Geochemical classification of groundwater using multivariate statistical analysis in Latvia

Inga Retike, Andis Kalvans, Konrads Popovs, Janis Bikse, Alise Babre and Aija Delina

ABSTRACT

Multivariate statistical methods – principal component analysis (PCA) and hierarchical cluster analysis (HCA) – are applied to identify geochemically distinct groundwater groups in the territory of Latvia. The main processes observed to be responsible for groundwater chemical composition are carbonate and gypsum dissolution, fresh and saltwater mixing and ion exchange. On the basis of major ion concentrations, eight clusters (C1–C8) are identified. C6 is interpreted as recharge water not in equilibrium with most sediment forming minerals. Water table aquifers affected by diffuse agricultural influences are found in C3. Groundwater in C4 reflects brine or seawater admixture and gypsum dissolution in C5. C7 and C2 belong to typical bicarbonate groundwater resulting from calcite and dolomite weathering. Extremely low Cl^- and SO_4^{2-} are observed in C8 and described as pre-industrial groundwater or a solely carbonate weathering result. Finally, C1 seems to be a poorly defined subgroup resulting from mixing between other groups. This research demonstrates the validity of applying multivariate statistical methods (PCA and HCA) on major ion chemistry to distribute characteristic trace elements in each cluster even when incomplete records of trace elements are present.

Key words | groundwater chemistry, hierarchical cluster analysis, principal component analysis, trace elements

Inga Retike (corresponding author)
Konrads Popovs
Janis Bikse
Alise Babre
Aija Delina
Faculty of Geography and Earth Sciences,
Centre of Geological Processes Research and
   Modelling, University of Latvia,
Jelgava Street 1,
Riga,
LV-1004,
Latvia
E-mail: inga.retike@lu.lv

Andis Kalvans
Faculty of Science and Technology,
Institute of Ecology and Earth Sciences, University
   of Tartu,
Tartu,
Estonia

INTRODUCTION

Groundwater is the most important water supply source in Latvia due to its overall good quality and sufficient quantity compared to the amount of water usage. The main processes controlling groundwater composition are water rock interaction and water mixing (Appelo & Postma 2005). Yet, other factors such as diffuse agricultural pollution or extensive groundwater extraction may intensify or change the processes in natural groundwater systems (Helena 2000; Levins & Gosk 2007). Due to natural and human induced variability in time and space the interpretation of groundwater composition can be complicated.

The most common groundwater type in the active water exchange zone is Ca-Mg-HCO₃ due to the omnipresent carbonate minerals in the Quaternary cover and the humid climate. The Ca-SO₄ groundwater is a result of gypsum dissolution and is regionally widespread. More locally, high salinity Na-Cl rich groundwater can be observed due to freshwater mixing with saltwater (Levins et al. 1998; Spalvins et al. 2004).

Large pumping of an overlaying freshwater aquifer can cause saltwater upward intrusion (Marandi & Karro 2008), which can be either naturally occurring and activated by groundwater extraction or uncharacteristic for an area. Intensive groundwater pumping started in the early 1960s until the 1990s, causing a significant change in piezometric levels in the Riga region. A noticeable increase of salinity in fresh groundwater around the city of Riga has been observed (Levins 1990).

Coastal groundwater aquifers sometimes are affected by seawater intrusion, in particular if large cities are using them

as a drinking water source (Klimas & Plankis 2006; Marandi & Karro 2008; Mukherjee & Fryar 2008). Two well-documented sites within the territory of Latvia are Liepaja (Levina & Levins 2001; Spalvins *et al.* 2005) and Riga (Spalvins 1997; Levins *et al.* 1998; Spalvins *et al.* 2004).

All the processes mentioned above affect not only the concentration of major anions and cations, but also many trace elements which can be useful indicators of geochemical processes and water residence time especially when complicated water mixing occurs (Edmunds & Smedley 2000; Helena 2000).

Most of the studies in the Baltic region have concentrated on particular problems and related trace elements (Klimas & Mališauskas 2008; Mokrik *et al.* 2009; Raidla *et al.* 2009; Hiiob & Karro 2012; Karro & Uppin 2013).

Apart from extensive study of trace elements in shallow groundwater in the agricultural territories (Levins & Gosk 2007) there is a lack of comprehensive overview of the trace elements in groundwater in Latvia. For the first time a comprehensive data set on trace element concentrations in groundwater have been made and analyzed. The main objective of this paper was to examine characteristic trace elements in each of the distributed groups and to propose an insight into major geochemical processes responsible for the evolution of each group. In this study, eight geochemically distinct groundwater groups were defined using multivariate statistical methods – principal component analysis (PCA) and hierarchical cluster analysis (HCA).

In this research, new information on baseline groundwater quality in Latvia is given which is useful for future trend assessment and related studies in the region. In a wider context, the results provide new knowledge on possible groundwater composition evolution paths in sedimentary basins. This research demonstrates the validity of applying multivariate statistical methods (PCA and HCA) on major ion chemistry even when incomplete records of trace elements are present.

STUDY AREA AND DATA

Hydrogeological setting

The study area covers the central part of the Baltic Artesian Basin. The thickness of the sedimentary cover varies from about 500 m in the northern part to more than 2,000 m in the southwestern part of Latvia (Lukševičs *et al.* 2012).

Three hydrodynamical and hydrochemical zones of groundwater are traditionally identified within the study area (Jodkazis 1989; Levins *et al.* 1998) (Figure 1):

- stagnation zone: Ediacaran-Cambrian aquifer complex with brines;
- passive (slow) water exchange zone: lower and middle Devonian aquifer complex with brackish groundwater;
- active water exchange zone: freshwater aquifers above Narva regional aquitard.

All zones present within the study area are separated by regional aquitards.

The Ediacaran-Cambrian aquifer complex lies on top of the crystalline basement and is composed of sandstones, siltstones, and clays (Lukševičs *et al.* 2012). The thickness of this complex varies from 50 to 150 m. Strong brine is encountered at this complex with Na^+ and Cl^- as the dominant ions (Kalvāns 2012).

Faults are important factors controlling the aquifer connectivity for the stagnation zone. The northern and southern parts of the aquifer are virtually separated by the major Pleskov–Liepaja fault zone (Brangulis & Kaņevs 2002), where contrasting water salinity often is observed at opposite flanks of major faults (Levins 1990).

The Ordovician–Silurian sedimentary sequence is composed of deep marine facies – marls and clays with occasional limestone and dolostone beds and forms a regional aquiclude separating the Ediacaran-Cambrian aquifer (Lukševičs *et al.* 2012; Virbulis *et al.* 2013). The thickness of this aquiclude varies between 80 m in the southeast to 800 m in the west, dipping from about 200 m below sea level in the north to more than 800 m below sea level in the south (Popovs *et al.* 2015).

Lower Devonian Gargzdu Fm to middle Devonian Parnu Fm forms the lower to middle Devonian aquifer system of the passive (brackish) water zone within the whole research territory (Levins *et al.* 1998). Predominantly, it is composed of sandstones, with siltstones, marls, and clays reaching a thickness of 200 m in the western part of the aquifer. This zone is dominated by brackish water with high SO_4^{2-} levels along with elevated contents of Cl^- and Na^+ ions (Kalvāns 2012).
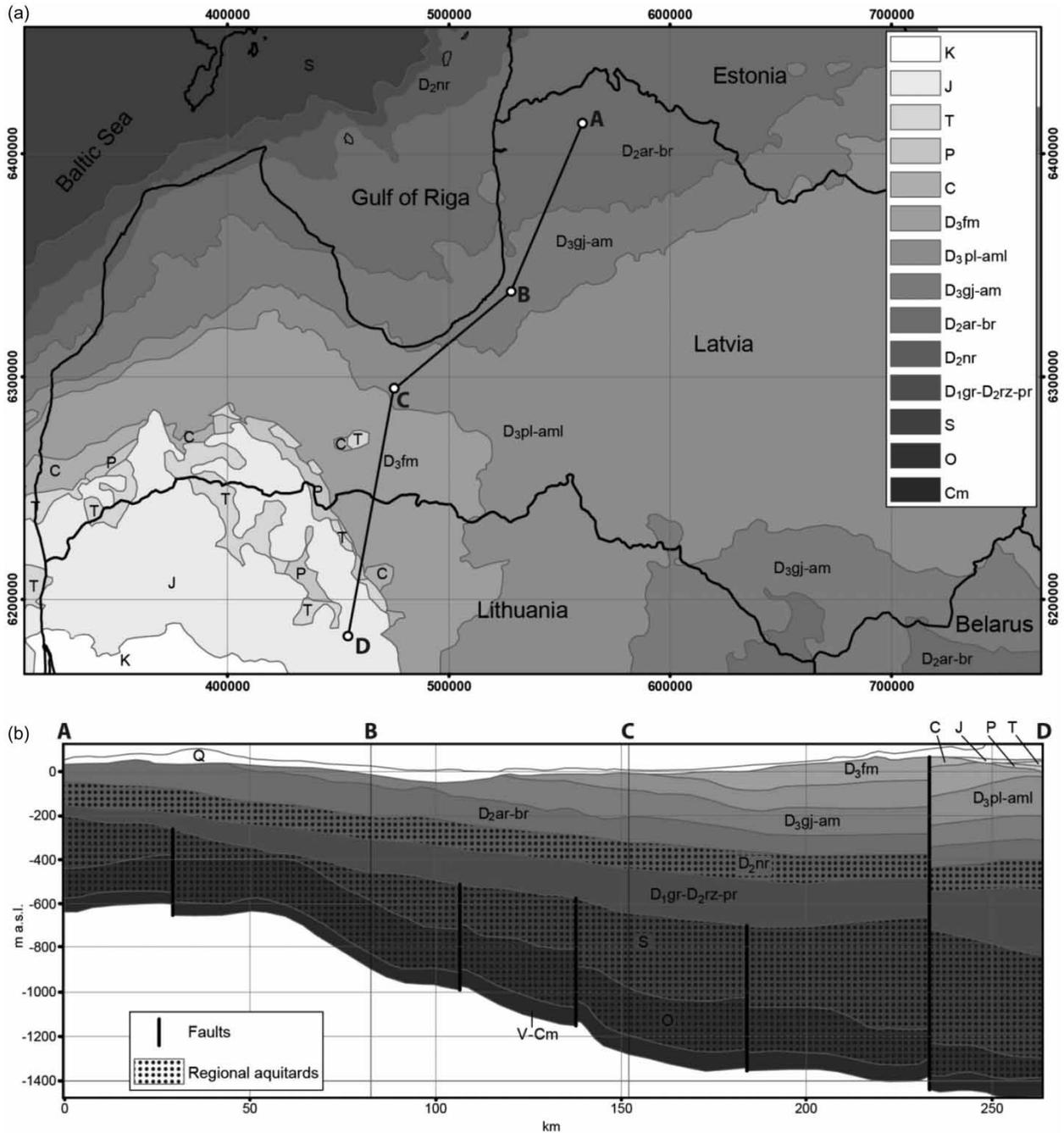
Figure 1 │ Geological map and geological cross-section of the study region without Quaternary cover (modified after Virbulis et al. 2013; Popovs et al. 2015). (a) Geological map. Labelled line denotes location of cross-section. V-Cm, Ediacaran-Cambrian sequence; O, Ordovician sequence; S, Silurian sequence; D_1gr-D_2rz-pr, lower Devonian Gargzdu Fm to middle Devonian Parnu Fm; D_2nr, middle Devonian Narva formation; D_2ar-br, middle Devonian Burnieki Fm to Arukila Fm; D_3gj-am, upper Devonian Gauja Fm to Amata Fm; D_3fm, upper Devonian Famena Fm; C, Carboniferous sequence; P, Permian sequence; T, Triassic sequence; J, Jurassic sequence; K, Cretaceous sequence. (b) Geological cross-section. Thick vertical lines denote major fault structures.

Narva formation is an important regional aquitard. Its thickness varies from 100 m in eastern Latvia to 200 m in western Latvia (Popovs et al. 2015). Sediments of middle and upper Devonian to Quaternary age form an active water exchange zone. A substantial part of this zone is formed by the sequence of clastic sediments that are

stratigraphically relevant to Arukila, Burtnieki, Gauja, and Amata formations. This sequence has a rhythmical structure where sandstones predominate at the base of each formation and fine-grained siltstones and clays dominate at the upper part (Lukševičs *et al.* 2012).

Above this terrigenous sequence of middle-upper Devonian, a pie of interlayered dolostones, clay dolomites, dolomitic marls, limestones, marls, clays, silts, sandstones, and occasional gypsum of the upper Devonian Frasnian and Famennian stages reside (Lukševičs *et al.* 2012). The complex is present in a large part of Latvia, missing only on its northern edges and the southeast. It gains particular importance at the southwestern edge of Latvia where its thickness is approaching 300 m.

The whole region is covered by Quaternary, mostly glacial and marine sediments, which discordantly lie atop of the Middle Devonian–Jurassic sequence. The almost omnipresent glaciotectonic dislocations and heterogeneous nature of the sediments denote a very complicated structure of the Quaternary sequence. From a hydrogeological point of view it is important in upland areas, where patches of glacial till loams (aquitards) and glaciofluvial sand and gravel (aquifers) sequences can exceed 200 m (Jodkazis 1989).

Subglacial valleys may be an important factor affecting vertical water exchange at the upper part of the sedimentary cover (Marandi *et al.* 2012), going in depth to more than a few hundred meters.

MATERIALS

In this research a large amount of existing data as well as new results about groundwater chemistry were summarized in the largest database on trace elements in groundwater in Latvia.

The database includes records about major ion chemistry (Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, Cl^-, SO_4^{2-}), nitrogen compounds (NO_3^-, NO_2^-, NH_4^+, N_{tot}), trace elements (F, Al, As, B, Ba, Br, Cd, Co, Cr, Cs, Cu, Fe_{tot}, Ge, Hf, Li, Mn, Mo, Ni, Pb, Rb, Sb, Se, Si, Sr, Th, U, V, Zn, Zr), field parameters (pH and temperature), and hydrogeological conditions (e.g., aquifer, aquifer material, sampling depth). However, the amount and type of parameters measured in groundwater depend on research interest and vary between analyses.

Existing data sources used in this study are as follows:

1. Data from national groundwater monitoring programs carried out by the Latvian Environment, Geology and Meteorology Centre in 2008, 2009, and 2013 when trace elements (As, Cu, and Pb) were included in the programme.
2. Data from groundwater extraction wells from 1998 to 2013. Due to legislation requirements, certain harmful trace elements should be measured at least once when groundwater extraction more than 100 m^3 per day is planned.
3. Data from large previous studies on trace element content in groundwater in Latvia (Gosk *et al.* 2006; Levins & Gosk 2007). As a result, more than 700 complete groundwater analyses were made containing large amounts of parameters measured.

All groundwater samples were taken according to LVS ISO 5667-1 standard. Chemical analysis of groundwater samples from national groundwater monitoring programs and groundwater extraction wells were performed in laboratories which are accredited according to the standard LVS EN ISO/IEC 17025 – General requirements for the competence of testing and calibration laboratories. Standard testing methods (Water Monitoring Programme 2015) were used which are in accordance with the procedure laid down in Article 21 of European Water Framework Directive (2000/60/EC) and meet the requirements of the European Commission 'Guidance on Groundwater Monitoring' (European Commission 2007). Measurements of pH were made according to LVS EN ISO 10523 standard.

The new data set

One hundred and seventeen groundwater samples were collected during the time period from year 2010 to year 2012 (Figure 2) to a maximum depth of 1,090 m. The samples preferably were taken from monitoring wells and water supply wells. Samples from springs were taken if none of the previously described sources were available.

Wells were pumped using GRUNDFOS SQ or GRUNDFOS MP1 groundwater pumps. Field parameters electric
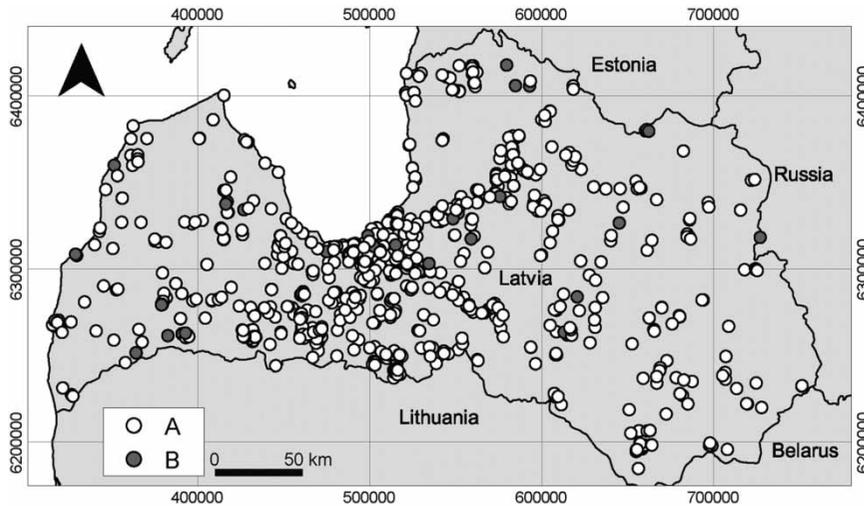
Figure 2 | Groundwater sampling sites. (A) sampling sites from previous studies (existing data); (B) samplings sites within this study (new data).

conductivity (µS/cm), dissolved oxygen, redox potential, and temperature (°C) were measured with WTW Multiline 3420 multimeter. Samples were collected after stabilization of all field parameters. Groundwater samples for major anions analysis were not filtered and were collected in 1-liter high density polyethylene (HDPE) bottles. Samples for trace element and cation analysis were passed through a 0.45 µm pore diameter filter and collected in two 50 mL HDPE bottles previously washed in ultrapure nitric acid and rinsed in distilled water. At the sampling, all bottles were rinsed three times using the groundwater to be sampled. Samples for cations and trace elements were acidified to pH <2. Then, all samples were stored in a field refrigerator below 4 °C until delivered to the laboratory.

METHODS

Analytical methods

Analyses of major cations and trace elements were carried out at the Faculty of Geography and Earth Sciences, University of Latvia, but analysis of major anions was carried out at the Faculty of Chemistry, University of Latvia.

Major cations Na^+ and K^+ were measured by flame emission spectrometry in line with ISO 9964-3:1993. Major cations Ca^{2+} and Mg^{2+} were determined by atomic absorption spectrometry in line with ISO 7980:1986 using

PerkinElmer atomic absorption spectrometer AAnalyst200. Major anions (Cl^-, SO_4^{2-} and alkalinity expressed as HCO_3^-) were determined according to standard analytical procedures (Andrew 2005; Grigorjevs & Kalvāns 2012).

Trace elements Fe_{tot}, Mn, Zn, and Cu were determined by atomic absorption spectrophotometry using a Perkin-Elmer atomic absorption spectrometer AAnalyst200.

Analyses of such trace elements as As, Ba, Br, Cr, Ni, Pb, Rb, Sr, and U were carried out on a benchtop total x-ray fluorescence spectrometer PicoTAX, Roentec (Berlin, Germany).

Synthetic quartz discs 30 ± 0.1 mm in diameter and 4 ± 0.1 mm thick were used as sample carriers. First, the set of sample carriers was pre-cleaned (Klockenkämper 1997). After, the carriers were left to dry in glass beakers covered by Petri dishes for 24 hours or until all carriers were dry. Second, 10 µL silicone SERVA in isopropanol solution was applied at the center of each carrier and placed in Petri dishes to dry for 48 hours. Finally, sample preparation was done by mixing 190 µL of groundwater sample with 10 µL Ga (10 mg/L Ga in 2% HNO_3, PerkinElmer) used as internal standard in 1,500 µL safe lock tubes and homogenized using an agitator. After homogenization, 10 µL of solution was transferred onto the center of a siliconized carrier and dried on a hot plate at 60 °C covered by a Petri plate until a thin film formed. For groundwater samples with very low mineralization the last step was repeated twice (Klockenkämper 1997; Stosnach 2005; Bruker 2007). All setup and calibration procedures were carried out in

line with the manufacturers' instructions (Bruker 2007). The measuring time for each sample was 1,000 seconds. The results were processed with PICOFOX 5.1.7.1.

Data preparation for multivariate statistical analysis

In total, 1,522 groundwater analyses were collected at the beginning. Fifty-seven samples were excluded due to having incomplete records of major ions (Ca^{2+}, Mg^{2+}, Na^+, K^+, Cl^-, SO_4^{2-}, HCO_3^-) chemistry. The ionic balance (Güler *et al.* 2002) was calculated to validate the analysis. Twenty-three samples having an ionic balance error greater that ±10% were rejected from further analysis. Multiple samples from the same locations comprised 16% of the data and were used in the analysis.

The majority of the multivariate statistical analysis assumes the data follow normal distribution (Güler *et al.* 2002). As most of the chemical parameters (except Mg^{2+} and HCO_3^-) were positively skewed, the data were log-transformed to achieve close to normal distribution. Then, standardization was applied on both log-transformed and non-transformed (for Mg^{2+} and HCO_3^-) data so that each variable weights equally (Güler *et al.* 2002; Cloutier *et al.* 2008). In total, a data set of 1,442 samples collected from monitoring springs and wells, springs, project wells, drainage, and water supply wells was used for further multivariate analysis. Groundwater hydrochemical groups were defined using HCA and PCA. The analyses were performed only on the basis of major ion concentrations. Trace elements were not included in multivariate statistical analysis because complete data matrix is required but most of the trace element measurements were made at different times and locations. Inclusion of trace elements in the analysis would significantly reduce the amount of data, thereby data points would not cover the whole territory of Latvia. Trace elements were later analyzed within each cluster.

Data pre-treatment, PCA, and HCA were performed using SPSS Statistics 22. Saturation indices of calcite, dolomite, gypsum, and halite minerals were calculated using software PHREEQC, version 3 (Parkhurst & Appelo 2013). Calculation was based on concentrations of major ions (Ca^{2+}, Mg^{2+}, Na^+, K^+, alkalinity as HCO_3^-, Cl^-, SO_4^{2-}), temperature and pH values. Temperature data were not available for 473 samples and the average groundwater temperature (8.5 °C, standard deviation 1.96 °C) obtained from the rest of the samples was used.

HCA

Cluster analysis is a technique for grouping observations in such a way that each group or cluster is homogeneous with respect to certain characteristics and distinct from other clusters regarding the same characteristics (Davis 2002). It is found that Euclidean distance as a similarity measure and Ward's method as a linkage method give the most efficient results for analysis of the groundwater chemical composition (Güler *et al.* 2002; Cloutier *et al.* 2008; Monjerezi *et al.* 2012; Surinaidu 2016) and is used in this study.

PCA

PCA is a dimension reduction technique (Davis 2002): a set of correlated variables is transformed into a set of uncorrelated principal components (PCs) (Farnham *et al.* 2002). In this study, PCs are obtained through eigenanalysis of the correlation matrix (Farnham *et al.* 2002). The number of components extracted is based on the Kaiser criterion (Kaiser 1958), which suggests that components with eigenvalue greater than 1 are the most appropriate ones for interpretation (Cloutier *et al.* 2008). Varimax rotation was used to increase the participation of the variables with higher contribution and reducing that of the variables with lesser contribution at the same time (Kaiser 1958; Cloutier *et al.* 2008).

RESULTS

PCA

Based on the Kaiser criterion only two PCs can be retained and have eigenvalues greater than 1 (Kaiser 1958). However, after a number of tests, three PCs were extracted explaining 84% of the total variance in the data set (Table 1).

PC1 explains the greatest variance and groups high positive loadings of Na^+, K^+, and Cl^- (Table 1). PC2 is characterized by highly positive loading of HCO_3^-, Ca^{2+},

Table 1 | PC loadings and explained variance for three components with Varimax rotation

Parameter	PC1	PC2	PC3
Ca^{2+}	0.168	**0.650**	**0.667**
Mg^{2+}	0.524	**0.625**	0.428
Na^+	**0.916**	0.104	0.182
K^+	**0.824**	0.139	0.123
HCO_3^-	0.007	**0.933**	−0.139
Cl^-	**0.783**	−0.001	0.381
SO_4^{2-}	0.331	−0.099	**0.878**
Eigenvalue	3.69	1.40	0.79
Explained variance (%)	52.67	20.10	11.30
Cumulative % of variance	52.67	72.72	83.98

Variables with PC loadings greater than 0.6 are considered to be significant and are marked in bold.

and Mg^{2+}. The last, PC3, explains the least amount of variance and contains highly positive loadings of SO_4^{2-} and Ca^{2+}.

PC1 reflects the Na-Cl rich groundwater. This type of water is generally found starting from middle-lower Devonian aquifers to Cambrian aquifer. PC2 is defined as Ca-Mg-HCO$_3$ water type and is the most common groundwater type in the active water exchange zone. PC3 accounts for the gypsum dissolution process and reflects the Ca-SO$_4$ water type characteristic to areas where gypsum is encountered

in upper and lower Devonian aquifers, but also can be found in other parts of the active water exchange zone (Levins et al. 1998).

HCA

The main result of HCA is a dendrogram (Figure 3) grouping groundwater samples based on their geochemical similarities and dissimilarities. At first, the number of clusters was visually selected by moving the Phenon line (Güler et al. 2002; Monjerezi et al. 2012) and then justified by best matching results. By observing the dendrogram four large groups can be easily identified (Figure 3, the dashed Phenon line). However, the median groundwater chemistry (Table A1, available with the online version of this paper) was analyzed in more detail for all eight possibly geochemically distinct clusters (Figure 3, the black Phenon line).

The first group consists of C1 and C3 clusters and appears to reflect bicarbonate waters with somewhat elevated Cl^- and SO_4^{2-} loading (Table A1). The second group of C8, C7, and C2 clusters is Ca-Mg-HCO$_3$ groundwater. The cluster C6, forming the third group, is groundwater with very low total dissolved solids (TDS) values. Finally, the fourth group formed by C5 and C4, is high mineralization Cl^- or SO_4^{2-} dominated water. The linkage distance
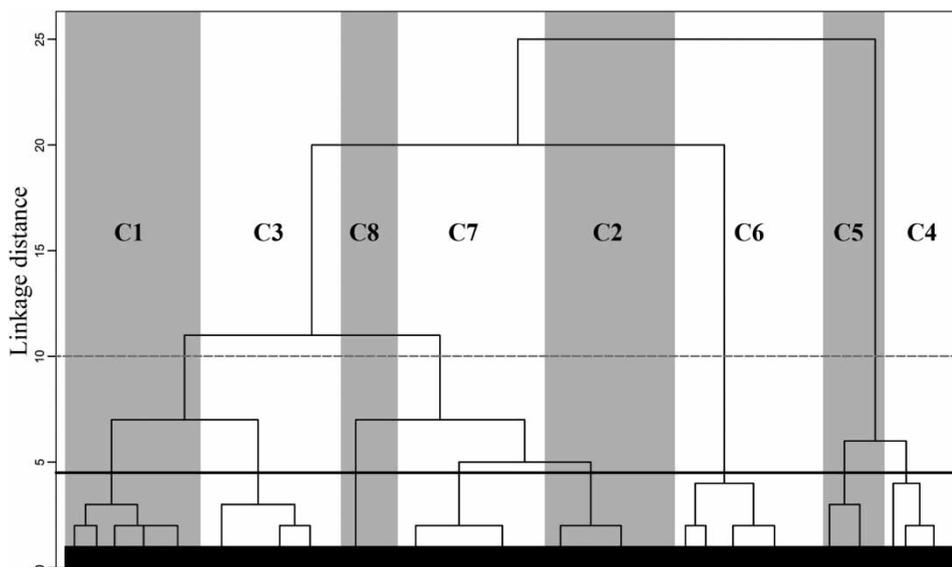

Figure 3 | Dendrogram from HCA showing division of groundwater samples. Dashed line reflects four major divisions and black line reflects the eight distributed clusters for further analysis.

indicates that the samples in the fourth group formed by C5 and C4 are more distinct from other clusters, suggesting they are more geochemically distinct. The largest compositional similarity is between clusters C7 and C2. Thus, the four cluster groups do not directly reflect the PCA results. However, the examination of groundwater chemistry within clusters gave a different insight.

Likewise, the four clusters seem to be geochemically distinct by comparing the median values of major ions used in HCA and calculated TDS (Table A1) (Gunnarsdottir *et al.* 2015). Yet, they do not directly reflect the four larger groups identified by higher level of the Phenon line at HCA (Figure 3). C4 indicates the highest median values of Mg^{2+}, Na^+, K^+, Cl^-, and TDS values are also high. This cluster reflects the Na-Cl rich groundwater with high salinity described by high PC1 values. C5 has the highest median values of Ca^{2+}, SO_4^{2-}, and TDS and is described by PC3. The lowest TDS, Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^- and second lowest Cl^- and SO_4^{2-} values are grouped in cluster C6. C6 belongs to Ca-Mg-HCO_3 water type with very low values

of all the PCs. C8 also describes the Ca-Mg-HCO_3 water type but differs from other clusters by having the lowest Cl^- and SO_4^{2-} median values, both less than 3 mg/L. Clusters 1, 2, 3, and 7 account for HCO_3^- dominant water type as well; however, with no significant differences in major ion distribution compared to each other. All clusters except C4 and C5 describe PC2.

The Piper diagram presents groundwater samples and their belonging to a certain cluster (Figure 4). Clusters C4, C5, and C1 can be easily visually separated in the diamond-shaped field, but others overlap each other significantly. The largest values of TDS typically show Na-Cl type groundwater samples (Levins & Gosk 2007; Cloutier *et al.* 2008); however, the smallest belongs to Ca-Mg-HCO_3 water type groundwater samples. The majority of samples fall into Ca-Mg-HCO_3 water type and belong to C2, C3, C6, C7, and C8. Na-Cl or Cl^- anion dominant groundwater samples are grouped in C4 where freshwater mixing with saltwater can also be observed. Ca-SO_4 water type is represented by the samples from C5. Median values
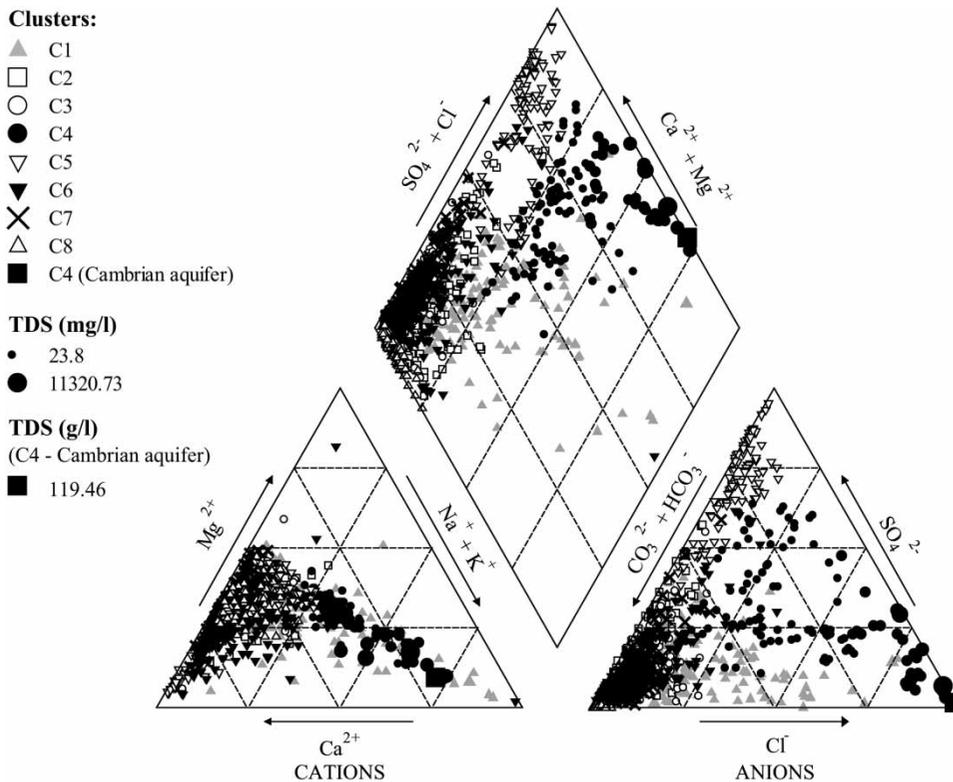
Figure 4 | Piper diagram showing the composition of groundwater samples used in this study labeled according to their clusters. Symbol size is associated with TDS.

(Table A1) and ratios of Cl^- and SO_4^{2-} ions among clusters where Ca-Mg-HCO_3 water type is dominant (C1–C3 and C6–C8) indicate that groundwater samples from C1 have enrichment in Cl^- ion along with no progressive addition of SO_4^{2-} ion (except C8 which has lowest TDS values), thus samples in this cluster may be influenced by anthropogenic sources. The results from PCA together with the results from HCA show that C1, which has elevated Cl^- concentrations, and C5, which reflects Ca-SO_4 water type groundwater, tend to be inversely related (Figure 5). A positive relation between PC1 and PC3 can be observed in C4, which has both high SO_4^{2-} and Cl^- values. Samples having the highest PC1 and lowest PC3 values in C1 are located in recharge areas. A similar trend of sample plotting can also be observed for C6 and C8 (Figure 5). Variance of trace element concentrations within clusters and in the whole data set is summarized in Table A2 (available with the online version of this paper). Further interpretation of the results is given in the Discussion section.

DISCUSSION

Multivariate statistical analysis suggests that the groundwater could be subdivided into eight compositionally distinctive groups. Although division was set out only on the basis of major ions, further analysis of the trace elements, NO_3^- and NH_4^+, showed significant dependence on the eight groups, thus proving appropriate division of them. The results are summarized in Table 2.

Characteristic geochemical parameters and hydrogeological conditions of sampling locations give an insight into the possible origin of each cluster. Thus they provide enough information to find the connections between
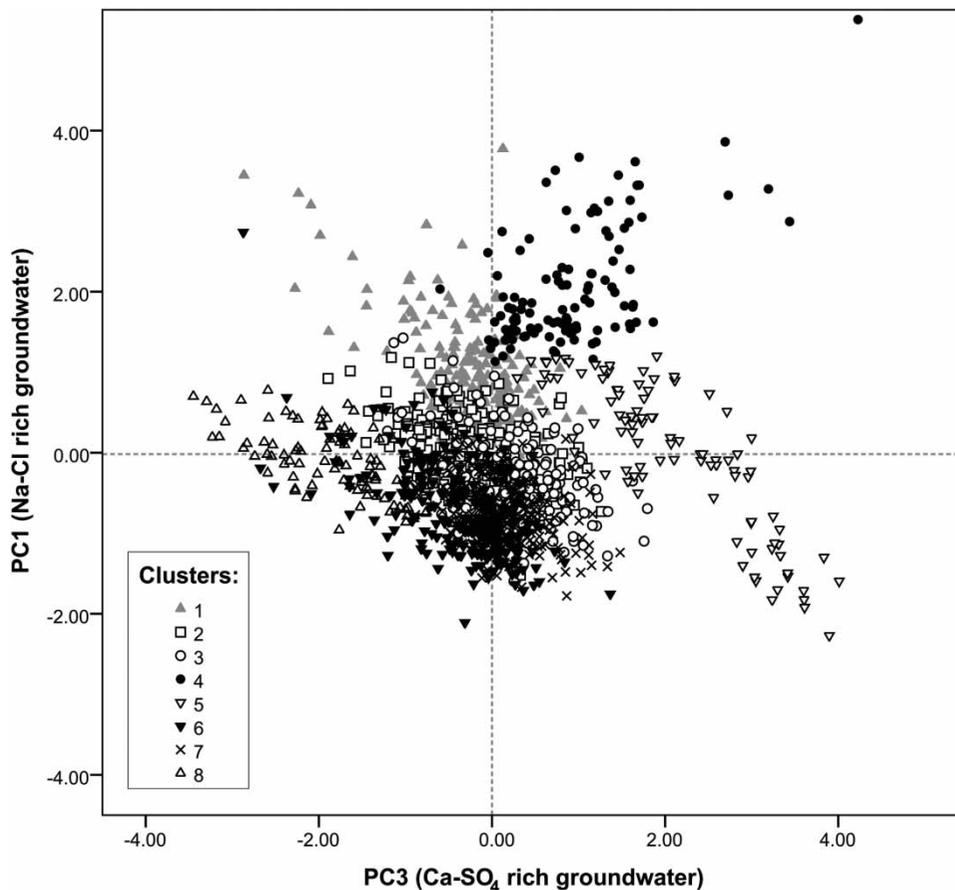
Figure 5 | Plot of loadings from PCA for the first and third PCs. Groundwater samples are grouped according to their clusters from HCA.

Table 2 │ Main geochemical and hydrogeological characteristics of each cluster

| Cluster (sample size)	Dominant sampling source	Positive PC scores	Average depth[a] (m)	Aquifer material	Dominant aquifers	Average TDS[1] (mg/l)	Median groundwater type	Characteristics parameters
C1 ($N = 218$)	MW, WS, SP, PW	PC1 and PC2	4–100	Sandstone, sand, dolomite, till	Q, D_3gj, middle Devonian	520–700	Ca–Mg–HCO_3	–
C2 ($N = 213$)	MW, WS, PW	PC2; PC1 or PC3	4–90	Sandstone, sand, dolomite	Q, D_3gj, middle Devonian	400–500	Ca–Mg–HCO_3	Highest median Al values
C3 ($N = 223$)	PW, SP, DR	PC2 and PC3 or PC2 and PC1	2–7	Till, sand, dolomite	Q, D_3pl-slp	570–750	Ca–Mg–HCO_3	Highest median Cd, Mn, Ni, Pb, U, Zn, NO_3^- values
C4 ($N = 115$)	WS, MW	PC1 and PC3	65–170	Sandstone	Middle and lower Devonian	780–1,520	Ca–Mg–Na–Cl–SO_4	Highest median B, Br, Rb, Sb, Se, V values
C5 ($N = 98$)	MW, WS, SP	PC3	15–100	Sandstone, dolomite, gypsum	D_3pl-slp, D_3gj	800–2,050	Ca–Mg–SO_4	Highest median Cu, F, Li, Sr values
C6 ($N = 242$)	PW, SP, MW	None or PC3	3–15	Sand, sandstone	Q, D_3gj, D_2br	170–300	Ca–Mg–HCO_3	Low median trace element, nitrogen compound, TDS values
C7 ($N = 240$)	SP, PW, WS	PC2 or PC2 and PC3	3–50	Sand, sandstone, dolomite	Q, upper and middle Devonian	410–490	Ca–Mg–HCO_3	–
C8 ($N = 93$)	MW, WS	PC2 or PC2 and PC1	25–75	Sandstone, dolomite	Upper and middle Devonian	420–570	Ca–Mg–HCO_3	Highest median As, Ba, Fe_{tot}, Si, NH_4^+ values (low SO_4^{2-}, Cl^-, NO_3^-)

[a]Expressed as 25th and 75th percentile.

MW, monitoring well; WS, water supply well; SP, spring; PW, project well (Levins & Gosk 2007); DR, drainage.

clusters and even speculate on the potential evolution path of each cluster (Figure 6).

Groundwater from C6 has the lowest TDS and major ion values as well as low concentrations of the majority of trace elements (Table 2; Tables A1 and A2, available with the online version of this paper). Considering the relatively shallow depth and the fact that groundwater is undersaturated with respect to calcite, C6 reflects slightly altered precipitation water. The samples are evenly distributed across all of Latvia. A study in Estonia has shown that the average dissolved solids load at the end of the 20th century was 16 mg/L on average (Treier *et al.* 2004), that is an order of magnitude less than the median value of C6. The absence of positive nor negative PC scores also reflects the very low loadings of dissolved salts

in this cluster. The greatest pH variability within clusters was observed in C6, from 4.6 to 9.0, and only in C6 negative Pearson correlation between Al and pH ($r^2 = -0.567$) appears. Al values noticeably rise when pH values fall under 6. Inversely related Al values to pH are explained by hydrolysis of aluminum silicate minerals by infiltrating water containing carbon dioxide and organic acids (Levins & Gosk 2007) as the mobility of Al dramatically increases in acidic and alkaline water (White 2013). The highest positive Pearson correlation between Al and Si ($r^2 = 0.458$) observed in C6 and dominant aquifer materials such as sand and sandstone complies with the conclusion from previous studies (Levins & Gosk 2007). As a result, C6 can be accepted as initial water for any of the following clusters (Figure 6).
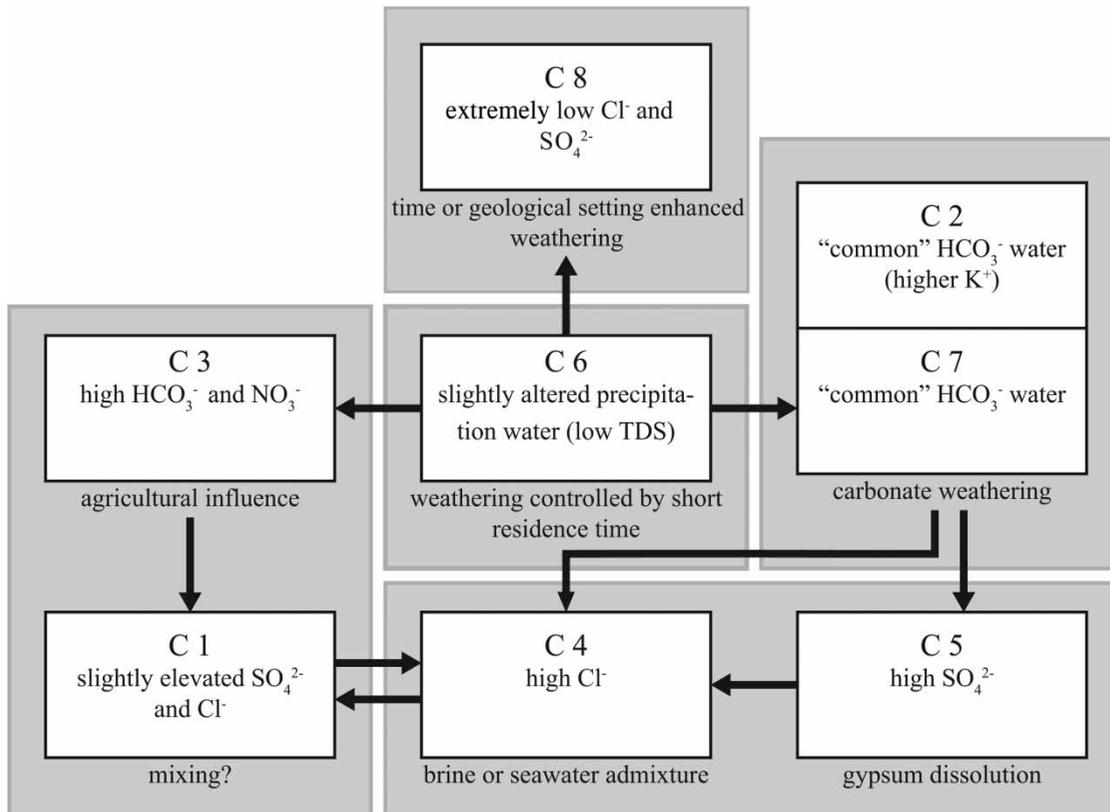
Figure 6 | Evolution of groundwater geochemistry. Grey areas reflect a close linkage distance observed in HCA.

Both groundwater samples in C7 and C2 belong to Ca-Mg-HCO$_3$ water type (Table 2) and are commonly observed in sandy Quaternary and upper and middle Devonian aquifers consisting of sandstone and dolomite. Samples are mainly taken from water supply wells due to overall good groundwater quality (Table A1). The dominance of positive PC2 and the fact that groundwater is mainly saturated with respect to calcite suggests that C7 and C2 is a result of carbonate dissolution (Cloutier *et al.* 2008). C2 differs from C7 by greater sampling depth (Table 2), proportionally higher major ions Na$^+$, K$^+$, and SO$_4^{2-}$ concentrations (Table A1) as well as trace elements Sr, Rb, and B values (Table A2). Strontium concentration increases along the flow path due to incongruent reactions with carbonates and can be used as residence time tracer, however, Sr also can be added from anhydrite or gypsum dissolution (Edmunds & Smedley 2000). Natural B sources in groundwater are water–rock interaction (carbonate rocks and evaporates) (Karro *et al.* 2009), therefore higher values

could also be associated with longer residence time in aquifers. Low major ion and trace element concentrations as well as shallow sampling depths (Table 2) in C6 suggest that C6 groundwater could be the initial water for both C2 and C7 groundwater. Neither this nor previous studies have observed influence of water-bearing rocks on groundwater composition except of gypsum and carbonates. The reason is the widespread carbonate cement for the sand grains in sandstones (Levins & Gosk 2007). However, also the widespread Ca-Mg-HCO$_3$ groundwater type observed in C2, C6, and C7 was not convincingly identified in previous studies (Levins & Gosk 2007). Probably the addition of trace elements and nitrogen compounds in statistical analysis overwhelmed the natural conditions. Also, the previous study (Levins & Gosk 2007) was concentrating on shallow quaternary groundwaters which are typically more affected by pollution than lower aquifers.

Groundwater samples from C8 have extremely low SO$_4^{2-}$ and Cl$^-$ concentrations, both under 3 mg/L. High

NH_4^+ and low NO_3^- values can be assumed as indicators of the reducing conditions in the aquifers. High Ba content can occur due to low SO_4^{2-} concentrations, otherwise Ba should be precipitated as barite (Mokrik *et al.* 2009). The distribution of samples in C8 through the territory of Latvia can be divided into three large groups: (1) fresh groundwater samples from lower Devonian aquifers in the northeast part of Latvia (aquifers contain brackish or saline water in other parts of Latvia); (2) sampling sites near the city of Daugavpils in the southeast part of Latvia where buried paleo valleys are present; and (3) samples from typical carbonate sediments in upper Devonian and Permian aquifers with no gypsum present. C8 is plotting slightly below the one-to-one equivalent line (Figure 7) and indicates that there is an excess of HCO_3^- and/or SO_4^{2-} ions that, in turn, can be interpreted as Ca^{2+} and Mg^{2+} replacement with Na^+ or K^+.

C8 groundwater could be formed from infiltration water during pre-industrial times because of very low initial Cl^- and SO_4^{2-} concentrations which are higher in modern times' waters due to human activity (Edmunds & Smedley 2000). In that case it is suggested that C2 and C7 reflect the modern groundwater and further studies that use groundwater age dating and stable isotopes are encouraged as these methods can help to distinguish pre-industrial water from modern water (Ženišová *et al.* 2015). The highest SO_4/Cl ratios among the clusters and relatively high Sr values support this assumption. Alternatively, C8 can reflect mature groundwater from well-washed rocks, e.g., local circulation systems, where all the easily soluble components such as Cl^- and SO_4^{2-} have been removed from the sediments. The high-Fe low-Mn association can support this opinion as well. Mn(IV) compounds are reduced before the Fe(III) compounds, giving us the chance to speculate that all the Mn has already been washed out of the sediments.

Groundwater samples from C3 reflect water table aquifers or, in some cases, samples taken from melioration
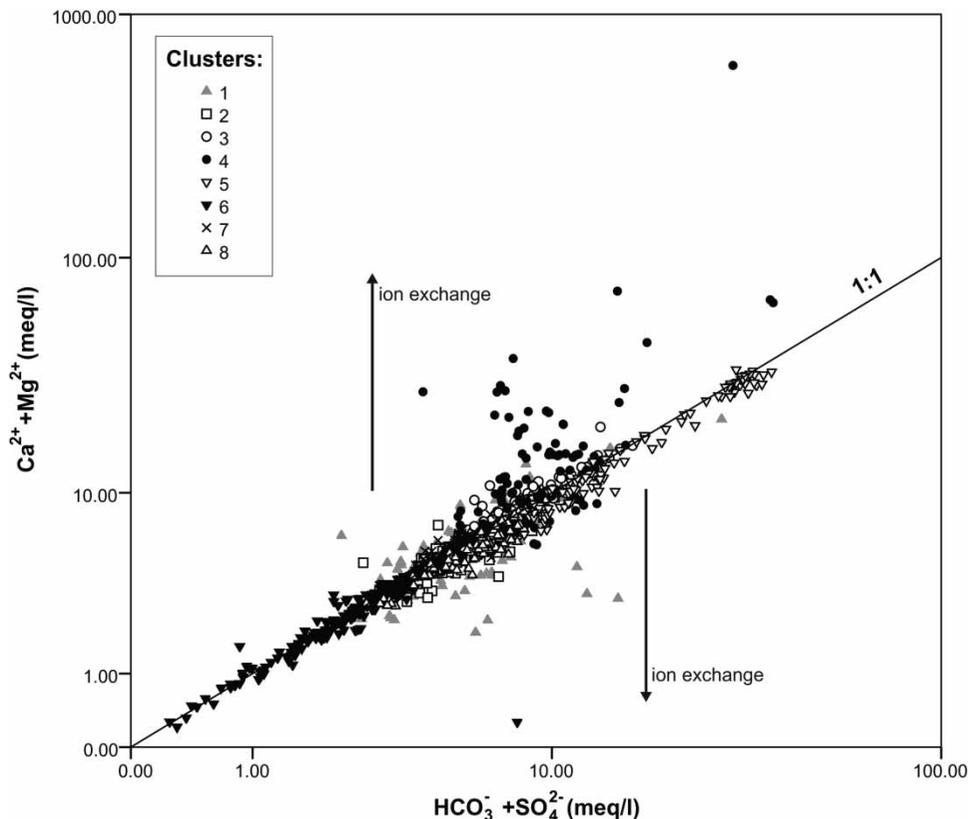

Figure 7 | Relation between $Ca^{2+} + Mg^{2+}$ and HCO_3^- and SO_4^{2-}.

drains. C3 is characterized by highlighted NO_3^- and U values which are mainly associated with agricultural influence (Helena 2000; Levins & Gosk 2007). The high NO_3^- values in shallow groundwater reflect diffuse contamination and are the result of the nitrification process (Valle Junior *et al.* 2014). Diffuse contamination has also been observed in previous studies (Levins & Gosk 2007). It was noticed that U tends to be mobile under oxidizing conditions. Drainage and irrigation processes may cause the increase of the groundwater aeration (Levins & Gosk 2007). Rather high TDS values in shallow groundwater and the fact that groundwater is saturated with respect to both calcite and dolomite suggest that ploughing may promote the dissolution of carbonate and gypsum in the soils (Valle Junior *et al.* 2014). The possible evolution for C3 is directly from C6 (Figure 6).

Samples from C4 reflect two main origins: (1) groundwater with high salinity from passive or stagnant water exchange zones from greater depth and (2) groundwater highly affected by seawater intrusion in the Riga and Liepaja region from upper and middle Devonian aquifers. Groundwater from both origins is saturated with respect to calcite and dolomite, however, only brines are also saturated with respect to gypsum and close to saturation index for halite. Both Cl^- and SO_4^{2-} concentrations are high which can be also observed from PCA results (Table 1). The PCA results also show a positive relation between PC1 and PC3 in C4 (Figure 5). The highest values of many trace elements observed in C4 (Table A2) are characteristic for waters with high salinity (Faye *et al.* 2005; Cloutier *et al.* 2008). The main processes controlling the chemistry of C4 are gypsum dissolution, groundwater mixing and ion exchange between Ca^{2+} and Na^+. It can be observed that due to ion exchange the Ca^{2+} amount in groundwater increases (Figure 7). Na-Cl rich groundwater in the study area had already been observed in previous studies (Levins & Gosk 2007), however with no signs for high SO_4^{2-} concentrations.

The Ca-SO$_4$ water type groundwater is described by C5. The dominant geochemical process is gypsum dissolution and can be justified by saturation with respect to calcite and gypsum and highest PC3 scores. All samples are located in areas with gypsum present in sediments. The characteristic trace elements (Table A2) for C5 are known to be incorporated in carbonates (Faye *et al.* 2005; Klimas &

Mališauskas 2008) or evaporites as secondary minerals, for example, Celestine (Klimas & Mališauskas 2008). Celestine is commonly found in association with gypsum in Latvia (Lukševičs *et al.* 2012). Very low Ba concentrations occur mainly because of barite precipitation (Monjerezi *et al.* 2012). Equally, the high presence of F in evaporites does not produce extremely high fluorine concentrations due to high Ca^{2+} presence in groundwater and fluorite precipitation (Karro & Uppin 2013). Ca-SO$_4$ water type groundwater typically evolves from Ca-Mg-HCO$_3$ groundwater, therefore the evolution path is from less mineralized bicarbonate waters from C2 or C7 (Figure 6).

C1 reflects the most diverse geochemical processes. Part of the samples belong to Ca-Mg-HCO$_3$ groundwater from confined aquifers with slightly to high elevated Cl^- ion concentrations and noticeable ion exchange process. Some of the sampling sites are located in the Riga and Liepaja regions where saltwater intrusion occurs, thus reflecting a connection between C1 and C4 (Figure 6). Placement of some samples close to Na-HCO$_3$ water type (Figure 4), together with sample plotting under the one-to-one equivalent line where Ca^{2+} deficiency can be observed (Figure 7), suggest that possible aquifer freshening is present. Few samples from C1 show very high Na^+ and Cl^- values and Na/Cl ratio close to 1. Those samples probably are the result of anthropogenic influence and halite dissolution delivered by roads de-icing (Cloutier *et al.* 2008). As a result, C1 shows two main origins: (1) anthropogenic influence and (2) water mixing.

CONCLUSIONS

For the first time a comprehensive data set of 1,442 groundwater samples from Latvia containing trace element analysis together with major ion chemistry was made and analyzed. Multivariate statistical methods – PCA and HCA – were used to identify distinct groundwater groups based on major ion chemistry. The distribution of trace element concentrations in each group was examined and factors controlling chemical composition evaluated.

Eight geochemically distinct groundwater groups (C1–C8) are observed characterized by particularly elevated or depressed major ion, trace elements and NO_3^- and NH_4^+

concentrations. The evolution of the groundwater composition is traced from recharge water not yet equilibrated with most of the sediment forming minerals (C6) to typical bicarbonate groundwater resulting from calcite and dolomite weathering (C7) that can have elevated K^+ concentrations (C2) or influenced by diffuse agricultural contamination in the water table aquifers (C3). A particular group of the bicarbonate groundwater has low Cl^- and SO_4^{2-} concentration (C8) which is interpreted as pre-industrial time water. The seemingly continuous bicarbonate waters are subdivided to form five distinct clusters. In addition, we found groundwater groups influenced by gypsum dissolution (C5), saltwater (sea water or brine) admixture (C4), and poorly defined mixtures of water types belonging to other groups (C1).

It is found that although trace elements and nitrogen compounds were not included in the multivariate statistical analysis, their variance in groundwater is remarkably aligned to the groups identified using the major ion chemistry.

ACKNOWLEDGEMENTS

REFERENCES

Andrew, D. E. 2005 *Standard Methods for the Examination of Water and Wastewater*, 21st edn (A. D. Eaton, L. S. Clesceri, E. W. Rice & A. E. Greenberg, eds). American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, USA.

Appelo, C. A. J. & Postma, D. 2005 *Geochemistry, Groundwater and Pollution*, 2nd edn. A.A.Balkema, Rotterdam, The Netherlands.

Brangulis, A. J. & Kaņevs, S. 2002 Latvijas tektonika [Tectonics of Latvia] State Geological Survey, Riga (in Latvian).

Bruker AXS Microanalysis GmbH 2007 *S2 PICOFOX User Manual*. Bruker AXS Microanalysis GmbH, Berlin, Germany.

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.

Davis, J. 2002 *Statistics and Data Analysis in Geology*, 3rd edn. Wiley, New York, USA.

Edmunds, W. M. & Smedley, P. L. 2000 Residence time indicators in groundwater: the East Midlands Triassic sandstone aquifer. *Appl. Geochem.* **15**, 737–752.

European Commission. 2007 Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 15, Guidance on Groundwater Monitoring. http://ec.europa.eu (accessed 8 November 2015).

Farnham, I. M., Singh, A. K., Stetzenbach, K. J. & Johannesson, K. H. 2002 Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemometr. Intell. Lab. Syst.* **60** (1–2), 265–281.

Faye, S., Maloszewski, P., Stichler, W., Trimborn, P., Faye, S. C. & Gaye, C. B. 2005 Groundwater salinization in the Saloum (Senegal) delta aquifer: minor elements and isotopic indicators. *Sci. Total Environ.* **343** (1–3), 243–259.

Gosk, E., Levins, I. & Jørgensen, L. F. 2006 Agricultural influence on groundwater in Latvia. Danmarks og Grønlands Geologiske Undersøgelse Rapport 2006/85, p. 98.

Grigorjevs, O. & Kalvāns, A. 2012 The sensibility analysis of Cl^- and SO_4^{2-} titration in groundwater samples. *Proceedings of the 70th Scientific Conference of the University of Latvia*, section Groundwater in Sedimentary Basins, Riga, Latvia, pp. 47–48.

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.

Gunnarsdottir, M. J., Gardarsson, S. M., Jonsson, G. St., Armannsson, H. & Bartram, J. 2015 Natural background levels for chemicals in Icelandic aquifers. *Hydrol. Res.* **46**, 647–660. doi:10.2166/nh.2014.123.

Helena, B. 2000 Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* **34** (3), 807–816.

Hiiob, M. & Karro, E. 2012 Iron in the Middle Devonian aquifer system and its removal at Võru County water treatment plants, Estonia. *Estonian J. Earth Sci.* **61** (3), 181–190.

Jodkazis, V. 1989 (Йодказис В.) *Региональная гидрогеология Прибалтики* [Regional hydrogeology of the Baltic region]. Vilnius (in Russian).

Kaiser, H. F. 1958 The Varimax criteria for analytical rotation in factor analysis. *Psychometrika* **23**, 187–200.

Kalvāns, A. 2012 A list of the factor controlling groundwater composition in the Baltic Artesian Basin. In: *Highlights of Groundwater Research in the Baltic Artesian Basin* (A. Dēliņa, A. Kalvāns, T. Saks, U. Bethers & V. Vircavs, eds), University of Latvia, Riga, pp. 91–105.

Karro, E. & Uppin, M. 2013 The occurrence and hydrochemistry of fluoride and boron in carbonate aquifer system, central and western Estonia. *Environ. Monit. Assess.* **185** (5), 3735–3748.

Karro, E., Marandi, A., Vaikmäe, R. & Uppin, M. 2009 Chemical peculiarities of the Silurian–Ordovician and Cambrian–Vendian aquifer systems in Estonia: an overview of hydrochemical studies. *Estonian J. Earth Sci.* **58** (4), 342–352.

Klimas, A. & Plankis, M. 2006 Identification of sources of chemical constituents in groundwater from the Vilnius wellfields, Lithuania. *Hydrogeol. J.* **14**, 785–794.

Klimas, A. & Mališauskas, A. 2008 Boron, fluoride, strontium and lithium anomalies in fresh groundwater of Lithuania. *Geologija* **50** (2), 114–124.

Klockenkämper, R. 1997 *Total Reflection X-ray Analysis*. Wiley, New York, USA.

Levins, I. 1990 *Гидрогеохимическая карта Латвии масштаба 1:500000 [Hydrogeochemical map of Latvia, scale 1:500,000].* State Geological Survey, Riga (in Russian).

Levina, N. & Levins, I. 2001 *Liepājas pilsētas centralizētās ūdensapgādes avotu novērtējums [The assessment of sources for centralized water supply in city Liepaja].* State Geological Survey, Riga (in Latvian).

Levins, I. & Gosk, E. 2007 Trace elements in groundwater as indicators of anthropogenic impact. *Environ. Geol.* **55** (2), 285–290.

Levins, I., Levina, N. & Gavena, I. 1998 *Latvijas pazemes ūdeņu resursi [Latvian groundwater resources].* State Geological Survey, Riga (in Latvian).

Lukševičs, E., Stinkulis, Ģ., Mūrnieks, A. & Popovs, K. 2012 Geological evolution of the Baltic Artesian Basin. In: *Highlights of Groundwater Research in the Baltic Artesian Basin* (A. Dēliņa, A. Kalvāns, T. Saks, U. Bethers & V. Vircavs, eds). University of Latvia, Riga, pp. 7–52.

Marandi, A. & Karro, E. 2008 Natural background levels and threshold values of monitored parameters in the Cambrian-Vendian groundwater body, Estonia. *Environ. Geol.* **54** (6), 1217–1225.

Marandi, A., Karro, E., Raidla, V. & Vaikmäe, R. 2012 Conceptual model of groundwater quality for the monitoring and management of the Voronka groundwater body, Estonia. *Estonian J. Earth Sci.* **61** (4), 328–339.

Mokrik, R., Karro, N., Savitskaja, L. & Drevaliene, G. 2009 The origin of barium in the Cambrian–Vendian aquifer system, North Estonia. *Estonian J. Earth Sci.* **58** (3), 193–208.

Monjerezi, M., Vogt, R. D., Gebru, A. G., Saka, J. D. K. & Aagaard, P. 2012 Minor element geochemistry of groundwater from an area with prevailing saline groundwater in Chikhwawa, lower Shire valley (Malawi). *Phys. Chem. Earth.* **50–52**, 52–63.

Mukherjee, A. & Fryar, A. E. 2008 Deeper groundwater chemistry and geochemical modeling of the arsenic affected western Bengal basin, West Bengal, India. *Appl. Geochem.* **23** (4), 863–894.

Parkhurst, D. L. & Appelo, C. A. J. 2013 Description of Input and Examples for PHREEQC Version 3 – a Computer Program for Speciation, Batch-reaction, One-dimensional Transport, and Inverse Geochemical Calculations. http://pubs.usgs.gov/tm/06/a43/pdf/tm6-A43.pdf (accessed 8 November 2015).

Popovs, K., Saks, T. & Jātnieks, J. 2015 A comprehensive approach to 3D geological modeling of sedimentary basins: example of Latvia, central part of Baltic Basin. *Estonian J. Earth Sci.* **64** (2), 173–188.

Raidla, V., Kirsimäe, K., Vaikmäe, R., Jõeleht, A., Karro, E., Marandi, A. & Savitskaja, L. 2009 Geochemical evolution of groundwater in the Cambrian–Vendian aquifer system of the Baltic Basin. *Chem. Geol.* **258** (3–4), 219–231.

Spalvins, A. 1997 Hydrogeological model 'Large Riga'. *J. State Geol. Surv. Latvia* **2**, 44–45.

Spalvins, A., Slangens, J., Janbickis, R., Lace, I., Eglite, I. & Skibelis, V. 2004 Hydrogeological model for well field Otanki of Liepaja, Latvia. *Proc. Riga Tech. Univ. Series 'Computer Science'* **21** (46), 162–171.

Spalvins, A., Slangens, J., Janbickis, R. & Lace, I. 2005 Preventing seawater intrusion into a well field of Liepaja. *Proceedings of the 6th International Conference on Environmental Engineering*, Vilnius, pp. 478–483.

Stosnach, H. 2005 Environmental trace-element analysis using a benchtop total reflection X-ray fluorescence spectrometer. *Analytical Sci.* **21**, 873–876.

Surinaidu, L. 2016 Role of hydrogeochemical process in increasing groundwater salinity in the central Godavari delta. *Hydrol. Res.* **47** (2), 373–389.

Treier, K., Pajuste, K. & Frey, J. 2004 Recent trends in chemical composition of bulk precipitation at Estonian monitoring stations 1994–2001. *Atmos. Environ.* **38** (40), 7009–7019.

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.

Virbulis, J., Bethers, U., Saks, T., Sennikovs, J. & Timuhins, A. 2013 Hydrogeological model of the Baltic Artesian Basin. *Hydrogeol. J.* **21** (4), 845–861.

Water Monitoring Programme 2015 *Pielikums Nr. 22. Virszemes ūdeņu, pazemes ūdeņu un sedimentu kvalitātes monitoringa rādītāju analīžu metodes. Iekš: II. Ūdeņu monitoringa programma [Annex 22. Methods for surface and groundwater quality parameters monitoring. In: Water monitoring program].* Latvian Environment, Geology and Meteorology Centre. http://www.daba.gov.lv (accessed 8 November 2015).

White, W. M. 2013 *Geochemistry*. John Wiley & Sons, Chichester, UK.

Ženišová, Z., Povinec, P. P., Šivo, A., Breier, R., Richtáriková, M., Ďuričková, A. & L'uptáková, A. 2015 Hydrogeochemical and isotopic characterization of groundwater at Žitný Island (SW Slovakia). *Hydrol. Res.* **46** (6), 929–942.