Fluoride contamination of groundwater resources in Ghana: Country-wide hazard modeling and estimated population at risk

Dahyann Araya [a,*], Joel Podgorski [a], Michael Kumi [b], Patrick A. Mainoo [b], Michael Berg [a,*]

[a] Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland
[b] CSIR-Water Research Institute, P. O. Box AH 38, Achimota, Accra, Ghana

ABSTRACT

Most people in Ghana have no or only basic access to safely managed water. Especially in rural areas, much of the population relies on groundwater for drinking, which can be contaminated with fluoride and lead to dental fluorosis. Children under the age of two are particularly susceptible to the adverse effects of fluoride and can retain 80–90% of a fluoride dose, compared to 60% in adults. Despite numerous local studies, no spatially continuous picture exists of the fluoride contamination across Ghana, nor is there any estimate of what proportion of the population is potentially exposed to unsafe fluoride levels. Here, we spatially model the probability of fluoride concentrations exceeding 1.0 mg/L in groundwater across Ghana to identify risk areas and estimate the number of children and adults exposed to unsafe fluoride levels in drinking water. We use a set of geospatial predictor variables with random forest modeling and evaluate the model performance through spatial cross-validation. We found that approximately 15% of the area of Ghana, mainly in the northeast, has a high probability of fluoride contamination. The total at-risk population is about 920,000 persons, or 3% of the population, with an estimated 240,000 children (0–9 years) in at-risk areas. In some districts, such as Karaga, Gushiegu, Tamale and Mion, 4 out of 10 children are potentially exposed to fluoride poisoning. Geology and high evapotranspiration are the main drivers of fluoride enrichment in groundwater. Consequently, climate change might put even greater pressure on the area's water resources. Our hazard maps should raise awareness and understanding of geogenic fluoride contamination in Ghana and can advise decision making at local levels to avoid or mitigate fluoride-related risks.

1. Introduction

Only 41% of the population in Ghana has access to safely managed water services and another 44% has only basic access, while the remaining 14% has limited to no access (WHO/UNICEF Joint Monitoring Program, 2020). In rural areas, and especially in the northern regions of the country, the population therefore depends directly on groundwater wells for drinking (Atipoka, 2009). However, the groundwater in Ghana is often naturally contaminated with fluoride, and its consumption can lead to dental fluorosis (Alfredo et al., 2014; Barbier et al., 2010; Smedley et al., 1995). The World Health Organization (WHO) has established a guideline maximum fluoride concentration of 1.5 mg/L for drinking water, which is based on an estimated drinking water intake of 2 L/day for adults, and, emphasizes that countries should identify locally relevant standards based on their characteristics (WHO, 2017). In countries with generally high

temperatures, such as Ghana where water consumption and thus exposure are higher, a lower threshold has been suggested, as is the case in India (Bureau of Indian Standards (BIS, 2012). Apambire et al. (1997) reported that the year-round high temperatures in the Upper Region of Northern Ghana, averaging 32 °C, can lead to a daily water intake as high as 3–4 L/day for adults; this is 1.5–2 times higher than the WHO-estimated daily water consumption (WHO, 2017). Therefore, Apambire et al. (1997) recommended an upper safe threshold of fluoride concentration in drinking water of 1.0 mg/L.

The high exposure of the population in Ghana to fluoride led Craig et al. (2015) to recommend an age-specific threshold for fluoride intake. The risk of developing non-carcinogenic effects, such as permanent dental fluorosis, is greater at an early age during the development of tooth enamel. Young children can also retain 80–90% of a given fluoride dose compared to 60% in adults (WHO, 2004), with about 90% of the fluoride absorbed from liquids compared to 30–40% absorbed from

---

solids (WHO, 1996). Hence, Craig et al. (2015) recommended an upper threshold of fluoride in drinking water of 1.0 mg/L for older children and adults, and an even lower level (0.6 mg/L) in the first two years of life (Kumar et al., 2019; Zango et al., 2019). This is particularly relevant for Ghana, where children (0–9 years old) account for approximately 25% of the total population (GSS, 2020). By contrast, in European countries children represent only about 12% of the population. Therefore, identifying children who are potentially exposed to unsafe fluoride levels in drinking water is critical in Ghana.

The spatial distribution of the fluoride contamination in groundwater throughout Ghana is related to the geological, topographical, and climatic characteristics of the country (Sunkari et al., 2020). Where groundwater testing has not been carried out, spatial environmental and socioeconomic variables are instrumental predictors for identifying areas at risk (Amini et al., 2008; Podgorski and Berg, 2020). Recently, machine-learning methods, such as random forest, have been adopted for spatial analysis for their advantages in evaluating complex and non-linear spatial relationships where a large number of variables might be needed to explain a given spatial phenomenon (Lary et al., 2016). However, for machine learning methods such as random forest, challenges arise when attempting to account for dependency structures associated with spatial data that differ from those of non-spatial datasets (Meyer et al., 2012; Pohjankukka et al., 2017; Valavi et al., 2019).

One challenge is that spatial autocorrelation (SAC) between variables can bias and inflate results; even though the phenomenon is a natural one occurring where related variables provide the same information in a given area. Another is that using traditional non-spatial validation methods to maintain the hypothesis of spatial independence is difficult throughout the validation process. These methods create a random set of validation points that may be very close to the training points, resulting in an overly optimistic view of a model's performance (Ploton et al., 2020). The use of spatial cross-validation can be valuable for maintaining independence between the training and testing data to help ensure reliable performance estimations in the context of spatial modeling (Araujo et al., 2005), thereby avoiding inflated estimates of prediction accuracy associated with traditional cross-validation methods (e.g., k-folds) (Hammond and Verbyla, 1996; Misiuk et al., 2019). Spatial cross-validation avoids problems related to inflated prediction accuracy caused by spatial autocorrelation by choosing a sufficiently large block size that avoids spatial autocorrelation between variables and sufficiently represents the classes (Roberts et al., 2017). Similarly, the area of applicability (AOA) (also called applicability domain) identifies the areas where the model has knowledge of environmental conditions, which are set by the range of the training data. The AOA is a valuable tool to identify areas where the model is able to learn from the training data and where its estimated performance is still reliable (Meyer and Pebesma, 2021).

Despite numerous local studies on fluoride contamination of groundwater in Ghana (Affam et al., 2012; Alfredo et al., 2014; Anim-Gyampo et al., 2012; Anku et al., 2009; Apambire, 1996; Atipoka, 2009; Craig et al., 2018; Firempong et al., 2013; Ganyaglo et al., 2019; Loh et al., 2012; Smedley et al., 2002; Yidana et al., 2012; Zango et al., 2019), no comprehensive assessment of fluoride hazard and risk yet exists for the whole country, including estimates of the number of children and adults potentially exposed to unsafe levels. Therefore, in the present study, we spatially model the probability of the occurrence of geogenic fluoride concentrations greater than 1.0 mg/L in groundwater in Ghana. A set of geospatial predictor variables with random forests is used. We address the abovementioned shortcomings of using random forests in a spatial context by evaluating the model performance through state-of-the-art spatial cross-validation and by assessing model uncertainties. We then use the resulting hazard map of fluoride contamination to estimate the population and the children potentially exposed to fluoride levels in drinking water that may affect their health. This hazard map and estimation of the at-risk population will be invaluable for guiding location-based, risk-reduction policy and intervention changes to improve the future outcomes of children living in high-risk fluoride contamination areas.

2. Materials and methods

2.1. Study area

Ghana is located in West Africa at the Gulf of Guinea, between $1°20'$ east to $3°25'$ west and $4°50'$ to $11°18'$ north (Fig. 1). Its population reached 31 million inhabitants in 2020. The country has a generally equatorial tropical climate in the south and a semi-arid climate in the north. The south has relatively high and stable temperatures throughout the year, with an average daily temperature ranging from 21–30 °C. The northeast has the most elevated temperatures in the country, reaching 35–40 °C during the hottest months of the year from February to April. Rainfall decreases with increasing latitude, ranging from about 1900 mm per year in the southwest to about 800 mm in the north (Asiamah et al., 1997).

The geology of the country is dominated by Voltaic, Birimian, and granitic geological formations in the north, while the south is dominated by Birimian and Middle Precambrian rock formations. An extended description of the geology is provided in Fig. S1. The state of the soil is strongly influenced by the climate of the country. In the north, Leptosols and Plinthosols are observed, which are semi-arid and poor soils. While in the south, the most characteristic soils are lateritic, such as Ochrosols and Oxysols, which are soils normally found in tropical rainforests (Jones et al., 2013). The topography of the country is gentle, with a low plain on the coast and an extensive plateau in the south-central part of the country and highlands mainly in the west. Less than 1% of the territory has slopes greater than 5%, and the altitude varies from sea level to 885 m on Mount Afadjato.

2.2. Fluoride data and predictor variables

The fluoride concentration dataset consisted of 611 of our own groundwater quality data measurements as well as 2623 collected from other sources (Abusa et al., 2018; Addo et al., 2011; Affam et al., 2012; Agyemang, 2020; Anornu et al., 2017; Arko et al., 2019; Avi et al., 2019; Boakye Opoku, 2013; Chegbeleh et al., 2020; CIDA, 2011; Egbi et al., 2019; Gastineau, 2015; Kulinkina et al., 2017; Mensah-Essilfie, 2013; Nkansah et al., 2019; Smedley, 1996; Smedley et al., 2002; Zango et al., 2019). More information is provided in supplementary Table S1. Of these data, 13% of the wells exceed the guideline value of 1.5 mg/L recommended by the WHO, and 22% exceed 1.0 mg/L, the guideline for tropical countries like Ghana with a higher than average water consumption and, therefore, greater exposure (Apambire et al., 1997; Craig et al., 2015). Sixty-five geospatial predictor variables of climate, geology, soil, topography, and ecology were collected based on their known or potential relationships with high fluoride levels in groundwater. Fig. 2 presents a selection of these explanatory variables, while supplementary Fig. S2 shows the 19 predictor variables ultimately selected. The spatial resolution of most of the explanatory variables was the same at 250 m, and a common spatial analysis unit was created by assigning the point data of fluoride concentrations to a grid of 250 m^2. When more than one point in a pixel was available, the geometric mean was calculated. The data were then categorized as zero for concentrations ≤1.0 mg/L, (i.e., class zero) and one for concentrations >1.0 mg/L, (i.e., class one).

2.3. Random forest modeling

Random forest is an ensemble method based on classification and regression trees that use recursive binary splitting to split a dataset into two sets for the selection of the optimal variables (Breiman, 2001). A reliable prediction is ensured by growing a large number of dissimilar trees based on a random resampling of the original data and a random
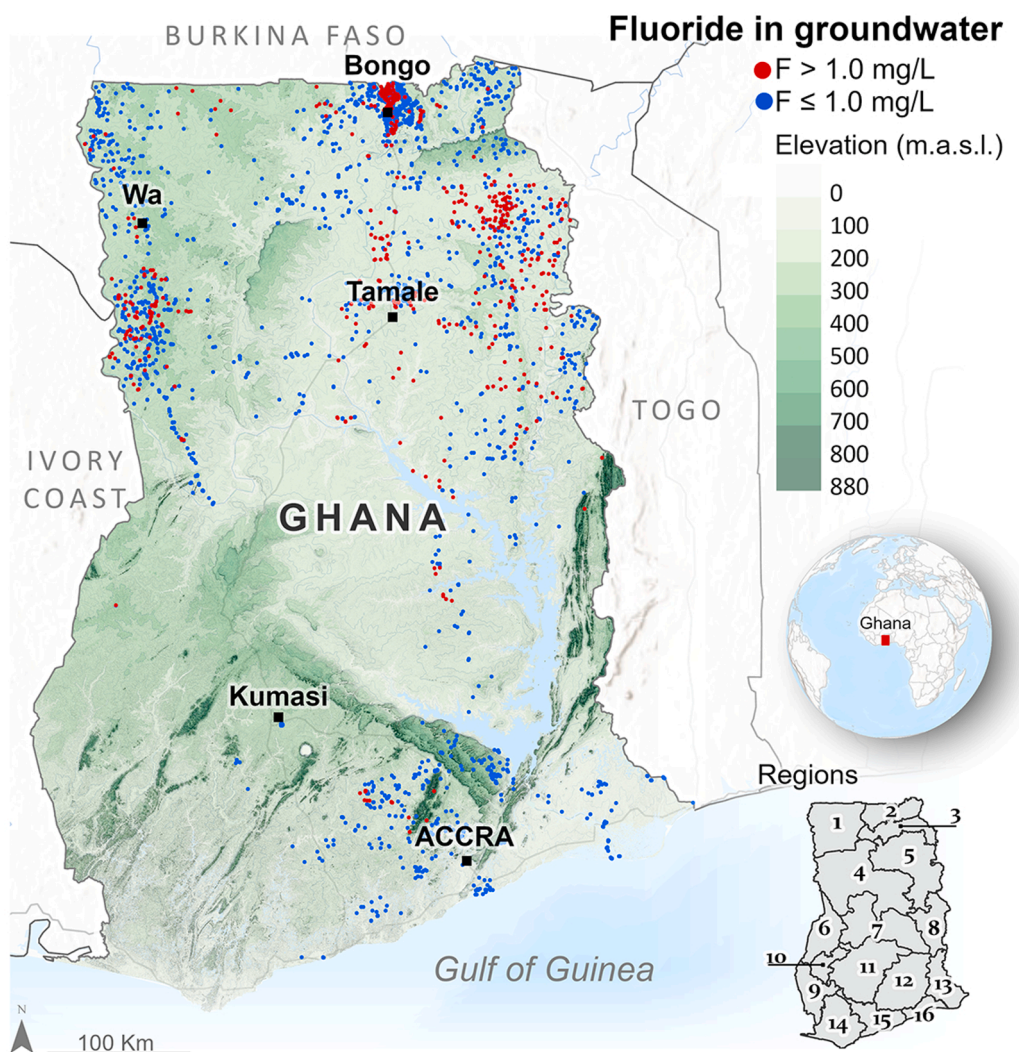
Fig. 1. Fluoride concentrations in groundwater from the data sources listed in supplementary Table S1 (n = 3234), plotted on a topographic map of Ghana. Red points represent fluoride concentrations >1.0 mg/L (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).
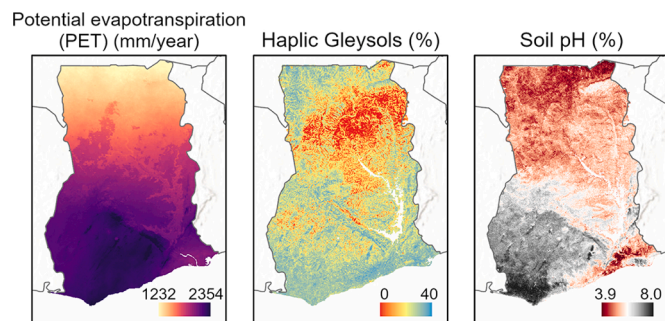
Fig. 2. Examples of predictor variables of the country of Ghana that were used in modeling, available at spatial resolutions of 250 m and 1 km; potential evapotranspiration (Trabucco and Zomer, 2019), haplic gleysols and pH (Hengl et al., 2015). The complete set of explanatory variables is shown in Fig. S2.

subset of the variables. The final prediction is the result of the average of all the trees. The present random forest model was created using the R statistical programming language (R Core Team, 2020).

Any bias produced by splitting data into particular training and testing data sets was avoided by performing 1000 iterations with different subsets of training and testing data. Since a large disparity in class frequencies can negatively influence the performance of a model, the data set was balanced for each iteration by randomly down-sampling the majority class in the training set to match the least frequent class (i. e., class one or sample points with a fluoride concentration over the threshold of 1.0 mg/L). Each model was produced by growing 1001 trees. The 19 predictors variables used in modeling were selected based on their contribution to the overall model performance as measured by accuracy. Therefore, variables that had a mean decrease in accuracy >0 on average over all random forest model iterations were retained. That is, only variables that systematically improved the predicted accuracy were kept.

2.4. Spatial cross-validation of the model

The performance of the model was verified on the test data by spatial cross-validation with the R package blockCV (Valavi et al., 2019). This step avoids the bias that occurs when the data used to train the model are not spatially independent from the data used to validate the model (Ambroise and McLachlan, 2002; Pohjankukka et al., 2017). The spatial independence of the training and validation data was achieved by spatially separating the training and validation data sets using 28 blocks of approximately 81 × 81 km in size. The block size was defined to ensure that both classes had sufficient representation (Roberts et al., 2017). Using a semivariogram of the predictor variables, the block size was also used to assess and avoid spatial autocorrelation of those variables. An example of the fold assignment and the semivariogram of the predictor variables is shown in Fig. S3.

The sensitivity, specificity, balanced accuracy, precision, Brier score, receiver operating characteristic (ROC) curve, and area under the curve (AUC) were calculated for cut-off probabilities between 0 and 1. Sensitivity provides the proportion of high measured fluoride concentrations that are correctly identified by the model. Conversely, specificity provides the proportion of low measured fluoride concentrations that are correctly identified by the model. Balanced accuracy describes the overall performance of the model through an average of sensitivity and specificity, while precision reports on the proportion of cases that were labelled with fluoride contamination that are indeed contaminated with fluoride. The Brier score verifies the accuracy of the prediction by showing how far the predictions are from the truth; Brier score values closer to zero indicate a better calibration of the model. The ROC depicts the relationship between the false positive rate (1-specificity) and sensitivity. The AUC measures the area under the ROC curve, with values closer to one indicating a better predictive capability of the model.

A better understanding and interpretation of the random forest model results were achieved using the importance of variables and partial dependence plots (PDP). The influence of the predictor variables on the prediction outcome of the model was determined by calculating the importance of the variables, although this still does not relate the predictor variable to the prediction outcome. For this reason, partial dependence plots (PDPs) were also calculated (Greenwell, 2017), which show the effect that changes in an explanatory variable have on the prediction.

2.5. Hazard map

The hazard probability map of fluoride in groundwater exceeding 1.0 mg/L in Ghana was calculated using the average of 1000 model predictions. Each run was created through the random forest model built with the final predictor variables in the compiled dataset of low and high fluoride concentrations. The averaged map is less sensitive to any arbitrary selection produced by the specific splitting of the training and test data sets. We then classified the fluoride hazard in groundwater as high or low using the average probability cut-off, which was taken at the point at which the specificity and sensitivity of a model are equal. In a balanced sample between low and high values, this usually corresponds to the highest overall accuracy over all cut-off points. We then created two additional maps that classify the fluoride hazard in groundwater as high or low using the minimum and maximum probability cut-off points from the 1000 iterations. The standard deviation for each pixel for the three hazard maps was then calculated. This represents the confidence in the prediction of the value of each pixel over the entire study area as well as the spatial stability of the models. The spatially uncertain areas were calculated using the AOA with the R package CAST (Meyer and Pebesma, 2021). The AOA is the area where the model can be applied with an expected average performance that is comparable to that estimated with the training data. In areas outside the AOA, the model predictions are more uncertain, as these areas contain predictor values

not found within the limits of the training data. The AOA is derived from the dissimilarity index (DI), which provides a threshold to define the AOA. The DI uses a unitless measure to determine how much each point outside the training data differs from the training data. For this, it evaluates the distances in a multidimensional predictor variable space by weighting the predictor variables by their importance derived from the random forest model. The AOA is determined by the 0.95 quantile threshold of the DI values of the training data. Using this threshold, a new data point is outside the AOA when the DI exceeds the 0.95 quantile. Subsequently, the pixels with a high standard deviation and uncertainty were removed from the final hazard map of high fluoride contamination. The indication of the regions of less certainty provides valuable information for the final interpretation of the results.

2.6. Estimation of the at-risk population

Our use of the global population age structure for the year 2020 (WorldPop, 2018) allowed us to identify the total population and children aged 0–9 years living in areas with groundwater fluoride concentrations exceeding 1.0 mg/L. Because not all the population relies on groundwater, we used national-level groundwater usage rates for urban and rural areas (WHO/UNICEF Joint Monitoring Program (JMP), 2019). Finally, to identify urban and rural areas in the country, we used the GHS-SMOD 06 of the Joint Research Centre (JRC) of the European Commission, which indicates the degree of urbanization in 2020 based on Landsat image data and global population grids (Pesaresi et al., 2019). We then estimated the potentially affected population by multiplying the pixel's groundwater-consuming population by its probability of having a high risk of fluoride contamination, as previously described (Podgorski and Berg, 2020). The population at risk was then broken down by district and region across the country.

3. Results

3.1. Predictor variables

We modeled the probability of groundwater fluoride exceeding 1.0 mg/L in Ghana by initially considering 65 geospatial variables. From these, 19 variables were kept, which are plotted in Fig. 3a according to their decreasing relative importance. The final set of variables is presented in descending order according to their importance in Fig. 3a. This diverse set of variables includes climatic, geological, soil, and topographic variables. The effects of geologic features and potential evapotranspiration (PET) on the probability model of high fluoride are shown in Fig. 3b. Note, however, that the model performance (see Section 3.2) is derived from the combination of explanatory variables representing the complexity and diversity of the system, and not just from any particular variable. The partial dependence plots provided in supplemental material Fig. S4 provide further insights into the relationships between the predictor variables and fluoride contamination.

3.2. Hazard map

The hazard probability map of groundwater fluoride contamination generated from the final random forest model is presented in Fig. 4. Several regions of high probability are evident across the country. The northeast is the most affected region, with probabilities varying between 50 and 90%. This region extends from Eastern Gonja to the northernmost district of Bongo, excluding a low-probability belt between the districts of Mamprugu Moagduri and Bunkpurugu Yonyo. In the northwest, the Sawla Tuna Kalba and Bole districts have a probability of around 50–70% of groundwater fluoride contamination. And an isolated area in the south near Accra has a probability of around 50–60%. As indicated by the histogram in Fig. 4, 1.2% and 16.5% of the country have a probability >80% and >50%, respectively, of having fluoride in groundwater exceeding 1.0 mg/L.

a)   Predictor variable importance

b)   Partial dependence plots
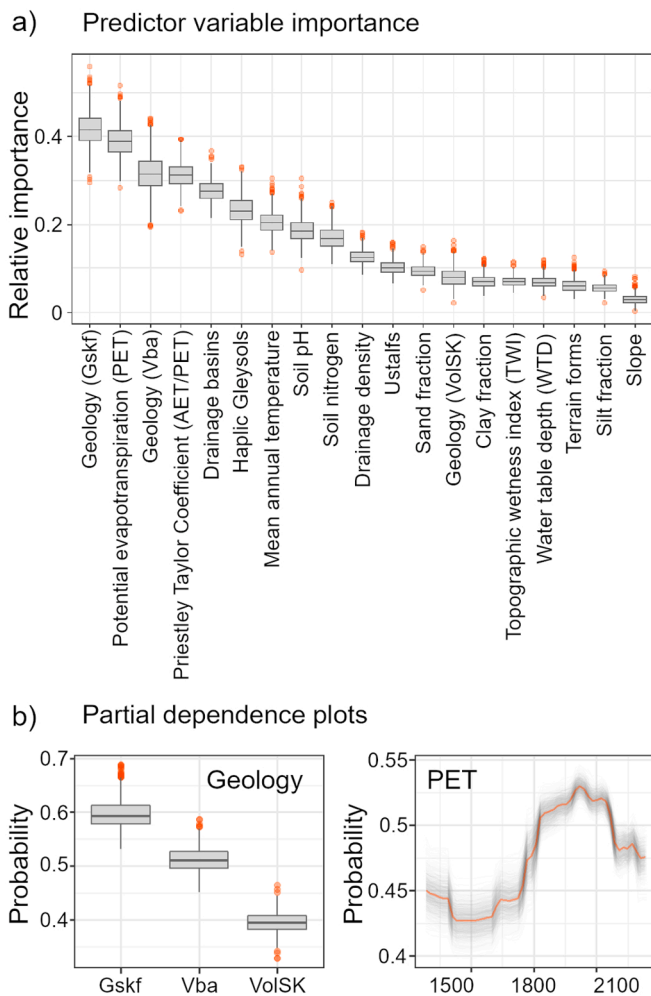
Fig. 3. Measures of the importance and partial dependence plots (PDP) of random forest variables. (a) Distribution of the relative importance of the 19 final predictor variables from 1000 iterations. Partial dependence plots (b) of the predictor variables of geological groups, and potential evapotranspiration (PET) (see supplementary Fig. S4 for more PDPs). The partial dependence represents the probability of exceeding a groundwater fluoride concentration of 1.0 mg/L. (Gskf = Eburnean Plutonic Suite – K-feldspar-rich granitoid, mainly granite and monzonite; Vba = Voltaian Supergroup, Oti-Pendjari Group – Sandstone, mudstone, siltstone, and carbonate; VolSK = Voltaian Supergroup, Kwahu-Morago – Sandstone). The box boundaries indicate the 25th percentile and the 75th percentile, while the middle line represents the 50th percentile or median, the whiskers represent 1.5 times the interquartile range and the orange points are outliers (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

3.3. Performance of the model

No significant variation beyond ~2% across the metrics was detected among the 1000 different models runs. The corresponding 1000 ROC curves shown in Fig. 4 have a standard deviation of 0.012, demonstrating consistent accuracies ranging between AUC values of 0.72 and 0.81 (mean 0.76). Sensitivity, specificity, precision, and balanced accuracy all have values around 70% at the optimal cutoff. For precision, for example, this means that the prediction was correct 7 out of 10 times where the model indicated a fluoride concentration exceeding the threshold. These metrics show that the model produced stable predictions across the 1000 iterations. The Brier score, which reports the quadratic error of probability and has a range of -∞ to 1, where values close to 0 represent a more accurate model, is around 0.2. Overall, these results indicate a good calibration of the model.

3.4. Estimation of the at-risk population

About 15.6% or ~37,300 km² of the total land area of Ghana may contain fluoride contamination above the threshold. These areas are mainly in the northeastern part of the country where 24% of Ghana's districts are located. To calculate the number of people potentially affected by high fluoride in drinking water, the population living in these hazard areas was adjusted by the proportion of untreated groundwater use. The total population that has a risk of ingesting high fluoride concentrations in drinking water is around 920,000 (Fig. 5a), or 3% of the population. Fig. 5c shows that Karaga, Gushiegu, Yendi, and Savelugu Nanton have the highest populations of potentially exposed individuals, comprising about 42% of those potentially exposed at the district level.

4. Discussion

4.1. Predictive variables

Geogenic fluoride contamination is a complex process, and our study of its predictors confirms that high fluoride concentrations in groundwater depend on an interplay of a number of variables (Amini et al., 2008; Frencken, 1992). Though all play a role, some variables do contribute more in predicting contamination, and we found the geology, climate, and soils of the area to be the main drivers of high fluoride levels (Fig. 3). This is consistent with high fluoride levels that are widely reported in the Bongo area, which are associated with the Eburnearn Supergroup and its K-feldspar-rich granitoids of mainly granite and monzonite (Apambire et al., 1997; Smedley et al., 1995). The highest levels of fluoride have also been reported in the rocks of the Voltaian Supergroup, Oti-Pendjari Group, which are composed of sandstone, mudstone, siltstone, and carbonate. In agreement with our findings on the role of climate, high fluoride levels have also been reported in more arid conditions, such as those observed in northern Ghana (see the partial dependence plot of PET in Fig. 3b). Here, higher temperatures and evaporation rates lead to high fluoride concentrations (CIDA, 2011; Edmunds and Smedley, 2013), which is especially apparent during the dry season when fluoride concentrations increase (Alfredo et al., 2014; Malago, 2017). This is a problem in areas like northern Ghana, where wells closed for safety reasons due to their high fluoride concentration have been reopened in periods of water stress to alleviate water shortages (Craig et al., 2015; Ganyaglo et al., 2019). In the face of future expected higher temperatures and increased water stress related to climate change in northern Ghana (EPA, 2020), we expect even more pressure and new challenges related to water safety for the population in that area.

4.2. Spatial cross-validation and model performance

The uncertainties of our model were determined using spatial cross-validation. This led to stable and reliable metrics and maps across all iterations. However, spatial cross-validation can also downplay the results by not accounting for the predictive capability of the model for combinations of predictor data that are similar but spatially distant from the training data. The possibility also exists that by separating the blocks for training, the diversity of training data is inadvertently reduced (see Fig. S3). We attempted to moderate this effect through the careful implementation of blocks and multiple iterations. This agrees with the literature consensus, which also suggests that a block-based spatial validation approach to building a predictive model is most appropriate (Roberts et al., 2017). Moreover, this approach bolsters the credibility of the resulting maps.

Overall, we reduced the biases produced by an unbalanced training sample by creating a balanced sample for the model by randomly down-sampling all classes in the training set to match the minority class (i.e., class one). Leaving this unbalanced, however, means that the majority
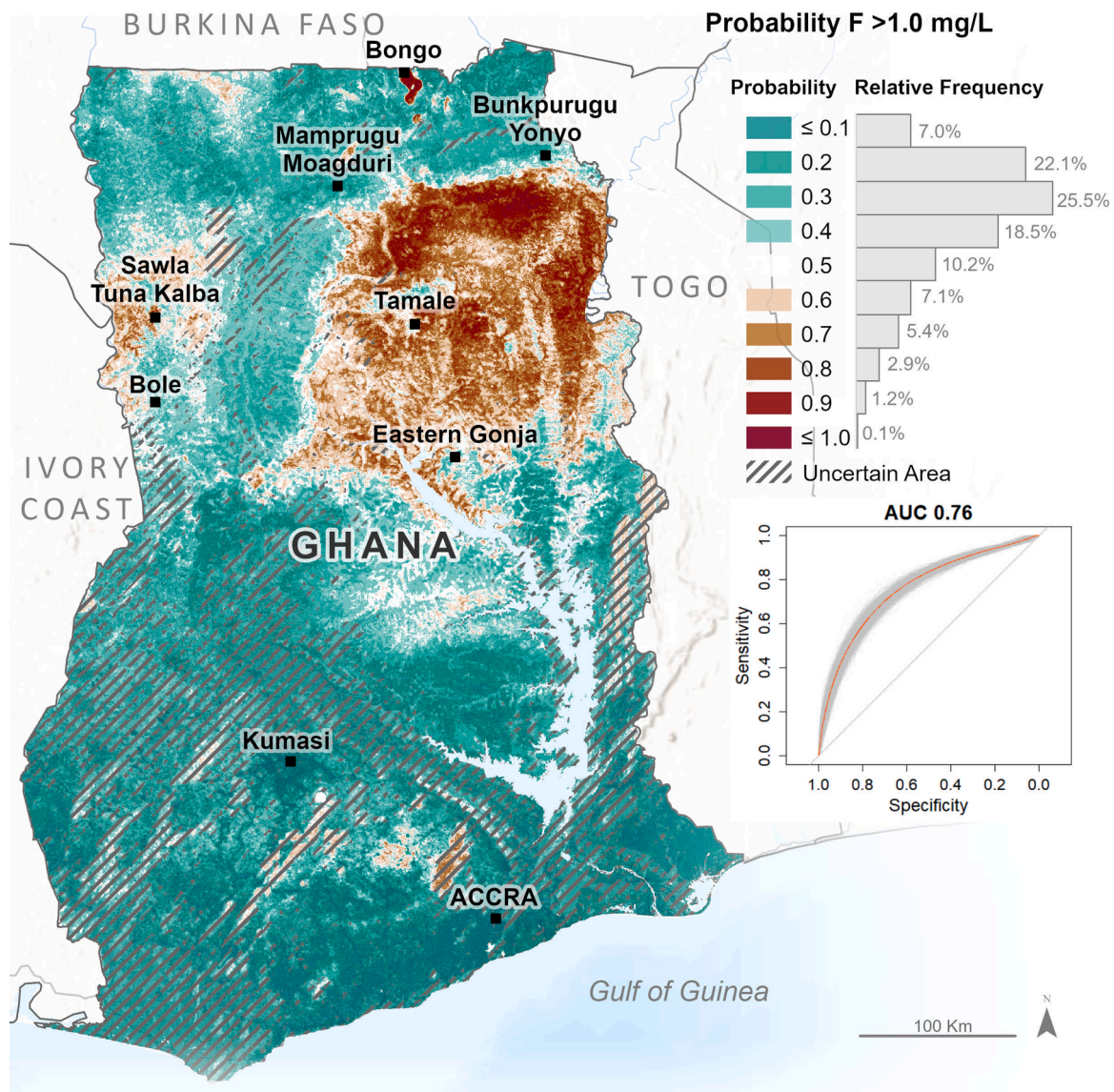
Fig. 4. Probability (hazard) map of fluoride exceeding 1.0 mg/L in groundwater of Ghana.. The hatched gray shading depicts areas outside the area of applicability (AOA), where the prediction is uncertain. The mean AUC is 0.76 (inset).

class would dominate the classification. This is especially relevant for the fluoride concentration because the over-represented class corresponds to cases where fluoride ≤1.0 mg/L is reported (i.e., class zero). Not correcting this bias would mean that sensitivity, a metric that informs about the correct prediction of positive cases, would be under-represented compared to specificity, which correspondingly affects the balanced accuracy (Evans and Cushman, 2009; Khalilia et al., 2011). Furthermore, by iterating the model 1000 times by randomly selecting the data from our balanced sample, we avoided any bias introduced by data splitting. Because of this calibration, metrics reporting the model's predictive performance with the new data indicate that the models generalize well. This highlights that the pixels assigned to areas with fluoride presence above 1.0 mg/L are well determined. This is also evident in the ability of the model to identify cases within the dataset around 70% of the time (sensitivity) and in the overall ability of the models to correctly identify both classes, which is also around 70% (at a cutoff of 0.49). Compared to other studies using surface predictor variables (e.g. Amini et al., 2008; Bindal and Singh, 2019; Cao et al., 2022), the metrics of this model are somewhat lower, but this is because we use spatial cross-validation, which provides a useful estimate of the model's predictive performance without an optimistic bias due to the SAC. For

example, Pohjankukka et al., 2017 has shown that the metrics produced by non-spatial cross validation can be up to 40% more optimistic than those of spatial-cross-validation. Similar results were reported by (Dolan et al., 2021; Ploton et al., 2020; Roberts et al., 2017).

4.3. Hazard map

We created a hazard map of geogenic groundwater fluoride contamination for the whole of Ghana (Fig. 4), where high concentrations of fluoride mainly affect the northeastern part of the country between the Northern region and the Upper East region. Fluoride contamination in this area is extensively documented in the literature (see supplementary Table S3). The Northern Region has been recognized for its high fluoride concentrations in relation to the rocks of the Oti-Pendjari Group, with reported fluoride concentrations of over 4.0 mg/L (Anim-Gyampo et al., 2012; CIDA, 2011). High fluoride concentrations are also observed in northwestern Ghana, mainly in districts of the Savannah region, with concentrations over 1.5 mg/L reported by some studies (Arhin and Affam, 2010; Loh et al., 2020).

In general, the model shows a very low probability (0–10%) of fluoride concentration above the threshold in the south of the country
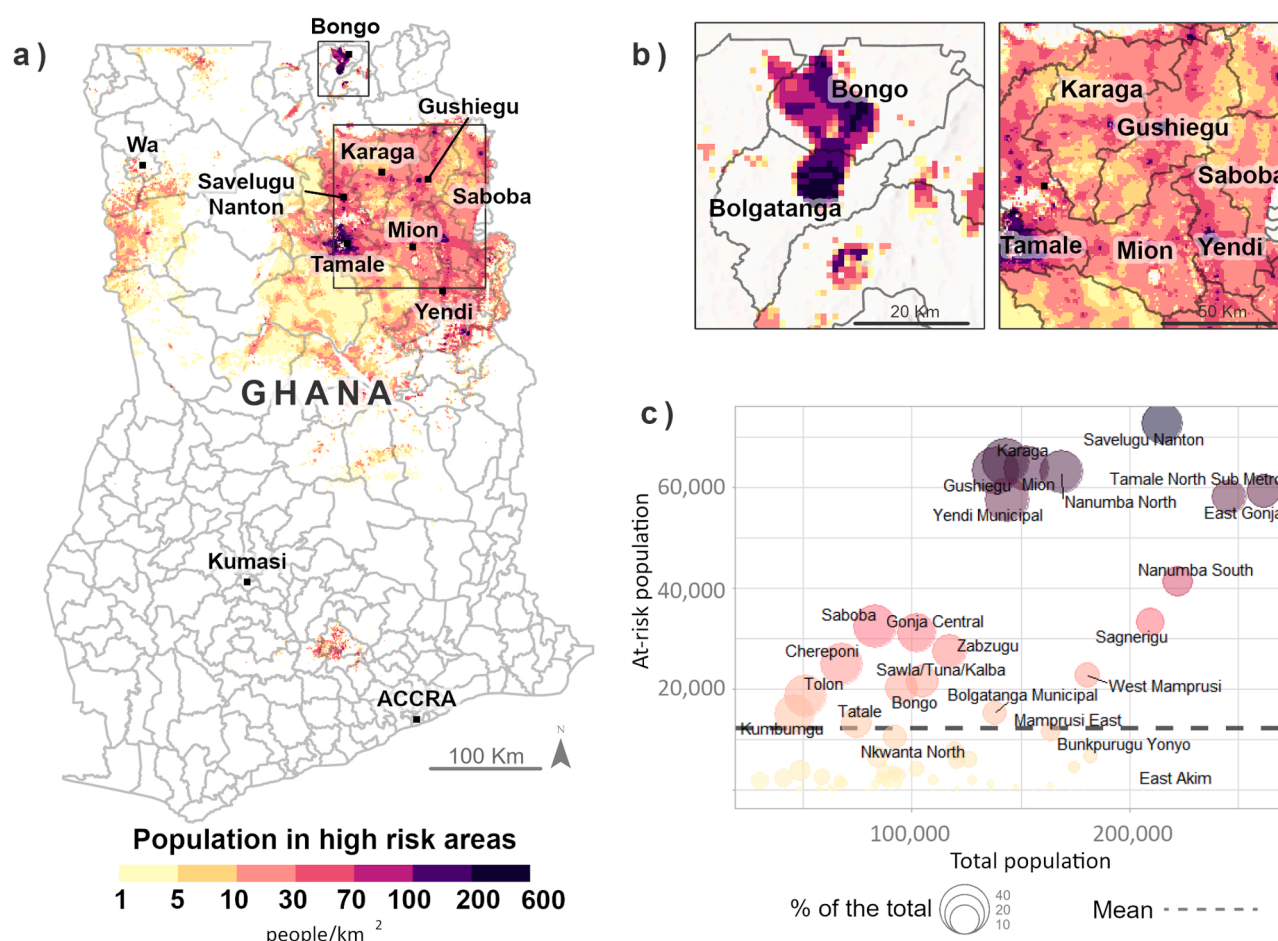
Fig. 5. Population of Ghana at risk of high fluoride ingestion through groundwater. (a) Population density of areas at risk of exposure to fluoride >1.0 mg/L at a probability cutoff 0.49. (b) Zoomed-in sections of the districts around Bongo, Karaga and Tamale. (c) Plot of the total population by district against the at-risk population, excluding the population in areas of high uncertainty (3% of at-risk population). The size of the bubbles shows the percentage of population at risk in each district out of the total population.

(see supplementary Table S3). Only in specific areas of the Eastern Region does the model have a 50–60% probability of fluoride concentrations above 1.0 mg/L, and these concentrations were mainly related to restricted areas within the Eburnearn supergroup. However, in the south of the country, the model shows less certainty (see area outside the AOA, Fig. S5d), especially in the Western, Western North, and Ahafo regions, where the reporting or measurement of fluoride levels has been sparse. This uncertainty is likely due to combinations of predictor values that are not found in the training data. Despite this, these results are most likely correct, as this area has environmental characteristics inversely related to higher fluoride concentration, including a more humid tropical climate and high rainfall that dilutes the chemical composition of the groundwater (Frencken, 1992). Furthermore, the authors who have reported high fluoride levels in the area have linked these findings to anthropogenic/agrochemical pollution or seawater intrusion (Yidana, 2010; Zango et al., 2019). Identifying areas of higher uncertainty where modeling is more complex is relevant for prioritizing resources in the design of future sampling campaigns. Here, future studies should look empirically into highlighted areas where more sampling is required to confirm results, refine the model, and create a more accurate representation of the geogenic hazard contamination. Future studies could also attempt modeling anthropogenic sources and salt-water intrusion.

We estimated the number of people potentially at risk from excess fluoride in drinking water by creating a risk map of fluoride contamination (>1.0 mg/L). We generated a reliable high fluoride hazard map by removing 0.7% of the pixels with standard deviation and 1.3% of the pixels with high uncertainty (supplementary Fig. S5). The exclusion of these pixels had only a limited effect on the fluoride hazard map. Given that the standard deviation was not significant and that the difference in pixels between the maps was located at the class boundaries, this behavior is expected, as modeling at class boundaries usually shows more discrepancies than within the class features (Foody, 2005; Steele et al., 1998). By contrast, the area of uncertainty (where the model was not trained with these environments) is highly concentrated in the south of the country, where almost no fluoride hazard was modeled.

4.4. Estimation of the at-risk population

Ghana faces a major challenge in providing drinking water to its most at-risk population, as the northeastern part of the country, which has the highest exposure to fluoride contamination, also contains a greater proportion of the national population of children aged 0–9 years (see Supplementary Fig. S6). Districts in the Northern Region have the largest exposed population (Fig. 6), though the exposure varies significantly. In districts such as Gushiegu, Karaga, and Mion, approximately 4 out of 10 children are potentially exposed to levels of fluoride that can affect their health. The Savelugu Nanton district has the largest exposed child population of about 19,000 children, followed by the Karaga district with about 17,000 children. In total, an estimated 920,000 people are potentially exposed in Ghana (Fig. 5). The northeastern part of the country has the highest proportion of the rural population and, therefore, the highest dependency on groundwater (WHO/UNICEF Joint
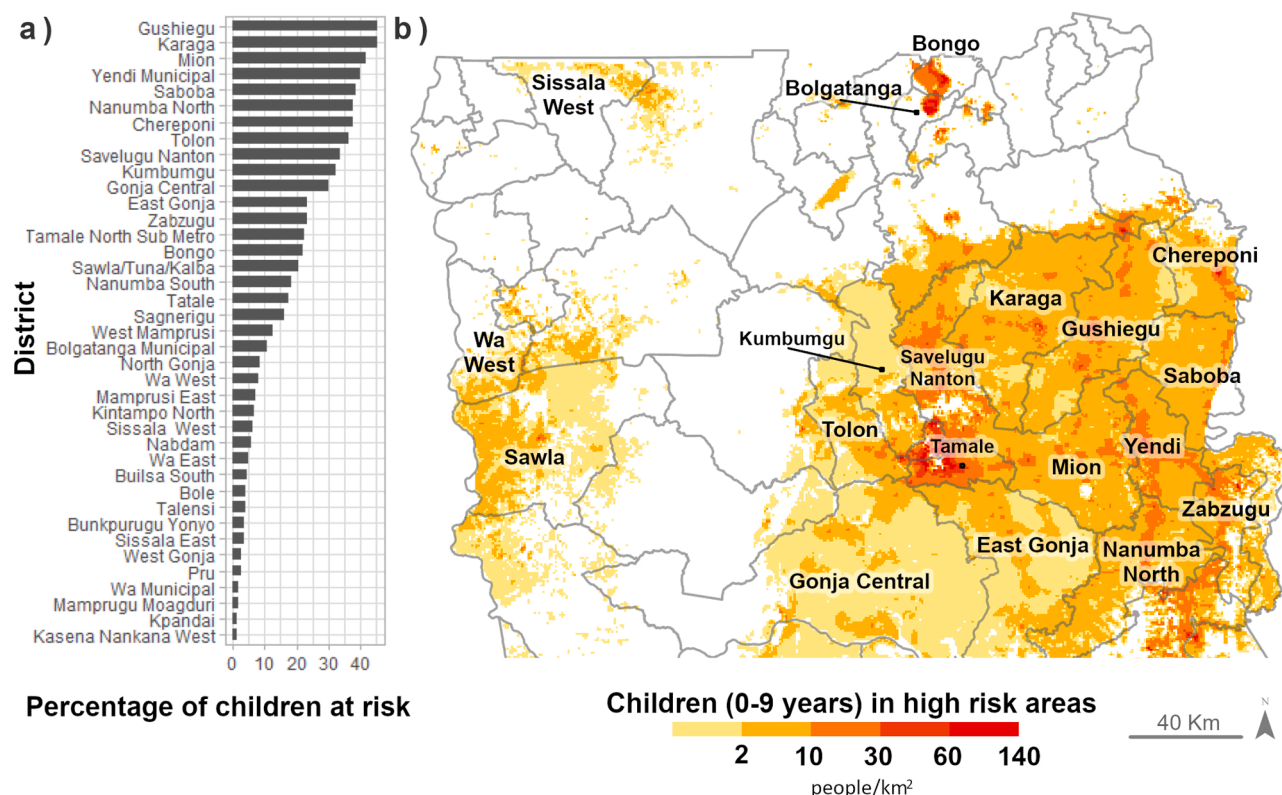
Fig. 6. Children in Ghana potentially exposed to high fluoride ingestion through the consumption of groundwater as drinking water. (a) Percentage of at-risk children per district. (b) Population density of children aged 0–9 years potentially exposed to high levels of fluoride.

Monitoring Program (JMP), 2019). Furthermore, this area, and particularly the Northern region, has a poverty level close to 50% of the population, and this level, unlike in other regions of Ghana, has not been reduced in recent decades (UNICEF, 2016). Children and adults must also often deal with psychological problems stemming from the social stigma of having teeth with fluorosis (Castilho, 2009; Dongzagla et al., 2019). Of note, Ghana conducted a census in 2021, such that the number of at-risk people can be updated when this new information becomes available. Regardless, new spatially disaggregated population databases would increasingly help to identify risk groups, rather than considering the population as a whole. This will provide a better understanding of the spatial patterns of the distribution and characteristics of these at-risk populations, thereby allowing the provision of information at more relevant administrative scales.

4.5. Applicability

The hazard and risk maps are at a scale that can serve the authorities as a basis for more detailed research on water quality in Ghana. Priority areas for further investigation and possible mitigation could be determined in part, for example, by the list of districts with a higher presence of young children potentially exposed to levels of fluoride that can harm their health (Fig. 6), or by the identification of fluoride hazard areas (Fig. 4). The results highlight not only hotspots of fluoride contamination and potentially exposed populations but also data gaps. Both the modeling approach and the choice of variables can serve as a basis for studying groundwater quality in other areas outside of Ghana. Furthermore, the spatial approach for model validation should be relevant to other machine learning applications that are conducted in a spatial context.

5. Conclusions

We present the first hazard map of fluoride in groundwater resources throughout Ghana, which was produced by geospatially modeling the fluoride concentration using a random forest machine-learning algorithm. The northeastern part of the country consistently exceeds a fluoride concentration of 1.0 mg/L in groundwater. Although the southern part of the country has a very low probability of reaching this concentration, a better and broader distribution of sampling data would be needed to confirm these results, refine the model, and create a more accurate model of contamination. We also identified that approximately 240,000 children and 680,000 adults are potentially exposed to levels of fluoride that may affect their health. An important point to be stressed is that the majority of the child population is located in the north of the country, where a higher probability exists for exposure to high fluoride levels in groundwater. This region of Ghana has a greater reliance on groundwater as well as the poorest population. Therefore, these concerns highlight the challenges the country faces in protecting its at-risk population. A remaining question is to what extent the expected increase in aridity in the northern region might affect fluoride concentrations. Overall, climate change is expected to make water scarcer and as a result increase fluoride concentrations in this area, which would force people to seek unsafe drinking water sources.

We have reduced the effect of spatial dependencies in the training and testing data by evaluating model performance through spatial cross-validation, thereby avoiding overly optimistic results. Using the area of applicability has allowed us to identify areas with greater confidence in the prediction. Partial dependence plots depict the relationships between variables and the probability of high fluoride contamination. These techniques assist the validation and interpretation of otherwise black-box models. Overall, the model created here is a valuable resource for estimating the presence and absence of fluoride contamination throughout Ghana, particularly in areas without widespread well

testing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.118083.

References

Abusa, Y., Anom, P.A., Doamekpor, M.E.A.M., Gyamfi, E.E., Doamekpor, L.K., 2018. Physico-chemical assessment of drinking water with special emphasis on fluoride concentration in the akatsi-north district in the volta region of Ghana. West Afr. J. Appl. Ecol. 26, 72–92. https://doi.org/10.4314/wajae.v26i2.

Addo, M.A., Darko, E.O., Gordon, C., Nyarko, B.J.B., 2011. Water quality analysis and human health risk assessment of groundwater from open-wells in the vicinity of a cement factory at Akporkloe, Southeastern Ghana. e-Journal Sci. Technol. 4, 16–30.

Affam, M., Arhin, E., Asamoah, D.N., 2012. Source of endemic fluorosis attributed to igneous granitiods in northern Ghana: a case study. In: 2nd UMaT Biennial International Mining and Mineral Conference, pp. 242–248.

Agyemang, V.O., 2020. Hydro chemical characterization and assessment of groundwater suitability for drinking, domestic and irrigation purposes in the agona east district, Ghana. IRE J. 3, 112–125.

Alfredo, K.A., Lawler, D.F., Katz, L.E., 2014. Fluoride contamination in the Bongo District of Ghana, West Africa: Geogenic contamination and cultural complexities. Water Int. 39, 486–503. https://doi.org/10.1080/02508060.2014.926234.

Ambroise, C., McLachlan, G.J., 2002. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc. Natl. Acad. Sci. USA 99, 6562–6566. https://doi.org/10.1073/pnas.102102699.

Amini, M., Mueller, K., Abbaspour, K.C., Rosenberg, T., Afyuni, M., Møller, K.N., Sarr, M., Johnson, C.A., 2008. Statistical modeling of global geogenic fluoride contamination in groundwaters. Environ. Sci. Technol. 42, 3662–3668. https://doi.org/10.1021/es071958y.

Anim-Gyampo, M., Zango, M.S., Apori, N., 2012. The origin of fluoride in groundwaters of the paleozoic sedimentary formations of Ghana-a preliminary study in Gushiegu District. Res. J. Environ. Earth Sci. 4, 546–552.

Anku, Y.S., Banoeng-Yakubo, B., Asiedu, D.K., Yidana, S.M., 2009. Water quality analysis of groundwater in crystalline basement rocks, Northern Ghana. Environ. Geol. 58, 989–997. https://doi.org/10.1007/s00254-008-1578-4.

Anornu, G., Gibrilla, A., Adomako, D., 2017. Tracking nitrate sources in groundwater and associated health risk for rural communities in the White Volta River basin of Ghana using isotopic approach (δ15N, δ18O[sbnd]NO3 and 3H). Sci. Total Environ. 603–604, 687–698. https://doi.org/10.1016/j.scitotenv.2017.01.219.

Apambire, W.B., 1996. Groundwater Geochemistry and the Genesis and Distribution of Groundwater Fluoride in the Bolgatanga and Bongo Districts. Carleton University, Ghana.

Apambire, W.B., Boyle, D.R., Michel, F.A., 1997. Geochemistry, genesis, and health implications of fluoriferous groundwaters in the upper regions of Ghana. Environ. Geol. 33, 13–24. https://doi.org/10.1007/s002540050221.

Araujo, M., Pearson, R., Thuiller, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Glob. Chang. Biol. 11, 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.001000.x.

Arhin, E., Affam, M., 2010. Fluoride in groundwater and its implications in west Gonja District of Ghana. Ghana Min. J. 11, 47–52. https://doi.org/10.4314/gm.v11i1.53272.

Arko, W.E., Hodgson, I.O.A., Nyame, F.K., 2019. Assessment of drinking water quality at Dodowa in the Dangbe West district of the Greater-Accra region, Ghana. Afr. J. Environ. Sci. Technol. 13, 181–190. https://doi.org/10.5897/ajest2019.2653.

Asiamah, R.D, Senayali, J.K, Agjei-Gyapong, T, Spaargaren, O.C, 1997. Ethno-pedology surveys in the semi-arid Savannah zone of Northern Ghana. Soil Research Institute, Kwadaso- Kumasi and International Soil Reference and Information Centre (ISRIC) 1–46. https://www.isric.org/sites/default/files/isric_report_1997_04.pdf.

Atipoka, F.A., 2009. Water supply challenges in rural Ghana. Desalination 248, 212–217. https://doi.org/10.1016/j.desal.2008.05.057.

Avi, C., Tettey, M., Ayitey, K., 2019. Evaluation of groundwater quality for drinking and irrigation in central Tongu District of the Volta Region of Ghana. J. Environ. Earth Sci. 9, 159–166. https://doi.org/10.7176/jees/9-10-16.

Barbier, O., Arreola-Mendoza, L., Del Razo, L.M., 2010. Molecular mechanisms of fluoride toxicity. Chem. Biol. Interact. 188, 319–333. https://doi.org/10.1016/j.cbi.2010.07.011.

Bindal, S., Singh, C.K., 2019. Predicting groundwater arsenic contamination : regions at risk in highest populated state of India. Water Res. 159, 65–76. https://doi.org/10.1016/j.watres.2019.04.054.

Boakye Opoku, R., 2013. Assessing the Water Supply Potential of Boreholes on Kwame Nkrumah University of Science and Technology Campus. Kwame Nkrumah University of Science and Technology.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1201/9780429469275-8.

Bureau of Indian Standards (BIS), 2012. Indian Standard Drinking Water-Specification. Bureau of Indian Standards (BIS), New Delhi, India (second revision).

Cao, H., Xie, X., Wang, Y., Liu, H., 2022. Predicting geogenic groundwater fluoride contamination throughout China. J. Environ. Sci. 115, 140–148. https://doi.org/10.1016/j.jes.2021.07.005.

De Castilho, L.S., 2009. Perceptions of adolescents and young people regarding endemic dental fluorosis in a rural area of Brazil : psychosocial suffering. Health Soc. Care Community 17, 557–563. https://doi.org/10.1111/j.1365-2524.2009.00859.x.

Chegbeleh, L.P., Akurugu, B.A., Yidana, S.M., 2020. Assessment of groundwater quality in the Talensi District, Northern Ghana. Sci. World J. 2020, 1–24. https://doi.org/10.1155/2020/8450860.

CIDA, 2011. Hydrogeological Assessment Project of the Northern Regions of Ghana (HAP). CIDA, Quebec, Canada.

Craig, L., Lutz, A., Berry, K.A., Yang, W., 2015. Recommendations for fluoride limits in drinking water based on estimated daily fluoride intake in the Upper East Region, Ghana. Sci. Total Environ. 532, 127–137. https://doi.org/10.1016/j.scitotenv.2015.05.126.

Craig, L., Thomas, J.M., Lutz, A., Decker, D.L., 2018. Determining the optimum locations for pumping low-fluoride groundwater to distribute to communities in a fluoridic area in the Upper East Region, Ghana. Chem. Geol. 476, 481–492. https://doi.org/10.1016/j.chemgeo.2017.12.001.

Dolan, M.F.J., Ross, R.E., Albretsen, J., Skarðhamar, J., Gonzalez-Mirelis, G., Bellec, V.K., Buhl-Mortensen, P., Bjarnadóttir, L.R., 2021. Using spatial validity and uncertainty metrics to determine the relative suitability of alternative suites of oceanographic data for seabed biotope prediction. A case study from the barents sea, Norway. Geoscience 11, 1–38. https://doi.org/10.3390/geosciences11020048.

Dongzagla, A., Jewitt, S., O'Hara, S., 2019. Assessment of fluoride concentrations in drinking water sources in the Jirapa and Kassena-Nankana Municipalities of Ghana. Groundw. Sustain. Dev. 9 https://doi.org/10.1016/j.gsd.2019.100272, 100272.

Edmunds, M., Smedley, P., 2013. Fluoride in natural waters, Essentials of Medical Geology. Elsevier Academic, London, UK. https://doi.org/10.1007/978-94-007-4375-5. Revised Edition.

Egbi, C.D., Anornu, G., Appiah-Adjei, E.K., Ganyaglo, S.Y., Dampare, S.B., 2019. Evaluation of water quality using hydrochemistry, stable isotopes, and water quality indices in the Lower Volta River Basin of Ghana. Environ. Dev. Sustain. 21, 3033–3063. https://doi.org/10.1007/s10668-018-0180-5.

Environmental Protection Agency (EPA), 2020. Ghana's fourth national communication to the United Nations framework convention on climate change. Environmental Protection Agency (EPA), Accra, Ghana project.

Evans, J.S., Cushman, S.A., 2009. Gradient modeling of conifer species using random forests. Landsc. Ecol. 24, 673–683. https://doi.org/10.1007/s10980-009-9341-0.

Firempong, C., Nsiah, K., Awunyo-Vitor, D., Dongsogo, J., 2013. Soluble fluoride levels in drinking water-a major risk factor of dental fluorosis among children in Bongo community of Ghana. Ghana Med. J. 47, 16–23.

Foody, G.M., 2005. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. Int. J. Remote Sens. 26, 1217–1228. https://doi.org/10.1080/01431160512331326521.

Frencken, J.E., 1992. Endemic Fluorosis in Developing Countries. NIPG-TNO, The Netherlands. AC Leiden.

Ganyaglo, S.Y., Gibrilla, A., Teye, E.M., Owusu-Ansah, E.D.G.J., Tettey, S., Diabene, P.Y., Asimah, S., 2019. Groundwater fluoride contamination and probabilistic health risk assessment in fluoride endemic areas of the Upper East Region, Ghana. Chemosphere 233, 862–872. https://doi.org/10.1016/j.chemosphere.2019.05.276.

Gastineau, K.M., 2015. Hydrogeologic controls on groundwater quality in the West Mamprusi area, Northern Ghana. Int. J. Soil Sci. https://doi.org/10.3923/ijss.2017.32.38.

Greenwell, B.M., 2017. pdp: An R package for constructing partial dependence plots. R J. 9, 421–436. https://doi.org/10.32614/rj-2017-016.

GSS, 2020. Ghana annual five-year age group population estimates by sex for 2000 to 2020: national, and first- and second-order administrative divisions. Ghana Statistical Services. https://www2.census.gov/programs-surveys/international-programs/tables/time-series/pepfar/ghana.xlsx.

Hammond, T.O., Verbyla, D.L., 1996. Optimistic bias in classification accuracy assessment. Int. J. Remote Sens. 17, 1261–1266. https://doi.org/10.1080/01431169608949085.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution : random forests significantly improve current predictions. PLoS One 1–26. https://doi.org/10.1371/journal.pone.0125814.

Jones, A., Al, E., 2013. Soil Atlas of Africa, European Commission. Publications Office of the European Union, Luxembourg. https://doi.org/10.2788/52319.

Khalilia, M., Chakraborty, S., Popescu, M., 2011. Predicting disease risks from highly imbalanced data using random forest. BMC Med. Inform. Decis. Mak. 11, 51. https://doi.org/10.1186/1472-6947-11-51.

Kulinkina, A.V., Plummer, J.D., Chui, K.K.H., Kosinski, K.C., Adomako-Adjei, T., Egorov, A.I., Naumova, E.N., 2017. Physicochemical parameters affecting the perception of borehole water quality in Ghana. Int. J. Hyg. Environ. Health 220, 990–997. https://doi.org/10.1016/j.ijheh.2017.05.008.

Kumar, S., Singh, R., Venkatesh, A.S., Udayabhanu, G., Sahoo, P.R., 2019. Medical Geological assessment of fluoride contaminated groundwater in parts of Indo-Gangetic Alluvial plains. Sci. Rep. 9, 1–16. https://doi.org/10.1038/s41598-019-52812-3.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3–10. https://doi.org/10.1016/j.gsf.2015.07.003.

Loh, Y.S., Yidana, S.M., Asiedu, D., Akabzaa, T., Jørgensen, N.O., 2012. Hydrochemical Characterisation of Groundwater in Parts of the Volta Basin, Northern Ghana. Ghana Min. J. 13, 24–32. –32.

Loh, Y.S.A., Akurugu, B.A., Manu, E., Aliou, A.S., 2020. Assessment of groundwater quality and the main controls on its hydrochemistry in some Voltaian and basement aquifers, northern Ghana. Groundw. Sustain. Dev. 10 https://doi.org/10.1016/j.gsd.2019.100296, 100296.

Malago, J., 2017. Fluoride levels in surface and groundwater in Africa: a review. Am. J. Water Sci. Eng. 3, 1. https://doi.org/10.11648/j.ajwse.20170301.11.

Mensah-Essilfie, I., 2013. Groundwater Quality Within Kwesiprah, Cape Coast, Central Region, Ghana. Kwame Nkrumah University of Science and Technology.

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633. https://doi.org/10.1111/2041-210X.13650.

Meyer, V., Kuhlicke, C., Luther, J., Fuchs, S., Priest, S., Dorner, W., Serrhini, K., Pardoe, J., McCarthy, S., Seidel, J., Palka, G., Unnerstall, H., Viavattene, C., Scheuer, S., 2012. Recommendations for the user-specific enhancement of flood maps. Nat. Hazards Earth Syst. Sci. 12, 1701–1716. https://doi.org/10.5194/nhess-12-1701-2012.

Misiuk, B., Diesing, M., Aitken, A., Brown, C.J., Edinger, E.N., Bell, T., 2019. A spatially explicit comparison of quantitative and categorical modelling approaches for mapping seabed sediments using random forest. Geoscience 9, 2017–2020. https://doi.org/10.3390/geosciences9060254.

Nkansah, M.A., Lartey, J., Agorsor, P.I., 2019. The quality of water sources from high schools in the Accra metropolis. SF J. Environ. Earth Sci. 2, 1–7. https://scienceforecastoa.com/Articles/SJEES-V2-E2-1032.pdf, 2.

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., Maffenini, L., 2019. GHS-SMOD R2019A - GHS settlement layers, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015). European Commission, Joint Research Centre (JRC) [Dataset] doi: https://doi.org/10.2905/42e8be89-54ff-464e-be7b-bf9e64da5218.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 1–11. https://doi.org/10.1038/s41467-020-18321-y, 4540.

Podgorski, J., Berg, M., 2020. Global threat of arsenic in groundwater. Science 368, 845–850. https://doi.org/10.1126/science.aba1510 (80-.).

Pohjankukka, J., Pahikkala, T., Nevalainen, P., Heikkonen, J., 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. Int. J. Geogr. Inf. Sci. 31, 2001–2019. https://doi.org/10.1080/13658816.2017.1346255.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881 (Cop.).

Smedley, P.L., 1996. Arsenic in rural groundwater in Ghana. J. Afr. Earth Sci. 22, 459–470. https://doi.org/10.1016/0899-5362(96)00023-1.

Smedley, P.L., Edmond, J.M., West, J.M., Gardner, S.J., Pelig-Ba, K.B., 1995. Vulnerability of Shallow Groundwater Quality Due to Natural Geochemical Environment: Health Problems Related to Groundwater in the Obuasi Ans Bolgatanga Areas, Ghana. British Geological Survey.

Smedley, P.L., Nkotagu, H., Pelig-Ba, K., MacDonald, A.M., Whittle, R.T., Whitehead, E.J., Kinniburgh, D.G., 2002. Fluoride in groundwater from high-fluoride areas of Ghana and Tanzania. Br. Geol. Surv. Nat. Environ. Res. Counc. 74, 1–74.

Steele, B.M., Chris Winne, J., Redmond, R.L., 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps. Remote Sens. Environ. 66, 192–202. https://doi.org/10.1016/S0034-4257(98)00061-3.

Sunkari, E.D., Abu, M., Zango, M.S., Lomoro Wani, A.M., 2020. Hydrogeochemical characterization and assessment of groundwater quality in the Kwahu-Bombouaka Group of the Voltaian Supergroup, Ghana. J. Afr. Earth Sci. 169 https://doi.org/10.1016/j.jafrearsci.2020.103899, 103899.

Trabucco, A., Zomer, R.. https://doi.org/10.6084/m9.figshare.7504448.v3.

UNICEF, 2016. The Ghana Poverty and Inequality Report, 2016. UNICEF.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol. Evol. 10, 225–232. https://doi.org/10.1111/2041-210X.13107.

WHO/UNICEF Joint Monitoring Program, (JMP), 2020. Joint Monitoring Programme for Water Supply, Sanitation and Hygiene, 2020. United Nations Children's Fund and World Health Organization UNICEF and WHO, New York.

WHO/UNICEF Joint Monitoring Program (JMP), 2019. Global data on Water Supply, Sanitation and Hygiene (WASH). WHO/UNICEF, New York.

WHO, 2017. Guidelines for Drinking-Water Quality, 4th ed. WHO, Geneva, Switzerland. incorporating the 1st addendum.

WHO, 2004. Guidelines for Drinking-water Quality, 3rd ed. WHO, Geneva, Switzerland.

WHO, 1996. Trace Elements in Human Nutrition and Health. WHO, Geneva, Switzerland.

WorldPop, 2018. Global High Resolution Population Denominators. WorldPop.

Yidana, S.M., 2010. Groundwater classification using multivariate statistical methods: Southern Ghana. J. Afr. Earth Sci. 57, 455–469. https://doi.org/10.1016/j.jafrearsci.2009.12.002.

Yidana, S.M., Banoeng-Yakubo, B., Aliou, A.S., Akabzaa, T.M., 2012. Groundwater quality in some Voltaian and Birimian aquifers in northern Ghana-application of mulitvariate statistical methods and geographic information systems. Hydrol. Sci. J. 57, 1168–1183. https://doi.org/10.1080/02626667.2012.693612.

Zango, M.S., Sunkari, E.D., Abu, M., Lermi, A., 2019. Hydrogeochemical controls and human health risk assessment of groundwater fluoride and boron in the semi-arid North East region of Ghana. J. Geochem. Explor. 207 https://doi.org/10.1016/j.gexplo.2019.106363, 106363.