Hydrogeological processes and geological settings over Europe controlling dissolved geogenic and anthropogenic elements in groundwater of relevance to human health and the status of dependent ecosystem

Deliverable 3.3

Proposing a common methodology to calculate the natural concentration of dissolved elements based on lithological/geological water families taking into account possible anthropogenic influences

Authors and affiliation:

Julie Lions, Eline Malcuit, Laurence Gourcy, (BRGM, France); Denitza Voutchkova, Birgitte Hansen, Jörg Schullehner (GEUS, Denmark); Elena Gimenez Forcada, Juan Grima Olmedo (IGME, Spain), Daniel Elster (Geologische Bundesanstalt, Austria); Victor Camps, Georgina Arno, Jignasi Herms, Georgina Arno (ICGC, Spain); Tanja Petrovic Pantic (GZS, Serbia); Nina Rman, Sonja Cerar (GEO-ZS, Slovenia); Katie Tedd, Monika Kabza (GSI, Ireland)

E-mail of lead author:
j.lions@brgm.fr
Version: 29-03-2021

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 731166.

Deliverable Data	
Deliverable number	D.3.3
Dissemination level	Public
Deliverable name	Supporting documents
Work package	WP3: Hydrogeochemistry and health: Mapping groundwater characteristics for the management of aquifers naturally enriched in dissolved elements Preparation and Launch Call
Lead WP/Deliverable beneficiary	D. Elster, GBA

Deliverable status		
Submitted (Author(s))	29.03.2021	- J.Lions (BRGM)
Verified (WP leader)	29.03.2021	- D. Elster (GBA)
Approved (Coordinator)	30.03.2021	- L. Gourcy (BRGM)

SUMMARY

The present document is deliverable D3.3 ” Data set of the results of the statistical data treatment allowing the preparation of the raw elements for the tasks 4 and 5 i.e. concentrations of elements of natural origin per typologies results of the task 3.3 “Proposing a common methodology to calculate the natural concentration of dissolved elements based on lithological/geological water families taking into account possible anthropogenic influences”.

The Water Framework Directive (WFD) requires the EU Member states to assess the chemical status of groundwater bodies, a status defined according to threshold values (TV) for harmful elements, taking into account the Natural Background Level (NBL). NBL corresponds to the expected value of the concentration of elements naturally present in the environment, a value that depends on the lithology of the geological formation in which the water circulates.

The purpose of the study is to propose and test a common methodology among several countries taking into account national specificities and available data linked to groundwater, lithologies and anthropic pressure. The methodology developed aims to work at large scale (river basin or national scale). Based on statistical approach, the methodology is aiming to determine the NBLs according to the lithology and hydrochemical conditions of the aquifers. Information from GW monitoring systems and corresponding hydrogeological setting descriptions is considered in a crossed way. Furthermore, 8 parameters were studied (SO_4, As, Cd, Cr, Cu, Ni, Zn and F) in detail and the methodology was applied in 7 EU countries to optimize the methodology for a wide range of European context.

For each element, average values were calculated for each selected water point and linked to a lithology (10 categories) and geographical area (country - region). Based on the dataset created, the significant differences between lithologies were tested using a nonparametric variance comparison test (Kruskal-Wallis). Then, within each lithology, the difference between anthropogenic influences was also tested in order to assess potential biases. Thus, it was possible to define sets of homogeneous lithologies. For each of these sets, the 90th percentiles were calculated to estimate the maximum possible natural concentrations (NBL).

Results from participating countries and regions show, that it is possible to conclude that prevailing pressures may impact trace elements concentrations in groundwater based on statistical significance among various groups of lithologies and anthropogenic influences. The application of statistical tests to distinguish the influence of prevailing pressure on trace contamination helps to evaluate the dataset confidence for NBLs calculation by discarding data when necessary. Nevertheless, in some cases, e.g. see case study for Denmark, sampling points without anthropogenic pressures are rare and adoptions of the methodology are needed in those cases.

It can be summarized that this work constitutes an important contribution to the definition of NBLs which should be further consolidated by regional studies and the acquisition of specific data dedicated to this purpose.

TABLE OF CONTENTS

1 INTRODUCTION

The objective of the task 3.3 is to propose a common methodology to identify the main geological factors and hydrogeological processes regulating the distribution of natural concentrations (Natural Background Level (NBL) of selected dissolved elements including Potentially Toxic Geogenic Trace Elements possibly affecting human heath

The methods used in the previous project BRIDGE (Wendland et al. 2007) consisted in using only sampling points exempt of anthropogenic influence (upward basin, Groundwaters with low NO_3 concentrations…). However, this approach limits the natural background level determination to some specific areas. Studying the anthropogenic pressure and relating activities to specific dissolved elements would permit to determine the expected NBL in some elements in area under agricultural, industrial (including mining) and urban influence.

The accuracy of this approach would depend on the available information on the anthropogenic activities (databases on typology of activities) and the association between activities and dissolved elements released.

Based on the aquifer/sectors of aquifer typologies determined in task 3.2 and pressure types, statistical data treatment would be used in order to select the most representative water points, eventually to group different typologies and to calculate the basic dataset statistics such as centiles, median, outliers for each element of interest (mainly trace elements) to determine NBL. Some sophisticated non-parametric statistical tests may be usefull, depending on the number of data available.

Finally, based on a common methodology applied in several countries with each specificity, we aim to propose **a common methodology to calculate the natural concentration of dissolved elements based on lithological/geological families.**

2 REVIEW OF EXITING APPROCHES FOR CALCULATING THE NATURAL CONCENTRATION OF DISSOLVED ELEMENTS

2.1 Overarching framework

2.1.1 *Definition of the natural background level*

A natural concentration or natural background level (NBL) is defined as an element acquired entirely from a natural source, whether geological, biological or atmospheric, under conditions that have not been disturbed by anthropogenic activities.

The origin of concentrations is considered entirely natural if neither the contamination itself nor the environmental conditions that render the contaminant mobile are the result of human intervention. The origin is semi-natural when the metals originate from the soil or aquifer matrix (their presence is natural), but are mobilized by a change in conditions as a result of human intervention, e.g. dewatering of the aquifer, an acid accident or a saltwater plume and/or water rich in organic matter.

2.1.2 *Which major, minor and trace elements are concerned*

As some elements are both naturally present in aquatic environments and potentially discharged by human activities, it appears essential to characterise the natural background levels for these elements.

The selection of the elements selected was based on the following criteria:

- Be an element of anthropic and natural origin;
- Have a sufficient number of potential data;
- Have good quality data (ambiguity on Fe and Mn between filtered and unfiltered which would make studying at EU level very complicated);
- Be an element that can cause health problems and / or exploitation of the resource (in connection with task 3.4);
- Do not be ubiquitous (e.g. Cl).

The list of elements selected for this study is therefore : SO_4, As, Cd, Cr, Cu, Ni, Zn and F.

Additional chemical parameters will be required in this study, notably pH, redox potential (Eh NHE), specific conductivity, temperature, dissolved oxygen (O_2), NO_3, Fe. These elements are relevant to determine the pH and redox conditions within groundwaters.

2.2 Objective

The aim is to obtain a homogeneous view of the NBLs over several countries and to provide methodological elements to be able to develop the method in several European countries.

The objective is therefore to try to go further in data processing than what was achieved in the BRIDGE project, by not only taking into account nitrates to identify water points contaminated by anthropogenic pollution and thus be able to keep a larger data set and cover a wider spectrum of aquifers, a larger territory.

2.3 Different methods for calculating the NBLs

Different methods are applied to assess the natural background levels. These methods include the use of historical monitoring, extrapolation from reference sectors with similar geologies, geochemical modelling and the use of statistical tools. These approaches can be applied separately or in a complementary manner in order to assess anthropogenic impacts and quantify natural concentrations.

Various regional, national and international projects have focused on defining methods and concentrations of natural background levels in groundwater. As a follow-up to the European project BaSeLiNe (EVK-CT-1999-00006), the European project BRIDGE (Background Criteria for the Identification of Groundwater Thresholds), carried out from 2005 to 2007, focused on procedures for estimating the chemical status of groundwater and the consideration of natural background levels under the Water Framework Directive (Blum et al. 2006 ; Wendland et al., 2008).

The methods used in BRIDGE (Wendland et al. 2008) consisted in using only sampling points exempt of anthropogenic influence (upward basin, GW with low NO_3 concentrations…). Therefore, this approach limits the natural background level determination to some specific areas.

Studying the anthropogenic pressure and relating activities to specific dissolved elements would permit to determine the expected NBL in some elements in area under agricultural, industrial (including mining) and urban influence.

The accuracy of this approach would depend on the information on the anthropogenic activities (databases on typology of activities) and the association between activities and dissolved elements released. Matrix linking specific activities to determined pollutants exists today in some countries and may be improved and adapted to other EU countries.

2.3.1 *Method used in BRIDGE project*

As a follow-up to the European project BaSeLiNe (EVK-CT-1999-00006), the European project BRIDGE (Background Criteria for the Identification of Groundwater Thresholds), carried out from 2005 to 2007, focused on procedures for estimating the chemical status of groundwater and the consideration of natural background levels under the Water Framework Directive (Blum et al. 2006 and Wendland et al., 2008).

BRIDGE is the acronym for the European FP 6 Project “Background criteria for identification of groundwater thresholds” in which scientists from eleven European countries (including representatives from several national GSOs) jointly defined a harmonized European aquifer typology (Wendland et al., 2008). The main application of the BRIDGE aquifer typology was for differentiation of natural background levels and threshold values of pollutants in groundwaters in Europe. Wendland et al. (2008) posited that petrography should be the prime criterion for developing such groundwater typologies on regional/continental scale, based on the generalization that aquifers with similar petrographic properties have similar composition when the hydrodynamic and hydrologic conditions are similar (Appelo and Postma, 2005).

2.3.2 *Other approaches*

Approach used by England (BGS and Environment Agency) in 2007

Reference: Shand et al., 2007.

The Natural Background Level (NBL) study was carried out over the period 1999 to 2005, synthesising groundwater chemistry data for 26 aquifers located in 23 zones across England and Wales.

The study was based on the collection of available information on regional hydrogeology, lithology, mineralogy and geochemistry of aquifers, precipitation chemistry, land use and groundwater chemistry.

Regional studies present water data through descriptive statistics, whisker plots and cumulative cumulative frequency plots. The descriptive statistics used include the minimum, median, maximum and the 97.7 percentile, the latter as a first approximation of the upper limit of the NBL. When the data set differed in a given study area, for example due to regional lithological or stratigraphic differences or due to differences between confined and unconfined aquifers, the data sets were often separated and descriptive statistics reported for each group (Smedley et al., 2003). Studies take into account data below detection limits that have been replaced by half that value (LQ/2). This is a simple way of dealing with non-detected data but it is a rather arbitrary approach depending on the analytical method and the analytical laboratory. Cumulative probability curves have been used as the preferred method to facilitate data comparisons as they are useful for evaluating different data populations. Spatial distributions were highlighted by mapping (GIS).

An important conclusion of the study is that the variability of groundwater chemistry is significant between and within aquifers. Groundwater chemistry varies depending on a number of factors such as precipitation composition, aquifer lithology, environment, groundwater flow pathways and residence time.

Therefore, there is a dependence depending on the scale studied and this implies considering a range of concentration rather than a single value. Upper limits for NBLs may exceed environmental or drinking water quality standards because they are independent from naturally occurring concentrations in natural systems.

Approach used by the United States (USGS) in 2011

Reference: Ayotte et al., 2011.

Trace element concentrations in groundwater were assessed for samples collected between 1992 and 2003 from aquifers across the United States. The data used in this study were taken from the USGS National Water-Quality Assessment (NAWQA) database in 2006 (Bell and Williamson, 2006). This database contains water quality data, measured field parameters and additional information for all sites sampled under the NAWQA program. Data from 5183 monitoring and drinking water wells representing more than 40 major aquifers in wet and dry regions and various land use contexts were used in the analysis.

This study used data collected by the USGS NAWQA Program to characterize trace-element concentrations in major aquifer groups across the United States. The data were analysed primarily by determining their distributions and comparing those distributions to factors such as climate, composition of geologic materials, well types, well depths, pH, and redox conditions. A variety of methods were used to develop the data used in the analyses and to identify factors influencing trace element concentrations in groundwater.

A large proportion of the trace-element and other water chemistry data were censored at various Laboratory reporting levels (LRLs). The censoring level for each trace element and Rn was adjusted to the respective Long Term Method Detection Limit (LT–MDL; a value that is usually one-half the LRL), thus providing a lower reporting level for each element (Childress et al., 1999). For this report, data censored to the LT–MDL were used for distributional data summaries.

Summary statistics, which include the minimum, maximum, and common percentiles, were computed for all trace elements by using regression-on-order statistics (ROS), statistics appropriate for data with a large proportion (up to 80 percent) of censored values (Helsel and Hirsch, 1992; Helsel, 2005). Data with concentrations above the limit of quantification (LOQ or LRL) are not affected by these corrections.

The study took into account the geological composition of the aquifers and the geochemistry, which are among the main factors affecting trace element occurrence.

Groundwater trace element concentrations were characterized in aquifers of eight major groups on the basis of 8 Major Aquifer Groups:

- Unconsolidated Sand and Gravel Aquifers (USG);
- Glacial Unconsolidated Sand and Gravel Aquifers (GLA);
- Semi consolidated Sand Aquifers (SCS);
- Sandstone Aquifers (SAN);
- Sandstone and Carbonate-Rock Aquifers (SCR);
- Carbonate-Rock Aquifers (CAR);
- Basaltic- and other Volcanic-Rock Aquifers (BAV);
- Crystalline-Rock Aquifers (CRL).

Non parametric statistics, which do not require distributional assumptions about the censored data, were used to describe distributions or compare data among groups. The Wilcoxon and Kruskal-Wallis tests were used to determine whether the distributions of the data from two or more groups were significantly different (Helsel and Hirsch, 1992; SAS Institute, 1999; Helsel, 2005). If multiple-comparison tests indicated significant differences, the Tukey test was used on the ranks of the data to identify which group means were significantly different (Helsel and Hirsch, 1992; SAS Institute, 1999; Helsel, 2005).

Logistic regression was used to help identify processes and factors affecting trace-element concentrations in groundwater for each aquifer group.

All tests and statistical models were calculated using SAS software (SAS Institute, 1999). The statistical records were calculated by robust ROS methods (Helsel, 2005) using a SAS macro for such a process (Practical Stats, 2007).

3 PROPOSED HOVER METHOD

3.1 Flowchart of the proposed method

The objective of the flowchart presented below is to explain the main steps of the proposed method and thus understand how we go from the available analytical results at the sampling points for a given date to a spatial approach taking into account the lithological families / geological and anthropic pressures.

The different steps are described in the following chapters.

Groundwater quality database

Step 1: Inventory of available data
Step 2: Selection of the reference period
Step 3: Extraction or synthesis of GW quality data
Step 4: Formatting the GW quality database
Step 5: Calculating for each sampling point an average value (median) for each parameter
Step 6: Determining Hover-redox and Hover-pH class for each sampling point

Hydrogeological characteristics of sampling points

Step 7: Synthesis of all technical information for the GW sampling points (name, location, detph, lithologies, age/stratigraphy, …)
Step 8: Attachment of each sampling point to a lithological/geological family
Step 9: Attachment of each sampling point to an underground watershed or, failing that to a surface watershed / GW basin (GIS approach)

Anthropogenic pressures assessing

Step 10: Pressure inventory (urban, industrial, agricultural, mining) and relevant information
Step 11: Creation of the pressures database
Step 12: Work at the watershed / GW basin scale to determine anthropogenic pressures and prevailing pressure for each sampling point (GIS approach)

Statistical treatments

Step 13: Synthesis of all data and descriptive statistics
Step 14: Discriminant Function Analysis (DFA) to test difference among GW families for further statistical treatments
Step 15: Non-parametric statistical treatments to determine whether sampling points under the influence of anthropogenic pressure have different distribution than other GW points

NBLs calculation

Step 16: Per element, remove sampling points for which the element concentration is significantly influenced by a prevailing pressure.
Step 17: **NBLs determination per GW family: 90th percentile**

3.2 Constitution of the groundwater quality dataset

3.2.1 *Step 1: Inventory data and relevant information*

The first step of the analysis is the **inventory of available and exploitable** data for the study, as well as the choice of relevant parameters for the characterization of the natural background levels.

3.2.2 *Step 2: Relevant period*

Analytical methods have evolved significantly in recent years, and the quality of analytical results has improved significantly, such as analytical uncertainties, limits of detection and/or limits of quantification ; therefore it is recommended that the last 5 or 10 years be used as the reference period for establishing the "groundwater quality" dataset.

The reference period chosen for the natural background levels study is therefore 2010-2020; this period may be reduced to 2015-2020, if the quality of the analytical results, limits of detection (LOD) and/or limits of quantification (LOQ) are considered too high for certain elements, especially trace elements.

Nota Bene: some countries do not necessarily have a lot of data, some sampling campaigns for example; in this case, depending on the number of data available and their quality, the reference period can be adjusted.

3.2.3 *Step 3 :Extraction or synthesis of quality data for major, minor and trace elements from the national or regional quality databases for the reference period*

This step consists in extracting from the national database or regional databases (depending on each country), all the available analytical results, qualification of the analysis (if the information exist), date of sampling, analytical result, unit of the parameter, analytical method, limit of detection (or quantification), etc...

The list of elements selected for this study are SO_4, As, Cd, Cr, Cu, Ni, Zn.

Additional chemical parameters are requested in this study to determine the hydrochemical conditons in grounbdwaters:pH, redox potential (Eh NHE), specific conductivity, temperature, dissolved Oxygen (O_2), NO_3, Fe.

The elements needed to calculate redox potential with a redox classification are : O_2, NO_3, Mn, SO_4, Fe.

Groundwater pH is an important geochemical parameter that control mineral solubility and mobility of trace metals, often through sorption/desorption processes.

Warning: do not take into account mineral and thermal waters in this study.

3.2.4 *Step 4: Formatting the groundwater quality database*

Eliminate input errors, including unit errors

Experience has shown that in databases there are often errors, such as errors in units (mg/L instead of µg/L for example), duplicates, etc.

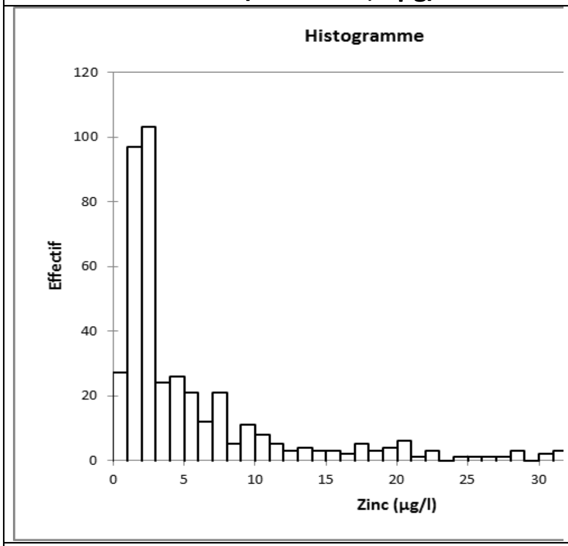
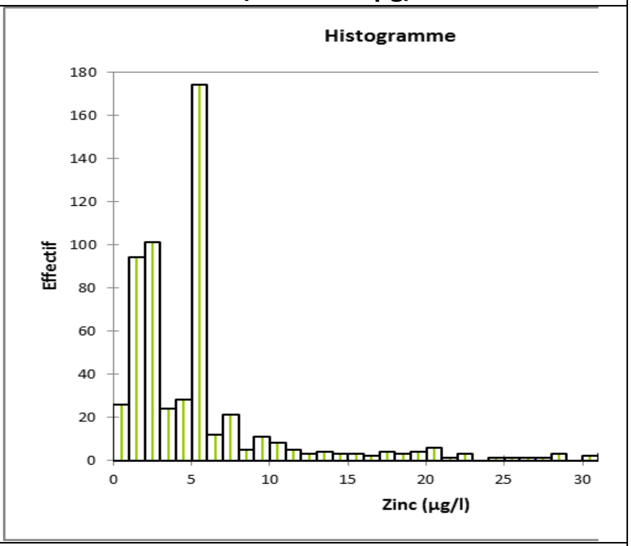
The first task is therefore to check the units and the corresponding values, but also to remove all duplicates of analyses that could have been banked; this is to ensure that each result is representative and not to assign additional weight to an analytical result.

Consideration of analytical limits, limit of detection (LOD) and/or limit of quantification (LOQ), especially for trace elements

The values of the LOD and /or LOQ should be considered to improve data quality. Some values could be too high. The dataset must then be modified so that for all the trace elements studied the limits (LOD or LOQ) are acceptable for further processing.

The figure illustrates the impact of censored value on the distribution of a dataset (Devau et al. 2017).

Figure 1: Histogramme for zinc with and without LOQ max cut-off (Devau et al. 2017).

<table>
<tr><th colspan="3">LOQ/2 max : 2,5 µg/L</th><th colspan="3">LOQ/2 max : 5 µg/L</th></tr>
<tr><td colspan="3">
</td><td colspan="3">
</td></tr>
<tr><td>Median</td><td>2</td><td>µg/L</td><td>median</td><td>5</td><td>µg/L</td></tr>
<tr><td>10th</td><td>1</td><td>µg/L</td><td>10th</td><td>1</td><td>µg/L</td></tr>
<tr><td>Q1</td><td>1</td><td>µg/L</td><td>Q1</td><td>2</td><td>µg/L</td></tr>
<tr><td>Q3</td><td>5</td><td>µg/L</td><td>Q3</td><td>5</td><td>µg/L</td></tr>
<tr><td>90th</td><td>13</td><td>µg/L</td><td>90th</td><td>18</td><td>µg/L</td></tr>
<tr><td>Number sampling points</td><td>130</td><td></td><td>Number sampling points</td><td>1608</td><td></td></tr>
<tr><td>Number analysis</td><td>736</td><td></td><td>Number analysis</td><td>949</td><td></td></tr>
</table>

Processing of censored data (below the LOQ or LOD)

Environmental data sets are frequently “left-censored”, indicating that some values are less than the limit of detection (LOD) or the limit of quantification (LOQ).

One of the difficulties in processing trace element data is the presence of assays with concentrations below the limit of quantification (<LOQ) , the limit of detection (LOD) or censored data.

There are various sophisticated <LOQ data processing methods that deserve to be evaluated for the large amount of <LOQ data encountered in the treatment of environmental data.

Indeed, for certain data sets concerning the elements traces, more than 50% of the data are below the LOQ.

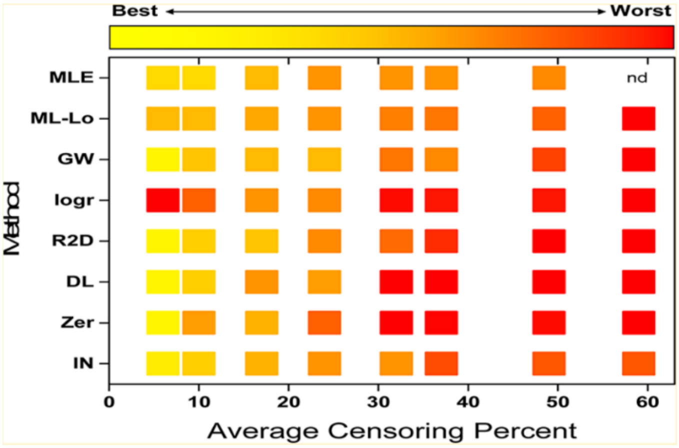
Sophisticated methods require, for example, work on the distribution law of the data (distributional methods), or an extrapolation from >LOQ data to the <LOQ (robust methods) and is therefore less direct than the substitution of the value by a constant (LOQ/2 or LOQ/√2) (Helsel and Hirsch, 2002).

The method of data substitution also depends on the amount of data below the LOQ (or LOD).

According to Antweiler (2015), it was found that technique "quality" was strongly dependent on the degree of censoring present in the groups.

- For low degrees of censoring, the Limit of detection (LOD) or LOD/2 or √2/2*LOD provide the best results
- By 25 % censoring, the Generalized Wilcoxon (GW) technique and substitution of √2/2*LOD gave overall the best results.
- For moderate degrees of censoring, MLE (maximum likelihood estimation) worked best, uncensored data are used to calculate parameters which represent the best fit to the distribution, from which the group comparison probabilities are calculated ; but only if the distribution could be estimated to be normal or log-normal prior to its application; otherwise, GW was a suitable alternative.
- For higher degrees of censoring (each group >40% censoring), no technique provided reliable estimates of the true probability. Group size did not appear to influence the quality of the result, and no technique appeared to become better or worse than other techniques relative to group size.

Figure 2: Statistical approach evaluation to treat <LOD data with reliability (Antweiler, 2015).



As part of the WFD and the assessment of good condition, it was recommended to use the LoQ/2 value.

3.2.5 *Step 5: Defining for each sampling point a representative value (median) for each parameter*

Depending on the sampling points, there is a variable numbers in each data set. The objective is to have only one value per water point so that each water sampling point has the same representativeness. When several analyses are available for the water points, we chose to calculate a median for these data set in order to obtain only one value per water sampling point. The median value is preferred because it is a more robust value than the mean because it is less affected by extreme values.

For each sampling point, it is possible to determine the following characteristics for each parameter (major, minor and trace elements):

- Parameter name
- first date
- last date
- Number of years for the reference period

- Number of analyses per parameter
- Frequency of quantification (number of analyses with data above the LOQ (or LOD depending on the country and/or laboratories) in relation to the total number of analyses)
- Unit
- Min LOD and/or Min LOQ
- Max LOD and/or Max LOQ
- Minimum value,
- Average value,
- Median value,
- Maximum value,
- Standard deviation
- Percentiles, 10, 25, 75 and 90, even 95.

Particular attention must be paid to the number of analyses available before carrying out this first statistical analysis, because the number of analyses per parameter must be sufficient to calculate this first statistical analysis.

Thereafter, the median value for each parameter will be assigned to each waterpoint. The median value is prefered to mean has it s not influenced by extreme values (latest).

3.2.6 *Step 6: Determining HOVER-Redox and HOVER pH classe for each GW point*

Acid/Base water types – pH is chemical widely available. As a first attempt at extending BRIDGE with pH, we propose to test two classifications with different number of groups :

5 groups, combining the classic definition and the drinking water limits (DWL) given in EU Directive 98/83/EC (Reimann & Birke, 2010):

- Acidic, below DWL: pH <6.5
- Acidic: pH $\in$ [6.5, 7)
- Neutral: pH=7
- Basic: pH $\in$ (7, 9.5]
- Basic, above DWL: pH>9.5

3 groups, based on the classic definition we can limit this definition to only 3 classes :

- Acidic: pH<7
- Neutral: pH=7-7.5
- Basic: pH>7.5

The **redox potential** is a measure of the oxidation-reduction state of the groundwater, which in natural systems can be measured by the hierarchical progression of terminal electron acceptors for the reduction of compounds in groundwater.

We can use a redox classification with threshold concentrations to determine the oxidation-reduction state of the groundwater. This classification has been described in the Deliverable 3.2 of HOVER WP3. The classification criteria are described in the following Figure 1 and Table 1.

For this determination, the following data are usefull : O_2, NO_3, Mn, Fe, SO_4

Figure 3: Flowchart of original method and the simplifications introduced in defining the HOVER-redox types for this report (modified from Hansen and Thorling, 2018)

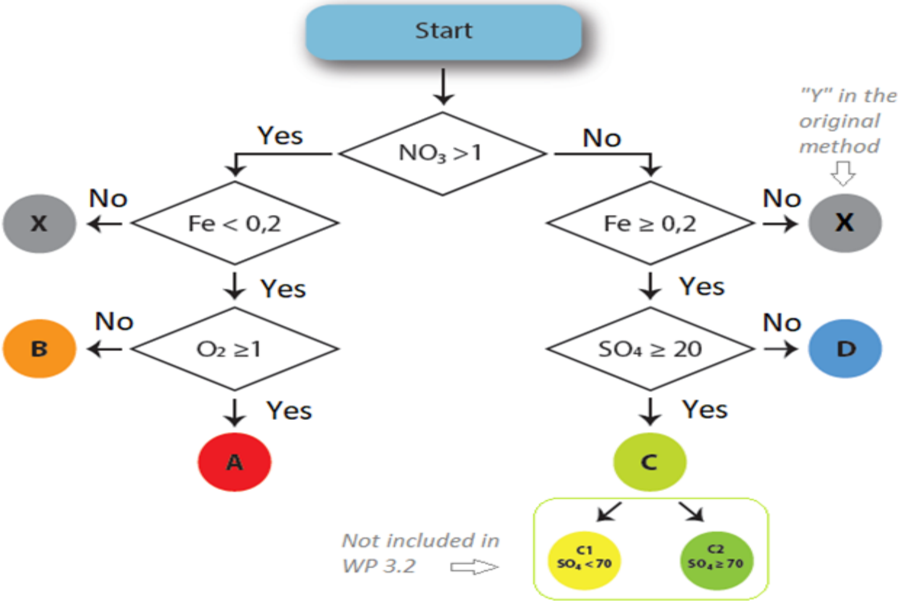


Table 1: Definition of HOVER-Redox types used in this report (modified from Hansen & Thorling, 2018)

Redox type	*Redox condition*	*NO_3* mg/l	*Fe* mg/l	*O_2* mg/l	*SO_4* mg/l
A	Oxic water	>1	<0.2	≥1	-
B	Nitrate-reducing anoxic water	>1	<0.2	<1	-
C	Weakly reduced water	≤1	≥0.2	-	≥20
D	Strongly reduced	≤1	≥0.2	-	<20
X	Unclassified samples				

In this project, we suggest to reduce the number of redox classes to 3:

HOVER-Redox (n=3)
Oxic or Anoxic (A, B redox types)
Weakly or Strongly reduced (C, D redox types)
Mixed (X redox type)

3.3 Hydrogeological characteristics of the sampling points

3.3.1 *Step 7: Synthesis of all technical properties for the GW sampling points*

The objective is to compile available hydrogeological data and information for sampling points that have chemical analyses available.

The data are: name of the sampling point, type of point (spring, well; borehole), location, depth of the borehole/well, age of formation /stratigraphy, BRIDGE and HOVER lithologies, aquifer-type.

3.3.2 *Step 8: Attachment of each sampling point to a lithological/geological family*

It is essential to correctly attach a water point to a lithological entity (aquifer).

The aquifer entities (units) used for the geological/lithological families were defined in Task 3.2 Hover - lithologies and Brigde - lithologies.

Here is a reminder of the different lithologies BRIDGE and HOVER retained in the report of WP3 task

BRIDGE-lithologies

The BRIDGE aquifer typology organized and simplified the complexity of individual aquifers into nine major aquifer rock types with specific ranges in porosity, permeability and petrochemistry (Wendland et al., 2008):

- Sands and gravels;
- Marls and clays;
- Sandstones;
- Chalk;
- Limestones;
- Volcanic rocks;
- Schist and shale;
- Crystalline rocks;
- Saline influence.

Further, based on data from 12 European countries, a European aquifer typology map was compiled (Figure 1).

In this map, to account for particular hydrochemical and hydrological factors, three of the nine typologies were further sub-divided, as follows:

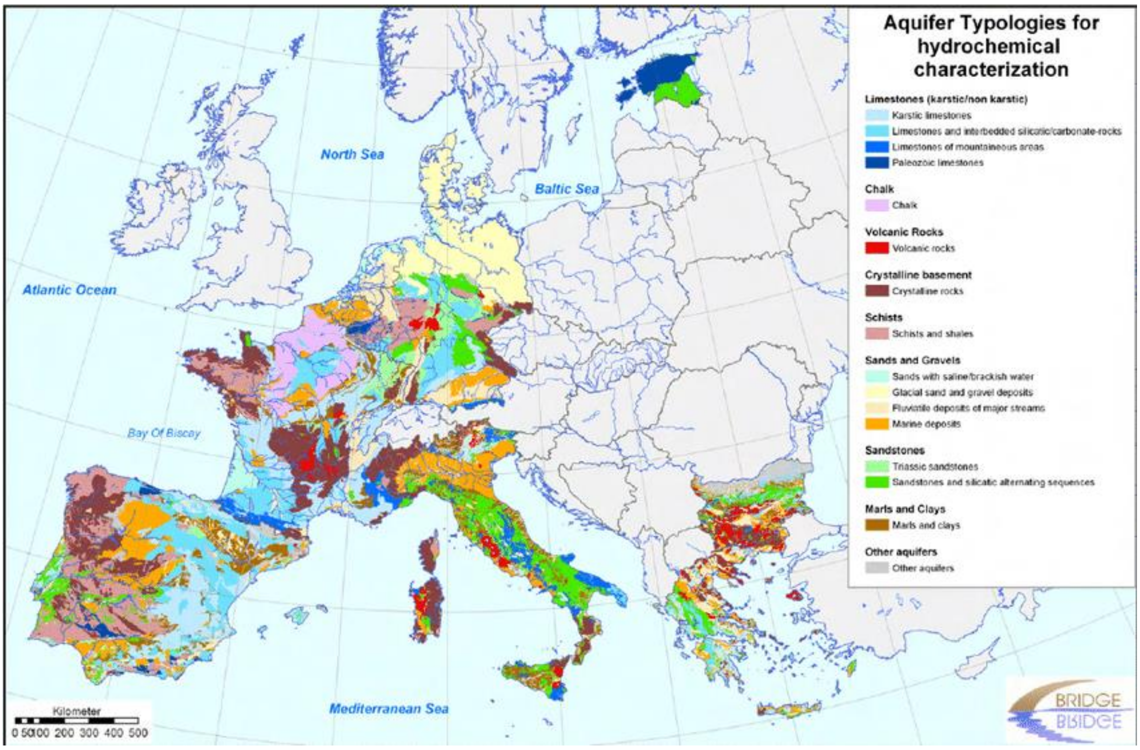
- Limestones:
  - Karstic limestones;
  - Limestones and interbedded silicatic/carbonate rocks;
  - Limestones of mountainous areas;
  - Paleozoic limestones;
- Sandstones:
  - Triassic sandstones;
  - Sandstones and silicatic alternating sequences;
- Sands and gravels:
  - Sands with saline/brakish water;
  - Glacial sand and gravel deposits;
  - Fluviatile deposits of major streams;
  - Marine depostis.

The nine major BRIDGE typologies were subdivided based on the following additional secondary criteria:

- Hydrodynamics, e.g. groundwater recharge, residence time, topography, leakage ;
- Redox conditions ;
- Particular occurrences, e.g. dykes, sulphide minerals, clays ;
- Geological age.

Wendland et al. (2008) developed BRIDGE as a consistent and simple framework for characterizing major groundwater composition patterns; however, they recognized that further refinements can be done to further enhance the accuracy and reliability of the proposed methodology.

Figure 4: European map of aquifer typologies for hydrochemical characterization, compiled as part of the European FP6 project BRIDGE (Wendland et al., 2008).



HOVER-lithologies defined in Task 3.2

In the framework of WP3 task 3.2, lithological/geological water families based on the extended BRIDGE typology were defined.

The key points are presented below, they are based on the BRIDGE-HOVER Lithology water families.

From the tested extended BRIDGE versions, BRIDGE and HOVER-pH seems to be the least suited for determining water families with high concentrations of trace elements, based on the VVM sub-set (As, Mn, Ni, Zn) and our definition of pH classes, which we propose to redefine for the pan-European assessment. BRIDGE and HOVER-pH alone are insufficient to represent the complex hydrogeological and geochemical conditions leading to elevated concentrations of these trace elements. However, a promising future direction for defining geological/lithological classes with high trace elements is to combine the BRIDGE-HOVER Lithology water families with a simplification/modification of the HOVER-pH and HOVER-Redox families (Table 2). It is, however, necessary to test this method proposal with a pan-European dataset, so most of the BRIDGE and HOVER lithology classes are represented (Table 3).

Table 2: Proposed modification/simplification of HOVER-pH and HOVER-Redox classes

HOVER-pH (n=3)	***HOVER-Redox (n=3)***
Acidic (pH <7)	Oxic or Anoxic (A, B redox types)
Neutral (pH ∈ [7, 7.5])	Weakly or Strongly reduced (C, D redox types)
Basic (pH >7.5)	Mixed (X redox type)

Table 3: List of BRIDGE and HOVER lithology classes. The underlined classes are represented in the VVM master dataset from task 3.2.

BRIDGE *(n=17)*	***HOVER Lithology*** *(n=10)*
Karstic limestones	Sedimentary: sand
Limestones and interbedded silicatic/carbonate-rocks	Sedimentary: gravel
Limestones of mountainous areas	Sedimentary: carbonates (limestone, chalk)
Paleozoic limestones	Sedimentary: clays and/or marls
Chalk	Sedimentary: other
Volcanic rocks	Volcanic rocks
Crystalline rocks	Crystalline bedrock
Schists and shales	Metamorphic rocks
Sands with saline/brackish water	Others
Glacial sand and gravel deposits	Unknown
Fluviatile deposits of major streams	
Marine deposits	
Triassic sandstones	
Sandstones and silicatic alternating sequences	
Marls and clays	
Others	
Unknown	

The new geological/lithological families can be defined by merging the BRIDGE & HOVER lithologies (so there is no redundancy) and abbreviating.

Table 4: HOVER lithological family

BRIDGE	HOVER-Lithology
Fluviatile deposits of major streams	Sedimentary: gravel
	Sedimentary: other
	Sedimentary: sand
Marine deposits	Others
	Sedimentary: carbonates (limestone, chalk)
	Sedimentary: clays and/or marls
	Sedimentary: other
	Sedimentary: sand
Others	Metamorphic rocks
	Others
	Sedimentary: other
	Sedimentary: sand

HOVER-lithologies to use for Task 3.3

We propose within the framework of Task 3.3 to take into account the HOVER –lithologies defined in the Task 3.2 enriched with a new class (alluvium) and to add a second criterion of age of the aquifer formation, depending on the geological eras.

HOVER - lithologies and Age formation / stratigraphy classifications are presented in Table 5.

Table 5: HOVER WP 3 - Task 3.3: lithological/geological/stratigraphic families

A5	**A6**
Lithology HOVER	**Age formation / Stratigraphy**
Sedimentary: sand	Quaternary
Sedimentary: gravel	Neogene
Sedimentary: carbonates (limestone, chalk, dolomite)	Paleogene
Sedimentary: clays and/or marls	Cretaceous
Sedimentary: other	Jurassic
Volcanic rocks	Triassic
Crystalline rocks	Permian
Metamorphic rocks	Carboniferous
Alluvium	Devonian
Other	Silurian
	Ordovician
	Cambrian

3.3.3 *Step 9: Attachment of each sampling point to a groundwater catchment area or failing that to a surface watershed (GIS)*

A work under GIS is necessary, the objective is that each water point is attached to a catchment area (watershed).

If possible, each groundwater point should be connected to an underground catchment area, but if this is not possible then the surface watershed will be chosen. No methodology is specified as each water point/ aquifer/geological settings may have different hydrogeological contexts (confined, unconfined, karstic…) or environmental conditions that could affect groundwater quality.

The objective will be to characterize the anthropogenic pressures (step 10) on each catchment area and to define for each water point if it is impacted by an anthropogenic pressure and by which type of pressure (urban, industrial, agricultural and mining).

3.4 Anthropogenic pressures / influences : diffuse or point pollutions

3.4.1 *Step 10: Inventory of data and relevant information (urban, industrial, agricultural, mining)*

To determine if the sampling points and/or concentrations are influenced by a diffuse or point source of anthropogenic pressure, different sources of information are necessary:

- information characterizing diffuse agricultural and urban sources,
- information characterizing anthropic point sources ;
- information characterizing anthropogenic mining activities.

The environment of each water point has to be determined on the basis of existing databases (CORINE Land Cover, database of industrial sites and service activities, register of polluting emissions, etc...) in order to define the potential anthropic pressures.

For this, depending on the available information, several sources of data are usefull.

It is therefore necessary to characterize the contributions of agricultur, industrial activities more specifically mining and urban influence. A database linking specific activities to pollutants may also be used to determine the nature of the parameters that could be modify by a water contamination.

All the data / relevant infloration should be synthesized in the form of database and GIS files.

Diffuse agricultural and urban pressures

Possible source of information:

Corine Land Cover data

Corine Land Cover (CLC) is a European biophysical land cover database. The project is led by the European Environment Agency (EEA) and covers 39 countries. This vector-based dataset includes 44 land cover and land use classes.

Urban influence

Characterize the urban pressure/influence; as well as the potential discharges of urbanization with the available data.

Fertilisation pressures

In order to determine organic and phosphate fertilization pressures, it is possible to use the following approach:

- synthesize information on the typology of farms defined at the municipality level;
- determine the production volumes of organic fertilizers from livestock for all municipalities;
- determine the volumes of phosphate/nitrate fertiliser consumption used by all the farms in a municipality.

Anthropic point sources

This part "characterisation of industrial pressures" is to be adapted according to the knowledge and data available in each country. For example, for France, the available data are data from the French databases BASIAS and BASOL and the facilities classified for environmental protection (ICPE).

The objective of the "characterization of industrial pressures" approach is to define, according to a pressure, a list of concerned elements, i.e. which elements are possibly impacted by an industrial pollution; it may be possible for this work to use for example an activity/pollutant matrix (see step 12).

French databases BASIAS, BASOL and ICPE references:

- BASIAS : « Base de données des Anciens Sites Industriels et Activités de Services »

https://www.georisques.gouv.fr/dossiers/inventaire-historique-des-sites-industriels-et-activites-de-service-basias#/

- BASOL, Base de données Basol, sur les sites et sols pollués ou potentiellement pollués appelant une action des pouvoirs publics, à titre préventif ou curatif

https://basol.developpement-durable.gouv.fr/

- Facilities classified for environmental protection (ICPE)

https://www.georisques.gouv.fr/dossiers/installations

Mining influence

For mining, whether underground or open pit quarries; it concerns the mining of heavy metals, gypsum, others salts, etc., when the exploited formation (lithology, mineralogy) can give information on the rock geochemistry and therefore the hydrogeochemistry.

The mining indices database can be used when available.

3.4.2 *Step 11: Creation of the pressures database and GIS*

Working with GIS :

- Determine for each water point an underground or surface watershed according to the data and the context;
- Synthesize all the anthropogenic pressure data in the form of a database and GIS files, at the scale of the communes and catchment basins for each sampling point for which chemical analyses are available.

3.4.3 *Step 12: Work at the watershed scale to determine anthropogenic pressures for each GW point (GIS) and determination of the prevailing pressure for each GW point*

The objective of this step is to define for each the sampling point, the potential dominant pressure and if chemicals parameters are potentially impacted by the anthropic pressure. This step is necessary to evaluate the bias on chemistry in the case of anthropic pressure on groundwater composition.

Working with GIS :

- For each watershed, determine the importance of each of the pressures (percentage per example) or determine the dominant pressure;
- Work per example with an activities/pollutants matrix to determine, based on pressures/industries, which element(s) may be impacted.

3.5 Statistical treatments

3.5.1 *Step 13: Synthesis of all necessary data and first results in a single database*

Before any statistical treatment of the data, it is necessary to compile all the necessary data.

A specific database must therefore be created and it must contain the following information:

- Name of the sampling point;
- Location;
- HOVER lithology;
- HOVER age formation / stratigraphy;
- HOVER-redox calculated;
- Relevant period / dates of analyses
- pH (median per point) -> determine the HOVER-pH classes ;
- Conductivity (median per point);
- Temperature (median per point);
- SO_4 (median per point);
- As (median per point);
- Cd (median per point);
- Cu (median per point);
- Ni (median per point);
- Zn (median per point);
- Prevaling pressure per point;
- Anthropogenic pressure or natural envrionment.

3.5.2 *Step 14: Discriminant function Analysis (DFA) per lithological/geological family to determine whether or not GW points of the same lithological/geological family, but geographically distant, can be grouped together for further statistical treatments*

This method can be used to separate the NBL zones. For example, when different data from various zones are available, discriminant analysis can find the pattern within the data and classify it effectively.

We propose a preliminary step before the pure statistical treatment, by carrying out an analysis by Discriminant function Analysis (DFA), in order to determine if water points which would have the same hover lithology and the same age formation/stratigraphy, but which would be geographically distant can be grouped in the same lithological / geological / stratigraphic entity.

3.5.3 *Step 15: Non-parametric hierarchical univariate statistical treatments*

Kruskal–Wallis test is a non-parametric method for testing whether samples originate from the same distribution. The parametric equivalent of the Kruskal–Wallis test is the one-way analysis of variance (ANOVA).

Non-parametric hierarchical univariate statistical treatments (Kruskal-Wallis test and Nemenyi post-hoc test) can be carried out to determine whether water points under the influence of anthropogenic pressure have different concentrations of major, minor or trace elements than other water points.

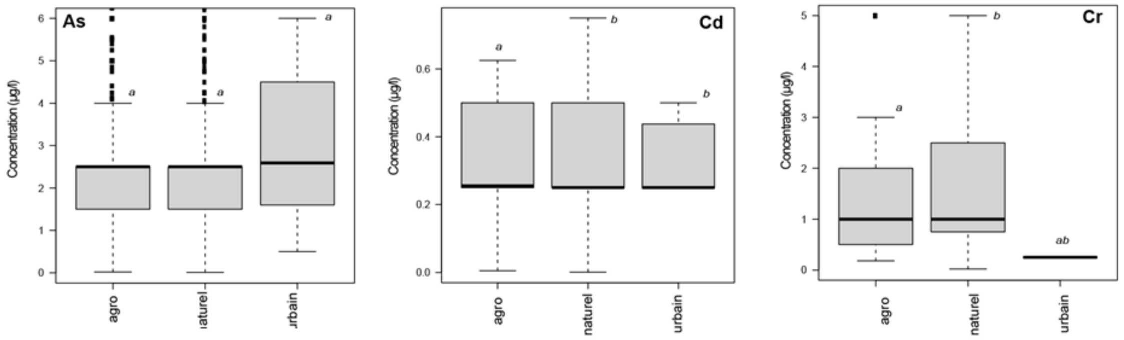
The first step in processing concentration data is to treat the data using a statistical methodology to determine whether the concentrations measured at a water point are influenced by a pressure source of anthropogenic origin. The second step is to characterize the variability of the concentrations measured at a water body as a function of lithological properties.

Non-parametric hierarchical univariate statistical treatments (Kruskal-Wallis test) can be carried out to determine whether water points under the influence of anthropogenic pressure have different concentrations of major, minor or trace elements than other water points.

It is possible to define discrete explanatory variables comprising several categories adapted to each type of anthropogenic pressures; categories defining whether or not a waterpoint is influenced by a point anthropogenic pressure.

Each of these sources of information is used to construct discrete explanatory variables to assign a categorical property to measurements of major, minor and trace element concentrations extracted from quality groundwater database. Univariate statistical treatments are then used to determine whether or not the distribution of concentrations between these categories differed significantly. The statistical treatments are performed independently for each of the discrete variables.

Figure 5: Concentrations of trace elements (As, Cd, Cu) in agricultural, urban and natural areas, defined according to the Corinne Land Cover in the Massif Central area. The bar indicates the median and the squares correspond to measures greater than the third quartile or less than the first quartile. The letters symbolize the results of the Kruskal-Wallis test and then the post-hoc Nemenyi test (p-value <0.05) (extract from Devau et al., 2017).



3.6 Removal of element concentrations influenced by contamination and NBL calculation

3.6.1 *Step 16: Remove sampling points for which the element concentration is significantly influenced by a source of contamination.*

This step consists in eliminating sampling points whose element concentration (SO_4, As, Cd, Cr, Cu, Ni, Zn) is significantly influenced by anthropic activity. This step could be done for each element, by removing the influenced data and retaining the others parameters.

3.6.2 *Step 17: NBLs determination per lithological/geological family: 90 percentile and range of concentrations*

Calculations of natural background levels are performed for the folowing elements: **SO_4, As, Cd, Cr, Cu, Ni, Zn and F** per each lithologcial/geological/stratigraphic family**.**

It is recommended to use the 90 percentile as the threshold value for defining reference concentrations by lithology. The 90 percentile makes it possible to estimate the maximum natural concentrations possible, particularly in particular geological contexts. The NBL is the highest of the values excluding anomalies.

In some specific contexts, the geogenic origins of these natural concentrations should be specified by local studies.

Within each of the lithological/geological/stratigraphic families, groundwater points whose element concentration is influenced by a diffuse or point source of contamination were removed prior to performing the 90 percentile calculation.

This step consists in defining by lithological/geolocial/stratigraphic family:

- the natural background level (90 percentile);
- a range of concentrations.

4 CASE STUDY - IRELAND

4.1 Previous studies on NBLs

The natural background levels (NBLs) for 39 parameters have been established previously in Irish groundwater in March 2017 by Environmental Protection (EPA). The assessment and development of the NBLs report can be access here. The methodology used to establish NBLs for Irish groundwater followed international best practice and made the best use of available data. The natural background levels were established at a national or regional scale and, as such, the results average out spatial variabilities.

The NBLs were established using EPA groundwater quality monitoring data. The location of the EPA's groundwater quality monitoring network's points used to establish NBLs are presented in Figure 6.

To best reflect groundwater chemistry, the monitoring points were initially grouped based on lithology. Details of the lithological groups used were based on the GSI's Groundwater Rock Units (GSI, 2005), which are shown in Table 6.

The natural background levels of parameters such as barium, radon and arsenic, the concentrations of which are unlikely to be affected by anthropogenic activities, were established using data from more than 97% of the EPA's monitoring network. In contrast, natural background levels for parameters such as nitrate, phosphorus and chloride, the concentrations of which are likely to be affected by anthropogenic activities, were established by data from less than 10% of the EPA's monitoring network. This reflects the extensive nature of anthropogenic activities in Ireland and their potential to alter groundwater chemistry.

The most likely hydrogeological/hydrochemical controlling processes for each parameter were considered and the natural background levels were defined accordingly. For example, the natural background levels for (1) parameters that are influenced by lithology (e.g. pH, electrical conductivity, calcium and alkalinity) were defined for simplified lithological groupings (Table 6); (2) parameters that are influenced by the degree of confinement (e.g. nitrate, ammonium, iron and manganese) were defined for unconfined conditions with a note on the likely impact of confined conditions; and (3) parameters that show little spatial variation, sometimes because of a high percentage of samples with concentrations below the limit of detection, were defined nationally.

Natural background levels are a range rather than a single value. In that study, the upper limits of the ranges of NBLs were defined as the 95th percentile and the lower limits of NBLs were defined as the 5th percentile. The NBLs upper limits (and lower limits where appropriate) are presented in Table 7.

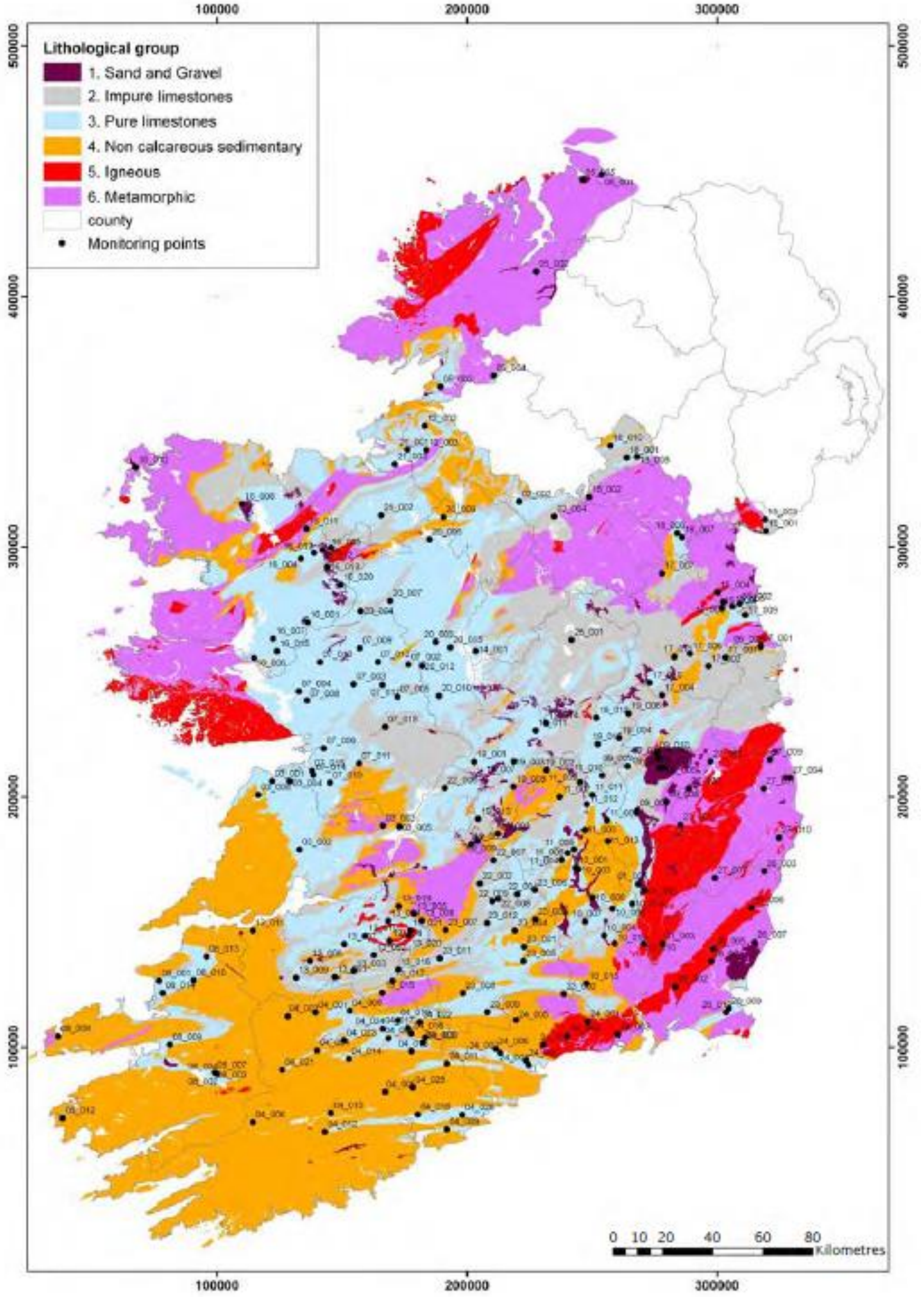


Figure 6: Location of the EPA' groundwater quality monitoring points used to establish NBLs (after EPA, 2017)

Table 6: Number of monitoring points in the lithological groups and detailed lithological groups (after EPA, 2017)

Lithological group	No of MPs	Detailed Lithological Group – GSI rock unit	No of MPs
1. Sand and gravel	30	1.1. Sand and gravel (limestone origin)	18
		1.2. Sand and gravel (other origin)	12
2. Impure limestones	30	2.1. Dinantian (early) sandstones, shales and limestones	3
		2.1. Dinantian lower impure limestones	6
		2.1. Dinantian mixed sandstones, shales and limestones	4
		2.1. Dinantian shales and limestones	1
		2.1. Dinantian upper impure limestones	16
3. Pure limestones	96	3.1. Dinantian pure bedded limestones	65
		3.1. Dinantian pure unbedded limestones	26
		3.2. Dinantian dolomitised limestones	5
4. Non-calcareous sedimentary	32	4.1. Devonian Kiltorcan-type sandstones	9
		4.2. Devonian Old Red sandstones	18
		4.3. Namurian sandstones	2
		4.3. Namurian undifferentiated	1
		4.3. Westphalian sandstones	2
5. Igneous	13	5.1. Basalts and other volcanic rocks	3
		5.2. Granites and other igneous intrusive rocks	2
		5.2. Ordovician volcanics	8
6. Metamorphic	16	6.1. Cambrian metasediments	1
		6.1. Ordovician metasediments	4
		6.1. Precambrian quartzites, gneisses and schists	4
		6.1. Silurian metasediments and volcanics	6
		6.2. Precambrian marble	1
Total No. of MPs	217		217

Table 7: Natural background levels established for parameters defined (a) by a lithological group; (b) by groundwater vulnerability categories; (c) by flow regime categories; (d) nationally for unconfined conditions; and (e) nationally for all pre-selected data (after EPA, 2017)

(a) Lithological groups

Parameter	Unit	1. Sand and gravel	2. Impure limestones	3. Pure limestones	4. Non-calcareous sedimentary	5. Igneous	6. Metamorphic
pH	pH units	*7.5*	*8.3*	7.9	7.1	*7.5*	*7.7*
Electrical conductivity	µS/cm	*583*	*554*	601	392	*60*	*546*
Turbidity	NTU	*1.7*	*3.1*	5.8	1.7	*1.9*	*4.7*
Calcium	mg/l Ca	*111*	*113*	117	64	*8.2*	*114*
Alkalinity	m/l $CaCO_3$	*350*	*290*	322	216	*43*	*310*
Silica	mg/l SiO_2	*13*	*17*	10	14	*19*	*16*
Aluminium	µg/l Al	15	19	26	16	131	19
Barium	µg/l Ba	586	195	179	427	179	114
Chromium	µg/l Cr	11	11	11	9.0	4.3	8.0
Radon	Bq/l	33	28	35	141	268	164
Strontium	µg/l Sr	651	2900	575	320	498	772
Total hardness	mg/l $CaCO_3$	*310*	*294*	328	212	*42*	*303*
Uranium	µg/l U	5.0	3.1	2.8	3.8	1.0	2.6
Zinc	µg/l Zn	*16*	*30*	31	67	*17*	*37*

(b) Groundwater vulnerability

Parameter	Unit		Extreme X	Extreme	High	Moderate	Low
Dissolved oxygen (mg/l O_2)	mg/l O_2	Lower limit	4.4	0.8	3.9	0.7	*0.5*
		Upper limit	12	11	12	8.1	*5.9*

(c) Flow regime

	Unit		Parameter	Inter-granular	Karstified	Poorly productive	Productive fissured
Total phosphorus (µg/l P)	µg/l P		Upper limit	13	26	16	-

(d) Influence of unconfined and confined aquifers

Parameter	Unit	Unconfined	Confined
Sodium	mg/l Na	13	*105*
Magnesium	mg/l Mg	17	*42*
Potassium	mg/l K	2.1	*5.9*
Sulphate	mg/l SO_4	13	*223*
Fluoride	mg/l F	0.2	*2.6*
Nitrate	mg/l NO_3	8.7	*1.5*
Ammonium	mg/l N	0.1	*0.4*
Boron	µg/l B	15	*573*
Iron	µg/l Fe	268	*762*
Manganese	µg/l Mn	23	*176*
Nickel	µg/l Ni	2.1	*4.7*

(e) National parameters

Parameter	Unit	All pre-selected data
Chloride	mg/l Cl	21
Molybdate reactive phosphorus	µg/l P	17
Antimony	µg/l Sb	0.5
Arsenic	µg/l As	2.0
Copper	µg/l Cu	16
Lead	µg/l Pb	2.9
Nitrite	mg/l NO_2	<0.05
Beryllium	µg/l Be	<1
Cadmium	µg/l Cd	<0.4
Cobalt	µg/l Co	<1
Mercury	µg/l Hg	<0.05
Molybdenum	µg/l Mo	<1

Values which are shown in blue italics are included for information only. They have been calculated using data from two or fewer monitoring points and are therefore not defined as natural background levels.

The methodology used to establish natural background levels for Irish groundwater followed international best practice and made the best use of available data. The results are comparable to natural background levels established for other countries or regions. However, the natural background levels were established at a national or regional scale and, as such, the results average out spatial variabilities.

4.2 Study with the proposed method

4.2.1 *4.2.1 Constitution of the groundwater quality dataset*

The inventory data was provided by the Environmental Protection Agency's (EPA's) groundwater quality monitoring programme. The EPA's groundwater monitoring programme includes a standard suite of 40 determinants, including field parameters, nutrients, major ions and certain minor and trace elements are analysed at each monitoring location (238 monitoring points) within the surveillance and operational monitoring networks three to four times a year.

It should be noted that, except for monitoring points in sand and gravel aquifers, the majority of the monitoring point installations are open hole. This means that the groundwater abstracted will be a mixture of deep and shallow (and therefore old and young groundwater). This has implications for the natural background levels calculated.

The groundwater quality data from the EPA's monitoring points were uploaded to a Microsoft Excel database for assessment. The dataset includes all chemical analyses for pH, Eh, conductivity, temperature, SO_4, As, Cd, Cr, Ni, Zn, F, O_2, NO_3, Fe, Cl HCO_3, Ca, Mg, Na, K.

The following steps were taken to improve the reliability of the data:

- removal of duplicate data;
- limits of detection (LOD) were replaced by half the limit detection (LOD/√2) (Table 8);
- removal of data points where the number of analyses for each of the parameter was less than 10.

The relevant period for the groundwater quality dataset is greater than nine years between 2009 and 2019. The groundwater quality dataset contains median concentration for 21 parameters assigned to each of the 238 monitoring points. Additionally, the dataset includes calculated HOVER-pH and HOVER-redox, hydrogeological characteristics and prevailing anthropogenic pressures for each monitoring point.

Table 8: The limits of detection for each element

Element	unit	Limits of detection
SO4	mg/l	1, 3, 10
As	µg/l	0.1, 0.2, 0.37, 0.5, 1
Cd	µg/l	0.06, 0.1, 0.4, 1
Cr	µg/l	0.1, 0.5, 0.7, 1
Cu	mg/l	0.5, 1, 2.7, 3, 10
Ni	mg/l	0.3, 0.5, 0.9, 1
Zn	mg/l	0.5, 1
F	mg/l	0.1, 0.15, 0.3
Cl	mg/l	2.6

4.2.2 *4.2.2 Hydrogeological characteristics of the sampling points*

The HOVER-lithology and HOVER-age formation/stratigraphy for each monitoring point was determined based on the Geological Survey Ireland's (GSI's) Groundwater Rock Units (GSI, 2015).

The vast majority of the monitoring points are located in the HOVER-lithology class 'Sedimentary: carbonates (limestone, chalk)'. To distinguish the pure (more karstified) limestones this class was divided into two classes: Sedimentary: pure carbonates and Sedimentary: impure carbonates (limestone, chalk).

Details of the GSI's Groundwater Rock Units with corresponding HOVER-age, HOVER-lithology and number of monitoring points in each group are shown in Table 9.

Table 9: GSI's Groundwater Rock Units with corresponding HOVER-age, HOVER-lithology and number of monitoring points in each group

GSI's Groundwater Rock Units		**HOVER Age**	**HOVER Lithology**	
Name	***No of MPs***		***Name***	***No of MPs***
Granites & other Igneous Intrusive rocks	9	Silurian	Crystalline rocks	9
Cambrian Metasediments	7	Cambrian	Metamorphic rocks	32
Ordovician Metasediments	9	Ordovician		
Precambrian Quartzites, Gneisses & Schists	10	Cambrian		
Silurian Metasediments and Volcanics	5	Silurian		
Dinantian Pure Bedded Limestone	65	Carboniferous	Sedimentary: pure carbonates	91
Dinantian Pure Unbedded Limestones	22			
Dinantian Dolomitised Limestones	4			
Dinantian (early) Sandstones, Shales and Limestones	2	Carboniferous	Sedimentary: impure carbonates (limestone, chalk)	32
Dinantian Lower Impure Limestones	12			
Dinantian Mixed Sandstones, Shales and Limestones	4			
Dinantian Shales and Limestones	1			
Dinantian Upper Impure Limestones	13			
Devonian Kiltorcan-type Sandstones	10	Devonian	Sedimentary: other	34
Devonian Old Red Sandstones	20			
Namurian Sandstones	2	Namurian		
Namurian Undifferentiated	1			
Westphalian Sandstones	1	Carboniferous		
Westphalian Shales	1			
Sand and Gravel (Limestone origin)	19	Quaternary	Sedimentary: sand	30
Sand and Gravel (other origin)	11			
Basalts & other Volcanic rocks	3	Carboniferous	Volcanic rocks	10
Ordovician Volcanics	7	Ordovican		

4.2.3 *4.2.3 Anthropogenic pressures / influences: diffuse or point pollutions*

The prevailing anthropognenic influences were identified using (a) the groundwater pressures file available on the EPA website provides information on the significant pressures identified for each Groundwater Body (GWB) and (b) the Corine landcover map 2018.

Information provided on the EPA website was used to determine which sampling points are influenced by a diffuse and/or point of anthropogenic pressure. While there is a multitude of pressures in every waterbody, the significant pressures are those pressures that need to be addressed to improve water quality. Many Irish waterbodies have multiple significant pressures. A robust scientific assessment process has been carried out to determine which pressures are significant pressures. This has incorporated over 140 datasets, a suite of modelling tools, and local knowledge from field and enforcement staff from the Local Authorities, Inland Fisheries Ireland and EPA.

Groundwater pressures identified for GWB are:

- Agriculture
- Anthropogenic Pressures
- Domestic Waste Water
- Extractive Industry
- Forestry
- Historically Polluted Sites
- Industry
- Waste

The first step taken to determine anthropogenic pressures for each monitoring point was to intersect GWB and monitoring point's ZOC shapefiles in ArcGIS. The output shapefile gave us information in which GWB each of the monitoring point's ZOC is located. The intersect shapefile was converted to an excel file and it was cross-checked with the GW pressure CVS file.

130 of the monitoring points within the EPA monitoring network have Zone of Contributions (ZOC) or Source Protection Zones (SPZ) delineated for them (available from www.gsi.ie). The Zones of Contribution for the remaining monitoring points were approximated by a 1km buffer.

The proportion of different CORINE land cover categories were identified for each monitoring point's ZOC / SPZ / buffer. Agricultural, Industrial, Mining and Urban pressures were identified from the CLC18 land cover categories as outlined in Table 10.

Table 10: CORINE land cover categories used to identify Agricultural, Industrial, Mining and Urban pressures within the ZOC / SPZ / buffer to the monitoring points within the EPA national groundwater monitoring network.

Pressure type	CLC18 land cover	CLC18 code
Agricultural pressures	Complex cultivation patterns	242
	Fruit trees and berry plantations	222
	Land principally occupied by agriculture, with significant areas of natural vegetation	243
	Non-irrigated arable land	211
	Pastures	231
Industrial pressures	Industrial or commercial units	121
	Road and rail networks and associated land	122
Mining pressure	Dump sites	132
	Mineral extraction sites	131
Urban pressure	Continuous urban fabric	111
	Discontinuous urban fabric	112
Natural setting	All other CORINE categories	

As expected, agricultural pressures were the most prevalent pressure within the catchments to the monitoring points (Table 11). Agricultural land cover categories were the majority land use categories in 89% (n=211) of the catchments. Urban land cover categories were the majority land use categories in 1% (n=3) of the catchments. Industrial and mining land cover categories were not the majority land use categories in any of the catchments.

Only 0.4 % of the catchments (n=1) had only natural land use covers categories. Meaning that 99.6% of the catchments had at least one of the anthropogenic land use cover categories outlined in Table 11.

Table 11: Number of ZOC / SPZ / buffer to the monitoring points within the EPA national groundwater monitoring network containing a majority of anthropogenic CORINE land cover categories used to identify Agricultural, Industrial, Mining and Urban pressures.

Pressure type	Value	n	%
Agricultural pressure	no	27	11
	yes	211	89
Urban pressure	no	235	99
	yes	3	1
Industrial pressure	no	238	100
	yes	0	0
Mining pressure	no	238	100
	yes	0	0

The non-parametric Kruskal-Wallis statistical test was carried out to determine if there was significant differences in the median of the populations from different pressures. Table 12 shows that only Copper and Zinc showed significant differences in the medians between the different data populations. Figure

7 and Figure 8 show a comparison of the concentration distribution for the Agricultural, Urban and Natural prevailing pressure categories using box and whisker plots and probability plots respectively.

Table 12: Results from the non-parametric Kruskal-Wallis statistical test to determine if there were significant differences in the median of the populations from different pressures.

Elements	**All**	**Pressure**						**Kruskal-Wallis rank sum test**
		Agriculture		**Urban**		**Natural**		
	n	n	%	n	%	n	%	
SO4	237	210	89	3	1	24	10	Not significantly different
As	237	210	89	3	1	24	10	Not significantly different
Cd	237	210	89	3	1	24	10	Not significantly different
Cr	237	210	89	3	1	24	10	Not significantly different
Cu	237	210	89	3	1	24	10	Significant (p < 0.05)
Ni	237	210	89	3	1	24	10	Not significantly different
Zn	237	210	89	3	1	24	10	Significant (p < 0.05)
F	237	210	89	3	1	24	10	Not significantly different
Cl	237	210	89	3	1	24	10	Not significantly different

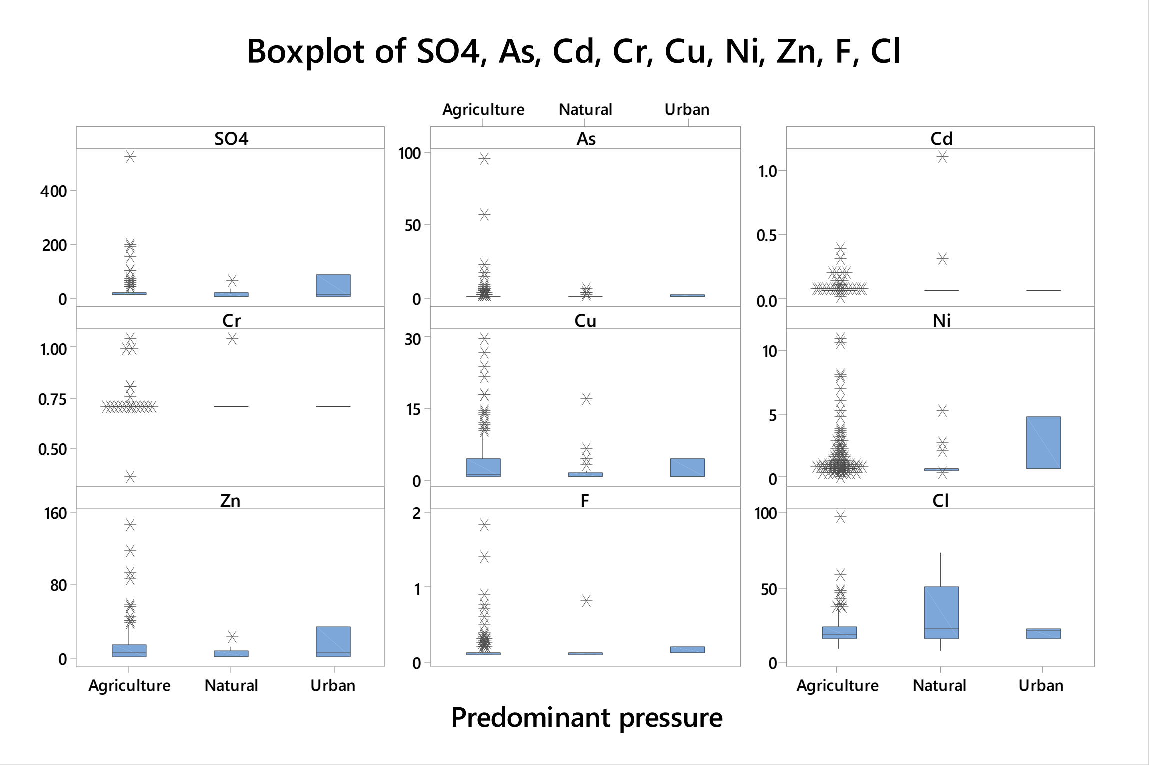


Figure 7: A comparison of the concentration distribution for the Agricultural, Urban and Natural prevailing pressure categories using box and whisker plots. Kruskal-Wallis tests showed that there were only significant differences in the medians of Cu and Zn.

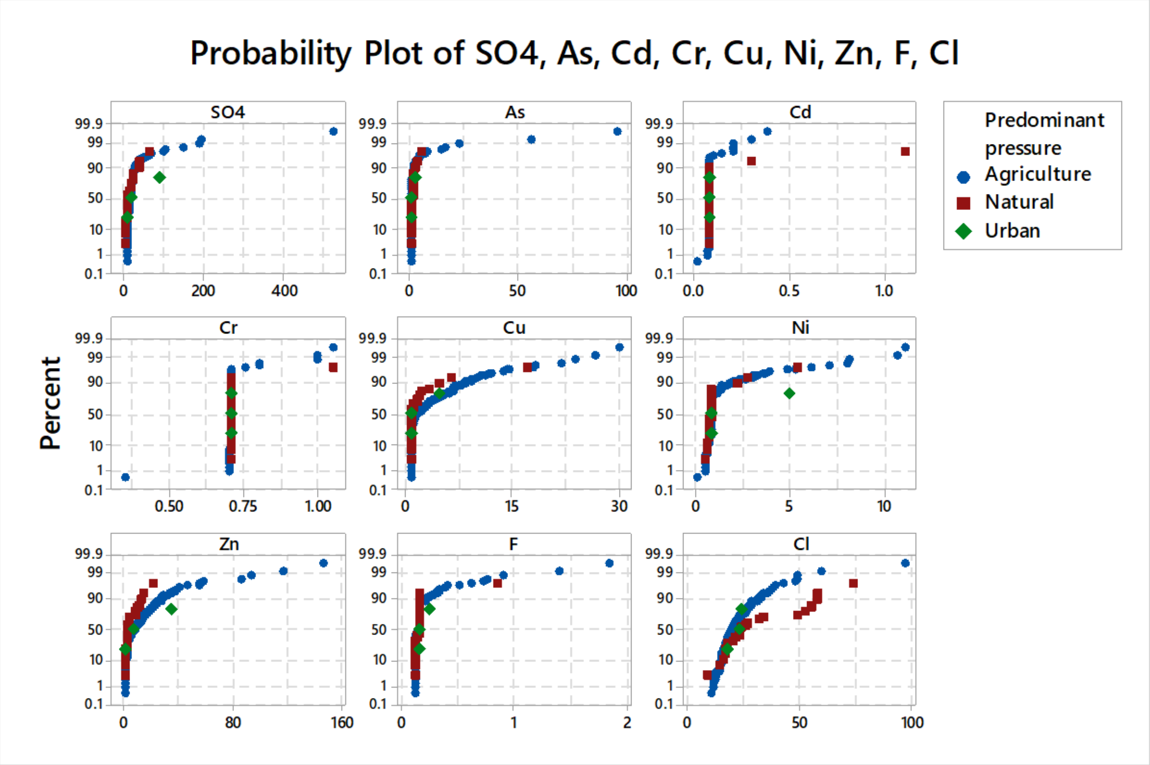


Figure 8: A comparison of the concentration distribution for the Agricultural, Urban and Natural prevailing pressure categories using probability plots. Kruskal-Wallis tests showed that there were only significant differences in the medians of Cu and Zn

Given the results from the Kruskal-Wallis tests were removed the “agricultural” and “urban” groups of data for Cu and Zn. This removed 90% of the datapoints (n=213) and left only 24 datapoints to calculate the NBLs for Cu and Zn. All data points were considered for the other elements.

4.2.4 *4.2.4 Statistical treatments*

Discriminant Factor Analysis (DFA) tests were not carried out as the test requires a normal distribution. Even with transformations carried out via Box-Cox transformation the distributions of the transformed data, with the exception of SO_4, were not normally distributed.

Probability plots (Figure 9 and 10) were used to inspect whether the distributions of lithologies from different parts of the country could be analysed together. The probability plots show single populations with the exception of Silurian Crystalline. Data from the Silurian Crystalline monitoring points show a bimodal distribution. However these MPs reflect only one setting (New Village poorly productive borehole cluster). Therefore not changes were made to the lithological groupings.

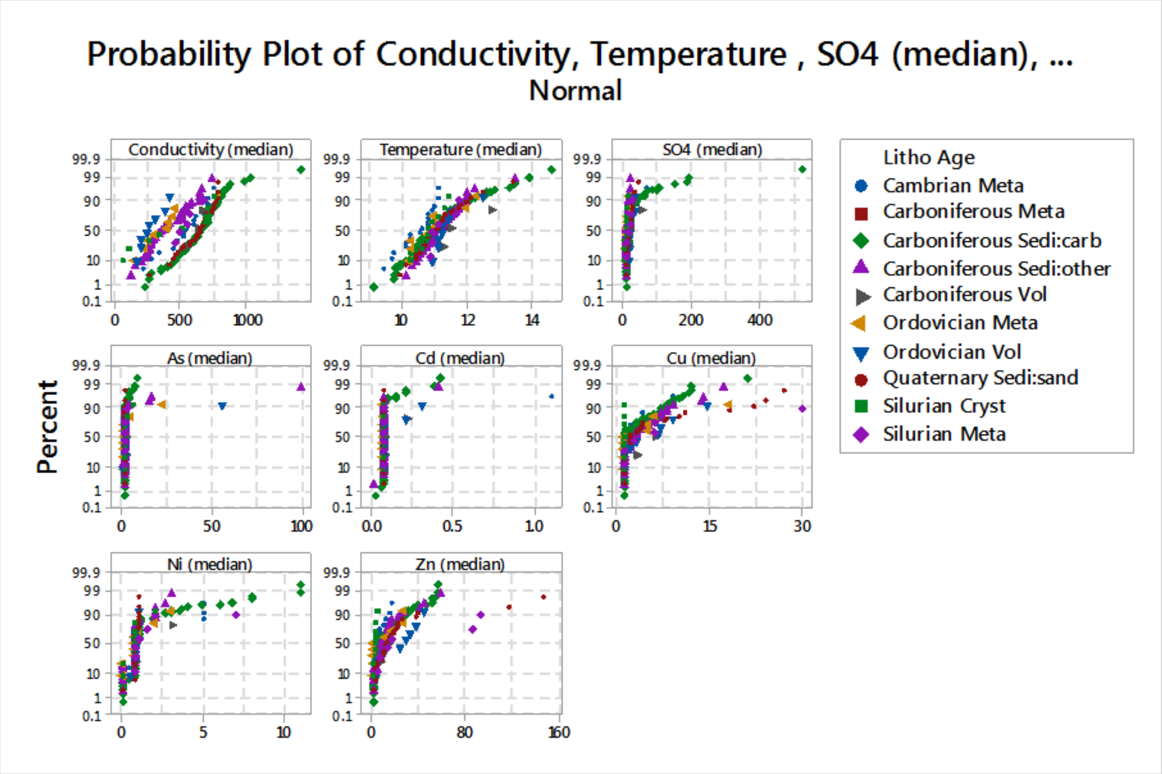


Figure 9: Probability plot of conductivity, temperature, SO4, As, Cd, Cu, Ni and Zn for lithological age.

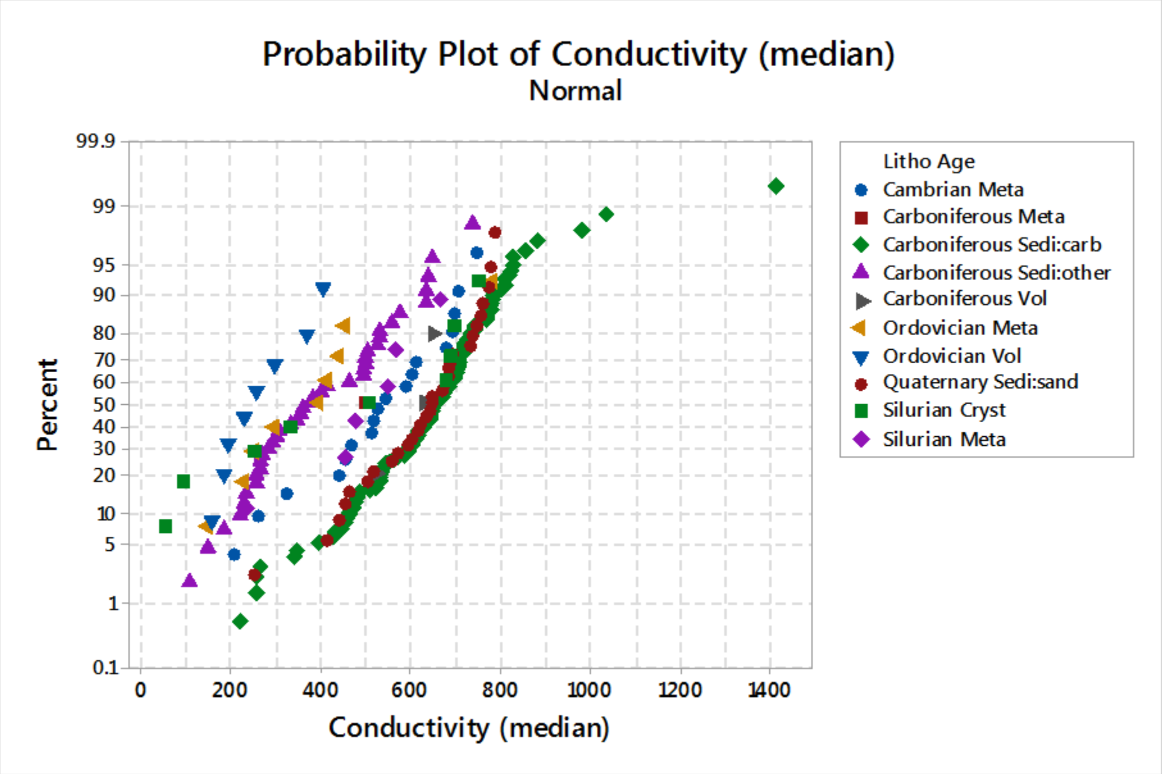


Figure 10: Probability plot of conductivity for lithological age.

4.2.5 *HOVER-pH and HOVER-redox importance*

Kruskal-Wallis tests were carried out to inspect if the groups based on HOVER-pH and HOVER-redox categories were significantly different and to define Water Families on which to base the calculation of NBLs.

Table 13 shows that the Kruskal-Wallis tests show that there are only significant differences between the medians of the different HOVER-pH categories for Cu and Zn. Figure 11 and 12 show a comparison of the concentration distributions for the Acidic, Basic and Neutral HOVER-pH categories using box and whisker plots and probability plots respectively.

Table 13: Results from the non-parametric Kruskal-Wallis statistical test to determine if there were significant differences in the median of the populations from HOVER-pH categories.

Elements	**All**	**HOVER-pH categories**						**Kruskal-Wallis rank sum test**
		Acidic		**Basic**		**Neutral**		
	n	n	%	n	%	n	%	
SO4	237	94	40	15	6	128	54	Not significantly different
As	237	94	40	15	6	128	54	Not significantly different
Cd	237	94	40	15	6	128	54	Not significantly different
Cr	237	94	40	15	6	128	54	Not significantly different
Cu	237	94	40	15	6	128	54	Significant ($p < 0.05$)
Ni	237	94	40	15	6	128	54	Not significantly different
Zn	237	94	40	15	6	128	54	Significant ($p < 0.05$)
F	237	94	40	15	6	128	54	Not significantly different
Cl	237	94	40	15	6	128	54	Not significantly different

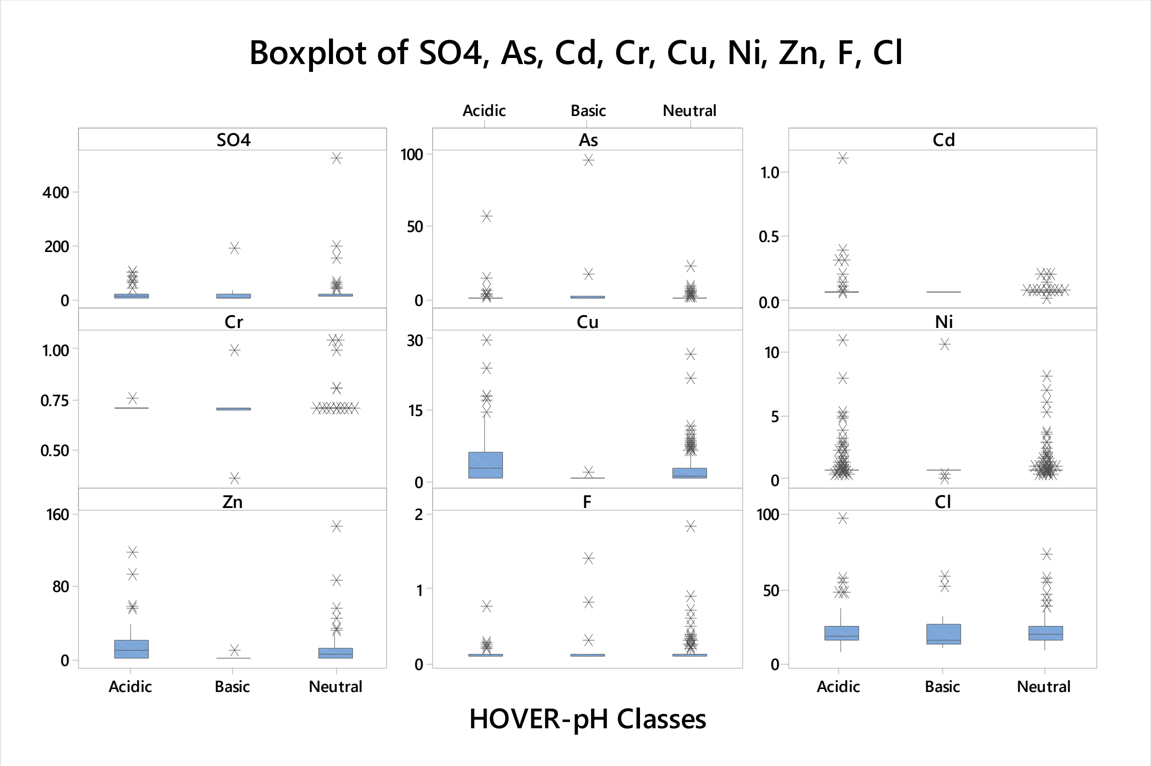


Figure 11: A comparison of the concentration distribution for the Acidic, Basic and Neutral HOVER-pH categories using box and whisker plots. Kruskal-Wallis tests showed that there were only significant differences in the medians of Cu and Zn.

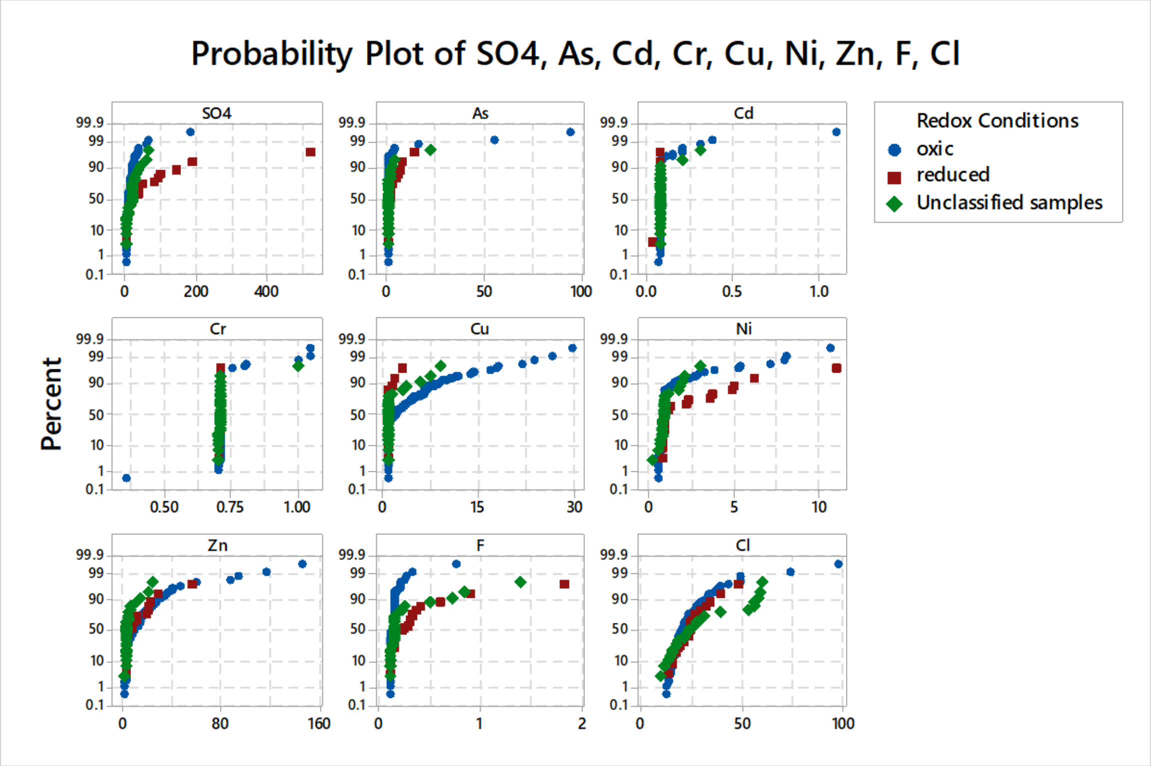


Figure 12: A comparison of the concentration distribution for the Acidic, Basic and Neutral HOVER-pH categories using probability plots. Kruskal-Wallis tests showed that there were only significant differences in the medians of Cu and Zn.

Table 14 shows that the Kruskal-Wallis tests show that there are significant differences between the medians of the different HOVER-redox categories for SO_4, As, Cu, Zn and F. Figure 13 and 14 show a comparison of the concentration distributions for the Oxic (A&B), Reduced (C&D) and Unclassified HOVER-redox categories using box and whisker plots and probability plots respectively.

Table 14: Results from the non-parametric Kruskal-Wallis statistical test to determine if there were significant differences in the median of the populations from Oxic (A&B), Reduced (C&D) and Unclassified HOVER-redox categories.

Elements	All	HOVER-redox categories						Kruskal-Wallis rank sum test
		A & B		C & D		Unknown		
	n	n	%	n	%	n	%	
SO4	237	188	79	22	9	27	11	Significant (p < 0.05)
As	237	188	79	22	9	27	11	Significant (p < 0.05)
Cd	237	188	79	22	9	27	11	Not significantly different
Cr	237	188	79	22	9	27	11	Not significantly different
Cu	237	188	79	22	9	27	11	Significant (p < 0.05)
Ni	237	188	79	22	9	27	11	Not significantly different
Zn	237	188	79	22	9	27	11	Significant (p < 0.05)
F	237	188	79	22	9	27	11	Significant (p < 0.05)
Cl	237	188	79	22	9	27	11	Not significantly different

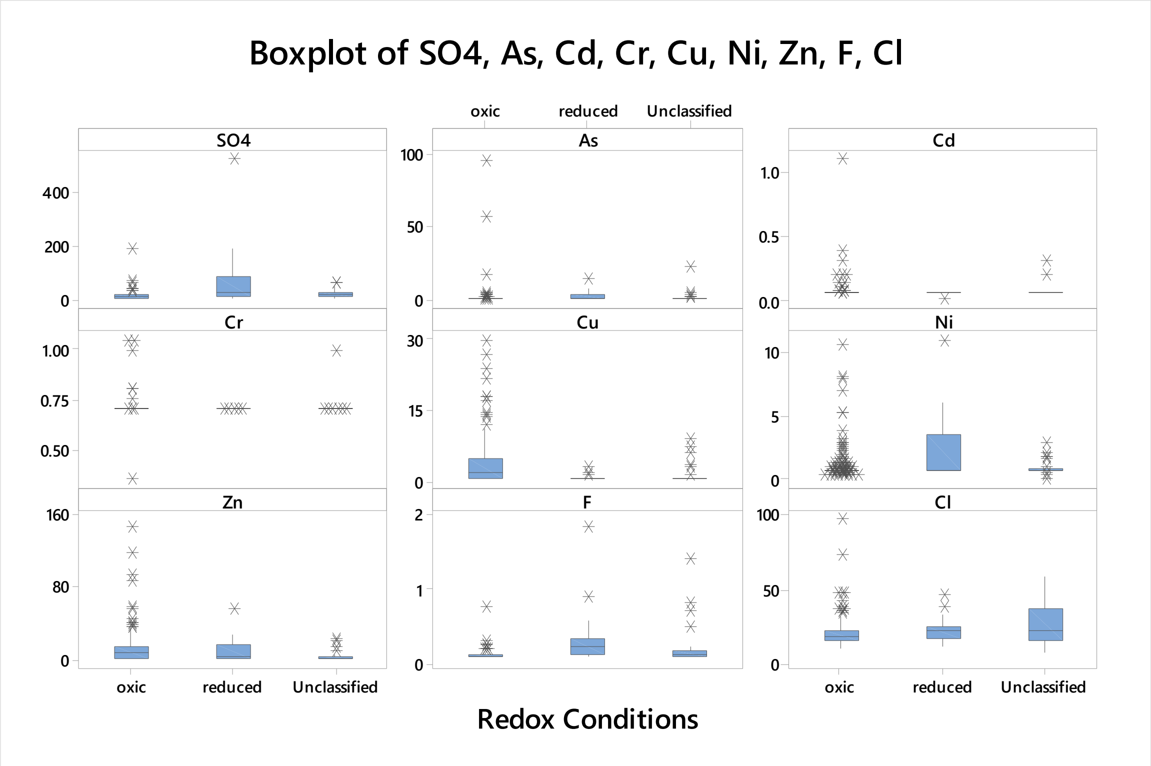


Figure 13: A comparison of the concentration distribution for the Oxic (A&B), Reduced (C&D) and Unclassified HOVER-redox categories using box and whisker plots. Kruskal-Wallis tests showed that there were significant differences in the medians of SO4, As, Cu, Zn and F.

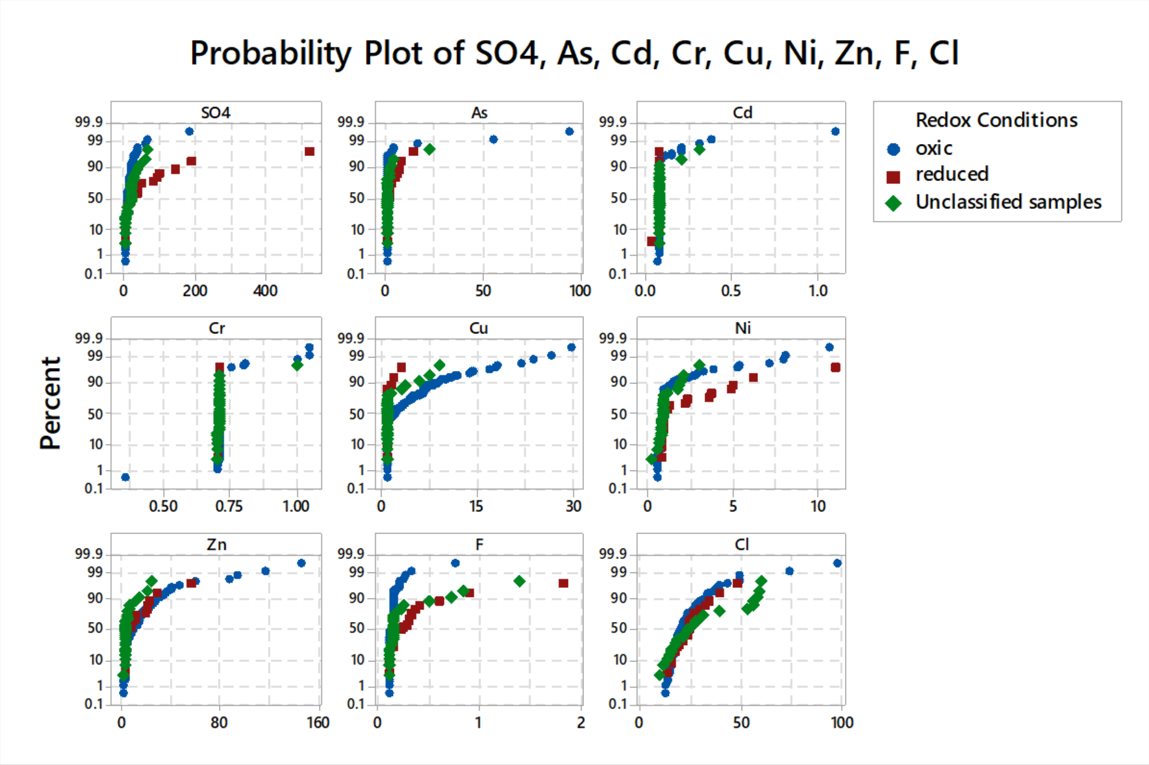


Figure 14: A comparison of the concentration distribution for the Oxic (A&B), Reduced (C&D) and Unclassified HOVER-redox categories using probability plots. Kruskal-Wallis tests showed that there were significant differences in the medians of SO4, As, Cu, Zn and F.

In addition to the consideration of the HOVER-pH and HOVER-redox categories, the influence of the concentrations which are less than the limit of detection. Figure 15 shows that Cd and Cr populations are dominated by concentrations less than the limit of detection indicated by the straight vertical line at one concentration. Therefore water families for these parameters are not consider further.

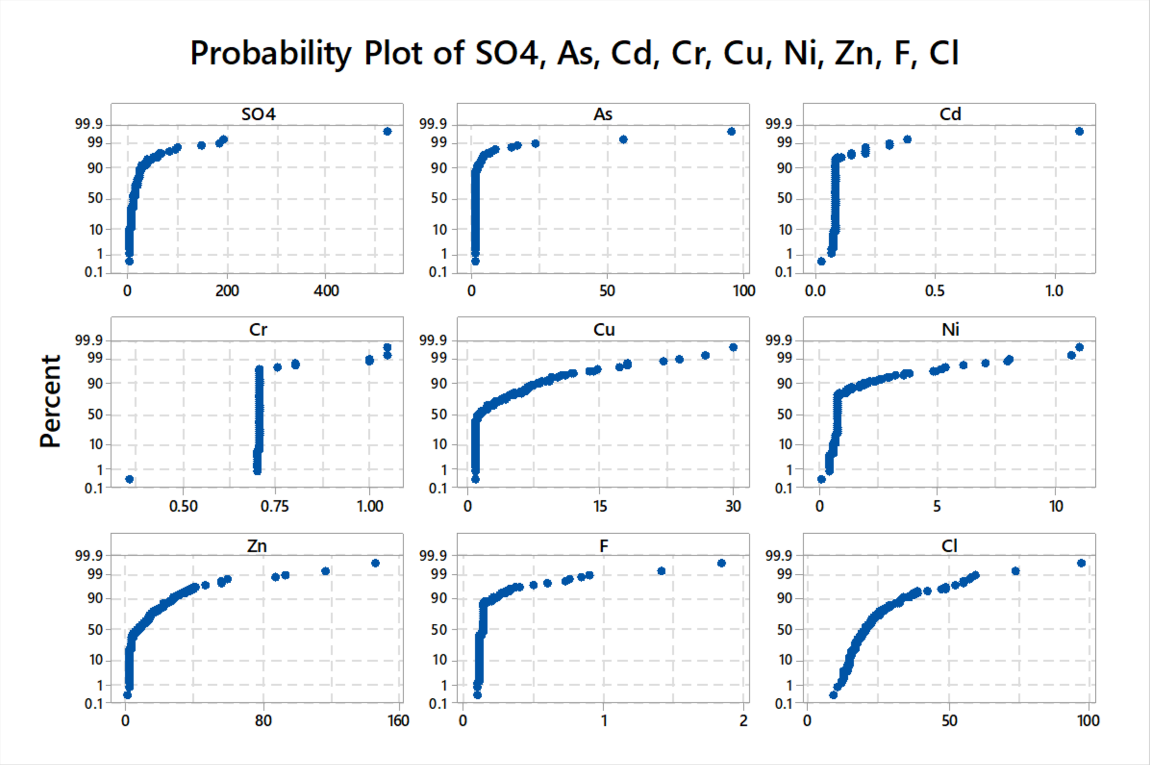


Figure 15: Probability plots showing the data distribution for the elements to be considered for NBL categorisation. It may be seen that the Cr and Cd populations are dominated by concentrations which are less than the limit of detection.

Table 15 shows a summary of the water families for which NBLs will be calculated for each parameter. NBLs for SO4, As, Ni and F will be calculated for a combination of lithology and HOVER-redox categories. NBLs for Cd and Cr will be calculated for the entire data population given the dominance of concentrations at less than the limit of detection. NBLs for Cu and Zn will be calculated for a combination of lithology, HOVER-redox and HOVER-pH categories. NBLs for Cl will be calculated for lithology only.

NBLs are calculated for all classes outlined in Table 16 with 10 or more monitoring points. As the monitoring points with Agricultural and Urban influence were removed for Cu and Zn, none of classes outlined in Table 15 met the threshold of 10 or more monitoring points, therefore NBLs are not calculated for these elements.

Table 15: Results from the non-parametric Kruskal-Wallis statistical test to determine if there were significant differences in the median of the populations from Oxic (A&B), Reduced (C&D) and Unclassified HOVER-redox categories.

Elements	Category				
	HOVER lithology	**HOVER redox**	**HOVER pH**	**pressures**	**Factor influencing hydrochemistry and NBL calculation**
SO4	s	s	ns	ns	lithology, redox
As	s	s	ns	ns	lithology, redox
Cd	ns	ns	ns	ns	High proportion of <LOD therefore NBL is calculated for entire data population (n=237)
Cr	ns	ns	ns	ns	High proportion of <LOD therefore NBL is calculated for entire data population (n=237)
Cu	s	s	s	s	lithology, redox, pH
Ni	ns	s	ns	ns	lithology, redox
Zn	s	s	s	s	lithology, redox, pH
F	s	s	ns	ns	lithology, redox
Cl	s	ns	ns	ns	lithology

4.2.6 *4.2.5 Removal of element concentrations influenced by contamination and NBL calculation*

The 90 percentile is used as the threshold value for defining reference concentrations by lithology. The 90 percentile makes it possible to estimate the maximum natural concentrations possible, particularly in particular geological contexts. The NBL is the highest of the values excluding anomalies.

The natural background level (90th percentile) calculated for Cd and Cr are shown in Table 16. The natural background level (90th percentile) calculated for SO_4, As, Ni, F and Cl per water family as outlined in Table 15 are shown in Table 17. Natural background levels were not calculated for Cu or Zn as there was not sufficient data (defined as less than 10 monitoring points) for any of the water families outlined for Cu and Zn in Table 15.

The ranges in natural background levels for 39 parameters in Irish groundwater were established previously by EPA. In that study, the upper limits of the ranges of NBLs were defined as the 95th percentile and the lower limits of NBLs were defined as the 5th percentile. A comparison table of the Irish NBLs (2017) is presented in Table 7.

Table 16: Table 12 The natural background levels (90 percentile) calculated for Cd and Cr

Elements	**Units**	**90^{th} percentile**
Cd	µg/l	<0.1
Cr	µg/l	<0.1

Table 17 : The natural background levels (90 percentile) calculated for SO4, As, Ni, F and Cl per water family as outlined in Table 15. “-“ indicates insufficient data available defined as less than 10 monitoring points.

HOVER-lithology	**Elements**	**units**	**HOVER-pH**	**HOVER-redox**			**90th percentile (only lithology)**
				A&B	**C & D**	**X**	
Sedimentary: sand	SO4	mg/l	all	22.1	-	-	22.1
	As	µg/l	all	0.48	-	-	0.6
	Ni	µg/l	all	0.78	-	-	0.8
	F	mg/l	all	<0.1	-	-	<0.1
	Cl	mg/l	all	27.0			27.0
Sedimentary: pure carbonates	SO4	mg/l	all	19.06	-	-	20.6
	As	µg/l	all	<1	-	-	0.7
	Ni	µg/l	all	1.3	-	-	1.3
	F	mg/l	all	<0.1	-	-	<0.1
	Cl	mg/l	all	31.7			31.7
Sedimentary: impure carbonates	SO4	mg/l	all	48.28	181.96	-	99.7
	As	µg/l	all	0.44	6.54	-	3.0
	Ni	µg/l	all	3.13	5.82	-	4.7
	F	mg/l	all	0.24	0.84	-	0.8
	Cl	mg/l	all	27.4			27.4
Sedimentary: other	SO4	mg/l	all	13.08	-	-	14.3
	As	µg/l	all	1.34	-	-	2.8
	Ni	µg/l	all	1.62	-	-	1.8
	F	mg/l	all	<0.1	-	-	<0.1
	Cl	mg/l	all	27.2			27.2
Volcanic rocks	SO4	mg/l	all	-	-	-	29.7
	As	µg/l	all	-	-	-	6.7
	Ni	µg/l	all	-	-	-	1.2
	F	mg/l	all	-	-	-	0.2
	Cl	mg/l	all	25.7			25.7
Metamorphic rocks	SO4	mg/l	all	32.54	-	33.22	34.6
	As	µg/l	all	1.4	-	3.56	2.5
	Ni	µg/l	all	5.19	-	0.71	4.6
	F	mg/l	all	<0.1	-	<0.1	<0.1
	Cl	mg/l	all	56.9			56.9

5 CASE STUDY - DENMARK

5.1 Previous studies on NBLs

5.1.1 *River Basin Management Plan 2015-2021 (Thorling and Sørensen, 2014)*

Natural background levels (NBLs) for the trace elements As and Ni were derived as part of the chemical status assessment of Danish groundwater bodies for the River Basin Management Plan 2015-2021 (Thorling and Sørensen, 2014). The derivation followed Guidance 18 (European Commission, 2009), thus the NBL were calculated based on the 90th percentile of the mean concentrations at water sampling points (well intakes) for the period 2007-2013, or if not available, for the period 2000-2006. The chemical data was from the nationwide database Jupiter, where all laboratory analyses of groundwater are reported. Both the geology and geography were taken into consideration: NBLs were calculated for each layer of the National Water Resources Model for Denmark (DK-model) and for ten geographical areas combining the main river basins (Figure 16). The geological model layers provide information on the type of aquifer (Quaternary sand, pre-Quaternary sand, chalk, and the various units on Bornholm) with multiple layers in depth for the sandy aquifers.

Groundwater redox conditions were also considered. Two redox water types were formulated based on nitrate concentration in the groundwater: oxic ($NO3 \geq 1$ mg/l) or reduced ($NO3 < 1$ mg/l) groundwater.

Water sampling points affected by known point-sources of anthropogenic pollution were excluded from the dataset prior to any analyses. The general NBLs were established only for the units where there were minimum 10 water sampling points with data, while the redox-dependent NBLs were calculated if there were more than 20 water sampling points.

Figure 17 shows the calculated 90th percentiles (Thorling and Sørensen, 2014) for the different river basins, geology, and redox-conditions. The boxplots summarise those values for the four different geological settings found in Denmark (Quaternary sand, pre-Quaternary sand, chalk/limestone, and the units on Bornholm). The median for different types of geology and redox conditions (same as in the boxplots) can be seen in Table 18.

The 90th percentiles for As were higher for reduced conditions, however there were not enough data for all DK-model layers, thus some of the highest values were calculated without taking into account the redox conditions. There were no NBLs above the current national threshold value (5 µg/l) for oxidised groundwaters, however for the rest (without differentiating based on redox and for the reduced waters) there were 90th percentiles exceeding the threshold for the Quaternary sand aquifers at the Limfjorden, Jylland øst, Fyn, Isefjord, and Østersøen, and in the chalk/limestone aquifers at Nordjylland and Østersøen.

Ni had higher 90th percentiles for oxidised conditions. The current national threshold value (10 µg/l) was exceeded for the Quaternary sands at Vadehavet, Limfjorden, Nordjylland, Isefjord, and Køge Bugt; the pre-Quaternary sands at Vadehavet and Nordjylland; and the chalk/limestones at Isefjord, Køge Bugt, and Østersøen. The reduced chalk/limestones at Køge Bugt had exceedance as well. However, if the redox conditions were not considered, there were NBLs exceeding the threshold for only the quaternary sands at Vadehavet, Limfjorden, and Køge Bugt, the units at Bornholm, and the chalk/limestone at Køge Bugt.

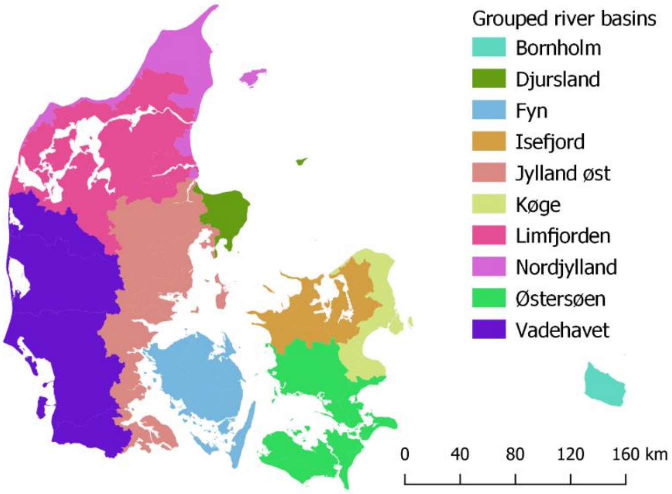


Figure 16 : Grouping of the main river basins in the ten geographic areas for which NBL were calculated

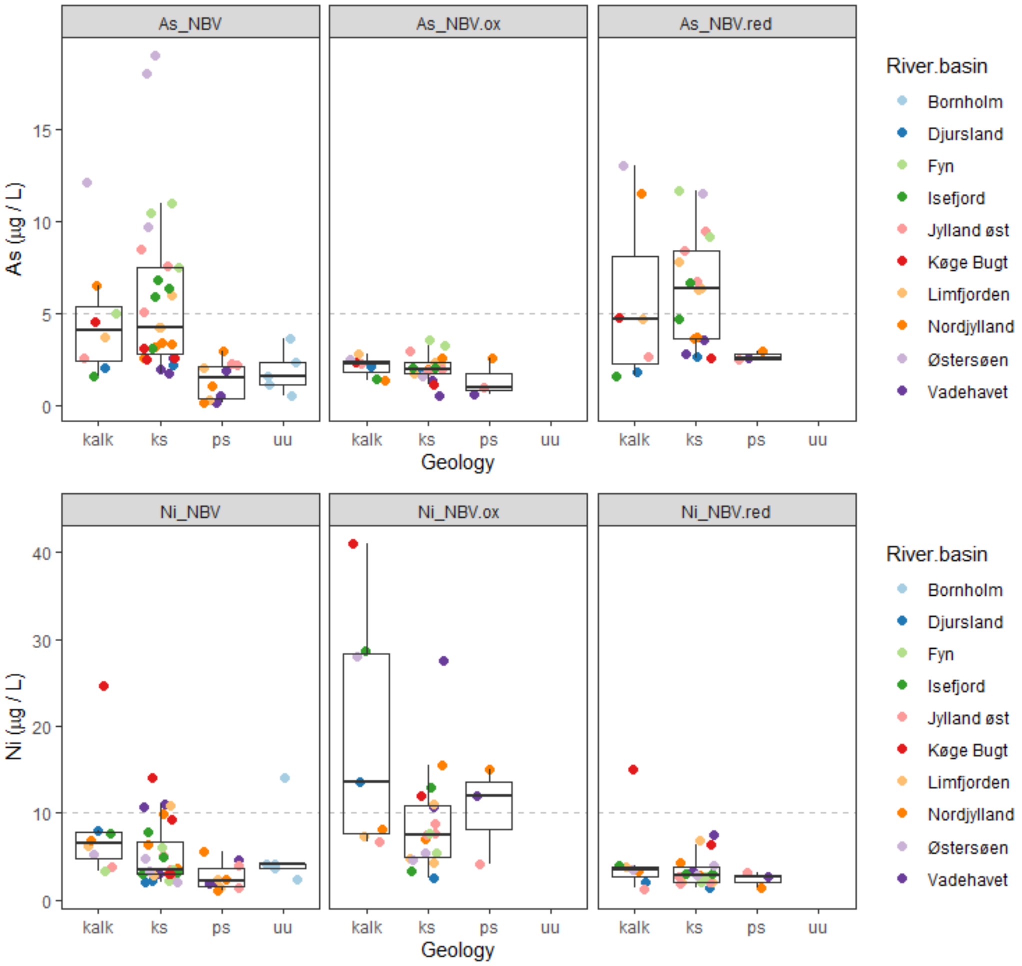


Figure 17: Natural background values (90th percentile) for As and Ni for the different river basins and different geological setting, where "kalk" is for chalk/limestone, "ks" is for Quaternary sand, "ps" is for pre-Quaternary sand, and "uu" is for all units at Bornholm. The 3 panels (horizontally) show the 90th percentiles calculated i) not considering the groundwater redox state (left), ii) only for the oxic groundwaters (middle), iii) and only for the reduced groundwaters (right). the horizontal dashed line is showing the current national threshold values for these trace elements; NBL results from: (Thorling and Sørensen, 2014)

Table 18 : Median of NBLs for different geology and redox conditions (based on the 90th percentiles calculated in Thorling and Sørensen (2014), visualized on Figure 17)

Geology	As (µg/l)			Ni (µg/l)		
	all	oxic	reduced	all	oxic	reduced
Chalk/limestone	4.2	2.3	4.8	8.2	**19.0**	4.7
Quaternary sand	4.3	2.0	**6.4**	5.2	8.8	3.4
Pre-Quaternary sand	1.5	1.0	2.6	2.6	**10.4**	2.4
Bornholm	1.7	-	-	5.6	-	-
Current national thresholds	5			10		

5.1.2 *River Basin Management Plan 2021-2027*

The national NBLs for trace elements (Al, As, Pb, Cd, Hg, Ni, Pb, Cu, Cr, Zn) are currently under revision as part of the work carried for the River Basin Management Plan 2021-2027. The results are not yet finalised, so we provide only a brief description of the methodology modifications from the previous version. In the current NBL calculation, a larger focus is placed on the different hydrogeochemical conditions in Danish aquifers. Instead of using the river basin grouping (10 areas), only five geographical areas are used (Jylland, Sjælland, Fynen, Bornholm, Islands near Jylland – Læsø, Samsø, and Anholt). At the same time, to distinguish between groundwater types, not only the redox conditions are considered (based on NO_3), but also the pH and the organic matter content (NVOC). Additionally, the geology classification was simplified, so it only includes four types (Chalk/limestone, Quaternary sand, Pre-Quaternary sand, all units in Bornholm, as used in Table 18 and Figure 17) without differentiating between the DK-model layers. The dataset also included only groundwater samples taken at wells part of the national groundwater monitoring program (GRUMO) and the waterworks wells used for drinking water production (BK) in the period 2000-2018. Values below the limit of detection (LOD) were substituted with ½ LOQ, where LOQ is defined as 3 times the limit of detection (LOD), thus the substitute values were calculated with the formula 1.5*LOD. The aggregation on water sampling point level was done by calculating mean of the annual means for each parameter and each location.

5.1.3 *Odense river basin pilot study (Hinsby et al., 2008)*

NBLs derived as both the 90th and the 97th percentile of pre-selected dataset (approximating a natural groundwater composition of a given aquifer based on the BRIDGE criteria) were calculated for the Odense river basin. Odense river basin is located on the isle of Fyn (see Figure 16). The groundwater bodies of this river basin consist of complex sequence of alternating glaciofluvial sand aquifers and clayey till confining layers. The focus of the assessment for Odense river basin, was on the derivation of threshold values for nitrogen and phosphorus based on environmental objectives for dependent ecosystems. However, NBLs (based on 90th percentiles) were reported also for Cl (83/64 mg/l), SO4 (114/117 mg/l), and As (11/8 µg/l) for two different sandy aquifer types in the basin.

5.2 Study with the proposed method

5.2.1 *Constitution of the groundwater quality dataset*

We use the Danish groundwater monitoring program dataset extracted in July 2019 (Thorling et al., 2019) from the national well database Jupiter[1]. From this raw dataset, only samples from the obligatory raw groundwater-monitoring program for drinking water production by public waterworks are used (dk: "Boringskontrol"). Only for pH, we used the previous-year version of the dataset, because pH was not present in Thorling et al. 2019 dataset.

Summary of HOVER WP 3-3 raw dataset:

- Period: 2009-2018 (incl. both years)
- Water sampling points: waterworks wells (at intake level) used for drinking water supply
- Contains all chemical analyses for pH, Eh, temperature, SO4, As, Cd, Cr, Cu, Ni, Zn, F, Cl, O2, NO3, Fe after the data-pretreatment procedures, as well as the calculated HOVER-pH and HOVER-redox
- Data-pretreatment included:
    - various element-specific quality checks,
    - treatment of all values below the limit of detection (< LOD)
    - Aggregation on sampling point level (see details below).

Element-specific quality checks

Negative pH and pH >10, temperature measurements <2 °C and >25 °C, and dissolved oxygen > 14 mg/l were considered erroneous and excluded. Zero concentrations for trace, minor, and major elements were removed from the dataset. For oxygen, however, zero concentrations were instead substituted with 0.01 mg/l, to account for measurement uncertainty. Field measurements for pH and dissolved oxygen were used when available; otherwise, lab measurements were used, if available. Electrical conductivity (EC) data did not pass our quality criteria assessment due to large number of mixed or wrong units and was not included. Single outliers for As, Ni, and Fe were excluded, due to obviously erroneous concentrations that were orders of magnitude higher than the next highest concentrations, most probably caused by unit reporting errors.

Treatment of values below the limits of detection (LOD)

The chemical laboratories report values below the limit of detection (LOD) with attribute "<". Before treating these censored values, we checked the ranges of LOD for each element and assessed if it was necessary to exclude any due to high LOD, i.e. low precision (high uncertainty). A single arsenic analysis <LOD with LOD= 4 µg/l was excluded prior to substitution, to avoid introducing artificially high concentrations to the dataset. No other DL were considered necessary to exclude at this stage.

All values < LOD were substituted with ½ LOQ, where LOQ = 3*LOD. Thus, the substitution formula is 1.5*LOD.

Table 19 shows the number of chemical analyses for each element, as well as an account of the LODs present in the dataset, including the range of LODs, the number and percentage of analyses < LOD, the max substitute value and the number of analyses with the max substitute.

---

[1] Jupiter database is hosted by GEUS and contains groundwater, drinking water, raw materials, environmental and geotechnical data for over 280000 wells throughout Denmark. The database is open to the public http://jupiter.geus.dk/

Aggregation

There were 885 samples with non-identical replicated analyses for some of the elements (same sample number, different concentration). Prior to aggregating on water-sampling point, we calculated sample-mean of these replicates. The aggregation on water sampling point level was calculated based on a median over the ten-year period of this study (2009-2018).

Table 19: Number and percentage of chemical analyses below the limit of detection (LOD), range of LODs in the dataset (after exclusion of 1 high DL for Arsenic), max substitute value and number of analyses with this max substitute; the last column shows the total number of chemical analyses after all data-pretreatment procedures, used in the aggregation step to calculate median for each water sampling point

Element	unit	All (n)	< LOD (n)	< LOD (%)	LOD range (min-max)	max substitute concentration	n	All analyses (n) for aggregation
SO4	mg/l	17977	404	2.2	0.2-1.5	2.25	4	17954
As	µg/l	17935	1491	8.3	0.01-1.0	1.5	1	17117
Cd	µg/l	660	559	84.7	0.003-0.5	0.75	1	628
Cr	µg/l	454	245	54.0	0.01-0.5	0.75	5	427
Cu	µg/l	616	206	33.4	0.03-5.0	7.5	1	578
Ni	µg/l	18906	3487	18.4	0.02-3.0	4.5	222	18027
Zn	µg/l	622	197	31.7	0.3-10.0	15	1	599
F	mg/l	17100	1132	6.6	0.02-0.5	0.75	6	17079
Cl	mg/l	18586	0	0	-	-	-	18528
O2	mg/l	18975	3464	18.3	0.01-2.0	6	1	17186
NO3	mg/l	17822	12666	71.1	0.002-1.0	1.5	323	17731
Fe	mg/l	17902	1136	6.3	0.0002-0.02	0.03	1	17780
pH	-	13971	-	-	-	-	-	13971
Eh	mV	1122	-	-	-	-	-	1122
Temperature	°C	14968	-	-	-	-	-	14968

Groundwater quality dataset

The groundwater quality dataset contains median concentrations for 15 chemical elements and parameters for 6509 water sampling points (intake level). Additionally, the HOVER-pH and HOVER-redox were calculated and added to the dataset. Table 20 and Table 21 show the distribution of water sampling points in the HOVER-pH and HOVER-redox classes.

The clean groundwater quality dataset was joined with the datasets containing the hydrogeological characteristics for each water sampling point and the dataset with prevailing anthropogenic pressures. The spatial distribution HOVER-pH, HOVER-redox, and the elemental concentrations will be presented for the final complete dataset. The clean groundwater quality dataset represents all water sampling points where data for at least one of the elements was present. However, for some of these locations, there may not be information about the geological setting.

Table 20 Number of water sampling points with chemical data for each HOVER-pH class; the unknown class shows the number of water sampling points without pH data (unable to classify in the 3 HOVER-pH classes)

HOVER-pH classes	SO4 (n)	As (n)	Cd (n)	Cr (n)	Cu (n)	Ni (n)	Zn (n)	F (n)	Cl (n)
Acidic (pH < 7)	342	342	19	8	12	342	23	341	342
Basic (pH > 7.5)	1769	1768	95	75	80	1769	99	1769	1769
Neutral (7 ≤ pH ≤ 7.5)	3682	3679	202	134	165	3681	200	3680	3682
unknown	686	682	44	38	37	686	46	676	687

Table 21 Number of water sampling points with chemical data for each of the HOVER-redox classes; the unknown class shows the number of water sampling points without data for O2, NO3, or Fe (unable to classify in the four HOVER-redox classes).

HOVER-redox classes	SO4 (n)	As (n)	Cd (n)	Cr (n)	Cu (n)	Ni (n)	Zn (n)	F (n)	Cl (n)
Oxic ("A")	513	513	16	11	15	513	15	513	513
Nitrate reducing, anoxic ("B")	234	233	12	7	7	234	12	233	234
Reduced ("C & D")	4902	4898	293	209	242	4900	302	4897	4902
Mixed ("X")	820	819	39	28	30	819	38	821	821
unknown	10	8	0	0	0	12	1	2	10

5.2.2 *Hydrogeological characteristics of the sampling points*

The purpose of this part of the study was to determine the HOVER-lithology and HOVER-age formation/stratigraphy for each of the water sampling points (n=6509). To be able to classify most of the water sampling points, we used primarily DK-model information, where the well intakes are coupled with a specific aquifer. All aquifers are classified in four geology classes (Table 22). Table 22 shows how the DK-model geological classes convert to the HOVER-lithology and HOVER-age classifications. Where direct conversion was impossible, we used additional information on the geology type at the intake depth (not available for all intakes). This was necessary especially for the HOVER-age classes, where only the Quaternary sand class could be directly classified as "Quaternary" age. For HOVER-lithology, we only needed to do this for the various units on Bornholm, classified as "uu" in the DK-model.

Table 22 Conversion of DK-model geology classes to HOVER-age and HOVER-lithology; if there is "x" no direct conversion was possible, so additional information on the geology type at intake level was used

DK-model geology	HOVER-age	HOVER-lithology
"qs" – Quaternary sand	Quaternary	Sedimentary: sand
"ps" – pre-Quaternary sand	x	Sedimentary: sand
"kalk" – chalk/limestone	x	Sedimentary: carbonates
"uu" – various geological units on Bornholm	x	x

It must be noted that the DK-model geology sand classes contain both sand and gravel. Thus, the HOVER-lithology class "Sedimentary: sand" is more precisely "Sedimentary: sand and gravel".

Table 23 HOVER-lithology class-representation of the water sampling points in the clean groundwater quality dataset; NA is for no data, for list with DGU mnemonic codes (e.g. as, bs..) see Appendix

HOVER-lithology classes	n	DK-model geology	DGU mnemonic codes (geology type at screen length)
Sedimentary: sand	4085	ks, ps	as, bs, dg, ds, dv, g, rg, rs, s, vs
Sedimentary: carbonates (limestone, chalk)	2262	kalk	
Sedimentary: other	12	uu	kj, kq, r
Crystalline rocks	8	uu	pa, a
Not specified (Bornholm)	21	uu	NA
NA	121	NA	

Table 24 HOVER-age formation class-representation of the water sampling points in the clean groundwater quality dataset; NA is for no data, NS for not specified; for list with DGU mnemonic codes, see appendix

HOVER-age formation/ stratigraphy	n	DK-model geology	DGU mnemonic codes (geology at the screen length) for the intakes with DK-model geology class		
			"ps"	"kalk"	"uu"
Quaternary	3394	ks, uu			dg, ds, dv, ml, di, dz
Neogene	296	ps, uu	kg, ks		kj
Paleogene	871	kalk		bk, lk, zk, kk, pk, ed	
Cretaceous	398	kalk, uu		sk, dk	as, bs, nl, vs, wl, bl
Cambrian	4	uu			kq,eq
Precambrian	7	uu			pa, pd
Classified	**4970**	**=76.4% of all water sampling points**			
Not-specific enough:					
Pre-quaternary	301	ps	NA/NS		
Neogene or Paleogene	77	ps	gi, gl, gs		
Paleogene or Cretaceous	1006	kalk		NA/NS	
Cretaceous or Jurassic	3	uu			rg, rs
NS (Bornholm)	31	uu			a, g, l, r, s, x, l, q, j
NA	121	NA			
Not classified	**1539**	**=23.6% of all water sampling points**			

The HOVER-age formation/stratigraphy classification is not used further in this study, because 23.6% of all water sampling points could not be properly classified. Instead we use the DK-model geology classes, so we can distinguish between Quaternary and pre-Quaternary sands. The chalk/limestone classes are of Palaeogene or Cretaceous ages, more specifically: Danian, Campanian-Maastrichtian, Selandian, and Eocene. The HOVER-lithology classes are used further as an alternative to the DK-model geology classes.

5.2.3 *Anthropogenic pressures / influences: diffuse or point pollutions*

To calculate the natural background levels (NBL) for the selected chemical elements we must first remove or minimize the influence of anthropogenic contamination. We address this issue by:

- Pre-selection of water sampling points limited only to the waterworks wells used for drinking water supply; Groundwater used for drinking water is generally of high quality, so only simple treatment is required (aeration and sand filtration). It is unlikely that these wells are affected by point-sources of pollution, because the usual practice of the water supply companies is to close the well, well field, or the specific waterworks and find another groundwater source complying with the drinking water standards. The main source of diffuse pollution is the agricultural sector.
- Testing the proposed methodology for identifying prevailing anthropogenic influences by using the CORINE landcover map[2]

We used a buffer of 1km around each water sampling point to determine the areal proportion of the land-cover types used in the CORINE land-cover 2012 (CLC-12) map (Figure 18). The 1km buffer is an approximation for the catchment area of the individual wells, where the water sampling locations are. It is possible that the actual catchment areas are either bigger, smaller, or not circular, depending on the local conditions. Since this is a national-level study, including more than 6350 wells (some with multiple water sampling depths), it was decided that the 1km-buffer would be an adequate proxy.

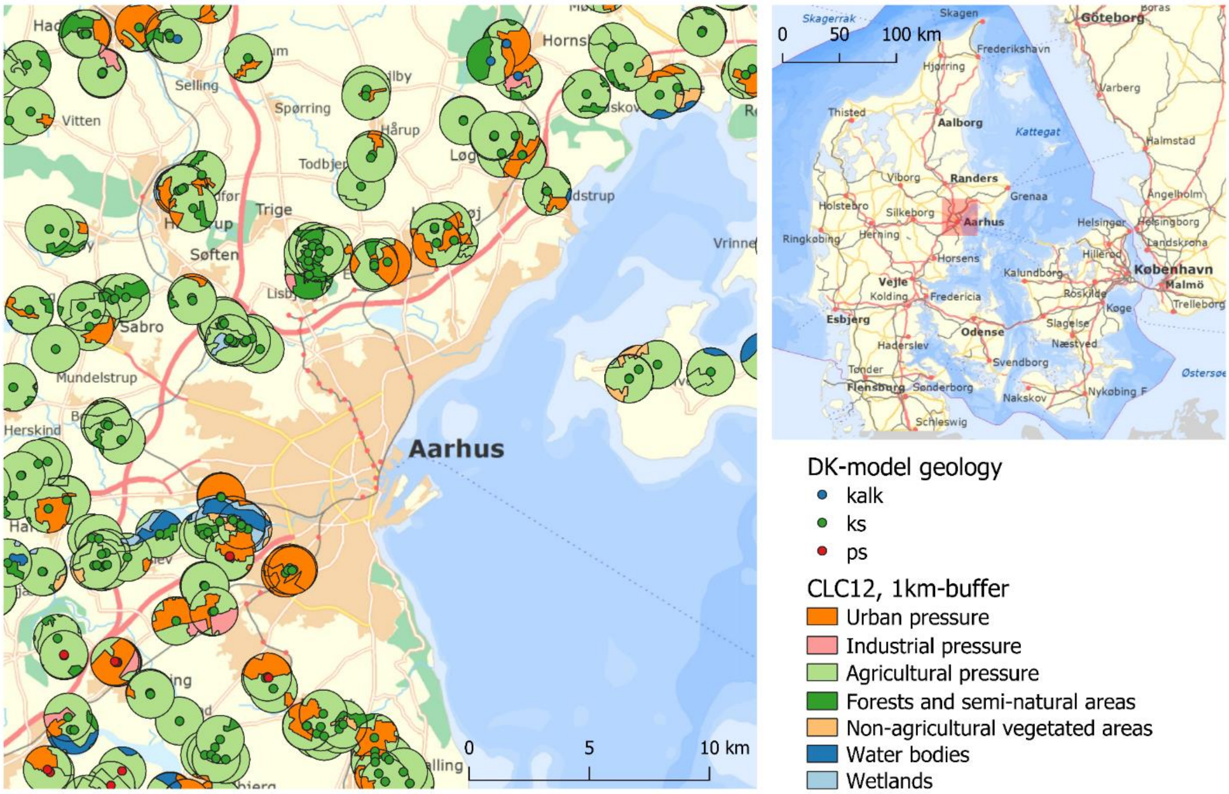


Figure 18 : Example of the area surrounding Aarhus (Denmark) with 1km buffers around the water sampling points (DK-model geology, where "kalk" is for chalk/limestone, "ks" for Quaternary sand, and "ps" for pre-Quaternary sand) and the land-cover types (CLC12 grouped by type). The red rectangle on the inset map shows the location of the area.

---

[2] Corine land cover (CLC) with reference year 2012 for Denmark (version 1, Oct. 2014) in 1: 100000 scale; downloaded from https://download.kortforsyningen.dk/content/corine-land-cover

The CLC-12 land-cover classes were grouped, so that the HOVER WP33 anthropogenic influence variables can be computed. For the “yes”/”no” variables (Table 25), if any of the listed CLC-12 codes was present within the 1km buffer irrespective of the areal proportion, the corresponding water sampling point was classified with “yes”, i.e. the water point is potentially influenced by the specific anthropogenic pressure.

The variable “Prevailing pressure” takes the values “urban”, “industrial”, “agricultural”, or “mining” based on the areal proportion of these individual anthropogenic pressures. For example, if the CLC-12 codes for agricultural pressure collectively have the largest proportion of the 1km buffer of a water sampling point, this water sampling point is assigned the values “agricultural”. If the 1km buffer area was dominated by other CLC-12 codes than the ones listed in Table 25, the water sampling point was assigned the value “natural or other”, because it could be any of the codes from Level 1 categories: Forests and semi natural areas, Wetlands, or Water bodies.

The variable “Anthropogenic pressure” takes the values “no: Natural” and “yes: Anthropogenic”. If any of the CLC-12 codes listed in Table 25 are present within the 1km buffer, the water sampling point is assigned the value “yes: Anthropogenic”.

From the 6509 water sampling points, 121 could not be classified in the HOVER-lithology class (due to no data), so these were excluded from the dataset. The rest of the water sampling points (n=6388) were classified as presented in Table 26, Table 27, and Table 28. These tables show that the most widespread anthropogenic pressure comes from the agriculture. Nearly all water sampling points have at least one type of anthropogenic pressure in the 1km buffer zone around them.

Table 25 HOVER WP 33 anthropogenic pressure “yes”/”no” variables; only the Level 1 codes for Agriculture that are present within our dataset are shown;

HOVER WP 33 class	Values	CLC-12 codes	CLC-12 meaning	Includes all codes at CLC level
Urban pressure	“yes”/”no”	111 112	Continuous urban fabric Discontinuous urban fabric	L2: Urban fabric
Industrial Pressure	“yes”/”no”	121 122 123 124	Industrial or commercial units Road and rail networks and associated land Port areas Airports	L2: Industrial, commercial and transport units
Agricultural Pressure	“yes”/”no”	211 222 231 242 243	Non-irrigated arable land Fruit trees and berry planation Pastures Complex cultivation pattens Land principally occupied by agriculture with significant areas of natural vegetation	L1: Agricultural areas
Mining Influence	“yes”/”no”	131 132 133	Mineral extraction sites Dump sites Construction sites	L2: Mine, dump, and construction sites

Table 26 Number and percentage of water sampling points in each anthropogenic "yes"/"no" pressure variable

HOVER WP 33 class	Value	Water sampling points (n=6388)	
		n	%
Urban pressure	no	2570	40.2
	yes	3818	59.8
Industrial Pressure	no	5506	86.2
	yes	882	13.8
Agricultural Pressure	no	237	3.7
	yes	6151	96.3
Mining Influence	no	6271	98.2
	yes	117	1.8

Table 27 Number and percentage of water sampling points in different classes of the variable "Prevailing pressure" (dominating pressure, area-wise)

Prevailing pressure (class)	Water sampling points (n=6388)	
	n	%
Agricultural	5477	85.7
Industrial	45	0.7
Mining	2	< 0.1
Urban	802	12.6
Natural or other	62	1.0

Table 28 Number and percentage of water sampling points in different classes of the variable "Anthropogenic pressure"

Anthropogenic pressure (class)	Water sampling points (n=6388)	
	n	%
no: Natural	48	0.8
yes: Anthropogenic	6340	99.2

5.2.4 *Master dataset (step 13)*

The clean groundwater quality dataset was joined with the datasets with hydrogeological characteristics and the anthropogenic pressures. This dataset is further referred to as "Master dataset" and contains 6388 water sampling points and the variables from Table 29.

Table 29 Master dataset used further in this study

Variable	Data type	Explanation
BORID	num	ID number of the well (used as key together with INDTNR)
INDTNR	num	ID of the well screen (used as key together with BORID)
CELLCODE	txt	Cell code for the EEA Reference grid 1 km
EOFORIGIN	Num	East of origin (EAA Reference grid 1km)
NOFORIGIN	Num	North of origin (EAA Reference grid 1km)

Variable	Data type	Explanation
DGUNR	txt	ID label of the well (can be used for searching in Jupiter)
XUTM32EUREF89	num	X coordinate in UTM 32N EUREF 89 coordinate system
YUTM32EUREF89	num	Y coordinate in UTM 32N EUREF 89 coordinate system
INDTTOP	num	Top of groundwater abstraction filter (meters below terrain)
INDTBUND	num	Bottom of groundwater abstraction filter (meters below terrain)
TERRAENKOTE	num	Terrain elevation
GVFOREKOM	txt	Groundwater body number, e.g. “dkmb_1795_uu”
nbl_unit	txt	Type of aquifer (location, geology), e.g. “dkmj_ks” stands for DK model Jutland, quaternary sand
geology	txt	DK-model geology, see Table 22
HOVER_lithology	txt	HOVER-lithology, see Table 23
HOVER_age	txt	HOVER-age formation / stratigraphy, see Table 24
urban	txt	Urban pressure (“yes”/”no”), see Table 25 and Table 26
mining	txt	Mining influence (“yes”/”no”), see Table 25 and Table 26
industrial	txt	Industrial pressure (“yes”/”no”), see Table 25 and Table 26
agricultural	txt	Agricultural pressure (“yes”/”no”), see Table 25 and Table 26
prevailing	txt	Prevailing pressure variable, see Table 27
anthropogenic_pressure	txt	Anthropogenic pressure variable, see Table 28
pH	num	Median of pH over the 10-year study period (-)
Eh	num	Median Eh over the 10-year study period (mV)
Temp	num	Median temperature over the 10-year study period (°C)
SO4	num	Median SO4 over the 10-year study period (mg/l)
As	num	Median As over the 10-year study period (µg/l)
Cd	num	Median Cd over the 10-year study period (µg/l)
Cr	num	Median Cr over the 10-year study period (µg/l)
Cu	num	Median Cu over the 10-year study period (µg/l)
Ni	num	Median Ni over the 10-year study period (µg/l)
Zn	num	Median Zn over the 10-year study period (µg/l)
Fl	num	Median F over the 10-year study period (mg/l)
Cl	num	Median Cl over the 10-year study period (mg/l)
O2	num	Median O2 over the 10-year study period (mg/l)
NO3	num	Median NO3 over the 10-year study period (mg/l)
Fe	num	Median Fe over the 10-year study period (mg/l)
HOVER_pH	txt	HOVER-pH, see Table 20
HOVER_redox	txt	HOVER-redox, see Table 21

5.2.5 *Statistical treatments*

Discriminant function analysis (Step 14)

There are different types of discriminant function analysis (DFA) methods, here we chose to perform the simplest one -- linear discriminant analysis (LDA). LDA uses linear combination of predictors to predict the class/category of given observations. The purpose of this analysis is to quantify how well the selected predictor variables (major, minor, trace elements etc.) discriminate between different groups (HOVER-lithology and DK-geology classes. The predictive power of the models is used to compare the different possible combinations of predictor variables.

No missing data is tolerated in this analysis, so the number of data-points is dependent on the chosen predictor variables (Table 30). The trace elements Cd, Cr, Cu, Zn were excluded from the dataset due to large number of missing values. Additionally, all observation of the HOVER-lithology classes “Crystalline rocks” (n=8), “Not specified (Bornholm)” (n=21), “Sedimentary: other” (n=12) were excluded from the models 1, 2, and 4 due to insufficient data. For model 6 we also excluded the Bornholm class “uu” from the DK-geology classes after it was seen that model 5 does not discriminate well “uu” class from the rest of the DK-geology classes (see Appendix).

The LDA method assumes that the predictor variables are normally distributed and that the classes have identical co-variances. All our predictor variables are left-skewed, thus failing the normality assumption, so we use a Box-Cox transformation. We also scaled and centered the data prior to analysis, because the LDA can be affected by the units of the variables. The data is split in training (80%) and test set (20%), so that an LDA model can be built with the training set and tested on the test set. Table 30 provides details for the LDA models and the number of data-points for each of the models.

We formulated three models for predicting HOVER-lithology classes (classifying data-points) and three for the DK-geology classes. The difference between model 1 and 2 is that model 1 has O2, NO3, and Fe as predictor variables, which are used for defining the HOVER-redox classes. The difference between Models 1 and 4 is that the latter also has pH as a predictive variable, but there are less data-points. Models 3 and 5 are predicting the DK-geology classes, but use the same predictor variables as Models 1 and 4. The predictor variables for model 6 are the same as for model 5, but the Bornholm class (“uu”) is excluded from the analysis.

Table 30 Linear discriminant analysis: model definition and comparison of model performance

LDA models	Predicting	Predictor variables	Data (n)	Accuracy	95% CI	Kappa
model 1	HOVER-lithology	SO4, As, Ni, Fl, Cl, O2, NO3, Fe	6292	81.8%	(79.6, 83.9)	0.591
model 2	HOVER-lithology	SO4, As, Ni, Fl, Cl	6300	80.8%	(78.5, 82.9)	0.565
model 3	DK-geology	SO4, As, Ni, Fl, Cl, O2, NO3, Fe	6333	75.7%	(73.3, 78.1)	0.567
model 4	**HOVER-lithology**	**pH, SO4, As, Ni, Fl, Cl, O2, NO3, Fe**	**5652**	**81.7%**	**(79.3, 83.9)**	**0.589**
model 5	DK-geology	pH, SO4, As, Ni, Fl, Cl, O2, NO3, Fe	5689	74.9%	(72.3, 77.4)	0.557
Model 6	**DK-geology (-uu)**	**pH, SO4, As, Ni, Fl, Cl, O2, NO3, Fe**	**5620**	**75.8%**	**(73.1, 78.2)**	**0.566**

Figures in Appendix provide information for number of data points in each HOVER-lithology and DK-geology class for each of the six LDA models. The model performance is compared based on overall accuracy and kappa statistics (Table 30). All models have nearly similar accuracies (i.e. classify correctly): 75% of all water sampling points for DK-geology classes and 81% of all water sampling points for HOVER-Lithology classes. We show and discuss the results for model 4 and 6 here, while the rest of the results are provided in Appendix.

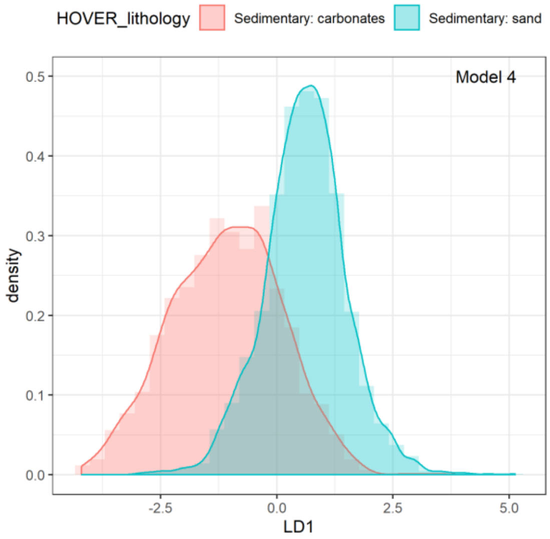


Figure 19 Linear discriminant function for model 4

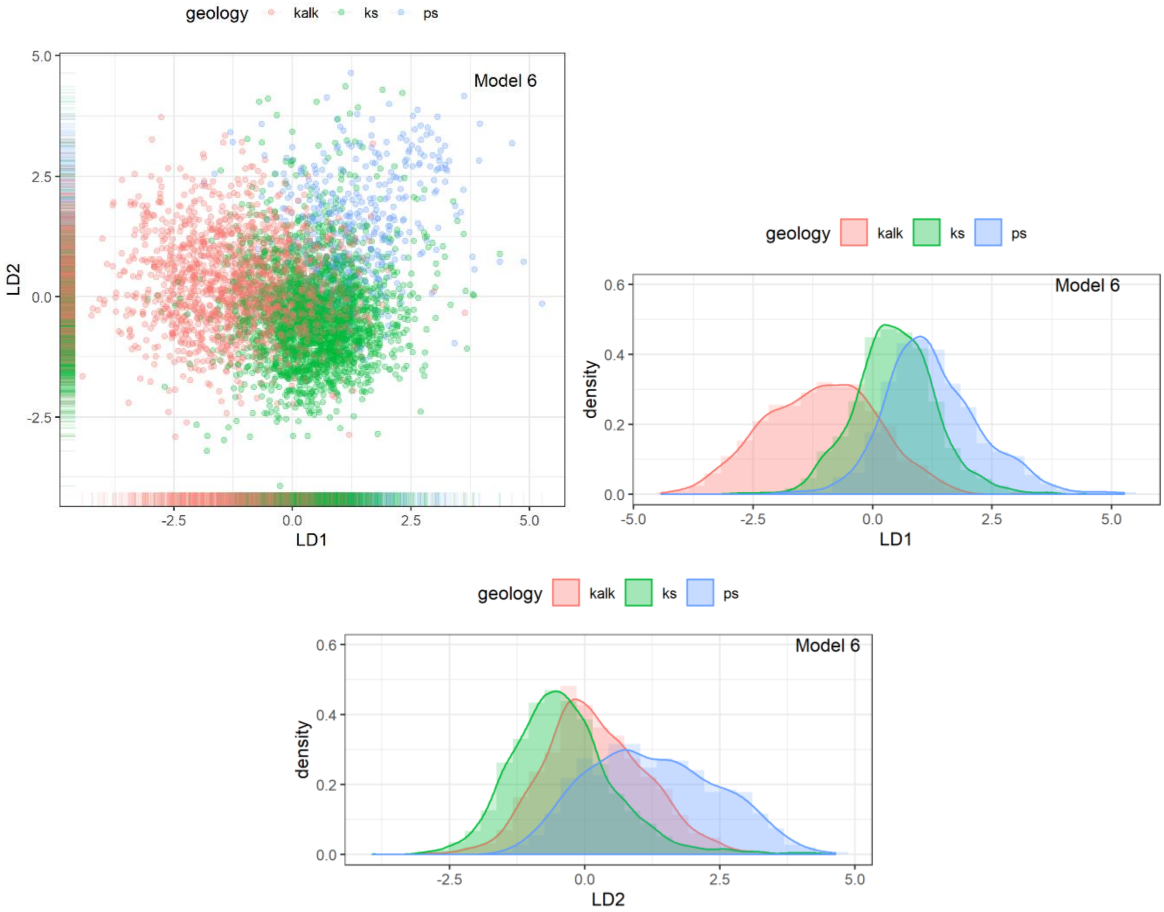


Figure 20 Linear discriminant functions for model 6, where the percentage separation achieved by the discriminant function LD1 is 72.2% and by LD2 is 27.8%

The results show that a better discrimination between types of geology/lithology is achieved when only the two HOVER-lithology classes are used. Model 6 manages to discriminate between the 3 DK-geology classes (chalk/limestone “kalk”, pre-Quaternary sand “ps”, Quaternary sand “ks”), but there is also a substantial overlap of the distributions. LD1 separates the sands from the chalk/limestone, while LD2 separates the pre-quaternary from the quaternary sand. From both LD1 and LD2 (Figure 20) we can also see that the chalk/limestone group is more similar to the quaternary sand than to the pre-quaternary sand. The overlap between the 3 distributions shows also that some water sampling points have very similar composition (concentration of predictor variables) even though they belong to different geological/lithological groups.

The conclusion of this discriminant function analysis is that based on the selected elements (pH, SO4, As, Ni, Fl, Cl, O2, NO3, Fe) there is significant difference between the two classes of HOVER-lithology (“Sedimentary: carbonates” and “Sedimentary: sand”) and the 3 classes of DK-geology (“kalk”, “ks”, “ps”). Therefore, it was decided that the 90th percentiles will be calculated for these classes only.

Prevailing pressures (Step 15)

Kruskal-Wallis rank sum test is a non-parametric method for testing if two or more groups originate from the same distribution. The alternative hypothesis is that at least one group differs from the rest. To determine which group(s) differ, we performed pairwise comparisons using Nemenyi-test (as a post-hoc) with Chi-squared approximation for independent samples (to account for ties). Statistical significance is assessed at the 95% confidence level, i.e. significant differences are found if $p < 0.05$. For this test, we used the categorical variable “Prevailing pressure”. The purpose of this analysis is to test if there is significant difference in the concentrations at sampling points with different prevailing anthropogenic pressures.

Table 31 Number of water sampling points with data for each element and each group of prevailing pressure and results from the Kruskal-Wallis rank sum test;

	all	Prevailing pressures										Kruskal-Wallis rank sum test
		agricultural		industrial		mining		urban		no or other		
	n	n	%	n	%	n	%	n	%	n	%	
SO4	6360	5454	85.8	45	0.7	2	0.03	797	12.5	62	0.97	Significant ($p < 0.05$)
As	6352	5447	85.8	45	0.7	2	0.03	796	12.5	62	0.98	Significant ($p < 0.05$)
Cd	356	287	80.6	17	4.8			52	14.6			Significant ($p < 0.05$)
Cr	250	187	74.8	7	2.8			56	22.4			Significant ($p < 0.05$)
Cu	289	217	75.1	7	2.4			64	22.1	1	0.35	Significant ($p < 0.05$)
Ni	6358	5452	85.8	45	0.7	2	0.03	797	12.5	62	0.98	Significant ($p < 0.05$)
Zn	363	281	77.4	19	5.2		0	62	17.1	1	0.28	Not significantly different
F	6349	5444	85.7	45	0.7	2	0.03	796	12.5	62	0.98	Significant ($p < 0.05$)
Cl	6361	5454	85.7	45	0.7	2	0.03	798	12.5	62	0.97	Significant ($p < 0.05$)

Table 31 shows that most of the water sampling points are with dominating agricultural pressures (range 74.8-85.8% of all locations). This means that within a buffer of 1km around the sampling point, the largest proportion of land is with agricultural land-use. This bias towards agriculturally dominated locations is because more than half of Denmark is arable land (62% in 2017, Danmarks Statistik[3]). It

---

[3] https://www.dst.dk/da/Statistik/nyt/NytHtml?cid=24323

should also be noted that the pre-selection of sampling locations, which included only waterworks wells used for drinking water supply, might be reflected in the low number of points with industrial, mining, or urban influences. The waterworks wells are usually in areas where there are no known point sources of pollution.

The Kruskal-Wallis results showed that except for Zn, all other elements have at least one group that is significantly different (Table 31). The results of the post-hoc Nemenyi test are presented on the following boxplot graphs (Figure 21). The prevailing pressure categories sharing letters (above the boxplot) are not significantly different. Interpretation of the element-specific results is provided after Figure 21.

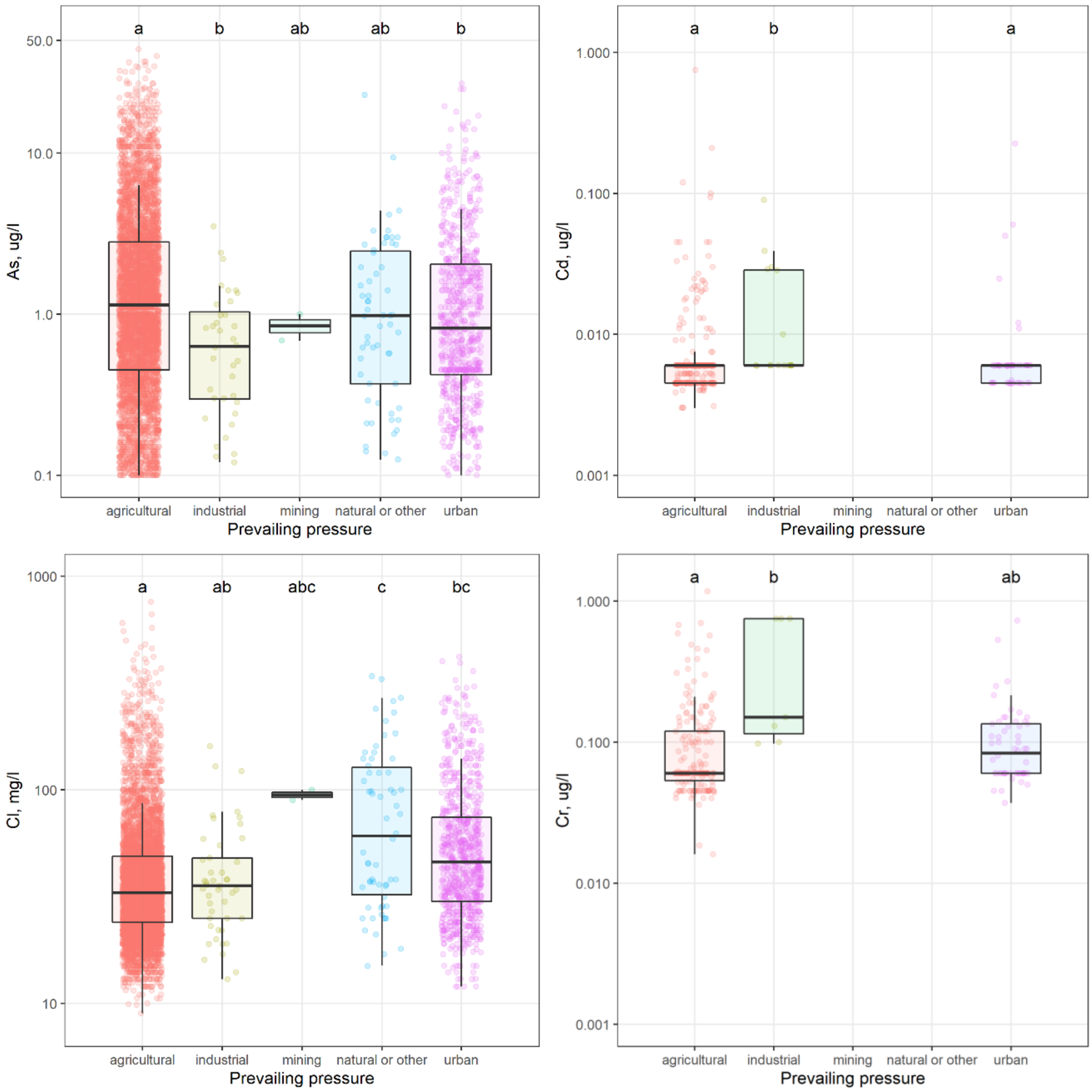
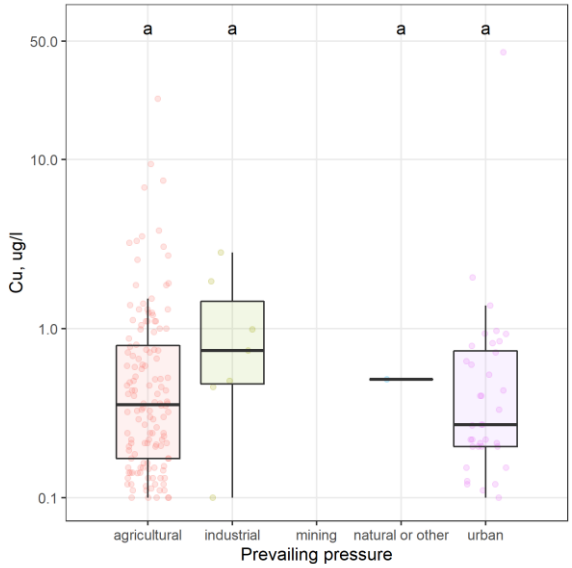
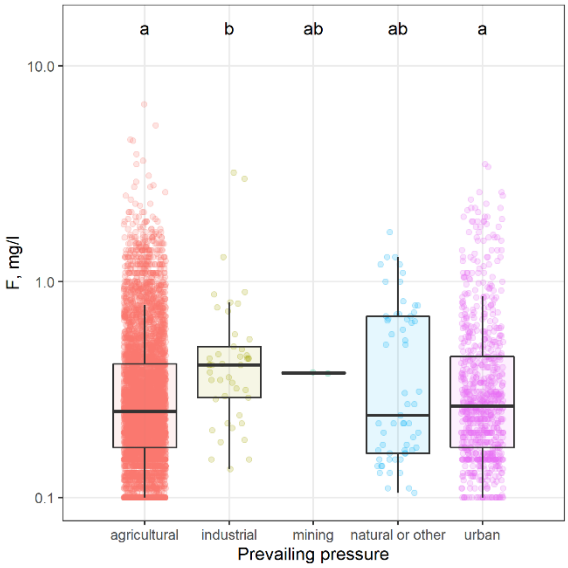
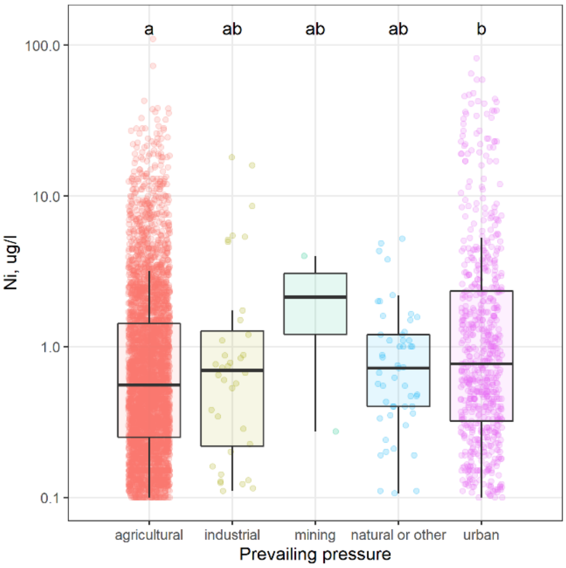
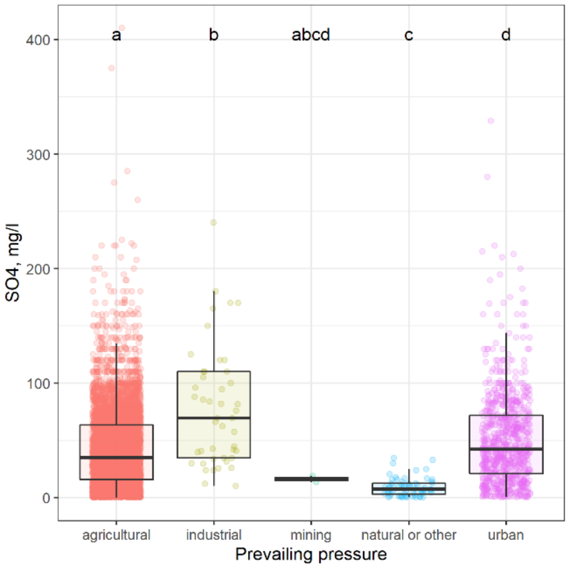


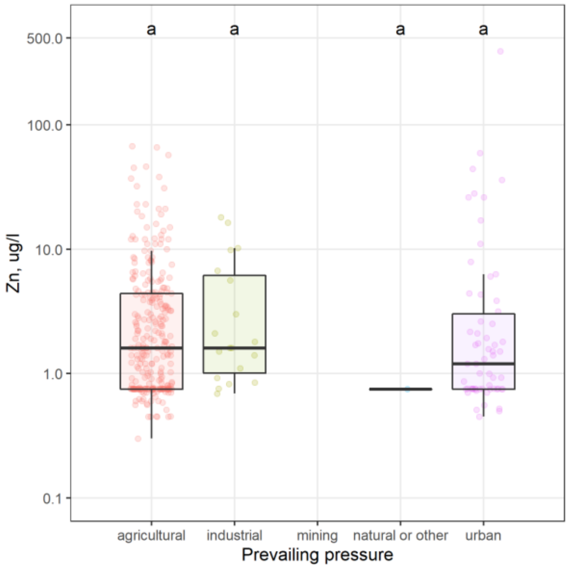
*Figure 21 Comparison of concentration distributions for the "Prevailing pressure" categories; Dots - median concentration at water sampling location, displaced horizontally only. Boxplot: thick horizontal line – median, the lower/upper hinges correspond to the Q1/Q3 (the 25th and 75th percentiles), the upper/lower whiskers extends to the largest/smallest value no further than 1.5 * IQR from the hinge. Categories with shared letters are not significantly different. All except SO4 are with log 10 transformed y-axis.*


a
a
a
a
50.0
10.0
1.0
0.1
Cu, ug/l
agricultural
industrial
mining
natural or other
urban
Prevailing pressure


a
b
ab
ab
a
10.0
1.0
0.1
F, mg/l
agricultural
industrial
mining
natural or other
urban
Prevailing pressure


a
ab
ab
ab
b
100.0
10.0
1.0
0.1
Ni, ug/l
agricultural
industrial
mining
natural or other
urban
Prevailing pressure


a
b
abcd
c
d
400
300
200
100
0
SO4, mg/l
agricultural
industrial
mining
natural or other
urban
Prevailing pressure


a
a
a
a
500.0
100.0
10.0
1.0
0.1
Zn, ug/l
agricultural
industrial
mining
natural or other
urban
Prevailing pressure

The "agricultural group" is used here as the background state to which all the rest of the groups are compared. We do this, because, as discussed previously, the agricultural pressure dominates the dataset (~80% of all water sampling points are potentially affected) and the "natural group" has less than 1% of the points.

In our application of the method, we only remove points belonging to the "industrial" and/or "urban" groups, based on their statistically significant difference from "agricultural". The "mining" group is not significantly different from the rest of the groups; this could be explained by the negligible number of water sampling points in this group (max 2).

Therefore, prior calculating the NBL for a specific element, we **exclude the following groups** of data (element-specific):

- for As and SO4 – both "industrial" and "urban"
- for Cd, Cr, and F – only "industrial"
- for Ni – only "urban"

For Cu and Zn, the test did not find significant difference between any of the groups, so we did not remove any group.

For Cl, we did not apply this method, because the most relevant anthropogenic influence would be the unsustainable pumping rates, which may cause up-coning (abstraction of older saline water, relevant in carbonate sediments in Denmark) or sea-water intrusion (near the coast or other saline environments). Additional diffuse source of pollution could be the use of road salt.

We discuss the application of this methodology with the Danish dataset and propose potential future improvements in section 5.3.

HOVER-pH and HOVER-redox importance

We used the same methodology (Kruskal-Wallis with post-hoc Nemenyi) to test if the groups based on HOVER-pH and HOVER-redox are significantly different. We performed this analysis to determine if and for which of the elements it is relevant to include pH and/or redox grouping when calculating the 90th percentiles, and to assess if it is necessary to use all four redox groups (A, B, C&D, X).

Table 32 shows that more than half of the data points are with neutral pH (neutral group with pH ≥7 & pH ≤7.5). About 1/3 is with basic (pH > 7.5), around 5% are acidic (pH < 7), and 10-14% cannot be classified because there is no pH data. Thus, the dataset is biased towards neutral and basic groundwaters. Table 33 shows that the dataset is also biased towards reduced groundwaters, as 3/4 of the data points are classified in the C & D redox type.

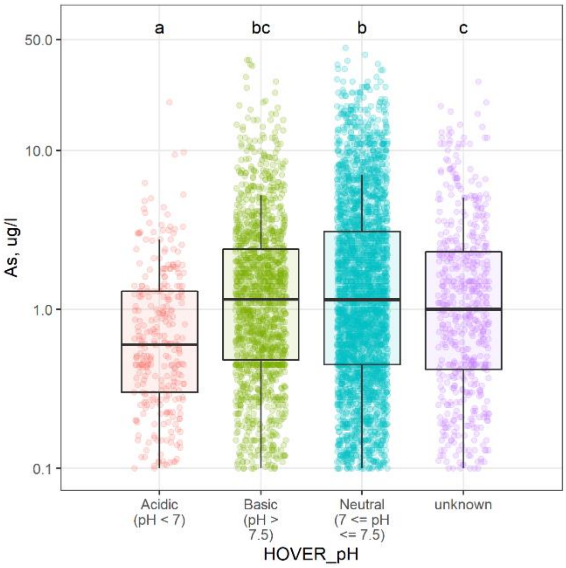
Table 32 Number and percentage of data-points (water sampling points) within each of the HOVER-pH groups, and results from the Kruskal-Wallis rank sum test

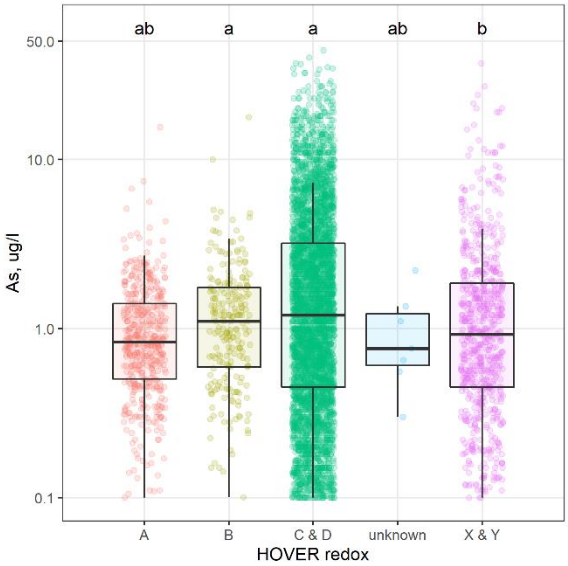
Elements	All	HOVER-pH categories								Kruskal-Wallis rank sum test
		Acidic		Basic		Neutral		Unknown		
	n	n	%	n	%	n	%	n	%	
SO4	6360	340	5.3	1722	27.1	3637	57.2	661	10.4	Significant (p < 0.05)
As	6352	340	5.4	1721	27.1	3634	57.2	657	10.3	Significant (p < 0.05)
Cd	356	19	5.3	94	26.4	201	56.5	42	11.8	Significant (p < 0.05)
Cr	250	8	3.2	74	29.6	132	52.8	36	14.4	Not significantly different
Cu	289	12	4.2	79	27.3	163	56.4	35	12.1	Not significantly different
Ni	6358	340	5.3	1722	27.1	3635	57.2	661	10.4	Significant (p < 0.05)
Zn	363	23	6.3	98	27.0	198	54.5	44	12.1	Significant (p < 0.05)
Fl	6349	339	5.3	1722	27.1	3635	57.3	653	10.3	Significant (p < 0.05)
Cl	6361	340	5.3	1722	27.1	3637	57.2	662	10.4	Significant (p < 0.05)

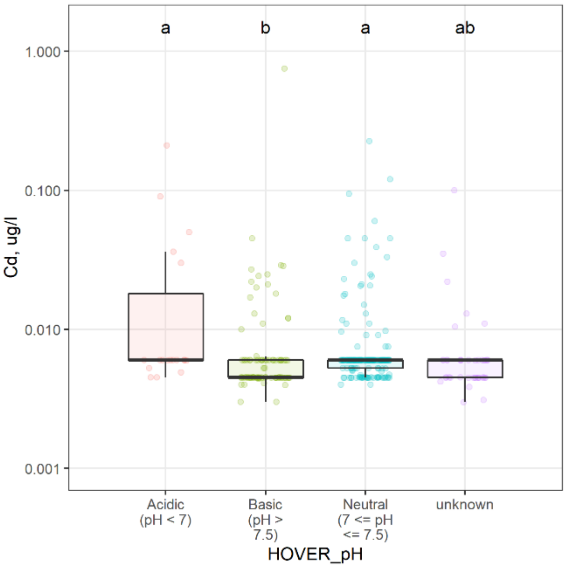
Table 33 Number and percentage of data-points (water sampling points) within each of the HOVER-redox groups, and results from the Kruskal-Wallis rank sum test

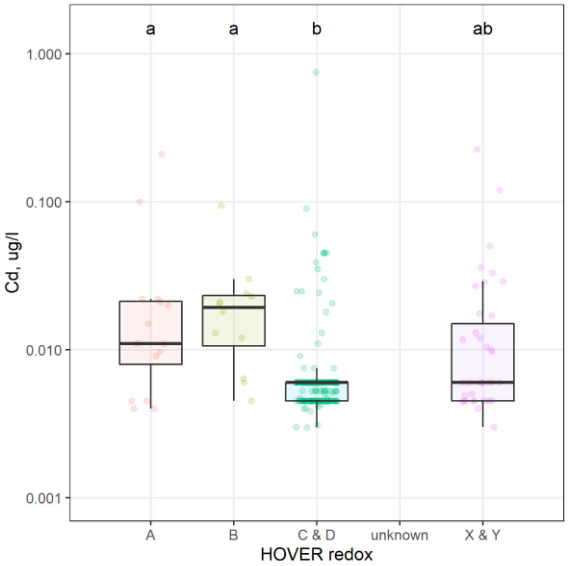
Elements	All	HOVER-redox categories										Kruskal-Wallis rank sum test
		A		B		C & D		Unknown		X		
	n	n	%	n	%	n	%	n	%	n	%	
SO4	6360	491	7.7	231	3.6	4824	75.8	10	0.16	804	12.6	Significant (p < 0.05)
As	6352	491	7.7	230	3.6	4820	75.9	8	0.13	803	12.6	Significant (p < 0.05)
Cd	356	16	4.5	12	3.4	289	81.2			39	11.0	Significant (p < 0.05)
Cr	250	11	4.4	7	2.8	204	81.6			28	11.2	Significant (p < 0.05)
Cu	289	15	5.2	7	2.4	237	82			30	10.4	Significant (p < 0.05)
Ni	6358	491	7.7	231	3.6	4822	75.8	11	0.17	803	12.6	Significant (p < 0.05)
Zn	363	15	4.1	12	3.3	297	81.8	1	0.28	38	10.5	Significant (p < 0.05)
Fl	6349	491	7.7	230	3.6	4821	75.9	2	0.03	805	12.7	Significant (p < 0.05)
Cl	6361	491	7.7	231	3.6	4824	75.8	10	0.16	805	12.7	Significant (p < 0.05)

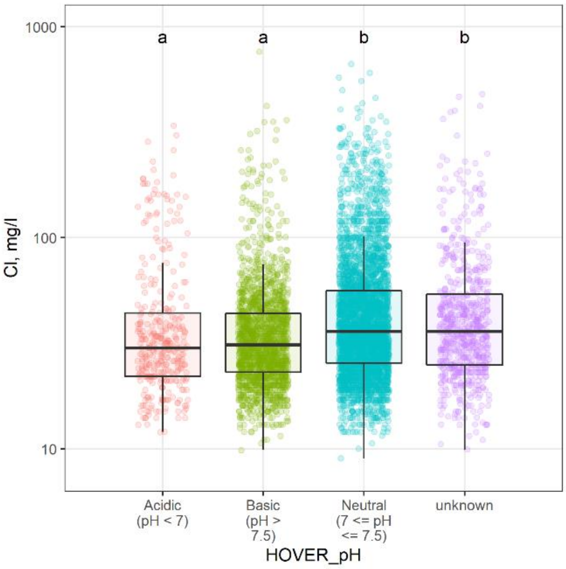
Figure 22 (next 3 pages) Comparison of concentration distributions for the HOVER-pH and HOVER-redox groups; for symbology explanation see Figure 21.

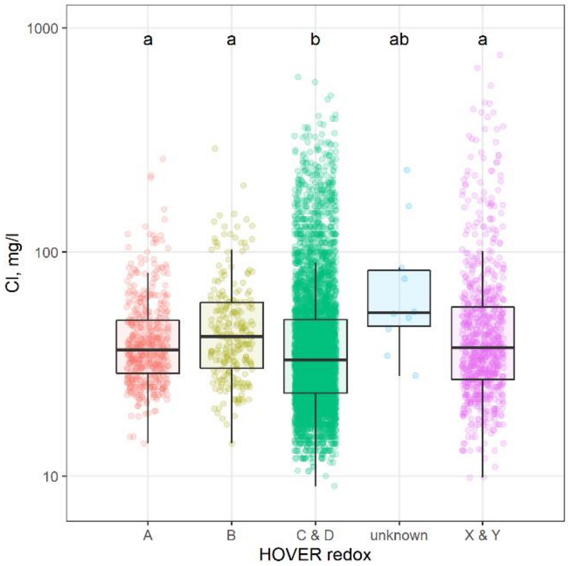

a
bc
b
c
50.0
10.0
1.0
0.1
As, ug/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

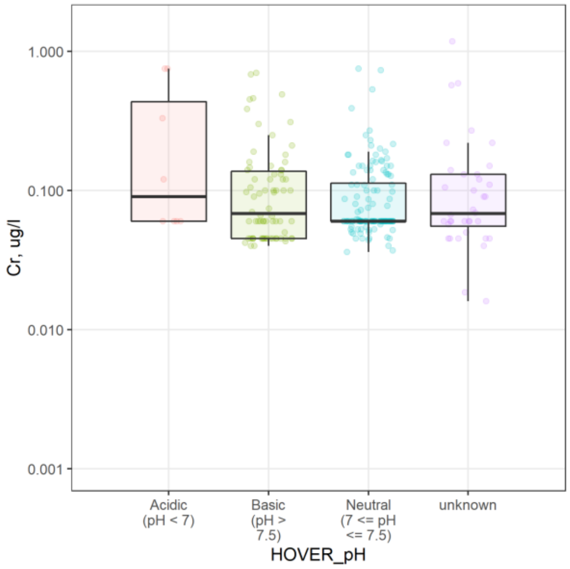

ab
a
a
ab
b
50.0
10.0
1.0
0.1
As, ug/l
A
B
C & D
unknown
X & Y
HOVER redox

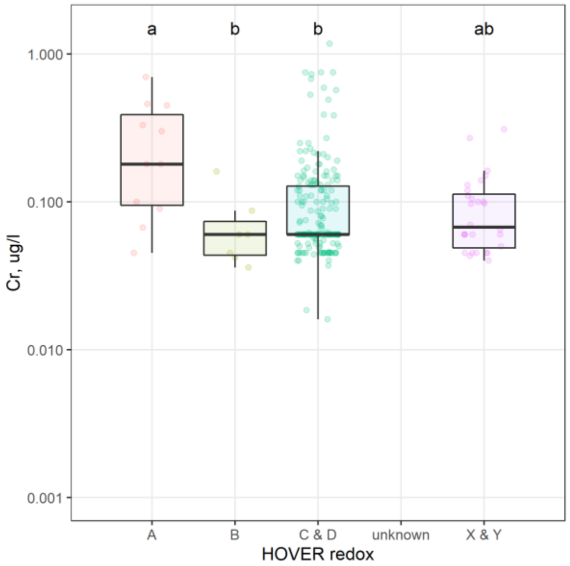

a
b
a
ab
1.000
0.100
0.010
0.001
Cd, ug/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

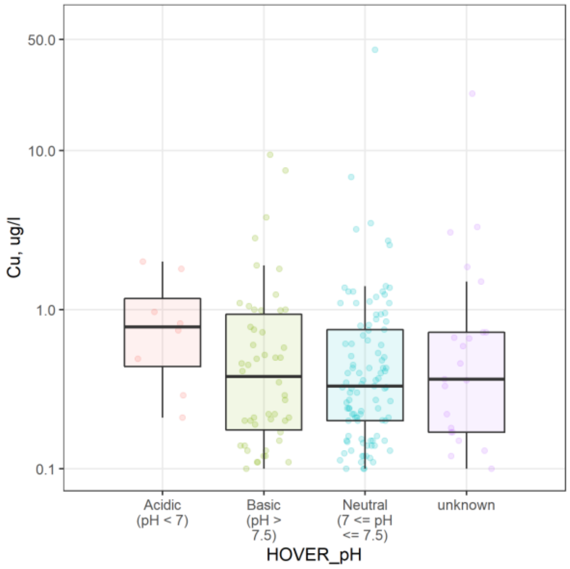

a
a
b
ab
1.000
0.100
0.010
0.001
Cd, ug/l
A
B
C & D
unknown
X & Y
HOVER redox

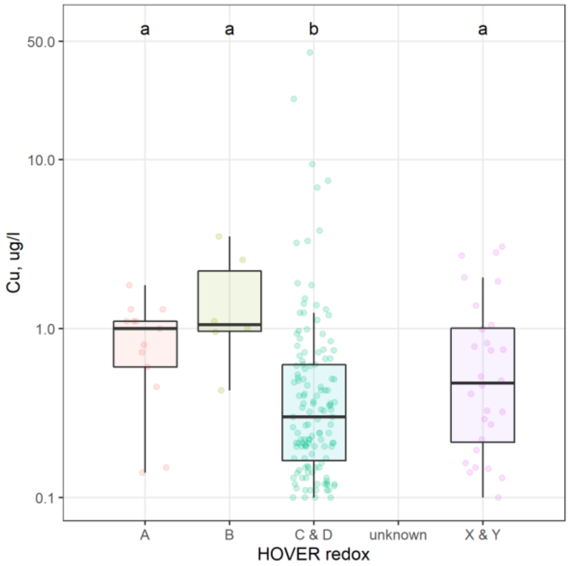

a
a
b
b
1000
100
10
Cl, mg/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

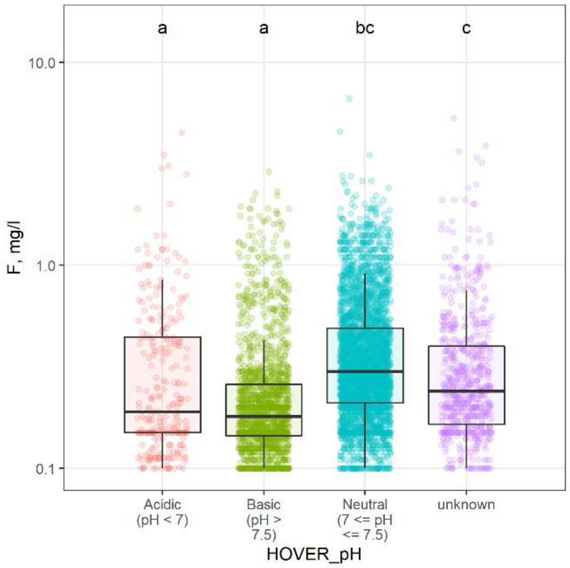

a
a
b
ab
a
1000
100
10
Cl, mg/l
A
B
C & D
unknown
X & Y
HOVER redox

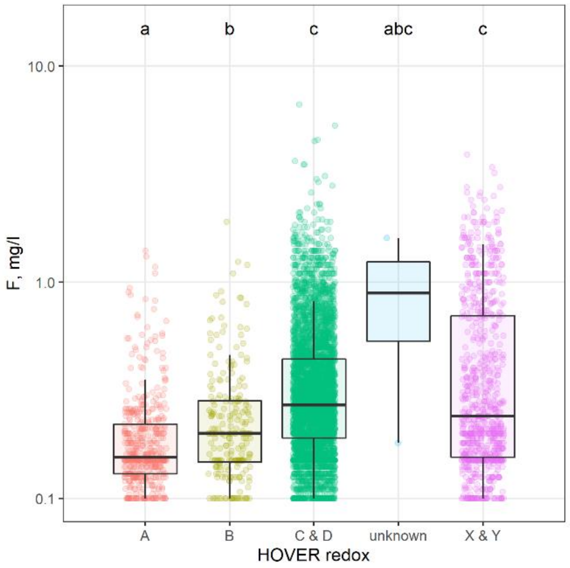

1.000
0.100
0.010
0.001
Cr, ug/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

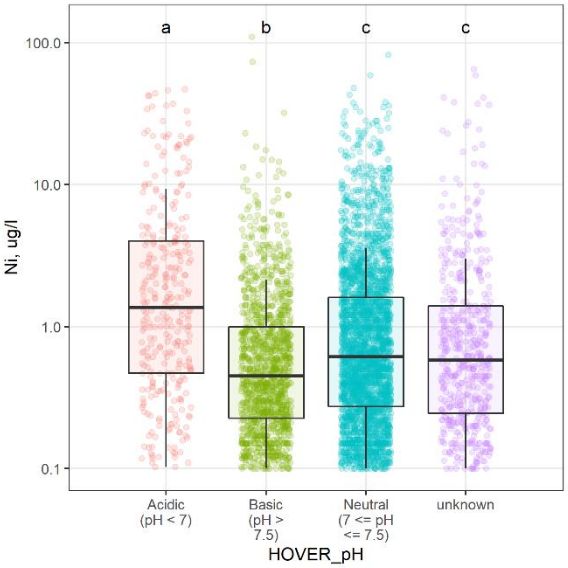

a
b
b
ab
1.000
0.100
0.010
0.001
Cr, ug/l
A
B
C & D
unknown
X & Y
HOVER redox

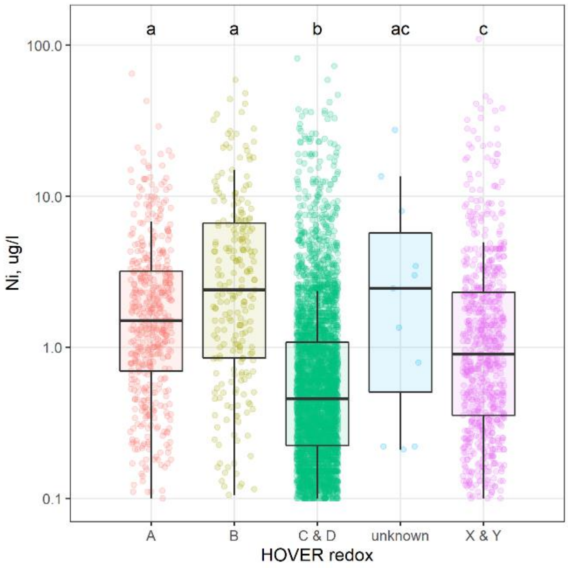

50.0
10.0
1.0
0.1
Cu, ug/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

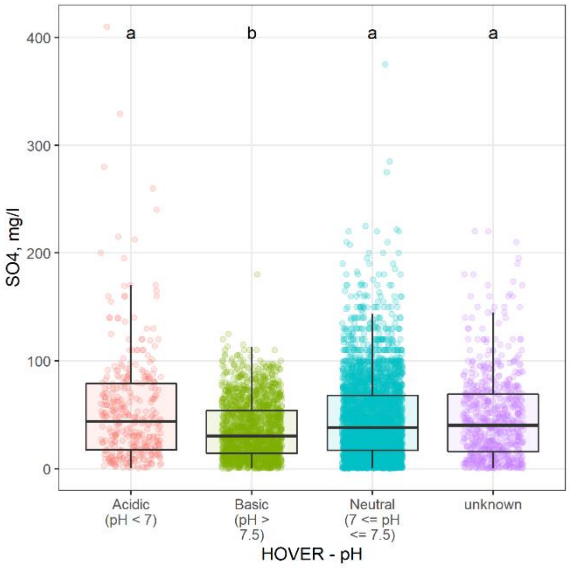

a
a
b
a
50.0
10.0
1.0
0.1
Cu, ug/l
A
B
C & D
unknown
X & Y
HOVER redox

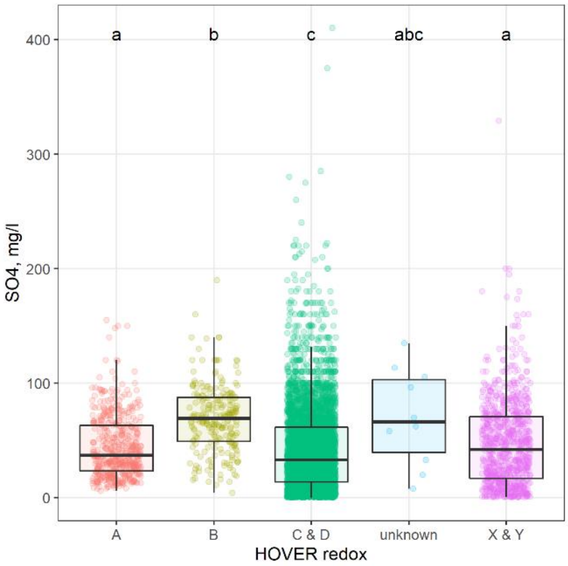

a
a
bc
c
10.0
1.0
0.1
F, mg/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

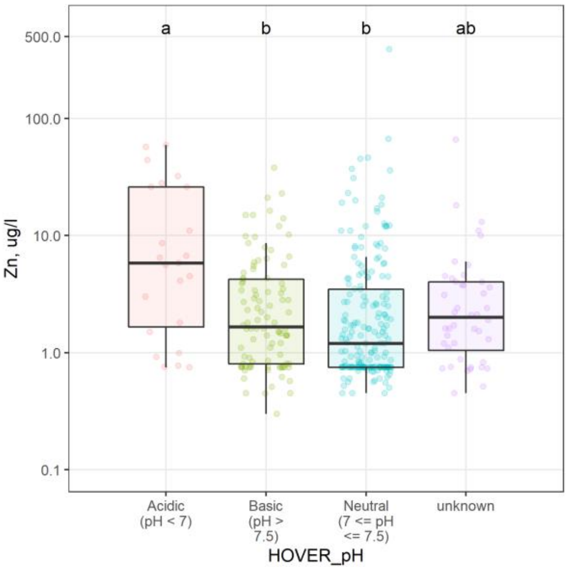

a
b
c
abc
c
10.0
1.0
0.1
F, mg/l
A
B
C & D
unknown
X & Y
HOVER redox


a
b
c
c
100.0
10.0
1.0
0.1
Ni, ug/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH


a
a
b
ac
c
100.0
10.0
1.0
0.1
Ni, ug/l
A
B
C & D
unknown
X & Y
HOVER redox


a
b
a
a
400
300
200
100
0
SO4, mg/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER - pH


a
b
c
abc
a
400
300
200
100
0
SO4, mg/l
A
B
C & D
unknown
X & Y
HOVER redox


a
b
b
ab
500.0
100.0
10.0
1.0
0.1
Zn, ug/l
Acidic (pH < 7)
Basic (pH > 7.5)
Neutral (7 <= pH <= 7.5)
unknown
HOVER_pH

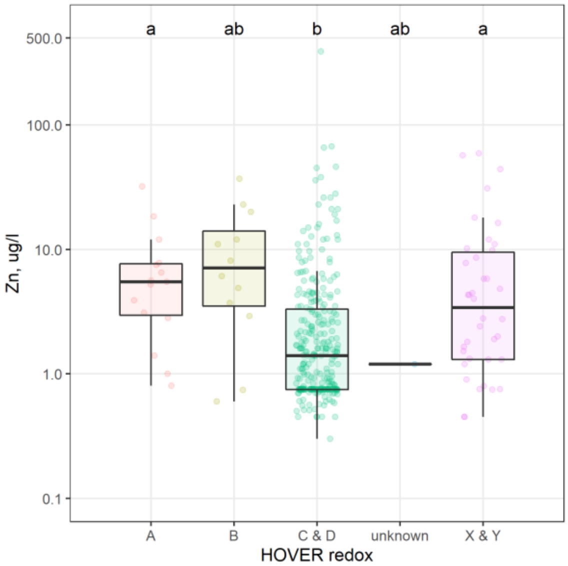

a
ab
b
ab
a
500.0
100.0
10.0
1.0
0.1
Zn, ug/l
A
B
C & D
unknown
X & Y
HOVER redox

Figure 22 results show that:

- HOVER-pH and HOVER-redox grouping cannot be used for Cd, Cr, Cu, and Zn because there is not enough data
- There is a significant difference between the acidic and basic groups for SO4, As, Cd, Ni; however, the neutral group is different from both acidic and basic groups only for Ni. For the rest of the elements, the neutral group is different from either the acidic or the basic groups.
- There is a significant difference between the reduced group (C &D) and the oxidised or the anoxic groups (A or B) for all elements, except for As. Additionally; there is significant difference between the oxic and anoxic groups for SO4 and F.

Therefore, it was decided to use the three HOVER-pH groups (acidic, basic, and neutral) and the four HOVER-redox groups (A, B, C&D, X) together with the lithology/geology when calculating the NBLs for SO4, As, Cd, and Ni.

For Cd, Cr, Cu, and Zn we will use only the HOVER-lithology and the DK-geological classes without taking into consideration redox and pH, because there is not enough data for this. After the initial 90th percentile calculation is done, we will evaluate if there is enough data in each combination group (lithology/geology + pH + redox).

5.2.6 *Additional removal of outliers -- element concentrations influenced by contamination*

After the removal of anthropogenic influenced water sampling points for each element, we identified some outliers, which may be influenced by contamination for Cd (n=1), Cu (n=2), SO4 (n=3), and Zn (n=1). These individual water sampling points are shown in the empirical cumulative distribution plots below (Figure 23).

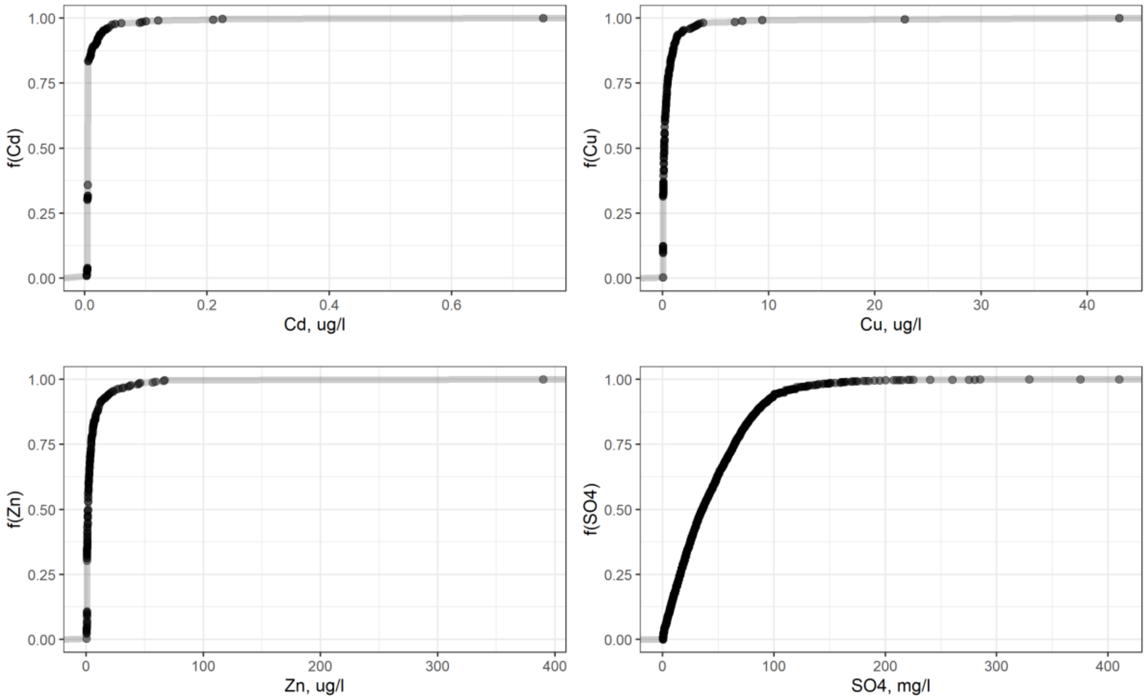


Figure 23 Empirical cumulative distribution plots for Cd, Cu, Zn, and SO4, showing the identified outliers.

We checked individually each of these sampling points to decide what action to take: delete the value, remove the water sampling point from the dataset, or ignore due to insufficient evidence. We used the online Jupiter portal in addition to the master dataset.

For Cd, we identified as an outlier the water-sampling point with concentration 0.75 µg/l (well with DGUNR 143. 63). This value is based on a single chemical analysis which was < LOD, thus this is not a pollution evidence, but the ½ LOQ substitute of a LOD = 0.5 µg/l. We are removing only the Cd value from the dataset (substitute with NA), because it is due to the high LOD. It must be noted that more than 80% of all individual analyses based on which we calculated the medians for each water sampling point were < LOD for Cd, so most of the values in this dataset are influenced by the ½ LOQ substitution.

For Cu, there are two water-sampling points identified as outliers:

- 43 µg/l (well with DGUNR 210. 655); this value is based on a single Cu analysis (2018), so it is not possible to assess Cu trends; the water sampling point is at 35-48 m dept (below terrain) in a quaternary sand (glacial sand) overlaid by 15.5 m of clayey till. It is with prevailing urban pressure, as it is in the town of Sorø (~8000 population), near Sorø lake. There is also agricultural pressure in the area. The same water point was also identified as an outlier for Zn with 390 µg/l, which is the highest concentration in the dataset. Thus, it was decided to remove the water sampling point from the dataset.
- 22.8 µg/l (well with DGUNR 9. 1150); the value is also based on a single Cu analysis (2018), so it is not possible to assess Cu trends; the water sampling point is at 26-32 m depth (below terrain) in a quaternary sand (glacial sand) overlaid by 10m of meltwater clay and 3.5m of meltwater silt. It is with prevailing agricultural pressure. The same water sampling point has also slightly elevated As concentration (5 µg/l) and Cr concentration (1.18 µg/l), so it was decided to remove the water sampling point from the dataset.

For Zn there was also one water sampling point identified as an outlier. This water sampling point was already discussed regarding Cu (well with DGUNR 210. 655) and it will be excluded from the final dataset.

For SO4, there were 3 water sampling points identified as possible outliers:

- 410 mg/l (well with DGUNR 212. 1029); the water sampling point is at 15-19 m depth (below terrain) in Paleogene chalk/limestone (Danian) overlaid by mostly sand, gravel, and stones (< 1m clay). The well was closed in 2018, so it is not used anymore for water supply. SO4 concentrations have been increasing steadily since year 2000. At the same time, a steady increase in Cl concentrations and a jump in the F concentrations was seen. It is with prevailing agricultural pressure, but also some urban pressure exists, as in the vicinity of the water sampling point is the small town of Lille-Skensved. It was decided to remove this water sampling point from the dataset.
- 375 mg/l (well with DGUNR 212. 471); the water sampling point is at 12-16 m depth (below terrain), also in Danian chalk and near by the Lille-Skensved. The well is also not used anymore for drinking water supply since 2018. SO4 concentrations have been at a similar level since 2003 (no earlier data). Additionally, some detergents have been detected in the groundwater. It was decided to remove this water sampling point from the dataset.
- 329 mg/l (well with DGUNR 201. 5311); the water sampling point is at 20.50-23.50 m depth in Danian calcarenite, protected by about 5m of clayey till. The well is in Frederiksberg (Copenhagen), with prevailing urban pressure, but industrial pressure also exists in the 1km buffer around the water sampling point. In 2016, MTBE, toluene, xylene, trichlorethylene and other organic micro-pollutants and halogenated compounds were detected as part of the municipality pollution control monitoring. The water sampling point will be removed from the

dataset. The same well has a deeper water sampling point, which will also be removed from the dataset.

To conclude, the following water sampling points will be removed due to pollution: DGUNR 201. 5311 (n=2), 212. 471 (n=1), 212. 1029 (n=1), 9. 1150 (n=1), 210. 655 (n=1). Additionally, we substitute the Cd concentration for 143. 63 (n=1) with NA (missing value).

5.2.7 *Calculating natural background levels with the final dataset*

Natural background levels (NBL) are determined as the 90th percentile of the median concentrations at water sampling points over the 10-year study period for a specific groundwater typology. The anthropogenically influenced water sampling points (element-specific, as described in section 0) and the outliers identified in section 5.2.6 were excluded prior to calculations. Descriptive statistics for each of the elements is presented in Table 109 (Appendix). Maps showing the spatial distribution of concentrations for each element and type of geology are also provided in the Appendix. The number of water sampling points in each geology/lithology subset is provided in Table 34. To assess the effects of the new method on the NBLs, we also calculated alternative NBLs without excluding the anthropogenic influenced points (results not shown here but see highlighted values in the following tables).

NBL for groundwater typologies based only on lithology/geology classification

The 90th percentiles were calculated first for the HOVER-Lithology and the DK-geology classes for all classes with 20 or more water sampling points (Table 27).

Table 34 Natural background levels (NBL) for the ***HOVER-lithology*** *and* ***DK-geology classes****; the number of water points on which the 90th percentile calculation for each class is provided as well (see “n”, with italic); the highlighting shows the effect of removing the anthropogenic influenced groups of water sampling points: green for increase in NBLs, yellow of decrease in NBLs; underlined NBLs are above the national threshold value*

		HOVER lithology				DK-geology						National threshold values[b]
		Sedimentary: carbonates[a]		Sedimentary: sand		Quaternary sand		pre-Quaternary sand		Bornholm		
		n	NBL	*n*	NBL	*n*	nbl	*n*	NBL	*n*	NBL	
As	µg/l	*1830*	3.9	*3639*	6.8	*3049*	7.7	*560*	2.5	*69*	1.3	5
Cd	µg/l	*77*	0.026	*259*	0.011	*197*	0.012	*56*	0.006	*<20*	-	0.5
Cr	µg/l	*55*	0.178	*186*	0.195	*155*	0.192	*31*	0.190	*<20*	-	25
Cu	µg/l	*67*	1.3	*218*	1.0	*182*	1.1	*36*	0.4	*<20*	-	100
Ni	µg/l	*1856*	5.3	*3661*	1.8	*3057*	1.9	*565*	1.4	*80*	1.9	10
Zn	µg/l	*85*	20	*271*	8.1	*200*	8.2	*56*	5.9	*<20*	-	100
SO4	mg/	*1834*	89	*3642*	87	*3052*	90	*560*	65	*69*	95	250[c]
F	mg/	*2213*	1.2	*4046*	0.4	*3356*	0.4	*660*	0.3	*69*	1.3	1.5[c]
Cl	mg/	*2242*	130	*4072*	70	*3368*	76	*665*	34	*80*	41	250[c]

[a] same as the DK-geology class “kalk” (chalk/limestone)

[b] these national threshold values are used in the River Basin Management Plan 2021-2027 for assessing the chemical status of groundwater bodies

[c] drinking water standard instead, as no national threshold value has been set

NBLs for the groundwater typologies combining redox and pH conditions with the lithology/geology

We calculated also the 90th percentiles for the groundwater typologies combining the HOVER-lithology/DK-geology with HOVER-pH and HOVER-redox. The NBLs were calculated only for typologies with 50 more water sampling points.

For the carbonates (chalk/limestone) there was enough data to calculate all NBLs, except for basic and anoxic (B). For the sandy aquifers, based on the DK-geology classification, there was enough data to calculate NBLs for all five elements for:

- Quaternary sand with:
    - Acidic and reduced (C&D) geochemical conditions
    - Neutral and oxic (A), reduced (C&D), and mixed redox types (X)
    - Basic and oxic (A), reduced (C&D), and mixed redox types (X)
- Pre-Quaternary sand
    - Reduced and acidic, neutral, and basic

Only for F and Cl, there was also enough data to calculate NBLs for Quaternary sand (DK-geology) and Sedimentary: sand for basic and anoxic conditions (B redox type). The NBL results for the groundwater typologies considering the geochemical conditions are presented in the following tables.

Table 35 Natural background levels for ***lithology/geology*** *combined with* ***redox and pH conditions****; bold values show classes where the geochemical class has higher NBL than the geological/lithological type (last column) and underlined NBLs are above the national threshold value. The highlighting shows the effect of removing the anthropogenically influenced groups of water sampling points: green for increase in NBLs, yellow of decrease in NBLs*

HOVER-lithology	Elements	units	HOVER-pH	HOVER-redox				NBLs[a] (only lithology)	National threshold values[b]
				A	B	C & D	X		
Sedimentary: carbonates	As	µg/l	Basic	2.2	-	**4.0**	3.3	3.9	5
			Neutral	2.0	2.7	**4.7**	**7.7**		
	Ni	µg/l	Basic	4.0	-	2.1	3.2	5.3	10
			Neutral	**12.0**	**20.0**	3.0	**7.7**		
	SO4	mg/l	Basic	43	-	63	48	89	250[c]
			Neutral	**92**	**100**	89	82		
	F	mg/l	Basic	0.3	-	0.9	**1.3**	1.2	1.5[c]
			Neutral	0.4	0.8	1.2	**1.7**		
	Cl	mg/l	Basic	55	-	64	96	130	250[c]
			Neutral	82	113	130	**203**		
Sedimentary: sand	As	µg/l	Acidic	-	-	3.0	-	6.8	5
			Basic	2.3	-	5.6	3.6		
			Neutral	1.3	-	**9.6**	3.8		
	Ni	µg/l	Acidic	-	-	**4.7**	-	1.8	10
			Basic	1.3	-	1.0	1.4		
			Neutral	**2.2**	-	1.3	**3.4**		
	SO4	mg/l	Acidic	-	-	87	-	87	250[c]
			Basic	68	**-**	74	86		
			Neutral	86	-	**89**	**111**		
	F	mg/l	Acidic	-	-	0.7	-	0.4	1.5[c]
			Basic	0.2	0.2	0.3	0.2		
			Neutral	0.3	-	0.4	0.3		
	Cl	mg/l	Acidic	-	-	46	-	70	250[c]
			Basic	60	69	58	60		
			Neutral	67	-	**78**	67		

[a] same as Table 36 (calculated without the HOVER-pH and HOVER-redox separation)

[b] these national threshold values are used in the River Basin Management Plan 2021-20272027 for assessing the chemical status of groundwater bodies

[c] drinking water standard instead, as no national threshold value has been set

Table 36 Natural background levels for the Quaternary and pre-Quaternary sands (DK-geology) combined with the HOVER-pH and HOVER-redox; bold values show classes where the geochemical class has higher NBL than the geological/lithological type (last column) and underlined NBLs are above the national threshold value. The highlighting shows the effect of removing the anthropogenically influenced groups of water sampling points: green for increase in NBLs, yellow of decrease in NBLs

DK-geology	Elements	units	HOVER-pH	HOVER-redox				**NBLs** [a] (only geology)	National threshold values[b]
				A	B	C & D	X		
Quaternary sand	As	µg/l	Acidic			3.3		7.7	5
			Basic	2.3		6.0	3.6		
			Neutral	1.3		**11.0**	3.9		
	Ni	µg/l	Acidic			**5.8**		1.9	10
			Basic	1.3		1.0	1.4		
			Neutral	**2.3**		1.4	**3.8**		
	SO4	mg/l	Acidic			**100**		90	250[c]
			Basic	68		76	86		
			Neutral	89		**94**	**115**		
	F	mg/l	Acidic			**0.8**		0.4	1.5[c]
			Basic	0.2	0.2	0.3	0.2		
			Neutral	0.3		0.4	0.3		
	Cl	mg/l	Acidic			**140**		76	250[c]
			Basic	60	69	62	60		
			Neutral	67		**87**	68		
Pre-Quaternary sand	As	µg/l	Acidic	-	-	1.5	-	2.5	5
			Basic	-	-	1.9	-		
			Neutral	-	-	**3.2**	-		
	Ni	µg/l	Acidic	-	-	**3.5**	-	1.4	10
			Basic	-	-	0.6	-		
			Neutral	-	-	1.1	-		
	SO4	mg/l	Acidic	-	-	**84**	-	65	250[c]
			Basic	-	-	46	-		
			Neutral	-	-	62	-		
	F	mg/l	Acidic	-	-	**0.5**	-	0.3	1.5[c]
			Basic	-	-	0.2	-		
			Neutral	-	-	0.3	-		
	Cl	mg/l	Acidic	-	-	32	-	34	250[c]
			Basic	-	-	32	-		
			Neutral	-	-	32	-		

[a] same as in Table 27 (calculated without the HOVER-pH and HOVER-redox separation)

[b] these national threshold values are used in the River Basin Management Plan 2021-20272027 for assessing the chemical status of groundwater bodies

[c] drinking water standard instead, as no national threshold value has been set

5.3 Comparison and contributions of the proposed method

The contribution of this land-use-based method for assessing anthropogenic influences is that it could identify statistical difference between water sampling points with prevailing agricultural pressure and those with industrial and/or urban pressures. In the Danish context, however, less than 1% of the water sampling points were without any anthropogenic pressure (i.e. natural) and around 80% are with prevailing agricultural pressure. This limits the application of the method.

There are two consequences of treating all water sampling points with similar prevailing anthropogenic pressure as a group (and excluding them as a group for specific elements). First, the method does not distinguish between polluted and potentially polluted water sampling points, i.e. it does not assess within-group differences. This is an especially important consequence for the Danish dataset, where the “agriculture” dominates the dataset and had to be used as the “normal state”. Thus, about 80% of the dataset could not be assessed for potential pollution. Second, because entire groups of data were excluded, irrespective of the actual observed concentrations, some of the NBLs increased after excluding the industrial and/or urban groups (those are highlighted with green in Table 34,

Table 35, and Table 36). By excluding entire groups of data, we also excluded a substantial amount of water sampling points with low concentrations (potentially the natural background), thus the resulting NBLs increased. It is of course debatable, if the observed increases are meaningful, as the precision of the chemical analyses should also be considered.

In conclusion, the proposed method could be used as an exploratory step. However, this analysis should also be supplemented with other more specific methods, that allow for distinguishing polluted or potentially polluted sites. In the Danish context, all groundwater wells are classified based on their use, so it is possible to select only the least affected group (the waterworks wells used for drinking water). However, there could still be sites, where there is anthropogenic pollution, which we did not account for. Thus, it is still valuable to identify potentially polluted water sampling points with the proposed method, but we should also supplement it with other methods in future. Examples of supplementary methods, which could be used together with the land-use method:

- Using pesticides, organic micropollutants, chlorinated solvents etc. as pollution indicators, to exclude polluted water sampling points. Such assessments should be done for each anthropogenic pressure group separately, e.g. for the agriculture group, pesticides would be a logical pollution indicator, while for the industrial and urban groups, we could select some of the most common organic micropollutants instead.
- For Cl and SO4, and other elements that are influenced by unsustainable pumping, it may be also beneficial to look at trends (if possible), so water sampling points with significant positive trends could be excluded from the dataset.
- Hinsby et al. (2008) differentiated between oxidised and reduced groundwaters and suggested that for oxidised groundwaters, water sampling points with median NO3 > 10 mg/l should be considered as polluted and excluded from NBL derivations. For reduced (anaerobic) groundwaters, SO4, the oxidation capacity, K, sum of cations, and the N2 gas pressure could be used as pollution indicators. As these also occur naturally, there should also be sound background understanding.

If groundwater age estimation is possible, it may be also used for identifying old, uninfluenced by modern anthropogenic pollution, water sampling points.

Lastly, on such national scale, data-analytical methods might also be accompanied by process-oriented analyses and supported by expert background knowledge at the local scale.

6 CASE STUDY – FRANCE

6.1 Previous studies on NBLs

In France (Metropolitan France), the first studies were carried out between 2004 and 2007 (Brenot et al., 2007). This work aims to delimit the zones exposed to a high risk level for natural background level with a degree of confidence given for each of these sectors and parameters. For this approach, in addition to the bibliographic synthesis, a statistical study on the chemical correlation of water and the type of lithology of aquifers were carried out on the major elements. However, the application of the statistical treatments to trace elements could not be carried out due to the lack of data and analytical methods with numerous high limits of quantification (LOQ). Thus, the studies consisted on the delimitation of high hydrogeochemical backgrounds and the determination of major reference concentrations. Sectors with exceptionally high values in trace elements were not validated due to the lack of data.

Since 2007, WFD monitoring networks have been set up in all French basins. In consequence, a lot of new data are available to complete the first work, in particular to define the concentration of trace elements such as: The investigated elements were : Al, As, Ba, B, Cd, Cr, Cu, F, Fe, Hg, Mn, Ni, Ptotal, Pb, Sb, Se, Zn, NH4, Cl, NO3, SO4. Natural background concentrations were calculated according to the aquifer typologies excluding geochemical anomalies or anthropic inputs. This approach was applied in recent studies, at different scales (basin, regional, local) for groundwater and surface water: in French Guyana (Lions et al., 2015), Martinique (Arnaud et al., 2013), Guadeloupe (Ratsimihara et al., 2014), Belfort (Doney et al. 2015), Loire-Bretagne (Devau et al, 2017), and France (Lions et al., 2016).

6.2 Study on Loire Bretagne Basin

The objective is to be able to define the ranges of natural concentrations of one element in groundwater, i.e. acquired entirely from a natural source, whether geological, biological or atmospheric, without the influence of human activities.

In the context of the present study, we have attempted to characterize natural concentrations by working at a water basin scale (Loire Bretagne) considering three geological regions (sedimentary, volcanic, bedrock). The chosen approach, adapted to the working scale, is a global statistical approach to calculate the reference concentrations by lithology/geology family or hydrogeochemical entity.

In France aquifers are distinguished in 5 geological types according to the BDLisa (ref):

- Sedimentary (Aquitanian Bassin, Paris Bassin) ;
- Basment (Armorican Massif, Massif Central) ;
- Alluvial;
- Volcanic ;
- Intensively folded (Mountains)

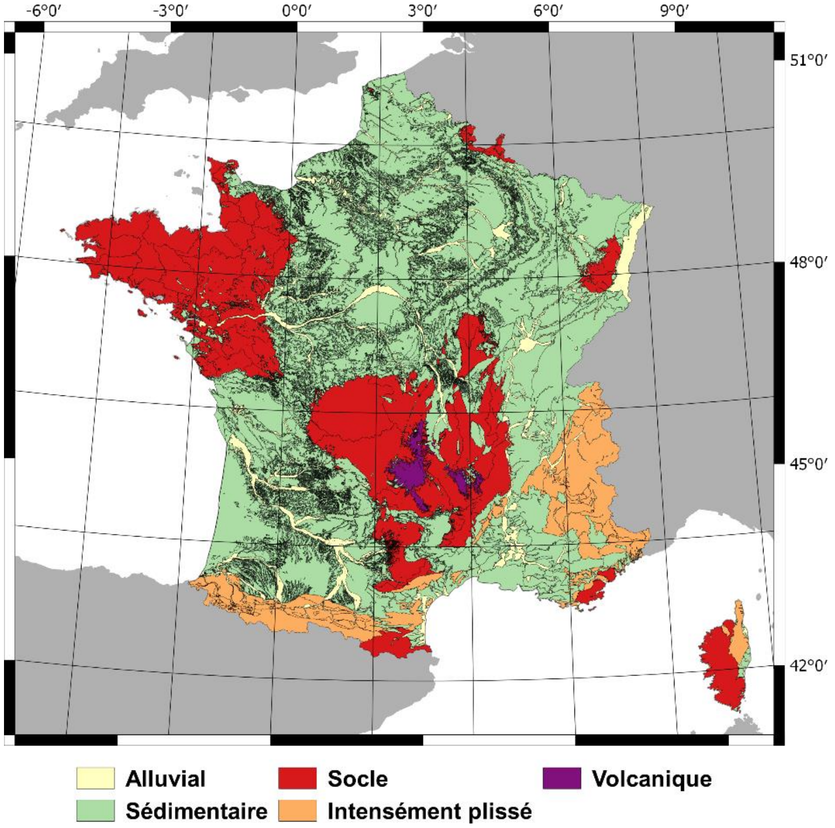


Figure 24 : Mapping of aquifers by geological formation typology (BD Lisa)

The method developed throughout the Loire-Bretagne basin took place in several stages:

The inventory of available and exploitable data for the study, as well as the choice of the priority or relevant factors and contexts for the characterization of the natural background is a preliminary step to select the representative points and to evaluate the natural background according to land use, natural geochemical background, anthropic activities (ICPE, BASIAS, etc.) and diffuse pollution.

For each element, the reference concentrations for a hydrogeochemical entity have been obtained by calculating the 90th percentile from the element concentrations measured within each entity. In each entity, the water points whose element concentration is influenced by a diffuse or point source of contamination are removed from the data set before carrying out the calculation of the 90 percentile. So that each water point present within an entity of hydrogeochemical background has the same weight in the calculation of the 90th percentile. For each point, the data used are the median of the concentrations monitored at the water sampling point. The hydrogeochemical entities are determined by grouping the lithological entities for which the element concentration distributions are identical (statistical treatment).

6.2.1 *Constitution of the groundwater quality dataset*

The river basin is composed by three geological domains, namely the Massif Central, the south of the Paris Basin and the Armorican Massif. Each domain has been treated separetly to optimize the statistics approach and NBLs are defined in each domain according to the main lithology:

- Basment (crystalline and metamrophic) for the Armorcan Massif and Massif Central;

- Sedimentary rocks for the Paris Basin
- Volcanic rosk for the Massif Central

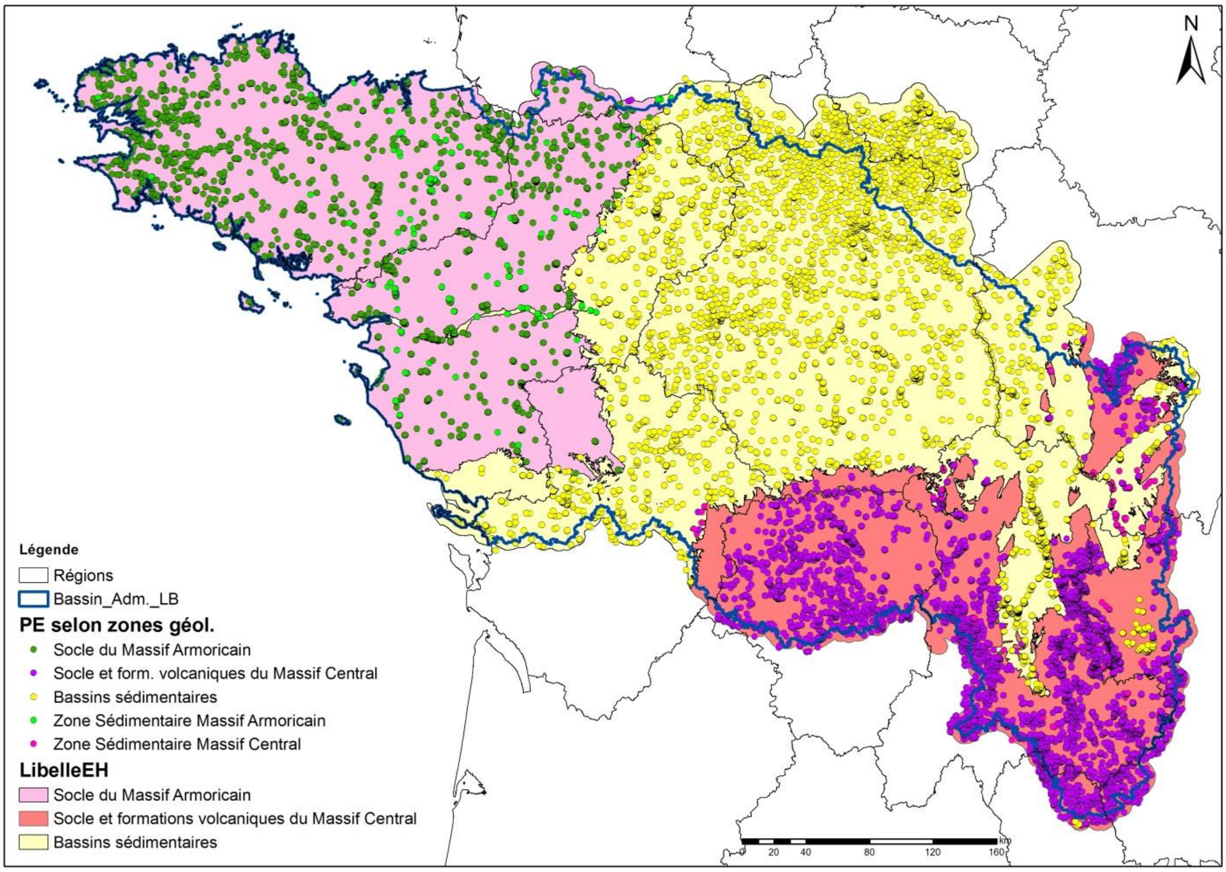


Figure 25: Map of the three geological domains of the Loire-Bretagne basin

The method is applied to the dataset available in ADES national portal (https://ades.eaufrance.fr/), which represents a national large dataset.

The first step of the analysis is the inventory of data available and usable for the statistical analyses. In fact, the available data are not acquired with the objective to characterize the natural background level but to monitor the groundwater quality as requested by the regulation (WFD, others…). Therefor several key-points have to be solved:

- The water points are not necessarily chosen to be representative of a natural context. Thus, it is necessary to select the most suitable water points given the information available.
- The quantification limits are not necessarily optimized in relation to the values of the natural background, they are more generally selected with respect to drinking water standards.
- The attachment of points to hydrogeological entities is not completed.

The approach adopted is a global approach by hydrogeochemical entity that allows extrapolating the results to groundwaters with comparable lithology not represented in the dataset.

The work provided here is a first step making it possible to define reference concentrations that will have to be changed based on experience feedback, the acquisition of new data, both qualitative and hydrogeological.

Data checking and correction

The qualitative data used are all the data extracted from the ADES database in October 2016 (excluding mining and industrial monitoring networks).

All the dataset has been checked in order to correct any data errors, including unit errors (factor x1000 µg or mg) common when banking trace element concentrations.

When the concentrations measured for a parameter are below the limit of quantification (<LOQ), the value of the concentration taken into account is the limit of quantification divided by two (LOQ / 2).

Data with LOQs too high for the definition of natural background level were removed from the dataset. For each element, a filter was applied in order to optimize the LOQ used for the calculations according to the number of data and the expected natural background level. This step follows on from the preliminary work carried out in 2014 and 2015 (Gourcy et al., 2015; Lions et al. 2016). The cut-off threshold was optimized according to the concentration levels and the distribution of the data according to the histograms.

According to the maximum LOQ / 2 retained, the distribution of the data is refined, as well as the calculation of the median and the percentiles which remain equal to the LOQ, which is lower. This makes it possible to offer lower concentrations and therefore closer to the real values. In this case, the 90 percentile is also lowered, which is very likely since data with an LOQ much greater than the natural background level disturbs the calculated references concentrations.

For large datasets, this approach is possible. For smaller datasets, this approach reduces the number of data and may overestimate calculated values greater than the LOQ, due to a lower number of <LOQ values. It is thus possible to overestimate values such as Q3 (2/3 or rank 33) and the 90th percentile in the event of extreme values in the data set.

Table 37: LOQ/2 selected for the present study

Parameter	LOQ/2 maximum (µg/l)
As	1
Ni	2
Cd	0,5
Cr	2
Cu	10
F	25
Zn	2

Table 38 : number of sampling points and chemical analyses for each element, account of value > LOQs in the dataset, range of LOQs, percentage of analyses < LOQ.

Element	No sampling points	No of analyses > LOQ	No of analyses	low LOQ	high LOQ	< LOQ (%)
As (µg/l)	3512	7283	16041	0.005	1.5	55%
Cd (µg/l)	3708	3187	3656	0.5	0.5	13%
Cr (µg/l)	426	5453	9928	0	1.25	45%
Cu (µg/l)	435	2920	17373	0.005	1	**83%**
Fe (µg/l)	3755	11971	26361	0	50	**55%**
Ni (µg/l)	2362	1694	4111	0.025	1	59%
Zn (µg/l)	495	3225	4378	0.05	0.5	26%
F (mg/l)	3313	7527	10316	0.005	5	27%
Cl (mg/l)	3819	29916	30545	0.5	2.5	2%
SO4 (mg/l)	3783	29021	29998	0.05	2.5	3%
NO3 (mg/l)	3893	43142	48554	0.05	2.5	11%

Mapping

The mapping of the dataset allows a regional approach of the natural background level, in particular for the bedrock regions. In sedimentary zones, the projection of data on a map does not make it possible to distinguish the waters of overlaid aquifers, which in many cases have different lithologies and hydrochemical facies.

This approach nevertheless makes it possible to distinguish regional variation for each element studied on a regional scale. To produce these maps (Illustration 7), 4 classes of concentrations were defined as follows:

- Class 1: below or close concentration of LOQ
- Class 2: average or median concentrations in natural waters (bulk dataset)
- Class 3: concentrations greater than the usual concentrations (Lions et al. 2016)
- Class 4: abnormal or higher concentrations to drinking water standards or Environmental Quality standard (EQS)

*Table 39: Parameter concentration classes (** EQS, * DWS)*

	classe1		**classe2**		**classe3**		**classe 4**	
As	>LQ	1	1	4,2	4,2	10*	> 10*	µg/l
Cd	>LQ	1	1	2,5	2,5	5*	> 5*	µg/l
Cl	>LQ	5	5	25	25	200	> 200	mg/l
Cr	>LQ	1	1	3,4**	3,4**	10	> 10	µg/l
Cu	>LQ	1,4	1,4	10	10	100	> 100	µg/l
F	>LQ	100	100	250	250	500	> 500	µg/l
Ni	>LQ	1	1	4**	4**	10	> 10	µg/l
SO4	>LQ	10	10	25	25	100	> 100	mg/l
Zn	>LQ	10	10	50	50	200	> 200	µg/l

These classes could adapted for each dataset, with the objective to highlight area with abnormal concentrations or exceed of DWS. The number of classes could reduce to 3 (below 1 µg/l for trace, average data, above DWS).

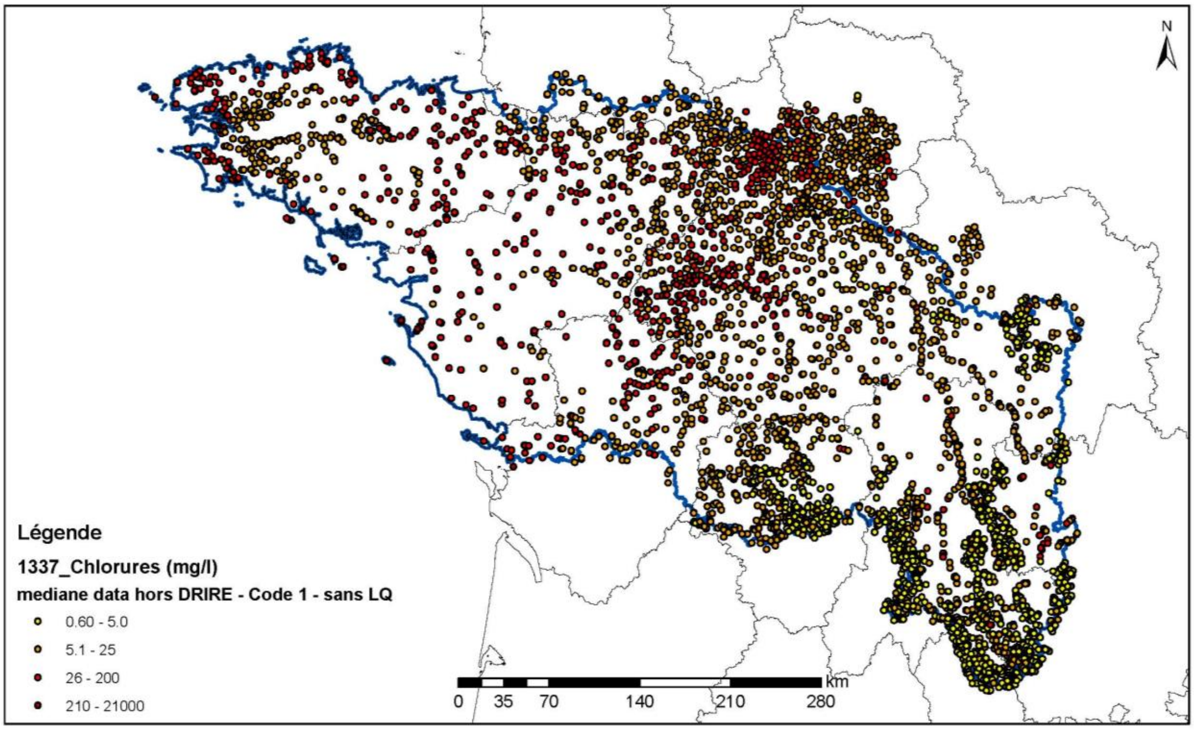


Figure 26: Chloride median concentrations (mg/l) for each water point

6.2.2 *Hydrogeological characteristics of the sampling points*

In order to identify the hydrogeological entities investigated through the dataset, it is necessary to know which entity is monitored for each water points. These hydrogeological entities are delimited in France by the BDLISA entities (Brugeron et al. 2018) or groundwater bodies. Then, it is possible to affect to each water point from aquifer monitored.

This step was carried out, for the sedimentary aquifers, based on the national project for priority boreholes or if necessary completed for other identified points. In the basement zone, the targeted entities correspond rather to lithological / typological groups.

All the information available for each of the water point has been banked in a database. The water points used for statistical processing have been systematically attached to the defined entities.

Properties of water points and geological data

Based on the selection of priority or relevant factors and contexts for the characterization of the natural background, the data collected come from several national databases:

- BSS WATER for information on water points and their connected groundwater,
- ADES for water quality,
- SIGMines for the mining inventory and the location of deposits and mining indices.

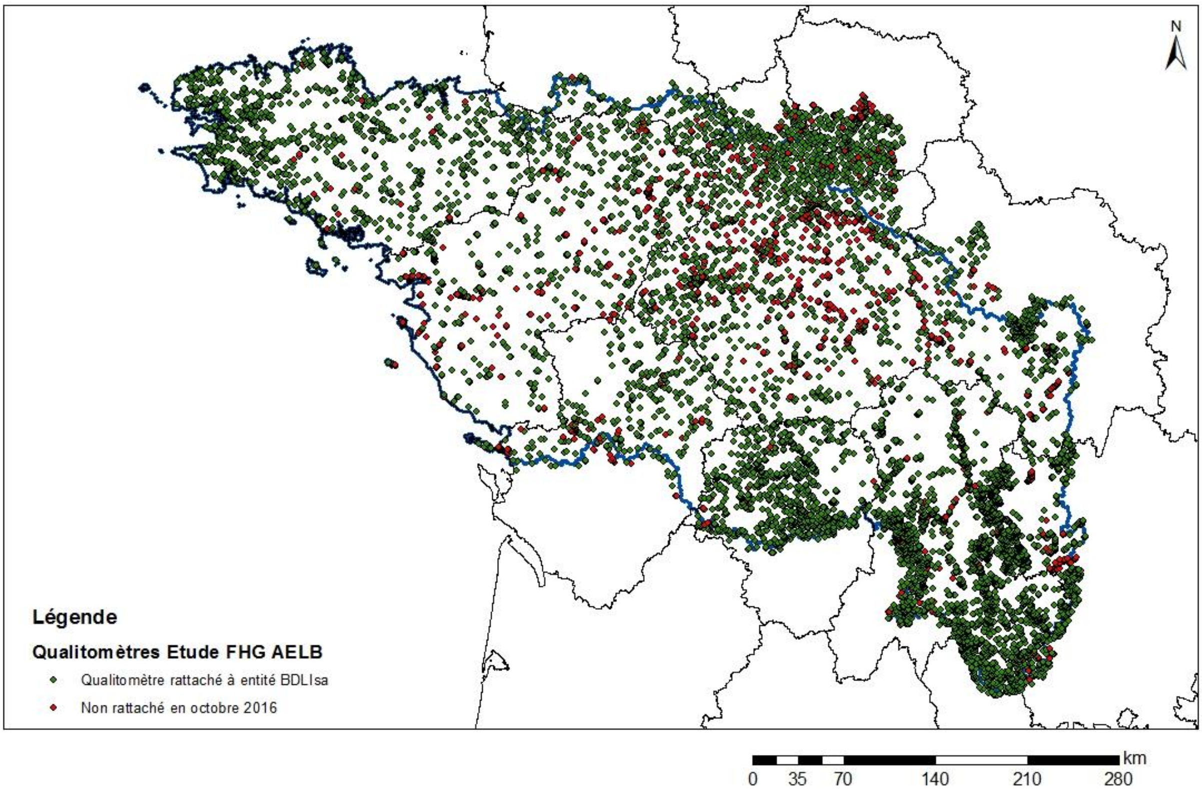


Figure 27: Location of 11,064 water points attached in October 2016 out of 14,041 water sampling point (qualitometer) identified in the basin as part of the study

To help define hydrogeochemical entities, the mine inventory and geological map (1: 50,000 or 1: 1,000,000) were used.

Delimitation of watersheds

The data were related to hydrogeochemical units, functional units or watershed (BV).

They are defined as watershed units divided longitudinally by the main river and then, into sub-units integrating one or more small tributaries located on the same bank. This work unit was used only to determine the anthropic pressures exerted at the surface of each observation points.

This division is interesting for our purposes because it corresponds to fairly small local units and, above all, their geometry is compatible with the methodology for taking samples from the mining inventory.

All the watersheds and sub-watersheds of the territory have been calculated using the “TAUDEM” module, which is an “Add-in” of ArcGis using a Digital Terrain Model (DEM) produced by the IGN. This DEM has a spatial resolution of 25 m.

The surface of the elementary watersheds (head of the basin) was set at 20km^2, ie 8,000 meshes. The territory was divided into 36,453 basins and sub-basins. These basins are hierarchical, that is to say that for each basin we know the basins located upstream and the one located downstream.

Each data (water point and pollution sources such as discharge points, industrial sites, mining index) has been linked to a watershed. At the level of aquifer monitoring water points, on the other hand, there is no concept of upstream or downstream. In addition, pollution sources can be linked on a surface watershed relatively far from the monitoring point but corresponding to the same underground watershed, i.e. having the potential to modify the quality of the underground water at a point of surveillance located in another watershed.

Surface data such as non-point sources or diffuse pollution per municipality or Corine Land Cover were linked to watersheds.

6.2.3 *Anthropogenic pressures / influences : diffuse or point pollutions*

Diffuse anthropogenic pressures have been determined based on Corine Land Cover dataset (2012) and agricultural information available.

Anthropogenic pressures and point sources (WWTP, facilities, land disposal site) were extracted from national databases or provided by the Water Agency. Point sources are known to be potential source of pollution that can reach the water table and influence the quality of groundwater.

Diffuse pollution

Information characterizing agricultural pressure and organic and phosphate fertilization pressure are from the following sources:

- Agreste - Database of the number and type of animals per municipality (Agricultural census 2000 and 2010)
- DISAR 2010 - Rate of manure and phosphorus per herd (Interactive Diffusion of Reference Agricultural Statistics)
- Agreste - Crop database by municipality (2011 crop practices survey - Main results.)
- SOGREAH (2007) - Treated surfaces and applied doses (“Balance sheet of contaminant flows entering agricultural soils in metropolitan France - Qualitative assessment of contamination by traced metallic elements and organic compounds and quantitative application for traced metallic elements ")

These data are defined areas (polygons) or identified municipalities. The calculation applied help to define for all the agricultural holdings present within a municipality:

- the consumption volumes of phosphate fertilizers used according to the crops,
- the volumes of organic fertilizer used,
- manure volumes.

Based on a rate of trace metal elements (TME) by type of manure, a volume of TME per municipality was calculated. For example, the volume of copper and zinc per municipality are calculated using levels of trace elements by type of crop.

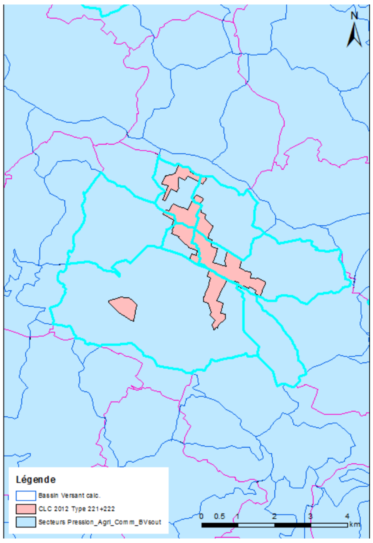


Figure 28: Geographical crossing between the CLC sectors (orange background), the calculated watersheds (blue lines) and the municipalities (fuschia pink lines).

Pressures and extension calculated by watersheds

The environment of each point was determined from existing databases (CORINE Land Cover, BASIAS, pollutant emissions register, ICPE, STEU, BASOL, RSDE) in order to define any pressures exerted.

Each point source such as WWTP discharge points, industrial sites (BASOL, BASIAS, ICPE), pollutant emissions, companies and basin treatment plants (RSDE) were linked to a calculated watershed.

Then, the presence or absence of one of these anthropic pressures attached to a calculated watershed were linked to the water points located in their respective watersheds.

Non-point sources are represented by polygons (agricultural activity by municipality or Corine Land Cover) and they are overlaid on the extension of watersheds and municipalities. If two agricultural municipalities are present in the same watershed, the highest ETM values per municipality are used for the entire basin. This maximum value was attached to the water points located in the respective watershed (Figure 28).

Table 40 : Number of water sampling points with data for each element and each group of prevailing pressure.

Element	No sampling points	agriculture (CLC)		natural area (CLC)		urban area (CLC)		industrial point source (BASIAS)		Mining and mineralization	
As	3512	2007	57%	1303	37%	202	6%	993	28%	442	13%
Zn	495	353	71%	93	19%	49	10%	182	37%	54	11%
Ni	2362	1413	60%	823	35%	126	5%	766	32%	458	19%
Cd	3708	2111	57%	1394	38%	203	5%	1056	28%	458	12%
Cr	426	294	69%	88	21%	44	10%	155	36%	59	14%
Cu	435	304	70%	86	20%	45	10%	159	37%	54	12%
F	3313	1953	59%	1160	35%	200	6%	974	29%	367	11%
Cl	3819	2171	57%	1444	38%	204	5%	1092	29%	480	13%
SO4	3783	2152	57%	1427	38%	204	5%	1081	29%	470	12%

6.2.4 *Removal of element concentrations influenced by contamination and NBL calculation*

For each element, the first step was to remove the water points for which the element concentration is significantly influenced by a source of contamination. This work was carried out on the basis of statistical processing. Thus, for each element, it was possible to define whether the concentration within a water point is influenced by: i) diffuse agricultural and / or urban pressure, ii) organic and / or phosphate fertilization pressure or iii) pressure linked to the presence of a site that could be a point source of contamination within the watershed.

Data treatment

In order to characterize anthropogenic non-point source or point source, three sources of information were used: i) the Corine Land Cover to characterize the diffuse agricultural and urban pressure, ii) the typology of farms inventoried at scale communal to characterize the organic and phosphate fertilization pressure and iii) the BASIAS, BASOL and ICPE databases to localize the industrial pressure. Each of these sources of information was used to construct discrete explanatory variables allowing a categorical property to be assigned to the measurements of major, minor and trace element concentrations extracted from ADES.

A more detailed analysis was carried out to characterize the impact of agricultural activities on the modification of the concentration of major, minor and trace elements in groundwater. The Kruskal-Wallis test was used to determine whether the concentrations of major, minor or trace elements are different between water sampling points under the influence of fertilization pressure (phosphates and / or organic) compared to other points of water. To do this, information on the typology of agricultural holdings has been defined at the commune level. Its application made it possible to calculate the production volumes of organic fertilizers from livestock for all the municipalities present in the Loire-Bretagne region. This methodology also makes it possible to define the consumption volumes of phosphate fertilizers used by all the farms present in a municipality.

It was thus possible to calculate, for each of the municipalities in the Loire-Bretagne territory, an organic and phosphate fertilization pressure. The watersheds were grouped into two groups, those with an organic fertilizer production volume below the median and those with a production volume above the median. A similar approach has been put in place to transform the quantitative values for the use of phosphate fertilizers. These two discrete explanatory variables were combined in order to define an overall fertilization pressure resulting from both the application of organic fertilizers and the application of phosphate fertilizers. This information defined at the watershed level was linked to the water sampling points.

Similar to the methodology developed to study anthropogenic non-point source, a nonparametric hierarchical univariate statistical treatment (Kruskal-Wallis test) was used to determine whether, water points on the influence of anthropogenic point source exhibit concentrations of elements major, minor or traces different from those of other water points. It was thus possible to define a discrete explanatory variable comprising three categories defining whether a water sampling point is influenced by at least one ICPE site, or by at least one site referenced in BASIAS outside the ICPE site or if the water sampling point is not subjected to anthropogenic pressure.

Selected dataset

Univariate statistical treatments were used to determine whether the distribution of concentrations between these categories differed significantly or not. Statistical processing has been performed independently for each of the discrete variables (Devau, 2017).

- For the Massif Central, the elements Cd, Cl, F, and SO4 have significantly higher concentrations for the water sampling points representing urban areas. Cu and Ni present the highest concentrations in agricultural areas. In addition, F is also higher in water point influence by anthropic activities.
- For the Paris Basin, As, Cl, F, Ni and Zn concentrations at water sampling points belonging to urban area are significantly different from those measured in at least one other area. In agricultural areas, As, Ni, and Zn stand out. In addition, As is also influenced by anthropic activities.
- In the Armorican Massif, the urban areas are characterized by higher concentrations of Cl, F, and SO4.

For each of these anthropogenic sources, the tables presented below summarize the choices for non-point sources (Table 41 : Major, minor and trace elements influenced by a anthropogenic non-point sources of urban and / or agricultural origin for the three areas.) and anthropic point source (Table 42).

	Areas	As	Cd	Cr	Cu	Ni	Zn	F	SO4
Massif Central	Agricultural				x	x			
	Urban	x						x	x
Paris Basin	Agricultural			x				x	x
	Urban			x			x	x	x
Armorican Massif	Agricultural								
	Urban							x	x

Table 41 : Major, minor and trace elements influenced by a anthropogenic non-point sources of urban and / or agricultural origin for the three areas.

	Areas	As	Cd	Cr	Cu	Ni	Zn	F	SO4
Massif Central	Anthropic						x		x
	Industrial								x
Paris Basin	Anthropic	x					x		
	Industrial								x
Armorican Massif	Anthropic								
	Industrial								

Table 42 : Major, minor and trace elements locally influenced by anthropogenic pressure for the three areas.

Mining area

In the case of the Massif Central and Massif Armoricain domains, mining index readings were also used to sort the concentration data extracted from the ADES database. Each of the information sources (lithology, mining indices) was used to construct discrete explanatory variables allowing a categorical property to be attributed to the measurements of concentrations of major, minor and trace elements. Then hierarchical univariate and bivariate statistical treatments were used to determine whether the distribution of concentrations between these categories differed significantly or not. Statistical processing was performed independently for each of the discrete variables.

This work made it possible to calculate an average concentration of a few trace elements (Pb, As, Cu, Zn) according to a lithological division. We noticed that As is influenced by mining in the Massif Central and Copper in the Armorican Massif.

6.2.5 *Statistical treatments*

The statistical analysis allows to define the hydrogeochemical background entities by grouping the lithologies with similar element concentration distributions. The statistical processing of the data was focused on the comparison of the distribution of the concentrations of major, minor and trace elements according to the different lithological contexts and pH and redox typologies.

Lithology

The results based on Kruskal-Wallis tests and Nemenyi post-hoc tests highlight that volcanic rocks belongs to individual groups for Nickel, Zinc, Fluoride and sulfates. Crystalline rocks are individualized for elements such as As, Cd, Cr while As, Cd, Ni and Zn are individualized for metamorphic rocks. Ni are individualized for sandy aquifers. Gravel, including alluvial aquifers are highlighted for several elements such as As. Carbonates include a wide range of concentration, therefore, the element are not clearly identified, therefore sedimentary rock will be defined only for the Paris Basin area as second criteria to distinguish the lithological facies.

The dataset for cupper is not relevant for lithological discrimination.

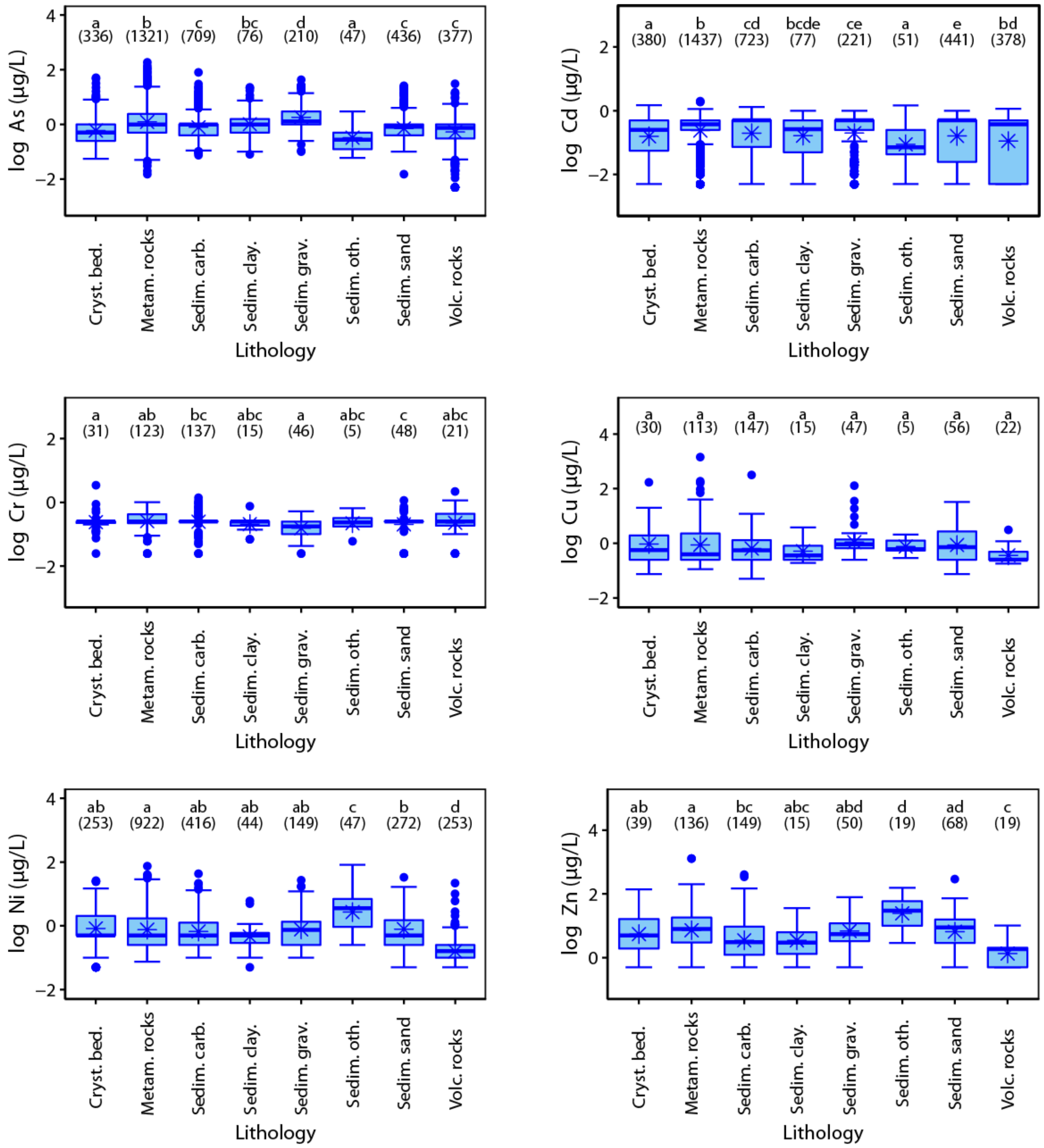


Figure 29: Results for Kruskal-Wallis test and Nemenyi post-hoc test for lithology families in Loire-Bretagne area (number of samples in brackets).

pH-HOVER

GW pH is partly dependent of lithology and hydrogeology. Therefore, looking at the distribuyion of lithology among acidic, basic and neutral water, it is possible to conclude that Crystalline and metamorphic rock mainly contains acidic waters (> 90% of sampling point). Sedimentary formations, classed as others are also acidic. The rocks classed as others sediments are seidenary formation mixed with cristalline rock or old sedimentary formations.

Carbonate, clay and marls contain mainly neutral pH, linked to the equilibrium with carbonates (~7-7.5) with basic water for carbonates (12%). Gravel are distributed between A and N classes while sand and volcanics rocks belong to the 3 groups (A, B, N).

Therefore, it seems to be not relevant to distinguish pH classes for Crystalline, metamorphic, carbonates, clays & marls and others.

Table 43: Distribution of lithology between pH-HOVER classes

Lithology	Nbr. of analyses	Mode	Classes	Frequency per class	Rel. frequency per class (%)
Crystalline bedrock	7076	A	**A**	6346	90.1
			B	94	1.3
			N	601	8.5
Metamorphic rocks	26777	A	**A**	24129	92.4
			B	498	1.9
			N	1476	5.7
Sedimentary: carbonates	13983	N	**A**	560	4.0
			B	1736	12.4
			N	11681	83.6
Sedimentary: clays and/or marls	1460	N	**A**	87	6.0
			B	78	5.3
			N	1295	88.7
Sedimentary: gravel	4302	N	**A**	1831	42.6
			B	150	3.5
			N	2321	54.0
Sedimentary: other	998	A	**A**	848	85.0
			B	37	3.7
			N	113	11.3
Sedimentary: sand	8412	N	**A**	1468	17.5
			B	2497	29.8
			N	4427	52.8
Volcanic rocks	6913	A	**A**	2947	46.7
			B	998	15.8
			N	2360	37.4

For pH classes, it is possible to conclude that Ni and Zn are influenced by pH, with higher concentrations in acidic conditions while cupper is less soluble in basic condtions. For arsenic, neutral conditions differs from acidic and basic waters with a wider range of concentrations. Sulfates and fluorides show higher concentrations in basic conditions. No differences are observed for Cd, Cr.

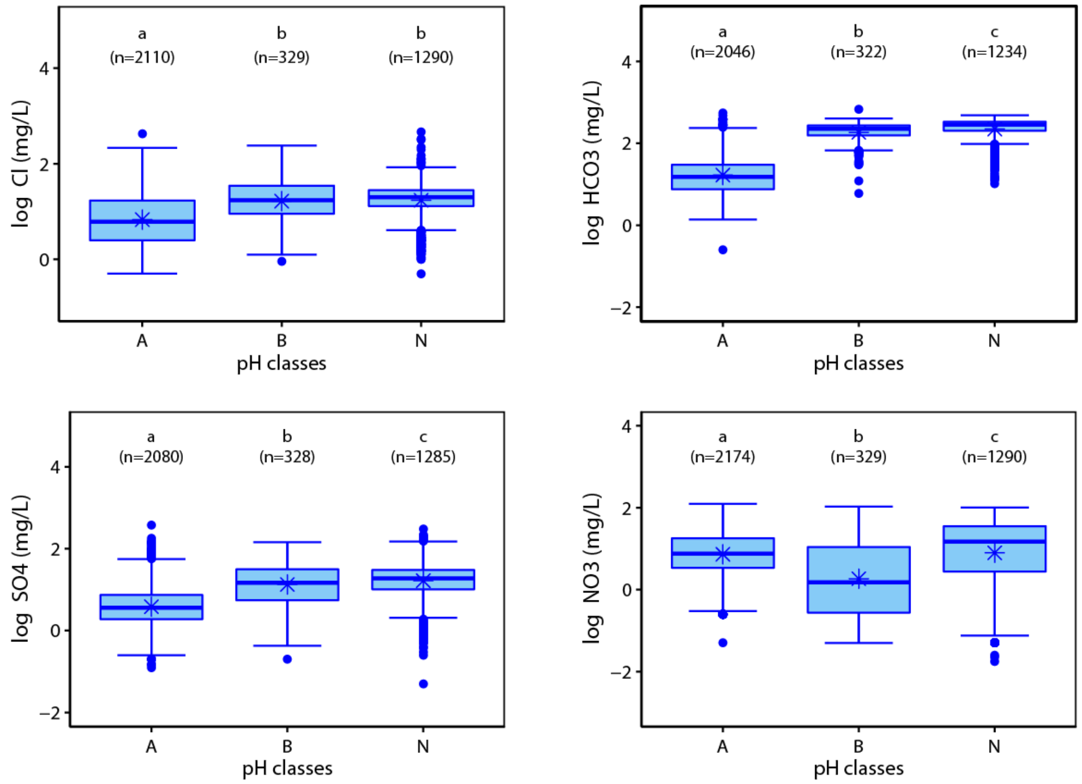


Figure 30: Results for Kruskal-Wallis test and Nemenyi post-hoc test for pH classes in Loire-Bretagne area (number of samples in brackets). – anions

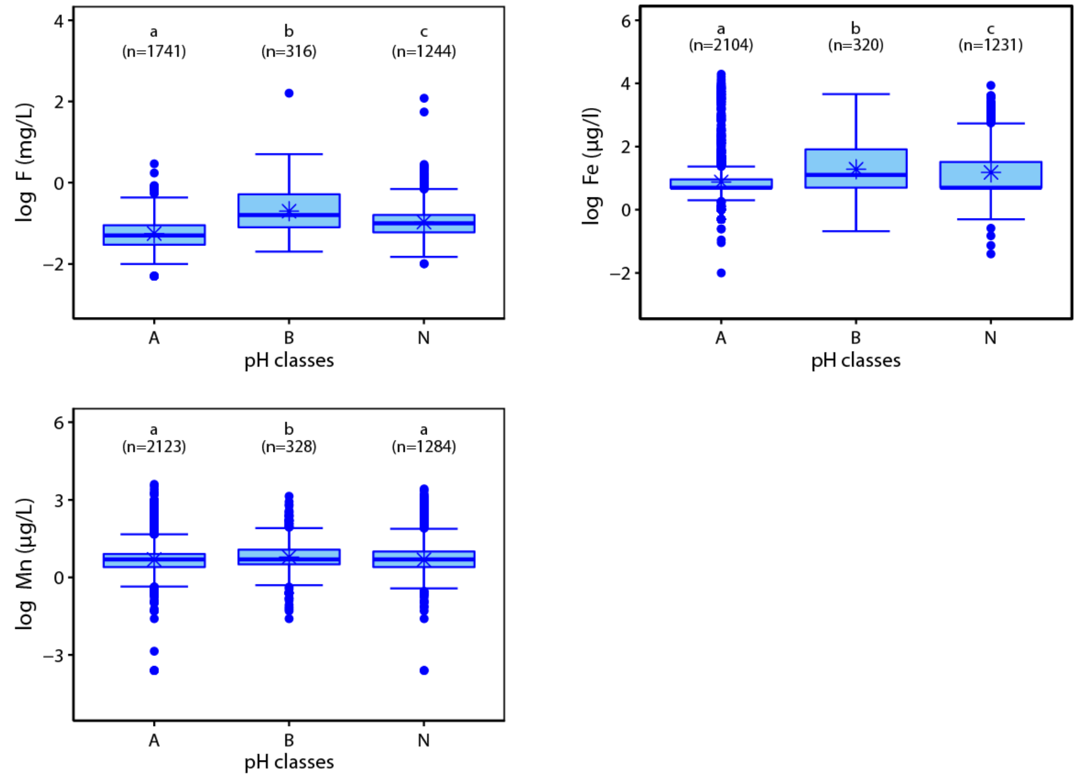


Figure 31: Results for Kruskal-Wallis test and Nemenyi post-hoc test for pH classes in Loire-Bretagne area (number of samples in brackets). – F, Fe Mn

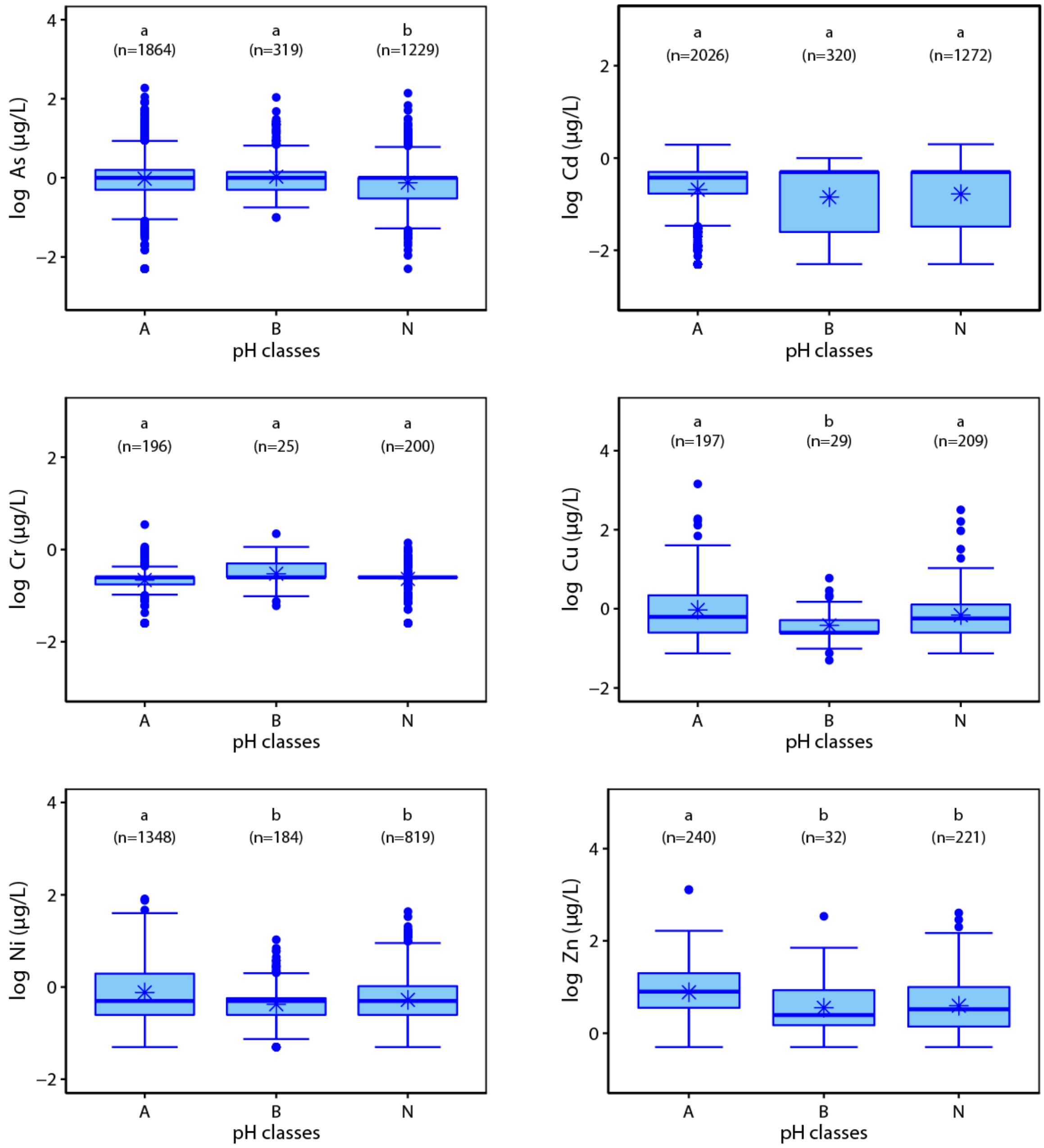


Figure 32: Results for Kruskal-Wallis test and Nemenyi post-hoc test for pH classes in Loire-Bretagne area (number of samples in brackets)- trace elements.

redox-HOVER

Regarding redox classes, the sampling points are mainly attached to oxic waters except for sand wichi is represented by both calluses (reduced and oxic).

Table 44:Distribution of lithology between redox-HOVER classes

Lithology	Nbr. of analyses	Mode	Classes	Frequency per class	Rel. frequency per class (%)
Crystalline bedrock	7076	A-B	A-B	6310	89.2
			C-D	766	10.8
			X	0	0.0
Metamorphic rocks	26777	A-B	A-B	24016	89.7
			C-D	2752	10.3
			X	9	0.0
Sedimentary: carbonates	13983	A-B	A-B	11301	80.8
			C-D	2678	19.2
			X	4	0.0
Sedimentary: clays and/or marls	1460	A-B	A-B	1278	87.5
			C-D	182	12.5
			X	0	0.0
Sedimentary: gravel	4302	A-B	A-B	3729	86.7
			C-D	573	13.3
			X	0	0.0
Sedimentary: other	998	A-B	A-B	746	74.7
			C-D	252	25.3
			X	0	0.0
Sedimentary: sand	8412	A-B	A-B	4740	56.3
			C-D	3670	43.6
			X	2	0.0
Volcanic rocks	6913	A-B	A-B	6306	91.2
			C-D	607	8.8
			X	0	0.0

The attribution of oxic and reduced class are well characterized as underline by the test applied on the geochemical parameters (Eh) and Mn that are not considred for the classificiation which is based on NO$_3$, Fe and O$_2$. This shows that reduced waters have higher pH and conductivity but also fluoride. As redox conditions controlled elements senttivie to redox process, we confirmed that reduced conditions induced high conditions in Fe, Mn, As. In addition, the tests higlitgh a stastitic difffrence among trace distributions under oxic and reduced water except for Zn.

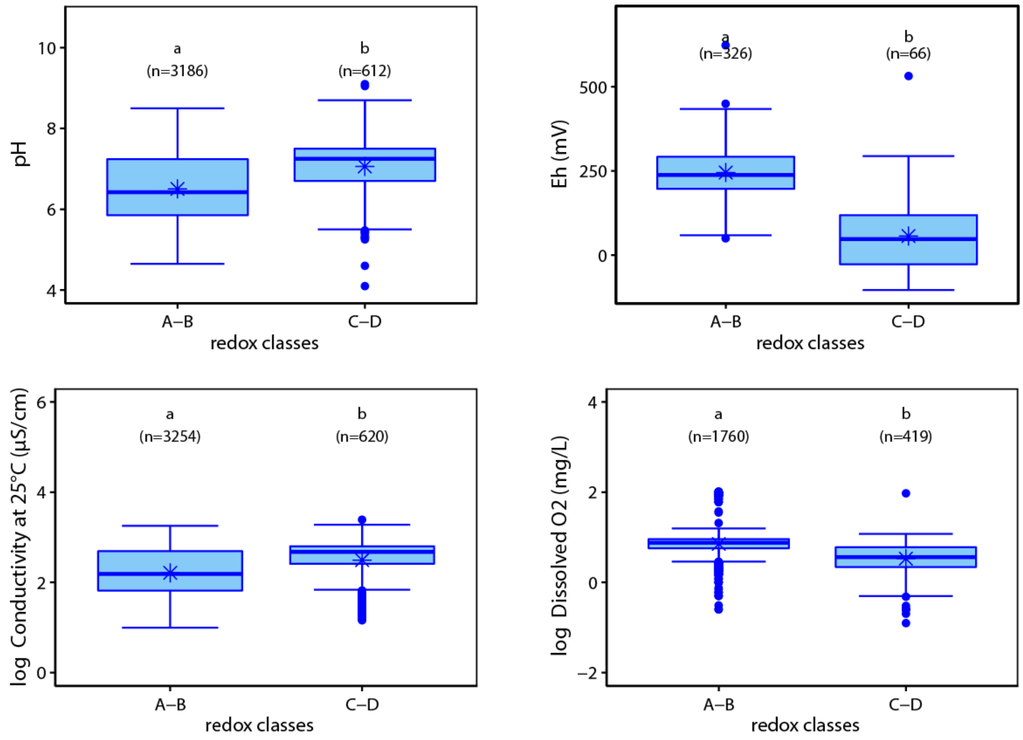


Figure 33: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (A-B) and reduced class (C-D) in Loire-Bretagne area (number of samples in brackets) – geochemical parameters.

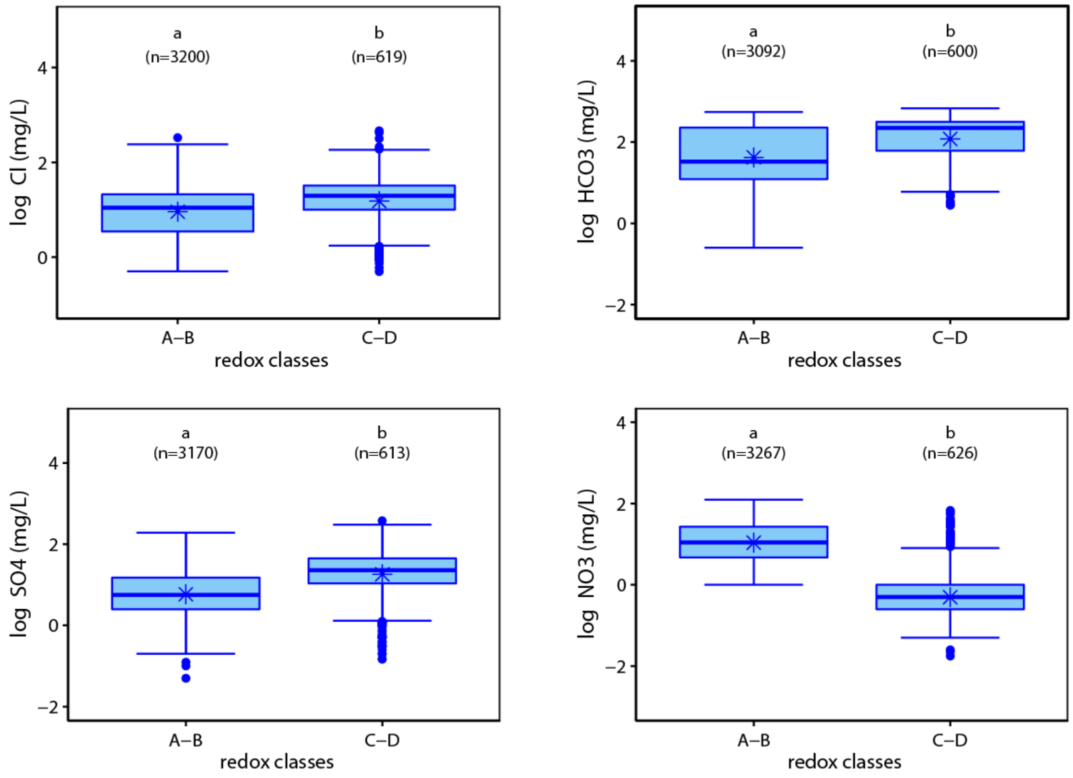


Figure 34: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (A-B) and reduced class (C-D) in Loire-Bretagne area (number of samples in brackets) – anions.

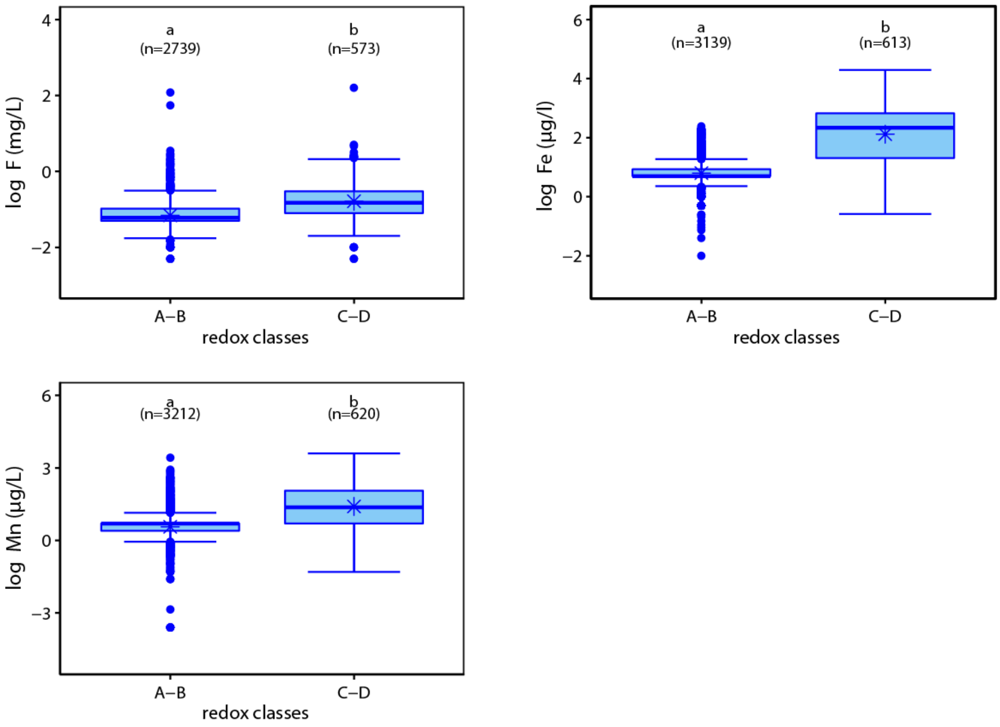


Figure 35: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (A-B) and reduced class (C-D) in Loire-Bretagne area (number of samples in brackets) – Fe, Mn and F.

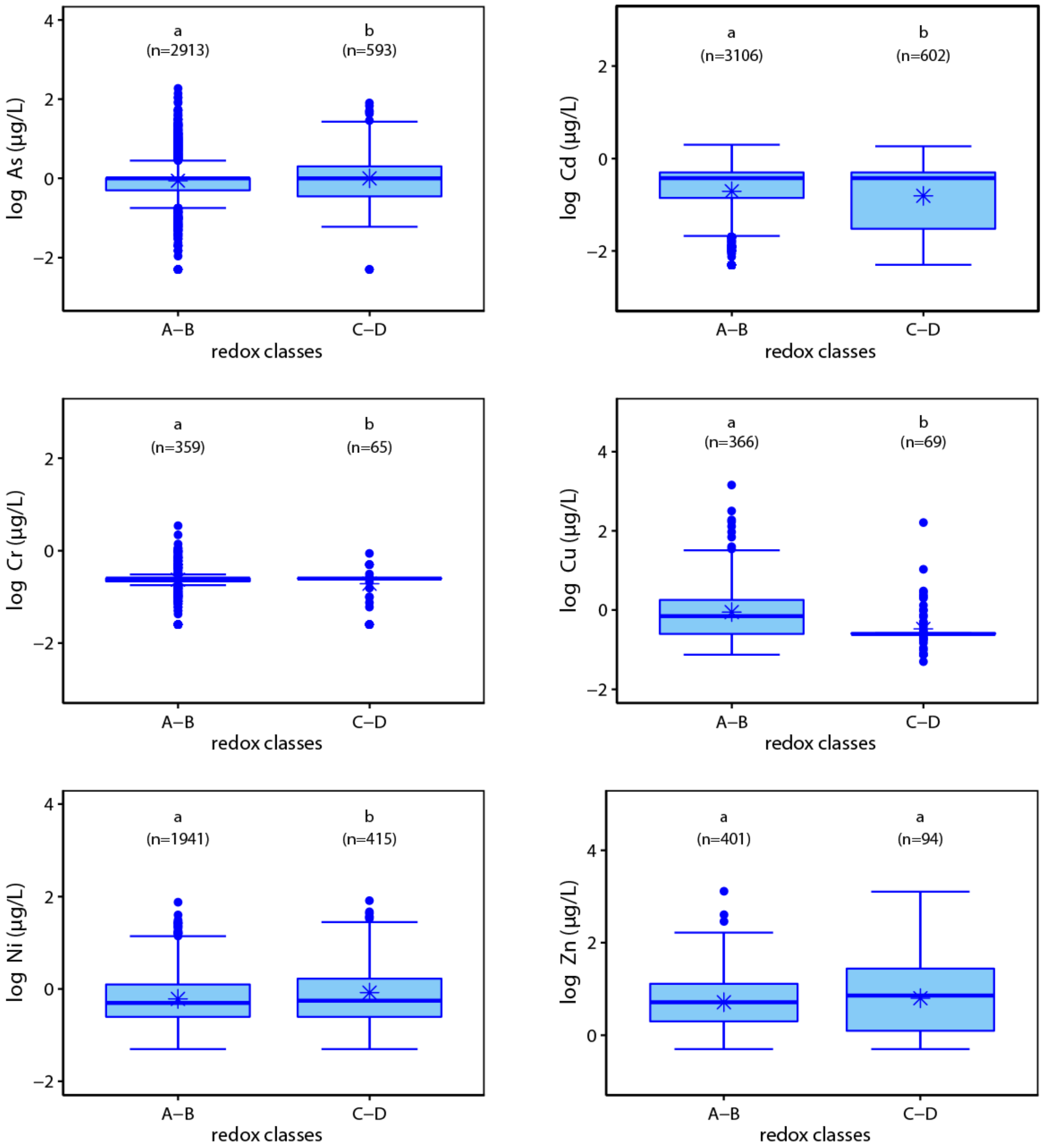


Figure 36: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (A-B) and reduced class (C-D) in Loire-Bretagne area (number of samples in brackets).

pH & redox classes

Looking at the conductivity, the following figure shows that conductivity is more dependent from pH, with lower value is acidic water. This is clearly explain by th e fact that acidic water are less mineralized, generally on crystalline or metamorphic roch while mineralized water are observed in carbonated water with higher pH value. Dissolved O2 is also reported for comparison, showing the fact that pH and redox are factors controlling its concentration in GW.

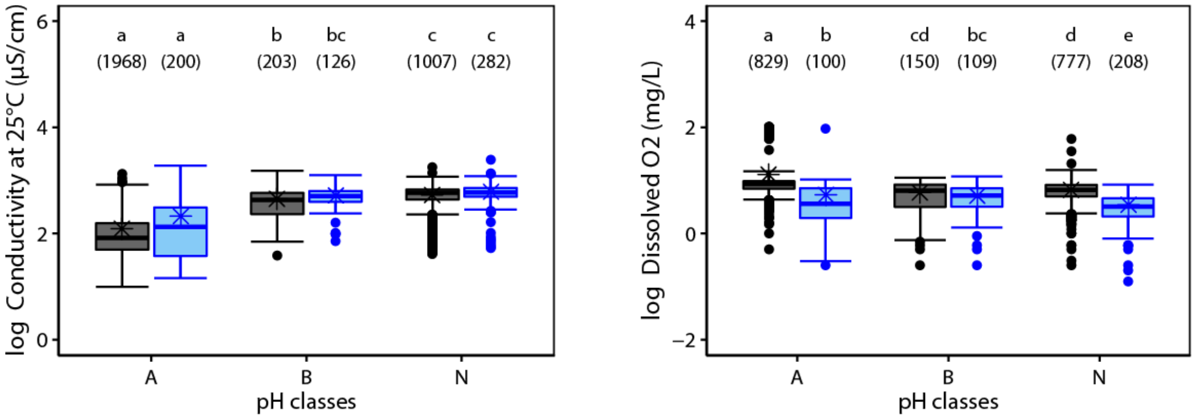


Figure 37: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (in black) and reduced class (in blue) in Loire-Bretagne area (number of samples in brackets) – geochemical parameters.

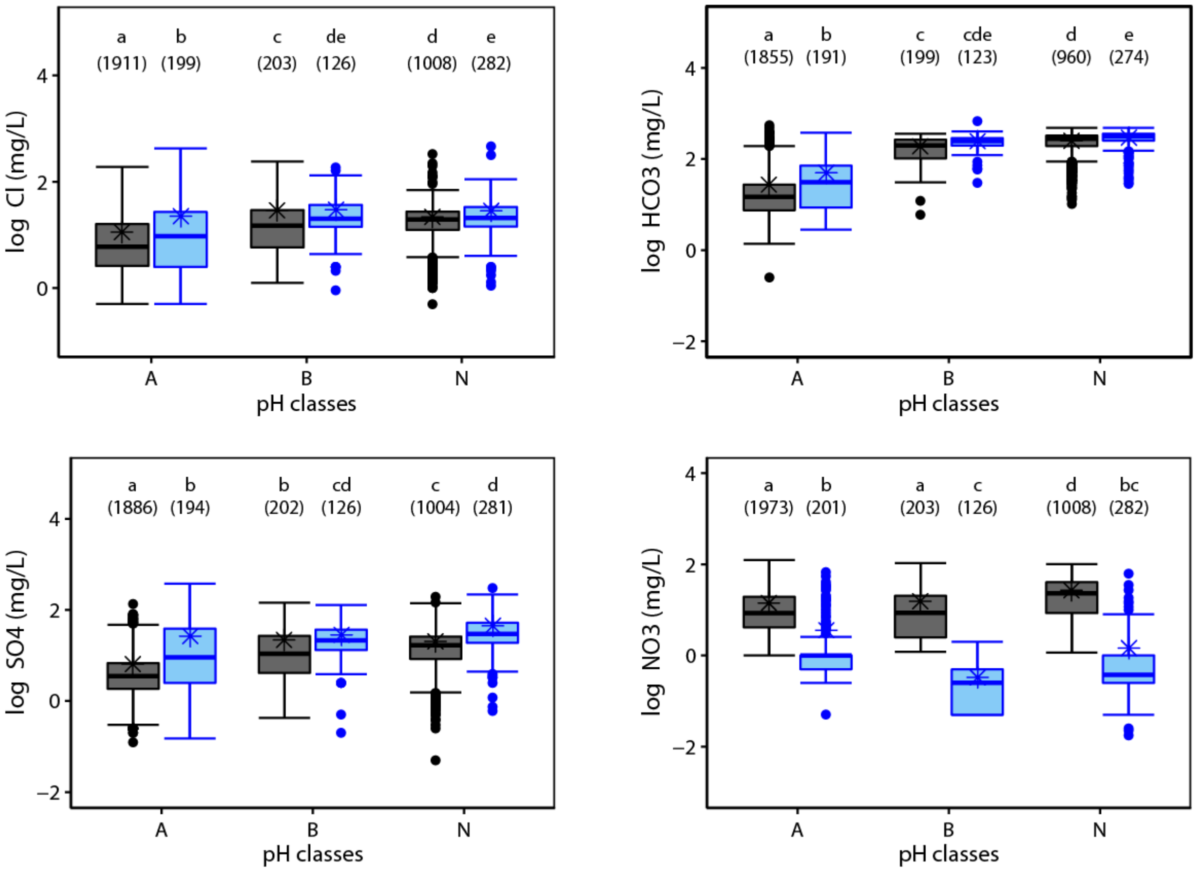


Figure 38: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (in black) and reduced class (in blue) in Loire-Bretagne area (number of samples in brackets) – anions.

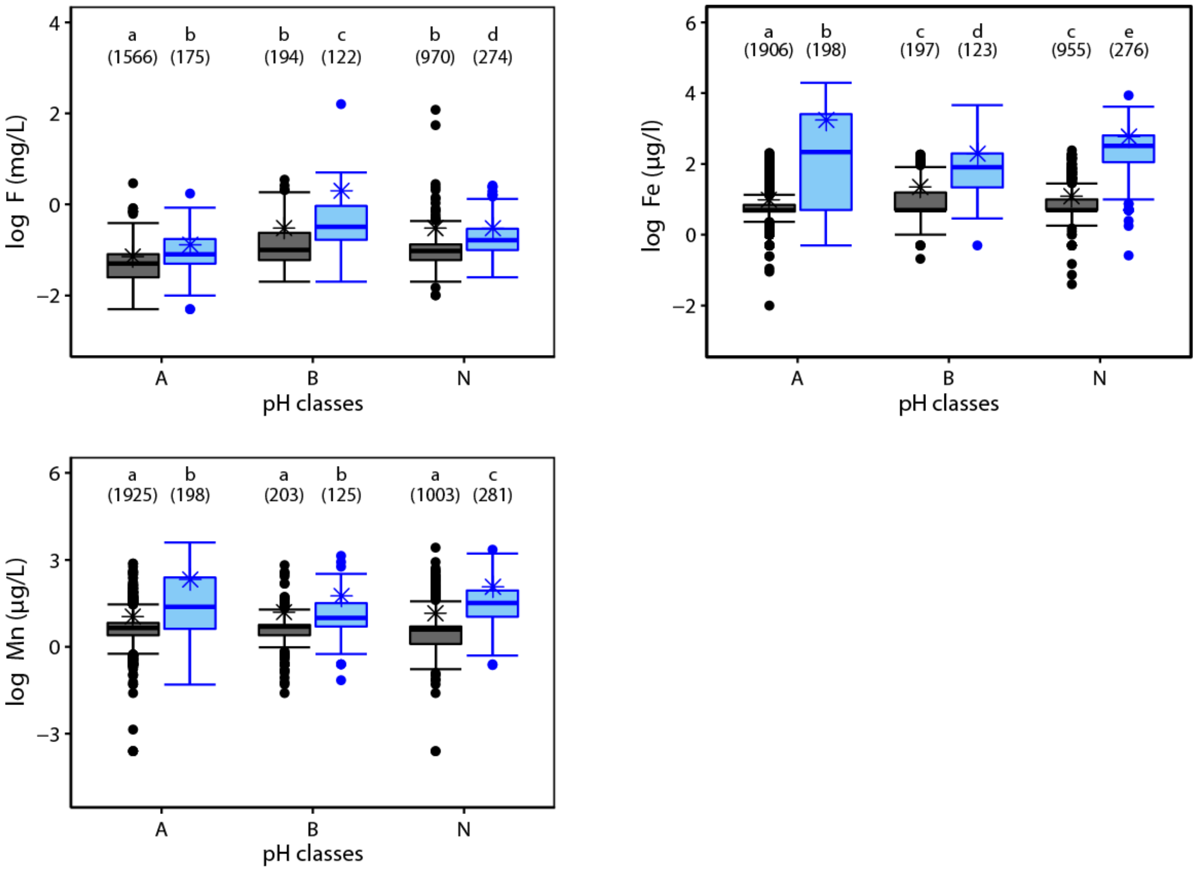


Figure 39: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (in black) and reduced class (in blue) in Loire-Bretagne area (number of samples in brackets) – minors elements.

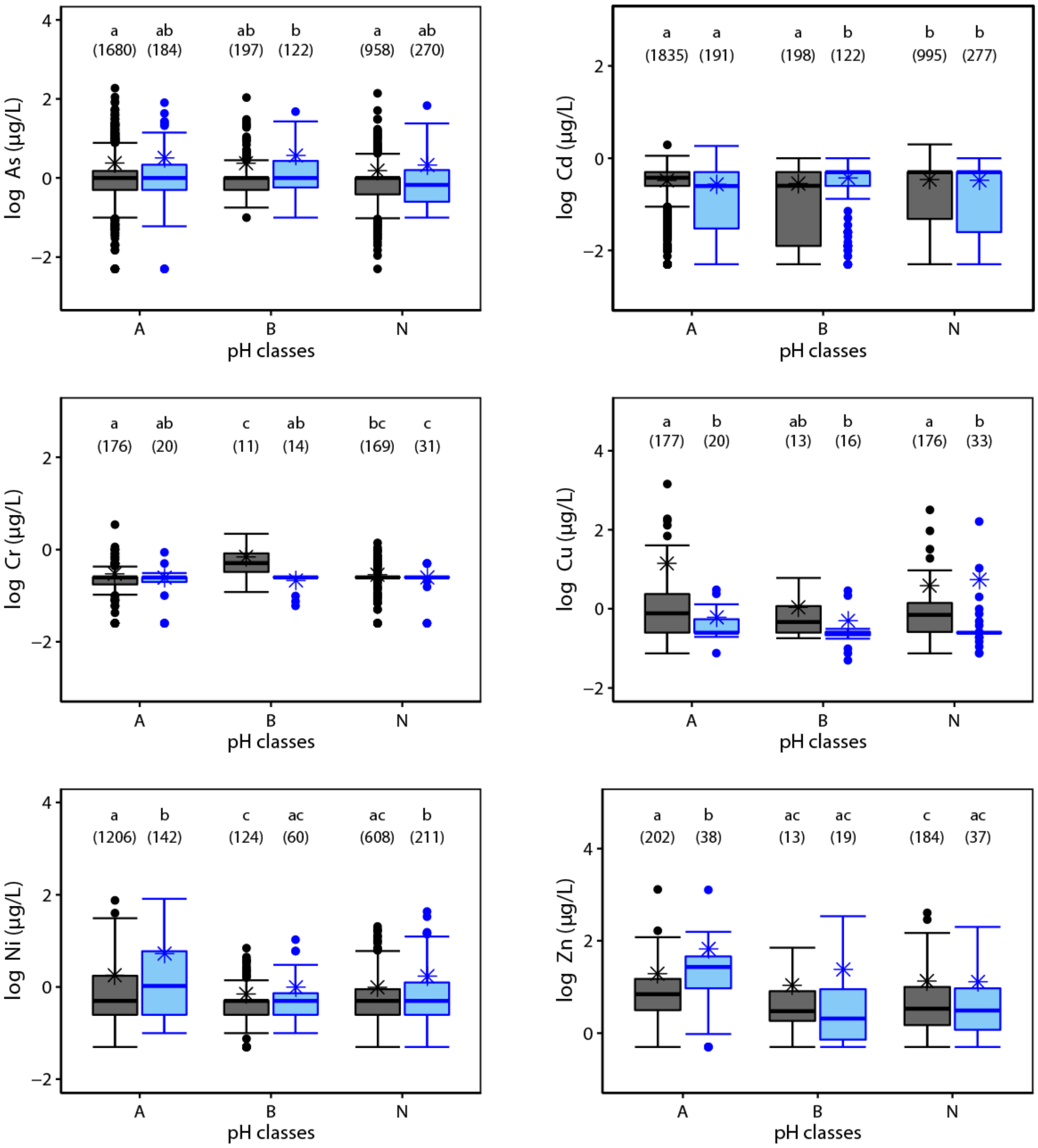


Figure 40: Results for Kruskal-Wallis test and Nemenyi post-hoc test for oxic class (in black) and reduced class (in blue) in Loire-Bretagne area (number of samples in brackets) – trace elements.

According to these tests it make possible to distinguish the water families for NBLs calculation. Based on thses results we select the following criteria for each elements.

For fluoride, redox is not directely involved in its mobilization as it is an element sentivie to redox process, howver, Fluoride is predominant in confined aquider and deep GW that is illustrated here by reduced waters (low NO_3, low Fe, low O_2 dissolved).

Table 45: Factors selected for the water familiy for each element.

Elements	**Factor Influencing concentration distribution**
As, Cd, Cr, F, SO4	Redox
Zn	pH
Ni, Cu	pH & Redox

Lithology, pH & redox classes

The following figures helps to select the water family for each class. In fact, regarding the number of data and the results from the Kruskal Wallis test, it is possible do detailed the water family for each lithology. For exempla, for arsenic, the number of data is not relevant to determine the NBLs for each class with no difference between oxic and reducd and basic or acidic waters. For carbontes, the differences is observed between reduced and oxidized but not according to the pH classes. This justify the choice made in Table 45.

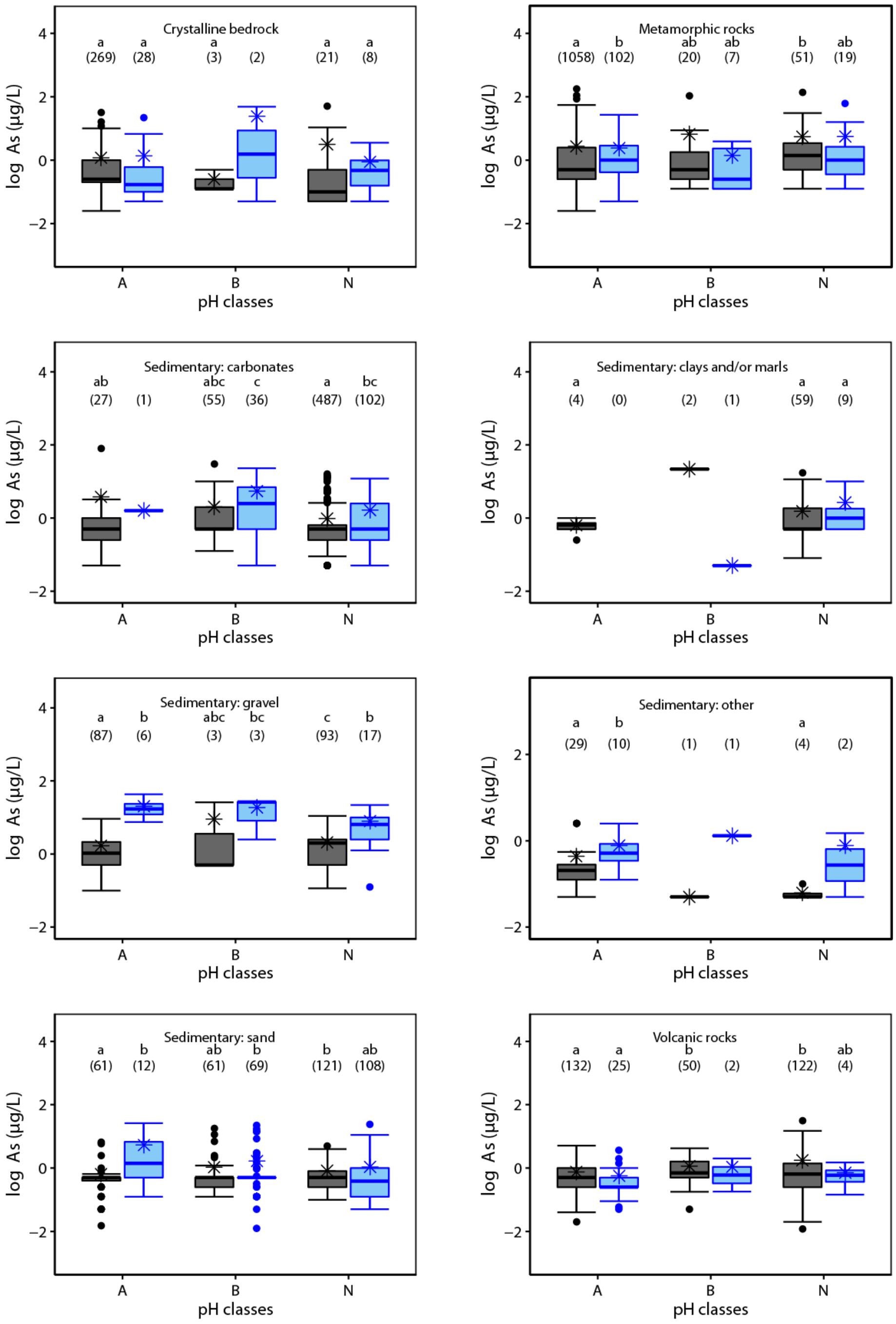


Figure 41: Results for Kruskal-Wallis test and Nemenyi post-hoc test for ***arsenic*** *- oxic class (in black) and reduced class (in blue) - (number of samples in brackets).*

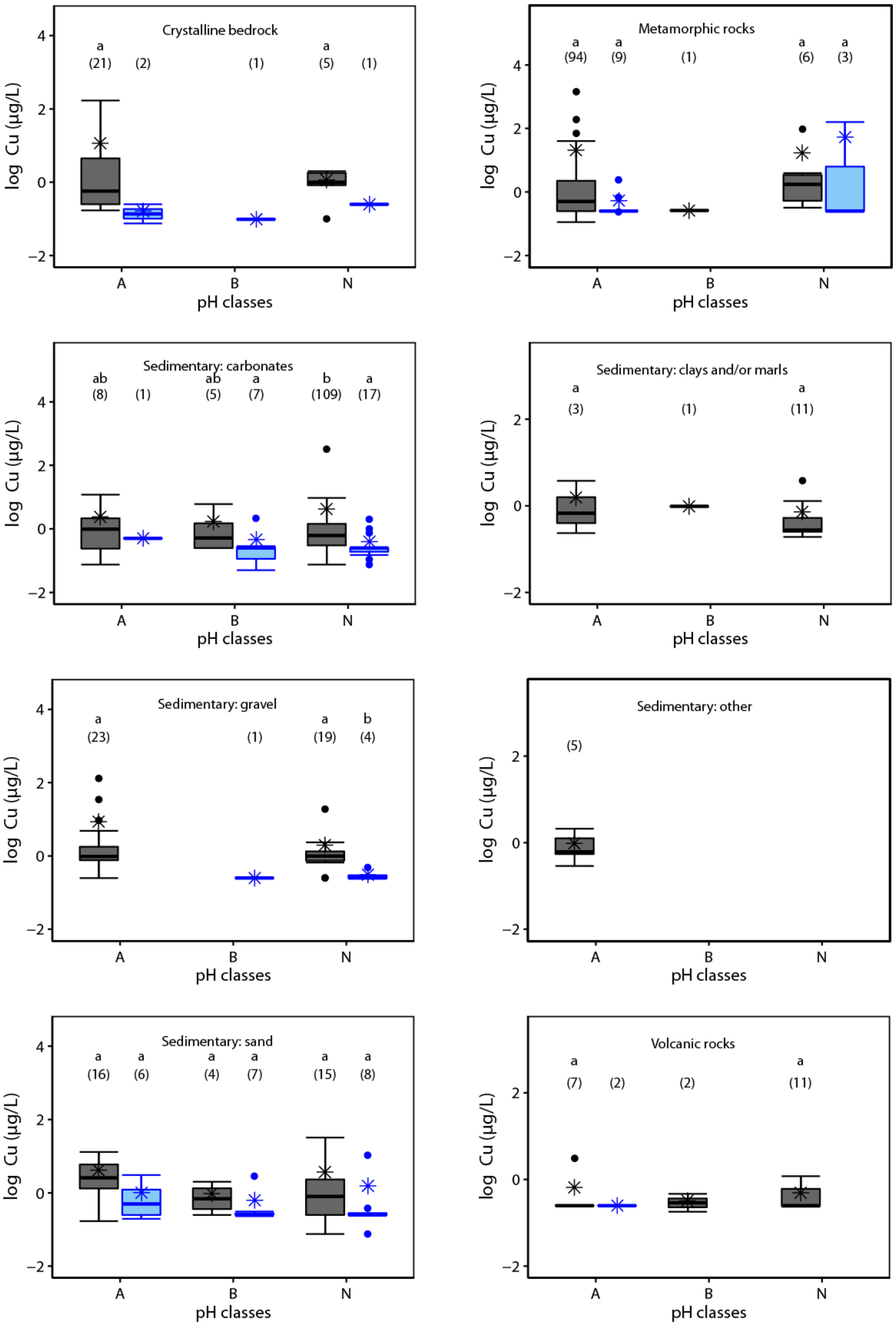


*Figure 42: Results for Kruskal-Wallis test and Nemenyi post-hoc test for **cupper** - oxic class (in black) and reduced class (in blue) - (number of samples in brackets).*

6.2.6 *Calculating natural background levels with the final dataset*

All these steps were necessary to prepare the data used to calculate the reference concentrations by lithology for the following major, minor and trace elements: As, Cd, Cr, Cu, F, Ni, Zn, Cl, and SO_4.

Within each of these areas, the water sampling points whose element concentration is considered by statistical approach to be influenced by a diffuse or point source of contamination has been removed on the basis of the results obtained in the previous steps.

For each distribution function, the 90th percentile and 10th percentile were calculated. The calculated 90th percentiles can be assimilated to the upper limit of the concentration that can be reached by the natural background level on all the defined lithology for each geological domain. For certain hydrogeological entities, the 90th percentiles calculated for certain elements, in particular nitrate, zinc and cadmium are very high and tend to overestimate the maximum concentration within a lithology. This may result from the inclusion of water sampling points under the influence of anthropogenic pressure in the calculation of the distribution function despite the methodology put in place to rule them out, or from generalized anthropogenic pressure in certain sectors.

The calculation of the 10 percentiles for each distribution function was carried out in order to exclude as much as possible the water sampling points not representative of the lithology.

Indeed, the assumption was made that all of the water sampling points from the 10th percentile are not influenced by anthropogenic pressure. However, calculating the 10th percentile tends to underestimate the natural variability of the natural background level within a lithology.

Thus, water sampling points with concentrations of major, minor or trace elements exceed the reference concentrations would not necessarily be under the influence of a pressure source but the measured concentrations could only be induced by the nature of the hydrogeological context. **The calculation of the 10th and 90th percentile values thus makes it possible to limit the variability of the natural background level for each lithology.** This approach allows to reduce the sources of error during the calculation of the distribution functions induced by the presence of non-representative data and anthropogenic pressure. The 10th and 90th percentiles of the medians for all the points not influenced by anthropogenic pressures are reported in three summary tables for each of the domains considered.

Table 46: NBLs established on the basis of the 90th percentile for the Loire Basin River – Sedimentary facies from Paris Basin.

	GW threshold values		HOVER lithology							
			Sedimentary		Sedimentary		Sedimentary		Sedimentary	
			Paris Basin		Paris Basin		Paris Basin		Paris Basin	
			Sand		Gravel		Carbonates		Clays/marls	
			n	NBL	n	NBL	n	NBL	n	NBL
As	10	µg/l	343	1.46	33	4.96	554	3	48	4
Cd	5	µg/l	221	<0.5	723	<0.5	723	<0.5	77	<0.5
Cr	50	µg/l	5	<0.45	<5	_	17	<0.4	<5	_
Cu	2000	µg/l	39	4.4	9	1.4	131	3.1	15	2.8
Ni	20	µg/l	231	2.3	27	7.8	375	2.8	44	0.9
Zn	5000	µg/l	38	12.3	10	65.1	133	24.1	15	26.7
F	1.5	mg/	388	0.9	51	0.2	673	0.3	75	0.3
Cl	200-250	mg/	394	53	52	27.9	686	38.8	77	23.1
SO4	250	mg/	394	69.9	52	56.2	686	44	77	35.6

Table 47: NBLs established on the basis of the 90th percentile for the Loire Basin River for Massif Central and Armorican domains.

	GW threshold values		HOVER lithology					
			Volcanic rocks Massif Central		Crystallines bedrocks		Metamorphic rocks	
			n	NBL	n	NBL	n	NBL
As	10		373	2.36	336	3.6	1321	6.7
Cd	5	µg/l	378	<0.5	380	<0.5	1437	<0.5
Cr	50	µg/l	21	1	31	<0.5	123	<0.5
Cu	2000	µg/l	13	<1	25	9.6	95	11
Ni	20	µg/l	139	0.3	211	5.1	741	5
Zn	5000	µg/l	19	<5	40	23.8	139	40.4
F	1.5	mg/	310	0.2	318	0.1	1146	0.1
Cl	200-250	mg/	392	11.4	383	27	1496	28.1
SO4	250	mg/	392	7.3	383	24	1469	18

Table 48: NBLs established on the basis of the 90th percentile – lithology classes combined with pH & redox classes.

HOVER-lithology	Elements	units	HOVER-pH	HOVER-redox			NBLs (only lithology)
				A & B	C&D	All redox	
sand	As	µg/l	All pH	1.1	1.5		1.46
	Ni	µg/l	Acidic	10.2	13.5		2.3
			Basic	1.3	2.6		
			Neutral	2.1	2.3		
	Zn	µg/l	Acidic			16.6	12.3
			Basic			11.8	
			Neutral			10	
	Cu	µg/l	Acidic	8			4.4
			Basic		1.3		
			Neutral	8.8	3.5		
carbonates	As	µg/l	All pH	1.2	6.4		3
			Basic	3.5	0.6		
			Neutral	2.2	5.5		
	Cu	µg/l	Basic	2.2	0.3		
			Neutral	3.2	0.6		
volcanics	As	µg/l	All pH	2.44	1.4		2.36

6.3 Comparison and contributions of the proposed method

The proposed method be based only on data using specific high-performance tests that can be transposed to different scales of work. The study required significant work in processing data from the ADES groundwater database in order to be able to carry out calculations of reference concentrations.

An innovative approach for linking spatial information automatically by determining the relation between watersheds and point of anthropogenic pressure. Likewise, a statistical treatment based on heterogeneous databases (Corine Land Cover, BASOL, etc.) was carried out in order to identify the parameretrs influence by anthropic activities. This selection cannot be done without are carefully analysis of data and it could be beter to work at the scale of GW body to highlight the well influence by contamination. At the end is to remove data for each of the dissolved elements where the water sampling points are influenced by agricultural, urban or industrial activities in order to improve NBLs determiantion. Therefore, the removal of data should not induce an increase of NBLs by removing data with value in the rable of NBLs. These data removal should be applied only to water sampling with outliers.

In addition, this work has limitations due to the overall approach chosen. Thus, some water sampling points under anthropic influence were not detected during statistical processing, for example due to the generalization of anthropic influence at the scale of the chosen work unit. This is the case, for example, for nitrate, zinc and cadmium in certain sectors and mainly alluvium. Thus, the calculation of the values of the 10 and 90 percentiles made it possible to limit the variability of the natural hydrogeochemical background within a lithology/geology family or hydrogeochemical entity. This approach helps to reduce the sources of error during the calculation of the distribution functions induced by the presence of water points influenced by a source of anthropogenic pressure in the dataset.

A global approach has been carried by lithology, however in Devau et al. (2017) the hydrogeochemical entities have been adapted to the context of each of the three major geology defined in Loire-Bretagne (Massif Central, Paris Basin, Armorican Massif). The proposed hydrogeochemical entities go beyond the use of a single database (such as BDLISA entities for example) and have made it possible to refine hydrogeological information with geological and lithological data. With this approach it is possible to refine NBLs for lithology such as crystalline or metamorphics rocks that are higly dependant of petrography. Ina ddition, sedimentary rocks such as sands or others are also dependent of the geological context and weathering of rocks. These lithologies present a huge variability controlled by geological context.

The statistical tests carried out (nonparametric hierarchical univariate analysis (Kruskal-Wallis test), statistical distributions of concentrations combined with the Post-hoc Nemenyi test, 90th percentiles of the medians) allows to determine the reference concentrations (or NBL) per hydrogeochemical entity for the elements major, minor and traces.

The calculations show that, depending on the trace and minor elements and the domains considered, between 1 and 10% of the monitoring data of the working area exceed the reference concentrations.

7 CASE STUDY – AUSTRIA

7.1 Previous studies on NBLs

Natural background levels of surface near Austrian groundwater bodies were first described by Hobiger et al. (2004) on a regional level. A follow-up study published in 2018 focused on specific geological units and hydrochemical time series obtained mainly from the Austrian groundwater monitoring network (Wasserinformationssystems Austria (WISA), see Gewässerzustandsüberwachungsverordnung, BGBl. II Nr. 479/06 i.d.g.F.) (Brielmann et al. 2018). The project considered 2.789 monitoring sites and hydrochemical data series between 1997 and 2017 with a total number of approx. 2,5 million analyses of single parameters (EC, pH, O_2, Ca, Cl, HCO_3, K, Mg, Na, SO4, NH_4, Fe, Mn, NO_3, NO_2, PO_4, Al, As, Pb, Cr, B, Cd, Ni, Cu, Hg, U and Zn).

Statistical methods were applied after a profound data treatment to obtain secured background levels. Methods comprised frequency distributions, probability plots, box plots and histograms. Concentrations deviating from the basic distribution (between 10% and 90% percentile) were excluded from the derivation of the background concentrations as locally elevated values (due to anthropogenic influences or e.g. mining).

7.2 Study with the proposed method

7.2.1 *Constitution of the groundwater quality dataset*

The Austrian dataset for Task 3-3 comprises publicly available data of the Austrian groundwater monitoring network (WISA Wasserinformationssystem; see Gewässerzustandsüberwachungsverordnung, BGBl. II Nr. 479/06 i.d.g.F.) which can be accessed online: https://wasser.umweltbundesamt.at/h2odb/index.xhtml.

The monitoring network covers the entire country and monitoring points observe surface near groundwater bodies. The majority of those monitoring points are also found in areas of major basins and inneralpine valleys.

Figure 43 provides an overview of the Austrian geology and the considered monitoring points for this study. Raw data of the Austrian groundwater monitoring network was also used for Task 3-4, however a different data processing approach took place. The complementary case study area is only relevant for Task 3-4, but can be seen as a link to this task.

A large number of geological units affect NBLs and it is important to obtain a brief overview of the geology of Austria. The nappes of the Penninicum are part of a former ocean floor that was situated between the former European continent in the north and the Austro-Alpine units as a part of Africa in the south, and represent an important suture. Today, the Penninicum appears in two geological zones: in the north towards the ‘Northern Calcareous Alps’ as a non-metamorphic ‘flysch zone’; and in the central zone of the Alps within metamorphic windows. The most expanded one of these is the ‘Tauern Window’. The Helveticum and the Subpenninicum are parts of the former European continent that were incorporated into the Alps. The non-metamorphic Helvetic zone underlays the non-metamorphic Penninicum in the ‘flysch zone’, and the metamorphic Subpenninicum underlays the metamorphic Penninicum in the ‘Tauern Window’, and vice versa. In the ‘Tauern Window’, the metamorphic Variscan granite intrusions called the ‘Central Gneiss’ belong to the Subpenninicum. The former European continent and the Penninicum were overthrusted by the Austro-Alpine nappe system, which is divided into the Lower Austro-Alpine and Upper Austro-Alpine nappes. The Austro-Alpine units and the Southern Alps originate from the microcontinent Apulia, which was a part of Africa. More details can be found in Schuster et al. (2014). Major basins with thick sedimentary covers are the Vienna Basin, the Styrian Basin and the Molasse Zone.

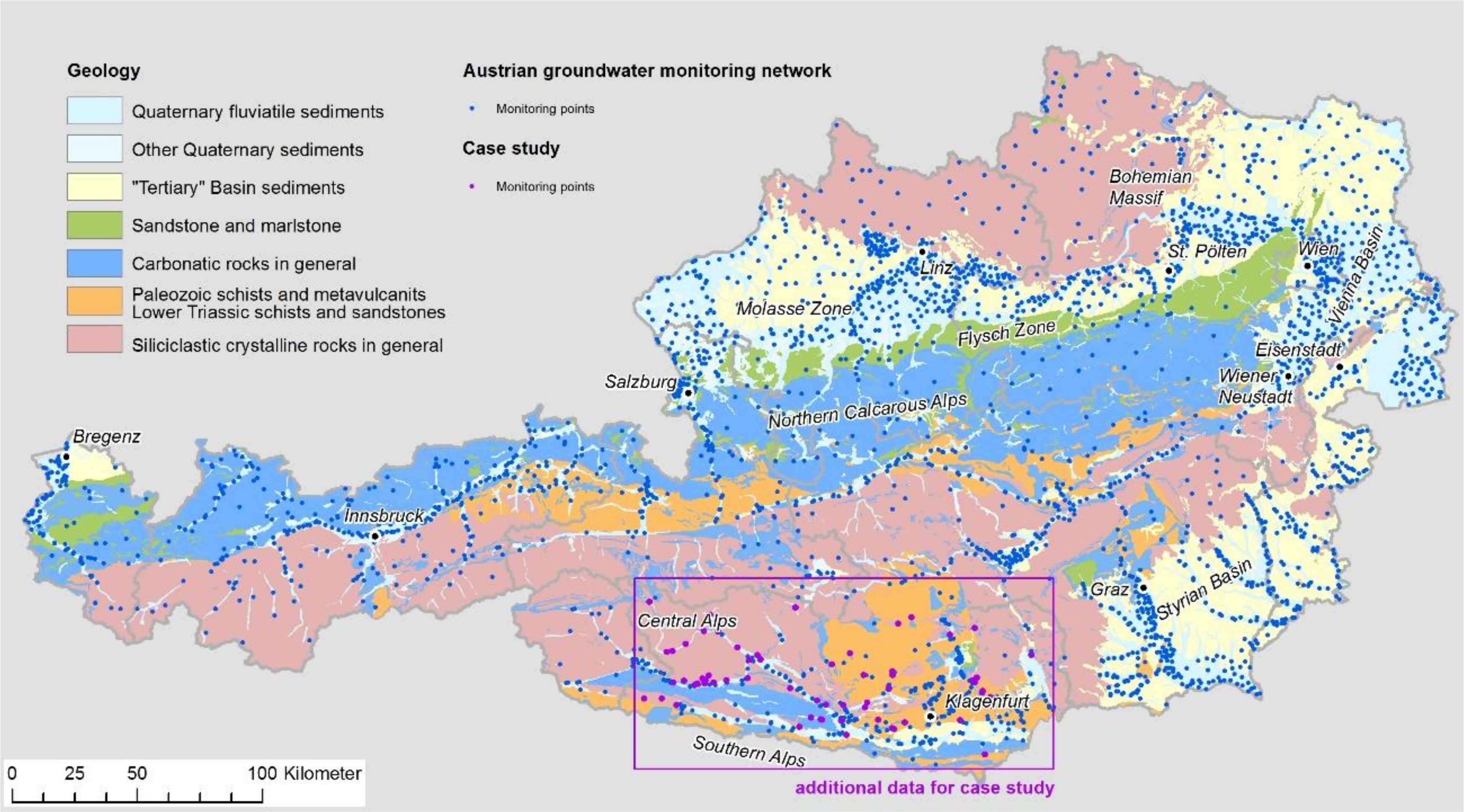


Figure 43: Overview of the geological situation in Austria, considered monitoring points of the Austrian groundwater network (Gewässerzustandsüberwachungsverordnung, BGBl. II Nr. 479/06 i.d.g.F.; https://wasser.umweltbundesamt.at/h2odb/index.xhtml) and an additional Austrian case study area considered for WP3 Task 3-4.

The provided dataset comprises 2.024 monitoring points for the time period 2010 to 2020 and considers in total 604.353 analyses for the following parameters: As, Ca, Cd, Cl, Cr, Cu, Fe, HCO_3, K, Mg, Na, Ni, NO_3, O_2, SO_4 and Zn (see Table 49). It is important to mention that the Austrian monitoring network does not cover the parameters fluoride and redox potential.

The number of analyses below LOD or LOQ are above 50 % for several minor and trace elements (Cd, Cr, As, Ni, Fe) (see Table 49). Those values below LOD or LOQ are processed with LOD/2 or LOQ/2, but minimum and maximum values are also considered to avoid unrealistic elevated concentrations. The maximum LOD and LOQ values for critical trace elements are between 0,001 and 0,0001 mg L^{-1}.

Table 49: Overview of the provided Austrian dataset, ranked by percentage of values below LOD or LOQ.

Parameter	Number of values above LOD or LOQ	Number of value below LOD or LOQ	Number of values / parameter	Percentage of values below LOD or LOQ	Max. LOD or LOQ (mg L^{-1})	Min. LOD or LOQ (mg L^{-1})
As	6290	21750	28040	77,57	0,001	1,16E-04
Ca	43743	9	43752	0,02	1	0,50
Cd	975	27062	28037	96,52	0,0001	1,34E-05
Cl	40002	3750	43752	8,57	1	0,14
Cr	4587	23452	28039	83,64	0,001	8,00E-05
Cu	12984	15054	28038	53,69	0,005	7,00E-05
Fe	14693	29049	43742	66,41	0,02	2,70E-04
HCO_3	43727	4	43731	0,01	1	0,50
K	37560	6178	43738	14,13	2	6,00E-02
Mg	43535	219	43754	0,50	1	0,50
Na	40932	2654	43586	6,09	1	0,2
Ni	6410	21632	28042	77,14	0,001	3,00E-05
NO_3	40739	3028	43767	6,92	1	0,14
O_2	42376	1189	43565	2,73	1,2	0,05
SO_4	43401	328	43729	0,75	1	0,2
Zn	20466	7575	28041	27,01	0,02	1,84E-04
Total	***442420***	***162933***	***605353***			

Box plots with median values are provided in Figure 44, the Austrian threshold values (red dots; Parameter- und Indikatorparameterwerte der TWV, BGBl. II Nr. 304/2001 i.d.g.F.) are also considered in this Figure. The number of samples per monitoring points for the time period 2010-2020 is commonly in the range of 20 for major elements and 14 for trace and minor elements. A reduction of monitoring points took place based on the availability of samples per monitoring sites (2071 to 2024).

Threshold exceedances are found in few cases for the following elements: Na, Cl, NO_3, Fe, SO_4, As and Ni.

The correlation matrix produced with median values for each monitoring sites shows correlations between major elements (e.g. Ca / HCO_3), but less for minor and trace elements (see Table 50).

Figure 44: Hydrochemical characteristics of the Austrian dataset (median values) taking account of national threshold values (red dots; Parameter- und Indikatorparameterwerte der TWV, BGBl. II Nr. 304/2001 i.d.g.F.).

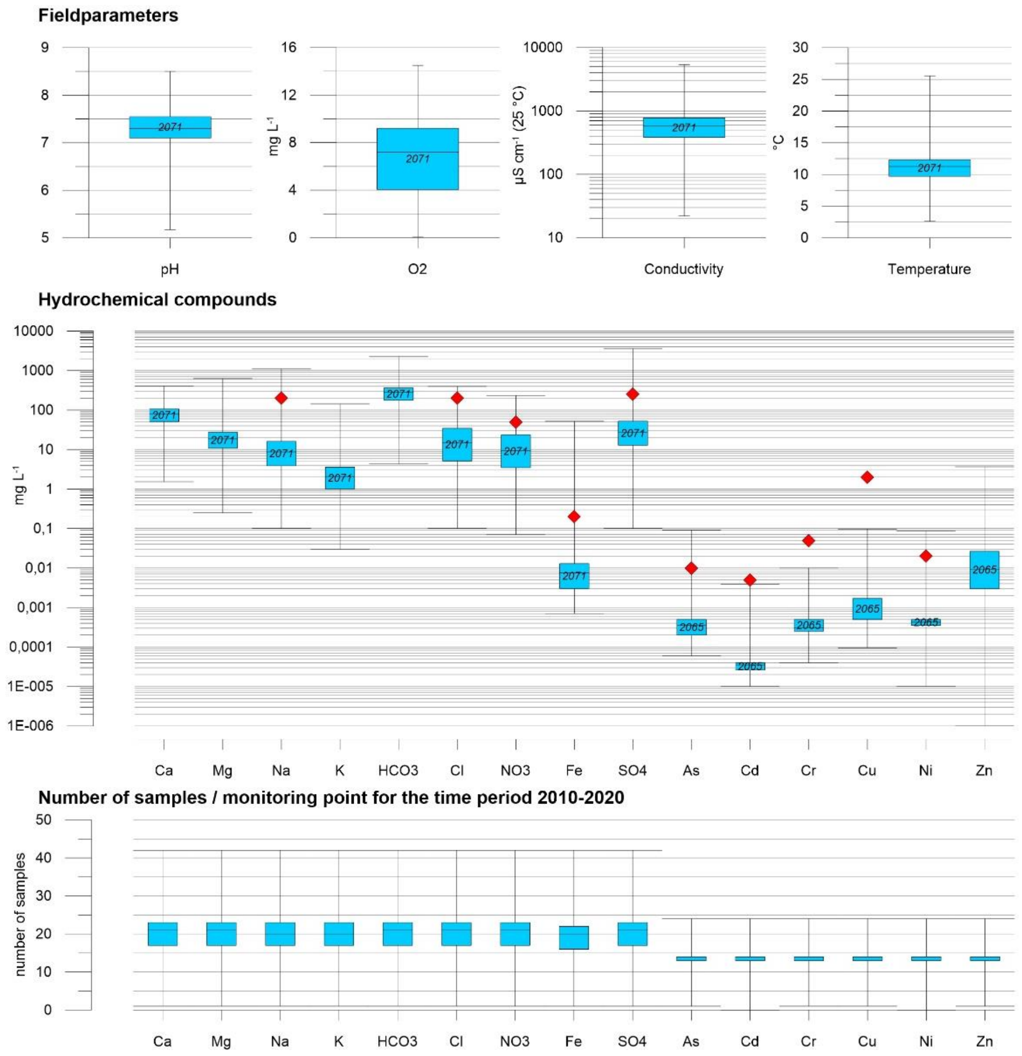


Table 50: Correlation matrix with median values of the Austrian dataset.

	T	EC	pH	O_2	Ca	Mg	Na	K	HCO_3	SO_4	NO_3	Cl	As	Cd	Cr	Cu	Fe	Ni	Zn
T	1,00	0,55	-0,42	-0,50	0,55	0,32	0,27	0,24	0,51	0,19	0,36	0,47	0,03	0,06	0,11	0,20	0,04	0,07	0,14
EC	0,55	1,00	-0,11	-0,39	0,84	0,82	0,63	0,39	0,83	0,68	0,48	0,72	0,02	0,02	0,16	0,09	0,01	0,06	0,07
pH	-0,42	-0,11	1,00	0,44	-0,08	-0,04	-0,07	-0,07	-0,01	-0,04	-0,17	-0,20	-0,10	-0,04	0,01	-0,14	-0,18	-0,20	-0,04
O_2	-0,50	-0,39	0,44	1,00	-0,31	-0,26	-0,26	-0,19	-0,34	-0,23	0,00	-0,32	-0,22	-0,04	0,06	-0,04	-0,24	-0,16	-0,06
Ca	0,55	0,84	-0,08	-0,31	1,00	0,62	0,24	0,18	0,80	0,48	0,43	0,52	-0,03	0,02	0,09	0,09	-0,03	0,01	0,08
Mg	0,32	0,82	-0,04	-0,26	0,62	1,00	0,37	0,28	0,57	0,84	0,39	0,48	-0,01	0,01	0,15	0,04	-0,03	0,04	0,04
Na	0,27	0,63	-0,07	-0,26	0,24	0,37	1,00	0,26	0,51	0,44	0,14	0,56	0,05	0,01	0,07	0,03	0,02	0,06	0,02
K	0,24	0,39	-0,07	-0,19	0,18	0,28	0,26	1,00	0,32	0,19	0,30	0,30	0,01	0,02	0,17	0,15	-0,02	0,06	0,03
HCO_3	0,51	0,83	-0,01	-0,34	0,80	0,57	0,51	0,32	1,00	0,30	0,30	0,45	0,02	0,02	0,14	0,10	0,01	-0,02	0,08
SO_4	0,19	0,68	-0,04	-0,23	0,48	0,84	0,44	0,19	0,30	1,00	0,20	0,33	0,00	0,00	0,02	0,00	-0,03	0,07	0,00
NO_3	0,36	0,48	-0,17	0,00	0,43	0,39	0,14	0,30	0,30	0,20	1,00	0,38	-0,10	0,03	0,24	0,10	-0,10	0,03	0,05
Cl	0,47	0,72	-0,20	-0,32	0,52	0,48	0,56	0,30	0,45	0,33	0,38	1,00	0,02	0,02	0,15	0,07	0,03	0,11	0,07
As	0,03	0,02	-0,10	-0,22	-0,03	-0,01	0,05	0,01	0,02	0,00	-0,10	0,02	1,00	0,01	-0,04	-0,03	0,69	0,03	-0,02
Cd	0,06	0,02	-0,04	-0,04	0,02	0,01	0,01	0,02	0,02	0,00	0,03	0,02	0,01	1,00	-0,01	0,04	0,01	0,02	0,59
Cr	0,11	0,16	0,01	0,06	0,09	0,15	0,07	0,17	0,14	0,02	0,24	0,15	-0,04	-0,01	1,00	0,01	-0,04	-0,02	-0,01
Cu	0,20	0,09	-0,14	-0,04	0,09	0,04	0,03	0,15	0,10	0,00	0,10	0,07	-0,03	0,04	0,01	1,00	-0,03	0,03	0,12
Fe	0,04	0,01	-0,18	-0,24	-0,03	-0,03	0,02	-0,02	0,01	-0,03	-0,10	0,03	0,69	0,01	-0,04	-0,03	1,00	0,04	0,02
Ni	0,07	0,06	-0,20	-0,16	0,01	0,04	0,06	0,06	-0,02	0,07	0,03	0,11	0,03	0,02	-0,02	0,03	0,04	1,00	0,15
Zn	0,14	0,07	-0,04	-0,06	0,08	0,04	0,02	0,03	0,08	0,00	0,05	0,07	-0,02	0,59	-0,01	0,12	0,02	0,15	1,00

7.2.2 *Hydrogeological characteristics of the sampling points*

The following steps were followed to derive the hydrogeological characteristics for each monitoring point:

1) Determine **HOVER and BRIDGE lithology classes** for each monitoring point. Consider HOVER lithology for further data processing. The data basis was derived from the national hydrogeological map at the scale 1 : 500.000.
2) Determine **Stratigraphy classes** for each monitoring point. The data basis was derived from the national hydrogeological map at the scale 1 : 500.000.
3) Determine **HOVER pH and HOVER Redox** classes from median values (pH, NO_3, Fe, O_2, SO_4) for each monitoring point.
4) Determine **HOVER water families** by merging **HOVER lithology, HOVER pH and HOVER Redox classes**.
5) It is not possible to share information on **type of sources (A1)** and depths of **monitoring points (A3)** within this project. **Aquifer types (A7)** are generalized as "unconfined", despite influences of confined groundwaters are possible in rare occasions.
6) Data is treated as a nationwide dataset and not subdivided on a **watershed or basin scale**. Instead, further focus is given to most dominant **HOVER water family** classes.

HOVER water family classes: 45 HOVER water family classes are found within the provided dataset (see Table 52 and

Figure 45).

Most dominant are **gravel aquifers with oxic and neutral pH conditions** (806 monitoring points), followed by **gravel aquifers with oxic and basic pH conditions** (265 monitoring points). **Clay and/or marl aquifers with oxic and neutral pH conditions** (159 montoring points) and **other sedimentary aquifers with oxic and basic water conditions** (141 monitoring points) are also present in larger numbers. The term "sedimentary: other" comprises a larger number of lithologies, see Table 52.

Table 51: HOVER water family classes within the Austrian dataset.

Nr.	Abbreviation	Explanation
1	CR_A_O	Crystalline rocks; Acidic; Oxic water
2	CR_A_SR	Crystalline rocks; Acidic; Strongly reduced
3	CR_A_US	Crystalline rocks; Acidic; Unclassified samples
4	CR_A_WRW	Crystalline rocks; Acidic; Weakly reduced water
5	CR_B_O	Crystalline rocks; Basic; Oxic water
6	CR_N_O	Crystalline rocks; Neutral; Oxic water
7	CR_N_US	Crystalline rocks; Neutral; Unclassified samples
8	MR_A_O	Metamorphic rocks; Acidic; Oxic water
9	MR_A_US	Metamorphic rocks; Acidic; Unclassified samples
10	MR_B_O	Metamorphic rocks; Basic; Oxic water
11	MR_N_O	Metamorphic rocks; Neutral; Oxic water
12	MR_N_US	Metamorphic rocks; Neutral; Unclassified samples
13	MR_N_WRW	Metamorphic rocks; Neutral; Weakly reduced water
14	SCA_A_O	Sedimentary: carbonates (limestone, chalk); Acidic; Oxic water
15	SCA_B_O	Sedimentary: carbonates (limestone, chalk); Basic; Oxic water
16	SCA_N_O	Sedimentary: carbonates (limestone, chalk); Neutral; Oxic water
17	SCM_A_O	Sedimentary: clays and/or marls; Acidic; Oxic water
18	SCM_A_SR	Sedimentary: clays and/or marls; Acidic; Strongly reduced
19	SCM_A_US	Sedimentary: clays and/or marls; Acidic; Unclassified samples
20	SCM_A_WRW	Sedimentary: clays and/or marls; Acidic; Weakly reduced water
21	SCM_B_O	Sedimentary: clays and/or marls; Basic; Oxic water
22	SCM_B_US	Sedimentary: clays and/or marls; Basic; Unclassified samples
23	SCM_N_O	Sedimentary: clays and/or marls; Neutral; Oxic water
24	SCM_N_SR	Sedimentary: clays and/or marls; Neutral; Strongly reduced
25	SCM_N_US	Sedimentary: clays and/or marls; Neutral; Unclassified samples
26	SCM_N_WRW	Sedimentary: clays and/or marls; Neutral; Weakly reduced water
27	SG_A_O	Sedimentary: gravel; Acidic; Oxic water
28	SG_A_SR	Sedimentary: gravel; Acidic; Strongly reduced
29	SG_A_US	Sedimentary: gravel; Acidic; Unclassified samples
30	SG_A_WRW	Sedimentary: gravel; Acidic; Weakly reduced water
31	SG_B_O	Sedimentary: gravel; Basic; Oxic water
32	SG_B_SR	Sedimentary: gravel; Basic; Strongly reduced
33	SG_B_US	Sedimentary: gravel; Basic; Unclassified samples
34	SG_B_WRW	Sedimentary: gravel; Basic; Weakly reduced water
35	SG_N_O	Sedimentary: gravel; Neutral; Oxic water
36	SG_N_SR	Sedimentary: gravel; Neutral; Strongly reduced
37	SG_N_US	Sedimentary: gravel; Neutral; Unclassified samples
38	SG_N_WRW	Sedimentary: gravel; Neutral; Weakly reduced water
39	SO_A_O	Sedimentary: other; Acidic; Oxic water
40	SO_B_O	Sedimentary: other; Basic; Oxic water
41	SO_N_O	Sedimentary: other; Neutral; Oxic water
42	SO_N_US	Sedimentary: other; Neutral; Unclassified samples
43	UN_A_O	Unclassified samples; Acidic; Oxic water
44	UN_B_O	Unclassified samples; Basic; Oxic water
45	UN_N_O	Unclassified samples; Neutral; Oxic water

HOVER and BRIDGE lithology and stratigraphy classes: BRIDGE and HOVER lithology classes were derived from the hydrogeological map of Austria 1 : 500.000 (Schubert et al., 2003). The content translation of the national hydrogeological map is feasible in most cases, but several fields remain unclassified due to missing entry options (see Stratigraphy and BRIDGE lithology) or complex conditions that can't be explained with the vocabulary to choose from (see Table 52). Most dominant are Quaternary gravels (BRIDGE: Glacial sand and gravel deposits; HOVER: Sedimentary gravel; Stratigraphy: Quaternary) with 1.360 monitoring sites, followed by Neogene marls and clays (n= 227) (see Figure 45). The classification might differ between BRIDGE and HOVER lithology depending on the available list entries. For example, crystalline rocks are the only option to choose from if metamorphic rocks are not available. This causes the misleading result, that more crystalline rocks (n= 168) are present when BRIDGE lithology is considered compared to n= 64 for HOVER lithology.

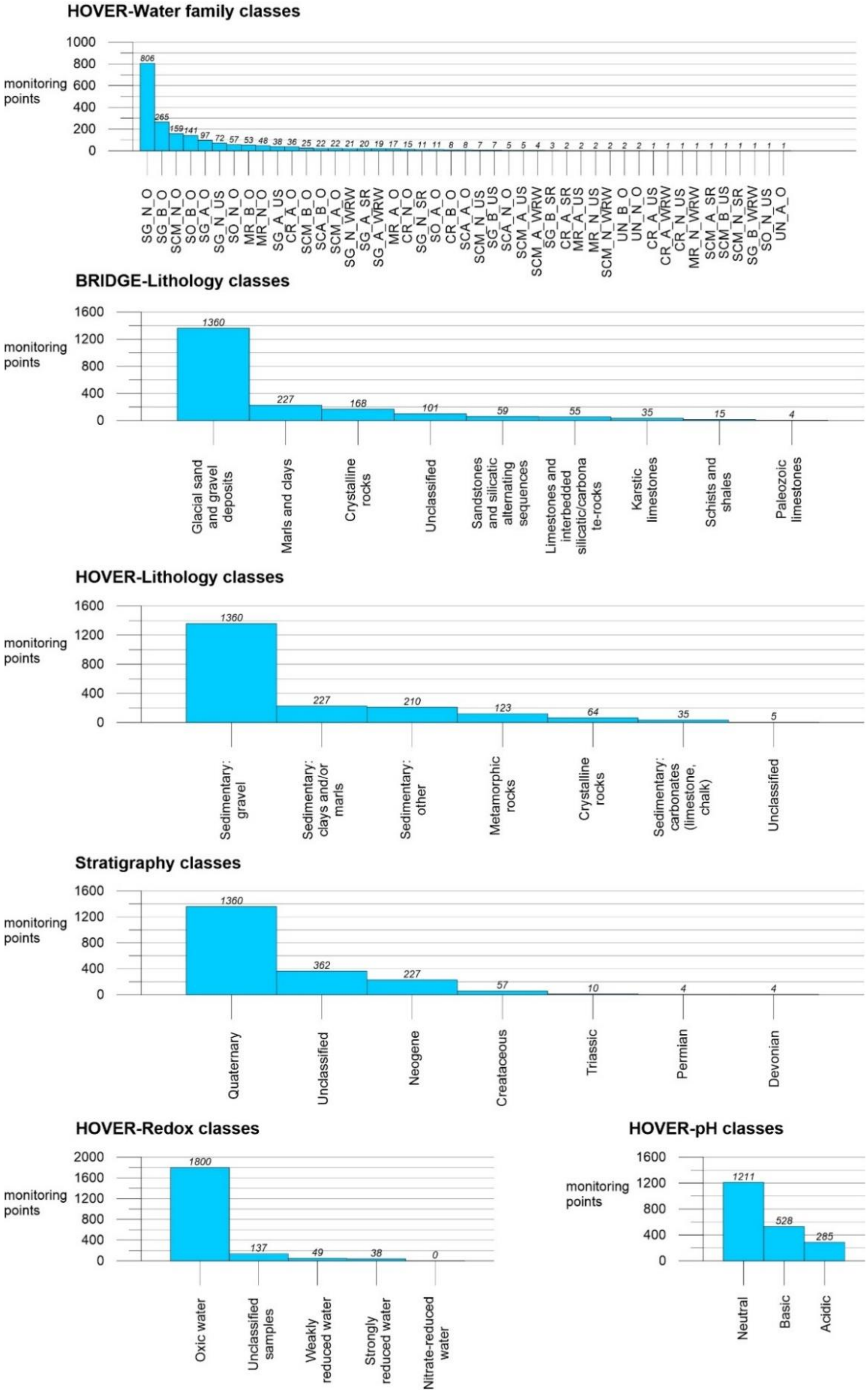
Table 52: Derived Stratigraphy, HOVER and BRIDGE lithology classes from the Hydrogeological map of Austria 1 : 500.000 (Schubert et al. 2003).

Considered legend entries of the hydrogeological map of Austria 1 : 500.000	Stratigraphy	BRIDGE Lithology	HOVER Lithology
Aquifers in which flow is mainly intergranular - Extensive and highly productive aquifers - Mainly gravel and sand	Quaternary	Glacial sand and gravel deposits	Sedimentary: gravel
Aquifers in which flow is mainly intergranular - Local or discontinuous productive aquifers or extensive but only moderately productive aquifers - Mainly gravel and sand, locally moraine	Quaternary	Glacial sand and gravel deposits	Sedimentary: gravel
Karstifiable aquifers - Extensive and highly productive aquifers - Limestone	-	Karstic limestones	Sedimentary: carbonates (limestone, chalk)
Karstifiable aquifers - Local or discontinuous productive aquifers or extensive but only moderately productive aquifers - Mainly carbonate rock	-		Sedimentary: other
Karstifiable aquifers - Local or discontinuous productive aquifers or extensive but only moderately productive aquifers - Dolomite	-		Sedimentary: other
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly carbonate rock	Devonian	Paleozoic limestones	Metamorphic rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly clay, marl and sand, locally gravel, sandstone and conglomerate	Neogene	Marls and clays	Sedimentary: clays and/or marls
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly marl and sandstone	Cretaceouse	Sandstones and silicatic alternating sequences	Sedimentary: other
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly slate and sandstone, locally with gypsum	Triassic	Limestones and interbedded silicatic/carbonate-rocks	Sedimentary: other
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly slate and sandstone, in the Northern Calcareous Alps locally with gypsum and halite	Triassic	Sandstones and silicatic alternating sequences	Sedimentary: other
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly slate and sandstone, locally metamorphic volcanic rock	-	Schists and shales	Metamorphic rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly greenschist, metamorphic black shale and calcareaous schist	Cretaceouse	Schists and shales	Metamorphic rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly quarzite and phyllite	Permian	Schists and shales	Metamorphic rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly phyllite and metamorphic volcanic rock	-	Limestones and interbedded silicatic/carbonate-rocks	Sedimentary: other
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Mainly paragneiss, micaschist and amphibolite, locally phyllite	-	Crystalline rocks	Metamorphic rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Granulite	-	Crystalline rocks	Metamorphic rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Granitoid, anatexite, migmatite, migmatic paragneiss, orthogneiss and tonalite	-	Crystalline rocks	Crystalline rocks
Aquifers (porous, fissured or karstified) with local and limited groundwater resources - Vulcanic rock	Neogene	Vulcanic rocks	Vulcanic rocks

HOVER-pH and Redox classes: For HOVER-Redox, most dominant are oxic waters (n= 1.800) followed by unclassified samples (n= 137), weakly reduced waters (n= 49) and strongly reduced waters (n= 38).

Furthermore nitrate reduced waters are not present. Most monitoring points show neutral pH conditions (n= 1.211), followed by basic (n= 528) and acidic (n=285).

Figure 45: Evaluation of the Austrian dataset regarding HOVER-Water families, HOVER- and BRIDGE-Lithology, Stratigraphy, HOVER-Redox and HOVER-pH.



HOVER-Water family classes

BRIDGE-Lithology classes

HOVER-Lithology classes

Stratigraphy classes

HOVER-Redox classes

HOVER-pH classes

7.2.3 *Anthropogenic pressures / influences : diffuse or point pollutions*

Corine Land Cover Data (CLC 2018) was used to determine potential anthropogenic pressures and influences. It shows that anthropogenic pressure is present in 78,6 % of all groundwater monitoring points. Most prevailing is agriculture (45,9 %), followed by urban (27,2 %), industrial (5,1 %) and mining (0,5 %).

Further focus is given to the most important HOVER water families of the dataset. It shows that again agricultural and urban are the most common pressures. Only a small number of monitoring points show no influence of anthropogenic pressure and correlations between specific pressures and median NO_3 concentrations are questionable.

- gravel aquifers – oxic –neutral pH: 7,8 % of 806 monitoring points show no pressures.
- gravel aquifers – oxic – basic pH: 21,5 % of 265 monitoring points) show no pressures.
- clay and/or marl aquifers – oxic – neutral pH: 5 % of 159 monitoring points show no pressures.

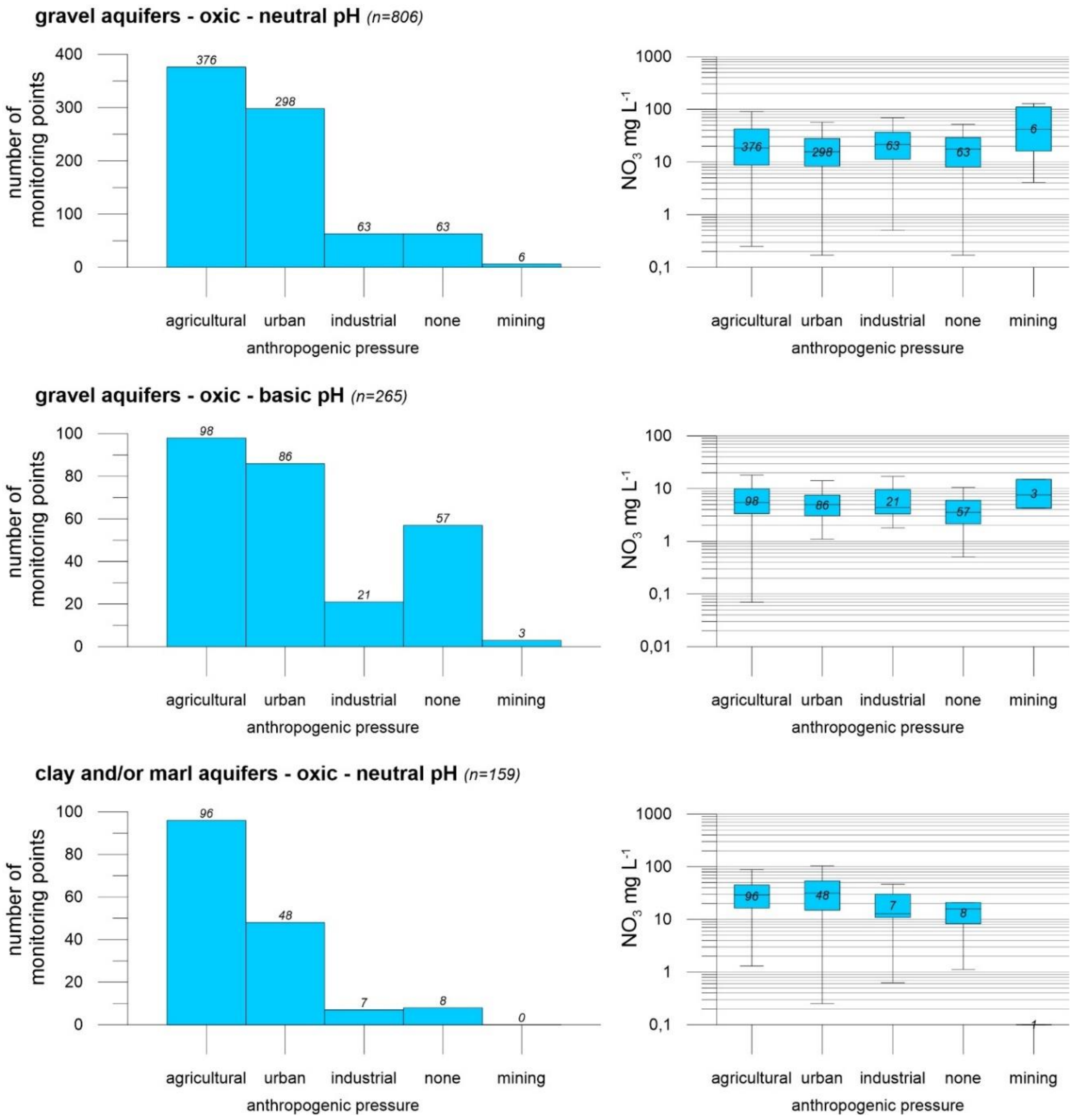


Figure 46: Anthropogenic pressures derived from Corine Land Cover Data (CLC 2018) for the most important HOVER water families of the dataset and the distribution of median NO_3 concentration for each pressure.

7.2.4 *Statistical treatments*

Dataset

First, a clean dataset with the following parameters and median values was created for further statistical data treatment.

- ID (num)
- EEA-cellcode
  (1 km grid) (txt)
- HOVER Water family class (txt)
- HOVER Redox (txt)
- HOVER pH (txt)
- BRIDGE Lithology (txt)
- HOVER Lithology (txt)
- Stratigraphy (txt)
- Depth of water source (data not shared)
- Type of aquifer (data not shared)
- Type of Pressure (derived from CLC2018) (txt)
- Anthropogenic pressure (yes/no) (txt)
- T (°C) (num)
- EC (25 °C) (num)
- pH (num)
- O_2 (mg L^{-1}) (num)
- Ca (mg L^{-1}) (num)
- Mg (mg L^{-1}) (num)
- Na (mg L^{-1}) (num)
- K (mg L^{-1}) (num)
- HCO_3 (mg L^{-1}) (num)
- SO_4 (mg L^{-1}) (num)
- NO_3 (mg L^{-1}) (num)
- Cl (mg L^{-1}) (num)
- As (mg L^{-1}) (num)
- Cd (mg L^{-1}) (num)
- Cr (mg L^{-1}) (num)
- Cu (mg L^{-1}) (num)
- Fe (mg L^{-1}) (num)
- Ni (mg L^{-1}) (num)
- Zn (mg L^{-1}) (num)

Discriminant analyses

The aim was to discriminate HOVER lithologies of most prevailing HOVER water families found in the dataset by chosen hydrochemical compounds. Water families that are present with less than 50 monitoring points or show unclassified compounds (either lithology, redox or pH) were excluded.

The following HOVER lithologies and HOVER water families were considered as qualitative variables:

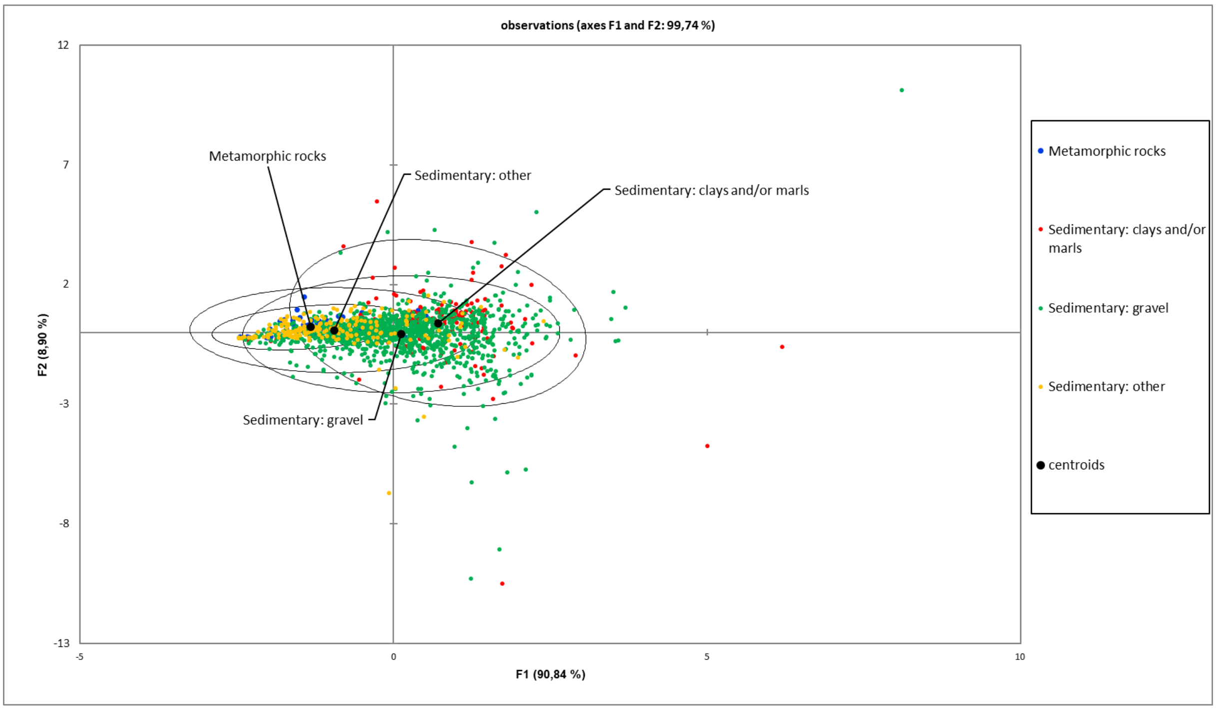
- Sedimentary gravels of SG_N_O, SG_B_O and SG_A_O (n=1.168 monitoring points)
- Sedimentary clays and/or marls of SCM_N_O (n=159)
- Sedimentary others of SO_B_O, and SO_N_O (n=198)
- Metamorphic rocks of MR_B_O (n=53)

First, the main cations Na, K, Ca and Mg were considered as quantitative variables. The confusion matrix shows, that metamorphic rocks are well classified (85 %) (see Table 53). This is also the case for sedimentary gravels (78 %), despite a large number of reclassifications to metamorphic rocks (18 %). Sedimentary clays and/or marls are commonly wrongly reclassified to sedimentary gravels (91 %). Finally, sedimentary others show 0 % of well classified monitoring points. The overall rate of correct predictions is 61 %. The chart of the monitoring points on the factor axes F1 and F2 is provided in Figure 47 and shows large overlaps.

Table 53: Confusion matrix of the discriminant analysis with the qualitative variables (sedimentary gravels, sedimentary clays and/or marks, sedimentary others and metamorphic rocks) and the quantitative variables Na, K, Ca and Mg.

from / to	**Metamorphic rocks**	**Sedimentary: clays and/or marls**	**Sedimentary: gravel**	**Sedimentary: others**	**True classification (n)**	**% correct prediction**
Metamorphic rocks	45	0	8	0	53	85%
Sedimentary: clays and/or marls	3	9	145	2	159	6%
Sedimentary: gravel	207	21	911	29	1168	78%
Sedimentary: other	135	2	61	0	198	0%
Total	390	32	1125	31	1578	61%

Figure 47: Monitoring points on factor axes F1 and F2 of the discriminant analysis with the qualitative variables (sedimentary gravels, sedimentary clays and/or marks, sedimentary others and metamorphic rocks) and the quantitative variables Na, K, Ca and Mg

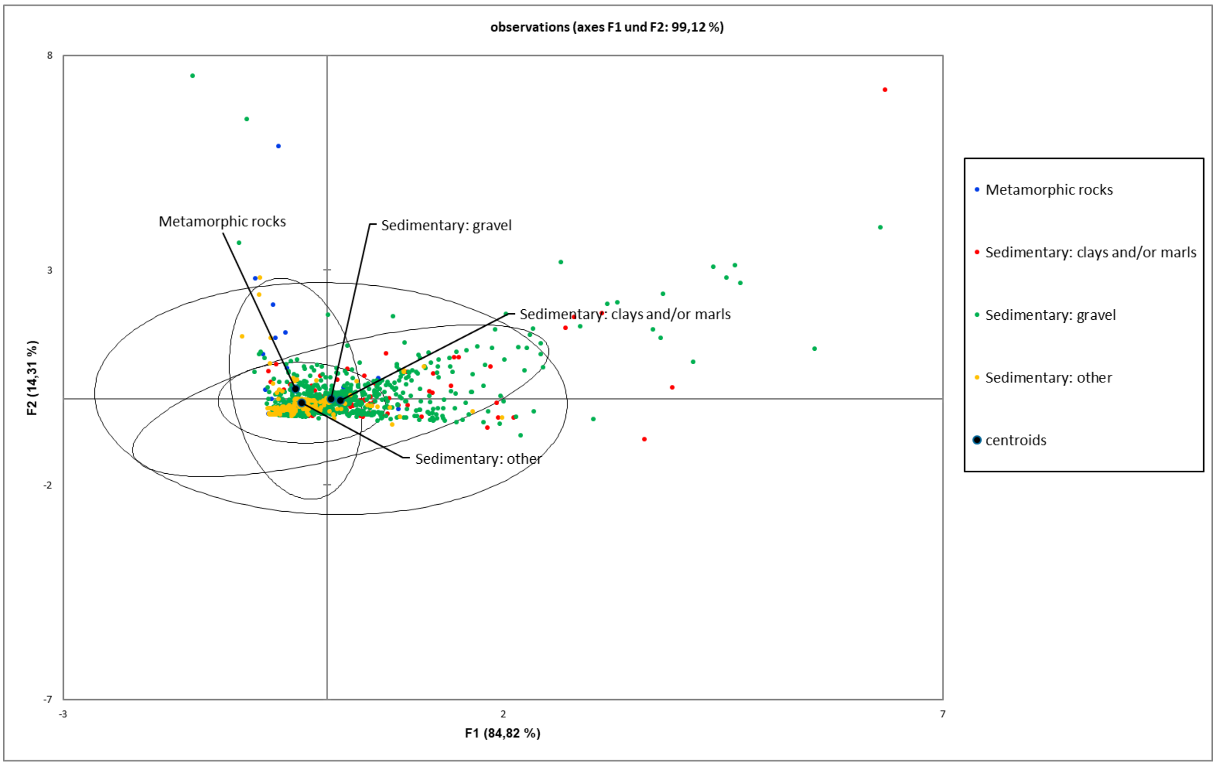


In the second approach, SO_4, Fe, As and Ni were considered as quantitative variables. The confusion matrix shows, that sedimentary others are very well classified (91 %). However, the other HOVER lithologies are commonly wrongly reclassified to sedimentary others. The overall rate of correct predictions is only 25 %, see Table 54, and the chart of the monitoring points on the factor axes F1 and F2 is provided in Figure 48 and illustrates the uncertainties.

Table 54: Confusion matrix of the discriminant analysis with the qualitative variables (sedimentary gravels, sedimentary clays and/or marks, sedimentary others and metamorphic rocks) and the quantitative variables SO_4, Fe, As and Ni.

from / to	Metamorphic rocks	Sedimentary: clays and/or marls	Sedimentary: gravel	Sedimentary: other	True classification (n)	% correct prediction
Metamorphic rocks	5	1	4	43	53	9%
Sedimentary: clays and/or marls	2	12	31	114	159	8%
Sedimentary: gravel	5	59	197	907	1168	17%
Sedimentary: other	3	3	11	181	198	91%
total number	15	75	243	1245	1578	25%

Figure 48: Monitoring points on factor axes F1 and F2 of the discriminant analysis with the qualitative variables (sedimentary gravels, sedimentary clays and/or marks, sedimentary others and metamorphic rocks) and the quantitative variables SO_4, Fe, As and Ni.



Kruskal-Wallis test

Kruskal-Wallis tests were carried out to determine, if monitoring points under the influence of anthropogenic pressures have different concentrations of chosen major, minor and trace elements.

Analogous to the discriminant analyses, a preselection of monitoring points was based on prevailing HOVER water families within the dataset.

- Sedimentary gravels of SG_N_O, SG_B_O and SG_A_O (n=1.168 monitoring points)
- Sedimentary clays and/or marls of SCM_N_O (n=159)
- Sedimentary others of SO_B_O, and SO_N_O (n=198)
- Metamorphic rocks of MR_B_O (n=53)

The results show that the p-values for the considered elements NO_3, SO_4, Cl As, Cd, Cr, Cu, Ni and Zn are all below 0,05. This means that the H0 hypothesis must be rejected and significant differences are found.

Table 55: Results from the Kruskal-Wallis test the influence of prevailing pressures.

	Kruskal-Wallis test (p value)	Explanation
NO_3	< 0,0001	The samples don't come from the same population (p<0,05)
SO_4	< 0,0001	
Cl	< 0,0001	
As	1,16E-04	
Cd	< 0,0001	
Cr	< 0,0001	
Cu	< 0,0001	
Ni	< 0,0001	
Zn	< 0,0001	

Dunn test with Bonferroni correction

Based on the result of the Kruskal-Wallis test, Dunn tests with Bonferroni correction were performed for each element to identify the differences between the types of prevailing pressures.

Results show that monitoring points without any prevailing anthropogenic pressure commonly show differences to monitoring points with prevailing pressures (see Table 56).

This entirely true for:

- Nitrate, Chloride, Copper, Zinc
- Sulfate, Nickel, Cadmium (except mining)
- Chromium (except agriculture and mining)

However, there is no difference observed for Arsenic. The low number of monitoring points found in areas influenced by mining have to be kept in mind, this might cause wrong interpretations.

Table 56: Results from the Dunn tests with Bonferroni correction.

Dunn test with Bonferri correction		significant differences				
		agricultural	industrial	urban	mining	none
NO_3	agricultural	no	no	no	no	yes
	industrial	no	no	no	no	yes
	urban	no	no	no	no	yes
	mining	no	no	no	no	yes
	none	yes	yes	yes	yes	no
SO_4	agricultural	no	yes	no	no	yes
	industrial	yes	no	no	no	yes
	urban	no	no	no	no	yes
	mining	no	no	no	no	no
	none	yes	yes	yes	no	no
Cl	agricultural	no	yes	yes	no	yes
	industrial	yes	no	yes	no	yes
	urban	yes	yes	no	no	yes
	mining	no	no	no	no	yes
	none	yes	yes	yes	yes	no
As	agricultural	no	yes	yes	no	no
	industrial	yes	no	no	no	no
	urban	yes	no	no	no	no
	mining	no	no	no	no	no
	none	no	no	no	no	no
Cd	agricultural	no	no	no	no	yes
	industrial	no	no	no	no	yes
	urban	no	no	no	no	yes
	mining	no	no	no	no	no
	none	yes	yes	yes	no	no

Dunn test with Bonferri correction		significant differences				
		agricultural	industrial	urban	mining	none
Cr	agricultural	no	**yes**	**yes**	no	no
	industrial	**yes**	no	no	no	**yes**
	urban	**yes**	no	no	no	**yes**
	mining	no	no	no	no	no
	none	no	**yes**	**yes**	no	no
Cu	agricultural	no	no	**yes**	no	**yes**
	industrial	no	no	no	no	**yes**
	urban	**yes**	no	no	no	**yes**
	mining	no	no	no	no	**yes**
	none	**yes**	**yes**	**yes**	**yes**	no
Ni	agricultural	no	no	no	no	**yes**
	industrial	no	no	no	no	**yes**
	urban	no	no	no	no	**yes**
	mining	no	no	no	no	no
	none	**yes**	**yes**	**yes**	no	no

7.2.5 *Removal of element concentrations influenced by contamination and NBL calculation*

NBL calculation was performed for:

- Selected trace and minor elements (As, Cd, Cr, Cu, Ni and Zn).
- HOVER lithologies of most prevailing HOVER water families found in the dataset.
- Water families that are present with less than 50 monitoring points or water families that show unclassified compounds (either lithology, redox or pH) were excluded.

The removal of element concentrations is only linked to the results of the Dunn test with Bonferroni correction.

- As: no removal of element concentrations.
- Cr: removal of element concentrations that are influenced by industrial and urban pressure.
- Cd, Ni: removal of element concentrations that are influenced by agricultural, industrial and urban pressure.
- Cu removal of element concentrations that are influenced by agricultural, industrial, urban and mining pressure.

Table 57: NBL calculation (90th percentiles) of selected minor and trace elements.

		Sedimentary: gravel		**Sedimentary: clays and/or marls**		**Sedimentary: other**		**Metamorphic rocks**	
element	*unit*	*value*	*n*	*value*	*n*	*value*	*n*	*value*	*n*
As	(µg L-1)	**1,05**	1168	**0,12**	159	**0,50**	198	**3,38**	53
Cd	(µg L-1)	**0,05**	135		<20	**0,05**	130	**0,04**	35
Cr	(µg L-1)	**0,6**	673	**2,37**	104	**0,5**	173	**0,5**	50
Cu	(µg L-1)	**2,15**	126		<20	**0,5**	130	**0,5**	34
Ni	(µg L-1)	**0,5**	135		<20	**0,5**	130	**0,5**	35

7.3 Summary and outlook

The provided dataset of the Austrian groundwater monitoring network comprises 2.024 monitoring points and 605.353 values of single parameters. The aim of this monitoring network is to observe surface near groundwater bodies. In detail, the vast majority of monitoring points can be linked to unconfined intergranular Quaternary gravel and sand aquifers.

This is reflected in the HOVER lithology classes, 1.360 monitoring points are attributed as “Sedimentary: gravel” followed by Neogene “Sedimentary: clays and/or marls” (n=227) and “Sedimentary: others” (n=210). “Sedimentary: others” cover a large number of lithologies including dolomites, carbonate rocks in general, slate and sandstone. Metamorphic rocks (n=123) comprise mainly quarzites, phyllites and shists, crystalline rocks (n=64) in most cases granitoids, migmatites and gneisses. It has to be added that the geological information for this task was entirely derived from the hydrogeological map of Austria (1 : 500.000) via spatial join.

The majority of monitoring points feature oxic conditions (n=1.800), followed by unclassified samples (n=137), weakly reduced (n=49) and strongly reduced (n=38). The predominant pH class is neutral (n=1.211), then basic (n=528) and acidic (n=285).

45 HOVER water family classes are found in the provided dataset. Most dominant are gravel aquifers with oxic and neutral pH conditions (n= 806), followed by gravel aquifers with oxic and basic pH conditions (n=265), clay and/or marl aquifers with oxic and neutral pH conditions (n=159) and other sedimentary aquifers with oxic and basic water conditions (n=141).

Corine Land Cover Data (CLC 2018) was used to determine potential anthropogenic pressures and influences. It shows that anthropogenic pressure is present at 78,6 % of all groundwater monitoring points. Most prevailing is agriculture (45,9 %), followed by urban (27,2 %), industrial (5,1 %) and mining (0,5 %).

Discriminant analyses were performed to discriminate HOVER lithologies (qualitative variables) of most prevailing HOVER water families present in the dataset by combinations of chosen main, trace and minor elements (quantitative variables). However, results show that distinctions are linked to noticeable uncertainties and should be avoided.

Kruskal-Wallis tests proved that monitoring points influenced by anthropogenic pressures (agricultural, industrial, urban, mining) show different concentrations of chosen major, minor and trace elements.

Dunn tests with the Bonferroni correction describe the influences of each pressure/element more detailed:

- **Arsenic:** No difference is observed between monitoring points without anthropogenic pressure and monitoring points with influence.
- **Nitrate, chloride, copper:** Monitoring points without any prevailing anthropogenic pressure show differences to monitoring points influenced by all pressures.
- **Sulfate, Nickel and Cadmium:** Monitoring points without any prevailing anthropogenic pressure show differences to all monitoring points influenced by pressures except mining. However, the low number of monitoring points attributed with "Mining" is likely to lead to a false interpretation.
- **Chromium:** Monitoring points without any prevailing anthropogenic pressure show differences to all monitoring points influenced by except agricultural and mining.

The removal of element concentration influenced by contamination is entirely based on the results of the statistical treatment. Results of the NBL calculation (90 % percentile) are found in Table 57. The Kruskal-Wallis test and Dunn test support the assumption, that using Corine Land Cover Data is a valid approach to define anthropogenic pressures.

However, Corine Land Cover information extracted for point data via spatial joins can't provide any information on a catchment scale. For example, monitoring points of this dataset that are not affected by anthropogenic pressures according to CLC2018 show elevated NO_3 levels in some cases. It would be reasonable to exclude those monitoring points, since they are most likely influenced by anthropogenic pollution.

8 CASE STUDY – DUERO RIVER BASIN, SPAIN

8.1 Legislative framework

In accordance with the provisions contemplated in the Royal Decree 1514/2009, of 2th October (last modified 16th December 2015), "**reference level"** means the concentration of a substance or the value of an indicator in a Groundwater Body (GWB) corresponding to no, or only very minor, anthropogenic alterations to undisturbed conditions.

On the other hand, **"basic level"** means the average value measured at least during the reference years 2007 and 2008, based on monitoring programmes implemented under Article 8 of Directive 2000/60/EC. In the case of substances identified after these reference years, the first period in which representative monitoring data exist is chosen as the basis for the average value calculation. It is required to define the baseline when relevant risks are identified during the process of characterization as set out by the Water Framework Directive.

To evaluate the chemical status of a GWB or group of GWBs, quality standards are used to determine its condition. **"Threshold values"** are groundwater quality standards set accordingly with the criteria regulated in article 3 of Royal Decree 1514/2009. Threshold values may be established at the level of River Basin Demarcation or of Groundwater Body or, where appropriate, at the State level.

It is possible to draw several conclusions from the above definitions. First, **Natural Background Levels** are similar to reference levels in Spanish legislation. Second, when human pressure does exist, and therefore, natural background levels are certainly difficult to achieve, the basic level is the tool to make a decision about a GWB chemical status.

8.2 Groundwater management

Spain has 25 river basin districts, out of which 6 are international sharing water courses with France to the northeast and Portugal to the west. River Basin Management Plans depend on the relevant Competent Authorities (Molinero et al., 2011): When a given hydrological basin is interregional, water management depends on the Ministry for the Ecological Transition and the Demographic Challenge of the Government of Spain. When the hydrological basin is entirely located in an Autonomous Community, water management is responsibility of the Regional Autonomous Government.

Groundwater bodies are complex systems and show very variable physical-chemical properties (Ballesteros et al, 2001). From a scientific point of view, therefore, it does not make sense to define one single concentration for an aquifer as a whole (Blum et al, 2009). However, referring to the European Directives requirements, using average concentrations is helpful.

Most River Basin Authorities in Spain have focused on the development of threshold values as part of the reporting requirements to the European Commission for an extensive evaluation of the existing policy. The establishment of these national standards is a first important step of the implementation of the Groundwater Directive at European level.

8.3 Available information

8.3.1 *Previous studies*

There exists an unequal degree of hydrogeological information among the GWs subject to exploitation and those of lesser entity (Fernandez-Ruiz, 2012), which have a significant lack, and even absence, of basic piezometric and quality data.

The greatest shortcomings in this regard come from information on pressures. While to some extent information to describe diffuse pressures exist, information to analyse point sources is not adequately covered. The reference source of information in this regard (sig.mapama.gob.es) is specifically designed for the treatment of pressures on surface water bodies, while the pressure analysis on groundwater suffers from a serious lack of adequacy. Most of the necessary information is dispersed in diverse and non-homogeneous documentary collections (mainly from the Autonomous Communities).

8.3.2 *Study area selection*

In addition, in Spain, there is a great variety of specificities in the genesis and extension of the different hydrographic basins. In order to verify the methodology for obtaining the background values, the Duero River Basin has been selected (Figure 49).

Figure 49: Overview of the Duero River Basin

8.3.3 *Groundwater database*

The starting point for the constitution of the data set that will serve as the basis of the study is the Official Control Network of Groundwater Chemical Status of the River Duero Basin.

In accordance with the provisions of the Water Framework Directive (DMA), it aims to provide a reliable assessment of the qualitative status of all water bodies (or groups) and protected areas. The dataset contains 303008 records and the variables from Table 58 . The information has been extracted on October 2020.[4]

---

[4] https://www.chduero.es/web/guest/red-control-estado-quimico

Variable	Data type	Explanation
Code	Character	Monitoring station identification
MASub	Character	Identification of Groundwater Body
TM	Character	Municipality
Provincia	Character	Province
UTM X	Numeric	Coordinate X UTM
UTM Y	Numeric	Coordinate Y UTM
Year	Numeric	Year
Date	date	Date with format "Y-m-d hh:mm:ss"
Par	Character	Measured parameter
Par2	Character	Abbreviation of the parameter
Valor	Numeric	Measurement value including less than
Valorn	Numeric	Measurement value substituting less than by 0
Unidad	Character	Units of measurement
QL	Character	Quantitation limit (when available)

Table 58: Raw groundwater database

Table 59 shows the number of records for those parameters with at least one thousand measurements.

Parameter	n	Parameter	n
Alkalinity	5133	Fluorides	4859
Aluminium	3571	Phosphates	5950
Ammonium	8841	Iron	5796
Dissolved carbon dioxide (field)	1062	Dissolved iron	2896
Arsenic	5931	Magnesium	8901
Barium	1087	Manganese	8723
Bicarbonates	8900	Nickel	1038
Boron	3508	Nitrates	8901
Cadmium	5353	Nitrites	8900
Calcium	8901	Orthophosphates	2896
Carbonates	8901	Dissolved oxygen (field)	15504
Total organic carbon	2498	pH (field)	8579
Chlorides	8902	pH (lab)	6007
Copper	7035	Lead	1863
Faecal coliforms	2224	Potassium	8902
Total coliforms	2240	Atmospheric pressure	2315
Electrical conductivity (field)	8429	Selenium	1087
Electrical conductivity (lab)	6005	Silica	7707
Chromium	1044	Sodium	8893
Biochemical Oxygen Demand	1617	Sulphates	8901
Chemical Oxygen Demand	1124	Ambient temperature (field)	4528
Hardness	5646	Water temperature (field)	8362
Total hardness	2896	Zinc	3611
Faecal streptococci	1916		

Table 59: Number of single records in groundwater database

8.3.4 *Approaches for derivation of NBLs*

As stated in the previous paragraphs, there is not a common methodology in Spain to obtain natural background levels. The basic level, that is to say, the average value measured at least during the reference years 2007 and 2008 is just a starting point. Any analysis should be done at monitoring station level and the precision of statistical analysis is no suitable to obtain background levels.

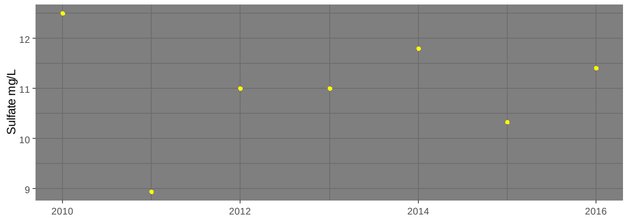


Figure 50: Scatterplot of Sulphate medians 2010-2016

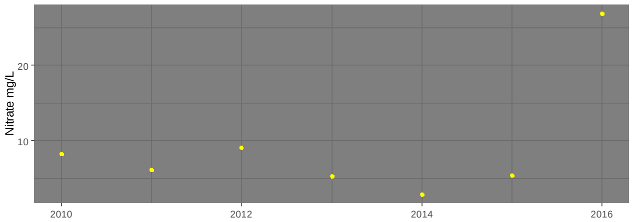


Figure 51: Scatterplot of Nitrate medians 2010-2016

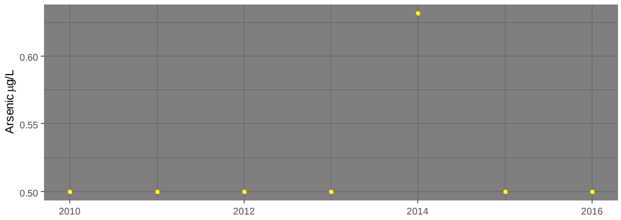


Figure 52: Scatterplot of Arsenic medians 2010-2016

Taking as an example the Duero River Basin, a range of concentrations depending on the different lithology is shown in Figure 53, *Figure 54* and Figure 55.

There is no prior information about point sources or the existing one is very scarce, so it has been no possible to evaluate anthropogenic pollution from the dataset prior to any analyses. The general NBLs were established for sites with a minimum of 10 or 20 water sampling analysis. Groundwater redox conditions have not been considered, as there is just 5 points out of 460 below the threshold of 1 mg/L for Nitrates concentration. Table 60 shows the number of points above and below the threshold in function of the selected Nitrate concentration value. Figure 53, *Figure 54* and Figure 55 show the range of log (concentrations) for Sulphates, Nitrates and Arsenic.

Table 60: Number of sampling points as a function of NO3 concentration threshold

Number of monitoring stations relative to threshold

Threshold	Below	Above
1 mg/L	5	455
5 mg/L	109	351
10 mg/L	193	267

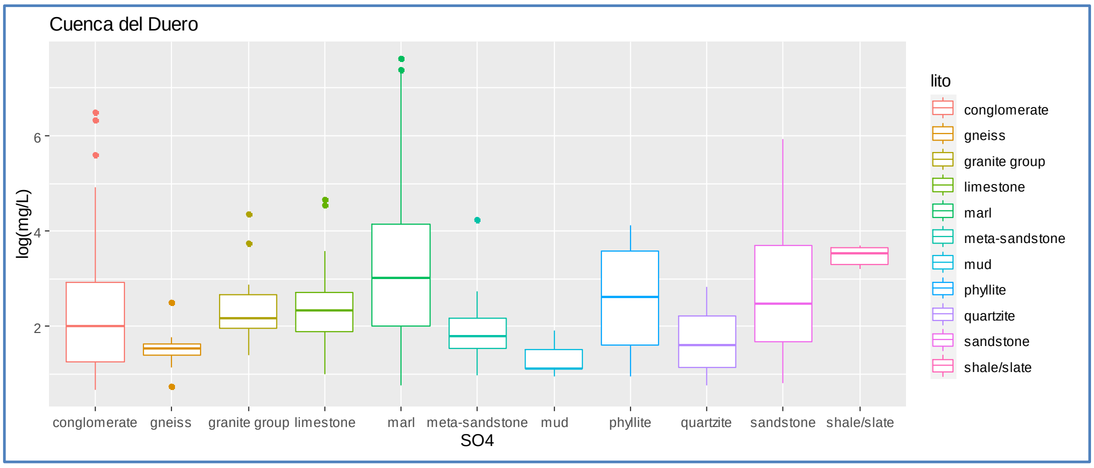


Figure 53: Range of log(concentration) for Sulfates in Duero River Basin depending on the different lithological units defined in IGME (2018)

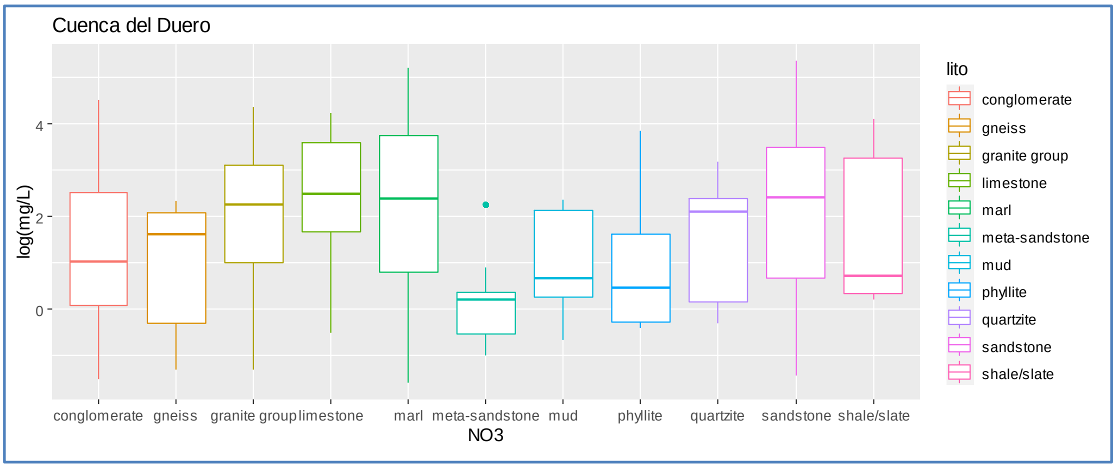


Figure 54: Range of log(concentration) for Nitrates in Duero River Basin depending on the different lithological units defined in IGME (2018)

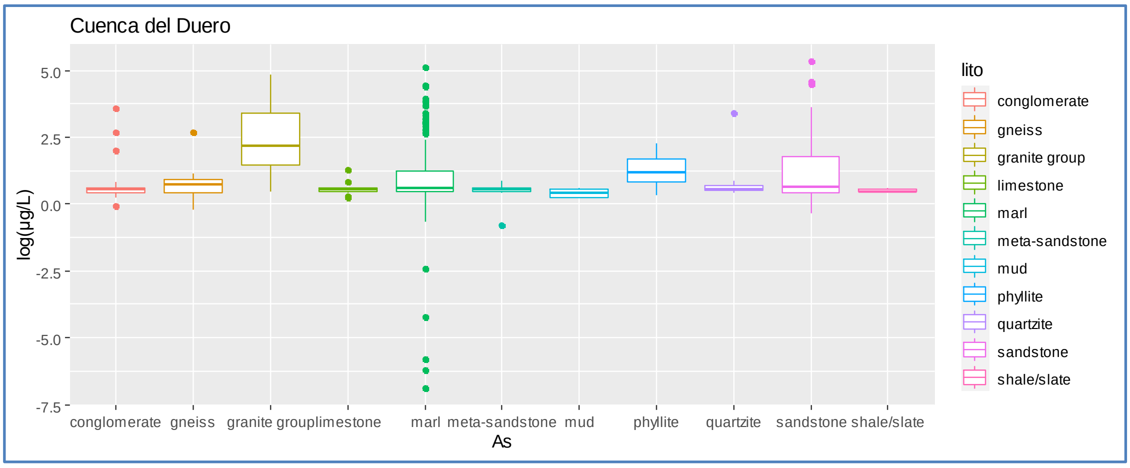


Figure 55: Range of log(concentration) for Arsenic in Duero River Basin depending on the different lithological units defined in IGME (2018)

The threshold values adopted in the Spanish Part of the catchment area regarding the pollutants used for the assessment of the chemical status of the GWBs have been calculated taking into account the provisions of article 3 of Royal Decree 1514/2009, of October 2, which regulates the protection of groundwater against contamination and deterioration, are indicated in Table 62.

According to the Guidance Document No. 18 of the Common Implementation Strategy for the Water Framework Directive (2000/60/EC), NBLs are calculated as the 90th percentile of the mean concentration at the official monitoring network stations. The computation is carried out during the period 2010-2020. As regards Arsenic, there are insufficient data to perform the analysis.

Percentile	**Stations selection criteria**		
	All	> 10 records	> 20 records
	Sulphates		
0.8	59.8	57.3	43.6
0.9	**186**	**194**	**122**
0.95	499	513	316
0.975	922	928	727
0.98	931	942	836
0.99	1142	1156	980
	Nitrates		
0.8	38.2	39.5	45.3
0.9	**66.5**	**69.8**	**74**
0.95	89.8	91.8	91.8
0.975	125	125	124
0.98	130	139	126
0.99	167	168	158

Table 61: Percentiles for Sulphates and Nitrates depending on the Station selection criteria

GWB CODE	**GWB name**	**Parameter**	**Selection criteria**	**Threshold value (mg/L)**
400038	Tordesillas	Ammonium	RD 140/2003	0,5
400045	Los Arenales			
400047	Medina del Campo			
400052	Salamanca			
400055	Cantimpalos			
400045	Los Arenales	Arsenic	Percentile 97,7	0,140
400047	Medina del Campo			0,079
400052	Salamanca			0,047
400053	Vitigudino		Percentile 90	0,204
400058	Campo Charro			0,027
400063	Ciudad Rodrigo			0,630
400031	Villafáfila	Chlorides	Percentile 97,7	303
400038	Tordesillas			441
400031	Villafáfila	Sodium	Percentile 98	295
400038	Tordesillas			522
400045	Los Arenales			1.040
400067	Terciario bajo Páramos			404
400016	Castrojeríz	Sulfates	Percentile 97,7	456
400045	Los Arenales			1.108
400067	Terciario bajo Páramos			1.548

Table 62: Threshold values of some parameters in Duero River Basin

8.3.5 *Discriminant function analysis*

In order to analyse differences between lithological groups defined in IGME, a linear discriminant function has been performed. The LDA method assumes that the predictor variables are normally distributed and that the classes have identical co-variances.

All our predictor variables are left-skewed, thus failing the normality assumption. In this situation it not possible to find the optimal solution, but still LDA is a pretty robust technique. We also scaled and centered the data prior to analysis, because the LDA can be affected by the units of the variables. The data is split in training (80%) and test set (20%), so that an LDA model can be built with the training set and tested on the test set.

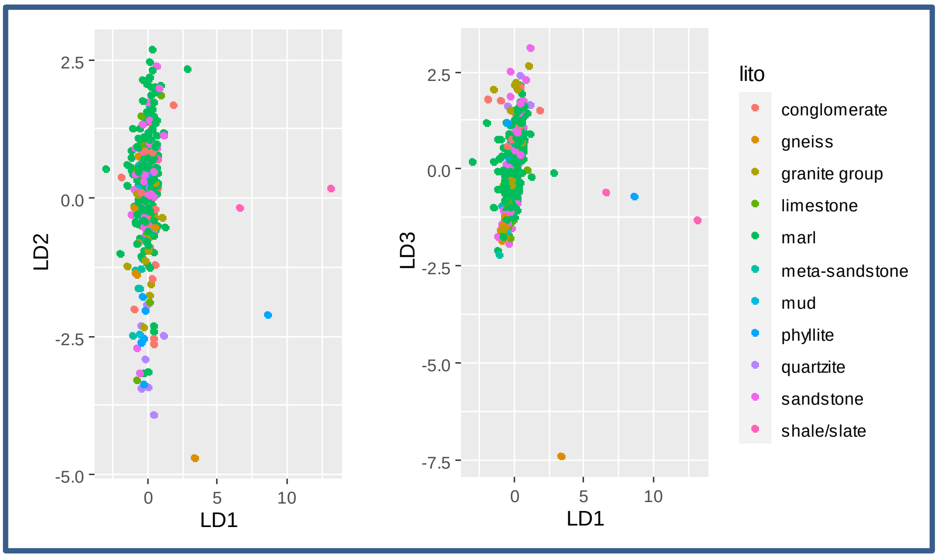


Figure 56: LDA applied to IGME lithological classes

8.4 Study with the proposed method

8.4.1 *Constitution of the groundwater quality dataset*

The starting point for the constitution of the data set that will serve as the basis of the study is the Official Control Network of Groundwater Chemical Status of the River Duero Basin[5]. In accordance with the provisions of the Water Framework Directive (DMA), it aims to provide a reliable assessment of the qualitative status of all water bodies (or groups) and protected areas. The information has been extracted on October 2020.

From this raw dataset, samples from stations completely or partially dedicated to drinking water production by public waterworks are used ("Abastecimiento", "Abastecimiento y Ganadero", "Abastecimiento de refuerzo (toman del río)", "Abastecimiento (verano)", "Abastecimiento y riego").

Summary of HOVER WP 3-3 raw dataset:

- Period: 2010-2020 (incl. both years)
- Water sampling points: waterworks wells (at intake level)

---

[5] https://www.chduero.es/web/guest/red-control-estado-quimico

- Contains chemical analyses for parameters listed in 14.4.1. It should be noted, however, that a single dataset containing all the information is provided; so many elements are measured in just a single monitoring station and/or in one or several periods.
- Data-pretreatment included:
    - various element-specific quality checks,
    - treatment of all values below the limit of detection (< LOD)
    - aggregation on sampling point level.

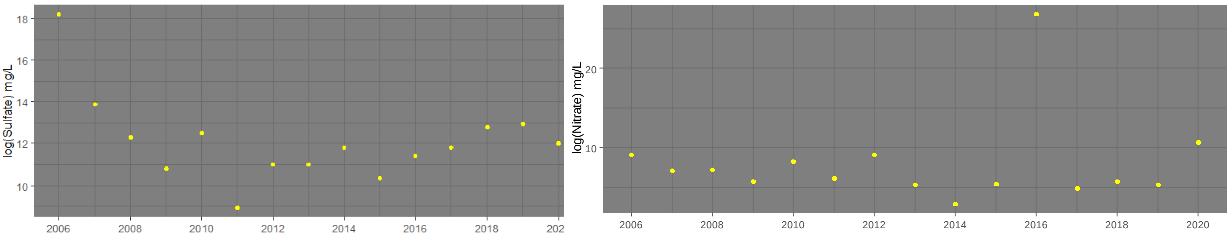


Figure 57: Available data of median values for Sulphates and Nitrates in Duero River Basin

8.4.1.1 Element-specific quality checks

Measurements of pH <2 and >12 and temperature measurements <2 °C and >25 °C were considered erroneous and excluded. Field measurements for pH and dissolved oxygen were used when available; otherwise, lab measurements were used, if available.

Treatment of values below the limits of detection (LOD). The chemical laboratories report values below the limit of detection (LOD) with attribute “<”. All values < LOD were substituted with ½ LOQ.

Table 19 shows the number of chemical analyses for each element, as well as an account of the LODs present in the dataset, including the range of LODs, the number and percentage of analyses < LOD, the max substitute value and the number of analyses with the max substitute.

Available information for Nickel and Chromium is very limited, with data ranging from 2011 to 2013.

Zinc data exist in the monitoring network during the period 2010-2016, but there are no data during 2014 and there is only one measurement (below the LOQ) in 2016.

Dissolved oxygen measurements are not considered reliables (there are values up to 6400 mg/L and no data in year 2015).

8.4.1.2 Aggregation

There were 465 samples. The aggregation on water sampling point level was calculated based on a median over the eleven-year period of this study (2010-2020).

Table 63: Number and percentage of chemical analyses below the limit of detection (LOD), range of LODs in the dataset, max substitute value and number of analyses with this max substitute; the last column shows the total number of chemical analyses after all data-pretreatment procedures, used in the aggregation step to calculate median for each water sampling point

Element	unit	All (n)	Period	< LOQ (n)	< LOQ (%)	LOQ range (min-max)	Min	Max	Mean	Med	SD
SO_4	mg/l	6341	10-20	908	14.65	1-10	1	2822.2	99.58	15.1	265.2
As	µg/l	4234	10-20	2642	62.40	0.1-5	0.102	801	15.94	5.25	41.68
Cd	mg/l	4456	11-20	3918	87.92	0.02-1	0.02	78	0.80	0.235	3.46
Cr	mg/l	807	11-13	279	34.57	1	1	28	2.49	1	3.02
Cu	mg/l	4332	10-20	2474	57.11	0.001-0.006	0.001	3.59	0.0079	0.0023	0.088
Ni	µg/l	799	11-13	646	80.85	1-7	0.59	19	1.79	1	2.05
Zn	mg/l	1328	10-16	535	40.2	0.002-0.06	0.002	6.06	0.065	0.01	0.35
F	mg/l	3783	10-20	1318	34.84	0.05-0.5	0.01	8.85	0.51	0.29	0.63
Cl	mg/l	6457	10-20	486	7.5	1-10	0.22	1152	45.88	18.65	85.49
O_2	mg/l	3091	10-20	-	-	-	0.04	17.4	5.50	5.63	2.66
NO_3	mg/l	6341	10-20	1317	20.77	0.15-5	0.21	1105	29.17	11.45	47.32
Fe	mg/l	6439	10-20	1935	30.05	0.01-0.08	0.003	132.6	2.34	0.19	7.22
pH	-	5887	10-20	-	-	-	2.11	11	7.67	7.6	0.76
T	°C	5935	10-20	-	-	-	3.60	25	14.48	14.60	3.11

8.4.1.3 Groundwater quality dataset

The groundwater quality dataset contains median concentrations for 14 chemical elements and parameters for 465 water sampling points (intake level). Additionally, the HOVER-pH was calculated and added to the dataset. Table 20 shows the distribution of water sampling points in the HOVER-pH classes. As no EH records are available, there is not information enough to make a classification of the proposed HOVER-redox classes.

The clean groundwater quality dataset was joined with the datasets containing the hydrogeological characteristics for each water sampling point and the dataset with prevailing anthropogenic pressures.

The HOVER-pH and the elemental concentrations will be presented for the final complete dataset. The clean groundwater quality dataset represents all water sampling points where data for at least one of the elements was present. However, for some of these locations, there may not be information about the geological setting.

Table 64 Number of water sampling points with chemical data for each HOVER-pH class

HOVER-pH classes	SO4 (n)	As (n)	Cd (n)	Cu (n)	F (n)	Cl (n)
Acidic (pH < 7)	48	45	44	48	45	48
Basic (pH > 7.5)	279	271	272	276	254	279
Neutral (7 ≤ pH ≤ 7.5)	147	146	142	146	143	147

8.4.2 *Hydrogeological characteristics of the sampling points*

The purpose of this part of the study was to determine the HOVER-age formation/stratigraphy for each of the water sampling points (n=465). To be able to classify most of the water sampling points, we used primarily IGME information, where the well intakes are coupled with a specific aquifer. Table 22 shows how the HOVER lithology has been simplified into five different HOVER-age classifications.

Table 65 Conversion and simplification of HOVER-age and HOVER-lithology

HOVER-lithology	HOVER-age
Sedimentary: sand	Cenozoic
Sedimentary: gravel	Cenozoic
Sedimentary: carbonates (limestone, chalk)	Mesozoic
Sedimentary: clays and/or marls	Cenozoic
Sedimentary: other	Cenozoic
Volcanic rocks	Igneous
Crystalline bedrock	Igneous
Metamorphic rocks	Metamorphic rocks
Others	Others

8.4.3 *Anthropogenic pressures / influences: diffuse or point pollutions*

To calculate the natural background levels (NBL) for the selected chemical elements we must first remove or minimize the influence of anthropogenic contamination. We address this issue by:

- Pre-selection of water sampling points limited only to the waterworks wells used for drinking water supply; Groundwater used for drinking water is generally of high quality, so it is unlikely that these wells are affected by point sources of pollution. In Spanish legislation, the River Basin Authority must close the well or wells affected by point-source pollution and provide another groundwater source complying with the drinking water standards. The main source of diffuse pollution is the agricultural sector.
- Testing the proposed methodology for identifying prevailing anthropogenic influences by using the CORINE landcover map[6]

We used a buffer of 1km around each water sampling point to determine the areal proportion of the land-cover types used in the CORINE land-cover 2018 (CLC-18) map (Figure 18). The 1km buffer is an approximation for the catchment area of the individual wells, where the water sampling locations are. It is possible that the actual catchment areas are either bigger, smaller, or not circular, depending on the local conditions. Since this is a regional-level study, including more than 465 wells (some with multiple water sampling depths), it was decided that the one km-buffer would be an adequate proxy.

The CLC-18 land-cover classes were grouped, resulting in pressures due to Agricultural areas, Forests and semi-natural areas and water bodies. It can be noticed that no anthropogenic pressures are detected in the water points. For the "yes"/"no" variables (Table 25), if any of the listed CLC-18 codes was present within the 1km buffer irrespective of the areal proportion, the corresponding water sampling point was classified with "yes", i.e. the water point is potentially influenced by the specific anthropogenic pressure.

The variable "Prevailing pressure" takes the value "agricultural" based on the areal proportion of these individual anthropogenic pressure. For example, if the CLC-18 codes for agricultural pressure collectively have the largest proportion of the 1km buffer of a water sampling point, this water sampling point is assigned the values "agricultural". If the 1km buffer area was dominated by other CLC-18 codes than the ones listed in Table 25, the water sampling point was assigned the value "natural or other", because it could be any of the codes from Level 1 categories: Forests and semi natural areas, Wetlands, or Water bodies.

---

[6] Corine land cover (CLC) with reference year 2018 for Spain in 1:100000 scale

https:// http://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=SIOSE

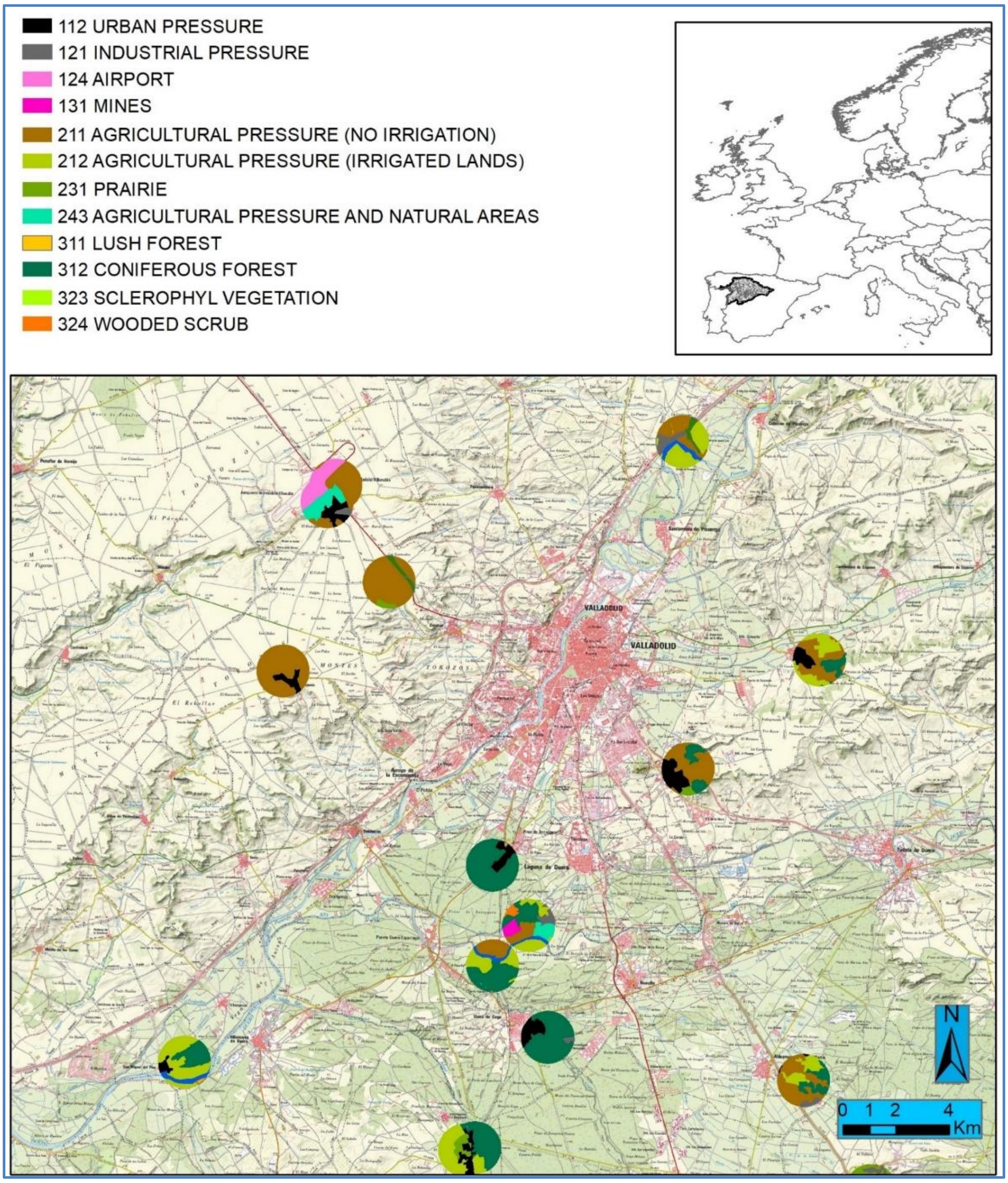


Figure 58 Example of the area surrounding Valladolid (Spain) with 1km buffers around the water sampling points, and the land-cover types (CLC18 grouped by type). The rectangle on the inset map shows the location of the area.

From the 467 water sampling points, 2 could not be classified in the HOVER-lithology class (due to no data), so these were excluded from the dataset. The rest of the water sampling points (n=465) were classified as presented in Table 25,

Table 67, and Table 68. These tables show that the most widespread anthropogenic pressure comes from the agriculture. Nearly 80% of all water sampling points have at least one type of anthropogenic pressure in the 1 km buffer zone around them (agricultural).

Table 66: HOVER WP 33 anthropogenic pressure "yes"/"no" variables

HOVER WP 33 class	Values	CLC-18 codes	CLC-18 meaning	Includes all codes at CLC level
Agricultural Pressure	Yes/no	211 212 231 243 244	Non-irrigated arable land Permanently irrigated land Pastures Land principally occupied by agriculture Agro-forestry areas	**L2. Agricultural areas**
Not applicable	Not applicable	311 312 321 322 323 324	Broad-leaved forest Coniferous forest Natural grassland Moors and heathland Sclerophyllous vegetation Transitional woodland shrub	**L3. Forests and semi-natural areas**
Not applicable	Not applicable	511	Water courses	**L5. Water bodies**

Table 67 Number and percentage of water sampling points in each anthropogenic "yes"/"no" pressure variable

HOVER WP 33 class	**Value**	**Water sampling points (n=465)**	
		n	%
Urban pressure	no	465	100
	yes	0	0.0
Industrial Pressure	no	465	100
	yes	0	0.0
Agricultural Pressure	no	104	22.36
	yes	361	77.63
Mining Influence	no	465	100
	yes	0	0.0

Table 68 Number and percentage of water sampling points in different classes of the variable "Prevailing pressure" (dominating pressure, area-wise)

Prevailing pressure (class)	**Water sampling points (n=465)**	
	n	%
Agricultural	361	77.63
Industrial	0	0.0
Mining	0	0.0
Urban	0	0.0
Natural or other	104	22.36

8.4.4 Master dataset (step 13)

The clean groundwater quality dataset was joined with the datasets with hydrogeological characteristics and the anthropogenic pressures. This dataset is further referred to as "Master dataset" and contains 465 water sampling points and the variables from

Table 69 Master dataset used further in this study

Variable	Data type	Explanation
Code	txt	ID label of the well (can be used for searching in Jupiter)
UTM X	num	Coordinate X UTM
UTM Y	num	Coordinate Y UTM
MASub	txt	Groundwater body number, e.g. " DU-400045 LOS ARENALES"
geology	txt	IGME-model geology, see Table 22
HOVER_lithology	txt	HOVER-lithology
HOVER_age	txt	HOVER-age formation / stratigraphy
urban	txt	Urban pressure ("yes"/"no")
agricultural	txt	Agricultural pressure ("yes"/"no")
prevailing	txt	Prevailing pressure variable
anthropogenic_pressure	txt	Anthropogenic pressure variable
pH	num	Median of pH over the 10-year study period (-)
Temp	num	Median temperature over the 10-year study period (°C)
SO4	num	Median SO4 over the 10-year study period (mg/l)
As	num	Median As over the 10-year study period (µg/l)
Cd	num	Median Cd over 2011-2020 (µg/l)
Cr	num	Median Cr over **2011-2013** (µg/l)
Cu	num	Median Cu over the 10-year study period (µg/l)
Ni	num	Median Ni over **2011-2013** (µg/l)
Zn	num	Median Zn over **2010-2016** (µg/l)
F	num	Median F over the 10-year study period (mg/l)
Cl	num	Median Cl over the 10-year study period (mg/l)
O2	num	Median O2 over the 10-year study period (mg/l)
NO3	num	Median NO3 over the 10-year study period (mg/l)
Fe	num	Median Fe over the 10-year study period (mg/l)
HOVER_pH	txt	HOVER-pH, see Table 20

8.4.5 ***Statistical treatments***

8.4.5.1 Discriminant function analysis (Step 14)

There are different types of discriminant function analysis (DFA) methods, here we chose to perform the simplest one -- linear discriminant analysis (LDA). LDA uses linear combination of predictors to predict the class/category of given observations. The purpose of this analysis is to quantify how well the selected predictor variables (major, minor, trace elements etc.) discriminate between different groups of HOVER-lithology classes. The predictive power of the models is used to compare the different possible combinations of predictor variables.

No missing data is tolerated in this analysis, so the number of data-points is dependent on the chosen predictor variables (Table 30). The trace elements Cr, Ni were excluded from the dataset because There are measurements just for the period between 2011 and 2013. Zn was excluded as well, because data exist in the monitoring network during the period 2010-2016, but there are no data during 2014 and there is only one measurement (below the LOQ) in 2016. Some other elements, like Arsenic, were excluded because of the large number of missing values.

The LDA method assumes that the predictor variables are normally distributed and that the classes have identical co-variances. We scaled and centered the data prior to analysis, because the LDA can be affected by the units of the variables. The data is split in training (80%) and test set (20%), so that an LDA model can be built with the training set and tested on the test set. Table 30 provides details for the LDA models and the number of data-points for each of the models.

We formulated four models for predicting HOVER-lithology classes (classifying data-points). The difference between Models 1 and 4 is that the latter also has pH as a predictive variable.

Table 70 Linear discriminant analysis: model definition and comparison of model performance

LDA models	Predicting	Predictor variables	Accuracy	95% CI
model 1	HOVER-lithology	SO4, As, F, Cl, O2, NO3, Fe	79.52%	(69.24, 87.59)
model 2	HOVER-lithology	SO4, As, F, Cl	83.13%	(73.32, 90.46)
model 3	**HOVER-lithology**	**pH, SO4, As, T, F, Cl, O2, NO3, Fe**	**83.13%**	**(73.32, 90.46)**
model 4	HOVER-lithology	pH, SO4, O2, NO3, Fe	83%	(73 – 90)

The model performance is compared based on overall accuracy (Table 30). All models have nearly similar accuracies (i.e. classify correctly): 80% of all water sampling points for for HOVER-Lithology classes. We show and discuss the results for model 1, 2, 3 and 4.

Model 1:

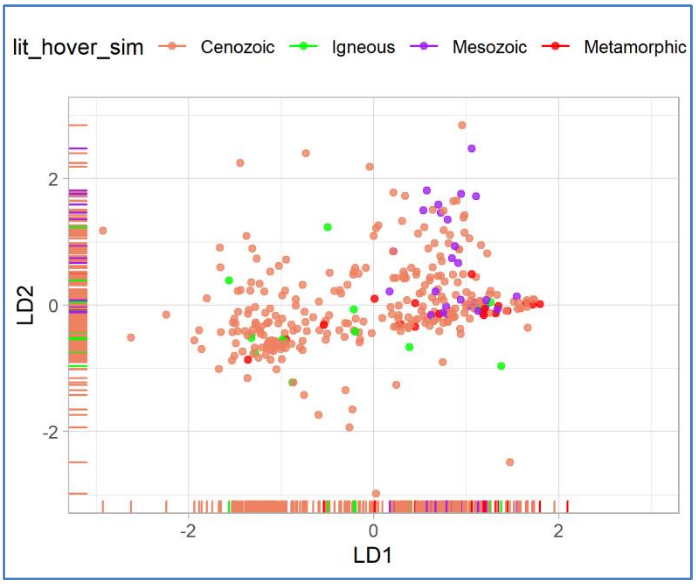


Figure 59: Scatter plot of LDA - Model 1

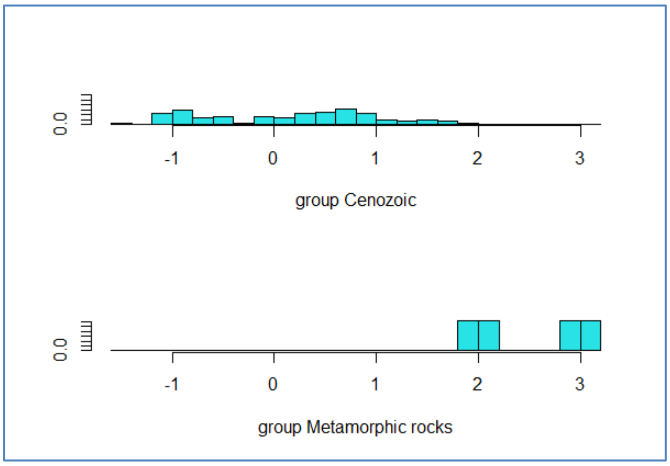


Figure 60: Histogram of Model 1 LDA components

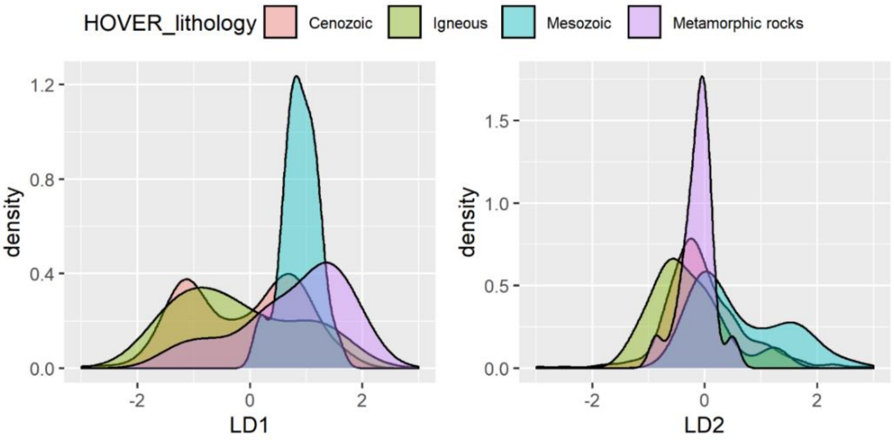


Figure 61 Linear discriminant functions for model 1, where the percentage separation achieved by the discriminant function LD1 is 63.05% and by LD2 is 33.02%

Model 2:

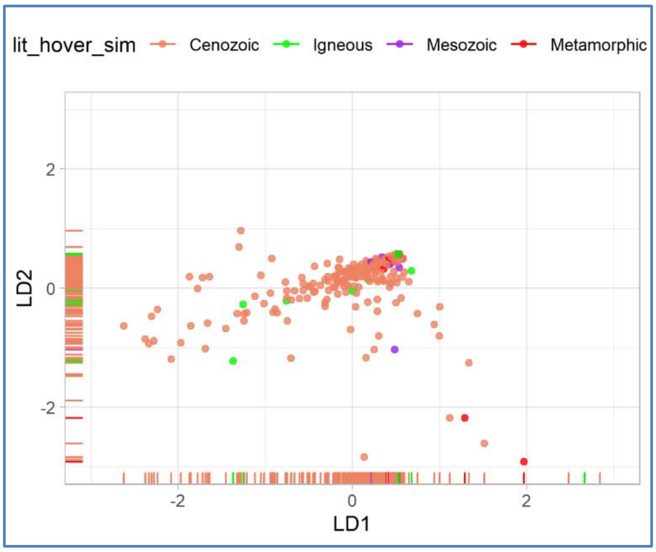


Figure 62: Scatter plot of LDA - Model 2

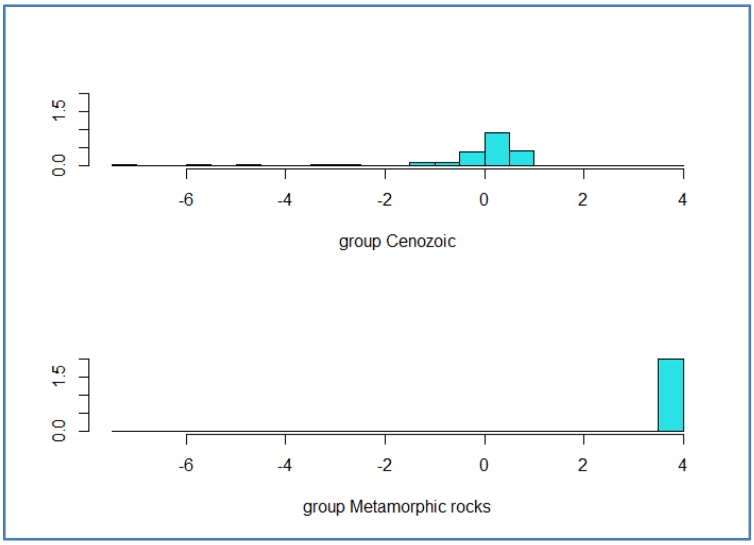


Figure 63: Histogram of Model 2 LDA components

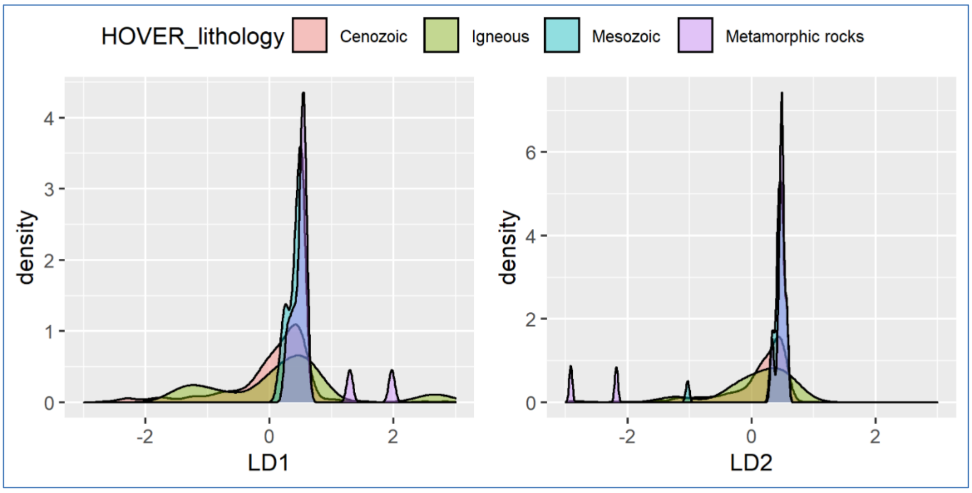


Figure 64: Linear discriminant functions for model 2, where the percentage separation achieved by the discriminant function LD1 is 70.76% and by LD2 is 23.39%

Model 3:

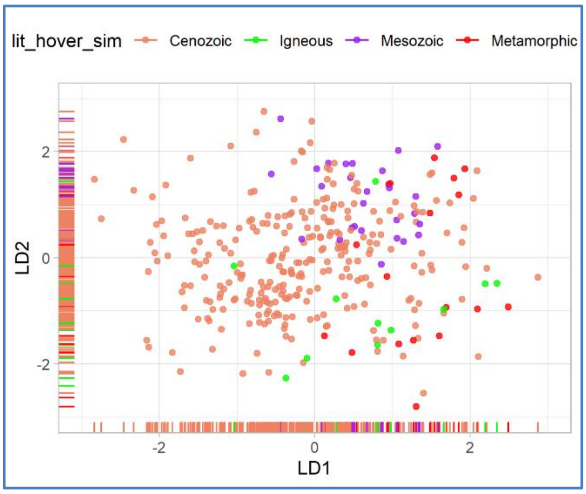


Figure 65: Scatter plot of LDA - Model 3

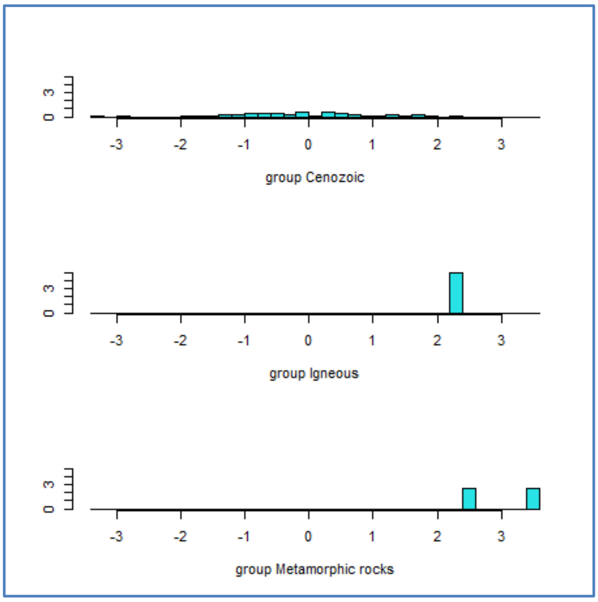


Figure 66: Histogram of Model 3 LDA components

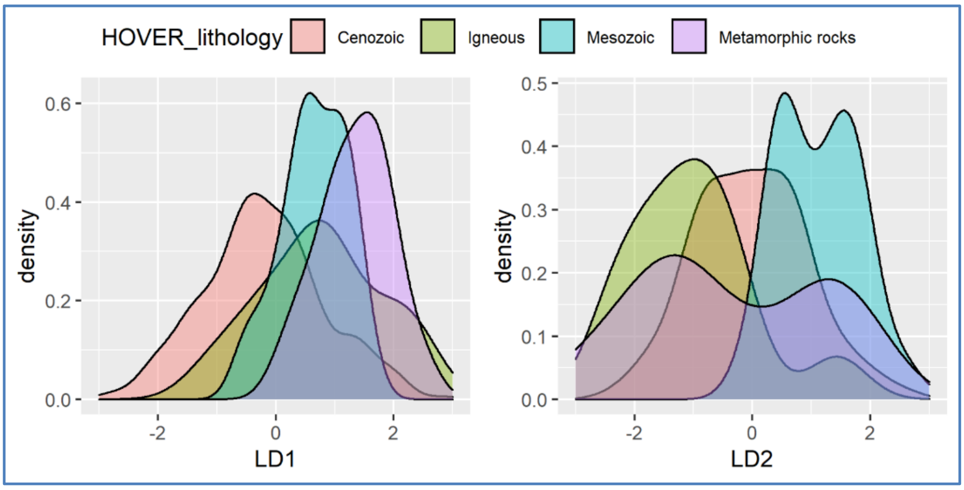


Figure 67: Linear discriminant functions for model 3, where the percentage separation achieved by the discriminant function LD1 is 57.90% and by LD2 is 35.22%

Model 4:

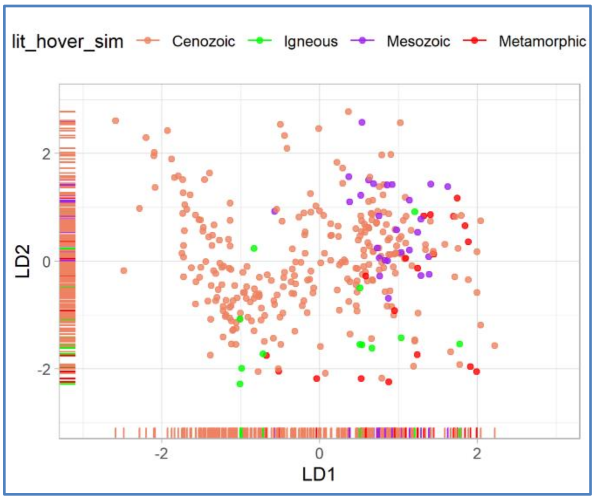


Figure 68: Scatter plot of LDA - Model 4

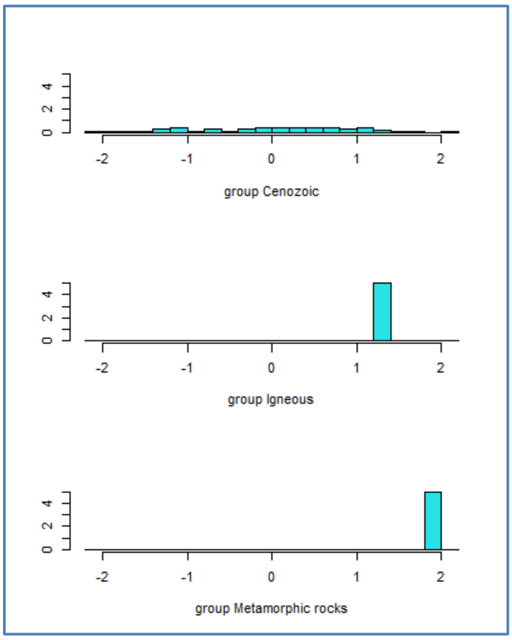


Figure 69: Histogram of Model 4 LDA components

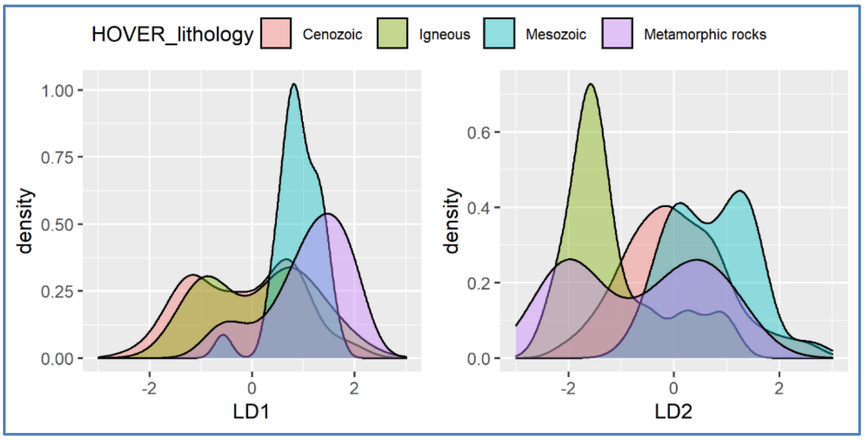


Figure 70: Linear discriminant functions for model 3, where the percentage separation achieved by the discriminant function LD1 is 57.90% and by LD2 is 35.22%

The results show that a better discrimination between types of geology/lithology is achieved when only the two HOVER-lithology classes are used.

The conclusion of this discriminant function analysis is that based on the selected elements (pH, SO4, As, Ni, Fl, Cl, O2, NO3, Fe) there is significant difference between the two classes of simplified HOVER-lithology ("Cenozoic" and "Metamorphic rocks"). Therefore, it was decided that the 90th percentiles will be calculated for these classes only.

8.4.5.2 Prevailing pressures (*Step 15*)

Kruskal-Wallis rank sum test is a non-parametric method for testing if two or more groups originate from the same distribution. The alternative hypothesis is that at least one group differs from the rest. To determine which group(s) differ, we performed pairwise comparisons using Nemenyi-test (as a post-hoc) with Chi-squared approximation for independent samples (to account for ties).

Statistical significance is assessed at the 95% confidence level, i.e. significant differences are found if $p < 0.05$. For this test, we used the categorical variable "Prevailing pressure". The purpose of this analysis is to test if there is significant difference in the concentrations at sampling points with different prevailing anthropogenic pressures.

Table 71 Number of water sampling points with data for each element and each group of prevailing pressure and results from the Kruskal-Wallis rank sum test;

Element	all	agricultural		no or other		Kruskal-Wallis rank sum test
	n	n	%	n	%	
SO4	465	361	77.63	104	22.36	Significant ($p < 0.05$)**
As	460	358	22.17	102	77.82	Not significant ($p > 0.05$)
Cd	449	353	21.38	96	78.62	Not significant ($p < 0.05$)
Cr	-	-	-	-	-	-
Cu	461	358	22.34	103	77.66	Not significant ($p < 0.05$)
Ni	-	-	-	-	-	-
Zn	-	-	-	-	-	-
F	440	338	76.82	102	23.18	Significant ($p < 0.05$)**
Cl	465	361	77.63	104	22.37	Significant ($p < 0.05$)**

Table 31 shows that most of the water sampling points are with dominating agricultural pressures (range 21.38-77.63 % of all locations). This means that within a buffer of 1km around the sampling point, the largest proportion of land is with agricultural land-use. It should also be noted that the pre-selection of sampling locations, which included only waterworks wells used for drinking water supply, might be reflected in the absence of points with industrial, mining, or urban influences. The waterworks wells are usually in areas where there are no known point sources of pollution. The Kruskal-Wallis results showed that Sulphates, Fluorides and Chlorides have at least one group significantly different (Table 31). As there are only two groups it is equivalent to a Wilcoxon rank sum test.

The "agricultural group" is used here as the background state to which all the rest of the groups are compared. We do this, because, as discussed previously, the agricultural pressure dominates the dataset (~80% of all water sampling points are potentially affected) and the "natural group" has less than 1% of the points. We discuss the application of this methodology with the Spanish dataset and propose potential future improvements in section 5.3.

8.4.5.3 HOVER-pH importance

We used the same methodology (Kruskal-Wallis with post-hoc Nemenyi) to test if the groups based on HOVER-pH are significantly different. We performed this analysis to determine if and for which of the elements it is relevant to include pH grouping when calculating the 90th percentiles.

Table 32 shows that more than half of the data points are with basic pH (pH ≥7.5). About 1/3 is with neutral (7< pH < 7.5) and around 10% are acidic (pH < 7).

Thus, the dataset is biased towards neutral and basic groundwaters.

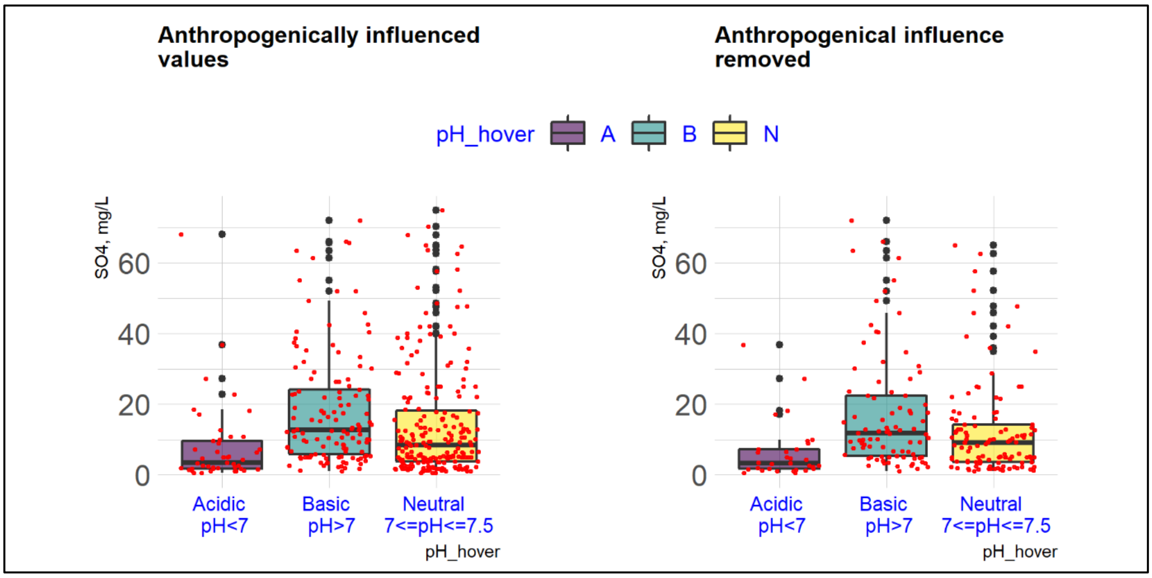
Table 72 Number and percentage of data-points (without removing Anthropogenical influence) within each of the HOVER-pH groups, and results from the Kruskal-Wallis rank sum test

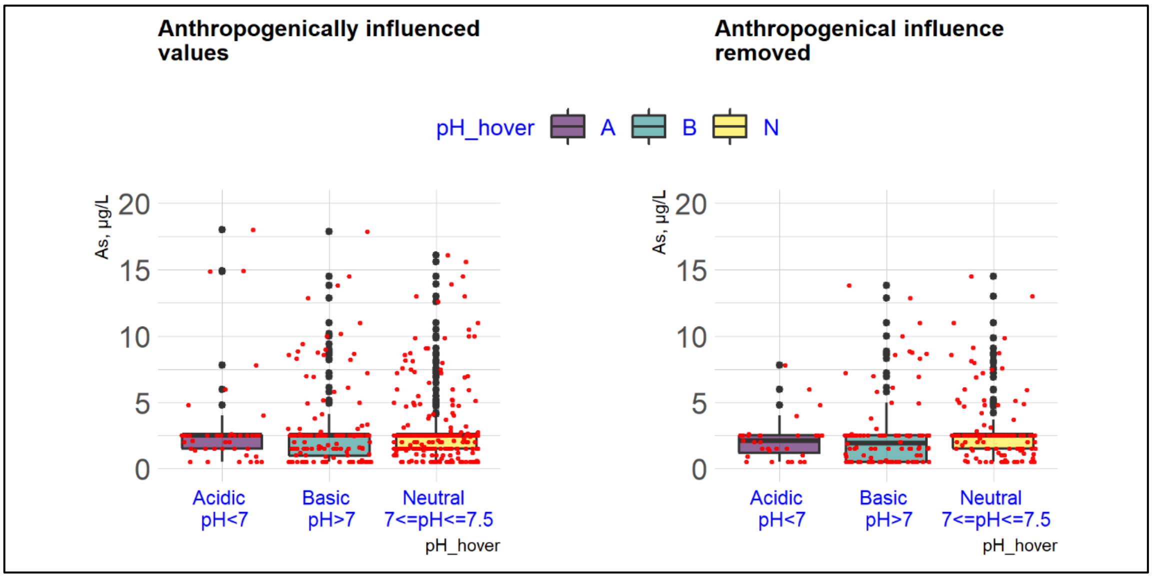
Anthropogenically influenced values								
Elements	All	HOVER-pH categories						Kruskal-Wallis rank sum test
		Acidic		Basic		Neutral		
	n	n	%	n	%	n	%	
SO_4	465	48	10.12	279	58.86	147	31.01	Significant difference(p < 0.05)
As	462	45	9.74	271	58.65	146	31.60	Not significant (p > 0.05)
Cd	458	44	9.61	272	59.39	142	31.00	Not significant (p > 0.05)
Cr	-	-	-	-	-	-	-	Records from 2011 to 2013
Cu	465	48	10.21	276	58.72	146	31.06	Not significant (p > 0.05)
Ni	-	-	-	-	-	-	-	Records from 2011 to 2013
Zn	-	-	-	-	-	-	-	Records from 2010 to 2016
F	442	45	10.18	254	57.46	143	32.35	Significant difference(p < 0.05)
Cl	465	48	10.12	279	58.86	147	31.01	Significant difference(p < 0.05)

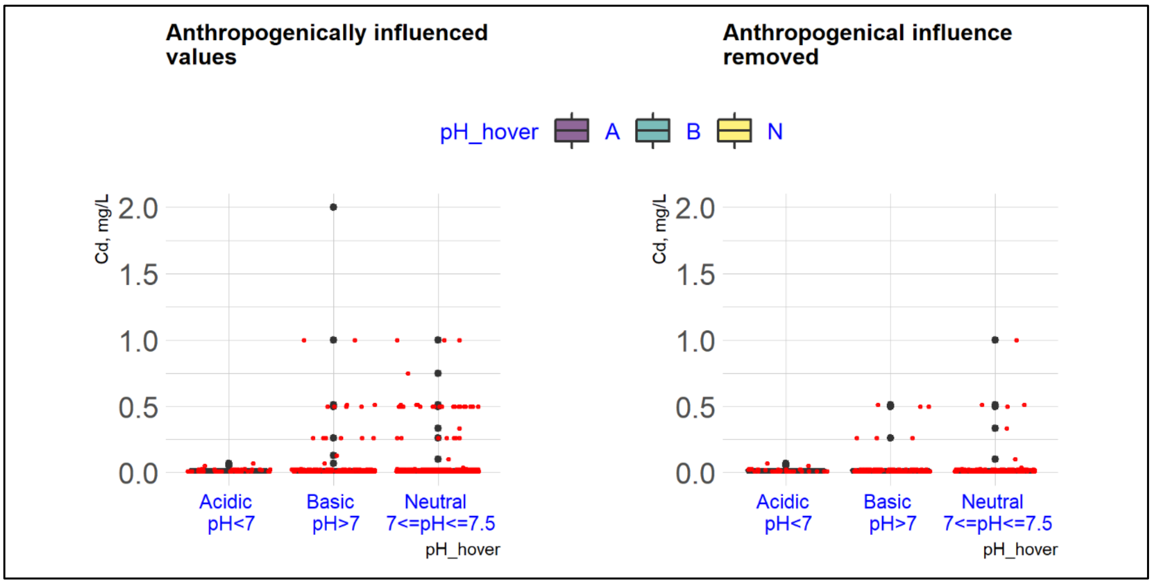
Table 73: Number and percentage of data-points (water sampling points) within each of the HOVER-pH groups, and results from the Kruskal-Wallis rank sum test

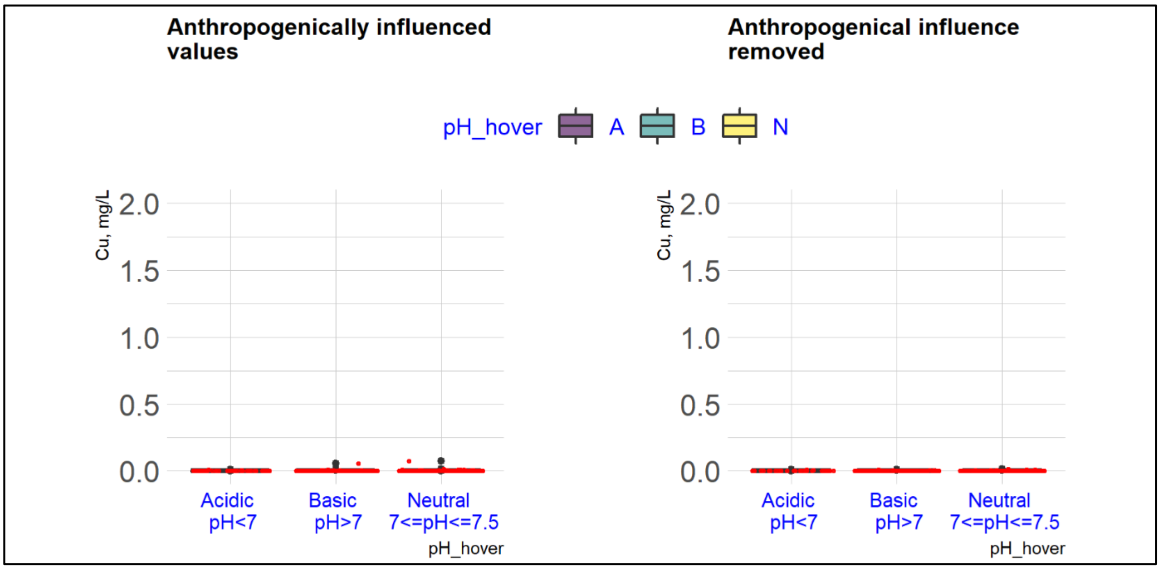
Anthropogenical influence removed								
Elements	All	HOVER-pH categories						Kruskal-Wallis rank sum test
		Acidic		Basic		Neutral		
	n	n	%	n	%	n	%	
SO_4	257	32	12.45	129	50.19	96	37.35	Significant difference(p < 0.05)
As								Significant difference(p < 0.05)
Cd								Not significant (p > 0.05)
Cr	-	-	-	-	-	-	-	-
Cu	257	32	12.45	129	50.19	96	37.35	
Ni	-	-	-	-	-	-	-	-
Zn	-	-	-	-	-	-	-	-
F	251	30	11.95	127	50.60	94	37.45	Significant difference(p < 0.05)
Cl	257	32	12.45	129	50.19	96	37.35	Significant difference(p < 0.05)

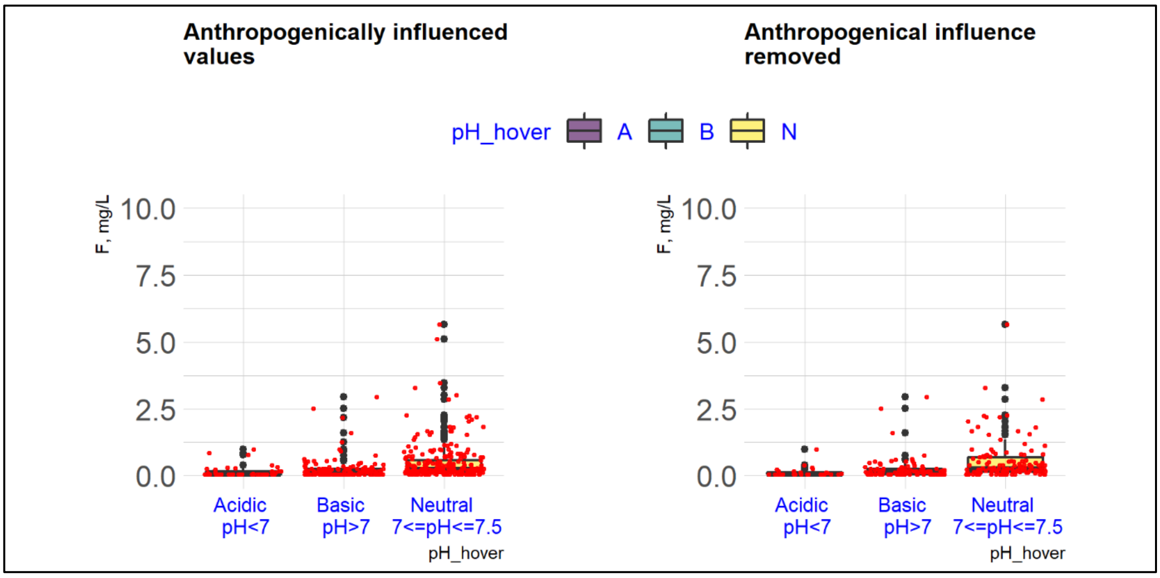
Figure 71 (next 2 pages) Comparison of concentration distributions for the HOVER-pH groups; for symbology explanation.


Anthropogenically influenced values
Anthropogenical influence removed
pH_hover ▨ A ▨ B ▨ N
SO4, mg/L
0
20
40
60
Acidic pH<7
Basic pH>7
Neutral 7<=pH<=7.5
pH_hover


Anthropogenically influenced values
Anthropogenical influence removed
pH_hover ▨ A ▨ B ▨ N
As, µg/L
0
5
10
15
20
Acidic pH<7
Basic pH>7
Neutral 7<=pH<=7.5
pH_hover


Anthropogenically influenced values
Anthropogenical influence removed
pH_hover ▨ A ▨ B ▨ N
Cd, mg/L
0.0
0.5
1.0
1.5
2.0
Acidic pH<7
Basic pH>7
Neutral 7<=pH<=7.5
pH_hover





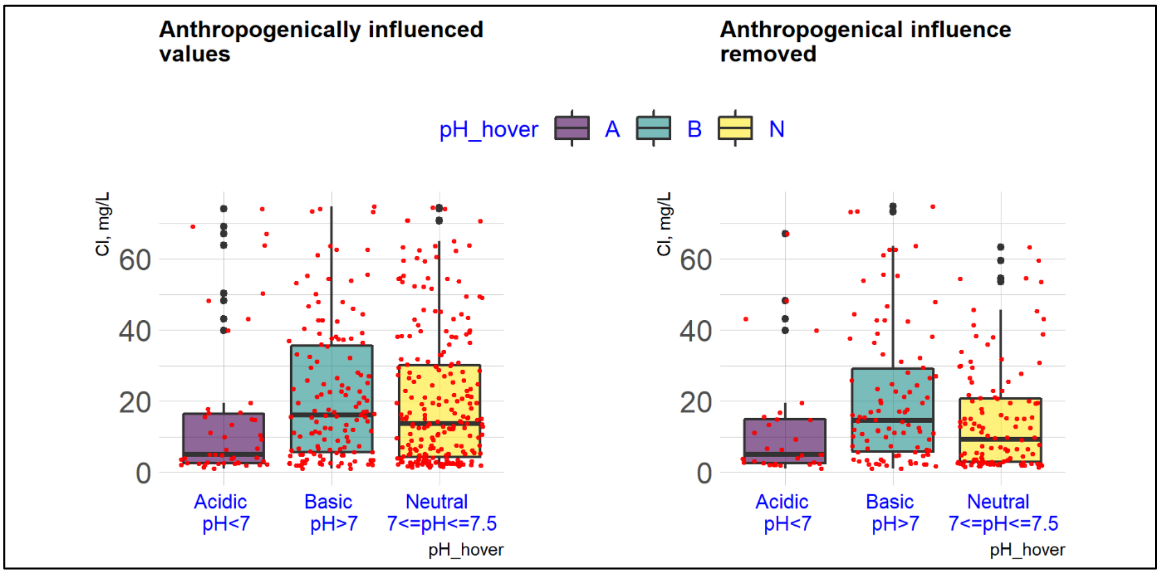


Figure 22 results show that:

- HOVER-pH cannot be used for Cd, Cr, Cu, and Zn because there is not enough data
- There is a significant difference between the acidic and basic groups for SO4 with both anthropogenic influence considered and removed. However, the basic and neutral groups are not statistically different.
- For As the only relevant difference occurs between the basic and neutral groups when anthropogenic influence is removed.
- For Cd, with more than 80% censored values, no difference is observed. There is not enough data to make an analysis.
- Cd, Ni; however, the neutral group is different from both acidic and basic groups only for Ni. For the rest of the elements, the neutral group is different from either the acidic or the basic groups.

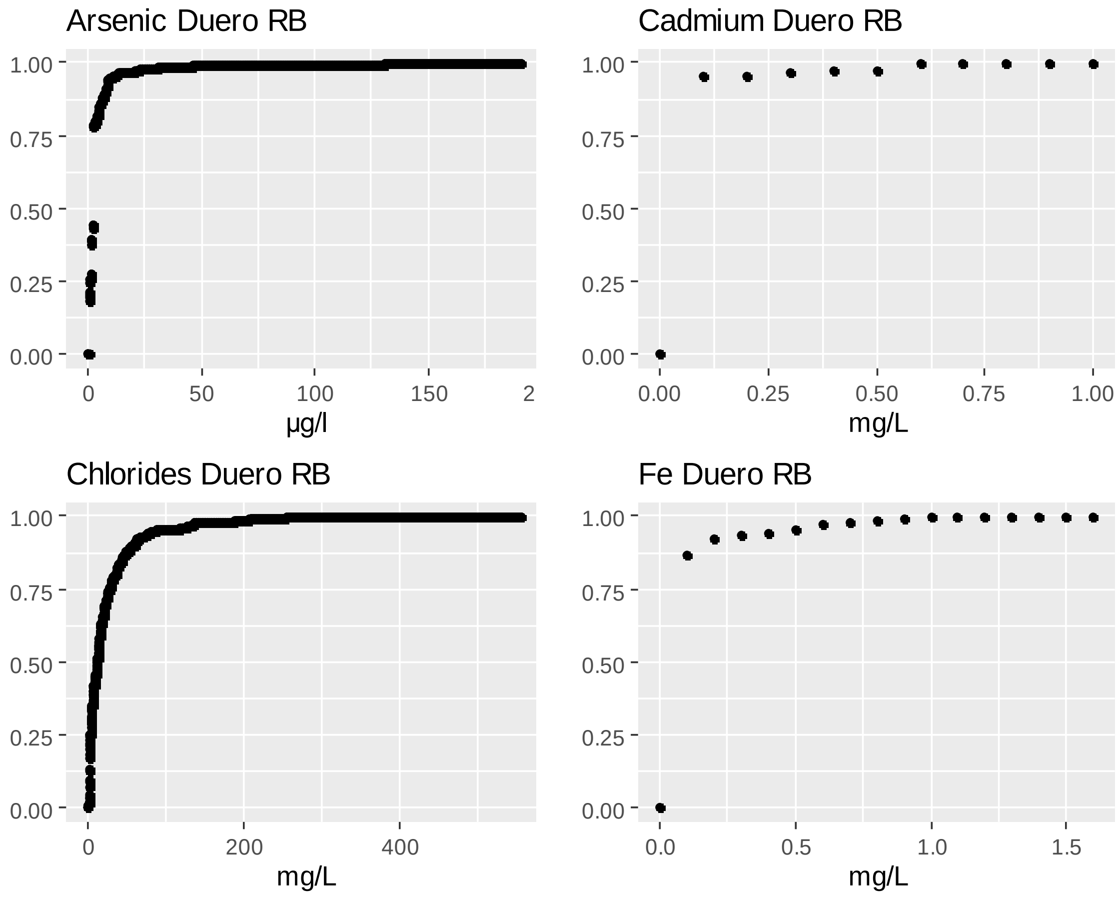
Therefore, it was decided to use the three HOVER-pH groups (acidic, basic, and neutral) together with the lithology/geology when calculating the NBLs for SO4, As. For Cd, Cr, Cu, and Zn we will use only the HOVER-lithology without taking into consideration pH, because there is not enough data for this. After the initial 90th percentile calculation is done, we will evaluate if there is enough data in each combination group (lithology/geology + pH).

8.4.6 *Additional removal of outliers -- element concentrations influenced by contamination*

After the removal of anthropogenically influenced water sampling points for each element, a cumulative distribution plot was made for some elements, attempting to identify additional points that could be identified as potential outliers (

Figure 72).

Cadmium and Fe were plotted just to show the limited amount of available information, as mentioned in previous sections.



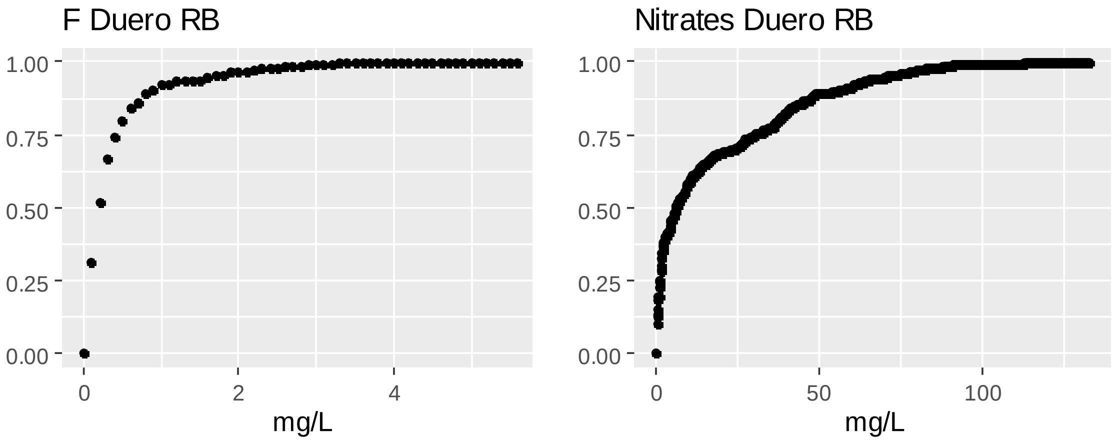


Figure 72: Empirical cumulative distribution plots for As, Cd, Cl, Fe, F and NO3.

We checked individually each of these sampling points to decide what action to take. Nevertheless, no point was identified by this method as a potential outlier. It is worth noting that for some elements more than 80% of all individual analyses based on which we calculated the medians for each water sampling point were < LOD, so most of their values in this dataset are influenced by the ½ LOQ substitution.

8.4.7 *Calculating natural background levels with the final dataset*

Natural background levels (NBL) are determined as the 90th percentile of the median concentrations at water sampling points over the 11-year study period for a specific groundwater typology. The anthropogenically influenced water sampling points (element-specific, as described in section 0and the outliers identified were excluded prior to calculations. Maps showing the spatial distribution of concentrations for each element and type of geology are attached in the Appendix. The number of water sampling points in each geology/lithology subset is given in Table 34. To assess the effects of the new method on the NBLs, we also calculated alternative NBLs without excluding the anthropogenically influenced points.

8.4.7.1 NBL for groundwater typologies based only on lithology/geology classification

The 90th percentiles were calculated first for the HOVER-Lithology classes for all classes with 16 or more water sampling points, due to the fact that the available information is very limited (Table 74).

Table 74 Natural background levels (NBL) for the ***HOVER-lithology****; the number of water points on which the 90th percentile calculation for each class is provided as well (see "n", with italic)*

		Anthropogenically influenced values				Anthropogenical influence removed			
HOVER lithology		Cenozoic		Metamorphic Rocks		Cenozoic		Metamorphic Rocks	
Number of points		*n*	NBL	*n*	NBL	*n*	nbl	*n*	NBL
As	µg/l	*386*	10	*23*	56	*202*	8.11	*16*	3.25
Cd	mg/l	*378*	0.26	*23*	0.029	*203*	0.025	*16*	0.046
Cr	-	*-*	-	*-*	-	*-*	-	*-*	-
Cu	mg/l	*385*	0.0027	*23*	0.0048	*203*	0.0027	*16*	0.0046
Fe		*389*	3.44	*23*	0.21	*203*	0.196	*16*	0.27
Ni	-	*-*	-	*-*	-	*-*	-	*-*	-
Zn	-	*-*	-	*-*	-	*-*	-	*-*	-
SO4	mg/L	*389*	241.16	*23*	23.63	*203*	100.30	*16*	30.86
F	mg/L	*364*	0.95	*23*	0.28	*197*	0.95	*16*	0.27
Cl	mg/L	*389*	119.03	*23*	24.3	*203*	62.58	*16*	28.91

8.4.7.2 NBLs for the groundwater typologies combining pH conditions with the lithology/geology

We calculated also the 90th percentiles for the groundwater typologies combining the HOVER-lithology with HOVER-pH.

Table 75 Natural background levels for ***lithology/geology*** *combined with* ***pH conditions***

HOVER lithology	Elements	units	HOVER-pH	NBLs	**NBLs** [a] (only geology)
Cenozoic	As	µg/l	Acidic	3.11	10
			Basic	77.08	
			Neutral	7.27	
	Cd	mg/l	Acidic	0.26	0.26
			Basic	-	
			Neutral	-	
	SO4	mg/l	Acidic	5.53	241.16
			Basic	413.52	
			Neutral	-	
	F	mg/l	Acidic	0.956	0.95
			Basic	-	
			Neutral	-	
	Cl	mg/l	Acidic	6.19	119.03
			Basic	150	
			Neutral	7.45	
Metamorphic Rocks	As	µg/l	Acidic	2.5	51.24
			Basic	178.8	
			Neutral	-	
	Cd	mg/l	Acidic	0.029	0.029
			Basic	-	
			Neutral	-	
	SO4	mg/l	Acidic	5.29	23.63
			Basic	37.30	
			Neutral	7.17	
	F	mg/l	Acidic	0.28	0.28
			Basic	-	
			Neutral	-	
	Cl	mg/l	Acidic	5.35	24.3
			Basic	38.92	
			Neutral	-	

[a] same as in Table 74(calculated without the HOVER-pH separation)

8.5 Comparison and contributions of the proposed method

The proposed method has been tested at a regional scale. Available data come from the official monitoring network of the Duero River Basin Authority. There is a series of limitations that condition the applicability of the method, among which we can cite the following:

- Mixture of measurements units, that makes necessary a thorough review of the database
- Precision of measurements, with very high LOQ in most cases
- Scarcity of records, especially for some trace elements. For some of them, just records from a couple of years of measurements exist

In addition, monitoring networks are located in areas with no significant industrial or mining activities. Agriculture is the dominant pressure, with about 80% of the total. This makes the calculated NBLs representative of the agricultural group, hampering the ability of the method to identify statistical differences between groups. In the case of urban, industrial and mining pressures it is not possible at all, due to the lack of data.

As a first example, the case of Arsenic is shown. NBLs obtained considering just the Hover-lithology range between **10 - 56 µg/l** for anthropogenically influenced values. When such an influence is removed, values range from **3.25 - 8.11 µg/l**. When compared to the threshold values provided by the Duero River Basin Authorities (Table 76), we observed that the threshold value is different for a number of GWs. For some of them, the percentile 90 is used, while for others, the 97.7 percentile is used instead. Taking into account that measurement units are mg/L, values obtained with the method are coherent with established threshold values.

GWB CODE	GWB name	Parameter	Selection criteria	Threshold value (mg/L)
400045	Los Arenales	Arsenic	Percentile 97,7	0,140
400047	Medina del Campo			0,079
400052	Salamanca			0,047
400053	Vitigudino		Percentile 90	0,204
400058	Campo Charro			0,027
400063	Ciudad Rodrigo			0,630

Table 76: Threshold values for As in Duero River Basin

As a second example, NBLs for Sulphates and Chlorides are analyzed. For SO4, like in the case of Cl and some other elements, overexploitation is a subject of possible concern, so an analysis at trends could supply additional insight into potentially affected water points, with the aim of excluding them from the final data set.

GWB CODE	GWB name	Parameter	Selection criteria	Threshold value (mg/L)
400031	Villafáfila	Chlorides	Percentile 97,7	303
400038	Tordesillas			441
400016	Castrojeríz	Sulfates	Percentile 97,7	456
400045	Los Arenales			1.108
400067	Terciario bajo Páramos			1.548

Table 77: Threshold values for SO5 and Cl in Duero River Basin

Going back to NBLs for Sulphates obtained when considering Hover-lithology, the maximum value is about **241 mg/L** (anthropogenically influenced values), while the value when the anthropogenically influence is removed is about **100 mg/L**. Specifically, 97.7 percentile for Cenozoic lithology of the final data set is about **900 mg/L**. This is quite consistent with the <u>threshold value</u> established by the competent body, which is **1108 mg/L**. Referred to chlorides, the <u>NBL is 94 mg/L</u> (97.7 percentile is **256 mg/L**). The threshold reference value (Table 77) ranges between 303 – 441 mg/L.

In conclusion, the proposed method is useful and provides coherent results when compared to the existing values. Therefore, it could be used as an exploratory step. However, the monitoring networks should be improved in order to provide additional needed information, mainly trace elements concentration and improve LOQs.

Lastly, on such regional scale, data-analytical methods might also be accompanied by process-oriented analyses and supported by expert background knowledge at the local scale.

9 CASE STUDY – WIDER AREA OF MT. FRUŠKA GORA, SERBIA

Interpretation: Tanja Petrović Pantić

Katarina Atanasković Samolov

9.1 Previous studies on NBLs

Serbia has 153 groundwater bodies. The Republic Hydrometeorological Service of Serbia, Environmental Protection Agency, Ministry of Environmental Protection, Institute for Public Health, and Faculty of Mining and Geology carry out quantitative and qualitative monitoring of groundwater.

Republic Hydrometeorological Service of Serbia controls quantitative monitoring network of groundwater, which cover 29 groundwater bodies or 53 piezometers mostly in alluvial sediments of great rivers and shallow aquifers.

Environmental Protection Agency and Ministry of Environmental Protection cover qualitative monitoring of groundwater on 26 groundwater bodies (53 piezometers).

Faculty of Mining and Geology, University of Belgrade established a monitoring network on 13 new groundwater bodies, or 35 water sampling points (Stevanović et al, 2019).

Groundwater is the main source of water supply. About 75% of groundwater is used for water supply in Serbia, while 25% is surface water. The Institute of Public Health analyses the quality of water for water supply assessing its hygiene levels and suitability for consumption. However, these samples are mostly taken from the taps; as a result, these data are not considered for this study.

The results of qualitative monitoring showed that the main problem of groundwater qualitative composition is in Vojvodina (part of Pannonian Basin) with high concentration of Mn, Fe, As, B, $NH_{3,}$ NO_3 and NO_4 and $KMnO_4$.

In order to create a hydrogeological map of Novi Sad, the one-year quantitative monitoring net was established in the wider area of Mt Fruška Gora (Petrović Pantić and Kostić, 2020). Qualitative monitoring was carried out for the study on groundwater reserves. During the monitoring period, the samples were taken four times per hydrologic year, or more. These results are analyzed for HOVER project.

9.2 Study with the proposed method

9.2.1 *Constitution of the groundwater quality dataset*

Data inventory and relevant information

The study area, a wider area of Mt. Fruška Gora is separated into two water bodies. Mountain Fruška Gora is on the right side of the Danube River, and on the left side is a plain. 33 sample points are shown on the Basic (Hydro) geological map scale 100.000, sheet Novi Sad, and some samples are on sheet Indjija (Figure 73). The most number of samples were taken from Neogene sediments. A few samples on Fruška Gora are taken from the Triassic limestones, and a few from Quaternary deposits (shallow aquifer).

Groundwater samples are mostly taken from groundwater sources of water supply:
- 28 samples from artesian wells/boreholes,
- 3 from shallow wells,
- one sample from a spring,
- one sample from a Ranney well.

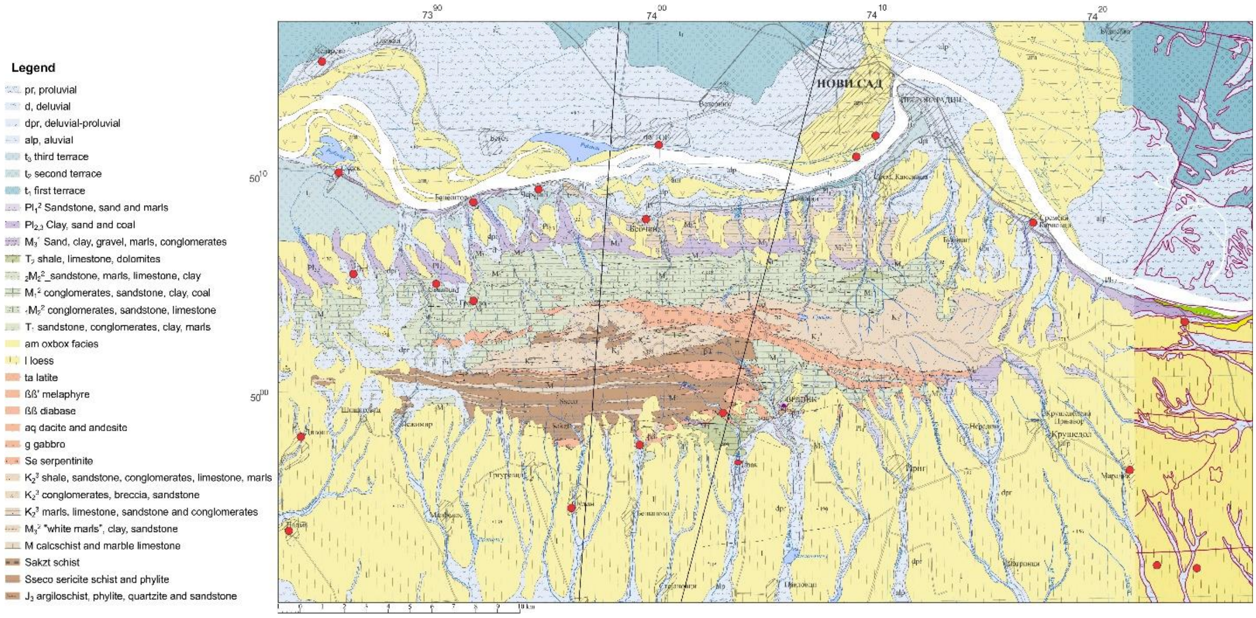


Figure 73: Hydrogeological map of wider area of Fruška Gora (source: basic hydrogeological map 1:100.000, sheet Novi Sad (Petrović Pantić and Kostić, 2020) and part of sheet Indjija (unfinished)). Categorization of lithostratigraphic units was made according to the guide and standard legend for hydrogeological maps 1:100.000 (Šarin, 1988).

Relevant period

The reference period chosen for the natural background levels is 2006-2017.

Extraction or synthesis of quality data for major, minor and trace elements from the national or regional quality databases for the reference period

We used chemical analysis, with available values of required elements for this study. Most of the analysis was carried out without the redox potential (Eh), and some of it did not contain the value of O_2.

Formatting the groundwater quality database

The database was checked for errors and any duplicates were removed from it. Average values of chemical parameters were calculated in accordance with the HOVER methodology.

Limits of detection (LOD) were considered to improve data quality and these detections were treated as LOD/2. Parameters with high value of LOD did not take in consideration.

Table 78 Number of analyses available for each element below the limit of detection (LOD).

Element	unit	All (n)	< LOD (n)	< LOD (%)	LOD_{max}	LOD_{min}
As	µg/L	31	13	41,94	5	0,75
Cd	µg/L	31	29	93,55	2	0,04
Cu	µg/L	18	4	22,22	3	1,62
Cr	µg/L	13	5	38,46	0,4	0,4
Ni	µg/L	15	7	46,67	2	2
Zn	µg/L	27	2	7,41	2,6	1
F	mg/l	33	3	9,09	0,1	0,1
SO4	mg/l	33	7	21,21	1	0,2

Defining for each sampling point a concentration (median) for each major, minor or trace element

According to HOVER methodology, median value for each sampling point was calculated. All median values are shown in table (excel sheet)

Determining HOVER-Redox and HOVER-pH classes

According to redox classification 16 water samples have oxic condition (“A”), 11 reduced (“C and D”) and 5 mixed (“X”).

According to HOVER-pH Classes water samples are basic pH>7,5 (16) to neutral 7 ≤ pH ≤ 7,5 (17 samples).

9.2.2 *Hydrogeological characteristics of the sampling points*

Quaternary sediments are dominant in the area around the Fruška Gora Mountain. Quaternary sediments are represented with channel and flood plain facies, proluvial-delluvial sediments, and river terraces on the northern slope of Fruška Gora as well as the area along the Danube River. Channel and flood plain facies are high potential aquifers, used for water supply in the city of Novi Sad and Beočin. Hydraulic conductivity is in the range between 2,3x10^{-2} and 7,1x10^{-3} m/s. An underlying bed is made of clay and sandy clay of Pleistocene age. Groundwater table is shallow and beneath the surface, and it is in the hydraulic connection with the Danube River (Petrović Pantić et al, 2017).

In the southern part of Fruška Gora, Quaternary sediments are also dominant on the surface, but the hydrogeological qualities of Neogene sediments mean they have significant distribution beneath Quaternary rocks. Neogene rocks are of Pliocene and Miocene age - sand, clay, marls, and gravel. There is a characteristic of shifting layers with different porosity. Hydraulic conductivity is in the range between 10^{-4} and 10^{-6} m/s. Intergranular porosity is dominant. Miocene sandstone, limestone, marls, and conglomerates formed a fracture aquifer with a low yield. Karst aquifer is formed in Triassic limestone, dolomite, sandstones, and conglomerates. Three water samples were obtained from this aquifer. Hydraulic conductivity is in the range between 8,5x10^{-5} and 4,4x10^{-6} m/s (Stoiljković, 2003).

On the Fruška Gora Mountain, low-impermeable or impermeable rocks are widely present. Low impermeable rocks occur on the south slope of the mountain with loess, then intrusive rocks (latite, diabase, dacite, andesite, gabbro) and serpentinite on the top of the mountain. Impermeable rocks are argiloshist, phylitte, quartz, sandstone, schist, and these rocks occur on the top of Fruška Gora.

9.2.3 *Anthropogenic pressures / influences: diffuse or point pollutions*

Mt. Fruška Gora is a National park, hence the area is being protected by the Law on National parks (Official Gazette of RS, no. 84/2015, 95/2018). According to the DRASTIC method for assessing the vulnerability of groundwater, the top of the mountain has a very low to low vulnerability index, while further down the slope the vulnerability index rises (Veljković et al, 2016; Petrović Pantić et al, 2017).

Fruška Gora recorded a population increase in the past twenty years due to its excellent strategic location, and proximity to large cities of Novi Sad and Belgrade (Petrović Pantić et al, 2017). Also, the mining activity exists with the exploitation of trachyte. Industrial pressure and/or mining influence could be a threat to the city of Beočin (Concrete plant). Urban influence is a dominant antropogenic pressure in the area (Novi Sad), and on shallow boreholes in Quaternary sediments. Agricultural activity exists on the slope of Fruška Gora. For indentifying prevailing antropogenic influences. Corine landover map were used. Non-irrigated arable land is dominant on the research field and forest on the highest elevation on the Mountain.

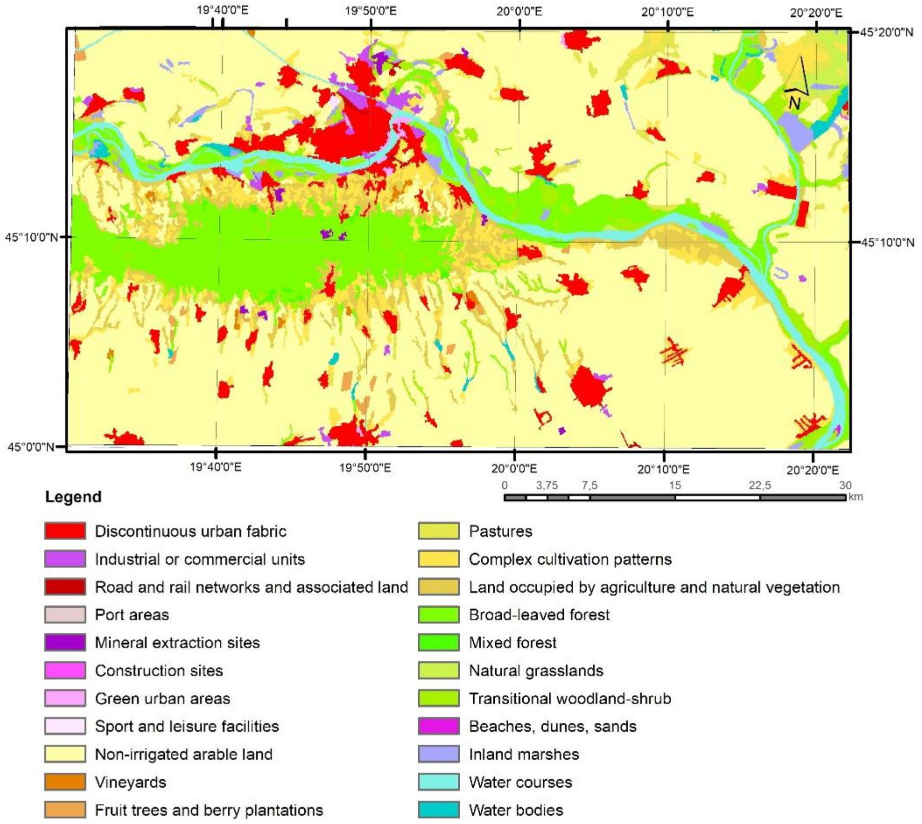


Figure 74: Corine land cover map for the research area (https://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-4)

9.2.4 *Stastistical treatment*

For all 33 samples point minimal, maximal, average, median value and standard deviation are calculated (Table 3). Values below LOD, calculated as LOD/2 (Table 72).

Table 79: Statistical values per parameter for all samples

	Min	Max	Average	Median	St.deviation
pH	7,15	8,16	7,57	7,46	0,3
Conductivity	0,54	1,3	0,75	0,7	0,2
Temperature	11,7	25,15	18,14	17,3	3,65
SO_4	0,1	104	19,9	13,2	23,4
As	0,38	106	10,3	2,5	20,6
Cd	0,02	1	0,54	0,5	0,4
Cr	0,2	30	6,97	0,6	9,7
Cu	0,2	15,32	2,57	1,5	3,5
Ni	1	11,5	3,32	1,4	3,62
Zn	0,12	293	38,8	12,5	71,5
F	0,05	0,85	0,24	0,19	0,19
Oxygen (O_2)	1	44	5,84	3,12	8,31
Nitrate (NO_3)	0,09	133,1	11,33	1,83	30,18
Fe	0,01	5,92	0,6	0,19	1,11
Cl	3,9	75,3	24	15,1	20,9
HCO_3	357,45	673,5	455,15	442,5	65,38
Ca	13,8	237	68,9	58,3	51,9
Mg	9,82	69,6	35,3	34,5	11,8
Na	7,1	263	64,3	37	62,7
K	0,4	25,1	2,45	1,51	4,2

Table 80: LOQ/2 selected for the present study

Parameter	LOQ/2 maximum (µg/l)
As	2,5
Ni	1
Cd	1
Cr	0,4
Cu	0,2
F	0,05
Zn	1,3

Discriminant function Analysis (DFA) per lithological/geological family to determine whether or not GW points of the same lithological/geological family, but geographically distant, can be grouped together for further statistical treatments

DFA classifies water points per lithology HOVER and age formation/stratigraphy. DFA classifies water point on wider area of Fruška Gora on 4 groups (table 3). Water samples were dominate sampled from Neogene sand.

Table 81: Discriminant function Analysis (DFA) per lithological/geological family

lithological/geological family	***age formation/stratigraphy***	***Number of samples***	***Percentage (%)***
sedimentary: sand	Neogene, Quaternary	25	75,75
sedimentary: other	Triassic, Neogene and Paleogene	5	15,15
sedimentary: carbonates (limestone, chalk)	Triassic	3	9,09
Alluvium	Quaternary	1	3,03

9.2.5 *Removal of element concentrations influenced by contamination and NBL calculation*

Remove sampling points for which the element concentration is significantly influenced by a source of contamination

Only concentration of As is elevated in few samples, but sampling points weren't remove, because As has geogenic origin. Almost in all Vojvodina area (north part of Fruška Gora) As is usually elevated in groundwater (Dalmacija et al, 2011, Jovanović et al, 2011, Petrović et al, 2012)

NBLs determination per lithological/geological family: 90 percentile and range of concentrations

Natural background levels are determinated for next parmeters: SO_4, As, Cd, Cr, Cu, Ni, Zn and F (table 4). The 90 percentiles as the threshold value as reference concentration by lithology is used. Method is not applicable for one sample, hence the sample point in alluvium is not considered.

Table 82 : The natural background level (90%) per lithological/geological family

lithological/geological family	***SO_4 (mg/l)***		***As (µg/l)***		***Cd (µg/l)***		***Cr (µg/l)***		***Cu (µg/l)***		***Ni (µg/l)***		***Zn (µg/l)***		***F (mg/l)***	
	NBL (90%)	*Range (10%-90%)*	*NBL (90%)*	*Range (10%-90%)*	*NBL (90%)*	*Range (10%-90%)*	*NBL (90%)*	*Range (10%-90%)*	*NBL (90%)*	*Range (10%-90%)*	*NBL (90%)*	*Range (10%-90%9*	*NBL (90%)*	*Range (10%-90%)*	*NBL (90%)*	*Range (10%-90%)*
sedimentary: sand	35,89	0,22 to 35,89	28,8	1 to 28,8	1	0,02 to 1	18,15	0,2 to 20,3	4,56	0,4 to 4,56	6,28	1 to 6,28	40,8	6,27 to 40,8	0,34	0,06 to 0,34
sedimentary: other	61,61	10,65 to 61,61	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	8,9	8,9 to 5	266,25	17,87 to 266,25	0,33	0,162 to 0,33
sedimentary: carbonates (limestone, chalk)	28,05	18,49 to 28,05	1	1 to 1	<LOD	<LOD	<LOD	<LOD	0,21	0,21 to 2,45	<LOD	<LOD	1,07	0,12 to 1,07	0,85	0,25 to 0,85

10 CASE STUDY – SLOVENIA

Interpretation: Nina Rman, Sonja Cerar

Technical support: Klemen Klančič, Simon Mozetič, Luka Serianz

10.1 Geological settings of Slovenia

The territory of Slovenia extends over 20,273 km^2. It is located at the junction of the Alps (N, W), the Dinaric Alps (S), the Pannonian Basin (E) and the Adriatic Sea Basin (SW), and a few geotectonic units: the Adriatic-Apulian foreland, the Dinarides, the Southern Alps, the Eastern Alps, and the Pannonian basin. Their present-day configurations were largely formed in Neogene times. Terrain rises from the coastline to the Triglav Mt. with 2864 m a.s.l.

Slovenia has mostly moderate, warm and humid climate, with mixing of continental, Alpine and coastal sub-Mediterranean climate patterns. Precipitation decreases towards east with average annual amounts extremes between roughly 3200 and 900 mm. Average annual ai temperature is between 8 and 12 °C, excluding high mountains.

The oldest rocks of the metamorphic complex of Pohorje, Kobansko and Strojna crop out within the so-called Periadriatic zone and at Pohorje. Metamorphic rocks constitute only 4% of the territory. They are mostly gneiss, micaschist, phyllite, slates, eclogite, marble and quartzite. Igneous and volcaniclastic rocks cover 3% of the territory; the central massif of the Pohorje (granodiorite, cezlakite, dacite) and Periadriatic igneous zone (sienogranite, tonalite). Volcanic (porphyrite, diabase and andesite) and volcaniclastic (tuffaceous) rocks are dispersed in central Slovenia and in the wider Savinja valley (so-called Smrekovec volcanism).

The most abundant are sediments and sedimentary rocks, which account for about 93% of the territory. Half of the territory (49%) is covered with sediments and clastic sedimentary rocks. Sediments (gravel, sand, silt, clay) fil riverbeds and young sedimentary basins of tectonic origin (Ljubljana, Celje and Krško-Brežice basin) and the Drava and Mura river plains. Clastic sedimentary rocks predominate in central Slovenia (Paleozoic in age), in the southwest (Eocene flysch rocks) and in the northeast (Neogene marlstone, mudstone and sandstone of the Pannonian basin). Carbonate rocks cover 39% of the territory. Limestone and dolomite form extensive Mesozoic massifs of southern Slovenia and the Alps in north-western and northern Slovenia. Regarding the age, Cenozoic and Mesozoic rocks comprise the largest part of Slovenia's surface, at 44% and 45% respectively; of the latter as much as 23% are of Triassic age.

10.2 Previous studies on NBLs in Slovenia

Slovenia has 21 (shallow) groundwater bodies. The Environmental Agency of Slovenia (slo. ARSO) performs systematic national groundwater quality monitoring on alluvial aquifers since 1987 while on karstic since 1990's. The first have much more monitoring sites than the latter. Two to four times per year about 160 parameters are analyzed (Kranjc et al. 2007) to observe their trends and report on current quality status according to WFD, however no detailed interpretation is provided on the source of origin or transport of substances.

Natural background levels in groundwaters in Slovenia were systematically investigated within the PhD survey of Mezga in 2014 for the first time. It was founded on a 3-year study (2009-2011) of 176 water samples from two sampling campaigns and taken from 87 sampling locations which were equally distributed in the country. The aim of the study was to determine the impact of aquifer lithology, climate and land use, and to identify NBLs of selected parameters. She used simplified lithological classification with 10 subunits:

- Clastic sedimentary rocks had subunits: flysch rocks; clay; gravel and sand; gravel, sand and clay; shale and sandstone.
- Carbonate rocks had subunits: limestone prevailing; dolomite prevailing; carbonates with clastics.
- Igneous and metamorphic rocks are composed of the two identically-named subunits.

Mezga determined NBL for parameters of which sources are most likely of anthropogenic origin. Within methodology, the mean values of measured parameters in groundwater that had up to 20% of anthropogenic influence in the recharge areas were considered as NBL.

Serianz et al. (2020) upgraded this approach by using longer time-series and more locations, all together 297 sites. Datasets from 2007 to 2017 were transformed to median values for 13 chemical parameters. After quality check, 213 sites, 96 springs and 117 wells were included. All samples were aerobic ($O_2 > 1$ mg/l) and samples with nitrate above 10 mg/l were excluded. BRIDGE methodology was applied and the 90th percentile was taken as the upper confidence limit of NBLs. The preselection (PS) and probability plot (PP) methods were used to determine NBLs. Where difference among the two was below 10%, average concentration was calculated. Where this was exceeded, NBL was selected according to hydrogeochemical characteristics.

Explanation of NBLs of Serianz et al. (2020) is as follows: As higher nitrate concentrations were omitted, the PS method NBL was selected in limestone. But as PP method for all samples showed change of slope at 6.0 mg/l, this value was selected for NBL. For sulphate, 26 mg/l was determined for all samples with PS method but in dolomite and calcareous rocks values of 9.3-11.2 mg/l were determined. The PP method suggested 17.0 mg/l, which we proposed at the end. For chloride, NBL in Quaternary clastics by the PS method is 37.8 mg/l and in dolomites 3.2 mg/l, while the entire set gives 30.8 mg/l. They propose 21.0 mg/l as the chloride NBL. They interpret that newly determined NBLs are higher than of Mezga (2014) because larger dataset was available and in which variations in groundwater composition over time and space were therefore more included.

Table 83: Archive determination of NBLs for few parameters in Slovenian groundwaters as interpreted by Mezga (2014) and Serianz et al. (2020). Not evaluated stands for samples where more than 1/3 was below LOQ.

	Mezga, 2014		Serianz et al. 2020	
Parameter	min-max / average (mg/l)	NBL (mg/l)	min-max / average (mg/l)	NBL PS method / NBL PP method = proposed NBL (mg/l)
SO4	0.75-67.40 / 10.05	6.93	0.5-75.5 / 13.1	26.0 / 17.0 = 17.0
Cr	0.5-42.0 / 1.7	Not evaluated		
Cl	0.11-36.70 / 4.99	1.72	0.5-100.0 / 11.9	30.8 / 21.0 = 21.0
NO3	0.33-92.56 / 9.56	3.81	1.1-89.0 / 18.2	6.3 / 6.0 = 6.0
Fe	0.010-0.132 / 0.036			
pH	6.4-8.3 / 7.6		6.2-8.5 / 7.4	
EC	28-882 / 368		138-976 / 444	
T			5.4-18.2 / 11.3	

10.3 Study with the proposed method

10.3.1 *Constitution of groundwater quality dataset*

Lithostratigraphy was determined according to Basic Geological Map in Scale 1:100.000 and 1:250.000 (Buser, 2010) and points were linked to delineated 21 groundwater bodies.

Chemical parameters were taken from the national monitoring stations of shallow groundwater aquifers managed by the Slovenian Environmental Agency (ARSO). Data is available for years between 2007 and

2019 at https://www.arso.gov.si/vode/podatki/. We used datasets from 2016 for all stations as the set of monitoring stations was the largest in that year.

GIS analysis on the activities within the recharge area was based on land use data from www.mkgp.gov and https://rkg.gov.si/vstop/.

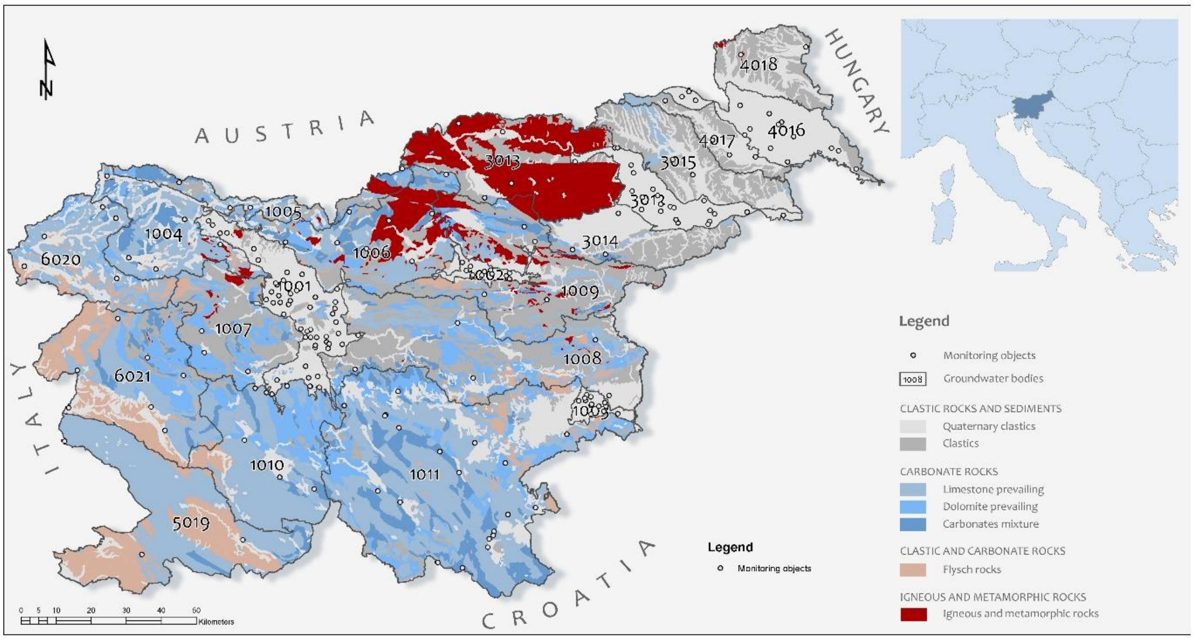


Figure 75: Used data points of the Slovenian national monitoring network with simplified geological map (modified from Buser 2010) and numbers of delineated 21 groundwater bodies (https://podatki.gov.si/dataset/vodna-telesa-podzemnih-voda).

10.3.2 *Constitution of groundwater quality dataset*

Data was available for 203 sampling sites all over Slovenia for year 2016, and we calculated average values when more than one was available.

10.3.3 *Data quality check*

Quality check consisted of inspection of data but it was impossible to realize whether individual outliers as a consequence of typing errors or actual chemical composition. Therefore, we removed whole analyses where parameters showed high anthropogenic influence which we defined as having nitrates, sulphates or chloride above 50 mg/l and, at the same time, showing anthropogenic activities in the recharge area when analyzing land use in GIS. We distinguished among urban and agricultural pressures only, and when one of them was identified we marked the dataset as being submerged to an anthropogenic pressure. Therefore, 24 of 203 original analyses were excluded from further analysis. We also excluded four outliers from the data set, one for As, Cd, Cr and Fe.

10.3.4 *Analysis of monitoring object types*

Among 179 sites, there were 39% springs, 33% shallow wells (below 30 m in depth), and 28% of boreholes with a maximum depth of 200 m.

10.3.5 *Processing of censored data*

Among the rest 179 analyses, one of largest issues using these data and performing statistical evaluation were different limits of quantification (LOQ). We have two sets of samples with two different LOQs for eight parameters: As, Cd, Cr, Cu, F, Fe, Ni and Zn (Table 84). In case of six parameters, As, Cd, Cr, F, Fe and Zn, more than half analyses are below LOQ. We calculated the values as half of LOQs and the maximum gained values (Table 84) are much lower than in French example, for example: 0.5 vs 1 µg/L As, 0.01 vs 0.5 µg/L Cd, 0.5 vs. 2 µg/L Cr, 0.5 vs. 10 µg/L Cu, 0.1 vs. 25 µg/L F, 0.5 vs. 2 µg/L Ni. However, we have higher value for zinc, being 5 vs. 2 µg/L.

Table 84: Two LOQs were reported for the whole set of samples in 2016

	Low LOQ	**% of samples below low LOQ**	**High LOQ**	**% of samples below high LOQ**	**Sum % of samples below LOQ**	max LOQ/2
As (µg/L)	<0.1	7	*<1*	*56*	**63**	0.5
Cd (µg/L)	**<0.01**	**31**	***<0.02***	***40***	**71**	0.01
Cr (µg/L)	*<0.4*	*31*	<1	22	**53**	0.5
Cu (µg/L)	<0.1	4	*<1*	*41*	45	0.5
F (mg/L)	**<0.04**	**18**	***<0.2***	***46***	**64**	0.1
Fe (mg/L)	**<0.04**	**35**	***<0.1***	***43***	**79**	0.05
Ni (µg/L)	<0.1	13	*<1*	*26*	38	0.5
NO3 (mg/L)			<2.2	4	**4**	1.1
Zn (µg/L)	**<9**	**35**	***<10***	***39***	**74**	5

10.3.6 *Determining HOVER-Redox and HOVER-pH classes*

Regarding aquifer types there is rather large uncertainty in type determination but in general we assume that most samples represent unconfined aquifers (93%) while the rest are probably semi-confined.

According to the proposed HOVER methodology we have determined HOVER-Redox and HOVER-pH classes on 179 samples.

- Among HOVER-pH classes we have predominant neutral waters (97; 54%), lots of basic also (75; 42%), while acidic are very rare (7; 4%).
- Among HOVER-Redox classes we identified predominant oxic conditions (171; 96%), few strongly reduced (5; 3%), and very rare unclassified (2; 1%) and nitrate reducing anoxic water (1; 1%).

10.3.7 *Determining BRIDGE and HOVER lithologies*

We applied both lithologies (Table 85). Most objects 52% tap fluviatile deposits where we differentiate among gravel (14%) and alluvium (86%) but one should not take this separation too precisely. That karstic areas are poorly covered by monitoring points is proved also by counted 37% objects (mostly springs) in carbonates, of which 54 are in karstic limestones and 12 in limestones of mountainous areas. Some wells, 10% of objects, tap (mostly neogene) sandstones, and only one sample represents crystaline and metamorphic rocks.

We decided not to apply dolomites separately, even though about 38% of so named carbonates are dolomites. We interpreted them together with limestone as for most NBLs determination up to now no separation was performed in Slovenia (Mezga, 2014, Serianz et al., 2020).

Table 85: Classification of 179 samples by HOVER and BRIDGE lithologies

		BRIDGE lithology				
lithology HOVER		Fluviatile deposits of major streams	Sandstones and silicatic alternating sequences	Karstic limestones	Limestones of mountainous areas	Crystalline rocks
Alluvium	80	93				
Sedimentary: sand	19		19			
Sedimentary: gravel	13					
Sedimentary: carbonates	66			54	12	
Metamorphic rocks	1					1

10.3.8 *Analysis of pressures in the recharge area*

Analysis of pressures shows that at about 44% of monitoring sites urban pressure is expected to occur in the recharge areas, agricultural pressure is expected at 17% of sites, so altogether about 61% of sites with data are expected to have some anthropogenic influence.

10.4 Results

Data processing was performed in Statistica software and Excel.

10.4.1 *Statistics of the whole dataset*

All parameters except for fluoride had 179 data available (Table 86). We have calculated basic statistic of the whole dataset and compared their ranges on Box and Whisker plots (Figure 76). It is evident that due to issues with lots of data below LOQ (see chapter 10.3.5), mean is mostly higher than the median, and median is the same as maximum replaced value of censored data for all parameters except for nitrate. Also 90th percentile is the same as maximum replaced value of censored data for As, F and Fe.

Table 86: Descriptive statistics of the whole dataset with 179 sites

	SO4 (mg/l)	As (µg/L)	Cd (µg/L)	Cr (µg/L)	Cu (µg/L)	F (mg/L)	Fe (mg/L)	Ni (µg/L)	NO3 (mg/L)	Zn (µg/L)
N	175	178	178	178	179	175	178	179	179	179
Mean	10.44	0.57	0.01	0.81	1.00	0.07	0.12	0.73	13.2	11.8
St.dev	7.95	1.61	0.01	0.79	3.43	0.04	0.55	0.87	12.9	38.0
Median	8.99	**0.5**	**0.01**	**0.5**	**0.5**	**0.1**	**0.05**	**0.5**	6.6	**5**
LOQ/2max	/	***0.5***	***0.01***	***0.5***	***0.5***	***0.1***	***0.05***	***0.5***	*1.1*	***5***
Min	0.50	0.05	0.005	0.20	0.05	0.02	0.02	0.05	1.1	4.5
P10	2.34	0.1	0.005	0.2	0.16	0.02	0.02	0.05	2.1	4.5
P25	3.73	0.14	0.005	0.2	0.35	0.032	0.02	0.15	3.8	4.5
P75	14.6	0.5	0.010	1.3	0.5	0.1	0.05	1.0	19.0	5.0
P90	***21.0***	***0.5***	***0.016***	***2.0***	***1.1***	***0.1***	***0.05***	***1.9***	***34.5***	***17***
Max	37.50	16.00	0.14	3.8	37.00	0.21	5.60	5.90	49.0	380.0

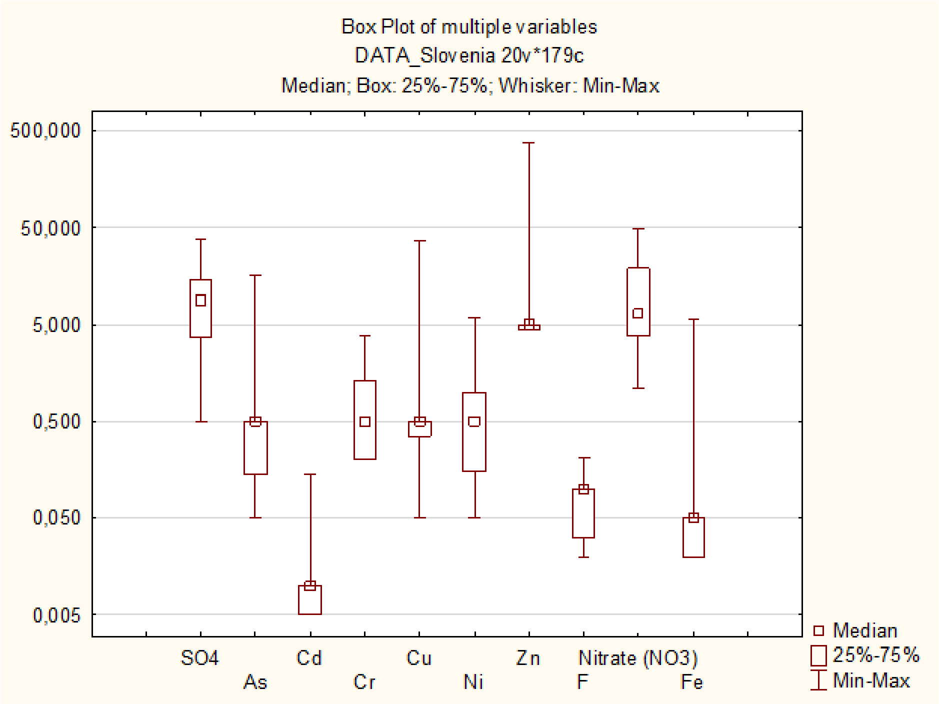
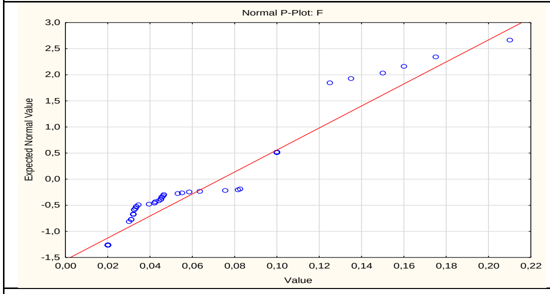
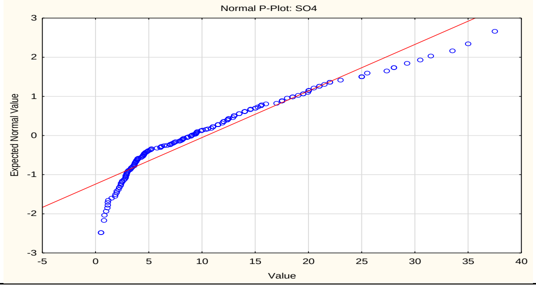
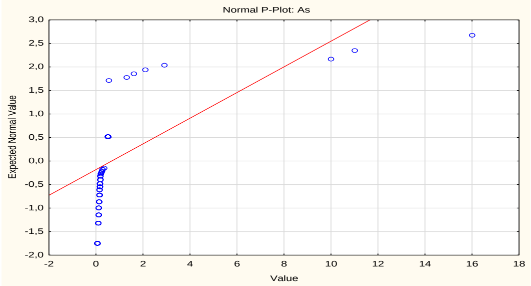
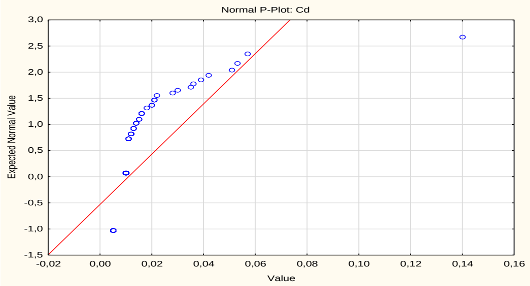
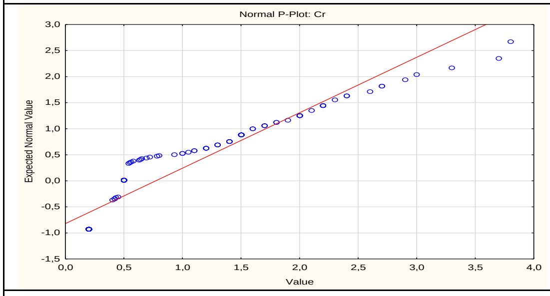
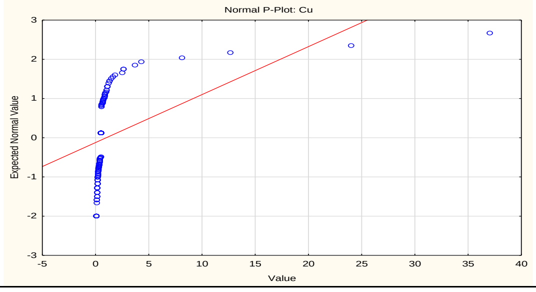
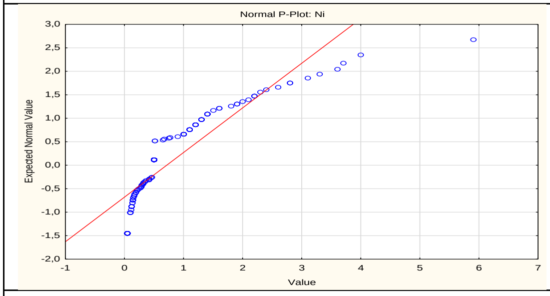
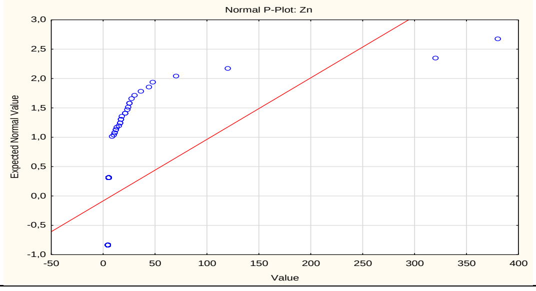
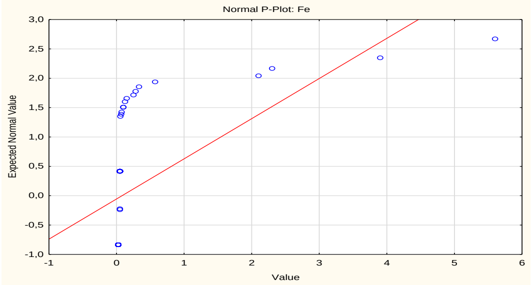
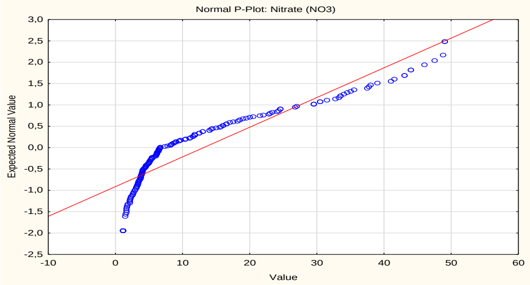


Figure 76: Box and Whisker plot of all samples and investigated parameters. Due to large range logaritmic scale is applied. Outliers are not shown.

10.4.2 *Data normality distribution*

Table 87: Normality distribution of all 179 samples for observed parameters. Cr, Ni, NO_3 and SO_4 are closest to normal distribution.

F	**SO4**
	
As	**Cd**
	
Cr	**Cu**
	
Ni	**Zn**
	
Fe	**NO3**
	

10.4.3 *F test and p(ANOVA) and Kruskal-Wallis test of parameters*

Influence of anthropic pressure on elements concentrations was assessed by the following tests. We performed F test and p(ANOVA) plus Kruskal-Wallis tests on the whole set of data.

Results show an influence of anthropic pressure (all pressures? urban, agricole?) on Cadmium for both tests, while As, Cu , Fe and Zn are distinguished with only the F test+Anova."

Table 88: Results of Kruskal-Wallis test. Results are bolded if they are statistically significant ($p<0.05$).

Test	**SO4 (mg/L)**	**As (µg/L)**	**Cd (µg/L)**	**Cr (µg/L)**	**Cu (µg/L)**	**F (mg/L)**	**Fe (mg/L)**	**Ni (µg/L)**	**NO3 (mg/L)**	**Zn (µg/L)**
F	(1;173) = 36.10	(1;177) = 1.72	(1;177) = 0.67	(1;177) = 6.64	(1;177) = 4.84	(1;173) = 19.05	(1;177) = 0.84	(1;177) = 14.95	(1;177) = 53.53	(1;177) = 0.75
KW-H	(1;175) = 30.28	(1;179) = 20.87	(1;179) = 0.07	(1;179) = 16.58	(1;179) = 19.00	(1;175) = 16.22	(1;179) = 11.32	(1;179) = 24.32	(1;179) = 34.23	(1;179) = 22.47

10.4.4 *Comparison of datasets regarding denoted anthropogenic influence*

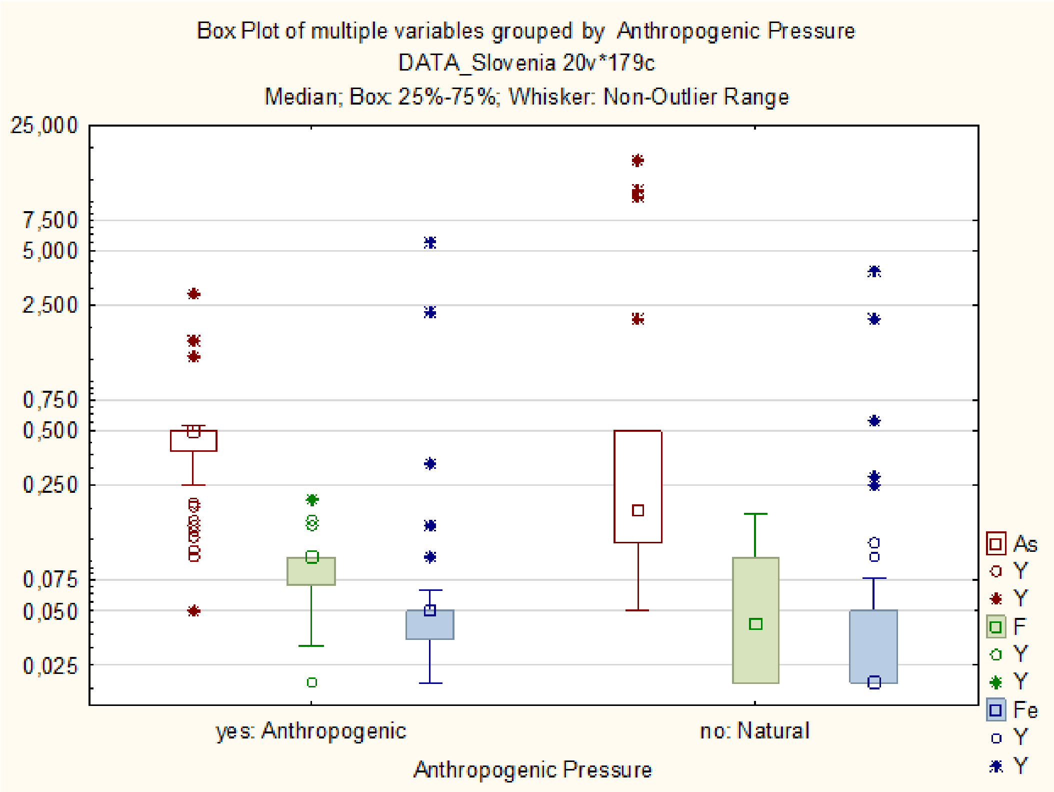


Figure 77: Comparison of sites with presumed anthropogenic pressures and without them for As, F and Fe. Note the logarithmic scale.

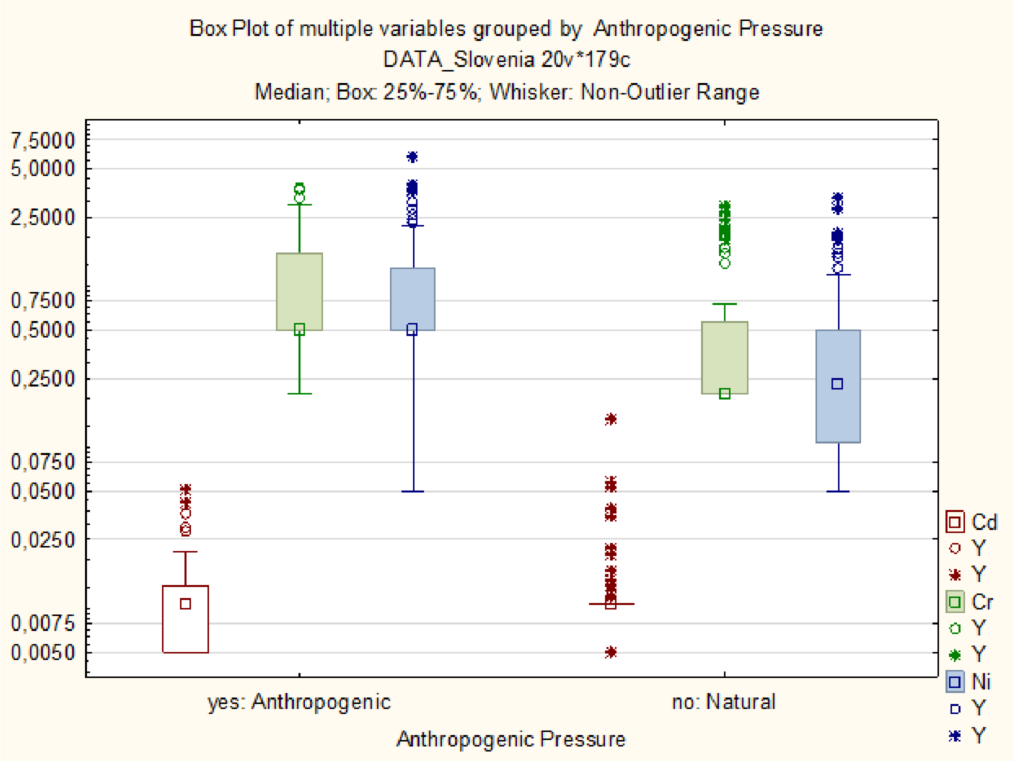


Figure 78: Comparison of sites with presumed anthropogenic pressures and without them for Cd, Cr and Ni.

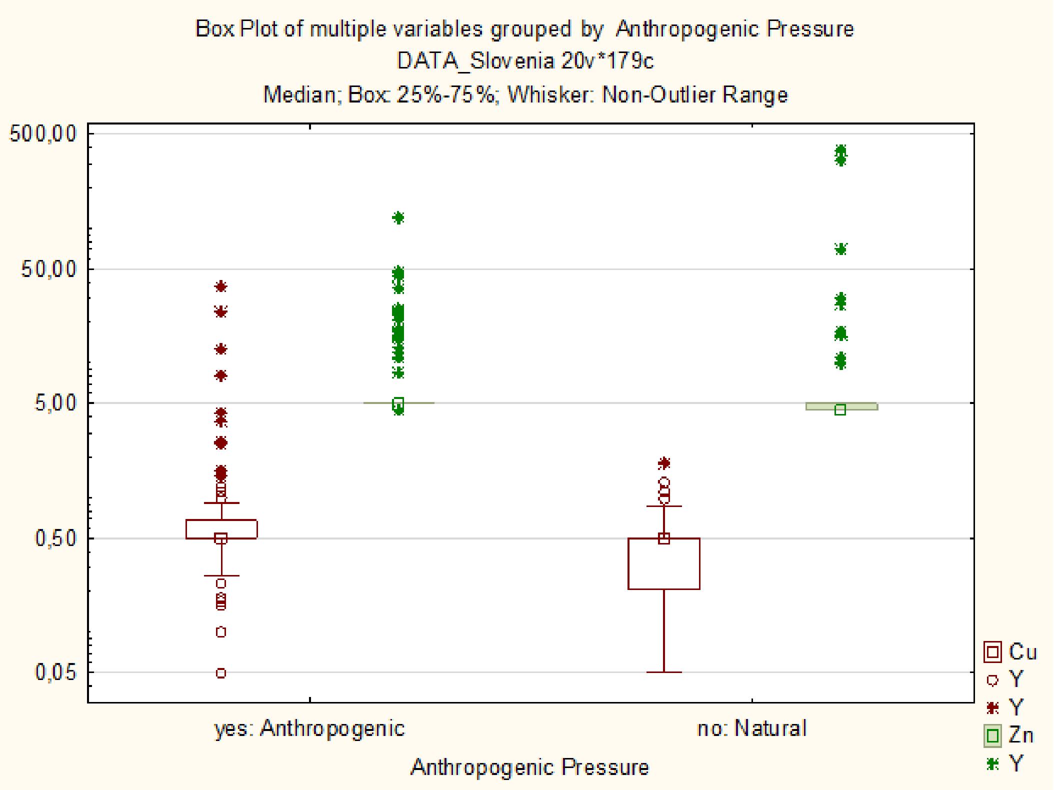


Figure 79: Comparison of sites with presumed anthropogenic pressures and without them for Cu and Zn, which show some very high outliers.

10.4.5 *Statistics of data without anthropogenic pressures in the recharge area*

Table 89: Descriptive statistics of the dataset without anthropogenic pressures, a total of 87 sites

	SO4 (mg/l)	As (µg/L)	Cd (µg/L)	Cr (µg/L)	Cu (µg/L)	F (mg/L)	Fe (mg/L)	Ni (µg/L)	NO3 (mg/L)	Zn (µg/L)
N	83	86	87	87	87	83	86	87	87	87
Mean	7.0	0.70	0.012	0.6	0.43	0.06	0.11	0.48	6.8	14.3
St.dev	5.7	2.29	0.016	0.7	0.30	0.04	0.47	0.62	5.7	52.6
Median	4.9	0.18	0.010	0.2	0.50	0.04	0.02	0.23	4.9	4.5
Min	0.5	0.05	0.005	0.2	0.05	0.02	0.02	0.05	1.1	4.5
P10	1.8	0.05	0.005	0.2	0.13	0.02	0.02	0.05	1.6	4.5
P25	3.0	0.12	0.010	0.2	0.21	0.02	0.02	0.10	2.9	4.5
P75	10.0	0.50	0.010	0.6	0.50	0.10	0.05	0.50	8.9	5.0
P90	**14.0**	**0.50**	**0.014**	**2.0**	**0.69**	**0.10**	**0.08**	**1.40**	**15.5**	**10.0**
LOQ/2$_{max}$	*/*	*0.5*	*0.01*	*0.5*	*0.5*	*0.1*	*0.05*	*0.5*	*1.1*	*5.0*
Max	31.5	16.00	0.140	2.9	1.80	0.18	3.90	3.30	26.8	380.0

10.4.6 *Statistics of data with anthropogenic pressures in the recharge area*

Table 90: Descriptive statistics of the dataset with anthropogenic pressures, a total of 92 sites

	SO4 (mg/l)	As (µg/L)	Cd (µg/L)	Cr (µg/L)	Cu (µg/L)	F (mg/L)	Fe (mg/L)	Ni (µg/L)	NO3 (mg/L)	Zn (µg/L)
N	92	92	91	91	92	92	92	92	92	92
Mean	13.6	0.46	0.011	1.0	1.55	0.08	0.13	0.96	19.2	9.4
St.dev	8.4	0.34	0.008	0.8	4.73	0.04	0.62	1.00	14.8	14.3
Median	12.4	0.50	0.010	0.5	0.50	0.10	0.05	0.50	16.5	5.0
Min	0.5	0.05	0.005	0.2	0.05	0.02	0.02	0.05	1.1	4.5
P10	2.8	0.13	0.005	0.2	0.26	0.02	0.02	0.15	3.2	4.5
P25	7.4	0.39	0.005	0.5	0.50	0.07	0.04	0.50	5.3	5.0
P75	19.3	0.50	0.013	1.5	0.68	0.10	0.05	1.20	32.1	5.0
P90	**25.0**	**0.50**	**0.018**	**2.2**	**1.60**	**0.10**	**0.05**	**2.20**	**41.5**	**21.0**
LOQ/2$_{max}$	*/*	*0.5*	*0.01*	*0.5*	*0.5*	*0.1*	*0.05*	*0.5*	*1.0*	*5.0*
Max	37.5	2.90	0.051	3.8	37.00	0.21	5.60	5.90	49.0	120.0

10.4.7 *Comparison of calculated 90th percentiles regarding all samples, sampled without and with anthropogenic pressures in the recharge area*

A comparison of determined values by separating data with presumably anthropogenic influence shows that:

- We cannot evaluate the NBLs for two parameters, As and F, as too much data is bellow detection limit and their 90th percentiles are identical to the half of max LOQ.
- The 90th percentile increases from the ‘natural’ group over the averaged ‘all samples’ group to the ‘anthropogenic’ group. This is obvious for sulfate, Cd, Cu, Ni, nitrate and Zn.
- In case of iron, only ‘natural’ group shows 90th percentile higher than half of max LOQ.

A comparison to previous archive studies for nitrate and sulfate shows:

- Mezga (2014), who used only sampling points with minimum possibility to be affected by anthropogenic activities, has determined much lower NBLs for nitrate and sulfate than are our values of the ‘natural’ group. This probably indicates that many points of the Slovenian national monitoring network are noticeable affected by anthropogenic pressures so by analyzing them we will get affected concentrations of at least some anthropogenic contaminants.

- Serianz et al. (2020) has applied two methods. Their NLB for sulphate (17.0 mg/l) gained by the Probability method is only slightly higher than our value of the 'natural' group (14.0 mg/l), while our 'anthropogenic' group is very similar to their max NBL gained by the Pre-selection method (26.0 mg/l). However, we see large difference in nitrate NBLs. The main reason is that their methodology based on removal of all values (not analyses but just nitrate concentration) which exceeded 10 mg/l of nitrate while we did that for values exceeding 50 mg/l only. Consequently, nitrate NBL (6.0 mg/l) was almost twice the one of Mezga (3.8 mg/l; 2014), while our from the 'natural' group (15.5 mg/l) was more than twice higher than of Serianz et al. From this we can assume, that nitrate NBLs are noticeably affected by anthropogenic (mainly agricultural) pressures and are not valid values for NBLs.

Table 91: Comparison of the 90th percentiles as determined by three datasets

	SO4 (mg/l)	**As (µg/L)**	**Cd (µg/L)**	**Cr (µg/L)**	**Cu (µg/L)**	**F (mg/L)**	**Fe (mg/L)**	**Ni (µg/L)**	**NO3 (mg/L)**	**Zn (µg/L)**
SI drink. wat. limit	*250*	*10*	*5.0*	*50*	*2000*	*1.5*	*0.200*	*20*	*50*	*/*
$LOQ/2_{max}$	*/*	***0.5***	***0.01***	***0.5***	***0.5***	***0.1***	***0.05***	***0.5***	*1.1*	***5***
$P90_{natur}$	**14.0**	**0.50**	**0.014**	**2.0**	**0.69**	**0.10**	**0.08**	**1.40**	**15.5**	**10.0**
$P90_{all}$	21.0	0.50	0.016	2.0	1.10	0.10	0.05	1.90	34.5	17.0
$P90_{anthr}$	25.0	0.50	0.018	2.2	1.60	0.10	0.05	2.20	41.5	21.0
Mezga, 2014	*6.93*								*3.81*	
Serianz et al. 2020	*26.0 / 17.0*								*6.3 / 6.0*	

10.5 Comparison of NBLs among various lithologies in Slovenia

We did not interpret the metamorphic rocks as only one sample fits this group. For the rest, NBLs of all samples (179 altogether) are shown in Table 92 and as determined for sites with lowest identified anthropogenic impact in Table 93.

Evaluating all samples it is evident that carbonates differ from the rest (sands, gravels and alluvium). They have much lower NBLs of sulfate (10.4 mg/l) and nitrate (6.6 mg/l) and very low metals which are just at LOQ (Cd, Ni, Zn, Fe) or slightly above (Cr) while Cu is at 0.98 µg/L. Sands and gravels are quite alike, and we are not totally certain that lithological classification is reliable, so the differences are not very certain. What is evident is that alluvium has very high nitrates (due to proven agricultural pollution), and Neogene sands elevated iron, which complied with hydrogeological conceptual models.

Table 92: Comparison of the 90th percentiles for four HOVER lithologies when all samples (179 sites) are used

	SO4 (mg/l)	**Cd (µg/L)**	**Cr (µg/L)**	**Cu (µg/L)**	**Ni (µg/L)**	**Zn (µg/L)**	**NO3 (mg/L)**	**Fe (mg/L)**
SI drinking water limit	*250*	*5.0*	*50*	*2000*	*20*	*/*	*50*	*0.200*
$LOQ/2_{max}$	*/*	*0.01*	*0.5*	*0.5*	*0.5*	*5*	*1.1*	*0.05*
N	62	66	66	66	66	66	66	65
$P90_{sedim\_carbonates}$	**10.4**	**0.011**	**0.6**	**0.98**	**0.51**	**5.0**	**6.6**	**0.05**
N	19	18	19	19	19	19	19	19
$P90_{sedim\_sands}$	**25.0**	**0.018**	**2.4**	**1.30**	**2.80**	**47.5**	**29.5**	**2.10**
N	13	13	13	13	13	13	13	13
$P90_{sedim\_gravel}$	**21.5**	**0.016**	**2.9**	**1.10**	**1.40**	**18.0**	**24.5**	**0.05**
N	80	80	79	80	80	80	80	80
$P90_{alluvium}$	**24.0**	**0.020**	**2.2**	**1.15**	**2.20**	**24.5**	**43.0**	**0.05**

Interpretation of solely ‘natural’ group reveals that NBLs do not noticeable change in the case of Cd and Cr.

Lower NBLs are evident for carbonates in comparison to clastic sediments, still, in general. NBL for sulphate is now evaluated in carbonates to be slightly lower - 8.2 mg/l, and much lower in clastics, where we now apply about 15.6 mg/l.

Regarding nitrates, carbonates have NBL concentrations 6.6 mg/l, which is comparable to Serianz et al. (2020), but in alluvium about 24.3 mg/l by this approach. Sand and gravel are similar, about 15 mg/l both.

NBL of iron in carbonates and gravel is the LOQ/2, slightly higher in alluvium with 0.12 mg/l, and the highest in sands, up to 3.9 mg/l. This is in accordance with the usually redox conditions in such clastic sediments. But it is worth mentioning here that only little natural samples are above LOQ: 76 samples below LOQ and only the rest 10 samples are between 0,057 and 3,9 mg/l; of this 2 lowest in carbonate rocks and the rest alluvium and sand. For comparison, in samples with anthropogenic pressure, 86 samples are below LOQ and 6 the rest with above from 0,065 to 5,6 mg/l: one in carbonates, most in sands, and the rest clastics.

Cupper has low concentrations in carbonates. NBL is determined to be 0.64 µg/L there, while it is not detectable in gravels and our best evaluation for NBL in alluvium is 0.70 µg/L.

Zinc is below LOQ (5 µg/L) in carbonates and gravel, NBL in alluvium is about 27.0 µg/L, and some local extreme values occur in sands with NBL of 380.0 µg/L.

Elevated NBL in sands are evident for Ni (3.3 µg/L), it is about 1.5 µg/L in alluvium, and at LOQ in carbonates.

Table 93: Comparison of the 90th percentiles for four HOVER lithologies when only samples with no presumed antropogenic impact (87 sites) are used

	SO4 (mg/l)	Cd (µg/L)	Cr (µg/L)	Cu (µg/L)	Ni (µg/L)	Zn (µg/L)	NO3 (mg/L)	Fe (mg/L)
SI drinking water limit	*250*	*5.0*	*50*	*2000*	*20*	*/*	*50*	*0.200*
LOQ/2$_{max}$	*/*	*0.01*	*0.5*	*0.5*	*0.5*	*5*	*1.1*	*0.05*
N	47	51	51	51	51	51	51	50
***P90**$_{sedim\_carbonates}$*	**8.2**	**0.010**	**0.5**	**0.64**	**0.50**	**4.5**	**6.6**	**0.05**
N	9	9	9	9	9	9	9	9
***P90**$_{sedim\_sands}$*	**15.5**	**0.014**	**2.4**	**1.30**	**3.30**	**380.0**	**16.5**	**3.9**
N	4	4	4	4	4	4	4	4
***P90**$_{sedim\_gravel}$*	**13.0**	**0.011**	**2.9**	**0.50**	**1.40**	**5.0**	**14.0**	**0.05**
N	22	22	22	22	22	22	22	22
***P90**$_{alluvium}$*	**15.6**	**0.020**	**2.0**	**0.70**	**1.50**	**27.0**	**24.3**	**0.12**

Statistical analysis of NBLs differences among various lithologies and anthropogenic influences in Slovenia

For testing the statistical significance of NBLs differences among various groups of lithology and anthropogenic influence we used nonparametric Mann–Whitney U test. The results of Mann–Whitney U test for NBLs at the 95-percent confidence level ($\alpha = 0.05$) are listed in Table 12.

Table 94: Results of the Mann-Whitney U test for NBLs in the groundwater between different lithology groups. The significant differences at 95 % confidence level (α = 0.05) are marked as red.

		Carb./Sand	Carb./Alluv.	Carb./Gravel	Sand/Alluv	Sand/Gravel	Alluv/Gravel	Anthr./Naur
SO4	Z	-2.27	-7.65	-4.01	-1.45	-0.06	1.06	5.50
	p-value	0.023	0.000	0.000	0.148	0.954	0.290	0.000
As	Z	-5.87	-7.78	-5.27	1.30	-0.02	-1.22	4.42
	p-value	0.000	0.000	0.000	0.194	0.984	0.221	0.000
Cd	Z	0.15	2.15	1.98	0.69	1.04	0.68	0.11
	p-value	0.883	0.031	0.047	0.491	0.298	0.496	0.910
Cr	Z	-4.37	-7.93	-5.28	-1.14	-2.24	-2.18	3.86
	p-value	0.000	0.000	0.000	0.254	0.025	0.029	0.000
Cu	Z	-3.53	-5.27	-3.13	-0.04	-0.15	-0.12	4.13
	p-value	0,000	0.000	0.002	0.968	0.878	0.903	0.000
F	Z	-5.71	-6.84	-4.91	0.92	-0.23	-1.00	3.68
	p-value	0.000	0.000	0.000	0.358	0.818	0.316	0.000
Fe	Z	-5.28	-7.04	-4.85	2.48	1.04	-1.09	3.20
	p-value	0,000	0.000	0.000	0.013	0.300	0.275	0.001
Ni	Z	-4.93	-7.76	-4.57	0.20	-0.17	-0.38	4.86
	p-value	0.000	0.000	0.000	0.845	0.863	0.706	0.000
NO3	Z	1.32	-8.35	-4.30	-4.54	-2.47	1.79	5.85
	p-value	0.186	0.000	0.000	0.000	0.013	0.074	0.000
Zn	Z	-4.81	-7.81	-4.89	0.16	-0.06	-0.19	4.37
	p-value	0.000	0.000	0.000	0.873	0.954	0.846	0.000

Carb. – Sedimentary: Carbonates; Sand – Sedimentary: Sands; Alluv. – Alluvium; Gravel – Sedimentary: gravel; Anthr. – anthropogenic influence; Natur – natural

In general, the results show significant differences between Carbonates and all other lithological groups at the 95% confidence level (α = 0.05). This stands for all NBLs except cadmium and nitrate in case of Carbonates and Sands.

Significant differences between Sands and Alluvium are found for iron and nitrate. The nitrate also distinguishes among Sands and Gravel. Differences are also observed for chromium between Sands and Gravel groups, and Alluvium and Gravel groups. Other NBLs within these groups do not show statistically significant differences at a 95% confidence level (α = 0.05).

According to the anthropogenic pressures results show significant differences between Anthropogenic influences and Natural state for all NBLs except for cadmium at the 95% confidence level (α = 0.05). Cadmium sources could be anthropogenic (mining, atmospheric deposition of combustion emissions, Cd-containing fertilizers) or natural (incorporated into sulfides, carbonates and phosphorites).

11 CASE STUDY - CATALONIA, SPAIN

11.1 Previous studies on NBLs

Studies for the calculation of the NBL in Catalonia are scarce to date. The Catalan Water Agency carries out an evaluation of the qualitative status of the groundwater bodies in Catalonia every four years following the indications of the Water Frame Directive (WFD). In these works, the NBL is calculated following the BRIDGE methodology (Muller et al, 2006) for each growndwater body. The Bridge methodology establishes a constraints for the selection points process. For some elements, such constraints may not be the most appropriate since they provide values above the values expected under natural conditions. Thus, e.g. values are provided for SO_4 that exceed 250 mg/l in several groundwater bodies.

In the framework of the GEOERA projects, the WP5 TACTIC project is foccused on comparing methodologies developed to establish the Cl NBL in vulnerable areas potentially affected by Salt Water Intrusion (SWI).

Other works has also been performed in more limited areas for the establishment of the NBL simultaneously with this work in high mountain karst aquifers. Some of the related references are:

- Herms, I; Jódar, J.; Soler, A.; Vadillo, I; Lambán, L.J.; Martos-Rosillo, S.; Custodio, E.; Jorge, J.; Agustí Núñez, J. (2019). Baseline hydrogeochemical characterisation of a vulnerable pristine highmountain karst aquifer in the southeastern Pyrenees, 46th IAH Congress,22 - 27th September 2019. Málaga, ESP
- Herms, I; Jódar, J.; Soler, A.; Lambán, L.J.; Custodio, E.; Agustí Núñez, J.; Arnó, G.; Ortego, M.I.; Parcerisa, D.; Jorge, J. (2020). Evaluation of natural background levels of high mountain karst aquifers in complex hydrogeological settings. A Gaussian mixture model approach in the Port del Comte (SE, Pyrenees) case study. November 2020 Science of The Total Environment. DOI: 10.1016/j.scitotenv.2020.143864

11.2 Study with the proposed method

11.2.1 *Constitution of the groundwater quality dataset (steps 1 – 6)*

The data used to build the growndwater quality database have been taken from the hydrogeological database (BDH) managed and nurtured by the Catalan Water Agency (ACA). The database includes data from supply wells, wells for several uses, and monitoring points. In addition, surface water data and groundwater level data are also included in BDH.

For this study, the water quality data available for the entire time series, in the internal basins of Catalonia have been taken. Therefore, the external basins that provide water to the Ebro basin have been excluded from this study. The work area and the initial availability of sampling points is shown in Figure 80.

Figure 80: Situation map of the internal basins in Catalonia, in the Northeast of the Iberian Peninsula

The reason for having taken the entire available time series and, therefore, not having restricted the study to the period of time indicated in the HOVER WP3 methodology for task 3.3 (2010-2020), is that in the previous works carried out by ACA (in which NBL are established for the groundwater bodies) it has been found that when making the data selection (following the BRIDGE method), in some water masses it has not been possible to establish a value due to the low number of data that have remained. Since the objective was to test the proposed method, it has been considered the best option to have a large number of data and a wide area.

Therefore, for the construction of the groundwater quality database, the following have been taken:

- Groundwater quality data for the period of time between 1994 and 2018.
- Data on supply wells, wells for all other uses uses and control points.

First quality datacheck

Subsequently, a data treatment was carried out to filter the data with a low quality or erroneous data. The filtering and treatment carried out is detailed below.

A first filtering based on the values of the physical-chemical parameters has been carried out.

- Analysis with pH values higher than 10 have been removed.
- As for the temperature, the analytical values with values outside the range between 30 and 2 C° have been removed.
- Regarding dissolved oxygen, values higher than 14 mg/l have been eliminated as they are considered erroneous. In this sense, the limited number of available data, added to some data of high uncertainty (numerous analyzes with values higher than 14 mg/l) has led to not taking into account the dissolved oxygen data.

On the other hand, analytics with anomalously high values have also been filtered for each element considered.

In addition, analytics that include values of type "Above a value" (>) have also been left out of this treatment. Table 95 shows the analytics available for each element and the proportion of left censored values or values below the Limit Of Detection (LOD).

Element	unit	All (n)	< LOD (n)	< LOD (%)
pH	pH units	2398	-	-
As	µg/L	2458	784	31.90
Cd	µg/L	7061	6554	92.82
Cu	µg/L	8243	2677	32.48
Cr	µg/L	8162	6862	84.07
Fe	µg/L	5963	2131	35.74
Mn	µg/L	8108	3377	41.65
Ni	µg/L	7077	4061	57.38
Pb	µg/L	6560	5262	80.21
Zn	µg/L	7322	1746	23.85
NO3	mg/l	8356	1252	14.98
Cl	mg/l	8338	208	2.49
Na	mg/l	8176	151	1.85
K	mg/l	8287	299	3.61
SO4	mg/l	8301	66	0.80

Table 95: Number of analytics available for each element and the proportion of left censored values (LOD=Limit Of Detection).

Treatment of left censored values (LOD).

Laboratories frequently report analytical results that include values below the limit of detection or quantification. In the case of the BBDD of the Catalan Water Agency (BDH), results of this type are very frequent, mainly for some trace elements. As shown in Table 95, for some elements included in the initial proposal, the number of values "<LOD" exceeds 90% of the total. Thus, the elements considered to apply the methodology have been those with a maximum number of "<LOD" values representing the 33% of all the available samples.

To carry out the transformation of the values, the ”Log-Ratio EM Algorithm” method has been used mainly with the lrEM function for R-studio software (package [zCompositions v1.3.4]). In short, this function assigns values below the established LOD value, considering multivariate character of the hydrogeochemical samples.

For some items, the number of left-censored values is around 30%. In addition, since the analytics come from different laboratories depending on the work area, there are different types of "<LOD" values for the same element.

Since each calculation with this function can only be performed for a single "<LOD" value of each variable, and there are many different "<LOD" values in the same variable, numerous combinations have been necessary performed .

Finally, for some analytics in which the rest of the values do not provide information to apply the lrEM function, the Wilcoxon transformation method has been applied by calculating "<LOD"/2.

Table 96 shows a summary of the "LCV" value transformations performed. It can be seen that the values obtained are not too far from those that would be obtained using Wilcoxon.

Element	LOD range (min-max)	max substitute concentration lrEM	max substitute concentration LOD/2
pH		-	
As	0.01 - 1	0.74	0.5
Cd	-	-	
Cu	0.5 - 10	4.13	2.5
Cr	1 - 5	-	
Fe	5 - 50	-	
Mn	1 - 20	-	
Ni	1 - 5	-	
Pb	1 - 5	-	
Zn	0.1 - 5	2.59	2
NO3	0.1 - 5	3.16	2.5
Cl	1 - 10	9.28	-
Na	5 - 50	26.91	-
K	1 - 10	5.78	-
SO_4	8 - 10	8.66	4

Table 96: LCV transformation mainly with lrEM function with R-studio. Maximum substituted values

Groundwater quality dataset

The groundwater quality dataset contain 11 variables. In addition, the physical-chemical parameters temperature and pH have been included. As it has been already commented, available data of dissolved oxygen and the redox potential are not enough to characterize each sampling point with the proposed REDOX conditions classification. In all the variables, the dissolved oxygen available data are only in less than 50% of available Samples. Furthermore, in several sampling points there is only one or two values of dissolved oxygen, wich is not enough to calculate a representative value.

With the available data of pH, the HOVER-pH was defined and added to the quality dataset. The following Table 97 shows the number of sampling points for each element in relation to the HOVER-pH classification.

HOVER-pH classes	HCO3	Ca	Mg	As	Cu	Zn	NO3	Cl	K	Na	SO4
Acidic (pH < 7)	88	88	85	54	88	83	88	87	87	85	87
Basic (pH > 7.5)	351	351	350	187	349	341	351	351	350	346	350
Neutral (7 ≤ pH ≤ 7.5)	564	564	559	357	564	540	564	564	561	551	561
unknown	355	355	354	149	340	305	356	355	355	348	354
unknown (%)	26.14	26.14	26.26	19.95	25.35	24.03	26.20	26.16	26.24	26.17	26.18

Table 97: HOVER-pH classification. Number of sampling points for each element.

11.2.2 *Hydrogeological characteristics of the sampling points (Steps 7 – 8)*

The next step has been to assign to each sampling point, a HOVER-lithological class and HOVER-age formation/Stratigraphy. Assignment for 1359 sample points has been made.

In this step, we have started from the aquifers map of Catalonia at 1:50.000 scale carried out by the Catalan Water Agency (2013). In this work, the aquifers of both quaternary and substrate materials are delimited.

Figure 81 shows the assignment of the HOVER lithology to each one of the delimited aquifers with the location of the available sampling points in the internal basins of Catalonia.

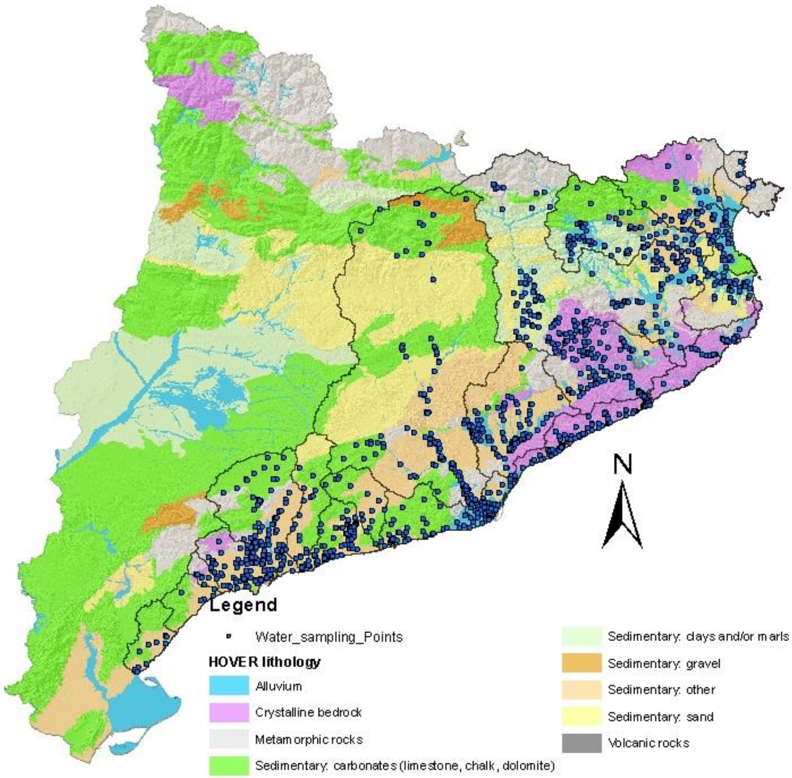


Figure 81: HOVER-lithology and available sampling points for internal basins in Catalonia.

The following Table 98 shows the categories assigned for HOVER-lithology and HOVER-Age formation/stratigraphy.

Lithology HOVER	Hover_strat
Sedimentary: sand	Quaternary
Sedimentary: gravel	Palaeozoic
Sedimentary: carbonates (limestone, chalk, dolomite)	Paleogene
Sedimentary: clays and/or marls	Neogene
Sedimentary: other	Triassic
Volcanic rocks	Mesozoic-Paleogene
Crystalline rocks	Jurassic-Cretaceous
Metamorphic rocks	Jurassic-Cretaceous-Triassic
Alluvium	
Other	

Table 98: Hover-lithology and HOVER-Age formation/stratigraphy assigned to sampling points in this work.

It has been difficult to assign the sampling points located in multilayer sedimentary aquifers. In most cases, these aquifers consist of very heterogeneous layers with a similar distribution, so it is difficult to assign a single HOVER lithological class to these formations. In this case, the "Sedimentary: other" group has been used. Furthermore, in some multilayer aquifers, layers of carbonate calcareous or detrital levels are also intercalated, which makes it difficult to separate carbonate and non-carbonate sedimentary lithologies. In this case it has been characterized as carbonated, if it is the predominant lithology

On the other hand, the granite massifs that contain metamorphic rocks (contact and regional) have not been distinguished if they are not delimitated separately the original aquifers map.

To summarize, given the breadth of the work area and its structural and lithological heterogeneity, it is difficult to assign a HOVER lithological class to each sampling point. Essentially because most of the sampling points exploit water from different lithologies, and in many cases, also from different aquifers.

Better results may be obtained with more detailed lithological discrimination, but a more detailed lithological assignment for each water point is not available for this work. If more samples were available at certain location, it would be necessary to carry out a deeply study in which the geology associated to each sampling point is known in more detail.

Regarding the assignment of the Age formation / stratigraphy classification, we have proceeded in a similar way. As in the HOVER lithology assignment, the difficulty has been to assign a single category to multilayer aquifers, many of which include layers of different ages and stratigraphic formations. Therefore, it has been necessary to group different classes or create a new category to be able to assign a class to some formations that are too wide in age. This classification has only been assigned to each sampling point but has not been used in the statistical treatment due to its high degree of uncertainty.

11.2.3 *Creation of the pressures database and GIS. Anthropogenic pressures / influences: diffuse or point pollutions (Steps 10 – 12).*

To be able to select and eliminate the sampling points with potential contamination by anthropogenic pressures, it is necessary to characterize the points subjected to the respective pressures. For this work, the option of characterizing the pressure to which each water point may be subjected (taking into account the use of the well described in the BDH) has not been considered, since it has been verified on several occasions that the use of the well (according to the BDH) may not be related to the main activity carried out in the area near the well.

To characterize the anthropogenic pressures for each sampling point, the method applied by GEUS has been followed to assign the main pressures to which each sampling point is subjected.

An area of influence of 1km buffer has been determined for each sampling point with the assumption that the real area may be larger, smaller, with a different shape, etc. It is still considered that it may be an approach given the regional scale of the work area.

The land-cover types have been taken from the CORINE Land Cover (CLC-2018) map. The land-cover classes included in CLC-2018 have been grouped (Table 99) and crossed with the 1 km buffer for each sampling point. Figure 82 shows a work area near the city of Girona, and part of the process of applying the method.

Table 99: Equivalences established between the CLC-18 map and the anthropogenic pressures defined for HOVER WP3.

Hover anthropogenic pressure	values in master dataset	CLC-18 codes	CLC-18 description
Urban pressure	Yes/ No	1.1.1 / 1.1.2	Continuous and discontinues urban areas
Industrial pressures	Yes/ No	1.2 Level	Industrials and transport areas
Agricultural pressure	Yes/ No	all codes included in level 2	Area with development of all kinds of agricultural activity
Mining influence	Yes/ No	1.3 level	Mining, landfill areas

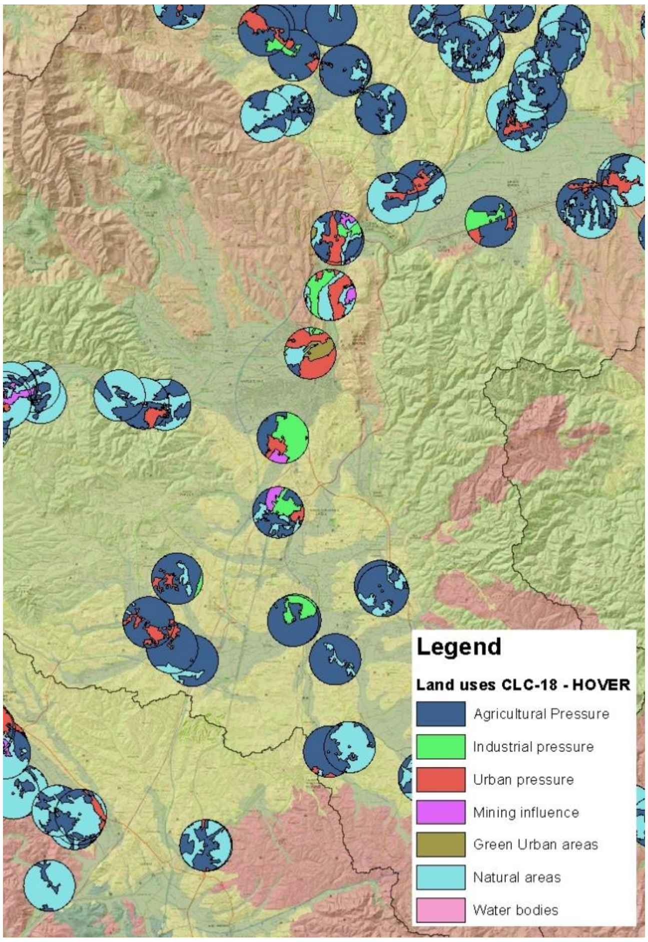


Figure 82: Characterization of the anthropogenic pressures included in a 1 km buffer for each sampling point in an area near the city of Girona.

Finally, for each point the anthropogenic activity with the largest area within the 1 km buffer zone has been calculated and this has been designated as "prevailing pressure".

HOVER WP Ant. Press	Detected	Water sampling points	
		Number	%
Urban pressure	no	431	31.71
	yes	928	68.29
Industrial Pressure	no	700	51.51
	yes	659	48.49
Agricultural Pressure	no	153	11.26
	yes	1206	88.74
Mining Influence	no	1210	89.04
	yes	149	10.96

Table 100: Anthropogenic pressures detected for available sampling points and total % of points at which each pressure was found.

Prevailing pressure (class)	Water sampling points	
	Number	%
Agricultural	407	30
Industrial	327	24
Mining	105	8
Urban	465	34
Natural or other	55	4

Table 101: Number of points and percentage of the total characterized with each "prevailing pressure".

To summarize, the pressure that is present in more sampling points is the agricultural pressure, which is in the area of influence of 88% of the points (Table 100). Even so, the prevailing pressure in the area of influence with the greatest presence is urban pressure and agricultural pressure. They are the "prevailing pressure" in 34% and 30% of the points respectively (Table 101).

11.2.4 *Synthesis of all necessary data and first results in a single database (Step 13)*

With the filtering treatments of the water quality data set and with the hydrogeological and anthropogenic characterization, a master table dataset of sampling points has been prepared with 1359 points and 31 variables (attributes). The type of variable and the description of each one of them is shown in Table 102.

The “master table” consists of 4 main parts. The first contains the attributes that intrinsically characterize each sampling point. The second part contains the median of the concentrations of the major and trace elements included in the study. The third part contains the hydrogeological attributes which characterize each sampling point. Finally, the last part contains information about the anthropogenic pressures that can influence each point and the prevailing pressure.

Variable	Attribute type	Attribute description
Codi PA	txt	ID of sampling point assigned by ACA
Massa d'aigua	txt	Water body in which sampling point is located
Aqüífer	txt	Aquifer associated to each sampling point
Codi Aqüífer	txt	Id of aquifer
Topònim	txt	Name of water sampling point
Municipi	txt	Municipality where sampling point is located
UTM X	num	X coordinate in ETRS1989 ETRS-TM31 coordinate system
UTM Y	num	Y coordinate in ETRS1989 ETRS-TM31 coordinate system
Cota Z	num	Terrain elevation
Date_first_An	date	Date of first analysis
Date_last_An	date	Date of last analysis
HCO$_3$	num	Median HCO$_3$ (mg/l)
Ca	num	Median Ca (mg/l)
Mg	num	Median Mg (mg/l)
pH	num	Median of pH
Hover_pH_Classes	txt	HOVER-pH class assigned
As	num	Median As (µg/l)
Cu	num	Median Cu (µg/l)
Zn	num	Median Zn (µg/l)
NO$_3$	num	Median NO$_3$ (mg/l)
Cl	num	Median Cl (mg/l)
K	num	Median K (mg/l)
Na	num	Median Na (mg/l)
SO$_4$	num	Median SO$_4$ (mg/l)
Hover_lit	txt	HOVER-lithology class
Hover_strat	txt	HOVER-age / stratigraphy
Agricultural	txt	Agricultural activities present in 1km. Buffer ("YES/NO")
Urban	txt	Urban pressure present in 1km. Buffer ("YES/NO")
Industrial	txt	Industrial activities present in 1km. Buffer ("YES/NO")
Mining	txt	Mining activities present in 1km. Buffer ("YES/NO")
Prevailing Pressure	txt	Prevailing pressure in 1 km. buffer

Table 102: Structure of the master dataset after applying the steps described in the previous sections.

11.2.5 *Statistical treatments*

Discriminant function analysis (Step 14)

Regarding the statistical treatment, the first part has consisted in applying a discriminant function analysis method, to evaluate the possibility of separating the sampling points based on the available variables and the assigned HOVER-lithology category. The linear discriminant analysis (LDA) function has been chosen. The aim, therefore, is to evaluate whether the predictor variables (in this case the major and selected trace elements) can be discriminated in the different assigned HOVER- lithology groups. Statistical treatment was performed with RStudio.

As mentioned in the previous steps, the trace elements with a proportion of "LCV" data greater than 33% have been excluded from the analysis. Thus, the major elements and the traces As, Cu and Zn have finally been included in the statistical analysis. Furthermore, the data to characterize the samples regarding their REDOX condition have also been excluded. Regarding the HOVER-lithology class variable, the classes with the lowest number of associated points have not been considered either. The “Alluvium”, “Crystalline rock”, “Sedimentary: carbonates (limestone, chalk, dolomite)” and “Sedimentary: other” groups have been included in the initial statistical analysis.

Given that the LDA requires a normal distribution of the predictor variables and it is observed that all the variables are left-skewed, the Box-Cox transformation has been applied to normalize the data. Data have been split in training (80%) and test (20%) samples to build LDA model with the training data and test it with the test data.

With these assumptions, three different data models have been performed according to the variables shown in Table 103. For the first model, all the variables mentioned have been included. In the second model, the Cl and Na variables have been excluded since high values have been detected in coastal areas that, independently of the assigned HOVER-lithology, are considered that could bias the result. In the third model, the group "Sedimentary: others" was excluded. This group was already considered a "warehouse" group since it groups the assignments that presented most difficulties and with greatest uncertainty. It is observed that the best results are obtained when this group is eliminated. Therefore, only three HOVER-lithology classes have finally been used.

LDA models	Predicting	Predictor variables	sampling points	Accuracy	95% CI	Kappa
LDA_1	HOVER-lithology, Alluvium, Crystalline bedrock, Sedimentary Carbonates, Sedimentary detrital	Majority ions (except Na and Cl), As, Cu, Zn	727	56%	(49.18, 62.35)	0.35
LDA_2	HOVER-lithology, Alluvium, Sedimentary Carbonates, Sedimentary detrital	Majority ions (except Na and Cl), As, Cu, Zn	627	67.70%	(63.03, 72.08)	0.48
LDA_2	HOVER-lithology, Alluvium, Crystalline bedrock, Sedimentary Carbonates,	Majority ions (except Na and Cl), As, Cu, Zn	431	69.30%	(63.77, 74.55)	0.51

Table 103: Linear discriminant analysis (LDA). Model's variables and comparisons. The evaluation of the suitability of the model has been carried out based on the accuracy and Kappa (variant of accuracy).

The table shows the variables included in the LDA for each model. It is assumed that the best discrimination of data is obtained in model 3. Even so, it should be noted that the accuracy achieved is less than desirable.

The results show that better discrimination is achieved with the three HOVER lithological groups included in model 3. Although, a wide overlap area is observed (which was to be expected with the non-very high value of the accuracy obtained). It is observed that the discriminant function 1 (LD1), separates alluvium from sedimentary carbonates and discriminant function 2 (LD2) better separates Crystalline rock from Sedimentary carbonates (Figure 83).

For LD1 the Crystalline rock and Sedimentary carbonates groups are very similar. On the other hand, in LD2 there are greater differences between the three groups.

The explanation for the wide area of overlap is that there are many sampling points with very similar characteristics regardless of the lithological group assigned, as also observed in the discriminant linear analysis performed for models 1 and 2.

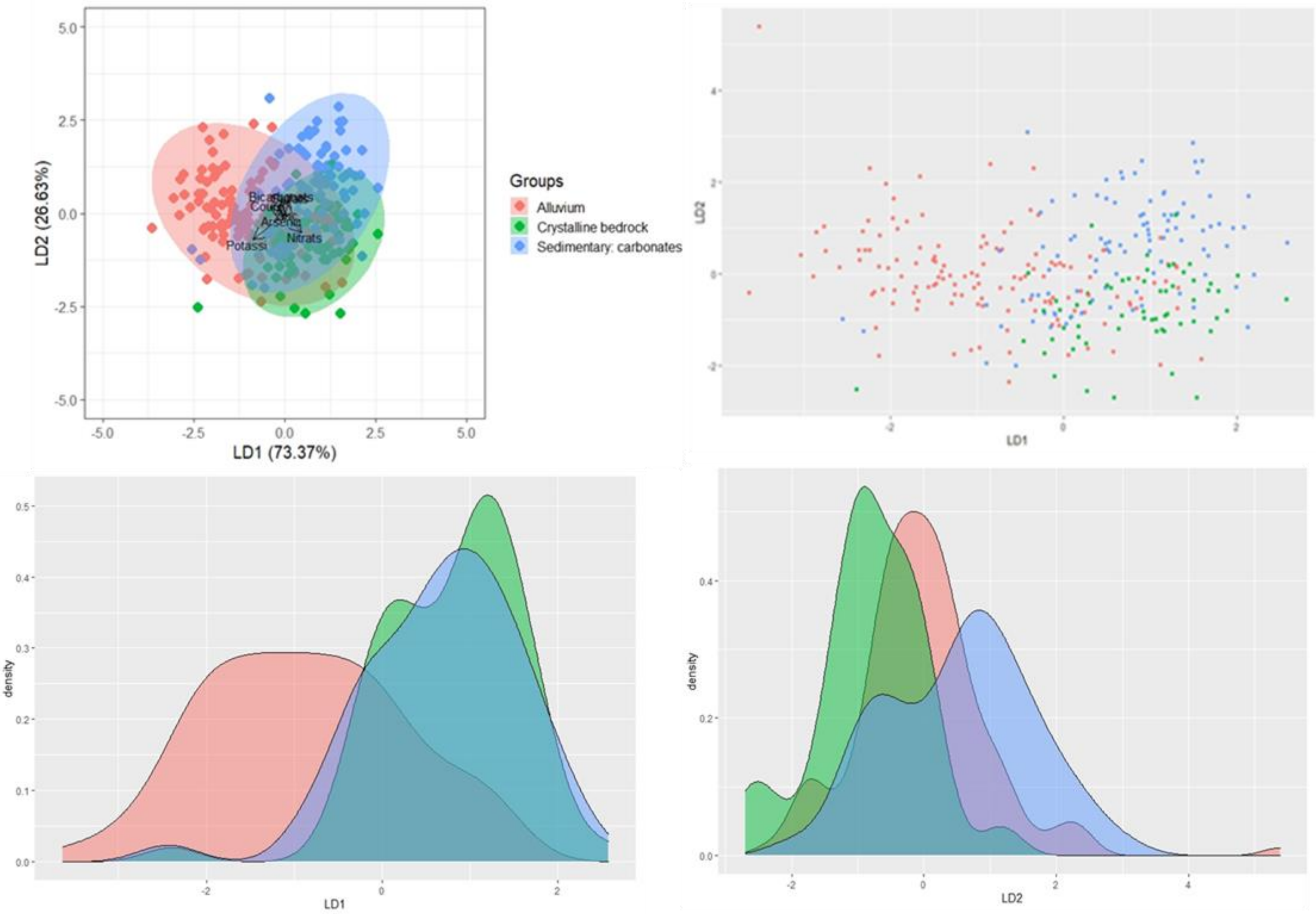


Figure 83: Linear discriminant functions for model 3. The percentage separation achieved by the LD1 is 73.37% and by LD2 is 26.63%. At the bottom left discriminant function 1 (LD1), separates alluvium from sedimentary carbonates and Crystalline rock. Discriminant function 2 (LD2), separates Crystalline rock from sedimentary carbonates.

With these results, it has been considered to calculate the NBL only for the HOVER lithological classes included in model 3: Alluvium, Sedimentary: Carbonates and Crystalline rock.

Non-parametric hierarchical univariate statistical treatments (Step 15)

With the non-parametric Kruskal-Wallis test it is possible to assess whether two or more groups come from the same data population i.e. if they have significant differences between them. The alternative hypothesis is that at least one of the groups has significant differences from the others. With Nemenyi post-hoc test it is possible to determine which group differs from the rest. Statistical significance is assessed at the 95% confidence level, i.e. significant differences are found if $p < 0.05$.

These tests have been applied with the categorical variables of the pH-HOVER class and the main anthropogenic pressure “Prevailing pressure” at each sampling point.

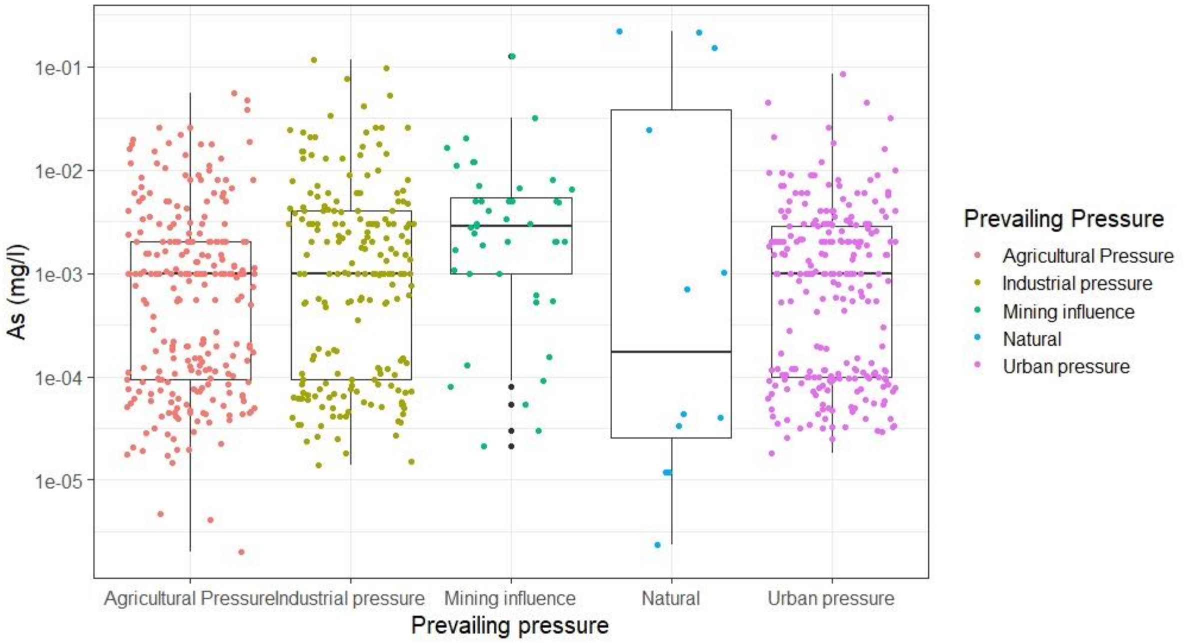
Table 104 shows that for the elements studied, in all cases there are significant differences between values, for at least one anthropogenic pressure. It is also observed that the three main "prevailing pressures” are urban and agricultural with similar percentages.

Element	Total	Agricultural Pressure		Industrial pressure		Mining influence		Natural		Urban pressure		Kruskal-Wallis test (p value <0.05)
		num	%	num	%	num	%	num	%	num	%	
As	747	251	33.6	201	26.9	44	5.9	12	1.6	239	32.0	p< 0.05 Significant difference
Cu	1341	406	30.3	323	24.1	105	7.8	55	4.1	452	33.7	p< 0.05 Significant difference
Zn	1269	385	30.3	309	24.3	95	7.5	53	4.2	427	33.6	p< 0.05 Significant difference
SO4	1352	407	32.1	322	23.8	104	7.7	55	4.1	464	34.3	p< 0.05 Significant difference

Table 104: Kruskal-Wallis test performed for a variable "prevailing pressure". Significant difference observed for all elements.

The results of the Nemenyi post-hoc test are shown below for As, Cu, Zn, SO4.

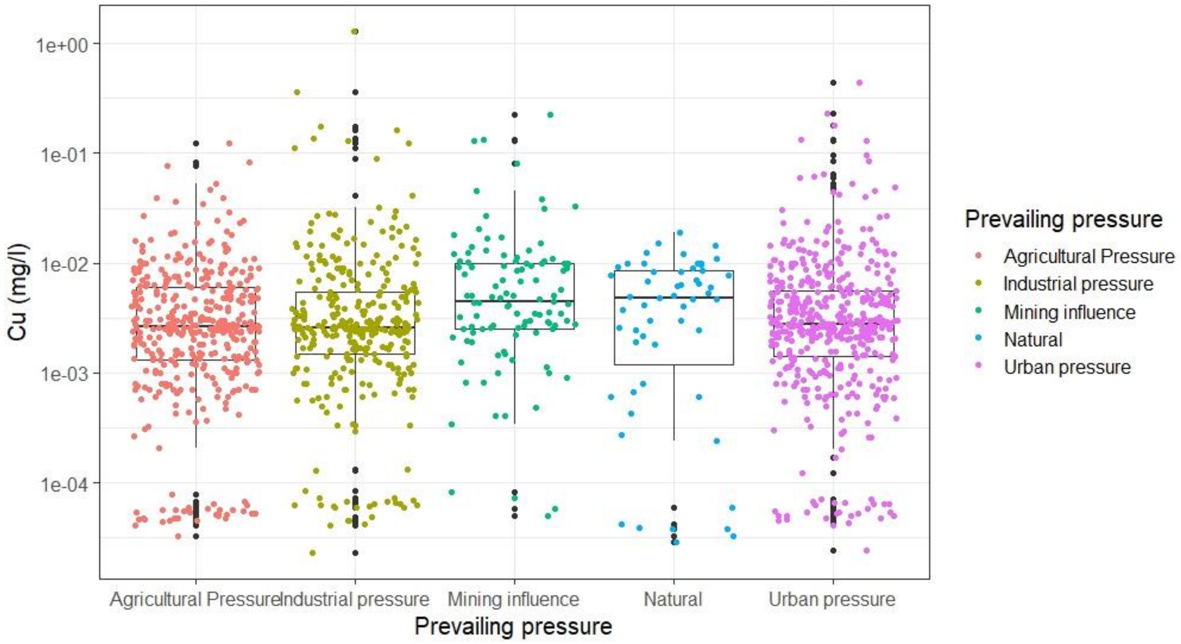
For Arsenic (Figure 84), the sampling points influenced by mining activities are those that show higher concentrations than the rest of the groups. Furthermore, in this case the group of points not influenced by anthropogenic pressures, has lower concentrations (median) but a greater dispersion of values. Furthermore, the values only represent 1.6%, so we have not considered them representative of values under natural conditions. Therefore, it has been decided to discard the Arsenic values from the "Mining influence" group.



	Agricultural Pressure	Industrial pressure	Mining influence	Natural	Urban pressure
Agricultural Pressure	**1**	0.259	**0.001**	0.964	0.947
Industrial pressure	0.259	**1**	0.095	0.895	0.612
Mining influence	**0.001**	0.095	**1**	0.741	**0.003**
Natural	0.964	0.895	0.741	**1**	0.927
Urban pressure	0.947	0.612	**0.003**	0.927	1

*Figure 84: Distribution of concentrations of As, for "Prevailing pressures" categories. Y-axis in log scale. Dots: Median concentration at each sampling point. Box Plot: The thick line corresponds to the median and the lower and upper limits correspond to Q1 (percentile 25th) and Q3 (percentile 75th). The upper and lower whiskers extend to the largest and smallest value (1.5 * IQR from the box limits). The table below shows the couples that have significant differences between them.*

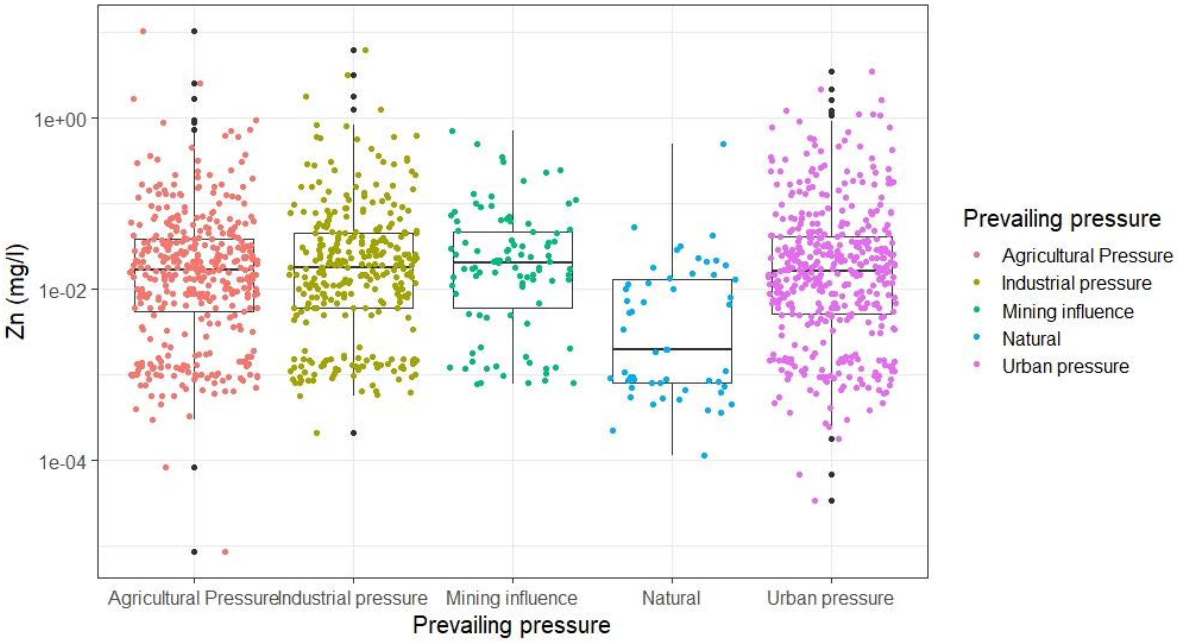
In the case of copper (Figure 85), analogously to arsenic, the group that presents differences with respect to the rest is that of "Mining influence". Therefore, it has been decided to discard the Copper concentration values from the "Mining influence" group too.



	Agricultural Pressure	Industrial pressure	Mining influence	Natural	Urban pressure
Agricultural Pressure	**1**	0.988	**0.000**	0.446	0.903
Industrial pressure	0.988	**1**	**0.001**	0.644	0.990
Mining influence	**0.000**	**0.001**	**1**	0.669	**0.001**
Natural	0.446	0.644	0.669	**1**	0.630
Urban pressure	0.903	0.990	**0.001**	0.630	1

*Figure 85: Distribution of concentrations of Cu, for “Prevailing pressures” categories. Y-axis in log scale. Dots: Median concentration at each sampling point. Box Plot: The thick line corresponds to the median and the lower and upper limits correspond to Q1 (percentile 25th) and Q3 (percentile 75th). The upper and lower whiskers extend to the largest and smallest value (1.5 * IQR from the box limits). The table below shows the couples that have significant differences between them.*

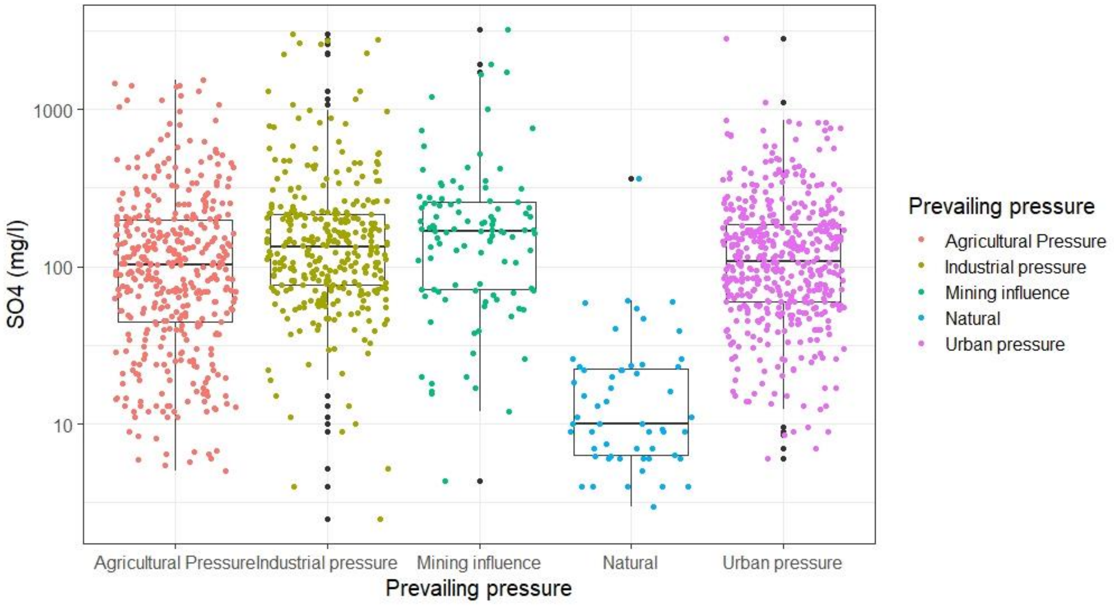
In the analysis carried out for zinc, the group that is different from the rest is the one that has not detected anthropogenic pressures, that is, the sampling points under natural conditions. In this case, they show lower concentrations than the rest. In this case, it has also been observed that most of the sampling points under natural conditions are points classified with lithology-HOVER Crystalline rock. Given that the population of the "natural" group is very small (4.1%) and that it could also be associated with a specific lithology and that there are no significant differences among the rest of the groups, it has been decided not to eliminate any group.



	Agricultural Pressure	Industrial pressure	Mining influence	Natural	Urban pressure
Agricultural Pressure	**1**	0.971	0.809	**<0,0001**	0.999
Industrial pressure	0.971	**1**	0.974	**<0,0001**	0.997
Mining influence	0.809	0.974	**1**	**<0,0001**	0.904
Natural	**<0,0001**	**<0,0001**	**<0,0001**	**1**	**<0,0001**
Urban pressure	0.999	0.997	0.904	**<0,0001**	1

*Figure 86: Distribution of concentrations of Zn, for "Prevailing pressures" categories. Y-axis in log scale. Dots: Median concentration at each sampling point. Box Plot: The thick line corresponds to the median and the lower and upper limits correspond to Q1 (percentile 25th) and Q3 (percentile 75th). The upper and lower whiskers extend to the largest and smallest value (1.5 * IQR from the box limits). The table below shows the couples that have significant differences between them.*

For the test carried out with the concentrations of sulphate ion SO_4, in this case differences are shown between practically all the groups with respect to the others. Similar pairs could only be established between agricultural and urban pressures, on the one hand, and mining influence and industrial pressure, on the other. The greatest difference is observed in the sampling points under natural conditions with respect to the others. In this case, given the notable difference and the high median values in the other groups (note that the 'y' axis is on a logarithmic scale) it has been considered that only the sampling points of the group "natural conditions" are not influenced by anthropogenic pressures.



	Agricultural Pressure	Industrial pressure	Mining influence	Natural	Urban pressure
Agricultural Pressure	**1**	**0.000**	**0.001**	**-<0,0001**	0.425
Industrial pressure	**0.000**	**1**	0.663	**-<0,0001**	**0.008**
Mining influence	**0.001**	0.663	**1**	**-<0,0001**	**0.007**
Natural	**-<0,0001**	**-<0,0001**	**-<0,0001**	**1**	**-<0,0001**
Urban pressure	0.425	**0.008**	**0.007**	**-<0,0001**	1

*Figure 87: Distribution of concentrations of SO4, for "Prevailing pressures" categories. Y-axis in log scale. Dots: Median concentration at each sampling point. Box Plot: The thick line corresponds to the median and the lower and upper limits correspond to Q1 (percentile 25th) and Q3 (percentile 75th). The upper and lower whiskers extend to the largest and smallest value (1.5 * IQR from the box limits). The table below shows the couples that have significant differences between them.*

On the other hand, in the case of applying the test with pH-HOVER as a categorical variable, no significant differences have been detected in any groups (Table 105, and Figure 88) into which the sampling points have been divided (pH-HOVER classes). Only between "Basic" and "neutral" groups for SO_4. In this case, neither of the two groups has been eliminated since it has not been identified with any contamination process. Thus, samples haven't been eliminated using the pH criteria.

Element	Total	Acidic		Basic		Neutral		Kruskal-Wallis test (p value <0.05)
		num	%	num	%	num	%	
As	747	54	7.2	187	25.0	357	47.8	p> 0.05
Cu	1341	88	6.6	349	26.0	564	42.1	p> 0.05
Zn	1269	83	6.5	341	26.9	540	42.6	p> 0.05
SO4	1352	87	6.4	350	25.9	561	41.5	p< 0.05 Significant difference

Table 105: Kruskal-Wallis test performed for a variable "pH-HOVER".

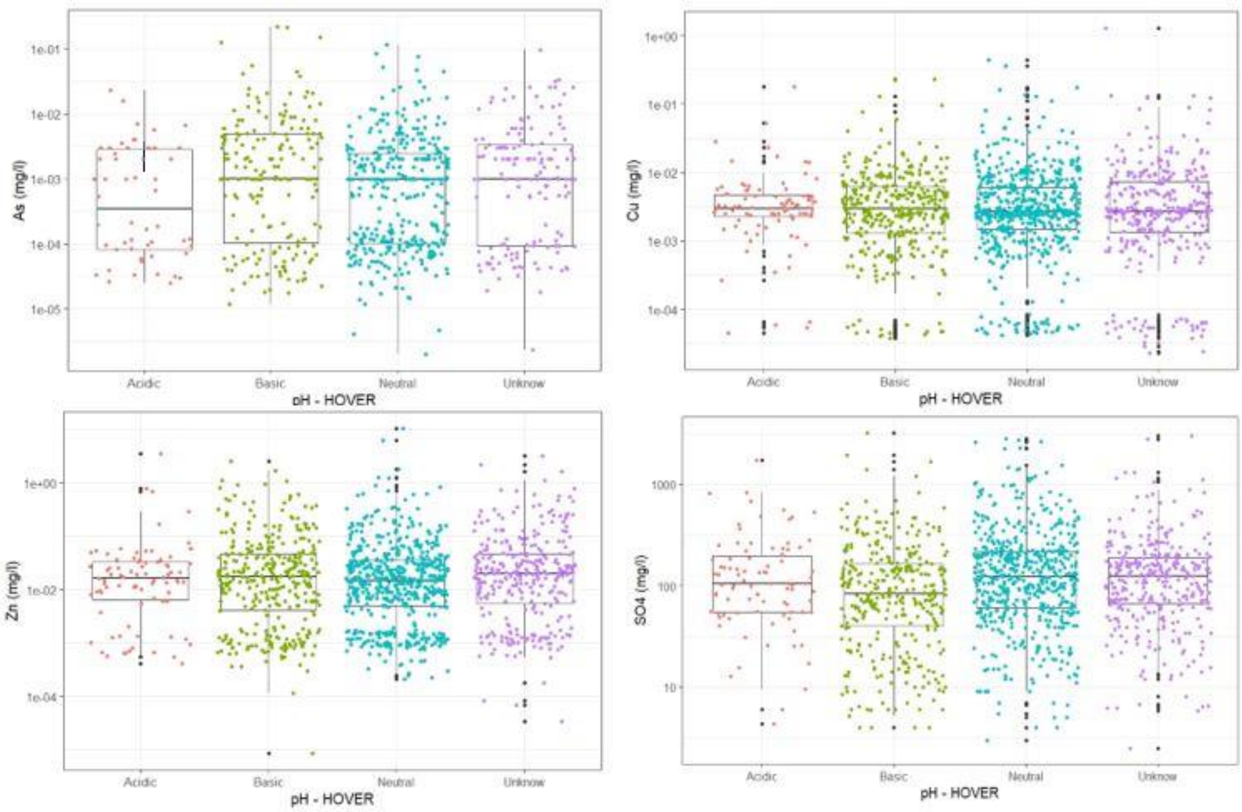


*Figure 88: Distribution of concentrations, for "pH-HOVER" categories. Y-axis in log scale. Dots: Median concentration at each sampling point. Box Plot: The thick line corresponds to the median and the lower and upper limits correspond to Q1 (percentile 25th) and Q3 (percentile 75th). The upper and lower whiskers extend to the largest and smallest value (1.5 * IQR from the box limits).*

11.2.6 *Removal of element concentrations influenced by contamination (Step 16)*

As explained, based on the method applied in the previous steps, it has been decided to remove the following groups of data from the calculation of the 90th percentile for the NBL for each element:

- As: The data of the group "mining influence" has been eliminated
- Cu: The data of the group "mining influence" has been eliminated
- Zn: No group has been deleted
- SO_4: All groups with anthropogenic pressures have been eliminated. NBL is only calculated from samples in the Natural state or not influenced.

Finally, to finish with the extraction of sampling points affected by anthropogenic contamination, a review of the remaining values has been carried out. Thus, taking into account that abnormally high values are detected, it has been decided to eliminate outliers that have not detected in the previous processes. In this sense and given the impossibility of carrying out a point-to-point analysis, due to lack of information in this regard, it has been decided to consider outliers the points with values above the 95th percentile. The objective is to obtain a less heterogeneous data population so that the final calculation of the 90th percentile is less biased by extreme values. Figure 89 show the accumulated distribution graphs of the elements studied.

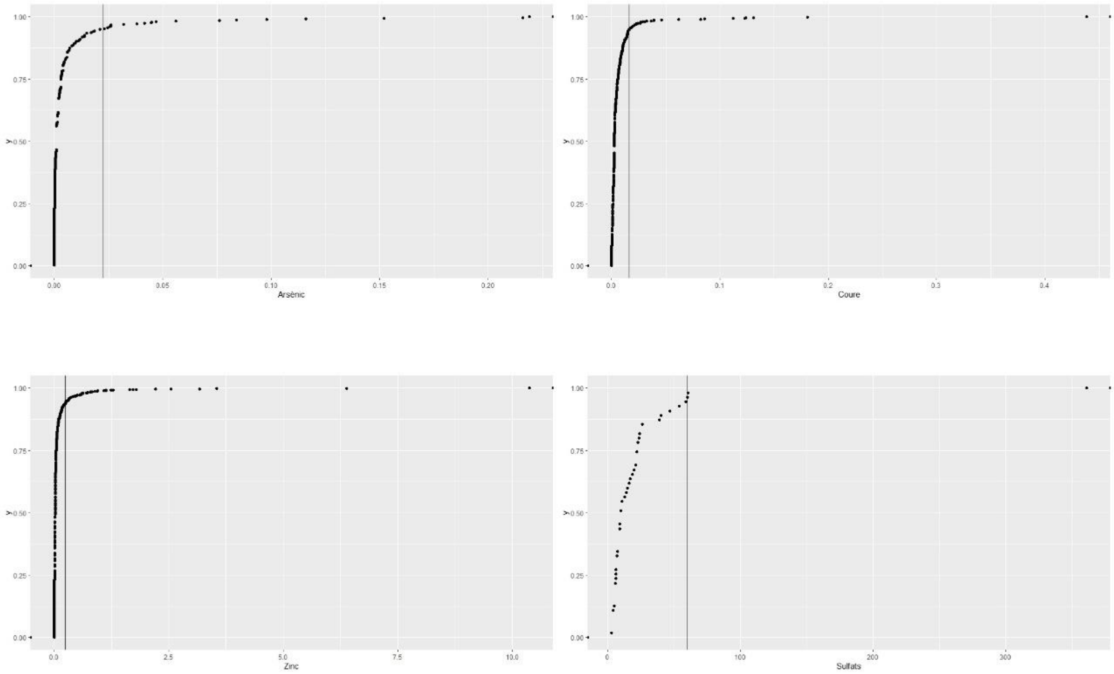


Figure 89: Cumulative distribution plots for As, Cu, Zn and SO4 concentrations. Vertical line: 90th percentile.

11.2.7 *Calculating NBL (Step 17)*

With the final dataset, the calculation was made for each element studied in each of the selected HOVER lithologies. The final calculation of the NBL has been performed with the 90th percentile. The values obtained are shown in Table 106.

As (µg/l)	**P-10**	**Q1**	**Q2**	**Q3**	**P90**	**n sampling points**
Sedimentary: carbonates (limestone, chalk, dolomite)	0.03	0.05	0.12	1.00	3.50	124
Alluvium	0.07	0.74	2.00	4.00	9.00	185
Crystalline bedrock	0.05	0.07	0.13	1.10	4.15	90
Cu (µg/l)	**P-10**	**Q1**	**Q2**	**Q3**	**P90**	**n sampling points**
Sedimentary: carbonates (limestone, chalk, dolomite)	0.07	0.70	1.75	2.98	8.00	229
Alluvium	0.36	1.37	2.57	4.80	9.89	260
Crystalline bedrock	0.62	2.00	3.04	6.35	10.84	237
Zn (µg/l)	**P-10**	**Q1**	**Q2**	**Q3**	**P90**	**n sampling points**
Sedimentary: carbonates (limestone, chalk, dolomite)	0.96	3.69	15.00	36.00	99.00	235
Alluvium	1.14	5.11	16.58	39.87	88.67	264
Crystalline bedrock	0.75	1.36	11.61	25.25	44.75	234
SO4 (mg/l)	**P-10**	**Q1**	**Q2**	**Q3**	**P90**	**n sampling points**
Sedimentary: carbonates (limestone, chalk, dolomite)	Range 6.23 - 60 mg/l					8
Alluvium	1 value: 47 mg/l					1
Crystalline bedrock	4.00	6.00	9.29	20.00	33.80	43

Table 106: NBL calculated with 90th percentile for As, Cu Zn, SO4 for HOVER lithologies classes selected.

The results show plausible values. For As concentration, the values are between 3.5 and 9 µg/l depending on the lithology.

For Cu concentration, the range varies between 8 µg/l and 10.84 µg/l showing very low variability from one lithology to another.

For Zn concentration, the variability between the values calculated for the lithology "Sedimentary: Carbonates" and "Alluvium is also very low (99 - 88.67) but it increases with respect to the values obtained for the" Crystalline rock "(44.75).

In the case of "SO_4", having eliminated numerous points potentially affected by anthropogenic pressures, the number of points under natural conditions on which to perform the calculation is too small. It has only been possible to establish the NBL for the group "Crystalline rock" with a value of 33.8 mg/l.

11.3 Comparison and contributions of the proposed method

It is difficult to establish a comparison with other studies or values because, as mentioned in the first section, in Catalonia no work has been done to determine the NBL for the whole internal basins of the territory. The Catalan Water Agency calculates NBL for each of the groundwater bodies separately. To perform the calculation, ACA uses the BRIDGE methodology, applied to all points of its control network in each period of time in each body of water. The period taken into account also makes a possible comparison difficult. Even so, it is possible to provide some data extracted from the report of the management plan of the district of the river basin of Catalonia (ACA, 2015) to compare them with those obtained.

The ACA values for As, in the water bodies of the internal basins (average calculated for 32 water bodies) is 10.9 µg/l. It is a value very similar than the obtained for the "Alluvium" lithology but higher than those obtained for the rest of HOVER-lithologies. Even so, regarding the values of each groundwater body, very high values are detected (e.g.196.8 µg/l in “Conca alta del Freser y Ter” basin). Even so, these values may be due to natural processes or even be modified by known mining activity in this basin.

In the case of SO_4, the average calculated for 35 bodies of water is 236.18 mg/l. In this case, the value is higher than that obtained with HOVER method, but it must be taken into account that the final number of points for the calculation of the NBL after applying the methodology, is too low. In this case the values of each body of water also reach very high values in some cases (e.g. 1033 mg/l in "Prades - Francolí" water body, or 633.2 mg/l for "Alluvial aquifer and other local aquifers"). Similarly, they may be due to natural processes, but they may be contaminations not eliminated from the calculation with the method used.

As conclusions, it can be said that the method works, but it has some limitations that can make it difficult to apply.

- **Work area, methodology and available data**

The selection of the work area has focused on obtaining an enough available data. In this sense, the main objective has been to test the method. Therefore, the values obtained should not be taken as characteristic of the area studied. These are plausible values, but they should not be given a hydrogeological sense with respect to the work area.

The method could fit better in areas with a greater hydrogeological sense. Zones with greater hydrogeological homogeneity instead of an area as hydrogeologically heterogeneous as the one worked.

It would be better, an application in which unit study was the aquifer. In this case, it would be easier to control lithological assignments, anthropogenic pressures and even, the role that the dynamics of the aquifer may play when considering or not some sampling points.

About the anthropogenic pressures and the influenced area of each sampling point, it is assumed that the area can be larger, smaller, with a different shape, etc. In addition, it is also assumed that the prevailing pressure in buffer 1 km, may not be the one that causes the greatest impact on groundwater. Furthermore, other very important variables such as the depth of the wells are not considered.

- **Statistical treatment**

As mentioned, in the statistical treatment section, there is a large number of points with similar hydrochemical characteristics, regardless of the lithology assigned. This fact could be mitigated with a more precise lithological assignment.

Regarding the analysis of anthropogenic pressures for each sampling point, all groups show a wide dispersion of values. This fact shows that eliminating the group showing the highest values does not guarantee that potentially contaminated sampling points will be eliminated.

12 SYNTHESIS & PERSPECTIVES

The approach adopted is a global approach by lithology, which makes it possible to extrapolate the results to aquifers with similar lithologies. The work provided here is therefore a first step making possible to define reference concentrations by lithology when no data are available. Generic lithology will produce generic NBLs that could be applied as guidelines for further studies but that cannot be applied locally without caution. The study thus makes it possible to define the reference concentrations expected by lithology by excluding any linked local natural anomaly, for example, to the presence of signs of mineralization (gypsum, evaporites, metals and metalloids) and of a mining deposit.

For each aquifer, this first estimation is open to the acquisition of new data, both qualitative and hydrogeological to confirm or precise the selected NBLs.

The approach used allows to discriminate, by lithology, the facies or geochemical context (pH, redox) for which higher NBLs is expected. For example, the distribution of one element between oxidized and reduced domain for elements sensitive to the redox potential (As).

The statistical tests applied are essentially based on the number of data available and the number of hydrogeological and lithological entities. NBLs could therefore be refined with the increasing number of data, the attachment of water points to lithology or GW entities and the precision of lithologies categories.

From the calculations performed in this study, it is possible to identify hydrogeological entities with anomalous concentrations when data is available. This can be done on the basis of the calculation of the median and percentiles for each entity. This exercise may for some entities be limited depending on the number of water sampling point available. These concentrations should therefore be taken with care for the entities described by little number of water points (<5). From these data and by comparing the concentrations measured with the concentrations calculated by lithology, it is possible to identify the entities for which additional investigations are necessary in order to better specify the NBLs and possible anthropogenic inputs.

The comparison tests are essentially based on the number of data available and the number of hydrogeological and lithological entities described by the dataset. These concentration ranges will therefore change with the increasing number of data, the knowledge of water point lithology or GW entities and the precision of lithologies categories by age, geological domain etc... This work therefore constitutes a first step for the definition of NBLs which will be strongly consolidated by regional studies and the acquisition of data dedicated to this purpose. This methodology will depend strongly on the selection of relevant water points and using sampling and analytical methods suitable for the characterization of NBLs.

Discriminant analyses were performed to discriminate HOVER lithologies (qualitative variables) of most prevailing HOVER water families present in the dataset by combinations of major, trace and minor elements (quantitative variables). However, results show that distinctions are linked to noticeable uncertainties and censored data. Uncertainty is generally of the order of 5 to 10% but for trace elements (<1 µg/l) it can rises from 10 to 35%.

Regarding statistical significance among various groups of lithology and anthropogenic influence, it is possible to conclude that prevailing pressures may impact trace elements concentration in groundwater. However trace element contamination, besides being dependent form the activities type and the properties of the elements, depends also of hydrogeological settings, with obvious contamination in alluvial aquifer while confined aquifer are naturally protected from surface infiltration. However, the application of statistical test to distinguish the influence of prevailing pressure on trace contamination helps to evaluate the dataset confidence for NBLs calculation by discarding data when necessary.

The contribution of land-use based method for assessing anthropogenic influences allow to identify statistical difference between water sampling points with prevailing agricultural pressure and those with industrial and/or urban pressures. In the Danish context, however, less than 1% of the water sampling points were without any anthropogenic pressure (i.e. natural) and around 80% are with prevailing agricultural pressure. This limits the application of the method. There are two consequences. First, the method does not distinguish between polluted and potentially polluted water sampling points. This is an especially important consequence for the Danish dataset, where the “agriculture” dominates the dataset and had to be used as the “normal state”. Second, because entire groups of data were excluded, irrespective of the actual observed concentrations, some of the NBLs increased after excluding the industrial and/or urban groups. By excluding substantial amount of water sampling points with low concentrations (potentially the natural background), thus the resulting NBLs increased.

From data provided by GW monitoring networks and by comparing the concentrations measured with the NBLs calculated by lithology, it is possible to identify the entities for which additional investigations are necessary in order to better specify the NBLs in case of natural anomalous concentration or anthropogenic inputs. The delineation of these entities and the interpretation of exceedances are not part of this study as they request local studies. Extreme values were not be investigated but they can potentially illustrate waters with a local high NBL due to specific geogenic features (mineralization, evaporites, thermal waters) that must be distinguished from the lithological NBL as defined in this study.

This work therefore constitutes a first step for the definition of NBLs which will be consolidated by regional studies and the acquisition of specific data dedicated to this purpose. On such scale, data-analytical methods might also be accompanied by process-oriented analyses and supported by expert background knowledge at the local scale. This methodology will depend strongly on the selection of relevant water points and using sampling and analytical methods suitable for the characterization of NBLs to reduce uncertainties. The 90th percentile makes it possible to estimate the possible natural concentrations in particular geological contexts that would have a natural origin. Values defined by higher percentiles, for example, 95 or 97.7% (Edmund and Shand, 2008) would be more appropriate as a reference for maximum natural concentrations but higher percentiles require larger data sets and refined water sampling points for the confidence intervals increase.

13 REFERENCES

ACA, 2013. Delimitació d'aqüífers de Catalunya escala 1:50.000.

ACA, 2015. Pla de Gestió del districte de la conca Fluvial de Catalunya. Període 2016 – 2021.

Agreste (2014). Enquête pratiques culturales (2011) – Principaux résultats. Les Dossiers Agreste No 21, juin 2014 : 70 p.

Antweiler R. C. (2015). Evaluation of Statistical Treatments of Left-Censored Environmental Data Using Coincident Uncensored Data Sets. II. Group Comparisons. Environ. Sci. Technol. 49(22): 13439-13446.

Appelo C.A.J., Postma D., 2005. Geochemistry, Groundwater, and Pollution, 2nd ed. A.A Balkema Publishers, Amsterdam (the Netherlands).

Ayotte J.D., Gronberg J.M., and Apodaca L.E., 2011. Trace elements and radon in groundwater across the United States, 1992–2003: U.S. Geological Survey Scientific Investigations Report 2011–5059, 115 p. (Also available at http://pubs.usgs.gov/sir/2011/5059.)

Ballesteros, B.J., Rodríguez, L., López, J. Grima, J, García, O., Cobos, J.R., Gómez, J.D. and De la Orden, J.A., 2001. Análisis y ordenación de recursos hídricos de la Marina Alta (Alicante). Alternativas y Directrices. Vol II. Evaluación de recursos subterráneos. Instituto Geológico y Minero de España-Diputación Provincial de Alicante. Fondo documental IGME. 179 pp. Alicante.

Bd Agreste (2010) - Les principaux résultats chiffrés au format tableur des recensements agricoles 2010, 2000 et 1988 par département, canton et commune. - Données en ligne - http://agreste.agriculture.gouv.fr/enquetes/structure-des-exploitations-964/recensement-agricole-2010/resultats-donnees-chiffrees/

Bell R.W., Williamson A.K., 2006. Data delivery and mapping over the Web-National Water-Quality Assessment Data Warehouse: U.S. Geological Survey Fact Sheet 2006–3101, 6 p.

Blum A., F. Wendland, R. Kunkel, M. Coestiers, M. Van Camp, K. Walraevens, R. Gorova, G. Berthold, G. Fritsche, R. Wolter, K. Hinsby, A. Marandi, Z. Simonffy, K. Kadunas, A. Arustiene, J. Griffioen, S. Witczac, J. Hookey, J. Gustafsson, 2006. Natural background levels. State of the art and review of existing methodology. In BRIDGE deliverable D10 "Impact of hydrogeological conditions on pollutant behaviour in groundwater and related ecosystems".

Blum, A., Pauwels, H., Wenland, F. and Grifioen, J. (2009). Background levels under the Water Framework Directive. Groundwater Monitoring (Water quality measurement series). John Wiley and Sons, Chipelham, Wiltshire, p.p. 145-153. ISBN: 978-0470-77809-8.

Brenot A., Gourcy L., Mascré C., et al. (2007) – Identification des zones à risque de fond géochimique élevé en éléments traces dans les cours d'eau et les eaux souterraines. Bassin Loire-Bretagne. BRGM/RP-55346-FR

Brielmann, H., Legerer, P., Schubert, G., Wemhöner, U., Philippitsch, R., Humer, F., Zieritz, I., Rosmann, T., Schartner, C., Scheidleder, A., Grath, J., Stadler, E. (2018): Hydrochemie und Hydrogeologie der österreichischen Grundwässer und deren natürliche Metall- und Nährstoffgehalte (Update Geohint 2018). Bundesministerium für Nachhaltigkeit und Tourismus, Wien.

Brugeron A., Paroissien J.B., Tillier L. (2018) - Référentiel hydrogéologique BDLISA version 2 : Principes de construction et évolutions. Rapport final BRGM/RP-67489-FR.

Buser, S. 2010: Geoloical map of Slovenia in scale 1: 250 000. GeoZS, Ljubljana.

Childress C.J.O., Foreman W.T., Connor B.F., Maloney T.J., 1999. New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey National Water Quality Laboratory: U.S. Geological Survey Open-File Report 99–193, 19 p.

Devau N., Lions J., Schomburgk S., Bertin C., Blanc, P., Mathurin F., Thinon-Larminach M., Lucassou F., Le Guern C., Tourlière B., Doney C., Salquebre D., Chretien P., Boisson A., Sedan O., Gourcy L., 2017. Etude par approche globale des fonds hydrogéochimiques des eaux souterraines sur le bassin Loire-Bretagne. BRGM/RP-67573-FR, 192 p. http://infoterre.brgm.fr/rapports//RP-67573-FR.pdf

DISAR (2010) - (Diffusion Interactive des Statistiques Agricoles de Référence) - Données en ligne - Available on internet on http://agreste.agriculture.gouv.fr/page-d-accueil/article/donnees-en-ligne.

DONEY.C., LIONS.J., DETANTE.B., DUYCK.G., 2015 Evaluation des concentrations naturelles et anthropiques en éléments traces dans les eaux et sols du Territoire de Belfort. Rapport. BRGM/RP-64747-FR.

Edmunds W.M., Shand P. (2008): Natural groundwater Quality. Blackwell Ed. 469 pp.

European Commission, 2009. Guidance on groundwater status and trend assessment (Guidance Document No. 18 of the Common Implementation Strategy for the Water Framework Directive (2000/60/EC), Technical Report. Office for Official Publications of the European Communities, Luxembourg.

European Commission, 2009. Guidance on groundwater status and trend assessment (Guidance Document No. 18 of the Common Implementation Strategy for the Water Framework Directive (2000/60/EC), Technical Report. Office for Official Publications of the European Communities, Luxembourg.

Fernandez-Ruiz, L. (ed.), 2012. Las Aguas Subterráneas en la Planificación Hidrogeológica. Madrid: Instituto Geológico y Minero de España. https://www.miteco.gob.es/es/agua/publicaciones/AGUAS%20SUBTERRANEAS%201de7_tcm30-215716.pdf

Ghestem JP, Crouzet C, Girardot N., 2008. Estimation des fonds géochimiques en éléments traces des eaux de surface continentales et des eaux souterraines : cahier des charges pour l'analyse et le prélèvement. Version 0, BRGM/RP-56835-FR, 38 p.

Gourcy L., Chery L., Lopez B., Auterives C., Brugeron A., Nicolas J., Sedan-Miegemolle O., Surdyk N. (2015) – Appui technique à la conception et la mise en œuvre de la politique de gestion et de protection des eaux souterraines. Rapport final. BRGM/RP-64572-FR. 15 p

Helsel D. R., Hirsch R. M., 2002. Statistical Methods in Water Resources (Chapter A3). Techniques of Water-Resources Investigations of the United States Geological Survey/Book 4, Hydrologic Analysis and Interpretation, U. S. Geological Survey (USGS).

Helsel D.R., 2005. Nondetects and data analysis: Statistics for censored environmental data (1st ed.): New York, John Wiley and Sons, 250 p.

Helsel D.R., Hirsch R.M., 1992. Statistical methods in water resources: New York, Elsevier Science Company, 522 p.

Helsel, D.R. and Hirsch, R.M., 2002. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation. Statistical Methods in Water Resources. Environmental science and pollution research international, 14(5):297-307.

Herms, I; Jódar, J.; Soler, A.; Lambán, L.J.; Custodio, E.; Agustí Núñez, J.; Arnó, G.; Ortego, M.I.; Parcerisa, D.; Jorge, J. (2020). Evaluation of natural background levels of high mountain karst aquifers in complex hydrogeological settings. A Gaussian mixture model approach in the Port del Comte (SE, Pyrenees) case study. November 2020Science of The Total Environment. DOI: 10.1016/j.scitotenv.2020.143864

Herms, I; Jódar, J.; Soler, A.; Vadillo, I; Lambán, L.J.; Martos-Rosillo, S.; Custodio, E.; Jorge, J.; Agustí Núñez, J. (2019). Baseline hydrogeochemical characterisation of a vulnerable pristine highmountain karst aquifer in the southeastern Pyrenees, 46th IAH Congress,22 - 27th September 2019. Málaga, ESP

Hinsby, K., Condesso de Melo, M.T., Dahl, M., 2008. European case studies supporting the derivation of natural background levels and groundwater threshold values for the protection of dependent ecosystems and human health. Science of The Total Environment 401, 1–20. https://doi.org/10.1016/j.scitotenv.2008.03.018

Hobiger, G., Klein, P., Denk, J., Grösel, K., Heger, H., Kohaut, St., Kollmann, W. F. H., Lampl, H., Lipiarsky, P., Pirkl, H., Schedl, A., Shadlau, S. & Winter, Ph. (2004): GEOHINT - Zl. 70.215/08-VII 1/03- Österreichweite Abschätzung von regionalisierten, hydrochemischen Hintergrundgehalten in oberflächennahen Grundwasserkörpern auf der Basis geochemischer und wasserchemischer Analysedaten zur Umsetzung der Wasserrahmenrichtlinie 2000/60/EG (Geogene Hintergrundgehalte oberflächennaher Grundwasserkörper). Endbericht. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft und Geologische Bundesanstalt, Wien 2004.

Kranjc, M., Gacin, M., Krsnik, P., Sodja, E., Kolenc, A. 2007: Groundwater quality in Slovenia assessed upon the results of national groundwater monitoring. European Water, 19/20: 37-46.

Law on National Parks (Official Gazette of RS, No. 84/2015, 95/2018).

Lions J.,Mauffert A., Devau N., 2016. Evaluation des concentrations de rérérence des fonds hydrogéochimiques des eaux souterraines par lithologie d'aquifères. Rapport BRGM/RP-65594-FR, 110 p.

LIONS.J., BLUM.A., COURBIN.A., JOSEPH.B. 2014. Caractérisation des fonds géochimiques en éléments traces des eaux douces de Guyane. Rapport. BRGM/RP-63670-FR.

Mezga, K. 2014: Natural hydrogeochemical background and dynamics of groundwater in Slovenia, PhD dissertation. University of Nova Gorica, Nova Gorica, 226 pp.

Molinero, Jorge & Custodio, E & Sahuquillo, A & Llamas, Manuel, 2011. Groundwater in Spain: Legal framework and management issues. Groundwater Management Practices. Balkema, Leiden, The Netherlands.

Müller et al. (2006). BRIIDGE. Background criteria for the Identification of Groundwater Thresholds. D18: FINAL PROPOSAL FOR A METHODOLOGY TO SET UP GROUNDWATER TRESHOLD VALUES IN EUROPE.

Novak, M., Bavec, M., Trajanova, M. 2014: Litološka karta Slovenije = Lithologica map of Slovenia. In: Novak, M., Rman, N. (eds.): Geološki atlas Slovenije = Geological atlas of Slovenia. Ljubljana : GeoZS.

Petrović Pantić, Veljković, Ž., Tomić, M., Samolov, K. 2017. Hydrogeology and vulnerability of groundwater to polluting in the area of national park „Fruška Gora“, 0350-0519, vol. 49 no. 288-290

Practical Stats, 2007. SAS macro for robust ROS (CROS): Practical Stats, accessed May 2009, at http://www.practicalstats.com/nada/nada/downloads.html.

Royal Decree 1514 / 2009, On 2 October, Which Regulates the Protection of Groundwater against Pollution and Deterioration.

SAS Institute, 1999. SAS/STAT online user's guide, Version 8: SAS Institute accessed October 17, 2006, at http://v8doc.sas.com/sashtml/.

Schubert, G., Lampl, H., Shadlau, S., Wurm, M., Pavlik, W., Pestal, G., Bayer, I., Freiler, M., Schild, A., Stöckl, W. (2003): Hydrogeologische Karte von Österreich 1 :500.000. Geologische Bundesanstalt, Wien.

Schuster, R., Daurer, A., Krenmayr, H.G., Linner, M., Mandl, G.W., Pestal, G. & Reitner, J.M. (2014): Rocky Austria. The Geology of Austria – Brief and Colourful. Geological Survey of Austria, Vienna.

Serianz, L., Cerar, S., Šraj, M. 2020: Hydrogeochemical characterization and determination of natural background levels (NBL) in groundwater within the main lithological units in Slovenia. Environmental Earth Sciences, 79/15: 373. doi: 10.1007/s12665-020-09112-1.

Shand P., Edmunds W. M., Lawrence A. R., Smedley P. L., Burke S., 2007. The natural (baseline) quality of groundwater in England and Wales. British Geological Survey and Envrionment Agency. Research Report No. RR/07/06.

Smedley P.L., Griffiths K., Tyler-Whittle R., Hargreaves R, Lawrence A R, Besien T., 2003. Baseline Report Series 5. The Chalk Aquifer of Kent and East Surrey. British Geological Survey and Environment Agency, Commissioned Report No. CR/03/033N.

SOGREAH 2007. Bilan des flux de contaminants entrant sur les sols agricoles de France métropolitaine - Bilan qualitatif de la contamination par les éléments tracés métalliques et les composés tracés organiques et application quantitative pour les éléments traces métalliques, Rapport final.

Stoiljković D.2003. Hidrogeološke odlike Fruške Gore. Letopis nauĉnih radova Poljoprivrednog fakulteta u Novom Sadu, 27 (1): 138-46

Thorling, L., Ditlefsen, C., Ernstsen, V., Hansen, B., Johnsen, A.R., Troldborg, L., 2019. Grundvandovervågning. Status og udvikling 1989 – 2018 (Teknisk rapport). GEUS, Copenhagen.

Thorling, L., Sørensen, B., 2014. Grundvandets kemiske tilstandsvurdering Vandområdeplan 2015-2021 - data og metodevalg: Kemisk tilstand af danske grundvandsforekomster, samlet rapportering (GEUS Rapport No. 2014/78). GEUS, Copenhagen.

Wendland F., Blum A., Coetsiers M., Gorova R., Griffioen J., Grima J., Hinsby K., Kunkel R., Marandi A., Melo T., Panagopoulos A., Pauwels H., Ruisi M., Traversa P., Vermooten J. S. A., Walraevens K., 2008. European aquifer typology: a practical framework for an overview of major groundwater composition at European scale. In Environmental Geology, v. 55: pp 77–85. https://doi.org/10.1007/s00254-007-0966-5

14 Appendices

14.1 NBLs per country

14.1.1 *Austria*

	Country or Domain	GW threshold values	Unit	Sedimentary Gravel		Sedimentary Clays/marls		Sedimentary Other		Metamorphic rocks	
				n	NBL	n	NBL	n	NBL	n	NBL
As	Austria - groundwater monitoring network	1	µg/l	1168	1.05	159	0.12	198	0.5	53	3.38
Cd	Austria - groundwater monitoring network	0.1	µg/l	135	0.05			130	0.05	35	0.04
Cr	Austria - groundwater monitoring network	1	µg/l	673	0.6	104	2.37	173	0.5	50	0.5
Cu	Austria - groundwater monitoring network	5	µg/l	126	2.15			130	0.5	34	0.5
Ni	Austria - groundwater monitoring network	1	µg/l	135	0.5			130	0.5	35	0.5

14.1.1 *Catalonia*

	Area	LOQ	Unit	Sedimentary Carbonates		Crystallines bedrocks		Other Lithologies Quaternary Alluvium	
				n	NBL	n	NBL	n	NBL
As	ICGC - Catalonia - Internal basins	0.01 - 1	µg/l	124	3.5	90	4.15	185	9
Cu	ICGC - Catalonia - Internal basins	0.5 - 10	µg/l	229	8	237	10.84	260	9.89
Zn	ICGC - Catalonia - Internal basins	0.1 - 5	µg/l	235	99	234	44.75	264	88.67
SO4	ICGC - Catalonia - Internal basins	8.0-10.0	mg/	8	6.23 - 60	43	33.8	-	-

14.1.1 *Denmark*

	Country or Domain	GW threshold values	Unit	LOQ	HOVER lithology Sedimentary Sand		Sedimentary Carbonates	
As	Denmark	5	µg/l	0.01-1.0	3639	6.8	1830	3.9
Cd	Denmark	0.5	µg/l	0.003-0.5	259	0.011	77	0.026
Cr	Denmark	25	µg/l	0.01-0.5	186	0.195	55	0.178
Cu	Denmark	100	µg/l	0.03-5.0	218	1	67	1.3
Ni	Denmark	10	µg/l	0.02-3.0	3661	1.8	1856	5.3
Zn	Denmark	100	µg/l	0.3-10.0	271	8.1	85	20
F	Denmark	1.5 [c]	mg/	0.02-0.5	4046	0.4	2213	1.2
Cl	Denmark	250	mg/	-	4072	70	2242	130
SO4	Denmark	250 [c]	mg/	0.2-1.5	3642	87	1834	89

HOVER-lithology	Elements	units	HOVER-pH	HOVER-redox				NBLs[a] (only lithology)	National threshold values[b]
				A	B	C & D	X		
Sedimentary: carbonates	As	µg/l	Basic	2.2	-	4	3.3	3.9	5
			Neutral	2	2.7	4.7	7.7		
	Ni	µg/l	Basic	4	-	2.1	3.2	5.3	10
			Neutral	12	20	3	7.7		
	SO4	mg/l	Basic	43	-	63	48	89	250[c]
			Neutral	92	100	89	82		
	F	mg/l	Basic	0.3	-	0.9	1.3	1.2	1.5[c]
			Neutral	0.4	0.8	1.2	1.7		
	Cl	mg/l	Basic	55	-	64	96	130	250[c]
			Neutral	82	113	130	203		
Sedimentary: sand	As	µg/l	Acidic	-	-	3	-	6.8	5
			Basic	2.3	-	5.6	3.6		
			Neutral	1.3	-	9.6	3.8		
	Ni	µg/l	Acidic	-	-	4.7	-	1.8	10
			Basic	1.3	-	1	1.4		
			Neutral	2.2	-	1.3	3.4		
	SO4	mg/l	Acidic	-	-	87	-	87	250[c]
			Basic	68	-	74	86		
			Neutral	86	-	89	111		
	F	mg/l	Acidic	-	-	0.7	-	0.4	1.5[c]
			Basic	0.2	0.2	0.3	0.2		
			Neutral	0.3	-	0.4	0.3		
	Cl	mg/l	Acidic	-	-	46	-	70	250[c]
			Basic	60	69	58	60		
			Neutral	67	-	78	67		

14.1.1 *France*

	Country	GW threshold values		HOVER lithology							
				Sedimentary		Sedimentary		Sedimentary		Sedimentary	
				Paris Basin		Paris Basin		Paris Basin		Paris Basin	
				Sand		Gravel		Carbonates		Clays/marls	
				n	NBL	n	NBL	n	NBL	n	NBL
As	France	10	µg/l	343	1.46	33	4.96	554	3	48	4
Cd	France	5	µg/l	221	<0.5	723	<0.5	723	<0.5	77	<0.5
Cr	France	50	µg/l	5	<0.45	<5	_	17	<0.4	<5	_
Cu	France	2000	µg/l	39	4.4	9	1.4	131	3.1	15	2.8
Ni	France	20	µg/l	231	2.3	27	7.8	375	2.8	44	0.9
Zn	France	5000	µg/l	38	12.3	10	65.1	133	24.1	15	26.7
F	France	1.5	mg/	388	0.9	51	0.2	673	0.3	75	0.3
Cl	France	200-250	mg/	394	53	52	27.9	686	38.8	77	23.1
SO4	France	250	mg/	394	69.9	52	56.2	686	44	77	35.6

	Country	GW threshold values		HOVER lithology					
				Volcanic rocks		Crystallines bedrocks		Metamorphic rocks	
				Massif Central					
				n	NBL	n	NBL	n	NBL
As	France	10		373	2.4	336	3.6	1321	6.7
Cd	France	5	µg/l	378	<0.5	380	<0.5	1437	<0.5
Cr	France	50	µg/l	21	1	31	<0.5	123	<0.5
Cu	France	2000	µg/l	13	<1	25	9.6	95	11
Ni	France	20	µg/l	139	0.3	211	5.1	741	5
Zn	France	5000	µg/l	19	<5	40	23.8	139	40.4
F	France	1.5	mg/	310	0.2	318	0.1	1146	0.1
Cl	France	200-250	mg/	392	11.4	383	27	1496	28.1
SO4	France	250	mg/	392	7.3	383	24	1469	18

HOVER-lithology	Elements	units	HOVER-pH	HOVER-redox			NBLs (only lithology)
				A & B	C&D		
sand	As	µg/l		1.06	1.5		1.46
	Ni	µg/l	Acidic	10.2	13.5		2.3
			Basic	1.3	2.6		
			Neutral	2.1	2.3		
	Zn	µg/l	Acidic			16.6	12.3
			Basic			11.8	
			Neutral			10	
	Cu	µg/l	Acidic	8			4.4
			Basic		1.3		
			Neutral	8.8	3.5		
carbonates	As	µg/l		1.2	6.4		3
	Ni	µg/l	Acidic				2.8
			Basic	3.5	0.6		
			Neutral	2.2	5.5		
	Cu	µg/l	Basic	2.2	0.3		
			Neutral	3.2	0.6		
volcanics	As	µg/l		2.44	1.4		2.36

14.1.2 *Serbia*

	Country or Domain	GW threshold values	Unit	LOQ	HOVER lithology					
					Sedimentary Sand		Sedimentary Carbonates		Sedimentary Other	
					n	NBL	n	NBL	n	NBL
As	Serbia	10	µg/l	2.5	24	28.8	1	1	5	2,5 (90th percentiles are identical to the half of max LOQ)
Cd	Serbia	3	µg/l	1	24	1 (90th percentiles are identical to the half of max LOQ)	1	1 (90th percentiles are identical to the half of max LOQ)	5	1 (90th percentiles are identical to the half of max LOQ)
Cr	Serbia	50	µg/l	0.2	13	20.3	NOT EVALUATED < LOQ		NOT EVALUATED < LOQ	
Cu	Serbia	2000	µg/l	1.5	15	4.6	3	1.3	NOT EVALUATED < LOQ	
Ni	Serbia	20	µg/l	1	14	6.3	NOT EVALUATED < LOQ		NOT EVALUATED < LOQ	
Zn	Serbia	3000	µg/l	5	19	40.8	3	1.1	4	266.3
F	Serbia	1.2	mg/	0.05	24	0.3	3	0.9	5	0.3
Cl	Serbia	200	mg/		24	53.4	3	26.1	5	38.4
SO4	Serbia		mg/		24	35.8	3	28.1	5	61.6

14.1.3 *Slovenia*

	Area	GW threshold values	Unit	LOQ	HOVER lithology					
					Sedimentary Sand		Sedimentary Gravel		Sedimentary Carbonates	
				LOQ/2 $_{max}$	n	NBL	n	NBL	n	NBL
As	Slovenia	10	µg/l	0.5	NOT EVALUATED = 90th percentiles are identical to the half of max LOQ					
Cd	Slovenia	5	µg/l	0.01	9	0.014	4	0.011	51	0.01
Cr	Slovenia	50	µg/l	0.5	9	2.4	4	2.9	51	0.5
Cu	Slovenia	2000	µg/l	0.5	9	1.3	4	0.5	51	0.64
Ni	Slovenia	20	µg/l	0.5	9	3.3	4	1.4	51	0.5
Zn	Slovenia	not defined	µg/l	5	9	380	4	5	51	4.5
F	Slovenia	1.5	mg/	0.1	NOT EVALUATED = 90th percentiles are identical to the half of max LOQ					
Cl	Slovenia	250	mg/		9	11	4	9.1	47	5.8
SO4	Slovenia	250	mg/		9	15.5	4	13	47	8.2

14.1.4 *Spain*

						Other Lithologies	
	Area	unit	LOQ	Metamorphic rocks		Cenozoic	
				n	NBL	n	NBL
As	**Duero River Basin - Spain**	µg/l	0.1-5	16	3.25	202	8.11
Cd	**Duero River Basin - Spain**	µg/l	0.02-1	16	0.046	203	0.025
Cr	**Duero River Basin - Spain**	µg/l	1	-	-	-	-
Cu	**Duero River Basin - Spain**	µg/l	0.001-0.006	16	0.0046	203	0.0027
Ni	**Duero River Basin - Spain**	µg/l	1.0-7.0	-	-		
Zn	**Duero River Basin - Spain**	µg/l	0.002-0.06	-	-	-	-
F	**Duero River Basin - Spain**	mg/	0.05-0.5	16	0.27	197	0.95
Cl	**Duero River Basin - Spain**	mg/	1.0-10.0	16	28.91	203	62.58
SO4	**Duero River Basin - Spain**	mg/	1.0-10.0	16	30.86	203	100.3

14.2 Appendix – Case study - Denmark

14.2.1 *LDA models*

Table 107 Number and percentage of data-points in each of the HOVER lithology categories

HOVER lithology	model 1		model 2		model 4	
	n	%	n	%	n	%
Sedimentary: carbonates	2230	35.4	2232	35.4	2017	35.7
Sedimentary: sand	4062	64.6	4068	64.6	3635	64.3

Table 108 Number and percentage of data-points in each of the DK geology categories

DK geology	model 3		model 5		Model 6	
	n	%	n	%	n	%
Chalk/Limestone (“kalk”)	2230	35.2	2017	35.5	2017	35.9
Quaternary sand (“ks”)	3359	53.0	2989	52.5	2989	53.2
Pre-quaternary sand (“ps”)	664	10.5	614	10.8	614	10.9
Bornholm units (“uu”)	80	1.3	69	1.2	-	-

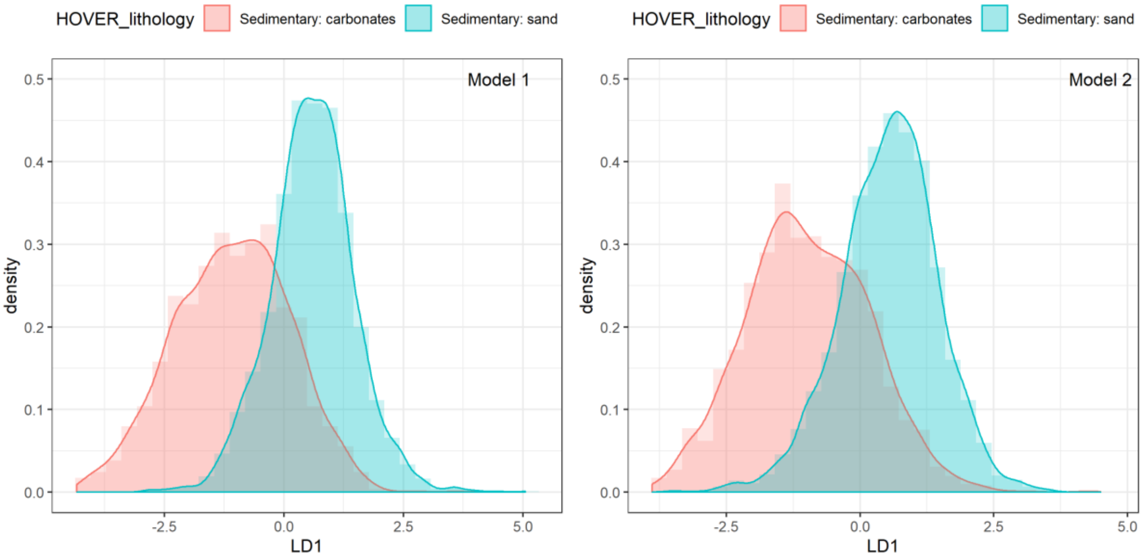


Figure 90 LD function (LD1) for model 1 and 2 (predicting HOVER-lithology)

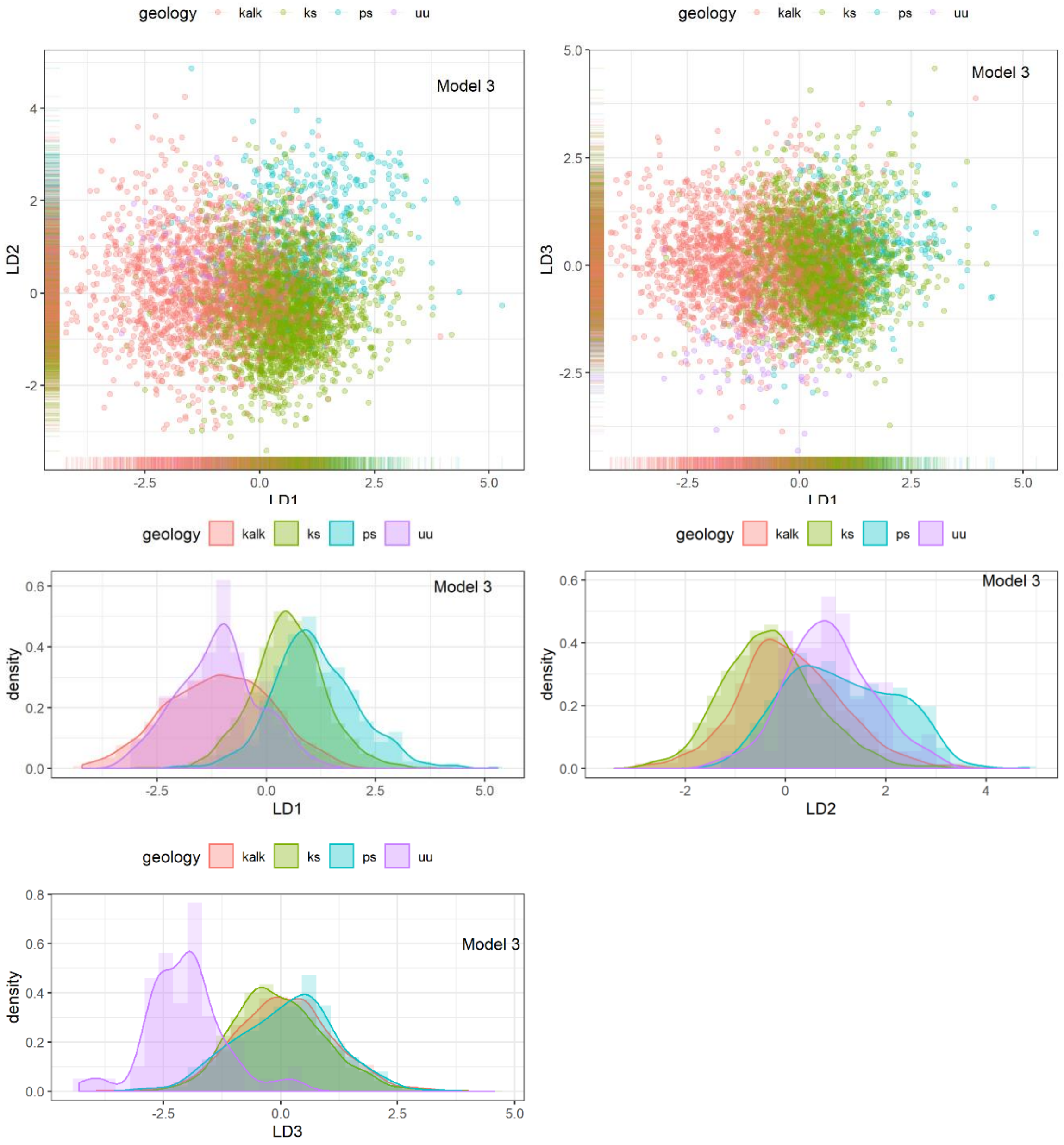


Figure 91 LD functions for model 3 (predicting DK-geology); Percentage separation achieved by each discriminant function: LD1: 73.7%, LD2: 20.2%, LD3: 6.1

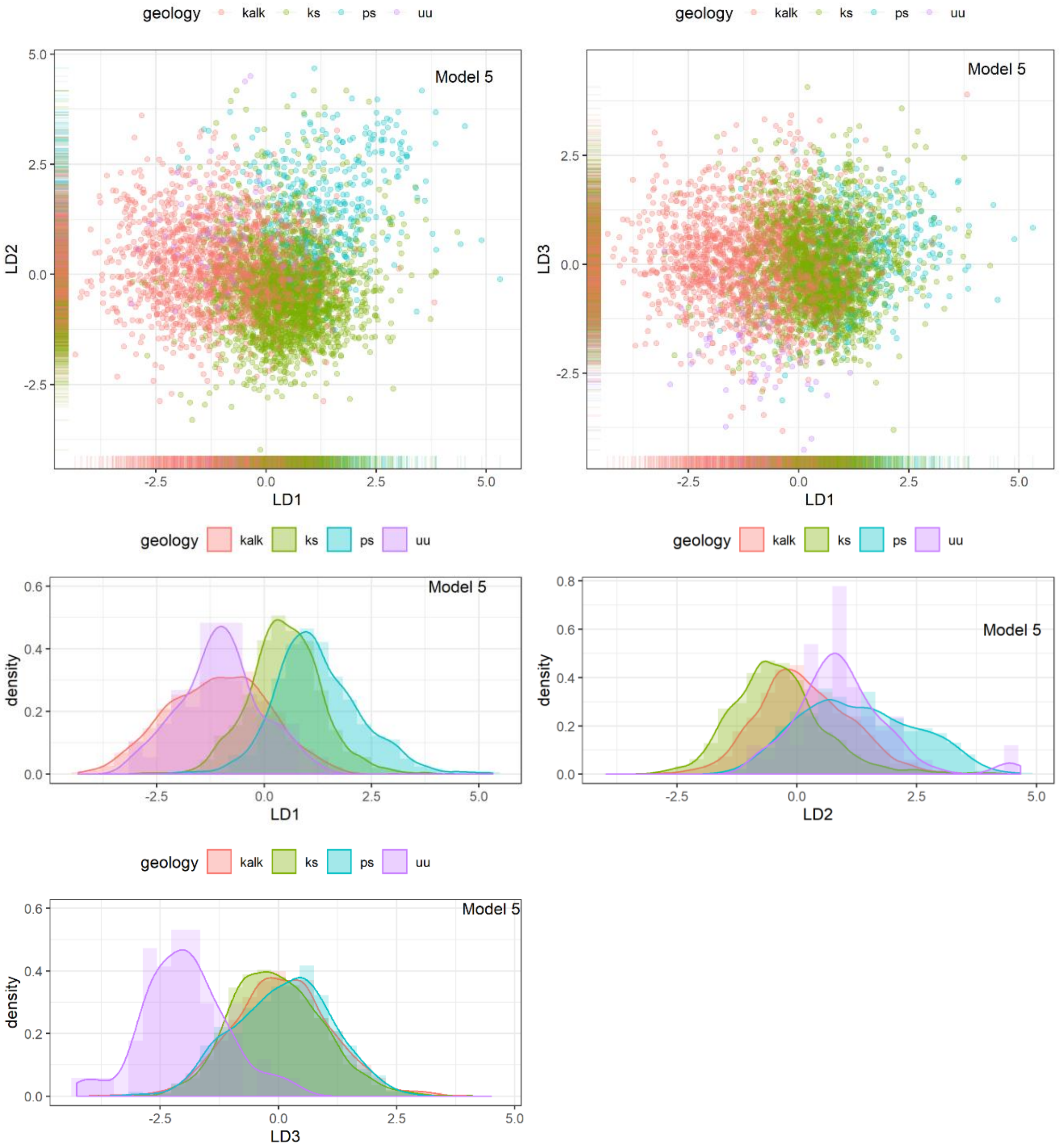


Figure 92 LD functions for model 5 (predicting DK-geology); Percentage separation achieved by each discriminant function: LD1: 68.3%, LD2: 26.7%, LD3: 5.0%

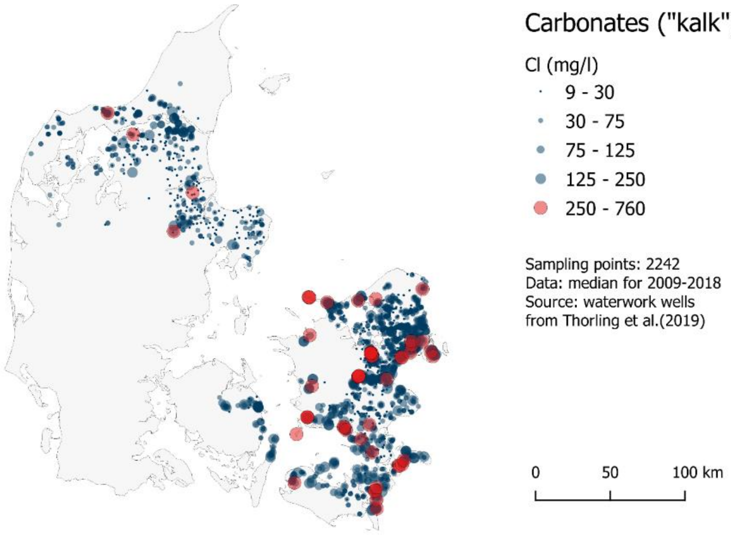
14.2.2 *Descriptive statistics*

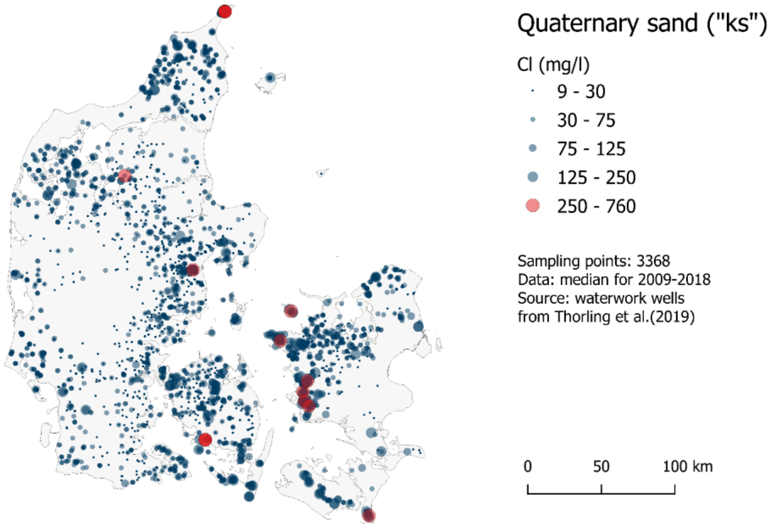
Table 109 Descriptive statistics for the data sub-sets for each element, used for the NBL calculation (after removal of anthropogenic pressures and outliers)

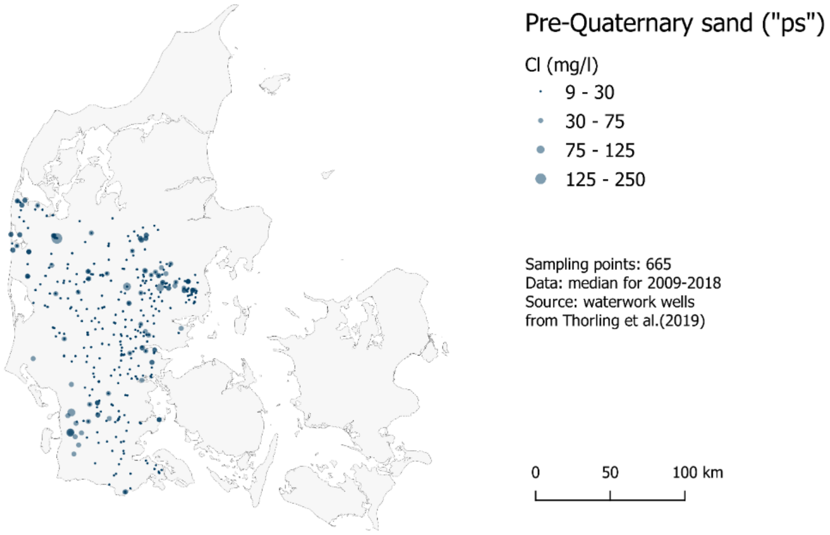
	unit	min	Q10	Q25	median	Q75	Q90	Q95	max	IQR	MAD	mean	SD	n
As	µg/l	0.018	0.08	0.3	0.9	2.4	5.9	9.7	44	2.1	0.7	2.3	3.9	5508
Cd	µg/l	0.003	0.005	0.005	0.006	0.006	0.013	0.025	0.225	0.001	0.001	0.010	0.020	337
Cr	µg/l	0.016	0.045	0.060	0.060	0.120	0.190	0.310	0.730	0.060	0.015	0.109	0.112	241
Cu	µg/l	0.039	0.05	0.1	0.2	0.5	1.1	1.7	9.4	0.4	0.1	0.5	1.0	285
Ni	µg/l	0.03	0.05	0.1	0.4	1.1	2.8	5.5	110.05	1.0	0.3	1.3	3.5	5558
Zn	µg/l	0.3	0.7	0.8	1.5	4.3	11	19	64	3.6	0.8	4.6	8.8	359
Cl	mg/l	9	20	25	34	52	89	133	760	28	12	49	49	6355
F	mg/l	0.03	0.13	0.16	0.25	0.42	0.76	1.1	6.6	0.26	0.10	0.37	0.37	6298
SO4	mg/l	0.23	5	15	34	63	87	102	285	48	22	42	35	5515

14.2.3 *Spatial distribution of elemental concentrations in Danish GW (only waterworks well) after excluding water sampling points with element-specific anthropogenic pressures*

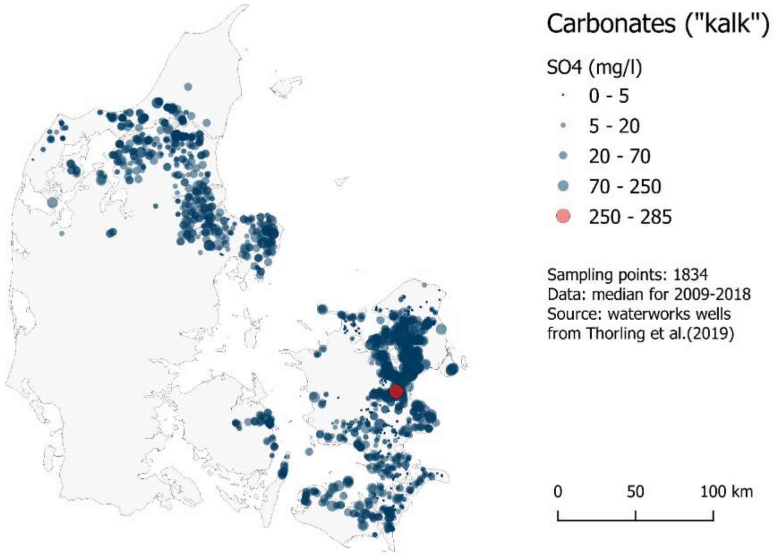
Chloride (Cl^-)



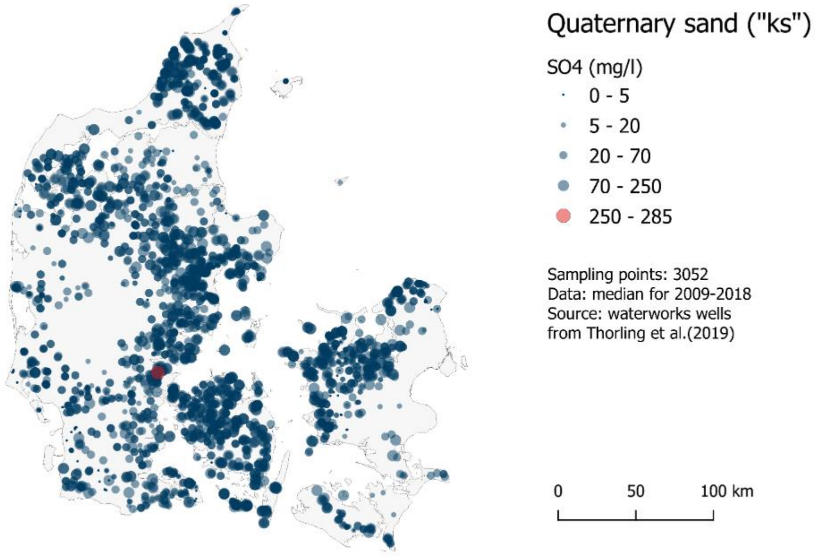




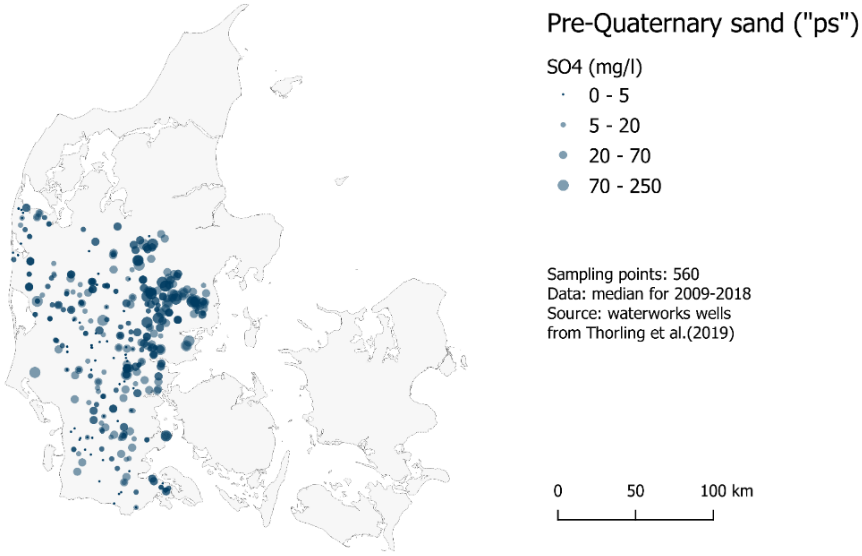
Sulphate (SO_4^-)


Carbonates ("kalk")

Sampling points: 1834
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

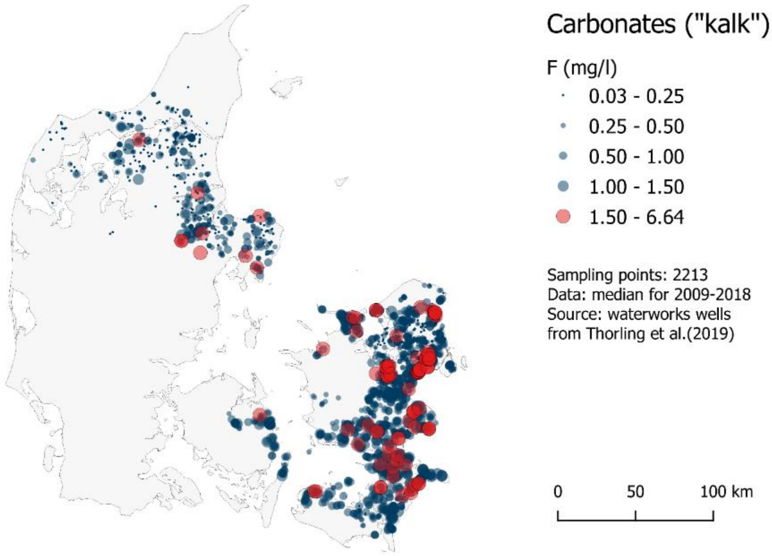

Quaternary sand ("ks")

Sampling points: 3052
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

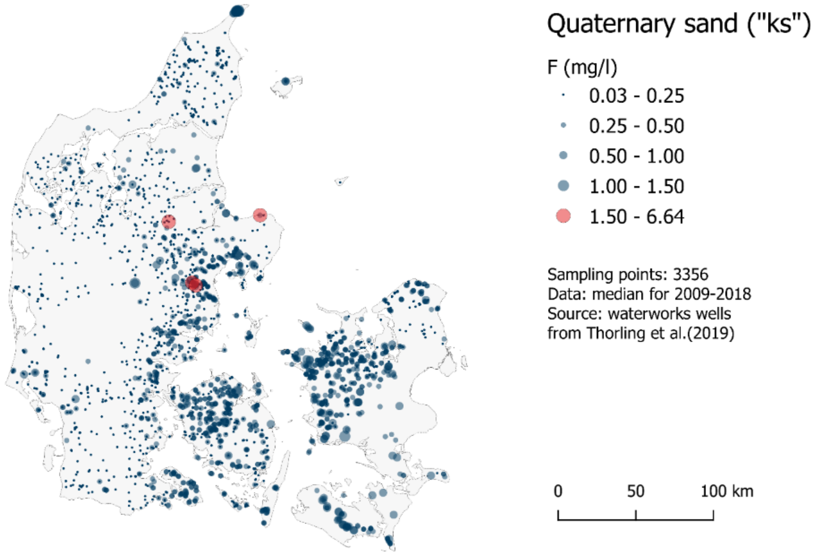

Pre-Quaternary sand ("ps")

Sampling points: 560
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

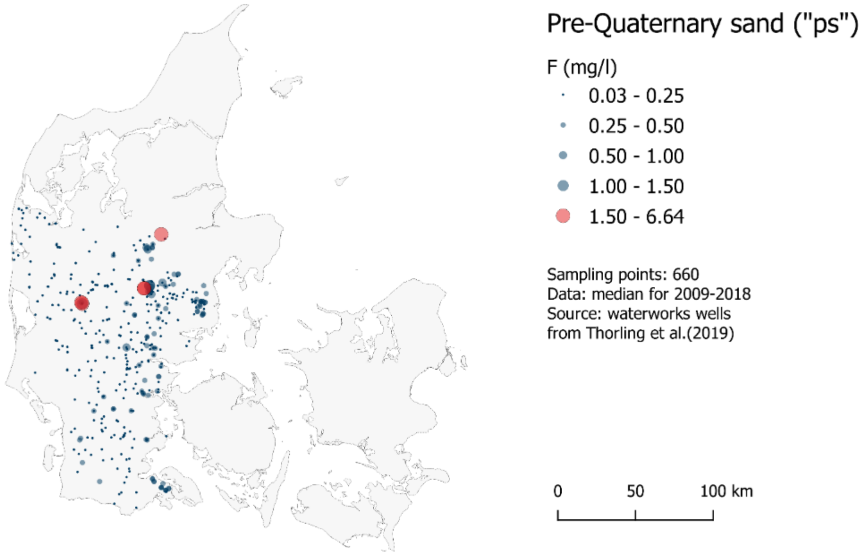
Fluoride (F$^-$)


Carbonates ("kalk")

Sampling points: 2213
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

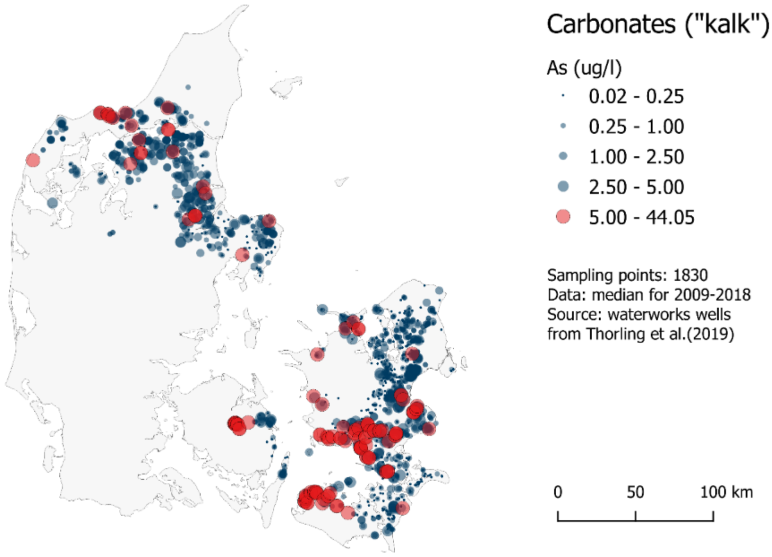

Quaternary sand ("ks")

Sampling points: 3356
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

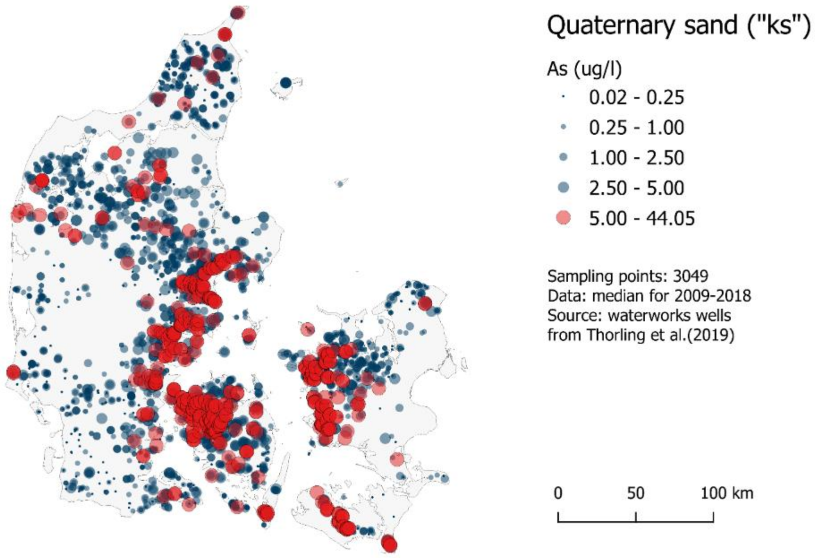

Pre-Quaternary sand ("ps")

Sampling points: 660
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

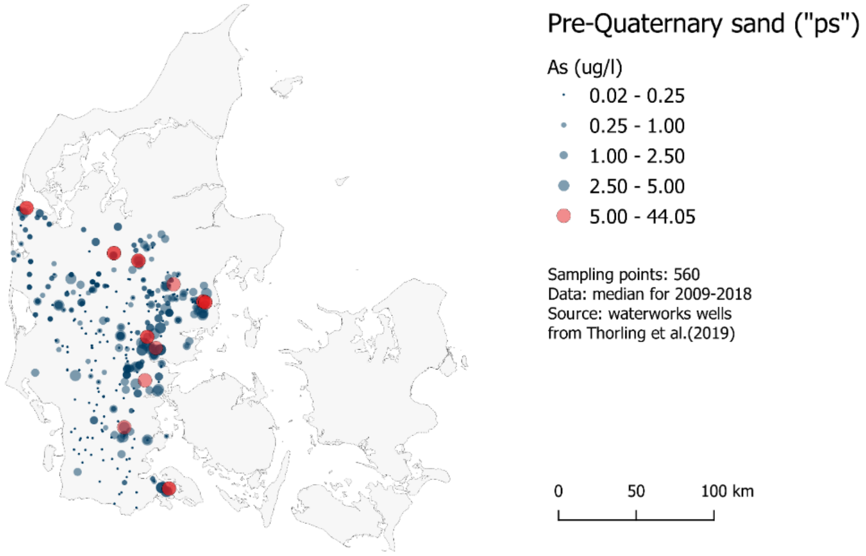
Arsenic (As)


Carbonates ("kalk")

Sampling points: 1830
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

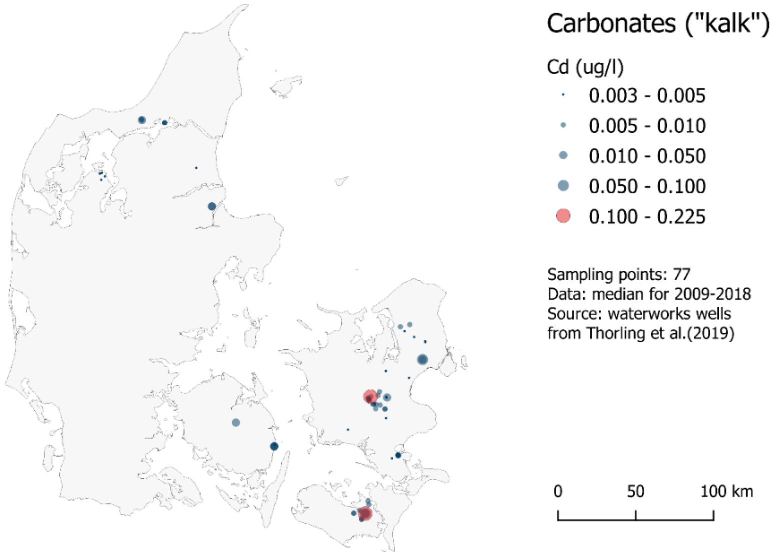

Quaternary sand ("ks")

Sampling points: 3049
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

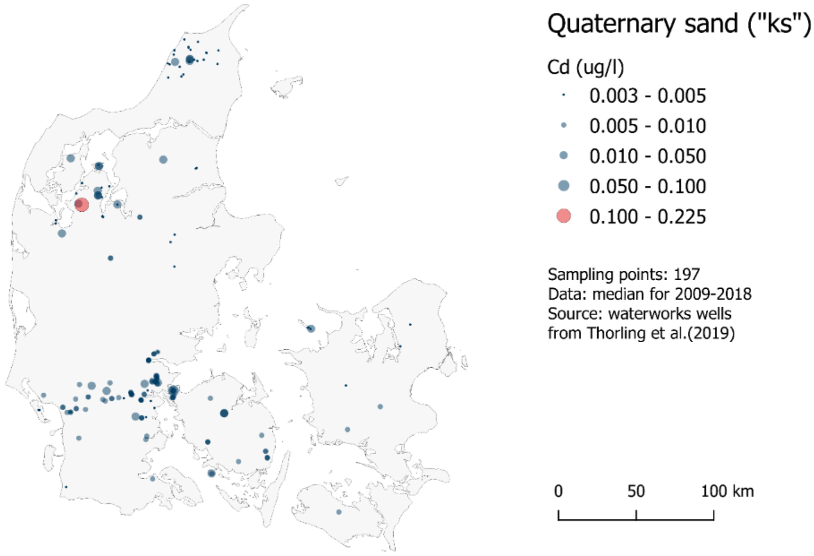

Pre-Quaternary sand ("ps")

Sampling points: 560
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

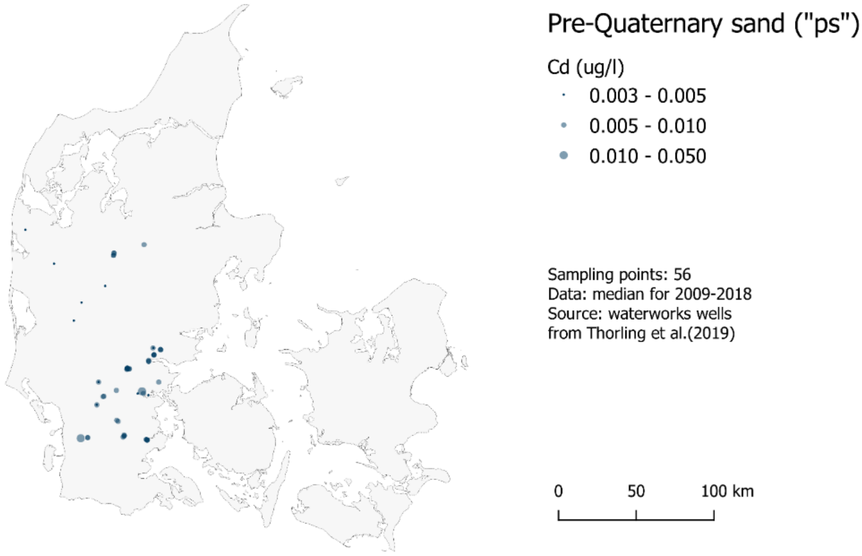
Cadmium (Cd)



Sampling points: 77
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

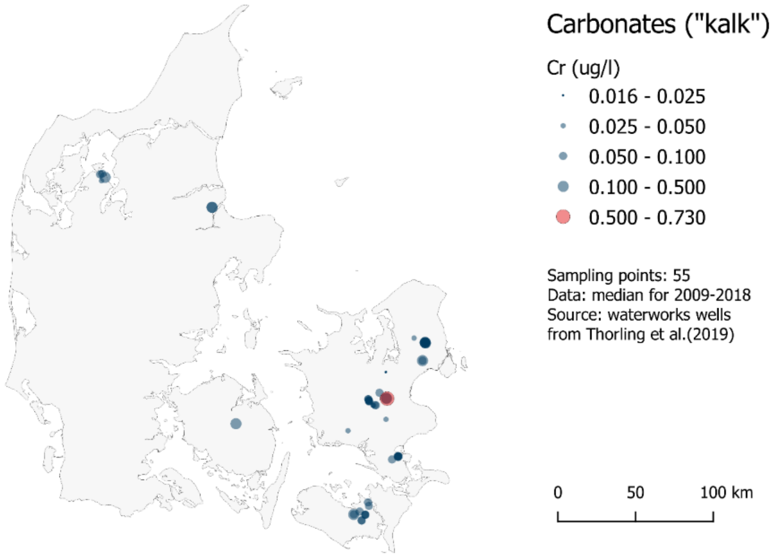


Sampling points: 197
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

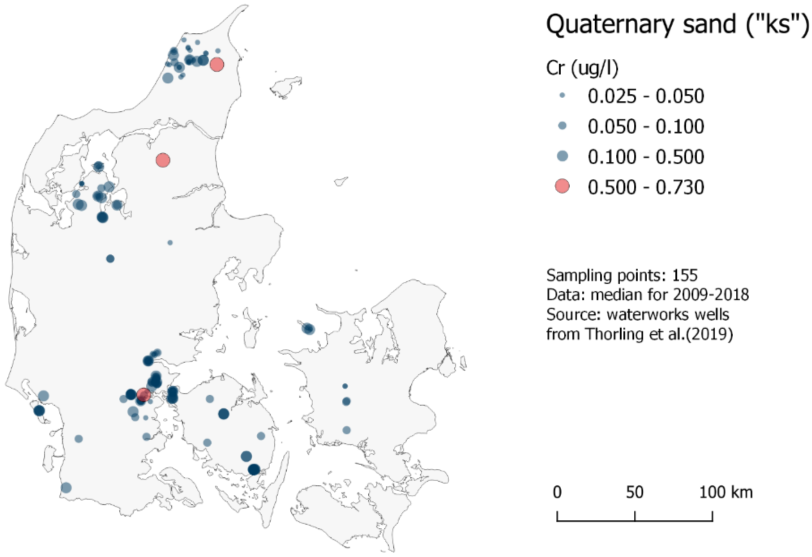


Sampling points: 56
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

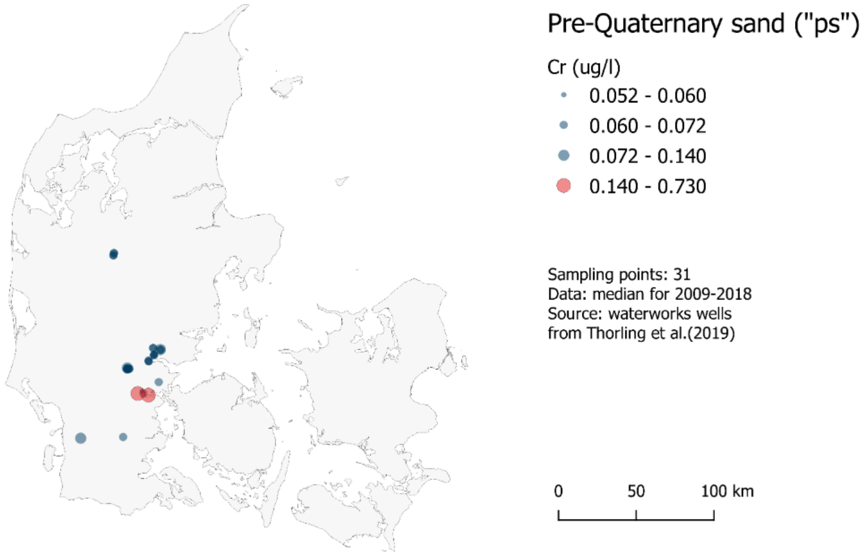
Chromium (Cr)


Carbonates ("kalk")

Sampling points: 55
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

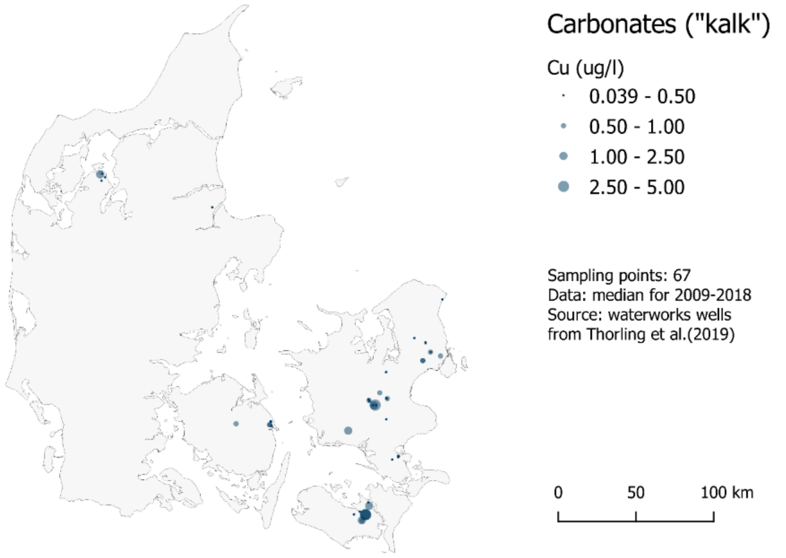

Quaternary sand ("ks")

Sampling points: 155
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)


Pre-Quaternary sand ("ps")

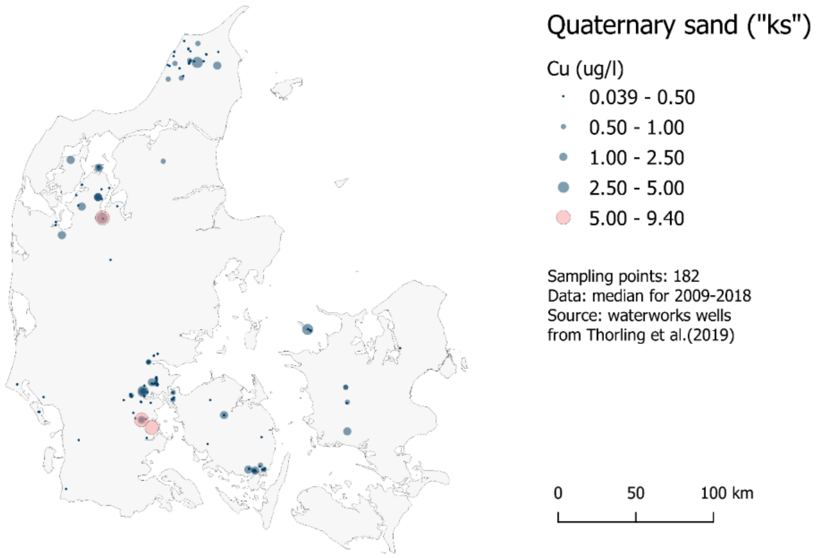
Sampling points: 31
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

Copper (Cu)



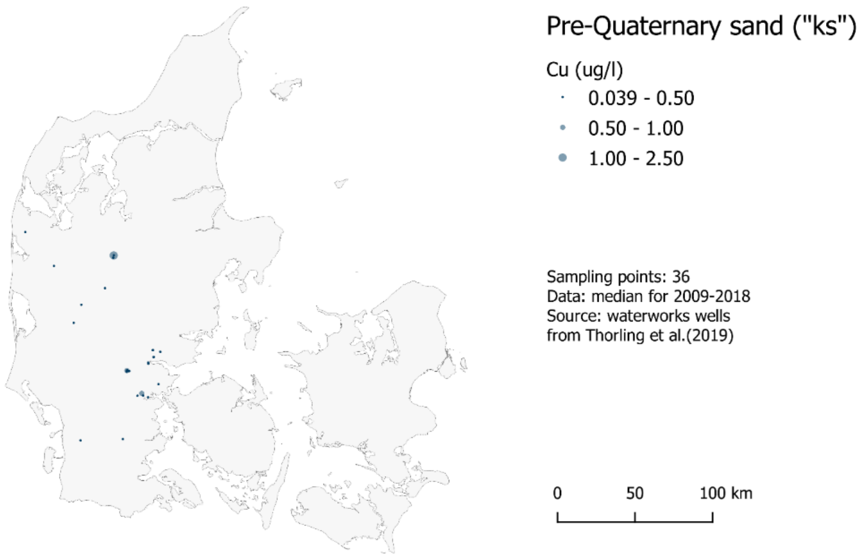
Carbonates ("kalk")

Sampling points: 67
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)



Quaternary sand ("ks")

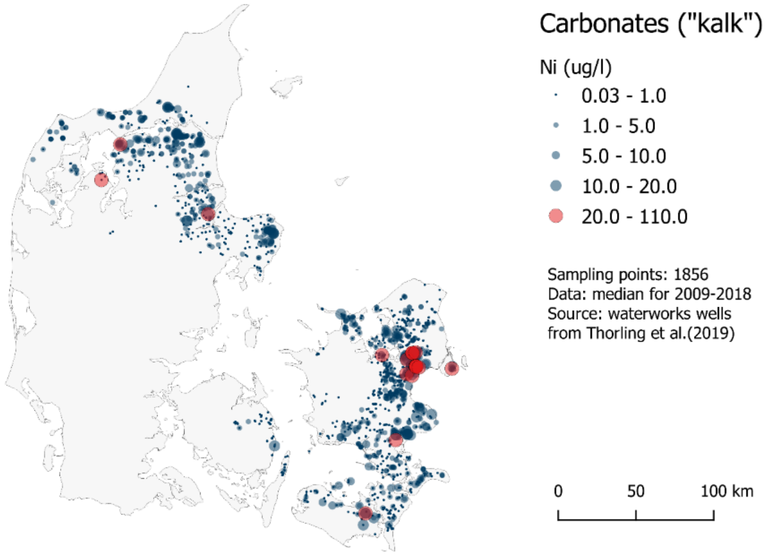
Sampling points: 182
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

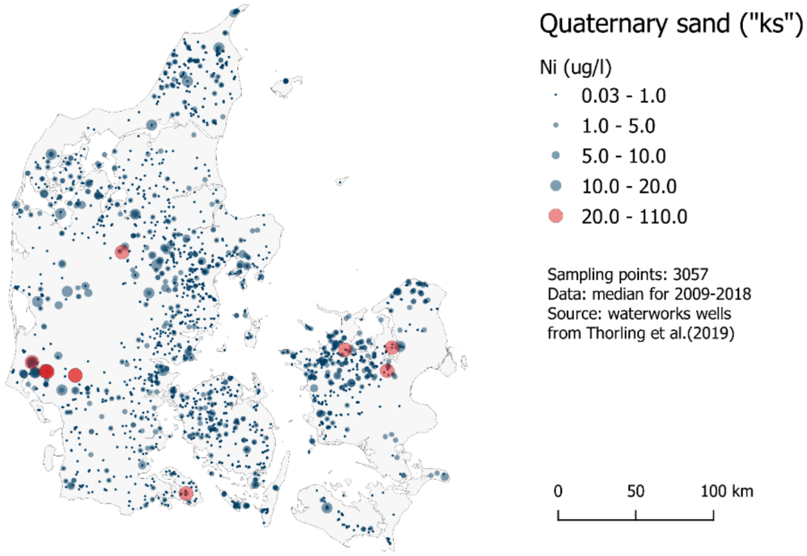


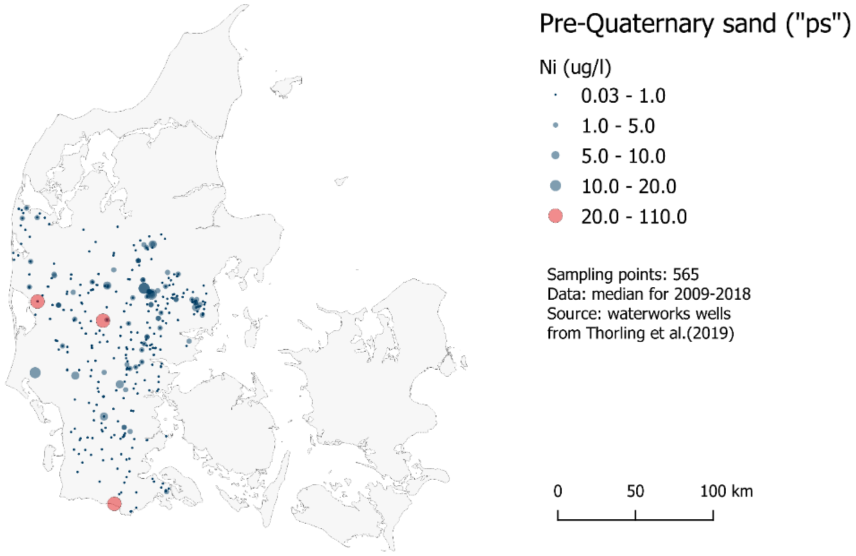
Pre-Quaternary sand ("ks")

Sampling points: 36
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

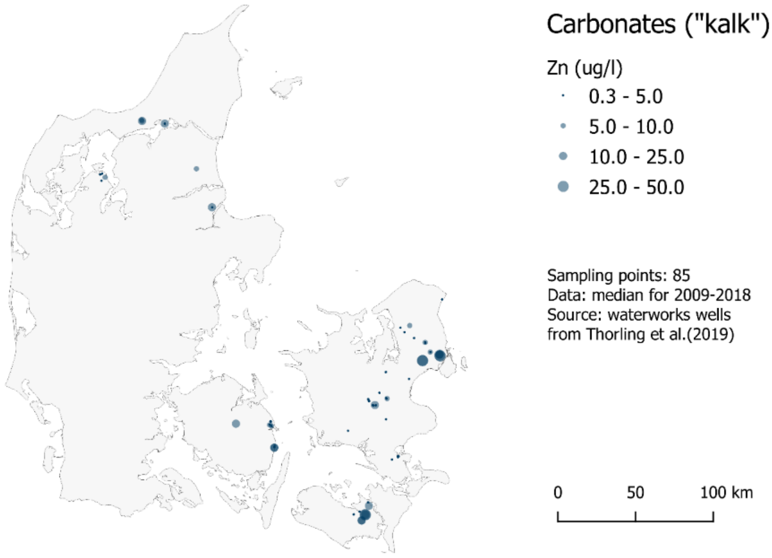
Nickel (Ni)


Carbonates ("kalk")
Ni (ug/l)
· 0.03 - 1.0
• 1.0 - 5.0
● 5.0 - 10.0
● 10.0 - 20.0
● 20.0 - 110.0
Sampling points: 1856
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)
0
50
100 km


Quaternary sand ("ks")
Ni (ug/l)
· 0.03 - 1.0
• 1.0 - 5.0
● 5.0 - 10.0
● 10.0 - 20.0
● 20.0 - 110.0
Sampling points: 3057
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)
0
50
100 km

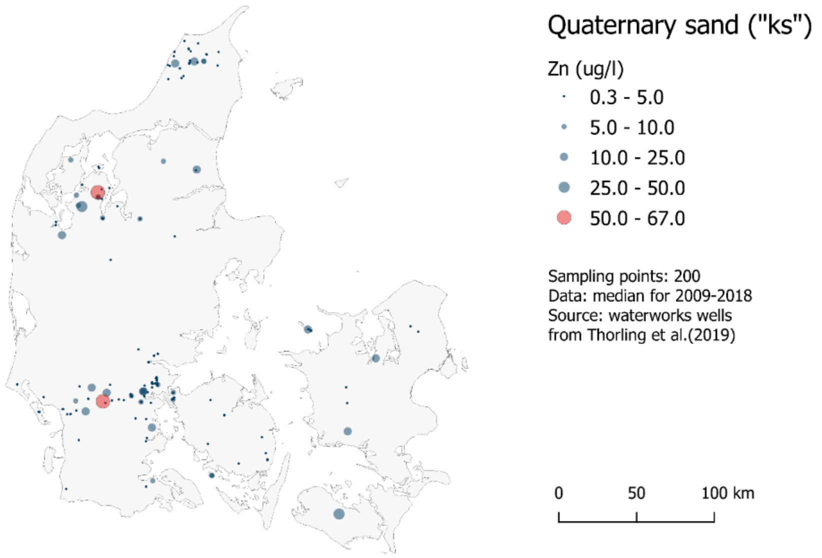

Pre-Quaternary sand ("ps")
Ni (ug/l)
· 0.03 - 1.0
• 1.0 - 5.0
● 5.0 - 10.0
● 10.0 - 20.0
● 20.0 - 110.0
Sampling points: 565
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)
0
50
100 km

Zink (Zn)



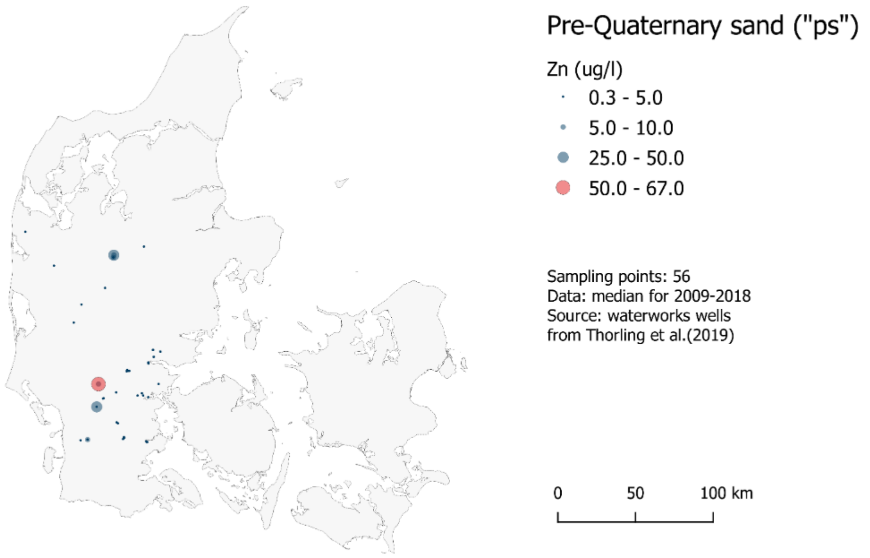
Carbonates ("kalk")

Sampling points: 85
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)



Quaternary sand ("ks")

Sampling points: 200
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)



Pre-Quaternary sand ("ps")

Sampling points: 56
Data: median for 2009-2018
Source: waterworks wells
from Thorling et al.(2019)

14.3 Appendix – Case study – Loire-Bretagne (France)

14.3.1 *Descriptive statistics by lithology, pH and redox classes*

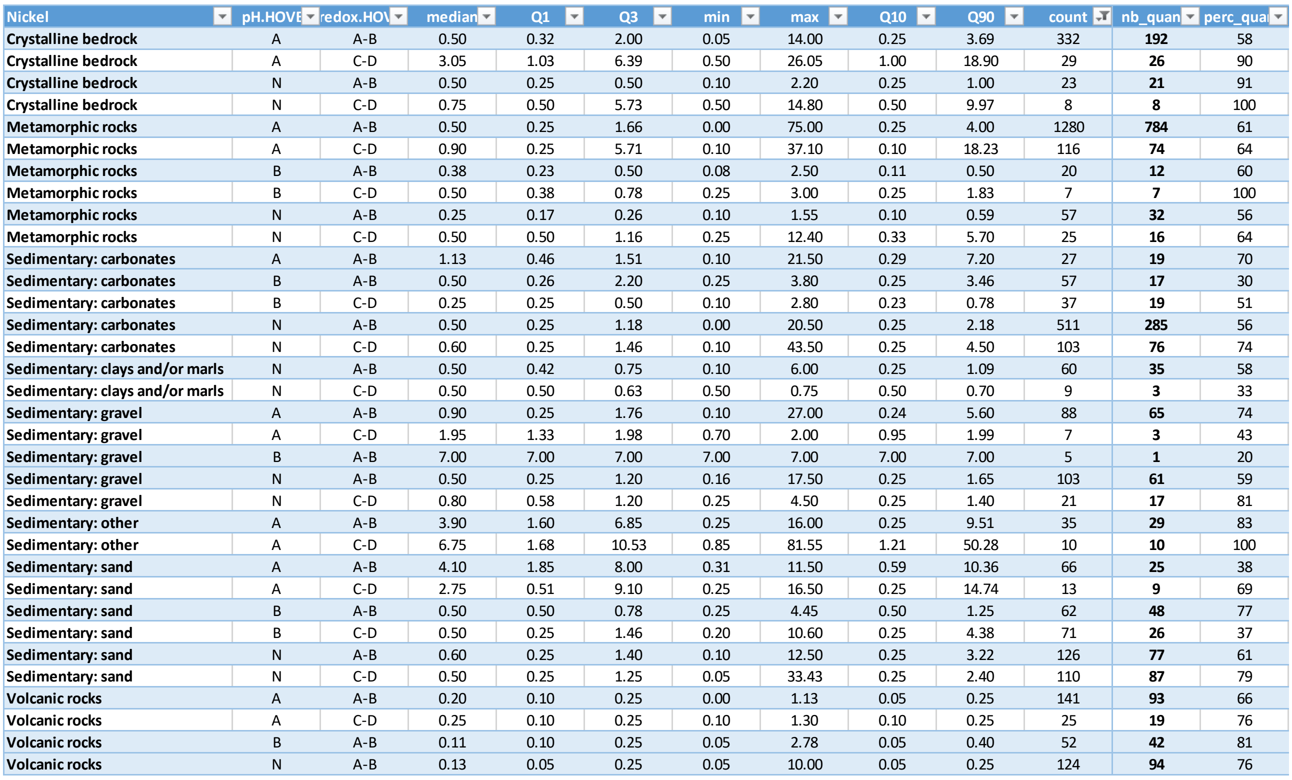
Treatment is applied to the whole dataset, results are presented only for count/analyses >5. Nb_quant is the number of analyses > LOQ and perc_quant the % of analyses > LOQ.

Arsenic	pH.HOVE	redox.HO	median	Q1	Q3	min	max	Q10	Q90	count	nb_quant	perc_qua
Crystalline bedrock	A	A-B	0.5	0.3	1.0	0.1	32.0	0.1	3.8	332	**272**	82
Crystalline bedrock	A	C-D	0.4	0.1	0.6	0.1	22.3	0.1	1.4	29	**28**	97
Crystalline bedrock	N	A-B	0.3	0.1	0.9	0.1	50.9	0.1	1.0	23	**21**	91
Crystalline bedrock	N	C-D	0.5	0.2	0.9	0.1	4.6	0.2	2.1	8	**8**	100
Metamorphic rocks	A	A-B	1.0	0.5	2.0	0.0	187.0	0.4	6.5	1280	**1068**	83
Metamorphic rocks	A	C-D	1.0	0.6	3.5	0.1	80.0	0.4	6.8	116	**102**	88
Metamorphic rocks	B	A-B	1.0	0.5	1.6	0.3	108.0	0.3	3.6	20	**20**	100
Metamorphic rocks	B	C-D	0.5	0.3	2.7	0.3	4.5	0.3	4.1	7	**7**	100
Metamorphic rocks	N	A-B	1.0	0.8	4.7	0.3	138.0	0.4	7.0	57	**51**	89
Metamorphic rocks	N	C-D	1.0	0.6	3.0	0.3	67.0	0.3	10.5	25	**19**	76
Sedimentary: carbonates	A	A-B	1.0	0.8	1.0	0.2	80.4	0.3	3.3	27	**27**	100
Sedimentary: carbonates	B	A-B	1.0	1.0	1.9	0.2	30.0	0.6	4.0	57	**55**	96
Sedimentary: carbonates	B	C-D	3.5	1.0	6.6	0.3	23.0	0.3	15.5	37	**36**	97
Sedimentary: carbonates	N	A-B	1.0	0.4	1.0	0.1	16.0	0.3	1.0	511	**487**	95
Sedimentary: carbonates	N	C-D	0.6	0.3	2.3	0.1	12.0	0.3	4.2	103	**102**	99
Sedimentary: clays and/or marls	N	A-B	1.0	0.5	1.1	0.1	17.2	0.3	4.0	60	**59**	98
Sedimentary: clays and/or marls	N	C-D	1.0	1.0	1.9	1.0	9.0	1.0	7.8	9	**9**	100
Sedimentary: gravel	A	A-B	1.0	1.0	2.0	0.1	9.5	0.7	4.1	88	**87**	99
Sedimentary: gravel	A	C-D	17.5	12.1	23.6	7.5	43.0	9.5	33.8	7	**6**	86
Sedimentary: gravel	B	A-B	1.0	1.0	13.5	1.0	26.0	1.0	21.0	5	**3**	60
Sedimentary: gravel	N	A-B	1.6	1.0	3.0	0.1	11.0	1.0	4.0	103	**93**	90
Sedimentary: gravel	N	C-D	6.6	4.6	10.8	0.3	21.5	1.3	17.4	21	**18**	86
Sedimentary: other	A	A-B	0.3	0.2	0.5	0.1	3.0	0.1	0.8	35	**29**	83
Sedimentary: other	A	C-D	0.5	0.3	0.9	0.3	1.9	0.3	1.6	10	**10**	100
Sedimentary: sand	A	A-B	1.0	0.4	1.0	0.0	6.6	0.3	1.0	66	**61**	92
Sedimentary: sand	A	C-D	1.4	0.9	7.0	0.2	26.0	0.3	13.0	13	**12**	92
Sedimentary: sand	B	A-B	0.5	0.5	1.0	0.3	18.9	0.5	1.2	62	**62**	100
Sedimentary: sand	B	C-D	1.0	0.6	1.0	0.3	22.5	0.3	2.1	71	**70**	99
Sedimentary: sand	N	A-B	1.0	0.4	1.0	0.1	5.0	0.3	1.9	126	**121**	96
Sedimentary: sand	N	C-D	0.5	0.3	1.0	0.1	24.0	0.3	1.7	110	**108**	98
Volcanic rocks	A	A-B	0.8	0.2	1.0	0.0	5.1	0.1	1.0	141	**132**	94
Volcanic rocks	A	C-D	0.8	0.2	1.0	0.0	3.6	0.0	1.0	25	**25**	100
Volcanic rocks	B	A-B	0.8	0.6	1.0	0.2	4.8	0.3	2.8	52	**50**	96
Volcanic rocks	N	A-B	0.8	0.2	1.0	0.0	31.0	0.0	3.9	124	**122**	98

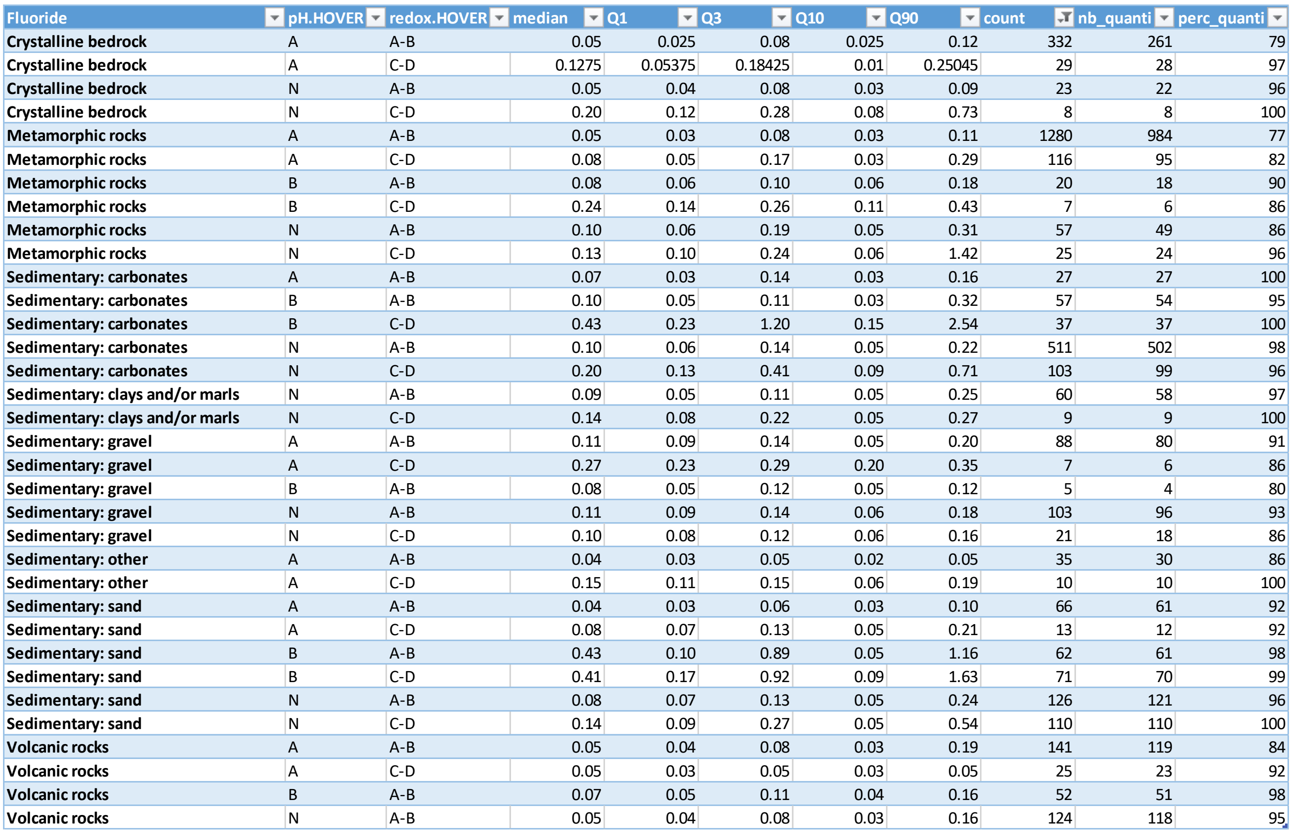
Cadmium	pH.HOVE	redox.HOV	median	Q1	Q3	min	max	Q10	Q90	count	nb_quan	perc_qua
Crystalline bedrock	A	A-B	0.25	0.10	0.50	0.01	1.10	0.04	0.50	332	**313**	94
Crystalline bedrock	A	C-D	0.05	0.01	0.08	0.01	1.49	0.01	0.50	29	**29**	100
Crystalline bedrock	N	A-B	0.06	0.03	0.50	0.01	1.00	0.01	0.70	23	**23**	100
Crystalline bedrock	N	C-D	0.02	0.01	0.04	0.01	0.40	0.01	0.17	8	**8**	100
Metamorphic rocks	A	A-B	0.38	0.25	0.50	0.01	1.95	0.05	0.50	1280	**1179**	92
Metamorphic rocks	A	C-D	0.26	0.06	0.50	0.01	1.85	0.01	0.50	116	**108**	93
Metamorphic rocks	B	A-B	0.31	0.13	0.50	0.01	0.50	0.01	0.50	20	**20**	100
Metamorphic rocks	B	C-D	0.13	0.02	0.19	0.01	0.25	0.01	0.25	7	**7**	100
Metamorphic rocks	N	A-B	0.50	0.19	0.50	0.01	2.00	0.01	1.00	57	**55**	96
Metamorphic rocks	N	C-D	0.25	0.03	0.50	0.01	0.50	0.01	0.50	25	**24**	96
Sedimentary: carbonates	A	A-B	0.38	0.25	0.50	0.01	1.00	0.04	0.50	27	**27**	100
Sedimentary: carbonates	B	A-B	0.50	0.40	0.50	0.01	1.00	0.01	0.50	57	**55**	96
Sedimentary: carbonates	B	C-D	0.50	0.25	0.50	0.01	1.00	0.02	0.50	37	**35**	95
Sedimentary: carbonates	N	A-B	0.50	0.05	0.50	0.01	1.30	0.01	0.50	511	**503**	98
Sedimentary: carbonates	N	C-D	0.50	0.05	0.50	0.01	1.00	0.02	0.50	103	**102**	99
Sedimentary: clays and/or marls	N	A-B	0.28	0.05	0.50	0.01	1.00	0.03	0.50	60	**60**	100
Sedimentary: clays and/or marls	N	C-D	0.50	0.18	0.50	0.01	0.50	0.04	0.50	9	**9**	100
Sedimentary: gravel	A	A-B	0.38	0.19	0.50	0.01	0.50	0.02	0.50	88	**88**	100
Sedimentary: gravel	A	C-D	0.25	0.23	0.25	0.07	0.38	0.15	0.31	7	**6**	86
Sedimentary: gravel	B	A-B	0.75	0.44	1.00	0.25	1.00	0.33	1.00	5	**4**	80
Sedimentary: gravel	N	A-B	0.50	0.25	0.50	0.01	1.00	0.01	0.50	103	**102**	99
Sedimentary: gravel	N	C-D	0.50	0.03	0.50	0.01	1.00	0.02	1.00	21	**18**	86
Sedimentary: other	A	A-B	0.16	0.06	0.28	0.01	0.50	0.04	0.48	35	**33**	94
Sedimentary: other	A	C-D	0.04	0.03	0.20	0.01	1.48	0.01	0.37	10	**10**	100
Sedimentary: sand	A	A-B	0.50	0.15	0.50	0.01	1.00	0.03	0.50	66	**60**	91
Sedimentary: sand	A	C-D	0.15	0.03	0.50	0.01	1.00	0.02	0.73	13	**12**	92
Sedimentary: sand	B	A-B	0.01	0.01	0.37	0.01	0.50	0.01	0.50	62	**62**	100
Sedimentary: sand	B	C-D	0.50	0.50	0.50	0.01	1.00	0.03	0.50	71	**71**	100
Sedimentary: sand	N	A-B	0.50	0.03	0.50	0.01	1.00	0.01	0.75	126	**125**	99
Sedimentary: sand	N	C-D	0.50	0.01	0.50	0.01	1.00	0.01	0.50	110	**110**	100
Volcanic rocks	A	A-B	0.38	0.01	0.50	0.01	0.50	0.01	0.50	141	**131**	93
Volcanic rocks	A	C-D	0.38	0.01	0.50	0.01	1.15	0.01	0.50	25	**25**	100
Volcanic rocks	B	A-B	0.25	0.01	0.47	0.01	0.50	0.01	0.50	52	**50**	96
Volcanic rocks	N	A-B	0.25	0.01	0.38	0.01	0.50	0.01	0.50	124	**123**	99

Chromium	pH.HOVE	redox.HOV	median	Q1	Q3	min	max	Q10	Q90	count	nb_quan	perc_qua
Crystalline bedrock	A	A-B	0.25	0.25	0.25	0.11	3.45	0.13	0.48	332	**22**	7
Crystalline bedrock	A	C-D	0.45	0.24	0.66	0.03	0.88	0.11	0.79	29	**2**	7
Crystalline bedrock	N	A-B	0.25	0.25	0.25	0.20	0.25	0.22	0.25	23	**5**	22
Crystalline bedrock	N	C-D	0.25	0.25	0.25	0.25	0.25	0.25	0.25	8	**1**	13
Metamorphic rocks	A	A-B	0.25	0.20	0.40	0.03	1.01	0.12	0.50	1280	**98**	8
Metamorphic rocks	A	C-D	0.25	0.25	0.25	0.03	0.50	0.03	0.28	116	**10**	9
Metamorphic rocks	B	A-B	0.73	0.61	0.84	0.50	0.95	0.55	0.91	20	**2**	10
Metamorphic rocks	B	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7	**0**	0
Metamorphic rocks	N	A-B	0.25	0.25	0.43	0.25	0.57	0.25	0.51	57	**5**	9
Metamorphic rocks	N	C-D	0.25	0.20	0.38	0.16	0.50	0.17	0.45	25	**3**	12
Sedimentary: carbonates	A	A-B	0.25	0.21	0.29	0.08	0.98	0.08	0.59	27	**8**	30
Sedimentary: carbonates	B	A-B	0.51	0.47	0.62	0.43	0.72	0.44	0.68	57	**3**	5
Sedimentary: carbonates	B	C-D	0.25	0.10	0.25	0.06	0.25	0.08	0.25	37	**5**	14
Sedimentary: carbonates	N	A-B	0.25	0.25	0.25	0.03	1.40	0.24	0.45	511	**105**	21
Sedimentary: carbonates	N	C-D	0.25	0.25	0.25	0.03	0.50	0.23	0.40	103	**15**	15
Sedimentary: clays and/or marls	N	A-B	0.25	0.19	0.25	0.07	0.75	0.07	0.25	60	**11**	18
Sedimentary: clays and/or marls	N	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9	**0**	0
Sedimentary: gravel	A	A-B	0.17	0.10	0.24	0.03	0.46	0.05	0.25	88	**23**	26
Sedimentary: gravel	A	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7	**0**	0
Sedimentary: gravel	B	A-B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5	**0**	0
Sedimentary: gravel	N	A-B	0.15	0.10	0.25	0.07	0.52	0.09	0.25	103	**18**	17
Sedimentary: gravel	N	C-D	0.25	0.25	0.25	0.25	0.25	0.25	0.25	21	**4**	19
Sedimentary: other	A	A-B	0.24	0.18	0.32	0.06	0.66	0.11	0.52	35	**5**	14
Sedimentary: other	A	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10	**0**	0
Sedimentary: sand	A	A-B	0.25	0.14	0.25	0.03	1.14	0.03	0.25	66	**11**	17
Sedimentary: sand	A	C-D	0.25	0.25	0.25	0.03	0.27	0.12	0.26	13	**5**	38
Sedimentary: sand	B	A-B	0.25	0.19	0.42	0.12	0.59	0.15	0.52	62	**3**	5
Sedimentary: sand	B	C-D	0.25	0.25	0.25	0.25	0.25	0.25	0.25	71	**7**	10
Sedimentary: sand	N	A-B	0.25	0.25	0.30	0.15	0.73	0.24	0.57	126	**14**	11
Sedimentary: sand	N	C-D	0.25	0.19	0.25	0.03	0.25	0.03	0.25	110	**8**	7
Volcanic rocks	A	A-B	0.25	0.24	0.26	0.03	0.44	0.13	0.35	141	**6**	4
Volcanic rocks	A	C-D	0.21	0.15	0.26	0.10	0.31	0.12	0.29	25	**2**	8
Volcanic rocks	B	A-B	1.67	1.41	1.94	1.14	2.20	1.25	2.09	52	**2**	4
Volcanic rocks	B	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3	**0**	0
Volcanic rocks	N	A-B	0.25	0.11	0.51	0.03	1.00	0.03	0.92	124	**11**	9

Copper	pH.HOVE	redox.HOV	median	Q1	Q3	min	max	Q10	Q90	count	nb_quan	perc_qua
Crystalline bedrock	A	A-B	0.58	0.25	4.50	0.17	170.00	0.25	20.00	332	**21**	6
Crystalline bedrock	A	C-D	0.16	0.12	0.21	0.08	0.25	0.09	0.23	29	**2**	7
Crystalline bedrock	N	A-B	1.00	0.84	1.80	0.10	2.00	0.40	1.92	23	**5**	22
Crystalline bedrock	N	C-D	0.25	0.25	0.25	0.25	0.25	0.25	0.25	8	**1**	13
Metamorphic rocks	A	A-B	0.50	0.25	2.27	0.11	1440.00	0.25	10.90	1280	**94**	7
Metamorphic rocks	A	C-D	0.25	0.25	0.25	0.24	2.40	0.25	1.02	116	**9**	8
Metamorphic rocks	B	A-B	0.26	0.26	0.26	0.26	0.26	0.26	0.26	20	**1**	5
Metamorphic rocks	B	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7	**0**	0
Metamorphic rocks	N	A-B	1.80	0.62	3.50	0.32	94.30	0.36	49.10	57	**6**	11
Metamorphic rocks	N	C-D	0.25	0.25	80.13	0.25	160.00	0.25	128.05	25	**3**	12
Sedimentary: carbonates	A	A-B	1.09	0.24	2.16	0.08	12.00	0.17	5.22	27	**8**	30
Sedimentary: carbonates	B	A-B	0.52	0.25	1.50	0.25	6.00	0.25	4.20	57	**5**	9
Sedimentary: carbonates	B	C-D	0.25	0.13	0.27	0.05	2.15	0.07	1.03	37	**7**	19
Sedimentary: carbonates	N	A-B	0.62	0.30	1.44	0.08	321.00	0.25	3.20	511	**109**	21
Sedimentary: carbonates	N	C-D	0.25	0.19	0.25	0.08	2.00	0.13	0.85	103	**17**	17
Sedimentary: clays and/or marls	N	A-B	0.28	0.25	0.54	0.19	3.80	0.19	1.29	60	**11**	18
Sedimentary: clays and/or marls	N	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9	**0**	0
Sedimentary: gravel	A	A-B	0.97	0.77	1.80	0.25	130.00	0.51	8.54	88	**23**	26
Sedimentary: gravel	A	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7	**0**	0
Sedimentary: gravel	B	A-B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5	**0**	0
Sedimentary: gravel	N	A-B	0.99	0.75	1.34	0.25	19.00	0.58	2.13	103	**19**	18
Sedimentary: gravel	N	C-D	0.25	0.25	0.31	0.25	0.49	0.25	0.41	21	**4**	19
Sedimentary: other	A	A-B	0.61	0.55	1.26	0.29	2.10	0.39	1.76	35	**5**	14
Sedimentary: other	A	C-D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10	**0**	0
Sedimentary: sand	A	A-B	2.60	1.32	6.08	0.17	13.00	0.22	9.80	66	**16**	24
Sedimentary: sand	A	C-D	0.63	0.25	1.23	0.20	3.05	0.22	2.18	13	**6**	46
Sedimentary: sand	B	A-B	0.79	0.37	1.38	0.25	2.00	0.30	1.75	62	**4**	6
Sedimentary: sand	B	C-D	0.25	0.25	0.28	0.23	2.85	0.24	1.33	71	**7**	10
Sedimentary: sand	N	A-B	0.80	0.25	2.36	0.08	32.30	0.25	6.90	126	**15**	12
Sedimentary: sand	N	C-D	0.25	0.25	0.28	0.08	10.69	0.20	3.47	110	**8**	7
Volcanic rocks	A	A-B	0.25	0.25	0.25	0.25	3.10	0.25	1.39	141	**7**	5
Volcanic rocks	A	C-D	0.25	0.25	0.25	0.25	0.25	0.25	0.25	25	**2**	8
Volcanic rocks	B	A-B	0.32	0.25	0.39	0.18	0.47	0.21	0.44	52	**2**	4
Volcanic rocks	N	A-B	0.25	0.25	0.62	0.24	1.20	0.25	1.00	124	**11**	9



Nickel	pH.HOVE	redox.HO\	median	Q1	Q3	min	max	Q10	Q90	count	nb_quan	perc_qua
Crystalline bedrock	A	A-B	0.50	0.32	2.00	0.05	14.00	0.25	3.69	332	**192**	58
Crystalline bedrock	A	C-D	3.05	1.03	6.39	0.50	26.05	1.00	18.90	29	**26**	90
Crystalline bedrock	N	A-B	0.50	0.25	0.50	0.10	2.20	0.25	1.00	23	**21**	91
Crystalline bedrock	N	C-D	0.75	0.50	5.73	0.50	14.80	0.50	9.97	8	**8**	100
Metamorphic rocks	A	A-B	0.50	0.25	1.66	0.00	75.00	0.25	4.00	1280	**784**	61
Metamorphic rocks	A	C-D	0.90	0.25	5.71	0.10	37.10	0.10	18.23	116	**74**	64
Metamorphic rocks	B	A-B	0.38	0.23	0.50	0.08	2.50	0.11	0.50	20	**12**	60
Metamorphic rocks	B	C-D	0.50	0.38	0.78	0.25	3.00	0.25	1.83	7	**7**	100
Metamorphic rocks	N	A-B	0.25	0.17	0.26	0.10	1.55	0.10	0.59	57	**32**	56
Metamorphic rocks	N	C-D	0.50	0.50	1.16	0.25	12.40	0.33	5.70	25	**16**	64
Sedimentary: carbonates	A	A-B	1.13	0.46	1.51	0.10	21.50	0.29	7.20	27	**19**	70
Sedimentary: carbonates	B	A-B	0.50	0.26	2.20	0.25	3.80	0.25	3.46	57	**17**	30
Sedimentary: carbonates	B	C-D	0.25	0.25	0.50	0.10	2.80	0.23	0.78	37	**19**	51
Sedimentary: carbonates	N	A-B	0.50	0.25	1.18	0.00	20.50	0.25	2.18	511	**285**	56
Sedimentary: carbonates	N	C-D	0.60	0.25	1.46	0.10	43.50	0.25	4.50	103	**76**	74
Sedimentary: clays and/or marls	N	A-B	0.50	0.42	0.75	0.10	6.00	0.25	1.09	60	**35**	58
Sedimentary: clays and/or marls	N	C-D	0.50	0.50	0.63	0.50	0.75	0.50	0.70	9	**3**	33
Sedimentary: gravel	A	A-B	0.90	0.25	1.76	0.10	27.00	0.24	5.60	88	**65**	74
Sedimentary: gravel	A	C-D	1.95	1.33	1.98	0.70	2.00	0.95	1.99	7	**3**	43
Sedimentary: gravel	B	A-B	7.00	7.00	7.00	7.00	7.00	7.00	7.00	5	**1**	20
Sedimentary: gravel	N	A-B	0.50	0.25	1.20	0.16	17.50	0.25	1.65	103	**61**	59
Sedimentary: gravel	N	C-D	0.80	0.58	1.20	0.25	4.50	0.25	1.40	21	**17**	81
Sedimentary: other	A	A-B	3.90	1.60	6.85	0.25	16.00	0.25	9.51	35	**29**	83
Sedimentary: other	A	C-D	6.75	1.68	10.53	0.85	81.55	1.21	50.28	10	**10**	100
Sedimentary: sand	A	A-B	4.10	1.85	8.00	0.31	11.50	0.59	10.36	66	**25**	38
Sedimentary: sand	A	C-D	2.75	0.51	9.10	0.25	16.50	0.25	14.74	13	**9**	69
Sedimentary: sand	B	A-B	0.50	0.50	0.78	0.25	4.45	0.50	1.25	62	**48**	77
Sedimentary: sand	B	C-D	0.50	0.25	1.46	0.20	10.60	0.25	4.38	71	**26**	37
Sedimentary: sand	N	A-B	0.60	0.25	1.40	0.10	12.50	0.25	3.22	126	**77**	61
Sedimentary: sand	N	C-D	0.50	0.25	1.25	0.05	33.43	0.25	2.40	110	**87**	79
Volcanic rocks	A	A-B	0.20	0.10	0.25	0.00	1.13	0.05	0.25	141	**93**	66
Volcanic rocks	A	C-D	0.25	0.10	0.25	0.10	1.30	0.10	0.25	25	**19**	76
Volcanic rocks	B	A-B	0.11	0.10	0.25	0.05	2.78	0.05	0.40	52	**42**	81
Volcanic rocks	N	A-B	0.13	0.05	0.25	0.05	10.00	0.05	0.25	124	**94**	76



Fluoride	pH.HOVER	redox.HOVER	median	Q1	Q3	Q10	Q90	count	nb_quanti	perc_quanti
Crystalline bedrock	A	A-B	0.05	0.025	0.08	0.025	0.12	332	261	79
Crystalline bedrock	A	C-D	0.1275	0.05375	0.18425	0.01	0.25045	29	28	97
Crystalline bedrock	N	A-B	0.05	0.04	0.08	0.03	0.09	23	22	96
Crystalline bedrock	N	C-D	0.20	0.12	0.28	0.08	0.73	8	8	100
Metamorphic rocks	A	A-B	0.05	0.03	0.08	0.03	0.11	1280	984	77
Metamorphic rocks	A	C-D	0.08	0.05	0.17	0.03	0.29	116	95	82
Metamorphic rocks	B	A-B	0.08	0.06	0.10	0.06	0.18	20	18	90
Metamorphic rocks	B	C-D	0.24	0.14	0.26	0.11	0.43	7	6	86
Metamorphic rocks	N	A-B	0.10	0.06	0.19	0.05	0.31	57	49	86
Metamorphic rocks	N	C-D	0.13	0.10	0.24	0.06	1.42	25	24	96
Sedimentary: carbonates	A	A-B	0.07	0.03	0.14	0.03	0.16	27	27	100
Sedimentary: carbonates	B	A-B	0.10	0.05	0.11	0.03	0.32	57	54	95
Sedimentary: carbonates	B	C-D	0.43	0.23	1.20	0.15	2.54	37	37	100
Sedimentary: carbonates	N	A-B	0.10	0.06	0.14	0.05	0.22	511	502	98
Sedimentary: carbonates	N	C-D	0.20	0.13	0.41	0.09	0.71	103	99	96
Sedimentary: clays and/or marls	N	A-B	0.09	0.05	0.11	0.05	0.25	60	58	97
Sedimentary: clays and/or marls	N	C-D	0.14	0.08	0.22	0.05	0.27	9	9	100
Sedimentary: gravel	A	A-B	0.11	0.09	0.14	0.05	0.20	88	80	91
Sedimentary: gravel	A	C-D	0.27	0.23	0.29	0.20	0.35	7	6	86
Sedimentary: gravel	B	A-B	0.08	0.05	0.12	0.05	0.12	5	4	80
Sedimentary: gravel	N	A-B	0.11	0.09	0.14	0.06	0.18	103	96	93
Sedimentary: gravel	N	C-D	0.10	0.08	0.12	0.06	0.16	21	18	86
Sedimentary: other	A	A-B	0.04	0.03	0.05	0.02	0.05	35	30	86
Sedimentary: other	A	C-D	0.15	0.11	0.15	0.06	0.19	10	10	100
Sedimentary: sand	A	A-B	0.04	0.03	0.06	0.03	0.10	66	61	92
Sedimentary: sand	A	C-D	0.08	0.07	0.13	0.05	0.21	13	12	92
Sedimentary: sand	B	A-B	0.43	0.10	0.89	0.05	1.16	62	61	98
Sedimentary: sand	B	C-D	0.41	0.17	0.92	0.09	1.63	71	70	99
Sedimentary: sand	N	A-B	0.08	0.07	0.13	0.05	0.24	126	121	96
Sedimentary: sand	N	C-D	0.14	0.09	0.27	0.05	0.54	110	110	100
Volcanic rocks	A	A-B	0.05	0.04	0.08	0.03	0.19	141	119	84
Volcanic rocks	A	C-D	0.05	0.03	0.05	0.03	0.05	25	23	92
Volcanic rocks	B	A-B	0.07	0.05	0.11	0.04	0.16	52	51	98
Volcanic rocks	N	A-B	0.05	0.04	0.08	0.03	0.16	124	118	95

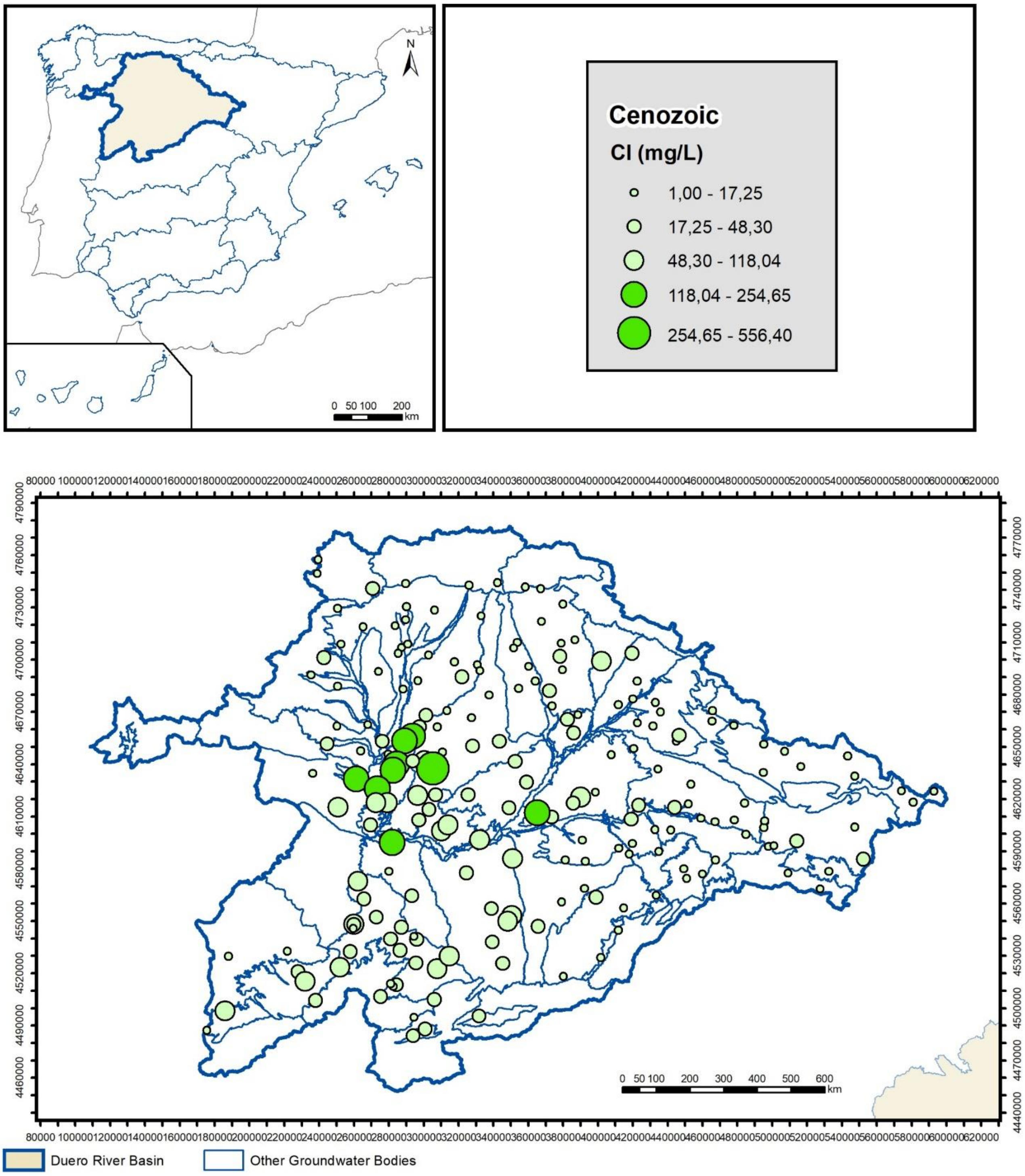
14.4 Appendix – Case study – Duero River Basin Spain

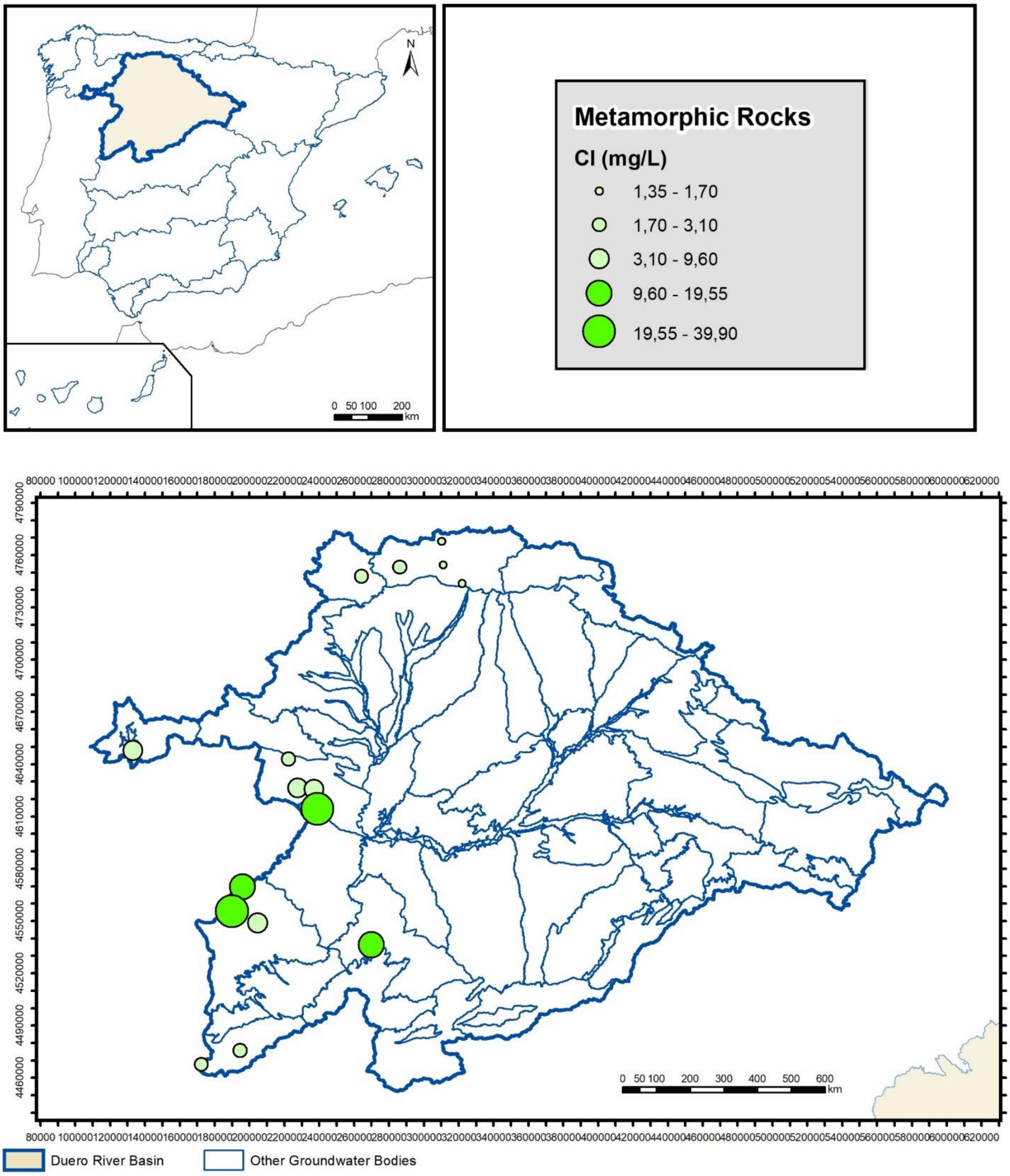
14.4.1 *Parameters measured in chemical analyses (including organic pollutants)*

"Alcalinidad", "Al", "NH4+", "As", "HCO3-", "Ca2+", "CO32-", "COT", "Cl-", "Cu", "C.E. in situ", "Dureza total", "PO43-", "Fe", "Mg2+", "Mn", "NO3-", "NO2-", "O2 disuelto in situ", "pH in situ", "K+", "SiO2", "Na+", "SO42-", "Tª agua", "Zn", "Aldrín", "a-HCH", "Ametrina", "Atrazina", "b-HCH", "Clorfenvinfos", "Clorpirifós", "Colif. Fec.", "Colif. Tot.", "d-HCH", "Diclorfentión", "Dieldrín", "Endosulfán I", "Endosulfán II", "Endosulfán sulfato", "Endrín", "Endrín Aldehido", "Escherichia coli", "Etil Bromofos", "Etil Paratión", "Fenclorfos", "Fenitrotión", "Lindano (g-HCH)", "Heptacloro", "Heptacloro epóxido", "Hexaclorobenceno", "Metidatión", "Metil Bromofos", "p,p'-DDD", "p,p'-DDE", "p,p'-DDT", "Prometrina", "Salmonelas", "Simazina", "Terbutilazina", "Terbutrina", "Tetraclorvinfos", "Trietazina", "Trifluralina", "Estrep. Fec.", "Oxidabilidad", "Cr", "Cd", "Enterococos", "Pb", "Cobre disuelto", "Hg", "F-", "Diazinón", "Endrín Cetona", "Etión", "Metil Paratión", "Metoxicloro", "Paratión", "Plaguicidas totales", "Propazina", "Tetracloroeteno", "Tricloroeteno", "Tetracloroetileno", "Tricloroetileno", "CO2 in situ", "Ba", "Be", "B", "Br-", "Co", "C.E.", "Cr VI", "DQO", "Dureza", "Ni", "pH", "Se", "Tª ambiente", "V", "Oxígeno disuelto (% satur.)", "1,1,1-Tricloroetano", "1,2,3-Triclorobenceno", "1,2,4-Triclorobenceno", "1,2-Diclorobenceno", "1,2-Dicloroetano", "1,3-Diclorobenceno", "1,4-Diclorobenceno", "Alacloro", "Antraceno", "Benceno", "Benzo(a)pireno", "Benzo(b)fluoranteno", "Benzo(g,h,i)perileno", "Benzo(k)fluoranteno", "Clorobenceno", "Cloroformo", "DEHP", "Diclorometano", "Diurón", "e-HCH", "Etilbenceno", "Fluoranteno", "HCBD", "Indeno(1,2,3-cd)pireno", "Isoproturón", "m,p-Xilenos", "Metolacloro", "Naftaleno", "Nonilfenol", "Octilfenol", "o-Xileno", "Pentaclorobenceno", "Pentaclorofenol", "Tolueno", "Eh (campo)", "4,4'-DDD", "4,4'-DDE", "Clodinafop prop", "Etil Azinfos", "Fonofos", "Formotión", "Isodrín", "Malatión", "Metamidofos", "Metil Azinfos", "Metil Clorpirifós", "Metil Pirimifos", "Naled", "Oxifluorfen", "Paratión (paratión etilo)", "Profenofos", "NH3", "DBO5", "P Tot.", "1,2,4-Trimetilbenceno", "1,3,5-Triclorobenceno", "Acenafteno", "Acenaftileno", "Benzo(a)antraceno", "Bromodiclorometano", "Criseno", "Dibenzo(ah)antraceno", "Dibromoclorometano", "Fenantreno", "Fluoreno", "Imazalil", "Isopropilbenceno", "Metazacloro", "Molinato", "o,p'-DDT", "Pireno", "CCl4", "Bromoformo", "Di(2-etilhexil)ftalato", "Fentión", "Fosalón", "Pendimetalina", "1,1,2-Tricloroetano", "N Tot.", "Q", "N.P.", "4,4'-DDT", "AMPA", "Benzo(a)Antraceno", "Bromacilo", "DDT (suma máxima)", "DDT (suma mínima)", "DesetilAtrazina", "Desispropilatrazina", "Dibenzo(ah)Antraceno", "Dimetoato", "Disulfotón", "Glifosato", "Linurón", "Metribuzina", "Pirimicarb", "Prometón", "Propizamida", "Secbumetón", "Tiabendazol", "Metil Cloropirifós", "THM", "Sb", "Sólidos en suspensión", "Turbidez", "P. Atm.", "AldrÍn", "N.P.I.", "NO32-", "N.P. final", "N.P.inicial", "pHi inicial", "Altura tapa", "NA"

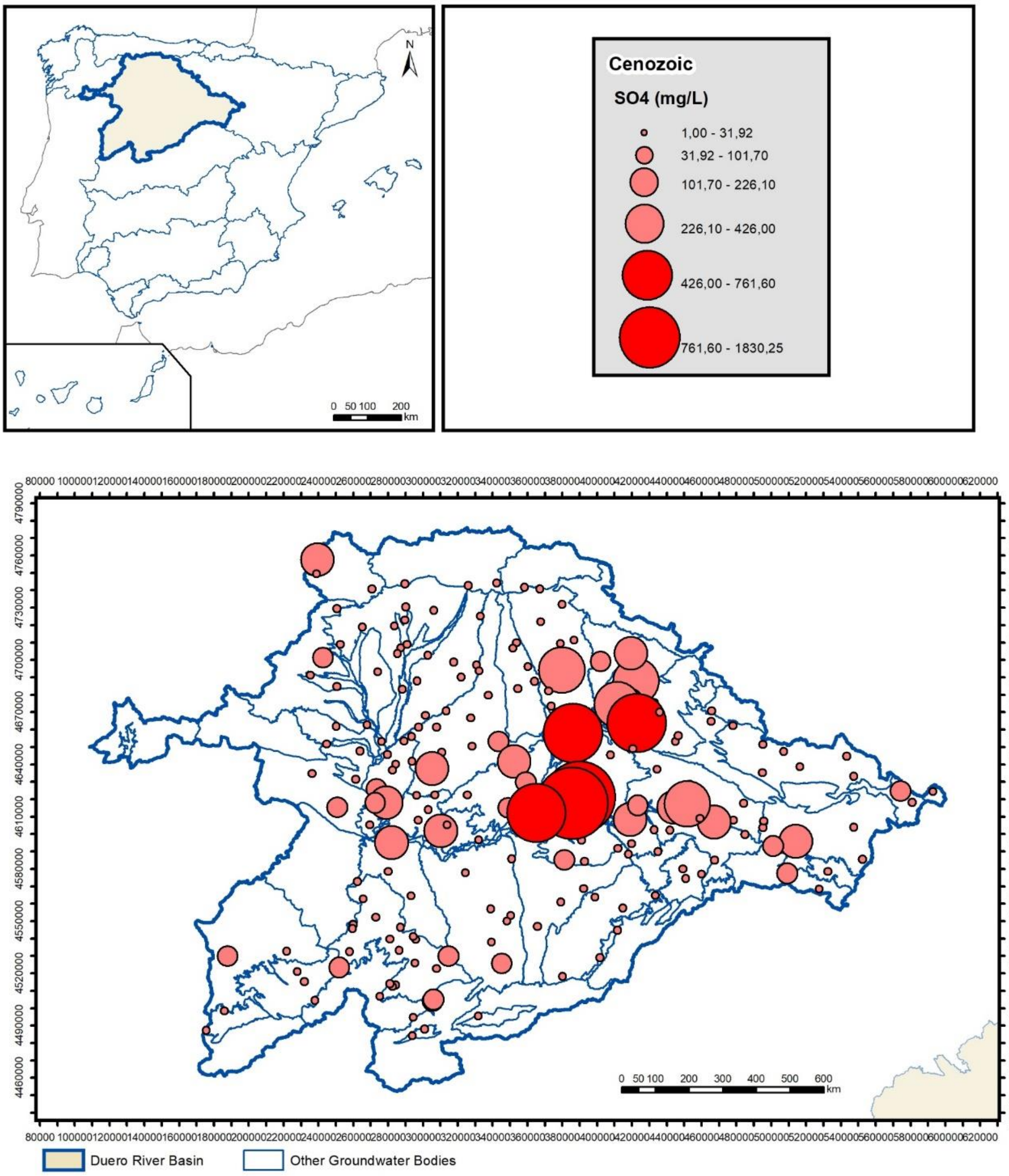
14.4.2 *Spatial distribution of elemental concentrations in Duero River Basin (Spain) groundwater (only waterworks well) after excluding water sampling points with element-specific anthropogenic pressures*

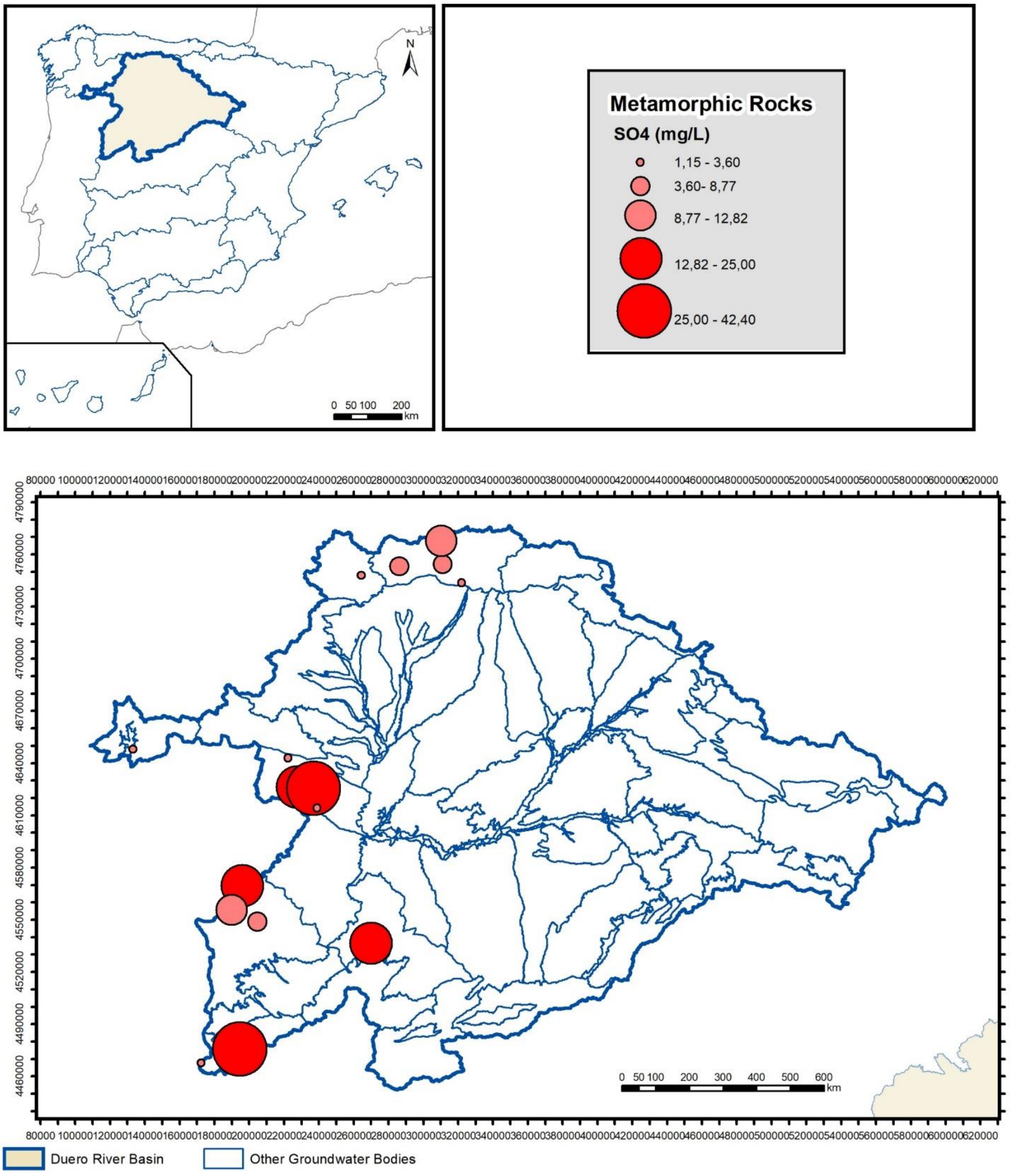
14.4.2.1 Chloride (Cl^-)



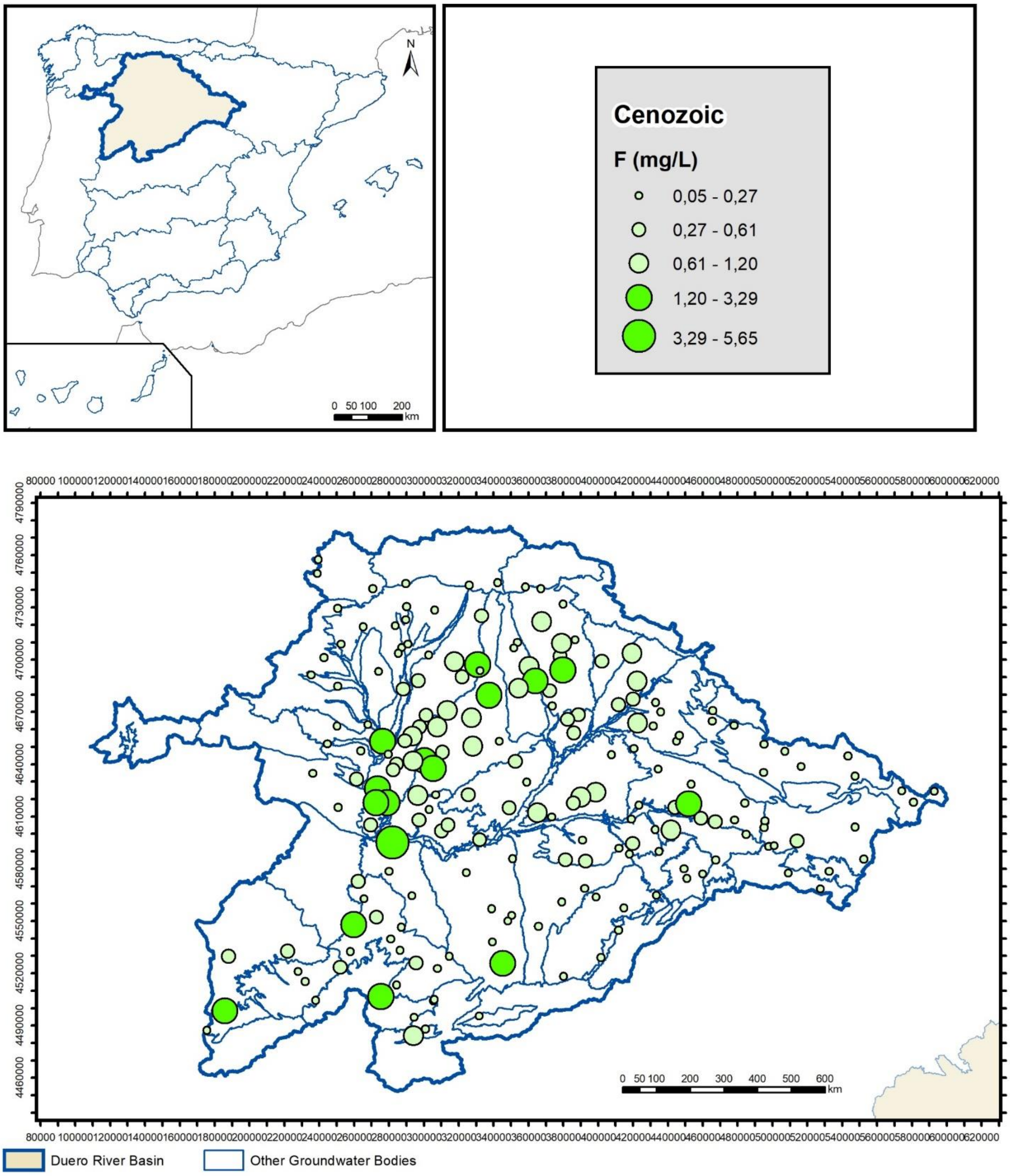

Metamorphic Rocks
Cl (mg/L)
1,35 - 1,70
1,70 - 3,10
3,10 - 9,60
9,60 - 19,55
19,55 - 39,90
N
0 50 100 200 km
0 50 100 200 300 400 500 600 km
80000 100000 120000 140000 160000 180000 200000 220000 240000 260000 280000 300000 320000 340000 360000 380000 400000 420000 440000 460000 480000 500000 520000 540000 560000 580000 600000 620000
4790000 4760000 4730000 4700000 4670000 4640000 4610000 4580000 4550000 4520000 4490000 4460000
4770000 4740000 4710000 4680000 4650000 4620000 4590000 4560000 4530000 4500000 4470000 4440000
Duero River Basin
Other Groundwater Bodies

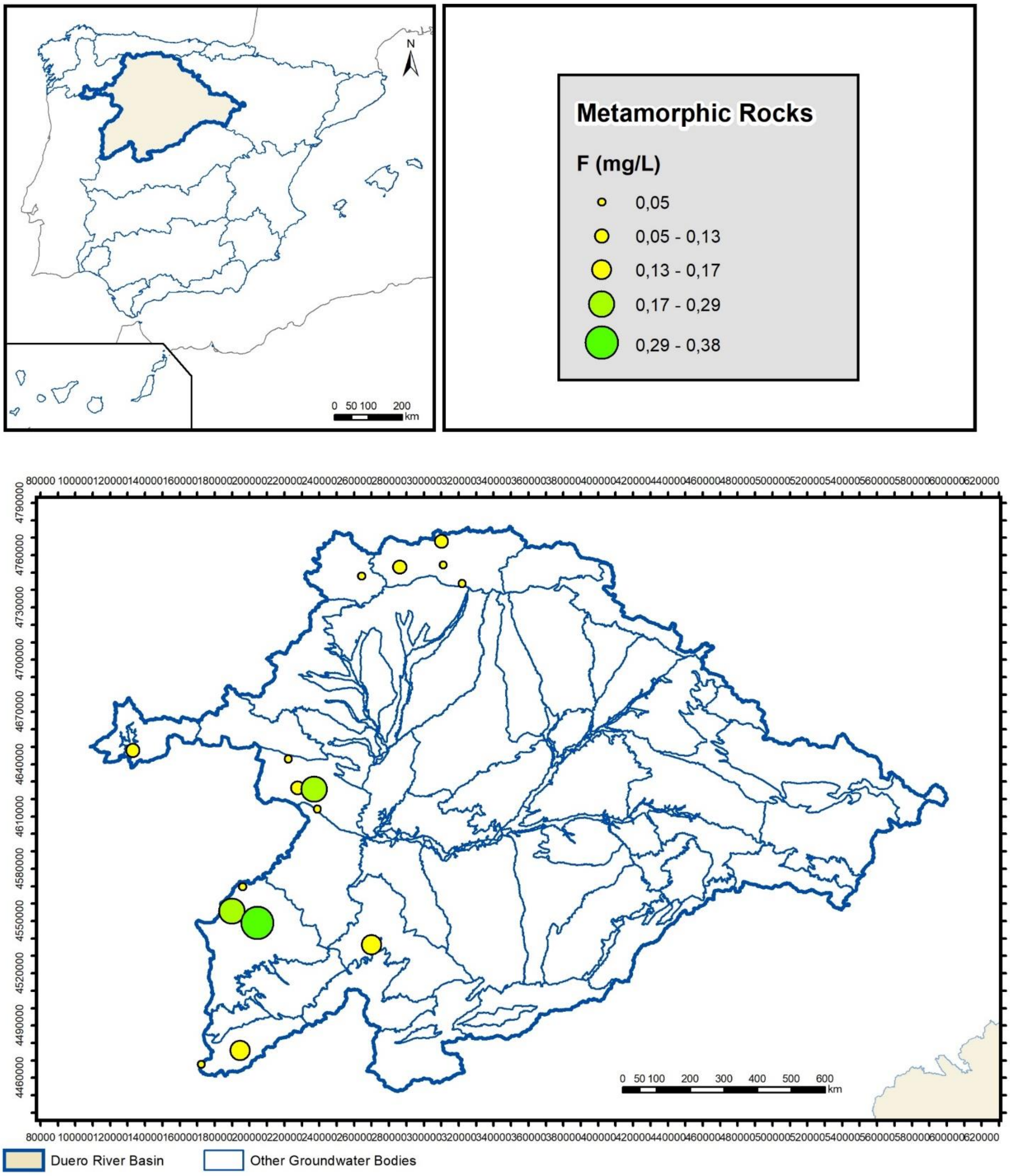
14.4.2.2 Sulphate (SO_4^-)




N
0 50 100 200
km
Metamorphic Rocks
SO4 (mg/L)
1,15 - 3,60
3,60- 8,77
8,77 - 12,82
12,82 - 25,00
25,00 - 42,40
80000 100000 120000 140000 160000 180000 200000 220000 240000 260000 280000 300000 320000 340000 360000 380000 400000 420000 440000 460000 480000 500000 520000 540000 560000 580000 600000 620000
4790000 4760000 4730000 4700000 4670000 4640000 4610000 4580000 4550000 4520000 4490000 4460000
4770000 4740000 4710000 4680000 4650000 4620000 4590000 4560000 4530000 4500000 4470000 4440000
0 50 100 200 300 400 500 600
km
Duero River Basin
Other Groundwater Bodies

14.4.2.3 Fluoride (F^-)




N
0 50 100 200 km
Metamorphic Rocks
F (mg/L)
0,05
0,05 - 0,13
0,13 - 0,17
0,17 - 0,29
0,29 - 0,38
0 50 100 200 300 400 500 600 km
Duero River Basin
Other Groundwater Bodies