Science of the Total Environment 735 (2020) 139584

Contents lists available at ScienceDirect

Science of the Total Environment

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

Spatial analysis and GIS mapping of regional hotspots and potential health risk of fluoride concentrations in groundwater of northern Tanzania

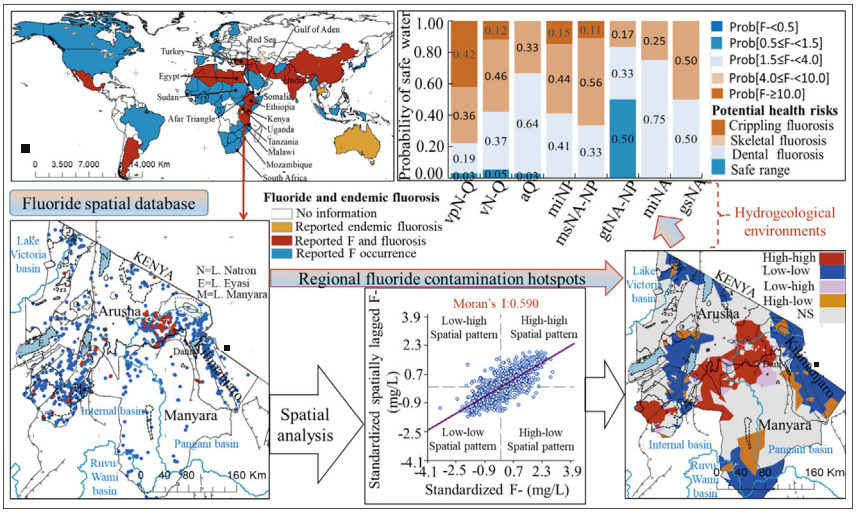
Julian Ijumulana [a,b,*], Fanuel Ligate [a,b,c], Prosun Bhattacharya [a,d], Felix Mtalo [b], Chaosheng Zhang [e]

[a] KTH-International Groundwater Arsenic Research Group, Department of Sustainable Development, Environmental Science and Engineering, KTH Royal Institute of Technology, Teknikringen 10B, SE-100 44 Stockholm, Sweden
[b] Department of Water Resources Engineering, College of Engineering and Technology, University of Dar es Salaam, Dar es Salaam, Tanzania
[c] Department of Chemistry, Mkwawa College of Education, University of Dar es Salaam, Tanzania
[d] KWR Watercycle Research Institute, Groningenhaven 7, 3433 PE Nieuwegein, The Netherlands
[e] International Network for Environment and Health (INEH), School of Geography and Archaeology & Ryan Institute, National University of Ireland, Galway, Ireland

HIGHLIGHTS

- Fluoride occurence in groundwater systems of the study area is space dependent.
- The univariate local Moran's I statistic has been used to identify significant regional hotspots and cool spots.
- Geoinformation on hotspots and cool spots is important for planning of new safe drinking water source development.
- Fluoride is mobilized from volcanic igneous rocks of intermediate chemical composition between mafic and felsic rocks.
- Potential processes include dissolution of fluoride bearing minerals mainly titanite, amphibole, hornblende and biotite.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 March 2020
Received in revised form 18 May 2020

ABSTRACT

Safe drinking water supply systems in naturally contaminated hydrogeological environments require precise geoinformation on contamination hotspots. Spatial statistical methods and GIS were used to study fluoride occurrence in groundwater and identify significant spatial patterns using fluoride concentrations. The global and local

Abbreviations: AMGC, African Minerals and Geosciences Centre; AUWASA, Arusha Urban Water Supply Authority; BH, borehole; DAFWAT, development of affordable adsorbent systems for arsenic and fluoride removal in the drinking water sources in Tanzania; DEM, digital elevation model; DQW, Division of Water Quality; DW, dug well; EARV, East African Rift Valley; ESDA, exploratory spatial data analysis; GFB, Global Fluoride Belt; GIS, Geographical Information Systems; GNSS, Global Navigation Satellite Systems; GST, Geological Survey of Tanzania; IQ, intelligent quotient; IQR, interquartile range; MoWI, Ministry of Water and Irrigation; NBS, National Bureau of Statistics; NDRS, Ngurdoto Defluoridation Research Station; NFB, National Fluoride Belt; SID, sample identification number; Sida, Swedish International Development Agency; SRTM, Shuttle Radar Topography Mission; SW, shallow well; WGS84, World Geodetic System of 1984; WHO, World Health Organization.

* Corresponding author.
E-mail address: julianij@kth.se (J. Ijumulana).

https://doi.org/10.1016/j.scitotenv.2020.139584

Accepted 19 May 2020
Available online 21 May 2020

Keywords:
Groundwater
Drinking water
Regional fluoride hotspots
Regional fluoride cool spots
Spatial analysis
GIS

Morans I indices were used. While the significant positive global Morans I index indicated spatial structure in fluoride occurrence, the significant spatial clusters were identified using local Morans I index and mapped at p-value of 0.05. The spatial clusters demonstrated patterns of drinking water sources with fluoride concentrations below or above WHO guideline and Tanzania standard for drinking water and were considered as 'regional fluoride cool spots' and 'regional fluoride contamination hotspots', respectively. Two regional fluoride contamination hotspots were identified and mapped around the Stratovolcano Mountains in the north-east and south-west of the study area; and along the Neogene Quaternary volcanic formations and Palaeo-Neoproterozoic East African Orogen (Mozambique Belt). The two largest regional fluoride cool spots dominated the major and minor rift escarpments in the west and east of the study area respectively while the small ones emerged around the volcanic mountains in the north and south. Furthermore, significant spatial outliers emerged at the boundary of regional fluoride hotspots and cool spots as an indication of the spatial processes controlling the mobilization of fluoride in groundwater. While all water sources in the cool spots had fluoride concentrations below 1.5 mg/L, some had extremely low concentrations below 0.5 mg/L which is not safe for human consumption. For hotspots, 96% of water sources had fluoride concentrations above 1.5 mg/L. The probability of having safe source of drinking water varied from one geological unit to another with sources in the Neogene Quaternary volcanic formations having least probabilities.

1. Introduction

Geogenic contamination in drinking water sources and associated health effects remain a global challenge to safe water supplying authorities (Bhattacharya and Bundschuh, 2015; Bundschuh et al., 2017). Despite groundwater remains the only source of fresh water, presence of low or high concentration of certain ions, mainly arsenic, fluoride, nitrate, iron, manganese, boron, most of heavy metals and radionuclides make them unsuitable for drinking (Ali et al., 2018; 2019; Brindha and Elango, 2011; Bundschuh et al., 2017; Chandrajith et al., 2020; Jacks et al., 2005; Jeihanipour et al., 2018, Kashyap et al., 2020; McMahon et al., 2020; Vithanage and Bhattacharya, 2015). Presence of high fluoride levels beyond the recommended World Health Organization (WHO) drinking water guideline and Tanzania standard value of 1.5 mg/L remains a setback to drinking water-quality in different countries in Africa (Alfredo et al., 2014; Brunt et al., 2004; Frencken, 1992). Tanzania is one of the countries with highest fluoride levels in groundwater ranging between 15 and 63 mg/L (Brindha and Elango, 2011; Nanyaro et al., 1984; Vuhahula et al., 2009). About one third of the country has water sources affected by high fluoride concentrations particularly in northern zone of its jurisdiction (Fawell and Bailey, 2006; Lathman and Grech, 1967; Smedley et al., 2002; Thole, 2005, 2013). Occurrence of low or high concentration of fluoride in groundwater is because of geogenic or anthropogenic causes or a combination of both.

The natural occurrence of fluoride in groundwater has been reported in many countries of the world with low income countries being prominent (Kimambo et al., 2019). For instance, 50% of the countries with population consuming water with fluoride concentration > 1.5 mg/L are from African continent (Kut et al., 2016). The natural fluoride sources in groundwater are linked to local geological setting of a country which varies from region to region depending on the heterogeneous rock formations. In Africa, over 435 million people depend on groundwater hosted in Precambrian basement rocks (~220 million), consolidated sedimentary rocks (~110 million), unconsolidated sedimentary rocks (~60 million) and young volcanic rocks (45 million) (MacDonald and Davies, 2000). The quality and quantity of groundwater in these hydrogeological environments is dependent on the physical and chemical properties of groundwater and host rocks in addition to residence time between the two. The most reported geological formations with host rocks bearing fluoride minerals include sellaite, fluorite(CaF_2), cryolite (Na_3AlF_6), fluorapatite ($Ca_5(PO_4)_3F$), villiaumite (NaF), topaz ($Al_2(SiO_4)F_2$), apatite ($Ca_5(PO_4)_3F$), fluormica, biotite, amphibole and hornblende [$(Ca,Na)_2(Mg,F,Al)_5(Si,Al)_8O_{22}(OH)_2$] (Brindha and Elango, 2011; Vithanage and Bhattacharya, 2015).

Fluoride contents of African groundwater vary from region to region depending on the geology, rock types, contact time of the water with aquifer rocks, regional climate, overall groundwater chemical composition, aquifer depth, and weathering intensity (Brunt et al., 2004; Chowdhury et al., 2019; Kimambo et al., 2019; Vithanage and Bhattacharya, 2015). The highest groundwater fluoride levels generally occur in alkaline volcanic areas of the East African Rift Valley. The presence of high fluoride concentrations is accelerated by young volcanic activities, occurrence of thermal waters especially those with high pH, gases emitted from earth's crust, granitic and gneissic rocks (Brunt et al., 2004; Malago et al., 2017; Thole, 2013; Ali et al., 2018, 2019).

Despite of health benefits of fluoride in human body, ingestion of excessive fluoride-rich water causes health effects collectively known as fluorosis (Chandrajith et al., 2012, 2020) whose severity depends on the level and period of exposure. The most vulnerable age groups include children and young people who are affected not only in terms of fluorosis but also decrease in intelligent quotient (IQ) particularly in primary school children (Hong et al., 2001; Lu et al., 2000; Mohanta and Mohanty, 2018; Shivaprakash et al., 2011). Several other health effects related to ingestion of fluoride contaminated waters have been reported among African population living in fluorotic regions (Fawell and Nieuwenhuijsen, 2003; Shrivastava and Vani, 2009; Thole, 2013).

Not only fluoride contamination in groundwater affects human health but also has socio-economic development implications. For instance, several boreholes have been abandoned in northern Tanzania within Rift Valley regions because of high levels of fluoride ions in water through the process of compliance with national standards and international guidelines (Gumbo and Mkongo, 1995). Likewise, the new development plans for water supply are slow in implementation because of uncertainty in fluoride contaminated hydrogeological environments (Kimambo et al., 2019).

Safe and adequate water supply in naturally contaminated hydrogeological environments requires precise and accurate information on the extent of contamination. Global models showing probability of contaminated groundwater resources in terms of fluoride concentrations exist (Amini et al., 2008; Brunt et al., 2004). These models provide global probability of contaminated groundwater resources, which may vary from one region to another because of different environmental conditions. Due to a large scale of variation in fluoride concentrations controlled by local geology, topography, climate and physio-chemical properties of water, regional models are required. Local spatial statistics and GIS tools when integrated together can provide means of handling large scale of variation in geogenic contaminants. In addition to characterization of spatial and temporal extent of contamination, local spatial statistical models indicate significant patterns technically known as contamination hotspots and cool spots in soil sciences (Zhang et al., 2008; Zhang and McGrath, 2004; Quino Lima et al., 2020).

Several methods for hotspot or spatial cluster identification are well documented including Gi and Gi* statistics (Getis and Ord, 1992), index for temporal clustering (Tango, 1995), spatial scan statistic (Kulldorff, 1997; Ishioka et al., 2007) and the local Morans I statistic (Anselin, 1995). Due to its simplicity and ability to test degree of association between individual observation and its neighbours, the local Moran's I statistical analysis has been used in several location-based studies (Nas, 2009; Zhang et al., 2008). However, the results of this technique are highly dependent on the neighbourhood function definition, presence of extreme values and non-normality of the data values in attribute space (Zhang et al., 2008). The problems can be overcome by adopting plausible tools such as exploratory spatial data analysis (ESDA) (Haining et al., 1998) and Geographical Information Systems (GIS) mapping techniques. ESDA tools provide means of understanding properties of the data to identify extreme values (spatial outliers) and overall distribution of the dataset in attribute and geographical space (Anselin, 1993; Symanzik, 2014). The GIS mapping capability provides means of integrating detected hotspots with other spatial datasets and hence facilitates interpretation of results.

The objective of this study was to conduct a systematic study on spatial variability of occurrence of fluoride in the drinking water sources in north-eastern Tanzania and map the significant spatial patterns (High-high, Low-low, Low-high and High-low) at 5% significance level using the fluoride levels in drinking water sources including boreholes, shallow wells and natural springs. Furthermore, the study aimed at mapping probability of targeting safe source following the significant spatial patterns under different natural geological conditions. The results of this research are vital to the district water supply authorities in the study area when planning for new safe drinking water source development.

2. Material and methods

2.1. Description of the study area

The study area is part of the earmarked national fluorotic regions in northern zone of Tanzania within the East African Rift Valley (EARV) that originates at Afar in Eritrea and extends through Ethiopia, Kenya, and Tanzania to Malawi where it divides into two branches (Fig. 1A and B). The area lies within the first Global Fluoride Belt (GFB) that originates from Republic of Turkey in South-East Europe and extends through Egypt, Sudan, Somalia, Ethiopia, Kenya, Tanzania, Mozambique and ends in South Africa (Brunt et al., 2004; Chowdhury et al., 2019; Kut et al., 2016) (Fig. 1A). According to Amini et al. (2008), the probability of fluoride concentrations in drinking water sources greater than WHO guideline and Tanzania standard value of 1.5 mg/L ranges between 0.6 and 0.8.

Our study area comprises part of two drainage basins, Internal and Pangani drainage basin, in northern zone of Tanzania extending from approximately 2° to 6° South of the Equator and from approximately 35° to 38° East of Greenwich (Fig. 1B). Specifically, the study was conducted within three administrative regions, namely, Kilimanjaro, Arusha and Manyara region (Fig. 1C). On average terms,

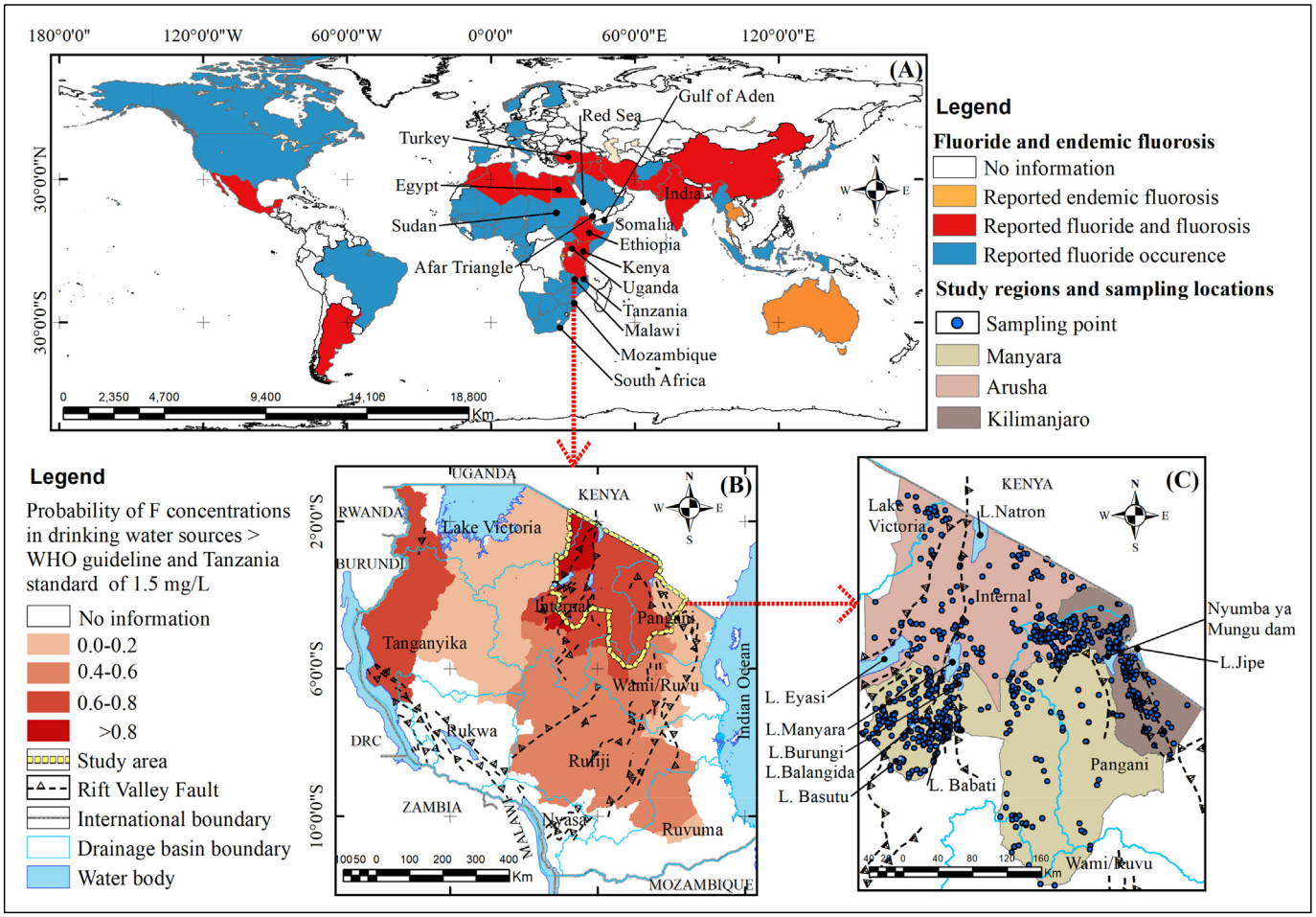


Fig. 1. (A) Map showing the global occurrences of elevated fluoride in groundwater (Amini et al., 2008) with reported endemic fluorosis (Mohanta and Mohanty, 2018) as well as the geographical location of Tanzania in the Global Fluoride Belt (GFB, Chowdhury et al., 2019), (B) Probability of fluoride concentration levels in groundwater-based drinking water sources in Tanzania (based on Amini et al., 2008) and (C) The study regions in northern Tanzania with sites of groundwater sampling locations.

the area is hot and semi-arid with an average annual rainfall of between 600 and 800 mm, falling mainly in October–December and March–May (Kijazi and Reason, 2009). Likewise, average monthly temperatures lie in the range of 20–26 °C in these regions. The area is characterized by volcanic activities mainly in the Rift Valley regions where Mount Kilimanjaro, the highest dormant stratovolcano in Africa, and other volcanic mountains such as Meru, Oldonyo Lengai and many others are located (Baudouin et al., 2016). According to National Bureau of Statistics (NBS) report in 2017 (NBS, 2018), approximately 10% of population resides in the study regions around the gentle slopes of the volcanic mountains. In these regions over 80% of the population depends on groundwater as a primary source of water for drinking and other socio-economic development activities both in urban and rural areas (Chacha et al., 2018; Kut et al., 2016).

2.2. General topography, geological and hydrogeological settings of the study area

Generally, large part of Tanzania is underlain by Precambrian basement rocks, mainly of granite, gabbro and meta-sediments with sedimentary basins of Karroo age to the south of the country (Smedley et al., 2002). Based on geological data obtained from the Geological Survey of Tanzania (GST) and through GIS analysis and mapping tools, our study area was geologically divided into four main lithostratigraphic units, namely, Palaeo-Neoproterozoic East African Orogen (Mozambique Belt), Neoarchaean granitic complex (Kavirondian-Nyanzian Supergroup), Neogene-Quaternary volcanic formations and Quaternary deposits, predominantly alluvial and eluvial sediments (Fig. 2). Table 1 provides summary of lithological formations and approximate coverage for each lithostratigraphic unit. The northern part of the study area is dominated by Neogene-Quaternary volcanic formations (23–0 Ma) and Quaternary deposits, predominantly alluvial and eluvial sediments (2.6–0 Ma) covering approximately 30.3% and 12.2% respectively. The southern part is covered by Palaeo-Neoproterozoic East African Orogen (Mozambique belt) with a total coverage of approximately 52.0% of the total study area. The unit covers part of the Internal Drainage Basin and approximately entire Pangani Drainage Basin.

Like in many other semi-arid southern and eastern African countries, unconsolidated sediments and volcanic rocks are potential hydrogeological environments with varying productivity capacity in the study (Robins et al., 2006). Substantial amount of groundwater occurs in the volcano-sedimentary sequences of Cenozoic era with main aquifer systems hosted in volcanic formations that occur singularly or superimposed upon one another (Ghiglieri et al., 2010). The present intrusive volcanic rocks are associated with the EARV formation and because of weathering of granitic and meta-sedimentary rocks, the surface is covered by young, unconsolidated alluvial, eluvial and lacustrine sediments especially in the northern part of Kilimanjaro region and around other strato-volcano mountains such as Meru, Hanang' and Oldoinyo Lengai. Denudational forces such as surface run offs during rainy season, rivers, and wind play great role in transporting the weathered rocks from source rock in elevated areas and deposited in depressions and grabens of the rift valley. The rifting and faulting processes, weathering, transportation and deposition of volcanic rocks leads to most potential groundwater resources to be hosted in a complex hydrogeological structure with average groundwater recharge between 2 and 20 mm/a (Taylor et al., 2013).

The present relief of the study area is associated with the local geological and hydrogeological setting. It consists of steep slopes at extrusive volcanic rocks (2900–5886 m above sea level) that change gently ending up into extended flat lands lying between 450 and 1476 m above sea level.

2.3. Data collection, fluoride database creation and quality assessment

Fluoride data was obtained from the Ngurdoto Defluoridation Research Station (NDRS) of the Ministry of Water and Irrigation (MoWI) in Tanzania. The data was collected between 2014 and 2016 by the task force comprising water quality experts from Division of Water Quality (DQW) in a campaign for national mapping of fluoride levels

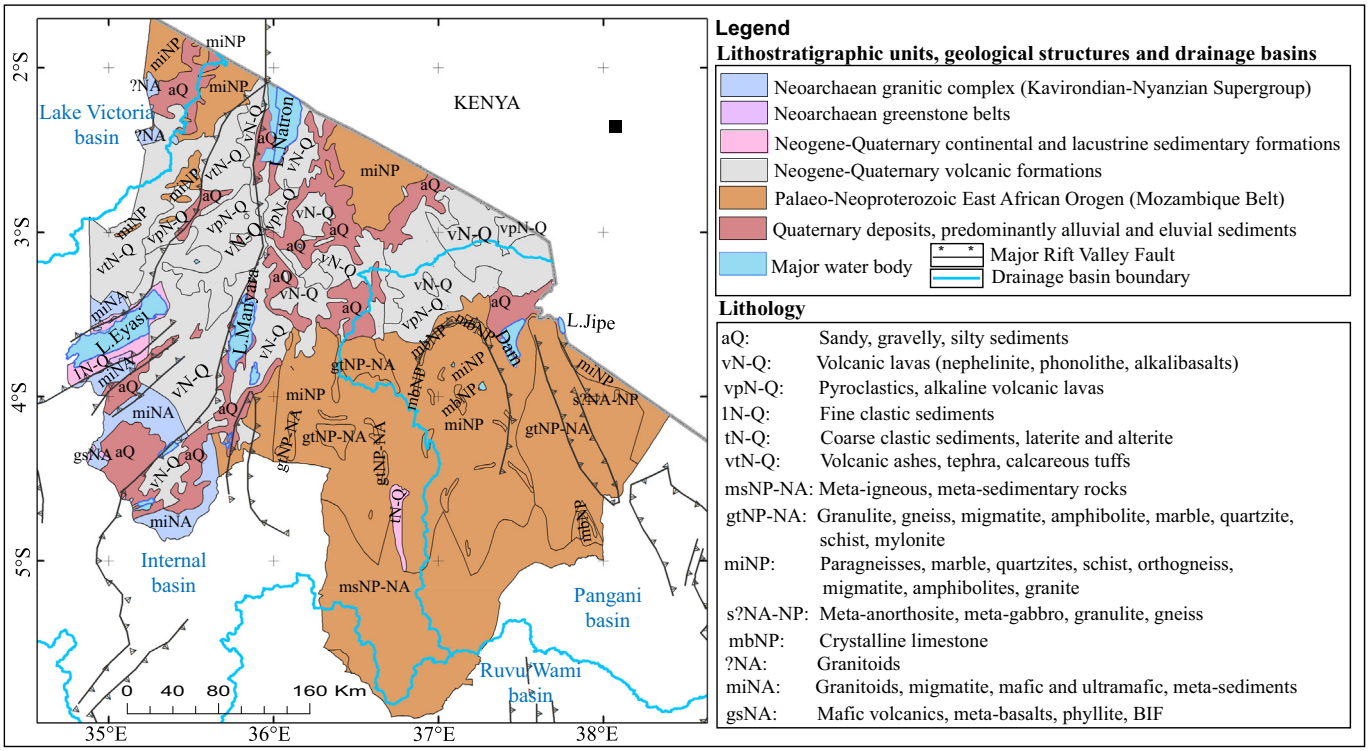


Fig. 2. Geological and hydrogeological setting of northern Tanzania (Data source: GST, 2018; MoIW, 2018).

Table 1
Geological setting and potential hydrogeological environments of northern Tanzania.

Lithostratigraphic unit	Coverage (sq. km)	Lithology	Coverage (sq. km)	Age (Ma)	% Coverage
Palaeo-Neoproterozoic East African Orogen (Mozambique Belt)	49,788				52.0
		gtNA-NP	9067	2800–850	9.4
		mbNP	930	1000–635	1.0
		miNP	25,802	1000–541	27.0
		msNA-NP	13,938	2800–541	14.5
		S?NA-NP	52	1000–850	0.1
Neoarchaean granitic complex(Kavirondian-Nyanzian Supergroup)	3564				3.6
		miNA	3127	2800–2500	3.3
		?NA	318	2800–2500	0.3
Neoarchaean greenstone belts					0.1
		gsNA	118	2800–2500	0.1
Water	498				0.5
		Water	498		0.5
Quaternary deposits, predominantly alluvial and eluvial sediments	11,667				12.2
		aQ	11,667	2.6–0	12.2
Neogene-Quarternary continental and lacustrine sedimentary formations	1203				1.3
		1 N-Q	863	23–0	0.9
		tN-Q	340	23–0	0.4
Neogene-Quaternary volcanic formations	28,979				30.3
		vN-Q	15,943	23–0	16.7
		vpN-Q	8044	23–0	8.4
		vtN-Q	4992	23–0	5.2
Total coverage	95,699				100

in drinking water sources. According to the sampling report (MoWI, 2016), water samples were collected using 1 L clean polythene bottles. Water samples were collected from both developed and undeveloped groundwater-based drinking water sources including boreholes (BH), shallow wells (SW), dug wells (DW) and springs. Fluoride concentrations in mg/L was analyzed in the laboratory at NDRS using fluoride ion selective electrode. The results of water quality analysis were recorded on a Microsoft excel spreadsheet with each water sample allocated a unique sample identification number (SID). In addition, position of each water sampling location was measured using a ± 3 m hand-held Global Navigation Satellite Systems (GNSS) receiver (Garmin GPSMap62s) with position recorded as latitude and longitude both in degree decimal units on a global World Geodetic System of 1984 (WGS84) datum.

Spatial database development involved conversion of water quality data on spreadsheet to GIS shapefile in ArcGIS 10.6 software. The shape file was further linked to elevation and geological data. The geological data was retrieved from the GST whereas the elevation data was downloaded from https://dwtkns.com/srtm30m/ site. The downloaded elevation data was 30 m resolution digital elevation model (DEM) collected by the Shuttle Radar Topography Mission (SRTM) and downloaded in form of tiles. A total of seven (7) tiles were downloaded and mosaicked to get the DEM covering the entire study area. To integrate the geological data with water quality database, the former on a local datum Arc1960 was transformed to a global datum WGS84 latitude and longitude. From this data integration, the water quality parameter spatial database was populated with information related to topography and geology that was used to interpret the spatial analysis results.

2.4. Exploratory spatial data analysis (ESDA)

2.4.1. Normality and homogeneity tests of fluoride concentrations

Graphical and statistical inference methods were used to study the nature of distribution of fluoride concentrations in attribute space. While the histogram and whisker-boxplot were used to understand the overall distribution of fluoride concentrations in drinking water sources, Shapiro test for normality (Shapiro and Wilk, 1965;) and Levene's test for constant variance (Brown and Forsythe, 1974) were used to make statistical inference on normality and homogeneity respectively in fluoride concentrations. Furthermore, boxplots were used to determine water samples with extreme fluoride concentrations. All water samples with fluoride concentrations beyond the minimum (Q_1–3.0 x IQR) and maximum (Q_3 + 3.0 x IQR) value; where Q_1 and Q_3 are first and third quartiles in the ordered dataset, IQR is interquartile range (Q_3-Q_1) and 3.0 is selected boxplot hinge were identified. These points were considered as outliers throughout the study process and some of them were verified by field resampling.

2.4.2. Spatial autocorrelation analysis using global Moran's I Index

The global Moran's I index is the most used indicator of global spatial autocorrelation initiated in 1950's and popularized in 1970's (Cliff and Ord, 1973; Moran, 1950). It measures spatial autocorrelation based on feature locations and attribute values using global Moran's I statistic calculated as (Cliff and Ord, 1981):

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad (1)$$

where n equals the number of observations; x_i represents fluoride concentration x at location i; $\bar{x}$ is the average over all locations of the variable (fluoride concentrations); x_j represents fluoride concentration x at all the other locations (where $j \neq i$); w_{ij} represents spatial weight function used to define the geography of fluoride concentration x at location i in this study.

The interpretation of the calculated Moran's I statistic in eq. (1) is based on its sign and magnitude. In practice, the Moran's I statistic lies between −1 and + 1. If it is high positive and significant, there is enough evidence that similar values (high or low) are clustering in space whereas the high negative and significant Moran's I statistic signifies dispersion (Moran, 1950). When the statistic is zero it implies the observed events are due to a completely spatial random process which is the null hypothesis in spatial dependence analysis.

When calculating the global Moran's I statistic, two parameters, z-score, and the p-value, are important as they provide statistical significance on the calculated Moran's I statistic (Anselin, 1995). The z-score

is calculated using eq. (2) (Cliff and Ord, 1981):

$$z-s-core = \frac{I-E(I)}{\sqrt{var(I)}} \quad (2)$$

where *E[I]* and *var(I)* is an expected and variance of the calculated global Moran's I statistic respectively. The expected I is calculated by eq. (3) whereas the variance is calculated using eq. (4) (Cliff and Ord, 1981):

$$E(I) = -(n-1)^{-1} \quad (3)$$

$$var(I) = \frac{1}{(n-1)(n+1)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2} \times \left[n^2 S_1 - nS_2 + 3\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2\right] - \frac{1}{(n-1)^2} \quad (4)$$

where

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2 \quad (5)$$

and

$$S_2 = \sum_{i=1}^{n}\sum_{j=1}^{n}(w_{i.} + w_{.i})^2 \quad (6)$$

From eqs. (1)–(5), calculation of parameters requires spatial weight matrix with weight elements w_{ij} corresponding to a pair of observations at locations *i* and *j*. While $w_{ij} \neq 0$ indicates the spatial interaction between two observations, $w_{ij} = 0$ indicates lack of interaction between two observations (Anselin, 1995). The spatial weight matrix w_{ij} can take many forms such as contiguity for areal units or distance-bands for point units (Anselin, 1995; Getis and Ord, 1992). However, the Rook's and Queen's (or King's)-based spatial weights are the most commonly used especially when observation locations are irregularly spaced (Ping et al., 2004). The two methods differ by the way neighbourhood j of i is conceptualized. For Rook's case, the w_{ij} is set to 1 if the pairs shares a common edge and 0 otherwise while in the Queen's case it is set to 1 if the pair shares either a common edge or a vertex and 0 otherwise (Berry and Marble, 1968). In practice, the Queen's method has less nonzero elements because of the definition of neighbourhood and was therefore used in this study.

The observation locations were converted into Thiessen polygons using proximity function in GeoDa 1.14.0 software (Anselin, 1995).The Thiessen polygons or Voronoi polygons are polygons generated around a set of points in a given space by assigning all locations in that space to the closet member of the point set (Yamada, 2016). For this study, observation location was considered as a realization of many observation locations within each polygon where fluoride concentrations did not vary insignificantly. The Queen's contiguity spatial weight matrix was used to calculate the Moran's I statistic whereas its statistical significance was tested through 999 randomization.

2.5. Spatial cluster and spatial outlier analysis

2.5.1. Moran's scatterplot and GIS analysis

Spatial patterns were studied using Moran's scatterplot and GIS. In this study, the spatial pattern refers to an individual water sampling location with fluoride concentration in the neighbourhood of other sampling locations similar or dissimilar fluoride concentrations. The Moran scatterplot is constructed based on the concept of bivariate linear regression analysis where the spatially lagged variable $\sum_j w_{ij}x_j$ is regressed on x_i giving the slope which is Moran's I (Anselin, 1996).In principle, the Moran's scatterplot is centred at the mean of zero and differences between observed and the sample mean are compared to the mean. Similarly, the differences between the spatially lagged and sample mean are compared to the mean. Through this comparison, the high-high and low-low (i.e, water samples with similar fluoride concentrations at neighbouring locations) spatial patterns were determined and the sampling locations were regarded as spatial clusters. In a similar way, low-high and high-low (i.e, water samples with dissimilar fluoride concentrations at neighbouring locations) were determined and regarded as spatial outliers. Since Moran's I statistic is calculated based on normality and homogeneity assumption in observed events, fluoride data was normalized through transformation using Box-Cox method (Box and Cox, 1964; Zhang et al., 2008). The spatial patterns were studied using both raw and transformed fluoride data.

2.5.2. Identification and GIS mapping of significant spatial patterns

The Moran's scatterplot classifies dataset into spatial patterns which can be clustered (spatial cluster) or exist individually (spatial outlier). The statistically significant spatial patterns can be identified using the local Moran's Index calculated as (Anselin, 1995; Getis and Ord, 1996):

$$Ii = \frac{x_i - \bar{x}}{\sigma^2}\sum_{j=1, j\neq i}^{n}\left[w_{ij}(x_j - \bar{x})\right] \quad (7)$$

where σ^2 represents the variance of fluoride concentrations *x* and the rest of the parameters are as defined in eq. (1).

In practice, the local Moran's I index classifies spatial patterns into two major categories, that is, spatial clusters indicated by a high positive local Moran's I statistic and spatial outliers indicated by a high negative local Moran's I statistic (Anselin, 1995). In regional groundwater contamination studies, low-low and high-high spatial clusters can be regarded as "regional cool spots" and "regional contamination hotspots" respectively. The regional cool spots are important places to be identified as they can be alternative sources of potable water with low fluoride content in the naturally contaminated hydrogeological environments (aquifers).

For practical purpose in scientific investigation using local spatial statistical indices, it is important to provide scientific inferences. In order to do so, the local Moran's I is usually standardised to ensure normal distribution when its significance level is tested (Anselin, 1995). However, the normality assumption may not be met when the dataset is heavily skewed. Anselin (1995) proposed conditional permutation method which works without assumption about the dataset. With this method, the value *x* at location *i* is fixed while its influence on the global trend, that is, Moran's I is evaluated by reshuffling randomly other neighbouring values *x* at location *j* where $j \neq i$. For every location, the local Moran's I index is calculated to form a reference distribution. In this approach, the reference distribution for statistic under the null hypothesis is calculated by randomly permuting the observed values over the locations and using the resulting distribution to calculate what Anselin (1995) calls the pseudo p-value. The pseudo p-value is calculated as $(R + 1)/M + 1$ where; R is the number of times the computed Moran's I from the spatial random datasets, that is, the permuted datasets is equal to or more extreme than the observed statistic and M is the number of permutations.

2.6. Identification of fluoride bearing minerals in rocks

The mineralogical analyses were performed by transmitted light petrographic microscope on fourteen (14) rock samples. The analyses were aimed at examining the presence of fluoride bearing mineral phases in the rock samples collected from north-east hotspot (11 samples) and east cool spot (3 samples). The study of the fluoride bearing mineral phases were done based on polished sections. The rock polished sections were prepared at the African Minerals and Geosciences Centre (AMGC) laboratories, Kunduchi, Dar es Salaam in Tanzania following

standard polished section making procedures. The petrographic examinations of the polished sections were carried out at the Department of Geology laboratories, University of Dar es Salaam, Tanzania. The analyses were done using a Carl ZEISS made Primotech petrographic microscope. The micrographs were acquired through an inbuilt camera in the Primotech microscope by MATSCOPE application which was also used to manipulate the images.

2.7. Potential health-based risk assessment in regional hotspots and cool spots

To evaluate risk levels in the regional fluoride hotspots and cool spots, all water samples in each category were classified according to lithological units and method through which groundwater is abstracted. For each class, fluoride concentrations were classified into five groups according to associated human health effects when consuming water with fluoride levels in that range (Dissanayake, 1991; WHO, 2011) and number of water samples in each group was determined. The risk level was studied based on probabilities calculated as number of water samples in that group divided by the total number of water samples in specific lithological unit.

2.8. Software used in data analysis and mapping

The calculation of Moran's I statistic both global and local, and identification of regional fluoride hotspots and cool spots were performed using GeoDa software (version 1.14.0). The statistical analysis and graphical visualization of the results were performed using RStudio software (version 1.1.463). Data preparation and mapping were done using ArcGIS software (version 10.6) and QGIS software (1.18.23).

3. Results and discussion

3.1. Fluoride concentrations in drinking water sources in northern Tanzania

Table 2 provides information on fluoride concentrations in drinking water sources in northern Tanzania. It further provides comparison between measured fluoride concentrations and WHO guideline value of 1.5 mg/L, which is same as adopted Tanzania standard value for drinking water in fluorotic regions. The median concentration for all water samples is less than WHO guideline and Tanzania standard value. However, it varies between water sources. The median concentration for shallow wells (SW), boreholes (BH) and springs is less than the recommended WHO guideline and the Tanzanian standard compared to that of the dug wells (DW) is approximately three times higher. On the other hand, the scale of variation in fluoride concentrations varied among different geological sources as indicated by the interquartile range. The largest variation was also in dug wells (7.77 mg/L) followed by shallow wells (4.48 mg/L) and boreholes (2.79 mg/L). The springs had the smallest variability of 1.77 mg/L, which was less than the overall interquartile range of 2.90 mg/L. It was further envisaged that each water source contained samples with fluoride concentrations that exceeded both the WHO recommended guideline as well as the Tanzanian standard Overall, 42% of water samples had fluoride concentrations >1.5 mg/L which exceeds the previously reported 30% (Thole, 2013) and it varied between water sources with dug wells having the highest percentage (69%) followed by shallow and boreholes which had 46% and 45% respectively. The springs had the least percentage (35%) which was smaller than the overall. The overall maximum concentration was 74.00 mg/L which was in boreholes. Likewise, each source contained maximum concentration which exceeded by a factor of >20 as compared to the WHO drinking water guideline of 1.5 mg/L as indicated in Table 2.

3.2. Statistical and spatial distribution of fluoride concentrations in drinking water sources

As shown in Table 2, the overall fluoride concentrations distribution was positively skewed as indicated by the mean greater than the median values. This was statistically inferred by Shapiro-Wilk test for normality where the p-value was <0.05 ($w = 0.508$, p-value<<0.000). From Whisker-boxplot for the entire dataset, extreme values were calculated using the 25th, 75th percentiles and interquartile range. A total of 118 water samples were identified as water sources with extreme fluoride concentrations. Fluoride concentrations ranged between 8.84 and 74.0 mg/L for extreme values whereas for standard distribution ranged between 0.01 and 8.84 mg/L (Fig. 3). While Fig. 3(a) shows statistical distribution of the standard range, (b) shows relative spatial distribution of water sources with extreme concentrations in relation to the standard range of the sample distribution.

Further study on the statistical distribution of the dataset was on homogeneity in fluoride concentrations. The robust Levene's test for homogeneity of variance indicated that the variance is not constant ($F = 15.148$, df = 3 and p-value < 0.000).

In geographical space, the global Moran's I index was used to study spatial aspects of fluoride concentrations. Using 1201 by 1201 spatial weight matrix of 0 and 1 spatial weights with first order queen's contiguity method and Thiessen polygons, the global Moran's I statistic (Eq. 1) was calculated. The statistical significance of the global Moran's I index was tested using conditional permutation method in which 999 conditional permutations were used. According to Anselin (1995), the large z-score and small pseudo p-value gives a strong evidence of the statistically significance of the calculated global Moran's I. In this study, a positive statistically significant global Moran's I statistic was obtained (Moran's $I = 0.514$, z-value =30.111 and p-value = 0.001) implying that fluoride occurrence in groundwater is not due to completely spatial random processes. Fig. 4 shows the Moran's scatterplot with a positive Moran's I statistic as an indication of spatial clustering in fluoride concentrations.

3.3. Exploratory analysis of potential spatial patterns using Moran's scatterplot

Before determining significant regional fluoride hotspots and cool spots, the Moran's scatterplot and GIS analysis were used to study potential spatial clusters (High-high and Low-low) and spatial outliers (Low-high and High-low). Since results of spatial cluster and spatial outlier analysis are affected by definition of spatial weight function, presence of extreme values and non-normality of the data values (Zhang et al., 2008), the study was based on the transformed dataset ($n = 1201$) and dataset without extreme values ($n = 1083$). The latter was obtained through calculation based on the whisker-boxplot

Table 2
Fluoride concentrations (mg/l) in drinking water sources in northern Tanzania.

Source	N	Min.	25%	Median	75%	IQR	Mean	SD	Maximum	N(F^- > 1.5)
All	1201	0.01	0.40	1.10	3.30	2.90	3.36	6.40	74.00	507(42%)
DW	72	0.27	0.72	4.45	8.50	7.77	7.55	11.15	62.00	50(69%)
SW	228	0.01	0.57	1.17	5.05	4.48	4.75	7.86	42.00	106(46%)
BH	436	0.09	0.52	1.21	3.32	2.79	3.11	6.05	74.00	190(45%)
Spring	465	0.03	0.23	0.81	2.00	1.77	2.27	4.21	31.00	161(35%)

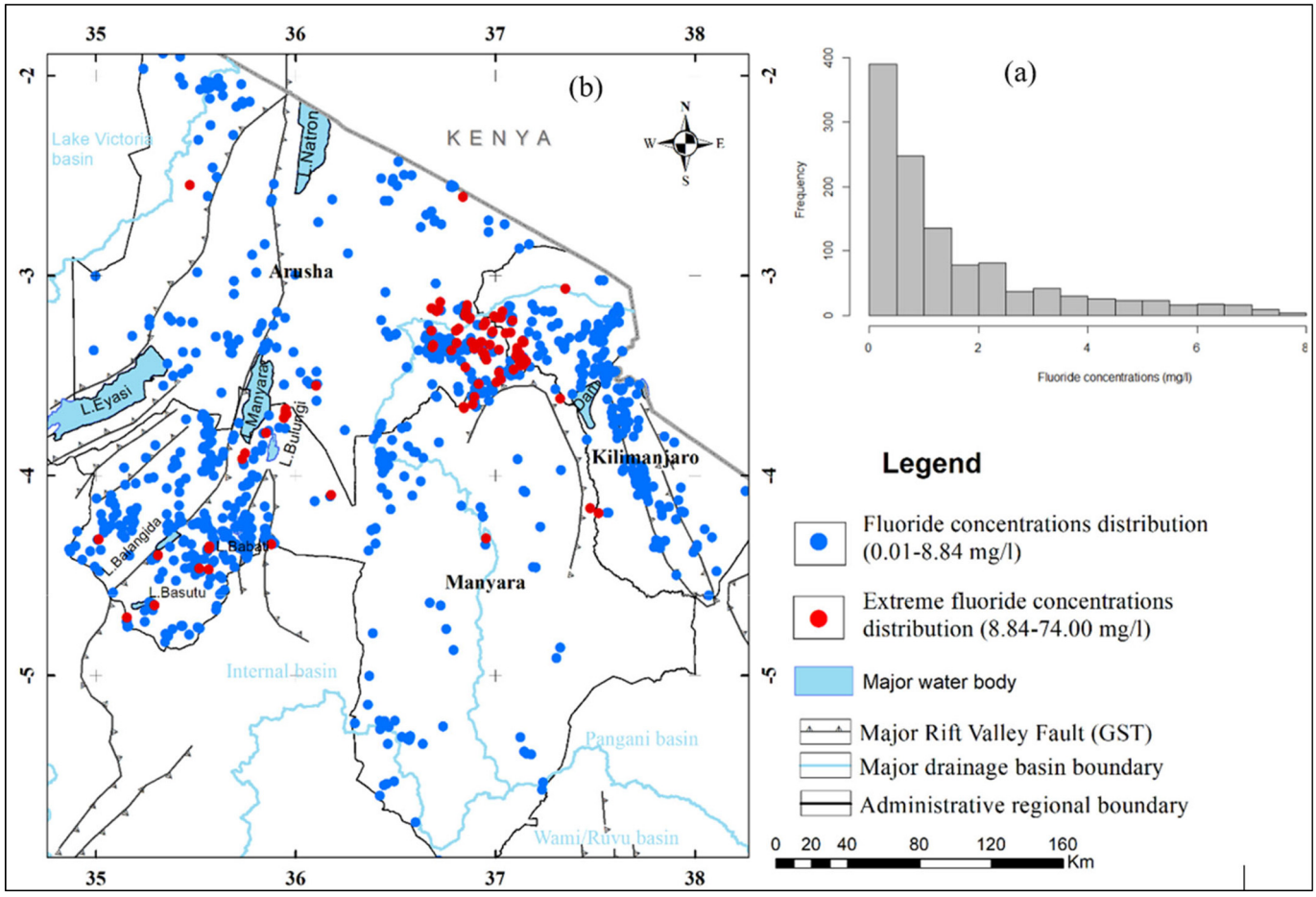


Fig. 3. (a) Statistical distribution of standard range and (b) Relative spatial distribution of extreme concentrations.

whereas the former was obtained through Box-Cox transformation. Potential patterns were studied based on Queen's spatial weight matrix first order (q1), second order (q2), third order (3) and fourth order (q2). Table 3 shows number of potential spatial patterns from which significant patterns were identified. The potential patterns in each data treatment were exported into ArcGIS software where they were integrated with tectonic unit boundaries, major rift valley faults and elevation data for interpretation.

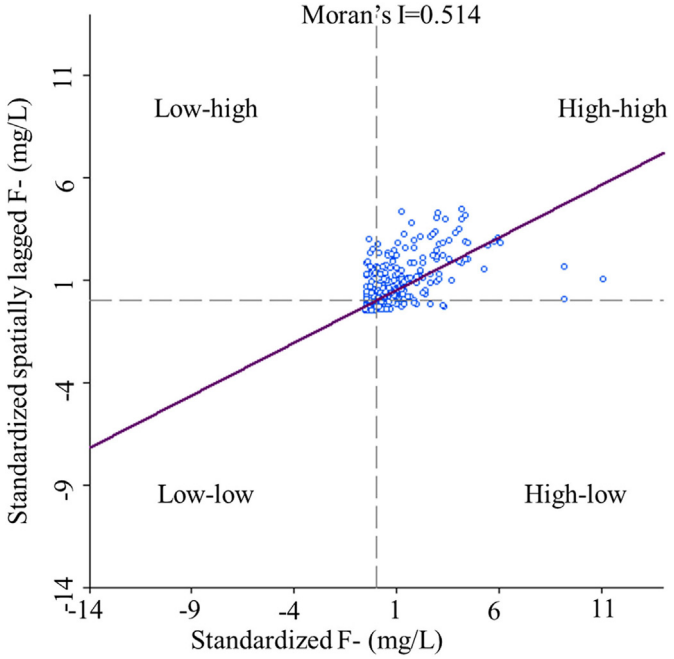


Fig. 4. Indicator of spatial autocorrelation in fluoride concentrations.

As indicated on Fig. 5, the high-high potential spatial patterns were located around major volcanic mountains particularly Mt. Meru (Me.) in the north-east and Mt. Hanang' (H.) in the south-west of the study area. Furthermore, the potential patterns in this category were concentrated along the tectonic unit boundaries. The major low-low potential spatial patterns were located along the major and minor Rift Escarpment in the west and eastern part of the study area, respectively. It was further envisaged that potential spatial patterns were affected by order of the spatial weight matrix. As shown in Fig. 5(a), there was a clear distinction between the high-high and low-low potential spatial patterns where the latter indicated a circular arc around the former as an indication of possible spatial processes aggravating fluoride contamination in groundwater sources. As the order increased, potential low-

Table 3
Number of potential spatial patterns in fluoride concentrations.

Data treatment	High-high	Low-low	Low-high	High-low	Total
Box-Cox transformed, q1	464	521	109	107	1201
Box-Cox transformed, q2	481	485	126	109	1201
Box-Cox transformed, q3	452	508	125	116	1201
Box-Cox transformed, q4	454	472	138	137	1201
Extreme values excluded, q1	273	568	152	90	1083
Extreme values excluded, q2	273	558	164	88	1083
Extreme values excluded, q3	266	540	202	75	1083
Extreme values excluded, q4	287	503	219	74	1083

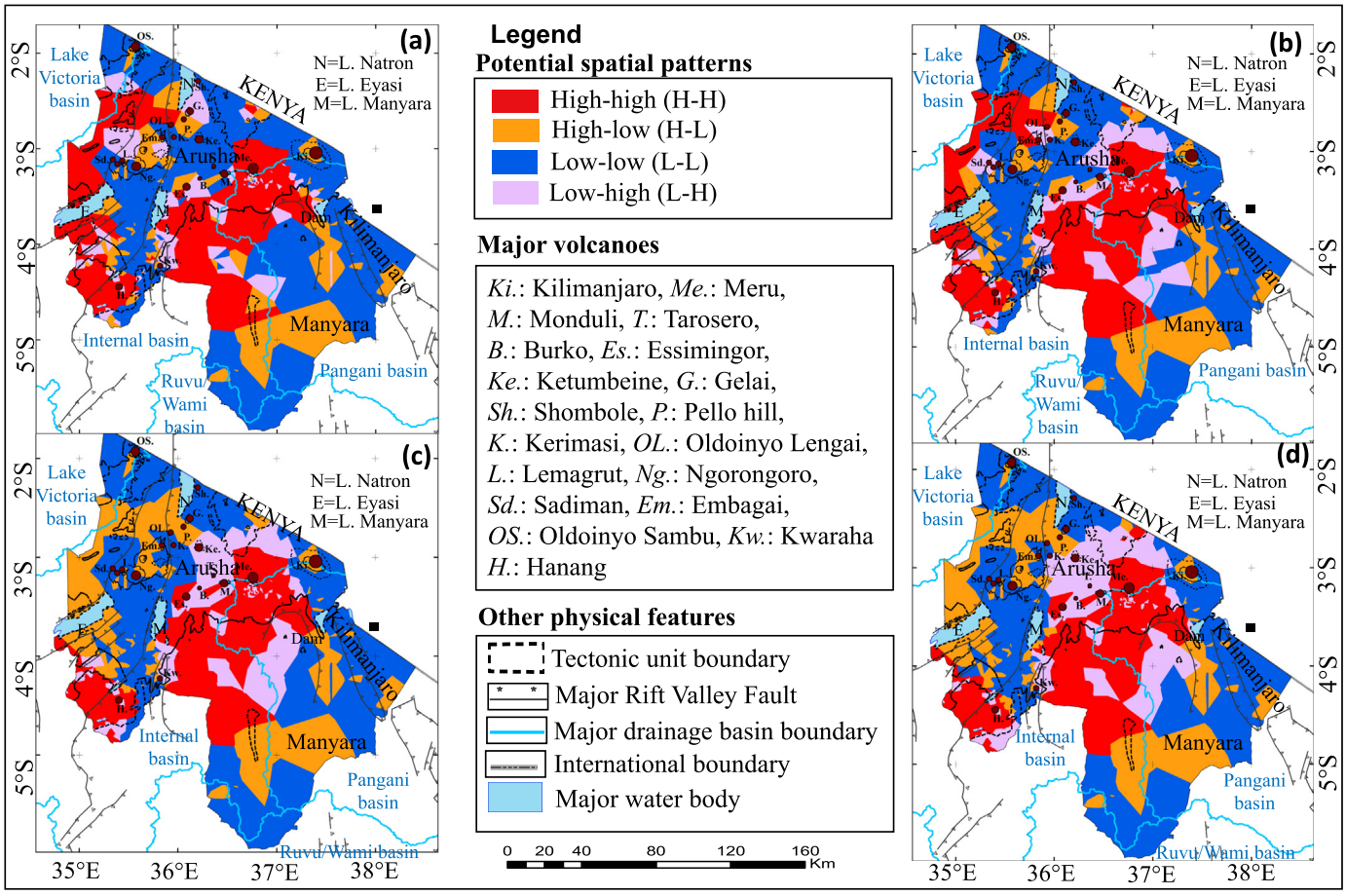


Fig. 5. Spatial distribution of potential spatial patterns using Queen's spatial weight matrix (a) first order –q1; (b) second order-q2; (c) third order-q2; and (d) fourth order-q4.

high spatial patterns emerged at the boundary of the high-high potential spatial pattern. This can be interpreted as an effect of natural blending between water with low and high fluoride concentrations through exchange of water from lateral movement of water between lithological units (Ghiglieri et al., 2010) though it may require further verification. Compared to other data treatments in Table 3, second order Queen's spatial weight matric with Box-Cox transformed data indicated many water sources as high-high potential spatial patterns although the difference was not significant. Therefore, significant spatial patterns were determined based on the four data treatments.

3.4. Spatial clusters and spatial outliers' identification and mapping

The univariate local Moran's I index was used to identify and map significant spatial clusters and spatial outliers. The calculation was based on eq. (7) and since the mean and variance are used in the calculation, the transformed version of the dataset was used to avoid problems associated with skewed data in spatial cluster and spatial outlier analysis (Anselin, 1995) The Queen's spatial weight function was used in the calculation of local Moran's I indices. Since the results of spatial analysis are affected by the definition of neighbourhood, the calculation of the local Moran's I statistic at each location was done even at higher orders. While Table 4 shows neighbourhood characteristics for each data treatment, Table 5 shows over all statistical significance test results when estimating spatial patterns through 999 permutations. From Table 4, it was observed that the irregularities in water sampling locations affected the cardinality of the neighbours. The cardinality decreased with an increase in the order. The same effect was also observed statistical significance test results of the estimated spatial clusters and spatial outliers in Fig. 6.

In comparison with results when analysing potential spatial patterns using Moran's scatterplot and GIS in Fig. 5, the number of significant spatial clusters (High-high and Low-low) increased with an increase in the order of spatial weight matrix. With Queen's first, second, third order and fourth order spatial weight matrix, 49% ($n = 226$), 70% ($n = 337$), 84%($n = 378$) and 87% ($n = 395$) of water samples were classified as significant high-high spatial clusters respectively. By comparing fluoride concentrations for every water sample with WHO guideline and Tanzania standard value of 1.5 mg/L, 98.6% ($n = 223$), 96.1% (n = 337), 93.6% and 92.9% ($n = 367$) of water samples with Queen's first, second, third and fourth order respectively. As shown on Fig. 6, water samples with fluoride concentrations <1.5 mg/L were either close to the volcanic mountains which are primary groundwater recharge areas or far away from the tectonic unit boundaries with a few exceptions. The occurrence of water sources with low fluoride concentrations in the proximity of volcanic mountains could be associated with local terrain slopes which might be controlling the residence time between groundwater and fluoride bearing rocks in the study area.

Table 4
Data treatment and neighbor characteristics when calculating local indicators of association.

Data treatment	Neighbor characteristics				
	Minimum	Maximum	Mean	Median	%non-zero
Box-Cox transformed, q1	2	13	5.78	6	0.48%
Box-Cox transformed, q2	4	43	19.15	18	1.59%
Box-Cox transformed, q3	9	104	42.51	40	3.54%
Box-Cox transformed, q4	14	188	77.35	74	6.44%

Table 5
Data treatment and statistical significance test of spatial clusters and spatial outliers.

Data treatment	Significance test				
	Moran's I	E[I]	Sd(I)	z-scores	p-value
Box-Cox transformed, q1	0.671	−0.0008	0.0177	38.03	0.001
Box-Cox transformed, q2	0.590	−0.0008	0.0093	63.37	0.001
Box-Cox transformed, q3	0.513	−0.0008	0.0063	81.91	0.001
Box-Cox transformed, q4	0.439	−0.0008	0.0045	97.36	0.001

For the significant spatial outliers (Low-high and High-low), the number of water samples in class increased with an increase in the order of the spatial weight matrix. All water samples in the significant low-high spatial outlier had fluoride concentrations <1.5 mg/L whereas 81.1%, 88.6%, 82.8% and 83.5% of water samples in significant high-low spatial outlier had fluoride concentrations above 1.5 mg/L in the four data treatments respectively. The significant low-high spatial outlier existed around the significant high-high spatial clusters which may be reflecting typical spatial process through which fluoride is mobilized into groundwater. As the order of spatial weight matrix increased, the size of the significance low-high spatial outlier increased as well forming a ring-like around the significant high-high spatial cluster. In a similar fashion, the significant high-low spatial outlier existed around the significant low-low spatial cluster (Fig. 6b-c). Most of the significant high-high spatial clusters existed around stratovolcano mountains mainly Mt. Meru in the north-east and Mt. Hanang'in the south-west of the study area. Similarly, substantial number of significant high-high spatial clusters existed along the Palaeo-Neoproterozoic East African Orogen (Mozambique Belt), Neogene-Quaternary volcanic formations and, Neoarchaean granitic complex (Kavirondian-Nyanzian Supergroup) and Neoarchaean greenstone belts. The significant low-low spatial clusters dominated the major and minor rift escarpments in the west and east of the study area respectively.

3.5. Spatial distribution of significant fluoride hotspots and cool spots

Despite the number of significant spatial patterns increased with an increase in the order of the spatial weight matrix, the emerging significant spatial outliers especially the low-high spatial outliers existed at the boundary of the significant high-high spatial cluster. In other words, the presence of spatial outliers at the boundary may imply two important phenomena: (i) the natural blending of high and low fluoride water through lateral movement of groundwater in which groundwater is exchanged between lithological units as previously reported by Ghiglieri et al. (2010) in their study to the north-east of Mt. Meru and (ii) ion exchange process in mineral phase bearing fluorine as groundwater flows (Jacks et al., 2005).

To map the significant spatial patterns, the second order based spatial weight matrix was used. The selection was based on the cardinality of neighbours used in the calculation of local indicators of spatial association and the percentage of water samples with fluoride concentrations >1.5 mg/L. Fig. 7 shows distribution of significant spatial patterns which are technically known as regional fluoride hotspots and cool spots in this study. Two and four distinct regional fluoride hotspots and cool spots respectively were identified. The largest regional fluoride hotspot originated around Mt. Meru in the west of Mt. Kilimanjaro and extended towards south-east of lake Manyara along Mozambique belt and Pliocene-recent volcanics boundary. Within this hotspot, not significant spatial patterns existed especially in near south-east of Mt. Meru where Arusha city

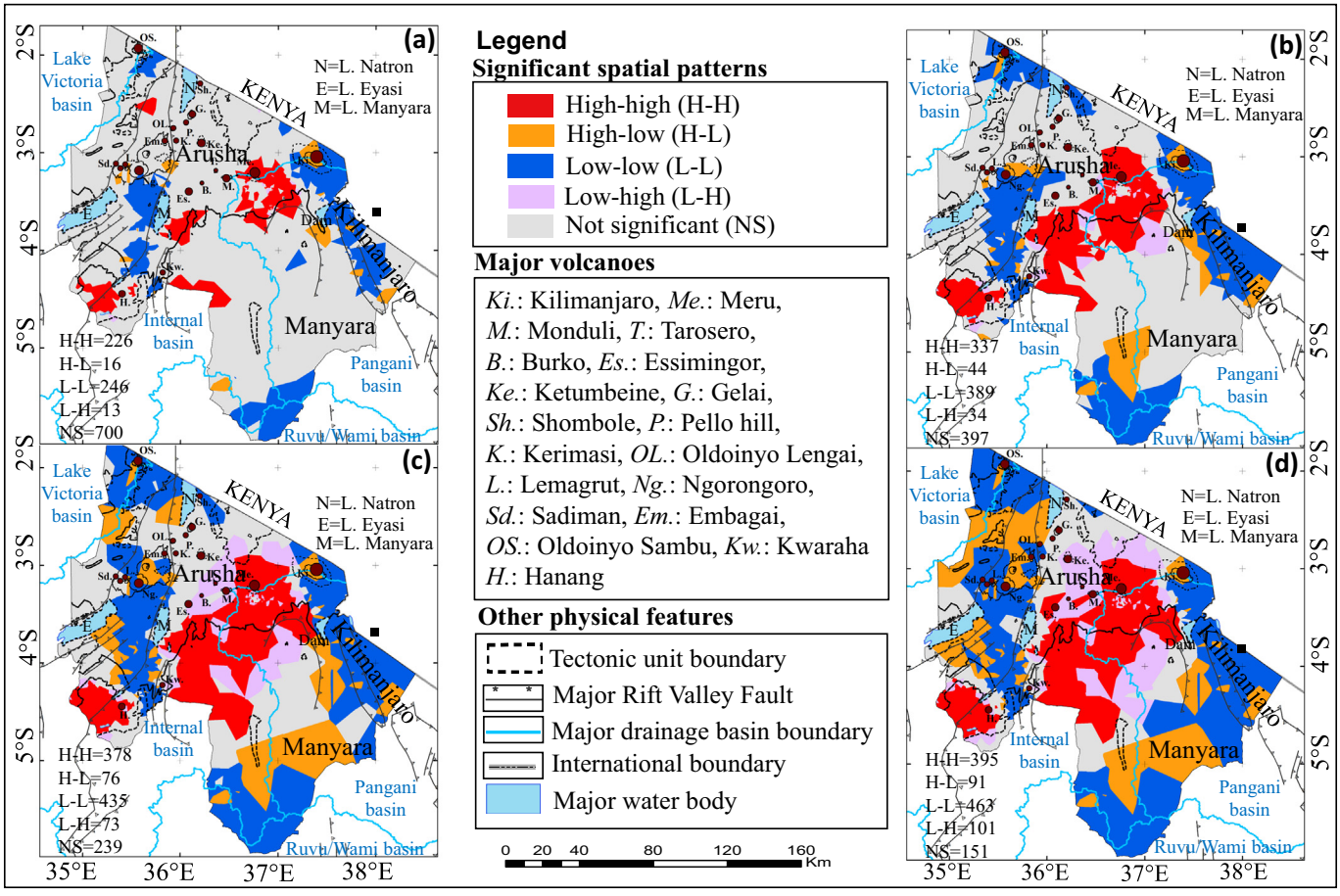


Fig. 6. Spatial distribution of significant fluoride hotspots and cool spots estimated using Queen's (a) first order, (b) second order, (c) third order and (4) fourth order with Box-Cox transformed data (p-value = .05).

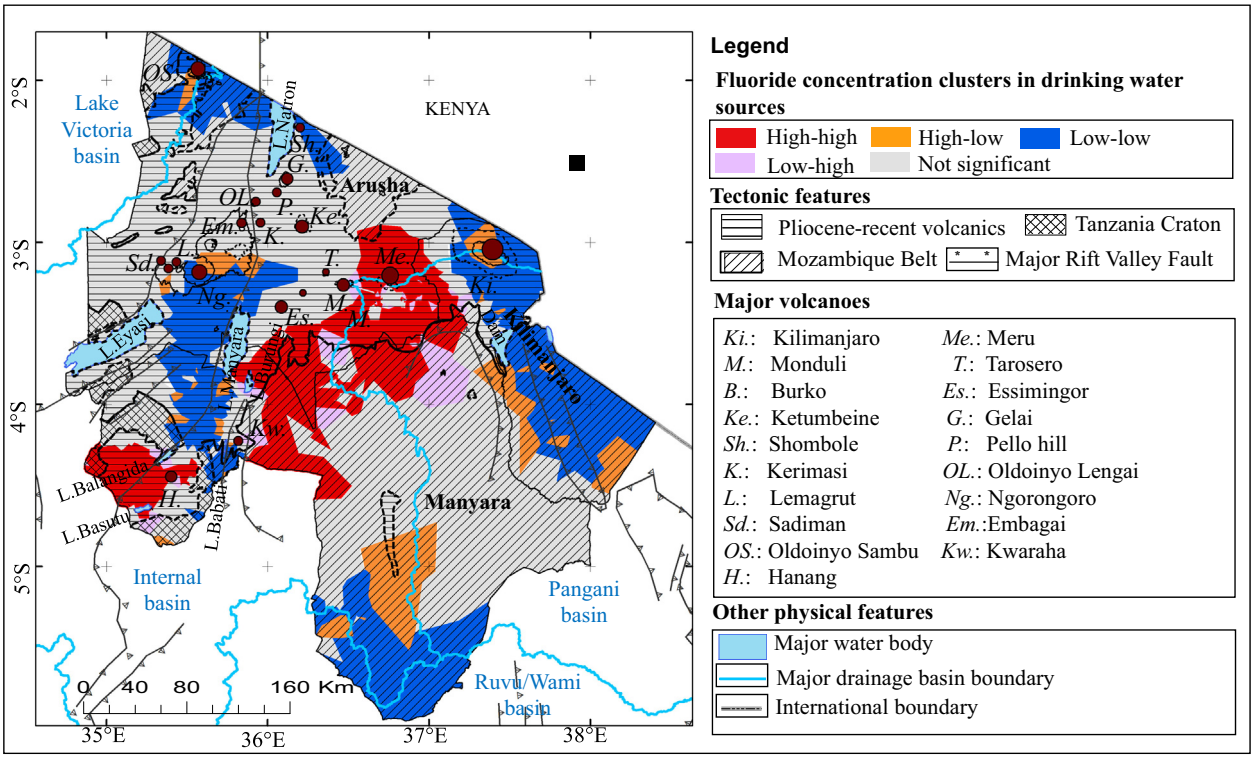


Fig. 7. Spatial distribution map of regional fluoride hotspots and cool spots in northern Tanzania.

is situated. The second regional fluoride hotspot existed around Mt. Hanang except in the south-east direction.

For the regional fluoride cool spots, the two largest cool spots were located along the major and minor rift escarpment in the west and east part of the study area, respectively. The west cool spot existed between Ngorongoro Crator where Ngorongoro Conservation Area is located and north of lake Manyara. The cool spot extended to the south along the recent rift escarpment that originates from lake Natron in the north at the border between Tanzania and Kenya and ends to the south where lake Babati, a fresh water lake, is situated (Baudouin et al., 2016).The eastern cool spot originated around Mt. Kilimanjaro and extended along Usambara block mountain ranges. The other two small cool spots existed in the north and south of the study area. The northern cool spot existed around Oldoinyo Sambu volcanic mountain at the border between Tanzania and Kenya. The southern cool spot existed within Manyara region and the confluence between Pangani, Wami/Ruvu and Internal drainage basin boundary. Fig. 8 gives statistical distribution of fluoride concentrations in each hotspot and cool spot.

The scale of variation in fluoride concentrations for the north-east fluoride hotspot was relatively high compared to that of the south-west as indicated by large interquartile range (Fig. 8a). The minimum fluoride concentration in the latter was equal to 1.5 mg/L while that for the former was below 1.5 m. The scale of variation for the cool spot was relatively high for the south cool spot and low for the northern cool spot whereas did not show significance difference between east and west cool spot (Fig. 8b). The scale of variation in the regional hotspots may be associated with local geological setting.

3.6. Rocks mineralogical composition and fluoride bearing phases

The abnormal fluoride concentrations within the East African Rift Valley have been linked to the development of hyper-alkaline volcanic rocks in the rift zone mainly nepheline and carbonatite magmas and associated ash deposits (Edmunds and Smedley, 2013). Petrographic study of 14 rock samples from the north-east hotspot (11 samples) and east cool spot (3 samples) indicated volcanic igneous rocks of intermediate chemical composition between mafic and felsic rocks. As indicated on Fig. 10, the dominant mineral composition include pyroxene, amphibole, plagioclase, feldspaphoids, biotite, titanite and hornblende (Fig. 9a) with major fluoride bearing phases being titanite, amphibole, hornblende and biotite (Fig. 9b).

3.7. Probability of safe drinking water source in regional fluoride hotspots and cool spots

Since the recommended safe range of fluoride lies between 0.5 and 1.5 mg/L (WHO, 2011), consumption of water with fluoride concentrations below 0.5 mg/L, 1.5–4.0 mg/L, 4.0–10.0 mg/L and above 10.0 mg/L may cause dental caries, dental fluorosis, skeletal fluorosis and crippling fluorosis respectively (Dissanayake, 1991). For each range, the number of samples in respective lithological unit was determined, and probability calculated. Fig. 10 shows calculated probabilities of having safe and unsafe source of drinking water in significant hotspots and cool spots respectively. Although the Neogene-Quartenary volcanic formations and Neogene-Quartenary deposits provide substantial quantities of groundwater, the probability of having safe water for drinking was very low, that is, 0.03, 0.05 and 0.03 in vpN-Q, vN-Q and aQ respectively (Fig. 10a). The presence of many sources with high and few with low fluoride concentrations is one of the challenges affecting the blending technology adopted by Arusha Urban Water Supply Authority (AUWASA). During field sampling campaign in 2018, fluoride concentration of 4.6 mg/L was measured at the blending tank as the net mass of fluoride in drinking water supplied throughout Arusha City in the south flanks of Mt. Meru. Because of few samples in other lithological units, the probabilities are not discussed in Fig. 10a.

Despite all samples in the cool spots had fluoride concentrations below 1.5 mg/L, the probability of accessing safe water ranged between 0.18 and 0.39 for water sources in Neogene-Quartenary volcanic formations. For Quartenary deposits

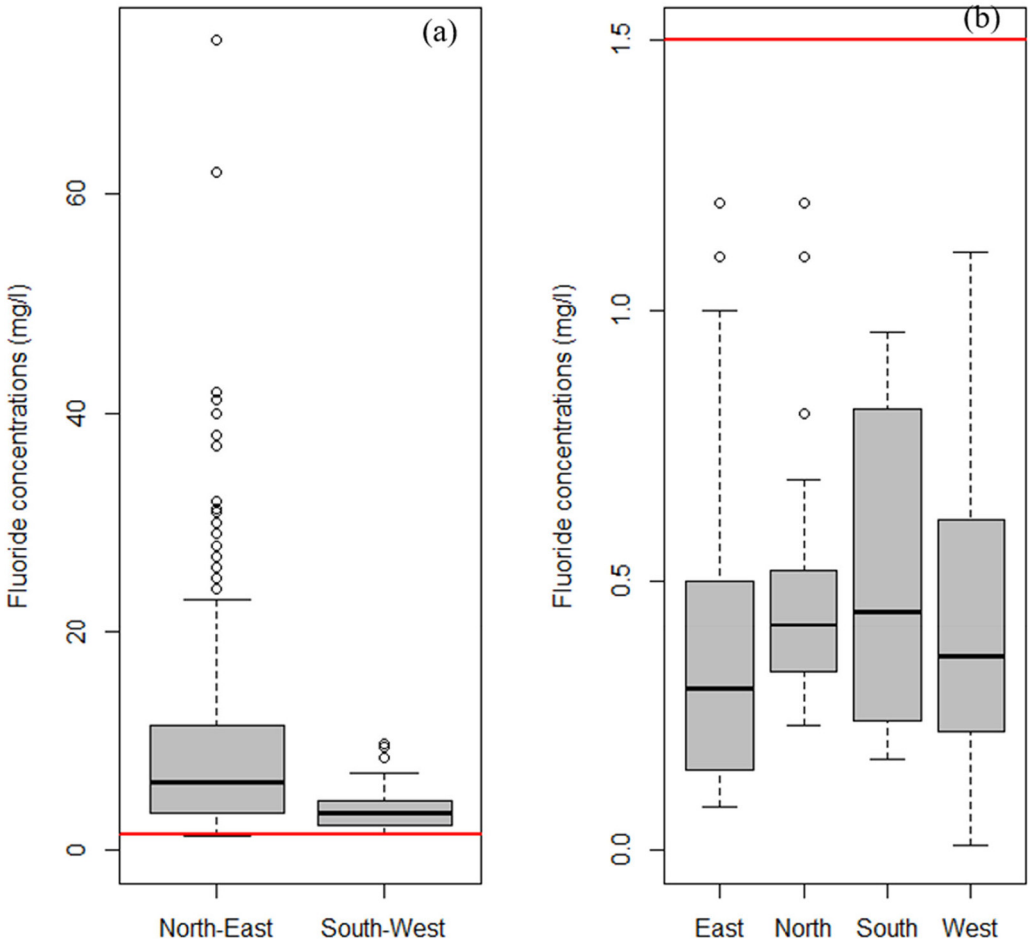


Fig. 8. Fluoride concentrations distribution in regional fluoride (a) hotspots and (b) cool spots.

predominantly alluvial and eluvial sediments, the probability was 0.25 whereas for Palaeo-Neoproterozoic East African Orogen-Mozambique belt it ranged between 0.35 and 0.39. Similarly, for Neoarchean granitic complex-Kavirondian-Nyanzian Supergroup, the probability was 0.62. On the other hand, there is a risk of dental caries among communities depending on water sources in these cool spots as a primary source of drinking water. The probabilities are indicated on Fig. 10b.

3.8. Probability of safe drinking water source based on mode of abstraction in regional fluoride hotspot and cool spot

Groundwater access in the study area is through hand dug wells (DW), shallow wells (SW), boreholes (BH) and springs. To evaluate possibility of targeting safe source, samples in hotspots and cool spots were grouped according to method of abstraction and probability of getting

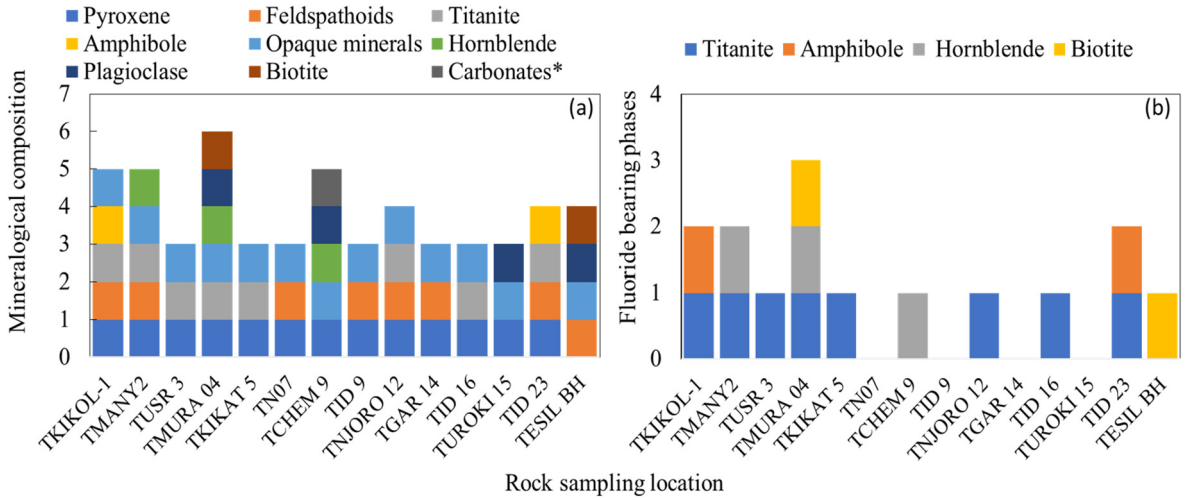


Fig. 9. Summary of (a) mineralogical composition of groundwater bearing rocks and (b) potential fluoride bearing phases.

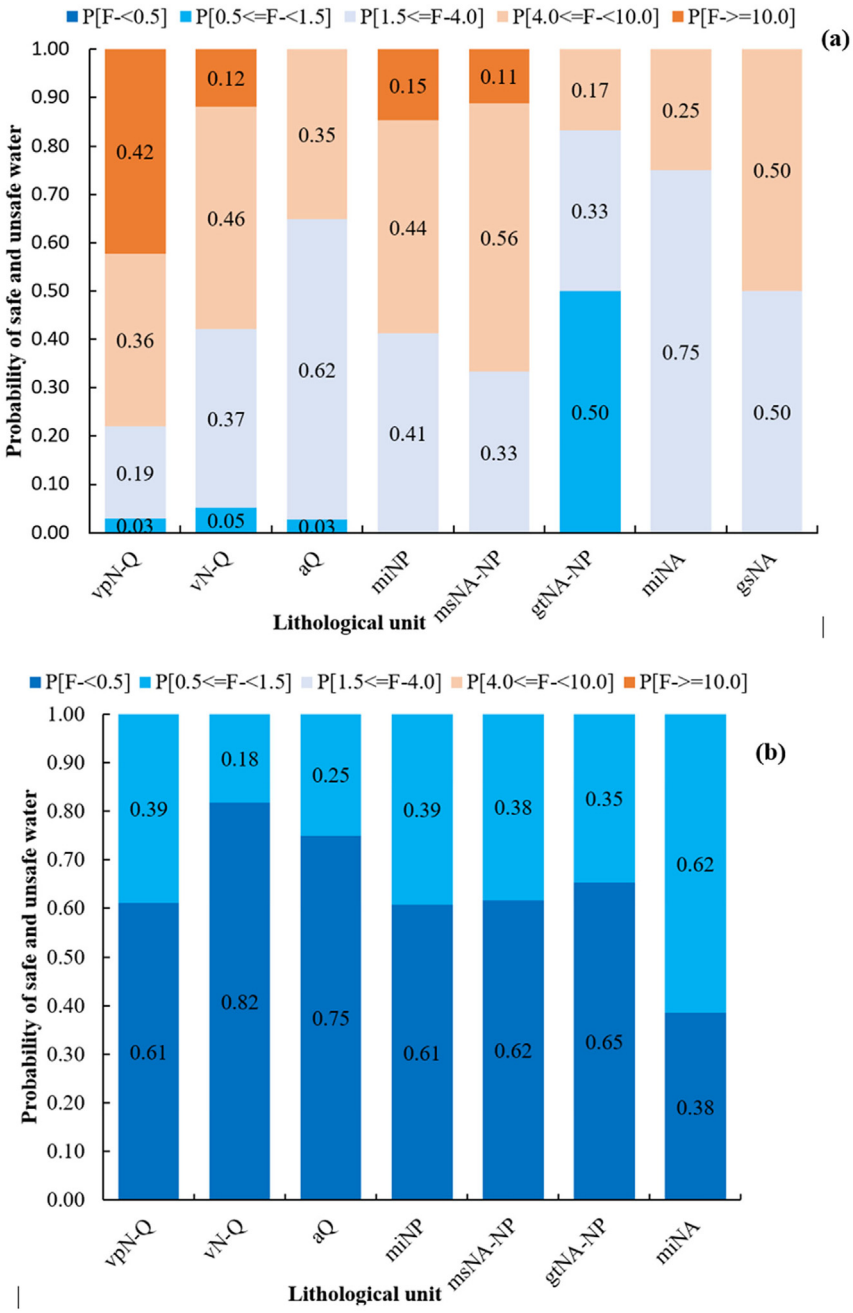


Fig. 10. Probability of safe drinking water sources in regional fluoride (a) hotspots and (b) cool spots according to local geological setting.

safe drinking water calculated. Fig. 11 shows probability of accessing safe water through DW, SW, BH and springs in significant hotspots, cool spots and insignificant patterns. The probability of accessing safe water for drinking is very low, that is, 0.01, 0.05 and 0.06 through shallow wells, boreholes and springs respectively while it is 0.00 through dug wells (Fig. 11a). On the other hand, not all drinking water sources in the significant cool spots are safe for human consumption although they contain fluoride concentrations below 1.5 mg/L. The probability was lowest for natural springs (0.13) while highest in the dug wells (0.60). For shallow wells and boreholes, the probability was 0.44 and 0.42 respectively (Fig. 11b).

4. Conclusions

With this regional scale study, we provide systematic spatial understanding of fluoride occurrence and substantial contamination in drinking water sources of northern Tanzania. Fluoride occurrence in groundwater systems of the study area is space dependent. Six significant spatial patterns were identified using local Moran's I index. Two of the spatial patterns consisted of water sources with high fluoride concentrations occurring in the same neighbourhood (High-high cluster). We technically termed this cluster as regional hotspot. Ninety six percent (96%) of water sources in the hotspots had fluoride concentrations above 1.5 mg/L The other four spatial patterns consisted of water sources with low fluoride concentrations occurring in the same neighbourhood (Low-low cluster). This was technically termed as regional cool spot. All water sources in these clusters had fluoride concentrations <1.5 mg/L.

The petrographic study of rock samples from the north-east hotspot and east cool spot suggest that fluoride in groundwater system originates from volcanic igneous rocks of intermediate chemical composition between mafic and felsic rocks. The dominant mineral

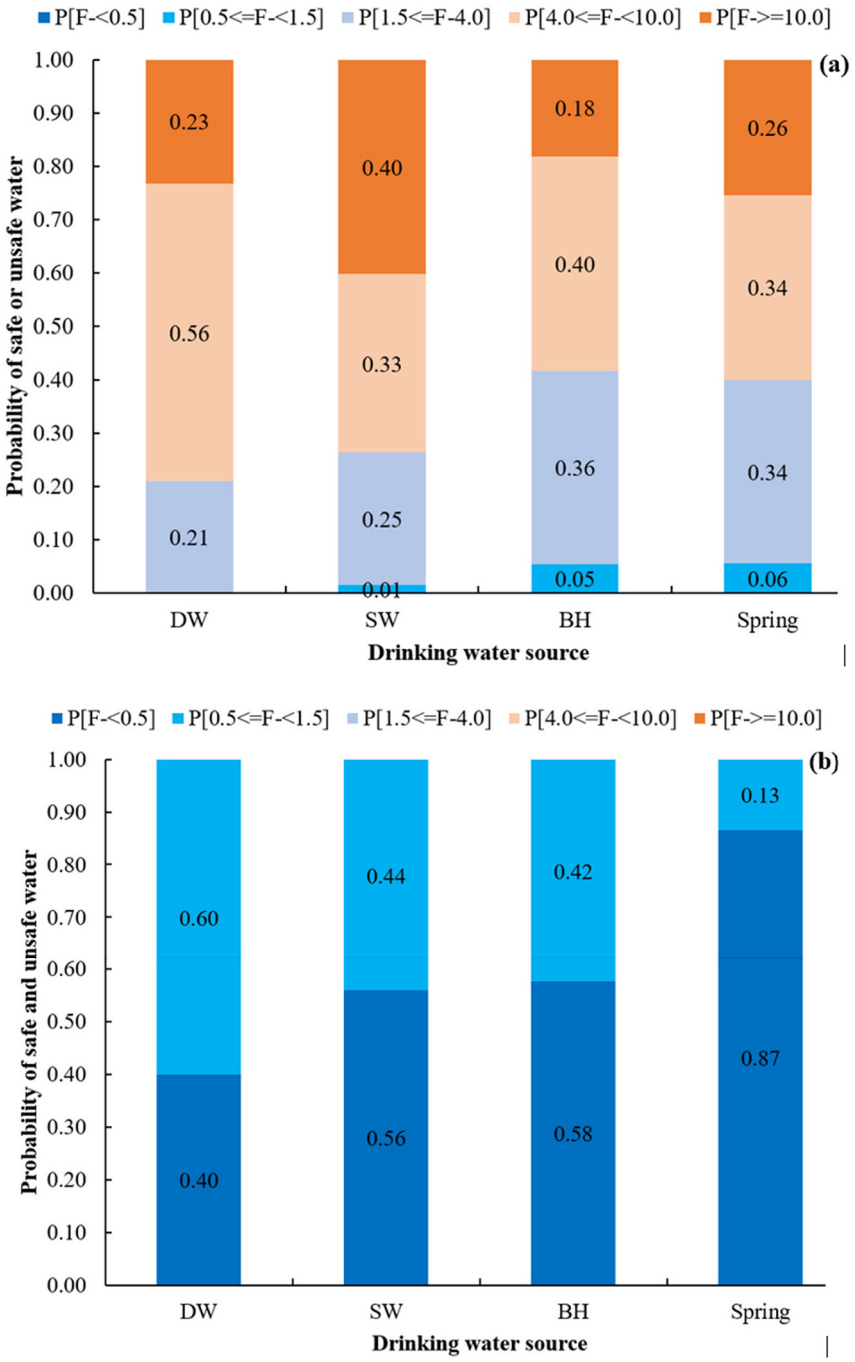


Fig. 11. Probability of accessing safe drinking water source through dug wells (DW), shallow wells (SW), boreholes (BH) and springs in regional fluoride (a) hotspots, and (b) cool spots.

composition include pyroxene, amphibole, plagioclase, feldspaphoids, biotite, titanite and hornblende with major fluoride bearing phases being titanite, amphibole, hornblende and biotite. Weathering of the rocks and dissolution of fluoride bearing minerals may be the main source of high levels of fluoride in drinking water sources.

The geographical distribution of regional cool spots around elevated areas suggests that variation in fluoride concentrations depends on water-rock contact time which is controlled by local topography. The presence of water sources with low or high fluoride concentrations in the neighbourhood of high or low concentrations suggest availability of two groundwater sources, that is, groundwater from precipitation and that from the rift system (geothermal waters).

Despite the East African Rift Valley region is regarded as one of global fluoride belts, the occurrence of fluoride in this region is space dependent. There are safe sources of drinking water in some areas while it is difficult to find safe sources in others. In this study we demonstrate the use of global and local Moran's I statistical methods to identify significant regional fluoride hotspots and cool spots. The identified hotspots and cool spots are crucial to the water supply authority when planning for new drinking water source development or managing existing drinking water sources.

CRediT authorship contribution statement

Julian Ijumulana: Conceptualization, Methodology, Formal analysis, Writing - original draft.**Fanuel Ligate:**Conceptualization, Formal

analysis, Writing - review & editing.**Prosun Bhattacharya:**Conceptualization, Methodology, Writing - review & editing, Project administration. **Felix Mtalo:**Supervision, Project administration.**Chaosheng Zhang:** Methodology, Writing - review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgement

The authors would like to extend their great appreciation to Director of Water Quality Division of the Ministry of Water and Irrigation of Tanzania for providing primary data on fluoride concentrations in drinking water sources from the National Fluoride Belt (NFB) in northern Tanzania. Further, the authors extend their heartfelt thanks to Swedish International Development Cooperation Agency (Sida) for funding the project, DAFWAT (Project 2235) through Sida Contribution 51170072.

References

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 (4), 486–503.
Ali, S., Shekhar, S., Bhattacharya, P., Verma, G., Chandrasekhar, T., Chandrashekhar, A.K., 2018. Elevated fluoride in groundwater of Siwani Block, Western Haryana, India: a potential concern for sustainable water supplies for drinking and irrigation. Groundw. Sustain. Dev. 7, 410–420. https://doi.org/10.1016/j.gsd.2018.05.008.
Ali, S., Fakhri, Y., Golbini, M., Thakur, S.K., Alinejad, A., Parseh, I., Shekhar, S., Bhattacharya, P., 2019. Concentration of fluoride in groundwater of India: a systematic review, metaanalysis and risk assessment. Groundw. Sustain. Dev. 9, 100224. https://doi.org/10.1016/j.gsd.2019.100224.
Amini, M., Mueller, K., Abbaspour, K.C., Rosenberg, T., Afyuni, M., Møller, K.N., Johnson, C.A., 2008. Statistical modeling of global geogenic fluoride contamination in groundwaters. Environmental Science & Technology 42 (10), 3662–3668.
Anselin, L., 1993. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. Regional Research Institute, West Virginia University, Morgantown, WV.
Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27 (2), 93–115.
Anselin, L., 1996. The Moran scatterplot as an ESDA tool to assess local instability in spatial. Spatial Analytical 4, 111.
Baudouin, C., Parat, F., Denis, C.M., Mangasini, F., 2016. Nephelinite lavas at early stage of rift initiation (Hanang volcano, North Tanzanian Divergence). Contributions to Mineralogy 171 (7), 64.
Berry, B.J.L., Marble, D.F., 1968. Spatial Analysis: A Reader in Statistical Geography. Prentice-Hall.
Bhattacharya, P., Bundschuh, J., 2015. Groundwater for sustainable development - cross cutting the UN sustainable development goals. Groundw. Sustain. Dev. 1 (1-2), 155–157. https://doi.org/10.1016/j.gsd.2016.04.004.
Box, G.E.P., Cox, D.R., 1964. An analysis of transformations (with discussion). J. R. Stat. Soc. Ser. B 26, 211–243.
Brindha, K., Elango, L., 2011. Fluoride in groundwater: causes, implications and mitigation measures. Fluoride properties, applications and environmental management. 1, pp. 111–136.
Brown, M.B., Forsythe, A.B., 1974. Robust tests for the equality of variances. J. Am. Stat. Assoc. 69 (346), 364–367. https://doi.org/10.1080/01621459.1974.10482955.
Brunt, R., Vasak, L., Griffioen, J., 2004. Fluoride in Groundwater: Probability of Occurrence of Excessive Concentration on Global Scale. IGRAC.
Bundschuh, J., Maity, J.P., Mushtaq, S., Vithanage, M., Seneweera, S., Schneider, J., Bhattacharya, P., Khan, N.I., Hamawand, I., Guilherme, L.R.G., Reardon-Smith, K., Parvez, F., Morales-Simfors, N., Ghaze, S., Pudmenzky, C., Kouadio, L., Chen, C.-Y., 2017. Medical geology in the framework of the sustainable development goals. Sci. Total Environ. 581–582, 87–104. https://doi.org/10.1016/j.scitotenv.2016.11.208.
Chacha, N., Njau, K.N., Lugomela, G.V., Muzuka, A.N., 2018. Hydrogeochemical characteristics and spatial distribution of groundwater quality in Arusha well fields, Northern Tanzania. Appl Water Sci 8 (4), 118.
Chandrajith, R., Padmasiri, J., Dissanayake, C., Prematilaka, K., 2012. Spatial distribution of fluoride in groundwater of Sri Lanka. Journal of the National Science Foundation of Sri Lanka 40 (4).
Chandrajith, R., Diyabalanage, S., Dissanayake, C.B., 2020. Geogenic fluoride and arsenic in groundwater of Sri Lanka and its implications to community health. Groundw. Sustain. Dev., 100359 https://doi.org/10.1016/j.gsd.2020.100359.
Chowdhury, A., Adak, M.K., Mukherjee, A., Dhak, P., Khatun, J., Dhak, D.J.J.o.H., 2019. A Critical Review on Geochemical and Geological Aspects of Fluoride Belts, Fluorosis and Natural Materials and Other Sources for Alternatives to Fluoride Exposure.
Cliff, D., Ord, J., 1973. Spatial Autocorrelation. Pion Limited, London.
Cliff, D., Ord, J., 1981. Spatial processes. Cliff Spatial Processes. Pion, London.
Dissanayake, C.B., 1991. The fluoride problem in the ground water of Sri Lanka — environmental management and health. Int. J. Environ. Stud. 38 (2–3), 137–155. https://doi.org/10.1080/00207239108710658.
Edmunds, W.M., Smedley, P.L., 2013. Fluoride in natural waters. Essentials of Medical Geology. Springer, pp. 311–336.
Fawell, J.K., Bailey, K., 2006. Fluoride in drinking-water. World Health Organization.
Fawell, J., Nieuwenhuijsen, M.J., 2003. Contaminants in drinking waterEnvironmental pollution and health. Br. Med. Bull. 68 (1), 199–208.
Frencken, J., 1992. Endemic Fluorosis in Developing Countries: Causes, Effects and Possible Solutions. TNO Institute for Preventive Health Care.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (3), 189–206.
Getis, A., Ord, J.K., 1996. Local spatial statistics: an overview. Spatial Analysis: Modelling in a GIS Environment. 374, pp. 261–277.
Ghiglieri, G., Balia, R., Oggiano, G., Pittalis, D., 2010. Prospecting for safe (low fluoride) groundwater in the Eastern African Rift: the Arumeru District (Northern Tanzania). Hydrol. Earth Syst. Sci. 14 (6), 1081.
GST, 2018. Lithological Data of Tanzania.
Gumbo, F., Mkongo, G., 1995. Water defluoridation for rural fluoride affected communities in Tanzania. First International Workshop on Fluorosis Prevention and Defluoridation of Water, Int. Soc. Fluoride Res.
Haining, R., Wise, S., Ma, J., 1998. Exploratory spatial data analysis. Journal of the Royal Statistical Society: Series D (The Statistician) 47 (3), 457–469.
Hong, F., Cao, Y., Yang, D., Wangb, H., 2001. Research on the effects of fluoride on child intellectual development under different environmental conditions. Chinese Primary Health Care 15 (3), 56–57.
Ishioka, F., Krihara, K., Suito, H., Horikawa, Y., Ono, Y., 2007. Detection of hotspots for three-dimensional spatial data and its application to environmental pollution data. Journal of Environmental Science for Sustainable Society 1, 15–24.
Jacks, G., Bhattacharya, P., Chaudhary, V., Singh, K.P., 2005. Controls on the genesis of some high-fluoride groundwaters in India. Appl. Geochem. 20 (2), 221–228. https://doi.org/10.1016/j.apgeochem.2004.07.002.
Jeihanipour, A., Shen, J., Abbt-Braun, G., Huber, S.A., Schäfer, A.I., 2018. Seasonal variation of organic matter characteristics and fluoride concentration in the Maji ya Chai River (Tanzania): Impact on treatability by nanofiltration/reverse osmosis. Sci. Total Environ. 637–638, 1209–1220. https://doi.org/10.1016/j.scitotenv.2018.05.113.
Kashyap, C.A., Ghosh, A., Singh, S., Ali, S., Singh, H.K., Chandrashekhar, H., Chandrasekharam, D., 2020. Distribution, genesis and geochemical modeling of fluoride in the water of tribal area of Bijapur district, Chhattisgarh, central India. Groundw. Sustain. Dev. 11, 100403. https://doi.org/10.1016/j.gsd.2020.100403.
Kijazi, A.L., Reason, C., 2009. Analysis of the 2006 floods over northern Tanzania. International Journal of Climatology: A Journal of the Royal Meteorological Society 29 (7), 955–970.
Kimambo, V., Bhattacharya, P., Mtalo, F., Mtamba, J., Ahmad, A., 2019. Fluoride occurrence in groundwater systems at global scale and status of defluoridation–state of the art. Groundwater for Sustainable Development, Groundwater for Sustainable Development 9, 100223.
Kulldorff, M., 1997. A spatial scan statistic. Communications in Statistics - Theory and Methods 26 (6), 1481–1496. https://doi.org/10.1080/03610929708831995.
Kut, K.M.K., Sarswat, A., Srivastava, A., Pittman Jr., C.U., Mohan, D., 2016. A review of fluoride in African groundwater and local remediation methods. Groundw. Sustain. Dev. 2, 190–212.
Lathman, M., Grech, P., 1967. The effects of excessive fluoride intake. American Journal of Public Health and the Nations Health 57 (4), 651–660.
Lu, Y., Sun, Z., Wu, L., Wang, X., Lu, W., Liu, S., 2000. Effect of high-fluoride water on intelligence in children. Fluoride 33 (2), 74–78.
MacDonald, A., Davies, J. (2000). A brief review of groundwater for rural water supply in sub-Saharan Africa. (Retrieved from).
McMahon, P.B., Brown, C.J., Johnson, T.D., Belitz, K., Lindsey, B.D., 2020. Fluoride occurrence in United States groundwater Sci. Total Environ. 732, 139217. https://doi.org/10.1016/j.scitotenv.2020.139217.
Malago, J., Makoba, E., Muzuka, A.N., 2017. Fluoride levels in surface and groundwater in Africa: a review. American Journal of Water Science and Engineering 3 (1), 1–17.
Mohanta, A., Mohanty, P.K., 2018. Dental fluorosis-revisited. Biomedical Journal of Scientific & Technological Research 2 (1). https://doi.org/10.26717/BJSTR.2018.02.000667.
MoIW, 2018. Major Drainage Basin Boundaries and Water Bodies of Tanzania.
Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1/2), 17–23. https://doi.org/10.2307/2332142.
Nanyaro, J.T., Aswathanarayana, U., Mungure, J.S., Lahermo, P.W., 1984. A geochemical model for the abnormal fluoride concentrations in waters in parts of northern Tanzania. Journal of African Earth Sciences (1983) 2 (2), 129–140. https://doi.org/10.1016/S0731-7247(84)80007-5.
Nas, B., 2009. Geostatistical approach to assessment of spatial distribution of groundwater quality. Pol. J. Environ. Stud. 18 (6).
NBS, 2018. Tanzania in Figures 2017. National Bureau of Statistics, Dar es Salaam, The United Republic of Tanzania 105p. https://www.nbs.go.tz/index.php/en/tanzania-in-figures/433-tanzania-in-figures-2017.
Ping, J.L., Green, C.J., Zartman, R.E., Bronson, K.F., 2004. Exploring spatial dependence of cotton yield using global and local autocorrelation statistics. Field Crop Res. 89 (2), 219–236. https://doi.org/10.1016/j.fcr.2004.02.009.
Quino Lima, I., Ramos Ramos, O.E., Ormachea Muñoz, M., Quintanilla Aguirre, J., Duwig, C., Maity, J.P., Sracek, O., Bhattacharya, P., 2020. Spatial dependency of arsenic, antimony, boron and other trace elements in the shallow groundwater systems of the Lower Katari Basin, Bolivian Altiplano. Sci. Total Environ. 719, 137505. https://doi.org/10.1016/j.scitotenv.2020.137505.

Robins, N., Davies, J., Farr, J., Calow, R.J.H.J., 2006. The changing role of hydrogeology in semi-arid southern and eastern Africa. 14 (8), 1483–1492.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52 (3/4), 591–611. https://doi.org/10.2307/2333709.
Shivaprakash, P., Ohri, K., Noorani, H., 2011. Relation between dental fluorosis and intelligence quotient in school children of Bagalkot district. Journal of Indian Society of Pedodontics and Preventive Dentistry 29 (2), 117.
Shrivastava, B.K., Vani, A., 2009. Comparative study of defluoridation technologies in India. Asian J. Exp. Sci. 23 (1), 269–274.
Smedley, P., Nkotagu, H., Pelig-Ba, K., MacDonald, A., Tyler-Whittle, R., Whitehead, E., Kinniburgh, D., 2002. Fluoride in groundwater from high-fluoride areas of Ghana and Tanzania. British Geological Survey Commissioned Report, CR/02/316 72 pp. http://resources.bgs.ac.uk/sadcreports/tanzania2002smedleyfluoridecr02316n.pdf.
Symanzik, J., 2014. Exploratory spatial data analysis. In: Fischer, M.M., Nijkamp, P. (Eds.), Handbook of Regional Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1295–1310.
Tango, T., 1995. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. Stat. Med. 14 (21—22), 2323–2334.
Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.-F., Holman, I., Treidel, H., 2013. Ground water and climate change. Nat. Clim. Chang. 3 (4), 322–329.
Thole, B., 2013. Ground water contamination with fluoride and potential fluoride removal technologies for East and Southern Africa. In: Dar, I.A., Dar, M.A. (Eds.), Perspectives in Water Pollution. InTech Open. https://doi.org/10.5772/50060 ISBN: 978-953-51-6332-9 (eBook (PDF).
Thole, B., 2013. Ground water contamination with fluoride and potential fluoride removal technologies for East and Southern Africa. Perspectives in Water Pollution. InTech.
Vithanage, M., Bhattacharya, P., 2015. Fluoride in the environment: sources, distribution and defluoridation. Environ. Chem. Lett. 13 (2), 131–147. https://doi.org/10.1007/s10311-015-0496-4.
Vuhahula, E.A.M., Masalu, J.R.P., Mabelya, L., Wandwi, W.B.C., 2009. Dental fluorosis in Tanzania Great Rift Valley in relation to fluoride levels in water and in 'Magadi' (Trona). Desalination 248 (1), 610–615. https://doi.org/10.1016/j.desal.2008.05.109.
WHO, 2011. Guidelines for drinking-water quality. Fourth Edition. WHO chronicle 38 (4), 104–108.
Yamada, I., 2016. Thiessen polygons. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology, pp. 1–6.
Zhang, C., McGrath, D., 2004. Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods. Geoderma 119 (3), 261–275.
Zhang, C., Luo, L., Xu, W., Ledwith, V., 2008. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. Sci. Total Environ. 398 (1), 212–221. https://doi.org/10.1016/j.scitotenv.2008.03.011.