Hydrogeology Journal
https://doi.org/10.1007/s10040-023-02646-1

PAPER

Tracking the sources and fate of nitrate pollution by combining hydrochemical and isotopic data with a statistical approach

Louis Christiaens[1,2] · Philippe Orban[2] · Serge Brouyère[2] · Pascal Goderniaux[1]

Received: 17 October 2022 / Accepted: 17 May 2023

Abstract

This study contributes to identifying and spatializing the different types of nitrate sources by combining hydrogeochemical and isotopic data with principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) multicriteria statistical methods. The methodology is applied to the strategic Mons Basin chalk aquifer (Belgium). The results are based on a whole dataset containing 72 water samples with analyses of the hydrogeochemical parameters (temperature, pH, electrical conductivity (EC), redox potential, dissolved O_2), alkalinity, total organic carbon (TOC), silica (SiO_2), major and minor ions (NO_3^-, NH^{4+}, Ca^{2+}, dissolved Fe and Mn, K^+, Mg^{2+}, Na^+, Sr^{2+}, Cl^-, F^-, SO_4^-, B) and multiple stable isotope ratios ($\delta^{11}B$, $\delta^{15}N$–NO_3^-, $\delta^{18}O$–NO_3^-). Compared to classical PCA, the recently developed t-SNE method, which considers nonlinear relationships between variables and preserves local-scale similarities in a low-dimensional space, showed much better performance in discriminating different groups of samples and related zones in the aquifer. t-SNE results combined with isotope ratios highlighted four zones in the aquifer (grouped as A–D) and the presence of denitrification fronts. Group A presents a manure signature ($\delta^{15}N$–NO_3^- – mean (μ) +12.78‰, standard deviation (σ) 6.48‰; $\delta^{11}B$ – μ 29.96‰, σ 6.91‰). Group B exhibits both manure and inorganic fertilizer signatures ($\delta^{15}N$–NO_3^- – μ 6.27‰, σ 2.55‰; $\delta^{11}B$ – μ 15.86‰, σ 9.69‰). Group C shows a contamination by sewage ($\delta^{15}N$–NO_3^- – μ 12.67‰, σ 5.60‰; $\delta^{11}B$ – μ 9.97‰, σ 7.08‰). Group D presents a mixed signature ($\delta^{15}N$–NO_3^- – μ 9.25‰, σ 2.94‰; $\delta^{11}B$ – μ 20.00‰, σ 6.70‰).

Keywords Groundwater sampling and monitoring · Nitrate · Isotopes · Denitrification · t-SNE

Introduction

Nitrate contamination of groundwater is a global and persistent problem. Thirty years ago, Spalding and Exner (1993) discussed the impact of nitrate on US aquifers. Abascal et al. (2022) showed that, driven by the growth of the world's population and the development of agriculture, nitrate pollution of groundwater affects all continents. The excess nitrate in groundwater induces different threats to human health and the environment. Methemoglobinemia is, for example, a severe disease that mainly affects infants and may occur when too much nitrate is ingested through drinking water (Canter 2019; Fan and Steinberg 1996; Sandor et al. 2001). Eutrophication of aquatic environments may be induced by high nutrient contents in water, including nitrate, leading to an imbalance between excessive development of algae and a decrease in available oxygen, inducing asphyxiation of the environment (Vitousek et al. 1997). In this context, pollution of groundwater and surface water by nitrate has remained a crucial issue for water and environmental managers (Orban et al. 2010; Sutton et al. 2011).

Mitigating nitrate concentrations in aquifers, however, is a complex process that involves identifying the source and fate of the pollutant. Agricultural activities with the application of inorganic fertilizers and manure are the most common source of nitrate considered in impact studies (Di and Cameron 2002; Jackson et al. 2008; Mitchell et al. 2003; Mohamed et al. 2003; Oren et al. 2004; Thorburn et al. 2003). Nevertheless, local wastewater intrants, a damaged or incomplete sewage system (Ronen and Magaritz 1985), or local industrial pollution (Wells et al. 2016) can also lead to groundwater contamination by nitrogen compounds. Those sources may be neglected in contexts where agriculture is strongly prevailing; however, in more urban or mixed

✉ Louis Christiaens
Louis.Christiaens@umons.ac.be

[1] University of Mons – Polytech Mons – Geology and Applied Geology, Mons, Belgium

[2] University of Liège – Urban and Environmental Engineering - Hydrogeology and Environmental Geology, Liège, Belgium

Published online: 12 July 2023

environments, they may become significant or even dominant (Wakida and Lerner 2005). Identifying those sources properly, although not easy, is crucial to implementing efficient and zone-specific mitigation measures.

Different studies have discriminated nitrate pollution sources and analyzed its evolution by measuring the stable isotope ratios of $^{15}N/^{14}N$ and $^{18}O/^{16}O$ in NO_3^-. For example, Böttcher et al. (1990), Chen and MacQuarrie (2005), Mariotti et al. (1988) and Mengis et al. (2005) used this technique to identify spontaneous denitrification along flow paths. Biddau et al. (2019), Böhlke and Denver (1995), Czekaj et al. (2016), Obeidat et al. (2021), Ogrinc et al. (2019), Pittalis et al. (2018), Torres-Martínez et al. (2021) and Vystavna et al. (2017) combined those measurements with broader hydrochemical interpretation (for example, chloride concentration, major ions, redox potential, etc.), considering land use and hydrogeological contexts to discriminate between different potential sources of nitrate in an aquifer, which allows some possible limitations related to a standalone nitrate isotope interpretation to be overcome. Xue et al. (2009) point out that fractionating processes such as nitrification and denitrification occurring within the aquifer may actually modify the isotopic signature inherited from the original source of the nitrate and lead to incorrect interpretation. As Bassett et al. (1995) showed that quantification of the stable isotope ratio of boron (^{11}B) enables the identification of the presence of wastewater in a sample, a complimentary method incorporating this measurement was proposed by Widory et al. (20052013) and Bronders et al. (2012) to refine the identification of nitrate sources based on isotopic measurements. Multiple examples showing the application of this technique are also provided by Accoe et al. (2008), Kruk et al. (2020), Martinelli et al. (2018), and Seiler (2005).

Dimensional reduction algorithms coupled with clustering methods are also used to identify similarities between samples and to determine which groups are likely to be impacted by a common pollution source. In a hydrogeological context, those methods are useful when a substantial dataset is available, usually including data such as physicochemical measurements, major element concentrations and isotope ratios. Methods applied to hydrochemical data include hierarchical clustering analysis (HCA) and principal component analysis (PCA; Su et al. 2020), Kohonen self-organizing maps (SOMs; Gamble and Babbar-Sebens 2011; Torres-Martínez et al. 2021) or dimensional reduction by t-distributed stochastic neighbor embedding (t-SNE; ex. Liu et al. 2021).

In this context, this work aims to study the complementarity of different methods in discriminating and characterizing the different sources and fate of nitrate when applied to the case of the Mons Sedimentary Basin chalk aquifer (Belgium). This strategic chalk aquifer provides approximately 50 million m^3 of groundwater each year for local distribution and for supplying drinking water to the city of Brussels. Approximately 76% of the official monitoring sites for this groundwater body present a nitrate concentration above the protection of ecosystems guideline threshold (25 mg/L; SPW and DEE 2022a; Gouvernement Wallon 2016), and 14% have a nitrate concentration value above the limit of 50 mg/L recommended by WHO (2017). In addition, some of the existing groundwater intakes are facing an increase in nitrate concentration, threatening their sustainability. The study area presents a strongly mixed environment with agricultural, residential, urban, and industrial land use areas, making the implementation of efficient mitigation measures difficult to set up. The methodology followed in this study combines hydrochemical data and isotopic analyses from 72 sampling sites, information about land use, geology and hydrogeology and multicriteria statistical analyses (PCA and t-SNE). Multicriteria statistical analyses are applied to identify samples likely to be impacted by the same context and pollution sources. The results of the $\delta^{15}N$, $\delta^{18}O$ and $\delta^{11}B$ stable isotope ratio measurements allow for the origin of the nitrate to be characterized, and for the possibility to make conclusions about the prevalence of denitrification reactions in the aquifer.

Methods and theoretical principles

The nitrogen cycle

The nitrogen cycle includes a set of transfers and interdependent chemical reactions (Fig. 1). This study is focused on the nitrate NO_3^- form in groundwater (N oxidation [+5]), considering (1) the multiple sources and chemical reactions leading to the formation of nitrate and (2) the denitrification reactions inducing the degradation of nitrate.

As described, for example, by Atteia (2015) and Abascal et al. (2022), Hiscock et al. (1991) and Rivett et al. (2008), regarding the contamination of groundwater by nitrate, the main process to be considered is the nitrification reaction, which transforms nitrogen in the form of ammonium (NH_4^+) into nitrate (NO_3^-)—see Eqs. (1) and (2) from Nikolenko et al. (2018). In this case, determining the source of nitrate is a matter of determining which sources of ammonium are involved. Multiple sources must be considered. Nitrate may directly originate from mineral N fertilizers spread on crops to promote plant development. Many fertilizers are also composed of already synthesized nitrate which, if not totally used by plants, percolates to the groundwater table (Di and Cameron 2002). Since the mineralization of organic matter generates ammonium, spreading of manure for agricultural production, decomposition of plant or household waste, leakage from sanitary sewage systems and industrial effluents are potential sources of nitrate contamination (Spalding and Exner 1993).

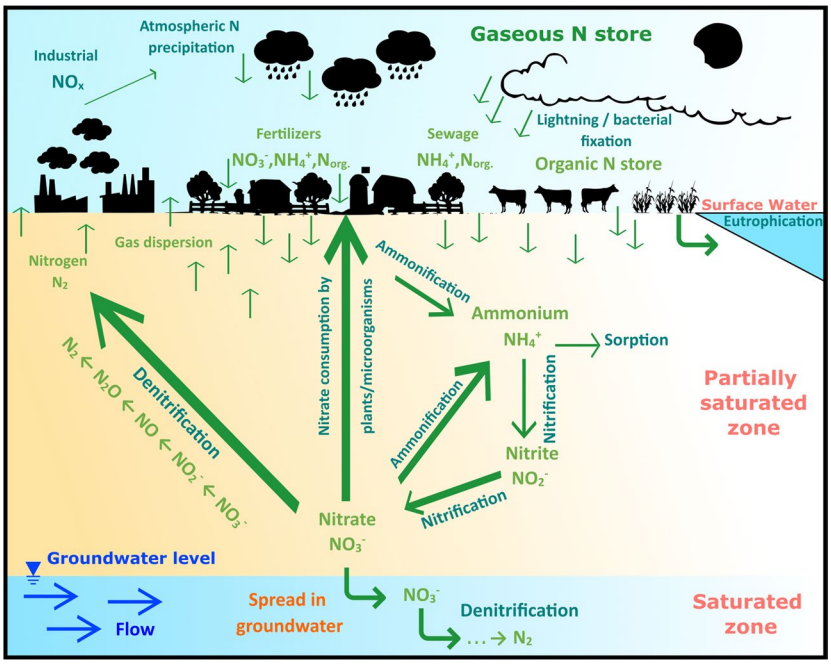

Fig. 1 Conceptual representation of the nitrogen cycle (based on Rivett et al. 2008)

$$NH_4^+ + \frac{3}{2}O_2 \rightarrow NO_2^- + H_2O + 2H^+ \quad (1)$$

$$NO_2^- + \frac{1}{2}O_2 \rightarrow NO_3^- \quad (2)$$

Nitrate degradation in groundwater mainly occurs through three processes. First, the 'reorganization' process consists of the conversion of the mineral nitrate into organic nitrogen by being consumed by plants or microorganism (Atteia 2015). Second, the 'dissimilar reduction of nitrate to ammonium' (DNRA - ammonification) is an anaerobic reduction reaction in which nitrate is converted to nitrite and then to ammonium ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$). Finally, the denitrification reactions include processes where nitrate (NO_3^-) acts as an electron acceptor in redox reactions that convert it into nitrogen gas—see the full reaction in Eq. (3) from Appelo and Postma (2004).

$$2NO_3^- + 6H^+ + 10e^- \rightarrow N_2 + 6H_2O \quad (3)$$

The various types of denitrification differ according to the element acting as an electron donor in the redox reactions. Several publications and reviews detail those processes (for example see Korom 1992; Postma et al. 1991; Rivett et al. 2008; Seitzinger et al. 2006). The most commonly considered denitrification reactions include (Appelo and Postma 2004):

Organic matter as an electron donor/heterotrophic denitrification

$$5CH_2O + 4NO_3^- \rightarrow 2N_2 + 4HCO_3^- + H_2CO_3 + 2H_2O \quad (4)$$

Pyrite and/or reduced iron as electron donor/autotrophic denitrification

$$5FeS_2 + 14NO_3^- + 4H^+ \rightarrow 7N_2 + 5Fe^{2+} + 10SO_4^{2-} + 2H_2O \quad (5)$$

And/or

$$5Fe^{2+} + NO_3^- + 7H_2O \rightarrow 5FeOOH + \frac{1}{2}N_2 + 9H^+ \quad (6)$$

Isotope ratios in nitrate

Analyses of the stable isotopes of nitrate and boron may be used to support the identification of nitrate sources. The methods applied in this study are described by Bronders et al. (2012), Kendall et al. (2007, 2015), Seiler (2005) and Widory et al. (2005, 2013).

The stable isotope abundance in a groundwater sample, δ_{sample} (in ‰) (Eq. 7), depends on the chemical processes occurring before sample collection and analysis. As an example, nitrate derived from the synthesis of mineral fertilizers presents a lower $^{15}N/^{14}N$ isotope ratio than nitrate derived from the degradation of manure. Considering the different isotopic ratios, it is then possible to attribute a specific signature to water samples and to interpret the origins of nitrate.

$$\delta_{sample} = \left[\frac{R_{sample} - R_{reference}}{R_{reference}}\right] \times 1000 \quad (7)$$

The nitrogen and boron minimum and maximum stable isotope abundances for different nitrate sources compiled

from previous studies (Aravena et al.) are listed in the following. These values should not be considered strict thresholds but can be used as a guide when interpreting isotope data. The $^{15}N/^{14}N - NO_3^-$ stable isotope ratio is typically used to discriminate mineral fertilizer sources (from –8 to +7‰) and organic sources such as manure (from +4.3 to +33.2‰) or sewage (from +4.3 to 14.3‰). The boron stable isotope ratio $^{11}B/^{10}B$ – B is not affected by the different processes that make up the nitrogen cycle. It can be used to discriminate the wastewater and manure sources in the aquifer (from –10 to +10‰ for sewage, from –2 to +14.8‰ for mineral fertilizer, and from +19.5 to +42.3‰ for manure). In addition to the source, chemical reactions in the nitrogen cycle influence the isotopic composition of nitrate. Nikolenko et al. (2018) present the evolution of the $\delta^{15}N–NO_3^-$ isotope ratio under nitrification (depletion, –5 to –35‰) and denitrification (enrichment, 5–40‰).

Multivariate analysis

The analysis of large datasets including many samples and variables can be performed using dimension reduction methods. Those methods aim to preserve the complexity and accuracy of the input dataset as much as possible, while at the same time, facilitating visualization and interpretation. In this study, the dataset consists of hydrochemical and isotopic analyses in different locations of a chalk aquifer. Two different dimensional reduction techniques are applied to the dataset: (1) classical "principal components analysis" (PCA), and (2) the "t-distributed stochastic neighbor embedding" (t-SNE).

Principal components analysis is a linear dimensional reduction method useful when the initial dataset contains correlated components. Redundant variables are replaced by synthetic, linearly independent variables, each of them maximizing variance in the whole dataset. PCA results are usually described using 2D representations, recalculated in the principal component space and used to identify groups including similarities. The theoretical and mathematical principles of PCA are presented and discussed by Rencher (2002). A case study showing the application of PCA to a geochemical dataset is presented by Gamble and Babbar-Sebens (2011) as an example.

The t-SNE was developed by van der Maaten and Hinton (2008) and is an evolution of the nonlinear stochastic neighbor embedding (SNE) dimensional reduction algorithm presented by Hinton and Roweis (2002). t-SNE aims to reduce an initial high-dimensional space dataset by replacing it with a new, completely redefined, low-dimensional space (usually 2D or 3D). The algorithm constructs a probability distribution (P1) for each element or sample pair (based on Euclidian distance) in the high-dimensional space so that similar samples are assigned a higher probability than others. A new 2D or 3D low-dimensional space is then created, and "duplicates" of the samples are positioned within it. A pairwise probability distribution (P2) is also determined in this subspace. Then, the arrangement of individuals in the 2D or 3D low-dimensional space is modified to minimize the difference between the probability distributions P1 and P2 as measured through the Kullback-Leibler divergence method. Compared to a classical PCA, t-SNE allows for the consideration of nonlinear relationships between variables and preserves local-scale similarities in the low-dimensional space. An application of the t-SNE algorithm to a geochemical dataset is, for example, presented by Liu et al. (2021).

The K-means algorithm, a classical clustering method described by Hartigan (1975), Hartigan and Wong (1979) and Bock (2007), aims to regroup several elements together in a way that minimizes the total distance between them and the centroids of their group. Combined with t-SNE dimensional reduction and general context knowledge, it simplifies the identification of subsets of similar samples. The t-SNE algorithm and K-means clustering are applied here using the Python scikit-learn package (Pedregosa et al. 2011).

Area of study, monitoring network and sampling campaigns

Geographical and hydrogeological contexts

The chalk aquifer of the Haine Valley is located in southern Belgium and covers 403 km^2 (Fig. 2a). The aquifer is mainly composed of Mesozoic chalk but also includes some Mesozoic calcarenite and tuffeau at the top of the series. The maximum thickness of the aquifer is up to 350 m in the central part of the studied area (Mengeot et al. 2017a). The base of the aquifer is made of very low permeability marls. Laterally and below the marls, less permeable layers are observed and consist of Upper Carboniferous (Namurian) pelites and sandstones to the north, Upper Carboniferous (Whestphalian) schisto-sandstone to the east, and Devonian quartzopelites to the southeast (Fig. 2a,b).

The aquifer is unconfined in the southern, northern and eastern areas. It is considered confined in the central part, and is covered by Cenozoic clay and sand layers. The confined part represents ~40% of the total area (Fig. 2a,b; Rorive and Goderniaux 2014).

The Haine River is the most important river, globally flowing from east to west in the central part of the aquifer area (Fig. 2b). The hydrogeological basin of the aquifer is included in the hydrographical basin of the rivers, which is slightly larger. Groundwater flows from the recharge zones, corresponding to the unconfined areas in the south, north and east of the basin, toward the center of the aquifer under the Cenozoic cover (Fig. 2a,b). Groundwater

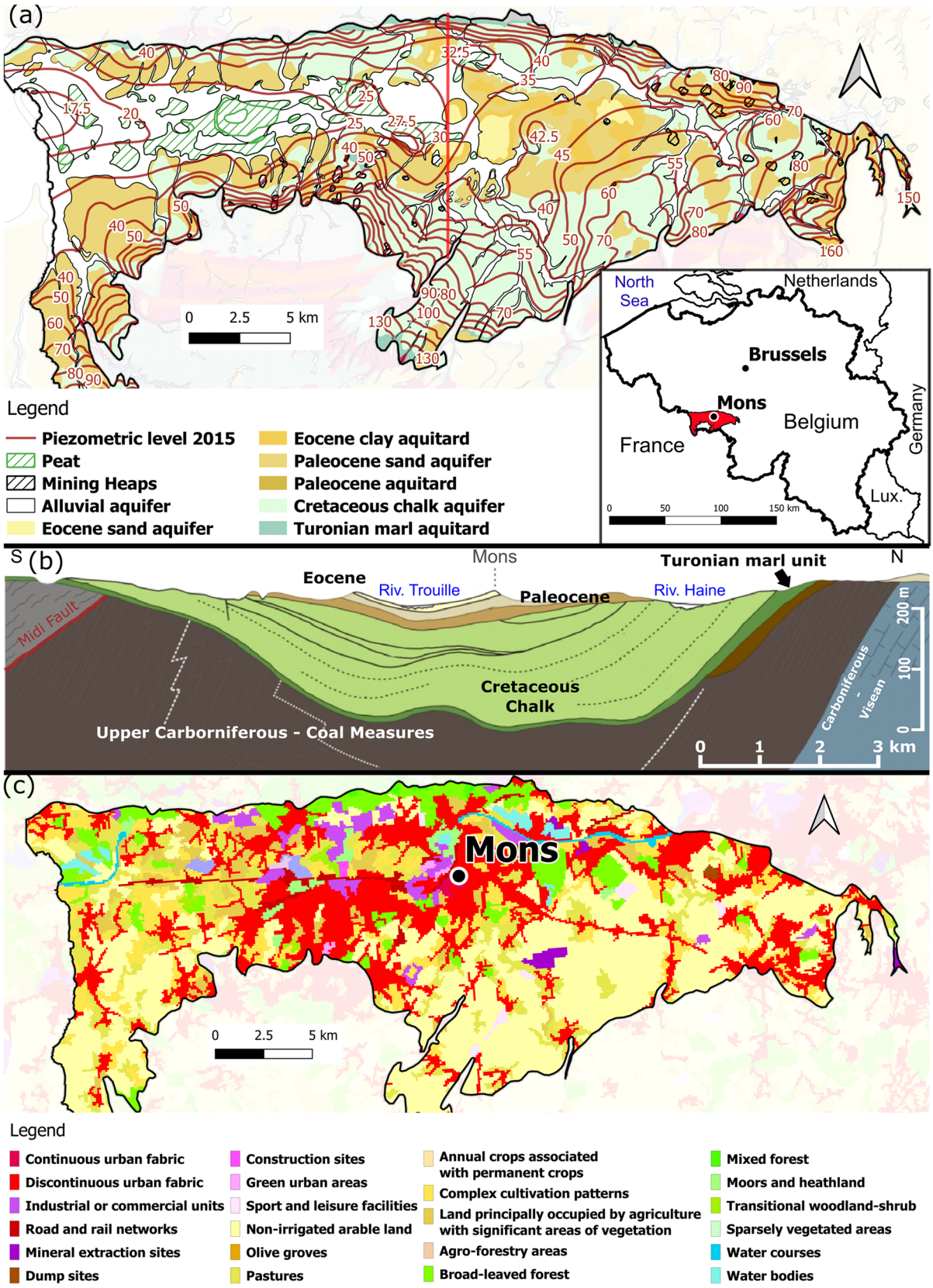

Fig. 2 **a** Hydrogeologic map (Bastien et al. 2017; Bougard et al. 2017; Habils et al. 2017a, b, 2018; Mengeot et al. 2017b) and 2015 piezometric level representation. **b** Hydrogeological vertical cross-section indicated by the red line (**a**) (modified from Pirson et al. 2008). **c** Land use distribution (from Corine Land Cover EEA 2018) and water production site locations. Coordinate system, EPSG: 31370

exfiltrates in the Haine River and some of its tributaries, where confining units are not present or characterized by a low thickness or a higher hydraulic conductivity.

The aquifer consists of an important and strategic reservoir of drinking water. Annually, approximately 50 million m^3 are abstracted (Rorive and Goderniaux 2014) to not only supply the local demand but also for exportation to neighboring regions and Brussels. Due to the subsidence induced by past mining exploitation, dewatering wells are permanently operated. Both classical production wells and dewatering facilities have a local impact on the piezometric levels.

The study area is characterized by a very heterogeneous land use distribution (Fig. 2c). The southern part, also corresponding to the unconfined zones, is largely devoted to open field agriculture. The central and eastern parts are significantly urbanized and densely populated. To the west, a greater proportion of grassland and mashes are observed.

Several local contamination sources are also identified within the studied area and must be considered when studying groundwater quality in the aquifer. Those contaminated sites include historical or current industrial activities (Wells et al. 2016), disused sites (SPW 2022; Thiernesse 1967), backfills related to past coal mine exploitation (SPW 2018), and the presence of landfills that may contain large quantities of organic matter and household waste as well as other possible pollutants (Navette et al. 2014, 2017; SPAQUE 2018).

Sampling locations and analyses

Two sampling campaigns were conducted in November 2020 and August 2021. Sampling sites were selected to be representative of the chalk aquifer within the whole area, considering agricultural and urban areas, as well as confined and unconfined zones of the aquifer. Additionally, samples from another concomitant project, focusing on a particularly polluted urban area, were also added (Balzani et al. 2022; Brouyère et al. 2022), thus making a total of 72 sampled sites. They include piezometers, water production wells, traditional masonry wells, springs and drainage galleries, inducing different investigation depths according to each specific point. The location, identifier and type of all sampling points are shown in Fig. 3.

A classical sampling procedure (ISSEP 2014) was applied. It includes a purge time systematically operated before sampling, with continuous in situ monitoring of the physico-chemical parameters (temperature, pH, redox potential, electrical conductivity and dissolved oxygen concentration). Samples were collected when those values were stable.

In situ physico-chemical parameters were measured using an HQ40D digital measuring instrument and appropriate probes (dissolved oxygen: IntelliCAL LDO10103 ±0.1 mg/L; acidity, pH: PHC10103 ±0.02; redox potential: MTC10103 ±0.02 mV/0.05%; electrical conductivity: CDC40101 ±0.5%; temperature: ±0.3 °C). Nitrate, ammonium and alkalinity were measured using flow injection analysis by the SWDE laboratory (Belgium). The quantification limits for nitrate and ammonium were 0.5 and 0.025 mg/L, respectively. Collected samples were kept refrigerated and analyzed within 24 h after collection. Total organic carbon was measured by the SWDE laboratory determining nonpurgeable organic carbon (ISO 8245). The quantification of ions was performed by the University

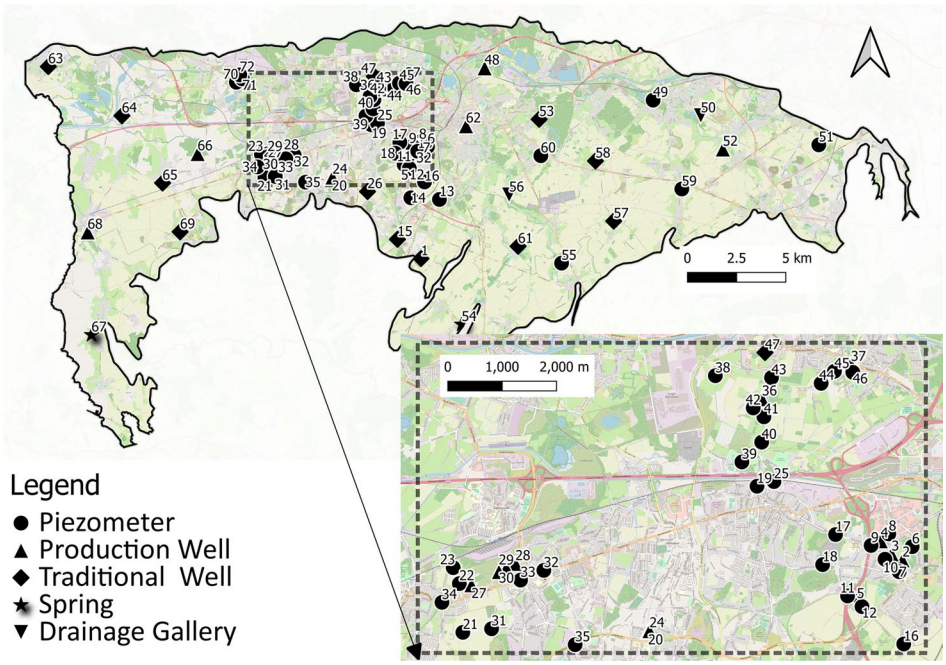

Fig. 3 Locations of the 72 sampling sites. Coordinate system, EPSG: 31370

of Liege (Belgium) using ion chromatography (Methrom – 850 Professional IC AnCat: K^+, Mg^{2+}, Na^+, Cl^-, SO_4^{2}, F^-, Sr^{2+}) or potentiometric titration (Methrom – Titrando 905: Ca^{2+}). The concentrations of dissolved iron, manganese and silica were determined by the University of Liege using atomic absorption (Analytik Jena AA400) on samples filtered at 0.45 µm and acidified in situ. Analysis of $\delta^{11}B$ was performed by VITO (Belgium) using high-resolution inductively coupled plasma mass spectrometry (HR-ICP-MS, ±3‰). Nitrate isotopic analyses were performed by the Helmholtz Centre for Environmental Research (Germany) using IRMS (isotope ratio mass spectrometer, nitrate denitrification and determination in nitrous oxide, ±1.5‰ $\delta^{18}O$, ±0,4‰ $\delta^{15}N$). Samples collected for nitrate isotopic analysis were filtered at 0.2 µm and kept frozen.

Results and source discrimination

Nitrate sources are discriminated through three interconnected steps:

- The interpretation of hydrochemical results (i.e., nitrate concentration, major elements) allows characterization of the magnitude of nitrate pollution, while identifying particular samples impacted by local pollution.
- The application of multicriteria statistical analysis (linear PCA, nonlinear t-SNE combined with K-means clustering) allows classification and clustering of the sampling sites and aquifer areas.
- The results of isotopic analyses on nitrate ($\delta^{15}N$, $\delta^{18}O$ – NO_3^-) and boron ($\delta^{11}B$) are used to refine and explain the classification, as well as to evidence denitrification processes.

A combination of hydrochemical data and isotope measurements is then employed to assess the occurrence of spontaneous denitrification at the interface between the confined and unconfined parts of the aquifer.

Discrimination based on major elements and nitrate concentrations

The first interpretation and discrimination are performed based on the chemical content of the groundwater samples regarding major elements, including nitrate. Figure 4 shows all samples represented in a Piper diagram (Domenico and Schwartz 1997). Considering the dataset as a whole, groundwater clearly exhibits a carbonated, calcic and magnesian character, as expected in a chalk aquifer. This Piper diagram and the specific composition of each groundwater sample are used here to identify 'outlier' sites affected by local contamination. Identified sites and related IDs are highlighted in white on the Piper diagram. Each of those sites is characterized by a singular composition compared to the dataset as a whole and can be related to well identified local contamination.

For the 72 analyzed samples, the nitrate concentration ranges between 0 and 150.0 mg/L. The mean and median concentrations are 35.09 and 31.75 mg/L, respectively, which exceeds the guide value of 25 mg/L (Gouvernement Wallon 2016; SPW and DEE 2022a). These results illustrate the general problem related to nitrate in the chalk aquifer, with 25% of the samples characterized by a nitrate

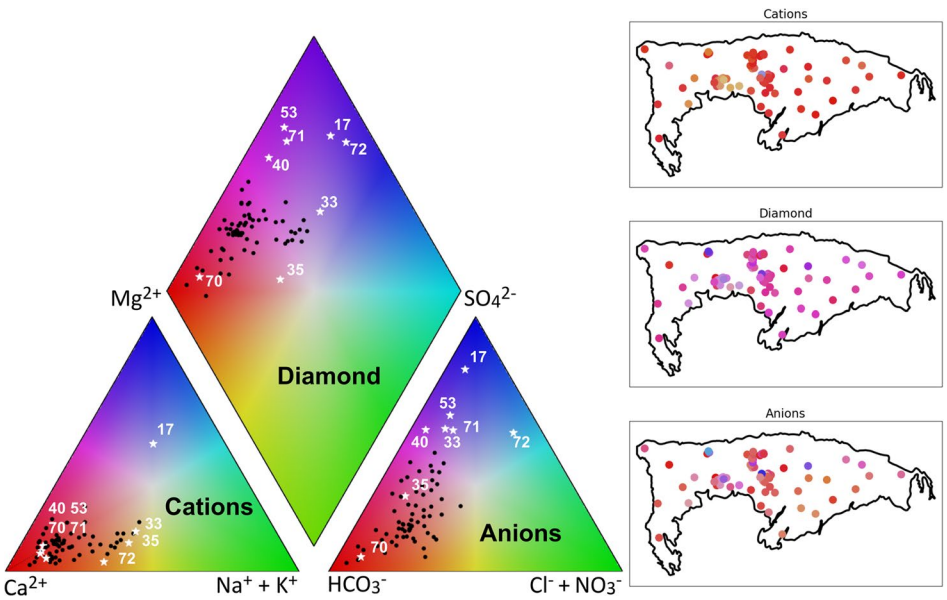

Fig. 4 Sampling sites represented in a Piper diagram. The background color distribution is based on Peeters (2013). White points correspond to samples considered outliers

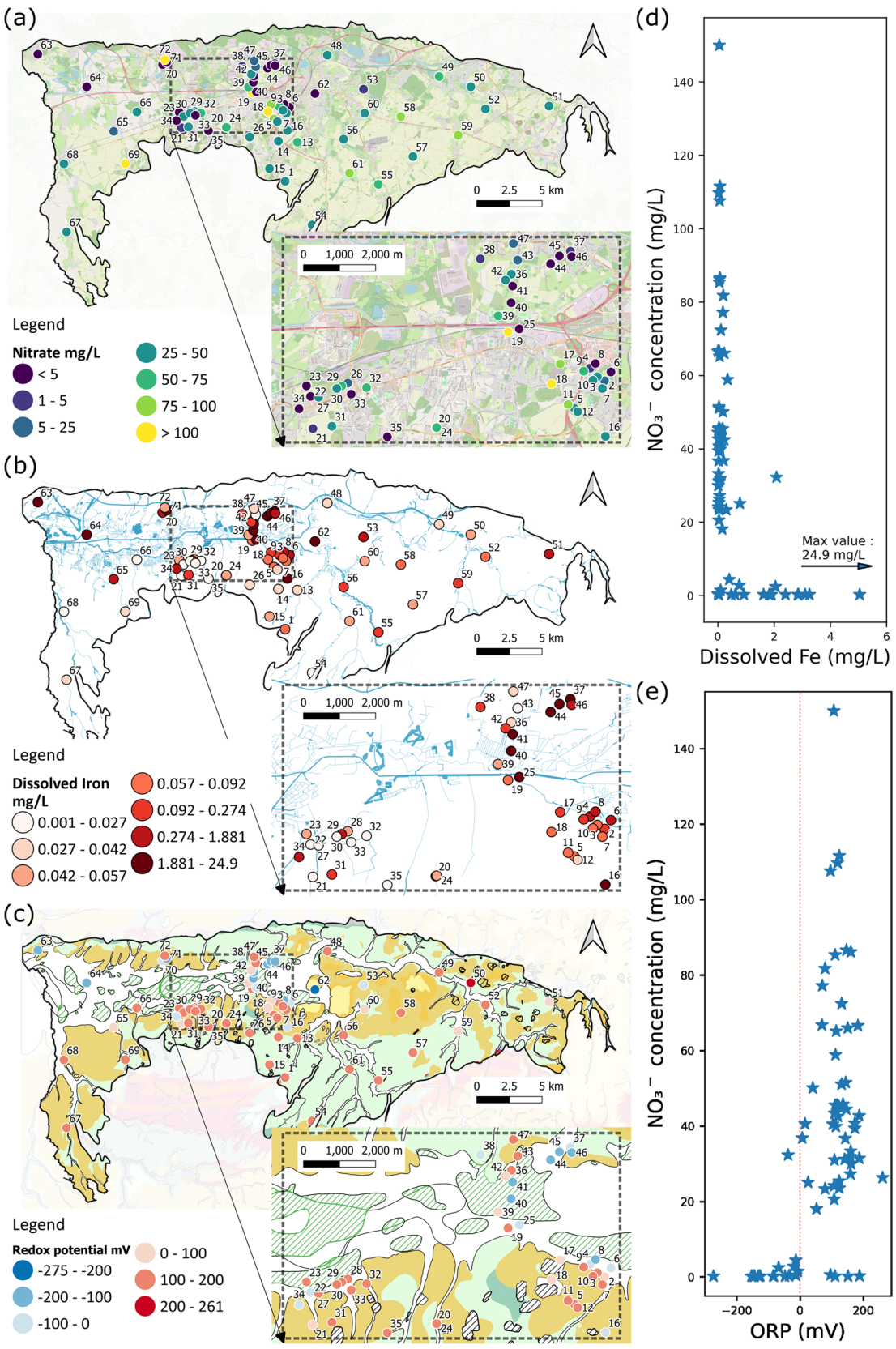

Fig. 5 **a** Nitrate concentration expressed in mg/L. The background corresponds to the OpenStreetMap raster. **b** Dissolved iron concentration expressed in mg/L, shown with the river network. **c** Redox potential expressed in mV, shown on the hydrogeological map. **d** Nitrate concentration vs. iron concentration. **e** Nitrate concentration vs. oxydo-reduction potential (ORP). Coordinate system, EPSG: 31370

concentration above the drinking-water standards (50 mg/L) as set by WHO (2017). High concentrations are mainly observed in agricultural areas to the southeast and west (Fig. 5a) and from sites affected by local pollution in urban or industrial areas (e.g., samples 17, 18, 19, 72). In contrast, approximately a quarter of the samples are characterized by a particularly low concentration (<2.73 mg/L). No nitrate was detected in 16 of them, making some of the isotopic measurements on these samples impossible. They mainly belong to the confined zones of the aquifer, to the west and in the center of the studied area, where land is generally used for meadows, woods, and urbanization (e.g., samples 6, 8, 63, 64).

Considering spatial distributions, bivariate graphs and correlation coefficients between hydrochemical variables, a clear relation between nitrate, dissolved iron (Fig. 5a,b,d) and redox potential (Fig. 5a,c,e) is observed. Looking at the dataset as a whole, the more reducing the conditions are, the higher the iron concentration, and the lower the nitrate concentration.

Based on these observations, a second subset of samples with low nitrate content is discriminated. Denitrification processes are highly suspected for those sites, favored by the presence in quantity of potential electron donors such as reduced iron or organic carbon, oxidizing nitrate in a reductive and anoxic environment corresponding to the confined zones of the aquifer. The land use of these areas, with fewer nitrogen intrants, may also contribute to explaining the absence of nitrate at those sites.

Discrimination based on multicriteria approaches (t-SNE, PCA, K-means)

Data preprocessing

The original dataset contains 72 samples, including 24 variables for each. One of the objectives of this study is to perform and compare interpretations using two different statistical methods (t-SNE and PCA). Preprocessing of the dataset is needed to guarantee an optimal and adequate functioning for each algorithm.

The subset of sampling sites identified as outliers or influenced by local pollution is removed from the whole dataset (see section 'Discrimination based on major elements and nitrate concentrations'). It includes the eight samples highlighted in white in the Piper diagram (Fig. 4) and it allows for the impact of extreme values on the results achieved from the applied statistical methods to be limited and further supports the identification of more regional trends and behavior.

All samples with nitrate concentrations that were too low (less than 0.5 mg/L) were removed from the whole dataset. The non- or low-detection of nitrate prevents any isotopic measurements on nitrate and keeping those samples (corresponding to the 13 samples, highlighted in dark blue in Fig. 5a) would therefore prevent those isotopic measurements from being incorporated as variables in the following multicriteria statistical analyses. Those samples globally correspond to confined and/or anoxic areas with highly suspected denitrification processes. In this study, the choice was made to keep isotopic measurements in the dataset rather than keeping all samples, as the study objective is to identify the origin of the nitrate pollution. Including those 13 samples and excluding nitrate isotopic measurements in the t-SNE dimensional reduction process has nevertheless been tested, but the results did not provide additional information. In this test with all samples, sampling sites without nitrate were separated into a distinct group, already identified without dimensional reduction, at the expense of a significantly less clear discrimination of the other samples.

Parameter values defined by the laboratory as below the detection or quantification thresholds are replaced by absolute quantities. Results indicated as "< detection threshold" are replaced by half of this threshold. Results indicated as "< quantification threshold" are replaced by the mean value of the two thresholds.

Finally, dataset values are transformed to limit the impact of data units. For PCA and the t-SNE algorithms, z-score standardization is used ($X' = [X\text{-}\mu]/\sigma$, where X' is the "new value", X is the measured value, μ is the mean value of a parameter, and σ is the standard deviation).

The resulting dataset used in the multicriteria analyses is a matrix (51×24) containing 51 sampling samples, each of which is characterized by 24 quantitative parameters (temperature, pH, dissolved O_2, electrical conductivity, redox potential, NO_3^-, NH_4^+, total organic carbon, alkalinity, Ca^{2+}, dissolved Fe, dissolved Mn, SiO_2, K^+, Mg^{2+}, Na^+, Cl^-, SO_4^{2-}, F^-, Sr^{2+}, B, $\delta\,^{11}B$, $\delta\,^{15}N$ and $\delta\,^{18}O - NO_3^-$).

t-SNE and PCA application

Figure 6 shows the results of the multicriteria statistical analyses. Figure 6a presents the sample groups obtained by combining dimensional reduction by t-SNE and K-means clustering. Figure 6b shows the 'within-cluster sum of squares' (WCSS) plotted as a function of the number of clusters considered. This graph helps identify an optimal number of clusters based on the semiempirical elbow method (Thorndike 1953). The optimal number of clusters corresponds to the "elbow" in the curve of this graph, which is the threshold at which the WCSS sharply increases. Figure 6c,d includes PCA results, with the circle of variables for the two first principal components and the graph of individuals, respectively. In Fig. 6d, colors and symbols of individuals relate to the groups determined using the t-SNE analysis. Finally, Fig. 6e shows the spatial distribution of the clusters

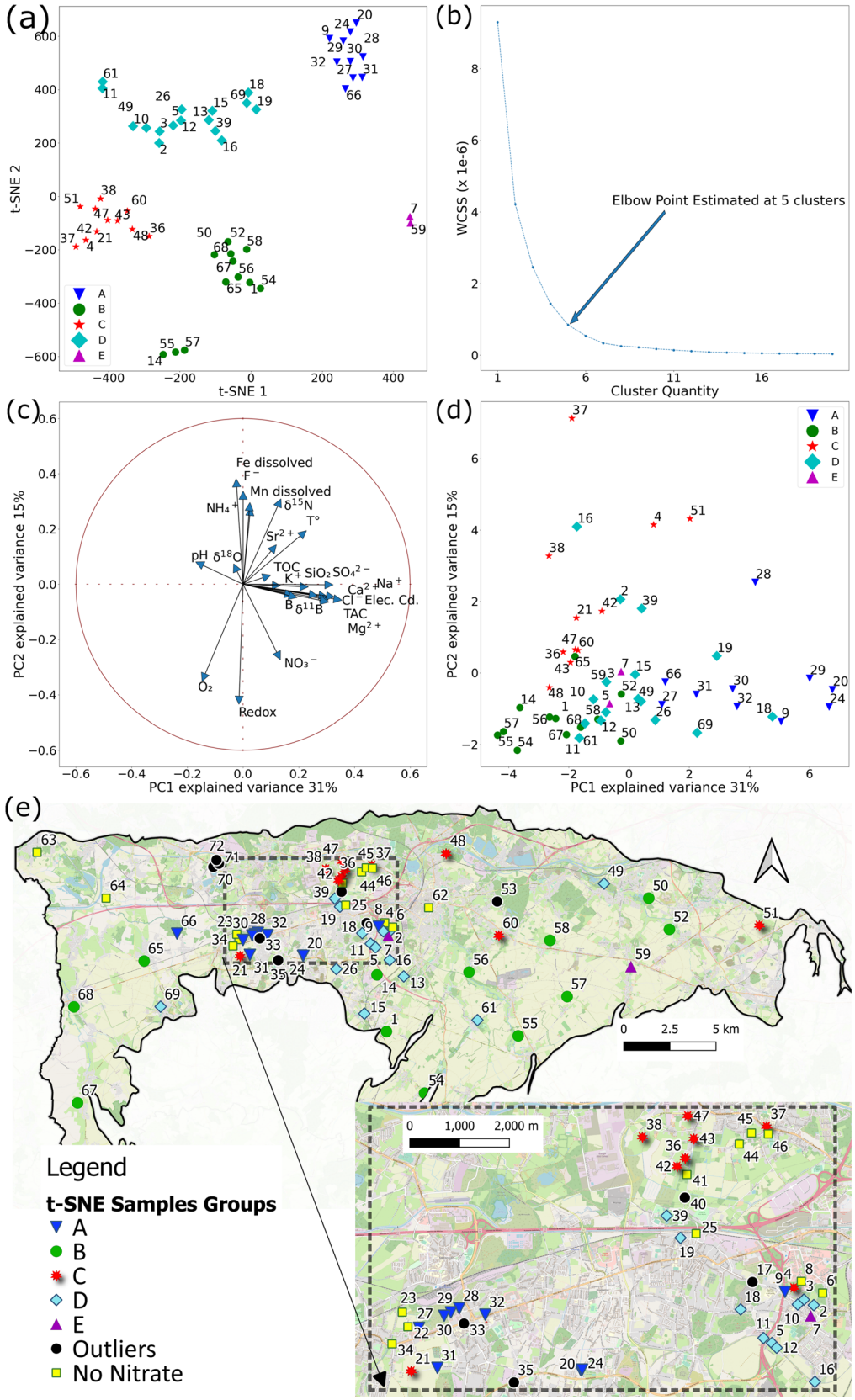

Fig. 6 **a** t-SNE reduction to a 2D space and clustering through K-means algorithm. **b** K-means optimal cluster number using the elbow method. **c** PCA variable circle for the first two components. **d** PCA graph of individuals for the first two components. **e** Spatial distribution of samples with the group colors attributed based on t-SNE results. Coordinate system, EPSG: 31370

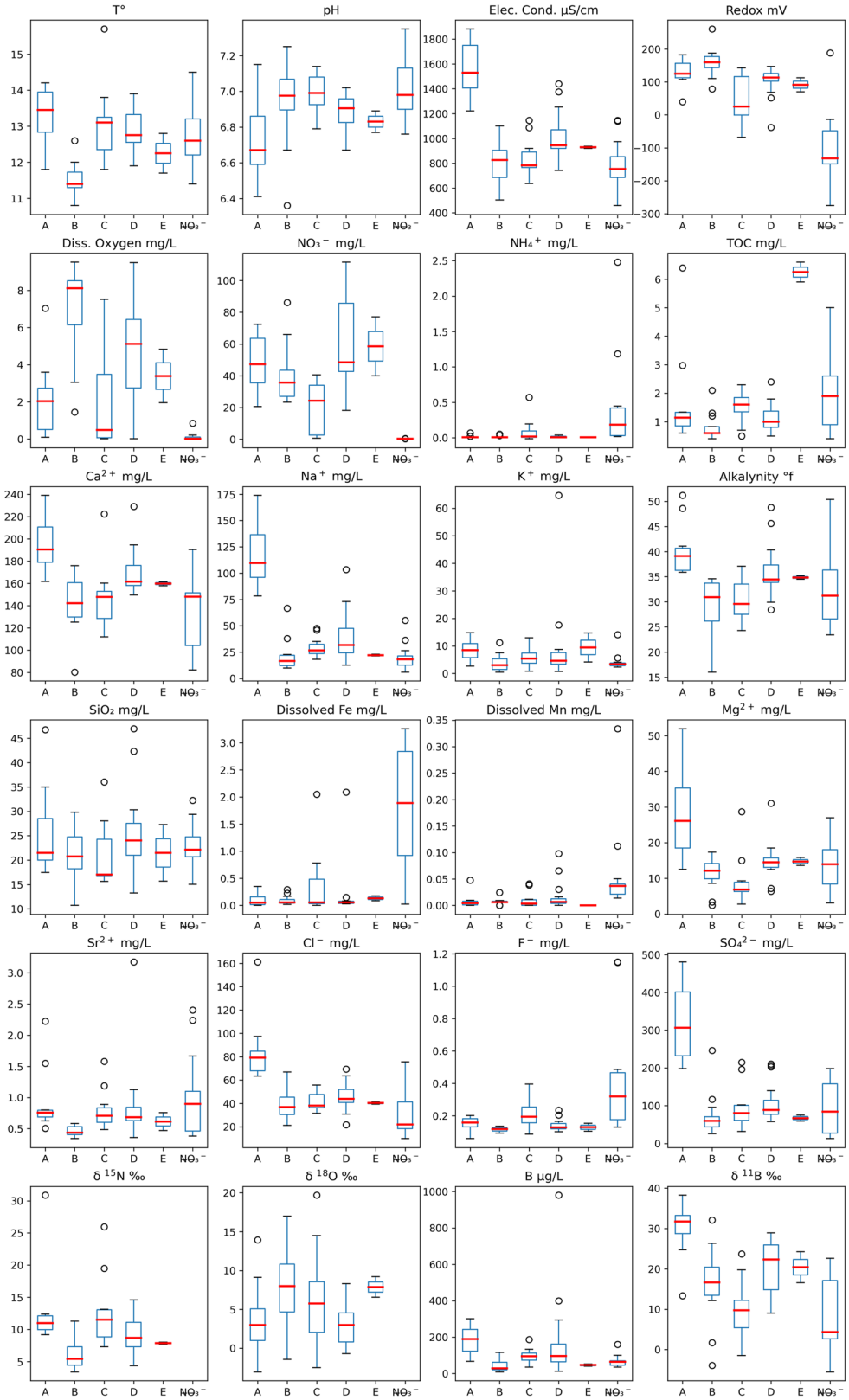

Fig. 7 Boxplots of t-SNE group statistical characteristics for each of the 24 parameters considered. Each boxplot column labelled as '~~NO3~~' corresponds to the group of samples characterized by the absence of nitrate. The red line corresponds to the median. The box extends between the first and third quartiles, and the whiskers extend between 1.5×[quartile 3 - quartile 1]. The black circles correspond to the extreme values that exceed the whiskers

identified following the dimensional reduction implemented by t-SNE.

The values set for the different parameters of the t-SNE and K-means algorithms are listed in the following. More information on the effect of each parameter is available in van der Maaten (2022) and in the sklearn.manifold.tsne documentation. Using the t-SNE algorithm, the 24 dimensions of the dataset are transposed into a 2D space. The initialization phase is based on the first two principal components (as opposed to a random initialization) to obtain a stable and reproducible result. The perplexity is a hyperparameter that influences the number of neighbors considered at each iteration. Although it was determined empirically, the classical values for this parameter are between $n/200$ and $n/10$ (n = number of samples); the algorithm for it was run with different values ([0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.50, 2.75, 3, 3.25, 3.50, 3.75, 4, 4.25, 4.50, 4.75, 5, 6, 7, 8, 9, 10]). The result obtained with a perplexity equal to 1.75 is presented. The early exaggeration is fixed to 100 to obtain a clear visualization of clusters. The conditional probabilities are based on Euclidean distances between neighbors, and the maximum number of iterations is arbitrarily set to 100,000. When applying the K-means clustering algorithm, the number of clusters is fixed at 5, as determined by the semiempirical elbow method (Fig. 6b). This method is not perfectly accurate, and the shape of the curve may suggest a number of clusters ranging from 4 to 6. However, combining this estimate with the general context of the study area, the number of 5 clusters seems appropriate. The other parameter values correspond to the default value given in the package documentation of sklearn.cluster.KMeans.

The 2D-transposed visualization of the samples by t-SNE allows for the identification of five groups of sites (Fig. 6a,e). This identification and related interpretations are performed based on the K-means clustering of the sampling sites and the hydrogeological context of the aquifer. A statistical overview of the values characterizing each group is shown in Fig. 7. The most important outcomes of each cluster are summarized in the following from a general perspective, integrating some possible singularities within the clusters.

- Group A (blue triangles) corresponds to the central-southern area located close to the limits between zones classified as urban or dedicated to livestock or agriculture (Fig. 2c). Land use is highly heterogeneous in this area, as confirmed by field observations. Samples from this area are generally characterized by a higher average electrical conductivity, higher temperature and higher acidity. This group also presents the highest mean concentration (excluding the outliers' group) of cations (calcium, magnesium, sodium, potassium), sulfate, boron, chlorine and mean $\delta\,^{15}N$ (+12.8‰) and $\delta\,^{11}B$ (+30.0‰). This group, which is probably impacted by different sources of pollution, also features a relatively high average nitrate concentration (48.4 mg/L).
- Group B (green circles) mostly includes samples in agricultural and unconfined areas located in the southeast and southwest areas. The average nitrate concentration is equal to 40.5 mg/L. Samples are characterized by the lowest mean $\delta\,^{15}N$ isotope ratio (mean = +6.3‰) and the highest mean $\delta\,^{18}O$ isotope ratio (mean = +7. ‰).
- Samples attributed to group C (red stars) are mainly located in the area northwest of the city of Mons, characterized by smaller agglomerations, pastures and cultivated lands. Three samples (21, 51 and 60) are nevertheless sporadically located in other zones of the aquifer. Except for the group without nitrate, group C is characterized by the lowest average nitrate concentration (20.5 mg/L), a low mean redox potential (+51.3 mV) and a high iron concentration (mean = 0.4 mg/L). From a hydrogeological point of view, in this group, groundwater originates from the northern part of the basin, and samples are taken at the border of the confined area.
- Group D (cyan diamonds) corresponds to samples mainly located southwest of the city of Mons, but also sporadically elsewhere. These sites are generally located in areas that combine a widely or locally urbanized character and locations downstream of agricultural areas. This group presents the highest average nitrate concentration (60.8 mg/L), which can be explained by the possible accumulation of different sources.
- Samples 7 and 59 form the separate cluster group E (purple triangles). Land use alone does not explain such segregation. Sample 7 is located in an urban area, whereas 59 is located near a water catchment on agricultural land. The variable that seems to cause this separation is the total organic carbon concentration, which is significantly higher for these two samples (5.9 and 6.6 mg/L).
- Samples that do not contain nitrate are characterized by a significantly lower redox potential (mean: –97.8 mV) linked to a low dissolved oxygen concentration (mean = 0.1 mg/L). These samples also have the lowest mean electrical conductivity (791.5 µS/cm) and the highest dissolved iron concentration (1.8 mg/L)

The application of the t-SNE algorithm to this dataset allows for a clear segregation of the samples (Fig. 6e) into different groups a priori exhibiting rather urban (group A), agricultural (group B) and mixed characteristics (groups D and C). Segregating using five clusters leads to the isolation of a group containing two unclassifiable samples (group E).

Correlations between variables are visible in Fig. 6c. Although those apparent correlations may be biased, because they are projected on the 2D space of the first two principal components, two groups of variables can be globally

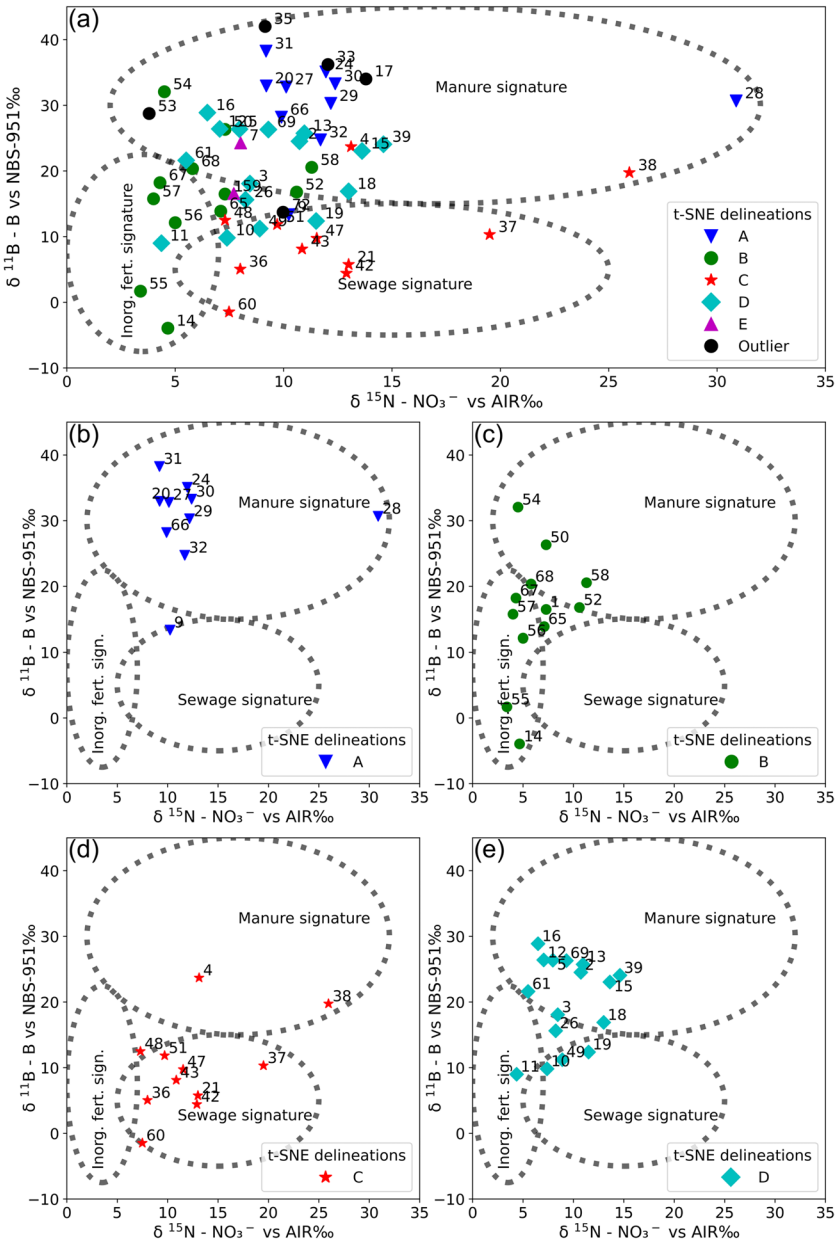

Fig. 8 Sample isotopic ratios and nitrate source signatures based on Widory et al. (2004). **a** Data for all samples. The other graphs show the samples for each group individually: **b** Samples gathered in group A by t-SNE dimensional reduction. **c** Samples gathered in group B by t-SNE dimensional reduction. **d** Samples gathered in group C by t-SNE dimensional reduction. **e** Samples gathered in group D by t-SNE dimensional reduction

distinguished. Vectors approximately parallel to the X axis (PC1) reflect the global mineralization of groundwater and general indicators of urban pollution (e.g., chloride, sulfate, electrical conductivity). Vectors globally oriented along the Y-axis (PC2) relate to reactions occurring within the nitrogen cycle, with elements and physico-chemical parameters involved and controlling denitrification processes (e.g., redox potential, iron ions concentration, see section 'The nitrogen cycle'). The vector related to nitrate concentration is in the fourth quadrant, showing an intermediate direction; therefore, nitrate concentration seems to be influenced by both controlling processes (mineralization and reactivity) in the chalk aquifer. Apparent correlations between variables are corroborated by bivariate graphs.

Figure 6d shows the PCA graph of individuals based on the first two principal components derived from the same dataset. For the sake of comparison, exhibited colors and symbols correspond to the groups previously identified using t-SNE. In the graph of Fig. 6d, the repartition of the samples according to their group remains consistent with the previous analyses. In particular, there is a clear repartition of samples along the X-axis, from typical agricultural to urban sampling sites at the extremes and mixed conditions in between; however, the clustering of samples is much less

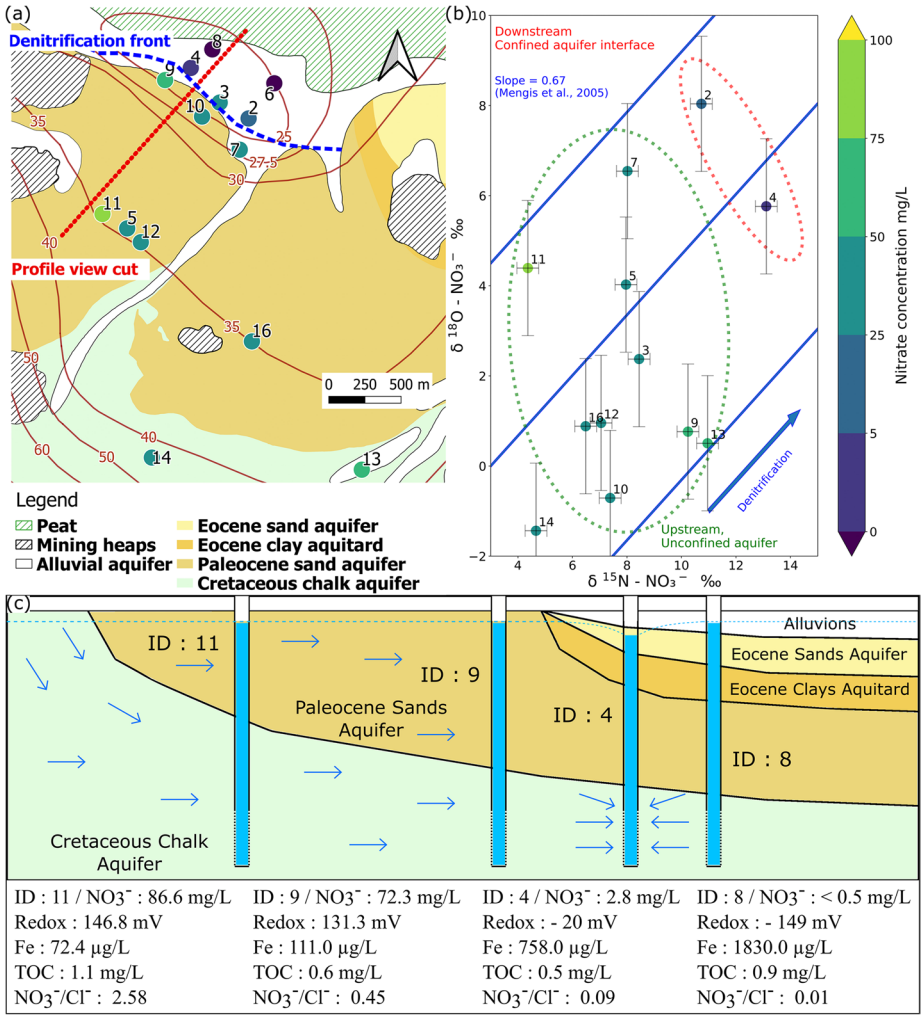

Fig. 9 Denitrification indicators in the area located southwest of the city of Mons (Belgium). **a** Nitrate concentration measured at each sampling point, hydrogeological background and iso-value piezometric lines. **b** Isotopic ratios ($\delta^{15}N$, $\delta^{18}O$) in nitrate samples. **c** Simplified conceptual profile view of the area between wells 11 and 8. NO_3^-, dissolved Fe, TOC and NO_3^-/Cl^- concentration values correspond to data obtained through sample analysis. Vertical and horizontal scales are not respected. Coordinate system, EPSG: 31370

obvious than using t-SNE. Partial superposition of groups in the PCA also demonstrates the difficulty of classifying some samples clearly characterized by a mix of different sources of pollution. However, it should be acknowledged that the two first principal components calculated for this dataset relate to a rather low explained variance (31 and 15%); therefore, those two components barely reflect only half of the whole dataset complexity. Using more principal components in the identification process may support and improve the discrimination of groups. Removing some of the extreme samples or parameters from the dataset could improve point segregation, but it would also reduce the total amount of information considered.

Discrimination based on isotope analysis

Previous statistical multicriteria analyses contributed to understanding general controlling factors explaining groundwater composition in the chalk aquifer. They also allowed to segregate the dataset into different clusters constructed according to the whole dataset including all hydrochemical variables. The analysis of specific isotope ratios allows further interpretation of nitrate.

Figure 8 shows the measured isotope ratios for each sample with a nitrate concentration higher than 0.5 mg/L. The symbols and colors correspond to the groups determined using the t-SNE algorithm (see section 'Discrimination based on multicriteria approaches (t-SNE, PCA, K-means)'). The signature domain delineations (dotted lines) are plotted according to Widory et al. (2004). Most of group A (central-southern located on the edge of an urban area with nested livestock and agriculture areas) and outliers present very high $\delta^{11}B$ and are located in the upper part of the manure signature domain. In accordance with land use, samples from group B (agricultural) exhibit signatures related to 'inorganic fertilizers', 'manure' or a mix of both. In the 'manure' domain, some samples correspond to groundwater collected in traditional wells within farmyards (ex. IDs 54, 58), possibly presenting a higher vulnerability to manure contamination. Group C (north urban area) presents a clear sewer leakage signature that coincides with a survey conducted

by one of the water producers active in the area (IDEA 2015), indicating that 19% of the residents living in the vicinity of the well ID 37 protection zone report the presence of a wastewater disposal pit unconnected to the sewer network on their property. Samples of group D, located in a periurban area downstream of agricultural zones, exhibit mixed isotopic signatures.

Denitrification markers

Within the dataset, many sampling sites present undetected or very low nitrate concentrations, although they are located in downstream parts of the aquifer. These sites are often located in areas with a low-permeability cover and confined conditions. Denitrification processes are highly suspected in those areas, as suggested by the increase in iron ions concentration and decrease in redox potential (see section 'Discrimination based on major elements and nitrate concentrations' and Fig. 5). The evolution of the isotopic ratios $\delta^{15}N$ and $\delta^{18}O - NO_3^-$ along streamlines is used here to demonstrate this hypothesis. As presented by a field study performed by Böttcher et al. (1990), the isotopes ^{15}N and ^{18}O are concurrently enriched in the residual nitrate pool during denitrification. The review of Chen and MacQuarrie (2005) reported fractionating ratios ($^{15}N/^{18}O$) ranging from 0.48 to 0.67.

The area located southwest of the city of Mons (Figs. 3 and 9a) is used for the demonstration. In this specific area, many sampling sites located upstream and downstream of the unconfined–confined transition zone are available (Fig. 9a). Points 2, 3 and 4 are dewatering and drinking-water production wells, while the other points are piezometers. Figure 9c shows a simplified profile of the area, with selected measured variables.

Along the streamlines, a significant drop in nitrate concentration is observed at the level of two of the three production wells, approximately located at the entrance of the confined area. While the concentration exceeds 50 mg/L upstream in the unconfined area, nitrate is not detected (<0.5 mg/L) in piezometers 6 and 8 in the confined zone. This sharp decrease in nitrate concentration, occurring close to the confined/unconfined interface, coincides with a decrease in redox potential (from +146.8 mV to –149 mV, Figs. 5b and 9b) and an increase in iron ions concentration (from 72.4 to 1,830.0 µg/L, Figs. 5c and 9b). In this subset of wells, the amount of total organic carbon (TOC) does not clearly exhibit similar trends or correlations with the nitrate concentration.

Figure 9b shows the isotopic ratios of the sites upstream and downstream of the assumed denitrification zone. Isotopic measurements were impossible for sites 6 and 8 due to nitrate concentrations that were too low. A clear average isotopic enrichment between the upstream and downstream areas is observed and supports the hypothesis of denitrification. This increase is comparable to the slope of the linear increase measured by Mengis et al. (2005), also plotted on the graph.

Figure 9c also shows the negative evolution of the NO_3^-/Cl^- ratio along the streamlines. This indicator is used by similar studies on nitrate contamination to differentiate the decrease in nitrate concentration due to mixing from that caused by denitrification reactions (Rezaei et al. 2017; Su et al. 2020; Vystavna et al. 2017). It is assumed that denitrification reactions do not affect the chloride concentration; therefore, a decrease in nitrate concentration caused by denitrification results in a decrease in the ratio between the two anions.

All aforementioned observations support the occurrence of denitrification processes in confined areas of the chalk aquifer. Autotrophic denitrification by sulfides (Eq. 5) or ferrous ions (Eq. 6) would apply here, as suggested by iron and sulfate concentrations (e.g., sulfate concentration: ID 4: 214.9 mg/L, ID 8: 158.8 mg/L; ID 9: 223.5 mg/L). Pyrite is found in some of the Cenozoic geological formations overlying the chalk aquifer, mainly in a clay layer (Orchie Member) located at the base of the Ypresian (Eocene) and containing pyrite and lignite. However, in the lithostratigraphic series, this member is not directly in contact with the chalk aquifer, separated by some medium- to low-permeability clayey sand. Additionally, the increase in sulfate, a product of the denitrification reaction (Eq. 5), is not observed in all suspected denitrification areas and appears more related to oxidation at the level of the coal mining waste heaps locally present throughout the catchment (see Figs. 3 and 9a). Higher dissolved iron concentrations may originate from the oxidation of pyrite or from other deposits containing iron minerals overlying the chalk aquifer in the central part of the catchment, including in the confined areas (Mengeot et al. 2017a). Denitrification by organic matter (Eq. 4) is also possible, although the total organic content (TOC) only slightly decreases along the streamlines from the unconfined to the confined areas (Fig. 9c).

Discussion

The methods used in this study allow conclusive results to be made. In particular, the t-SNE algorithm coupled with a K-means analysis enabled a clear clusterization of samples and locations. These clusters are representative of the whole dataset and are calculated considering all hydrochemical variables, and not only nitrogen components. In regards to clusterization and visual simplification, t-SNE shows much better performances compared to more classical methods such as PCA.

The clusters calculated using t-SNE globally correspond to well-delineated spatial areas, suggesting specific

hydrochemical characteristics, stresses and controlling factors that are representative of some specific areas over the catchment. While t-SNE is highly successful in such a clusterization, it must be complemented by other methods to understand the origin of those clusters. For this purpose, classical hydrochemical analyses and correlation analyses help identify relevant processes, including those explaining the fate of nitrate within the aquifer. Analysis of variance (Fig. 6c) clearly highlights the influence of mineralization processes probably related to various anthropic contaminations and the influence of denitrification. The existence of active denitrification processes in the aquifer is confirmed by isotopic analyses of $\delta^{15}N$, $\delta^{18}O - NO_3^-$ (Fig. 9b). Isotopic analyses also allow differentiation of the main drivers of nitrate in the different identified groups of samples. The most important insight of this study regarding isotopic analyses relies in the characterization and discrimination of nitrate sources in more urban environments. Isotopic analyses allowed for the clear identification of the zones where nitrate is predominantly explained by manure decomposition (group A), agricultural activities (group B), wastewater intrusion (group C) or by a mixed origin accumulated along groundwater streamlines (group D). Those insights could not be expected just based on land use or major element concentrations. In the group A samples, for example, the presence of other elements such as chloride or boron strongly suggests urban-related contamination of the groundwater; however, the results of this study indicate that nitrate pollution is related to the presence and use of manure in this area. Note that boron is also widely used in industry (e.g., glass production, detergents, fertilizers), making confident identification of the precise anthropogenic source difficult without additional information (Palmucci and Rusi 2014; Parks and Edwards 2005).

The methods used succeed in discriminating zones throughout the catchment and in providing qualitative information about groundwater pollution, particularly here, about the origin of nitrate. Nevertheless, those methods hardly provide quantitative information about the part of each nitrate source in the total content. As an illustration, samples to the north of the city of Mons (group C) are characterized by a clear 'wastewater' signature and medium nitrate concentrations, whereas south of Mons (group D), samples are characterized by a mixed signature, while nitrate concentrations are much higher. In this configuration, the absolute contribution of wastewater to nitrate pollution could hypothetically be similar in both zones but would relatively decrease in the associated with group D because of a concurrence with an agricultural contribution. Those quantitative aspects can possibly be caught through the construction of Bayesian isotope mixing models such as SIAR (Duan et al. 2022; Li et al. 2019; Meghdadi and Javar 2018) or MixSIAR (Kaown et al. 2023; Torres-Martínez et al. 2020). Such models incorporate quantitative aspects by assessing the percentage contribution of specific sources within the sample set. However, as highlighted by Xu et al. (2016), the reliability of this approach is highly dependent on the isotopic values associated with the sources under consideration and the different fractionation processes, such as denitrification, occurring in the aquifer. In the study area, the expected sources of nitrate are difficult to characterize exhaustively, and their heterogeneity is probably high. Considering these difficulties and uncertainties, the authors think that applying those isotope mixing models would not bring a reliable quantification of the nitrate sources' contributions.

Conclusion

The approach followed and described in this paper remains qualitative but combines tools, data, and aquifer understanding, which brings robustness to the conclusions. A relatively recent dimensional reduction method (t-SNE) is applied to a hydrochemical and isotopic dataset. This approach succeeded in classifying samples with respect to the whole dataset more efficiently than other commonly used methods. A dominant nitrate source is then related to the different groups based on the isotopic data, and areas more likely to be impacted by an excess of inorganic fertilizer, manure or sewage are identified. The methodology applied here is based on a comprehensive hydrochemical dataset, hydrogeological knowledge, land use, and not only on isotopes. It allows for the understanding and consideration of the physical processes occurring in the aquifer and it efficiently leads to qualitative but robust indications related to the dominant nitrate sources, which provide useful insights for implementing mitigation measures. Thus, this combination constitutes an efficient tool for groundwater resource management in the context of nitrate contamination. In the case of the chalk aquifer of the Mons Basin, it clearly allowed for the identification of the different sources of pollution and for their spatial distribution.

The new insights from this study provide important clues for mitigating the nitrate concentration in groundwater. The obtained results show the relevant mitigation drivers according to the different zones of the catchment and drinking-water production facilities. As an example, the study highlights the zones where improving wastewater collection, either by repairing old sewage system networks or removing direct infiltration through soak pits, is expected to significantly decrease the nitrate contamination in groundwater. The same applies to any agricultural practice that would contribute to reducing nitrogen intrants in the south and southwest areas of the aquifer. The question still arises about the exact impact of measures on each specific driver and the timing required for observing visible effects. Those questions can, for example, be addressed by aquifer modeling, provided that nitrate intrants and transfers are well quantified, which remains challenging.

Acknowledgements We thank the water production companies IDEA, SWDE, Vivaqua and Farys for sharing some data and making some facilities available in the framework of this project.

Funding information This research was financially supported by SPGE (Société Publique de Gestion de l'Eau – Public Water Management Company), which is a public limited company set up by the Walloon Region. Some of the data collected are linked to the CASPER project, which is also funded by the SPGE (Brouyère et al. 2022).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abascal E, Gómez-Coma L, Ortiz I, Ortiz A (2022) Global diagnosis of nitrate pollution in groundwater and review of removal technologies. Sci Total Environ 810:152233. https://doi.org/10.1016/j.scitotenv.2021.152233

Accoe F, Berglund M, Duta S, Hennessy C, Taylor P, Van Hoof K, De Smedt S (2008) Source apportionment of nitrate pollution in surface water using stable isotopes of N and O in nitrate and B: a case study in Flanders (Belgium). JRC Scientific and Technical Reports, Joint Research Centre, Geel, Belgium, 25 pp

Appelo, CAJ, Postma D (2004) Redox processes. In: Geochemistry, groundwater and pollution. CRC, Boca Raton, FL, pp 415–487

Aravena R, Evans M, Cherry J (1993) Stable isotopes of oxygen and nitrogen in source identification of nitrate from septic systems. Groundwater 31:180–186

Atteia O (2015) Chimie et pollution des eaux souterraines [Groundwater chemistry and pollution]. Tec & Doc. Lavoisier, Paris

Balzani L, Orban P, Brouyère S (2022) Protection of peri-urban groundwater catchments: a multi-tracer approach for the identification of urban pollution sources. Presented at the EGU General Assembly 2022, Vienna, May 2022

Bassett R, Buszka P, Davidson G, Chong-Diaz D (1995) Identification of groundwater solute sources using boron isotopic composition. Environ Sci Technol 29:2915–2922. https://doi.org/10.1021/es00012a005

Bastien J, Roland S, Rorive A (2017) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Planchettes Binche – Morlanwelz no. 46/5-6, SPW, Wallonia, Belgium

Biddau R, Cidu R, Da Pelo S, Carletti A, Ghiglieri G, Pittalis D (2019) Source and fate of nitrate in contaminated groundwater systems: assessing spatial and temporal variations by hydrogeochemistry and multiple stable isotope tools. Sci Total Environ 647:1121–1136

Bock H-H (2007) Clustering methods: a history of k-means algorithms. In: Brito P, Cucumel G, Bertrand P, de Carvalho F (eds) Selected contributions in data analysis and classification. Springer, Heidelberg, Germany, pp 161–172. https://doi.org/10.1007/978-3-540-73560-1_15

Böhlke JK, Denver JM (1995) Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic Coastal Plain, Maryland. Water Resour Res 31:2319–2339. https://doi.org/10.1029/95WR01584

Böttcher J, Strebel O, Voerkelius S, Schmidt H-L (1990) Using isotope fractionation of nitrate-nitrogen and nitrate-oxygen for evaluation of microbial denitrification in a sandy aquifer. J Hydrol 114:413–424. https://doi.org/10.1016/0022-1694(90)90068-9

Bougard G, Roland S, Rorive A (2017) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Planchettes Quiévrain – Saint-Ghislain no. 45/5-6, SPW, Wallonia, Belgium

Bronders J, Tirez K, Desmet N, Widory D, Petelet-Giraud E, Bregnot A, Boeckx P (2012) Use of compound-specific nitrogen ($d^{15}N$), oxygen ($d^{18}O$), and bulk boron ($d^{11}B$) isotope ratios to identify sources of nitrate-contaminated waters: a guideline to identify polluters. Environ Forensics 13:32–38. https://doi.org/10.1080/15275922.2011.643338

Brouyère S, Balzani L, Orban P (2022) The CASPER project: an integrated approach for pollution risk assessment in peri-urban groundwater catchment areas. Copernicus Meetings. https://doi.org/10.5194/adgeo-59-45-2022

Canter LW (2019) Nitrates in groundwater. Routledge, Abingdon-on-Thames, UK

Chen DJZ, MacQuarrie KTB (2005) Correlation of $\delta^{15}N$ and $\delta^{18}O$ in NO_3^- during denitrification in groundwater. J Environ Eng Sci 4:221–226. https://doi.org/10.1139/S05-002

Czekaj J, Jakóbczyk-Karpierz S, Rubin H, Sitek S, Witkowski AJ (2016) Identification of nitrate sources in groundwater and potential impact on drinking water reservoir (Goczałkowice Reservoir, Poland). Phys Chem Earth Parts ABC 94:35–46

Di HJ, Cameron KC (2002) Nitrate leaching in temperate agroecosystems: sources, factors and mitigating strategies. Nutr Cycl Agroecosyst 64:237–256. https://doi.org/10.1023/A:1021471531188

Domenico PA, Schwartz FW (1997) Physical and chemical hydrogeology, 2nd edn. Wiley, New York

Duan L, Wu Y, Fan J, Ye F, Xie C, Fu X, Sun Y (2022) Identification of nitrogen pollution sources and transport transformation processes in groundwater of different landforms using C, H, N, and O isotope techniques: an example from the lower Weihe River. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-022-24337-2

EEA C (2018) Corine land cover (CLC) 2018, version 2020_20u1. https://land.copernicus.eu/pan-european/corine-land-cover/clc2018. Accessed May 2023

Fan AM, Steinberg VE (1996) Health implications of nitrate and nitrite in drinking water: an update on methemoglobinemia occurrence and reproductive and developmental toxicity. Regul Toxicol Pharmacol 23:35–43. https://doi.org/10.1006/rtph.1996.0006

Gamble A, Babbar-Sebens M (2011) On the use of multivariate statistical methods for combining in-stream monitoring data and spatial analysis to characterize water quality conditions in the White River Basin, Indiana, USA. Environ Monit Assess 184:845–875. https://doi.org/10.1007/s10661-011-2005-y

Gouvernement Wallon (2016) Code de l'Eau [Water Code], Annexe XIV, [A.G.W. 03.05.2007] [A.G.W. 12.02.2009], Evaluation de la qualité des masses d'eay souterraine [Evaluation of the quality of undergound water]. Gouvernement Wallon, Wallonia, Belgium

Habils F, Roland S, Rorive A (2017a) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Planchettes Beloeil – Baudour no. 45/1-2, SPW, Wallonia, Belgium

Habils F, Roland S, Rorive A (2017b) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Planchettes Jurbise – Obourg no. 45/3-4, SPW, Wallonia, Belgium

Habils F, Roland S, Rorive A (2018) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Planchettes Le Roeulx – Seneffe no. 46/1-2, SPW, Wallonia, Belgium

Hartigan JA (1975) Clustering algorithms, 99th edn. Wiley, Chichester, UK

Hartigan JA, Wong MA (1979) A K-means clustering algorithm. J R Stat Soc Ser C Appl Stat 28:100–108. https://doi.org/10.2307/2346830

Heaton TH (1986) Isotopic studies of nitrogen pollution in the hydrosphere and atmosphere: a review. Chem Geol Isot Geosci Sect 59:87–102

Hinton GE, Roweis S (2002) Stochastic neighbor embedding. In: Becker S, Thrun S, Obermayer K (eds) Advances in neural information processing systems. MIT Press, Cambridge, MA

Hiscock K, Lloyd J, Lerner D (1991) Review of natural and artificial denitrification of groundwater. Water Res 25:1099–1111

IDEA (2015) CHYDRO-004 – Délimitation des zones de prévention du puits BRASSICO à GHLIN (MONS) [Delimitation of the protection zones of the BRASSICO well in GHLIN (Mons)]. SPW, Wallonia, Belgium

ISSEP (2014) Compendium Wallon d'échantillonnage et d'analyses, méthode de prélèvement des eaux souterraines dans les aquifères non superficiels [Walloon compendium of sampling and analysis methods]. ISSEP, Liege, Belgium

Jackson B, Browne C, Butler A, Peach D, Wade AJ, Wheater H (2008) Nitrate transport in Chalk catchments: monitoring, modelling and policy implications. Environ Sci Policy 11:125–135

Kaown D, Koh D-C, Mayer B, Mahlknecht J, Ju Y, Rhee S-K, Kim J-H, Park DK, Park I, Lee H-L, Yoon Y-Y, Lee K-K (2023) Estimation of nutrient sources and fate in groundwater near a large weir-regulated river using multiple isotopes and microbial signatures. J Hazard Mater 446:130703. https://doi.org/10.1016/j.jhazmat.2022.130703

Kendall C, Elliott EM, Wankel SD (2007) Tracing anthropogenic inputs of nitrogen to ecosystems. Stable Isot Ecol Environ Sci 2:375–449

Kendall C, Young MB, Silva SR, Kraus TEC, Peek S, Guerin M (2015) Tracing nutrient and organic matter sources and biogeochemical processes in the Sacramento River and Northern Delta: proof of concept using stable isotope data. US Geol Surv Data Release. http://dx.doi.org/10.5066/F7QJ7FCM. Accessed May 2023

Komor S (1997) Boron contents and isotopic compositions of hog manure, selected fertilizers, and water in Minnesota (no. 0047–2425). J Environ Qual. https://doi.org/10.2134/jeq1997.00472425002600050004x. Accessed May 2023

Korom SF (1992) Natural denitrification in the saturated zone: a review. Water Resour Res 28:1657–1668. https://doi.org/10.1029/92WR00252

Kruk M, Mayer B, Nightingale M, Laceby J (2020) Tracing nitrate sources with a combined isotope approach (δ15NNO3, δ18ONO3 and δ11B) in a large mixed-use watershed in southern Alberta, Canada. Sci Total Environ 703:135043

Li C, Li S-L, Yue F-J, Liu J, Zhong J, Yan Z-F, Zhang R-C, Wang Z-J, Xu S (2019) Identification of sources and transformations of nitrate in the Xijiang River using nitrate isotopes and Bayesian model. Sci Total Environ 646:801–810. https://doi.org/10.1016/j.scitotenv.2018.07.345

Liu H, Yang J, Ye M, James SC, Tang Z, Dong J, Xing T (2021) Using t-distributed Stochastic Neighbor Embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data. J Hydrol 597:126146. https://doi.org/10.1016/j.jhydrol.2021.126146

Mariotti A, Landreau A, Simon B (1988) 15N isotope biogeochemistry and natural denitrification process in groundwater: application to the chalk aquifer of northern France. Geochim Cosmochim Acta 52:1869–1878. https://doi.org/10.1016/0016-7037(88)90010-5

Martinelli G, Dadomo A, De Luca DA, Mazzola M, Lasagna M, Pennisi M, Pilla G, Sacchi E, Saccon P (2018) Nitrate sources, accumulation and reduction in groundwater from northern Italy: insights provided by a nitrate and boron isotopic database. Appl Geochem 91:23–35. https://doi.org/10.1016/j.apgeochem.2018.01.011

Meghdadi A, Javar N (2018) Quantification of spatial and seasonal variations in the proportional contribution of nitrate sources using a multi-isotope approach and Bayesian isotope mixing model. Environ Pollut 235:207–222. https://doi.org/10.1016/j.envpol.2017.12.078

Mengeot A, Roland S, Rorive A (2017a) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Mons - Givry, 45 7/8, Notice explicative, SPW, Wallonia, Belgium

Mengeot A, Roland S, Rorive A (2017b) Carte hydrogéologique de Wallonie [Hydrogeological map of Wallonia]. Planchettes Mons – Givry no. 45/7-8, SPW, Wallonia, Belgium

Mengis M, Schif S, Harris M, English M, Aravena R, Elgood R, Maclean A (2005) Multiple geochemical and isotopic approaches for assessing ground water NO_3^- elimination in a riparian zone. Ground Water 37:448–457. https://doi.org/10.1111/j.1745-6584.1999.tb01124.x

Mitchell RJ, Babcock RS, Gelinas S, Nanus L, Stasney DE (2003) Nitrate distributions and source identification in the Abbotsford–Sumas aquifer, northwestern Washington State. J Environ Qual 32:789–800

Mohamed MA, Terao H, Suzuki R, Babiker IS, Ohta K, Kaori K, Kato K (2003) Natural denitrification in the Kakamigahara groundwater basin, Gifu prefecture, central Japan. Sci Total Environ 307:191–201

Navette E, Bietlot E, Collart C (2014) Réseau de contrôle des C.E.T. en région Wallone C.E.T. de Cronfestu Quatrième campagne de contrôle [Monitoring network for landfills in the Walloon region]. No. 1370/2014. SPW, Wallonia, Belgium

Navette E, Bietlot E, Collart C (2017) Evaluation de la situation environnementale des eaux souterraines 2017 C.E.T. de classe 3 "La Morette" à Flénu [Assessment of the environmental situation of the groundwater 2017 class 3 landfill "La Morette"]. No. 1851/2017, ISSEP, Liège, Belgium

Nikolenko O, Jurado A, Borges AV, Knöller K, Brouyère S (2018) Isotopic composition of nitrogen species in groundwater under agricultural areas: a review. Sci Total Environ 621:1415–1432. https://doi.org/10.1016/j.scitotenv.2017.10.086

Obeidat M, Awawdeh M, Al-Kharabsheh N, Al-Ajlouni A (2021) Source identification of nitrate in the upper aquifer system of the Wadi Shueib catchment area in Jordan based on stable isotope composition. J Arid Land 13:350–374

Ogrinc N, Tamše S, Zavadlav S, Vrzel J, Jin L (2019) Evaluation of geochemical processes and nitrate pollution sources at the Ljubljansko polje aquifer (Slovenia): a stable isotope perspective. Sci Total Environ 646:1588–1600. https://doi.org/10.1016/j.scitotenv.2018.07.245

Orban P, Brouyère S, Batlle-Aguilar J, Couturier J, Goderniaux P, Leroy M, Maloszewski P, Dassargues A (2010) Regional transport modelling for nitrate trend assessment and forecasting in a chalk aquifer. J Contam Hydrol 118:79–93

Oren O, Yechieli Y, Böhlke J, Dody A (2004) Contamination of groundwater under cultivated fields in an arid environment, central Arava Valley, Israel. J Hydrol 290:312–328

Palmucci W, Rusi S (2014) Boron-rich groundwater in central eastern Italy: a hydrogeochemical and statistical approach to define origin and distribution. Environ Earth Sci 72:5139–5157. https://doi.org/10.1007/s12665-014-3384-5

Parks JL, Edwards M (2005) Boron in the environment. Crit Rev Environ Sci Technol 35:81–114

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Peeters L (2013) A background color scheme for piper plots to spatially visualize hydrochemical patterns. Ground Water 52. https://doi.org/10.1111/gwat.12118

Pirson S, Spagna P, Baele J-M, Damblon F, Gerrienne P, Vanbrabant Y, Yans J (2008) An overview of the geology of Belgium. Presented at the Memoirs of the Geological Survey of Belgium, Geological Survey of Belgium, Brussels, pp 5–26

Pittalis D, Carrey R, Da Pelo S, Carletti A, Biddau R, Cidu R, Celico F, Soler A, Ghiglieri G (2018) Hydrogeological and multi-isotopic approach to define nitrate pollution and denitrification processes in a

coastal aquifer (Sardinia, Italy). Hydrogeol J 26:2021–2040. https://doi.org/10.1007/s10040-018-1720-7
Postma D, Boesen C, Kristiansen H, Larsen F (1991) Nitrate reduction in an unconfined sandy aquifer: water chemistry, reduction processes, and geochemical modeling. Water Resour Res 27:2027–2045. https://doi.org/10.1029/91WR00989
Rencher AC (2002) Methods of multivariate analysis, 2nd edn. Wiley, New York
Rezaei M, Nikbakht M, Shakeri A (2017) Geochemistry and sources of fluoride and nitrate contamination of groundwater in Lar area, south Iran. Environ Sci Pollut Res 24:15471–15487. https://doi.org/10.1007/s11356-017-9108-0
Rivett MO, Buss SR, Morgan P, Smith JWN, Bemment CD (2008) Nitrate attenuation in groundwater: a review of biogeochemical controlling processes. Water Res 42:4215–4232. https://doi.org/10.1016/j.watres.2008.07.020
Ronen D, Magaritz M (1985) High concentration of solutes at the upper part of the saturated zone (water table) of a deep aquifer under sewage-irrigated land. J Hydrol 80:311–323
Rorive A, Goderniaux P (2014) L'aquifère du Crétacé de la vallée de la Haine [The Haine Valley Cretaceous aquifer]. In: Dassargues A, Walraevens K (eds) Watervoerende Lagen En Grondwater in België/Aquiferes et Eaux Souterraines En Belgique [Aquifers and groundwater in Belgium]. Academia Press, Gent, Belgium
Sandor J, Kiss I, Farkas O, Ember I (2001) Association between gastric cancer mortality and nitrate content of drinking water: ecological study on small area inequalities. Eur J Epidemiol 17:443–447
Seiler R (2005) Combined use of 15N and 18O of nitrate and 11B to evaluate nitrate contamination in groundwater. Appl Geochem 20:1626–1636. https://doi.org/10.1016/j.apgeochem.2005.04.007
Seitzinger S, Harrison JA, Böhlke JK, Bouwman AF, Lowrance R, Peterson B, Tobias C, Van Drecht G (2006) Denitrification across landscapes and waterscapes: a synthesis. Ecol Appl 16:2064–2090. https://doi.org/10.1890/1051-0761(2006)016[2064:DALAWA]2.0.CO;2
Spalding RF, Exner ME (1993) Occurrence of nitrate in groundwater: a review. J Environ Qual 22:392–402. https://doi.org/10.2134/jeq1993.00472425002200030002x
SPAQUE (2018) Ht0806-009 – "Crayère des Fonds de Morvau" à Binche Surveillance environnementale: Bilan 2018 [Environmental monitoring]. SPAQUE, Liège, Belgium
SPW (2018) Terrils (version 2018) [Mining spoil heaps]. Série, SPW, Wallonia, Belgium
SPW (2022) Banque de Données de l'État des Sols (BDES) - Inventaire par parcelle des informations en lien avec l'état des sols [Soil condition database: inventory by plot of information related to the state of the soil]. SPW, Wallonia, Belgium
SPW, DEE (Direction des Eaux souteraines) (2022a) Etat des nappes et des masses d'eau souterraine de Wallonie [State of the water tables and groundwater bodies in Wallonia]. SPW, Wallonia, Belgium
Su C, Zhang F, Cui X, Cheng Z, Zheng Z (2020) Source characterization of nitrate in groundwater using hydrogeochemical and multivariate statistical analysis in the Muling-Xingkai Plain, Northeast China. Environ Monit Assess 192. https://doi.org/10.1007/s10661-020-08347-6
Sutton MA, Howard CM, Erisman JW, Billen G, Bleeker A, Grennfelt P, Van Grinsven H, Grizzetti B (2011) The European nitrogen assessment: sources, effects and policy perspectives. Cambridge University Press, Cambridge, UK
Thiernesse L (1967) Problèmes de reconversion et d'aménagement de la région boraine [Problems of conversion and development of the Boraine region]. Hommes Terres Nord 1:10–25
Thorburn PJ, Biggs JS, Weier KL, Keating BA (2003) Nitrate in groundwaters of intensive agricultural areas in coastal Northeastern Australia. Agric Ecosyst Environ 94:49–58
Thorndike RL (1953) Who belongs in the family? Psychometrika 18:267–276. https://doi.org/10.1007/BF02289263
Torres-Martínez JA, Mora A, Knappett PSK, Ornelas-Soto N, Mahlknecht J (2020) Tracking nitrate and sulfate sources in groundwater of an urbanized valley using a multi-tracer approach combined with a Bayesian isotope mixing model. Water Res 182. https://doi.org/10.1016/j.watres.2020.115962
Torres-Martínez JA, Mora A, Mahlknecht J, Kaown D, Barceló D (2021) Determining nitrate and sulfate pollution sources and transformations in a coastal aquifer impacted by seawater intrusion: a multi-isotopic approach combined with self-organizing maps and a Bayesian mixing model. J Hazard Mater 417. https://doi.org/10.1016/j.jhazmat.2021.126103
van der Maaten LJP (2022) t-SNE. https://lvdmaaten.github.io/tsne/. Accessed May 2023
van der Maaten LJP, Hinton GE (2008) Visualizing high-dimensional data using t-SNE. J Mach Learn Res 9:2579–2605
Vengosh A, Barth S, Heumann KG, Eisenhut S (1999) Boron isotopic composition of freshwater lakes from Central Europe and possible contamination sources. Acta Hydrochim Hydrobiol 27:416–421
Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH, Tilman DG (1997) Technical report: human alteration of the global nitrogen cycle—sources and consequences. Ecol Appl 7:737–750. https://doi.org/10.2307/2269431
Vystavna Y, Diadin D, Valeriy Y, Hejzlar J, Vadillo I, Huneau F, Lehmann M (2017) Nitrate contamination in a shallow urban aquifer in East Ukraine: evidence from hydrochemical, stable nitrate isotope, and land use analysis. Environ Earth Sci 76. https://doi.org/10.1007/s12665-017-6796-1
Wakida FT, Lerner DN (2005) Non-agricultural sources of groundwater nitrate: a review and case study. Water Res 39:3–16. https://doi.org/10.1016/j.watres.2004.07.026
Wells NS, Hakoun V, Brouyère S, Knöller K (2016) Multi-species measurements of nitrogen isotopic composition reveal the spatial constraints and biological drivers of ammonium attenuation across a highly contaminated groundwater system. Water Res 98:363–375. https://doi.org/10.1016/j.watres.2016.04.025
WHO (2017) Guidelines for drinking-water quality: fourth edition incorporating the first addendum. World Health Organization, Geneva
Widory D, Kloppmann W, Chéry L, Bonnin J, Rochdi H, Guinamant J-L (2004) Nitrate in groundwater: an isotopic multi-tracer approach. J Contam Hydrol 72:165–188. https://doi.org/10.1016/j.jconhyd.2003.10.010
Widory D, Petelet-Giraud E, Brenot A, Bronders J, Tirez K, Boeckx P (2013) Improving the management of nitrate pollution in water by the use of isotope monitoring: the $\delta^{15}N$, $\delta^{18}O$ and $\delta^{11}B$ triptych. Isotopes Environ Health Stud 49:29–47
Widory D, Petelet-Giraud E, Négrel P, Ladouche B (2005) Tracking the sources of nitrate in groundwater using coupled nitrogen and boron isotopes: a synthesis. Environ Sci Technol 39:539–548. https://doi.org/10.1021/es0493897
Xu S, Kang P, Sun Y (2016) A stable isotope approach and its application for identifying nitrate source and transformation process in water. Environ Sci Pollut Res 23:1133–1148
Xue D, Botte J, De Baets B, Accoe F, Nestler A, Taylor P, Van Cleemput O, Berglund M, Boeckx P (2009) Present limitations and future prospects of stable isotope methods for nitrate source identification in surface- and groundwater. Water Res 43:1159–1170