



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΧΗΜΙΚΩΝ ΜΗΧΑΝΙΚΩΝ &
ΜΗΧΑΝΙΚΩΝ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

«ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΗΣ ΛΕΙΤΟΥΡΓΙΑΣ ΜΙΑΣ ΕΓΚΑΤΑΣΤΑΣΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΝΕΡΟΥ»

Διδακτορική Διατριβή

“MODELLING OF A WATER TREATMENT PLANT OPERATION”

Thesis

ΓΥΠΑΡΑΚΗΣ Γ. ΣΤΥΛΙΑΝΟΣ

ΔΙΠΛ. ΧΗΜΙΚΟΣ ΜΗΧΑΝΙΚΟΣ ΕΜΠ

ΜΑΡΤΙΟΣ 2025

ΚΑΘΗΓΗΤΕΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ

1. ΟΜ. ΚΑΘ. ΕΥΑΓΓΕΛΟΣ ΔΙΑΜΑΝΤΟΠΟΥΛΟΣ, ΣΧΟΛΗ ΧΗΜΗΠΕΡ, ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ (ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ),
2. ΚΑΘ. ΓΕΩΡΓΙΟΣ ΚΑΡΑΤΖΑΣ, ΣΧΟΛΗ ΧΗΜΗΠΕΡ, ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ,
3. ΑΝΑΠΛ. ΚΑΘ. ΤΡΥΦΩΝ ΔΑΡΑΣ, ΣΧΟΛΗ ΧΗΜΗΠΕΡ, ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ,
4. ΑΝΑΠΛ. ΚΑΘ. ΓΕΩΡΓΙΟΣ ΑΡΑΜΠΑΤΖΗΣ, ΣΧΟΛΗ ΜΠΔ, ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ,
5. ΕΠΙΚ. ΚΑΘ. ΕΜΜΑΝΟΥΗΛ ΒΑΡΟΥΧΑΚΗΣ, ΣΧΟΛΗ ΜΗΧΟΠ, ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ,
6. ΚΑΘ. ΑΘΑΝΑΣΙΟΣ ΣΤΑΣΙΝΑΚΗΣ, ΤΜΗΜΑ ΠΕΡΙΒΑΛΛΟΝΤΟΣ, ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ,
7. ΔΡ. ΙΩΑΝΝΗΣ ΤΡΙΧΑΚΗΣ, SCIENCE FOR POLICY RESEARCHER, EUROPEAN COMMISSION, JOINT RESEARCH CENTRE, ISPRA, ITALY

Στην κόρη μου Νεφέλη...

ΠΡΟΛΟΓΟΣ

Η παρούσα Διδακτορική Διατριβή (ΔΔ) ξεκίνησε το 2017 με στόχο την μοντελοποίηση μιας Εγκατάστασης Επεξεργασίας Νερού (EEN), αξιοποιώντας την εμπειρία και τα δεδομένα από την EEN Αποσελέμη του Οργανισμού Ανάπτυξης Κρήτης Α.Ε. (ΟΑΚ Α.Ε.) στα Αγριανά του Δ. Χερσονήσου, Ηρακλείου Κρήτης. Το αρχικό όραμα και ο σχεδιασμός ήταν η μοντελοποίηση να στηρίζεται σε σύγχρονα και αξιόπιστα εργαλεία μοντελοποίησης, για τις κύριες και κρίσιμες μεταβλητές λειτουργίας, η οποία θα βοηθούσε στη λήψη κατάλληλων αποφάσεων στην επιλογή των δόσεων χημικών πεδίου, απαραίτητα εργαλεία στα χέρια ενός οποιοδήποτε λειτουργού μιας τέτοιου είδους εγκατάστασης.

Το όραμα αυτό υλοποιήθηκε με τη συνεχή επιστημονική υποστήριξη του επιβλέποντα καθηγητή κ. Ε. Διαμαντόπουλου, ομότιμου καθηγητού Πολυτεχνείου Κρήτης, με τη συμβουλευτική υποστήριξη των λοιπών μελών της τριμελούς επιτροπής καθηγητού κ. Γ. Καρατζά και αναπληρωτή καθηγητού κ. Τρ. Δάρα και με την αμέριστη και αδιάλειπτη επιστημονική και τεχνική υποστήριξη του Δρ. Ι. Τριχάκη, Science for Policy Researcher, European Commission, Joint Research Centre. Στην παρούσα φάση που η ΔΔ έχει ολοκληρωθεί, αισθάνομαι την ανάγκη να ευχαριστήσω θερμά το σύνολο των προαναφερόμενων για την επικοδομητική συνεργασία έως την ολοκλήρωση αυτού

του δύσκολου αλλά παράλληλα χρήσιμου και καινοτόμου έργου.

Επιπρόσθετα, ευχαριστώ θερμά τον κ. Ε. Μαμαγκάκη, Δ/ντη Διαχείρισης των Συγκοινωνιακών και Υδραυλικών Έργων του ΟΑΚ Α.Ε. για την έγκριση της υλοποίησης της παρούσας διατριβής με χρήση δεδομένων από την ΕΕΝ Αποσελέμη.

Τέλος, αναμφίβολα, πολλά θερμά ευχαριστώ αξίζουν στη σύζυγό μου Φανή για την έμπνευση έναρξης του εγχειρήματος εκπόνησης διδακτορικής διατριβής και τη συνεχή υποστήριξη μέχρι την ολοκλήρωση της. Είθε, αυτή η διδακτορική διατριβή, που αφιερώνεται στη μονάκριβη κόρη μου Νεφέλη, να αποτελέσει κίνητρο στη δική της ζωή, με γνώμονα τα λόγια του Αριστοτέλη «τὰ ἀγαθὰ κόποις κτῶνται».

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	VII
ΠΕΡΙΕΧΟΜΕΝΑ	IX
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	XI
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ.....	XII
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	XIII
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ ΠΑΡΑΡΤΗΜΑΤΟΣ.....	XIV
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ ΠΑΡΑΡΤΗΜΑΤΟΣ	XIV
ΛΙΣΤΑ ΣΥΝΤΟΜΕΥΣΕΩΝ	XV
ΠΕΡΙΛΗΨΗ.....	XVII
ABSTRACT.....	XXI
1 ΕΙΣΑΓΩΓΗ	2
1.1 ΕΙΣΑΓΩΓΙΚΑ ΣΤΟΙΧΕΙΑ	2
1.2 Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ (ARTIFICIAL INTELLIGENCE).....	6
1.3 ΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (ARTIFICIAL NEURAL NETWORKS, ANNs)	6
1.3.1 ΙΔΙΟΤΗΤΕΣ ΚΑΙ ΔΥΝΑΤΟΤΗΤΕΣ ΤΩΝ ΤΝΔ	9
1.3.2 Ο ΑΝΘΡΩΠΙΝΟΣ ΕΓΚΕΦΑΛΟΣ.....	11
1.3.3 ΜΟΝΤΕΛΑ ΕΝΟΣ ΝΕΥΡΩΝΑ	14
1.4 ΣΚΟΠΟΙ ΤΗΣ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ.....	18
2 ΔΕΔΟΜΕΝΑ	20
2.1 ΕΓΚΑΤΑΣΤΑΣΗ ΕΠΕΞΕΡΓΑΣΙΑΣ ΝΕΡΟΥ (ΕΕΝ) ΑΠΟΣΕΛΕΜΗ	20
2.2 ΠΕΡΙΓΡΑΦΗ ΛΕΙΤΟΥΡΓΙΑΣ ΕΕΝ ΑΠΟΣΕΛΕΜΗ.....	22
2.2.1 ΟΖΟΝΩΣΗ	24
2.2.2 ΚΡΟΚΙΔΩΣΗ, ΚΑΘΙΖΗΣΗ.....	25
2.2.3 ΔΙΗΘΗΣΗ	25
2.2.4 ΧΛΩΡΙΩΣΗ	26
2.2.5 ΔΕΞΑΜΕΝΗ ΚΑΘΑΡΟΥ ΝΕΡΟΥ	26
2.2.6 ΕΡΓΑΣΤΗΡΙΟ ΕΛΕΓΧΟΥ ΠΟΙΟΤΗΤΑΣ ΝΕΡΟΥ	26
2.2.7 ΠΑΧΥΝΣΗ ΚΑΙ ΑΦΥΔΑΤΩΣΗ ΙΛΥΟΣ	27
2.2.8 ΜΟΝΑΔΑ ΕΝΕΡΓΟΥ ΑΝΘΡΑΚΑ	27
2.3 ΚΥΡΙΕΣ ΛΕΙΤΟΥΡΓΙΚΕΣ ΜΕΤΑΒΑΝΤΕΣ ΚΑΙ ΜΕΤΑΒΑΝΤΕΣ ΠΟΙΟΤΗΤΑΣ ΝΕΡΟΥ	27
2.3.1 ΚΥΡΙΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΠΑΡΑΜΕΤΡΟΙ ΜΕΤΑΒΛΗΤΩΝ	33
2.3.2 ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ	33
3 ΜΕΘΟΔΟΙ	36

3.1	ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ, ΠΓΠ (MULTIPLE LINEAR REGRESSION, MLR)	36
3.1.1	ΜΕΘΟΔΟΣ ΒΗΜΑΤΙΚΗΣ ΕΠΙΛΟΓΗΣ ΜΕΤΑΒΛΗΤΩΝ (STEPWISE)	38
3.2	ΑΝΑΠΤΥΞΗ ΤΝΔ	40
3.2.1	ΣΕΝΑΡΙΑ ΑΝΑΠΤΥΞΗΣ ΤΝΔ	40
3.2.2	ΚΩΔΙΚΑΣ ΣΤΗ MATLAB	43
3.3	ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ (FACTOR ANALYSIS)	46
3.3.1	Η ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PRINCIPAL COMPONENTS ANALYSIS- PCA)	49
4	ΑΠΟΤΕΛΕΣΜΑΤΑ	52
4.1	ΑΝΑΛΥΣΗ ΠΟΛΛΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (MULTIPLE LINEAR REGRESSION ANALYSIS)	52
4.1.1	ΣΥΣΧΕΤΙΣΗ ΜΕΤΑΒΛΗΤΩΝ	53
4.2	ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΤΝΔ	65
4.2.1	ΕΦΑΡΜΟΓΗ ΚΑΙ ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΤΝΔ	65
4.2.2	ΑΠΑΙΤΟΥΜΕΝΟΣ ΧΡΟΝΟΣ ΑΝΑΠΤΥΞΗΣ ΜΟΝΤΕΛΩΝ ΤΝΔ	66
4.2.3	ΒΕΛΤΙΣΤΟ ΣΕΝΑΡΙΟ ΑΝΑΠΤΥΞΗΣ ΤΝΔ	68
4.2.4	ΠΑΡΑΜΕΤΡΟΙ ΕΞΟΔΟΥ (OUTPUTS) ΤΝΔ	70
4.2.5	ΕΞΙΣΩΣΕΙΣ ΑΠΟ-ΚΑΝΟΝΙΚΟΠΟΙΗΣΗΣ ΜΕΤΑΒΛΗΤΩΝ	75
4.3	ΣΥΓΚΡΙΣΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΤΝΔ ΜΕ MLR ANALYSIS	76
4.4	ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ	78
4.5	ΔΙΕΡΕΥΝΗΤΙΚΗ ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ (EXPLORATORY FACTOR ANALYSIS, EFA) ...	80
5	ΣΥΖΗΤΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ, ΣΥΜΠΕΡΑΣΜΑΤΑ, ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ	96
5.1	ΣΥΖΗΤΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ	96
5.1.1	ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (MULTIPLE LINEAR REGRESSION)	96
5.1.2	ΑΝΑΠΤΥΞΗ ΤΝΔ	97
5.1.3	ΣΥΓΚΡΙΣΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΤΝΔ ΜΕ MLR ANALYSIS	100
5.1.4	ΔΙΕΡΕΥΝΗΤΙΚΗ ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ (EFA)	102
5.2	ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ	104
	ΒΙΒΛΙΟΓΡΑΦΙΑ	106
	ΠΑΡΑΡΤΗΜΑΤΑ	111
	ΠΑΡΑΡΤΗΜΑ Ι – ΔΙΑΓΡΑΜΜΑΤΑ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕΤΑΒΛΗΤΩΝ	112
	ΠΑΡΑΡΤΗΜΑ ΙΙ – ΣΥΝΤΕΛΕΣΤΕΣ ΣΥΣΧΕΤΙΣΗΣ PEARSON	117
	ΠΑΡΑΡΤΗΜΑ ΙΙΙ – ΚΩΔΙΚΑΣ ΣΤΗ MATLAB	118

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Η αρχιτεκτονική ενός απλού ΤΝΔ με ένα επίπεδο εισόδου, δύο κρυφά επίπεδα και επίπεδο εξόδου (Alam et al., 2022).....	9
Σχήμα 2: Ανθρώπινο νευρικό σύστημα (Haykin and Haykin, 2009)	12
Σχήμα 3: Σχήμα ενός πυραμιδικού κυττάρου (Haykin and Haykin, 2009)	13
Σχήμα 4: Ιεραρχία επιπέδων οργάνωσης (Haykin and Haykin, 2009)	14
Σχήμα 5: Μη γραμμικό μοντέλο ενός νευρώνα (Haykin and Haykin, 2009)	14
Σχήμα 6: Μετασχηματισμός συγγένειας που παράγεται από την παρουσία bias, $v_k=b_k$ στο $u_k=0$ (Haykin and Haykin, 2009).....	16
Σχήμα 7: Ένα άλλο μη γραμμικό μοντέλο ενός νευρώνα. Το w_{k0} αντιπροσωπεύει την παράμετρο bias b_k (Haykin and Haykin, 2009).....	17
Σχήμα 8: Κάτοψη της Εγκατάστασης Επεξεργασίας Νερού Αποσελέμη.....	23
Σχήμα 9: Διάγραμμα Ροής παραγωγικής διαδικασίας EEN Αποσελέμη.....	24
Σχήμα 10: Θολότητα ανεπεξέργαστου νερού	32
Σχήμα 11: pH ανεπεξέργαστου νερού	32
Σχήμα 12: Διάγραμμα ροής κώδικα στη Matlab για την ανάπτυξη μοντέλων ΤΝΔ	45
Σχήμα 13: Κανονική κατανομή τιμών τυποποιημένων υπολοίπων μεταβλητής παροχής αερίου χλωρίου	55
Σχήμα 14: Προβλεπόμενες τιμές (Regression Standardized) της παροχής $Cl_{2(g)}$ σε σχέση με Regression Studentized Deleted Residual	56
Σχήμα 15: Παρατηρούμενες κανονικοποιημένες τιμές των μεταβλητών (observed values) και οι τιμές των προσομοιώσεων (simulated values) των ΤΝΔ	65
Σχήμα 16: Απόδοση της προσομοίωσης των ΤΝΔ των 16 διαφορετικών περιπτώσεων του εκάστοτε σεναρίου, με βάση τον αριθμό των κόμβων	66
Σχήμα 17: Χρόνος ανάπτυξης μοντέλων ΤΝΔ, ανά αριθμό Νευρωνικών Δικτύων (ΝΔ), διατηρώντας σταθερό τον αριθμό των κόμβων ($K= 10$)	67
Σχήμα 18: Χρόνος ανάπτυξης μοντέλων ΤΝΔ, ανά αριθμό Κόμβων (Κ), διατηρώντας σταθερό τον αριθμό των Νευρωνικών Δικτύων ($ΝΔ=100$).....	68
Σχήμα 19: Δομή βέλτιστου επιλεγμένου ΤΝΔ	69
Σχήμα 20: Σχηματική απεικόνιση δομής βέλτιστου επιλεγμένου μοντέλου ΤΝΔ.....	70
Σχήμα 21: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του O_3	71
Σχήμα 22: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του AN PE	71
Σχήμα 23: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του PACI.....	72
Σχήμα 24: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του $Cl_{2(g)}$	72
Σχήμα 25: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του O_3	73
Σχήμα 26: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του AN PE	73

Σχήμα 27: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του PACI.....	74
Σχήμα 28: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) κανονικοποιημένες τιμές του $Cl_{2(g)}$	74
Σχήμα 29: Διάγραμμα Scree Plot.....	91
Σχήμα 30: Αναπαράσταση των μεταβλητών στο τρισδιάστατο σύστημα αξόνων.....	95

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Χωροθέτηση έργων Αποσελέμη	20
Εικόνα 2: Χάρτης αποτύπωσης θέσης φράγματος, ταμιευτήρα, υδραγωγείου και EEN Αποσελέμη	21
Εικόνα 3: Αποψη του φράγματος και του ταμιευτήρα Αποσελέμη (1/2)	22
Εικόνα 4: Αποψη του φράγματος και του ταμιευτήρα Αποσελέμη (2/2)	22
Εικόνα 5: Εγκατάσταση Επεξεργασίας Νερού Αποσελέμη.....	23

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Πίνακας συντομεύσεων	xv
Πίνακας 2: Κύριες στατιστικές παράμετροι μεταβλητών.....	33
Πίνακας 3: Κανονικοποιημένες κύριες στατιστικές παράμετροι μεταβλητών	34
Πίνακας 4: Κατηγοριοποίηση μεταβλητών	43
Πίνακας 5: Καμβάς περιπτώσεων συνδυασμών για τις 4 μεταβλητές που πιθανά να είναι απαραίτητες στην ανάπτυξη του μοντέλου TND.....	44
Πίνακας 6: Σενάρια TND	44
Πίνακας 7: Συντελεστής πολλαπλού προσδιορισμού (R^2) ανά μελετώμενο σενάριο.....	53
Πίνακας 8: Περιγραφικά στατιστικά στοιχεία μεταβλητών σεναρίου $Cl_{2(g)}$	54
Πίνακας 9: Συσχετίσεις σεναρίου $Cl_{2(g)}$	57
Πίνακας 10: Μοντέλο MLR Analysis σεναρίου $Cl_{2(g)}$	60
Πίνακας 11: Συντελεστές μοντέλου MLR Analysis σεναρίου $Cl_{2(g)}$	61
Πίνακας 12: Πίνακας ανάλυσης διασποράς (ANOVA ^a) μοντέλου MLR Analysis σεναρίου $Cl_{2(g)}$	63
Πίνακας 13: Χρόνος ανάπτυξης μοντέλων TND, ανά αριθμό Νευρωνικών Δικτύων (ΝΔ), διατηρώντας σταθερό τον αριθμό των κόμβων ($K=10$)	66
Πίνακας 14: Χρόνος ανάπτυξης μοντέλων TND, ανά αριθμό Κόμβων (K), διατηρώντας σταθερό τον αριθμό των Νευρωνικών Δικτύων ($ΝΔ=100$).....	67
Πίνακας 15: Δείκτης καλύτερης απόδοσης δοκιμών (Best testing performance) ανά σενάριο και περίπτωση TND	69
Πίνακας 16: Τιμές παραμέτρου RMSE	77
Πίνακας 17: Τιμές συντελεστή προσδιορισμού R^2	77
Πίνακας 18: Τιμές συντελεστή συσχέτισης Pearson R.....	78
Πίνακας 19: Εφαρμογή βέλτιστου επιλεγμένου TND σε μια μελέτη περίπτωσης	78
Πίνακας 20: Αποτελέσματα εφαρμογής βέλτιστου επιλεγμένου TND σε μια μελέτη περίπτωσης	79
Πίνακας 21: KMO and Bartlett's Test	81
Πίνακας 22: Anti- image Matrices	83
Πίνακας 23: Descriptive Statistics.....	87
Πίνακας 24: Communalities	88
Πίνακας 25: Total Variance Explained.....	90
Πίνακας 26: Component Matrix	91
Πίνακας 27: Rotated Component Matrix.....	94
Πίνακας 28: Component Transformation Matrix	95

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ ΠΑΡΑΡΤΗΜΑΤΟΣ

Σχήμα Π 1: Ημερήσια διαφορά ύψους νερού στον Ταμιεντήρα	112
Σχήμα Π 2: Παροχή ανεπεξέργαστου νερού	112
Σχήμα Π 3: Θολότητα επεξεργασμένου νερού.....	113
Σχήμα Π 4: pH επεξεργασμένου νερού	113
Σχήμα Π 5: Συγκέντρωση ελεύθερου υπολειμματικού χλωρίου επεξεργασμένου νερού	113
Σχήμα Π 6: Συγκέντρωση υπολειμματικού Αργιλίου επεξεργασμένου νερού.....	114
Σχήμα Π 7: Ημερήσια καταναλισκόμενη ηλεκτρική ενέργεια EEN Αποσελέμη	114
Σχήμα Π 8: Συγκέντρωση υπολειμματικού όζοντος (O_3).....	114
Σχήμα Π 9: Δόση ανιονικού πολυηλεκτρολύτη (AN PE)	115
Σχήμα Π 10: Δόση χλωριούχου θεικού πολυ- αργιλίου (PACl).....	115
Σχήμα Π 11: Παροχή Διαλύματος χλωρίου ($Cl_{2(g)}$).....	115
Σχήμα Π 12: Θολότητα νερού εισόδου κλινών διήθησης	116

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ ΠΑΡΑΡΤΗΜΑΤΟΣ

Πίνακας Π 1: Συντελεστές συσχέτισης Pearson των μελετώμενων μεταβλητών.....	117
---	-----

ΛΙΣΤΑ ΣΥΝΤΟΜΕΥΣΕΩΝ

Πίνακας 1: Πίνακας συντομεύσεων

Σύντμηση	Πλήρης ονομασία
ΑΠΓΠ	Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης
Α.Κ.Σ	Ανάλυση Κύριων Συνιστωσών
ΔΔ	Διδακτορική Διατριβή
ΔΕΥΑ	Δημοτική Επιχείρηση Ύδρευσης- Αποχέτευσης
ΕΕ	Ευρωπαϊκή Ένωση
ΕΚ	Ευρωπαϊκό Κοινοβούλιο
ΕΕΝ	Εγκατάσταση Επεξεργασίας Νερού
ΚΥΑ	Κοινή Υπουργική Απόφαση
ΝΔ	Νευρωνικό Δίκτυο
ΟΑΚ	Οργανισμός Ανάπτυξης Κρήτης
Π.Α.	Παραγοντική Ανάλυση
ΠΓΠ	Πολλαπλή Γραμμική Παλινδρόμηση
ΤΝ	Τεχνητή Νοημοσύνη
ΤΝΔ	Τεχνητό Νευρωνικό Δίκτυο
AI	Artificial Intelligence
AN PE	Anionic Polyelectrolyte
ANN	Artificial Neural Network
CFN	Cascade Forward Network
EFA	Exploratory Factor Analysis
FA	Factor Analysis
GRNN	General Regression Neural Network
ISO	International Organization for the Standardization
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAX	Maximum
MIN	Minimum
ML	Machine Learning
MSA	Measure of Sampling Adequacy
MSE	Mean Square Error
MLR	Multiple Linear Regression
MLP	Multi- Layer Perceptron
N	Node
NFTOOL	Neural Fitting Tool
NN	Neural Network

Σύντμηση	Πλήρης ονομασία
No	Number
NSE	Nash- Sutcliffe
PAC	Powdered Activated Carbon
PACl	Poly- Aluminum chloride hydroxide sulfate
PCA	Principal Components Analysis
R	Pearson Correlation Coefficient
R ²	Coefficient of Determination
RBF	Radial Basis Function
RMSE	Root-Mean-Square Error
SCADA	Supervisory Control and Data Acquisition
SPSS	Statistical Package for the Social Sciences
SOM	Self- Organizing Map
SSE	Sum of Squared Error
STD	Standard Deviation
TOC	Total Organic Carbon
VIF	Variance Inflation Factor
VLSI	Very Large- Scale Integration

ΠΕΡΙΛΗΨΗ

Η παρακολούθηση των κύριων λειτουργικών μεταβλητών μιας EEN και των μεταβλητών ποιότητας του νερού είναι κρίσιμο ζήτημα για όλους τους λειτουργούς εγκαταστάσεων επεξεργασίας νερού (EEN) για την παραγωγή πόσιμου (ανθρώπινης κατανάλωσης) νερού. Οι λειτουργοί των EEN αναζητούν συχνά μια γρήγορη, εύχρηστη και αξιόπιστη μέθοδο για την πρόβλεψη των ημερήσιων δόσεων των χρησιμοποιούμενων χημικών επεξεργασίας του νερού, που είναι το κύριο και καθημερινό μέλημά τους.

Η παρούσα Διδακτορική Διατριβή επικεντρώνεται στη μοντελοποίηση της λειτουργίας μιας EEN με χρήση Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) και ανάλυσης Πολλαπλής Γραμμικής Παλινδρόμησης (Multiple Linear Regression, MLR), τη συσχέτιση των ποιοτικών χαρακτηριστικών του νερού με τις κύριες λειτουργικές μεταβλητές μιας EEN, την αλληλεπίδραση των παραπάνω μεταβλητών μεταξύ τους, και κυρίως την πρόβλεψη των δόσεων των απαραίτητων χημικών επεξεργασίας του νερού. Η σημασία και η καινοτομία αυτής της μελέτης έγκειται στο γεγονός ότι λαμβάνει υπόψη τη μακρόχρονη εμπειρία του λειτουργού μιας EEN, πράγμα που σε μεγάλο βαθμό απουσίαζε από την υπάρχουσα βιβλιογραφία.

Στην παρούσα ΔΔ, η εξεταζόμενη μελέτη περίπτωσης προέρχεται από την EEN Αποσελέμη, κατά την οποία πραγματοποιείται μια πρόβλεψη των μεταβλητών εξόδου

μοντέλων TND και MLR, αναφορικά με τις απαιτούμενες δόσεις χημικών επεξεργασίας του νερού, με βάση την ποιότητα του νερού που παρατηρήθηκε και άλλες λειτουργικές μεταβλητές. Οι εκτιμώμενες κύριες λειτουργικές μεταβλητές της EEN περιλαμβάνουν: τις δόσεις υπολειμματικού όζοντος (O_3), ανιονικού πολυηλεκτρολύτη (ANPE), χλωριούχου θεικού πολυ-αργιλίου (PACI) και αερίου χλωρίου ($Cl_{2(g)}$). Ως παράμετροι εισόδου για το TND χρησιμοποιήθηκαν ημερήσια αποτελέσματα αναλύσεων δειγμάτων νερού και καταγραφές από το Supervisory Control and Data Acquisition System (SCADA) της EEN, που καλύπτουν περίοδο 38 μηνών (1.188 τιμές για καθεμία από τις 14 μετρήσιμες μεταβλητές). Συγκεκριμένα, οι παράμετροι εισόδου του TND περιλαμβάνουν: την παροχή ακατέργαστου νερού (Q), τη θολότητα ακατέργαστου νερού (T_1), τη θολότητα επεξεργασμένου νερού (T_2), το υπολειμματικό ελεύθερο χλώριο επεξεργασμένου νερού (Cl_2), τη συγκέντρωση υπολειμματικού αργιλίου επεξεργασμένου νερού (Al), τη θολότητα του νερού στην είσοδο των κλινών διήθησης (T_3), την ημερήσια διαφορά ύψους νερού στον ταμιευτήρα του φράγματος Αποσελέμη (ΔH), την τιμή pH ακατέργαστου νερού (pH_1), την τιμή pH επεξεργασμένου νερού (pH_2) και την ημερήσια κατανάλωση της ηλεκτρικής ενέργειας (EI) στην EEN Αποσελέμη. Οι παράμετροι εξόδου του TND ήταν: η συγκέντρωση του υπολειμματικού όζοντος (O_3) μετά τη διεργασία της οζόνωσης, η δόση του ανιονικού πολυηλεκτρολύτη (ANPE), η δόση του χλωριούχου θεικού πολυ-αργιλίου (PACI) και η παροχή αερίου χλωρίου ($Cl_{2(g)}$).

Συνολικά δοκιμάστηκαν 304 διαφορετικά μοντέλα TND και με βάση την καλύτερη τιμή του δείκτη της απόδοσης της δοκιμής (test performance, $tperf$) των TND, επιλέχθηκε τελικά το σενάριο με 100 νευρωνικά δίκτυα, 100 κόμβους, 42 κρυφούς κόμβους, 10 εισόδους και 4 εξόδους. Το συγκεκριμένο μοντέλο TND πέτυχε πολύ καλά αποτελέσματα προσομοίωσης ($best\ tper = 0.008848$), γεγονός που υποδηλώνει ότι τα TND είναι δυνητικά χρήσιμα εργαλεία για την πρόβλεψη των κύριων λειτουργικών μεταβλητών μιας EEN.

Επίσης, εξετάστηκαν 4 διαφορετικά σενάρια με Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης (MLR) με εξαρτημένες μεταβλητές: το υπολειμματικό όζον (O_3), τη δόση ανιονικού πολυηλεκτρολύτη (ANPE), τη δόση του θεικού χλωριούχου πολυ-

Αργιλίου (PACl) και την παροχή του αερίου χλωρίου ($\text{Cl}_{2(g)}$), καθώς και δέκα (10) ανεξάρτητες μεταβλητές λειτουργίας και ποιότητας νερού.

Σύμφωνα με τα αποτελέσματα του R^2 (Coefficient of Determination) και R (Pearson Correlation Coefficient), το μοντέλο TND είχε καλύτερη απόδοση σε σύγκριση με το μοντέλο ανάλυσης MLR για την πρόβλεψη των δόσεων των χημικών και των τεσσάρων χημικών επεξεργασίας του νερού. Με βάση το κριτήριο $R^2 > 0,5$, η απόδοση του TND ήταν ικανοποιητική στην πρόβλεψη των δόσεων τριών χημικών επεξεργασίας του νερού: ANPE ($R^2 = 0,772$), PACl ($R^2 = 0,742$) και $\text{Cl}_{2(g)}$ ($R^2 = 0,838$, +23% σε σύγκριση με αντίστοιχη τιμή του MLR μοντέλου και $R = 0,95$, +11% σε σύγκριση με αντίστοιχη τιμή του MLR μοντέλου). Αντίστοιχα, η πρόβλεψη του μοντέλου MLR, αξιολογήθηκε ως ικανοποιητική για την πρόβλεψη της δόσης μόνο του $\text{Cl}_{2(g)}$ ($R^2 = 0,681$, $R = 0,82500$).

Σύμφωνα με τα αποτελέσματα του RMSE, το μοντέλο MLR είχε καλύτερη απόδοση για τα τρία ($\text{RMSE}_{\text{ANPE}} = 0.05 \text{ mg/L}$, $\text{RMSE}_{\text{PACl}} = 0.08 \text{ mg/L}$ και $\text{RMSE}_{\text{Cl}_{2(g)}} = 0.10 \text{ kg/h}$) από τα τέσσερα χρησιμοποιημένα χημικά επεξεργασίας του νερού, σε σύγκριση με το μοντέλο TND, το οποίο είχε καλύτερη απόδοση μόνο για ένα χημικό επεξεργασίας του νερού ($\text{RMSE}_{\text{O}_3} = 0.02 \text{ mg/L}$).

Γενικά, εάν κάποιος θέλει να χρησιμοποιήσει τα σενάρια και τα μοντέλα πρόβλεψης (TND ή MLR) για να προβλέψει τις δόσεις $\text{Cl}_{2(g)}$, τότε είναι προτιμότερο να χρησιμοποιήσει αυτό με το μικρότερο RMSE. Εάν ενδιαφέρεται κάποιος να έχει μια μελλοντική πρόβλεψη δόσεων χημικών επεξεργασίας του νερού με μοντέλο πρόβλεψης με την καλύτερη προσαρμογή, τότε είναι προτιμότερο να χρησιμοποιηθεί αυτό το μοντέλο με την μεγαλύτερη τιμή του R^2 . Επίσης, η μεταβλητή δόσης του όζοντος (O_3) παρουσίασε χαμηλές τιμές του R^2 , σε όλες τις περιπτώσεις, πιθανώς λόγω της μεγάλης διακύμανσης των τιμών του.

Αυτή η ΔΔ ενισχύει περαιτέρω την άποψη ότι τα TND είναι χρήσιμα εργαλεία υποστήριξης λήψης αποφάσεων για έναν λειτουργό μιας EEN πόσιμου νερού, τα οποία δύναται να μιμηθούν με μεγάλη ακρίβεια και επαρκώς τις αποφάσεις σχετικά με τις χρησιμοποιούμενες δόσεις των χημικών επεξεργασίας του νερού, που είναι το κύριο και καθημερινό μέλημα του λειτουργού μια τέτοιου είδους εγκατάστασης.

Συνιστάται η εκπόνηση αντίστοιχων μελλοντικών ερευνών, για την περαιτέρω αύξηση

της γνώσης σχετικά με την πρόβλεψη των χημικών επεξεργασίας του νερού, χρησιμοποιώντας μοντέλα, όπως τα ΤΝΔ, ως ακριβή μοντέλα πρόβλεψης αλλά και μοντέλα ανάλυσης MLR, ως ευέλικτα, γρήγορα και αξιόπιστα μοντέλα πρόβλεψης. Συγκεκριμένα, περαιτέρω έρευνα θα μπορούσε να διεξαχθεί στην πρόβλεψη των χημικών επεξεργασίας του νερού σε μια EEN, χρησιμοποιώντας ΤΝΔ με μικρότερο αριθμό μεταβλητών για εξασφάλιση μεγαλύτερης ευελιξίας, χωρίς να μειώνεται, ωστόσο, ουσιαστικά η αξιοπιστία του μοντέλου πρόβλεψης.

Επίσης, προτείνεται να καταβληθεί ακόμη μεγαλύτερη προσπάθεια για την καθιέρωση των ΤΝΔ ως μοντέλων πρόβλεψης στον τομέα του νερού και στην καθημερινή λειτουργία των EEN. Επιπρόσθετα, οι μελλοντικές έρευνες θα μπορούσαν να περιλαμβάνουν τη διερεύνηση της χρήσης και άλλων κριτηρίων σύγκρισης, όπως το MAE (Mean Absolute Error, μέσο απόλυτο σφάλμα), το MAPE (Mean Absolute Percentage Error, μέσο απόλυτο ποσοστό σφάλματος) ή τον δείκτη απόδοσης NSE (Nash- Sutcliffe). Επιπλέον, προτείνεται η διεξαγωγή αναλύσεων ευαισθησίας και αβεβαιότητας στις μεταβλητές με τη μεγαλύτερη επιρροή, οι οποίες θα μπορούσαν να βελτιώσουν περαιτέρω τη διαδικασία μοντελοποίησης. Τέλος, δεδομένου ότι ο κύριος περιορισμός της τρέχουσας ΔΔ είναι ότι τα μοντέλα έχουν εκπαιδευτεί με δεδομένα από μία μόνο EEN, προτείνεται ως μελλοντική εργασία την συμπερίληψη δεδομένων από περισσότερες αντίστοιχες EEN, προκειμένου να αυξηθεί η στιβαρότητα των μοντέλων και η καθολική εφαρμογή τους.

ABSTRACT

The monitoring of the main operational variables and water quality characteristics of a Water Treatment Plan (WTP) is a critical issue for all WTP operators, for the production of human consumption (drinking) water. WTP operators often seek a quick, reliable and easy-to-use method for predicting the dosages of the water treatment chemicals used, which is their main daily concern.

This PhD thesis focuses on the modeling of a WTP operation, using ANN and MLR analysis models. Also, examines the water quality characteristics and the main WTP operational variables correlation, their interaction and mainly focuses on the prediction of the necessary water treatment chemicals dosages in a WTP. The innovation of this study lies in the fact that it takes into account the extended experience of a WTP operator.

In this Thesis, the studied case study comes from the Aposelemis WTP, in which a prediction of the output variables of the ANN and MLR models is made, regarding the required dosages of the water treatment chemicals, based on the observed water quality and other operational variables. The estimated main operational variables of the WTP include: the dosages of residual ozone (O_3), anionic polyelectrolyte (ANPE), polyaluminum chloride sulfate (PACl) and chlorine gas ($Cl_{2(g)}$). Daily water sample analysis results and recordings from the WTP SCADA, covering a period of 38 months

(1,188 values for each of the 14 measurable variables), were used as input parameters for the ANN modelling. Specifically, the input parameters of the ANN model include: the raw water flow (Q), the raw water turbidity (T_1), the treated water turbidity (T_2), the residual free chlorine of the treated water (Cl_2), the residual aluminum concentration of the treated water (Al), the water turbidity at the inlet of the filtration beds (T_3), the daily water height difference in the reservoir of the Aposelemis dam (ΔH), the raw water pH value (pH_1), the treated water pH value (pH_2) and the daily electricity consumption (El) at the Aposelemis WTP. The output parameters of the ANN include: the concentration of residual ozone (O_3) after the ozonation process, the dosage of anionic polyelectrolyte (ANPE), the dosage of polyaluminum chloride sulfate (PACl) and the supply of chlorine gas ($Cl_{2(g)}$).

A total of 304 different ANN models were constructed and based on the best value of the test performance index (tperf) of them, the scenario with 100 neural networks, 100 nodes, 42 hidden nodes, 10 inputs and 4 outputs was finally selected. This ANN model achieved very good simulation results, which suggests that ANNs are potentially useful tools for predicting the main WTP operational variables.

Also, four (4) different scenarios were examined using Multiple Linear Regression Analysis (MLR) with dependent variables: the residual ozone (O_3), anionic polyelectrolyte dosage (ANPE), poly-Aluminum chloride sulfate (PACl) dosage and chlorine gas flow ($Cl_{2(g)}$), as well as were used ten (10) independent operational and water quality variables.

According to the results of R^2 and R , the ANN model had a better performance compared to the MLR analysis model for predicting the dosages of the used water treatment chemicals. Based on the criterion $R^2 > 0.5$, the ANN performance was satisfactory in predicting the dosages of the three water treatment chemicals: ANPE ($R^2 = 0.772$), PACl ($R^2 = 0.742$) and $Cl_{2(g)}$ ($R^2 = 0.838$, +23% compared to the corresponding value of the MLR model and $R = 0.95$, +11% compared to the corresponding value of the MLR model). Accordingly, the prediction of the MLR model was evaluated as satisfactory for predicting the dosage of $Cl_{2(g)}$ only ($R^2 = 0.681$, $R = 0.82500$).

According to the RMSE results, the MLR model performed better for three

($RMSE_{ANPE} = 0.05$ mg/L, $RMSE_{PACl} = 0.08$ mg/L and $RMSE_{Cl2(g)} = 0.10$ kg/h) of the four dependant variables (drinking water added chemicals), than the ANN model, which performed better for only one water treatment chemical ($RMSE_{O3} = 0.02$ mg/L).

In general, if someone wants to use the scenarios and prediction models (ANN or MLR) to predict $Cl_{2(g)}$ dosages, then it is preferable to use the one with the smallest RMSE. If one is interested in having a future prediction of water treatment chemical dosages with a prediction model with the best fit, then it is preferable to use the model with the largest R^2 value. Also, the ozone dosage variable (O_3) presented low R^2 values, in all cases, probably due to the large variation of its values.

This study further reinforces the point of view that ANNs are useful decision support tools for a WTP operator, which can simulate with great accuracy and adequacy the decisions regarding the dosages of the water treatment chemicals used, which is the main and daily concern of the operator of such a facility.

It is recommended future research to be conducted to further increase knowledge on the prediction of water treatment chemicals, using models such as ANNs, as accurate prediction models, and MLR analysis models as flexible, fast and reliable prediction models. In particular, further research could be conducted on the prediction of chemicals used in a WTP, using ANNs with a smaller number of variables to ensure greater flexibility, without, however, substantially reducing the reliability of the prediction model.

This will enable to establish ANNs as forecasting models in the water sector and in the daily operation of the WTPs. In addition, future research could include investigating the use of other comparison criteria, such as MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) or the NSE (Nash-Sutcliffe) performance index. Furthermore, it is suggested to conduct sensitivity and uncertainty analyses on the most influential variables, which could further improve the modeling process. Finally, since the main limitation of the current study is that the models have been trained with data from a single WTP, it is suggested as future work to include data from more corresponding WTPs, in order to increase the robustness of the models and their universal applicability.

1 ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγικά στοιχεία

Ο 21^{ος} αιώνας αντιμετωπίζει παγκοσμίως αυξανόμενη εξάντληση και ρύπανση των υδάτινων πόρων, λόγω της πληθυσμιακής αύξησης, της έντονης εκβιομηχάνισης αλλά και της κλιματικής αλλαγής. Η μεγάλη σημασία που έχει η ασφαλής κατανάλωση και η βιώσιμη διαχείριση των υδάτινων πόρων, ασκεί πιέσεις στις EEN να αναπτύσσουν προηγμένες τεχνολογίες για την τήρηση των υψηλότερων προτύπων ποιότητας νερού. Ωστόσο, η πολυπλοκότητα των διεργασιών επεξεργασίας νερού απαιτεί καινοτόμες μεθόδους για τον έλεγχο της ποιότητας και τη βελτιστοποίηση τους (Lin et al., 2023a). Στις μέρες μας, η τέταρτη βιομηχανική επανάσταση φέρνει αλλαγές στη βιομηχανία νερού, απαιτώντας έξυπνη τεχνολογία και συστήματα νερού. Τα έξυπνα συστήματα νερού που βασίζονται σε Τεχνητή Νοημοσύνη (TN) (Artificial Intelligence, AI) αναμένεται να παρέχουν καλύτερα συστήματα ελέγχου και βελτιστοποίηση της αποδοτικότητας των διεργασιών επεξεργασίας νερού, όπως και της σχέσης κόστους-αποτελεσματικότητας (Lowe et al., 2022).

Οι λειτουργοί των EEN αναζητούν συχνά μια γρήγορη, εύχρηστη και αξιόπιστη μέθοδο για την πρόβλεψη των ημερήσιων δόσεων των χρησιμοποιούμενων χημικών επεξεργασίας του νερού. Η παρούσα ΔΔ εστιάζει στην κάλυψη αυτής της ανάγκης χρησιμοποιώντας τα ΤΝΔ, ως ένα χρήσιμο και αξιόπιστο εργαλείο για τους λειτουργούς

EEN. Αυτή η μελέτη χρησιμοποιεί ιστορικά δεδομένα σχετικά με την ποιότητα και τις λειτουργικές μεταβλητές μιας εγκατάστασης επεξεργασίας επιφανειακών υδάτων, η οποία παράγει νερό για ανθρώπινη κατανάλωση (πόσιμο νερό). Η σημασία και η καινοτομία αυτής της μελέτης έγκειται στο γεγονός ότι λαμβάνει υπόψη την εμπειρία του λειτουργού μιας EEN, πράγμα που σε μεγάλο βαθμό απουσίαζε από την υπάρχουσα βιβλιογραφία.

Η TN αναγνωρίζεται ως ένα ισχυρό εργαλείο για την επίλυση πολλών βιομηχανικών λειτουργικών προβλημάτων και έχει εφαρμοστεί σε διάφορους τομείς, όπως οι μεταφορές, η οικονομική διαχείριση και η υγειονομική περίθαλψη (Alam et al., 2022; Bhattacharya et al., 2021; Wongburi and Park, 2023). Η TN έχει επίσης βρει εφαρμογές στον τομέα της περιβαλλοντικής παρακολούθησης, όπως στην πρόβλεψη των βροχοπτώσεων και την παρακολούθηση της επεξεργασίας του νερού ή των υγρών αποβλήτων (Chen et al., 2020; Maloney et al., 2022; Nasir et al., 2022; Papailiou et al., 2022; Stylianoudaki et al., 2022; Wongburi and Park, 2023). Πρόσφατα, η ονομαζόμενη Μηχανική Μάθηση (Machine Learning, ML), η οποία είναι ένα μέρος της τεχνητής νοημοσύνης, έχει επίσης χρησιμοποιηθεί σε διαδικασίες διαχείρισης και επεξεργασίας των υδάτινων πόρων (Dunnington et al., 2021; Huang et al., 2021; Kim et al., 2023; Li et al., 2021a; Lowe et al., 2022).

Ένα αξιόπιστο μοντέλο για την πρόβλεψη των μεταβλητών λειτουργίας μιας EEN είναι απαραίτητο για τον έλεγχο της λειτουργίας της μονάδας και τη διασφάλιση της παροχής ασφαλούς πόσιμου νερού για τους καταναλωτές. Οι μεταβλητές ποιότητας του νερού, όπως η θολότητα, το pH και η θερμοκρασία παρακολουθούνται συχνά και υπάρχει σημαντική συσχέτιση μεταξύ των προαναφερθέντων μεταβλητών και των ποσοτήτων κροκιδωτικών και λοιπών ουσιών, που χρησιμοποιούνται στις διαδικασίες επεξεργασίας του νερού (Dadebo et al., 2023; Ghaedi and Vafaei, 2017; Kim et al., 2023; Wu and Lo, 2010, 2008).

Τα ΤΝΔ, ως κλάδος της τεχνητής νοημοσύνης (Alprol et al., 2024) είναι μοντέλα σχεδιασμένα να μοιάζουν με τις λειτουργίες του εγκεφάλου (Alprol et al., 2024; Haykin and Haykin, 2009). Τα ΤΝΔ είναι πολύ χρήσιμα εργαλεία με υψηλή απόδοση στην σύνθετη αντιστοίχιση σχέσεων και προβλέψεων στη ρύθμιση της λειτουργίας μιας EEN.

Τα ΤΝΔ μπορούν να επεξεργαστούν μη γραμμικά δεδομένα που είναι πολύπλοκα και δύσκολο να προσομοιωθούν με απλά μαθηματικά μοντέλα (Dadebo et al., 2023; Li et al., 2021a; Wu and Lo, 2010). Τα ΤΝΔ βασίζονται σε μηχανισμούς μάθησης, εκπαίδευσης και ελέγχου και δεν προγραμματίζονται όπως τα συμβατικά προγράμματα υπολογιστών. Η εκπαίδευση ενός ΤΝΔ επιτυγχάνεται με την προσθήκη των παραπάνω συνδέσεων μέσω ενός αλγόριθμου εκπαίδευσης. Η μοντελοποίηση με χρήση ΤΝΔ επιτυγχάνεται με βάση τα ακόλουθα κύρια στάδια: 1. συλλογή των δεδομένων, 2. ανάλυση των δεδομένων και 3. εκπαίδευση των νευρωνικών δικτύων.

Ένα νευρωνικό δίκτυο (ΝΔ) είναι ικανό να μαθαίνει και επομένως να γενικεύει. Αυτοί είναι οι δύο κύριοι λόγοι που ένα ΝΔ μπορεί να κάνει ακριβείς προβλέψεις. Η γενίκευση είναι η παραγωγή λογικών δεδομένων εξόδου για δεδομένα εισόδου που δεν χρησιμοποιούνται όσο διαρκεί η εκπαίδευση του ΤΝΔ. Αυτός είναι και ο λόγος που ένα ΤΝΔ είναι σε θέση να βρει κατάλληλες λύσεις προσομοίωσης σε πολύπλοκα, πλούσια σε δεδομένα και διαφορετικά δυσεπίλυτα προβλήματα (Haykin and Haykin, 2009).

Οι πιο συχνά χρησιμοποιούμενοι τύποι ΤΝΔ, είναι: τα MLPs (Multi-layer Perceptrons), τα RBF (Radial Basis Function), τα GRNNs (General Regression Neural Networks), τα CFNs (Cascade Forward Networks) και τα Kohonen's SOM (Self-Organizing Maps) (Farmaki et al., 2013; Haykin and Haykin, 2009; O'Reilly et al., 2018; Wu et al., 2014). Αρκετές μελέτες έχουν δημοσιευθεί σχετικά με τη μοντελοποίηση και τη βελτιστοποίηση των διαδικασιών επεξεργασίας νερού με τη χρήση ΤΝΔ. Τα προαναφερόμενα MLP ΤΝΔ έχουν χρησιμοποιηθεί για παλινδρόμηση για τον προσδιορισμό των δόσεων κροκιδωτικών και της θολότητας του επεξεργασμένου νερού (Griffiths and Andrews, 2011; Kim and Parnichkun, 2017a). Μια άλλη προσέγγιση, χρησιμοποιώντας μηχανική όραση, περιλάμβανε τη χρήση ενός νευρωνικού δικτύου για την πρόβλεψη των δόσεων κροκιδωτικών αναλύοντας εικόνες κροκίδωσης (Yamamura et al., 2020).

Οι τιμές των παραμέτρων του R^2 , του MSE (Mean Square Error), του SSE (Sum of Squared Error) και του RMSE (Root- Mean Square Error) χρησιμοποιούνται συνήθως στην επικύρωση των αποτελεσμάτων πρόβλεψης των μοντέλων. Συνήθως

παρατηρούνται μικρές αποκλίσεις μεταξύ των προβλέψεων των μοντέλων και των πραγματικών μετρήσεων (Alam et al., 2022; Tabari and Hosseinzadeh Talaee, 2015).

Σύμφωνα με προηγούμενες έρευνες (Alam et al., 2022; Dadebo et al., 2023; Tabari and Hosseinzadeh Talaee, 2015; Wu and Lo, 2010; Yateh et al., 2023), τα πλεονεκτήματα από τη μοντελοποίησης με χρήση ΤΝΔ στον τομέα του νερού περιλαμβάνουν μεταξύ των άλλων:

- (i) την έλλειψη αλγορίθμου που απαιτείται για την κατασκευή του μοντέλου ΤΝΔ, καθιστώντας τη μοντελοποίηση μια γρήγορη και ευέλικτη διαδικασία,
- (ii) την ικανότητα χειρισμού μη γραμμικών σχέσεων με ευκολία,
- (iii) την ενσωμάτωση της εμπειρίας και της γνώσης του λειτουργού της εγκατάστασης στην κατασκευή του μοντέλου,
- (iv) τη βελτιστοποίηση των διαδικασιών επεξεργασίας νερού,
- (v) τις πρακτικές λύσεις σε θέματα ρύπανσης των υδάτων,
- (vi) τη μείωση των λειτουργικών δαπανών, μέσω της βελτιστοποίησης της χρήσης των χημικών επεξεργασίας του νερού και
- (vii) την έγκαιρη δημιουργία αποτελεσμάτων μοντελοποίησης και πρόβλεψης.

Οι περιορισμοί της μοντελοποίησης με χρήση ΤΝΔ στον τομέα του νερού, μεταξύ των άλλων, είναι:

- (i) η διαθεσιμότητα των δεδομένων,
- (ii) η κακή αναπαραγωγικότητα των δεδομένων,
- (iii) η ανάγκη για επαρκή δεδομένα για εκπαίδευση (training) και δοκιμές (testing),
- (iv) η εξάρτηση της απόδοσης πρόβλεψης του μοντέλου από συγκεκριμένες συνθήκες (π.χ. μεγάλη αβεβαιότητα σε ξαφνικές αλλαγές),
- (v) τα μειονεκτήματα που σχετίζονται με την τυχαία επιλογή δεδομένων και
- (vi) οι υψηλές υπολογιστικές απαιτήσεις (Alam et al., 2022; Alprol et al., 2024; Zhao et al., 2020).

1.2 Η Τεχνητή Νοημοσύνη (Artificial Intelligence)

Η τεχνητή νοημοσύνη έχει αναπτυχθεί ως ένα πολύ χρήσιμο εργαλείο για πρακτικά προβλήματα και έχει τύχει μεγάλης προσοχής για τις εφαρμογές της σε διάφορους τομείς όπως π.χ υγειονομική περίθαλψη, μεταφορές και οικονομικής διαχείρισης (Alam et al., 2022). Η τεχνητή νοημοσύνη έχει υιοθετηθεί με επιτυχία και στον τομέα της περιβαλλοντικής παρακολούθησης, όπως για πρόβλεψη των βροχοπτώσεων και για την παρακολούθηση της επεξεργασίας του νερού (Chen et al., 2020; Maloney et al., 2022; Nasir et al., 2022). Πρόσφατα, ένας κλάδος της τεχνητής νοημοσύνης, η μηχανική μάθηση, έχει επεκταθεί και στις διεργασίες επεξεργασίας νερού (Dunnington et al., 2021; Huang et al., 2021; Li et al., 2021a; Lowe et al., 2022).

Η τεχνητή νοημοσύνη έχει αποδείξει όλο και περισσότερο τις δυνατότητες που έχει να επιλύει τις προκλήσεις που αντιμετωπίζει η επεξεργασία του πόσιμου νερού, λόγω της ισχυρής αυτόνομης μάθησης και της ικανότητάς της να αντιμετωπίζει σύνθετα προβλήματα. Η τεχνολογία της τεχνητής νοημοσύνης προσφέρει τεχνική υποστήριξη για τη διαχείριση και τη λειτουργία των διαδικασιών της επεξεργασίας του νερού, κάτι που είναι περισσότερο αποτελεσματικό από το να βασίζεται αποκλειστικά σε ανθρώπινες λειτουργίες και αποφάσεις. Η ανάλυση δεδομένων με βάση την τεχνητή νοημοσύνη και οι μηχανισμοί εξελικτικής μάθησης είναι σε θέση να πραγματοποιούν διάγνωση της ποιότητας νερού, αυτόνομη λήψη αποφάσεων και βελτιστοποίηση των διεργασιών επεξεργασίας του νερού και έχουν τη δυνατότητα να δημιουργήσουν μια διεθνής πλατφόρμα ανάλυσης διεργασιών και προγνωστικών μοντέλων επεξεργασίας του νερού (Li et al., 2021a).

1.3 Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks, ANNs)

Τα Νευρωνικά Δίκτυα (ΝΔ), ή ακριβέστερα τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), είναι κλάδος της τεχνητής νοημοσύνης (Gardner and Dorling, 1998). Τα ΤΝΔ ξεκίνησαν από την αναγνώριση ότι ο εγκέφαλος ενός ανθρώπου κάνει υπολογισμούς με έναν τελείως διαφορετικό τρόπο από τον ψηφιακό συμβατικό υπολογιστή.

Ο ανθρώπινος εγκέφαλος είναι ένας εξαιρετικά πολύπλοκο και μη γραμμικό

σύστημα επεξεργασίας πληροφοριών. Ο ανθρώπινος εγκέφαλος οργανώνει τα δομικά του συστατικά, τα οποία είναι γνωστά ως νευρώνες, για να εκτελεί ορισμένους υπολογισμούς (π.χ. να αναγνωρίζει πρότυπα, να έχει αντίληψη και να ασκεί έλεγχο) πολλές φορές με μεγαλύτερη ταχύτητα από τον ταχύτερο υπολογιστή των ημερών μας.

Όταν γεννιέται ένας άνθρωπος, ο εγκέφαλός του ήδη έχει μεγάλο μέρος της δομής του και την ικανότητα να χτίζει τους δικούς του κανόνες συμπεριφοράς μέσω της «εμπειρίας». Η εμπειρία αυξάνεται όσο περνάει ο χρόνος και το μεγαλύτερο μέρος του ανθρώπινου εγκεφάλου δημιουργείται κατά τα δύο πρώτα έτη από τη γέννηση και η ανάπτυξη συνεχίζεται πολύ πιο πέρα από το στάδιο αυτό (Haykin and Haykin, 2009).

Ένα νευρικό σύστημα που αναπτύσσεται προσομοιάζεται με έναν εγκέφαλο με πλαστικότητα (*plastic brain*). Η πλαστικότητα (*plasticity*) δίνει τη δυνατότητα, στο νευρικό σύστημα που αναπτύσσεται, να έχει πολύ καλή προσαρμογή στο περιβάλλον του. Όπως συμβαίνει με την πλαστικότητα, η οποία είναι απαραίτητη για τη λειτουργία των εγκεφαλικών νευρώνων, αντίστοιχο συμβαίνει και με τους τεχνητούς νευρώνες που συνθέτουν τα ΤΝΔ.

Γενικά, θα μπορούσαμε να πούμε ότι το νευρωνικό δίκτυο ότι είναι μια μηχανή, η οποία μοντελοποιεί τον τρόπο που ο ανθρώπινος εγκέφαλος λειτουργεί για να εκτελέσει μια συγκεκριμένη εργασία ή λειτουργία. Εάν θέλουμε να δώσουμε έναν γενικό ορισμό, θα λέγαμε ότι:

- ✓ Ένα ΤΝΔ είναι ένας επεξεργαστής που αποτελείται από απλές μονάδες επεξεργασίας και αποθηκεύει τη γνώση, που με βιωματικό τρόπο αποκτιέται, έτσι ώστε να μπορεί να τη χρησιμοποιεί. Προσομοιάζει τον ανθρώπινο εγκέφαλο γιατί:

1. Η βιωματική γνώση του ΤΝΔ αποκτιέται από το περιβάλλον του μέσω μιας διαδικασίας μάθησης.
2. Τα συναπτικά βάρη, τα οποία είναι οι δυνάμεις σύνδεσης των ενδο-νευρώνων, χρησιμοποιούνται για την αποθήκευση της βιωματικής γνώσης που αποκτήθηκε (Haykin and Haykin, 2009).

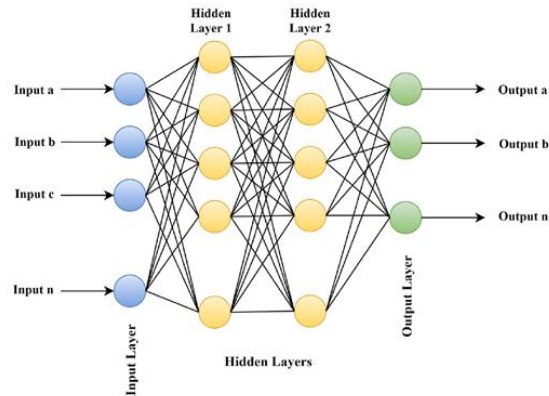
Αλγόριθμος μάθησης, ονομάζεται η διαδικασία που χρησιμοποιείται για τη μάθηση και τροποποιεί κατάλληλα τα συναπτικά βάρη του ΤΝΔ, έτσι ώστε να επιτευχθεί ο επιθυμητός στόχος. Τα ΤΝΔ σχεδιάζονται μέσω της τροποποίησης των συναπτικών βαρών. Η προσέγγιση αυτή είναι η πιο κοντινή στη γραμμική προσαρμοστική θεωρία φίλτρων (*linear adaptive filter theory*), που έχει ήδη καθιερωθεί και εφαρμόζεται με επιτυχία σε πολλούς διαφορετικούς τομείς. Ωστόσο, δύναται ένα ΤΝΔ να τροποποιήσει τη δική του τοπολογία, όπως γίνεται και στον ανθρώπινο εγκέφαλο, όταν οι εγκεφαλικοί νευρώνες πεθαίνουν και αναπτύσσονται νέες συναπτικές συνδέσεις.

Είναι προφανές ότι η υπολογιστική ισχύς ενός ΤΝΔ προέρχεται πρώτα από τη δομή του και, κατά δεύτερον, από την ικανότητα που έχει να μαθαίνει και επομένως να μπορεί να γενικεύει.

Η **γενίκευση** (*generalization*) είναι η παραγωγή λογικών εξόδων (outputs) για εισροές που δεν χρησιμοποιούνται κατά τη διαδικασία της εκπαίδευσης (μάθησης). Οι προαναφερόμενες δύο δυνατότητες επεξεργασίας των πληροφοριών, κάνουν τα ΤΝΔ να βρίσκουν λύσεις με καλή προσέγγιση, σε προβλήματα πολύπλοκα και μεγάλης κλίμακας που παλαιότερα ήταν δυσεπίλυτα (Haykin and Haykin, 2009).

Τα ΤΝΔ βασίζονται σε υπολογιστικές τεχνικές που αντιστοιχούν σε ορισμένα κύρια λειτουργικά χαρακτηριστικά του ανθρώπινου εγκεφάλου. Τα ΤΝΔ βασίζονται σε μηχανισμούς **μάθησης**, **εκπαίδευσης** και **ελέγχου** και δεν προγραμματίζονται όπως τα συμβατικά προγράμματα υπολογιστών. Στα ΤΝΔ υπάρχει μια σύνδεση μεταξύ των δεδομένων η οποία βασίζεται σε **βάρη** παράλληλα με τις **συνάψεις**, όπως στον ανθρώπινο εγκέφαλο. Η εκπαίδευση ενός τεχνητού νευρωνικού δικτύου επιτυγχάνεται με την προσθήκη αυτών των προαναφερόμενων συνδέσεων μέσω ενός **αλγόριθμου εκπαίδευσης**.

Το **Σχήμα 1** δείχνει την αρχιτεκτονική ενός απλού ΤΝΔ, το οποίο περιλαμβάνει ένα επίπεδο εισόδου, δύο κρυφά επίπεδα και το επίπεδο εξόδου (Alam et al., 2022; Basheer and Hajmeer, 2000; Ding et al., 2013).



Σχήμα 1: Η αρχιτεκτονική ενός απλού ΤΝΔ με ένα επίπεδο εισόδου, δύο κρυφά επίπεδα και επίπεδο εξόδου (Alam et al., 2022)

1.3.1 Ιδιότητες και δυνατότητες των ΤΝΔ

Τα ΤΝΔ έχουν τις παρακάτω ιδιότητες και δυνατότητες (Haykin and Haykin, 2009):

1. **Μη γραμμικότητα (Nonlinearity).** Τα ΤΝΔ είναι μη γραμμικά, καθώς αποτελούνται από συνδέσεις μη γραμμικών νευρώνων. Η ιδιότητα των ΤΝΔ της μη γραμμικότητας είναι πολύ σημαντική, ειδικά όταν ο μηχανισμός παραγωγής των σημάτων εισόδου είναι μη γραμμικός.
2. **Χαρτογράφηση εισόδου-εξόδου.** Κατά την εποπτευόμενη μάθηση (ή μάθηση με δάσκαλο), τα συναπτικά βάρη τροποποιούνται με την εφαρμογή ενός συνόλου παραδειγμάτων εκπαίδευσης. Κάθε παράδειγμα εκπαίδευσης αποτελείται από ένα σήμα εισόδου, το οποίο είναι μοναδικό και ένα σήμα στόχου, το οποίο είναι η επιθυμητή απόκριση του συστήματος. Κάθε φορά τα συναπτικά βάρη του ΤΝΔ τροποποιούνται, έτσι ώστε να ελαχιστοποιηθεί η διαφορά μεταξύ της επιθυμητής απόκρισης (τιμή στόχου) και της πραγματικής απόκρισης του ΤΝΔ, που παράγεται από το σήμα εισόδου, επιλέγοντας ένα κατάλληλο στατιστικό κριτήριο. Η διαδικασία εκπαίδευσης του ΤΝΔ επαναλαμβάνεται για πολλά αντίστοιχα παραδείγματα, έως ότου το ΤΝΔ φτάσει σε μια σταθερή κατάσταση, όπου τα συναπτικά βάρη δεν διαφοροποιούνται σημαντικά.

3. **Προσαρμοστικότητα.** Τα ΤΝΔ, όπως προαναφέρθηκε, έχουν την ικανότητα να προσαρμόζουν τα συναπτικά βάρη τους. Ωστόσο, η προσαρμοστικότητα αυτή, δεν έχει πάντα τα βέλτιστα αποτελέσματα, αλλά μπορεί να έχει ακριβώς αντίθετα αποτελέσματα. Παραδείγματος χάριν, ένα ΤΝΔ, με τιμές που αλλάζουν απότομα, μπορεί, λόγω της προσαρμοστικότητας του, να αλλάξει άμεσα, έτσι ώστε να ανταποκριθεί σε ψευδείς διαταραχές, το οποίο αυτό έχει ως αποτέλεσμα τη δραστική μείωση της απόδοσής του. Το πρόβλημα αυτό στα ΤΝΔ ονομάζεται δίλημμα σταθερότητας- πλαστικότητας.
4. **Αποδεικτική απόκριση (*Evidential Response*).** Ένα ΤΝΔ δύναται να κατασκευαστεί για να βοηθήσει στη λήψη μιας κατάλληλης απόφασης, όταν υπάρχουν διφορούμενα μοτίβα.
5. **Πληροφορίες συμφραζόμενων (*Contextual Information*).** Κάθε νευρώνας σε ένα ΤΝΔ δύναται να επηρεαστεί από τη συμπεριφορά του συνόλου των νευρώνων του δικτύου. Αυτό έχει ως αποτέλεσμα, οι πληροφορίες των συμφραζόμενων, όπως ονομάζονται, να αντιμετωπίζονται με φυσικό τρόπο σε ένα ΤΝΔ.
6. **Ανεκτικότητα σε σφάλματα.** Ένα ΤΝΔ δύναται να έχει ανεκτικότητα σε σφάλματα. Για να εξασφαλισθεί ότι το ΤΝΔ έχει ανεκτικότητα σε σφάλματα, θα πρέπει να ληφθούν κατάλληλα διορθωτικά μέτρα όταν σχεδιάζεται ο αλγόριθμος για την εκπαίδευσή του.
7. **Πολύ μεγάλης κλίμακας ολοκληρωμένη εφαρμοστικότητα (*very-large-scale-integrated Implementability*).** Τα ΤΝΔ δίνουν πολύ γρήγορα υπολογισμούς για διάφορες εργασίες και αυτό έχει βοηθήσει έτσι ώστε να βρουν εφαρμογή σε συστήματα περίπλοκης συμπεριφοράς, χρησιμοποιώντας τεχνολογία πολύ μεγάλης ολοκληρωμένης κλίμακας (Very-large-scale integration, VLSI).

8. Ομοιομορφία Ανάλυσης και Σχεδιασμού. Τα ΤΝΔ έχουν βρει ευρεία εφαρμογή ως επεξεργαστές πληροφοριών. Η ιδιότητα των ΤΝΔ ότι εμφανίζουν ομοιομορφία, συναντάται με διάφορους τρόπους:

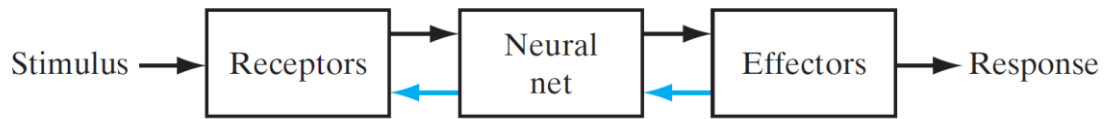
- Οι νευρώνες αποτελούν ένα συστατικό το οποίο είναι κοινό σε όλα τα ΤΝΔ.
- Είναι δυνατή η κοινή χρήση των αλγορίθμων μάθησης σε διαφορετικές μορφές των ΤΝΔ.
- Τα αρθρωτά ΤΝΔ δύναται να κατασκευαστούν με ενοποίηση αντίστοιχων μονάδων.

9. Νευροβιολογική Αναλογία (Neurobiological Analogy). Τα ΤΝΔ σχεδιάζονται με βάση τις λειτουργίες του ανθρώπινου εγκέφαλου. Η λειτουργία του ανθρώπινου εγκέφαλου αποδεικνύει περίτρανα ότι η ανεκτική σε σφάλματα παράλληλη επεξεργασία μπορεί να γίνει με φυσικό τρόπο, γρήγορα και αποτελεσματικά. Από τη μια, τα ΤΝΔ εξετάζονται από τους νευρο-βιολόγους για την ερμηνεία νευρο-βιολογικών φαινομένων και από την άλλη, οι μηχανικοί ελπίζουν και βασίζονται στη νευρο-βιολογία για την επίλυση πολύπλοκων προβλημάτων.

1.3.2 Ο ανθρώπινος εγκέφαλος

Το νευρικό σύστημα ενός ανθρώπινου οργανισμού δύναται να παρασταθεί ως ένα σύστημα αποτελούμενο από τρία στάδια (**Σχήμα 2**). Βασικό στοιχείο του προαναφερόμενου συστήματος είναι ο ανθρώπινος εγκέφαλος, ο οποίος αντιπροσωπεύεται από το νευρικό δίκτυο (neural net), το οποίο λαμβάνει συνεχώς πληροφορίες, τις οποίες αντιλαμβάνεται και στη συνέχεια παίρνει τις κατάλληλες αποφάσεις. Στο **Σχήμα 2**, τα μαύρου χρώματος βέλη, από αριστερά προς δεξιά, υποδεικνύουν την εμπρός μετάδοση σημάτων. Τα μπλε χρώματος βέλη, από δεξιά προς αριστερά, δείχνουν την ανάδραση του συστήματος. Οι υποδοχείς (Receptors) έχουν τη δυνατότητα να μετατρέπουν τα ερεθίσματα του ανθρώπινου σώματος ή του

εξωτερικού περιβάλλοντος σε ερεθίσματα ηλεκτρικά, τα οποία μεταφέρουν τις πληροφορίες στον εγκέφαλο (νευρικό δίκτυο).

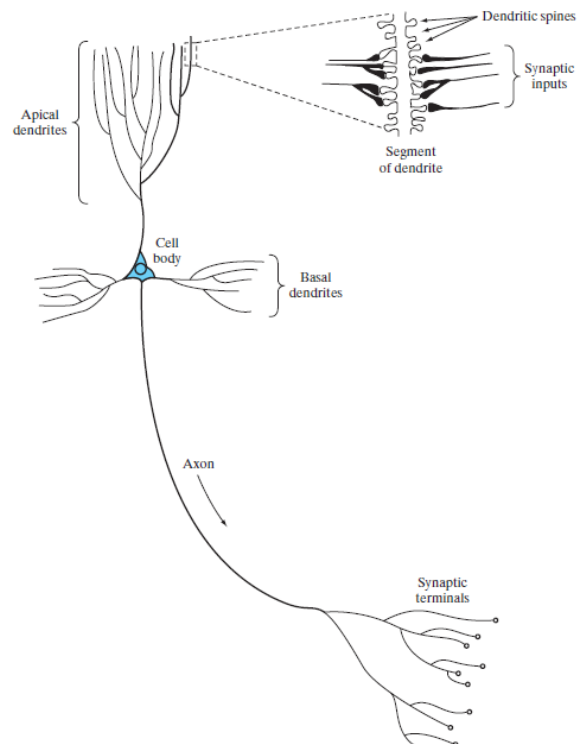


Σχήμα 2: Ανθρώπινο νευρικό σύστημα (Haykin and Haykin, 2009)

Οι νευρώνες είναι πιο αργοί από τις πύλες πυριτίου (συνήθως πέντε έως έξι τάξεις πιο αργοί). Συγκεκριμένα, οι αναδράσεις σε ένα τσιπ πυριτίου συμβαίνουν στο εύρος νανοδευτερόλεπτου, ενώ στους νευρώνες τα αντίστοιχα συμβάντα στο εύρος του χιλιοστού του δευτερολέπτου.

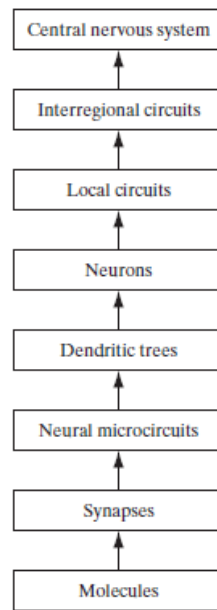
Οι στοιχειώδεις δομικές και λειτουργικές μονάδες, οι οποίες μεσολαβούν στις αλληλεπιδράσεις μεταξύ των νευρώνων, είναι οι συνάψεις ή νευρικές απολήξεις. Η πλαστικότητα του εγκεφάλου ενός ενήλικα ανθρώπου δύναται να οφείλεται είτε στη δημιουργία νέων συναπτικών συνδέσεων μεταξύ των νευρώνων, είτε στην τροποποίηση των υφιστάμενων συνάψεων.

Οι νευρώνες του ανθρώπινου εγκεφάλου διακρίνονται για τη μεγάλη τους ποικιλία σε σχήμα και μέγεθος. Στο **Σχήμα 3** απεικονίζεται ένα τυπικό πυραμιδικό κύτταρο, το οποίο συνήθως συναντάται στον φλοιό του ανθρώπινου εγκεφάλου.



Σχήμα 3: Σχήμα ενός πυραμιδικού κυττάρου (Haykin and Haykin, 2009)

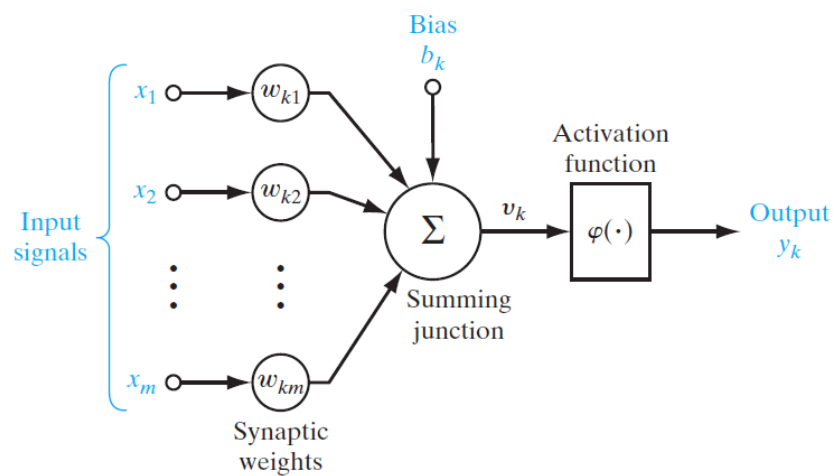
Στο **Σχήμα 4** απεικονίζεται η ιεραρχία των επιπέδων οργάνωσης του ανθρώπινου εγκεφάλου. Το βασικότερο επίπεδο οργάνωσης αποτελούν οι συνάψεις. Άλλα επίπεδα οργάνωσης αποτελούν τα νευρωνικά μικροκυκλώματα, τα δενδριτικά δέντρα και οι νευρώνες.



Σχήμα 4: Ιεραρχία επιπέδων οργάνωσης (Haykin and Haykin, 2009)

1.3.3 Μοντέλα ενός νευρώνα

Ο θεμελιώδης λίθος της λειτουργίας ενός νευρωνικού δικτύου είναι ο νευρώνας, ο οποίος στην ουσία είναι μια μονάδα επεξεργασίας των πληροφοριών. Στο **Σχήμα 5** απεικονίζεται ένα τυπικό μοντέλο ενός νευρώνα. Οι νευρώνες αποτελούν τα βασικά στοιχεία μιας ομάδας Νευρωνικών Δικτύων.



Σχήμα 5: Μη γραμμικό μοντέλο ενός νευρώνα (Haykin and Haykin, 2009)

Όπως αποτυπώνεται στο **Σχήμα 5**, τα τρία (3) θεμελιώδη στοιχεία ενός νευρωνικού μοντέλου είναι:

1. **Συνάψεις ή συνδεδετικοί κρίκοι.** Κάθε μία σύναψη χαρακτηρίζεται από ένα δικό της βάρος ή δύναμη. Συγκεκριμένα, ένα σήμα εισόδου x_j της σύναψης j , το οποίο συνδέεται με τον νευρώνα k , πολλαπλασιάζεται με το συναπτικό βάρος w_{kj} , όπου το k αναφέρεται στον νευρώνα και το j στο βάρος. Στα ΤΝΔ τα συναπτικά βάρη δύναται να πάρουν και αρνητικές τιμές, το οποίο δεν παρατηρείται το ίδιο στον ανθρώπινο εγκέφαλο.
2. **Αθροιστής,** ο οποίος αθροίζει τα σήματα εισόδου, σταθμισμένος με τις αντίστοιχες συναπτικές δυνάμεις του νευρώνα.
3. **Συνάρτηση ενεργοποίησης,** η οποία περιορίζει το πλάτος της εξόδου ενός νευρώνα. Πολλές φορές η συνάρτηση ενεργοποίησης αναφέρεται και ως συνάρτηση σύνθλιψης. Αυτό οφείλεται στο γεγονός ότι περιορίζει το εύρος του πλάτους σήματος εξόδου σε κάποια συγκεκριμένη τιμή.

Το εύρος του πλάτους της εξόδου ενός νευρώνα, όταν έχει υποστεί κανονικοποίηση, είναι το διάστημα $[0,1]$ ή εναλλακτικά, $[-1,1]$. Στο **Σχήμα 5** απεικονίζεται, επίσης, μια παράμετρο $bias$, η οποία εφαρμόζεται εξωτερικά και συμβολίζεται ως b_k . Η προαναφερόμενη παράμετρος b_k αυξάνει ή μειώνει την είσοδο στη συνάρτηση ενεργοποίησης, αντίστοιχα εάν είναι θετική ή αρνητική. Οι μαθηματικές εξισώσεις που περιγράφουν τον νευρώνα k του **Σχήματος 5**, είναι οι (1) και (2):

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (1)$$

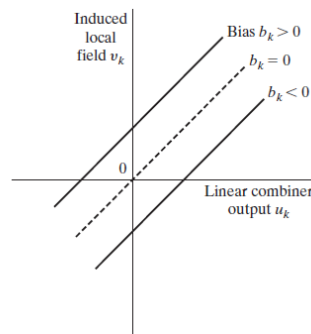
$$y_k = \varphi(u_k + b_k) \quad (2)$$

Όπου: x_1, x_2, \dots, x_m τα σήματα εισόδου, $w_{k1}, w_{k2}, \dots, w_{km}$ τα αντίστοιχα συναπτικά βάρη του νευρώνα k , u_k η έξοδος του γραμμικού συνδυαστή, λόγω των σημάτων εισόδου, b_k

η παράμετρος $bias^1$, $\phi(\cdot)$ η προαναφερόμενη συνάρτηση ενεργοποίησης και y_k το σήμα εξόδου του νευρώνα. Η χρήση της παραμέτρου b_k έγκειται στην εφαρμογή ενός μετασχηματισμού στην έξοδο u_k , η οποία περιγράφεται από την εξίσωση (3).

$$U_k = u_k + b_k \quad (3)$$

Ανάλογα με τις θετικές ή αρνητικές τιμές της παραμέτρου b_k , τροποποιείται η σχέση μεταξύ του δυναμικού ενεργοποίησης, v_k του νευρώνα k και της εξόδου u_k , όπως απεικονίζεται στο **Σχήμα 6**. Αυτό έχει ως αποτέλεσμα, η γραφική παράσταση μεταξύ του v_k έναντι του u_k , να μην περνάει πλέον από την αρχή των αξόνων.



Σχήμα 6: Μετασχηματισμός συγγένειας που παράγεται από την παρουσία $bias$, $v_k=b_k$ στο $u_k=0$ (Haykin and Haykin, 2009)

Όπως έχει προαναφερθεί, η παράμετρος b_k εφαρμόζεται εξωτερικά στον νευρώνα k . Η παρουσία της απεικονίζεται στην εξίσωση (2). Από τις εξισώσεις (1) έως (3) οδηγούμαστε στις εξισώσεις (4) έως (5), για τιμή εισόδου την εξίσωση (6) και για τιμή βάρους την εξίσωση (7):

$$u_k = \sum_{j=0}^m w_{kj} x_j \quad (4)$$

και

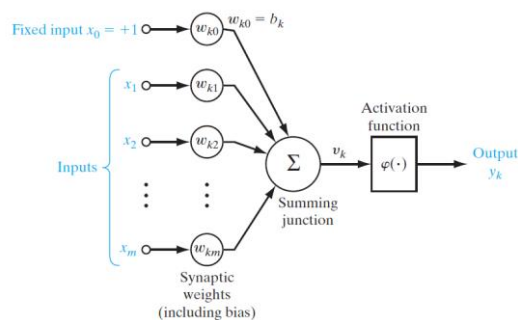
¹ Ο όρος $bias$ στα νευρωνικά δίκτυα εξυπηρετεί το σκοπό να επιτρέπει στη συνάρτηση ενεργοποίησης να ταιριάζει καλύτερα στα δεδομένα.

$$y_k = \varphi(u_k) \quad (5)$$

$$x_0 = +1 \quad (6)$$

$$w_{k0} = b_k \quad (7)$$

Σύμφωνα με τα παραπάνω, το μοντέλο του προαναφερόμενου νευρώνα k , αναδιατυπώνεται όπως απεικονίζεται στο **Σχήμα 7**. Το αποτέλεσμα της πόλωσης υπολογίζεται ως εξής: (1) προσθέτοντας ένα νέο σήμα εισόδου σταθερό στο 1 και (2) προσθέτοντας ένα νέο συναπτικό βάρος ίσο με την παράμετρο b_k . Παρατηρείται ότι, αν και τα μοντέλα των **Σχημάτων 5** και **7** είναι διαφορετικά στην εμφάνιση, στην ουσία είναι ισοδύναμα μαθηματικά.



Σχήμα 7: Ένα άλλο μη γραμμικό μοντέλο ενός νευρώνα. Το w_{k0} αντιπροσωπεύει την παράμετρο b_k (Haykin and Haykin, 2009)

1.4 Σκοποί της Διδακτορικής Διατριβής

Οι άξονες που επικεντρώνεται η παρούσα ΔΔ, είναι οι εξής:

- μοντελοποίηση της λειτουργίας μιας EEN με χρήση ΤΝΔ και ανάλυσης MLR ,
- συσχέτιση των ποιοτικών χαρακτηριστικών του νερού και των κύριων λειτουργικών μεταβλητών μιας EEN και
- κυρίως η πρόβλεψη των δόσεων των απαραίτητων χημικών επεξεργασίας του νερού.

Με βάση προηγούμενες μελέτες, οι οποίες έχουν αποδείξει την αξιοπιστία της πρόβλεψης που βασίζεται σε μοντέλα ΤΝΔ και MLR ανάλυσης στον τομέα της επεξεργασίας επιφανειακών υδάτων, η πρόβλεψη των δόσεων των απαραίτητων χημικών επεξεργασίας του νερού, θα μπορούσε να βοηθήσει τους λειτουργούς EEN, ιδιαίτερα στις σύνθετες φυσικοχημικές διεργασίες που λαμβάνουν χώρα κατά την επεξεργασία του επιφανειακού νερού ταμιευτήρα. Η μοντελοποίηση της λειτουργίας μιας EEN βοηθάει στη λήψη κατάλληλων αποφάσεων στην επιλογή των δόσεων χημικών πεδίου και αποτελεί χρήσιμο εργαλείο στα χέρια ενός οποιοδήποτε λειτουργού μιας EEN, για τον οποίο, ο καθορισμός των δόσεων των χημικών επεξεργασίας νερού αποτελεί κύριο καθημερινό μέλημά του.

Η παρούσα ΔΔ και η έρευνα που έχει διεξαχθεί στο πλαίσιο αυτής, διαφοροποιείται από όσα έχουν διερευνηθεί στην υπάρχουσα διεθνή βιβλιογραφία, καθώς η μοντελοποίηση βασίζεται σε μεγάλο αριθμό δεδομένων από τα πρόσφατα έτη (2019-2022) από μία EEN με μέγιστη δυναμικότητα 110,600 m³/ημέρα και δεν βασίζεται μόνο σε πειραματικά δεδομένα εργαστηριακής κλίμακας (π.χ. από τη διεξαγωγή jar tests). Επίσης, καινοτομία αποτελεί το γεγονός ότι στη μοντελοποίηση για την πρόβλεψη των δόσεων των χημικών επεξεργασίας του νερού λαμβάνεται υπόψη και αξιολογείται η εμπειρία ετών του λειτουργού μιας EEN.

2 ΔΕΔΟΜΕΝΑ

2.1 Εγκατάσταση Επεξεργασίας Νερού (ΕΕΝ) Αποσελέμη

Η Εγκατάσταση Επεξεργασίας Νερού (ΕΕΝ) Αποσελέμη βρίσκεται στο Δ.Δ. Αγριανών του Δήμου Χερσονήσου στον Ν. Ηρακλείου Κρήτης και αποτελεί μέρος των Έργων Αποσελέμη. Σύμφωνα με τον **Νόμο 4138/2013** (ΦΕΚ 72Α/19.03.2013) ο **ΟΑΚ Α.Ε.** είναι ο αρμόδιος φορέας διαχείρισης και λειτουργίας των Έργων Αποσελέμη. Στην **Εικόνα 1** αποτυπώνεται η θέση των έργων Αποσελέμη στον χάρτη της Ελλάδας και της Κρήτης.



Εικόνα 1: Χωροθέτηση έργων Αποσελέμη

Στην **Εικόνα 2** αποτυπώνεται η θέση του φράγματος, του ταμιευτήρα, του

υδραγωγείου και της ΕΕΝ Αποσελέμη σε απόσπασμα του χάρτη του Ν. Ηρακλείου Κρήτης.



Εικόνα 2: Χάρτης αποτύπωσης θέσης φράγματος, ταμιευτήρα, υδραγωγείου και ΕΕΝ Αποσελέμη

Η ΕΕΝ Αποσελέμη επεξεργάζεται το νερό που προέρχεται από την τεχνητή λίμνη του ταμιευτήρα του φράγματος Αποσελέμη (Εικόνες 3- 4), ο οποίος έχει χωρητικότητα $25.3 \times 10^6 \text{ m}^3$. Η ΕΕΝ Αποσελέμη είναι μία συμβατικού τύπου εγκατάσταση φυσικοχημικής επεξεργασίας νερού, η οποία αποτελείται από τις διεργασίες της οζόνωσης, της κροκιδο-καθίζησης, της διήθησης με φίλτρα άμμου και της απολύμανσης με χρήση αερίου χλωρίου. Η μέγιστη ημερήσια δυναμικότητα επεξεργασίας νερού της ΕΕΝ Αποσελέμη ανέρχεται στα: **110.600 m³/d**.



Εικόνα 3: Άποψη του φράγματος και του ταμιευτήρα Αποσελέμη (1/2)



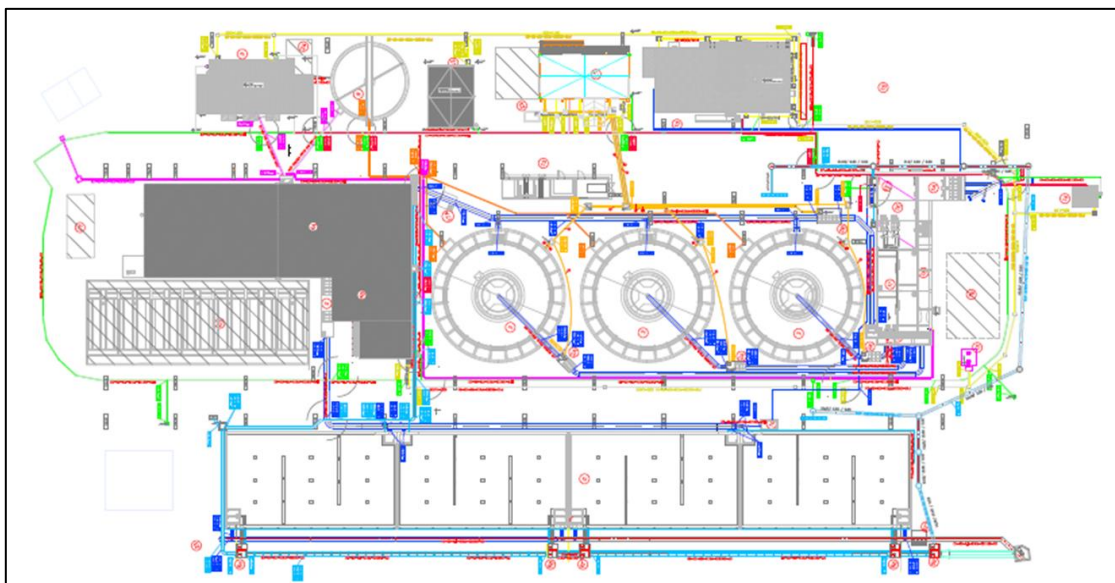
Εικόνα 4: Άποψη του φράγματος και του ταμιευτήρα Αποσελέμη (2/2)

2.2 Περιγραφή λειτουργίας ΕΕΝ Αποσελέμη

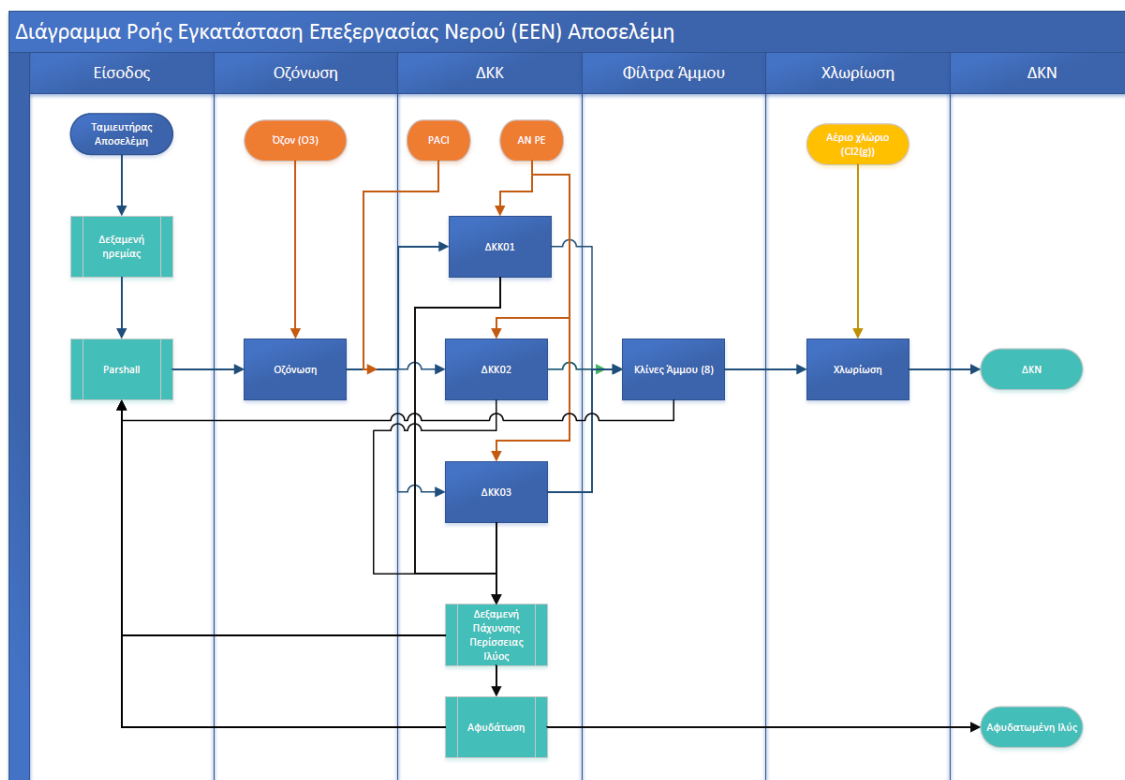
Οι κύριες διεργασίες επεξεργασίας του νερού στην ΕΕΝ Αποσελέμη, όπως προαναφέρθηκε, είναι τέσσερις: οζόνωση, κροκιδο-καθίζηση, διήθηση με φίλτρα άμμου και απολύμανση με χρήση αερίου χλωρίου. Στην **Εικόνα 5** αποτυπώνεται μια πανοραμική άποψη της ΕΕΝ Αποσελέμη και στο **Σχήμα 8** η κάτοψη της εγκατάστασης. Επίσης, στο **Σχήμα 9** αποτυπώνεται το διάγραμμα ροής της παραγωγικής διαδικασίας της ΕΕΝ Αποσελέμη.



Εικόνα 5: Εγκατάσταση Επεξεργασίας Νερού Αποσελέμη



Σχήμα 8: Κάτοψη της Εγκατάστασης Επεξεργασίας Νερού Αποσελέμη



Σχήμα 9: Διάγραμμα Ροής παραγωγικής διαδικασίας ΕΕΝ Αποσελέμη

Οι διαδικασίες οζονισμού, κροκίδωσης και απολύμανσης δεν προσομοιώνονται εύκολα, χρησιμοποιώντας κλασικές μεθόδους μοντελοποίησης, λόγω των πολύπλοκων φυσικών και χημικών μηχανισμών που εμπλέκονται. Η κροκίδωση, είναι μια κοινή μέθοδος για την επεξεργασία του νερού εδώ και δεκαετίες, ο κύριος σκοπός της οποίας είναι η απομάκρυνση των κολλοειδών σωματιδίων (Dadebo et al., 2023; Li et al., 2021a; Lin et al., 2023b; Yateh et al., 2023).

2.2.1 Οζόνωση

Η είσοδος του νερού στην ΕΕΝ Αποσελέμη πραγματοποιείται από τον αγωγό του ταμιευτήρα Αποσελέμη με χρήση ηλεκτροκίνητων δικλίδων. Το νερό εισέρχεται στη δεξαμενή ηρεμίας και στη συνέχεια εισέρχεται στο κανάλι εισόδου, όπου πραγματοποιείται μέτρηση της παροχής με παροχόμετρο τύπου Parshall. Στη συνέχεια, το νερό εισέρχεται στη δεξαμενή επαφής νερού- όζοντος της μονάδας οζόνωσης. Το σύστημα παραγωγής όζοντος αποτελείται από δύο παράλληλες μονάδες (η μία

εφεδρική) δυναμικότητας 12 kg O₃/h έκαστη με χρήση αέρα. Η κάθε μονάδα αποτελείται από το σύστημα παραγωγής και επεξεργασίας αέρα, τη γεννήτρια παραγωγής όζοντος και τα όργανα μέτρησης και λειτουργίας της κάθε μονάδας.

2.2.2 Κροκίδωση, Καθίζηση

Στη διεργασία της κροκίδωσης πραγματοποιείται η προσθήκη χημικών στο νερό για την αποσταθεροποίηση των κολλοειδών σωματιδίων, έτσι ώστε να σχηματίζονται ακόμη μεγαλύτερα συσσωματώματα και να καθιζάνουν στη διεργασία της καθίζησης, η οποία ακολουθεί. Η διεργασία της κροκιδο- καθίζησης στην EEN Αποσελέμη πραγματοποιείται σε τρεις κυλινδρικές δεξαμενές κροκίδωσης και καθίζησης με περιστρεφόμενα ξέστρα. Ως κύριο κροκιδωτικό έχει επιλεγεί το θειικό αργίλιο [Al₂(SO₄)₃ 14*H₂O)] ή εναλλακτικά το χλωριούχο θειικό πολυ- αργίλιο (Poly- Aluminum chloride hydroxide sulfate, PACl). Ως ενισχυτικό της καθίζησης για τη δημιουργία ακόμη μεγαλύτερων συσσωματωμάτων χρησιμοποιείται ανιονικός πολυηλεκτρολύτης (Anionic Polyelectrolyte, AN PE), ο οποίος είναι ένα ανιονικό πολυακρυλαμίδιο, μεσαίου ιοντικού φορτίου και υψηλού μοριακού βάρους και χρησιμοποιείται σε συνδυασμό με τα ανόργανα κροκιδωτικά για την επεξεργασία πόσιμου νερού.

Η τροφοδοσία του πολυηλεκτρολύτη προς τη μονάδα κροκίδωσης γίνεται για την υποβοήθηση της διαδικασίας της κροκίδωσης. Στις δεξαμενές κροκιδοκαθίζησης εγκάρσιας ροής, η διεργασία της συσσωμάτωσης πραγματοποιείται στο ανάντη τμήμα της δεξαμενής. Καθώς οι κροκίδες που σχηματίζονται γίνονται όλο και μεγαλύτερες και βαρύτερες, καθιζάνουν στον πυθμένα της δεξαμενής καθίζησης.

2.2.3 Διήθηση

Η διήθηση (διύλιση) του νερού στην EEN Αποσελέμη πραγματοποιείται από οκτώ (8) μονοστρωματικές δίδυμες κλίνες με φίλτρα άμμου διατεταγμένες σε παράλληλη λειτουργία. Το πληρωτικό υλικό των κλινών διήθησης είναι φυσική χαλαζιακή άμμος με ύψος στρώματος άμμου 1,4 m.

2.2.4 Χλωρίωση

Το νερό μετά τη διήθηση υφίσταται μετα- χλωρίωση με πυκνό διάλυμα χλωρίου. Το αέριο χλώριο ($\text{Cl}_{2(g)}$) μεταφέρεται και αποθηκεύεται στην εγκατάσταση σε υγροποιημένη μορφή σε συνολικά έξι τυποποιημένους κυλίνδρους χωρητικότητας ενός τόνου. Επιπρόσθετα, στην ΕΕΝ Αποσελέμη έχουν εγκατασταθεί και λειτουργούν δύο μονάδες ηλεκτροχλωρίωσης για την παραγωγή διαλύματος υποχλωριώδους νατρίου, μέσω της διεργασίας της ηλεκτρόλυσης κεκορεσμένου διαλύματος άλμης χλωριούχου νατρίου (NaCl). Η μονάδα ηλεκτροχλωρίωσης χρησιμοποιείται εναλλακτικά της χρήσης αερίου χλωρίου.

2.2.5 Δεξαμενή καθαρού νερού

Στη Δεξαμενή Καθαρού Νερού (ΔΚΝ) αποθηκεύεται προσωρινά προς διανομή το τελικά επεξεργασμένο νερό. Η ΔΚΝ αποτελείται από τέσσερις θαλάμους, έκαστος αποτελούμενος από τρία διαμερίσματα μαιανδρικής διαμόρφωσης, συνολικής χωρητικότητας 22.500 m^3 .

2.2.6 Εργαστήριο ελέγχου ποιότητας νερού

Στην ΕΕΝ Αποσελέμη λειτουργεί το **Εργαστήριο Ελέγχου Ποιότητας Νερού Αποσελέμη**, το οποίο αποτελεί ένα σύγχρονο εργαστήριο ανάλυσης φυσικοχημικών και μικροβιολογικών παραμέτρων ποιότητας νερού. Είναι εξοπλισμένο με φασματοφωτόμετρα, θολερόμετρα, ατομική απορρόφηση, ανάστροφο μικροσκόπιο, αναλυτή TOC (Total Organic Carbon Analyzer, TOC Analyzer), τοξικόμετρο και πλήθος άλλων οργάνων και εξοπλισμού, απαραίτητα για την καθημερινή παρακολούθηση και τον ποιοτικό έλεγχο του νερού σε όλα τα στάδια επεξεργασίας και τελικής διανομής του στους τελικούς καταναλωτές των Δημοτικών Επιχειρήσεων Ύδρευσης-Αποχέτευσης (ΔΕΥΑ) Ηρακλείου, Αγίου Νικολάου και Χερσονήσου.

2.2.7 Πάχυνση και αφυδάτωση ιλύος

Επίσης, στην EEN Αποσελέμη υπάρχει βαρυτικός παχυντής ιλύος. Πρόκειται για μια κυλινδρική δεξαμενή με ξέστρο, στην οποία οδηγείται η ιλύς από τον πυθμένα των τριών δεξαμενών κροκίδο- καθίζησης. Στην ιλύ αυξάνεται βαρυτικά η συγκέντρωση των στερεών, έτσι ώστε να αφυδατωθεί στη συνέχεια στη μονάδα αφυδάτωσης. Η μονάδα αφυδάτωσης αποτελείται από δύο ταινιο-φιλτρόπρεσες. Στη μονάδα αυτή η παχυμένη πλέον ιλύς, η οποία αντλείται από τον πυθμένα του παχυντή ιλύος, οδηγείται σε πρέσες όπου αφυδατώνεται.

2.2.8 Μονάδα ενεργού άνθρακα

Στην EEN Αποσελέμη, με στόχο την απομάκρυνση , μέσω προσρόφησης, των ουσιών που προκαλούν οσμή και γεύση στο νερό, έχει εγκατασταθεί και λειτουργεί, όταν απαιτείται, μονάδα δοσιμέτρησης ενεργού άνθρακα σε μορφή σκόνης (*Powdered Activated Carbon, PAC*).

2.3 Κύριες λειτουργικές μεταβλητές και μεταβλητές ποιότητας νερού

Η παρακολούθηση των κύριων λειτουργικών μεταβλητών μιας EEN και των μεταβλητών ποιότητας του νερού είναι κρίσιμο ζήτημα για όλους τους λειτουργούς εγκαταστάσεων επεξεργασίας νερού για την παραγωγή πόσιμου (ανθρώπινης κατανάλωσης) νερού.

Η επεξεργασία του νερού αποτελείται από μια ακολουθία σύνθετων φυσικών και χημικών διεργασιών. Στις EEN η ρύθμιση των διεργασιών, γενικά, επιτυγχάνεται αξιολογώντας την ποιότητα του εισερχόμενου και του τελικά παραγόμενου νερού και προσαρμόζοντας τις διεργασίες μέσω της εμπειρίας του λειτουργού της εγκατάστασης. Οι διεργασίες οξονισμού, κροκίδωσης και απολύμανσης περιλαμβάνουν πολλούς πολύπλοκους φυσικούς και χημικούς μηχανισμούς, οι οποίοι είναι δύσκολο να μοντελοποιηθούν χρησιμοποιώντας παραδοσιακές μεθόδους μοντελοποίησης (Wu and Lo, 2010).

Τα χαρακτηριστικά του νερού, όπως η θερμοκρασία, η θολότητα και το pH είναι σημαντικές παράμετροι ποιότητας του νερού και υπάρχει μια σημαντική σχέση μεταξύ αυτών των μεταβλητών και των ποσοτήτων χημικών κροκιδωτικών και απολυμαντικών, που χρησιμοποιούνται στις διεργασίες επεξεργασίας του νερού. Οι βροχοπτώσεις μπορούν να προκαλέσουν πλήθος μεταβολών στα χαρακτηριστικά του νερού εισόδου μιας ΕΕΝ, όπως στη θολότητα, και να μειώσουν την αποδοτικότητα της απολύμανσης.

Στην ΕΕΝ Αποσελέμη οι ακόλουθοι παράγοντες θεωρήθηκαν σημαντικοί, σύμφωνα με τη μακρόχρονη εμπειρία του λειτουργού της εγκατάστασης, για τον έλεγχο των διεργασιών:

1. Ημερήσια διαφορά ύψους νερού στον ταμιευτήρα, Daily difference in water height in reservoir (ΔH)
2. Παροχή νερού προς επεξεργασία, Raw water supply (Q)
3. Θολότητα ανεπεξέργαστου νερού, Raw water turbidity (T_1)
4. pH ανεπεξέργαστου νερού, Raw water pH (pH_1)
5. Θολότητα επεξεργασμένου νερού, Treated water turbidity (T_2)
6. pH επεξεργασμένου νερού, Treated water pH (pH_2)
7. Ελεύθερο υπολειμματικό χλώριο επεξεργασμένου νερού, Treated water residual chlorine (Cl_2)
8. Υπολειμματικό αργίλιο επεξεργασμένου νερού, Treated water concentration of residual Aluminium (Al)
9. Ημερήσια καταναλισκόμενη ηλεκτρική ενέργεια ΕΕΝ, Daily consumption of DWTP electricity (EI)
10. Υπολειμματικό όζον, Residual O_3 after ozonation process (O_3)
11. Δόση ανιονικού πολυηλεκτρολύτη, Anionic polyelectrolyte (ANPE) dosage
12. Δόση χλωριούχου θειικού πολύ- αργιλίου, Poly- Aluminum chloride hydroxide sulfate (PACl) dosage
13. Παροχή αερίου χλωρίου, Chlorine gas supply ($Cl_{2(g)}$)
14. Θολότητα νερού στην είσοδο των κλινών διήθησης, Filtration Beds inlet water turbidity (T_3)

Στη διεργασία της κροκίδωσης, η τιμή του pH είναι η σημαντική παράμετρος που καθορίζει το φορτίο των κύριων ειδών κροκίδωσης και της συγκέντρωσης των ιζημάτων υδροξειδίου. Η χαμηλή θολότητα του ανεπεξέργαστου νερού μπορεί να έχει επίδραση στον πολύ μικρό αριθμό σωματιδίων για καλή κροκίδωση. Η παροχή του εισερχόμενου νερού στην εγκατάσταση, άρα και ο ρυθμός υπερχείλισης της δεξαμενής κροκιδοκαθίξεσης, μπορεί να έχει επιπτώσεις στην απόδοση των διεργασιών. Από τη μια πλευρά, οι υψηλοί ρυθμοί υπερχείλισης της δεξαμενής κροκιδοκαθίξεσης, όχι μόνο μειώνουν την αποτελεσματικότητα της καθίξεσης αλλά δυσχεραίνουν και τη λειτουργία και την απόδοση των υπόλοιπων διεργασιών που ακολουθούν. Ωστόσο, ένας χαμηλός ρυθμός υπερχείλισης μπορεί να μην είναι οικονομικός για τη λειτουργία της εγκατάστασης. Η δοσολογία χλωριούχου θεικού πολύ-αργιλίου, το οποίο είναι το βασικό κροκιδωτικό που χρησιμοποιείται στην εγκατάσταση, αποτελεί την κύρια παράμετρο για τη διεργασία της κροκιδοκαθίξεσης.

Επειδή η διεργασία της κροκίδωσης και της καθίξεσης πραγματοποιούνται σε μία δεξαμενή κροκιδο-καθίξεσης εγκάρσιας ροής, αυτές οι διεργασίες μελετήθηκαν ως μία διεργασία ελέγχου (control unit).

Στην ανάπτυξη ενός συστήματος ελέγχου και ρύθμισης των διεργασιών, η προγενέστερη γνώση της κύριας μονάδας ελέγχου είναι πολύ σημαντική. Κατά πρώτον, η προγενέστερη γνώση μπορεί να χρησιμοποιηθεί για να αναγνωρισθούν οι παράγοντες κλειδιά (key-parameters) εισόδου και εξόδου. Κατά δεύτερον, μία ξεκάθαρη κατανόηση των μηχανισμών των διεργασιών θα βοηθήσει στον προσδιορισμό των άμεσων αλληλεπιδράσεων μεταξύ των μεταβλητών εισόδου και εξόδου και της έμμεσης σχέσης όπως στη συσχέτιση των χρονοσειρών.

Υπάρχουν τρεις (3) συνεχείς διεργασίες που λαμβάνουν μέρος στη μονάδα ελέγχου. Η διεργασία της κροκίδωσης, η οποία αναφέρεται στην προσθήκη χημικών στο νερό για την αποσταθεροποίηση των κolloειδών σωματιδίων, που επιτρέπει σε αυτά στη συνέχεια να σχηματίσουν ακόμη μεγαλύτερα συσσωματώματα. Στις δεξαμενές κροκιδοκαθίξεσης εγκάρσιας ροής, η διεργασία της συσσωμάτωσης πραγματοποιείται στο ανάντη τμήμα της δεξαμενής και καθώς οι κροκίδες που σχηματίζονται γίνονται όλο και μεγαλύτερες και βαρύτερες, καθιζάνουν στον πυθμένα

της δεξαμενής καθίζησης.

Η αναγνώριση των σημαντικών/κύριων λειτουργικών μεταβλητών στις διεργασίες επεξεργασίας του νερού της ΕΕΝ Αποσελέμη ήταν απαραίτητη για την προσέγγιση της μοντελοποίησης με χρήση ΤΝΔ και συγκεκριμένα για τον προσδιορισμό των κατάλληλων μεταβλητών εισόδου και εξόδου.

Οι προαναφερόμενες κύριες λειτουργικές μεταβλητές και οι μεταβλητές ποιότητας νερού ταξινομήθηκαν σε δύο κατηγορίες μεταβλητών:

1. Σε αυτές που δύναται να χρησιμοποιηθούν για την παρακολούθηση λειτουργίας της ΕΕΝ και την παρακολούθηση των σταδίων επεξεργασίας του νερού, και
2. σε αυτές που δύναται να χρησιμοποιηθούν για τη ρύθμιση και τον έλεγχο των επιτελούμενων φυσικο- χημικών διεργασιών επεξεργασίας του νερού.

Η πρώτη κατηγορία κυρίως περιλαμβάνει τις **μεταβλητές ποιότητας του νερού, όπως και διάφορες λειτουργικές μεταβλητές**, όπως είναι το pH, η θολότητα εισερχόμενου νερού και η θολότητα σε όλα τα στάδια επεξεργασίας αυτού, η παροχή του υπό επεξεργασία νερού, η συγκέντρωση ελεύθερου υπολειμματικού χλωρίου και αργιλίου, όπως και η καταναλισκόμενη ηλεκτρική ενέργεια.

Η δεύτερη κατηγορία μεταβλητών που χρησιμοποιούνται για τη ρύθμιση των διεργασιών, περιλαμβάνει **λειτουργικές μεταβλητές**, οι τιμές των οποίων καθορίζονται από τον λειτουργό της εγκατάστασης, όπως είναι οι δόσεις των χημικών και συγκεκριμένου του όζοντος (O_3), ως ισχυρό οξειδωτικό και απολυμαντικό μέσου, του χλωριούχου θεικού πολύ- αργιλίου, PACl, του ανιονικού πολυηλεκτρολύτη, AN PE και του αερίου χλωρίου ($Cl_{2(g)}$). Αυτές οι μεταβλητές ρυθμίζονται από τους λειτουργούς μιας εγκατάστασης επεξεργασίας νερού έτσι ώστε να ικανοποιούνται τα κριτήρια της απόδοσης που έχουν τεθεί για την εγκατάσταση, τα οποία, δεν είναι άλλα, από το τα όρια και οι προδιαγραφές της πρόσφατης νομοθεσίας περί νερού ανθρώπινης κατανάλωσης, σύμφωνα με την **ΚΥΑ Δ1(δ)/ΓΠοικ.27829 (ΦΕΚ 3525'Β, 25/05/2023)** σχετικά με την **ποιότητα νερού ανθρώπινης κατανάλωσης σε συμμόρφωση προς τις**

διατάξεις της Οδηγίας (ΕΕ) 2020/2184 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 16ης Δεκεμβρίου 2020 (L435/1, 23/12/2020).

Τα δεδομένα που χρησιμοποιήθηκαν συλλέχθηκαν από:

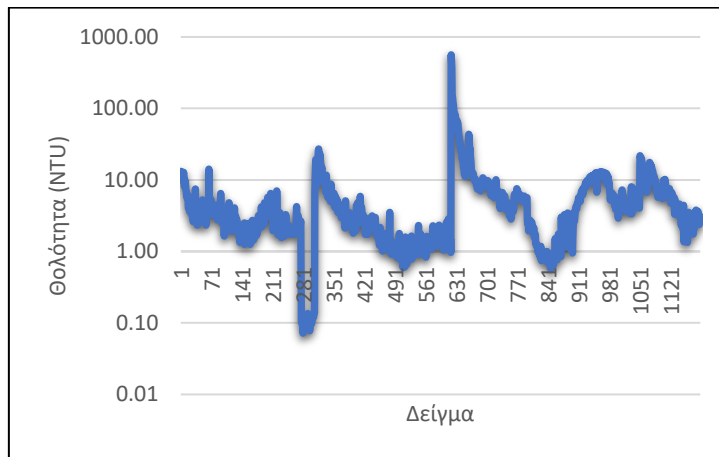
- ✓ τις ημερήσιες αναλύσεις δειγμάτων νερού από όλα τα στάδια επεξεργασίας του στην EEN Αποσελέμη. Οι αναλύσεις διεξήχθησαν στο Εργαστήριο Ελέγχου Ποιότητας Νερού Αποσελέμη σύμφωνα με πρότυπες μεθόδους ανάλυσης νερού (Bridgewater et al., 2017), όπως και άλλες πρότυπες διαπιστευμένες μεθόδους, σύμφωνα με τον Διεθνή Οργανισμό ISO (*International Organization for the Standardization*)
- ✓ και από το εγκατεστημένο SCADA (*Supervisory Control and Data Acquisition*) της EEN Αποσελέμη.

Συγκεκριμένα, συλλέχθηκαν **1.188 τιμές** για καθεμία από τις **14** προαναφερόμενες μεταβλητές, ήτοι συνολικά **16.632** τιμές για χρονικό διάστημα **38 μηνών (01.03.2019 έως 31.05.2022)**.

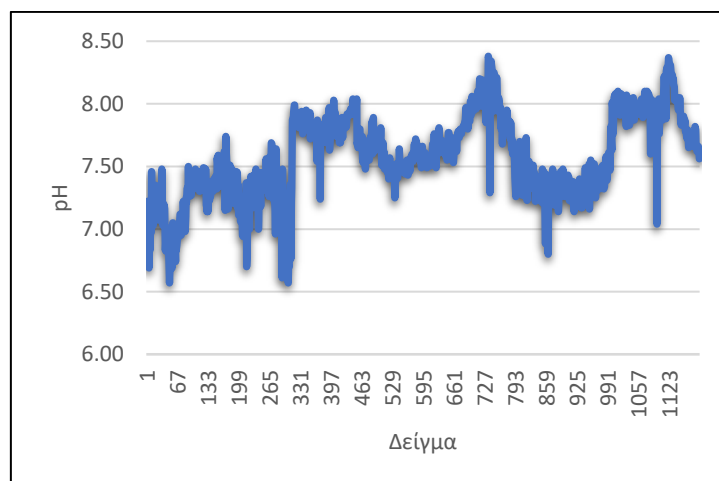
Τα μοντέλα πρόβλεψης ποιότητας νερού προβλέπουν την ποιότητα του νερού για μελλοντικές περιόδους που χρησιμοποιούν ιστορικά δεδομένα σε πραγματικό χρόνο και δεδομένα από απομακρυσμένα συστήματα παρακολούθησης της ποιότητας του νερού (Kim et al., 2023). Σύμφωνα με τη βιβλιογραφία, αν και έχουν γίνει πολλές προσπάθειες για την κατασκευή μαθηματικών μοντέλων πρόβλεψης της χρήσης των χημικών επεξεργασίας του νερού σε EEN, αυτά τα μοντέλα δεν μπορούν να προβλέψουν σύνθετες φυσικοχημικές διεργασίες και μπορεί να αντιμετωπίσουν δυσκολίες στην ανάλυση μη γραμμικών σχέσεων μεταξύ των συστατικών της διεργασίας, για παράδειγμα, στην κροκίδωση (Li et al., 2021b; Lin et al., 2023b).

Κατά τη συλλογή των δεδομένων, παρατηρήθηκαν εποχιακές διακυμάνσεις στα δεδομένα ποιότητας του ακατέργαστου νερού. Συγκεκριμένα, κατά τη διάρκεια του φθινοπώρου και του χειμώνα, που είναι περίοδοι των βροχών, παρατηρήθηκε

αυξημένη διακύμανση στη θολότητα και στην τιμή του pH του ακατέργαστου νερού. Ενδεικτικά, η πορεία της θολότητας του ανεπεξέργαστου νερού και της αντίστοιχης τιμής του pH, ανά δείγμα, απεικονίζεται στα **Σχήματα 10** και **11**, αντίστοιχα. Στο **Παράρτημα Ι** παρατίθενται τα υπόλοιπα αντίστοιχα διαγράμματα όλων των μελετώμενων μεταβλητών στο πλαίσιο της παρούσας ΔΔ.



Σχήμα 10: Θολότητα ανεπεξέργαστου νερού



Σχήμα 11: pH ανεπεξέργαστου νερού

2.3.1 Κύριες στατιστικές παράμετροι μεταβλητών

Οι κύριες στατιστικές παράμετροι (MIN, MAX, AVERAGE, STD) των δεδομένων που χρησιμοποιήθηκαν αποτυπώνονται στον Πίνακα 2.

Πίνακας 2: Κύριες στατιστικές παράμετροι μεταβλητών

No	Variable	Unit	MIN	MAX	AVERAGE	STDEV
1	Ημερήσια διαφορά ύψους νερού στον ταμιευτήρα (ΔΗ)	m	-1,91	3,55	0,00	0,18
2	Παροχή ανεπεξέργαστου νερού (Q)	m ³ /d	4.271	71.858	39.062	9.492
3	Θολότητα ανεπεξέργαστου νερού (T ₁)	NTU	0,07	562,00	6,61	22,12
4	pH ανεπεξέργαστου νερού (pH ₁)		6,57	8,38	7,58	0,34
5	Θολότητα επεξεργασμένου νερού (T ₂)	NTU	0,01	0,74	0,16	0,08
6	pH επεξεργασμένου νερού (pH ₂)		6,42	8,03	7,30	0,32
7	Συγκέντρωση ελ. υπολειμμ. χλωρίου επεξεργασμένου νερού (Cl ₂)	mg/L	0,02	0,90	0,44	0,11
8	Συγκέντρωση υπολειμμ. Αργιλίου επεξεργασμένου νερού (Al)	μg/L	7,00	146,00	41,97	21,36
9	Ημερήσια κατ/μενη ηλεκτρική ενέργεια στην EEN (EI)	kWh	1.060	19.788	9.424	2.892
10	Συγκέντρωση υπολειμματικού όζοντος (O ₃)	mg/L	0,00	0,20	0,05	0,02
11	Δόση ανιονικού πολυηλεκτρολύτη (ANPE)	mg/L	0,20	0,80	0,40	0,15
12	Δόση χλωριούχου θεικού πολύ- αργιλίου (PACl)	mg/L	7,00	100,00	17,99	11,64
13	Παροχή αερίου χλωρίου (Cl _{2(g)})	kg/h	0,70	8,00	2,46	1,27
14	Θολότητα νερού εισόδου κλινών διήθησης (T ₃)	NTU	0,17	7,25	1,29	0,84

2.3.2 Κανονικοποίηση δεδομένων

Για να χρησιμοποιηθούν τα δεδομένα στην ανάλυση πολλαπλής γραμμικής παλινδρόμησης και στην ανάπτυξη ΤΝΔ, ήταν απαραίτητη η κανονικοποίηση των δεδομένων, λόγω του διαφορετικού εύρους κλίμακας μεταξύ των διαφορετικών μεταβλητών. Συγκεκριμένα, τα δεδομένα κλιμακώθηκαν γραμμικά μεταξύ 0.0 και 1.0 χρησιμοποιώντας την εξίσωση:

$$Normalized\ Data = \frac{L - Min}{Max - Min} \quad (8)$$

Όπου:

L: η ανεπεξέργαστη τιμή της εκάστοτε μεταβλητής, Min: η ελάχιστη τιμή της μεταβλητής, Max: η μέγιστη τιμή της μεταβλητής.

Οι κανονικοποιημένες τιμές των κύριων στατιστικών παραμέτρων (MIN, MAX, AVERAGE, STD) των μεταβλητών που χρησιμοποιήθηκαν αποτυπώνονται στον **Πίνακα 3**.

Πίνακας 3: Κανονικοποιημένες κύριες στατιστικές παράμετροι μεταβλητών

No	Variable	MIN	MAX	AVERAGE	STDEV
1	Ημερήσια διαφορά ύψους νερού στον ταμιευτήρα (ΔΗ)	0,00	1,00	0,35	0,03
2	Παροχή ανεπεξέργαστου νερού (Q)	0,00	1,00	0,51	0,14
3	Θολότητα ανεπεξέργαστου νερού (T ₁)	0,00	1,00	0,01	0,04
4	pH ανεπεξέργαστου νερού (pH ₁)	0,00	1,00	0,56	0,19
5	Θολότητα επεξεργασμένου νερού (T ₂)	0,00	1,00	0,21	0,11
6	pH επεξεργασμένου νερού (pH ₂)	0,00	1,00	0,55	0,20
7	Συγκέντρωση ελ. υπολειμ. χλωρίου επεξεργασμένου νερού (Cl ₂)	0,00	1,00	0,48	0,13
8	Συγκέντρωση υπολειμ. Αργιλίου επεξεργασμένου νερού (Al)	0,00	1,00	0,25	0,15
9	Ημερήσια κατ/μενη ηλεκτρική ενέργεια στην EEN (EI)	0,00	1,00	0,45	0,15
10	Συγκέντρωση υπολειμματικού όζοντος (O ₃)	0,00	1,00	0,23	0,10
11	Δόση ανιονικού πολυηλεκτρολύτη (ANPE)	0,00	1,00	0,34	0,26
12	Δόση χλωριούχου θεικού πολύ- αργιλίου (PACI)	0,00	1,00	0,12	0,13
13	Παροχή αερίου χλωρίου (Cl _{2(g)})	0,00	1,00	0,24	0,17
14	Θολότητα νερού εισόδου κλινών διήθησης (T ₃)	0,00	1,00	0,16	0,12

Στο **Παράρτημα II** παρέχονται οι Συντελεστές Συσχέτισης Pearson, από τους οποίους υποδεικνύεται ότι η συσχέτιση ανάμεσα στις μεταβλητές εισόδου και στις μεταβλητές εξόδου στα μοντέλα είναι γενικά χαμηλή και καμία άμεση γραμμική συσχέτιση δεν ανιχνεύεται μεταξύ τους.

3 ΜΕΘΟΔΟΙ

3.1 Πολλαπλή Γραμμική Παλινδρόμηση, ΠΓΠ (*Multiple Linear Regression, MLR*)

Στην **ανάλυση της απλής γραμμικής παλινδρόμησης** εξετάζετε η (γραμμική) σχέση μεταξύ μεταβλητών: μιας **εξαρτημένης** και μιας **ανεξάρτητης μεταβλητής**. Πραγματοποιείται μια προσπάθεια για πρόβλεψη των τιμών μιας μεταβλητής με τις τιμές της άλλης μεταβλητής. Σε πραγματικά όμως προβλήματα, για να προβλεφθεί η τιμή μιας εξαρτημένης μεταβλητής απαιτούνται περισσότερες από μια ανεξάρτητες μεταβλητές.

Για την πρόβλεψη αυτή πραγματοποιείται μια προσπάθεια κατασκευής ενός μαθηματικού μοντέλου, το οποίο ονομάζεται **μοντέλο πολλαπλής γραμμικής παλινδρόμησης**. Η συνολική τεχνική της ανάπτυξης του μοντέλου ονομάζεται **ανάλυση πολλαπλής γραμμικής παλινδρόμησης (*multiple linear regression analysis*)**. Η γενική εξίσωση ενός μοντέλου πολλαπλής γραμμικής παλινδρόμησης είναι μια εξίσωση της μορφής:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \dots + \beta_p \cdot X_p + e \quad (9)$$

όπου $\beta_0, \beta_1, \dots, \beta_p$ είναι οι **συντελεστές μερικής (πολλαπλής γραμμικής) παλινδρόμησης (partial regression coefficients)**, οι X_1, X_2, \dots, X_p είναι οι **ανεξάρτητες μεταβλητές (predictor variables)** και το e το **σφάλμα (error)** στο γραμμικό μοντέλο. Μια εξίσωση της παραπάνω μορφής λέμε ότι παριστά μια ευθεία πολλαπλής γραμμικής παλινδρόμησης. Η μέθοδος που χρησιμοποιείται για την εκτίμηση των προαναφερόμενων συντελεστών είναι η **μέθοδος των ελαχίστων τετραγώνων**. Σύμφωνα με τη μέθοδο αυτή, επιλέγονται εκείνοι οι συντελεστές που δίνουν το μικρότερο άθροισμα των τετραγώνων των διαφορών ανάμεσα στις παρατηρούμενες και τις προβλεπόμενες τιμές.

Στην ανάλυση πολλαπλής γραμμικής παλινδρόμησης, η εξαρτημένη μεταβλητή θα πρέπει να είναι αναλογικού ή αριθμητικού τύπου, ενώ οι ανεξάρτητες μεταβλητές δύναται να είναι ή αριθμητικού τύπου ή κατηγορικού τύπου με ορισμένες προϋποθέσεις. Αναφορικά με το πλήθος των παρατηρήσεων (cases), κατ' ελάχιστον η απαίτηση είναι ο αριθμός των παρατηρήσεων να είναι πενταπλάσιος του αριθμού των ανεξάρτητων μεταβλητών (predictors).

Επίσης, στην ανάλυση πολλαπλής γραμμικής παλινδρόμησης, τα δεδομένα θα πρέπει να ικανοποιούν τις παρακάτω προϋποθέσεις:

- **Ανεξαρτησία.** Το αν θα συμπεριληφθεί μια παρατήρηση στο δείγμα, δεν πρέπει να επηρεάζει το αν θα συμπεριληφθεί κάποια άλλη.
- **Κανονικότητα.** Θα πρέπει να ακολουθείται κανονική κατανομή της εξαρτημένης μεταβλητής για κάθε συνδυασμό τιμών των ανεξαρτήτων μεταβλητών.
- **Ισότητα των διακυμάνσεων ή ομοσκεδαστικότητα (homoscedasticity).** Για οποιοδήποτε συνδυασμό των τιμών των ανεξαρτήτων μεταβλητών, η κατανομή της εξαρτημένης μεταβλητής θα πρέπει να παραμένει η ίδια.
- **Γραμμικότητα.** Η μελετώμενη εξαρτημένη μεταβλητή θα πρέπει να συνδέεται γραμμικά με κάθε μια από τις ανεξάρτητες μεταβλητές. Μικρές αποκλίσεις από τη γραμμικότητα, δεν πρέπει να μας ανησυχούν. Μέτριες ή ακραίες αποκλίσεις από την γραμμικότητα, συνήθως οδηγούν σε υποτίμηση της σχέσης των μεταβλητών που μας ενδιαφέρουν.

- **Πολύ- συγγραμμικότητα (multicollinearity) και ιδιομορφία (singularity).** Πολύ- συγγραμμικότητα παρατηρείται όταν υπάρχει ισχυρή συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών, ενώ εάν οι ανεξάρτητες μεταβλητές συσχετίζονται τέλεια μεταξύ τους, τότε παρατηρείται το φαινόμενο της ιδιομορφίας. Τέτοιου είδους φαινόμενα καλό είναι να αποφεύγονται και να μην υπάρχουν στο υπό ανάπτυξη μοντέλο.

Στο πλαίσιο κατασκευής ενός παλινδρομικού μοντέλου με χρήση του λογισμικού SPSS, οι διαθέσιμες μέθοδοι επιλογής των ανεξάρτητων μεταβλητών, είναι ενδεικτικά οι εξής: **Forward, Backward, Stepwise, Enter, Remove, Test.** Στις συγκεκριμένες μεθόδους πραγματοποιείται πρόσθεση ή/ και αφαίρεση μεταβλητών, με βάση την αλλαγή που επιφέρεται στον **συντελεστή πολλαπλού προσδιορισμού R^2 (coefficient of determination)**. Η πλέον δημοφιλής μέθοδος από τις προαναφερόμενες είναι η μέθοδος βηματικής επιλογής μεταβλητών (Stepwise).

3.1.1 Μέθοδος βηματικής επιλογής μεταβλητών (Stepwise)

Η **Μέθοδος Stepwise** (μέθοδος βηματικής επιλογής μεταβλητών) ξεκινά επιλέγοντας τη μεταβλητή που επιφέρει τη μεγαλύτερη αύξηση στον R^2 .

Στο επόμενο βήμα προστίθεται μια ακόμα μεταβλητή που επιφέρει τη μεγαλύτερη αύξηση στον R^2 , δεδομένου ότι η αλλαγή στον R^2 είναι αρκετά μεγάλη, ώστε να απορριφθεί η μηδενική υπόθεση ότι η πραγματική αλλαγή είναι μηδέν, χρησιμοποιώντας ένα προκαθορισμένο επίπεδο στατιστικής σημαντικότητας (κριτήριο εισόδου, το οποίο είναι συνήθως 0,05).

Στη συνέχεια πραγματοποιείται έλεγχος των δύο μεταβλητών, για το εάν για κάποια από αυτές βρίσκεται ισχύ το κριτήριο εξόδου. Στην ουσία, αφαιρείται η μεταβλητή που επιφέρει τη μικρότερη αύξηση στον R^2 , δεδομένου ότι η αλλαγή που επιφέρει είναι τόσο μικρή, ώστε να μην απορρίπτεται η μηδενική υπόθεση, ότι η πραγματική αλλαγή είναι μηδέν χρησιμοποιώντας ένα προκαθορισμένο επίπεδο (κριτήριο εξόδου, το οποίο

είναι συνήθως 0,10). Η διαδικασία συνεχίζεται μέχρι ότου να μην υπάρχουν άλλες μεταβλητές που να συνεισφέρουν σημαντική αύξηση στον R^2 .

Όταν πραγματοποιείται ανάλυση πολλαπλής γραμμικής παλινδρόμησης, βασικός στόχος είναι η κατασκευή του «καλύτερου δυνατού» παλινδρομικού μοντέλου. Γίνεται προσπάθεια να κατασκευαστεί ένα μοντέλο που θα έχει το μικρότερο δυνατό αριθμό ανεξαρτήτων μεταβλητών, αλλά από το οποίο δεν θα απουσιάζουν σημαντικές ανεξάρτητες μεταβλητές. Ο λόγος που είναι επιθυμητό αυτό, είναι για να αποφευχθούν φαινόμενα **underfitting** (μεροληπτική πρόβλεψη) και **overfitting** (με μεγάλες διακυμάνσεις). Υπάρχουν δείκτες που ελαχιστοποιούν τα ποσά τέτοιων καταστάσεων, όπως ο **δείκτης (κριτήριο) Mallows**. Το SPSS, δεν έχει τη δυνατότητα να δώσει, μέσω ενός μενού, την τιμή του δείκτη. Μπορεί να υπολογισθεί «τρέχοντας» ένα αρχείο syntax.

Τέλος, κατά τη διεξαγωγή μιας ανάλυσης πολλαπλής γραμμικής παλινδρόμησης, θα πρέπει να συμπεριληφθούν άλλοι ακόμη δύο παράγοντες/προϋποθέσεις:

1. **Ακραίες τιμές (outliers) και παρατηρήσεις επίδρασης (influential points)**. Πολλές φορές η ανάλυση παλινδρόμησης είναι ευαίσθητη σε παρατηρήσεις, οι τιμές των οποίων διαφέρουν σημαντικά από τις τιμές των υπόλοιπων παρατηρήσεων (ακραίες τιμές ή outliers). Μερικές φορές οι τιμές αυτές φτάνουν σε σημείο να επηρεάζουν σημαντικά ή ακόμα και να ανατρέπουν τα αποτελέσματα της ανάλυσης (παρατηρήσεις επίδρασης ή influential points). Η διαγραφή των τιμών αυτών δεν είναι πάντα η καλύτερη δυνατή λύση. Το SPSS διαθέτει δείκτες, όπως οι αποστάσεις Mahalanobis, Leverage και Cook οι οποίες ανιχνεύουν την ύπαρξη τέτοιων τιμών. Τόσο όμως η ανίχνευση όσο και η αντιμετώπιση τέτοιων τιμών είναι τις περισσότερες φορές διαδικασία περίπλοκη.
2. Ο **βαθμός γενίκευσης** του παλινδρομικού μοντέλου. Θα πρέπει να εξετάζετε εάν η παλινδρομική εξίσωση είναι σε θέση να προβλέψει το

ίδιο καλά, εφόσον εφαρμοστεί σε διάφορα άλλα ανεξάρτητα δείγματα δεδομένων. Ένας τέτοιου είδους έλεγχος μπορεί να γίνει με τη βοήθεια του SPSS και της λεγόμενης **μεθόδου διαχωρισμού των δεδομένων (data splitting)**.

3.2 Ανάπτυξη ΤΝΔ

3.2.1 Σενάρια ανάπτυξης ΤΝΔ

Τα βήματα που ακολουθήθηκαν κατά τη διάρκεια αυτής της ΔΔ περιλαμβάνουν τη συλλογή πραγματικών δεδομένων, τη στατιστική ανάλυσή τους και τέλος την κατασκευή μοντέλων πρόβλεψης ΤΝΔ. Η κατασκευή ενός μοντέλου χρησιμοποιώντας ένα τεχνητό νευρωνικό δίκτυο απαιτεί μικρό υπολογιστικό χρόνο και η νευρωνική προσέγγιση διαχειρίζεται επιτυχώς μη γραμμικές σχέσεις μεταξύ των εισόδων και των εξόδων του συστήματος (Wu and Lo, 2010).

Δεκαεννέα διαφορετικά σενάρια ΤΝΔ εξετάστηκαν χρησιμοποιώντας το **Neural Fitting Tool (nftool)** της **MATLAB R2019a**. Συγκεκριμένα, χρησιμοποιήθηκαν **1,188 τιμές ανά μεταβλητή** με τυχαία διαίρεση μεταξύ των διαδικασιών εκπαίδευσης, επικύρωσης και δοκιμής των ΤΝΔ. Τα διαθέσιμα δεδομένα κατανεμήθηκαν:

- 70% για **εκπαίδευση/ training** (832 μεμονωμένες τιμές),
- 15% για **επικύρωση/ validation** (178 μεμονωμένες τιμές) και
- 15% για **δοκιμή/testing** (178 μεμονωμένες τιμές) των αναπτυσσόμενων μοντέλων ANN.

Κατά την διάρκεια εκπαίδευσης, το ΤΝΔ χρησιμοποίησε τα δεδομένα εκπαίδευσης για να προσαρμόσει τα **βάρη**, τα οποία συνδέουν το **κόμβους**, με στόχο την ελαχιστοποίηση του **σφάλματος** μεταξύ στόχων και προσομοιωμένων τιμών. Για να αποφευχθεί η **υπερ- προσαρμογή (overfitting)**, ο αλγόριθμος χρησιμοποίησε την **πρώρη διακοπή (early stopping)**, η οποία κάνει χρήση των δεδομένων επικύρωσης και το αντίστοιχο σφάλμα τους. Όταν το σφάλμα των δεδομένων επικύρωσης αρχίζει να αυξάνεται, ενώ το σφάλμα των δεδομένων εκπαίδευσης συνεχίζει να μειώνεται,

τότε θεωρείται ότι η γενίκευση του ΤΝΔ έχει συγκλίνει και η εκπαίδευση σταματά. Άλλα κριτήρια πρόωρης διακοπής περιλαμβάνουν επίσης ένα μέγιστο αριθμός **εποχών/epochs** (για να αποφευχθούν άπειρα **runs/τρεξίματα**) και μια ελάχιστη τιμή κλίσης (για αποφυγή τρεξίματος/ running μετά τις ρυθμίσεις των βαρών σταματούν να τροποποιούνται με την πάροδο του χρόνου). Τα δεδομένα δοκιμών δεν έχουν καμία επίδραση στην εκπαίδευση και την επικύρωση του ΤΝΔ, γι' αυτό τα σφάλματα εκπαίδευσης παρέχουν μια ανεξάρτητη μέτρηση της απόδοσης του ΤΝΔ και της ικανότητας γενίκευσης.

Για την εκπαίδευση των ΤΝΔ, επιλέχθηκε ο αλγόριθμος **Levenberg–Marquardt**, καθώς είναι ένας γρήγορος αλγόριθμος σύγκλισης, που χρησιμοποιείται ευρέως από την επιστημονική κοινότητα (Aalipour et al., 2022; Azeem et al., 2023; Hassan et al., 2024; Mu'azu, 2022; Özdoğan et al., 2024; Žic and Pereverzyev, 2023).

Αυτός ο αλγόριθμος τυπικά απαιτεί περισσότερη μνήμη, αλλά στο πλαίσιο αυτής της ΔΔ, οι ανάγκες μνήμης δεν υπερέβαιναν τις προδιαγραφές ενός προσωπικού υπολογιστή μεσαίας κατηγορίας. Ο αλγόριθμος υπερέχει στην ελαχιστοποίηση των σφαλμάτων εκπαίδευσης του ΤΝΔ, αλλά για να αποφευχθεί η υπερ-εκπαίδευση και να εξασφαλιστούν οι δυνατότητες παρεμβολής και παρεκβολής του μοντέλου, χρησιμοποιήθηκε ένας μηχανισμός πρόωρης διακοπής. Αυτός ο μηχανισμός πρόωμης διακοπής χρησιμοποίησε το σύνολο των δεδομένων επικύρωσης για να ελέγξει πώς τα σφάλματα επικύρωσης εξελίσσονται κατά τη διάρκεια της εκπαίδευσης. Γενικά, το σφάλμα επικύρωσης μειώνεται στις πρώτες επαναλήψεις του αλγορίθμου, και σε ένα ορισμένο σημείο φτάνει στο ελάχιστο, και μετά αρχίζει να αυξάνεται και πάλι, λόγω υπερ-εκπαίδευσης του ΤΝΔ. Για την παρούσα ΔΔ, η εκπαίδευση του ΤΝΔ σταμάτησε αυτόματα, όταν το μέσο τετραγωνικό σφάλμα των δεδομένων επικύρωσης αυξήθηκε για έξι (6) συνεχόμενες φορές, και το τελικό ΤΝΔ για αυτήν την επανάληψη ήταν αυτό με τα βάρη που παρήγαγαν το ελάχιστο σφάλμα επικύρωσης.

Στο πλαίσιο εκπόνησης της παρούσας ΔΔ, χρησιμοποιήθηκαν σύνολα ΤΝΔ αντί για μεμονωμένα ΤΝΔ, για να επιτευχθούν πιο στιβαρά αποτελέσματα. Αυτά τα σύνολα αποτελούνταν από πολλά ΤΝΔ που κυμαίνονταν από 10 έως 100 ΤΝΔ με προσauξήσεις κάθε φορά των 10 δικτύων. Κάθε σύνολο είχε έναν μέγιστο αριθμό κόμβων για τα ΤΝΔ

του, που κυμαίνονται από 10 έως 100 μέγιστους κόμβους, με προσαυξήσεις των 10 κόμβων. Για κάθε μεμονωμένο ΤΝΔ, σε οποιοδήποτε από αυτά τα σύνολα, οι κόμβοι του κρυφού επιπέδου επιλέχθηκαν στη συνέχεια τυχαία μεταξύ ενός κόμβου και του μέγιστου αριθμού κόμβων (100) σε αυτό το συγκεκριμένο σύνολο.

Αναφορικά με τα διαθέσιμα δεδομένα, οι μεταβλητές ποιότητας του νερού και της λειτουργίας της EEN κατηγοριοποιήθηκαν σε τρεις (3) κατηγορίες (**Πίνακας 4**), λαμβάνοντας υπόψη τη μακρόχρονη εμπειρία του λειτουργού της εγκατάστασης:

1. Η πρώτη κατηγορία περιλαμβάνει τις μεταβλητές, η παρουσία των οποίων, με βάση την εμπειρία του λειτουργού της εγκατάστασης, απαιτείται να είναι **οπωσδήποτε παρούσες**, ως μεταβλητές εισόδου στο μοντέλο του ΤΝΔ. Οι μεταβλητές που περιλαμβάνονται στην πρώτη κατηγορία, είναι: η παροχή ακατέργαστου νερού (Q), η θολότητα ακατέργαστου νερού (T_1), η θολότητα επεξεργασμένου νερού (T_2), η συγκέντρωση του ελεύθερου υπολειμματικού χλωρίου στο επεξεργασμένο νερό (Cl_2), η συγκέντρωση Αργιλίου (Al) στο επεξεργασμένο νερό και η θολότητα του νερού εισόδου στις κλίνες διήθησης (T_3).
2. Η δεύτερη ομάδα περιλαμβάνει εκείνες τις μεταβλητές που είναι **πιθανώς απαραίτητες** ως είσοδοι στο μοντέλο του ΤΝΔ. Οι μεταβλητές που περιλαμβάνονται στη δεύτερη ομάδα είναι: η ημερήσια διαφορά ύψους νερού του ταμιευτήρα (ΔH), η τιμή pH του ακατέργαστου νερού (pH_1), η τιμή pH επεξεργασμένου νερού (pH_2) και η καθημερινή κατανάλωση ηλεκτρικής ενέργειας στην EEN (EI).
3. Η τρίτη ομάδα περιλαμβάνει τις **μεταβλητές εξόδου** του μοντέλου ΤΝΔ. Οι παράμετροι που περιλαμβάνονται στην τρίτη ομάδα είναι: η συγκέντρωση του υπολειμματικού όζοντος (O_3) μετά τη διεργασία της οζόνωσης, η δόση του ανιονικού πολυηλεκτρολύτη (ANPE), η δόση του χλωριούχου θεικού πολύ- αργιλίου (PACI) και η παροχή αερίου χλωρίου ($Cl_{2(g)}$).

Πίνακας 4: Κατηγοριοποίηση μεταβλητών

No	Variable	Οπωσδήποτε παρούσες μεταβλητές	Πιθανώς απαραίτητες μεταβλητές	Μεταβλητές Εξόδου
1	Ημερήσια διαφορά ύψους νερού στον ταμιευτήρα (ΔH)		<input checked="" type="checkbox"/>	
2	Παροχή ανεπεξέργαστου νερού (Q)	<input checked="" type="checkbox"/>		
3	Θολότητα ανεπεξέργαστου νερού (T_1)	<input checked="" type="checkbox"/>		
4	pH ανεπεξέργαστου νερού (pH_1)		<input checked="" type="checkbox"/>	
5	Θολότητα επεξεργασμένου νερού (T_2)	<input checked="" type="checkbox"/>		
6	pH επεξεργασμένου νερού (pH_2)		<input checked="" type="checkbox"/>	
7	Συγκέντρωση ελ. υπολειμμ. χλωρίου επεξεργασμένου νερού (Cl_2)	<input checked="" type="checkbox"/>		
8	Συγκέντρωση υπολειμμ. Αργιλίου επεξεργασμένου νερού (Al)	<input checked="" type="checkbox"/>		
9	Ημερήσια κατ/μενη ηλεκτρική ενέργεια στην EEN (EI)		<input checked="" type="checkbox"/>	
10	Συγκέντρωση υπολειμματικού όζοντος (O_3)			<input checked="" type="checkbox"/>
11	Δόση ανιονικού πολυηλεκτρολύτη (ANPE)			<input checked="" type="checkbox"/>
12	Δόση χλωριούχου θεικού πολύ- αργιλίου (PACI)			<input checked="" type="checkbox"/>
13	Παροχή αερίου χλωρίου ($Cl_{2(g)}$)			<input checked="" type="checkbox"/>
14	Θολότητα νερού εισόδου κλινών διήθησης (T_3)	<input checked="" type="checkbox"/>		

3.2.2 Κώδικας στη Matlab

Σχεδιάσθηκε και αναπτύχθηκε κατάλληλος κώδικας στη **Matlab**, ο οποίος έτρεχε δεκαέξι (16) διαφορετικούς συνδυασμούς των μεταβλητών της πρώτης (1^{ης}) ομάδας με τις μεταβλητές της δεύτερης (2^{ης}) ομάδας του **Πίνακα 4**, σύμφωνα με τον καμβά συνδυασμών, που προέκυψε και αποτυπώνεται στον **Πίνακα 5**. Το διάγραμμα ροής του αλγόριθμου (κώδικα) που αναπτύχθηκε στη **Matlab**, απεικονίζεται στο **Σχήμα 12** και ο κώδικας παρατίθεται στο **Παράρτημα II** της παρούσας ΔΔ.

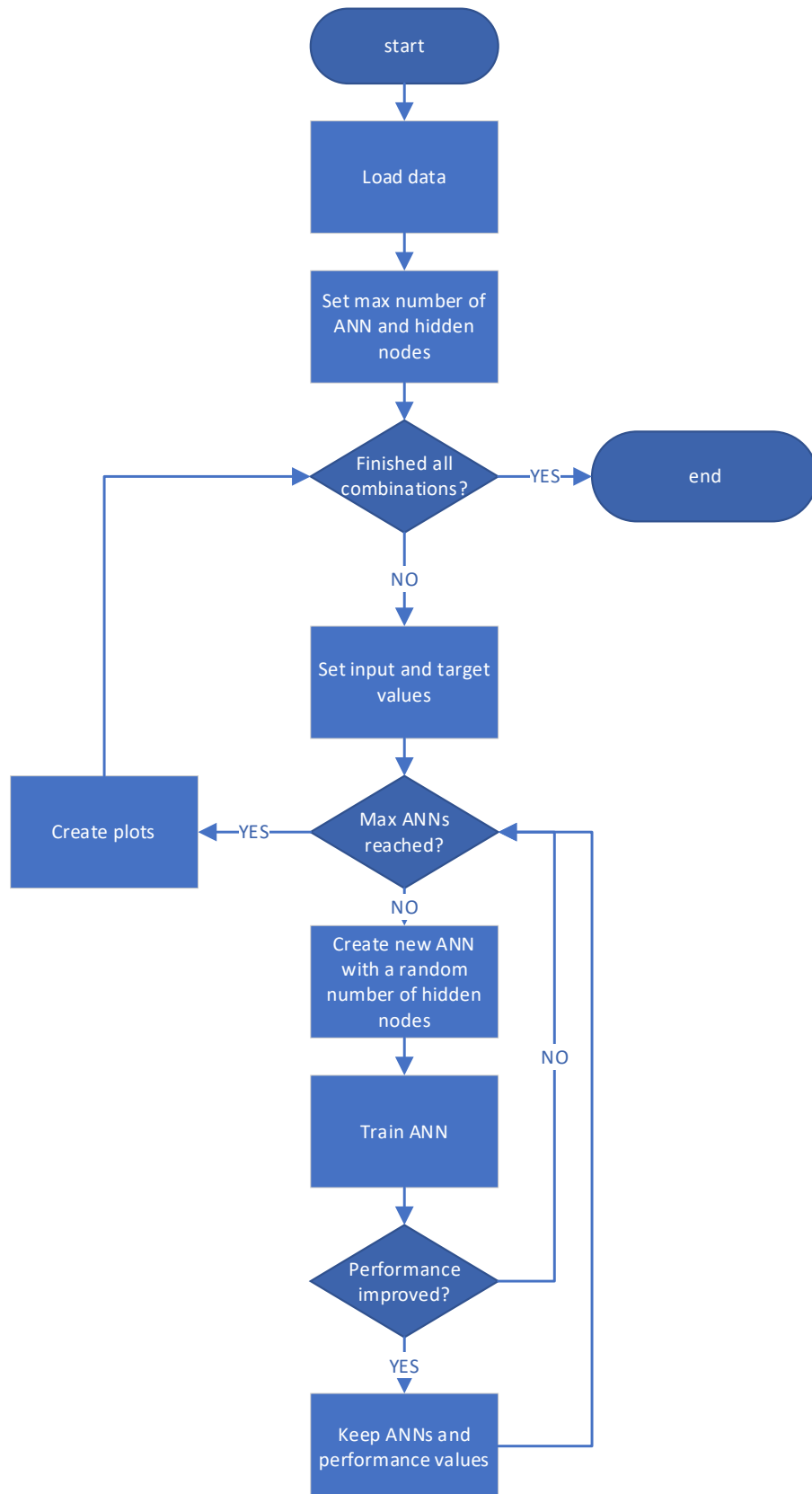
Σε αυτή τη μελέτη συγκριτικής αξιολόγησης, στο πλαίσιο της παρούσας ΔΔ, εξετάσθηκαν δεκαεννέα (19) διαφορετικά σενάρια (**Πίνακας 6**), τα οποία βασιζόταν στον αριθμό των νευρωνικών δικτύων και στον αριθμό των κόμβων, με 16 διαφορετικούς συνδυασμούς των μεταβλητών (**Πίνακας 5**), σε κάθε σενάριο ΤΝΔ. Συνολικά, εξετάσθηκαν 304 διαφορετικές περιπτώσεις ΤΝΔ.

Πίνακας 5: Καμβάς περιπτώσεων συνδυασμών για τις 4 μεταβλητές που πιθανά να είναι απαραίτητες στην ανάπτυξη του μοντέλου ΤΝΔ

No	ΔΗ	pH ₁	pH ₂	ΕΙ
1	1	1	1	1
2	0	1	1	1
3	1	0	1	1
4	1	1	0	1
5	1	1	1	0
6	0	0	1	1
7	0	1	0	1
8	0	1	1	0
9	1	0	1	0
10	1	1	0	0
11	1	0	0	1
12	0	0	0	1
13	0	0	1	0
14	0	1	0	0
15	1	0	0	0
16	0	0	0	0

Πίνακας 6: Σενάρια ΤΝΔ

ΑΑ	Σενάριο ΤΝΔ	Αριθμός Νευρωνικών Δικτύων (NN)	Αριθμός Κόμβων (N)	Συνδυασμός μεταβλητών
1	NN10N10	10	10	16 διαφορετικές περιπτώσεις, σύμφωνα με τον Πίνακα 5, για κάθε σενάριο
2	NN20N10	20	10	
3	NN30N10	30	10	
4	NN40N10	40	10	
5	NN50N10	50	10	
6	NN60N10	60	10	
7	NN70N10	70	10	
8	NN80N10	80	10	
9	NN90N10	90	10	
10	NN100N10	100	10	
11	NN100N20	100	20	
12	NN100N30	100	30	
13	NN100N40	100	40	
14	NN100N50	100	50	
15	NN100N60	100	60	
16	NN100N70	100	70	
17	NN100N80	100	80	
18	NN100N90	100	90	
19	NN100N100	100	100	



Σχήμα 12: Διάγραμμα ροής κώδικα στη Matlab για την ανάπτυξη μοντέλων ΤΝΔ

3.3 Παραγοντική Ανάλυση (Factor Analysis)

Κατά τη διαδικασία συλλογής πολύ- μεταβλητών δεδομένων δεν είναι ασυνήθιστο να διαπιστωθεί ότι υπάρχει συσχέτιση μεταξύ των μεταβλητών. Η συσχέτιση αυτή ενδεχομένως να οδηγεί σε φαινόμενα πλεονασμού των πληροφοριών που παρέχονται. Στην θεωρητική ιδανική περίπτωση όπου δύο μεταβλητές συσχετίζονται τέλεια, τότε η μία από τις δύο θεωρείται πλεονάζουσα, διότι όταν είναι γνωστές οι τιμές της μίας, είναι γνωστές και οι τιμές της άλλης, χωρίς να υπάρχει καμία ελευθερία και αντίστροφα.

Στην **παραγοντική ανάλυση (Factor Analysis, FA)** πραγματοποιείται διερεύνηση του προαναφερόμενου πλεονασμού σε πολύ-μεταβλητά δεδομένα και πραγματοποιείται εξαγωγή των χαρακτηριστικών της δομής τους και μείωση του μεγέθους των δεδομένων, χωρίς παράλληλα να μειώνεται η πληροφορία που παρέχεται.

Στην περίπτωση αυτή ο αριθμός των μεταβλητών θα μπορούσε να μειωθεί με μικρή απώλεια στην πληροφορία που παρέχουν. Οι νέες αυτές μεταβλητές, δεν παρατηρούνται (**λανθάνουσες μεταβλητές/ latent variables**), αλλά παράγονται από τις αρχικές μεταβλητές και ονομάζονται **παράγοντες (factors)**.

Αυτό έχει σαν αποτέλεσμα:

α) να έχουμε τη δυνατότητα να αποθηκεύουμε ένα μικρότερο αριθμό (νέων) μεταβλητών (παράγοντες),

β) να μειώνονται οι διαστάσεις του προβλήματος,

γ) μας επιτρέπει να αναδείξουμε νέες μεταβλητές, δίνοντας τους νέα ονόματα αλλά ταυτόχρονα συσχετίζοντάς τις με κάποιες από τις αρχικές,

δ) να ερμηνεύσουμε τον κάθε παράγοντα σύμφωνα με τη σημασία των αρχικών μεταβλητών,

ε) οι παράγοντες που προκύπτουν είναι ασυσχέτιστοι με αποτέλεσμα να βοηθούν στην αντιμετώπιση του προβλήματος της πολύ- συγγραμμικότητας που μπορεί να υπήρχε κατά την ανάλυση πολλαπλής παλινδρόμησης των αρχικών μεταβλητών,

στ) οι μεταβλητές του ενδιαφέροντός μας ταξινομούνται σε συναφείς ομάδες που σχετίζονται με κάποιο χαρακτηριστικό/ έννοια που ενδιαφερόμαστε, με αποτέλεσμα να εξετάζεται η λεγόμενη εγκυρότητα κατασκευής ενός ερωτηματολογίου (*construct validity*).

Υπάρχει μία διαδικασία χειρισμού πινάκων η οποία διαδραματίζει σημαντικό ρόλο σε αρκετές πολύ- μεταβλητές μεθόδους. Η διαδικασία αυτή περιλαμβάνει τον υπολογισμό των ιδιοτιμών (*eigenvalues*) των πινάκων (γνωστές και ως *λανθάνουσες ρίζες/ latent roots*) και αντίστοιχων ιδιο- διανυσμάτων (*eigenvectors*) τους. Στην παραγοντική ανάλυση ενδιαφερόμαστε για τις ιδιοτιμές του πίνακα των συσχετίσεων των μεταβλητών. Κάθε μια από τις ιδιοτιμές του αντιστοιχίζεται σ' ένα παράγοντα.

Επιλέγοντας λιγότερες μεταβλητές (παράγοντες) από όσες είχαμε αρχικά, χάνουμε αναγκαστικά μέρος της πληροφορίας με κέρδος όμως τη μείωση των διαστάσεων του εξεταζόμενου προβλήματος και επικεντρώνεται το ενδιαφέρον σε έναν μικρότερο αριθμό παραγόντων.

Τα κριτήρια, που υπάρχουν βιβλιογραφικά, αναφορικά με τον καθορισμό του αριθμού των παραγόντων, είναι:

1. Το **ποσοστό της συνολικής διακύμανσης των δεδομένων**, για τα οποία δίνουν εξήγηση οι παράγοντες. Συνήθως θέτουμε κάποιο όριο π.χ. 70% και επιλέγουμε τόσους παράγοντες ώστε το άθροισμά τους να δίνει ερμηνεία για πιο μεγάλο ποσοστό (διακύμανσης) από το στόχο που θέσαμε.
2. Το **κριτήριο του Kaiser**. Το συγκεκριμένο κριτήριο αναφέρει ότι, επιλέγουμε τόσους παράγοντες, όσων οι ιδιοτιμές είναι μεγαλύτερες από τη μέση τους τιμή. Στην περίπτωση του λεγόμενου «πίνακα συσχετίσεων» αυτή η μέση τιμή είναι ίση με 1 και επομένως επιλέγουμε τόσους παράγοντες όσες και οι ιδιοτιμές μεγαλύτερες του 1.
3. **Ποσοστό της διακύμανσης των αρχικών μεταβλητών που ερμηνεύεται**. Το κριτήριο αυτό επιλέγει τόσους παράγοντες ώστε, για κάθε μεταβλητή να ερμηνεύεται τουλάχιστον ένα υψηλό ποσοστό της διακύμανσης της (που εμείς καθορίζουμε).
4. Το **Scree plot (διάγραμμα κρημνού)**: (διάγραμμα του αύξοντα αριθμού μιας

ιδιοτιμής σε σχέση με το μέγεθός της κατά φθίνουσα τάξη μεγέθους). Το κριτήριο αυτό προτείνει να πάρουμε τόσους παράγοντες μέχρι που το γράφημα γίνεται περίπου παράλληλο με τον οριζόντιο άξονα.

5. Bootstrap. Στην περίπτωση που δεν εμπιστευόμαστε το κριτήριο του Kaiser, τότε προχωράμε στην κατασκευή διαστημάτων εμπιστοσύνης από την κατανομή των ιδιοτιμών και διατηρούμε μόνο τις ιδιοτιμές εκείνες, για τις οποίες όλο το διάστημα εμπιστοσύνης είναι πάνω από το όριο που ορίζει το κριτήριο του Kaiser.

6. Μέθοδος του σπασμένου ραβδιού (*Broken Stick*).

Η έννοια της **περιστροφής (*Rotation*)** των αξόνων αφορά τεχνικές (αναφέρονται ως ***Varimax, Direct Oblimin, Quatrimax, Equamax, Promax***) συνυφασμένες με την παραγοντική ανάλυση. Οι τεχνικές αυτές μας βοηθούν ώστε να αναγνωρίσουμε ευκολότερα την αντιστοίχιση των αρχικών μεταβλητών με τους παράγοντες που προέκυψαν δηλαδή φροντίζει να υπάρχουν λίγες μεταβλητές, με συντελεστές (συντεταγμένες ως προς τους παράγοντες) μεγάλες απόλυτες τιμές και πολλές μεταβλητές που οι συντελεστές τους είναι κοντά στο μηδέν. Επομένως έχουμε συγκέντρωση πολλών μεταβλητών σε λίγους παράγοντες για καθένα από τους οποίους υπολογίζονται με κατάλληλες τεχνικές (αναφέρονται ως ***Regression, Bartlett, Anderson-Rubin***) τα σκορ (***factor Scores***), και τα οποία ως μια νέα μεταβλητή δύναται να χρησιμοποιηθούν για t- test, για ανάλυση διακύμανσης κ.λπ.

Η γραφική αναπαράσταση πολυδιάστατων δεδομένων είναι κάτι αρκετά δύσκολο. Αν αντί για τα αρχικά δεδομένα είναι δυνατή η αναπαράσταση των παραγόντων, επιτυγχάνεται αξιόλογη οπτική παρουσίαση των δεδομένων. Επιπλέον, η παραγοντική ανάλυση προσφέρει σημαντικά στην **εξόρυξη γνώσης (*data mining*)**. Στο ***data mining*** ενδιαφερόμαστε να εξάγουμε γνώση από μεγάλες βάσεις δεδομένων. Η δυνατότητα που μας προσφέρει η παραγοντική ανάλυση να συμπίεσουμε την παρεχόμενη πληροφορία σε όσο το δυνατόν μικρότερη διάσταση είναι πολύ σημαντική.

Αναγκαίες προϋποθέσεις στην παραγοντική ανάλυση είναι:

1. Το μέγεθος του εξεταζόμενου δείγματος να μην είναι πολύ μικρότερο των 50 περιπτώσεων και κατά προτεραιότητα μικρότερο των 100 περιπτώσεων,
2. αν οι μεταβλητές κατανέμονται κανονικά, η λύση που δίνεται ισχυροποιείται σημαντικά,
3. οι ακραίες τιμές (outliers, μεταξύ περιπτώσεων και μεταβλητών) καλό είναι να βρεθούν και να απομακρυνθούν από τα δεδομένα και
4. να υπάρχει ανεξαρτησία μεταξύ των περιπτώσεων,
5. οι μεταβλητές θα πρέπει να σχετίζονται, όχι όμως ισχυρά,
6. εφαρμόζεται σε ποσοτικές μεταβλητές (διαστημάτων ή αναλογικής κλίμακας) και σε ειδικές περιπτώσεις μεταβλητών τακτικής κλίμακας.

Τα είδη παραγοντικής ανάλυσης είναι δύο (2):

1. **Η Διερευνητική παραγοντική ανάλυση (*exploratory factor analysis*):** Εάν δοθεί ένα σύνολο συσχετιζόμενων μεταβλητών, η ανάλυση ερευνά την ύπαρξη ενός (πολύ) μικρότερου αριθμού (νέων) ασυσχέτιστων μεταβλητών (παραγόντων) οι οποίες προκύπτουν από τις αρχικές μεταβλητές.
2. **Η Επικυρωτική παραγοντική ανάλυση (*confirmatory παραγοντική ανάλυση*):** Εάν δοθεί μια υπόθεση/ θεωρία αναφορικά με ένα σύστημα/φαινόμενο η οποία βασίζεται σε m έννοιες, η ανάλυση ερευνά την απόδειξη ότι οι μεταβλητές που χρησιμοποιήθηκαν για την περιγραφή της συνοψίζονται σε αυτές τις m έννοιες (παράγοντες).

3.3.1 Η Ανάλυση κύριων συνιστωσών (*Principal Components Analysis-PCA*)

Μια από τις μεθόδους εξαγωγής παραγόντων, από ένα σύνολο πολυμεταβλητών δεδομένων, είναι και η **Ανάλυση Κύριων Συνιστωσών, ΑΚΣ (*Principal Components Analysis-PCA*)**. Επομένως η παραγοντική ανάλυση (Π.Α.) και η ανάλυση κύριων συνιστωσών (Α.Κ.Σ.) είναι δυο μέθοδοι που προσπαθούν να ανακαλύψουν μοτίβα στις συσχετίσεις ενός συνόλου μεταβλητών. Τα μοτίβα αυτά χρησιμοποιούνται

για να διαπιστώσουμε την ύπαρξη (ενός μικρότερου συνόλου) «**κρυφών**» μεταβλητών (**latent variables**) που χαρακτηρίζουν τα δεδομένα. Οι (νέες) αυτές μεταβλητές ονομάζονται στη μεν Π.Α. **παράγοντες (factors)** στη δε Α.Κ.Σ. **συνιστώσες (components)**.

Υπάρχουν όμως διαφορές ανάμεσα στις δύο τεχνικές:

1. Η ανάλυση σε κύριες συνιστώσες είναι μια σύνθετη μαθηματική διαδικασία στην οποία οι συνιστώσες υπολογίζονται ως γραμμικοί συνδυασμοί των αρχικών μεταβλητών. Αναζητούμε ποιοί είναι αυτοί οι γραμμικοί συνδυασμοί και πως συνεισφέρει κάθε μια από τις μεταβλητές στους εν λόγω συνδυασμούς. Η παραγοντική ανάλυση είναι μια σαφώς δομημένη στατιστική τεχνική, υπάρχουν δε σε αυτή, περισσότερες από μια μέθοδοι εξαγωγής παραγόντων (πέρα από την Α.Κ.Σ.). Εδώ οι αρχικές μεταβλητές θεωρούνται γραμμικοί συνδυασμοί των παραγόντων.
2. Η Π.Α. υποθέτει ρητά την ύπαρξη παραγόντων στα δεδομένα μας, ενώ η Α.Κ.Σ. αναζητά μεταβλητές που είναι σύνθεση των αρχικών μεταβλητών μας.
3. Στην Α.Κ.Σ. το ενδιαφέρον μας είναι να επεξηγήσουμε όσο μεγαλύτερο μέρος της διακύμανσης των μεταβλητών, ενώ στην Π.Α. να εξηγήσουμε την κοινή διακύμανση μεταξύ των μεταβλητών (συνδιακύμανση των μεταβλητών).
4. Χρησιμοποιούμε την Α.Κ.Σ. για να μειώσουμε τον αριθμό των αρχικών μεταβλητών (μείωση των διαστάσεων ενός προβλήματος), ενώ την Π.Α. για να κατανοήσουμε τους παράγοντες/δομή που χαρακτηρίζουν τα δεδομένα.

Η διαφορά στις λύσεις που παίρνουμε με τις δύο τεχνικές είναι (συνήθως) μικρή (ιδιαίτερα όταν το μέγεθος του δείγματος είναι μεγαλύτερο από 30 και τα **communalities** μεγαλύτερα από 0,70). Διαφορά ίσως υπάρχει για δείγματα μεγέθους μικρότερο από 20 και communalities μικρότερα από 0,40. Αν και με τις δυο μεθόδους λοιπόν μπορεί να πάρουμε διαφορετικά αποτελέσματα, μοιάζουν στο σημείο ότι το SPSS, για την εφαρμογή της Π.Α. χρησιμοποιεί ως default επιλογή εξαγωγής παραγόντων την Α.Κ.Σ.

4 ΑΠΟΤΕΛΕΣΜΑΤΑ

4.1 Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης (*Multiple Linear Regression Analysis*)

Εξετάστηκαν **τέσσερα (4)** διαφορετικά σενάρια πολλαπλής γραμμικής παλινδρόμησης, με τη μέθοδο βηματικής επιλογής μεταβλητών (**Stepwise**), με χρήση του λογισμικού SPSS (IBM SPSS Statistics, version 26):

1. Στο **1^ο σενάριο** επιλέχθηκε ως εξαρτημένη μεταβλητή το υπολειμματικό όζον (O_3) και ως ανεξάρτητες μεταβλητές όλες οι υπόλοιπες από τις μελετώμενες μεταβλητές του **Πίνακα 3**.
2. Στο **2^ο σενάριο** επιλέχθηκε ως εξαρτημένη μεταβλητή η δόση του ανιονικού πολυηλεκτρολύτη (AN PE) και ως ανεξάρτητες μεταβλητές όλες οι υπόλοιπες από τις μελετώμενες μεταβλητές του **Πίνακα 3**.
3. Στο **3^ο σενάριο** επιλέχθηκε ως εξαρτημένη μεταβλητή η δόση του χλωριούχου θειικού πολυαργιλίου (PACl) και ως ανεξάρτητες μεταβλητές όλες οι υπόλοιπες από τις μελετώμενες μεταβλητές του **Πίνακα 3**.
4. Στο **4^ο σενάριο** επιλέχθηκε ως εξαρτημένη μεταβλητή η παροχή αερίου χλωρίου ($Cl_{2(g)}$) και ως ανεξάρτητες μεταβλητές όλες οι υπόλοιπες από τις μελετώμενες μεταβλητές του **Πίνακα 3**.

Στον **Πίνακα 7** αποτυπώνονται οι τιμές του **συντελεστή πολλαπλού προσδιορισμού R^2 (coefficient of determination)** ανά σενάριο που μελετήθηκε (Gyparakis et al., 2025).

Πίνακας 7: Συντελεστής πολλαπλού προσδιορισμού (R^2) ανά μελετώμενο σενάριο

ΑΑ	Εξαρτημένη μεταβλητή	R^2
1	O ₃	0,156
2	AN PE	0,236
3	PACl	0,501
4	Cl _{2(g)}	0,681

4.1.1 Συσχέτιση μεταβλητών

Σύμφωνα με τα αποτελέσματα του **Πίνακα 7**, με βάση τις τιμές του συντελεστή πολλαπλού προσδιορισμού (R^2), από τα **τέσσερα (4)** διαφορετικά σενάρια πολλαπλής γραμμικής παλινδρόμησης, με τη μέθοδο βηματικής επιλογής μεταβλητών (**Stepwise**) που εξετάστηκαν, αξίζει να σχολιασθεί το τέταρτο (4^ο) σενάριο. Ο συντελεστής πολλαπλού προσδιορισμού (R^2), σε ένα μοντέλο παλινδρόμησης, εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής, το οποίο εξηγείται από μια ανεξάρτητη μεταβλητή. Οι τιμές του R^2 κοντά στο 1 υποδεικνύουν πολύ καλή εφαρμογή μοντέλου. Σύμφωνα με βιβλιογραφικές αναφορές, τιμές του R^2 μεγαλύτερες από 0,5 υποδηλώνουν υψηλή συσχέτιση (Baouab and Cherif, 2018; D. N. Moriasi et al., 2007; Gyparakis et al., 2024; Kim and Parnichkun, 2017a; Wang et al., 2024; Wongburi and Park, 2023).

Σύμφωνα με αυτό, είχε επιλεγεί ως εξαρτημένη μεταβλητή η παροχή αερίου χλωρίου (Cl_{2(g)}) και ως ανεξάρτητες μεταβλητές όλες οι υπόλοιπες από τις μελετώμενες μεταβλητές του **Πίνακα 3**. Στην ουσία, θέλαμε να εξετάσουμε, εάν η παροχή αερίου χλωρίου, κατά το στάδιο της χλωρίωσης του επεξεργασμένου νερού πριν αποθηκευτεί ως τελικά επεξεργασμένο νερό στη δεξαμενή καθαρού νερού (ΔΚΝ), συνδέεται με κάποια από τις υπόλοιπες 13 μεταβλητές που μελετάμε και διαθέτουμε δεδομένα.

Τα περιγραφικά στατιστικά στοιχεία των μεταβλητών που χρησιμοποιήθηκαν αποτυπώνονται στον **Πίνακα 8** που ακολουθεί.

Πίνακας 8: Περιγραφικά στατιστικά στοιχεία μεταβλητών σεναρίου $Cl_{2(g)}$

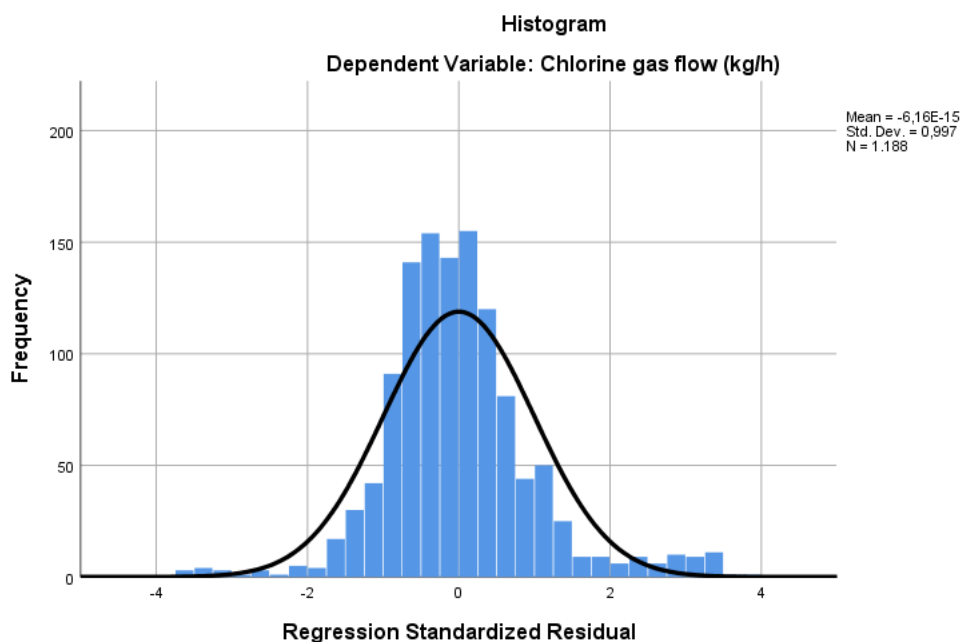
Descriptive Statistics			
	Mean	Std. Deviation	N
Chlorine gas flow (kg/h)	0,2410	0,17426	1.188
Daily reservoir water height difference (m)	0,3499	0,03229	1.188
Water flow at the entrance of the DWTP (m ³ /d)	0,5148	0,14044	1.188
Inlet water turbidity (NTU)	0,0116	0,03937	1.188
Inlet water pH	0,5590	0,18816	1.188
Outlet water turbidity (NTU)	0,2114	0,10700	1.188
Outlet water pH	0,5455	0,19884	1.188
Outlet water residual free chlorine (mg/L)	0,4798	0,12607	1.188
Outlet water residual aluminum (μg/L)	0,2516	0,15368	1.188
Daily electricity consumption (kWh)	0,4466	0,15441	1.188
Residual ozone (mg/L)	0,2254	0,10134	1.188
Anionic polyelectrolyte dosage (ppm)	0,0740	0,06096	1.188
Polyaluminum sulfate chloride dosage (ppm)	0,1181	0,12511	1.188
Filtration bed inlet water turbidity (NTU)	0,1581	0,11911	1.188

Όπως αποτυπώνεται στο **Σχήμα 13**, οι τιμές τυποποιημένων υπολοίπων της μεταβλητής της παροχής αερίου χλωρίου, ακολουθούν κανονική κατανομή, που είναι απαίτηση της πολλαπλής γραμμικής παλινδρόμησης. Επίσης, στην περίπτωση αυτή πληρούνται και όλες οι υπόλοιπες απαραίτητες προϋποθέσεις (π.χ. πολύ-συγγραμμικότητα, ομοσκεδαστικότητα κ.ά.) της πολλαπλής γραμμικής παλινδρόμησης (Gyparakis et al., 2025).

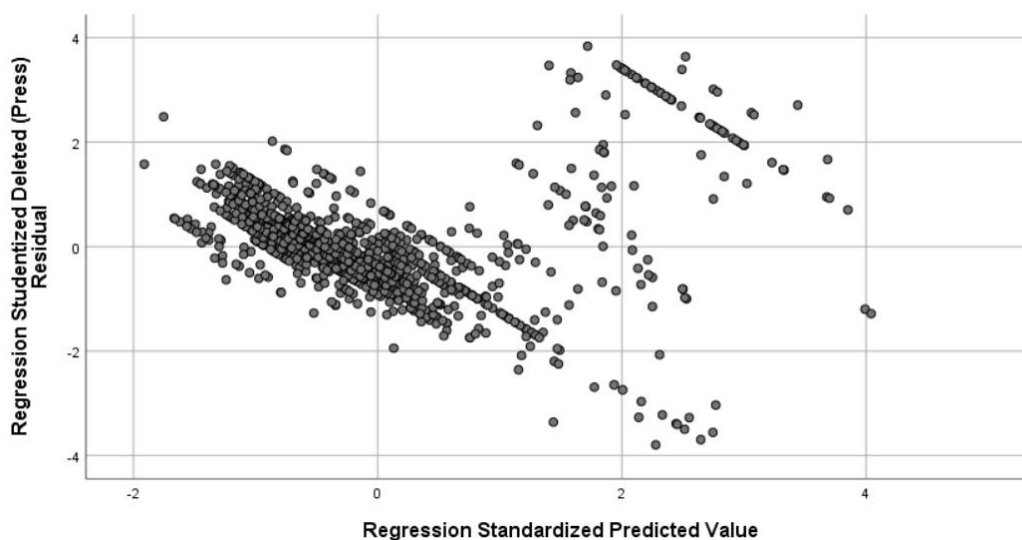
Η πολύ- συγγραμμικότητα (δηλαδή το γεγονός ότι οι ανεξάρτητες μεταβλητές

είναι (κατά ζεύγη) γραμμικά συσχετισμένες αλλά όχι έντονα ελέγχεται χρησιμοποιώντας τον δείκτη Variance Inflation Factor (VIF) ή Tolerance. Τιμές VIF μεγαλύτερες από 10 και ανοχή μικρότερες από 0,20 δείχνουν σημαντική πολύ-συγγραμμικότητα μεταξύ των μεταβλητών που αυτό θα πρέπει να διορθωθεί (δεν ισχύει στην περίπτωσή μας).

Αναφορικά με την ομοσκεδαστικότητα, η οποία αναφέρεται στην κατανομή της εξαρτημένης μεταβλητής, η οποία πρέπει να παραμένει ίδια για οποιοδήποτε συνδυασμό τιμών των ανεξάρτητων μεταβλητών, παρατηρούμε τις τιμές των διαγραμμένων υπολειμμάτων Student (**Σχήμα 14**), με μερικές εξαιρέσεις, βρίσκονται στο κλειστό διάστημα $[-2, 2]$ και κατανέμονται σχεδόν ομοιόμορφα (τυχαία) σε όλο το εύρος των προβλεπόμενων συνολικών τιμών. Άρα, η υπόθεση της ομοσκεδαστικότητας ή της ισότητας των αποκλίσεων ικανοποιείται (Gyparakis et al., 2025).



Σχήμα 13: Κανονική κατανομή τιμών τυποποιημένων υπολοίπων μεταβλητής παροχής αερίου χλωρίου



Σχήμα 14: Προβλεπόμενες τιμές (Regression Standardized) της παροχής $Cl_{2(g)}$ σε σχέση με Regression Studentized Deleted Residual

Στην πολλαπλή γραμμική παλινδρόμηση απαιτείται οι μεταβλητές να μην συνδέονται γραμμικά ισχυρά και ο συντελεστής συσχέτισης κατά Pearson θα πρέπει να είναι $<0,75$. Στον **Πίνακα 9** αποτυπώνεται η συσχέτιση κατά Pearson, στον οποίο φαίνεται ότι οι μεταβλητές δεν συνδέονται γραμμικά ισχυρά και ο συντελεστής συσχετισμού κατά Pearson είναι $<0,75$. Επίσης, παρατηρείται ότι οι προαναφερόμενες συσχετίσεις είναι παράλληλα και στατιστικά σημαντικές καθώς $\text{sig.} < 0.05$.

Επίσης, η ανάλυση πολλαπλής γραμμικής παλινδρόμησης με τη μέθοδο βηματικής επιλογής μεταβλητών (Stepwise) με χρήση του λογισμικού SPSS, πραγματοποιήθηκε σε **οκτώ (8) διαφορετικά βήματα** και στο όγδοο βήμα η τιμή του R^2 (συντελεστή πολλαπλού προσδιορισμού) ήταν ίση με 0,681, όπως αποτυπώνεται στον **Πίνακα 10**. Άρα, η συσχέτιση μεταξύ της μελετώμενης εξαρτημένης μεταβλητής, αυτής της παροχής αερίου χλωρίου και των οκτώ ανεξάρτητων μεταβλητών, είναι ισχυρή και στατιστικά σημαντική ($\text{sig.} < 0.05$).

Πίνακας 9: Συσχετίσεις σεναρίου $Cl_{2(g)}$

		Correlations													
		Chlorine gas flow (kg/h)	Daily reservoir water height difference (m)	Water flow at the entrance of the DWTP (m ³ /d)	Inlet water turbidity (NTU)	Inlet water pH	Outlet water turbidity (NTU)	Outlet water pH	Outlet water residual free chlorine (mg/L)	Outlet water residual aluminum (μg/L)	Daily electricity consumption (kWh)	Residual ozone (mg/L)	Anionic polyelectrolyte dosage (ppm)	Polyaluminum sulfate chloride dosage (ppm)	Filtration bed inlet water turbidity (NTU)
Pearson Correlation	Chlorine gas flow (kg/h)	1,000	-0,088	0,477	0,032	-0,269	0,541	-0,415	-0,258	-0,206	0,704	-0,155	-0,132	-0,264	-0,156
	Daily reservoir water height difference (m)	-0,088	1,000	-0,117	0,251	0,110	-0,069	0,077	0,077	0,022	-0,078	-0,021	0,080	0,062	0,169
	Water flow at the entrance of the DWTP (m ³ /d)	0,477	-0,117	1,000	-0,049	-0,009	0,286	-0,033	-0,127	0,056	0,490	0,062	-0,356	-0,354	-0,083
	Inlet water turbidity (NTU)	0,032	0,251	-0,049	1,000	0,070	0,091	-0,082	-0,037	-0,123	0,106	-0,032	-0,006	0,229	0,499
	Inlet water pH	-0,269	0,110	-0,009	0,070	1,000	-0,236	0,703	0,187	0,345	-0,278	0,233	-0,095	-0,281	0,002
	Outlet water turbidity (NTU)	0,541	-0,069	0,286	0,091	-0,236	1,000	-0,268	-0,148	-0,081	0,461	-0,073	-0,060	-0,059	-0,013
	Outlet water pH	-0,415	0,077	-0,033	-0,082	0,703	-0,268	1,000	0,257	0,474	-0,459	0,201	-0,063	-0,367	-0,135
	Outlet water residual free chlorine (mg/L)	-0,258	0,077	-0,127	-0,037	0,187	-0,148	0,257	1,000	0,086	-0,211	0,082	0,069	0,116	0,140
	Outlet water residual aluminum (μg/L)	-0,206	0,022	0,056	-0,123	0,345	-0,081	0,474	0,086	1,000	-0,065	-0,018	0,006	-0,373	-0,079

	Daily electricity consumption (kWh)	0,704	-0,078	0,490	0,106	-0,278	0,461	-0,459	-0,211	-0,065	1,000	-0,201	-0,237	-0,111	0,125
	Residual ozone (mg/L)	-0,155	-0,021	0,062	-0,032	0,233	-0,073	0,201	0,082	-0,018	-0,201	1,000	-0,181	0,050	0,045
	Anionic polyelectrolyte dosage (ppm)	-0,132	0,080	-0,356	-0,006	-0,095	-0,060	-0,063	0,069	0,006	-0,237	-0,181	1,000	0,136	-0,152
	Polyaluminum sulfate chloride dosage (ppm)	-0,264	0,062	-0,354	0,229	-0,281	-0,059	-0,367	0,116	-0,373	-0,111	0,050	0,136	1,000	0,458
	Filtration bed inlet water turbidity (NTU)	-0,156	0,169	-0,083	0,499	0,002	-0,013	-0,135	0,140	-0,079	0,125	0,045	-0,152	0,458	1,000
Sig. (1-tailed)	Chlorine gas flow (kg/h)		0,001	0,000	0,135	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Daily reservoir water height difference (m)	0,001		0,000	0,000	0,000	0,009	0,004	0,004	0,228	0,003	0,235	0,003	0,016	0,000
	Water flow at the entrance of the DWTP (m3/d)	0,000	0,000		0,044	0,385	0,000	0,128	0,000	0,027	0,000	0,016	0,000	0,000	0,002
	Inlet water turbidity (NTU)	0,135	0,000	0,044		0,008	0,001	0,002	0,100	0,000	0,000	0,132	0,414	0,000	0,000
	Inlet water pH	0,000	0,000	0,385	0,008		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,478
	Outlet water turbidity (NTU)	0,000	0,009	0,000	0,001	0,000		0,000	0,000	0,003	0,000	0,006	0,020	0,021	0,326
	Outlet water pH	0,000	0,004	0,128	0,002	0,000	0,000		0,000	0,000	0,000	0,000	0,015	0,000	0,000

Outlet water residual free chlorine (mg/L)	0,000	0,004	0,000	0,100	0,000	0,000	0,000		0,002	0,000	0,002	0,009	0,000	0,000
Outlet water residual aluminum (µg/L)	0,000	0,228	0,027	0,000	0,000	0,003	0,000	0,002		0,013	0,263	0,417	0,000	0,003
Daily electricity consumption (kWh)	0,000	0,003	0,000	0,000	0,000	0,000	0,000	0,000	,0013		0,000	0,000	0,000	0,000
Residual ozone (mg/L)	0,000	0,235	0,016	0,132	0,000	0,006	0,000	0,002	0,263	0,000		0,000	0,043	0,062
Anionic polyelectrolyte dosage (ppm)	0,000	0,003	0,000	0,414	0,000	0,020	0,015	0,009	0,417	0,000	0,000		0,000	0,000
Polyaluminum sulfate chloride dosage (ppm)	0,000	0,016	0,000	0,000	0,000	0,021	0,000	0,000	0,000	0,000	0,043	0,000		0,000
Filtration bed inlet water turbidity (NTU)	0,000	0,000	0,002	0,000	0,478	0,326	0,000	0,000	0,003	0,000	0,062	0,000	0,000	

Πίνακας 10: Μοντέλο MLR Analysis σεναρίου $Cl_{2(g)}$

Model Summary ⁱ										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	Change Statistics		Sig. F Change	Durbin-Watson
							df1	df2		
1	0,704 ^a	0,495	0,495	0,12384	0,495	1.164,538	1	1.186	0,000	
2	0,745 ^b	0,556	0,555	0,11625	0,060	160,779	1	1.185	0,000	
3	0,779 ^c	0,606	0,605	0,10950	0,050	151,557	1	1.184	0,000	
4	0,796 ^d	0,633	0,632	0,10571	0,027	87,586	1	1.183	0,000	
5	0,813 ^e	0,661	0,659	0,10169	0,028	96,331	1	1.182	0,000	
6	0,819 ^f	0,671	0,669	0,10024	0,010	35,414	1	1.181	0,000	
7	0,823 ^g	0,677	0,675	0,09928	0,007	23,971	1	1.180	0,000	
8	0,825 ^h	0,681	0,679	0,09879	0,003	12,828	1	1.179	0,000	0,425

a. Predictors: (Constant), Daily electricity consumption (kWh)

b. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU)

c. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU)

d. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$)

e. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm)

f. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm), Outlet water pH

g. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm), Outlet water pH, Water flow at the entrance of the DWTP (m^3/d)

h. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm), Outlet water pH, Water flow at the entrance of the DWTP (m^3/d), Inlet water turbidity (NTU)

i. Dependent Variable: Chlorine gas flow (kg/h)

Πίνακας 11: Συντελεστές μοντέλου *MLR Analysis* σεναρίου $Cl_{2(g)}$

Model		Coefficients ^a		Standardized Coefficients Beta	t	Sig.	Collinearity Statistics	
		Unstandardized Coefficients B	Std. Error				Tolerance	VIF
1	(Constant)	-0,114	0,011		-10,339	0,000		
	Daily electricity consumption (kWh)	0,794	0,023	0,704	34,125	0,000	1,000	1,000
2	(Constant)	-0,072	0,011		-6,652	0,000		
	Daily electricity consumption (kWh)	0,829	0,022	0,735	37,651	0,000	0,984	1,016
	Filtration bed inlet water turbidity (NTU)	-0,362	0,029	-0,247	-12,680	0,000	0,984	1,016
3	(Constant)	-0,104	0,011		-9,844	0,000		
	Daily electricity consumption (kWh)	0,695	0,023	0,616	29,620	0,000	0,770	1,298
	Filtration bed inlet water turbidity (NTU)	-0,335	0,027	-0,229	-12,429	0,000	0,978	1,022
	Outlet water turbidity (NTU)	0,413	0,034	0,254	12,311	0,000	0,782	1,278
4	(Constant)	-0,047	0,012		-4,016	0,000		
	Daily electricity consumption (kWh)	0,691	0,023	0,612	30,498	0,000	0,770	1,299
	Filtration bed inlet water turbidity (NTU)	-0,354	0,026	-0,242	-13,551	0,000	0,972	1,028
	Outlet water turbidity (NTU)	0,394	0,032	0,242	12,129	0,000	0,779	1,283
	Outlet water residual aluminum (µg/L)	-0,188	0,020	-0,166	-9,359	0,000	0,987	1,013
5	(Constant)	0,005	0,013		,427	0,669		
	Daily electricity consumption (kWh)	0,645	0,022	0,572	28,967	0,000	0,737	1,357
	Filtration bed inlet water turbidity (NTU)	-0,216	0,029	-0,147	-7,485	0,000	0,740	1,352
	Outlet water turbidity (NTU)	0,397	0,031	0,244	12,695	0,000	0,779	1,283
	Outlet water residual aluminum (µg/L)	-0,270	0,021	-0,238	-12,829	0,000	0,831	1,204
	Polyaluminum sulfate chloride dosage (ppm)	-0,289	0,029	-0,208	-9,815	0,000	0,640	1,562
6	(Constant)	0,099	0,020		4,953	0,000		
	Daily electricity consumption (kWh)	0,571	0,025	0,506	22,601	0,000	0,557	1,796
	Filtration bed inlet water turbidity (NTU)	-0,193	0,029	-0,132	-6,742	0,000	0,727	1,376

7	Outlet water turbidity (NTU)	0,387	0,031	0,237	12,527	0,000	0,777	1,287
	Outlet water residual aluminum (μg/L)	-0,220	0,022	-0,194	-9,821	0,000	0,713	1,402
	Polyaluminum sulfate chloride dosage (ppm)	-0,358	0,031	-0,257	-11,453	0,000	0,552	1,810
	Outlet water pH	-0,122	0,021	-0,140	-5,951	0,000	0,506	1,977
	(Constant)	0,063	0,021		2,988	0,003		
	Daily electricity consumption (kWh)	0,515	0,027	0,457	18,782	0,000	0,462	2,163
	Filtration bed inlet water turbidity (NTU)	-0,192	0,028	-0,131	-6,773	0,000	0,727	1,376
	Outlet water turbidity (NTU)	0,374	0,031	0,229	12,182	0,000	0,771	1,297
	Outlet water residual aluminum (μg/L)	-0,211	0,022	-0,186	-9,488	0,000	0,708	1,411
	Polyaluminum sulfate chloride dosage (ppm)	-0,321	0,032	-0,231	-10,072	0,000	0,521	1,918
8	Outlet water pH	-0,136	0,021	-0,155	-6,602	0,000	0,497	2,012
	Water flow at the entrance of the DWTP (m ³ /d)	0,124	0,025	0,100	4,896	0,000	0,653	1,532
	(Constant)	0,070	0,021		3,284	0,001		
	Daily electricity consumption (kWh)	0,511	0,027	0,453	18,700	0,000	0,461	2,167
	Filtration bed inlet water turbidity (NTU)	-0,243	0,032	-0,166	-7,690	0,000	0,581	1,721
	Outlet water turbidity (NTU)	0,362	0,031	0,222	11,802	0,000	0,763	1,311
	Outlet water residual aluminum (μg/L)	-0,202	0,022	-0,178	-9,064	0,000	0,699	1,430
	Polyaluminum sulfate chloride dosage (ppm)	-0,319	0,032	-0,229	-10,046	0,000	0,521	1,919
	Outlet water pH	-0,141	0,021	-0,161	-6,864	0,000	0,495	2,021
	Water flow at the entrance of the DWTP (m ³ /d)	0,130	0,025	0,105	5,127	0,000	0,650	1,537
	Inlet water turbidity (NTU)	0,305	0,085	0,069	3,582	0,000	0,729	1,371

a. Dependent Variable: Chlorine gas flow (kg/h)

Σύμφωνα με τον **Πίνακα 11**, ο οποίος περιλαμβάνει τους αντίστοιχους συντελεστές των ανεξάρτητων μεταβλητών, η εξίσωση του μαθηματικού μοντέλου πολλαπλής γραμμικής παλινδρόμησης δίνεται από την εξίσωση 10 (Gyarakis et al., 2025):

$$Y = 0.070 + 0.511 * X_1 - 0.243 * X_2 + 0.362 * X_3 - 0.202 * X_4 - 0.319 * X_5 - 0.141 * X_6 + 0.130 * X_7 + 0.305 * X_8 \quad (10)$$

Όπου:

- ✓ Y: Παροχή αερίου Cl₂ (kg/h),
- ✓ X1: Καθημερινή κατανάλωση ηλεκτρικής ενέργειας (kWh),
- ✓ X2: Θολότητα νερού εισόδου κλινών διήθησης (NTU),
- ✓ X3: Θολότητα επεξεργασμένου νερού (NTU),
- ✓ X4: Υπολειμματικό Αργίλιο επεξεργασμένου νερού (μg/L),
- ✓ X5: Δόση PACl (ppm),
- ✓ X6: pH επεξεργασμένου νερού,
- ✓ X7: Παροχή ανεπεξέργαστου νερού (m³/d),
- ✓ X8: Θολότητα ανεπεξέργαστου νερού (NTU).

Σύμφωνα με τον **Πίνακα 12** της ανάλυσης διασποράς (ANOVA) που ακολουθεί, όλο το μοντέλο είναι στατιστικά σημαντικό (Sig.<0.05).

Πίνακας 12: Πίνακας ανάλυσης διασποράς (ANOVA^a) μοντέλου MLR Analysis σεναρίου Cl_{2(g)}

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	17.859	1	17.859	1.164,538	0,000 ^b
	Residual	18.188	1.186	0,015		
	Total	36.047	1.187			
2	Regression	20.032	2	10.016	741,103	0,000 ^c
	Residual	16.015	1.185	0,014		
	Total	36.047	1.187			

3	Regression	21.849	3	7.283	607,360	0,000 ^d
	Residual	14.198	1.184	0,012		
	Total	36.047	1.187			
4	Regression	22.828	4	5,707	510.729	0,000 ^e
	Residual	13.219	1.183	0,011		
	Total	36.047	1.187			
5	Regression	23.824	5	4.765	460.775	0,000 ^f
	Residual	12.223	1.182	0,010		
	Total	36.047	1.187			
6	Regression	24.180	6	4.030	401.061	0,000 ^g
	Residual	11.867	1.181	0,010		
	Total	36.047	1.187			
7	Regression	24.416	7	3.488	353.877	0,000 ^h
	Residual	11.631	1.180	0,010		
	Total	36.047	1.187			
8	Regression	24.541	8	3.068	314.350	0,000 ⁱ
	Residual	11.505	1.179	0,010		
	Total	36.047	1.187			

a. Dependent Variable: Chlorine gas flow (kg/h)

b. Predictors: (Constant), Daily electricity consumption (kWh)

c. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU)

d. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU)

e. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$)

f. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm)

g. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm), Outlet water pH

h. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm), Outlet water pH, Water flow at the entrance of the DWTP (m^3/d)

i. Predictors: (Constant), Daily electricity consumption (kWh), Filtration bed inlet water turbidity (NTU), Outlet water turbidity (NTU), Outlet water residual aluminum ($\mu\text{g/L}$), Polyaluminum sulfate chloride dosage (ppm), Outlet water pH, Water flow at the entrance of the DWTP (m^3/d), Inlet water turbidity (NTU)

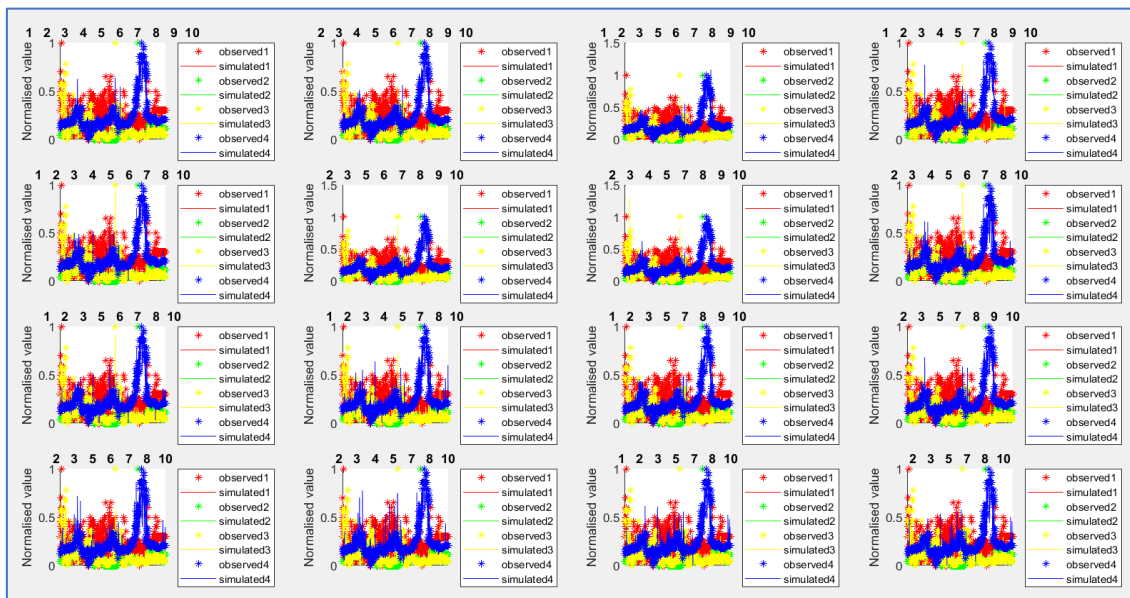
4.2 Ανάπτυξη μοντέλων ΤΝΔ

4.2.1 Εφαρμογή και ανάπτυξη μοντέλων ΤΝΔ

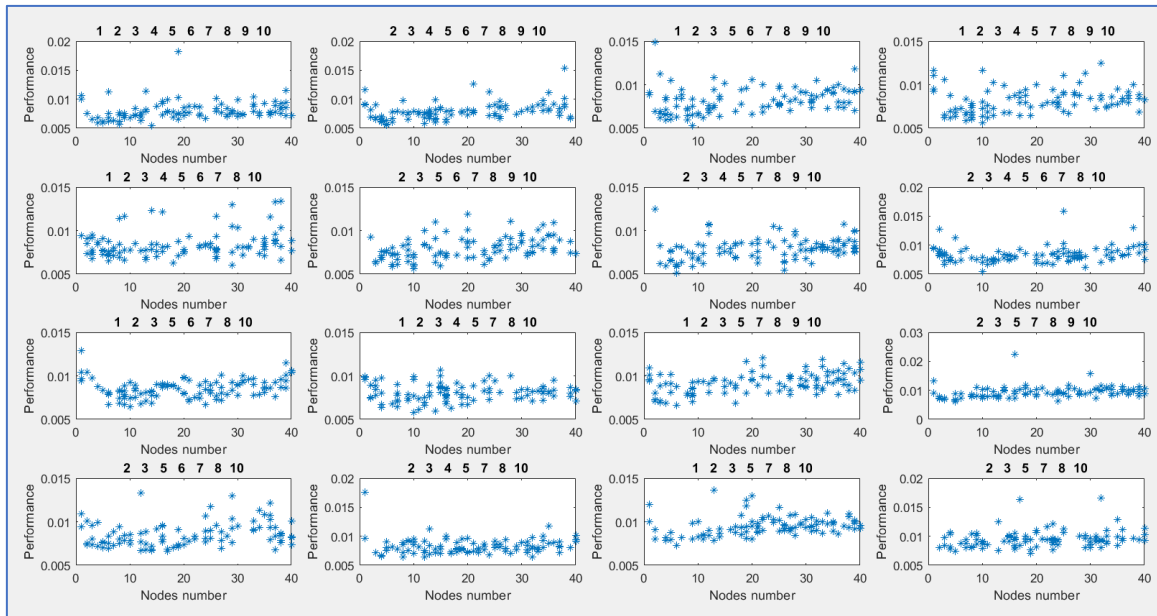
Κατά την εφαρμογή του κώδικα που αναπτύχθηκε στη Matlab για τα 19 διαφορετικά βασικά σενάρια (Πίνακας 6) με 16 διαφορετικές περιπτώσεις ανά βασικό σενάριο, προέκυψαν τα Σχήματα 15 και 16.

Στο Σχήμα 15 παρουσιάζονται, ενδεικτικά για ένα σενάριο, για κάθε μία από τις τέσσερις (4) μεταβλητές εξόδου (*outputs*) των ΤΝΔ που αναπτύχθηκαν, οι παρατηρούμενες κανονικοποιημένες τιμές των μεταβλητών (*observed values*) και οι τιμές των προσομοιώσεων (*simulated values*).

Στο Σχήμα 16 παρουσιάζονται, ενδεικτικά για ένα σενάριο, η απόδοση της προσομοίωσης των ΤΝΔ των 16 διαφορετικών περιπτώσεων του εκάστοτε σεναρίου, με βάση τον αριθμό των κόμβων.



Σχήμα 15: Παρατηρούμενες κανονικοποιημένες τιμές των μεταβλητών (*observed values*) και οι τιμές των προσομοιώσεων (*simulated values*) των ΤΝΔ



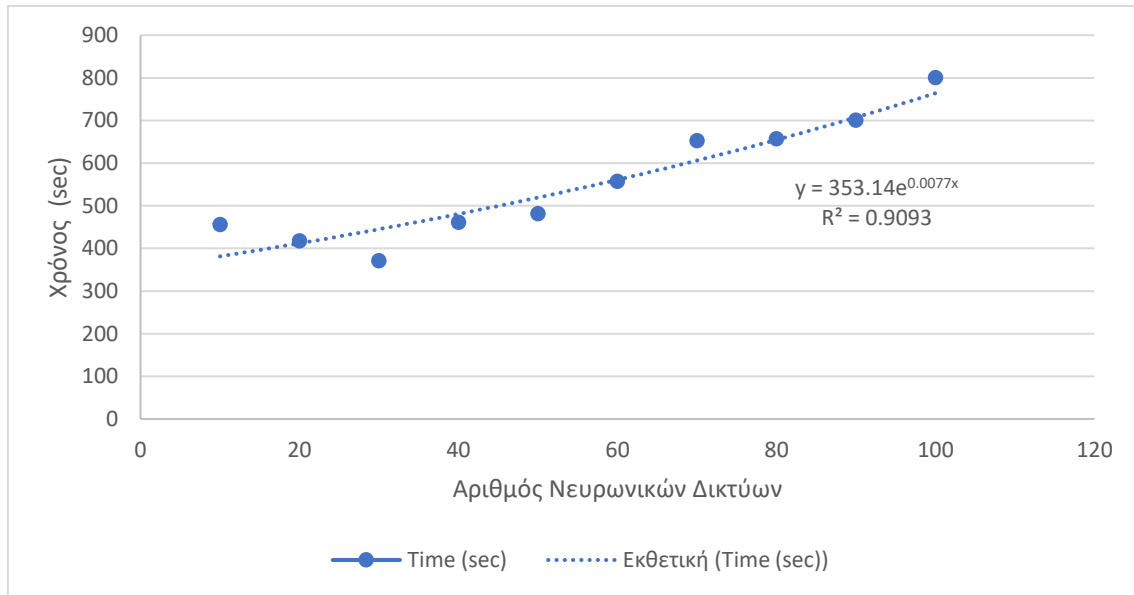
Σχήμα 16: Απόδοση της προσομοίωσης των ΤΝΔ των 16 διαφορετικών περιπτώσεων του εκάστοτε σεναρίου, με βάση τον αριθμό των κόμβων

4.2.2 Απαιτούμενος χρόνος ανάπτυξης μοντέλων ΤΝΔ

Ο χρόνος που απαιτήθηκε για την ανάπτυξη μοντέλων ΤΝΔ, ανά αριθμό Νευρωνικών Δικτύων (ΝΔ), διατηρώντας σταθερό τον αριθμό των κόμβων ($K=10$), αποτυπώνεται στον Πίνακα 13 και στο Σχήμα 17.

Πίνακας 13: Χρόνος ανάπτυξης μοντέλων ΤΝΔ, ανά αριθμό Νευρωνικών Δικτύων (ΝΔ), διατηρώντας σταθερό τον αριθμό των κόμβων ($K=10$)

ΑΑ	Σενάριο ΤΝΔ	Αριθμός Νευρωνικών Δικτύων (ΝΔ)	Αριθμός Κόμβων (Κ)	Χρόνος (sec)
1	NN10N10	10	10	456,09
2	NN20N10	20	10	417,4
3	NN30N10	30	10	371,03
4	NN40N10	40	10	461,49
5	NN50N10	50	10	481,14
6	NN60N10	60	10	557,47
7	NN70N10	70	10	652,62
8	NN80N10	80	10	657,06
9	NN90N10	90	10	700,47
10	NN100N10	100	10	800,62

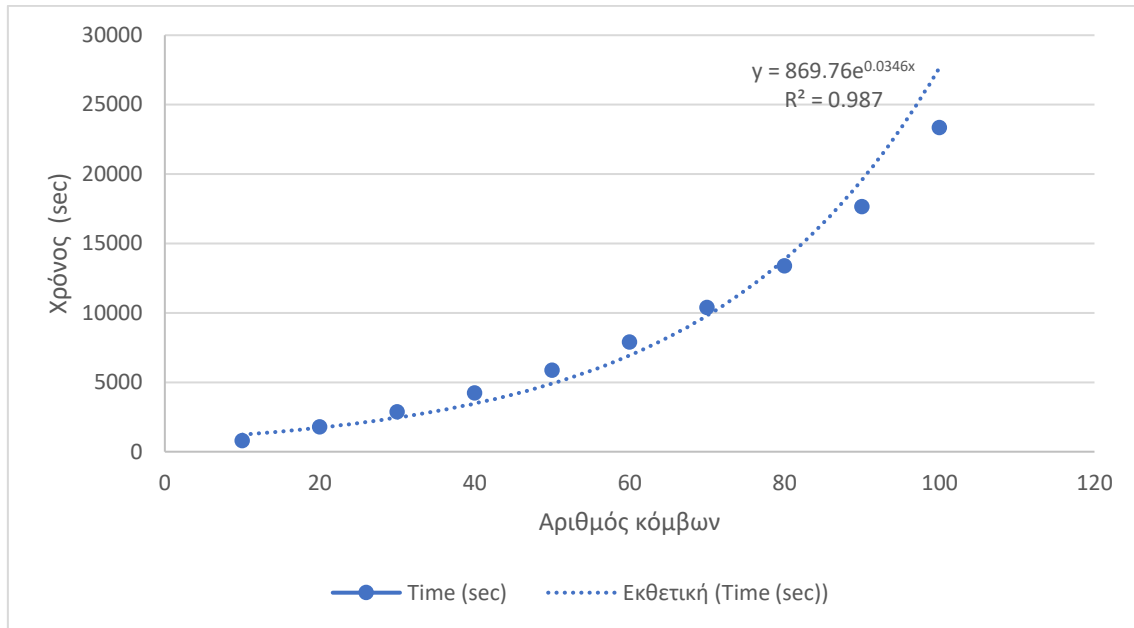


Σχήμα 17: Χρόνος ανάπτυξης μοντέλων ΤΝΔ, ανά αριθμό Νευρωνικών Δικτύων (ΝΔ), διατηρώντας σταθερό τον αριθμό των κόμβων (Κ= 10)

Ο χρόνος που απαιτήθηκε για την ανάπτυξη μοντέλων ΤΝΔ, ανά αριθμό Κόμβων (Κ), διατηρώντας σταθερό τον αριθμό των Νευρωνικών Δικτύων (ΝΔ=100), αποτυπώνεται στον **Πίνακα 14** και στο **Σχήμα 18**.

Πίνακας 14: Χρόνος ανάπτυξης μοντέλων ΤΝΔ, ανά αριθμό Κόμβων (Κ), διατηρώντας σταθερό τον αριθμό των Νευρωνικών Δικτύων (ΝΔ=100)

ΑΑ	Σενάριο ΤΝΔ	Αριθμός Νευρωνικών Δικτύων (ΝΔ)	Αριθμός Κόμβων (Κ)	Χρόνος (sec)
1	NN100N10	100	10	800,62
2	NN100N20	100	20	1.784,8
3	NN100N30	100	30	2.862,7
4	NN100N40	100	40	4.221,4
5	NN100N50	100	50	5.860,3
6	NN100N60	100	60	7.898,3
7	NN100N70	100	70	10.383
8	NN100N80	100	80	13.377
9	NN100N90	100	90	17.648
10	NN100N100	100	100	23.346



Σχήμα 18: Χρόνος ανάπτυξης μοντέλων ΤΝΔ, ανά αριθμό Κόμβων (Κ), διατηρώντας σταθερό τον αριθμό των Νευρωνικών Δικτύων (ΝΔ=100)

4.2.3 Βέλτιστο σενάριο ανάπτυξης ΤΝΔ

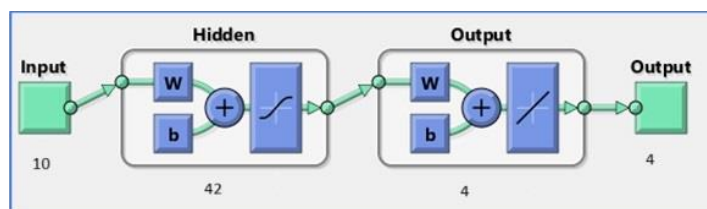
Η εκτέλεση 19 διαφορετικών σεναρίων ANN με 16 διαφορετικές περιπτώσεις ανά σενάριο (304 διαφορετικά περιπτώσεις), οδήγησε στη δημιουργία του **Πίνακα 15** με τις τιμές του δείκτη απόδοσης δοκιμής (tperf) ανά σενάριο και περίπτωση.

Ο δείκτης απόδοσης δοκιμής (tperf) είναι ο προτιμώμενος δείκτης απόδοσης, αφού δεν είναι προκατειλημμένος και τα δεδομένα που χρησιμοποιήθηκαν για τον υπολογισμό αυτού δεν χρησιμοποιήθηκαν προηγουμένως ούτε για εκπαίδευση ούτε για επικύρωση του ΤΝΔ. Υπολογίζεται ως το MSE (μέσο τετραγωνικό σφάλμα) του συνόλου των τιμών των δεδομένων δοκιμής. Το MSE είναι μια στατιστική παράμετρος και αποτελείται από τη μέση τετραγωνική διαφορά μεταξύ τιμών εκροών και στόχων. Όσο μικρότερη είναι η τιμή του δείκτη tperf, τόσο καλύτερη απόδοση, ενώ οι μηδενικές τιμές του tperf υποδηλώνουν καμία απολύτως διαφορά. Ο **Πίνακας 15** περιλαμβάνει μόνο εκείνα τα σενάρια με την ελάχιστη τιμή του δείκτη tperf, για χάριν συντομίας. Η τελική επιλογή του βέλτιστου σεναρίου μοντέλου ANN γίνεται μεταξύ αυτών με τη μικρότερη τιμή του δείκτη tperf.

Πίνακας 15: Δείκτης καλύτερης απόδοσης δοκιμών (Best testing performance) ανά σενάριο και περίπτωση ΤΝΔ

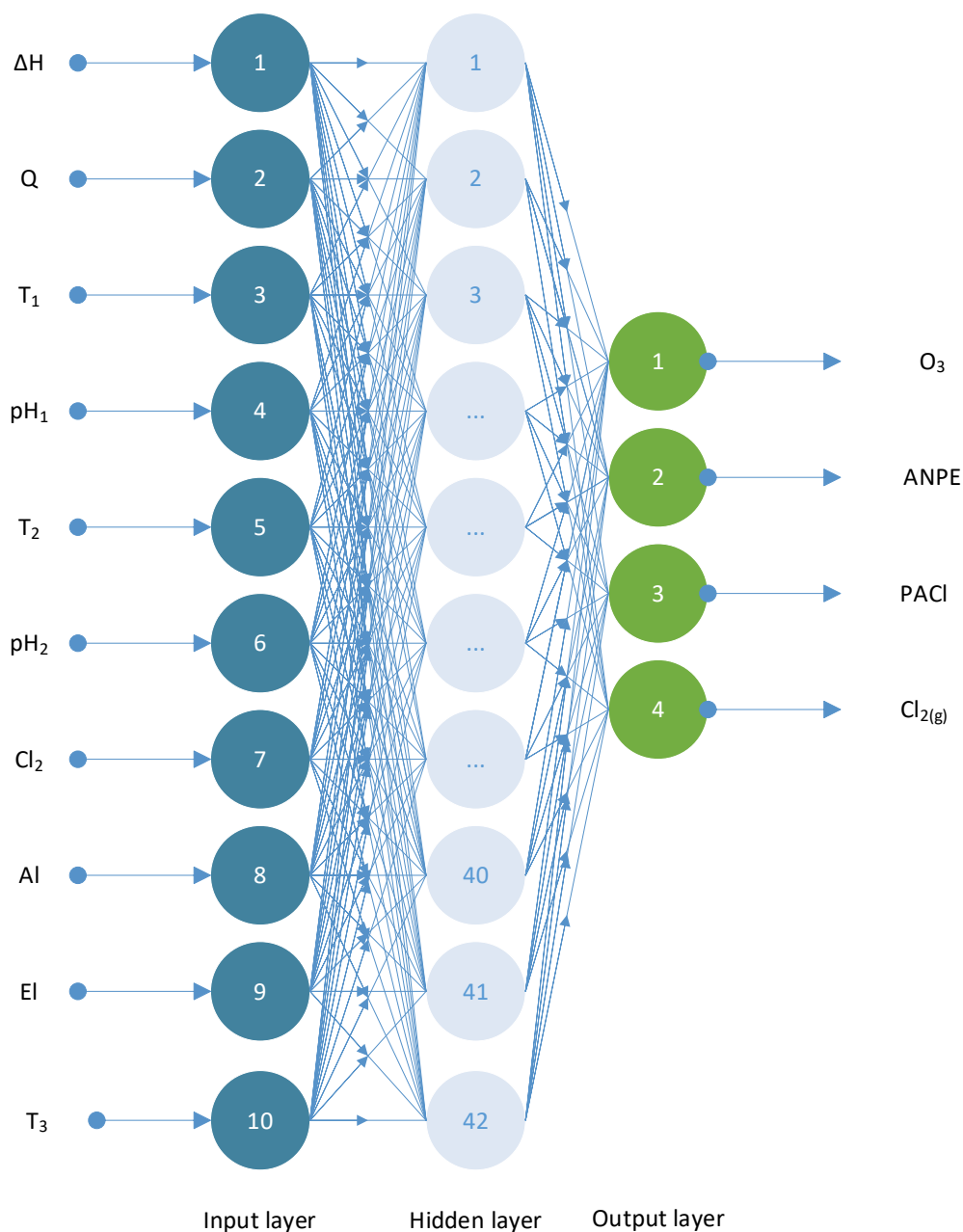
ΑΑ	Νευρωνικά Δίκτυα	Κόμβοι	Περίπτωση	Κρυφοί Κόμβοι	Outputs	Best t Perf	Time (s)
1	100	100	1	42	4	0,008848	23.346
2	40	10	7	8	4	0,008993	461
3	100	70	7	11	4	0,009211	10.383
4	100	50	13	46	4	0,009423	5.860
5	100	20	4	13	4	0,009552	17.848
6	100	10	2	8	4	0,00957	801
7	100	80	6	72	4	0,009594	13.377
8	100	30	13	19	4	0,009642	2.863
9	100	10	4	9	4	0,009658	801
10	100	60	1	12	4	0,009842	7.898
11	100	30	1	8	4	0,009897	2.863
12	80	10	1	8	4	0,009942	657

Σύμφωνα με τον **Πίνακα 15**, το βέλτιστο σενάριο που επιλέχθηκε αποτελούνταν από 100 νευρωνικά δίκτυα, 100 κόμβους, 42 κρυφούς κόμβους σε 1 κρυφό επίπεδο και ανήκε στην περίπτωση 1 [επιλέγονται δηλαδή και οι 4 παράμετροι εισόδου (ΔΗ, pH₁, pH₂, EI)]. Το βέλτιστο σενάριο ΤΝΔ απεικονίζεται στο **Σχήμα 19**.



Σχήμα 19: Δομή βέλτιστου επιλεγμένου ΤΝΔ

Το βέλτιστο μοντέλο ΤΝΔ που επιλέχθηκε βάσει τη μικρότερη τιμή του δείκτη απόδοσης δοκιμής αποτελείται από 10 εισόδους (ΔΗ, Q, T₁, pH₁, T₂, pH₂, Cl₂, AI, EI, T₃), 100 κόμβους, 42 κρυφούς κόμβους και 4 στόχους/ targets (O₃, ANPE, PACl, Cl_{2(g)}) και απεικονίζεται στο **Σχήμα 20**.

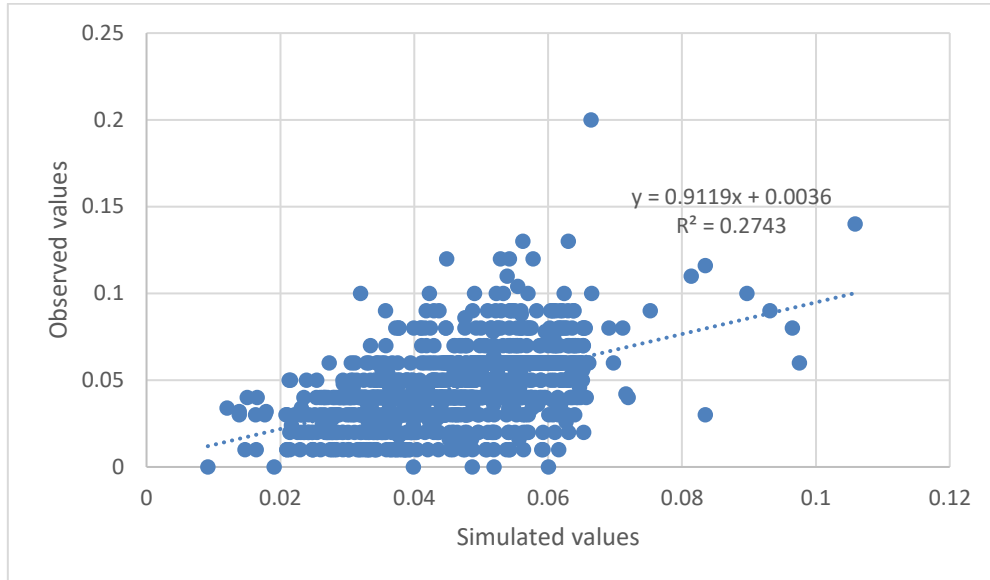


Σχήμα 20: Σχηματική απεικόνιση δομής βέλτιστου επιλεγμένου μοντέλου ΤΝΔ

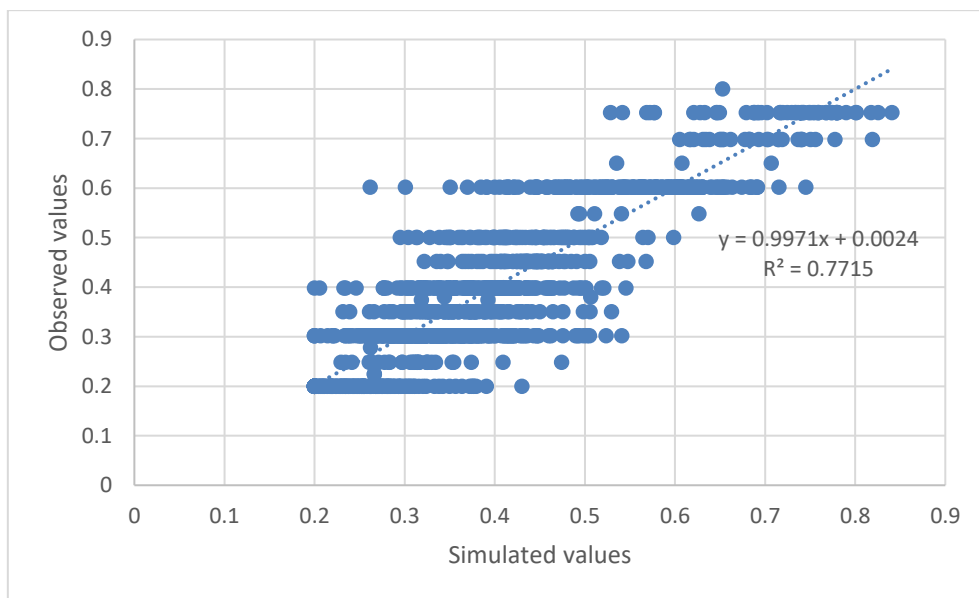
4.2.4 Παράμετροι εξόδου (outputs) ΤΝΔ

Στη συνέχεια κατασκευάστηκαν τα διαγράμματα μεταξύ των κανονικοποιημένων παρατηρούμενων (*observed*) και προσομοιωμένων (*simulated*) με το μοντέλο ΤΝΔ τιμών των τεσσάρων (4) μεταβλητών στόχων: O_3 , $ANPE$, $PACl$ και $Cl_{2(g)}$ τα οποία δίνονται στα **Σχήματα 21- 24**, αντίστοιχα. Ομοίως, τα **Σχήματα 25- 28** δείχνουν τις προσομοιωμένες (*simulated*) και παρατηρούμενες (*observed*) κανονικοποιημένες

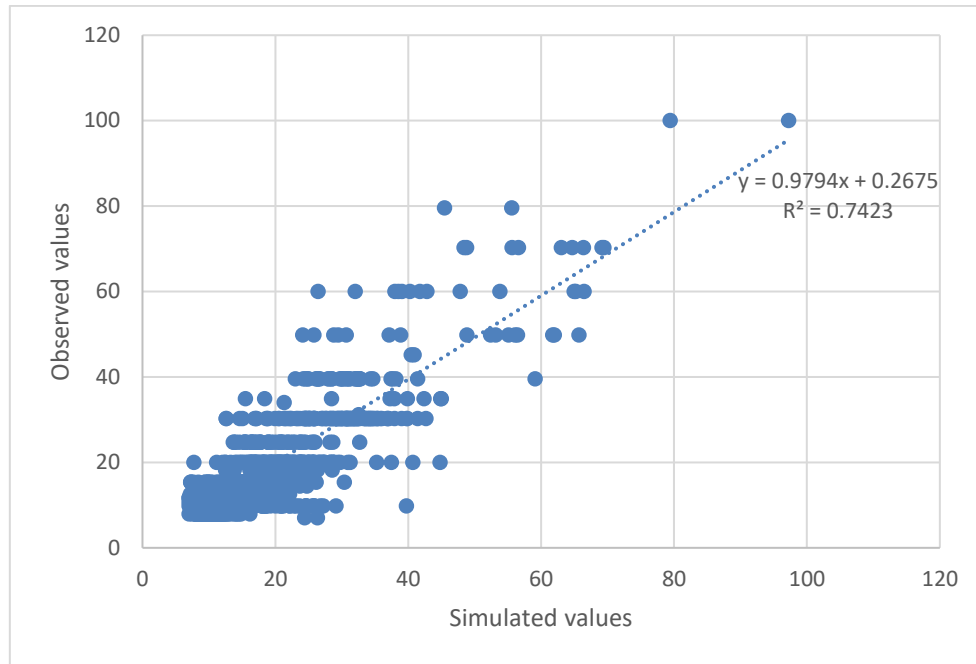
τιμές των κύριων λειτουργικών μεταβλητών, που είναι οι τέσσερις (4) έξοδοι του βέλτιστου επιλεγμένου μοντέλου ΤΝΔ.



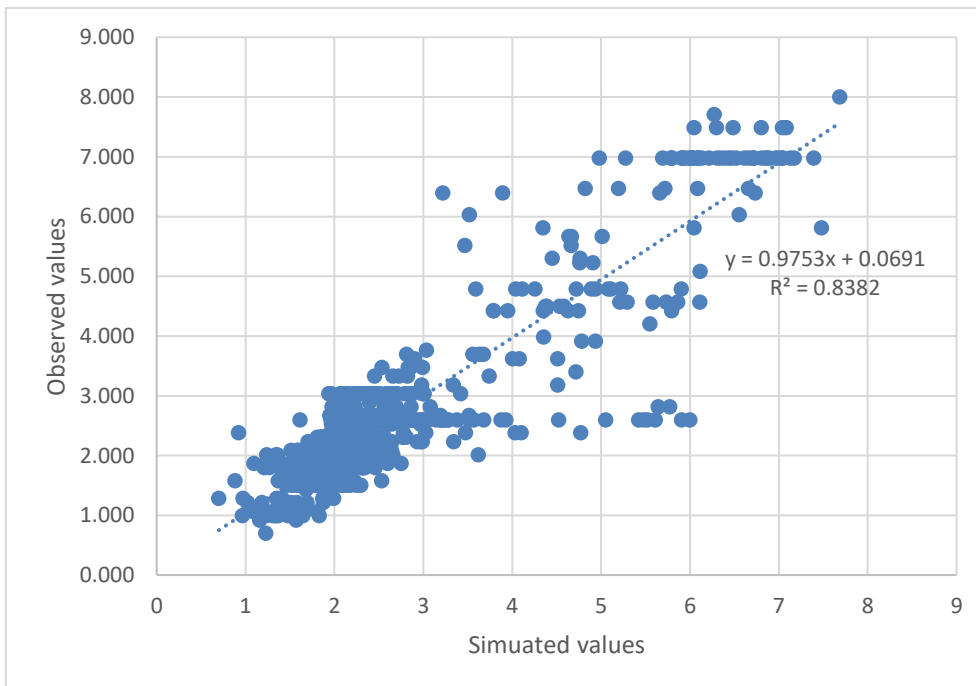
Σχήμα 21: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του O_3



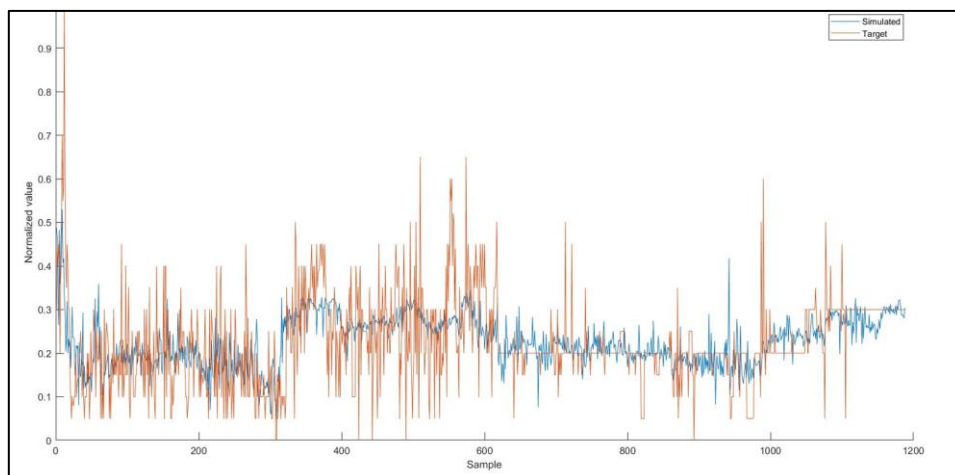
Σχήμα 22: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του AN PE



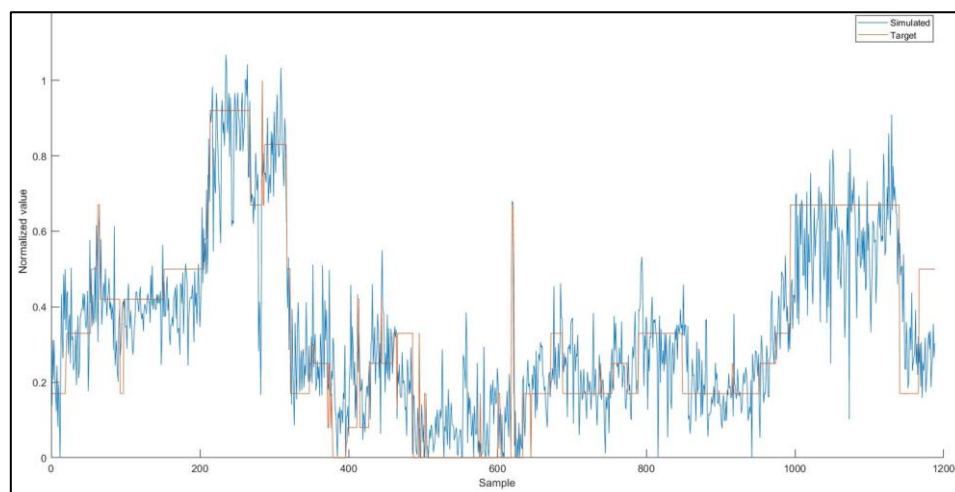
Σχήμα 23: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του PACI



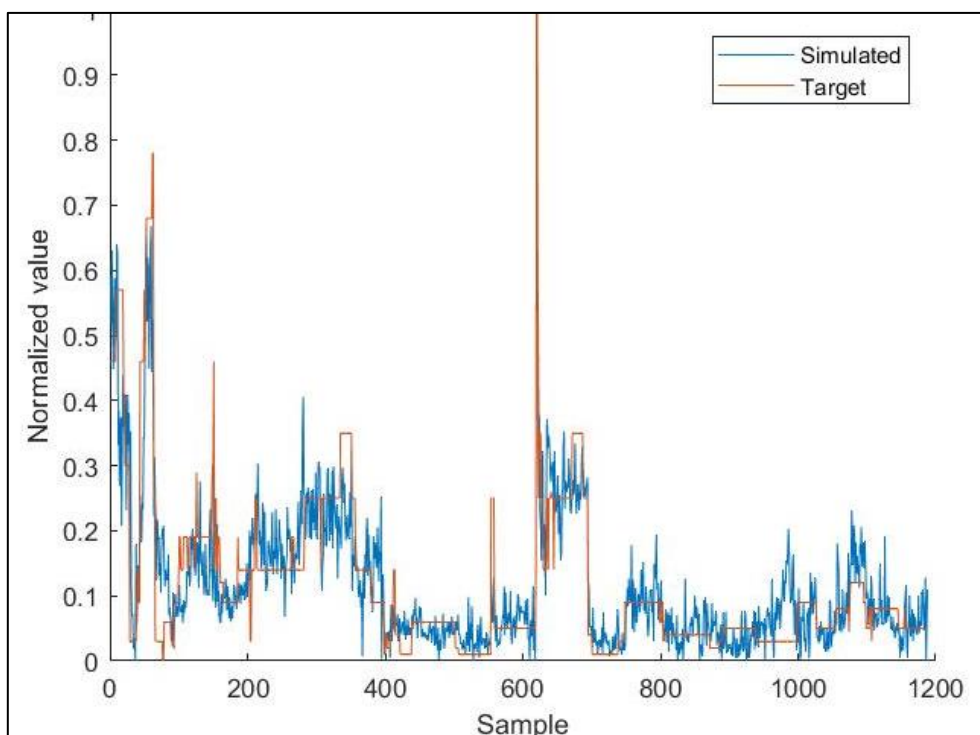
Σχήμα 24: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του $Cl_{2(g)}$



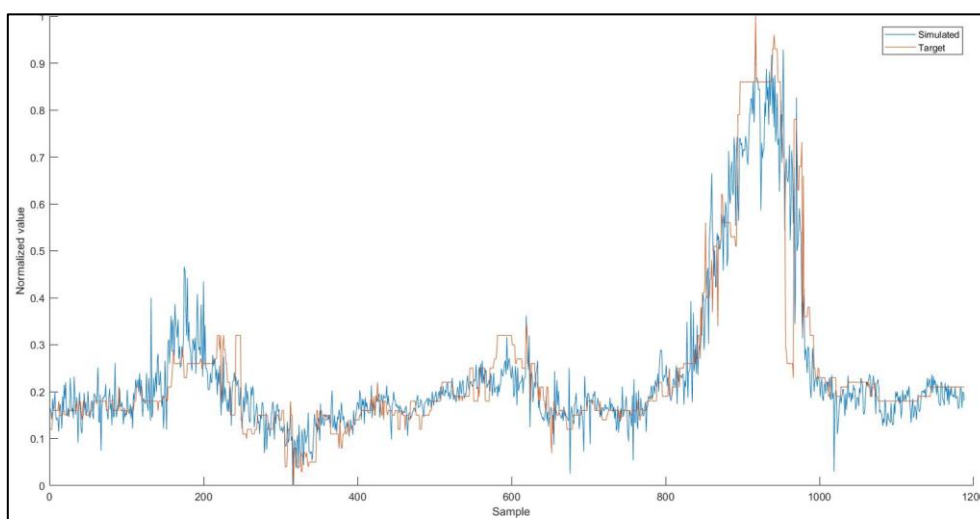
Σχήμα 25: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του O_3



Σχήμα 26: Προσομοιωμένες (simulated) και παρατηρούμενες (observed) τιμές του AN PE



Σχήμα 27: Προσομοιωμένες (*simulated*) και παρατηρούμενες (*observed*) τιμές του PACI



Σχήμα 28: Προσομοιωμένες (*simulated*) και παρατηρούμενες (*observed*) κανονικοποιημένες τιμές του $Cl_{2(g)}$

4.2.5 Εξισώσεις απο-κανονικοποίησης μεταβλητών

Όπως είχε προαναφερθεί στην παράγραφο **2.3.2 Κανονικοποίηση δεδομένων**, για να χρησιμοποιηθούν τα δεδομένα στην ανάπτυξη των ΤΝΔ, ήταν απαραίτητη η κανονικοποίηση των δεδομένων, λόγω του διαφορετικού εύρους κλίμακας μεταξύ των διαφορετικών μεταβλητών. Συγκεκριμένα, τα δεδομένα εισόδου του ΤΝΔ κλιμακώθηκαν γραμμικά μεταξύ 0.0- 1.0 χρησιμοποιώντας την **εξίσωση (8)**. Λύνοντας την **εξίσωση (8)** ως προς L: ανεπεξέργαστη τιμή της εκάστοτε μεταβλητής και χρησιμοποιώντας τις στατικές παραμέτρους ανά παράμετρο του **Πίνακα 2**, υπολογίζεται η εξίσωση από- κανονικοποίησης των μεταβλητών εξόδου (outputs) του ΤΝΔ.

Συγκεκριμένα, για κάθε μία από τις τέσσερις (4) μεταβλητές εξόδου του ΤΝΔ, οι εξισώσεις από- κανονικοποίησης των τιμών, είναι, αντίστοιχα, οι **εξισώσεις 11- 14**.

$$O_3 = 0.2 * NV_{O_3} \quad (11)$$

$$SP_{ANPE} = 0.2 + 0.6 * NV_{SP_{ANPE}} \quad (12)$$

$$SP_{PACl} = 7 + 93 * NV_{SP_{PACl}} \quad (13)$$

$$SP_{Cl_{2(g)}} = 0.7 + 7.30 * NV_{SP_{Cl_{2(g)}}} \quad (14)$$

Όπου:

- NV: κανονικοποιημένη τιμή (*Normalized Value*),
- SP: σημείο ρύθμισης (*Set Point*)

4.3 Σύγκριση μοντελοποίησης ΤΝΔ με *MLR Analysis*

Πραγματοποιήθηκε σύγκριση μεταξύ μετρούμενων και προβλεπόμενων τιμών των μοντέλων πρόβλεψης ΤΝΔ και *MLR Analysis*, χρησιμοποιώντας τα πιο συχνά χρησιμοποιούμενα μαθηματικά κριτήρια: RMSE (Root Mean Square Error), R^2 (Coefficient of Determination) και ο συντελεστής συσχέτισης Pearson R (Pearson Correlation Coefficient) (Εξισώσεις 15- 17) (Gyparakis et al., 2025).

Η **Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE)** δείχνει πόσο διαφορετικές είναι, κατά μέσο όρο, οι προβλέψεις από τις μετρούμενες τιμές, που κυμαίνονται από το μηδέν έως το θετικό άπειρο. Γενικά, μια χαμηλότερη τιμή της RMSE υποδηλώνει καλύτερη προσαρμογή του μοντέλου στα δεδομένα. Η τιμή 0 σημαίνει ότι οι προβλεπόμενες τιμές ταιριάζουν απόλυτα με τις πραγματικές τιμές.

Ο **Συντελεστής Προσδιορισμού (R^2)**, σε ένα μοντέλο παλινδρόμησης, εκφράζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής, το οποίο εξηγείται από μια ανεξάρτητη μεταβλητή. Οι τιμές του R^2 κοντά στο 1 υποδεικνύουν πολύ καλή εφαρμογή μοντέλου.

Ο **Συντελεστής Συσχέτισης Pearson (R)** δείχνει την ισχύ της γραμμικής σχέσης μεταξύ μιας εξαρτημένης και ενός συνόλου ανεξάρτητων μεταβλητών και τιμές κοντά στο 1 υποδεικνύουν μια καλή προσαρμογή του μοντέλου. Γενικά, η καλύτερη απόδοση του μοντέλου επιτυγχάνεται με το χαμηλότερο RMSE (κοντά στο 0) και το υψηλότερο R και R^2 (κοντά στο 1) (Dadebo et al., 2023; Lin et al., 2023b).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{p_i} - y_i)^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum (y_i - y_{p_i})^2}{\sum (y_i - \frac{\sum y_i}{n})^2} \quad (16)$$

$$R = \frac{\sum_{i=1}^n (y_{p_i} - \bar{y}_{p_i})(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (y_{p_i} - \bar{y}_{p_i})^2 \sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (17)$$

Όπου: y_{p_i} : η προβλεπόμενη τιμή, y_i : η μετρούμενη τιμή, $\overline{y_{p_i}}$, $\overline{y_i}$: η μέση τιμή των προβλεπόμενων και των μετρούμενων τιμών, αντίστοιχα και n : ο αριθμός των παρατηρήσεων/ μετρήσεων.

Ο **Πίνακας 16** περιέχει τις τιμές του RMSE για το μοντέλο πρόβλεψης ΤΝΔ και MLR ανάλυσης ανά μελετώμενη μεταβλητή (Gyparakis et al., 2025). Σημειώνεται ότι η φυσική μονάδα του RMSE είναι mg/L για τη δόση O_3 , ANPE και PACl και kg/h για την παροχή του $Cl_{2(g)}$. Οι παράμετροι R^2 και R είναι αδιάστατες ποσότητες.

Πίνακας 16: Τιμές παραμέτρου RMSE

Μεταβλητή	ΤΝΔ	MLR Analysis
O_3	0,02	0,09
ANPE	0,07	0,05
PACl	5,93	0,08
$Cl_{2(g)}$	0,51	0,10

Ο **Πίνακας 17** περιέχει τις τιμές του συντελεστή προσδιορισμού R^2 για το μοντέλο πρόβλεψης ΤΝΔ και MLR ανάλυσης ανά μελετώμενη μεταβλητή (Gyparakis et al., 2025).

Πίνακας 17: Τιμές συντελεστή προσδιορισμού R^2

Μεταβλητή	ΤΝΔ	MLR Analysis	ΤΝΔ vs MLR
O_3	0,274	0,156	
ANPE	0,772	0,236	
PACl	0,742	0,501	
$Cl_{2(g)}$	0,838	0,681	+23%

Ο **Πίνακας 18** περιέχει τις τιμές του συντελεστή συσχέτισης Pearson R για το μοντέλο πρόβλεψης ΤΝΔ και MLR ανάλυσης ανά μελετώμενη μεταβλητή (Gyparakis et al., 2025).

Πίνακας 18: Τιμές συντελεστή συσχέτισης Pearson R

Μεταβλητή	TNΔ	MLR Analysis	TNΔ vs MLR
Residual O ₃	0,52374	0,40100	
ANPE dosage	0,87835	0,55200	
PACl dosage	0,86157	0,73500	
Cl _{2(g)} dosage	0,91553	0,82500	+11%

4.4 Μελέτη περίπτωσης

Το επιλεγμένο βέλτιστο μοντέλο TNΔ εφαρμόστηκε χρησιμοποιώντας πραγματικές ημερήσιες τιμές των λειτουργικών μεταβλητών της EEN Αποσελέμη (Πίνακας 19). Τα αποτελέσματα της εφαρμογής του επιλεγμένου μοντέλου TNΔ εμφανίζεται στον Πίνακα 20.

Πίνακας 19: Εφαρμογή βέλτιστου επιλεγμένου TNΔ σε μια μελέτη περίπτωσης

ΑΑ	Μεταβλητή Εισόδου	Μονάδα	Τιμή	Κανονικοποιημένη Τιμή
1	Daily difference in water height in reservoir (ΔH)	m	0,024	0,35
2	Raw water supply (Q)	m ³ /d	35.340	0,46
3	Raw water turbidity (T ₁)	NTU	7,85	0,01
4	Raw water pH (pH ₁)		7,9	0,73
5	Treated water turbidity (T ₂)	NTU	0,257	0,34
6	Treated water pH (pH ₂)		7,7	0,80
7	Treated water residual chlorine (Cl ₂)	mg/L	0,496	0,54
8	Treated water concentration of residual Aluminum (Al)	μg/L	84	0,55
9	Daily consumption of DWTP electricity (EI)	kWh	6.269	0,28
14	Filtration beds inlet water turbidity (T ₃)	NTU	0,95	0,11

Πίνακας 20: Αποτελέσματα εφαρμογής βέλτιστου επιλεγμένου ΤΝΔ σε μια μελέτη περίπτωσης

ΑΑ	Παράμετροι Στόχοι/ Έξοδοι ΤΝΔ	Μονάδα	Προβλεπόμενες τιμές των μεταβλητών στόχων από το επιλεγμένο ΤΝΔ	Πραγματικές τιμές των μεταβλητών, όπως χρησιμοποιήθηκαν στην ΕΕΝ Αποσελέμη
10	Residual O ₃ after ozonation process (O ₃)	mg/L	0,04	0,03
11	Anionic polyelectrolyte (ANPE)	mg/L	0,66	0,5
12	Poly-aluminum chloride hydroxide sulfate (PACl)	mg/L	11,05	10
13	Chlorine gas supply (Cl _{2(g)})	kg/h	1,67	1,50

Ο στόχος της μελέτης περίπτωσης ήταν να δείξει πόσο κοντά προσεγγίζουν οι προβλεπόμενες τιμές από το μοντέλο, τις τιμές των μεταβλητών που επιλέγονται από τους λειτουργούς των εγκαταστάσεων, με βάση στην εμπειρία τους. Τα αποτελέσματα της εφαρμογής του μοντέλου ΤΝΔ αξιολογούνται ως αρκετά ικανοποιητικά, καθώς οι πραγματικές τιμές των μεταβλητών, όπως χρησιμοποιήθηκαν στην ΕΕΝ Αποσελέμη, είναι αρκετά κοντά με τις αντίστοιχες προβλεπόμενες τιμές από το επιλεγμένο μοντέλο ΤΝΔ (Gyparakis et al., 2024).

4.5 Διερευνητική Παραγοντική Ανάλυση (Exploratory Factor Analysis, EFA)

Αρχικά πραγματοποιήθηκε έλεγχος, εάν όλες οι 14 μεταβλητές που μελετάμε, θα πρέπει να συμμετέχουν στη στατιστική διαδικασία της (διερευνητικής) παραγοντικής ανάλυσης. Στο πλαίσιο αυτό, υπολογίσθηκε:

1. ο δείκτης **Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy**: ο οποίος αποτελεί μέτρο της επάρκειας του δείγματος (sampling adequacy) για την εφαρμογή της παραγοντικής ανάλυσης, και ελέγχει ένα οι μερικές συσχετίσεις ανάμεσα στις μεταβλητές είναι μικρές,
2. **Bartlett's test of sphericity**: το οποίο εξετάζει την παρουσία υψηλών συσχετίσεων μεταξύ των μεταβλητών, δηλαδή την ύπαρξη πολύ-συγγραμμικότητας (η μη ύπαρξη αποτελεί κριτήριο για την εφαρμογή της ανάλυσης), αλλά και
3. **Anti-image Matrices**: υπολογίζονται σαν διαγώνια στοιχεία του πίνακα, τιμές KMO για κάθε μια από τις μεταβλητές.

Στον **Πίνακα 21**, δίνεται ο δείκτης KMO-Kaiser-Meyer-Olkin Measure of Sampling Adequacy. Οι τιμές του εν λόγω δείκτη κυμαίνονται από 0 έως 1. Τιμές μικρότερες του 0,5 θεωρούνται μη αποδεκτές και δεν συνιστάται η συνέχιση της παραγοντικής ανάλυσης.

Επίσης, στον **Πίνακα 21** υπάρχει και η τιμή του στατιστικού Bartlett's Test of Sphericity το οποίο όπως αναφέραμε εξετάζει την παρουσία υψηλών συσχετίσεων μεταξύ των μεταβλητών (ακολουθεί προσεγγιστικά χι-τετράγωνο κατανομή). Αν η τιμή Sig. είναι μικρότερη του 0,05, δηλαδή το τεστ είναι στατιστικά σημαντικό, απορρίπτεται η υπόθεση της μη ύπαρξης σημαντικών συσχετίσεων σε επίπεδο σημαντικότητας 5%.

Στην περίπτωση των 14 μελετώμενων μεταβλητών της EEN, ο δείκτης KMO = 0,703 τιμή **αποδεκτή** για τη συνέχιση της διαδικασίας της παραγοντικής ανάλυσης (πολύ καλές τιμές είναι αυτές που είναι > 0,8 ενώ τιμές < 0,5 μας αποτρέπουν από τη συνέχιση της ανάλυσης). Επιπλέον, από το ότι το Sig. του Bartlett's Test of Sphericity έχει τιμή

πολύ μικρή (στον **Πίνακα 21** φαίνεται ως 0,000, τιμή που συνηθίζεται να αναγράφεται ως 0,0005), προκύπτει ότι **υπάρχουν μεταβλητές οι οποίες συσχετίζονται, όχι όμως ισχυρά**. Αυτό μας οδηγεί στον έλεγχο του δείκτη **MSA (Measures of Sampling Adequacy)**, για κάθε μεταβλητή ξεχωριστά (**Πίνακας 22**).

Πίνακας 21: KMO and Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,703
Bartlett's Test of Sphericity	Approx. Chi-Square	5.551,494
	df	91
	Sig.	0,000

Στο μέρος Anti-Image Correlation του **Πίνακα 22** οι τιμές που εμφανίζονται διαγώνια είναι οι τιμές του **MSA (Measure of Sampling Adequacy)** για κάθε μεταβλητή. Τιμές κοντά στη μονάδα (1), φανερώνουν την αναγκαιότητα χρησιμοποίησης της μεταβλητής αυτής στη διαδικασία της παραγοντικής ανάλυσης. Αντίθετα μεταβλητές με τιμή του $MSA < 0,5$ θα πρέπει να απαλειφθούν στη συνέχεια από την ανάλυσή μας. Στην περίπτωση μας, όλες οι τιμές του MSA είναι $> 0,5$, οπότε όλες οι μεταβλητές θα πάρουν μέρος στην εν λόγω ανάλυση.

Στη συνέχεια, στο SPSS (IBM SPSS Statistics, version 26) χρησιμοποιώντας την επιλογή *Univariate descriptives* της Factor Analysis, παράγονται περιγραφικά στατιστικά μέτρα για τα δεδομένα όπως: μέση τιμή, τυπική απόκλιση και ο αριθμός των έγκυρων περιπτώσεων (valid cases) για κάθε μεταβλητή. Επίσης, χρησιμοποιώντας την επιλογή *Initial solution* παρουσιάζονται τα communalities αρχικά (initial) και μετά την εξαγωγή των παραγόντων (Extraction), δηλαδή τιμές που φανερώνουν το ποσοστό της διασποράς της κάθε μεταβλητής που ερμηνεύεται από τους παράγοντες, πριν και μετά την εξαγωγή τους.

Χρησιμοποιώντας κατόπιν το *Extraction* της Factor Analysis του SPSS, επιλέγουμε τη μέθοδο εξαγωγής (νέων μεταβλητών) παραγόντων **Principal components**, μέθοδος που χρησιμοποιείται για να δώσει (σαν παράγοντες) ασυσχέτιστους γραμμικούς συνδυασμούς των αρχικών μεταβλητών. Ο πρώτος

παράγοντας ερμηνεύει το μεγαλύτερο, σε σχέση με τους άλλους, μέρος της διασποράς των δεδομένων. Οι επόμενοι (διαδοχικά) παράγοντες εξηγούν μικρότερα ποσά της διασποράς τους.

Πίνακας 22: Anti- image Matrices**Anti-image Matrices**

		Daily reservoir water height difference (m)	Water flow at the entrance of the DWTP (m ³ /d)	Inlet water turbidity (NTU)	Inlet water pH	Outlet water turbidity (NTU)	Outlet water pH	Outlet water residual free chlorine (mg/L)	Outlet water residual aluminum (μg/L)	Daily electricity consumption (kWh)	Residual ozone (mg/L)	Anionic polyelectrolyte dosage (ppm)	Polyaluminum sulfate chloride dosage (ppm)	Chlorine gas flow (kg/h)	Filtration bed inlet water turbidity (NTU)
Anti-image Covariance	Daily reservoir water height difference (m)	0,904	0,037	-0,144	-0,022	0,029	-0,014	-0,034	-0,007	0,015	0,024	-0,052	0,012	-0,014	-0,047
	Water flow at the entrance of the DWTP (m ³ /d)	0,037	0,581	0,015	0,022	-0,020	-0,044	-0,003	-0,001	-0,115	-0,101	0,145	0,088	-0,067	0,001
	Inlet water turbidity (NTU)	-0,144	0,015	0,663	-0,077	-0,055	-0,010	0,101	0,066	0,004	0,045	-0,059	-0,020	-0,037	-0,255
	Inlet water pH	-0,022	0,022	-0,077	0,462	0,068	-0,204	-0,012	-0,015	-0,018	-0,097	0,010	0,041	-0,016	-0,031
	Outlet water turbidity (NTU)	0,029	-0,020	-0,055	0,068	0,669	-0,030	0,006	-0,048	-0,047	-0,039	-0,028	-0,045	-0,150	0,006
	Outlet water pH	-0,014	-0,044	-0,010	-0,204	-0,030	0,307	-0,092	-0,115	0,092	-0,010	0,049	0,096	0,054	0,024
	Outlet water residual free chlorine (mg/L)	-0,034	-0,003	0,101	-0,012	0,006	-0,092	0,850	0,015	0,003	-0,003	-0,069	-0,069	0,010	-0,097
	Outlet water residual aluminum (μg/L)	-0,007	-0,001	0,066	-0,015	-0,048	-0,115	0,015	0,639	-0,108	0,050	-0,084	0,144	0,117	-0,046

	Daily electricity consumption (kWh)	0,015	-0,115	0,004	-0,018	-0,047	0,092	0,003	-0,108	0,340	0,088	0,078	0,004	-0,155	-0,097
	Residual ozone (mg/L)	0,024	-0,101	0,045	-0,097	-0,039	-0,010	-0,003	0,050	0,088	0,840	0,130	-0,070	0,003	-0,016
	Anionic polyelectrolyte dosage (ppm)	-0,052	0,145	-0,059	0,010	-0,028	0,049	-0,069	-0,084	0,078	0,130	0,762	-0,055	-0,022	0,132
	Polyaluminum sulfate chloride dosage (ppm)	0,012	0,088	-0,020	0,041	-0,045	0,096	-0,069	0,144	0,004	-0,070	-0,055	0,462	0,104	-0,164
	Chlorine gas flow (kg/h)	-0,014	-0,067	-0,037	-0,016	-0,150	0,054	0,010	0,117	-0,155	0,003	-0,022	0,104	0,318	0,082
	Filtration bed inlet water turbidity (NTU)	-0,047	0,001	-0,255	-0,031	0,006	0,024	-0,097	-0,046	-0,097	-0,016	0,132	-0,164	0,082	0,512
Anti-image Correlation	Daily reservoir water height difference (m)	0,749^a	0,051	-0,186	-0,033	0,038	-0,026	-0,039	-0,009	0,026	0,027	-0,063	0,019	-0,027	-0,070
	Water flow at the entrance of the DWTP (m ³ /d)	0,051	0,805^a	0,024	0,042	-0,032	-0,105	-0,004	-0,001	-0,259	-0,144	0,218	0,170	-0,155	0,001
	Inlet water turbidity (NTU)	-0,186	0,024	0,581^a	-0,139	-0,083	-0,022	0,134	0,102	0,008	0,060	-0,082	-0,035	-0,082	-0,438
	Inlet water pH	-0,033	0,042	-0,139	0,732^a	0,122	-0,541	-0,019	-0,028	-0,045	-0,156	0,017	0,089	-0,042	-0,063
	Outlet water turbidity (NTU)	0,038	-0,032	-0,083	0,122	0,827^a	-0,067	0,009	-0,073	-0,099	-0,052	-0,039	-0,081	-0,326	0,010
	Outlet water pH	-0,026	-0,105	-0,022	-0,541	-0,067	0,707^a	-0,180	-0,261	0,286	-0,020	0,101	0,255	0,173	0,062

Outlet water residual free chlorine (mg/L)	-0,039	-0,004	0,134	-0,019	0,009	-0,180	0,766^a	0,020	0,005	-0,004	-0,085	-0,111	0,020	-0,147
Outlet water residual aluminum (μg/L)	-0,009	-0,001	0,102	-0,028	-0,073	-0,261	0,020	0,653^a	-0,233	0,069	-0,120	0,265	0,259	-0,080
Daily electricity consumption (kWh)	0,026	-0,259	0,008	-0,045	-0,099	0,286	0,005	-0,233	0,725^a	0,165	0,152	0,011	-0,471	-0,233
Residual ozone (mg/L)	0,027	-0,144	0,060	-0,156	-0,052	-0,020	-0,004	0,069	0,165	0,634^a	0,163	-0,112	0,006	-0,024
Anionic polyelectrolyte dosage (ppm)	-0,063	0,218	-0,082	0,017	-0,039	0,101	-0,085	-0,120	0,152	0,163	0,606^a	-0,093	-0,045	0,211
Polyaluminum sulfate chloride dosage (ppm)	0,019	0,170	-0,035	0,089	-0,081	0,255	-0,111	0,265	0,011	-0,112	-0,093	0,683^a	0,272	-0,337
Chlorine gas flow (kg/h)	-0,027	-0,155	-0,082	-0,042	-0,326	0,173	0,020	0,259	-0,471	0,006	-0,045	0,272	0,724^a	0,202
Filtration bed inlet water turbidity (NTU)	-0,070	0,001	-0,438	-0,063	0,010	0,062	-0,147	-0,080	-0,233	-0,024	0,211	-0,337	0,202	0,554^a

a. Measures of Sampling Adequacy (MSA)

Στη συνέχεια, από την ενότητα *Analyze* επιλέγουμε το πλαίσιο *Correlation matrix* (δίνεται ο πίνακας των συσχετίσεων) και από την ενότητα *Extract* μπορούμε είτε να διατηρήσουμε τους παράγοντες εκείνους που θα προκύψουν και των οποίων τα αντίστοιχα ιδιο-διανύσματα (*eigenvalues*) υπερβαίνουν μια συγκεκριμένη τιμή (συνήθως την τιμή 1) ή έναν συγκεκριμένο αριθμό παραγόντων (αριθμό που πρέπει να αναφέρουμε).

Τέλος, από την ενότητα *Display* επιλέγουμε *Unrotated factor solution*, ώστε να προκύψει, χωρίς περιστροφή των νέων αξόνων, η λύση και *Screeplot* (χρησιμοποιείται αρχικά για τη διερεύνηση του αριθμού των παραγόντων, κατόπιν επιλέγουμε το συγκεκριμένο αυτό αριθμό παραγόντων στο *Fixed number of factors*). Η επιλογή *Maximum Iterations for Convergence* επιτρέπει να προσδιοριστεί ο μέγιστος αριθμός βημάτων που ο αλγόριθμος μπορεί να κάνει για να υπολογίσει τη λύση (επιλέχθηκε η τιμή 25).

Στην ενότητα *Method* επιλέγουμε *Varimax*, (είναι μια ορθογώνια μέθοδος περιστροφής των αξόνων που ελαχιστοποιεί τον αριθμό μεταβλητών που έχουν υψηλές φορτίσεις (συντεταγμένες) σε περισσότερους από έναν παράγοντες και έτσι βοηθά στην απλοποίηση της ερμηνείας τους).

Στη συνέχεια, και στην οθόνη *Scores* επιλέγοντας *Save as variables* αποθηκεύονται τα *scores* κάθε μιας από τις περιπτώσεις (*cases*) ως προς τους παράγοντες, σαν νέες μεταβλητές, ενώ στην ενότητα *Method* που ενεργοποιείται επιλέγουμε το πλαίσιο *Regression* (μέθοδος για τον υπολογισμό των συντελεστών των παραγόντων).

Τέλος, από την ενότητα *Coefficient Display Format* επιλέγουμε *Sorted by size* (για την ταξινόμηση των συντελεστών των παραγόντων με βάση την (φθίνουσα) τιμή τους, όσες μεταβλητές σχετίζονται εμφανίζονται μαζί) και στο πλαίσιο διαλόγου *Suppress absolute values less than* πληκτρολογούμε την τιμή 0.33 (δεν εμφανίζει τους συντελεστές των παραγόντων που έχουν τιμή < 0.33).

Κάνοντας τις παραπάνω επιλογές, προκύπτουν τα αποτελέσματα της ανάλυσης σε μια σειρά από πίνακες που παρουσιάζονται παρακάτω μαζί με την ερμηνεία τους.

Στον **Πίνακας 23** δίνονται η μέση τιμή, η τυπική απόκλιση και ο αριθμός των

περιπτώσεων για κάθε μεταβλητή (1,188 δεδομένα/τιμές).

Πίνακας 23: Descriptive Statistics

Descriptive Statistics			
	Mean	Std. Deviation	Analysis N
Daily reservoir water height difference (m)	-0,0017	0,17640	1.188
Water flow at the entrance of the DWTP (m3/d)	39.062,18	9.492,131	1.188
Inlet water turbidity (NTU)	6,6074	22,12048	1.188
Inlet water pH	7,5818	0,34058	1.188
Outlet water turbidity (NTU)	0,1622	0,07840	1.188
Outlet water pH	7,2983	0,32013	1.188
Outlet water residual free chlorine (mg/L)	0,4412	0,11119	1.188
Outlet water residual aluminum (µg/L)	41,9691	21,36114	1.188
Daily electricity consumption (kWh)	9.424,48	2.891,806	1.188
Residual ozone (mg/L)	0,0451	0,02027	1.188
Anionic polyelectrolyte dosage (ppm)	0,4072	0,17068	1.188
Polyaluminum sulfate chloride dosage (ppm)	17,9860	11,63555	1.188
Chlorine gas flow (kg/h)	2,4596	1,27212	1.188
Filtration bed inlet water turbidity (NTU)	1,2878	0,84297	1.188

Ο Πίνακας 24 περιέχει τα λεγόμενα *communalities (Initial and Extraction)*. Τα *communalities (Initial and Extraction)* παίρνουν τιμές από 0 έως 1. Το 0 σημαίνει ότι οι παράγοντες ερμηνεύουν ποσοστό 0% της διασποράς της αντίστοιχης μεταβλητής, ενώ η τιμή 1 δηλώνει ότι το 100% της διασποράς της μεταβλητής ερμηνεύεται από τους παράγοντες που προέκυψαν. Δηλαδή, το *Extraction* εκφράζει το ποσοστό της διασποράς της κάθε μεταβλητής, ξεχωριστά, που μπορεί να ερμηνευτεί από τους παράγοντες που προέκυψαν. Στη μέθοδο *Principal components* οι τιμές *Initial* είναι πάντα 1 (γιατί σαν παράγοντες αρχικά θεωρούνται όλες οι μεταβλητές, οπότε και ερμηνεύουν το 100% της διασποράς).

Πίνακας 24: Communalities

Communalities		
	Initial	Extraction
Daily reservoir water height difference (m)	1,000	0,432
Water flow at the entrance of the DWTP (m ³ /d)	1,000	0,654
Inlet water turbidity (NTU)	1,000	0,661
Inlet water pH	1,000	0,693
Outlet water turbidity (NTU)	1,000	0,445
Outlet water pH	1,000	0,819
Outlet water residual free chlorine (mg/L)	1,000	0,197
Outlet water residual aluminum (μg/L)	1,000	0,527
Daily electricity consumption (kWh)	1,000	0,756
Residual ozone (mg/L)	1,000	0,562
Anionic polyelectrolyte dosage (ppm)	1,000	0,607
Polyaluminum sulfate chloride dosage (ppm)	1,000	0,759
Chlorine gas flow (kg/h)	1,000	0,768
Filtration bed inlet water turbidity (NTU)	1,000	0,736

Extraction Method: Principal Component Analysis.

Στον **Πίνακα 25**, η πρώτη στήλη *Component*, δίνει το πλήθος των παραγόντων και οι οποίοι είναι σε πλήθος όσοι και οι αρχικές μεταβλητές, δηλαδή στην περίπτωση μας 14.

Η στήλη *Initial Eigenvalues-Total* δίνει τις **ιδιοτιμές (*eigenvalues*)** των παραγόντων ταξινομημένες κατά φθίνουσα τάξη. Η κάθε μία από αυτές τις τιμές εκφράζει τη συνολική διασπορά (των δεδομένων) η οποία ερμηνεύεται από τον αντίστοιχο παράγοντα.

Η στήλη *Initial Eigenvalues-% of Variance* δίνει το ποσοστό της διασποράς το οποίο ερμηνεύεται από τον (αντίστοιχο) παράγοντα, ενώ η στήλη *Initial Eigenvalues-Cumulative %* παρέχει αθροιστικά τα ποσοστά της προηγούμενης στήλης (συνολική μεταβλητότητα που ερμηνεύεται από τον εν λόγω παράγοντα και τους προηγούμενους από αυτόν).

Οι στήλες *Extraction Sums of Squared Loadings –Total*, *Extraction Sums of Squared Loadings -% of Variance* και *Extraction Sums of Squared Loadings - Cumulative %*, περιέχουν τα ίδια αποτελέσματα με αυτά των *Initial eigenvalues*, αλλά μόνον για

τους (τέσσερις) παράγοντες που σχηματίζονται.

Η στήλη *Rotation Sums of Squared Loadings –Total* δίνει τις τέσσερις ιδιοτιμές που έχουν τιμή μεγαλύτερη του 1 μετά την περιστροφή (*rotation*) και οι οποίες έχουν διαφορετικές τιμές από προηγούμενες αντίστοιχες στήλες. Η στήλη *Rotation Sums of Squared Loadings –% of Variance* δίνει το ποσοστό της διασποράς που ερμηνεύεται από τους παράγοντες μετά την περιστροφή, ενώ η στήλη *Rotation Sums of Squared Loadings –Cumulative %* παρέχει αθροιστικά τα ποσοστά της προηγούμενης στήλης. Στην περίπτωση μας βλέπουμε ότι οι τέσσερις παράγοντες ερμηνεύουν το **61,5%** της μεταβλητότητας των μεταβλητών/ δεδομένων.

Στη μέθοδο *Principal Components* οι τιμές στις περιοχές Initial και Extraction του **Πίνακα 25**, είναι ίδιες. Επιπλέον, η περιοχή *Rotation* αν και έχει διαφορετικές τιμές από τις αντίστοιχες της περιοχής *Extraction*, όμως η συνολική μεταβλητότητα όπως αυτή εκφράζεται στο τέλος της στήλης *Cumulative* είναι η ίδια (με αυτή των προηγούμενων περιοχών), που σημαίνει ότι συνολικά οι τέσσερις παράγοντες πριν και μετά την περιστροφή ερμηνεύουν το ίδιο ποσοστό της διασποράς των μεταβλητών.

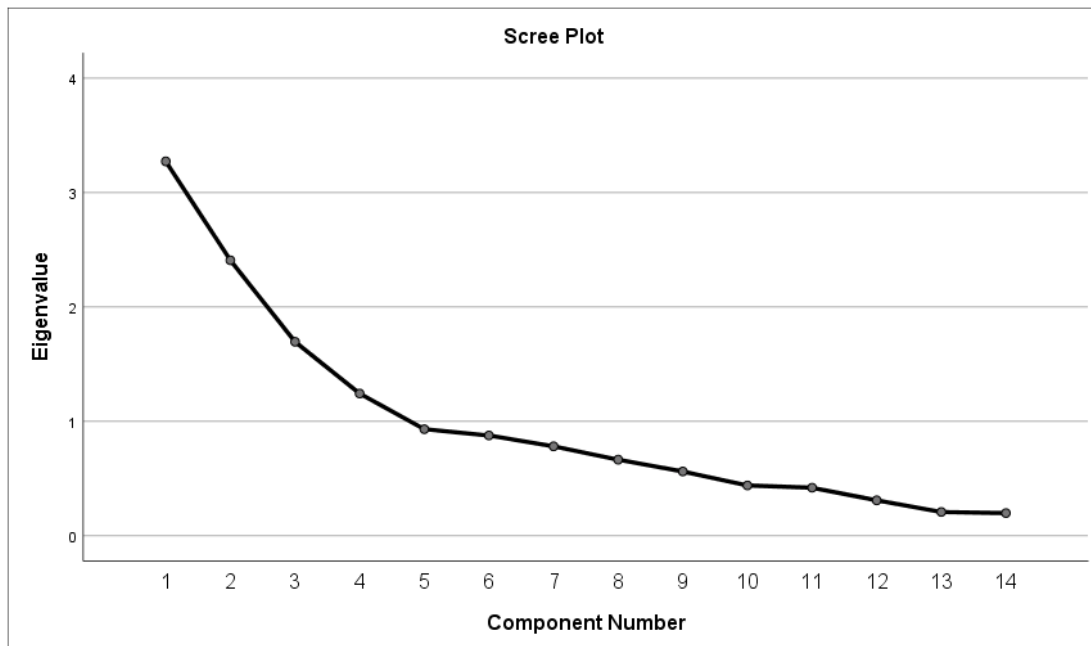
Σε μια παραγοντική ανάλυση, ο καθορισμός του αριθμού των παραγόντων γίνεται μεταξύ άλλων είτε με βάση το μέγεθος της (αντίστοιχης) ιδιοτιμής (συνήθως θέλουμε τιμές της > 1), είτε με βάση το συνολικό ποσοστό διασποράς των δεδομένων που θέλουμε να ερμηνεύεται από τους παράγοντες. Υπάρχουν όμως περιπτώσεις που ο ερευνητής επιθυμεί συγκεκριμένο αριθμό παραγόντων με σκοπό να κάνει σύγκριση των αποτελεσμάτων της έρευνάς του με άλλη, αντίστοιχη, άλλου ερευνητή.

Πίνακας 25: Total Variance Explained

Component	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,272	23,374	23,374	3,272	23,374	23,374	3,023	21,592	21,592
2	2,407	17,193	40,567	2,407	17,193	40,567	2,357	16,836	38,427
3	1,695	12,108	52,675	1,695	12,108	52,675	1,814	12,959	51,386
4	1,242	8,873	61,548	1,242	8,873	61,548	1,423	10,162	61,548
5	0,931	6,651	68,199						
6	0,875	6,254	74,452						
7	0,781	5,578	80,031						
8	0,664	4,744	84,775						
9	0,561	4,008	88,783						
10	0,439	3,132	91,915						
11	0,419	2,993	94,908						
12	0,309	2,207	97,115						
13	0,207	1,476	98,591						
14	0,197	1,409	100,000						

Extraction Method: Principal Component Analysis.

Το διάγραμμα Scree Plot (**Σχήμα 29**) μπορεί να αποτελέσει (ένα αρχικό διερευνητικό) κριτήριο επιλογής του αριθμού των παραγόντων. Προτείνει να πάρουμε τόσες συνιστώσες/ παράγοντες όσα και τα σημεία της γραμμής του γραφήματος των ιδιοτιμών πριν αυτή γίνει περίπου παράλληλη με τον οριζόντιο άξονα (*Component Number*). Για τα δεδομένα μας, το *Scree plot* φαίνεται να προτείνει ίσως και έναν πέμπτο παράγοντα.



Σχήμα 29: Διάγραμμα Scree Plot

Στον **Πίνακα 26**, έχουμε πλέον τους τέσσερις παράγοντες (component 1, 2, 3, 4) και τις αντίστοιχες φορτίσεις (loadings) των αρχικών μεταβλητών σε αυτούς τους, πριν την περιστροφή. Κάθε μια από τις φορτίσεις αυτές είναι στην ουσία ο συντελεστής συσχέτισης της μεταβλητής με τον αντίστοιχο παράγοντα.

Πίνακας 26: Component Matrix

	Component ^a			
	1	2	3	4
Chlorine gas flow (kg/h)	0,819			
Daily electricity consumption (kWh)	0,805			
Outlet water pH	-0,730	0,485		
Outlet water turbidity (NTU)	0,628			
Inlet water pH	-0,603	0,404	0,402	
Outlet water residual free chlorine (mg/L)	-0,403			
Polyaluminum sulfate chloride dosage (ppm)		-0,840		

Water flow at the entrance of the DWTP (m ³ /d)	0,478	0,552		
Outlet water residual aluminum (μg/L)	-0,372	0,504		0,348
Filtration bed inlet water turbidity (NTU)		-0,545	0,662	
Inlet water turbidity (NTU)		-0,410	0,626	
Anionic polyelectrolyte dosage (ppm)			-0,502	0,486
Residual ozone (mg/L)				-0,641
Daily reservoir water height difference (m)			0,332	0,498

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Στον **Πίνακα 27**, έχουμε τους τέσσερις παράγοντες και τις αντίστοιχες φορτίσεις των αρχικών μεταβλητών σε αυτούς, **μετά** όμως την περιστροφή τους. Η περιστροφή έχει ως στόχο την αύξηση των μεγάλων φορτίσεων και τη μείωση των μικρών. Από τον πίνακα αυτό καταλήγουμε στη σύνθεση των παραγόντων με κριτήριο ότι, η κάθε μεταβλητή ανήκει στον παράγοντα με τη μεγαλύτερη φόρτιση.

Στη μελέτη περίπτωσης μας:

- οι μεταβλητές: ***Chlorine gas flow (kg/h), Daily electricity consumption (kWh), Water flow at the entrance of the DWTP (m³/d), Outlet water turbidity (NTU), Outlet water residual free chlorine (mg/L)*** εντάσσονται στον **πρώτο (1^ο) παράγοντα**. Οι μεταβλητές ομαδοποιούνται σε μια ομάδα, η οποία θα μπορούσε να ονομαστεί **Ομάδα Ποσοτικών Χαρακτηριστικών νερού**, καθώς όταν αυξάνεται η παροχή του υπό επεξεργασία νερού, τότε αυξάνεται η συνολική κατανάλωση ηλεκτρικής ενέργειας της EEN, η κατανάλωση αερίου χλωρίου για την τελική απολύμανση του νερού και ενδεχομένως να υπάρχει μεταβολή της θολότητας του επεξεργασμένου νερού και της ποσότητας που αυτό περιέχει σε ελεύθερο υπολειμματικό χλώριο.
- Οι μεταβλητές: ***Outlet water pH, Inlet water pH, Outlet water residual***

aluminum (μg/L), Polyaluminum sulfate chloride dosage (ppm) εντάσσονται στον **δεύτερο (2^ο) παράγοντα**. Οι προαναφερόμενες μεταβλητές εντάσσονται σε μια ομάδα μεταβλητών, η οποία θα μπορούσε να ονομαστεί **Ομάδα Χημικών Χαρακτηριστικών νερού**, καθώς οι μεταβλητές σχετίζονται με τη δόση PACl και το αντίστοιχο υπολειμματικό αργίλιο στο τελικά επεξεργασμένο νερό, τα οποία σχετίζονται με την τιμή του pH του ανεπεξέργαστου και του τελικά επεξεργασμένου νερού.

- Οι μεταβλητές: *Inlet water turbidity (NTU), Filtration bed inlet water turbidity (NTU), Daily reservoir water height difference (m)* εντάσσονται στον **τρίτο παράγοντα**. Οι προαναφερόμενες μεταβλητές εντάσσονται σε μια ομάδα μεταβλητών, η οποία θα μπορούσε να ονομαστεί **Ομάδα Ύψους νερού ταμιευτήρα**, καθώς, πόσο ομαλά ή απότομα, αυξάνεται ή μειώνεται, καθημερινά, το ύψος του νερού στον Ταμιευτήρα Αποσελέμη, επηρεάζει τη θολότητα του εισερχόμενου νερού στην EEN και τη θολότητα του τελικά επεξεργασμένου νερού.
- Οι μεταβλητές: *Anionic polyelectrolyte dosage (ppm), Residual ozone (mg/L)* εντάσσονται στον **τέταρτο παράγοντα**. Οι προαναφερόμενες μεταβλητές εντάσσονται σε μια ομάδα μεταβλητών, η οποία θα μπορούσε να ονομαστεί **Ομάδα Αποτελεσματικότητας Οζόνωσης**, καθώς η ποσότητα του υπολειμματικού όζοντος στο νερό, μετά τη διεργασία της οζόνωσης, δύναται να επηρεάσει αντιστρόφως ανάλογα την ποσότητα του AN PE, που χρησιμοποιείται για την καλύτερη συσσωμάτωση των παραγόμενων κροκίδων στη διεργασία της κροκίδο- καθίζησης.

Πίνακας 27: Rotated Component Matrix

	Component			
	1	2	3	4
Chlorine gas flow (kg/h)	0,866			
Daily electricity consumption (kWh)	0,847			
Water flow at the entrance of the DWTP (m3/d)	0,664			0,394
Outlet water turbidity (NTU)	0,655			
Outlet water residual free chlorine (mg/L)	-0,376			
Outlet water pH	-0,409	0,788		
Inlet water pH		0,721		
Outlet water residual aluminum (μg/L)		0,716		
Polyaluminum sulfate chloride dosage (ppm)	-0,354	-0,700	0,378	
Inlet water turbidity (NTU)			0,806	
Filtration bed inlet water turbidity (NTU)			0,785	
Daily reservoir water height difference (m)			0,554	
Anionic polyelectrolyte dosage (ppm)				-0,721
Residual ozone (mg/L)				0,700

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.^a

a. Rotation converged in 7 iterations.

Ο **Πίνακας 28** είναι ο πίνακας που χρησιμοποιήθηκε για το μετασχηματισμό των φορτίσεων του **Πίνακα 26** στις φορτίσεις του **Πίνακα 27**. Παρατηρούμε ότι οι συσχετίσεις μεταξύ των παραγόντων είναι μικρές και επομένως η ορθογώνια περιστροφή κρίνεται ως ικανοποιητική. Αν ο **Πίνακας 28** έδινε υψηλές συσχετίσεις θα έπρεπε να κάνουμε μη ορθογώνια περιστροφή (*Direct Oblimin Method*).

Πίνακας 28: Component Transformation Matrix

Component Transformation Matrix				
Component	1	2	3	4
1	0,887	-0,460	-0,026	-0,030
2	0,380	0,747	-0,500	0,218
3	0,186	0,279	0,785	0,522
4	0,186	0,391	0,365	-0,824

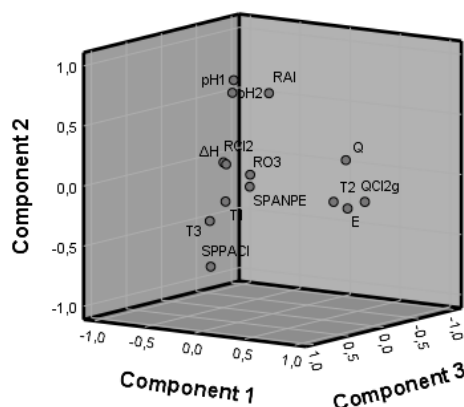
Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Στο **Σχήμα 30** δίνεται η αναπαράσταση των μεταβλητών στο τρισδιάστατο σύστημα αξόνων των τριών παραγόντων.

Τέλος, στον *Data Editor* έχουν αποθηκευτεί 4 νέες μεταβλητές με τα ονόματα FAC1_1, FAC2_1, FAC3_1 και FAC4_1 στο τέλος του αρχείου, οι οποίες προέκυψαν με χρήση αφ' ενός των φορτίσεων και αφ' ετέρου των τιμών των αρχικών μεταβλητών. Είναι οι συντεταγμένες κάθε μιας από τις περιπτώσεις (cases) ως προς τους παράγοντες που προέκυψαν. Οι νέες αυτές μεταβλητές μπορούν να αντικαταστήσουν τις αρχικές σε παραπέρα μελέτη και σε διαδικασίες όπως t- test, Ανάλυση διακύμανσης κ.α.

Component Plot in Rotated Space



Σχήμα 30: Αναπαράσταση των μεταβλητών στο τρισδιάστατο σύστημα αξόνων

5 ΣΥΖΗΤΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ, ΣΥΜΠΕΡΑΣΜΑΤΑ, ΜΕΛΛΟΝΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

5.1 Συζήτηση Αποτελεσμάτων και Συμπεράσματα

5.1.1 Πολλαπλή Γραμμική Παλινδρόμηση (*Multiple Linear Regression*)

Σύμφωνα με την εμπειρία από τη λειτουργία μια Εγκατάστασης Επεξεργασίας Νερού (EEN), τα αποτελέσματα του μαθηματικού μοντέλου με βάση την ανάλυση πολλαπλής παλινδρόμησης (MLR), αξιολογούνται ικανοποιητικά και αιτιολογούνται, για τους κύριους συντελεστές της εξίσωσης (10) που περιγράφει το προαναφερόμενο μοντέλο, ως εξής:

- ✓ Αυξημένη κατανάλωση ηλεκτρικής ενέργειας [(συντελεστής μοντέλου: +0.511 εξίσωσης (10)] παρατηρείται όταν λειτουργεί με αυξημένη απόδοση, ο κύριος ενεργοβόρος εξοπλισμός της EEN, που είναι η γεννήτρια όζοντος (O_3). Για να συμβεί αυτό σημαίνει ότι το εισερχόμενο νερό είναι επιβαρυνμένο και τα ποιοτικά χαρακτηριστικά του, τόσο τα φυσικοχημικά, όσο και τα μικροβιολογικά, είναι επιβαρυνμένα.
- ✓ Επίσης, λόγω των αυξημένων αυτών τιμών αιτιολογείται η αυξημένη παροχή αερίου χλωρίου, που προβλέπει το προαναφερόμενο μοντέλο, έτσι ώστε να

επιτευχθεί η επιθυμητή τιμή ελεύθερου υπολειμματικού χλωρίου, που έχει ορισθεί από τον λειτουργό της εγκατάστασης για να καλύπτει το τελικά επεξεργασμένο νερό τις προδιαγραφές του νερού ανθρώπινης κατανάλωσης.

- ✓ Αναφορικά με τη θολότητα του νερού εισόδου των κλινών (συντελεστής μοντέλου: -0,243), όταν αυξάνεται η θολότητα του νερού εισόδου των κλινών, οι κλίνες διήθησης φράζουν (φαινόμενο έμφραξης κλινών διήθησης), με αποτέλεσμα το εξερχόμενο νερό να έχει μειωμένη θολότητα, άρα και μειωμένο ρυπαντικό φορτίο και γι' αυτό το λόγο είναι λογικό να απαιτείται μικρότερη ποσότητα παροχής αερίου χλωρίου για την απολύμανσή του.
- ✓ Επίσης, η αύξηση της παροχής του αερίου χλωρίου δύναται να προκαλέσει αύξηση της θολότητας του νερού εξόδου (συντελεστής μοντέλου: +0.362), καθώς οξειδώνονται τα διαλυμένα μέταλλα που υπάρχουν στο υπό εξεργασία νερό, τα οποία μετατρέπονται σε αδιάλυτα και αυτό προκαλεί την αύξηση της θολότητας του τελικά επεξεργασμένου νερού.
- ✓ Αναφορικά με την τιμή του pH του τελικά επεξεργασμένου νερού (συντελεστής μοντέλου: -0.141), η αύξηση της παροχής αερίου χλωρίου, οδηγεί σε αύξηση της συγκέντρωσης ελεύθερου υπολειμματικού χλωρίου και αυτό σε μείωση της τιμής του pH του τελικά επεξεργασμένου νερού.
- ✓ Τέλος, η αύξηση της παροχής του νερού στην είσοδο της EEN νερού (συντελεστής μοντέλου: +0.130), όπως και η αύξηση της θολότητας του εισερχόμενου νερού (συντελεστής μοντέλου: +0,305), προκαλούν αύξηση της παροχής αερίου χλωρίου και αυτή είναι μια λογική σχέση που περιγράφεται στο μοντέλο πρόβλεψης.

5.1.2 Ανάπτυξη TNA

Η παρούσα μελέτη ενισχύει την άποψη ότι τα TNA είναι χρήσιμα εργαλεία για μια EEN, με υψηλή αποτελεσματικότητα στην αντιστοίχιση και πρόβλεψη σύνθετων σχέσεων. Το συμπέρασμα ότι τα μοντέλα TNA παρέχουν ακριβή αποτελέσματα πρόβλεψης έχει αντληθεί από πολλές άλλες παρόμοιες μελέτες (Alam et al., 2022;

Bhattacharya et al., 2021; Dadebo et al., 2023; Li et al., 2021a; Wongburi and Park, 2023).

Στο πλαίσιο της παρούσας ΔΔ, τελικά επιλέχθηκε το σενάριο μοντέλου ΤΝΔ, ανάμεσα από 304 σενάρια που εξετάστηκαν, το οποίο αποτελείται από 100 νευρωνικά δίκτυα, 100 κόμβους και 42 κρυφούς κόμβους. Συγκεκριμένα, το επιλεγμένο μοντέλο ΤΝΔ κατασκευάστηκε με 10 μεταβλητές εισόδου (ΔH , Q , T_1 , pH_1 , T_2 , pH_2 , Cl_2 , Al , El , T_3) και 4 μεταβλητές εξόδου (O_3 , $ANPE$, $PACl$, $Cl_{2(g)}$). Η επιλογή βασίστηκε στο μικρότερο τιμή του δείκτη απόδοσης δοκιμής (t_{perf}). Αντίστοιχες μελέτες (Chen et al., 2020; Dadebo et al., 2023; Ghaedi and Vafaei, 2017; Kim et al., 2023; Wu and Lo, 2010, 2008) έχουν χρησιμοποιήσει παρόμοιες μεταβλητές ποιότητας ακατέργαστου νερού (όπως pH , θολότητα και χρώμα) με αρκετά ικανοποιητικά αποτελέσματα.

Επίσης, όπως ήταν αναμενόμενο, η αύξηση του αριθμού των νευρωνικών δικτύων, με σταθερό αριθμό κόμβων και η αύξηση του αριθμού των κόμβων, με σταθερό αριθμό νευρωνικών δικτύων, έχουν ως αποτέλεσμα μια εκθετική αύξηση του χρόνου εκτέλεσης των μοντέλων ΤΝΔ. Η ενσωμάτωση όλων των διαθέσιμων μετρήσιμων μεταβλητών στο μοντέλο πρόβλεψης βελτιώνει την απόδοσή του, σε βάρος του χρόνου.

Από την άλλη, η δεύτερη καλύτερη επιλογή είναι με δύο μεταβλητές λιγότερες και η οποία έχει ελαφρώς λιγότερο ικανοποιητικά αποτελέσματα, αν και η εκτέλεση του κώδικα γίνεται σε σαφώς πολύ μικρότερο χρονικό διάστημα. Είναι πιθανό η σχέση μεταξύ των μεταβλητών ΔH και pH_2 και οι δόσεις των χημικών που χρησιμοποιούνται σε μια ΕΕΝ, να είναι πολύ πιο περίπλοκη και μόνο όταν χρησιμοποιείται μεγάλος αριθμός νευρωνικών δικτύων, να μπορεί να αποτυπωθεί αυτή η πολυπλοκότητα. Οι επιπλέον πληροφορίες, που παρέχονται με τις δύο προαναφερθείσες μεταβλητές, βελτιώνουν την ακρίβεια πρόβλεψης του επιλεγμένου μοντέλου κατά 1,65%. Ωστόσο, το ελαφρώς πιο ακριβές μοντέλο χρησιμοποιεί 42 κρυφούς κόμβους και διαρκεί 23.346 sec, σε σύγκριση με το δεύτερο καλύτερο, το οποίο χρησιμοποιεί 8 κρυφούς κόμβους και το τρέξιμο του κώδικα διαρκεί μόλις 461 sec. Όπως έχουν καταλήξει πολλές άλλες μελέτες (Chen et al., 2020; Dunnington et al., 2021; Kim and Parnichkun, 2017a; Lin et al., 2023a), η αύξηση της πολυπλοκότητας ενός ΤΝΔ αυξάνει την ακρίβειά του,

αυξάνοντας παράλληλα τον αριθμό των κρυφών κόμβων και τελικά τον απαιτούμενο χρόνο. Σε περιπτώσεις όπου ο χρόνος και η υπολογιστική ισχύς είναι σημαντικές παράμετροι για τη λήψη αποφάσεων, φαίνεται ότι μπορούν να ληφθούν ικανοποιητικά αποτελέσματα χρησιμοποιώντας λιγότερο πολύπλοκα μοντέλα με λιγότερα νευρωνικά δίκτυα.

Με εξαίρεση τη δόση του όζοντος (O_3), το επιλεγμένο μοντέλο ΤΝΔ προβλέπει αποτελεσματικά τις τιμές των μεταβλητών εξόδου για τις δόσεις των χημικών που χρησιμοποιούνται σε μια EEN, οι οποίες περιλαμβάνουν τις δόσεις $Cl_{2(g)}$, ANPE και PACl. Το μοντέλο ΤΝΔ ακολουθεί την τάση αύξησης ή μείωσης των χημικών πεδίου και προσεγγίζει τις ακραίες τιμές σε μεγάλο βαθμό. Η τιμή του συντελεστή R^2 επιτυγχάνεται καλύτερα για το μοντέλο πρόβλεψης του $Cl_{2(g)}$, ακολουθούμενο από αυτό του ANPE, μετά από του PACl, και τέλος από του O_3 . Η μεγάλη διακύμανση των τιμών του όζοντος μπορεί να δικαιολογήσει τη συγκεκριμένη χαμηλή τιμή του συντελεστή R^2 . Αντίστοιχες μελέτες έχουν δείξει ακόμη καλύτερα αποτελέσματα σχετικά με τον συντελεστή αυτό (Dadebo et al., 2023; Dunnington et al., 2021; Kim and Parnichkun, 2017b; Lin et al., 2023a; Papailiou et al., 2022).

Το μοντέλο ΤΝΔ είναι σε θέση να υποδεικνύει τις βέλτιστες δόσεις $Cl_{2(g)}$, ANPE και PACl, με βάση την εμπειρία καθημερινών μετρήσεων 38 μηνών. Με αυτόν τον τρόπο, οι προβλέψεις του μοντέλου μπορούν να βοηθήσουν τους νέους λειτουργούς EEN, ιδίως για τον καθορισμό των απαιτούμενων δόσεων χημικών επεξεργασίας νερού, εξοικονομώντας έτσι χρόνο και βοηθώντας τους να αποκτήσουν πρακτική τεχνογνωσία.

Στην περίπτωση της εφαρμογής του στην EEN Αποσελέμη, ο λειτουργός της εγκατάστασης θα έχει μια αξιόπιστη εκτίμηση των μεταβλητών εξόδου του ΤΝΔ, αναφορικά με τις δόσεις των χημικών επεξεργασίας του νερού που θα πρέπει να εφαρμόζει, ανάλογα με την τρέχουσα ποιότητα του εισερχόμενου νερού και άλλες διαθέσιμες λειτουργικές μεταβλητές. Το προτεινόμενο μοντέλο ΤΝΔ πρόβλεψης ανταποκρίνεται ικανοποιητικά στην πρόβλεψη βασικών λειτουργικών μεταβλητών, όπως έχει αποδειχθεί και από παρόμοιες μελέτες (Alam et al., 2022; Dadebo et al., 2023; Kim and Parnichkun, 2017b; Wu and Lo, 2010; Yamamura et al., 2020).

Στη μελέτη περίπτωσης και με βάση την εμπειρία του λειτουργού της EEN

Αποσελέμη, οι διαφορές μεταξύ των δόσεων των χημικών επεξεργασίας του νερού που χρησιμοποιήθηκαν στην πράξη και που προβλέφθηκαν από το μοντέλο ΤΝΔ, αξιολογούνται ως ασήμαντες. Η περίοδος πρόβλεψης ορίζεται ως ο χρόνος κατά τη διάρκεια του οποίου, οι τιμές των λειτουργικών μεταβλητών εισόδου του μοντέλου δεν αλλάζουν σημαντικά ($\pm 10\%$), π.χ. εποχιακά, μετά από έντονα καιρικά φαινόμενα, μετά από μια σημαντική αλλαγή στην ποσότητα του νερού στον ταμιευτήρα κ.λπ. Οι προβλέψεις από το μοντέλο ΤΝΔ θα μπορούσαν να βοηθήσουν τους λειτουργούς των εγκαταστάσεων να βελτιστοποιήσουν τους χρησιμοποιούμενους πόρους, συμπεριλαμβανομένης της μείωσης της κατανάλωσης των χημικών επεξεργασίας του νερού και της αποφυγής περιττών δοκιμών (π.χ. διεξάγοντας jar tests). Αυτό οδηγεί σε εξοικονόμηση χρόνου και χρήματος, διασφαλίζοντας παράλληλα ότι το πόσιμο νερό που παράγεται συμμορφώνεται με τις προδιαγραφές του νερού ανθρώπινης κατανάλωσης.

5.1.3 Σύγκριση μοντελοποίησης ΤΝΔ με MLR Analysis

Πραγματοποιήθηκε σύγκριση μεταξύ μετρούμενων και προβλεπόμενων τιμών των μοντέλων πρόβλεψης ΤΝΔ και MLR Analysis, χρησιμοποιώντας τα πιο συχνά χρησιμοποιούμενα μαθηματικά κριτήρια: RMSE, R^2 και R (Εξισώσεις 15- 17).

Σύμφωνα με τις τιμές του R^2 , παρατηρείται ότι τρεις από τις τέσσερις δόσεις των χημικών επεξεργασίας νερού που μελετήθηκαν, προβλέπονται σχετικά ικανοποιητικά από το επιλεγμένο μοντέλο ΤΝΔ ($\text{Cl}_{2(\text{g})}$: $R^2= 0,838$, δόση ANPE: $R^2= 0,772$, δόση PACl: $R^2= 0.742$ σε φθίνουσα σειρά), ενώ με βάση το μοντέλο της MLR Analysis μόνο μία δόση ($\text{Cl}_{2(\text{g})}$: $R^2= 0.681$). Η μεγάλη διακύμανση των τιμών του όζοντος μπορεί να δικαιολογήσει τη συγκεκριμένη χαμηλή τιμή του R^2 , σε όλες τις περιπτώσεις. Ωστόσο, ακόμη και για ένα μόνο χημικό επεξεργασίας του νερού, η πρόβλεψη με χρήση μοντέλου MLR Analysis, αξιολογείται ως ικανοποιητική. Παρόμοια ή ακόμα καλύτερα αποτελέσματα έχουν βρεθεί από αντίστοιχες μελέτες (Alnuwaiser et al., 2022; Dadebo et al., 2023; Li et al., 2021b; Lin et al., 2023b, 2023a).

Αναφορικά με τις τιμές του RMSE των προβλεπόμενων και παρατηρούμενων

τιμών των μοντέλων πρόβλεψης με TND και MLR Analysis, το μοντέλο MLR έχει καλύτερη απόδοση για τα τρία (RMSE ANPE= 0,05 mg/L, RMSE PACl= 0,08 mg/L και RMSE Cl_{2(g)}= 0,10 kg/h) από τα τέσσερα χρησιμοποιημένα χημικά επεξεργασίας του νερού, από το μοντέλο TND που έχει μόνο για το ένα από αυτά (RMSE O₃= 0,02 mg/L). Πρακτικά, στη μοντελοποίηση η τιμή του RMSE είναι η τυπική απόκλιση των σφαλμάτων πρόβλεψης (Wongburi and Park, 2023). Η τιμή αυτής της πρόβλεψης, χρησιμοποιώντας το μοντέλο MLR, ενισχύεται περαιτέρω εάν λάβουμε υπόψη ότι η μέση τιμή της δόσης ANPE είναι 0,40 mg/L, της δόσης PACl 17,99 mg/L και της δόσης Cl_{2(g)} 2,46 kg/h, αντίστοιχα.

Επιπλέον, υπάρχουν μελέτες για τον προσδιορισμό των δόσεων κροκιδωτικών EEN στις οποίες τα αποτελέσματα της μοντελοποίησης της ανάλυσης MLR φαίνεται να έχουν ελαφρώς καλύτερη απόδοση (μικρό RMSE και υψηλό R²) από τη μοντελοποίηση TND (Shi et al., 2022). Σύμφωνα με τη βιβλιογραφία, στη μοντελοποίηση πρόβλεψης, όσο χαμηλότερες τιμές του RMSE και όσο υψηλότερες είναι οι τιμές των R² και R, τόσο πιο κοντά είναι οι προβλεπόμενες τιμές στις μετρούμενες τιμές (Azeem et al., 2023; Lin et al., 2023b).

Πρέπει να επισημάνουμε ότι, εάν κάποιος θέλει να χρησιμοποιήσει τα μοντέλα πρόβλεψης που περιεγράφηκαν παραπάνω (TND ή MLR) για την πρόβλεψη δόσεων Cl_{2(g)}, είναι προτιμότερο να χρησιμοποιήσει αυτό με το μικρότερο RMSE. Εάν το ενδιαφέρον είναι περισσότερο για την καλύτερη προσαρμογή των προβλεπόμενων τιμών στις παρατηρούμενες τιμές, τότε θα πρέπει να χρησιμοποιηθεί καλύτερα, το μοντέλο αυτό με τις μεγαλύτερες τιμές R².

Σύμφωνα με τον συντελεστή συσχέτισης Pearson (R), καλύτερη πρόβλεψη επιτυγχάνεται από το μοντέλο TND για τα τρία από τα τέσσερα χρησιμοποιούμενα χημικά επεξεργασίας νερού (δόση ANPE, PACl, Cl_{2(g)}), ενώ με την ανάλυση MLR μόνο για ένα (Cl_{2(g)}). Σύμφωνα με τα αποτελέσματα των R² και R, η πρόβλεψη της δόσης Cl_{2(g)} φαίνεται να είναι 23% και 11% καλύτερη χρησιμοποιώντας μοντέλο πρόβλεψης TND σε σύγκριση με το μοντέλο πρόβλεψης με MLR Analysis.

Συμπερασματικά, οι τρεις χρησιμοποιούμενες παράμετροι σύγκρισης (RMSE, R² και R), δείχνουν ικανοποιητικά αποτελέσματα σχετικά με το αρχικό ερώτημα: ποιο

μοντέλο πρόβλεψης αποδίδει καλύτερα στην πρόβλεψη των δόσεων των χημικών που χρησιμοποιούνται σε μια EEN. Τα μοντέλα ΤΝΔ είναι πολύ καλύτερα από τα αντίστοιχα μοντέλα πρόβλεψης ανάλυσης MLR, αν μας ενδιαφέρει κυρίως η προσαρμοστικότητα της πρόβλεψης, ενώ αν μας ενδιαφέρει να έχουμε όσο το δυνατόν λιγότερα σφάλματα στις προβλεπόμενες τιμές, τότε η ανάλυση MLR δύναται σε κάποιες περιπτώσεις να είναι καλύτερη.

Γενικά, τα μοντέλα ΤΝΔ μπορούν να προβλέψουν τα περισσότερα από τα χρησιμοποιούμενα χημικά επεξεργασίας νερού και με μεγαλύτερη ακρίβεια σε σχέση με το μοντέλο πρόβλεψης με MLR ανάλυση. Παρόμοιες μελέτες πρόβλεψης χημικών δόσεων σε μια EEN έχουν δείξει ικανοποιητικά αποτελέσματα όσον αφορά την απόδοση των προβλέψεων και την υψηλή ακρίβειά τους. Συγκεκριμένα, για προβλέψεις χημικών δόσεων με χρήση ΤΝΔ, οι τιμές του RMSE κυμαίνονται από 0,64-5,93 mg/L και του R^2 από 0,742-0,940. Οι αντίστοιχες τιμές για προβλέψεις χημικών δόσεων με χρήση ανάλυσης MLR, το RMSE κυμαίνεται από 0,085-4,31 mg/L και το R^2 κυμαίνεται από 0,63-0,9 (Dadebo et al., 2023; Gyparakis et al., 2024; Haghiri et al., 2018; Lin et al., 2023b; Sharafi et al., 2024; Shi et al., 2022).

Τέλος, στο πλαίσιο της παρούσας ΔΔ, ενισχύεται περαιτέρω η άποψη ότι τα ΤΝΔ είναι χρήσιμα εργαλεία υποστήριξης αποφάσεων (Gyparakis et al., 2024; Lamrini et al., 2014; O'Reilly et al., 2018) για έναν λειτουργό μιας EEN και μπορούν να μιμηθούν με ακρίβεια και επαρκώς τις αποφάσεις, σχετικά με τις χρησιμοποιούμενες δόσεις των χημικών επεξεργασίας του νερού, που τον ενδιαφέρουν κάθε ημέρα.

5.1.4 Διερευνητική Παραγοντική Ανάλυση (EFA)

Πραγματοποιήθηκε διερευνητική παραγοντική ανάλυση των 14 μελετώμενων μεταβλητών της EEN Αποσελέμη, αφού εξετάσθηκε η κάλυψη των αναγκαίων προϋποθέσεων, από την οποία εξάγονται τα παρακάτω συμπεράσματα:

- Από τις 14 μελετώμενες μεταβλητές προκύπτει ότι υπάρχουν μεταβλητές οι οποίες συσχετίζονται, όχι όμως ισχυρά,
- όλες οι τιμές του δείκτη MSA (Measure of Sampling Adequacy) είναι $> 0,5$, οπότε

όλες οι μεταβλητές παίρνουν μέρος στην εν λόγω ανάλυση,

- στη μελέτη περίπτωσης μας, οι μεταβλητές: Chlorine gas flow (kg/h), Daily electricity consumption (kWh), Water flow at the entrance of the DWTP (m³/d), Outlet water turbidity (NTU), Outlet water residual free chlorine (mg/L) εντάσσονται στον **πρώτο (1^ο) παράγοντα**. Οι μεταβλητές ομαδοποιούνται σε μια ομάδα, η οποία θα μπορούσε να ονομαστεί **Ομάδα Ποσοτικών Χαρακτηριστικών νερού**, καθώς όταν αυξάνεται η παροχή του υπό επεξεργασία νερού, τότε αυξάνεται η συνολική κατανάλωση ηλεκτρικής ενέργειας της EEN, η κατανάλωση αερίου χλωρίου για την τελική απολύμανση του νερού και ενδεχομένως να υπάρχει μεταβολή της θολότητας του επεξεργασμένου νερού και της ποσότητας που αυτό περιέχει σε ελεύθερο υπολειμματικό χλώριο.
- Οι μεταβλητές Outlet water pH, Inlet water pH, Outlet water residual aluminum (μg/L), Polyaluminum sulfate chloride dosage (ppm) εντάσσονται στον **δεύτερο (2^ο) παράγοντα**. Οι προαναφερόμενες μεταβλητές εντάσσονται σε μια ομάδα μεταβλητών, η οποία θα μπορούσε να ονομαστεί **Ομάδα Χημικών Χαρακτηριστικών νερού**, καθώς οι μεταβλητές σχετίζονται με τη δόση PACl και το αντίστοιχο υπολειμματικό αργίλιο στο τελικά επεξεργασμένο νερό, τα οποία σχετίζονται με την τιμή του pH του ανεπεξέργαστου και του τελικά επεξεργασμένου νερού.
- Οι μεταβλητές Inlet water turbidity (NTU), Filtration bed inlet water turbidity (NTU), Daily reservoir water height difference (m) εντάσσονται στον **τρίτο (3^ο) παράγοντα**. Οι προαναφερόμενες μεταβλητές εντάσσονται σε μια ομάδα μεταβλητών, η οποία θα μπορούσε να ονομαστεί **Ομάδα Ύψους νερού ταμιευτήρα**, καθώς, πόσο ομαλά ή απότομα, αυξάνεται ή μειώνεται, καθημερινά, το ύψος του νερού στον Ταμιευτήρα Αποσελέμη, επηρεάζει τη θολότητα του εισερχόμενου νερού στην EEN και τη θολότητα του τελικά επεξεργασμένου νερού.

- Οι μεταβλητές Anionic polyelectrolyte dosage (ppm), Residual ozone (mg/L) εντάσσονται στον **τέταρτο (4^ο) παράγοντα**. Οι προαναφερόμενες μεταβλητές εντάσσονται σε μια ομάδα μεταβλητών, η οποία θα μπορούσε να ονομαστεί **Ομάδα Αποτελεσματικότητας Οζόνωσης**, καθώς η ποσότητα του υπολειμματικού όζοντος στο νερό, μετά τη διεργασία της οζόνωσης, δύναται να επηρεάσει αντιστρόφως ανάλογα την ποσότητα του AN PE, που χρησιμοποιείται για την καλύτερη συσσωμάτωση των παραγόμενων κροκίδων στη διεργασία της κροκίδο- καθίζησης.

5.2 Μελλοντικές Εργασίες

Οι προκλήσεις που αντιμετωπίζουν οι τεχνολογίες τεχνητής νοημοσύνης, όπως είναι τα ΤΝΔ, αναφορικά με μοντέλα πρόβλεψης στον τομέα του νερού και γενικότερα τα θέματα που χρήζουν περαιτέρω μελέτης, μπορούν να συνοψιστούν συνοπτικά ως:

1. Δημιουργία ολιστικών μοντέλων πρόβλεψης και λήψης αποφάσεων τεχνητής νοημοσύνης για την υποστήριξη ολόκληρου του συστήματος διαχείρισης και παροχής νερού ανθρώπινης κατανάλωσης (Li et al., 2021a).
2. Τα μοντέλα πρόβλεψης που έχουν δημιουργηθεί έως σήμερα, λαμβάνουν υπόψη τη σχέση μεταξύ των μεταβλητών ποιότητας του νερού και της εκροής σε μια εγκατάσταση επεξεργασίας νερού, αγνοώντας τις πληροφορίες που δίνονται από τα δεδομένα ακολουθιών χρονικής κλίμακας (Dunnington et al., 2021). Η μακρόχρονη παρακολούθηση και η παρατηρούμενη χρονική εξάρτηση μπορεί να παρέχει πρόσθετες πληροφορίες και να βελτιώσει την ακρίβεια των προβλέψεων.
3. Στις μέρες μας, ενδεχομένως λόγω της κλιματικής αλλαγής, παρατηρούνται ακραία καιρικά φαινόμενα. Ως εκ τούτου, είναι σημαντικό να χρησιμοποιούνται μακροπρόθεσμα δεδομένα πολλών ετών για την εκπαίδευση των αλγορίθμων των μοντέλων, με δεδομένα που να περιέχουν τέτοιου είδους ακραία επεισόδια (Yamamura et al., 2020).
4. Συνίσταται η εκπόνηση μελλοντικών μελετών για την αύξηση της γνώσης

σχετικά με την πρόβλεψη των χημικών επεξεργασίας του νερού που χρησιμοποιούνται σε μια EEN, χρησιμοποιώντας μοντέλα, όπως τα TND, ως ακριβή μοντέλα πρόβλεψης αλλά και μοντέλα ανάλυσης MLR, ως ευέλικτα, γρήγορα και αξιόπιστα μοντέλα πρόβλεψης.

5. Συγκεκριμένα, περαιτέρω έρευνα θα μπορούσε να διεξαχθεί στην πρόβλεψη των χημικών επεξεργασίας του νερού σε μια EEN, χρησιμοποιώντας TND με μικρότερο αριθμό μεταβλητών για εξασφάλιση μεγαλύτερης ευελιξίας, χωρίς να μειώνεται ουσιαστικά η αξιοπιστία του μοντέλου πρόβλεψης.
6. Επίσης, προτείνεται να καταβληθεί ακόμη μεγαλύτερη προσπάθεια για την καθιέρωση των TND ως μοντέλων πρόβλεψης στον τομέα του νερού και στην καθημερινή λειτουργία των EEN (Doorn, 2021; Xiang et al., 2021).
7. Επιπρόσθετα, οι μελλοντικές έρευνες θα μπορούσαν να περιλαμβάνουν τη διερεύνηση της χρήσης και άλλων κριτηρίων σύγκρισης, όπως το MAE (μέσο απόλυτο σφάλμα), το MAPE (μέσο απόλυτο ποσοστό σφάλματος) ή τον δείκτη απόδοσης Nash- Sutcliffe (NSE).
8. Αναφορικά, με τα αποτελέσματα της διερευνητικής παραγοντικής ανάλυσης, προτείνεται οι προτεινόμενες νέες μεταβλητές να αντικαταστήσουν τις αρχικές σε παραπέρα μελέτη, σε διαδικασίες όπως t- test, ανάλυση διακύμανσης κ.α., όπως και σε μοντελοποίηση EEN με TND και MLR ανάλυση.
9. Επιπλέον, προτείνεται η διεξαγωγή αναλύσεων ευαισθησίας και αβεβαιότητας στις μεταβλητές με τη μεγαλύτερη επιρροή, οι οποίες θα μπορούσαν να βελτιώσουν περαιτέρω τη διαδικασία μοντελοποίησης.
10. Τέλος, δεδομένου ότι ο κύριος περιορισμός της τρέχουσας ΔΔ είναι ότι τα μοντέλα έχουν εκπαιδευτεί με δεδομένα από μία μόνο EEN, προτείνεται ως μελλοντική εργασία την συμπερίληψη δεδομένων από περισσότερες EEN, προκειμένου να αυξηθεί η στιβαρότητα των μοντέλων και η καθολική εφαρμογή τους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Aalipour, M., Šťastný, B., Horký, F., Jabbarian Amiri, B., 2022. Scaling an Artificial Neural Network-Based Water Quality Index Model from Small to Large Catchments. *Water* 14, 920. <https://doi.org/10.3390/w14060920>
- Alam, G., Ihsanullah, I., Naushad, Mu., Sillanpää, M., 2022. Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes: Recent advances and prospects. *Chem. Eng. J.* 427, 130011. <https://doi.org/10.1016/j.cej.2021.130011>
- Alnuwaiser, M.A., Javed, M.F., Khan, M.I., Ahmed, M.W., Galal, A.M., 2022. Support vector regression and ANN approach for predicting the ground water quality. *J. Indian Chem. Soc.* 99, 100538. <https://doi.org/10.1016/j.jics.2022.100538>
- Alprol, A.E., Mansour, A.T., Ibrahim, M.E.E.-D., Ashour, M., 2024. Artificial Intelligence Technologies Revolutionizing Wastewater Treatment: Current Trends and Future Prospective. *Water* 16, 314. <https://doi.org/10.3390/w16020314>
- Azeem, A., Mai, W., Tian, C., Javed, Q., 2023. Dry Weight Prediction of *Wedelia trilobata* and *Wedelia chinensis* by Using Artificial Neural Network and MultipleLinear Regression Models. *Water* 15, 1896. <https://doi.org/10.3390/w15101896>
- Baouab, M.H., Cherif, S., 2018. Prediction of the optimal dose of coagulant for various potable water treatment processes through artificial neural network. *J. Hydroinformatics* 20, 1215–1226. <https://doi.org/10.2166/hydro.2018.014>
- Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods* 43, 3–31. [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3)
- Bhattacharya, A., Sahu, S., Telu, V., Duttagupta, S., Sarkar, S., Bhattacharya, J., Mukherjee, A., Ghosal, P.S., 2021. Neural Network and Random Forest-Based Analyses of the Performance of Community Drinking Water Arsenic Treatment Plants. *Water* 13, 3507. <https://doi.org/10.3390/w13243507>
- Bridgewater, L.L., Baird, R.B., Eaton, A.D., Rice, E.W., *American Public Health*

- Association, American Water Works Association, Water Environment Federation (Eds.), 2017. Standard methods for the examination of water and wastewater, 23rd edition. ed. American Public Health Association, Washington, DC.
- Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., Ren, H., 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* 171, 115454. <https://doi.org/10.1016/j.watres.2019.115454>
- D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, T. L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE* 50, 885–900. <https://doi.org/10.13031/2013.23153>
- Dadebo, D., Obura, D., Etyang, N., Kimera, D., 2023. Economic and social perspectives of implementing artificial intelligence in drinking water treatment systems for predicting coagulant dosage: A transition toward sustainability. *Groundw. Sustain. Dev.* 23, 100987. <https://doi.org/10.1016/j.gsd.2023.100987>
- Ding, S., Li, H., Su, C., Yu, J., Jin, F., 2013. Evolutionary artificial neural networks: a review. *Artif. Intell. Rev.* 39, 251–260. <https://doi.org/10.1007/s10462-011-9270-6>
- Doorn, N., 2021. Artificial intelligence in the water domain: Opportunities for responsible use. *Sci. Total Environ.* 755, 142561. <https://doi.org/10.1016/j.scitotenv.2020.142561>
- Dunnington, D.W., Trueman, B.F., Raseman, W.J., Anderson, L.E., Gagnon, G.A., 2021. Comparing the Predictive Performance, Interpretability, and Accessibility of Machine Learning and Physically Based Models for Water Treatment. *ACS EST Eng.* 1, 348–356. <https://doi.org/10.1021/acsestengg.0c00053>
- Farmaki, E.G., Thomaidis, N.S., Simeonov, V., Efsthathiou, C.E., 2013. Comparative use of artificial neural networks for the quality assessment of the water reservoirs of Athens. *J. Water Supply Res. Technol.-Aqua* 62, 296–308. <https://doi.org/10.2166/aqua.2013.108>
- Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos. Environ.* 32, 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0)
- Ghaedi, A.M., Vafaei, A., 2017. Applications of artificial neural networks for adsorption removal of dyes from aqueous solution: A review. *Adv. Colloid Interface Sci.* 245, 20–39. <https://doi.org/10.1016/j.cis.2017.04.015>
- Griffiths, K.A., Andrews, R.C., 2011. The application of artificial neural networks for the optimization of coagulant dosage. *Water Supply* 11, 605–611. <https://doi.org/10.2166/ws.2011.028>
- Gyparakis, S., Trichakis, I., Daras, T., Diamadopoulos, E., 2025. Artificial Neural Networks (ANNs) and Multiple Linear Regression (MLR) Analysis Modelling for Predicting Chemical Dosages of a Water Treatment Plant (WTP) of Drinking Water. *Water* 17, 227. <https://doi.org/10.3390/w17020227>
- Gyparakis, S., Trichakis, I., Diamadopoulos, E., 2024. Using Artificial Neural Networks to Predict Operational Parameters of a Drinking Water Treatment Plant (DWTP). *Water* 16, 2863. <https://doi.org/10.3390/w16192863>
- Haghir, S., Daghighi, A., Moharramzadeh, S., 2018. Optimum coagulant forecasting by modeling jar test experiments using ANNs. *Drink. Water Eng. Sci.* 11, 1–8.

- <https://doi.org/10.5194/dwes-11-1-2018>
- Hassan, E.S., Alharbi, A.A., Oshaba, A.S., El-Emary, A., 2024. Enhancing Smart Irrigation Efficiency: A New WSN-Based Localization Method for Water Conservation. *Water* 16, 672. <https://doi.org/10.3390/w16050672>
- Haykin, S.S., Haykin, S.S., 2009. *Neural networks and learning machines*, 3rd ed. ed. Prentice Hall, New York.
- Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water systems. *Water Res.* 205, 117666. <https://doi.org/10.1016/j.watres.2021.117666>
- Kim, C.M., Parnichkun, M., 2017a. MLP, ANFIS, and GRNN based real-time coagulant dosage determination and accuracy comparison using full-scale data of a water treatment plant. *J. Water Supply Res. Technol. - Aqua* 66, 49–61. <https://doi.org/10.2166/aqua.2016.022>
- Kim, C.M., Parnichkun, M., 2017b. MLP, ANFIS, and GRNN based real-time coagulant dosage determination and accuracy comparison using full-scale data of a water treatment plant. *J. Water Supply Res. Technol. - Aqua* 66, 49–61. <https://doi.org/10.2166/aqua.2016.022>
- Kim, Y., Kwak, S., Lee, M., Jeong, M., Park, M., Park, Y.-G., 2023. Determination of Optimal Water Intake Layer Using Deep Learning-Based Water Quality Monitoring and Prediction. *Water* 16, 15. <https://doi.org/10.3390/w16010015>
- Lamrini, B., Lakhal, E.K., Le Lann, M.V., 2014. A decision support tool for technical processes optimization in drinking water treatment. *Desalination Water Treat.* 52, 4079–4088. <https://doi.org/10.1080/19443994.2013.803327>
- Li, L., Rong, S., Wang, R., Yu, S., 2021a. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chem. Eng. J.* 405, 126673. <https://doi.org/10.1016/j.cej.2020.126673>
- Li, L., Rong, S., Wang, R., Yu, S., 2021b. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chem. Eng. J.* 405, 126673. <https://doi.org/10.1016/j.cej.2020.126673>
- Lin, S., Kim, J., Hua, C., Kang, S., Park, M.-H., 2023a. Comparing artificial and deep neural network models for prediction of coagulant amount and settled water turbidity: Lessons learned from big data in water treatment operations. *J. Water Process Eng.* 54, 103949. <https://doi.org/10.1016/j.jwpe.2023.103949>
- Lin, S., Kim, J., Hua, C., Park, M.-H., Kang, S., 2023b. Coagulant dosage determination using deep learning-based graph attention multivariate time series forecasting model. *Water Res.* 232, 119665. <https://doi.org/10.1016/j.watres.2023.119665>
- Lowe, M., Qin, R., Mao, X., 2022. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water* 14, 1384. <https://doi.org/10.3390/w14091384>
- Maloney, K.O., Buchanan, C., Jepsen, R.D., Krause, K.P., Cashman, M.J., Gressler, B.P., Young, J.A., Schmid, M., 2022. Explainable machine learning improves interpretability in the predictive modeling of biological stream conditions in the Chesapeake Bay Watershed, USA. *J. Environ. Manage.* 322, 116068. <https://doi.org/10.1016/j.jenvman.2022.116068>
- Mu'azu, N.D., 2022. Insight into ANN and RSM Models' Predictive Performance for Mechanistic Aspects of Cr(VI) Uptake by Layered Double Hydroxide

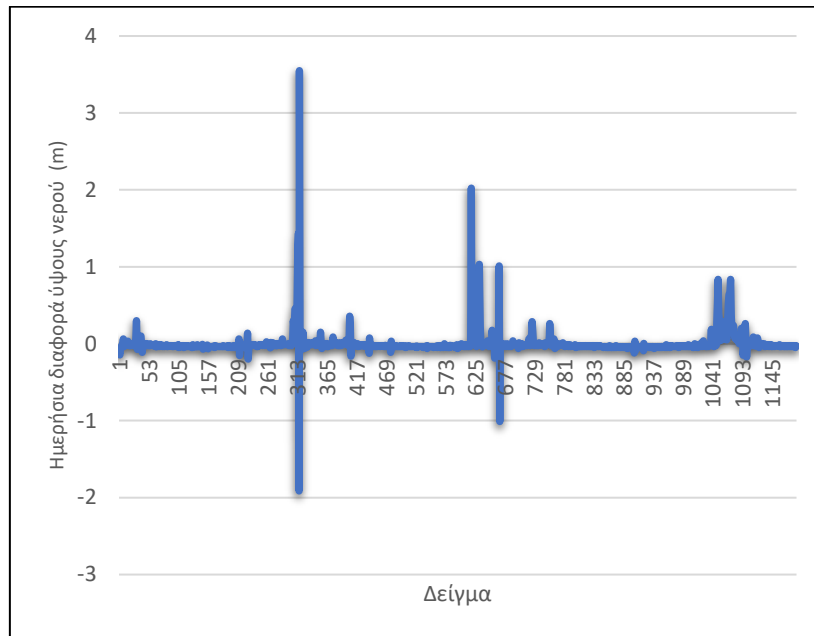
- Nanocomposites from Water. Water 14, 1644. <https://doi.org/10.3390/w14101644>
- Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., Al-Shamma'a, A., 2022. Water quality classification using machine learning algorithms. J. Water Process Eng. 48, 102920. <https://doi.org/10.1016/j.jwpe.2022.102920>
- O'Reilly, G., Bezuidenhout, C.C., Bezuidenhout, J.J., 2018. Artificial neural networks: applications in the drinking water sector. Water Supply 18, 1869–1887. <https://doi.org/10.2166/ws.2018.016>
- Özdoğan, H., Üncü, Y.A., Şekerci, M., Kaplan, A., 2024. Neural network predictions of (α, n) reaction cross sections at 18.5 ± 3 MeV using the Levenberg-Marquardt algorithm. Appl. Radiat. Isot. 204, 111115. <https://doi.org/10.1016/j.apradiso.2023.111115>
- Papailiou, I., Spyropoulos, F., Trichakis, I., Karatzas, G.P., 2022. Artificial Neural Networks and Multiple Linear Regression for Filling in Missing Daily Rainfall Data. Water 14, 2892. <https://doi.org/10.3390/w14182892>
- Sharafi, M., Rezaverdinejad, V., Behmanesh, J., Samadianfard, S., 2024. Development of long short-term memory along with differential optimization and neural networks for coagulant dosage prediction in water treatment plant. J. Water Process Eng. 65, 105784. <https://doi.org/10.1016/j.jwpe.2024.105784>
- Shi, Z., Chow, C.W.K., Fabris, R., Liu, J., Sawade, E., Jin, B., 2022. Determination of coagulant dosages for process control using online UV-vis spectra of raw water. J. Water Process Eng. 45, 102526. <https://doi.org/10.1016/j.jwpe.2021.102526>
- Stylianoudaki, C., Trichakis, I., Karatzas, G.P., 2022. Modeling Groundwater Nitrate Contamination Using Artificial Neural Networks. Water 14, 1173. <https://doi.org/10.3390/w14071173>
- Tabari, H., Hosseinzadeh Talaei, P., 2015. Reconstruction of river water quality missing data using artificial neural networks. Water Qual. Res. J. 50, 326–335. <https://doi.org/10.2166/wqrjc.2015.044>
- Wang, A., Wang, J., Luan, B., Wang, S., Yang, D., Wei, Z., 2024. Classification of Pollution Sources and Their Contributions to Surface Water Quality Using APCS-MLR and PMF Model in a Drinking Water Source Area in Southeastern China. Water 16, 1356. <https://doi.org/10.3390/w16101356>
- Wongburi, P., Park, J.K., 2023. Prediction of Wastewater Treatment Plant Effluent Water Quality Using Recurrent Neural Network (RNN) Models. Water 15, 3325. <https://doi.org/10.3390/w15193325>
- Wu, G.-D., Lo, S.-L., 2010. Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neural network. Expert Syst. Appl. 37, 4974–4983. <https://doi.org/10.1016/j.eswa.2009.12.016>
- Wu, G.-D., Lo, S.-L., 2008. Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system. Eng. Appl. Artif. Intell. 21, 1189–1195. <https://doi.org/10.1016/j.engappai.2008.03.015>
- Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. Environ. Model. Softw. 54, 108–127. <https://doi.org/10.1016/j.envsoft.2013.12.016>
- Xiang, X., Li, Q., Khan, S., Khalaf, O.I., 2021. Urban water resource management for

- sustainable environment planning using artificial intelligence techniques. *Environ. Impact Assess. Rev.* 86, 106515. <https://doi.org/10.1016/j.eiar.2020.106515>
- Yamamura, H., Putri, E.U., Kawakami, T., Suzuki, A., Ariesyady, H.D., Ishii, T., 2020. Dosage optimization of polyaluminum chloride by the application of convolutional neural network to the floc images captured in jar tests. *Sep. Purif. Technol.* 237, 116467. <https://doi.org/10.1016/j.seppur.2019.116467>
- Yateh, M., Lartey-Young, G., Li, F., Li, M., Tang, Y., 2023. Application of Response Surface Methodology to Optimize Coagulation Treatment Process of Urban Drinking Water Using Polyaluminium Chloride. *Water* 15, 853. <https://doi.org/10.3390/w15050853>
- Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J., Yang, Y., 2020. Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. *Process Saf. Environ. Prot.* 133, 169–182. <https://doi.org/10.1016/j.psep.2019.11.014>
- Žic, M., Pereverzyev, S., 2023. Application of self-adapting regularization, machine learning tools and limits in Levenberg–Marquardt algorithm to solve CNLS problem. *J. Electroanal. Chem.* 939, 117420. <https://doi.org/10.1016/j.jelechem.2023.117420>

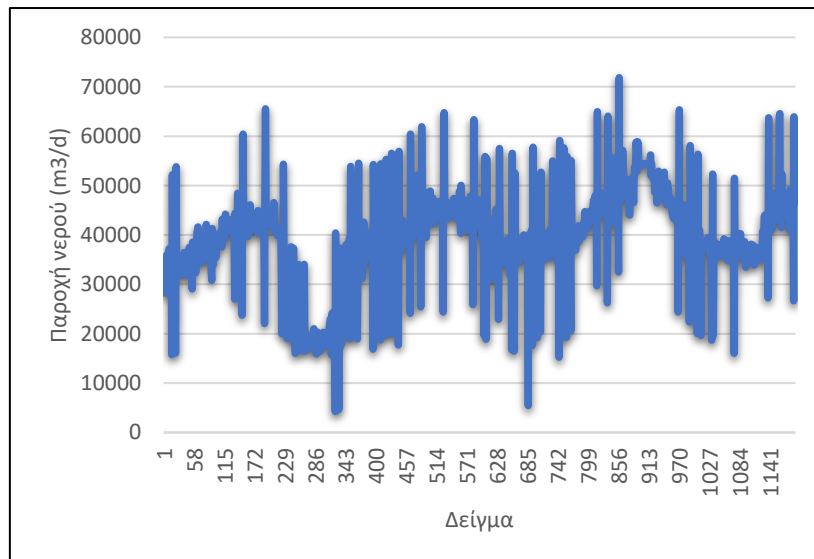
ΠΑΡΑΡΤΗΜΑΤΑ

ΠΑΡΑΡΤΗΜΑ Ι – ΔΙΑΓΡΑΜΜΑΤΑ ΧΡΟΝΟΣΕΙΡΩΝ

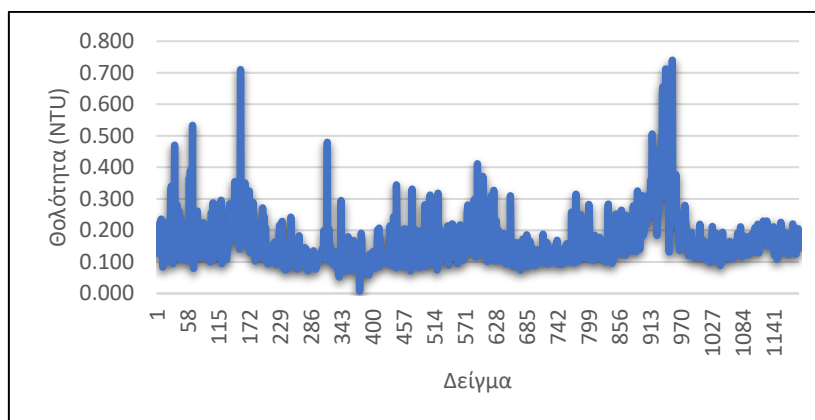
ΜΕΤΑΒΛΗΤΩΝ



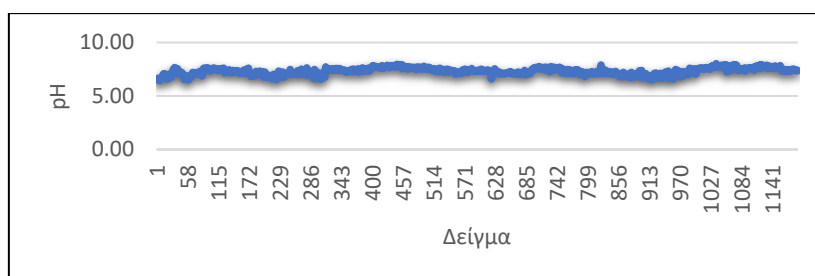
Σχήμα Π 1: Ημερήσια διαφορά ύψους νερού στον Ταμιευτήρα



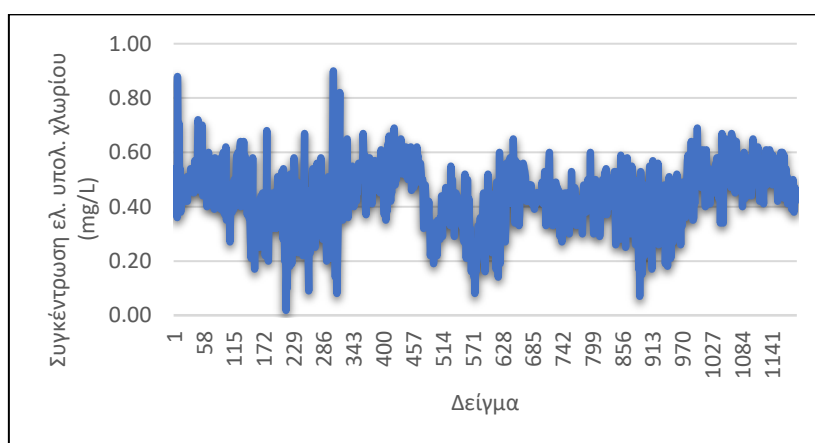
Σχήμα Π 2: Παροχή ανεπεξέργαστου νερού



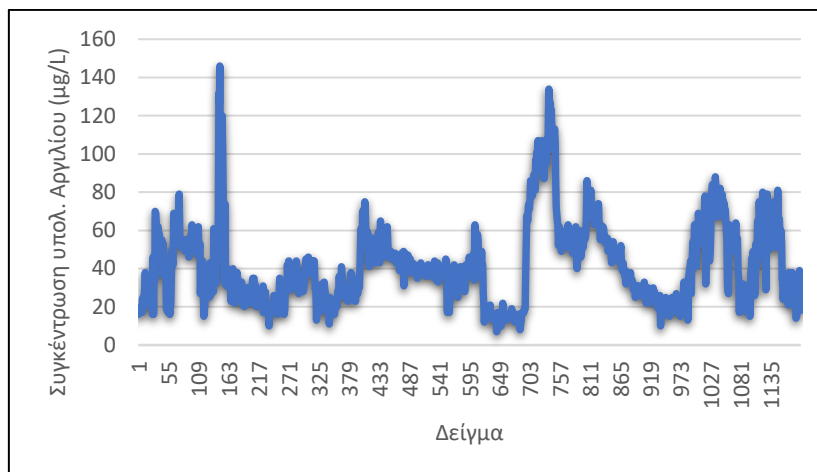
Σχήμα Π 3: Θολότητα επεξεργασμένου νερού



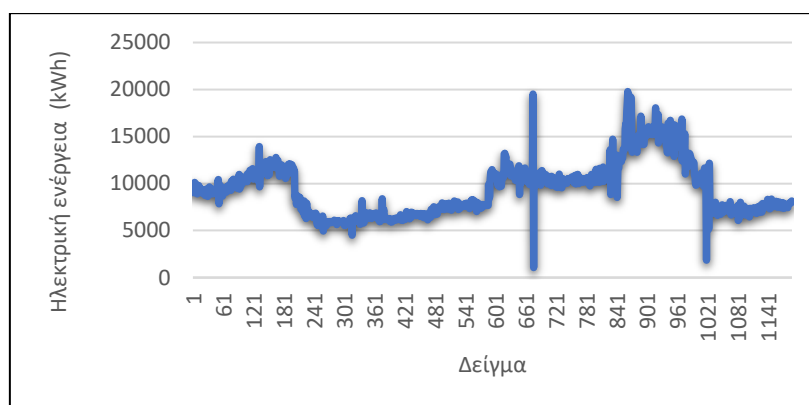
Σχήμα Π 4: pH επεξεργασμένου νερού



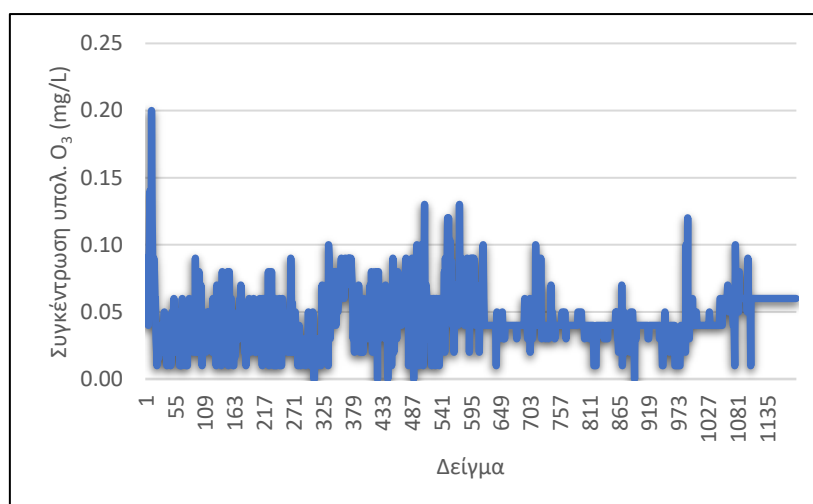
Σχήμα Π 5: Συγκέντρωση ελεύθερου υπολειμματικού χλωρίου επεξεργασμένου νερού



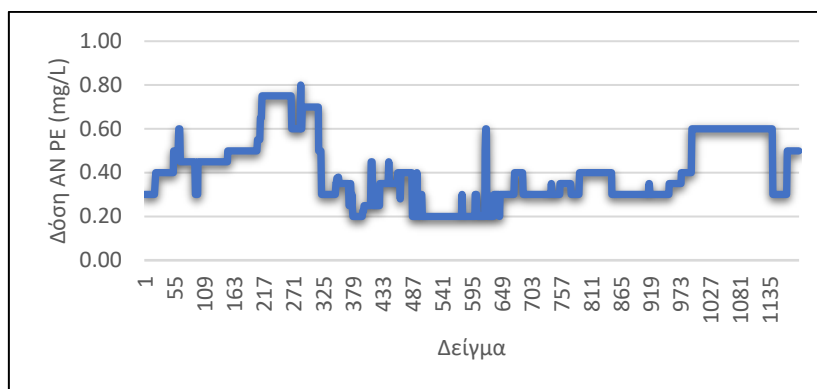
Σχήμα Π 6: Συγκέντρωση υπολειμματικού Αργιλίου επεξεργασμένου νερού



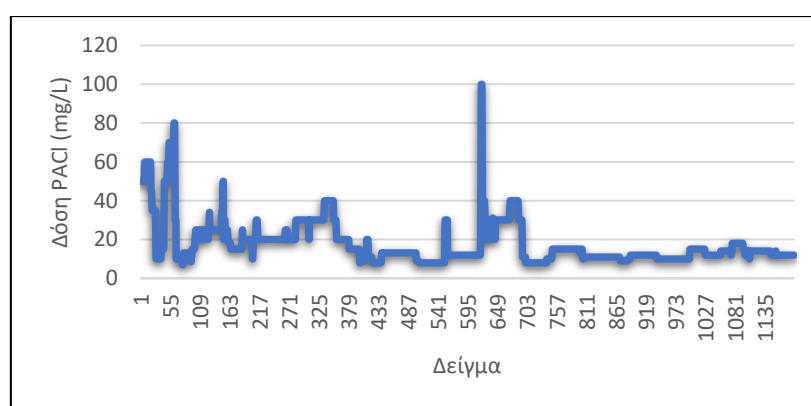
Σχήμα Π 7: Ημερήσια καταναλισκόμενη ηλεκτρική ενέργεια ΕΕΝ Αποσελέμη



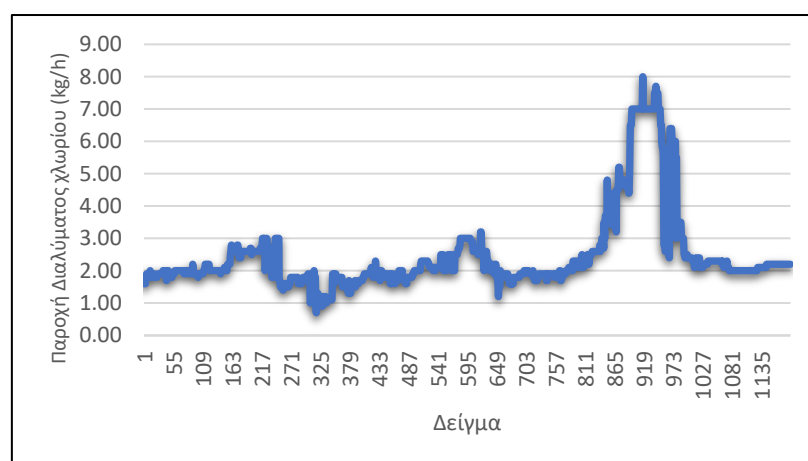
Σχήμα Π 8: Συγκέντρωση υπολειμματικού όζοντος (O₃)



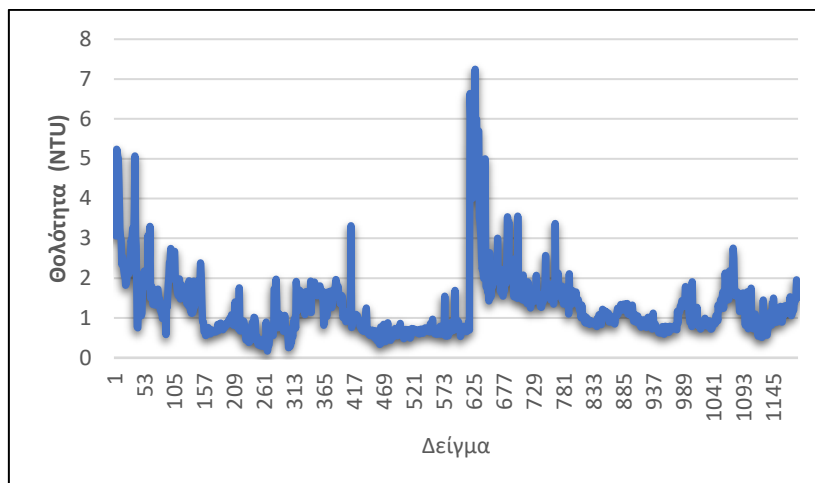
Σχήμα Π 9: Δόση ανιονικού πολυηλεκτρολύτη (AN PE)



Σχήμα Π 10: Δόση χλωριούχου θειικού πολύ- αργιλίου (PACl)



Σχήμα Π 11: Παροχή Διαλύματος χλωρίου ($Cl_{2(g)}$)



Σχήμα Π 12: Θολότητα νερού εισόδου κλινών διήθησης

ΠΑΡΑΡΤΗΜΑ II – ΣΥΝΤΕΛΕΣΤΕΣ ΣΥΣΧΕΤΙΣΗΣ

PEARSON

Πίνακας Π 1: Συντελεστές συσχέτισης Pearson των μελετώμενων μεταβλητών

Pearson Correlation	Residual ozone (mg/L)	Anionic polyelectrolyte dosage (ppm)	Polyaluminum sulfate chloride dosage (ppm)	Chlorine gas flow (kg/h)
Daily reservoir water height difference (m)	-0,021	0,080	0,062	-0,088
Water flow at the entrance of the DWTP (m ³ /d)	0,062	-0,356	-0,354	0,477
Inlet water turbidity (NTU)	-0,032	-0,006	0,229	0,032
Inlet water pH	0,233	-0,095	-0,281	-0,269
Outlet water turbidity (NTU)	-0,073	-0,060	-0,059	0,541
Outlet water pH	0,201	-0,063	-0,367	-0,415
Outlet water residual free chlorine (mg/L)	0,082	0,069	0,116	-0,258
Outlet water residual Aluminum (μg/L)	-0,018	0,006	-0,373	-0,206
Daily electricity consumption (kWh)	-0,201	-0,237	-0,111	0,704
Filtration bed inlet water turbidity (NTU)	0,045	-0,152	0,458	-0,156

ΠΑΡΑΡΤΗΜΑ ΙΙΙ – ΚΩΔΙΚΑΣ ΣΤΗ MATLAB

```
tic
load('full_aposelemis.mat')
clear bestperf bestnet besttr y net perf leg
leg={};
imax=100; %number of ANNs run for each combination
maxnodes=60; %maximum allowed number of hidden nodes in each ANN
pmax=size(inputs_combinations,1); %number of possible combinations
perf=zeros(pmax, imax);
nodesnumber=zeros(pmax, imax);
%The following loop goes through all possible combinations, creates the
%predetermined number of ANNs for each combination and keeps the best,
%together with all the results in a separate .mat file in the current
%folder
for p=1:pmax
    bestperf=Inf;
    target=data_norm(:,11:14)'; %Transpose values (ANN format requirement)
    %The following selection creates the input matrix according to the
    %current combination. If an input parameter is missing, it is not
    %included in the input matrix for this combination.
    input=data_norm(:,logical(inputs_combinations(p,:)))';
    for i=1:imax %number of ANNs
        disp(i)
        nodesnumber(p,i)=floor(maxnodes*rand(1)+1);
        net = feedforwardnet(nodesnumber(p,i)); %max number of hidden nodes
        %net.layers{1}.transferFcn = 'poslin';
```

```
[net, tr] = train(net,input,target);  
y{i} = net(input);  
perf(p,i) = perform(net,y{i},target)+0.0001*sum(sum(y{i}<0));  
dummy=y{i};  
dummy(dummy<0)=0;  
y{i}=dummy;  
if perf(p,i)<bestperf  
    bestperf=perf(p,i);  
    bestnet=net;  
    besttr=tr;  
    besty=y{i};  
    besti=i;  
end  
end  
%figure()  
for n=1:size(target,1)  
    DV= bestnet(input);  
    DV(DV<0)=0;  
    figure(1)  
    subplot(4,4,p)  
    palette='brgy';  
    hold on  
    %dates=1:1188;  
    opto=[palette(mod(n,4)+1), '*'];  
    opts=[palette(mod(n,4)+1), '-'];  
    plot(dates,target(n,:),opto,dates,DV(n,:),opts)  
    ob=['observed',num2str(n)];  
    si=['simulated',num2str(n)];
```

```
leg{2*n-1}=ob;
leg{2*n}=si;
title(num2str(find(inputs_combinations(p,:))))
legend(leg, 'location','NorthEastOutside')
datetick('x','yyyy')
xlim([min(dates),max(dates)])
ylabel('Normalised value')
figure(2)
subplot(4,4,p)
plot(nodesnumber(p,:), perf(p,:), '*')
title(num2str(find(inputs_combinations(p,:))))
xlabel('Nodes number')
ylabel('Performance')
end
T= ['resultnet',num2str(p),'.mat'];
save(T,'best*','y','target')
clear input target bestperf bestnet besttr besty besti
leg={};
end
toc
```