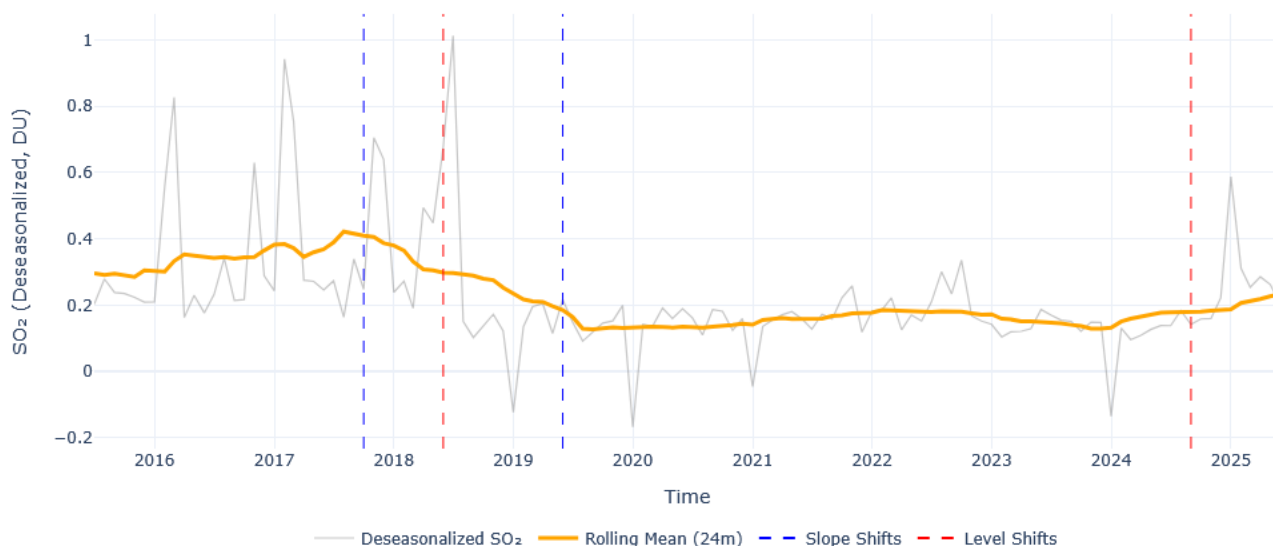
Γράφημα 40: Συνολική κατεύθυνση μεταβολών τάσης για τις ΗΠΑ.

όπου φαίνεται ξεκάθαρα ότι χρόνο με το χρόνο αλλάζει συστηματικά το πλήθος των μεταβολών ακολουθώντας μία κυκλική συμπεριφορά.

Παρόμοια εικόνα παρατηρούμε και με τις τιμές του δείκτη συνέπειας κατεύθυνσης, όπου τόσο σε επίπεδο slope shifts όσο και σε level shifts, παρουσιάζει μικτή εικόνα με σημαντικές αυξομειώσεις μεταξύ των ετών. Ειδικότερα τα έτη 2019 και 2024 χαρακτηρίζονται από σχεδόν μηδενικές τιμές του δείκτη συνέπειας κατεύθυνσης για τις μεταβολές στις κλίσεις, αλλά ταυτόχρονα εξαιρετικά υψηλές θετικές τιμές για τις μεταβολές στις μέσες τιμές συγκεντρώσεων διοξειδίου του θείου.

Μία χαρακτηριστική περίπτωση που συνδυάζει την ανωτέρω περιγραφόμενη συμπεριφορά είναι η πόλη Hilo της Χαβάης, στην οποία αναφερθήκαμε και στην αρχή της παρούσας ενότητας, κατά τη βασική ανάλυση της τάσης.

Όπως φαίνεται και στο παρακάτω σχεδιάγραμμα,



Γράφημα 41: Απεικόνιση σημαντικών γεγονότων που μετέβαλαν την τάση για την πόλη Hilo.

κατά την περίοδο 2018 - 2019 παρατηρούμε μία απότομη και μόνιμη μείωση της μέσης τιμής της αποεποχικοποιημένης χρονοσειράς, η οποία συνοδεύεται και από σχετική αλλαγή στο ρυθμό μεταβολής των συγκεντρώσεων διοξειδίου του θείου. Η χρονική αυτή μεταβολή συμπίπτει με τη λήξη της έντονης ηφαιστειακής δραστηριότητας του ηφαιστείου Kīlauea, που όπως ήδη έχουμε αναφέρει, τον Αύγουστο του 2018, σταμάτησε απότομα η ηφαιστειακή δραστηριότητα και ο κρατήρας Halema'ūma'ū κατέρρευσε, με αποτέλεσμα οι εκπομπές διοξειδίου του θείου να μειωθούν κατακόρυφα. Στις 23 Δεκεμβρίου του 2024, μετά από αύξηση της σεισμικής δραστηριότητας, έγινε μια έκρηξη στην καλντέρα της κορυφής, η οποία περιλάμβανε εκτοξεύσεις λάβας που κάλυψαν μεγάλο μέρος του πυθμένα του κρατήρα Halema'ūma'ū. Η δραστηριότητα σταμάτησε και επανεκκίνησε αρκετές φορές τις επόμενες ημέρες, με τις εκτοξεύσεις λάβας να ξεπερνούν τα 80 μέτρα.

Αυτή λοιπόν η εικόνα, δηλαδή, η απότομη παύση της ηφαιστειακής δραστηριότητας το 2018 και η επανεκκίνηση το 2024 αποτυπώνονται εμφανώς στο παραπάνω σχεδιάγραμμα.

Καταλήγοντας, τα αποτελέσματα της ανάλυσής μας έδειξαν ότι τόσο στην Ευρώπη όσο στις ΗΠΑ, σημαντικές αλλαγές στη συμπεριφορά των συγκεντρώσεων του διοξειδίου του θείου εμφανίζονται σε σχετικά μικρό αριθμό πόλεων. Πρόσθετα, δεν παρατηρούνται χωρικά συνεκτικά ή γενικευμένα μοτίβα, αλλά μεμονωμένα επεισόδια μεταβολών, συνήθως βραχύχρονης διάρκειας και τοπικού χαρακτήρα. Ο δείκτης συνέπειας κατεύθυνσης, αναφορικά με τις μεταβολές στις μέσες τιμές των συγκεντρώσεων διοξειδίου του θείου, παρουσιάζει σε ορισμένα έτη υψηλές θετικές τιμές, ενώ παράλληλα ο ίδιος δείκτης για το ρυθμό μεταβολής είναι χαμηλός ή αρνητικός. Αυτό υποδηλώνει ότι σε κάποιες πόλεις εμφανίστηκαν έκτακτα γεγονότα τοπικής κλίμακας, με άλλοτε παροδικές επιδράσεις και άλλοτε πιο παρατεταμένες, όπως άλλωστε παρουσιάσαμε για δύο πόλεις, μία από κάθε ήπειρο.

4.7 Ανάλυση καταλοίπων αποσύνθεσης χρονοσειρών

Στόχος της ανάλυσης των καταλοίπων, δηλαδή των βραχυπρόθεσμων έκτακτων διακυμάνσεων που δεν εξηγούνται από την εποχικότητα και την τάση, είναι να αξιολογηθεί η επάρκεια της διαδικασίας αποσύνθεσης, που εφαρμόστηκε στις χρονοσειρές των συγκεντρώσεων διοξειδίου του θείου και να διαπιστωθεί αν τα κατάλοιπα προσεγγίζουν τη συμπεριφορά του λευκού θορύβου (white noise). Στην αρχή της ενότητας της τεχνικής διαδικασίας αποσύνθεσης, ασχοληθήκαμε εκτενώς με το έλεγχο των καταλοίπων και μάλιστα αναφερθήκαμε και σε συγκεκριμένα παραδείγματα ερμηνείας και αξιολόγησής τους. Συνεπώς, παρακάτω θα παραθέσουμε τη μεθοδολογία που ακολουθήσαμε και συνοπτικά τα αποτελέσματα αυτής.

4.7.1 Υπολογισμός στατιστικών δεικτών καταλοίπων

Όπως περιγράψαμε εκτενώς σε προηγούμενη ενότητα, στις χρονοσειρές των οντοτήτων του δείγματός μας εφαρμόστηκαν οι διαγνωστικοί έλεγχοι: α) Ljung-Box test, όπου εξετάζει αν τα κατάλοιπα παρουσιάζουν αυτοσυσχέτιση, β) στασιμότητας Augmented Dickey - Fuller (ADF), όπου εξετάει αν τα κατάλοιπα έχουν μοναδιαία ρίζα, δηλαδή αν η μέση τιμή και η διακύμανσή τους μεταβάλλονται στο χρόνο και γ) του δείκτη Durbin-Watson που αξιολογεί την αυτοσυσχέτιση πρώτης τάξης στα κατάλοιπα, δηλαδή τη συσχέτιση της κάθε τιμής με την αμέσως προηγούμενή της.

Για την καλύτερη ερμηνεία των αποτελεσμάτων ταξινομήσαμε τις μετρικές των ελέγχων σε 3 κατηγορίες, ήτοι,

Clean: πλήρως επαρκής αποσύνθεση, δηλαδή τα κατάλοιπα αποτελούν λευκό θόρυβο,

Moderate: μερική επάρκεια αποσύνθεσης, όπου εμφανίζονται ήπιες υπολειπόμενες δομές και

Poor: ύπαρξη συστηματικών μοτίβων ή αυτοσυσχέτισης.

Η κατάταξη ανά χωρικό επίπεδο, παρατίθενται αμέσως παρακάτω,

Europe

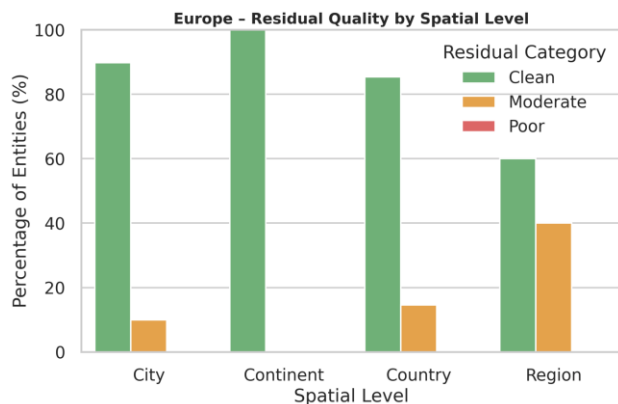
Level	n	Clean	Moderate	Poor
city	410	368	41	1
country	41	35	6	0
region	5	3	2	0
continent	1	1	0	0

USA

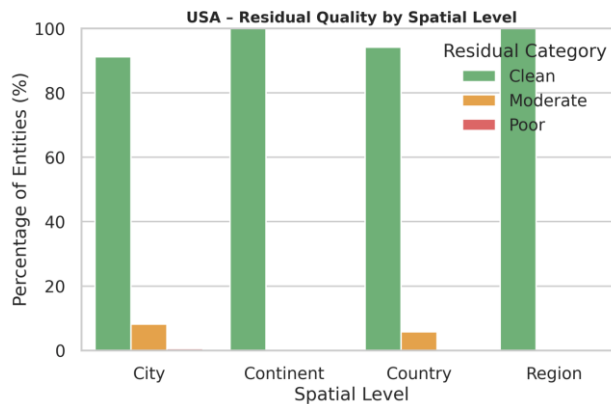
Level	n	Clean	Moderate	Poor
city	511	466	42	3
state	52	49	3	0
region	5	5	0	0
continent	1	1	0	0

Η συντριπτική πλειοψηφία των οντοτήτων, τόσο στην Ευρώπη όσο και στις ΗΠΑ ταξινομήθηκε στην κατηγορία ‘Clean’, γεγονός που υποδηλώνει ότι η διαδικασία της αποσύνθεσης απομόνωσε επιτυχώς την τάση και την εποχικότητα. Στην Ευρώπη, εμφανίζονται 41 πόλεις και 6 χώρες με ήπιες υπολειπόμενες δομές, ενώ μόλις 1 πόλη χαρακτηρίστηκε ως ‘Poor’. Στις ΗΠΑ, 42 πόλεις και 3 πολιτείες εμφάνισαν ήπιες υπολειπόμενες δομές μετά την αποσύνθεση, ενώ 3 εμφάνισαν μοτίβα που ίσως θα έπρεπε να έχουν εξηγηθεί από τις συνιστώσες της εποχικότητας ή τάσης.

Οπτικά, φαίνονται ακόμη πιο ξεκάθαρα, τα αποτελέσματα των στατιστικών ελέγχων,



Γράφημα 42: Στατιστικοί έλεγχοι καταλοίπων, ανά χωρικό επίπεδο Ευρώπης.



Γράφημα 43: Στατιστικοί έλεγχοι καταλοίπων, ανά χωρικό επίπεδο ΗΠΑ.

όπου κυριαρχούν οι πράσινες μπάρες που υποδηλώνουν την απουσία αυτοσυσχέτισης στα κατάλοιπα.

Τέλος, επισημαίνουμε ότι, πέραν των αριθμητικών δεικτών και όπως περιγράψαμε στην αρχή της ενότητας της διαδικασίας αποσύνθεσης, πραγματοποιήθηκε και οπτικός έλεγχος των καταλοίπων μέσω διαγνωστικών διαγραμμάτων, ήτοι,

- ✓ Residuals over Time, για τυχόν μοτίβα ή συστηματικές διακυμάνσεις,
- ✓ Histogram + KDE, για εξέταση συμμετρίας και κανονικότητας,
- ✓ Q-Q Plot, για σύγκριση της εμπειρικής με την θεωρητική κανονική κατανομή [81],
- ✓ ACF / PACF, για έλεγχο ύπαρξης αυτοσυσχέτισης ή μερικής αυτοσυσχέτισης αντίστοιχα και
- ✓ Residuals vs Fitted (Trend + Seasonal), για την ανίχνευση τυχόν σχέσης μεταξύ καταλοίπων και των συνιστωσών εποχικότητας και τάσης.

4.7.2 Ανάλυση κατανομής και ακραίων τιμών καταλοίπων

Σε όλες τις οντότητες (πλην εξαιρέσεων που θα παραθέσουμε παρακάτω), πραγματοποιήθηκε περαιτέρω ανάλυση της κατανομής των καταλοίπων, με σκοπό τον εντοπισμό τυχόν ακραίων αποκλίσεων και την εκτίμηση του βαθμού συμμετρίας και συγκέντρωσης της κατανομής τους. Πιο συγκεκριμένα προβήκαμε σε έλεγχο κανονικότητας Shapiro-Wilk [93], προκειμένου να διαπιστώσουμε αν τα κατάλοιπα ακολουθούν την κανονική κατανομή ή αποκλίνουν από αυτή. Παρότι, η κανονικότητα των καταλοίπων δεν αποτελεί προϋπόθεση για την επιτυχή αποσύνθεση των χρονοσειρών, η εξέτασή της μας παρέχει πληροφορίες για τη μορφή της κατανομής τους και τη συχνότητα εμφάνισης τυχαιών ακραίων γεγονότων.

Ο υπόψη έλεγχος, εξετάζει κατά πόσο η κατανομή ενός δείγματος προσεγγίζει την κανονική, συγκρίνοντας τις ταξινομημένες τιμές του δείγματος με τις αναμενόμενες τιμές μιας θεωρητικής κανονικής κατανομής [81], ιδίου μεγέθους.

Η στατιστική του ελέγχου υπολογίζεται ως:

$$W = \frac{\left(\sum_{i=1}^n a_i y_{(i)}\right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Όπου $y_{(i)}$ είναι οι ταξινομημένες παρατηρήσεις του δείγματός μας, \bar{y} ο μέσος όρος τους και a_i οι συντελεστές που υπολογίζονται από τις αναμενόμενες τιμές και τη συνδιακύμανση μιας κανονικής κατανομής.

Η τιμή του στατιστικού W κυμαίνεται μεταξύ 0 και 1, όπου τιμές κοντά στο 1 υποδηλώνουν υψηλό βαθμό προσαρμογής στην κανονική κατανομή, ενώ χαμηλές τιμές υποδηλώνουν σημαντική απόκλιση από αυτήν. Το αντίστοιχο p-value εκτιμάται με βάση την κατανομή του W όπου αν $p\text{-value} < 0.05$, η μηδενική υπόθεση H_0 , δηλαδή ότι τα δεδομένα προέρχονται από κανονική κατανομή, απορρίπτεται, οπότε η κατανομή των δεδομένων διαφέρει στατιστικά σημαντικά από την κανονική.

Για την καλύτερη ερμηνεία των αποτελεσμάτων καταρτίσαμε τις παρακάτω κατηγορίες, συνδυάζοντας τόσο την τιμή του W όσο και του p-value. Στο σημείο αυτό να σημειώσουμε ότι δεν εντοπίσαμε στη βιβλιογραφία παρόμοια κατηγοριοποίηση, ωστόσο, αυτή μας βοήθησε να ερμηνεύσουμε καλύτερα τα αποτελέσματα της εφαρμογής του παραπάνω ελέγχου. Αναλυτικότερα,

Perfect Normality: ($W \geq 0.99$ και $p > 0.10$), όπου τα κατάλοιπα ακολουθούν απόλυτα την κανονική κατανομή, χωρίς ενδείξεις απόκλισης,

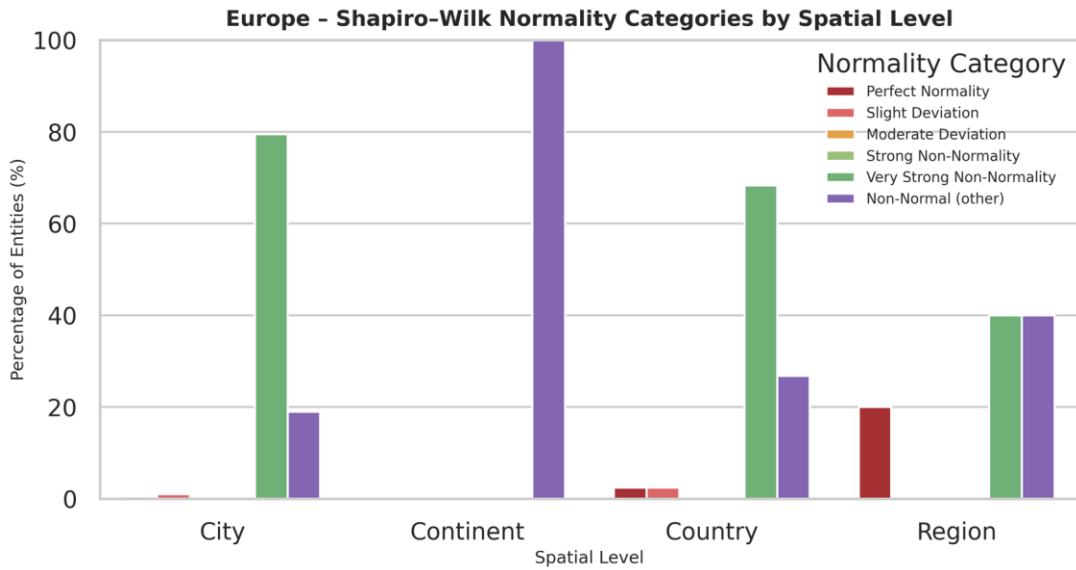
Slight Deviation: ($0.97 \leq W < 0.99$ και $0.05 < p < 0.10$), όπου τα κατάλοιπα παρουσιάζουν ελαφριά απόκλιση από την κανονικότητα, η οποία όμως θεωρείται στατιστικά αμελητέα,

Moderate Deviation: ($0.94 \leq W < 0.97$ και $0.01 < p < 0.05$), για τις περιπτώσεις όπου παρατηρείται μέτρια απόκλιση από την κανονική κατανομή, πιθανών λόγω κάποιων ακραίων τιμών ή μικρής ασυμμετρίας,

Strong or Very Strong Non-Normality: ($W < 0.94$ και $p < 0.001$), όπου υποδεικνύει σαφής απουσία κανονικότητας στα κατάλοιπα, με έντονη ασυμμετρία ή ακόμη και μεγάλη συγκέντρωση ακραίων τιμών και

Non-Normal (other): αφορά σε περιπτώσεις όπου οι τιμές του στατιστικού W και του p-value δεν εμπίπτουν σε κάποια από τις άλλες κατηγορίες και εμφανίζουν μικρή ασυμφωνία μεταξύ των δύο κριτηρίων (W και p-value).

Τα υπόψη αποτελέσματα, επιλέξαμε να τα παρουσιάσουμε μέσω σχετικών διαγραμμάτων. Αναλυτικότερα,

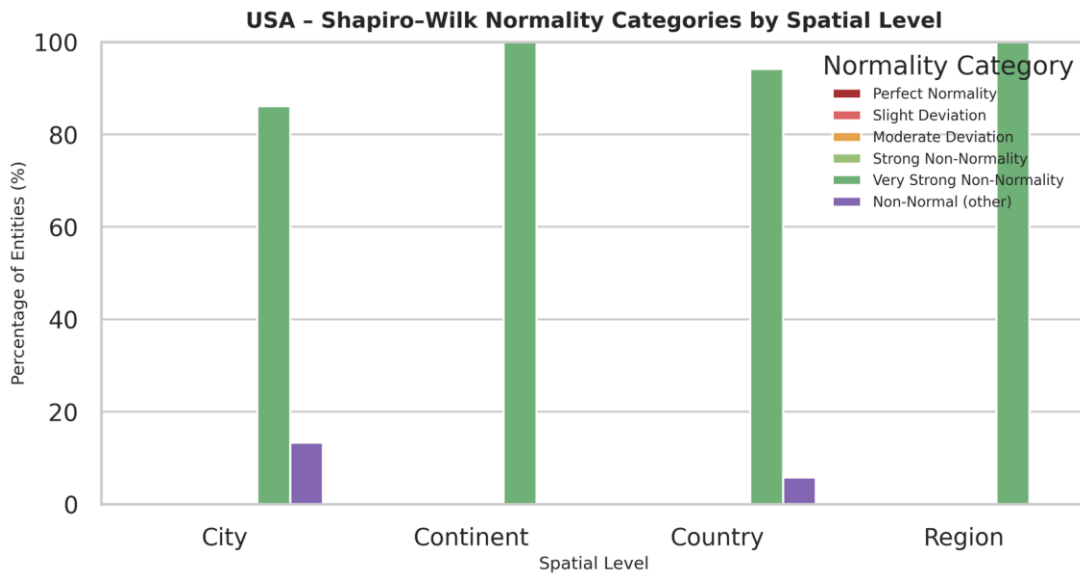


Γράφημα 44: Ταξινόμηση αποτελεσμάτων ανά χωρικό επίπεδο Ευρώπης, από την εφαρμογή Shapiro-Wilk normality test.

Στην Ευρώπη, παρατηρούμε πως το μεγαλύτερο πλήθος των οντοτήτων, εμφανίζουν ασυμμετρίες ή ακραίες τιμές στα κατάλοιπα, κάτι αναμενόμενο άλλωστε για περιβαλλοντικά δεδομένα, όπου πιθανότατα προέρχονται από έκτακτα επεισόδια φυσικών ή ανθρωπογενών παραγόντων.

Να σημειώσουμε εδώ, ότι οι υπολογισμοί στα επίπεδα κρατών, περιφερειών και ηπείρου, εφαρμόστηκαν στις αντίστοιχες χρονοσειρές που διαμορφώθηκαν από τα σταθμισμένα μέσα των μηνιαίων χρονοσειρών των πόλεων. Συνεπώς, περιλαμβάνουν εξομαλυμένες τιμές.

Στις ΗΠΑ, η εικόνα είναι πιο σταθερή μεταξύ των οντοτήτων,



Γράφημα 45 Ταξινόμηση αποτελεσμάτων ανά χωρικό επίπεδο των ΗΠΑ, από την εφαρμογή Shapiro-Wilk normality test.

όπου παρατηρούμε πως περίπου το 90% να μην εμφανίζει ενδείξεις κανονικότητας στα κατάλοιπα. Το γεγονός αυτό, υποδηλώνει την ύπαρξη περισσότερων ακραίων γεγονότων και βραχυπρόθεσμων επεισοδίων ρύπανσης έναντι όσων παρατηρήσαμε στην Ευρώπη. Συνδυαστικά και με όσα αναφέραμε σε προηγούμενες ενότητες, οι ΗΠΑ χαρακτηρίζονται από μία κοινή συμπεριφορά μεταξύ όλων των οντοτήτων, κάτι που αποτυπώνεται εμφαντικά και εδώ.

Ο έλεγχος κανονικότητας Shapiro-Wilk ανέδειξε ότι τα κατάλοιπα των χρονοσειρών διοξειδίου του θείου παρουσιάζουν σημαντικές αποκλίσεις από την κανονική κατανομή. Για να διερευνηθεί περαιτέρω η φύση αυτής της μη κανονικότητας, καταφύγαμε στην εύρεση των ακραίων τιμών (outliers) και στον υπολογισμό των περιγραφικών δεικτών συμμετρίας και κυρτότητας (Skewness και Kurtosis).

Ο δείκτης Outlier Ratio εκφράζει το ποσοστό των καταλοίπων που απέχουν περισσότερο από το κατάφλι που θέσαμε, ήτοι, τριών τυπικών αποκλίσεων από το μέσο όρο τους, δηλαδή, πόσες φορές οι τιμές των καταλοίπων ξεπερνούν σε μέγεθος τις τρεις τυπικές αποκλίσεις από το μέσο όρο των τιμών των καταλοίπων της χρονοσειράς [94]. Ο υπόψη δείκτης, μας δίνει μια σημαντική πληροφορία για τη συχνότητα με την οποία εμφανίζονται έκτακτα ακραία γεγονότα, όπως για παράδειγμα μία έκρηξη ηφαιστείου ή μία βραχυπρόθεσμη σημαντική αύξηση των συγκεντρώσεων διοξειδίου του θείου, εξαιτίας κάποιου έκτακτου φαινομένου. Συνήθως, τιμές μέχρι 1% θεωρούνται εντός φυσιολογικών ορίων υποδεικνύοντας την ύπαρξη περιορισμένου πλήθους έκτακτων γεγονότων.

Ο συντελεστής ασυμμετρίας (Skewness) ποσοτικοποιεί το βαθμό και την κατεύθυνση ασυμμετρίας της κατανομής των καταλοίπων ως προς το μέσο όρο τους. Σε μια τέλεια συμμετρική κατανομή, οι θετικές και αρνητικές αποκλίσεις εμφανίζονται με ίση συχνότητα και ένταση, οπότε Skewness = 0. Αν η κατανομή παρουσιάζει μεγάλη συγκέντρωση τιμών κάτω από το μέσο όρο των καταλοίπων και μερικές μεγάλες θετικές αποκλίσεις, δηλαδή έκτακτα επεισόδια συγκεντρώσεων διοξειδίου του θείου, τότε Skewness > 0 (θετικά ασύμμετρα). Στην αντίθετη περίπτωση, δηλαδή αν οι περισσότερες τιμές συγκεντρώνονται πάνω από το μέσο όρο και υπάρχουν κάποιες έντονες αρνητικές αποκλίσεις, δηλαδή αιφνίδια μείωση διοξειδίου του θείου, τότε Skewness < 0 (αρνητικά ασύμμετρα), κάτι που σε φαινόμενο δεν το συναντάμε συχνά.

Η μαθηματική του μορφή [95] είναι,

$$Skewness = \frac{\frac{1}{n} \sum (Y_i - \bar{Y})^3}{\left(\frac{1}{n} \sum (Y_i - \bar{Y})^2\right)^{3/2}}$$

όπου,

Y_i , είναι τα κατάλοιπα της χρονοσειράς,

\bar{Y} , ο μέσος των καταλοίπων και

n το πλήθος των παρατηρήσεων.

Ο συντελεστής κυρτότητας (Kurtosis) ποσοτικοποιεί το βαθμό συγκέντρωσης των τιμών γύρω από το μέσο όρο και τη συχνότητα εμφάνισης ακραίων αποκλίσεων, σε σύγκριση με τη θεωρητική κανονική κατανομή. Θετικές τιμές (leptokurtic), υποδεικνύουν υψηλή συγκέντρωση των καταλοίπων κοντά στο μέσο όρο και σπάνια εμφάνιση αλλά ιδιαίτερης έντασης ακραίων τιμών. Δηλαδή, η χρονοσειρά παρουσιάζει σταθερή συμπεριφορά το μεγαλύτερο διάστημα, με σπάνια επεισόδια έντονων όμως διακυμάνσεων. Αντίθετα, αρνητικές τιμές (platykurtic) υποδηλώνουν πιο ομοιόμορφη κατανομή των αποκλίσεων, χωρίς την παρουσία έντονων φαινομένων, αλλά λιγότερες συγκεντρωμένες τιμές γύρω από το μέσο όρο των καταλοίπων.

Μαθηματικά [95], ορίζεται ως,

$$Kurtosis = \frac{\frac{1}{n} \sum (Y_i - \bar{Y})^4}{\left(\frac{1}{n} \sum (Y_i - \bar{Y})^2\right)^2} - 3$$

όπου,

Y_i είναι οι παρατηρήσεις ή τα κατάλοιπα στην περίπτωσή μας,

\bar{Y} είναι το μέσο των καταλοίπων και

n το πλήθος των παρατηρήσεων.

Από την εφαρμογή των ανωτέρω συντελεστών προέκυψαν τα ακόλουθα,

Europe

Region	Mean Outlier	Mean Skewness	Mean Kurtosis
East	0.0233	0.91	5.85
Middle	0.0228	0.93	5.71
North	0.0188	1.12	8.37
South	0.0188	1.08	7.30
West	0.0257	0.78	5.10

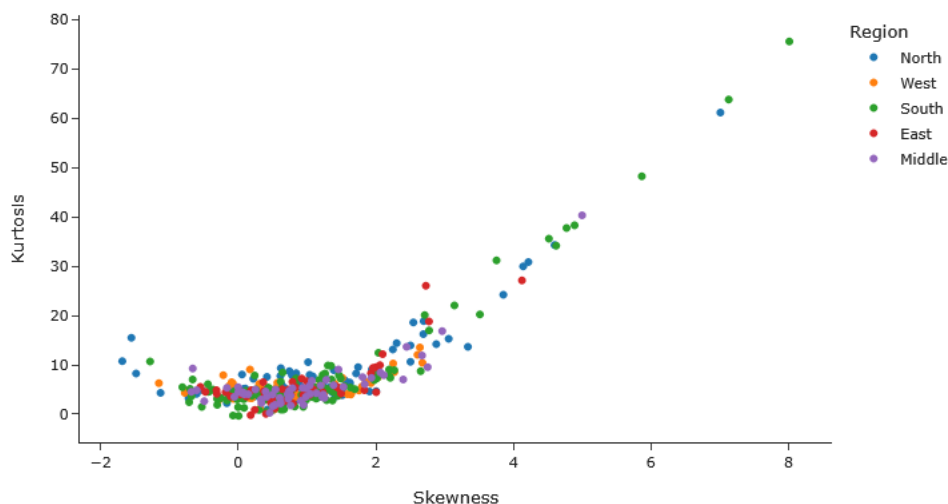
USA

Region	Mean Outlier	Mean Skewness	Mean Kurtosis
East	0.0244	0.94	7.00
Middle	0.0239	0.74	5.71
North	0.0214	1.21	7.91
South	0.0166	1.39	11.15
West	0.0111	3.10	24.66

Στις Ευρωπαϊκές περιφέρειες, οι μέσες τιμές του δείκτη Outlier Ratio κυμαίνονται μεταξύ 1,88 % έως 2,57 %, υπερβαίνουν δηλαδή το θεωρητικά φυσιολογικό 1% που θα αναμενόταν για μία κανονική κατανομή. Αυτό σημαίνει συχνότερη εμφάνιση ακραίων τιμών από ότι θα παρατηρούσαμε σε μία κανονική κατανομή και ιδίως στις πόλεις της Δυτικής περιφέρειας. Οι θετικές τιμές του δείκτη Skewness (0.78 - 1.12), δείχνουν ότι οι κατανομές παρουσιάζουν μία μέτρια δεξιά ασυμμετρία, δηλαδή μεγαλύτερη συχνότητα θετικών αποκλίσεων, ενώ οι τιμές του δείκτη κυρτότητας, υποδηλώνουν την ύπαρξη περιστασιακών έντονων επεισοδίων, με πιο συχνή εμφάνιση στη Βόρεια και Νότια περιφέρεια.

Στο κάτωθι γράφημα παρουσιάζεται η σχέση μεταξύ ασυμμετρίας και κυρτότητας των καταλοίπων στις πόλεις της Ευρώπης, ανά περιφέρεια,

SO: Residual Distribution Shape (Skewness vs Kurtosis) – Cities (Europe)



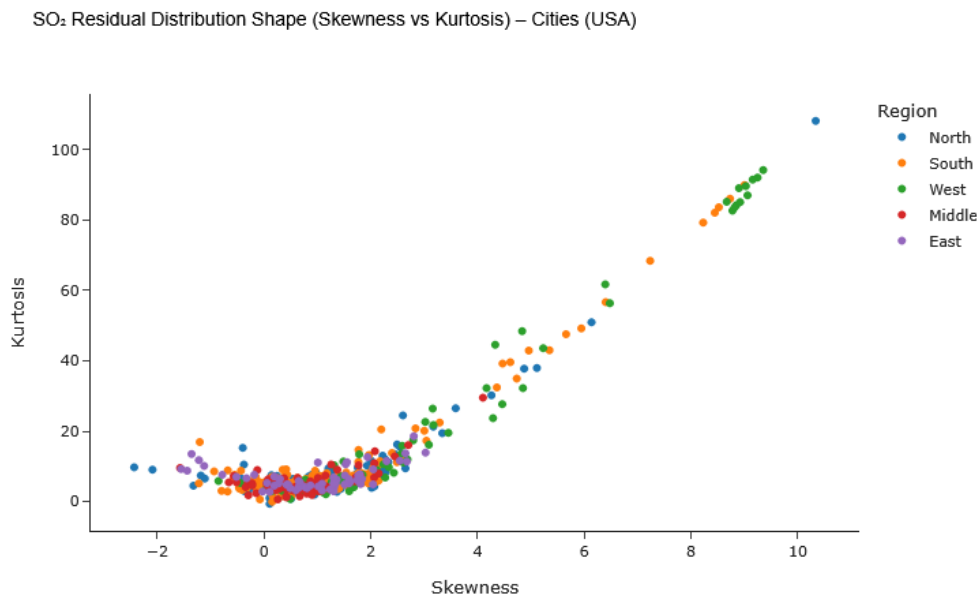
Γράφημα 46: Συσχέτιση ασυμμετρίας και κυρτότητας στα κατάλοιπα των Ευρωπαϊκών πόλεων ανά περιφέρεια.

από όπου, παρατηρούμε ότι στο μεγαλύτερο πλήθος των Ευρωπαϊκών πόλεων υπάρχει μία θετική σχέση μεταξύ ασυμμετρίας και κυρτότητας, δηλαδή, όσο αυξάνεται η θετική ασυμμετρία των καταλοίπων τόσο αυξάνεται και η κυρτότητα. Αυτό σημαίνει ότι οι πόλεις που εμφανίζουν περισσότερες θετικές αποκλίσεις από το 0, είναι επίσης εκείνες όπου καταγράφουν λίγα αλλά ισχυρότερα μεμονωμένα επεισόδια.

Στις ΗΠΑ, οι μέσες τιμές του Outlier Ratio κυμαίνονται από 1,1 % έως 2,4 %, υπερβαίνοντας και εδώ το θεωρητικά φυσιολογικό 1%, που θα αναμενόταν για μία κανονική κατανομή [81]. Αντίθετα με ότι παρατηρήσαμε στην Ευρώπη, η περιφέρεια με τη συχνότερη εμφάνιση ακραίων περιστατικών είναι η Ανατολική. Οι συνολικά θετικές τιμές του δείκτη Skewness (0.74 - 3.10), δείχνουν ότι οι κατανομές παρουσιάζουν μία μέτρια δεξιά ασυμμετρία, δηλαδή μεγαλύτερη συχνότητα θετικών αποκλίσεων. Η Δυτική περιφέρεια, από τη μία εμφανίζει τα λιγότερες ακραίες τιμές (0.111), από την άλλη τους υψηλότερους δείκτες ασυμμετρίας και κυρτότητας. Το γεγονός αυτό, υποδηλώνει ότι τα ακραία γεγονότα είναι ελάχιστα αλλά όταν συμβούν η έντασή τους είναι εξαιρετικά

μεγάλη. Μια πιθανή ερμηνεία, θα μπορούσε να είναι η επιρροή που ασκείται στη Δυτική περιφέρεια από μεγάλες δασικές πυρκαγιές και ηφαιστειακές εκρήξεις.

Οπτικά, η σχέση ασυμμετρίας και κυρτότητας, για τις πόλεις των ΗΠΑ ανά περιφέρεια, απεικονίζεται στο ακόλουθο διάγραμμα,



Γράφημα 47: Συσχέτιση ασυμμετρίας και κυρτότητας στα κατάλοιπα των πόλεων των ΗΠΑ ανά περιφέρεια.

όπου, όπως και στην Ευρώπη, στο συντριπτικά μεγαλύτερο πλήθος των πόλεων υπάρχει μία θετική σχέση μεταξύ ασυμμετρίας και κυρτότητας, δηλαδή, όσο αυξάνεται η θετική ασυμμετρία τόσο αυξάνονται και οι τιμές του δείκτη κυρτότητας. Όπως ήδη αναφέραμε, αρκετές πόλεις της Δυτικής περιφέρειας (πράσινες κουκίδες), εμφανίζουν τις υψηλότερες τιμές ασυμμετρίας και κυρτότητας και απεικονίζονται σε περίοπτη θέση στο επάνω δεξί τμήμα του διαγράμματος.

4.8 Ανάλυση Επιπτώσεων Βραχυπρόθεσμων Συμβάντων

Στις προηγούμενες ενότητες, μελετήσαμε της μακροχρόνια συμπεριφορά των συγκεντρώσεων του διοξειδίου θείου, δίδοντας βαρύτητα στην εποχικότητα, την τάση και στις έκτακτες διακυμάνσεις που δεν εξηγούνται από αυτές. Στην παρούσα υποενότητα, εστιάζουμε στην ανάλυση των επιπτώσεων βραχυπρόθεσμων γεγονότων, στη συμπεριφορά των συγκεντρώσεων διοξειδίου του θείου. Ένα χαρακτηριστικό τέτοιο γεγονός είναι η πανδημία COVID-19, η οποία ξέσπασε στα τέλη του 2019 και οδήγησε σε εκτεταμένους περιορισμούς μετακινήσεων καθώς και στη μείωση της βιομηχανικής δραστηριότητας παγκοσμίως.

Η νόσος COVID-19 είναι η νόσος που προκαλείται από τον ιό που ονομάζεται κορονοϊός σοβαρού οξέος αναπνευστικού συνδρόμου 2 (SARS-CoV-2). Ο SARS-CoV-2 είναι ένα νέο στέλεχος κορονοϊού, το οποίο δεν είχε εντοπιστεί στον άνθρωπο πριν από το Δεκέμβριο του 2019. Στις 11 Μαρτίου του 2020, ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) κήρυξε τη νόσο COVID-19 σε πανδημία και συνέστησε διεθνώς τη λήψη πρωτοφανών έκτακτων μέτρων για τον περιορισμό της.

Όπως αναφέραμε και στην αρχή της παρούσης διπλωματικής εργασίας, μεταξύ των βασικών ρύπων, το διοξείδιο του θείου ξεχωρίζει λόγω της προέλευσής του από ανθρωπογενείς δραστηριότητες, όπως η χρήση ορυκτών καυσίμων και η βιομηχανική παραγωγή. Παρακάτω θα παραθέσουμε τις μεθόδους που χρησιμοποιήσαμε και τα αποτελέσματα αυτών, προκειμένου να μελετήσουμε εάν οι περιορισμοί, ελέω πανδημίας, επηρέασαν και σε ποιο βαθμό τις συγκεντρώσεις του διοξειδίου του θείου, στις πόλεις της Ευρώπης και των ΗΠΑ.

4.8.1. Ανάλυση συγκεντρώσεων διοξειδίου του θείου τους μήνες Μάρτιο - Απρίλιο - Μάιο

Για να διερευνηθεί η επίδραση της πανδημίας COVID-19 στις συγκεντρώσεις διοξειδίου του θείου, υπολογίστηκε ο μέσος όρος των τιμών του διοξειδίου του θείου για τους μήνες Μάρτιο, Απρίλιο και Μάιο (MAM) κάθε έτους από το 2015 έως το 2025. Η περίοδος MAM 2020 αντιστοιχεί στο χρονικό διάστημα του πρώτου γενικευμένου lockdown, κατά το οποίο καταγράφηκε σημαντική μείωση της οικονομικής δραστηριότητας και των μετακινήσεων.

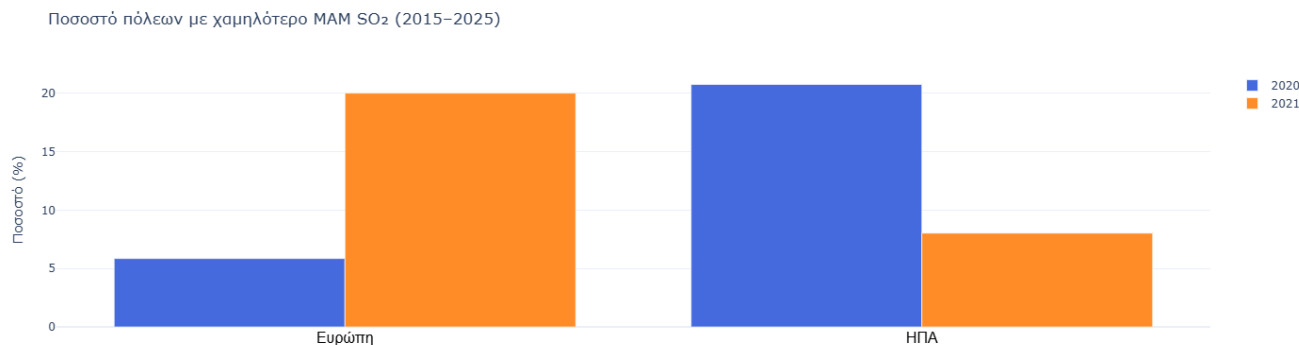
Για κάθε οντότητα, υπολογίστηκε η μέση τιμή MAM ανά έτος και συγκρίθηκαν οι τιμές των ετών 2020 και 2021 με τις αντίστοιχες τιμές MAM των προγενέστερων και μεταγενέστερων ετών, ώστε να διαπιστωθεί αν πράγματι κατά τη διάρκεια των περιορισμών παρατηρήθηκαν σημαντικές μειώσεις στις συγκεντρώσεις διοξειδίου του θείου (σ.σ. η διαδικασία έγινε σε αποεποχικοποιημένες χρονοσειρές). Στην υπόψη ανάλυση, η έννοια της σημαντικότητας προσδιορίστηκε όχι με κάποιο στατιστικό έλεγχο, αλλά με την απευθείας σύγκριση των μέσων MAM. Οι διαφορές εκφράστηκαν τόσο σε απόλυτες μονάδες Dobson Units όσο και ποσοστιαία (%). Ακολούθως, ελέγχθηκε σε πόσες οντότητες το έτος 2020 αποτέλεσε τη χρονιά με τη χαμηλότερη τιμή MAM, γεγονός που θα υποδείκνυε σαφή επίδραση των περιοριστικών μέτρων και ομοίως για το 2021.

Τα αποτελέσματα από αυτή τη διαδικασία παρατίθενται συνοπτικά παρακάτω:

<u>Europe</u>						<u>USA</u>					
Level	n	Lowest MAM 2020		Lowest MAM 2021		Level	n	Lowest MAM 2020		Lowest MAM 2021	
		n	%	n	%			n	%		
city	410	24	5.85	82	20.00	city	511	106	20.74	41	8.02
country	41	4	9.76	7	17.07	state	52	18	34.61	3	5.77

Παρά το γεγονός ότι το 2020 χαρακτηρίστηκε από τα πιο μαζικά και αυστηρά lockdown παγκοσμίως, μόλις το 6% των πόλεων του δείγματός μας κατέγραψε τις χαμηλότερες συγκεντρώσεις διοξειδίου του θείου της περιόδου 2015 - 2025. Ωστόσο, το 2021, μια χρονιά με πιο στοχευμένα, επαναλαμβανόμενα και συχνά τοπικά μέτρα, το ποσοστό των πόλεων που παρουσίασαν το χαμηλότερο τρίμηνο Μαρτίου - Μαΐου της δεκαετίας, ανήλθε στο 20% του δείγματός μας (82 πόλεις το 2021 έναντι μόλις 24 το 2020).

Εκ διαμέτρου αντίθετη είναι η εικόνα στις πόλεις των ΗΠΑ, όπου το 2020 συνολικά 106 πόλεις (21% του δείγματός μας) παρουσίασαν τις χαμηλότερες συγκεντρώσεις διοξειδίου του θείου, έναντι μόλις 41 (8%) το 2021.



Γράφημα 48: Ποσοστό πόλεων ανά ήπειρο με τις χαμηλότερες συγκεντρώσεις κατά της περιόδους εφαρμογής περιοριστικών μέτρων 2020 και 2021.

Αν και η ανάλυσή μας, περιορίστηκε μονάχα στον υπολογισμό των μέσων συγκεκριμένων μηνών κάθε έτους της τελευταίας δεκαετίας, η έντονη αντίθεση μεταξύ Ευρώπης και ΗΠΑ, θα μπορούσε να αποδοθεί στο διαφορετικό

τρόπο εφαρμογής των περιορισμών ακόμη και σε τοπικό επίπεδο. Στις ΗΠΑ, φαίνεται να υπήρξε ένα πιο συγχρονισμένο μοτίβο περιορισμών το 2020, με έντονα μέτρα ευρείας εφαρμογής, οπότε οι επιπτώσεις στις συγκεντρώσεις του διοξειδίου του θείου να ήταν τόσο σημαντικές, ώστε να 'κάλυψαν' τις επιπτώσεις από τα πιο στοχευμένα και μέτριας έντασης μέτρα που λήφθηκαν το 2021.

Αντίθετα, στην Ευρώπη, το 2020, ίσως εφαρμόστηκαν διαφορετικής έντασης και τύπου μέτρα, τα οποία, να μην είχαν άμεση επίπτωση στις συγκεντρώσεις διοξειδίου του θείου, (πχ. στη Σουηδία και την Ολλανδία τα μέτρα ήταν ήπια σε σχέση με άλλες χώρες της Ευρώπης) ή ακόμη και οι όποιες επιπτώσεις σε επίπεδο διοξειδίου του θείου να παρουσιάστηκαν εκτός των μηνών που επιλέξαμε προς έλεγχο.

Σε κάθε περίπτωση, γίνεται εμφανές ότι για μια ορθή και επιστημονική ερμηνεία, είναι αναγκαία η σε βάθος συνδυαστική εξέταση πολλαπλών παραγόντων, όπως η ένταση των μέτρων, ο χρονικός προσδιορισμός εκκίνησης, η διάρκεια εφαρμογής τους, οι κλιματικές συνθήκες κα.

Προς επίρρωση της άποψή μας αυτής, ανατρέξαμε στα στοιχεία του Oxford COVID-19 Government Response Tracker 'A global panel database of pandemic policies - Nature Human Behaviour' [96], από όπου διαπιστώσαμε ότι η Ευρώπη και οι ΗΠΑ, για τις ίδιες περιόδους (MAM 2020 και 2021) εφάρμοσαν παρόμοιας έντασης μέτρα, ήτοι,

Μέσο Stringency Index MAM 2020 - ΗΠΑ: 63.65
Ημέρες ΗΠΑ: 92
Μέσο Stringency Index MAM 2020 - Ευρώπη: 64.99
Ημέρες Ευρώπη: 92
Διαφορά (ΗΠΑ - Ευρώπη): -1.34

Μέσο Stringency Index MAM 2021 - ΗΠΑ: 58.25
Ημέρες ΗΠΑ: 92
Μέσο Stringency Index MAM 2021 - Ευρώπη: 61.70
Ημέρες Ευρώπη: 92
Διαφορά (ΗΠΑ - Ευρώπη): -3.45

κάτι που επιβεβαιώνει την αρχική μας εκτίμηση, ότι δεν επαρκεί η απλή μέθοδος σύγκρισης των μέσων τιμών συγκεκριμένων περιόδων, αλλά πρέπει να εξεταστούν και άλλοι παράγοντες ώστε να μελετηθεί ο τυχόν αντίκτυπος στις συγκεντρώσεις διοξειδίου του θείου από ένα τέτοιο σημαντικό γεγονός όπως η πανδημία COVID-19.

4.8.2. Ανάλυση επιρροής συγκεντρώσεων διοξειδίου του θείου από εξωγενείς παράγοντες

Για να ελέγξουμε αν πράγματι υπήρξαν σημαντικές μεταβολές στις συγκεντρώσεις διοξειδίου του θείου, εξαιτίας των περιοριστικών μέτρων αντιμετώπισης της πανδημίας, καταφύγαμε σε πιο εξειδικευμένες τεχνικές, εξετάζοντας τόσο τη συνολική συμπεριφορά των συγκεντρώσεων SO₂ όσο και την τυχόν επιρροή της από εξωγενείς παράγοντες. Συγκεκριμένα, εφαρμόσαμε ένα σεναριακού τύπου έλεγχο, όπου εξετάσαμε όχι μόνο την επιρροή των συγκεντρώσεων διοξειδίου του θείου κατά τις περιόδους των lockdowns αλλά και κατά τις μετέπειτα περιόδους επανεκκίνησης των δραστηριοτήτων. Πρόσθετα, εξετάσαμε αν οι μετρικές των συγκεντρώσεων διοξειδίου του θείου επηρεάστηκαν από άλλους μετεωρολογικούς παράγοντες όπως η θερμοκρασία και η ταχύτητα του ανέμου.

Διαδικαστικά,

α) ορίσαμε τις εξής περιόδους ελέγχου (Events):

Lockdown 2020: 01.03.2020 έως 31.05.2020,
Rebound 2020: 01.06.2020 έως 31.08.2020,
Lockdown 2021: 01.02.2021 έως 30.04.2021 και
Rebound 2021: 01.05.2021 έως 31.07.2021.

β) υπολογίσαμε τις μέσες τιμές των συγκεντρώσεων διοξειδίου του θείου (σε αποεποχικοποιημένες χρονοσειρές) τόσο για τις ανωτέρω περιόδους όσο και για τις αντίστοιχες περιόδους, των ετών 2016 - 2019 (Baseline).

γ) υπολογίσαμε την ποσοστιαία απόκλιση των συγκεντρώσεων διοξειδίου του θείου μεταξύ των περιόδων ελέγχου και των αντίστοιχων περιόδων των ετών 2016 - 2019, ως,

$$\Delta SO_2 \% = 100 \times \frac{\text{Event Mean} - \text{Baseline Mean}}{\text{Baseline Mean}}$$

δ) ορίσαμε ως σημαντική απόκλιση $\Delta SO_2 \% = \geq \pm 10\%$ και υπολογίσαμε σχετικούς δείκτες (decrease και increase) για κάθε οντότητα.

ε) υπολογίσαμε την ποσοστιαία απόκλιση της θερμοκρασίας αέρα (μετρική T2M από σύνολο δεδομένων ERA5) και της ταχύτητάς του [$\text{wind_speed} = \sqrt{u10^2 + v10^2}$] (U10: οριζόντια συνιστώσα του ανέμου, δηλαδή από δυτικά προς ανατολικά σε m/s και V10: κατακόρυφη συνιστώσα του ανέμου, δηλαδή από νότο προς βορρά σε m/s), συγκρίνοντας τις περιόδους ελέγχου με τις αντίστοιχες περιόδους των ίδιων ως άνω ετών (2016 - 2019), ως,

$$\Delta T2M \% = 100 \times \frac{\text{Event} - \text{Baseline}}{\text{Baseline}}$$

$$\Delta WIND \% = 100 \times \frac{\text{Event} - \text{Baseline}}{\text{Baseline}}$$

Από την εφαρμογή των ανωτέρω, λάβαμε μεικτά και μη αναμενόμενα αποτελέσματα. Συγκεκριμένα,

Europe

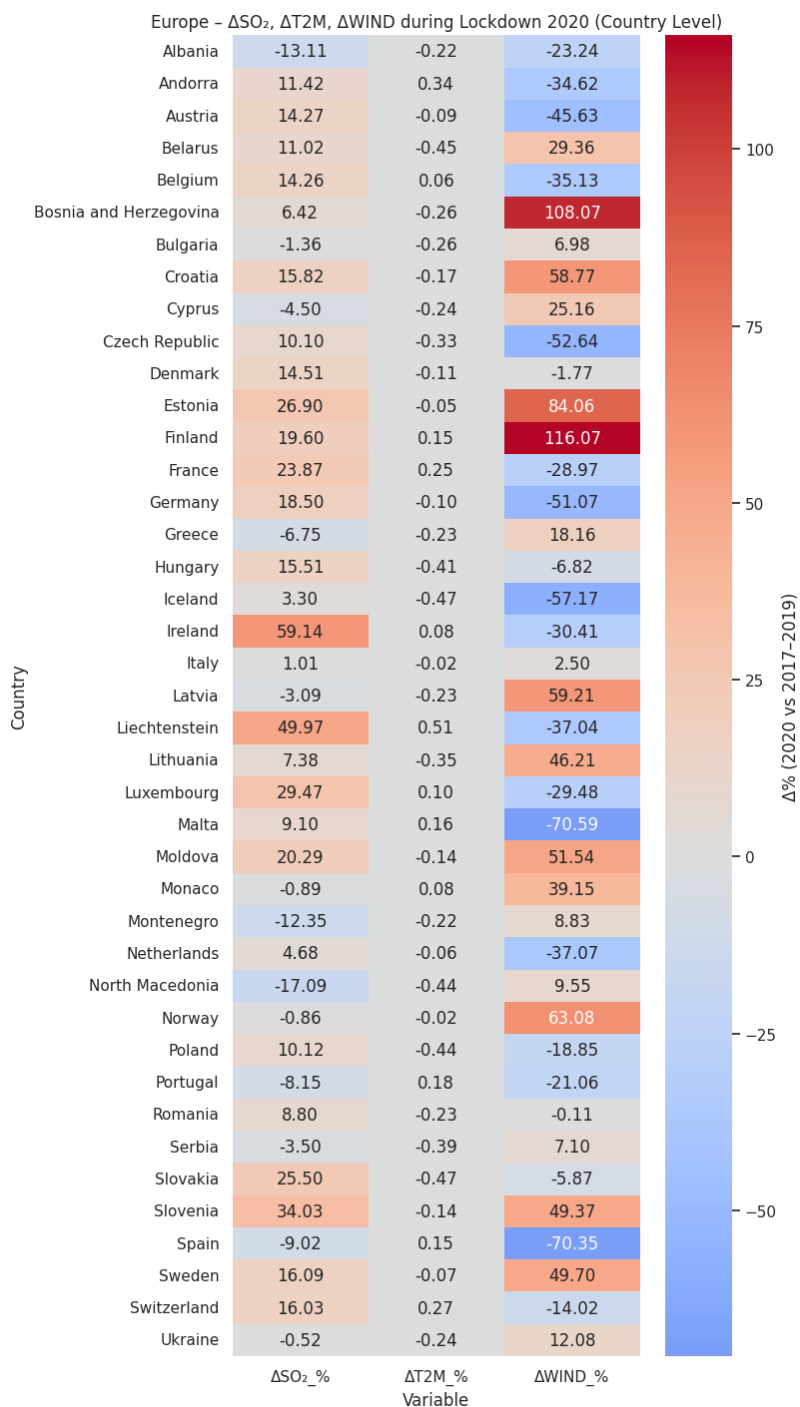
Level	n entities	Event	SO2 Change %	T2M Change %	WIND Change %
city	410	Lockdown_2020	12,05	-0,08	9,46
		Rebound_2020	-5,27	-0,14	-3,95
		Lockdown_2021	5,68	-0,43	-4,00
		Rebound_2021	2,39	-0,05	-2,81
country	41	Lockdown_2020	11,11	-0,09	11,70
		Rebound_2020	-5,57	-0,14	-5,21
		Lockdown_2021	5,50	-0,43	-5,15
		Rebound_2021	1,54	-0,05	-7,04

Στην Ευρώπη, καμία από τις περιόδους lockdown δε συνοδεύτηκε από στατιστικά σημαντική μείωση των συγκεντρώσεων διοξειδίου του θείου, αντίθετα, κατά το Lockdown 2020, παρατηρήθηκε αύξηση κατά 12,05%, παρά την αύξηση της ταχύτητας ανέμου κατά 9,46%, παράγοντας όπου θεωρητικά θα έπρεπε να μειώσει τις συγκεντρώσεις.

Κατά το Rebound 2020, αντί της αναμενόμενης αύξησης, καταγράφηκε μείωση κατά 5,27% και παρά τη μείωση του ανέμου κατά περίπου 4%, δηλαδή, αν οι άνεμοι ήταν κυμαίνονταν στα ίδια επίπεδα, θα μπορούσε η παρατηρούμενη μείωση να ήταν μεγαλύτερη.

Το 2021, παρατηρούμε ελαφρώς πιο ήπιες μεταβολές, με εκείνη των συγκεντρώσεων διοξειδίου του θείου να παραμένει σε θετικό επίπεδο αλλά μειωμένη έναντι του 2020, ίσως εξαιτίας αντίστοιχης μείωσης της ταχύτητας του ανέμου. Η μόνη αναμενόμενη συμπεριφορά, είναι εκείνη της ανάκαμψης μετά το lockdown του 2021, όπου παρατηρείται μία ήπια θετική μεταβολή των συγκεντρώσεων διοξειδίου του θείου κατά 2.4 ποσοστιαίες μονάδες, με ταυτόχρονη μείωση της ταχύτητας του ανέμου.

Συνοπτικά τα αποτελέσματα για το 2020, σε επίπεδο κρατών απεικονίζονται στην κάτωθι εικόνα:



Γράφημα 49: Συγκεντρωτικά στοιχεία παραγόντων κατά την περίοδο εφαρμογής περιοριστικών μέτρων 2020, ανά χώρα της Ευρώπης.

Από την παραπάνω εικόνα, θα αναφερθούμε σε 3 χώρες, την Ελλάδα, το Μαυροβούνιο και τη Βόρεια Μακεδονία, οι οποίες εμφάνισαν σημαντική μείωση των συγκεντρώσεων του διοξειδίου του θείου, παράλληλα όμως συνοδεύτηκε από αύξηση της ταχύτητας του ανέμου. Ειδικότερα για την Ελλάδα, όπου τα μέτρα περιορισμού της πανδημίας ήταν ιδιαίτερης έντασης και σε συνδυασμό με το γεγονός ότι η Ελλάδα δε διαθέτει βαριά βιομηχανία, φαίνεται πως ο συνδυασμός των καιρικών συνθηκών με τους περιορισμούς μετακινήσεων επηρέασαν σε σημαντικό βαθμό τις συγκεντρώσεις διοξειδίου του θείου.

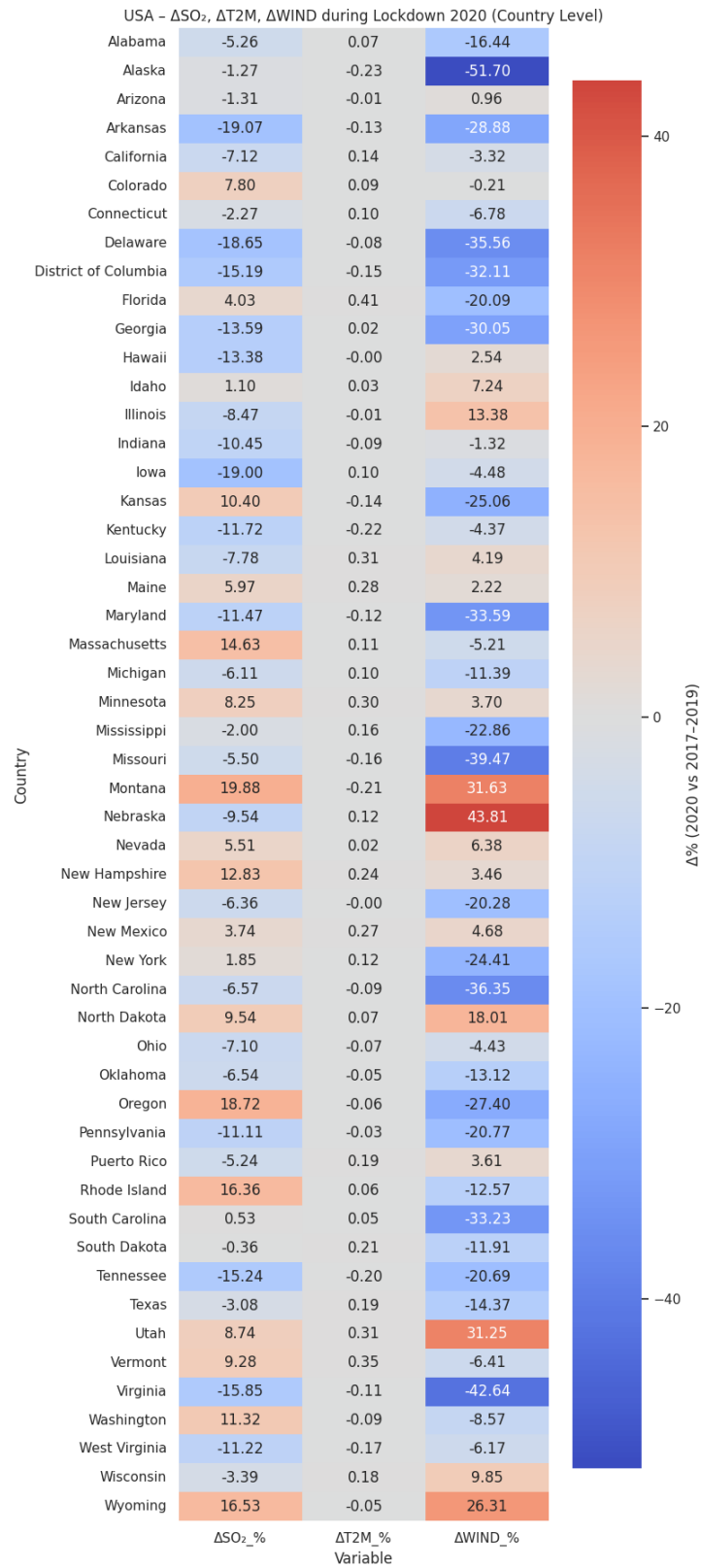
USA

Level	n entities	Event	SO2 Change %	T2M Change %	WIND Change %
city	511	Lockdown_2020	-1,13	-0,02	-5,20
		Rebound_2020	-1,61	0,11	1,87
		Lockdown_2021	6,50	-0,31	3,77
		Rebound_2021	1,59	-0,03	7,90
state	52	Lockdown_2020	-1,53	-0,03	-4,98
		Rebound_2020	-2,44	0,11	2,62
		Lockdown_2021	6,39	-0,31	3,23
		Rebound_2021	0,13	-0,03	8,69

Στις ΗΠΑ, σε αντίθεση με ότι παρατηρήσαμε στην Ευρώπη, κατά το lockdown του 2020 καταγράφηκε ήπια μείωση των συγκεντρώσεων διοξειδίου του θείου, η οποία μάλιστα ίσως ήταν να ήταν μεγαλύτερη αν δεν επικρατούσαν συνθήκες μείωσης κατά 5 ποσοστιαίες μονάδες της ταχύτητας του ανέμου. Η ανάκαμψη μετά το lockdown του 2020, συνοδεύτηκε με περεταίρω μείωση των συγκεντρώσεων του διοξειδίου του θείου, υποδεικνύοντας πως ίσως τα μέτρα που λήφθηκαν να ήταν πιο παρατεταμένα από την περίοδο του ελέγχου μας.

Αντίθετα, το 2021 συνοδεύτηκε από αύξηση των συγκεντρώσεων κατά 6.5% την περίοδο του Lockdown 2021, ενώ η πολύ ήπια αύξηση των συγκεντρώσεων, κατά την περίοδο ανάκαμψης, σε συνδυασμό με την αύξηση της ταχύτητας του ανέμου, μάλλον υποδηλώνει μια σχετική σταθερότητα. Σε κάθε περίπτωση, θα μπορούσαμε να πούμε πως τα αποτελέσματα για τις πόλεις και χώρες των ΗΠΑ είναι πιο αναμενόμενα από εκείνα της Ευρώπης.

Συνοπτικά και για τις πολιτείες των ΗΠΑ,



Γράφημα 50: Συγκεντρωτικά στοιχεία παραγόντων κατά την περίοδο εφαρμογής περιοριστικών μέτρων 2020 ανά πολιτεία των ΗΠΑ.

φαίνεται ότι κατά το Lockdown 2020, οι περισσότερες πολιτείες είχαν ήπια μείωση των συγκεντρώσεων, παρά τη μειωμένη ταχύτητα του ανέμου, όπου θα έπρεπε να επιφέρει αντίθετο αποτέλεσμα. Αυτό δείχνει ότι πράγματι οι συγκεντρώσεις του διοξειδίου του θείου παρουσίασαν μείωση πιθανότατα εξαιτίας των περιορισμών που εφαρμόστηκαν στο πλαίσιο της πανδημίας.

4.9 Συσχέτιση διοξειδίου του θείου με μετεωρολογικούς παράγοντες και το όζον (O₃)

Για να μελετήσουμε τη συσχέτιση των συγκεντρώσεων διοξειδίου του θείου με μετεωρολογικούς παράγοντες αλλά και με έναν άλλο ρύπο, το όζον (O₃) καταφύγαμε σε διάφορες μεθόδους που θα αναλύσουμε στις αμέσως επόμενες ενότητες.

4.9.1. Ανάλυση συσχέτισης συγκεντρώσεων διοξειδίου του θείου με το όζον (O₃).

Στην υποενότητα αυτή εξετάζουμε τη σχέση των συγκεντρώσεων του διοξειδίου του θείου με τις συγκεντρώσεις του όζοντος στην ατμόσφαιρα, για όλες τις οντότητες και χωρικά επίπεδα του δείγματός μας. Σημειώνουμε πως για το σκοπό αυτό εφαρμόσαμε στις χρονοσειρές του όζοντος την ίδια μέθοδο αποσύνθεσης STL, με εκείνη που χρησιμοποιήσαμε για τις χρονοσειρές διοξειδίου του θείου.

Σύμφωνα με τους Chris Chatfield και Hairpeng Xing στο βιβλίο τους ‘The Analysis of Time Series: An Introduction with R’ [55], μία στατιστική μέθοδος για τη διερεύνηση της σχέσης μεταξύ δύο χρονοσειρών αποτελεί η συνάρτηση ετεροσυσχέτισης (CCF). Η CCF υπολογίζει για κάθε χρονική μετατόπιση (lag) πόσο ισχυρά και προς ποια κατεύθυνση συσχετίζονται οι τιμές της μίας μεταβλητής με τις τιμές της άλλης. Όταν, για παράδειγμα, η τιμή της συνάρτησης στο lag 0 είναι θετική, αυτό σημαίνει ότι οι δύο χρονοσειρές μεταβάλλονται ταυτόχρονα προς την ίδια κατεύθυνση. Αν η συσχέτιση κορυφώνεται σε αρνητικό lag, τότε η δεύτερη χρονοσειρά όπου στην περίπτωση μας είναι εκείνη των συγκεντρώσεων του όζοντος προηγείται χρονικά της πρώτης (SO₂), ενώ σε θετικό lag το αντίθετο.

Μαθηματικά, ορίζεται ως η κανονικοποιημένη συνδιακύμανση των δύο χρονοσειρών (X, Y) για κάθε χρονική καθυστέρηση k από τον τύπο:

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sqrt{\gamma_X(0)\gamma_Y(0)}} = \frac{\gamma_{XY}(k)}{\sigma_X\sigma_Y}$$

όπου,

$$\gamma_{XY}(k) = Cov(X_t, Y_{t+k}) = E[(X_t - \mu_X)(Y_{t+k} - \mu_Y)]$$

και εκφράζει τη γραμμική συσχέτιση μεταξύ του X_t και του Y_{t+k} .

Αν $k > 0$, το X_t προηγείται του Y_t κατά k χρονικά βήματα και αν $k < 0$, το Y_t προηγείται του X_t ομοίως.

Οι τιμές της $\rho_{XY}(k)$ κυμαίνονται από -1 έως $+1$ και υπολογίζονται στην περίπτωση μας για lag = ± 12 μήνες.

Οι Chatfield και Xing τονίζουν ότι, στην κλασική περίπτωση γραμμικής ετεροσυσχέτισης Pearson, αν δύο χρονοσειρές είναι εντελώς ανεξάρτητες, τότε η αναμενόμενη τιμή της CCF (δηλαδή η μέση τιμή όλων των $r_{XY}(k)$) είναι περίπου μηδέν $E[r_{XY}(k)] \approx 0$ και η διασπορά της (δηλαδή πόσο μεταβάλλονται οι τιμές της γύρω από το μηδέν) είναι περίπου $Var[r_{XY}(k)] \approx \frac{1}{N}$, όπου N είναι το πλήθος των παρατηρήσεων. Δηλαδή, οι τιμές της CCF θα κινούνται τυχαία γύρω από το μηδέν και σπάνια θα ξεπερνούν τα θεωρητικά όρια εμπιστοσύνης $\pm 2/\sqrt{N}$. Αντίθετα, αν κάποια τιμή της ετεροσυσχέτισης πέσει έξω από αυτό το διάστημα (πχ. είναι πάνω από $+2/\sqrt{N}$ ή κάτω από $-2/\sqrt{N}$), τότε θεωρείται στατιστικά σημαντική, δηλαδή υπάρχει ένδειξη πραγματικής σχέσης ανάμεσα στις δύο χρονοσειρές στο συγκεκριμένο lag.

Για τις ανάγκες όμως της παρούσης διπλωματικής εργασίας, δεν εφαρμόσαμε την κλασική συνάρτηση ετεροσυσχέτισης CCF που βασίζεται στο γραμμικό συντελεστή Pearson ($\frac{\gamma_{XY}(k)}{\sigma_X\sigma_Y}$). Και τούτο διότι οι χρονοσειρές των συγκεντρώσεων του διοξειδίου του θείου και του όζοντος δεν ακολουθούν την κανονική κατανομή και δεν παρουσιάζουν απαραίτητα γραμμική σχέση, γεγονός που την καθιστούν λιγότερο κατάλληλη για τις χρονοσειρές διοξειδίου του θείου και όζοντος.

Προς τούτο, επιλέξαμε να χρησιμοποιήσουμε μία μέθοδο ανάλυσης χρονικά μετατοπισμένων συσχετίσεων με το συντελεστή Spearman (ρ ή rho) [97], η οποία είναι εννοιολογικά συναφής με τη CCF, χωρίς όμως να ακολουθεί τον κλασικό γραμμικό ορισμό της. Ουσιαστικά, η μέθοδος αυτή χρησιμοποιεί τις ταξινομημένες τιμές (ranks) των

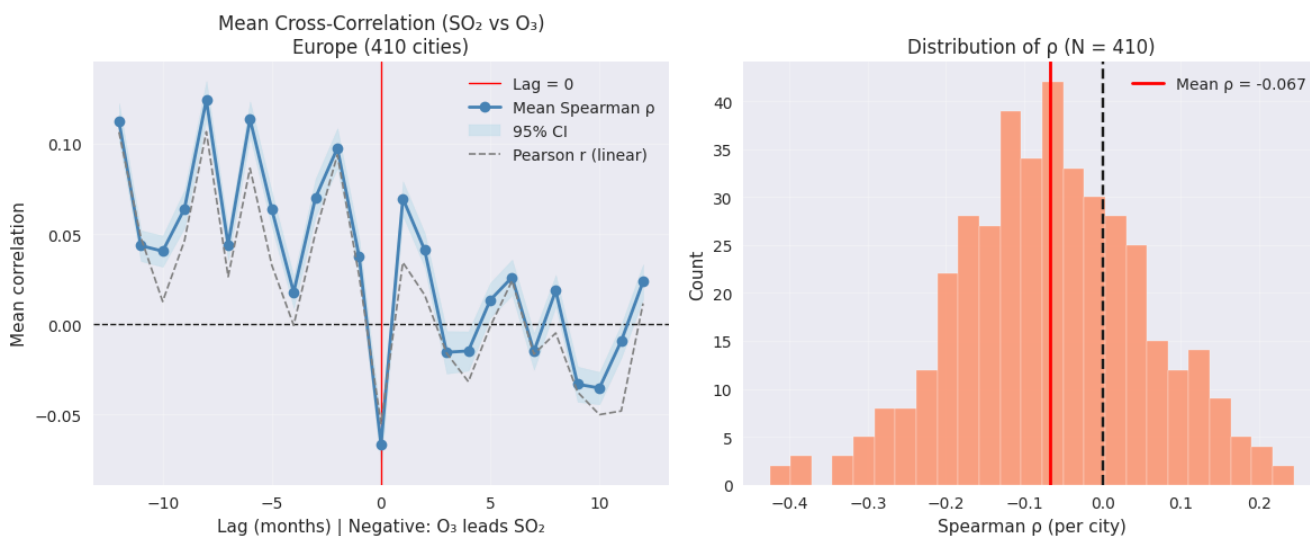
παρατηρήσεων και εξετάζει το βαθμό μονοτονικής σχέσης μεταξύ των δύο ρύπων, δηλαδή το κατά πόσο οι συγκεντρώσεις τους αυξάνονται ή μειώνονται μαζί, χωρίς να απαιτείται η ύπαρξη γραμμικότητας.

Ορίζεται δε ως ο συντελεστής Pearson εφαρμοσμένος στις ταξινομημένες τιμές, δηλαδή,

$$\rho = r_{\text{Pearson}}(\text{rank}(X), \text{rank}(Y))$$

Πρόσθετα σημειώνουμε πως ο συντελεστής Spearman είναι ανθεκτικός σε ακραίες τιμές (outliers) και κατάλληλος για δεδομένα όπου δεν ακολουθούν κανονική κατανομή, ενώ οι τιμές που παράγονται από την παραπάνω σχέση κυμαίνονται μεταξύ -1 και +1, όπου +1 δηλώνει τέλεια θετική μονοτονική σχέση, -1 δηλώνει τέλεια αρνητική μονοτονική σχέση και 0 υποδηλώνει την πλήρη απουσία μονοτονικής σχέσης. Η αξιολόγηση της συσχέτισης ανά χρονική μετατόπιση πραγματοποιείται μέσω της μέσης τιμής του συντελεστή και των αντίστοιχων διαστημάτων εμπιστοσύνης.

Οπτικά, που είναι και πιο εύκολο να παρουσιαστούν τα αποτελέσματα της υπόψη διαδικασίας,



Γράφημα 51: Ανάλυση συσχέτισης διοξειδίου του θείου και όζοντος και κατανομή των Ευρωπαϊκών πόλεων βάσει τιμών CCF

στο αριστερό διάγραμμα, απεικονίζεται η μέση τιμή του συντελεστή Spearman σε κάθε χρονική μετατόπιση (± 12 μηνών). Οι θετικές τιμές στα αρνητικά lags υποδεικνύουν ότι το O_3 προηγείται χρονικά του SO_2 , με τη μέγιστη μέση συσχέτιση να εμφανίζεται στις υστερήσεις -8 και -6 μηνών ($\rho \approx 0,12$), ένδειξη ύπαρξης μιας ασθενούς αλλά σταθερής μονοτονικής χρονικής σχέσης σε αυτές τις χρονικές υστερήσεις.

Για θετικές υστερήσεις, δηλαδή όταν το διοξείδιο του θείου προηγείται χρονικά, οι τιμές της συσχέτισης παραμένουν πολύ κοντά στο 0 και δεν εμφανίζουν κάποιο σαφές ή συνεπές στατιστικό μοτίβο. Αντίθετα, στο lag 0 παρατηρείται ασθενής αλλά στατιστικά σημαντική αρνητική συσχέτιση ($\rho \approx -0,07$), γεγονός που υποδηλώνει ότι σε ταυτόχρονη χρονική κλίμακα οι αποκλίσεις των δύο ρύπων τείνουν να κινούνται αντίθετα.

Στο σημείο αυτό, πρέπει να επισημάνουμε ότι το όζον αποτελεί έναν από τους πιο σημαντικούς οξειδωτικούς παράγοντες στην τροπόσφαιρα, καθώς συμμετέχει άμεσα ή έμμεσα σε αντιδράσεις που οδηγούν στην αποδόμηση και απομάκρυνση του διοξειδίου του θείου. Μάλιστα, σε περιόδους υψηλών συγκεντρώσεων όζοντος, αυξάνεται η παραγωγή ριζών υδροξυλίου ($\bullet OH$) και υδροπεροξυλίου ($HO_2\bullet$), οι οποίες οξειδώνουν το διοξείδιο του θείου σε θειικό οξύ (H_2SO_4). Στην υγρή φάση δε, το όζον μπορεί να οξειδώσει άμεσα το διοξείδιο του θείου (πχ. σε υδροσταγονίδια ή σύννεφα), επιταχύνοντας έτσι το σχηματισμό θειικών ιόντων (SO_4^{2-}). Ως αποτέλεσμα, αυτής της χημικής διαδικασίας, οι συγκεντρώσεις διοξειδίου του θείου στην ατμόσφαιρα μειώνονται, ενώ ταυτόχρονα αυξάνονται οι συγκεντρώσεις δευτερογενών σωματιδίων θειικών ενώσεων.

Ωστόσο, κάτι τέτοιο δεν εξηγείται απόλυτα από τα ανωτέρω αποτελέσματα. Δηλαδή, παρότι το όζον αποτελεί ισχυρό οξειδωτικό παράγοντα που συμμετέχει σε αντιδράσεις αποδόμησης του διοξειδίου του θείου, η

παρατηρούμενη ασθενής θετική ετεροσυσχέτιση στις αρνητικές υστερήσεις, πιθανότατα να αντανακλά την επίδραση κοινών μετεωρολογικών και εποχικών παραγόντων. Αυτοί οι παράγοντες μάλλον ευνοούν την ταυτόχρονη αύξηση και των δύο ρύπων σε ορισμένες περιόδους, με αποτέλεσμα η καθαρά χημική αντικρουόμενη σχέση τους να μην αποτυπώνεται ισχυρά στη στατιστική ανάλυσή μας.

Στο δεξιό τμήμα παρουσιάζεται η κατανομή των επιμέρους συντελεστών Spearman (ρ) ανά πόλη, όπου η πλειονότητα των τιμών συγκεντρώνεται γύρω από το $-0,08$, με σχεδόν συμμετρική κατανομή και μικρή διασπορά, γεγονός που επιβεβαιώνει ότι η συνολική σχέση μεταξύ των δύο ρύπων είναι πολύ ασθενής και ελαφρώς αρνητική.

Προκειμένου να εντοπίσουμε αν οι παρατηρούμενες χρονικές εξαρτήσεις μεταξύ συγκεντρώσεων διοξειδίου του θείου και όζοντος υποδηλώνουν πιθανή προγνωστική αιτιότητα, εφαρμόστηκε η ανάλυση αιτιότητας κατά Granger (Granger, C. W. J. (1969)), η οποία εξετάζει αν οι παρελθούσες τιμές μιας μεταβλητής περιέχουν στατιστικά σημαντική πληροφορία για την πρόβλεψη μιας άλλης. Η εφαρμογή της διαδικασίας αυτής, πραγματοποιήθηκε στις αποεποχικοποιημένες χρονοσειρές των συγκεντρώσεων των δύο ρύπων, για όλες τις πόλεις της Ευρώπης και με μέγιστο lag 12 μηνών.

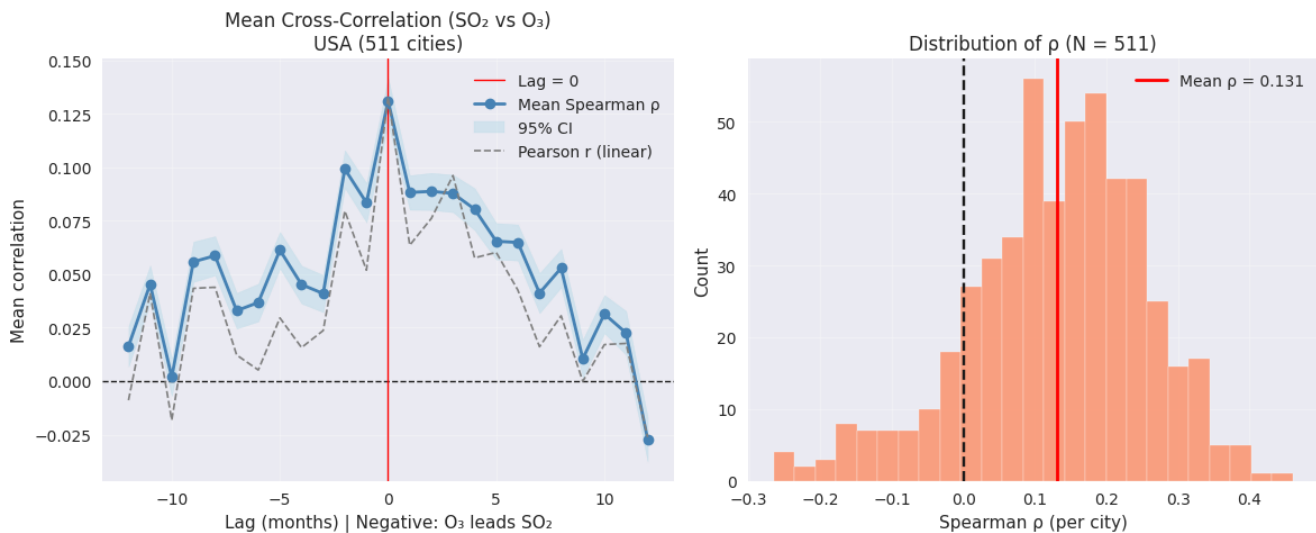
Συνοπτικά, τα αποτελέσματα είναι τα ακόλουθα,

Europe					
Direction	N	Lag No O₃ → SO₂	Lag No SO₂ → O₃	Percentage	
Bidirectional	32	4,5	5	7,8	
No significant causality	218	5,5	4,6	53,17	
O ₃ → SO ₂	109	4,8	5	26,59	
SO ₂ → O ₃	51	4,1	4,1	12,44	

σύμφωνα με τα οποία στο 53.2% των πόλεων, δεν εντοπίζεται στατιστικά σημαντική προγνωστική αιτιότητα (κατά Granger) μεταξύ των συγκεντρώσεων διοξειδίου του θείου και του όζοντος. Δηλαδή, οι όποιες μεταβολές των δύο ρύπων εξελίσσονται ανεξάρτητα. Ωστόσο, σε ένα σημαντικό ποσοστό των πόλεων, ήτοι 26.6%, οι μεταβολές των συγκεντρώσεων του όζοντος φαίνεται να προηγούνται χρονικά εκείνων του διοξειδίου του θείου, με μέση καθυστέρηση περίπου κατά 5 μήνες. Δηλαδή, οι συγκεντρώσεις του όζοντος παρέχουν στατιστικά προγνωστική πληροφορία για τις μελλοντικές συγκεντρώσεις του διοξειδίου του θείου. Σε ένα μικρότερο ποσοστό των πόλεων της τάξης του 12.4% παρατηρείται ακριβώς το αντίθετο, δηλαδή οι μεταβολές των συγκεντρώσεων διοξειδίου του θείου ενδέχεται να αποτελούν προγνωστικό πρόδρομο των συγκεντρώσεων του όζοντος (με καθυστέρηση 4 μηνών).

Για να εξηγηθεί αυτή η συμπεριφορά, απαιτούνται εξειδικευμένες γνώσεις και περαιτέρω ανάλυση, καθώς δεν είμαστε σε θέση να γνωρίζουμε αν πράγματι υπάρχουν συνθήκες που ωριμάζουν σε ένα τέτοιο χρονικό διάστημα και μπορούν να επηρεάσουν τις συγκεντρώσεις των δύο ρύπων. Μία πιο απλή ερμηνεία, θα μπορούσε να ήταν ότι συνδυαστικά κοινói παράγοντες επηρεάζουν διαφορετικά το διοξείδιο του θείου και το όζον ή απλώς πρόκειται για ένα αποτέλεσμα που θα μπορούσε να αποδοθεί στην τυχαιότητα. Σε κάθε περίπτωση, τα υπόψη αποτελέσματα δεν είναι ικανά να μας δώσουν μια ξεκάθαρη εικόνα για την επιρροή μεταξύ των δύο ρύπων.

Στην αντίπερα όχθη, σε επίπεδο πόλεων, παρατηρούμε μία εντελώς διαφορετική εικόνα,



Γράφημα 52: Ανάλυση συσχέτισης διοξειδίου του θείου και όζοντος και κατανομή των πόλεων των ΗΠΑ, βάσει τιμών CCF

όπου, η μέση τιμή του συντελεστή Spearman ρ στο lag 0 είναι θετική ($\rho \approx 0.13$), υποδηλώνοντας ότι οι συγκεντρώσεις του διοξειδίου του θείου και του όζοντος μεταβάλλονται κατά κανόνα προς την ίδια κατεύθυνση χωρίς καν χρονική καθυστέρηση. Παρόμοια θετικά επίπεδα συσχέτισης ($\rho = 0.08 - 0.10$) διατηρούνται και για μικρά θετικά lags (1 - 4 μήνες), γεγονός που δείχνει ότι σε αρκετές περιπτώσεις οι μεταβολές του διοξειδίου του θείου ακολουθούν εκείνες του όζοντος με καθυστέρηση ολίγων μηνών.

Η απουσία αρνητικών τιμών στο lag 0 και η γενικότερη θετική πορεία της συνάρτησης σε όλο το εύρος των lag, μαρτυρούν ότι οι δύο ρύποι δεν παρουσιάζουν αντικρουόμενη συμπεριφορά, αλλά μάλλον επηρεάζονται από κοινούς μηχανισμούς (πχ. φωτοχημικές ενώσεις, θερμοκρασία, μεταφορά αερίων μαζών, στασιμότητα ατμόσφαιρας κλπ.), οι οποίοι ευνοούν τις ταυτόχρονες μεταβολές (αυξήσεις ή μειώσεις) και στους δύο ρύπους.

Στο δεξιό διάγραμμα απεικονίζεται η κατανομή των συντελεστών Spearman, με μέση τιμή mean $\rho = 0.131$ και φαίνεται ξεκάθαρα πως το μεγαλύτερο πλήθος των πόλεων εμφανίζει θετικές τιμές, επιβεβαιώνοντας την παραπάνω ερμηνεία μας.

USA				
Direction	N	Lag No O ₃ → SO ₂	Lag No SO ₂ → O ₃	Percentage
Bidirectional	36	5,6	3,9	7,05
No significant causality	303	5	4,7	59,3
O ₃ → SO ₂	61	5,3	5,4	11,94
SO ₂ → O ₃	111	5,3	4,7	21,72

Η συμπληρωματική ανάλυση αιτιότητας κατά Granger, ενισχύει ακόμη περισσότερο τα συμπεράσματά μας, αποκαλύπτοντας ότι η πλειονότητα των πόλεων (59,3%) δεν παρουσιάζει στατιστικά σημαντική προγνωστική αιτιότητα κατά Granger μεταξύ SO₂ και O₃, δηλαδή απουσία σταθερής κατεύθυνσης πρόβλεψης. Ωστόσο, σε ποσοστό 21,7% των πόλεων, το διοξείδιο του θείου φαίνεται να προηγείται χρονικά του όζοντος με μέσο lag περίπου 5 μήνες, αντίθετα δηλαδή από ότι παρατηρήσαμε στις πόλεις της Ευρώπης.

Η συμπεριφορά αυτή ίσως να συνδέεται με τις διαφορετικές ατμοσφαιρικές και χημικές συνθήκες που επικρατούν στις ΗΠΑ. Σε πολλές αστικές και βιομηχανικές περιοχές, η παραγωγή όζοντος σχετίζεται με υψηλά επίπεδα πρόδρομων ρύπων, όπως οξειδία του αζώτου και πτητικές ενώσεις, οι οποίες, κάτω από συγκεκριμένες συνθήκες, οδηγούν στην παραγωγή όζοντος. Αντίθετα, οι οξειδωτικές διεργασίες που αναφέραμε και προηγουμένως μπορεί να είναι πιο περιορισμένες λόγω διαφορετικών μετεωρολογικών και άλλων συνθηκών.

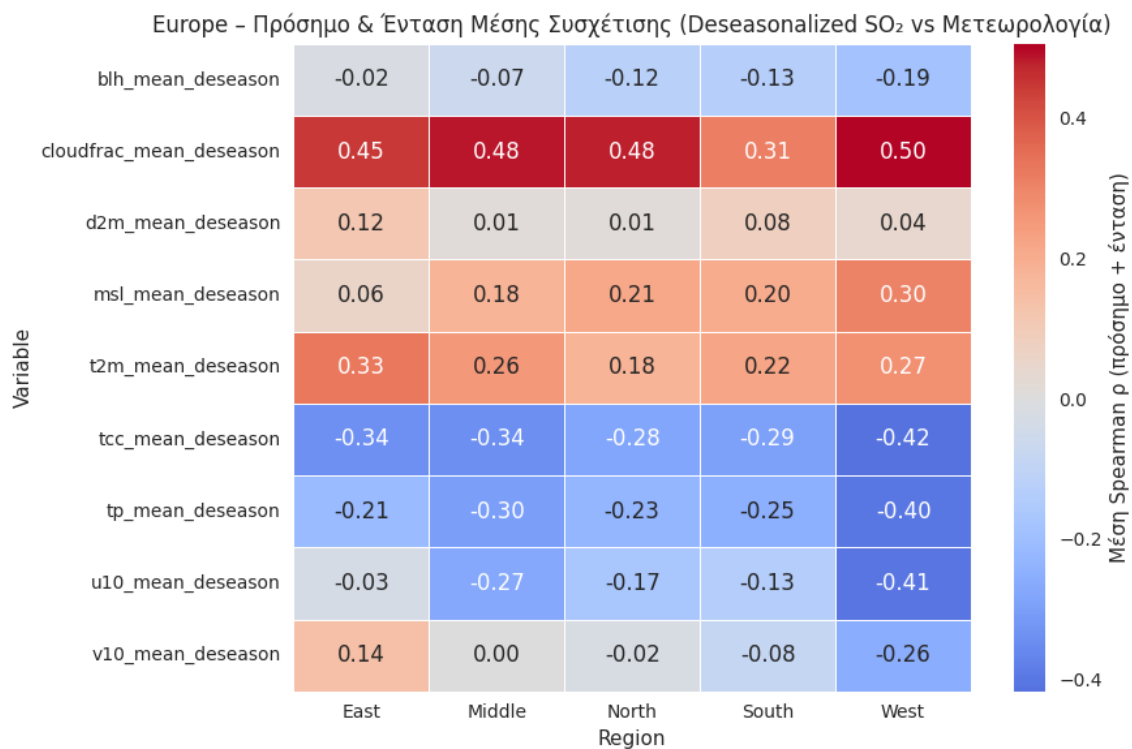
4.9.2 Ανάλυση συσχέτισης συγκεντρώσεων διοξειδίου του θείου με μετεωρολογικούς παράγοντες.

Οι συγκεντρώσεις του διοξειδίου του θείου στην ατμόσφαιρα επηρεάζονται σε μεγάλο βαθμό από τις επικρατούσες μετεωρολογικές συνθήκες, οι οποίες ρυθμίζουν τόσο τη διασπορά όσο και την απομάκρυνσή του από το οριακό στρώμα του αέρα. Παράγοντες όπως η θερμοκρασία, η υγρασία, το ύψος του οριακού στρώματος (BLH), η ταχύτητα και η διεύθυνση των ανέμων, η νεφοκάλυψη και η βροχόπτωση, καθορίζουν σε σημαντικό βαθμό τη χρονική διακύμανση των συγκεντρώσεων του διοξειδίου του θείου, καθώς επηρεάζουν την ανάμειξη, τη μεταφορά, τη διάλυση και την υγρή ή ξηρή εναπόθεσή του.

Στην παρούσα ενότητα θα μελετήσουμε τη σχέση μεταξύ των συγκεντρώσεων του διοξειδίου του θείου και των κυριότερων μετεωρολογικών παραγόντων σε επίπεδο πόλεων, με προβολή τους στις περιφέρειες, ώστε να εντοπίσουμε τυχόν μοτίβα εξαιτίας γεωγραφικών θέσεων. Η ανάλυσή μας πραγματοποιήθηκε σε αποεποχικοποιημένες χρονοσειρές, ώστε να εξαλειφθεί η επίδραση των εποχικών μοτίβων και να αναδειχθεί η πραγματική συσχέτιση μεταξύ των μετεωρολογικών παραγόντων και του διοξειδίου του θείου. Για το σκοπό αυτό χρησιμοποιήθηκε ο συντελεστής Spearman (ρ), ο οποίος, όπως αναφέραμε και στην προηγούμενη ενότητα, αποτυπώνει τη μονοτονική σχέση μεταξύ χρονοσειρών, ανεξάρτητα από τη γραμμικότητά τους.

Ως μετεωρολογικούς παράγοντες, 9 στο σύνολο, εξετάσαμε εκείνους που αντλήσαμε από το σύνολο δεδομένων ERA5 και περιγράφονται αναλυτικά σε προηγούμενη ενότητα.

Τα αποτελέσματα από τη διαδικασία που ακολουθήσαμε παρουσιάζονται παρακάτω.



Γράφημα 53: Συσχέτιση μεταξύ SO₂ και μετεωρολογικών παραγόντων, στην Ευρώπη.

Από τα παραπάνω αποτελέσματα φαίνεται πως οι μετεωρολογικοί παράγοντες που επηρεάζουν περισσότερο τις συγκεντρώσεις διοξειδίου του θείου σχετίζονται με τη νεφοκάλυψη, κάτι που ήταν αναμενόμενο καθώς η νεφοκάλυψη ασκεί μια σύνθετη επίδραση στις συγκεντρώσεις του διοξειδίου του θείου. Αν και περιορίζει την κατακόρυφη ανάδευση και μπορεί προσωρινά να εγκλωβίσει ρύπους κοντά στο έδαφος, η παράλληλη αύξηση της σχετικής υγρασίας και η δημιουργία υγρής φάσης επιταχύνουν τη διάλυση και την οξείδωση του SO₂ σε θειικά ιόντα. Ωστόσο, θα πρέπει να αναφέρουμε πως οι μετρήσεις των συγκεντρώσεων διοξειδίου του θείου που χρησιμοποιούμε συνολικά στην παρούσα διπλωματική εργασία, προέρχονται από δορυφορικές παρατηρήσεις και

όχι από μετρήσεις επίγειων σταθμών. Αν και λάβαμε τα προτεινόμενα από τη NASA μέτρα για την κατά το δυνατό ποιοτικότερη επιλογή των μετρήσεων, ο παράγοντας νεφοκάλυψη δεν παύει να επηρεάζει συνολικά τις δορυφορικές μετρήσεις.

Υπό το πρίσμα αυτό, παρατηρούμε μια αντίθεση μεταξύ των παραγόντων ολικής νεφοκάλυψης (tcc), δηλαδή νεφοκάλυψη σε όλη τη στήλη της ατμόσφαιρας και νεφοκάλυψη σε υψηλά στρώματα (cloudfrac). Ο πρώτος εξ αυτών των παραγόντων (tcc) εμφανίζει την πιο έντονη αρνητική συσχέτιση από όλους του παράγοντες και σε όλες τις περιφέρειες, υποδηλώνοντας ότι όσο αυξάνονται οι τιμές της ολικής νεφοκάλυψης, τόσο μειώνονται οι συγκεντρώσεις διοξειδίου του θείου. Και τούτο διότι περιορίζεται η ηλιακή ακτινοβολία (σημαντική μεταβολή των φωτοχημικών ενώσεων) και συνήθως ενισχύονται οι υγρές διεργασίες, όπως η συμπύκνωση και η βροχόπτωση, διευκολύνοντας έτσι την απομάκρυνσή του.

Αντίθετα, παρατηρούμε τις πιο έντονες θετικές συσχετίσεις για τη νεφοκάλυψη στα υψηλά στρώματα της ατμόσφαιράς. Το αποτέλεσμα αυτό δεν υποδηλώνει άμεση επίδραση των νεφών στο διοξείδιο του θείου αλλά μάλλον συνδέεται με την παρουσία λεπτών υψηλών νεφών σε θερμές και σταθερές ατμοσφαιρικές συνθήκες. Σε τέτοιες περιπτώσεις απουσιάζει η υγρή εναπόθεση και περιορίζεται η κατακόρυφη ανάμειξη, με αποτέλεσμα τη σταδιακή συσσώρευση του διοξειδίου του θείου στο κατώτερο στρώμα της ατμόσφαιρας. Επομένως, η έντονη θετική στατιστική συσχέτιση δεν υποδηλώνει απαραίτητα σχέση εξάρτησης, αλλά εκφράζει το γεγονός ότι οι αυξημένες τιμές νεφοκάλυψης στα ανώτερα στρώματα, δημιουργούν ή ακόμη και συνυπάρχουν με συνθήκες που ευνοούν τη διατήρηση υψηλότερων συγκεντρώσεων διοξειδίου του θείου ή περιορίζουν τους παράγοντες απομάκρυνσής του διοξειδίου του θείου.

Παρόμοια εικόνα με τον παράγοντα ολικής νεφοκάλυψης παρατηρούμε και για τη μετρική της ολικής κατακρήμνισης (tp), κάτι αναμενόμενο άλλωστε, αφού κατά διάρκεια των βροχοπτώσεων, το διοξείδιο του θείου διαλύεται σε υδροσταγονίδια και οξειδώνεται προς θειικά ιόντα (SO_4^{2-}), τα οποία απομακρύνονται αποτελεσματικά από την ατμόσφαιρα μέσω της καθίζησης. Μάλιστα, η βροχόπτωση αποτελεί έναν από τους πιο σημαντικούς φυσικούς μηχανισμούς απομάκρυνσης του διοξειδίου του θείου.

Το ύψος του οριακού στρώματος (BLH) καθορίζει το βαθμό κατακόρυφης ανάμειξης και κατά συνέπεια τη διασπορά των ρύπων στο κατώτερο τμήμα της τροπόσφαιρας. Όταν το BLH είναι χαμηλό, τότε οι ρύποι εγκλωβίζονται κοντά στο έδαφος, οδηγώντας σε αυξημένες συγκεντρώσεις τους, ενώ όταν είναι υψηλό έχει ως αποτέλεσμα την εντονότερη ανάμειξη και αραιώση των ρύπων. Για το λόγο αυτό, στις περισσότερες πόλεις της Ευρώπης, βλέπουμε μία αρνητική συσχέτιση μεταξύ BLH και SO_2 . Το φαινόμενο αυτό είναι πιο έντονο στις θαλάσσιες και παράκτιες περιοχές, όπου οι θερμικές αντιθέσεις ξηράς και θάλασσας ενισχύουν τη μεταφορά και τη διάχυση του αέρα, μειώνοντας έτσι τις συγκεντρώσεις του διοξειδίου του θείου.

Η ατμοσφαιρική πίεση στη μέση στάθμη της θάλασσας (mean sea level pressure - msl) είναι η πίεση που θα μετρούσαμε αν η ατμόσφαιρα εξισωνόταν σε μηδενικό ύψος, δηλαδή στο επίπεδο της θάλασσας, εξ' ου και η ονομασία του εν λόγω παράγοντα. Ουσιαστικά αποτελεί ένα δείκτη της επικρατούσας ατμοσφαιρικής σταθερότητας. Περιοχές με υψηλή πίεση χαρακτηρίζονται από καθοδικά ρεύματα, σταθερή ατμόσφαιρα και περιορισμένη κατακόρυφη ανάμειξη, ενώ περιοχές χαμηλής πίεσης συνδέονται με ανοδικά ρεύματα, αστάθεια και αυξημένη νεφοκάλυψη ή ακόμη και βροχόπτωση. Βάσει των αποτελεσμάτων της ανάλυσής μας, παρατηρούμε μία ασθενή προς μέτρια θετική συσχέτιση με τις συγκεντρώσεις διοξειδίου του θείου, η οποία, όσο πλησιάζουμε προς τη δυτική περιφέρεια της Ευρώπης γίνεται εντονότερη. Αυτό συμβαίνει πιθανότατα γιατί στις δυτικές ακτές της Ευρώπης επικρατούν συνθήκες που εγκλωβίζουν το διοξείδιο του θείου, όπως ξηρός καιρός και ασθενείς άνεμοι.

Η θερμοκρασία αέρα στα 2 μέτρα (t2m) σχετίζεται με τις θερμικές συνθήκες και τη σταθερότητα στο κατώτερο στρώμα της τροπόσφαιρας. Από την ανάλυσή μας προκύπτει μια μέτρια θετική συσχέτιση με το διοξείδιο του θείου σε όλες τις περιφέρειες, γεγονός που υποδηλώνει ότι οι θερμές και ξηρές ατμοσφαιρικές συνθήκες ευνοούν τη συσσώρευση του ρύπου. Σε τέτοιες περιόδους, η κατακόρυφη ανάμειξη και η απουσία υγρασίας, νεφοκάλυψης και βροχοπτώσεων, μειώνει τις διεργασίες απομάκρυνσης του διοξειδίου του θείου, δηλαδή, την υγρή εναπόθεση και τη χημική του μετατροπή σε θειικά ιόντα (SO_4^{2-}). Ως αποτέλεσμα, οι συγκεντρώσεις του διοξειδίου του θείου

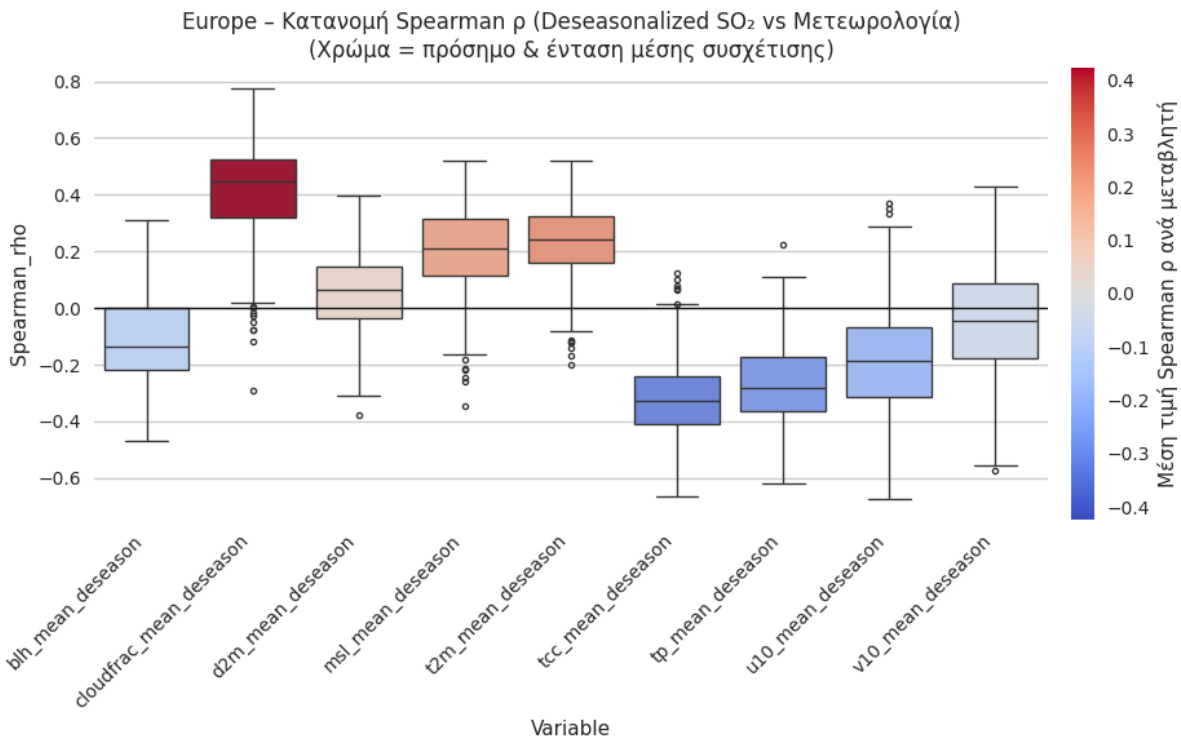
να παραμένουν αυξημένες ή καλύτερα να μην υπάρχουν κατάλληλες συνθήκες που να συμβάλλουν ουσιαστικά στη μείωσή τους.

Η θερμοκρασία δρόσου (d_{2m}) είναι η θερμοκρασία στην οποία ο αέρας φτάνει σε κορεσμό, δηλαδή σε σημείο να αρχίσει να σχηματίζει σταγονίδια νερού (υγρασία 100%). Ουσιαστικά η υπόψη μετρική δείχνει έμμεσα το επίπεδο υγρασίας, οπότε όταν η θερμοκρασία δρόσου είναι υψηλή, αυτό σημαίνει ότι υπάρχει πολλή υγρασία, άρα περισσότερα σταγονίδια και συνεπώς το διοξείδιο του θείου μετατρέπεται σε θειικά ιόντα. Όταν η θερμοκρασία δρόσου είναι χαμηλή, τότε ο αέρας είναι ξηρός και συνεπώς περιορίζονται οι συνθήκες απομάκρυνσής του, με αποτέλεσμα να παρατηρούνται αυξημένες συγκεντρώσεις. Ωστόσο, από την ανάλυσή μας, προέκυψε ότι ο υπόψη μετεωρολογικός παράγοντας έχει εξαιρετικά μικρή θετική συσχέτιση με τις συγκεντρώσεις του διοξειδίου του θείου σε όλες τις περιφέρειες, σχεδόν αμελητέα. Το φαινόμενο αυτό, έρχεται σε αντίθεση με ότι θα περιμέναμε, συνεπώς η επιστημονική εξήγηση του φαινομένου, μάλλον απαιτεί σε βάθος ανάλυση και ξεφεύγει από τα όρια της διπλωματικής εργασίας.

Ο παράγοντας u_{10} , ορίζει την οριζόντια ταχύτητα του ανέμου, μετρούμενη στα 10 μέτρα, όπου θετικές τιμές υποδεικνύουν ανέμους από τα δυτικά προς τα ανατολικά και αρνητικές τιμές από τα ανατολικά προς τα δυτικά. Τα αποτελέσματα της ανάλυσης δείχνουν ότι, στις δυτικές περιφέρειες της Ευρώπης, η ενίσχυση των ανέμων (είτε από δυτικά προς ανατολικά είτε το αντίθετο) συνδέεται με τη μείωση των συγκεντρώσεων διοξειδίου του θείου. Και τούτο διότι οι άνεμοι που προέρχονται από θαλάσσιες περιοχές, μεταφέρουν καθαρότερο αέρα με υψηλή υγρασία και, ανάλογα την έντασή τους, ευνοούν την οριζόντια και κατακόρυφη ανάμειξη, οδηγώντας σε αραίωση γενικά των ρύπων. Από την άλλη, παρατηρούμε μία αμελητέα συσχέτιση με την περιφέρεια της ανατολικής Ευρώπης, κάτι που απαιτεί περαιτέρω ανάλυση για την κατά το δυνατό ορθότερη ερμηνεία του.

Ο παράγοντας v_{10} , αποτελεί τη βόρεια συνιστώσα του ανέμου, όπου θετικές τιμές υποδεικνύουν ανέμους με φορά από το νότο προς το βορρά και αρνητικές τιμές, από το βορρά προς το νότο. Εδώ, παρατηρούμε μία διαφορετική εικόνα από ότι στην ανατολική συνιστώσα. Συγκεκριμένα, από την ανάλυσή μας προκύπτει ότι η συσχέτιση του υπόψη παράγοντα με το διοξείδιο του θείου είναι σχεδόν αμελητέα στις κεντρικές, βόρειες και νότιες περιφέρειες, ενώ παρουσιάζεται μία ήπια θετική συσχέτιση στην ανατολική περιφέρεια και αντίστοιχα μια ήπια προς μέτρια αρνητική συσχέτιση στη δυτική περιφέρεια της Ευρώπης. Για την τελευταία, (δυτική περιφέρεια) η πιο πιθανή εξήγηση είναι ότι οι νοτιάδες άνεμοι που προερχόμενοι από τη Μεσόγειο, συνδέονται με αυξημένη υγρασία, κάτι που ευνοεί τη μείωση των συγκεντρώσεων διοξειδίου του θείου. Αντίθετα, οι βόρειοι άνεμοι σχετίζονται με ψυχρές και ξηρές συνθήκες, που περιορίζουν την ανάμειξη και κατ' επέκταση τις συνθήκες απομάκρυνσης του διοξειδίου του θείου.

Πέραν της ανάλυσης των συσχετίσεων, ανά γεωγραφική περιοχή, παρακάτω παραθέτουμε τη συνολική εικόνα της διασποράς των τιμών Spearman,

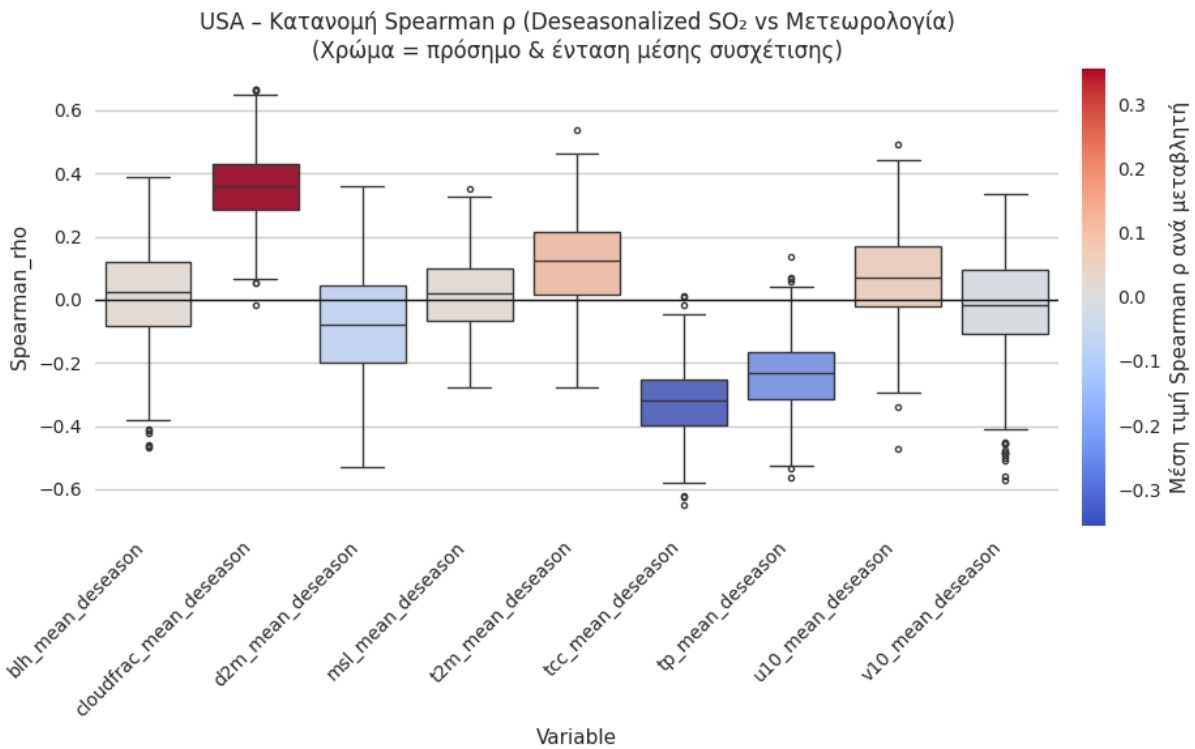


Γράφημα 54: box-and-whisker plot κατανομής συσχετίσεων διοξειδίου του θείου με μετεωρολογικούς παράγοντες στην Ευρώπη.

από όπου μπορούμε να διακρίνουμε πως οι παράγοντες που σχετίζονται με θερμές και σταθερές ατμοσφαιρικές συνθήκες (όπως η θερμοκρασία του αέρα t2m και η ατμοσφαιρική πίεση msl) εμφανίζουν θετικές μέσες συσχετίσεις, υποδηλώνοντας ότι σε περιόδους αυξημένης θερμοκρασίας ή πίεσης οι συγκεντρώσεις του διοξειδίου του θείου τείνουν να αυξάνονται. Αντίθετα, οι μεταβλητές που σχετίζονται με δυναμικές και υγρές συνθήκες, όπως η συνολική νεφοκάλυψη tcc, το ύψος βροχόπτωσης tp και οι συνιστώσες του ανέμου (u10, v10), παρουσιάζουν αρνητικές συσχετίσεις με το διοξείδιο του θείου, υποδηλώνοντας πως όσο αυξάνονται, οι συγκεντρώσεις τους, τείνουν να μειώνονται οι συγκεντρώσεις του διοξειδίου του θείου.

Ο παράγοντας νεφοκάλυψη στα υψηλά στρώματα της ατμόσφαιρας, όπως προαναφέραμε, παρουσιάζει την υψηλότερη θετική συσχέτιση μεταξύ των υπολοίπων μετεωρολογικών παραγόντων, υποδηλώνοντας πως όσο αυξάνονται οι τιμές του, αυξάνονται παράλληλα και οι τιμές των συγκεντρώσεων διοξειδίου του θείου, πιθανότατα εξαιτίας δημιουργίας συνθηκών που επηρεάζουν την απομάκρυνσή τους.

Τα αποτελέσματα της ανάλυσής μας για τις πόλεις των ΗΠΑ έχουν αρκετές ομοιότητες με εκείνα της Ευρώπης και λίγες διαφορές, κυρίως ως προς την ένταση των συσχετίσεων, κάτι που αντικατοπτρίζει την ανομοιογένεια του κλίματος μεταξύ των δύο ηπείρων.



Γράφημα 55: box-and-whisker plot κατανομής συσχετίσεων διοξειδίου του θείου με μετεωρολογικούς παράγοντες στις ΗΠΑ.

Από την ανωτέρω κατανομή των τιμών του συντελεστή Spearman, είναι εμφανής και εδώ η σημαντική επιρροή των παραγόντων νεφοκάλυψης cloudfrac και tcc στις συγκεντρώσεις διοξειδίου του θείου, όπου η πρώτη παρουσιάζει τη μεγαλύτερη θετική συσχέτιση από όλους τους μετεωρολογικούς παράγοντες και μάλιστα με μικρότερο εύρος τιμών μεταξύ των πόλεων, κάτι που υποδηλώνει ότι οι περισσότερες πόλεις ακολουθούν την ίδια συμπεριφορά. Αντίθετα, η συνολική νεφοκάλυψη (tcc) παρουσιάζει τη μεγαλύτερη αρνητική συσχέτιση, επιβεβαιώνοντας τη σημασία των συνθηκών ακτινοβολίας και παραγώγων της νεφοκάλυψης (πχ. βροχόπτωση) στη συμπεριφορά των συγκεντρώσεων διοξειδίου του θείου.

Σε επίπεδο έντασης, ακολουθούν οι παράγοντες βροχόπτωσης (tp) με μέτρια αρνητική συσχέτιση και οι θερμοκρασίας (t2m) και ανατολικής συνιστώσας ανέμου στα 10 μέτρα (u10), που εμφανίζουν μια ελαφριά προς μέτρια θετική συσχέτιση με τις συγκεντρώσεις διοξειδίου του θείου. Ειδικότερα, για τον παράγοντα της ανατολικής συνιστώσας του ανέμου (u10), σημειώνουμε πως έχει την ακριβώς αντίθετη συμπεριφορά από εκείνη της Ευρώπης, κάτι που αντανακλά τη διαφορετικότητα ως προς τις πηγές εκπομπών του διοξειδίου του θείου, όπου στις ΗΠΑ εντοπίζονται κυρίως στο εσωτερικό της ηπειρωτικής χώρας, οπότε οι ισχυρότεροι οριζόντιοι άνεμοι δεν επιδρούν το ίδιο αποτελεσματικά στη διασπορά γενικά των ρύπων, όπως συμβαίνει στην Ευρώπη, όπου η επίδραση των θαλάσσιων αέριων μαζών είναι πιο άμεση.

Σε αντίθεση με ότι παρατηρήσαμε στην Ευρώπη, ο παράγοντας ατμοσφαιρικής πίεσης, φαίνεται να μην έχει σαφή συσχέτιση με τις συγκεντρώσεις διοξειδίου του θείου, ενώ και ο παράγοντας της βόρειας συνιστώσας του ανέμου (v10), παρότι δεν έχει μια σταθερά αρνητική ή θετική συσχέτιση με το διοξείδιο του θείου, εμφανίζει πλήθος πόλεων με ασυνήθιστα ισχυρή αρνητική συσχέτιση (outliers), δηλαδή, στις πόλεις αυτές, όσο αυξάνεται η ταχύτητα των ανέμων παρατηρείται μια έντονη μείωση των συγκεντρώσεων διοξειδίου του θείου, κάτι που θα περιμέναμε να ισχύει για το σύνολο των πόλεων του δείγματός μας.

Συνολικά, η ανάλυση των συσχετίσεων μεταξύ των συγκεντρώσεων διοξειδίου του θείου και των κύριων μετεωρολογικών παραγόντων, ανέδειξε την πολυπαραγοντική φύση της ατμοσφαιρικής ρύπανσης, τις πολύπλοκες φωτοχημικές συνθήκες αλλά και την καθοριστική σημασία που διαδραματίζουν οι μετεωρολογικές συνθήκες στις συγκεντρώσεις διοξειδίου του θείου, με πιο σημαντικές τη νεφοκάλυψη και τη μετρική της βροχόπτωσης.

Κεφάλαιο 5: Πρόβλεψη συγκεντρώσεων διοξειδίου του θείου

Το τελευταίο στάδιο της μελέτης μας αφορά στην εκτίμηση της μελλοντικής πορείας των συγκεντρώσεων του διοξειδίου του θείου, με τη χρήση τόσο παραδοσιακών όσο και σύγχρονων μεθόδων πρόβλεψης, ώστε να αξιολογηθεί συγκριτικά η απόδοσή τους. Προς τούτο, επιλέξαμε τη χρήση των μοντέλων,

- i. Holt - Winters Exponential Smoothing, για την απλότητα, την υπολογιστική του αποδοτικότητα και την προσαρμοστικότητά του σε έντονα εποχικά χρονοσειρές με σταθερό μοτίβο,
- ii. SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors), που επεκτείνει το κλασικό μοντέλο ARIMA λαμβάνοντας υπόψη του την εποχικότητα και εξωγενείς παράγοντες, καθιστώντας το ιδανικό για τη διενέργεια μακροπρόθεσμων προβλέψεων και
- iii. LSTM (Long Short-Term Memory Neural Networks), ένα τύπο αναδρομικών νευρωνικών δικτύων, ικανό να συλλαμβάνει μη γραμμικές σχέσεις, μακροχρόνιες εξαρτήσεις και πολλαπλά μοτίβα, ιδανικό για προβλέψεις σε πολυσύνθετες χρονοσειρές.

Ως πεδίο εφαρμογής, χρησιμοποιήσαμε αυτούσιες τις παρατηρήσεις των συγκεντρώσεων διοξειδίου του θείου στις πόλεις του δείγματός μας σε Ευρώπη και ΗΠΑ. Σε προηγούμενη ενότητα αναλύθηκε η θεωρητική βάση των υπόψη μοντέλων, οπότε στην παρούσα ενότητα θα επικεντρωθούμε στην τεχνική εφαρμογή, την παραμετροποίησή τους και τη συγκριτική αξιολόγηση της απόδοσής τους, μέσω κατάλληλων δεικτών.

5.1 Εφαρμογή μοντέλου Holt - Winters Exponential Smoothing

Αρχικά επιλέξαμε μία πόλη (Αθήνα) από το δείγμα μας, προκειμένου να αξιολογήσουμε το βέλτιστο συνδυασμό των παραμέτρων για τη διαχείριση της εποχικότητας και της τάσης. Συγκεκριμένα, δοκιμάστηκαν διαφορετικοί συνδυασμοί των τύπων προσθετικός και πολλαπλασιαστικός, όπου ο προσθετικός υποθέτει σταθερή εποχική διακύμανση ανεξαρτήτως επιπέδου, ενώ ο πολλαπλασιαστικός επιτρέπει την αυξομείωση της εποχικότητας αναλογικά με την τάση.

Για κάθε συνδυασμό έγινε προσαρμογή των δεδομένων σε ετήσια εποχικότητα και υπολογίστηκαν αυτόματα οι παράμετροι εξομάλυνσης, μέσω βελτιστοποίησης ελαχίστων τετραγώνων (optimized = True). Σε περιπτώσεις αδυναμίας σύγκλισης εφαρμόστηκε μηχανισμός επαναδοκιμής (use_brute = True).

Ως κριτήριο επιλογής του βέλτιστου μοντέλου, χρησιμοποιήθηκε ο δείκτης Root Mean Square Error (RMSE) και για τη συνολική αποτίμηση της προσαρμογής, υπολογίστηκαν επιπλέον οι δείκτες Mean Absolute Error (MAE), Akaike Information criterion (AIC) και Bayesian Information Criterion (BIC).

Ερμηνεία δεικτών αξιολόγησης:

- RMSE: Υπολογίζει τη μέση τετραγωνική απόκλιση των προβλεπόμενων τιμών από τις πραγματικές, όπου, χαμηλές τιμές (στην περίπτωσή μας DU) υποδηλώνουν υψηλότερη ακρίβεια πρόβλεψης.
- MAE: Εκφράζει τη μέση απόλυτη απόκλιση μεταξύ πρόβλεψης και πραγματικής τιμής, όπου και εδώ, χαμηλότερες τιμές υποδηλώνουν μεγαλύτερη ακρίβεια πρόβλεψης.
- AIC και BIC: Στατιστικά κριτήρια που συνδυάζουν την καλή προσαρμογή του μοντέλου με την πολυπλοκότητά του. Χαμηλότερες τιμές υποδηλώνουν καλύτερη ισορροπία μεταξύ ακρίβειας προσαρμογής και απλότητας, καθιστώντας το μοντέλο πιο γενικεύσιμο.

Από την παραπάνω διαδικασία, προέκυψαν τα ακόλουθα αποτελέσματα,

Athens

Συνδυασμός	AIC	BIC	RMSE	MAE
Τάση = Add, Εποχικότητα = Add	-576.741	-534.741	0.061033	0.048522
Τάση = Add, Εποχικότητα = Mul	-576.696	-534.697	0.061040	0.049029
Τάση = Mul, Εποχικότητα = Add	-576.741	-534.741	0.061034	0.048526
Τάση = Mul, Εποχικότητα = Mul	-576.745	-534.746	0.061310	0.048963

όπου, φαίνεται πως όλοι οι συνδυασμοί αποδίδουν σχεδόν το ίδιο, με τις όποιες διαφορές να εντοπίζονται σε επίπεδο τρίτου δεκαδικού ψηφίου.

Τα σφάλματα προσαρμογής του μοντέλου, παρουσίασαν,

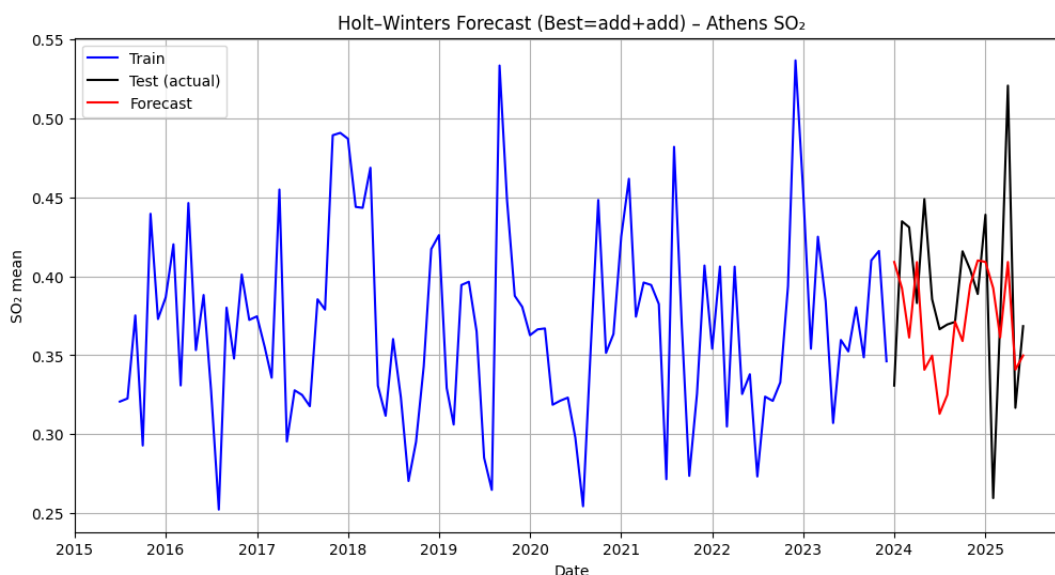
α) μέση τιμή -0.0000011 DU, δηλαδή, πολύ κοντά στο 0 (απειροελάχιστη και φυσικά αμελητέα τάση να υπερεκτιμά τις πραγματικές τιμές), ένδειξη απουσίας μεροληψίας (bias) και

β) τυπική απόκλιση 0.0508 DU, δηλαδή, η διαφορά μεταξύ πραγματικής και προβλεπόμενης τιμής,

Ο έλεγχος Ljung-Box είχε αποτέλεσμα $p\text{-value} = 0.0041$, δηλαδή, ένδειξη ότι παραμένει υπολειπόμενη αυτοσυσχέτιση στα σφάλματα προσαρμογής, οπότε το μοντέλο δεν κατάφερε να απορροφήσει πλήρως τη χρονική εξάρτηση των δεδομένων (σ.σ. υπάρχουν συστηματικά μοτίβα που δεν εξηγούνται πλήρως).

Ωστόσο, ο έλεγχος White (συνάρτηση `het_white` της βιβλιοθήκης `statsmodels.stats.diagnostic` [98]) παρήγαγε $p\text{-value} = 0.816$, δηλαδή δεν απορρίπτεται η μηδενική υπόθεση ομοσκεδαστικότητας. Επομένως, η διακύμανση των σφαλμάτων προσαρμογής μπορεί να θεωρηθεί ως σταθερή (σ.σ. δεν υπάρχουν ενδείξεις ετεροσκεδαστικότητας) και κατ' επέκταση το μοντέλο διατηρεί παρόμοια ακρίβεια τόσο σε χαμηλές όσο και υψηλές συγκεντρώσεις διοξειδίου του θείου.

Οπτικά, η εφαρμογή του μοντέλου Holt-Winters με παράμετρο `additive` για εποχικότητα και τάση μας έδωσε την κάτωθι εικόνα,



Γράφημα 56: Πρόβλεψη συγκεντρώσεων SO_2 για την πόλη της Αθήνας, με χρήση του μοντέλου Holt-Winters.

Προκειμένου να επιβεβαιώσουμε την επιλογή παραμέτρων, για την τάση `Additive` και για την εποχικότητα, επίσης `Additive` ως βέλτιστες, εφαρμόσαμε την πιο πάνω διαδικασία και σε άλλες πόλεις, από όπου λάβαμε παρόμοια αποτελέσματα. Ενδεικτικά, για την πόλη Helsinki της Φινλανδίας,

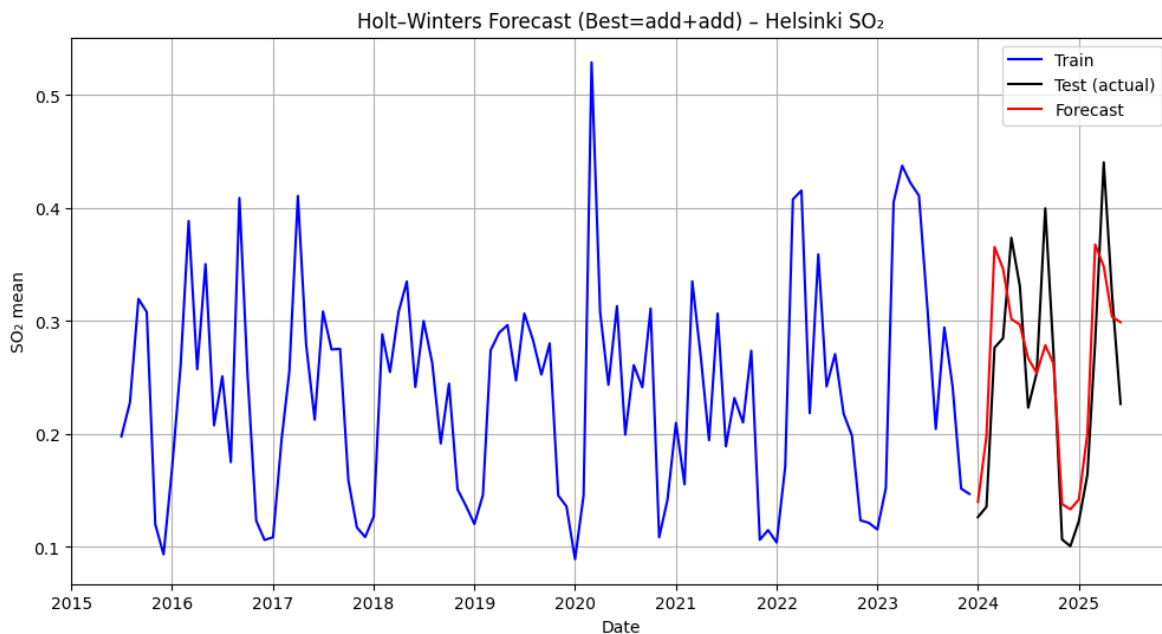
Helsinki

Συνδυασμός	AIC	BIC	RMSE	MAE
Τάση = Add, Εποχικότητα = Add	-562.434	-520.434	0.059725	0.049878
Τάση = Add, Εποχικότητα = Mul	-563.639	-521.640	0.061170	0.050027
Τάση = Mul, Εποχικότητα = Add	-562.466	-534.741	0.059831	0.050028
Τάση = Mul, Εποχικότητα = Mul	-563.729	-521.729	0.061511	0.050270

όπου κατάλοιπα παρουσιάζουν μέση τιμή 0.0000026 DU και τυπική απόκλιση 0.0545 DU, γεγονός που δείχνει απουσία μεροληψίας και σχετικά σταθερή διασπορά των σφαλμάτων προσαρμογής του μοντέλου (έγινε και

σχετικός έλεγχος). Το Ljung-Box test (για lag = 12) παρήγαγε p-value = 0.805, ένδειξη απουσίας αυτοσυσχέτισης στα σφάλματα προσαρμογής, δηλαδή οι διαφορές μεταξύ πραγματικών και τιμών πρόβλεψης δεν έχουν χρονική σειριακή εξάρτηση.

Οπτικά, η εφαρμογή του μοντέλου Holt-Winters για την πόλη Helsinki, με παράμετρο additive για εποχικότητα και τάση, μας έδωσε την κάτωθι εικόνα,



Γράφημα 57: Πρόβλεψη συγκεντρώσεων SO₂ για την πόλη Ελσίνκι, με χρήση του μοντέλου Holt-Winters.

Κατόπιν αυτών, επιλέχθηκε ως βέλτιστο το μοντέλο Holt-Winters με παραμέτρους additive για τάση και εποχικότητα.

Κατά την εφαρμογή του στο σύνολο των οντοτήτων, πραγματοποιήθηκαν δύο διαφορετικές προσεγγίσεις πρόβλεψης, με στόχο τη σύγκριση της απόδοσής του σε στατικό (Batch) αλλά και σε δυναμικό πλαίσιο (Rolling).

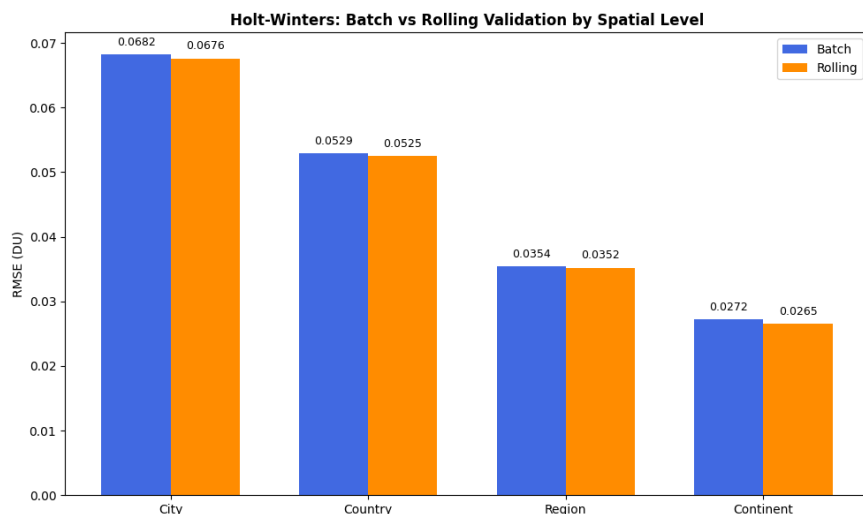
Πιο συγκεκριμένα, εφαρμόστηκαν,

α) Batch πρόβλεψη με ορίζοντα 24 μηνών, όπου το μοντέλο εκπαιδεύτηκε μόνο μία φορά στο σύνολο των παρατηρήσεων για την περίοδο 2015-2023. Με τον τρόπο αυτό, παρήγαγε απευθείας προβλέψεις για ολόκληρη την περίοδο ελέγχου που ορίσαμε για το διάστημα 2024-2025. Το πλεονέκτημα αυτής της προσέγγισης είναι η απλότητα και η συνέπεια στη χρήση των ίδιων παραμέτρων σε όλο το διάστημα πρόβλεψης. Ωστόσο, δεν προσαρμόζεται σε νέα δεδομένα, κάτι που είναι αναγκαίο για το σκοπό μας.

β) Rolling πρόβλεψη με βήμα 1 μήνα, όπου το μοντέλο επανεκπαιδεύεται σε κάθε νέο μήνα, δηλαδή, μετά από κάθε πρόβλεψη ενός μηνός, το αποτέλεσμα ενσωματώνεται στη χρονοσειρά και το μοντέλο προσαρμόζεται εκ νέου. Με τον τρόπο αυτό, βελτιώνει σταδιακά την εκτίμηση της εποχικότητας και της τάσης βάσει των πιο πρόσφατων παρατηρήσεων. Το σφάλμα υπολογίζεται με τη μέθοδο 1-step-ahead forecasting, όπου η πρόβλεψη κάθε περιόδου συγκρίνεται με την πραγματική τιμή που τελικά εμφανίστηκε.

Η ως άνω μεθοδολογία εφαρμόστηκε στο σύνολο των οντοτήτων του δείγματός μας, ενώ για την αποφυγή αστοχιών προσαρμογής, συμπεριλάβαμε μηχανισμό επαναδοκιμής (brute-force retry) σε περίπτωση μη σύγκλισης. Ο χρόνος εκτέλεσης για κάθε ήπειρο (με παραλληλισμό) ανήλθε σε 4 λεπτά.

Παρακάτω, παραθέτουμε τα αποτελέσματα από την εφαρμογή του μοντέλου Holt-Winters,



Γράφημα 58: Συγκριτικά αποτελέσματα εφαρμογής Holt -Winters μεταξύ των μεθόδων Batch και Rolling, σε όλα τα χωρικά επίπεδα της Ευρώπης.

από όπου φαίνεται πως οι διαφορές είναι μεν μικρές μεταξύ Batch και Rolling πρόβλεψης, ωστόσο η τελευταία, έστω και ελάχιστα, δίδει μικρότερο σφάλμα. Έτσι, επιλέξαμε να κρατήσουμε την καλύτερη εκ των δύο μεθόδων για κάθε οντότητα.

Για την ευκολότερη ερμηνεία των αποτελεσμάτων, προβήκαμε στις ακόλουθες κατηγοριοποιήσεις,

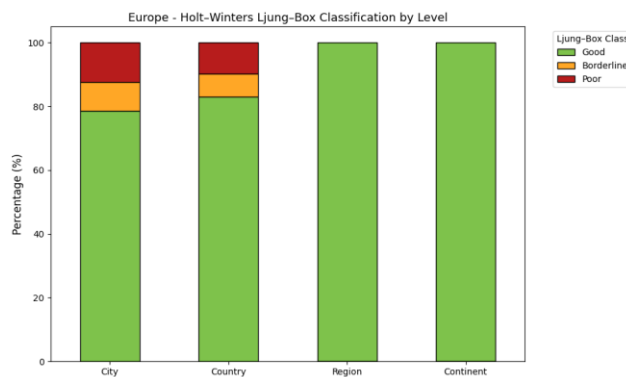
RMSE	Ερμηνεία
< 0.03	Excellent
0.03 ≤ RMSE < 0.07	Good
0.07 ≤ RMSE < 0.12	Moderate
> 0.12	Poor

Ljung p-value	Ερμηνεία
> 0.10	Good
0.05 ≤ p-value < 0.10	Borderline
< 0.05	Poor

Παρακάτω παρουσιάζονται οπτικά τα αποτελέσματα από την εφαρμογή του μοντέλου Holt Winters στις χρονοσειρές του διοξειδίου του θείου.



Γράφημα 59: Ταξινόμηση χωρικών επιπέδων Ευρώπης, ανά κατηγορία δείκτη RMSE.

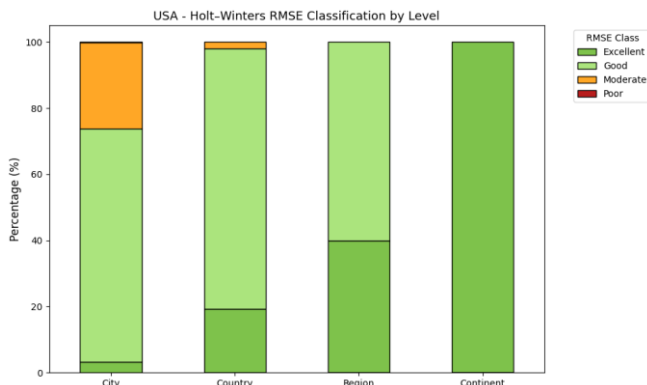


Γράφημα 60: Ταξινόμηση χωρικών επιπέδων Ευρώπης, ανά κατηγορία ελέγχου Ljung-Box

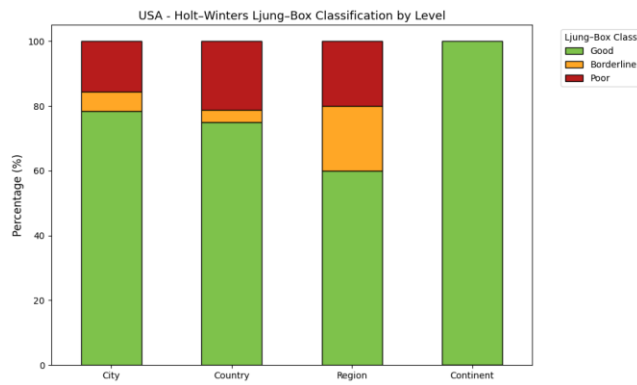
Στην Ευρώπη, περίπου το 60% των πόλεων ταξινομήθηκε στην κατηγορία Good, δηλαδή παρουσίασε τιμές RMSE μεταξύ 0.03 και 0.07 DU. Ένα σημαντικό ποσοστό (περίπου 39 %) εμφάνισε τιμές RMSE μεταξύ 0.07 και 0.12 DU, γεγονός που υποδηλώνει ότι, αν και το μοντέλο κατόρθωσε να αποτυπώσει τη γενική τάση και την εποχικότητα, δεν κατάφερε να προβλέψει με υψηλή ακρίβεια τις βραχυπρόθεσμες διακυμάνσεις των συγκεντρώσεων του διοξειδίου του θείου. Επιπλέον, από τον έλεγχο των σφαλμάτων προσαρμογής (μέσω Ljung-

Box test) προέκυψε ότι στο μεγαλύτερο ποσοστό των περιπτώσεων ($p\text{-value} > 0.10$), τα σφάλματα είναι ασυσχέτιστα, γεγονός που δείχνει πως το μοντέλο περιέγραψε ικανοποιητικά τη δομή των χρονοσειρών.

Για τις ΗΠΑ, με χρήση του ίδιο μοντέλου, προέκυψαν τα ακόλουθα,



Γράφημα 61: Ταξινόμηση χωρικών επιπέδων ΗΠΑ, ανά κατηγορία δείκτη RMSE.



Γράφημα 62: Ταξινόμηση χωρικών επιπέδων ΗΠΑ, ανά κατηγορία ελέγχου Ljung-Box

από όπου φαίνεται πως σε επίπεδο πόλεων, περίπου το 86% αυτών, κατατάσσεται στην κατηγορία ‘Good’ (RMSE μεταξύ 0.03 και 0.07 DU), επίδοση καλύτερη από εκείνη της Ευρώπης. Ο έλεγχος Ljung-Box έδειξε ότι σε περισσότερο από το 75% των χρονοσειρών δεν εντοπίζεται στατιστικά σημαντική αυτοσυσχέτιση στα σφάλματα προσαρμογής ($p\text{-value} > 0.10$), ένδειξη ότι και εδώ το μοντέλο κατάφερε να περιγράψει ικανοποιητικά τη δομή των χρονοσειρών. Οι μεμονωμένες περιπτώσεις που κατατάσσονται στις βαθμίδες ‘Borderline’ και ‘Poor’, υποδηλώνουν την ύπαρξη ορισμένων τοπικών μοτίβων ή ανωμαλιών που δεν μπορούν να περιγραφούν πλήρως από την απλή εποχική δομή του μοντέλου.

Συνοπτικά,

<u>Europe</u>				<u>USA</u>			
level	Mean RMSE	Mean MAE	Mean SMAPE	level	Mean RMSE	Mean MAE	Mean SMAPE
City	0.066802	0.051010	21.15%	City	0.058963	0.045043	20.06%
Country	0.051767	0.040060	15.68%	Country	0.043594	0.033289	10.33%
Region	0.034569	0.026123	10.15%	Region	0.031701	0.022351	8.62%
Continent	0.026541	0.018342	6.019%	Continent	0.027903	0.018929	6.80%

για την Ευρώπη, οι μέσες τιμές RMSE, MAE και Symmetric Mean Absolute Percentage Error (SMAPE), σε επίπεδο πόλεων, εμφανίζουν καλή ακρίβεια (μέση απόκλιση από τις πραγματικές τιμές περίπου 0.067 DU), ενώ όσο ανεβαίνουμε χωρικό επίπεδο (χώρα, περιφέρεια, ήπειρος), οι υπόψη δείκτες μειώνονται αισθητά, αντανakλώντας την εξομάλυνση των τοπικών διακυμάνσεων, αλλά και το γεγονός ότι οι ίδιες οι χρονοσειρές των ανώτερων επιπέδων προκύπτουν ως σταθμισμένοι μέσοι όροι (weighted means) των αρχικών μηνιαίων χρονοσειρών των πόλεων. Αποδεκτή προσαρμογή απεικονίζει και ο δείκτης SMAPE, όπου κατά μέσο όρο η πρόβλεψη διαφέρει κατά 21.15% από την πραγματική τιμή και φτάνει σε μονοψήφιο νούμερο στο ανώτατο χωρικό επίπεδο.

Στις ΗΠΑ, η εικόνα είναι καλύτερη, όπου στο επίπεδο πόλης το μέσο τετραγωνικό σφάλμα είναι ελαφρώς μειωμένο κατά 0.008 DU έναντι εκείνου της Ευρώπης. Ελαφρώς καλύτερος είναι και ο δείκτης SMAPE, με τις τιμές στις ανώτερες κλίμακες να είναι μονοψήφιες, ένδειξη ότι το μοντέλο έχει καλή προσαρμογή στα δεδομένα.

5.2 Εφαρμογή μοντέλου SARIMAX

Η εφαρμογή του μοντέλου SARIMAX, πραγματοποιήθηκε στο σύνολο των οντοτήτων και ανάλογα το επίπεδο επιλέχθηκαν οι κατάλληλες μεταβλητές εξωγενών παραγόντων, όπου στην περίπτωση μας είναι οι

μετεωρολογικοί παράγοντες του συνόλου δεδομένων ERA5. Για να αποφευχθούν ζητήματα κλίμακας μεταξύ των μεταβλητών, εφαρμόστηκε ομαλοποίηση μέσω z-score, με μια ελαφρώς διαφορετική προσέγγιση μεταξύ του επιπέδου πόλεων και των ανώτερων επιπέδων. Συγκεκριμένα, για τις πόλεις εφαρμόστηκε ομαλοποίηση εντός κάθε μίας από αυτές, ώστε να διατηρηθούν οι τοπικές διαφοροποιήσεις. Αντίθετα, στα υπόλοιπα επίπεδα, έγινε συνολικά ανά επίπεδο, προκειμένου να είναι συγκρίσιμα τα αποτελέσματα μεταξύ τους.

Πριν την εφαρμογή του μοντέλου, πραγματοποιήθηκε η διαδικασία της αποσύνθεσης με σκοπό τον εντοπισμό της επικρατέστερης εποχικής περιόδου. Ουσιαστικά υπολογίστηκε η διακύμανση της εποχικής συνιστώσας για τις υποψήφιες περιόδους 6 και 12 μηνών, επιλέγοντας τελικά εκείνη με τη μεγαλύτερη διακύμανση, δηλαδή, εκείνης που εξηγεί το μεγαλύτερο μέρος της χρονοσειράς. Η επιλογή των υποψήφιων περιόδων προέκυψε βάσει των αποτελεσμάτων της διαδικασίας αποσύνθεσης που αναλύσαμε διεξοδικά σε προηγούμενο κεφάλαιο, από όπου εντοπίσαμε περιπτώσεις, όπου πέραν των ετήσιων εμφάνιζαν είτε διπλό μοτίβο εποχικότητας (6 και 12 μηνών) είτε 6 μηνών.

Για την εύρεση των βέλτιστων παραμέτρων του μοντέλου ARIMA, χρησιμοποιήσαμε τη συνάρτηση 'auto_arima' από τη βιβλιοθήκη της rpython 'pmdarima', όπου για τις πόλεις καθορίσαμε το εύρος των τιμών $p = 0$ έως 2 και για τα υπόλοιπα χωρικά επίπεδα $p = 0$ έως 3.

Το μοντέλο SARIMAX ορίζεται ως SARIMA(p,d,q)(P,D,Q)m, όπου,

Παράμετρος	Ονομασία	Ερμηνεία
p	<i>autoregressive order</i>	Πόσες προηγούμενες τιμές της χρονοσειράς χρησιμοποιεί το μοντέλο (AR). Όσο μεγαλύτερο είναι το p τόσο μεγαλύτερη η εξάρτηση από παρελθούσες παρατηρήσεις.
d	<i>degree of differencing</i>	Ορίζει, πόσες φορές γίνεται διαφοροποίηση (I) για να επιτευχθεί η στασιμότητα. Για παράδειγμα d=1 σημαίνει ότι γίνεται μία φορά διαφοροποίηση, προκειμένου η χρονοσειρά να καταστεί στάσιμη.
q	<i>moving average order</i>	Πόσα προηγούμενα σφάλματα θα ληφθούν υπόψη από το μοντέλο (MA). Για παράδειγμα αν $q = 1$ τότε η πρόβλεψη για τον τρέχοντα μήνα διορθώνεται λαμβάνοντας υπόψη το σφάλμα που προέκυψε του προηγούμενου μήνα.
P	<i>seasonal autoregressive order</i>	Ομοίως με το p αλλά αφορά σε εποχικά βήματα. Δηλαδή πόσα εποχικά βήματα χρησιμοποιεί το μοντέλο. Στην περίπτωση μας όπου το $m=6$ ή 12 μήνες (ανάλογα το αποτέλεσμα της αποσύνθεσης) το κάθε βήμα είναι επίσης 6 ή 12 μήνες.
D	<i>seasonal differencing order</i>	Ομοίως με το d αλλά η διαφοροποίηση γίνεται στην εποχική συνιστώσα.
Q	<i>seasonal moving average order</i>	όπως το q, αλλά ως όροι τίθενται τα σφάλματα ανά εποχικό βήμα.
m	<i>seasonal period</i>	το μήκος του εποχικού κύκλου (πχ. $m = 12$ για ετήσιο κύκλο και $m = 6$ για εξαμηνιαίο), ανάλογα με το αποτέλεσμα της αποσύνθεσης.

Κατά τη διαδικασία εύρεσης των βέλτιστων παραμέτρων ακολουθήσαμε τα εξής βήματα,

α) έλεγχος στασιμότητας βάσει του Augmented Dickey-Fuller (ADF) test, όπου αν διαπιστωθεί πως η χρονοσειρά δεν είναι στάσιμη, αυξάνει κατά μία μονάδα την παράμετρο d , δηλαδή την απλή διαφοροποίηση, μέχρι το όριο που θέσαμε,

β) έλεγχος εποχικότητας, βάσει του Osborn-Chui-Smith-Birchebhall (OCSB) test [99], όπου αν εντοπιστεί ότι η χρονοσειρά διαθέτει έντονη και σταθερή εποχικότητα, δηλαδή οι όποιες αυξομειώσεις επαναλαμβάνονται με τον ίδιο ρυθμό και την ίδια ένταση κάθε έτος, τότε αυξάνει το D κατά μία μονάδα (εποχική διαφοροποίηση),

γ) για κάθε πιθανό συνδυασμό των παραμέτρων (p , q , P και Q), εντός των καθορισμένων από την αρχή ορίων, εφαρμόζει το μοντέλο σε κάθε χρονοσειρά,

δ) για κάθε μοντέλο που δοκιμάζεται, υπολογίζεται το κριτήριο BIC (Bayesian Information Criterion), δίδοντας μεγαλύτερη βαρύτητα στην απλότητα, δηλαδή προτιμώνται οι λιγότεροι παράμετροι, εκτός αν η προσθήκη επιπλέον παραμέτρων μειώνει σημαντικά το σφάλμα. Το μοντέλο με την μικρότερη τιμή του κριτηρίου BIC επιλέγεται ως βέλτιστο.

Μόλις επιλεγούν οι βέλτιστοι παράμετροι του μοντέλου ARIMA, γίνεται η εκπαίδευση κάθε οντότητας με το μοντέλο SARIMAX, με τις εξής τεχνικές παραμέτρους,

α) η αρχική εκτίμηση γίνεται με τον όρο $trend = c$, δηλαδή η μέση τιμή που θα λάμβανε η χρονοσειρά αν αφαιρούσαμε την τάση, την εποχικότητα και τους εξωγενείς παράγοντες.

β) οι παράμετροι `enforce_stationarity` και `enforce_invertibility` ορίζονται σταθερά ως `False`, δηλαδή, δεν επιβάλλονται τεχνητοί περιορισμοί στη στασιμότητα ή στην αναστρεψιμότητα. Με τον τρόπο αυτό το μοντέλο είναι πιο ευέλικτο, επιτρέποντας περιπτώσεις όπου μικρές αποκλίσεις μπορεί να βελτιώνουν την προσαρμογή του.

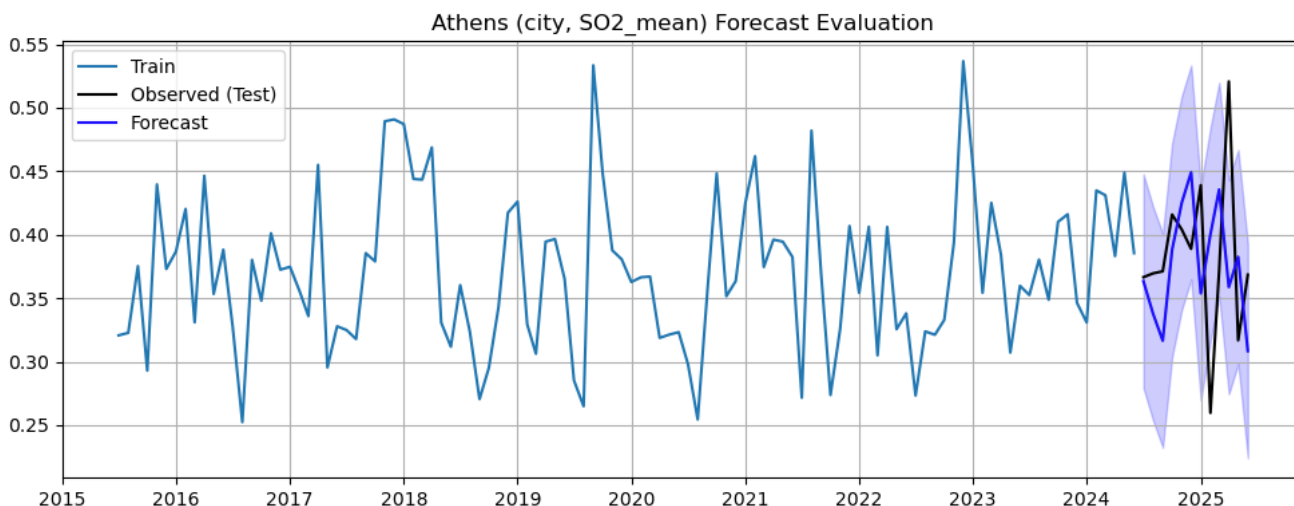
Στις περιπτώσεις των χωρικών επιπέδων, χώρα, περιφέρεια και ήπειρος, αν για οποιοδήποτε λόγο αποτύχει η διαδικασία (πχ. δε συγκλίνει η βελτιστοποίηση, προκύψει `exceprtion`, ή το αποτέλεσμα είναι `NaN` κλπ.) τότε επαναλαμβάνεται η διαδικασία αλλά με διαφορετική μέθοδο (`method = rowell`) και μέγιστες επαναλήψεις 500. Μετά την επανεκτίμηση, ελέγχονται οι διαφορές μεταξύ πραγματικών τιμών και τιμών πρόβλεψης με χρήση του Ljung-Box test. Αν το αποτέλεσμα του ελέγχου είναι 'Poor' (δηλαδή μικρότερο από 0.01) τότε εκκινεί νέα διαδικασία, αυτή τη φορά αλλάζοντας τον όρο `trend` από τιμή 'c' σε τιμή 'ct', δηλαδή γίνεται επανεκτίμηση με γραμμική τάση και επανεκτιμάται με νέο κάθε φορά έλεγχο. Εάν προκύψει βελτίωση τότε επιλέγεται αυτό το μοντέλο, αλλιώς καταγράφεται ως `Fail`.

Δεδομένης της πολυπλοκότητας, η ως άνω διαδικασία καταγράφεται σε σχετικά logs, μαζί με το τελικό αποτέλεσμα για κάθε χρονοσειρά.

Πρόσθετα, σημειώνουμε ότι για σύνολο δεδομένων εκπαίδευσης επιλέχθηκαν οι χρονοσειρές από 1/7/2015 έως 30/6/2024, ενώ ως σύνολο δεδομένων ελέγχου, οι χρονοσειρές από 1/7/2024 μέχρι 30/6/2025. Με τον τρόπο αυτό εξασφαλίσαμε ότι η εκπαίδευση του μοντέλου θα γίνει σε απολύτως ανεξάρτητες χρονοσειρές από εκείνες της διαγνωστικής εφαρμογής του.

Για κάθε οντότητα, υπολογίστηκαν οι μετρικές RMSE, MAE και SMAPE σε %, ενώ σχεδιάστηκε και διάγραμμα με την εκτίμηση της πρόβλεψης. Συνολικά, πέραν των εν λόγω μετρικών, υπολογίστηκαν και άλλες στατιστικές μετρικές όπως IQR κλπ. και σαφέστατα όλο το υποστηρικτικό υλικό είναι στη διάθεση του αναγνώστη.

Ενδεικτικά, για την πόλη της Αθήνας, από την εφαρμογή του μοντέλου SARIMAX προέκυψαν: $RMSE_{train} = 0.053$, $RMSE_{test} = 0.078$, $MAE_{test} = 0.071$, $SMAPE = 18.80\%$ (δηλαδή, κατά μέσο όρο η πρόβλεψη διαφέρει κατά 18.8% από την πραγματική τιμή) και $Ljung-Box\ p\text{-value}\ test = 0.898$.



Γράφημα 63: Αποτέλεσμα εφαρμογής μοντέλου πρόβλεψης SARIMAX για την πόλη της Αθήνας.

Από τα ως άνω αποτελέσματα, η προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης είναι καλή, με αποδεκτή απόδοση στα δεδομένα ελέγχου, υποδηλώνοντας καλή γενίκευση ($RMSE_{train} = 0.053$, ενώ $RMSE_{test} = 0.078$). Ο δείκτης SMAPE παρουσιάζει μέτρια σχετική ακρίβεια, ενώ η υψηλή τιμή του Ljung-Box test επιβεβαιώνει την απουσία αυτοσυσχέτισης στο σφάλμα προσαρμογής. Οπτικά, φαίνεται ότι οι πραγματικές τιμές, κινούνται εντός του διαστήματος prediction interval (95%) με μικρές εξαιρέσεις στις αιχμές, ένδειξη, επιβεβαιώνοντας ότι η εκτίμηση της αβεβαιότητας είναι ρεαλιστική και αποτελεί ισχυρή ένδειξη της στατιστικής εγκυρότητας του μοντέλου.

Η διαδικασία που εφαρμόστηκε στο σύνολο των οντοτήτων του δείγματός μας, εκτελέστηκε αναγκαστικά σε πλατφόρμα νέφους με 32 πυρήνες και παραλληλία, ενώ για την ολοκλήρωση της απαιτήθηκαν περίπου 9 ώρες για κάθε ήπειρο (συνολικά 18 ώρες).

Τα αποτελέσματα τόσο για την Ευρώπη όσο και για τις ΗΠΑ, ανά χωρικό επίπεδο, παρουσιάζονται παρακάτω:

Europe

level	Mean RMSE	Mean MAE	Mean SMAPE
City	0.0822	0.0632	26.31%
Country	0.0697	0.0544	21.08%
Region	0.0523	0.0407	15.39%
Continent	0.0458	0.0341	11.22%

USA

level	Mean RMSE	Mean MAE	Mean SMAPE
City	0.0649	0.0509	23.02%
Country	0.0493	0.0390	17.38%
Region	0.0385	0.0306	12.02%
Continent	0.0304	0.0247	8.98%

Στην Ευρώπη και σε επίπεδο πόλεων, παρατηρούμε ότι το μοντέλο καταφέρνει να προβλέψει με μέση προς ικανοποιητική ακρίβεια τις τιμές των συγκεντρώσεων διοξειδίου του θείου. Η υψηλή τιμή του δείκτη SMAPE φανερώνει από τη μία ότι υπάρχουν περιθώρια βελτίωσης της προβλεπτικής ικανότητας του μοντέλου, από την άλλη, αντανακλά τη φυσική πολυπλοκότητα των αστικών χρονοσειρών, όπου οι συγκεντρώσεις διοξειδίου του θείου επηρεάζονται έντονα από παράγοντες μικρής χρονικής κλίμακας, όπως οι μεταβολές στις μετεωρολογικές συνθήκες και η ένταση των πηγών εκπομπών διοξειδίου του θείου. Όπως έχουμε παρατηρήσει ήδη, όσο προχωράμε σε ανώτερα χωρικά επίπεδα, η εξομάλυνση και η απορρόφηση των τοπικών έντονων μεταβολών, βοηθάει στον περιορισμό των σφαλμάτων πρόβλεψης φτάνοντας σε ποσοστό ακόμη και περίπου 50% βελτίωση από εκείνο των πόλεων.

Στις ΗΠΑ, η εικόνα είναι καλύτερη, με το μοντέλο να αποδίδει ικανοποιητικά, περιγράφοντας με μεγαλύτερη ακρίβεια τη διαχρονική μεταβλητότητα των συγκεντρώσεων διοξειδίου του θείου. Ομοίως η εικόνα γίνεται ακόμη καλύτερη όσο αυξάνεται το χωρικό επίπεδο, καταλήγοντας σε εκείνο της ηπείρου με μονοψήφιο δείκτη SMAPE.

Συνοπτικά, η μέθοδος που ακολουθήσαμε σε σύγκριση με εκείνη του Holt-Winters, είναι σαφώς πιο πολύπλοκη και ιδιαίτερα χρονοβόρα, τόσο σε επίπεδο υλοποίησης όσο και εφαρμογής. Αντιθέτως, παρά τη συνδρομή εξωγενών μετεωρολογικών παραγόντων, δε φάνηκε να προσεγγίζει την επιθυμητή προβλεπτική απόδοση. Σαφώς, υπάρχουν περιθώρια βελτίωσης, ωστόσο, οι τεχνικοί περιορισμοί που αντιμετωπίσαμε, δεν μας επέτρεψαν να εξαντλήσουμε κάθε τέτοια δυνατότητα.

5.3 Εφαρμογή μοντέλου LSTM

Πέραν των κλασσικών μεθόδων πρόβλεψης, προβήκαμε στην ανάπτυξη ενός πολυπαραγοντικού δικτύου LSTM, αξιοποιώντας ταυτόχρονα τη χρονική εξάρτηση και τους εξωγενείς μετεωρολογικούς παράγοντες. Η μεθοδολογία που ακολουθήσαμε παρατίθενται συνοπτικά παρακάτω,

Προεπεξεργασία δεδομένων:

Ως σύνολο δεδομένων χρησιμοποιήθηκαν οι μηνιαίες παρατηρήσεις των συγκεντρώσεων διοξειδίου του θείου των πόλεων, καθώς και οι αντίστοιχες χρονοσειρές των μετεωρολογικών παραγόντων που αντλήσαμε από το ERA5. Για κάθε πόλη δημιουργήθηκε ένας μοναδικός κωδικός (city identifier), ο οποίος μετέπειτα μετατρέπεται σε διάνυσμα, μέσω ενός embedding layer, ώστε αφενός να μπορεί να χρησιμοποιηθεί από το μοντέλο, να ταιριάζει τα χαρακτηριστικά του και αφετέρου να αναγνωρίζει χωρικές ομοιότητες μεταξύ πόλεων, χωρίς να έχει γεωγραφικά ή άλλα στοιχεία για αυτές.

Πριν τη διαδικασία της εκπαίδευσης, όλες οι μεταβλητές κανονικοποιήθηκαν με τη χρήση της μεθόδου Standard Scaler, ώστε να έχουν μηδενικό μέσο και τυπική απόκλιση = 1. Η κανονικοποίηση αυτή εφαρμόστηκε ξεχωριστά για τις μεταβλητές εισόδου και ξεχωριστά για την τιμή στόχο, που δεν είναι άλλος από τις τιμές των συγκεντρώσεων διοξειδίου του θείου.

Για την αποτύπωση της ετήσιας εποχικότητας, ο αριθμητικός μήνας μετασχηματίστηκε σε δύο κυκλικές μεταβλητές, με χρήση ημίτονου και συνημίτονου (sin/cos), επιτρέποντας στο μοντέλο να κατανοεί τη συνεχή, επαναλαμβανόμενη φύση του ετήσιου κύκλου, για παράδειγμα, ότι ο μήνας Ιανουάριος είναι γειτονικός με το μήνα Δεκέμβριο. Παράλληλα ορίσαμε τις μεταβλητές χρονικής καθυστέρησης (lags) σε 1, 2, 3 και 12 μήνες, ώστε το δίκτυο να αξιοποιεί τόσο τη βραχυπρόθεσμη όσο και την ετήσια πληροφορία της χρονοσειράς.

Μετά τις ανωτέρω προπαρασκευαστικές διαδικασίες, το σύνολο των δεδομένων μετατράπηκε σε ακολουθίες μήκους 12 μηνών, ώστε κάθε ακολουθία να χρησιμοποιηθεί για την πρόβλεψη των επόμενων 12 μηνών (horizon = 12). Ουσιαστικά, με τον τρόπο αυτό, παράγονται τρισδιάστατοι πίνακες (σ.σ. εκείνο δηλαδή που απαιτεί το δίκτυο για την εκπαίδευση) που περιλαμβάνουν, την ακολουθία (πχ. για την πόλη Αθήνα, από Ιανουάριο του 2023 μέχρι Δεκέμβριο του 2023), τα χρονικά βήματα και φυσικά τις μεταβλητές (πχ. συγκεντρώσεις διοξειδίου του θείου, ταχύτητα ανέμου, θερμοκρασία κλπ).

Για την εκπαίδευση, την παρακολούθηση της απόδοσης και την τελική αξιολόγηση, το σύνολο δεδομένων χωρίστηκε σε train set, που αποτελείται από το 90% των παρατηρήσεων από 1/7/2015 μέχρι 30/6/2024, validation set (χρησιμοποιείται εσωτερικά για μηχανισμούς που αναφέρουμε παρακάτω), που αποτελείται από το 10% του ίδιου χρονικού διαστήματος και test set που αποτελείται από τις παρατηρήσεις των συγκεντρώσεων διοξειδίου του θείου κατά το χρονικό διάστημα από 1/7/2024 μέχρι 30/6/2025 (τελευταίοι 12 μήνες του δείγματός μας). Με τον τρόπο αυτό εξασφαλίσαμε ότι η αξιολόγηση του μοντέλου θα γίνει σε αθέατα δεδομένα.

Εκπαίδευση:

Το δίκτυο αποτελείται από δύο διαδοχικά bidirectional LSTM layers των 64 νευρώνων το κάθε ένα. Τα layers αυτά επιτρέπουν στο μοντέλο να αναγνωρίζει μοτίβα τόσο από το παρελθόν προς το παρόν όσο και από το παρόν παρελθόν, ώστε να ενισχυθεί κατά το δυνατό η κατανόηση της εποχικής δομής και των χρονικών εξαρτήσεων. Μεταξύ των δύο επιπέδων εφαρμόσαμε dropout (0.2), ώστε να απενεργοποιείται τυχαία το 20% των νευρώνων, σε κάθε επανάληψη (epoch). Με τον τρόπο αυτό ελαττώνεται ο κίνδυνος υπερπροσαρμογής και βελτιώνεται η γενίκευση του μοντέλου σε νέα δεδομένα (σ.σ. σε κάθε επανάληψη χρησιμοποιείται ένα ελαφρώς διαφορετικό υποσύνολο νευρώνων). Επισημαίνεται πως στο στάδιο της πρόβλεψης, απενεργοποιείται ο υπόψη μηχανισμός, προκειμένου να συμμετέχουν όλοι οι νευρώνες του δικτύου.

Η έξοδος των layers του δικτύου συνδυάζεται με το διάνυσμα χωρικής ενσωμάτωσης της εκάστοτε πόλης και στη συνέχεια περνάει από ένα Dense layer, το οποίο είναι υπεύθυνο για την πρόβλεψη των καθορισμένων μηνών, όπου στην περίπτωση μας είναι 12.

Η εκπαίδευση του μοντέλου πραγματοποιήθηκε με τον αλγόριθμο Adam (Adaptive Moment Estimation), με ρυθμό εκμάθησης 10^{-4} , ενώ ως συνάρτηση απώλειας χρησιμοποιήθηκε η MSE (Mean Squared Error). Η MSE υπολογίζει το μέσο όρο των τετραγώνων των διαφορών μεταξύ πραγματικών τιμών και προβλέψεων και εκφράζει το συνολικό σφάλμα του μοντέλου. Το αποτέλεσμα της συνάρτησης απώλειας χρησιμοποιείται από τον αλγόριθμο Adam για την ενημέρωση των βαρών μέσω backpropagation. Ο αλγόριθμος Adam προσαρμόζει την ποσότητα της μεταβολής των βαρών ανά βήμα, βάσει ενός εξομαλυμένου μέσου όρου των βαθμίδων και της μεταβλητότητάς τους, ώστε η εκπαίδευση να είναι κατά το δυνατό πιο σταθερή και αποτελεσματική. Ο ρυθμός εκμάθησης (learning rate) καθορίζει το μέγεθος του βήματος με το οποίο ενημερώνονται τα βάρη σε κάθε στάδιο της εκπαίδευσης.

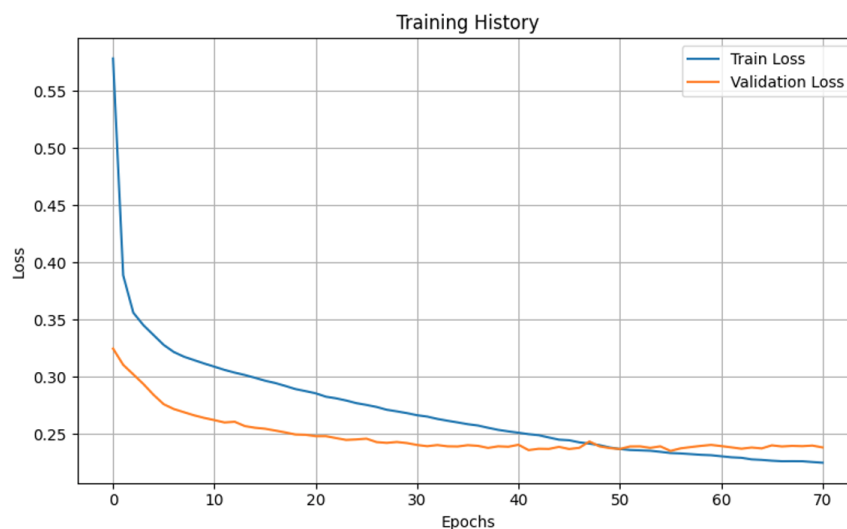
Για τη βελτίωση της σταθερότητας της εκπαίδευσης, εφαρμόστηκαν οι μηχανισμοί α) early stopping, όπου τερματίζει τη διαδικασία της εκπαίδευσης όταν δε βελτιώνεται η απόδοση του μοντέλου για 15 εποχές και β) ReduceLROnPlateau, όπου μειώνεται αυτόματα ο ρυθμός μάθησης όταν η πρόοδος επιβραδύνεται (σ.σ. όταν το validation loss παύσει να μειώνεται). Πρόσθετα, μέσω του μηχανισμού ModelCheckpoint, σε κάθε βήμα αποθηκεύεται η καλύτερη εκδοχή του μοντέλου, εξασφαλίζοντας έτσι ότι τα τελικά αποτελέσματα αντιστοιχούν στη μέγιστη απόδοση.

Εξαγωγή μετρικών απόδοσης

Μετά την εκπαίδευση, το μοντέλο αξιολογήθηκε στο test set (12 τελευταίοι μήνες του δείγματός μας), υπολογίζοντας τις μετρικές RMSE, MAE, SMAPE, R^2 , που δείχνει ποιο ποσοστό της διακύμανσης των πραγματικών τιμών εξηγείται από το μοντέλο και skill score που απεικονίζει την απόδοση του μοντέλου σε σχέση με ένα μοντέλο αναφοράς (persistence ή baseline model που προβλέπει ότι η τιμή του επόμενου μήνα θα είναι ίση με την τελευταία παρατηρούμενη). Επισημαίνουμε πως οι ίδιες μετρικές αποθηκεύονται ξεχωριστά για κάθε πόλη και για κάθε ορίζοντα πρόβλεψης, μαζί βέβαια και με το σχετικό log της όλης διαδικασίας. Ο χρόνος εκτέλεσης, ανήλθε σε περίπου 30 λεπτά για κάθε ήπειρο.

Από την εφαρμογή του δικτύου LSTM, στις μηνιαίες παρατηρήσεις των συγκεντρώσεων διοξειδίου του θείου στην Ευρώπη και τις ΗΠΑ, προέκυψαν τα ακόλουθα:

Για την απεικόνιση της διαδικασίας εκπαίδευσης του δικτύου για της πόλεις της Ευρώπη, παραθέτουμε παρακάτω το σχετικό διάγραμμα εκπαίδευσης, που απεικονίζει τη μεταβολή της συνάρτησης απώλειας (loss function) MAE, στα train και validation sets.



Γράφημα 64: Διαδικασία εκπαίδευσης δικτύου LSTM σε δεδομένα Ευρώπης.

Η καμπύλη εκπαίδευσης, όπως απεικονίζεται στο παραπάνω διάγραμμα, παρουσιάζει ομαλή και ταχεία πτώση του σφάλματος προσαρμογής (loss) τόσο για τα δεδομένα εκπαίδευσης (Train Loss) όσο και για το σύνολο επικύρωσης (Validation Loss) κατά τις πρώτες περίπου 20 εποχές, με απότομη βελτίωση από περίπου 0.57 σε 0.28. Από την 20ή εποχή και μετά, η μείωση του σφάλματος γίνεται προοδευτικά πιο αργή, με την τελική σύγκλιση και των δύο καμπυλών να επιτυγχάνεται μετά την 48ή εποχή σε περίπου 0.23. Η σχεδόν πλήρης επικάλυψη της πορτοκαλί καμπύλης (Validation) με τη μπλε (Train), χωρίς ανοδική απόκλιση ή συνεχή διεύρυνση της απόστασής τους, αποτελεί χαρακτηριστικό απουσίας υπερπροσαρμογής (overfitting) και της πολύ καλής γενίκευσης σε άγνωστα δεδομένα παρατηρήσεων. Παρά την εφαρμογή του μηχανισμού early stopping, η εκπαίδευση ολοκληρώθηκε στις 70 εποχές, καθώς το validation loss παρουσίαζε σταθερή, αν και μικρή, βελτίωση μέχρι το τέλος της εκπαίδευσης.

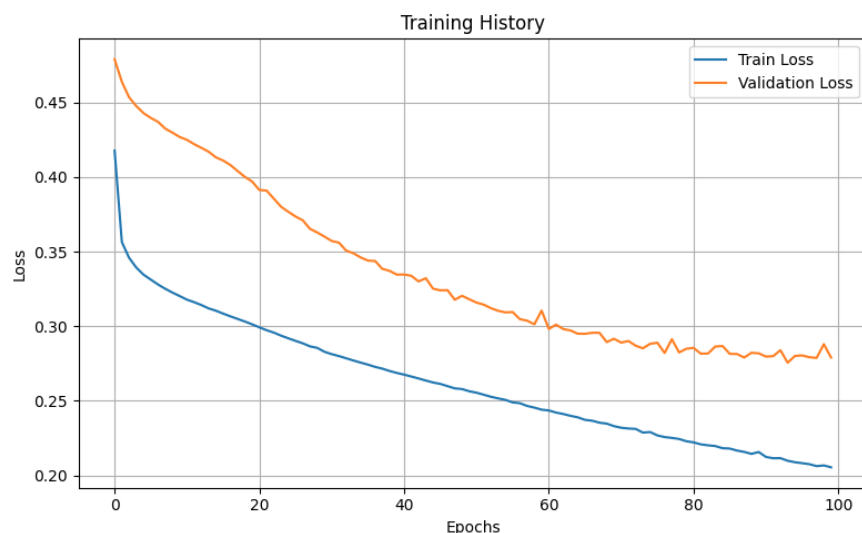
Προς επίρρωση της επιτυχούς εκπαίδευσης του μοντέλου, τα αποτελέσματα των προβλέψεων ανά ορίζοντα, για την Ευρώπη, παρατίθενται στον κάτωθι πίνακα,

Europe

Horizon	Mean RMSE	Mean MAE	Mean SMAPE (%)	R ²	Skill Score
1	0,0423	0,0357	15,85	0,0529	0,2987
2	0,0469	0,0412	18,09	-0,1438	0,2164
3	0,0442	0,0365	16,33	-0,2103	0,2493
4	0,0393	0,0329	14,51	0,1908	0,3103
5	0,0417	0,0327	13,99	-0,0485	0,3118
6	0,0496	0,0386	16,74	-0,4784	-0,0414
7	0,0333	0,0233	10,45	0,1448	0,3440
8	0,0282	0,0213	9,33	0,6635	0,5280
9	0,0298	0,0216	9,36	0,6232	0,5131
10	0,0384	0,0257	10,71	0,4335	0,3786
11	0,0277	0,0206	8,72	0,7507	0,6110
12	0,0387	0,0304	12,64	0,5028	0,5076

από όπου παρατηρούμε, αρχικά μια χαμηλή απόδοση στους ορίζοντες 1 - 6 και σημαντική βελτίωση σε μακροπρόθεσμο ορίζοντα (7 - 11 μηνών). Ο δείκτης RMSE εμφανίζει ελάχιστες τιμές στους μήνες 8 - 11 (0.027 - 0.029 DU), ενώ ο δείκτης SMAPE διατηρείται σε χαμηλά επίπεδα με ανώτερο το 18%. Ο συντελεστής προσδιορισμού R² είναι αρνητικός στους πρώτους 6 μήνες (εκτός του πρώτου και του τέταρτου), υποδηλώνοντας ότι το μοντέλο δεν εξηγεί ικανοποιητικά τη διακύμανση των πραγματικών τιμών, αλλά εκτοξεύεται σε 0.75 τον 11^ο μήνα, που καταφέρνει να εξηγήσει το 75% της διακύμανσης. Ο δείκτης δεξιοτήτας (skill score) παραμένει θετικός σε όλους τους ορίζοντες (με εξαίρεση τον ορίζοντα των 6 μηνών), με μέγιστη τιμή 0.61 κατά τον 11^ο μήνα, επιβεβαιώνοντας ότι το μοντέλο υπερτερεί σημαντικά εκείνου της αναφοράς σε μακροχρόνια πρόβλεψη. Η βελτίωση αυτή φαίνεται να οφείλεται στην ικανότητα του LSTM να συλλαμβάνει εποχικά μοτίβα και μακροχρόνια τάση, που κυριαρχούν πέραν των 6 μηνών. Συμπερασματικά, το μοντέλο είναι ιδιαίτερα αποτελεσματικό στη μεσομακροπρόθεσμη πρόβλεψη των συγκεντρώσεων διοξειδίου του θείου, αλλά λιγότερο κατάλληλο για βραχυπρόθεσμη πρόβλεψη.

Από την εφαρμογή του δικτύου LSTM στις πόλεις των ΗΠΑ, προέκυψαν τα ακόλουθα,



Γράφημα 65: Διαδικασία εκπαίδευσης δικτύου LSTM σε δεδομένα των ΗΠΑ.

Σε αντίθεση με ότι παρατηρήσαμε στην Ευρώπη, η εφαρμογή του μοντέλου LSTM στις πόλεις των ΗΠΑ, εμφανίζει καλή αρχική σύγκλιση αλλά ήπια ένδειξη υπερπροσαρμογής (overfitting). Η ‘train loss’ μειώνεται ομαλά καταλήγοντας στην τιμή 0.21, δείχνοντας τις δυνατότητες του μοντέλου να προσαρμόζεται καλά σε βραχυπρόθεσμα γεγονότα. Ωστόσο, η ‘validation loss’, παρουσιάζει μία ήπια υπερπροσαρμογή περίπου κατά την 60^η εποχή, όμως επανέρχεται σύντομα και τελικά καταλήγει στο 0.27, έναντι 0.21 της ‘training loss’. Γενικά φαίνεται ότι υπάρχει ανάγκη για περαιτέρω μελέτη στην εύρεση των βέλτιστων παραμέτρων του μοντέλου, ωστόσο, ακόμη και σε αυτή τη μορφή, διατηρεί μια αποδεκτή ικανότητα να προβλέπει τις συγκεντρώσεις του διοξειδίου του θείου στις ΗΠΑ.

Προς επιβεβαίωση της ικανοποιητικής λειτουργίας του δικτύου LSTM, προβήκαμε στην πρόβλεψη των συγκεντρώσεων διοξειδίου του θείου σε ορίζοντα 12 μηνών, σε αθέατα δεδομένα, από όπου προέκυψαν τα ακόλουθα,

USA

Horizon	Mean RMSE	Mean MAE	Mean SMAPE (%)	R ²	Skill Score
1	0,0802	0,0580	15,58	0,0965	0,2960
2	0,0652	0,0439	11,57	0,4173	0,4279
3	0,0822	0,0628	16,74	0,1246	0,3371
4	0,0782	0,0640	17,30	0,2041	0,3651
5	0,0771	0,0643	17,30	0,2272	0,3446
6	0,0789	0,0604	16,61	0,0088	0,2493
7	0,0797	0,0646	18,64	0,1785	0,0902
8	0,0536	0,0432	13,44	0,3265	0,2517
9	0,0561	0,0401	12,34	0,2185	0,1980
10	0,0543	0,0426	13,44	0,2901	0,2392
11	0,0585	0,0490	15,76	0,3320	0,2068
12	0,0546	0,0482	15,74	0,4160	0,3184

Τόσο ο συντελεστής προσδιορισμού R² όσο και το skill score, παραμένουν θετικά σε όλους τους ορίζοντες πρόβλεψης, ενώ και ο δείκτης SMAPE παρουσιάζει σημαντική μείωση μεταξύ 8^{ου} και 11^{ου} ορίζοντα. Οι τιμές του δείκτη RMSE μειώνονται σταθερά φτάνοντας στις 0.0546 DU στο 12^ο ορίζοντα. Γενικά, παρατηρούμε μία ελαφρώς μικρότερη αλλά πιο σταθερή απόδοση έναντι των αποτελεσμάτων της Ευρώπης. Δηλαδή, το ίδιο δίκτυο LSTM έχει καλή προσαρμογή στα δεδομένα, αλλά σίγουρα επιδέχεται βελτίωση, προκειμένου να εντοπίζει πιο αποτελεσματικά τις τοπικές διακυμάνσεις.

Συνολικά, από την εφαρμογή των τριών μοντέλων, στις παρατηρήσεις των συγκεντρώσεων διοξειδίου του θείου στις πόλεις της Ευρώπης και των ΗΠΑ, προέκυψαν οι ακόλουθες μετρικές απόδοσης,

Europe

Holt Winters			SARIMAX			LSTM		
Mean RMSE	Mean MAE	Mean SMAPE	Mean RMSE	Mean MAE	Mean SMAPE	Mean RMSE	Mean MAE	Mean SMAPE
0.0668	0.0510	21.15%	0.0822	0.0632	26.31%	0.0383	0.0300	13.06%

Στα δεδομένα των Ευρωπαϊκών πόλεων, το δίκτυο LSTM εμφάνισε την καλύτερη απόδοση, με δεύτερο κατά σειρά το μοντέλο Holt-Winters και τρίτο το SARIMAX. Το δίκτυο LSTM, κατάφερε να προσαρμοστεί πολύ καλά στα δεδομένα, με σχετικά σύντομο χρόνο εκτέλεσης, κάνοντάς το αξιόλογη επιλογή για χρήση σε παραγωγικά συστήματα παρακολούθησης των συγκεντρώσεων διοξειδίου του θείου.

Οι αντίστοιχοι δείκτες από την εφαρμογή των μοντέλων στα δεδομένα των πόλεων των ΗΠΑ,

USA

Holt Winters			SARIMAX			LSTM		
Mean RMSE	Mean MAE	Mean SMAPE	Mean RMSE	Mean MAE	Mean SMAPE	Mean RMSE	Mean MAE	Mean SMAPE
0.0590	0.0450	20.06%	0.0649	0.0509	23.02%	0.0682	0.0534	15.37%

δεν έχουν πολύ σημαντικές διαφορές μεταξύ τους. Το μοντέλο Holt - Winters εμφάνισε μεν μικρότερη τιμή ρίζας μέσου τετραγωνικού σφάλματος (RMSE), ωστόσο, το δίκτυο LSTM εμφάνισε το μικρότερο μέσο σχετικό σφάλμα (SMAPE). Συνεπώς, θα μπορούσαμε να πούμε ότι το μοντέλο Holt - Winters είναι πιο αξιόπιστο όταν μας ενδιαφέρει η συνολική εικόνα, ενώ το δίκτυο LSTM είναι πιο ακριβές σε τοπικό επίπεδο πόλης.

Κεφάλαιο 6. Σύνοψη μελέτης και γενικά συμπεράσματα

Η παρούσα διπλωματική εργασία επικεντρώθηκε στη μελέτη της μακροχρόνιας συμπεριφοράς του διοξειδίου του θείου, σε 410 πόλεις της Ευρώπης και 511 των ΗΠΑ, καθώς και στην αξιολόγηση μεθόδων πρόβλεψης της μελλοντικής του πορείας. Αρχικά, παρουσιάστηκαν οι χημικές ιδιότητες του διοξειδίου του θείου, οι κύριες ανθρωπογενείς και φυσικές πηγές εκπομπές του, οι εφαρμογές του στη βιομηχανία και οι επιπτώσεις που προκαλεί τόσο στο περιβάλλον όσο και στην ανθρώπινη υγεία. Στη συνέχεια, αναπτύχθηκε το θεωρητικό υπόβαθρο των μεθόδων ανάλυσης χρονοσειρών και πρόβλεψης που χρησιμοποιήσαμε και παρουσιάστηκαν τα αποτελέσματα της μελέτης μας.

Στην προσπάθειά μας να προσεγγίσουμε ολιστικά το θέμα της εργασίας, βασιστήκαμε σε επιλεγμένα συγγράμματα από διάφορες πηγές και το διαδίκτυο. Τεχνικά, αντλήσαμε ένα μεγάλο σύνολο δεδομένων, κυρίως από το Data Center της NASA και το European Centre for Medium-Range Weather Forecasts (ECMWF), η ανάλυση του οποίου πραγματοποιήθηκε με χρήση κώδικα Python, σε πλατφόρμα νέφους.

Οι απαιτήσεις τόσο σε θεωρητική τεκμηρίωση όσο και σε τεχνική επεξεργασία ήταν σημαντικές και εκτείνονταν από την κατανόηση των ατμοσφαιρικών διεργασιών του διοξειδίου του θείου έως τη διαχείριση και μετατροπή μεγάλου όγκου δεδομένων διαφόρων τύπων και από πολλαπλές πηγές.

Αφού συγκεντρώθηκαν τα απαραίτητα δεδομένα, καταρτίστηκαν οι αντίστοιχες χρονοσειρές και εφαρμόστηκε σε αυτές μία τροποποιημένη εκδοχή της μεθόδου STL, ώστε να διαχωριστούν επιτυχώς το εποχικό μοτίβο, η μακροχρόνια τάση και η βραχυπρόθεσμη μεταβλητότητα. Οι υπόψη συνιστώσες εξετάστηκαν με εξειδικευμένες τεχνικές, ανά χωρικό επίπεδο (πόλη, χώρα, περιφέρεια και ήπειρο) και παρατέθηκαν τα σχετικά αποτελέσματα, τόσο αριθμητικά όσο και διαγραμματικά για την πληρέστερη ερμηνεία τους.

Σε γενικές γραμμές, οι συγκεντρώσεις διοξειδίου του θείου, στις περισσότερες πόλεις του δείγματός μας, εμφάνισαν μέτρια εποχικότητα με ήπια ανοδική τάση. Στην Ευρώπη καταγράφηκαν μεγαλύτερες διαφοροποιήσεις μεταξύ των πόλεων, ενώ στις ΗΠΑ η εικόνα ήταν πιο σταθερή, καταδεικνύοντας την ανομοιογένεια μεταξύ των δύο ηπείρων τόσο σε γεωγραφικό επίπεδο όσο και σε πηγές εκπομπών διοξειδίου του θείου.

Η μελέτη μας επεκταθήκαμε και στη διερεύνηση της επιρροής των συγκεντρώσεων διοξειδίου του θείου από εξωγενείς παράγοντες, όπως μετεωρολογικούς και το όζον. Όπως καταδείχθηκε, ο παράγοντας με την εντονότερη συσχέτιση ήταν η νεφοκάλυψη και μάλιστα με διαφοροποίηση του προσήμου, ανάλογα την απόσταση από το έδαφος. Όσον αφορά το όζον, δε φάνηκε να υπάρχει κάποια σημαντική συσχέτιση μεταξύ των δύο ρύπων, παρά μόνον κάποιες ελάχιστες ενδείξεις ότι οι συγκεντρώσεις του όζοντος ίσως να δημιουργούν συνθήκες που ευνοούν την εμφάνιση υψηλότερων τιμών συγκεντρώσεων διοξειδίου του θείου μετά από παράθυρο 5-6 μηνών. Η ερμηνεία ενός τέτοιου γεγονότος εκφεύγει του πλαισίου της παρούσης διπλωματικής εργασίας, καθώς δεν είμαστε σε θέση να γνωρίζουμε εάν η ωρίμανση συγκεκριμένων συνθηκών εντός τέτοιου χρονικού διαστήματος, μπορούν πράγματι να επηρεάζουν τις συγκεντρώσεις διοξειδίου του θείου ή θα πρέπει να αποδοθεί στην τυχαιότητα.

Ένα εξίσου σημαντικό πεδίο μελέτης, υπήρξε η επίδραση των περιοριστικών μέτρων της πανδημίας COVID-19. Τα αποτελέσματα της ανάλυσής μας, έδειξαν μία σημαντική διαφορά στη συμπεριφορά των συγκεντρώσεων διοξειδίου του θείου, κατά τις περιόδους εφαρμογής περιοριστικών μέτρων, μεταξύ Ευρώπης και ΗΠΑ. Στη μεν πρώτη παρατηρήθηκε μείωση των συγκεντρώσεων διοξειδίου του θείου το 2021, στη δε δεύτερη παρόμοια μείωση καταγράφηκε το 2020. Μάλιστα, από τα στοιχεία που αντλήσαμε, μέσω του Oxford COVID-19 Government Response Tracker, φάνηκε ότι η ένταση των μέτρων που ελήφθησαν στις δύο ηπείρους και κατά τις δύο περιόδους εφαρμογής περιορισμών (2020, 2021), ήταν παρόμοια.

Τελικό στάδιο της μελέτης μας, ήταν η ανάπτυξη μοντέλων για την πρόβλεψη της πορείας των συγκεντρώσεων διοξειδίου του θείου. Προς τούτο, καταρτίστηκαν 3 μοντέλα διαφορετικού τύπου και τρόπου λειτουργίας (Holt-Winters, SARIMAX, LSTM), με το νευρωνικό δίκτυο LSTM να εμφανίζει την καλύτερη μεσομακροπρόθεσμη προβλεπτική ικανότητα, ιδίως για τις Ευρωπαϊκές πόλεις.

Περιορισμοί μελέτης

Πέραν των τεχνικών προκλήσεων που αντιμετωπίσαμε, εξαιτίας του αρχικού όγκου των δεδομένων, κρίνεται σκόπιμη η αναφορά μας στην ποιότητα των δεδομένων και τις ενέργειες στις οποίες καταφύγαμε για τη διασφάλισή της. Συγκεκριμένα, μετά την εφαρμογή των κανόνων που συστήνει η NASA, το 46% των παρατηρήσεων κατέστη αναξιόπιστο. Προς αντικατάστασή του, χρησιμοποιήθηκαν μετρήσεις, από 3 διαφορετικά σύνολα, προερχόμενα μεν από την ίδια πηγή, που όμως έχουν υποστεί επεξεργασία από το σύστημα reanalysis της NASA Global Model and Assimilation Office (GMAO). Τόσο οι μονάδες μέτρησης όσο και η χωρική ανάλυση, έπρεπε να υποστούν κατάλληλους μετασχηματισμούς, ώστε να είναι συμβατές με το αρχικό σύνολο δεδομένων. Παρά τους δειγματοληπτικούς ελέγχους που πραγματοποιήθηκαν, δεν μπορεί να αποκλειστεί το ενδεχόμενο, σε κάποιες περιπτώσεις, οι μετρήσεις των συγκεντρώσεων διοξειδίου του θείου (και του όζοντος) να μη διατηρούν την ίδια βάση και κλίμακα σε όλο το σύνολο δεδομένων. Παρά ταύτα, η τυχόν ύπαρξη τέτοιων περιπτώσεων θα πρέπει να θεωρηθεί περιορισμένη.

Προτάσεις για μελλοντική έρευνα

Από τα αποτελέσματα της παρούσης μελέτης προκύπτουν κατευθύνσεις που θα μπορούσαν να αποτελέσουν αντικείμενο μελλοντικής έρευνας, με στόχο την πληρέστερη κατανόηση της μεταβλητότητας των συγκεντρώσεων διοξειδίου του θείου και κατ' επέκταση την ανάπτυξη τεχνικών για τον περιορισμό των εκπομπών του. Ας μην ξεχνάμε ότι το διοξείδιο του θείου αποτελεί, πέραν των άλλων, κι έναν έμμεσο παράγοντα περιορισμού του έντονου φαινομένου του θερμοκηπίου, όποτε η πλήρως ελεγχόμενη εκπομπή του είναι το κλειδί για την προστασία του περιβάλλοντος και της ανθρώπινης υγείας.

Ένας σημαντικός άξονας αφορά στη χρήση δεδομένων υψηλότερης χρονικής ανάλυσης, όπως ημερήσιων ή ακόμη και ωριαίων παρατηρήσεων, τόσο για το διοξείδιο του θείου όσο και για ένα ευρύτερο σύνολο μετεωρολογικών και περιβαλλοντικών παραγόντων. Μέσω αυτής της προσέγγισης, μπορούν να αξιολογηθούν οι μετρήσεις σε τοπικό επίπεδο, η επιρροή που ενδέχεται να ασκείται σε γειτονικές περιοχές (μέσω της διασποράς και μεταφοράς

του) και να εξεταστεί η αποτελεσματικότητα των μέτρων που λαμβάνονται για τον περιορισμό των εκπομπών του.

Προς αυτή την κατεύθυνση μάλιστα, θα ήταν εξαιρετικό σε ακαδημαϊκό επίπεδο να πραγματοποιηθούν μελέτες με την ταυτόχρονη συμμετοχή ειδικών όχι μόνο από το πεδίο της πληροφορικής αλλά και τεχνικών με εξειδίκευση στη μηχανική της ατμόσφαιρας και στη χημεία περιβάλλοντος.

Γενικό Συμπέρασμα

Από όσα προεκτήθηκαν, αναδείχθηκε η πολυπλοκότητα στην ανάλυση και ερμηνεία της συμπεριφοράς ενός περιβαλλοντικού ρύπου όπως το διοξείδιο του θείου. Η μελέτη μας βασίστηκε σε εξειδικευμένες τεχνικές, προκειμένου να κατανοήσουμε την πορεία του στο χρόνο, τις διαφορές μεταξύ χωρικών οντοτήτων και τις αλληλεπιδράσεις με μετεωρολογικούς παράγοντες και άλλους ρύπους.

Σε γενικές γραμμές, προέκυψε ότι την τελευταία δεκαετία οι συγκεντρώσεις του διοξειδίου του θείου εμφανίζουν μία σταθερή προς ήπια ανοδική τάση, με περιορισμένο αριθμό έκτακτων γεγονότων που όμως δεν επηρεάζουν σημαντικά την εν γένει συμπεριφορά του. Αξιοσημείωτες είναι οι διαφορές που εντοπίστηκαν μεταξύ της Ευρώπης και των ΗΠΑ, τόσο σε επίπεδο εποχικότητας και ρυθμού μεταβολής όσο και συνεκτικότητας μεταξύ των οντοτήτων που τις απαρτίζουν.

Τέλος, οι πολιτικές που λαμβάνονται διαχρονικά, φαίνεται πως αποδίδουν, όταν εφαρμόζονται σωστά, εκτίμηση η οποία προκύπτει από τη συνολική εικόνα που αποκομίσαμε καθ' όλη τη διάρκεια εκπόνησης της παρούσας μελέτης.

Βιβλιογραφία – Παραπομπές

- [1] J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 3rd ed. Wiley, 2016.
- [2] S. J. Smith, J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. D. Arias, “Anthropogenic sulfur dioxide emissions: 1850–2005,” *Atmospheric Chem. Phys.*, vol. 11, no. 3, pp. 1101–1116, 2011.
- [3] B. J. Finlayson-Pitts and J. N. Pitts, *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications*. Academic Press, 2000.
- [4] World Health Organization, “WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide,” 2021.
- [5] S. A. Carn, V. E. Fioletov, C. A. McLinden, C. Li, and N. A. Krotkov, “A decade of global volcanic SO₂ emissions measured from space,” *Sci. Rep.*, vol. 7, p. 44095, 2017.
- [6] M. O. Andreae and P. J. Crutzen, “Atmospheric aerosols: Biogeochemical sources and role in atmospheric chemistry,” *Science*, vol. 276, no. 5315, pp. 1052–1058, 1997.
- [7] D. G. Streets, C. Yu, Y. Wu, and M. Chin, “Global sulfur dioxide emissions and the contribution of China,” *Atmos. Environ.*, vol. 43, no. 22–23, pp. 3019–3022, 2009.
- [8] V. Eyring *et al.*, “Transport impacts on atmosphere and climate: Shipping,” *Atmos. Environ.*, vol. 44, no. 37, pp. 4735–4771, 2010.
- [9] European Environment Agency, “European Union emission inventory report 1990–2018 under the UNECE Convention on Long-range Transboundary Air Pollution (LRTAP).” 2020. [Online]. Available: <https://www.eea.europa.eu/publications/european-union-emission-inventory-report-1990-2018>
- [10] V. Masson-Delmotte *et al.*, Eds., *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021. doi: 10.1017/9781009157896.
- [11] M. Cheng and *et al.*, “Secondary inorganic aerosols formation during haze episodes at an urban site in Beijing, China,” *Atmospheric Chem. Phys.*, vol. 16, pp. 10743–10758, 2016.
- [12] Concawe, “Review of recent health effect studies with sulphur dioxide,” Concawe, Report No. 16/4, 2016.
- [13] P. Orellano, N. Quaranta, J. Reynoso, B. Balbi, and J. Vasquez, “Short-term exposure to sulphur dioxide (SO₂) and all-cause and respiratory mortality: A systematic review and meta-analysis,” *Environ. Res.*, vol. 192, p. 110279, 2021.
- [14] R. O’Brien and *et al.*, “Short-term association between sulfur dioxide and mortality across 399 cities,” *Environ. Health Perspect.*, vol. 131, no. 4, 2023.
- [15] E. Samoli and *et al.*, “Short-term effects of SO₂ on mortality in 12 European cities: Results from the APHEA project,” *Epidemiology*, vol. 17, no. 1, pp. 75–81, 2006.
- [16] C. A. Pope and D. W. Dockery, “Health effects of fine particulate air pollution: lines that connect,” *J. Air Waste Manag. Assoc.*, vol. 56, no. 6, 2006, doi: 10.1080/10473289.2006.10464485.
- [17] S. A. Meo and *et al.*, “Effect of air pollutants (PM_{2.5}, PM₁₀, SO₂) on lung function,” *J. Clin. Med.*, vol. 13, no. 2, p. 377, 2024.
- [18] R. J. Delfino, C. Sioutas, and S. Malik, “Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health,” *Environ. Health Perspect.*, vol. 113, no. 8, pp. 934–946, 2005, doi: 10.1289/ehp.7938.
- [19] R. D. Brook and *et al.*, “Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association,” *Circulation*, vol. 121, no. 21, pp. 2331–2378, 2010.
- [20] U. Pöschl, “Atmospheric aerosols: Composition, transformation, climate and health effects,” *Angew. Chem. Int. Ed.*, vol. 44, no. 46, pp. 7520–7540, 2005.
- [21] M. L. Block and L. Calderón-Garcidueñas, “Air pollution: Mechanisms of neuroinflammation and CNS disease,” *Trends Neurosci.*, vol. 32, no. 9, pp. 506–516, 2009.
- [22] L. Calderón-Garcidueñas and *et al.*, “Air pollution, cognitive deficits and brain abnormalities: A pilot study with children and dogs,” *Brain Cogn.*, vol. 68, no. 2, pp. 117–127, 2008.
- [23] M. Elser and *et al.*, “Global sulfuric acid production and emissions,” *Nat. Geosci.*, vol. 9, pp. 782–785, 2016.
- [24] S. L. Taylor, N. A. Higley, and R. K. Bush, “Sulphites in foods: Uses, analytical methods, residues, fate, exposure, assessment and hypersensitivity,” *Adv. Food Res.*, vol. 30, pp. 1–76, 1986.

- [25] Essential Chemical Industry, “Sulfuric Acid.” 2012. [Online]. Available: <https://www.essentialchemicalindustry.org/chemicals/sulfuric-acid.html>
- [26] ChemAnalyst, “Sulphuric Acid Market Analysis: 2015–2035.” 2024. [Online]. Available: <https://www.chemanalyst.com/industry-report/sulphuric-acid-market-604>
- [27] KBV Research, “Global Sulfuric Acid Market Size, Share & Trends Analysis Report.” 2023. [Online]. Available: <https://www.kbvresearch.com/sulfuric-acid-market/>
- [28] Grand View Research, “Sulfuric Acid Market Size, Share & Trends Analysis Report 2023–2030.” 2023. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/sulfuric-acid-market>
- [29] M. L. Bell and D. L. Davis, “Reassessment of the lethal London fog of 1952: Novel indicators of acute and chronic consequences of acute exposure to air pollution,” *Environ. Health Perspect.*, vol. 109, no. Suppl 3, pp. 389–394, 2001.
- [30] A. Ciocco and D. J. Thompson, “A follow-up of Donora ten years after: Methodology and findings,” *Am. J. Public Health*, vol. 51, no. 2, pp. 155–164, 1961.
- [31] European Union, “Directive 2001/80/EC of the European Parliament and of the Council of 23 October 2001 on the limitation of emissions of certain pollutants into the air from large combustion plants.” 2001. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32001L0080>
- [32] US EPA, “Overview of the Clean Air Act and Air Pollution.” 2019. [Online]. Available: <https://www.epa.gov/clean-air-act-overview>
- [33] Z. Klimont, S. J. Smith, and J. Cofala, “The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions,” *Environ. Res. Lett.*, vol. 8, no. 1, p. 014003, 2013.
- [34] United Nations Economic Commission for Europe (UNECE), “Convention on Long-range Transboundary Air Pollution (LRTAP).” 1979. [Online]. Available: <https://unece.org/environment-policyair/lrtap>
- [35] D. Archer and R. Pierrehumbert, Eds., *The Warming Papers: The Scientific Foundation for the Climate Change Forecast*. Wiley-Blackwell, 2011.
- [36] J. T. Kiehl and K. E. Trenberth, “Earth’s annual global mean energy budget,” *Bull. Am. Meteorol. Soc.*, vol. 78, no. 2, pp. 197–208, 1997, doi: 10.1175/1520-0477(1997)078%3C0197:EAGMEB%3E2.0.CO;2.
- [37] Α. Φωτιάδη, “Φαινόμενο του Θερμοκηπίου και Κλιματικές Μεταβολές, Ενότητα 3: Παράγοντες προσδιορισμού της ανθρωπογενούς επίδρασης στο ενεργειακό ισοζύγιο του πλανήτη.” Πανεπιστήμιο Πατρών, 2014.
- [38] R. J. Charlson *et al.*, “Climate forcing by anthropogenic aerosols,” *Science*, vol. 255, no. 5043, pp. 423–430, 1992, doi: 10.1126/science.255.5043.423.
- [39] M. O. Andreae, C. D. Jones, and P. M. Cox, “Strong present-day aerosol cooling implies a hot future,” *Nature*, vol. 435, pp. 1187–1190, 2005, doi: 10.1038/nature03671.
- [40] B. A. Albrecht, “Aerosols, cloud microphysics, and fractional cloudiness,” *Science*, vol. 245, no. 4923, pp. 1227–1230, 1989, doi: 10.1126/science.245.4923.1227.
- [41] S. Twomey, “The influence of pollution on the shortwave albedo of clouds,” *J. Atmospheric Sci.*, vol. 34, no. 7, pp. 1149–1152, 1977, doi: 10.1175/1520-0469(1977)034%3C1149:TIOPOT%3E2.0.CO;2.
- [42] S. I. Rasool and S. H. Schneider, “Atmospheric carbon dioxide and aerosols: Effects of large increases on global climate,” *Science*, vol. 173, no. 3992, pp. 138–141, 1971, doi: 10.1126/science.173.3992.138.
- [43] B. H. Samset *et al.*, “Climate impacts from a removal of anthropogenic aerosol emissions,” *Geophys. Res. Lett.*, vol. 45, no. 2, pp. 1020–1029, 2018, doi: 10.1002/2017GL076079.
- [44] D. M. Westervelt, L. W. Horowitz, V. Naik, J.-C. Golaz, and D. L. Mauzerall, “Radiative forcing and climate response to projected 21st century aerosol decreases,” *Atmospheric Chem. Phys.*, vol. 15, no. 22, pp. 12681–12703, 2015, doi: 10.5194/acp-15-12681-2015.
- [45] G. Myhre, D. Shindell, J. Pongratz, and *et al.*, “Multi-model simulations of aerosol and ozone radiative forcing due to anthropogenic emission changes during the period 1990–2015,” *Atmospheric Chem. Phys.*, vol. 17, pp. 2709–2720, 2017, doi: 10.5194/acp-17-2709-2017.
- [46] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Chapman & Hall/CRC, 2003.
- [47] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley, 2016.
- [48] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts, 2018. [Online]. Available: <https://otexts.com/fpp2/>
- [49] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, 4th ed. Springer, 2017.

- [50] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “STL: A Seasonal-Trend Decomposition Procedure Based on Loess,” *J. Off. Stat.*, vol. 6, no. 1, pp. 3–73, 1990.
- [51] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “STL: A Seasonal-Trend Decomposition Procedure Based on Loess,” *J. Off. Stat.*, vol. 6, pp. 3–73, 1990.
- [52] C. C. Holt, “Forecasting Seasonals and Trends by Exponentially Weighted Averages,” Office of Naval Research, Research Memorandum, 1957.
- [53] P. R. Winters, “Forecasting Sales by Exponentially Weighted Moving Averages,” *Manag. Sci.*, vol. 6, no. 3, pp. 324–342, 1960.
- [54] J. D. Hamilton, *Time Series Analysis*. Princeton University Press, 1994.
- [55] C. Chatfield and H. Xing, *The Analysis of Time Series: An Introduction*, 7th ed. Chapman and Hall/CRC, 2019.
- [56] D. A. Dickey and W. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *J. Am. Stat. Assoc.*, vol. 74, no. 366a, pp. 427–431, 1979.
- [57] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, 1994, doi: 10.1109/72.279181.
- [58] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <https://www.deeplearningbook.org/>
- [59] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [60] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with LSTM,” *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000, doi: 10.1162/089976600300015015.
- [61] J. Brownlee, *Long Short-Term Memory Networks with Python*. Machine Learning Mastery, 2017. [Online]. Available: <https://machinelearningmastery.com/lstms-with-python/>
- [62] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Netw.*, vol. 18, no. 5–6, pp. 602–610, July 2005, doi: 10.1016/j.neunet.2005.06.042.
- [63] A. S. Foundation, “Apache Parquet: A Columnar Storage Format.” 2025. [Online]. Available: <https://parquet.apache.org/>
- [64] Global Modeling and Assimilation Office (GMAO), “MERRA-2 tavg1_2d_adg_Nx: 2d, 1-Hourly, Time-averaged, Single-Level, Assimilation, Aerosol Diagnostics (extended) V5.12.4.” Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, 2015. doi: 10.5067/HM00OHQBHKTP.
- [65] Global Modeling and Assimilation Office (GMAO), “MERRA-2 tavg1_2d_aer_Nx: 2d, 1-Hourly, Time-averaged, Single-Level, Assimilation, Aerosol Diagnostics V5.12.4.” Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, 2015. doi: 10.5067/KLICLTZ8EM9D.
- [66] Global Modeling and Assimilation Office (GMAO), “MERRA-2 tavg1_2d_chm_Nx: 2d, 1-Hourly, Time-averaged, Single-Level, Assimilation, Carbon Monoxide and Ozone Diagnostics V5.12.4.” Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, 2015. doi: 10.5067/3RQ5YS674DGQ.
- [67] R. W. Sinnott, “Virtues of the Haversine,” *Sky Telesc.*, vol. 68, no. 2, p. 159, 1984.
- [68] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. OTexts, 2021. [Online]. Available: <https://otexts.com/fpp3/>
- [69] G. M. Ljung and G. E. P. Box, “On a measure of lack of fit in time series models,” *Biometrika*, vol. 65, no. 2, 1978.
- [70] D. A. Dickey and W. A. Fuller, “Distribution of the Estimators for Autoregressive Time Series With a Unit Root,” *J. Am. Stat. Assoc.*, vol. 74, no. 366, pp. 427–431, 1979.
- [71] J. Durbin and G. S. Watson, “Testing for Serial Correlation in Least Squares Regression: I,” *Biometrika*, vol. 37, no. 3/4, pp. 409–428, 1950.
- [72] H. B. Mann, “Nonparametric Tests Against Trend,” *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945.
- [73] M. G. Kendall, *Rank Correlation Methods*. C. Griffin, 1948. [Online]. Available: <https://books.google.gr/books?id=hiBMAAAAMAAJ>
- [74] P. M. Kendall, *Rank correlation methods*, 5th ed. in Charles Griffin Book. London, England: Oxford University Press, 1990.

- [75] R. O. Gilbert, *Statistical Methods for Environmental Pollution Monitoring*. New York: John Wiley & Sons, 1987.
- [76] “EPA - Air Pollutant Emissions Trends Data - National Tier 1 CAPS Trends.” [Online]. Available: https://www.epa.gov/system/files/other-files/2025-04/national_tier1_caps_21feb2025.xlsx
- [77] H. Theil, “A rank-invariant method of linear and polynomial regression analysis,” *Ned. Tijdschr. Psychol.*, vol. 5, pp. 386–392, 1950.
- [78] P. K. Sen, “Estimates of the Regression Coefficient Based on Kendall’s Tau,” *J. Am. Stat. Assoc.*, vol. 63, no. 324, pp. 1379–1389, 1968.
- [79] “theilslopes, scipy.stats.” [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.theilslopes.html>
- [80] “Copernicus Atmosphere Monitoring Service - CAMS - Fagradalsfjall - Iceland.” [Online]. Available: <https://www.copernicus.eu/en/media/image-day-gallery/sulfur-dioxide-emitted-fagradalsfjall-volcanic-eruption-iceland>
- [81] Μ. Φιλιππάκης, *Θεωρία Πιθανοτήτων & Στοιχεία Στατιστικής Ανάλυσης*. Τσότρας, 2019.
- [82] W. H. Kruskal and W. A. Wallis, “Use of Ranks in One-Criterion Variance Analysis,” *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952.
- [83] M. G. Kendall and B. B. Smith, “The Problem of m Rankings,” *Ann. Math. Stat.*, vol. 10, no. 3, pp. 275–287, 1939.
- [84] A. Field, J. Miles, and Z. Field, *Discovering Statistics Using R*. London: SAGE Publications, 2012.
- [85] O. J. Dunn, “Multiple comparisons using rank sums,” *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964, doi: 10.1080/00401706.1964.10490181.
- [86] S. Holm, “A Simple Sequentially Rejective Multiple Test Procedure,” *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.
- [87] H. Levene, “Robust tests for equality of variances,” in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, Ed., Stanford: Stanford University Press, 1960, pp. 278–292.
- [88] W. M. Organization and R. Sneyers, *On the statistical analysis of series of observations. Technical Note No. 143, WMO-No. 415*. Geneva: World Meteorological Organization, 1990.
- [89] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal detection of changepoints with a linear computational cost,” *J. Am. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012, doi: 10.1080/01621459.2012.737745.
- [90] “costRbf, ruptures py library.” [Online]. Available: <https://centre-borelli.github.io/ruptures-docs/user-guide/costs/costrbf/>
- [91] N. R. Zhang and D. O. Siegmund, “A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data,” *J. Stat. Plan. Inference*, vol. 136, no. 11, pp. 3959–3968, 2007, doi: 10.1016/j.jspi.2006.07.018.
- [92] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” *Ann. Math. Stat.*, vol. 18, no. 1, pp. 50–60, 1947, doi: 10.1214/aoms/1177730491.
- [93] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3–4, pp. 591–611, 1965, doi: 10.1093/biomet/52.3-4.591.
- [94] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 6th ed. Hoboken, NJ: John Wiley & Sons, 2014.
- [95] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. New York: Springer, 2016.
- [96] T. Hale *et al.*, “A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker),” *Nat. Hum. Behav.*, vol. 5, no. 4, pp. 529–538, 2021, doi: 10.1038/s41562-021-01079-8.
- [97] C. Spearman, “The proof and measurement of association between two things,” *Am. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904, doi: 10.2307/1412159.
- [98] “het_white, py library statsmodels.stats.diagnostic.” [Online]. Available: https://www.statsmodels.org/dev/generated/statsmodels.stats.diagnostic.het_white.html
- [99] D. R. Osborn, A. P. L. Chui, J. Smith, and C. R. Birchenhall, “Seasonality and the order of integration for consumption,” *Oxf. Bull. Econ. Stat.*, vol. 50, no. 4, pp. 361–377, 1988, doi: 10.1111/j.1468-0084.1988.mp50004002.x.