R & D),

,

e-mail: nthanas@eydap.gr, tsorovad@eydap.gr

Σε μία προσπάθεια αξιοποίησης απεικονιστικών εργαλείων από τον τομέα της επιχειρηματικής ευφυΐας και προκειμένου να αποτιμηθεί η ποιότητα του ποσίμου νερού του δικτύου ύδρευσης της ΕΥΔΑΠ ΑΕ, ανασύρθηκαν από τις βάσεις δεδομένων της Υπηρεσίας Ποιότητας Νερού αποτελέσματα περίπου 127000 προσδιορισμών που διενεργήθηκαν σε περισσότερα από 81000 δείγματα νερού κατά τα έτη 2010 – 2015. Για την επεξεργασία των δεδομένων χρησιμοποιήθηκαν το εμπορικό προϊόν Microsoft Excel 2016 και η γλώσσα προγραμματισμού R. Η προετοιμασία και ο έλεγχος των αποτελεσμάτων έγιναν με χρήση τεχνικών φιλτραρίσματος, ιστογραμμάτων, θηκογραμμάτων (box–plots) και διαγραμμάτων συσχέτισης. Εξ αιτίας της έλλειψης κανονικότητας στις κατανομές των δεδομένων, κρίθηκε σκόπιμο να αποφευχθούν παραμετρικές στατιστικές τεχνικές. Λόγω του ότι στο σύνολο των προσδιορισμών, 62% των τιμών αφορούσαν μετρήσεις υπολειμματικού απολυμαντικού και 16% αφορούσαν μετρήσεις θολότητας, δόθηκε ιδιαίτερη βαρύτητα στη μελέτη αυτών των παραμέτρων. Η χωροχρονική απεικόνιση των δεδομένων απολυμαντικού, θολότητας και ενός γενικού δείκτη ποιότητας που υπολογίστηκε με χρήση κατάλληλων συναρτήσεων (desirability functions) έδειξε ότι στο δίκτυο ύδρευσης δεν παρουσιάστηκαν προβλήματα καταλληλότητας του ποσίμου νερού. Η εφαρμογή των εργαλείων της επιχειρηματικής ευφυΐας αποδείχθηκε ικανοποιητική για την εκτίμηση της εξέλιξης της ποιότητας του νερού στο χώρο και το χρόνο. Ως συμπέρασμα προτείνεται τόσο η διασύνδεση των εργαλείων αυτών με online συστήματα μέτρησης για τον έγκαιρο εντοπισμό προβλημάτων σε πραγματικό χρόνο όσο και η χρήση τους ως εποπτικά μέσα κατά τη φάση αποτίμησης των αποτελεσμάτων.

Λέξεις – Κλειδιά: ποιότητα πόσιμου νερού, επιχειρηματική ευφυΐα, αναλυτική μεγάλων δεδομένων, desirability functions

Abstract

In this work, we present the utilization of certain visualization techniques from the area of business intelligence (BI) with a view to evaluating the quality of the drinking water in the EYDAP SA distribution network. Data were gathered during the years 2010-2015 and regarded the inorganic profile of water samples (apart from metals). All data treatment was carried out using Microsoft Excel 2016 and the R programming language. During the time period mentioned above, more than 81000 samples were analyzed in the laboratory and over

127000 results were stored in a relational database management system (RDBMS). Data checks and pretreatment were carried out by means of data filtering, histograms, box–plots and correlation matrix charts. As most chemical parameters did not follow the normal distribution, the use of parametric statistics was avoided. The determination of residual chlorine contributed by 62% to the total number of measurements followed by turbidity which contributed by 16%. The spatio–temporal visualization of the data with regard to residual chlorine and turbidity as well as an appropriately constructed quality index (calculated by means of desirability like functions) showed that no problems were detected in the overall water quality. Since the use of BI tools proved to be an efficient means for the estimation of drinking water quality in time and space, we propose the integration of these tools with online monitoring systems that will allow the prompt identification of problems followed by the necessary corrective actions so that high water quality standards are constantly met.

Keywords: drinking water quality, business intelligence, big data analytics, desirability functions

1.

Στο πλαίσιο των αρμοδιοτήτων της Υπηρεσίας Ποιότητας Νερού (ΥΠΝ) της ΕΥΔΑΠ ΑΕ και συμμορφούμενο με το ελληνικό νομοθετικό πλαίσιο, το Χημικό Εργαστήριο Γαλατσίου (ΧΕΓ) διεξάγει αδιάλειπτους χημικούς ελέγχους σε δείγματα νερού που προέρχονται από τους ταμιευτήρες απόληψης νερού, από τα ανεπεξέργαστα νερά που εισέρχονται στις Μονάδες Επεξεργασίας Νερού (ΜΕΝ), από τα επεξεργασμένα νερά των δεξαμενών των ανωτέρω διυλιστηρίων και από το δίκτυο ύδρευσης που καλύπτει μία μεγάλη περιοχή του νομού Αττικής. Διεξάγονται επίσης επιπλέον δοκιμές που εξυπηρετούν την ορθή λειτουργία των ΜΕΝ. Η διεξαγωγή αυτών των ελέγχων έχει ως αποτέλεσμα την παραγωγή ενός μεγάλου αριθμού δεδομένων, τα οποία εκτός της καθημερινής ερμηνείας τους απαιτούν τη χρήση κατάλληλων εργαλείων για τη συνολική εκτίμηση της ποιότητας του νερού και τη λήψη ορθών αποφάσεων όπου και όταν απαιτείται αυτό.

Το προαναφερθέν πρόβλημα της συσσώρευσης εκτεταμένου πλήθους δεδομένων και της γρήγορης και αποτελεσματικής τους ανάλυσης αποτελεί το αντικείμενο μελέτης της επιστήμης δεδομένων και της αναλυτικής μεγάλων δεδομένων. Ήδη υφίστανται πολλαπλά και ετερόκλητα δεδομένα από το πεδίο της διαχείρισης υδάτινων πόρων (Kambatla, Kollias, Kumar, & Grama, 2014), τον τομέα διανομής νερού για ανθρώπινη κατανάλωση (Kshetri, 2014), τις διαδικασίες παρακολούθησης της ποιότητας νερού μέσω αισθητήρων (Reis et al., 2015), τις τεχνικές εντοπισμού πηγών επιμόλυνσης πόσιμων υδάτων (Hu, Yang, Warty, & Xu, 2015) κ.α. Πέραν της ανάγκης κατανόησης των υπαρχόντων δεδομένων από περιβαλλοντικής πλευράς, στην περίπτωση των εταιρειών ύδρευσης, εμπλέκονται επίσης οικονομικοί παράγοντες οι οποίοι αφορούν τη βελτιστοποίηση των χημικών και μικροβιολογικών χαρακτηριστικών του προϊόντος (πόσιμο νερό) με ταυτόχρονη μείωση κόστους παραγωγής και συντήρησης των ανάλογων συστημάτων επεξεργασίας ανεπεξέργαστου νερού (Boulil, Le Ber, Bimonte, Grac, & Cernesson, 2014; Mounce, Pedraza, Jackson, Linford, & Boxall, 2015). Σε αυτή τη πρόκληση, προστίθεται και το πρόβλημα κατανόησης της ικανοποίησης του πελάτη, ο οποίος διατηρεί το δικαίωμα της

πρόσβασης στην πληροφορία που αφορά την ποιότητα του προσφερόμενου σε αυτόν αγαθού (Fróes Lima & Portillo Navas, 2012). Ο κλάδος της επιχειρηματικής ευφυίας καλείται συνεπώς να αξιοποιήσει μία σειρά εργαλείων προκειμένου να περιγράψει και να αναλύσει τις προαναφερθείσες ανάγκες βασιζόμενος σε πληθώρα δεδομένων από διαφορετικές πηγές. Τα αποτελέσματα της ανάλυσης θα πρέπει να αποδίδουν πληροφορίες σε μία μορφή κατανοητή τόσο από επιστήμονες διαφορετικών ειδικοτήτων όσο και από μη ειδήμονες (ευρύτερο καταναλωτικό κοινό) (Wang & Guo, 2013).

Βάσει των προαναφερθέντων προβληματισμών, στην τρέχουσα εργασία παρουσιάζουμε μια προσπάθεια αξιοποίησης απεικονιστικών εργαλείων από τον τομέα της επιχειρηματικής ευφυΐας για την αποτίμηση της ποιότητας του ποσίμου νερού του δικτύου ύδρευσης της ΕΥΔΑΠ ΑΕ κατά τα έτη 2010 – 2015. Για την επεξεργασία των δεδομένων χρησιμοποιήθηκαν το εμπορικό προϊόν Microsoft Excel 2016 και η ανοικτού κώδικα γλώσσα προγραμματισμού R. Η χρήση των δύο αυτών εργαλείων αποδεικνύει ότι η εφαρμογή σύγχρονων τεχνικών αναλυτικής μεγάλων δεδομένων δεν θα πρέπει απαραίτητα να συνεπάγεται αυξημένο κόστος υποδομών (λογισμικό). Τα αποτελέσματα που αφορούσαν περίπου 127000 προσδιορισμούς σε περισσότερα από 81000 δείγματα προετοιμάστηκαν και ελέγχθηκαν με χρήση τεχνικών φιλτραρίσματος, ιστογραμμάτων, θηκογραμμάτων (box–plots) και διαγραμμάτων συσχέτισης. Η έλλειψη κανονικότητας στις κατανομές των δεδομένων οδήγησε στην απόφαση αποφυγής χρήσης παραμετρικών στατιστικών τεχνικών. Λόγω της σημαντικότητας του προσδιορισμού των παραμέτρων του υπολειμματικού απολυμαντικού και της θολότητας για την εκτίμηση της υγειονομικής καταλληλότητας του πόσιμου νερού έγινε ειδική χωροχρονική μελέτη των τιμών των παραμέτρων αυτών προκειμένου να εκτιμηθεί η ποιότητα του νερού στο δίκτυο. Επιπλέον, υπολογίστηκε ένας συνολικός δείκτης ποιότητας νερού με χρήση συναρτήσεων επαναπεικόνισης (desirability functions) και γεωμετρικού μέσου όρου. Με τον τρόπο αυτό υπήρξε δυνατή και η χωροχρονική απεικόνιση της συνολικής ποιότητας του πόσιμου νερού. Δεδομένου ότι η εφαρμογή των εργαλείων της επιχειρηματικής ευφυΐας αποδείχθηκε ικανοποιητική για την εκτίμηση της εξέλιξης της ποιότητας του νερού στο χώρο και το χρόνο, προτείνεται συμπερασματικά η διασύνδεση των εργαλείων αυτών με online συστήματα μέτρησης για την έγκαιρη διάγνωση προβλημάτων σε πραγματικό χρόνο.

2.

2.1

Για την προετοιμασία και ανάλυση των δεδομένων χρησιμοποιήθηκε το εμπορικό προϊόν Microsoft Excel 2016 ως μέρος του Office 365. Μέρος της μελέτης υλοποιήθηκε με χρήση της γλώσσας προγραμματισμού R (version 3.2.3 – © 2015 The R Foundation for Statistical Computing) μέσω του ολοκληρωμένου περιβάλλοντος ανάπτυξης (Integrated Development Environment, IDE) RStudio (version 0.99.489 – © 2009-2015 RStudio, Inc.). Για τη μελέτη των δεδομένων, πλέον των βασικών πακέτων της R, απαιτήθηκε η εγκατάσταση των ακόλουθων πακέτων: *tidyr*, *treemap*, *ggmap*, *mapproj*, *ggplot2*, *colorspace*, *gridExtra* και *igraph*.

2.2

Τα δεδομένα της εργασίας αντλήθηκαν από την σχεσιακή βάση δεδομένων (RDBMS) Leviath 65, (version 2010-IAN-19) της Υπηρεσίας Ποιότητας Νερού της ΕΥΔΑΠ ΑΕ με κατάλληλη μορφοποίησή τους ώστε να μπορούν να εισαχθούν στο πρόγραμμα Excel 2016. Η περίοδος ενδιαφέροντος κάλυπτε πλήρως τα έτη 2010-2014 καθώς και το πρώτο εξάμηνο του έτους 2015. Από τα 111887 δείγματα της περιόδου αυτής απομονώθηκαν 81392 δείγματα που αφορούσαν το δίκτυο ύδρευσης της ΕΥΔΑΠ ΑΕ. Από τις 107 διαθέσιμες παραμέτρους που περιέγραφαν κάθε δείγμα, διατηρήθηκαν μόνο οι 22, οι οποίες αφορούσαν τις αναλύσεις που διεξάγονται στο Χημικό Εργαστήριο Γαλατσίου και περιλάμβαναν τις παραμέτρους του pH, της αγωγιμότητας, της αλκαλικότητας, του αμμωνίου, του αργιλίου, της γεύσης, των θειικών, της θολότητας, των κυανιούχων, των νιτρικών, των νιτρωδών, της οσμής, του διοξειδίου του χλωρίου, του χλωρίου, των φθοριούχων, των φωσφορικών, των χλωρικών, των χλωριωδών, των χλωριούχων και του χρώματος.

2.3

Οι γεωγραφικές συντεταγμένες των σημείων δειγματοληψίας για το δίκτυο ύδρευσης ελήφθησαν από την Υπηρεσία Εφαρμογών Γεωγραφικών Πληροφοριών G.I.S. της ΕΥΔΑΠ ΑΕ σε μορφή ζευγών γεωγραφικού πλάτους και μήκους με αντιστοίχιση ως προς τον κωδικό των δειγμάτων όπως αυτός έχει οριστεί από την Υπηρεσία Ποιότητας Νερού της ΕΥΔΑΠ ΑΕ. Οι χωρικές και χρονικές απεικονίσεις των αποτελεσμάτων έγιναν είτε με εισαγωγή των αντιστοίχων πινάκων στο μοντέλο δεδομένων του Excel 2016, είτε με απευθείας συγχώνευση των δεδομένων με χρήση της R.

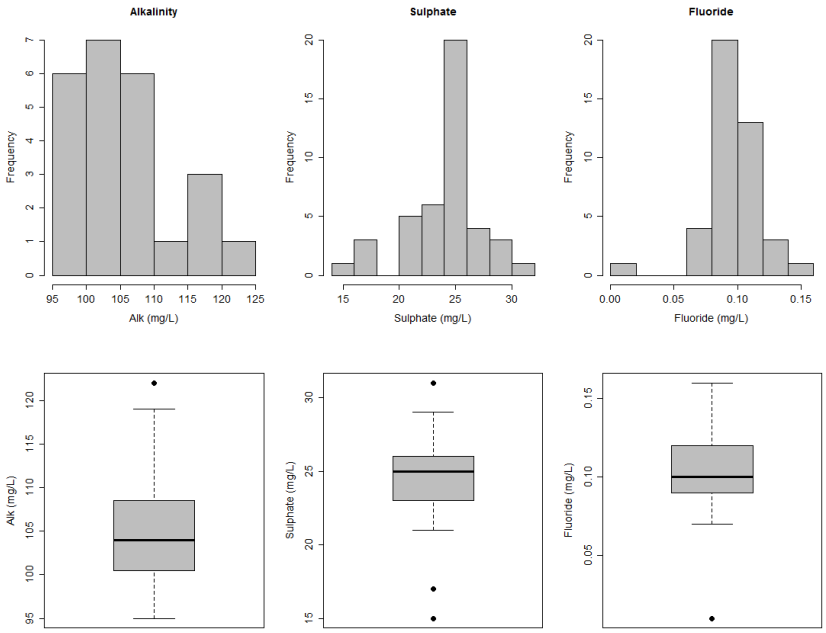
3.

3.1

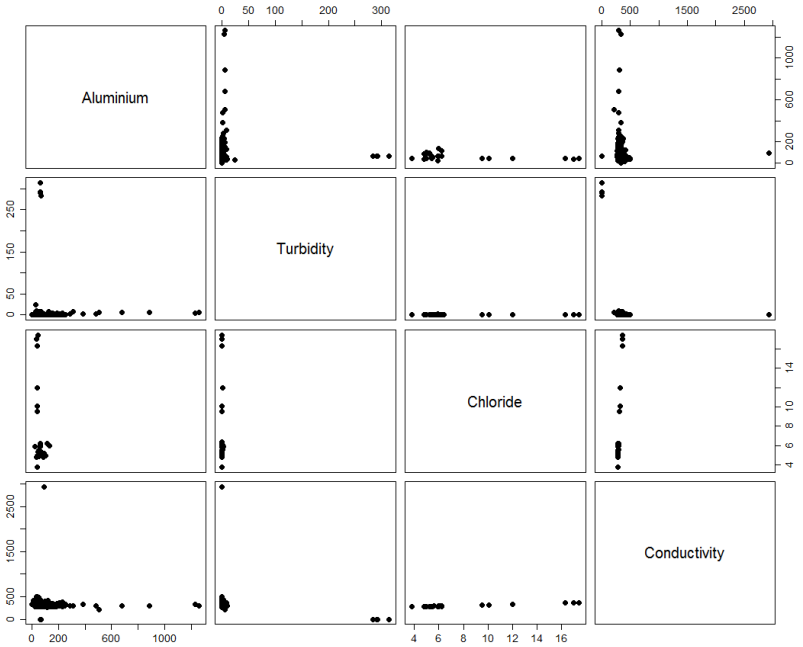
Ο έλεγχος της ορθότητας των δεδομένων έγινε με τη χρήση κατάλληλων φίλτρων στο Excel 2016 ώστε να εντοπιστούν ύποπτες τιμές των παραμέτρων (π.χ. pH<6.5). Χρησιμοποιήθηκαν επίσης ιστογράμματα και θηκογράμματα (box–plots) για τον εντοπισμό ακραίων τιμών καθώς και διαγράμματα συσχέτισης κατάλληλα επιλεγμένων ζευγών παραμέτρων οι οποίες αναμένονταν να συσχετίζονται θεωρητικά (π.χ. αγωγιμότητα και χλωριούχα).

Ενδεικτικά δίνονται εδώ ιστογράμματα και θηκογράμματα για την περίπτωση της αλκαλικότητας, των θειικών και των φθοριούχων (Σχήμα 1) καθώς και διαγράμματα συσχέτισης που αφορούν τις τιμές αργιλίου, θολότητας, χλωριούχων και αγωγιμότητας (Σχήμα 2). Όλες οι ύποπτες τιμές που εντοπίστηκαν με τις παραπάνω μεθόδους ελέγχθηκαν ανατρέχοντας στα αρχικά έντυπα καταγραφής δεδομένων που τηρούνται από το εργαστήριο όπως προβλέπεται από το ISO 17025 το οποίο εφαρμόζεται στα εργαστήρια της Υπηρεσίας Ποιότητας Νερού της ΕΥΔΑΠ ΑΕ. Δεδομένα για τα οποία δεν κατέστη δυνατόν να αποφασιστεί αν αποτελούσαν πραγματικά ακραίες τιμές ή σφάλματα εισαγωγής

απορρίφθηκαν από το τελικό αρχείο με διαγραφή της πλήρους σειράς του αντιστοίχου δείγματος. Το τελικό αρχείο το οποίο χρησιμοποιήθηκε για την εργασία αποτελείτο από 81346 σειρές (δείγματα) και 22 στήλες (παραμέτρους).



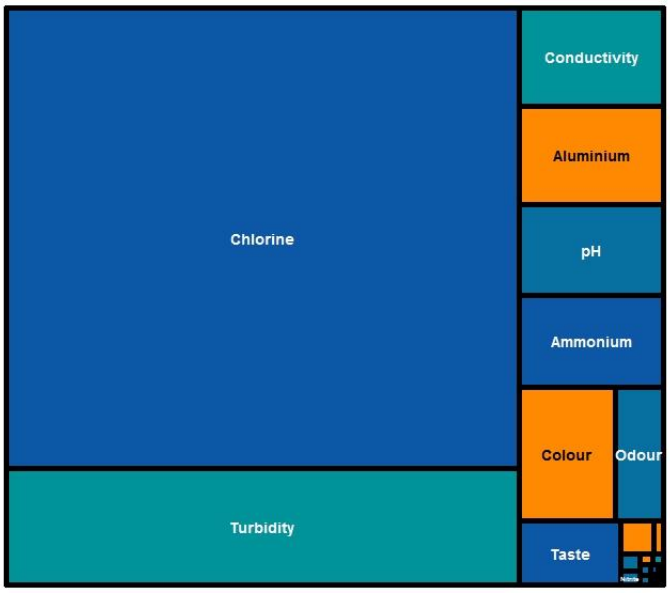
Σχήμα 1: Ενδεικτικά ιστογράμματα και θηκογράμματα για τον εντοπισμό ύποπτων τιμών



Σχήμα 2: Ενδεικτικά διαγράμματα συσχέτισης για τον εντοπισμό ύποπτων τιμών

3.2

Η ανάλυση της συμμετοχής κάθε προσδιορισμού στο σύνολο των μετρήσεων που διεξήχθησαν στο Χημικό Εργαστήριο Γαλατσίου έδειξε ότι ο προσδιορισμός υπολειμματικού απολυμαντικού υπήρξε ο συχνότερα εκτελούμενος προσδιορισμός με συνεισφορά 62% ακολουθούμενος από τον προσδιορισμό θολότητας με συνεισφορά 16% (Σχήμα 3). Οι δύο αυτοί προσδιορισμοί όντως αποτελούν το πρώτο στάδιο ελέγχου της υγειονομικής καταλληλότητας του νερού λόγω του ότι ικανοποιητικά επίπεδα υπολειμματικού απολυμαντικού και χαμηλές τιμές θολότητας αποτελούν σημαντικές προϋποθέσεις για την μείωση του κινδύνου μικροβιακής επιμόλυνσης. Σε κάθε περίπτωση, ο μικροβιολογικός έλεγχος αποτελεί το τελικό κριτήριο εκτίμησης της μικροβιολογικής ποιότητας του πόσιμου νερού.

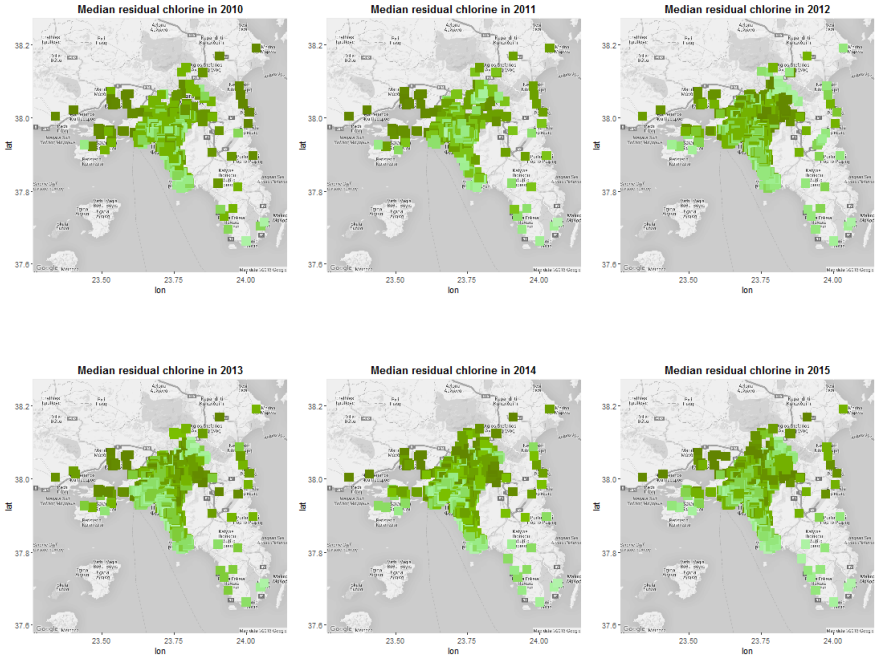


Σχήμα 3: Συμβολή προσδιορισμών στο σύνολο των μετρήσεων

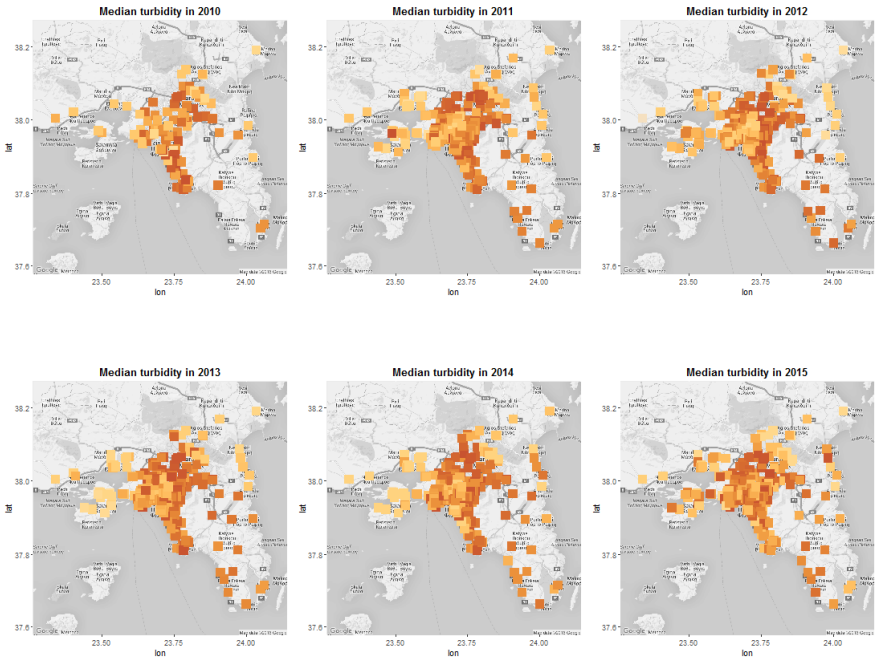
Λόγω της σημαντικότητας των τιμών υπολειμματικού απολυμαντικού και θολότητας, διεξήχθη κατάλληλη χωροχρονική μελέτη των παραμέτρων αυτών λαμβάνοντας υπόψη την διάμεσο τιμή των παραμέτρων ανά χρονική περίοδο (Σχήμα 4, Σχήμα 5). Η εικόνα των σχετικών μετρήσεων στο χώρο και τον χρόνο δεν φάνηκε να παρουσιάζει ουσιαστικές διαφορές υποδεικνύοντας έτσι μια σταθερή ποιότητα του πόσιμου νερού ως προς αυτά τα χαρακτηριστικά.

Προκειμένου να μελετηθεί χωροχρονικά η συνολική ποιότητα νερού στο δίκτυο ύδρευσης, απαιτήθηκε ο υπολογισμός ενός κατάλληλου δείκτη ποιότητας. Για τον υπολογισμό του δείκτη ήταν απαραίτητη η επαναπεικόνιση των αρχικών τιμών στο διάστημα [0,1]. Για τις τιμές θολότητας, αγωγιμότητας, αμμωνίου, αργιλίου και χρώματος χρησιμοποιήθηκε ο τύπος

$$x_{i,j}^{norm} = \frac{x_{i,j} - max_j}{min_j - max_j} \quad (1)$$



Σχήμα 4: Διάμεση τιμή υπολειμματικού απολυμαντικού ανά έτος



Σχήμα 5: Διάμεση τιμή θολότητας ανά έτος

όπου $x_{i,j}^{norm}$ το κανονικοποιημένο i στοιχείο της παραμέτρου j, $x_{i,j}$ το αρχικό i στοιχείο της παραμέτρου j, min_j η ελάχιστη τιμή της παραμέτρου j και max_j η μέγιστη τιμή της παραμέτρου j.

Για τις τιμές pH και υπολειμματικού απολυμαντικού χρησιμοποιήθηκε ο τύπος

$$x_{i,j}^{norm} = \begin{cases} \dfrac{x_{i,j} - min_j}{med_j - min_j}, & x_{i,j} \leq med_j \\ \dfrac{x_{i,j} - max_j}{med_j - max_j}, & x_{i,j} > med_j \end{cases} \quad (2)$$

όπου $x_{i,j}^{norm}$ το κανονικοποιημένο i στοιχείο της παραμέτρου j, $x_{i,j}$ το αρχικό i στοιχείο της παραμέτρου j, min_j η ελάχιστη τιμή της παραμέτρου j, med_j η διάμεσος της παραμέτρου j και max_j η μέγιστη τιμή της παραμέτρου j.

Η συνολική ποιότητα του κάθε δείγματος υπολογίστηκε ως ο γεωμετρικός μέσος όρος των επιμέρους κανονικοποιημένων τιμών βάσει του τύπου

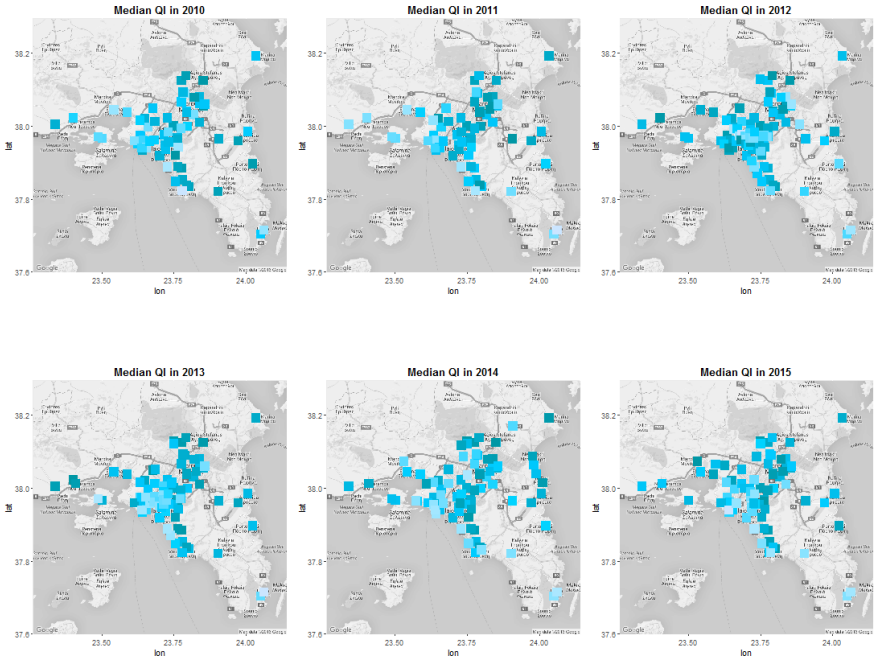
$$QI_i = \sqrt[7]{\prod_{j=1}^{7} x_{i,j}^{norm}} \quad (3)$$

όπου QI_i ο δείκτης ποιότητας του δείγματος i και $x_{i,j}^{norm}$ το κανονικοποιημένο i στοιχείο της παραμέτρου j.

Η χωροχρονική απεικόνιση του δείκτη ποιότητας έδειξε ότι δεν υφίστανται σημαντικές μεταβολές στο χρόνο όσον αφορά την ποιότητα του νερού δικτύου ύδρευσης της ΕΥΔΑΠ ΑΕ (Σχήμα 6).

4.

Από τη μελέτη που παρουσιάζεται εδώ φάνηκε ότι η χρήση «έξυπνων» εργαλείων απεικόνισης από τον τομέα της επιχειρηματικής ευφυίας μπορεί να επιτρέψει την εποπτική παρακολούθηση και ανάλυση της ποιότητας νερού σε ένα δίκτυο ύδρευσης. Βασική προϋπόθεση για να επιτευχθεί αυτό είναι η χρήση (ημι)αυτοματοποιημένων συστημάτων καταγραφής δεδομένων τα οποία μπορούν να τροφοδοτούνται είτε απ' ευθείας από τα σχετικά εργαστηριακά όργανα, είτε να ενημερώνονται από το προσωπικό του εργαστηρίου αμέσως μόλις παράγονται τα ανάλογα αποτελέσματα. Ιδανική περίπτωση αποτελεί η δυνατότητα σύζευξης των αναλυτικών οργάνων με τις αντίστοιχες βάσεις δεδομένων, σε τοπικό ή διαδικτυακό περιβάλλον (cloud), ώστε να αποφεύγονται σφάλματα κατά την αντιγραφή των αποτελεσμάτων. Τέτοιες βάσεις θα πρέπει να έχουν την δυνατότητα να δέχονται δεδομένα είτε σε batch μορφή, είτε σε συνεχή ροή και σε πραγματικό χρόνο. Η καταγραφή δεδομένων σε πραγματικό χρόνο θα παρείχε το πλεονέκτημα της on line παρακολούθησης της ποιότητας του νερού επιτρέποντας έτσι την έγκαιρη ενημέρωση των αρμοδίων τεχνικών τμημάτων ώστε τα προβλήματα να αντιμετωπίζονται το συντομότερο δυνατόν.



Σχήμα 6: Διάμεσος δείκτης ποιότητας ανά έτος

Η χρήση εργαλείων ανάλυσης όπως του Microsoft Excel 2016 και η γλώσσα προγραμματισμού R αποτελεί παράδειγμα οικονομικών μεθόδων πρόσβασης στην πληροφορία. Και τα δύο αυτά εργαλεία διαθέτουν τη δυνατότητα επικοινωνίας με βάσεις δεδομένων (ακόμα και αν αυτές βρίσκονται στο διαδίκτυο) και ειδικά η γλώσσα R παρέχει την ευελιξία τροποποίησης των αλγορίθμων ώστε κάθε φορά να επιτυγχάνεται το τρέχον ζητούμενο.

Αν και η αναλυτική μεγάλων δεδομένων έχει σχετικά πρόσφατα αρχίσει να χρησιμοποιείται σε συχνή βάση, ευελπιστούμε ότι σχετικά σύντομα η χρήση της, στα πλαίσια της έννοιας του Internet of Things (IoT), θα βρει εκτεταμένη εφαρμογή στην άντληση πληροφορίας από τα σχετικά συστήματα και θα συνεισφέρει στην λήψη αποφάσεων με σκοπό τη βελτίωση των διαδικασιών, της ποιότητας και του κόστους.

Η εργασία αυτή εκπονήθηκε και παρουσιάζεται με την έγκριση της ΕΥΔΑΠ ΑΕ. Ευχαριστούμε θερμά τις συναδέλφους του Χημικού Εργαστηρίου Γαλατσίου κα. Α. Καγιάρα, κα. Ε. Γεωργοπούλου, κα. Σ. Δήμου και τη φοιτήτρια κα. Σ. Μαυρομάτη για τον έλεγχο των αποτελεσμάτων καθώς και την Προϊσταμένη της Υπηρεσίας Εφαρμογών Γεωγραφικών Πληροφοριών G.I.S. κα. Δ. Ζάννου για την παροχή των γεωγραφικών συντεταγμένων των σημείων δειγματοληψίας. Θα θέλαμε επίσης να εκφράσουμε τις ευχαριστίες μας προς την Αναπληρώτρια Διευθύντρια και Προϊσταμένη της Υπηρεσίας Ποιότητας Νερού κα. Ζ. Μελαμπιανάκη και τη Διευθύντρια Ποιότητας, Έρευνας και Ανάπτυξης (R&D) κα. Μ. Ξανθάκη για την αμέριστη συμπαράστασή τους στην εκπόνηση και ανακοίνωση της παρούσας εργασίας.

5.

Boulil, K., Le Ber, F., Bimonte, S., Grac, C., & Cernesson, F. (2014). Multidimensional modeling and analysis of large and complex watercourse data: an OLAP-based solution. *Ecological Informatics*, *24*, 90–106. http://doi.org/10.1016/j.ecoinf.2014.07.001

Fróes Lima, C. A., & Portillo Navas, J. R. (2012). Smart metering and systems to support a conscious use of water and electricity. *Energy*, *45*(1), 528–540. http://doi.org/10.1016/j.energy.2012.02.033

Hu, J., Yang, K., Warty, C., & Xu, K. (2015). Special issue on big data inspired data sensing, processing and networking technologies. *Ad Hoc Networks*, *35*, 1–2. http://doi.org/10.1016/j.adhoc.2015.07.014

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561–2573. http://doi.org/10.1016/j.jpdc.2014.01.003

Kshetri, N. (2014). Big data′s impact on privacy, security and consumer welfare. *Telecommunications Policy*, *38*(11), 1134–1145. http://doi.org/10.1016/j.telpol.2014.10.002

Mounce, S. R., Pedraza, C., Jackson, T., Linford, P., & Boxall, J. B. (2015). Cloud Based Machine Learning Approaches for Leakage Assessment and Management in Smart Water Networks. *Procedia Engineering*, *119*, 43–52. http://doi.org/10.1016/j.proeng.2015.08.851

Reis, S., Seto, E., Northcross, A., Quinn, N. W. T., Convertino, M., Jones, R. L., … Wimberly, M. C. (2015). Integrating modelling and smart sensors for environmental and human health. *Environmental Modelling & Software*, *74*, 238–246. http://doi.org/10.1016/j.envsoft.2015.06.003

Wang, H. C., & Guo, J.-L. (2013). Constructing a water quality 2.0 OLAP system in Taiwan. *Journal of Cleaner Production*, *40*, 40–45. http://doi.org/10.1016/j.jclepro.2011.04.019