



Source characterization and human health risk assessment of nitrate in groundwater of middle Gangetic Plain, India

Deepak Kumar¹ · Anshuman Singh¹ · Rishi Kumar Jha² · Bibhuti Bhushan Sahoo¹ · Sunil Kumar Sahoo³ · Vivekanand Jha³

Received: 18 July 2018 / Accepted: 9 May 2019 / Published online: 21 May 2019
© Saudi Society for Geosciences 2019

Abstract

Nitrate (NO_3^-) pollution is a global concern as it affects the whole ecosystem: human, livestock, economy, and environment. The elevated levels of nitrate in groundwater can directly pose risks to population. A total of 156 representative groundwater samples were collected from groundwater sources such as hand pumps and bore wells across the study area. To identify the source of nitrate with its associated attributes, multivariate statistical methods (factor analysis (FA), sparse principal component analysis (SPCA)) were used in this study. In addition, empirical Bayesian kriging (EBK) modeling was used to predict the nitrate at ungauged locations of the study area. From the analysis of results, it was found that 5% of the groundwater samples exceeded the acceptable limit (50 mg l^{-1}) of nitrate as specified by the World Health Organization (WHO). The first principal component (PC) indicated by the SPCA was salinity factor, which was significantly contributed by electrical conductivity followed by sulfate. The fourth PC represented the nitrate as a factor and positive loading of nitrate was strongly associated with chloride, sulfate, and calcium. The associated loading of nitrate with water quality attributes indicated that elevated level of nitrate in groundwater may be due to external sources that came through anthropogenic activity. A similar conclusion was drawn from factor analysis as well, indicating that SPCA can be applied as a new method for groundwater geochemistry. Hazard index calculations showed that infants of the study region were at a higher risk compared to the adults and children.

Keywords Groundwater · Nitrate · Health risk · Groundwater quality · Groundwater pollution · Gaya

Introduction

Due to high crop yield, middle Gangetic plain is considered an important agricultural region of India. The agricultural activities in the area leaves behind contaminant residues in the environment. Such residues are leached out during precipitation and irrigation, resulting in changes of physiochemical

environmental characteristics. Groundwater is one of the most important water resources for agricultural, domestic, and industrial activities of the region. Nitrate contamination in groundwater in such scenario becomes a major concern across many similar regions of the world (Kim et al. 2015; Su et al. 2013; Zhang et al. 2018). In rural agricultural areas where groundwater is shallow and primarily used for domestic water supplies, this becomes an important consideration (Gu et al. 2013; He and Wu 2018; He et al. 2018; Zhang et al. 2013). The nitrate (NO_3^-) contamination in groundwater is mainly governed by livestock activity, fertilizer application to farmland, and wastewater disposal including septic tanks (Chesnaux and Allen 2008; Gu et al. 2013; Kaown et al. 2009; WHO 2008; Zhang et al. 2013). Three forms of inorganic nitrogen (NO_3^- , NO_2^- , and NH_4^+) exist in the soil and NO_3^- and NH_4^+ are freely available form for the plants. In groundwater system, the concentrations of NH_4^+ and NO_2^- are very low, as they are easily converted to NO_3^- under oxidizing conditions. The specific physicochemical properties (e.g., high solubility rate, movability, and more stable

Editorial handling: Haroun Chenchouni

✉ Deepak Kumar
decage007@gmail.com

¹ Department of Civil Engineering, National Institute of Technology, Patna, Bihar 800005, India

² Department of Mathematics, National Institute of Technology, Patna, Bihar 800005, India

³ Environmental Assessment Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400085, India

oxidation state) categorize nitrate as a potentially contaminant (Arabgol et al. 2016; WHO 2008; Zhai et al. 2017a).

From the viewpoint of human health risk assessment, the exposure to higher nitrate concentration via ingestion can cause adverse effects on human health (Wu and Sun 2016). Methemoglobinemia is a predominant health effect for bottle-fed infants as a result of high-nitrate-contaminated water. Methemoglobinemia is a disease in which the level of oxygen depletes in blood cells as a result of nitrite (after reduction of nitrate to nitrite in reducing condition) reaction with hemoglobin. The higher levels of methemoglobin (> 10%) can result in cyanosis (blue-baby syndrome) (Gao et al. 2012). Methemoglobinemia effects can also be seen as gastrointestinal infections and combatting dehydration (Almasri 2007). Apart from these health effects due to elevated nitrate levels in water, researchers have also reported nitrosamines (Elisante and Muzuka 2016), non-Hodgkin lymphoma (NHL) (Chang et al. 2009), and multiple sclerosis (Fabro et al. 2015). The World Health Organization guidelines (WHO 2011) recommends 50 mg l^{-1} as acceptable level of NO_3^- concentration in drinking water, whereas the Bureau of Indian Standards (BIS) recommends maximum acceptable limit of 45 mg l^{-1} .

Source identification by adopting various statistical techniques has gained popularity in recent times. Source identification essentially correlates geological and climate attributes of environment like groundwater depth, direction of water flow, and land use activities (Behera and Das 2018; Chen et al. 2005; Gu et al. 2013) to groundwater quality parameters. Recently, Liao et al. (2016) have used hierarchical cluster analysis as a source identification technique for nitrate investigation in groundwater and illustrated that agricultural fertilizers are the main sources of nitrate pollution in shallow groundwater. Kim and Park (2016) and Zhai et al. (2017b) have used principal component analysis for source identification of nitrate contamination in groundwater for agricultural area. Furthermore, Wang et al. (2018) have used popular statistical tool of factor analysis (FA) to investigate the source of nitrate in groundwater pollution. This study applied an advanced statistical tool called sparse principal component analysis (SPCA) to investigate the source for nitrate pollution in groundwater and compare with factor analysis to check the adaptability of the method in the field of water quality assessment. In addition, empirical bayesian kriging (EBK) interpolation model was also adopted to find out the concentrations at the unsampled locations in the study area.

Factor analysis is a dimension reduction technique in multivariate analysis which can effectively reduce numerous variables into a few factors (Wu et al. 2014). It is used to analyze the interrelationship among a set of objects (Cao et al. 2016). Factor analysis reduces large numbers of variables into linear function (as factors) based on variance explained by the each attribute (Davis 1973). The variance contributed in each factor by individual variable is called loading. Factor analysis helps

in extracting unobservable (latent) underlying process which cannot be seen in raw dataset. In this study, an attempt has been made to extract the latent factor from hydrogeochemical data and to classify the original data using factor analysis. Factor analysis can find hidden patterns, extract patterns overlap, and show characteristics of input features in terms of loading. The different orthogonal rotation gives a different loading of individual input variables. The orthogonal rotations give a clear pattern of each individual loading based on maximization of the variance on the first extracted principal axes in terms of first factor (Adams et al. 2001; Aiuppa et al. 2003; Behera and Das 2018; Jayakumar and Siraz 1997). The second factor explains the second maximum residual variability after orthogonal rotation. In this research, factor analysis is used to extract the pollution factors and their involving attributes of geochemical data. The interested reader can find more details about factor analysis in Jolliffe (1986).

Sparse principal component analysis (SPCA) has emerged as a powerful statistical technique in multivariate analysis. It provides simple interpretable modes with localized spatial support. SPCA is an extension of the principal component analysis. The basic framework of this method is parsimonious decomposition. The parsimonious decomposition uses prior information in the form of sparsity to regularize the loading of each input's variables, which means each weight vector has only a few "active" loadings and at the same time other loadings by other variables are constrained to zero. The generated sparse principal components (PCs) are sparsely weighted linear combination of the input variables. The sparse coefficient of input variable is calculated through L1 penalty. The term penalty is a regularization term which is used to avoid the overfitting of the model (generally used in Lasso regression). The term L1 limits the size of coefficient to yield sparsity (i.e., model with few coefficient). In SPCA loading is depending upon the L1 constraint. This L1 constraint maximizes the variance for the input variables on PC axis. In this study, SPCA is used to extract the pollution factors and involved components which control the process. A brief review of this concept can be found in Zou et al. (2006).

To access the geographical distribution of any features requires spatial geochemical data. Collecting sample data from every location is an expensive, time-consuming, and exhaustive process. Therefore, to overcome such problems and analyze geographical distribution of any features, different geostatistical models are available that can predict data values as intermediate locations with certain uncertainty. The basic concept behind any interpolation methods is that it requires distance function between collected samples and unknown location. Classical interpolation methods use a single variogram model to predict the values at unknown location. EBK is a powerful approach, which uses several-variograms model instead of single-variogram model (Gribov and Krivoruchko 2012). EBK works on autocorrelation to produce prediction surface with minimum

uncertainty based on lowest mean square error. Furthermore, this algorithm calculates the local trends when data coverage is good. For poor data coverage, it allows bending using prior background mean (Krivoruchko 2012; Krivoruchko and Butler 2013) to avoid the distortion to derived semi-variogram. Here, EBK statistical interpolation method was used to interpolate concentration of nitrate over others' unsampled location to generate a raster map and to find the elevated level of nitrates in the study region.

The objective of this study was to investigate the source of nitrate using factor analysis and compare the same with sparse principal component analysis. EBK modeling was used to investigate the distribution of nitrate pollution in the study area. USEPA assessment method was adopted to assess the human health risk due to exposure to nitrate-contaminated groundwater. The results of this investigation are expected to provide accurate results and evidence supporting adoption of the abovementioned multivariate statistical tools for hydro-geochemical data and improve the future research.

Study area and regional hydrogeology

The study was conducted in Gaya district of Bihar (Fig. 1) which falls under Punpun sub-basin of the Ganga basin. Bihar state is situated in the Middle Gangetic Plain of India. The Gaya district spreads over an area of 4976 km², which lies between north latitudes 24° 30': 25° 06' and east longitudes

84° 24': 85° 30'. The drainage pattern of this study area is primarily controlled by four parallel streams, the Morhar, the Phalgu, the Paimar, and the Dhadhar, all stemming from the southern plateau of Jharkhand to north and northeasterly (CGWB 2013). The whole study area experiences a continental monsoon with extreme climate. The summer months are characterized by hot blasts (46 °C) of westerly winds (known as "loo") to mercury drops down to as low as 4 °C. The monsoon sets at the end of June which receives annual rainfall between 568.5 mm and 1109 mm (CGWB 2013).

The soil type of this region consists of younger alluvial, sandy, red, and yellow soil. The sandy, red, and yellow soil is restricted to the southern and northeastern parts of the area which consists of less amounts of nitrogen and organic matter. The younger alluvial soil is restricted to mid-north and north-west parts which have loam with a small proportion of sand and clay (locally called "Kewal"). The younger alluvial soil is rich in nitrogen and calcium. The groundwater aquifers of this region was categorized between unconfined and semi-confined. The movement of groundwater in such aquifer is primarily controlled by joints, fissures, and other planes of weak structure like granite gneisses, mica-schist, quartzite, and their other associated rocks of Pre-Cambrian age (CGWB 2013). The exploratory drilling results show that there is good groundwater development prospect up to 100 m below the ground level. The semi-unconfined aquifer consists sandy horizon which gradually increase towards eastwards (CGWB 2013). The groundwater depth level varies between 5 and 10

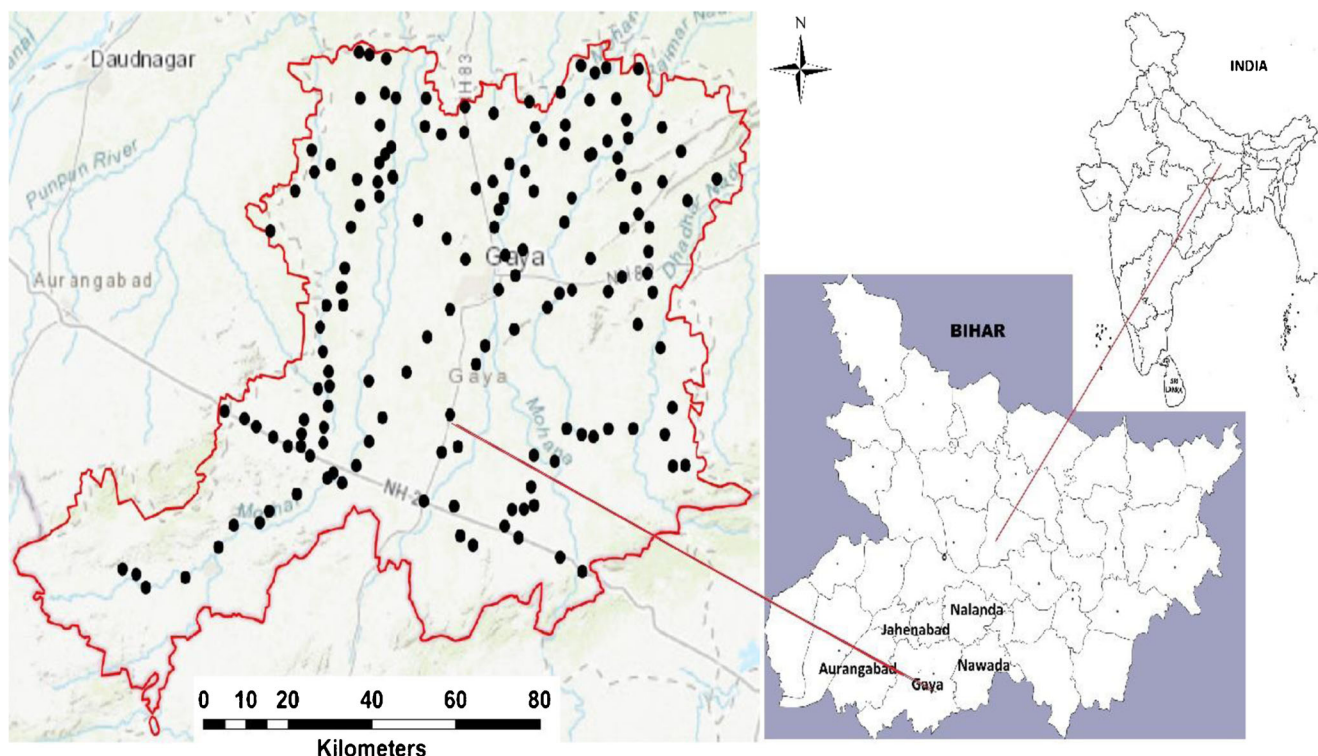


Fig. 1 Study area with sampling location

Table 1 Health risk parameter value

	IR (L d ⁻¹)	ED (a)	A _{SD} (cm ²)	BW (kg)	FE (t day ⁻¹)	AT (day)	ET (h)	<i>f</i>	k (cm h ⁻¹)
Infants	0.65	0.5	3416	6.94	1	182.5	0.29	0.5	0.001
Children	1.5	6	9035.2	25.9	1	2190	0.29	0.5	0.001
Adults	2	30	16600	57.3	1	10950	0.29	0.5	0.001

m below ground level in the major portion of the district during pre-monsoon period, whereas the groundwater levels (> 10 m) are found in the eastern and western parts of the district (CGWB 2013).

Methods and materials

Water sampling and analysis

A total of 156 representative groundwater samples were collected from underground sources (mainly bore wells and hand pumps) across the study area (Fig. 1). All the underground sources (bore wells/hand pumps) were freely accessible in the community for the consumption. Sampling site and geo-positions (latitude and longitude) were found using a Global Positioning System (GPS) (model: Garmin GPS 72H). Groundwater samples were collected after 2–5 min of flushing time. For each location, 1-l high-density polyethylene (Thermo Scientific, 1131000BPC) sample bottles pre-treated with dilute nitric acid overnight and rinsed with ultrapure water (Milli-Q Ultrapure Water Purification System, Model Z00QSVC01) were used. Sample bottles were washed thoroughly with the sampled bore well/hand-pump water before collecting the water samples. Samples were collected during the pre-monsoon season from marked locations (Fig. 1). Immediately after collection of water samples at the site, in situ water quality parameters such as pH and conductivity were estimated in the field. Collected samples were filtered

using 0.45- μ m membrane filter (syringe filters). To avoid wall deposition, a portion of the samples was acidified after filtration. To avoid the matrix decomposition, the unacidified samples were stored at 4 °C in the dark until completion of analysis (APHA 2005).

Before analysis, water samples were filtered using 0.45- μ m cellulose filters (syringe filters) in spectrophotometer equipment (Thermo Scientific Evolution 201). The reagents and redistilled water used for the solution preparation (or dilution) were, respectively, analytical grade (Merk Millipore) and highest purity Millipore water. The in situ parameters such as pH and electrical conductivity (EC) were measured immediately after sampling using Thermo Scientific Orion VERSA STAR pH/ISE/Conductivity/RDO/Dissolved Oxygen Multiparameter Meter Kit. Three-point calibration method, using pH buffer (4, 7, 10) and two-point calibration of potassium chloride strength solutions (1413 μ S, 0.1 mS) were used to calibrate the instrument. The instrument accuracy for pH and EC meter was ± 0.002 pH and 0.5% of reading ± 1 digit > 3 μ S. The selected anions, i.e., chloride, sulfate, and phosphate, were determined using standard AgNO₃ titration method, turbidimetric method, and colorimetric method respectively using UV spectrophotometer (Adimalla et al. 2018b). The nitrate (NO₃⁻) concentration in water samples was estimated by a method adopted by Ganesh et al. (2012). The standard EDTA titration method was used for estimating calcium and magnesium (Adimalla et al. 2018b). All procedures followed for analysis were as per standard methods for the examination of water and wastewater (APHA 2005).

Table 2 Descriptive statistics of groundwater samples across middle Gangetic plains of Gaya district

Parameters	NO ₃ ⁻ (mg/l)	pH	EC (μ S/cm)	Cl ⁻ (mg/l)	SO ₄ ²⁻ (mg/L)	PO ₄ ³⁻ (mg/l)	Calcium (mg/l)	Magnesium (mg/l)	Bicarbonate (mg/l)
Minimum	0.7	5.8	311.4	14.4	0.2	0.1	28	2	91.7
Maximum	54.9	7.6	3891	514	316.3	1.4	496	1127	787.4
1st quartile	8.2	6.9	555.8	36	7.8	0.2	137.8	45	288.6
Median	14.3	7	713.3	52.8	18.1	0.4	181	78	347.8
3rd quartile	30.1	7.2	996.8	88.9	47.3	0.5	237.3	119	412.8
Mean	19.9	7	872.5	84	36.8	0.4	192.6	103.3	362.3
Standard deviation	15.1	0.3	529.52	83.9	46	0.2	88.1	116.3	122.8
Skewness	0.8	-1.6	2.4	2.5	2.7	1.3	1	5.1	0.7
Kurtosis	-0.6	5.4	8.5	6.6	9.9	3	1.4	37.8	1.1
Geometric mean	14.2	7	738.3	60.8	18.2	0.3	173.1	69.5	340.8

Table 3 Total variance explained based on selected factors

	F1	F2	F3	F4
SS loadings	2.89	2.2	1.76	1.56
Proportion Var	0.24	0.18	0.15	0.13
Cumulative Var	0.24	0.42	0.57	0.7
Proportion explained	0.34	0.26	0.21	0.19
Cumulative proportion	0.34	0.61	0.81	1

Human health risk assessment

The USEPA health risk assessment method was used to assess the human health of three age groups (infants, children, and adults) of the study area, who are exposed to nitrate-contaminated groundwater. According to the International Agency for Research on Cancer (IARC), nitrate (NO₃⁻) is categorized under group D carcinogenic pollutants. Therefore, only non-carcinogenic risks were calculated in this study. The health risk due to exposure of nitrate-contaminated water was evaluated via two pathways (oral and dermal). The oral and dermal exposure is estimated using Eqs. (1), (3), and (4). In addition, the hazard index (HI) which is the sum of THQ_{ingestion} + THQ_{dermal} is also estimated considering Eqs. (2) and (5) (Li et al. 2017; Li et al. 2016).

$$CDI_{oral} = c_w \times IR \times EF \times ED / (BW \times AT) \quad (1)$$

$$HQ_{oral} = CDI_{oral} / RfD_{oral} \quad (2)$$

$$CDI_{dermal} = I \times A_{SD} \times EF \times FE \times ED / (BW \times AT \times f) \quad (3)$$

$$I = 10^{-3} \times k \times c_w \times ET \quad (4)$$

$$HQ_{dermal} = CDI_{dermal} / RfD_{dermal} \quad (5)$$

where CDI_{oral} and CDI_{dermal}, respectively, represents the daily average exposure dosage through the oral and dermal

Table 4 Factor loading matrix of four factors of groundwater dataset

	F1	F2	F3	F4
Nitrate	-0.04	-0.01	0.07	0.37
pH	-0.13	0.42	-0.22	-0.18
EC	1	0.01	-0.03	-0.03
Chloride	0.49	0.1	0.12	0.30
Sulfate	0.65	-0.02	0.14	0.17
Phosphate	0.21	-0.23	0.01	-0.19
Calcium	0.07	0.04	-0.09	0.91
Magnesium	0.03	0.03	1	-0.18
Bicarbonate	0.01	0.98	0.01	0

Table 5 Factor correlation matrix

	MR1	MR2	MR3	MR4
MR1	1	0.52	0.35	0.42
MR2	0.52	1	0.3	0.22
MR3	0.35	0.3	1	0.19
MR4	0.42	0.22	0.19	1

pathways in per-unit weight (mg l⁻¹ day⁻¹), C_w is the concentration of the nitrate in the groundwater, IR indicates intake rate of groundwater for drinking (l day⁻¹). EF and ED are the exposure frequency (day a⁻¹) and exposure duration (a), respectively; BW is the body weight of a person (kg), AT is the average time for non-carcinogenic effect (day), FE is the daily exposure frequency of dermal contact event (t day⁻¹), A_{SD} is the body surface area (cm²), I indicates the exposure dosage of every single event (mg cm²), k is the coefficient of permeability of the skin (cm h⁻¹), ET is the contact time for a single shower (h day⁻¹), and the reference dose for NO₃⁻ is RfD_{oral} and RfD_{dermal} are 1.6 (mg kg⁻¹ day⁻¹) and 0.8 (mg kg⁻¹ day⁻¹) respectively. The input values used to calculate the health risk for three age groups are presented Table 1.

The hazard index (HI) of pollutant, if less than 1, means that risk is within the recommended limit, while HI > 1 indicates that the risk is beyond the acceptable level.

Results and Discussion

General groundwater chemistry

Table 2 shows the descriptive statistics of groundwater quality in the study area. The pH value ranged from 5.8 to 7.6, which clearly indicated that the groundwater of the study region is slightly acidic in nature. The EC is the basic water quality parameter which represents the total dissolved ions in the groundwater samples. The EC values ranged from 311.4 to

Table 6 SPCA loading matrix of four component in the groundwater dataset

	PC1	PC2	PC3	PC4
Nitrate	0.00	0.00	0.00	0.49
pH	0.00	-0.24	-0.01	0.00
EC	0.61	0.00	0.00	0.00
Chloride	0.00	0.00	0.00	0.27
Sulfate	0.27	0.00	0.00	0.12
Phosphate	0.00	0.05	0.00	0.00
Calcium	0.00	0.00	0.00	0.82
Magnesium	0.00	0.00	0.78	0.00
Bicarbonate	0.00	0.00	0.00	0.00

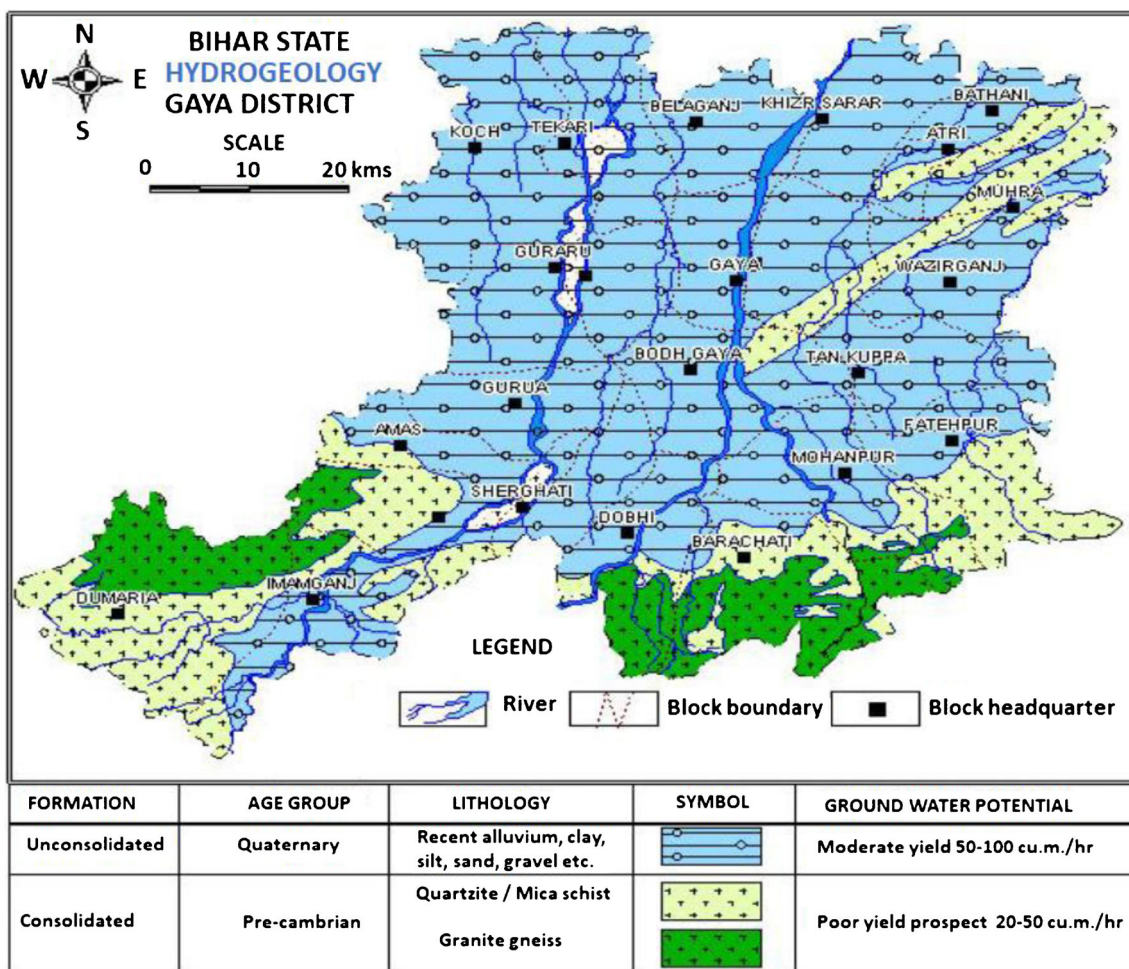


Fig. 2 Hydrogeological map of Gaya district (source: CGWB 2013))

3891 mg l⁻¹. The higher variation of EC value might be due to random variation of hydrogeology. The water quality parameters like bicarbonate, chloride, and sulfate were found well below the BIS acceptable limit (BIS 2012). From the analysis, it was found that major anions in the groundwater were chloride (3rd quartile = 88.9 mg l⁻¹) followed by sulfate (3rd quartile = 47.3 mg l⁻¹), nitrate (3rd quartile = 30.1 mg l⁻¹), and phosphate (3rd quartile = 0.5 mg l⁻¹). The higher amount of chloride ions shows its conservative nature in groundwater system. Chloride is not lost from physical or chemical process like sorption and precipitation, causing higher concentrations among various water quality attributes (Gascoyne 1989). The shallow groundwater level (< 200 ft) also favors the calcium-magnesium-sulfate-bicarbonate-type groundwater of this region.

The nitrate concentrations in samples ranged from 0.7 to 54.9 mg l⁻¹ with a mean value of 19.9 mg l⁻¹, and 12% of the groundwater samples exceeded the BIS acceptable limit (45 mg l⁻¹). Also, 5% of groundwater samples exceeded the WHO acceptable limit of nitrate (50 mg l⁻¹). Threshold concentrations of 1 mg l⁻¹ and 3 mg l⁻¹ were also considered to

represent possibility of anthropogenic contamination of groundwater due to the surface origins (Madison 1984; Nolan and Hitt 2006). Considering the above thresholds, 99% and 94.87% of the samples exceeded the limit of 1 mg l⁻¹ and 3 mg l⁻¹ of threshold limits of nitrate, respectively. That clearly indicated some initiation of nitrate contamination due to anthropogenic activities. Gaya is an important grain commodity base of South Bihar, where extensive agriculture is the primarily land use and groundwater is the main source of water demand for irrigation. The anthropogenic activity like fertilizer and manure application and high nutrient content irrigation water might be the reason for elevated level of nitrate in this region. Furthermore, high chloride content in groundwater (10% samples exceeded the BIS acceptable limit) also indicate the source of contamination from anthropogenic intervention.

Figure 3 shows the Spearman's correlation coefficient between different water quality attributes of groundwater samples. The histogram and bell-shape density curve show that pH, calcium, and magnesium were approximately normally distributed. However, the nitrate and phosphate data were

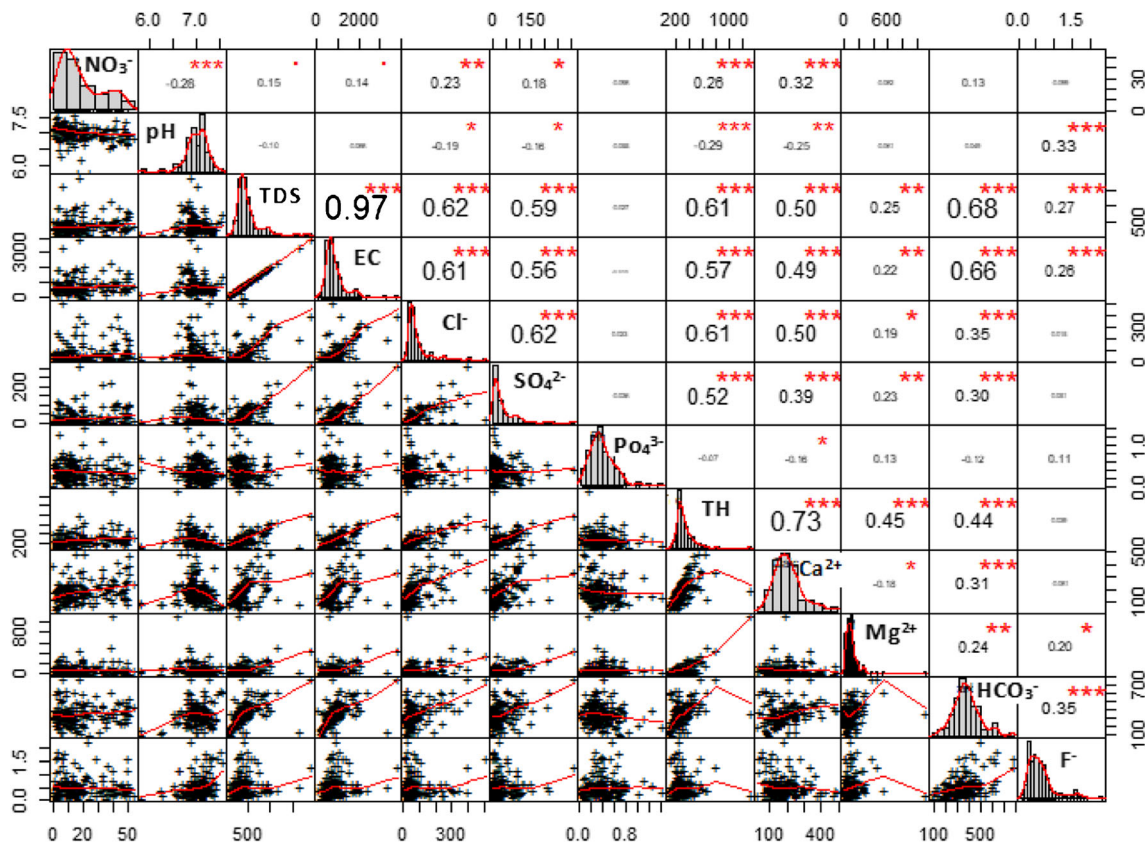


Fig. 3 Spearman's correlation coefficient between measured parameters of groundwater sample

rightly skewed indicating existence of some abnormality in the dataset. In addition, some other attributes like EC, chloride, sulfate, and magnesium were found highly right skewed following approximately lognormal distribution (Fig. 3). The poor correlation of nitrate (NO₃⁻) ($r < 0.50$) with others' water quality attributes was observed for this geochemical data. The correlation of nitrate with chloride represents the initiation of nitrate contamination by anthropogenic intervention like application of fertilizer or percolation of domestic sewage water

or nutrient-rich irrigation water (Kim et al. 2015). Insignificant correlation ($r < 0.18$) of nitrate with sulfate might be due to sulfide oxidation. Furthermore, there were insignificant one by one correlations ($p < 0.05$, $r = 0.39$, 0.23) between sulfate with calcium and magnesium. That reflects either slow decomposition of calcium-bearing minerals like calcite and dolomite or congruent dissolution of magnesium containing silicate and carbonate minerals in the study area as results of the water-rock interaction (Kumar et al. 2006). Thus, the elevated level of nitrate in groundwater is mainly controlled by the external sources.

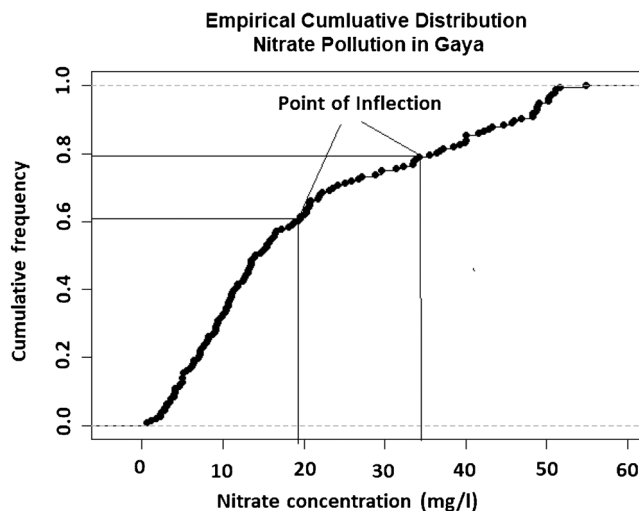


Fig. 4. The empirical cumulative frequency plot for nitrate

Source identification of nitrate in groundwater

The graphical representation of cumulative frequency curve (CFC) for nitrate concentration is shown in Fig. 4. Cumulative frequency shows the running total of frequencies (number of times an event occurs within a given scenario). The CFC plot represents the two step inflection points (the minimum gap between actual and simulated) around 20 and 35 mg l⁻¹. This clearly indicates that different factors are controlling the abnormal behavior of nitrate concentration in groundwater system for this region (Matschullat et al. 2000). To understand the underlying but unobservable relationship among their

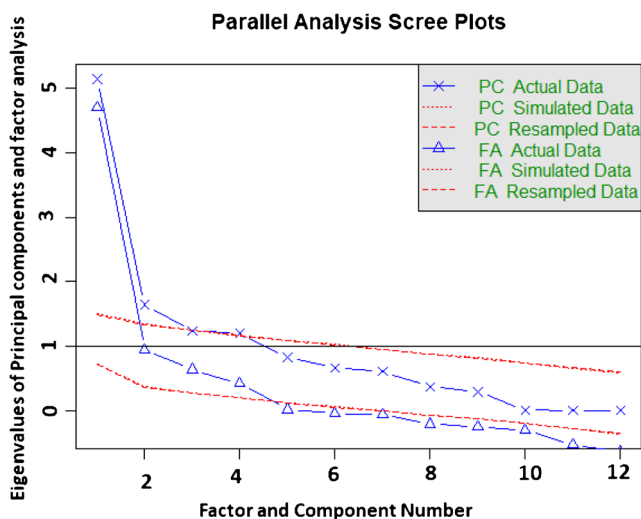


Fig. 5. Scree plots (selection of factor based on the eigenvalue)

attributes for nitrate contamination, factor analysis technique is used. The factor analysis was performed using free open source R software with “Psych” package (Revelle 2011).

The scree plot was drawn to extract the number of factors (Fig. 5). The blue and two red lines (on top of each other) represent the actual, simulated, and resampled data, which is estimated based on the parallel analysis (Revelle 2011). From the scree plot, the selected number of factors in between 2 and 4 can be a good choice to determine the factor loading for dataset. This study preferred to select four factors, which explain about 70% of the total variance of the dataset (Table 3). Furthermore, to obtain the factor loading, the oblique rotation was preferred based on the assumption that factors have a certain correlation. This was observed while computing

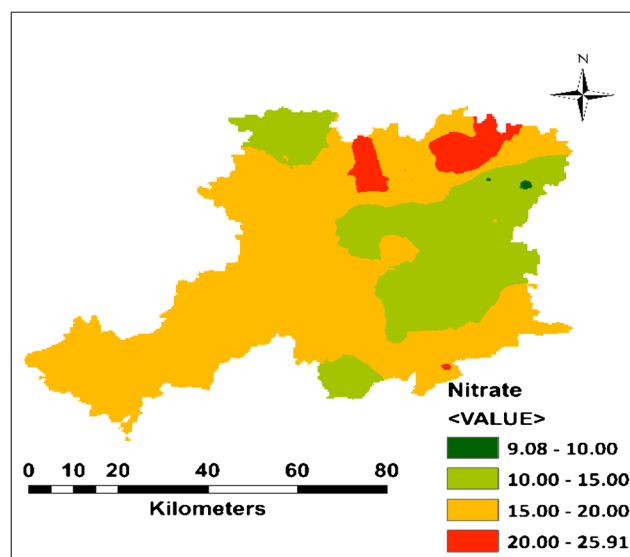


Fig. 7 Spatial prediction of nitrate using EBK modeling

Spearman’s correlations. This method derives solutions through iterative eigenvalue decomposition without considering prior normal distribution. All the geochemical data attributes do not follow normal distribution (pH, calcium, and magnesium) as stated earlier.

From the analysis, factor 1 explains 24% of the total variance in the dataset and has significant variance contribution from EC followed by chloride, sulfate, and phosphate (Table 4). Therefore, factor 1 can be considered “mixed factor” from natural and anthropogenic influences. Factor 2 explains 18% of the total variance, that is primarily contributed by pH and bicarbonate. As these two geochemical attributes control the redox environment in the groundwater system, factor 2 can be

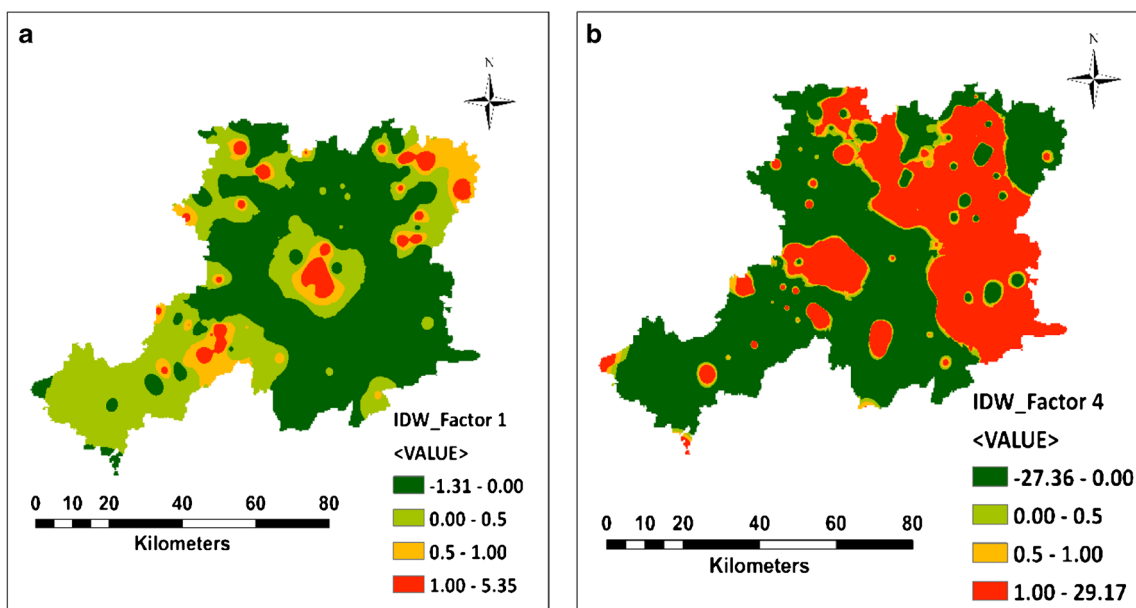


Fig. 6 a, b Spatial distribution of factors 1 and 4

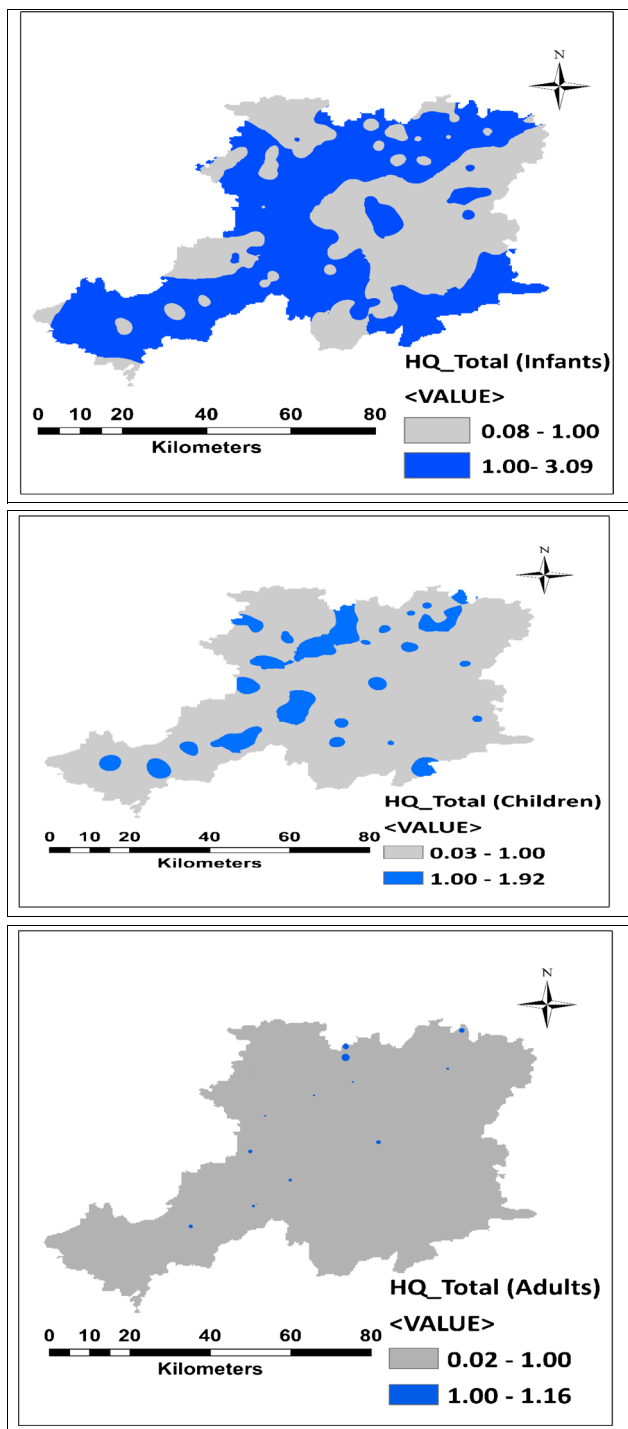


Fig. 8 Spatial distribution map of total hazard index of nitrate in groundwater of Gaya district using IDW interpolation

considered redox factor. In addition, factor 2 is primarily controlled by HCO_3^- ($\text{pH} < 8.5$), whereas factor 3 contributes 15% of the total variance contributed from magnesium followed by sulfate and chloride. The positive loading of sulfate and chloride infer the same source of origin, which was derived from domestic sewage and industrial sewage discharge in terms of anthropogenic intervention. Factor 4 shows the significant

association of nitrate with calcium and can be considered “nitrate” factor, which contributes 13% of the total variance of the geochemical data. The association of nitrate with chloride and calcium represents that nitrate is controlled by anthropogenic intervention. The above said finding is supported by factor correlation matrix as factors 1 and 4 were significantly correlated (Table 5).

For better understanding, SPCA was applied using “elasticnet” package (Friedman et al. 2009) in R open source software. Four principal components were selected based on scree plot (Fig. 5). From analysis, it was clear that the first principal component (PC) was significantly contributed by EC followed by sulfate. Therefore, similar to factor analysis, first PC represented the “salinity factor” also indirectly representing the total ions present in the groundwater samples. The second and third PCs have little contribution for underlying processes as similar conclusion was found with factor analysis. The fourth PC represents the “nitrate” factor and water quality attributes responsible for the presence of elevated level of nitrate in the groundwater system. It was found that nitrate was strongly associated in terms of loading with calcium followed by chloride and sulfate. The positive loading on nitrate with chloride indicated that nitrate comes from same sources of chloride, that may be derived from domestic and industrial sewage. Furthermore, the association of nitrate with sulfate showed the application of fertilizer is the reason behind elevated nitrate level of nitrate in the groundwater system (Cao et al. 2016). These all are external sources indicate that anthropogenic activity might be the reason for the elevated nitrate.

To understand, the groundwater samples variability in the different regions of the study area, based on the factor scores (1 and 4), inverse distance weighted (IDW) interpolation method was used (Fig. 6). Samples in the NE, middle, west, and SW parts of study area with factor loading > 0 indicated high salinity and anthropogenic pollution (Fig. 6a) contributed by chloride, sulfate, and phosphate. Samples in N to SE and middle of SW parts of study area with factor 4 score > 0 indicated nitrate enrichment, which can be due to association of chloride, sulfate, and calcium as a result of anthropogenic intervention (Fig. 6b). A large portion of the study area has exposure to recent alluvium enriched with clay and silt (Fig. 2) leading to poor connectivity between the unsaturated zone of soil and atmosphere. This condition assists microbial action for reduction of nitrate to nitrite and leaching into groundwater resulting in elevated nitrite concentration in groundwater.

Modeling of nitrate using empirical Bayesian

Semi-variogram-based model was used to estimate the nitrate (NO_3^-) concentrations at unsampled locations. Different EBK

Table 7 Results of cross-validation for evaluating the performance of six EBK models for nitrate datasets

EBK models	Mean error	Mean standardized error	Root-mean-square standardized error	Mean standard error (mm)	Root mean square
Nitrate					
Linear	- 0.5318	- 0.033	0.969	15.94	15.539
Power	- 0.5374	- 0.033	0.977	15.74	15.446
Thin plate spline	- 0.4168	- 0.014	0.967	16.89	16.485
Exponential no.	0.017	- 0.003	0.9696	15.50	15.026 [#]
Whittle	- 0.052	- 0.008	0.9732	15.46	15.028
K-Bessel	- 0.017	- 0.017	1.003	15.318	15.111

[#] Comparatively favorable values among employed empirical Bayesian kriging models

models were tested to identify the best fitted semivariogram for the nitrate dataset. There is no specific rule for selecting the best model for any data set. It is a trade-off process resulting in lowest root mean square error. From the results, the exponential model semivariogram was best fit for nitrate data after setting subset size (70), overlap factor (1.5), and major semi-axis (0.123) on 500 simulations. The final model was selected considering the lowest root mean square error (Table 7). The spatial distribution of nitrate map was plotted to determine the locations of elevated levels of concentration of nitrate in the study area.

The distribution map of nitrate shows the highest concentration value in the mid-north portion of the study region (Fig. 7). The shallow groundwater and younger alluvial soil type of this region assists the groundwater nitrate contamination through surface water interaction. The raster map of nitrate concentration reveals that concentrations are well below the acceptable limit in this study region (BIS 2012; WHO 2011).

Table 8 Descriptive statistics of risk assessment in terms of total hazard quotient

Statistic	Hazard index (HI)		
	Infants	Children	Adults
Minimum	0.04	0.02	0.01
Maximum	3.10	1.92	1.16
1st quartile	0.46	0.29	0.17
Median	0.81	0.50	0.30
3rd quartile	1.70	1.05	0.64
Mean	1.13	0.70	0.42
Variance (<i>n</i>)	0.73	0.28	0.10
Standard deviation (<i>n</i>)	0.85	0.53	0.32
Skewness (Pearson)	0.81	0.81	0.81
Kurtosis (Pearson)	- 0.63	- 0.63	- 0.63
Geometric mean	0.80	0.50	0.30

Human risk assessment

The human risk assessments via two pathways (oral and dermal) for three age groups were performed. The hazard quotient was calculated based on Eqs. (2) and (5) consisting of both oral and dermal exposure to nitrate in the groundwater. The value for HQ_{oral} ranged from 0.037 to 3.08, for infants; 0.02 to 1.90, for children; and 0.0136 to 1.1476 for adults. The value for HQ_{dermal} ranged from 2×10^{-4} to 1.8×10^{-2} for infants; 1.6×10^{-4} to 1.3×10^{-2} for children; and 1.3×10^{-4} to 1.1×10^{-2} for adults. Table 8 shows the descriptive statistics of HI, considering different age groups (infants, children, and adults). If the calculated HI is less than 1, then no adverse health effects are expected due to exposure of nitrate. The hazard index value for HI_{infants}, HI_{children}, and HI_{adults} are ranged from 0.04 to 3.10, 0.02 to 1.92, and 0.01 to 1.16 respectively. From the results, it was observed that the mean value for HI has slightly exceeded for infants in this region, whereas for other cases (children and adults), they are well below 1. This investigation shows that NO₃⁻ is an accountable element for health risk to infants and need to be examined. Young people (infants and children) are usually associated with higher health risks to contaminants than adults, and similar conclusions have been reported widely by Wu and Sun (2016), Li et al. (2018), Li et al. (2014), and Adimalla et al. (2018a).

To understand the spatial distribution of risk variability (in terms of HI) for different groups, the IDW interpolation method was used. The spatial distribution of hazard index for different groups is shown in Fig. 8. From Fig. 8, it was observed that residents in the northwest and southwest region pose health risk. Especially infants were the most vulnerable group due to exposure of drinking water at a same level of pollution followed by children. Adults were mostly safe in different areas except for a few elevated pockets. The finding of result can be helpful to the population of the region for taking safeguard against utilizing fertilizer and improving groundwater quality. In addition, there is a need for water resource management at a local scale to improve the groundwater quality by preventing the leaching of nitrate to the groundwater system.

Conclusions

- This investigation reveals that the groundwater of this region is on the slightly acidic side. The major dominant ions in the groundwater are $\text{HCO}_3^- > \text{Cl}^- > \text{SO}_4^{2-} > \text{PO}_4^{3-}$ and $\text{Ca}^{2+} > \text{Mg}^{2+}$ showing samples of this region fall under the weak mineralized zone.
- The measured concentration of nitrate in 95% water samples from the middle Gangetic plains is well below the acceptable limit of WHO (2008), whereas 88% of samples were below the BIS safe limit.
- The results of factor analysis identified various underlying processes that controlled the nitrate concentration in groundwater. Two identified factors were selected to understand the nitrate process. Factor 1 represented by the mixed factor indicated external sources of ions influencing regional hydrochemistry. The nitrate factor (factor 4), showed association of nitrate with chloride and calcium that represented nitrate concentrations controlled by anthropogenic intervention.
- The groundwater of region was of $\text{Ca}^{2+} - \text{Cl}^- (\text{NO}_3^- + \text{SO}_4^{2-})$ type groundwater with elevated level of NO_3^- .
- EBK modeling shows that nitrate concentration of this region varied from 9.07 to 25.91 mg l^{-1} . The nitrate concentration of this region was well within the acceptable levels of BIS and WHO.
- Health risk assessment due to exposure of nitrate in drinking water shows that infants' risk was highest followed by that in children and adults. Therefore, water quality management needed for drinking water should put more emphasis on effects on infants at local scale.

Acknowledgments The authors are profoundly grateful to the reviewers and the associate editor for the careful examination of the draft of the manuscript and their many valuable comments and suggestions to help improve the manuscript.

Funding information This work was supported by the Board of Research and Nuclear Sciences under Department of Atomic Energy, India, for providing financial assistance under the National Uranium project (NUP) (BRNS Project Ref. No.: 36(4)/14/10/2014-BRNS).

References

- Adams S, Titus R, Pietersen K, Tredoux G, Harris C (2001) Hydrochemical characteristics of aquifers near Sutherland in the Western Karoo, South Africa. *J Hydrol* 241:91–103
- Adimalla N, Li P, Qian H (2018a) Evaluation of groundwater contamination for fluoride and nitrate in semi-arid region of Nirmal Province, South India: a special emphasis on human health risk assessment (HHRA). *Hum Ecol Risk Assess Int J*:1–18. <https://doi.org/10.1080/10807039.2018.1460579>
- Adimalla N, Vasa SK, Li P (2018b) Evaluation of groundwater quality, Peddavagu in Central Telangana (PCT), South India: an insight of controlling factors of fluoride enrichment. *Modeling Earth Syst Environ* 4:841–852
- Aiuppa A, Bellomo S, Brusca L, d'Alessandro W, Federico C (2003) Natural and anthropogenic factors affecting groundwater quality of an active volcano (Mt. Etna, Italy). *Appl Geochem* 18:863–882
- Almasri MN (2007) Nitrate contamination of groundwater: a conceptual management framework. *Environ Impact Assess Rev* 27:220–242
- APHA (2005) Standard methods for the examination of water and wastewater. American Public Health Association (APHA), Washington, DC
- Arabgol R, Sartaj M, Asghari K (2016) Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (SVMs) model. *Environ Model Assess* 21:71–82
- Behera B, Das M (2018) Application of multivariate statistical techniques for the characterization of groundwater quality of Bachel and Kirandul area, Dantewada district, Chattisgarh. *J Geol Soc India* 91:76–80
- BIS (2012) Indian Standards Specifications for Drinking Water vol IS: 10500. Bureau of Indian Standards, New Delhi
- Cao Y, Tang C, Song X, Liu C, Zhang Y (2016) Identifying the hydrochemical characteristics of rivers and groundwater by multivariate statistical analysis in the Sanjiang Plain, China. *Appl Water Sci* 6:169–178
- CGWB (2013) Ground Water Information Booklet, Gaya District, Bihar State. Central Ground water Board. Ministry of water Resources (Govt. of India)
- Chang C-C, Tsai S-S, Wu T-N, Yang C-Y (2009) Nitrates in municipal drinking water and non-Hodgkin lymphoma: an ecological cancer case-control study in Taiwan. *J Toxic Environ Health A* 73:330–338
- Chen J, Tang C, Sakura Y, Yu J, Fukushima Y (2005) Nitrate pollution from agriculture in different hydrogeological zones of the regional groundwater flow system in the North China Plain. *Hydrogeol J* 13: 481–492
- Chesnaux R, Allen D (2008) Simulating nitrate leaching profiles in a highly permeable vadose zone. *Environ Model Assess* 13:527–539
- Davis J (1973) Statistics and Data Analysis in Geology. Wiley, New York
- Elisante E, Muzuka AN (2016) Assessment of sources and transformation of nitrate in groundwater on the slopes of Mount Meru, Tanzania. *Environ Earth Sci* 75:277
- Fabro AYR, Ávila JGP, Alberich MVE, Sansores SAC, Camargo-Valero MA (2015) Spatial distribution of nitrate health risk associated with groundwater use as drinking water in Merida, Mexico. *Appl Geogr* 65:49–57
- Friedman J, Hastie T, Tibshirani R (2009) glmnet: lasso and elastic-net regularized generalized linear models. R package version 1
- Ganesh S, Khan F, Ahmed M, Velavendan P, Pandey N, Mudali UK (2012) Spectrophotometric determination of trace amounts of phosphate in water and soil. *Water Sci Technol* 66:2653–2658
- Gao Y, Yu G, Luo C, Zhou P (2012) Groundwater nitrogen pollution and assessment of its health risks: a case study of a typical village in rural-urban continuum, China. *PLoS One* 7:e33982
- Gascoyne M (1989) High levels of uranium and radium in groundwaters at Canada's Underground Research Laboratory, Lac du Bonnet, Manitoba, Canada. *Appl Geochem* 4:577–591
- Gribov A, Krivoruchko K (2012) New flexible non-parametric data transformation for trans-gaussian kriging. In: *Geostatistics Oslo 2012*. Springer, pp 51–65
- Gu B, Ge Y, Chang SX, Luo W, Chang J (2013) Nitrate in groundwater of China: sources and driving forces. *Glob Environ Chang* 23:1112–1121
- He S, Wu J (2018) Hydrogeochemical characteristics, groundwater quality, and health risks from hexavalent chromium and nitrate in groundwater of Huanhe formation in Wuqi County, Northwest China. *Expo Health*. <https://doi.org/10.1007/s12403-018-0289-7>
- He X, Wu J, He S (2018) Hydrochemical characteristics and quality evaluation of groundwater in terms of health risks in Luohe aquifer

- in Wuqi County of the Chinese Loess Plateau, northwest China. *Hum Ecol Risk Assess Int J*:1–20. <https://doi.org/10.1080/10807039.2018.1531693>
- Jayakumar R, Siraz L (1997) Factor analysis in hydrogeochemistry of coastal aquifers—a preliminary study. *Environ Geol* 31:174–177
- Jolliffe IT (1986) Principal component analysis and factor analysis. In: *Principal component analysis*. Springer, pp 115–128
- Kaown D, Koh D-C, Mayer B, Lee K-K (2009) Identification of nitrate and sulfate sources in groundwater using dual stable isotope approaches for an agricultural area with different land use (Chuncheon, mid-eastern Korea). *Agric Ecosyst Environ* 132:223–231
- Kim H-S, Park S-R (2016) Hydrogeochemical characteristics of groundwater highly polluted with nitrate in an agricultural area of Hongseong, Korea. *Water* 8:345
- Kim K-H, Yun S-T, Mayer B, Lee J-H, Kim T-S, Kim H-K (2015) Quantification of nitrate sources in groundwater using hydrochemical and dual isotopic data combined with a Bayesian mixing model. *Agric Ecosyst Environ* 199:369–381
- Krivoruchko K (2012) Empirical bayesian kriging *ArcUser* Fall:6–10
- Krivoruchko K, Butler K (2013) Unequal probability-based spatial mapping. Esri, Redlands Availableonline:<http://www.esri.com/esr/news/arcuser/spring2013/~media/Files/Pdfs/news/arcuser/0313/unequal.pdf>
- Kumar M, Ramanathan A, Rao M, Kumar B (2006) Identification and evaluation of hydrogeochemical processes in the groundwater environment of Delhi, India. *Environ Geol* 50:1025–1039
- Li P, Wu J, Qian H, Lyu X, Liu H (2014) Origin and assessment of groundwater pollution and associated health risk: a case study in an industrial park, northwest China. *Environ Geochem Health* 36:693–712
- Li P, Li X, Meng X, Li M, Zhang Y (2016) Appraising groundwater quality and health risks from contamination in a semiarid region of northwest China. *Expo Health* 8:361–379
- Li P, Feng W, Xue C, Tian R, Wang S (2017) Spatiotemporal variability of contaminants in lake water and their risks to human health: a case study of the Shahu Lake tourist area, northwest China. *Expo Health* 9:213–225
- Li P, He X, Li Y, Xiang G (2018) Occurrence and health implication of fluoride in groundwater of loess aquifer in the Chinese Loess Plateau: a case study of Tongchuan, Northwest China. *Expo Health*. <https://doi.org/10.1007/s12403-018-0278-x>
- Liao L, He JT, Zeng Y, et al. 2016. A study of nitrate background level of shallow groundwater in the Liujiang Basin. *Geol in China* 43(2): 671–82
- Madison RJ (1984) Overview of the occurrence of nitrate in ground water of the United States. National Water Summary 1984-Water-Quality Issues. US Geol Surv Water-Supply Paper 2275:93–103
- Matschullat J, Ottenstein R, Reimann C (2000) Geochemical background—can we calculate it? *Environ Geol* 39:990–1000
- Nolan BT, Hitt KJ (2006) Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. *Environ Sci Technol* 40:7834–7840
- Revelle W (2011) An overview of the psych package. Department of Psychology Northwestern University Accessed on March 3:2012
- Su X, Wang H, Zhang Y (2013) Health risk assessment of nitrate contamination in groundwater: a case study of an agricultural area in Northeast China. *Water Resour Manag* 27:3025–3034
- Wang H, Gu H, Lan S, Wang M, Chi B (2018) Human health risk assessment and sources analysis of nitrate in shallow groundwater of the Liujiang basin, China. *Hum Ecol Risk Assess Int J*:1–17
- WHO (2008) Guidelines for drinking-water quality, third edition, incorporating first and second addenda. Geneva
- WHO (2011) Guidelines for drinking-water quality. *WHO Chronicle* 38: 104–108
- Wu J, Sun Z (2016) Evaluation of shallow groundwater contamination and associated human health risk in an alluvial plain impacted by agricultural and industrial activities, mid-west China. *Expo Health* 8:311–329
- Wu J, Li P, Qian H, Duan Z, Zhang X (2014) Using correlation and multivariate statistical analysis to identify hydrogeochemical processes affecting the major ion chemistry of waters: a case study in Laoheba phosphorite mine in Sichuan, China. *Arab J Geosci* 7: 3973–3982
- Zhai Y, Lei Y, Wu J, Teng Y, Wang J, Zhao X, Pan X (2017a) Does the groundwater nitrate pollution in China pose a risk to human health? A critical review of published data. *Environ Sci Pollut Res* 24:3640–3653
- Zhai Y, Zhao X, Teng Y, Li X, Zhang J, Wu J, Zuo R (2017b) Groundwater nitrate pollution and human health risk assessment by using HHRA model in an agricultural area, NE China. *Ecotoxicol Environ Saf* 137:130–142
- Zhang X, Xu Z, Sun X, Dong W, Ballantine D (2013) Nitrate in shallow groundwater in typical agricultural and forest ecosystems in China, 2004–2010. *J Environ Sci* 25:1007–1014
- Zhang Y, Wu J, Xu B (2018) Human health risk assessment of groundwater nitrogen pollution in Jinghui canal irrigation area of the loess region, northwest China. *Environ Earth Sci* 77:273
- Zou H, Hastie T, Tibshirani R (2006) Sparse principal component analysis. *J Comput Graph Stat* 15:265–286