JOURNAL OF GEOINTERFACE, Vol.1, No.1, July 2022 pp.2-7

Groundwater Contamination: A Review

B. K. Sahu*
Department of Earth Sciences, IIT Bombay, Mumbai- 400076. India
*Corresponding author: bas.sahu37@gmail.com

Abstract: All contaminants in groundwater are not necessarily harmful but are essential for human growth and good health. But if the level of contaminant concentrations exceeds the safe limits, as set by WHO, then these contaminants are termed as pollutants and thereby can have adverse health and other effects, if used in domestic, industrial, and agricultural sectors. Such water pollutants must be removed by suitable physico-chemical remediation processes before application for any purpose. Aquifers are heterogeneous at microscopic to megascopic scales and often display unconnected compartments. So, a statistically homogeneous (EPM) model with a dynamic approach provides more accurate and realistic estimation and optimal inferences. Contaminant concentrations are log-normally distributed, and the parameters of a lognormal model can be estimated from adequate independent (random) sample volume (REV), where the sample sites are widely spaced. Hypotheses tests such as mean concentration of samples = population concentration; sample variance=population variance, are performed by Univariate Statistical analysis. However, there is a variety of contaminants (cations, anions, insecticides, pesticides, etc), hence a multivariate (MND) model is appropriate for better statistical analysis and geochemical inference. Factor model for a single population and a single set of variables can reduce the Principal Components (p) to a smaller number of informative components (m << p) using the explained variance of m principal component to be 85% or so; the remaining components (errors due to measurement, sampling and model approximations) due to insufficient information. A second-order criterion of inflexion point or minima point on a cumulative variance of principal components vs number of principal components along with the above first-order criterion of 85% signal correctly identify the exact number of common components (m). Orthogonal (Varimax) rotation of retained significant components(m<<p) provides an identification tool for recognizing the process causing the rotated factor on basis of significantly high absolute values of rotated loadings (partial correlations) between rotated factor and the set of highly loaded variables. Causation of different rotated factors can be (i) natural weathering and infiltration, rock-water interaction, (ii) domestic use, (iii) industrial effluents, (iv) agricultural use of chemical fertilizers, (v) transport industry, (vi) well drilling for oil & gas; (vii) mining industry, and, (vii)Power generation. MND theory can also predict the levels of pollution at different sites and at different times to take optimal decisions.
Keywords: Groundwater, Contamination, EPM model.

Introduction

Groundwater is very important for socio-economic and industrial growth in a developing country like India. About 85% of domestic and 55% of irrigation demands in rural areas and 50% of demands for urban development and industries are met by groundwater resources. However, since groundwater is essential for drinking purposes, domestic use, and the very survival of humankind, its overexploitation and misuse for irrigation and/or industrial purposes cannot continue if the quantity and quality of water is to be maintained for drinking purposes. In the past, anthropogenic degradation of natural resources (including groundwater) has remained within the natural carrying capacity of Earth`s natural processes and hence, simple linear and deterministic solutions like Darcy`s Law solved most of the problems since there was little contamination, pollution or waste. However, during the last three decades or so, an adequate supply of good quality groundwater has demanded much greater accuracy and precision in estimations for control of quality remediation which can be achieved through interdisciplinary, non-linear, and/or stochastic models.

A confined aquifer is largely deterministic in geometry but unconfined aquifers can be highly varied in their saturated zones, upper and lower surfaces of saturation zones. Both confined and unconfined aquifers can be homogeneous and/or heterogeneous in terms of their characteristics such as porosity, permeability, connectivity, and also on their levels of contamination/pollution. Hence, we can estimate only their average/stable values in samples of suitably large volume (representative elementary volume or REV, Bear, 1972) of these aquifers. These estimates are made in micro-, meso-, macro-, and even mega/basin scales as required. WHO guidelines for safe and permissible limits of water quality are as follows;

• pH 6.5-8.5; mean 7.5 for drinking; 6.68-8.00 mean 7.15) for domestic/industrial use and
• Electrical Conductivity, EC: 1400 μS/cm; (iii) TDS: 500 mg/l.

Similarly, the Bureau of Indian Standards (BIS), in 1991, set guidelines for safe limits for drinking water in 100 ml samples as 95% of samples to be free of gases, coliform organisms; No sample to have E. Coli; E Coli should not be detected in two consecutive supplies. pH should be within 6.5-7.5; TH < 300; Fe < 0.3 mg/l; Cl- from 2500 mg/l to 0.2mg/l. It has set safe limits for groundwater for irrigation use.

However, since EC is perfectly correlated to TDS (total cations/anions), it is best to use only one of these two random variables (rvs) (e.g. TDS) for statistical analysis. Similarly, total hardness (TH) is completely dependent on anion concentrations of HCO_3, CO_3, and SO_4 and hence TH can be dropped for later statistical analysis. Groundwater is contaminated/polluted by soluble cations and/or anions, but rarely polluted by natural rock(aquifer)-water interactions except in rare localities having toxic elements such as F, As, S, Se, U, etc. in the aquifers/host rocks. The chemical quality of groundwater can be affected by evapotranspiration (in summer), precipitation (in monsoons), and also by the composition of connate waters in deep aquifers. About 1/3 of the world`s population living in underdeveloped countries live with permanent water scarcity. In arid and semi-arid regions, Rajasthan, central and south India, rainfall is extremely irregular and very low. This induces problems for food security, ecology, environment, and survival of mankind. Hence, groundwater use must be restricted only to drinking and domestic purposes, and domestic wastes must be properly treated and recycled. Realistic geochemical models must be developed for parameter estimation, hypotheses testing, and taking optimal decisions for the management of pollution and remedial measures to control it.

Groundwater aquifers (soil, weathered rocks, alluvium, sedimentary rocks) conform to simplistic assumptions of (i) uncorrelated-ness among sample values; (ii) homogeneous pore distribution systems in holding and transmitting fluids through them. However, sample values display spatial dependency at all scales of pores (micro-, meso-, and macro-levels) if sampling spacing is very small. Impermeable rocks and structures (faults and joints) separate the aquifer into isolated/unconnected compartments which are not independent or random. Flow, dispersion, and displacement processes in fluids can be studied as a homogeneous continuum model using the representative elementary volume (REV, Bear,1972) of sample sizes or as a discrete heterogeneous equivalent network model in fractured rocks. It is pragmatic and more economical to adopt a macroscopic view with a length scale much larger than the dimensions of pores/grains or individual fractures in the aquifers. Stable sample statistics can be obtained using samples taken during Pre-Monsoon and Post-Monsoon periods as sample statistics during Rainy seasons (Monsoon period) are generally inconsistent and unstable due to irregular and inconsistent precipitations. Characterization of granular and fractured aquifers can be achieved through classical models such as (i) Hg – porosimeter; (ii) adsorption-desorption using gases such as N, Ar, He; (iii) constant and variable head permeameter; although more accurate modern methods include, (i) NMR;(ii) small angle X-ray scattering; (iii) fractals; and (iv) air permeameter for shale-gas reservoirs. General information on groundwater contamination can be found in Fetter (2001) and especially in Fetter et al. (2018).

Groundwater contaminants

Groundwater is the main and/or only source of drinking water especially in arid, rural, and suburban areas. About 35% of groundwater is used for human consumption, about 36% for irrigation, and 28% for industries. However, over-exploitation of limited groundwater resources in recent years has stressed the groundwater supply and affects about 2 billion people. The quality of groundwater is crucial for reducing health hazards, its natural sources being, TDS, sulphates, chlorides; and, its anthropogenic sources being, dissolved inorganic cations and anions; organic compounds, pathogens, and radio-nuclides. These contaminants need to be remedied for safe use in domestic, agricultural, and industrial purposes which is rather expensive. The main sources of contaminants are (i) septic tanks, open/unlined drains, injection wells, and water spreading tanks for recharge; (ii) landfills, graveyards, and underground storage tanks, (iii) pipelines for metallic waste transport; (iv) irrigation including pesticides, insecticides, fertilizers, farm refuse. The main inorganic cations and anions which are contaminants are (a) Cations: As, Pb, Hg, Cd, Cr^{6+}, Be (b) Anions: F, Cl, Nitrates, Sulphates, Phosphates. Natural processes of chemical rock weathering (hydrolysis, carbonation, etc.) release soluble cations and anions which become contaminants in groundwater: (a). Cations: H, K, Na, Ca, Mg, Pb, As and (b) Anions: OH, CO_2, HCO_3, Cl, F, SO_4. This paper focuses mainly on natural contaminants/pollutants in groundwater.

Morphology of aquifers

Aquifers are soils, weathered rocks, alluvium, sedimentary rocks (sandstones, carbonates), fractured rocks, etc. Accessible porosity (effective porosity) can be defined as; porosity (phi) = 1 – (v(1)/v(s)) – (v(2)/v(s)), where v (1), v (2) and v(s) are volumes of total voids, isolated voids, and solids in the aquifer. Permeability results from effective/connected porosity. Tortuosity, T(p), is the ratio of the total length of the flow path(streamline) to the straight line length of the path from starting to finishing points. Adsorption measures specific surface defined as the ratio of the surface area of connected pores to the volume of the aquifer. Pore/ grain size distribution is modelled as log-normal (or, phi-normal; where phi =-log (to base 2) (pore/grain sizes in mm). Hence, parameters of the log-normal distribution for each pore/grain size mode can theoretically predict the permeability of the aquifer (Sahu, 2020). Fractured rocks are characterized at two different scales: (i) single fracture which need not be smooth or have parallel upper and lower surfaces, (ii) network of intersecting fractures with complex geometry and orientation due to tectonic stresses (Sahu, 2006). Therefore, the identification of co-genetic and parallel fracture sets is very important for inference and flow modeling. Fracture trace lengths are modelled as log- normal density. However, measurement of fracture lengths is often biased because of (i) exposure may not include both the end-points, (ii) available exposure may not be the desired plane but at random in 3D, and (iii) the possible presence of anthropogenic fractures. Some fractures follow fractal distributions which could be due to diagenesis.

Geochemical characterization

Groundwater from soil aquifers could be renewable through annual precipitation and some subsurface aquifers may have a renewable surface source(s) but are generally considered non-renewable since annual precipitation is fixed (on an average annually). In arid and semi-arid regions, surface water may not be available, and people completely depend on groundwater for consumption. Hence, groundwater quality must be suitable for drinking and domestic use (free of pollutants). Chemical equilibrium is heterogeneous due to the presence of multiple gases CO_2, O_2); liquid water with dissolved salts; solids such as clays, colloids, etc.). Sample characterization is, therefore, site-specific and deals with several random variables in spatial (1D,2D,3D) and time domains. From statistical theory, samples from wider intervals are almost independent, but samples from closer intervals in spatial or time domains are correlated. Statistical analyses and inferences differ greatly in these two different situations and univariate/multivariate analyses are appropriate in the first situation whereas time (spatial) series or geostatistical/fractal methods may be needed in the second situation for correct and optimal decisions (Sahu, 2003, 2005). The sample size for sound parameter estimation and inference varies with different methods, e.g., Univariate statistical methods require more than 20, multivariate statistical methods more than 33, and time series methods more than 100 samples. These sample sizes are for analyses in 1D, while for 2D and 3D analyses, the sample numbers must be adequately increased to at least double for 2D and at least triple for 3D problems. Geostatistical methods have technical problems of non-zero Nugget Effect (Co) at the origin of variograms which should theoretically be zero in the case of homogeneous domains and/or REV sampling procedure). Therefore, the author does not suggest using geostatistical approaches. Linear statistical models are preferred as these are amenable to mathematical analysis over nonlinear approaches which are more complex and very difficult to analyses and interpret. However, truly nonlinear models can be approximated as locally linear and analysed and inferred for the locality (extending to a large number of different localities in the domain of interest). Linear models also require a few pre-requisites for applicability as follows:

(a) Samples must be independent (wide intervals in spatial/temporal domains).

(b) Random variables are continuous random variables and possess Gaussian or Normal PDF/CDF.

The first pre-requisite can be achieved by sampling randomly at sufficient wider intervals. Each sample size must be sufficient (> or equal to REV size) so that we get a homogeneous domain. The second pre-requisite is seldom true (except pH which is -log hydrogen ion concentration). All other cation or anion concentrations (c) are NOT Gaussian/Normal and must be transformed by a suitable log transform (such as log (c/(1-c)) that reduces to log (c) for trace/very low concentrations as then (1-c) tends to 1) to make the contaminant concentration distribution Normal. Since contaminant concentrations in groundwater is very low we can use log (c) transform for normalization before any statistical analyses (Sahu, 1982a,b, 2003, 2005). Geochemical modelling of groundwater contaminants permits the prediction of not only the geological processes that cause these contaminants but also the evolution of these geological processes in the near future. The quality of groundwater must be within the safe (at least permissible) limits for drinking, domestic, irrigation, and industrial uses. Although simpler deterministic solutions provide solutions to contaminant/pollutant problems, realistic solutions are possible through the application of appropriate statistical and or time series approaches. Input

Normalized random variables (rvs) for Univariate, Multivariate, and/or Time series modeling are: pH, log (TDS/EC), log (Mg^{2+}), log (Ca^{2+}), log (Na^{+}), log(K^{+}), log (Cl^{-}), log (HCO_3^{-}), log (SO_4^{2-}), log (NO_3 -), log (CO_3^{-}), log (PO_4^{3-}), log (F^{-}). We can thus use a 13 x 13 covariance/correlation matrix for multivariate statistical analyses/inferences. The linear MND model is a very useful tool for geochemical inference, especially regarding groundwater quality such as contaminants (Sahu, 2005).

Linear multivariate normal distribution (MND) theory

This method is appropriate for analyzing and inferring multiple-correlated random vectors (measurements of contaminant concentrations) in one or more samples from one or more homogenous populations/groups. MND methods are classified on the basis number of populations and on the number of sets of rvs as given below:

Population (s)	One variable	More than ONE Variable
One Set	Principal Component (PCA)	MANOVA, Two- & multi-group
Factor (FA)	Cluster (CA)	Discriminants (LDF), QDF; Classification
More sets	Multiple Regression	Partial MANCOVA correlation; Canonical correlation

MND theory is robust especially if sample sizes are adequate and large. The two parameters are mean vector (mu) and population variance (sigma square; or its positive square root population standard deviation). It is more useful to analyses the dimensionless correlation matrix (R) rather than the covariance matrix for contaminant/pollution problems (Sahu, 2005). Factor Analysis with Orthogonal Rotation: This method applies to a single population and one set of normalized random variables (rvs). Original p-dimensional space is linearly transformed to a smaller dimensional subspace (m) of principal (mutually) orthogonal components. Mathematically this is equivalent to diagonalizing a real symmetric covariance or correlation matrix such that the principal diagonals yield eigenvalues (variance of a linear compound vector) along the orthogonal eigenvectors (new directions). In Factor Analysis (FA), some of the smaller non-significant eigenvalues are dropped (eliminated) as errors without losing any information (signal). The sum of error variances could be due to measurement (2%), sampling (3%), and model (10%); a total of about 15% (Rao, 1973). However, Sahu (1973) introduced a second-order criterion based on inflexion point/ minima point on a graph of cumulative variance on Y- axis vs. eigen-numbers as determined from subsamples from a given population. However, the variance of any rotated factor is different from the corresponding variance of the Principal Component. The retained eigenvectors are standardized by subtracting its mean value and dividing it by its eigenvalue, so each new factors are of equal importance with mean vector zero and variance unity. The retained eigenvectors can be identified as processes causing them through the high absolute values of rotated loadings for each retained factor. Cluster analysis is an empirical method using similarity (strongly related) or distance (highly unrelated) matrices but this is not very accurate. The correlation matrix can be computed by summing over the samples (R mode) or by summing over the input variables (Q mode), both methods having the same information, but R mode analysis is preferred since it is of order $p << N$. Although the individual rotated eigenvalues are not equal to principal component eigenvalues; the total sum of eigenvalues of both original and rotated loading matrices are equal as desired.

Geochemical inference

Inference is based on equilibrium relations between different ions in relation to varying physico-chemical characteristics of groundwater and its host rocks/aquifers. Under acidic conditions (pH 4-6.5) CO_2 from the atmosphere will be dissolved and this induces carbonate dissolutions. Under neutral conditions (pH 6.5 – 7.5), bicarbonates (HCO_3^{-}) become the dominant anion leading to greater carbonate dissolutions. Under alkaline conditions (pH 7.5 – 12), carbonate ions become dominant leading to the precipitation of carbonate minerals/cements. Oxidation -Reduction Potential (Eh) also affects the equilibrium and reaction rates which are also affected by pH. Stronger cations can displace weaker cations as $Na^{+} > K^{+} > Mg^{2+} > Ca^{2+}$. The sodium absorption ratio (SAR) is defined as Na^{+}/(Squre root ($Ca^{2+} + Mg^{2+}$) is useful as a contaminant index. Values of 2-10 have little danger; 7-18 have medium hazard; 18-26 have high hazard and > 26 have very high hazard. Piper`s triangular diagrams are useful for visual comparison and classification of groundwater. Water quality or its contamination (pollution) is based on both physical as well as chemical properties as follows:

Freshwater	Brackish water	Saline water	Brine
0 – 1000	1000-10000	10000-100000	> 100000

Multivariate statistical methodology and inferences are dealt in more detail by Sahu (2005). Here, we give a few general inferences based on chemical changes in groundwater due to, (i) dissolution, (ii) hydrolysis, and (iii) evapotranspiration processes.

- Dissolution: Soluble salts like halite, sylvite in host rocks (aquifers) get dissolved and release cations and anions in same proportions as present in salts.
- Hydrolysis: It is a process of substitution of H+ ions for alkali cations (Na^+ and/or K^+). Different situations occur with acid, neutral or alkaline conditions as given by pH.
- Evapotranspiration: It is dominant in arid and semiarid regions. The contaminant/pollution process is aggravated since much water is transmitted out.
- Orthogonally rotated factor analysis provides sets of homogeneous variables which can be identified as a natural and/or anthropogenic process based on high absolute values of their rotated loadings. Specific sources are: (i) Natural weathering/infiltration; (ii) domestic loaded on Pb, Zn, and bacteria, (iii) industrial effluents (toxic base metals); (iv) agriculture (fertilizers, pesticides, NO_3^-, SO_4, and PO_4)
- Significant multiple regression or partial regression can be used for quality predictions and mapping quality over the 2D space. Correlation analysis (equation) is non-dimensional and hence preferred over multiple/partial regression equations.
- Multiple linear discriminants (M-LDF; Sahu, 1982b) simultaneously delineate each homogeneous population and also for classification.

In conclusion, MND theory and applications are valuable for finding homogeneous sets of contaminants and identifying the source of each set. However, the following are the prerequisites for using MND theory: (i) Input chemical constituents should be randomly located (independent), (ii) chemical random variables must be normalized by suitable log transform (say, log (c), (iii) Homogeneity of covariance matrices should be checked, but for analysis, it is preferable to use a correlation matrix (non-dimensional).

Conclusions

Based on the study, these conclusions can be made:

a. Groundwater fills the pore spaces in soils and porous/fractured rocks as precipitation infiltrates into these aquifers. However, contamination of surface water is almost instantaneous whereas groundwater contamination is a slow process (1 to 10 years). Contaminants in groundwater are harmless, if their level is below safe limits, as set by the WHO or the National Standards, but above these limits they become pollutants and water may be unfit for drinking, household, agricultural or industrial uses.
b. Aquifer characterization should be realistic and accurate by using a suitable stochastic model (Normal, Log-Normal) and parameters of these models can be estimated in large random samples where each sample volume is greater than or equal to REV for the homogeneous aquifer. These estimated parameters are most useful for statistical tests of hypotheses, and predictions in spatial and/time domains of interest.
c. Groundwater contaminants originate from natural weathering of rocks/soils, treatment of wastewater, leakage from drains/pipelines, underground storage tanks, recharge by pumps and surface infiltration ponds, agricultural activities, such as., fertilizers, crop rotations, and flushing; insecticides, urbanization, industrial effluents, landfills, etc. Inorganic contaminants are cations, such as H, Na, K, Ca, Mg, U, and anions e.g., HCO_3, NO_3, PO_4, SO_4. Health hazards are caused by As, Cd, Cr^{6+}, Pb, Hg, Ni, Na, Zn, organisms, viruses, bacteria, etc.
d. Factor analysis with orthogonal rotation of retained significant principal components provides an easy identification tool for recognizing the processes causing the contaminant/pollution sets

on the basis of high absolute values of loadings of any Rotated Factor to the chemical ions. So, we can identify the processes causing the contaminants as (a) Weathering/Infiltration: Na, K, Ca, Mg, HCO_3, Cl, CO_3, SO_4, (b)Domestic sources: Pb, Zn, Cl, insecticides, (c) Industrial Effluents: Cu, Pb, Zn, Cr^{6+}, Ni, Hg, Cl, bacteria,

e. Agricultural K, NO_3, SO_4, PO_4, pesticides, insecticides contaminants are termed undesirable pollutants if concentrations are above safe usage limits. These pollutants must be remedied by a sequence of suitable physico-chemical processes for safe use and to avoid future hazards. Predictions of pollution levels at different spatial and/or temporal locations is possible through MND theory (multiple regression or preferably multiple correlation equation) using the estimated parameters. This can help in pollution remediation processes to operate at required polluted sites/times.

References

Bear, J. (1972) Dynamics of fluids in porous media. Am. Elsevier, New York, USA, 792p.
Fetter, C.W. (2001) Applied Hydrogeology, 4th ed., Prentice Hall, New Jersey, 651p.
Fetter, C.W. (2018) Contaminant Hydrogeology, 3rd ed., Waveland Press, USA, 663p.
Kaiser, H.F. (1958) Varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, pp.187-200.
Rao, C.R (1973) Linear statistical inference and applications. 2nd edn., Wiley, New York, 625p.
Sahu, B.K. (1973) Factor model applied to sphericity and roundness data of quartz grains. Bull. Ind. Geol. Asson. Chandigarh, v.1, pp.73-83.
Sahu, B.K. (1982a) Mineral deposit modeling. Mineralium Deposita, v.17, pp.99-107.
Sahu, B.K. (1982b) Multigroup discrimination of river, beach and dune sands. Math. Geol., v.14, pp.577-586.
Sahu, B.K. (2003) Time series modeling in Earth sciences. Balkema, Lisse, Netherlands, 284p.
Sahu, B.K. (2005) Statistical models in Earth Sciences. BS Publ. Hyderabad, A.P., 211p.
Sahu, B.K. (2006) Groundwater modeling in hard rock terrains. In: Ghosh, N.C. and Sharma, K.D. (Eds.), Groundwater modeling and management. Capital Publ., New Delhi, pp.368-380.
Sahu, B.K. (2011) Groundwater Reservoir (Aquifer) characterization: Geostatistical Approach. In: A Primeron Mineral Geostatistics, DST Natl Training Program, ISM Dhanbad, pp.149-165.
Sahu, B.K. (2019) Distribution Law of mineral and chemical constituent proportions in rocks and ores. In: Essa (ed.) Minerals. InTech OPEN, London, UK, pp.93-107.
Sahu, B.K. (2019) Geochemical modeling of groundwater using MND Theory. In: Eds. M. Kumar et al. Emerging issues in Water Environment during Anthropocene, Springer, Trans Civil & Env Eng. pp.93-106.
Sahu, B.K. (2020) Permeability of clastic reservoirs and aquifers. Jour. Anal Tech. Res., Houston, v.1, pp.1-7.

Manuscript received: 25-05-2022
Manuscript accepted: 14-06-2022

