QUATERNARY GROUNDWATER VULNERABILITY ASSESSMENT IN LATVIA USING MULTIVARIATE STATISTICAL ANALYSIS

Inga Retike[1,2], Aija Delina[1], Janis Bikse[1], Andis Kalvans[1], Konrads Popovs[1], Daiga Pipira[1,2]
[1]University of Latvia
[2]Latvian Environment, Geology and Meteorology Centre
inga.retike@lu.lv

Abstract
Groundwater is the main drinking water source in Latvia, and Quaternary groundwater is widely used in households due to shallow occurrence. The identification of vulnerable areas is important for better water management and protection of deeper, more intensively used aquifers. The existing groundwater vulnerability map of Latvia does not take into account land use which can be an important factor affecting natural groundwater quality. Multivariate statistical methods - principal component analysis (PCA) and hierarchical cluster analysis (HCA) - were applied to identify groundwater groups with distinct water quality in Quaternary sediments in Latvia. On the basis of major ion concentrations and nitrogen compounds four distinct groundwater groups were identified. First group represents natural and most common calcium- magnesium bicarbonate water type in Latvia with low nitrate and ammonium concentrations. Samples from second and third group both reflect anthropogenic influence: diffuse agricultural contamination mostly with nitrates and/or contamination derived from artificial surfaces. Fourth group belongs to calcium bicarbonate water type and is characterised as a very young groundwater formed in sandy deposits. The results show that the highest concentrations of nitrogen compounds can be found in areas with agricultural land use or in artificial surfaces which are often classified as medium low or low vulnerability areas (mostly samples from group two and three). Meanwhile the lowest values of nitrogen compounds are present in areas where dominant land covers are forests and semi-natural areas or wetlands, and groundwater vulnerability classes are medium to high (samples from the first and fourth group).
Key words: Groundwater chemistry, principal component analysis, hierarchical cluster analysis, land use.

Introduction

The thickness of Quaternary deposits in Latvia varies from a few to 200 m and can reach even more in the areas where buried valleys occur. Due to shallow occurrence and ease of access Quaternary groundwater is widely used for small household needs in the whole territory of Latvia (Dēliņa, 2007).

Groundwater vulnerability reflects the ability of groundwater system to maintain natural condition and its sensitivity to contamination. Various parameters can be used as vulnerability indicators (for instance, site lithology, hydraulic conductivity, surface/ groundwater interaction, groundwater flow directions) and there are many approaches to evaluate aquifer vulnerability (Saidi *et al*., 2010; Valle Junior *et al*., 2014). Such factors as land use, type of contamination and groundwater exploitation are important factors affecting natural groundwater quality and should be taken into account.

Groundwater vulnerability map (Dēliņa & Prols, 2008) of Latvia outlines five vulnerability classes of water table aquifer based on lithological composition, hydraulic conductivity of the sediments, specific yield and recharge. It was made on the basis of Quaternary sediments map. As a result, five groups according to five vulnerability classes (starting from lowest) were outlined: 1) glaciolacustrine clay, 2) glacial till loam and sandy loam, 3) water saturated peat, 4) sand of different origin, 5) sand- gravel- pebbles sediments, mainly of glaciofluvial origin. The map of groundwater recharge modulus was used to adjust the contours of areas of different vulnerability classes. The map shows intrinsic vulnerability of water table aquifer, but it does not take into account land use or presence of sporadic shallow groundwater and its quality.

There have been several extensive studies on Quaternary groundwater in Latvia (Dēliņa, 2007; Levins & Gosk, 2007; Retike *et al*., 2016); however, there is a lack of research dealing with land use and groundwater vulnerability. The identification of vulnerable areas is necessary for sustainable groundwater management, in particular to meet the requirements of the Water Framework Directive. **The aim** of the study is to identify groundwater groups with distinct water quality in Quaternary sediments in Latvia using multivariate statistical analysis (principal component analysis (PCA) and hierarchical cluster analysis (HCA)) and compare the results with existing groundwater vulnerability map and land use data. This research demonstrates the validity of applying multivariate statistical methods (PCA and HCA) on groundwater chemistry to assess the compliance of the existing groundwater vulnerability map with observation data.

Materials and Methods

In this research, available data about Quaternary groundwater chemistry from previous studies (Levins & Gosk, 2007; Retike *et al*., 2016) and new data from national groundwater monitoring programme

Table 1

Principal component loadings and explained variance for three components with Varimax rotation

Parameter	PC1	PC2	PC3
Ca^{2+}	**0.89**	0.19	0.07
Mg^{2+}	**0.88**	0.28	0.08
Na^{+}	0.40	**0.74**	-0.08
K^{+}	0.20	**0.68**	0.22
HCO_3^-	**0.89**	0.09	0.07
Cl^-	0.40	**0.69**	0.11
SO_4^{2-}	0.32	0.49	0.23
NH_4^+	-0.19	**0.70**	-0.11
NO_2^-	-0.12	0.19	**0.82**
NO_3^-	0.29	-0.08	**0.82**
Eigenvalue	4.03	1.46	1.34
Explained variance (%)	40.32	14.63	13.36
Cumulative % of variance	40.32	54.95	68.30

Variables with principal component loadings greater than 0.5 are considered to be significant and are marked in bold.

carried out by the Latvian Environment, Geology and Meteorology Centre in years 2014 and 2015 were analysed.

In the beginning, samples having incomplete records of major ions (Ca^{2+}, Mg^{2+}, Na^+, K^+, Cl^-, SO_4^{2-}, HCO_3^-) were excluded from further analysis. Then samples having an ionic balance error greater that ± 10% were rejected from further analysis (Güler, Thyne, & McCray, 2002). For multiple samples from the same location median values were calculated. Data preparation for multivariate statistical analysis consisted of two steps: firstly data were log-transformed (except HCO_3^-) to achieve close to normal distribution, then data were standardized (Cloutier *et al.*, 2008; Güler, Thyne, & McCray, 2002). In total, a data set of 650 samples collected from monitoring springs and wells, springs, project wells, drainage, and water supply wells was used for further multivariate analysis. Hierarchical cluster analysis was made using Squared Euclidean distance as a similarity measure and Ward's method as a linkage method. The number of principal components was extracted based on the Kaiser criterion (Kaiser, 1958), and Varimax rotation was used (Cloutier *et al.*, 2008; Güler, Thyne, & McCray, 2002). Multivariate statistical analysis was performed on the basis of major ions (Ca^{2+}, Mg^{2+}, Na^+, K^+, Cl^-, SO_4^{2-}, HCO_3-) and nitrogen compounds (NO_3^-, NO_2^- and NH_4^+). Data pre-treatment and multivariate statistical analysis were performed using SPSS Statistics 22 and MS Excel 2013. CORINE land cover (2012) data were used to analyse the distribution of first level land use classes within clusters.

Results and Discussion

Based on Kaiser criterion (Kaiser, 1958), three principal components (PC) having eigenvalues greater than 1 were retained explaining 68% of the total variance in the data set (Table 1). PC1 explains the greatest amount of the variance and groups the high positive loadings of Ca^{2+}, Mg^{2+} and HCO_3^-. This component reflects the most common Ca-Mg-HCO_3 water type in Quaternary sediments in Latvia (Dēliņa, 2007). Highest positive loadings of parameters Na^+, K^+, Cl^- and NH_4^+ suggest that PC2 outline groundwater samples affected by mixing with Na-Cl dominated water (Retike *et al.*, 2016). PC3 groups highest positive loadings of NO_2^- and NO_3^-. Probably the PC3 indicate active nitrification process (Valle Junior *et al.*, 2014).

Dendrogram is the main result of hierarchical cluster analysis and the number of clusters was visually selected by moving the Phenon line (Güler, Thyne, & McCray, 2002). As a result, all samples were divided into four groups based on their geochemical similarities and dissimilarities, distribution of clusters can be seen in Figure 1. Median major ion and nitrogen compound concentrations as well as pH, total dissolved solids (TDS) and PC loadings in each of the distributed clusters are summarised in Table 2.

It can be observed that the highest median values of major ions (except HCO_3^-) can be found in C2. The highest median positive loading of PC2 and highest median TDS confirm the hypothesis of groundwater mixing with NaCl dominated water. Similar results were obtained in previous studies (Levins & Gosk, 2007; Retike *et al.*, 2016), although they were mostly

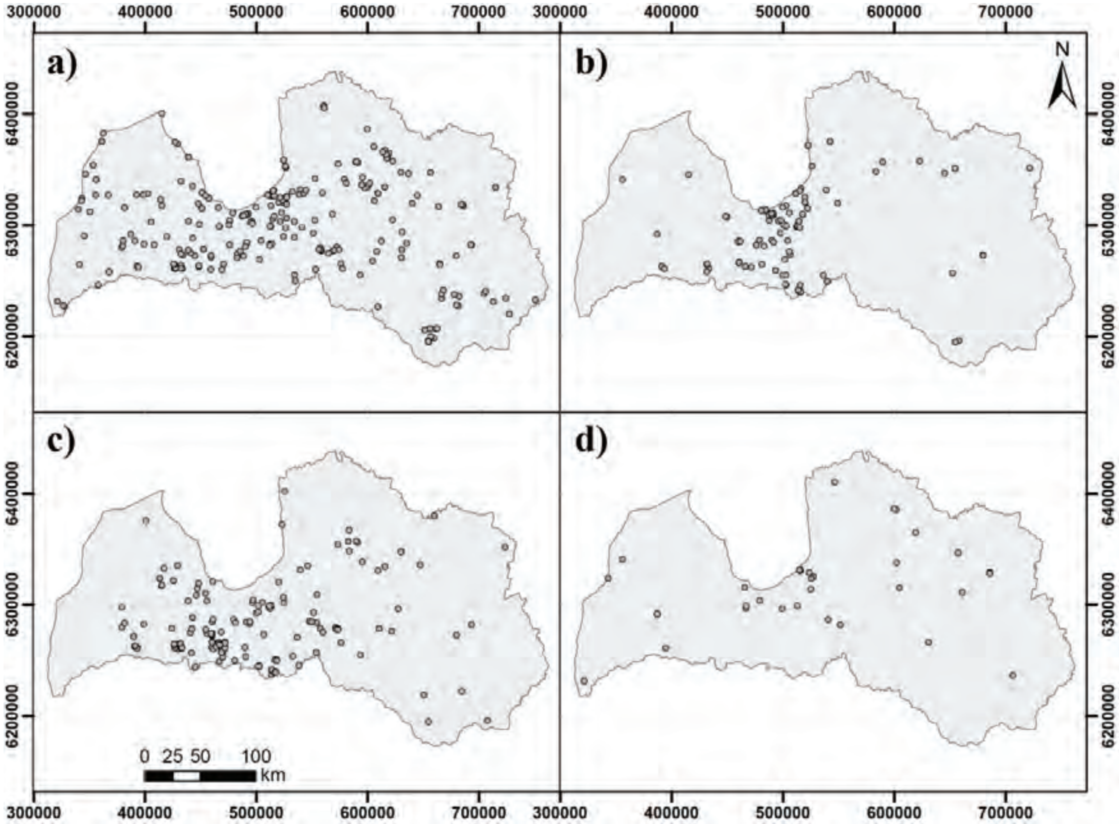

Figure 1. Distribution of sampling sites for four clusters obtained using hierarchical cluster analysis a) cluster 1; b) cluster 2; c) cluster 3; d) cluster 4.

from aquifers of greater depth. In some cases very high Na^+ and Cl^- values and Na/Cl ratio close to 1 were found in Quaternary groundwater as well (Retike *et al*., 2016) and can be explained as a result of halite dissolution delivered by roads de-icing (Cloutier *et al*., 2008). According to Dēliņa (2007), higher Na^+, Cl^- and K^+ values in Quaternary groundwater were found only in sandy deposits near the coastline and the distribution of sampling sites for C2 often match the previous observations (Figure 1). However, majority of the sampling sites in C2 are placed in the area near the capital city Riga and in Lielupe river basin, which is the most intensively used area for agricultural needs.

C4 shows the lowest median concentrations of all major ions, TDS and pH, as well as lowest median PC1 loading compared with other clusters (Table 2). All listed parameters in C4 have concentrations much lower than in slightly altered precipitation water found in earliest study (Retike *et al*., 2016). It can be seen (Figure 2) that C4 reflects Ca-HCO_3 water type and there are relatively low Mg^{2+} concentrations compared to Ca^{2+} concentrations (Table 2). For this reason C4 can be interpreted as very young groundwater formed in sandy deposits. C1 groups Ca-Mg-HCO_3 type groundwater with most typical chemical characteristic for Quaternary sediments in Latvia (Dēliņa, 2007) and sampling sites are distributed widely (Figure 1). Groundwater samples from C3 also belong to Ca-Mg-HCO_3 water type and C3 has highest median Ca^{2+}, HCO_3^- and NO_3^- values, highest positive PC1 and high positive PC3 loading. Moreover, very similar results were obtained in previous study (Retike *et al*., 2016) suggesting that such high median values of Ca^{2+}, HCO_3^- and NO_3^- as well as highlighted SO_4^{2-} and Cl^- values are not typical for groundwater and may indicate to diffuse agricultural influence. For example, Valle Junior *et al*. (2014) suggest that ploughing may promote the dissolution of carbonate and gypsum in the soils and increase the value of TDS in groundwater. Likewise, the placement of sampling points for C3 (Figure 1) in most intensively used area for agricultural needs- Lielupe river basin- supports the theory.

As it can be seen in Table 2, the highest NO_3^- concentrations can be found in C2 and C3. According to CORINE land cover data (Figure 3), C3 mostly represents agricultural areas and then artificial surfaces, but C2- artificial surfaces and then agricultural areas. Consequently, this explains the highest Na^+ and Cl^- concentrations in C2 and supports the theory of possible road de-icing influence. Likewise, the C3 shows the highest NO_3^- concentrations most likely produced via

Table 2

Chemical characteristics of four clusters and whole data set (median values)

Parameter	C1 (N = 298)	C2 (N = 121)	C3 (N = 194)	C4 (N = 37)	All samples (N = 650)
Ca^{2+} (mg l^{-1})	65.5	**105.0**	**105.0**	18.0	81.0
Mg^{2+} (mg l^{-1})	16.0	**34.0**	30.0	3.1	22.6
Na^{+} (mg l^{-1})	4.1	**20.0**	7.2	2.2	5.7
K^{+} (mg l^{-1})	1.5	**9.9**	2.5	1.4	2.2
HCO_3^- (mg l^{-1})	252.5	360.0	**412.5**	60.0	315.0
Cl^- (mg l^{-1})	6.0	**33.0**	15.0	4.0	11.0
SO_4^{2-} (mg l^{-1})	13.0	**52.0**	26.5	8.2	20.0
NH_4^+ (mg l^{-1})	0.18	**0.48**	0.14	0.19	0.18
NO_2^- (mg l^{-1})	0.01	0.01	0.01	0.01	0.01
NO_3^- (mg l^{-1})	1.20	10.23	**11.60**	1.68	3.19
pH	**7.4**	7.3	**7.4**	6.5	**7.4**
TDS[1] (mg l^{-1})	383.8	**776.1**	630.5	100.9	495.7
PC1	-0.29	0.37	**0.78**	-2.27	0.10
PC2	-0.50	**1.46**	-0.16	-0.33	-0.16
PC3	-0.41	**0.57**	0.34	0.38	0.02

The highest values are marked in bold and the lowest are underlined; N- number of samples; [1]TDS calculated as the sum of major ions Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, Cl^-, SO_4^{2-} and NO_3^-; PC- principal component loading.

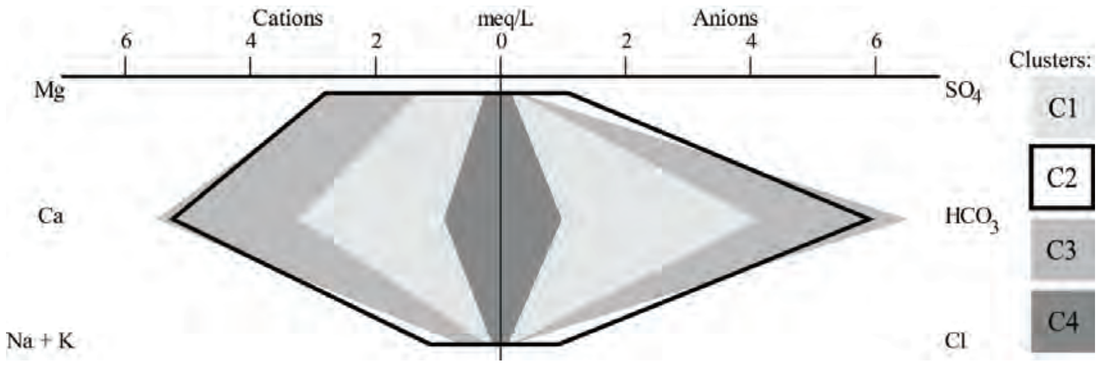

Figure 2. Median values of four clusters on Stiff diagram (meq/L).

the nitrification of N-fertilizers (Valle Junior *et al*., 2014). The lowest median NO_3^- concentrations are found in C1 and C4. In both clusters dominant land covers are forests and semi-natural areas or wetlands.

However, the distribution of existing vulnerability classes within four clusters shows unexpected results. For example, groundwater from C1 and C4, which are considered to be of good quality and non-affected by pollution are placed in the areas with high to medium groundwater vulnerability (Figure 3). Unlike the samples in C3 and C4, which have the highest nitrate concentration and indicate anthropogenic influence, they are placed in areas with medium to low vulnerability, especially for C3. In conclusion, the existing vulnerability classes at this point are useless as a tool for water management. More research should be carried out to improve the methodology how groundwater vulnerability classes are defined and the first step should be taking into account the pressures from diverse land use.

Conclusions

1. Four distinct groundwater groups were identified using multivariate statistical analysis. All groups belong to bicarbonate water type. The first group represents natural and most common calcium-magnesium bicarbonate water type in Latvia with low nitrate and ammonium concentrations that can be found in the whole territory of Latvia. Samples from second and third group both reflect anthropogenic influence: diffuse agricultural contamination mostly with nitrates produced

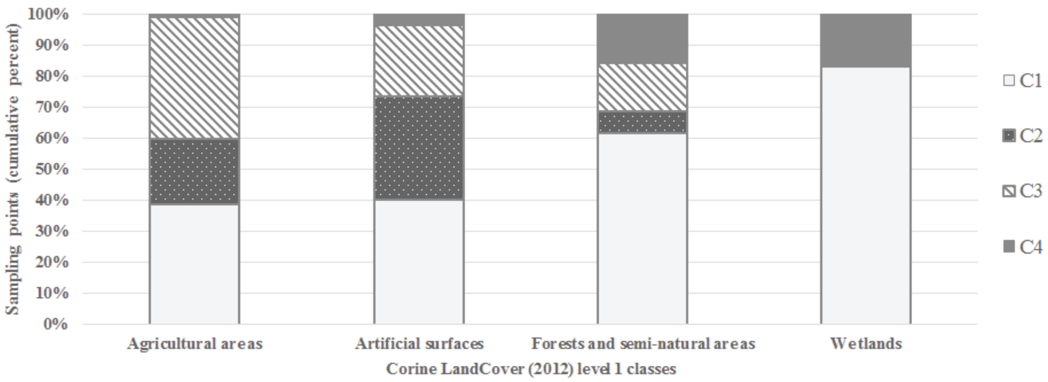

Figure 3. CORINE land cover (2012) level 1 classes within distributed four clusters.

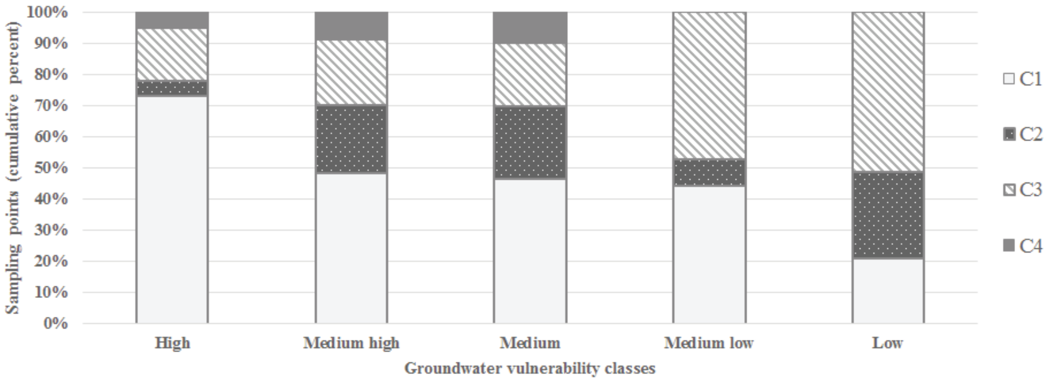

Figure 4. Groundwater vulnerability classes within distributed four clusters.

via the nitrification of N-fertilizers and / or contamination derived from artificial surfaces, for example, high Na^+ and Cl^- concentrations from road de-icing. Fourth group belongs to calcium bicarbonate water type and is characterised as very young groundwater located mostly in sandy deposits.

2. The results show that the highest concentrations of nitrogen compounds can be found in areas with agricultural land use or artificial surfaces which mostly are classified as medium low or low vulnerability areas. Meanwhile, the lowest values are present where dominant land covers are forests and semi-natural areas or wetlands and groundwater vulnerability classes are medium to high.
3. Fertility of the soil depends on the amount of clay minerals in it; therefore, areas with the highest clay content are mostly used in agriculture. Likewise, areas with more clay in soil are considered to be of lower vulnerability. As a result, areas with lowest groundwater vulnerability are mostly used in agriculture and have the highest anthropogenic pressure on groundwater.
4. The results show that all Quaternary groundwater in Latvia becomes vulnerable at a certain level of pressure and that the existing knowledge on groundwater vulnerability assessment in Latvia is insufficient; therefore, further researches are encouraged.

Acknowledgements

This research was supported by NRP project EVIDENnT project ‘Groundwater and climate scenarios’ subproject ‘Groundwater Research’ and by Science-based funding of Latvian Ministry of Education and Science via ‘Attraction of Human Resources to Development of Scientific Study in the area of Earth and Environmental Sciences’ programme.

References

1. Cloutier, V., Lefebvre, R., Therriern, R., & Savard, M.M. (2008). Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J. Hydrol.* 353(3–4), 294-313. DOI: 10.1016/j.jhydrol.2008.02.015.

2. Dēliņa, A., & Prols, J. (1998). Latvijas pazemes ūdeņu aizsargātības karte (*Groundwater vulnerability map of Latvia*). SIA Geo Consultants, Rīga, 34. lpp. (in Latvian).
3. Dēliņa, A. (2007). *Kvartārsegas pazemes ūdeņi Latvijā (Quaternary groundwater in Latvia).* Summary of doctoral thesis, University of Latvia, Riga. (in Latvian).
4. Güler, C., Thyne, G.D., & McCray, J.E. (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol. J.* 10(4), 455-474. DOI: 10.1007/s10040-002-0196-6.
5. Kaiser, H.F. (1958). The Varimax criteria for analytical rotation in factor analysis. *Psychometrika*. 23(3), 187-200. DOI: 10.1007/BF02289233.
6. Levins, I., & Gosk, E. (2007). Trace elements in groundwater as indicators of anthropogenic impact. *Environ. Geol.* 55(2), 285-290. DOI: 10.1007/s00254-007-1003-4.
7. Retike, I., Kalvans, A., Popovs, K., Bikse, J., Babre, A., & Delina, A. (2016). Geochemical classification of groundwater using multivariate statistical analysis in Latvia. *Hydrology Research*. In Print. DOI: 10.2166/nh.2016.020.
8. Saidi, S., Bouri, S., & Ben Dhia, H. (2010). Groundwater vulnerability and risk mapping of the Hajeb–Jelma aquifer (Central Tunisia) using a GIS-based DRASTIC model. *Environ Earth Sci.* 59(7), 1579-88. DOI: 10.1007/s12665-009-0143-0.
9. Valle Junior, R.F., Varandas, S.G.P., Sanches Fernandes, L.F., & Pacheco, F.A.L. (2014). Groundwater quality in rural watersheds with environmental land use conflicts. *Sci. Total Environ.* 493, 812-827. DOI: 10.1016/j.scitotenv.2014.06.068.