



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

Δημιουργία προγνωστικών μοντέλων ατμοσφαιρικής
ρύπανσης στην ευρύτερη περιοχή της πόλης του Πεκίνο,
Κίνα, με την ανάπτυξη και εφαρμογή Τεχνητών
Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Φαζάκη Παναγιώτη

Επιβλέποντες :

Δρ. Μουστρής Κωνσταντίνος

Καθηγητής

Δρ. Σπυρόπουλος Γεώργιος

Ε.Δ.Ι.Π.

Αθήνα, Ιούλιος 2024



University of West Attica

School of Engineering

Department of Mechanical Engineering

Development of air pollution forecasting models applying
Artificial Neural Networks in the greater area of Beijing city,
China

Bachelor thesis
of
Fazakis Panagiotis

Supervisors :

Dr. Moustris Konstantinos

Professor

Dr. Spyropoulos George

Faculty Member

Τριμελής Εξεταστική Επιτροπή

<i>A/A</i>	ΕΠΩΝΥΜΟ ΟΝΟΜΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
<i>1</i>	ΜΟΥΣΤΡΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ (Καθηγητής)	
<i>2</i>	ΣΠΥΡΟΠΟΥΛΟΣ ΓΕΩΡΓΙΟΣ (Ε.ΔΙ.Π.)	
<i>3</i>	ΖΑΦΕΙΡΑΚΗΣ ΔΗΜΗΤΡΙΟΣ (Επίκουρος Καθηγητής)	

Copyright © - 10/8/2024

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Φαζάκης Παναγιώτης του Γεωργίου, με αριθμό μητρώου 19392159 φοιτητής του Πανεπιστημίου Δυτικής Αττικής, της Σχολής Μηχανικών, του τμήματος Μηχανολόγων Μηχανικών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



ΕΥΧΑΡΙΣΤΙΕΣ

Φτάνοντας στο τέλος εκπόνησης της συγγραφής Διπλωματικής Εργασίας, θα ήθελα να ευχαριστήσω του υπεύθυνους καθηγητές μου Μουστή Κωνσταντίνο και Σπυρόπουλο Γεώργιο για την εξαιρετική συνεργασία και γνώσεις που προσέφεραν καθ' όλη τη διάρκεια. Τέλος θα ήθελα να ευχαριστήσω όλα τα άτομα τα οποία στάθηκαν δίπλα μου σε αυτή τη διαδικασία, επιτρέποντας μου να ανταπεξέλθω πλήρως στις υποχρεώσεις αυτής.

ΠΕΡΙΛΗΨΗ

Η συνεχώς αυξανόμενη βιομηχανοποίηση ορισμένων περιοχών του πλανήτη σε συνδυασμό με την ταυτόχρονη υποβάθμιση του φυσικού περιβάλλοντος αποτελούν ανησυχητικά φαινόμενα ιδιαίτερα όσον αφορά τον τομέα της ανθρώπινης υγείας. Η συγκέντρωση σωματιδίων αεροδυναμικής διαμέτρου 2.5 μm (PM_{2.5}) και 10 μm (PM₁₀), τα οξείδια του Αζώτου (NO_x), το μονοξείδιο του άνθρακα (CO), του διοξειδίου του θείου (SO₂), αλλά και του Όζοντος (O₃) χρήζουν συνεχούς παρακολούθησης, καθώς αποτελούν κύριες αιτίες για πληθώρα νοσημάτων. Βάσει της ύπαρξης θεσμοθετημένων ορίων, από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ.), για τη συγκέντρωση καθενός εκ των προαναφερθέντων ρύπων στην ατμόσφαιρα κρίνεται απαραίτητη η ανάπτυξη προγνωστικών συστημάτων που θα έχουν την ικανότητα συσχέτισης των εκάστοτε μετεωρολογικών δεδομένων με τις συγκεντρώσεις των παραπάνω ρύπων. Στην παρούσα διπλωματική εργασία γίνεται επιτυχώς η προσπάθεια πρόβλεψης των παραγόμενων αυτών ρύπων στην ευρύτερη περιοχή του Πεκίνο, στην Κίνα, με χρήση υπολογιστικών μοντέλων. Κατά την εκπόνηση της εργασίας αναπτύχθηκε σημαντικός αριθμός Τεχνητών Νευρωνικών Δικτύων , για την δημιουργία των οποίων αξιοποιήθηκαν “ανοιχτά-ελεύθερα” μετεωρολογικά δεδομένα και δεδομένα ατμοσφαιρικής ρύπανσης, καθώς και ποικίλες υπολογιστικές μέθοδοι για την συμπλήρωση τους, όπου αυτά ήταν ελλιπή. Τέλος γίνεται στατιστική αξιολόγηση των μοντέλων και έλεγχος σχετικά με την ορθότητα των αποτελεσμάτων.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Τεχνητά Νευρωνικά Δίκτυα ,Ατμοσφαιρική Ρύπανση, Μοντέλο Πρόγνωσης , Πρόγνωση Ρύπων.

ABSTRACT

The ever-increasing industrialization of certain areas of the planet combined with the simultaneous degradation of the natural environment are alarming phenomena, especially in the field of human health. The concentration of Particulate Matter with an aerodynamic diameter of $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$) and $10\ \mu\text{m}$ (PM_{10}), nitrogen oxides (NO_x), carbon monoxide (CO), sulfur dioxide (SO_2), and ozone (O_3) needs constant monitoring, as they consist the main cause of many diseases. Based on the existence of statutory limits, by the World Health Organization (WHO), for the concentration of each of the aforementioned pollutants in the atmosphere, it is considered necessary to develop forecasting systems that will have the ability to correlate the current meteorological data with the concentrations of the above pollutants. In this thesis, the attempt to predict the produced pollutants in the wider area of Beijing, in China, is successfully carried using computer models. During the preparation of the work, a significant number of Artificial Neural Networks were developed, for the creation of which open-access meteorological data and air pollution data were used, as well as various computational methods to complete them, where these were incomplete. Finally, a statistical evaluation of the models is carried out and a check on the correctness of the results.

KEYWORDS

Artificial Neural Networks, Atmospheric Pollution, Predictive Model, Pollutant Forecast.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	8
ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ	8
ABSTRACT.....	9
KEYWORDS.....	9
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	11
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ.....	12
1. ΕΙΣΑΓΩΓΗ.....	13
2. ΑΤΜΟΣΦΑΙΡΙΚΗ ΡΥΠΑΝΣΗ.....	15
2.1. Βασικές Έννοιες.....	15
2.2. Επιπτώσεις Ατμοσφαιρικής Ρύπανσης.....	18
3. ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	23
3.1. Βασικές Έννοιες.....	23
3.2. Τεχνητά Νευρωνικά Δίκτυα	25
3.2.1. Ορισμός & Βασικές Έννοιες.....	25
3.2.2. Τεχνητά Νευρωνικά Δίκτυα Πολλαπλών Στρωμάτων (MLP)	27
3.2.3. Levenberg – Marquardt και Bayesian Regularization.....	31
3.2.4. Μοντέλα Πρόγνωσης.....	32
4. ΜΕΘΟΔΟΛΟΓΙΑ	34
4.1. Περιοχή Μελέτης.....	34
4.2. Κάλυψη Κενών Στοιχείων.....	36
4.2.1. Μέθοδος 1 : Χρήση ήδη υπαρχουσών τιμών και αξιοποίηση εντολών Microsoft Excel.	43
4.2.2. Μέθοδος 2 : Χρήση κώδικα εντός προγραμματιστικού περιβάλλοντος MATLAB.....	45
4.2.3. Μέθοδος 3 : Χρήση γραμμικής παρεμβολής	47
4.3. Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων.....	50
4.3.1. Προετοιμασία δεδομένων	50
4.3.2. Δημιουργία σεναρίων.....	51
4.3.3. Λογισμικά και Υποδομές	54
5. ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	55
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	68
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	69

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

<i>A/A</i>	ΤΙΤΛΟΣ ΠΙΝΑΚΑ	ΣΕΛ
1	Χημικές εξισώσεις αλληλεπίδρασης ρύπων και ουσιών που συμμετέχουν στο σχηματισμό O ₃ και NO ₂	17
2	Οριακές τιμές-επίπεδα συγκέντρωσης ατμοσφαιρικών ρύπων με βάση την οδηγία GB 3095-2012 του State Council of the People's Republic of China.	20
3	Οριακές τιμές και τιμές στόχοι της συγκέντρωσης ατμοσφαιρικών ρύπων για την προστασία της δημόσιας υγείας και του περιβάλλοντος, με βάση την Ευρωπαϊκή Οδηγία 2008/50/EK και το Υπουργείο Περιβάλλοντος, Ενέργειας και Κλιματικής Αλλαγής (ΚΥΑ ΗΠ 14122/549/Ε103, ΦΕΚ 488Β/30.3.11).	21
4	Οριακές τιμές-επίπεδα συγκέντρωσης ατμοσφαιρικών ρύπων με βάση τη νομοθεσία του Παγκόσμιου Οργανισμού Υγείας (WHO).	22
5	Συναρτήσεις εκπαίδευσης ΤΝΔ και ο συνήθης τρόπος αξιοποίησής τους.	28
6	Ενδεικτικό πλήθος κενών στοιχείων στις μεταβλητές εισόδου για την Περιοχή 3.	34
7a	Πλήθος κενών στοιχείων, βάσει σεναρίου 1, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.	36
7b	Πλήθος κενών στοιχείων, βάσει σεναρίου 2, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.	37
7c	Πλήθος κενών στοιχείων, βάσει σεναρίου 3, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.	38
7d	Πλήθος κενών στοιχείων, βάσει σεναρίου 4, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.	39
7e	Πλήθος κενών στοιχείων, βάσει σεναρίου 5, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.	40
7f	Πλήθος κενών στοιχείων, βάσει σεναρίου 6, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.	41
8	Πίνακας αντιστοίχισης φοράς ανέμου με τον αντίστοιχο ακέραιο αριθμό.	43
9	Ποσοστό επιτυχίας της 1ης μεθόδου κάλυψης κενών στοιχείων.	43
10	Αριθμοί προσθαφαίρεσης για κάθε ρύπο και μετεωρολογικό παράγοντα.	43
11	Συσχέτιση επίδρασης περιοχών μελέτης ως προς τη φορά του ανέμου και αναγραφή εγγύτητας κατά αύξουσα σειρά.	45
12	Γραμμικές εξισώσεις συσχέτισης περιοχών και οι αντίστοιχοι δείκτες R ² .	47
13	Πίνακας μέσων συντελεστών συσχέτισης.	50
14	Πίνακας σεναρίων και πλήθους δεδομένων για κάθε ένα από αυτά.	51
15	Πίνακας μεταβλητών εκπαίδευσης κάθε ΤΝΔ.	53
16a	Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα PM _{2.5} .	56
16b	Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα PM _{2.5} .	56
16c	Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα PM ₁₀ .	56
16d	Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα PM ₁₀ .	56
16e	Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα SO ₂ .	57
16f	Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα SO ₂ .	57
16g	Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα NO ₂ .	57
16h	Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα NO ₂ .	57
16i	Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα CO.	58
16j	Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα CO.	58
16k	Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα O ₃ .	58
16l	Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα O ₃ .	58

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

<i>A/A</i>	ΤΙΤΛΟΣ ΠΙΝΑΚΑ	ΣΕΛ
1	Χημική αλληλεξάρτηση ρύπων και ουσιών που συμμετέχουν στο σχηματισμό O ₃ και NO ₂	17
2	Πρόωροι θάνατοι λόγω έκθεσης σε σωματίδια PM _{2.5} σε κρατικό επίπεδο για τα έτη 2005 – 2021	19
3	Επιπτώσεις της όξινης βροχής σε αγάλματα (γυψοποίηση), στο φυσικό περιβάλλον και σε βαφή αυτοκινήτου	21
4	Γραφική αναπαράσταση του αλγόριθμου ID3, σχετικά με τον έλεγχο ρίσκου (Risk Assessment) σχετικά με την παραχώρηση δανείου σε πελάτη τράπεζας (Credit Analysis)	25
5	Σχηματική αναπαράσταση νευρώνα ενός ΤΝΔ.	26
6	Σχηματική αναπαράσταση ενός ΤΝΔ πολλαπλών στρωμάτων (MLP) με δομή (1-15-10-1).	30
7	Λογικό διάγραμμα της διαδικασίας εξαγωγής των προγνωστικών μοντέλων.	34
8	Απεικόνιση μέρους της βόρειο-ανατολικής Κίνας μέσω Google Maps (αριστερά) και της πόλης του Πεκίνο όπου φαίνονται οι επιμέρους περιοχές μελέτης (δεξιά).	36
9	Γραφική απεικόνιση της αλληλεπίδρασης των επιμέρους περιοχών εντός ακτίνας 10 χιλιομέτρων μεταξύ τους. Τα γκρι νούμερα αντιστοιχούν στις περιοχές και τα πράσινα στους ακέραιους αριθμούς που αντιπροσωπεύουν την αντίστοιχη διεύθυνση αέρα.	45
10	Ενδεικτική απεικόνιση των γραμμικών σχέσεων των συγκεντρώσεων των PM _{2.5} μεταξύ των 4 περιοχών κοντά στην περιοχή 3 και της περιοχής 3.	48
11	Γραφική αναπαράσταση μέσων ωριαίων τιμών συγκέντρωσης για κάθε τα PM ₁₀ της περιοχής 1 σε χρονικό διάστημα 5 ημερών πριν και μετά της ημερομηνίας αναφοράς	54
12	Διάγραμμα Box & Whisker για τον δείκτη IA για όλα τα σενάρια εκπαίδευσης.	61
13	Διάγραμμα Box & Whisker για τον δείκτη SI για όλα τα σενάρια εκπαίδευσης.	61
14	Ποσοστό εμφάνισης κάθε Α.Ε. σε βέλτιστο σενάριο για (α) Στατιστικούς Δείκτες αξιοπιστίας και (β) Στατιστικούς Δείκτες ψευδών προβλέψεων	62
15	Ποσοστό εμφάνισης κάθε σεναρίου μεταβλητών εισόδου, ως βέλτιστο σενάριο για (α) Στατιστικούς Δείκτες αξιοπιστίας και (β) Στατιστικούς Δείκτες ψευδών προβλέψεων	63
16a	Διάγραμμα ROC για κάθε σενάριο και κάθε περιοχή για τα αιωρούμενα σωματίδια αεροδυναμικής διαμέτρου 2.5 και 10 μm.	63
16b	Διάγραμμα ROC για κάθε σενάριο και κάθε περιοχή για τα όζον και μονοξείδιο του άνθρακα	64
16c	Διάγραμμα ROC για κάθε σενάριο και κάθε περιοχή για διοξείδια του θείου και διοξείδια του αζώτου.	64

1. ΕΙΣΑΓΩΓΗ

Ο σημερινός τρόπος ζωής, όπως αυτός έχει διαμορφωθεί από τις υπάρχουσες κοινωνικές και οικονομικές συνθήκες, απαιτεί τη συνεχή έκθεση του ανθρώπου σε μια ατμόσφαιρα με δυσμενείς επιπτώσεις για την υγεία του. Η ραγδαία εκβιομηχάνιση πολλών χωρών, η οποία επικροτεί την άκρατη ανάπτυξη εις βάρος του περιβάλλοντος και οι γρήγοροι ρυθμοί που αυτή επιβάλλει αποσπών την προσοχή του κάθε πολίτη από θέματα τα οποία δρουν επιβαρυντικά στην υγεία του. Βιομηχανίες, εργοστάσια παραγωγής ενέργειας, αυτοκίνητα και γενικότερα μέσα μεταφοράς αποτελούν βασικές πηγές μόλυνσης του ατμοσφαιρικού αέρα, οδηγώντας σε μια όλο και πιο υποβαθμισμένη ατμόσφαιρα. Τα πορίσματα ερευνών σε όλο τον κόσμο συγκλίνουν ως προς τις καταστροφικές συνέπειες, που έχει ο ατμοσφαιρικός αυτός αέρας τόσο στον άνθρωπο, όσο και στις αγροτικές καλλιέργειες αλλά και τα κτήρια. Η συνεχής αυτή υποβάθμιση του κλίματος καθιστά επιτακτική την εύρεση ενός μέσου πρόβλεψης της ποιότητας της ατμόσφαιρας, καθώς και αντιμετώπισης της παραγωγής των ατμοσφαιρικών ρύπων.

Ο ατμοσφαιρικός αέρας αποτελεί ένα μίγμα αέριων όπως N_2 78%, O_2 21%, Ar 0.93%, CO_2 0.031% και άλλα, τα οποία όταν παρατηρούνται σε αυτές τις αναλογίες, γίνεται λόγος για καθαρό ατμοσφαιρικό αέρα. Ωστόσο στα αστικά περιβάλλοντα κάτι τέτοιο αποτελεί στην καλύτερη περίπτωση ένα σπάνιο φαινόμενο, καθώς οι προσθήκες ξένων ουσιών στο μίγμα αυτό αποτελούν απόρροια του σημερινού τρόπου ζωής. Αιωρούμενα σωματίδια αεροδυναμικής διαμέτρου 2.5 και 10 μm , αυξημένες συγκεντρώσεις O_3 και άλλων χημικών ουσιών αποτελούν βασικούς παράγοντες ρύπανσης. Η ατμοσφαιρική ρύπανση αποτελεί ένα φαινόμενο τόσο φυσικό όσο και ανθρωπογενές, με το δεύτερο όμως να γίνεται όλο και πιο έντονο. Φυσικά μια αλλοίωση του φυσικού περιβάλλοντος σε τέτοια επίπεδα είναι βέβαιο πως θα φέρει μια πληθώρα προβλημάτων, τα οποία θα αποβούν επιβαρυντικά τόσο για την ανθρώπινη υγεία (αναπνευστικά και καρδιαγγειακά προβλήματα, νευρολογικές και συμπεριφορικές διαταραχές και άλλα), όσο και για τομείς όπως οι υποδομές (όξινη βροχή και φαινόμενο γυψοποίησης), αλλά και ο αγροτικός τομέας, με τη σταδιακή μείωση των διαθέσιμων σιτηρών.

Ωστόσο η σημερινή τεχνολογική ανάπτυξη, έχει εξοπλίσει τον επιστημονικό κλάδο με πληθώρα εργαλείων, τα οποία με την κατάλληλη χρήση τους μπορούν να συμβάλλουν στην αντιμετώπιση της ατμοσφαιρικής ρύπανσης και στη βαρύτητα των επιπτώσεων αυτής στην ανθρώπινη υγεία. Η μηχανική μάθηση παρότι δεν αποτελεί μια καινούρια ανακάλυψη, έχει γνωρίσει ραγδαία ανάπτυξη την τελευταία δεκαετία, επιτρέποντας την δημιουργία αλγορίθμων, ειδικά σχεδιασμένων για την πρόβλεψη ρύπων. Η διαδικασία της συσχέτισης μεταξύ ρύπων και μετεωρολογικών συνθηκών αποτελεί μια επίπονη μαθηματική διαδικασία, καθώς οι σχέσεις είναι μη γραμμικές και υπό περιπτώσεις μεταβαλλόμενες. Ωστόσο με τη βοήθεια συγκεκριμένων αλγορίθμων η εξαγωγή των κρυφών, υπό άλλες περιστάσεις, σχέσεων γίνεται με λιγότερη δυσκολία. Με τη χρήση Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ), ένα από τα πολλά παρακλάδια της μηχανικής μάθησης, η διαδικασία πρόγνωσης ρύπων γίνεται μια απλή διαδικασία, χωρίς αυτό να σημαίνει ότι δεν εγκυμονεί κινδύνους. Η προ-επεξεργασία των δεδομένων εισόδου, η εισαγωγή τους με ορθό τρόπο στα

προγράμματα εκπαίδευσης και η αρχιτεκτονική δομή των ΤΝΔ επηρεάζουν άμεσα την ποιότητα των αποτελεσμάτων.

Η περιοχή μελέτης είναι η πόλη Beijing, China και συγκεκριμένα 12 επιμέρους περιοχές εντός αυτής. Τα δεδομένα μελέτης απαρτίζονται από ωριαίες μετρήσεις για την περίοδο 2013 έως και 2017, η ανάλυση των οποίων γίνεται στην **Ενότητα 4**. Η παρούσα διπλωματική εργασία καλείται να δημιουργήσει και να αξιολογήσει ένα ικανό πλήθος ΤΝΔ, προκειμένου να επιλεγθούν αυτά τα οποία παρουσιάζουν την βέλτιστη ικανότητα πρόβλεψης. Η συγκεκριμένη εργασία αποσκοπεί μέσω της εκπαίδευσης και της αξιολόγησης των παραγόμενων ΤΝΔ, να απομονώσει τον ακριβέστερο και απλούστερο συνδυασμό μεταβλητών εισόδου, καθώς και να δημιουργήσει ένα καθολικό μοντέλο πρόγνωσης ρύπων. Ύστερα της εξαγωγής των προγνώσεων, γίνεται ο έλεγχος ορθότητας των δεδομένων, αξιοποιώντας οκτώ (8) στατιστικούς δείκτες, προκειμένου να εξασφαλιστεί η ορθότητα των αποτελεσμάτων.

Στα παρακάτω κεφάλαια θα γίνει μια εισαγωγή στις βασικές συνιστώσες του προβλήματος και αναλυτική παρουσίαση της μεθοδολογίας, βάσει της οποίας έγινε η προ-επεξεργασία των δεδομένων εισόδου, η εκπαίδευση των νευρωνικών και η αξιολόγησή τους.

2. ΑΤΜΟΣΦΑΙΡΙΚΗ ΡΥΠΑΝΣΗ

2.1. Βασικές Έννοιες

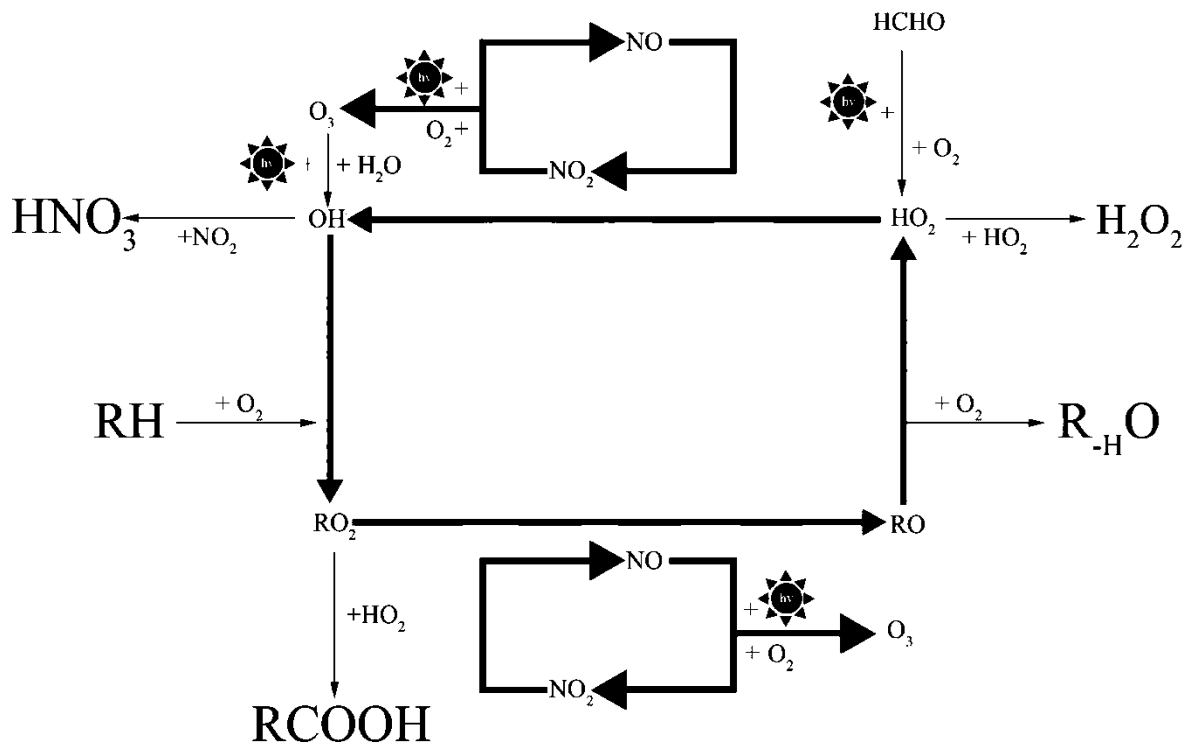
Ως ατμοσφαιρική ρύπανση ορίζεται «η παρουσία στο περιβάλλον ρύπων, δηλαδή κάθε είδους ουσιών, θορύβου, ακτινοβολίας ή άλλων μορφών ενέργειας σε ποσότητα, συγκέντρωση ή διάρκεια που μπορούν να προκαλέσουν αρνητικές επιδράσεις στην υγεία, στους ζωντανούς οργανισμούς και στα οικοσυστήματα ή υλικές ζημιές και γενικά να καταστήσουν το περιβάλλον ακατάλληλο για τις επιθυμητές χρήσεις του.» [1]. Η παρούσα διπλωματική εργασία απευθύνεται σε ένα μικρό φάσμα των ατμοσφαιρικών ρύπων, συγκεκριμένα στις συγκεντρώσεις των αιωρούμενων σωματιδίων με αεροδυναμική διάμετρο 2.5 μm (PM_{2.5}) και 10 μm (PM₁₀), στις συγκεντρώσεις του διοξειδίου του θείου (SO₂), του διοξειδίου του αζώτου (NO₂), του μονοξειδίου του άνθρακα (CO) και του όζοντος (O₃). Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας [2], [3] οι παραπάνω ρύποι ορίζονται ως εξής :

- 1) Σωματίδια (Particulate Matter) : «Εισπνεόμενα σωματίδια τα οποία συνίστανται από θειικό, νιτρικά, αμμωνία, χλωριούχο νάτριο, μαύρο άνθρακα, ορυκτή σκόνη ή νερό.». Ανάλογα με την αεροδυναμική τους διάμετρο διαχωρίζονται σε PM_{2.5} και PM₁₀ με αεροδυναμικές διαμέτρους ίσες με 2.5 και 10 μm αντίστοιχα. Αποτελούν εξαιρετικά επικίνδυνο ρύπο για τον άνθρωπο, όπως θα αναφερθεί παρακάτω.
- 2) NO₂ : Αποτελεί πρωτογενή ρύπο και βασικό παράγοντα στην δημιουργία του Όζοντος (O₃). Παράλληλα επηρεάζει αρνητικά το αναπνευστικό σύστημα, ενώ η ικανότητα μετασχηματισμού του σε διαφορετικές ουσίες, όπως το νιτρικό οξύ και μετέπειτα σε άλας εξουδετέρωσης αμμωνίου, το καθιστούν ιδιαίτερα επικίνδυνο.
- 3) CO : Αποτελεί έναν άοσμο, άχρωμο, άγευστο και μη ερεθιστικό αέριο ρύπο, ο οποίος είναι εξαιρετικά επιβλαβής για τον άνθρωπο προκαλώντας ακόμη και θάνατο. Κύρια πηγή συνιστούν οι μηχανές καύσης όταν πραγματοποιούν ατελή καύση.
- 4) SO₂ : Αποτελεί πρωτογενή ρύπο με καίριες επιπτώσεις στην υγεία, καθώς οι συγκεντρώσεις του σχετίζονται με την αύξηση περιστατικών νοσηλείας λόγω ασθματικού επεισοδίου στις μονάδες έκτακτης ανάγκης, τους πρόωρους θανάτους και την θνησιμότητα λόγω αναπνευστικών επιπλοκών. Το θείο (S) αποτελεί βασικό στοιχείο στη σύσταση του γαιάνθρακα και κάποιων «βαρέων» καυσίμων, υλικά τα οποία χρησιμοποιούνται σε μεγάλες ποσότητες σε εργοστάσια παραγωγής ηλεκτρικής ενέργειας, αλλά και σε βιομηχανίες γενικότερα. Κατά την επεξεργασία και καύση του γαιάνθρακα το θείο που εμπεριέχεται εντός αυτού μεταβαίνει σε μια σειρά μετασχηματισμών όπου παράγεται ένα αέριο μίγμα αποτελούμενο από διοξείδιο και τριοξείδιο του Θείου (SO₂, SO₃)[4]. Η βασική χημική εξίσωση που οδηγεί στο σχηματισμό SO₃ είναι η οξείδωση του SO₂: $SO_2 + O + M \rightarrow SO_3 + M$. Όσο το τριοξείδιο του θείου βρίσκεται κοντά στην εστία φωτιάς, συμβαίνει ραγδαίος μετασχηματισμός του σε διοξείδιο του θείου (SO₂) και σε ρίζα υδροξυ σουλφονυλίου (HOSO₂).
- 5) O₃ : Αποτελεί βασικό δευτερογενή ρύπο, υπεύθυνο για ένα μεγάλο πλήθος αναπνευστικών προβλημάτων. Βασικοί πρωτογενείς ρύποι που συντελούν στην αύξηση αυτού, είναι το διοξείδιο

του αζώτου και οι οργανικές πτητικές ενώσεις (VOCs). Βασική προϋπόθεση για την δημιουργία αυτού είναι η ύπαρξη ηλιακής ακτινοβολίας.

Η διαδικασία δημιουργίας κάθε ρύπου αποτελεί ένα δύσκολο πεδίο μελέτης, καθώς συνιστά ένα πολυμεταβλητό πρόβλημα. Πολλοί ρύποι αλληλοεπιδρούν μεταξύ τους είτε συνεργατικά είτε αποτρεπτικά. Χαρακτηριστικό παράδειγμα αποτελεί το όζον με τις πρόδρομες ουσίες του (precursors), όπως το διοξείδιο του αζώτου (NO_2) και τις πτητικές οργανικές ενώσεις.

Το όζον αποτελεί μια χημική ένωση η οποία συναντάται στην φύση και συντελεί έναν πολύ ωφέλιμο για τον άνθρωπο, αλλά και για την υπόλοιπη χλωρίδα και πανίδα της γης, ρόλο. Όσο το όζον βρίσκεται στην ατμόσφαιρα αποτρέπει βλαβερές ακτινοβολίες από το να φτάσουν στην επιφάνεια της Γης, προκαλώντας πληθώρα προβλημάτων. Συγκεκριμένα το όζον συμβάλει στη μείωση της υπεριώδους ακτινοβολίας (UV-B) που φτάνει στην επιφάνεια της Γης, προκαλώντας μεταξύ άλλων καρκίνο του δέρματος, βλάβες στα οπτικά νεύρα και επιβάρυνση του ανοσοποιητικού συστήματος [5]. Ωστόσο όταν το όζον βρίσκεται στα κατώτερα επίπεδα της ατμόσφαιρας, ειδικά στην τροπόσφαιρα, αποτελεί βασικό παράγοντα πολλών προβλημάτων. Ενδιαφέρον παρουσιάζει ο τρόπος δημιουργίας του καθώς είναι αποτέλεσμα πολλών επιμέρους παραγόντων. Οι Michael E. Jenkin και Kevin C. Clemitshaw[6] δημοσίευσαν το 2000 μια αναλυτική βιβλιογραφική ανασκόπηση σχετικά με την επιρροή δευτερογενών ρύπων στη συγκέντρωση του όζοντος. Αυτό το οποίο διαπιστώθηκε ήταν η πολύπλοκη αλληλεπίδραση μεταξύ πολλών διαφορετικών χημικών ουσιών, εξαρτώμενων τόσο η μια από την άλλη, όσο και από περιβαλλοντικούς και



ΣΧΗΜΑ 1 : Χημική αλληλεξάρτηση ρύπων και ουσιών που συμμετέχουν στο σχηματισμό O_3 και NO_2

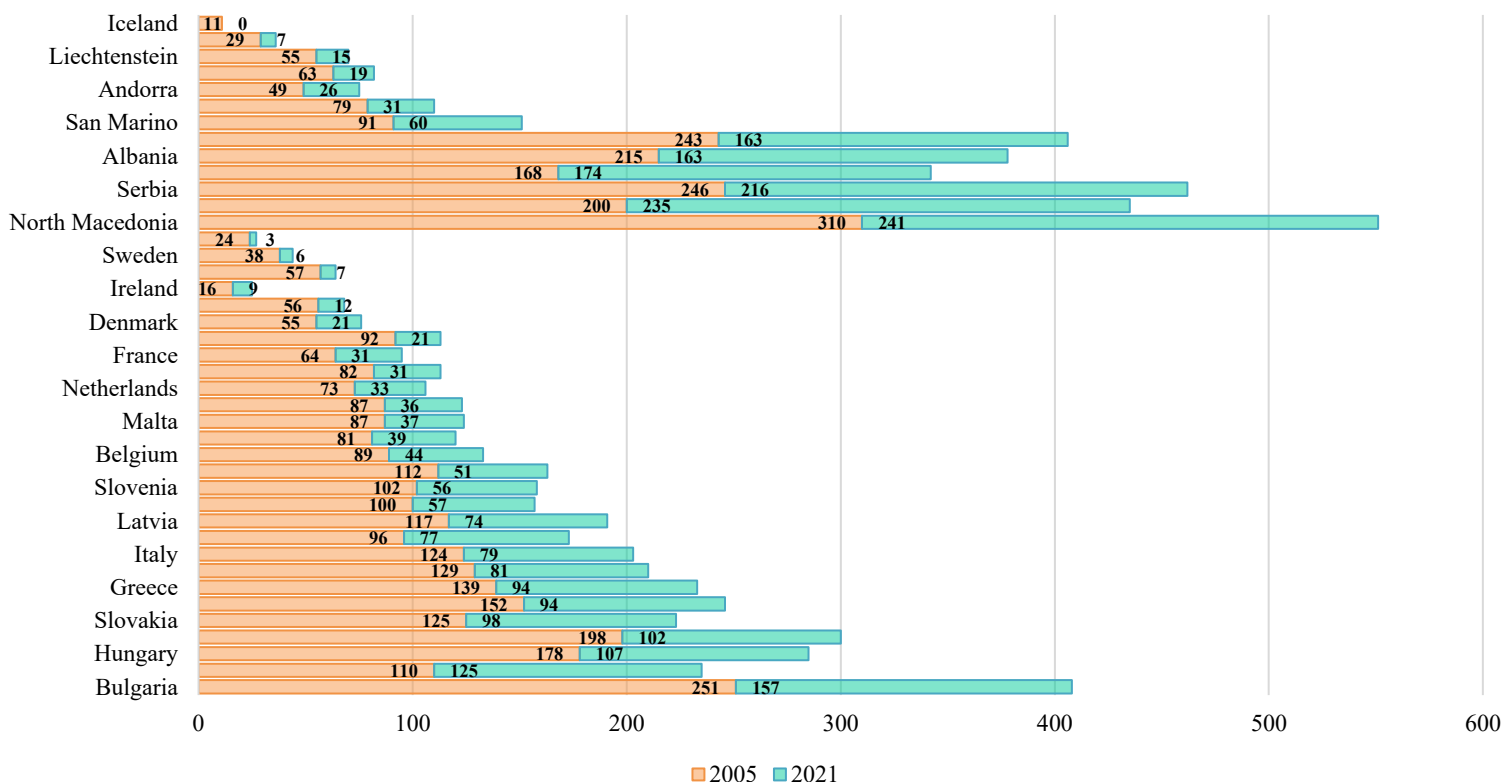
μετεωρολογικούς παράγοντες. Η παρούσα διπλωματική εργασία δεν καλείται να ερμηνεύσει τις σχέσεις αυτές ή την χημική αλληλεξάρτηση μεταξύ των ρύπων. Ωστόσο θα παρουσιαστούν οι βασικές εξισώσεις, που διέπουν τα φαινόμενα αυτά, με απώτερο στόχο την καλύτερη κατανόηση του βασικού ερευνητικού πεδίου. Στον [Πίνακα 1](#) αναγράφονται κάποιες από τις εξισώσεις σχετικά με τους ρύπους και τον τρόπο με τον οποίο επηρεάζει έκαστος τη συγκέντρωση του Όζοντος (O_3) και του Διοξειδίου του Αζώτου (NO_2). Αρκετές από τις παρακάτω χημικές ενώσεις παράγουν τα ίδια προϊόντα ασχέτως της χρήσης διαφορετικών αντιδρώντων. Αξίζει να σημειωθεί ότι κάποιες από τις παρακάτω αντιδράσεις είναι αυτοκαταλυόμενες όπως παραδείγματος χάριν η χημική αντίδραση $RH + 4O_2 \rightarrow R-HO + H_2O + 2O_3$, την οποία οι Michael E. Jenkin και Kevin C. Clementshaw [6] χαρακτηρίζουν ως αυτοκαταλυτική. Πηγές των χημικών αυτών ουσιών αποτελούν τόσο τα φυσικά (νιτροποίηση) όσο και τα ανθρωπογενή φαινόμενα (εργοστάσια, μέσα μεταφοράς κ.α.). Όπως αναφέρθηκε οι διάφοροι ρύποι μπορούν να λειτουργήσουν συνεργατικά, ωστόσο μπορούν και επιβραδυντικά. Ένα τέτοιο παράδειγμα αποτελεί η αλληλεξάρτηση μεταξύ των πτητικών οργανικών ενώσεων και των διοξειδίων του αζώτου, για σταθερές συγκεντρώσεις όζοντος, όπως φαίνεται στο [Σχήμα 1](#). Τέλος ένας σημαντικός παράγοντας αποτελεί ο χρόνος ζωής του κάθε ρύπου, μια παράμετρος η οποία επηρεάζεται από πληθώρα συνθηκών και επηρεάζει με τη σειρά της τα παραγόμενα αντιδρώντα.

	$2NO + O_2 \rightarrow 2NO_2$	(2.1.1)
	$NO + O_3 \rightarrow NO_2 + O_2$	(2.1.2)
$HO_2 + NO \rightarrow OH + NO_2$	} όπου $O_2 + hv (\lambda < 420 \text{ nm}) \rightarrow NO + O(^3P)$ όπου $O(^3P) + O_2(+M) \rightarrow O_3(+M)$	(2.1.3), (2.1.4),
$RO_2 + NO \rightarrow RO + NO_2$		(2.1.5), (2.1.6)
	$OH + NO_2(+M) \rightarrow HNO_3(+M)$	(2.1.7)
	$HO_2 + NO_2(+M) \rightleftharpoons HO_2NO_2(+M)$	(2.1.8)
	$OH + RH \rightarrow R + H_2O$	(2.1.9)
	$R + O_2(+M) \rightarrow RO_2(+M)$ όπου $RO_2 + NO \rightarrow RO + NO_2$	(2.1.10), (2.1.6)
	$RO \rightarrow \text{carbonyl product}(s) + HO_2$ όπου $HO_2 + NO \rightarrow OH + NO_2$	(2.1.11), (2.1.3)
	$RH + 4O_2 \rightarrow R-HO + H_2O + 2O_3$	(2.1.12)
	$OH + CO \rightarrow H + CO_2$	(2.1.13)
	$H + O_2(+M) \rightarrow HO_2(+M)$	(2.1.14)
	$OH + CO(+M) \rightarrow HOCO(+M)$	(2.1.15)
	$HOCO + O_2 \rightarrow HO_2 + CO_2$	(2.1.16)
	$O_3 + hv \rightarrow O(^1D) + O_2$	(2.1.17)
	$O(^1D) + H_2O \rightarrow 2OH$	(2.1.18)
	$O(^1D)(+M) \rightarrow O(^3P)(+M)$ όπου $O(^3P) + O_2(+M) \rightarrow O_3(+M)$	(2.1.19), (2.1.5)
	$RCHO + hv \rightarrow R + HCO$ όπου $HCO + O_2 \rightarrow HO_2 + CO$	(2.1.20), (2.1.21)
	$NO_2 + O_3 \rightarrow NO_3 + O_2$	(2.1.22)
	$NO_3 + hv \rightarrow NO_2 + O(^3P)$	(2.1.23)
	$NO_3 + NO \rightarrow 2NO_2$	(2.1.24)
	$NO_3 + NO_2(+M) \leftrightarrow N_2O_5(+M)$	(2.1.25)
	$HO_2 + NO_3 \rightarrow OH + NO_2 + O_2$	(2.1.26)

ΠΙΝΑΚΑΣ 1 : Χημικές εξισώσεις αλληλεπίδρασης ρύπων και ουσιών που συμμετέχουν στο σχηματισμό O_3 και NO_2

2.2. Επιπτώσεις Ατμοσφαιρικής Ρύπανσης

Οι επιπτώσεις της ατμοσφαιρικής ρύπανσης έχουν μελετηθεί σε πολλές χώρες, τόσο εντός όσο και εκτός της Ε.Ε., με κάθε έρευνα να καταλήγει στον καταλυτικό ρόλο αυτής στην υποβάθμιση της ποιότητας ζωής. Η έκθεση σε αιωρούμενα σωματίδια PM₁₀ βρέθηκε να αυξάνει τις πιθανότητες νοσηλείας για συμπτώματα βρογχίτιδας, τόσο σε ενήλικες όσο και σε παιδιά [7], ενώ η έκθεση σε υπέρβαση συγκέντρωσης σωματιδίων PM_{2.5} βρέθηκε ότι αποτελεί αιτία νοσηλείας για καρδιαγγειακά και αναπνευστικά αίτια, καθώς και αιτία θανάτων [8], [9]. Επίσης η έκθεση σε αυτού του είδους τα σωματίδια μπορεί να επηρεάσει αρνητικά τα βρέφη, ακόμα και σε προγεννητικό στάδιο, αυξάνοντας τις πιθανότητες για εκ γενετής ανωμαλίες, παχυσαρκία, διαβήτη τύπου I, νευρολογικές και συμπεριφορικές δυσλειτουργίες, πρόωρη γέννα ακόμη και νεογνικό θάνατο [10], [11], [12], [13]. Ανάλογης σημασίας είναι και οι υπερβάσεις στη συγκέντρωση των διοξειδίων του θείου (SO₂), καθώς μια υπέρβαση της τάξης των 10 ppb μπορεί να αυξήσει το πλήθος νοσηλειών κατά 1.7% [10]. Η προγεννητική και μεταγεννητική έκθεση σε διοξείδιο του αζώτου (NO₂) οδηγεί σε αυξημένες πιθανότητες για βρογχοπνευμονικές μολύνσεις, πνευμονία καθώς και εμφάνιση παχυσαρκίας [10], [14]. Έρευνα η οποία έγινε σε περιοχές της Αθήνας (Θρακομακεδόνες και Κέντρο) για τη χρονική περίοδο 2001-2018, έδειξε ότι το 6% και 7.5% των θανάτων στις περιοχές αυτές αντίστοιχα, οφείλεται στις αυξημένες συγκεντρώσεις Οζοντος (O₃) [15].



ΣΧΗΜΑ 2 : Πρόωροι θάνατοι λόγω έκθεσης σε σωματίδια PM_{2.5} σε κρατικό επίπεδο για τα έτη 2005 – 2021[14]

Ωστόσο οι επιπτώσεις της ατμοσφαιρικής ρύπανσης δεν περιορίζονται μονάχα στην ανθρώπινη υγεία. Εκτενείς έρευνες που έγιναν στην πόλη Beijing, China έδειξαν άμεση συσχέτιση μεταξύ της συγκέντρωσης του όζοντος (O₃) καθώς και των οξειδίων του Αζώτου (NO_x) και της μείωσης της ποσότητας σιτηρών. Τα ποσοστά ανέρχονται μέχρι και το 15% ενώ προβλέπεται ότι θα αγγίξουν το ύψος του 23%, μέσα στα επόμενα χρόνια [16]. Αντίστοιχα βλαβερές είναι οι επιπτώσεις και στα κτήρια λόγω της όξινης βροχής στα κτήρια αλλά και στον τομέα μεταφορών. Συγκεκριμένα ύστερα από έρευνες που έγιναν από τους Hernández et al [17] σχετικά με την επίδραση αυτής στη βαφή των αυτοκινήτων, διαπιστώθηκε ότι μεταξύ των δυο μεθόδων γήρανσης, Ξένο και μίγμα όξινης βροχής, η τελευταία είχε τις πιο καταστρεπτικές συνέπειες. Τέλος οι Ibrahim et al. [18] στην έρευνά τους σχετικά με την επιφανειακή επεξεργασία σκυροδέματος, ύστερα από έκθεση του σε όξινο περιβάλλον, αναφέρουν το φαινόμενο της γυψοποίησης και της καταστρεπτικές ιδιότητες του πάνω στα δομικά υλικά, εξασθενώντας σταδιακά της ικανότητα αντοχής του τσιμέντου σε θλίψη.

Οι παραπάνω ρύποι προκαλούν πληθώρα προβλημάτων σε ατομικό, κρατικό και διεθνές επίπεδο, καθιστώντας τη μείωση ή την ελάττωση τους ευθύνη κάθε πολίτη. Ο παγκόσμιος Οργανισμός Υγείας προκειμένου να αποτρέψει την πλήρη υποβάθμιση του ατμοσφαιρικού αέρα, έχει θεσπίσει κάποιες οριακές τιμές τις οποίες κάθε κράτος καλείται να τηρήσει. Αντίστοιχη νομοθεσία έχει θεσπιστεί και από την Ελληνική κυβέρνηση. Κάθε χώρα έχει αναπροσαρμόσει τις τιμές του Π.Ο.Υ. κατά τέτοιο τρόπο ώστε να μπορεί να τις ακολουθήσει με επιτυχία, με όσον το δυνατόν μικρότερες αποκλίσεις. Παράλληλα όλα τα έθνη προσπαθούν να φτάσουν τα δεδομένα που αυτός θέτει. Μια από τις χώρες παγκοσμίως που παρουσιάζει διαχρονικά πλήρη αδυναμία να ανταπεξέλθει σε αυτά τα όρια είναι η Κίνα. Η συνεχώς αυξανόμενη βιομηχανία, σε συνδυασμό με τον υπερπληθυσμό έχουν θέσει την ποιότητα του ατμοσφαιρικού αέρα της χώρας αυτής σε εξαιρετικά χαμηλά επίπεδα. Παρακάτω παρατίθενται τα νομοθετικά πλαίσια όπως αυτά έχουν οριστεί από τον Π.Ο.Υ., την Ελληνική Κυβέρνηση και το υπουργείο περιβάλλοντος της Κίνας.



ΣΧΗΜΑ 3: Επιπτώσεις της όξινης βροχής σε αγάλματα (γυψοποίηση), στο φυσικό περιβάλλον και σε βαφή αυτοκινήτου [19], [20], [21]

Ρύπος	Μονάδα Μέτρησης Συγκέντρωσης	Μέσο Ετήσιο Όριο	Μέσο Εποχιακό Όριο	Μέσο Ημερήσιο Όριο	Μέσο Ημερήσιο Όριο Κυλιόμενου Οχτάωρου	Μέσο Ωριαίο Όριο
SO ₂	μg/m ³	60	-	150	-	500
NO ₂	μg/m ³	40	-	80	-	200
CO	μg/m ³	-	-	4000	-	10000
O ₃	μg/m ³	-	-	-	160	200
PM ₁₀	μg/m ³	70	-	150	-	-
PM _{2.5}	μg/m ³	35	-	-	-	75
C ₆ H ₆	μg/m ³	-	-	-	-	-
Pb	μg/m ³	0.5	1	-	-	-

ΠΙΝΑΚΑΣ 2 : Οριακές τιμές-επίπεδα συγκέντρωσης ατμοσφαιρικών ρύπων με βάση την οδηγία GB 3095-2012 του State Council of the People's Republic of China.[28]

Ρύπος	Μονάδα Μέτρησης Συγκέντρωσης	Τιμή στόχος για την προστασία της δημόσιας υγείας	Όριο ενημέρωσης του κοινού	Όριο συναγερμού και λήψης εκτάκτων μέτρων
SO ₂	μg/m ³	Η μέση ωριαία τιμή, να μην υπερβαίνει τα 350 μg/m ³ περισσότερο από 24 ώρες το χρόνο.	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για το SO ₂	Όταν κατά τη διάρκεια του 24ωρου, η τιμή συγκέντρωσης του SO ₂ , ξεπεράσει τα 500 μg/m ³ για τρεις διαδοχικές ώρες
		Επίσης, η μέση ημερήσια τιμή του 24ωρου, να μην υπερβαίνει τα 125 μg/m ³ περισσότερο από 3 ημέρες το χρόνο.		
NO ₂	μg/m ³	Η μέση ωριαία τιμή, να μην υπερβαίνει τα 200 μg/m ³ περισσότερο από 18 ώρες το χρόνο.	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για το NO ₂	Όταν κατά τη διάρκεια του 24ωρου, η τιμή συγκέντρωσης του NO ₂ ξεπεράσει τα 400 μg/m ³ για τρεις διαδοχικές ώρες
		Επίσης, η μέση ετήσια τιμή συγκέντρωσης να μην υπερβαίνει τα 40 μg/m ³ .		
CO	mg/m ³	Ο μέγιστος ημερήσιος μέσος όρος κυλιόμενου δώρου, να μην υπερβαίνει τα 10 mg/m ³ περισσότερο από μια φορά το χρόνο.	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για το CO	Δεν έχει θεσπιστεί όριο συναγερμού για το CO
O ₃	μg/m ³	Ο μέγιστος ημερήσιος μέσος όρος κυλιόμενου δώρου, να μην υπερβαίνει τα 120 μg/m ³ περισσότερο από 25 ημέρες ανά ημερολογιακό έτος κατά μέσο όρο σε διάστημα τριών διαδοχικών ετών.	Όταν κατά τη διάρκεια του 24ωρου, η τιμή συγκέντρωσης του όζοντος ξεπεράσει τα 180 μg/m ³ για τουλάχιστον μια ώρα	Όταν κατά τη διάρκεια του 24ωρου, η τιμή συγκέντρωσης του όζοντος ξεπεράσει τα 240 με τι για τρεις διαδοχικές ώρες
PM ₁₀	μg/m ³	Η μέση ημερήσια τιμή του 24ωρου να μην υπερβαίνει τα 50 μg/m ³ περισσότερο από 35 ημέρες ανά ημερολογιακό έτος	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για τα PM ₁₀	Δεν έχει θεσπιστεί όριο συναγερμού για τα PM ₁₀
		Επίσης, η μέση ετήσια τιμή συγκέντρωσης να μην υπερβαίνει τα 40 μg/m ³		
PM _{2.5}	μg/m ³	Η μέση ετήσια τιμή συγκέντρωσης να μην υπερβαίνει τα 25 μg/m ³	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για τα PM _{2.5}	Δεν έχει θεσπιστεί όριο συναγερμού για τα PM _{2.5} .
C ₆ H ₆	μg/m ³	Η μέση ετήσια τιμή συγκέντρωσης να μην υπερβαίνει τα 5 μg/m ³	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για το βενζόλιο	Δεν έχει θεσπιστεί όριο συναγερμού για το βενζόλιο
Pb	μg/m ³	Η μέση ετήσια τιμή συγκέντρωσης να μην υπερβαίνει τα 0.5 μg/m ³	Δεν έχει θεσπιστεί όριο ενημέρωσης του κοινού για τον Pb	Δεν έχει θεσπιστεί όριο συναγερμού για τον Pb

ΠΙΝΑΚΑΣ 3 : Οριακές τιμές και τιμές στόχοι της συγκέντρωσης ατμοσφαιρικών ρύπων για την προστασία της δημόσιας υγείας και του περιβάλλοντος, με βάση την Ευρωπαϊκή Οδηγία 2008/50/EK και το Υπουργείο Περιβάλλοντος, Ενέργειας και Κλιματικής Αλλαγής (ΚΥΑ ΗΠ 14122/549/Ε103, ΦΕΚ 488Β/30.3.11)/22]

Ρύπος	Μονάδα Μέτρησης Συγκέντρωσης	Περίοδος Μέτρησης	Οδηγός Ποιότητας Ατμόσφαιρας (AQG)	Όριο αναφοράς (RL)	Σχόλια
SO ₂	μg/m ³	10 λεπτά	500	-	99th percentile (3-4 exceedance days per year). New 2021 guideline
		1 ημέρα	40		
NO ₂	μg/m ³	1 ώρα	200	-	99th percentile (3-4 exceedance days per year). New 2021 guideline
		1 ημέρα	25		
		Ετησίως	10		
CO	mg/m ³	1 ώρα	30	-	99th percentile (3-4 exceedance days per year). New 2021 guideline
		Μέγιστη μέση τιμή 8-ώρου ημερησίως	10		
		1 ημέρα	4		
O ₃	μg/m ³	Μέγιστη μέση τιμή 8-ώρου ημερησίως	100	-	99th percentile (3-4 exceedance days per year). New 2021 guideline
		Μέγιστη εποχική τιμή	60		
PM ₁₀	μg/m ³	1 ημέρα	45	-	99th percentile (3-4 exceedance days per year). Updated 2021 guideline
		Ετησίως	15		
PM _{2.5}	μg/m ³	1 ημέρα	15	-	99th percentile (3-4 exceedance days per year). Updated 2021 guideline
		Ετησίως	5		
C ₆ H ₆	μg/m ³	Ετησίως	-	1.7	
Pb	μg/m ³	Ετησίως	0.5		
As	ng/m ³	Ετησίως	-	6.6	
Cd	ng/m ³	Ετησίως	5		
Ni	ng/m ³	Ετησίως	-	25	
BaP	ng/m ³	Ετησίως	-	0.12	

ΠΙΝΑΚΑΣ 4: Οριακές τιμές-επίπεδα συγκέντρωσης ατμοσφαιρικών ρύπων με βάση τη νομοθεσία του Παγκόσμιου Οργανισμού Υγείας (WHO)[23]

3. ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

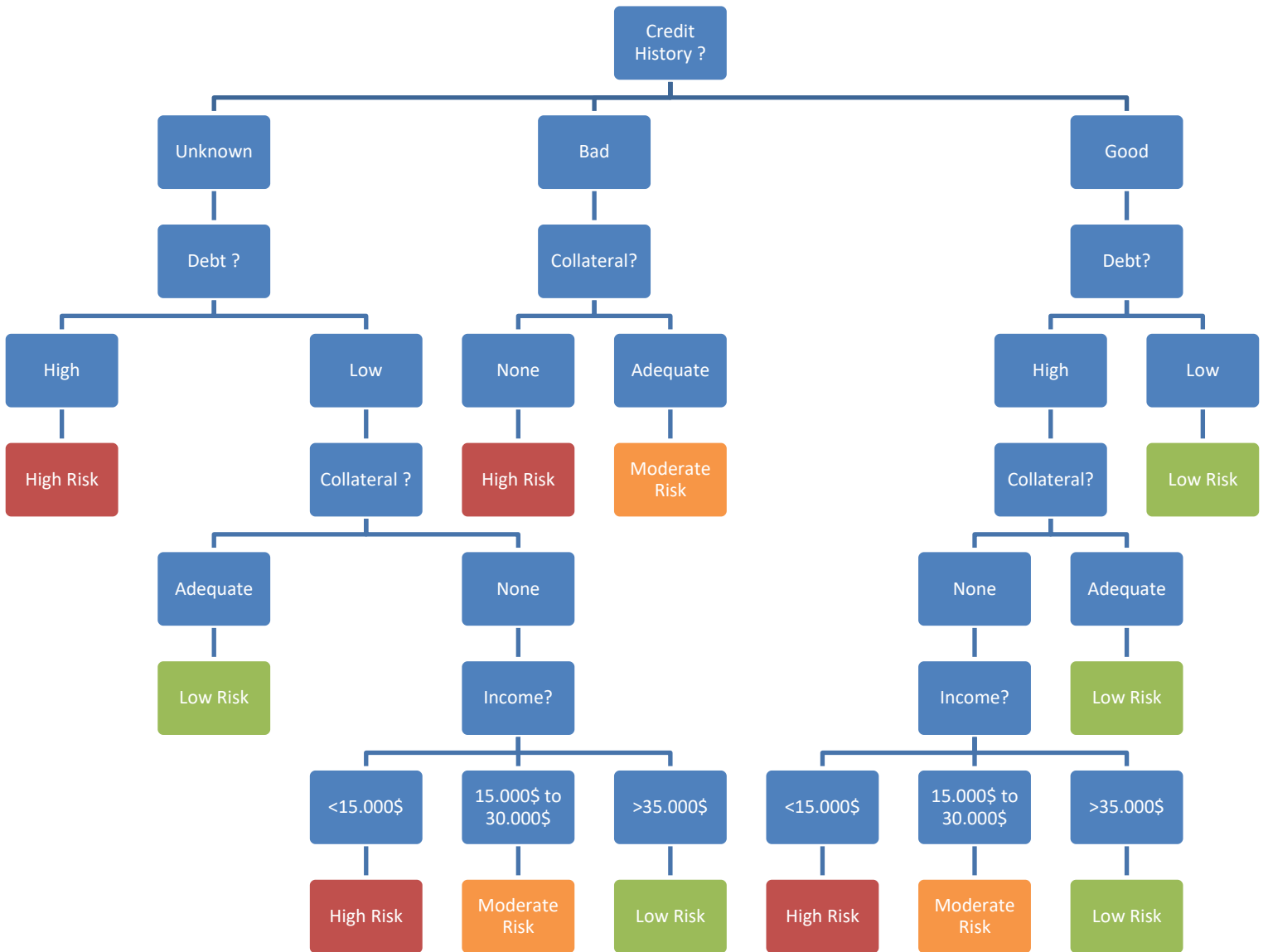
3.1. Βασικές Έννοιες

Η Μηχανική μάθηση (Machine Learning, ML) αποτελεί το θεμελιώδη λίθο για κάθε εφαρμογή η οποία περιλαμβάνει με τον ένα ή τον άλλο τρόπο τη Τεχνητή Νοημοσύνη. Σκοπός της συγκεκριμένης ενότητας είναι η παρουσίαση κάποιων βασικών εννοιών, προκειμένου η συγκεκριμένη εργασία να καταστεί πιο κατανοητή για τον αναγνώστη.

Ως Μηχανική μάθηση θεωρείται η διαδικασία κατά την οποία ένας αλγόριθμος φαίνεται να «μαθαίνει» μέσω των δεδομένων που εισάγονται σε αυτόν. Υπάρχουν πολλοί ορισμοί για το τι είναι η Μηχανική Μάθηση, ωστόσο ο πιο απλός είναι αυτός των I.Βλαχάβας et al [24]: «Μηχανική μάθηση (Machine Learning) είναι η δημιουργία μοντέλων ή προτύπων με βάση ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα». Σκοπός των συστημάτων αυτών είναι η συνεχής βελτιστοποίηση των δεδομένων εξόδου (αποτελέσματα), είτε αυτό συμβαίνει μέσω μείωσης της ταχύτητας επεξεργασίας είτε μέσω μεγαλύτερης ακρίβειας αυτών. Υπάρχουν τρία είδη μηχανικής μάθησης με καθένα από αυτά να περιέχει μια ποικιλομορφία αλγόριθμων διαθέσιμων για την επεξεργασία δεδομένων και την ανάκτηση «κρυφών» σχέσεων μεταξύ αυτών. Το πρώτο είδος είναι η «Μάθηση με επίβλεψη» (Supervised Learning, SL), όπου ο χρήστης χρειάζεται να παρέχει στον αλγόριθμο ένα μέρος από τα δεδομένα εξόδου. Ύστερα ο αλγόριθμος θα μεταβεί σε μια διαδικασία συσχέτισεων (εκπαίδευση) μεταξύ δεδομένων εισόδου (inputs) και δεδομένων εξόδου (outputs), καταλήγοντας σε μια τελική σχέση. Η τελική αυτή σχέση μπορεί, αν η εκπαίδευση γίνει σωστά, να αξιοποιηθεί και για δεδομένα εκτός του εύρους των αρχικών δεδομένων εισόδου. Επόμενο είδος μάθησης αποτελεί η «Μάθηση χωρίς επίβλεψη» (Unsupervised Learning, UL) κατά την οποία ο αλγόριθμος καλείται να δημιουργήσει μόνος του τις συσχετίσεις μεταξύ των μεγεθών, ενώ δεν υπάρχει ικανότητα γενίκευσης του μοντέλου καθώς παρουσιάζει τοπικό χαρακτήρα [24], περιγράφοντας σχέσεις μονάχα μεταξύ δεδομένων εντός του ίδιο σετ. Τέλος υπάρχει η «Ενισχυτική μάθηση» (Reinforcement Learning, RL) κατά την οποία το σύστημα αλληλοεπιδρά με το περιβάλλον και διαμορφώνει τα βάρη των δεδομένων μέσω της αλληλεπίδρασης αυτής. Σαν μέθοδος εκμάθησης η ενισχυτική μάθηση έχει γνωρίσει ραγδαία ανάπτυξη και σε συνδυασμό με την μάθηση χωρίς επίβλεψη, είναι δυνατόν να σχεδιαστούν συστήματα υψηλής ποιότητας και λειτουργικότητας. Ωστόσο η συγκεκριμένη εργασία περιορίζεται στο πρώτο είδος μάθησης, το οποίο είναι εξίσου πολύπλοκο και χρηστικό.

Η μέθοδος της παρεμβολής αποτελεί το βασικότερο είδος SL, είτε αυτή είναι γραμμική (Linear Regression, LR), είτε πολυωνυμική (Newton, Lagrange). Παρότι μπορεί να εξάγει αποτελέσματα με σχετική ακρίβεια ωστόσο είναι αξιοποιήσιμη όσο τα δεδομένα παρουσιάζουν εμφανείς συσχετίσεις μεταξύ τους. Μια άλλη μέθοδος είναι τα δένδρα αποφάσεων (Decision Trees, DT) τα οποία αποτελούν μια απλούστερη διαδικασία εξαγωγής διακριτών αποτελεσμάτων. Τα DT έχουν γνωρίσει εφαρμογή σε πληθώρα πεδίων, όπως ιατρική, χρηματοοικονομικά κλπ, λόγω της απλής όπως αναφέρθηκε δομής τους (if-then statements). Κάποιοι ενδεικτικοί αλγόριθμοι μάθησης είναι ο ID3, ο C4.5, ο CART, ο CHAID, ο Condition Inference Tree και ο Random Forest. Ερευνητές έχουν προσπαθήσει με επιτυχία να χρησιμοποιήσουν τέτοιου είδους μοντέλα

προκειμένου να επιτύχουν παρόμοιους στόχους με αυτούς της συγκεκριμένης εργασίας, ωστόσο η μεθοδολογία εδώ θα εστιάσει κυρίως στα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), τα οποία περιγράφονται εκτενώς στην επόμενη ενότητα.

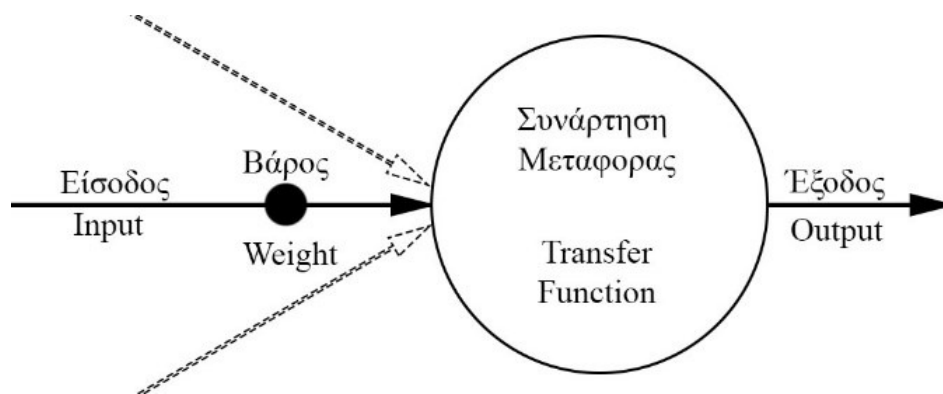


ΣΧΗΜΑ 4 : Γραφική αναπαράσταση του αλγόριθμου ID3, σχετικά με τον έλεγχο ρίσκου (Risk Assessment) σχετικά με την παραχώρηση δανείου σε πελάτη τράπεζας (Credit Analysis) [22]

3.2. Τεχνητά Νευρωνικά Δίκτυα

3.2.1. Ορισμός & Βασικές Έννοιες

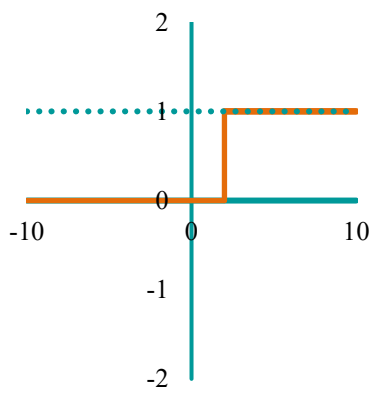
Τα ΤΝΔ εντάσσονται στον τομέα της μηχανικής μάθησης και αυτό το οποίο τα ξεχωρίζει από τα υπόλοιπα συστήματα μάθησης, είναι ο ιδιαίτερος τρόπος με τον οποίο δομούνται. Τα δίκτυα αυτά, όπως προδίδει και το όνομα τους, προσπαθούν να αντιγράψουν τόσο τη δομή όσο και τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου [24]. Συγκεκριμένα αποτελούνται από τεχνητούς νευρώνες, οι οποίοι μεταφέρουν τα δεδομένα εισόδου (inputs), τα οποία πολλαπλασιάζονται με κάποια «βάρη» (συντελεστές οι οποίοι αυξομειώνουν την επίδραση που έχει η κάθε είσοδος στο τελικό αποτέλεσμα) και αφού διέλθουν μέσω μίας συνάρτησης ενεργοποίησης (activation function), εξέρχονται από αυτόν (output) [24]. Όπως και οι κανονικοί νευρώνες έτσι και εδώ, κάθε νευρώνας μπορεί να έχει πολλαπλές τιμές εισόδου, αλλά μονάχα μια τιμή εξόδου. Τα δεδομένα ύστερα από την είσοδο τους στον νευρώνα πολλαπλασιάζονται με τα αντίστοιχα βάρη τους και προστίθενται μεταξύ τους. Ο αθροιστής αυτός εισάγεται στη συνάρτηση ενεργοποίησης και εξάγεται το τελικό σήμα εξόδου, το οποίο συγκρίνεται με το επιθυμητό σήμα. Σε περίπτωση που η τιμή εξόδου είναι ίδια με την επιθυμητή τιμή τότε δεν πραγματοποιείται καμία μεταβολή στο βάρος με το οποίο πολλαπλασιάζεται κάθε τιμή εισόδου, σε αντίθετη περίπτωση ο αλγόριθμος αυξομειώνει τα βάρη και επαναλαμβάνει την ίδια διαδικασία. Φυσικά οι παραπάνω διεργασίες δεν επαναλαμβάνονται άπειρες φορές, αλλά μέχρι να ικανοποιηθεί κάποια συνθήκη, η οποία αν δεν έχει τεθεί από τον χειριστή, είναι η μείωση του σφάλματος κάτω από μια ορισμένη τιμή. Η σχηματική αναπαράσταση της παραπάνω διαδικασίας παρουσιάζεται στο Σχήμα 5. Ο τρόπος με τον οποίο διεξάγεται η επεξεργασία των δεδομένων εισόδου και η συσχέτιση τους με τα δεδομένα εξόδου, καθιστούν τη συγκεκριμένη μέθοδο μάθησης ιδανική για περιπτώσεις δημιουργίας μοντέλων ταξινόμησης, παρεμβολής, ομαδοποίησης, βελτιστοποίησης και πρόβλεψης ενώ παράλληλα μπορούν εφαρμοστούν και σε εφαρμογές όπως η μηχανική όραση και η επεξεργασία φυσικής γλώσσας.



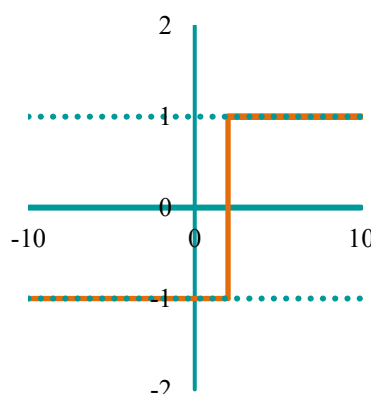
ΣΧΗΜΑ 5 : Σχηματική αναπαράσταση νευρώνα ενός ΤΝΔ.

Ένα ΤΝΔ είναι τόσο καλό, όσο τα δεδομένα, η τοπολογία και η διαδικασία εκπαίδευσης του. Η επιλογή μιας σωστής συνάρτησης ενεργοποίησης μπορεί να παρουσιάσει ουσιώδη διαφορά μεταξύ των δεδομένων, καθιστώντας τη διαδικασία της εκπαίδευσης επιτυχημένη ή μη. Κάποιες από τις βασικότερες συναρτήσεις ενεργοποίησης είναι :

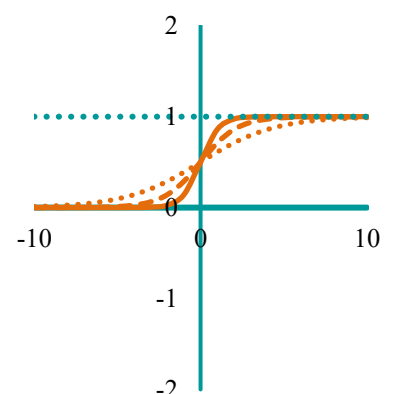
- Η βηματική απόκριση η οποία αναγκάζει την τιμή εξόδου να ισούται με 0 ή 1, ανάλογα με το αν η τιμή του αθροιστή βρίσκεται κάτω ή άνω του κατωφλίου που ορίστηκε.
- Η συνάρτηση πρόσημου η οποία αναγκάζει την τιμή εξόδου να ισούται -1 ή 1, ανάλογα με την τιμή του αθροιστή.
- Η λογιστική συνάρτηση η οποία φράζει την τιμή εξόδου μεταξύ 0 και 1, ανάλογα με την τιμή του αθροιστή.
- Η συνάρτηση υπερβολικής εφαπτομένης η οποία φράζει την τιμή εξόδου μεταξύ -1 και 1, ανάλογα με την τιμή του αθροιστή.
- Η συνάρτηση ανόρθωσης η οποία θέτει την τιμή εξόδου ίση με 0 για τιμή του αθροιστή είναι κάτω από μηδέν και ίση με την ίδια όταν είναι μεγαλύτερη του μηδενός.
- Η συνάρτηση ομαλής ανόρθωσης η οποία κάνει ακριβώς το ίδιο με τη συνάρτηση ανόρθωσης, αλλά με μια πιο ομαλή μετάβαση.



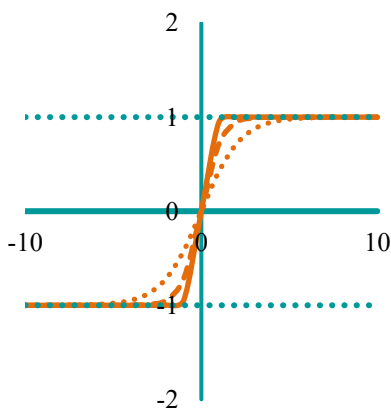
(a)



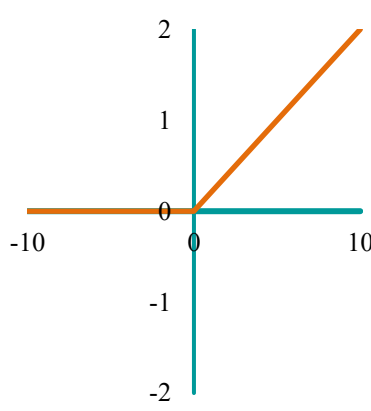
(b)



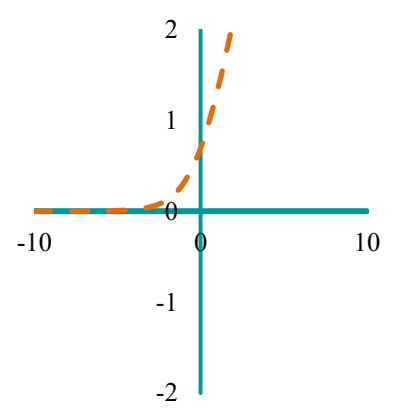
(c)



(d)



(e)



(f)

Γραφικές απεικονίσεις των συναρτήσεων ενεργοποίησης (a) βηματική, (b) πρόσημου, (c) λογιστική, (d) υπερβολική εφαπτομένη, (e) ανόρθωσης, (f) ομαλή ανόρθωση

3.2.2. Τεχνητά Νευρωνικά Δίκτυα Πολλαπλών Στρωμάτων (MLP)

Όπως αναφέρθηκε ένας νευρώνας μπορεί να έχει πολλές τιμές εισόδου αλλά μονάχα μια τιμή εξόδου, η οποία μπορεί να διακλαδίζεται σε παραπάνω από ένα παράλληλους νευρώνες. Η βασική τοπολογία ενός ΤΝΔ απαρτίζεται από τρεις περιοχές (α) την είσοδο, (β) τα κρυφά στρώματα και (γ) την έξοδο του δικτύου. Η τοπολογία αποτελεί καθοριστικό παράγοντα για την εξασφάλιση ενός σωστού ΤΝΔ, το οποίο καταφέρνει να συσχετίσει τα δεδομένα με ακρίβεια. Οι ύπαρξη κρυφών στρώσεων μπορεί να βοηθήσει στην βελτίωση της απόδοσης του δικτύου. Τα ΤΝΔ για αυτό τον λόγο χωρίζονται σε 2 κατηγορίες. Στην πρώτη κατηγορία βρίσκονται τα perceptrons, τα οποία είναι η απλούστερη μορφή ΤΝΔ χωρίς κανένα κρυφό στρώμα. Σαν τεχνική θεωρείται απαρχαιωμένη και πλέον δεν έχει κάποια πρακτική εφαρμογή. Η δεύτερη κατηγορία ΤΝΔ είναι τα δίκτυα πολλαπλών στρωμάτων/επιπέδων (Multi-Layer Perceptron, MLP), τα οποία εμπεριέχουν πλήθος επιμέρους δικτύων, όπως τα ΤΝΔ εμπρόσθιας τροφοδότησης (FFNN), οπισθοδιάδοσης σφαλμάτων (BPNN), βαθιά νευρωνικά δίκτυα (DNN), συνελκτικά νευρωνικά δίκτυα (CNN) και πολλά ακόμη. Η εκτενής ανάλυση όλων των παραπάνω δικτύων δεν θα γίνει στην παρούσα εργασία, ωστόσο θα γίνει ανάλυση των βασικών συνιστωσών εκπαίδευσης των νευρωνικών δικτύων που εκπαιδεύτηκαν.

- Νευρώνες Εισόδου : Το πλήθος των νευρώνων εισόδου είναι άμεσα εξαρτώμενο τόσο από το πλήθος των δεδομένων, όσο και από το σκοπό δημιουργίας του δικτύου. Ιδανικά το πλήθος των νευρώνων πρέπει να είναι ίσο με το πλήθος των κλάσεων στην έξοδο, ειδικά αν το δίκτυο αποσκοπεί σε λειτουργίες ταξινόμησης, χωρίς αυτό να μειώνει απαραίτητα το πλήθος των σφαλμάτων. Κάθε νευρώνας μπορεί να χρησιμοποιήσει διαφορετική συνάρτηση ενεργοποίησης, ωστόσο κάτι τέτοιο συχνά αποφεύγεται, όπως επίσης αποφεύγεται και η χρήση γραμμικών συναρτήσεων ενεργοποίησης όπως η βηματική ή πρόσημου, καθώς δεν συμβάλουν θετικά στην εκπαίδευση.

- Κρυφά Επίπεδα/Στρώματα : Το πλήθος των κρυφών στρωμάτων διαδραματίζει καίριο ρόλο στην ποιότητα των ΤΝΔ. Νευρωνικά δίκτυα με παραπάνω από 2 κρυφά στρώματα θεωρούνται «βαθιά» και παρουσιάζουν καλύτερη απόδοση, σε σχέση με δίκτυα λιγότερων στρωμάτων. Η ανάγκη ύπαρξης παραπάνω των δύο κρυφών στρωμάτων προέκυψε από την ανάγκη προσέγγισης περίπλοκων συναρτήσεων, χωρίς όμως την ταυτόχρονη αύξηση των νευρώνων εντός ενός μονάχα επιπέδου. Γενικά όσο πιο περίπλοκες οι συναρτήσεις μελέτης, τόσο μεγαλύτερο και το πλήθος των νευρώνων, ωστόσο με την πληθυσμιακή αυτή αύξηση ανάλογη συμπεριφορά παρουσιάζει και η πιθανότητα υπερπροσαρμογής.

- Υπερπροσαρμογή : Ως υπερπροσαρμογή (over-fitting) ορίζεται το φαινόμενο κατά το οποίο το ΤΝΔ απομνημονεύει τις εξόδους για κάθε συγκεκριμένη είσοδο. Φαινομενικά ένα τέτοιο δίκτυο αποδίδει εξαιρετικά για τις συγκεκριμένες τιμές εισόδου, αλλά αποτυγχάνει πλήρως στην πρόβλεψη άλλων τιμών εισόδου. Για αυτόν το λόγο, τα δεδομένα εκπαίδευσης χωρίζονται σε τρία διαφορετικά σετ (1) Σετ εκπαίδευσης (training), (2) Σετ επιβεβαίωσης (validation) και (3) Σετ ελέγχου (testing). Το συγκεκριμένο σφάλμα παρατηρείται έντονα όταν τίθεται ως αποκλειστική συνθήκη τερματισμού τις διαδικασίας εκπαίδευσης, η ελαχιστοποίηση του σφάλματος.

- Αλγόριθμος Εκπαίδευσης (A.E.) : Προκειμένου να εξαχθούν τα μοντέλα, δεν αρκεί μονάχα ο προσδιορισμός των συναρτήσεων ενεργοποίησης, του πλήθους των νευρώνων και των

επιπέδων. Αναγκαίος είναι και ο προσδιορισμός ενός Α.Ε., ανάλογα με το είδος του προβλήματος και τη δομή του δικτύου. Υπάρχουν πολλοί τέτοιοι αλγόριθμοι και κάποιος αρκετά εξοικειωμένος με τον τομέα αυτό δύναται να δημιουργήσει εξατομικευμένους κώδικες. Ωστόσο η εκπαίδευση των συγκεκριμένων νευρωνικών έγινε με τη βοήθεια του MATLAB, συνεπώς επιλέχθηκαν ορισμένοι έτοιμοι κώδικες προκειμένου να υλοποιηθεί η εκπαίδευση των ΤΝΔ. Αξίζει να αναφερθεί ότι κάθε συνάρτηση αποτελεί μια ξεχωριστή και ιδιαίτερη μέθοδο εκπαίδευσης των ΤΝΔ καθώς έχει ορισμένα δυνατά, αλλά και αδύναμα, σημεία ως προς αυτή. Οι Α.Ε. χωρίζονται συνήθως σε δύο κατηγορίες.

Η πρώτη κατηγορία αποσκοπεί στην προσέγγιση μιας συνάρτησης και στην εξαγωγή μια συνάρτησης παρεμβολής. Συγκεκριμένα :

«Προσέγγιση μιας συνάρτησης $f(x)$ από μια άλλη συνάρτηση $g(x)$ που να είναι γενικά πιο εύχρηστη και που να είναι κοντά στην $f(x)$ στο κλειστό διάστημα $[a,b]$ θα ονομάζεται οποιαδήποτε συνάρτηση $g(x)$ για την οποία ισχύει η σχέση: $|f(x)-g(x)| \leq \epsilon$ [25], όπου η $g(x)$ ονομάζεται συνάρτηση παρεμβολής και το ϵ είναι το μέγιστο επιτρεπόμενο σφάλμα μεταβολής σε κάθε σημείο x .

Η δεύτερη είναι η αναγνώριση μοτίβων, κατά την οποία οι κατάλληλα δομημένοι αλγόριθμοι ψάχνουν για «κανονικότητες», διακριτά δηλαδή στοιχεία, μέσα στο σύνολο των δεδομένων, με συνέπεια την κατηγοριοποίηση αυτών [26]. Φυσικά πέραν των παραπάνω βασικών εννοιών υπάρχουν πολλές ακόμα παράμετροι που επηρεάζουν τη διαδικασία εκπαίδευσης ενός ΤΝΔ, ωστόσο δεν κρίνεται αναγκαία η ανάλυση τους για την καλύτερη κατανόηση της παρούσας εργασίας. Οι διαθέσιμοι αλγόριθμοι κατηγοριοποιούνται όπως φαίνεται στον [Πίνακα 5](#) [27].

Κλασσικοί Αλγόριθμοι Εκπαίδευσης ΤΝΔ

Όνομα	Συντομογραφία	Εντολή	Τρόπος Αξιοποίησης
Levenberg-Marquardt	LM	trainlm	Προσέγγιση Συνάρτησης
Bayesian Regularization	BR	trainbr	Αναγνώριση Μοτίβων
BFGS Quasi-Newton	BFG	trainbfg	Προσέγγιση Συνάρτησης
Conjugate gradient backpropagation with Powell-Beale restarts	CGB	traincgb	Αναγνώριση Μοτίβων
Fletcher-Powell Conjugate Gradient	CGF	traincgf	Αναγνώριση Μοτίβων
Polak-Ribière Conjugate Gradient	CGP	traincgp	Αναγνώριση Μοτίβων
Gradient descent backpropagation	GD	traingd	Αναγνώριση Μοτίβων
Gradient descent with adaptive learning rate backpropagation	GDA	traingda	Αναγνώριση Μοτίβων
Gradient descent with momentum backpropagation	GDM	traingdm	Αναγνώριση Μοτίβων
Gradient descent with momentum and adaptive learning rate backpropagation	GDX	traingdx	Αναγνώριση Μοτίβων
One-step secant backpropagation	OSS	trainoss	Προσέγγιση Συνάρτησης
Resilient backpropagation	RP	trainrp	Αναγνώριση Μοτίβων
Scaled conjugate gradient backpropagation	SCG	trainscg	Ευρείας χρήσης, χρειάζεται μεγάλα δίκτυα

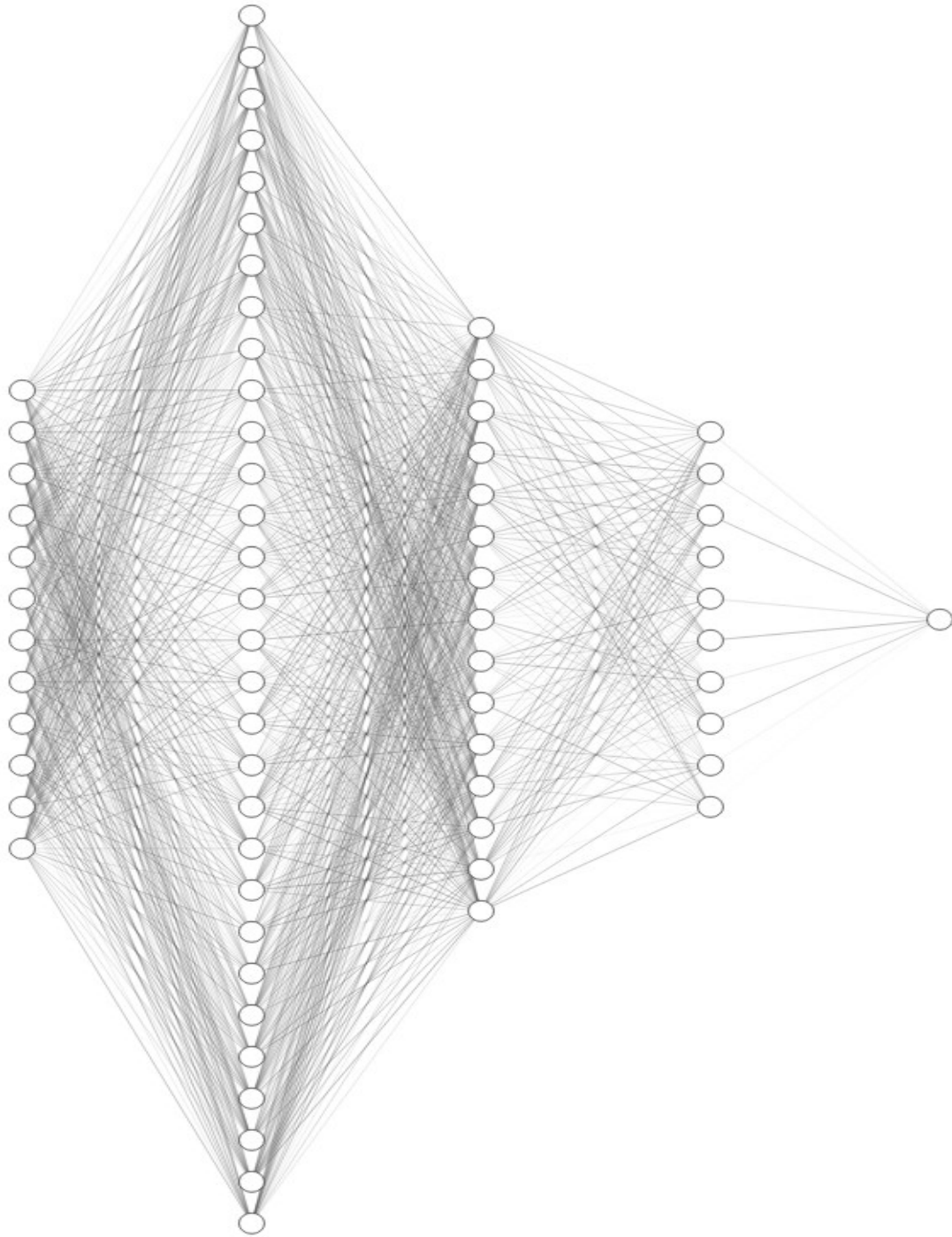
Αλγόριθμοι Εκπαίδευσης ΤΝΔ υπό Επίβλεψη για Βαθιά Δίκτυα (Deep Learning)

Όνομα		Εντολή	Τρόπος Αξιοποίησης
Batch training with weight and bias learning rules	B	trainb	-
Cyclical order weight/bias training	C	trainc	-
Random order incremental training with learning functions	R	trainr	-
Sequential order incremental training with learning functions	S	trains	-

Αλγόριθμοι Εκπαίδευσης ΤΝΔ χωρίς Επίβλεψη για Βαθιά Δίκτυα (Deep Learning)

Όνομα		Εντολή	Τρόπος Αξιοποίησης
Batch unsupervised weight/bias training	BU	trainbu	-
Unsupervised random order weight/bias training	RU	trainru	-

ΠΙΝΑΚΑΣ 5 : Συναρτήσεις εκπαίδευσης ΤΝΔ και ο συνήθης τρόπος αξιοποίησης τους.



Input Layer
(12 Neurons)

Hidden Layer 1
(30 Neurons)

Hidden Layer 2
(15 Neurons)

Hidden Layer 3
(10 Neurons)

Output Layer

ΣΧΗΜΑ 6 : Σχηματική αναπαράσταση ενός ΤΝΔ πολλαπλών στρωμάτων (MLP) με δομή (12-30-15-10-1).

3.2.3. Levenberg – Marquardt και Bayesian Regularization

Οι βασικές συναρτήσεις εκπαίδευσης που χρησιμοποιήθηκαν στη συγκεκριμένη εργασία είναι η συνάρτηση Levenberg – Marquardt (LM) και Bayesian Regularization (BR). Φυσικά, όπως αναφέρθηκε χρησιμοποιήθηκαν επτά συνολικά συναρτήσεις εκπαίδευσης, ωστόσο οι υπόλοιπες έδρασαν επικουρικά, χωρίς να παρέχουν μοντέλα μεγάλης αξιοπιστίας. Συνεπώς δεν κρίνεται αναγκαία η περιγραφή των αρχών λειτουργίας των υπόλοιπων πέντε. Παρακάτω θα γίνει μια συνοπτική ανάλυση των βασικών αρχών λειτουργίας των συναρτήσεων LM και BR.

«Ο αλγόριθμος LM αποτελεί μια επαναληπτική μέθοδο εντοπισμού τοπικών ελαχίστων μιας πολύ-μεταβλητής συνάρτησης, εκφραζόμενης ως άθροισμα των τετραγώνων πολλών μη γραμμικών, πραγματικών συναρτήσεων.»[28]. Ο συγκεκριμένος αλγόριθμος αποτελεί συνήθη πρακτική για την επίλυση πολλών προβλημάτων προσαρμογής καμπύλης (Fitting). Αποτελεί ένα συνδυασμό δύο μεθόδων, της μεθόδου Gauss-Newton και της μεθόδου της απότομης κατάβασης (Steepest Descent). Μέσω της αξιοποίησης των δύο παραπάνω τεχνικών ο αλγόριθμος συμπεριφέρεται όπως η μέθοδος Gauss-Newton όταν βρίσκεται κοντά στη λύση του συστήματος, με αποτέλεσμα να συγκλίνει ραγδαία. Ωστόσο όταν βρίσκεται μακριά από το τοπικό ελάχιστο, συμπεριφέρεται όπως η μέθοδος της απότομης κατάβασης, η οποία αποτελεί μια επέκταση της μεθόδου Laplace, με αποτέλεσμα να συγκλίνει αλλά με αργό ρυθμό [28]. Η μέθοδος αυτή επιτρέπει στο χρήστη τη μετάβαση στο μιγαδικό επίπεδο, με στόχο τον προσδιορισμό ενός στατικού σημείου, επιτρέποντας τον υπολογισμό της ασυμπτωτικής επέκτασης του ολοκληρώματος μορφής [29] $F(\lambda) = \int_{\gamma} f(z) \cdot e^{\lambda \cdot S(z)} dz$ (3.2.1), όπου γ : η κλίση της ευθείας. Σε αντίθεση με τη προηγούμενη μέθοδο η οποία βασίζεται στη δημιουργία παραγώγων, η μέθοδος Gauss-Newton, αποτελώντας και αυτή μια επέκταση της μεθόδου Newton, δεν απαιτεί την ύπαρξη δεύτερης παραγώγου και αποσκοπεί στην εύρεση ενός ελαχίστου αθροίσματος τετραγώνων, των μη γραμμικών εξισώσεων όπου εφαρμόζεται [30]. Είναι μια επαναληπτική μέθοδος, η οποία παρουσιάζει κάποιες ιδιαιτερότητες ως προς το είδος των μητρών που μπορεί να διαχειριστεί και δεδομένου ότι αποτελεί άθροισμα μη αρνητικών στοιχείων, η προσέγγιση της βέλτιστης λύσης γίνεται με το σταδιακό μηδενισμό του αθροίσματος αυτού. Η μέθοδος Gauss-Newton αποτελεί μια από τις βασικότερες μεθόδους επίλυσης προβλημάτων μη γραμμικής παλινδρόμησης. Η μέθοδος LM παρουσιάζει εξαιρετικές ικανότητες γενίκευσης, με γρήγορες ταχύτητες και μικρές πιθανότητες υπερπροσαρμογής, ως προς άλλες συμβατικές μεθόδους.

Ο αλγόριθμος BR βασίζεται πάνω σε δύο βασικές μαθηματικές – στατιστικές μεθόδους, αυτή της κανονικοποίησης και αυτή των στατιστικών πιθανοτήτων βάσει του κανόνα Bayes, αποτελώντας μια μη γραμμική παραλλαγή της μεθόδου αναζήτησης με αναρρίχηση λόφων (Ridge Regression)[31]. Κατά την εκπαίδευση ενός αλγορίθμου που χρησιμοποιεί το άθροισμα των τετραγώνων ως δείκτη επιτυχίας, στόχος είναι η ελαχιστοποίηση της τιμής αυτής. Επομένως προκύπτει η σχέση (αντικειμενική συνάρτηση)[32]:

$$S = S_{\text{errors}} = \sum_{i=1}^n [(obs_i - pred_i)^2] \quad (3.2.2)$$

Ωστόσο κατά την κανονικοποίηση η παραπάνω σχέση (3.2.2) μεταβάλλεται ως εξής :

$$S = a \cdot S_{\text{errors}} + b \cdot S_{\text{weights}} \quad (3.2.3), \text{ όπου } a, b : \text{πραγματικοί αριθμοί.}$$

Στην σχέση (3.2.3) η αντικειμενική συνάρτηση αποτελεί άθροισμα των σφαλμάτων και των βαρών του συστήματος. Αν το $a > b$ τότε η εκπαίδευση θα δώσει έμφαση στην μείωση των βαρών ενώ σε αντίθετη περίπτωση θα δώσει έμφαση στη μείωση των σφαλμάτων. Η βασική καινοτομία του συγκεκριμένου αλγόριθμου, πέραν από την εξάλειψη της ανάγκης επαλήθευσης των αποτελεσμάτων της εκπαίδευσης (validation), είναι ότι τα βάρη αποτελούν τυχαίες μεταβλητές, στην αρχή, ενώ στη συνέχεια δομούνται βάση του κανόνα πιθανοτήτων κατά Bayes, όπου :

$$P(\vec{w} | D, a, b, M) = \frac{P(D | \vec{w}, b, M) \cdot P(\vec{w} | a, M)}{P(D | a, b, M)} \quad (3.2.4) \text{ με } \vec{w}: \text{το διάνυσμα βαρών}$$

D: το σετ δεδομένων
M: το μοντέλο ΤΝΔ

Η ανάλυση των εννοιών του νόμου Bayes αποτελεί ένα χρονοβόρο και περίπλοκο εγχείρημα, οπότε σε μία προσπάθεια απλούστευσης, προκειμένου να γίνει κατανοητός στον αναγνώστη, χωρίς την ταυτόχρονη αναγραφή τύπων μπορεί να περιγραφεί ως εξής. « Ο νόμος Bayes επιτρέπει τον υπολογισμό πιθανοτήτων υπό συνθήκη με χρήση άλλων πιθανοτήτων που είναι ευκολότερο να υπολογιστούν.» [24]. Γενικά ο αλγόριθμος BR προσφέρει πολλά πλεονεκτήματα όπως τη χρήση μεγάλων σετ δεδομένων, ευελιξία σχεδίασης λόγω της εκτενούς βιβλιογραφίας τους, παρέχει βελτιωμένη ανάλυση αβεβαιότητας, προσφέρει αυξημένες επιδόσεις και αποφεύγεται η υπερπροσαρμογή, ενώ τέλος μπορεί να λάβει υπόψη εμπειρικά δεδομένα [33], [34].

3.2.4. Μοντέλα Πρόγνωσης

Όπως γίνεται αντιληπτό, είναι επιτακτική η ανάγκη σχεδιασμού ενός μέσου πρόγνωσης των συγκεντρώσεων των ρύπων. Διαχρονικά πολλοί ερευνητές προσπάθησαν να σχεδιάσουν τέτοιου είδους προγνωστικά μοντέλα. Το 2015 οι Xiao Feng et al. [35] δημιούργησαν ένα μοντέλο το οποίο συνδύαζε ένα νευρωνικό δίκτυο MLP με ένα γεωγραφικό μοντέλο με βάση την τροχιά του αέρα και την αξιοποίηση της μεθόδου μετασχηματισμού κυματιδίων (wavelet transformation) για την πρόγνωση της μέσης ημερήσιας συγκέντρωσης $PM_{2.5}$ για τις επόμενες δύο ημέρες στην πόλη Beijing, China. Αντίστοιχες προσπάθειες έγιναν και από τους Madhavi Anushka Elangasinghe et al. [36], το 2014, για την πρόβλεψη της ωριαίας συγκέντρωσης του διοξειδίου του Αζώτου (NO_2), με την αξιοποίηση νευρωνικού δικτύου MLP με ένα κρυφό στρώμα, εκπαιδευμένο με αλγόριθμο Levenburg Marquardt, για την περιοχή Auckland, New Zealand. Το 2016 οι Yun Bai et al. [37] αξιοποιώντας τη μέθοδο μετασχηματισμού στατικών κυματιδίων και την εκπαίδευση ενός νευρωνικού δικτύου με αλγόριθμο οπίσθιας διάδοσης, προσπάθησαν να προβλέψουν τις ημερήσιες συγκεντρώσεις σωματιδίων PM_{10} και τις συγκεντρώσεις διοξειδίου του θείου (SO_2) και διοξειδίου του αζώτου (NO_2). Το 2017 οι Fabio Biancofiore et al. [38] προσπάθησαν και αυτοί να προβλέψουν τις μέσες ημερήσιες συγκεντρώσεις σωματιδίων $PM_{2.5}$ και PM_{10} για τις επόμενες μια έως και τρεις μέρες, ενώ το 2018 οι Fabiana Franceschi et al. [39] προσπάθησαν μέσω στατιστικής μελέτης και ενσωμάτωσης ΤΝΔ να προβλέψουν τις συγκεντρώσεις των σωματιδίων $PM_{2.5}$ και PM_{10} στην Bogotá, Colombia. Αντίστοιχα με έρευνες που δημοσιεύθηκαν εντός 2024 οι Quanchao

Chen et al. [40] σχεδίασαν ένα καινοτόμο μοντέλο πρόβλεψης ωριαίων συγκεντρώσεων για σωματίδια $PM_{2.5}$, PM_{10} καθώς και για την ωριαία συγκέντρωση του όζοντος (O_3), στην πόλη του Πεκίνο, με την δημιουργία ενός αλγόριθμου AAMGCRN (Adaptive Adjacency Matrix-Based Graph Convolutional Recurrent Network). Αντίστοιχα οι Sarmad Dashti Latif et al. [41] εφάρμοσαν πολλαπλές μεθόδους μηχανικής μάθησης προκειμένου να προβλέψουν την συγκέντρωση του όζοντος (O_3) για τις επόμενες 1,3,5 και 7 ώρες, στην περιοχή Klang Valley, Malaysia.

Η παρούσα ερευνητική εργασία διεξήχθη σε μια προσπάθεια να σχεδιαστεί ένα καθολικό προγνωστικό μοντέλο, καθώς κανένας από τους παραπάνω ερευνητές δεν έχει σχεδιάσει κάτι εξίσου συμπεριληπτικό. Αναφέρεται ακόμα ότι σε βάθος δεκαετίας δεν βρέθηκαν αντίστοιχες μελέτες οι οποίες χρησιμοποιούν ως δεδομένα εκπαίδευσης, αυστηρά μετεωρολογικές μεταβλητές και όχι άλλους ρύπους. Η τελευταία αναφορά είναι το 2014, στο προγνωστικό μοντέλο που δημιούργησαν οι Madhavi Anushka Elangasinghe et al. [36], οι οποίοι υποστήριξαν ότι μοντέλα στα οποία τα δεδομένα εισόδου (inputs) απαρτίζονται μεταξύ άλλων και από συγκεντρώσεις ρύπων, έχουν μικρή δυνατότητα αξιοποίησης.

4. ΜΕΘΟΔΟΛΟΓΙΑ

Όπως αναφέρθηκε και στην [Ενότητα 3.2](#) την τελευταία δεκαετία έχουν υπάρξει πολλές προσπάθειες για την δημιουργία ενός μοντέλου πρόγνωσης συγκεντρώσεων ρύπων. Ωστόσο η συγκεκριμένη έρευνα πραγματοποιείται την δημιουργία ενός ενιαίου μοντέλου πρόβλεψης της συγκέντρωσης των PM_{2.5}, PM₁₀, CO, SO₂, NO₂ και O₃, για τις επόμενες 24 ώρες. Στόχος είναι η πρόβλεψη και κατά συνέπεια η ενημέρωση των πολιτών καθώς και η υποβοήθηση των κρατικών φορέων στην ρύθμιση διαφόρων παραγόντων που οδηγούν στην έξαρση των ρύπων αυτών, καθώς επίσης και η έγκαιρη ενημέρωση των φορέων υγείας ώστε να μπορέσουν να προετοιμαστούν για την αύξηση των αυξημένων νοσηλειών.

4.1. Περιοχή Μελέτης

Η περιοχή μελέτης είναι η πόλη του Πεκίνο, Κίνα (39° 54' 13" N, 116° 23' 17" E) για την οποία συλλέχθηκαν ωριαίες μετρήσεις των συγκεντρώσεων των ρύπων ενδιαφέροντος, καθώς και μετεωρολογικών συνθηκών για την χρονική περίοδο 1/3/2013 έως 28/2/2017. Συγκεκριμένα δόθηκαν τα παρακάτω δεδομένα:

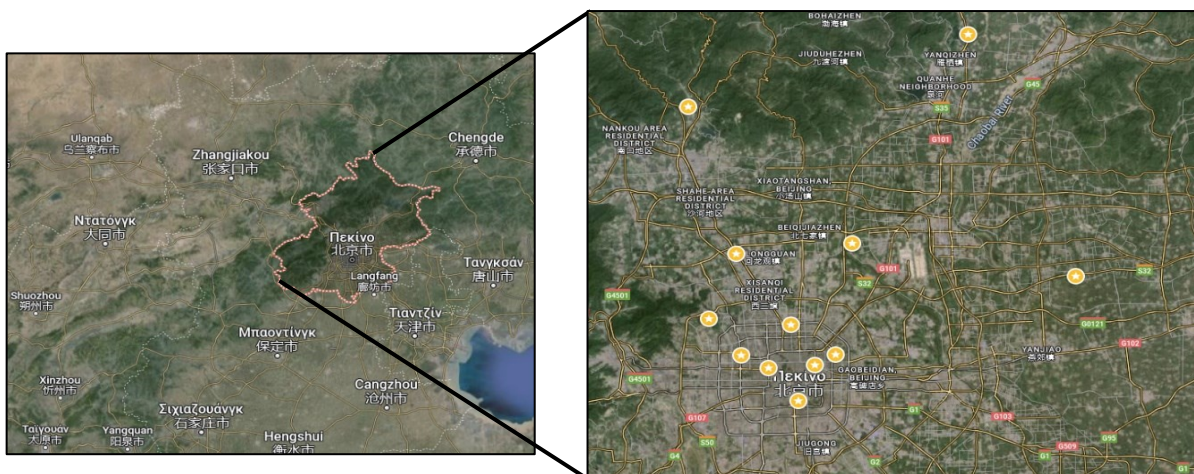
- Συγκέντρωση σωματιδίων αεροδυναμικής διαμέτρου έως και 2.5 μικρόμετρα, PM_{2.5}.
- Συγκέντρωση σωματιδίων αεροδυναμικής διαμέτρου έως και 10 μικρόμετρα, PM₁₀.
- Συγκέντρωση διοξειδίου του θείου, SO₂.
- Συγκέντρωση διοξειδίου του αζώτου, NO₂.
- Συγκέντρωση μονοξειδίου του άνθρακα, CO.
- Συγκέντρωση όζοντος, O₃.
- Θερμοκρασία ατμοσφαιρικού αέρα, TEMP.
- Θερμοκρασία δρόσου, DEWP.
- Ατμοσφαιρική πίεση, PRES.
- Ύψος βροχής, RAIN.
- Ταχύτητα ανέμου, WSPM.
- Διεύθυνση ανέμου, Wd.

Οι παραπάνω τιμές ήταν διαθέσιμες για δώδεκα (12) συνολικά περιοχές της πόλης του Πεκίνο, συγκεκριμένα :

Περιοχή 1	Aotizhongxin
Περιοχή 2	Changping
Περιοχή 3	Dongsi
Περιοχή 4	Guanyuan
Περιοχή 5	Gucheng
Περιοχή 6	Nongzhanguan
Περιοχή 7	Shunyi
Περιοχή 8	Tiantan
Περιοχή 9	Wanliu
Περιοχή 10	Wanshouxigong
Περιοχή 11	Dingling
Περιοχή 12	Huaiyou



ΣΧΗΜΑ 7 : Λογικό διάγραμμα της διαδικασίας εξαγωγής των προγνωστικών μοντέλων.



ΣΧΗΜΑ 8 : Απεικόνιση μέρους της βόρειο-ανατολικής Κίνας μέσω Google Maps (αριστερά) και της πόλης του Πεκίνο όπου φαίνονται οι επιμέρους περιοχές μελέτης (δεξιά).

Οι παραπάνω περιοχές παρουσιάζονται και στο **Σχήμα 8** (Περιοχές με αστερίσκο). Τα δεδομένα αποτελούν μετρήσεις του Chen Song [42] και αποκτήθηκαν δωρεάν μέσω του κάτωθι συνδέσμου «<https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data>». Λόγω της φύσης της εργασίας, επιλέχθηκε η απόσταση των υπό μελέτη περιοχών να μην υπερβαίνει τα 20 χιλιόμετρα (km), εξασφαλίζοντας μεγαλύτερη ακρίβεια αποτελεσμάτων. Κρίθηκε σκόπιμος ο αποκλεισμός τριών (3) περιοχών: Περιοχή 7 (Shunyi), Περιοχή 11 (Dingling) και Περιοχή 12 (Huairou), καθώς θεωρήθηκε ότι οι μετρήσεις τους θα επηρέαζαν αρνητικά τη διαδικασία με την οποία ο αλγόριθμος εκπαίδευσης των ΤΝΔ θα συσχετιζε τα δεδομένα. Πριν από οποιαδήποτε επεξεργασία είναι απαραίτητος ο έλεγχος των δεδομένων για κενά στοιχεία. Όλος ο προκαταρκτικός έλεγχος έγινε με τη χρήση του Microsoft Excel 2021, τα αποτελέσματα του οποίου βρίσκονται στον **Πίνακα 6**. Συνολικά και στις εννέα (9) περιοχές βρέθηκαν 7600 κενά στοιχεία σε πλήθος 420780 στοιχείων, ανά περιοχή. Πριν γίνει η προ-επεξεργασία των μετρήσεων, έγινε έλεγχος για «ίδιες» τιμές μεταξύ των περιοχών. Αυτό το οποίο βρέθηκε ήταν ότι οι περιοχές 1 και 4, καθώς και οι 3,6 και 8, είχαν ακριβώς τις ίδιες τιμές στα μετεωρολογικά τους δεδομένα (Temperature, Pressure, Dew Point, Rain, Wind Direction και Wind Speed).

Μεταβλητή	Κενά	Μεταβλητή	Κενά	Σύνολο Κενών
Year	0	CO	3197	7600
Month	0	O ₃	664	
Day	0	TEMP	20	
Hour	0	PRES	20	
PM _{2.5}	750	DEWP	20	
PM ₁₀	553	RAIN	20	
SO ₂	663	wd	78	
NO ₂	1601	WSPM	14	

ΠΙΝΑΚΑΣ 6 : Ενδεικτικό πλήθος κενών στοιχείων στις μεταβλητές εισόδου για την Περιοχή 3.

4.2. Κάλυψη Κενών Στοιχείων

Η συνήθης μεθοδολογία υπαγορεύει ότι κατά τον έλεγχο των δεδομένων στο ενδεχόμενο εύρεσης κενών στοιχείων, ο ερευνητής επιλέγει να διαγράψει ολόκληρη τη σειρά, υπό προϋποθέσεις. Οι προϋποθέσεις αυτές ορίζονται συνήθως από τους ίδιους τους ερευνητές, αλλά παρουσιάζουν παρόμοιο μοτίβο. Συνήθως γίνεται έλεγχος για το αν υπάρχουν πάνω από X συνεχόμενες κενές τιμές ή πάνω από Ψ κενά στοιχεία εντός της ημέρας, όπου $X < \Psi$ και αν ισχύει κάποια από τις δύο συνθήκες, διαγράφεται ολόκληρη η ημέρα. Ωστόσο μια τέτοια αντιμετώπιση είναι εφικτή όταν υπάρχει μεγάλο πλήθος διαθέσιμων δεδομένων ή/και μικρό πλήθος κενών στοιχείων. Συγκεκριμένα αν ακολουθηθεί η παραπάνω μεθοδολογία, ορίζοντας ως συνθήκες $X \geq 4$ και $\Psi \geq 6$, τότε μπορούμε να υπολογίσουμε τις εξής θεωρητικές περιπτώσεις :

A) Τα 7600 κενά στοιχεία βρίσκονται κατανεμημένα ημερησίως έτσι ώστε $X=4$. Τότε το πλήθος των ημερών που θα διαγραφούν θα ανέρχεται στις $7600/4 = 1900$ ημέρες και το σύνολο των υπό διαγραφή δεδομένων ίσο με $1900 \text{ ημέρες} \cdot 24 \text{ ώρες} \cdot 12 \text{ μετρήσεις} = 547200$ μετρήσεις, δηλαδή το 130% των δεδομένων, κάθε περιοχής. Προφανώς ένα τέτοιο σενάριο είναι αδύνατο.

B) Τα 7600 κενά στοιχεία βρίσκονται κατανεμημένα ημερησίως έτσι ώστε $X < 4$ και $\Psi = 6$. Τότε το πλήθος των ημερών που θα διαγραφούν θα ανέρχεται στις $7600/6 \approx 1226$ ημέρες και το σύνολο των υπό διαγραφή δεδομένων ίσο με $1226 \text{ ημέρες} \cdot 24 \text{ ώρες} \cdot 12 \text{ μετρήσεις} = 353088$ μετρήσεις, δηλαδή το 83.91% των δεδομένων, κάθε περιοχής.

Γ) Τα 7600 κενά στοιχεία βρίσκονται κατανεμημένα ημερησίως έτσι ώστε $X < 4$ και $\Psi < 6$. Τότε το πλήθος των κενών είναι ίσο με $7600 \cdot 12 = 91200$, δηλαδή το 21.67% των δεδομένων, κάθε περιοχής.

Οι παραπάνω θεωρητικές περιπτώσεις παρουσιάζουν τα ακραία σενάρια, στα οποία η απώλωση δεδομένων θα ήταν είτε η μέγιστη είτε η ελάχιστη δυνατή. Φυσικά τα πραγματικά σενάρια διαφέρουν από τα παραπάνω προσεγγιστικά, καθώς οι κενές τιμές παρουσιάζουν μια τυχαιότητα ως προς την εμφάνισή τους. Ωστόσο ακόμα και για το ιδανικό σενάριο, όταν το σύνολο δεδομένων δεν είναι επαρκούς μεγέθους, τότε η εφαρμογή της συγκεκριμένης μεθοδολογίας θα οδηγούσε σε πιθανή μείωση δεδομένων κατά 20%. Παρακάτω παρουσιάζονται τα 6 πιθανά σενάρια μείωσης των διαθέσιμων δεδομένων βάσει της κατανομής αυτών, εντός κάθε σετ δεδομένων. Τα πρώτα 3 σενάρια δεν αναπαριστούν πραγματικές μετρήσεις και δεν έχουν θεωρηθεί τιμές X και Ψ . Στα επόμενα 3 σενάρια αναπαρίστανται πραγματικά δεδομένα και έχει γίνει η θεώρηση των τιμών $X \geq 4$ και $\Psi \geq 6$.

Σενάριο 1 (Βέλτιστο Υποθετικό Σενάριο) : Όλη η σειρά απαρτίζεται (ή τουλάχιστον δομείται σε μεγάλο βαθμό) από κενές τιμές. Σε αυτό το σενάριο παρατηρείται η ελάχιστη μείωση δεδομένων, περίπου της τάξης του 1.8% ($7600/420780 = 1.8\%$).

a/a	Year	Month	Day	Hour	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	WD	WSPM
1	2013	3	1	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	2013	3	1	1	8	8	4	7	0.30	77	-1.1	1023.2	-18.2	0	4	4.7
3	2013	3	1	2	7	7	5	10	0.30	73	-1.1	1023.5	-18.2	0	4	5.6
4	2013	3	1	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5	2013	3	1	4	3	3	12	12	0.30	72	-2	1025.2	-19.5	0	1	2
6	2013	3	1	5	5	5	18	18	0.40	66	-2.2	1025.6	-19.6	0	16	3.7
7	2013	3	1	6	3	3	18	32	0.50	50	-2.6	1026.5	-19.1	0	16	2.5
8	2013	3	1	7	3	6	19	41	0.50	43	-1.6	1027.4	-19.1	0	15	3.8
9	2013	3	1	8	3	6	16	43	0.50	45	0.1	1028.3	-19.2	0	16	4.1
10	2013	3	1	9	3	8	12	28	0.40	59	1.2	1028.5	-19.3	0	16	2.6
11	2013	3	1	10	3	6	9	12	0.40	72	1.9	1028.2	-19.4	0	1	3.6
12	2013	3	1	11	3	6	9	14	0.40	71	2.9	1028.2	-20.5	0	2	3.7
13	2013	3	1	12	3	6	7	13	0.30	74	3.9	1027.3	-19.7	0	1	5.1
14	2013	3	1	13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
15	2013	3	1	14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
16	2013	3	1	15	8	15	7	14	0.40	76	6.2	1025.7	-18.6	0	1	2.8
17	2013	3	1	16	9	19	9	13	0.40	76	5.9	1025.6	-18.1	0	3	3.9
18	2013	3	1	17	10	23	11	15	0.40	74	4.3	1026.3	-18.7	0	4	2.8
19	2013	3	1	18	11	20	8	20	0.50	70	3.1	1027.4	-18.4	0	3	2.1
21	2013	3	1	19	8	14	12	30	0.50	60	2.3	1028.3	-18.4	0	4	2.8
22	2013	3	1	20	11	17	13	33	0.60	55	1.7	1029.1	-17.3	0	3	2.1
23	2013	3	1	21	12	18	16	35	0.50	50	0.6	1030.1	-16.7	0	5	0.8
24	2013	3	1	22	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
25	2013	3	1	23	24	24	26	54	0.60	36	-0.2	1030.5	-17.4	0	3	1.4
26	2013	3	2	0	22	24	24	44	0.50	44	-0.4	1031	-17.6	0	11	1.4
27	2013	3	2	1	14	17	21	36	0.40	50	-1	1031.3	-17.3	0	10	1.1
28	2013	3	2	2	13	13	20	37	0.40	47	-1.5	1030.9	-16.9	0	10	1.7
29	2013	3	2	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
30	2013	3	2	4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
31	2013	3	2	5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
32	2013	3	2	6	4	10	28	46	0.50	39	-2.5	1029.6	-17.7	0	2	0.7
33	2013	3	2	7	3	11	34	58	0.50	27	-1.7	1029.8	-17	0	16	1.2
34	2013	3	2	8	3	7	21	49	0.50	43	-0.4	1029.6	-17.6	0	2	1.8
35	2013	3	2	9	10	14	47	62	0.70	29	0.6	1029.7	-16.7	0	12	1.7
36	2013	3	2	10	11	19	36	60	0.60	34	1.7	1028.9	-16.3	0	11	1.6
37	2013	3	2	11	18	29	36	70	0.80	27	2.2	1028.2	-16.8	0	6	3.1
38	2013	3	2	12	26	30	25	76	0.90	22	2.7	1027.3	-16.4	0	12	2.7
39	2013	3	2	13	25	33	20	57	0.70	39	3.3	1025.7	-16.4	0	7	1.5
40	2013	3	2	14	26	35	22	54	0.70	43	3.8	1024.8	-16	0	9	1.3
41	2013	3	2	15	37	40	24	55	0.70	44	3.9	1024.1	-16.5	0	12	1.2
42	2013	3	2	16	37	46	29	63	1.00	38	3.5	1023.6	-15.2	0	14	2.5
43	2013	3	2	17	44	58	45	77	1.20	30	2.2	1023.4	-13.6	0	5	0.4
44	2013	3	2	18	54	79	50	80	1.10	24	1.2	1023.2	-12.3	0	3	0.9
45	2013	3	2	19	61	86	50	97	1.30	10	1.5	1022.9	-13.8	0	3	1.2
46	2013	3	2	20	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
47	2013	3	2	21	74	103	72	107	1.40	2	0.6	1022.6	-14.2	0	1	1.3
48	2013	3	2	22	81	113	84	110	1.50	3	-0.6	1022.3	-13.9	0	2	1.4
49	2013	3	3	23	112	130	73	100	1.90	2	-1.4	1020.4	-13	0	1	1.2

ΠΙΝΑΚΑΣ 7α : Πλήθος κενών στοιχείων, βάσει σεναρίου 1, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες.

Σενάριο 2 (Ρεαλιστικό Υποθετικό Σενάριο) : Όλη η σειρά απαρτίζεται μερικώς από κενές τιμές .
 Σε αυτό το σενάριο παρατηρείται μικρή μείωση δεδομένων αλλά, μεγαλύτερη της τάξης του 1.8%.

a/a	Year	Month	Day	Hour	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	WD	WSPM
1	2013	3	1	0	NA	NA	NA	7	0.30	77	-0.7	1023	-18.8	0	5	4.4
2	2013	3	1	1	8	8	4	7	0.30	77	-1.1	1023.2	-18.2	0	4	4.7
3	2013	3	1	2	7	7	5	10	0.30	73	-1.1	1023.5	-18.2	0	4	5.6
4	2013	3	1	3	6	6	11	11	0.30	72	-1.4	1024.5	-19.4	0	2	3.1
5	2013	3	1	4	NA	3	12	12	NA	NA	NA	NA	-19.5	0	1	2
6	2013	3	1	5	5	5	18	18	0.40	66	-2.2	1025.6	-19.6	0	16	3.7
7	2013	3	1	6	3	3	18	32	0.50	50	-2.6	1026.5	-19.1	0	16	2.5
8	2013	3	1	7	3	6	19	41	0.50	43	-1.6	1027.4	-19.1	0	15	3.8
9	2013	3	1	8	3	6	16	43	0.50	45	0.1	1028.3	-19.2	0	16	4.1
10	2013	3	1	9	3	8	12	28	0.40	59	1.2	1028.5	-19.3	0	16	2.6
11	2013	3	1	10	3	6	9	12	0.40	72	1.9	1028.2	-19.4	0	1	3.6
12	2013	3	1	11	3	6	9	14	0.40	71	2.9	1028.2	-20.5	0	2	3.7
13	2013	3	1	12	3	6	7	13	0.30	74	3.9	1027.3	-19.7	0	1	5.1
14	2013	3	1	13	3	6	7	12	0.40	76	5.3	1026.2	-19.3	0	3	4.3
15	2013	3	1	14	6	9	7	11	0.40	77	6	1025.9	-19.6	0	1	4.4
16	2013	3	1	15	8	15	7	14	0.40	76	6.2	1025.7	-18.6	0	1	2.8
17	2013	3	1	16	9	19	9	13	0.40	76	5.9	1025.6	-18.1	0	3	3.9
18	2013	3	1	17	10	NA	NA	NA	NA	74	4.3	1026.3	NA	0	4	2.8
19	2013	3	1	18	11	20	8	20	0.50	70	3.1	1027.4	-18.4	0	3	2.1
21	2013	3	1	19	8	14	12	30	0.50	60	2.3	1028.3	-18.4	0	4	2.8
22	2013	3	1	20	11	17	13	33	0.60	55	1.7	1029.1	-17.3	0	3	2.1
23	2013	3	1	21	12	18	16	35	0.50	50	0.6	1030.1	-16.7	NA	NA	NA
24	2013	3	1	22	NA	19	21	57	0.70	32	0.9	1030.5	-17.4	0	5	1.8
25	2013	3	1	23	NA	24	26	54	0.60	36	NA	1030.5	-17.4	0	3	1.4
26	2013	3	2	0	NA	24	24	44	0.50	44	NA	1031	-17.6	0	11	1.4
27	2013	3	2	1	14	17	21	36	0.40	50	-1	1031.3	-17.3	0	10	1.1
28	2013	3	2	2	13	13	20	37	0.40	47	-1.5	1030.9	-16.9	0	10	1.7
29	2013	3	2	3	3	9	13	34	0.40	52	-1.4	1030.6	-17.6	0	11	1.4
30	2013	3	2	4	3	7	NA	43	0.40	43	-1.5	1030.8	-17.7	0	15	0.9
31	2013	3	2	5	9	11	NA	70	0.50	20	-1.8	NA	NA	NA	1	2
32	2013	3	2	6	4	10	28	46	0.50	39	-2.5	1029.6	-17.7	0	2	0.7
33	2013	3	2	7	3	11	34	58	0.50	27	-1.7	1029.8	-17	0	16	1.2
34	2013	3	2	8	3	7	21	49	0.50	43	-0.4	1029.6	-17.6	0	2	1.8
35	2013	3	2	9	10	14	47	62	0.70	29	0.6	1029.7	-16.7	0	12	1.7
36	2013	3	2	10	11	19	36	60	0.60	34	1.7	1028.9	-16.3	0	11	1.6
37	2013	3	2	11	18	29	36	70	0.80	27	2.2	1028.2	-16.8	0	6	3.1
38	2013	3	2	12	26	30	25	76	0.90	22	2.7	1027.3	-16.4	0	12	2.7
39	2013	3	2	13	25	33	20	57	0.70	39	3.3	1025.7	-16.4	0	7	1.5
40	2013	3	2	14	26	35	22	54	0.70	43	3.8	1024.8	-16	0	9	1.3
41	2013	3	2	15	37	40	24	55	0.70	44	3.9	1024.1	-16.5	0	12	1.2
42	2013	3	2	16	37	46	29	63	1.00	38	NA	NA	NA	NA	14	2.5
43	2013	3	2	17	44	58	45	77	1.20	30	2.2	1023.4	-13.6	0	5	0.4
44	2013	3	2	18	54	79	50	80	1.10	24	1.2	1023.2	NA	NA	3	0.9
45	2013	3	2	19	61	86	50	97	1.30	10	1.5	1022.9	-13.8	0	3	1.2
46	2013	3	2	20	67	96	54	108	1.40	2	1.2	1022.9	-14.1	0	3	1.2
47	2013	3	2	21	74	103	72	107	1.40	2	0.6	1022.6	-14.2	0	1	1.3
48	2013	3	2	22	81	113	84	110	1.50	3	-0.6	1022.3	-13.9	0	2	1.4
49	2013	3	3	23	112	130	73	100	1.90	2	-1.4	1020.4	-13	0	1	1.2

ΠΙΝΑΚΑΣ 7b: Πλήθος κενών στοιχείων, βάσει σεναρίου 2, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες..

Σενάριο 3 (Χείριστο Υποθετικό Σενάριο) : Όλη η σειρά περιέχει μονάχα μια κενή τιμή . Σε αυτό το σενάριο παρατηρείται μεγάλη μείωση δεδομένων, πολύ μεγαλύτερη της τάξης του 1.8%, έως και 20% ($12 \cdot 7600/420780 = 21.67\%$).

α/α	Year	Month	Day	Hour	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	WD	WSPM
1	2013	3	1	0	NA	4	4	7	0.30	77	-0.7	1023	-18.8	0	5	4.4
2	2013	3	1	1	8	8	4	7	0.30	77	NA	1023.2	-18.2	0	4	4.7
3	2013	3	1	2	7	7	5	NA	0.30	73	-1.1	1023.5	-18.2	0	4	5.6
4	2013	3	1	3	6	6	11	11	0.30	72	-1.4	1024.5	-19.4	0	2	3.1
5	2013	3	1	4	3	3	12	12	0.30	72	-2	NA	-19.5	0	1	2
6	2013	3	1	5	5	5	NA	18	0.40	66	-2.2	1025.6	-19.6	0	16	3.7
7	2013	3	1	6	3	3	18	32	0.50	50	-2.6	1026.5	-19.1	0	16	2.5
8	2013	3	1	7	3	NA	19	41	0.50	43	-1.6	1027.4	-19.1	0	15	3.8
9	2013	3	1	8	3	6	16	43	0.50	45	0.1	1028.3	-19.2	0	16	4.1
10	2013	3	1	9	3	8	12	28	0.40	59	1.2	1028.5	-19.3	0	16	2.6
11	2013	3	1	10	3	6	9	12	0.40	72	1.9	1028.2	-19.4	0	1	3.6
12	2013	3	1	11	3	6	9	14	0.40	71	2.9	1028.2	NA	0	2	3.7
13	2013	3	1	12	3	6	7	13	0.30	74	3.9	1027.3	-19.7	0	1	5.1
14	2013	3	1	13	3	6	7	12	0.40	76	5.3	1026.2	-19.3	0	3	4.3
15	2013	3	1	14	6	9	7	11	0.40	77	6	1025.9	-19.6	0	1	4.4
16	2013	3	1	15	8	15	7	14	0.40	76	6.2	1025.7	-18.6	0	1	2.8
17	2013	3	1	16	9	19	9	13	0.40	76	5.9	1025.6	-18.1	0	3	3.9
18	2013	3	1	17	10	23	11	15	0.40	74	4.3	1026.3	-18.7	0	4	2.8
19	2013	3	1	18	11	20	8	20	0.50	70	3.1	1027.4	-18.4	0	3	2.1
21	2013	3	1	19	8	14	12	30	0.50	60	2.3	1028.3	-18.4	0	4	2.8
22	2013	3	1	20	11	17	13	33	0.60	55	1.7	1029.1	-17.3	0	3	2.1
23	2013	3	1	21	12	18	NA	35	0.50	50	0.6	1030.1	-16.7	0	5	0.8
24	2013	3	1	22	15	19	21	57	0.70	32	0.9	1030.5	-17.4	0	5	1.8
25	2013	3	1	23	24	24	26	54	0.60	36	-0.2	1030.5	-17.4	0	3	1.4
26	2013	3	2	0	22	24	24	44	0.50	44	-0.4	1031	-17.6	0	11	NA
27	2013	3	2	1	14	17	21	36	0.40	50	-1	1031.3	-17.3	0	10	NA
28	2013	3	2	2	13	13	20	37	0.40	47	-1.5	1030.9	-16.9	0	10	NA
29	2013	3	2	3	3	9	13	34	0.40	52	-1.4	1030.6	-17.6	0	11	1.4
30	2013	3	2	4	NA	7	18	43	0.40	43	-1.5	1030.8	-17.7	0	15	0.9
31	2013	3	2	5	NA	11	19	70	0.50	20	-1.8	1030.1	-17.5	0	1	2
32	2013	3	2	6	NA	10	28	46	0.50	39	-2.5	1029.6	-17.7	0	2	0.7
33	2013	3	2	7	NA	11	34	58	0.50	27	-1.7	1029.8	-17	0	16	1.2
34	2013	3	2	8	3	7	21	49	0.50	43	-0.4	1029.6	-17.6	0	2	1.8
35	2013	3	2	9	10	14	47	62	0.70	29	0.6	1029.7	-16.7	0	12	1.7
36	2013	3	2	10	11	19	36	60	0.60	34	1.7	1028.9	-16.3	0	11	1.6
37	2013	3	2	11	18	29	36	70	0.80	27	2.2	1028.2	-16.8	0	6	3.1
38	2013	3	2	12	26	30	25	76	0.90	22	NA	1027.3	-16.4	0	12	2.7
39	2013	3	2	13	25	33	20	57	0.70	39	3.3	1025.7	-16.4	0	7	1.5
40	2013	3	2	14	26	35	22	54	0.70	43	3.8	1024.8	-16	0	9	1.3
41	2013	3	2	15	37	40	24	55	0.70	44	3.9	1024.1	-16.5	0	12	1.2
42	2013	3	2	16	37	46	29	63	1.00	38	3.5	1023.6	-15.2	0	14	2.5
43	2013	3	2	17	44	58	45	77	1.20	30	2.2	1023.4	-13.6	0	5	0.4
44	2013	3	2	18	54	79	50	80	1.10	24	1.2	1023.2	-12.3	0	3	0.9
45	2013	3	2	19	61	86	50	97	1.30	10	1.5	1022.9	-13.8	0	3	1.2
46	2013	3	2	20	67	96	54	108	1.40	2	1.2	1022.9	-14.1	0	3	1.2
47	2013	3	2	21	74	103	72	107	1.40	2	0.6	1022.6	-14.2	0	1	1.3
48	2013	3	2	22	81	113	84	110	1.50	3	-0.6	1022.3	-13.9	0	2	1.4
49	2013	3	2	23	93	120	87	107	1.70	2	-0.8	1021.1	-13.4	0	2	1.3

ΠΙΝΑΚΑΣ 7c: Πλήθος κενών στοιχείων, βάσει σεναρίου 3, στις μεταβλητές εισόδου για την Περιοχή 3, για τις πρώτες δυο μέρες..

Σενάριο 4 (Βέλτιστο Πραγματικό Σενάριο) : Ταυτόχρονη ικανοποίηση και των δύο συνθηκών για πολλαπλούς ρύπους

a/a	Year	Month	Day	Hour	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	WD	WSPM
3625	2013	7	30	0	118.0	119.0	10.9	54.8	NA	50.0	26.5	994.2	23.6	0.0	11	1.0
3626	2013	7	30	1	103.0	115.0	10.0	52.4	NA	47.0	24.8	993.9	23.4	0.0	12	0.5
3627	2013	7	30	2	101.0	84.0	9.1	48.7	NA	50.0	24.4	993.4	23.5	0.0	13	1.5
3628	2013	7	30	3	101.0	97.0	9.4	65.3	NA	26.0	24.1	993.3	23.4	0.0	5	1.1
3629	2013	7	30	4	77.0	74.0	9.7	87.3	NA	2.0	23.8	993.7	23.3	0.0	3	0.8
3630	2013	7	30	5	78.0	136.0	10.9	88.7	NA	2.0	23.8	994.2	23.5	0.0	3	0.9
3631	2013	7	30	6	133.0	197.0	13.1	84.0	NA	2.0	25.2	994.4	24.2	0.0	4	0.8
3632	2013	7	30	7	146.0	209.0	13.1	87.5	NA	3.0	26.7	994.9	24.5	0.0	4	1.6
3633	2013	7	30	8	161.0	215.0	12.0	100.6	NA	12.0	28.3	994.9	24.9	0.0	4	1.4
3634	2013	7	30	9	190.0	226.0	11.7	100.8	NA	35.0	29.8	994.6	25.5	0.0	4	0.6
3635	2013	7	30	10	182.0	171.0	11.4	81.5	NA	98.0	31.4	994.2	23.2	0.0	5	1.7
3636	2013	7	30	11	184.0	188.0	13.1	93.2	NA	128.0	32.1	993.6	24.4	0.0	4	2.8
3637	2013	7	30	12	216.0	188.0	14.6	113.3	NA	156.0	33.0	992.9	22.8	0.0	12	2.1
3638	2013	7	30	13	215.0	193.0	15.1	76.0	NA	254.0	34.2	992.3	20.9	0.0	7	2.2
3639	2013	7	30	14	215.0	215.0	16.9	39.6	NA	289.0	35.0	991.8	20.9	0.0	6	1.2
3640	2013	7	30	15	90.0	35.0	13.1	25.5	NA	244.0	34.8	991.7	22.2	0.0	6	2.7
3641	2013	7	30	16	82.0	54.0	12.6	25.9	NA	239.0	34.3	991.1	21.7	0.0	15	2.6
3642	2013	7	30	17	82.0	82.0	NA	NA	NA	NA	33.7	991.0	21.2	0.0	9	1.8
3643	2013	7	30	18	NA	NA	NA	NA	NA	NA	32.1	991.9	21.9	0.0	10	1.6
3644	2013	7	30	19	NA	NA	NA	NA	NA	NA	31.2	992.4	23.3	0.0	11	1.3
3645	2013	7	30	20	NA	NA	NA	NA	NA	NA	30.4	993.4	22.6	0.0	9	0.8
3646	2013	7	30	21	NA	NA	NA	NA	NA	NA	29.8	994.0	22.8	0.0	8	1.1
3647	2013	7	30	22	NA	NA	NA	NA	NA	NA	29.3	995.6	24.0	0.0	10	1.8
3648	2013	7	30	23	NA	NA	NA	NA	NA	NA	21.7	997.6	20.3	6.7	10	6.7
3649	2013	7	31	0	NA	NA	NA	NA	NA	NA	21.7	996.8	21.0	8.2	4	0.8
3650	2013	7	31	1	NA	NA	NA	NA	NA	NA	21.9	995.1	21.1	1.5	14	3.6
3651	2013	7	31	2	NA	NA	NA	NA	NA	NA	21.9	995.1	21.2	0.0	10	1.8
3652	2013	7	31	3	NA	NA	NA	NA	NA	NA	22.2	995.2	21.5	0.0	6	1.3
3653	2013	7	31	4	NA	NA	NA	NA	NA	NA	22.2	995.1	21.4	0.1	14	1.1
3654	2013	7	31	5	NA	NA	NA	NA	NA	NA	22.3	995.4	21.8	0.0	13	1.1
3655	2013	7	31	6	NA	NA	NA	NA	NA	NA	23.4	995.7	22.4	0.0	12	1.4
3656	2013	7	31	7	NA	NA	NA	NA	NA	NA	24.8	996.0	22.3	0.0	5	1.2
3657	2013	7	31	8	NA	NA	NA	NA	NA	NA	26.7	996.2	21.2	0.0	5	1.2
3658	2013	7	31	9	NA	NA	NA	NA	NA	NA	28.2	996.1	21.2	0.0	7	1.4
3659	2013	7	31	10	NA	NA	NA	NA	NA	NA	30.0	995.8	22.2	0.0	8	1.4
3660	2013	7	31	11	NA	NA	16.0	66.9	NA	84.0	31.2	995.6	20.2	0.0	9	1.1
3661	2013	7	31	12	30.0	30.0	9.4	50.7	NA	144.0	32.0	994.8	20.6	0.0	8	1.0
3662	2013	7	31	13	46.0	46.0	10.6	34.7	NA	212.0	32.6	994.6	21.8	0.0	12	2.7
3663	2013	7	31	14	57.0	57.0	12.0	28.5	NA	238.0	32.5	993.9	22.6	0.0	11	2.7
3664	2013	7	31	15	75.0	75.0	11.1	31.0	NA	249.0	32.3	994.0	23.3	0.0	7	1.3
3665	2013	7	31	16	78.0	78.0	10.9	31.6	NA	249.0	32.2	993.8	23.4	0.0	7	1.7
3666	2013	7	31	17	81.0	81.0	10.6	27.9	NA	234.0	30.5	993.6	21.9	0.0	8	1.4
3667	2013	7	31	18	107.0	107.0	10.6	33.9	NA	234.0	29.5	994.0	22.7	0.0	8	1.5
3668	2013	7	31	19	123.0	9.0	10.3	48.5	NA	221.0	24.8	996.0	20.7	0.0	8	5.4
3669	2013	7	31	20	123.0	123.0	7.4	41.7	NA	179.0	23.5	997.5	21.8	1.1	6	2.8
3670	2013	7	31	21	26.0	26.0	2.0	17.7	NA	143.0	23.2	998.5	22.2	1.5	15	2.5
3671	2013	7	31	22	22.0	22.0	0.6	23.2	NA	120.0	22.4	998.0	22.1	3.1	15	3.1
3672	2013	7	31	23	33.0	33.0	0.3	45.2	NA	83.0	22.1	998.3	21.4	0.0	5	1.5

ΠΙΝΑΚΑΣ 7d: Πλήθος κενών στοιχείων, βάσει σεναρίου 4, στις μεταβλητές εισόδου για την Περιοχή 3, για δυο μέρες..

Σενάριο 5 (Σύνηθες Πραγματικό Σενάριο) : Ικανοποίηση μιας εκ των δύο συνθηκών για πολλούς ρύπους. Ικανοποιείται μόνο η συνθήκη $X \geq 4$, με αποτέλεσμα να διαγραφεί ολόκληρη η ημέρα (19/12/2014).

α/α	Year	Month	Day	Hour	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	WD	WSPM
15769	2014	12	18	0	115	189	54	92	3000	2	-5	1023	-16	0	5	1.7
15770	2014	12	18	1	96	156	55	84	2700	2	-6	1023	-15.3	0	5	0.8
15771	2014	12	18	2	109	185	72	82	3100	2	-7	1026	-14.8	0	4	1
15772	2014	12	18	3	108	174	85	84	2100	2	-6	1028	-15	0	3	0.8
15773	2014	12	18	4	96	171	88	79	2500	4	-6	1024	-13.6	0	6	0.9
15774	2014	12	18	5	95	164	101	76	4200	4	-7	1025	-14.3	0	1	0.1
15775	2014	12	18	6	99	172	103	72	5000	4	-8	1028	-14.4	0	2	1.2
15776	2014	12	18	7	95	177	103	69	4900	3	-8	1024	-14.2	0	6	1.2
15777	2014	12	18	8	87	176	97	71	4500	4	-9	1028	-14.8	0	12	0.9
15778	2014	12	18	9	82	162	91	68	4200	4	-5	1026	-12.7	0	3	0.7
15779	2014	12	18	10	84	140	99	64	4100	7	-1	1027	-10.1	0	5	1.3
15780	2014	12	18	11	89	145	92	64	3900	11	0	1025	-8.7	0	5	2.9
15781	2014	12	18	12	99	155	89	68	3800	14	0	1027	-8.7	0	6	2.5
15782	2014	12	18	13	100	163	81	69	3400	18	1	1024	-8	0	4	1.8
15783	2014	12	18	14	107	159	78	74	3200	17	2	1024	-7.6	0	5	2.3
15784	2014	12	18	15	121	183	84	82	3400	12	1	1022	-8.3	0	6	2.1
15785	2014	12	18	16	135	206	82	86	3800	8	1	1022	-7.5	0	9	1.3
15786	2014	12	18	17	137	217	72	86	4100	5	0	1021	-7.3	0	8	0.7
15787	2014	12	18	18	175	266	71	90	4500	4	-1	1021	-7	0	9	0.4
15788	2014	12	18	19	194	270	74	91	4500	4	-1	1021	-6.8	0	7	0.7
15789	2014	12	18	20	179	258	72	89	4600	4	-2	1023	-7.1	0	5	0.9
15790	2014	12	18	21	186	249	70	90	4800	4	-2	1021	-6.6	0	4	1
15791	2014	12	18	22	NA	NA	NA	NA	NA	NA	-4	1021	-8	0	3	1
15792	2014	12	18	23	NA	NA	NA	NA	NA	NA	-5	1021	-7.9	0	3	0.9
15793	2014	12	19	0	NA	NA	NA	NA	NA	NA	-6	1021	-8.4	0	7	0.3
15794	2014	12	19	1	NA	NA	NA	NA	NA	NA	-6	1016	-8	0	6	0.6
15795	2014	12	19	2	NA	NA	NA	NA	NA	NA	-6	1019	-8.1	0	5	0.2
15796	2014	12	19	3	NA	NA	NA	NA	NA	NA	-5	1019	-8.2	0	6	0.8
15797	2014	12	19	4	144	207	77	101	4800	3	-6	1015	-9.4	0	3	0.6
15798	2014	12	19	5	145	218	73	101	5100	3	-7	1016	-9.7	0	4	1
15799	2014	12	19	6	143	213	73	98	4800	3	-7	1017	-10.7	0	14	0.3
15800	2014	12	19	7	56	74	45	73	2300	8	0	1016	-15.2	0	15	2.2
15801	2014	12	19	8	8	49	16	34	600	37	1	1016	-14.3	0	1	3.8
15802	2014	12	19	9	8	56	23	38	600	38	2	1022	-14.2	0	16	4.8
15803	2014	12	19	10	5	43	16	27	500	51	3	1023	-13.8	0	16	3.5
15804	2014	12	19	11	8	42	12	14	400	67	3	1021	-14.7	0	14	6
15805	2014	12	19	12	8	24	14	11	400	69	3	1017	-15.2	0	15	6.1
15806	2014	12	19	13	4	19	13	12	300	68	4	1020	-14.8	0	14	4.6
15807	2014	12	19	14	6	14	9	12	300	69	4	1019	-17.6	0	16	4.8
15808	2014	12	19	15	4	27	5	12	300	71	3	1022	-18.4	0	16	5.6
15809	2014	12	19	16	3	27	5	12	300	71	3	1020	-19.1	0	16	4.3
15810	2014	12	19	17	4.3	29	4	11	200	74	2	1024	-19.3	0	16	4.4
15811	2014	12	19	18	3	35	4	21	400	55	1	1020	-19.5	0	15	3.9
15812	2014	12	19	19	6	38	5	20	400	55	0	1023	-18.2	0	1	2.5
15813	2014	12	19	20	4	33	5	21	400	49	0	1023	-19.2	0	16	2.9
15814	2014	12	19	21	4	31	5	18	400	47	-1	1026	-19.1	0	1	3.4
15815	2014	12	19	22	5	25	7	21	400	41	-2	1024	-19.5	0	1	3
15816	2014	12	19	23	6	20	10	30	600	29	-2	1024	-19	0	2	3.9

ΠΙΝΑΚΑΣ 7ε: Πλήθος κενών στοιχείων, βάσει σεναρίου 5, στις μεταβλητές εισόδου για την Περιοχή 3, για δυο μέρες..

Σενάριο 6 (Χείριστο Πραγματικό Σενάριο) : Ικανοποίηση μιας εκ των δύο συνθηκών για ένα ρύπο.
 Ικανοποιείται μόνο η συνθήκη $X \geq 4$, με αποτέλεσμα να διαγραφεί ολόκληρη η ημέρα
 (12/05/2016).

α/α	Year	Month	Day	Hour	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	WD	WSPM
28033	2016	5	12	0	240	240	5	68	1400	59	21.5	993.9	17.7	0.4	13	1.3
28034	2016	5	12	1	293	293	5	68	1400	48	21.2	993.7	18.2	0.2	11	1.5
28035	2016	5	12	2	237	237	3	34	1200	98	19.3	995.3	15.8	2.8	16	3.3
28036	2016	5	12	3	93	93	2	16	500	88	17.8	997	12.3	4.9	16	3.6
28037	2016	5	12	4	3	5	2	8	300	91	18	997.6	11.5	0	1	3
28038	2016	5	12	5	NA	5	2	8	300	89	18.1	998.1	10.7	0	16	2.5
28039	2016	5	12	6	NA	53	2	9	400	80	15	1001.7	12.1	3.1	2	1.6
28040	2016	5	12	7	3	5	2	27	500	52	14.4	1001.6	12.8	1.7	5	0.8
28041	2016	5	12	8	NA	5	2	32	600	51	15.2	1002.9	12	0.7	4	1.3
28042	2016	5	12	9	NA	5	2	26	500	57	17	1004.1	8.2	0.1	15	3.1
28043	2016	5	12	10	NA	22	2	10	400	55	16.8	1006.7	4.1	0	15	4.2
28044	2016	5	12	11	NA	21	2	12	300	54	16.6	1008.6	3.6	0	15	3.2
28045	2016	5	12	12	3	11	2	13	300	74	16.7	1009.5	0.3	0	16	4.9
28046	2016	5	12	13	3	42	2	6	300	86	17.1	1010	-1.1	0	15	5.1
28047	2016	5	12	14	27	27	2	8	300	83	17.8	1010.2	0	0	15	3.1
28048	2016	5	12	15	19	33	2	8	200	86	20.7	1009.7	-1.2	0	15	3.5
28049	2016	5	12	16	19	19	2	6	200	95	21.1	1010	-4.2	0	16	4.3
28050	2016	5	12	17	19	19	2	8	200	93	21.4	1010.4	-3.9	0	1	3.5
28051	2016	5	12	18	3	41	2	8	200	95	21.3	1010.9	-5.6	0	16	3.9
28052	2016	5	12	19	4	43	2	16	300	87	20.3	1011.7	-5.6	0	16	2.2
28053	2016	5	12	20	33	63	2	34	500	65	17.4	1012.6	0.1	0	16	0.6
28054	2016	5	12	21	66	66	NA	NA	NA	NA	14.7	1013.8	1.6	0	11	0.5
28055	2016	5	12	22	54	74	2	83	900	6	12.8	1014.4	2.6	0	0	0
28056	2016	5	12	23	48	48	2	77	800	8	11.2	1014.6	4.6	0	14	0.4
28057	2016	5	13	0	90	170	2	74	1500	2	10.6	1015.3	3.8	0	12	0.6
28058	2016	5	13	1	47	52	2	79	1300	2	9.6	1015.4	5	0	4	1.4
28059	2016	5	13	2	32	45	2	80	1400	2	9	1015.6	5.2	0	5	1.2
28060	2016	5	13	3	57	99	2	78	1700	2	8.3	1015.8	4.5	0	4	1.3
28061	2016	5	13	4	40	40	2	76	600	2	7.7	1016.6	5	0	10	0.3
28062	2016	5	13	5	36	43	2	68	800	2	7.3	1017.2	4.8	0	2	0.4
28063	2016	5	13	6	40	68	2	67	700	2	9.7	1017.7	6	0	3	1.4
28064	2016	5	13	7	48	78	2	70	900	4	13.5	1018.5	4.9	0	4	1.2
28065	2016	5	13	8	52	61	2	65	800	10	15.9	1018.9	5.2	0	5	1.2
28066	2016	5	13	9	118	141	2	68	800	16	18.3	1018.9	3	0	6	1
28067	2016	5	13	10	55	96	3	98	900	22	20.6	1018.2	3.8	0	7	2
28068	2016	5	13	11	36	77	4	31	300	83	22.5	1017.6	0.3	0	11	4.3
28069	2016	5	13	12	30	34	3	17	200	101	23.4	1016.9	-0.2	0	10	4.6
28070	2016	5	13	13	36	52	4	16	300	114	23.7	1015.9	0.7	0	11	4
28071	2016	5	13	14	41	41	NA	NA	NA	NA	24.2	1015.1	1.7	0	9	3.2
28072	2016	5	13	15	53	186	14	16	400	131	24.2	1014.4	1.7	0	10	3.8
28073	2016	5	13	16	49	95	14	17	500	146	23.6	1014.1	1.2	0	10	4.4
28074	2016	5	13	17	36	40	12	20	400	136	22.9	1013.7	1.2	0	9	4.2
28075	2016	5	13	18	35	35	4	19	200	127	22.3	1013.9	1.3	0	11	4.6
28076	2016	5	13	19	24	38	3	23	200	114	21.7	1014.2	1.9	0	10	2.3
28077	2016	5	13	20	26	38	3	29	200	104	21.2	1014.6	2.9	0	9	1.6
28078	2016	5	13	21	37	40	3	41	300	84	20.8	1015	2.6	0	9	1.7
28079	2016	5	13	22	42	50	4	56	300	62	20.3	1015.2	2.6	0	10	2.3
28080	2016	5	13	23	39	48	5	53	300	59	20	1014.9	3.2	0	10	2.4

ΠΙΝΑΚΑΣ 7f: Πλήθος κενών στοιχείων, βάσει σεναρίου 6, στις μεταβλητές εισόδου για την Περιοχή 3, για δυο μέρες..

Στα παραπάνω σενάρια φαίνεται η αναπόφευκτη απώλεια δεδομένων εισόδου, κάτι τέτοιο όμως θα επηρέαζε αρνητικά την εκπαίδευση του νευρωνικού δικτύου. Συνεπώς κρίνεται αναγκαία, η δημιουργία ενός νέου σεναρίου, στο οποίο υπάρχει μηδενική απώλεια δεδομένων. Ένας πιθανός τρόπος κάλυψης των κενών είναι η αντικατάσταση των τιμών «NA» με μηδενική τιμή, ή με τιμή πολλές τάξεις μεγαλύτερη από αυτές των υπόλοιπων δεδομένων. Ύστερα από αρκετές προσπάθειες και οι δύο παραπάνω ιδέες εγκαταλείφθηκαν, καθώς παρουσιάζονταν μεγάλα σφάλματα κατά τη διάρκεια της εκπαίδευσης των ΤΝΔ. Επομένως για την δημιουργία ενός σεναρίου με μηδενικές κενές τιμές, αποφασίστηκε να ακολουθηθούν οι παρακάτω 3 ξεχωριστές μέθοδοι κάλυψης τιμών.

4.2.1. Μέθοδος 1 : Χρήση ήδη υπαρχουσών τιμών και αξιοποίηση εντολών Microsoft Excel.

Το υπολογιστικό περιβάλλον του Microsoft Excel παρέχει τη δυνατότητα γρήγορου εντοπισμού κενών τιμών με τη χρήση των κατάλληλων φίλτρων. Προτού αξιοποιηθεί οποιαδήποτε εντολή είναι αναγκαία η μετατροπή όλων των μεταβλητών σε αριθμούς. Η διαδικασία αυτή είναι αρκετά απλή για την πλειοψηφία των αριθμών με μοναδική εξαίρεση 2 στήλες. Η πρώτη είναι η στήλη των συγκεντρώσεων του μονοξειδίου του άνθρακα, η οποία βρίσκεται σε διαφορετική κλίμακα μεγέθους ως προς τις άλλες συγκεντρώσεις. Η δεύτερη στήλη, η οποία χρειάστηκε την περισσότερη τροποποίηση ήταν η στήλη της κατεύθυνσης του ανέμου (Wind Direction), καθώς οι κατευθύνσεις αναγράφονταν με λατινικούς χαρακτήρες. Συνολικά υπήρξαν 16 διευθύνσεις, οι οποίες αντικαταστάθηκαν με έναν ακέραιο αριθμό όπως φαίνεται στον [Πίνακα 8](#).

Προκειμένου να εφαρμοστεί η μέθοδος αυτή γίνεται η βασική παραδοχή, ότι οι διαδοχικές τιμές εντός 5 ωρών δεν παρουσιάζουν μεγάλη απόκλιση μεταξύ τους. Φυσικά η παραπάνω παραδοχή έρχεται σε πλήρη αντίθεση με τη μεθοδολογία που ακολουθείται μέχρι τώρα, ωστόσο εδώ βρίσκεται και η πρώτη καινοτομία της παρούσας εργασίας. Έχοντας κάνει αυτή την παραδοχή ο τρόπος εργασίας είναι αρκετά απλός. Αρχικά γίνεται έλεγχος για την ύπαρξη τιμών εντός των προηγούμενων 5 ωρών και εντός των επόμενων 4 ωρών. Από τον έλεγχο αυτό προκύπτουν τέσσερα (4) επιμέρους σενάρια ως εξής :

1. Ύπαρξη και των δύο τιμών, εντός των χρονικών ορίων. Σε αυτή την περίπτωση η τιμή που λείπει είναι ο μέσος όρος των δύο τιμών.

2. Ύπαρξη μόνο της προηγούμενης τιμής, εντός των χρονικών ορίων. Σε αυτή την περίπτωση η τιμή που λείπει είναι ίση με την προηγούμενη τιμή αυξομειωμένη κατά ένα συγκεκριμένο νούμερο, το οποίο αναγράφεται στον Πίνακα 2 παρακάτω.

3. Ύπαρξη μόνο της επόμενης τιμής, εντός των χρονικών ορίων. Σε αυτή την περίπτωση η τιμή που λείπει είναι ίση με την επόμενη τιμή αυξομειωμένη κατά ένα συγκεκριμένο νούμερο, το οποίο αναγράφεται στον [Πίνακα 10](#) παρακάτω.

4. Ταυτόχρονη απουσία των δύο τιμών, εντός των χρονικών ορίων. Σε αυτή την περίπτωση το στοιχείο λαμβάνει την τιμή «NA», έτσι ώστε να γίνει μετέπειτα επεξεργασία του.

Η αυξομείωση των τιμών βρέθηκε μέσω υπολογισμού των αυξομειώσεων μεταξύ προηγούμενης και επόμενης τιμής, ακολουθούμενη από γραφική αναπαράσταση των τιμών αυτών σε ιστόγραμμα, με πλάτος ίσο με 1. Το πλάτος βρέθηκε πειραματικά, μεταβάλλοντας την τιμή του μέχρι η γραφική παράσταση να αποκτήσει την επιθυμητή μορφή μιας καμπάνας κανονικής κατανομής. Η ύπαρξη μιας γραφικής παράστασης υπό μορφή κανονικής κατανομής επιτρέπει τον ισομοιασμό των σφαλμάτων με το μεγαλύτερο πλήθος αυτών να εμφανίζεται στο κέντρο αυτής. Δεδομένου ότι το κέντρο της συγκεκριμένης καμπάνας είναι το ± 1 , γίνεται αντιληπτό το μικρό μέγεθος των σφαλμάτων. Η παραπάνω μέθοδος παρότι είναι αρκετά αποτελεσματική δεν είναι ικανή να καλύψει το 100% των κενών στοιχείων, ωστόσο παρουσιάζει μια μέση απόδοση άνω του 60%. Οι αποδόσεις παρουσιάζονται αναλυτικά στον [Πίνακα 9](#).

ΚΑΤΕΥΘΥΝΣΗ ΑΝΕΜΟΥ	ΑΚΕΡΑΙΟΣ	ΚΑΤΕΥΘΥΝΣΗ ΑΝΕΜΟΥ	ΑΚΕΡΑΙΟΣ
N	1	S	9
N-NE	2	S-SW	10
NE	3	SW	11
E-NE	4	W-SW	12
E	5	W	13
E-SE	6	W-NW	14
SE	7	NW	15
S-SE	8	N-NW	16

ΠΙΝΑΚΑΣ 8 : Πίνακας αντιστοίχισης φοράς ανέμου με τον αντίστοιχο ακέραιο αριθμό.

ΠΕΡΙΟΧΗ	1	2	3	4	5	6	7	8	9	10
ΡΥΠΟΣ	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ [%]									
PM _{2.5}	46.2	55.6	63.6	66.1	77.4	83.9	60.5	92.3	90.6	77.3
PM ₁₀	42.1	59.3	64.6	75.5	85.8	81.4	83.9	92.1	90.1	74.2
SO ₂	43.0	68.2	69.4	67.1	70.2	79.6	64.7	47.6	89.2	54.9
NO ₂	51.6	70.5	25.5	62.5	68.1	71.7	56.8	81.0	43.9	59.3
CO	46.7	66.9	19.3	39.9	47.5	58.8	54.6	74.7	31.3	63.8
O ₃	62.4	71.7	70.6	67.3	54.5	84.2	58.4	79.6	67.3	65.4

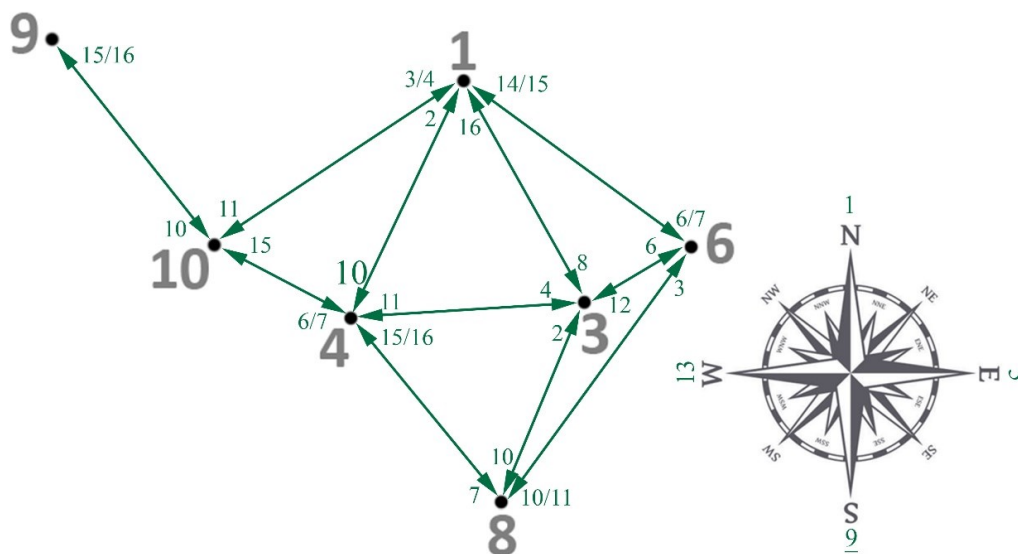
ΠΙΝΑΚΑΣ 9 : Ποσοστό επιτυχίας της 1ης μεθόδου κάλυψης κενών στοιχείων.

Μεταβλητή	Αριθμός Προσθαφαίρεσης	Μεταβλητή	Αριθμός Προσθαφαίρεσης
PM _{2.5}	1	TEMP	Δεν χρειάστηκαν κάποια προσθαφαίρεση.
PM ₁₀	1	PRES	
SO ₂	0.5	DEWP	
NO ₂	0.5	RAIN	
CO	50	wd	
O ₃	0.5	WSPM	

ΠΙΝΑΚΑΣ 10 : Αριθμοί προσθαφαίρεσης για κάθε ρύπο και μετεωρολογικό παράγοντα.

4.2.2. Μέθοδος 2 : Χρήση κώδικα εντός προγραμματιστικού περιβάλλοντος MATLAB

Η μέθοδος αυτή απαιτεί επίσης μια παραδοχή, ότι οι συγκεντρώσεις των ρύπων και οι τιμές των μετεωρολογικών παραγόντων δεν μεταβάλλονται σημαντικά σε περιοχές με απόσταση μικρότερη των δέκα (10) χιλιομέτρων. Αρχικά υπολογίζεται η απόσταση μεταξύ των εννέα (9) περιοχών μελέτης και για κάθε μια επιλέγονται οι πιο κοντινές σε αυτή περιοχές. Ας θέσουμε ως παράδειγμα την τοποθεσία 3 (Dongsì), αν υπολογιστούν οι αποστάσεις των άλλων περιοχών ως προς αυτή διαπιστώνεται ότι οι περιοχές 6,4,8 και 1 (κατά σειρά εγγύτητας) βρίσκονται σε κοντινή απόσταση. Συνεπώς είναι λογικό οι τιμές των πιο κοντινών σε αυτή περιοχών να επηρεάζουν περισσότερο τις τιμές ενδιαφέροντος. Σε επόμενο βήμα χρειάζεται να προσδιοριστεί η σχετική θέση των τοποθεσιών 1-4-6-8 ως προς την περιοχή 3 (Βόρεια, Ανατολικά, Βόρειο-Ανατολικά κλπ.). Προκειμένου να γίνει με ακρίβεια ο καθορισμός της σχετικής θέσης των περιοχών μεταξύ τους, είναι απαραίτητη η χρήση σχεδιαστικών προγραμμάτων. Αξιοποιώντας μια πυξίδα με 16 ακμές, τοποθετείται η κάθε περιοχή στο κέντρο αυτής. Ύστερα σημειώνεται η θέση των περιοχών, εντός ακτίνας 10 χιλιομέτρων, γύρω από αυτή. Με την εκτέλεση της παραπάνω διαδικασίας είναι δυνατή η συμπλήρωση του [Πίνακα 11](#). Στον παρακάτω πίνακα δεν περιλαμβάνονται οι περιοχές 2 και 5 καθώς η απόσταση τους από τις υπόλοιπες περιοχές μελέτης ξεπερνούσε τα δέκα (10) χιλιόμετρα. Στο [Σχήμα 9](#) φαίνεται η αλληλεπίδραση των περιοχών μεταξύ τους. Τέλος ακολουθεί η συγγραφή κατάλληλου κώδικα στο περιβάλλον MATLAB, έκδοση R2022a, βάσει του οποίου γίνεται έλεγχος της κατεύθυνσης του ανέμου και βάσει αυτής επιλέγεται η κατάλληλη περιοχή. Το κενό στοιχείο της περιοχής 3 λαμβάνει την ίδια ακριβώς τιμή με το αντίστοιχο στοιχείο της περιοχής που επιλέχθηκε. Η μέθοδος αυτή παρότι είναι αρκετά πιο περίπλοκη και αναλυτική από την πρώτη δεν καλύπτει μεγάλο πλήθος κενών, με ποσοστό επιτυχίας μικρότερο του 20%. Τα υπολειπόμενα κενά στοιχεία λαμβάνουν για ακόμα μια φορά την τιμή «NA», ώστε να τεθούν σε επεξεργασία σε επόμενο χρόνο.



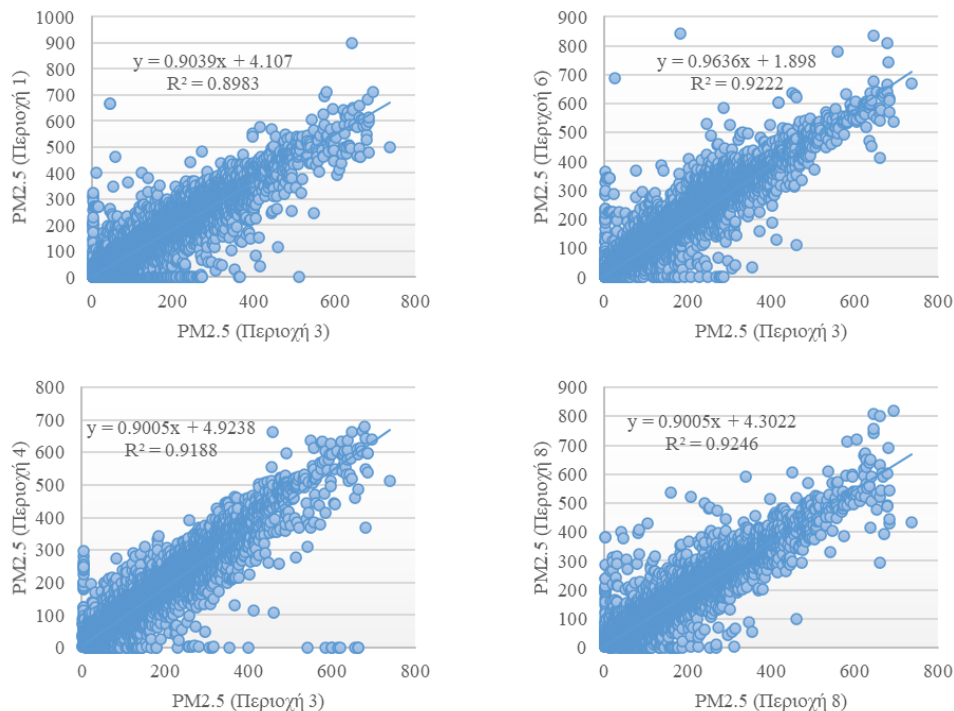
ΣΧΗΜΑ 9 : Γραφική απεικόνιση της αλληλεπίδρασης των επιμέρους περιοχών εντός ακτίνας 10 χιλιομέτρων μεταξύ τους. Τα γκρι νούμερα αντιστοιχούν στις περιοχές και τα πράσινα στους ακέραιους αριθμούς που αντιπροσωπεύουν την αντίστοιχη διεύθυνση αέρα.

Περιοχή	Επηρεαζόμενη Περιοχή	Διεύθυνση Ανέμου	Ακέραιος Αριθμός	Σειρά Εγγύτητας
1	3	S-SE	8	3>6>4>10
	6	E-SE/SE	6 / 7	
	4	S-SW	10	
	10	SW	11	
3	6	E-SE	6	9>1
	8	S-SW	10	
	4	SW	11	
	1	N-NW	16	
4	10	NW	15	6>8>4>1
	3	E-NE	4	
	8	SE	7	
	1	N-NE	2	
6	1	W-NW/NW	14 / 15	3>1>8
	3	W-SW	12	
	8	S-SW/SW	10 / 11	
8	3	N-NE	2	3>4>6>1
	4	NW/N-NW	15 / 16	
	6	NE	3	
9	10	SE/S-SE	7 / 8	4>9>1>3
10	4	E-SE/ SE	6 / 7	
	9	NW/N-NW	15 / 16	
	1	NE/E-NE	3 / 4	

ΠΙΝΑΚΑΣ 11 : Συσχέτιση επίδρασης περιοχών μελέτης ως προς τη φορά του ανέμου και αναγραφή εγγύτητας κατά αύξουσα σειρά.

4.2.3. Μέθοδος 3 : Χρήση γραμμικής παρεμβολής

Η ιδεολογία αυτής της μεθόδου σχετίζεται άμεσα με την βασική ιδέα πίσω από τη Μέθοδο 2. Όπως και πριν έτσι και εδώ γίνεται η βασική θεώρηση πως οι κλιματικές και ατμοσφαιρικές συνθήκες που επικρατούν σε κοντινές περιοχές περιγράφουν με σχετική ακρίβεια τις συνθήκες που επικρατούν στην περιοχή μελέτης. Έχοντας κάνει την παραπάνω παραδοχή, η συσχέτιση των περιοχών μεταξύ τους απλοποιείται σε μερικά μόνο βήματα. Ωστόσο προτού γίνει η ανάλυση των επιμέρους βημάτων, χρειάζεται να αναφερθεί ότι ο μοναδικός παράγοντας συσχέτισης ήταν η απόσταση και όχι άλλοι παράγοντες όπως η κατεύθυνση/ταχύτητα του ανέμου, το υψόμετρο, ή η βιομηχανική ανάπτυξη αυτής. Αρχικά κατηγοριοποιούνται οι περιοχές κατά σειρά εγγύτητας. Έπειτα γίνεται η ταξινόμηση των δεδομένων των επιμέρους περιοχών ως προς τα δεδομένα της περιοχής μελέτης, τα οποία τοποθετούνται κατά αύξουσα σειρά. Στόχος είναι η εξαγωγή γραμμικών σχέσεων. Στην παραπάνω προσπάθεια φαίνεται να δρα θετικά η συγκεκριμένη ταξινόμηση, καθώς τα δεδομένα δεν τοποθετούνται «άτακτα» εντός του γραφήματος. Η γραφική τους αναπαράσταση επιτρέπει την εξαγωγή γραφημάτων διασποράς κατάλληλα δομημένων ώστε να είναι φανερό η γραμμική συσχέτιση μεταξύ των τιμών. Αξιοποιώντας την δυνατότητα χάραξης γραμμής τάσης επιτυγχάνεται η εξαγωγή των γραμμικών εξισώσεων και η ύστερη αξιοποίηση τους στα κενά στοιχεία της κάθε περιοχής. Οι εξισώσεις αυτές χρησιμοποιούνται με σειρά προτεραιότητας η οποία ορίζεται βάσει της απόστασης της περιοχής την οποία αντιπροσωπεύουν ως προς την περιοχή μελέτης. Συγκεκριμένα έστω ότι θεωρείται ως παράδειγμα η περιοχή 1, τότε οι περιοχές κοντά σε αυτή είναι οι 3, 4, 6, 10 ή αν αναφερθούν κατά σειρά εγγύτητας 3-6-4-10. Σε περίπτωση ύπαρξης κενού στοιχείου σε ένα από τα δεδομένα της περιοχής 1, τότε αξιοποιείται η σχέση (1.1) για την κάλυψη του κενού.



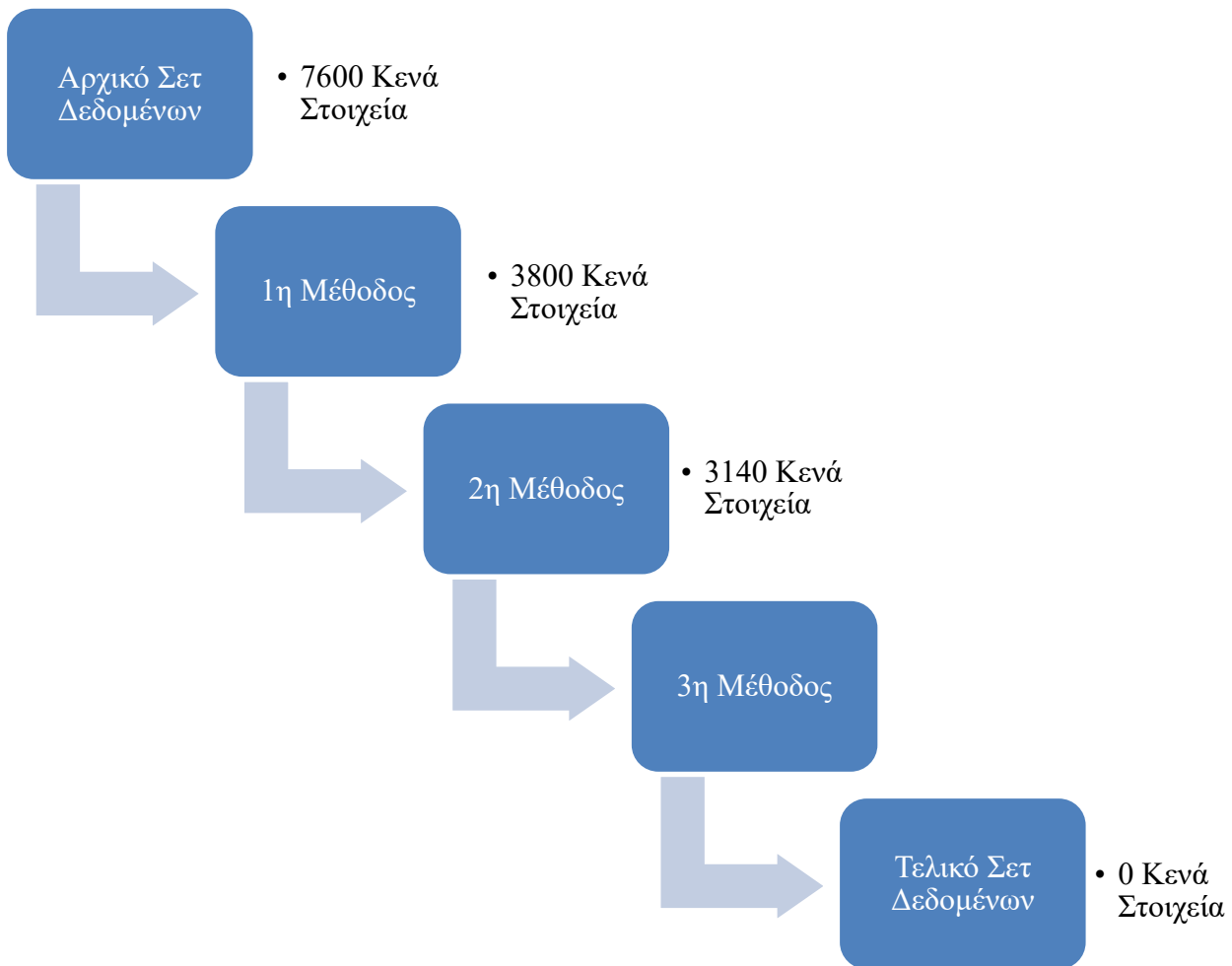
ΣΧΗΜΑ 10 : Ενδεικτική απεικόνιση των γραμμικών σχέσεων των συγκεντρώσεων των PM2.5 μεταξύ των 4 περιοχών κοντά στην περιοχή 3 και της περιοχής 3.

Ωστόσο υπάρχει το ενδεχόμενο να παρουσιάζεται ταυτόχρονα κενή τιμή και στην περιοχή 3, τότε γίνεται χρήση της σχέσης (1.3), η οποία αντιστοιχεί στην 2^η κοντινότερη περιοχή (περιοχή 6), ώστε να καλυφθεί το κενό. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να μην υπάρχουν άλλες διαθέσιμες σχέσεις. Η μέθοδος αυτή προσφέρει κάλυψη των εναπομεινάντων κενών στοιχείων σε ποσοστό 100%. Στο **Σχήμα 10** παρουσιάζονται ενδεικτικά τα γραφήματα για την συσχέτιση μεταξύ των τιμών στην περιοχή 3 σε σχέση με τις περιοχές 1-4-6-8.

Περιοχή	Γραμμική Εξίσωση		R ² [%]					
			PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃
1	$y = 1.0026 \cdot x + 3.2005$	για την περιοχή 3 (1.1)	91	85	84	71	72	73
	$y = 0.941 \cdot x + 4.6084$	για την περιοχή 4 (1.2)	91	85	83	75	81	89
	$y = 0.999 \cdot x + 2.1118$	για την περιοχή 6 (1.3)	90	85	87	77	85	97
	$y = 0.9648 \cdot x + 4.7272$	για την περιοχή 10 (1.4)	85	78	82	72	79	88
2	$y = 0.9585 \cdot x + 13.412$	για την περιοχή 1 (2.1)	72	65	63	41	57	76
	$y = 0.9856 \cdot x + 12.9410$	για την περιοχή 9 (2.2)	76	68	70	43	60	75
3	$y = 0.9039 \cdot x + 4.1070$	για την περιοχή 1 (3.1)	90	83	80	76	76	74
	$y = 0.9005 \cdot x + 4.9238$	για την περιοχή 4 (3.2)	92	87	85	81	80	78
	$y = 0.9636 \cdot x + 1.8988$	για την περιοχή 6 (3.3)	92	89	92	83	84	80
	$y = 0.9005 \cdot x + 4.3022$	για την περιοχή 8 (3.4)	93	86	69	75	83	76
4	$y = 0.9631 \cdot x + 2.0824$	για την περιοχή 1 (4.1)	91	83	80	72	81	89
	$y = 1.0229 \cdot x + 0.9658$	για την περιοχή 3 (4.2)	93	86	87	72	75	76
	$y = 0.9492 \cdot x + 3.1852$	για την περιοχή 8 (4.3)	91	85	66	74	86	89
	$y = 1.0002 \cdot x + 1.7004$	για την περιοχή 10 (4.4)	89	83	86	80	85	91
5	$y = 0.8929 \cdot x + 6.8053$	για την περιοχή 1 (5.1)	81	70	70	61	69	83
	$y = 0.7432 \cdot x + 8.0100$	για την περιοχή 2 (5.2)	72	61	71	46	62	76
6	$y = 0.8975 \cdot x + 5.6863$	για την περιοχή 1 (6.1)	88	81	81	73	81	90
	$y = 0.9562 \cdot x + 4.4986$	για την περιοχή 3 (6.2)	92	88	91	75	73	77
	$y = 0.894 \cdot x + 5.9297$	για την περιοχή 8 (6.3)	81	84	70	82	85	90
8	$y = 0.9509 \cdot x + 4.0874$	για την περιοχή 1 (8.1)	86	76	66	61	76	86
	$y = 1.0279 \cdot x + 1.5135$	για την περιοχή 3 (8.2)	92	85	71	65	72	75
	$y = 0.9692 \cdot x + 3.2804$	για την περιοχή 4 (8.3)	91	84	66	74	81	88
	$y = 1.03 \cdot x + 0.5538$	για την περιοχή 6 (8.4)	91	85	73	72	85	91
9	$y = 0.9414 \cdot x + 3.4712$	για την περιοχή 1 (9.1)	87	79	73	64	73	89
	$y = 0.7751 \cdot x + 6.1161$	για την περιοχή 2 (9.2)	74	66	69	45	61	74
	$y = 0.9573 \cdot x + 4.7721$	για την περιοχή 10 (9.3)	82	77	75	67	75	87
10	$y = 0.8785 \cdot x + 7.2847$	για την περιοχή 1 (10.1)	84	76	79	70	74	89
	$y = 0.9527 \cdot x + 4.7818$	για την περιοχή 3 (10.2)	90	83	87	70	69	76
	$y = 0.8693 \cdot x + 6.4018$	για την περιοχή 4 (10.3)	89	84	85	81	79	91
	$y = 0.8756 \cdot x + 8.9021$	για την περιοχή 10 (10.4)	84	78	78	68	69	88

ΠΙΝΑΚΑΣ 12 : Γραμμικές εξισώσεις συσχέτισης περιοχών και οι αντίστοιχοι δείκτες R².

Συνεπώς προκειμένου να εκπαιδευτούν τα ΤΝΔ, είναι αναγκαία η προ-επεξεργασία των δεδομένων εισαγωγής, ώστε να διορθωθούν τυχόν σφάλματα. Βασικό πρόβλημα αποτελούν οι κενές τιμές, οι οποίες μπορεί να οφείλονται σε αστοχίες των οργάνων, βλάβες, προγραμματισμένες συντηρήσεις ή/και διακοπές ρεύματος και άλλους ανθρωπογενείς παράγοντες. Τα κενά στοιχεία αν δεν λάβουν την απαραίτητη μεταχείριση δύνανται να προκαλέσουν σοβαρά προβλήματα στην εκπαίδευση, καθιστώντας κάθε προσπάθεια ανούσια. Η άτακτη τοποθέτηση τιμών ή ακόμα και η διαγραφή ολόκληρων σειρών δεδομένων μπορεί να οδηγήσει στην αλλοίωση αλλά και στην απώλεια μεγάλου όγκου δεδομένων. Οι τρεις (3) διακριτές μέθοδοι που αξιοποιήθηκαν καταφέρνουν να επιτύχουν ένα τελικό ποσοστό κάλυψης κενών στοιχείων 100%, βελτιώνοντας τη διαδικασία εκπαίδευσης των ΤΝΔ. Οι μέθοδοι εξελίσσονται σε περιπλοκότητα, με την πρώτη να καλύπτει το μεγαλύτερο μέρος των κενών στοιχείων (με ποσοστά κοντά στο 50%), η δεύτερη να καλύπτει ένα ποσοστό του 20% των εναπομεινασών κενών τιμών, ενώ το υπόλοιπο 80% καλύπτεται από την τρίτη.



4.3. Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων

4.3.1. Προετοιμασία δεδομένων

Προτού αρχίσει η διαδικασία εκπαίδευσης των ΤΝΔ είναι αναγκαία η κατάλληλη επεξεργασία των δεδομένων εισόδου. Αρχικά είναι αναγκαίος ο έλεγχος συσχέτισης των μεταβλητών μεταξύ τους, ώστε να αποκαλυφθούν οι όποιες σχέσεις υπάρχουν. Προκειμένου να επιτευχθεί αυτό γίνεται χρήση του συντελεστή συσχέτισης (Correlation Factor)[43] μεταξύ των μεταβλητών, για κάθε περιοχή. Ο συντελεστής συσχέτισης ενός σετ μεταβλητών εκφράζει την εξάρτηση μιας μεταβλητής ως προς ένα σύνολο παραγόντων. Πρακτικά αυτό σημαίνει ότι η πιθανότητα μεταβολής μιας μεταβλητής δεν είναι άμεσα εξαρτώμενη από μια μονάχα άλλη μεταβλητή. Η ανάλυση αυτή προσφέρει χρήσιμες πληροφορίες για την αλληλεξάρτηση των μεταβλητών τόσο τοπικά (εντός της περιοχής μελέτης) όσο και ολιστικά (εντός της πόλης του Πεκίνο). Παρότι η ολιστική συμπεριφορά των μεταβλητών παρουσιάζει ιδιαίτερο ενδιαφέρον, στο πλαίσιο της συγκεκριμένης διπλωματικής εργασίας θα ληφθεί υπόψιν στη δημιουργία των σεναρίων, μονάχα ο τοπικός χαρακτήρας. Ο μέσος όρος όλων των συντελεστών συσχέτισης παρουσιάζεται στον **Πίνακα 13**. Ωστόσο παρά τη προσπάθεια εξαγωγής «κρυφών» σχέσεων μέσω του συντελεστή συσχέτισης, ο **Πίνακας 13** δεν προσφέρει ιδιαίτερη βοήθεια. Ως εναλλακτική μέθοδος προσδιορισμού βέλτιστων συνδυασμών πραγματοποιείται η Ανάλυση των Κύριων Στοιχείων (Principal Components Analysis, PCA), σε συνδυασμό με την μέθοδο k-means clustering. Η PCA αποτελεί μια από τις παλαιότερες μεθόδους γραφικής αναπαράστασης μεγάλων σετ δεδομένων, καθώς μειώνει την «διαστατικότητα» των τιμών αυτών διατηρώντας παράλληλα τη «μεταβλητότητα» αυτών [44]. Αυτό επιτυγχάνεται συνήθως μέσω δημιουργίας νέων μεταβλητών γραμμικά εξαρτώμενων από τις μεταβλητές του συστήματος, αλλά χωρίς να παρουσιάζουν κάποια εξάρτηση μεταξύ τους. Οι δύο παραπάνω μέθοδοι λειτουργούν συνεργατικά με την πρώτη να επιτρέπει την γραφική αναπαράσταση της συσχέτισης παραπάνω των τριών (3) μεταβλητών, κάτι το οποίο δεν είναι συμβατό με άλλα οπτικά μέσα καθώς δεν υπάρχει δυνατότητα δημιουργίας 4^{ου} άξονα τιμών. Η δεύτερη μέθοδος αξιοποιείται πριν την PCA, προκειμένου να διαχωρίσει το σετ δεδομένων, το οποίο αποτελείται από 16 στήλες των 35024 σειρών, σε ξεχωριστά «πακέτα» (clusters) δεδομένων. Τα δεδομένα μπορούν να χωριστούν σε 3 πακέτα (Ημερομηνίες, Ρύποι, Μετεωρολογικά Δεδομένα), σε 4 πακέτα (Ημερομηνίες, Ρύποι, Συγκέντρωση CO, Μετεωρολογικά Δεδομένα), σε 5 με τον περαιτέρω διαχωρισμό της ατμοσφαιρικής πίεσης (Pressure), σε 6 με το διαχωρισμό και της διεύθυνσης του ανέμου (Wind Direction) και τέλος σε 13 διαφορετικά πακέτα, ένα για κάθε μεταβλητή. Το τελευταίο πακέτο, πρακτικά αναιρεί πλήρως τη χρησιμότητα της PCA, καθώς το μόνο μέγεθος το οποίο εντάσσεται σε cluster είναι οι χρονικές μεταβλητές. Το κριτήριο βάσει του οποίου χωρίζονται τα πακέτα δεδομένων είναι συνήθως είτε το είδος είτε η τάξη των δεδομένων. Στα παραπάνω πακέτα για παράδειγμα, η συγκέντρωση του CO ανήκει στους ρύπους, ωστόσο καταγράφεται σε διαφορετική κλίμακα δεδομένων, ομοίως με την πίεση η οποία είναι δυο τάξεις μεγαλύτερη από τις άλλες μετεωρολογικές τιμές. Τέλος η κατεύθυνση του ανέμου αποτελεί ένα σύνολο ακέραιων μεταβλητών, φραγμένων μεταξύ του μηδέν (0) και του δεκαέξι (16). Η PCA φανέρωσε πως το 40% της πληροφορίας μπορεί να περιγραφεί μονάχα με το σύνολο των θερμοκρασιών. Πρακτικά

ωστόσο κάτι τέτοιο δεν είναι βάσιμο, καθώς υπάρχει πληθώρα μεταβλητών όπου παρότι εξαρτώνται από την θερμοκρασία, αποτελούν παράλληλα και παράγοντα αυξομείωσης συγκεντρώσεων άλλων μεταβλητών. Χαρακτηριστικό αποτελεί η περίπτωση του Όζοντος και του διοξειδίου του αζώτου, όπου παρότι εξαρτώνται από την θερμοκρασία, η συγκέντρωση κάθε ενός από αυτά επηρεάζει, άλλοτε θετικά, άλλοτε αρνητικά τη συγκέντρωση του άλλου. Παρότι βάσει μεθοδολογίας θα έπρεπε να αρκεστούμε μονάχα στα αποτελέσματα της ανάλυσης PCA, στην παρούσα Διπλωματική εργασία αναλύονται παραπάνω σενάρια, τα οποία αποδείχτηκαν καλύτερα του προτεινόμενου. Άλλωστε ερευνητές όπως οι Fabiana Franceschi et al. [39] εντόπισαν σε παρόμοιες έρευνες, θετική συσχέτιση μεταξύ της συγκέντρωσης PM₁₀ και της κατεύθυνσης του ανέμου, ενώ παρατηρήθηκε αρνητική συσχέτιση μεταξύ της συγκέντρωσης του ίδιου ρύπου ως προς την θερμοκρασία και την ταχύτητα του αέρα. Τα ευρήματα αυτά ενισχύουν περαιτέρω οι Zhang et al. [45] σε παλαιότερη έρευνα, στη πόλη Beijing, China. Αντίστοιχα για τα PM_{2.5} είχε παρατηρηθεί θετική συσχέτιση μεταξύ αυτών και της σχετικής υγρασίας [39], κάτι το οποίο επαληθεύει και η συγκεκριμένη ερευνητική εργασία.

	PM _{2.5}	PM ₁₀	SO ₂	NO ₂	CO	O ₃	TEMP	PRES	DEWP	RAIN	wd	WSPM	RH
PM _{2.5}	1.00	0.88	0.48	0.68	0.79	-0.16	-0.14	0.01	0.11	-0.01	-0.12	-0.27	0.39
PM ₁₀		1.00	0.46	0.66	0.70	-0.13	-0.11	-0.02	0.06	-0.02	-0.07	-0.18	0.26
SO ₂			1.00	0.49	0.54	-0.17	-0.34	0.22	-0.28	-0.03	-0.05	-0.10	-0.08
NO ₂				1.00	0.71	-0.50	-0.29	0.14	-0.03	-0.03	-0.15	-0.42	0.33
CO					1.00	-0.32	-0.34	0.18	-0.07	-0.01	-0.14	-0.29	0.34
O ₃						1.00	0.60	-0.45	0.32	0.02	0.14	0.29	-0.27
TEMP							1.00	-0.83	0.82	0.06	0.02	0.03	0.10
PRES								1.00	-0.77	-0.01	0.00	0.08	-0.24
DEWP									1.00	0.10	-0.11	-0.28	0.63
RAIN										1.00	-0.01	0.12	0.10
Wd											1.00	0.24	-0.22
WSPM												1.00	-0.52
RH													1.00

ΠΙΝΑΚΑΣ 13: Πίνακας μέσων συντελεστών συσχέτισης.

4.3.2. Δημιουργία σεναρίων

Τα σενάρια τα οποία δημιουργήθηκαν ήταν τα ίδια για κάθε περιοχή (1 έως 6 και 8 έως 10), ωστόσο η κάθε μια από αυτές μελετήθηκε ξεχωριστά. Στόχος είναι η εκπαίδευση 6 «Βέλτιστων» ΤΝΔ για κάθε περιοχή και η συγχώνευση τους σε έναν καθολικό αλγόριθμο. Κάθε περιοχή περιλαμβάνει 6 ρύπους (PM_{2.5}, PM₁₀, SO₂, NO₂, CO και O₃), για κάθε ρύπο σχεδιάστηκαν 8 σενάρια με κάθε σενάριο να περιλαμβάνει 10 ΤΝΔ, έκαστο με διαφορετικές μεταβλητές εκπαίδευσης. Το σύνολο των ΤΝΔ που δημιουργούνται συνολικά ανέρχεται στα 4320.

Τα δεδομένα εκπαίδευσης περιλαμβάνουν διαφορετικό ανά σενάριο συνδυασμό μετεωρολογικών μεταβλητών και προβλέπουν τις 24 επόμενες ωριαίες τιμές. Για την καλύτερη μελέτη και ανάλυση των δεδομένων κρίθηκε σκόπιμη η προσθήκη μίας ακόμη μεταβλητής, της σχετικής

υγρασίας (RH), καθώς συνδυάζει τα δεδομένα της θερμοκρασίας δρόσου (DEWP) και της θερμοκρασίας ξηρού βολβού (TEMP). Ο υπολογισμός αυτής έγινε με την χρήση της εξίσωσης 2.1 [46]:

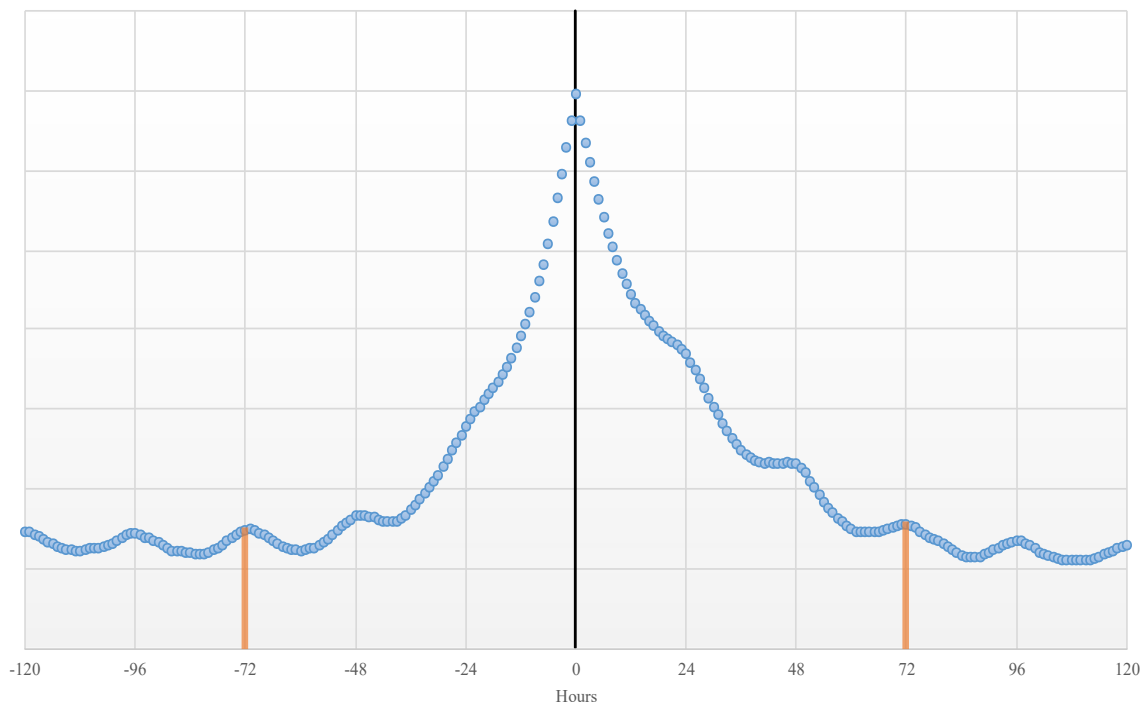
$$RH = \frac{e^{\frac{17.625 \cdot D_p}{243.04 + D_p}}}{e^{\frac{17.625 \cdot T}{243.04 + T}}} \quad (4.1)$$

Τα σενάρια τα οποία σχεδιάστηκαν ακολούθησαν συγκεκριμένες αρχές και μεθοδολογίες, ωστόσο πρέπει να αναφερθεί ότι ένας από τους παράγοντες ήταν η κρίση του συγγραφέα. Παρότι κάτι τέτοιο μπορεί να θεωρηθεί μη δόκιμο, δυστυχώς αποτέλεσε αναγκαία συνθήκη, καθώς λόγω των ασαφών δεδομένων από τη PCA, θα χρειαζόταν να δημιουργηθούν σενάρια με κάθε πιθανό συνδυασμό μετεωρολογικών παραγόντων. Φυσικά το πλήθος των σεναρίων θα καθιστούσε την έγκαιρη παράδοση της εργασίας αδύνατη. Συνεπώς βάσει ερευνών όπως αναφέρονται στην παραπάνω ενότητα (Fabiana Franceschi et al. [39] και Zhang et al. [45]), καθώς και υποθέσεων για την πιθανή ορθή συσχέτιση μεταξύ μετεωρολογικών συνθηκών και ρύπων, σχεδιάστηκαν 8 διακριτά σενάρια, τα οποία εφαρμόστηκαν σε κάθε ρύπο. Τα σενάρια, οι μεταβλητές καθώς και το πλήθος των δεδομένων εισόδου εκάστου αναγράφονται στον [Πίνακα 14](#).

<i>A/A</i>	Μεταβλητές	Πλήθος Δεδομένων
1	Temperature - Pressure - Dew Point - Rain - Wind Direction - Wind Speed	350640
2	Temperature - Pressure - Dew Point - Rain - Wind Direction - Wind Speed - Relative Humidity	685704
3	Temperature - Pressure - Wind Direction - Wind Speed - Relative Humidity	315576
4	Temperature - Pressure - Wind Direction - Wind Speed	280512
5	Temperature - Wind Direction - Wind Speed	245448
6	Wind Direction - Wind Speed	210384
7	Temperature - Pressure	
8	Temperature	175320

ΠΙΝΑΚΑΣ 14: Πίνακας σεναρίων και πλήθους δεδομένων για κάθε ένα από αυτά.

Προκειμένου να πραγματοποιηθεί η εκπαίδευση των ΤΝΔ χρειάζεται πρώτα να γίνει η επιλογή των δεδομένων εισόδου. Παραπάνω αναφέρθηκαν τα σενάρια βάσει των οποίων έγινε η εκπαίδευση, ωστόσο αναγκαίος είναι και ο προσδιορισμός του πλήθους αυτών. Όπως αναφέρθηκε και στο θεωρητικό υπόβαθρο, προκειμένου να εκπαιδευτεί ένα ΤΝΔ πρέπει να δοθούν δεδομένα εισόδου και εξόδου. Τα δεδομένα εισόδου γίνεται να δοθούν είτε με ημερήσιο είτε με ωριαίο βήμα. Στα πλαίσια της συγκεκριμένης εργασίας επιλέχθηκε να δοθούν με ημερήσιο βήμα, οπότε τα δεδομένα εισόδου αποτελούνται από $\frac{16 \cdot \nu + 10}{17 \cdot \nu + 10}$ στήλες, όπου ν είναι το πλήθος των ημερών που προηγούνται της χρήσης του νευρωνικού και τα νούμερα 10 και 11 αντιστοιχούν στο άθροισμα των στηλών που περιέχουν τις χρονικές και μετεωρολογικές μεταβλητές (χωρίς και με την σχετική υγρασία). Το ν έγινε απόπειρα να προσδιοριστεί γραφικά, μέσω της μεθόδου SEA (Superposed Epoch Analysis), και όπως φαίνεται και στα παρακάτω διαγράμματα η πρώτη κοιλιάδα παρουσιάζεται στην χρονική στιγμή $t = \pm 60$. Ωστόσο ο προσδιορισμός της τιμής του ν είναι αναγκαίο να γίνει σε ημέρες, συνεπώς ακέραια πολλαπλάσια του 24, προκειμένου να ισχύει ο αρχικός σκοπός εκπόνησης της εργασίας. Το πλήθος των ημερών επιλέχθηκε να ισούται με 3 ημέρες (72 ώρες), καθώς το συγκεκριμένο χρονικό πλαίσιο συνδυάζει την ύπαρξη ικανού πλήθους τιμών εκπαίδευσης, χωρίς όμως να αυξάνει ιδιαίτερα τις υπολογιστικές απαιτήσεις της εκπαίδευσης. Ωστόσο η παραπάνω θεώρηση αποτελεί παραδοχή, καθώς στη βιβλιογραφία παρατηρούνται ποικίλες επιλογές προηγούμενων ημερών ή ακόμα και ωρών.



ΣΧΗΜΑ 11 : Γραφική αναπαράσταση μέσω ωριαίων τιμών συγκέντρωσης για κάθε τα PM10 της περιοχής 1 σε χρονικό διάστημα 5 ημερών πριν και μετά της ημερομηνίας αναφοράς.

Όπως αναφέρθηκε για κάθε σενάριο εκπαιδεύτηκαν 10 διαφορετικά ΤΝΔ τα οποία αξιολογήθηκαν σε δεύτερο χρόνο και επιλέχθηκε το καλύτερο για κάθε ρύπο και για κάθε περιοχή μελέτης. Για λόγους επαναληψιμότητας του πειράματος παρατίθενται οι μεταβλητές εκπαίδευσης των νευρωνικών δικτύων στον **Πίνακα 15**. Οι αλγόριθμοι εκπαίδευσης επιλέχθηκαν ώστε να μπορέσει να γίνει και μια δευτερεύουσα αξιολόγηση του τρόπου με τον οποίο δύναται να βελτιστοποιηθεί η διαδικασία εκπαίδευσης προγνωστικών μοντέλων. Συγκεκριμένα ως πρώτος αλγόριθμος επιλέχθηκε η προεπιλογή του λογισμικού (Levenburg Marquardt) για την δημιουργία αλγόριθμων προσαρμογής, ενώ οι επιμέρους μεταβλητές του προσαρμόστηκαν ώστε να απαιτούν «μεσαία» υπολογιστική ισχύ. Για τα επόμενα τέσσερα (4) ΤΝΔ επιλέχθηκε να εκπαιδευτούν με τον αλγόριθμο Bayesian Regularization, καθώς ενδείκνυται για εκπαίδευση ΤΝΔ με στόχο την προσέγγιση μη γραμμικών εξισώσεων. Τα νευρωνικά 6 έως 10 αποτελούν παραδείγματα άλλων αλγόριθμων, οι οποίοι επιλέχθηκαν σε μια προσπάθεια εύρεσης πιθανών συναρτήσεων εκπαίδευσης με καλύτερη επίδοση. Ανάμεσα στους δέκα (10) αλγόριθμους οι πρώτοι 5 αποδείχθηκαν ως οι πιο «απαιτητικοί», καθώς χρειάζονται αρκετή υπολογιστική ισχύ. Ωστόσο όπως θα παρουσιαστεί στην επόμενη ενότητα προσέφεραν σταθερά τα εγκυρότερα αποτελέσματα. Αντίθετα το δεύτερο μισό του Πίνακα 5 παρότι ήταν λιγότερο «απαιτητικό» παρείχε πολύ ασταθή αποτελέσματα με μηδενική ικανότητα αξιοποίησης. Ωστόσο αξίζει να αναφερθεί ότι κανένας από τους πέντε (5) τελευταίους αλγόριθμους δεν ενδείκνυται για την εκπαίδευση νευρωνικών δικτύων τέτοιου είδους.

A/A	Αλγόριθμος	Κρυφά Στρώματα	Νευρώνες (Input Layer)	Νευρώνες (Hidden Layer)	Εποχές	Train	Test	Validation
1	LM	2	10	30-15	200	70%	15%	15%
2	BR ₁	2	10	30-15				
3	BR ₂	2	25	30-15				
4	BR ₃	3	10	30-15-10				
5	BR ₄	3	25	30-15-10				
6	CGB	3	30	30-15-10				
7	CGF	3	30	30-15-10				
8	CGP	3	30	30-15-10				
9	OSS	3	30	30-15-10				
10	SCG	3	30	30-15-10				

ΠΙΝΑΚΑΣ 15: Πίνακας μεταβλητών εκπαίδευσης κάθε ΤΝΔ.

4.3.3. Λογισμικά και Υποδομές

Για την εκπόνηση της συγκεκριμένης εργασίας αξιοποιήθηκαν τα παρακάτω λογισμικά :

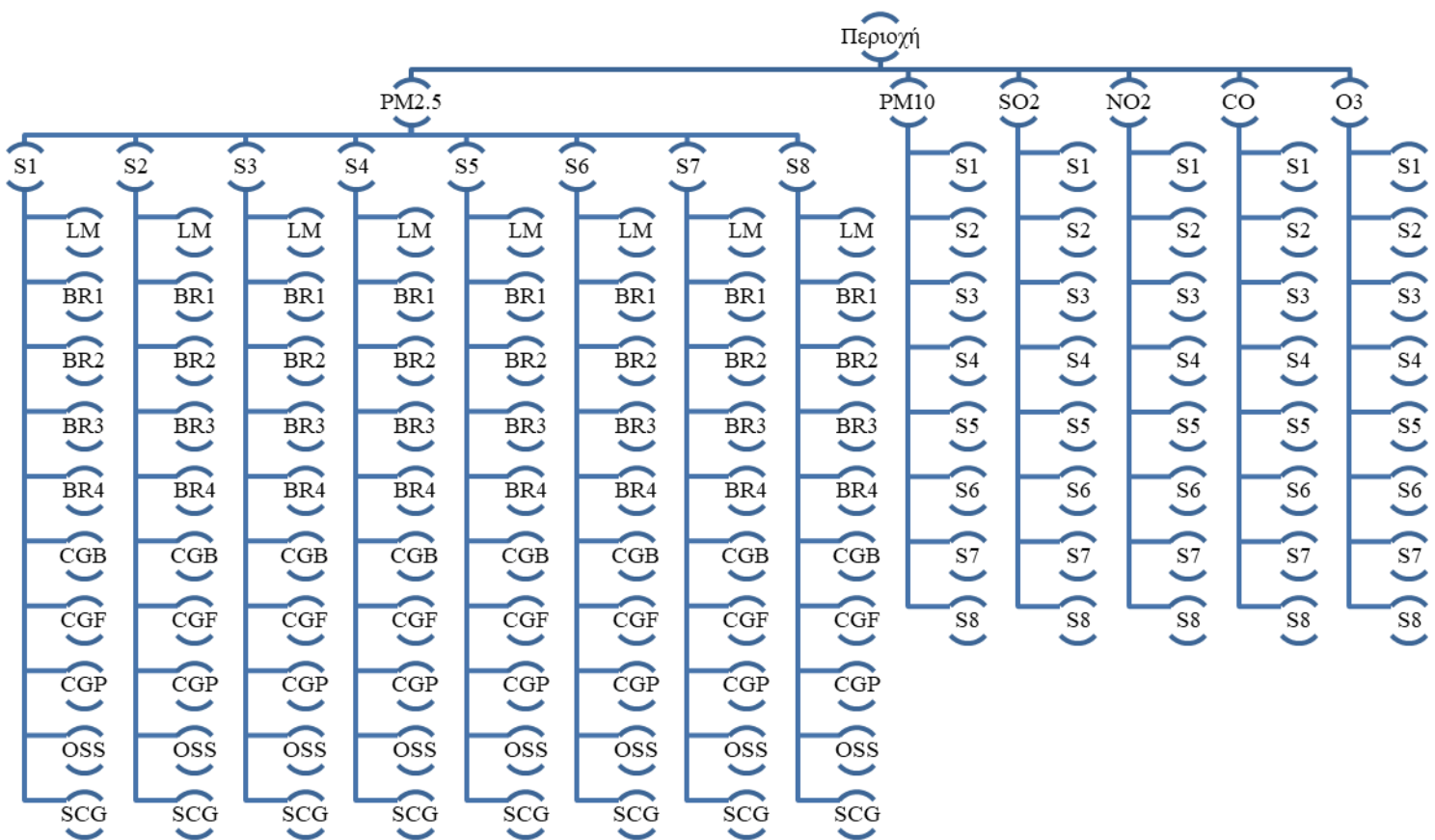
- MATLAB R2022a
- Microsoft Excel 2021

Η εκπαίδευση των ΤΝΔ πραγματοποιήθηκε σε οικιακό σταθερό υπολογιστή με τις εξής προδιαγραφές:

- CPU : AMD Ryzen 7 5700G
- RAM : G.Skill RipjawsV 16GB DDR4-3200MHz
- GPU : N/A
- Αποθηκευτικό Μέσο : SSD Kingston NV1 500GB M.2 NVMe (SNVS/500G).

5. ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Στο παραπάνω κεφάλαιο αναφέρθηκε ότι προκειμένου να βρεθεί ο βέλτιστος συνδυασμός σεναρίων δεδομένων εισόδου και Α.Ε., χρειάστηκε να δημιουργηθούν 4320 ΤΝΔ. Από τα 4320 αυτά ΤΝΔ επιλέχθηκαν συνολικά 54, τα οποία παρουσίασαν τους καλύτερους δείκτες αξιοπιστίας. Παρακάτω θα γίνει αναλυτική παρουσίαση των δεικτών και του τρόπου υπολογισμού τους, καθώς και αναφορά των αποδόσεων των ΤΝΔ για κάθε περιοχή και κάθε ρύπο. Στο παρακάτω σχήμα παρουσιάζονται γραφικά όλα τα ΤΝΔ που σχεδιάστηκαν για μία μονάχα περιοχή. Για λόγους αναγνωσιμότητας παρουσιάζονται όλες οι συναρτήσεις εκπαίδευσης ανά σενάριο μονάχα για τα αιωρούμενα σωματίδια αεροδυναμικής διαμέτρου έως και 2.5μm (PM_{2.5}), ωστόσο η εικόνα είναι ακριβώς ίδια για κάθε ρύπο.



Προκειμένου να μπορέσει να γίνει η αξιολόγηση των παραπάνω νευρωνικών δικτύων, κρίθηκε αναγκαίος ο υπολογισμός οκτώ (8) στατιστικών δεικτών. Οι δείκτες αυτοί καθώς και οι εξισώσεις υπολογισμού αυτών αναγράφονται παρακάτω [47] :

Γενικοί Στατιστικοί Δείκτες Αξιοπιστίας

Στατιστικοί Δείκτες Αληθούς / Ψευδούς Αποτελέσματος

1) Mean Absolute Error (MAE) $MAE = \frac{1}{n} \cdot \sum_{i=1}^n (P_i - O_i) $	(5.1)	5) True Prediction Rate (TPR) $TPR = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$	(5.5)
2) Root Mean Square Error (RMSE) $RMSE = \sqrt{\sum_{i=1}^n \frac{(P_i - O_i)^2}{n}}$	(5.2)	6) False Prediction Rate (FPR) $FPR = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$	(5.6)
3) Pearson's Correlation Coefficient (R) $R = \frac{1}{n} \cdot \frac{\sum_{i=1}^n [(P_i - P_{\text{mean}}) \cdot (O_i - O_{\text{mean}})]}{\sqrt{\sum_{i=1}^n [(O_i - O_{\text{ave}})^2]} \cdot \sqrt{\sum_{i=1}^n [(P_i - P_{\text{ave}})^2]}}$	(5.3)	7) False Alarm Rate (FAR) $FAR = \frac{\text{False Positives}}{\text{False Positives} + \text{True Positives}}$	(5.7)
4) Index of Agreement (IA) $IA = \frac{\sum_{i=1}^n [(P_i - O_i)^2]}{\sum_{i=1}^n [(P_i - O_{\text{mean}} + O_i - O_{\text{mean}})^2]}$	(5.4)	8) Success Index (SI) $SI = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$	(5.8)

Οι παραπάνω μεταβλητές επεξηγούνται αναλυτικά παρακάτω

P_i	Η προβλεπόμενη τιμή i	True Positives	Οι πραγματικές υπερβάσεις που προβλέπει σωστά το προγνωστικό μοντέλο
P_{mean}	Ο μέσος όρος των προβλεπόμενων τιμών	True Negatives	Οι μη υπερβάσεις που παρατηρήθηκαν και προβλέφθηκαν σωστά από το μοντέλο
O_i	Η παρατηρούμενη τιμή i	False Positives	Οι υπερβάσεις που προβλέφθηκαν αλλά δεν παρατηρήθηκαν
O_{mean}	Ο μέσος όρος των παρατηρούμενων τιμών	False Negatives	Οι υπερβάσεις που παρατηρήθηκαν αλλά δεν προβλέφθηκαν από το μοντέλο
n	Το συνολικό πλήθος των τιμών		

Ο υπολογισμός των παραπάνω δεικτών επιτρέπει την αξιολόγηση κάθε ΤΝΔ. Εφόσον σχεδιάστηκαν 4320 ΤΝΔ, χρειάστηκε να υπολογιστούν συνολικά 34560 τιμές. Η αναφορά όλων αυτών των τιμών κρίθηκε μη παραγωγική, καθώς θα καταλάμβανε υπερβολικό χώρο. Ωστόσο κρίθηκε σκόπιμο να γίνει αναφορά κάποιων χαρακτηριστικών τιμών όπως φαίνονται παρακάτω στον [Πίνακα 16a](#) έως [16l](#). Συγκεκριμένα επιλέχθηκε να αναφερθούν οι βέλτιστες τιμές κάθε σεναρίου για κάθε περιοχή ανά ρύπο, επιτρέποντας την ταυτόχρονη σύγκριση της ελεγχιμότητας των περιοχών, του βέλτιστου σεναρίου, του βέλτιστου Α.Ε. ανά ρύπο. Συγκεκριμένα:

Περιοχή	Σενάριο	A.E.	MAE($\mu\text{g}/\text{m}^3$)	RMSE($\mu\text{g}/\text{m}^3$)	R	IA (%)
1	All Meteorological & RH	BR ₃	0.16	29.17	0.936	96.64
2	All Meteorological	BR ₃	0.21	26.03	0.935	96.58
3	All Meteorological	BR ₃	0.29	28.77	0.943	97.03
4	All Meteorological & RH	BR ₄	0.28	26.70	0.945	97.13
5	All Meteorological	BR ₃	0.15	27.65	0.943	97.03
6	All Meteorological	BR ₄	0.05	29.31	0.941	96.95
8	All Meteorological & RH	BR ₄	0.11	27.47	0.941	96.91
9	All Meteorological & RH	BR ₄	0.31	28.44	0.938	96.80
10	All Meteorological	BR ₃	0.06	30.41	0.936	96.65

ΠΙΝΑΚΑΣ 16a: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα PM_{2.5}.

Περιοχή	Σενάριο	A.E.	TPR (%)	FPR (%)	FAR (%)	SI (%)
1	All Meteorological	BR ₄	97.35	78.26	1.27	96.16
2	All Meteorological	BR ₁	96.93	42.78	6.11	91.83
3	All Meteorological & RH	BR ₂	97.89	48.84	4.63	93.75
4	All Meteorological	BR ₁	98.43	49.58	4.29	94.51
5	All Meteorological	BR ₁	98.04	50.78	4.75	93.75
6	All Meteorological	BR ₁	98.18	40.29	4.15	94.51
8	All Meteorological	BR ₂	97.65	49.64	5.09	93.14
9	All Meteorological & RH	BR ₂	98.41	43.61	4.26	94.57
10	All Meteorological	BR ₁	98.20	48.82	4.53	94.10

ΠΙΝΑΚΑΣ 16b: Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα PM_{2.5}.

Στους πίνακες 16a & 16b παρουσιάζεται ως βέλτιστη λύση το σενάριο 1 (σε ποσοστό 67%) σε συνδυασμό με τον A.E. BR₁ & BR₃. Ως περιοχή με βέλτιστη προβλεπτική ικανότητα παρουσιάζεται η περιοχή 1 και 4, η πρώτη παρουσιάζοντας την υψηλότερη ικανότητα πρόγνωσης των υπερβάσεων και η δεύτερη παρουσιάζοντας την μέγιστη ικανότητα προσαρμογής των δεδομένων. Τα μέσα απόλυτα σφάλματα παρουσιάζουν τιμές μικρότερες του 0.5 $\mu\text{g}/\text{m}^3$, με την ελάχιστη τιμή συγκέντρωσης PM_{2.5} να ανέρχεται στα 1.0 $\mu\text{g}/\text{m}^3$. Οι στατιστικοί δείκτες Pearson βρίσκονται άνω του 0.90 παρουσιάζοντας μια εξαιρετική συσχέτιση των δεδομένων μεταξύ τους, συμπεράσμα το οποίο επίσης επιβεβαιώνεται και από τον δείκτη IA, με αντίστοιχες τιμές άνω του 95%. Όσον αφορά τους δείκτες αληθούς/ψευδούς αποτελέσματος, παρατηρούνται τιμές άνω του 95%, συνεπώς είναι ασφαλές να ειπωθεί ότι είναι δυνατή η σχεδόν πλήρης πρόβλεψη των υπερβάσεων. Ως οριακή τιμή της συγκέντρωσης των PM_{2.5} θεωρήθηκαν, βάσει του Παγκόσμιου Οργανισμού Υγείας, τα 15 $\mu\text{g}/\text{m}^3$ ημερησίως.

Περιοχή	Σενάριο	A.E.	MAE($\mu\text{g}/\text{m}^3$)	RMSE($\mu\text{g}/\text{m}^3$)	R	IA (%)
1	All Meteorological	BR ₃	<u>0.03</u>	42.57	0.897	94.47
2	All Meteorological & RH	BR ₄	0.11	<u>36.78</u>	0.900	94.65
3	All Meteorological	BR ₄	0.48	40.61	<u>0.911</u>	<u>95.31</u>
4	All Meteorological & RH	BR ₄	0.01	39.29	0.905	94.95
5	All Meteorological & RH	BR ₄	0.01	44.05	0.893	94.20
6	All Meteorological & RH	BR ₁	0.24	40.91	0.905	94.87
8	All Meteorological	BR ₄	0.15	38.65	0.904	94.84
9	All Meteorological	BR ₄	0.14	39.40	0.907	95.03
10	All Meteorological	BR ₃	0.17	44.20	0.894	94.37

ΠΙΝΑΚΑΣ 16c: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα PM₁₀.

Περιοχή	Σενάριο	A.E.	TPR (%)	FPR (%)	FAR (%)	SI (%)
1	All Meteorological & RH	BR ₃	97.00	30.25	8.19	90.94
2	All Meteorological & RH	BR ₂	95.91	24.94	9.11	90.11
3	All Meteorological & RH	BR ₄	97.16	32.02	8.83	90.53
4	All Meteorological	BR ₂	96.61	31.27	7.95	90.73
5	All Meteorological & RH	BR ₂	96.91	33.63	5.99	92.17
6	All Meteorological	BR ₂	97.08	24.85	6.87	92.18
8	All Meteorological & RH	BR ₁	97.28	29.02	7.66	91.56
9	All Meteorological	BR ₄	<u>97.75</u>	<u>24.67</u>	<u>6.15</u>	<u>93.13</u>
10	All Meteorological	BR ₁	97.25	32.31	7.75	91.28

ΠΙΝΑΚΑΣ 16d: Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα PM₁₀.

Στους πίνακες 16c & 16d παρουσιάζεται ως βέλτιστη λύση το σενάριο 1 και 2 (σε ποσοστά 50%) σε συνδυασμό με τον A.E. BR₂ & BR₄. Ως περιοχή με βέλτιστη προβλεπτική ικανότητα παρουσιάζεται η περιοχή 3 και 9, η πρώτη παρουσιάζοντας την μέγιστη ικανότητα προσαρμογής των δεδομένων και η δεύτερη παρουσιάζοντας την υψηλότερη ικανότητα πρόγνωσης των υπερβάσεων. Τα μέσα απόλυτα σφάλματα παρουσιάζουν τιμές μικρότερες του 0.5 $\mu\text{g}/\text{m}^3$, με την ελάχιστη τιμή συγκέντρωσης PM₁₀ να ανέρχεται στα 2.0 $\mu\text{g}/\text{m}^3$. Οι στατιστικοί δείκτες Pearson βρίσκονται, στη συντριπτική πλειοψηφία τους, άνω του 0.90 παρουσιάζοντας μια ικανοποιητική συσχέτιση των δεδομένων μεταξύ τους, συμπεράσμα το οποίο επίσης επιβεβαιώνεται και από τον δείκτη IA, με αντίστοιχες τιμές κοντά στο 95%. Όσον αφορά τους δείκτες αληθούς/ψευδούς αποτελέσματος, παρατηρούνται τιμές άνω του 90%, συνεπώς μπορεί να ειπωθεί ότι είναι δυνατή η σχεδόν πλήρης πρόβλεψη των υπερβάσεων. Ως οριακή τιμή της συγκέντρωσης των PM₁₀ θεωρήθηκαν, βάσει του Παγκόσμιου Οργανισμού Υγείας, τα 45 $\mu\text{g}/\text{m}^3$ ημερησίως.

Περιοχή	Σενάριο	A.E.	MAE($\mu\text{g}/\text{m}^3$)	RMSE($\mu\text{g}/\text{m}^3$)	R	IA (%)
1	All Meteorological	BR ₃	0.08	8.51	0.929	96.32
2	All Meteorological & RH	BR ₄	0.04	7.50	0.934	96.59
3	All Meteorological	BR ₃	0.06	8.39	0.931	96.40
4	All Meteorological	BR ₄	0.05	8.72	0.930	96.36
5	All Meteorological & RH	BR ₃	0.26	8.21	0.928	96.17
6	All Meteorological	BR ₃	0.00	8.57	0.936	96.69
8	All Meteorological & RH	BR ₄	0.12	18.26	0.948	97.28
9	All Meteorological & RH	BR ₃	0.01	7.13	0.950	97.40
10	All Meteorological	BR ₄	0.01	8.41	0.937	96.72

ΠΙΝΑΚΑΣ 16e: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα SO₂.

Περιοχή	Σενάριο	A.E.	TPR (%)	FPR (%)	FAR (%)	SI (%)
1	All Meteorological & RH	BR ₄	90.51	0.54	4.67	98.49
2	All Meteorological & RH	BR ₄	94.31	0.22	2.52	99.31
3	All Meteorological	BR ₃	91.52	0.23	1.95	98.83
4	All Meteorological & RH	BR ₄	87.58	0.46	4.08	98.22
5	Only TEMP & WD & WSPM	BR ₄	84.92	0.15	1.83	98.56
6	All Meteorological & RH	BR ₂	90.34	0.23	1.85	98.63
8	All Meteorological & RH	BR ₃	88.16	1.02	7.46	97.62
9	All Meteorological & RH	BR ₄	96.91	0.23	1.88	99.45
10	Only TEMP & PRES	BR ₄	90.97	0.31	2.76	98.76

ΠΙΝΑΚΑΣ 16f: Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα SO₂.

Στους πίνακες 16e & 16f παρουσιάζεται ως βέλτιστη λύση το σενάριο 2 (σε ποσοστό 56%) σε συνδυασμό με τον A.E. BR₃ & BR₄. Ως περιοχή με βέλτιστη προβλεπτική ικανότητα παρουσιάζεται η περιοχή 9, παρουσιάζοντας την υψηλότερη ικανότητα πρόγνωσης των υπερβάσεων καθώς και τη μέγιστη ικανότητα προσαρμογής των δεδομένων. Τα μέσα απόλυτα σφάλματα παρουσιάζουν τιμές μικρότερες του 0.3 $\mu\text{g}/\text{m}^3$, με την ελάχιστη τιμή συγκέντρωσης SO₂ να ανέρχεται στα 0.0 $\mu\text{g}/\text{m}^3$. Οι στατιστικοί δείκτες Pearson βρίσκονται άνω του 0.90 παρουσιάζοντας μια ικανοποιητική συσχέτιση των δεδομένων μεταξύ τους, συμπέρασμα το οποίο επίσης επιβεβαιώνεται και από τον δείκτη IA, με αντίστοιχες τιμές άνω του 95%. Όσον αφορά τους δείκτες αληθούς/ψεύδους αποτελέσματος, παρατηρούνται τιμές άνω του 90%, συνεπώς μπορεί να ειπωθεί ότι είναι δυνατή η σχεδόν πλήρης πρόβλεψη των υπερβάσεων. Ως οριακή τιμή της συγκέντρωσης των SO₂ θεωρήθηκαν, βάσει του Παγκόσμιου Οργανισμού Υγείας, τα 40 $\mu\text{g}/\text{m}^3$ ημερησίως.

Περιοχή	Σενάριο	A.E.	MAE($\mu\text{g}/\text{m}^3$)	RMSE($\mu\text{g}/\text{m}^3$)	R	IA(%)
1	All Meteorological & RH	BR ₄	0.03	16.92	0.890	94.05
2	All Meteorological & RH	BR ₄	0.13	14.14	0.879	93.34
3	All Meteorological	BR ₃	0.28	15.44	0.892	94.16
4	All Meteorological	BR ₃	0.11	15.93	0.892	94.13
5	All Meteorological & RH	BR ₄	0.03	16.67	0.890	94.04
6	All Meteorological & RH	BR ₄	0.05	16.62	0.890	94.05
8	All Meteorological	BR ₃	0.02	15.23	0.880	93.45
9	All Meteorological & RH	BR ₄	0.12	18.26	0.948	97.28
10	All Meteorological & RH	BR ₃	0.10	16.04	0.895	94.32

ΠΙΝΑΚΑΣ 16g: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα NO₂.

Περιοχή	Σενάριο	A.E.	TPR (%)	FPR (%)	FAR (%)	SI (%)
1	All Meteorological & RH	BR ₄	57.78	0.07	31.58	99.82
2	Only WD & WSPM	BR ₄	50.00	0.03	76.92	99.96
3	Only TEMP & WD & WSPM	BR ₃	65.00	0.03	25.71	99.93
4	All Meteorological	BR ₃	47.06	0.05	44.19	99.87
5	All Meteorological	BR ₄	60.00	0.04	25.00	99.89
6	All Meteorological	BR ₃	40.68	0.04	36.84	99.86
8	All Meteorological	BR ₄	67.74	0.02	22.22	99.95
9	All Meteorological	BR ₃	64.84	0.07	22.43	99.80
10	All Meteorological & RH	BR ₃	88.16	1.02	7.46	97.62

ΠΙΝΑΚΑΣ 16h: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα NO₂.

Στους πίνακες 16g & 16h παρουσιάζεται ως βέλτιστη λύση το σενάριο 1 και 2 (σε ποσοστά 44%) σε συνδυασμό με τον A.E. BR₃ & BR₄. Ως περιοχή με βέλτιστη προβλεπτική ικανότητα παρουσιάζεται η περιοχή 9 και 10, η πρώτη παρουσιάζοντας την μέγιστη ικανότητα προσαρμογής των δεδομένων και η δεύτερη παρουσιάζοντας την υψηλότερη ικανότητα πρόγνωσης των υπερβάσεων. Τα μέσα απόλυτα σφάλματα παρουσιάζουν τιμές μικρότερες του 0.3 $\mu\text{g}/\text{m}^3$, με την ελάχιστη τιμή συγκέντρωσης SO₂ να ανέρχεται στα 1.0 $\mu\text{g}/\text{m}^3$. Οι στατιστικοί δείκτες Pearson βρίσκονται άνω του 0.85 παρουσιάζοντας τη συσχέτιση των δεδομένων μεταξύ τους, συμπέρασμα το οποίο επίσης επιβεβαιώνεται και από τον δείκτη IA, με αντίστοιχες τιμές κάτω του 90%. Όσον αφορά τους δείκτες αληθούς/ψευδούς αποτελέσματος, παρατηρούνται τιμές άνω του 90%, με μοναδική εξαίρεση την περιοχή 10, συνεπώς μπορεί να ειπωθεί ότι είναι δυνατή η μερική πρόβλεψη των υπερβάσεων. Ως οριακή τιμή της συγκέντρωσης των NO₂ θεωρήθηκαν, βάσει του Παγκόσμιου Οργανισμού Υγείας, τα 25 $\mu\text{g}/\text{m}^3$ ημερησίως.

Περιοχή	Σενάριο	A.E.	MAE(mg/m ³)	RMSE(mg/m ³)	R	IA (%)
1	All Meteorological & RH	BR ₄	0.011	0.480	0.924	96.01
2	All Meteorological & RH	BR ₄	0.003	0.455	0.915	95.47
3	All Meteorological & RH	BR ₁	0.001	0.475	0.918	95.63
4	All Meteorological & RH	BR ₄	0.003	0.446	0.925	96.05
5	All Meteorological & RH	BR ₃	0.007	0.442	0.932	96.39
6	All Meteorological & RH	BR ₄	0.003	0.450	0.934	96.57
8	All Meteorological	BR ₄	0.002	0.438	0.929	96.26
9	All Meteorological	BR ₃	0.000	0.441	0.939	96.84
10	All Meteorological & RH	BR ₄	0.002	0.447	0.932	96.44

ΠΙΝΑΚΑΣ 16i: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα CO.

Περιοχή	Σενάριο	A.E.	TPR (%)	FPR (%)	FAR (%)	SI (%)
1	No RAIN & DEWPOINT	BR ₂	94.87	0.14	5.13	99.73
2	Only TEMP & WD & WSPM	BR ₁	95.24	0.00	0.00	99.93
3	All Meteorological & RH	BR ₁	93.55	0.00	0.00	99.86
4	No RAIN & DEWPOINT	BR ₁	96.97	0.00	0.00	99.93
5	No RAIN & DEWPOINT	BR ₁	94.44	0.07	2.86	99.79
6	All Meteorological & RH	BR ₄	97.44	0.07	2.56	99.86
8	All Meteorological	BR ₂	96.97	0.00	0.00	99.93
9	All Meteorological & RH	BR ₃	93.62	0.00	0.00	99.79
10	All Meteorological	BR ₃	94.74	0.14	5.26	99.73

ΠΙΝΑΚΑΣ 16j: Δείκτες στατιστικής αξιολόγησης ψεύδους των καλύτερων σεναρίων ανά περιοχή για τα CO.

Στους πίνακες 16i & 16j παρουσιάζεται ως βέλτιστη λύση το σενάριο 2 (σε ποσοστό 56%) σε συνδυασμό με τον A.E. BR₃ & BR₄. Ως περιοχή με βέλτιστη προβλεπτική ικανότητα παρουσιάζεται η περιοχή 2,3,4,8 και 9, οι πρώτες τέσσερις παρουσιάζοντας την υψηλότερη ικανότητα πρόγνωσης των υπερβάσεων και η δεύτερη παρουσιάζοντας την μέγιστη ικανότητα προσαρμογής των δεδομένων. Τα μέσα απόλυτα σφάλματα παρουσιάζουν τιμές μικρότερες του 0.02 mg/m³, με την ελάχιστη τιμή συγκέντρωσης CO να ανέρχεται στα 0.05 mg/m³. Οι στατιστικοί δείκτες Pearson βρίσκονται άνω του 0.90 παρουσιάζοντας ικανοποιητικά τη συσχέτιση των δεδομένων μεταξύ τους, συμπεράσμα το οποίο επίσης επιβεβαιώνεται και από τον δείκτη IA, με αντίστοιχες τιμές άνω του 95%. Όσον αφορά τους δείκτες αληθούς/ψευδούς αποτελέσματος, παρατηρούνται τιμές άνω του 90% συνεπώς μπορεί να ειπωθεί ότι είναι δυνατή η σχεδόν πλήρης πρόβλεψη των υπερβάσεων. Ως οριακή τιμή της συγκέντρωσης του CO θεωρήθηκαν, βάσει του Παγκόσμιου Οργανισμού Υγείας, τα 4 mg/m³ ημερησίως.

Περιοχή	Σενάριο	A.E.	MAE($\mu\text{g}/\text{m}^3$)	RMSE($\mu\text{g}/\text{m}^3$)	R	IA (%)
1	All Meteorological	BR ₃	0.20	18.21	0.948	97.33
2	All Meteorological & RH	BR ₄	0.06	17.31	0.948	97.29
3	All Meteorological & RH	BR ₃	0.11	19.15	0.945	97.16
4	All Meteorological & RH	BR ₄	0.12	18.26	0.948	97.28
5	All Meteorological & RH	BR ₄	0.11	18.11	0.948	97.31
6	All Meteorological & RH	BR ₃	0.07	17.95	0.952	97.49
8	All Meteorological	BR ₄	0.15	18.87	0.948	97.30
9	All Meteorological & RH	BR ₃	0.01	16.26	0.954	97.64
10	All Meteorological	BR ₄	0.10	17.68	0.951	97.46

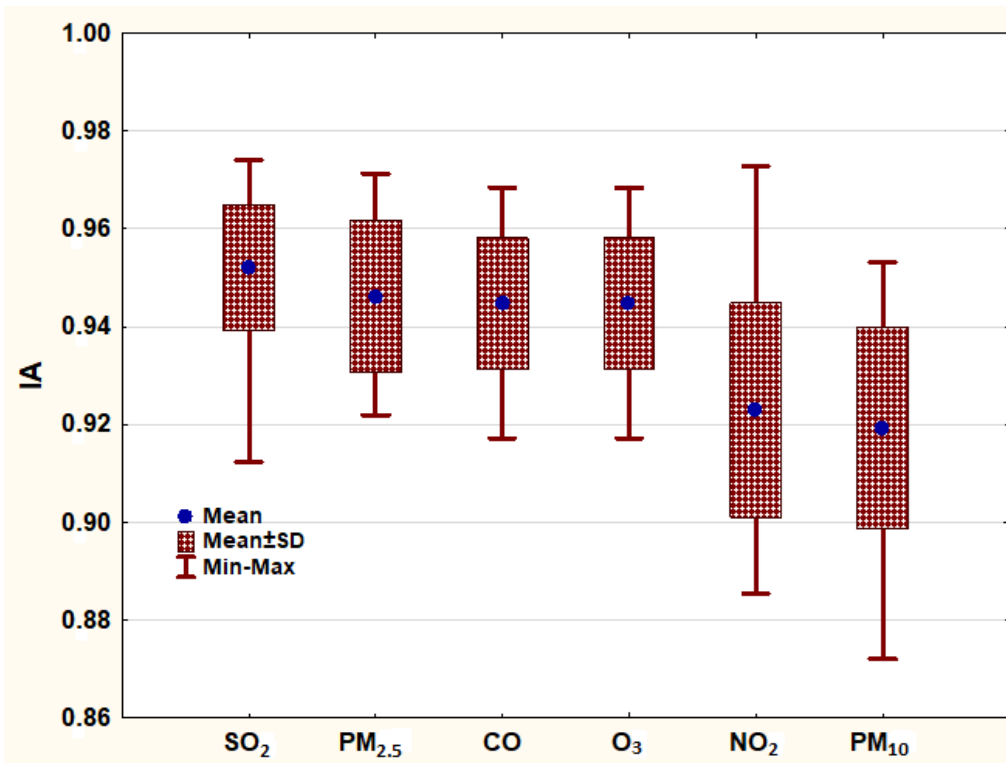
ΠΙΝΑΚΑΣ 16k: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα O₃.

Περιοχή	Σενάριο	A.E.	TPR (%)	FPR (%)	FAR (%)	SI (%)
1	All Meteorological	BR ₄	88.22	0.89	6.72	97.78
2	All Meteorological	BR ₃	90.46	0.99	7.42	97.99
3	All Meteorological & RH	BR ₃	89.29	1.56	11.21	97.32
4	All Meteorological & RH	BR ₃	88.16	1.02	7.46	97.62
5	All Meteorological & RH	BR ₃	90.00	1.29	8.86	97.60
6	All Meteorological	BR ₄	88.89	1.06	7.04	97.57
8	All Meteorological & RH	BR ₃	87.83	1.11	7.68	97.44
9	All Meteorological & RH	BR ₄	89.88	0.89	8.39	98.22
10	All Meteorological	BR ₃	90.68	1.31	8.99	97.67

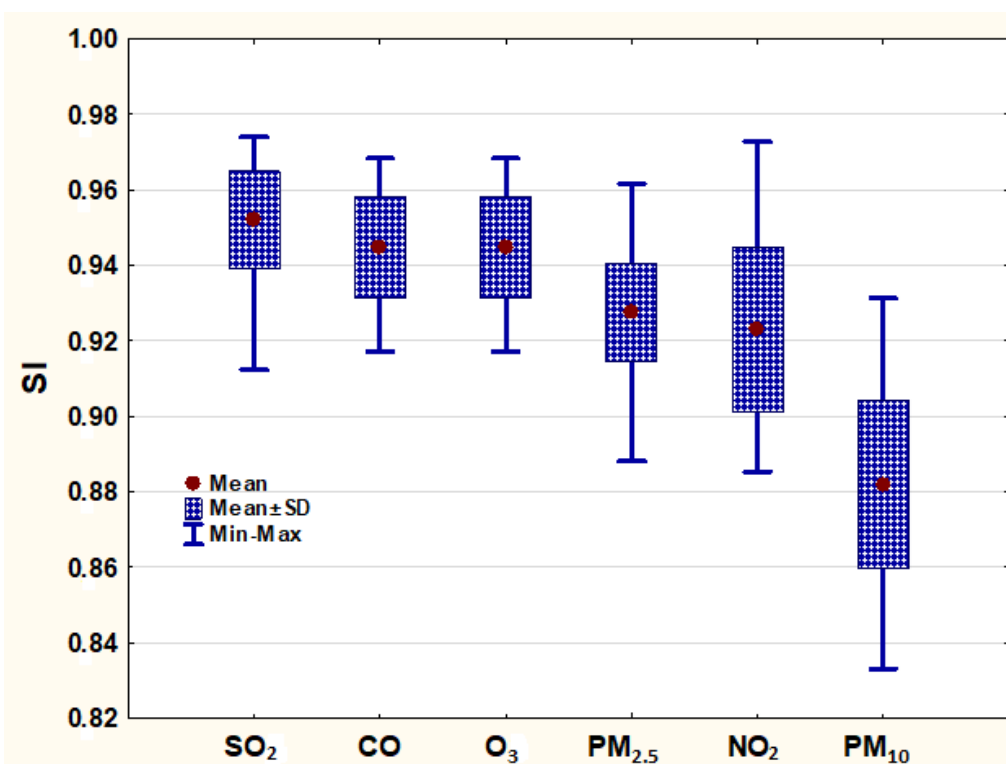
ΠΙΝΑΚΑΣ 16l: Δείκτες στατιστικής αξιολόγησης των καλύτερων σεναρίων ανά περιοχή για τα O₃.

Στους πίνακες 16k & 16l παρουσιάζεται ως βέλτιστη λύση το σενάριο 2 (σε ποσοστό 61%) σε συνδυασμό με τον A.E. BR₃ & BR₄. Ως περιοχή με βέλτιστη προβλεπτική ικανότητα παρουσιάζεται η περιοχή 1 και 9, η πρώτη παρουσιάζοντας την υψηλότερη ικανότητα πρόγνωσης των υπερβάσεων και η δεύτερη παρουσιάζοντας την μέγιστη ικανότητα προσαρμογής των δεδομένων. Τα μέσα απόλυτα σφάλματα παρουσιάζουν τιμές μικρότερες του 0.25 $\mu\text{g}/\text{m}^3$, με την ελάχιστη τιμή συγκέντρωσης O₃ να ανέρχεται στα 0.0 $\mu\text{g}/\text{m}^3$. Οι στατιστικοί δείκτες Pearson βρίσκονται τους, άνω του 0.90 παρουσιάζοντας ικανοποιητικά τη συσχέτιση των δεδομένων μεταξύ τους, συμπεράσμα το οποίο επίσης επιβεβαιώνεται και από τον δείκτη IA, με αντίστοιχες τιμές άνω του 95%. Όσον αφορά τους δείκτες αληθούς/ψευδούς αποτελέσματος, παρατηρούνται τιμές άνω του 85% συνεπώς μπορεί να ειπωθεί ότι είναι δυνατή η σχεδόν πλήρης πρόβλεψη των υπερβάσεων. Ως οριακή τιμή της συγκέντρωσης του O₃ θεωρήθηκαν, βάσει του Παγκόσμιου Οργανισμού Υγείας, τα 100 $\mu\text{g}/\text{m}^3$ εντός ενός κυλιόμενου οχταώρου.

Προκειμένου να γίνει εφικτή η παρουσίαση της γενικότερης εικόνας της εκπαίδευσης των ΤΝΔ, σχεδιάστηκαν τα δύο παρακάτω γραφήματα, μορφής Box & Whisker στα οποία απεικονίζονται οι δείκτες IA και SI για κάθε σενάριο και κάθε περιοχή ανά ρύπο (Σχήμα 12 και Σχήμα 13).



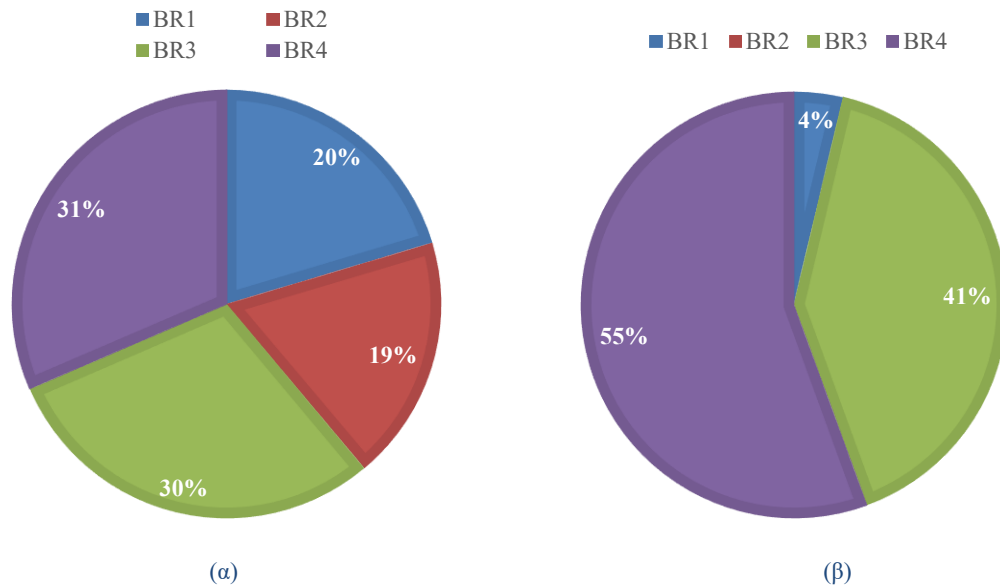
ΣΧΗΜΑ 12 : Διάγραμμα Box & Whisker για τον δείκτη IA για όλα τα σενάρια εκπαίδευσης.



ΣΧΗΜΑ 13 : Διάγραμμα Box & Whisker για τον δείκτη SI για όλα τα σενάρια εκπαίδευσης.

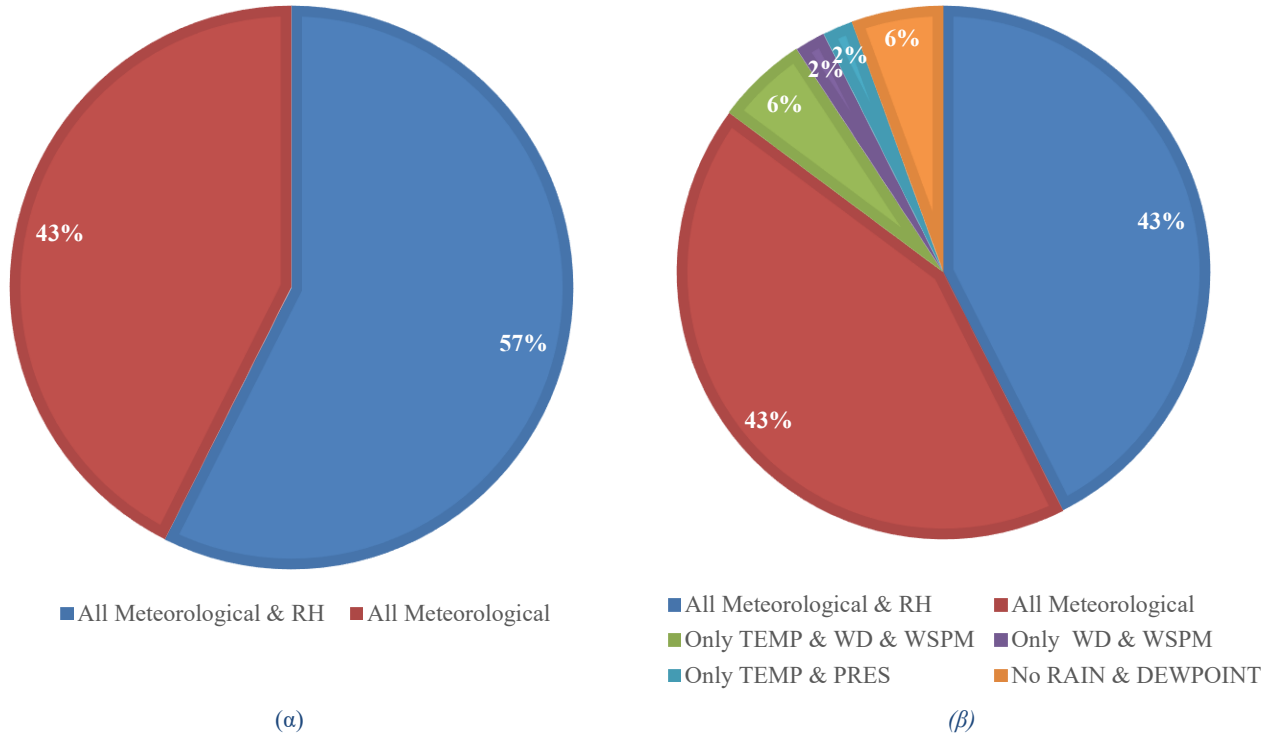
Στην προσπάθεια εξαγωγής της γενικότερης συμπεριφοράς των ΤΝΔ, ως προς την προγνωστική τους ικανότητα, αποφασίστηκε η δημιουργία γραφημάτων Box & Whisker. Απαραίτητα δεδομένα για τον σχεδιασμό των γραφημάτων αυτών είναι η μέγιστη, μέση και ελάχιστη τιμή κάθε μεγέθους, καθώς και η σχετική απόκλιση που παρουσιάζουν οι τιμές. Τα δεδομένα αποτελούνται από τις βέλτιστες τιμές για κάθε ρύπο και για βέλτιστο Α.Ε., συνεπώς από 72 συνολικά τιμές (9 περιοχές από 8 σενάρια δεδομένων εισόδου). Τα δεδομένα παρουσιάζουν ενθαρρυντικά αποτελέσματα, ιδιαίτερα για τις συγκεντρώσεις των PM_{2.5} και του SO₂, με τις πρώτες να παρουσιάζουν μικρότερη διακύμανση και τις τελευταίες τη μεγαλύτερη μέγιστη τιμή.

Βάσει των δεικτών των Πινάκων 16a έως 16l γίνεται εμφανές ότι η εκπαίδευση ΤΝΔ με τη χρήση του αλγόριθμου Bayesian Regularization αποδίδει συνεχώς τα καλύτερα αποτελέσματα, ενώ όσον αφορά τις τέσσερις διαφορετικές αρχιτεκτονικές αυτής της μεθόδου παρατηρήθηκαν ανάμεικτα αποτελέσματα. Συγκεκριμένα όπως φαίνεται στα παρακάτω σχήματα, η δομή BR₄ φαίνεται να παρουσιάζει τα καλύτερα αποτελέσματα όσον αφορά τους στατιστικούς δείκτες αξιοπιστίας (MAE, RMSE, R και IA). Παρόμοια αποτελέσματα παρουσιάζει ο BR₄ και για τους στατιστικούς δείκτες ψευδών προβλέψεων, με βασική διαφορά ότι εδώ συναγωνίζεται με τον BR₃ σε πολύ κοντινά ποσοστά (31% και 30% αντίστοιχα). Ωστόσο μεταξύ δυο Α.Ε. αυτός με την απλούστερη δομή πάντα θα υπερτερεί έναντι μιας πιο περίπλοκης μεθόδου.



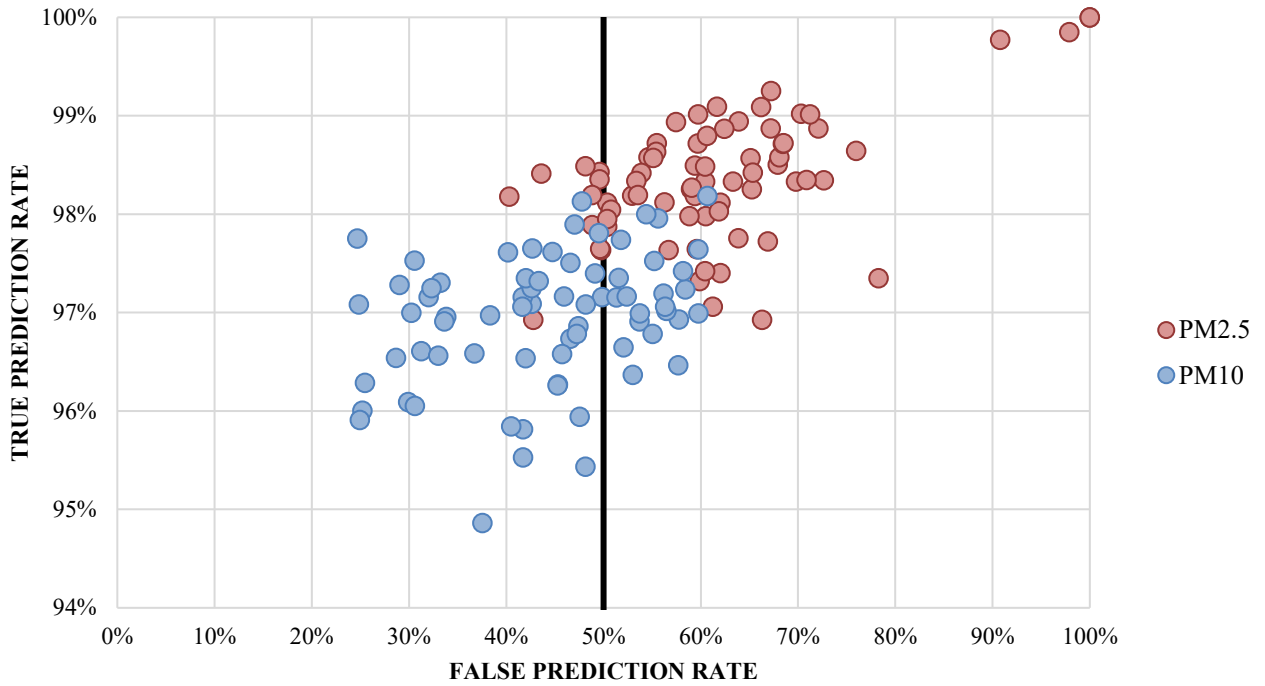
ΣΧΗΜΑ 14 : Ποσοστό εμφάνισης κάθε Α.Ε. σε βέλτιστο σενάριο για (α) Στατιστικούς Δείκτες αξιοπιστίας και (β) Στατιστικούς Δείκτες ψευδών προβλέψεων

Όσον αφορά τα δεδομένα εκπαίδευσης στο ευρύτερο φάσμα της εκπαίδευσης φάνηκε να παρουσιάζονται τα καλύτερα αποτελέσματα όταν αξιοποιούνταν το σενάριο 1 και 2 (με ποσοστό 43% και 57% αντίστοιχα), για το πρώτο σετ στατιστικών δεικτών, ενώ στο δεύτερο παρότι υπήρχε παρουσία περισσότερων σεναρίων ως βέλτιστες λύσεις, μονάχα τα σενάρια 1 και 2 πέτυχαν ποσοστά χρήσης άνω του 6% (43% έκαστο).

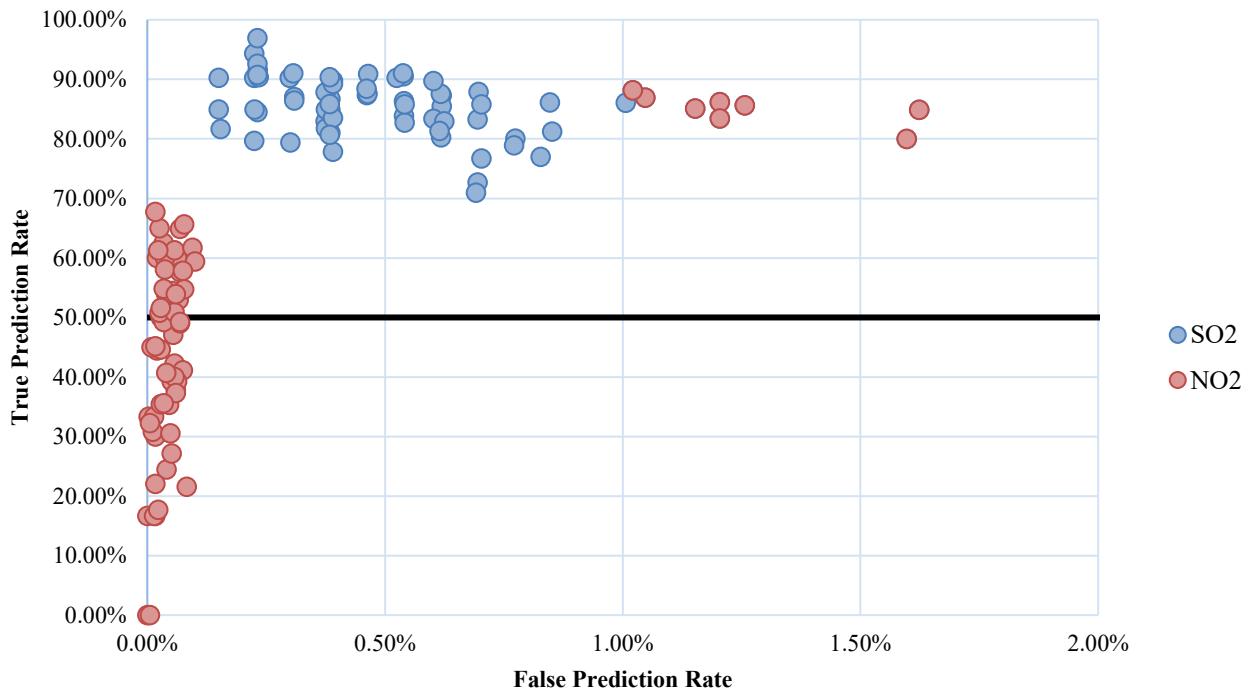


ΣΧΗΜΑ 15 : Ποσοστό εμφάνισης κάθε σεναρίου μεταβλητών εισόδου, ως βέλτιστο σενάριο για (α) Στατιστικούς Δείκτες αξιοπιστίας και (β) Στατιστικούς Δείκτες ψευδών προβλέψεων

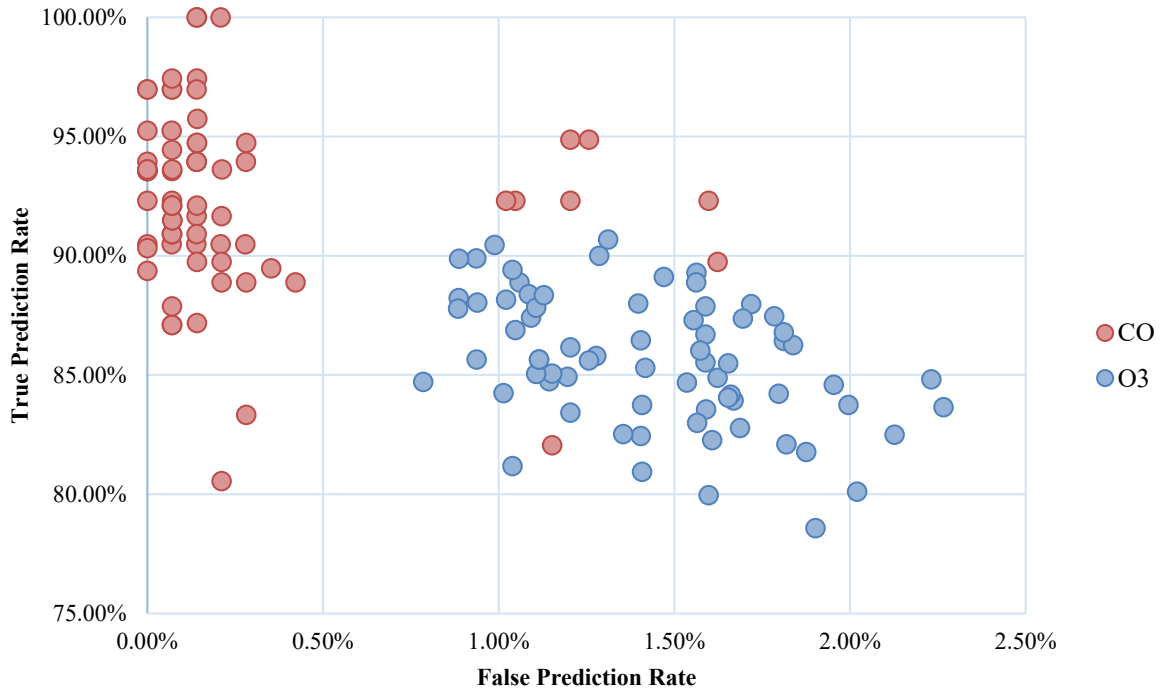
Προκειμένου να αξιολογηθούν τα ΤΝΔ περεταίρω, καθώς και προκειμένου να εξαχθεί η μέγιστη ποσότητα διαθέσιμης πληροφορίας, κρίθηκε σκόπιμη η δημιουργία διαγραμμάτων ROC (Receiver Operating Characteristic) για τις τιμές των TPR και FPR κάθε ρύπου ανά περιοχή και ανά βέλτιστο σενάριο. Με τη χρήση του παρακάτω διαγράμματος γίνεται εφικτός ο γραφικός προσδιορισμός του βέλτιστου σεναρίου. Στα διαγράμματα τύπου ROC, παρουσιάζονται οι τιμές των πραγματικών προβλέψεων συναρτήσει των τιμών ψευδών προβλέψεων. Συνεπώς σε αυτά τα διαγράμματα σημείο ενδιαφέροντος αποτελεί η περιοχή του 1^{ου} τεταρτημόριου φραγμένη για τιμές $x < 0.5$ και τιμές $y > 0.5$. Είναι προφανές πως η ιδανική λύση θα είναι αυτή που βρίσκεται επάνω στον άξονα $y'y$ και όσο το δυνατόν πιο κοντά στο $y=1$. Τα παραπάνω προγνωστικά μοντέλα όπως γίνεται εμφανές παρουσιάζουν πολύ καλές προγνωστικές ικανότητες με την πλειοψηφία των μοντέλων (βέλτιστων και μη) να βρίσκονται εντός των επιθυμητών ορίων.



ΣΧΗΜΑ 16α : Διάγραμμα ROC για κάθε σενάριο και κάθε περιοχή για τα αιωρούμενα σωματίδια αεροδυναμικής διαμέτρου 2.5 και 10 μm.



ΣΧΗΜΑ 16β : Διάγραμμα ROC για κάθε σενάριο και κάθε περιοχή για διοξείδια του θείου και του αζώτου.



ΣΧΗΜΑ 16c : Διάγραμμα ROC για κάθε σενάριο και κάθε περιοχή για όζον και μονοξείδιο του άνθρακα.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στόχος της παρούσας διπλωματικής εργασίας ήταν η δημιουργία ενός καθολικού μοντέλου πρόγνωσης των συγκεντρώσεων των ατμοσφαιρικών ρύπων και συγκεκριμένα των σωματιδίων PM_{2.5} και PM₁₀, καθώς και των SO₂, NO₂, CO και O₃. Βασική προϋπόθεση για την εκπαίδευση των μοντέλων αυτών είναι η κατάλληλη προ-επεξεργασία των δεδομένων εισόδου (inputs) προκειμένου να εξαχθούν τυχόν κρυφές σχέσεις, οποίες θα διευκολύνουν τη μετέπειτα δημιουργία των σεναρίων. Η διαδικασία κάλυψης των κενών αποτέλεσε μια καινοτομία η οποία δεν βρέθηκε σε κάποια βιβλιογραφία και προσέφερε ικανοποιητικά αποτελέσματα. Οι αλγόριθμοι που δημιουργήθηκαν για αυτή τη διαδικασία αποτελούν επίσης καθολικά μοντέλα και η εφαρμογή είναι εφικτή σε οποιοδήποτε άλλο σετ δεδομένων, με μικρές αναπροσαρμογές. Τέλος ακολούθησε η σύγκριση οκτώ πιθανών σεναρίων τιμών εισόδου και η εκπαίδευση καθενός με δέκα διαφορετικούς Α.Ε. Η σύγκριση των αποτελεσμάτων, προσέφερε πολύτιμη γνώση σχετικά με τη βελτιστοποίηση της εκπαίδευσης των Τεχνητών Νευρωνικών Δικτύων έχοντας ως παράλληλη μεταβλητή και το είδος του προς μελέτη ρύπου. Τα αποτελέσματα ήταν αποδεκτά στην πλειοψηφία των ρύπων με μέσες τιμές πρόβλεψης άνω του 80%. Μοναδικό έλλειμμα της παρούσας εργασίας αποτελεί η αδυναμία εύρεσης κατάλληλου συνδυασμού μεταβλητών εισόδου και αρχιτεκτονικής ΤΝΔ, ώστε να επιτευχθούν υψηλότερα ποσοστά πρόβλεψης των υπερβάσεων. Ωστόσο η εκπαίδευση περαιτέρω ΤΝΔ δεν ήταν δυνατή, λόγω χρονικών περιορισμών, καθώς και την περιορισμένη υπολογιστική ισχύ των υποδομών μελέτης. Μέσω περαιτέρω στατιστικής ανάλυσης αποδείχθηκε ότι η αξιοποίηση όλων των μετεωρολογικών μεταβλητών ενισχύει την επίδοση της εκπαίδευσης των ΤΝΔ και δεν τα «μπερδεύει». Ακόμη διαπιστώθηκε ότι η δομή BR₄ αποτελεί την πλέον κατάλληλη αρχιτεκτονική εκπαίδευσης ΤΝΔ.

Πράγματι η πρόβλεψη των συγκεντρώσεων των ατμοσφαιρικών ρύπων τυπικά αποτελεί μια επίπονη και χρονοβόρα διαδικασία, ωστόσο μέσω της συγκεκριμένης εργασίας γίνεται αντιληπτό ότι με την κατάλληλη επιλογή των μεταβλητών εισόδου και των αλγορίθμων εκπαίδευσης, η διαδικασία της εκπαίδευσης ενός προγνωστικού μοντέλου μπορεί να βελτιστοποιηθεί, χωρίς την ανάγκη ύπαρξης μεγάλης υπολογιστικής ισχύος. Το εξαγόμενο μοντέλο μπορεί να χρησιμοποιηθεί τόσο από δημόσιους, όσο και ιδιωτικούς φορείς έτσι ώστε να επιτευχθεί η άμεση ενημέρωση αυτών αλλά και του κοινού, σε μέτρα προφύλαξης και αντιμετώπισης.

Οι εξελίξεις του κλάδου της Μηχανικής Μάθησης είναι ολοένα και αυξανόμενες και ο ρυθμός ανάπτυξης αυτής ραγδαίος, με καινούργιες προσθήκες ημερησίως. Όπως κάθε αλγόριθμος έτσι και αυτός επιδέχεται βελτιστοποίησης, καθώς θα ήταν ενδιαφέρον να ερευνηθεί πως θα επηρέαζε την απόδοση των ΤΝΔ η ύπαρξη τιμών άλλων ρύπων εντός των δεδομένων εκπαίδευσης, ή η αύξηση του χρονικού πλαισίου των τιμών εισόδου από τρεις ημέρες σε περισσότερες. Τέλος, ενδιαφέρον θα παρουσίαζε η συμπεριφορά ενός μοντέλου πρόβλεψης, στα δεδομένα εισόδου του οποίου θα υπήρχαν, μεταξύ άλλων, τα δεδομένα εξόδου ενός άλλου μοντέλου.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Αθανάσιος Γ. Κούγκολος, *Περιβαλλοντική Μηχανική Ρύπανση & Προστασία Περιβάλλοντος*, 3rd ed. Εκδόσεις ΤΖΙΟΛΑ, 2021.
- [2] “Types of pollutants.” Accessed: Apr. 15, 2024. [Online]. Available: <https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants>
- [3] World Health Organization, *WHO global air quality guidelines. Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization. [Online]. Available: <https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdf?sequence=1>
- [4] R. K. Srivastava, C. A. Miller, C. Erickson, and R. Jambhekar, “Emissions of Sulfur Trioxide from Coal-Fired Power Plants,” *Journal of the Air & Waste Management Association*, vol. 54, no. 6, pp. 750–762, Jun. 2004, doi: 10.1080/10473289.2004.10470943.
- [5] UNEP, Ed., *Action for ozone: 2000 edition*. Nairobi: United Nations Environment Programme, 2000.
- [6] M. E. Jenkin and K. C. Clemitshaw, “Ozone and other secondary photochemical pollutants: chemical processes governing their formation in the planetary boundary layer,” *Atmospheric Environment*, vol. 34, no. 16, pp. 2499–2527, Jan. 2000, doi: 10.1016/S1352-2310(99)00478-1.
- [7] K. Ntourou, K. Moustiris, G. Spyropoulos, K.-M. Fameli, and N. Manousakis, “Adverse Health Effects (Bronchitis Cases) Due to Particulate Matter Exposure: A Twenty-Year Scenario Analysis for the Greater Athens Area (Greece) Using the AirQ+ Model,” *Atmosphere*, vol. 14, no. 2, p. 389, Feb. 2023, doi: 10.3390/atmos14020389.
- [8] X. Mo, H. Li, L. Zhang, and Z. Qu, “Environmental impact estimation of PM_{2.5} in representative regions of China from 2015 to 2019: policy validity, disaster threat, health risk, and economic loss,” *Air Qual Atmos Health*, vol. 14, no. 10, pp. 1571–1585, Oct. 2021, doi: 10.1007/s11869-021-01040-8.
- [9] P. Amoatey, P. Sicard, A. De Marco, and Y. O. Khaniabadi, “Long-term exposure to ambient PM_{2.5} and impacts on health in Rome, Italy,” *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 531–535, Jun. 2020, doi: 10.1016/j.cegh.2019.11.009.
- [10] M. M. Veras, S. C. L. Farhat, A. C. Rodrigues, D. Waked, and P. H. N. Saldiva, “Beyond respiratory effects: Air pollution and the health of children and adolescents,” *Current Opinion in Environmental Science & Health*, vol. 32, p. 100435, Apr. 2023, doi: 10.1016/j.coesh.2022.100435.
- [11] M. Veras, D. Waked, and P. Saldiva, “Safe in the womb? Effects of air pollution to the unborn child and neonates,” *Jornal de Pediatria*, vol. 98, pp. S27–S31, Mar. 2022, doi: 10.1016/j.jpmed.2021.09.004.
- [12] M. Elten *et al.*, “Ambient air pollution and incidence of early-onset paediatric type 1 diabetes: A retrospective population-based cohort study,” *Environmental Research*, vol. 184, p. 109291, May 2020, doi: 10.1016/j.envres.2020.109291.
- [13] R. Raz *et al.*, “Autism Spectrum Disorder and Particulate Matter Air Pollution before, during, and after Pregnancy: A Nested Case–Control Analysis within the Nurses’ Health Study II Cohort,” *Environ Health Perspect*, vol. 123, no. 3, pp. 264–270, Mar. 2015, doi: 10.1289/ehp.1408133.
- [14] V. Simoncic, C. Enaux, S. Deguen, and W. Kihal-Talantikite, “Adverse Birth Outcomes Related to NO₂ and PM Exposure: European Systematic Review and Meta-Analysis,” *IJERPH*, vol. 17, no. 21, p. 8116, Nov. 2020, doi: 10.3390/ijerph17218116.
- [15] K. Ntourou, K.-M. Fameli, K. Moustiris, A. Augoustinos, and C. Tsitsis, “The Influence of Ozone Concentrations on Public Health over the Greater Athens Area, Greece,” in *16th International Conference on Meteorology, Climatology and Atmospheric Physics—COMECAP 2023*, MDPI, Aug. 2023, p. 107. doi: 10.3390/environsciproc2023026107.
- [16] Z. Feng, E. Hu, X. Wang, L. Jiang, and X. Liu, “Ground-level O₃ pollution and its impacts on food crops in China: A review,” *Environmental Pollution*, vol. 199, pp. 42–48, Apr. 2015, doi: 10.1016/j.envpol.2015.01.016.

- [17] A. Hernández-Peña, E. A. Gallardo-Hernández, L. I. Farfan-Cabrera, M. Vite-Torres, and J. Muñoz-Saldaña, “Solid particle erosion evaluation of automotive paint coatings under the influence of artificial weathering,” *Wear*, vol. 532–533, p. 205105, Nov. 2023, doi: 10.1016/j.wear.2023.205105.
- [18] A. M. Ibrahim, M. T. Bassuoni, J. Carroll, and A. Ghazy, “Performance of concrete superficially treated with nano-modified coatings under sulfuric acid exposures,” *Journal of Building Engineering*, vol. 86, p. 108957, Jun. 2024, doi: 10.1016/j.jobbe.2024.108957.
- [19] “What Happened to Acid Rain? | Britannica.” Accessed: May 25, 2024. [Online]. Available: <https://www.britannica.com/story/what-happened-to-acid-rain>
- [20] “The end of acid rain - Works in Progress.” Accessed: May 25, 2024. [Online]. Available: <https://worksinprogress.co/issue/the-end-of-acid-rain/>
- [21] “Acid Rain,” Young People’s Trust For the Environment. Accessed: May 25, 2024. [Online]. Available: <https://ypte.org.uk/factsheets/acid-rain/forests>
- [22] Μέτρα για τη βελτίωση της ποιότητας της ατμόσφαιρας, σε συμμόρφωση με τις διατάξεις της οδηγίας 2008/50/EK «για την ποιότητα του ατμοσφαιρικού αέρα και καθαρότερο αέρα για την Ευρώπη» του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της Ευρωπαϊκής ένωσης της 21ης Μαΐου 2008».
- [23] *World Health Organization (WHO) air quality guidelines (AQGs) and estimated reference levels (RLs)*.
- [24] Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, and Η. Σακελλαρίου, *Τεχνητή Νοημοσύνη*, 4th ed. 2020.
- [25] Ιωάννης Ε. Σαρρής and Θεόδωρος Καρακασίδης, *Αριθμητικές Μέθοδοι και Εφαρμογές για Μηχανικούς με Παραδείγματα στο MATLAB*, vol. 4η. 2017.
- [26] C. M. Bishop, *Pattern recognition and machine learning*. in Information science and statistics. New York: Springer, 2006.
- [27] “Choose a Multilayer Neural Network Training Function - MATLAB & Simulink.” Accessed: May 16, 2024. [Online]. Available: <https://www.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html>
- [28] M. L. A. Lourakis and A. A. Argyros, “Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment?,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, Beijing, China: IEEE, 2005, pp. 1526-1531 Vol. 2. doi: 10.1109/ICCV.2005.128.
- [29] “Saddle point method - Encyclopedia of Mathematics.” Accessed: May 19, 2024. [Online]. Available: https://encyclopediaofmath.org/index.php?title=Saddle_point_method
- [30] W. F. Mascarenhas, “The divergence of the BFGS and Gauss Newton methods,” *Math. Program.*, vol. 147, no. 1–2, pp. 253–276, Oct. 2014, doi: 10.1007/s10107-013-0720-6.
- [31] H. Okut, D. Gianola, G. J. M. Rosa, and K. A. Weigel, “Prediction of body mass index in mice using dense molecular markers and a regularized neural network,” *Genet. Res.*, vol. 93, no. 3, pp. 189–201, Jun. 2011, doi: 10.1017/S0016672310000662.
- [32] F. Dan Foresee and M. T. Hagan, “Gauss-Newton approximation to Bayesian learning,” in *Proceedings of International Conference on Neural Networks (ICNN’97)*, Houston, TX, USA: IEEE, 1997, pp. 1930–1935. doi: 10.1109/ICNN.1997.614194.
- [33] N. G. Polson and V. Sokolov, “Bayesian regularization: From Tikhonov to horseshoe,” *WIREs Computational Stats*, vol. 11, no. 4, p. e1463, Jul. 2019, doi: 10.1002/wics.1463.
- [34] M. Kayri, “Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data,” *MCA*, vol. 21, no. 2, p. 20, May 2016, doi: 10.3390/mca21020020.
- [35] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, “Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation,” *Atmospheric Environment*, vol. 107, pp. 118–128, Apr. 2015, doi: 10.1016/j.atmosenv.2015.02.030.
- [36] M. A. Elangasinghe, N. Singhal, K. N. Dirks, and J. A. Salmond, “Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis,” *Atmospheric Pollution Research*, vol. 5, no. 4, pp. 696–708, Oct. 2014, doi: 10.5094/APR.2014.079.

- [37] Y. Bai, Y. Li, X. Wang, J. Xie, and C. Li, “Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions,” *Atmospheric Pollution Research*, vol. 7, no. 3, pp. 557–566, May 2016, doi: 10.1016/j.apr.2016.01.004.
- [38] F. Biancofiore *et al.*, “Recursive neural network model for analysis and forecast of PM10 and PM2.5,” *Atmospheric Pollution Research*, vol. 8, no. 4, pp. 652–659, Jul. 2017, doi: 10.1016/j.apr.2016.12.014.
- [39] F. Franceschi, M. Cobo, and M. Figueredo, “Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering,” *Atmospheric Pollution Research*, vol. 9, no. 5, pp. 912–922, Sep. 2018, doi: 10.1016/j.apr.2018.02.006.
- [40] Q. Chen, R. Ding, X. Mo, H. Li, L. Xie, and J. Yang, “An adaptive adjacency matrix-based graph convolutional recurrent network for air quality prediction,” *Sci Rep*, vol. 14, no. 1, p. 4408, Feb. 2024, doi: 10.1038/s41598-024-55060-2.
- [41] S. D. Latif *et al.*, “Ozone concentration forecasting utilizing leveraging of regression machine learnings: A case study at Klang Valley, Malaysia,” *Results in Engineering*, vol. 21, p. 101872, Mar. 2024, doi: 10.1016/j.rineng.2024.101872.
- [42] S. Chen, “Beijing Multi-Site Air Quality.” [object Object], 2017. doi: 10.24432/C5RK5G.
- [43] M. Hazewinkel, *Encyclopaedia of mathematics*. Berlin: Springer-Verlag, 2002.
- [44] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Phil. Trans. R. Soc. A.*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.
- [45] Z. Zhang *et al.*, “Evolution of surface O3 and PM2.5 concentrations and their relationships with meteorological conditions over the last decade in Beijing,” *Atmospheric Environment*, vol. 108, pp. 67–75, May 2015, doi: 10.1016/j.atmosenv.2015.02.071.
- [46] “Relative Humidity Calculator,” Omni Calculator. Accessed: Apr. 27, 2024. [Online]. Available: <https://www.omnicalculator.com/physics/relative-humidity>
- [47] K. P. Moustris, I. C. Ziomas, and A. G. Paliatsos, “3-Day-Ahead Forecasting of Regional Pollution Index for the Pollutants NO2, CO, SO2, and O3 Using Artificial Neural Networks in Athens, Greece,” *Water Air Soil Pollut*, vol. 209, no. 1–4, pp. 29–43, Jun. 2010, doi: 10.1007/s11270-009-0179-5.
- [48] *NATIONAL STANDARD OF THE PEOPLE’S REPUBLIC OF CHINA.*