Science of the Total Environment 963 (2025) 178412

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Addressing drivers and data gaps in Spain's non-compliance of drinking water quality standards

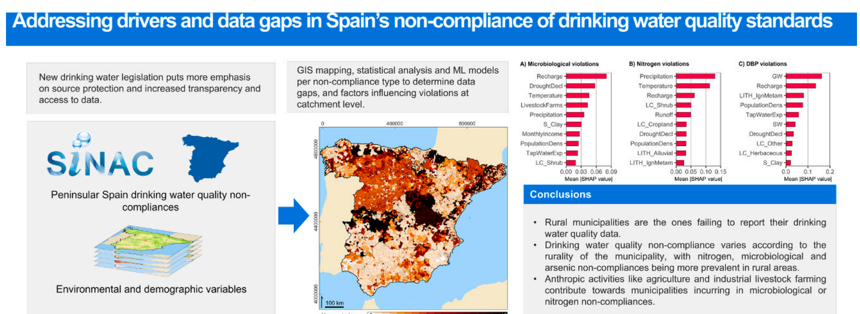
Delia M. Andries [a,*], Alberto Garrido [b], Lucia De Stefano [a]

[a] Departamento de Geodinámica, Estratigrafía y Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, Calle José Antonio Nováis 12, 28040 Madrid, Spain
[b] CEIGRAM, Universidad Politécnica de Madrid, Paseo Senda del Rey 13, 28040 Madrid, Spain

HIGHLIGHTS

- In Spain, rural municipalities are the ones failing to report their drinking water quality data.
- At least 47.2 % of the Spanish population is supplied from a water source that fails to achieve the 'global good status'.
- Nitrogen, microbiological and arsenic drinking water quality non-compliances are more prevalent in rural areas.
- Agriculture and confined livestock farming contribute towards microbiological or nitrogen non-compliances.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Damià Barceló

Keywords:
Random forest
Rural water supply
Drinking water directive
Data quality reporting
Water quality

ABSTRACT

In 2021 Spain passed a new law regarding the quality of drinking water, increasing the transparency and introducing a risk assessment approach to the catchment area, which ascribe to the shift in recent years in how drinking water management is understood in the European Union legislation. Good data quality is important to ensure the correct implementation of policies. We used the drinking quality data uploaded to Spain's National Drinking Water Information System to gauge the state of the drinking water reporting in Spain, the differences between rural urban and rural areas in both quality and reporting and identify which variables at catchment level influence the probability of a municipality incurring in drinking water quality non-compliance. Random forest modeling was used to assess the drivers of non-compliance, including environmental (e.g., land cover, lithology, climate, state of the water supply source) and demographic (e.g., tap water expenditure, population density) data.

We found that rural municipalities are more vulnerable both because of a lack of reporting but also because they have higher non-compliance rates for arsenic, microbiological and contaminants and nitrogen compounds (e.g. nitrate). We also found different spatial patterns of non-compliance according to each group of contaminants (e.g., microbiological violations are widespread in the northern half of Spain). The random forest model suggests that agriculture and confined livestock farming are behind nitrogen and microbiological non-compliances. Climate drivers have also emerged for all the groups of contaminants, which underscores the importance of studying drinking water quality non-compliance on a case-by-case basis in order to properly adapt to local realities and enhance compliance across Spain.

* Corresponding author at: Facultad de Ciencias Geológicas, C/ Jose Antonio Novais 12, 28040 Madrid, Spain.
E-mail address: deliandr@ucm.es (D.M. Andries).

https://doi.org/10.1016/j.scitotenv.2025.178412
Received 5 August 2024; Received in revised form 23 December 2024; Accepted 5 January 2025
Available online 17 January 2025

1. Introduction

Drinking water is a service enabled not only by sufficient supply, but also by the quality of the serviced water. As a vector or cause of diseases, it has numerous socio-economic implications, the main one being the health impacts on the population (Li and Wu, 2019), and secondly, on the citizens' trust. Being this the case, drinking water supply is subject to regulations that other public utilities do not face (Okun, 2003). In the European Union (EU), the main regulation is the Drinking Water Directive (DWD), which is incorporated into national legislation of the Member States. Since its first version in 1980, this Directive has moved from the 'end-of-pipe' solutions that focus only on the testing of the final product to a more holistic concept of drinking water management. The latest version, approved in 2020, underscores the importance of the catchment area of the abstraction point as part of the drinking water supply cycle (Andries et al., 2024).

Although treatment can sometimes compensate for the loss of water quality, the condition of the water body from which it is drawn ultimately determines the quality of tap water (R. Das et al., 2021; Paun et al., 2017). Diffuse pollution in the form of presence of pesticides and nitrogen (commonly as nitrates) from agriculture has been recognized as one of the biggest threats to drinking water resources in the European Union (Wuijts et al., 2021), with a special emphasis on their contribution to the pollution of groundwater resources (European Commission, 2014; Kristensen et al., 2018). In the EU, 65 % of the drinking water comes from groundwater (European Environment Agency, 2022). The second most common reason for poor groundwater chemical status identified in the Second River Basin Management plans (2016 – 2021) was failing to meet the requirements for drinking water protected areas (Kristensen et al., 2018). In the case of surface waters, nitrogen discharges also include atmospheric deposition or wastewater and industrial discharges (Erisman et al., 2011), and their contamination can be exacerbated by climate change and thus carried into drinking water (European Commission, 2014). Floods and extreme rainfalls intensify the runoff in the catchment area, and thus the incorporation of possible contaminants to water courses (Yard et al., 2014) and groundwater (Corada-Fernández et al., 2017; Geris et al., 2022); low flows during dry periods limit the dilution of chemical contaminants (Boxall., 2011; Osorio et al., 2014; Ricart et al., 2010; Sjerps et al., 2017). Anthropogenic compounds like *per*- and polyfluoroalkyl substances (PFAS) and endocrine-disrupting chemicals (EDCs) are considered of emerging concern globally as well as in the EU (Halleux, 2023), as proved by their introduction in both the 2020 DWD and the list of priority substances that must be monitored under the EU Water Framework Directive (WFD). Other threats to drinking water quality come from ubiquitous elements such as arsenic, recognized worldwide as one of the most serious inorganic contaminants (Smedley and Kinniburgh, 2002). Its presence in the drinking water supply has not been a subject of interest at EU-wide level, despite concentrations exceeding the legal limit in the drinking water of several of its countries (Hungary, Serbia, Croatia, Greece, Italy, Spain, Slovakia and Romania) (van Halem et al., 2009).

It is expected that large water supplies (those providing >1000 m^3/day on average or serving a population of >5000, as per the DWD) should have the means to overcome the abovementioned challenges (WHO, 2011). However, small water supplies (SWS), most of them situated in rural areas, require a different consideration. In the EU and from the DWD legislative point of view, minimum monitoring requirements for SWS are laxer in comparison with their larger counterparts, since monitoring is done based on the population served (Roig et al., 2014), and they are excluded from the mandatory reporting to the European Commission (European Commission, 2014). Thus, available data for these zones is usually scarce, which prevents assessing SWS performance and safety. This problem is compounded by the lack of personnel with specialized knowledge, making SWS less resilient to possible climate change impacts, and more vulnerable to contamination, especially microbial and nitrates (WHO, 2011). As a matter of fact, in 2008 the EC found that compliance with the DWD parametric limits in the SWS was down to 30 % (European Commission, 2011), and in 2011 the Commission was looking at implementing a risk-based approach for SWS for more effective quality control (European Commission, 2011). Despite this, literature regarding insecure access to water in rural areas in Europe remains scarce (Meehan et al., 2020). Moreover, the abovementioned challenges, and others pertaining to environmental justice (Anderson, 2010; Balazs and Ray, 2014), remain masked by reports that aggregate data at large scales, as it is the case of the WHO/UN Joint Monitoring Program (Lee et al., 2023; Meehan et al., 2020) or the annual reports submitted by Member States to the European Commission (EC) regarding the level of compliance with the DWD parameters (Grizzetti et al., 2011). This data gap further leaves unchecked the assumption that water access in high-income countries (the so-called 'Global North') is universal, safe, and uniformly or equitably governed (Meehan et al., 2020). At the same time, water governance, understood both as the inter-institutional agreements that regulate the flows of water and as the civil society actors exercising control over the water services (Zwarteveen et al., 2017), relies on accurate data to create a baseline for diagnostics, which is especially important for investment in infrastructure (Lee et al., 2023). In the policymaking sphere, the data that results from the monitoring programs established in the legislation can shed light on the performance of the policy, i.e. whether it is fit for purpose, and gauge the state of policy implementation, as shown in the evaluation of the 1998 DWD (European Commission, 2016) or the Urban Wastewater Directive (European Commission, 2019).

This paper's goal is twofold. First, it focuses on the quality of the drinking water in Spain and on how it is reported by official sources, with an emphasis on the differences between rural and urban areas. Second, we aim to provide a nationwide assessment of the factors that influence non-compliance at catchment- and municipal-level, thus contributing to the implementation of the DWD. To achieve this, we apply a first-time approach to the use of Machine Learning models as risk assessment tools that could potentially predict areas at national scale with more risks of drinking water quality non-compliance by using, mostly, natural and anthropic factors.

2. Methodology

2.1. Study area

Our study case concerns peninsular Spain, excluding the Balearic and Canary Archipelagos, and the African city enclaves of Ceuta and Melilla. The population of Spain is concentrated in urban areas; out of the 43 million inhabitants of peninsular Spain, 86 % of them live in urban or semi-urban municipalities, which represent 15 % of the 7974 peninsular municipalities, organized at the top level in 15 regions (Fig. 1).

Under Spanish law (LBRL, Ley 7/1985; LRSAL, Ley 27/2013), supplying the population with drinking water is a competence reserved to municipalities. This implies the ownership of the supply service and its management, which can be direct (e.g., in-house provision, public company owned by the municipality) or indirect (e.g., contractual or institutional public-private partnerships, leasing agreements). As it happens in all the EU countries, Spanish drinking water supply is organized in supply zones, i.e. "geographically defined areas within which water intended for human consumption comes from one or more sources and within which water quality may be considered as being approximately uniform" (European Commission, 2014 p. 3). In terms of abstraction of raw water intended for consumption, the ownership of the supply system presupposes that the drinking water manager has been granted a concession by the River Basin District (RBD) authorities (RDPH, Real Decreto 849/1986); RBDs in Spain can be intraregional or interregional, and their competences are regulated by the Water Law (TRLA, Real Decreto 1/2001).

The piece of legislation that transposes the DWD to national law and regulates drinking water quality is the Royal Decree for Water for

Human Consumption (RDAC, Real Decreto 3/2023). The RDAC regulations include microbiologic, chemical, indicator, organoleptic and radioactive parameters, which must be monitored in different types of drinking water quality controls that have to be performed by drinking water managers and regional health authorities. The responsibility for interrupting the supply when a health hazard non-compliance occurs rests with the regional health authorities. The RDAC also mandates that all analytic results from the drinking water quality monitoring must be reported to the Drinking Water Information National System (SINAC in its Spanish acronym). The SINAC was established in 2003, when the first RDAC was passed, and further developed by specific legislation in 2005 (OM SINAC, Orden SCO/1591/2005). It is hosted and maintained by the Ministry of Health, whereas the responsibility of reporting the results of drinking water quality reporting falls on the municipality.

2.2. Data collection and preparation

The data used for this research comes from the SINAC reporting for the years 2004–2021 and was provided by the Ministry of Health on May 2023 upon request. All the monitored parameters were classified in different types (Table 1). Additionally, based on expert knowledge indicating that efforts to enforce reporting to SINAC significantly increased from 2016 onwards, we created a subset of reported data from 2016 to 2021. This subset includes municipalities that reported any parameter for at least five years during this period. The 2004–2021 data was used to study (a) the quality and characteristics of the reporting to SINAC, and (b) the evolution of non-compliance per municipality type (Fig. 2). The 2016–2021 subset database was utilized to (a) study the spatial distribution of water quality non-compliance, (b) assess the differences in the percentage of non-compliant samples according to the status of the water source, and (c) the drivers of non-compliance (Fig. 2).

To analyze strengths and gaps of the drinking water quality reporting we looked at the number of unreported years per municipality using the 18 years of existing data. We performed a Kruskal-Wallis and a Pairwise Wilcoxon test with Benjamini-Hochberg adjustment to assess whether the number of unreported years was different per municipality type (H_1, population medians are not equal). We also performed a Z-test of homogeneity of proportions to analyze if the percentage of non-compliant drinking water samples of different group of contaminants was higher if the drinking water was abstracted from a water body that failed to achieve global status according to the EU WFD (H_0).

Table 1
Classification of parameters on different non-compliance types. Organic non-compliances refer to those compounds in RDAC that contain a carbon-hydrogen or a carbon-carbon link, whereas inorganic contaminants are the ones lacking that bond. All the groups except health non-compliances are mutually exclusive.

Non-compliance type	Parameters associated to non-compliance type
Microbiological	*Escherichia coli*, *Clostridium perfigens*, Enterococci
Nitrogen	Nitrate, Nitrite
Disinfection by-product	Trihalomethanes, Bromate, Chlorate, Chlorite
Lead and copper	Lead, Copper
Inorganic	Antimony, Cadmium, Fluoride, Selenium, Mercury, Chromium, Nickel, Cyanide,
Organic	Benzene, 1,2-Dicloroetene, Benzo(*a*)pirene, Microcystine LR, Vinyl chloride, Acrylamide, Epichloridyne
Pesticides	Sum of pesticides, and all individual pesticides
Arsenic	Arsenic
Radioactive	Tritium, Indicative Dose, Gross beta activity, Residual alpha activity, Residual beta activity, Radon
Health non-compliance	All the other non-compliance types, excluding indicator non-compliances: Microbiological, Nitrogen, DBPs, LC, Inorganic, Organic, Pesticides, Arsenic and Radioactive types
Indicator parameters	Aluminum, Ammonium, Chloride, Color, Conductivity, pH, Iron, Manganese, Odor, Oxidisability, Sulphate, Sodium, Taste, Colony count at 22 °C, Coliform bacteria

Random forest (RF) is an ensemble method based on multiple decision trees (Breiman, 2001). Random forest trains a set of randomized decision trees, whose output is aggregated into a single output by using voting (in the case of classification) or averaging (in the case of regression). The randomization is implemented first by sampling the data set with replacement (bagging) and secondly, at the decision node level, where a random subset of predictors is chosen (Breiman, 2001; Rigatti, 2017). This algorithm has been used for some years in water science due to its suitability when non-linearity exists, which is common in the processes studied in water science (Tyralis et al., 2019).

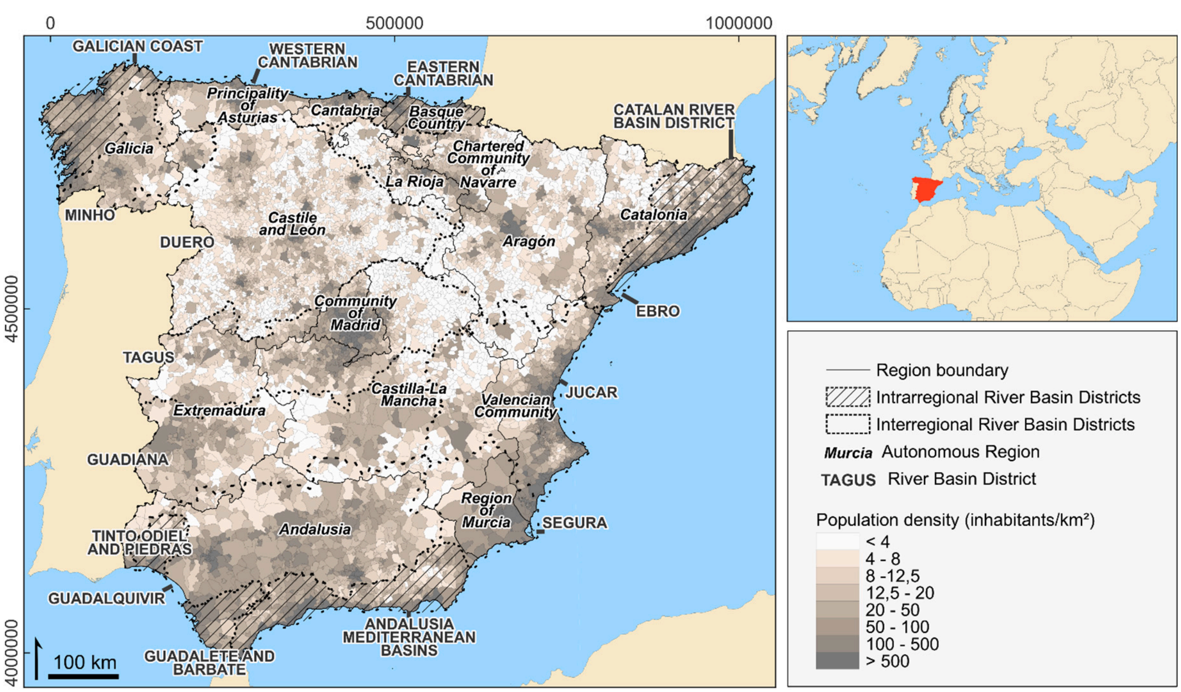


Fig. 1. Administrative divisions of Peninsular Spain (municipalities, regions) and River Basin Districts.

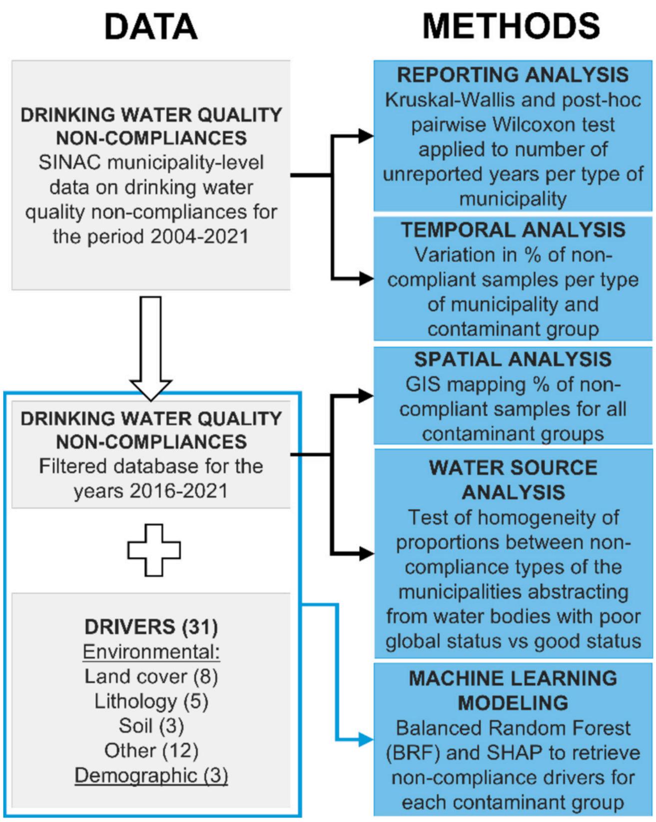


Fig. 2. Flowchart describing the data and methods employed.

Particularly, random forest has become popular for predicting water quality (Zavareh et al., 2024), being applied to predict a variety of indicators and compounds, such as pH, chloride, nitrate, arsenic, iron, or total phosphorus (Ganga Devi, 2020; Tesoriero et al., 2017; Wang et al., 2021). Random forest has also been used to predict Safe Drinking Water Act violations in the United States (Pennino et al., 2020; Scanlon et al., 2022). Building on this knowledge, we decided to apply a RF algorithm to each group of contaminants.

2.3. Balanced random forest classification

As stated, the dataset consisted of aggregated data of all the municipalities from 2016 to 2021 that reported any parameter for 5 or more years. The response variable is binary: for each non-compliance type (e.g. microbiological, nitrogen), 1 represents the existence of at least one non-compliance event for that municipality in any given moment of the selected time period, whereas 0 means that there were not any non-compliances reported.

Data exploration showed that the dataset was heavily unbalanced. Due to this characteristic, we opted for a Balanced Random Forest (BRF) classifier, which is able to deal with the unbalanced panel data (Scanlon et al., 2022) and works well with possible collinearity (Cutter, 2007). The unbalanced panel data was split with a stratified split, and the BRF was applied to 80 % of the data (training set), and the resulting model was applied to the remaining 20 % (test set) to assess model performance. Hyperparameter tuning was done via a Randomized Search with a stratified ten-fold cross-validation, searching for the best Precision-Recall Area Under Curve (AUC PR), since it performs better for binary classifiers applied to heavily imbalanced datasets (Davis and Goadrich, 2006; Saito and Rehmsmeier, 2015). Other important metrics for classifiers, like Balanced Accuracy (ACC) and Area Under the Receiver Operating Characteristic (AUROC) were also calculated. All operations were performed in Python 3.7. Unbalanced random forest modeling was done with the help of the *imbalanced-learn* toolbox (Lemaître et al., 2017), with supporting operations, and metrics were calculated with the package package scikit-learn (Pedregosa et al., 2011).

To assess the contribution of each explanatory variable to the predicted probability of a municipality not complying with the drinking water quality standards, the output of the BRF was passed through Shapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). SHAP is model agnostic, and its measure of variable importance allows to investigate the impact (positive or negative) of each variable on the prediction, i.e., not complying with the standards, and the magnitude of said impact (Lundberg et al., 2019).

2.3.1. Explanatory variables

Explanatory variables (Table 2; SM Fig. 1, SM Fig. 2) were selected based on literature review of previous studies (see Allaire et al., 2018; Pennino et al., 2017, 2020; Scanlon et al., 2022); their selection was also determined by data availability (e.g. many municipalities have their Gini inequality index hidden). Explanatory variables were tested for collinearity (SM Fig. 3); variables with a Pearson correlation coefficient higher than 0.75 and lower than −0.75 were discarded, with the exception of climate and climate-derived variables (i.e. precipitation, runoff, temperature, and aquifer recharge).

The quantitative and qualitative status of the water bodies can determine the quality of the drinking water; at its core, this means that what happens at catchment level can influence the drinking water quality. This has been studied in the contiguous United States (CONUS) by Pennino et al. (2020) in the case of nitrates (using data aggregated at National Hydrography Dataset catchment level) and Scanlon et al. (2022) (using data aggregated at county level), but to the knowledge of the authors, no similar study at a national scale has been conducted in Europe.

Spain does not have a dataset with the origin of drinking water at a national scale, nor a delineation of the boundaries of drinking water systems. Hence, the most direct approach to getting a comprehensive database of catchments associated with a municipality was to request the water abstraction points associated with each municipality. However, only three of the 14 River Basin Authorities contacted for this study (Eastern Cantabric, Ebro and Jucar River Basin) provided this information, and due to its structure, the Eastern Cantabric database could not be used. For the rest of the river basins, we could only identify the abstraction origin (i.e., the water body -or bodies- from which water intended for human consumption is abstracted) of Urban Demand Units (UDUs, i.e., groups of municipalities aggregated per common origin) from the Third Cycle River Basin Management Plans (RBMPs), which are less accurate than abstraction points. Also, since accuracy varies between RBMPs, the UDU information was completed using a) the Water Points Database (Instituto Geológico y Minero de España, n.d.), b) web pages of large water operators, c) local newspapers, b) the Local Equipment Survey from the year 2017 (Secretaría de Estado de Política Territorial, 2017), c) aerial photography and topographic map inspection, and d) Special Drought Management Plans (MITERD, 2023c; 2023d).

With this information, we compiled a database of the water bodies used by each municipality to abstract drinking water and calculated the catchment area of drinking water abstraction. Catchment area for a given municipality (MCA_i) was calculated as the sum of the catchment areas of all surface water bodies that supply the municipality and its groundwater catchment area, as follows:

$$MCA_i = \sum\nolimits_0^{n_i} SWC_{n_i} + GWC_i \quad (1)$$

Where i is the municipality, n_i the number of surface water bodies that supply the municipality. Surface catchment area of a surface water body (SWC) is calculated by spatially intersecting the water body used for drinking water abstraction with a 100 m-resolution DEM catchment shapefile (MITERD, 2006). If the water was abstracted from a water body not defined by the RBMPs, the municipal area was assigned as the

Table 2
Explanatory variables used in the model. Spatial distribution of the variables and their histograms can be found in SI Figs. 1 and 2. All variables except the number of drought declarations, origin of abstraction (SW/GW), state of water bodies and demographic variables refer to the catchment area of the municipalities.

Name of variable	Description	Variable range	Data source
Land cover			
LC_Cropland	% of cropland	0–100	Corine Land Cover 2018 (Copernicus Land Monitoring Service, 2020)
LC_Developed	% of developed land	0–88.63	
LC_Forest	% of forest cover	0–93.77	
LC_Shrub	% of shrub cover	0–91.21	
LC_Herbaceous	% of herbaceous cover	0–78.49	
LC_Other	% of other land cover (e.g. bare terrain)	0–64.44	
LC_Wetland	% of wetland cover	0–29.39	
Irrigation_cov	% of irrigated land cover	0–95.11	SIOSE 2014 (Instituto Geográfico Nacional, 2014)
Lithology			
LITH_Alluvial	% of alluvial deposits	0–100	Geologic map of the Iberian Peninsula (Instituto Geológico y Minero de España, 2015)
LITH_Carbonated	% of carbonated rocks	0–100	
LITH_Sedimentarian	% of sedimentary rocks	0–100	
LITH_IgnMetam	% of igneous and metamorphic rocks	0–100	
LITH_Volcanic	% of volcanic rocks	0–15.58	
Soil			
S_Clay	Mean subsoil clay content (%)	0–46.94	European Soil Database Derived data (Hiederer, 2013)
S_OC	Mean subsoil organic carbon content (%)	0–1.51	
S_BulkDensity	Mean subsoil bulk density	0–1.70	
Other			
Temperature	Mean temperature for the 1980–2017 series (°C)	4.32–18.73	Assessment of natural regime water resources (SIMPA) (Centro de Estudios y Experimentación de Obras Públicas, 2020)
Precipitation	Mean precipitation for the 1980–2017 series (mm)	267.78–2097.50	
Recharge	Mean recharge for the 1980–2017 series (mm)	0–931.78	
Runoff	Mean runoff for the 1980–2017 series (mm)	0.08–1070.39	
DroughtDecl	Number of times "long drought" was declared in the municipality	0–30 −1 for unknown municipalities	Drought monitoring maps (Ministerio para la Transición Ecológica y el Reto Demográfico [MITERD], 2023b)
LivestockFarms	Density of confined livestock farms per km^2	0–0.76	European Pollutant Release and Transfer Register (*E*-PRTR) (European Environment Agency, 2020)
IndustrialSites	Density of industrial sites per km^2	0–1.85	
UrbanDischargeP	Density of authorized urban discharge points per km^2	0–2.28	National Discharge Census (MITERD, 2023a)
IndustrialDischargeP	Density of authorized industrial discharge points per km^2	0–1.07	
GW	Groundwater abstraction	0–1	Third River Basin Management Plans (MITERD, 2023e)
SW	Surface water abstraction	0–1	
WB_GlStatus	Usage of water body in 'failing to achieve good' global state	0–1	
Demographic			
PopulationDens	Log10 of population density (inhabitants/km^2)	−0.73–4.43	2020 census (Instituto Nacional de Estadística, 2020b)
MonthlyIncome	Average monthly income per family for the years 2016–2019 (€)	992.12–3785.75	2019 Household Budget Survey (Instituto Nacional de Estadística, 2020a)
TapWaterExp	Annual average expenditure on tap water per family for the years 2016–2019 (€)	65.55–392.20	

SWC. In the case of groundwater abstraction, groundwater catchment area (GWC) was also assimilated to the area of the municipality.

3. Results and discussion

3.1. SINAC and spatial variability in drinking water quality non-compliance

The spatial analysis of the entire SINAC dataset (2004–2021) shows that there are several regions in Spain with chronic non-reporting to SINAC (i.e., more than half of the 18 years without uploading the mandatory data to the information system) (Fig. 3 A). Lack of reporting is especially prevalent in central Spain, some northern parts of the country and the interior of the East coast. Kruskal-Wallis test reveals differences between the number of unreported years per municipality type ($p < 0.001$). Whereas the median of unreported years for urban and semi-urban municipalities is 1 year (over a 18-year long period), rural municipalities have a median of 6 unreported years (Fig. 3B).

The vast majority of urban and semi-urban municipalities joined the reporting effort of SINAC since it was established in 2004, whereas the number of rural municipalities reporting to the national database has been growing gradually (Fig. 3C). Thus, despite the mandatory nature of

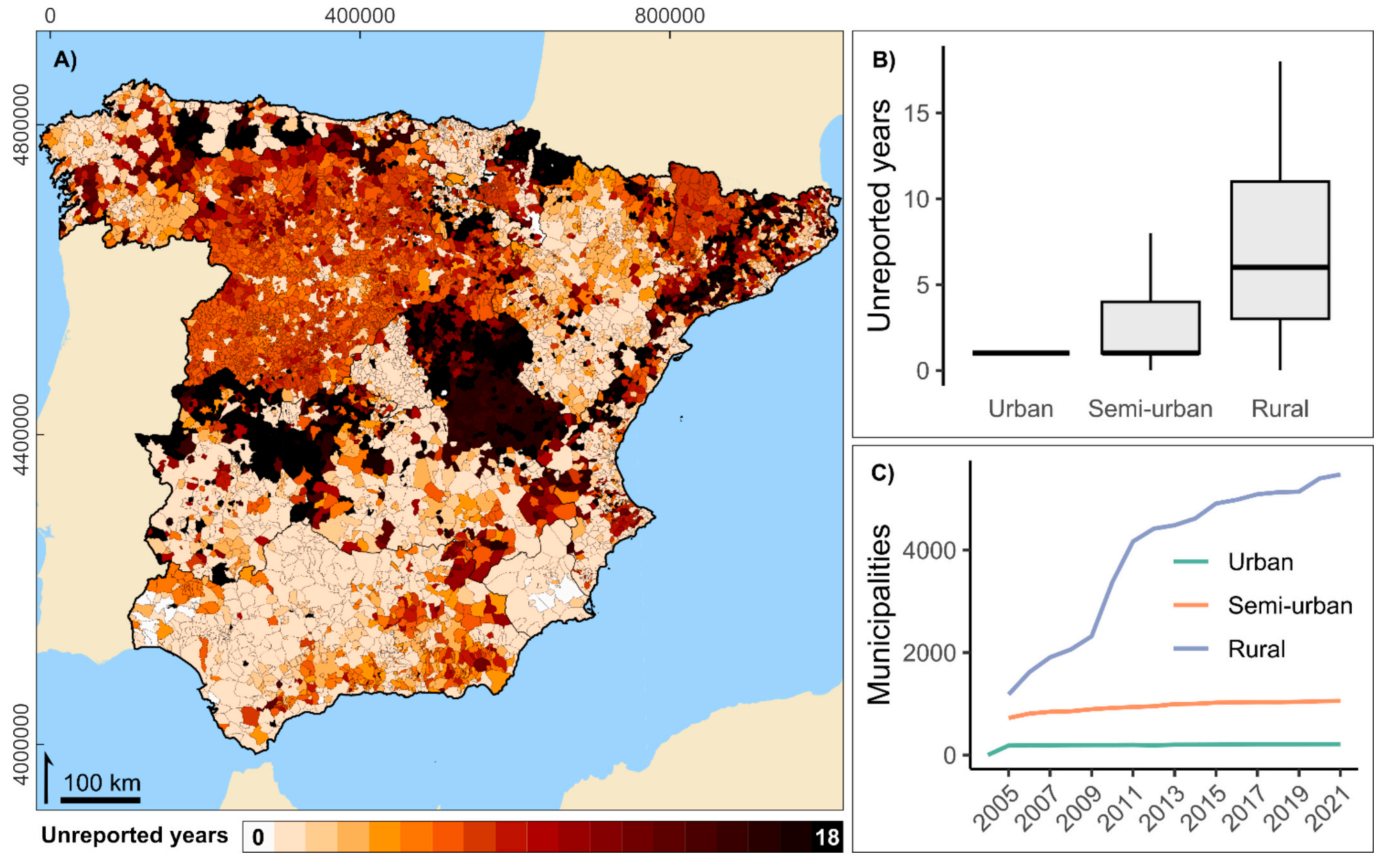


Fig. 3. SINAC reporting at municipal level in peninsular Spain. A) Geographic distribution of unreported years, per municipality; B) Median of unreported years per municipality type; C) Time progression of the number of municipalities reporting their data to SINAC, per type.

SINAC reporting and the efforts to improve it, there are numerous rural municipalities in Spain whose drinking water quality is still unreported, sometimes forming spatial clusters (see Fig. 3 A).

Fig. 4 shows the evolution of the percentage of non-compliant samples through the years. Overall, health-related drinking water non-compliances at a national level hardly reach over 1 %, with the exception of disinfection by-products, arsenic and nitrogen (Fig. 4 A). However, a much more nuanced reality of drinking water quality emerges when considering data disaggregated by type of municipality and group of contaminants (Fig. 4). Thus, non-compliance levels increase in their magnitude and number of contaminants the more rural the municipality is; this especially affects arsenic, nitrogen compounds (nitrate and nitrites), and microbiological contaminants (Fig. 4D).

Disinfection by-products have the highest non-compliance percentage, to the point that they are the ones contributing the most to the overall number of health non-compliance (those that group all the health parameters). In the year 2016, after disinfection by-product non-compliances experienced a decrease in their numbers in the year 2009, chlorite and chlorate were introduced in the monitoring routine by some municipalities, which then led to another increase in the reporting of DBP non-compliances (Fig. 4E, 4F, and 4G) for all the municipal types. The influence of DBPs on the general ratio of health non-compliances can be seen in the decline in the percentage of non-compliant samples for all health parameters in 2020 for urban (Fig. 4B) and semiurban (Fig. 4C) municipalities, related to a decrease in the percentage of DBP non-compliant samples (Figs. 4E, 4F), and the following increase in the following year. Besides, rural areas have not experienced a decrease in the percentage of non-compliant samples in 2020; in fact, in that year, DBP non-compliance continued to increase in them (Fig. 4G). Another decline can also be seen for the indicator parameters of urban and semi-urban municipalities, and more slightly on rural ones.

Lead and copper, related to pipe maintenance, experienced a decrease in 2013–2014 in all the municipalities (Figs. 4B, 4C, and 4D), and non-compliance levels have remained low until the end of the study period.

For the 2016–2021 period, health parameters (Fig. 5 A) and indicator parameters (Fig. 5B) non-compliances can be found across peninsular Spain, while spatial patterns emerge for some specific contaminant groups. For instance, microbiological non-compliances appear in the northern half of the Peninsula (Fig. 5B). Nitrogen compounds non-compliances can be found in three large regions: one around the upper half of Spain, and two in Eastern Spain (Fig. 5C), whereas disinfection by-products non-compliance is more concentrated in central and southwestern Spain (Fig. 5D). On the other hand, arsenic non-compliances are restricted to certain zones in the west of Spain (Fig. 6E).

3.2. Drinking water sources in Spain

Fig. 6 depicts the administrative, population and spatial distribution of the tap water origin in Spain, according to official data. Overall, 42.7 % of the population of peninsular Spain relies solely on surface water for human consumption (Fig. 6B), while 9.7 % is supplied only by aquifers and 12.6 % use mixed sources. However, due to the distribution of population, with most of it concentrated in a few urban areas and the rest dispersed in rural areas, these figures change when considering the number and types of municipalities relying on each type of water sources. As a matter of fact, from the 9.7 % of the population that relies solely on groundwater, only 1.1 % is urban (Fig. 6B), whereas this number increases when we look at mixed water source, with 7.4 % of urban population out of the 12.6 % total.

Thus, in terms of administrative divisions, most municipalities take their water from groundwater, with surface water being the secondary

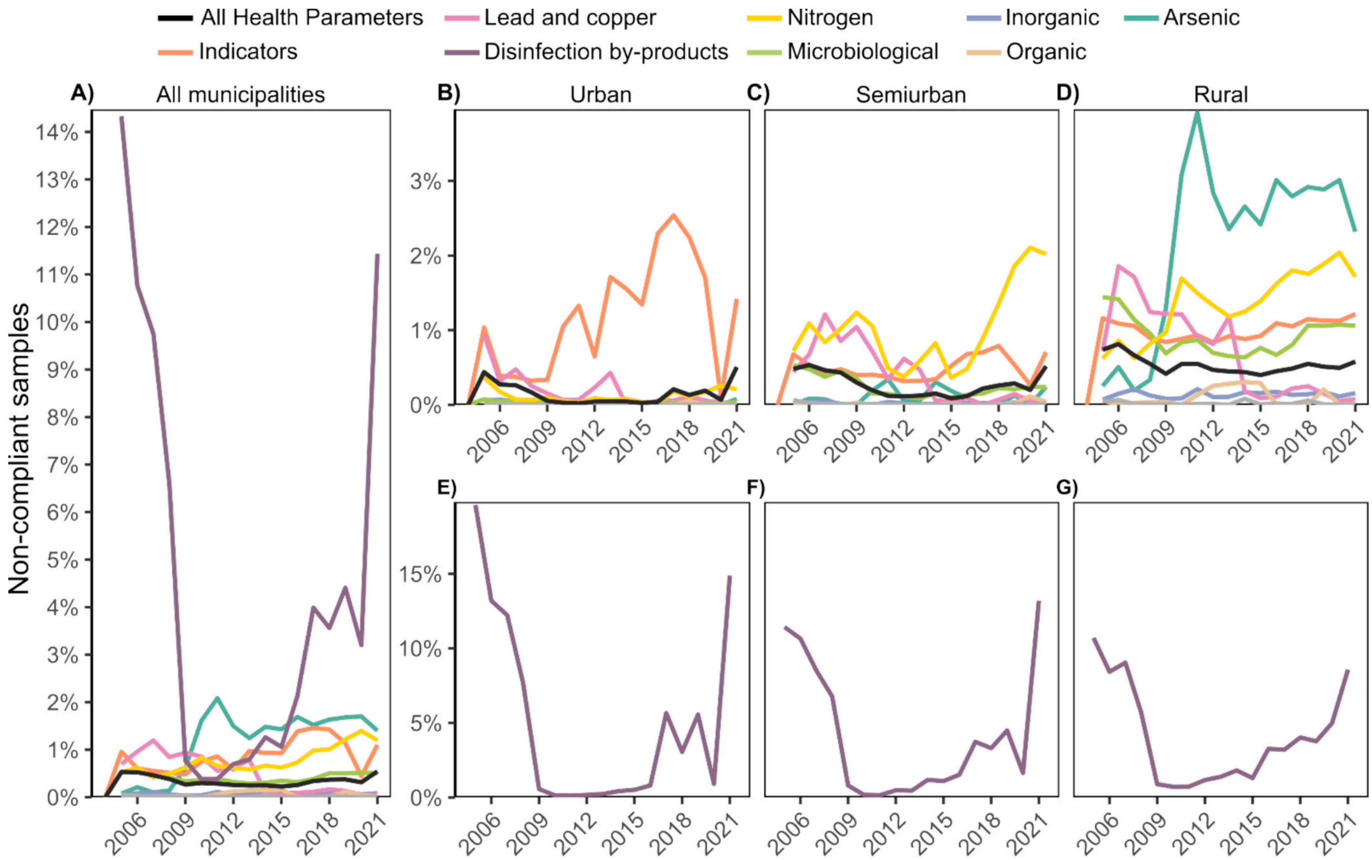


Fig. 4. Time series of percentage of drinking water quality non-compliant samples (over the total number of samples analyzed each year) grouped into different types of contaminants, per municipality type. Disinfection by-products non-compliances are presented separately for clarity.

source, sometimes mixed with groundwater. More than half of our studied rural municipalities have concessions solely on groundwater bodies (Fig. 6 A), making it the most used source for drinking water in rural areas. On the other hand, more than half of the identified urban and semi-urban municipalities (respectively, 67.9 % and 56.5 %) have surface water as their only origin for abstraction (Fig. 6 A).

Lastly, a small portion of the population from urban and semi-urban areas has desalination as a secondary source beside surface and/or groundwater bodies; these municipalities are situated in the Mediterranean coast, where access to water can be especially challenging depending on the infrastructure, the location of the municipality and the pressure from tourism in summer.

The source of abstraction varies across the Spanish geography (Fig. 6C). Southwestern Spain relies more on surface water for drinking water, with large distribution systems dependent on reservoirs. Northern Spanish municipalities, and especially northwestern ones rely on groundwater, the exception being clusters of surface water use or mixed use in urban areas close to large rivers like the Ebro and the Douro. The abstraction origin of 33.1 % of the population, distributed in 1312 municipalities, could not be accounted for via the hydrological plans, corresponding to those River Basin Districts (RBD) dependent on the Regions whose descriptions of urban water demands were insufficient.

We identified 252 lake-type surface water bodies used for drinking water abstraction, which represent 28.44 % of the total WFD lake-type bodies that appear in the third cycle of river basin management planning (SM Fig. 5 A). Out of those water bodies, 32.54 % fail to achieve the global good status' under the WFD requirements. Regarding river-like surface water bodies, 13.40 % (546 rivers or river segments) of them are being used for abstraction of drinking water; out of those, 31.14 % fail to achieve the 'global good status'. Moreover, 35.98 % of the 403 groundwater bodies used for drinking water (59.25 % of the total, SM Fig. 5B) do not achieve the 'global good status'.

Based on this data, we found that at least 47.2 % of the Spanish population is serviced from a source (either surface water or groundwater) that fails to achieve the 'global good status', 14 % abstracts drinking water from a water body with 'poor chemical status'; 38.5 % uses a water body with 'unacceptable ecological status' and 9.99 % abstracts from a water body with 'poor quantitative status'.

We have also studied the difference between the percentage of non-compliances (years 2016–2021) of or each group of contaminants according to the status of the water body used for drinking water abstraction. We found that for nitrogen, DBP, arsenic, pesticides and indicator, non-compliance rates for the years 2016–2021 were higher for drinking water that comes from one or more water bodies with poor global status. Results are reported in Table 3.

By matching the census data with the drinking water source data we found that at least 1.91 % of the total population of Spain, all of it rural, takes its drinking water from one—or more—sources that are not recognized as a water body in the corresponding RBD management plans. This means that water is abstracted from intermittent streams, small local aquifers or springs. More importantly, these represent water bodies that are not being monitored by RBDs as per the WFD guidelines.

3.3. Balanced random forest classifier for drinking water quality non-compliance

For each group of parameters, a BRF model was trained (10 models in total). The complete dataset has 5776 municipalities in total; this number varies across the different models because not all municipalities reported for all contaminant groups during the study period. The number of municipalities and the percentage of non-compliances (baseline) can be found in Table 4. While the overall performance of

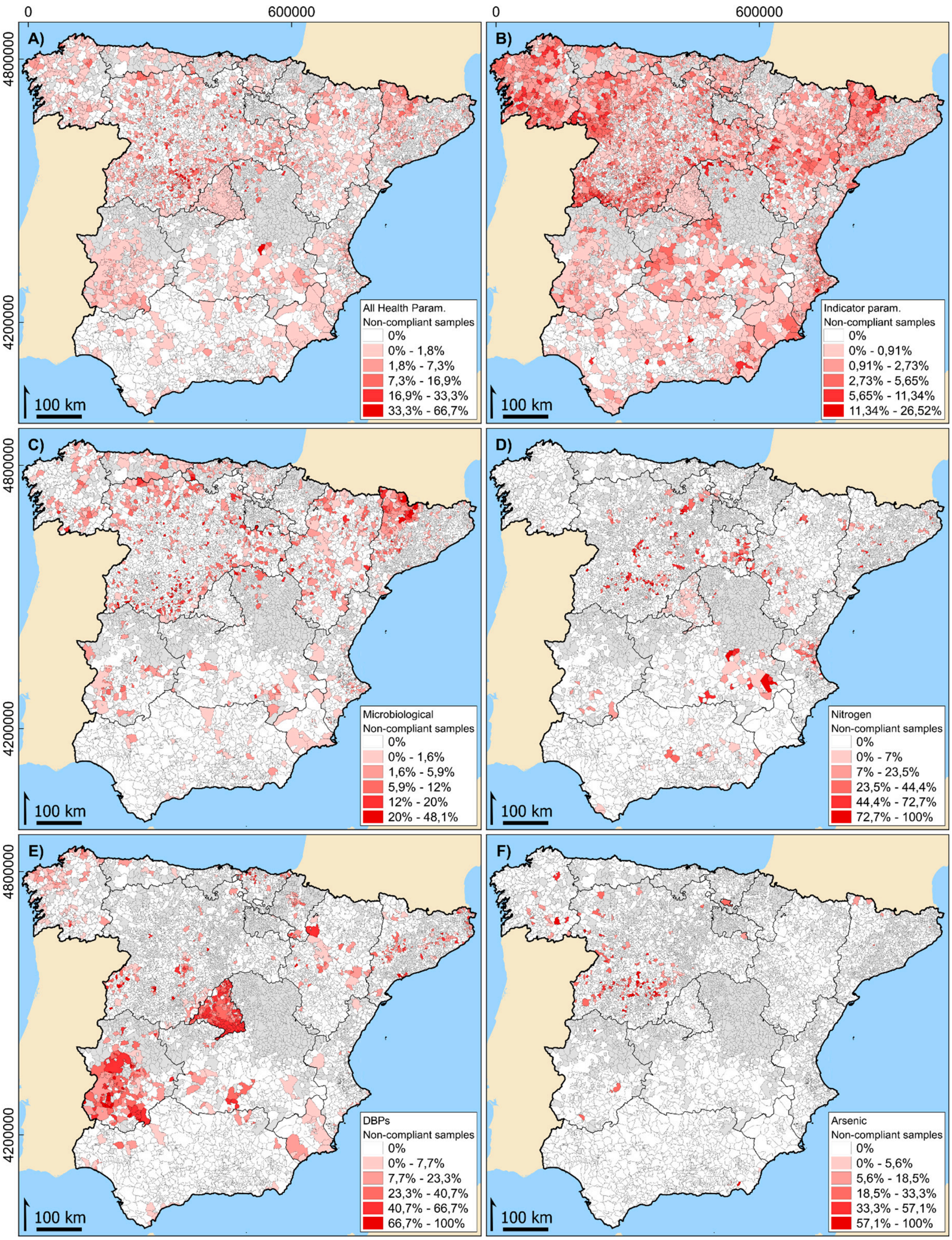


Fig. 5. Maps of the percentage of non-compliant samples per municipality for the 2016–2021 period for: A) All health parameters (HPs); B) Indicator parameters (HP); C) Microbiological parameters (HP); D) Nitrogen-related parameters (HP); E) Disinfection by-products (DBPs; HP); F) Arsenic. Breaks are based on Jenks natural breaks classification method. Grey areas represent municipalities without SINAC data for the given period. Maps for the rest of the models can be found in SM Fig. 4.

the models is good (i.e. there is an improvement of the AUC PR over the baseline value, Chybowski et al., 2024), the specific performance of each model depends on the number of non-compliances and the homogeneity of the groups of parameters. Higher baseline values, homogenous groups of contaminants (like nitrogen and DBP non-compliances) and the consideration of contaminants that are more related to the catchment area activities than socioeconomic or demographic variables have higher balanced accuracy and AUROC. In the next pages we focus on those models that performed better (i.e. baseline values higher than 0.1, AUROC and accuracy higher than 0.70), namely microbiological,

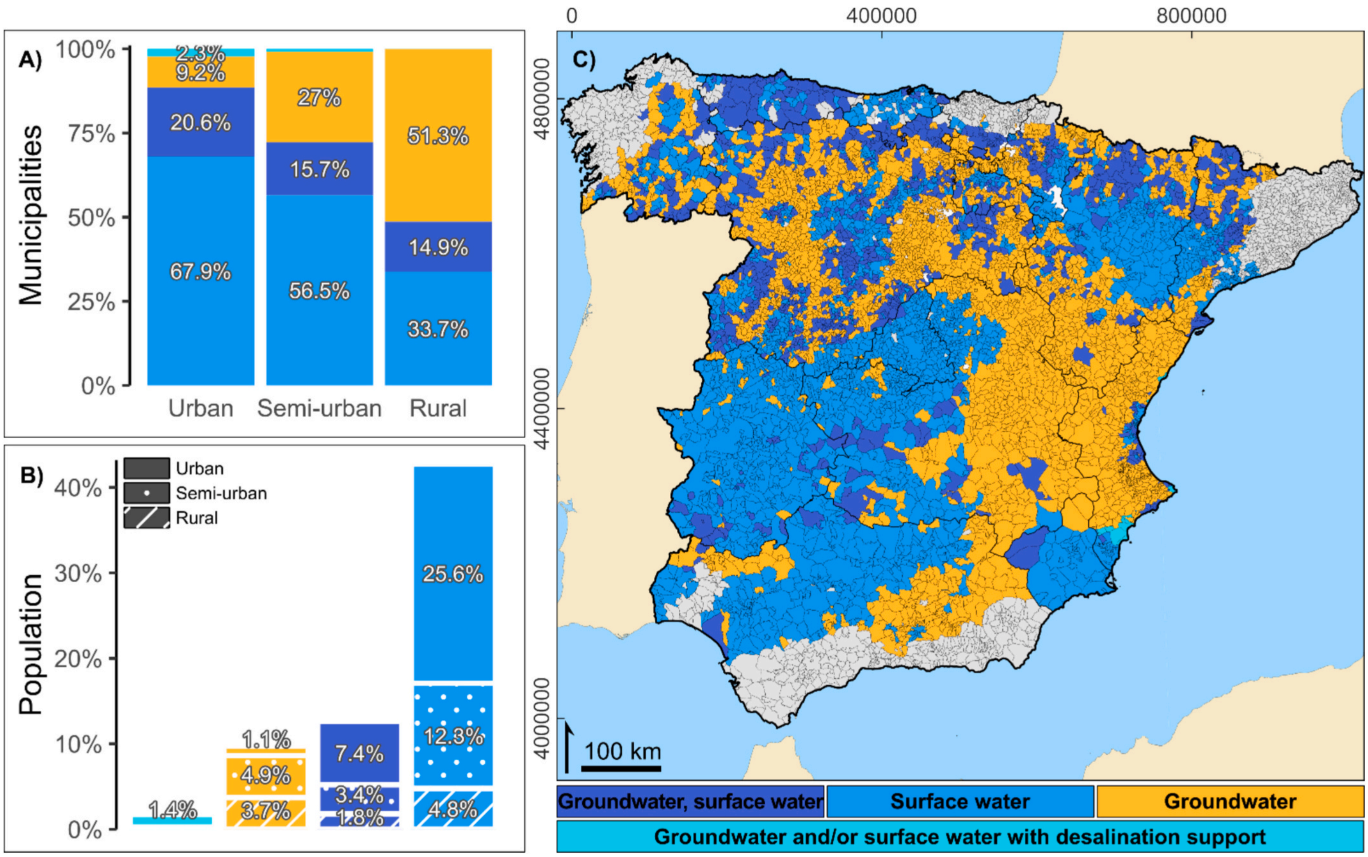


Fig. 6. Drinking water abstraction origin in Spain. A) Type of water source used by municipalities, per type, out of the total analyzed; B) Abstraction source per population type, out of the total population of Spain; C) Geographical distribution of the abstraction type by municipality. Grey municipalities represent those with unavailable drinking water origin data.

Table 3
Results of the *Z*-test of the homogeneity of proportions for the % of non-compliant samples of each type of non-compliance (see Table 1 for more details on the parameters that make each non-compliance type) according to the status of the source of abstraction. Decimals are rounded up.

Non-compliance type	% of non-compliance when drinking water is abstracted from one or more water bodies with poor global status	Standard error	% of non-compliance when drinking water is abstracted from one or more water bodies with good global status	Standard error	*p*-value	
Microbiological	0.27	9.64×10^{-5}	1.18	0.00	1	
Nitrogen	1.27	0.00	1.02	0.00	< 0.001	***
Disinfection by-product	6.80	0.00	4.30	0.00	< 0.001	***
Lead and copper	0.11	0.00	0.12	0.00	0.55	
Inorganic	0.00	5.69×10^{-5}	0.11	9.73×10^{-5}	1	
Organic	6.07×10^{-5}	2.48×10^{-5}	3.09×10^{-4}	6.75×10^{-5}	0.99	
Arsenic	2.22	0.00	1.53	0.00	< 0.001	***
Pesticides	1.65×10^{-4}	1.68×10^{-5}	3.88×10^{-5}	1.08×10^{-5}	< 0.001	***
Indicators	1.34	5.31×10^{-5}	0.94	9.49×10^{-5}	< 0.001	***

nitrogen, DBP, arsenic and indicator non-compliances. Information on the rest of the models, as well as complete partial dependance plots of the top 10 variables of all of them and prediction error maps can be found in the supplementary materials (SM Fig. 6, SM Figs. 7 to 16, and SM Fig. 17). Error maps reflect the balanced accuracy score from Table 3, and they are spatially distributed across the Spanish geography, with false negatives being predicted in more heterogenous models, like the one for indicator parameters and all health non-compliances. The rest of the models manage to predict true positives with a sensitivity that ranges from 0.46 (lead and copper) to 0.88 (pesticides).

Prediction maps for municipalities that have missing contaminant data for the years 2016–2021 for microbiological, nitrogen compounds, arsenic, and DBPs, and that have known abstraction origin can be found in SM Fig. 18. These maps show that, according to our models, 8.7 % of the municipalities without reporting data would have a DBP non-compliance. This percentage is higher in the case of nitrogen and arsenic non-compliance, with 50 % and 45 %, respectively, of non-reporting municipalities predicted to not comply with the drinking water quality regulations.

Overall, environmental variables are the most important ones when it comes to explaining non-compliance, as we can see them ranking among the top 10 (or even 5) drivers (Fig. 7). Out of them, climate-relate variables (temperature, precipitation, and precipitation-related variables like aquifer recharge) are more important to microbiological and nitrogen non-compliance than they are to the rest of contaminants. Microbiological and nitrogen non-compliances are also the ones affected by human activity, confined livestock farming and agriculture, respectively. It is worth noting that, with varying degrees, the number of

Table 4
Performance metrics of BRF for different types of non-compliance (see Table 1 for more details on the parameters that make each non-compliance type). Health parameter non-compliances include all the indented non-compliance types. Acronyms: Precision-Recall Area Under Curve (AUC PR); Balanced accuracy (ACC.); AUROC (Area Under the Receiver Operating Characteristic).

Non-compliance type	Number of municipalities	Baseline	AUC PR	ACC.	AUROC
Health parameters	5776	0.3175	0.5698	0.6510	0.7230
Microbiological	5767	0.1664	0.3737	0.6706	0.7267
Nitrogen	4877	0.0838	0.2913	0.7613	0.8254
Disinfection by-product	4650	0.1249	0.6897	0.8327	0.9027
Lead and copper	4731	0.0141	0.0813	0.5241	0.5979
Inorganic	4801	0.0226	0.1743	0.6197	0.6695
Organic	4614	0.0039	0.3637	0.6639	0.6928
Arsenic	4751	0.0334	0.2381	0.8065	0.8784
Pesticides	4565	0.0103	0.7464	0.7464	0.8600
Indicator parameters	5775	0.6332	0.8385	0.6518	0.7260

drought declarations is relevant for all contaminants. In the case of DBP non-compliance, the source of abstraction, which does not appear in the rest of the models, is the most important variable. For both DBP and arsenic violations, lithology, as well as demographic variables have higher relevance than for other contaminants, with annual average tap water expenditure per family being first driver when it comes to arsenic non-compliance. Indicator parameters, include several non-hazardous contaminants of diverse origin, which seems to translate to a diverse range of drivers, like tap water expenditure or presence of urban water authorized discharge points.

Going into more detail for each contaminant (Fig. 8), we can see that for microbiological non-compliances, high recharge, a high number of drought declarations, as well as low mean annual temperatures (Figs. 8 A, 8B, and 8C) contribute towards non-compliance. These results are consistent with the distribution of microbiological non-compliances presented in Fig. 5B, which are found in the more humid parts of Spain. A high density of confined livestock facilities (Fig. 8D), most of them of pig farms, within the catchment area also increases the probability of not complying with microbiological standards of drinking water quality.

In the case of nitrogen non-compliance, it is more likely to be found

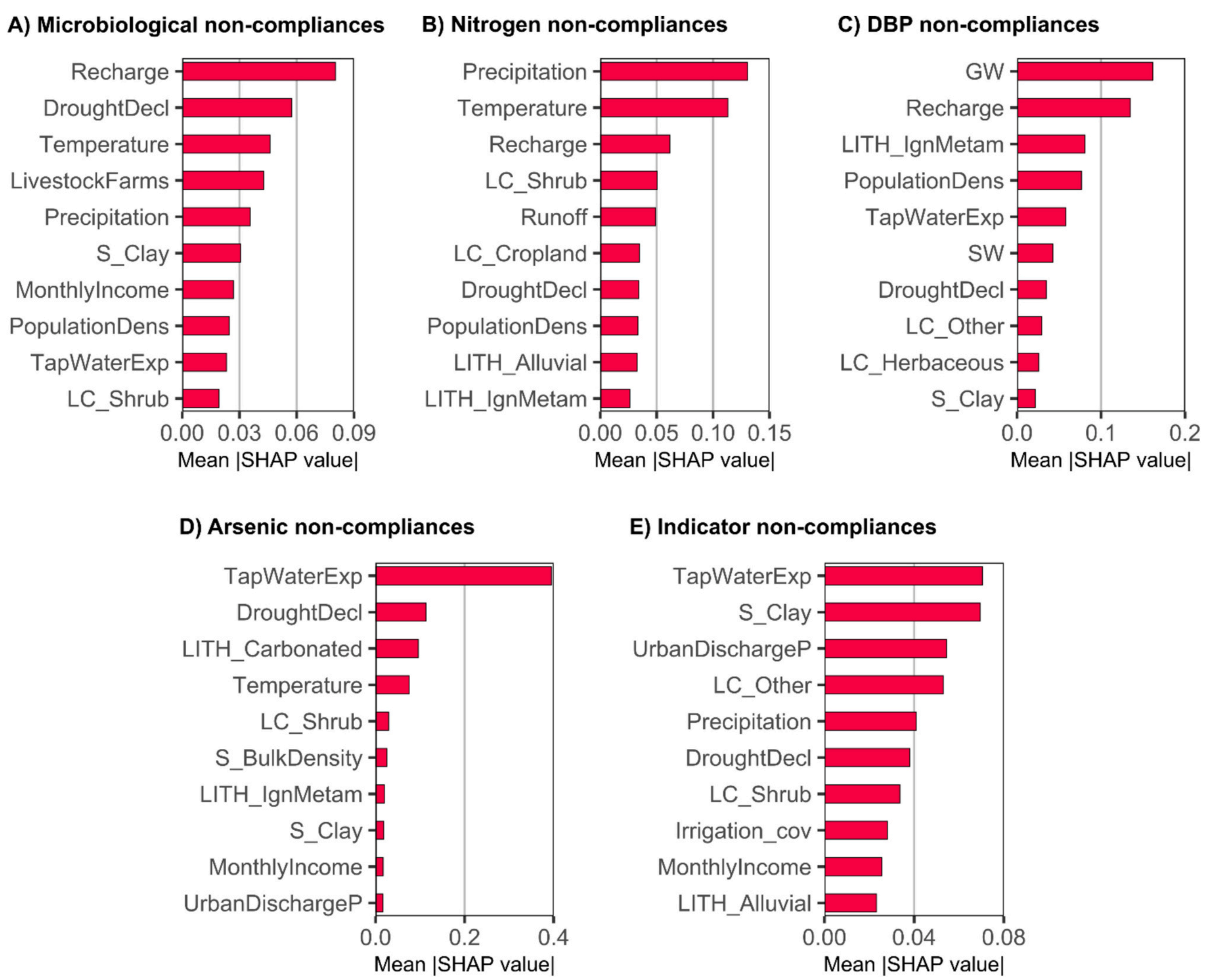


Fig. 7. Top ten driver violations for the main health non-compliances A) Microbiological; B) Nitrogen (nitrate and nitrite); C) Disinfection-by-product; D) Arsenic) and E) Indicator parameter non-compliances. Acronyms: drought declarations (DroughtDecl); subsoil clay content (S_Clay); tap water expenditure (TapWaterExp); population density (PopulationDens); % shrub cover (LC_Shrub); % cropland cover (LC_Cropland); % alluvial deposits (LITH_Alluvial); % igneous and metamorphic rocks (LITH_IgnMetam); surface water (SW); % other land cover (LC_Other); % herbaceous cover (LC_Herbaceous); % carbonated rocks (LITH_Carbonated); subsoil bulk density (S_BulkDensity); authorized urban discharge points (UrbanDischargeP); % irrigated land cover (Irrigation_cov).

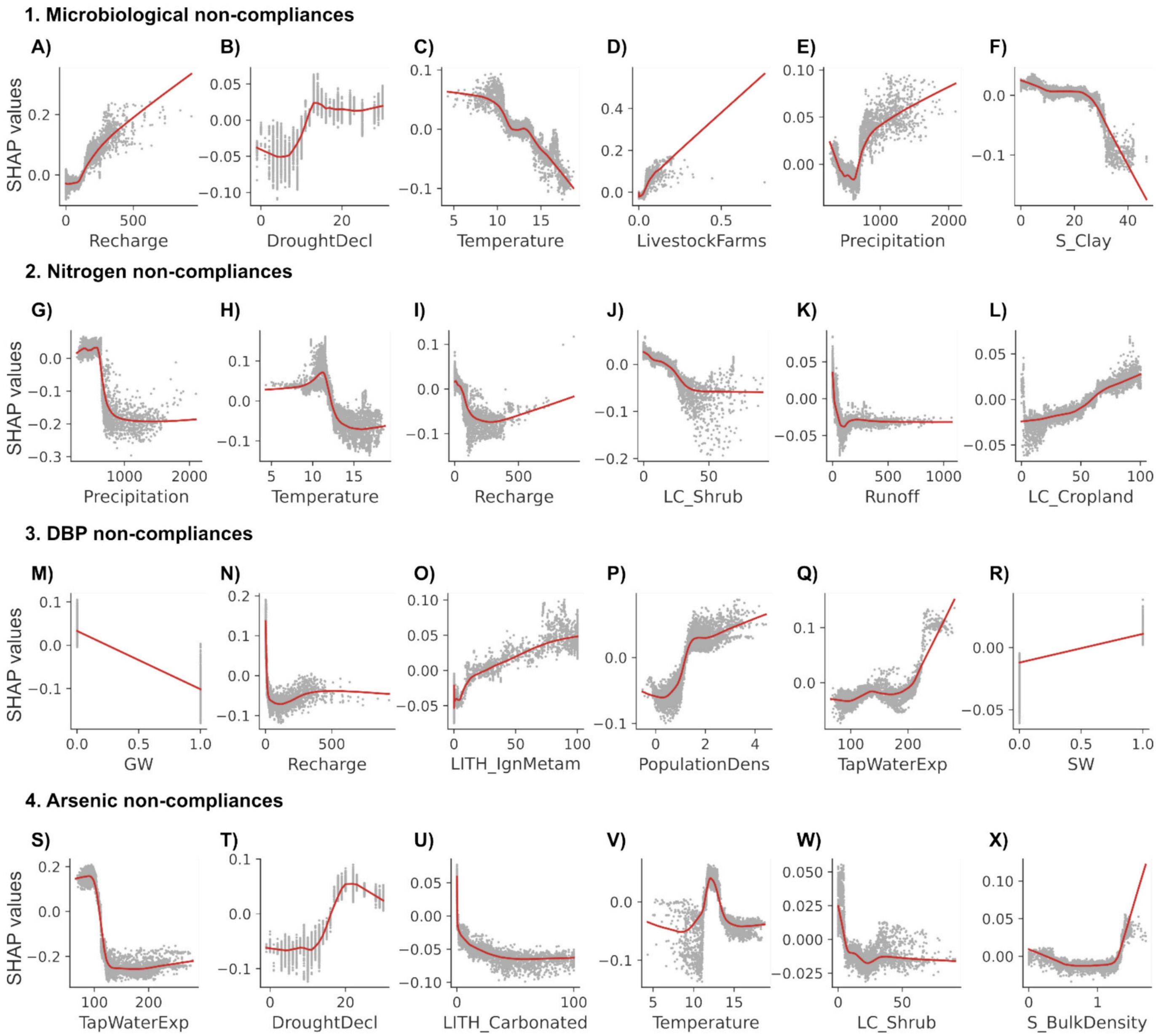


Fig. 8. Top 6 drivers for the main types of health non-compliances: microbiological (A-F), nitrogen (G-L), DBP (M-R) and arsenic (S-X). Acronyms: drought declarations (DroughtDecl); subsoil clay content (S_Clay); tap water expenditure (TapWaterExp); population density (PopulationDens); % shrub cover (LC_Shrub); % cropland cover (LC_Cropland); % alluvial deposits (LITH_Alluvial); % igneous and metamorphic rocks (LITH_IgnMetam); surface water (SW); % other land cover (LC_Other); % herbaceous cover (LC_Herbaceous); % carbonated rocks (LITH_Carbonated); subsoil bulk density (S_BulkDensity); authorized urban discharge points (UrbanDischargeP); % irrigated land cover (Irrigation_cov).

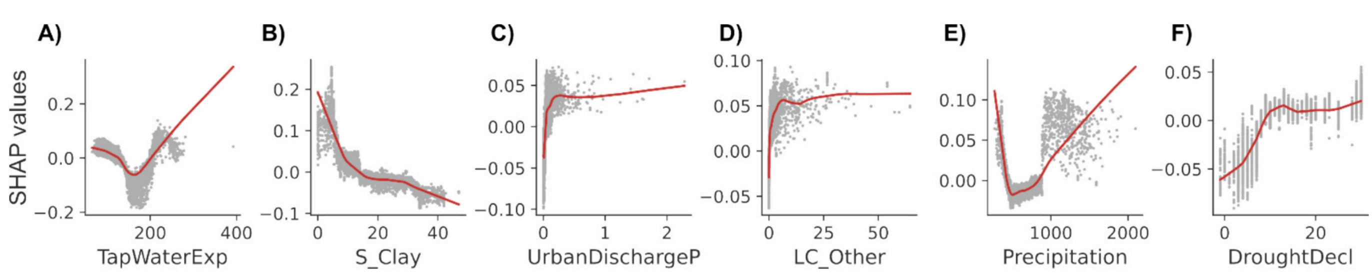


Fig. 9. Top 6 drivers for indicator parameter non-compliance. This model includes all the parameters classified as 'indicator' in Annex B of the Drinking Water Directive that have a parametric value in the SINAC database: Aluminum, Ammonium, Chloride, Color, Conductivity, pH, Iron, Manganese, Odor, Oxidisability, Sulphate, Sodium, Taste, Colony count at 22 °C, and Coliform bacteria. Acronyms: tap water expenditure (TapWaterExp); subsoil clay content (S_Clay); authorized urban discharge points (UrbanDischargeP); % other land cover (LC_Other); drought declarations (DroughtDecl).

in areas with low precipitation and relatively low temperatures (Figs. 8G and 8H). While low values of runoff and recharge seem to also be associated with a non-compliance of this type, the partial dependence plot shows a positive trend (Fig. 8I) that is consistent with alluvial lithology also being one of the drivers. According to the model output, a high coverage of cropland (above 60 %, Fig. 9 L) is associated with an increase in the probability of non-compliancen.

In the case of the DBPs, using surface water for abstraction is associated with non-compliance (i.e. not using groundwater, as shown in Fig. 8 M). Low recharge (Fig. 8 N) is also associated with DBP non-compliance, as well as high values of tap water expenditure and population density (Figs. 8P and 8Q). The latter implies that DBP non-compliances are more likely to happen in urban, or semi-urban, areas than in rural ones. Drought and a low percentage of clay content are also associated with a higher probability of not meeting the drinking water quality standards (SM Fig. 10G and 10 J).

For arsenic, low tap water expenditure is associated with a higher probability of non-compliance (Fig. 8S). A rich presence of carbonates also drives the model towards a low probability of not complying with the standards (Fig. 8 U), whereas a high number of drought declarations (Fig. 8 T) increases the probability of it. High values of bulk density are also associated with a higher non-compliance probability (Fig. 8X).

Moving onto indicator parameters, we can see that tap water expenditure, albeit not monotonic, is positively related to an indicator violation (Fig. 9 A; the probability of a violation escalates, after decreasing, at approximatively 190€/year). Low values of soil clay content are also associated with an increased probability of violation (Fig. 9B), as well as a high density of urban discharge points (Fig. 9C). There is also a relationship between a high number of drought declarations (Fig. 9F) and a higher probability of indicator violation.

4. Discussion

This is the first study of drinking water quality at the Spanish national scale that focuses on the differences between urban and rural areas and on the source area of the drinking water. Reports at the EU (see the latest DWD Synthesis report for the years 2017–2019; European Commission, 2024) and national level (e.g. Ministerio de Sanidad, 2022) focus on overall quality at a country scale. In the latter, the closest data disaggregation can be found as a number of analytics per drinking water supply zone size, but such supply zones are neither geographically localized, nor the results of monitoring discussed.

In EU synthesis reports, the threshold at which a country is considered non-compliant with a parameter or group of parameters is 1 % or more of samples above the parametric levels of the total collected ones (e.g. see European Commission, 2024a). Considering this, our results show that overall drinking water quality in Spain is good except for DBPs, arsenic and nitrogen compounds (especially nitrate). Regarding the temporal evolution of the non-compliance of certain parameters, they can be explained from at least two perspectives: a regulatory one, and one pertaining to external factors, where the state of the water body used for drinking water abstraction should be included. From a regulatory standpoint, we have the case of DPBs and lead. In the former, the introduction of new parameters to the monitoring resulted in an increase in the reporting of DBP non-compliances. In the case of lead, the stricter limits for this element that had to be complied with from 2014 onwards (from 25 μg/L to 10 μg /L, as stated in Annex I, Part B of the RDAC) required a replacement of lead pipelines by that year, which led to a drop in the number non-compliant samples. With regard to the external factors that also influence the evolution of non-compliance, we have the case DPBs, nitrogen, and indicator parameters. In 2020, coinciding with the outburst of the COVID-19 pandemic, DPBs experienced a decline in urban areas, with a steep increase the following year. That could be explained by the abundant use of disinfectants during the first year that were carried into the ecosystem via wastewater and stormwater (Parveen et al., 2022). On the other hand, nitrogen non-compliances spiked during the 2020 year, with one of the explanations being that the influx of chlorine disinfectant could have affected nitrogen removal during the sewage treatment process (Dang et al., 2023).

As good as the overall compliance might be, the more complex reality of drinking water quality remains masked when the data is aggregated. For instance, rural arsenic and nitrogen non-compliances are the ones contributing towards the total percentage of national non-compliance, whereas disinfection by-products violation ratios are high across all types of municipalities, but higher in urban areas when compared to rural ones. Moreover, arsenic, nitrogen compounds, and microbiological parameters have higher non-compliance rates in rural municipalities than in urban or semi-urban ones, a trend found across small water systems in rural areas, which are known to be especially vulnerable to those contaminants (Bain et al., 2014; WHO, 2011).

Rural municipalities are also the ones with poorer reporting, which is reflected in data gaps spread across the Spanish geography. This happens either because they fail to monitor or because, even if they monitor the quality, the resulting information is not being uploaded. Both explanations have emerged in informal talks with different regional health authorities and drinking water managers. They emphasized that the municipalities that do not report lack technical personnel, are oftentimes geographically isolated (i.e., mountainous regions), and generally have old and/or dwindling population with a distrust towards water treatments such as chlorination. González-Gómez et al. (2014) pointed out that reporting failure being more prevalent in smaller municipalities was related to fragmentation of management, and that implementing multi-municipal systems could reduce the number of municipalities that do not report.

As public sector decisions rely on good data (Gallaher and Heikkila, 2014; Laituri and Sternlieb, 2014), this lack of data has implications at a regulatory level, because it hinders the assessment of the extent of possible problems, which ultimately impact informed decision-making. Lack of water quality data is considered one of the barriers to the implementation of Sustainable Development Goal 6 "Ensure water and sanitation for all" (High-Level Panel on Water, 2018), one of which even Global North countries struggle with (Josset et al., 2019).

Lack of readily available data also affects locating the source of drinking water of the municipalities. Despite our best efforts, not all Spain could be covered because this information was not available for all the River Basin Districts. It should also be noted that official data might not necessarily match the reality, especially in isolated areas, where adaptation usually occurs at the sidelines of river basin management planning; one municipality might be listed as using surface water in the official registries but could have switched to groundwater abstraction because of a drought. For example, Molinero et al. (2011) mentions that rural water supply relies on groundwater in humid northwestern Spain, which matches the distribution of our maps, but they also emphasize that this dependence on groundwater is not reflected in the official inventories because of the dispersion of small wells and springs. With the new RBMP these inventories have improved, but the limitations of the data should still be taken into account.

Most of the population is supplied from surface water bodies. According to the latest EU drinking water synthesis report, Spain is the second country in the EU-27 in terms of abstracted surface water, with Ireland being first (European Commission, 2024). This is partly explained by the fact that Spain is the EU country with the highest number of dams and the fifth in the world (Spanish National Committee on Large Dams, 2022) with historical roots that stem from the centralization of water management in Spain from its onset (Llamas and Garrido, 2007). Other countries, for example Denmark, rely entirely on groundwater (Cabrera, 2024; European Commission, 2024). As per our results, Spanish rural municipalities are the ones that depend more on groundwater for drinking water provision, something found also by other authors or in older reports based on secondary data (Llamas and Garrido, 2007; Ministerio de Medio Ambiente, 2000). Groundwater use makes those municipalities more resilient to drought (Food and

Agriculture Organization, 2016), but more vulnerable to non-point source pollution, one of the biggest threats to groundwater quality in the EU (European Environment Agency, 2022).

Part of the population is serviced by abstracting water from water bodies that are in poor status according to the WFD standards; this poses a challenge in two directions. In the first place, it is a challenge in terms of availability, since there are water bodies in poor quantitative status that are used as drinking water source; this puts the population at risk of running out of water but can also possibly subject it to higher concentration of contaminants due to the low dilution (Boxall., 2011; Lasagna et al., 2013). Secondly, if the water body being used is declared as having poor chemical status, there is a potential health risk should the purification treatment fail or be insufficient. It is also important to note that there are parts of rural Spain where the drinking water is being obtained from sources that are not even monitored by RBDs as per the WFD guidelines. Such sources are intermittent streams, small local aquifers or springs, the former being one of the most vulnerable resources in the face of climate change due to changes in flow reductions (i.e., an increase in zero-flow days; T. Das et al., 2011; Döll and Schmied, 2012; Jaeger et al., 2014; McKerchar and Schmidt, 2007). There have already been cases in central Spain regarding municipalities that have been forced to switch to another resource due to the increased ephemerality of their streams and springs (F.C.V., 2022).

The importance of these statistics resides in the shift in the European Union from an end-of-pipe approach to a source-to-mouth approach, in line with multi-barrier approaches adopted in other countries (Andries et al., 2024), in which the state of the water source is seen as another part of the supply chain. This risk assessment approach necessarily requires examining the activities that occur in the catchment that lead to overexploitation (and ultimately, depletion) and pollution. Climate change will exacerbate these existing problems (Intergovernmental Panel on Climate Change, 2007; Whitehead et al., 2009), with it ultimately increasing the price of water, as it will need a more intensive treatment to be apt for human consumption or will need to come from alternative sources (e.g. desalination, water imports), with the consequent investments in infrastructure change (Cashman and Ashley, 2008).

To contribute to the implementation of the Drinking Water Directive from a risk assessment perspective, we investigated the spatial and demographic drivers that may lead to non-compliance with drinking water quality standards. Our model represents a first approach to this idea in Spain, and as such, some caveats must be presented. In the first place, while widely accepted, the SINAC database has been criticized due to its difficult access, missing data, and limited control over what is uploaded to the system. Despite this, SINAC is the only source of information regarding the quality of drinking water in Spain. Moreover, access to high-resolution data about relevant explanatory variables remains a challenge, despite improvements in recent years; important demographic data such as inequality indexes, water tariffs, or public/private ownership of the supply, distribution or treatment are not readily available. Regarding the nature of the missing data in SINAC, our prediction maps reveal that there might be unreported non-compliances, especially affecting Castille and Leon, where most of the non-reported data is localized, especially in rural areas (as seen in the population density map from Fig. 1).

The lack of data about the applied potabilization treatment makes it difficult to discern whether a municipality is not complying with the standards because of the treatment (or even blending of water) or because the source water is not contaminated. We tried to control the latter by adding the global status of the source of abstraction. However, as we have also mentioned, not all drinking water origins could be identified, which leaves important zones such as the Greater Metropolitan Area of Barcelona out of our model. While there is no data or analysis in Spain focused on the influence of treatment, it is known that arsenic violations have decreased in the United States partly due to the implementation of treatment techniques (Foster et al., 2019), and that treatment implementation (as in, nitrate removal) is effective for systems that were previously in nitrate violation (Pennino et al., 2017). The influence of treatment on models has been discussed by other authors (Pennino et al., 2020; Scanlon et al., 2022) who also faced lack of treatment data. Specifically, this data gap could skew the model towards a) false positives, i.e. predicting a high non-compliance risk, due to environmental conditions, in areas where treatment is eliminating the contaminant, or b) false negative, that is, a prediction of a low probability of non-compliance in areas where there are factors that should lead to quality non-compliance. In the case of our models, false negatives are rare, appearing more in heterogeneous models, but false positives are more widespread, with one of the reasons possibly being the existence of treatment or the ability to blend or switch the water source. Treatment has not been included in the model due to the lack of this information at the time database was compiled and the methodology being designed around the municipality as the study unit. Presently, information on the treatment used is presented per distribution network, and can be found on the SINAC webpage, as mandated by the 2020 DWD (European Commission, 2020).

Overall, Balanced Random Forest algorithm performed better when it had more data available and when the groups of parameters were homogeneous (e.g., nitrogen-based non-compliance, or disinfection-by-products). This is consistent with similar analysis performed by Scanlon et al. (2022) for the United States. Top drivers for each model we selected were also consistent with the mapped distribution of non-compliances across the Spanish geography, as shown in Fig. 5.

We have found that several climate and climate-derived variables have a great influence in the probability of non-compliance of microbiological (*E. coli*, *C. perfigens* and enterococci) and nitrogen-related parameters, with two anthropic activities contributing towards the probability of non-compliance: confined livestock farming, and agriculture, respectively. Manure from livestock is usually used as fertilizer, but its improper handling, storage and application can lead to pollution of water sources and drinking water (Gagliardi and Karns, 2000; Kumar et al., 2013). Pathogen survival in soil is known to increase with soil moisture (Mubiru et al., 2000) and low temperatures (Ogden et al., 2001), and *E. coli* leaching depends on the time that has passed from the moment of the application of manure and the rainfall event (Saini et al., 2003). Our maps show that microbiological non-compliance reported to SINAC is more widespread in northern Spain, where the climate is a bit colder than in the rest of the country, even if high temperatures are known to aid the proliferation of microorganisms (Blaustein et al., 2013). Michielssen et al. (2020) found that coliform violations in the United States were high in the Great Lakes region (another region with a relatively colder climate), and also associated them with smaller systems, which is the case of those in the north of Spain, as opposed to relatively bigger systems in southern Spain. Confined livestock farming is also very prevalent in northern Spain, as it can be seen in the density of European Pollutant Release and Transfer Register sites in SM Fig. 1. In the case of nitrogen, in groundwater, it has been found that at high temperatures, nitrate leaching is reduced due to increased evapotranspiration and biomass production, whereas high precipitation dilutes nitrate in the soil and fosters nitrogen uptake by the crop (Wick et al., 2012). Low values of runoff pushing the model towards non-compliance could mean most of the rainfall feeds aquifer recharge; this, paired with alluvial lithology also appearing as important, suggests that most of the nitrogen non-compliances come from use of groundwater, despite this variable not appearing in the top drivers. On the other hand, as we mentioned, the presence of agriculture in the catchment area contributes to non-compliance, which is consistent with the non-point pollution of agricultural origin that can be observed across all Europe (Esteban and Albiac, 2012).

In the case of DBP non-compliance, we find that the most important variables are related to the source of abstraction and its location. This is likely because DBPs are formed when chemical oxidants react with dissolved organic matter (Rougé et al., 2022), which in general has low concentrations in groundwater compared to inland surface waters (Harjung et al., 2023) and is very difficult to remove from reservoir water (Xu et al., 2022). Disinfection by-products can also form after the primary disinfection process has occurred, i.e. within the water distribution system (Dong et al., 2023). Factors that can influence this are the use of secondary chlorination due to the length of the distribution system (Jiang et al., 2017), the pipe materials (He et al., 2016; Ye et al., 2020), or the presence of biofilms within the distribution system pipes (Shi et al., 2022). Low recharge is also associated with DBP non-compliance, probably because of the low permeability of crystalline rocks, where most of the Spanish dams—which provide drinking water to urban populations—are located. According to our model, DBP non-compliances are more likely to happen in urban, or semiurban, areas than in rural ones, which contrasts with what happens in the United States, where rural populations have a higher likelihood of a DBP violation of the Safe Drinking Water Act (Allaire et al., 2018; Scanlon et al., 2022). Given the significance of climate change and increasing droughts in Spain (Vicente-Serrano et al., 2014), our findings indicate that drought conditions—identified as a driver of DBP non-compliance—can concentrate DBP precursors and nutrients in surface waters. These nutrient concentrations may also promote algal blooms, which are associated with certain types of DBP precursors (Xiao et al., 2023).

Despite the analysis of non-compliance evolutions indicating that arsenic is mostly a rural issue, our model does not point at lower population densities being a factor in the non-compliance for this contaminant. Groundwater use, mostly associated with geogenic arsenic, has also not appeared as an important variable. We expected this to appear as a variable because when the geospatial distribution of the breaches was cross-checked with the Arsenic Isovalues Map from the Geochemical Atlas of Spain (IGME, 2012), we found that they overlapped with the areas with greater As concentration in the soil, thus suggesting the geogenic origin of it. This is likely due to the spatial distribution of the municipalities affected by arsenic presence, which are very few, to which we should also add the complex mechanisms that lead to arsenic mobilization (Herath et al., 2016). Lower values of tap water expenditure associated with non-compliancecould be indicating that treatment techniques for arsenic removal (i.e., reverse osmosis) might not have been implemented, as they are costly, and arsenic is usually associated with rural areas where advanced water treatment techniques are not common. Prolonged drought has already been associated with a higher probability of exposure to arsenic (Lombard et al., 2021). The type of lithology is relevant to arsenic presence, as it appears either in sedimentary rocks or igneous ones, with the latter being the case in Spain. High values of soil bulk density being associated with a higher non-compliance probability is also consistent with the fact that at high bulk density values, the reduction of iron oxides responsible for As retention is triggered by rich organic matter, which leads to its mobilization (Tran et al., 2023).

The last of the models pertains to indicator parameters. While they are not necessarily hazardous, they are key to the consumer's perception of drinking water (De França Doria, 2010; Kelly and Pomfret, 1997). This could explain why tap water expenditure is important in our model, as the necessity to have acceptable and palatable tap water for the consumer can lead to more intensive treatments; for example, a one-percent increase in turbidity increases the costs in 0.1162 % (Danelon et al., 2021). Both drought and precipitations can put the water treatment facilities under stress, the former because contaminants are concentrated during drought episodes and the latter because, if surface water is used, it increases turbidity (Chou and Wu, 2010). The relationship between urban discharge points and indicator parameters should be further studied. We also found higher probability of indicator non-compliance in low-income brackets, which has also been reported in other parts of the world (Allaire et al., 2018).

A variable of low importance is the overall status of the water body, which, based on the findings of our exploratory analysis, we anticipated would have a greater influence on the probability of non-compliance. There could be multiple reasons for this and elucidating which ones are at work requires a case-by-case study. Reasons for this result range from imprecise water origin designation, as we have already discussed, to the dynamic nature of water and contamination, which is difficult to capture in a static model that groups 5 years of non-compliance. It could also mean that the technological paradigm, in which water treatment can make up for the pollution of the source, still applies in Spain.

Regarding the other variables, we want to emphasize that, with the exception of land use and anthropic activities like agriculture and confined livestock farming, the rest of them are not under the control of the authorities, be it River Basin Management or local and regional governments. This underscores the importance of spatial planning and land use management to protect the sources of drinking water, which should be done at local scale and considering that different types of municipalities face different challenges (e.g. microbiological pollution in rural areas vs disinfection by-products in high population density zones). As more data is made available to the public and scientists, more efficient models can be made to assess not only the catchment area, but also household economics and the nature of the water management of a municipality.

5. Conclusions

This paper presents a baseline of the state of the reporting of drinking water quality in Spain, the source of drinking water, and the factors that might influence quality non-compliance at catchment or municipal level. While further research is needed, especially one that should incorporate new demographics data that as of today are unavailable at a national level (e.g., ownership of the system, or water tariffs), some conclusions can be extracted from this study.

Overall, drinking water quality in Spain is good, but smaller municipalities from rural areas have a higher exposure of their domestic water supply to certain pollutants. This underscores the importance of disaggregating drinking water quality data, as aggregating national non-compliance totals masks problems that are specific to rural (like arsenic, nitrogen compounds as nitrate, or microbiological pathogens as *E. coli*) or urban municipalities (fundamentally, disinfection by-products). On top of this, rural areas also struggle with the mandatory reporting to the health authorities, which places them in a position of vulnerability, as lack of data hinders diagnoses and the development of solutions to possible problems. This could be attributed to a lack of resources—economic, but also of specialized personnel. While the importance of SINAC for the drinking water quality monitoring of Spain cannot be disputed, its opacity and the uneven enforcement of reporting obligations that have led to these data gaps point at a governance problem at its core.

Most of the Spanish population receives its drinking water from a surface source, which is the preferred source for urban areas. On the other hand, rural areas rely on groundwater and, sometimes, non-monitored sources such as local aquifers or local creeks—this, too, makes these zones more vulnerable to pollution, whereas the reliance of urban zones on surface water has implications in the light of climate change.

We also found that climate is also an important factor for non-compliance, including droughts. Temperature, precipitation, or recharge or the aforementioned droughts are factors over which local and regional authorities have little to no control. Our results also suggest that anthropic activities in the catchment area, such as agriculture and confined livestock farming—over which authorities have more control—are behind drinking quality non-compliance of nitrogen and microbiological nature. Moreover, disinfection by-product non-compliance is influenced by the source of abstraction, as surface water makes the municipality more susceptible to breach the quality standards. These results point to the importance of applying a risk-based approach to the catchment area and focusing management efforts to improve drinking water quality at the local scale, making it possible to target local factors and adapt to the local reality and its unique challenges.

CRediT authorship contribution statement

Delia M. Andries: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Alberto Garrido:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Lucia De Stefano:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by a Botín Foundation fellowship under its Water Observatory program. The authors want to thank the Spanish Ministry of Health for providing the full dataset used in this research. We are also grateful to Dr. Martínez-Muñoz and Dr. Gómez Escalonilla-Canales for their help with the machine learning models, and to the anonymous reviewers for their insightful suggestions and comments that helped improve this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2025.178412.

Data availability

Data will be made available on request.

References

Allaire, M., Wu, H., Lall, U., 2018. National trends in drinking water quality violations. Proc. Natl. Acad. Sci. 115 (9), 2078–2083. https://doi.org/10.1073/pnas.1719805115.

Anderson, M.W., 2010. *Cities inside out: race, poverty, and exclusion at the urban fringe* (SSRN scholarly paper 1560065). https://papers.ssrn.com/abstract=1560065.

Andries, D.M., Garrido, A., De Stefano, L., 2024. The evolution of the EU drinking water policy towards a source-to-mouth approach. International Journal of Water Resources Development, in press. https://doi.org/10.1080/07900627.2024.2318248.

Bain, R., Cronk, R., Hossain, R., Bonjour, S., Onda, K., Wright, J., Yang, H., Slaymaker, T., Hunter, P., Prüss-Ustün, A., Bartram, J., 2014. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. Trop. Med. Int. Health 19 (8), 917–927. https://doi.org/10.1111/tmi.12334.

Balazs, C.L., Ray, I., 2014. The drinking water disparities framework: on the origins and persistence of inequities in exposure. Am. J. Public Health 104 (4), 603–611. https://doi.org/10.2105/AJPH.2013.301664.

Blaustein, R.A., Pachepsky, Y., Hill, R.L., Shelton, D.R., Whelan, G., 2013. *Escherichia coli* survival in waters: temperature dependence. Water Res. 47 (2), 569–578. https://doi.org/10.1016/j.watres.2012.10.027.

Boxall., 2011. Hazardous Substances in Europe's Fresh and Marine Waters: A Overview. Publications Office of the European Union. https://doi.org/10.2800/78305.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Cabrera, E., 2024. *Directrices para una gestión sostenible del agua urbana* (1st Edition) (FACSA).

Cashman, A., Ashley, R., 2008. Costing the long-term demand for water sector infrastructure. Foresight 10 (3), 9–26. https://doi.org/10.1108/14636680810883099.

Centro de Estudios y Experimentación de Obras Públicas. (2020). *Evaluación de Recursos en Régimen Hídrico Natural en España (1940/41–2017/18)* [Raster map]. https://www.miteco.gob.es/en/agua/temas/evaluacion-de-los-recursos-hidricos/evaluacion-recursos-hidricos-regimen-natural.html.

Chou, F., Wu, C.-W., 2010. Reducing the impacts of flood-induced reservoir turbidity on a regional water supply system. Advances in Water Resources - ADV WATER RESOUR 33, 146–157. https://doi.org/10.1016/j.advwatres.2009.10.011.

Chybowski, B., Klimes, P., Cimbalnik, J., Travnicek, V., Nejedly, P., Pail, M., Peter-Derex, L., Hall, J., Dubeau, F., Jurak, P., Brazdil, M., Frauscher, B., 2024. Timing matters for accurate identification of the epileptogenic zone. Clin. Neurophysiol. 161, 1–9. https://doi.org/10.1016/j.clinph.2024.01.007.

Copernicus Land Monitoring Service, 2020. *Corine Land Cover 2018* (V2020_20u1) [Map; Vector 1:100000]. https://doi.org/10.2909/71c95a07-e296-44fc-b22b-415f42acfdf0.

Corada-Fernández, C., Candela, L., Torres-Fuentes, N., Pintado-Herrera, M.G., Paniw, M., González-Mazo, E., 2017. Effects of extreme rainfall events on the distribution of selected emerging contaminants in surface and groundwater: the Guadalete River basin (SW, Spain). Sci. Total Environ. 605–606, 770–783. https://doi.org/10.1016/j.scitotenv.2017.06.049.

Cutter, W.B., 2007. Valuing groundwater recharge in an urban context. Land Econ. 83 (2), 234–252. https://doi.org/10.3368/le.83.2.234.

Danelon, A.F., Augusto, F.G., Spolador, H.F.S., 2021. Water resource quality effects on water treatment costs: an analysis for the Brazilian case. Ecol. Econ. 188, 107134. https://doi.org/10.1016/j.ecolecon.2021.107134.

Dang, C., Zhang, Y., Zheng, M., Qiyue, M., Wang, J., Zhong, Y., Wu, Z., Liu, B., Fu, J., 2023. Effect of chlorine disinfectant influx on biological sewage treatment process under the COVID-19 pandemic: performance, mechanisms and implications. Water Res. 244, 120453. https://doi.org/10.1016/j.watres.2023.120453.

Das, R., Krishnakumar, A., Kumar, M.R., Thulseedharan, D., 2021. Water quality assessment of three tropical freshwater lakes of Kerala, SW India, with special reference to drinking water potential. Environmental Nanotechnology, Monitoring & Management 16, 100588. https://doi.org/10.1016/j.enmm.2021.100588.

Das, T., Pierce, D.W., Cayan, D.R., Vano, J.A., Lettenmaier, D.P., 2011. The importance of warm season warming to western U.S. streamflow changes. Geophys. Res. Lett. 38 (23). https://doi.org/10.1029/2011GL049660.

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In *proceedings of the 23rd international conference on machine learning, ACM* (Vol. 06). doi:https://doi.org/10.1145/1143844.1143874.

De França Doria, M., 2010. Factors influencing public perception of drinking water quality. Water Policy 12 (1), 1–19. https://doi.org/10.2166/wp.2009.051.

Döll, P., Schmied, H.M., 2012. How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. Environmental Research Letters 7 (1), 014037. https://doi.org/10.1088/1748-9326/7/1/014037.

Dong, F., Zhu, J., Li, J., Fu, C., He, G., Lin, Q., Li, C., Song, S., 2023. The occurrence, formation and transformation of disinfection byproducts in the water distribution system: a review. Sci. Total Environ. 867, 161497. https://doi.org/10.1016/j.scitotenv.2023.161497.

Erisman, J.W., van Grinsven, H., Grizzetti, B., Bouraoui, F., Powlson, D., Sutton, M.A., Bleeker, A., Reis, S., 2011. The European nitrogen problem in a global perspective. In: Bleeker, A., Grizzetti, B., Howard, C.M., Billen, G., van Grinsven, H., Erisman, J. W., Sutton, M.A., Grennfelt, P. (Eds.), The European Nitrogen Assessment: Sources, Effects and Policy Perspectives. Cambridge University Press, pp. 9–31. https://doi.org/10.1017/CBO9780511976988.005.

Esteban, E., Albiac, J., 2012. Assessment of nonpoint pollution instruments: the case of Spanish agriculture. International Journal of Water Resources Development 28(1), Article 1. https://doi.org/10.1080/07900627.2012.640878.

European Commission. (2011, February 22). Draft minutes of the meeting of the Committee under Directive 98/83/EC (Drinking Water Committee). https://circabc.europa.eu/ui/group/65764c73-4a57-45dc-8199-473014cf65bf/library/2cc31937-0a2e-40ed-97ef-f4fcb48246b3/details.

European Commission, 2014. REPORT FROM THE COMMISSION synthesis report on the quality of drinking water in the EU examining the member States' reports for the period 2008-2010 under directive 98/83/EC (COM/2014/0363). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2014%3A0363%3AFIN.

European Commission, 2016. *COMMISSION STAFF WORKING DOCUMENT REFIT EVALUATION of the drinking water directive 98/83/EC* (SWD/2016/428; Commission Staff Working document). https://ec.europa.eu/environment/water/water-drink/pdf/SWD_2016_428_F1.pdf.

European Commission, 2019. *COMMISSION STAFF WORKING DOCUMENT EVALUATION of the council directive 91/271/EEC of 21 may 1991, concerning urban waste-water treatment* (Commission Staff Working document SWD/2019/0700). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019SC0700&qid=1690378833708.

European Commission, 2020. Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the Quality of Water Intended for Human Consumption (Recast) (Text with EEA Relevance), CONSIL, EP, 435 OJ L. http://data.europa.eu/eli/dir/2020/2184/oj/eng.

European Commission, 2024a. Synthesis Report on the Quality of Drinking Water in the European Union examining Member States' reports for the 2014–2016 period, foreseen under Article 13(5) of Directive 98/83/EC. European Commission. https://circabc.europa.eu/ui/group/65764c73-4a57-45dc-8199-473014cf65bf/library/ef1e69cb-a703-4d09-a312-cf5853fd8396/details.

European Commission, 2024. Synthesis report on the quality of drinking water in the European Union examining member States' reports for the 2017-2019 period, foreseen under article 13(5) of directive 98/83/EC. https://circabc.europa.eu/ui/group/65764c73-4a57-45dc-8199-473014cf65bf/library/c65cc53d-fc1c-47b2-b5a0-0ebc853d9d43/details.

European Environment Agency, 2020. *European pollutant release and transfer register (E-PRTR)* [CSV]. https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-23/european-pollutant-release-and-transfer-register-e-prtr-data-base/eprtr_v9_csv.zip.

European Environment Agency, 2022. Europe's groundwater—a key resource under pressure. https://www.eea.europa.eu/publications/europes-groundwater/europes-groundwater.

F.C.V. (2022, August 21). Los ribereños ya tienen a su alcance su 'sueño' de beber agua del Tajo. Nueva Alcarria. https://nuevaalcarria.com/articulos/los-riberenos-ya-tienen-a-su-alcance-su-sueno-de-beber-agua-del-tajo_1.

Food and Agriculture Organization, 2016. Global Diagnostic on Groundwater Governance (Food and Agriculture Organization of the United Nations).

Foster, S.A., Pennino, M.J., Compton, J.E., Leibowitz, S.G., Kile, M.L., 2019. Arsenic drinking water violations decreased across the United States following revision of the maximum contaminant level. Environ. Sci. Technol. 53 (19), 11478–11485. https://doi.org/10.1021/acs.est.9b02358.

Gagliardi, J.V., Karns, J.S., 2000. Leaching of Escherichia coli O157:H7 in diverse soils under various agricultural management practices. Appl. Environ. Microbiol. 66 (3), 877–883. https://doi.org/10.1128/AEM.66.3.877-883.2000.

Gallaher, S., Heikkila, T., 2014. Challenges and opportunities for collecting and sharing data on water governance institutions. Journal of Contemporary Water Research and Education 153, 66–78. https://doi.org/10.1111/j.1936-704X.2014.03181.x.

Ganga Devi, S.V.S., 2020. Analysing ground water quality in the regions of Kadapa District using supervised learning methods. In: Jyothi, S., Mamatha, D.M., Satapathy, S.C., Raju, K.S., Favorskaya, M.N. (Eds.), Advances in Computational and Bio-Engineering. Springer International Publishing, pp. 305–313. https://doi.org/10.1007/978-3-030-46943-6_34.

Geris, J., Comte, J.-C., Franchi, F., Petros, A.K., Tirivarombo, S., Selepeng, A.T., Villholth, K.G., 2022. Surface water-groundwater interactions and local land use control water quality impacts of extreme rainfall and flooding in a vulnerable semi-arid region of sub-Saharan Africa. J. Hydrol. 609, 127834. https://doi.org/10.1016/j.jhydrol.2022.127834.

González-Gómez, F., García-Rubio, M.A., González-Martínez, J., 2014. Beyond the public–private controversy in urban water management in Spain. Util. Policy 31, 1–9. https://doi.org/10.1016/j.jup.2014.07.004.

Grizzetti, B., Bouraoui, F., Billen, G., van Grinsven, H., Cardoso, A.C., Thieu, V., Garnier, J., Curtis, C., Howarth, R., Johnes, P., 2011. Nitrogen as a threat to European water quality. In: Bleeker, A., Grizzetti, B., Howard, C.M., Billen, G., van Grinsven, H., Erisman, J.W., Sutton, M.A., Grennfelt, P. (Eds.), The European Nitrogen Assessment: Sources, Effects and Policy Perspectives. Cambridge University Press, pp. 379–404. https://doi.org/10.1017/CBO9780511976988.020.

Halleux, V. (2023). Pollutants in EU waters: Update of chemical substances listed for control | Think Tank | European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)749772.

Harjung, A., Schweichhart, J., Rasch, G., Griebler, C., 2023. Large-scale study on groundwater dissolved organic matter reveals a strong heterogeneity and a complex microbial footprint. Sci. Total Environ. 854, 158542. https://doi.org/10.1016/j.scitotenv.2022.158542.

He, G., Li, C., Dong, F., Zhang, T., Chen, L., Cizmas, L., Sharma, V.K., 2016. Chloramines in a pilot-scale water distribution system: transformation of 17β-estradiol and formation of disinfection byproducts. Water Res. 106, 41–50. https://doi.org/10.1016/j.watres.2016.09.047.

Herath, I., Vithanage, M., Bundschuh, J., Maity, J.P., Bhattacharya, P., 2016. Natural arsenic in global Groundwaters: distribution and geochemical triggers for mobilization. Curr. Pollut. Rep. 2 (1), 68–89. https://doi.org/10.1007/s40726-016-0028-2.

Hiederer, R., 2013. Mapping soil properties for Europe: spatial representation of soil database attributes. Publications Office of the European Union. https://doi.org/10.2788/94128.

High-Level Panel on Water, 2018. Making every Drop Count: An Agenda for Water Action. High-Level Panel on Water Outcome Document, United Nations. https://sustainabledevelopment.un.org/content/documents/17825HLPW_Outcome.pdf.

Instituto Geográfico Nacional, 2014. *Sistema de Ocupación del Suelo de España (SIOSE)* (Version 2014) [Map; Vector 1:25000]. https://www.siose.es/web/guest/base-de-datos.

Instituto Geológico y Minero de España. (2012). Mapa de isovalores de Arsénico. Atlas Geoquímico de España. [Map; 1km grid]. https://catalogo.igme.es/geonetwork/srv/api/records/ESPIGMEATLASGEOQUIMICAAS20210531.

Instituto Geológico y Minero de España. (2015). *Mapa Geológico de la Península Ibérica, Baleares y Canarias a escala 1:1.000.000* (Version 2015) [Map; Vector 1:100000]. https://info.igme.es/cartografiadigital/geologica/Geologicos1MMapa.aspx?Id=Geologico1000_(2015)&language=es#metadatos.

Instituto Geológico y Minero de España. (n.d.). *Base de datos de puntos del agua* (2.0) [Dataset]. http://info.igme.es/BDAguas/.

Instituto Nacional de Estadística, 2020a. *Encuesta de Presupuestos Familiares 2019* [CSV]. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176806&menu=resultados&idp=1254735976608.

Instituto Nacional de Estadística, 2020b. *Nomenclátor o Población del Padrón Continuo por unidad poblacional* [Excel]. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177010&menu=resultados&idp=1254734710990#_tabs-1254736195532.

Intergovernmental Panel on Climate Change, 2007. Climate Change 2007 - the Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC. Cambridge University Press.

Jaeger, K.L., Olden, J.D., Pelland, N.A., 2014. Climate change poised to threaten hydrologic connectivity and endemic fishes in dryland streams. Proc. Natl. Acad. Sci. 111 (38), 13894–13899. https://doi.org/10.1073/pnas.1320890111.

Jiang, J., Zhang, X., Zhu, X., Li, Y., 2017. Removal of intermediate aromatic halogenated DBPs by activated carbon adsorption: a new approach to controlling halogenated DBPs in chlorinated drinking water. Environ. Sci. Technol. 51 (6), 3435–3444. https://doi.org/10.1021/acs.est.6b06161.

Josset, L., Allaire, M., Hayek, C., Rising, J., Thomas, C., Lall, U., 2019. The U.S. water data gap—a survey of state-level water data platforms to inform the development of a National Water Portal. Earth's Future 7 (4), 433–449. https://doi.org/10.1029/2018EF001063.

Kelly, M.G., Pomfret, J.R., 1997. Tastes and odours in potable water: Perception versus reality. In: Sutcliffe, D. (Ed.), The Microbiological Quality of Water. Freshwater Biological Association, pp. 71–80.

Kristensen, P., Whalley, C., Nery, F., Christiansen, T., Schmedtje, U., Solheim, A., Austnes, K., Kampa, E., Rouillard, J., Prchalová, H., Klančnik, K., Völker, J., Peterlin, M., Fribourg-Blanc, B., Prins, T., Kodes, V., Persson, J., Baltas, E., Bariamis, G., 2018. 2018 EEA European waters assessment (Report No. 7/2018). European Environment Agency. https://www.eea.europa.eu/publications/state-of-water.

Kumar, R.R., Park, B.J., Cho, J.Y., 2013. Application and environmental risks of livestock manure. Journal of the Korean Society for Applied Biological Chemistry 56 (5), 497–503. https://doi.org/10.1007/s13765-013-3184-8.

Laituri, M., Sternlieb, F., 2014. Water data systems: science, practice, and policy. Journal of Contemporary Water Research & Education 153 (1), 1–3. https://doi.org/10.1111/j.1936-704X.2014.03174.x.

Lasagna, M., De Luca, D.A., Debernardi, L., Clemente, P., 2013. Effect of the dilution process on the attenuation of contaminants in aquifers. Environ. Earth Sci. 70 (6), 2767–2784. https://doi.org/10.1007/s12665-013-2336-9.

Lee, D., Gibson, J.M., Brown, J., Habtewold, J., Murphy, H.M., 2023. Burden of disease from contaminated drinking water in countries with high access to safely managed water: a systematic review. Water Res. 242, 120244. https://doi.org/10.1016/j.watres.2023.120244.

Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res. 18 (17), 1–5. http://jmlr.org/papers/v18/16-365.html.

Li, P., Wu, J., 2019. Drinking water quality and public health. Expo. Health 11 (2), 73–79. https://doi.org/10.1007/s12403-019-00299-8.

Llamas, R., Garrido, A., 2007. Lessons from Intensive Groundwater Use in Spain: Economic and Social Benefits and Conflicts.

Lombard, M.A., Daniel, J., Jeddy, Z., Hay, L.E., Ayotte, J.D., 2021. Assessing the impact of drought on arsenic exposure from private domestic Wells in the conterminous United States. Environ. Sci. Technol. 55 (3), 1822–1831. https://doi.org/10.1021/acs.est.9b05835.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2019). *Explainable AI for trees: from local explanations to global understanding* (arXiv:1905.04610; version 1). arXiv. Doi:10.48550/arXiv.1905.04610.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Proces. Syst. 30. In: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

McKerchar, A.I., Schmidt, J., 2007. Decreases in low flows in the lower Selwyn River? J. Hydrol. N. Z. 46 (2), 63–72.

Meehan, K., Jepson, W., Harris, L.M., Wutich, A., Beresford, M., Fencl, A., London, J., Pierce, G., Radonic, L., Wells, C., Wilson, N.J., Adams, E.A., Arsenault, R., Brewis, A., Harrington, V., Lambrinidou, Y., McGregor, D., Patrick, R., Pauli, B., Young, S., 2020. Exposing the myths of household water insecurity in the global north: a critical review. WIREs Water 7 (6), e1486. https://doi.org/10.1002/wat2.1486.

Michielssen, S., Vedrin, M.C., Guikema, S.D., 2020. Trends in microbiological drinking water quality violations across the United States. Environ. Sci.: Water Res. Technol. 6 (11), 3091–3105. https://doi.org/10.1039/D0EW00710B.

Ministerio de Medio Ambiente, 2000. Water in Spain. Secretaría de Estado de Aguas y Costas, Dirección General de Obras Hidráulicas y Calidad de las Aguas. https://www.miteco.gob.es/content/dam/miteco/es/agua/temas/planificacion-hidrologica/water-in-spain_tcm30-527170.pdf.

Ministerio de Sanidad, 2022. *Calidad del agua de consumo en España 2022* 17; Calidad Del Agua de Consumo En España, p. In: 239. Ministerio de Sanidad, Secretaría General Técnica. https://www.sanidad.gob.es/areas/sanidadAmbiental/calidadAguas/aguasConsumoHumano/publicaciones/docs/2022_INFORME_ANUAL_AGUAS_DE_CONSUMO.pdf.

Ministerio para la Transición Ecológica y el Reto Demográfico. (2006). *Subcuencas de los cauces de la red hidrográfica básica* [Vector 1:25000]. https://www.mapama.gob.es/ide/metadatos/index.html?srv=metadata.show&uuid=e89e9aee-1fcb-4ae8-83d5-a9eca2d98597.

Ministerio para la Transición Ecológica y el Reto Demográfico. (2023a). *Censo Nacional de Vertidos* [Shapefile 1:25000]. https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/censo-nacional-vertidos.html.

Ministerio para la Transición Ecológica y el Reto Demográfico. (2023b). *Informes y mapas de seguimiento de la sequía en España* [Dataset]. https://www.miteco.gob.es/es/agua/temas/observatorio-nacional-de-la-sequia/informes-mapas-seguimiento.html.

Ministerio para la Transición Ecológica y el Reto Demográfico, 2023c. Plan Especial de Sequía de la Demarcación Hidrográfica del Miño-Sil. Ministerio para la Transición Ecológica y el Reto Demográfico. https://www.chminosil.es/es/chms/planificacionhidrologica/nuevo-plan-especial-de-sequia.

Ministerio para la Transición Ecológica y el Reto Demográfico, 2023d. Plan Especial de Sequía de la Demarcación Hidrográfica del Segura. Ministerio para la Transición Ecológica y el Reto Demográfico. https://www.chsegura.es/es/cuenca/planificacion/plan-especial-sequias/.

Ministerio para la Transición Ecológica y el Reto Demográfico. (2023e). Planes Hidrológicos del tercer ciclo de planificación (2022–2027). https://www.miteco.go b.es/es/agua/temas/planificacion-hidrologica/planificacion-hidrologica/pphh_tercer_ciclo.html.

Molinero, J., Sahuquillo, A., Llamas, M., 2011. Groundwater in Spain: Legal Framework and Management Issues. Groundwater Management Practices, Balkema, Leiden, The Netherlands.

Mubiru, D.N., Coyne, M.S., Grove, J.H., 2000. Mortality of Escherichia coli O157:H7 in two soils with different physical and chemical properties. J. Environ. Qual. 29 (6), 1821–1825. https://doi.org/10.2134/jeq2000.00472425002900060012x.

Ogden, I.D., Fenlon, D.R., Vinten, A.J.A., Lewis, D., 2001. The fate of *Escherichia coli* O157 in soil and its potential to contaminate drinking water. Int. J. Food Microbiol. 66 (1), 111–117. https://doi.org/10.1016/S0168-1605(00)00508-0.

Okun, D.A., 2003. Drinking water and public health protection. In: Drinking Water Regulation and Health. John Wiley & Sons, Ltd., pp. 1–24. https://doi.org/10.1002/0471721999.ch1

Osorio, V., Proia, L., Ricart, M., Pérez, S., Ginebreda, A., Cortina, J.L., Sabater, S., Barceló, D., 2014. Hydrological variation modulates pharmaceutical levels and biofilm responses in a Mediterranean river. Sci. Total Environ. 472, 1052–1061. https://doi.org/10.1016/j.scitotenv.2013.11.069.

Parveen, N., Chowdhury, S., Goel, S., 2022. Environmental impacts of the widespread use of chlorine-based disinfectants during the COVID-19 pandemic. Environ. Sci. Pollut. Res. Int. 29 (57), 85742–85760. https://doi.org/10.1007/s11356-021-18316-2.

Paun, I., Chiriac, F.L., Marin, N., Cruceru, L., Pascu, L., Lehr, C., Ene, C., 2017. Water quality index, a useful tool for Evaluation of Danube River raw water. Rev. Chim. 68, 1732–1739. https://doi.org/10.37358/RC.17.8.5754.

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, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12 (85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html.

Pennino, M.J., Compton, J.E., Leibowitz, S.G., 2017. Trends in drinking water nitrate violations across the United States. Environ. Sci. Technol. 51 (22), Article 22. https://doi.org/10.1021/acs.est.7b04269.

Pennino, M.J., Leibowitz, S.G., Compton, J.E., Hill, R.A., Sabo, R.D., 2020. Patterns and predictions of drinking water nitrate violations across the conterminous United States. Sci. Total Environ. 722, 137661. https://doi.org/10.1016/j.scitotenv.2020.137661.

Ricart, M., Guasch, H., Barceló, D., Brix, R., Conceição, M. H., Geiszinger, A., Alda, M. J. L. de, López-Doval, J. C., Muñoz, I., Postigo, C., Romaní, A. M., Villagrasa, M., & Sabater, S. (2010). Primary and complex stressors in polluted mediterranean rivers: pesticide effects on biological communities. J. Hydrol., 383(1), 52–61. doi:https://doi.org/10.1016/j.jhydrol.2009.08.014.

Rigatti, S.J., 2017. Random Forest. Journal of Insurance Medicine 47 (1), 31–39. https://doi.org/10.17849/insm-47-01-31-39.1.

Roig, B., Baures, E., Olivier, T., 2014. Perspectives on drinking water monitoring for small scale water systems. Water Sci. Technol. Water Supply 14 (1), 1–12. https://doi.org/10.2166/ws.2013.211.

Rougé, V., Lee, Y., von Gunten, U., Allard, S., 2022. Kinetic and mechanistic understanding of chlorite oxidation during chlorination: optimization of ClO2 pre-oxidation for disinfection byproduct control. Water Res. 220, 118515. https://doi.org/10.1016/j.watres.2022.118515.

Saini, R., Halverson, L.J., Lorimor, J.C., 2003. Rainfall timing and frequency influence on leaching of Escherichia coli RS2G through soil following manure application. J. Environ. Qual. 32 (5), 1865–1872. https://doi.org/10.2134/jeq2003.1865.

Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10 (3), e0118432. https://doi.org/10.1371/journal.pone.0118432.

Scanlon, B.R., Fakhreddine, S., Reedy, R.C., Yang, Q., Malito, J.G., 2022. Drivers of spatiotemporal variability in drinking water quality in the United States. Environ. Sci. Technol. 56 (18), 12965–12974. https://doi.org/10.1021/acs.est.1c08697.

Secretaría de Estado de Política Territorial, 2017. *Encuesta de Infraestructura y Equipamientos Locales* [Dataset]. https://mpt.gob.es/portal/politica-territorial/local/coop_econom_local_estado_fondos_europeos/eiel.html.

Shi, X., Clark, G.G., Huang, C., Nguyen, T.H., Yuan, B., 2022. Chlorine decay and disinfection by-products formation during chlorination of biofilms formed with simulated drinking water containing corrosion inhibitors. Sci. Total Environ. 815, 152763. https://doi.org/10.1016/j.scitotenv.2021.152763.

Sjerps, R.M.A., ter Laak, T.L., Zwolsman, G.J.J.G., 2017. Projected impact of climate change and chemical emissions on the water quality of the European rivers Rhine and Meuse: a drinking water perspective. Sci. Total Environ. 601–602, 1682–1694. https://doi.org/10.1016/j.scitotenv.2017.05.250.

Smedley, P.L., Kinniburgh, D.G., 2002. A review of the source, behaviour and distribution of arsenic in natural waters. Appl. Geochem. 17 (5), 517–568. https://doi.org/10.1016/S0883-2927(02)00018-5.

Spanish National Committee on Large Dams, 2022. *Governance of dams and reservoirs. Technical Committee on Water Resources Planning Engineer Activities*. https://www.spancold.org/wp-content/uploads/2022/06/Gobernanza-de-Presas-y-Embalses-EN-US.pdf.

Tesoriero, A.J., Gronberg, J.A., Juckem, P.F., Miller, M.P., Austin, B.P., 2017. Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification. Water Resour. Res. 53 (8), 7316–7331. https://doi.org/10.1002/2016WR020197.

Tran, T.H.H., Kim, S.H., Lee, H., Jo, H.Y., Chung, J., Lee, S., 2023. Variable effects of soil organic matter on arsenic behavior in the vadose zone under different bulk densities. J. Hazard. Mater. 447, 130826. https://doi.org/10.1016/j.jhazmat.2023.130826.

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, *11*(5), article 5. doi:https://doi.org/10.3390/w11050910.

van Halem, D., Bakker, S.A., Amy, G.L., van Dijk, J.C., 2009. Arsenic in drinking water: a worldwide water quality concern for water supply companies. Drinking Water Engineering and Science 2 (1), 29–34. https://doi.org/10.5194/dwes-2-29-2009.

Vicente-Serrano, S.M., Lopez-Moreno, J.-I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J.M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., Espejo, F., 2014. Evidence of increasing drought severity caused by temperature rise in southern Europe. Environ. Res. Lett. 9 (4), 044001. https://doi.org/10.1088/1748-9326/9/4/044001.

Wang, F., Wang, Y., Zhang, K., Hu, M., Weng, Q., Zhang, H., 2021. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. Environ. Res. 202, 111660. https://doi.org/10.1016/j.envres.2021.111660.

Whitehead, P.G., Wilby, R.L., Battarbee, R.W., Kernan, M., Wade, A.J., 2009. A review of the potential impacts of climate change on surface water quality. Hydrol. Sci. J. 54 (1), 101–123. https://doi.org/10.1623/hysj.54.1.101.

WHO, 2011. Small-Scale Water Supplies in the pan-European Region: Background, Challenges, Improvements. World Health Organization, Regional Office for Europe. https://apps.who.int/iris/handle/10665/326401.

Wick, K., Heumesser, C., Schmid, E., 2012. Groundwater nitrate contamination: factors and indicators. J. Environ. Manag. 111, 178–186. https://doi.org/10.1016/j.jenvman.2012.06.030.

Wuijts, S., Claessens, J., Farrow, L., Doody, D.G., Klages, S., Christophoridis, C., Cvejić, R., Glavan, M., Nesheim, I., Platjouw, F., Wright, I., Rowbottom, J., Graversgaard, M., van den Brink, C., Leitão, I., Ferreira, A., Boekhold, S., 2021. Protection of drinking water resources from agricultural pressures: effectiveness of EU regulations in the context of local realities. J. Environ. Manag. 287, 112270. https://doi.org/10.1016/j.jenvman.2021.112270.

Xiao, R., Deng, Y., Xu, Z., Chu, W., 2023. Disinfection byproducts and their precursors in drinking water sources: origins, influencing factors, and environmental insights. Engineering. https://doi.org/10.1016/j.eng.2023.08.017.

Xu, T., Zhang, W., Gómez-Hernández, J.J., Xie, Y., Yang, J., Chen, Z., Lu, C., 2022. Non-point contaminant source identification in an aquifer using the ensemble smoother with multiple data assimilation. J. Hydrol. 606, 127405. https://doi.org/10.1016/j.jhydrol.2021.127405.

Yard, E.E., Murphy, M.W., Schneeberger, C., Narayanan, J., Hoo, E., Freiman, A., Lewis, L.S., Hill, V.R., 2014. Microbial and chemical contamination during and after flooding in the Ohio River—Kentucky, 2011. J. Environ. Sci. Health A 49 (11), 1236–1243. https://doi.org/10.1080/10934529.2014.910036.

Ye, X., Wang, P., Wu, Y., Zhou, Y., Sheng, Y., Lao, K., 2020. Microplastic acts as a vector for contaminants: the release behavior of dibutyl phthalate from polyvinyl chloride pipe fragments in water phase. Environ. Sci. Pollut. Res. 27 (33), 42082–42091. https://doi.org/10.1007/s11356-020-10136-0.

Zavareh, M., Maggioni, V., Zhang, X., 2024. Assessing the efficiency of a random forest regression model for estimating water quality indicators. Meteorology Hydrology and Water Management 11 (2), 52–69. https://doi.org/10.26491/mhwm/183734.

Zwarteveen, M., Kemerink-Seyoum, J.S., Kooy, M., Evers, J., Guerrero, T.A., Batubara, B., Biza, A., Boakye-Ansah, A., Faber, S., Cabrera Flamini, A., Cuadrado-Quesada, G., Fantini, E., Gupta, J., Hasan, S., ter Horst, R., Jamali, H., Jaspers, F., Obani, P., Schwartz, K., Wesselink, A., 2017. Engaging with the politics of water governance. WIREs Water 4 (6), e1245. https://doi.org/10.1002/wat2.1245.