

Evaluation of Potable Groundwater Quality Using Environmetrics. The case of Nestos and Strymon River Regions, Northern Greece

Thomas Spanos^{1,*}, Nikolaos Mittas¹, Christina Chatzichristou¹, Konstantinos Dermentzis¹, Vilson Topi¹, Despina Selina Spanou², Antoaneta Ene³, Liliana Teodorof⁴, Elena Zubcov⁵ and Oleg Bogdevich⁶

¹International Hellenic University (IHU), Faculty of Natural Sciences, Department of Chemistry, St. Loukas, 65404 Kavala, Greece

²University of Thessaly, Department of Agriculture, Ichthyology and Aquatic Environment, Fytoko Street, 38 446 Volos, Greece

³Dunarea de Jos University of Galati, INPOLDE research center, Faculty of Sciences and Environment, Department of Chemistry, Physics and Environment, 47 Domneasca Street, 800008 Galati, Romania

⁴Danube Delta National Institute, 165 Babadag Street, Tulcea, Romania

⁵Institute of Zoology, 1 Academy Street, Chisinau, MD-2028, Republic of Moldova

⁶Laboratory of Geochemistry, Institute of Geology and Seismology, 3 Academy Street, Chisinau, MD-2028, Republic of Moldova

Received 12 January 2021; Accepted 26 February 2021

Abstract

In the present study potable water samples were regularly collected from two geographical regions in Northern Greece. The two regions are close to the Nestos and Strymon Rivers. Throughout 2015, samples from 13 sites were collected and analyzed with a frequency of three times per each quarter for the following parameters: chloride, calcium, nitrate, potassium, electric conductivity and total hardness. The main objective of this study was to investigate the effects of the geographical region and seasonal influence on the above parameters regarding the limits of the 98/83/EC Directive. The findings of the Linear Mixed Effects modeling revealed statistically significant main effects of both factors on the examined parameters.

Keywords: Northern Greece, Potable water, Groundwater quality, Environmetrics, Linear Mixed Effects Models

1. Introduction

Water for public consumption plays undoubtedly an important role for public health, human development and well-being. For these reasons potable water must meet distinct quality standards corresponding to various norms, regulations which are legislated and applied by the European Union (EU) as well as from institutions and national authorities of the countries whether they are members of the EU or not. These directives- regulations are aimed towards the insurance of the quality of potable water for example directive [1-5], just to mention some of the well-established common acts of EU.

The directives of water for the public consumption contain tables with parameters that must be examined such as: chemical, physical, microbiological and radiological, and contain the maximum admissible limits so the water may be characterized as potable. Also, the regulations refer to several other obligations such as: the number of analyses for each parameter per year in relation to the population, the obligatory water disinfection and controlling, and other obligations from the regional and national authorities which are charged with the control of quality and distribution of potable water to the population.

The limits of these parameters are legislated after extensive research and observations of the results that were provoked in the health of the population. This way it is logical for every new EU norm to modify some of the limits of the

parameters (e.g. NO₃), or to add the control of new parameters, due to the insertion of the potable water (e.g. the fabrication of pipeline for water distribution from new materials) or to remove some parameters because of their non-importance to the human health (e.g. water hardness).

Additionally, in order to treat specific problems which cause the limits to be surpassed in some areas when it is due to intensive meteorological phenomenon (climate change) or to the composition of the underground soil from which the potable water comes from, or in local problems as hot spot emissions, that are effected by the influence of air and soil pollution, waste disposal both domestic and industrial, organic components, fertilizers and organic pesticides used in agriculture, etc. [6-8], there is the opportunity for the authorities to use derogations. Derogation means in case of emergencies like the above, the local authorities ask for permission by the national authorities for the use of these waters for a limited time period (usually 2 years) by parallel declaration of the necessary work in the water body in order to reduce the parameter within the maximum admissible concentration by the end of the time period.

By systematically investigating the water quality in an area large data set may be collected. The data set of the drinking water supplied contains very important information. The classification, modeling and interpretation of the monitoring data could be a significantly important step in the complete assessment of the quality of potable water. By use of multivariate statistical approaches, such as linear mixed effects models, one is able to determine hidden information from the data as to possible environmental influences

*E-mail address: tspanos@chem.ihu.gr

ISSN: 1791-2377 © 2021 School of Science, IHU. All rights reserved.

doi:10.25103/jestr.141.13

(anthropogenic, naturally occurring, time influence) and find sources of the pollution in various environmental system [7,9,10].

The statistical methodologies can be an exceptional exploratory mechanism for interpreting datasets of complex water quality and for understanding spatial variations. This is useful and effective for the management of water quality. Therefore, information obtained by the application of multivariate data analysis methods will be a very useful addition to the required concentration values compared to the admissible threshold.

The aim of this study is to provide a multivariate statistical strategy in order to study the effects of the region and time period on the examined parameters regarding the limits authorized by the 98/83/EC Directive [11]. Multivariate statistical approach applied on data collected from different region sites (n=13) in the vicinity of the two rivers of the region, Nestos and Strymon. To this regard, six quality parameters (Chloride, Calcium, Nitrate, Potassium, Electric Conductivity, Total Hardness) from the above 13 region sampling sites of drinking water were monitored within a period of one year (2015), whereas samples were collected once every three to four months.

2. Experimental

2.1 Study Area and Sampling

This study was conducted in Kavala, Northern Greece and water quality monitoring data were collected once every three to four months from February 2015 to November 2015. The selection of examined parameters (Table 1) was based on requirements from the European directives 98/83/EC [11] and 2000/60/EU [12]. The collection of the samples was made from 13 region sampling sites from two different regions of Kavala located across the Nestos and Strymon rivers (Fig. 1). The 13 regions were selected due to their different geographical and geological characteristics. The location of the potable water region sampling sites (1. Ofrinio, 2. Galipsos, 3. Orfani, 4. Kokkinochori, 5. Kariani, 6. Podochori, 7. Akropotamos, 8. Piges, 9. Agiasma, 10. Nea Karia, 11. Chaidefto, 12. Monastiraki, 13. Keramoti).

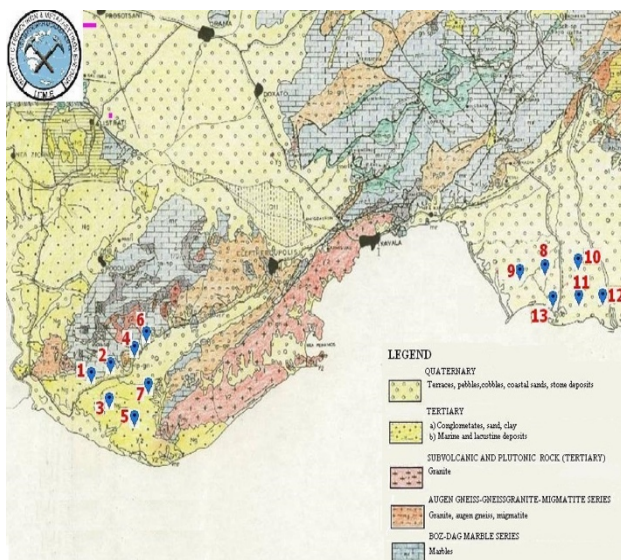


Fig. 1. The 13 regions sampling sites

2.2. Parameters and analytical methods

Table 1. Description of the examined parameters

Parameter	Description
Chloride (Cl)	Due to anthropogenic, agricultural, industrial influence
Calcium (Ca)	Due mainly to CaCO ₃ rock passages
Sodium (Na)	Natural factor, increases greatly with sea water contact
Potassium (K)	Usually in low levels in all natural, clean waters
Electric Conductivity (EC)	Due to the totality of dissolved ions in water
Total Hardness (TH)	Due mainly to CaCO ₃ rock passages

2.3. Statistical Methods

Univariate analysis for each parameter was carried out by descriptive statistics and results were expressed as mean (*M*), standard deviation (*SD*), median (*Mdn*), minimum (*min*) and maximum (*max*). In order to study the effect of factors Region (Nestos/Strymon) and Time (Feb/June/Nov) on the mean values of the examined parameters (Table 1), the *Linear Mixed Effects* (LME) modeling (Pinheiro and Bates, 2006) [13], which is a statistical model containing both fixed effects and random effects, were used. These models are useful in experimental designs, where repeated measurements are collected on the same statistical units, whereas they can also provide specialized treatment for data organized in a hierarchically or nested structure. In our study, samples were collected from specific sites that are nested within two regions and the parameters were measured repeatedly every four months for a one year follow-up period.

Regarding the fixed component structure, which describes the main and interaction terms that will be included in the final model, the optimal structure was defined through the protocol proposed by Zuur et al. (2009) [14]. Describing briefly, a model (defined as the *beyond model*) examining all factors of interest and their possible interactions is fitted and tested against a second model after omitting the higher order interaction term through the *Likelihood Ratio* (LR) test. In case of an insignificant finding, the selection is based on the *principle of parsimony*, which practically means that simpler models with similar explanatory power are preferred over more complex models with more parameters but slightly better fit. To this end, the *Akaike Information Criterion* (AIC) is used for the comparison process, while the model with the lowest AIC value was preferred over the competitive ones. Finally, we have also to point out that in cases, where the final model presented significant departure from homoscedasticity assumption of the residuals, the logarithmic transformation was applied to the raw measurements and the model was refitted on the transformed variable.

All statistical analyses were conducted using the statistical language R (R Core Team, 2013) [15] and the function *lme* from package *nlme*, respectively (Pinheiro and Bates, 2006) [13]. Furthermore, the Tukey's HSD procedure was used in multiple comparisons in order to control the family-wise error rate (FWER) using function *glht* from package *multcomp* (Bretz et al., 2016) [16] of the R statistical language. In all tests a difference was considered as statistically significant when *p*-value (significance) was less than 0.05. All the tests conducted were two-tailed (non-directional) in the sense that

the alternative hypothesis is that the measures tested are not equal.

Table 2 summarizes the descriptive statistics for each level of factors Region and Time from which, we can observe that there are differences between the two regions and among the examined time periods.

3. Results and Discussion

Table 2. Descriptive statistics of the examined parameters for each level of factors Region and Time

Parameter	Region	Time	N	M	SD	Mdn	min	max
Cl	Nestos	Feb	6	76.50	40.86	75.50	21.00	120.00
		June	6	80.32	45.00	78.90	19.40	130.00
		Nov	6	88.20	52.72	86.25	21.00	157.20
		Total	18	81.67	43.92	79.75	19.40	157.20
	Strymon	Feb	7	9.11	2.63	10.00	5.00	12.00
		June	7	8.80	1.66	8.90	5.60	10.60
		Nov	7	12.33	4.68	13.80	3.10	15.60
		Total	21	10.08	3.48	10.00	3.10	15.60
Ca	Nestos	Feb	6	3.17	0.41	3.00	3.00	4.00
		June	6	2.83	0.75	3.00	2.00	4.00
		Nov	6	2.57	0.78	2.30	2.00	4.00
		Total	18	2.86	0.68	3.00	2.00	4.00
	Strymon	Feb	7	53.71	7.59	52.00	47.00	70.00
		June	7	45.14	4.67	44.00	40.00	55.00
		Nov	7	41.71	3.04	41.00	39.00	48.00
		Total	21	46.86	7.30	45.00	39.00	70.00
Na	Nestos	Feb	6	191.83	26.48	175.50	173.00	226.00
		June	6	192.17	24.25	180.00	171.00	224.00
		Nov	6	186.83	27.05	172.50	166.00	223.00
		Total	18	190.28	24.51	176.00	166.00	226.00
	Strymon	Feb	7	15.97	9.27	13.50	8.30	36.00
		June	7	19.29	10.26	16.00	12.00	42.00
		Nov	7	16.43	10.95	13.00	10.00	41.00
		Total	21	17.23	9.78	14.00	8.30	42.00
K	Nestos	Feb	6	5.72	7.31	2.50	2.00	20.50
		June	6	5.90	6.94	2.90	2.30	19.90
		Nov	6	5.85	6.69	2.85	2.30	19.30
		Total	18	5.82	6.56	2.65	2.00	20.50
	Strymon	Feb	7	1.60	0.95	1.30	1.00	3.70
		June	7	1.94	0.87	1.60	1.50	3.90
		Nov	7	1.93	0.88	1.60	1.40	3.90
		Total	21	1.82	0.87	1.60	1.00	3.90
EC	Nestos	Feb	6	864.33	165.49	811.00	685.00	1068.00
		June	6	883.17	159.76	844.50	698.00	1076.00
		Nov	6	876.00	168.29	845.00	680.00	1071.00
		Total	18	874.50	154.77	817.00	680.00	1076.00
	Strymon	Feb	7	551.71	107.62	542.00	439.00	752.00
		June	7	557.29	87.84	520.00	498.00	752.00
		Nov	7	555.43	99.56	541.00	460.00	752.00
		Total	21	554.81	93.64	541.00	439.00	752.00
TH	Nestos	Feb	6	5.55	1.98	5.75	2.80	8.60
		June	6	5.66	2.12	5.65	2.60	8.90
		Nov	6	5.07	2.13	5.10	2.50	8.60
		Total	18	5.43	1.97	5.25	2.50	8.90
	Strymon	Feb	7	26.43	3.52	25.00	22.50	33.00
		June	7	23.28	2.51	23.10	19.50	26.70
		Nov	7	22.30	2.65	22.20	17.70	25.10
		Total	21	24.00	3.31	23.95	17.70	33.00

Table 3 summarizes the findings from the application of the protocol for each examined parameter, in which the beyond model (**Model A**) incorporating the main effects of Region and Time and their interaction term (Region×Time) was compared against the model (**Model B**) incorporating only the main effects (Region and Time). The findings

indicate that in the majority of the examined parameters (five out of six parameters), there was not noted a statistically significant interaction term, which practically means that the effect of Region factor on the changes of the examined parameters did not depend on the levels of factor Time. In contrast, for the case of Na parameter, there was noted a

statistically significant interaction term, $\chi^2(2) = 13.771$, $p = 0.001$ and the interaction term was retained in the final model.

Table 3. Application of the study protocol (LME models) for the examined parameters

Parameter	LMEM	df	AIC	Comparison
Cl	Model A	8	308.329	$\chi^2(2) = 4.852$, $p = 0.088$
	Model B	6	309.181	
Ca (log transformation)	Model A	8	-35.692	$\chi^2(2) = 0.564$, $p = 0.754$
	Model B	6	-39.416	
Na	Model A	8	-30.726	$\chi^2(2) = 13.771$, $p = 0.001$
	Model B	6	-20.955	
K	Model A	8	16.127	$\chi^2(2) = 1.164$, $p = 0.559$
	Model B	6	13.291	
EC	Model A	8	427.413	$\chi^2(2) = 5.879$, $p = 0.053$
	Model B	6	423.986	
TH (log transformation)	Model A	8	-33.993	$\chi^2(2) = 0.276$, $p = 0.871$
	Model B	6	-32.114	

Based on the application of the study protocol, the results of the Wald's tests are presented in Table 4 for all the examined parameters of selected final models and also the estimates for the main effects (and possible interaction terms), in which the reference categories for Region and Time were Nestos region and February, respectively. Due to space limitation and in order to better illustrate the results presented in this table, we provide details for an indicative parameter, i.e. Cl. The final model (fixed effects columns in Table 4) indicates statistically significant main effects of both Region, $F(1, 11) = 17.080$, $p = 0.002$ and Time, $F(2, 24) = 6.395$, $p = 0.006$ on the mean values of Cl measurements. Regarding the main effect of factor Region, the estimated coefficient for the region of Strymon ($b = -71.591$, $SE = 17.323$, $t(11) = -4.133$, $p = 0.002$) demonstrates that the mean value of Cl evaluated from the sites along Strymon river was 71.591 lower than the corresponding mean value evaluated from the samples collected from Nestos region. As far as the main effect of factor Time concerns, the post-hoc analysis through Tukey's HSD procedure (Table 5) reveals statistically significant differences between specific time periods, i.e. between November and February ($p = 0.002$) and between November and June ($p = 0.022$).

The inspection of the fixed effects from the set of the examined LME models can be summarized into the following points:

- The factor Region presented statistically significant main effects on five out of six examined parameters (except K parameter). More specifically, the samples derived from the sites located at Strymon region presented lower mean values for Cl and EC and higher mean values for Ca and TH parameters. The lower mean values for Cl is due to the fact that these waters are not affected by human, agricultural or industrial activities and the lower EC shows that the clean water has no contact with seawater as is the case for the Strymon region sites. The higher mean values for Ca and TH most often characterize water that passes through mountains rich in calcium carbonate rocks.
- For the case of K parameter, there was noted only a marginally significant difference between the mean values of the two regions.
- Regarding the effect of factor Time, the LME models revealed statistically significant main effect for five out of six parameters. In contrast, time seems not to present a statistically significant main effect on the mean values of EC parameter. The EC value corresponds to the amount of the total ion dissolved in the water. Characteristically underground water travels the same passage and the dissolved solids are always in a constant level thus the EC value is stable and independent of the factor of Time.

Table 4. Final LME models for the examined parameters and their fixed effects

Parameter	Factor	Fixed Effects				Estimates of fixed effects					
		num df	den df	F	p	Coefficient	Estimate	S.E.	df	t	p
Cl	Time	2	24	6.395	0.006	(Intercept)	78.765	12.769	24	6.169	< 0.001
	Region	1	11	17.080	0.002	Time:June	1.592	2.093	24	0.761	0.454
						Time:Nov	7.131	2.093	24	3.407	0.002
						Region:Strymon	-71.591	17.323	11	-4.133	0.002
Ca (log transformation)	Time	2	24	23.011	< 0.001	(Intercept)	1.155	0.067	24	17.221	< 0.001
	Region	1	11	1051.595	< 0.001	Time:June	-0.154	0.036	24	-4.252	< 0.001
						Time:Nov	-0.243	0.036	24	-6.704	< 0.001
						Region:Strymon	2.815	0.087	11	32.428	< 0.001
Na (log transformation)	Time	2	22	11.730	< 0.001	(Intercept)	5.249	0.141	22	37.247	< 0.001
	Region	1	11	174.864	< 0.001	Time:June	0.003	0.041	22	0.070	0.944

	Time:Region	2	22	7.682	0.003	Time:Nov	-0.027	0.041	22	-0.658	0.518
						Region:Strymon	-2.586	0.192	11	-	<
						Time:June:RegionStrymon	0.208	0.056	22	3.698	0.001
						Time:Nov:RegionStrymon	0.041	0.056	22	0.723	0.477
K	Time	2	22	49.458	< 0.001	(Intercept)	2.773	0.465	22	5.959	<
	Region	1	10	3.778	0.081	Time:June	0.342	0.041	22	8.401	<
						Time:Nov	0.358	0.041	22	8.811	<
EC	Time	2	24	0.813	0.455	Region:Strymon	-1.183	0.609	10	-1.944	0.081
	Region	1	11	19.174	0.001	(Intercept)	868.141	53.841	24	16.124	<
						Time:June	11.692	9.273	24	1.261	0.219
TH (log transformation)	Time	2	24	10.584	0.001	Time:Nov	7.385	9.273	24	0.796	0.434
	Region	1	11	94.452	< 0.001	Region:Strymon	-319.690	73.009	11	-4.379	0.001
						(Intercept)	1.691	0.118	24	14.314	<
						Time:June	-0.062	0.031	24	-2.009	0.056
						Time:Nov	-0.141	0.031	24	-4.589	<
						Region:Strymon	1.546	0.159	11	9.719	<
											0.001

Table 5. Post-hoc analysis for LME models presenting statistically significant main effect for factor Time

Parameter	Comparison	Estimate	p
Cl	June - Feb	1.592	0.727
	Nov - Feb	7.131	0.002
	Nov - June	5.538	0.022
Ca (log transformation)	June - Feb	-0.154	< 0.001
	Nov - Feb	-0.243	< 0.001
	Nov - June	-0.089	0.038
K	June - Feb	0.342	< 0.001
	Nov - Feb	0.358	< 0.001
	Nov - June	0.017	0.912
TH (log transformation)	June - Feb	-0.062	0.110
	Nov - Feb	-0.141	< 0.001
	Nov - June	-0.080	0.027

Acknowledgments

The investigations were done in the frame of the BSB27 MONITOX, Joint operational Programme Black Sea Basin (2014-2020). The authors thank the Municipal Authorities who financed this research program for the 13 sampling sites of the Nestos and Paggeio areas.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License.



References

- Council Directive 90/778/EEC, European Union Directive, (1990).
- Council Directive 91/692/EEC, European Union Directive, (1991).
- Council Decision 95/1/EC, Euratom, ECSC, (1995).
- Council Directive 94/567/EEC, European Union Directive, (1994).
- Directive of Greek Hygienic Office A5/228, (1986).
- T. Spanos, A. Ene, I.B. Karadjova, Assessment of toxic elements Cu, Cr, Ni, Pb, Cd, Hg, Zn, As and hexavalent chromium in sewage sludge from municipal wastewater treatment plants by combined spectroscopic techniques, Rom. Journ. Phys. 60(1-2), (2015).
- A. Ene, N. Tigau, M. Praisler, L. Moraru, Principal component analysis of physicochemical parameters of river water, Annals Univ. Dunarea de Jos Galati, Fasc. II. 33(1), 130 (2010).
- A. Ene, A. Bosneaga, L. Georgescu, Determination of heavy metals in soils using XRF technique, Rom. Journ. Phys. 55(7-8), 815, (2010).
- A. Ene, T. Spanos, Study of distribution and sources of polycyclic aromatic hydrocarbons (PAHs) in topsoils from SE Romania using multivariate statistical analysis, Annals Univ. Dunarea de Jos Galati, Fasc. II. 35, 15 (2012).
- V. Lovchinov, S. Tsakovski, Multivariate statistical approaches as applied to environmental physics studies, Cent. Eur. J. Phys. 4(2), 277 (2006).
- Council Directive 98/83/EC, Off. J. Eur. Comm. L330/32, (1998).
- Council Directive 2000/60/EC, Off. J. Eur. Comm. L 327/1, (2000).
- Pinheiro, J., & Bates, D. Mixed-effects models in S and S-PLUS. Springer Science & Business Media, (2006).
- Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., Smith, G. M., Mixed effects models and extensions in ecology with R. Springer Science & Business Media. (2009).
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (2013). URL <http://www.R-project.org/>.
- Bretz, F., Hothorn, T., & Westfall, P. Multiple comparisons using R. Chapman and Hall/CRC, (2016).