



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΔΙΟΙΚΗΣΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΔΙΟΙΚΗΣΗΣ**

**Αξιολόγηση της Ποιότητας του Πόσιμου Νερού της Αθήνας,
με Στόχο την Ολοκληρωμένη Διαχείρισή του,
μέσω Στατιστικών Μεθόδων**

**Ελένη Σμέτη
Φυσικός – MSc Στατιστικής**

Διδακτορική διατριβή

**Επιβλέπων
Επίκ. Καθ. κ. Σπυρίδων Γκολφινόπουλος**

Ιούνιος 2010

*Στη μνήμη των γονιών μου
Μιλτιάδη και Αγορίτσας*

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ θερμά τον Επίκ. Καθηγητή κ. Σπυρίδωνα Γκολφινόπουλο για την ανάθεση της παρούσας Διδακτορικής Διατριβής καθώς και για την καθοδήγησή του, τις χρήσιμες συμβουλές του και τον χρόνο που αφιέρωσε για την συνεχή παρακολούθηση και έλεγχο της. Τον ευχαριστώ για την υποστήριξη του σε όλη τη διάρκεια εκπόνησης της διατριβής μου.

Ευχαριστώ θερμά τα μέλη της συμβουλευτικής επιτροπής Καθηγητή κ. Ιωάννη Γκιάλα και Αναπλ. Καθηγητή κ. Επαμεινώνδα Κυριακίδη για την παρακολούθηση της εργασίας μου και τις χρήσιμες παρατηρήσεις τους.

Είναι τιμή μου που συμμετείχαν στην εξεταστική επιτροπή ο Καθηγητής κ. Θεμιστοκλής Λέκκας και οι Επίκ. Καθηγητές κ.κ. Δημήτριος Καρλής, Ιωάννης Ντζούφρας και Αθανάσιος Βλεσσίδης. Ευχαριστώ θερμά όλα τα μέλη της εξεταστικής επιτροπής για τις ερωτήσεις και τις εύστοχες παρατηρήσεις τους κατά την διαδικασία εξέτασης της Διδακτορικής μου Διατριβής.

Ευχαριστώ ιδιαίτερα τον Επίκ. Καθηγητή κ. Ιωάννη Ντζούφρα για την εποικοδομητική μας συζήτηση πριν την εκπόνηση της διατριβής.

Εκφράζω τις ευχαριστίες μου προς στον Διευθυντή Ποιότητας της ΕΥΔΑΠ Χημικό κ. Φίλιππο Τζουμέρκα για την παροχή δεδομένων ποιότητας του νερού και τις χρήσιμες πληροφορίες του για θέματα ποιότητας και επεξεργασίας του νερού.

Ευχαριστώ την Υπηρεσία Ελέγχου Ποιότητας Ύδατος της ΕΥΔΑΠ για την παροχή δεδομένων ποιότητας του νερού. Ευχαριστώ τον Βιολόγο – Υγειονολόγο της ΕΥΔΑΠ κ. Δημήτριο Κορωνάκη για την παροχή δεδομένων τοξικότητας και τον Χημικό της ΕΥΔΑΠ Δρ. Νικόλαο Θανασούλια για τις συζητήσεις μας και την ανταλλαγή απόψεων.

Θα ήθελα τέλος να εκφράσω τις ευχαριστίες μου στον σύζυγό μου, Παναγιώτη Σπιθουράκη, και στα παιδιά μου Γιώργο και Μαριάννα για την κατανόησή τους για τις επιστημονικές μου ανησυχίες, την συμπαράσταση και την ενθάρρυνσή τους. Συγχρόνως, θα ήθελα να τους ζητήσω συγνώμη για όσα τους στέρησα κατά τη διάρκεια εκπόνησης της Διδακτορικής μου Διατριβής.

ΣΥΝΘΕΣΗ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ

- Σ. Γκολφινόπουλος Επίκουρος Καθηγητής Τμήματος Μηχανικών Οικονομίας και Διοίκησης Πανεπιστημίου Αιγαίου (Επιβλέπων)
- Ι. Γκιάλας Καθηγητής Τμήματος Μηχανικών Οικονομίας και Διοίκησης Πανεπιστημίου Αιγαίου (Μέλος Συμβουλευτικής Επιτροπής)
- Ε. Κυριακίδης Αναπληρωτής Καθηγητής Τμήματος Στατιστικής Οικονομικού Πανεπιστημίου Αθηνών (Μέλος Συμβουλευτικής Επιτροπής)
- Θ. Λέκκας Καθηγητής Τμήματος Περιβάλλοντος Πανεπιστημίου Αιγαίου
- Δ. Καρλής Επίκουρος Καθηγητής Τμήματος Στατιστικής Οικονομικού Πανεπιστημίου Αθηνών
- Ι. Ντζούφρας Επίκουρος Καθηγητής Τμήματος Στατιστικής Οικονομικού Πανεπιστημίου Αθηνών
- Α. Βλεσσίδης Επίκουρος Καθηγητής Τμήματος Χημείας Πανεπιστημίου Ιωαννίνων

ΠΕΡΙΛΗΨΗ

Σκοπός της διατριβής ήταν η ανάπτυξη μεθοδολογίας για την εκτίμηση της ποιότητας του νερού μέσω της ολοκληρωμένης στατιστικής ανάλυσης των χημικών, των μικροβιολογικών και των τοξικολογικών παραμέτρων του νερού, από την ακατέργαστη μορφή του, στις πηγές υδροληψίας, μέχρι και την επεξεργασμένη μορφή του, ως νερό προς ανθρώπινη κατανάλωση, που θα οδηγήσει στην ολοκληρωμένη αξιολόγηση και διαχείριση ενός συστήματος υδατικών πόρων.

Η πρωτοτυπία της παρούσας διατριβής έγκειται στα εξής:

- Στη διερεύνηση των ιδιαίτερων χαρακτηριστικών των παραμέτρων ποιότητας του νερού και στην κατάλληλη επιλογή καθώς και προσαρμογή των στατιστικών μεθόδων που χρησιμοποιήθηκαν, με σκοπό την εξασφάλιση της μέγιστης πληροφορίας.
- Στην ολοκληρωμένη στατιστική ανάλυση των παραμέτρων ποιότητας του νερού ενός περίπλοκου υδρευτικού συστήματος όπως αυτό των Αθηνών.
- Στην ανάπτυξη επικυρωμένων μοντέλων τα οποία προσφέρουν τη δυνατότητα ανάδρασης όταν αυτό κρίνεται απαραίτητο.

Διεξήχθη ενδεδειγμένη στατιστική ανάλυση των τιμών τοξικότητας (εκφρασμένης ως ποσοστό αναστολής της φωταύγειας των βακτηρίων *Vibrio fischeri*) στους ταμιευτήρες Μαραθώνα και Μόρνου καθώς στη λίμνη Υλίκη για τη χρονική περίοδο 2000-2006, με τη χρήση εξειδικευμένων μη-παραμετρικών στατιστικών μεθόδων. Οι τιμές της τοξικότητας στους τρεις ταμιευτήρες / λίμνες, ήταν αρκετά χαμηλές και συγκρίσιμες με την τοξικότητα που συνήθως παρατηρείται σε πόσιμα νερά. Στο διάστημα αυτό, δεν παρουσιάστηκαν στατιστικά σημαντικές ισχυρές ενδείξεις εποχικότητας. Αποκαλύφθηκε μικρού μεγέθους σειριακή συσχέτιση των μηνιαίων τιμών τοξικότητας σε κάθε ταμιευτήρα, η οποία και ελήφθη υπόψη κατά τη στατιστική ανάλυση, φάνηκε όμως ότι δεν επηρεάζει τα αποτελέσματα. Μέσω του ελέγχου τάσεων Mann-Kendall, διαπιστώθηκε η έλλειψη ανοδικής ή πτωτικής τάσης των τιμών της τοξικότητας το συγκεκριμένο χρονικό διάστημα. Ο έλεγχος για ύπαρξη πιθανών διαφορών των κεντρικών μέτρων θέσης της τοξικότητας μεταξύ των τριών ταμιευτήρων / λιμνών, κατέληξε σε μη στατιστική σημαντικότητα. Συνεπώς, το νερό

των τριών ταμιευτήρων μπορεί να χαρακτηριστεί ως χαμηλής τοξικότητας και σχετικά σταθερής τόσο όσον αφορά στην χρονική μεταβλητότητα της τοξικότητας όσο και στις μεταξύ τους διαφορές. Το γεγονός αυτό αποτελεί ένδειξη υψηλής ποιότητας του ανεπεξέργαστου νερού που χρησιμοποιείται για την παραγωγή πόσιμου. Επιπλέον, δεδομένου ότι η τοξικότητα των ταμιευτήρων/λίμνων μελετήθηκε για πρώτη φορά, τίθεται η βάση για περαιτέρω παρακολούθηση των πηγών υδροληψίας και σύγκρισης των τιμών τοξικότητας στο μέλλον.

Η μικροβιακή φόρτιση του νερού των ποταμών που εκβάλλουν στους τρεις ταμιευτήρες / λίμνες ελέγχθηκε μέσω των βακτηρίων – δεικτών μικροβιακής φόρτισης *enterococci*. Η επιλογή των βακτηρίων εντεροκόκκου (*enterococci*) ως δείκτη μικροβιακής φόρτισης έγινε βάσει της σταθερότητας των βακτηρίων, της πλήρους εργαστηριακής ανάπτυξης της μεθόδου ανάλυσής τους και της σημαντικότητας της παραμέτρου. Για την αποτίμηση της μικροβιακής φόρτισης σε βακτήρια *enterococci* των ποταμών/ρεμάτων που εκβάλλουν στους ταμιευτήρες Μαραθώνα και Μόρνου καθώς και τη λίμνη Υλίκη, χρησιμοποιήθηκαν οι μηνιαίοι προσδιορισμοί των βακτηρίων κατά τη διάρκεια μιας περιόδου δεκατριών ετών (1997 - 2009). Τα μηνιαία δεδομένα εντεροκόκκων υπεβλήθησαν, για πρώτη φορά, σε μια σειρά στατιστικών αναλύσεων με σκοπό τη διερεύνηση εποχικότητας, ύπαρξης τάσεων στα βακτήρια εντεροκόκκων και τον έλεγχο της ομοιογένειας των τάσεων. Σε δύο από τους ποταμούς που εκβάλλουν στον ταμιευτήρα Μόρνου, στους ποταμούς Κόκκινο και Άβορο, το ποσοστό των κάτω περικεκομμένων τιμών *enterococci* βρέθηκε να είναι εξαιρετικά μεγάλο. Επομένως μπορούν να χαρακτηριστούν ως χαμηλού μικροβιακού φορτίου όσον αφορά στα βακτήρια εντεροκόκκων. Αντίθετα, ο μοναδικός ποταμός (Βοιωτικός Κηφισός) που εκβάλλει στην λίμνη Υλίκη, παρουσίασε υψηλό ποσοστό (μεγαλύτερο από 50%) άνω περικεκομμένων τιμών. Συνεπώς δεν είναι εφικτή η αξιολόγηση της μικροβιακής φόρτισης του ποταμού ως προς τα βακτήρια εντεροκόκκων. Σε καμία από τις υπόλοιπες θέσεις δειγματοληψίας δεν εντοπίστηκε στατιστικά σημαντική ανοδική τάση για τους *εντεροκόκκους*.

Διερευνήθηκε εκτενώς η χρησιμότητα και η αποτίμηση των πολυμεταβλητών στατιστικών τεχνικών στην ερμηνεία της δομής των φυσικο-χημικών ποιοτικών παραμέτρων του νερού κάθε ταμιευτήρα/λίμνης και στην αποτίμηση των χρονικών διακυμάνσεων της ποιότητάς του. Εξετάστηκαν οι συσχετίσεις μεταξύ των

παραμέτρων και δημιουργήθηκε μοντέλο για την επεξήγησή τους. Αναπτύχθηκαν και επικυρώθηκαν μοντέλα ομαδοποίησης των δειγμάτων του νερού ανάλογα με την χρονική περίοδο δειγματοληψίας. Επιτεύχθηκε σημαντική μείωση των διαστάσεων του πολύπλοκου σετ των δεδομένων (μεγάλος αριθμός παραμέτρων και αριθμού δειγμάτων) και αναδείχθηκαν λίγες μόνο παράμετροι - δείκτες ως κύριες υπεύθυνες για τις διακυμάνσεις στην ποιότητα του νερού στη διάρκεια του χρόνου. Τα μοντέλα που κατασκευάστηκαν, λαμβάνουν υπόψη τους την πολυπλοκότητα της πραγματικότητας και προβλέπουν τη μελλοντική συμπεριφορά του συστήματος των φυσικοχημικών παραμέτρων των ταμιευτήρων/λιμνών. Τα υψηλά ποσοστά ορθής ταξινόμησης των μοντέλων που αναπτύχθηκαν για κάθε ταμιευτήρα υποδεικνύουν την καλή ακρίβειά τους. Οι πολυμεταβλητές στατιστικές τεχνικές αναδείχθηκαν σε ισχυρά εργαλεία για την επεξήγηση πολύπλοκων σετ δεδομένων της ποιότητας του νερού και την κατανόηση των χρονικών διακυμάνσεων της. Επομένως οι τεχνικές αυτές μπορούν να προσφέρουν χρήσιμες πληροφορίες για την προσαρμογή της επεξεργασίας του ανεπεξέργαστου νερού βάσει των ιδιαίτερων χαρακτηριστικών ποιότητάς του.

Διερευνήθηκε, για πρώτη φορά, η χρησιμότητα των πολυμεταβλητών στατιστικών μεθόδων στη διάκριση του επεξεργασμένου νερού διαφορετικών Μονάδων Επεξεργασίας Νερού (MEN) καθώς και στην πρόβλεψη της προέλευσης του νερού του δικτύου ύδρευσης. Η εφαρμογή ανάλυσης κύριων συνιστωσών/παραγοντικής ανάλυσης στα δεδομένα των προσδιορισμών εννέα ποιοτικών παραμέτρων του νερού από τις δεξαμενές επεξεργασμένου νερού των MEN Αχαρνών, Γαλατσίου και Πολυδενδρίου επέτρεψε σημαντική μείωση των διαστάσεων του σετ των δεδομένων σε δύο παράγοντες οι οποίοι ερμηνεύουν το 75,6% της συνολικής μεταβλητότητας των δεδομένων. Μέσω της διαχωριστικής ανάλυσης, παρατηρήθηκαν σημαντικές διαφορές στα χαρακτηριστικά του επεξεργασμένου νερού λόγω των διαφορετικών πηγών προέλευσης του επιφανειακού νερού (Μόρνος ή μεικτό νερό από Μαραθώνα – Μόρνο) καθώς και διαφορές που αποδίδονται στην τεχνική κατασκευή και τις λειτουργικές διαδικασίες κάθε μονάδας επεξεργασίας νερού (MEN). Το μοντέλο της βηματικής διαχωριστικής ανάλυσης απέφερε κατά 97,8% ορθή ταξινόμηση των δειγμάτων στις δεξαμενές επεξεργασμένου νερού των MEN. Το μοντέλο χρησιμοποιήθηκε στη συνέχεια για την εκτίμηση της ορθότητας ταξινόμησης των δειγμάτων από δεξαμενές του δικτύου ύδρευσης ως προς τη MEN από την οποία

προέρχονταν. Η πρόβλεψη της προέλευσης του νερού στις δεξαμενές του δικτύου ύδρευσης ήταν επίσης 97,8%. Το υψηλό ποσοστό ορθής ταξινόμησης των δειγμάτων νερού των δεξαμενών δικτύου καθώς και η εγγύτητα των ποσοστών ορθής ταξινόμησης των δειγμάτων νερού MEN (βάσει των οποίων κατασκευάστηκε το μοντέλο) και δεξαμενών δικτύου πιστοποιεί την υψηλή ακρίβειά του. Συνεπώς, μπορεί να χρησιμοποιηθεί ως μέσον περιγραφής των δικτύων διανομής και πρόβλεψης της προέλευσης του νερού. Η αξιοποίηση τέτοιου είδους στατιστικών εργαλείων θα μπορούσε να βοηθήσει επικουρικά στις παραδοσιακές υδραυλικές τεχνικές κατά την περιγραφή των δικτύων ύδρευσης, παρέχοντας χρήσιμες πληροφορίες στους διαχειριστές των δικτύων, με αποτέλεσμα την αύξηση του βαθμού ασφαλείας. Μικρότερο αλλά αρκετά ικανοποιητικό (88,3%) ήταν και το ποσοστό ορθής ταξινόμησης των δειγμάτων δικτύου ύδρευσης βάσει του εναλλακτικού προς τη διαχωριστική ανάλυση αλγορίθμου των δέντρων ταξινόμησης και παλινδρόμησης, τα οποία αποτελούν μία προσέγγιση πλησιέστερη προς την ανθρώπινη λογική.

Κατασκευάστηκαν, για πρώτη φορά, διαγράμματα ελέγχου της τοξικότητας του επεξεργασμένου νερού σε MEN. Η παρακολούθηση της τοξικότητας του πόσιμου νερού παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς αποτελεί έναν συνολικό δείκτη της ποιότητάς του. Οι μέθοδοι στατιστικού ελέγχου διεργασιών αποσκοπούν στον περιορισμό της μεταβλητότητας των παραμέτρων με επακόλουθο τη βελτιστοποίηση της ποιότητας (υψηλότερη και σταθερότερη). Για τον στατιστικό έλεγχο διεργασιών όσον αφορά στις ημερήσιες τιμές της τοξικότητας του επεξεργασμένου νερού τριών δεξαμενών MEN, διερευνήθηκαν οι τυπικές τεχνικές στατιστικού ελέγχου διεργασιών και οι συνδυασμοί τους με περισσότερο περίπλοκες στατιστικές μεθόδους. Η προσέγγιση των Alwan και Roberts που βασίζεται στη χρήση υποδειγμάτων χρονολογικών σειρών, όχι μόνο απέδωσε ρεαλιστικά διαγράμματα ελέγχου της τοξικότητας, τα οποία μπορούν να αποτελέσουν χρήσιμα εργαλεία για τη βελτίωση της ποιότητας του πόσιμου νερού, αλλά επιπλέον έδωσε και τη δυνατότητα για βραχυχρόνιες προβλέψεις των τιμών της τοξικότητας. Η δυνατότητα πρόβλεψης της συμπεριφοράς του συστήματος για την αμέσως επόμενη χρονική περίοδο, χρησιμεύει στην επαναπροσαρμογή της διεργασίας εάν αυτό κρίνεται απαραίτητο και είναι εφικτό.

Στην παρούσα διατριβή, διερευνήθηκε η αποτελεσματικότητα, η επάρκεια και οι αναγκαίες προσαρμογές, λόγω των ιδιαίτερων χαρακτηριστικών των ποιοτικών παραμέτρων του νερού, ενός μεγάλου αριθμού σύνθετων στατιστικών μεθόδων. Η κατανόηση των μηχανισμών που επηρεάζουν την ποιότητα του νερού από την ακατέργαστη πρωταρχική του μορφή έως την τελική του που προορίζεται για ανθρώπινη κατανάλωση, αποτέλεσε τον κύριο στόχο της έρευνας. Η μεθοδολογία που ακολουθήθηκε μπορεί να αποτελέσει έναν οδηγό προς άντληση πληροφοριών για την ποιότητα του νερού και σε άλλα συστήματα υδατικών πόρων και ύδρευσης.

ABSTRACT

During the recent decades there has been constantly increasing concern for the quality of surface water intended for human consumption as well as for drinking water. Special bodies have been founded to ensure that surface water meets high standards and the procedures used for the treatment of raw water result in a product that complies with legislation and safety rules. The statistical analyses techniques of the water quality data can provide important and useful information for the management and quality control of the water. However, due to the inherent characteristics of the water quality data, sophisticated statistical techniques for their analysis may be required. The aim of this thesis was to explore statistical methods for the integrated analysis of the water quality parameters at all stages (raw water – finished water). In this thesis, a great variety of univariate and multivariate, parametric and non-parametric, statistical methods were extensively explored and then applied to the water quality data of the complicated water supply system of Athens with a view to gain insight knowledge of the water quality behavior and the system's integrated management.

The originality of this thesis is based on:

- The exploration of the specific characteristics of the water quality parameters and the appropriate selection and adjustment of the statistical methods that have been used in order to gain the ultimate information.
- The integrated statistical analysis of the water quality parameters of a complicated system such as the Athens water supply system.
- The development of validated prediction models that offer the ability to feedback when it is needed.

The data that were obtained from the monthly toxicological analysis of the raw water from the reservoirs/lakes (monitoring time period: 2000-2006) were subjected to exhaustive statistical analysis by the usage of specialized non-parametric statistical methods. A small amount of autocorrelation was observed for each time-series implying that corrective remedies should be used in the statistical analyses. The toxicity rates from the three reservoirs/lakes were low enough and comparable with the toxicity that is usually observed in potable waters. The study of seasonality (based on Kruskal-Wallis test) of the toxicity data for each reservoir/lake resulted in no

statistical significance. Trend analysis (based on Mann-Kendall test) resulted in no statistical significant upward or downward trends. Moreover, there was not observed statistical significant difference of the central tendency measures for the toxicity data between the three reservoirs/lakes.

The microbial load of the rivers, that flow in Marathon and Mornos reservoirs as well as in Yliki lake, was investigated by means of seasonal and local variability as well as trend analysis tests. The available data were raw water samples from six different sampling sites, collected on a monthly basis, over a 13-year period (1997-2009), and analyzed for the bacteria indicators of *enterococci*. Trend analysis to the monthly quality data for each one of the sampling sites was applied to investigate for possible bacteria trends and the homogeneity of trends. The nonparametric Seasonal Kendall test and the Van Belle and Hughes test were used for the trend analysis and the homogeneity of trends respectively. Two of the rivers that flow into Mornos reservoir were characterized by exceptionally small microbial load with regard to the *enterococci* bacteria and they were excluded from the analysis of trends. On the contrary, the only river (Viotikos Kifisos) that flow into the lake Yliki, was excluded from the analysis of trends due to high amount (more than 50%) of censored data at the highest reporting limit. Consequently is not feasible the evaluation of microbial load of this river with regard to *enterococci*. No trends for the aforementioned parameter were found in the rest of the sampling sites.

Multivariate statistical techniques, such as principal components (PCA) / factor analysis (FA), cluster analysis (CA), discriminant analysis (DA) and classification trees (CART) were applied to the surface water quality data of the Marathon reservoir, Yliki lake and Mornos reservoir to explain the data structure and to assess temporal variations in the water quality. Water samples were analyzed for 16 physico-chemical parameters over a 4 or 5-year period (2003-2008), on a monthly basis, in order to ensure the surface water quality. The results were subjected to a PCA/FA and 3, 4 and 6 latent factors were extracted for Marathon, Yliki and Mornos with 75.6%, 80.4% and 80.4% of the total variance, in each reservoir/lake, being explained respectively. Cluster analysis resulted in two major clusters for each reservoir/lake. Discriminant analysis as well as CART were used for determining which variables discriminate between the different seasons originated by CA. Stepwise-mode DA and

CART allowed a reduction in the dimensionality of the complex dataset, delineating a few indicator parameters responsible for the variations in temporal water quality. The post-hoc reclassification of each dataset as well as the cross-validation was very satisfactory for both DA and CART (over 90%). During 2003-2007, the Marathon reservoir was reinforced by water transferred only from the Mornos reservoir. A stepwise -mode DA showed that the discrimination of samples from the Mornos and the Marathon reservoirs was based mainly on Mg^{2+} and Cl^- . The Marathon reservoir presents quite higher values of both parameters. This study illustrates that multivariate statistical techniques are powerful tools for interpreting complex water quality datasets and for understanding temporal variations. Therefore, these techniques can provide useful information for water quality management. The insight knowledge of the specific characteristics of the surface water sources is necessary for the follow-up of their quality characteristics. Furthermore, this can allow the establishment of protocols to ensure that the raw water reaching all the water treatment plants is of similar quality or to plan a more effective water treatment depending on the particular raw water characteristics.

Throughout the year 2007, eighty-nine treated water samples from three water treatment plants (WTPs) of the Athens Water Supply and Sewerage Company (EYDAP S.A.) and 180 samples from network tanks (NWTs) were analyzed for 9 physico-chemical parameters. The results regarding the WTPs were subjected to a PCA with 75% of the total variance being explained. Significant differences in the treated water characteristics due to different sources of surface water and differences in the water treatment facilities and operational procedures were observed. These differences were further utilized for the development of discriminative models for the treated water produced at three water treatment plants. Besides this, the predictive models were evaluated for the discrimination of network water samples according to their theoretically expected origin. A stepwise linear discriminant analysis (LDA) model constructed from the 89 treated water samples was used to predict class membership of the samples from the NWTs with a view to estimating the propagation of a possible water quality deterioration originating from the WTPs. The model utilized Cl^- , Al and EC and yielded a 96% correct classification of the training dataset, whereas the cross-validation yielded a 94% correct classification. Network tank samples were 95% correctly classified with regard to their theoretically expected

origin. Discriminant analysis was shown to be robust to violations of its assumptions. However, the application of appropriate remedies improved its discriminating power. The stepwise discriminant analysis based on separate covariance matrices of the canonical discriminant functions yielded a 98% correct classification of both the training dataset and the network tank samples. The classification and regression trees resulted in a more human-like approach as they deal with one parameter at a time. The CART algorithm showed that the main parameters used in the discrimination of the WTP samples were EC and Al. The post-hoc classification of the training dataset was 99%, whereas 88% of NWT samples were correctly classified. Discriminant analysis and classification and regression trees were shown to be powerful multivariate techniques for the discrimination of treated water samples originating from different water treatment plants. Treated water samples from the Galatsi water treatment plant were easily discriminated because of their chemical characteristics summarized under the term salinity. The discrimination of samples from the Acharnai and Polydendri water treatment plants, although less clear, was based on procedural features (addition of coagulant). The prediction of the water origin in network tanks was very satisfactory (over 88% for all algorithms). Pattern recognition techniques can be used in the discrimination of treated water samples with regard to their origin and therefore in the establishment of safety protocols in the case of an emergency, where a possible deterioration of water quality in one of the plants could degrade the water quality of the related water supply network sections.

Finally, control charts were constructed for the data that were obtained from the daily toxicological analysis of finished water (that is the water which has already been treated for human consumption) from the treated-water tanks of the Water Supply & Sewerage Corporation of Athens. The basic idea of the control charts is to test the hypothesis that there are only common causes of variability versus the alternative that there are special causes. The typical SPC control charts were ineffective, because of the existence of autocorrelation in the toxicity data for the three treated-water tanks. The serious amount of autocorrelation that was present in the data dramatically reduced the performance of the charts, giving a large number of false alarms when σ was estimated via the average moving range \overline{MR} and covering real alarms when σ was estimated via the sample standard deviation s . The problem of the autocorrelation was overcome by the usage of a more sophisticated method based on time – series

ARIMA models. The more sophisticated Alwan and Roberts (1988) approach eliminated the autocorrelation of the data. The special cause charts (SCC), which are Shewhart charts of individual observations applied to the residuals which were obtained after fitting the processes with an ARIMA model, gave 3 out-of-control clues the one treated-water tank and none for the other two. For fast localization of small shifts, EWMA and CUSUM charts applied to the residuals are more suitable. In any case of the three examined treated-water tanks, the appropriate models to explain the toxicity procedure were ARIMA(0,1,1). The time-series based method captures the dynamic structure of the data and gives reasonable control charts that can be useful tools for the improvement of the water quality.

ΠΕΡΙΕΧΟΜΕΝΑ

ΣΥΝΘΕΣΗ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ	vii
ΠΕΡΙΛΗΨΗ	ix
ABSTRACT	xv
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	xxix
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	xxxiii
ΟΝΟΜΑΤΟΛΟΓΙΑ	xxxv
ΣΥΜΒΟΛΑ ΣΤΑΤΙΣΤΙΚΗΣ	xxxv
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ ΣΤΑΤΙΣΤΙΚΗΣ	xxxvii
ΠΡΟΛΟΓΟΣ	1
ΚΕΦΑΛΑΙΟ 1: ΝΕΡΟ ΑΝΘΡΩΠΙΝΗΣ ΚΑΤΑΝΑΛΩΣΗΣ	7
1.1. ΕΙΣΑΓΩΓΗ	7
1.2. ΥΔΑΤΙΚΟΙ ΠΟΡΟΙ - ΡΥΠΑΝΣΗ	7
1.3. ΝΟΜΟΘΕΤΙΚΟ ΠΛΑΙΣΙΟ	10
1.4. ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥ ΝΕΡΟΥ	11
ΚΕΦΑΛΑΙΟ 2: ΤΟ ΥΔΡΕΥΤΙΚΟ ΣΥΣΤΗΜΑ ΤΗΣ ΑΘΗΝΑΣ	15
2.1. Η ΥΔΡΕΥΣΗ ΤΗΣ ΑΘΗΝΑΣ ΤΗΝ ΣΥΓΧΡΟΝΗ ΕΠΟΧΗ	15
2.2. ΤΑΜΙΕΥΤΗΡΕΣ – ΛΙΜΝΕΣ	18
2.2.1. Ταμειυτήρας Μαραθώνα	18
2.2.2. Λίμνη Υλίκη	20
2.2.3. Ταμειυτήρας Μόρνου	22
2.2.4. Ταμειυτήρας Εύηνου	24
2.3. ΓΕΩΤΡΗΣΕΙΣ	26
2.4. ΜΕΘΟΔΟΛΟΓΙΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥ ΝΕΡΟΥ	27
2.5. ΜΟΝΑΔΕΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΝΕΡΟΥ	29

2.6.	ΑΝΑΛΥΤΙΚΟΣ ΕΛΕΓΧΟΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΝΕΡΟΥ	30
ΚΕΦΑΛΑΙΟ 3: ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΔΕΔΟΜΕΝΩΝ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΝΕΡΟΥ – ΠΡΟΒΛΗΜΑΤΑ ΚΑΙ ΠΡΟΣΕΓΓΙΣΕΙΣ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ		35
3.1.	ΕΙΣΑΓΩΓΗ	35
3.2.	ΑΚΡΑΙΕΣ ΤΙΜΕΣ	35
3.3.	ΕΛΛΕΠΟΥΣΕΣ ΤΙΜΕΣ	36
3.4.	ΠΕΡΙΚΕΚΟΜΜΕΝΑ ΔΕΔΟΜΕΝΑ	37
3.5.	ΕΛΛΕΙΨΗ ΚΑΝΟΝΙΚΟΤΗΤΑΣ	39
3.6.	ΕΠΟΧΙΚΟΤΗΤΑ	42
3.7.	ΣΕΙΡΙΑΚΗ ΣΥΣΧΕΤΙΣΗ (ΑΥΤΟΣΥΣΧΕΤΙΣΗ)	44
ΚΕΦΑΛΑΙΟ 4: ΑΝΑΛΥΣΗ ΤΑΣΕΩΝ		47
4.1.	ΕΙΣΑΓΩΓΗ	47
4.2.	ΠΡΟΣΕΓΓΙΣΕΙΣ ΣΤΗΝ ΑΝΑΛΥΣΗ ΤΑΣΕΩΝ	47
4.2.1.	4.2.1. Παραμετρική ανάλυση τάσης	47
4.2.2.	4.2.2. Μη παραμετρική ανάλυση τάσης	49
ΚΕΦΑΛΑΙΟ 5: ΠΟΛΥΜΕΤΑΒΛΗΤΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ		53
5.1.	ΕΙΣΑΓΩΓΗ – ΣΤΟΧΟΙ ΤΗΣ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ	53
5.2.	ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ / ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ	54
5.2.1.	Εισαγωγή	54
5.2.2.	Δημιουργία του πίνακα δεδομένων	55
5.2.3.	Βασική ιδέα της μεθόδου	56
5.2.4.	Επιλογή του πίνακα που θα χρησιμοποιηθεί για τη φασματική ανάλυση	56
5.2.5.	Φασματική ανάλυση του πίνακα συνδιακυμάνσεων ή του πίνακα συσχετίσεων	57
5.2.6.	Ποσοστό της συνολικής διακύμανσης που εξηγούν οι κύριες συνιστώσες	59
5.2.7.	Επιλογή του αριθμού κύριων συνιστωσών που επιλέγονται	59

5.2.8.	Ερμηνεία των κύριων συνιστωσών	60
5.2.9.	Περιστροφή των αξόνων	61
5.2.10.	Ανακατασκευή των αρχικών μεταβλητών από τις κύριες συνιστώσες	62
5.2.11.	Προϋποθέσεις για την ανάλυση σε κύριες συνιστώσες και την παραγοντική ανάλυση	63
5.2.12.	Διαφορές μεταξύ ανάλυσης σε κύριες συνιστώσες και παραγοντικής ανάλυσης	63
5.3.	ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ	64
5.3.1.	Εισαγωγή	64
5.3.2.	Απόσταση και ομοιότητα	65
5.3.3.	Ιεραρχική ομαδοποίηση	66
5.3.4.	Η μέθοδος k-means	67
5.4.	ΔΙΑΧΩΡΙΣΤΙΚΗ ΑΝΑΛΥΣΗ	67
5.4.1.	Αντικείμενο της διαχωριστικής ανάλυσης	67
5.4.2.	Ερευνητικός σχεδιασμός για τη διαχωριστική ανάλυση	68
5.4.3.	Προϋποθέσεις για τη διαχωριστική ανάλυση και διορθωτικές ενέργειες	68
5.4.4.	Εκτίμηση των διαχωριστικών συναρτήσεων και αποτίμηση της καταλληλότητας του μοντέλου	70
5.4.5.	Ερμηνεία των αποτελεσμάτων της διαχωριστικής ανάλυσης	72
5.4.6.	Επικύρωση των αποτελεσμάτων της διαχωριστικής ανάλυσης	73
5.5.	ΔΕΝΤΡΑ ΤΑΞΙΝΟΜΗΣΗΣ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	74
5.5.1.	Εισαγωγή	74
5.5.2.	Κανόνες διαχωρισμού (Splitting rules) για τα δέντρα ταξινόμησης	75
5.5.3.	Περικοπή του δέντρου (Pruning)	76
ΚΕΦΑΛΑΙΟ 6: ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΔΙΕΡΓΑΣΙΩΝ (SPC)		81
6.1.	ΕΙΣΑΓΩΓΗ	81
6.2.	ΔΙΑΓΡΑΜΜΑΤΑ SHEWHART	82
6.3.	ΔΙΑΓΡΑΜΜΑΤΑ ΕΚΘΕΤΙΚΑ ΣΤΑΘΜΙΣΜΕΝΩΝ ΜΕΣΩΝ (EWMA)	86

6.4.	ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΣΣΩΡΕΥΜΕΝΩΝ ΑΘΡΟΙΣΜΑΤΩΝ (CUSUM)	89
6.5	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΩΝ ΔΙΑΓΡΑΜΜΑΤΩΝ ΕΛΕΓΧΟΥ	91

ΚΕΦΑΛΑΙΟ 7: ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΓΙΑ ΑΥΤΟΣΥΣΧΕΤΙΣΜΕΝΑ ΔΕΔΟΜΕΝΑ

7.1.	ΕΙΣΑΓΩΓΗ	97
7.2.	ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΤΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ	98
7.3.	ΥΠΟΔΕΙΓΜΑΤΑ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ	101
	7.3.1. Εισαγωγή	101
	7.3.2. Αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως AR(1)	101
	7.3.3. Αυτοπαλίνδρομα υποδείγματα μεγαλύτερης τάξεως	102
	7.3.4. Υποδείγματα κινητών μέσων	103
	7.3.5. Μεικτά υποδείγματα ARMA(p,q)	104
	7.3.6. Υποδείγματα ARIMA(p,d,q)	105
7.4.	ΕΥΡΕΣΗ ΚΑΤΑΛΛΗΛΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ ARIMA	106
7.5.	ΔΙΑΓΝΩΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΚΑΤΑΛΛΗΛΟΤΗΤΑΣ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ	107
7.6.	ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΔΙΕΡΓΑΣΙΩΝ ΜΕ ΧΡΗΣΗ ΥΠΟΔΕΙΓΜΑΤΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ	107
7.7.	ΣΥΓΚΡΙΣΗ ΔΙΑΓΡΑΜΜΑΤΩΝ ΕΛΕΓΧΟΥ ΤΩΝ ΚΑΤΑΛΟΙΠΩΝ ΚΑΙ ΤΥΠΙΚΩΝ ΔΙΑΓΡΑΜΜΑΤΩΝ ΕΛΕΓΧΟΥ SPC ΤΩΝ ΑΡΧΙΚΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ	109
7.8.	ΑΛΛΕΣ ΤΕΧΝΙΚΕΣ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΑΥΤΟΣΥΣΧΕΤΙΣΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ	112

ΚΕΦΑΛΑΙΟ 8: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΤΟΞΙΚΟΤΗΤΑΣ ΑΝΑ ΤΑΜΙΕΥΤΗΡΑ / ΛΙΜΝΗ

8.1.	ΕΙΣΑΓΩΓΗ	115
8.2.	ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ ΤΟΞΙΚΟΤΗΤΑΣ	118
8.3.	ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ	119
8.4.	ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ	123
	8.4.1. Προκαταρκτική στατιστική διερεύνηση των δεδομένων	123
	8.4.2. Ανάλυση τάσεων	129

8.4.3. Μεταβλητότητα της τοξικότητας μεταξύ των ταμειυτήρων / λιμνών	130
8.5. ΣΥΜΠΕΡΑΣΜΑΤΑ	132
ΚΕΦΑΛΑΙΟ 9: ΑΝΑΛΥΣΗ ΤΑΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΟΤΗΤΑ ΜΙΚΡΟΒΙΟΛΟΓΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ / ΛΙΜΝΩΝ	133
9.1. ΕΙΣΑΓΩΓΗ	133
9.2. ΠΕΡΙΓΡΑΦΗ ΘΕΣΕΩΝ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΔΕΔΟΜΕΝΩΝ	135
9.3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ	136
9.4. ΑΠΟΤΕΛΕΣΜΑΤΑ & ΣΥΖΗΤΗΣΗ	137
9.5. ΣΥΜΠΕΡΑΣΜΑΤΑ	141
ΚΕΦΑΛΑΙΟ 10: ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΦΥΣΙΚΟΧΗΜΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ/ΛΙΜΝΩΝ	143
10.1. ΕΙΣΑΓΩΓΗ	143
10.2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	144
10.2.1. Περιγραφή των δεδομένων	144
10.2.2. Πολυμεταβλητές στατιστικές τεχνικές	145
10.3. ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΘΕΣΕΩΝ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ / ΛΙΜΝΩΝ ΒΑΣΕΙ ΤΩΝ ΦΥΣΙΚΟ-ΧΗΜΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΟΥ ΝΕΡΟΥ	147
10.4. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ ΓΙΑ ΤΟΝ ΤΑΜΙΕΥΤΗΡΑ ΜΑΡΑΘΩΝΑ	150
10.4.1. Προκαταρκτική εξέταση των παραμέτρων και διαχείριση των δεδομένων	150
10.4.2. Ανάλυση σε κύριες συνιστώσες/Παραγοντική ανάλυση	154
10.4.3. Ανάλυση κατά συστάδες	157
10.4.4. Διαχωριστική ανάλυση	158
10.4.5. Δέντρα ταξινόμησης	160
10.5. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ ΓΙΑ ΤΗ ΛΙΜΝΗ ΥΛΙΚΗ	161
10.5.1. Προκαταρκτική εξέταση των παραμέτρων και διαχείριση των δεδομένων	161
10.5.2. Ανάλυση σε κύριες συνιστώσες/Παραγοντική ανάλυση	164

10.5.3.	Ανάλυση κατά συστάδες	166
10.5.4.	Διαχωριστική ανάλυση	167
10.5.5.	Δέντρα ταξινόμησης	168
10.6.	ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ ΓΙΑ ΤΟΝ ΤΑΜΙΕΥΤΗΡΑ ΜΟΡΝΟΥ	169
10.6.1.	Προκαταρκτική εξέταση των παραμέτρων και διαχείριση των δεδομένων	169
10.6.2.	Ανάλυση σε κύριες συνιστώσες/Παραγοντική ανάλυση	172
10.6.3.	Ανάλυση κατά συστάδες	174
10.6.4.	Διαχωριστική ανάλυση	175
10.6.5.	Δέντρα ταξινόμησης	176
10.7.	ΔΙΑΚΡΙΣΗ ΜΕΤΑΞΥ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ ΜΑΡΑΘΩΝΑ ΚΑΙ ΜΟΡΝΟΥ ΒΑΣΕΙ ΤΩΝ ΦΥΣΙΚΟΧΗΜΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΝΕΡΟΥ	177
10.8.	ΣΥΜΠΕΡΑΣΜΑΤΑ	179
 ΚΕΦΑΛΑΙΟ 11: ΔΙΑΣΦΑΛΙΣΗ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΕΠΕΞΕΡΓΑΣΜΕΝΟΥ ΝΕΡΟΥ ΚΑΙ ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΔΙΚΤΥΟΥ ΥΔΡΕΥΣΗΣ ΜΕ ΤΗ ΧΡΗΣΗ ΠΟΛΥΜΕΤΑΒΛΗΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΤΕΧΝΙΚΩΝ		181
11.1.	ΕΙΣΑΓΩΓΗ	181
11.2.	ΜΕΘΟΔΟΛΟΓΙΑ	183
11.2.1.	Μονάδες επεξεργασίας νερού, δεξαμενές δικτύου	183
11.2.2.	Περιγραφή των δεδομένων	184
11.2.3.	Χειρισμός των δεδομένων και πολυμεταβλητές στατιστικές μέθοδοι	187
11.3.	ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ	188
11.3.1.	Ανάλυση κύριων συνιστωσών	188
11.3.2.	Διαχωριστική ανάλυση	190
11.3.3.	Δέντρα ταξινόμησης και παλινδρόμησης	196
11.3.4.	Σύγκριση των προβλεπτικών μεθόδων	198
11.4.	ΣΥΜΠΕΡΑΣΜΑΤΑ	200

ΚΕΦΑΛΑΙΟ 12: ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΓΙΑ ΤΗΝ	203
ΤΟΞΙΚΟΤΗΤΑ ΤΟΥ ΠΟΣΙΜΟΥ ΝΕΡΟΥ –	
ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΗΣ ΔΟΜΗΣ ΤΗΣ ΤΟΞΙΚΟΤΗΤΑΣ	
12.1. ΕΙΣΑΓΩΓΗ	203
12.2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	204
12.2.1. Αναλυτικός τοξικολογικός έλεγχος του επεξεργασμένου νερού	204
12.2.2. Στατιστική ανάλυση	205
12.3. ΑΠΟΤΕΛΕΣΜΑΤΑ & ΣΧΟΛΙΑ	206
12.3.1. Τυπικά διαγράμματα ελέγχου SPC για τα δεδομένα τοξικότητας	206
12.3.2. Προσαρμογή καταλλήλων υποδειγμάτων ARIMA στα δεδομένα	209
12.4. ΣΥΜΠΕΡΑΣΜΑΤΑ	221
ΚΕΦΑΛΑΙΟ 13: ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ	223
ΒΙΒΛΙΟΓΡΑΦΙΑ	231
ΔΗΜΟΣΙΕΥΣΕΙΣ	265
ΠΑΡΑΡΤΗΜΑ	267

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

2.1	Σημεία δειγματοληψίας των ταμιευτήρων / λιμνών ΕΥΔΑΠ	32
5.1	Προσαρμογή των δεδομένων σε πίνακα για τη διεξαγωγή της ανάλυσης σε κύριες συνιστώσες	56
6.1	Τιμές του k και οι αντίστοιχες τιμές του h που δίνουν $ARL_0 = 370$ για το αμφίπλευρο CUSUM	93
8.1	Περιγραφικά στατιστικά μέτρα των δεδομένων τοξικότητας ανά ταμιευτήρα / λίμνη	123
8.2	Έλεγχος ομοιογένειας των διακυμάνσεων (Levene's test) ανά ταμιευτήρα / λίμνη	126
8.3	Περίληπτικά στατιστικά των ελέγχων Kruskal-Wallis για κάθε ταμιευτήρα / λίμνη (μεταβλητή ομαδοποίησης: Μήνας)	126
8.4	Έλεγχος ομοιογένειας των διακυμάνσεων (Levene's test) των προ-λευκασμένων χρονοσειρών κάθε ταμιευτήρα / λίμνης	129
8.5	Περίληπτικά στατιστικά των ελέγχων Kruskal-Wallis για τις προλευκασμένες χρονοσειρές κάθε ταμιευτήρα / λίμνης	129
8.6	Περίληπτικά στατιστικά για τους ελέγχους Mann-Kendall και τις εκτιμήσεις κλίσης του Sen για την τοξικότητα ανά ταμιευτήρα/λίμνη	130
8.7	Πίνακας συνάφειας προ λευκασμένων τιμών τοξικότητας των ταμιευτήρων / λιμνών σε σχέση με τη γενική διάμεσο	131
8.8	Περίληπτικά στατιστικά του ελέγχου διαμέσων (Median test) για τις προ-λευκασμένες τιμές τοξικότητας (μεταβλητή ομαδοποίησης: Ταμιευτήρας / λίμνη)	131
9.1	Περιγραφή χαρακτηριστικών των δεδομένων <i>Enterococci</i> ανά θέση δειγματοληψίας	137
9.2	Διάμεσες τιμές των μηνιαίων βακτηριδίων <i>Enterococci</i> ανά σημείο δειγματοληψίας το χρονικό διάστημα 1997-2009	138
9.3	Περίληπτικά στατιστικά για τους εποχικούς ελέγχους Kendall για τα βακτηρίδια <i>enterococci</i>	140
9.4	Περίληπτικά στοιχεία των ελέγχων ομοιογένειας	140
10.1	Διάμεσες τιμές των παραμέτρων την τριετία 2004-2006 για κάθε θέση δειγματοληψίας στους ταμιευτήρες/λίμνες	148

10.2	Περιγραφικά στατιστικά μέτρα των παραμέτρων για τον ταμιευτήρα Μαραθώνα	151
10.3	Μέτρα ασυμμετρίας και κύρτωσης των \log_{10} μετασχηματισμένων μεταβλητών για τον ταμιευτήρα Μαραθώνα	151
10.4	Συντελεστές συσχέτισης Spearman's rho για τις παραμέτρους του ταμιευτήρα Μαραθώνα	153
10.5	1 ^η προσέγγιση παραγοντικής ανάλυσης παραμέτρων του ταμιευτήρα Μαραθώνα (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization).	154
10.6	Παραγοντική ανάλυση των μετασχηματισμένων σε τάξεις μεγέθους παραμέτρων του ταμιευτήρα Μαραθώνα (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization)	157
10.7	Συντελεστές συσχέτισης μεταβλητής-διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για τον Μαραθώνα	159
10.8	Πίνακας ταξινόμησης για τη βηματική DA της χρονικής διακύμανσης του νερού στον Μαραθώνα	160
10.9	Περιγραφικά στατιστικά μέτρα των παραμέτρων για τη λίμνη Υλίκη	162
10.10	Μέτρα ασυμμετρίας και κύρτωσης των \log_{10} μετασχηματισμένων μεταβλητών για τη λίμνη Υλίκη	162
10.11	Συντελεστές συσχέτισης Spearman's rho για τις παραμέτρους της λίμνης Υλίκης	163
10.12	Παραγοντική ανάλυση των παραμέτρων της λίμνης Υλίκη (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization).	165
10.13	Συντελεστές συσχέτισης μεταβλητής-διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για την Υλίκη	167
10.14	Πίνακας ταξινόμησης για τη βηματική DA της χρονικής διακύμανσης του νερού στην Υλίκη	168
10.15	Περιγραφικά στατιστικά μέτρα των παραμέτρων για τον ταμιευτήρα Μόρνου	170
10.16	Μέτρα ασυμμετρίας και κύρτωσης των \log_{10} μετασχηματισμένων μεταβλητών για τον ταμιευτήρα Μόρνου	170

10.17	Συντελεστές συσχέτισης Spearman's rho για τις παραμέτρους του ταμιευτήρα Μόρνου	171
10.18	Παραγοντική ανάλυση των παραμέτρων του ταμιευτήρα Μόρνου (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization)	173
10.19	Συντελεστές συσχέτισης μεταβλητής-διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για τον Μόρνο	175
10.20	Πίνακας ταξινόμησης για τη βηματική DA της χρονικής διακύμανσης του νερού στον Μόρνο	176
10.21	Συντελεστές συσχέτισης μεταβλητής - διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για των ταμιευτήρων Μαραθώνα - Μόρνου	178
11.1	Αριθμητικοί μέσοι των ημερήσιων δοσολογιών θειϊκού αργιλίου και χλωρίου ανά ΜΕΝ το έτος 2007	183
11.2	Περιληπτικά περιγραφικά στατιστικά μέτρα και χαρακτηριστικά κατανομών των χημικών παραμέτρων ανά ΜΕΝ	186
11.3	Πίνακας φορτίσεων των παραγόντων για τα δεδομένα επεξεργασμένου νερού των ΜΕΝ	189
11.4	Περιληπτικά περιγραφικά στατιστικά μέτρα των χημικών παραμέτρων στους ταμιευτήρες Μόρνου και Μαραθώνα το έτος 2007	190
11.5	Συσχετίσεις μεταξύ των διαχωριστικών μεταβλητών και των τυποποιημένων κανονικών διαχωριστικών συναρτήσεων (βηματική μέθοδος)	193
11.6	Συντελεστές συναρτήσεων ταξινόμησης	195
11.7	Πίνακας ταξινόμησης που προέκυψε από την τυπική διαχωριστική ανάλυση της χωρικής διακύμανσης του επεξεργασμένου νερού	195
11.8	Πίνακας ταξινόμησης που προέκυψε από την διαχωριστική ανάλυση, βάσει ξεχωριστών πινάκων συνδιακύμανσης των κανονικών διαχωριστικών συναρτήσεων, της χωρικής διακύμανσης του επεξεργασμένου νερού	196
11.9	Πίνακας ταξινόμησης των δειγμάτων μέσω CART	198
12.1	Περιγραφικά στατιστικά μέτρα των δεδομένων τοξικότητας για PDM, NDM και ASP	206

12.2	Στατιστικά σφαλμάτων για τα επιλεγμένα υποδείγματα χρονολογικών σειρών των δεδομένων τοξικότητας των PDM, NDM και ASP	212
------	--	-----

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

2.1	Εξέλιξη κατανάλωσης νερού καθώς και πληθυσμού - υδρευτικών έργων στο υδρευτικό σύστημα της Αθήνας, από το 1927 έως το 2008	15
2.2	Γενική διάταξη του υδροδοτικού συστήματος της Αθήνας	18
2.3	Θέσεις δειματοληψίας ταμιευτήρων / λιμνών	33
5.1	Παράδειγμα δομής δέντρου ταξινόμησης	76
5.2	Σχέση μεταξύ της πολυπλοκότητας του δέντρου και του κόστους απόφασης	79
6.1	Διάγραμμα ελέγχου Shewhart μεμονωμένων ημερήσιων τιμών θολότητας (σε φυσική λογαριθμική κλίμακα) επεξεργασμένου νερού της ΜΕΝ Πολυδενδρίου	84
8.1	Οργανισμοί βιοδείκτες	117
8.2	Θηκογράμματα για την τοξικότητα ανά ταμιευτήρα / λίμνη	124
8.3	Διαγράμματα χρονολογικών σειρών τοξικότητας ανά ταμιευτήρα/λίμνη	125
8.4	Διαγράμματα συναρτήσεων δειγματικής αυτοσυσχέτισης για την τοξικότητα ανά ταμιευτήρα/λίμνη	127
9.1	Θηκογράμματα των <i>Enterococci</i> ανά μήνα σε κάθε θέση δειματοληψίας	139
10.1	Ιεραρχικό δεντρόγραμμα για τις θέσεις δειματοληψίας των τριών ταμιευτήρων/λιμνών	149
10.2	Χρονική διακύμανση των σκορ του πρώτου παράγοντα για τον Μαραθώνα	156
10.3	Ιεραρχικό δεντρόγραμμα για τις περιόδους παρακολούθησης του ταμιευτήρα Μαραθώνα	158
10.4	Δέντρο ταξινόμησης για τη χρονική διακύμανση των δεδομένων του Μαραθώνα	161
10.5	Χρονική διακύμανση των σκορ του πρώτου παράγοντα για την Υλίκη	166
10.6	Ιεραρχικό δεντρόγραμμα για τις περιόδους παρακολούθησης της Υλίκης	166
10.7	Δέντρο ταξινόμησης για την χρονική διακύμανση των δεδομένων της Υλίκης (αλγόριθμος CART)	169

10.8	Χρονική διακύμανση των σκορ του πρώτου παράγοντα για τον Μόρνο	174
10.9	Ιεραρχικό δεντρόγραμμα για τις περιόδους παρακολούθησης του Μόρνου	174
10.10	Δέντρο ταξινόμησης για την χρονική διακύμανση των δεδομένων του Μόρνου (αλγόριθμος CART)	177
11.1	Μονάδες επεξεργασίας νερού της ΕΥΔΑΠ και δεξαμενές δικτύου ύδρευσης που θεωρήθηκαν στη μελέτη	185
11.2	Κατανομή των δύο κύριων συνιστωσών που παρατηρήθηκαν για τις τρεις μονάδες επεξεργασίας νερού	188
11.3	Διάγραμμα διασποράς των δεδομένων επεξεργασμένου νερού των ΜΕΝ ως προς τις δύο διαχωριστικές συναρτήσεις	194
11.4	Δέντρο ταξινόμησης (αλγόριθμος CART) - Δείγμα εκμάθησης (ΜΕΝ)	197
12.1	Διαγράμματα ελέγχου Shewhart (I Charts) των αρχικών δεδομένων για τις δεξαμενές PDM, NDM και ASP	207
12.2	Συναρτήσεις δειγματικής αυτοσυσχέτισης (autocorrelation) και δειγματικής μερικής αυτοσυσχέτισης (partial autocorrelation) για τις PDM, NDM και ASP	210
12.3	Συναρτήσεις δειγματικής αυτοσυσχέτισης (autocorrelation) και δειγματικής μερικής αυτοσυσχέτισης (partial autocorrelation) για τις dPDM, dNDM και dASP	211
12.4	Διαγράμματα συναρτήσεων δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης για τις PDMf, NDMf και ASPf	213
12.5	Διαγράμματα CCC για τις PDM, NDM και ASP	215
12.6	Διαγράμματα ειδικών αιτιών SCC για τις PDMf, NDMf και ASPf	216
12.7	Διαγράμματα EWMA για τις PDMf, NDMf και ASPf	218
12.8	Διαγράμματα CUSUM για τις PDMf, NDMf και ASPf	219

ΟΝΟΜΑΤΟΛΟΓΙΑ

ASP	η χρονοσειρά των δεδομένων τοξικότητας της δεξαμενής επεξεργασμένου νερού του Ασπροπύργου
dASP	στάσιμη χρονοσειρά των πρώτων διαφορών της ASP
ASPf	η χρονοσειρά των καταλοίπων για την τοξικότητα στη δεξαμενή επεξεργασμένου νερού του Ασπροπύργου
NDM	η χρονοσειρά των δεδομένων τοξικότητας της νέας δεξαμενής επεξεργασμένου νερού του Μενιδίου (Αχαρνών)
dNDM	στάσιμη χρονοσειρά των πρώτων διαφορών της NDM
NDMf	η χρονοσειρά των καταλοίπων για την τοξικότητα στη νέα δεξαμενή επεξεργασμένου νερού του Μενιδίου (Αχαρνών)
PDM	η χρονοσειρά των δεδομένων τοξικότητας της παλαιάς δεξαμενής επεξεργασμένου νερού του Μενιδίου (Αχαρνών)
dPDM	στάσιμη χρονοσειρά των πρώτων διαφορών της PDM
PDMf	η χρονοσειρά των καταλοίπων για την τοξικότητα στην παλαιά δεξαμενή επεξεργασμένου νερού του Μενιδίου (Αχαρνών)
MEN	Μονάδες Επεξεργασίας Νερού
ΑΣΠ	Δεξαμενή επεξεργασμένου νερού της MEN Ασπροπύργου
ΑΧ	Δεξαμενή επεξεργασμένου νερού της MEN Αχαρνών
ΓΑΛ	Δεξαμενή επεξεργασμένου νερού της MEN Γαλατσίου
ΠΟΛ	Δεξαμενή επεξεργασμένου νερού της MEN Πολυδενδρίου

ΣΥΜΒΟΛΑ ΣΤΑΤΙΣΤΙΚΗΣ

c_4	συνάρτηση του μεγέθους του δείγματος, η οποία χρησιμοποιείται για την εκτίμηση της τυπικής απόκλισης μέσω της δειγματικής τυπικής απόκλισης
d	η τάξη των μη-εποχικών διαφορών των στοιχείων μίας χρονοσειράς για την επίτευξη στασιμότητας
d_2	συνάρτηση των διαγραμμάτων ελέγχου, που χρησιμοποιείται για την εκτίμηση της τυπικής απόκλισης μέσω των κινητών ευρών

e_i	κατάλοιπο ή τυχαίο λάθος, δηλ. η διαφορά μεταξύ της παρατηρηθείσας τιμής και της προβλεφθείσας τιμής για το χρονικό διάστημα i
h	πολλαπλασιαστής της τυπικής απόκλισης σε ένα διάγραμμα ελέγχου συσσωρευμένων αθροισμάτων
H	το διάστημα απόφασης σε ένα διάγραμμα ελέγχου συσσωρευμένων αθροισμάτων
K	η τιμή αναφοράς για το διάγραμμα ελέγχου συσσωρευμένων αθροισμάτων
MR	κινητό εύρος
\overline{MR}	ο μέσος των κινητών ευρών
r	συντελεστής συσχέτισης
s	η δειγματική τυπική απόκλιση
S	στατιστικό του ελέγχου Mann-Kendall
S_{Hi}	το ανώτερο συσσωρευτικό άθροισμα για το δείγμα i σε ένα διάγραμμα ελέγχου συσσωρευμένων αθροισμάτων
S_{Li}	το κατώτερο συσσωρευτικό άθροισμα για το δείγμα i σε ένα διάγραμμα ελέγχου συσσωρευμένων αθροισμάτων
X_i	η τιμή της παρατήρησης τη χρονική περίοδο i
\bar{X}	ο δειγματικός μέσος (αριθμητικός μέσος)
ε	τυχαίο λάθος
λ	παράμετρος που χρησιμοποιείται στον τύπο της στατιστικής συνάρτησης του διαγράμματος ελέγχου εκθετικά σταθμισμένου κινητού μέσου
λ_i	ιδιοτιμή της i κύριας συνιστώσας
μ	μέση τιμή
ρ	δειγματικός συντελεστής συσχέτισης
σ^2	διακύμανση
σ	τυπική απόκλιση

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ ΣΤΑΤΙΣΤΙΚΗΣ

ACF	Συνάρτηση Αυτοσυσχέτισης (Autocorrelation Function)
AIC	κριτήριο πληροφοριών Akaike (Akaike's information criterion)
ARL	μέσο μήκος ροής (Average Run Length)
ARL ₀	εντός-ελέγχου μέσο μήκος ροής
BIC	Μπεϋζιανό κριτήριο πληροφοριών (Bayesian information criterion ή κριτήριο Schwartz)
CCC	διάγραμμα κοινών αιτιών (Common Cause Chart)
CA	Ανάλυση κατά συστάδες (Cluster Analysis)
CART	Δέντρα Ταξινόμησης και Παλινδρόμησης (Classification and Regression Trees)
Cov	Συνδιακύμανση (Covariance)
CUSUM	διάγραμμα ελέγχου συσσωρευμένων αθροισμάτων (Cumulative Sum Chart)
CV	Συντελεστής Μεταβλητότητας (Coefficient of Variation)
DA	Διακριτή ή Διαχωριστική Ανάλυση (Discriminant Analysis)
EWMA	διάγραμμα ελέγχου εκθετικά σταθμισμένου κινητού μέσου (Exponentially Weighted Moving Average)
FA	Παραγοντική Ανάλυση (Factor Analysis)
MAE	Μέσο Απόλυτο Σφάλμα (Mean Absolute Error)
ME	Μέσο Σφάλμα (Mean Error)
MSE	Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error)
PACF	Συνάρτηση Μερικής Αυτοσυσχέτισης (Partial Autocorrelation Function)
PCA	Ανάλυση σε κύριες Συνιστώσες (Principal Components Analysis)
RMSE	τετραγωνική ρίζα του τετραγωνικού σφάλματος (Root Mean Square Error)
QC	Έλεγχος Ποιότητας (Quality Control)
SCC	διάγραμμα ειδικών αιτιών (Special Cause Chart)
SPC	Στατιστικός Έλεγχος Διεργασιών (Statistical Process Control)
Var	Διακύμανση (Variance)

ΠΡΟΛΟΓΟΣ

Το νερό αποτελεί έναν από τους πιο σημαντικούς φυσικούς πόρους. Πέρα από τη ζωτική του σημασία για τη διατήρηση της ανθρώπινης ζωής, είναι απαραίτητο στη φύση για τη διατήρηση των υπολοίπων ζωικών και φυτικών οργανισμών. Χρησιμοποιείται από τον άνθρωπο όχι μόνο προς πόση, αλλά και ως μέσο προς εξυπηρέτηση άλλων αναγκών του που αφορούν τη βελτίωση των όρων της διαβίωσής του ή την εξασφάλιση αγαθών. Ιστορικά, η εξέλιξη και η ιστορία του ανθρώπου ήταν συνδεδεμένη με τη διαχείριση του νερού. Σε πολλές περιπτώσεις όμως η διαχείριση της ποιότητας και ποσότητας του νερού δεν είναι πάντα η καλύτερη δυνατή.

Το πόσιμο νερό είναι το πολυτιμότερο αγαθό και η ποιότητά του απασχολεί την επιστημονική κοινότητα όλο και περισσότερο καθώς τα τελευταία χρόνια επιβαρύνεται με ανεπιθύμητες χημικές ενώσεις που εμφανίζουν δυσμενείς επιπτώσεις στην ανθρώπινη υγεία. Επιπλέον είναι γνωστό ότι μέσω του νερού μεταδίδονται πολυάριθμες υδατογενείς ασθένειες που στο παρελθόν κλόνισαν πολιτισμούς.

Ένα βασικό πρόβλημα για πολλές περιοχές είναι η ελλιπής πληροφόρηση σχετικά με τους υδατικούς πόρους. Οι μετρήσεις για τους υδατικούς πόρους πρέπει να είναι ακριβείς και να έχουν ικανοποιητική συχνότητα ενώ παράλληλα να παρουσιάζουν την κατάλληλη γεωγραφική κατανομή-πυκνότητα και να παρέχουν κάλυψη για μεγάλα χρονικά διαστήματα. Η γνώση της διακύμανσης των ποιοτικών παραμέτρων του νερού, με τον τόπο και τον χρόνο, κρίνεται απαραίτητη διότι χαρακτηρίζει την ποιότητα του νερού.

Η εκτίμηση της διαθεσιμότητας του νερού βασίζεται σε μια σειρά πληροφορίες μεταξύ των οποίων σημαντική θέση κατέχει η ποιότητα του νερού η οποία θα καθορίσει στη συνέχεια και την τελική του χρήση.

Ο έλεγχος της ποιότητας του νερού είναι σημαντικός και στο μέλλον θα γίνει πιο επιτακτικός με δεδομένη την αύξηση των ανεπιθύμητων χημικών ουσιών στο περιβάλλον και την εισροή μέρους αυτών στους υδάτινους αποδέκτες. Οι διαρκείς πιέσεις που δέχεται το γλυκό νερό ως φυσικός πόρος εξαιτίας της ρύπανσης και της υπέρμετρης χρήσης δημιουργεί κινδύνους που αυξάνουν το εξωτερικό κόστος χρήσης

του νερού στο σπίτι, στην βιομηχανία, στην γεωργία και στην παραγωγή τροφίμων. Χημικώς καθαρό νερό δεν υπάρχει στη φύση διότι το νερό είναι ισχυρό διαλυτικό μέσο. Συνεπώς εμπλουτίζεται κατά την επαφή του με το περιβάλλον με διάφορες ουσίες που διαλύονται απ' αυτό. Παρουσιάζει μια μεγάλη ποικιλία ποιοτικών χαρακτηριστικών ανάλογα με τις περιεχόμενες ξένες προσμίξεις.

Το πόσιμο νερό θα έπρεπε να είναι το καλύτερα ελεγχόμενο μέσο διατροφής. Η νομοθεσία προσδιορίζει τις συγκεντρώσεις διαφόρων ουσιών, που επιτρέπεται να υπάρχουν μέσα στο πόσιμο νερό, ώστε να ανταποκρίνεται στις υψηλές ποιοτικές προδιαγραφές, που απαιτούνται σε σχέση με το σημαντικό για τη ζωή μας αγαθό.

Κατά τις τελευταίες δεκαετίες η φυσική ποιότητα των υδατικών πόρων μεταβλήθηκε σημαντικά εξ' αιτίας των διαφόρων ανθρώπινων δραστηριοτήτων και χρήσεων του νερού. Οι περισσότερες περιπτώσεις ρύπανσης αναπτύχθηκαν βαθμιαία μέχρις ότου έγιναν φανερές και μετρήσιμες. Χρειάστηκε αρκετός χρόνος μέχρι να αναγνωριστούν τα προβλήματα ρύπανσης και ακόμα περισσότερος για να γίνουν οι απαραίτητες μετρήσεις και οι έλεγχοι.

Στην σύγχρονη εποχή, επιβάλλεται ο έλεγχος ποιότητας του νερού για ένα μεγάλο αριθμό παραμέτρων, όπως οι φυσικές, χημικές, μικροβιολογικές, βιολογικές. Η συμμόρφωση με τις προδιαγραφές της νομοθεσίας θεωρείται απαραίτητη για τη διασφάλιση της ποιότητας του νερού. Επιπλέον, καθώς νέες χημικές ενώσεις δημιουργούνται διαρκώς με άγνωστες ακόμα συνέπειες για το περιβάλλον, οι μετρήσεις τοξικότητας του νερού, δηλαδή η έκφραση της επίδρασης του νερού σε ζωντανούς οργανισμούς-δείκτες, είναι ιδιαίτερα χρήσιμες. Ωστόσο, η αξιολόγηση των αποτελεσμάτων των αναλύσεων των ποιοτικών χαρακτηριστικών του νερού δεν πρέπει να περιορίζεται απλά στην τήρηση των προδιαγραφών. Η αναζήτηση πληροφοριών για τη συμπεριφορά των ποιοτικών παραμέτρων του νερού επιβάλλεται να είναι συνεχής. Η συμβολή της στατιστικής μπορεί να αποδειχθεί πολύτιμη για την μελέτη της συμπεριφοράς των ποιοτικών παραμέτρων του νερού, τη διερεύνηση της μεταβλητότητάς τους, την χρονική και εποχική εξέλιξη των τιμών τους, την κατανόηση των αλληλεπιδράσεων μεταξύ τους και τις μελλοντικές προβλέψεις των τιμών τους. Σε κάθε περίπτωση όμως, επιβάλλεται η αρχική διερεύνηση των στατιστικών ιδιοτήτων των παραμέτρων ποιότητας του νερού, ώστε να επιλέγονται

στη συνέχεια οι κατάλληλες στατιστικές μέθοδοι για το είδος των διαθέσιμων δεδομένων.

Στην παρούσα διατριβή, διερευνάται η αποτελεσματικότητα, η επάρκεια και οι αναγκαίες προσαρμογές, λόγω των ιδιαίτερων χαρακτηριστικών των ποιοτικών παραμέτρων του νερού, ενός μεγάλου αριθμού σύνθετων στατιστικών μεθόδων. Σκοπός της διατριβής ήταν η ανάπτυξη μεθοδολογίας για την εκτίμηση της ποιότητας του νερού μέσω της ολοκληρωμένης στατιστικής ανάλυσης των χημικών, των μικροβιολογικών και των τοξικολογικών παραμέτρων του νερού, από την ακατέργαστη μορφή του στις πηγές υδροληψίας μέχρι και την επεξεργασμένη μορφή του ως νερό προς ανθρώπινη κατανάλωση, που θα οδηγήσει στην ολοκληρωμένη αξιολόγηση και διαχείριση ενός συστήματος υδατικών πόρων. Κύριο στόχο της έρευνας αποτέλεσε η κατανόηση των μηχανισμών που επηρεάζουν τα ποιοτικά χαρακτηριστικά του νερού από την πρωταρχική του μορφή έως την τελική που προορίζεται για ανθρώπινη κατανάλωση. Επιπλέον στόχο αποτέλεσε, η ανάπτυξη και ο έλεγχος καταλληλότητας ποικίλων συστημάτων τα οποία θα διαθέτουν δυνατότητα αντίδρασης σε διάφορες κλίμακες χρόνου και χώρου και θα επικεντρώνονται σε συστήματα έγκαιρης προειδοποίησης για τον έλεγχο των εγκαταστάσεων επεξεργασίας νερού.

Για την εφαρμογή των παραπάνω, επιλέχτηκε το υδρευτικό σύστημα της Αθήνας για τους παρακάτω λόγους:

- Έχει πολύπλοκη σύνθεση. Το ακατέργαστο νερό προέρχεται από διαφορετικούς ταμιευτήρες / λίμνες οι περισσότεροι από τους οποίους βρίσκονται σε μεγάλη απόσταση από τις μονάδες επεξεργασίας. Διαθέτει τέσσερεις μονάδες επεξεργασίας νερού και το δίκτυο ύδρευσης έχει μήκος περίπου 8.000 km.
- Είναι ιδιαίτερα σημαντική η ποιότητα του νερού των Αθηνών δεδομένου ότι υδροδοτεί περίπου 4.000.000 καταναλωτές.

Παρόλο που οι ποιότητα του νερού ελέγχεται σύμφωνα με τη νομοθεσία σε τακτική βάση, δεν έχει γίνει στο παρελθόν ολοκληρωμένη στατιστική ανάλυση των αποτελεσμάτων των αναλυτικών προσδιορισμών. Η μεθοδολογία που ακολουθείται μπορεί να εφαρμοστεί και σε άλλα υδρευτικά συστήματα με τις κατάλληλες

προσαρμογές λόγω των ιδιαίτερων χαρακτηριστικών του νερού ή των λειτουργικών διαδικασιών τους. Η ολοκληρωμένη αυτή στατιστική προσέγγιση δεν έχει παρατηρηθεί σε άλλη περίπτωση.

Η διάρθρωση της παρούσας διατριβής έχει ως εξής:

Στο 1^ο Κεφάλαιο, γίνεται συνοπτική αναφορά στην προέλευση του νερού που προορίζεται προς ανθρώπινη κατανάλωση, στα κυριώτερα σύγχρονα προβλήματα ποιότητας του νερού στις πηγές υδροληψίας, στο ισχύον ευρωπαϊκό νομοθετικό πλαίσιο για την ποιότητα του νερού και στις μεθόδους επεξεργασίας του. Στο 2^ο Κεφάλαιο, περιγράφεται το υδρευτικό σύστημα της Αθήνας, η μεθοδολογία επεξεργασίας και το δειγματοληπτικό σχήμα του αναλυτικού ελέγχου ποιότητας του νερού που ακολουθεί η ΕΥΔΑΠ συμμορφούμενη προς τη νομοθεσία. Στο 3^ο Κεφάλαιο, παρουσιάζονται τα ιδιαίτερα χαρακτηριστικά των παραμέτρων ποιότητας του νερού τα οποία δημιουργούν δυνητικά προβλήματα κατά την στατιστική ανάλυση. Επιπλέον, μέσω της ανασκόπησης της επιστημονικής βιβλιογραφίας, προτείνονται εναλλακτικοί τρόποι στατιστικής προσέγγισης.

Στα κεφάλαια 4 έως 7, παρουσιάζονται στατιστικές τεχνικές οι οποίες χρησιμοποιούνται στην παρούσα έρευνα. Συγκεκριμένα, στο 4^ο Κεφάλαιο, περιγράφονται τεχνικές ανάλυσης τάσεων, που αφορούν στον έλεγχο πιθανής μεταβολής (αυξητικής ή πτωτικής) των συγκεντρώσεων των ποιοτικών παραμέτρων του νερού στη διάρκεια του χρόνου. Στο 5^ο Κεφάλαιο παρουσιάζονται πολυμεταβλητές στατιστικές τεχνικές. Οι τεχνικές αυτές παρέχουν τη δυνατότητα του συνδυασμού πληροφοριών πολλών αλληλοεξαρτώμενων παραμέτρων ποιότητας του νερού και η κατάλληλη χρησιμοποίησή τους μπορεί να δώσει πολύτιμες πληροφορίες. Στο 6^ο Κεφάλαιο παρουσιάζονται μέθοδοι Στατιστικού Ελέγχου Διεργασιών (SPC) ο οποίος αποσκοπεί στη μείωση της μεταβλητότητας με στόχο τη βελτίωση της ποιότητας και τον εντοπισμό ασυνήθιστων τιμών των ποιοτικών παραμέτρων του νερού. Στο 7^ο Κεφάλαιο, παρουσιάζονται τεχνικές αναπροσαρμογής του SPC, οι οποίες είναι απαραίτητες όταν τα δεδομένα δεν είναι ανεξάρτητα μεταξύ τους, λόγω μικρού χρονικού διαστήματος, που μεσολαβεί μεταξύ των διαδοχικών δειγματοληψιών. Λόγω των απαιτήσεων της κύριας τεχνικής, παρουσιάζονται επιπλέον οι βασικές έννοιες και τα υποδείγματα των χρονολογικών σειρών.

Στο 8^ο Κεφάλαιο, διερευνάται κατά πόσο υπάρχει τυχόν αυξητική τάση του επιπέδου τιμών της τοξικότητας στους ταμιευτήρες / λίμνες του υδροδοτικού συστήματος της Αθήνας σε μία περίοδο επτά ετών. Επιπλέον ελέγχονται τυχόν εποχικές διακυμάνσεις της τοξικότητας και η μεταβλητότητά της μεταξύ των ταμιευτήρων / λιμνών. Στο 9^ο Κεφάλαιο, εξετάζεται η μικροβιακή φόρτιση των ποταμών / ρεμάτων που εκβάλλουν στους ταμιευτήρες / λίμνες ως προς το βακτήριο – δείκτη *enterococci* και διερευνάται η εποχικότητα και η ύπαρξη τάσεων για μία περίοδο δεκατριών ετών. Στο 10^ο Κεφάλαιο, χρησιμοποιούνται πολυμεταβλητές στατιστικές τεχνικές για την ερμηνεία της δομής των φυσικοχημικών ποιοτικών παραμέτρων του νερού των ταμιευτήρων / λιμνών και την αποτίμηση των χρονικών διακυμάνσεων της ποιότητάς του. Στο 11^ο Κεφάλαιο, διευρενήθηκε η χρησιμότητα των πολυμεταβλητών στατιστικών τεχνικών ως μέσον διάκρισης μεταξύ του νερού που προέρχεται από διαφορετικές μονάδες επεξεργασίας νερού (MEN). Επιπλέον, οι τεχνικές αυτές χρησιμοποιούνται για την ανάπτυξη μοντέλων πρόβλεψης της προέλευσης του νερού των δεξαμενών δικτύου ύδρευσης. Τα μοντέλα αυτά επιτρέπουν την ιχνηλάτιση της ροής του νερού στο δίκτυο ύδρευσης και τη διαχείριση προβλημάτων, που τυχόν παρουσιάζονται στις MEN, πριν φτάσει το νερό στους καταναλωτές. Στο 12^ο Κεφάλαιο εφαρμόζονται τεχνικές SPC σε δεδομένα τοξικότητας του επεξεργασμένου νερού των MEN με στόχο τη βελτιστοποίηση της ποιότητας και τη δυνατότητα βραχυχρόνιων προβλέψεων ώστε, εάν χρειαστεί, να είναι δυνατή η επαναπροσαρμογή της διεργασίας.

ΚΕΦΑΛΑΙΟ 1

ΝΕΡΟ ΑΝΘΡΩΠΙΝΗΣ ΚΑΤΑΝΑΛΩΣΗΣ

1.1. ΕΙΣΑΓΩΓΗ

Το νερό είναι το πολυτιμότερο συστατικό στοιχείο του φυσικού περιβάλλοντος. Είναι το απαραίτητο οικονομικό, κοινωνικό και περιβαλλοντικό αγαθό για την επιβίωση του ανθρώπου, τη διατήρηση του οικοσυστήματος και την ανάπτυξη στον πλανήτη μας. Ωστόσο, η μη ορθολογική διαχείριση, η υπερκατανάλωση, οι κλιματικές αλλαγές και η ρύπανση, επιφέρουν μείωση των διαθέσιμων αποθεμάτων και υποβάθμιση της ποιότητάς του.

Ως νερό ανθρώπινης κατανάλωσης, θεωρείται (98/83/ΕΚ):

α) το νερό το οποίο προορίζεται για πόση, μαγείρεμα, προπαρασκευή τροφής ή άλλες οικιακές χρήσεις, είτε στη φυσική του κατάσταση είτε μετά από επεξεργασία και ανεξάρτητα από την προέλευσή του και από το αν παρέχεται από δίκτυο διανομής, από βυτίο, ή σε φιάλες ή δοχεία.

β) το νερό που χρησιμοποιείται στις επιχειρήσεις παραγωγής τροφίμων για την παρασκευή, επεξεργασία, συντήρηση ή εμπορία προϊόντων ή ουσιών, που προορίζονται για ανθρώπινη κατανάλωση.

Το πόσιμο νερό θα πρέπει να είναι καθαρό από φυσική, χημική, βιολογική και μικροβιολογική άποψη και να μην να εγκυμονεί κινδύνους για την υγεία των καταναλωτών.

1.2. ΥΔΑΤΙΚΟΙ ΠΟΡΟΙ - ΡΥΠΑΝΣΗ

Το πόσιμο νερό μπορεί να προέλθει από διάφορες πηγές. Οι τυπικοί υδατικοί πόροι περιλαμβάνουν το γλυκό νερό των λιμνών, των ποταμών και των υπογείων υδροφορέων σε μικρά και μεσαία βάθη και αποτελούν ποσοστό μικρότερο του 1% του συνόλου των νερών του πλανήτη (Ξανθόπουλος και Κουτσογιάννης, 1997, Λέκκας, 2005). Ως γλυκό θεωρείται το νερό που έχει συγκέντρωση ολικών

διαλυμένων στερεών μικρότερη από 1000 mg/l (Τσώνης, 2003). Για την παραλαβή πόσιμου νερού χρησιμοποιούνται σχεδόν αποκλειστικά τα υδατικά αποθέματα γλυκού νερού (Τσώνης, 2003). Λόγω της μείωσης των αποθεμάτων των υδατικών πόρων, σε αρκετές περιπτώσεις χρησιμοποιείται η αφαλάτωση ενώ εξετάζεται η ανακύκλωση των υγρών αποβλήτων με στόχο την επαναχρησιμοποίηση του νερού.

Τα κυριότερα σύγχρονα προβλήματα για την ποιότητα των υδατικών πόρων είναι (Αντωνόπουλος, 2001; Νταρακάς, 2009, Μαμάης, 2006):

- Η μόλυνση του νερού από παθογόνους μικροοργανισμούς (κύρια απειλή στις υποανάπτυκτες και αναπτυσσόμενες χώρες). Η μικροβιακή μόλυνση του νερού οφείλεται κυρίως σε παθογόνους μικροοργανισμούς που περιέχονται στα περιττώματα ανθρώπων και ζώων και εισέρχονται στα υδατικά συστήματα μέσω αστικών και κτηνοτροφικών αποβλήτων αντίστοιχα.
- Εποχική μείωση του οξυγόνου λόγω υπερφόρτωσης των ποταμών με αποικοδομούμενα οργανικά λύματα αστικής και βιομηχανικής προέλευσης. Τα φυσικά νερά που δέχονται υψηλά φορτία οργανικής ύλης έχουν χαμηλή συγκέντρωση κορεσμού σε οξυγόνο η οποία μειώνεται περαιτέρω με την άνοδο της θερμοκρασίας. Οι συνέπειες είναι πολύ σοβαρές για τους υδρόβιους οργανισμούς οι οποίοι κινδυνεύουν από ασφυξία.
- Η υπερφόρτιση των υδατορευμάτων με βιοαποικοδομήσιμα οργανικά απόβλητα από τους παρόχθιους οικισμούς και βιομηχανίες.
- Ο ευτροφισμός, ο οποίος ορίζεται ως η υπέρμετρη αύξηση της πρωτογενούς παραγωγικότητας (φυτοπλαγκτόν, υδρόβια και υδροχαρής βλάστηση) σε ένα επιφανειακό υδάτινο σώμα, με δυσμενή αποτελέσματα στα φυσικοχημικά και βιολογικά χαρακτηριστικά των υδάτων και της χρήσης τους.
- Βαθμιαία αύξηση των βαρέων μετάλλων στα ιζήματα και στο νερό των ποταμών και λιμνών.
- Παρουσία συνθετικών οργανικών ουσιών στα υπόγεια και επιφανειακά νερά, λόγω της εκτεταμένης παραγωγής και χρήσης τους.
- Ατμοσφαιρική μεταφορά των αερίων ρύπων από τις καύσεις των ορυκτών καυσίμων, οξίνιση των λιμνών και των ποταμών και η μεταφορά των ρύπων αυτών στα υπόγεια νερά.

- Υπερβάσεις των συνιστώμενων ορίων για τα νιτρικά στα υπόγεια και επιφανειακά νερά, λόγω της εκτεταμένης χρήσης των αζωτούχων λιπασμάτων και των στερεών αποβλήτων (ζώων και λάσπης βιολογικών σταθμών).

Ο ευτροφισμός αποτελεί φυσιολογική διαδικασία γήρανσης ενός επιφανειακού αποδέκτη. Ωστόσο, ο ρυθμός της διαδικασίας επηρεάζεται από διάφορες ανθρώπινες δραστηριότητες. Το φαινόμενο του ευτροφισμού αποτελεί πολύ σοβαρή διαταραχή των επιφανειακών υδάτων, με πολλές δυσμενείς συνέπειες όπως η αποξυγόνωση, η μείωση της διαφάνειας και η δυσοσμία του νερού. Στο φαινόμενο του ευτροφισμού, συμβάλλουν κυρίως τα θρεπτικά άλατα του αζώτου (αμμωνιακά, νιτρώδη, νιτρικά) και του φωσφόρου (φωσφορικά) τα οποία οφείλονται κυρίως σε ανθρωπογενείς δραστηριότητες και εισέρχονται στα υδατικά συστήματα μέσω των γεωργικών απορροών (αποπλύσεις καλλιεργούμενων εκτάσεων - λιπάσματα), των αστικών λυμάτων, των κτηνοτροφικών αποβλήτων και ορισμένων βιομηχανικών αποβλήτων (Νταρακάς, 2009, Μαμάης, 2006). Ανάλογα με την τροφική τους κατάσταση, τα υδάτινα σώματα διακρίνονται σε oligοτροφικά, μεσοτροφικά και ευτροφικά.

Ως βαρέα μέταλλα χαρακτηρίζονται τα στοιχεία με ειδικό βάρος μεγαλύτερο ή ίσο των 5 g/cm^3 όπως ο μόλυβδος (Pb), το κάδμιο (Cd), ο υδράργυρος (Hg), ο ψευδάργυρος (Zn), ο χαλκός (Cu), το χρώμιο (Cr), ο σίδηρος (Fe) κ.λ.π. Συχνά σε αυτούς τους ρύπους, περιλαμβάνονται και ελαφρύτερα στοιχεία όπως αργίλιο Al και βηρύλλιο Be και μεταλλοειδή όπως αρσενικό As, σελήνιο Se και αντιμόνιο Sb (Νταρακάς, 2009). Η τοξικότητά τους διαφοροποιείται ανάλογα με το είδος του μετάλλου και τη μορφή με την οποία είναι διαθέσιμα στο περιβάλλον. Ο υδράργυρος (Hg), το κάδμιο (Cd), ο μόλυβδος (Pb), το νικέλιο (Ni) και το αρσενικό (As), παρουσιάζουν πολύ υψηλή τοξικότητα ακόμη και σε χαμηλές συγκεντρώσεις. Η εισροή των μετάλλων στο νερό αποδίδεται σε φυσικές πηγές (διάβρωση πετρωμάτων, ηφαιστειακή δραστηριότητα) ή σε ανθρωπογενείς δραστηριότητες (βιομηχανία, μεταλλεία, αστικές απορροές).

Οι κυριότερες συνθετικές οργανικές ενώσεις είναι τα παρασιτοκτόνα, τα εντομοκτόνα, τα ζιζανιοκτόνα, οι διοξίνες, οι υδρογονάνθρακες του πετρελαίου, οι χλωριωμένοι υδρογονάνθρακες, τα πολυχλωριωμένα διαφαινύλια (PCB), οι φαινόλες, οι πολυκυκλικοί αρωματικοί υδρογονάνθρακες (PAH) κ.λ.π.. Οι ενώσεις αυτές

χαρακτηρίζονται από σημαντική βιολογική δράση και παρουσιάζουν υψηλή τοξικότητα (καρκινογόνες, τερατογόνες κ.λ.π.). (Μαμάης, 2006). Οι τοξικές οργανικές ουσίες καταλήγουν στο νερό λόγω της ευρείας χρήσης τους στη γεωργία και στη βιομηχανία. Οι διοξίνες παράγονται εκεί όπου υπάρχουν καύσεις ή διεργασίες με χλώριο και μεταφέρονται στη συνέχεια στο νερό. Οι διαρροές πετρελαίου αποτελούν επίσης έναν πολύ σπουδαίο ρύπο για τα φυσικά νερά προκαλώντας ρύπανση με χλωριωμένους υδρογονάνθρακες (Νταρακάς, 2009).

Ανάλογα με την προέλευση του φυσικού νερού, αυτό υποβάλλεται σε επεξεργασία ώστε να μην περιέχει μολυσματικούς παράγοντες και να είναι κατάλληλο προς ανθρώπινη χρήση. Η ποιότητα των υπόγειων νερών είναι συνήθως ανώτερη από εκείνη των επιφανειακών ως προς τους μικροοργανισμούς, τη θολότητα και τον Ολικό Οργανικό Άνθρακα. Ένας μεγάλος αριθμός αιωρούμενων σωματιδίων και παθογόνων του νερού που προέρχεται από πηγές ή γεωτρήσεις αφαιρείται λόγω της διήθησής του δια μέσου των υπερκείμενων πετρωμάτων του εδάφους (Γκολφινόπουλος, 2005). Τα βασικά προβλήματα της ποιότητας των υπόγειων νερών είναι η υψηλή σκληρότητα, ο σίδηρος και το μαγγάνιο. Τα υπόγεια νερά συνήθως ικανοποιούν τις υγειονομικές απαιτήσεις και δεν χρειάζονται ιδιαίτερη επεξεργασία πλην ίσως της απολύμανσης. Ωστόσο, σε περίπτωση που ορισμένα ποιοτικά χαρακτηριστικά τους (πχ. σίδηρος, μαγγάνιο, σκληρότητα) παρεκκλίνουν των αποδεκτών ορίων, χρειάζεται κατάλληλη επεξεργασία για τη βελτίωσή τους (Λέκκας, 2005). Αντίθετα, το νερό επιφάνειας, δεδομένου ότι είναι εκτεθειμένο σε πολλούς δυνητικούς παράγοντες ρύπανσης, πρέπει να υποβληθεί σε πολύ περισσότερα βήματα καθαρισμού από τα υπόγεια νερά για να γίνει κατάλληλο για πόση. Για το λόγο αυτό ο καθαρισμός του νερού που προέρχεται από τα νερά επιφάνειας είναι πιο δαπανηρός. Τα βασικά προβλήματα ποιότητας των επιφανειακών νερών είναι τα αιωρούμενα στερεά, το χρώμα, η γεύση, η οσμή, μικροβιακή μόλυνση και η χημική μόλυνση (Λέκκας, 2005).

1.2. ΝΟΜΟΘΕΤΙΚΟ ΠΛΑΙΣΙΟ

Οι πρώτες νομοθετικές ρυθμίσεις της ΕΕ για το νερό ανάγονται στην δεκαετία του 1970 όταν εισήχθησαν οδηγίες που αφορούσαν κυρίως στο πόσιμο νερό, στα υπόγεια

νερά και στις επικίνδυνες ουσίες. Στη δεκαετία του 1990 η νομοθεσία αυτή συμπληρώθηκε από ένα δεύτερο κύμα οδηγιών που αφορούσαν στα αστικά λύματα, στη ρύπανση από νιτρικά και στη βιομηχανική ρύπανση, και βελτιώθηκε με αναθεωρήσεις των οδηγιών για το πόσιμο νερό και τα νερά για την κολύμβηση. Η Οδηγία-Πλαίσιο 2000/60/EK θέτει ένα πιο ορθολογικό πλαίσιο στην πολιτική της Κοινότητας για το νερό, εισάγοντας την έννοια του ολοκληρωμένου συστήματος διαχείρισης.

Σύμφωνα με την οδηγία 2000/60/EK, «για τη θέσπιση πλαισίου κοινοτικής δράσης στον τομέα της πολιτικής των υδάτων» του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 23^{ης} Οκτωβρίου 2000, κάθε ταμιευτήρας πρέπει να υπόκειται σε κατάλληλη διαχείριση κατόπιν παρακολούθησης των σημείων απόληξης ύδατος και των πιθανών πηγών επιμόλυνσής του. Οι ρύποι που πρέπει να προσδιορίζονται στα σημεία αυτά χωρίζονται στις ουσίες προτεραιότητας και σε λοιπούς ρύπους. Ο αριθμός των δειγμάτων που πρέπει να λαμβάνονται ανά σημείο ετησίως εξαρτάται άμεσα από τον αριθμό των καταναλωτών που εξυπηρετούνται.

Η οδηγία του Συμβουλίου των Ευρωπαϊκών Κοινοτήτων 75/440/ΕΟΚ, ορίζει σχήματα δειγματοληψίας και αναλύσεις που πρέπει να εφαρμόζονται σε επιφανειακά νερά που προορίζονται για ανθρώπινη κατανάλωση κατόπιν κατάλληλης επεξεργασίας.

Οι χώρες μέλη της Ευρωπαϊκής Ένωσης οφείλουν να λαμβάνουν όλα τα αναγκαία μέτρα ώστε να εξασφαλίσουν ότι παρακολουθείται τακτικά η ποιότητα του νερού ανθρώπινης κατανάλωσης, προκειμένου να ελέγχεται αν το διατιθέμενο στους καταναλωτές νερό πληροί τις απαιτήσεις της οδηγίας 98/83/EK του Συμβουλίου της Ευρωπαϊκής Ένωσης της 3^{ης} Νοεμβρίου 1998 για την «Ποιότητα του νερού ανθρώπινης κατανάλωσης», και ιδίως τις παραμετρικές τιμές που τίθενται στην συγκεκριμένη οδηγία.

1.4. ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥ ΝΕΡΟΥ

Η μεθοδολογία επεξεργασίας του νερού, αποσκοπεί στο να γίνει το νερό ένα άριστο προϊόν προς ανθρώπινη κατανάλωση, διαυγές, υγιεινό, με ευχάριστη γεύση,

μικροβιολογικά ασφαλές και απαλλαγμένο από κάθε ουσία που θα μπορούσε να επηρεάσει την ανθρώπινη υγεία. Το είδος της επεξεργασίας, που απαιτείται, διαφοροποιείται ανάλογα με τα ποιοτικά χαρακτηριστικά του νερού στην πηγή υδροδότησης.

Βάσει της οδηγία 75/440/ΕΟΚ, τα γλυκά επιφανειακά ύδατα κατατάσσονται σε τρεις κατηγορίες Α1, Α2, Α3, σύμφωνα με τα ποιοτικά τους χαρακτηριστικά, οι οποίες αντιστοιχούν σε κατάλληλες διαδικασίες επεξεργασίας. Ειδικότερα:

- **Κατηγορία Α1:** Απλή φυσική επεξεργασία και απολύμανση ιδίως με ταχεία διήθηση και απολύμανση.
- **Κατηγορία Α2:** Κανονική φυσική και χημική επεξεργασία και απολύμανση (ιδίως με προχλωρίωση, συσσωμάτωση, κροκίδωση, καθίζηση, διήθηση, και απολύμανση).
- **Κατηγορία Α3:** Εντατική φυσική και χημική επεξεργασία και προχωρημένη απολύμανση ιδίως με χλωρίωση μέχρι σημείου θραύσεως, συσσωμάτωση, κροκίδωση, καθίζηση, διήθηση, προσρόφηση (ενεργός άνθρακας), και απολύμανση (όζον, τελική χλωρίωση).

Οι κλασσικές μέθοδοι επεξεργασίας του νερού περιλαμβάνουν συνήθως τα παρακάτω στάδια: Κροκίδωση – συσσωμάτωση, καθίζηση, διήθηση και απολύμανση. Σε ορισμένες περιπτώσεις απαιτείται και η απομάκρυνση ορισμένων χημικών στοιχείων (πχ. σιδήρου, μαγγανίου).

Κύριος σκοπός της απολύμανσης του νερού είναι η αποτροπή της εξάπλωσης παθογόνων που σχηματίζονται στο νερό. Η χλωρίωση αποτέλεσε την πρώτη μέθοδο που χρησιμοποιήθηκε για την απολύμανση τόσο των υγρών αποβλήτων, όσο και του πόσιμου νερού. Ξεκίνησε από το New Jersey των Η.Π.Α το 1908 και εξακολουθεί να είναι η πιο διαδεδομένη μέθοδος απολύμανσης του νερού (Rook, 1974). Ωστόσο, τα πρώτα προβλήματα της χλωρίωσης ανέκυψαν στις αρχές της δεκαετίας του '40 στις Η.Π.Α. (δυσάρεστη γεύση και οσμή). Επιπλέον, σε συστήματα ύδρευσης στα οποία χρησιμοποιείται η τεχνική της χλωρίωσης και ειδικότερα σε εκείνα στα οποία η τροφοδοσία γίνεται από επιφανειακά νερά, ανιχνεύονται ως παραπροϊόντα τα τριαλογονομένα παράγωγα μεθανίου (Trihalomethanes-THMs), τα οποία έχουν

κατηγορηθεί για καρκινογέννηση (Morris και λοιποί, 1992; Vogt και Regli, 1981; Means και Krasner, 1993). Τα THMs είναι η πιο σημαντική κατηγορία μιας ευρύτερης ομάδας τοξικών οργανικών ενώσεων η οποία συμπεριλαμβάνει τα αλογονομένα οξικά οξέα, τα αλογονομένα ακετονιτρίλια, τη χλωροπικρίνη, τον υδρίτη της χλωράλης και τις αλογονομένες κετόνες. Για την ελάττωση των επιπτώσεων των παραπροϊόντων απολύμανσης και την αντιμετώπιση θεμάτων που σχετίζονται με την ασφαλή διαχείριση των εγκαταστάσεων απολύμανσης, τη μείωση του λειτουργικού κόστους και φυσικά τη βελτίωση της απολυμαντικής δράσης, αναπτύχθηκαν, παράλληλα και με τις τεχνολογικές εξελίξεις, νέες μέθοδοι απολύμανσης, ή και παραλλαγές της χλωρίωσης (διοξειδίου του χλωρίου). Σήμερα οι πιο διαδεδομένες μέθοδοι απολύμανσης είναι:

- Χλωρίωση (αέριο χλώριο, υποχλωριώδες νάτριο, διοξείδιο του χλωρίου)
- Έκθεση σε Υπεριώδη Ακτινοβολία (U.V.)
- Οζόνωση (με παραγωγή όζοντος από ατμοσφαιρικό αέρα ή οξυγόνο).

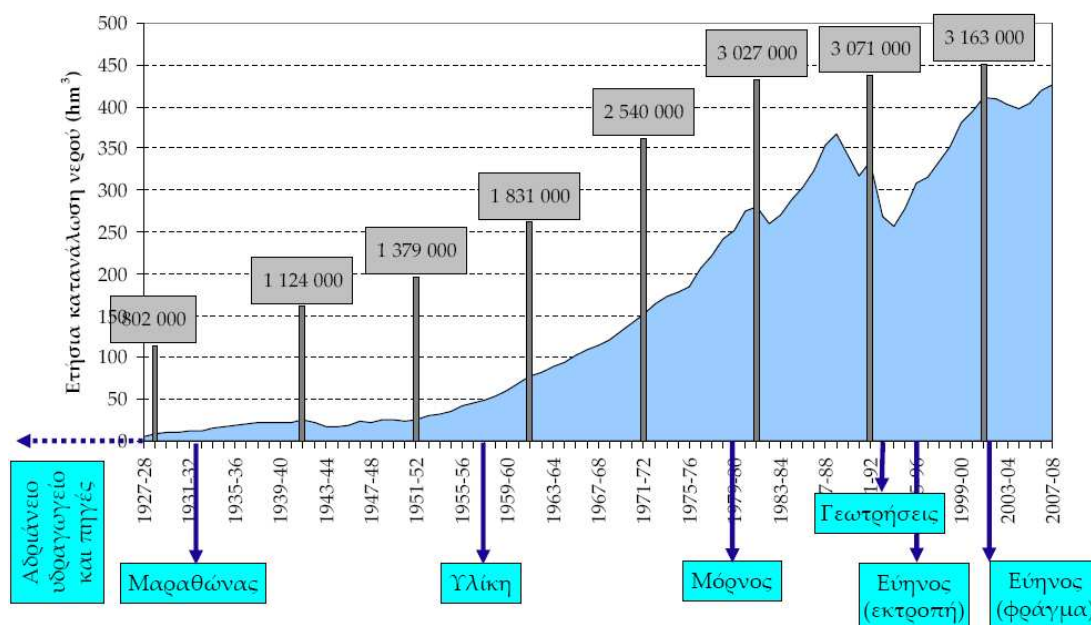
Η αρχή των πολλαπλών φραγμών στη μετάδοση των παθογόνων μικροβίων ορίζει ότι, η αξιοπιστία ενός υδρευτικού συστήματος πρέπει να βασίζεται σε περισσότερες από μία διεργασίες για να είναι εγγυημένη η προστασία από επιδημίες υδάτινης πρέλευσης. Η παραπάνω αρχή αποτελεί την θεμελιώδη αρχή σχεδιασμού των υδρευτικών συστημάτων. Η επεξεργασία των δημοτικών λυμάτων μαζί με την προστασία των φυσικών υδατικών αποθεμάτων από λύματα αποτελούν τους πρώτους φραγμούς. Η απλή απολύμανση δεν παρέχει την επιθυμητή προστασία αν δεν ενισχύεται από κροκίδωση – καθίζηση και διήθηση του πόσιμου νερού, διεργασίες οι οποίες αφαιρούν από το νερό σημαντικότατο αριθμό παθογόνων μικροβίων (Τζουμέρκας, 2002; Γκολφινόπουλος, 2005). Πολλά από αυτά τα μικρόβια είναι ιδιαίτερα ανθεκτικά στην συνήθη απολύμανση με χλώριο. Άλλωστε, ένας μόνο ή και περισσότεροι φραγμοί μπορεί να φανούν αναποτελεσματικοί από τυχαία γεγονότα. Η απολύμανση (συνηθέστερα η χλωρίωση) αποτελεί τον τελευταίο φραγμό και γίνεται συνήθως σε περισσότερα από ένα σημεία των μονάδων επεξεργασίας νερού.

ΚΕΦΑΛΑΙΟ 2

ΤΟ ΥΔΡΕΥΤΙΚΟ ΣΥΣΤΗΜΑ ΤΗΣ ΑΘΗΝΑΣ

2.1. Η ΥΔΡΕΥΣΗ ΤΗΣ ΑΘΗΝΑΣ ΤΗΝ ΣΥΓΧΡΟΝΗ ΕΠΟΧΗ (1926 – ΣΗΜΕΡΑ)

Ο ανησυχία για τις επιπτώσεις της λειψυδρίας, οι διαρκώς αυξανόμενες ανάγκες για πόσιμο νερό και η δημιουργία υποδομών που αρμόζουν στην πρωτεύουσα ενός κράτους, ήταν οι αρχικοί βασικοί λόγοι που επέβαλλαν τη δημιουργία ενός σύγχρονου υδρευτικού συστήματος στην περιοχή της πρωτεύουσας, στις αρχές του προηγούμενου αιώνα. Η συνεχόμενη αύξηση του πληθυσμού της Αθήνας θεωρείται ο κύριος παράγοντας αύξησης της ζήτησης και κατανάλωσης νερού στην σύγχρονη εποχή. Επιπλέον, σταδιακά συνετέλεσαν και άλλοι παράγοντες όπως για παράδειγμα ο βαθμός ανάπτυξης άλλων χρήσεων νερού (βιομηχανικές, δημόσιες, δημοτικές), η αύξηση του βιοτικού επιπέδου και η υδροδότηση νέων περιοχών.



Σχήμα 2.1. Εξέλιξη κατανάλωσης νερού καθώς και πληθυσμού - υδρευτικών έργων στο υδρευτικό σύστημα της Αθήνας, από το 1927 έως το 2008.

Πηγή: Ευστρατιάδης, 2008

Στο Σχήμα 2.1 παρουσιάζεται η εξέλιξη της ετήσιας κατανάλωσης νερού στην περιοχή της πρωτεύουσας, σε συνδυασμό με την αύξηση του πληθυσμού και την ένταξη των υδρευτικών έργων στο υδροδοτικό της σύστημα, από το υδρολογικό έτος 1927-28 μέχρι και το υδρολογικό έτος 2007-08.

Η αύξηση του πληθυσμού της Αθήνας, κυρίως μετά την μικρασιατική καταστροφή, δημιουργούσε διαρκώς νέες ανάγκες. Το 1925 άρχισαν να κατασκευάζονται τα πρώτα σύγχρονα έργα ύδρευσης στην περιοχή της Πρωτεύουσας. Τη χρονιά αυτή υπογράφηκε σύμβαση μεταξύ του Ελληνικού Δημοσίου, της Αμερικανικής Εταιρείας ULEN και της Τράπεζας Αθηνών για τη χρηματοδότηση και κατασκευή έργων ύδρευσης της Πρωτεύουσας από τη λεκάνη απορροής της Πάρνηθας. Τα έργα θα επόπτευε κατασκευαστικά η Ανώνυμος Ελληνική Εταιρεία Υδάτων (ΕΕΥ) η οποία συστάθηκε για το σκοπό αυτό. Το πρώτο μεγάλο έργο ήταν η κατασκευή του φράγματος του Μαραθώνα.

Για την κατασκευή του φράγματος -που ξεκίνησε τον Οκτώβριο του 1926 και ολοκληρώθηκε το 1929- εργάστηκαν περίπου 900 άνθρωποι. Για τη μεταφορά του νερού από το Μαραθώνα στην Αθήνα κατασκευάστηκε η σήραγγα του Μπογιατίου μήκους 13,4 χλμ.. Τα δύο παραπάνω έργα μαζί με τη Μονάδα Επεξεργασίας Νερού (ΜΕΝ) Γαλατσίου και το δίκτυο διανομής Αθήνας και Πειραιά αποτελούν την πρώτη οργανωμένη προσπάθεια για την ορθολογική ύδρευση της Αθήνας.

Αργότερα, λόγω της συνεχιζόμενης αύξησης του πληθυσμού της Αθήνας, κρίθηκε αναγκαία η χρησιμοποίηση των νερών της φυσικής λίμνης Υλίκης που βρίσκεται στη Βοιωτία. Η Υλίκη έχει την ιδιαιτερότητα να βρίσκεται σε περιοχή χαμηλού υψομέτρου. Έτσι για να γίνει εφικτή η άντληση του νερού από τη λίμνη λειτουργούν πλωτά και χερσαία αντλιοστάσια. Το κεντρικό αντλιοστάσιο της Υλίκης είναι σήμερα το μεγαλύτερο στην Ευρώπη (ΕΥΔΑΠ, 2010).

Το 1974 οι αρμοδιότητες για την υδροδότηση της Αθήνας που είχε ως τότε η εταιρεία ULEN, μεταβιβάζονται εξ ολοκλήρου στην ΕΕΥ, η οποία γίνεται ο αποκλειστικός πλέον φορέας διαχείρισης της ύδρευσης της πόλης. Το 1980 αλλάζει οριστικά το θεσμικό πλαίσιο που περιχαράκωνε τη διαχείριση της ύδρευσης και της αποχέτευσης της Αθήνας, στα όρια των αρμοδιοτήτων των δύο ανεξάρτητων μεταξύ τους οργανισμών, της ΕΕΥ και του ΟΑΠ (Οργανισμός Αποχέτευσης Πρωτεύουσας)

αντίστοιχα. Οι δύο αυτοί οργανισμοί συγχωνεύονται στον ενιαίο πλέον φορέα διαχείρισης ύδρευσης και αποχέτευσης της Πρωτεύουσας, την Εταιρεία Ύδρευσης και Αποχέτευσης Πρωτεύουσας (ΕΥΔΑΠ).

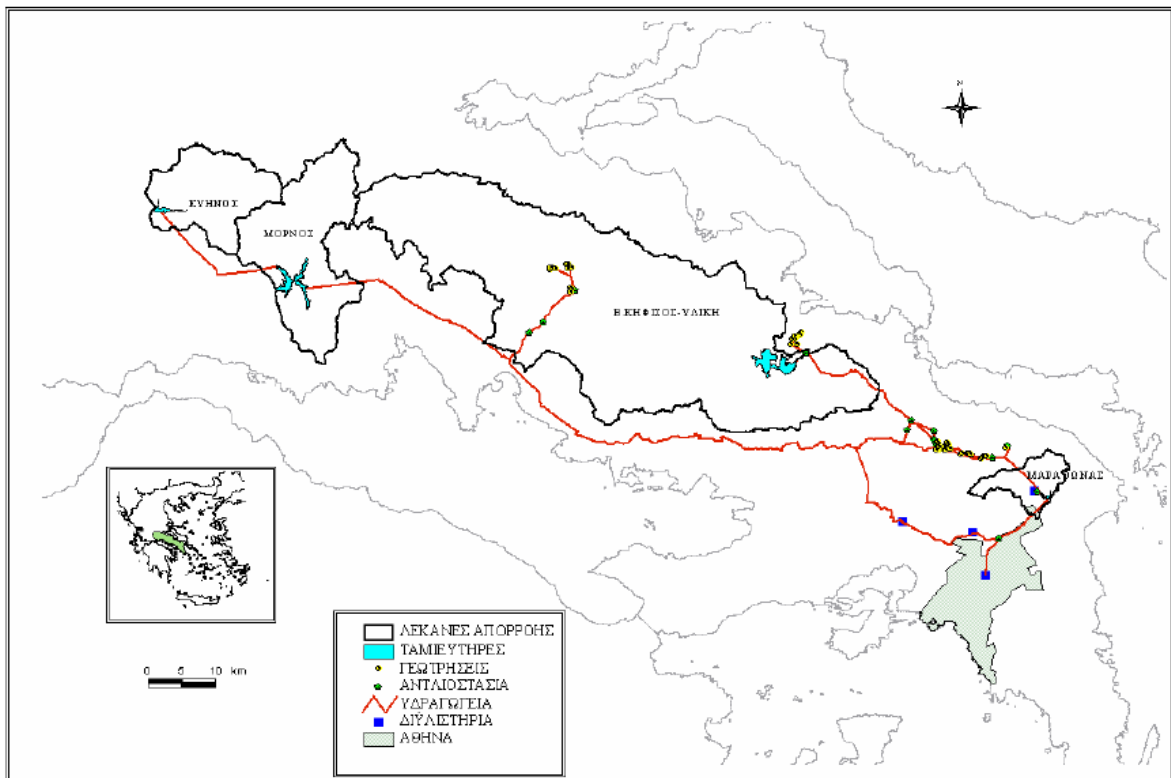
Ένα νέο σημαντικό έργο, το οποίο ενίσχυσε την υδροδότηση της Αθήνας, ήταν το τεχνικό έργο που έγινε στον ποταμό Μόρνο. Το φράγμα που βρίσκεται επί του ποταμού Μόρνου, είναι το ψηλότερο χωμάτινο φράγμα της Ευρώπης, ύψους 126 μ.. Το νερό φτάνει την Αθήνα διαμέσου του υδραγωγείου του Μόρνου, το οποίο είναι το δεύτερο μεγαλύτερο υδραγωγείο στην Ευρώπη.

Ένα άλλο σημαντικό έργο που ενισχύει την υδροδότηση της Αθήνας είναι η εκτροπή του ποταμού Εύηνου προς τον ταμιευτήρα του Μόρνου με την κατασκευή φράγματος και σήραγγας. Η έναρξη των εργασιών στον Εύηνο έγινε το 1992. Η ενωτική σήραγγα προσαγωγής που φέρνει τα νερά του Εύηνου στον ταμιευτήρα του Μόρνου ολοκληρώθηκε σε διάστημα λιγότερο των δύο ετών, γεγονός που αποτελεί παγκόσμιο επίτευγμα για την ολοκλήρωση σήραγγας μεγάλου μήκους.

Τα υδραγωγεία του Μόρνου και της Υλίκης επικοινωνούν μεταξύ τους με ενωτικά υδραγωγεία. Η ύπαρξη των ενωτικών υδραγωγείων επιτρέπει τον έλεγχο, τη συντήρηση και την επισκευή των εγκαταστάσεων των δύο υδραγωγείων με τη δυνατότητα παύσης του ενός από τα δύο.

Μέσω των υδραγωγείων του Μόρνου και της Υλίκης το ακατέργαστο νερό μεταφέρεται στις τέσσερις ΜΕΝ, του Γαλατσίου, του Πολυδενδρίου, των Αχαρνών και του Ασπροπύργου. Στις εγκαταστάσεις των ΜΕΝ το νερό καθαρίζεται μέσω των διαδικασιών της κροκίδωσης, της καθίζησης, της διύλισης και απολυμαίνεται με την προσθήκη χλωρίου που το απαλλάσσει από μικρόβια και μικροοργανισμούς. Στη συνέχεια το νερό διοχετεύεται από τις ΜΕΝ στις δεξαμενές πόλεως, οι οποίες βρίσκονται διασπαρμένες σε διάφορα σημεία της πόλης, και σήμερα ανέρχονται σε 45. Από τις δεξαμενές το νερό διανέμεται στους καταναλωτές μέσα από ένα εκτενές δίκτυο σωληνώσεων μήκους 8078 εκατομμυρίων μέτρων, το οποίο συνεχώς αναβαθμίζεται και επεκτείνεται.

Στο Σχήμα 2.2 παρουσιάζεται η γενική διάταξη του υδροδοτικού συστήματος της Αθήνας.



Σχήμα 2.2. Γενική διάταξη του υδροδοτικού συστήματος της Αθήνας.

Πηγή: Κουτσογιάννης και λοιποί, 2000

2.2. ΤΑΜΙΕΥΤΗΡΕΣ – ΛΙΜΝΕΣ

Οι τέσσερις ταμιευτήρες που αποτελούν το υδροδοτικό σύστημα της Αθήνας είναι ο ταμιευτήρας του Μαραθώνα, η φυσική λίμνη Υλίκη, ο ταμιευτήρας του Μόρνου και ο ταμιευτήρας του Εύηνου.

2.2.1. Ταμιευτήρας Μαραθώνα

Ο Ταμιευτήρας του Μαραθώνα δημιουργήθηκε με την κατασκευή Φράγματος αμέσως κατάντη της συμβολής των ρεμάτων του Χάραδρου και του Βαρνάβα. Ο ποταμός Χάραδρος πηγάζει από τις ανατολικές απολήξεις της Πάρνηθας, διασχίζει την Αττική και εκβάλλει στον όρμο του Μαραθώνα. Η λεκάνη απορροής του ορίζεται από τα όρη Πάρνηθα και Πεντέλη και έχει έκταση 232,6 km². Το υδρογραφικό του

δίκτυο είναι πολύπλοκο, ιδιαίτερα στον άνω ρου, όπου δέχεται μεγάλο πλήθος από χείμαρρους. Στο μέσο ρου του ποταμού έχει κατασκευαστεί ο ταμιευτήρας του Μαραθώνα με σκοπό την ύδρευση της Αθήνας. Οι εργασίες για την κατασκευή του Φράγματος άρχισαν το 1925 και ολοκληρώθηκαν το 1929. Το Φράγμα είναι τοξωτό, βαρύτητας από σκυρόδεμα και είναι εξ ολοκλήρου επενδεδυμένο από Πεντελικό μάρμαρο, γεγονός που του προσδίδει μοναδικότητα σε παγκόσμιο επίπεδο.

Τα χαρακτηριστικά του ταμιευτήρα Μαραθώνα είναι (ΕΥΔΑΠ, 2010):

- Επιφάνεια στη στάθμη υπερχειλίσης: 2,45 τετραγωνικά χιλιόμετρα
- Μέγιστο βάθος: 54 μ.
- Λεκάνη απορροής: 118 τετραγωνικά χιλιόμετρα (ανάντη του φράγματος)
- Μέση απορροή: 14,4 εκατ. μ^3 /έτος
- Μέση βροχόπτωση: 580mm/έτος (τυπική απόκλιση 135mm)
- Μέση απόληψη: 12 εκατ. μ^3 /έτος
- Μέγιστη χωρητικότητα: 41 μ^3 /έτος
- Μέγιστος ωφέλιμος όγκος: 34 εκατ. μ^3

Το φράγμα του Μαραθώνα βρίσκεται σε απόσταση 25 χιλιομέτρων από την πόλη των Αθηνών και σε υψόμετρο 224 m, σε μια περιοχή με μεγάλο ποσοστό δασικών εκτάσεων και ανεξέλεγκτων βοσκοτόπων. Η λεκάνη απορροής του Μαραθώνα ανάντη του ταμιευτήρα περιλαμβάνει τον Άγιο Στέφανο, τις Αφίδνες, το Βαρνάβα, το Καπανδρίτι και το Πολυδένδρι. Σύμφωνα με στοιχεία της Εθνικής Στατιστικής Υπηρεσίας (ΕΣΥΕ), από την απογραφή του 2001, ο πληθυσμός των περιοχών στην περιοχή μελέτης του Μαραθώνα, ανέρχεται σε 17337 κατοίκους. Το μεγαλύτερο ποσοστό της έκτασης της περιοχής καταλαμβάνεται από δασικές εκτάσεις και ανεξέλεγκτους βοσκοτόπους (Λέκκας, 2002). Η κτηνοτροφία στην περιοχή είναι περιορισμένη, αλλά υπάρχουν κτηνοτροφικές μονάδες σε μικρή απόσταση από τα όρια του ταμιευτήρα (εντός της περιοχής του 1 km). Οι υπάρχουσες κτηνοτροφικές μονάδες βρίσκονται σε περιοχές πολύ κοντά στον ταμιευτήρα και αποτελούν μία από τις κύριες πηγές ρύπανσης λόγω των υγρών και στερεών αποβλήτων τους.

Επιπρόσθετες πηγές ρύπανσης αποτελούν οι μικροί ανεξέλεγκτοι χώροι διάθεσης απορριμμάτων και τα όμβρια ύδατα των παραλίμνιων οικισμών και κυρίως του Αγίου Στεφάνου. Κίνδυνος ρύπανσης υπάρχει κι από ουσίες που μπορεί να διοχετευτούν στο περιβάλλον μετά από ατύχημα στην Εθνική οδό η οποία διασχίζει την περιοχή ή στο φράγμα του Μαραθώνα καθώς ο δρόμος περνάει πάνω από αυτό (Λέκκας, 2002). Ο Ταμιευτήρας Μαραθώνα χρησιμοποιείται κυρίως για την αποθήκευση νερού για λόγους ασφαλείας λόγω της εγγύτητας του με την Αθήνα, τροφοδοτούμενος από τον Ταμιευτήρα Υλίκης καθώς και από τον Ταμιευτήρα Μόρνου μέσω του υδραγωγείου Υλίκης και του ενωτικού υδραγωγείου Μόρνου - Υλίκης.

2.2.2. Λίμνη Υλίκη

Η φυσική λίμνη Υλίκη του Νομού Βοιωτίας εντάχθηκε στο υδροδοτικό σύστημα το 1956 προκειμένου να καλυφθούν οι ανάγκες κατανάλωσης λόγω του αλματώδους ρυθμού ανάπτυξης και της αύξησης του πληθυσμού του Λεκανοπεδίου. Με την αποστράγγιση της λίμνης Κωπαΐδας, η Υλίκη αποτέλεσε τον κύριο αποδέκτη των επιφανειακών υδάτων της ευρύτερης περιοχής μέσω του ποταμού Βοιωτικού Κηφισού, ο οποίος καταλήγει με τεχνητή διαμόρφωση και σήραγγα στο ΒΔ τμήμα της Υλίκης αφού διασχίζει την πεδιάδα της Κωπαΐδας, η οποία καλλιεργείται εντατικά. Μέσω της διώρυγας Καρδίτσας παροχετεύεται το σύνολο των απορροών του Β. Κηφισού στη λίμνη της Υλίκης (αποτελούν πάνω από το 90% των συνολικών εισροών στη λίμνη).

Τα χαρακτηριστικά της λίμνης Υλίκη είναι (ΕΥΔΑΠ, 2010):

- Επιφάνεια στη στάθμη υπερχειλίσης: 28 τετραγ. χιλιόμετρα
- Λεκάνη απορροής: 2.432 τετραγ. χιλιόμετρα
- Μέση απορροή: 295 εκατ. μ³ / έτος
- Στάθμη υπερχειλίσης: 79,80 μ.υ.θ.
- Μέγιστη χωρητικότητα: 600 εκατ. μ³
- Μέγιστος ωφέλιμος όγκος: 580 εκατ. μ³
- Κατωτάτη στάθμη υδροληψίας: + 43,50 μ.υ.θ.

Η λίμνη Υλίκη βρίσκεται σε απόσταση 90 χιλιομέτρων από την πόλη των Αθηνών και σε υψόμετρο 78 m. Η ευρύτερη περιοχή που επηρεάζει την ποιότητα των υδάτων της Υλίκης περιλαμβάνει τα αστικά κέντρα της Θήβας και της Λιβαδειάς. Σύμφωνα με την απογραφή του 2001, ο πληθυσμός της περιοχής μελέτης της Υλίκης ανέρχεται σε 92632 κατοίκους. Το μεγαλύτερο τμήμα της περιοχής καλύπτεται από εκτάσεις καλλιεργούμενης γης και θαμνωδών εκτάσεων που χρησιμοποιούνται ως βοσκότοποι (Λέκκας, 2002).

Η εντατική καλλιέργεια έχει ως αποτέλεσμα την εφαρμογή μεγάλων ποσοτήτων λιπασμάτων και ζιζανιοκτόνων στις καλλιεργούμενες εκτάσεις και συνεπώς την γεωργική ρύπανση της περιοχής. Σύμφωνα με μελέτη για την προστασία του περιβάλλοντος που έγινε το 1987 (Ανωμερίτης και λοιποί, 1987), μετά από εκτίμηση των ρυπαντικών φορτίων, (ποσότητες λιπασμάτων και ζιζανιοκτόνων που εφαρμόζονταν ανά καλλιεργούμενη έκταση) δεν υφίσταται άμεσος κίνδυνος ευτροφισμού ή συγκεντρώσεις τοξικών γεωργικών φαρμάκων στη περιοχή. Σε αναλύσεις που έγιναν στα νερά της Υλίκης την περίοδο 1998-2000 από το Εργαστήριο Ποιότητας Υδάτων και Αέρα του Τμήματος Περιβάλλοντος, ανιχνεύθηκαν ποσότητες φυτοφαρμάκων και ζιζανιοκτόνων (Λέκκας, 1998;1999;2000). Κατά τους ελέγχους της ΕΥΔΑΠ, οι συγκεντρώσεις των φυτοπροστατευτικών ουσιών σε δείγματα νερού της Υλίκης είναι συνήθως κάτω από τα όρια ανίχνευσης ενώ σπάνια ανιχνεύονται σε εξαιρετικά χαμηλές τιμές (Πεπραγμένα Υπηρεσίας Ελέγχου Ποιότητας Ύδατος, 2005, 2006, 2007, 2008, 2009; Λύτρας, 2009). Σύμφωνα με το ισχύον θεσμικό πλαίσιο ΚΥΑ Α5/2280/83 που αφορά επιφανειακά νερά που προορίζονται για την ύδρευση της πρωτεύουσας, απαγορεύεται η χρήση χημικών λιπασμάτων και φυτοφαρμάκων μέσα στην ζώνη 1 km. Η κτηνοτροφία αποτελεί σοβαρή απειλή για ρύπανση των ρεμάτων της περιοχής και της λίμνης Υλίκης (Ανωμερίτης και λοιποί, 1987). Ο ζωικός πληθυσμός στην περιοχή αποτελείται κυρίως από ορνιθοειδή, χοίρους και αιγοπρόβατα. Ο μεγάλος αριθμός ζώων, η ελεύθερη βόσκηση καθώς και οι δραστηριότητες επεξεργασίας των κτηνοτροφικών προϊόντων επηρεάζουν αρνητικά την ποιότητα των υδάτων της λίμνης. Σημαντική έκταση καταλαμβάνουν και οι βιομηχανικές δραστηριότητες κατά μήκος των οδικών αξόνων. Οι περισσότερες βιομηχανικές μονάδες διαθέτουν κάποια μορφή επεξεργασίας των αποβλήτων τους, χωρίς ωστόσο η απομάκρυνση ρυπογόνων

ουσιών να είναι ικανοποιητική. Το κύριο χαρακτηριστικό των αποβλήτων των βιομηχανιών της περιοχής είναι το υψηλό οργανικό φορτίο.

Από τους ταμιευτήρες της ΕΥΔΑΠ μόνο αυτός της Υλίκης είναι φυσικός και μετά την ολοκλήρωση των έργων του Μόρνου, που αποτελεί πλέον την κύρια πηγή υδροδότησης της Αθήνας, χρησιμοποιείται ως βοηθητικός υδατικός πόρος σε επείγουσες ανάγκες και συμπληρωματικά στην αιχμή της ζήτησης κατά τη θερινή περίοδο. Οι κύριοι λόγοι είναι:

- Η λίμνη Υλίκη αποτελεί έναν ασταθή παράγοντα προγραμματισμού για την υδροδότηση της Αθήνας, λόγω των μεγάλων απωλειών νερού από τον πυθμένα της μέσω καταβόθρων.
- Λόγω της χαμηλής υψομετρικής θέσης της Υλίκης, η υδροληψία γίνεται μέσω αντλητικών συγκροτημάτων. Τα αντλιοστάσια καταναλώνουν μεγάλες ποσότητες ηλεκτρικής ενέργειας, γι' αυτό η υδροληψία και η μεταφορά νερού από την Υλίκη στον Μαραθώνα παρουσιάζει υψηλό λειτουργικό κόστος.

2.2.3. Ταμιευτήρας Μόρνου

Στην κοίτη του ποταμού Μόρνου 7 χιλιόμετρα δυτικά του Λιδορικού στο Νομό Φωκίδος δημιουργήθηκε με την κατασκευή χωμάτινου φράγματος ο Ταμιευτήρας του Μόρνου. Το έργο ξεκίνησε να κατασκευάζεται τον Μάιο του 1972, ολοκληρώθηκε το 1979, ενώ η κανονική λειτουργία του άρχισε το 1981. Το φράγμα, ένα από τα μεγαλύτερα της Ευρώπης, αποτελείται από αδιαπέραστο αργιλικό πυρήνα, μεταβατικές ζώνες φίλτρων εκατέρωθεν του πυρήνα, σώματα στήριξης του πυρήνα ανάντη - κατάντη από αμμοχάλικο και προστατευτική λιθορριπή στην ανάντη πλευρά. Ο ανάντη πόδας του Φράγματος είναι διαμορφωμένος σε πρόφραγμα με στεγανό αργιλικό πυρήνα.

Τα χαρακτηριστικά του ταμιευτήρα Μόρνου είναι (ΕΥΔΑΠ, 2010):

- Επιφάνεια στη στάθμη περχειίσης: 19,9 τετραγωνικά χιλιόμετρα
- Λεκάνη απορροής: 560 τετραγωνικά χιλιόμετρα
- Μέσο υψόμετρο λεκάνης απορροής: 1.082 μ.υ.θ.

- Μέση βροχόπτωση: 950 mm/έτος (τυπική απόκλιση 210 mm)
- Μέση απορροή: 235 εκατ. μ^3 /έτος
- Μέση απόληψη: 210 εκατ. μ^3 /έτος
- Μέγιστη χωρητικότητα: 780 εκατ. μ^3
- Μέγιστος ωφέλιμος όγκος: 670 εκατ. μ^3
- Απολήψιμος Όγκος : 113 μ^3 /έτος
- Κανονική στάθμη λειτουργίας πύργου υδροληψίας: +394 μ.υ.θ.
- Ελάχιστη στάθμη Υδροληψίας: + 378 μ.υ.θ.

Η περιοχή κατάκλισης γενικά θεωρείται στεγανή, πλην της κλιτύος του όρους "Πύρνος", λόγω των διαπερατών καρστικών ασβεστολιθικών πετρωμάτων αυτού, όπου και κατασκευάστηκε έργο στεγανοποίησης της επιφάνεια, συνολικού μήκους 2,5 χλμ. και μέσου πλάτους 100 μ.

Η περιοχή άμεσης επιρροής – λεκάνη απορροής του Μόρνου ανάντη του ταμιευτήρα περιλαμβάνει το Δήμο Λιδορικού και 28 κοινότητες. Σύμφωνα με την απογραφή της ΕΣΥΕ του 2001, ο συνολικός πληθυσμός στην περιοχή είναι 4092 άτομα. Η διακύμανση του πληθυσμού είναι έντονη κατά τη διάρκεια του έτους. Συγκεκριμένα, υποδιπλασιάζεται το χειμώνα ενώ το καλοκαίρι τετραπλασιάζεται. Στην περιοχή, μεταξύ των κοινοτήτων Βράιλα και Μαλανδρίνου, λειτουργεί σωφρονιστικό κατάστημα δυναμικότητας 500 ατόμων. Το σωφρονιστικό κατάστημα κείται παραπλεύρως χειμάρρου σε απόσταση μικρότερης των 2 km από τον ταμιευτήρα ενώ δεν διαθέτει μονάδα επεξεργασίας υγρών αποβλήτων. Το μεγαλύτερο ποσοστό της ευρύτερης περιοχής, σύμφωνα με στοιχεία της ΕΣΥΕ, χαρακτηρίζεται ως μικτές γεωργικές / κτηνοτροφικές (Λέκκας, 2002). Ωστόσο, από αυτοψία που έγινε στην περιοχή, φαίνεται ότι οι βοσκότοποι εκτείνονται σε βάρος των καλλιεργούμενων εκτάσεων και οι αμιγείς γεωργικές εκτάσεις είναι περιορισμένες (Λέκκας, 2002). Στην περιοχή υπάρχουν σημαντικός αριθμός ζώων, κυρίως μόσχων, που βόσκουν ελεύθερα σε περιοχές που βρίσκονται πολύ κοντά στον ταμιευτήρα.

Η περιοχή υδρεύεται από πηγές ενώ οι ανάγκες για αποχέτευση καλύπτονται με τη χρήση βόθρων ή με απ' ευθείας διάθεση σε χείμαρρους που καταλήγουν μέσα στον ταμιευτήρα, όπως στην περίπτωση του Λιδορικού (κάτι που δεν είναι σύμφωνο με την ΚΥΑ Α5/2280/83). Είναι αξιοσημείωτο το γεγονός ότι στο Λιδορικό ενώ υπάρχει

μονάδα επεξεργασίας υγρών αποβλήτων, που ανήκει στην ΕΥΔΑΠ, δεν έχει λειτουργήσει ποτέ από το 1998, έτος που κατασκευάστηκε. Από τη Νομαρχιακή Αυτοδιοίκηση Φωκίδας δημοπρατήθηκε το έργο «Μελέτη αποχέτευσης και επεξεργασίας λυμάτων οικισμών περιοχής Μόρνου Δήμου Λιδορικίου», τον Δεκέμβριο του 2009, με συμβατικό χρόνο αποπεράτωσης 10 μήνες.

Σε καμία από τις κοινότητες και τον Δήμο δεν υπάρχει πρόγραμμα ολοκληρωμένης διαχείρισης των απορριμμάτων. Παλαιότερα, η πρακτική που ακολουθείτο, περιελάμβανε κυρίως ανεξέλεγκτη διάθεση, σε έναν ή περισσότερους χώρους διάθεσης εντός των διοικητικών ορίων κάθε κοινότητας, ενώ σε μερικές περιπτώσεις γινόταν μερική διευθέτηση, περιοδική καύση και στοιχειώδης επιχωμάτωση των απορριμμάτων (Παρασκευόπουλος, 1997). Συνολικά υπήρχαν 47 ανεξέλεγκτοι χώροι διάθεσης που βρίσκονταν κυρίως σε ρέματα, πρηνή δρόμων, βοσκοτόπια και δασικές εκτάσεις. Χαρακτηριστικό παράδειγμα είναι ο ένας από τους χώρους διάθεσης των απορριμμάτων του Δήμου Λιδορικίου που γειτνιάζε με τον ταμιευτήρα του Μόρνου. Ο συγκεκριμένος χώρος, καταργήθηκε το 2006, όπως και όλοι οι υπόλοιποι χώροι ανεξέλεγκτης διάθεσης απορριμμάτων (ΧΑΔΑ) του Δήμου Λιδορικίου, και χρησιμοποιείται ένας μόνο ΧΑΔΑ ο οποίος βρίσκεται εκτός της λεκάνης απορροής του Μόρνου. Ήδη, δρομολογείται η παύση λειτουργίας του και η διάθεση των απορριμμάτων στον χώρο υγειονομικής ταφής απορριμμάτων (ΧΥΤΑ) στην Ναύπακτο. Αξίζει να σημειωθεί ότι είχε γίνει μελέτη για την συνολική διαχείριση των απορριμμάτων στην λεκάνη του Μόρνου ανάντη του ταμιευτήρα που προέβλεπε κατασκευή ΧΥΤΑ ανάντη του φράγματος. Η μελέτη αυτή η οποία έγινε για λογαριασμό της ΕΥΔΑΠ δεν εγκρίθηκε (Λέκκας, 2002).

2.2.4. Ταμιευτήρας Εύηνου

Στον ποταμό Εύηνο, στην περιοχή Αγίου Δημητρίου του Νομού Αιτωλοακαρνανίας (Ορεινή Ναυπακτία), κατασκευάστηκε το Φράγμα και η Σήραγγα Ευήνου - Μόρνου. Το έργο του Ευήνου (Φράγμα και Σήραγγα) αποτελεί την 2η φάση του συστήματος υδροδότησης της μείζονος περιοχής της Αθήνας, η οποία προέβλεπε την ενίσχυση του βασικού Ταμιευτήρα του Μόρνου από τη γειτονική λεκάνη του ποταμού Ευήνου. Η έναρξη των εργασιών στον Εύηνο έγινε το 1992 με την κατασκευή χωμάτινου φράγματος το οποίο περατώθηκε τον Ιούνιο του 2001 οπότε και ξεκίνησε η

διαδικασία της πρώτης πλήρωσης του ταμιευτήρα του Φράγματος, η οποία ολοκληρώθηκε τον Οκτώβριο του 2002.

Τα χαρακτηριστικά του ταμιευτήρα Ευήνου είναι (ΕΥΔΑΠ, 2010):

- Επιφάνεια: 3,6 τετραγωνικά χιλιόμετρα
- Λεκάνη απορροής: 352 τετραγωνικά χιλιόμετρα
- Μέση βροχόπτωση: 1.230 mm/έτος (τυπική απόκλιση 260mm)
- Μέση απορροή: 280 εκατ. μ^3 /έτος
- Μέση απόληψη: 200 εκατ. μ^3 /έτος
- Μέγιστη χωρητικότητα: 140 εκατ. μ^3
- Απολήψιμος Όγκος : 113 μ^3 /έτος
- Κανονική στάθμη λειτουργίας πύργου υδροληψίας: 630 εκατ. μ^3
- Απολήψιμος Όγκος : 113 μ^3 /έτος
- Κανονική στάθμη λειτουργίας πύργου υδροληψίας: + 458,00 μ.υ.θ.
- Ελάχιστη στάθμη λειτουργίας Πύργου Υδροληψίας: + 448,60 μ.υ.θ.

Η Σήραγγα Ευήνου - Μόρνου αποτελεί στην ουσία την οδό μεταφοράς των υδάτων του ποταμού Ευήνου, από το ομώνυμο Φράγμα στον Άγιο Δημήτριο Αιτωλοακαρνανίας, στον Ταμιευτήρα του Μόρνου. Η λειτουργία της Σήραγγας είναι υπό πίεση με παροχή 27 μ^3 /δευτ.. Η σήραγγα Ευήνου - Μόρνου έχει συνολικό μήκος 29.393 μέτρα και εσωτερική διάμετρο 3,5 μέτρα. Η διάνοιξη της σήραγγας ξεκίνησε το 1992 και ολοκληρώθηκε σε διάστημα δύο ετών. Από το 1995 οπότε και ολοκληρώθηκε η κατασκευή Σήραγγας Ευήνου - Μόρνου καθώς και η κατασκευή του έργου προσωρινής υδροληψίας, έγινε δυνατή η μεταφορά νερού από τον Εύηνο στον Ταμιευτήρα του Μόρνου πριν την ολοκλήρωση του Φράγματος Ευήνου, της τάξεως των 100 εκατ. μ^3 /έτος, ώστε να αντιμετωπισθεί η ανομβρία που παρατηρείτο κατά το διάστημα αυτό.

Το μέσο υψόμετρο της λεκάνης απορροής του Ευήνου είναι 1.194 μ. Το 80% λεκάνης απορροής του Ευήνου έχει ορεινό χαρακτήρα, με κατά περιοχές δασικές εκτάσεις, απόκρημνες γυμνές ορεινές περιοχές και ανάλογο υδρογραφικό δίκτυο (Λέκκας, 2002). Τα χωριά της περιοχής είναι αραιοκατοικημένα και απομακρυσμένα μεταξύ τους, ενώ οι γεωργικές καλλιέργειες είναι περιορισμένης μορφής και

κατανέμονται γύρω από τα χωριά και τους οικισμούς. Κατά την απογραφή του 2001, στην άμεση περιοχή του ταμιευτήρα Αγ. Δημητρίου (στις κοινότητες Άγ. Δημήτριος, Αράχοβα, Κλέπα, Νεοχώριο) καταγράφηκαν περί τα 632 άτομα. Το μεγαλύτερο τμήμα της περιοχής καλύπτεται από εκτάσεις μη καλλιεργούμενες οι είναι κυρίως δασικές (Λέκκας, 2002).

Η κύρια δραστηριότητα των κατοίκων είναι η γεωργία. Ο ζωικός πληθυσμός στην περιοχή αποτελείται κυρίως από αιγοπρόβατα. Οι υφιστάμενες πηγές ρύπανσης στην περιοχή αφορούν κυρίως στα λύματα και τα απορρίμματα των οικισμών, στα απόβλητα των ποιμνιοστασίων και στα κοπάδια γιδοπροβάτων τα οποία βόσκουν στην περιοχή. Ελάχιστη έως μηδενική είναι η επιβάρυνση από άποψη ρυπογόνων φορτίων που προέρχεται από τις γεωργικές καλλιέργειες (Λέκκας, 2002).

2.3. ΓΕΩΤΡΗΣΕΙΣ

Η ΕΥΔΑΠ Α.Ε. έχει εγκαταστήσει περισσότερες από εκατό γεωτρήσεις που λειτουργούν σε ομάδες και χρησιμοποιούνται σήμερα εφεδρικά. Οι γεωτρήσεις έχουν συνολική ισχύ 25.000 HP και συνολική αντλητική ικανότητα 800.000 μ³/ημέρα. Η ασφαλής τους απόδοση εκτιμάται σε 70 - 125 εκατ.μ³/έτος. Οι βασικότερες ομάδες εκ των κυρίων γεωτρήσεων είναι:

- Β.Α. Πάρνηθας, με 43 γεωτρήσεις, εγκατεστημένη ισχύ 8.340HP και αντλητική ικανότητα 210.000 μ³/ημέρα.
- Μαυροσουβάλας, με 15 γεωτρήσεις, εγκατεστημένη ισχύ 8.400HP και αντλητική ικανότητα 100.000 μ³/ημέρα.
- Ούγγρων, με 10 γεωτρήσεις, εγκατεστημένη ισχύ 1960 HP και αντλητικής ικανότητας 60.000 μ³/ημέρα.
- Βοιωτικού Κηφισού, με 28 γεωτρήσεις, εγκατεστημένη ισχύ 2.350 Hp3 και αντλητικής ικανότητας 260.000 μ³/ημέρα.
- Ν.Δ. Υλίκης, με 14 γεωτρήσεις, εγκατεστημένη 2500HP και αντλητικής ικανότητας 100.000 μ³/ημέρα.

2.4. ΜΕΘΟΔΟΛΟΓΙΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥ ΝΕΡΟΥ

Το ανεπεξέργαστο νερό, όταν φτάνει στις εγκαταστάσεις ΜΕΝ, περιέχει όλες τις συνηθισμένες για ένα επιφανειακό νερό ουσίες, όπως δύσσομα αέρια, μικρόβια, αιωρούμενα σωματίδια, κολλοειδή σωματίδια, άλγες, πρωτόζωα, αδρανή στερεά κ.λ.π.. Τα ανεπεξέργαστα νερά που χρησιμοποιεί η ΕΥΔΑΠ χαρακτηρίζονται ως νερά κατηγορίας Α2, δηλαδή νερά για τα οποία η επεξεργασία τους περιλαμβάνει συνήθη φυσική και χημική επεξεργασία (προχλωρίωση, συσσωμάτωση, κροκίδωση, καθίζηση, διήθηση και απολύμανση). Η βασική επεξεργασία του νερού περιλαμβάνει τις ακόλουθες διαδικασίες ΕΥΔΑΠ, 2010; Λέκκας, 2005):

1. Προεπεξεργασία – Αφαίρεση φερτών υλικών μέσω εσχάρωσης.

Στο στάδιο αυτό επιτυγχάνεται η αφαίρεση φερτών υλικών (κλαδιά, φύλλα, υδρόβια φυτά κ.λ.π.) τα οποία παρασύρονται από το νερό. Τα υλικά αυτά κατακρατούνται σε εσχάρες με διάκενο ράβδων περίπου 15 mm.

2. Προσθήκη Υπερμαγγανικού Καλίου.

Η προσθήκη του υπερμαγγανικού καλίου γίνεται για να οξειδωθούν ορισμένες ουσίες, που προκαλούν δυσάρεστη οσμή και γεύση στο παραγόμενο πόσιμο νερό. Η προσθήκη δεν γίνεται σε μόνιμη βάση αλλά μόνο σε περιόδους αύξησης ορισμένων ρύπων στο ανεπεξέργαστο νερό. Το υπερμαγγανικό κάλιο αντιδρά με τις οξειδώσιμες ουσίες και μετατρέπεται σε στερεό διοξείδιο του μαγγανίου, που αφαιρείται από το νερό με την καθίζηση και τη διήθηση.

3. Προχλωρίωση.

Το χλώριο διαθέτει οξειδωτική και απολυμαντική ικανότητα. Οξειδώνει τις ανόργανες και οργανικές ουσίες (σίδηρο, υδρόθειο, αμμωνία, φυτική οργανική ύλη) και αδρανοποιεί ή θανατώνει τους μικροοργανισμούς (βακτηρίδια, φύκη, πρωτόζωα, ιούς κ.λ.π.). Η μικρή περίσσεια χλωρίου, η οποία παραμένει στο νερό, εμποδίζει την ανάπτυξη μικροοργανισμών στα επόμενα τμήματα της εγκατάστασης επεξεργασίας.

4. Προσθήκη θειικού αργιλίου και ανιονικού πολυηλεκτρολύτη

Η προσθήκη του θειικού αργιλίου γίνεται για την αφαίρεση των κολλοειδών σωματιδίων του ανεπεξέργαστου νερού και τη διαύγασή του. Το θειικό αργίλιο εξουδετερώνει τα αρνητικά φορτία των κολλοειδών σωματιδίων, αντιδρά χημικά με τα δισσάνθρακικά άλατα του νερού και παράγει ίζημα υδροξειδίου του

αργιλίου. Τα κολλοειδή παύουν πλέον να αιωρούνται (φαινόμενο θρόμβωσης) και συσσωματώνονται σχηματίζοντας με το υδροξείδιο του αργιλίου μεγαλύτερου όγκου σωματίδια (φαινόμενο κροκίδωσης). Οι παραγόμενες έτσι κροκίδες είναι πυκνότερες από το νερό (μεγαλύτερο ειδικό βάρος), καθιζάνουν ταχύτερα και αφαιρούνται πιο εύκολα με την καθίζηση και τη διήθηση, που επακολουθούν. Η ποσότητα του θεικού αργιλίου που χρησιμοποιείται είναι μεταβλητή. Εξαρτάται από την ποιότητα του προς επεξεργασία νερού και προσδιορίζεται με δοκιμή κροκίδωσης – καθίζησης. Ο ανιονικός πολυηλεκτρολύτης λειτουργεί σαν βοηθητικό, ενισχύοντας τη δράση του θεικού αργιλίου. Η χρήση του δημιουργεί μεγάλες και βαρύτερες κροκίδες που καθιζάνουν πολύ εύκολα. Έτσι φτάνει στη διήθηση νερό με μικρή θολότητα. Η προσθήκη του πολυηλεκτρολύτη γίνεται μόνο σε ειδικές περιπτώσεις και όταν η προσθήκη του θεικού αργιλίου δεν επιτυγχάνει την επιθυμητή κροκίδωση (χαμηλές θερμοκρασίες το χειμώνα, μεγάλη περιεκτικότητα σε φύκη το καλοκαίρι και το φθινόπωρο, μεγάλη θολότητα από αδρανή μετά από βροχές κ.λ.π.). Η ακριβής ποσότητα του κροκιδωτικού που πρέπει να προστεθεί, προσδιορίζεται με δοκιμή κροκίδωσης – καθίζησης, μαζί με θεικό αργίλιο και δεν ξεπερνά το 1 mg/lit. Τα στάδια της κροκίδωσης και καθίζησης, αποτελούν την προετοιμασία για το κατώτερο στάδιο που είναι η διήθηση. Στο στάδιο της κροκίδωσης επιδιώκεται η συνεχής και ήπια κίνηση του νερού, με τη βοήθεια χωρισμάτων ή αναδευτήρων, με σκοπό την αύξηση του αριθμού των συγκρούσεων μεταξύ των κροκίδων και την συσσωμάτωσή τους. Αντίθετα στο στάδιο της καθίζησης επιδιώκεται η εξαιρετικά ήρεμη οριζόντια ή ανοδική ροή του νερού, ώστε να επιτυγχάνεται η καθίζηση πολύ μεγάλου ποσοστού των κροκίδων. Οι κροκίδες, μετά την καθίζησή τους, έχουν τη μορφή ιζήματος (ιλύος) υδροξειδίου του αργιλίου, που αφαιρείται περιοδικά από τον πυθμένα των δεξαμενών. Με αυτόν τον τρόπο το νερό καθαρίζεται σε ποσοστό 80%.

5. Διήθηση του νερού.

Το τελευταίο στάδιο της διήθησης θεωρείται πολύ σημαντικό για την επεξεργασία νερού επιφανειακής προέλευσης. Στο στάδιο αυτό, το νερό περνά μέσα από ειδικά αμμόφιλτρα τα οποία κατακρατούν τα πολύ μικρά και ελαφρά σωματίδια που δεν έγιναν κροκίδες. Έτσι το νερό εξάγεται πια καθαρό για να δοθεί στην κατανάλωση. Η διήθηση του νερού πραγματοποιείται από το σύνολο των κόκκων του διηθητικού υλικού, όταν συσσωρευτεί ποσότητα κροκίδων η οποία αποτελεί

το μέσο συγκράτησης των κολλοειδών και αιωρούμενων σωματιδίων. Η δεξαμενή κροκίδωσης – καθίζησης τροφοδοτεί μια συστοιχία μονοστρωματικών φίλτρων. Η ρύθμιση της παροχής του νερού επιτυγχάνεται με τη διατήρηση σταθερής στάθμης στη δεξαμενή καθενός φίλτρου. Το διηθητικό στρώμα των φίλτρων αποτελείται από χαλαζιακή άμμο με μέγεθος κόκκων από 0,8 mm έως 1,4mm. Η υπόβαση του φίλτρου απέχει από τον πυθμένα της δεξαμενής και συνήθως αποτελείται από ψευδοπυθμένα από πλάκες σκυροδέματος, οι οποίες έχουν στη βάση τους διαμορφωμένο σύστημα τοιχίων από σκυρόδεμα με οπές για την ομοιόμορφη διανομή του νερού και του αέρα. Το σύστημα αυτό χρησιμοποιείται για τον καθαρισμό των φίλτρων με ανάστροφη ροή νερού και αέρα, όταν το στρώμα της άμμου κορεσθεί και δυσκολεύεται η ροή του νερού μέσα απ' αυτό.

6. Μεταχλωρίωση.

Εφόσον η προχλωρίωση δεν είναι ικανοποιητική, προστίθεται συμπληρωματικό χλώριο κατά την έξοδο του νερού από τις κλειστές δεξαμενές αποθήκευσης και πριν την είσοδό του στο δίκτυο ύδρευσης. Για την ακρίβεια, γίνεται έγχυση χλωριωμένου νερού, ώστε να διατηρείται υπολειμματικό χλώριο στο δίκτυο διανομής, με στόχο την αποφυγή ανάπτυξης μικροβίων κατά την πορεία του καθαρού νερού στον καταναλωτή.

2.4. ΜΟΝΑΔΕΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΝΕΡΟΥ

Η ΕΥΔΑΠ είναι υπεύθυνη για την παροχή πόσιμου νερού σε περίπου 4.000.000 καταναλωτές στην περιφέρεια της Αττικής. Για την παραγωγή του πόσιμου νερού, η εταιρεία διαθέτει τέσσερις ΜΕΝ οι οποίες βρίσκονται στο Γαλάτσι, τις Αχαρνές, το Πολυδένδρι και τον Ασπρόπυργο. Οι μονάδες επεξεργασίας Γαλατσίου, Αχαρνών και Πολυδενδρίου προμηθεύουν νερό τους δήμους της Αθήνας και του Πειραιά, τις υψηλότερες περιοχές της Αττικής, τις ανατολικές ακτές της Αττικής, τις νότιες ακτές της Αττικής καθώς και τα βόρεια προάστια της Αθήνας. Η μονάδα επεξεργασίας νερού του Γαλατσίου έχει ημερήσια συνολική ικανότητα επεξεργασίας 540.000 m³ νερού. Η τρέχουσα συνολική ικανότητα επεξεργασίας νερού στη μονάδα επεξεργασίας των Αχαρνών ανέρχεται σε 800.000 m³/ημέρα. Η μονάδα επεξεργασίας του Πολυδενδρίου έχει συνολική ημερήσια ικανότητα επεξεργασίας 200.000 m³ νερού ενώ κατά τις περιόδους υψηλής ζήτησης φτάνει τις 300.000 m³. Η μονάδα

επεξεργασίας νερού του Ασπροπύργου έχει συνολική ημερήσια ικανότητα επεξεργασίας 200.000 m³ νερού και προμηθεύει νερό στη δυτική Αττική (Θριάσιο Πεδίο) και το νησί της Σαλαμίνας.

2.6. ΑΝΑΛΥΤΙΚΟΣ ΕΛΕΓΧΟΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΝΕΡΟΥ

Ο έλεγχος των ποιοτικών χαρακτηριστικών του νερού είναι απαραίτητος για τη διασφάλιση της συμμόρφωσης προς την ισχύουσα νομοθεσία αλλά και την παρακολούθηση της διαχρονικής ποιότητάς του. Ο αναλυτικός έλεγχος ποιότητας του νερού αφορά όχι μόνο στην επεξεργασμένη μορφή του (πόσιμο νερό) αλλά και την ανεπεξέργαστη πρωταρχική μορφή του. Κύριος στόχος κάθε φορέα ύδρευσης είναι να προμηθεύει τους καταναλωτές με νερό υψηλών προδιαγραφών ποιότητας και να συμμορφώνεται με τη σχετική νομοθεσία. Στη συγκεκριμένη περίπτωση προκειμένου να επιτευχθεί αυτός ο στόχος η ΕΥΔΑΠ διαθέτει εργαστήρια με ανώτατο επίπεδο τεχνικής στο Γαλάτσι, το Πολυδένδρι και τις Αχαρνές. Στα εργαστήρια αυτά πραγματοποιούνται οι χημικές, μικροβιολογικές και τοξικολογικές αναλύσεις του νερού σε δείγματα που προέρχονται από όλα τα στάδια επεξεργασίας του τηρώντας τις προδιαγραφές που θέτει η νομοθεσία.

Η ΕΥΔΑΠ και οι συναρμόδιες αρχές, οφείλουν να λαμβάνουν όλα τα αναγκαία μέτρα ώστε να εξασφαλίσουν ότι παρακολουθείται τακτικά η ποιότητα του νερού ανθρώπινης κατανάλωσης, προκειμένου να ελέγχεται αν το διατιθέμενο στους καταναλωτές νερό πληροί τις απαιτήσεις της ΚΥΑ Υ2/2600/2001 για την «Ποιότητα του νερού ανθρώπινης κατανάλωσης», σε συμμόρφωση προς την οδηγία 98/83/ΕΚ του Συμβουλίου της Ευρωπαϊκής Ένωσης της 3^{ης} Νοεμβρίου 1998 (ΦΕΚ 892/Β/11.7.2001) και της τροποποίησης της υγειονομικής διάταξής της (ΦΕΚ 630/Β/26.4.2007), και ιδίως τις παραμετρικές τιμές που τίθενται στην παρούσα Απόφαση. Επιπλέον, σύμφωνα με το ΦΕΚ 438/Β/3.7.1988 (Υπουργικές Αποφάσεις και Εγκρίσεις αριθ. οικ. 46399/1352 Απαιτούμενη ποιότητα των επιφανειακών νερών που προορίζονται για «πόσιμα», «κολύμβηση», «διαβίωση ψαριών σε γλυκά νερά» και «καλλιέργεια και αλιεία οστρακοειδών», μέθοδοι μέτρησης, συχνότητα δειγματοληψίας και ανάλυση των επιφανειακών νερών που προορίζονται για πόσιμα, σε συμμόρφωση με τις οδηγίες του Συμβουλίου των Ευρωπαϊκών Κοινοτήτων

75/440/ΕΟΚ, 78/659/ΕΟΚ, 79/923/ΕΟΚ και 79/869/ΕΟΚ), καθορίζεται το σχήμα δειγματοληψίας και οι αναλύσεις που πρέπει να εφαρμόζονται στα επιφανειακά νερά που προορίζονται για ανθρώπινη κατανάλωση κατόπιν κατάλληλης επεξεργασίας.

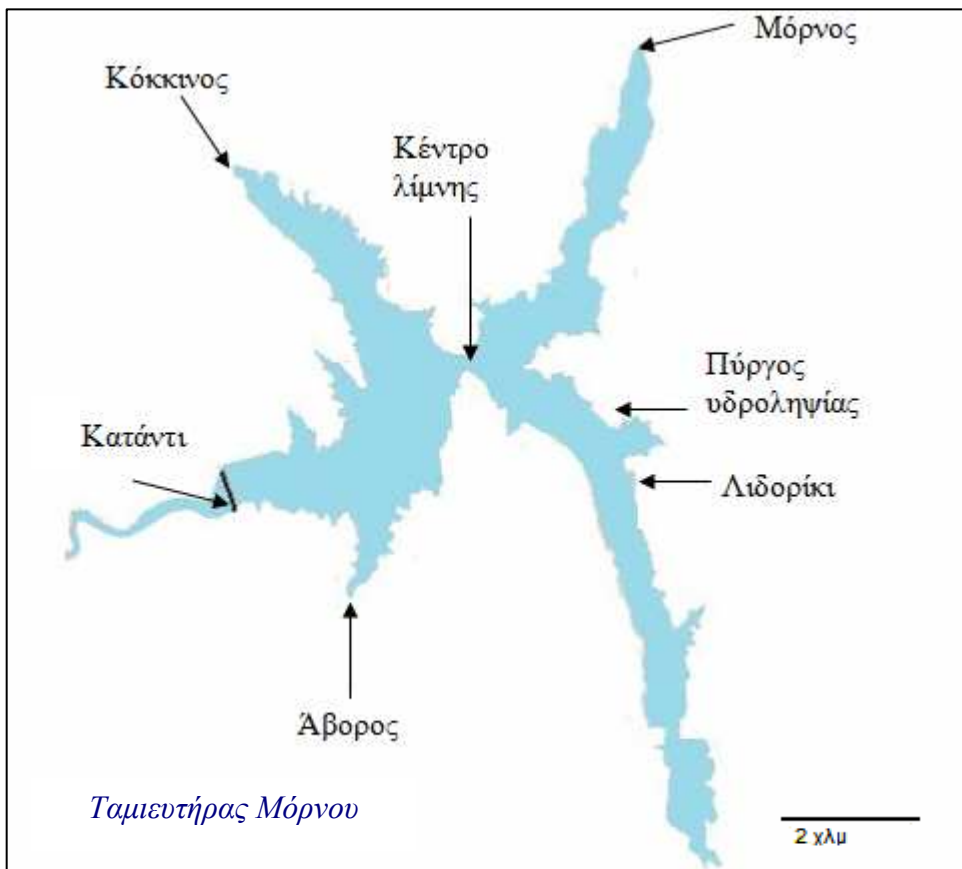
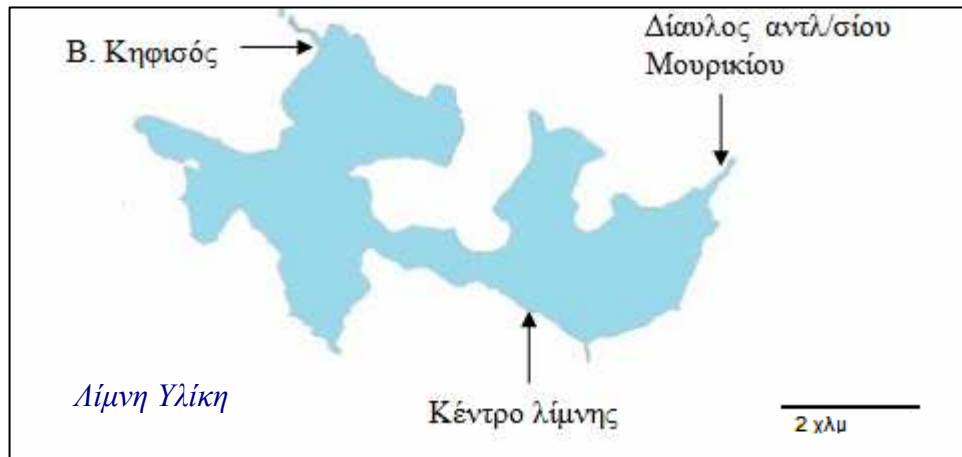
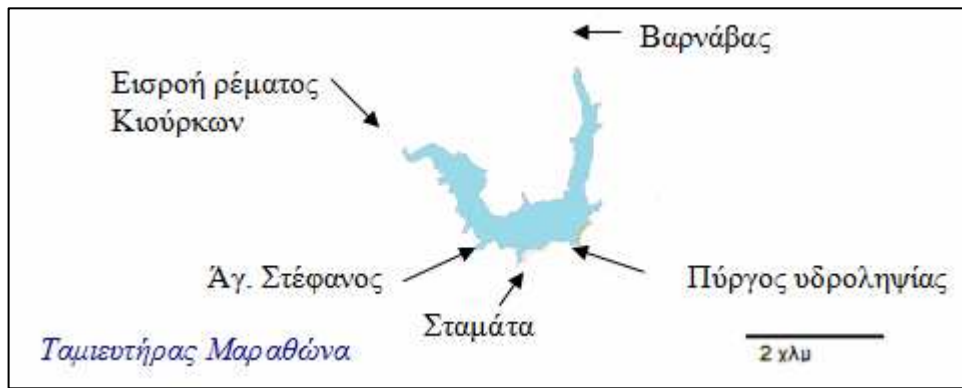
Συμμορφούμενη λοιπόν με την νομοθεσία, η Υπηρεσία Ελέγχου Ποιότητας Ύδατος της ΕΥΔΑΠ διενεργεί ετησίως έναν μεγάλο αριθμό αναλύσεων. Η ποιότητα του νερού ελέγχεται σε όλο το υδρευτικό σύστημα της ΕΥΔΑΠ. Ο έλεγχος επιβάλλεται από την ανάγκη συνεχούς προσαρμογής των ενεργειών των Τεχνικών Υπηρεσιών των ΜΕΝ και του Δικτύου Ύδρευσης της ΕΥΔΑΠ ώστε το παραγόμενο νερό να είναι αποδεδειγμένα πάντοτε σύμφωνο με την ισχύουσα Υγειονομική Νομοθεσία. Ο αναλυτικός έλεγχος ποιότητας του νερού γίνεται εργαστηριακά με χημικούς και μικροβιολογικούς ελέγχους που εκτελούνται περιοδικά από προκαθορισμένα σημεία των ταμιευτήρων, των μονάδων επεξεργασίας νερού και του δικτύου ύδρευσης. Ο έλεγχος αφορά τον προσδιορισμό χημικών ομάδων ή μικροοργανισμών, που είτε είναι επιβλαβή για την ανθρώπινη υγεία είτε είναι ενδεικτικά της ρύπανσης του νερού από λύματα, απόβλητα, φυτοφάρμακα ή άλλες ανεπιθύμητες δραστηριότητες.

Για την απόληψη ύδατος χρησιμοποιούνται τρεις ταμιευτήρες / λίμνες (Μαραθώνα, Υλίκη και Μόρνος), οι οποίοι σύμφωνα με την νομοθεσία πρέπει να υφίστανται κατάλληλη διαχείριση και έλεγχο ώστε να εξασφαλίζεται ότι η ποιότητα του παραλαμβανόμενου προς επεξεργασία νερού είναι κατάλληλη για την μετατροπή του σε πόσιμο ως τελικό προϊόν. Σε κάθε έναν από αυτούς, αντιστοιχεί ένα γεωγραφικό σημείο απόληξης ύδατος (πύργος υδροληψίας). Επίσης ορίζονται γεωγραφικά σημεία εισόδου νερού στις λίμνες τα οποία πρέπει να θεωρούνται πιθανές πηγές επιμόλυνσης και επιβάρυνσης των ταμιευτήρων. Στον πίνακα 1.1 και στο Σχήμα 2.3 παρουσιάζονται τα σημεία δειγματοληψίας για την φυσική λίμνη Υλίκη και για τους ταμιευτήρες Μόρνου και Μαραθώνα. Σε όλα τα σημεία της λίμνης Υλίκη και του ταμιευτήρα Μόρνου η δειγματοληψία γίνεται σε βάθος 20 – 30 cm από την επιφάνεια του νερού και σε απόσταση περίπου ενός μέτρου από την όχθη. Η δειγματοληψία των ποταμών γίνεται κοντά στα σημεία εκβολής τους. Όσον αφορά στα σημεία δειγματοληψίας του ταμιευτήρα Μαραθώνα, το Βεντούρι βρίσκεται στο νότιο τμήμα του ταμιευτήρα Μαραθώνα, κοντά στο φράγμα. Στο σημείο αυτό το νερό συλλέγεται με αντλία σε βάθος 45 μέτρων από την επιφάνεια και προέρχεται από τον υπόγειο υδροφόρο ορίζοντα του ταμιευτήρα, όπου καταλήγουν ύδατα από τον Μόρνο, την

Υλίκη και τους γειτονικούς προς τον ταμιευτήρα υδάτινους φορείς. Στον ποταμό Βαρνάβα και τα ρέματα Κιούρκα, Σταμάτα, Άγ. Στέφανος και Καταρράκτης, η δειγματοληψία γίνεται σε βάθος 20 – 30 cm από την επιφάνεια του νερού και σε απόσταση περίπου ενός μέτρου από την όχθη. Τα ρέματα Σταμάτας και Άγ. Στέφανος διατρέχουν τις ομώνυμες περιοχές και καταλήγουν στο ΝΔ τμήμα του ταμιευτήρα. Η δειγματοληψία γίνεται κοντά στα σημεία εκβολής τους. Τα ρέματα Κιούρκα και Βαρνάβας εκβάλουν αντίστοιχα στο ΒΔ και ΒΑ τμήμα του ταμιευτήρα. Η δειγματοληψία σε αυτά γίνεται σε σημεία που απέχουν περίπου 600 και 500 μέτρα αντίστοιχα από τον ταμιευτήρα. Αξίζει να σημειωθεί ότι μέσω του ρέματος Κιούρκων εισέρχεται στον ταμιευτήρα νερό από τον ταμιευτήρα Μόρνου είτε από την Υλίκη.

Πίνακας 2.1. Σημεία δειγματοληψίας των ταμιευτήρων / λιμνών ΕΥΔΑΠ

Λίμνη / Ταμιευτήρας	Κωδικός σημείου	Περιγραφή θέσης δειγματοληψίας
Υλίκη	9501	Μπούκα : Σημείο εισροής του Βοιωτικού Κηφισού
	9502	Κέντρο λίμνης (παραλιακό σημείο)
	9503	Δίαυλος αντλιοστασίου Μουρικού
Μόρνος	9601	Εκβολή ποταμού Μόρνου
	9602	Εκβολή ποταμού Άβορου
	9603	Κέντρο λίμνης (παραλιακό σημείο)
	9604	Λύματα Λιδωρικού
	9606	Πύργος υδροληψίας
	9607	Κατάντι (φράγμα)
	9608	Εκβολή ποταμού Κόκκινου
Μαραθώνας	9701	Πύργος υδροληψίας (Βεντούρι)
	9702	Εισροή ρέματος Κιούρκων
	9703	Εισροή ποταμού Βαρνάβα (συνεχής ροή)
	9704	Εισροή ρέματος Καταρράκτης (μη συνεχής ροή)
	9705	Εισροή ρέματος Σταμάτας (συνεχής ροή)
	9706	Εισροή ρέματος Αγ. Στεφάνου (συνεχής ροή το χειμώνα)



Σχήμα 2.3. Θέσεις δειγματοληψίας ταμιευτήρων / λιμνών.

Η ποιότητα του ανεπεξέργαστου νερού ελέγχεται επιπροσθέτως στην είσοδο κάθε ΜΕΝ. Η ποιότητα του επεξεργασμένου νερού ελέγχεται στις δεξαμενές αποθήκευσης των ΜΕΝ και σε αντιπροσωπευτικά σημεία που καλύπτουν όλες τις περιοχές του λεκανοπεδίου Αττικής και το νησί της Σαλαμίνας. Σκοπός της δοκιμαστικής παρακολούθησης είναι να παρέχονται, σε τακτική βάση, στοιχεία για την οργανοληπτική και μικροβιολογική ποιότητα του νερού που διατίθεται για ανθρώπινη κατανάλωση καθώς και πληροφορίες για την αποτελεσματικότητα της επεξεργασίας του πόσιμου ύδατος (κυρίως της απολύμανσης), ώστε να διαπιστωθεί κατά πόσον το νερό ανθρώπινης κατανάλωσης τηρεί τις σχετικές παραμετρικές τιμές της οδηγίας 98/83 του Συμβουλίου της Ευρωπαϊκής Ένωσης. Σκοπός της ελεγκτικής παρακολούθησης είναι να παρέχονται τα στοιχεία που απαιτούνται για να διαπιστωθεί κατά πόσον τηρούνται όλες οι παραμετρικές τιμές της ίδιας οδηγίας.

Το τρέχων σχήμα των προσδιορισμών που διενεργούνται, έχει ως εξής:

Οι χημικές παράμετροι που προσδιορίζονται σε δείγματα νερού είναι: Θολότητα, Πυκνότητα ιόντων Υδρογόνου (pH), Αγωγιμότητα, Ολικά διαλυτά στερεά, θερμοκρασία, χρώμα, Αλκαλικότητα Ηλιανθίνης, Σκληρότητα, Χλωριούχα, Φθοριούχα, Βρωμιούχα, Θεϊικά, Νιτρικά, Νιτρώδη, Αμμώνιο, Ασβέστιο, Μαγνήσιο, Φωσφορικά, Κάλιο, Νάτριο, κυανιούχα, Ολικός Οργανικός Άνθρακας, Τριαλογονομεθάνια και χλωροφύλλη-α. Επιπλέον αυτών, γίνονται προσδιορισμοί 15 μετάλλων: Βόριο, Χρώμιο, Μαγγάνιο, Νικέλιο, Χαλκός, Ψευδάργυρος, Αρσενικό, Σελήνιο, Άργυρος, Κάδμιο, Κασσίτερος, Αντιμόνιο, Υδράργυρος, Μόλυβδος και Σίδηρος. Επίσης, διενεργούνται χημικές αναλύσεις για παρασιτοκτόνα (φυτοπροστατευτικές ουσίες) και πολυκυκλικούς αρωματικούς υδρογονάνθρακες. Οι παράμετροι που περιλαμβάνονται στον μικροβιολογικό έλεγχο του νερού είναι: Ολικά κολοβακτηριοειδή, Στελέχη *E.coli*, Εντερόκοκκοι, Ετερότροφα βακτήρια στους 35°C και Ετερότροφα βακτήρια στους 22°C. Το δειγματοληπτικό σχήμα και επιμέρους προσδιορισμοί που διενεργούνται διαφοροποιούνται ανάλογα με τη θέση δειγματοληψίας (Ταμειυτήρες/λίμνες - Ανεπεξέργαστο Νερό ΜΕΝ - Επεξεργασμένο Νερό ΜΕΝ - Δίκτυο Ύδρευσης).

ΚΕΦΑΛΑΙΟ 3

ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΔΕΔΟΜΕΝΩΝ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΝΕΡΟΥ – ΠΡΟΒΛΗΜΑΤΑ ΚΑΙ ΠΡΟΣΕΓΓΙΣΕΙΣ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ

3.1. ΕΙΣΑΓΩΓΗ

Οι υποθέσεις των κλασικών στατιστικών παραμετρικών μεθόδων συνήθως δεν ικανοποιούνται από τα δεδομένα ποιότητας νερού των οποίων μερικές φορές μπορεί να εκλείπουν στοιχεία ως ένα ορισμένο βαθμό ή και να είναι περικεκομμένα (Van Belle και Hughes, 1984). Επιπλέον, οι κατανομές των ποιοτικών δεδομένων του επιφανειακού νερού είναι συχνά ασύμμετρες λόγω ακραίων τιμών με αποτέλεσμα να παραβιάζονται συχνά οι υποθέσεις των παραμετρικών στατιστικών ελέγχων. Αυτά τα ζητήματα απαιτούν περιπλοκότερες στατιστικές τεχνικές συμπεριλαμβανομένων της θεώρησης των μη παραμετρικών στατιστικών προσεγγίσεων. Οι μη παραμετρικές μέθοδοι μπορούν να είναι μια καταλληλότερη επιλογή για την ανάλυση των ποιότητας στοιχείων του επιφανειακού νερού. Οι μη παραμετρικές στατιστικές τεχνικές δεν υποθέτουν μια ιδιαίτερη μορφή κατανομής, και μπορούν να χειριστούν τις απομακρυσμένες τιμές και τις περικεκομμένες που είναι συνήθεις στα ποιοτικά δεδομένα του νερού.

3.2. ΑΚΡΑΙΕΣ ΤΙΜΕΣ

Οι ακραίες τιμές είναι τιμές οι οποίες είναι φανερά μεγαλύτερες ή μικρότερες από την πλειονότητα των δεδομένων. Οι ακραίες τιμές μπορεί να οφείλονται (Helsel και Hirsch, 1992) σε:

- Λανθασμένη μέτρηση ή λάθος στην καταγραφή της τιμής.
- Μία παρατήρηση από έναν πληθυσμό μη παρόμοιο προς αυτόν της πλειονότητας των δεδομένων.
- Ένα σπάνιο γεγονός από έναν μοναδικό ασύμμετρο πληθυσμό.

Οι ακραίες τιμές οι οποίες είναι αποτέλεσμα κάποιας από τις δυο τελευταίες αιτίες θεωρούνται ως πραγματικές (αληθείς) παρατηρήσεις και δεν πρέπει να εξαιρούνται.

Οι λανθασμένες παρατηρήσεις μπορεί να είναι το αποτέλεσμα αρκετών παραγόντων όπως η επιμόλυνση του δείγματος, η ανεπάρκεια του εξοπλισμού του εργαστηρίου, λάθη των αναλυτών, μη προσεκτική καταχώρηση των δεδομένων. Όταν συμπεριλαμβάνονται στην ανάλυση δεδομένων λανθασμένες παρατηρήσεις, οι στατιστικές μέθοδοι μπορεί να δώσουν παραπλανητικές πληροφορίες. Όταν μια ακραία τιμή αποδεικνύεται ότι είναι λανθασμένη παρατήρηση, πρέπει να εξαιρείται από την ανάλυση των δεδομένων. Σε κάθε περίπτωση, ο αριθμός των λανθασμένων παρατηρήσεων μπορεί να μειωθεί σημαντικά όταν ακολουθείται πρόγραμμα διασφάλισης ποιότητας και ελέγχου ποιότητας (Quality Assurance – QA, Quality Control – QC).

3.3. ΕΛΛΕΙΠΟΥΣΕΣ ΤΙΜΕΣ

Οι ελλείπουσες τιμές μπορεί να είναι τυχαίες ή συστηματικές. Οι τυχαίες ελλείπουσες τιμές μπορεί να οφείλονται σε διάφορους παράγοντες όπως αποτυχία του εξοπλισμού, χαμένα δείγματα ή αποτελέσματα της ανάλυσης, έντονα καιρικά φαινόμενα που δεν επιτρέπουν τη διεξαγωγή δειγματοληψίας, έκτακτη έλλειψη προσωπικού κλπ. Οι συστηματικές ελλείπουσες τιμές επιφανειακών δειγμάτων νερού οφείλονται συχνά στο κλίμα μιας περιοχής. Για παράδειγμα, ρέματα και χείμαρροι μπορεί να είναι στεγνά κατά τη διάρκεια του θέρους και απρόσιτα κατά τη χειμερινή περίοδο. Επιπροσθέτως, οι αλλαγές στην συχνότητα δειγματοληψίας επιφέρουν συστηματικές ελλείπουσες τιμές (Lettenmaier και λοιποί, 1991).

Αρκετές τεχνικές στατιστικές αναλύσεις απαιτούν οι δειγματοληψίες να είναι σε ίσα χρονικά διαστήματα με αποτέλεσμα να μην είναι δυνατόν να εφαρμοστούν σε αρχεία δεδομένων τα οποία έχουν ελλείπουσες τιμές. Σε αυτή την περίπτωση, οι μη παραμετρικές μέθοδοι μπορούν συνήθως να διευθετήσουν το πρόβλημα των τυχαίων ελλειπουσών τιμών. Παρόλα αυτά, οι συστηματικές ελλείπουσες τιμές μπορεί να παρουσιάσουν ιδιαίτερα προβλήματα (Lettenmaier και λοιποί, 1991).

Η αντικατάσταση των ελλειπουσών τιμών με αριθμητικές εκτιμήσεις τους δεν συνιστάται λόγω πιθανών σφαλμάτων (μεροληψίας) των στατιστικών

αποτελεσμάτων. Η χρησιμοποίηση μη παραμετρικών στατιστικών μεθόδων είναι η καλύτερη εναλλακτική λύση για την πλειονότητα των περιστάσεων. Στην περίπτωση που το ποσοστό των ελλειπουσών τιμών είναι μεγάλο ή οι ελλείπουσες τιμές είναι συστηματικές, η καλύτερη εναλλακτική μπορεί να είναι η μείωση της συχνότητας των δεδομένων πριν την στατιστική ανάλυση.

Ο Harcum (1990) διερεύνησε την επίδραση των ελλειπουσών τιμών στον εντοπισμό τάσεων. Το ερώτημα που έθεσε αφορούσε στο ποσοστό των ελλειπουσών τιμών των δεδομένων που χρειαζόταν ώστε να ήταν αναγκαία η μείωση των δεδομένων από μηνιαία σε τριμηνιαία. Τα συμπεράσματα στα οποία κατέληξε ήταν:

- Κατά την εφαρμογή του ελέγχου Mann-Kendall ή του εποχικού ελέγχου Kendall, απαιτείται η μείωση των δεδομένων από μηνιαία σε τριμηνιαία όταν το ποσοστό των μηνιαίων δεδομένων που λείπουν είναι μεγαλύτερο από 50%.
- Κατά την εφαρμογή του εποχικού ελέγχου Kendall με διόρθωση για την σειριακή συσχέτιση, απαιτείται η μείωση των δεδομένων από μηνιαία σε τριμηνιαία όταν το ποσοστό των μηνιαίων δεδομένων που λείπουν είναι μεγαλύτερο από 40%.
- Στην περίπτωση που τα καταγεγραμμένα δεδομένα είναι για διάστημα μόλις πέντε ετών και επιπλέον το ποσοστό των ελλειπουσών τιμών είναι μεγαλύτερο από 50% δεν υπάρχει καλή εναλλακτική λύση.

Οι υποδείξεις του Harcum είναι οι πλέον κατάλληλες στην προσπάθεια ανάλυσης ήδη διαθέσιμων δεδομένων. Σε κάθε περίπτωση όμως, ένα ορθά δομημένο και λειτουργικό πρόγραμμα παρακολούθησης οφείλει να προλαμβάνει την εμφάνιση μεγάλου ποσοστού ελλειπουσών τιμών.

3.4. ΠΕΡΙΚΕΚΟΜΜΕΝΑ ΔΕΔΟΜΕΝΑ

Περικοπή ή λογοκρισία τιμών (censoring) είναι η αντικατάσταση αριθμητικών μετρήσεων με ποιοτικές επεξηγήσεις όπως «δεν ανιχνεύθηκε», «μικρότερο από ...», «μικρότερο από το όριο ανίχνευσης». Γενικά, στην ανάλυση της ποιότητας του νερού η περικοπή των τιμών εμφανίζεται στα χαμηλά επίπεδα συγκέντρωσης όπου η αξιοπιστία των μετρήσεων τίθεται υπό αμφισβήτηση. Η περικοπή των δεδομένων

είναι αποτέλεσμα της έλλειψης εμπιστοσύνης στο αριθμητικό αποτέλεσμα ή και του φόβου ότι τα αβέβαια αριθμητικά αποτελέσματα μπορεί να χρησιμοποιηθούν ή να ερμηνευθούν λανθασμένα (Ward, 1999). Αρκετοί ερευνητές προτείνουν ότι δεν πρέπει να επιτρέπεται η περικοπή των τιμών από τα αναλυτικά εργαστήρια αλλά να παρέχονται οι μετρούμενες τιμές συγκεντρώσεων και τα όρια ανίχνευσης ή τα διαστήματα εμπιστοσύνης. Οι Porter και Ward (1991) συγκρίνανε την απώλεια πληροφορίας που αποδίδεται στην περικοπή των δεδομένων με αυτήν που οφείλεται στο θόρυβο της μέτρησης και τη βαθμονόμηση. Συμπεράναν ότι όταν η περικοπή των δεδομένων υπερβαίνει περίπου το 50%, τα μη περικομμένα δεδομένα προσφέρουν περισσότερη πληροφορία για την κεντρική θέση του υπό εξέταση πληθυσμού απ' ό,τι τα περικομμένα. Οι Gilliom και λοιποί (1984) διερεύνησαν την επίδραση της περικοπής των δεδομένων στην ικανότητα εντοπισμού τάσεων. Κατέληξαν ότι οι τάσεις εντοπίζονται πιο αποτελεσματικά σε μη περικομμένα δεδομένα απ' ό,τι σε περικομμένα ακόμα κι όταν τα δεδομένα που είχαν περικοπεί ήταν πολύ ανακριβή.

Η επίδραση των περικομμένων τιμών στην στατιστική ανάλυση δεδομένων ποιότητας του νερού εξαρτάται από το ποσοστό των περικομμένων τιμών (Gilliom και λοιποί, 1984; Porter και λοιποί, 1988; Porter και Ward, 1991; Helsel, 1990), τη στατιστική προσέγγιση που επιλέγεται για το χειρισμό τους (Helsel, 1990) και την ποιότητα των περικομμένων δεδομένων (Gilliom και λοιποί, 1984; Taylor, 1987).

Ο Helsel (1990) πραγματοποιεί μια πολύ ενδιαφέρουσα αναφορά σχετική με τις κατάλληλες στατιστικές προσεγγίσεις για τη διαχείριση περικομμένων δεδομένων. Για τους ελέγχους υποθέσεων προτείνει τους μη παραμετρικούς ελέγχους οι οποίοι χρησιμοποιούν τις τάξεις μεγέθους (ranks) των δεδομένων αντί για τα ίδια τα δεδομένα. Όλες οι περικομμένες τιμές αντιπροσωπεύονται από τάξεις μεγέθους οι οποίες ισοβαθμούν με τιμές οι οποίες είναι μικρότερες από την χαμηλότερη παρατηρούμενη τιμή πάνω από το όριο αναφοράς. Στην περίπτωση που υπάρχουν περισσότερα του ενός όρια αναφοράς, οι έλεγχοι υποθέσεων όπως οι συγκρίσεις μεταξύ διαφορετικών ομάδων δεδομένων είναι πολύ περίπλοκοι. Ο Helsel προτείνει ότι μια ανθεκτική μέθοδος, που μπορεί πάντα να χρησιμοποιηθεί, είναι να περικοπούν όλα τα δεδομένα στο υψηλότερο όριο αναφοράς, και στη συνέχεια να εφαρμοστεί ο κατάλληλος μη παραμετρικός έλεγχος. Επιπλέον, όταν εμφανίζεται μεγάλο ποσοστό περικομμένων τιμών (κοντά στο 50% μεγαλύτερο), ακόμη και οι μη παραμετρικοί

έλεγχοι έχουν μικρή ικανότητα να ανιχνεύσουν διαφορές στις κεντρικές τιμές των δεδομένων. Σε αυτήν την περίπτωση, οι πίνακες συνάφειας μπορούν να χρησιμοποιηθούν για τον έλεγχο ύπαρξης διαφορών στα ποσοστά των δεδομένων που υπερβαίνουν το όριο ανίχνευσης σε κάθε ομάδα (Conover, 1980). Ο έλεγχος προσδιορίζει εάν το ποσοστό των δεδομένων που εμπίπτει σε κάθε κατηγορία διαφέρει συναρτήσει των διαφορετικών επεξηγηματικών ομάδων, π.χ. διαφορετικές θέσεις δειγματοληψίας.

3.5. ΕΛΛΕΙΨΗ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

Τα δεδομένα των ποιοτικών χαρακτηριστικών του νερού παρουσιάζουν συχνά θετική ασυμμετρία και συνεπώς παραβιάζουν την υπόθεση της κανονικότητας (Adkins, 1993; Twenter και Loftis, 2003). Οι περισσότερες παραμετρικές στατιστικές τεχνικές οι οποίες χρησιμοποιούνται για την ανάλυση δεδομένων ποιότητας του νερού υποθέτουν ότι τα δεδομένα ακολουθούν κανονική κατανομή. Στην περίπτωση που παραβιάζεται η υπόθεση της κανονικότητας η ισχύς των παραμετρικών μεθόδων μειώνεται δραματικά. Για την αποφυγή του προβλήματος της μείωσης της ισχύος λόγω της μη κανονικότητας των δεδομένων, προτείνεται η χρήση μη παραμετρικών στατιστικών τεχνικών. Παρόλα αυτά, δεν υπάρχουν διαθέσιμες μη παραμετρικές τεχνικές για όλες τις εφαρμογές. Επιπλέον, σε κάποιες περιπτώσεις υπάρχουν οδηγίες σύμφωνα με τις οποίες απαιτείται η χρήση παραμετρικών μεθόδων. Στην περίπτωση που πρέπει να χρησιμοποιηθούν παραμετρικές τεχνικές, χρειάζεται να διερευνηθεί η κατανομή των δεδομένων.

Γενικά, θεωρείται ότι είναι σύνηθες για τα δεδομένα των ποιοτικών χαρακτηριστικών του νερού να μην κατανέμονται κανονικά. Τα δεδομένα παρουσιάζουν συχνά θετική ασυμμετρία λόγω του κατώτατου ορίου (μηδέν) ή της παρουσίας τιμών κάτω από το όριο ανίχνευσης και της εμφάνισης σπάνιων υψηλών τιμών. Οι ιδιότητες κατανομής των ποιοτικών δεδομένων του νερού έχουν διερευνηθεί από αρκετούς ερευνητές και τα αποτελέσματά τους παρουσιάζονται στη συνέχεια.

Οι Montgomery και λοιποί (1987) έλεγξαν 172 αρχεία δεδομένων ποιότητας υπόγειων νερών όσον αφορά στην κανονικότητα των παραμέτρων. Χρησιμοποίησαν ιστογράμματα συχνοτήτων, διαγράμματα κανονικής πιθανότητας (normal probability

plots) τον έλεγχο X^2 καλής προσαρμογής (chi-squared goodness-of-fit test) και τον έλεγχο ασυμμετρίας. Το επίπεδο σημαντικότητας που χρησιμοποίησαν στους δυο ελέγχους υποθέσεων ήταν 5%. Σύμφωνα με τα ευρήματά τους, στο 62% των αρχείων τα δεδομένα δεν ακολουθούσαν την κανονική κατανομή ενώ παρουσίαζαν δεξιά κλίση (θετική ασυμμετρία) με βαθμό κυμαινόμενο σε μεγάλο εύρος.

Οι Gilliom και Helsel (1986) εξέτασαν τα στατιστικά χαρακτηριστικά των συγκεντρώσεων ιχνοστοιχείων σε δείγματα που συλλέχθηκαν σε σταθμούς παρακολούθησης της ποιότητας νερού ποταμών από την γεωλογική επιθεώρηση των ΗΠΑ (U.S. Geological Survey). Σε 482 σετ μη περικομμένων δεδομένων, ο συντελεστής ασυμμετρίας κυμαίνονταν από -0,8 έως 5,2 με διάμεση τιμή 1,8. Το ποσοστό των συντελεστών ασυμμετρίας που είχαν αρνητική τιμή ήταν μόλις 6%. Δεδομένου ότι ο συντελεστής ασυμμετρίας σε μια κανονική κατανομή (η οποία είναι εξ ορισμού συμμετρική) είναι 0, οι Gilliom και Helsel συμπέραναν ότι τα δεδομένα ποιότητας του νερού παρουσιάζουν συχνά θετική ασυμμετρία.

Οι Loftis και λοιποί (1989), εξέτασαν την κανονικότητα των ποιοτικών παραμέτρων νερού λιμνών. Χρησιμοποίησαν αμφίπλευρους ελέγχους υποθέσεων για τους συντελεστές ασυμμετρίας και κύρτωσης σε δυο επίπεδα σημαντικότητας (10% και 2%) σε κάθε έλεγχο. Σύμφωνα με τα ευρήματά τους, στο 20% των αρχείων τα δεδομένα παρέκλιναν σημαντικά από την κανονικότητα. Επανέλαβαν την στατιστική ανάλυση στις δεκαδικά λογαριθμισμένες τιμές των δεδομένων (ο λογαριθμικός μετασχηματισμός μπορεί να φέρει τα δεδομένα πιο κοντά στην κανονικότητα) καθώς επίσης και μετά την απαλοιφή των τριμηνιαίων μέσων (η απαλοιφή των τριμηνιαίων μέσων μπορεί να μειώσει την επίδραση της εποχικότητας). Καμία από τις δυο αυτές ενέργειες δεν επέφερε σημαντική αλλαγή στο ποσοστό των παραμέτρων οι οποίες δεν ακολουθούσαν κανονική κατανομή.

Η εφαρμογή στατιστικών ελέγχων οι οποίοι υποθέτουν κανονικότητα σε μη κανονικά κατανομημένα δεδομένα μπορεί να επιφέρει δυο ανεπιθύμητα αποτελέσματα. Το πρώτο είναι η διαστροφή της ονομαστικής τιμής του λάθους τύπου I (Snedecor και Cochran, 1989), ενώ το δεύτερο είναι η μείωση της ισχύος (Conover και Iman, 1976; Helsel, 1987). Η μείωση της ισχύος των ελέγχων είναι ενδεχομένως πολύ σοβαρό πρόβλημα γιατί μπορεί να μην αναγνωριστούν σημαντικές αλλαγές ή διαφορές στην

ποιότητα του νερού. Γενικά, όταν ένας έλεγχος θεωρείται ότι είναι ανθεκτικός στην έλλειψη κανονικότητας σημαίνει ότι η έλλειψη κανονικότητας δεν προκαλεί μεγάλη διαστροφή στο λάθος τύπου I (Conover και Iman, 1976). Για παράδειγμα, ο έλεγχος t συχνά θεωρείται ανθεκτικός στην έλλειψη κανονικότητας γιατί το ονομαστικό επίπεδο σημαντικότητας του διατηρείται για μεγάλο μέγεθος δείγματος. Αυτός όμως ο ισχυρισμός αγνοεί την απώλεια της ισχύος η οποία μπορεί να εμφανιστεί όταν δεν ικανοποιείται η υπόθεση της κανονικότητας (Helsel, 1987).

Η μεταχείριση δεδομένων που ενδεχομένως δεν ακολουθούν κανονική κατανομή μπορεί να γίνει με τους παρακάτω τρόπους:

- Με τη χρησιμοποίηση μη παραμετρικών μεθόδων στατιστικής ανάλυσης. Η αποτελεσματικότητα των μη παραμετρικών ελέγχων είναι ελαφρώς μικρότερη των αντίστοιχων παραμετρικών όταν οι υποκείμενοι πληθυσμοί κατανέμονται κανονικά ενώ είναι πολύ μεγαλύτερη όταν οι υποκείμενοι πληθυσμοί δεν κατανέμονται κανονικά (Hollander και Wolfe, 1973; Bradley, 1968; Hirsch και λοιποί, 1991). Επιπλέον οι μη παραμετρικές μέθοδοι είναι ανθεκτικές στις ακραίες τιμές, μπορούν να χρησιμοποιηθούν για την ανάλυση δεδομένων στα οποία εμπεριέχονται αρκετές περικομμένες τιμές, να χειριστούν τις ελλείπουσες τιμές και να εφαρμοστούν σε διατεταγμένα δεδομένα και ενίοτε σε κατηγορικά δεδομένα (Daniel, 1991; Gibbons, 1985; Helsel, 1987; Hollander και Wolfe, 1973).
- Με τη διεξαγωγή ελέγχων κανονικότητας και, εφόσον τα δεδομένα ακολουθούν κανονική κατανομή, τη χρησιμοποίηση παραμετρικών μεθόδων.
- Με τη διεξαγωγή ελέγχων κανονικότητας και, εφόσον τα δεδομένα δεν ακολουθούν κανονική κατανομή, τη χρησιμοποίηση μετασχηματισμών. Με έλεγχο της κανονικότητας των μετασχηματισμένων δεδομένων και, εφόσον ακολουθούν κανονική κατανομή, τη χρησιμοποίηση παραμετρικών μεθόδων στα μετασχηματισμένα δεδομένα (Helsel, 1987). Το πρόβλημα όμως είναι ότι συχνά οι απλοί μετασχηματισμοί (π.χ. λογαρίθμηση, τετραγωνική ρίζα, αντιστροφή) ενώ μπορεί να έχουν φυσική ερμηνεία σε αρκετές περιπτώσεις, δεν οδηγούν απαραίτητα στην κανονικότητα. Από την άλλη, πιο πολύπλοκοι μετασχηματισμοί που προσεγγίζουν την κανονικότητα, παράγουν συχνά νέες μεταβλητές που δεν έχουν καμία φυσική σημασία. Η χρησιμοποίηση μετασχηματισμών είναι γενικά

επίπονη και χρονοβόρα διαδικασία στην περίπτωση πολλαπλών σετ δεδομένων (Hirsch και λοιποί, 1991). Κάθε σετ δεδομένων πρέπει να εξεταστεί ξεχωριστά για την επιλογή του καλύτερου δυνατού μετασχηματισμού. Επιπλέον, στην περίπτωση που επιλέγονται αρκετοί διαφορετικοί μετασχηματισμοί, υπάρχει δυσκολία στη σύγκριση των αποτελεσμάτων της στατιστικής ανάλυσης μεταξύ των σετ δεδομένων.

- Με τη διεξαγωγή ανάλυσης ευαισθησίας για την ποσοτικοποίηση των επιδράσεων της έλλειψης κανονικότητας στον παραμετρικό έλεγχο που πρόκειται να χρησιμοποιηθεί (Hirsch και Slack, 1984). Αν οι επιδράσεις είναι αποδεκτές, ο έλεγχος μπορεί να προχωρήσει. Η συγκεκριμένη προσέγγιση θεωρείται δαπανηρή και έχει πρακτική αξία όταν πρόκειται να αναλυθούν πολύ σημαντικά δεδομένα με μεθόδους για τις οποίες δεν υπάρχουν μη παραμετρικές εναλλακτικές μέθοδοι.

3.6. ΕΠΟΧΙΚΟΤΗΤΑ

Η εποχικότητα είναι ένα χαρακτηριστικό των δεδομένων ποιότητας του νερού το οποίο απεικονίζει μια αναγνωρίσιμη κυκλική διάταξη που εμφανίζεται στα δεδομένα (Ward, 1999). Με άλλα λόγια, η εποχικότητα είναι η αλλαγή στις κατανομές των ποιοτικών παραμέτρων του νερού η οποία μπορεί να αποδοθεί στην χρονική περίοδο του έτους. Ως εποχή μπορεί να θεωρηθεί οποιαδήποτε χρονική περίοδος αλλά γενικά είναι ένας μήνας (12 εποχές ανά έτος) ή τρεις μήνες (4 εποχές ανά έτος).

Η εποχικότητα μεγαλώνει τη διακύμανση των ποιοτικών δεδομένων του νερού με συνέπεια την αύξηση του πλάτους των διαστημάτων εμπιστοσύνης στις διαδικασίες εκτίμησης και τη μείωση της ισχύος στους ελέγχους υποθέσεων.

Η εποχικότητα για μια μεταβλητή μπορεί να ελεγχθεί με κάποιον από τους παρακάτω τρόπους:

- Οπτική επιθεώρηση του διαγράμματος των συγκεντρώσεων έναντι του χρόνου (Montgomery και λοιποί, 1987).
- Εξέταση των ετήσιων κύκλων σε ένα διάγραμμα αυτοσυσχετίσεων (correlogram) (Loftis και λοιποί, 1989).

- Με την κατασκευή θηκογραμμάτων (box plots) για κάθε εποχή (Montgomery και λοιποί, 1987). Στην περίπτωση που τα ορθογώνια πλαίσια δεν επικαλύπτονται, θεωρείται ότι η εποχικότητα πιθανόν να αποτελεί σημαντική πηγή της διακύμανσης των δεδομένων (Ward και Loftis, 1989).
- Με ομαδοποίηση των δεδομένων σύμφωνα με την εποχή και τον υπολογισμό του αριθμητικού μέσου για κάθε εποχή. Στη συνέχεια, με τον υπολογισμό του λόγου του μέγιστου προς τον ελάχιστο αριθμητικό μέσο. Η εποχικότητα αυξάνει ανάλογα με τον λόγο αυτόν. Αντίστοιχη διαδικασία μπορεί να γίνει χρησιμοποιώντας τις τυπικές αποκλίσεις (Loftis και λοιποί, 1989).
- Με ομαδοποίηση των δεδομένων σύμφωνα με την εποχή και διεξαγωγή ανάλυσης διακύμανσης (ANOVA) ή του αντίστοιχου μη παραμετρικού ελέγχου Kruskal-Wallis (Montgomery και λοιποί, 1987).
- Με την εξέταση του διαγράμματος της συνάρτησης αυτοσυσχέτισης (Reckhow και λοιποί, 1993; Yürekli και Kurunc, 2005). Η ύπαρξη εποχικότητας στα δεδομένα διαπιστώνεται με την εμφάνιση στατιστικά σημαντικών αυτοσυσχετίσεων στις εποχικές υστερήσεις.

Στην περίπτωση ύπαρξης εποχικότητας στα δεδομένα, υπάρχουν δυο συνήθεις τρόποι στατιστικής προσέγγισης:

- Η αποεποχικοποίηση, δηλαδή ο μετασχηματισμός των δεδομένων ώστε να εξαλειφθεί ο ποσοτικός εποχικός κύκλος. Αυτό επιτυγχάνεται με την αφαίρεση των εποχικών διαμέσων από όλα τα δεδομένα μέσα σε μια εποχή. Στη συνέχεια ακολουθεί η στατιστική ανάλυση των αποεποχικοποιημένων δεδομένων.
- Η χρησιμοποίηση στατιστικών μεθόδων οι οποίες λαμβάνουν υπόψη τους την εποχικότητα των δεδομένων.

Στην περίπτωση που ο σκοπός της στατιστικής ανάλυσης είναι ο εντοπισμός τάσεων σε δεδομένα τα οποία είναι ύποπτα για την ύπαρξη εποχικότητας, συνιστάται ο εποχικός έλεγχος Kendall (Seasonal Kendall test), ο οποίος μπορεί επιπλέον να χειριστεί ελλείπουσες τιμές, περικεκομμένες τιμές και ισοβαθμούσες τιμές δεδομένων. Ο εποχικός έλεγχος Kendall χρησιμοποιείται σε μεγάλο βαθμό στην ανάλυση δεδομένων ποιότητας του νερού.

3.7. ΣΕΙΡΙΑΚΗ ΣΥΣΧΕΤΙΣΗ (ΑΥΤΟΣΥΣΧΕΤΙΣΗ)

Συνήθως, στα προγράμματα παρακολούθησης της ποιότητας του νερού η δειγματοληψία γίνεται ανά τακτά χρονικά διαστήματα (πχ. ημερησίως, μηνιαίως κλπ). Η σειριακή συσχέτιση ή αυτοσυσχέτιση μπορεί να θεωρηθεί γενικά ως περιττή πληροφορία μεταξύ διαδοχικών παρατηρήσεων μιας χρονολογικής σειράς που προκύπτει από δείγματα τα οποία λαμβάνονται ανά μικρά χρονικά διαστήματα σε σχέση με την ενδιαφέρουσα χρονική περίοδο. Όταν η σειριακή συσχέτιση είναι θετική, παρουσιάζεται η τάση υψηλές τιμές να ακολουθούν υψηλές τιμές και χαμηλές τιμές να ακολουθούν χαμηλές τιμές. Αντίθετα, αρνητική σειριακή συσχέτιση παρουσιάζεται όταν υψηλές τιμές ακολουθούν χαμηλές και χαμηλές ακολουθούν υψηλές. Τα δεδομένα ποιότητας του νερού παρουσιάζουν κυρίως θετική σειριακή συσχέτιση.

Οι Loftis και λοιποί (1991) δηλώνουν ότι το διάστημα εμπιστοσύνης της μέσης τιμής για ένα συγκεκριμένο χρονικό διάστημα (πχ. ένα έτος), όταν εκτιμάται μέσω ενός δεδομένου αριθμού ισαπέχουσων χρονικά παρατηρήσεων που καλύπτουν χρονικά το υπό εξέταση διάστημα, είναι στενότερο όταν τα δεδομένα είναι σειριακά συσχετισμένα απ' ότι όταν είναι ανεξάρτητα. Στην περίπτωση όμως που αποτελεί ενδιαφέρον η εξαγωγή συμπερασμάτων για μια μακρά χρονική περίοδο χρησιμοποιώντας ένα συγκεκριμένο χρονικό διάστημα της περιόδου αυτής, μια σειρά σειριακά συσχετισμένων δεδομένων θα περιέχει μικρότερη ποσότητα πληροφορίας σε σχέση με μια σειρά ανεξάρτητων δεδομένων του ίδιου μεγέθους. Στην περίπτωση αυτή, οι έλεγχοι για εντοπισμό τάσης ή για αλλαγές στη διάρκεια του χρόνου τείνουν να απορρίπτουν την μηδενική υπόθεση, υποδηλώνοντας ότι έχει συμβεί κάποια αλλαγή, συχνότερα απ' ότι θα έπρεπε. Επιπλέον, οι Loftis και λοιποί (1991) επισήμαναν ότι το κατά πόσο το παρατηρούμενο πρότυπο μιας δεδομένης σειράς δεδομένων ισαπέχουσων χρονικά παρατηρήσεων μπορεί να αποδοθεί σε σειριακή συσχέτιση ή τάση εξαρτάται από την χρονική κλίμακα.

Η ύπαρξη σειριακής συσχέτισης στα δεδομένα μπορεί να διαστρέψει τα αποτελέσματα οποιουδήποτε στατιστικού ελέγχου συμπεριλαμβανομένων των μη παραμετρικών ελέγχων λόγω της παραβίασης της υπόθεσης ανεξαρτησίας των παρατηρήσεων. Η συνηθέστερη επίδραση της θετικής σειριακής συσχέτισης στους

ελέγχους υποθέσεων θεωρείται ότι είναι η αύξηση του λάθους τύπου I συγκριτικά με την ονομαστική του τιμή (Harcum, 1990; El-Shaarawi και Damsleth, 1988; Keller-McNulty και McNulty, 1987; Pavur, 1988; Scariano και Davenport, 1987).

Η σειριακή συσχέτιση είναι πιθανότατα το πιο δύσκολο χαρακτηριστικό όσον αφορά τόσο στην κατανόησή του όσο και στον χειρισμό του που μπορεί να εμφανίζουν τα δεδομένα. Η διερεύνηση για την ύπαρξη σειριακής συσχέτισης στα δεδομένα, η μοντελοποίησή της, ο υπολογισμός του μεγέθους της και ο έλεγχος σημαντικότητάς της αναλύονται περαιτέρω στο Κεφάλαιο 7 (§7.2 έως §7.5).

ΚΕΦΑΛΑΙΟ 4

ΑΝΑΛΥΣΗ ΤΑΣΕΩΝ

4.1. ΕΙΣΑΓΩΓΗ

Η ανάλυση τάσεων επιτρέπει τον προσδιορισμό της γενικής αύξησης ή μείωσης των τιμών μιας σειράς παρατηρήσεων μιας τυχαίας μεταβλητής οι οποίες έχουν συλλεχθεί στη διάρκεια μιας χρονικής περιόδου. Σε όρους στατιστικής, πρόκειται για τον καθορισμό του κατά πόσο έχει συμβεί μεταβολή στην κατανομή πιθανότητας, από την οποία προκύπτουν οι παρατηρήσεις, με την πάροδο του χρόνου. Οι έλεγχοι για ύπαρξη τάσης στα δεδομένα χρησιμοποιούν ως μηδενική την υπόθεση H_0 : δεν υπάρχει τάση. Η μη απόρριψη της μηδενικής υπόθεσης δεν σημαίνει την απόδειξη της μη ύπαρξης τάσης αλλά τη δήλωση ότι δεν υπάρχουν αρκετές ενδείξεις για την εξαγωγή του συμπεράσματος ότι υπάρχει τάση.

Ο έλεγχος για ύπαρξη τάσης που επιλέγεται θα πρέπει να έχει υψηλή ισχύ για το είδος των δεδομένων στα οποία πρόκειται να εφαρμοστεί. Τα δεδομένα ποιότητας του νερού εμφανίζουν συχνά χαρακτηριστικά όπως έλλειψη κανονικότητας, ακραίες τιμές, εποχικότητα, ελλείπουσες τιμές, περικεκομμένες τιμές, σειριακή συσχέτιση. Επομένως, ο έλεγχος που επιλέγεται θα πρέπει να είναι ανθεκτικός, δηλαδή να έχει σχετικά υψηλή ισχύ για όλες τις καταστάσεις και τύπους δεδομένων που αναμένεται να εμφανιστούν.

4.2. ΠΡΟΣΕΓΓΙΣΕΙΣ ΣΤΗΝ ΑΝΑΛΥΣΗ ΤΑΣΕΩΝ

4.2.1. Παραμετρική ανάλυση τάσης

Η παραμετρική μέθοδος για τον έλεγχο τάσης είναι η απλή γραμμική παλινδρόμηση της τυχαίας μεταβλητής απόκρισης Y πάνω στον χρόνο X (ο οποίος συνήθως εκφράζεται σε έτη): $Y = \alpha + \beta * X + \varepsilon$, όπου ε αντιπροσωπεύει τα τυχαία λάθη ενώ α

και β είναι οι συντελεστές της παλινδρόμησης. Η μέση τιμή της τυχαίας μεταβλητής Y για δοθέν x συνδέεται γραμμικά με το x . Δηλαδή:

$$\mu_{Y|x_i} = \alpha + \beta x_i \quad (\text{ευθεία παλινδρόμησης})$$

Η σημειακή εκτίμηση της ευθείας παλινδρόμησης που βασίζεται σε ένα τυχαίο δείγμα παρατηρήσεων (x_i, y_i) θα δίνεται από τον τύπο:

$$\hat{y} = a + bx$$

Όπου a και b είναι οι σημειακές εκτιμήσεις των εκτιμητριών $\hat{\alpha}$ και $\hat{\beta}$ των α και β αντίστοιχα. Οι παρατηρούμενες τιμές e_i των ε_i θα δίνονται από τον τύπο:

$$e_i = y_i - \hat{y}_i$$

Η εκτίμηση των τιμών a και b , ώστε αυτές να είναι ικανοποιητικές προσεγγίσεις των α και β , στηρίζεται στη μέθοδο ελαχίστων τετραγώνων σύμφωνα με την οποία οι τιμές αυτές θα πρέπει να εξασφαλίζουν την ελάχιστη τιμή της παράστασης:

$$\sum_{i=1}^n e_i^2$$

Οι εκτιμήσεις ελαχίστων τετραγώνων των συντελεστών γραμμικής παλινδρόμησης δίνουν:

$$a = \bar{y} - b\bar{x} \quad \text{και} \quad b = \frac{n \sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Όπου \bar{y} και \bar{x} οι αριθμητικοί μέσοι των y_i και x_i αντίστοιχα και n το μέγεθος του δείγματος.

Η ανάλυση παλινδρόμησης χρησιμοποιεί ισχυρές υποθέσεις για την κατανομή της Y στην πάροδο του χρόνου. Θα πρέπει να ελέγχεται η Y για γραμμικότητα στη διάρκεια του χρόνου και τα κατάλοιπα της παλινδρόμησης για τυχαιότητα, κανονικότητα και ομοσκεδαστικότητα. Στην περίπτωση που η Y δεν είναι γραμμική στη διάρκεια του χρόνου, μπορεί να είναι απαραίτητος κάποιος μετασχηματισμός της. Με την προϋπόθεση ότι ισχύουν οι υποθέσεις της παλινδρόμησης, γίνεται ο έλεγχος με μηδενική υπόθεση ότι η κλίση $\beta=0$. Για να ελεγχθεί εάν το β διαφέρει στατιστικά σημαντικά από το μηδέν, αν δηλαδή υπάρχει γραμμική σχέση μεταξύ των μεταβλητών X και Y , εξετάζεται το στατιστικό T , το οποίο ακολουθεί κατανομή t (student) με $n-2$ βαθμούς ελευθερίας και δίνεται από τον τύπο:

$$T = \frac{\hat{\beta}}{\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2 / (n-2)}{\sum_i (x_i - \bar{X})^2}}}$$

4.2.2. Μη παραμετρική ανάλυση τάσης

Η ευρύτατα χρησιμοποιημένη μη παραμετρική μέθοδος για τον έλεγχο τάσης είναι ο έλεγχος Mann-Kendall, γνωστός και ως Kendall's tau statistic. Αυτή η τεχνική έχει χρησιμοποιηθεί ευρύτατα για τον έλεγχο της τυχαιότητας έναντι της ύπαρξης τάσης στις κλιματολογικές χρονολογικές σειρές (Zhang και λοιποί, 2001; Kahya και Kalaycı, 2004). Η μηδενική υπόθεση του ελέγχου είναι «H₀: Δεν υπάρχει τάση» ενώ η εναλλακτική: «Υπάρχει ανοδική ή πτωτική τάση». Η μοναδική προϋπόθεση του ελέγχου είναι ότι η τάση είναι μονότονη και οι παρατηρήσεις ανεξάρτητες. Το στατιστικό του ελέγχου Mann-Kendall δίνεται από τον τύπο:

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

Όπου,

$$\text{sgn}(x_j - x_k) = \begin{cases} +1 & \alpha\nu \quad (x_k - x_j) > 0 \\ 0 & \alpha\nu \quad (x_k - x_j) = 0 \\ -1 & \alpha\nu \quad (x_k - x_j) < 0 \end{cases}$$

Και, x_j είναι η τιμή της παρατήρησης, j^{th} το έτος και x_k η τιμή της παρατήρησης το k^{th} έτος (με $k > j$).

Η διακύμανση του S , στην περίπτωση που μπορεί να υπάρχουν ισοβαθμίες (δηλαδή ίσες τιμές στην υπό εξέταση μεταβλητή X), δίνεται από τον τύπο:

$$\text{Var}(S) = \frac{1}{18} [n(n-1)(2n+5) - \sum_{i=1}^g t_i(t_i-1)(2t_i+5)]$$

Όπου, n είναι το μέγεθος του δείγματος, g είναι ο αριθμός των ομάδων με ισοβαθμίες στο σύνολο των δεδομένων, και t_i ο αριθμός των ισοβαθμισμένων παρατηρήσεων στην i^{th} ομάδα.

Ακόμη και για μέγεθος δείγματος $n = 10$, κάτω από τη μηδενική υπόθεση της τυχαιότητας των δεδομένων ενάντια στην ύπαρξη τάσης, η ποσότητα z , που υπολογίζεται από την ακόλουθη εξίσωση (Douglas και λοιποί, 2000), ακολουθεί προσεγγιστικά την τυποποιημένη κανονική κατανομή.

$$z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & , \text{αν } S > 0 \\ 0 & , \text{αν } S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}} & , \text{αν } S < 0 \end{cases}$$

Στον αμφίπλευρο έλεγχο τάσης, η μηδενική υπόθεση απορρίπτεται, σε επίπεδο σημαντικότητας α , εάν $|z| > z_{1-\alpha/2}$. Θετική τιμή του S υποδεικνύει ανοδική τάση ενώ αρνητική τιμή υποδεικνύει πτωτική τάση.

Για την εκτίμηση της συνολικής κλίσης, ακολουθείται μια μη παραμετρική μέθοδος, γνωστή ως «εκτιμήτρια κλίσης του Sen» (Sen, 1968b). Σύμφωνα με αυτήν, χρειάζεται να υπολογιστούν οι κλίσεις για όλα τα ζεύγη των n χρονικών σημείων σύμφωνα με τον τύπο:

$$Q_i = (x_k - x_j) / (k - j)$$

Όπου, x_j και x_k είναι οι τιμές των δεδομένων τις χρονικές στιγμές j και k (με $k > j$) αντίστοιχα. Η διάμεσος όλων των τιμών Q_i είναι η εκτιμήτρια Sen's της συνολικής κλίσης. Αν δεν υπάρχει μια υποκείμενη τάση, τότε προσεγγιστικά ο αριθμός των θετικών κλίσεων θα ισούται με τον αριθμό των αρνητικών κλίσεων και συνεπώς η διάμεσος θα έχει τιμή κοντά στη μηδενική.

Οι Hirsch και λοιποί (1982) πρότειναν μια τροποποιημένη έκδοση του ελέγχου Mann-Kendall, η οποία είναι πιο κατάλληλη για εποχικά δεδομένα και αποκαλείται εποχικός έλεγχος Kendall (Seasonal Kendall test). Αυτός ο έλεγχος προορίζεται να αποτιμήσει την τυχαιότητα ενός συνόλου στοιχείων $X = (X_1, \dots, X_m)$ και $X_i = (x_{i1}, x_{i2}, \dots, x_{im})$, όπου X είναι ο πίνακας των συνολικών εποχικών δεδομένων κατά τη διάρκεια n ετών για μια μόνο παράμετρο σε μια θέση δειγματοληψίας. Η μηδενική υπόθεση του ελέγχου δηλώνει ότι οι παρατηρήσεις είναι ανεξάρτητες από τον χρόνο (εποχή και έτος) συλλογής των δειγμάτων. Η εναλλακτική υπόθεση δηλώνει ότι για

μια ή περισσότερες εποχές τα δεδομένα δεν είναι ανεξάρτητα από τον χρόνο. Το στατιστικό S' για τον εποχικό έλεγχο Kendall είναι ένα άθροισμα των στατιστικών S_i των ελέγχων Mann-Kendall που έχουν υπολογιστεί για κάθε εποχή. Δηλαδή:

$$S' = \sum_{i=1}^m S_i$$

Η διακύμανση του S' είναι:

$$Var(S') = \sum_{i=1}^m Var(S_i)$$

Στην περίπτωση μηνιαίων δεδομένων έχουμε ότι $m=12$.

Η ερμηνεία του υπολοίπου ελέγχου είναι παρόμοια με αυτήν του ελέγχου Mann-Kendall.

Στην περίπτωση που απορρίπτεται η μηδενική υπόθεση, το μέγεθος της κλίσης μπορεί να εκτιμηθεί με την τροποποίηση της εκτιμήτριας κλίσης του Sen ώστε να λαμβάνει υπόψη της τους εποχικούς κύκλους. Για παράδειγμα, αν τα διαθέσιμα δεδομένα είναι μηνιαίες τιμές της μεταβλητής για έναν αριθμό ετών, τότε θα πρέπει να προσδιοριστούν δώδεκα διαφορετικά σετ κλίσεων. Η διάμεσος όλων των κλίσεων θα είναι η εκτιμήτρια της μακροχρόνιας τάσης.

Η γενίκευση της εκτιμήτριας κλίσης του Sen οδηγεί στην εποχική Kendall εκτιμήτρια κλίσης. Σύμφωνα με αυτή την μέθοδο υπολογίζονται οι μεμονωμένες εκτιμήτριες κλίσης για κάθε εποχή. Για παράδειγμα, στην περίπτωση μηνιαίων δεδομένων, θα είναι:

$$Q_i = (x_{ik} - x_{ij}) / (k - j)$$

Όπου $i = 1, 2, \dots, 12$, x_{ik} η τιμή των δεδομένων τον μήνα i του έτους k και x_{ij} η τιμή των δεδομένων τον μήνα i του έτους j με $k > j$. Η διάμεσος των δώδεκα μεμονωμένων εκτιμητριών κλίσης, θα είναι η εποχική Kendall εκτιμήτρια κλίσης (Yu και λοιποί., 1993).

Το πλεονέκτημα του εποχικού ελέγχου Kendall είναι ότι οι τάσεις στις εποχές με τις μικρές τιμές δεν κυριαρχούνται από τις μεγαλύτερες τιμές σε άλλες εποχές (Lettenmaier και λοιποί., 1994). Επιπλέον, έχει τη δυνατότητα να λαμβάνει υπόψη του την σειριακή συσχέτιση με την ενσωμάτωση ενός όρου συνδιακύμανσης στον

υπολογισμό της διακύμανσης των στατιστικών Mann-Kendall (Hirsch και Slack, 1984). Εάν τα σφάλματα είναι ασυσχέτιστα, τότε η τυπική διαδικασία του εποχικού ελέγχου Kendall, χωρίς τον όρο συνδιακύμανσης, έχει καλή απόδοση (Taylor και Loftis, 1989). Αξίζει να σημειωθεί ότι ο προσαρμοσμένος για τη σειριακή συσχέτιση εποχικός έλεγχος Kendall μπορεί να αγνοεί τις βραχυχρόνιες τάσεις μετρίου μεγέθους και να τις αποδίδει σε σειριακή συσχέτιση (Loftis και Taylor, 1989).

Όλοι οι έλεγχοι που προαναφέρθηκαν περιλαμβάνουν την υπόθεση ότι η τάση είναι ομοιογενής μεταξύ των εποχών. Όταν η τάση είναι ετερογενής μεταξύ των εποχών, παραδείγματος χάριν, μια ανοδική τάση σε μια εποχή και μια προς καθοδική τάση σε άλλη, ένας συνολικός έλεγχος για την κατεύθυνση της τάσης καθώς και ο εκτιμητής κλίσης θα δίνουν παραπλανητικά αποτελέσματα (Yu και λοιποί, 1993). Επομένως, πριν από τη διεξαγωγή οποιουδήποτε από τους προαναφερθέντες ελέγχους, είναι απαραίτητος ένας προκαταρκτικός έλεγχος για την ομοιογένεια της κατεύθυνσης της τάσης. Για την εξέταση της ομοιογένειας στις εποχικές τάσεις χρησιμοποιείται ο έλεγχος Van Belle and Hughes (Van Belle και Hughes, 1984). Αυτή η διαδικασία περιλαμβάνει τον υπολογισμό του στατιστικού $X^2_{homogeneous}$ το οποίο προσεγγιστικά κατανέμεται ως μία X^2 (Chi-Square) τυχαία μεταβλητή με $m-1$ βαθμούς ελευθερίας, όπου m το πλήθος των εποχών ($m=12$ για μηνιαία δεδομένα). Το στατιστικό $X^2_{homogeneous}$ δίνεται από τον τύπο:

$$X^2_{homogeneous} = \sum Z_i^2 - m\bar{Z}^2$$

Όπου οι τιμές των Z_i και \bar{Z} υπολογίζονται, αντίστοιχα, ως:

$$Z_i = \frac{S_i}{\sqrt{Var(S_i)}} \quad \text{και} \quad \bar{Z} = \frac{1}{m} \sum_{i=1}^m Z_i$$

Όπου S_i είναι το στατιστικό Mann-Kendall για την i εποχή. Εάν η τιμή του $X^2_{homogeneous}$ είναι μεγαλύτερη από την κρίσιμη τιμή της κατανομής X^2 με $m-1$ βαθμούς ελευθερίας, σε επίπεδο σημαντικότητας α , απορρίπτεται η μηδενική υπόθεση της ομογένειας της κατεύθυνσης των εποχικών τάσεων.

ΚΕΦΑΛΑΙΟ 5

ΠΟΛΥΜΕΤΑΒΛΗΤΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ

5.1. ΕΙΣΑΓΩΓΗ – ΣΤΟΧΟΙ ΤΗΣ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ

Οι πολυμεταβλητές στατιστικές τεχνικές έχουν πεδίο εφαρμογής την ταυτόχρονη ανάλυση πολυμεταβλητών δεδομένων, τα οποία προκύπτουν όταν αρκετές αλληλοσχετιζόμενες μεταβλητές μετρώνται για κάθε διερευνώμενη μονάδα (Hair, 1998).

Οι στόχοι της Πολυμεταβλητής Ανάλυσης είναι (Καρλής, 2005; Ντζούφρας, 2001):

1. Μελέτη των σχέσεων μεταξύ των μεταβλητών. Εξετάζονται οι σχέσεις μεταξύ των μεταβλητών για την κατανόηση διάφορων φαινομένων.
2. Μείωση των υπό διερεύνηση δεδομένων/ μεταβλητών. Η πληροφορία που περιέχεται σε πολλές μεταβλητές συμπυκνώνεται σε λίγες περιληπτικές μεταβλητές, με τις οποίες παρέχεται η περαιτέρω δυνατότητα της ερμηνείας αθέατων χαρακτηριστικών της πραγματικότητας.
3. Δημιουργία ομάδων. Με κατάλληλες προσεγγίσεις, τα υπό μελέτη αντικείμενα (παρατηρήσεις ή μεταβλητές) μπορούν να ταξινομηθούν σε ομοειδείς ομάδες ως προς κάποια χαρακτηριστικά, με στόχο την καλύτερη διαχείριση των δεδομένων.
4. Πρόβλεψη. Δημιουργία κατάλληλων μοντέλων που λαμβάνουν υπόψη τους την πολυπλοκότητα της πραγματικότητας και προβλέπουν τη μελλοντική συμπεριφορά ενός συστήματος.
5. Έλεγχος πολυμεταβλητών υποθέσεων και πολύπλοκων επιστημονικών σεναρίων (υποθέσεων). Μέσω σύνθετων μαθηματικών υποδειγμάτων μπορούν να επεκταθούν οι έλεγχοι υποθέσεων σε πολυμεταβλητά δεδομένα με σκοπό τον έλεγχο πιο ρεαλιστικών υποθέσεων.

5.2. ΑΝΑΛΥΣΗ ΣΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ / ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

5.2.1. Εισαγωγή

Η ανάλυση σε κύριες συνιστώσες (Principal components analysis, PCA) και η παραγοντική ανάλυση (Factor analysis, FA) είναι τεχνικές που χρησιμοποιούνται για να συνοψίσουν έναν μεγάλο αριθμό στοιχείων (μεταβλητών) με έναν μικρότερο αριθμό παραγόμενων στοιχείων, αποκαλούμενο συνιστώσες ή παράγοντες. Η ανάλυση σε κύριες συνιστώσες περιλαμβάνει τον σχηματισμό νέων ασυσχέτιστων μεταβλητών οι οποίες είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών. Η παραγοντική ανάλυση, επιπλέον, προσπαθεί να προσδιορίσει μια υποκείμενη δομή των δεδομένων προσπαθώντας να ερμηνεύσει τη συσχέτιση των προκαθορισμένων μεταβλητών. Η ανάλυση σε κύριες συνιστώσες χρησιμοποιείται συχνά σαν ένα κύριο στάδιο στον καθορισμό του αριθμού των κύριων συνιστωσών που θα χρησιμοποιηθούν στην παραγοντική ανάλυση. Οι κύριες συνιστώσες που τελικά επιλέγονται υποβάλλονται συνήθως σε περιστροφή των αξόνων οπότε προκύπτουν οι παράγοντες οι οποίοι παρέχουν τη δυνατότητα ερμηνείας των συσχετίσεων των μεταβλητών που περιέχουν.

Η ανάλυση σε κύριες συνιστώσες έχει ως σκοπό να δημιουργήσει γραμμικούς συνδυασμούς των αρχικών μεταβλητών (κύριες συνιστώσες) έτσι ώστε οι γραμμικοί αυτοί συνδυασμοί να είναι ασυσχέτιστοι μεταξύ τους αλλά να περιέχουν όσο γίνεται μεγαλύτερο μέρος της διακύμανσης των αρχικών μεταβλητών.

Τα οφέλη από την εφαρμογή της ανάλυσης κυρίων συνιστωσών μπορούν να συνοψιστούν στα εξής (Καρλής, 2005; Ντζούφρας, 2001):

- Από ένα σύνολο συσχετισμένων μεταβλητών λαμβάνεται τελικά ένα σύνολο ασυσχέτιστων μεταβλητών. Η δυνατότητα αυτή είναι ιδιαίτερα χρήσιμη για ορισμένες στατιστικές μεθόδους. Με αυτόν τον τρόπο επιλύεται το πρόβλημα της πολυσυγγραμμικότητας στην παλινδρόμηση, όπου αν χρησιμοποιηθούν οι αρχικές συσχετισμένες μεταβλητές οι εκτιμήσεις που θα δοθούν δεν θα είναι συνεπείς.
- Στην περίπτωση κατά την οποία οι κύριες συνιστώσες ερμηνεύουν ένα μεγάλο ποσοστό της διακύμανσης, αντί για τις αρχικές μεταβλητές, μπορούν να χρησιμοποιηθούν οι κύριες συνιστώσες, οι οποίες είναι λιγότερες, με κόστος βέβαια

ότι χάνεται κάποιο ποσοστό της συνολικής μεταβλητότητας. Πολλές φορές, η δυνατότητα αυτή είναι ιδιαίτερα σημαντική. Για παράδειγμα σε μια τεράστια βάση δεδομένων αντί να αποθηκεύονται όλες τις μεταβλητές μπορεί να αποθηκεύεται μόνο κάποιος αριθμός κυρίων συνιστωσών. Παρόλο που χάνεται κάποιο μέρος της πληροφορίας, το κέρδος σε χώρο αλλά και ταχύτητα επεξεργασίας μπορεί να είναι τεράστιο. Επιπλέον, η μείωση των διαστάσεων του προβλήματος επιλύει το πρόβλημα της στατιστικής ανάλυσης δεδομένων με μικρό πλήθος παρατηρήσεων και μεγάλο αριθμό μεταβλητών.

- Ένα άλλο μεγάλο πλεονέκτημα είναι ότι με τη μέθοδο των κυρίων συνιστωσών μπορούν να εξετασθούν οι συσχετίσεις ανάμεσα στις μεταβλητές και να διαπιστωθεί κατά πόσο οι μεταβλητές μοιάζουν ή όχι. Επίσης η μέθοδος επιτρέπει το χαρακτηρισμό των καινούριων μεταβλητών (κύριων συνιστωσών) παρατηρώντας ποιες από τις αρχικές μεταβλητές έχουν μεγάλη επίδραση σε αυτές. Με αυτόν τον τρόπο καθίσταται δυνατή η ποσοτικοποίηση μη μετρήσιμων αφηρημένων εννοιών. Από την άλλη πλευρά, πολλοί συγγραφείς επικρίνουν την διαδικασία ερμηνείας των κύριων συνιστωσών, θεωρώντας ότι εμπεριέχει σε μεγάλο βαθμό υποκειμενικά κριτήρια.

5.2.2. Δημιουργία του πίνακα δεδομένων

Το πρώτο βήμα για την ανάλυση κυρίων συνιστωσών είναι η προσαρμογή των δεδομένων σε έναν πίνακα $n \times p$, όπου n είναι ο αριθμός των δειγμάτων και p ο αριθμός των μεταβλητών (Πίνακας 5.1). Τα στοιχεία του πίνακα είναι οι τιμές του κάθε δείγματος ως προς κάθε μεταβλητή. Στην περίπτωση m επαναληπτικών μετρήσεων σε κάθε δείγμα i για κάθε μεταβλητή j , οι τιμές του πίνακα μπορούν να προκύψουν ως αριθμητικοί μέσοι των παρατηρήσεων των m επαναληπτικών μετρήσεων του δείγματος i για κάθε μεταβλητή j .

Τα στοιχεία του πίνακα μπορούν να παρασταθούν γραφικά με ένα γράφημα p διαστάσεων στο οποίο κάθε σειρά του πίνακα να δίνει τις συντεταγμένες ενός σημείου του γραφήματος. Επομένως, κάθε μεταβλητή (στήλη του πίνακα) θα αποτελεί μία διάσταση σε έναν p -διάστατο χώρο και κάθε σειρά του πίνακα θα δίνει τις συντεταγμένες κάθε σημείου στον χώρο αυτόν.

Πίνακας 5.1. Προσαρμογή των δεδομένων σε πίνακα για τη διεξαγωγή της ανάλυσης σε κύριες συνιστώσες

Δείγματα	Μεταβλητές			
	X_1	X_2	...	X_p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
.	.	.		.
.	.	.		.
.	.	.		.
n	x_{n1}	x_{n2}	...	x_{np}

5.2.3. Βασική ιδέα της μεθόδου

Η βασική ιδέα της μεθόδου βασίζεται στην θεώρηση του ποσοστού της συνολικής διακύμανσης (του αθροίσματος των διακυμάνσεων των p αρχικών μεταβλητών) που ερμηνεύεται από κάθε μία από τις νέες μεταβλητές (Bartholomew και λοιποί, 2002). Η ανάλυση σε κύριες συνιστώσες μετατρέπει ένα σύνολο συσχετισμένων μεταβλητών (X_1, \dots, X_p) σε ένα σύνολο ασυσχέτιστων μεταβλητών (Y_1, \dots, Y_p) οι οποίες καλούνται κύριες συνιστώσες (principal components), έτσι ώστε η πρώτη κύρια συνιστώσα (Y_1) να ερμηνεύει το μεγαλύτερο μέρος της συνολικής διακύμανσης, η δεύτερη κύρια συνιστώσα (Y_2) να ερμηνεύει το μεγαλύτερο μέρος της υπόλοιπης διακύμανσης κ.ο.κ. (Beebe και λοιποί, 1998). Γεωμετρικά, οι κύριες συνιστώσες είναι οι άξονες ενός νέου συστήματος συντεταγμένων το οποίο προκύπτει από την περιστροφή των αξόνων του αρχικού συστήματος των μεταβλητών. Οι νέοι άξονες αντιπροσωπεύουν τις διευθύνσεις της μεγαλύτερης μεταβλητότητας.

5.2.4. Επιλογή του πίνακα που θα χρησιμοποιηθεί για τη φασματική ανάλυση

Η ανάλυση σε κύριες συνιστώσες βασίζεται στη φασματική ανάλυση του πίνακα συνδιακυμάνσεων (covariance matrix) ή του πίνακα συσχετίσεων (correlation matrix) των αρχικών μεταβλητών ο οποίος στην ουσία είναι ο πίνακας συνδιακυμάνσεων των τυποποιημένων μεταβλητών (Massart και λοιποί, 1988). Με την τυποποίηση η κάθε μεταβλητή μετασχηματίζεται έτσι ώστε να έχει αριθμητικό μέσο μηδέν και

διακύμανση ίση με την μονάδα. Η τυποποίηση (standardization) μιας μεταβλητής εκφράζεται με την παρακάτω σχέση:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$$

όπου, x_{ij} είναι το στοιχείο της σειράς j της μεταβλητής i , \bar{x}_i είναι ο αριθμητικός μέσος της μεταβλητής, s_i και z_{ij} είναι η τιμή της τυποποιημένης μεταβλητής.

Η συνδιακύμανση των τυποποιημένων μεταβλητών κυμαίνεται μεταξύ -1 και 1 και αντιστοιχεί στη συσχέτιση των αρχικών μεταβλητών.

Ο πίνακας των συσχετίσεων (ή ισοδύναμα ο πίνακας συνδιακυμάνσεων των τυποποιημένων μεταβλητών) προτιμάται έναντι του πίνακα συνδιακυμάνσεων των αρχικών μεταβλητών στην περίπτωση που οι κλίμακες μέτρησης των μεταβλητών διαφέρουν ή υπάρχει διαφορά μεταξύ των τάξεων μεγέθους ή των διακυμάνσεων των μεταβλητών (Davis, 1986). Αυτή η προσέγγιση είναι η καθιερωμένη επιλογή των περισσότερων στατιστικών (Reimann και λοιποί, 2002). Με τον τρόπο αυτόν δίνονται ίσα βάρη σε όλες τις μεταβλητές και αποφεύγεται η μεγάλη επιρροή μιας μεταβλητής, η οποία έχει πολύ μεγαλύτερη διακύμανση από τις υπόλοιπες, στη δημιουργία των κύριων συνιστωσών (Manley, 1994).

5.2.5. Φασματική ανάλυση του πίνακα συνδιακυμάνσεων ή του πίνακα συσχετίσεων

Οι κύριες συνιστώσες είναι, όπως προαναφέρθηκε, γραμμικός συνδυασμός των αρχικών μεταβλητών και υπολογίζονται από τους τύπους:

$$Y_1 = \alpha_{11}X_1 + \alpha_{21}X_2 + \dots + \alpha_{p1}X_p$$

$$Y_2 = \alpha_{12}X_1 + \alpha_{22}X_2 + \dots + \alpha_{p2}X_p$$

.

.

.

$$Y_p = \alpha_{1p}X_1 + \alpha_{2p}X_2 + \dots + \alpha_{pp}X_p$$

όπου Y_i είναι οι κύριες συνιστώσες, X_i είναι οι μεταβλητές και τα α_{ij} είναι τα βάρη ή συντελεστές για τη μεταβλητή i στην συνιστώσα j και ικανοποιούν τις ακόλουθες συνθήκες:

$$\sum_{i=1}^p \alpha_{ij}^2 = 1, \quad (j = 1, 2, \dots, p)$$

και
$$\sum_{i=1}^p \alpha_{ij} \alpha_{ik} = 0 \quad (j \neq k, \quad j = 1, 2, \dots, p \text{ και } k = 1, 2, \dots, p)$$

Η πρώτη συνθήκη σημαίνει ότι το άθροισμα των τετραγώνων των συντελεστών για κάθε γραμμική εξίσωση πρέπει να ισούται με τη μονάδα. Η δεύτερη συνθήκη εξασφαλίζει ότι οι κύριες συνιστώσες είναι ασυσχέτιστες μεταξύ τους (οι νέοι άξονες είναι κάθετοι μεταξύ τους).

Επιπλέον η πρώτη κύρια (Y_1) συνιστώσα θα πρέπει να ερμηνεύει το μέγιστο της διακύμανσης στα δεδομένα, η δεύτερη (Y_2) να ερμηνεύει τη μέγιστη διακύμανση που δεν ερμηνεύεται από την Y_1 κ.ο.κ..

Επομένως το πρόβλημα εύρεσης των κύριων συνιστωσών ανάγεται στην εύρεση των ιδιοδιανυσμάτων του τετραγωνικού $p \times p$ πίνακα των συνδιακυμάνσεων (\mathbf{C}) ή των συσχετίσεων (\mathbf{R}) των μεταβλητών.

Μετά λοιπόν από την επιλογή του πίνακα \mathbf{C} ή \mathbf{R} πάνω στον οποίο θα στηριχθεί η ανάλυση, υπολογίζονται οι ιδιοτιμές (eigenvalues) και τα ιδιοδιανύσματα (eigenvectors) του πίνακα.

Οι ιδιοτιμές λ του τετραγωνικού $p \times p$ πίνακα \mathbf{P} είναι οι ρίζες της εξίσωσης:

$$|\mathbf{P} - \lambda \mathbf{I}_p| = 0$$

όπου \mathbf{P} είναι ο πίνακας \mathbf{C} ή \mathbf{R} και \mathbf{I}_p είναι ο μοναδιαίος πίνακας. Ο πίνακας \mathbf{P} έχει μέχρι p ιδιοτιμές των οποίων το άθροισμα ισούται με το ίχνος του.

Σε κάθε ιδιοτιμή λ του πίνακα \mathbf{P} αντιστοιχεί και ένα διαφορετικό ιδιοδιάνυσμα, το οποίο είναι ένας πίνακας στήλη \mathbf{A} και υπολογίζεται ως η λύση της εξίσωσης:

$$\mathbf{P}\mathbf{A} = \lambda\mathbf{A}$$

5.2.6. Ποσοστό της συνολικής διακύμανσης που εξηγούν οι κύριες συνιστώσες

Οι ιδιοτιμές λ_i του πίνακα \mathbf{C} ή \mathbf{R} είναι οι διακυμάνσεις των κύριων συνιστωσών. Σε κάθε κύρια συνιστώσα i αντιστοιχεί μία ιδιοτιμή λ_i . Μετά τον υπολογισμό τους, οι ιδιοτιμές διατάσσονται κατά φθίνουσα σειρά. Όταν χρησιμοποιείται ο πίνακας \mathbf{C} τότε το άθροισμα των ιδιοτιμών, δηλαδή το άθροισμα των διακυμάνσεων των κύριων συνιστωσών, ισούται με το άθροισμα των διακυμάνσεων των αρχικών μεταβλητών. Στην περίπτωση κατά την οποία χρησιμοποιείται ο πίνακας συσχετίσεων για τον υπολογισμό των ιδιοτιμών (ή, ισοδύναμα, έχει προηγηθεί τυποποίηση των αρχικών μεταβλητών), τότε το άθροισμα των ιδιοτιμών, δηλαδή το άθροισμα των διακυμάνσεων των κύριων συνιστωσών, ισούται με το άθροισμα των διακυμάνσεων των τυποποιημένων μεταβλητών και συνεπώς ισούται με το πλήθος των μεταβλητών p . Επομένως, το σύνολο των p κύριων συνιστωσών ερμηνεύει πλήρως την συνολική διακύμανση:

$$\sum_{j=1}^p \text{Var}(y_j) = \sum_{i=1}^p \text{Var}(x_i)$$

Η ποσότητα $\frac{\lambda_j}{\sum_{i=1}^p \lambda_i}$ δίνει το ποσοστό της συνολικής διακύμανσης που εξηγεί η j

συνιστώσα. Και η ποσότητα $\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i}$ δίνει το ποσοστό της συνολικής διακύμανσης που

εξηγούν οι πρώτες k συνιστώσες.

5.2.7. Επιλογή του αριθμού κύριων συνιστωσών που επιλέγονται

Σκοπός είναι να επιλεγούν όσο το δυνατόν λιγότερες κύριες συνιστώσες οι οποίες όμως να προσφέρουν μία ικανοποιητική αναπαράσταση των αρχικών δεδομένων. Με άλλα λόγια το κέρδος από τη μείωση των διαστάσεων του προβλήματος να αντισταθμίζει το κόστος από τη μείωση της πληροφορίας. Δυστυχώς δεν υπάρχει μία συγκεκριμένη και κοινώς αποδεκτή λύση για τον αριθμό των κύριων συνιστωσών που

πρέπει να επιλεγθούν. Τα δημοφιλέστερα κριτήρια που χρησιμοποιούνται είναι τα παρακάτω :

- Κριτήριο του Kaiser (1960). Σύμφωνα με το κριτήριο αυτό, πρέπει να επιλέγονται μόνο οι k πρώτες κύριες συνιστώσες των οποίων οι ιδιοτιμές είναι μεγαλύτερες από τη μέση τιμή των ιδιοτιμών $\bar{\lambda} = \sum_{j=1}^k \lambda_j$. Στην περίπτωση κατά την οποία η ανάλυση βασίζεται στον πίνακα συσχετίσεων, τότε $\bar{\lambda} = 1$, οπότε επιλέγονται μόνο οι k πρώτες κύριες συνιστώσες οι οποίες έχουν ιδιοτιμές μεγαλύτερες από 1. Η λογική πάνω στην οποία στηρίζεται αυτός ο εμπειρικός κανόνας είναι ότι μία κύρια συνιστώσα με ιδιοτιμή μεγαλύτερη από 1 εξηγεί το ίδιο ποσό διακύμανσης με μία αρχική μεταβλητή X_i . Παρόλα αυτά ο Jolliffe (1972), θεωρεί ότι είναι προτιμότερο να επιλέγονται οι κύριες συνιστώσες με ιδιοτιμές μεγαλύτερες από 0,7 παρά μόνο αυτές με ιδιοτιμές μεγαλύτερες από 1. Επιπλέον, ο Cliff (1988) θεωρεί ότι το συγκεκριμένο κριτήριο ενδέχεται να οδηγήσει στην επιλογή περισσότερων ή λιγότερων κύριων συνιστωσών απ' ό,τι είναι απαραίτητο και συνεπώς πρέπει να χρησιμοποιείται σε συνδυασμό με άλλα κριτήρια.
- Γράφημα των ιδιοτιμών (scree plot). Το scree plot είναι το γράφημα των ιδιοτιμών έναντι του αριθμού των κύριων συνιστωσών και αρχικά προτάθηκε από τον Cattell (1966). Σύμφωνα με το κριτήριο αυτό, επιλέγονται τόσες συνιστώσες μέχρι του αριθμού για τον οποίο αρχίζει να αλλάζει η κλίση του γραφήματος.
- Επιλέγονται οι k πρώτες κύριες συνιστώσες έτσι ώστε το ποσοστό της συνολικής διακύμανσης που εξηγούν να είναι μεγαλύτερο από κάποιο συγκεκριμένο όριο (συνήθως 70 – 80%).

5.2.8. Ερμηνεία των κύριων συνιστωσών

Τα σχετικά μεγέθη των τιμών των ιδιοδιανυσμάτων a_{ij} 's απεικονίζουν τη σχετική συνεισφορά κάθε μίας από τις μεταβλητές i στην j κύρια συνιστώσα. Επομένως, η ερμηνεία μιας κύριας συνιστώσας βασίζεται στο πρότυπο των συντελεστών για την συνιστώσα αυτή.

Συχνά συνηθίζεται η αλλαγή κλίμακας των συντελεστών, έτσι ώστε οι συντελεστές των περισσότερο σημαντικών κύριων συνιστωσών (δηλαδή αυτών που εξηγούν το

μεγαλύτερο μέρος της διακύμανσης) να είναι μεγαλύτεροι από τους συντελεστές των λιγότερο σημαντικών κύριων συνιστωσών (Bartholomew και λοιποί, 2002). Οι νέοι συντελεστές που προκύπτουν ονομάζονται «φορτίσεις των συνιστωσών» (component loadings), μπορούν να χρησιμοποιηθούν για την ανακατασκευή των αρχικών μεταβλητών από τις κύριες συνιστώσες και υπολογίζονται σύμφωνα με τη σχέση:

$$\alpha_{ij}^* = \sqrt{\lambda_j} \alpha_{ij}, \text{ όπου } i = 1, \dots, p \text{ και } j = 1, \dots, p.$$

Όταν η ανάλυση βασίζεται στον πίνακα συσχετίσεων των αρχικών μεταβλητών, τότε τα α_{ij}^* μπορούν να θεωρηθούν ως οι συντελεστές συσχέτισης μεταξύ της μεταβλητής i και της συνιστώσας j . Αυτό είναι ιδιαίτερα χρήσιμο για την ερμηνεία των συνιστωσών. Όσο μεγαλύτερη είναι η συσχέτιση μιας μεταβλητής με μία κύρια συνιστώσα, τόσο περισσότερο συνεισφέρει η μεταβλητή στο σχηματισμό της συγκεκριμένης συνιστώσας. Επομένως, οι συσχετίσεις μπορούν να χρησιμοποιηθούν για τον εντοπισμό των μεταβλητών οι οποίες εκφράζονται και περιγράφονται από μία κύρια συνιστώσα και στη συνέχεια να αποδοθεί κάποιο όνομα στην κύρια συνιστώσα βάσει του συνόλου των μεταβλητών που συσχετίζονται ισχυρά με αυτήν.

5.2.9. Περιστροφή των αξόνων

Προκειμένου να επιτευχθεί καλύτερη ερμηνεία των συνιστωσών, συνίσταται η περιστροφή των αξόνων. Η περιστροφή είναι ουσιαστικά ο πολλαπλασιασμός του πίνακα των συντελεστών που υπολογίστηκε με έναν ορθογώνιο πίνακα. Από τους άπειρους ορθογώνιους πίνακες μπορεί να επιλεγεί κάποιος με βάση κριτήρια βελτιστοποίησης, όπως για παράδειγμα κάθε συνιστώσα να έχει όσο γίνεται λιγότερες μεταβλητές με μεγάλους συντελεστές (Καρλής, 2005; Ντζούφρας, 2001). Με την περιστροφή συνήθως επιτυγχάνεται σε κάθε συνιστώσα να υπάρχουν λίγες μεταβλητές με μεγάλες απόλυτες τιμές των συντελεστών (φορτίσεις των συνιστωσών) ενώ οι υπόλοιπες μεταβλητές να τείνουν να έχουν συντελεστές κοντά στο μηδέν.

Η περιστροφή των αξόνων διατηρεί το συσσωρευτικό ποσοστό της διακύμανσης που εξηγείται από τις εξαγόμενες συνιστώσες, αλλά τώρα η διακύμανση είναι εξαπλωμένη πιο ομοιόμορφα στις συνιστώσες. Οι μεγάλες αλλαγές στα επιμέρους

σύνολα προτείνουν ότι η περιστροφή κάνει ευκολότερη την ερμηνεία των συνιστωσών.

5.2.10. Ανακατασκευή των αρχικών μεταβλητών από τις κύριες συνιστώσες

Οι αρχικές μεταβλητές μπορούν να ανακατασκευαστούν από τις κύριες συνιστώσες βάσει της σχέσης:

$$X_i = \alpha_{i1}^* Y_1 + \dots + \alpha_{ip}^* Y_p, \quad (i = 1, \dots, p)$$

Αν υποθεθεί ότι επιλέγονται μόνο οι k πρώτες κύριες συνιστώσες, τότε οι εκτιμήσεις των ανακατασκευασμένων μεταβλητών X_i θα δίνονται από τον τύπο:

$$\hat{X}_i = \alpha_{i1}^* Y_1 + \dots + \alpha_{ik}^* Y_k, \quad (i = 1, \dots, k)$$

Ένας τρόπος για να διαπιστωθεί πόσο ικανοποιητικά αναπαράγεται κάθε αρχική τυποποιημένη μεταβλητή X_i από τις πρώτες k συνιστώσες, είναι μέσω των συμμετοχικοτήτων (communalities), δηλαδή του αθροίσματος των τετραγώνων των φορτίσεων των k πρώτων συνιστωσών. Για την μεταβλητή X_i η συμμετοχικότητα δίνεται από τον τύπο:

$$\alpha_{i1}^{*2} + \dots + \alpha_{ik}^{*2}, \quad (i = 1, \dots, k)$$

Οι συμμετοχικότητες δείχνουν το ποσό της διακύμανσης κάθε μεταβλητής που εξηγείται από τις κύριες συνιστώσες. Οι αρχικές συμμετοχικότητες είναι εκτιμήσεις της διακύμανσης κάθε μεταβλητής που εξηγείται από όλες τις συνιστώσες. Όταν η ανάλυση βασίζεται στον πίνακα συσχετίσεων, οι αρχικές συμμετοχικότητες είναι πάντοτε ίσες με τη μονάδα. Οι εξαγόμενες συμμετοχικότητες είναι οι εκτιμήσεις της διακύμανσης κάθε μεταβλητής που εξηγείται από τις k πρώτες κύριες συνιστώσες. Δεδομένου ότι οι φορτίσεις των συνιστωσών είναι οι συσχετίσεις μεταξύ των μεταβλητών και των συνιστωσών και οι συνιστώσες είναι ορθογώνιες, η συμμετοχικότητα μιας μεταβλητής αντιπροσωπεύει τον συντελεστή προσδιορισμού R^2 της μεταβλητής X_i που εκτιμάται από τις συνιστώσες Y_1, \dots, Y_k . Για να αντιπροσωπεύουν οι επιλεγμένες κύριες συνιστώσες καλά τις αρχικές μεταβλητές θα πρέπει οι εξαγόμενες συμμετοχικότητες να είναι υψηλές (μεγαλύτερες από 0,5). Εάν κάποια εξαγόμενη συμμετοχικότητα είναι μικρή, πιθανόν να χρειάζεται να εξαχθεί και επιπλέον συνιστώσα.

5.2.11. Προϋποθέσεις για την ανάλυση σε κύριες συνιστώσες και την παραγοντική ανάλυση

Η ανάλυση σε κύριες συνιστώσες είναι μία μαθηματική μέθοδος για τον περιορισμό των δεδομένων και δεν προϋποθέτει κάποια συγκεκριμένη κατανομή για τα δεδομένα. Για την παραγοντική ανάλυση που βασίζεται στην ανάλυση σε κύριες συνιστώσες (PCA/FA), παρόλο που δεν θεωρείται απαραίτητη η κανονική κατανομή των δεδομένων, η μέθοδος βασίζεται στον πίνακα συνδιακυμάνσεων ή τον πίνακα συσχετίσεων οι οποίοι επηρεάζονται ισχυρά από την έλλειψη κανονικότητας των δεδομένων καθώς και την ύπαρξη ακραίων τιμών. Επομένως, συνιστάται οι μεταβλητές που θα χρησιμοποιηθούν σε PCA/FA να προσεγγίζουν όσο το δυνατόν γίνεται την κανονική κατανομή (Reimann και Filzmoser, 2000). Ωστόσο, γενικά τα περιβαλλοντικά δεδομένα δεν ακολουθούν κανονική κατανομή (Reimann and Filzmoser, 2000). Ο λογαριθμικός μετασχηματισμός χρησιμοποιείται ευρύτατα ώστε οι μεταβλητές που αφορούν σε ποιοτικά χαρακτηριστικά του νερού να προσεγγίσουν την κανονική κατανομή.

5.2.12. Διαφορές μεταξύ ανάλυσης σε κύριες συνιστώσες και παραγοντικής ανάλυσης

Γενικά, υπάρχει συχνά σύγχυση μεταξύ της ανάλυσης σε κύριες συνιστώσες (PCA) και της παραγοντικής ανάλυσης (FA). Πολλοί στατιστικοί χρησιμοποιούν τις δύο μεθόδους χωρίς να θεωρούν ότι υπάρχει κάποια διάκριση μεταξύ τους. Το πρόβλημα επιδεινώνεται κυρίως στην περίπτωση, που αναλύθηκε παραπάνω, κατά την οποία, ως μέθοδος εξαγωγής των παραγόντων κατά την παραγοντική ανάλυση, χρησιμοποιείται η ανάλυση κύριων συνιστωσών (PCA/FA). Στην, PCA οι συνιστώσες υπολογίζονται ως γραμμικοί συνδυασμοί των αρχικών μεταβλητών. Αντίθετα, στην FA, οι αρχικές μεταβλητές ορίζονται ως γραμμικοί συνδυασμοί των παραγόντων. Στόχος της PCA είναι η επεξήγηση όσον το δυνατόν μεγαλύτερου ποσοστού της συνολικής διακύμανσης των μεταβλητών, ενώ το αντικείμενο της FA είναι η επεξήγηση της συνδιακύμανσης ή των συσχετίσεων μεταξύ των μεταβλητών. Συχνά, διαχωρίζεται η PCA/FA από την επονομαζόμενη «κοινή παραγοντική ανάλυση» (Common Factor Analysis) (Bartholomew και λοιποί, 2002; SYSTAT 12, 2007; Hair και λοιποί, 1998). Στην «κοινή παραγοντική ανάλυση», λαμβάνεται υπόψη μόνο η κοινή διακύμανση

μεταξύ των μεταβλητών και οι συνηθέστερες μέθοδοι εκτίμησης των παραγόντων είναι η μέθοδος των κυρίων αξόνων (Principal Axis) και η μέθοδος μεγίστης πιθανοφάνειας (Maximum Likelihood). Μερικοί υποστηρίζουν ότι η χρήση της PCA για την εξαγωγή των παραγόντων δεν αποτελεί αληθινή μέθοδο για FA και συνιστούν τον περιορισμό της χρήσης της και την προτίμηση μίας αληθινής μεθόδου FA (Bentler και Kano, 1990; Floyd και Widaman, 1995; Ford και λοιποί, 1986; Gorsuch, 1990; Loehlin, 1990; MacCallum και Tucker, 1991; Mulaik, 1990; Snook και Gorsuch, 1989; Widaman, 1990; Widaman, 1993; Costello και Osborne, 2005). Άλλοι διαφωνούν, και επισημαίνουν ότι δεν υπάρχει σχεδόν καμία διαφορά μεταξύ PCA και FA ή ότι είναι προτιμότερη η PCA (Arrindell και van der Ende, 1985; Guadagnoli και Velicer, 1988; Schoenmann, 1990; Steiger, 1990; Velicer και Jackson, 1990).

Στην στατιστική ανάλυση δεδομένων ποιότητας του νερού η ευρύτερα χρησιμοποιούμενη μέθοδος είναι η PCA/FA (Kannel και λοιποί, 2006; Najafpour και λοιποί, 2008; Lambrakis και λοιποί, 2004; De Ceballos και λοιποί, 1998; Voutsas και λοιποί, 2001; Vega και λοιποί, 1998; Simeonov και λοιποί, 2002; Kotti και λοιποί, 2005; Shrestha και λοιποί, 2007; Paratheodorou και λοιποί, 2006; Kumar και Riyazuddin, 2008; Wuderlin και λοιποί, 2001; Spanos και λοιποί, 2002; Qadir και λοιποί, 2008; Ouyang και λοιποί, 2006; Baralkiewicz και λοιποί, 2008; Singh και λοιποί, 2004; Simeonov και λοιποί, 2001; Singh και λοιποί, 2005; Helena και λοιποί, 2000).

5.3. ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ

5.3.1. Εισαγωγή

Η ανάλυση κατά συστάδες (Cluster analysis, CA) είναι μια πολυμεταβλητή μέθοδος για τον εντοπισμό φυσικών ομαδοποιήσεων των δεδομένων. Οι παρατηρήσεις ομαδοποιούνται σε ομάδες – συστάδες κατά τέτοιο τρόπο ώστε κάθε ομάδα να είναι όσο το δυνατόν πιο ομοιογενής (παρόμοια αντικείμενα να ανήκουν στην ίδια ομάδα) και οι ομάδες να διαφέρουν μεταξύ τους όσο το δυνατόν περισσότερο.

Η ανάλυση κατά συστάδες δεν έχει προϋποθέσεις για τις κατανομές των μεταβλητών. Στην περίπτωση διαφορετικών μονάδων μέτρησης στις μεταβλητές, συνιστάται η

τυποποίηση των δεδομένων πριν την ανάλυση κατά συστάδες (Rogerson, 2001). Η τυποποίηση των δεδομένων εξασφαλίζει ότι κάθε μεταβλητή έχει την ίδια επίδραση στην ανάλυση.

Η ανάλυση κατά συστάδες αποσκοπεί στην κατάταξη υπάρχουσών παρατηρήσεων σε ομάδες χρησιμοποιώντας την πληροφορία που υπάρχει σε κάποιες μεταβλητές. Η μέθοδος, εξετάζοντας πόσο όμοιες είναι κάποιες παρατηρήσεις ως προς κάποιο αριθμό μεταβλητών, τείνει να δημιουργεί ομάδες από παρατηρήσεις που μοιάζουν μεταξύ τους (Massart και Kaufman, 1983). Μια επιτυχημένη ανάλυση καταλήγει σε ομάδες για τις οποίες οι παρατηρήσεις μέσα σε κάθε ομάδα να είναι όσο γίνεται πιο ομοιογενείς αλλά οι παρατηρήσεις διαφορετικών ομάδων διαφέρουν όσο γίνεται περισσότερο. Οι βασικότερες προσεγγίσεις για την ομαδοποίηση των δεδομένων είναι οι ιεραρχικές μέθοδοι, η μέθοδος K-Means και παραλλαγές τους.

Η σημαντική διαφορά της μεθόδου από τη διαχωριστική ανάλυση, η οποία εξετάζεται σε επόμενη ενότητα, είναι πως στη διαχωριστική ανάλυση είναι γνωστή η ομαδοποίηση ως προς κάποιο χαρακτηριστικό των παρατηρήσεων και επιζητείται η δημιουργία συγκεκριμένου κανόνα που θα βοηθήσει στην κατάταξη καινούριων παρατηρήσεων. Καθώς λοιπόν οι δύο μέθοδοι έχουν κάποια κοινά χαρακτηριστικά ως προς τον τρόπο που λειτουργούν, μπορούν να λειτουργήσουν συμπληρωματικά.

5.3.2. Απόσταση και ομοιότητα

Δύο πολύ βασικές έννοιες για την ανάλυση κατά συστάδες είναι οι έννοιες της απόστασης και της ομοιότητας. Οι δύο αυτές έννοιες είναι αντίθετες. Παρατηρήσεις που είναι όμοιες θα έχουν μεγάλη ομοιότητα και μικρή απόσταση. Οι έννοιες αυτές είναι πολύ χρήσιμες καθώς επιτρέπουν τη μέτρηση της ομοιότητας των παρατηρήσεων μεταξύ τους και επομένως την τοποθέτηση τους στην ίδια ομάδα. Η επιλογή της απόστασης έχει να κάνει με τη μέθοδο που θα χρησιμοποιηθεί, τον τύπο των δεδομένων καθώς και τα ίδια τα δεδομένα. Για δεδομένα συνεχής η ευκλείδεια απόσταση ή η τετραγωνική ευκλείδεια απόσταση είναι οι συχνότερα χρησιμοποιούμενες. Αν μία από τις μεταβλητές έχει όμως τεράστια διακύμανση σε σχέση με τις υπόλοιπες, αυτή θα παίζει σπουδαιότερο ρόλο και άρα θα κατευθύνει και

τα αποτελέσματα. Σε αυτή την περίπτωση καλό είναι τα δεδομένα να τυποποιηθούν ώστε να έχουν ίδια μέση τιμή και διακύμανση (άρα και ειδικό βάρος).

5.3.3. Ιεραρχική ομαδοποίηση

Οι πιο διαδεδομένοι ιεραρχικοί αλγόριθμοι είναι οι συσσωρευτικοί (agglomerative). Στις συσσωρευτικές ιεραρχικές μεθόδους, αρχικά κάθε παρατήρηση είναι από μόνη της μια ομάδα. Σε κάθε βήμα ενώνονται δύο παρατηρήσεις ή μία παρατήρηση με μία προϋπάρχουσα ομάδα ή δύο προϋπάρχουσες ομάδες, που έχουν την πιο μικρή απόσταση. Η μορφή των αποτελεσμάτων προσδιορίζει τον αριθμό των ομάδων που θα επιλεγούν. Στην πραγματικότητα τα κριτήρια επιλογής του αριθμού των ομάδων είναι πολλά, αλλά μερικές φορές η ερμηνεία που μπορεί να δοθεί είναι ο καλύτερος οδηγός για την επιλογή του αριθμού των ομάδων. Το κρίσιμο σημείο για τον ιεραρχικό αλγόριθμο είναι πως θα υπολογιστεί η απόσταση της ομάδας που κατασκευάστηκε (είτε από συγχώνευση άλλων ομάδων είτε από συγχώνευση παρατηρήσεων). Υπάρχουν πολλές μέθοδοι, όπως:

- Η μέθοδος του κοντινότερου γείτονα (nearest neighbour or single linkage)
- Η μέθοδος του μακρινότερου γείτονα (furthest neighbour or complete linkage)
- Η μέθοδος του μέσου ανάμεσα στις ομάδες (Average between groups)
- Η μέθοδος του μέσου μέσα στις ομάδες (Average within groups)
- Η μέθοδος του Ward

Από αυτές, η πιο απλή είναι η μέθοδος του κοντινότερου γείτονα η οποία όμως έχει το μειονέκτημα πως δίνει ομάδες με μεγάλες διαφορές ως προς το μέγεθός τους. Η μέθοδος του Ward έχει το πλεονέκτημα ότι δίνει περίπου ισοπληθείς ομάδες και για αυτό χρησιμοποιείται συχνά στην πράξη. Από πειράματα προσομοίωσης έχει βρεθεί ότι η μέθοδοι με την καλύτερη επίδοση είναι η Ward's και η Average Linkage ενώ η μέθοδος του κοντινότερου γείτονα παρουσιάζει τη χειρότερη επίδοση (Καρλής, 2005).

5.3.4. Η μέθοδος k-means

Στην μέθοδο K-Means, ο αριθμός των ομάδων πρέπει να είναι γνωστός εκ των προτέρων. Η μέθοδος δουλεύει επαναληπτικά. Χρησιμοποιεί την έννοια του κέντρου της ομάδας (centroid) και στη συνέχεια κατατάσσει τις παρατηρήσεις ανάλογα με την απόστασή τους από τα κέντρα όλων των ομάδων. Το κέντρο της ομάδας δεν είναι τίποτα άλλο από τη μέση τιμή, ή άλλο μέτρο θέσης, για κάθε μεταβλητή όλων των παρατηρήσεων της ομάδας. Συνήθως χρησιμοποιείται η ευκλείδεια απόσταση. Ο αλγόριθμος K-Means δουλεύει ικανοποιητικά για μεγάλα σετ δεδομένων επειδή σε αυτή την περίπτωση είναι πολύ πιο γρήγορος από την ιεραρχική ομαδοποίηση.

5.4. ΔΙΑΧΩΡΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

5.4.1. Αντικείμενο της διαχωριστικής ανάλυσης

Η διαχωριστική (Discriminant Analysis, DA) είναι μια πολυμεταβλητή στατιστική μέθοδος, η οποία σχετίζεται με την πολλαπλή παλινδρόμηση, μέσω της οποίας προσδιορίζονται γραμμικές εξισώσεις οι οποίες διαχωρίζουν με τον καλύτερο δυνατό τρόπο τις παρατηρήσεις σε δύο ή περισσότερες φυσικά προκαθορισμένες ομάδες. Η διαχωριστική ανάλυση χρησιμοποιείται για τους παρακάτω ερευνητικούς σκοπούς:

- Τον προσδιορισμό τυχόν στατιστικών σημαντικών διαφορών που υπάρχουν μεταξύ των μέσων επιδόσεων δύο ή περισσότερων προκαθορισμένων ομάδων δεδομένων.
- Τον προσδιορισμό των ανεξάρτητων μεταβλητών που ευθύνονται περισσότερο για τις διαφορές των μέσων επιδόσεων δύο ή περισσότερων προκαθορισμένων ομάδων δεδομένων.
- Για τον καθορισμό διαδικασιών για την ταξινόμηση αντικειμένων σε ομάδες βάσει των σκορ που επιδεικνύουν σε ένα σύνολο ανεξαρτήτων μεταβλητών.
- Για τον καθορισμό του αριθμού και τη σύνθεση των διαστάσεων της διάκρισης μεταξύ ομάδων, βάσει ενός συνόλου ανεξάρτητων μεταβλητών.

Συνεπώς, η διαχωριστική ανάλυση είναι χρήσιμη τόσο στην κατανόηση των διαφορών μεταξύ ομάδων όσο και στην ορθή ταξινόμηση αντικειμένων σε ομάδες ή κλάσεις.

5.4.2. Ερευνητικός σχεδιασμός για τη διαχωριστική ανάλυση

Η διαχωριστική ανάλυση είναι αρκετά ευαίσθητη όσον αφορά στο λόγο του μεγέθους του δείγματος προς τον αριθμό των ανεξάρτητων μεταβλητών. Τα αποτελέσματα τείνουν να είναι μη σταθερά όταν το μέγεθος του δείγματος μικραίνει σε σχέση με τον αριθμό των ανεξάρτητων μεταβλητών. Το ελάχιστο μέγεθος του δείγματος συνιστάται να είναι τριπλάσιο του αριθμού των ανεξάρτητων μεταβλητών (Massart και λοιποί, 1988). Επιπλέον, το μέγεθος της μικρότερης ομάδας θα πρέπει να υπερβαίνει τον αριθμό των ανεξάρτητων μεταβλητών. Σαν ένας πρακτικός οδηγός, κάθε ομάδα θα πρέπει να έχει τουλάχιστον 20 παρατηρήσεις. Επιπροσθέτως, τα σχετικά μεγέθη των επιμέρους ομάδων δεν θα πρέπει να διαφέρουν κατά πολύ καθώς οι μεγάλες σχετικές διαφορές μπορεί να επηρεάσουν την εκτίμηση των διαχωριστικών συναρτήσεων και την ταξινόμηση των παρατηρήσεων (οι μεγαλύτερες ομάδες παρουσιάζουν δυσανάλογα μεγαλύτερη τυχαία ταξινόμηση των παρατηρήσεων σε αυτές).

5.4.3. Προϋποθέσεις για τη διαχωριστική ανάλυση και διορθωτικές ενέργειες

Η σωστή εφαρμογή της διαχωριστικής ανάλυσης προϋποθέτει οι ανεξάρτητες μεταβλητές να έχουν πολυμεταβλητή κανονική κατανομή και άγνωστες πλην όμως ίσες διακυμάνσεις και συνδιακυμάνσεις σε κάθε ομάδα που προκαθορίζεται από την εξαρτημένη κατηγορική μεταβλητή (Green, 1978; Harris, 1975; Hair και λοιποί, 1998). Η έλλειψη πολυμεταβλητής κανονικότητας μπορεί να προκαλέσει προβλήματα στην εκτίμηση των διαχωριστικών συναρτήσεων. Παρόλα αυτά, η προϋπόθεση της πολυμεταβλητής κανονικότητας δεν είναι κρίσιμα σημαντική. Σύμφωνα με τον Ainsworth (2004), εάν στη μικρότερη ομάδα υπάρχουν τουλάχιστον 20 παρατηρήσεις, η διαχωριστική ανάλυση είναι ανθεκτική στις παραβιάσεις της πολυμεταβλητής κανονικότητας των δεδομένων ακόμα και στην περίπτωση που οι ομάδες είναι άνισου μεγέθους. Όταν σκοπός της διαχωριστικής ανάλυσης είναι η ταξινόμηση, τότε η ανάλυση επηρεάζεται από παραβιάσεις της υπόθεσης της

ομοιογένειας επειδή τα αντικείμενα τείνουν να ταξινομούνται στις ομάδες με τη μεγαλύτερη διασπορά. Αυτό μπορεί να αποτιμηθεί σχεδιάζοντας το διάγραμμα των διαχωριστικών συναρτήσεων για τις δύο πρώτες τουλάχιστον συναρτήσεις και συγκρίνοντας τις ομάδες προκειμένου να φανεί αν έχουν περίπου το ίδιο μέγεθος και διάχυση. Ο έλεγχος της υπόθεσης ομοιογένειας μπορεί επίσης να ελεγχθεί με τον στατιστικό έλεγχο Box's M, ο οποίος ελέγχει την μηδενική υπόθεση ότι οι πίνακες διακύμανσης – συνδιακύμανσης των ομάδων είναι ίσοι. Θα πρέπει να επισημανθεί ότι ο έλεγχος Box's M είναι ευαίσθητος στην έλλειψη κανονικότητας των μεταβλητών και το μεγάλο μέγεθος του δείγματος με αποτέλεσμα να δίνει συχνά στατιστικά σημαντικό αποτέλεσμα. Για αυτόν το λόγο, συνήθως χρησιμοποιείται σαν επίπεδο σημαντικότητας του ελέγχου το $\alpha=0,001$ αντί του συνηθισμένου $\alpha=0,05$. Στην περίπτωση παραβίασης της υπόθεσης της ομοιογένειας, συνιστάται ο μετασχηματισμός των δεδομένων, η χρησιμοποίηση ξεχωριστών πινάκων συνδιακύμανσης, η τετραγωνική διαχωριστική ανάλυση (quadratic discriminant analysis – QDA) ή μη παραμετρικές προσεγγίσεις για την ταξινόμηση των παρατηρήσεων (Ainsworth, 2004; Hair και λοιποί, 1998).

Η επίδραση των άνισων πινάκων συνδιακύμανσης μπορεί να ελαχιστοποιηθεί με την αύξηση του μεγέθους του δείγματος καθώς επίσης με τη χρησιμοποίηση ξεχωριστών πινάκων συνδιακύμανσης της κάθε ομάδας κατά το στάδιο της ταξινόμησης (Hair και λοιποί, 1998). Εάν η ταξινόμηση χρησιμοποιώντας ξεχωριστούς πίνακες συνδιακύμανσης είναι κατά 2% ή περισσότερο ακριβής σε σχέση με την ταξινόμηση που βασίζεται στον κοινό πίνακα συνδιακύμανσης μέσα στις ομάδες, θα πρέπει να αναφέρονται τα αποτελέσματα της ταξινόμησης που βασίζεται στο μοντέλο των ξεχωριστών πινάκων συνδιακύμανσης (Schwab, 2007). Η τετραγωνική διαχωριστική ανάλυση (quadratic discriminant analysis – QDA), κατά την οποία δεν χρησιμοποιείται κοινός πίνακας συνδιακύμανσης, παρέχεται από αρκετά στατιστικά προγράμματα (Gessner και λοιποί, 1988; Huberty, 1984; Johnson και Wichern, 1998).

Οι δυσκολίες λόγω των μη ευνοϊκών στατιστικών χαρακτηριστικών επηρεάζουν μόνο τα όρια και επομένως την ταξινόμηση με τη διαχωριστική ανάλυση και όχι τον προσδιορισμό των διαχωριστικών συναρτήσεων. Επομένως οι διαχωριστικές συναρτήσεις μπορούν να χρησιμοποιηθούν τελείως ικανοποιητικά για την παρουσίαση των δεδομένων (Massart και λοιποί, 1988).

Επιπλέον η διαχωριστική ανάλυση προϋποθέτει την έλλειψη πολυσυγγραμμικότητας μεταξύ των ανεξάρτητων μεταβλητών ιδιαίτερα όταν χρησιμοποιείται η βηματική μέθοδος για τον προσδιορισμό των διαχωριστικών συναρτήσεων. Η βηματική διαχωριστική ανάλυση προσδιορίζει τις μεταβλητές ή τους δείκτες που επιτρέπουν τον βέλτιστο διαχωρισμό μεταξύ δύο ή περισσότερων προκαθορισμένων ομάδων δειγμάτων (Hair και λοιποί, 1998; Johnson και Wichern, 1998).

5.4.4. Εκτίμηση των διαχωριστικών συναρτήσεων και αποτίμηση της καταλληλότητας του μοντέλου

Η διαχωριστική ανάλυση εμπεριέχει την εξαγωγή των διαχωριστικών συναρτήσεων οι οποίες είναι γραμμικοί συνδυασμοί δύο ή περισσότερων ανεξάρτητων μεταβλητών που διαχωρίζουν καλύτερα τις προκαθορισμένες ομάδες. Μία διαχωριστική συνάρτηση λαμβάνεται σύμφωνα με τον παρακάτω τύπο:

$$Z_{jk} = \alpha + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk}$$

Όπου,

Z_{jk} = η διαχωριστική τυποποιημένη τιμή της *διαχωριστικής συνάρτησης j* για το αντικείμενο *k*,

α = σταθερά,

W_i = ο διαχωριστικός συντελεστής της ανεξάρτητης μεταβλητής *i*,

X_{ik} = η τιμή της ανεξάρτητης μεταβλητής *i* για το αντικείμενο *k*

Ο αριθμός των διαχωριστικών συναρτήσεων οι οποίες εκτιμώνται με τη διαχωριστική ανάλυση είναι κατά μία μονάδα μικρότερος από τον αριθμό των ομάδων. Μία διαχωριστική συνάρτηση εκτιμάται χρησιμοποιώντας τις εξής δύο προσεγγίσεις: Την ταυτόχρονη εισαγωγή όλων των μεταβλητών και τη βηματική προσέγγιση (stepwise). Με την προσέγγιση της ταυτόχρονης εκτίμησης, η διαχωριστική συνάρτηση εκτιμάται κατά τέτοιο τρόπο ώστε όλες οι ανεξάρτητες μεταβλητές να θεωρείται ότι δρουν ταυτόχρονα. Συνεπώς ο υπολογισμός της διαχωριστικής συνάρτησης βασίζεται σε ολόκληρο το σετ των ανεξάρτητων μεταβλητών ανεξάρτητα από τη διαχωριστική ισχύ κάθε ανεξάρτητης μεταβλητής (Hair και λοιποί, 1998). Η προσέγγιση αυτή είναι κατάλληλη, όταν ο ερευνητής, για

θεωρητικούς λόγους, θέλει να συμπεριλάβει όλες τις ανεξάρτητες μεταβλητές στην ανάλυση και δεν ενδιαφέρεται να δει ενδιάμεσα αποτελέσματα που βασίζονται μόνο στις μεταβλητές με τη μεγαλύτερη διαχωριστική ικανότητα. Εναλλακτικά, κατά τη βηματική προσέγγιση, οι ανεξάρτητες μεταβλητές εισέρχονται στη διαχωριστική συνάρτηση μία κάθε φορά βάσει της διαχωριστικής τους ισχύος. Κατά τη βηματική προσέγγιση, επιλέγεται αρχικά η μία μοναδική καλύτερη διαχωριστική μεταβλητή. Στη συνέχεια, η αρχική αυτή μεταβλητή συνδυάζεται ξεχωριστά με κάθε μία από τις υπόλοιπες ανεξάρτητες μεταβλητές και επιλέγεται από αυτές η μεταβλητή η οποία είναι η πιο ικανή, σε συνδυασμό με την πρώτη μεταβλητή, να βελτιώσει τη διαχωριστική ικανότητα της συνάρτησης. Η τρίτη καθώς και κάθε επόμενη μεταβλητή επιλέγονται με παρόμοιο τρόπο. Καθώς επιπρόσθετες μεταβλητές συμπεριλαμβάνονται, κάποιες από τις μεταβλητές που είχαν επιλεγεί σε προηγούμενο στάδιο μπορεί να απομακρυνθούν εάν οι πληροφορίες που περιέχουν σχετικά με τις διαφορές μεταξύ των ομάδων είναι διαθέσιμες σε κάποιο συνδυασμό των άλλων μεταβλητών που έχουν συμπεριληφθεί σε επόμενα στάδια. Τελικά, ή όλες οι ανεξάρτητες μεταβλητές θα έχουν εισαχθεί στην συνάρτηση ή οι μεταβλητές που έχουν εξαιρεθεί θα έχουν κριθεί ως μη σημαντικά συμβάλλουσες σε επιπλέον διαχωρισμό ή διάκριση. Η βηματική προσέγγιση είναι χρήσιμη όταν ο ερευνητής θέλει να θεωρήσει έναν σχετικά μεγάλο αριθμό ανεξάρτητων μεταβλητών προς εισαγωγή στην συνάρτηση. Η τεχνική αυτή προσδιορίζει τις μεταβλητές που επιτρέπουν τον καλύτερο διαχωρισμό δύο ή περισσότερων ομάδων (Dixon και λοιποί, 1981; Dillon και λοιποί, 1984; Hair και λοιποί, 1998; Johnson και Wichern, 1998; Sharma, 1996). Σύμφωνα με αυτήν την προσέγγιση, σε κάθε βήμα επιλέγεται η αμέσως επόμενη μεταβλητή που συμβάλλει σε καλύτερο διαχωρισμό έτσι ώστε οι μεταβλητές που δεν είναι χρήσιμες στον διαχωρισμό ανάμεσα στις ομάδες να αποκλείονται και τελικά να προσδιορίζεται ένα μειωμένο σετ μεταβλητών. Το μειωμένο αυτό σετ μεταβλητών είναι τυπικά σχεδόν τόσο καλό ή και μερικές φορές καλύτερο από το συνολικό σετ των μεταβλητών. Παρόλα αυτά αξίζει να σημειωθεί ότι η βηματική προσέγγιση μπορεί να δώσει λιγότερο σταθερά και γενικεύσιμα αποτελέσματα όταν το μέγεθος του δείγματος είναι μικρό σε σχέση με τον αριθμό των ανεξάρτητων μεταβλητών. Συνεπώς, θεωρείται σημαντική η επικύρωση των αποτελεσμάτων.

Τα στατιστικά κριτήρια τα οποία χρησιμοποιούνται για την εισαγωγή ή την απομάκρυνση μεταβλητών είναι τα: Wilks' lambda, unexplained variance, Mahalanobis D^2 , μικρότερος F λόγος και Rao's V. Σύμφωνα με τη μέθοδο Wilks' lambda, οι μεταβλητές επιλέγονται προς εισαγωγή στη διαχωριστική συνάρτηση βάσει του πόσο πολύ μειώνουν το Wilks' lambda (Petalas και Anagnostopoulos, 2006; Αποστολάκης και λοιποί, 2003; Σιάρδος, 2000). Σε κάθε βήμα εισάγεται η μεταβλητή που ελαχιστοποιεί το συνολικό Wilks' lambda. Με τη μέθοδο της ανεξήγητης διακύμανσης, σε κάθε βήμα εισέρχεται η μεταβλητή που ελαχιστοποιεί το άθροισμα της ανεξήγητης διακύμανσης ανάμεσα στις ομάδες. Το κριτήριο επιλογής μεταβλητών με τη μέθοδο Mahalanobis D^2 είναι η μεγιστοποίηση του Mahalanobis D^2 ανάμεσα στις ομάδες. Με τη μέθοδο του μικρότερου F λόγου επιλέγονται οι μεταβλητές που μεγιστοποιούν τον F λόγο που υπολογίζεται από την απόσταση Mahalanobis ανάμεσα στις ομάδες. Το Rao's V είναι ένα μέτρο των διαφορών μεταξύ των ομάδων. Συχνά αποκαλείται και Lawley-Hotelling ίχνος. Σύμφωνα με τη μέθοδο αυτή, σε κάθε βήμα, εισέρχεται η μεταβλητή που μεγιστοποιεί την αύξηση του Rao's V. Η περισσότερο συχνά χρησιμοποιούμενη μέθοδος είναι η Wilks' lambda.

Η στατιστική σημαντικότητα της διαχωριστικής ισχύος κάθε διαχωριστικής συνάρτησης αποτιμάται μέσω των τιμών κατάλληλων κριτηρίων, όπως τα Wilks' lambda, Hottelling's trace και Pillai's, με περισσότερο συχνά χρησιμοποιούμενο το Wilks' lambda το οποίο ισούται με το ποσοστό της συνολικής διακύμανσης στις τυποποιημένες τιμές της διαχωριστικής συνάρτησης που δεν εξηγείται από τις διαφορές ανάμεσα στις ομάδες. Μικρότερες τιμές του Wilks' lambda υποδεικνύουν μεγαλύτερη διαχωριστική ικανότητα της συνάρτησης. Ο έλεγχος X^2 που συνοδεύει το Wilks' lambda ελέγχει τη μηδενική υπόθεση ότι οι μέσοι της διαχωριστικής συνάρτησης είναι ίσοι ανάμεσα στις ομάδες. Η απόρριψη της μηδενικής υπόθεσης υποδεικνύει ότι η διαχωριστική συνάρτηση συμβάλλει στατιστικά σημαντικά στο διαχωρισμό των ομάδων.

5.4.5. Ερμηνεία των αποτελεσμάτων της διαχωριστικής ανάλυσης

Η ερμηνεία των αποτελεσμάτων της διαχωριστικής ανάλυσης μπορεί να βασιστεί:

α) Στο απόλυτο μέγεθος των τυποποιημένων διαχωριστικών συντελεστών. Οι ανεξάρτητες μεταβλητές με σχετικά μεγαλύτερους συντελεστές κατά απόλυτη τιμή συνεισφέρουν περισσότερο στη διαχωριστική ισχύ της συνάρτησης απ' ό,τι οι μεταβλητές με μικρότερους συντελεστές. Η ερμηνεία των διαχωριστικών συντελεστών είναι παρόμοια με την ερμηνεία των συντελεστών των ανεξάρτητων μεταβλητών στη γραμμική παλινδρόμηση και συνεπώς υποβάλλεται στις ίδιες κριτικές. Ένας μικρός συντελεστής μπορεί να υποδηλώνει ή ότι η σχετική ανεξάρτητη μεταβλητή είναι άσχετη με τον προσδιορισμό μιας σχέσης ή ότι έχει μερικώς εξαιρεθεί από την σχέση λόγω πολυσυγγραμμικότητας.

β) Στις φορτίσεις διάκρισης ή διαχωρισμού (discriminant loadings) οι οποίες ονομάζονται επίσης και συντελεστές δομής (structure coefficients). Οι φορτίσεις είναι μέτρο της γραμμικής συσχέτισης μεταξύ κάθε ανεξάρτητης μεταβλητής και της διαχωριστικής συνάρτησης και μπορούν να ερμηνευθούν όπως οι παραγοντικές φορτίσεις στην αποτίμηση της σχετικής συνεισφοράς κάθε ανεξάρτητης μεταβλητής στη διαχωριστική συνάρτηση.

Παρόλο που τόσο οι φορτίσεις όσο και οι διαχωριστικοί συντελεστές μπορεί να μην είναι σταθεροί, οι φορτίσεις θεωρούνται περισσότερο έγκυρες για την ερμηνεία της διαχωριστικής ισχύος των ανεξάρτητων μεταβλητών λόγω της φύσης τους να συνδέονται με τη συσχέτιση.

5.4.6. Επικύρωση των αποτελεσμάτων της διαχωριστικής ανάλυσης

Η αποτίμηση της ακρίβειας εκτίμησης της ορθής ταξινόμησης, μέσω της διαχωριστικής ανάλυσης, γίνεται βάσει του πίνακα ταξινόμησης. Ένα υψηλό ποσοστό ορθής ταξινόμησης υποδεικνύει καλή ακρίβεια του μοντέλου. Παρόλα αυτά, η χρησιμοποίηση των ίδιων δεδομένων τόσο στην κατασκευή του μοντέλου όσο και στον προσδιορισμό των αποτελεσμάτων της ταξινόμησης έχει ως αποτέλεσμα τις επονομαζόμενες «τιμές σφάλματος επαναληπτικής αντικατάστασης» (reconstitution error rates). Αυτές τείνουν να υποεκτιμούν τις πραγματικές τάξεις σφάλματος του μοντέλου. Περισσότερο ακριβείς εκτιμήσεις μπορούν να ληφθούν με την εκτίμηση της ταξινόμησης δεδομένων τα οποία δεν είχαν χρησιμοποιηθεί στην κατασκευή του μοντέλου. Η μέθοδος αυτή απαιτεί μεγάλο μέγεθος δείγματος και χρησιμοποιεί ένα

μέρος των δεδομένων για την κατασκευή του μοντέλου (σετ εκμάθησης) ενώ ένα άλλο μέρος (σετ επικύρωσης) για την αποτίμηση της ικανότητας διαχωρισμού του μοντέλου. Μία άλλη μέθοδος επικύρωσης των αποτελεσμάτων είναι αυτή της διασταυρούμενης επικύρωσης (cross-validation) (Crask και λοιποί, 1977; Dillon και λοιποί, 1984). Σύμφωνα με τη μέθοδο αυτή, το μοντέλο κατασκευάζεται παρακρατώντας ένα από τα δεδομένα και στη συνέχεια εκτιμάται η ομάδα στην οποία ανήκει αυτό. Η ίδια διαδικασία επαναλαμβάνεται για καθένα από τα δεδομένα.

5.5. ΔΕΝΤΡΑ ΤΑΞΙΝΟΜΗΣΗΣ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

5.5.1. Εισαγωγή

Τα δέντρα ταξινόμησης και παλινδρόμησης (Classification and regression trees - CART) χρησιμοποιούνται ως εναλλακτική μη-παραμετρική μέθοδος προς τη διαχωριστική ανάλυση. Ο αλγόριθμος των δέντρων ταξινόμησης και παλινδρόμησης χρησιμοποιείται για τον προσδιορισμό ενός σετ από «εάν – τότε» (*if-then*) λογικών συνθηκών (διαχωριστικό σημείο - split) οι οποίες επιτρέπουν την ακριβή πρόβλεψη ή ταξινόμηση των περιπτώσεων (Breiman και λοιποί, 1984). Σκοπός των δέντρων ταξινόμησης και παλινδρόμησης είναι η πρόβλεψη ή η επεξήγηση της απόκρισης μιας κατηγορικής μεταβλητής (δέντρα ταξινόμησης) ή μιας συνεχούς μεταβλητής (δέντρα παλινδρόμησης) από ένα σετ επεξηγηματικών μεταβλητών με τη χρησιμοποίηση δυαδικών περιοδικά επαναλαμβανόμενων διαχωριστικών κανόνων, οι οποίοι βασίζονται σε όρια – κατώφλια (thresholds) στις επεξηγηματικές μεταβλητές (Breiman και λοιποί, 1984; Steinberg και Cola, 1997).

Η μέθοδος διάκρισης που χρησιμοποιείται από τα δέντρα ταξινόμησης μπορεί να θεωρηθεί ότι προσεγγίζει καλύτερα την ανθρώπινη λογική καθώς χρησιμοποιείται ξεχωριστά η κάθε μεταβλητή κάθε φορά.

Υπάρχουν δύο υποκείμενες βασικές ιδέες για την μέθοδο CART των δέντρων ταξινόμησης. Η πρώτη αφορά στον επαναλαμβανόμενο διαχωρισμό του διαστήματος των ανεξάρτητων μεταβλητών. Η δεύτερη αφορά στην περικοπή του δέντρου με τη χρησιμοποίηση των δεδομένων επικύρωσης ή ελέγχου.

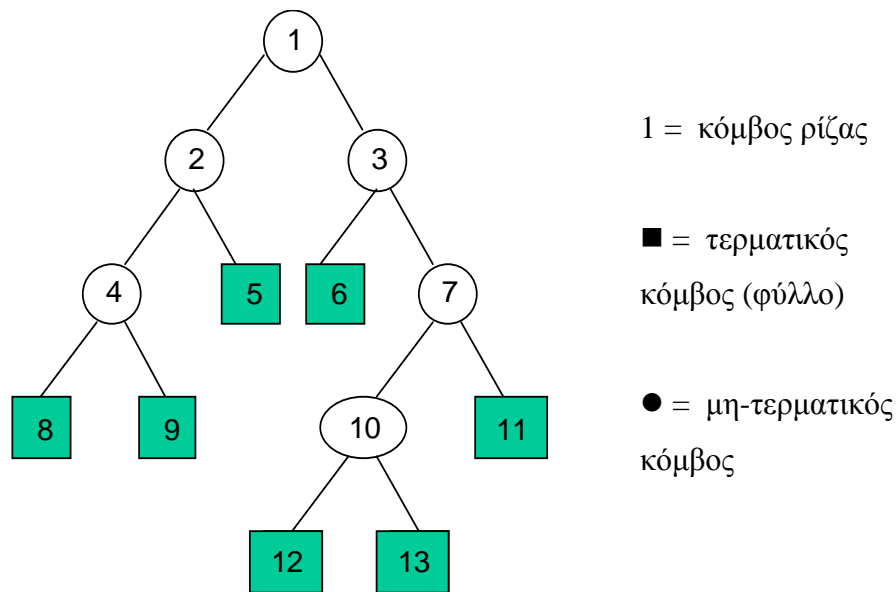
5.5.2. Κανόνες διαχωρισμού (Splitting rules) για τα δέντρα ταξινόμησης

Η πρώτη βασική ιδέα της μεθόδου CART για την ανάπτυξη ενός δέντρου ταξινόμησης πραγματοποιείται μέσω της παρακάτω διαδικασίας. Καθεμία μεταβλητή και όλες οι πιθανές διαχωριστικές τιμές για κάθε μεταβλητή εξετάζονται ώστε να βρεθεί ο καλύτερος διαχωρισμός. Οι πιθανές διαχωριστικές τιμές για μία μεταβλητή είναι απλώς τα ενδιάμεσα σημεία μεταξύ ζευγών διαδοχικών τιμών της μεταβλητής. Αυτά τα διαχωριστικά σημεία ταξινομούνται σύμφωνα με το πόσο πολύ μειώνουν τη μη-αγνότητα (impurity) δηλαδή την ετερογένεια της σύστασης του κόμβου. Η μείωση της μη-αγνότητας ορίζεται ως η μη - αγνότητα του κόμβου πριν τη διάσπαση μείον το ποσόν των μη -αγνοτήτων των δύο κόμβων που προκύπτουν από μια διάσπαση. Το κριτήριο διαχωρισμού (διαχωριστικός κανόνας) θα είναι αυτό που επιφέρει τη μεγαλύτερη μείωση στην μη - αγνότητα ενός κόμβου. Υπάρχουν διάφοροι τρόποι με τους οποίους μπορεί να μετρηθεί η μη-αγνότητα ενός κόμβου. Το δημοφιλέστερο μέτρο μη-αγνότητας είναι ο δείκτης μη - αγνότητας Gini (Gini impurity index) (Moisen, 2008). Εάν οι κλάσεις δηλωθούν ως k με $k=1, 2, \dots, J$, όπου J είναι ο συνολικός αριθμός των κλάσεων για την εξαρτημένη μεταβλητή, ο δείκτης μη-αγνότητας Gini για έναν κόμβο προσδιορίζεται από τον τύπο:

$$I(A) = 1 - \sum_{k=1}^J p_k^2$$

Όπου, p_k είναι το κλάσμα των παρατηρήσεων του κόμβου που ανήκουν στην κλάση k . Στην περίπτωση κατά την οποία όλες οι παρατηρήσεις ανήκουν σε μία μοναδική κλάση θα είναι $I(A) = 0$. Η τιμή του $I(A)$ μεγιστοποιείται όταν όλες οι κλάσεις εμφανίζονται στον κόμβο με ίσες αναλογίες. Η μέγιστη τιμή του δείκτη είναι $(J-1)/J$. Ο δείκτης Gini είναι ένα μέτρο του συνολικού σφάλματος σε κάθε κόμβο (Loh και Shih, 1997). Η μεταβλητή που επιλέγεται θα είναι αυτή που μειώνει περισσότερο το σφάλμα μεταξύ του γονικού κόμβου και του ποσού των σφαλμάτων των δύο νέων κόμβων τέκνων.

Ο λόγος για τον οποίο η μέθοδος καλείται αλγόριθμος δέντρων ταξινόμησης είναι ότι κάθε διάσπαση μπορεί να απεικονιστεί ως διάσπαση ενός κόμβου σε δύο κόμβους διαδόχων (κόμβοι-τέκνα). Η πρώτη διάσπαση παρουσιάζεται ως διακλάδωση του κόμβου - ρίζας ενός δέντρου. Στο Σχήμα 5.1 παρουσιάζεται ένα παράδειγμα δέντρου ταξινόμησης.



Σχήμα 5.1. Παράδειγμα δομής δέντρου ταξινόμησης.

5.5.3 Περικοπή του δέντρου (Pruning)

Η ακρίβεια είναι το σπουδαιότερο χαρακτηριστικό ενός δέντρου ταξινόμησης. Ο καλύτερος τρόπος για να ελεγχθεί η προβλεπτική ακρίβεια ενός δέντρου ταξινόμησης βασίζεται στον προσδιορισμό του ποσοστού λαθεμένης ταξινόμησης των περιπτώσεων ενός ανεξάρτητου σετ δεδομένων ελέγχου, με γνωστές κατανομές κλάσεων, το οποίο διατρέχει το δέντρο ταξινόμησης (Yohannes και Webb, 1999). Τα μεγάλα δέντρα ταξινόμησης παρουσιάζουν δύο προβλήματα. Πρώτον, παρόλο που είναι υψηλής ακρίβειας, με χαμηλά ή μηδενικά ποσοστά λαθεμένης ταξινόμησης, τα αποτελέσματα που παρουσιάζουν όταν εφαρμόζονται σε νέα δεδομένα είναι ανεπαρκή (Steinberg και Colla, 1997). Δεύτερον, η κατανόηση και επεξήγηση των δέντρων με μεγάλο αριθμό τερματικών κόμβων αποτελεί μια πολύπλοκη διαδικασία. Επομένως, τα μεγάλα δέντρα αναφέρονται ως πολύπλοκα δέντρα. Η πολυπλοκότητα ενός δέντρου μετρείται με τον αριθμό των τερματικών κόμβων.

Η δεύτερη βασική ιδέα της μεθόδου CART, ήταν η χρησιμοποίηση των δεδομένων επικύρωσης για την εκ των υστέρων «περικοπή» του δέντρου το οποίο αναπτύχθηκε από τα δεδομένα εκμάθησης. Η μέθοδος CART χρησιμοποιεί τα δεδομένα

επικύρωσης για την εκ των υστέρων «περικοπή» του δέντρου το οποίο έχει υπερ-αναπτυχθεί σκοπίμως. Η ιδέα πίσω από τη διαδικασία «περικοπής» είναι να αναγνωριστεί ότι ένα πολύ μεγάλο δέντρο είναι πιθανόν να κάνει υπερ-προσαρμογή των δεδομένων εκμάθησης. Για παράδειγμα, οι τελευταίες διασπάσεις μπορεί να οδηγούν σε ορθογώνια (κόμβους) με πολύ λίγα σημεία (ακόμη και μόνο ένα). Τότε, αυτοί οι τελευταίοι διαχωρισμοί είναι πιθανόν να συλλαμβάνουν απλώς τον «θόρυβο» στο σετ εκμάθησης παρά να απεικονίζουν τα πρότυπα που θα απεικονίζονταν σε μελλοντικά δεδομένα όπως είναι τα δεδομένα επικύρωσης.

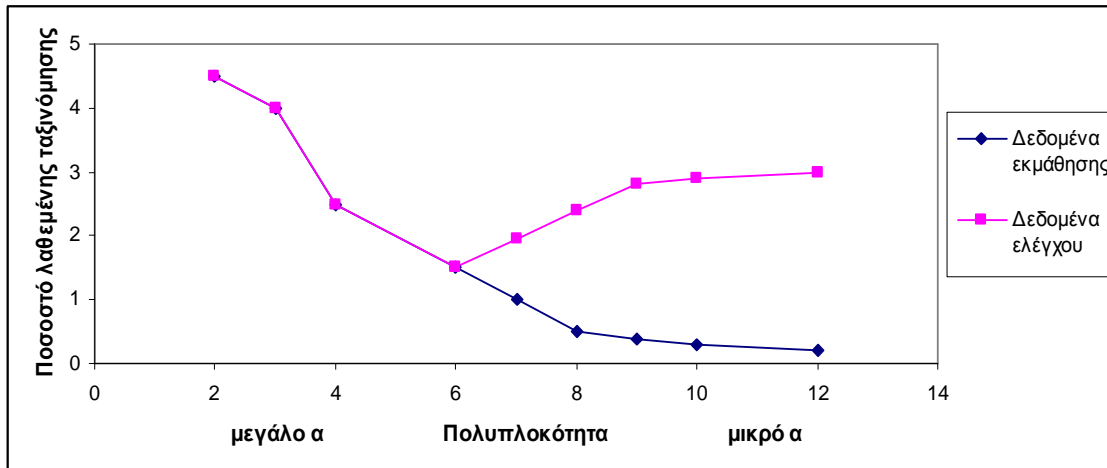
Η «περικοπή» αναφέρεται στη διαδοχική επιλογή ενός κόμβου απόφασης και τον επαναπροσδιορισμό του σαν έναν κόμβο φύλλου. Με αυτόν τον τρόπο «περικόπτονται» οι κλάδοι που εκτείνονται πέρα από αυτόν τον κόμβο απόφασης (δηλαδή το υπο-δέντρο του) με αποτέλεσμα τη μείωση του μεγέθους του δέντρου. Με τη διαδικασία της «περικοπής» γίνεται ουσιαστικά μια αντιστάθμιση μεταξύ του σφάλματος της λαθεμένης ταξινόμησης των δεδομένων του σετ επικύρωσης έναντι του αριθμού των κόμβων απόφασης στο «περικεκομμένο» δέντρο μέχρι να προσδιοριστεί ένα δέντρο που να συλλαμβάνει τα πρότυπα αλλά όχι τον «θόρυβο» στα δεδομένα εκμάθησης.

Η V-πλή διασταυρούμενη επικύρωση (V-fold cross validation) είναι μία εντατικά υπολογιστική μέθοδος για την επικύρωση μίας διαδικασίας για την κατασκευή μοντέλου, με την οποία αποφεύγεται η απαίτηση για ένα νέο ή ανεξάρτητο σετ δεδομένων επικύρωσης. Σύμφωνα με τη μέθοδο αυτή, το σετ δεδομένων εκμάθησης διαχωρίζεται τυχαία σε V τμήματα ώστε να διασφαλιστεί ότι η κατανομή των αποτελεσμάτων είναι παρόμοια σε καθένα από τα V υπο-τμήματα των δεδομένων. Ένα από αυτά τα υπο-τμήματα των δεδομένων παρακρατείται για να χρησιμοποιηθεί ως ανεξάρτητο σετ δεδομένων ελέγχου, ενώ τα υπόλοιπα V-1 υπο-τμήματα ενώνονται για να χρησιμοποιηθούν ως σετ δεδομένων εκμάθησης κατά τη διαδικασία κατασκευής του μοντέλου. Όλη η διαδικασία κατασκευής του μοντέλου επαναλαμβάνεται V φορές κρατώντας κάθε φορά ένα διαφορετικό υπο-τμήμα ως σετ δεδομένων ελέγχου. Με αυτόν τον τρόπο, παράγονται V διαφορετικά μοντέλα, καθένα από τα οποία μπορεί να ελεγχθεί έναντι ενός ανεξάρτητου υπο-τμήματος των δεδομένων. Η μέση επίδοση αυτών των V μοντέλων είναι μία εξαιρετική εκτίμηση της επίδοσης του αρχικού μοντέλου (που παράγεται με την χρησιμοποίηση

ολόκληρου του σετ των δεδομένων εκμάθησης) σε ένα μελλοντικό ανεξάρτητο σετ δεδομένων, ως μία συνάρτηση του αριθμού των τερματικών κόμβων ή της πολυπλοκότητας. Αυτό επιτρέπει μία εκτίμηση της πολυπλοκότητας του δέντρου, βασισμένη στα δεδομένα, η οποία καταλήγει στην βέλτιστη επίδοση όσον αφορά σε ένα ανεξάρτητο σετ δεδομένων. Με τη χρησιμοποίηση αυτής της μεθόδου, εμφανίζεται ένα ελάχιστο ποσοστό λαθεμένης ταξινόμησης όταν το δέντρο είναι αρκετά σύνθετο ώστε να προσαρμόσει την πληροφορία του σετ δεδομένων εκμάθησης, αλλά όχι τόσο σύνθετο ώστε να προσαρμόσει και το «θόρυβο» που υπάρχει στα δεδομένα.

Προκειμένου να δημιουργηθεί μια ακολουθία από ολοένα και απλούστερα δέντρα, καθένα από τα οποία είναι το καλύτερο μεταξύ όσων έχουν το ίδιο μέγεθος, χρησιμοποιείται το κριτήριο «κόστους πολυπλοκότητας» (cost complexity). Το κριτήριο «κόστους πολυπλοκότητας» εκφράζεται ως $Err(T) + \alpha |L(T)|$, όπου $Err(T)$ είναι το σφάλμα λαθεμένης ταξινόμησης ενός δέντρου T (βάσει των δεδομένων εκμάθησης), $|L(T)|$ είναι ο αριθμός φύλλων (τερματικών κόμβων) στο δέντρο T και α είναι το κόστος ποινής ανά κόμβο (ένας αριθμός που θα μεταβάλλεται προς τα πάνω από το μηδέν). Όταν $\alpha = 0$, δεν υπάρχει καμία ποινή για την ύπαρξη πάρα πολλών κόμβων σε ένα δέντρο και το καλύτερο δέντρο με τη χρησιμοποίηση του κριτηρίου «πολυπλοκότητας κόστους» είναι το πλήρως αναπτυγμένο δέντρο (χωρίς περικοπή). Όταν η τιμή του α αυξάνεται πάρα πολύ, η συνιστώσα κόστους ποινής αντισταθμίζει τη συνιστώσα σφάλματος λαθεμένης ταξινόμησης της συνάρτησης του κριτηρίου «πολυπλοκότητας κόστους» και το καλύτερο δέντρο είναι απλά το δέντρο με τα λιγότερα φύλλα, δηλαδή το δέντρο με απλά έναν κόμβο. Οι Breiman και λοιποί, (1984), έδειξαν ότι για κάθε $\alpha \geq 0$ υπάρχει ένα μοναδικό δέντρο ελάχιστου μεγέθους το οποίο ελαχιστοποιεί την ποσότητα $Err(T) + \alpha |L(T)|$. Καθώς η τιμή του α αυξάνεται από το 0 σε κάποια τιμή, αρχικά για κάποιο δέντρο T που διαμορφώνεται με την περικοπή του υποδέντρου σε έναν κόμβο απόφασης, εξισορροπείται ακριβώς το πρόσθετο κόστος του αυξανόμενου σφάλματος λαθεμένης ταξινόμησης (λόγω λιγότερων φύλλων) ενάντι του κόστους ποινής που κερδίζεται από την ύπαρξη λιγότερων φύλλων. Το πλήρες δέντρο «περικόπτεται» στον συγκεκριμένο κόμβο απόφασης με την περικοπή του υποδέντρου του και τον αναπροσδιορισμό αυτού του κόμβου απόφασης ως κόμβο φύλλου. Έστω T το νέο δέντρο που προκύπτει. Η

διαδικασία επαναλαμβάνεται στο νέο δέντρο T με την περαιτέρω αύξηση της τιμής του a . Συνεχίζοντας με τον ίδιο τρόπο, δημιουργείται μία ακολουθία δέντρων με διαρκώς μειωμένο αριθμό κόμβων μέχρις του δέντρου που αποτελείται από μόνο έναν κόμβο.



Σχήμα 5.2. Σχέση μεταξύ της πολυπλοκότητας του δέντρου και του κόστους απόφασης

Στο Σχήμα 5.2 παρουσιάζεται ένα παράδειγμα για τη σχέση μεταξύ της πολυπλοκότητας του δέντρου, εκφρασμένη με τον αριθμό των τερματικών κόμβων, και του ποσοστού λαθεμένης ταξινόμησης τόσο για το αρχικό σετ δεδομένων εκμάθησης όσο και για ένα ανεξάρτητο σετ δεδομένων ελέγχου ή επικύρωσης. Για το σετ των δεδομένων εκμάθησης, καθώς ο αριθμός των κόμβων αυξάνεται, το κόστος απόφασης μειώνεται μονότονα. Αυτό εναρμονίζεται με το γεγονός ότι το πλήρες δέντρο δίνει πάντοτε την καλύτερη προσαρμογή στο σετ των δεδομένων εκμάθησης. Αντίθετα, το αναμενόμενο ποσοστού λαθεμένης ταξινόμησης για ένα ανεξάρτητο σετ δεδομένων φτάνει σε μια ελάχιστη τιμή και στην συνέχεια αυξάνεται καθώς αυξάνεται η πολυπλοκότητα. Αυτό αντικατοπτρίζει το γεγονός ότι ένα υπερπροσαρμοσμένο και πολύ σύνθετο δέντρο δεν θα έχει καλή επίδοση σε ένα νέο σετ δεδομένων.

Το κριτήριο «κόστους πολυπλοκότητας» είναι ιδιαίτερα χρήσιμο στην περίπτωση κατά την οποία χρησιμοποιείται η διαδικασία της διασταυρούμενης επικύρωσης (Moisen, 2008; De'Ath and Fabricius, 2000). Με αυτόν τον τρόπο, η V-πλή διασταυρούμενη επικύρωση εφαρμόζεται σε ένα σχετικά μικρότερο σετ δέντρων. Από την ακολουθία των δέντρων φαίνεται φυσικό να επιλεγεί αυτό που έδωσε το

ελάχιστο σφάλμα λαθεμένης ταξινόμησης στο σύνολο των δεδομένων επικύρωσης. Το δέντρο αυτό καλείται δέντρο ελάχιστου σφάλματος. Εναλλακτικά, οι Breiman και λοιποί (1984) πρότειναν τον κανόνα του ενός τυπικού σφάλματος (1 – SE Rule) για την επιλογή του βέλτιστου μεγέθους του δέντρου. Σύμφωνα με τον κανόνα αυτόν επιλέγεται το μικρότερο από τα δέντρα των οποίων το ποσοστό λαθεμένης ταξινόμησης απέχει μέχρι ένα τυπικό σφάλμα από το δέντρο ελάχιστου σφάλματος. Με αυτόν τον τρόπο επιλέγονται ακόμα πιο φειδωλά δέντρα (parsimonious) με μικρή «θυσία» όσον αφορά στο σφάλμα ταξινόμησης.

ΚΕΦΑΛΑΙΟ 6

ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΔΙΕΡΓΑΣΙΩΝ (SPC)

6.1. ΕΙΣΑΓΩΓΗ

Ο Στατιστικός Έλεγχος Διεργασιών (SPC) χρησιμοποιεί στατιστικές μεθόδους για τη βελτίωση της ποιότητας μιας διεργασίας, η οποία επιτυγχάνεται με την συστηματική μείωση της μεταβλητότητας. Όλες οι διεργασίες, περιλαμβανομένων των χημικών αναλυτικών διεργασιών, υπόκεινται σε βέβαιη μεταβλητότητα. Υπάρχουν δύο αιτίες μεταβλητότητας: Οι κοινές αιτίες και οι ειδικές αιτίες. Οι πρώτες αφορούν στη φυσική μεταβλητότητα της διεργασίας και δεν μπορούν να αποφευχθούν. Οι ειδικές αιτίες, όμως, δεν αποτελούν ένα ενυπάρχον χαρακτηριστικό της διεργασίας και επομένως μπορούν να ταυτοποιηθούν. Η ύπαρξη μόνο κοινών αιτιών μεταβλητότητας, χαρακτηρίζει τις σταθερές διεργασίες οι οποίες θεωρείται ότι βρίσκονται υπό στατιστικό έλεγχο. Οι ειδικές αιτίες έχουν σαν αποτέλεσμα μη σταθερή διεργασία που είναι εκτός στατιστικού ελέγχου. Ο στατιστικός έλεγχος διεργασιών αποσκοπεί ακριβώς στο διαχωρισμό των κοινών (τυχαίων) αιτιών μεταβλητότητας από τις ειδικές αιτίες οι οποίες πρέπει να εξαλειφθούν μετά από τον εντοπισμό τους. Η στρατηγική για την εξάλειψη των ειδικών αιτιών μεταβλητότητας αφορά στη χρήση έγκαιρων προειδοποιήσεων για ειδικές μορφές μεταβλητότητας, την άμεση έρευνα για τις αιτίες του προβλήματος όταν υπάρχει ένδειξη ειδικής αιτίας (τι διαφορετικό υπήρχε σ' αυτή την περίπτωση;), και τέλος την εξάλειψη του ιδιαίτερου περιστασιακού ή τοπικού προβλήματος.

Κύρια εργαλεία του Στατιστικού Ελέγχου Διεργασιών είναι τα Διαγράμματα Ελέγχου. Η βασική ιδέα των διαγραμμάτων αυτών είναι να ελεγχθεί η υπόθεση ότι υπάρχουν μόνο κοινές αιτίες μεταβλητότητας έναντι της εναλλακτικής ότι υπάρχουν και ειδικές αιτίες. Δεδομένου ότι τα ποιοτικά χαρακτηριστικά του νερού αφορούν συνήθως σε μετρήσιμα μεγέθη, το ενδιαφέρον περιορίζεται σε διαγράμματα ελέγχου μεταβλητών. Όλα τα διαγράμματα ελέγχου έχουν κάποια κοινά χαρακτηριστικά:

Η κεντρική γραμμή του διαγράμματος αντιστοιχεί στη μέση επίδοση ενώ τα όρια ελέγχου αντιστοιχούν στο αναμενόμενο εύρος της μεταβλητότητας βάσει της διεργασίας. Εάν όλα τα σημεία του διαγράμματος είναι μεταξύ των ορίων ελέγχου και δεν παρουσιάζουν κάποια συστηματική μορφή, η διεργασία είναι εντός ελέγχου.

Τα διαγράμματα ελέγχου είναι κατάλληλα για εφαρμογή σε δεδομένα ανεξάρτητα που πανομοιότυπα ακολουθούν την κανονική κατανομή. Οι τρεις πιο σημαντικοί τύποι διαγραμμάτων για μεταβλητές, είναι τα: Διαγράμματα ελέγχου Shewhart, τα διαγράμματα εκθετικά σταθμισμένων κινητών μέσων (EWMA) και τα διαγράμματα CUSUM. Στη συνέχεια περιγράφονται οι τρεις τύποι διαγραμμάτων ελέγχου για μεταβλητές και συγκρίνονται μεταξύ τους.

6.2. ΔΙΑΓΡΑΜΜΑΤΑ SHEWHART

Ο πρώτος τύπος διαγραμμάτων ελέγχου αναπτύχθηκε το 1920 από τον W.A. Shewhart. Το πιο απλό διάγραμμα ελέγχου είναι η γραφική απεικόνιση των μετρήσεων πάνω στα δεδομένα, οι οποίες σημειώνονται στον κατακόρυφο άξονα, έναντι της σειράς (ή του χρόνου) κατά τον οποίο αποκτήθηκαν τα δεδομένα, που σημειώνεται στον οριζόντιο άξονα. Η κεντρική γραμμή και τα όρια για αυτά τα διαγράμματα διαφέρουν αν οι παρατηρήσεις αφορούν μεμονωμένες τιμές (διάγραμμα X των μεμονωμένων τιμών) ή υποομάδες (διάγραμμα \bar{X} των μέσων των υποομάδων).

Το πλέον συνηθισμένο διάγραμμα ελέγχου στην περίπτωση μεμονωμένων παρατηρήσεων, είναι το διάγραμμα X ή I chart, το οποίο είναι ένα διάγραμμα Shewhart των παρατηρήσεων. Επιπροσθέτως με τα μεμονωμένα σημεία, X_i , στο διάγραμμα X σχεδιάζεται μια κεντρική γραμμή στο μ και όρια ελέγχου στα $\pm 3\sigma$, όπου μ και σ είναι αντίστοιχα ο μέσος και η τυπική απόκλιση του πληθυσμού. Σημείο X_i εκτός των ορίων ελέγχου, είναι ένδειξη ειδικής αιτίας μεταβλητότητας και συνεπώς εκτός στατιστικού ελέγχου κατάσταση.

Συχνά, στην πράξη, τα μ και σ είναι άγνωστα και πρέπει να εκτιμηθούν. Συνήθως το μ εκτιμάται από τον δειγματικό μέσο $\hat{\mu} = \bar{X}$. Συχνά προτείνεται να

χρησιμοποιούνται τουλάχιστον 50 παρατηρήσεις (Ryan, 1989). Υπάρχουν πολλές διαδικασίες προσέγγισης για την εκτίμηση της σ . Η πιο συνηθισμένη προσέγγιση, είναι η «δημιουργία» ευρών λαμβάνοντας τις απόλυτες τιμές των διαφορών των διαδοχικών παρατηρήσεων (η δεύτερη μείον την πρώτη, η τρίτη μείον τη δεύτερη, κλπ). Ο μέσος αυτών των «κινητών ευρών» (moving ranges) διάστασης 2, χρησιμοποιείται για την εκτίμηση της σ . Συγκεκριμένα είναι:

$$\hat{\sigma} = \overline{MR} / d_2$$

όπου \overline{MR} είναι ο μέσος των κινητών ευρών. Το d_2 αναφέρεται σε ειδικούς πίνακες για $n=2$ και η ακριβής τιμή του είναι 1,128 (Ryan, 1989). Αυτή η προσέγγιση εκτίμησης της σ , είναι προβληματική όταν τα δεδομένα έχουν μια ενυπάρχουσα τάση. Τότε η εκτιμήτρια της σ , που βασίζεται στα κινητά εύρη, θα υποεκτιμά την σ (Ryan, 1989). Δεδομένου ότι ένα κινητό εύρος 2^{ης} τάξης εξομαλύνει οποιαδήποτε τάση που μπορεί να υπάρχει στα δεδομένα, τα όρια ελέγχου για το διάγραμμα X, που υπολογίζονται με βάση το κινητό εύρος θα είναι πολύ στενότερα. Επομένως μπορεί να λαμβάνονται λανθασμένες ενδείξεις ότι η διεργασία βρίσκεται εκτός ελέγχου.

Μία άλλη προσέγγιση για την εκτίμηση της σ , είναι μέσω της δειγματικής τυπικής απόκλισης s (Ryan, 1989), ως εξής:

$$\hat{\sigma} = s / c_4$$

όπου s είναι η τυπική απόκλιση του δείγματος τουλάχιστον 50 συνεχόμενων παρατηρήσεων και c_4 μια τιμή που μπορεί να υπολογιστεί από την επόμενη εξίσωση:

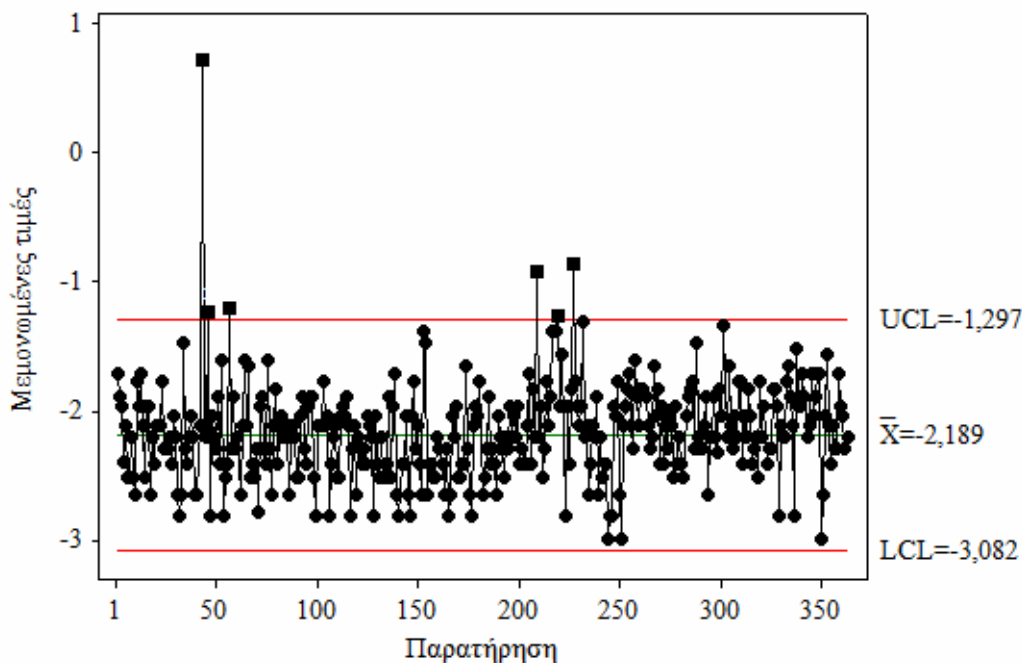
$$c_4 = \sqrt{\frac{2}{n-1} * \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}}$$

όπου n είναι ο αριθμός των παρατηρήσεων και Γ είναι η συνάρτηση γάμμα. Οι τιμές του c_4 που υπολογίζονται με αυτόν τον τρόπο προσεγγίζουν τη μονάδα καθώς αυξάνεται το πλήθος n των μεμονωμένων παρατηρήσεων. Για $n > 25$, $c_4 \cong 4(n-1)/(4n-3)$.

Αυτός ο τρόπος εκτίμησης της σ δίνει έναν αμερόληπτο εκτιμητή της όταν οι παρατηρήσεις είναι ανεξάρτητες. Όταν οι παρατηρήσεις δεν είναι ανεξάρτητες, μπορεί να εξασφαλίσει μια καλύτερη εκτίμηση της σ απ' ό,τι η \overline{MR} / d_2 , στην

περίπτωση που όλες οι παρατηρήσεις έχουν τα τυπικά χαρακτηριστικά των δεδομένων που παράγονται από τη διεργασία. Εάν όμως, η s έχει υπολογιστεί από δεδομένα που περιέχουν μια τάση λόγω μιας ειδικής αιτίας η οποία μπορεί (άμεσα) να απαλειφθεί, δεν συνιστάται η χρησιμοποίησή της για την εκτίμηση της σ , (Ryan, 1989).

Στο Σχήμα 6.1 παρουσιάζεται διάγραμμα ελέγχου Shewhart των μεμονωμένων ημερήσιων τιμών θολότητας (σε φυσική λογαριθμική κλίμακα) του επεξεργασμένου νερού της MEN Πολυδενδρίου στη διάρκεια του έτους 2004. Η εκτίμηση της σ έχει γίνει μέσω των κινητών ευρών και τα δεδομένα έχουν ελεγχθεί ώστε να επιβεβαιωθεί ότι τηρούνται οι προϋποθέσεις του διαγράμματος (κανονικότητα – ανεξαρτησία). Στο διάγραμμα αποτυπώνονται τα όρια ελέγχου τριών τυπικών αποκλίσεων και η κεντρική στον αριθμητικό μέσο των δεδομένων.



Σχήμα 6.1. Διάγραμμα ελέγχου Shewhart μεμονωμένων ημερήσιων τιμών θολότητας (σε φυσική λογαριθμική κλίμακα) επεξεργασμένου νερού της MEN Πολυδενδρίου.

Με το βασικό κριτήριο των διαγραμμάτων Shewhart, μία διεργασία θεωρείται ότι είναι εκτός ελέγχου όταν ένα ή περισσότερα σημεία είναι εκτός των ορίων των τριών τυπικών αποκλίσεων. Για την επαύξηση της αποτελεσματικότητας των διαγραμμάτων Shewhart, έχουν προταθεί κριτήρια τα οποία ανιχνεύουν μικρές τάσεις στα δεδομένα

προς εκτός ελέγχου κατάσταση. Τα κριτήρια αυτά, που συχνά αποκαλούνται ως κανόνες ροής για διαδοχική σειρά σημείων (run rules), βασίζονται σε διαδοχές παρατηρήσεων του ίδιου τύπου (όσον αφορά στη διεύθυνση των σημείων). Το βασικό κριτήριο προσδιορίζει ότι η διεργασία είναι εκτός ελέγχου όταν εντοπίζεται μία ακραία τιμή, με την έννοια ότι τότε αυτή θα προέρχεται από μια διαφορετική κατανομή πιθανότητας από την κατανομή της διεργασίας όταν είναι εντός ελέγχου. Οι κανόνες ροής κατά κάποιο τρόπο προβλέπουν μια τάση της διεργασίας προς κατάσταση εκτός ελέγχου. Για παράδειγμα, μία συνεχής σειρά μήκους 8 ή περισσότερων διαδοχικών σημείων που βρίσκονται προς την ίδια πλευρά της κεντρικής γραμμής, έχει μικρή πιθανότητα να εμφανισθεί σε ένα τυχαίο δείγμα σημείων. Η χρήση των κανόνων ροής, για την βελτίωση της ικανότητας των διαγραμμάτων ελέγχου να εντοπίζουν μικρότερες μετατοπίσεις, πρέπει να γίνεται με μεγάλη προσοχή γιατί μπορούν να οδηγήσουν σε ψευδείς ενδείξεις για εκτός ελέγχου κατάσταση (Champ και Woodall, 1987).

Για τον έλεγχο της μεταβλητότητας μπορεί να χρησιμοποιηθεί το διάγραμμα R (διάγραμμα κινητού εύρους). Στο διάγραμμα αυτό, η κεντρική του γραμμή σχεδιάζεται στο:

$$\overline{MR} = \frac{\sum_{i=2}^n MR_i}{n-1}$$

όπου $MR_i = |x_i - x_{i-1}|$, και x_i είναι η παρατήρηση τη χρονική στιγμή i . Τα όρια ελέγχου γι' αυτό το διάγραμμα R είναι:

$$UCL = D_4^* \overline{MR} \text{ (άνω όριο ελέγχου) και } LCL = D_3^* \overline{MR} \text{ (κάτω όριο ελέγχου)}$$

Οι τιμές για τα D_3 και D_4 υπάρχουν σε ειδικούς πίνακες, στους οποίους για την περίπτωση μεμονωμένων παρατηρήσεων, αναφέρονται σε δείγματα μεγέθους $n = 2$.

Το διάγραμμα R έχει αποδειχθεί ότι δεν μπορεί να συνεισφέρει πραγματικά στην αναγνώριση μιας μετακίνησης στη μεταβλητότητα της διεργασίας (Sullivan και Woodal, 1996a).

Η έλλειψη κανονικότητας των δεδομένων, αποτελεί σημαντικό πρόβλημα για τα διαγράμματα Shewhart μεμονωμένων παρατηρήσεων.

6.3. ΔΙΑΓΡΑΜΜΑΤΑ ΕΚΘΕΤΙΚΑ ΣΤΑΘΜΙΣΜΕΝΩΝ ΚΙΝΗΤΩΝ ΜΕΣΩΝ (EWMA)

Το διάγραμμα EWMA (Exponentially Weighted Moving Average) ορίστηκε από τον Roberts το 1960 ως ένα εναλλακτικό διάγραμμα των Shewhart διαγραμμάτων. Χρησιμοποιείται για να γίνονται αντιληπτές μικρές μεταβολές στην διαδικασία. Γενικά, χρησιμοποιείται για μεμονωμένες παρατηρήσεις γεγονός που το καθιστά ιδιαίτερα ενδιαφέρον για την αναλυτική χημεία.

Η στατιστική συνάρτηση του διαγράμματος EWMA ορίζεται σύμφωνα με την παρακάτω εξίσωση:

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1}$$

Όπου λ είναι σταθερά, $0 < \lambda \leq 1$ και η αρχική τιμή z_0 , που απαιτείται για την πρώτη παρατήρηση (τη χρονική στιγμή $i = 1$), είναι η τιμή στόχος της διαδικασίας, δηλαδή $z_0 = \mu_0$. Ως αρχική τιμή μπορεί να χρησιμοποιηθεί, αν είναι διαθέσιμος, ο μέσος \bar{x} προηγούμενων δειγμάτων, που είχαν ληφθεί όταν θεωρείτο ότι η διαδικασία ήταν υπό έλεγχο. Εάν δεν υπάρχουν προηγούμενα στοιχεία, ο μέσος εκτιμάται από τα δεδομένα.

Αποδεικνύεται ότι η στατιστική συνάρτηση του EWMA είναι ένας σταθμισμένος μέσος όλων των προηγούμενων παρατηρήσεων. Πράγματι, με διαδοχικές αντικαταστάσεις στην προηγούμενη σχέση, λαμβάνεται:

$$\begin{aligned} z_i &= \lambda x_i + (1 - \lambda) z_{i-1} = \lambda x_i + (1 - \lambda)[\lambda x_{i-1} + (1 - \lambda) z_{i-2}] = \\ &= \lambda x_i + \lambda(1 - \lambda) x_{i-1} + (1 - \lambda)^2 z_{i-2} = \dots = \\ &= \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0 \end{aligned}$$

Προκύπτει λοιπόν ένας αναδρομικός τύπος που δίνει βάρος λ στην τελευταία παρατήρηση, $\lambda(1 - \lambda)$ στην προτελευταία, $\lambda(1 - \lambda)^2$ στην αμέσως προηγούμενη κ.ο.κ. Επειδή τα βάρη $\lambda(1 - \lambda)^j$, φθίνουν γεωμετρικά (κάθε παρατήρηση έχει μεγαλύτερο βάρος στον υπολογισμό απ' ότι η προηγούμενή της), το διάγραμμα EWMA, αποκαλείται συχνά ως διάγραμμα γεωμετρικού κινητού μέσου. Επιπροσθέτως, τα βάρη αθροίζονται στην μονάδα. Πράγματι,

$$\lambda \sum_{j=0}^{i-1} (1 - \lambda)^j + (1 - \lambda)^i = \lambda \left[\frac{1 - (1 - \lambda)^i}{1 - (1 - \lambda)} \right] + (1 - \lambda)^i = 1 - (1 - \lambda)^i + (1 - \lambda)^i = 1$$

Θεωρώντας ότι οι παρατηρήσεις x_i είναι ανεξάρτητες τυχαίες μεταβλητές με διακύμανση σ^2 , τότε η διακύμανση της z_i είναι:

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]$$

Στην περίπτωση που δεν είναι γνωστή η διακύμανση σ^2 του πληθυσμού, χρησιμοποιείται η εκτίμησή της από προηγούμενα δεδομένα και ελλείψει αυτών, εκτιμάται από τα τρέχοντα δεδομένα, είτε μέσω του μέσου κινητού εύρους είτε μέσω της δειγματικής διακύμανσης s (όπως στα διαγράμματα Shewhart).

Το διάγραμμα EWMA λοιπόν, κατασκευάζεται σχεδιάζοντας τις τιμές z_i έναντι της χρονικής σειράς i με την οποία προέκυψαν οι παρατηρήσεις. Η κεντρική γραμμή και τα όρια ελέγχου του διαγράμματος EWMA, δίνονται από τις παρακάτω σχέσεις:

$$UCL = \mu_0 + L\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]}$$

$$CL = \mu_0$$

$$LCL = \mu_0 - L\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]}$$

Όπου L είναι το πλάτος των ορίων ελέγχου – η τιμή που θα πολλαπλασιαστεί με την τυπική απόκλιση στα όρια (για $L = 3$, προκύπτουν όρια τριών τυπικών αποκλίσεων).

Από τις εξισώσεις των ορίων παρατηρείται ότι αυτά είναι μεταβλητά. Όσο όμως αυξάνει το i , η παράσταση $[1 - (1 - \lambda)^{2i}]$ τείνει στη μονάδα. Συνεπώς, μετά από αρκετές χρονικές περιόδους, τα όρια του διαγράμματος EWMA, θα δίνονται προσεγγιστικά από τους παρακάτω τύπους:

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$$

$$LCL = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$$

Στην περίπτωση που το EWMA δεν χρησιμοποιείται για μεμονωμένες παρατηρήσεις αλλά για υποομάδες ($n > 1$), αρκεί στους τύπους που προηγήθηκαν να αντικατασταθεί το x_i με \bar{x}_i (τον μέσο κάθε υποομάδας) και η σ με σ/\sqrt{n} .

Οι τιμές των παραμέτρων λ και L ενός διαγράμματος EWMA, προσδιορίζονται ακολουθώντας την γενική ιδέα κάθε διαγράμματος ελέγχου, σύμφωνα με την οποία πρέπει συγχρόνως:

- Να αποκαλύπτει μια μεταβολή στη διαδικασία όσον το δυνατόν γρηγορότερα.
- Η πιθανότητα λανθασμένων ενδείξεων, ότι η διαδικασία είναι εκτός ελέγχου όταν δεν ισχύει κάτι τέτοιο στην πραγματικότητα, να είναι μικρή.

Επειδή το EWMA είναι ένας σταθμισμένος μέσος όλων των προηγούμενων παρατηρήσεων και της τρέχουσας παρατήρησης, δεν είναι ευαίσθητο στην υπόθεση της κανονικότητας, η οποία όπως έχει ήδη αναφερθεί, είναι πολύ σημαντική στην περίπτωση των μεμονωμένων παρατηρήσεων.

Μια ευρύτερη θεώρηση του διαγράμματος EWMA είναι η ικανότητά του να προβλέπει το επίπεδο της διεργασίας. Το διάγραμμα EWMA, δίνει μια πρόβλεψη για το που θα βρίσκεται ο μέσος της διεργασίας την επόμενη χρονική περίοδο. Δηλαδή η στατιστική συνάρτηση z_i , είναι μια πρόβλεψη της τιμής του μέσου της διεργασίας μ τη χρονική στιγμή $i + 1$. Έτσι, το EWMA μπορεί να χρησιμοποιηθεί σαν μια βάση για έναν αλγόριθμο δυναμικού ελέγχου διεργασιών.

6.4. ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΣΣΩΡΕΥΜΕΝΩΝ ΑΘΡΟΙΣΜΑΤΩΝ (CUSUM)

Το διάγραμμα συσσωρευμένων αθροισμάτων (CUSUM - Cumulative Sum Chart), πρωτοσχεδιάστηκε από τον Page το 1950 με σκοπό να βελτιώσει την αποτελεσματικότητα των διαγραμμάτων ελέγχου σε μικρές μετακινήσεις. Τα συσσωρευμένα αθροίσματα υπολογίζονται ως εξής:

$$C_i = \sum_{j=1}^i (X_j - \bar{X})$$

Όπου, X_i είναι η δειγματική μέτρηση τη χρονική στιγμή i και \bar{X} είναι η δειγματική εκτίμηση του μέσου του πληθυσμού. Στο διάγραμμα CUSUM, τα C_i σημειώνονται στον άξονα των τεταγμένων, ενώ στον άξονα των τεταγμένων σημειώνεται ο αριθμός του δείγματος.

Παρακάτω εξετάζεται το τυποποιημένο CUSUM στο οποίο χρησιμοποιούνται οι τυποποιημένες τιμές της μεταβλητής X_i . Οι z - τιμές ενός διαγράμματος CUSUM, υπολογίζονται ως:

$$Z_i = \frac{X_i - \bar{X}}{\hat{\sigma}}$$

όπου, X_i είναι η δειγματική μέτρηση τη χρονική στιγμή i και \bar{X} είναι η δειγματική εκτίμηση του μέσου του πληθυσμού και η τυπική απόκλιση σ των μεμονωμένων παρατηρήσεων εκτιμάται είτε μέσω της \overline{MR}/d_2 είτε μέσω της s / c_4 . Στην περίπτωση που υπάρχει κάποια τάση στα δεδομένα, το διάγραμμα CUSUM χρησιμοποιείται μόνο για τον εντοπισμό της και μάλιστα εκτιμώντας την σ μέσω της s / c_4 .

Για κάθε παρατήρηση, υπολογίζονται δύο συσσωρευτικά αθροίσματα, από τα οποία το πρώτο εντοπίζει θετικές μετακινήσεις του μέσου και το δεύτερο αρνητικές:

$$S_{Hi} = \max[0, (z_i - k) + S_{Hi-1}]$$

και

$$S_{Li} = \max[0, (-z_i - k) + S_{Li-1}]$$

Όπου η τιμή της k συνήθως επιλέγεται να είναι το μισό της αλλαγής του μέσου – εκφρασμένη σε μονάδες z (τυποποιημένες) που είναι επιθυμητό να προσδιοριστεί, δηλαδή $k = \delta/2$ όπου δ η τυποποιημένη αλλαγή στον μέσο που επιθυμείται να προσδιοριστεί. Τα S_{Hi} και S_{Li} ονομάζονται ανώτερο και κατώτερο CUSUMs αντίστοιχα και συσσωρεύουν τις αποκλίσεις από τον μέσο \bar{X} , που είναι μεγαλύτερες από k . Και οι δύο ποσότητες τίθενται ως μηδέν όταν γίνονται αρνητικές. Η συνήθης επιλογή για την k είναι 0.5 και είναι κατάλληλη για τον εντοπισμό μιας 1σ αλλαγής στον μέσο της διεργασίας. Τα παραπάνω αθροίσματα, αρχίζουν με $S_{Ho} = 0$ και $S_{Lo} = 0$. Η τυπική οριακή τιμή h (όρια ελέγχου) που όταν την υπερβεί κάποιο από τα δύο αθροίσματα υπάρχει ένδειξη ότι η διεργασία είναι εκτός ελέγχου, είναι είτε 4 είτε 5 με πιο συνηθισμένη στην πράξη την τιμή $h=5$. Εάν είτε το S_{Hi} είτε το S_{Li} υπερβαίνει το διάστημα απόφασης h , τότε η διεργασία θεωρείται εκτός ελέγχου. Στην περίπτωση που δεν χρησιμοποιούνται οι τυποποιημένες τιμές της X_i αλλά οι ίδιες οι τιμές της, τότε τα CUSUMs ορίζονται ως εξής:

$$S_{Hi} = \max[0, X_i - (\bar{X} + K) + S_{Hi-1}]$$

και

$$S_{Li} = \max[0, (\bar{X} - K) - X_i + S_{Li-1}]$$

Όπου $K=k\sigma$ και το διάστημα απόφασης είναι $H=h\sigma$.

Ένας γραφικός τρόπος για να αποφασιστεί αν η διαδικασία είναι εκτός ελέγχου, είναι η χρησιμοποίηση μιας V μάσκας πάνω σε ένα διάγραμμα CUSUM. Η V μάσκα έχει αρκετά μειονεκτήματα και δεν συνιστάται η χρήση της (Montgomery, 2001).

Στην περίπτωση που το διάγραμμα Cusum χρησιμοποιείται για τους μέσους υποομάδων ($n>1$) θα πρέπει στους τύπους να αντικατασταθεί το X_i με \bar{X}_i (τον μέσο της υποομάδας) και η σ με σ/\sqrt{n} .

Θα πρέπει να σημειωθεί ότι τα διαγράμματα CUSUM επηρεάζονται, επίσης, από την έλλειψη κανονικότητας των δεδομένων και την συσχέτισή τους, μιας και τα δύο συσσωρευτικά αθροίσματα στηρίζονται στην υπόθεση ότι οι z – τιμές που υπολογίζονται, έχουν κατανομή $N(0,1)$ (Ryan, 1989).

6.5. ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΩΝ ΔΙΑΓΡΑΜΜΑΤΩΝ ΕΛΕΓΧΟΥ

Το μήκος ροής (run length) είναι μία πολύ σημαντική ιδιότητα κάθε διαγράμματος του SPC. Το μήκος ροής είναι ο αριθμός των παρατηρήσεων που απαιτούνται ώστε να ληφθεί μία παρατήρηση εκτός ορίων ελέγχου για μία δεδομένη μετακίνηση στον μέσο της διεργασίας (Wadsworth, Stephens, και Godfrey 1986). Σύμφωνα με τους Ewan και Kemp (1960), όταν η ποιότητα παραμένει σταθερή, ο αναμενόμενος αριθμός δειγμάτων που έχουν ληφθεί από τη δειγματοληψία μέχρι να εμφανιστεί εκτός ελέγχου κατάσταση, καθορίζεται μέσω του ARL. Εννοείται ότι οι ιδανικές τιμές για το ARL θα πρέπει είναι μεγάλες όταν δεν έχει υπάρξει κάποια ειδική αιτία μεταβλητότητας ενώ σε αντίθετη περίπτωση είναι επιθυμητό να είναι μικρές. Επιζητείται δηλαδή, να υπάρχει μικρή πιθανότητα ενός ψευδούς σήματος εκτός ελέγχου ενώ παράλληλα η αλλαγή στη διεργασία να γίνεται αντιληπτή όσον το δυνατόν ταχύτερα.

Όταν οι παρατηρήσεις είναι ανεξάρτητες και κατανέμονται κανονικά, σε ένα διάγραμμα Shewhart των τριών τυπικών αποκλίσεων, η πιθανότητα ενός σημείου να πέφτει εκτός των ορίων ελέγχου, όταν η διεργασία είναι εντός ελέγχου, είναι $(1 - 0,99865)^2 = 0,0027$, μιας και $\Phi(3) = P(Z \leq 3) = 0,99865$ (οπότε η πιθανότητα να βρίσκεται ένα σημείο στη δεξιά ουρά της κανονικής κατανομής, είναι $1 - 0,99865 = 0,00135$).

Για κάθε Shewhart διάγραμμα ελέγχου το ARL ορίζεται ως το αντίστροφο της πιθανότητας ένα σημείο να βρίσκεται εκτός ορίων ελέγχου, δηλαδή $ARL = 1/p$, όπου p η παράμετρος της γεωμετρικής κατανομής που ακολουθεί το μήκος ροής. Στην περίπτωση που ενδιαφέρει ο αναμενόμενος αριθμός δειγμάτων που πρέπει να ληφθούν για να υπάρξει λανθασμένο σήμα ότι η διεργασία βρίσκεται εκτός ελέγχου ενώ στην πραγματικότητα είναι υπό έλεγχο, το ARL ισούται με $ARL_0 = 1/\alpha$, όπου α είναι η πιθανότητα σφάλματος τύπου I. Στην περίπτωση που αναφέρεται στον αναμενόμενο αριθμό δειγμάτων που πρέπει να ληφθούν για να εντοπιστεί πραγματική αλλαγή στη διεργασία, ισούται με $ARL_1 = 1/(1 - \beta)$, όπου β είναι η πιθανότητα σφάλματος τύπου II. Έτσι λοιπόν τα ARL_0 και ARL_1 αποτελούν τυπικό μέτρο σύγκρισης της αποτελεσματικότητας των διαγραμμάτων ελέγχου. Η SRL (τυπική

απόκλιση του μήκους ροής) δίνει μια ένδειξη της μεταβλητότητας της κατανομής του μήκους ροής. Για παράδειγμα, στο τυπικό διάγραμμα Shewhart, των τριών τυπικών αποκλίσεων, εν απουσία μιας μετακίνησης του μέσου το ARL είναι $1/p = 1/0,0027 \cong 370$ ($ARL_0 \cong 370$). Όμως, η SRL είναι επίσης περίπου 370, όπως υπολογίζεται από τον τύπο $\sqrt{(1-p)}/p$ της τυπικής απόκλισης της γεωμετρικής κατανομής του μήκους ροής. Έτσι είναι πιθανόν να υπάρχουν παρατηρήσεις εκτός ελέγχου πολύ συχνότερα ή πολύ αργότερα απ' ότι αναμενόταν, ακόμα κι αν δεν υπάρχει μετακίνηση στον μέσο (Wardell, Moskowitz, και Plante 1994). Το διάγραμμα Shewhart μεμονωμένων παρατηρήσεων X υστερεί σε σχέση με το διάγραμμα \bar{X} υποομάδων ίσου μεγέθους ως προς την ικανότητα γρήγορου εντοπισμού αλλαγών στην διαδικασία, και συνιστάται να χρησιμοποιείται μόνο στην περίπτωση που δεν είναι δυνατόν να χρησιμοποιηθούν δείγματα (υποομάδες) μεγέθους $n>1$ για σχεδιασμό διαγράμματος \bar{X} . Αξίζει να σημειωθεί σε ότι για ίδιου μεγέθους μετακίνηση του μέσου της διεργασίας, το ARL_1 μικραίνει όταν αυξάνεται το μέγεθος των δειγμάτων.

Όταν τα σημεία του διαγράμματος σχεδιάζονται ένα προς ένα με την σειρά που συλλέγονται από τα δεδομένα, τα διαγράμματα Shewhart χρησιμοποιούν για τη διεργασία μόνο την πληροφορία που προέρχεται από το τελευταίο σημείο που σχεδιάζεται και αγνοούν την πληροφόρηση που μπορεί να δώσει ολόκληρη η ακολουθία των σημείων. Το αποτέλεσμα είναι, γενικά, τα διαγράμματα Shewhart να αντιλαμβάνονται γρήγορα τις μεγάλες αλλαγές στη διεργασία ενώ να μην είναι ευαίσθητα στις μικρές. Αντίθετα τα διαγράμματα CUSUM, με τη σχεδίαση των συσσωρευμένων αθροισμάτων των αποκλίσεων των δειγματικών τιμών από την τιμή στόχο, συγχωνεύουν όλη την πληροφορία της ακολουθίας των δειγματικών τιμών. Το γεγονός αυτό τα καθιστά πιο ευαίσθητα από τα διαγράμματα Shewhart στον εντοπισμό μικρών αλλαγών στη διεργασία. Οι τιμές ARL_0 είναι μεγαλύτερες για ένα διάγραμμα CUSUM με $k=0,5$ και $h=5$ και οι τιμές ARL_1 , για αλλαγές στη διαδικασία μεγέθους μικρότερου από δύο τυπικές αποκλίσεις, είναι μικρότερες απ' ότι αντίστοιχα σε ένα διάγραμμα Shewhart μεμονωμένων παρατηρήσεων. Η γενική άποψη (Champ και Woodall, 1987) είναι ότι το CUSUM εντοπίζει τις μικρές αποκλίσεις γρηγορότερα από το Shewhart, ακόμα κι αν χρησιμοποιηθούν στο τελευταίο συμπληρωματικοί κανόνες ροής (run rules). Φυσικά αυτό εξαρτάται από την επιλογή της παραμέτρου δ (τιμή της αλλαγής σε μονάδες τυπικής απόκλισης)

όταν υπολογίζονται τα S_H , S_L και τα όρια ελέγχου του Cusum. Για $k = 0,5$ η τιμή $h = 4,77$ εξασφαλίζει ένα CUSUM με $ARL_0 = 370$ δείγματα, που εναρμονίζεται με την τιμή ARL_0 για ένα Shewhart διάγραμμα ελέγχου με τα συνηθισμένα όρια των τριών τυπικών αποκλίσεων, όπως φαίνεται από τον παρακάτω πίνακα 6.1. (Hawkins, 1993α).

Πίνακας 6.1. Τιμές του k και οι αντίστοιχες τιμές του h που δίνουν $ARL_0 = 370$ για το αμφίπλευρο CUSUM

k	0,25	0,5	0,75	1,0	1,25	1,5
h	8,01	4,77	3,34	2,52	1,99	1,61

Προκειμένου να μειωθεί το ARL_1 χωρίς να μειωθεί σημαντικά το ARL_0 , οι Lucas και Croiser (1982) πρότειναν μία παραλλαγή του CUSUM που καλείται FIR CUSUM (CUSUM γρήγορης αρχικής απόκρισης).

Τα διαγράμματα EWMA, διαφέρουν από τους δύο άλλους τύπους διαγραμμάτων ελέγχου κυρίως ως προς τον τρόπο με τον οποίο αξιοποιούνται οι προηγούμενες πληροφορίες για τον υπολογισμό του στατιστικού. Το τελευταίο σημείο σε ένα διάγραμμα Shewhart, εξαρτάται μόνο από την τελευταία μέτρηση, ενώ σε ένα CUSUM εξαρτάται από όλες τις προηγούμενες μετρήσεις και σε ένα EWMA εξαρτάται από την τιμή της παραμέτρου λ . Όσο μεγαλύτερη είναι η τιμή του λ , τόσο μεγαλύτερη είναι η επίδραση των τελευταίων παρατηρήσεων. Καθώς το λ τείνει στο μηδέν, το διάγραμμα EWMA, παίρνει τη μορφή του CUSUM, ενώ καθώς το λ τείνει στην μονάδα, παίρνει τη μορφή του Shewhart. Υπάρχουν θεωρητικές έρευνες (Lucas και Saccucci, 1990) σχετικές με την ARL συμπεριφορά του διαγράμματος EWMA σε σχέση με τις τιμές των λ και L . Αποδεικνύεται λοιπόν, ότι ο βέλτιστος σχεδιασμός ενός τέτοιου διαγράμματος, επιτυγχάνεται αφού προηγουμένως έχουν καθοριστεί τα επιθυμητά εντός – ελέγχου και εκτός – ελέγχου ARL καθώς και το μέγεθος της αλλαγής του μέσου της διαδικασίας που επιθυμείται να εντοπιστεί και στη συνέχεια, από τα διάφορα ζεύγη λ και L που έχουν την ίδια εντός ελέγχου τιμή ARL, επιλεγεί ο συνδυασμός εκείνος που δίνει, για τη συγκεκριμένη αλλαγή στον μέσο, την μικρότερη τιμή ARL. Γενικά, στην πράξη, έχει αποδειχθεί ότι οι τιμές του λ που ανήκουν στο κλειστό διάστημα $[0,05, 0,25]$ δίνουν καλά αποτελέσματα.

Συνηθέστερες είναι οι επιλογές $\lambda = 0,05$, $\lambda = 0,10$ και $\lambda = 0,20$. Ένας εμπειρικός κανόνας για να γίνονται αντιληπτές οι μικρές αλλαγές, είναι να χρησιμοποιούνται μικρές τιμές του λ . Επίσης η τιμή του $L = 3$ (τα συνηθισμένα όρια των τριών τυπικών αποκλίσεων), δίνει ικανοποιητικά αποτελέσματα, ειδικά για μεγάλες τιμές του λ . Όταν το λ έχει τιμή μικρότερη ή ίση του $0,1$, είναι προτιμότερο να μειώνεται το πλάτος των ορίων, χρησιμοποιώντας για το L κάποια τιμή μεταξύ $2,6$ και $2,8$ (Montgomery, 2001). Όταν οι παρατηρήσεις είναι ανεξάρτητες και $\lambda = 0,2$ τότε για $L = 2,859$ θα δώσει το ίδιο εντός – ελέγχου ARL με ένα διάγραμμα Shewhart με $c = 3$, δηλαδή τριών τυπικών αποκλίσεων, και αυτό θα είναι $ARL = 370,4$ (Lu και Reynolds, 1999).

Σε αντίθεση με τα διαγράμματα Shewhart για μεμονωμένες παρατηρήσεις, για τα οποία, στην περίπτωση που οι παρατηρήσεις δεν κατανέμονται κανονικά, οι πραγματικές τιμές του ARL_0 είναι αξιοσημείωτα μικρότερες από αυτές που βασίζονται σε κανονική κατανομή, τα διαγράμματα EWMA δείχνουν να είναι ανθεκτικά στην έλλειψη κανονικότητας (Montgomery, 2001). Με άλλα λόγια, για μια κατανομή που αποκλίνει από την κανονικότητα, το διάγραμμα Shewhart μεμονωμένων παρατηρήσεων θα δίνει συγκριτικά με το αντίστοιχο διάγραμμα EWMA, πολύ περισσότερες λανθασμένες ενδείξεις ότι η διαδικασία είναι εκτός ελέγχου ενώ αυτό δεν αληθεύει. Υπάρχει λοιπόν συγκριτικό πλεονέκτημα του διαγράμματος EWMA έναντι του Shewhart για μεμονωμένες παρατηρήσεις, στην περίπτωση έλλειψης κανονικότητας των παρατηρήσεων.

Όπως έχει ήδη αναφερθεί, το διάγραμμα EWMA, όπως και το CUSUM, αντιλαμβάνεται καλά τις μικρές μεταβολές στον μέσο της διαδικασίας, όμως δεν ανταποκρίνεται τόσο γρήγορα στις μεγάλες αλλαγές όπως τα διαγράμματα Shewhart. Ειδικά για $\lambda > 0,10$ το EWMA είναι συχνά ανώτερο του CUSUM για μεγάλες αλλαγές (Montgomery, 2001). Ένας τρόπος για να βελτιωθεί η ευαισθησία του ελέγχου της διαδικασίας, όταν αυτή παρακολουθείται, στις μεγάλες μεταβολές χωρίς όμως να θυσιαστεί η ικανότητα γρήγορης αποκάλυψης μικρών μεταβολών, είναι να συνδυαστούν ένα διάγραμμα Shewhart με ένα διάγραμμα EWMA (Combined Shewhart – EWMA control procedures). Αυτός ο συνδυασμός είναι αποτελεσματικός έναντι τόσο των μεγάλων όσο και των μικρών μεταβολών. Όταν χρησιμοποιείται αυτός ο συνδυασμός, έχει αποδειχθεί ότι είναι προτιμότερο να χρησιμοποιούνται για

το διάγραμμα Shewhart ελαφρώς πλατύτερα όρια από τα συνηθισμένα (π.χ. 3,25 ή ακόμα και 3,5 τυπικών αποκλίσεων) (Montgomery, 2001). Κατά ανάλογο τρόπο, ένας συνδυασμός Shewhart με όρια προσεγγιστικά στις 3,5 τυπικές αποκλίσεις από την κεντρική γραμμή και ενός CUSUM, στον on – line έλεγχο, μπορεί να βελτιώσει την ικανότητα του CUSUM να αποκαλύπτει μεγάλες μετακινήσεις στη διεργασία, μειώνοντας σχετικά λίγο το εντός ελέγχου ARL (Lucas, 1982). Τότε θα λαμβάνονται εκτός ελέγχου σήματα είτε όταν η τυποποιημένη μετατόπιση είναι μεγαλύτερη από 3,5 είτε όταν τα S_H ή S_L υπερβαίνουν το h . Στα συνδυασμένα διαγράμματα Shewhart – EWMA και Shewhart – CUSUM χρησιμοποιούνται ευρύτερα όρια (από τα συνήθη των τριών τυπικών αποκλίσεων) για το Shewhart ώστε το ARL_0 να μην πέφτει κάτω από 370 που είναι το ARL_0 για ένα διάγραμμα Shewhart τριών τυπικών αποκλίσεων. Για $k=1/2$ και $h=5$ στο CUSUM (το οποίο μόνο του είχε $ARL_0 = 465$) και τυποποιημένα όρια Shewhart στο 3,5 το συνδυασμένο διάγραμμα Shewhart – CUSUM έχει $ARL_0 = 391$ (Lucas, 1982).

ΚΕΦΑΛΑΙΟ 7

ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΓΙΑ ΑΥΤΟΣΥΣΧΕΤΙΣΜΕΝΑ ΔΕΔΟΜΕΝΑ

7.1. ΕΙΣΑΓΩΓΗ

Τα τυπικά διαγράμματα ελέγχου που παρουσιάστηκαν στο προηγούμενο κεφάλαιο, αφορούν δεδομένα που διέπονται από την υπόθεση της ανεξαρτησίας. Στην πράξη όμως, η υπόθεση των ανεξάρτητων παρατηρήσεων δεν ικανοποιείται ούτε καν προσεγγιστικά σε μερικές βιομηχανικές διαδικασίες. Στις περιπτώσεις αυτές συμπεριλαμβάνονται οι χημικές διαδικασίες, όπου οι συνεχείς διαδοχικές μετρήσεις πάνω στη διαδικασία ή στα ποιοτικά χαρακτηριστικά του προϊόντος, είναι συχνά πολύ συσχετισμένες (Montgomery, 1991). Γενικά, οι χημικές διαδικασίες καθοδηγούνται από αδρανειακά στοιχεία. Συνεπώς, όταν το διάστημα μεταξύ των δειγματοληψιών γίνει σχετικά μικρό, οι παρατηρήσεις πάνω στη διαδικασία θα συσχετίζονται διαμέσου του χρόνου.

Η αυτοσυσχέτιση έχει σημαντικές επιπτώσεις στις ιδιότητες των τυπικών διαγραμμάτων ελέγχου του στατιστικού ελέγχου διεργασιών (SPC), τα οποία αναπτύσσονται και αποτιμώνται κάτω από την υπόθεση της ανεξαρτησίας. Ένα τυπικό αποτέλεσμα της αυτοσυσχέτισης, σ' αυτά τα διαγράμματα, είναι η δραματική μείωση του εντός – ελέγχου ARL. Όταν τα τυπικά διαγράμματα ελέγχου του SPC εφαρμόζονται σε αυτοσυσχετισμένα δεδομένα, οδηγούν σε πολλές ψευδείς θετικές ενδείξεις ότι η διαδικασία είναι εκτός ελέγχου ή/και σε ανεπάρκεια εντοπισμού πραγματικών θετικών ενδείξεων για εκτός ελέγχου κατάσταση (Harris και Ross, 1991; Lu και Reynolds, 1999b).

Στο Κεφάλαιο αυτό παρουσιάζονται οι κύριες μέθοδοι προσέγγισης στατιστικού ελέγχου αυτοσυσχετισμένων δεδομένων. Ιδιαίτερη έμφαση δίνεται στην κυριότερη από αυτές, η οποία είναι αρκετά πολύπλοκη και εξειδικευμένη καθώς απαιτεί τη

χρήση υποδειγμάτων (μοντέλων) χρονολογικών σειρών. Αρχικά λοιπόν, περιγράφονται οι βασικές έννοιες καθώς και τα υποδείγματα χρονολογικών σειρών.

7.2. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΤΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Η ανάλυση χρονολογικών σειρών αποσκοπεί στην κατασκευή ενός υποδείγματος, το οποίο σύμφωνα με τη θεωρία των πιθανοτήτων, έχει παρόμοιες ιδιότητες με αυτές του μηχανισμού που παράγει τη στοχαστική διαδικασία, της οποίας μια πραγματοποίηση είναι η χρονολογική σειρά των διαδοχικών παρατηρήσεων (Box και Λοιποί, 1994; Δημέλη, 1996). Για να επιλεγεί το κατάλληλο υπόδειγμα, αρχικά θα πρέπει να εξεταστεί η μορφή της δομής της χρονολογικής σειράς. Να διαπιστωθεί εάν η χρονολογική σειρά είναι στάσιμη ή όχι και να μελετηθούν οι συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.

Μια στοχαστική διαδικασία χαρακτηρίζεται ως στάσιμη όταν οι στατιστικές της ιδιότητες δεν επηρεάζονται από μια μεταβολή στην αρχή του χρόνου. Δηλαδή, οι στατιστικές ιδιότητες των N παρατηρήσεων με αρχή t ($x_t, x_{t+1}, \dots, x_{t+N-1}$) είναι οι ίδιες με τις στατιστικές ιδιότητες των N παρατηρήσεων με αρχή την περίοδο $t+k$ ($x_{t+k}, x_{t+k+1}, \dots, x_{t+k+N-1}$). Γενικά μια χρονολογική σειρά είναι στάσιμη εάν ο μέσος και η διακύμανσή της δεν μεταβάλλονται με το χρόνο και η συνδιακύμανση μεταξύ των τιμών της σε δύο χρονικά σημεία εξαρτάται μόνο από την απόσταση ανάμεσα σ' αυτά τα χρονικά σημεία, και όχι από τον ίδιο το χρόνο. Σε μία στάσιμη χρονολογική σειρά, ο μέσος και η διακύμανσή της μπορούν να εκτιμηθούν από τον μέσο και τη διακύμανση αντίστοιχα του δείγματος των παρατηρήσεων της χρονολογικής σειράς. Μη στάσιμες είναι οι σειρές που παρουσιάζουν τάση ή εποχική διακύμανση. Αν σε μια χρονολογική σειρά ο μέσος μετατοπίζεται διαχρονικά τότε μιλάμε για σειρά μη στάσιμη ως προς τον μέσο ενώ εάν η διακύμανσή της δεν παραμένει σταθερή τότε μιλάμε για μη στασιμότητα ως προς τη διακύμανση.

Στις χρονολογικές σειρές ποιοτικών χαρακτηριστικών χημικών διαδικασιών, οι παρατηρήσεις παρουσιάζουν θετική κυρίως αυτοσυσχέτιση. Εμπειρικά, θα μπορούσαμε να περιγράψουμε την κατάσταση αυτή ως εξής: Έστω για παράδειγμα ότι υπάρχουν αρκετές παρατηρήσεις για ένα ποιοτικό χαρακτηριστικό x_t μιας

διαδικασίας και από το διάγραμμα τους αποκαλύπτεται ότι η συμπεριφορά της μεταβλητής δεν είναι τυχαία με την έννοια ότι μια τιμή της x_t που είναι πάνω από τον μακροπρόθεσμο μέσο όρο τείνει να ακολουθείται από άλλες τιμές που να βρίσκονται πάνω από τον μέσο όρο, ενώ μια τιμή κάτω από τον μέσο τείνει να ακολουθείται από άλλες παρόμοιες τιμές. Το γεγονός αυτό αντανακλάται και στο διάγραμμα διασποράς της x_t (η παρατήρηση τη χρονική στιγμή t) ως προς την x_{t-1} (η παρατήρηση τη χρονική στιγμή $t-1$, δηλαδή μια περίοδο νωρίτερα). Σ' αυτό θα παρατηρήσουμε ότι οι παρατηρήσεις μαζεύονται γύρω από μια ευθεία γραμμή με θετική κλίση. Δηλαδή μια σχετικά χαμηλή τιμή παρατήρησης πάνω στην x τη χρονική στιγμή $t-1$ τείνει να ακολουθείται από μια άλλη χαμηλή τιμή τη χρονική στιγμή t , ενώ μια σχετικά μεγάλη παρατήρηση τη χρονική στιγμή $t-1$ τείνει να ακολουθείται από μια άλλη μεγάλη τιμή στον χρόνο t . Αυτός ο τύπος συμπεριφοράς δηλώνει θετική αυτοσυσχέτιση στις παρατηρήσεις. Υπάρχει επίσης η δυνατότητα να μετρηθεί ο βαθμός της αυτοσυσχέτισης αναλυτικά. Η αυτοσυσχέτιση σε παρατηρήσεις μιας χρονολογικής σειράς, μετριέται με την συνάρτηση αυτοσυσχέτισης:

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{V(X_t)}, \quad k = 0, 1, \dots$$

Όπου $Cov(x_t, x_{t-k})$ είναι η συνδιακύμανση των παρατηρήσεων που απέχουν k χρονικές περιόδους, και έχουμε υποθέσει ότι οι παρατηρήσεις έχουν σταθερή διακύμανση $V(x_t)$. Οι τιμές του ρ_k βρίσκονται μέσα στα όρια $-1 \leq \rho_k \leq 1$. Για $k = 0$, ο συντελεστής αυτοσυσχέτισης ρ_0 ισούται με τη μονάδα μιας και η συνδιακύμανση υπολογίζεται για κάθε παρατήρηση με τον εαυτό της. Αν δύο παρατηρήσεις που απέχουν μεταξύ τους k χρονικές περιόδους έχουν μεγάλη σχέση μεταξύ τους, τότε η τιμή του ρ_k θα είναι κοντά στη μονάδα. Συνήθως, οι τιμές της ρ_k εκτιμώνται με τη δειγματική συνάρτηση αυτοσυσχέτισης:

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad k = 1, 2, \dots, K$$

Σαν ένα γενικό κανόνα, συνήθως χρειάζεται να υπολογιστούν τιμές του r_k για μερικές τιμές του k , $k \leq n/4$. Αν ο αριθμός των παρατηρήσεων της χρονολογικής σειράς είναι μεγάλος, τότε η εκτιμηθείσα συνάρτηση αυτοσυσχέτισης θα προσεγγίζει

ικανοποιητικά την αληθινή αυτοσυσχέτιση. Το διάγραμμα της δειγματικής συνάρτησης αυτοσυσχέτισης ACF είναι χρήσιμο για την αποκάλυψη των μη μηδενικών αυτοσυσχετίσεων. Στην πράξη, εάν η δειγματική αυτοσυσχέτιση υπερβαίνει τα όρια των δύο τυπικών αποκλίσεων για την παράμετρο ρ_k στην υστέρηση k , τότε η αντίστοιχη παράμετρος ρ_k είναι πιθανώς μη μηδενική. Το διάγραμμα της δειγματικής συνάρτησης αυτοσυσχέτισης έναντι της υστέρησης k , βοηθάει στο να διαπιστωθεί εάν μία χρονολογική σειρά είναι στάσιμη ή όχι. Στις μη στάσιμες χρονολογικές σειρές η συνάρτηση αυτοσυσχέτισης τείνει προς το μηδέν με σχετικά αργό ρυθμό, ενώ όταν η σειρά είναι στάσιμη φθίνει γοργά.

Μια άλλη συνάρτηση που χρησιμοποιείται στη μελέτη των χαρακτηριστικών μιας χρονολογικής σειράς, είναι η συνάρτηση μερικής αυτοσυσχέτισης (PACF). Στην περίπτωση μιας χρονολογικής σειράς, ο συντελεστής μερικής αυτοσυσχέτισης μεταξύ x_t και x_{t+k} ορίζεται ως ο συντελεστής συσχέτισης μεταξύ τους όταν έχουν ληφθεί υπόψη οι συσχετίσεις όλων των ενδιάμεσων τιμών $x_{t+1}, x_{t+2}, \dots, x_{t+k-1}$. Η έννοια της μερικής αυτοσυσχέτισης μπορεί να εξηγηθεί καλύτερα μέσα από την παλινδρόμηση. Γενικότερα, ο συντελεστής μερικής αυτοσυσχέτισης p – τάξης συμβολίζεται ϕ_{pp} και είναι ο συντελεστής του x_{t-p} στην παλινδρόμηση:

$$x_t = \phi_{1p} x_{t-1} + \phi_{2p} x_{t-2} + \dots + \phi_{pp} x_{t-p} + \varepsilon_t$$

Ο πρώτος υποδείκτης του ϕ δηλώνει τη χρονική υστέρηση της μεταβλητής και ο δεύτερος τη μέγιστη τάξη της παλινδρόμησης. Οπότε ο συντελεστής ϕ_{pp} του x_{t-p} μετρά τον συντελεστή μερικής αυτοσυσχέτισης p τάξης, μιας και δείχνει τη συσχέτιση μεταξύ x_t και x_{t-p} όταν έχουν ήδη συμπεριληφθεί στην παλινδρόμηση οι $x_{t-1}, \dots, x_{t-p+1}$.

Οι τιμές των ϕ_{pp} για διάφορες τιμές του $p = 1, 2, \dots$ αποτελούν τη συνάρτηση μερικής αυτοσυσχέτισης (PACF). Η μορφή αυτής της συνάρτησης καθώς και της συνάρτησης αυτοσυσχέτισης παίζουν καθοδηγητικό ρόλο στον προσδιορισμό της μορφής της стоχαστικής διαδικασίας που δημιούργησε τη υπό μελέτη χρονολογική σειρά. Εννοείται ότι στην περίπτωση που τα δεδομένα είναι ασυσχέτιστα, οι τιμές και των δύο συναρτήσεων δεν διαφέρουν στατιστικά σημαντικά από το μηδέν. Τότε βέβαια, μπορούν να εφαρμοστούν τα τυπικά διαγράμματα ελέγχου του στατιστικού ελέγχου

διεργασιών. Διαφορετικά, εάν τα δεδομένα είναι συσχετισμένα, οι τιμές αυτών των συναρτήσεων διαφέρουν στατιστικά σημαντικά από το μηδέν.

7.3. ΥΠΟΔΕΙΓΜΑΤΑ ARIMA ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ARIMA

7.3.1. Εισαγωγή

Τα υποδείγματα ARIMA λαμβάνουν την ονομασία τους από τις τρεις συνιστώσες που έχουν, οι οποίες αποκαλούνται ως: Αυτοπαλίνδρομη (AutoRegressive), Ολοκληρωμένη (Integrated) και Κινητού Εύρους (Moving Average). Η τάξη της αυτοπαλίνδρομης συνιστώσας συμβολίζεται ως p , η τάξη της συνιστώσας κινητού μέσου ως q και η τάξη της ολοκλήρωσης (διαφορών) ως d . Συνεπώς η γενική μορφή ενός υποδείγματος ARIMA είναι $ARIMA(p, d, q)$. Όταν η τάξη κάποιας από τις συνιστώσες είναι μηδέν, το αντίστοιχο συνθετικό και η τιμή της τάξης συνήθως παραλείπονται χάριν συντομίας. Στην περίπτωση ύπαρξης τυχόν εποχικών διακυμάνσεων, τα υποδείγματα ARIMA μπορούν να γενικευτούν έτσι ώστε να ενσωματώσουν το εποχικό πρότυπο των δεδομένων (Box και λοιποί, 1994). Στην ενότητα αυτή περιγράφονται τα βασικά υποδείγματα χρονολογικών σειρών ARIMA και η μεθοδολογία κατασκευής τους.

7.3.2. Αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως AR(1)

Σε ένα αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως, κάθε μεμονωμένη τιμή x_t παλινδρομείται πάνω στην προηγούμενή της. Η γενική μορφή του υποδείγματος είναι:

$$x_t = \delta + \alpha x_{t-1} + \varepsilon_t$$

όπου δ και α ($-1 < \alpha < 1$) είναι άγνωστες σταθερές και ε_t μετρά τα τυχαία σφάλματα τα οποία θεωρούνται λευκός θόρυβος δηλαδή ανεξάρτητες τυχαίες μεταβλητές που κατανέμονται κανονικά με μέσο το μηδέν και σταθερή διακύμανση σ^2 .

Οι παρατηρήσεις x_t από ένα τέτοιο μοντέλο, έχουν μέσο $\mu = \delta / (1 - \alpha)$, τυπική απόκλιση $\sigma / (1 - \alpha^2)^{1/2}$, και οι παρατηρήσεις που απέχουν μεταξύ τους k περιόδους (x_t και x_{t-k}) έχουν συντελεστή συσχέτισης α^k . Δηλαδή, η συνάρτηση αυτοσυσχέτισης για το αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως, αρχίζει από τη μονάδα, $\rho_0 = 1$, και

μειώνεται γεωμετρικά τείνοντας προς το μηδέν καθώς το k μεγαλώνει. Είναι προφανές ότι ο συντελεστής μερικής αυτοσυσχέτισης πρώτης τάξης ϕ_{11} ταυτίζεται με τον απλό συντελεστή αυτοσυσχέτισης ρ_1 .

Η ανάλυση που προηγήθηκε για το AR(1) υπόδειγμα βασίζεται στην υπόθεση ότι η χρονολογική σειρά είναι στάσιμη. Για να είναι μια AR(1) διαδικασία στάσιμη, θα πρέπει $|a| < 1$.

7.3.3.. Αυτοπαλίνδρομα υποδείγματα μεγαλύτερης τάξεως

Μια προφανής επέκταση στην εξίσωση του αυτοπαλίνδρομου υποδείματος πρώτης τάξης, είναι η:

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t$$

η οποία περιγράφει το αυτοπαλίνδρομο μοντέλο δευτέρας τάξης. Για την στασιμότητα του AR(2) υποδείματος, οι παράμετροι α_1 και α_2 θα πρέπει να παίρνουν τέτοιες τιμές ώστε:

$$\alpha_2 + \alpha_1 < 1$$

$$\alpha_2 - \alpha_1 < 1$$

$$|\alpha_2| < 1$$

Η θεωρητική συνάρτηση αυτοσυσχέτισης, μπορεί να γραφεί σαν συνάρτηση των παραμέτρων του AR(2) υποδείματος, ως:

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} \quad , \quad k > 0$$

Ο συντελεστής μερικής αυτοσυσχέτισης δευτέρας τάξεως ϕ_{22} συμπίπτει με τον συντελεστή α_2 στο AR(2) υπόδειγμα.

Γενικά, στο αυτοπαλίνδρομο υπόδειγμα p τάξης, AR(p), η μεταβλητή x_t εξαρτάται άμεσα από τις προηγούμενες παρατηρήσεις x_{t-1} , x_{t-2} , ..., x_{t-p} .

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t$$

Παρόμοια με τα AR(1) και AR(2) υποδείγματα, μπορούμε να εκφράσουμε τις παραμέτρους ενός AR(p) υποδείγματος σε όρους των συντελεστών αυτοσυσχέτισης ρ_k , $k = 1, 2, \dots, p$, αλλά ισχύει και το αντίθετο. Οι σχέσεις αυτές είναι γνωστές ως εξισώσεις των Yule – Walker, και προκύπτουν λύνοντας το σύστημα των εξισώσεων:

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \dots + \alpha_p \rho_{k-p}, \quad \text{για } k = 1, 2, \dots, p$$

Στην πράξη οι συντελεστές μερικής αυτοσυσχέτισης ϕ_{pp} υπολογίζονται σε όρους των συντελεστών αυτοσυσχέτισης ρ_k επιλύοντας το σύστημα των Yule – Walker για $k = 1, 2, \dots, p$. Η εξίσωση που αντιστοιχεί στην παράμετρο α_p είναι ο συντελεστής μερικής αυτοσυσχέτισης ϕ_{pp} .

Για μια διαδικασία AR(p) οι αυτοσυσχετίσεις ρ_j φθίνουν γεωμετρικά από το ρ_p , ενώ οι μερικές αυτοσυσχετίσεις είναι μηδέν μετά το ϕ_{pp} .

7.3.4. Υποδείγματα κινητών μέσων

Η γενική μορφή ενός τέτοιου υποδείγματος είναι:

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

όπου τα θ είναι σταθερές παράμετροι και ε_t είναι λευκός θόρυβος. Στο υπόδειγμα αυτό υποθέτουμε ότι η χρονολογική σειρά x_t δημιουργείται σαν ένας σταθμικός μέσος των τυχαίων σφαλμάτων των q προηγούμενων περιόδων και ονομάζεται υπόδειγμα κινητού μέσου τάξεως q , MA(q). Κάθε υπόδειγμα MA(q) είναι πάντα στάσιμο για πεπερασμένο q . Η συνάρτηση αυτοσυσχέτισης γενικά για το υπόδειγμα MA(q) δίνεται από τον τύπο:

$$\rho_k = \begin{cases} \frac{\sum_{i=1}^q \theta_i \theta_{i+k} - \theta_k}{1 + \sum_{i=1}^q \theta_i^2} & \text{για } k = 1, 2, \dots, q \\ 0 & \text{για } k > q \end{cases}$$

Επομένως η τάξη ενός υποδείγματος MA ανταποκρίνεται στο μέγιστο k για το οποίο η αυτοσυσχέτιση ρ_k είναι μη μηδενική. Ενώ σε μια διαδικασία MA(q) οι αυτοσυσχετίσεις ρ_k μηδενίζονται για $k > q$, οι μερικές αυτοσυσχετίσεις φθίνουν από το ϕ_{qq} .

Στο υπόδειγμα κινητού μέσου πρώτης τάξης MA(1), $x_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}$, η διακύμανση είναι $\text{Var}(x_t) = \sigma^2(1 + \theta^2)$. Η συσχέτιση μεταξύ x_t και x_{t-1} είναι $\rho_1 = -\theta / (1 + \theta^2)$ και είναι μηδέν για όλες τις άλλες υστερήσεις, δηλαδή εκτείνεται προς τα πίσω μόνο μία χρονική περίοδο.

7.3.5. Μεικτά υποδείγματα ARMA(p,q)

Αν τα δεδομένα μιας χρονολογικής σειράς έχουν συναρτήσεις αυτοσυσχέτισης ή μερικής αυτοσυσχέτισης που δεν φαίνονται να μηδενίζονται μετά από κάποιο σημείο αλλά φθίνουν και οι δύο με αργό ρυθμό, τότε η χρονολογική σειρά έχει στοιχεία και των δύο μορφών AR και MA. Τότε, χρειάζονται υποδείγματα που να περιέχουν και τα δύο παραπάνω μέρη, οπότε προκύπτουν τα μεικτά υποδείγματα ARMA. Η γενική μορφή ενός ARMA(p,q) υποδείγματος είναι:

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Σε ένα τέτοιο υπόδειγμα οι αυτοσυσχετίσεις φθίνουν γεωμετρικά από το ρ_q ενώ οι μερικές αυτοσυσχετίσεις φθίνουν από το ϕ_{pp} .

Το υπόδειγμα ARMA(1,1), $x_t = \delta + \alpha x_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$, συναντάται αρκετά συχνά στις βιομηχανίες χημικών διεργασιών. Εάν η υποκείμενη μεταβλητή x_t της διεργασίας είναι αυτοπαλίνδρομη πρώτης τάξης (δηλαδή η x_t παλινδρομείται πάνω στην x_{t-1}) και μια συνιστώσα τυχαίου λάθους προστίθεται στην x_t , το αποτέλεσμα είναι το μεικτό μοντέλο ARMA(1,1) (Montgomery, 2001). Επιπλέον, το ποιοτικό χαρακτηριστικό μετριέται συχνά σε ένα εργαστήριο (ή με ένα on – line όργανο) που έχει σφάλμα μέτρησης, το οποίο συνήθως να θεωρηθεί σαν τυχαίο ή ασυσχέτιστο. Τότε, η αναφερόμενη ή παρατηρούμενη μέτρηση, παραπέμπει σε μια αυτοπαλίνδρομη συνιστώσα συν την τυχαία κύμανση κι έτσι το μεικτό υπόδειγμα ARMA(1,1)

απαιτείται σαν υπόδειγμα της διεργασίας. Η διακύμανση των παρατηρήσεων και η συνάρτηση αυτοσυσχέτισης στο ARMA(1,1) υπόδειγμα για $|\alpha| < 1$ και $|\theta| < 1$ είναι:

$$\text{Var}(x_t) = (1 + \theta^2 - 2\alpha\theta)\sigma^2 / (1 - \alpha^2)$$

$$\text{και } \rho_k = \begin{cases} \alpha - \theta\sigma^2 / \text{Var}(x_t) & , \text{ για } k = 1 \\ \alpha\rho_{k-1} & , \text{ για } k > 1 \end{cases}$$

όπου σ^2 είναι η διακύμανση των τυχαίων λαθών ε_t (τα ε_t κατανέμονται κανονικά με μέσο μηδέν και διακύμανση σ^2).

7.3.6. Υποδείγματα ARIMA(p,d,q)

Τα υποδείγματα που προαναφέρθηκαν βασίζονται στην υπόθεση ότι οι χρονολογικές σειρές είναι στάσιμες (δηλαδή x_t να ‘περιπλανάται’ γύρω από έναν ‘σταθερό’ μέσο). Αν σε μια χρονολογική σειρά ο μέσος μετατοπίζεται διαχρονικά τότε η σειρά θεωρείται μη στάσιμη ως προς τον μέσο ενώ εάν η διακύμανσή της δεν παραμένει σταθερή τότε υπάρχει μη στασιμότητα ως προς τη διακύμανση. Σύμφωνα με τους Box και Jenkins (1994) οι μη στάσιμες σειρές μπορούν να μετατραπούν σε στάσιμες, παίρνοντας τις διαφορές πρώτης, δεύτερης ή και μεγαλύτερης τάξης d , ανάλογα με τον αν παρουσιάζουν γραμμική τάση ή δευτεροβάθμια τάση ή τάση πολυωνυμικής μορφής d αντίστοιχα. Αφού εξασφαλιστεί η στασιμότητα της σειράς, με τις διαφορές, στη συνέχεια ακολουθεί η προσαρμογή του κατάλληλου ARMA(p,q) υποδείματος στη μετασχηματισμένη σειρά. Τότε το υπόδειγμα που προσαρμόζεται στην αρχική σειρά, ονομάζεται αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητών μέσων τάξεως (p,d,q) και συμβολίζεται ως ARIMA(p,d,q). Αν η χρονολογική σειρά δείχνει να έχει αυξανόμενη διακύμανση (μη σταθερή ως προς τη διακύμανση), τότε χρησιμοποιείται λογαριθμικός μετασχηματισμός πριν τις διαφορές (χρησιμοποιούνται δηλαδή προσεγγιστικά οι ρυθμοί μεταβολής της σειράς). Τα υποδείγματα ARIMA, παρόλο που φαίνονται πολύ διαφορετικά από τα υποδείγματα Shewhart ($x_t = \mu + \varepsilon_t$), είναι σχετικά παρόμοια και περιλαμβάνουν τα υποδείγματα Shewhart σαν ειδικές περιπτώσεις (π.χ. για $\alpha = 0$ ή για $\theta = 0$ από το AR(1) ή το MA(1) υπόδειγμα αντίστοιχα, έχουμε σαν αποτέλεσμα Shewhart υπόδειγμα).

Το υπόδειγμα $ARIMA(0,1,1)$, $x_t = \delta + x_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$ (πρώτης τάξης ολοκληρωμένο κινητών μέσων), εμφανίζεται συχνά σε εργοστάσια χημικών διεργασιών, όταν δεν γίνονται ενέργειες ελέγχου ώστε η μεταβλητή x_t να παίρνει τιμές κοντά στην τιμή στόχο.

7.4. ΕΥΡΕΣΗ ΚΑΤΑΛΛΗΛΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ ARIMA

Για την εύρεση κατάλληλου υποδείγματος $ARIMA(p,d,q)$ ακολουθείται η Μεθοδολογία Box–Jenkins (Box, Jenkins, και Reinsel, 1994) η οποία περιλαμβάνει τα εξής στάδια:

- Εξειδίκευση του αριθμού d των διαφορών της αρχικής σειράς για την επίτευξη στασιμότητας.
- Ταυτοποίηση δυνητικών υποδειγμάτων με εξέταση των δειγματικών συναρτήσεων αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) (εξειδίκευση των τάξεων p, q).
- Εκτίμηση των παραμέτρων στα δυνητικά υποδείγματα.
- Διαγνωστικός έλεγχος: είναι τα κατάλοιπα λευκός θόρυβος (ασυσχέτιστα και κανονικά κατανομημένα με σταθερή διακύμανση);

Η επιλογή του βέλτιστου υποδείγματος γίνεται με χρήση κατάλληλων κριτηρίων όπως τα: Μέσο τετραγωνικό σφάλμα (Mean Squared Error, MSE), το κριτήριο Akaike (Akaike's Information Criterion, AIC), και το κριτήριο Schwartz (Schwartz Bayesian Information Criterion, BIC). Το MSE είναι ο αριθμητικός μέσος των τετραγωνικών σφαλμάτων (καταλοίπων). Κατά τη σύγκριση δύο ή περισσότερων υποδειγμάτων, το υπόδειγμα με την καλύτερη προσαρμογή είναι αυτό με το μικρότερο MSE. Η μεθοδολογία Box-Jenkins προτάσσει ότι το επιλεγέν υπόδειγμα πρέπει να έχει ως όσο το δυνατόν λιγότερες παραμέτρους από την άποψη της φειδωλότητας (απλότητας). Τα κριτήρια AIC (Akaike, 1974) και BIC (Schwartz, 1978) λαμβάνουν υπόψη τους τον αριθμό των όρων στα υποδείγματα. Τα κριτήρια αυτά αντισταθμίζουν τη μείωση των καταλοίπων, από την προσθήκη περισσότερων παραμέτρων, με την αρχή της «οικονομίας» των υποδειγμάτων, δηλαδή της όσο δυνατόν μικρότερης τάξης τους (DeLurgio, 1998). Σύμφωνα με τα κριτήρια αυτά, καταλληλότερο είναι το υπόδειγμα με την μικρότερη τιμή. Εντούτοις, ο όρος

«ποινικής ρήτρας» του κριτηρίου BIC είναι πιο αυστηρός από αυτόν του κριτηρίου AIC. Επομένως, το κριτήριο BIC τείνει να επιλέγει υποδείγματα τα οποία είναι περισσότερο φειδωλά (parsimonious) από αυτά που επιλέγονται με το κριτήριο AIC.

7.5. ΔΙΑΓΝΩΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΚΑΤΑΛΛΗΛΟΤΗΤΑΣ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Μετά την ταυτοποίηση (εξειδίκευση του αριθμού d των διαφορών της αρχικής σειράς που απαιτούνται για την επίτευξη στασιμότητας του ARIMA υποδείγματος και των υστερήσεων p και q των AR και MA) και την εκτίμηση του υποδείγματος, ελέγχεται η καταλληλότητά του με την έννοια του πόσο καλά προσαρμόζεται στα δεδομένα (Box, Jenkins και Reinsel, 1994).

Αν \hat{x}_t είναι η εκτιμηθείσα τιμή της x_t η οποία προκύπτει από την ανάλυση του δείγματος των δεδομένων από τη διαδικασία, τότε τα κατάλοιπα είναι $e_t = x_t - \hat{x}_t$. Το υπόδειγμα θεωρείται κατάλληλο όταν τα κατάλοιπα προσεγγιστικά είναι κανονικά και ανεξάρτητα κατανομημένα με μέσο μηδέν και σταθερή διακύμανση. Επομένως οι τιμές της δειγματικής συνάρτησης αυτοσυσχέτισης των καταλοίπων, δεν θα πρέπει να διαφέρουν στατιστικά σημαντικά από το μηδέν. Εάν εντούτοις δεν συμβαίνει κάτι τέτοιο, η διαδικασία δεν έχει ταυτοποιηθεί σωστά και θα πρέπει να δοκιμαστούν άλλα υποδείγματα.

7.6. ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΔΙΕΡΓΑΣΙΩΝ ΜΕ ΧΡΗΣΗ ΥΠΟΔΕΙΓΜΑΤΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Η κυριότερη μέθοδος στατιστικού ελέγχου διεργασιών για αυτοσυσχετισμένα δεδομένα, προτάθηκε αρχικά από τους Alwan και Roberts το 1988. Κατά την προσέγγιση αυτή, αρχικά προσαρμόζεται στα δεδομένα ένα κατάλληλο υπόδειγμα ARIMA χρονολογικών σειρών με το οποίο περιγράφεται η συστηματική διακύμανση των δεδομένων. Στη συνέχεια χρησιμοποιούνται δύο διαγράμματα για την παρακολούθηση της διεργασίας. Το πρώτο από αυτά, αναφέρεται ως «διάγραμμα

κοινών αιτιών» (Common Cause Chart - CCC) και το δεύτερο ως «διάγραμμα ειδικών αιτιών» (Special Cause Chart - SCC).

Το CCC είναι ένα διάγραμμα των εκτιμηθεισών τιμών βάσει του υποδείγματος ARIMA. Δεν έχει όρια ελέγχου και συνεπώς δεν είναι διάγραμμα ελέγχου. Δίνει απλώς μια εικόνα του τρέχοντος επιπέδου της διεργασίας και την εξέλιξή της στο χρόνο. Στην ουσία, το διάγραμμα CCC ερμηνεύει τη συστηματική διακύμανση της διεργασίας (Wardell, Moskowitz και Plante 1994). Όταν τα δεδομένα είναι αυτοσυσχετισμένα, η συστηματική διακύμανση σ' αυτά είναι πολύ μεγαλύτερη απ' αυτή που οφείλεται σε ειδικές αιτίες και συνεπώς περισσότερο σημαντική όσον αφορά την επιρροή στην ποιότητα παραγωγής.

Το SCC είναι ένα τυπικό διάγραμμα ελέγχου Shewhart που εφαρμόζεται στα κατάλοιπα (δηλαδή τη διαφορά μεταξύ των πραγματικών τιμών της διεργασίας και των εκτιμήσεών τους). Με τη μεθοδολογία αυτή, ουσιαστικά αφαιρείται η αυτοσυσχέτιση από τα δεδομένα, αφού τα κατάλοιπα που εντέλει χρησιμοποιούνται για τα διαγράμματα ελέγχου, εφόσον το μοντέλο εξηγεί επαρκώς τα δεδομένα, είναι ασυσχέτιστα. Ουσιαστικά δηλαδή, το κατάλληλο υπόδειγμα χρονολογικής σειράς λειτουργεί σαν ένα φίλτρο που μετατρέπει τα αρχικά συσχετισμένα δεδομένα σε ασυσχέτιστα, οπότε να μπορούν στη συνέχεια να εφαρμοστούν τα τυπικά διαγράμματα ελέγχου του SPC.

Στο διάγραμμα SCC η κεντρική γραμμή είναι μηδέν, μιας και ο μέσος των καταλοίπων είναι μηδέν. Η τυπική απόκλιση που χρησιμοποιείται σ' αυτή την περίπτωση είναι η εκτίμηση της τυπικής απόκλισης των καταλοίπων σ και όχι της τυπικής απόκλισης των παρατηρήσεων που για διάκριση μπορεί να συμβολίζεται με σ_x . Επομένως τα όρια ελέγχου του διαγράμματος SCC θα είναι:

$$UCL = + c\bar{\sigma}$$

$$CL = 0$$

$$LCL = - c\bar{\sigma}$$

Όπου ο πολλαπλασιαστής c είναι και εδώ συνήθως ίσος με 3.

Σημεία εκτός ελέγχου ή ασυνήθιστες μορφές σε τέτοια διαγράμματα, δηλώνουν ότι οι παράμετροι του υποδείγματος έχουν αλλάξει και συνεπώς η αρχική μεταβλητή x_i ήταν εκτός ελέγχου.

Δεδομένου ότι το διάγραμμα SCC βασίζεται στη υπόθεση ότι τα κατάλοιπα είναι τυχαία, τηρούνται όλες οι προϋποθέσεις των τυπικών διαγραμμάτων ελέγχου του SPC και συνεπώς μπορεί να χρησιμοποιηθεί κάθε τυπική μέθοδος του SPC, συμπεριλαμβανομένων των κανόνων ροής (run rules) για αποκλίσεις από την τυχειότητα, διαγράμματα CUSUM κλπ. (Wardell και λοιποί, 1994). Επιπλέον, σε αντίθεση με άλλες μεθόδους για αυτοσυσχετισμένα δεδομένα, οι οποίες έχουν περιοριστεί σε υποδείγματα AR(1) και MA(1) χρονολογικών σειρών, η προσέγγιση των Alwan και Roberts μπορεί να εφαρμοστεί σε οποιοδήποτε υπόδειγμα χρονολογικών σειρών.

Το βασικό μειονέκτημα της μεθόδου των Alwan & Roberts είναι ότι προαπαιτούνται γνώσεις χρονολογικών σειρών. Επίσης, η διαδικασία εξεύρεσης κατάλληλου υποδείγματος είναι χρονοβόρα και επιπλέον χρειάζεται μεγάλος αριθμός δεδομένων (Lu και Reynolds, 1999a) για να βρεθεί το κατάλληλο υπόδειγμα χρονολογικής σειράς που κρύβεται πίσω από τα δεδομένα.

7.7. ΣΥΓΚΡΙΣΗ ΔΙΑΓΡΑΜΜΑΤΩΝ ΕΛΕΓΧΟΥ ΤΩΝ ΚΑΤΑΛΟΙΠΩΝ ΚΑΙ ΤΥΠΙΚΩΝ ΔΙΑΓΡΑΜΜΑΤΩΝ ΕΛΕΓΧΟΥ SPC ΤΩΝ ΑΡΧΙΚΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

Οι Lu και Reynolds (1999a) μελέτησαν τη διεργασία που μπορεί να εκφραστεί μέσω ενός αυτοπαλίνδρομου υποδείγματος πρώτης τάξης με προστιθέμενο λευκό θόρυβο (ασυσχέτιστο λάθος). Αυτό είναι ισοδύναμο με το πρώτης τάξης μεικτό υπόδειγμα ARMA(1,1). Ασχολήθηκαν εκτενώς με το διάγραμμα EWMA που εφαρμόζεται απευθείας στα δεδομένα και το EWMA των καταλοίπων. Επίσης ασχολήθηκαν και με διαγράμματα ελέγχου των αρχικών παρατηρήσεων με όρια ελέγχου που είναι προσαρμοσμένα να εξηγήσουν τους λόγους της αυτοσυσχέτισης. Πρόκειται για την ακρίβεια, για διαγράμματα Shewhart και EWMA στα οποία χρησιμοποιείται η εκτίμηση της τυπικής απόκλισης των παρατηρήσεων, η οποία υπολογίζεται βάσει των

εκτιμημένων παραμέτρων κατάλληλου υποδείγματος χρονολογικών σειρών που έχει ήδη προσαρμοστεί στα δεδομένα. Η εκτίμηση αυτή της τυπικής απόκλισης των παρατηρήσεων, θα είναι πιο κοντά στην πραγματική τιμή απ' ό,τι η εκτίμηση μέσω των κινητών μέσων. Τα όρια όλων των διαγραμμάτων που μελέτησαν, προσαρμόστηκαν ρυθμίζοντας την σταθερά c (Shewhart) ή L (EWMA) έτσι ώστε το εντός – ελέγχου ARL να είναι το ίδιο (370,4) και συγκρίθηκαν τα ARL για διάφορες μετατοπίσεις του μέσου της διεργασίας καθώς και τα πραγματικά εντός – ελέγχου ARL (βάσει των στοιχείων που προκύπτουν από τα διαγράμματα). Μερικά από τα συμπεράσματά τους είναι:

- Όταν υπάρχει σημαντική αυτοσυσχέτιση στα δεδομένα της διεργασίας και αυτή η αυτοσυσχέτιση είναι ένα ενυπάρχον μέρος της διεργασίας, δεν θα πρέπει να χρησιμοποιούνται παραδοσιακές μέθοδοι εκτίμησης των παραμέτρων της διεργασίας και κατασκευής των διαγραμμάτων ελέγχου. Αντί αυτών, θα πρέπει να γίνει προσπάθεια μοντελοποίησης της αυτοσυσχέτισης ώστε να μπορούν να κατασκευαστούν αξιόπιστα διαγράμματα ελέγχου.
- Απαιτείται να χρησιμοποιείται ένα μεγάλο σετ δεδομένων στη διεργασία για την προσαρμογή ενός υποδείγματος για τις παρατηρήσεις της διεργασίας και την εκτίμηση των παραμέτρων του μοντέλου αυτού. Εάν πρέπει να κατασκευαστεί ένα διάγραμμα ελέγχου χρησιμοποιώντας ένα μικρό σετ δεδομένων, τότε οι ενδείξεις του διαγράμματος πρέπει να διερμηνεύονται με προσοχή και η διαδικασία προσαρμογής του μοντέλου και εκτίμησης των παραμέτρων να επαναλαμβάνεται μόλις υπάρχουν διαθέσιμα πρόσθετα δεδομένα.
- Για χαμηλά έως μέτρια επίπεδα συσχέτισης, ένα διάγραμμα Shewhart των παρατηρήσεων, θα είναι πολύ καλύτερο στην ανακάλυψη των μετακινήσεων του μέσου της διεργασίας απ' ό,τι ένα διάγραμμα Shewhart των καταλοίπων. Ένα διάγραμμα EWMA θα είναι καλύτερο απ' ό,τι ένα διάγραμμα Shewhart, εκτός κι αν ενδιαφερόμαστε μόνο στην ανακάλυψη μεγάλων μετακινήσεων. Για μεγάλες μετακινήσεις ένα διάγραμμα EWMA των καταλοίπων θα είναι καλύτερο από ένα διάγραμμα EWMA των παρατηρήσεων, ενώ για μικρές μετακινήσεις θα είναι λίγο καλύτερο το EWMA των παρατηρήσεων. (Εννοείται ότι για τα διαγράμματα των αρχικών παρατηρήσεων, τα όρια ελέγχου έχουν προσαρμοστεί ώστε να λαμβάνεται υπόψη η αυτοσυσχέτιση).

Οι Wardell, Moskowitz και Plante (1992), χρησιμοποιώντας προσομοίωση συγκρίναν την επίδοση των τυπικών διαγραμμάτων Shewhart και EWMA στις αρχικές παρατηρήσεις με το διάγραμμα SCC των καταλοίπων. Για λόγους συγκρισιμότητας οι πολλαπλασιαστικοί παράγοντες των ορίων ελέγχου σε κάθε διάγραμμα (c, L και K) τροποποιήθηκαν έτσι ώστε το ARL, όταν δεν υπάρχει μετακίνηση στον μέσο, να είναι το ίδιο. Τότε το διάγραμμα με το μικρότερο ARL, όταν έχει συμβεί μια μετακίνηση του μέσου, θα θεωρείται ως καλύτερο. Για τα όρια ελέγχου των διαγραμμάτων των αρχικών παρατηρήσεων χρησιμοποιήθηκε η τυπική απόκλιση σ_x της διεργασίας που προκύπτει από το ARMA(1,1) υπόδειγμα. Τα αποτελέσματα αυτής της μελέτης έδειξαν ότι τα συμβατικά διαγράμματα ελέγχου δεν είναι εντελώς ανθεκτικά όταν οι παρατηρήσεις είναι συσχετισμένες. Το διάγραμμα EWMA είναι πολύ καλό στην αποκάλυψη μικρών μετατοπίσεων, και συμπεριφέρεται ικανοποιητικά για μεγάλες μετατοπίσεις, όταν η αυτοπαλίνδρομη παράμετρος είναι αρνητική και η παράμετρος του κινητού μέσου είναι θετική (τότε $\alpha_1 - \theta_1 < 0$ και καθορίζει σε μεγάλο βαθμό ότι η αυτοσυσχέτιση είναι αρνητική). Συνεπώς, στην περίπτωση αυτή, το διάγραμμα EWMA είναι μια καλή εναλλακτική λύση έναντι της πιο σύνθετης μεθόδου των Alwan και Roberts. Παρόλα αυτά, επειδή στην πράξη επιζητείται συνήθως η αποκάλυψη των μεγαλύτερων μετατοπίσεων, η προσέγγιση των Alwan και Roberts είναι πιο ελκυστική και αξίζει την επιπλέον προσπάθεια που πρέπει να καταβληθεί.

Οι Wardell, Moskowitz, και Plante (1994) έδειξαν ότι για διάφορους συνδυασμούς των παραμέτρων του ARMA(1,1) υποδείγματος, όταν έχει συμβεί κάποια μετακίνηση στη διεργασία, το ARL στο διάγραμμα Shewhart των καταλοίπων είναι μικρότερο όταν η διεργασία είναι αρνητικά αυτοσυσχετισμένη απ' ό,τι όταν είναι θετικά αυτοσυσχετισμένη. Αυτό συμβαίνει λόγω του ότι σε μια αρνητικά αυτοσυσχετισμένη διεργασία, όταν ο μέσος μετατοπίζεται, η κατά ένα βήμα μπροστά πρόβλεψη μετακινείται προς την αντίθετη κατεύθυνση της μετατόπισης. Έτσι το κατάλοιπο, δηλαδή η διαφορά μεταξύ της παρατήρησης και της πρόβλεψης, είναι πολύ μεγάλο και επομένως η μετατόπιση αποκαλύπτεται γρηγορότερα. Το διάγραμμα Shewhart των καταλοίπων έχει μικρή ευαισθησία στις μικρές μετατοπίσεις, όταν η διεργασία είναι θετικά αυτοσυσχετισμένη. Θεωρούν, λοιπόν, ότι η χρησιμοποίηση CUSUM ή EWMA διαγραμμάτων ελέγχου στα κατάλοιπα αντί για τα διαγράμματα ελέγχου Shewhart που δεν είναι ευαίσθητα σε μικρές μετακινήσεις της διεργασίας μπορεί να δώσουν καλύτερα αποτελέσματα, όπως στην περίπτωση των τυπικών διαγραμμάτων

ελέγχου του SPC. Οι Runger, Willemain και Prabhu (1995), προκειμένου να βελτιωθεί η ευαισθησία, προτείνουν τη χρήση του CUSUM διαγράμματος στην παρακολούθηση των καταλοίπων.

7.8. ΑΛΛΕΣ ΤΕΧΝΙΚΕΣ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΔΙΕΡΓΑΣΙΩΝ ΑΥΤΟΣΥΣΧΕΤΙΣΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Η ανάπτυξη ενός σαφούς υποδείγματος χρονολογικής σειράς απαιτεί εξειδικευμένες γνώσεις και είναι παρά την εξέλιξη των υπολογιστικών στατιστικών πακέτων συνεχίζει να είναι χρονοβόρα. Για αυτόν τον λόγο, αναφέρονται παρακάτω τεχνικές στατιστικού ελέγχου διεργασιών για αυτοσυσχετισμένα δεδομένα οι οποίες δεν απαιτούν την εκτίμηση υποδείγματος χρονολογικής σειράς.

Στην περίπτωση που η αυτοσυσχέτιση είναι μικρού βαθμού, μπορούν να εφαρμοστούν τα τυπικά διαγράμματα ελέγχου του SPC στις αρχικές παρατηρήσεις, αφού προηγουμένως γίνει προσαρμογή των ορίων ελέγχου ώστε να λαμβάνεται υπόψη η αυτοσυσχέτιση (Vassilopoulos και Stamboulis, 1978; Wardell και λοιποί, 1992, 1994; Zhang, 1998; VanBranckle και Reynolds, 1997; Lu και Reynolds, 2001).

Μία άλλη μέθοδος στατιστικού ελέγχου αυτοσυσχετισμένων δεδομένων, είναι το διάγραμμα ελέγχου UBM (Unweighted Batch Means). Η μέθοδος προτάθηκε από τους Runger και Willemain (1995) και βασίζεται στη διαπίστωση ότι οι μέσοι αρκετά μεγάλων συνεχόμενων και ισομεγέθων παρτίδων (batches) από τα δεδομένα, προσεγγιστικά είναι ανεξάρτητοι και ακολουθούν κανονική κατανομή για κάθε υποκείμενο υπόδειγμα χρονολογικής σειράς. Επομένως σ' αυτούς μπορούν να εφαρμοστούν τα τυπικά διαγράμματα ελέγχου του SPC. Βασικό μειονέκτημα της μεθόδου είναι, ότι για να επιτευχθεί η ανεξαρτησία και η κανονικότητα των μέσων, το μέγεθος των παρτίδων απαιτείται συχνά να είναι πολύ μεγάλο. Ο καθορισμός του μεγέθους κάθε παρτίδας είναι εμπειρικός και δεν εξαρτάται από την ταυτοποίηση και εκτίμηση ενός υποδείγματος χρονολογικής σειράς. Επιπροσθέτως, ο μέσος μιας παρτίδας χρησιμοποιεί πληροφορίες απ' όλες τις παρατηρήσεις στην παρτίδα.

Οι Montgomery και Mastrangelo (1991) πρότειναν μία προσεγγιστική μέθοδο που βασίζεται στο διάγραμμα EWMA. Εκμεταλλεύονται το γεγονός ότι το διάγραμμα EWMA μπορεί να εφαρμοσθεί σε κάποιες περιπτώσεις όπου τα δεδομένα είναι αυτοσυσχετισμένα, όπως όταν η διεργασία μπορεί να αποδοθεί με ένα υπόδειγμα «πρώτης τάξης ολοκληρωμένο κινητών μέσων» ($x_t = x_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$). Τότε αποδεικνύεται ότι ένα διάγραμμα EWMA, με $\lambda = 1 - \theta$, είναι η προτιμότερη επόμενη πρόβλεψη για αυτή τη διεργασία. Δηλαδή, αν $\hat{x}_{t+1}(t)$ είναι η πρόβλεψη, η οποία γίνεται στο τέλος της περιόδου t , για την παρατήρηση την περίοδο $t+1$, τότε $\hat{x}_{t+1}(t) = z_t$, όπου $z_t = \lambda x_t + (1 - \lambda)z_{t-1}$ είναι το διάγραμμα EWMA. Η ακολουθία των λαθών πρόβλεψης της επόμενης τιμής, είναι:

$$e_t = x_t - \hat{x}_t(t-1)$$

Τα λάθη πρόβλεψης του EWMA είναι ανεξάρτητα και ακολουθούν κανονική κατανομή με μέσο μηδέν. Έτσι, μπορούν να εφαρμοστούν διαγράμματα ελέγχου σε αυτά. Η παράμετρος λ (ή ισοδύναμα η θ) μπορεί να βρεθεί με την ελαχιστοποίηση των αθροισμάτων των τετραγώνων των λαθών e_t . Αν η διεργασία δεν ακολουθεί το υπόδειγμα «πρώτης τάξης ολοκληρωμένο κινητών μέσων», αλλά οι παρατηρήσεις από τη διεργασία είναι θετικά αυτοσυσχετισμένες και ο μέσος της διεργασίας δεν συμπαρασύρεται πάρα πολύ γρήγορα, το διάγραμμα EWMA με μια κατάλληλη τιμή για το λ , σύμφωνα με τους Montgomery και Mastrangelo (1991), θα δίνει μια εξαιρετική επόμενη πρόβλεψη. Συνεπώς, αναμένεται πολλές διεργασίες οι οποίες υπακούουν σε δυναμικές πρώτης τάξης (δηλαδή ακολουθούν μια αργή τάση) να αντιπροσωπεύονται καλά από το διάγραμμα EWMA. Συνεπώς, υπό αυτές τις συνθήκες, το διάγραμμα EWMA μπορεί προσεγγίσει το υπόδειγμα της χρονολογικής σειράς. Συνεπώς, το διάγραμμα ελέγχου για μεμονωμένες παρατηρήσεις που εφαρμόζεται στα λάθη πρόβλεψης του EWMA, θα διαφέρει ελαφρά αλλά όχι σημαντικά από το διάγραμμα ελέγχου των καταλοίπων του ακριβούς υποδείγματος της χρονολογικής σειράς. Ωστόσο, σύμφωνα με τους Tseng και Adams (1994) για τις διεργασίες που δεν προσεγγίζονται από το υπόδειγμα «πρώτης τάξης ολοκληρωμένο κινητών μέσων», η χρήση της τεχνικής που βασίζεται στο EWMA δεν θα εξηγεί ικανοποιητικά την αυτοσυσχέτιση, και αυτό μπορεί να επηρεάσει την στατιστική

παρουσίαση των διαγραμμάτων ελέγχου που βασίζονται στα λάθη πρόβλεψης του EWMA.

ΚΕΦΑΛΑΙΟ 8

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΤΟΞΙΚΟΤΗΤΑΣ ΑΝΑ ΤΑΜΙΕΥΤΗΡΑ / ΛΙΜΝΗ

8.1. ΕΙΣΑΓΩΓΗ

Τα ποιοτικά δεδομένα του επιφανειακού νερού ενδέχεται να παρουσιάζουν υψηλή μεταβλητότητα σε πολλές παραμέτρους. Επομένως, ένα μεγαλύτερο μέγεθος δειγμάτων, που βασίζεται σε αποτελέσματα πολλών ετών παρακολούθησης, αυξάνει την πιθανότητα ότι η κατανομή μιας ποιοτικής παραμέτρου του νερού αντιπροσωπεύεται κατάλληλα, ενώ παράλληλα μειώνει την επιρροή ανώμαλων παρατηρήσεων. Κατά τον υπολογισμό των περιγραφικών στατιστικών μέτρων ή τους στατιστικούς ελέγχους για διαφοροποιήσεις χρησιμοποιώντας δεδομένα από προγράμματα τυπικής παρακολούθησης των ποιοτικών παραμέτρων, συνιστάται να αθροίζονται τρία έως πέντε έτη στοιχείων (συγκεντρωτικά) (Ministry of the Environment, Ontario, 2006). Εντούτοις, σε αυτήν την περίπτωση, υποτίθεται ότι δεν έχει υπάρξει καμία σημαντική αλλαγή στην υπό εξέταση ποιοτική παράμετρο του νερού μέσα στα έτη που αθροίζονται. Γενικά, είναι αποδεκτό να συγκεντρώνονται τρία έως πέντε έτη δεδομένων. Στην περίπτωση που συναθροίζονται στοιχεία μεγαλύτερων χρονικών περιόδων, πρέπει να χρησιμοποιηθούν τεχνικές ανάλυσης τάσης (κάτι που πραγματοποιείται στην παρούσα έρευνα) για να επιβεβαιωθεί η απουσία μιας τάσης.

Οι δοκιμές τοξικότητας είναι απαραίτητες για την ολοκληρωμένη και έγκαιρη διάγνωση και πρόληψη της ρύπανσης του νερού και αυτό γιατί από μόνος του ο χημικός έλεγχος (Γενικό Χημείο Κράτους-Κύπρος, 2010, Βυλλιώτη, 2009, Καραδήμα, 2009):

- Καλύπτει συνήθως μόνο περίπου το 10-20% των τοξικολογικά σημαντικών ρυπαντών στο νερό (EPA, 1992), ενώ παράλληλα, ο αριθμός νέων χημικών ουσιών, που χρησιμοποιούνται και μπορούν να εισέλθουν στο νερό, αυξάνεται συνεχώς.

- Δεν μπορεί να προσδιορίσει τις επιπτώσεις που δημιουργούνται από την ταυτόχρονη έκθεση σε αριθμό τοξικών ουσιών και τις συνεργητικές και ανταγωνιστικές αντιδράσεις τους (Ribo και Rogers, 1990; Canna-Michaelidou και λοιποί, 2000; Kungolos, A., Samaras, P., Kiporoulou, A. M., Zoumboulis, A. and Sakellaropoulos, G. P.).
- Το όριο τοξικής επίδρασης ιδιαίτερα των καρκινογόνων και των γονοτοξικών ουσιών μπορεί να είναι μικρότερο του ορίου αναλυτικού προσδιορισμού.

Είναι επομένως απαραίτητο να εφαρμόζονται οικοτοξικολογικές μέθοδοι, δηλαδή έλεγχοι τοξικότητας με οργανισμούς, οι οποίες ερευνούν τις επιπτώσεις και τις αλληλεπιδράσεις των ρυπαντών παράλληλα με τις χημικές αναλύσεις (Canna-Michaelidou και λοιποί, 2000). Οι έλεγχοι τοξικότητας αποτελούν απαραίτητο συμπλήρωμα σε κάθε χημική ανάλυση και χρησιμοποιούνται ως ένα αναλυτικό εργαλείο που σηματοδοτεί, μετράει και προβλέπει τις περιβαλλοντικές επιπτώσεις (Καραδήμα, 2009). Στα πειράματα ελέγχου τοξικότητας εκτιμάται η επίδραση ενός δείγματος στην επιβίωση ή στο ρυθμό ανάπτυξης ζωντανών οργανισμών (βιοδείκτες) με σκοπό την αξιολόγηση της επικινδυνότητας του. Η επίδραση προσδιορίζεται συνήθως:

- Μέσω του ποσοστού θνησιμότητας ή αναστολής της ανάπτυξης / φυσιολογικής απόκρισης του πληθυσμού του ζωντανού οργανισμού
- Μέσω της συγκέντρωσης που προκαλεί 20% ή 50 % θνησιμότητα (LC₂₀ και LC₅₀ αντίστοιχα) ή αναστολή στην ανάπτυξη / φυσιολογική απόκριση στον πληθυσμό του ζωντανού οργανισμού (EC₂₀ ή IC₂₀ και EC₅₀ ή IC₅₀ αντίστοιχα).

Η οδηγία πλαίσιο για τα νερά 2000/60/EK αναφέρεται, εκτός των χημικών αναλύσεων, και στον προσδιορισμό της οικολογικής ποιότητας των επιφανειακών νερών με τη χρήση υδρόβιων οργανισμών καθώς και κατάλληλων βιοδεικτών. Ορισμένα ευρωπαϊκά κράτη, όπως η Ιταλία και η Ισπανία, περιλαμβάνουν στη νομοθεσία τους και οικοτοξικολογικές αναλύσεις (Κούγκολος, 2005). Στην Ελλάδα, μόνο η ΕΥΔΑΠ μετρούσε συστηματικά την τοξικότητα του νερού το χρονικό διάστημα 2000 - 2006. Ο έλεγχος της τοξικότητας είναι απαραίτητος όχι μόνο για το ανεπεξέργαστο νερό αλλά και για το επεξεργασμένο καθώς κατά την επεξεργασία ενδέχεται να προκύψουν ενδιάμεσα τοξικά προϊόντα. Οι οργανισμοί που

χρησιμοποιούνται για τον έλεγχο της τοξικότητας προέρχονται από διάφορα στάδια της τροφικής αλυσίδας και συνήθως περιλαμβάνουν βακτήρια, πρωτόζωα, τροχόζωα, βραγχιόποδα και ψάρια καθώς επίσης και μικροφύκη και ανώτερους φυτικούς οργανισμούς (Κούγκολος, 2005). Μερικά από τα πιο διαδεδομένα πειράματα ελέγχου τοξικότητας είναι (Κούγκολος, 2005; Newman και Unger, 2002):

- Πείραμα ακινησίας των βραγχιόποδων *Daphnia* (είδη *D. magna* και *D. pulex*) μετά από 24 ή 48 ώρες έκθεσης.
- Πείραμα αναστολής της βιοφωταύγειας των βακτηρίων *Vibrio fischeri* (του γένους NRRL-B-11177) συνήθως μετά από 5, 15 ή 30 λεπτά έκθεσης.
- Πείραμα αναστολής της ανάπτυξης των μικροφυκών *Pseudokirchneriella subcapitata* συνήθως μετά από 72 ώρες έκθεσης.

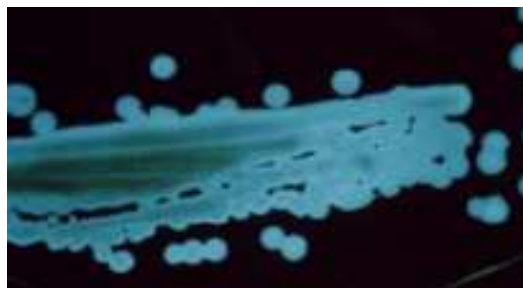
Στο Σχήμα 8.1 παρουσιάζονται οι εικόνες των παραπάνω οργανισμών.



Daphnia magna



Pseudokirchneriella subcapitata



Vibrio fischeri

Σχήμα 8.1. Οργανισμοί βιοδείκτες.

Τα πλεονεκτήματα των πειραμάτων με τα βακτήρια *Vibrio fischeri*, σε σχέση με τις εναλλακτικές επιλογές, είναι η απλότητα των διαδικασιών, η ταχύτητα και το χαμηλό κόστος (Ribo και λοιποί, 2001). Τα πειράματα με τα ψάρια έχουν το πλεονέκτημα ότι

χρησιμοποιούν οργανισμούς κατά πολύ πλησιέστερους προς τον ανθρώπινο (σπονδυλωτά). Βασικό όμως μειονέκτημά τους είναι ο μεγάλος χρόνος διάρκειας του πειράματος.

Τα τελευταία χρόνια υπάρχει αυξημένο ενδιαφέρον για την ανάλυση τάσεων των παραμέτρων ποιότητας του νερού. Στην επιστημονική βιβλιογραφία υπάρχουν αρκετές μελέτες που αφορούν τα επιφανειακά ύδατα (Murdoch και Shanley, 2006; Rogora και λοιποί, 2008; Qian και λοιποί, 2007; Bouraoui και λοιποί, 1999; Mitchell και λοιποί, 2001; Davies, 2006; Evans and Jenkins, 2000; Donohue και λοιποί, 2001; Peters και Donohue, 2001; Johnson και λοιποί, 2009; Qian και λοιποί, 2007; Sprague και Lorenz, 2009; Reynolds και λοιποί, 1997). Ωστόσο, το ενδιαφέρον περιορίζεται κυρίως στις φυσικο-χημικές παραμέτρους του νερού. Στον ελληνικό χώρο, οι μελέτες που αφορούν σε ανάλυση τάσεων αφορούν σε υδρολογικά χαρακτηριστικά (Koutsoyiannis, 2005) ενώ δεν έχουν διερευνηθεί τυχόν τάσεις ποιοτικών παραμέτρων των υδάτων μας.

Στο παρόν κεφάλαιο γίνεται μία αναλυτική περιγραφή των στατιστικών ιδιοτήτων των δεδομένων που αφορούν στα αποτελέσματα των δοκιμών τοξικότητας ανά ταμιευτήρα / λίμνη και συγχρόνως διερευνάται η ύπαρξη μονότονων τάσεων καθώς και η μεταβλητότητα της τοξικότητας ανά ταμιευτήρα / λίμνη που τροφοδοτούν τις ΜΕΝ της ΕΥΔΑΠ.

8.2. ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ ΤΟΞΙΚΟΤΗΤΑΣ

Οι μετρήσεις τοξικότητας πραγματοποιήθηκαν σε μηνιαία βάση, κατά τη διάρκεια μιας χρονικής περιόδου επτά ετών (2000 – 2006). Τα δεδομένα τοξικότητας από τους πύργους υδροληψίας των ταμιευτήρων Μαραθώνα και Μόρνου καθώς και τον διάυλο αντλιοστασίου Υλίκης καταχωρήθηκαν χρησιμοποιώντας ένα τεστ βιολογικής ακτινοβολίας, το οποίο βασίζεται στη συσχέτιση μεταξύ της τοξικότητας του δείγματος νερού και της επίδρασής της στην ένταση του εκπεμπόμενου φωτός των θαλάσσιων βακτηρίων *Vibrio fischeri* (παλαιότερα ονομαζόμενα *Photobacterium phosphoreum*), μετρούμενη με το bioluminometer Microtox (Strategic Diagnostics Microtox 1010, 2003). Οι τοξικές ουσίες προκαλούν μια διαταραχή στον κανονικό

μεταβολισμό των βακτηριδίων με αποτέλεσμα τη μείωση του εκπεμπόμενου φωτός. Ο βαθμός τοξικότητας είναι ανάλογος προς τη μετρούμενη απώλεια φωτός (ISO 11348-3, 1998). Τα δείγματα ρυθμίστηκαν ώστε να περιέχουν 2% NaCl, το οποίο παρέχει ωσμωτική προστασία στους θαλάσσιους μικροοργανισμούς. Γενικά, θεωρείται ότι υπάρχει καλή συσχέτιση μεταξύ των προσδιορισμών της τοξικότητας με χρήση των *Vibrio fischeri* και αυτής με χρήση ανώτερων οργανισμών όπως η *Daphnia* και η πολύχρωμη πέστροφα (rainbow trout) (EVS Consultants, 1989; Ribo και Kaiser, 1983). Τα βακτήρια *Vibrio fischeri* έχουν χρησιμοποιηθεί σε αρκετές δοκιμές τοξικότητας επιφανειακών νερών (π.χ. Simeonov και λοιποί, 2007; Canna-Michaelidou και λοιποί 2000; Kaiser, Lum και Palabrica, 1988; Kaiser, Ribo και Kwasniewska, 1988; Ribo και λοιποί, 1985).

Ως έκφραση της τοξικότητας λαμβάνεται η ανασταλτική επίδραση ενός δείγματος δοκιμής μετά από το χρόνο επαφής 30 δευτερολέπτων (H_{30}). Αυτή η έκφραση δίνεται από τον κάτωθι τύπο (ISO 11348-3, 1998):

$$H_t = [(I_{ct} - I_{Tt}) / I_{ct}] \times 100,$$

όπου, H_t είναι η επί της εκατό ανασταλτική επίδραση ενός δείγματος μετά από το χρόνο επαφής 30 δευτερολέπτων, I_{Tt} είναι η ένταση της φωταύγειας του δείγματος μετά από χρόνο επαφής 30 δευτερολέπτων σε σχετικές μονάδες φωταύγειας και I_{ct} είναι η ένταση της φωταύγειας του εναιωρήματος ελέγχου (μάρτυρα) σε σχετικές μονάδες φωταύγειας. Για την τοξικότητα των επιφανειακών νερών δεν έχει οριστεί κάποιο ανώτερο όριο από τη νομοθεσία. Ωστόσο, εάν η αναστολή της φωταύγειας είναι μεγαλύτερη από 20% το δείγμα θεωρείται τοξικό (Joret και λοιποί, 1986; Léni και λοιποί, 1989; Niemirycz, 2007).

Όλες οι δοκιμές πραγματοποιήθηκαν βάσει των τυποποιημένων πρωτόκολλων που περιγράφονται στο εγχειρίδιο οξείας τοξικότητας της Microtox (Microtox Acute Toxicity Users' Guide).

8.3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

Αρχικά, τα δεδομένα τοξικότητας κάθε ταμειυτήρα / λίμνης ελέγχθηκαν ως προς τα βασικά στατιστικά χαρακτηριστικά τους προκειμένου να αποφασιστεί η

χρησιμοποίηση παραμετρικής ή μη παραμετρικής ανάλυσης τάσεων. Γενικά, στους στατιστικούς που ασχολούνται με τις μεταβλητές ποιότητας του νερού, υπάρχει έλλειψη εμπιστοσύνης όσον αφορά στην αποτελεσματικότητα των μη παραμετρικών μεθόδων στην περίπτωση ασήμαντων αποκλίσεων από την κανονικότητα και/ή δειγμάτων μικρού μεγέθους. Συνεπώς, είναι ιδιαίτερα σημαντική η μελέτη περιπτώσεων στις οποίες η απόκλιση από την κανονικότητα είναι επαρκώς μικρή ώστε η οπτική επιθεώρηση της κατανομής των δεδομένων ή οι τυπικοί έλεγχοι κανονικότητας είναι απίθανο να εξασφαλίσουν αποδείξεις για έλλειψη κανονικότητας. Οι Hirsch και λοιποί (1991), ακολουθώντας ανάλυση Monte Carlo, σύγκριναν την απόδοση παραμετρικών και μη-παραμετρικών μεθόδων ανάλυσης τάσεων στην περίπτωση μικρών αποκλίσεων από την κανονικότητα και/ή δειγμάτων μικρού μεγέθους. Τα αποτελέσματα της ανάλυσης έδειξαν εμφανώς ότι, για σετ δεδομένων που αποκλίνουν ελαφρώς από την κανονικότητα (ακόμα και ανεπαίσθητα), οι μη παραμετρικές μέθοδοι παρουσιάζουν μέτρια πλεονεκτήματα όσον αφορά στην αποτελεσματικότητα και την ισχύ συγκριτικά με τις παραμετρικές μεθόδους, ενώ παρουσιάζουν μόνο μικρά μειονεκτήματα στην περίπτωση όπου όντως υπάρχει κανονικότητα. Οι Reckhow και λοιποί (1993) πρότειναν να χρησιμοποιούνται συστηματικά μη-παραμετρικές μέθοδοι για τον εντοπισμό τάσεων ενώ οι παραμετρικές μέθοδοι να χρησιμοποιούνται μόνο όταν η χρήση τους δικαιολογείται πλήρως. Η υπόδειξη αυτή βασίζεται στις επιπλοκές των επιδράσεων της μη-κανονικότητας, των περιστασιακών ακραίων τιμών στα δεδομένα ποιότητας του νερού και στην αντίληψη ότι οι μη-παραμετρικές μέθοδοι αναπτύσσονται σε «αναγνωρισμένη πρακτική» στις μελέτες για τον προσδιορισμό τάσεων στις παραμέτρους ποιότητας του νερού.

Επιπλέον, τα δεδομένα ελέγχθηκαν ως προς την ύπαρξη εποχικότητας και αυτοσυσχέτισης προκειμένου να γίνουν οι κατάλληλες διορθωτικές ενέργειες κατά την εφαρμογή των μεθόδων ανάλυσης τάσης ή την αξιολόγηση των αποτελεσμάτων. Μελέτες προσομοίωσης δείχνουν ότι όταν τα δεδομένα δεν παρουσιάζουν εποχικότητα, ο εποχικός έλεγχος Kendall δεν είναι τόσο ισχυρός όσο ο έλεγχος τάσης Mann-Kendall (McLeod και λοιποί, 1991; Hipel και McLeod, 1994). Συνεπώς, πριν την εφαρμογή ενός ελέγχου εποχικής τάσης όπως ο εποχικός έλεγχος Kendall προτείνεται να ελέγχεται η ύπαρξη ή μη εποχικότητας. Η διερεύνηση για τυχόν ύπαρξη εποχικότητας έγινε μέσω του μη-παραμετρικού στατιστικού ελέγχου

Kruskall-Wallis (Conover, 1980; Yu και λοιποί, 1993) και των διαγραμμάτων δειγματικής αυτοσυσχέτισης (Reckhow και λοιποί, 1993; Yürekli, 2005). Για τον έλεγχο Kruskal-Wallis, τα δεδομένα ταξινομήθηκαν από την μικρότερη τιμή προς τη μεγαλύτερη και οι υπολογισμοί έγιναν στις τάξεις μεγέθους των δεδομένων. Αυτός ο έλεγχος εξετάζει κατά πόσο υπάρχουν διαφορές στις μέσες τάξεις μεγέθους, παρά στις μέσες τιμές, μεταξύ των διαφορετικών ομάδων. Το στατιστικό που υπολογίζεται με τον έλεγχο Kruskal-Wallis test καλείται H. Στις περιπτώσεις που υπήρχαν ταυτισμένες τάξεις μεγέθους, το στατιστικό H και η p-τιμή που το συνοδεύει διορθώθηκαν κατάλληλα για τις ταυτίσεις (Wheater και Cook, 2000).

Σύμφωνα με τον von Storch (1995), η ύπαρξη θετικής αυτοσυσχέτισης στα δεδομένα αυξάνει την πιθανότητα ότι ο έλεγχος Mann-Kendall εντοπίζει την σημαντικότητα μιας τάσης. Αυτό μπορεί να οδηγήσει στην απόρριψη της μηδενικής υπόθεσης μη ύπαρξης τάσης, ενώ η μηδενική υπόθεση είναι στην πραγματικότητα αληθής. Προκειμένου να εξαληφθεί η επίδραση της αυτοσυσχέτισης στον έλεγχο Mann-Kendall, προτάθηκε από τους Kulkarni και von Storch (1995) καθώς και από τον von Storch (1995) η προ-λεύκανση (pre-whitening) της χρονολογικής σειράς πριν την εφαρμογή του ελέγχου Mann-Kendall. Σύμφωνα με τη μέθοδο αυτή, μία χρονολογική σειρά δεδομένων X_i με μη μηδενικό συντελεστή αυτοσυσχέτισης ρ_1 μπορεί να μοντελοποιηθεί με ένα αυτοπαλίνδρομο μοντέλο πρώτης τάξης (Box και λοιποί, 1994) και στη συνέχεια να εφαρμοστεί ο έλεγχος Mann-Kendall στην προ-λευκασμένη χρονολογική σειρά που δίνεται από την εξίσωση:

$$X'_i = X_i - \rho_1 X_{i-1}$$

Η μέθοδος αυτή έχει εφαρμοστεί σε διάφορες μελέτες προσδιορισμού τάσεων, όπως για παράδειγμα στις μελέτες των Douglas και λοιποί (2000), Zhang και λοιποί (2000, 2001a,b), Burn και Hag Elnur (2002). Παρόλα αυτά, ενώ στην περίπτωση που δεν υπάρχει πραγματική τάση στα δεδομένα της χρονολογικής σειράς η μέθοδος της προ-λεύκανσης είναι ικανή να μειώσει την επίδραση της σειριακής συσχέτισης στον έλεγχο Mann-Kendall (von Storch, 1995), στην περίπτωση που όντως υπάρχει τάση μπορεί να οδηγήσει στη μείωση της ικανότητας προσδιορισμού της σημαντικής τάσης με τον έλεγχο Mann-Kendall (Douglas και λοιποί, 2000; Yue και λοιποί, 2002; Yue και λοιποί, 2003). Συγκεκριμένα, στην περίπτωση ύπαρξης μίας πραγματικής θετικής τάσης η μέθοδος της προ-λεύκανσης θα αφαιρέσει ένα μέρος της τάσης με

αποτέλεσμα τη μείωση της πιθανότητας απόρριψης της μηδενικής υπόθεσης ενώ θα μπορούσε να ήταν ψευδής, ενώ στην περίπτωση ύπαρξης μίας πραγματικής αρνητικής τάσης θα διογκώσει την υπάρχουσα τάση και θα οδηγήσει σε αύξηση της πιθανότητας απόρριψης της μηδενικής υπόθεσης ενώ θα μπορούσε να ήταν αληθής. Προκειμένου να αρθούν τα κωλύματα της προ-λεύκανσης, οι Yue και λοιποί (2002) πρότειναν μία τροποποίηση της μεθόδου, η οποία ονομάζεται «προ-λεύκανση ελεύθερη τάσης» (Trend-free pre-whitening, TFPW). Σύμφωνα με την μέθοδο αυτή, αρχικά εκτιμάται η κλίση της τάσης (b). Στη συνέχεια, αφαιρείται η τάση T_t από τις τιμές του αρχικού δείγματος των δεδομένων ως εξής:

$$Y_t = X_t - T_t = X_t - bt$$

Εφόσον ο συντελεστής δειγματικής αυτοσυσχέτισης (r_1) της νέας χρονοσειράς που προκύπτει είναι στατιστικά σημαντικά διάφορος του μηδενός, εφαρμόζεται η μέθοδος της προ-λεύκανσης στη νέα αυτή χρονοσειρά:

$$Y_t' = Y_t - r_1 Y_{t-1}$$

Η χρονοσειρά των καταλοίπων Y_t' θεωρείται πλέον ως ανεξάρτητη. Στη συνέχεια, η τάση T_t και τα κατάλοιπα Y_t' συνδυάζονται ως:

$$Y_t'' = Y_t' + T_t$$

Τέλος, για την αποτίμηση της σημαντικότητας της τάσης, εφαρμόζεται η μέθοδος Mann-Kendall στην σειρά Y_t'' , η οποία περιέχει την τάση και έναν θόρυβο ο οποίος δεν επηρεάζεται πλέον από σειριακή συσχέτιση.

Για τον έλεγχο διαφορών στις κεντρικές τιμές των δεδομένων μεταξύ των τριών ταμειυτήρων / λιμνών, χρησιμοποιήθηκε ο μη-παραμετρικός έλεγχος Kruskal-Wallis ή/και ο έλεγχος διαμέσων (Median test) (Conover, 1980; Pett, 1997). Ο έλεγχος διαμέσων, είναι επίσης μη-παραμετρικός, με λιγότερες προϋποθέσεις έναντι του ελέγχου Kruskal-Wallis (δεν περικλείει την υπόθεση της ισότητας των διακυμάνσεων των διαφορετικών ομάδων) και δεν λαμβάνει υπόψη την απόσταση από τη γενική διάμεσο, παρά μόνο σε ποια πλευρά της διαμέσου βρίσκονται οι παρατηρήσεις. Η μηδενική υπόθεση του ελέγχου είναι ότι όλες οι διάμεσοι των ομάδων είναι ίσες μεταξύ τους (ίσες με τη γενική διάμεσο του συνόλου των παρατηρήσεων όλων των ομάδων) ενώ η εναλλακτική ότι μία τουλάχιστον διάμεσος είναι διαφορετική. Για την εκτέλεση του ελέγχου των διαμέσων, οι παρατηρήσεις

μέσα σε κάθε ομάδα ταξινομήθηκαν ανάλογα με το αν βρίσκονταν πάνω ή κάτω από τη γενική διάμεσο. Οι αντίστοιχες συχνότητες τοποθετήθηκαν στα κελιά ενός πίνακα συνάφειας $2 \times c$ (όπου c ο αριθμός των διαφορετικών ομάδων) και χρησιμοποιήθηκε ο έλεγχος χ^2 για την αποτίμηση της σημαντικότητας των διαφορών μεταξύ των διαμέσων. Ο έλεγχος των διαμέσων θεωρείται λιγότερο ισχυρός από τον έλεγχο Kruskal-Wallis με την προϋπόθεση ότι τηρούνται όλες οι προϋποθέσεις του ελέγχου Kruskal-Wallis.

8.4. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ

8.4.1. Προκαταρκτική στατιστική διερεύνηση των δεδομένων

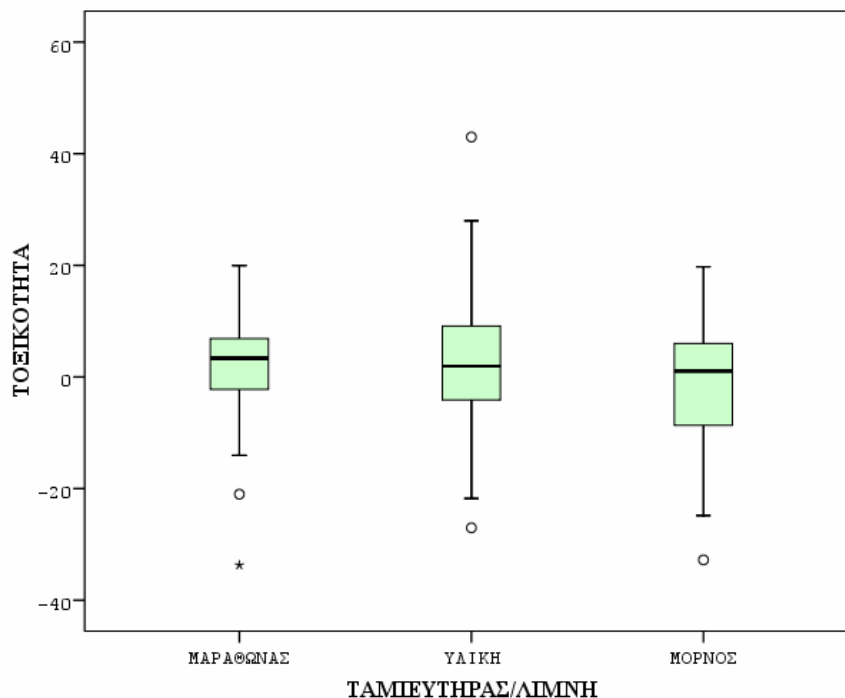
Στον Πίνακα 8.1 που ακολουθεί παρουσιάζονται περιληπτικά περιγραφικά στατιστικά μέτρα των δεδομένων τοξικότητας ανά ταμιευτήρα / λίμνη. Παρόλο που δεν υπάρχουν από την νομοθεσία άνω όρια για την αναστολή της φωταύγειας του επιφανειακού νερού που προορίζεται για ανθρώπινη κατανάλωση μετά από επεξεργασία, παρατηρείται ότι η μέγιστη τιμή για τους ταμιευτήρες Μαραθώνα και Υλίκης είναι μικρότερες από το όριο του 20%.

Πίνακας 8.1. Περιγραφικά στατιστικά μέτρα των δεδομένων τοξικότητας ανά ταμιευτήρα / λίμνη

	Μαραθώνας	Υλίκη	Μόρνος
Αριθμός δειγμάτων	84	84	84
Ελάχιστη τιμή	-33,7	-27,03	-32,78
Μέγιστη τιμή	19,95	43	19,72
Εύρος	53,65	70,03	52,5
Διάμεσος	3,355	1,935	1,05
Αριθμητικός μέσος	2,545	2,428	-0,349
Τυπική απόκλιση	8,554	11,342	10,509
Ασυμμετρία	-1,004	0,3	-0,224
Τυπικό σφάλμα ασυμμετρίας	0,263	0,263	0,263
Κύρτωση	3,447	1,537	0,151
Τυπικό σφάλμα κύρτωσης	0,52	0,52	0,52
Shapiro-Wilk στατιστικό	0,937	0,973	0,98
Shapiro-Wilk p-Value	<0,001	0,074	0,224

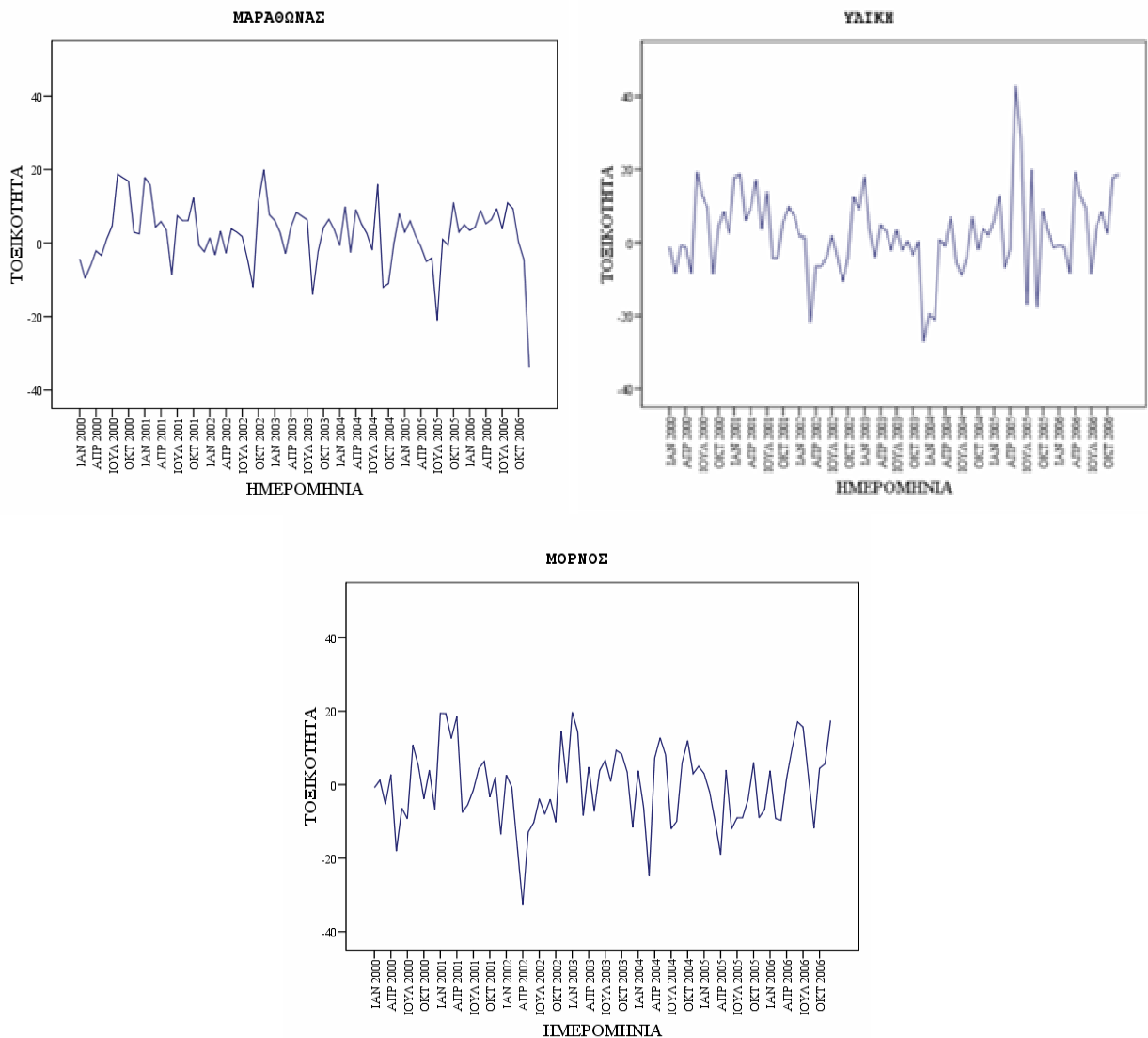
Η ύπαρξη αρνητικών τιμών στα δεδομένα μπορεί να αποδοθεί στο φαινόμενο της όρμεσης (Hormesis) (Charman, 2002), σύμφωνα με το οποίο όχι μόνο δεν αναστέλλεται η φωταύγεια αλλά ευνοείται η εκπομπή φωτός από τα βακτήρια. Όταν τα δείγματα περιέχουν ρύπους σε πολύ χαμηλές συγκεντρώσεις, προκαλούν διέγερση των φωτοβακτηριδίων με αποτέλεσμα την αύξηση της φωταύγειας σε σχέση με αυτήν του δείγματος μάρτυρα (Biggs, 1994). Συνεπώς το νερό των ταμιευτήρων μπορεί να θεωρηθεί ως χαμηλής τοξικότητας.

Ο έλεγχος κανονικότητας Shapiro-Wilk (Πίνακας 8.1) απέρριψε την υπόθεση της κανονικότητας μόνο των δεδομένων τοξικότητας που προέρχονται από τον ταμιευτήρα Μαραθώνα (p -τιμή $<0,001$). Παρόλα αυτά, συνεκτιμώντας την εικόνα των θηκογραμμάτων (box-plots) ανά ταμιευτήρα / λίμνη (Σχήμα 8.2), λόγω της ύπαρξης ακραίων τιμών κρίθηκε προτιμότερη η επιλογή μη παραμετρικών τεχνικών για την ανάλυση τάσεων (Wang, 2006; Van Gelder και λοιποί, 2008). Οι μη-παραμετρικές μέθοδοι για τον προσδιορισμό τάσεων είναι λιγότερο ευαίσθητες στις ακραίες τιμές σε σύγκριση με τις παραμετρικές μεθόδους. Επιπροσθέτως, ένας μη-παραμετρικός έλεγχος, βασισμένος στις τάξεις μεγέθους, μπορεί να ελέγξει για την ύπαρξη τάσης σε μία χρονοσειρά χωρίς να χρειάζεται να καθοριστεί κατά πόσο η τάση είναι γραμμική ή μη-γραμμική (Wang και λοιποί, 2005; Van Gelder και λοιποί, 2008).



Σχήμα 8.2. Θηκογράμματα για την τοξικότητα ανά ταμιευτήρα / λίμνη.

Τα χρονοδιαγράμματα της τοξικότητας ανά ταμιευτήρα / λίμνη (Σχήμα 8.3) δεν παρουσίασαν ισχυρές ενδείξεις ύπαρξης εποχικότητας στα δεδομένα καθώς δεν υπάρχουν εμφανή κυκλικά πρότυπα. Ωστόσο, η ύπαρξη ή όχι εποχικότητας ελέγχθηκε και μέσω του μη παραμετρικού στατιστικού ελέγχου - Kruskal-Wallis.



Σχήμα 8.3. Διαγράμματα χρονολογικών σειρών τοξικότητας ανά ταμιευτήρα/λίμνη.

Ο έλεγχος Kruskal-Wallis προϋποθέτει την ομοιογένεια των διακυμάνσεων των διαφορετικών ομάδων που συγκρίνονται ως προς την κεντρική θέση των τιμών τους. Η ομοιογένεια των διακυμάνσεων των ομάδων (μηνών) διαπιστώθηκε με τον στατιστικό έλεγχο του Levene (1960). Ο έλεγχος του Levene (Levene's test) είναι μια επαγωγική στατιστική μέθοδος που χρησιμοποιείται για να αξιολογήσει την ισότητα

των διακυμάνσεων των πληθυσμών από τους οποίους προέρχονται τα διαφορετικά δείγματα. Ο έλεγχος του Levene εξετάζει την μηδενική υπόθεση ότι οι διακυμάνσεις των πληθυσμών είναι ίσες. Ένα πλεονέκτημα του ελέγχου Levene είναι ότι δεν είναι ευαίσθητος στην έλλειψη κανονικότητας των δεδομένων (Shoukri και Pause, 1999). Ως επίπεδο σημαντικότητας για τον έλεγχο θεωρήθηκε το $\alpha = 0,05$. Κατά συνέπεια, η μηδενική υπόθεση των ίσων διακυμάνσεων απορρίπτεται εάν p -τιμή $< \alpha$ και τότε συνάγεται το συμπέρασμα ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των διακυμάνσεων των πληθυσμών. Στον Πίνακα 8.2 παρουσιάζονται τα αποτελέσματα των ελέγχων Levene. Είναι εμφανές ότι η υπόθεση της ισότητας των διακυμάνσεων δεν απορρίπτεται σε κανέναν από τους ταμιευτήρες/λίμνες.

Πίνακας 8.2. Έλεγχος ομοιογένειας των διακυμάνσεων (Levene's test) ανά ταμιευτήρα / λίμνη

Ταμιευτήρας / λίμνη	Levene Statistic	df1	df2	p-τιμή
ΜΑΡΑΘΩΝΑΣ	1,021	11	72	0,437
ΥΛΙΚΗ	0,874	11	72	0,569
ΜΟΡΝΟΣ	1,155	11	72	0,334

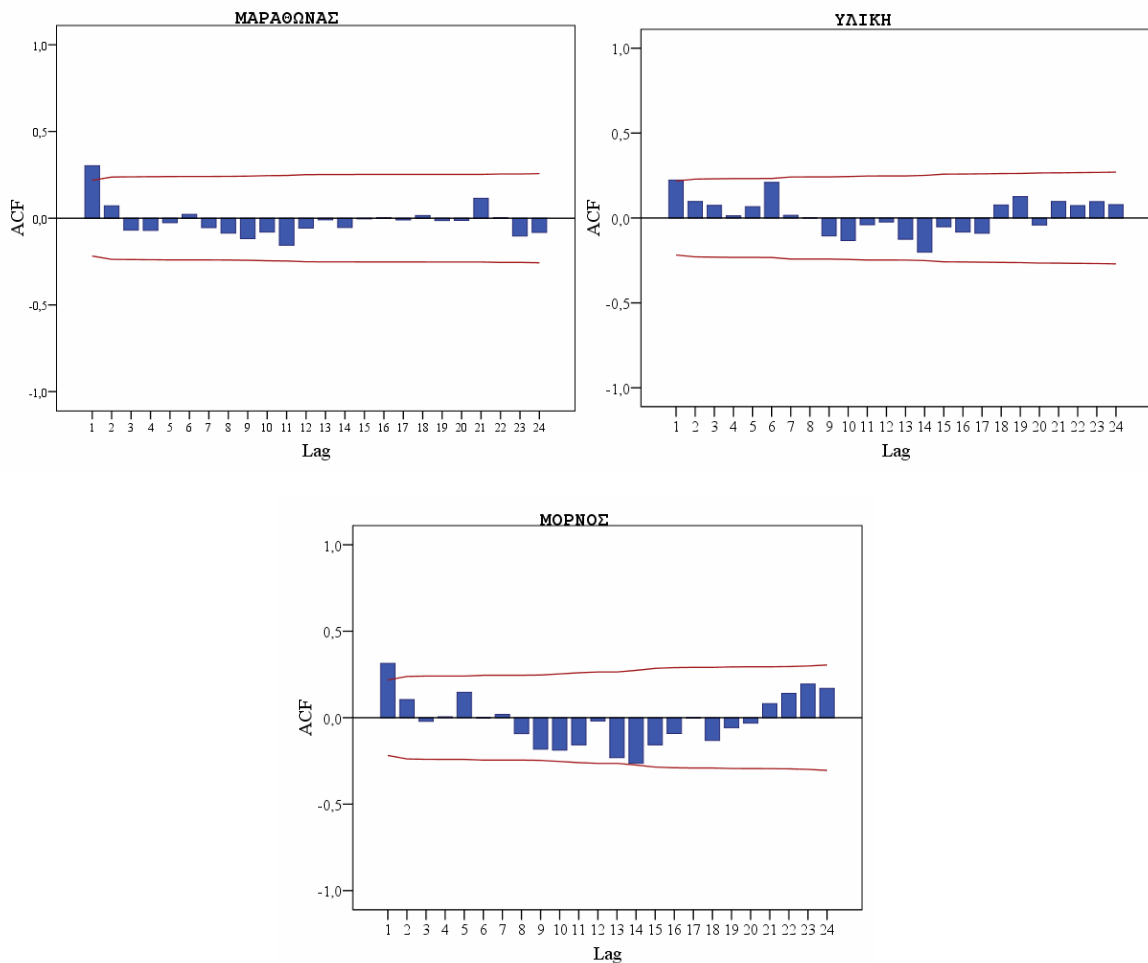
Στον Πίνακα 8.3 παρουσιάζονται τα αποτελέσματα των ελέγχων Kruskal-Wallis για τον έλεγχο της εποχικότητας σε κάθε ταμιευτήρα / λίμνη. Βάσει των p -τιμών των ελέγχων, σε επίπεδο σημαντικότητας $\alpha=0,05$, δεν απορρίπτεται η μηδενική υπόθεση της ισότητας των διαμέσων των τιμών της τοξικότητας ανά μήνα σε κάθε ταμιευτήρα / λίμνη.

Πίνακας 8.3. Περίληπτικά στατιστικά των ελέγχων Kruskal-Wallis για κάθε ταμιευτήρα / λίμνη (μεταβλητή ομαδοποίησης: Μήνας)

	ΜΑΡΑΘΩΝΑΣ	ΥΛΙΚΗ	ΜΟΡΝΟΣ
Chi-Square	3,970	10,741	10,573
df	11	11	11
p-τιμή	0,971	0,465	0,480

Στο ίδιο συμπέρασμα, όσον αφορά στην εποχικότητα, κατέληξε και η εξέταση των διαγραμμάτων των συναρτήσεων δειγματικής αυτοσυσχέτισης των δεδομένων ανά ταμιευτήρα / λίμνη (Σχήμα 8.4). Η συνάρτηση δειγματικής αυτοσυσχέτισης (ACF) παρέχει σημαντικές πληροφορίες για την συσχέτιση μεταξύ των ζευγών των παρατηρήσεων που απέχουν κατά k μονάδες χρόνου (υστερήσεις – lags).

Γενικά, η ύπαρξη θετικής αυτοσυσχέτισης στην δωδέκατη υστέρηση μηνιαίων δεδομένων ποιότητας του νερού, δηλώνει μία κυκλική συμπεριφορά η οποία επαναλαμβάνεται κάθε δώδεκα μήνες (εποχικότητα) (Reckhow και λοιποί, 1993). Επιπλέον, η ύπαρξη αρνητικής αυτοσυσχέτισης στην έκτη υστέρηση δηλώνει επίσης έναν δωδεκάμηνο (εποχικό) κύκλο, καθώς υποδεικνύει μία αντίθετη ανταπόκριση των δεδομένων που απέχουν χρονικά κατά έξι μήνες.



Σχήμα 8.4. Διαγράμματα συναρτήσεων δειγματικής αυτοσυσχέτισης για την τοξικότητα ανά ταμιευτήρα/λίμνη.

Οι συναρτήσεις δειγματικής αυτοσυσχέτισης των δεδομένων ανά ταμιευτήρα / λίμνη (Σχήμα 8.4) δεν παρουσίασαν στατιστικά σημαντικές τιμές (εκτός των ορίων πιθανότητας 95%) (Lehmann and Rode, 2001) στις υστερήσεις οι οποίες είναι πολλαπλάσιες του 12 και επιπλέον δεν παρουσίασαν κυματοειδή μορφή. Επομένως, βάσει του διαγράμματος συναρτήσεων δειγματικής αυτοσυσχέτισης, για τις μηνιαίες

τιμές των δεδομένων τοξικότητας θεωρείται ότι υπάρχει έλλειψη εποχικής περιοδικότητας (Reckhow και λοιποί, 1993; Yürekli, 2005).

Η έλλειψη εποχικότητας στις τιμές της τοξικότητας μπορεί να αποδοθεί στις πιθανόν μικρές τιμές των παραμέτρων που μπορεί να την επηρεάζουν, καθώς επίσης και στην πιθανή μικρή διακύμανση τους εντός του ετήσιου κύκλου.

Επιπροσθέτως όμως, τα διαγράμματα συναρτήσεων δειγματικής αυτοσυσχέτισης για την τοξικότητα ανά ταμιευτήρα/λίμνη υπέδειξαν την ύπαρξη σχετικά μικρής θετικής αυτοσυσχέτισης των δεδομένων καθώς ο συντελεστής αυτοσυσχέτισης για την πρώτη υστέρηση υπερέβαινε το άνω 95% όριο πιθανότητας. Επομένως, κάθε παρατήρηση συσχετίζεται με την προηγούμενή της. Αυτό σημαίνει ότι ένα μέρος της πληροφορίας που μεταφέρει μια τρέχουσα παρατήρηση έχει ήδη μεταφερθεί από την προηγούμενη παρατήρηση.

Παρόλα αυτά, σε αυτό το στάδιο της ανάλυσης, τα διαγράμματα αυτοσυσχέτισης απλώς βοηθούν στην διάγνωση της εποχικότητας. Η ύπαρξη αυτοσυσχέτισης στα δεδομένα τυγχάνει ενδιαφέροντος μόνο μετά την απαλοιφή όλων των ντετερμινιστικών προτύπων (Reckhow και λοιποί, 1993).

Λόγω της έστω και μικρού μεγέθους ύπαρξης αυτοσυσχέτισης στα δεδομένα, οι μη παραμετρικοί έλεγχοι Kruskal-Wallis για την διερεύνηση της εποχικότητας, επαναλήφθηκαν για κάθε ταμιευτήρα / λίμνη στις προ-λευκασμένες χρονοσειρές $X'_t = X_t - r_1 X_{t-1}$, ώστε να τα δεδομένα κάθε μήνα να θεωρούνται πλήρως ανεξάρτητα μεταξύ τους. Οι τιμές της δειγματικής συνάρτησης αυτοσυσχέτισης r_1 υπολογίστηκαν ως 0,303 για τον Μαραθώνα, 0,223 για την Υλίκη και 0,314 για τον Μόρνο. Στον Πίνακα 8.4 παρουσιάζονται τα αποτελέσματα του ελέγχου Levene ομοιογένειας των διακυμάνσεων για τις προ-λευκασμένες χρονοσειρές κάθε ταμιευτήρα / λίμνης. Τα αποτελέσματα των ελέγχων Kruskal-Wallis παρουσιάζονται στον Πίνακα 8.5 και επιβεβαιώνουν ότι η υπόθεση της ισότητας των διαμέσων κάθε μήνα δεν απορρίπτεται σε επίπεδο σημαντικότητας $\alpha = 0,05$ για κανένα από τους ταμιευτήρες / λίμνες.

Πίνακας 8.4. Έλεγχος ομοιογένειας των διακυμάνσεων (Levene's test) των προ-
λευκασμένων χρονοσειρών κάθε ταμιευτήρα / λίμνης

Ταμιευτήρας / λίμνη	Levene Statistic	df1	df2	p-τιμή
ΜΑΡΑΘΩΝΑΣ	1,030	11	72	0,430
ΥΛΙΚΗ	0,628	11	72	0,799
ΜΟΡΝΟΣ	1,071	11	72	0,396

Πίνακας 8.5. Περιληπτικά στατιστικά των ελέγχων Kruskal-Wallis για τις
προλευκασμένες χρονοσειρές κάθε ταμιευτήρα / λίμνης

	ΜΑΡΑΘΩΝΑΣ	ΥΛΙΚΗ	ΜΟΡΝΟΣ
Chi-Square	6,114	10,496	13,781
df	11	11	11
p-τιμή	0,866	0,486	0,245

(μεταβλητή ομαδοποίησης: Μήνας)

8.4.2. Ανάλυση τάσεων

Η προκαταρκτική στατιστική ανάλυση των δεδομένων τοξικότητας ανά ταμιευτήρα / λίμνη υπέδειξε ως καταλληλότερη μέθοδο για τον έλεγχο τάσεων τον μη-παραμετρικό έλεγχο Mann-Kendall με την επιφύλαξη ότι παρατηρήθηκε μικρή θετική αυτοσυσχέτιση στα δεδομένα. Καθώς υπάρχει ενδιαφέρον στο εάν οι αλλαγές των τιμών της τοξικότητας στον χρόνο είναι αποτέλεσμα σειριακής συσχέτισης ή μίας ντετερμινιστικής τάσης, ακολουθήθηκε η πρόταση των Ward και λοιπών (2003) οι οποίοι προτείνουν τη χρήση του ελέγχου Mann-Kendall ως αρχικό διαγνωστικό μέσο ύπαρξης τάσης στα δεδομένα. Στην περίπτωση απόρριψης της μηδενικής υπόθεσης προτείνουν την περαιτέρω εξέταση της δομής της σειριακής συσχέτισης ως πιθανής αιτίας της φαινομενικής τάσης.

Πίνακας 8.6. Περιληπτικά στατιστικά για τους ελέγχους Mann-Kendall και τις εκτιμήσεις κλίσης του Sen για την τοξικότητα ανά ταμιευτήρα/λίμνη

	Μαραθώνας	Υλίκη	Μόρνος
Στατιστικό (S)	-6	57	87
Τυπική απόκλιση του S	258,856	258,831	258,849
Z	-0,019	0,216	0,332
p-τιμή*	0,985	0,829	0,74
Εκτίμηση κλίσης Sen	-0,001	0,008	0,015
95% κάτω όριο ΔΕ κλίσης	-0,079	-0,092	-0,072
95% άνω όριο ΔΕ κλίσης	0,06	0,108	0,11

*H₀: Δεν υπάρχει τάση, έναντι H₁: Υπάρχει τάση

Στον Πίνακα 8.6 παρουσιάζονται τα αποτελέσματα των ελέγχων Mann-Kendall καθώς και οι εκτιμήσεις κλίσης του Sen για την τοξικότητα ανά ταμιευτήρα/λίμνη. Οι εκτιμήσεις κλίσης του Sen και τα στατιστικά S των ελέγχων Mann-Kendall υπέδειξαν ενδεχόμενη πτωτική τάση για τον Μαραθώνα και αυξητική τάση για την Υλίκη και τον Μόρνο. Ωστόσο, σε κανέναν ταμιευτήρα / λίμνη δεν απορρίφθηκε η μηδενική υπόθεση ότι δεν υπάρχει τάση σε κανένα από τα συνηθισμένα επίπεδα σημαντικότητας (p-τιμή: 0,985 για τον Μαραθώνα, 0,829 για την Υλίκη και 0,74 για τον Μόρνο). Συνεπώς, συμπεραίνεται ότι δεν υπάρχουν αρκετές ενδείξεις για την απόρριψη της μηδενικής υπόθεσης και δεν χρειάζεται περαιτέρω διερεύνηση για τις επιπτώσεις της ύπαρξης αυτοσυσχέτισης στα δεδομένα κάθε ταμιευτήρα / λίμνης στους ελέγχους Mann-Kendall. Μία περαιτέρω διερεύνηση θα είχε νόημα μόνο στην περίπτωση που οι έλεγχοι Mann-Kendall ήταν στατιστικά σημαντικοί (απέρριπταν τη μηδενική υπόθεση).

8.4.3. Μεταβλητότητα της τοξικότητας μεταξύ των ταμιευτήρων / λιμνών

Ήδη το διάγραμμα θηκογραμμάτων της τοξικότητας (Σχήμα 8.2) δίνει μία πρώτη ένδειξη ότι μάλλον δεν υπάρχουν διαφορές στις διαμέσους των τιμών τοξικότητας μεταξύ των τριών ταμιευτήρων / λιμνών. Λόγω της ύπαρξης μικρής αυτοσυσχέτισης στα δεδομένα τοξικότητας των ταμιευτήρων / λιμνών, οι παρατηρήσεις μέσα σε κάθε ομάδα δεν είναι πλήρως ανεξάρτητες. Η σειριακή συσχέτιση μπορεί να επηρεάσει τα αποτελέσματα της σύγκρισης των κεντρικών θέσεων των ομάδων λόγω της υποεκτίμησης των διακυμάνσεων τους (Shoukri και Pause, 1999). Για τον λόγο αυτόν, η ανάλυση των κεντρικών θέσεων των ταμιευτήρων / λιμνών έγινε στις προ-

λευκασμένες χρονοσειρές τους. Πριν την εφαρμογή της τεχνικής ανάλυσης διακύμανσης Kruskal-Wallis, που χρησιμοποιεί τις τάξεις μεγέθους στις προ-λευκασμένες χρονοσειρές για να εντοπίσει εάν υπάρχει μια σημαντική διαφορά μεταξύ Μαραθώνα, Υλίκης και Μόρνου όσον αφορά στην τοξικότητα, χρησιμοποιήθηκε ο έλεγχος Levene για τη διερεύνηση της ισότητας των διακυμάνσεων. Η p-τιμή του ελέγχου υπολογίστηκε ίση με 0,018 υποδηλώνοντας ότι σε επίπεδο σημαντικότητας $\alpha = 0,05$ απορρίπτεται η υπόθεση της ομοιογένειας των διακυμάνσεων των τριών ταμιευτήρων / λιμνών. Εναλλακτικά λοιπόν του ελέγχου Kruskal-Wallis, χρησιμοποιήθηκε ο έλεγχος των διαμέσων (Median test) των προλευκασμένων δεδομένων των ταμιευτήρων / λιμνών, ο οποίος είναι επίσης μη παραμετρικός και δεν εμπεριέχει την προϋπόθεση της ομοιογένειας των διακυμάνσεων των διαφορετικών ομάδων. Η p-τιμή του ελέγχου υπολογίστηκε ίση με 0,276 υποδηλώνοντας ότι σε επίπεδο σημαντικότητας $\alpha = 0,05$ δεν υπάρχει στατιστικά σημαντική διαφορά στην τοξικότητα μεταξύ των ταμιευτήρων / λιμνών. Στον Πίνακα 8.7 παρουσιάζονται οι συχνότητες των προ-λευκασμένων τιμών τοξικότητας για κάθε ταμιευτήρα / λίμνη που είναι μεγαλύτερες ή μικρότερες από τη γενική διάμεσο. Στον Πίνακα 8.8 παρουσιάζονται τα αποτελέσματα του ελέγχου διαμέσων.

Πίνακας 8.7. Πίνακας συνάφειας προ λευκασμένων τιμών τοξικότητας των ταμιευτήρων / λιμνών σε σχέση με τη γενική διάμεσο

	Ταμιευτήρας / λίμνη		
	Μαραθώνας	Υλίκη	Μόρνος
> Γενική διάμεσο	48	39	39
<= Γενική διάμεσο	36	45	45

Πίνακας 8.8. Περιληπτικά στατιστικά του ελέγχου διαμέσων (Median test) για τις προ-λευκασμένες τιμές τοξικότητας (μεταβλητή ομαδοποίησης: Ταμιευτήρας / λίμνη)

	Τοξικότητα
Αριθμός δειγμάτων	252
Διάμεσος	1,096
Chi-Square	2,571 ^a
df	2
p-τιμή	0,276

^aΚανένα κελί (0%) δεν είχε αναμενόμενη συχνότητα μικρότερη από 5. Η ελάχιστη αναμενόμενη συχνότητα ήταν 42.

8.5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα μελέτη έγινε ενδεδειγμένη στατιστική ανάλυση των τιμών τοξικότητας (εκφρασμένης ως ποσοστό αναστολής της φωταύγειας των βακτηρίων *Vibrio fischeri*) στους ταμιευτήρες Μαραθώνα και Μόρνου καθώς στη λίμνη Υλίκη για τη χρονική περίοδο 2000-2006, με τη χρήση εξειδικευμένων μη-παραμετρικών στατιστικών μεθόδων.

Οι τιμές της τοξικότητας στους τρεις ταμιευτήρες / λίμνες ήταν αρκετά χαμηλές και συγκρίσιμες με την τοξικότητα που συνήθως παρατηρείται στο επεξεργασμένο πόσιμο νερό (Κεφάλαιο 12). Στο διάστημα των επτά ετών που κάλυπταν τα διαθέσιμα δεδομένα, δεν παρουσιάστηκαν ισχυρές ενδείξεις εποχικότητας και ανοδικής ή πτωτικής τάσης των τιμών της τοξικότητας. Επιπλέον, δεν παρουσιάστηκε διαφοροποίηση των κεντρικών μέτρων θέσης της τοξικότητας μεταξύ των τριών ταμιευτήρων / λιμνών.

Αξίζει να σημειωθεί ότι η μέτρηση της τοξικότητας βάσει ενός είδους οργανισμού, υποδεικνύει την επίδραση του δείγματος στον συγκεκριμένο οργανισμό. Η εκτίμηση της τοξικότητας μέσω και άλλων οργανισμών, που ανήκουν σε διαφορετικά επίπεδα της τροφικής αλυσίδας, θα επέτρεπε μία περισσότερο ολοκληρωμένη αποτίμησή της. Ωστόσο, μόνο η παρακολούθηση και καταγραφή δεδομένων τοξικότητας χρησιμοποιώντας τον ίδιο οργανισμό-δείκτη, παρέχει τη δυνατότητα σύγκρισης των τιμών της (Ribo, 1997). Μέσω των αναλύσεων τοξικότητας ανιχνεύεται η συνδυασμένη επίδραση των τοξικών ουσιών. Σε περίπτωση εμφάνισης αυξημένων τιμών τοξικότητας, οι χημικές αναλύσεις είναι απαραίτητες για τον προσδιορισμό του προβλήματος.

ΚΕΦΑΛΑΙΟ 9

ΑΝΑΛΥΣΗ ΤΑΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΟΤΗΤΑ ΜΙΚΡΟΒΙΟΛΟΓΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ / ΛΙΜΝΩΝ

9.1. ΕΙΣΑΓΩΓΗ

Η μόλυνση από βακτηριακές πηγές έχει προσδιοριστεί ως τρίτη κύρια αιτία της ρύπανσης των ποταμών στις Η.Π.Α., μετά από την απόθεση ιλύος (siltation) και τις θρεπτικές ουσίες (USEPA, 1999). Είναι ευρύτατα αποδεκτό ότι το γένος *enterococcus* και τα *Escherichia coli* (*E.coli*) είναι πολύτιμοι βακτηριακοί δείκτες για τον καθορισμό της περιττωματικής μόλυνσης στο νερό λόγω της υψηλής αφθονίας τους στα περιττώματα θερμόαιμων ζώων.

Για τα επιφανειακά νερά που προορίζονται για ανθρώπινη κατανάλωση, πρέπει να εξασφαλίζεται ότι η ποιότητά τους συνάδει με τις οδηγίες που τίθενται από την Ευρωπαϊκή Ένωση. Σύμφωνα με το ΦΕΚ 438/Β/3.7.1988 και την οδηγία 75/440/EEC (περί της απαιτούμενης ποιότητας των υδάτων επιφανείας που προορίζονται για την παραγωγή πόσιμου νερού) για τα επιφανειακά νερά κατηγορίας Α2 τα ανώτερα επιθυμητά όρια τίθενται στις 2000 (colony forming units ανά 100 ml) cfu/100ml και 1000 cfu/100ml για τα *Escherichia coli* και τον *enterococci* αντίστοιχα. Εντούτοις, η παρακολούθηση της ποιότητας νερού οφείλει να είναι περισσότερο απαιτητική από τον απλό έλεγχο για τη διασφάλιση ότι δεν παραβιάζονται τα πρότυπα ποιότητας. Η ενδεδειγμένη παρακολούθηση είναι η διαδικασία της αναζήτησης πληροφοριών για τη συμπεριφορά των ποιοτικών μεταβλητών του νερού στο περιβάλλον (π.χ. μέση κατάσταση, τάσεις, ακραίες τιμές). Οι εποχικές αλλαγές μπορούν επίσης να επηρεάσουν τους βακτηριακούς πληθυσμούς (π.χ. Howell και λοιποί, 1995; Edwards και λοιποί, 1997).

Η έρευνα σχετικά με τη μικροβιακή φόρτιση του ανεπεξέργαστου νερού είναι χρήσιμο να μην περιορίζεται μόνο στην καταγραφή των τιμών των μικροβιολογικών

παραμέτρων και την σύγκρισή τους με τα ανώτερα επιθυμητά όρια της νομοθεσίας αλλά και να εξετάζονται οι τυχόν εποχικές διακυμάνσεις καθώς και η πιθανή ύπαρξη τάσης των τιμών. Σκοπός της παρούσας μελέτης είναι ακριβώς η λεπτομερειακή εξέταση της μικροβιακής φόρτισης των νερών των ποταμών που εκβάλλουν στους ταμειυτήρες / λίμνες που χρησιμοποιούνται για την ύδρευση της Αθήνας, για μια μακρά χρονική περίοδο, όσον αφορά στα βακτήρια *enterococci*, με τη χρήση εξειδικευμένων στατιστικών μεθόδων για τη συγκεκριμένη μορφή των διαθέσιμων δεδομένων. Τα βακτήρια *enterococci* ενδημούν στα εντερικά ίχνη (κόπρανα) ζώων, πτηνών και ανθρώπων. Η παρουσία *enterococci* στο νερό, είναι ένδειξη κοπρανώδους ρύπανσης. Πρέπει όμως να διευκρινιστεί ότι μικρός αριθμός *enterococci* που βρίσκονται στο νερό μπορεί επίσης να προέρχεται από άλλους βιότοπους (ISO/DIS 7899-2, 1998). Το πλεονέκτημα των *enterococci*, ως δείκτη κοπρανώδους ρύπανσης του νερού έναντι των *E.coli*, είναι ότι επιβιώνουν για μεγαλύτερο χρονικό διάστημα, με αποτέλεσμα τη μείωση του κινδύνου για ψευδώς αρνητικά αποτελέσματα (Forsythe, 2010; Edberg και λοιποί, 1997).

Η προκαταρκτική εξέταση των διαθέσιμων δεδομένων υποδήλωσε την αναγκαιότητα χρήσης περιγραφικών και μη παραμετρικών στατιστικών προσεγγίσεων για την αξιολόγηση της εποχικότητας σε κάθε ποταμό και τον προσδιορισμό τυχόν μονότονων χρονολογικών τάσεων. Οι μη παραμετρικές στατιστικές τεχνικές κρίθηκαν κατάλληλες προκειμένου να μειωθεί η επίδραση των απομακρυσμένων ή και ακραίων τιμών καθώς και των ελλειπυσών τιμών, για τη διαχείριση των περικομμένων τιμών και για να αποφευχθούν οι συνέπειες παραβιάσεων των προϋποθέσεων που απαιτούνται για τους παραμετρικούς στατιστικούς ελέγχους.

Ο έλεγχος και η στατιστική ανάλυση της συγκέντρωσης της προαναφερθείσας μικροβιολογικής παραμέτρου κατά τη διάρκεια μιας μεγάλης περιόδου εξασφαλίζει σημαντικές και χρήσιμες πληροφορίες στην υπηρεσία ύδρευσης με σκοπό τη διαχείριση και τον ποιοτικό έλεγχο των υδάτινων πόρων.

9.2. ΠΕΡΙΓΡΑΦΗ ΘΕΣΕΩΝ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΔΕΔΟΜΕΝΩΝ

Τα δείγματα νερού συλλέχθηκαν από τους ποταμούς / ρέματα, με συνεχή ροή, που εκβάλλουν στους ταμιευτήρες Μαραθώνα και Μόρνου καθώς και στη λίμνη Υλίκη, σε μηνιαία βάση, κατά τη διάρκεια μιας περιόδου δεκατριών ετών (1997 - 2009), και αναλύθηκαν για την ανίχνευση και απαρίθμηση των βακτηρίων *enterococci*. Συνολικά, οι θέσεις δειγματοληψίας ήταν έξι. Ειδικότερα οι θέσεις δειγματοληψίας ήταν: Για τον ταμιευτήρα Μαραθώνα, οι εκβολές του ποταμού Βαρνάβα και του ρέματος της Σταμάτας. Για τη λίμνη Υλίκη, η εκβολή του ποταμού Βοιωτικού Κηφισού. Και τέλος, για τον ταμιευτήρα Μόρνου, οι εκβολές των ποταμών Μόρνου, Κόκκινου και Άβορου.

Οι κοπρανώδεις εντερόκοκκοι είναι βακτήρια ικανά να ανάγουν το 2,3,5 – triphenyltetrazolium chloride (χλωριούχο 2,3,5-τριφαινυλτετραζόλιο) σε φορμαζίνη και να υδρολύουν την εσκουλίνη στους 44° C σε κατάλληλα θρεπτικά υλικά. Επιπλέον είναι κατά Gram-θετικά, καταλάσης-αρνητικά βακτήρια, που σχηματίζουν συνήθως αλυσίδες. Έχουν σχήμα ωοειδές έως σφαιρικό και διαθέτουν το αντιγόνο D (ISO 7899-2, 2000). Για την ανίχνευση και την αρίθμηση των αποικιών βακτηρίων *enterococci* χρησιμοποιήθηκε η μέθοδος διήθησης ορισμένου όγκου δείγματος νερού (100 ml) σε μεμβράνη. Το εκλεκτικό θρεπτικό υλικό ήταν το ‘Membrane filter *Enterococcus* Selective Agar acc. to Slanetz and Bartley’. Η μέθοδος περιλαμβάνει τα στάδια της διήθησης, επώασης, επιβεβαίωσης και απαρίθμησης και περιγράφεται πλήρως στο ISO 7899-2, 2000. Για την ανίχνευση και απαρίθμηση των *enterococci* δεν προηγείτο καμία αραίωση των δειγμάτων νερού. Τα δεδομένα περιείχαν σποραδικές ελλείπουσες τιμές καθώς και κάτω ή / και άνω περικομμένες τιμές. Οι άνω περικομμένες τιμές προσδιορίστηκαν σε δύο διαφορετικά επίπεδα μέγιστου επιπέδου απαρίθμησης των αποικιών τα οποία ήταν τα 300 cfu/100ml και τα 100 cfu/100ml. Συνεπώς χρησιμοποιήθηκε το μικρότερο από τα δύο άνω όρια απαρίθμησης. Επιπλέον, βάσει του ISO 8199 (2005), για τις μεθόδους διήθησης σε μεμβράνη τίθεται ως άνω όριο απαρίθμησης το 100 cfu. Οι κάτω περικομμένες τιμές αφορούσαν σε αποτελέσματα απαρίθμησης με λιγότερο από τέσσερις τυπικές αποικίες *enterococci*. Παρόλο που το όριο ανίχνευσης της μεθόδου, βάσει δοκιμών του μικροβιολογικού εργαστηρίου, είναι το 1 cfu/100 ml, το χαμηλότερο όριο προσδιορισμού σε συγκεκριμένο όγκο δείγματος V, λόγω της κατανομής Poisson,

είναι $(4/V) \cdot V_s$ όπου V_s είναι ο όγκος αναφοράς που έχει επιλεγθεί για την έκφραση της συγκέντρωσης των μικροοργανισμών στο δείγμα. Συνεπώς, για όγκο δείγματος 100 ml και όγκο αναφοράς επίσης 100 ml, το χαμηλότερο όριο προσδιορισμού είναι 4 cfu/100ml. Για τιμές αποικιών μεταξύ 1 και 3, η ακρίβεια του αποτελέσματος είναι εξαιρετικά μικρή (ISO 8199, 2005).

9.3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

Τα δεδομένα περιείχαν ελλείπουσες, ακραίες ή και περικομμένες τιμών *enterococci*. Για τον λόγο αυτόν κρίθηκε προτιμότερη η χρησιμοποίηση μιας μη παραμετρικής προσέγγισης. Ο μεγάλος αριθμός περικομμένων τιμών σε μερικές από τις μηνιαίες χρονολογικές σειρές βακτηρίων *enterococci* είχε σαν αποτέλεσμα την εξαίρεσή τους από την περαιτέρω στατιστική ανάλυση μονότονων τάσεων. Για τη διερεύνηση της επίδρασης της εποχικότητας στα βακτήρια *enterococci* σε κάθε θέση δειγματοληψίας χρησιμοποιήθηκε η γραφική μέθοδος των θηκογραμμάτων ανά εποχή (μήνα) στις τάξεις μεγέθους των τιμών *enterococci* (McLeod και λοιποί, 1991). Στις κάτω περικομμένες τιμές αντιστοιχήθηκε η τιμή 2 cfu/100ml (το ήμισυ του κατώτερου ορίου προσδιορισμού) και συνεπώς οι τάξεις μεγέθους τους σε κάθε θέση δειγματοληψίας ήταν ταυτισμένες (ties). Με ανάλογο τρόπο αντιμετωπίστηκαν και οι άνω περικομμένες τιμές με τη διαφορά ότι οι ταυτισμένες τάξεις μεγέθους τους ήταν μεγαλύτερες από τις τάξεις μεγέθους όλων των λοιπών παρατηρούμενων τιμών. Για αυτόν το σκοπό, τα δεδομένα κάθε διαφορετικής θέσης δειγματοληψίας ταξινομήθηκαν από την μικρότερη τιμή προς τη μεγαλύτερη και οι υπολογισμοί έγιναν στις τάξεις μεγέθους των δεδομένων. Η ανάλυση τάσεων σε κάθε θέση δειγματοληψίας έγινε βάσει του ελέγχου Mann- Kendall ή του εποχικού ελέγχου Kendall ανάλογα με την διαπίστωση έλλειψης ή ύπαρξης εποχικότητας αντίστοιχα. Λόγω της ύπαρξης ελλειπουσών τιμών, δεν ήταν δυνατή η ακριβής αποτίμηση της σειριακής συσχέτισης των δεδομένων σε κάθε θέση δειγματοληψίας. Συνεπώς, οι εποχικοί έλεγχοι Kendall διεξήχθησαν τόσο στην τυπική μορφή τους όσο και στην προσαρμοσμένη για την περίπτωση σειριακής συσχέτισης ώστε να είναι δυνατή μία σύγκριση των μεταξύ τους αποτελεσμάτων. Όλοι οι έλεγχοι τάσεων που διεξήχθησαν αφορούν στον εντοπισμό τυχόν μονότονων τάσεων στις παρατηρούμενες τιμές των *enterococci*, δηλαδή συμπεριλαμβάνουν τόσο τις ανθρωπογενείς όσο και τις φυσικές

επιδράσεις στις συγκεντρώσεις της παραμέτρου. Ορισμένες ποιοτικές παράμετροι του νερού ποταμών εξαρτώνται από τη ροή του ποταμού κατά έναν από δύο δυνητικούς τρόπους (Helsel και Hirsch, 1992). Ο ένας αφορά στη μείωση της συγκέντρωσης μίας παραμέτρου λόγω «διάλυσης» όταν αυξάνει η ροή, ενώ ο άλλος στην αύξηση της συγκέντρωσης λόγω «ξεπλύματος» (wash-off). Τα αποτελέσματα της ανάλυσης τάσεων στα δεδομένα μετά την απομάκρυνση της μεταβλητότητας που οφείλεται στη ροή (flow-adjustment) υποδεικνύουν κατά πόσο υπάρχει κάποια υποκείμενη άλλη αιτία μίας πιθανής ύπαρξης τάσης. Η έλλειψη δεδομένων για τη ροή των ποταμών, δεν επέτρεψε την προσαρμογή των συγκεντρώσεων *enterococci* βάσει της ροής και τη διενέργεια ανάλυσης τάσεων στις προσαρμοσμένες τιμές. Επομένως, τα αποτελέσματα της ανάλυσης τάσεων υποδεικνύουν την ύπαρξη ή μη μίας παρατηρούμενης μονότονης τάσης (McBride και Smith, 1997).

9.4. ΑΠΟΤΕΛΕΣΜΑΤΑ & ΣΥΖΗΤΗΣΗ

Στον Πίνακα 9.1 παρουσιάζονται τα ποσοστά ελλειψουσών και περικομμένων τιμών των βακτηρίων *enterococci* ανά θέση δειγματοληψίας στην περίοδο των 13 ετών.

Πίνακας 9.1. Περιγραφή χαρακτηριστικών των δεδομένων *Enterococci* ανά θέση δειγματοληψίας

Ταμειωτήρας / λίμνη	Θέση δειγ/ψίας	% Ελλειψουσών τιμών	% Κάτω περικομμένων τιμών	% Άνω περικομμένων τιμών	% Συνόλου περικομμένων τιμών
Μαραθώνας	π. Βαρνάβας	11,5	3,6	29,7	33,3
	ρ. Σταμάτας	9,0	16,9	17,6	34,5
Υλίκη	π. Β. Κηφισός	17,3	4,7	59,7	64,3
Μόρνος	π. Μόρνος	5,1	27,0	11,5	38,5
	π. Κόκκινος	7,1	40,7	9,7	50,3
	π. Άβωρος	5,3	45,6	3,2	48,8

Από τον Πίνακα 9.1 παρατηρείται ότι το 59,7% των τιμών *enterococci* στον Βοιωτικό Κηφισό είναι μεγαλύτερες από το άνω όριο προσδιορισμού της μεθόδου. Είναι εμφανές ότι η διάμεσος των τιμών *enterococci* στον ποταμό Βοιωτικό Κηφισό, το

χρονικό διάστημα 1997 - 2009, βρίσκεται εντός του συνόλου των άνω περικομμένων τιμών δηλαδή είναι μεγαλύτερη από 100 cfu/100ml και δεν υπάρχει δυνατότητα αριθμητικής έκφρασής της. Επομένως, προτείνεται η λήψη μεγαλύτερης ποσότητας δειγμάτων νερού ώστε, στην περίπτωση που οι τυπικές αποικίες υπερβαίνουν το 100, να υπάρχει η δυνατότητα επανάληψης της απαρίθμησης τους σε διαδοχικά αραιωμένα δείγματα ώστε να είναι δυνατή η απαρίθμηση των αποικιών και στη συνέχεια η αναγωγή των αποτελεσμάτων. Με αυτόν τον τρόπο θα περιοριστεί το πρόβλημα του μεγάλου αριθμού περικομμένων τιμών και θα είναι δυνατή η παρακολούθηση των τάσεων του μικροβιολογικού φορτίου των *enterococci* στον ποταμό. Παρόλα αυτά, αξίζει να σημειωθεί ότι η διάμεσος τιμή των μηνιαίων συγκεντρώσεων *enterococci* στο κέντρο της λίμνης Υλίκης, κατά το χρονικό διάστημα 2004–2005 ήταν μόλις 12 cfu/100 ml. Στο ίδιο χρονικό διάστημα, 2004-2005, οι διάμεσες τιμές των μηνιαίων συγκεντρώσεων *enterococci* στο πύργο υδροληψίας του Μαραθώνα (Βεντούρι) ήταν 24 cfu/100 ml ενώ στον πύργο υδροληψίας του ταμιευτήρα Μόρνου ήταν μικρότερη από 4 cfu/100 ml. Οι ποταμοί Κόκκινος και Άβωρος, που εκβάλλουν στον ταμιευτήρα Μόρνου, παρουσιάζουν μικρό μικροβιολογικό φορτίο όσον αφορά στα βακτήρια *enterococci* (η διάμεσος τιμή των μηνιαίων τιμών 1997-2009 είναι 8 cfu/100ml στον Κόκκινο και η διάμεσος των μηνιαίων τιμών 1997-2007 είναι 4 cfu/100ml στον Άβωρο) και έχουν υψηλό ποσοστό κάτω περικομμένων τιμών. Λόγω του υψηλού συνολικού ποσοστού των περικομμένων τιμών (σχεδόν 50% των δεδομένων ή και περισσότερο) *enterococci*, οι ποταμοί Βοιωτικός Κηφισός, Κόκκινος και Άβωρος εξαιρέθηκαν από την περαιτέρω στατιστική ανάλυση τάσεων. Στον Πίνακα 9.2 παρουσιάζονται οι διάμεσες τιμές των βακτηριδίων *enterococci* για τα υπόλοιπα σημεία δειγματοληψίας, τα οποία έχουν τις προϋποθέσεις για τη διενέργεια ανάλυσης τάσεων.

Πίνακας 9.2. Διάμεσες τιμές των μηνιαίων βακτηριδίων *Enterococci* ανά σημείο δειγματοληψίας το χρονικό διάστημα 1997-2009

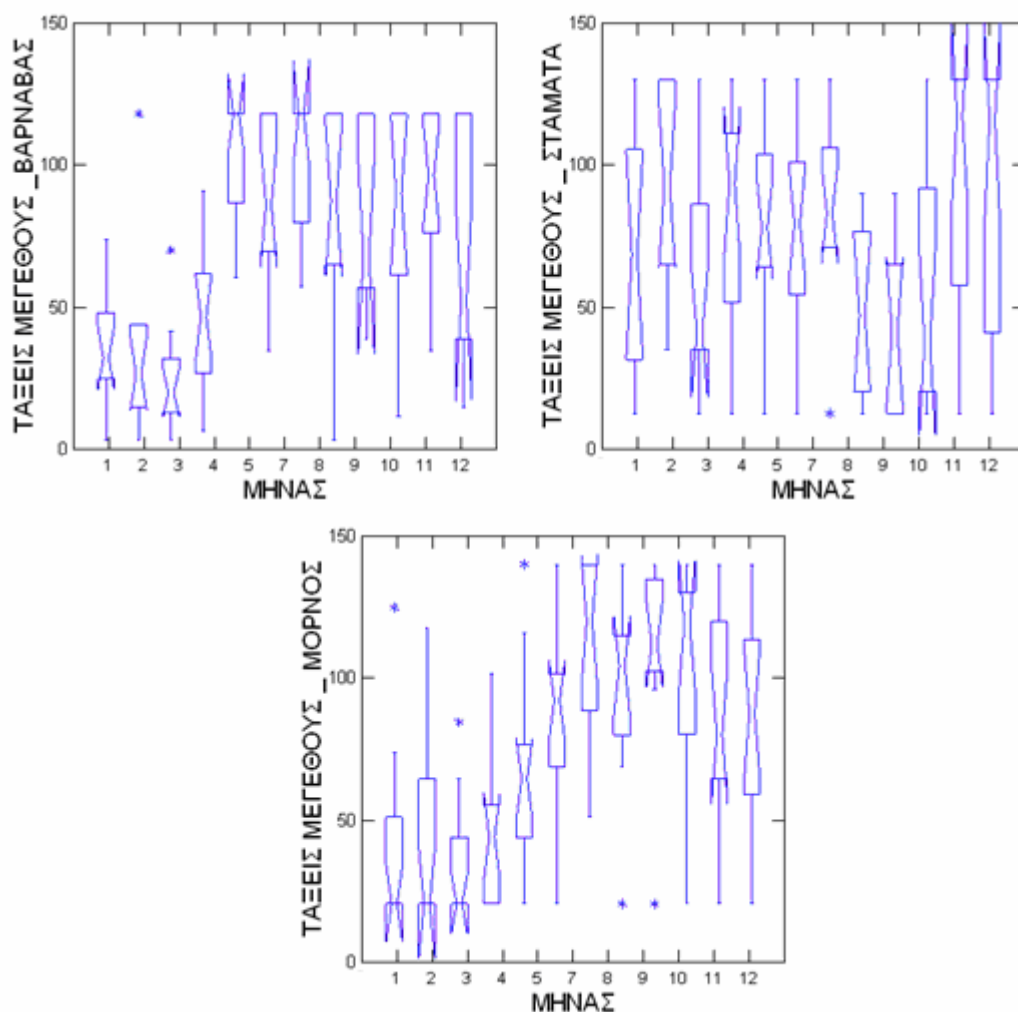
Ταμιευτήρας / λίμνη	Θέση δειγματοληψίας	Διάμεσος <i>Enterococci</i>
	π. Βαρνάβας	44,5
Μαραθώνας	ρ. Σταμάτας	12,0
Μόρνος	π. Μόρνος	11,0

Τα μηνιαία θηκογράμματα των τάξεων μεγέθους των τιμών *enterococci* χρησιμοποιήθηκαν για τη διερεύνηση ύπαρξης εποχικότητας στα δεδομένα σε κάθε

μία θέση δειγματοληψίας (Σχήμα 9.1). Στα διαγράμματα αυτά, όταν υπάρχει επικάλυψη των εγκοπών (notches) των πλαισίων στον κατακόρυφο άξονα, θεωρείται ότι προσεγγιστικά τα 95% διαστήματα εμπιστοσύνης των διαμέσων ανά κατηγορία αλληλεπικαλύπτονται και συνεπώς οι διάμεσοι δεν διαφέρουν στατιστικά σημαντικά μεταξύ τους σε επίπεδο σημαντικότητας περίπου $\alpha=0,05$. Το ύψος των εγκοπών πάνω και κάτω της διαμέσου υπολογίστηκαν με τον προσεγγιστικό τύπο (McGill και λοιποί, 1978; Reckhow και λοιποί, 1993):

$$\text{Όρια εγκοπών} = \text{Διάμεσος} \pm (1,57 * \text{Ενδοτεταρτημοριακό εύρος} / \sqrt{n}),$$

Όπου n το μέγεθος του δείγματος.



Σχήμα 9.1. Θηκογράμματα των *Enterococci* ανά μήνα σε κάθε θέση δειγματοληψίας.

Βάσει των θηκογραμμάτων αυτών, γίνεται εμφανής η ύπαρξη εποχικότητας για τους ποταμούς Μόρνο και Βαρνάβα ενώ για το ρέμα της Σταμάτας δεν είναι ευδιάκριτη κάποια ύπαρξη εποχικότητας.

Για τη διαχείριση των ελλειπουσών τιμών κατά τη διενέργεια των εποχικών ελέγχων Kendall ή του ελέγχου Mann-Kendall, δεδομένου ότι δεν μπορεί να γνωρίζει κανείς αν η ελλείπουσα τιμή είναι μεγαλύτερη ή μικρότερη από οποιαδήποτε πραγματική τιμή, θεωρείται δεν ισχύει τίποτα από τα δύο. Επομένως η συνάρτηση προσήμου (Κεφάλαιο 6, παράγραφος 6.2.2.) $\text{sgn}(x_j - x_k)$ θεωρήθηκε ότι λαμβάνει την τιμή μηδέν αν είναι ελλείπουσα η τιμή x_j ή x_k (Hirsch και Slack, 1984). Στον Πίνακα 9.3 παρουσιάζονται τα αποτελέσματα των εποχικών αμφίπλευρων ελέγχων Kendall, με ή χωρίς τροποποίηση για πιθανή ύπαρξη σειριακής συσχέτισης, για τα βακτηρίδια *enterococci* στους ποταμούς Βαρνάβα και Μόρνο καθώς και στο ρέμα της Σταμάτας. Σε καμία από τις θέσεις δειγματοληψίας δεν παρουσιάστηκε στατιστικά σημαντική τάση σε επίπεδο σημαντικότητας $\alpha=0,05$ είτε με τον τυπικό έλεγχο είτε με τον προσαρμοσμένο για την σειριακή συσχέτιση.

Ο έλεγχος Van Belle και Hughes χρησιμοποιήθηκε για τον έλεγχο της ομοιογένειας στις εποχικές τάσεις. Για όλες τις θέσεις δειγματοληψία, τα αποτελέσματα έδειξαν ότι δεν υπήρξαν αρκετές ενδείξεις για να απορριφθεί η μηδενική υπόθεση της ομοιογένειας σε επίπεδο σημαντικότητας $\alpha=0,05$ (Πίνακας 9.4).

Πίνακας 9.3. Περίληπτικά στατιστικά για τους εποχικούς ελέγχους Kendall για τα βακτηρίδια *enterococci*

Θέση Δειγματοληψίας	S'	p-τιμή (SK)	p-τιμή (SKC)
π. Βαρνάβας	-44	0,358	0,664
ρ. Σταμάτας	-1	0,999	0,999
π. Μόρνος	56	0,278	0,519

Πίνακας 9.4. Περίληπτικά στοιχεία των ελέγχων ομοιογένειας

Θέση δειγματοληψίας	Στατιστικό (X_h^2)	df	p-τιμή
π. Βαρνάβας	12,321	11	0,340
ρ. Σταμάτας	6,659	11	0,826
π. Μόρνος	16,053	11	0,139

Ειδικά για το ρέμα της Σταμάτας, όπου η ύπαρξη εποχικότητας είναι αμφισβητήσιμη, διεξήχθη επιπλέον ο έλεγχος Mann-Kendall. Το στατιστικό του ελέγχου υπολογίστηκε ίσο με -79 και η p-τιμή του ίση με 0,890 υποδηλώνοντας ότι δεν υπάρχει στατιστικά σημαντική τάση σε επίπεδο σημαντικότητας $\alpha = 0,05$ ή και οποιοδήποτε άλλο από τα συνηθισμένα επίπεδα σημαντικότητας.

9.5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα μελέτη διερευνήθηκαν τα αποτελέσματα των προσδιορισμών των βακτηρίων – δεικτών μικροβιακής φόρτισης *enterococci* των ποταμών που εκβάλουν στους τρεις ταμιευτήρες / λίμνες. Προκειμένου να μειωθεί η επιρροή των ακραίων και των ελλειπουσών τιμών, για τη διαχείριση των περικεκομμένων τιμών και για να αποφευχθούν οι παραβιάσεις των υποθέσεων των παραμετρικών ελέγχων, χρησιμοποιήθηκαν γραφικές και μη παραμετρικές στατιστικές μέθοδοι για την αξιολόγηση των διαφορών μεταξύ των θέσεων δειγματοληψίας, της εποχικότητας και τον προσδιορισμό μονότονων χρονικών τάσεων. Τα μηνιαία δεδομένα εντεροκόκκων υπεβλήθησαν σε μια σειρά στατιστικών αναλύσεων με σκοπό τη διερεύνηση εποχικότητας, ύπαρξης τάσεων στα βακτήρια εντεροκόκκων και τον έλεγχο της ομοιογένειας των τάσεων.

Δύο από τους ποταμούς που εκβάλουν στον ταμιευτήρα Μόρνου χαρακτηρίζονται από εξαιρετικά μικρό μικροβιακό φορτίο όσον αφορά στα βακτήρια εντεροκόκκων και εξαιρέθηκαν της ανάλυσης τάσεων. Αντίθετα ο μοναδικός ποταμός (Βοιωτικός Κηφισός) που εκβάλλει στην λίμνη Υλίκη, εξαιρέθηκε της ανάλυσης τάσεων λόγω της παρουσίας υψηλού ποσοστού άνω περικομμένων τιμών. Συνεπώς δεν είναι εφικτή η αξιολόγηση της μικροβιακής φόρτισης του ποταμού ως προς τα βακτήρια εντεροκόκκων. Γενικότερα, δεν μπορεί κανείς να αποφανθεί αν οι άνω περικομμένες τιμές των βακτηρίων είναι κάτω από τα όρια που αναφέρονται στην οδηγία 75/440/EEC του Ευρωπαϊκού Συμβουλίου η οποία αφορά στην ποιότητα των επιφανειακών υδάτων που προορίζονται για την παρασκευή νερού για ανθρώπινη κατανάλωση. Σε καμία από τις υπόλοιπες θέσεις δειγματοληψίας δεν εντοπίστηκε στατιστικά σημαντική ανοδική τάση των παρατηρούμενων τιμών εντεροκόκκων.

ΚΕΦΑΛΑΙΟ 10

ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΦΥΣΙΚΟΧΗΜΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ/ΛΙΜΝΩΝ

10.1. ΕΙΣΑΓΩΓΗ

Η παρακολούθηση της ποιότητας του επιφανειακού νερού, που προορίζεται για ανθρώπινη κατανάλωση, είναι απαραίτητη για την αποτίμηση της ποιότητάς του. Ωστόσο, παρέχει μεγάλα σύνολα δεδομένων τα οποία συχνά είναι δύσκολο να ερμηνευθούν. Οι πολυμεταβλητές στατιστικές τεχνικές βοηθούν στην ερμηνεία των σύνθετων συνόλων δεδομένων, όπως εκείνα που δημιουργούνται από τα μακροπρόθεσμα προγράμματα ελέγχου ποιότητας νερού, επιτρέποντας μια καλύτερη κατανόηση των χρονικών ή/και των χωρικών διακυμάνσεων στην ποιότητα του νερού. Στην τελευταία δεκαετία, η εφαρμογή των πολυμεταβλητών στατιστικών τεχνικών έχει χρησιμοποιηθεί ευρέως για την ανάλυση δεδομένων του νερού και την εξαγωγή σημαντικών πληροφοριών (Singh και λοιποί., 2004; Shrestha και Kazama, 2007; Simeonon και λοιποί., 2001; Simeonon και λοιποί., 2003; Wunderlin και λοιποί., 2001; Paratheodorou και λοιποί., 2006).

Στο κεφάλαιο αυτό διερευνήθηκε εκτενώς η χρησιμότητα ενός μεγάλου αριθμού πολυμεταβλητών στατιστικών τεχνικών - όπως είναι η ανάλυση κύριων συνιστωσών (PCA), η παραγοντική ανάλυση (FA), η ανάλυση κατά συστάδες (CA), η διαχωριστική ανάλυση (DA) και τα δέντρα ταξινόμησης και παλινδρόμησης (CART) - στην ερμηνεία της δομής των φυσικοχημικών ποιοτικών παραμέτρων του νερού της λίμνης Υλίκης και των ταμιευτήρων Μόρνου και Μαραθώνα και στην αποτίμηση των χρονικών διακυμάνσεων της ποιότητάς τους. Εξετάστηκαν οι συσχετίσεις μεταξύ των παραμέτρων και δημιουργήθηκε μοντέλο για την επεξήγησή τους. Αναπτύχθηκαν και επικυρώθηκαν μοντέλα ομαδοποίησης των δειγμάτων του νερού ανάλογα με την χρονική περίοδο δειγματοληψίας. Επιτεύχθηκε σημαντική μείωση των διαστάσεων του πολύπλοκου σετ των δεδομένων (μεγάλος αριθμός παραμέτρων και αριθμού

δειγμάτων) και αναδείχθηκαν λίγες μόνο ποιοτικές παράμετροι ως κύριες υπεύθυνες για τις διακυμάνσεις στην ποιότητα του νερού στη διάρκεια του χρόνου. Τα μοντέλα που κατασκευάστηκαν, λαμβάνουν υπόψη τους την πολυπλοκότητα της πραγματικότητας και προβλέπουν τη μελλοντική συμπεριφορά του συστήματος των φυσικοχημικών παραμέτρων κάθε ταμιευτήρα/λίμνης.

10.2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

10.2.1 Περιγραφή των δεδομένων

Τα διαθέσιμα δεδομένα αφορούσαν σε δείγματα νερού των τριών ταμιευτήρων/λιμνών, τα οποία αναλύθηκαν για 16 φυσικοχημικές παραμέτρους στη διάρκεια τεσσάρων ή πέντε ετών σε μηνιαία βάση, προκειμένου να επιβεβαιωθεί η συμμόρφωση των ποιοτικών παραμέτρων του επιφανειακού νερού, που προορίζεται για την παραγωγή του πόσιμου, με τις οδηγίες της Ευρωπαϊκής Ένωσης και της Ελληνικής Νομοθεσίας. Τα δείγματα της λίμνης Υλίκης κάλυπταν τη χρονική περίοδο από τον Σεπτέμβριο του 2003 έως τον Μάιο του 2008, του ταμιευτήρα Μόρνου τα έτη από 2003 έως και 2006, ενώ τα δείγματα του ταμιευτήρα Μαραθώνα την περίοδο από Σεπτέμβριο του 2003 έως τον Σεπτέμβριο του 2007. Στην διάρκεια της περιόδου αυτής, ο ταμιευτήρας του Μαραθώνα ενισχύονταν με νερό αποκλειστικά από τον ταμιευτήρα του Μόρνου. Τα δείγματα συνελέγησαν από τους πύργους υδροληψίας των ταμιευτήρων Μαραθώνα και Μόρνου και το κέντρο της λίμνης Υλίκη (σε βάθος 30 εκατοστών περίπου από την επιφάνεια). Οι παράμετροι για τις οποίες αναλύθηκαν ήταν: Θολότητα (TU), πυκνότητα συγκέντρωσης υδρογόνου (pH), ηλεκτρική αγωγιμότητα (EC), ολική αλκαλικότητα (TA), ολική σκληρότητα (TH), ολικά διαλελυμένα στερεά (TDS), διαλελυμένο οξυγόνο (DO), χημικά απαιτούμενο οξυγόνο (COD), νιτρικά (NO_3^-), ασβέστιο (Ca^{2+}), αμμώνιο (NH_4^+), χλωριούχα (Cl^-), κατιόντα νατρίου (Na^+), θειικά (SO_4^{2-}), φωσφορικά (PO_4^{3-}) και κατιόντα μαγνησίου (Mg^{2+}). Όλες οι παραπάνω παράμετροι ποιότητας του νερού εκφράζονται σε mg/l, εκτός από την πυκνότητα συγκέντρωσης υδρογόνου, την θολότητα (σε NTU) και την ηλεκτρική αγωγιμότητα (σε $\mu\text{S}/\text{cm}$). Οι αναλύσεις των προαναφερθεισών παραμέτρων διεξήχθησαν στο χημικό εργαστήριο της ΕΥΔΑΠ με τη χρησιμοποίηση μεθόδων που περιγράφονται στο «Τυπικές μέθοδοι για την

εξέταση νερού και αποβλήτων» (Standard Methods for the Examination of Water and Wastewater) (APHA, AWWA και WEF, 1992). Η ποιότητα των μετρήσεων εξασφαλίστηκε μέσω της προσεκτικής τυποποίησης των μεθόδων, της διαδικασίας των τυφλών δειγμάτων, των αναλύσεων προτύπων και διπλών δειγμάτων. Η περιγραφή των παραμέτρων, οι μέθοδοι προσδιορισμού και τα όρια της νομοθεσίας παρουσιάζονται στο Παράρτημα I (Πίνακες Π1, Π2 και Π3 αντίστοιχα).

10.2.2. Πολυμεταβλητές στατιστικές τεχνικές

Τόσο η ανάλυση σε κύριες συνιστώσες (PCA) όσο και η παραγοντική ανάλυση (FA) χρησιμοποιήθηκαν στα δεδομένα κάθε ταμειυτήρα ως τεχνικές μείωσης των δεδομένων, αποσκοπώντας στη μείωση των μεγάλων συνόλων συσχετισμένων μεταβλητών σε μερικούς μόνο σύνθετους δείκτες. Και οι δύο τεχνικές περιλαμβάνουν το σχηματισμό νέων ασυσχέτιστων μεταβλητών (κύριες συνιστώσες {PCs} ή παραγόντων {VFs} αντίστοιχα) που είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών. Ωστόσο, η ανάλυση παραγόντων πηγαίνει ένα βήμα περαιτέρω και προσπαθεί να προσδιορίσει μια θεμελιώδη δομή ή κατασκευή των δεδομένων. Με την παραγοντική ανάλυση αναζητήθηκε να ερμηνευτεί πώς συσχετίζονται ορισμένες μεταβλητές. Η ανάλυση κύριων συνιστωσών χρησιμοποιήθηκε ως αρχικό βήμα για να καθορίσει πόσοι παράγοντες θα χρησιμοποιηθούν στην παραγοντική ανάλυση. Η ανάλυση κύριων συνιστωσών, ως μια μαθηματική μέθοδος για τη μείωση δεδομένων, δεν υποθέτει οποιαδήποτε ιδιαίτερη κατανομή των δεδομένων. Για την παραγοντική ανάλυση που βασίζεται στην ανάλυση κυρίων συνιστωσών (PCA/FA), παρότι η κανονική κατανομή δεν είναι απαραίτητη, αυτή η μέθοδος βασίζεται στον πίνακα συσχετίσεων ή συνδιακυμάνσεων οι οποίοι επηρεάζονται έντονα από την έλλειψη κανονικότητας των δεδομένων και την παρουσία ακραίων τιμών. Δεδομένου ότι οι κατανομές πιθανότητας πολλών παραμέτρων ποιότητας του νερού είναι συχνά μη-κανονικά κατανεμημένες (Loftis και λοιποί, 1989; Montgomery και λοιποί, 1987; Giliom και Helsel, 1986) - περιορίζονται στο μηδέν και είναι συχνά ιδιαίτερα ασύμμετρες, περιλαμβάνοντας σπάνιες υψηλές τιμές (Helsel, 1987) - πριν τη διεξαγωγή PCA/FA, επιδιώχτηκε όλες οι μεταβλητές να έρθουν όσο το δυνατόν περισσότερο κοντά σε μια κανονική κατανομή (Reimann και Filzmoser, 2000). Ως εκ τούτου, όλες οι μεταβλητές εξετάστηκαν για την κανονικότητα και εκείνες που

παραβίασαν αυτήν την υπόθεση, υπεβλήθησαν σε δεκαδικό λογαριθμικό μετασχηματισμό και επανεξετάστηκαν ως προς την κανονικότητα πριν την εισαγωγή τους στις αναλύσεις που επακολούθησαν. Προκειμένου να αποφευχθούν πιθανές συνέπειες λόγω των διαφορετικών τάξεων μεγέθους καθώς και του εύρους της διακύμανσης των μεταβλητών, προηγήθηκε των PCA/FA η τυποποίηση των δεδομένων (Liu και λοιποί, 2003; Singh και λοιποί, 2004). Για τους σκοπούς της ερμηνείας των δεδομένων, η «υψηλή» φόρτιση ορίστηκε ως μεγαλύτερη από 0,75, η μέτρια φόρτιση ως 0,50-0,75 και η ασθενής φόρτιση ως 0,30-0,50 (Liu και λοιποί, 2003). Προκειμένου να αποφασιστεί ο αριθμός των εξαγόμενων παραγόντων, χρησιμοποιήθηκαν το γράφημα των ιδιοτιμών (scree-plot) καθώς και το κριτήριο Kaiser των ιδιοτιμών.

Η πολυμεταβλητή μέθοδος της ανάλυσης κατά συστάδες χρησιμοποιήθηκε για τον εντοπισμό φυσικών συνόλων ομαδοποιημένων στοιχείων στα δεδομένα κάθε ταμειυτήρα. Οι παρατηρήσεις ομαδοποιήθηκαν σε συστάδες έτσι ώστε κάθε συστάδα να είναι όσο το δυνατόν περισσότερο ομοιογενής (παρόμοια αντικείμενα να είναι στην ίδια κλάση) όσον αφορά στις μεταβλητές και οι ομάδες να είναι όσο το δυνατόν πιο διαφορετικές. Οι συστάδες σχηματίστηκαν με τη χρησιμοποίηση της ιεραρχικής συσσωρευτικής ανάλυσης κατά συστάδες, η οποία δεν συνδέεται με την εκ των προτέρων γνώση του αριθμού των κλάσεων. Ως μέτρο της ομοιότητας μεταξύ των δειγμάτων χρησιμοποιήθηκε η τετραγωνική ευκλείδεια απόσταση ενώ ως μέτρο των αποστάσεων μεταξύ των συστάδων χρησιμοποιήθηκε η μέθοδος σύνδεσης του Ward (Ward's linkage method). Η ανάλυση κατά συστάδες δεν έχει καμία υπόθεση για την κατανομή των δεδομένων. Δεδομένου όμως ότι οι μεταβλητές των ποιοτικών χαρακτηριστικών του νερού έχουν διαφορετικές μονάδες μέτρησης, τα πρωταρχικά δεδομένα τυποποιήθηκαν πριν από τη διεξαγωγή των αναλύσεων κατά συστάδες (Rogerson, 2001). Η τυποποίηση των δεδομένων εξασφάλισε ότι κάθε μεταβλητή έχει την ίδια επιρροή στην ανάλυση (Massart και Kaufman, 1983). Ως μέτρο απόστασης θα μπορούσε να χρησιμοποιηθεί η απόσταση Mahalanobis (Mahalanobis distance) η οποία λαμβάνει υπόψη τις συσχετίσεις μεταξύ των μεταβλητών και δεν επηρεάζεται από την κλίμακα μέτρησης τους (Καρλής, 2005; Hair, 1998).

Η διαχωριστική ανάλυση χρησιμοποιήθηκε στα δεδομένα κάθε ταμειυτήρα για τον προσδιορισμό των μεταβλητών που διαχωρίζουν τις διαφορετικές εποχικές ομάδες,

που προέκυψαν από την PCA/FA και την ανάλυση κατά συστάδες, για να αξιολογήσει τη συμβολή κάθε μεταβλητής στο μοντέλο και για να χρησιμοποιήσει αυτές τις μεταβλητές για να προβλέψει την ταξινόμηση μελλοντικών δειγμάτων με μεγαλύτερη ακρίβεια (Hair και λοιποί., 1998; Johnson και Wichern, 1998). Για τον προσδιορισμό των μεταβλητών που θα έπρεπε να συμπεριλαμβάνονται στα μοντέλα και για την κατασκευή των διαχωριστικών συναρτήσεων (DFs) με σκοπό την αποτίμηση των χρονικών διακυμάνσεων της ποιότητας νερού, χρησιμοποιήθηκε η διαδικασία της βηματικής διαχωριστικής ανάλυσης (stepwise DA). Η διαχωριστική ανάλυση υποθέτει ότι οι μεταβλητές πρέπει να ακολουθούν πολυμεταβλητή κανονική κατανομή μέσα σε κάθε ομάδα και ότι οι πίνακες διακυμάνσεων – συνδιακυμάνσεων μέσα σε κάθε ομάδα πρέπει να είναι ίσοι μεταξύ τους. Για το λόγο αυτόν, πέρα από τον έλεγχο κανονικότητας, χρησιμοποιήθηκαν και οι κατάλληλοι έλεγχοι για την ισότητα των πινάκων διακυμάνσεων – συνδιακυμάνσεων.

Τα δέντρα ταξινόμησης χρησιμοποιήθηκαν ως μη-παραμετρική μέθοδος εναλλακτική της διαχωριστικής ανάλυσης. Ο αλγόριθμος των «δέντρων ταξινόμησης και παλινδρόμησης» (CART) χρησιμοποιήθηκε για τον προσδιορισμό ενός συνόλου «εάν-τότε» (*if-then*) λογικών (διαχωριστικών) συνθηκών που επιτρέπουν την ακριβή πρόβλεψη ή την ταξινόμηση των περιπτώσεων (Breiman και λοιποί, 1984). Η προσέγγιση που χρησιμοποιείται για τη διάκριση στα δέντρα ταξινόμησης πλησιάζει περισσότερο την ανθρώπινη λογική διότι χρησιμοποιούν μια παράμετρο τη φορά.

Για τις στατιστικές αναλύσεις των δεδομένων χρησιμοποιήθηκαν τα στατιστικά πακέτα SPSS 15 για Windows και Statistica 7.

10.3. ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΘΕΣΕΩΝ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ / ΛΙΜΝΩΝ ΒΑΣΕΙ ΤΩΝ ΦΥΣΙΚΟ-ΧΗΜΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΟΥ ΝΕΡΟΥ

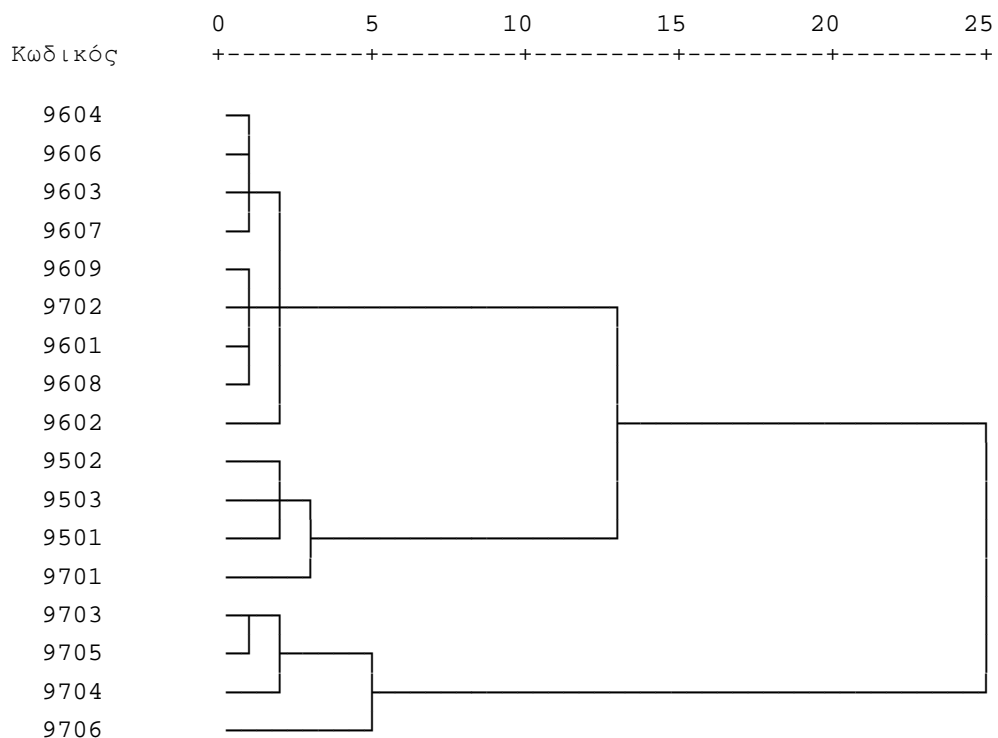
Η ανάλυση κατά συστάδες χρησιμοποιήθηκε για την ταξινόμηση των θέσεων δειγματοληψίας των τριών ταμιευτήρων/λιμνών βάσει των φυσικο-χημικών χαρακτηριστικών του νερού. Χρησιμοποιήθηκαν οι διάμεσες τιμές των παραμέτρων για το χρονικό διάστημα 2004-2006 σε κάθε θέση δειγματοληψίας (Πίνακας 10.1).

Πίνακας 10.1. Διάμεσες τιμές των παραμέτρων την τριετία 2004-2006 για κάθε θέση δειγματοληψίας στους ταμιευτήρες/λίμνες

Ταμιευτήρας / λίμνη	Θέση δειγματοληψίας	Κωδικός θέσης	Αριθμός δειγμάτων	EC	TDS	TU	pH	TA	TH	Cl ⁻	SO ₄ ²⁻	Na ⁺	Ca ²⁺	Mg ²⁺	COD	DO	NH ₄ ⁺	NO ₃ ⁻
Υλίκη	Β. Κηφισός	9501	35	531	310	4,6	8,2	252	277	11,9	33,5	10,7	76	22,9	1,7	8,9	0,074	7,47
Υλίκη	Κέντρο λίμνης	9502	35	415	244	2,6	8,4	164	197	15,5	39,5	12,7	43	22,9	1,7	8,7	0,042	0,41
Υλίκη	Δίαυλος αντλιοστασίου	9503	35	412	258	5,7	8,3	165	196	15,6	39,8	12,8	41	23,6	1,8	7,8	0,078	0,42
Μόρνος	π. Μόρνος	9601	34	314	187	1,3	8,2	153	165	2,9	16,8	4,3	55	7,4	0,5	9,7	0,019	0,26
Μόρνος	π. Άβορος	9602	34	442	275	0,7	8,1	188	225	5,8	43,8	9,9	72	11,1	0,4	9,4	0,018	0,02
Μόρνος	Κέντρο λίμνης	9603	35	268	157	1,5	8,3	122	136	2,8	17,4	4,3	46	5,3	0,6	9,1	0,021	0,03
Μόρνος	Λύματα Λιδωρικίου	9604	35	266	156	2,1	8,3	122	135	2,9	17,3	4,3	46	5,3	0,6	8,9	0,020	0,02
Μόρνος	Πύργος υδροληψίας	9606	35	266	156	2,2	8,3	123	136	2,8	17,2	4,3	46	5,3	0,6	9,0	0,021	0,03
Μόρνος	Κατάντι	9607	35	267	154	1,4	8,2	122	137	2,8	17,4	4,2	46	5,4	0,6	8,9	0,023	0,10
Μόρνος	π. Κόκκινος	9608	35	302	176	2,8	8,1	146	156	2,6	15,3	4,2	55	4,7	0,4	9,5	0,017	0,14
Μόρνος	π. Εύηνος	9609	17	281	164	4,5	8,2	128	135	3,0	19,9	4,8	46	4,7	0,7	9,7	0,019	0,10
Μαραθώνας	Πύργος υδροληψίας	9701	35	388	219	9,1	8,0	149	174	17,7	27,4	12,6	56	7,8	1,1	8,6	0,040	2,42
Μαραθώνας	ρ. Κιούρκων	9702	34	279	162	3,5	8,2	130	142	3,0	18,0	4,5	49	5,5	0,6	9,8	0,024	0,32
Μαραθώνας	π. Βαρνάβας	9703	34	587	351	0,6	7,9	214	262	35,8	35,5	21,6	94	6,7	0,4	9,0	0,021	10,80
Μαραθώνας	ρ. Καταρράκτης	9704	13	683	448	0,6	8,0	210	304	48,0	58,2	26,0	113	7,2	0,5	9,7	0,030	32,20
Μαραθώνας	ρ. Σταμάτας	9705	35	728	440	0,9	7,9	237	314	55,2	53,6	31,6	107	11,2	0,4	9,0	0,024	15,36
Μαραθώνας	ρ. Αγ. Στεφάνου	9706	22	910	583	1,7	8,1	287	365	78,4	83,8	59,7	119	21,1	1,5	9,6	0,029	9,80

Το δεντρόγραμμα του Σχήματος 10.1 δίνει μια οπτική αντιπροσώπευση των βημάτων σε μια ιεραρχική λύση ομαδοποίησης η οποία παρουσιάζει τις συστάδες που συνδυάζονται και τις τιμές των συντελεστών απόστασης σε κάθε βήμα. Οι συνδεδεμένες κάθετες γραμμές υποδεικνύουν τις ενωμένες περιπτώσεις. Στο δεντρόγραμμα έχει γίνει αναπροσαρμογή των πραγματικών αποστάσεων σε αριθμούς μεταξύ του 0 και του 25 διατηρώντας την αναλογία των αποστάσεων μεταξύ των βημάτων.

▽



Σχήμα 10.1. Ιεραρχικό δεντρόγραμμα για τις θέσεις δειγματοληψίας των τριών ταμιευτήρων/λιμνών.

Όπως ήταν αναμενόμενο, διακρίνονται τρεις κλάσεις στις οποίες ταξινομούνται γενικά οι θέσεις δειγματοληψίας για κάθε ταμιευτήρα/λίμνη. Εξάιρεση αποτελεί το ρέμα των Κιούρκων (κωδικός 9702) το οποίο ταξινομείται στην κλάση θέσεων δειγματοληψίας του Μόρνου και όχι του Μαραθώνα. Η εξήγηση αποδίδεται στο γεγονός ότι κατά το συγκεκριμένο χρονικό διάστημα ο Μαραθώνας ενισχύετο με νερό του Μόρνου, το οποίο, μετά τη διέλευσή του από το διωλιστήριο Πολυδενδρίου και την επεξεργασία ενός μέρους του, καταλήγει στον ταμιευτήρα Μαραθώνα μέσω

της σήραγγας Κιούρκων ή ρέματος Κιούρκων. Επιπλέον, ο πύργος υδροληψίας του Μαραθώνα (κωδικός 9701) συμπεριλαμβάνεται στη κλάση που περιέχει τις θέσεις δειγματοληψίας της Υλίκης, κέντρο λίμνης (9502), διάυλος αντλιοστασίου (9503) και π. Β. Κηφισός (9501), αν και παρουσιάζεται διάκριση όσον αφορά την απόστασή του από αυτές. Οι υπόλοιπες θέσεις δειγματοληψίας στον ταμιευτήρα Μαραθώνα αφορούν στα ρέματα που εκβάλουν στον ταμιευτήρα - Βαρνάβα (9703), Σταμάτα (9705), Καταρράκτη (9704) και Αγ. Στέφανο (9706) - συγκροτούν μία κλάση και διαχωρίζονται πρώτες από όλες τις υπόλοιπες θέσεις δειγματοληψίας. Το ρέμα του Αγίου Στεφάνου παρουσιάζει τη μεγαλύτερη απόσταση από τα υπόλοιπα ρέματα του ταμιευτήρα βάσει των φυσικο-χημικών ιδιοτήτων του νερού του.

Στον ταμιευτήρα Μόρνου, οι εσωτερικές θέσεις δειγματοληψίας (λύματα Λιδορικού (9604), πύργος υδροληψίας (9606), κέντρο ταμιευτήρα (9603) και Κατάντι (9607), παρουσιάζουν υψηλή ομοιότητα μεταξύ τους όσον αφορά στις φυσικο-χημικές ιδιότητες του νερού. Το ίδιο ισχύει για τις θέσεις δειγματοληψίας του ταμιευτήρα Μόρνου που αφορούν στα κύρια σημεία εισόδου του νερού στον ταμιευτήρα (π. Εύηνος, π. Μόρνος και π. Κόκκινος). Ο ποταμός Άβωρος διαφοροποιείται ελαφρώς σε σχέση με τις υπόλοιπες θέσεις δειγματοληψίας στον ταμιευτήρα Μόρνου. Εντούτοις οι θέσεις δειγματοληψίας του ταμιευτήρα συγκροτούν μία κλάση με μικρές αποστάσεις μεταξύ των διαφορετικών θέσεων.

10.4. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΧΟΛΙΑ ΓΙΑ ΤΟΝ ΤΑΜΙΕΥΤΗΡΑ ΜΑΡΑΘΩΝΑ

10.4.1. Προκαταρκτική εξέταση των παραμέτρων και διαχείριση των δεδομένων

Οι τιμές των δεδομένων για τα φωσφορικά (PO_4^{3-}) ήταν σε υψηλό ποσοστό περικεκομμένες. Συγκεκριμένα, το ποσοστό των δειγμάτων για τα οποία οι τιμές των PO_4^{3-} ήταν κάτω του ορίου ανίχνευσης, ήταν 82,2%. Συνεπώς, η παράμετρος PO_4^{3-} εξαιρέθηκε από τις περαιτέρω αναλύσεις. Στον Πίνακα 10.2 παρουσιάζονται τα στατιστικά περιγραφικά μέτρα για κάθε μία παράμετρο. Σε τρεις από τις παραμέτρους (TU , COD και NH_4^+) παρατηρήθηκε θετική ασυμμετρία ενώ για μία μεταβλητή (SO_4^{2-}) παρατηρήθηκε μέτρια αρνητική ασυμμετρία. Για τις θετικά ασύμμετρες

μεταβλητές, χρησιμοποιήθηκε ο δεκαδικός λογαριθμικός μετασχηματισμός (\log_{10}) ώστε να βελτιωθούν τα χαρακτηριστικά αυτών των μεταβλητών (Paratheodorou και λοιποί, 2006; Çamdevýren και λοιποί, 2005). Μετά τον δεκαδικό λογαριθμικό μετασχηματισμό τους οι μεταβλητές TU και COD, όπως φαίνεται στον Πίνακα 10.3, ακολουθούσαν προσεγγιστικά την κανονική κατανομή ενώ η μεταβλητή NH_4^+ βελτίωσε αισθητά τα χαρακτηριστικά της. Ωστόσο, ο μετασχηματισμός αυτός δεν χρησιμοποιήθηκε για τα SO_4^{2-} διότι θα επέτεινε το πρόβλημα της μέτριας αρνητικής ασυμμετρίας της κατανομής τους (από μέτρια σε μεγάλη).

Πίνακας 10.2. Περιγραφικά στατιστικά μέτρα των παραμέτρων για τον ταμιευτήρα Μαραθώνα

Παράμετρος	Μέσος	Τυπική απόκλιση	Ασυμμετρία ^A	Κύρτωση ^B
EC	369,43	34,75	-0,21	-1,00
TDS	213,19	27,67	-0,05	1,26
TU	16,68	19,03	2,21	4,75
pH	7,95	0,22	0,00	-1,08
TA	145,70	12,90	0,32	1,03
TH	169,98	12,86	0,16	-0,54
Cl^-	16,25	4,05	0,25	-0,96
SO_4^{2-}	24,42	7,02	-1,66	1,82
Na^+	11,48	2,33	0,24	-1,25
Ca^{2+}	54,88	5,21	-0,14	-0,85
Mg^{2+}	7,68	0,84	0,38	-0,42
COD	1,11	0,37	2,00	6,79
DO	8,09	1,78	-0,35	-0,67
NH_4^+	0,45	1,01	2,51	5,14
NO_3^-	2,16	1,47	0,80	1,18

A: Τυπικό σφάλμα ασυμμετρίας = 0,347 και B: Τυπικό σφάλμα κύρτωσης = 0,681

Πίνακας 10.3. Μέτρα ασυμμετρίας και κύρτωσης των \log_{10} μετασχηματισμένων μεταβλητών για τον ταμιευτήρα Μαραθώνα

Παράμετρος (σε \log_{10})	Ασυμμετρία ^A	Κύρτωση ^B
TU	0,15	-0,36
COD	0,58	1,27
NH_4^+	1,43	0,81

A: Τυπικό σφάλμα ασυμμετρίας = 0,347 και B: Τυπικό σφάλμα κύρτωσης = 0,681

Δεδομένου ότι η παραγοντική ανάλυση, που βασίζεται στην ανάλυση κύριων συνιστωσών (PCA/FA), χρησιμοποιεί τον πίνακα συσχετίσεων ή συνδιακυμάνσεων, η μη – κανονικότητα των δεδομένων και η παρουσία ακραίων τιμών ενδέχεται να επηρεάσουν τα αποτελέσματα της (Reimann και Filzmoser, 2000; Baxter, 1995; Baxter και Freestone, 2006). Κατά συνέπεια, χρησιμοποιήθηκαν δύο διαφορετικές προσεγγίσεις για την PCA/FA. Στην πρώτη προσέγγιση, χρησιμοποιήθηκαν οι μετασχηματισμένες σε δεκαδικούς λογαρίθμους τιμές των παραμέτρων TU, COD και NH_4^+ και οι αρχικές τιμές των υπόλοιπων μεταβλητών. Στην δεύτερη προσέγγιση, προκειμένου να επιβεβαιωθούν τα αποτελέσματα της πρώτης προσέγγισης, χρησιμοποιήθηκε για όλες τις μεταβλητές ο μετασχηματισμός τους σε τάξεις μεγέθους (rank transformation) (Baxter, 1995). Ο μετασχηματισμός των μεταβλητών στις τάξεις μεγέθους των τιμών τους μπορεί να εφαρμοστεί σε όλες τις κατανομές με εξίσου καλή επίδοση. Παρόλα αυτά η χρήση του δεν είναι ευρεία διαδεδομένη. Η τεχνική αυτή προτάθηκε αρχικά για τη διαχωριστική ανάλυση από τους Conover και Iman (1980). Η τεχνική του μετασχηματισμού των μεταβλητών στις τάξεις μεγέθους των τείνει να ελαχιστοποιήσει το πρόβλημα των ακραίων τιμών και ως εκ τούτου την αιτία της μη-κανονικότητας. Ο μετασχηματισμός σε τάξεις μεγέθους είναι μία δεικτική τεχνική που συνδυάζει τη γενική δυνατότητα εφαρμογής των μη-παραμετρικών μεθόδων με την ισχύ των παραμετρικών στατιστικών αναλύσεων (Conover και Iman 1981; Helsel, 1987).

Για τη μελέτη των συσχετίσεων μεταξύ των μεταβλητών, χρησιμοποιήθηκαν οι συντελεστές συσχέτισης Spearman rho, οι οποίοι υπολογίζονται βάσει των τάξεων μεγέθους των τιμών των δεδομένων παρά από τις ίδιες τις τιμές τους. Συνεπώς, είναι λιγότερο ευαίσθητοι από τους συντελεστές συσχέτισης Pearson στην ύπαρξη ακραίων τιμών (Kannel και λοιποί, 2007). Επιπλέον, οι συντελεστές συσχέτισης Spearman ξεπερνούν το πρόβλημα της μη-κανονικότητας των δεδομένων. Στην περίπτωση ίσων τιμών χρησιμοποιήθηκε η μέθοδος της μέσης τάξης μεγέθους (Zar, 1999). Από τον Πίνακα 10.4, γίνεται εμφανές ότι υπάρχουν αρκετές στατιστικά σημαντικές συσχετίσεις μεταξύ των μεταβλητών.

Πίνακας 10.4. Συντελεστές συσχέτισης Spearman's rho για τις παραμέτρους του ταμιευτήρα Μαραθώνα

	EC	TDS	TU	pH	TA	TH	Cl ⁻	SO ₄ ²⁻	Na ⁺	Ca ²⁺	Mg ²⁺	COD	DO	NH ₄ ⁺
EC	1													
TDS	0,657**	1												
TU	-0,127	-0,107	1											
pH	-0,174	0,080	-0,070	1										
TA	0,763**	0,485**	-0,024	-0,475**	1									
TH	0,864**	0,643**	-0,114	-0,026	0,719**	1								
Cl ⁻	0,850**	0,612**	-0,216	0,057	0,499**	0,721**	1							
SO ₄ ²⁻	0,531**	0,502**	-0,249	0,492**	0,059	0,549**	0,610**	1						
Na ⁺	0,797**	0,658**	-0,244	0,177	0,410**	0,687**	0,968**	0,708**	1					
Ca ²⁺	0,887**	0,672**	-0,198	-0,159	0,735**	0,810**	0,788**	0,523**	0,781**	1				
Mg ²⁺	0,452**	0,442**	-0,154	0,452**	0,092	0,491**	0,486**	0,835**	0,568**	0,338*	1			
COD	0,011	-0,241	0,373**	-0,278	0,107	-0,050	-0,082	-0,237	-0,111	-0,079	-0,126	1		
DO	-0,175	0,036	-0,012	0,846**	-0,371*	0,000	0,005	0,361*	0,119	-0,116	0,215	-0,232	1	
NH ₄ ⁺	-0,242	-0,385**	0,135	-0,358*	0,117	-0,202	-0,290*	-0,669**	-0,377**	-0,190	-0,576**	0,259	-0,204	1
NO ₃ ⁻	0,633**	0,504**	-0,216	0,385**	0,170	0,556**	0,747**	0,762**	0,818**	0,589**	0,595**	-0,205	0,357*	-0,481**

** Στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,01$ (αμφίπλευρος έλεγχος)

* Στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,05$ (αμφίπλευρος έλεγχος)

10.4.2. Ανάλυση σε κύριες συνιστώσες/Παραγοντική ανάλυση

Στην πρώτη προσέγγιση για τη διεξαγωγή PCA/FA και βάσει του γραφήματος των ιδιοτιμών (scree-plot) καθώς και του κριτηρίου Kaiser των ιδιοτιμών εξήχθησαν τρεις παράγοντες, οι οποίοι επεξηγούν το 75,0% της συνολικής διακύμανσης των δεδομένων. Οι συμμετοχικότητες ευρέθησαν να είναι υψηλές (μεγαλύτερες από 0,50) για όλες τις μεταβλητές (Πίνακας 10.5). Επομένως, το μοντέλο των τριών παραγόντων επεξηγεί ικανοποιητικά τη μεταβλητότητα όλων των μεταβλητών και μπορεί να χρησιμοποιηθεί για να υποδείξει τις κύριες υδροχημικές διαδικασίες που ελέγχουν τη σύσταση του νερού, χωρίς την απώλεια κάποιων σημαντικών χαρακτηριστικών. Για την περιστροφή των αξόνων χρησιμοποιήθηκε η μέθοδος Varimax με Kaiser Normalization.

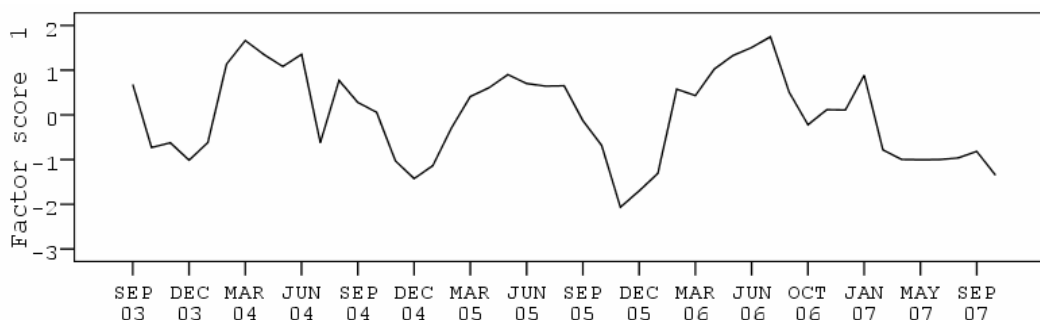
Πίνακας 10.5. 1^η προσέγγιση παραγοντικής ανάλυσης παραμέτρων του ταμιευτήρα Μαραθώνα (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization)

Παράμετροι	Παράγοντας			Συμμετοχικότητες
	1	2	3	
EC	0,97	-0,12	0,00	0,95
TDS	0,68	0,26	-0,18	0,56
TU	-0,11	0,03	0,79	0,64
pH	-0,09	0,87	-0,03	0,76
TA	0,68	-0,62	0,05	0,85
TH	0,90	-0,02	0,02	0,81
Cl ⁻	0,89	0,11	-0,07	0,82
SO ₄ ²⁻	0,30	0,84	-0,24	0,85
Na ⁺	0,88	0,27	-0,09	0,87
Ca ²⁺	0,92	-0,12	-0,10	0,87
Mg ²⁺	0,49	0,54	-0,04	0,53
COD	0,00	-0,22	0,78	0,66
DO	-0,07	0,77	0,09	0,61
NH ₄ ⁺	-0,16	-0,82	0,24	0,75
NO ₃ ⁻	0,67	0,52	0,00	0,72
% της Διακύμανσης	39,5	26,0	9,5	

Οι τιμές με έντονους χαρακτήρες υποδεικνύουν ισχυρή συσχέτιση μεταξύ μιας μεταβλητής και ενός παράγοντα.

Ο πρώτος παράγοντας επεξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης (39,5%) και παρουσιάζει υψηλές θετικές φορτίσεις στις μεταβλητές EC, TH, Cl⁻, Na⁺ και Ca²⁺, ενώ παρουσιάζει μέτριες θετικές φορτίσεις στις μεταβλητές TDS, TA και NO₃⁻ (Πίνακας 3). Αυτός ο παράγοντας μπορεί να θεωρηθεί ως παράγοντας «ιοντικής δύναμης του νερού». Ο δεύτερος παράγοντας επεξηγεί το 26,0% της συνολικής διακύμανσης και εμφανίζεται με τη μορφή διπόλου παρουσιάζοντας υψηλές θετικές φορτίσεις στις μεταβλητές pH, DO και SO₄²⁻ και υψηλή αρνητική φόρτιση στην μεταβλητή NH₄⁺. Η στενή σχέση μεταξύ pH και DO έχει παρατηρηθεί και σε άλλες μελέτες επιφανειακών υδάτων όπως για παράδειγμα αυτές των Paratheodorou και λοιπών (2006), Romero και λοιπών (2002) και Perona και λοιπών (1999). Οι Wayland και λοιποί (2003) θεωρούν ότι υπάρχουν τόσες πολλές διαδικασίες που επηρεάζουν τη σχέση μεταξύ του pH και άλλων φυσικο-χημικών παραμέτρων που ακόμα και η πολυμεταβλητή προσέγγιση αδυνατεί να τις αναγνωρίσει. Οι Paratheodorou και λοιποί (2006), σε μελέτη για τη λίμνη Παμβώτιδα, ερμήνευσαν τον παράγοντα ως παράγοντα «οξύτητας». Οι Perona και λοιποί (1999) ερμήνευσαν τη σχέση μεταξύ pH και DO ως έναν «φωτοσυνθετικό» παράγοντα. Παράλληλα, δεδομένου ότι η θέση δειγματοληψίας αφορούσε στα κατώτερα στρώματα του ταμιευτήρα, ανοξικές καταστάσεις λόγω μείωσης της συγκέντρωσης του DO, ιδιαίτερα κατά τους φθινοπωρινούς μήνες, επιφέρουν αύξηση της συγκέντρωσης NH₄⁺ και μείωση στην συγκέντρωση SO₄²⁻. Ο τρίτος παράγοντας επεξηγεί το 9,5% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση στις μεταβλητές TU και COD. Επομένως, ο συγκεκριμένος παράγοντας θα μπορούσε να ερμηνευθεί ως παράγοντας «αιωρούμενης οργανικής ύλης».

Το διάγραμμα της χρονολογικής σειράς των σκορ του πρώτου παράγοντα (Σχήμα 10.2) δεν παρουσιάζει ένα απολύτως ευκρινές πρότυπο για τη χρονική διακύμανση.



Σχήμα 10.2. Χρονική διακύμανση των σκορ του πρώτου παράγοντα για τον Μαραθώνα.

Στην δεύτερη προσέγγιση για τη διεξαγωγή PCA/FA (των μετασχηματισμένων σε τάξεις μεγέθους παραμέτρων) και βάσει του κριτηρίου Kaiser των ιδιοτιμών εξήχθησαν επίσης τρεις παράγοντες, οι οποίοι επεξηγούν το 75,6% της συνολικής διακύμανσης των δεδομένων και παρουσιάζουν μόνο μικροδιαφορές σε σχέση με την 1^η προσέγγιση. Οι συμμετοχικότητες ευρέθησαν να είναι υψηλές (μεγαλύτερες από 0,50) σχεδόν για όλες τις μεταβλητές εκτός από το αμμώνιο (NH_4^+) (Πίνακας 10.6).

Ο πρώτος παράγοντας επεξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης (42,6%) και παρουσιάζει υψηλές θετικές φορτίσεις στις μεταβλητές EC, TA, TH Cl^- , Na^+ και Ca^{2+} , ενώ παρουσιάζει μέτριες θετικές φορτίσεις στις μεταβλητές TDS και NO_3^- (Πίνακας 10.6). Αυτός ο παράγοντας μπορεί να θεωρηθεί ως παράγοντας «ιοντικής δύναμης του νερού». Ο δεύτερος παράγοντας επεξηγεί το 23,1% της συνολικής διακύμανσης και παρουσιάζει υψηλή θετική φόρτιση στις μεταβλητές μεταβλητές pH και DO και μέτρια θετική φόρτιση στη μεταβλητή SO_4^{2-} και μέτρια αρνητική φόρτιση στην NH_4^+ . Ο τρίτος παράγοντας επεξηγεί το 9,8% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση στις μεταβλητές TU και COD (παράγοντας «αιωρούμενης οργανικής ύλης»).

Πίνακας 10.6. Παραγοντική ανάλυση των μετασχηματισμένων σε τάξεις μεγέθους παραμέτρων του ταμιευτήρα Μαραθώνα (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization).

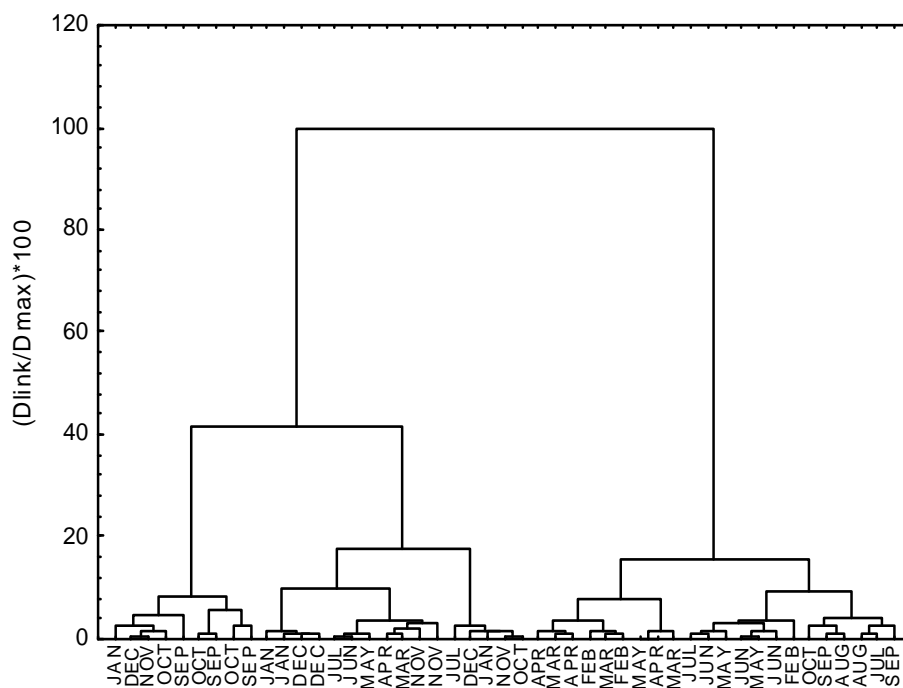
Παράμετροι (τάξεις μεγέθους)	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Συμμετοχικότητες
EC	0,97	-0,04	0,01	0,94
TDS	0,73	0,16	-0,16	0,59
TU	-0,14	0,00	0,83	0,70
pH	-0,15	0,91	-0,07	0,85
TA	0,76	-0,48	0,08	0,81
TH	0,89	0,05	0,01	0,79
Cl ⁻	0,89	0,18	-0,09	0,83
SO ₄ ²⁻	0,58	0,70	-0,16	0,86
Na ⁺	0,86	0,32	-0,11	0,86
Ca ²⁺	0,93	-0,07	-0,11	0,87
Mg ²⁺	0,50	0,63	-0,04	0,65
COD	0,01	-0,22	0,80	0,69
DO	-0,16	0,79	-0,01	0,65
NH ₄ ⁺	-0,27	-0,60	0,20	0,47
NO ₃ ⁻	0,66	0,56	-0,13	0,77
% της Διακύμανσης	42,6	23,1	9,8	

Οι τιμές με έντονους χαρακτήρες υποδεικνύουν ισχυρή συσχέτιση μεταξύ μιας μεταβλητής και ενός παράγοντα

10.4.3. Ανάλυση κατά συστάδες

Η χρονική μεταβλητότητα της ποιότητας νερού στον ταμιευτήρα Μαραθώνα, προσδιορίστηκε μέσω της ανάλυσης κατά συστάδες.

Το δεντρόγραμμα (Σχήμα 10.3) παρέχει μια οπτική περίληψη των διαδικασιών ομαδοποίησης, παρουσιάζοντας μια απεικόνιση των ομάδων και την εγγύτητά τους, με μια δραματική μείωση των διαστάσεων των αρχικών δεδομένων. Χρησιμοποιήθηκε η διαδικασία αναπροσαρμογής των αποστάσεων σύνδεσης των ομάδων ως $100 \cdot D_{\text{link}} / D_{\text{max}}$, όπου D_{link} είναι η απόσταση σύνδεσης για μια ειδική περίπτωση και D_{max} η μέγιστη απόσταση. Με αυτόν τον τρόπο τυποποιείται η απόσταση σύνδεσης των ομάδων η οποία αντιπροσωπεύεται στον άξονα y (Wunderlin και λοιποί, 2001; Simeonov και λοιποί, 2003; Singh και λοιποί, 2004; Shrestha και Kazama, 2007).



Σχήμα 10.3. Ιεραρχικό δεντρόγραμμα για τις περιόδους παρακολούθησης του ταμιευτήρα Μαραθώνα.

Η ανάλυση κατά συστάδες κατέληξε σε δύο κύριες ομάδες. Η πρώτη ομάδα περιλάμβανε τις περισσότερες παρατηρήσεις για τους μήνες Σεπτέμβριο, Οκτώβριο, Νοέμβριο, Δεκέμβριο και Ιανουάριο (κλάση 1: Σεπτέμβριος – Ιανουάριος). Η δεύτερη ομάδα περιλάμβανε τις περισσότερες παρατηρήσεις για τους υπόλοιπους μήνες του έτους (κλάση 2: Φεβρουάριος έως Αύγουστος). Λόγω της ιδιομορφίας του ταμιευτήρα Μαραθώνα αφενός να τροφοδοτείται με νερό από τον ταμιευτήρα Μόρνου (το συγκεκριμένο χρονικό διάστημα από το οποίο προέρχονται τα δεδομένα) και αφετέρου να εισρέουν σε αυτόν ρέματα και χείμαροι με μη συνεχή ροή, δεν είναι εύκολη μία άμεση σύνδεση των κλάσεων με τις εποχές του χρόνου και το υδρολογικό έτος. Παρόλα αυτά, η προσπάθεια ομαδοποίησης των δειγμάτων νερού παρουσιάζει ενδιαφέρον καθώς μέσω αυτής απεικονίζεται η διακύμανση των φυσικοχημικών ιδιοτήτων του νερού του ταμιευτήρα σε ένα διάστημα τεσσάρων ετών.

10.4.4. Διαχωριστική ανάλυση

Η παρατηρούμενη χρονική μεταβλητότητα των παραμέτρων ποιότητας του νερού αποτιμήθηκε με τη χρησιμοποίηση της διαχωριστικής ανάλυσης. Η συγκεκριμένη

τεχνική διαχωριστικής ανάλυσης, προσδιορίζει τις μεταβλητές ή τους δείκτες που επιτρέπουν την καλύτερη διάκριση μεταξύ δύο ή περισσότερων προκαθορισμένων ομάδων δειγμάτων (Hair και λοιποί, 1998; Johnson και Wichern, 1998).

Η χρονική διαχωριστική ανάλυση εκτελέστηκε στο σύνολο των δεδομένων τα οποία στη συνέχεια ομαδοποιήθηκαν σε δύο περιόδους (Σεπτέμβριος-Ιανουάριος και Φεβρουάριος-Αύγουστος), βάσει των αποτελεσμάτων της χρονικής ανάλυσης κατά συστάδες. Η ομοιογένεια των διακυμάνσεων - συνδιακυμάνσεων των δύο ομάδων ελέγχθηκε με τον έλεγχο Box's M, ο οποίος ελέγχει τη μηδενική υπόθεση ότι οι πίνακες διακύμανσης-συνδιακύμανσης των ομάδων είναι ίσοι. Η p-τιμή του ελέγχου υπολογίστηκε ίση με 0,004.

Η βηματική μέθοδος διαχωριστικής ανάλυσης οδήγησε σε μια στατιστικά σημαντική διαχωριστική συνάρτηση (DF) (Wilks' lambda=0,247, p-τιμή <0,001) και χρησιμοποίησε τις μεταβλητές NO_3^- , Ca^{2+} , TH, COD και Mg^{2+} . Οι τιμές ανοχής (tolerance) για κάθε μία από τις ανεξάρτητες μεταβλητές ήταν πολύ μεγαλύτερες από 0,10, γεγονός που υποδεικνύει την έλλειψη πολυσυγγραμμικότητας των μεταβλητών. Οι συντελεστές συσχέτισης (r_{DF}) των παραπάνω μεταβλητών με τη διαχωριστική συνάρτηση παρουσιάζονται στον Πίνακα 10.7 που ακολουθεί.

Πίνακας 10.7. Συντελεστές συσχέτισης μεταβλητής-διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για τον Μαραθώνα

Μεταβλητή	r_{DF}
NO_3^-	0,531
Ca^{2+}	0,411
TH	0,205
COD	-0,182
Mg^{2+}	0,081

Παρατηρείται λοιπόν μέτρια θετική συσχέτιση της διαχωριστικής συνάρτησης με τη μεταβλητή NO_3^- , ασθενής θετική συσχέτιση με την Ca^{2+} και εξαιρετικά ασθενής με τις TH, COD και Mg^{2+} . Το κεντροειδές της DF για την ομάδα 1 (Σεπτέμβριος-Ιανουάριος) υπολογίστηκε ίσο με -1,820 ενώ το κεντροειδές για την ομάδα 2 (Φεβρουάριος-Αύγουστος) υπολογίστηκε ίσο με 1,602. Επομένως, συμπεραίνεται ότι οι συγκεντρώσεις NO_3^- και Ca^{2+} ήταν χαμηλότερες την περίοδο «Σεπτέμβριος έως Ιανουάριος» συγκριτικά με την περίοδο «Φεβρουάριος έως Αύγουστος». Η βηματική

διαχωριστική ανάλυση απέφερε 97,9% ορθή ταξινόμηση του συνόλου των δεδομένων. Η διασταυρούμενη επικύρωση του μοντέλου οδήγησε σε 95,7% ορθή ταξινόμηση (Πίνακας 10.8). Ο πίνακας ταξινόμησης των ορθών και των ανακριβών εκτιμήσεων είναι απαραίτητος για την αξιολόγηση της αποτελεσματικότητας της διαδικασίας της διαχωριστικής ανάλυσης. Το υψηλό ποσοστό ορθής ταξινόμησης αποτελεί ένδειξη της καλής ακρίβειας του μοντέλου.

Πίνακας 10.8. Πίνακας ταξινόμησης για τη βηματική DA της χρονικής διακύμανσης του νερού στον Μαραθώνα

Δείγμα	Παρατηρούμενη κατηγορία	Εκτιμηθείσα κατηγορία		% Ορθής ταξινόμησης
		Σεπ-Ιαν	Φεβ-Αυγ	
Πρωτότυπο ^α	Σεπ-Ιαν	21	1	95,5
	Φεβ-Αυγ	0	25	100,0
Διασταυρούμενης επικύρωσης ^β	Σεπ-Ιαν	21	1	95,5
	Φεβ-Αυγ	1	24	96,0

α: 97,9% ορθή ταξινόμηση των πρωτότυπων δεδομένων

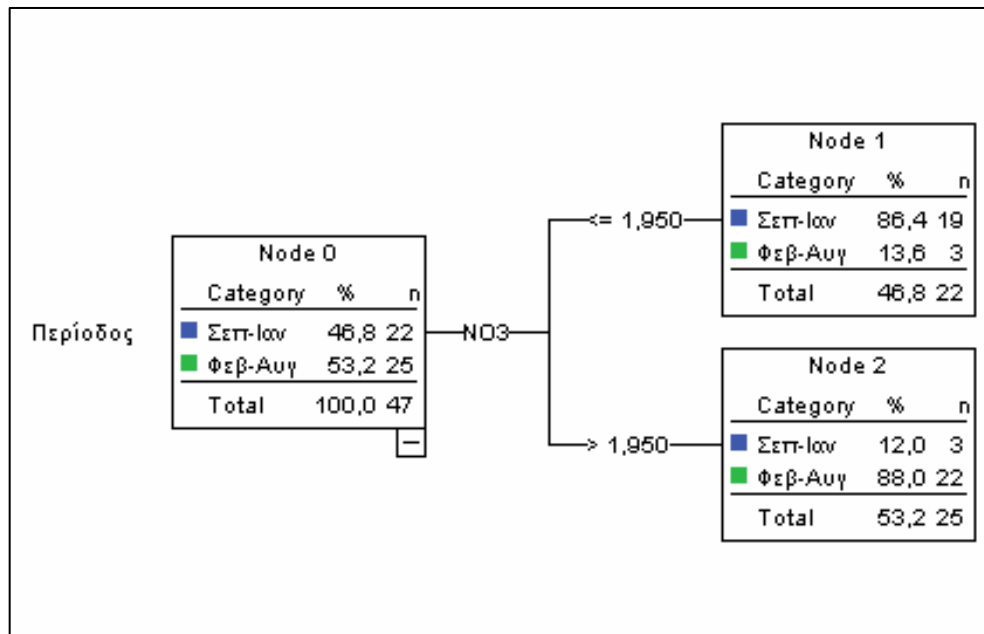
β: 95,7% ορθή ταξινόμηση των δεδομένων διασταυρούμενης επικύρωσης (cross-validation)

10.4.5. Δέντρα ταξινόμησης

Με τη διαδικασία των δέντρων ταξινόμησης και παλινδρόμησης (CART) αποτιμήθηκαν όλες οι μεταβλητές ώστε να καθοριστούν αυτές οι οποίες παρέχουν τον καλύτερο διαχωρισμό (δηλαδή, οι μεταβλητές που διαχωρίζουν τον «γονικό» κόμβο στους δύο αγνότερους κόμβους «τέκνα») χρησιμοποιώντας τον δείκτη μη-καθαρότητας του Gini (Breiman και λοιποί, 1984; Loh and Shih, 1997). Κατά τη διαδικασία της περικοπής χρησιμοποιήθηκε το κριτήριο «πολυπλοκότητας κόστους» (cost complexity), για να παραχθεί μία ακολουθία με δέντρα τα οποία είναι διαδοχικά μικρότερα μέχρι του σημείου στο οποίο παρήχθη ένα δέντρο με μόνο τον ριζικό κόμβο. Η επιλογή του βέλτιστου δέντρου στηρίχθηκε στα αποτελέσματα της 10-πλής διασταυρούμενης επικύρωσης (10-fold cross-validation) και τον κανόνα του ενός τυπικού σφάλματος (1SE rule).

Ο αλγόριθμος CART κατέληξε σε ένα δέντρο με ένα επίπεδο ανάπτυξης κάτω από τον αρχικό κόμβο και υπέδειξε ότι η κύρια παράμετρος που χρησιμοποιήθηκε για τον

διαχωρισμό των δειγμάτων στις δυο χρονικές περιόδους ήταν τα NO_3^- (Σχήμα 10.4). Το ποσοστό της συνολικής ορθής ταξινόμησης των δειγμάτων υπολογίστηκε 87,2% . Η 10-πλή διασταυρούμενη επικύρωση του μοντέλου είχε ποσοστό ορθής ταξινόμησης, ίσο με 83,0%.



Σχήμα 10.4. Δέντρο ταξινόμησης για τη χρονική διακύμανση των δεδομένων του Μαραθώνα.

10.5. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΧΟΛΙΑ ΓΙΑ ΤΗ ΛΙΜΝΗ ΥΛΙΚΗ

10.5.1 Προκαταρκτική εξέταση των παραμέτρων και διαχείριση των δεδομένων

Οι τιμές των δεδομένων για τα φωσφορικά (PO_4^{3-}) ήταν σε υψηλό ποσοστό περικεκομμένες. Συγκεκριμένα, το ποσοστό των δειγμάτων για τα οποία οι τιμές των PO_4^{3-} ήταν κάτω του ορίου ανίχνευσης, ήταν 95,6%. Συνεπώς, η παράμετρος PO_4^{3-} εξαιρέθηκε από τις περαιτέρω αναλύσεις. Στον πίνακα 10.9 παρουσιάζονται τα στατιστικά περιγραφικά μέτρα για κάθε μία παράμετρο. Σε τέσσερις από τις παραμέτρους (TU , COD , NH_4^+ και NO_3^-) παρατηρήθηκε θετική ασυμμετρία. Για αυτόν το λόγο, χρησιμοποιήθηκε δεκαδικός λογαριθμικός μετασχηματισμός τους (\log_{10}) ώστε να βελτιωθούν τα χαρακτηριστικά αυτών των μεταβλητών (Paratheodorou και λοιποί, 2006). Μετά τον δεκαδικό λογαριθμικό μετασχηματισμό

τους οι μεταβλητές αυτές, όπως φαίνεται στον Πίνακα 10.10, ακολουθούσαν προσεγγιστικά την κανονική κατανομή.

Πίνακας 10.9. Περιγραφικά στατιστικά μέτρα των παραμέτρων για τη λίμνη Υλίκη

Παράμετρος	Μέσος	Τυπική απόκλιση	Ασυμμετρία ^A	Κύρτωση ^B
EC	404,79	49,53	-0,30	-0,96
TDS	235,47	38,54	-0,14	-0,14
TU	2,85	1,66	1,64	3,39
pH	8,47	0,30	0,08	0,64
TA	159,15	28,71	-0,25	-1,16
TH	192,11	31,20	-0,15	-1,20
Cl ⁻	15,59	1,12	-0,30	-0,15
SO ₄ ²⁺	38,87	2,23	-0,26	-0,31
Na ⁺	12,74	0,77	0,07	-0,63
Ca ²⁺	40,56	12,29	-0,12	-1,23
Mg ²⁺	22,98	0,68	0,24	-0,49
COD	1,86	0,55	0,67	0,39
DO	8,76	1,05	0,11	-0,42
NH ₄ ⁺	0,06	0,05	1,80	2,82
NO ₃ ⁻	0,74	0,98	1,41	0,79

A: Τυπικό σφάλμα ασυμμετρίας = 0,327 και B: Τυπικό σφάλμα κύρτωσης = 0,644

Πίνακας 10.10. Μέτρα ασυμμετρίας και κύρτωσης των log₁₀ μετασχηματισμένων μεταβλητών για τη λίμνη Υλίκη

Παράμετρος (σε log ₁₀)	Ασυμμετρία ^A	Κύρτωση ^B
TU	0,212	-0,369
COD	-0,157	-0,081
NH ₄ ⁺	0,235	-0,350
NO ₃ ⁻	-0,200	-1,016

A: Τυπικό σφάλμα ασυμμετρίας = 0,327 και B: Τυπικό σφάλμα κύρτωσης = 0,644

Στον Πίνακα 10.11 παρουσιάζονται τα αποτελέσματα της ανάλυσης συσχέτισεων Spearman rho μεταξύ των μεταβλητών. Παρατηρείται ότι υπάρχουν αρκετές στατιστικά σημαντικές συσχετίσεις μεταξύ των μεταβλητών.

Πίνακας 10.11. Συντελεστές συσχέτισης Spearman's rho για τις παραμέτρους της λίμνης Υλίκης

	EC	TDS	TU	pH	TA	TH	Cl ⁻	SO ₄ ²⁻	Na ⁺	Ca ²⁺	Mg ²⁺	COD	DO	NH ₄ ⁺
EC	1													
TDS	0,817**	1												
TU	-0,272*	-0,253	1											
pH	-0,704**	-0,501**	0,461**	1										
TA	0,978**	0,832**	-0,235	-0,648**	1									
TH	0,953**	0,790**	-0,246	-0,673**	0,958**	1								
Cl ⁻	-0,311*	-0,151	-0,25	0,145	-0,363**	-0,391**	1							
SO ₄ ²⁻	0,200	0,261	0,203	0,050	0,173	0,168	-0,316*	1						
Na ⁺	-0,156	-0,097	-0,203	-0,061	-0,199	-0,215	0,832**	-0,350*	1					
Ca ²⁺	0,978**	0,790**	-0,213	-0,672**	0,979**	0,954**	-0,434**	0,260	-0,259	1				
Mg ²⁺	-0,021	0,017	0,098	0,059	-0,060	-0,132	0,472**	0,103	0,632**	-0,103	1			
COD	-0,539**	-0,531**	0,554**	0,481**	-0,540**	-0,548**	0,070	0,003	0,001	-0,525**	0,144	1		
DO	0,561**	0,505**	0,031	-0,237	0,613**	0,561**	-0,362**	0,097	-0,317*	0,603**	-0,158	-0,276*	1	
NH ₄ ⁺	0,476**	0,330*	-0,237	-0,526**	0,441**	0,468**	-0,368**	0,014	-0,221	0,480**	-0,282*	-0,359**	0,300*	1
NO ₃ ⁻	0,738**	0,588**	0,035	-0,359**	0,732**	0,727**	-0,550**	0,441**	-0,461**	0,791**	-0,172	-0,252	0,509**	0,333*

** Στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,01$ (αμφίπλευρος έλεγχος)

* Στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,05$ (αμφίπλευρος έλεγχος)

10.5.2. Ανάλυση σε κύριες συνιστώσες/Παραγοντική ανάλυση

Προκειμένου να αποφασιστεί ο αριθμός των εξαγόμενων παραγόντων, χρησιμοποιήθηκαν το γράφημα των ιδιοτιμών (scree-plot) καθώς και το κριτήριο Kaiser των ιδιοτιμών. Το αποτέλεσμα ήταν η εξαγωγή τεσσάρων παραγόντων οι οποίοι επεξηγούν το 80,4% της συνολικής διακύμανσης των δεδομένων. Οι συμμετοχικότητες ευρέθησαν να είναι υψηλές (μεγαλύτερες από 0,50) σχεδόν για όλες τις μεταβλητές εκτός από το αμμώνιο (NH_4^+) (Πίνακας 10.12). Επομένως, το μοντέλο των τεσσάρων παραγόντων επεξηγεί ικανοποιητικά τη μεταβλητότητα σχεδόν όλων των μεταβλητών.

Στον Πίνακα 10.12 παρουσιάζονται τα αποτελέσματα της PCA/FA στην οποία ως μέθοδος περιστροφής χρησιμοποιήθηκε η Varimax με Kaiser Normalization. Ο πρώτος παράγοντας επεξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης (39,2%) και παρουσιάζει υψηλές θετικές φορτίσεις στις μεταβλητές EC, TDS, TA, TH και Ca^{2+} , μέτριες θετικές φορτίσεις στις μεταβλητές NO_3^- και DO ενώ παρουσιάζει μέτρια αρνητική φόρτιση στο pH (Πίνακας 10.12). Αυτός ο παράγοντας μπορεί να θεωρηθεί ως παράγοντας «σκληρότητας του νερού – αλατότητας» και βασίζεται κυρίως στην προέλευση του φυσικού νερού (Simeonov και λοιποί, 2001). Ο δεύτερος παράγοντας επεξηγεί το 17,3% της συνολικής διακύμανσης και παρουσιάζει υψηλές θετικές φορτίσεις στις μεταβλητές Cl^- , Na^+ και Mg^{2+} . Ο γραμμομοριακός λόγος του Na^+ προς το Cl^- υπολογίστηκε ότι είναι μικρότερος από 1. Αυτό θα μπορούσε να είναι το αποτέλεσμα ενός αριθμού διεργασιών όπως είναι η ανταλλαγή μεταξύ ιόντων νατρίου και καλίου (Paratheodorou και λοιποί, 2006). Ο τρίτος παράγοντας επεξηγεί το 14,9% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση στις μεταβλητές TU και COD. Επομένως, ο συγκεκριμένος παράγοντας θα μπορούσε να ερμηνευθεί ως παράγοντας «αιωρούμενης οργανικής ύλης». Ο τέταρτος παράγοντας επεξηγεί το 8,9% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση μόνο στα θειϊκά (SO_4^{2-}). Αξίζει να σημειωθεί ότι η ερμηνεία των παραγόντων που φορτίζουν σε μία μόνο μεταβλητή είναι δύσκολη. Θα μπορούσε να υποδηλώνει ότι η συγκεκριμένη μεταβλητή μεταβάλλεται ανεξάρτητα από τις άλλες μεταβλητές (Clow και λοιποί, 1996). Οι Suk και Lee (1999) πρότειναν ότι ένας παράγοντας, ο οποίος κυριαρχείται από τα θειϊκά, είναι πολύ δύσκολο να επεξηγηθεί με κάποια

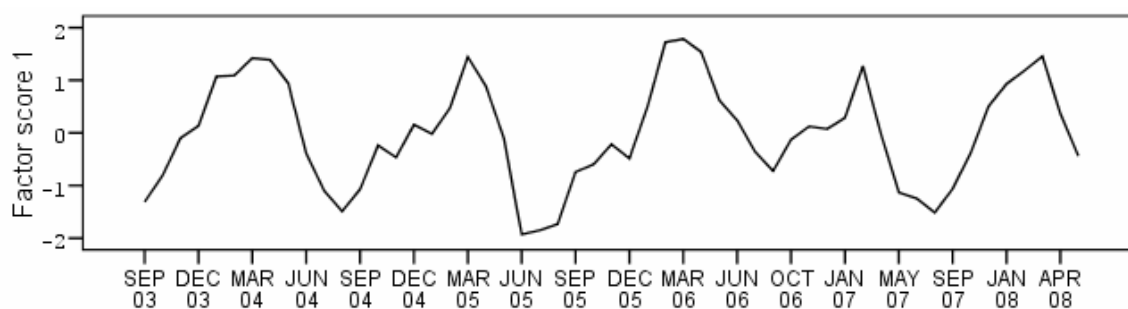
συγκεκριμένη χημική αντίδραση και μπορεί να θεωρηθεί ως παράγοντας μιας πολλαπλής χημικής διαδικασίας.

Πίνακας 10.12. Παραγοντική ανάλυση των παραμέτρων της λίμνης Υλίκη (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization)

Παράμετροι	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Παράγοντας 4	Συμμετοχικότητες
EC	0,93	-0,06	-0,31	0,07	0,97
TDS	0,80	0,08	-0,23	0,21	0,75
pH	-0,59	-0,07	0,56	0,15	0,69
TA	0,94	-0,11	-0,26	0,04	0,97
TH	0,91	-0,15	-0,28	0,05	0,94
Cl ⁻	-0,36	0,79	-0,18	-0,22	0,83
SO ₄ ²⁻	0,14	-0,08	0,06	0,94	0,92
Na ⁺	-0,14	0,87	-0,18	-0,32	0,92
Ca ²⁺	0,93	-0,18	-0,23	0,11	0,97
Mg ²⁺	0,08	0,86	0,12	0,19	0,79
DO	0,69	-0,27	0,20	-0,06	0,60
TU	-0,06	-0,08	0,88	0,09	0,79
COD	-0,33	0,00	0,74	-0,07	0,66
NH ₄ ⁺	0,34	-0,45	-0,38	-0,15	0,48
NO ₃ ⁻	0,73	-0,34	0,03	0,36	0,77
% της Διακύμανσης	39,1	17,3	15,0	8,9	

Οι τιμές με έντονους χαρακτήρες υποδεικνύουν ισχυρή συσχέτιση μεταξύ μιας μεταβλητής και ενός παράγοντα

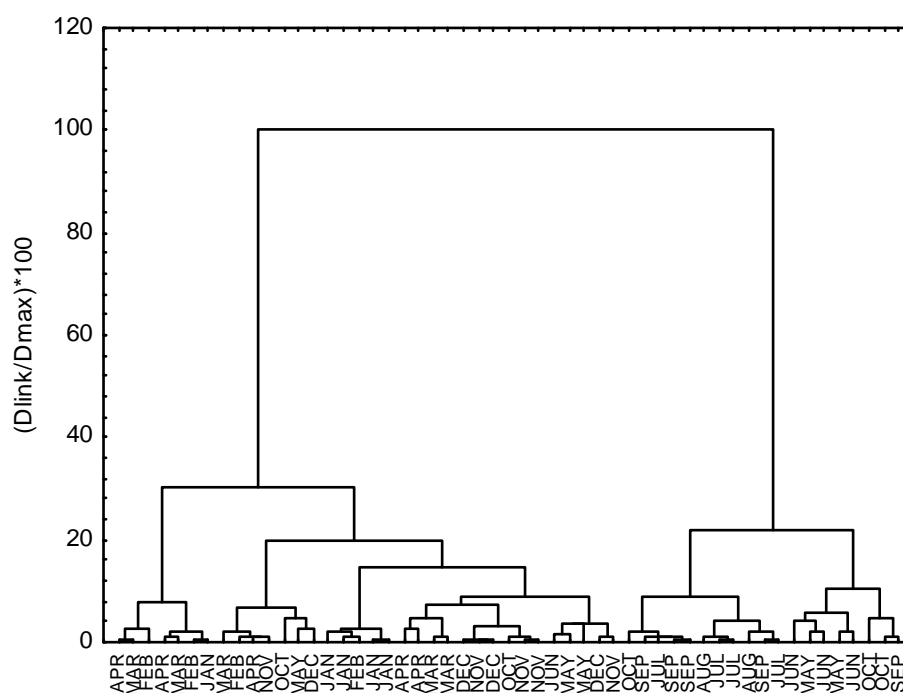
Καθώς ο πρώτος παράγοντας επεξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης των δεδομένων, θα μπορούσε να δώσει χρήσιμες πληροφορίες για το χρονικό πρότυπο της διακύμανσης. Το διάγραμμα της χρονολογικής σειράς των σκορ του πρώτου παράγοντα (Σχήμα 10.5) παρουσιάζει ένα σχεδόν ημιτονοειδές πρότυπο. Οι υψηλότερες τιμές παρατηρούνται από τον Νοέμβριο έως το Μάιο και οι χαμηλότερες από τον Ιούνιο έως τον Οκτώβριο.



Σχήμα 10.5. Χρονική διακύμανση των σκορ του πρώτου παράγοντα για την Υλίκη.

10.5.3 Ανάλυση κατά συστάδες

Στο Σχήμα 10.6 παρουσιάζεται το δεντρόγραμμα, το οποίο παρέχει μια οπτική περίληψη των διαδικασιών ομαδοποίησης για την λίμνη Υλίκη με την απεικόνιση των ομάδων και του τρόπου της μεταξύ τους σύνδεσης.



Σχήμα 10.6. Ιεραρχικό δεντρόγραμμα για τις περιόδους παρακολούθησης της Υλίκης

Η ανάλυση κατά συστάδες κατέληξε σε δύο κύριες ομάδες. Η πρώτη ομάδα περιλάμβανε τους μήνες Ιούνιο, Ιούλιο, Αύγουστο, Σεπτέμβριο και Οκτώβριο, οι

οποίοι προσεγγιστικά αντιστοιχούν στην ξηρή και θερμή περίοδο του έτους (Ιούνιος – Οκτώβριος). Η δεύτερη ομάδα περιλάμβανε τους υπόλοιπους μήνες του έτους (Νοέμβριο έως Μάιο) και αντιστοιχεί προσεγγιστικά στην υγρή και ψυχρή περίοδο του έτους.

10.5.4. Διαχωριστική ανάλυση

Η χρονική μεταβλητότητα των παραμέτρων ποιότητας του νερού αποτιμήθηκε με τη χρησιμοποίηση της διαχωριστικής ανάλυσης. Η χρονική διαχωριστική ανάλυση εκτελέστηκε στο σύνολο δεδομένων μετά την ομαδοποίηση τους σε δύο περιόδους/εποχές (Ιούνιος-Οκτώβριος και Νοέμβριος-Μάιος), βασιζόμενη στα αποτελέσματα της χρονικής ανάλυσης κατά συστάδες. Η ομοιογένεια των διακυμάνσεων - συνδιακυμάνσεων των δύο ομάδων ελέγχθηκε με τον έλεγχο Box's M, ο οποίος ελέγχει τη μηδενική υπόθεση ότι οι πίνακες διακύμανσης-συνδιακύμανσης των ομάδων είναι ίσοι. Ο έλεγχος δεν ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας $\alpha=0,05$ (p-τιμή=0,632).

Η βηματική μέθοδος διαχωριστικής ανάλυσης οδήγησε σε μια στατιστικά σημαντική διαχωριστική συνάρτηση (DF) (Wilks' lambda=0,283, p-τιμή <0,001) και χρησιμοποίησε τις μεταβλητές Ca^{2+} , DO και SO_4^{2-} . Οι τιμές ανοχής (tolerance) για κάθε μία από τις ανεξάρτητες μεταβλητές ήταν πολύ μεγαλύτερες από 0,10, γεγονός που υποδεικνύει την έλλειψη πολυσυγγραμμικότητας των μεταβλητών. Οι συντελεστές συσχέτισης (r_{DF}) των παραπάνω μεταβλητών με τη διαχωριστική συνάρτηση παρουσιάζονται στον Πίνακα 10.13 που ακολουθεί.

Πίνακας 10.13. Συντελεστές συσχέτισης μεταβλητής-διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για την Υλίκη

Μεταβλητή	r_{DF}
Ca^{2+}	0,724
DO	0,497
SO_4^{2-}	-0,082

Παρατηρείται λοιπόν ισχυρή/μέτρια θετική συσχέτιση της διαχωριστικής συνάρτησης με τις μεταβλητές Ca^{2+} και DO αντίστοιχα ενώ η συσχέτισή της με την μεταβλητή SO_4^{2-} είναι εξαιρετικά ασθενής. Το κεντροειδές της DF για την ομάδα 1 (Ιούνιος-

Οκτώβριος) υπολογίστηκε ίσο με -2,0 ενώ το κεντροειδές για την ομάδα 2 (Νοέμβριος-Μάιος) υπολογίστηκε ίσο με 1,2. Επομένως, συμπεραίνεται ότι οι συγκεντρώσεις Ca^{2+} και DO είναι χαμηλότερες την περίοδο «Ιούνιο έως Οκτώβριο» συγκριτικά με την περίοδο «Νοέμβριο έως Μάιο». Η βηματική διαχωριστική ανάλυση απέφερε 92,5% ορθή ταξινόμηση του συνόλου δεδομένων. Η διασταυρούμενη επικύρωση του μοντέλου οδήγησε επίσης σε 92,5% ορθή ταξινόμηση (Πίνακας 10.14). Το υψηλό ποσοστό ορθής ταξινόμησης αποτελεί ένδειξη της καλής ακρίβειας του μοντέλου.

Πίνακας 10.14. Πίνακας ταξινόμησης για τη βηματική DA της χρονικής διακύμανσης του νερού στην Υλίκη

Δείγμα	Παρατηρούμενη κατηγορία	Εκτιμηθείσα κατηγορία		% Ορθής ταξινόμησης
		Ιουν-Οκτ	Νοεμβρ-Μάης	
Πρωτότυπο ^α	Ιουν-Οκτ	19	1	95,0
	Νοεμβρ-Μάης	3	30	90,9
Διασταυρούμενης επικύρωσης ^β	Ιουν-Οκτ	19	1	95,0
	Νοεμβρ-Μάης	3	30	90,9

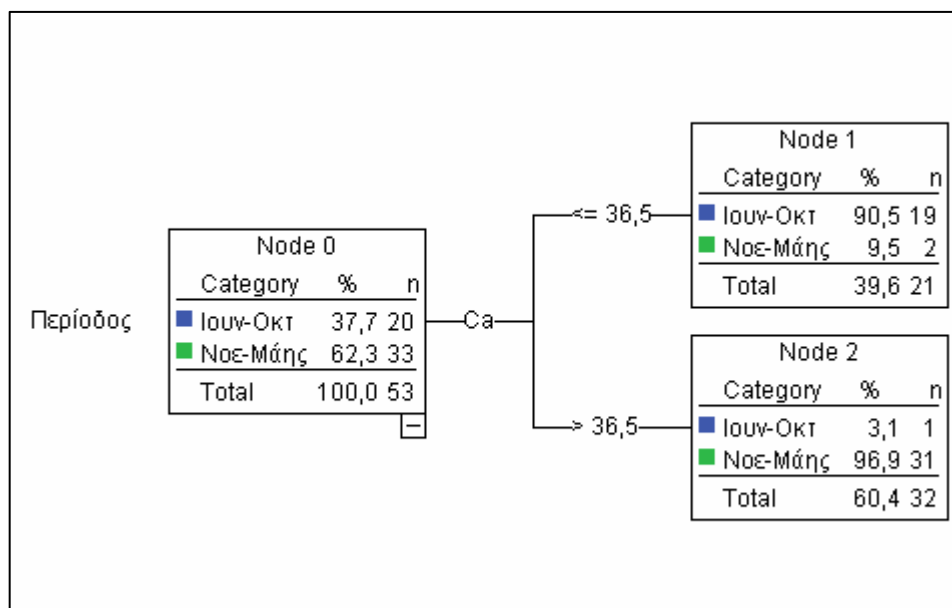
α: 92,5% ορθή ταξινόμηση των πρωτότυπων δεδομένων

β: 92,5% ορθή ταξινόμηση των δεδομένων διασταυρούμενης επικύρωσης

10.5.5. Δέντρα ταξινόμησης

Για την κατασκευή των δέντρων ταξινόμησης και παλινδρόμησης (CART), ακολουθήθηκε η ίδια ακριβώς διαδικασία που περιγράφηκε στην περίπτωση του ταμειυτήρα Μαράθωνα (δείκτης μη-καθαρότητας του Gini, περικοπή δέντρου και επιλογή βέλτιστου δέντρου βάσει της 10-πλής διασταυρούμενη επικύρωσης και τον κανόνα του ενός τυπικού σφάλματος).

Ο αλγόριθμος CART κατέληξε σε ένα δέντρο με ένα επίπεδο ανάπτυξης κάτω από τον αρχικό κόμβο και υπέδειξε ότι η κύρια παράμετρος που χρησιμοποιήθηκε για τον διαχωρισμό των δειγμάτων στις δυο χρονικές περιόδους ήταν το Ca^{2+} (Σχήμα 10.7). Το ποσοστό της συνολικής ορθής ταξινόμησης των δειγμάτων υπολογίστηκε σε 94,3% . Η 10-πλή διασταυρούμενη επικύρωση του μοντέλου είχε επίσης υψηλό ποσοστό ορθής ταξινόμησης, ίσο με 90,6%.



Σχήμα 10.7. Δέντρο ταξινόμησης για την χρονική διακύμανση των δεδομένων της Υλίκης (αλγόριθμος CART).

10.6. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΧΟΛΙΑ ΓΙΑ ΤΟΝ ΤΑΜΙΕΥΤΗΡΑ ΜΟΡΝΟΥ

10.6.1. Προκαταρκτική εξέταση των παραμέτρων και διαχείριση των δεδομένων

Οι τιμές των δεδομένων για τα φωσφορικά (PO_4^{3-}) ήταν σε υψηλό ποσοστό περικεκομμένες. Συγκεκριμένα, το ποσοστό των δειγμάτων για τα οποία οι τιμές των PO_4^{3-} ήταν κάτω του ορίου ανίχνευσης, ήταν 95,7%. Συνεπώς, η παράμετρος PO_4^{3-} εξαιρέθηκε από τις περαιτέρω αναλύσεις. Στον Πίνακα 10.15 παρουσιάζονται τα στατιστικά περιγραφικά μέτρα για κάθε μία παράμετρο. Σε πέντε από τις παραμέτρους (TU , Cl^- , Na^+ , NH_4^+ και NO_3^-) παρατηρήθηκε θετική ασυμμετρία. Για αυτόν το λόγο, χρησιμοποιήθηκε δεκαδικός λογαριθμικός μετασχηματισμός τους (\log_{10}) ώστε να βελτιωθούν τα χαρακτηριστικά αυτών των μεταβλητών (Paratheodorou και λοιποί, 2006). Μετά τον δεκαδικό λογαριθμικό μετασχηματισμό τους οι μεταβλητές αυτές, όπως φαίνεται στον Πίνακα 10.16, ακολουθούσαν προσεγγιστικά την κανονική κατανομή.

Πίνακας 10.15. Περιγραφικά στατιστικά μέτρα των παραμέτρων για τον ταμιευτήρα Μόρνου

Παράμετρος	Μέσος	Τυπική απόκλιση	Ασυμμετρία	Κύρτωση
EC	263,68	11,37	0,55	0,51
TDS	153,09	17,59	-0,58	0,59
TU	4,42	5,85	2,69	7,58
pH	8,32	0,10	0,16	-0,55
TA	123,49	6,43	0,49	-0,34
TH	136,19	7,60	0,22	-0,80
Cl ⁻	2,84	0,19	0,71	2,01
SO ₄ ²⁻	17,56	1,04	0,60	1,10
Na ⁺	4,31	0,17	0,79	0,27
Ca ²⁺	45,61	2,49	-0,44	-0,87
Mg ²⁺	5,42	0,24	0,51	-0,61
COD	0,70	0,21	0,38	-0,36
DO	8,87	1,04	-0,17	-1,25
NO ₃ ⁻	0,09	0,14	2,01	3,09
NH ₄ ⁺	0,02	0,01	0,79	0,63

A: Τυπικό σφάλμα ασυμμετρίας = 0,347 και B: Τυπικό σφάλμα κύρτωσης = 0,681

Πίνακας 10.16. Μέτρα ασυμμετρίας και κύρτωσης των log₁₀ μετασχηματισμένων μεταβλητών για τον ταμιευτήρα Μόρνου

Παράμετρος (σε log ₁₀)	Ασυμμετρία ^A	Κύρτωση ^B
TU	0,720	-0,133
Cl ⁻	0,406	1,412
Na ⁺	0,706	0,049
NO ₃ ⁻	0,211	-0,851
NH ₄ ⁺	-0,232	-0,730

A: Τυπικό σφάλμα ασυμμετρίας = 0,347 και B: Τυπικό σφάλμα κύρτωσης = 0,681

Στον Πίνακα 10.17 παρουσιάζονται τα αποτελέσματα της ανάλυσης συσχετίσεων Spearman rho μεταξύ των μεταβλητών. Παρατηρείται ότι υπάρχουν αρκετές στατιστικά σημαντικές συσχετίσεις μεταξύ των μεταβλητών.

Πίνακας 10.17. Συντελεστές συσχέτισης Spearman's rho για τις παραμέτρους του ταμιευτήρα Μόρνου

	EC	TDS	TU	pH	TA	TH	Cl ⁻	SO ₄ ²⁻	Na ⁺	Ca ²⁺	Mg ²⁺	COD	DO	NH ₄ ⁺
EC	1													
TDS	0,519**	1												
TU	0,172	0,254	1											
pH	-0,132	-0,320*	-0,454**	1										
TA	0,619**	0,564**	0,391**	-0,103	1									
TH	0,642**	0,574**	0,347*	-0,210	0,777**	1								
Cl ⁻	-0,371*	-0,211	-0,297*	0,233	-0,164	-0,414**	1							
SO ₄ ²⁻	0,401**	0,418**	0,240	-0,270	0,347*	0,588**	-0,266	1						
Na ⁺	-0,101	-0,149	0,180	-0,119	-0,072	-0,047	0,038	0,334*	1					
Ca ²⁺	0,731**	0,564**	0,300*	-0,130	0,868**	0,654**	-0,266	0,297*	-0,092	1				
Mg ²⁺	0,431**	0,462**	0,348*	-0,310*	0,408**	0,596**	-0,462**	0,717**	0,168	0,334*	1			
COD	0,173	0,303*	0,424**	-0,131	0,285	0,272	-0,091	0,383**	-0,036	0,153	0,471**	1		
DO	0,361*	0,526**	0,468**	-0,339*	0,652**	0,553**	-0,21	0,197	0,036	0,606**	0,241	0,259	1	
NH ₄ ⁺	0,062	0,270	-0,129	-0,011	0,066	0,163	0,145	0,077	-0,368*	0,082	0,026	0,044	-0,067	1
NO ₃ ⁻	0,080	0,280	0,390**	-0,415**	0,346*	0,202	0,115	0,104	-0,018	0,313*	0,108	0,302*	0,488**	0,000

** Στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,01$ (αμφίπλευρος έλεγχος)

* Στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας $\alpha=0,05$ (αμφίπλευρος έλεγχος)

10.6.2. Ανάλυση σε κύριες συνιστώσες/Παραγοντική ανάλυση

Με συνεκτίμηση της μορφής του γραφήματος των ιδιοτιμών (scree-plot) και του κριτηρίου Kaiser των ιδιοτιμών, εξήχθησαν έξι παράγοντες οι οποίοι επεξηγούν το 80,4% της συνολικής διακύμανσης των δεδομένων. Οι συμμετοχικότητες ευρέθησαν να είναι υψηλές (μεγαλύτερες από 0,50) για όλες τις μεταβλητές (Πίνακας 10.18). Επομένως, το μοντέλο των έξι παραγόντων επεξηγεί ικανοποιητικά τη μεταβλητότητα όλων των μεταβλητών και μπορεί να χρησιμοποιηθεί για να υποδείξει τις κύριες υδροχημικές διαδικασίες που ελέγχουν τη σύσταση του νερού, χωρίς την απώλεια κάποιων σημαντικών χαρακτηριστικών. Για την περιστροφή των αξόνων χρησιμοποιήθηκε, όπως και στους άλλους ταμιευτήρες/λίμνες, η μέθοδος Varimax με Kaiser Normalization.

Ο πρώτος παράγοντας επεξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης (27,3%) και παρουσιάζει υψηλές (ή σχεδόν υψηλές) θετικές φορτίσεις στις μεταβλητές TA, TH, Ca^{2+} και μέτριες θετικές φορτίσεις στις μεταβλητές EC, TDS και DO (Πίνακας 10.18). Αυτός ο παράγοντας μπορεί να θεωρηθεί ως παράγοντας «σκληρότητας του νερού» και βασίζεται κυρίως στην προέλευση του φυσικού νερού (Simeonov και λοιποί, 2001). Η «σκληρότητα» του νερού φαίνεται να συµμεταβάλλεται με το ασβέστιο ενώ η συµμετοχή του μαγνησίου στη διακύμανση της σκληρότητας είναι ελάχιστη. Ο δεύτερος παράγοντας επεξηγεί το 12,6% της συνολικής διακύμανσης και παρουσιάζει υψηλή θετική φόρτιση στη μεταβλητή SO_4^{2-} και μέτρια θετική φόρτιση στις μεταβλητές Mg^{2+} και Na^+ . Ο τρίτος παράγοντας επεξηγεί το 11,4% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση στη μεταβλητή COD και μέτρια θετική φόρτιση στη μεταβλητή TU. Ο τέταρτος παράγοντας επεξηγεί το 10,8% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή αρνητική φόρτιση στο pH και μέτρια θετική φόρτιση στα NO_3^- . Ο πέμπτος παράγοντας επεξηγεί το 9,3% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση μόνο στο αμμώνιο (NH_4^+) και μέτρια αρνητική φόρτιση στο Na^+ , αν και το Na^+ φαίνεται ότι φορτίζει κυρίως τον δεύτερο παράγοντα. Τέλος, ο έκτος παράγοντας επεξηγεί το 9,1% της συνολικής διακύμανσης των δεδομένων και παρουσιάζει υψηλή θετική φόρτιση μόνο στα χλωριούχα, γεγονός που πιθανόν υποδηλώνει ότι η συγκεκριμένη

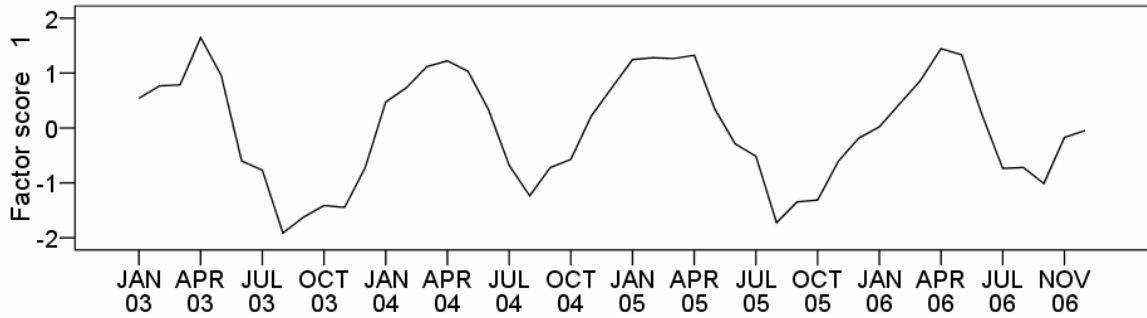
μεταβλητή μεταβάλλεται ανεξάρτητα από τις υπόλοιπες μεταβλητές (Clow και λοιποί,1996).

Πίνακας 10.18. Παραγοντική ανάλυση των παραμέτρων του ταμιευτήρα Μόρνου (Μέθοδος εξαγωγής: Ανάλυση κυρίων συνιστωσών / Μέθοδος περιστροφής: Varimax με Kaiser Normalization)

Παρ/τροι	Π/ντας 1	Π/ντας 2	Π/ντας 3	Π/ντας 4	Π/ντας 5	Π/ντας 6	Συμμ/τες
EC	0,73	0,20	0,01	-0,06	0,14	-0,27	0,67
TDS	0,60	0,11	0,11	0,25	0,34	-0,31	0,65
pH	-0,01	-0,23	-0,02	-0,90	-0,05	0,16	0,89
TA	0,89	0,13	0,28	0,02	0,06	0,05	0,89
TH	0,77	0,36	0,19	0,03	0,15	-0,19	0,83
SO ₄ ²⁻	0,32	0,82	0,20	0,12	0,07	0,05	0,83
Ca ²⁺	0,91	0,03	0,02	0,06	-0,06	-0,11	0,85
Mg ²⁺	0,29	0,67	0,44	0,12	0,13	-0,31	0,85
COD	0,15	0,20	0,89	0,02	0,10	0,05	0,87
DO	0,73	-0,09	0,17	0,35	-0,17	0,06	0,73
TU	0,19	0,06	0,60	0,43	-0,24	-0,21	0,68
Cl ⁻	-0,27	0,04	-0,05	-0,07	0,14	0,88	0,87
Na ⁺	-0,16	0,67	-0,13	0,10	-0,55	0,21	0,85
NH ₄ ⁺	0,05	0,03	-0,03	0,01	0,89	0,15	0,82
NO ₃ ⁻	0,33	-0,07	0,37	0,62	-0,03	0,38	0,78
% της Διακ/νσης	27,3	12,6	11,4	10,8	9,3	9,1	

Οι τιμές με έντονους χαρακτήρες υποδεικνύουν ισχυρή συσχέτιση μεταξύ μιας μεταβλητής και ενός παράγοντα

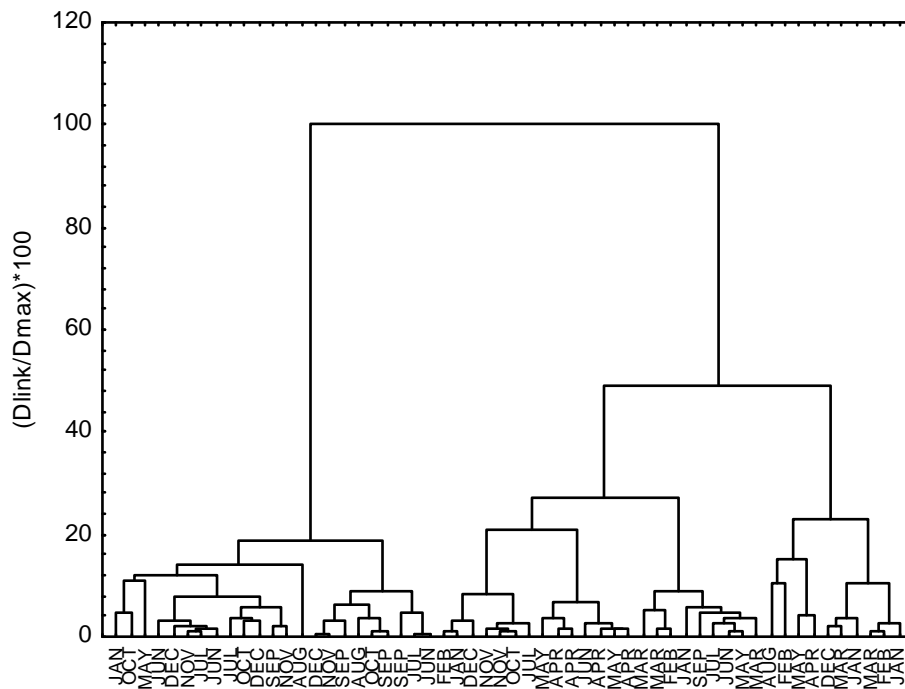
Δεδομένου ότι ο πρώτος παράγοντας επεξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης των δεδομένων, συγκριτικά με τους υπόλοιπους, θα μπορούσε να δώσει χρήσιμες πληροφορίες για το χρονικό πρότυπο της διακύμανσής. Το διάγραμμα της χρονολογικής σειράς των σκορ του πρώτου παράγοντα (Σχήμα 10.8) παρουσιάζει ένα σχεδόν ημιτονοειδές πρότυπο. Οι υψηλότερες τιμές παρατηρούνται από τον Ιανουάριο έως τον Μάιο και οι χαμηλότερες από τον Ιούνιο έως τον Δεκέμβριο.



Σχήμα 10.8. Χρονική διακύμανση των σκορ του πρώτου παράγοντα για τον Μόρνο.

10.6.3. Ανάλυση κατά συστάδες

Το δεντρόγραμμα (Σχήμα 10.9) που προέκυψε από την ιεραρχική ανάλυση κατά συστάδες παρέιχε μια οπτική περίληψη των διαδικασιών ομαδοποίησης, παρουσιάζοντας μια απεικόνιση των ομάδων και την εγγύτητά τους.



Σχήμα 10.9. Ιεραρχικό δεντρόγραμμα για τις περιόδους παρακολούθησης του Μόρνου

Η ανάλυση κατά συστάδες κατέληξε σε δύο κύριες ομάδες. Η πρώτη ομάδα περιλάμβανε τους μήνες Ιούνιο, Ιούλιο, Αύγουστο, Σεπτέμβριο και Οκτώβριο, Νοέμβριο και Δεκέμβριο οι οποίοι προσεγγιστικά αντιστοιχούν στους θερινούς και

φθινοπωρινούς μήνες (Ιούνιος – Δεκέμβριος). Η δεύτερη ομάδα περιλάμβανε τους υπόλοιπους μήνες του έτους (Ιανουάριο έως Μάιο) και αντιστοιχεί προσεγγιστικά στους χειμερινούς και εαρινούς μήνες του έτους.

10.6.4. Διαχωριστική ανάλυση

Η χρονική μεταβλητότητα των παραμέτρων ποιότητας του νερού αποτιμήθηκε με τη χρησιμοποίηση της διαχωριστικής ανάλυσης. Η χρονική διαχωριστική ανάλυση εκτελέστηκε στο σύνολο δεδομένων μετά την ομαδοποίηση τους σε δύο περιόδους/εποχές (Ιούνιος-Δεκέμβριος και Ιανουάριος-Μάιος), βασιζόμενη στα αποτελέσματα της χρονικής ανάλυσης κατά συστάδες. Η ομοιογένεια των διακυμάνσεων - συνδιακυμάνσεων των δύο ομάδων ελέγχθηκε με τον έλεγχο Box's M. Ο έλεγχος δεν ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας $\alpha=0,05$ (p-τιμή=0,12).

Η βηματική μέθοδος διαχωριστικής ανάλυσης οδήγησε σε μια στατιστικά σημαντική διαχωριστική συνάρτηση (DF) (Wilks' lambda=0,164, p-τιμή <0,001) και χρησιμοποίησε τις μεταβλητές Ca^{2+} , DO, NO_3^- και NH_4^+ . Οι τιμές ανοχής (tolerance) για κάθε μία από τις ανεξάρτητες μεταβλητές ήταν πολύ μεγαλύτερες από 0,10, γεγονός που υποδεικνύει την έλλειψη πολυσυγγραμμικότητας των μεταβλητών. Οι συντελεστές συσχέτισης (r_{DF}) των παραπάνω μεταβλητών με τη διαχωριστική συνάρτηση παρουσιάζονται στον Πίνακα 10.19 που ακολουθεί.

Πίνακας 10.19. Συντελεστές συσχέτισης μεταβλητής-διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για τον Μόρνο

Μεταβλητή	r_{DF}
Ca^{2+}	0,562
DO	0,464
NO_3^-	0,313
NH_4^+	0,106

Παρατηρείται λοιπόν μέτρια θετική συσχέτιση της διαχωριστικής συνάρτησης με τις μεταβλητές Ca^{2+} και DO, ασθενής θετική συσχέτιση με τη μεταβλητή NO_3^- και εξαιρετικά ασθενής συσχέτιση με τη μεταβλητή NH_4^+ . Το κεντροειδές της DF για την ομάδα 1 (Ιούνιος-Δεκέμβριος) υπολογίστηκε ίσο με -1,9 ενώ το κεντροειδές για την

ομάδα 2 (Ιανουάριος-Μάιος) υπολογίστηκε ίσο με 2,6. Επομένως, συμπεραίνεται ότι οι συγκεντρώσεις Ca^{2+} και DO είναι χαμηλότερες την περίοδο «Ιούνιο έως Δεκέμβριο» συγκριτικά με την περίοδο «Ιανουάριο έως Μάιο». Η βηματική διαχωριστική ανάλυση απέφερε 100% ορθή ταξινόμηση του συνόλου δεδομένων. Η διασταυρούμενη επικύρωση του μοντέλου οδήγησε επίσης σε 100% ορθή ταξινόμηση (Πίνακας 10.20). Το υψηλό ποσοστό ορθής ταξινόμησης αποτελεί ένδειξη άριστης ακρίβειας του μοντέλου.

Πίνακας 10.20. Πίνακας ταξινόμησης για τη βηματική DA της χρονικής διακύμανσης του νερού στον Μόρνο

Δείγμα	Παρατηρούμενη κατηγορία	Εκτιμηθείσα κατηγορία		% Ορθής ταξινόμησης
		Ιουν-Δεκ	Ιαν-Μάης	
Πρωτότυπο ^α	Ιουν-Δεκ	27	0	100
	Ιαν-Μάης	0	20	100
Διασταυρούμενης επικύρωσης ^β	Ιουν-Δεκ	27	0	100
	Ιαν-Μάης	0	20	100

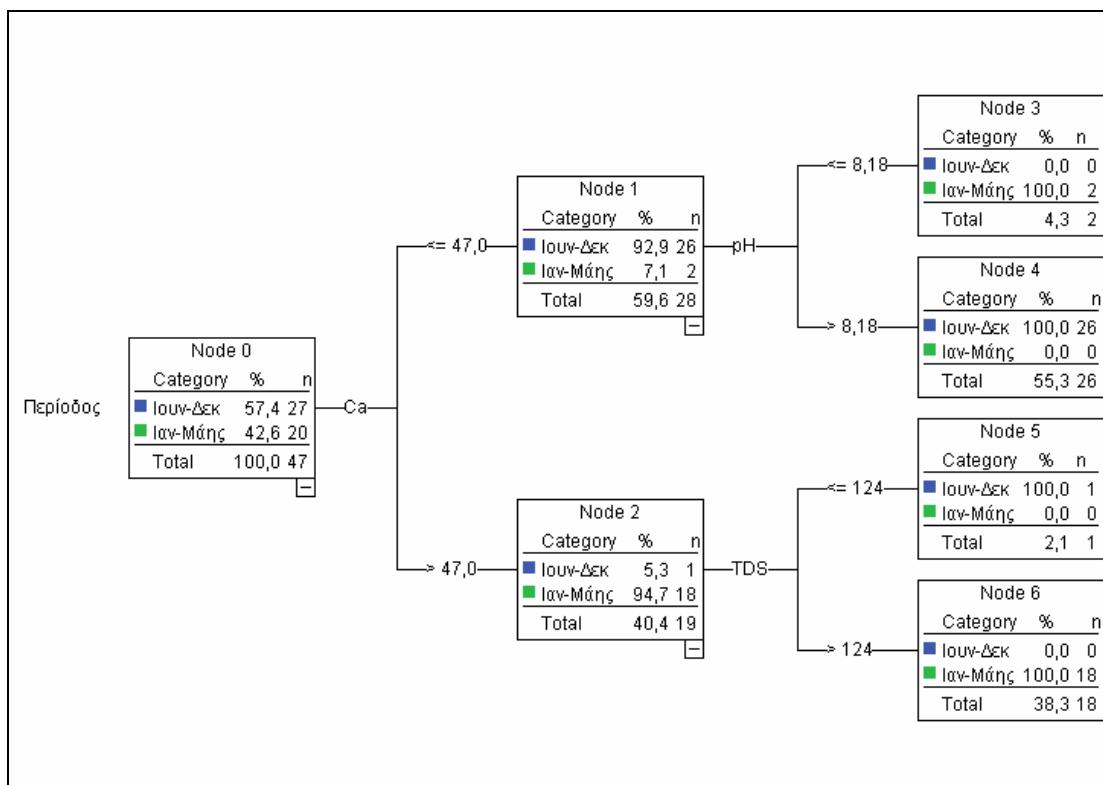
α: 100% ορθή ταξινόμηση των πρωτότυπων δεδομένων

β: 100% ορθή ταξινόμηση των δεδομένων διασταυρούμενης επικύρωσης

10.6.5. Δέντρα ταξινόμησης

Για την κατασκευή των δέντρων ταξινόμησης και παλινδρόμησης (CART), ακολουθήθηκε η ίδια ακριβώς διαδικασία που περιγράφηκε στην περίπτωση του ταμιευτήρα Μαράθωνα (δείκτης μη-καθαρότητας του Gini, περικοπή δέντρου και επιλογή βέλτιστου δέντρου βάσει της 10-πλής διασταυρούμενη επικύρωσης και τον κανόνα του ενός τυπικού σφάλματος).

Ο αλγόριθμος CART κατέληξε σε ένα δέντρο με δύο επίπεδα ανάπτυξης κάτω από τον αρχικό κόμβο και υπέδειξε ότι οι κύριες παράμετροι που χρησιμοποιήθηκαν για τον διαχωρισμό των δειγμάτων στις δυο χρονικές περιόδους ήταν οι Ca^{2+} , pH και TDS. (Σχήμα 10.10). Το ποσοστό της συνολικής ορθής ταξινόμησης των δειγμάτων υπολογίστηκε ίσο με 100%. Η 10-πλή διασταυρούμενη επικύρωση του μοντέλου είχε ποσοστό ορθής ταξινόμησης ίσο με 91,5%.



Σχήμα 10.10. Δέντρο ταξινόμησης για την χρονική διακύμανση των δεδομένων του Μόρνου (αλγόριθμος CART).

10.7. ΔΙΑΚΡΙΣΗ ΜΕΤΑΞΥ ΤΩΝ ΤΑΜΙΕΥΤΗΡΩΝ ΜΑΡΑΘΩΝΑ ΚΑΙ ΜΟΡΝΟΥ ΒΑΣΕΙ ΤΩΝ ΦΥΣΙΚΟΧΗΜΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΝΕΡΟΥ

Δεδομένου ότι, κατά τη χρονική περίοδο που εξετάζεται, μόνο οι ταμιευτήρες Μαραθώνα και Μόρνου χρησιμοποιούνταν ως υδατικοί πόροι, παρουσιάζει ενδιαφέρον η αξιολόγηση των ποιοτικών παραμέτρων του νερού που αποφέρουν την καλύτερη διάκριση μεταξύ των δειγμάτων νερού των δύο ταμιευτήρων. Για αυτόν τον σκοπό, χρησιμοποιήθηκε η βηματική διαχωριστική ανάλυση. Χρησιμοποιήθηκαν τα δεδομένα που καλύπτουν κοινή χρονική περίοδο για τους δύο ταμιευτήρες (Σεπτέμβριος 2003 – Δεκέμβριος 2006). Συνολικά χρησιμοποιήθηκαν 78 δείγματα νερού (39 από κάθε ταμιευτήρα). Καθώς για κάποιες από τις μεταβλητές δεν ήταν δυνατόν ένας κοινός μετασχηματισμός για την ίδια μεταβλητή να βελτιώσει τα χαρακτηριστικά της, όσον αφορά στην κανονικότητα στους δύο ταμιευτήρες, χρησιμοποιήθηκε η μέθοδος διαχωριστικής ανάλυσης στις τάξεις μεγέθους των

μεταβλητών του συνόλου των δεδομένων (Conover και Iman, 1980; Conover και Iman, 1981; Ogorman and Woolson, 1993). Η ομοιογένεια των διακυμάνσεων - συνδιακυμάνσεων των δύο ομάδων ελέγχθηκε με τον έλεγχο Box's M. Το παρατηρούμενο επίπεδο σημαντικότητας του ελέγχου υπολογίστηκε να είναι p -τιμή=0,033. Επομένως σε επίπεδο σημαντικότητας $\alpha=0,01$ δεν απορρίπτεται η υπόθεση της ομοιογένειας.

Η βηματική μέθοδος διαχωριστικής ανάλυσης οδήγησε σε μια στατιστικά σημαντική διαχωριστική συνάρτηση (DF) (Wilks' $\lambda=0,075$, p -τιμή $<0,001$) και χρησιμοποίησε τις μεταβλητές Mg^{2+} , Cl^- , NH_4^+ , TU και COD. Οι τιμές ανοχής (tolerance) για κάθε μία από τις ανεξάρτητες μεταβλητές ήταν πολύ μεγαλύτερες από 0,10, γεγονός που υποδεικνύει την έλλειψη πολυσυγγραμμικότητας των μεταβλητών. Οι συντελεστές συσχέτισης (r_{DF}) όλων των μεταβλητών με τη διαχωριστική συνάρτηση παρουσιάζονται στον Πίνακα 10.21 που ακολουθεί.

Πίνακας 10.21. Συντελεστές συσχέτισης μεταβλητής - διαχωριστικής συνάρτησης (r_{DF}) στην βηματική διαχωριστική ανάλυση για των ταμειυτήρων Μαραθώνα - Μόρνου

Μεταβλητή	r_{DF}
Mg^{2+}	0,503
Cl^-	0,500
EC (a)	0,488
Ca^{2+} (a)	0,408
Na^+ (a)	0,404
NO_3^- (a)	0,356
TDS (a)	0,343
TA (a)	0,339
TH (a)	0,338
SO_4^{2-} (a)	0,312
COD	0,294
TU	0,213
NH_4^+	0,180
pH (a)	-0,136
DO (a)	-0,043

(a) Η μεταβλητή δεν χρησιμοποιήθηκε από το μοντέλο βηματικής διαχωριστικής ανάλυσης.

Στον Πίνακα 10.21, οι μεταβλητές έχουν διαταχθεί κατά απόλυτο μέγεθος συσχέτισης με τη διαχωριστική συνάρτηση. Επομένως, παρέχεται η εικόνα της χρησιμότητας κάθε μεταβλητής στην διαχωριστική συνάρτηση. Παρατηρείται μέτρια θετική συσχέτιση της διαχωριστικής συνάρτησης με τις μεταβλητές Mg^{2+} και Cl^- , και εξαιρετικά ασθενής συσχέτιση με τις μεταβλητές COD, TU και NH_4^+ . Το κεντροειδές της DF για τον ταμιευτήρα Μαραθώνα υπολογίστηκε ίσο με 3,5 ενώ το κεντροειδές για τον ταμιευτήρα Μόρνου ίσο με -3,5. Επομένως, συμπεραίνεται ότι οι συγκεντρώσεις Mg^{2+} και Cl^- είναι υψηλότερες στον Μαραθώνα σε σύγκριση με τον Μόρνο. Επιπλέον, οι δύο αυτές παράμετροι είναι βασικές στον διαχωρισμό του νερού μεταξύ των δύο ταμιευτήρων. Η βηματική διαχωριστική ανάλυση στις τάξεις μεγέθους των ποιοτικών χαρακτηριστικών του νερού απέφερε 100% ορθή ταξινόμηση του συνόλου δεδομένων. Στο μονομεταβλητό επίπεδο, ο έλεγχος Levene για την ισότητα των διακυμάνσεων μίας μεταβλητής στους δύο ταμιευτήρες δεν ήταν στατιστικά σημαντικός για καμία μεταβλητή σε επίπεδο σημαντικότητας $\alpha=0,05$ (p-τιμές μεταξύ 0,062 για το DO και 0,994 για την TH). Οι μονομεταβλητοί έλεγχοι στις τάξεις μεγέθους των παραμέτρων (ανάλυση διακύμανσης), εκτός από το DO (p-τιμή=0,225), ήταν για όλες τις υπόλοιπες παραμέτρους στατιστικά σημαντικοί (p-τιμές<0,001). Όλες οι παράμετροι πλην του διαλελυμένου οξυγόνου βρέθηκαν να είναι στατιστικά σημαντικά υψηλότερες στον Μαραθώνα σε σύγκριση με τον Μόρνο.

10.8. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ανακεφαλαιώνοντας, η συγκεκριμένη μελέτη υποδεικνύει εμφανώς ότι οι πολυμεταβλητές στατιστικές τεχνικές, όπως οι PCA/FA, CA, DA και CART, είναι ισχυρά εργαλεία για την επεξήγηση πολύπλοκων σετ δεδομένων της ποιότητας του νερού, την κατανόηση των χρονικών διακυμάνσεων της και τον εντοπισμό και την ταυτοποίηση συνόλων ομαδοποιημένων στοιχείων στα δεδομένα. Επομένως οι τεχνικές αυτές μπορούν να προσφέρουν χρήσιμες πληροφορίες για την προσαρμογή της επεξεργασίας του ανεπεξέργαστου νερού βάσει των ιδιαίτερων χαρακτηριστικών ποιότητάς του.

Η ανάλυση κύριων συνιστωσών/παραγοντική ανάλυση (PCA/FA) επέτρεψε τη μείωση των δεκαπέντε αρχικών μεταβλητών σε τρεις, τέσσερις και έξι σημαντικούς

λανθάνοντες παράγοντες για τον Μαραθώνα, την Υλίκη και τον Μόρνο αντίστοιχα οι οποίοι επεξηγούν το μεγαλύτερο μέρος της μεταβλητότητας στην ποιότητα νερού κάθε λίμνης/ταμιευτήρα. Επιπλέον, η PCA/FA έδωσε ικανοποιητικά αποτελέσματα για την ανίχνευση της χρονικής-εποχικής συμπεριφοράς στο σύνολο των δεδομένων για την Υλίκη και τον Μόρνο. Η ανάλυση συστάδων ομαδοποίησε τους δώδεκα μήνες σε δύο περιόδους για κάθε λίμνη/ταμιευτήρα βασισμένη στην ομοιότητα των χαρακτηριστικών ποιότητας νερού. Η βηματική μέθοδος διαχωριστικής ανάλυσης και ο αλγόριθμος CART των δέντρων ταξινόμησης επέτρεψαν μείωση των διαστάσεων του πολύπλοκου συνόλου δεδομένων, αναδεικνύοντας λίγες μόνο παραμέτρους-δείκτες ως υπεύθυνες για τις μεγάλες διακυμάνσεις στην ποιότητα νερού στη διάρκεια του χρόνου. Το υψηλό ποσοστό της ορθής ταξινόμησης και των δύο μεθόδων έδειξε την καλή ακρίβεια των μοντέλων.

Οι τιμές των ποιοτικών χαρακτηριστικών του ταμιευτήρα Μόρνου (εκτός από το διαλελυμένο οξυγόνο), στο χρονικό διάστημα κατά το οποίο ο Μαραθώνας ενισχύεται με νερό του Μόρνου, βρέθηκαν στατιστικά σημαντικά υψηλότερες από αυτές του Μόρνου, με σημαντικότερες παραμέτρους διαφοροποίησης τα ιόντα μαγνησίου και χλωρίου.

ΚΕΦΑΛΑΙΟ 11

ΔΙΑΣΦΑΛΙΣΗ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΕΠΕΞΕΡΓΑΣΜΕΝΟΥ ΝΕΡΟΥ ΚΑΙ ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΔΙΚΤΥΟΥ ΥΔΡΕΥΣΗΣ ΜΕ ΤΗ ΧΡΗΣΗ ΠΟΛΥΜΕΤΑΒΛΗΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΤΕΧΝΙΚΩΝ

11.1. ΕΙΣΑΓΩΓΗ

Τις τελευταίες δεκαετίες υπάρχει μια συνεχής αύξηση του ενδιαφέροντος για την ποιότητα του νερού που προορίζεται για ανθρώπινη κατανάλωση (Rizak και Hrudey, 2008; Zhou και λοιποί, 1997). Έχουν ιδρυθεί ειδικοί φορείς προκειμένου να διασφαλίζεται ότι το πόσιμο νερό ακολουθεί υψηλές προδιαγραφές και ότι οι διαδικασίες που χρησιμοποιούνται για την επεξεργασία του ανεπεξέργαστου νερού έχουν σαν αποτέλεσμα ένα προϊόν το οποίο συμμορφώνεται με τη νομοθεσία και τους κανόνες ασφαλείας (Οδηγία 98/83/EEC σχετικά με την ποιότητα του νερού που προορίζεται για ανθρώπινη κατανάλωση, 1998). Παρόλο που η ασφάλεια των σύγχρονων δικτύων επεξεργασμένου νερού θεωρείται δεδομένη, είναι δυνατόν να ακόμα και στην εποχή μας να συμβούν περιστατικά μόλυνσης του νερού σε διάφορα στάδια της παραγωγής πόσιμου νερού. Οι ταμειυτήρες, στους οποίους συγκεντρώνεται το ανεπεξέργαστο νερό και στη συνέχεια οδηγείται στις μονάδες επεξεργασίας, μπορεί να μολυνθούν από μια πληθώρα ουσιών. Αυτό το είδος μόλυνσης μπορεί να οφείλεται σε φυσικούς παράγοντες ή στην ανθρώπινη δραστηριότητα (Brookes και λοιποί, 2004; Prudham, 2004). Οι μονάδες επεξεργασίας νερού μπορεί επίσης να είναι μια πηγή πιθανής μόλυνσης του πόσιμου νερού. Αυτό μπορεί να συμβεί όταν προκαθορισμένες διεργασίες τίθενται εκτός ελέγχου ή όταν συμβούν ατυχήματα κατά το στάδιο της επεξεργασίας (Winston και Leventhal, 2008). Τέλος, τα ίδια τα δίκτυα που χρησιμοποιούνται για τη διανομή του νερού στους καταναλωτές, μπορεί να μην είναι αρκετά ασφαλή ώστε να εγγυώνται τις υψηλές προδιαγραφές του τελικού προϊόντος όπως αυτές διαπιστώθηκαν στις μονάδες επεξεργασίας (Jakoranec και λοιποί, 2008; O'Donnell και λοιποί, 2000).

Τέτοιου είδους περιστατικά μόλυνσης πρέπει να διαχειρίζονται άμεσα (Mahmud και λοιποί, 2007) καθώς η υποβάθμιση της ποιότητας του νερού θα διαδοθεί γρήγορα από το σημείο προέλευσής της προς τους καταναλωτές. Επομένως, είναι σημαντική η γνώση της πλήρους περιγραφής ενός συστήματος επεξεργασίας και διανομής νερού (Huck και Coffey, 2004; Yanga και λοιποί, 2008). Η διεξοδική ανάλυση των πηγών κινδύνου είναι επίσης επιβεβλημένη, για την αναγνώριση των κρίσιμων σημείων κατά μήκος των σταδίων παραγωγής, έτσι ώστε να ελαχιστοποιούνται οι κίνδυνοι (Damikouka και λοιποί, 2007; Hrudey και λοιποί, 2006; Pollard και λοιποί, 2004; Sun και λοιποί, 2007). Σε περίπτωση εμφάνισης τέτοιων περιστατικών, πρέπει να είναι διαθέσιμα κατάλληλα πρωτόκολλα τα οποία να επιτρέπουν, στους διαχειριστές των μονάδων επεξεργασίας και του δικτύου ύδρευσης, την άμεση παρέμβαση και τη λήψη όλων των απαραίτητων ενεργειών για την πρόληψη κάθε αρνητικής επίπτωσης στην υγεία των καταναλωτών (Mondal και λοιποί, 2006).

Παραδοσιακά τα δίκτυα διανομής νερού περιγράφονται με βάση την υδραυλική τους. Αυτή η προσέγγιση είναι κατάλληλη στις περιπτώσεις μικρών δικτύων διανομής ενώ μπορεί να είναι ανεπαρκής για δίκτυα πολύπλοκων δομών στα οποία η προέλευση του επεξεργασμένου νερού μπορεί να διαφοροποιείται λόγω της ζήτησης των καταναλωτών. Οι πολυμεταβλητές στατιστικές μέθοδοι έχουν χρησιμοποιηθεί για την αποτίμηση της ποιότητας του νερού σε όλα τα στάδια παραγωγής, από τους ταμειυτήρες μέχρι τις βρύσες των καταναλωτών (Astel και λοιποί, 2006; Ayoko και λοιποί, 2007; Ismail και Ramadan, 1995; Platikanov και λοιποί, 2007; Stanimirova και λοιποί, 2007). Παρόλα αυτά, δεν έχει εξερευνηθεί έως τώρα η χρησιμότητα των πολυμεταβλητών στατιστικών μεθόδων, ως μέσον περιγραφής των δικτύων διανομής και πρόβλεψης της προέλευσης του νερού. Η αξιοποίηση τέτοιου είδους στατιστικών εργαλείων θα μπορούσε να βοηθήσει επικουρικά στην περιγραφή των δικτύων ύδρευσης παρέχοντας χρήσιμες πληροφορίες στους διαχειριστές των δικτύων με αποτέλεσμα την αύξηση του βαθμού ασφαλείας.

Στην συγκεκριμένη μελέτη, διερευνάται η χρησιμότητα των πολυμεταβλητών στατιστικών μεθόδων, όπως η ανάλυση κύριων συνιστωσών (PCA), η διαχωριστική ανάλυση (DA) και τα δέντρα ταξινόμησης, ως μέσα διάκρισης μεταξύ των πόσιμων νερών που προέρχονται από τρεις μονάδες επεξεργασίας νερού (MEN) της ΕΥΔΑΠ. Επιπλέον, οι τεχνικές αυτές χρησιμοποιούνται για την πρόβλεψη της προέλευσης του

πόσιμο νερού των δεξαμενών δικτύου (ΔΥ) έτσι ώστε να καθίσταται δυνατή η πρόβλεψη της ροής του νερού στο δίκτυο και η διαχείριση προβλημάτων που τυχόν δημιουργούνται στις μονάδες επεξεργασίας πριν φτάσει το νερό στους καταναλωτές.

11.2. ΜΕΘΟΔΟΛΟΓΙΑ

11.2.1 Μονάδες επεξεργασίας νερού, δεξαμενές δικτύου

Το ανεπεξέργαστο νερό, που προέρχεται κυρίως από τους ταμιευτήρες Μόρνου και Μαραθώνα ή σε περιόδους αυξημένης ζήτησης και από τη φυσική λίμνη Υλίκη καθώς και από συγκεκριμένες γεωτρήσεις, υποβάλλεται σε κατάλληλη επεξεργασία στις τέσσερις ΜΕΝ της ΕΥΔΑΠ. Η επεξεργασία περιλαμβάνει τα στάδια της προχλωρίωσης, κροκίδωσης, καθίζησης, διήθησης και μεταχλωρίωσης. Στον Πίνακα 11.1 παρουσιάζονται οι αριθμητικοί μέσοι των ημερήσιων δοσολογιών θειϊκού αργιλίου και χλωρίου σε κάθε ΜΕΝ το έτος 2007.

Πίνακας 11.1. Αριθμητικοί μέσοι των ημερήσιων δοσολογιών θειϊκού αργιλίου και χλωρίου ανά ΜΕΝ το έτος 2007

MEN	Θειϊκό αργίλιο (mg/l)	Χλώριο (mg/l)
Αχαρνών	12	1,5
Γαλατσίου	14	1,1
Πολυδενδρίου	17	1,6
Ασπροπύργου	14	1,6

Το τελικό προϊόν των μονάδων επεξεργασίας νερού (επεξεργασμένο νερό) παραδίδεται στους καταναλωτές μέσω ενός δικτύου διανομής συνολικού μήκους περίπου 8.000 km. Εκτός των μεγάλων δεξαμενών των ΜΕΝ, στο δίκτυο ύδρευσης υπάρχουν διασκορπισμένες οι επονομαζόμενες δεξαμενές δικτύου ύδρευσης (ΔΔΥ). Η ποιότητα του νερού αυτών των μικρότερων δεξαμενών σχετίζεται άμεσα με την ποιότητα του νερού των δεξαμενών ΜΕΝ με τις οποίες συνδέονται. Παρόλο που κάθε ΔΔΥ αναμένεται να λαμβάνει νερό από μία αντίστοιχη δεξαμενή ΜΕΝ, η ζήτηση καθώς και η πολυπλοκότητα του δικτύου διανομής επιτρέπει την είσοδο νερού στις

ΔΔΥ το οποίο προέρχεται από διαφορετικές δεξαμενές ΜΕΝ. Προφανώς, κάθε αλλαγή της ποιότητας του επεξεργασμένου νερού των δεξαμενών ΜΕΝ θα επιδρά στην ποιότητα του νερού των μικρότερων ΔΔΥ. Επομένως, είναι ιδιαίτερα σημαντική για τους διαχειριστές του δικτύου ύδρευσης η πληροφόρηση για την προέλευση του νερού, ιδιαίτερα σε περίπτωση επείγουσας ανάγκης, ώστε να γνωρίζουν ποια μέρη του δικτύου ύδρευσης θα επηρεαστούν.

Στην παρούσα μελέτη δεν συμπεριλήφθηκε η ΜΕΝ Ασπροπύργου. Η συγκεκριμένη ΜΕΝ προμηθεύει νερό μια καλά οριοθετημένη περιοχή (Δυτική Αττική – Θριάσιο Πεδίο καθώς και το νησί της Σαλαμίνας) η οποία συνήθως δεν λαμβάνει νερό από τις υπόλοιπες ΜΕΝ.

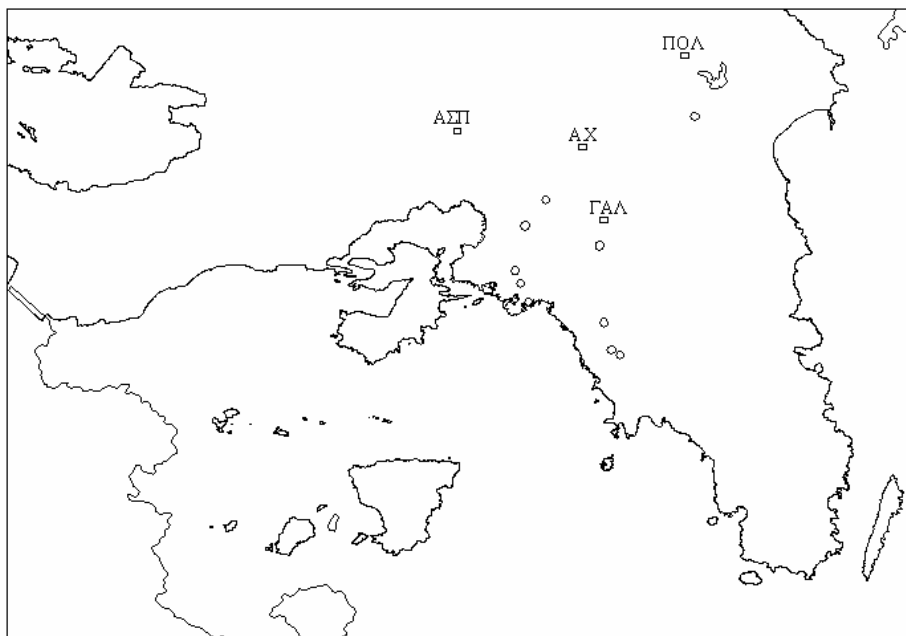
11.2.2. Περιγραφή των δεδομένων

Κατά τη διάρκεια του έτους 2007, αναλύθηκαν δείγματα νερού των δεξαμενών των τριών ΜΕΝ της ΕΥΔΑΠ που βρίσκονται στο Γαλάτσι (ΓΑΛ), στις Αχαρνές (ΑΧ) και στο Πολυδένδρι (ΠΟΛ), τρεις φορές κάθε μήνα, για τις παρακάτω παραμέτρους:

Ηλεκτρική αγωγιμότητα (EC), ολική αλκαλικότητα (TA), πυκνότητα συγκέντρωσης υδρογόνου (pH), αργίλιο (Al), ολική σκληρότητα (TH), χλωριούχα (Cl⁻), υπολειμματικό χλώριο (free Cl), ασβέστιο (Ca²⁺) και μαγνήσιο (Mg²⁺). Οι παραπάνω προσδιορισμοί διεκπεραιώθηκαν στα χημικά εργαστήρια της ΕΥΔΑΠ με τη χρησιμοποίηση μεθόδων που περιγράφονται στις «Τυπικές μεθόδους για την εξέταση νερού και αποβλήτων» (Standard Methods for the Examination of Water and Wastewater) (APHA, AWWA, WEF, 1992). Η ποιότητα των μετρήσεων εξασφαλίστηκε μέσω της προσεκτικής τυποποίησης των μεθόδων, της διαδικασίας των τυφλών δειγμάτων, των αναλύσεων προτύπων και διπλών δειγμάτων. Μετά την εξαίρεση των δειγμάτων τα οποία δεν αναλύθηκαν για όλες τις προαναφερόμενες παραμέτρους, απέμειναν συνολικά 89 δείγματα επεξεργασμένου νερού των ΜΕΝ για περαιτέρω στατιστική ανάλυση. Τα εναπομείναντα δείγματα, για κάθε μία ΜΕΝ, δεν ισαπέχουν ακριβώς μεταξύ τους στη διάρκεια του χρόνου. Επομένως, δεν θα μπορούσε να αποτιμηθεί η σειριακή συσχέτιση των μετρήσεων μεταξύ των διαδοχικών δειγμάτων. Σύμφωνα με τους Loftis και Ward (1980), όταν η συχνότητα της δειγματοληψίας είναι μεταξύ περίπου 10 και 30 δείγματα ανά έτος, οι επιδράσεις της εποχικής διακύμανσης και της σειριακής συσχέτισης τείνουν να

«αλληλοαναιρούνται». Συνεπώς, για την εκτίμηση των διαστημάτων εμπιστοσύνης των μέσων των μεταβλητών, θα πρέπει ή να ληφθούν υπόψη τόσο η εποχική διακύμανση όσο και η σειριακή συσχέτιση ή να αγνοηθούν και οι δύο. Η θεώρηση μόνο της εποχικής διακύμανσης εισάγει μεγαλύτερο λάθος απ'ότι αν αγνοηθεί. Συνεπώς, για τις στατιστικές αναλύσεις που ακολούθησαν, δεν θεωρήθηκε ούτε η εποχική διακύμανση ούτε η σειριακή συσχέτιση των δειγμάτων επεξεργασμένου νερού των ΜΕΝ. Επιπλέον, αυτή η απλοποίηση επιτρέπει μια περισσότερο άμεση διαδικασία ταξινόμησης των δειγμάτων των ΔΔΥ. Στον Πίνακα 11.2 παρουσιάζονται περιγραφικά στατιστικά μέτρα και τα χαρακτηριστικά των κατανομών των αποτελεσμάτων για κάθε μία από τις δεξαμενές επεξεργασμένου νερού ΜΕΝ.

Οι ίδιες ακριβώς παράμετροι μετρήθηκαν επίσης, μία φορά ή δύο ανά μήνα, σε 180 δείγματα νερού από 10 ΔΔΥ, για τις οποίες ήταν γνωστή εξ αρχής η προέλευση του νερού με το οποίο γινόταν ο εφοδιασμός τους. Οι έξι από τις ΔΔΥ λάμβαναν νερό από τη ΜΕΝ Αχαρνών, οι τρεις από τη ΜΕΝ Γαλασίου και μία από τη ΜΕΝ Πολυδενδρίου. Στο Σχήμα 11.1 παρουσιάζονται οι μονάδες επεξεργασίας νερού της ΕΥΔΑΠ και οι δεξαμενές δικτύου ύδρευσης που θεωρήθηκαν στην συγκεκριμένη μελέτη.



Σχήμα 11.1. Μονάδες επεξεργασίας νερού (□:ΜΕΝ) της ΕΥΔΑΠ και δεξαμενές δικτύου ύδρευσης που θεωρήθηκαν στη μελέτη (○:ΔΔΥ).

Πίνακας 11.2. Περιληπτικά περιγραφικά στατιστικά μέτρα και χαρακτηριστικά κατανομών των χημικών παραμέτρων ανά ΜΕΝ

ΜΕΝ	Παράμετρος	Μονάδες	Αριθμός δειγμάτων	Αριθμητικός μέσος	Ελάχιστη τιμή	Μέγιστη τιμή	Τυπική απόκλιση	Ασυμμετρία		Κύρτωση	
								Στατιστικό	Τυπικό σφάλμα	Στατιστικό	Τυπικό σφάλμα
ΜΕΝ	EC	μS/cm	29	279,2	272	288	4,2	0,19	0,43	-0,69	0,85
	TA	mg/l	29	120,0	114	126	2,7	0,40	0,43	-0,02	0,85
	pH		29	7,9	7,7	8	0,1	0,54	0,43	-0,39	0,85
	Al	μg/l	29	138,1	100	178	19,9	0,07	0,43	-0,09	0,85
	TH	mg/l	29	134,7	130	138	2,2	0,02	0,43	-1,07	0,85
	Cl ⁻	mg/l	29	3,9	3,7	4,3	0,2	0,64	0,43	0,29	0,85
	Free Cl	mg/l	29	0,55	0,34	0,82	0,1	0,33	0,43	0,50	0,85
	Ca ²⁺	mg/l	29	46,1	44	48	1,2	0,00	0,43	-0,42	0,85
ΑΧ	Mg ²⁺	mg/l	29	5,3	5,1	5,4	0,1	-0,66	0,43	-0,88	0,85
ΜΕΝ	EC	μS/cm	30	309,7	295	322	6,9	0,13	0,43	-0,46	0,83
	TA	mg/l	30	123,9	117	132	4,4	0,08	0,43	-0,95	0,83
	pH		30	7,7	7,6	7,8	0,1	-0,12	0,43	-0,30	0,83
	Al	μg/l	30	74,5	40	131	24,4	0,29	0,43	-0,87	0,83
	TH	mg/l	30	146,0	140	154	3,5	0,49	0,43	-0,20	0,83
	Cl ⁻	mg/l	30	7,9	5,6	10,5	1,1	0,22	0,43	0,51	0,83
	Free Cl	mg/l	30	0,47	0,3	0,65	0,1	0,15	0,43	-0,89	0,83
	Ca ²⁺	mg/l	30	49,0	47	51	1,2	-0,07	0,43	-0,85	0,83
ΓΑΛ	Mg ²⁺	mg/l	30	6,0	5,3	7	0,4	0,94	0,43	0,60	0,83
ΜΕΝ	EC	μS/cm	30	282,5	271	289	4,5	-0,65	0,43	0,14	0,83
	TA	mg/l	30	118,0	111	127	3,2	0,50	0,43	1,54	0,83
	pH		30	7,8	7,6	8	0,1	0,52	0,43	1,53	0,83
	Al	μg/l	30	91,2	53	112	14,0	-0,71	0,43	0,37	0,83
	TH	mg/l	30	136,5	133	141	2,3	0,15	0,43	-0,96	0,83
	Cl ⁻	mg/l	30	4,1	3,8	4,7	0,2	0,55	0,43	0,94	0,83
	Free Cl	mg/l	30	0,48	0,29	0,7	0,1	0,17	0,43	0,20	0,83
	Ca ²⁺	mg/l	30	46,4	44	49	1,2	0,29	0,43	-0,63	0,83
ΠΟΛ	Mg ²⁺	mg/l	30	5,4	5,1	5,8	0,1	0,44	0,43	1,13	0,83

11.2.3. Χειρισμός των δεδομένων και πολυμεταβλητές στατιστικές μέθοδοι

Τα αποτελέσματα των μετρήσεων των παραμέτρων ποιότητας του επεξεργασμένου νερού των ΜΕΝ υπεβλήθησαν αρχικά σε ανάλυση κυρίων συνιστωσών (PCA) με σκοπό την επεξήγηση της διακύμανσής τους και την κατανόηση της βασικής δομής τους.

Η διαχωριστική ανάλυση (DA) χρησιμοποιήθηκε με τη μορφή της βηματικής μεθόδου με σκοπό την αξιολόγηση της σημαντικότητας των μεταβλητών στη διάκριση μεταξύ των τριών ΜΕΝ και την εκτίμηση της ορθότητας ταξινόμησης των δειγμάτων των ΔΔΥ ως προς τη ΜΕΝ από την οποία προέρχονταν.

Τα δέντρα ταξινόμησης και παλινδρόμησης (CART) χρησιμοποιήθηκαν ως μη παραμετρική στατιστική μέθοδος εναλλακτική της διαχωριστικής ανάλυσης. Ο αλγόριθμος κατασκευής των «δέντρων ταξινόμησης και παλινδρόμησης», χρησιμοποιήθηκε για τον καθορισμό ενός συνόλου λογικών συνθηκών, (διαχωριστικά σημεία - splits) - της μορφής «εάν – τότε», οι οποίες επιτρέπουν την ακριβή πρόβλεψη ή ταξινόμηση των περιπτώσεων (Breiman και λοιποί, 1984).

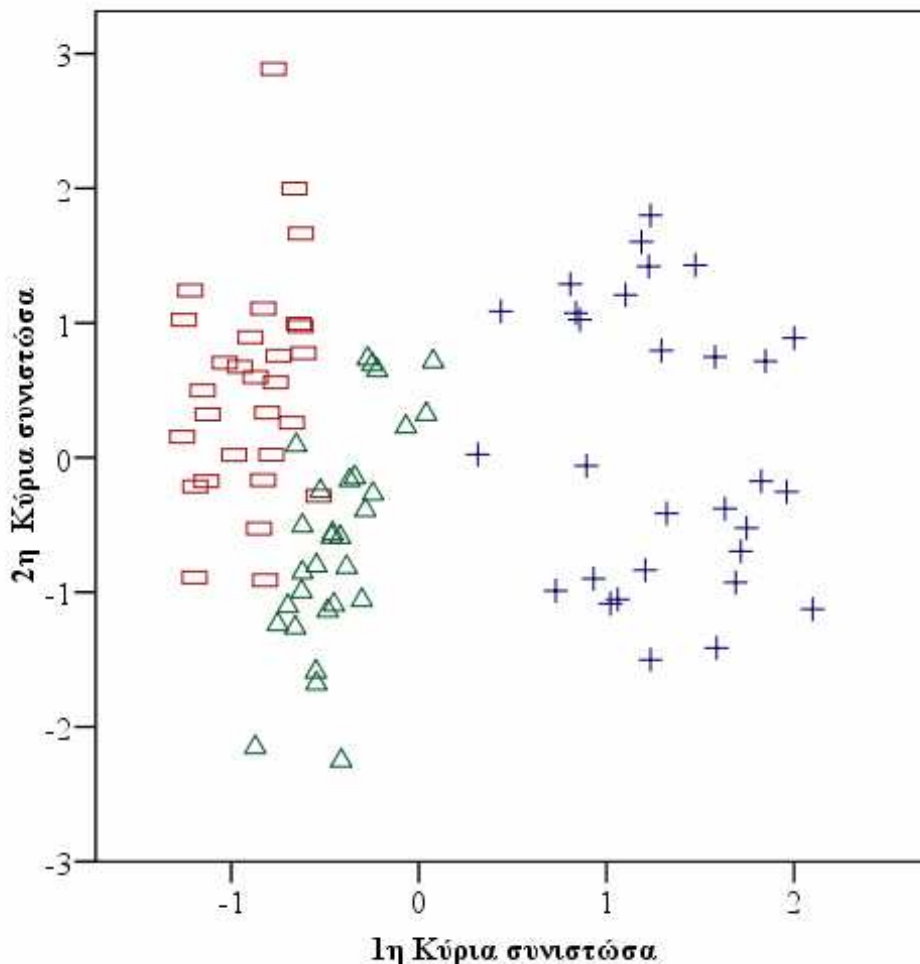
Τόσο η διαχωριστική ανάλυση όσο και τα δέντρα ταξινόμησης και παλινδρόμησης εφαρμόστηκαν στις πρωταρχικές τιμές των δεδομένων ενώ η ανάλυση κύριων συνιστωσών εφαρμόστηκε στις τυποποιημένες τιμές των δεδομένων (z-τιμές) ώστε να εξαλειφθεί η επίδραση των διαφορετικών μονάδων μέτρησης των προσδιοριζόμενων παραμέτρων (Simeonov και λοιποί, 2003; Liu και λοιποί, 2003).

Για την στατιστική ανάλυση των δεδομένων χρησιμοποιήθηκε το στατιστικό πρόγραμμα «SPSS v15.0 for Windows».

11.3. ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ

11.3.1 Ανάλυση κύριων συνιστωσών

Η ανάλυση κύριων συνιστωσών είναι μια ισχυρή τεχνική αναγνώρισης προτύπων της οποίας στόχος είναι να εξηγηθεί η διακύμανση ενός μεγάλου συνόλου δεδομένων διά-συσχετισμένων μεταβλητών και να μετασχηματιστούν σε ένα μικρότερο σύνολο ανεξάρτητων μεταβλητών (κύριες συνιστώσες). Η ανάλυση κύριων συνιστωσών εκτελέστηκε σε όλο το σύνολο δεδομένων επεξεργασμένου νερού των τριών ΜΕΝ ώστε να συγκριθούν τα συνθετικά πρότυπα μεταξύ αυτών και για να προσδιοριστούν οι παράγοντες που επηρεάζουν κάθε ομάδα (Σχήμα 11.2).



Σχήμα 11.2. Κατανομή των δύο κύριων συνιστωσών που παρατηρήθηκαν για τις τρεις μονάδες επεξεργασίας νερού (□: Αχαρναί, +: Γαλάτσι, Δ: Πολυδένδρι).

Τόσο η μέθοδος περιστροφής των αξόνων (Varimax) όσο και η μέθοδος χωρίς περιστροφή των αξόνων έδωσαν παρόμοια αποτελέσματα για το μοντέλο των φορτίσεων. Για το λόγο αυτόν, παρουσιάζεται μόνο η αρχική επίλυση, χωρίς περιστροφή, ως μέσο απεικόνισης της αρχικής δομής δεδομένων. Εξήχθησαν δύο κύριες συνιστώσες (PCs) με ιδιοτιμές μεγαλύτερες από τη μονάδα οι οποίες επεξηγούν το 75,55% της συνολικής διακύμανσης του σετ δεδομένων επεξεργασμένου νερού (Πίνακας 11.3).

Πίνακας 11.3. Πίνακας φορτίσεων των παραγόντων για τα δεδομένα επεξεργασμένου νερού των ΜΕΝ

Παράμετρος	PC1	PC2
EC	0,96	0,14
TA	0,55	0,66
pH	-0,61	0,26
Al	-0,70	0,58
TH	0,91	0,10
Cl ⁻	0,96	-0,01
Free Cl	-0,23	0,76
Ca ²⁺	0,82	0,33
Mg ²⁺	0,80	-0,14
% της διακύμανσης	57,95	17,60

Στον Πίνακα 11.3, οι τιμές με έντονους χαρακτήρες υποδεικνύουν ισχυρή συσχέτιση μεταξύ μιας μεταβλητής και μιας κύριας συνιστώσας.

Η πρώτη κύρια συνιστώσα, ερμηνεύει το 57,95% της συνολικής διακύμανσης και παρουσιάζει υψηλή θετική συσχέτιση με τις μεταβλητές EC, TH, Cl⁻, Ca²⁺ και Mg²⁺ ενώ παρουσιάζει αρνητική συσχέτιση με το Al. Αυτή η διπολική κύρια συνιστώσα αντιπροσωπεύει την αντίθεση μεταξύ του παράγοντα της αλατότητας και του αργιλίου το οποίο χρησιμοποιείται για το στάδιο της κροκίδωσης (προσθήκη θειϊκού αργιλίου) κατά την επεξεργασία του νερού. Η δεύτερη κύρια συνιστώσα ερμηνεύει το 17,60% της συνολικής διακύμανσης και συσχετίζεται ισχυρά μόνο με το υπολειμματικό χλώριο, το οποίο προέρχεται από το στάδιο της χλωρίωσης για την απολύμανση του νερού.

Τα δείγματα επεξεργασμένου νερού από τη ΜΕΝ Γαλασίου παρουσίασαν υψηλότερες αλατότητες που αποδόθηκαν στο γεγονός ότι η συγκεκριμένη ΜΕΝ τροφοδοτείται με μικτό νερό που προέρχεται από τους ταμιευτήρες του Μαραθώνα (ο οποίος παρουσιάζει υψηλότερη αλατότητα) και του Μόρνου. Οι άλλες δύο ΜΕΝ τροφοδοτούνται με νερό αποκλειστικά από τον ταμιευτήρα Μόρνου. Στον Πίνακα 11.4 παρουσιάζονται περιγραφικά στατιστικά μέτρα των παραμέτρων ποιότητας του νερού από τους πύργους υδροληψίας των ταμιευτήρων Μόρνου και Μαραθώνα το έτος 2007.

Πίνακας 11.4. Περιληπτικά περιγραφικά στατιστικά μέτρα των χημικών παραμέτρων στους ταμιευτήρες Μόρνου και Μαραθώνα το έτος 2007

	Παράμετρος	Αριθμός δειγμάτων	Αριθμητικός μέσος	Τυπική απόκλιση	Ελάχιστη τιμή	Μέγιστη τιμή
Ταμιευτήρας Μόρνου	EC	9	270,8	4,24	262	275
	pH	9	8,2	0,11	8	8,4
	TA	9	124,4	3,32	120	130
	TH	9	137,1	3,76	132	144
	Cl ⁻	9	2,9	0,23	2,6	3,4
	Ca ²⁺	9	46,6	1,74	44	49
	Mg ²⁺	9	5,3	0,15	5,1	5,6
Ταμιευτήρας Μαραθώνα	EC	10	342,1	31,15	300	414
	pH	10	7,8	0,20	7,6	8,1
	TA	10	142,2	18,55	123	185
	TH	10	158,3	12,24	142	190
	Cl ⁻	10	11,9	2,29	8,8	16,2
	Ca ²⁺	10	52	4,50	47	63
	Mg ²⁺	10	6,8	0,46	6,4	7,9

11.3.2 Διαχωριστική ανάλυση

Η διαχωριστική ανάλυση (DA) είναι μια πολυμεταβλητή στατιστική τεχνική, σχετιζόμενη με την πολλαπλή παλινδρόμηση, μέσω της οποίας βρίσκονται γραμμικές εξισώσεις οι οποίες παρέχουν την καλύτερη διάκριση των παρατηρήσεων σε δύο ή περισσότερες φυσικά εμφανιζόμενες ομάδες. Με βάση τα δεδομένα των ποιοτικών χαρακτηριστικών του επεξεργασμένου νερού των τριών ΜΕΝ, κατασκευάστηκε ένα

μοντέλο βηματικής διαχωριστικής ανάλυσης. Η συγκεκριμένη τεχνική διαχωριστικής ανάλυσης, προσδιορίζει τις μεταβλητές ή τους δείκτες που επιτρέπουν την καλύτερη διάκριση μεταξύ δύο ή περισσότερων προκαθορισμένων ομάδων δειγμάτων (Hair και λοιποί, 1998; Johnson και Wichern, 1998).

Η διαχωριστική ανάλυση υποθέτει ότι οι κατανομές των πληθυσμών είναι κανονικές, οι πίνακες διακύμανσης-συνδιακύμανσης αυτών των πληθυσμών είναι ομοιογενείς και δεν υπάρχει πολυσυγγραμμικότητα μεταξύ των ανεξάρτητων μεταβλητών.

Για τον έλεγχο της κανονικότητας των κατανομών των μεταβλητών σε κάθε MEN ακολουθήθηκε η παρακάτω διαδικασία: Οι συντελεστές ασυμμετρίας και κύρτωσης (Πίνακας 11.2) μετασχηματίστηκαν στις αντίστοιχες τυποποιημένες τιμές τους (z -τιμές), με τη χρησιμοποίηση των τύπων (Hair et al., 1998) $z_{ασυμμετρία} = ασυμμετρία / \sqrt{6/N}$ και $z_{κύρτωση} = κύρτωση / \sqrt{24/N}$ αντίστοιχα, (όπου N το μέγεθος του δείγματος). Στη συνέχεια επαληθεύτηκε ότι οι τυποποιημένες αυτές τιμές δεν υπερέβαιναν την τιμή 1,96 η οποία είναι η κρίσιμη τιμή για την απόρριψη της μηδενικής υπόθεσης της προσέγγισης της κανονικής κατανομής. Ως αποτέλεσμα λοιπόν, θεωρήθηκε ότι τα δεδομένα ακολουθούν προσεγγιστικά την κανονική κατανομή για κάθε μεταβλητή και κάθε δεξαμενή επεξεργασμένου νερού των MEN σε επίπεδο σημαντικότητας 0,05. Επιπροσθέτως, υπήρχαν περισσότερα από 20 δείγματα σε κάθε ομάδα, γεγονός το οποίο, σύμφωνα με τον Ainsworth (2004), καθιστά την DA ανθεκτική σε πιθανές παραβιάσεις της πολυμεταβλητής κανονικότητας.

Δεδομένου ότι ο στόχος ήταν η ταξινόμηση, η οποία επηρεάζεται από μια πιθανή ανομοιογένεια των πινάκων διακύμανσης-συνδιακύμανσης των ομάδων, σχεδιάστηκε το διάγραμμα των δύο διαχωριστικών συναρτήσεων ώστε να ελεγχθεί κατά πόσο οι ομάδες έχουν περίπου ίδιο μέγεθος και διάχυση (Σχήμα 11.3). Λόγω των διαφορών που παρατηρήθηκαν, στη συνέχεια εκτελέστηκε ο έλεγχος Box's M, προκειμένου να ελεγχθεί η μηδενική υπόθεση σύμφωνα με την οποία οι πίνακες διακύμανσης-συνδιακύμανσης των τριών ομάδων είναι ίσοι. Ο έλεγχος Box's M ήταν στατιστικά σημαντικός σε επίπεδο σημαντικότητας 0,05 (p -τιμή < 0,001). Ως εκ τούτου, στην ταξινόμηση χρησιμοποιήθηκε εκτός της τυπικής τεχνικής που βασίζεται στη θεώρηση

ενός συνολικού κοινού πίνακα συνδιακύμανσης και η τεχνική που βασίζεται σε ξεχωριστούς πίνακες συνδιακύμανσης ώστε να ελεγχθεί εάν βελτιώνεται ή όχι η ακρίβεια της ταξινόμησης (Hair και λοιποί. 1998). Παρατηρήθηκε λοιπόν, ότι η ταξινόμηση χρησιμοποιώντας ξεχωριστούς πίνακες συνδιακύμανσης ήταν κατά περισσότερο από 2% ακριβέστερη απ' ότι χρησιμοποιώντας έναν συνολικό κοινό πίνακα. Για αυτόν το λόγο, στη συνέχεια αναφέρονται και τα δύο αποτελέσματα ταξινόμησης (Schwab, 2007).

Οι τιμές ανοχής (tolerance) για κάθε μία από τις ανεξάρτητες μεταβλητές ήταν πολύ μεγαλύτερες από 0,10, γεγονός που υποδεικνύει την έλλειψη πολυσυγγραμμικότητας των μεταβλητών.

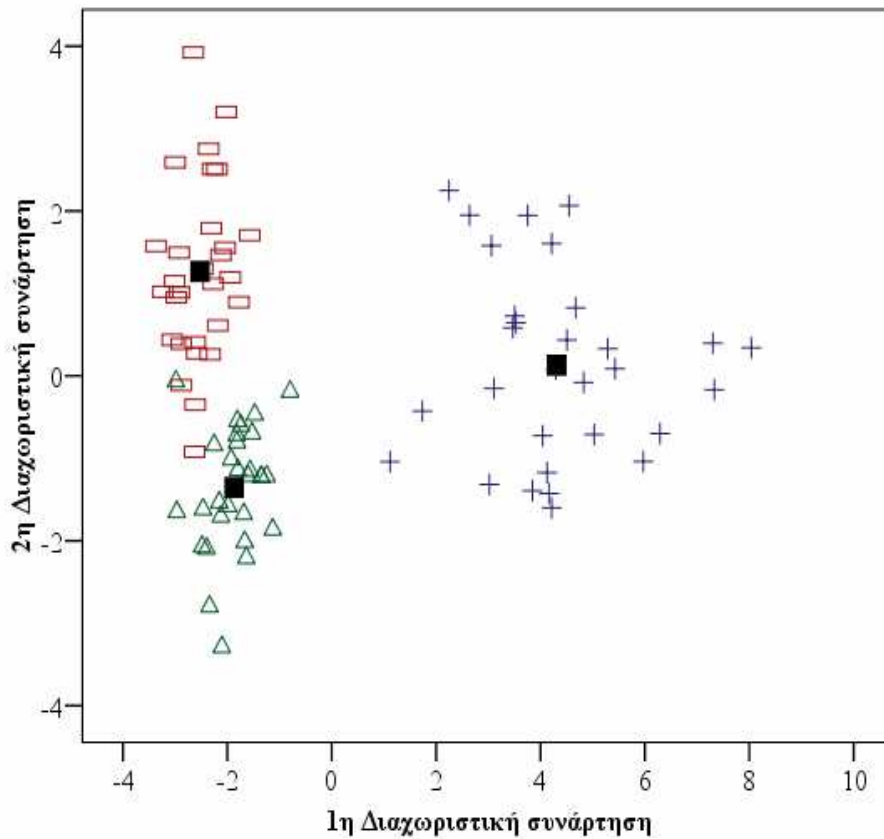
Η βηματική γραμμική DA οδήγησε στον προσδιορισμό δύο διαχωριστικών συναρτήσεων (DFs). Η στατιστική σημαντικότητα των διαχωριστικών συναρτήσεων ελέγχθηκε με τη χρησιμοποίηση του Wilks' lambda (0,042 για την DF₁ και 0,457 για την DF₂) και ελέγχους X^2 (p-τιμή<0.001 και για τις δύο διαχωριστικές συναρτήσεις). Η πρώτη διαχωριστική συνάρτηση ερμήνευσε το 89,3% της συνολικής διασποράς. Οι διαχωριστικές συναρτήσεις, χρησιμοποίησαν τελικά τις παραμέτρους CI (συσχέτιση με DF₁: r=0,923), EC (συσχέτιση με DF₁: r=0,827) και AI (συσχέτιση με DF₂: r=0,816) (Πίνακας 11.5).

Στον Πίνακα 11.5, οι τιμές με έντονους χαρακτήρες υποδεικνύουν τις μεγαλύτερες απόλυτες συσχετίσεις μεταξύ κάθε μεταβλητής και κάθε διαχωριστικής συνάρτησης. Οι παράμετροι με αστερίσκο δεν χρησιμοποιήθηκαν από το μοντέλο της βηματικής διαχωριστικής ανάλυσης.

Πίνακας 11.5. Συσχετίσεις μεταξύ των διαχωριστικών μεταβλητών και των τυποποιημένων κανονικών διαχωριστικών συναρτήσεων (βηματική μέθοδος).

Παράμετρος	Διαχωριστική συνάρτηση	
	1 ^η	2 ^η
Cl ⁻	0,923	0,093
EC	0,827	-0,025
TH*	0,383	0,144
Ca ^{2+*}	0,359	0,060
Mg ^{2+*}	0,298	0,150
Al	-0,332	0,816
pH*	-0,221	0,282
Free Cl*	0,073	0,265
TA*	0,088	0,236

Στο διάγραμμα διασποράς των δύο διαχωριστικών συναρτήσεων του Σχήματος 11.3, αναδεικνύονται οι συνάψεις των τριών ομάδων δειγμάτων των MEN με τις διαχωριστικές συναρτήσεις. Τα δείγματα της MEN Γαλατσίου είχαν θετικές τιμές για την DF₁ (οριζόντιος άξονας), ενώ τα δείγματα των δύο άλλων MEN είχαν αρνητικές τιμές για την DF₁. Οι προβλεπτικές μεταβλητές που συσχετίζονται ισχυρά με την DF₁ (EC και Cl⁻) διέκριναν τα δείγματα επεξεργασμένου νερού της MEN Γαλατσίου από τα δείγματα των δύο άλλων MEN. Η διαχωριστική συνάρτηση 2, η οποία συσχετίζεται ισχυρά με το Al, διέκρινε τα δείγματα επεξεργασμένου νερού της MEN Αχαρνών από τα δείγματα της MEN Πολυδενδρίου. Τα δείγματα της MEN Αχαρνών παρουσίασαν εν γένει θετικές τιμές για την DF₂ ενώ της MEN Πολυδενδρίου αρνητικές. Η στατιστικά σημαντικά υψηλότερη συγκέντρωση αργιλίου στο επεξεργασμένο νερό των Αχαρνών σε σχέση με αυτό του Πολυδενδρίου, μπορεί να αποδοθεί κυρίως στην αυξημένη παροχή ανεπεξέργαστου νερού στη MEN Αχαρνών, η οποία έχει ως αποτέλεσμα μικρό χρόνο παραμονής στις δεξαμενές κροκίδωσης και καθίζησης.



Σχήμα 11.3. Διάγραμμα διασποράς των δεδομένων επεξεργασμένου νερού των ΜΕΝ ως προς τις δύο διαχωριστικές συναρτήσεις (□: Αχαρναί, +: Γαλάτσι, Δ: Πολυδένδρι, ■ κέντρο ομάδας).

Στον Πίνακα 11.6, παρουσιάζονται οι συντελεστές των συναρτήσεων ταξινόμησης (Classification Function Coefficients) για κάθε δεξαμενή ΜΕΝ, βάσει του μοντέλου της βηματικής διαχωριστικής ανάλυσης. Οι συναρτήσεις είναι γραμμικές ως προς τις ανεξάρτητες μεταβλητές και οι συντελεστές έχουν υπολογιστεί με τη μέθοδο του Fischer. Κάθε «άγνωστο» νέο δείγμα νερού, κατατάσσεται στην ομάδα (ΜΕΝ) για την οποία υπολογίζεται ότι έχει το μέγιστο σκορ, βάσει της συνάρτησης ταξινόμησης:

$$f(G_i) = k_i + \sum_{j=1}^n w_{ij} p_{ij}$$

Όπου k_i είναι η σταθερά που αντιστοιχεί στην ομάδα i , n είναι ο αριθμός των μεταβλητών που χρησιμοποιούνται για την ταξινόμηση των δεδομένων σε μία ομάδα, w_{ij} είναι ο συντελεστής βαρύτητας για μία συγκεκριμένη παράμετρο p_i .

Πίνακας 11.6. Συντελεστές συναρτήσεων ταξινόμησης

	ΑΧ	ΓΑΛ	ΠΟΛ
EC	13,271	13,963	13,535
AI	-0,274	-0,358	-0,424
CI	-54,835	-49,467	-57,365
(Σταθερά)	-1727,590	-1954,245	-1774,560

Η τυπική τεχνική ταξινόμησης, απέδωσε κατά 95,5% ορθή ταξινόμηση του σετ δεδομένων των MEN (σετ εκμάθησης του μοντέλου), ενώ η διασταυρούμενη επικύρωση του μοντέλου απέδωσε κατά 94,4% ορθή ταξινόμηση (Πίνακας 11.7). Για τα δείγματα νερού των δεξαμενών δικτύου ύδρευσης (σετ ελέγχου του μοντέλου), το ποσοστό ορθής ταξινόμησής τους, όσον αφορά στην θεωρητικά αναμενόμενη προέλευσή τους, ήταν 95%.

Πίνακας 11.7. Πίνακας ταξινόμησης που προέκυψε από την τυπική διαχωριστική ανάλυση της χωρικής διακύμανσης του επεξεργασμένου νερού

Δείγμα	Παρατηρούμενη κατηγορία	Εκτιμηθείσα κατηγορία			% Ορθό	
		ΑΧ	ΓΑΛ	ΠΟΛ		
Εκμάθησης	Πρωτότυπο ^α	ΑΧ	27	0	2	93,1
		ΓΑΛ	0	29	1	96,7
		ΠΟΛ	1	0	29	96,7
	Διασταυρούμενης επικύρωσης ^β	ΑΧ	26	0	3	89,7
		ΓΑΛ	0	29	1	96,7
		ΠΟΛ	1	0	29	96,7%
Ελέγχου	Πρωτότυπο ^γ	ΑΧ	117	0	9	92,9
		ΓΑΛ	0	35	0	100,0
		ΠΟΛ	0	0	19	100,0

α: 95,5% ορθή ταξινόμηση των δεδομένων εκμάθησης

β: 94,4% ορθή ταξινόμηση των δεδομένων εκμάθησης διασταυρούμενης επικύρωσης

γ: 95,0% ορθή ταξινόμηση των δεδομένων ελέγχου

Η ταξινόμηση με τη χρησιμοποίηση ξεχωριστών πινάκων συνδιακύμανσης, ήταν περισσότερο ακριβής, συγκριτικά με την τυπική μέθοδο, κατά 2,3% για το σετ δεδομένων των MEN (97,8% ορθή ταξινόμηση) και κατά 2,8% για το σετ δεδομένων των δεξαμενών ΔΥ (97,8% ορθή ταξινόμηση) (Πίνακας 11.8).

Πίνακας 11.8. Πίνακας ταξινόμησης που προέκυψε από την διαχωριστική ανάλυση, βάσει ξεχωριστών πινάκων συνδιακύμανσης των κανονικών διαχωριστικών συναρτήσεων, της χωρικής διακύμανσης του επεξεργασμένου νερού.

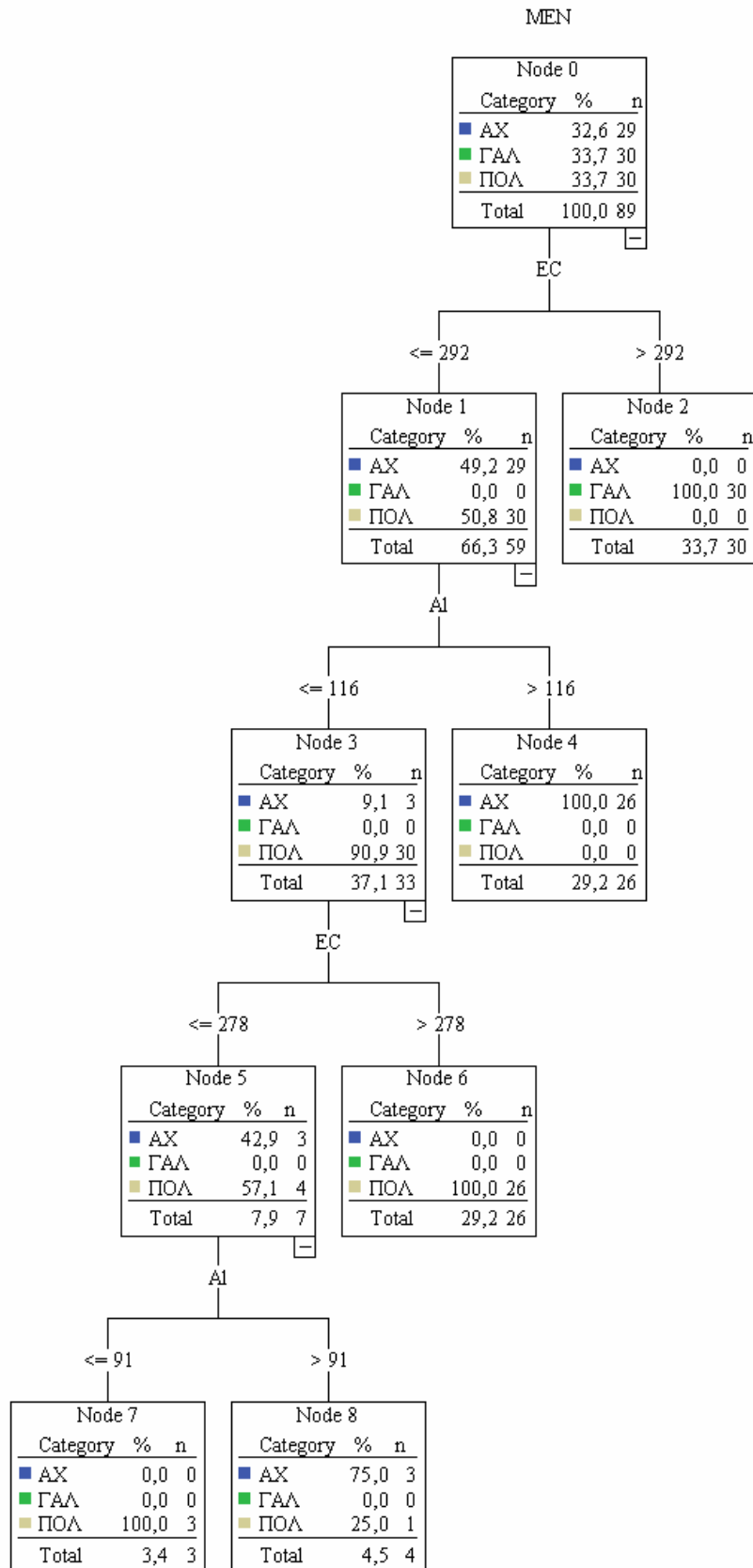
Δείγμα	Παρατηρούμενη κατηγορία	Εκτιμηθείσα κατηγορία			
		ΑΧ	ΓΑΛ	ΠΟΛ	% Ορθό
Εκμάθησης ^α	ΑΧ	28	0	1	96,6
	ΓΑΛ	0	30	0	100,0
	ΠΟΛ	1	0	29	96,7
Ελέγχου ^β	ΑΧ	122	0	4	96,8
	ΓΑΛ	0	35	0	100,0
	ΠΟΛ	0	0	19	100,0

α: 97,8% ταξινόμηση των δεδομένων εκμάθησης

β: 97,8% ορθή ταξινόμηση των δεδομένων ελέγχου

11.3.3. Δέντρα ταξινόμησης και παλινδρόμησης

Τα δείγματα επεξεργασμένου νερού των τριών ΜΕΝ συνέθεσαν το δείγμα εκμάθησης βάσει του οποίου κατασκευάστηκε το δέντρο ταξινόμησης. Με τη διαδικασία CART αποτιμήθηκαν επανακολουθητικά όλες οι ανεξάρτητες μεταβλητές για να προσδιοριστεί η μεταβλητή που έδινε τον καλύτερο διαχωρισμό (δηλαδή, η μεταβλητή που διαχώριζε τον γονικό κόμβο στους δύο αγνότερους κόμβους τέκνα) με τη χρησιμοποίηση του δείκτη μη-αγνότητας Gini (Breiman και λοιποί, 1984). Η κατασκευή του βέλτιστου δέντρου ταξινόμησης βασίστηκε στην παρακάτω προσέγγιση. Αρχικά, αναπτύχθηκε ένα πλήρες δέντρο (μέγιστου βάθους) όπου κάθε γονικός κόμβος (μη-τερματικός) επιτρεπόταν να έχει τουλάχιστον δύο μέλη. Με την διαδοχική απαλοιφή των επιπέδων του δέντρου, επιλέχτηκε τελικά ως βέλτιστο δέντρο της ακολουθίας που προέκυψε, αυτό το οποίο έδωσε το μικρότερο ποσοστό σφάλματος λαθεμένης ταξινόμησης στο σετ δεδομένων ελέγχου (δείγματα των ΔΥ).



Σχήμα 11.4. Δέντρο ταξινόμησης (αλγόριθμος CART) - Δείγμα εκμάθησης (MEN).

Ο αλγόριθμος CART κατέληξε σε ένα δέντρο με τέσσερα επίπεδα ανάπτυξης κάτω από τον ριζικό κόμβο και έδειξε ότι οι κύριες παράμετροι που χρησιμοποιήθηκαν για τον διαχωρισμό των δειγμάτων ήταν οι EC και AI (Σχήμα 11.4). Το ποσοστό ορθής ταξινόμησης του σετ δεδομένων εκπαίδευσης του μοντέλου ήταν 98,9% (Πίνακας 11.9) ενώ η 25-fold διασταυρούμενη επικύρωση του μοντέλου κατέληξε σε ένα ποσοστό 94,4% ορθής ταξινόμησης. Το γεγονός ότι τα δείγματα των ΔΔΥ ταξινομήθηκαν ορθά μόνο κατά 88,3%, σημαίνει ότι στην συγκεκριμένη περίπτωση τα δέντρα ταξινόμησης μπορεί να μην είναι τόσο επιτυχή στην ταξινόμηση αγνώστων δειγμάτων. Οποσδήποτε όμως, η προσέγγιση των δέντρων ταξινόμησης με την εξέταση μίας-μίας ανεξάρτητης μεταβλητής κάθε φορά, είναι ακόμα εξαιρετικά χρήσιμη για την σκιαγράφηση των δειγμάτων επεξεργασμένου νερού.

Πίνακας 11.9. Πίνακας ταξινόμησης των δειγμάτων μέσω CART

Δείγμα	Παρατηρούμενη κατηγορία	Εκτιμηθείσα κατηγορία			% Ορθό
		ΑΧ	ΓΑΛ	ΠΟΛ	
Εκμάθησης	ΑΧ	29	0	0	100,0
	ΓΑΛ	0	30	0	100,0
	ΠΟΛ	1	0	29	96,7
	Συνολικό ποσοστό	33,7	33,7	32,6	98,9
Ελέγχου	ΑΧ	110	0	16	87,3
	ΓΑΛ	0	30	5	85,7
	ΠΟΛ	0	0	19	100,0
	Συνολικό ποσοστό	61,1	16,7	22,2	88,3

11.3.4 Σύγκριση των προβλεπτικών μεθόδων

Τα αποτελέσματα υποστηρίζουν τον ισχυρισμό ότι η DA είναι αρκετά ανθεκτική στις παραβιάσεις των υποθέσεων (Feldesman, 2002). Ο ισχυρισμός αυτός συμβαδίζει με τον ορισμό της ανθεκτικότητας, όπως διατυπώθηκε από τους Johnson and Wichern (1998), ως «η αντίδραση στην χειροτέρευση στις τάξεις σφαλμάτων που προκαλούνται με τη χρησιμοποίηση μίας διαδικασίας ταξινόμησης σε δεδομένα τα οποία δεν συμμορφώνονται με τις υποθέσεις στις οποίες βασιζόταν η διαδικασία». Παρόλο που ο Feldesman (2002) αναφέρει ότι είναι δελεαστικό να συμπεράνει

κανείς ότι η DA είναι αρκετά ανθεκτική για όλα τα προβλήματα ταξινόμησης, η αναφορά αυτή δεν θα μπορούσε να χρησιμοποιηθεί χωρίς προηγούμενη συγκριτική ανάλυση των δεδομένων.

Όσον αφορά στα δεδομένα εκμάθησης, τα CART απέφεραν καλύτερα αποτελέσματα από την τυπική DA καθώς και την DA που βασίζεται σε ξεχωριστούς πίνακες συνδιακύμανσης, όμως η επίδοσή τους ήταν χειρότερη στον διαχωρισμό του σετ των δεδομένων ελέγχου. Παρόλα αυτά, το εξαγόμενο των CART είναι πιο εύκολο να ερμηνευθεί σε σύγκριση με τις αυστηρά αριθμητικές μεθόδους (Spruill και λοιποί, 2002).

Η διαχωριστική ανάλυση, βασισμένη σε ξεχωριστούς πίνακες συνδιακύμανσης των διαχωριστικών συναρτήσεων, απέφερε αποτελέσματα συγκρίσιμα με αυτά των CART για τον διαχωρισμό των δειγμάτων νερού των MEN και τα καλύτερα αποτελέσματα για τον διαχωρισμό των δειγμάτων νερού των δεξαμενών ΔΥ όσον αφορά στην θεωρητικά αναμενόμενη προέλευσή τους.

Τα μοντέλα που περιγράφηκαν σε αυτή την μελέτη θα μπορούσαν να θεωρηθούν ανθεκτικά υπό τον όρο ότι τόσο η ποιότητα του ανεπεξέργαστου νερού που εισέρχεται στις MEN όσο και η διαδικασίες επεξεργασίας του νερού παραμένουν σταθερές. Στην συγκεκριμένη περίπτωση, τα δεδομένα που χρησιμοποιήθηκαν απεικονίζουν τις τυπικές ετήσιες διακυμάνσεις που παρατηρούνται στην σύσταση του ανεπεξέργαστου νερού των κύριων ταμιευτήρων που τροφοδοτούν τις MEN της ΕΥΔΑΠ. Επιπλέον, μέσω της διεύθυνσης επεξεργασίας νερού, διακριβώθηκε ότι δεν λήφθηκαν ασυνήθιστα μέτρα για την επεξεργασία του ανεπεξέργαστου νερού στη διάρκεια του έτους. Ως αποτέλεσμα, τα ευρήματα της μελέτης θα μπορούσαν να θεωρηθούν αντιπροσωπευτικά της κατάστασης που διέπει το σύστημα τροφοδοσίας νερού της ΕΥΔΑΠ. Ωστόσο, όταν πρόκειται για περιπτώσεις ταμιευτήρων οι οποίοι επηρεάζονται σε μεγάλη έκταση από τις κλιματικές συνθήκες, λόγω της ασυνήθιστης μεταβλητότητας των εισροών και της ρύπανσης, η οποία με τη σειρά της απαιτεί τροποποιήσεις στη διαδικασία επεξεργασίας, τα μοντέλα θα έπρεπε σταθερά να επανεκπαιδευτούν και να επικυρωθούν.

Περαιτέρω έρευνα θα μπορούσε επίσης να συμπεριλάβει τη μελέτη περισσότερο

ασυνήθιστων παραμέτρων ειδικών προς τα τοπικά χαρακτηριστικά των νερών που μελετούνται. Ένα άλλο θέμα που θα μπορούσε να ελεγχθεί αφορά στην εποχικότητα και τη σειριακή συσχέτιση των δεδομένων. Στην περίπτωση αυτή, θα ήταν απαραίτητο ένα περισσότερο επιμηκυμένο και αυστηρό σχήμα δειγματοληψίας. Επίσης, θα παρουσίαζε ενδιαφέρον η εφαρμογή άλλων πολυμεταβλητών μεθόδων, όπως είναι τα νευρωνικά δίκτυα.

11.4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Παρατηρήθηκαν σημαντικές διαφορές στα χαρακτηριστικά του επεξεργασμένου νερού λόγω των διαφορετικών πηγών προέλευσης του επιφανειακού νερού καθώς και διαφορές στις μονάδες επεξεργασίας νερού και των λειτουργικών διαδικασιών. Αυτές οι διαφορές χρησιμοποιήθηκαν περαιτέρω για την ανάπτυξη μοντέλων διάκρισης για το επεξεργασμένο νερό που παρήχθη σε τρεις μονάδες επεξεργασίας νερού. Επιπροσθέτως, τα μοντέλα πρόβλεψης που αναπτύχθηκαν, αξιολογήθηκαν για τη διάκριση των δειγμάτων νερού του δικτύου ύδρευσης, σύμφωνα με την θεωρητικά αναμενόμενη προέλευσή τους.

Η διαχωριστική ανάλυση και τα δέντρα ταξινόμησης και παλινδρόμησης αποδείχθηκε ότι είναι ισχυρές πολυμεταβλητές τεχνικές για τη διάκριση των δειγμάτων επεξεργασμένου νερού που προέρχονται από διαφορετικές μονάδες επεξεργασίας νερού. Ο διαχωρισμός των δειγμάτων επεξεργασμένου νερού της μονάδας επεξεργασίας νερού του Γαλασίου, αποδείχθηκε ότι αποδίδεται στα χημικά χαρακτηριστικά τους τα οποία συνοψίστηκαν υπό τον όρο «αλατότητα». Η διάκριση μεταξύ των δειγμάτων επεξεργασμένου νερού από τις μονάδες επεξεργασίας νερού των Αχαρνών και του Πολυδενδρίου, αν και ήταν λιγότερο προφανής, λόγω της υδροληψίας και των δύο μονάδων από τον ίδιο ταμιευτήρα, βασίστηκε σε διαφοροποιήσεις των λειτουργικών χαρακτηριστικών των μονάδων στο στάδιο της κροκίδωσης. Η πρόβλεψη της προέλευσης του νερού στις δεξαμενές του δικτύου ύδρευσης ήταν πολύ ικανοποιητική (πάνω από 88% ορθή ταξινόμηση για όλους τους αλγορίθμους).

Η διαχωριστική ανάλυση αποδείχθηκε να είναι ανθεκτική στις παραβιάσεις των υποθέσεών της. Εντούτοις, η εφαρμογή των κατάλληλων επιδιορθώσεων βελτίωσε τη διακριτική δύναμή της. Τα δέντρα ταξινόμησης και παλινδρόμησης κατέληξαν σε μία περισσότερο ανθρώπινη προσέγγιση καθώς εξετάζουν μία-μία παράμετρο κάθε φορά.

Τα μοντέλα που αναπτύχθηκαν από αυτήν την μελέτη θα μπορούσαν να είναι χρήσιμα στην επαλήθευση της δομής του δικτύου υπό τον όρο ότι ορισμένες συνθήκες παραμένουν ίδιες, δηλαδή καμία αλλαγή στην προέλευση ανεπεξέργαστου νερού και σταθερές διαδικασίες επεξεργασίας στις μονάδες επεξεργασίας νερού. Διαφορετικά, πρέπει να επαναληφθεί το τρέξιμο των αλγορίθμων.

Οι τεχνικές αναγνώρισης προτύπων μπορούν να χρησιμοποιηθούν στη διάκριση των δειγμάτων επεξεργασμένου νερού, όσον αφορά στην προέλευσή τους, και επομένως στην καθιέρωση πρωτοκόλλων ασφάλειας στην περίπτωση μιας έκτακτης ανάγκης, όπου μια πιθανή επιδείνωση της ποιότητας νερού σε μια από τις εγκαταστάσεις θα μπορούσε να υποβιβάσει την ποιότητα του νερού των σχετικών τμημάτων του δικτύου παροχής νερού.

ΚΕΦΑΛΑΙΟ 12

ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΓΙΑ ΤΗΝ ΤΟΞΙΚΟΤΗΤΑ ΤΟΥ ΠΟΣΙΜΟΥ ΝΕΡΟΥ – ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΗΣ ΔΟΜΗΣ ΤΗΣ ΤΟΞΙΚΟΤΗΤΑΣ

12.1. ΕΙΣΑΓΩΓΗ

Η οικοτοξικολογία εξετάζει τα ενδεχομένως επιβλαβή αποτελέσματα των ανθρωπογενών χημικών ουσιών, που απελευθερώνονται στη βίωση, στους οργανισμούς στα λαμβάνοντα περιβάλλοντα. Οι οικοτοξικολογικές τεχνικές ελέγχου διαδραματίζουν έναν όλο και περισσότερο σημαντικό ρόλο στην αξιολόγηση της ποιότητας νερού. Παραδοσιακά οι απαραίτητες δοκιμές περιλαμβάνουν τη χρήση των ψαριών. Εντούτοις, η χρήση των ψαριών στη δοκιμή τοξικότητας είναι δαπανηρή και χρονοβόρα, έτσι απαιτούνται εναλλακτικές λύσεις. Οι μικροβιακές δοκιμές έχουν παρόμοιες σύνθετες βιοχημικές λειτουργίες με εκείνες των ανώτερων οργανισμών.

Η αποτίμηση της τοξικότητας του πόσιμου νερού είναι σημαντική καθώς παρέχει έναν δείκτη για την συνολική επίδραση του νερού σε ζώντες οργανισμούς. Παρόλο που δεν επιβάλλεται από την νομοθεσία, η εφαρμογή μεθόδων στατιστικού ελέγχου διεργασιών (SPC), για τις παραμέτρους ποιότητας του νερού, είναι ιδιαίτερα χρήσιμη καθώς οδηγεί στη μείωση της ανεπιθύμητης μεταβλητότητας, τον έγκαιρο εντοπισμό προβληματικών καταστάσεων και στην συστηματική βελτίωση της ποιότητας. Ο στατιστικός έλεγχος διεργασιών χρησιμοποιεί τις στατιστικές μεθόδους για να βελτιώσει την ποιότητα μιας διαδικασίας. Αυτό μπορεί να επιτευχθεί από τη συστηματική μείωση της ειδικών αιτιών μεταβλητότητας, οι οποίες είναι οι μόνες που μπορούν να ταυτοποιηθούν και εν συνεχεία να εξαλειφθούν, καθώς οι κοινές αιτίες μεταβλητότητας ενυπάρχουν σε κάθε διαδικασία και δεν μπορούν να αποφευχθούν. Ο στατιστικός έλεγχος διεργασιών στοχεύει ακριβώς στο διαχωρισμό των δύο τύπων μεταβλητότητας.

Σε αυτήν την μελέτη, αναλύονται τα δεδομένα τοξικότητας του νερού για ανθρώπινη κατανάλωση από τρεις δεξαμενές επεξεργασμένου νερού της ΕΥΔΑΠ. Δύο από αυτές τις δεξαμενές είναι στο Μενίδι (Αχαρνές) (παλαιά και νέα δεξαμενή επεξεργασμένου νερού) και η τρίτη είναι στον Ασπρόπυργο. Το επεξεργασμένο νερό σε κάθε μια από αυτές τις δεξαμενές προέρχεται από τις εγκαταστάσεις επεξεργασίας νερού της "παλαιάς μονάδας Μενιδίου", της "νέας μονάδας Μενιδίου" και της "μονάδας Ασπροπύργου" αντίστοιχα. Το ανεπεξέργαστο νερό στις ανωτέρω εγκαταστάσεις επεξεργασίας νερού προέρχεται αποκλειστικά από τον ταμιευτήρα του Μόρνου.

Οι τυπικές τεχνικές SPC και η μέθοδος που βασίζεται στις χρονολογικές σειρές εφαρμόστηκαν στα δεδομένα. Αυτή η εφαρμογή ανέδειξε τις σοβαρές επιδράσεις της αυτοσυσχέτισης όταν τα τυπικά διαγράμματα ελέγχου SPC εφαρμόζονται σε αυτοσυσχετισμένες παρατηρήσεις. Το πρόβλημα της αυτοσυσχέτισης ξεπεράστηκε με τη χρήση της μεθόδου των Alwan και Roberts (1988).

12.2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

12.2.1. Αναλυτικός τοξικολογικός έλεγχος του επεξεργασμένου νερού

Τα δεδομένα τοξικότητας από την παλαιά δεξαμενή επεξεργασμένου νερού του Μενιδίου (μεταβλητή PDM), την νέα δεξαμενή επεξεργασμένου νερού του Μενιδίου (μεταβλητή NDM) και τη δεξαμενή επεξεργασμένου νερού του Ασπροπύργου (μεταβλητή ASP) καταχωρήθηκαν χρησιμοποιώντας ένα τεστ βιολογικής ακτινοβολίας, το οποίο βασίζεται στη συσχέτιση μεταξύ της τοξικότητας του δείγματος νερού και της επίδρασής της στην ένταση του εκπεμπόμενου φωτός των θαλάσσιων βακτηρίων *Vibrio fischeri* (παλαιότερα ονομαζόμενα *Photobacterium phosphoreum*), μετρούμενη με το bioluminometer Microtox (Strategic Diagnostics Microtox 1010, 2003). Η αρχές της μεθόδου μέτρησης της τοξικότητας, έχουν ήδη περιγραφεί στο Κεφ. 8, παρ. 8.2.

Τα διαθέσιμα δεδομένα αφορούσαν σε 163 ημερήσια δείγματα (διαδοχικών ημερών) μεγέθους $n=1$ για κάθε μία από τις δεξαμενές του Μενιδίου και 91 δείγματα από τη δεξαμενή Ασπροπύργου. Όλες οι δοκιμές πραγματοποιήθηκαν βάσει των

τυποποιημένων πρωτόκολλων που περιγράφονται στο εγχειρίδιο οξείας τοξικότητας της Microtox (Microtox Acute Toxicity Users' Guide). Χρησιμοποιήθηκε η διαδικασία της διπλής δοκιμής. Με τη διπλή δοκιμή, η ποιότητα των δεδομένων βελτιώνεται μέσω της διασταυρούμενης σύγκρισης. Τα αποτελέσματα του διπλού ελέγχου θεωρούνται μη αποδεκτά όταν οι μετρήσεις της φωταύγειας διαφέρουν μεταξύ τους περισσότερο από 20%. Η μέση τιμή των δύο αναγνώσεων χρησιμοποιείται σε όλους τους επόμενους υπολογισμούς. Οι δοκιμές πραγματοποιήθηκαν από τον ίδιο αναλυτή υπό τις ίδιες λειτουργικές συνθήκες και την ίδια ημέρα παραλαβής του δείγματος νερού. Σε συμμόρφωση με το τεστ επαναληψιμότητας (ενδο-εργαστηριακή ακρίβεια) της Microtox, χρησιμοποιήθηκε θειικό άλας ψευδάργυρου (ως τοξικό διάλυμα αναφοράς) για την παραγωγή των στοιχείων εσωτερικού ποιοτικού ελέγχου. Ο γενικός μέσος όρος των διπλών μετρήσεων τοξικού διαλύματος αναφοράς ήταν 83,5 με μια συνολική τυπική απόκλιση (“εντός” {within} και “μεταξύ” {between}) 5,9. Αυτή είναι μια εκτίμηση της μεταβλητότητας της δοκιμής βιολογικής ακτινοβολίας κατά τη διάρκεια του χρόνου. Η τυπική απόκλιση εντός των διπλών μετρήσεων των δειγματοληψιών ύδατος ήταν 1,7.

12.2.2. Στατιστική ανάλυση

Στα δεδομένα τοξικότητας του νερού εφαρμόστηκαν αρχικά τα τυπικά διαγράμματα ελέγχου Shewhart για μεμονωμένες παρατηρήσεις (κεφάλαιο 6, υποκεφάλαιο 6.2.), με σκοπό τη διερεύνηση της αποτελεσματικότητας του συγκεκριμένου τύπου διαγραμμάτων και της ικανότητάς τους να αναδεικνύουν την ύπαρξη τυχόν συστηματικής συμπεριφοράς των χρονολογικών σειρών των δεδομένων. Διερευνήθηκε η ύπαρξη αυτοσυσχέτισης στα δεδομένα καθώς μπορεί να επηρεάσει σημαντικά τις ιδιότητες των τυπικών διαγραμμάτων ελέγχου. Στη συνέχεια τα δεδομένα αναλύθηκαν διεξοδικά υπό το πρίσμα του στατιστικού ελέγχου διεργασιών για αυτοσυσχετισμένα δεδομένα (Κεφάλαιο 7). Αναπτύχθηκαν υποδείγματα χρονολογικών σειρών με δυνατότητα βραχυχρόνιων προβλέψεων για την τοξικότητα. Επιπροσθέτως, διερευνήθηκε η αποτελεσματικότητα των διαγραμμάτων Shewhart, EWMA και CUSUM των καταλοίπων.

12.3. ΑΠΟΤΕΛΕΣΜΑΤΑ & ΣΧΟΛΙΑ

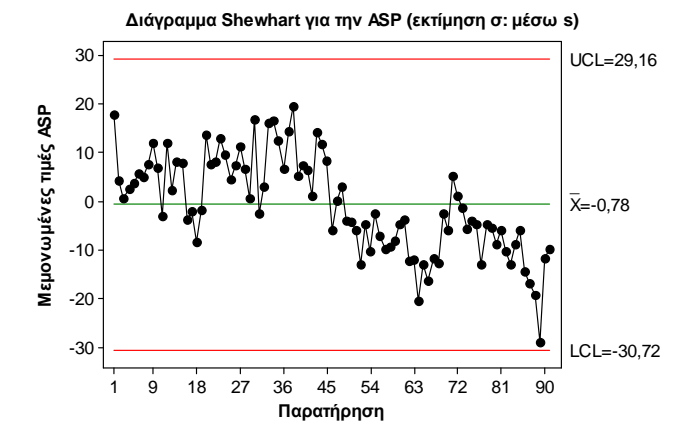
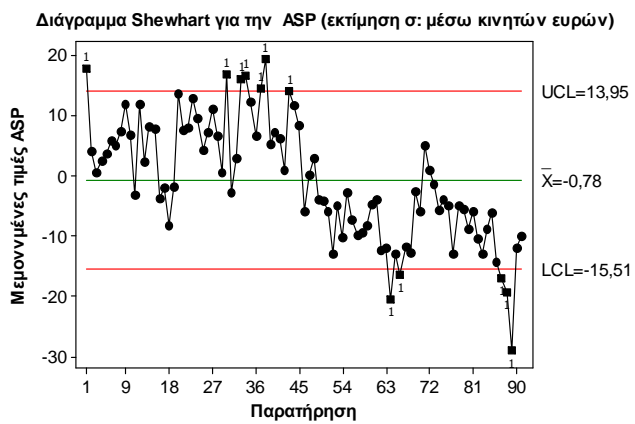
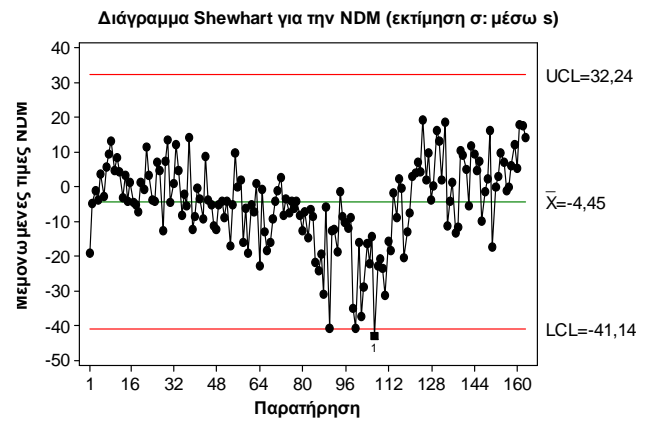
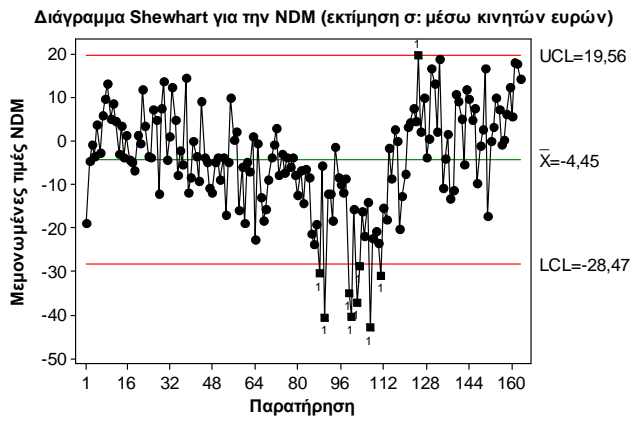
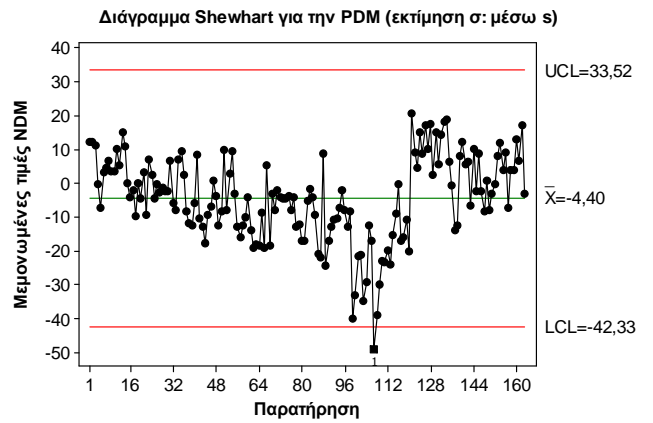
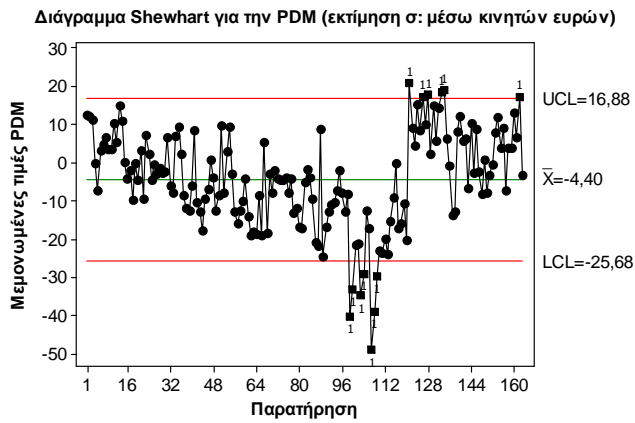
12.3.1. Τυπικά διαγράμματα ελέγχου SPC για τα δεδομένα τοξικότητας

Στον Πίνακα 12.1 παρουσιάζονται τα περιγραφικά στατιστικά μέτρα των δεδομένων τοξικότητας για τις τρεις δεξαμενές επεξεργασμένου νερού.

Πίνακας 12.1. Περιγραφικά στατιστικά μέτρα των δεδομένων τοξικότητας για PDM, NDM και ASP

	PDM	NDM	ASP
Αριθμός δειγμάτων	163	163	91
Αριθμητικός μέσος	-4,4	-4,5	-0,8
Τυπική απόκλιση	12,64	12,23	9,98
Ελάχιστη τιμή	-49	-43	-29
Διάμεσος	-4	-4	-2
Μέγιστη τιμή	20	19	19
Εύρος	69	62	48

Αρχικά, εφαρμόστηκαν στα δεδομένα τα τυπικά διαγράμματα Shewhart, για μεμονωμένες παρατηρήσεις, από τα οποία είναι δυνατόν να φανεί εάν υπάρχει κάποια συστηματική συμπεριφορά στα δεδομένα. Δεδομένου ότι δεν είναι γνωστές οι τιμές του μέσου και της τυπικής απόκλισης για καμία από τις τρεις μεταβλητές και ελλείψει εκτιμήσεων από προηγούμενα δείγματα που θεωρούντο ότι ήταν εντός ελέγχου, οι τιμές των δύο αυτών παραμέτρων, για την κάθε μεταβλητή, εκτιμώνται από τα διαθέσιμα δεδομένα.



Σχήμα 12.1. Διαγράμματα ελέγχου Shewhart (I Charts) των αρχικών δεδομένων για τις δεξαμενές PDM, NDM και ASP. Η εκτίμηση της σ βασίζεται στο μέσο κινητό εύρος ή στη δειγματική τυπική απόκλιση.

Τα τυπικά διαγράμματα Shewhart των αρχικών δεδομένων (Σχήμα 12.1), χρησιμοποιώντας τα κινητά εύρη (\overline{MR}/d_2) για την εκτίμηση της τυπικής απόκλισης σ , εντόπισαν αρκετά σημεία εκτός ελέγχου σε καθεμία δεξαμενή επεξεργασμένου νερού (13 στην παλαιά δεξαμενή Μενιδίου, 9 στη νέα δεξαμενή Μενιδίου και 12 στη δεξαμενή Ασπροπύργου). Από την άλλη, τα τυπικά διαγράμματα Shewhart των αρχικών δεδομένων, χρησιμοποιώντας τη δειγματική τυπική απόκλιση (s/c_4) για την εκτίμηση της τυπικής απόκλισης σ , έδωσαν μόνο από ένα σημείο εκτός ορίων ελέγχου στις PDM και NDM και κανένα στην ASP (Σχήμα 12.1). Στην πρώτη περίπτωση υπάρχουν πολλά ψευδώς θετικά σήματα ενώ στη δεύτερη μπορεί να εμφανίζονται μερικά ψευδώς αρνητικά σήματα. Στην περίπτωση κατά την οποία τα δεδομένα δεν είναι αυτοσυσχετισμένα, δεν υπάρχει ουσιαστική διαφορά μεταξύ των δύο προσεγγίσεων εκτίμησης της σ (Alwan, 1992). Στην παρουσία αυτοσυσχέτισης, η εκτίμηση της σ που βασίζεται στα κινητά εύρη υποεκτιμά την σ (Ryan, 1989) δίνοντας ως αποτέλεσμα πολλές λανθασμένες εκτός ελέγχου ενδείξεις ενώ η εκτίμηση της σ που βασίζεται στη δειγματική τυπική απόκλιση υπερεκτιμά τη σ καταλήγοντας σε μη-ευαίσθητα διαγράμματα ελέγχου ακόμα κι αν υπάρχουν μεγάλες μετατοπίσεις στη διαδικασία (Wardell και λοιποί, 1994). Έτσι δεν είναι δυνατόν, με βάση τα διαγράμματα αυτά, να αποφανθεί κανείς για το αν η διεργασία είναι εντός ελέγχου. Δίνουν μόνο μια εικόνα για την τάση και την αυτοσυσχέτιση των δεδομένων.

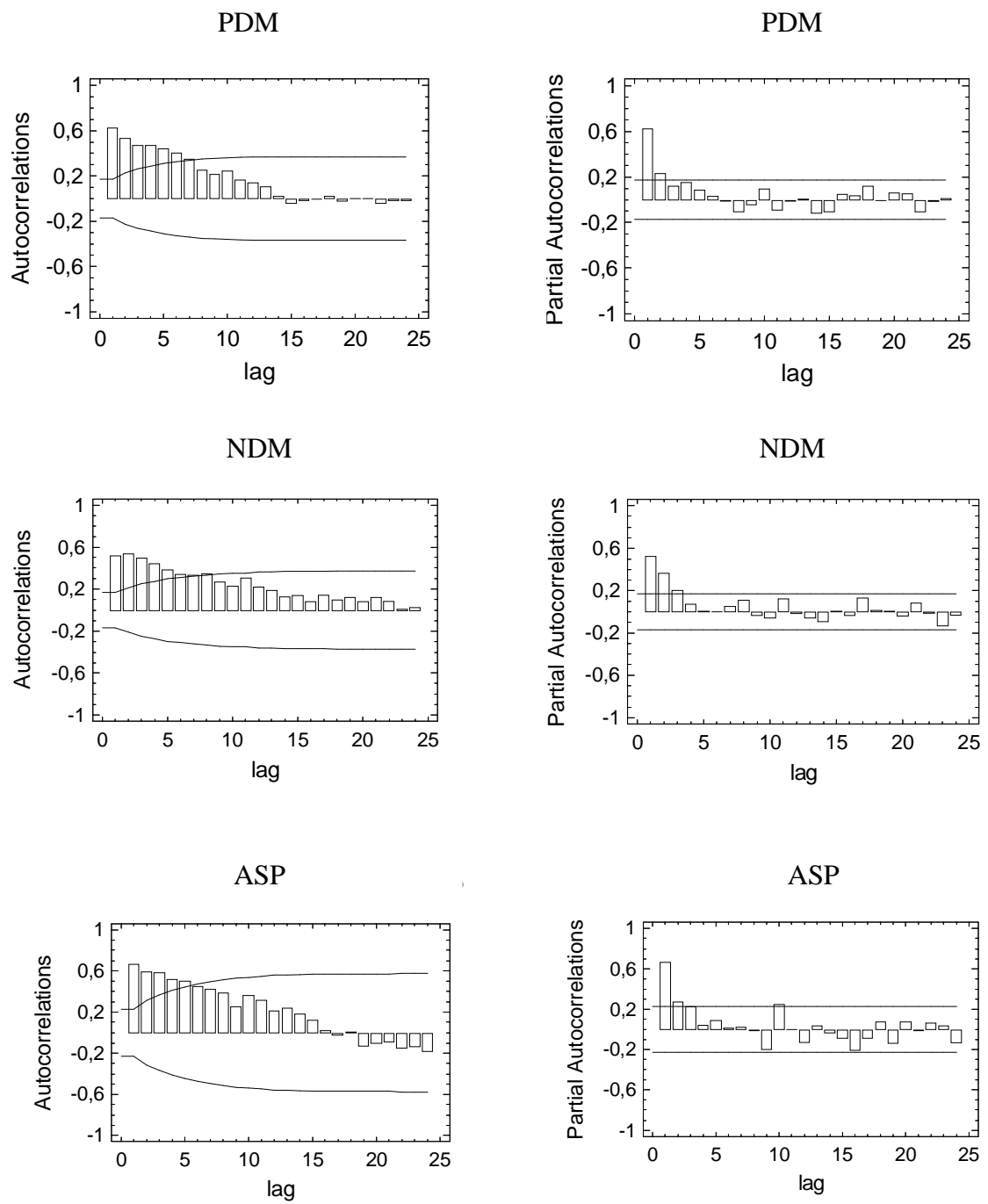
Στην συγκεκριμένη περίπτωση, η μορφή των δεδομένων στα τυπικά διαγράμματα ελέγχου SPC υποδεικνύει ότι υπάρχει ένα σημαντικό ποσό αυτοσυσχέτισης στα δεδομένα μιας και υπάρχουν αρκετές συνεχόμενες τιμές (πάνω από 8 στη σειρά) που βρίσκονται από την ίδια μεριά ως προς την κεντρική γραμμή. Η αναζήτηση των ειδικών αιτιών μεταβλητότητας γίνεται ακόμα πιο περίπλοκη εάν επιπροσθέτως εφαρμοστούν σε αυτά τα διαγράμματα ελέγχου οι κανόνες ροής (run rules) (Alwan, 1992). Επομένως, τα διαγράμματα ελέγχου Shewhart στα συγκεκριμένα δεδομένα είναι ανεπαρκή για να διακρίνουν τις κοινές αιτίες από τις ειδικές αιτίες μεταβλητότητας.

12.3.2. Προσαρμογή καταλλήλων υποδειγμάτων ARIMA στα δεδομένα

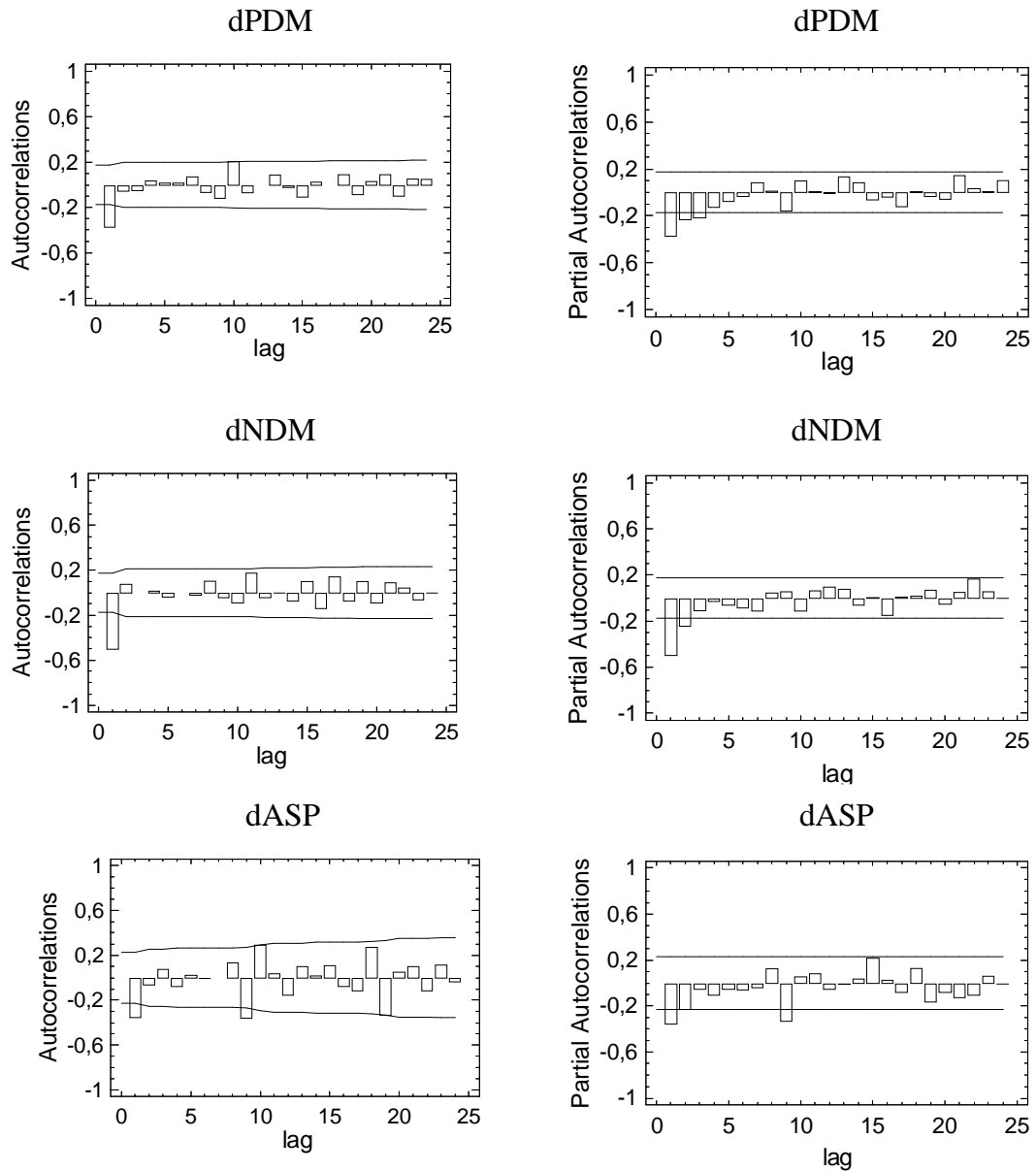
Οι πρώτες 80% εκ των συνολικών μετρήσεων κάθε μεταβλητής χρησιμοποιήθηκαν για τον προσδιορισμό και την εκτίμηση των παραμέτρων των υποδειγμάτων χρονολογικών σειρών (περίοδος εκτίμησης). Το υπόλοιπο των δεδομένων (20%) παρακρατήθηκε για να χρησιμοποιηθεί για την επικύρωση των υποδειγμάτων (περίοδος επικύρωσης). Η φύλαξη δεδομένων για λόγους επικύρωσης είναι μια πολύ σημαντική διαγνωστική δοκιμή ενός υποδείγματος. Δίνει την καλύτερη ένδειξη της ακρίβειας που μπορεί να αναμένεται κατά την πρόβλεψη μελλοντικών τιμών. Ένα καλό υπόδειγμα παρουσιάζει παρόμοια μέτρα λάθους στην περίοδο επικύρωσης με εκείνα της περιόδου εκτίμησης (αν και συχνά είναι τουλάχιστον ελαφρώς μεγαλύτερα (Nau, 2005)).

Τα διαγράμματα των συναρτήσεων δειγματικής αυτοσυσχέτισης (ACF) και των συναρτήσεων δειγματικής μερικής αυτοσυσχέτισης (PACF) με 95% όρια πιθανότητας (Σχήμα 12.2) επιβεβαιώνουν ότι οι μετρήσεις τοξικότητας του νερού είναι μέτρια έως υψηλά θετικά αυτοσυσχετισμένες σε κάθε δεξαμενή επεξεργασίας νερού.

Τα διαγράμματα των ACF και PACF δείχνουν ότι η διαδικασία δεν είναι στάσιμη για τις PDM, NDM και ASP. Η στασιμότητα επιτυγχάνεται λαμβάνοντας τις πρώτες διαφορές των χρονολογικών σειρών (διαφορές κατά διαδοχικές υστερήσεις - lags). Αυτό σημαίνει τη δημιουργία μιας νέας χρονολογικής σειράς των διαφορών πρώτης τάξης ($X_i - X_{i-1}$, όπου X_i είναι η παρατήρηση τη χρονολογική στιγμή i και X_{i-1} είναι η παρατήρηση την προηγούμενη χρονολογική περίοδο). Οι νέες χρονολογικές σειρές που παράγονται από τις πρώτες διαφορών των PDM, NDM και ASP ονομάζονται dPDM, dNDM και dASP αντίστοιχα. Οι συναρτήσεις ACF και PACF για τις νέες αυτές μεταβλητές παρουσιάζονται στο Σχήμα 12.3.



Σχήμα 12.2. Συναρτήσεις δειγματικής αυτοσυσχέτισης (autocorrelation) και δειγματικής μερικής αυτοσυσχέτισης (partial autocorrelation) για τις PDM, NDM και ASP.



Σχήμα 12.3. Συναρτήσεις δειγματικής αυτοσυσχέτισης (autocorrelation) και δειγματικής μερικής αυτοσυσχέτισης (partial autocorrelation) για τις dPDM, dNDM και dASP.

Οι συναρτήσεις ACF και PACF για τις νέες αυτές μεταβλητές (dPDM, dNDM και dASP) είναι χρήσιμα εργαλεία για να δώσουν ενδείξεις για την εξειδίκευση των υστερήσεων p και q , αυτοπαλίνδρομου και κινητού μέσου αντίστοιχα, του υποδείγματος ARIMA που πιθανόν προσαρμόζει ικανοποιητικά τα δεδομένα καθεμιάς από τις χρονολογικές σειρές. Μόνο ο πρώτος συντελεστής αυτοσυσχέτισης είναι στατιστικά σημαντικός για καθεμία από τις συναρτήσεις δειγματικής αυτοσυσχέτισης των χρονολογικών σειρών των πρώτων διαφορών, ενώ οι άλλοι δεν

διαφέρουν στατιστικά σημαντικά από το μηδέν καθώς βρίσκονται μέσα στα όρια των δύο τυπικών αποκλίσεων. Οι συντελεστές μερικής αυτοσυσχέτισης φθίνουν γρήγορα μετά τον πρώτο από αυτούς. Επομένως, ένα υπόδειγμα ARIMA(0,1,1) θα ήταν ένα ενδεδειγμένο πρότυπο προκειμένου να επεξηγηθεί η διαδικασία. Εντούτοις, διάφορα εναλλακτικά υποδείγματα (στάσιμα και μη στάσιμα) εξετάστηκαν και η τελική επιλογή των κατάλληλων υποδειγμάτων ARIMA βασίστηκε στο κριτήριο BIC και στην επίδοση των υποδειγμάτων στην περίοδο επικύρωσης. Χαμηλές τιμές του κριτηρίου BIC σε σύγκριση με τα άλλα υποδείγματα είναι ενδεικτικές ενός καλού υποδείγματος. Στον Πίνακα 12.2 παρουσιάζεται περιληπτικά η επίδοση των επιλεγέντων υποδειγμάτων τόσο για την περίοδο εκτίμησης όσο και για την περίοδο επικύρωσης. Στον πίνακα αυτόν εμφανίζονται η τετραγωνική ρίζα του τετραγωνικού σφάλματος (root squared error, RMSE), το μέσο απόλυτο σφάλμα (mean absolute error, MAE) και το μέσο σφάλμα (mean error, ME). Τα επιλεγέντα υποδείγματα έδωσαν παρόμοια αποτελέσματα για τις περιόδους εκτίμησης και επικύρωσης. Επομένως, τα επιλεγέντα υποδείγματα αναμένεται να έχουν καλή προβλεπτική ικανότητα.

Πίνακας 12.2. Στατιστικά σφαλμάτων για τα επιλεγμένα υποδείγματα χρονολογικών σειρών των δεδομένων τοξικότητας των PDM, NDM και ASP

Στατιστικό	PDM: ARIMA(0,1,1) χωρίς σταθερά		NDM: ARIMA(0,1,1) χωρίς σταθερά		ASP: ARIMA(0,1,1) χωρίς σταθερά	
	Περίοδος εκτίμησης	Περίοδος επικύρωσης	Περίοδος εκτίμησης	Περίοδος επικύρωσης	Περίοδος εκτίμησης	Περίοδος επικύρωσης
RMSE	9,33073	9,05085	9,28666	9,79153	6,41239	5,80558
MAE	6,98312	7,62017	7,32305	7,78385	5,03069	4,78155
ME	0,022535	-0,483222	0,435758	0,573015	-0,465342	-1,75819

Τα εκτιμηθέντα υποδείγματα ARIMA(0,1,1) για καθεμία από τις τρεις μεταβλητές είναι:

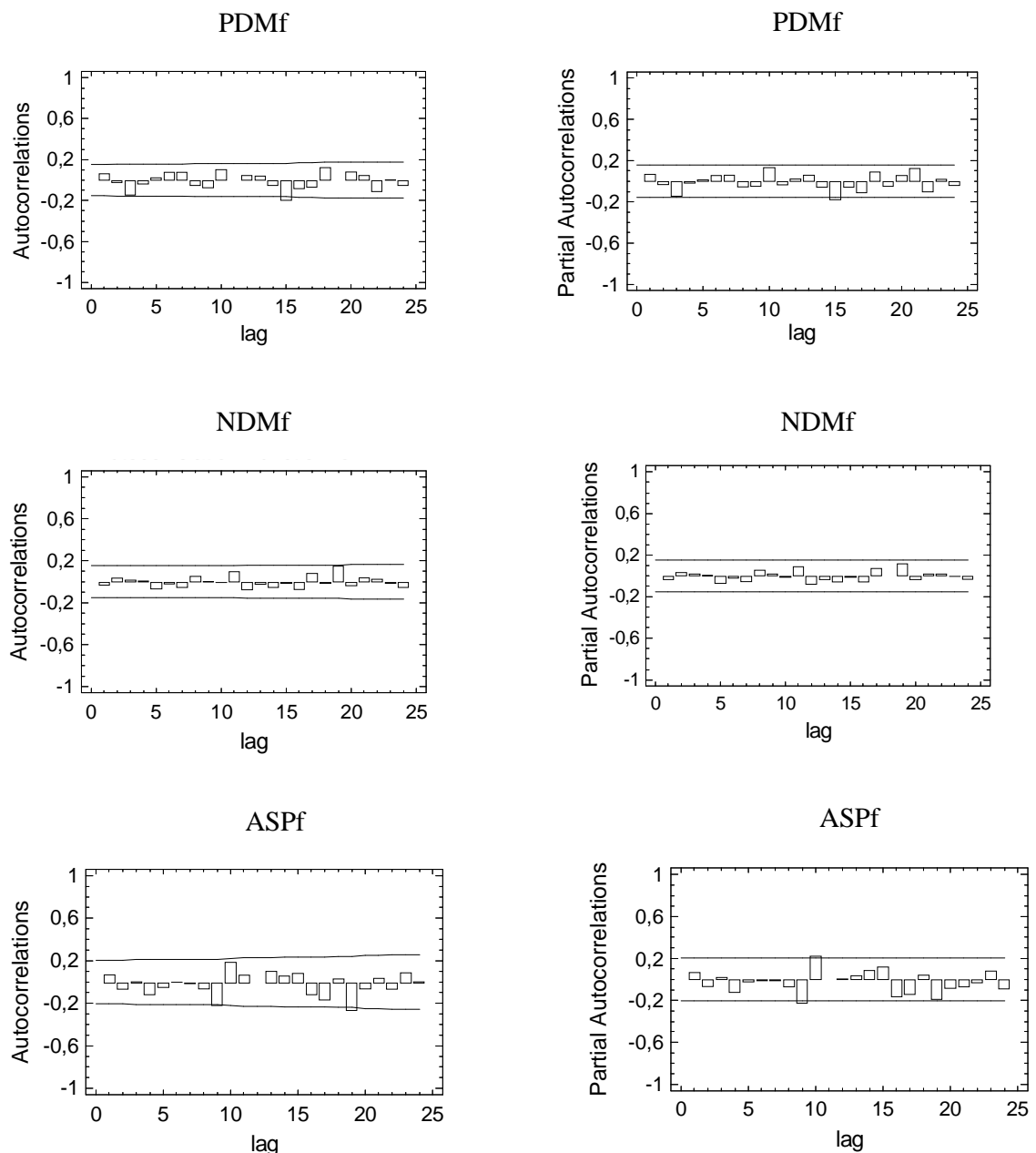
$$\text{PDM} : X_i = X_{i-1} - 0,6294e_{i-1} + e_i$$

$$\text{NDM} : X_i = X_{i-1} - 0,6954e_{i-1} + e_i$$

$$\text{ASP} : X_i = X_{i-1} - 0,5883e_{i-1} + e_i$$

Όπου, X_t είναι η τιμή τη χρονική περίοδο t , X_{t-1} είναι η τιμή της προηγούμενης χρονικής περιόδου, e_t είναι τα κατάλοιπα και e_{t-1} είναι τα κατάλοιπα της προηγούμενης χρονικής περιόδου.

Οι διαφορές μεταξύ των εκτιμηθέντων υποδειγμάτων για την τοξικότητα στις τρεις δεξαμενές επεξεργασμένου νερού μπορούν να εξηγηθούν από το γεγονός ότι, αν και το αρχικό ανεπεξεργαστο νερό προέρχεται από τον ίδιο ταμιευτήρα, υπάρχουν μικρές διαφορές στον τρόπο επεξεργασίας του νερού στις τρεις μονάδες επεξεργασίας καθώς και στο χρόνο παραμονής του επεξεργασμένου νερού σε κάθε δεξαμενή.



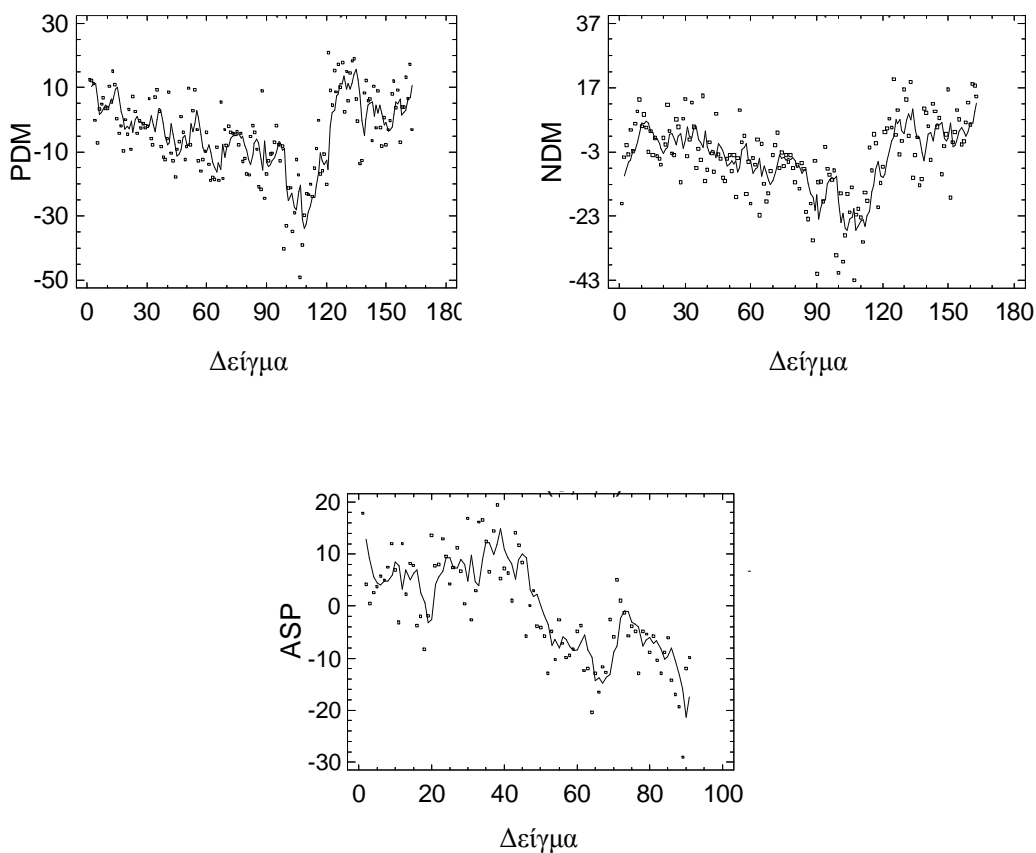
Σχήμα 12.4. Διαγράμματα συναρτήσεων δειγματικής αυτοσυσχέτισης και δειγματικής μερικής αυτοσυσχέτισης για τις PDMf, NDMf και ASPf.

Οι χρονολογικές σειρές των καταλοίπων των υποδειγμάτων που προσαρμόστηκαν στις PDM, NDM και ASP ονομάζονται PDMf, NDMf και ASPf αντίστοιχα. Τα κατάλοιπα εκπλήρωσαν όλους τους διαγνωστικούς ελέγχους.

Η p-τιμή του στατιστικού Ljung –Box ισούται με 0,934, 0,836 και 0,175 για τις PDMf, NDMf και ASPf αντίστοιχα και δηλώνει ότι δεν απορρίπτεται η μηδενική υπόθεση ότι οι αυτοσυσχετίσεις, για όλες τις υστερήσεις μέχρι την 48^η, ισούνται με μηδέν. Οι συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των καταλοίπων (Σχήμα 12.4) επιβεβαιώνουν το συμπέρασμα ότι τα κατάλοιπα δεν αυτοσυσχετίζονται. Συνεπώς, συμπεραίνεται ότι η αυτοσυσχέτιση εξαλείφθηκε. Οι έλεγχοι Kolmogorov Smirnov- Lilliefors correction και Shapiro Wilks, για την κανονικότητα των καταλοίπων, δεν ήταν στατιστικά σημαντικοί σε επίπεδο σημαντικότητας $\alpha=0,05$. Μη στατιστικά σημαντικοί ήταν και οι έλεγχοι Breusch και Pagan για την ομοσκεδαστικότητα των καταλοίπων. Επομένως, τα κατάλοιπα μπορούν να θεωρηθούν ως «λευκός θόρυβος» και συνεπώς τηρούν τις προϋποθέσεις των τυπικών διαγραμμάτων ελέγχου του SPC.

Στο Σχήμα 12.5 παρουσιάζονται τα διαγράμματα κοινών αιτιών (Common-Cause Control Chart, CCC), τα οποία είναι διαγράμματα των εκτιμηθεισών τιμών (ή προβλέψεων) που έδωσαν τα υποδείγματα ARIMA για την τοξικότητα σε καθεμία δεξαμενή επεξεργασμένου νερού. Τα διαγράμματα αυτά δίνουν μια εκτίμηση του τοπικού-τρέχοντος επιπέδου των διεργασιών. Στα ίδια διαγράμματα, απεικονίζονται οι πραγματικές τιμές των παρατηρήσεων, παρόλο που κάτι τέτοιο δεν απαιτείται από τα CCC.

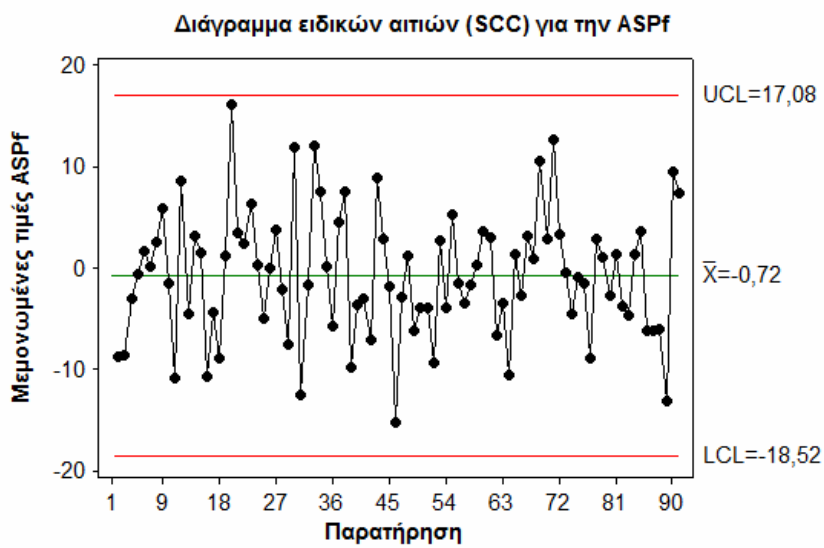
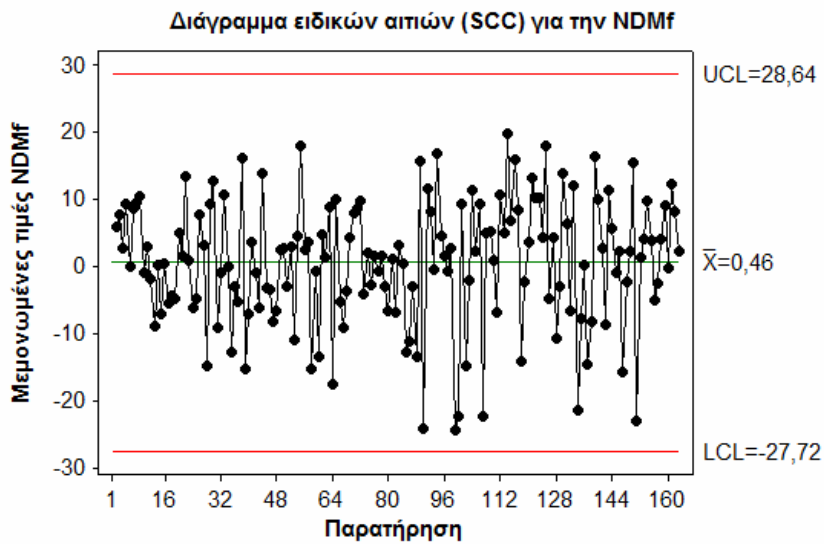
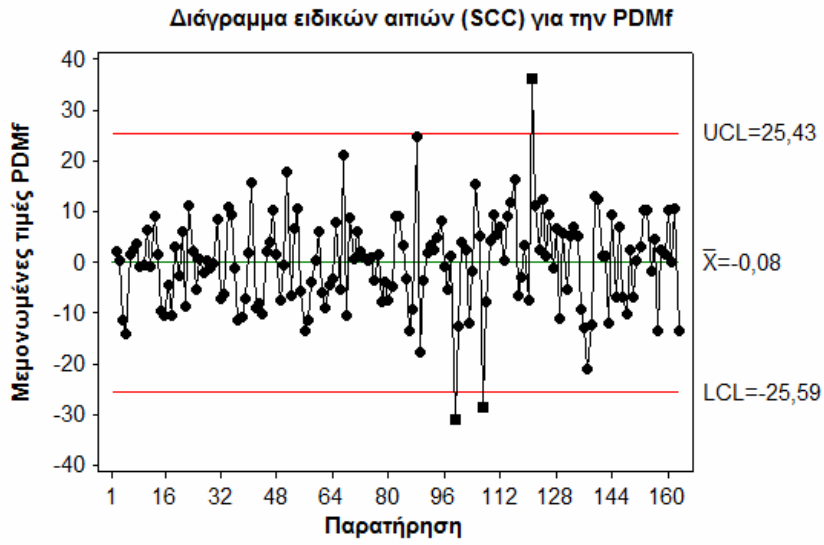
Το διάγραμμα CCC μπορεί να φανεί χρήσιμο για τον έλεγχο της διεργασίας θέτοντας σ' αυτό όρια δράσης κάτω και πάνω από το επιθυμητό επίπεδο της διεργασίας και στην περίπτωση που ληφθεί κάποιο σήμα ότι η διαδικασία τα υπερέβη, τότε θα πρέπει να γίνει διορθωτική κίνηση ώστε να επανακεντραριστεί η διεργασία στην επιθυμητή τιμή. Τα όρια αυτά διαφέρουν από τα όρια ελέγχου και μπορούν να υπολογιστούν έτσι ώστε να εξισορροπήσουν την πιθανή αναμενόμενη οικονομική απώλεια από την παραγωγή ενός «κακού» προϊόντος για κάποια συγκεκριμένη περίοδο έναντι του κόστους της επαναφοράς της διεργασίας στην επιθυμητή τιμή.



Σχήμα 12.5. Διαγράμματα CCC για τις PDM, NDM και ASP.

- Παρατηρηθείσες τιμές
- Εκτιμηθείσες τιμές

Το διάγραμμα ειδικών αιτιών (Special Cause Chart, SCC) είναι ένα διάγραμμα Shewhart για μεμονωμένες παρατηρήσεις το οποίο εφαρμόζεται στα καταλόγια. Στο Σχήμα 12.6 παρουσιάζονται τα διαγράμματα SCC των καταλοίπων (PDMf, NDMf και ASPf για τις PDM, NDM και ASP αντίστοιχα), χρησιμοποιώντας τα κινητά εύρη για την εκτίμηση της τυπικής απόκλισης. Στα διαγράμματα αυτά υπάρχουν τρία σημεία εκτός των ορίων των τριών τυπικών αποκλίσεων για την PDMf και κανένα για τις NDMf και ASPf.

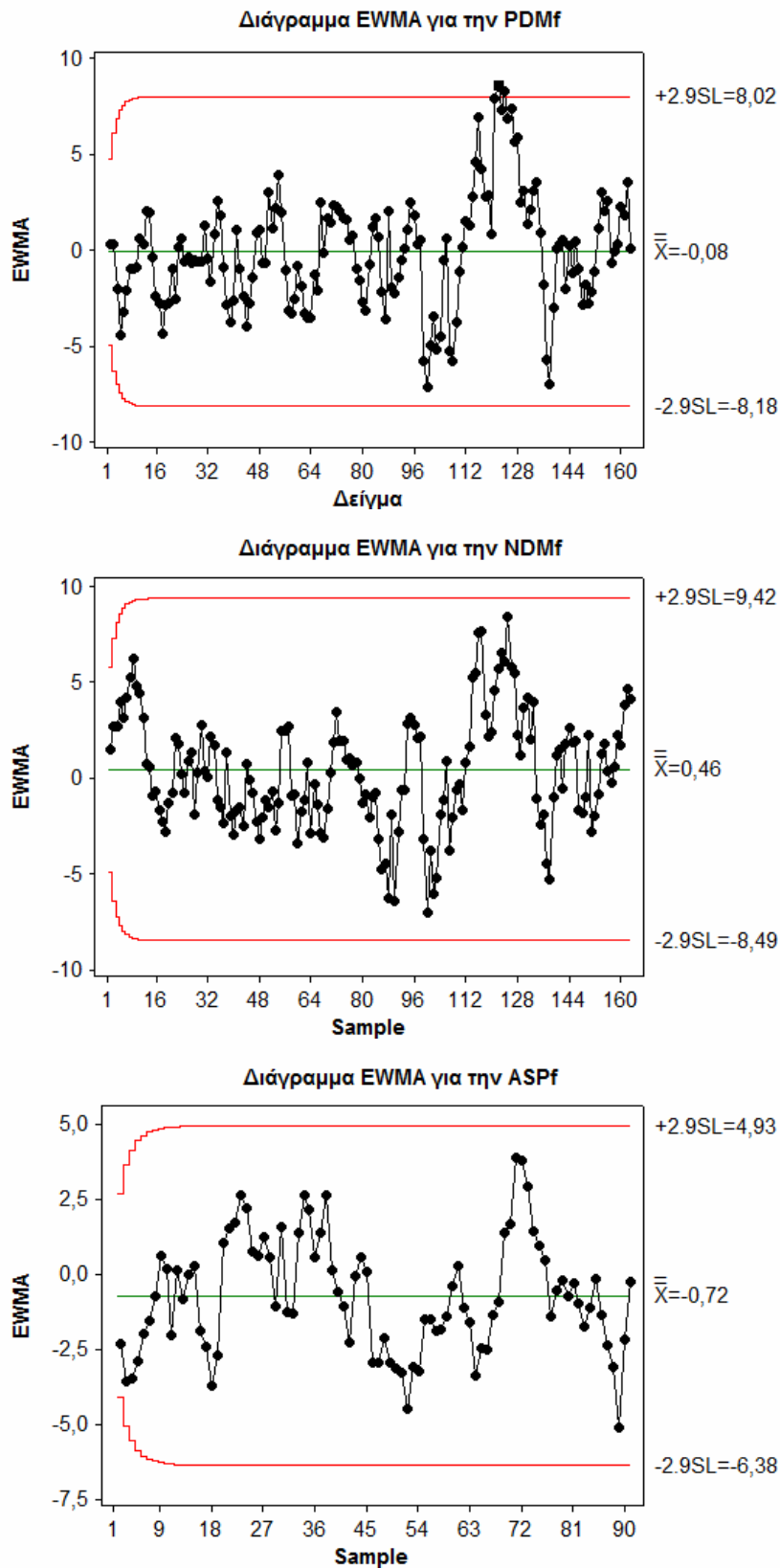


Σχήμα 12.6. Διαγράμματα ειδικών αιτιών SCC για τις PDMf, NDMf και ASPf.

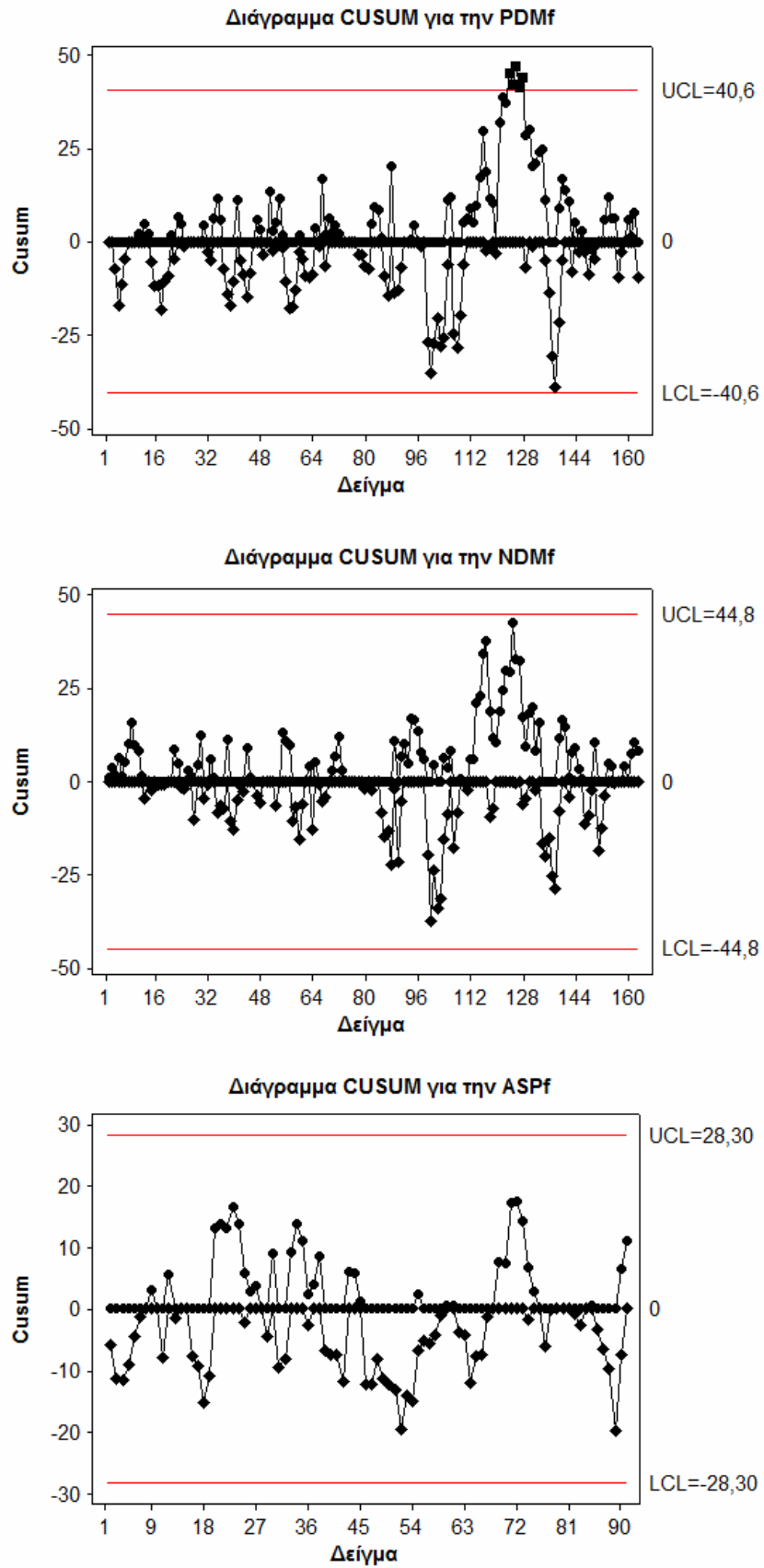
Ακόμα κι αν χρησιμοποιηθεί η δειγματική τυπική απόκλιση s για να εκτιμηθεί η τυπική απόκλιση σ , δεν υπάρχει καμία ουσιαστική διαφορά σε αυτά τα διαγράμματα. Επομένως, οι διαφορές μεταξύ των δύο τρόπων εκτίμησης των τυπικών αποκλίσεων, που παρατηρήθηκαν κατά την εφαρμογή των διαγραμμάτων Shewhart στα αρχικά δεδομένα, εξαλείφθηκαν.

Το διάγραμμα Shewhart των καταλοίπων έχει μικρή ευαισθησία στις μικρές μετατοπίσεις, όταν η διεργασία είναι θετικά αυτοσυσχετισμένη (Wardell και λοιποί, 1994). Τα διαγράμματα EWMA ή CUSUM στα κατάλοιπα αντί των Shewhart μπορούν να δώσουν καλύτερα αποτελέσματα όταν ενδιαφέρουν οι μικρές μετατοπίσεις, όπως και στην περίπτωση των τυπικών διαγραμμάτων ελέγχου του SPC (Wardell και λοιποί, 1994; Runger και λοιποί, 1995). Το διάγραμμα μεμονωμένων παρατηρήσεων Shewhart δίνει συγκριτικά πολύ μεγαλύτερες εκτός ελέγχου τιμές του ARL, για όλες τις αλλαγές στη διεργασία, σε σύγκριση με τα διαγράμματα EWMA και CUSUM με εξαίρεση μόνο τις πολύ μεγάλες αλλαγές στη διαδικασία (της τάξεως σχεδόν των τριών τυπικών αποκλίσεων), στις οποίες ανταποκρίνεται γρηγορότερα. Όρια στενότερα των τριών τυπικών αποκλίσεων στο διάγραμμα των μεμονωμένων παρατηρήσεων, δεν βελτιώνουν την ικανότητά του στον εντοπισμό μικρών μετακινήσεων της διεργασίας, καθώς μειώνεται δραματικά το εντός ελέγχου ARL, με αποτέλεσμα την μεγάλη αύξηση λανθασμένων εκτός ελέγχου ενδείξεων. Αντίστοιχη επίδραση στο εντός ελέγχου ARL έχει και η χρησιμοποίηση συμπληρωματικών κανόνων ροής στο διάγραμμα μεμονωμένων παρατηρήσεων των τριών τυπικών αποκλίσεων. Σε κάθε περίπτωση, όταν ενδιαφέρουν οι μικρές μετατοπίσεις στη διεργασία είναι προτιμότερο να χρησιμοποιείται διάγραμμα EWMA ή CUSUM (Montgomery, 2001; Champ και Woodall, 1987).

Στο Σχήμα 12.7 παρουσιάζονται τα διαγράμματα EWMA των καταλοίπων PDMf, NDMf και ASPf. Τα διαγράμματα αυτά σχεδιάστηκαν με παράμετρο $\lambda=0,2$ και για λόγους συγκρισιμότητας, χρησιμοποιήθηκαν όρια ελέγχου $L=2,859$ τυπικών αποκλίσεων ώστε να δίνουν το ίδιο εντός ελέγχου ARL με τα διαγράμματα μεμονωμένων παρατηρήσεων Shewhart των τριών τυπικών αποκλίσεων (Lu και Reynolds, 1999b), δηλαδή $ARL_0=370,4$. Τα διαγράμματα EWMA των καταλοίπων έδωσαν δύο σημεία εκτός ορίων ελέγχου για την PDMf και κανένα για τις NDMf and ASPf.



Σχήμα 12.7. Διαγράμματα EWMA για τις PDMf, NDMf και ASPf.



Σχήμα 12.8. Διαγράμματα CUSUM για τις PDMf, NDMf και ASPf.

Στο Σχήμα 12.8 παρουσιάζονται τα διαγράμματα CUSUM των καταλοίπων για τις μεταβλητές PDMf, NDMf and ASPf. Τα διαγράμματα αυτά σχεδιάστηκαν χρησιμοποιώντας $K = 0,5\sigma$. Επομένως, είναι κατάλληλα για τον εντοπισμό μετατοπίσεων τάξης μεγέθους μιας τυπικής απόκλισης. Το διάστημα απόφασης είναι $H = 4,77\sigma$ ώστε να έχουμε το ίδιο εντός ελέγχου ARL με το διάγραμμα μεμονωμένων παρατηρήσεων των τριών τυπικών αποκλίσεων (Hawkins, 1993a). Τα αμφίπλευρα διαγράμματα CUSUM των καταλοίπων έδωσαν πέντε τιμές εκτός ορίων ελέγχου για την PDMf και καμία για τις NDMf και ASP. Τα διαγράμματα CUSUM υπέδειξαν περισσότερα εκτός ορίων ελέγχου σημεία σε σύγκριση με τα διαγράμματα EWMA. Αυτά τα επιπρόσθετα εκτός ορίων ελέγχου σημεία βρίσκονταν κοντά στα όρια ελέγχου των διαγραμμάτων EWMA. Επομένως, η σύγκριση των διαγραμμάτων CUSUM με τα EWMA αναδεικνύει ότι τα διαγράμματα CUSUM είναι αποτελεσματικότερα στην ανίχνευση των μικρών μετατοπίσεων από ότι τα διαγράμματα EWMA.

Δεν υπάρχει λόγος ανησυχίας για τις εκτός ορίων ελέγχου ενδείξεις που εντοπίστηκαν από τα διαγράμματα ελέγχου, καθώς όλες οι αρχικές μετρήσεις τοξικότητας σε καθεμία δεξαμενή επεξεργασμένου νερού ήταν κάτω από 20%. Παρόλα αυτά, η απλή συμμόρφωση με τα όρια προδιαγραφής δεν είναι επαρκής για να εξασφαλιστούν καλά ποιοτικά χαρακτηριστικά (Deming, 1986). Ένας σύγχρονος ορισμός της ποιότητας είναι ότι «η ποιότητα είναι αντιστρόφως ανάλογη προς τη μεταβλητότητα» (Montgomery, 2001). Με αυτή την έννοια, η βελτίωση της ποιότητας του πόσιμου νερού σημαίνει τη μείωση της ανεπιθύμητης μεταβλητότητας. Δεδομένου ότι η μεταβλητότητα μπορεί να εκφραστεί μόνο με στατιστικούς όρους, οι στατιστικές μέθοδοι παίζουν κεντρικό ρόλο στην βελτίωση της ποιότητας. Τα διαγράμματα ελέγχου αποσκοπούν ακριβώς στον εντοπισμό των ειδικών αιτιών μεταβλητότητας. Μία ένδειξη εκτός ορίων ελέγχου υποδεικνύει ότι κάτι ασυνήθιστο συμβαίνει στη διεργασία. Με την ενδελεχή διερεύνηση της διεργασίας είναι δυνατή η ιχνηλάτιση του προβλήματος. Με αυτόν τον τρόπο μπορούν να ληφθούν διορθωτικές ενέργειες ώστε να αποτραπεί η συσσώρευση των προβλημάτων.

12.4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα διαγράμματα του τυπικού Στατιστικού Ελέγχου Διεργασιών αποδείχθηκαν αναποτελεσματικά ενώ παράλληλα οδήγησαν σε παραπλανητικά αποτελέσματα, λόγω της ύπαρξης αυτοσυσχέτισης στα δεδομένα τοξικότητας στις τρεις δεξαμενές επεξεργασμένου νερού. Κρίνονται επομένως ως ακατάλληλα για τον διαχωρισμό των ειδικών αιτιών μεταβλητότητας από τις κοινές αιτίες. Λόγω της σημαντικής αυτοσυσχέτισης των δεδομένων, μειώθηκε δραστικά η επίδοση των διαγραμμάτων του τυπικού Στατιστικού Ελέγχου Διεργασιών. Τα διαγράμματα αυτά, όταν η τυπική απόκλιση εκτιμήθηκε μέσω του μέσου κινητού εύρους, οδήγησαν σε ένα μεγάλο αριθμό ψευδών θετικών ενδείξεων ότι η διεργασία βρίσκεται εκτός ελέγχου. Αντιθέτως, όταν η τυπική απόκλιση εκτιμήθηκε μέσω της δειγματικής τυπικής απόκλισης, παρουσιάστηκε το πρόβλημα της πιθανής απόκρυψης πραγματικών θετικών ενδείξεων για εκτός ελέγχου κατάσταση.

Η πιο σύνθετη προσέγγιση των Alwan και Roberts (1988) εξάλειψε την αυτοσυσχέτιση των δεδομένων. Το διάγραμμα ειδικών αιτιών (SCC), το οποίο είναι ένα διάγραμμα μεμονωμένων παρατηρήσεων των καταλοίπων, έδωσε τρεις εκτός ορίων ελέγχου ενδείξεις για την τοξικότητα στην παλαιά δεξαμενή Μενιδίου και κανένα στις άλλες δύο δεξαμενές. Για τον γρήγορο εντοπισμό μικρών μετατοπίσεων στη διεργασία, τα διαγράμματα EWMA και CUSUM των καταλοίπων φάνηκαν περισσότερο κατάλληλα.

Τα κατάλληλα υποδείγματα για την επεξήγηση των δεδομένων τοξικότητας ήταν ARIMA(0,1,1) για καθεμία από τις τρεις δεξαμενές επεξεργασμένου νερού.

Η προσέγγιση των Alwan and Roberts approach (1988) εκμεταλλεύτηκε το γεγονός ότι τα δεδομένα είναι αυτοσυσχετισμένα ώστε να μπορεί να προβλέψει τη μελλοντική συμπεριφορά του συστήματος (Wardell et al.,1994). Η δυνατότητα πρόβλεψης της συμπεριφοράς του συστήματος για την αμέσως επόμενη χρονική περίοδο, μπορεί να χρησιμεύσει στην επαναπροσαρμογή της διεργασίας εάν αυτό κρίνεται απαραίτητο και είναι εφικτό. Επιπλέον, το διάγραμμα ειδικών αιτιών μπορεί να χρησιμοποιηθεί για τον εντοπισμό ειδικών αιτιών μεταβλητότητας, συμπεριλαμβανομένων πιθανών αλλαγών στη δομή της χρονολογικής σειράς. Η μέθοδος αυτή, που βασίζεται σε

υποδείγματα χρονολογικών σειρών, συλλαμβάνει τη δυναμική δομή των δεδομένων και δίνει λογικά διαγράμματα ελέγχου τα οποία μπορούν να αποτελέσουν χρήσιμα εργαλεία για τη βελτίωση της ποιότητας του πόσιμου νερού.

Από την ανασκόπηση της επιστημονικής βιβλιογραφίας, δεν έχει παρατηρηθεί αντίστοιχη μελέτη ανάπτυξης μεθόδων Στατιστικού Ελέγχου Διεργασιών για την παρακολούθηση της τοξικότητας του πόσιμου νερού. Ο έλεγχος της τοξικότητας μέσω των βακτηρίων *Vibrio fischeri*, αποτελεί ένα βιολογικό σύστημα έγκαιρης προειδοποίησης (Biological Early Warning System) για την ποιότητα του πόσιμου νερού (Ferguson, 2003; Gonzalez και λοιποί, 2007; Greenwood και λοιποί, 2007; Kramer, 2009; Roig και λοιποί, 2007; US Environmental Protection Agency, 2005). Υπό αυτό το πρίσμα, ο Στατιστικός Έλεγχος Διεργασιών για την τοξικότητα είναι απαραίτητος για την αποκάλυψη ασυνήθιστων μεταβολών του επιπέδου τιμών της.

ΚΕΦΑΛΑΙΟ 13

ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

Σε ένα περιβάλλον που διαρκώς αλλάζει, είτε λόγω φυσικών φαινομένων είτε λόγω ανθρωπίνων παρεμβάσεων, είναι απαραίτητη η διαρκής παρακολούθηση των υδάτινων πόρων ώστε να προληφθεί έγκαιρα οποιαδήποτε πιθανή υποβάθμιση της ποιότητάς τους. Οι δοκιμές τοξικότητας είναι πολύ σημαντικές για την ολοκληρωμένη και έγκαιρη διάγνωση και πρόληψη της ρύπανσης του νερού καθώς αποτελούν έναν δείκτη της επίδρασης του νερού σε βιολογικούς οργανισμούς. Για το λόγο αυτόν είναι κρίσιμης σημασίας η παρακολούθηση της εξέλιξης της τοξικότητας, τόσο χρονικά όσο και μεταξύ των διαφορετικών πηγών υδροληψίας του ανεπεξέργαστου νερού που προορίζεται προς ανθρώπινη κατανάλωση. Η ενδεδειγμένη στατιστική ανάλυση των τιμών τοξικότητας (εκφρασμένης ως ποσοστό αναστολής της φωταύγειας των βακτηρίων *Vibrio fischeri*) στους ταμιευτήρες Μαραθώνα και Μόρνου καθώς στη λίμνη Υλίκη για τη χρονική περίοδο 2000-2006, διεξήχθη με τη χρήση εξειδικευμένων μη-παραμετρικών στατιστικών μεθόδων. Οι τιμές της τοξικότητας στους τρεις ταμιευτήρες / λίμνες, ήταν αρκετά χαμηλές και συγκρίσιμες με την τοξικότητα που συνήθως παρατηρείται σε πόσιμα νερά. Στο διάστημα αυτό, δεν παρουσιάστηκαν στατιστικά σημαντικές ισχυρές ενδείξεις εποχικότητας. Η έλλειψη εποχικότητας στις τιμές της τοξικότητας μπορεί να αποδοθεί στις πιθανόν μικρές τιμές των παραμέτρων που μπορεί να την επηρεάζουν καθώς επίσης και στην πιθανή μικρή διακύμανσή τους εντός του ετήσιου κύκλου. Αποκαλύφθηκε μικρού μεγέθους σειριακή συσχέτιση των μηνιαίων τιμών τοξικότητας σε κάθε ταμιευτήρα, η οποία και ελήφθη υπόψη κατά τη στατιστική ανάλυση, φάνηκε όμως ότι δεν επηρεάζει τα αποτελέσματα. Μέσω του ελέγχου τάσεων Mann-Kendall, διαπιστώθηκε η έλλειψη ανοδικής ή πτωτικής τάσης των τιμών της τοξικότητας το συγκεκριμένο χρονικό διάστημα. Ο έλεγχος για ύπαρξη πιθανών διαφορών των κεντρικών μέτρων θέσης της τοξικότητας μεταξύ των τριών ταμιευτήρων / λιμνών, κατέληξε σε μη στατιστική σημαντικότητα. Συνεπώς, το νερό των τριών ταμιευτήρων μπορεί να χαρακτηριστεί ως χαμηλής τοξικότητας και σχετικά σταθερής τόσο όσον αφορά στην χρονική μεταβλητότητα της τοξικότητας όσο και στις μεταξύ τους διαφορές. Το γεγονός αυτό αποτελεί ένδειξη υψηλής ποιότητας του ανεπεξέργαστου νερού που χρησιμοποιείται για την παραγωγή πόσιμου. Επιπλέον, δεδομένου ότι η

τοξικότητα των ταμιευτήρων/λιμνών μελετήθηκε για πρώτη φορά, τίθεται η βάση για περαιτέρω παρακολούθηση των πηγών υδροληψίας και σύγκρισης των τιμών τοξικότητας στο μέλλον.

Η μικροβιακή φόρτιση του νερού των ποταμών που εκβάλλουν στους τρεις ταμιευτήρες / λίμνες ελέγχθηκε μέσω των βακτηρίων – δεικτών μικροβιακής φόρτισης *enterococci*. Η επιλογή των βακτηρίων *enterococci* ως δείκτη μικροβιακής φόρτισης έγινε βάσει της σταθερότητας των βακτηρίων, της πλήρους εργαστηριακής ανάπτυξης της μεθόδου ανάλυσής τους και της σημαντικότητας της παραμέτρου. Σύμφωνα με το ΦΕΚ 438/Β/3.7.1988 και την οδηγία 75/440/EEC (περί της απαιτούμενης ποιότητας των υδάτων επιφανείας που προορίζονται για την παραγωγή πόσιμου νερού) για τα επιφανειακά νερά κατηγορίας Α2, όπως αυτά που χρησιμοποιούνται από την ΕΥΔΑΠ, τα ανώτερα επιθυμητά όρια τίθενται στις 2000 cfu/100ml. Για την αποτίμηση της μικροβιακής φόρτισης σε βακτήρια *enterococci* των ποταμών/ρεμάτων που εκβάλλουν στους ταμιευτήρες Μαραθώνα και Μόρνου καθώς και τη λίμνη Υλίκη, χρησιμοποιήθηκαν οι μηνιαίοι προσδιορισμοί των βακτηρίων κατά τη διάρκεια μιας περιόδου δεκατριών ετών (1997 - 2009). Σε δύο από τους ποταμούς που εκβάλλουν στον ταμιευτήρα Μόρνου, στους ποταμούς Κόκκινο και Άβορο, το ποσοστό των κάτω περικεκομμένων τιμών *enterococci* (μικρότερων από 4 cfu/100ml) βρέθηκε να είναι εξαιρετικά μεγάλο (40,7% και 45,6% αντίστοιχα). Επομένως μπορούν να χαρακτηριστούν ως χαμηλού μικροβιακού φορτίου όσον αφορά στα βακτήρια εντεροκόκκων. Αντίθετα, ο μοναδικός ποταμός (Βοιωτικός Κηφισός) που εκβάλλει στην λίμνη Υλίκη, παρουσίασε υψηλό ποσοστό άνω περικεκομμένων τιμών (μεγαλύτερων από 100 cfu/100ml). Συνεπώς δεν είναι εφικτή η αξιολόγηση της μικροβιακής φόρτισης του ποταμού ως προς τα βακτήρια εντεροκόκκων. Γενικότερα, δεν μπορεί κανείς να αποφανθεί αν οι άνω περικεκομμένες τιμές των βακτηρίων είναι κάτω από τα όρια που αναφέρονται στην οδηγία 75/440/EEC. Οι τρεις παραπάνω ποταμοί εξαιρέθηκαν από τις υπόλοιπες στατιστικές αναλύσεις, που ακολούθησαν για τους υπόλοιπους, λόγω του υψηλού συνολικού ποσοστού περικεκομμένων τιμών (μεγαλύτερο από 50%). Τα μηνιαία δεδομένα εντεροκόκκων υπεβλήθησαν σε μια σειρά στατιστικών αναλύσεων με σκοπό τη διερεύνηση εποχικότητας, ύπαρξης τάσεων στα βακτήρια εντεροκόκκων και τον έλεγχο της ομοιογένειας των τάσεων. Από τους ποταμούς/ρέματα που εκβάλλουν στον ταμιευτήρα Μαραθώνα, ο ποταμός Βαρνάβας παρουσίασε αρκετά υψηλότερη

διάμεση τιμή *enterococci* (44,5 cfu/100ml) σε σχέση με το ρέμα της Σταμάτας (12 cfu/100ml). Επιπλέον, στον ποταμό Βαρνάβα υπήρχε έντονη διαφοροποίηση των *enterococci* ανά μήνα. Συγκεκριμένα, τους μήνες από Μάιο έως Νοέμβριο παρατηρήθηκαν αυξημένες τιμές αποικιών *enterococci* συγκριτικά με τους υπόλοιπους μήνες (Δεκέμβριο έως Απρίλιο). Ωστόσο με τον έλεγχο ομοιογένειας των εποχικών τάσεων (Van Belle και Hughes) δεν απορρίφθηκε η υπόθεση της ομοιογένειας των τάσεων σε κανένα από τα συνηθισμένα επίπεδα σημαντικότητας. Επιπλέον, ο εποχικός έλεγχος τάσεων Kendall για τα δεδομένα των *enterococci* στον ποταμό Βαρνάβα, ήταν μη στατιστικά σημαντικός σε οποιοδήποτε από τα συνηθισμένα επίπεδα σημαντικότητας. Για το ρέμα της Σταμάτας δεν διαπιστώθηκε ευδιάκριτη εποχικότητα στις τιμές των *enterococci*. Οι έλεγχοι για ύπαρξη τάσης, για τα δεδομένα των *enterococci* στο ρέμα της Σταμάτας, ήταν μη στατιστικά σημαντικοί σε οποιοδήποτε από τα συνηθισμένα επίπεδα σημαντικότητας. Στον ποταμό Μόρνο που εκβάλλει στον αντίστοιχο ταμιευτήρα, παρατηρήθηκε εποχικότητα στις μηνιαίες τιμές των *enterococci*. Συγκεκριμένα, παρατηρήθηκαν χαμηλές τιμές κατά τους μήνες από Ιανουάριο έως Μάιο και υψηλότερες τιμές από Ιούνιο έως Δεκέμβριο. Ο έλεγχος ομοιογένειας και τάσεων για τους εντεροκόκκους στον ποταμό Μόρνο ήταν μη στατιστικά σημαντικοί. Συνολικά λοιπόν, δεν εντοπίστηκε στατιστικά σημαντική ανοδική τάση σε κανέναν από τους ποταμούς/ρέματα για τους οποίους διερευνήθηκε η ύπαρξη τάσεων για τα βακτήρια εντεροκόκκων. Επιπλέον, το μικροβιακό φορτίο στο εσωτερικό των ταμιευτήρων/λιμνών, που οφείλεται στους εντεροκόκκους, χαρακτηρίζεται γενικά ως πολύ χαμηλό, με αύξουσα σειρά κατάταξης: Μόρνος, Υλίκη, Μαραθώνας. Πέρα όμως, από την καλή εκτίμηση όσον αφορά στην μικροβιακή φόρτιση σε εντεροκόκκους του συστήματος των τριών ταμιευτήρων, θα ήταν χρήσιμη η παράλληλη συλλογή δεδομένων ροής των εισρεόντων ποταμών/ρεμάτων για τη μελλοντική παρακολούθηση του συστήματος και τον έλεγχο τυχόν ανοδικών τάσεων που θα μπορούσαν να αποδοθούν σε ανθρωπογενείς παράγοντες. Επιπλέον, η διαδικασία απαρίθμησης των εντεροκόκκων που βασίζεται σε διαδοχικές αραιώσεις κρίνεται αναγκαία για τον προσδιορισμό των αποικιών σε δείγματα περισσότερο πυκνά σε μικροβιακό φορτίο.

Για την ερμηνεία της δομής των φυσικοχημικών ποιοτικών παραμέτρων του νερού της λίμνης Υλίκης και των ταμιευτήρων Μόρνου και Μαραθώνα και την αποτίμηση των χρονικών διακυμάνσεων της ποιότητας τους, χρησιμοποιήθηκε ένας μεγάλος

αριθμός πολυμεταβλητών στατιστικών τεχνικών - όπως είναι η ανάλυση κύριων συνιστωσών (PCA), η παραγοντική ανάλυση (FA), η ανάλυση κατά συστάδες (CA), η διαχωριστική ανάλυση (DA) και τα δέντρα ταξινόμησης και παλινδρόμησης (CART) - με ταυτόχρονη διερεύνηση της χρησιμότητάς των μεθόδων. Τα δείγματα νερού της λίμνης Υλίκης κάλυπταν τη χρονική περίοδο από τον Σεπτέμβριο του 2003 έως τον Μάιο του 2008, του ταμιευτήρα Μόρνου τα έτη από 2003 έως και 2006, ενώ τα δείγματα του ταμιευτήρα Μαραθώνα την περίοδο από Σεπτέμβριο του 2003 έως τον Σεπτέμβριο του 2007. Στην διάρκεια της περιόδου αυτής, ο ταμιευτήρας του Μαραθώνα ενισχύονταν με νερό αποκλειστικά από τον ταμιευτήρα του Μόρνου. Τα δείγματα προέρχονταν από τους πύργους υδροληψίας των ταμιευτήρων Μαραθώνα και Μόρνου και το κέντρο της λίμνης Υλίκη. Οι παράμετροι ποιότητας που εξετάστηκαν ήταν: Θολότητα (TU), πυκνότητα συγκέντρωσης υδρογόνου (pH), ηλεκτρική αγωγιμότητα (EC), ολική αλκαλικότητα (TA), ολική σκληρότητα (TH), ολικά διαλελυμένα στερεά (TDS), διαλελυμένο οξυγόνο (DO), χημικά απαιτούμενο οξυγόνο (COD), νιτρικά (NO_3^-), ασβέστιο (Ca^{2+}), αμμώνιο (NH_4^+), χλωριούχα (Cl^-), κατιόντα νατρίου (Na^+), θειικά (SO_4^{2-}), φωσφορικά (PO_4^{3-}) και κατιόντα μαγνησίου (Mg^{2+}). Οι τιμές των δεδομένων για τα φωσφορικά (PO_4^{3-}) ήταν σε υψηλό ποσοστό κάτω περικεκομμένες. Συγκεκριμένα, το ποσοστό των δειγμάτων για τα οποία οι τιμές των PO_4^{3-} ήταν κάτω του ορίου ανίχνευσης, ήταν 82,2% για τον Μαραθώνα, 95,6% για την Υλίκη και 95,7% για τον Μόρνο. Το γεγονός αυτό είναι ιδιαίτερα σημαντικό και αποτελεί ένδειξη καλής ποιότητας του νερού των ταμιευτήρων/λιμνών καθώς τα φωσφορικά ανήκουν στην κατηγορία των θρεπτικών και συνιστούν τον σημαντικότερο παράγοντα (με αμέσως επόμενο τα νιτρικά ιόντα) που προκαλεί το φαινόμενο του ευτροφισμού. Λόγω όμως των υψηλών ποσοστών περικεκομμένων τιμών, η παράμετρος PO_4^{3-} εξαιρέθηκε από τις περαιτέρω αναλύσεις. Η ανάλυση κύριων συνιστωσών/παραγοντική ανάλυση (PCA/FA) επέτρεψε τη μείωση των δεκαπέντε αρχικών μεταβλητών σε τρεις, τέσσερις και έξι σημαντικούς λανθάνοντες παράγοντες για τον Μαραθώνα, την Υλίκη και τον Μόρνο αντίστοιχα οι οποίοι επεξηγούν το μεγαλύτερο μέρος της μεταβλητότητας στην ποιότητα νερού κάθε λίμνης/ταμιευτήρα (75,6% για τον Μαραθώνα, 80,4% για την Υλίκη και 80,4% για τον Μόρνο). Επιπλέον, η PCA/FA έδωσε ικανοποιητικά αποτελέσματα για τον ανίχνευση της χρονικής-εποχικής συμπεριφοράς στο σύνολο δεδομένων για την Υλίκη και τον Μαραθώνα. Η ανάλυση συστάδων ομαδοποίησε τους δώδεκα μήνες σε δύο περιόδους για κάθε λίμνη/ταμιευτήρα βασιζόμενη στην ομοιότητα των

χαρακτηριστικών ποιότητας του νερού. Η βηματική μέθοδος διαχωριστικής ανάλυσης και ο αλγόριθμος CART των δέντρων ταξινόμησης επέτρεψαν μείωση των διαστάσεων του πολύπλοκου συνόλου δεδομένων, αναδεικνύοντας λίγες μόνο παραμέτρους-δείκτες ως κύριες υπεύθυνες για τις μεγάλες διακυμάνσεις στην ποιότητα νερού στη διάρκεια του χρόνου. Συγκεκριμένα για τον Μαραθώνα, η διάκριση των δειγμάτων νερού μεταξύ των περιόδων «Σεπτεμβρίου – Ιανουαρίου» και «Φεβρουαρίου – Αυγούστου» βάσει της βηματικής διαχωριστικής ανάλυσης επέφερε κατά 97,9% ορθή ταξινόμηση και ανέδειξε τις διακυμάνσεις των νιτρικών και του ασβεστίου ως κύριες υπεύθυνες παραμέτρους για τη διαφοροποίηση των ποιοτικών χαρακτηριστικών του νερού (μικρότερες τιμές των συγκεκριμένων παραμέτρων κατά την πρώτη περίοδο και υψηλότερες κατά τη δεύτερη). Οι τιμές των ποιοτικών χαρακτηριστικών του ταμιευτήρα Μόρνου (εκτός από το διαλελυμένο οξυγόνο), στο χρονικό διάστημα κατά το οποίο ο Μαραθώνας ενισχύεται με νερό του Μόρνου, βρέθηκαν στατιστικά σημαντικά υψηλότερες από αυτές του Μόρνου, με σημαντικότερες παραμέτρους διαφοροποίησης τα ιόντα μαγνησίου και χλωρίου. Στον ταμιευτήρα Μόρνου, αναδείχθηκαν δύο εποχικές περιόδοι, «Ιούνιος έως Δεκέμβριος» και «Ιανουάριος έως Μάϊος», οι οποίες προσεγγιστικά αντιστοιχούν στους «θερινούς – φθινοπωρινούς μήνες» και στους «χειμερινούς – εαρινούς μήνες» αντίστοιχα. Οι κύριες ποιοτικές παράμετροι που θεωρούνται ως υπεύθυνες, βάσει του μοντέλου της βηματικής διαχωριστικής ανάλυσης, για τη συνολική μεταβλητότητα της ποιότητας μεταξύ των δύο περιόδων ήταν το ασβέστιο και το διαλελυμένο οξυγόνο (χαμηλότερες τιμές κατά την περίοδο «Ιουνίου – Δεκεμβρίου» και υψηλότερες κατά την περίοδο «Ιανουαρίου – Μαΐου»). Η ορθότητα της ταξινόμησης των δειγμάτων ήταν 100%. Για τη λίμνη Υλίκη προσδιορίστηκαν επίσης δύο εποχικές περιόδοι, «Ιούνιος έως Οκτώβριος» και «Νοέμβριος έως Μάϊος», οι οποίες προσεγγιστικά αντιστοιχούν στην «ξηρή και θερμή» περίοδο του έτους και στην «υγρή και ψυχρή» περίοδο. Οι κύριες ποιοτικές παράμετροι που θεωρούνται ως υπεύθυνες, βάσει του μοντέλου της βηματικής διαχωριστικής ανάλυσης, για τη συνολική μεταβλητότητα της ποιότητας μεταξύ των δύο περιόδων ήταν το ασβέστιο και το διαλελυμένο οξυγόνο (χαμηλότερες τιμές κατά την περίοδο «Ιουνίου – Οκτωβρίου» και υψηλότερες κατά την περίοδο «Νοεμβρίου – Μαΐου»). Η ορθότητα της ταξινόμησης των δειγμάτων ήταν 92,5%. Τα υψηλά ποσοστά ορθής ταξινόμησης των μοντέλων που αναπτύχθηκαν για κάθε ταμιευτήρα υποδεικνύουν την καλή ακρίβειά τους.

Η εφαρμογή ανάλυσης κύριων συνιστωσών/παραγοντικής ανάλυσης στα δεδομένα των προσδιορισμών εννέα ποιοτικών παραμέτρων του νερού (ηλεκτρική αγωγιμότητα (EC), ολική αλκαλικότητα (TA), πυκνότητα συγκέντρωσης υδρογόνου (pH), αργίλιο (Al), ολική σκληρότητα (TH), χλωριούχα (Cl), υπολειμματικό χλώριο (free Cl), ασβέστιο (Ca²⁺) και μαγνήσιο (Mg²⁺) από τις δεξαμενές επεξεργασμένου νερού των διυλιστηρίων Αχαρνών, Γαλασίου και Πολυδενδρίου επέτρεψε σημαντική μείωση των διαστάσεων του σετ των δεδομένων σε δύο παράγοντες οι οποίοι ερμηνεύουν το 75,6% της συνολικής μεταβλητότητας των δεδομένων. Μέσω της διαχωριστικής ανάλυσης, παρατηρήθηκαν σημαντικές διαφορές στα χαρακτηριστικά του επεξεργασμένου νερού λόγω των διαφορετικών πηγών προέλευσης του επιφανειακού νερού (Μόρνος ή μεικτό νερό από Μαραθώνα – Μόρνο) καθώς και διαφορές που αποδίδονται στην τεχνική κατασκευή και τις λειτουργικές διαδικασίες κάθε μονάδας επεξεργασίας νερού (MEN). Οι κύριες παράμετροι διάκρισης των δειγμάτων επεξεργασμένου νερού της MEN Γαλασίου (στην οποία το ανεπεξέργαστο νερό που εισέρχεται είναι μεικτό Μαραθώνα – Μόρνου) από τα δείγματα των άλλων MEN (προέλευση ανεπεξέργαστου νερού: Μόρνος) ήταν οι EC και Cl⁻ (παράμετροι που ανήκουν στο σύνολο των παραμέτρων που αφορούν στην «αλατότητα» του νερού). Στο επεξεργασμένο νερό των Αχαρνών διαπιστώθηκε στατιστικά σημαντικά υψηλότερη συγκέντρωση αργιλίου σε σχέση με αυτό του Πολυδενδρίου. Το μοντέλο της βηματικής διαχωριστικής ανάλυσης απέφερε κατά 97,8% ορθή ταξινόμηση των δειγμάτων στις δεξαμενές επεξεργασμένου νερού των MEN. Το μοντέλο χρησιμοποιήθηκε στη συνέχεια για την εκτίμηση της ορθότητας ταξινόμησης των δειγμάτων από δεξαμενές του δικτύου ύδρευσης ως προς τη MEN από την οποία προέρχονταν. Η πρόβλεψη της προέλευσης του νερού στις δεξαμενές του δικτύου ύδρευσης ήταν επίσης 97,8%. Το υψηλό ποσοστό ορθής ταξινόμησης των δειγμάτων νερού των δεξαμενών δικτύου καθώς και η εγγύτητα των ποσοστών ορθής ταξινόμησης των δειγμάτων νερού MEN (βάσει των οποίων κατασκευάστηκε το μοντέλο) και δεξαμενών δικτύου πιστοποιεί την υψηλή ακρίβειά του. Συνεπώς, μπορεί να χρησιμοποιηθεί ως μέσον περιγραφής των δικτύων διανομής και πρόβλεψης της προέλευσης του νερού. Η αξιοποίηση τέτοιου είδους στατιστικών εργαλείων θα μπορούσε να βοηθήσει επικουρικά στις καθιερωμένες υδραυλικές τεχνικές στην περιγραφή των δικτύων ύδρευσης, παρέχοντας χρήσιμες πληροφορίες στους διαχειριστές των δικτύων, με αποτέλεσμα την αύξηση του βαθμού ασφαλείας. Μικρότερο αλλά αρκετά ικανοποιητικό (88,3%) ήταν και το ποσοστό ορθής

ταξινόμησης των δειγμάτων δικτύου ύδρευσης βάσει του εναλλακτικού προς τη διαχωριστική ανάλυση αλγορίθμου των δέντρων ταξινόμησης και παλινδρόμησης, τα οποία αποτελούν μία προσέγγιση πλησιέστερη προς την ανθρώπινη λογική.

Η παρακολούθηση της τοξικότητας του πόσιμου νερού παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς αποτελεί έναν συνολικό δείκτη της ποιότητάς του. Οι μέθοδοι στατιστικού ελέγχου διεργασιών αποσκοπούν στον περιορισμό της μεταβλητότητας των παραμέτρων με επακόλουθο τη βελτιστοποίηση της ποιότητας (υψηλότερη ποιότητα και σταθερότερη). Για τον στατιστικό έλεγχο διεργασιών όσον αφορά στις ημερήσιες τιμές της τοξικότητας του επεξεργασμένου νερού των ΜΕΝ Αχαρνών και Ασπροπύργου, διερευνήθηκαν οι τυπικές τεχνικές στατιστικού ελέγχου διεργασιών και οι συνδυασμοί τους με περισσότερο περίπλοκες στατιστικές μεθόδους. Η προσέγγιση των Alwan and Roberts που βασίζεται στη χρήση υποδειγμάτων χρονολογικών σειρών, όχι μόνο απέδωσε ρεαλιστικά διαγράμματα ελέγχου της τοξικότητας, τα οποία μπορούν να αποτελέσουν χρήσιμα εργαλεία για τη βελτίωση της ποιότητας του πόσιμου νερού, αλλά επιπλέον έδωσε και τη δυνατότητα για βραχύχρονες προβλέψεις των τιμών της τοξικότητας. Η δυνατότητα πρόβλεψης της συμπεριφοράς του συστήματος για την αμέσως επόμενη χρονική περίοδο, μπορεί να χρησιμεύσει στην επαναπροσαρμογή της διεργασίας εάν αυτό κρίνεται απαραίτητο και είναι εφικτό.

Στην παρούσα διατριβή, διερευνήθηκε η αποτελεσματικότητα, η επάρκεια και οι αναγκαίες προσαρμογές, λόγω των ιδιαίτερων χαρακτηριστικών των ποιοτικών παραμέτρων του νερού, ενός μεγάλου αριθμού σύνθετων στατιστικών μεθόδων με σκοπό την κατανόηση των μηχανισμών που επηρεάζουν την ποιότητα του νερού από την ακατέργαστη πρωταρχική του μορφή έως την τελική του που προορίζεται για ανθρώπινη κατανάλωση. Η μεθοδολογία που ακολουθήθηκε μπορεί να αποτελέσει έναν οδηγό προς άντληση πληροφοριών για την ποιότητα του νερού και σε άλλα συστήματα υδατικών πόρων και ύδρευσης.

Καθώς τόσο οι υδατικοί πόροι όσο και τα υδρευτικά συστήματα ενδέχεται να παρουσιάζουν δυναμικότητα ως προς τα ποιοτικά χαρακτηριστικά του νερού, η αναζήτηση πληροφοριών σε σχέση με αυτά οφείλει να είναι συνεχής. Στην μελλοντική έρευνα θα ήταν χρήσιμο να συμπεριληφθούν κι άλλες παράμετροι

ποιότητας του νερού όπως για παράδειγμα τα οργανικά παραπροϊόντα χλωρίωσης και τα βαρέα μέταλλα. Επίσης, θα παρουσίαζε ενδιαφέρον η εφαρμογή και άλλων πολυμεταβλητών μεθόδων, όπως είναι η πολυμεταβλητή ανάλυση τάσεων, ο πολυμεταβλητός έλεγχος διεργασιών και τα νευρωνικά δίκτυα.

BIBΛΙΟΓΡΑΦΙΑ

Adkins, N.C., 1993. A Framework for Development of Data Analysis Protocols for Ground Water Quality Monitoring. Technical Report No. 60, Colorado Water Resources Research Institute, Colorado State University, Fort Collins, Colorado.

Ainsworth, A. 2004. Discriminant Function Analysis. Wilkinson, L., 2000. Classification and regression trees in SYSTAT10. Vol. I. SPSS, Chicago.

Akaike, H., 1974. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, Vol. AC-19, 716-723.

Alwan, L. C., 1992. Effects of Autocorrelation on Control Chart Performance. Communications in Statistics – Theory and Methods 21, 1025-1049

Alwan, L. C. and Roberts, H.V., 1988. Time-Series Modeling for Statistical Process Control. Journal of Business and Economic Statistics 6, 87-95

APHA, AWWA, WEF, 1992. Standard Methods for the Examination of Water and Wastewater, 18th ed. American Public Health Association, Washington, DC.

Arrindell, W. A., & van der Ende, J. (1985). An Empirical-Test of the Utility of the Observations to Variables Ratio in Factor and Components Analysis. Applied Psychological Measurement, 9(2), 165- 178.

Astel, A., Biziuk, M., Przyjazny, A. and Namiesnik, J., 2006. Chemometrics in monitoring spatial and temporal variations in drinking water quality. Water Research 40(8), 1706-1716.

Ayoko, G.A., Singh, K., Balarea, S. and Kokot, S., 2007. Exploratory multivariate modeling and prediction of the physico-chemical properties of surface water and groundwater. Journal of Hydrology 336(1), 115-124.

Baralkiewicz, D., Gramowska, H., Kanecka, A., Krzyżaniak, I., Goldyn, R., 2008. Spatial distribution of major and trace elements in the water of Swarzędzkie Lake (Poland). *Environmental Monitoring and Assessment*, 143(1-3), 327-336.

Bartholomew, D. J., Steele, F., Moustaki, I., Galbraith, J. I., 2002. *The analysis and interpretation of multivariate data for social sciences*. Chapman & Hall/CRC, Florida.

Baxter, M. J., 1995. Standardization and Transformation in Principal Component Analysis, with Applications to Archaeometry. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(4), 513-527.

Baxter, M. J. and Freestone, I. C., 2006. Log-ratio Compositional Data Analysis in Archaeometry. *Archaeometry*, 48, 511–531.

Beebe, K. R., Pell, R. J., Seasholtz, M. B., 1998. *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc., USA.

Bentler, P. M., & Kano, Y., 1990. On the Equivalence of Factors and Components. *Multivariate Behavioral Research*, 25(1), 67-74.

Biggs, L. N., 1994. An investigation of hormesis using the Microtox assay. MSc Thesis, University of Illinois, Urbana, Illinois.

Bouraoui, F., Turpin, N., and Boerlen, P., 1999. Trend analysis of nutrient concentrations and loads in surface water in an intensively fertilized watershed. *Journal of Environmental Quality*, 28: 1878-1885.

Box, G. E. P., Jenkins, G. M., and Reinsel G. C., 1994. *Time Series Analysis, Forecasting and Control*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Bradley, J.V., 1968. *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, NJ.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J., 1984. Classification and Regression Trees, Chapman and Hall (Wadsworth, Inc.), New York, US.

Brookes, J.D., Antenucci, J., Hipse, M., Burch, M.D., Ashbolt, N.J. and Ferguson, C., 2004. Fate and transport of pathogens in lakes and reservoirs. *Environment International* 30(5), 741-759.

Burn, D. H. and Hag Elnur, M. A., 2002. Detection of hydrological trends and variability. *Journal of Hydrology*, 255(1-4), 107-122.

Burr, I. W., 1967. The effect of non-normality on constants for \bar{X} and R charts. *Industrial Quality Control* 23, 563-569

Çamdevýren, H., Demýr, N., Kanik, A., Keskýn, S., 2005. Use of principal component scores in multiple linear regression models for prediction of *Chlorophyll-a* in reservoirs. *Ecological Modelling* 181, 581-589.

Canna-Michaelidou S., Nicolaou A.S., Neophytou E., Christodoulidou M., 2000. The use of battery of microbiotests as a tool for integrated pollution control: evaluation and perspectives in Cyprus. Persoone G., Janssen C. R. and De Coen W. (eds). *New Microbiotests for Routine Toxicity Screening and Biomonitoring*, pp. 39-48. Kluwer Academic / Plenum Publishers, New York.

Cattel, R. B., 1966. The scree test for the number of factors. *J. Multiv. Behav. Res.*, 1, 245-276.

Champ, C.W., and Woodall W.H., 1987. Exact Results for Shewhart Control Charts with Supplementary Runs Rules. *Technometrics* 29, 393-399.

Chapman, P.M., 2002. Ecological risk assessment (ERA) and hormesis. *The Science of the Total Environment*, 288, 131-140.

Charnes, J. M., 1995. Tests for Special Causes with Multivariate Autocorrelated Data. *Computers Ops Res.*, 443-453

Cliff, N., 1988. The eigen-value greater-than-one rule and the reliability of components. *Psychological Bulletin*, 103(2), 276-279.

Clow, D. W., Mast, M. A., Campbell, D. H., 1996. Controls on surface water chemistry in the Upper Merced River Basin, Yosemite National Park, California. *Hydrological Processes*, 10, 727-746.

Conover, W. J., 1980. *Practical Nonparametric Statistics*. John Wiley & Sons, 2nd ed., New York.

Conover, W.J. and Iman, R.L., 1976. On some alternative procedures using ranks for the analysis of experimental designs. *Communication in Statistics – Theory and Methods*. A5(14):1349-1368.

Conover, W. J., Iman, R. L., 1981. Rank Transformation as a Bridge Between Parametric and Nonparametric Statistics. *The American Statistician*, 35(3), 124-129.

Conover, W.J. and Iman, Ronald L., 1980. The rank transformation as a method of discrimination with some examples. *Communications in Statistics - Theory and Methods*, 9:5, 465 — 487.

Costello, Anna B. & Jason Osborne, 2005. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7). Available online: <http://pareonline.net/getvn.asp?v=10&n=7>

Crask, M., and Perreault, W., 1977. Validation of Discriminant Analysis in Marketing Research. *Journal of Marketing Research*, 14 (February): 60-68.

Damikouka, I., Katsiri, A. and Tzia, C., 2007. Application of HACCP principles in drinking water treatment. *Desalination* 210(1), 138-145.

Daniel, W.W., 1991. *Biostatistics, a foundation for analysis in the health sciences*. 5th Ed. John Wiley and Sons, New York, NY.

Davies, J.M., 2006. 'Application and Tests of the Canadian Water Quality Index for Assessing Changes in Water Quality in Lakes and Rivers of Central North America'. *Lake and Reservoir Management*, 22: 4, 308 —320.

Davis, J.C. 1986. *Statistics and Data Analysis in Geology*. Wiley, New York.

De'Ath, G. and Fabricius, K. E., 2000. Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis. *Ecology* (by the Ecological Society of America), 81(11), 3178–3192.

De Ceballos, B. S. O., König, A., De Oliveira, J. F., 1998. Dam reservoir eutrophication : A simplified technique for a fast diagnosis of environmental degradation. *Water Research*, 32(11), 3477-3483.

Dillon, W. R., and Goldstein, M., 1984. *Multivariate Analysis: Methods and Applications*. New York: Willey.

DeLurgio, S. A., 1998. *Forecasting Principles and Applications*. Irwin/McGraw-Hill, Missouri

Deming, W.E., 1986. *Out of Crisis*. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, Mass.

Directive 98/83/EEC about the quality of water for human consumption (1998).

Directive 75/440/EEC regarding The Quality of Surface Water for the Production of Drinking Water (1975).

Dixon, W. J., Brown, M. B., Engelman, L., Fme, J. W., Hill, M. A., Jennrich, R. I., and Toporek, J. D., 1981. *BMDP Statistical Software*: Berkely. University of California Press, 725 p.

Donohue,R., Davidson, W. A., Peters, N. E., Nelson, S., and Jakowyna, B., 2001. Trends in total phosphorus and total nitrogen concentrations of tributaries to the Swan–Canning Estuary, 1987 to 1998. *Hydrological Processes*, 15, 2411–2434.

Douglas, E.M., Vogel, R.M., Kroll, C.N., 2000. Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology*, 240, 90-105.

Edberg, S. C., Leclerc, H. and Robertson, J., 1997, 'Natural protection of spring and well drinkingwater against surface microbial contamination. II. Indicators and monitoring parameters for parasites', *Critical Reviews in Microbiology* 23(2), 179–206.

Edwards, D. R., M. S. Coyne, T. C. Daniel, P. F. Vendrell, J. F. Murdoch, and P. A. Moore, Jr., 1997. Indicator Bacteria Concentrations of Two Northeast Arkansas Streams in Relation to Flow and Season. *Transactions of the ASAE* 40(1):103-109.

El-Shaarawi,A.H. and Damsleth, E.,1988. Parametric and nonparametric tests for dependent data. *Water Resources Bulletin*, 24(3):513-519.

EPA, 1992. Introduction to water quality based toxics control for the NPDES program. Environmental protection agency, Washington, DC 20460, USA, pp. 1-9.

Evans, C. D. and Jenkins, A., 2000. Surface water acidification in the South Pennines II. Temporal trends. *Environmental Pollution*, Vol. 109, Issue 1, 21-34.

EVS Consultants, 1989. 'An evaluation of the sensitivity of microassays relative to trout and daphnia acute lethality', Final Report, EVS Project No 3/064-30.

Feldesman, M.R., 2002. Classification Trees as an Alternative to Linear Discriminant Analysis. *American journal of physical anthropology* 119:257–275.

Ferguson, R., 2003. On the Alert. Reprint from Environmental Protection - June 2003 <http://www.sdimicrotox.com/pdf/On-the-alert.pdf>

Floyd, F. J., & Widaman, K. F., 1995. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.

Ford, J. K., MacCallum, R. C., & Tait, M., 1986. The Application of Exploratory Factor-Analysis in Applied- Psychology - a Critical-Review and Analysis. *Personnel Psychology*, 39(2), 291-314.

Forsythe, S. I., 2010. *Microbiology of safe food*. 2nd ed., Blackwell Publishing Ltd.

Gessner, Guy, Maholtra, N. K., Kamakura, W. A., and Zmijewski, M. E., 1988. Estimating Models with Binary Depended Variables: Some Theoretical and Empirical Observations. *Journal of Business Research*, 16(1): 49-65.

Gibbons, J.D., 1985. *Nonparametric Methods for Quantitative Analysis*. 2nd Ed. American Sciences Press, Columbus, OH.

Gilliom, R.J. and Helsel, D.R., 1986. Estimation of distributional parameters for censored trace level water quality data; 1. Estimation techniques. *Water Resources Research*, 22(2):135-146.

Gilliom, R.J., Hirsch, R.M. and Gilroy, E.J., 1984. Effect of censoring trace-level water-quality data on trend-detection capability. *Environmental Science and Technology*, 18(7):530-535.

Gonzalez, C., Prichard, E., Spinelli, S., Gille, J., Touraud, E., 2007. Validation procedure for existing and emerging screening methods *Trends in Analytical Chemistry*, Vol. 26, No. 4, 315-322.

Gorsuch, R. L., 1990. Common Factor-Analysis Versus Component Analysis - Some Well and Little Known Facts. *Multivariate Behavioral Research*, 25(1), 33-39.

Green, P. E., 1978. *Analyzing Multivariate Data*. Hinsdale, III.: Holt, Rinehart, and Winston.

Greenwood, R., Mills, G.A., Roig B., 2007. Introduction to emerging tools and their use in water monitoring. *Trends in Analytical Chemistry*, Vol. 26, No. 4, 263-267.

Guadagnoli, E., & Velicer, W. F., 1988. Relation of Sample-Size to the Stability of Component Patterns. *Psychological Bulletin*, 103(2), 265-275.

Hair, J. F. Jr, Anderson, R. E., Tatham, R. L., Black, W. C., 1998. *Multivariate data analysis*. Upper Saddle River, N.J.: Prentice-Hall. Inc., 5th edition.

Harcum, J.B., 1990. Water-quality data analysis protocol development. Ph.D. Dissertation, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.

Harris, R. J., 1975. *A Primer of Multivariate Statistics*, New York: Academic Press.

Harris, J. H. and Ross, W. H, 1991. Statistical Process Control Procedures for Correlated Observations. *The Canadian Journal of Chemical Engineering* 69, 48-57.

Hawkins, D.M., 1993a. Cumulative Sum Control Charting: An Underutilized SPC Tool. *Quality Engineering* 5, 463-467.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., Fernandez, L., 2000. Temporal evaluation of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, 34(3), 807-816.

Helsel, D. R., 1990. Less than Obvious – Statistical Treatment of Data below the Detection Limit. *Environmental Science Technology*, Vol. 24, No. 12, 1766-1774.

Helsel, D.R. and Hirsch, R.M., 1992. *Statistical Methods in Water Resources*. Elsevier, Amsterdam.

Helsel, D.R. and Hirsch, R.M., 2002. *Statistical Methods in Water Resources*. Techniques of Water-Resources Investigations of the United States Geological Survey, chap. A3, book 4.

Helsel, D.R., 1987. Advantages of nonparametric procedures for analysis of water quality data. *Hydrological Sciences - Journal*, 32(2):179-190.

Hipel, K.W., McLeod, A.I., 1994. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier Science B.V., The Netherlands.

Hirsch, R.M., Alexander, R.B. and Smith, R.A., 1991. Selection of Methods for the Detection and Estimation of Trends in Water Quality. *Water Resources Research*, 27(5):803-813.

Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982. Techniques of Trend Analysis for Monthly Water Quality Data. *Water Resources Research*, 18(1): 107-121.

Hirsch, R.M. and Slack J.R., 1984. A Nonparametric Trend Test for Seasonal Data with Serial Dependence. *Water Resources Research*, 20 (6), 727–732.

Hollander, M. and Wolfe, D.A., 1973. *Nonparametric Statistical Methods*. John Wiley and Sons, New York, NY.

Howell, J. M., M. S. Coyne, and P. Cornelius, 1995. Fecal Bacteria in Agricultural Waters of the BlueGrass Region of Kentucky. *Journal of Environmental Quality* 24: 411-419.

Hren, J., Wilson, K.S. and Helsel, D.R., 1984. A statistical approach to evaluate the relation of coal mining, land reclamation, and surface-water quality in Ohio. USGS Water-Resources Investigations Report 84-4117.

Hrudey, S.E., Hrudey, E.J. and Pollard, S.J.T., 2006. Risk management for assuring safe drinking water. *Environment International* 32(8), 948-957.

Huberty, C. J., Wisenbaker, J. W., and Smith, J. C., 1987. Assessing Predictive Accuracy in Discriminant Analysis. *Multivariate Behavioral Research* 22 (July): 307-29.

Huck, P.M. and Coffey, B.M., 2004. The importance of robustness in drinking-water systems. *Journal of toxicology and environmental health. Part A* 67(20-22), 1581-1590.

Ismail, S.S. and Ramadan, A., 1995. Characterisation of Nile and drinking water quality by chemical and cluster analysis. *Science of the Total Environment*, The 173, 69-81.

ISO 7899-2, 2000. Water quality – Detection and enumeration of intestinal enterococci – Part 2: Membrane filtration method.

ISO 8199, 2005. Water quality – General guidance on the enumeration of micro-organisms by culture.

ISO 11348-3, 1998. Water quality -- Determination of the inhibitory effect of water samples on the light emission of *Vibrio fischeri* (Luminescent bacteria test) - Part 3: Method using freeze-dried bacteria.

Jakopanec, I., Borgen, K., Vold, L., Lund, H., Forseth, T., Hannula, R. and Nygerd, K., 2008. A large waterborne outbreak of campylobacteriosis in Norway: The need to focus on distribution system safety. *BMC Infectious Diseases* 8(1), 128.

Johnson, R. A. and Wichern, D. W., 1998. *Applied Multivariate Statistical Analysis*. Prentice Hall, Inc. New Jersey.

Jolliffe, I., 1972. Discarding variables in a principal components analysis, I: artificial data. *Applied Statistics* 21, 160-173

Johnson, H. O., Gupta, S. C. Vecchia A. V. and Zvomuya F., 2009. Assessment of Water Quality Trends in the Minnesota River using Non-Parametric and Parametric Methods. *Journal of Environmental Quality*, 38: 1018-1030.

Joret, J. C., Lévi, Y., Berger, R., Nakache, F., Gilbert, M., 1986. Use of the Microtox test in monitoring the quality of raw water for the production of potable water. *J. Fr. Hydrol.*, 17(2), 143-152.

Kahya, E., Kalayci, S., 2004. Trend analysis of streamflow in Turkey. *Journal of Hydrology*, 289:128-144.

Kaiser, H. F., 1960. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.*, 20, 141-151.

Kannel, P.,R, Lee, S., Kanel, S.R., Khan, S.P., 2007. Chemometric application in classification and assessment of monitoring locations of an urban river system. *Analytica Chimica Acta* 582, 390-399.

Kaiser, K.L.E., Lum, K.R., Palabrica, V.S., 1988. Review of field applications of the Microtox test in Great Lakes and Saint Lawrence river waters. *Water Pollution Research Journal Canada* 23(2), 270-278.

Kaiser, K.L.E., Ribo, J.M. and Kwasniewska, K., 1988. A Microtox Test Survey of Lake St. Clair Water. *Water Poll. Res. J. Canada* 23 (3), pp. 356–359.

Keller-McNulty, S. and McNulty, M., 1987. The independent pairs assumption in hypothesis tests based on rank correlation coefficients. *The American Statistician*, 41(1):40-41.

Kotti, M. E., Vlessidis, A., G., Thanasoulas, N. C., Evmiridis, N. P., 2005. Assessment of River Water Quality in Northwestern Greece. *Water Resources Management*, 19: 77-94.

Koutsoyiannis, D., 2005. Nonstationarity versus scaling in hydrology. *Journal of Hydrology*, 324, 239-254.

Kramer, K. J. M., 2009. Continuous Monitoring of Waters by Biological Early Warning Systems. Rapid Chemical and Biological Techniques for Water Monitoring. Edited by Gonzalez, C., Greenwood, R. and Quevauviller, P. P. John Wiley & Sons Ltd.

Kulkarni, A. and von Storch, H., 1995. Monte Carlo experiments on the effect of serial correlation on the Mann-Kendall test of trend. *Meteorol. Zeitschrift* 4 NF 82-85.

Kumar, A., R. and Riyazuddin, P., 2008. Application of chemometric techniques in the assessment of groundwater pollution in a suburban area of Chennai city, India. *Current Science*, 94(8), 1012-1022.

Kungolos, A., Samaras, P., Kipopoulou, A. M., Zoumboulis, A. and Sakellariopoulos, G. P., 1999. Interactive toxic effects of agrochemicals on aquatic organisms. *Water Science and Technology*, 40(1), 357-364.

Lambrakis, N., Antonakos, A., Panagopoulos, G., 2004. The use of multicomponent statistical analysis in hydrogeological environmental research. *Water Research*, 38, 1862-1872.

Lehmann, A. and Rode, M., 2001. Long-term behaviour and cross-correlation water quality analysis of the river Elbe, Germany. *Water Research*, 35(9), 2153-2160.

Lettenmaier, D.P., Hooper, E.R., Wagoner, C. and Faris, K.B., 1991. Trends in stream quality in the continental United States. *Water Resources Research*, 27(3): 327-339.

Lettenmaier, D.P., Wood, E.F., Wallis, J.R., 1994. Hydroclimatological trends in the continental United States 1948-1988. *Journal of Climate* 7, 586-607.

Levene, H., 1960. "Robust tests for equality of variances". in Ingram Olkin, Harold Hotelling et al. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press. pp. 278-292.

Lévi, Y., Grimaud, A., Dutang, M., 1989. Protection of Surface Water Resources by Real Time Automated Toxicity Tests. Presented at the 5th IAWPRC Workshop, Kyoto, Japan. Microbics Reference Library #215.

Liu, C. W., Lin, K. H., Kuo, Y. M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment*, 313, 77-89.

Loehlin, J. C., 1990. Component Analysis Versus Common Factor-Analysis - a Case of Disputed Authorship. *Multivariate Behavioral Research*, 25(1), 29-31.

Loftis, J.C., McBride, G.B. and Ellis, J.C., 1991. Considerations of scale in water quality monitoring and data analysis. *Water Resources Bulletin*, 27(2):255-264.

Loftis, J. C., Porter, P. S. and Settembre G., 1987. Statistical analysis of industrial wastewater monitoring data. *Journal Water Pollution Control Federation*, 59(3): 145-151.

Loftis, J.C., Ward, R.C., Phillips, R.D. and Taylor, C.H., 1989. An evaluation of trend detection techniques for use in water quality monitoring programs. U.S. Environmental Protection Agency Report No. EPA/600/3-89/037.

Loh, W. Y., Shih, Y. S., 1997. Split selection methods for classification trees. *Stat. Sinica* 7:825-840.

Lu, C. W and Reynolds, M. R.. Jr, 1999a. EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology* 31, 166-188

Lu, C.W. and Reynolds, M.R. Jr, 1999b. Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes. *Journal of Quality Technology* 31, 259-274.

Lu, C.W. and Reynolds, M.R. Jr, 2001. CUSUM Charts for Monitoring an Autocorrelated Process. *Journal of Quality Technology*, 33, 316-334.

Lucas, J. M., 1982. Combined Shewhart-CUSUM Quality Control Schemes. *Journal of Quality Technology* 14, 51-59.

Lucas, J. M., and Crosier, R. B., 1982. Fast Initial Response for CUSUM Quality Control Schemes. *Technometrics* 24, 199-205.

Lucas, J.M. and Saccucci, M.S., 1990. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32, 1-12.

MacCallum, R. C., & Tucker, L. R., 1991. Representing Sources of Error in the Common-Factor Model - Implications for Theory and Practice. *Psychological Bulletin*, 109(3), 502-511.

Mahmud, S.G., Shamsuddin, S.A.J., Ahmed, M.F., Davison, A., Deere, D. and Howard, G. (2007) Development and implementation of water safety plans for small water supplies in Bangladesh: benefits and lessons learned. *Journal of water and health* 5(4), 585-597.

Manley, B., 1994. *Multivariate statistical methods*. 2nd edition. Chapman & Hall, London.

Massart D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L., 1988. *Chemometrics: a text book*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands.

Massart, D. L., and Kaufman, L., 1983. *The interpretation of analytical chemical data by the use of cluster analysis*. John Wiley & Sons, New York.

Mazlum, N., Özer, A., Mazlum, S., 1999. Interpretation of Water Quality Data by Principal Components Analysis. *Turkish Journal of Engineering and Environmental Science*, 23, 19-26.

McBride, G.B. and Smith, D.G., 1997. Results of a Trend Assessment of New Zealand's National River Water Quality Network. Proceedings of Monitoring Tailor_Made II: An international workshop on information strategies in water management, Data Management, 135-142.

McGill, R., Tukey, J.W., and Larsen, W.A., 1978. Variations of box plots. American Statistician, Vol. 32, No. 1, 12-16.

McLeod, A.I., Hipel, K.W., and Bodo, B., A., 1991. Trend Analysis Methodology for Water Quality Time Series. Environmetrics, 2(2), 169-200.

Means, E.G. III and Krasner, S.W., 1993. D-DBP Regulation: Issues and Ramifications. Journal AWWA, 85(2), 68-73.

Ministry of the Environment, Ontario, 2006. Assessment Report: Draft guidance modules – For use with Source Protection Technical Studies.

Mitchell, A.W., Reghenzani, J.R., and Furnas, M.J., 2001. Nitrogen levels in the Tully River – A long-term view. Water Science and Technology, 43: 99-105.

Moisen, G. G., (2008). Classification and Regression Trees. Ecological Informatics, 582-588.

Mondal, P., Majumder, C.B. and Mohanty, B., 2006. Laboratory based approaches for arsenic remediation from contaminated water: Recent developments. Journal of Hazardous Materials 137(1), 464-479.

Montgomery, D.C., 2001. Introduction to Statistical Quality Control, 4th ed. John Wiley and Sons, Inc., N. York.

Montgomery, D.C. and Mastrangelo C.M., 1991. Some Statistical Process Control Methods for Autocorrelated Data. Journal of Quality Technology 23, 179-193.

Montgomery, R.H., Loftis, J.C. and Harris, J., 1987. Statistical characteristics of ground-water quality variables. *Ground Water*, 25(2):176-184.

Morris, R.D., Audet, A.M., Angelillo, I.F., Chalmers, T.C. and Mosteller, F., 1992. Chlorination, chlorination by-products, and cancers: a meta-analysis. *American Journal of Public Health*, 82(7), 955-963.

Mucha, H.J., Bartel, H.G., and Dolata, J., 2008. Effects of Data Transformation on Cluster Analysis of Archaeometric Data. *Studies in Classification, Data Analysis, and Knowledge Organization. Data Analysis, Machine Learning and Applications. XI*, 681-688.

Mulaik, S. A., 1990. Blurring the Distinctions between Component Analysis and Common Factor-Analysis. *Multivariate Behavioral Research*, 25(1), 53-59.

Murdoch, P.S. and Shanley, J.B., 2006. Detection of water quality trends at high, median, and low flow in a Catskill Mountain stream, New York, through a new statistical method. *Water Resources Research*, Vol. 42, W08407, doi: 10.1029/2004WR003892.

Najafpour, Sh., Alkarkhi, A. F. M., Kadir, M. O. A. and Najafpour, Gh. D., 2008. Evaluation of Spatial and Temporal Variation in River Water Quality. *International Journal of Environmental Research*, 2(4), 349-358.

Nau, R.F., 2005. Three types of forecasts: estimation period, validation period, and “the future”. *Statistical Forecasting (MBA Course)*, <http://www.duke.edu/~rnau/three.htm>

Newman, M.C., Unger, M.A, 2002. *Fundamentals of Ecotoxicology*. 2nd Ed. Lewis Publishers, Florida.

Niemirycz, E., Nithhauser, J., Staniszevska, M., Nałęcz-Jawecki, G., Bolałek, J., 2007. The Microtox® biological test: Application in toxicity evaluation of surface waters and sediments in Poland. *International Journal of Oceanography and Hydrobiology*, Vol. XXXVI, No. 4, 151-163.

O'Donnell, M., Platt, C. and Aston, R., 2000. Effect of a boil water notice on behaviour in the management of a water contamination incident. *Communicable disease and public health / PHLS* 3(1), 56-59.

Ogorman, T. W. and Woolson, R. F., 1993. On the efficacy of the rank transformation in stepwise logistic and discriminant analysis. *Statistics in Medicine*, 12, 143-151.

Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., Huang, C. H., 2006. Assessment of seasonal variations in surface water quality. *Water Research*, 40, 3800-3810.

Page, E.S., 1954. Continuous Inspection Schemes. *Biometrika*, 41, 100-115.

Roberts, S.W., 1959. Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1, 239-250.

Papatheodorou, G., Demopoulou, G., Lamprakis, N., 2006. A long-term study of temporal hydrochemical data in a shallow lake using multivariate statistical techniques. *Ecological Modelling*, 193, 759-776.

Pavur, R., 1988. Type I error rates for multiple comparison procedures with dependent data. *The American Statistician*, 42(3):171-173.

Perona, E., Bonilla, I., Mateo, P., 1999. Spatial and temporal changes in water quality in a Spanish river. *The Science of the Total Environment*, Volume 241, Number 1, pp. 75-90.

Petalas, C., and Anagnostopoulos, K., 2006. Application of Stepwise Discriminant Analysis for the Identification of Salinity Sources of Groundwater. *Water Resources Management* 20: 681–700.

Peters, N. E., and Donohue, R., 2001. Nutrient transport to the Swan–Canning Estuary, Western Australia. *Hydrological Processes*, 15, 2555–2577.

Pett, M.A., 1997. *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions*. SAGE Publications, Inc, California, U.S.A.

Platikanov, S., Puig, X., Martin, J. and Tauler, R., 2007. Chemometric modeling and prediction of trihalomethane formation in Barcelona's water works plant. *Water Research* 41(15), 3394-3406.

Pollard, S.J.T., Strutt, J.E., Macgillivray, B.H., Hamilton, P.D. and Hrudey, S.E., 2004. Risk Analysis and Management in the Water Utility Sector: A Review of Drivers, Tools and Techniques. *Process Safety and Environmental Protection* 82(6), 453-462.

Porter, P.S. and Ward, R.C., 1991. Estimating central tendency from unsensored trace level measurements. *Water Resources Bulletin*, 27(4):687-700.

Porter, P.S., Ward, R.C. and Bell, H.F., 1988. The detection limit. *Environmental Science and Technology*, 22(8):856-861.

Prudham, S., 2004. Poisoning the well: neoliberalism and the contamination of municipal water in Walkerton, Ontario. *Geoforum* 35(3), 343-359.

Qadir, A., Malik, R. N., Husain, S. Z., 2008. Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. *Environmental Monitoring and Assessment*, 140(1-3), 43-59.

Qian, Y., Migliaccio, K. W., Wan, Y., Li, Y. C., and Chin, D., 2007. Seasonality of Selected Surface Water Constituents in the Indian River Lagoon, Florida. *Journal of Environmental Quality*, 36: 416-425.

Qian, Y., Migliaccio, K. W., Wan Y., and Li, Y., 2007. Trend Analysis of Nutrient Concentrations and Loads in Selected Canals of the Southern Indian River Lagoon, Florida. *Trend Analysis of Surface Water Nutrients. Water Air Soil Pollution*, 186:195–208.

Reckhow, K.H., Kepford, K., and Warren Hicks, W., 1993. *Methods for the Analysis of Lake Water Quality Trends. EPA 841-R-93-003.*

Reimann, C., Filzmoser, P., Garrett, R.G., 2002. Factor analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry* 17, 185-206.

Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: death of myth. Consequences for statistical treatment of geochemical and environmental data. *Environmental Geology*, Vol. 39(9), pp. 1001-1014.

Reynolds, B., Renshaw, M., Sparks, T.H., Crane, S., Hughes, S., Brittain, S.A. and Kennedy, V.H., 1997. Trends and seasonality in stream water chemistry in two moorland catchments of the Upper River Wye, Plynlimon. *Hydrology and Earth System Sciences*, 1(3), 571-581.

Ribo, J. M., 1997. Interlaboratory Comparison Studies of the Luminescent Bacteria Toxicity Bioassay. *Environmental Toxicology and Water Quality*, 12(4), 283-294.

Ribo J.M and K.L.E Kaiser, 1983. Effects of selected chemicals to photoluminescent bacteria and their correlations with acute and sublethal effects on other organisms' *Chemosphere*, Vol 12, pp 1421-1942.

Ribo, J. M., Canela, M., Griful, E., 2001. Repeatability and Reproducibility of the Luminescent Bacteria Bioassay. *Environmental Toxicology*, 16(2), 127-135.

Ribo, J. M., Rogers, F., 1990. Toxicity of mixtures of aquatic contaminants using the luminescent bacteria bioassay. *Toxicity Assessment: An International Journal*, 5(2), 135-152.

Ribo J.M, Zaruk, B.M., Hunter, H.and Kaiser, K.L.E., 1985. Microtox toxicity test results for water samples from the Detroit river. *Journal of Great Lakes Research*, Vol 11, pp. 297-304.

Rius, A., Ruisanchez, I., Gallao M. P. C., and Rius, F. X., 1998. Reliability of analytical systems: use of control charts, time series models and recurrent neural networks (RNN). *Chemometrics and Intelligent Laboratory Systems* 40, 1-18

Rizak, S. and Hrudey, S.E., 2008. Drinking-water safety: challenges for community-managed systems. *Journal of water and health* 6(1), 33-41.

Rogerson, P.A., 2001. *Statistical methods for Geography*. Sage Publications, London.

Rogora, M., Arese, C., Balestrini, R. And Marchetto, A., 2008. Climate control on sulphate and nitrate concentrations in alpine streams of Northern Italy along a nitrogen saturation gradient. *Hydrology and Earth System Sciences*, 12, 371-381.

Roig, B., Valat, C., Berho, C., Allan, I.J., Guigues, N., Mills, G.A., Ulitzur, N., Greenwood, R., 2007. The use of field studies to establish the performance of a range of tools for monitoring water quality. *Trends in Analytical Chemistry*, Vol. 26, No. 4, 274-282.

Romero, J., Kagalou, I., Imberger, J., Hela, D., Kotti, M., Bartzokas, A., Aldanis, T., Emvirides, N., Karkabounas, S., Papagiannis, J., Bithava, A., 2002. Seasonal water quality of shallow and eutrophic Lake Pamvotis, Greece: implications for restoration. *Hydrobiologia*, 474, 91-105.

Runger, G. C., Willemain, T. R., 1995. Model-Based and Model-Free Approaches for Control of Autocorrelated Processes. *Journal of Quality Technology* 27, 283-292

Runger, G.C., Willemain, T.R., and Prabhu, S., 1995. Average Run Lengths for CUSUM Control Charts Applied to Residuals. *Communications in Statistics – Theory and Methods* 24, 273-283.

Ryan, T.P., 1989. *Statistical Methods for Quality Improvement*, John Wiley and Sons, Inc., N. York

Sullivan, J.H. and Woodall, W.H., 1996a. A Control Chart for Preliminary Analysis of Individual Observations. *Journal of Quality Technology* 28, 265-278.

Scariano, S.M. and Davenport, J.M., 1987. The effects of violations of independence assumptions in the one-way ANOVA. *The American Statistician*, 41(2):123-129.

Schilling, E. G., and Nelson, P. R., 1976. The Effect of Nonnormality on the Control Limits of \bar{X} Charts. *Journal of Quality Technology* 8, 183-188

Schonemann, P. H., 1990. Facts, Fictions, and Common-Sense About Factors and Components. *Multivariate Behavioral Research*, 25(1), 47-51.

Schwab, A. J., 2007. *Strategy for Complete Discriminant Analysis*. University of Texas at Austin. <http://www.utexas.edu/courses/schwab>

Schwartz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, Vol. 5, No. 2, 461-464.

Sen P.K., 1968b. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* 39: 1379–1389.

Sharma, S., 1996. *Applied Multivariate Techniques*. John Wiley & Sons Inc., New York.

Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, 22(4), 464-475.

Shoukri, M.M. and Pause, C.A., 1999. *Statistical Methods for Health Sciences*. CRC Press LLC, 2nd ed.

Simeonov, V., Einax, J. W., Stanimirova, I., Kraft, J., 2002. Environmetric modeling and interpretation of river water monitoring data. *Analytical and Bioanalytical Chemistry*, 374(5), 898-905.

Simeonov, V., Sarbu, C., Massart, D. L., Tsakovski, S., 2001. Danube River Water Data Modelling by Multivariate Data Analysis. *Mikrochimica Acta*, 137, 243-248.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsas, D., Anthemidis, A., Sofoniou, M., Kouimtzis, T., 2003. Assessment of the surface water quality in Northern Greece. *Water Research* 37, 4119-4124.

Simeonov, V., Wolska, L., Kuczyńska, A., Gurwin, J., Tsakovski, S. and Namieśnik, J., 2007. 'Chemometric Estimation of Natural Water Samples Using Toxicity Tests and Physicochemical Parameters', *Critical Reviews in Analytical Chemistry*, 37: 2, 81-90.

Singh, K. P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gompti River (India) – a case study. *Water Research*, 38, 3980-3992.

Singh, K. P., Malik, A., Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gompti river (India) using multivariate statistical techniques – a case study. *Analytica Chimica Acta*, 538, 355-374.

Snedecor, G.W. and Cochran, W.G., 1989. *Statistical Methods*. 8th Ed. Iowa State University Press, Ames, IA.

Snook, S. C., & Gorsuch, R. L., 1989. Component Analysis Versus Common Factor-Analysis – a Monte- Carlo Study. *Psychological Bulletin*, 106(1), 148-154.

Spanos, T., Simeonov, V., Stratis, J., and Xatzixristou, X., 2003. Assessment of Water Quality for Human Consumption. *Microchimica Acta*, 141(1-2), 35-40.

Sprague L. A. and Lorenz D. L., 2009. Regional Nutrient Trends in Streams and Rivers of the United States, 1993–2003. *Environ. Science and Technology*, 2009, 43 (10), 3430–3435.

Spruill, T. B., Showers, W. J., Howe, S.S., 2002. Ground Water Quality. Application of Classification-Tree Methods to Identify Nitrate Sources in Ground Water. *J. Environ. Qual.* 31:1538–1549

Stanimirova, I., PoBowniak, M., Skorek, R., Kita, A., John, E., Buhl, F. and Walczak, B., 2007. Chemometric analysis of the water purification process data. *Talanta* 74(1), 153-162.

Steiger, J. H., 1990. Some Additional Thoughts on Components, Factors, and Factor-Indeterminacy. *Multivariate Behavioral Research*, 25(1), 41-45.

Steinberg, D., and Cola, P., 1997. *CART-Classification and Regression Trees*, Salford Systems, San Diego, CA.

Strategic Diagnostics Microtox 1010, 2003. Method for Measuring the Acute Toxicity of Wastewater and Receiving Water with *Vibrio fischeri* (NRRL B-11177) Microtox[®] Test System, Strategic Diagnostics, Inc.

Suk, H., Lee, K., 1999. Characterization of a ground water hydrochemical system through multivariate analysis: clustering into ground water zones. *Ground Water* 37(3), 358-366.

Sullivan, J.H. and Woodall, W.H., 1996a. A Control Chart for Preliminary Analysis of Individual Observations. *Journal of Quality Technology* 28, 265-278.

Sun, F., Chen, J., Tong, Q. and Zeng, S., 2007. Integrated risk assessment and screening analysis of drinking water safety of a conventional water supply system. *Water science and technology: a journal of the International Association on Water Pollution Research* 56(6), 47-56.

Taylor, J.K., 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Chelsea, MI.

Taylor, C.H. and Loftis, J.C., 1989. Testing for trend in lake and ground water quality time series. *Water Resources Bulletin*, 25(4): 715-726.

Tseng, S., and Adams, B. M., 1994. Monitoring Autocorrelated Processes with an Exponentially Weighted Moving Average Forecast. *Journal of Statistical Computation and Simulation* 50

Twenter, J.C. and Loftis, J.C., 2003. Design of a water quality information system for source water assessment: A Denver water case study, Denver, Colorado. Completion Report No.197, Colorado Water Resources Research Institute, Colorado State University, Fort Collins, Colorado.

USEPA (U.S. Environmental Protection Agency), 1999. Nonpoint Source Pollution: The Nation's Largest Water Quality Problem, Pointer No. 1. Available at <http://www.epa.gov/OWOW/NPS/facts/point1.html>.

US Environmental Protection Agency, 2005. Technologies and Techniques for Early Warning Systems to Monitor and to Evaluate Drinking Water Monitoring Quality: a State of the Art Review. EPA/600/R-05/156.

Van Belle, G., Hughes, J.P., 1984. Nonparametric tests for trend in water quality. *Water Resources Research* 20 (1), 127-136.

VanBrackle, L.N. and Reynolds M.R. JR., 1997. EWMA and CUSUM Control Charts in the Presence of Autocorrelation. *Communications in Statistics-Simulation and Computation* 26, 979-1008.

Van Gelder P.H.A.J.M., Mai, C.V., Wang, W., Shams, G., Rajabalinejad, M., Burgmeijer, M., 2008. *Journal of Hydraulic Research* Vol. 46, Extra Issue 2 (2008), pp. 191–210. International Association of Hydraulic Engineering and Research

Vasilopoulos, A.V. and Stamboulis, A.P., 1978. Modification of Control Chart limits in the presence of Data Correlation. *Journal of Quality Technology* 10, 20-30.

Vega, M., Pardo, R., Barrado, E. And Debán, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32(12), 3581-3592.

Velicer, W. F., & Jackson, D. N., 1990. Component Analysis Versus Common Factor-Analysis – Some Further Observations. *Multivariate Behavioral Research*, 25(1), 97-114.

Vogt, C. and Regli, S., 1981. Controlling Trihalomethanes while Attaining Disinfection. *Journal AWWA*, 73(1), 33-40.

Von Storch, H., 1995. Misuses of statistical analysis in climate research. In: *Analysis of Climate Variability: Applications of Statistical Techniques* (ed. by H. von Storch and A. Navarra), 11-26. Springer-Verlag, Berlin, Germany.

Voutsas, D., Manoli, E., Samara, C., Sofoniou, M. and Stratis, I., 2001. A study of surface water quality in Macedonia, Greece: Speciation of nitrogen and phosphorous. *Water, Air and Soil Pollution*, 129, 13-32.

Wadsworth, H. M., Stephens, K. S., and Godfrey, A. B., 1986. *Modern Methods for Quality Control and Improvement*, John Wiley and Sons, Inc., N. York

Wang, W., 2006. Stochasticity, nonlinearity and forecasting of streamflow processes PhD Thesis, Delft Uni. of Technology. ISBN: 978-1-58603-621-8.

Wang, W., Van Gelder, P.H.A.J.M., and Vrijling, J.K., 2005. Trend and Stationarity Analysis for Streamflow Processes of Rivers in Western Europe in the 20th Century. *Proceedings of the IWA International Conference on Water Economics, Statistics, and Finance*, Rethymno, Greece, 8-10 July 2005, p. 451-461.

Ward, R.C., 1999. Management and monitoring of water quality: class notes. Colorado State University, Fort Collins, CO.

Ward, R. C., Loftis, J. C., McBride, G. B., 2003. Design of Water Quality Monitoring Systems. Wiley & Sons, Inc., Hoboken, New Jersey.

Wardell, D.G., Moskowitz, H., and Plante, R.D., 1992. Control Charts in the Presence of Data Correlation. *Management Science* 38, 1084-1105.

Wardell, D.G., Moskowitz, H., and Plante, R.D., 1994. Run-Length Distributions of Special-Cause Control Charts for Correlated Processes. *Technometrics* 36, 3-27.

Wayland, K., Long, D., Hyndman, D., Pijanowski, B., Woodhams, S., Haack, K., 2003. Identifying relationships between baseflow geochemistry and land use with synoptic sampling and R-Mode factor analysis. *Journal of Environmental Quality*, 32, 180-190.

Wheater, C. P. and Cook, P. A., 2000. Using Statistics to Understand the Environment. Routledge, Taylor & Francis Group, London and New York.

Widaman, K. F., 1990. Bias in Pattern Loadings Represented by Common Factor-Analysis and Component Analysis. *Multivariate Behavioral Research*, 25(1), 89-95.

Widaman, K. F., 1993. Common Factor-Analysis Versus Principal Component Analysis – Differential Bias in Representing Model Parameters. *Multivariate Behavioral Research*, 28(3), 263-311.

Winston, G. and Leventhal, A., 2008. Unintentional drinking-water contamination events of unknown origin: surrogate for terrorism preparedness. *J Water Health* 6(Suppl 1), 11-19.

Woodall, W.H., 1985. The Statistical Design of Quality Control Charts. *The Statistician* 34, 155-160.

Wunderlin, D.A., Diaz, M.P., Amé, M.V., Pesce, S.F., Hued, A.C., Bistoni, M.A., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía river basin (Córdoba, Argentina). *Water Research* 35(12), 2881-2894.

Yanga, Y.J., Goodricha, J.A., Clarkb, R.M. and Lic, S.Y., 2008. Modeling and testing of reactive contaminant transport in drinking water pipes: Chlorine response and implications for online contaminant detection. *Water Research* 42(6-7), 1397-1412.

Yohannes, Y., and Webb, P., (1999). Classification and Regression Trees, CART™: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity. International Food Policy Research Institute, Washington D.C., U.S.A.

Yu, Y.-S., Zou, S., Whittemore, D., 1993. Non-parametric trend analysis of water quality data of rivers in Kansas. *Journal of Hydrology*, 150, 61-80.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, 16(9), 1807-1829.

Yue, S., Pilon, P., 2003. Canadian streamflow trend detection: impacts of serial and cross-correlation. *Hydrological Sciences-Journal*, 48(1), 51-63.

Yürekli, K., Kurunç, 2005. Testing the Residuals of an ARIMA Model on the Çekerek Stream Watershed in Turkey. *Turkish J. Eng. Env. Sci.*, 29, 61-74.

Zar, J. H. 1999. *Biostatistical analysis*, 4th Edition. Prentice Hall, Inc. Upper Saddle River, New Jersey, USA.

Zhang, N.F., 1998. A Statistical Control Chart for Stationary Process Data. *Technometrics* 40, 24-38.

Zhang, X., Vincent, L.A., Hogg, W.D. and Niitsoo, A., 2000. Temperature and Precipitation Trends in Canada During the 20th Century. *Atmosphere - Ocean* 38(3), 395-429.

Zhang, X., Harvey, K. D., Hogg, W. D., Yuzyk, T. R., 2001a. Trends in Canadian streamflow. *Water Resources Research*, 37(4): 987-998.

Zhang, X., Hogg, W. D. and Mekis, E., 2001b. Spatial and Temporal Characteristics of Heavy Precipitation Events over Canada. *Journal of Climate* 14, 1923-1936.

Zhou, S.W., Xu, F.D., Li, S.M., Song, R.X., Qi, S., Zhang, Y. and Bao, Y.P., 1997. Major origin of mutagenicity of chlorinated drinking water in China: humic acid or pollutants. *The Science of the Total Environment* 196(3), 191-196.

Αντωνόπουλος, Β., 2001. Ποιότητα Νερού και Ρύπανση Υδατικών Πόρων. Συμπόσιο: “Αιγαίο – Νερό - Βιώσιμη Ανάπτυξη” (υπό την Αιγίδα του Υπουργείου Αιγαίου). Πάρος, 6 & 7 Ιουλίου 2001.

Ανωμερίτης, Β., Βάζιος, Γ., Δημακόπουλος, Ι., Διαμαντόπουλος, Γ., Διαμαντόπουλος, Μ., Ζαννάκη, Κ., Κατσίκη, Τ., Καφές, Δ., Κούκιος, Μ., Κιτσίδης, Θ., Μερτζάνης, Α., Μπάζος, Α., Μπέλλου, Α., Παπαϊωάννου, Δ., Παναγιωτίδης, Π., Χαντζημπίρος, Κ., 1987. Προστασία περιβάλλοντος και οικονομική ανάπτυξη στην περιοχή της Υλίκης. Ευρωπαϊκή Οικονομική Κοινότητα – Δήμος Θηβαίων.

Αποστολάκης, Ι., Καστανιά, Α., Πιερράκου, Χ., 2003. Στατιστική επεξεργασία δεδομένων στην υγεία. Εκδόσεις Παπαζήση, Αθήνα.

Βυλλιώτη, Α. Ε., 2009. Ανάπτυξη προχωρημένων οξειδωτικών διεργασιών για την απομάκρυνση οργανικών τοξικών ουσιών σε υδατικά συστήματα. Διδακτορική διατριβή. Τμήμα Χημείας, Σχολή Θετικών Επιστημών, Πανεπιστήμιο Ιωαννίνων

Γενικό Χημείο Κράτους - Κύπρος, 2010. <http://www.moh.gov.cy/moh/sgl.nsf>

Γκολφινόπουλος, Σ., 2005. Σημειώσεις του μαθήματος «Επιστήμη Περιβάλλοντος. Πανεπιστήμιο Αιγαίου, Τμήμα Μηχανικών Οικονομίας και Διοίκησης.

Δημέλη, Σ., 1996. Σύγχρονες μέθοδοι ανάλυσης χρονολογικών σειρών. Εκδόσεις Ευγ. Μπένου, Αθήνα.

ΕΥΔΑΠ, 2010. Ύδρευση. Πηγές υδροληψίας. Επεξεργασία νερού.
http://www.eydap.gr/index.asp?a_id=43

ΕΥΔΑΠ, 2005. Πεπραγμένα Υπηρεσίας Ελέγχου Ποιότητας Ύδατος, Αθήνα.

ΕΥΔΑΠ, 2006. Πεπραγμένα Υπηρεσίας Ελέγχου Ποιότητας Ύδατος, Αθήνα.

ΕΥΔΑΠ, 2007. Πεπραγμένα Υπηρεσίας Ελέγχου Ποιότητας Ύδατος, Αθήνα.

ΕΥΔΑΠ, 2008. Πεπραγμένα Υπηρεσίας Ελέγχου Ποιότητας Ύδατος, Αθήνα.

ΕΥΔΑΠ, 2009. Πεπραγμένα Υπηρεσίας Ελέγχου Ποιότητας Ύδατος, Αθήνα.

Ευστρατιάδης, Α., 2008. Προσομοίωση και βελτιστοποίηση διαχείρισης υδροδοτικού συστήματος Αθήνας, Τομέας Υδατικών Πόρων και Περιβάλλοντος-ΕΜΠ, Αθήνα.

Καραδήμα, Κ., 2009. Εκτίμηση της τοξικότητας διαφόρων σταδίων επεξεργασίας αποβλήτων τυροκομικών μονάδων με χρήση βιοδεικτών. Διδακτορική διατριβή. Τομέας Βιολογίας Ζώων, Τμήμα Βιολογίας, Πανεπιστήμιο Πατρών.

Καρλής, Δ., 2005. Πολυμεταβλητή Στατιστική Ανάλυση. Εκδόσεις Σταμούλη Α.Ε., Αθήνα.

Κούγκολος Α., 2005. Εισαγωγή στην Περιβαλλοντική Μηχανική. Εκδόσεις Τζιόλα, Θεσσαλονίκη.

Κουτσογιάννης, Δ., Ευστρατιάδης, Α., Καραβοκυρός, Γ., Κουκουβίνος, Α., Μαμάσης, Ν., Ναλμπάντης, Ι., Γκριντζιά Δ., Δαμιανόγλου, Ν., Ξανθάκης Α., Πολιτάκη, Σ., Τσουκαλά, Β., 2000. Σχέδιο διαχείρισης του υδροδοτικού συστήματος της Αθήνας - Έτος 2000-2001, Εκσυγχρονισμός της εποπτείας και διαχείρισης του συστήματος των υδατικών πόρων ύδρευσης της Αθήνας, Τεύχος 5, ΕΜΠ, ΕΥΔΑΠ, Αθήνα.

Λέκκας, Θ., 2005. Περιβαλλοντική Μηχανική Ι. Διαχείριση Υδατικών Πόρων. Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 2002. Πρώτη Έκθεση Προόδου, Ποιότητα Πόσιμου Ύδατος ΕΥΔΑΠ – Εναλλακτικές Μέθοδοι Απολύμανσης. Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 1998. Αναγνώριση της κατάστασης ρύπανσης των επιφανειακών νερών της Ελλάδας από ουσίες του καταλόγου ΙΙ, υποψήφιος για τον κατάλογο Ι της οδηγίας 76/464/EEC και οργάνωση – λειτουργία δικτύου παρακολούθησης της ποιότητας των επιφανειακών νερών, σύμφωνα με τις ουσίες που θα προκύψουν από την αναγνώριση (Φάση Α). Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 1999. Ενδιάμεση Έκθεση, Αναγνώριση της κατάστασης ρύπανσης των επιφανειακών νερών της Ελλάδας από ουσίες του καταλόγου ΙΙ, υποψήφιος για τον κατάλογο Ι της οδηγίας 76/464/EEC και οργάνωση – λειτουργία δικτύου παρακολούθησης της ποιότητας των επιφανειακών νερών, σύμφωνα με τις ουσίες που θα προκύψουν από την αναγνώριση. Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 1999. Τελική Έκθεση, Λειτουργία δικτύου παρακολούθησης των επιφανειακών νερών αναφορικά με τις ουσίες του καταλόγου Ι της οδηγίας 76/464/EEC. Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 1999. Τελική Έκθεση, Λειτουργία δικτύου παρακολούθησης των επιφανειακών νερών αναφορικά με τις ουσίες του καταλόγου II της οδηγίας 76/464/EEC. Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 2000. Εθνικό πρόγραμμα μείωσης της ρύπανσης για τις ουσίες του καταλόγου II, υποψήφιος για τον κατάλογο I, C176, 14/07/1982/EEC, σύμφωνα με το άρθρο 7 της Οδηγίας του Συμβουλίου 76/464/EEC. Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λέκκας, Θ., 2000. Εθνικό πρόγραμμα μείωσης της ρύπανσης για τις ουσίες του καταλόγου II, (Μέταλλα), C176, 14/07/1982/EEC, σύμφωνα με το άρθρο 7 της Οδηγίας του Συμβουλίου 76/464/EEC. Εργαστήριο Ποιότητας Υδάτων και Αέρα, Πανεπιστήμιο Αιγαίου, Μυτιλήνη.

Λύτρας, Ε., 2009. Η ποιότητα του ανεπεξέργαστου και πόσιμου νερού από το 2000 μέχρι σήμερα. Συνέδριο ΕΥΔΑΠ «Νερό: Η Επόμενη Μέρα», Ζάππειο Μέγαρο, 20 Μαρτίου 2009.

Μαμάης, Δ., 2006. Ρύπανση Υδατικών Οικοσυστημάτων. Αθήνα. Ε.Μ.Π. Σημειώσεις του Μαθήματος "Υδατικό Περιβάλλον και Ανάπτυξη" του Δ.Π.Μ.Σ. "Περιβάλλον και Ανάπτυξη".

Μπάζος, Α., Μπέλλου, Α., Παπαϊωάννου, Δ., Παναγιωτίδης, Π., Χαντζημπίρος, Κ., 1987. Προστασία περιβάλλοντος και οικονομική ανάπτυξη στην περιοχή της Υλίκης. Ευρωπαϊκή Οικονομική Κοινότητα – Δήμος Θηβαίων.

Νταρακάς, Ε. Ν., 2009. Βασικές παράμετροι ελέγχου ποιότητας νερού και λυμάτων. Σημειώσεις για τη θεωρία και τα εργαστήρια των μαθημάτων Υδρεύσεις – Αποχετεύσεις και Τεχνική Περιβάλλοντος. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Πολιτικών Μηχανικών, Τομέας Υδραυλικής & Τεχνικής Περιβάλλοντος.

Ντζούφρας, Ι., 2001. Στοιχεία Πολυμεταβλητής Ανάλυσης Δεδομένων. Σημειώσεις για το μάθημα Ανάλυση Δεδομένων Ι. Πανεπιστήμιο Αιγαίου, Τμήμα Διοίκησης Επιχειρήσεων.

Ξανθόπουλος, Θ. και Κουτσογιάννης, Δ., 1997. Υδατικοί Πόροι, Τεχνολογία & Πληροφορική, Εκπαιδευτική Ελληνική Εγκυκλοπαίδεια: Εκδοτική Αθηνών.

Παρασκευόπουλος, Α., 1997. Διαχείριση απορριμμάτων στη λεκάνη του Μόρνου ανάντη του ταμιευτήρα. ΕΥΔΑΠ Διεύθυνση έργων αποχέτευσης – Υπηρεσία μελέτης.

Σιάρδος, Γ., 2000. Μέθοδοι Πολυμεταβλητής Στατιστικής Ανάλυσης: Με την επίλυση ασκήσεων μέσω του στατιστικού προγράμματος SPSS. Μέρος δεύτερο: Διερεύνηση εξάρτησης μεταξύ μεταβλητών. Αυτοέκδοση, Θεσσαλονίκη, ISBN 960-91293-0-7.

Τζουμέρκας, Φ., 2002. Η διασφάλιση της ποιότητας του πόσιμου νερού κατά τη διανομή του με δίκτυο σωληνώσεων. Ημερίδα: Η Ποιότητα του Νερού Ανθρώπινης Κατανάλωσης, η Οδηγία 98/83 ΕΟΚ και η Ελληνική Πραγματικότητα. Γενικό Χημείο του Κράτους, Ιούλιος 2002.

Τσώνης, Σ.Π., 2003. Καθαρισμός νερού. 1^η Έκδοση, Εκδόσεις Παπασωτηρίου, Αθήνα.

ΦΕΚ 720/Β/13-12-83. Προστασία των νερών που χρησιμοποιούνται για την ύδρευση της περιοχής Πρωτεύουσας από ρυπάνσεις και μολύνσεις (ΚΥΑ Α5/2280/83).

ΦΕΚ 892/Β/11.7.2001. «Ποιότητα του νερού ανθρώπινης κατανάλωσης», σε συμμόρφωση προς την οδηγία 98/83/ΕΚ του Συμβουλίου της Ευρωπαϊκής Ένωσης της 3^{ης} Νοεμβρίου 1998. (Κοινή Υπουργική Απόφαση Υ2/2600/2001).

ΦΕΚ 630/Β/26.4.2007. Τροποποίηση της Υγειονομικής Διάταξης κοινής υπουργικής απόφασης Υ2/2600/2001 «Ποιότητα νερού ανθρώπινης κατανάλωσης», σε συμμόρφωση προς την οδηγία 98/83/ΕΚ του Συμβουλίου της Ευρωπαϊκής Ένωσης της 3^{ης} Νοεμβρίου 1988.

ΦΕΚ 54/Α/8.3.2007. Προεδρικό Διάταγμα υπ' αριθμ. 51. Καθορισμός μέτρων και διαδικασιών για την ολοκληρωμένη προστασία και διαχείριση των υδάτων σε συμμόρφωση με τις διατάξεις της Οδηγίας 2000/60/ΕΚ «για της θέσπιση πλαισίου κοινοτικής δράσης στον τομέα της πολιτικής των υδάτων» του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 23^{ης} Οκτωβρίου 2000.

ΦΕΚ 438/Β/3.7.1988. Υπουργικές Αποφάσεις και Εγκρίσεις αριθ. οικ. 46399/1352. Απαιτούμενη ποιότητα των επιφανειακών νερών που προορίζονται για «πόσιμα», «κολύμβηση», «διαβίωση ψαριών σε γλυκά νερά» και «καλλιέργεια και αλιεία οστρακοειδών», μέθοδοι μέτρησης, συχνότητα δειγματοληψίας και ανάλυση των επιφανειακών νερών που προορίζονται για πόσιμα, σε συμμόρφωση με τις οδηγίες του Συμβουλίου των Ευρωπαϊκών Κοινοτήτων 75/440/ΕΟΚ, 78/659/ΕΟΚ, 79/923/ΕΟΚ και 79/869/ΕΟΚ.

**ΔΗΜΟΣΙΕΥΣΕΙΣ ΤΗΣ ΠΑΡΟΥΣΑΣ ΔΙΑΤΡΙΒΗΣ ΣΕ ΔΙΕΘΝΗ
ΕΠΙΣΤΗΜΟΝΙΚΑ ΠΕΡΙΟΔΙΚΑ ΚΑΙ ΠΡΑΚΤΙΚΑ ΔΙΕΘΝΩΝ
ΕΠΙΣΤΗΜΟΝΙΚΩΝ ΣΥΝΕΔΡΙΩΝ (ΜΕ ΚΡΙΤΕΣ)**

A. ΕΠΙΣΤΗΜΟΝΙΚΑ ΠΕΡΙΟΔΙΚΑ

Smeti, E., Thanasoulis, N., Kousouris, L., Tzoumerkas, P., 2006. An approach for the application of statistical process control techniques for quality improvement of treated water. *Desalination*, 213, 273-281.

Smeti, E., Koronakis, D., Golfinopoulos, S., 2007. Control charts for the toxicity of finished water – Modeling the structure of toxicity. *Water Research*, 41, 2679-2689.

Smeti, E., Thanasoulis, N., Lytras, E., Tzoumerkas, P., Golfinopoulos, S., 2009. Treated water quality assurance and description of distribution networks by multivariate chemometrics. *Water Research*, 43, 4676-4684.

B. ΔΙΕΘΝΗ ΣΥΝΕΔΡΙΑ ΜΕ ΚΡΙΤΕΣ

Koronakis, D., Melabianaki, Z., Smeti, E., 2005. «Ecotoxicological Monitoring of Raw and Finished Water by Bioluminescent Bacteria *Vibrio fischeri*». *Proceedings of 12th International Symposium on Toxicity Assessment*.

Smeti, E., Koronakis, D., Kousouris, L., 2005. «Statistical Process Control Techniques on Water Toxicity Data». *Proceedings of 7th Hellenic European Research on Computer Mathematics & its Applications (HERCMA) Conference*.

Smeti, E. M., Thanasoulis, N. C., Kousouris, L. P., Tzoumerkas, P. C., 2005. «Chemical Profile of Potable Water Produced at the Athens Water Supply and Sewerage Company Water Treatment Plants – Statistical Process Control for Quality Improvement». *Proceedings of 1st International Conference on new water culture of South-East European Countries (Aqua 2005)*.

Smeti, E. M., Kousouris, L. P., Tzoumerkas, P. C., Golfinopoulos, S. K., 2006. «Statistical process control techniques on autocorrelated turbidity data from finished water tank». *Proceedings of 2nd International Conference: Water Science and Technology – Integrated Management of Water Resources (Aqua 2006)*.

Smeti, E. M., Kousouris, L. P., Tzoumerkas, P. C., Golfinopoulos, S. K., 2007. «Trend analysis and variability of microbiological parameters of the Mornos Reservoir (Greece)». *Proceedings of the 10th International Conference on Environmental Science and Technology, Kos island, Greece, 5-7 September 2007*.

Smeti, E. M., Thanasoulis, N. C., Lytras, E. S., Tzoumerkas, P. C., Golfinopoulos, S. K., 2008. «Multivariate chemometrics for the discrimination of treated water samples and network tank classification». *Proceedings of the 3rd International Conference: Water Science and Technology - Integrated Management of Water Resources (AQUA 2008), Athens, 16-19 October 2008*.

Smeti, E. M., Lytras, E. S., Tzoumerkas, P. C., Golfinopoulos, S. K., 2009. «Multivariate statistical analysis of surface water intended for human consumption». *Proceedings of the 11th International Conference on Environmental Science and Technology – 2009 (CEST 2009), Chania, Greece, 3-5 September 2009*.

ΠΑΡΑΡΤΗΜΑ

Πίνακας Π1. Χημικές παράμετροι ποιότητας του νερού

Πίνακας Π2. Μέθοδοι προσδιορισμού των χημικών παραμέτρων

Πίνακας Π3. Όρια νομοθεσίας ανά χημική παράμετρο

Πίνακας Π1. Χημικές παράμετροι ποιότητας του νερού

Παράμετρος	Περιγραφή	Λόγοι παρακολούθησης
Ηλεκτρική αγωγιμότητα (μS/cm)	Γενική	Παρέχει τη συνάφεια μεταξύ των ολικών διαλυτών στερεών στο νερό και των κύριων ιόντων.
Ολικά διαλυτά στερεά (mg/l NaCl)	Γενική	Υψηλά επίπεδα υποδεικνύουν «σκληρό» νερό το οποίο μπορεί να προκαλέσει προβλήματα εναπόθεσης αλάτων στους σωλήνες προκαλώντας αύξηση του κόστους συντήρησης του συστήματος.
Θολότητα (NTU)	Γενική	Δείκτης βιολογικής δραστηριότητας στην υδάτινη στήλη.
pH	Γενική	Σημαντική παράμετρος στην αποτίμηση της ποιότητας του νερού καθώς πολλές βιολογικές και χημικές διεργασίες, που εμπλέκονται στους υδάτινους όγκους, επηρεάζονται από το pH.
Αλκαλικότητα, Ολική (mg/l CaCO ₃)	Γενική	Δείκτης των ειδών ανθρακικών συγκεντρώσεων. Επίδραση στην ουδετεροποίηση του pH.
Σκληρότητα, Ολική (mg/l CaCO ₃)	Γενική	Επιπτώσεις κατά την χρήση. Το σκληρό νερό προκαλεί εναπόθεση αλάτων στους θερμοσίφωνες και τους λέβητες. Το μαλακό νερό θεωρείται διαβρωτικό.
Χλωριούχα (mg/l Cl ⁻)	Κύριο ιόν	Δείκτης (σε υψηλές συγκεντρώσεις) εκρών βιομηχανικών αποβλήτων και λυμάτων.

Παράμετρος	Περιγραφή	Λόγοι παρακολούθησης
Θειικά (mg/l SO ₄ ²⁻)	Κύριο ιόν	Επιπτώσεις στην χρήση (γεύση και οσμή). Σε υψηλά επίπεδα υποδεικνύει εκροές βιομηχανικών αποβλήτων και διοχετεύσεων ορυχείων.
Νάτριο (mg/l Na ⁺)	Κύριο ιόν	Υψηλά επίπεδα στο επιφανειακό νερό μπορεί να οφείλονται σε λύματα και βιομηχανικά απόβλητα καθώς και σε άλατα που χρησιμοποιούνται στους δρόμους. Μπορεί επίσης να επηρεάσει την αποτελεσματικότητα της άρδευσης.
Ασβέστιο (mg/l Ca ²⁺)	Κύριο ιόν	Δείκτης σκληρότητας (επηρεάζει την σκληρότητα του νερού).
Μαγνήσιο (mg/l Mg ²⁺)	Κύριο ιόν	Δείκτης σκληρότητας (επηρεάζει την σκληρότητα του νερού).
Χημικά απαιτούμενο οξυγόνο (mg/l O ₂)	Γενική	Αποτελεί έμμεση μέτρηση της ποσότητας των οργανικών ενώσεων στο νερό.
Διαλελυμένο οξυγόνο (mg/l O)	Γενική	Ουσιαστικό για την υδρόβια ζωή. Δείκτης της οργανικής ρύπανσης, της καταστροφής των οργανικών ουσιών, και του επιπέδου αυτο-καθαρισμού στο φυσικό ύδωρ (το οξυγόνο περιλαμβάνεται μέσα, ή επηρεάζει, όλες τις χημικές/βιολογικές διαδικασίες μέσα στους υδάτινους όγκους).
Αμμώνιο (mg/l NH ₄ ⁺)	Θρεπτικά	Δείκτης οργανικής ρύπανσης από λύματα ή βιομηχανικές εκροές, γεωργικά απόβλητα και λιπάσματα.

Παράμετρος	Περιγραφή	Λόγοι παρακολούθησης
Νιτρικά (mg/l NO ₃ ⁻)	Θρεπτικά	Σε αυξημένα επίπεδα, πιθανός κίνδυνος για την υγεία (ειδικά για τα βρέφη)
Φωσφορικά (mg/l PO ₄ ³⁻)	Θρεπτικά	Υψηλές συγκεντρώσεις υποδεικνύουν ρύπανση. Δείκτης του επιπέδου των θρεπτικών (αύξηση των φυκών).
Αργίλιο (μg/l Al)	Γενική	
Υπολειματικό χλώριο (mg/l Cl ₂)	Γενική	

Πίνακας Π2. Μέθοδοι προσδιορισμού των χημικών παραμέτρων

ΠΑΡΑΜΕΤΡΟΣ	ΜΟΝΑΔΕΣ ΜΕΤΡΗΣΗΣ	ΟΡΙΟ ΑΝΙΧΝΕΥΣΗΣ (MDL)	ΜΕΘΟΔΟΣ ΔΟΚΙΜΗΣ
ΑΓΩΓΙΜΟΤΗΤΑ, 25 °C	μS/cm		SMWW 21st ed., 2005, 2510, Ηλεκτρομετρική
ΟΛΙΚΑ ΔΙΑΛΥΤΑ ΣΤΕΡΕΑ	mg/l NaCl		Ηλεκτρομετρική μέθοδος
ΘΟΛΟΤΗΤΑ	NTU	0,05	SMWW 21st ed., 2005, 2130 B, Νεφελομετρική
pH (Θερμοκρασία μέτρησης)	μονάδες pH (°C)		SMWW 21st ed., 2005, 4500-H, Ηλεκτρομετρική
ΑΛΚΑΛΙΚΟΤΗΤΑ (Ηλιανθίνης)	mg/l CaCO ₃		SMWW 21st ed., 2005, 2320 B, Ογκομετρική
ΣΚΛΗΡΟΤΗΤΑ (Ογκομετρ.)	mg/l CaCO ₃		SMWW 21st ed., 2005, 2340 C, Ογκομετρική
ΑΜΜΩΝΙΟ	mg/l NH ₄ ⁺	0,005	SMWW 18 th ed., 1992, 4500-NH ₃ C, Φωτομετρική
ΧΛΩΡΙΟΥΧΑ	mg/l Cl ⁻	0,027	Ιοντική Χρωματογραφία
ΝΙΤΡΙΚΑ	mg/l NO ₃ ⁻	0,006	Ιοντική Χρωματογραφία
ΦΩΣΦΟΡΙΚΑ	mg/l PO ₄ ³⁻	0,010	Ιοντική Χρωματογραφία
ΘΕΠΚΑ	mg/l SO ₄ ²⁻	0,009	Ιοντική Χρωματογραφία
ΝΑΤΡΙΟ	mg/l Na ⁺	0,075	Ιοντική Χρωματογραφία
ΑΣΒΕΣΤΙΟ	mg/l Ca ²⁺		SMWW 21st ed., 2005, 3500Ca B, Ογκομετρική
ΜΑΓΝΗΣΙΟ	mg/l Mg ²⁺		SMWW 21st ed., 2005, 3500Mg B, Ογκομετρική
ΧΗΜΙΚΑ ΑΠΑΙΤΟΥΜΕΝΟ ΟΞΥΓΟΝΟ	mg/l O ₂		Εσωτερική μέθοδος με KMnO ₄
ΔΙΑΛΕΔΥΜΕΝΟ ΟΞΥΓΟΝΟ	mg/l O		SMWW 21 st ed., 2005, 4500-O D, Ογκομετρική
ΑΡΓΙΛΙΟ	μg/l Al	12	SMWW 21st ed., 2005, 3500-Al B, Φωτομετρική
ΥΠΟΛΕΙΜΜΑΤΙΚΟ ΧΛΩΡΙΟ	mg/l Cl ₂		Φωτομετρική

Πίνακας Π3. Όρια νομοθεσίας ανά χημική παράμετρο

ΠΑΡΑΜΕΤΡΟΣ	ΜΟΝΑΔΕΣ ΜΕΤΡΗΣΗΣ	ΟΡΙΑ ΟΙΚ 46399/1352 (Α2) /1986 *	ΟΡΙΑ ΚΥΑ Υ2/2600/2001
ΑΓΩΓΙΜΟΤΗΤΑ, 25 °C	μS/cm	1000	2500
ΟΛΙΚΑ ΔΙΑΛΥΤΑ ΣΤΕΡΕΑ	mg/l NaCl		
ΘΟΛΟΤΗΤΑ	NTU		1
pH (Θερμοκρασία μέτρησης)	μονάδες pH (°C)	5,5-9,0	6,5-9,5
ΑΛΚΑΛΙΚΟΤΗΤΑ (Ηλιανθίνης)	mg/l CaCO ₃		
ΣΚΛΗΡΟΤΗΤΑ (Ογκομετρ.)	mg/l CaCO ₃		
ΑΜΜΩΝΙΟ	mg/l NH ₄ ⁺	1 / 1,5	0,5
ΧΛΩΡΙΟΥΧΑ	mg/l Cl ⁻	200	250
ΝΙΤΡΙΚΑ	mg/l NO ₃ ⁻	50	50
ΦΩΣΦΟΡΙΚΑ	mg/l PO ₄ ³⁻	0,7 mg/l P ₂ O ₅	
ΘΕΪΚΑ	mg/l SO ₄ ²⁻	150 / 200	250
ΝΑΤΡΙΟ	mg/l Na ⁺		200
ΑΣΒΕΣΤΙΟ	mg/l Ca ²⁺		
ΜΑΓΝΗΣΙΟ	mg/l Mg ²⁺		
ΧΗΜΙΚΑ ΑΠΑΙΤΟΥΜΕΝΟ ΟΞΥΓΟΝΟ	mg/l O ₂		5
ΔΙΑΛΕΛΥΜΕΝΟ ΟΞΥΓΟΝΟ	mg/l O		
ΑΡΓΙΑΙΟ	μg/l Al		200
ΥΠΟΛΕΙΜΜΑΤΙΚΟ ΧΛΩΡΙΟ	mg/l Cl ₂		

* Αναφέρονται τα ανώτατα επιτρεπόμενα όρια για επιφανειακά νερά κατηγορίας Α2 και με πλαγιογραφημένους χαρακτήρες τα επιθυμητά όρια.