

# ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΠΡΟΕΛΕΥΣΗΣ ΥΠΟΓΕΙΩΝ ΔΙΑΡΡΟΩΝ ΜΕ ΓΡΑΜΜΙΚΗ ΔΙΑΧΩΡΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ε. Γ. Φαρμάκη<sup>1</sup>, Ν. Κ. Θανασούλιας<sup>2</sup>, Α. Χ. Καγιάρα, Σ. Α. Σάμιος, Δ. Θ. Τσοροβά,  
Φ. Τ. Μισκάκη, Ε. Σ. Λύτρας

Εταιρεία Ύδρευσης και Αποχέτευσης Πρωτεύουσας (ΕΥΔΑΠ ΑΕ), Διεύθυνση Ποιότητας,  
Έρευνας και Ανάπτυξης (R&D), Υπηρεσία Ποιότητας Νερού, Ωρωπού 156, 11146, Γαλάτσι,  
Αθήνα

<sup>1</sup>[hfarmaki@eydap.gr](mailto:hfarmaki@eydap.gr), <sup>2</sup>[nthanas@eydap.gr](mailto:nthanas@eydap.gr)

## Περίληψη

105 δείγματα νερών από διαρροές σε υπόγειους χώρους ακινήτων αναλύθηκαν από το Χημικό Εργαστήριο Γαλασίου της Υπηρεσίας Ποιότητας Νερού της ΕΥΔΑΠ ΑΕ για την ταυτοποίησή τους ως δείγματα προερχόμενα από το δίκτυο ύδρευσης, το δίκτυο αποχέτευσης ή από άγνωστη πηγή. Στα δείγματα προσδιορίστηκαν φωτομετρικά οι παράμετροι  $Al$ ,  $NO_2^-$ ,  $NH_4^+$ ,  $Cl^-$  και  $NO_3^-$  καθώς και η αγωγιμότητά τους (ηλεκτροχημική μέθοδος). Ως πραγματική ταυτότητα των δειγμάτων χρησιμοποιήθηκε η γνωμάτευση των Τεχνικών Τομέων Ύδρευσης και Αποχέτευσης της ΕΥΔΑΠ ΑΕ η οποία προέκυψε από επιτόπιους ελέγχους. Τα δείγματα ελέγχθησαν ως προς την συνολική τους διασπορά με την απεικόνισή τους σε καρτεσιανό επίπεδο συντεταγμένων οριζόμενο από τις δύο πρώτες βασικές συνιστώσες που εξήχθησαν με ανάλυση βασικών συνιστωσών (PCA). Προσδιορίστηκαν δύο παράγοντες που ονομάστηκαν «παράγοντας αλατότητας» και «παράγοντας παρουσίας λυμάτων». Τα δεδομένα υποβλήθηκαν στη συνέχεια σε γραμμική διαχωριστική ανάλυση (LDA) με στόχο τη βελτιστοποίηση του διαχωρισμού τους. Το σύνολο επαλήθευσης ταξινομήθηκε επιτυχώς κατά 74% με τη βοήθεια πρωτίστως της πρώτης ρίζας («παράγοντας παρουσίας λυμάτων») και δευτερευόντως λόγω της δεύτερης ρίζας («παράγοντας αλατότητας»). Η επαλήθευση έγινε με τη μέθοδο της διασταυρούμενης επικύρωσης (cross-validation ή jackknife). Και οι δύο ρίζες βρέθηκαν στατιστικά σημαντικές σε επίπεδο  $p=0,05$ . Παρατηρήθηκε αύξηση του συνολικού ποσοστού επιτυχίας ταξινόμησης κατά 21% σε σχέση με την εκτίμηση προέλευσης των δειγμάτων λόγω εμπειρίας των αναλυτών. Παράλληλα, με την επεξεργασία της ίδιας βάσης δεδομένων κατασκευάστηκαν μοντέλα Νευρωνικών Δικτύων (ANNs), κάνοντας χρήση των πιο δημοφιλών τεχνικών (MLPs και Linear). Το καλύτερο μοντέλο πέτυχε ποσοστό 87% για την «άγνωστη» ομάδα ελέγχου με μειωμένα φαινόμενα υπέρ-προσαρμογής όπως φάνηκε από τη σύγκριση των ποσοστών στις ομάδες εκπαίδευσης και επικύρωσης.

*Λέξεις-κλειδιά: υπόγειες διαρροές, ταξινόμηση, ανάλυση βασικών συνιστωσών, γραμμική διαχωριστική ανάλυση, νευρωνικά δίκτυα*

## Abstract

105 water samples located in basements and other places in urban areas were analyzed photometrically for  $Al$ ,  $NO_2^-$ ,  $NH_4^+$ ,  $Cl^-$  and  $NO_3^-$ . The sample conductivities were also recorded. Samples were taken to belong to one of three classes, namely drinking water (DW), wastewater (WW) and of unknown origin (U) according to the data provided by the relevant technical departments of The Athens Water Supply and Sewerage Company,

EYDAP SA. Principal Component Analysis (PCA) was used to visualize the data on a plane defined by the first two principal components. The factors extracted were named “salinity” and “waste”. Linear Discriminant Analysis (LDA) produced two canonical roots that were found statistically significant at the  $p=0.05$  level. The first root was attributed to “waste” and the second one was assigned to “salinity”. Cross-validation of the model produced a 74% correct classification of the test set when the jackknife method was employed. This corresponded to a 21% increase in correct overall classification when the human empirical approach was taken into account. The same database was used for the construction of ANNs models. Two of the most popular architectures were used (MLPs and Linear). The best ANNs model achieved a correct classification percentage of 87% in the “unknown” group of the test samples. The overtraining phenomena were minimal as it was shown by the comparison of the success percentages in the training and validation samples.

*Keywords: water leakage, classification, principal component analysis, linear discriminant analysis, artificial neural networks*

## 1. Εισαγωγή

Συχνά, σε οικιστικές περιοχές της δικαιοδοσίας της ΕΥΔΑΠ ΑΕ, εμφανίζεται το φαινόμενο της συσσώρευσης υδάτων σε υπόγειους χώρους ακινήτων ως αποτέλεσμα αφανών διαρροών του δικτύου ύδρευσης και αποχέτευσης ή εισχώρησης υδάτων από άλλες πηγές όπως ο υδροφόρος ορίζοντας, τα όμβρια ύδατα κλπ. (Piaia, Cheriaf et al. 2013). Η συσσώρευση των υδάτων αυτών προκαλεί σοβαρά προβλήματα τόσο στους καταναλωτές (λόγω φθοράς περιουσιακών τους στοιχείων) όσο και στην ΕΥΔΑΠ ΑΕ (λόγω οικονομικών επιπτώσεων, στην περίπτωση που τα διαρρέοντα νερά προέρχονται από αφανείς διαρροές των δικτύων της) (Ashton and Hope 2001; Morais and de Almeida 2007; Delgado-Galván, Pérez-García et al. 2010).

Είναι προφανές λοιπόν ότι η αντιμετώπιση του προβλήματος των διαρροών είναι μεγίστης σημασίας (Li, Ling et al. 2011; Britton, Stewart et al. 2013; Candelieri, Conti et al. 2014). Στην περίπτωση της ΕΥΔΑΠ ΑΕ, επεμβαίνει πρωτίστως ο αντίστοιχος Τομέας Ύδρευσης ή Αποχέτευσης (ανάλογα με τα χαρακτηριστικά της κάθε περίπτωσης) και προσπαθεί μέσω παρατήρησης και φυσικών τεχνικών (π.χ. ακρόαση εδάφους) να εντοπίσει την πιθανή προέλευση του συσσωρευόμενου νερού και στην περίπτωση που αυτό προέρχεται από το δίκτυο ύδρευσης ή αποχέτευσης να προβεί στην αποκατάσταση του προβλήματος.

Στην περίπτωση που οι ενέργειες του Τομέα δεν αποβούν καρποφόρες, κατάλληλο δείγμα νερού συλλέγεται από τους τεχνικούς και αποστέλλεται στο Χημικό Εργαστήριο Γαλασίου (ΧΕΓ) της Υπηρεσίας Ποιότητας Νερού, το οποίο καλείται να γνωμοδοτήσει σχετικά με την προέλευση του δείγματος βασιζόμενο στην φυσικοχημική του εξέταση. Τα δείγματα τα οποία καταλήγουν στο Χημικό Εργαστήριο αποτελούν ένα σημαντικό αριθμό επί του συνόλου των δειγμάτων (περίπου 600 δείγματα ετησίως). Η συγκεκριμένη διαδικασία που εφαρμόζεται μέχρι σήμερα στο εργαστήριο περιλαμβάνει διήθηση του δείγματος με ηθμό σύριγγας και προσδιορισμό  $Al$ ,  $NO_2^-$ ,  $NH_4^+$ ,  $Cl^-$  και  $NO_3^-$  στο διήθημα. Ταυτόχρονα προσδιορίζεται και η αγωγιμότητα του δείγματος ως έχει. Τα δείγματα κατηγοριοποιούνται σε δείγματα δικτύου ύδρευσης, αποχέτευσης, αγνώστου προελεύσεως ή μίγματα των παραπάνω με εμπειρικό τρόπο βάσει κυρίως της αγωγιμότητάς τους και του περιεχομένου τους σε  $NO_2^-$  και  $NH_4^+$  (δείκτες νερού αποχέτευσης) και  $Al$  (δείκτης νερού ύδρευσης).

Η μέθοδος αυτή της κατηγοριοποίησης των δειγμάτων αν και βασίζεται στη χημική εμπειρία του αναλυτή παρουσιάζει το μειονέκτημα της έλλειψης μοντέλου για την ταξινόμηση και επιπλέον απαιτεί την εκτεταμένη εκπαίδευση νέων αναλυτών οι οποίοι καλούνται να αποφασίσουν για την προέλευση των δειγμάτων (Bermúdez and García-García 2012; Bleam 2012). Για το λόγο αυτό, στην παρούσα εργασία μελετάται η δυνατότητα ταξινόμησης δειγμάτων υπογείων διαρροών με τη μέθοδο της γραμμικής διαχωριστικής ανάλυσης (Linear Discriminant Analysis, LDA) και τη μέθοδο των νευρωνικών δικτύων (Artificial Neural Networks, ANNs). Και για τις δύο μεθόδους, χρησιμοποιήθηκαν τα δεδομένα αναλύσεων 105 δειγμάτων τα οποία παρελήφθησαν στο ΧΕΓ κατά το έτος 2013 και αναλύθηκαν ως προς τις παραμέτρους: αγωγιμότητα, Al, NO<sub>2</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>, Cl<sup>-</sup> και NO<sub>3</sub><sup>-</sup>. Βάσει των αυτοψιών των Τομέων, τα δείγματα θεωρήθηκαν ως προερχόμενα από το δίκτυο ύδρευσης (Drinking Water, DW), το δίκτυο αποχέτευσης (Wastewater, WW) ή αγνώστου προέλευσης (Unknown, U). Αρχικά, τα δείγματα απεικονίστηκαν σε επίπεδο με τη μέθοδο της ανάλυσης βασικών συνιστωσών (Principal Component Analysis, PCA) ώστε να είναι δυνατή μία πρώτη εκτίμηση της εικόνας που παρουσίαζαν ως προς τη διασπορά τους. Εξήχθησαν δύο συνιστώσες οι οποίες εξήγησαν το 58% της διακύμανσης ενώ το διάγραμμα φόρτισης των συνιστωσών έδειξε ότι υφίστανται δύο κύριοι παράγοντες οι οποίοι ονομάστηκαν «παράγοντας αλατότητας» και «παράγοντας παρουσίας λυμάτων». Το μοντέλο της LDA σχηματίστηκε με χρήση όλων των δειγμάτων και για όλες τις φυσικοχημικές παραμέτρους ενώ έγινε επαλήθευσή του με τη μέθοδο της διασταυρούμενης επικύρωσης (cross-validation ή jackknife). Υπολογίστηκαν δύο κανονικές ρίζες οι οποίες βρέθηκαν σημαντικές σε επίπεδο αξιοπιστίας p=0,05. Η πρώτη ρίζα συσχετίστηκε με τον «παράγοντα παρουσίας λυμάτων», ενώ η δεύτερη με τον «παράγοντα αλατότητας». Ο πίνακας επαλήθευσης έδειξε συνολικά 74% ορθή ταξινόμηση των δειγμάτων επαλήθευσης με ορθή ταξινόμηση κατά 90% για τα πόσιμα δείγματα, 48% για τα δείγματα αποχέτευσης και 18% για τα δείγματα αγνώστου προελεύσεως. Τα ίδια δεδομένα χρησιμοποιήθηκαν για την κατασκευή μοντέλων ANNs: Γραμμικά (Linear) και Πολυστιβαδικά Νευρωνικά Δίκτυα (Multi-Layers Perceptrons, MLPs). Πολλά μοντέλα κατασκευάστηκαν με τα καλύτερα εξ' αυτών να «εκμεταλλεύονται» από 2 ως 6 κρισιμότερες μεταβλητές. Οι πίνακες επαλήθευσης έδειξαν μέχρι 85 % και 87 % ορθή ταξινόμηση των δειγμάτων εκπαίδευσης και ελέγχου αντίστοιχα.

## 2. Μεθοδολογία

Οι αναλύσεις των δειγμάτων υπογείων διαρροών έγιναν κατόπιν διήθησης περίπου 10 ml δείγματος με ηθμό σύριγγας. Το διήθημα αναλύθηκε για τις παραμέτρους Al, NO<sub>2</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>, Cl<sup>-</sup> και NO<sub>3</sub><sup>-</sup> σε αυτόματα φωτομετρικό αναλυτή Aquakem 250 της εταιρίας Thermo Fisher Scientific σύμφωνα με τις ισχύουσες οδηγίες εργασίας του Χημικού Εργαστηρίου Γαλασίου. Για κάθε δείγμα καταγράφηκε επίσης η αγωγιμότητά του. Οι γνωματεύσεις των Τομέων σχετικά με την προέλευση του δείγματος (DW: νερό δικτύου υδρεύσεως ΕΥΔΑΠ, WW: νερό δικτύου αποχέτευσης ΕΥΔΑΠ, U: νερό αγνώστου προελεύσεως) αντλήθηκαν από τη βάση αποτελεσμάτων του Τμήματος Παραπόνων (1022) της ΕΥΔΑΠ ΑΕ. Για τη στατιστική ανάλυση των αποτελεσμάτων χρησιμοποιήθηκε το στατιστικό πακέτο STATISTICA 7 της εταιρίας StatSoft Inc.

### 3. Αποτελέσματα και συζήτηση

#### 3.1 Ανάλυση βασικών συνιστωσών (PCA)

Προς εκτίμηση της εικόνας των δεδομένων και των σχέσεων που διέπουν τις προσδιοριζόμενες σε αυτά παραμέτρους εφαρμόστηκε ανάλυση βασικών συνιστωσών (PCA) (Syms 2008; Esbensen and Geladi 2009). Για τον σκοπό αυτό εξήγησαν οι δύο πρώτες συνιστώσες οι οποίες εξήγησαν το 58% της συνολικής διακύμανσης των δεδομένων (Πίνακας 1).

Πίνακας 1: Βασικές συνιστώσες και ιδιοτιμές τους

Συνιστώσα	Ποσοστό διακύμανσης που ερμηνεύεται	Συνολικό ποσοστό διακύμανσης που ερμηνεύεται	Ιδιοτιμές
1	0,373	0,373	2,238
2	0,206	0,579	1,238

Οι φορτίσεις των συνιστωσών (Πίνακας 2) κατόπιν περιστροφής τους (μέθοδος Varimax) έδειξαν ότι η πρώτη συνιστώσα αφορούσε τις παραμέτρους αγωγιμότητα,  $Cl^-$  και  $NO_3^-$  και ως εκ τούτου χαρακτηρίστηκε ως «παράγοντας αλατότητας», ενώ η δεύτερη συνιστώσα συσχετίστηκε με τις παραμέτρους  $NH_4^+$  και  $NO_2^-$  και χαρακτηρίστηκε ως «παράγοντας παρουσίας λυμάτων». Η παράμετρος Al δεν φάνηκε να συσχετίζεται με καμία από τις δύο πρώτες συνιστώσες κάτι που ήταν αναμενόμενο λόγω του ότι τις περισσότερες φορές βρίσκεται να έχει μηδενική τιμή.

Πίνακας 2: Φορτίσεις βασικών συνιστωσών

Μεταβλητή	Συνιστώσα 1	Συνιστώσα 2
$Cl^-$	<b>0,831</b>	-0,176
Αγωγιμότητα	<b>0,953</b>	-0,059
$NH_4^+$	0,357	<b>0,755</b>
$NO_2^-$	0,351	<b>0,641</b>
$NO_3^-$	<b>0,620</b>	-0,460
Al	-0,065	0,103

Από την απεικόνιση των δειγμάτων στο επίπεδο των δύο πρώτων συνιστωσών (Σχήμα 1) φάνηκε ότι οι κατηγορίες των δειγμάτων δεν ήταν σαφώς διαχωρισμένες και θα ήταν εύλογο να υπάρξουν προβλήματα στις μεθόδους της DA και των νευρωνικών δικτύων.

#### 3.2 Γραμμική διαχωριστική ανάλυση (LDA)

Κατά την εφαρμογή της LDA (Pella and Masuda 2005; Lavine and Rayens 2009) υπολογίστηκαν δύο ρίζες οι οποίες βρέθηκαν στατιστικά σημαντικές σε επίπεδο  $p=0,05$  (Πίνακας 3). Το ποσοστό της διακύμανσης που παρέμεινε ανεξήγητο από την πρώτη και τη δεύτερη ρίζα ήταν 61% και 89% αντίστοιχα.

Πίνακας 3: Κανονικές ρίζες και σημαντικότητά τους

Ρίζα	Ιδιοτιμή	Κανονικό $R_C$	Wilks $\Lambda$	Δοκιμή $\chi^2$	B.E.	p-τιμή*
1	0,459	0,561	0,610	49,209	12	0,000
2	0,124	0,332	0,890	11,646	5	0,040

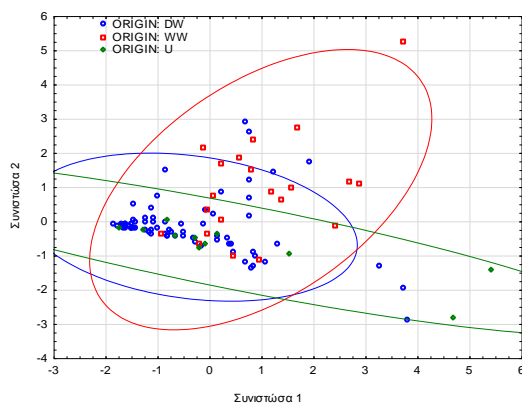
\* Η μηδενική υπόθεση  $H_0$  αφορά την ισότητα των μέσων τιμών των υπολογιζόμενων τιμών των ομάδων (DW, WW, U).

Η εξέταση της δομής των παραγόντων (Πίνακας 4) έδειξε σημαντική συσχέτιση της πρώτης ρίζας με τις παραμέτρους  $NH_4^+$  και  $NO_2^-$  («παράγοντας παρουσίας λυμάτων») ενώ η δεύτερη ρίζα συσχετίστηκε με τις παραμέτρους  $Cl^-$  και αγωγιμότητα («παράγοντας αλατότητας»).

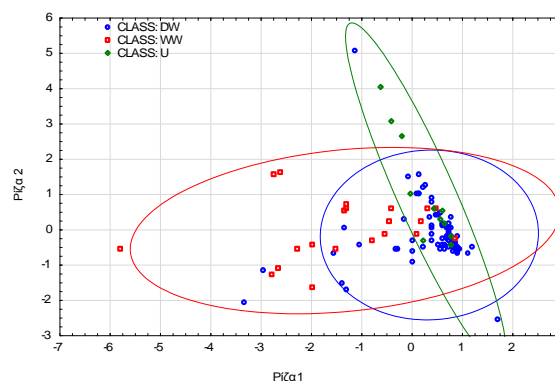
Πίνακας 4: Συντελεστές δομής παραγόντων

Παράμετρος	Ρίζα 1	Ρίζα 2
$Cl^-$	-0,347	<b>0,864</b>
Αγωγιμότητα	-0,400	<b>0,537</b>
$NH_4^+$	<b>-0,814</b>	-0,332
$NO_2^-$	<b>-0,570</b>	-0,119
$NO_3^-$	-0,040	0,299
Al	0,119	-0,244

Η απεικόνιση των δειγμάτων στο επίπεδο των δύο ριζών (Σχήμα 2) έδειξε ότι τα δείγματα δεν εμφανίζονται σαφώς διαχωρισμένα. Η πρώτη ρίζα («παράγοντας παρουσίας λυμάτων») επιτυγχάνει το διαχωρισμό μεταξύ DW (ύδρευση) και WW (αποχέτευση) ενώ η δεύτερη ρίζα («παράγοντας αλατότητας») συνεισφέρει στο διαχωρισμό των U (αγνώστου προέλευσης) από τα DW-WW.



Σχήμα 1: Απεικόνιση δειγμάτων στο επίπεδο των δύο πρώτων βασικών συνιστωσών.



Σχήμα 2: Απεικόνιση δειγμάτων στο επίπεδο των δύο κανονικών ριζών.

Η συνάρτηση ταξινόμησης για την εκτίμηση της ομάδας στην οποία ανήκει ένα δείγμα περιγράφεται από την γενική εξίσωση (1) όπου  $b_0 \dots b_n$  είναι οι συντελεστές που προκύπτουν από το μοντέλο και  $x_1 \dots x_n$  είναι οι τιμές των μεταβλητών που χρησιμοποιούνται στο μοντέλο.

$$F_{CLASS} = b_0 + b_1x_1 + \dots + b_nx_n \quad (1)$$

Οι συντελεστές  $b$  της εξίσωσης για κάθε ομάδα (Πίνακας 5) χρησιμοποιούνται για την ταξινόμηση των δειγμάτων. Τα δεδομένα κάθε δείγματος εισάγονται στις τρεις εξισώσεις και η εξίσωση που δίνει την μεγαλύτερη τιμή ορίζει την ομάδα στην οποία ανήκει το δείγμα.

Πίνακας 5: Συντελεστές εξισώσεων ταξινόμησης

Συντελεστής	DW	WW	U
$b_o$	-1,702	-5,179	-4,816
$b_{CHLORIDE}$	-0,014	0,003	0,008
$b_{CONDUCTIVITY}$	0,004	0,003	0,003
$b_{AMMONIA}$	-0,045	0,385	-0,178
$b_{NITRITE}$	-0,023	0,119	-0,019
$b_{NITRATE}$	-0,004	-0,003	-0,004
$b_{ALUMINIUM}$	0,006	0,001	0,000

Το μέχρι τώρα ποσοστό επιτυχούς ταξινόμησης που προκύπτει από την ερμηνεία των αποτελεσμάτων των αναλύσεων του εργαστηρίου με μεσολάβηση του ανθρώπινου παράγοντα ανέρχεται συνολικά στο 53% (Πίνακας 6).

Πίνακας 6: Πίνακας ταξινόμησης ομάδας εκπαίδευσης βάσει ερμηνείας εργαστηριακών αποτελεσμάτων (ανθρώπινος παράγοντας)

Πραγματική προέλευση		Εκτιμώμενη προέλευση		
		DW	WW	U
Ομάδα εκπαίδευσης	DW	42,5%	17,8%	39,7%
	WW	0,0%	76,2%	23,8%
	U	18,2%	0,0%	81,8%

Συνολικό ποσοστό επιτυχίας: 53,4%

Τα επί μέρους ποσοστά επιτυχίας δείχνουν ότι νερά τα οποία προέρχονται από το δίκτυο ύδρευσης ταξινομούνται ορθά κατά περίπου 40% ενώ ένα άλλο 40% χαρακτηρίζονται ως αγνώστου προελεύσεως. 20% περίπου των δειγμάτων ύδρευσης ταξινομούνται εσφαλμένα ως νερά δικτύου αποχέτευσης. Στην περίπτωση νερών που προέρχονται από το δίκτυο αποχέτευσης 75% περίπου των δειγμάτων ταξινομούνται ορθά, ενώ τα υπόλοιπα ταξινομούνται εσφαλμένα ως νερά αγνώστου προελεύσεως. Οι εσφαλμένες ταξινομήσεις των δειγμάτων που είναι στην ευθύνη της ΕΥΔΑΠ ΑΕ θα μπορούσαν να έχουν ως αποτέλεσμα την έλλειψη δράσης του αντίστοιχου Τομέα εφόσον δεν κατατεθεί νέο παράπονο από τον καταναλωτή.

Ο πίνακας ταξινόμησης (Πίνακας 7), όπως προέκυψε από την εφαρμογή του μοντέλου της LDA, έδειξε 78% συνολική επιτυχή επαναταξινόμηση της ομάδας εκπαίδευσης και 74% συνολική επιτυχή ταξινόμηση της ομάδας ελέγχου. Η επαλήθευση έγινε με τη μέθοδο jackknife (Lee, Jogesh Babu et al. 2012; Zhang, Wan et al. 2013).

Πίνακας 7: Πίνακας ταξινόμησης των ομάδων εκπαίδευσης και ελέγχου

Πραγματική προέλευση		Εκτιμώμενη προέλευση		
		DW	WW	U
Ομάδα εκπαίδευσης	DW	91,8%	6,8%	1,4%
	WW	42,9%	57,1%	0,0%
	U	72,7%	0,0%	27,3%
Ομάδα ελέγχου	DW	90,4%	8,2%	1,4%
	WW	52,4%	47,6%	0,0%
	U	81,8%	0,0%	18,2%

Συνολικό ποσοστό επιτυχίας για την ομάδα εκπαίδευσης: 78,1%

Συνολικό ποσοστό επιτυχίας για την ομάδα ελέγχου: 74,3%

Από τα αποτελέσματα της επαλήθευσης, τα οποία είναι και τα σημαντικότερα φάνηκε ότι το μοντέλο ταξινόμησε επιτυχώς τα δείγματα ύδρευσης (90%) ενώ είχε σχεδόν 50% επιτυχία στην περίπτωση των δειγμάτων αποχέτευσης. Αν και αυτό είναι σημαντικό μειονέκτημα του μοντέλου πρέπει να αναφερθεί ότι δείγματα τα οποία προέρχονται από το δίκτυο αποχέτευσης και δεν ταξινομούνται επιτυχώς, εναλλακτικά, αναγνωρίζονται ως δείγματα ύδρευσης και συνεπώς, αν και εσφαλμένη, μία τέτοια εργαστηριακή γνωμάτευση θα είχε και πάλι σαν αποτέλεσμα τη δρομολόγηση ενεργειών επισκευής από τον αντίστοιχο Τομέα οπότε και θα ήταν πιθανότερος ο εντοπισμός του πραγματικού προβλήματος.

### 3.3 Νευρωνικά δίκτυα (ANNs)

Κατά την εφαρμογή των ANNs, δοκιμάστηκαν μοντέλα MLPs (η δημοφιλέστερη ίσως τεχνική των ANNs) και απλά γραμμικά μοντέλα (Zupan and Gasteiger 1993; Farmaki, Thomaidis et al. 2010). Ο πίνακας αξιολόγησης (Πίνακας 8) των μοντέλων καταγράφει τα ποσοστά επιτυχίας των μοντέλων για τις ομάδες εκπαίδευσης, ελέγχου και επικύρωσης, καθώς και τις χρησιμοποιούμενες μεταβλητές με την σειρά κρισιμότητας αυτών.

Πίνακας 8: Πίνακας αξιολόγησης μοντέλων ANNs

Μοντέλο ANN	Ομάδα εκπαίδευσης	Ομάδα ελέγχου	Ομάδα επικύρωσης	Αριθμός μεταβλητών: σειρά κρισιμότητας
MLP 6-4-3	85,3%	86,7%	80,0%	6: Cl <sup>-</sup> > NH <sub>4</sub> <sup>+</sup> > NO <sub>2</sub> <sup>-</sup> > NO <sub>3</sub> <sup>-</sup> >Cond.>Al
MLP 6-11-3	85,3%	86,7%	73,3%	6: Cl <sup>-</sup> > NH <sub>4</sub> <sup>+</sup> > Cond.> NO <sub>2</sub> <sup>-</sup> > Al> NO <sub>3</sub> <sup>-</sup>
MLP 5-12-3	84,9%	76,9%	69,2%	5: NH <sub>4</sub> <sup>+</sup> >Cl <sup>-</sup> > Al> NO <sub>3</sub> <sup>-</sup> > NO <sub>2</sub> <sup>-</sup>
MLP 4-12-3	73,6%	73,1%	80,8%	4: NH <sub>4</sub> <sup>+</sup> >Cl <sup>-</sup> >Al> NO <sub>2</sub> <sup>-</sup>
MLP 5-8-3	83,0%	76,9%	69,2%	5: NH <sub>4</sub> <sup>+</sup> >Cl <sup>-</sup> > Al> NO <sub>3</sub> <sup>-</sup> > NO <sub>2</sub> <sup>-</sup>
Linear 2-3	75,5%	73,1%	69,2%	2: NH <sub>4</sub> <sup>+</sup> >Cl <sup>-</sup>
MLP 5-15-3	77,4%	69,2%	80,8%	5: NH <sub>4</sub> <sup>+</sup> >Cond.>Al>Cl <sup>-</sup> > NO <sub>2</sub> <sup>-</sup>
Linear 3-3	71,7%	73,1%	76,9%	5: NO <sub>2</sub> <sup>-</sup> >Cond.> NH <sub>4</sub> <sup>+</sup>
MLP 6-9-3	79,2%	80,8%	84,6%	6: NH <sub>4</sub> <sup>+</sup> >Cond.> NO <sub>3</sub> <sup>-</sup> >Al>Cl <sup>-</sup> > NO <sub>2</sub> <sup>-</sup>
Linear 4-3	73,6%	80,8%	76,9%	4: NH <sub>4</sub> <sup>+</sup> >Cl <sup>-</sup> > NO <sub>2</sub> <sup>-</sup> > NO <sub>3</sub> <sup>-</sup>

Για τα καλύτερα από αυτά (Πίνακας 9, Πίνακας 10) παρατίθενται αναλυτικά τα αποτελέσματα για τις τρεις ομάδες: DW, WW και U.

E. Γ. Φαρμάκη, Ν. Κ. Θανασούλιας, Α. Χ. Καγιάρα, Σ. Α. Σάμιος, Δ. Θ. Τσοροβά, Φ. Τ. Μισκάκη, Ε. Σ. Λύτρας, ΕΥΔΑΠ ΑΕ  
Διερεύνηση της Προέλευσης Υπογείων Διαρροών με Γραμμική Διαχωριστική Ανάλυση και Τεχνητά Νευρωνικά Δίκτυα

Πίνακας 9: Συνολικός πίνακας ταξινόμησης των τριών ομάδων για το μοντέλο MLP 6-4-3

Πραγματική προέλευση	Εκτιμώμενη προέλευση		
	DW	WW	U
DW	97,3%	2,7%	0,0%
WW	28,6%	71,4%	0,0%
U	72,7%	0,0%	27,3%

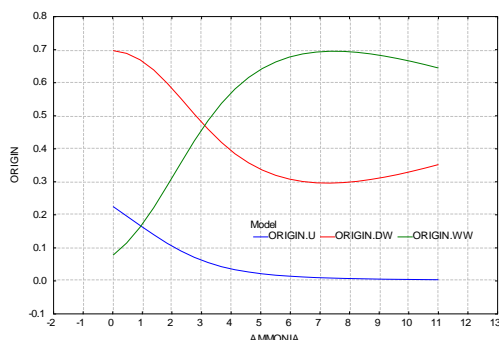
Συνολικό ποσοστό επιτυχίας: 73,3%

Πίνακας 10: Συνολικός πίνακας ταξινόμησης των τριών ομάδων για το μοντέλο Linear 2-3

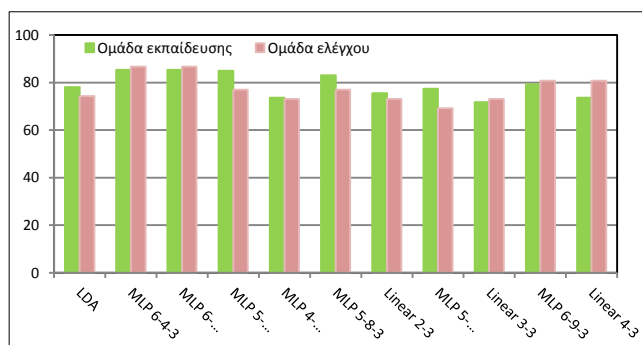
Πραγματική προέλευση	Εκτιμώμενη προέλευση		
	DW	WW	U
DW	94,5%	4,1%	1,4%
WW	66,7%	33,3%	0,0%
U	90,9%	0,0%	9,1%

Συνολικό ποσοστό επιτυχίας: 84,8%

Το σχήμα Σχήμα 3 απεικονίζει το διάγραμμα απόκρισης για το μοντέλο MLP 6-4-3 και την παράμετρο  $\text{NH}_4^+$  που αποδεικνύεται να είναι μια από τις κρισιμότερες. Από τα αποτελέσματα της ταξινόμησης των μοντέλων ANNs, φάνηκε ότι το καλύτερο από αυτά (MLP 6-4-3), ταξινόμησε επιτυχώς τα δείγματα ύδρευσης (97%), αλλά και τα προερχόμενα από την αποχέτευση (71%). Είναι φανερό ότι τα άγνωστα δείγματα (U) δεν αποτελούν κάποια αδυναμία των μοντέλων, εφόσον στην πραγματικότητα περικλείουν μια μεγάλη ετερογενή κατηγορία δειγμάτων (ύδατα από τον υδροφόρο ορίζοντα, τα όμβρια ύδατα κλπ). Γενικά, είναι συχνό φαινόμενο να παρατηρούνται σε υπόγεια οικιών εισροές υδάτων που προέρχονται από τον υπόγειο υδροφόρο, ως συνέπεια της κατά τόπους άναρχης δόμησης της πόλης των Αθηνών που έχει μεταβάλλει τις φυσικές εξόδους των υπόγειων υδάτων. Επομένως κανένα μοντέλο δεν θα μπορούσε να επιτύχει συνολικά τον απόλυτο χαρακτηρισμό μιας τόσο ανομοιογενούς κατηγορίας. Επίσης, είναι φανερό ότι στην κατηγορία αυτή περιλαμβάνονται και δείγματα προερχόμενα από το δίκτυο ύδρευσης (DW) ή το δίκτυο αποχέτευσης (WW) που δεν μπόρεσαν να ταυτοποιηθούν από τον αντίστοιχο Τομέα. Συγκριτικά τα μοντέλα ANNs δίνουν καλύτερα αποτελέσματα από την LDA (Σχήμα 4) λόγω του ότι διαθέτουν μεγάλη ευελιξία (ποικιλία αρχιτεκτονικών και βαθμών ελευθερίας στις παραμέτρους που τα χαρακτηρίζουν).



Σχήμα 3: Διάγραμμα απόκρισης για το μοντέλο MLP 6-4-3 και την παράμετρο  $\text{NH}_4^+$ .



Σχήμα 4: Συγκριτικό διάγραμμα μοντέλων LDA και ANNs.

#### 4. Συμπεράσματα



Από τη στατιστική μελέτη των δειγμάτων φάνηκε ότι τα δείγματα περιγράφονται από δύο κύριους παράγοντες οι οποίοι, με τη βοήθεια της ανάλυσης βασικών συνιστωσών (PCA), ονομάστηκαν «παράγοντας αλατότητας» και «παράγοντας παρουσίας λυμάτων».

Η γραμμική διαχωριστική ανάλυση (LDA) έδειξε ότι είναι δυνατόν να επιτευχθεί κατά 74% ορθή ταξινόμηση των δειγμάτων με τη βοήθεια δύο κανονικών ριζών οι οποίες εκφράζουν τον «παράγοντα παρουσίας λυμάτων» και τον «παράγοντα αλατότητας». Το ποσοστό αυτό αντιστοιχεί σε αύξηση της ορθής ταξινόμησης κατά 21% σε σχέση με τον τρόπο που ταξινομούν οι αναλυτές τα δείγματα βασιζόμενοι στο φυσικοχημικό τους προφίλ.

Τα νευρωνικά δίκτυα (ANNs) πέτυχαν 87% ορθή ταξινόμηση των δειγμάτων ελέγχου. Τα μοντέλα ANN δίνουν καλύτερα αποτελέσματα από την LDA λόγω της μεγάλης ευελιξίας τους.

Σε κάθε περίπτωση, δεν θα πρέπει να παραβλεφθεί πως η διερεύνηση της προέλευσής των υδάτων στα υπόγεια των οικιών βάσει των ποιοτικών τους χαρακτηριστικών υπόκειται σε μια σειρά παραγόντων που είναι πολύ δύσκολο έως αδύνατο να περιγραφούν όπως η ταχύτητα ροής τους, το μήκος της διαδρομής τους έως το σημείο εξόδου τους, ο πιθανός χρόνος παραμονής τους σε στάσιμη κατάσταση (με συνέπεια να λαμβάνουν χώρα βιοχημικές αντιδράσεις που αλλοιώνουν τα αρχικά τους χαρακτηριστικά) και το ποσοστό εκχύλισης διαλυτών ουσιών κατά τη διαδρομή του νερού στο υπέδαφος. Συνεπώς, για οποιαδήποτε γνωμάτευση που εκδίδεται από το Χημικό Εργαστήριο σχετικά με την προέλευση ενός αγνώστου δείγματος, είτε αυτή βασίζεται στην εμπειρική μέθοδο ερμηνείας είτε προκύπτει από τα μοντέλα που αναπτύχθηκαν εδώ, θα πρέπει να γίνεται κατανοητό ότι αυτή διέπεται από ορισμένη αβεβαιότητα και τα αποτελέσματα θα πρέπει να ερμηνεύονται με προσοχή.

Γενικά, από τη χημειομετρική προσέγγιση που ακολουθήθηκε στην παρούσα εργασία, φάνηκε ότι τα αποτελέσματα είναι αισιόδοξα και μπορούν να εγερθούν πολύ μεγαλύτερα ποσοστά επιτυχούς ταξινόμησης των δειγμάτων με τη χρήση μιας ευρύτερης βάσης δεδομένων και τη μέτρηση επιπλέον παραμέτρων.

## 5. Βιβλιογραφία

- Ashton, C. H. and V. S. Hope (2001). "Environmental valuation and the economic level of leakage." Urban Water **3**(4): 261-270.
- Bermúdez, A. and L. M. García-García (2012). "Mathematical modeling in chemistry. Application to water quality problems." Applied Numerical Mathematics **62**(4): 305-327.
- Bleam, W. F. (2012). Chapter 5 - Water Chemistry. Soil and Environmental Chemistry. W. F. Bleam. Boston, Academic Press: 151-208.
- Britton, T. C., R. A. Stewart, et al. (2013). "Smart metering: enabler for rapid and effective post meter leakage identification and water loss management." Journal of Cleaner Production **54**(0): 166-176.
- Candelieri, A., D. Conti, et al. (2014). "Improving Analytics in Urban Water Management: A Spectral Clustering-based Approach for Leakage Localization." Procedia - Social and Behavioral Sciences **108**(0): 235-248.
- Delgado-Galván, X., R. Pérez-García, et al. (2010). "An analytic hierarchy process for assessing externalities in water leakage management." Mathematical and Computer Modelling **52**(7-8): 1194-1202.

- Esbensen, K. H. and P. Geladi (2009). 2.13 - Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice. Comprehensive Chemometrics. S. D. Brown, R. Tauler and B. Walczak. Oxford, Elsevier: 211-226.
- Farmaki, E. G., N. S. Thomaidis, et al. (2010). "Artificial Neural Networks in Water Analysis: Theory and Applications." International Journal of Environmental Analytical Chemistry **90**: 85-105.
- Lavine, B. K. and W. S. Rayens (2009). 3.16 - Statistical Discriminant Analysis. Comprehensive Chemometrics. S. D. Brown, R. Tauler and B. Walczak. Oxford, Elsevier: 517-540.
- Lee, H., G. Jogesh Babu, et al. (2012). "A jackknife type approach to statistical model selection." Journal of Statistical Planning and Inference **142**(1): 301-311.
- Li, W., W. Ling, et al. (2011). "Development of systems for detection, early warning, and control of pipeline leakage in drinking water distribution: A case study." Journal of Environmental Sciences **23**(11): 1816-1822.
- Morais, D. C. and A. T. de Almeida (2007). "Group decision-making for leakage management strategy of water network." Resources, Conservation and Recycling **52**(2): 441-459.
- Pella, J. and M. Masuda (2005). Chapter 25 - Classical Discriminant Analysis, Classification of Individuals, and Source Population Composition of Mixtures. Stock Identification Methods. S. X. Cadrin, K. D. Friedland and J. R. Waldman. Burlington, Academic Press: 517-552.
- Piaia, J. C. Z., M. Cheriaf, et al. (2013). "Measurements of water penetration and leakage in masonry wall: Experimental results and numerical simulation." Building and Environment **61**(0): 18-26.
- Syms, C. (2008). Principal Components Analysis. Encyclopedia of Ecology. S. E. Jørgensen and B. D. Fath. Oxford, Academic Press: 2940-2949.
- Zhang, X., A. T. K. Wan, et al. (2013). "Model averaging by jackknife criterion in models with dependent data." Journal of Econometrics **174**(2): 82-94.
- Zupan, J. and J. Gasteiger (1993). Neural Networks for Chemists: An Introduction. New York, Wiley-VCH.