Stoch Environ Res Risk Assess (2005)
DOI 10.1007/s00477-005-0012-7

ORIGINAL PAPER

Faisal Hossain · Bellie Sivakumar

Spatial pattern of arsenic contamination in shallow wells of Bangladesh: regional geology and nonlinear dynamics

Abstract Since the discovery of large-scale arsenic contamination of groundwater in Bangladesh more than a decade ago, studies related to its spatial characterization have relied on geostatistical approaches and the classical notion of linear stochastic dynamics. This study explores an alternative nonlinear approach, with a motivation to possibly achieve more cost-effective solutions for Bangladesh. It investigates the existence of nonlinear deterministic and chaotic dynamic behavior in the spatial pattern of arsenic contamination in the shallow wells (depth < 150 m). The database comprises the nationwide arsenic survey completed in 1999 by the British Geological Survey (BGS) in collaboration with the Department of Public Health Engineering (DPHE) of Bangladesh. Distinction is made in terms of regional geology (Pleistocene vs. Holocene deposits/Northwest vs. Southwest) to understand the geologic dependency. Identification of possible presence of nonlinear deterministic and chaotic patterns is made via the Grassberger-Procaccia correlation dimension algorithm. The analysis yields correlation dimension values ranging anywhere from 8 to 11 depending on the region, suggesting that the arsenic contamination in space, from a chaotic dynamic perspective, is a medium- to high-dimensional problem. The dimension results also indicate that the spatial dynamics of arsenic may be moderately sensitive to geology, with Pleistocene aquifers appearing to require a minimum of about two less dominant processes/variables for its description when compared to that required by the Holocene aquifers. Based on these results, a qualitative discussion is also cast on the potential opportunities offered by a nonlinear deterministic and chaotic dynamic approach towards improving cost-effectiveness in siting new safe wells.

Keywords Groundwater contamination · Bangladesh · Shallow tube wells · Nonlinear dynamics · Regional geology · Remediation drilling · Cost-effectiveness

F. Hossain (✉)
Department of Civil and Environmental Engineering,
Tennessee Technological University, Box 5015,
Cookeville, TN 38505-0001, USA
E-mail: fhossain@tntech.edu
Tel.: +1-931-3723257
Fax: +1-931-3726239

B. Sivakumar
Department of Land, Air and Water Resources,
University of California, Davis, CA 95616, USA

Introduction

The health risk posed by dissolved arsenic in groundwater has been reported in many countries around the world. Concentrations exceeding the World Health Organization's (WHO) safe limit of 10 parts per billion (ppb) have been found in, among others, Bangladesh (Karim 2000), West Bengal in India (Mazumder et al. 1998); Taiwan (Tseng et al. 1968); Vietnam (Berg et al. 2001); Mexico (Del Razo et al. 1990) and regions of the United States (Welch et al. 2000). However, in terms of relative proportion of population at risk, arsenic contamination in Bangladesh represents a major calamity in modern history. It is estimated that about 80% of the population in Bangladesh (about 103 million) depend on shallow tube wells that have been excavated at a depth of less than 150 m (hereafter called 'shallow wells') (Ahmed 2002, 2003). An exposure distribution study by Yu et al. (2003) predicts that more than one million-per-year cases of arsenic-induced ailments are likely to evolve in the near future.

Since 1993, when it was first discovered that the alluvial Ganges aquifers of Bangladesh were contaminated with arsenic, numerous studies have been conducted to better understand the contamination scenario (e.g., Biswas et al. 1998; Burgess et al. 2000; Bhattacharya et al. 2002; Mukherjee and Bhattacharya 2002; Harvey et al. 2002; van Geen et al. 2002; Meharg and Rahman 2003; Yu et al. 2003). These studies indicate that the arsenic contamination is mostly unique to shallow wells where both the WHO limit and the Bangladesh limit (50 ppb) are consistently exceeded up to a depth of 150 m. The first countrywide study towards accurate spatial (horizontal) characterization of the calamity was conducted in 1998 by the British Geological Survey (BGS) in collaboration with the Department of Public Health and Engineering (DPHE) of Bangladesh (hereafter, this survey is referred to as 'BGS-DPHE'). This survey revealed that 46% of shallow wells exceeded the WHO safe limit, while about 27% exceeded the Bangladesh limit.

There have been other studies as well that have attempted a spatial description of arsenic contamination by either alluding to the BGS-DPHE (2001) survey for benchmarking and/or conducting independent small-scale surveys. Notable examples include McArthur et al. (2001), van Geen et al. (2003a), Yu et al. (2003), and Hossain et al. (2005). Central in all these studies addressing the 'spatial' character of arsenic is the use of classical geostatistical tools. For example, the BGS-DPHE (2001) study reports the application of geostatistics involving three steps and the assumption that the arsenic concentration could be treated as a 'regionalized' random variable in space. These three steps are: (1) computation and modeling of the variogram; (2) prediction of concentrations at nonsampled locations by kriging; and (3) statistical analysis of errors. Yu et al. (2003) used a variogram analysis to characterize the spatial variability of arsenic at three spatial scales (1 km, 10 km and 100 km) nationwide. The study by van Geen et al. (2003a) employed simple classical error statistics to quantify the vertical (depth) aspect of arsenic variability over a 25 km^2 area with high-resolution measurements (6,000 wells). More recently, Hossain et al. (2005) have also applied the variogram method to quantify the spatial variability of arsenic as a function of geology. Knowledge of anisotropy due to geology was also used therein to understand the implications for enhancing the cost-effectiveness of remediation drilling of safe wells on a regional basis in Western Bangladesh.

While there is no structural, or even philosophical, flaw in using the conventional geostatistical approach, there is indeed ample room to argue that the geostatistical treatment of arsenic contamination in space as a regionalized random (or stochastic) variable may constitute only an incomplete analysis of its spatial variability (even if system-dependent). Incompleteness can potentially arise from the fact that geostatistics often fails to recognize the random looking but deterministic behavior (hereafter interchanged with 'chaotic behavior') that may be present due to self-similar (scale-invariant) factors in the continuum of the sub-surface. For example, it is generally accepted that arsenic in groundwaters of Bangladesh is geologic in origin, deriving from the sediments transported from the upland Himalayan catchments (BGS-DPHE 2001; McArthur et al. 2001; Yu et al. 2003). The BGS-DPHE study clearly indicated that, contrary to the purely random phenomenon observed at the village-scale (<5 km), there exists distinct spatial averages of arsenic contamination, as indicated by geostatistics, in the regional scale that is 50–100 km scale at which geologic characteristics vary in Bangladesh (see Fig. 1). The association of low levels of arsenic is found in relatively oxic, uplifted old Pleistocene aquifers, and high arsenic concentrations in reducing young Holocene aquifers (Nickson et al. 1998). Most (but not all) of the Pleistocene deposits are located in the Northern region (comprising Madhupur clay, Barind clay and Alluvial fan deposits), while the majority of young Holocene deposits are located in the floodplains in the South (Deltaic deposits and Alluvial deposits) (see Fig. 1; Alam et al. 1990). Because Bangladesh is essentially a riverine (and dendritic) country with numerous 'small' floodplains, and further because the geology shows presence of pockets of Holocene-like and Pleistocene-like deposits scattered throughout the country (see Fig. 1), there is adequate reason to anticipate chaotic behavior in the spatial pattern of arsenic contamination.

However, a more physical argument in favor of expecting deterministic chaos in the spatial variation of arsenic can be argued as follows. Despite the apparently 'random' variability observed in the spatial structure of arsenic contamination (magnified further at scales smaller than 5 km; see BGS-DPHE 2001 and Yu et al. 2003 for details), field studies so far indicate evidence in support of a limited number of competing theories/hypotheses behind the mobilization of arsenic (Burgess et al. 2000, 2002; McArthur et al. 2001; Harvey et al. 2002; van Geen et al. 2003b). Each of these theories can, in principle, be mathematically represented as the cumulative effect of a finite number of dominant processes modeled by 3 or more partial differential equations (note: a minimum of 3 PDEs is required for a deterministic system to exhibit chaotic behavior; Hao 1984). As an example, the theory put forward by Harvey et al. (2002) states that groundwater arsenic may have increased as a result of increased water withdrawal for irrigation. The three core (necessary but not sufficient) processes that make up this theory are: (1) groundwater extraction by irrigation during winter/nonrainy season—a process of porous media flow; (2) recharge of groundwater by rainwater, carrying along surface organic matter to the subsurface aquifer—a process of infiltration (Richards Equation); and (3) microbial activity in the aquifer zone leading to reduction in conditions for arsenic mobilization—a process involving microbial kinetics and diffusion. It may be possible to construct similar lines of argument for other competing theories/hypotheses to argue that a simple physically based arsenic mobilization model (with finite degrees of freedom) can produce apparently 'random' spatial patterns of arsenic contamination in Bangladesh.

Taking note of the potential limitations of the geostatistical approaches and the possible nonlinear and chaotic nature of groundwater flow and transport phenomena, recent studies have also suggested consideration of an alternate (non-geostatistical) paradigm for analysis of groundwater resources. For example, in a review of research on nonlinear deterministic dynamics

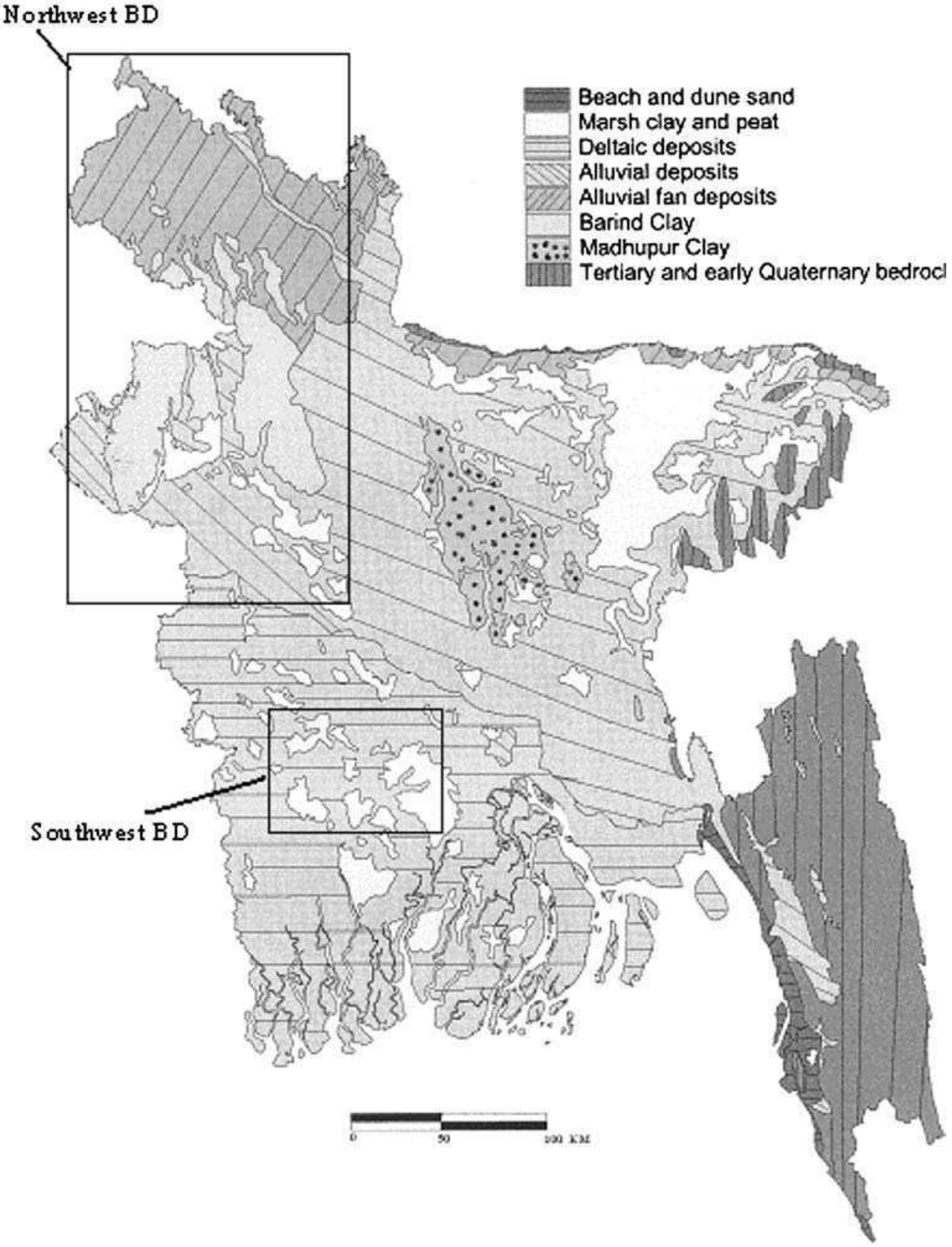


Fig. 1 Geology of Bangladesh (after Alam et al. 1990)

in porous media flow, Faybishenko (2004) noted the following relevant observation: “For many years the general approach to flow investigations in a fractured environment has been based on using stochastic methods to describe random-looking data sets (e.g., Gelhar 1993), without considering that deterministic chaotic processes could cause apparent randomness of experimentally observed data.” Similar concerns on the use of purely stochastic methods have been echoed by a few other studies as well (e.g. Faybishenko 2002; Sivakumar 2004; Sivakumar et al. 2005), which have indicated the potential of nonlinear deterministic approach either independently or in combination with a stochastic approach.

Granted that the categorical absence of chaotic behavior in groundwater flow and contamination phenomenon cannot therefore be theoretically established, the fundamental question that this study seeks to answer is as follows: What is the degree of nonlinear and chaotic behavior observed in the spatial pattern of arsenic contamination of shallow wells in Bangladesh? The study is motivated by the argument that the traditional approach of using geostatistics may be inadequate in the context of cost-effective solutions (e.g., remediation drilling) for a resource-poor country like Bangladesh. Traditional geostatistical methods, such as kriging, only solve the pattern completion problem (i.e., spatial interpolation), but not the (complementary) pattern recognition problem (e.g., fractals and chaos). As a result, the 'field' of arsenic estimated in this fashion from a finite amount of field information is subject to uncertainty due to measurement and sampling errors of in situ arsenic tests. On the other hand, a pattern recognition method seeks to associate the sampled field with one or more describable 'memories' (e.g., similar to recognizing a letter from a hand-written text) and can be analogous to a nonstochastic approach, such as the nonlinear deterministic dynamic (chaotic) approach.

Our investigation of nonlinear deterministic and chaotic behavior is, however, not directed towards replacement of the conventional geostatistical characterization techniques, but rather to eventually strengthen them by proposing a synergistic use that minimizes the individual limitations. Our study represents only a preliminary exploration of chaotic behavior, and we intend to employ in the future appropriately more sophisticated methods, in the spirit of increasing cost-effectiveness of remediation solutions for Bangladesh. The study is based on data from the BGS-DPHE (2001) survey and makes geologic distinctions (Holocene vs. Pleistocene; Northwest vs. Southwest) to characterize the geologic dependency of the chaotic property. Results from this study are eventually expected to initiate exploration on the usefulness of a nonlinear deterministic and chaotic approach in complementing the purely geostatistical approach, and, more specifically, to provide a conceptual framework to improve the problem definition for questions, such as: (1) what are the implications of the chaotic property in improving the cost-effectiveness of remediation drilling? and (2) How can guidelines based on conventional geostatistical approach be improved for a more effective water resources strategy in Bangladesh?

The paper is organized as follows. Section 2 presents the study region and dataset, while Sect. 3 describes the correlation dimension method used for identification of chaos. In Sect. 4, we discuss the results and the implications for more cost-effective remediation strategies vis-à-vis conventional geostatistical approaches. Finally, Sect. 5 presents the conclusions and recommendations for field-scale investigations to explore further the merit of the chaotic approach.

Study region and data

We choose to study the entire region of Bangladesh as had been first surveyed by the BGS-DPHE (2001) study comprising 3534 wells (see Fig. 1). The dataset is available at http://www.bgs.ac.uk/arsenic/Bangladesh.html. The wells deeper than 150 m (and consistently below the safe limits) are excluded from the analysis, thus resulting in a set of 3,085 shallow wells. This further implies that the vertical variability of arsenic concentration is insignificant compared to its horizontal variability and hence will have negligible effect on the chaotic analysis conducted herein. Sample wells are systematically and uniformly selected with approximately one well per 37 km^2 ($\sim$ 6 km×6 km). The arsenic measurements of the BGS-DPHE (2001) survey were based on Atomic Absorption Spectro-photometry (AAS), which is currently considered the most reliable technique for benchmarking arsenic measurements (Rahman et al. 2002). We assume a minimum detection limit of arsenic concentration as 1 ppb and, hence, all nondetection wells are assigned a value of 1 ppb. The advantage of this adjustment is that it preserves the sanity of log-transformation of data that is considered a necessary element, as arsenic concentrations are known to vary over 3–4 orders of magnitude in Bangladesh (Yu et al. 2003; Hossain et al. 2004). It must be noted that, due to the spatial resolution of the BGS-DPHE (2001) survey, the study is limited to the scale of about 6–7 km (also note that this is the scale at which villages are clustered under the smallest administrative unit called a 'Union'). As with any type of field investigation, certain limitations (such as inaccessibility of sampling locations and local lack of familiarity with randomized sampling) existed with this BGS-DPHE survey as well. However, in the overall scheme of our investigations, such limitations are considered insignificant due to the fact that the BGS-DPHE survey currently represents the most quality-controlled database of arsenic measurements available countrywide.

To study the role played by geology, we further classified our arsenic database into three categories: (A) Whole Bangladesh (no distinction made in geology); (B) Holocene deposits of Southwest Bangladesh (BD) (geologic distinction—those regions usually high in arsenic); and (C) Pleistocene deposits of Northwest BD (geologic distinction—those regions usually low in arsenic). We first defined Western BD as the region west of the *Brahmaputra-Meghna* River system (see Fig. 1). The region is then geographically subdivided into two parts based on major Holocene/Pleistocene differences reported by Alam et al. (1990): (B) Southwest BD; and (C) Northwest BD. Hereafter regions A, B and C shall be conveniently interchanged with 'Whole BD', 'Southwest BD (or Holocene deposits)' and 'Northwest BD (or Pleistocene deposits)', respectively. The Northwest BD is defined as the region bound by 24.0°N–26.7°N latitude and 88.0°E–89.5°E longitude representing an area of about 35,000 km^2 with 872 shallow wells (Fig. 1). The Southwest BD is bound by 22.49°N–23.79°N latitude and 89.0°E–90.0°E longitude representing an exclusively Holocene area of about 13,000 km^2 with 848 shallow wells.

The chaotic approach

Many methods have been formulated for the identification of chaotic behavior in a data series. One such method used herein is the 'Correlation Dimension' (CD) method, which attempts to measure the extent to which the presence of a data point affects the position of the other points on the attractor. The concept is analogous to the classical notion of auto-covariance function, with the exception that the dependency of a point in the series (in the continuum of space or time) is cast from the perspective of nonlinear determinism exhibited by a local attractor. The CD method uses the correlation integral or function (Grassberger and Procaccia 1983) for distinguishing between chaotic and stochastic behavior (more specifically, between low-dimensional and high-dimensional systems). The concept of the correlation integral is that even when a process may look irregular (i.e., 'random'), if it comes from deterministic dynamics, it will have a limited number of degrees of freedom equal to the smallest number of first-order differential equations that capture the most important features of the dynamics. For further details on the relevant issues in the application of CD in hydrology and related fields, the reader is referred to the studies by Tsonis et al. (1993), Sivakumar (2000, 2005), and Sivakumar et al. (2002a, b), among others.

It is appropriate to mention, at this stage, that there is a fundamental difference in the manner in which the Grassberger–Procaccia algorithm is employed in the present study when compared to its application in the past. Traditionally, the algorithm has been applied to data series in the continuum of time (e.g. Theiler 1987), whereas herein it is applied to data series in space. While this deviation may raise concerns, which are understandable, we would also like to emphasize that there is no convincing reason to believe that the algorithm cannot be used in the space domain, even involving unequal delay distances. We admit that the phase-space reconstruction for 'irregular-interval data' (regardless of time or space) may have its own limitations, but we believe that such potential limitations alone should not hamper our ability to investigate the usefulness of the algorithm, and this is particularly so when our intention is to primarily conduct a preliminary exploration in a spatial context. In addition, there are two caveats of the Grassberger–Procaccia algorithm that the reader should be forewarned of. The first is causality—there is no reason to expect that causality will hold for spatial series. The second is that there are 1–3 independent variables, rather than 1, affecting arsenic variability in space. Using just the distance as the independent variable, rather than 2 spatial coordinates, implies some type of isotropy in the spatial pattern and may bias results. We believe that the weaknesses of this algorithm, if any, may be revealed in our results, and consequently, we may also employ a more appropriate phase-space reconstruction method in the future.

With the above limitations in mind, each selected well is therefore considered a focal point, and the intra-well distances between all other wells are computed. The arsenic concentrations of each well with respect to the focal well are arranged in the order of increasing intra-well distance. Thus, in essence, Region A (Whole BD) comprises of 3,085 spatial series, while Region B and C have 848 and 872 spatial series, respectively, each with the same number of data points. The algorithm uses the phase-space reconstruction of these spatial series. For a scalar spatial series X_i, where $i = 1, 2, 3, \ldots, N$, (and X_i is the arsenic concentration at well i), the phase-space can be constructed using the method of delays (distances) given by,

$$\boldsymbol{Y}_j = (X_j, X_{j+\tau}, X_{j+2\tau}, \ldots, X_{j+(m-1)\tau/\Delta s}), \tag{1}$$

where $j = 1, 2, \ldots, N - (m-1)\tau/\Delta s$; m is the dimension of the vector $\boldsymbol{Y}_j$, also called the embedding dimension; and τ is the delay distance taken to be some suitable multiple of the average intra-well distances Δs. For an m-dimensional phase-space, the correlation integral $C(r)$ is given by (Theiler 1987),

$$C(r) = \lim_{(N\to\infty)} \frac{2}{N(N-1)} \sum_{i,j} H\left(r - \left|Y_i - Y_j\right|\right). \tag{2}$$

Here, $1 \le i < j \le N$; H is the Heaviside step function with $H(u) = 1$ for $u > 0$ and $H(u) = 0$ for $u \le 0$, where $u = r - |\boldsymbol{Y}_i - \boldsymbol{Y}_j|$, and r is the radius of sphere centered on $\boldsymbol{Y}_i$ or $\boldsymbol{Y}_j$; and N is the number of data points (wells) in the spatial series.

If the spatial series is characterized by an attractor, then for positive values of r, the correlation integral $C(r)$ is related to the radius r by the following relation:

$$C(r) \approx \alpha r^{v}, \tag{3}$$

where α is constant; and v is the correlation exponent or the slope of Log $C(r)$ versus Log r plot given by:

$$v = \lim_{(r\to 0, N\to\infty)} \frac{\operatorname{Log} C(r)}{\operatorname{Log} r}. \tag{4}$$

The slope is generally estimated by a least-squares fit of a straight line over a certain range r, called the scaling region.

To observe whether or not a chaotic pattern exists in the spatial property of arsenic, the correlation exponent values are plotted against the corresponding embedding dimension values. If the correlation exponent leads to a finite value (i.e., saturation of slopes), then the system is often considered dominated by nonlinear deterministic and chaotic dynamics. If the value of the correlation exponent is small, then the system is generally thought as dominated by a low-dimensional dynamics (spatial) governed by the properties of an attractor. The saturation value of the correlation exponent is defined as the correlation dimension of the attractor of the spatial series. In contrast, for systems dominated by stochastic processes, the correlation exponent is supposed to increase without any bound. While this type of interpretation is generally accepted for distinguishing between chaotic and stochastic behaviors, there may also be certain exceptions, since finite correlation dimensions may also result for stochastic systems with power-law spectra.

Some studies have questioned the studies that reported existence of chaos in hydrologic data, but such have been raised mostly on the ground of data size (e.g. Schertzer et al. 2002). While the correlation dimension method may indeed possess certain limitations, any claim and counterclaim on the presence/absence of chaos in a time series needs careful interpretation [see, for instance, Sivakumar (2000, 2005) and Sivakumar et al. (2002a, b) for details].

The CD of an attractor provides information on the dimension of the phase-space required for embedding the attractor, which, in turn, provides information on the number of variables present in the spatial pattern of the corresponding arsenic-contaminated hydro-system. According to Fraedrich (1986), the nearest integer above the correlation dimension value provides the minimum dimension of the phase-space essential to embed the attractor, while the value of the embedding dimension at which the saturation of the correlation exponent occurs provides an upper bound on the dimension of the phase-space sufficient to describe the motion of the attractor.

Results and discussion

Correlation dimension analysis

Preliminary results of correlation dimension analysis reveal insignificant differences among the closely clustered wells. This is expected, since only the overall variability of the spatial series is reflected in the CD analysis, rather than that between individual and closely spaced wells. Hence, for convenience, we demonstrate the CD analysis for a finite number of focal wells for each region (A, B and C). We select 10 wells for each of regions A, B and C. Table 1 summarizes the location of each focal well (note: focal wells were also included in the CD analyses), its depth and corresponding arsenic concentration. Figure 2 shows the relationship between correlation function $C(r)$ and radius r for the focal well A-1 (see Table 1) in region A (no geologic distinction). Large scaling regions are observed in Fig. 2, which allow us fairly reasonable estimations of the correlation exponent. In Figs. 3 and 4, we show similar relationships for other focal wells, B-1 and C-1 for regions B (Holo-

Table 1 Description of the 10 selected focal wells for regions A, B and C

Focal well	Location (°)		Depth (m)	Arsenic concentration (ppb)	Comment
	Latitude	Longitude			
Whole BD (Region A)					
A-1	24.303	91.450	35.0	18.7	
A-2	23.851	88.654	39.0	58.4	
A-3	24.547	88.608	38.0	4.6	
A-4	24.211	89.419	39.0	1.0	Nondetection well
A-5	24.151	89.280	34.0	1.0	Nondetection well
A-6	22.709	89.635	23.0	15.0	
A-7	23.018	89.139	34.0	224.0	
A-8	24.336	88.749	33.0	1.0	Nondetection well
A-9	24.054	89.373	45.0	1.0	Nondetection well
A-10	22.558	89.007	35.0	192.0	
Holocene (Region B)					
B-1	23.607	90.991			
B-2	23.946	90.115	55.0	29.9	
B-3	22.488	90.067	22.0	6.0	
B-4	22.629	89.610	22.0	424.0	
B-5	22.901	88.949	46.0	88.0	
B-6	22.704	89.688	15.0	234.0	
B-7	23.542	90.607	13.0	120.0	
B-8	23.144	89.763	39.0	76.0	
B-9	22.751	89.714	16.0	571.0	
B-10	23.368	89.556	57.0	1.0	Nondetection well
Pleistocene (Region C)					
C-1	23.968		89.830		
C-2	25.776	88.565	19.5	1.0	Nondetection well
C-3	25.855	88.539	18.9	5.4	
C-4	25.639	88.671	32.6	5.6	
C-5	24.811	89.485	22.9	8.6	
C-6	24.292	89.313	38.0	1.0	Nondetection well
C-7	24.418	89.001	41.0	11.6	
C-8	24.395	89.057	39.0	1.0	Nondetection well
C-9	24.809	88.935	30.0	1.0	Nondetection well
C-10	24.814	88.881	39.0	3.2	

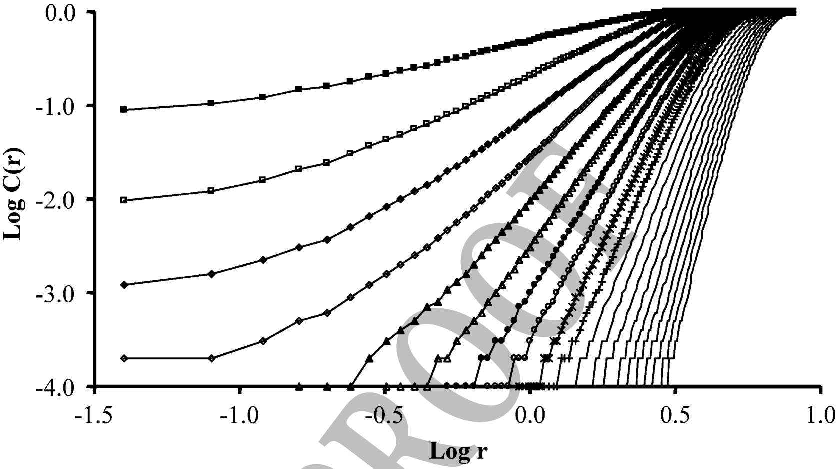


Fig. 2 The relationship between Log *C(r)* and radius (Log *r*) for focal A-1 (region A)

cene) and C (Pleistocene), respectively. However, distinctions due to geologic property are hard to establish qualitatively in these figures. Since the Log *C(r)* versus Log (*r*) analyses from all other focal wells appear very similar, they are not shown herein.

To analyze the relationship between correlation exponent and embedding dimension, we show Figs. 5, 6 and 7 for regions A, B and C, respectively. It must be noted, however, that the exact delineation of the scaling region (from Figs. 2, 3 and 4, for example) can be a difficult task often requiring semiquantitative methods such as visual inspection. Hence, the subsequent estimation of correlation exponent with respect to embedding dimensions should be assessed as an empirical exercise subject to the limitations of the semiquantitative method employed herein. A saturation of the slope Log *C(r)*/Log *(r)* is observed, indicating evidence towards possible nonlinear deterministic and chaotic dynamic behavior. This saturation value of the correlation exponent (also known as Correlation Dimension, CD) appears to show moderate variability across geologic property. When no distinction is made in terms of geology, the nationwide CD value appears to lie in the ranges of 10–11 while the embedding dimension at which this saturation occurs is found to be about 12 (see Fig. 5). We take this finding as a preliminary indication of the medium-to-high level of dimensionality that exists in the mobilization mechanisms of arsenic. Consequently, it is an indication that a medium-to-large model structure (perhaps with 10–12 model parameters) is required to adequately capture this spatial variability at the regionalized scale (~ 6 km×6 km). Across geologic regions, Pleistocene deposits appear to require a minimum of two less variables/parameters for its spatial description than their Holocene counterpart. Also note that the CD for Pleistocene region (Northwest BD) ranges around 8–9, while Holocene deposits correspond much closer to the nationwide value of 10–11. At this stage, we speculate that this difference is perhaps due to the predominance of oxic conditions in the Pleistocene shallow wells that imply the absence of one or more arsenic mobilization mechanisms (i.e., no reduction mechanism). It is currently unknown as to how these saturation values will vary as higher resolution arsenic measurements (scales < 6 km×6 km) become available along with more up-to-date groundwater chemistry data. Hence, we stress the need for more detailed investigation involving higher resolution data to quan-

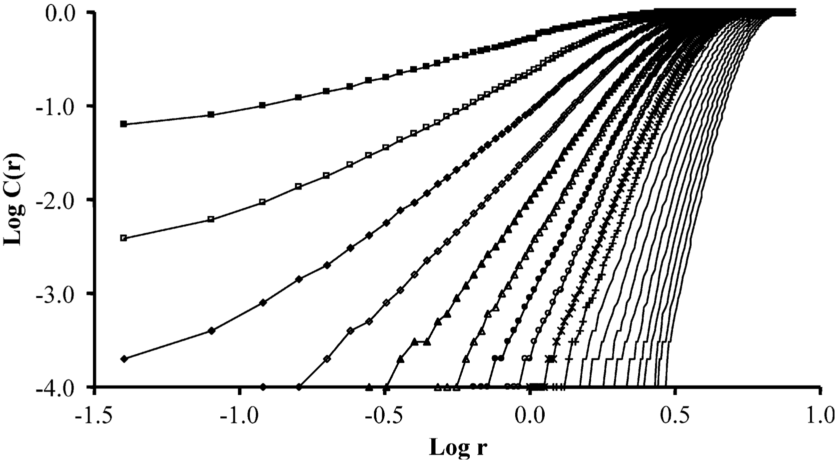


Fig. 3 Same as Fig. 3 but for focal well B-1 (region B)

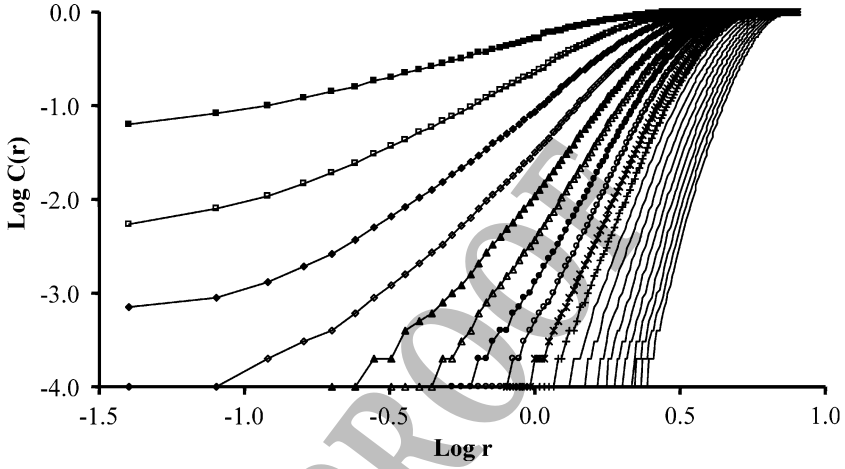


Fig. 4 Same as Fig. 3 but for focal well C-1 (region C)

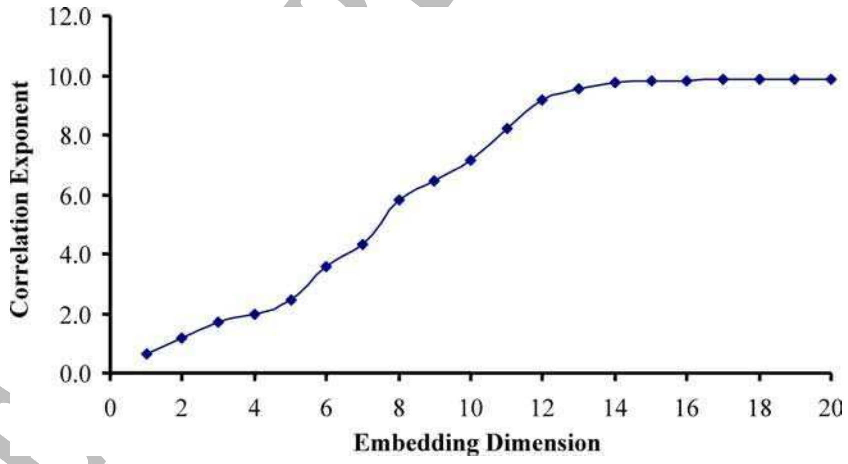


Fig. 5 Relationship between correlation exponent and embedding dimension for Region A (no geologic distinction) (focal well A-1)

tify more definitively the forcing role played by geology. In Table 2, we summarize the mean CD values for the selected 10 focal points over each region.

It is appropriate, at this stage, to highlight spatial variability analyses recently reported by Hossain et al. (2005) on the basis of a purely geostatistical framework (variogram analyses). Hossain et al. (2005) reported that regional anisotropy in the spatial dependence of arsenic for Northwest region of Bangladesh was found to be stronger than that in the Southwest. The correlation length for arsenic concentration in the East–West direction of Northwest Bangladesh (i.e., across major river floodplains, Fig. 1) was found to be almost twice (158.80 km) that of the North–South direction (along the major axis of Pleistocene deposits) (78.21 km). For the Southwest region, the ratio of East–West to North–South correlation lengths ranged from 1.40 to 1.51.

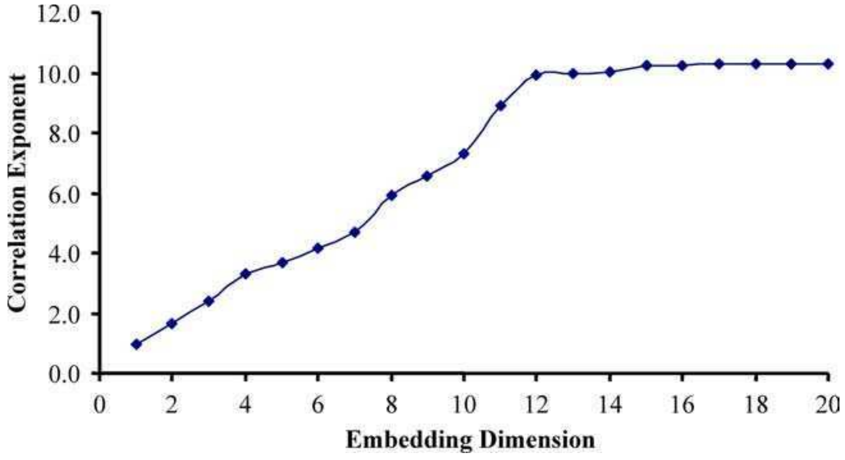
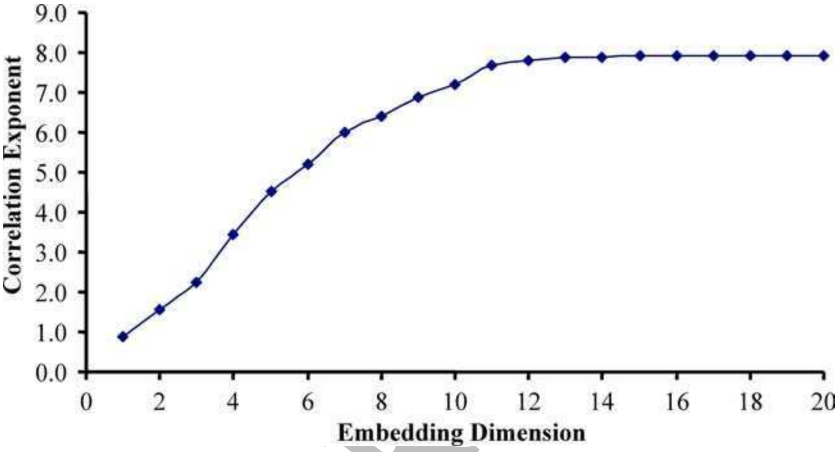


Fig. 6 Same as Fig. 6 but for region B (focal well B-1)

Fig. 7 Same as Fig. 6 but for Region C (focal well, C-1)



Implications of chaotic behavior for cost-effective siting of remediation wells

For siting of new (safe) wells, it is currently considered a very expensive, and even almost-impossible, task to reconcile the need for detailed knowledge of hydrogeology with the tremendous spatial variability (centimeters vertically; meters horizontally) that is observed in the Bangladesh sediments. Nevertheless, detailed field studies can always lead to some useful ‘rules of thumb’ for well-siting at the village level. The indications of such an approach can be found in Hossain et al. (2004). However, initial experiences of the BGS-DPHE (2001, see page 240) over a small study region have exemplified the difficulty in accurately quantifying the likelihood of identifying ‘uncontaminated wells’ from a geostatically derived spatial map of arsenic contamination. A major inadequacy of geostatistical approaches, in our opinion, arises from the observation that a geostatistically identified (i.e., ‘kriged’) drilling location of a remediation well in the neighborhood of a contaminated well is associated with an uncertainty (equivalent to the minimized error variance). In Bangladesh, where public health is at risk, such uncertainty needs to be reduced and, hence, conventional geostatistical approaches (such as kriging) should be explored for ‘enhancement’ with the less conventional paradigms such as the nonlinear deterministic chaos theory to minimize the limitations.

Let us now consider, as an example, the recent report on implementing a safe water supply program for Srinagar—a sub-district of Bangladesh comprising about 30,000 (mostly rural) people (Hoque et al. 2004). Experience from this program clearly recommended that, for arsenic-affected areas, a cluster-based piped water system be given proper consideration when selecting appropriate water options rather than household-based options or the development of new low-cost options (e.g. filters, rain-water harvesting etc.) (Hoque et al. 2004). To effectively construct a cluster-based piped water system, a network of wells would need to be drilled and centrally connected through a piped network in a fashion that the overall quality of water being pumped is within the Bangladesh safe limit of 50 ppb for arsenic. This requires drilling the maximum number of safe wells, although a few unsafe wells may also be acceptable. For cost-effective implementation of this program, there are now two competing aspects interacting at the level of risk management and decision-making: (1) alternative but safe drinking water supply needs to be ensured rapidly; and (2) remediation drilling should be preferably shallow (given that deep-well drilling is a time-consuming and very costly option) and planned in surrounding locations that are known to be ‘probabilistically’ low in arsenic.

Under such a situation, a chaotic approach has the potential to augment the geostatistical approach. The chaotic approach can identify the minimum number of model (physical or black-box) parameters required to adequately model the spatial variability of arsenic in space for the affected region that has already been mapped for its CD value. One suggested example is the use of a neural-network (NN) model with the number of nodes equaling the number of variables (embedding dimensions) obtained from the CD analysis that spatially ‘forecasts’ safe zones. The inputs to the NN-chaotic model would be the arsenic concentration of the focal well (point of application) and the distance from the well (i.e., point of prediction). This NN-chaotic model can then be used to identify the spots that are likely to be safe in arsenic concentration (in recognition of the self-similarity) and subsequently used as an additional constraint to the geostatistical siting of wells

Table 2 Summary of correlation dimension value for regions A, B and C

	Correlation dimension (Range)	
	Min	Max
Region A	10	11
Region B (Holocene)	10	11
Region C (Pleistocene)	8	9

(i.e., both kriging and the NN-chaotic model should agree on the location's safe/unsafe likelihood). Despite the traditional criticism of neural-network approaches (or any 'time series' approach, for that matter) as lacking potential to promote the scientific insight of a natural phenomenon, recent work indicates that substantial understanding has been gained in designing its architecture (nodes and hidden layers) based on the physical process it models [see for example: Sudheer and Jain (2003) and Jain et al. (2004) for river flow].

Currently, there are a number of maps available that characterize the probability of arsenic contamination in nonsampled regions based on kriging (see BGS-DPHE 2001 and McArthur et al. 2001, for example). Hence, a combination of geostatistics with a chaotic dynamic approach can intuitively be expected to refine the safe drilling spots with considerably lower probability of failure (i.e., by reducing the number of false hopes and false alarms) than what would have otherwise been indicated by kriging alone. Although it remains to be seen if there exists a (suggested) physical connection between the minimum number of NN nodes and the minimum embedding dimension from the CD method at the region of application [see, however, Sivakumar et al. (2002) for such a connection for river flow], the suggestion, which has insignificant start-up costs compared to laboratory studies, is certainly worth investigating until proven wholly ineffective.

Conclusion

This paper presented a preliminary investigation of the existence of nonlinear deterministic and chaotic behavior in arsenic contamination in shallow tube wells of Bangladesh. The correlation dimension method revealed convincing medium-to-high dimensional chaotic pattern with a countrywide dimension value ranging between 8 and 11. The minimum number of variables and, hence, the number of dominant processes required to model the spatial variability of arsenic were also identified. It appeared that chaotic behavior of arsenic is moderately sensitive to geology (Holocene vs. Pleistocene). The study indicated that Pleistocene aquifers would require two less minimum number of variables for its spatial description compared to the Holocene counterpart. However, higher resolution data may be required to explore this issue further. Finally, the paper discussed the potential opportunities offered by the chaotic approach towards better cost-effective remediation strategies than that possible by a purely geostatistical framework.

As part of a proposed (and needed) 'enhancement' of geostatistical approaches for spatial characterization towards increasing cost-effectiveness, we list the following as natural extensions to this study: (1) Investigate the scale-invariant or fractal behavior of arsenic with high-resolution data, with due consideration for the possible influence of (small) data size, thresholds and presence of zeros (or any other single value) on the outcomes of fractal/chaos methods (e.g., Harris et al. 1997; Sivakumar 2001); (2) Consider the effects of depth of wells on chaotic characterization and explore fuzzy logic approaches of membership functions (e.g., Klir and Yuan 1995) to categorize wells as deep or shallow; and (3) Study the impact of measurement error on the identification of chaotic behavior [see, for example, Sivakumar et al. (1999)] using the semiquantitative field kits (Rahman et al. 2002).

Among the suggested extensions, (3) has particular significance to increase cost-effectiveness of (and reducing risks for) siting of safe shallow wells through our proposed paradigmatic approach (geostatistical-cum-chaotic). The majority of arsenic measurements in Bangladesh on a large scale (totalling about 1.3 million wells) are available from the cheap Field Kits (FK) (Rahman et al. 2002). Although FK measurements are subject to large errors with a highly complex error structure, Hossain et al. (2005) have demonstrated that FK measurements can still adequately characterize the modal depth (depth of highest arsenic contamination) for shallow wells. This probably indicates that the database of FK-arsenic measurements that are currently available countrywide may indeed be sufficient to characterize the probable depth of highest arsenic contamination for local regions, even though the modal arsenic contamination rate (i.e., "highest fraction of contaminated wells") may have large errors. Hence, one current challenge is to investigate the efficacy of using FKs for mapping the chaotic behavior of arsenic and (otherwise) propose new desired performance levels for FKs that would make them as useful as AAS method in our proposed geostatistical-cum-chaotic siting of new wells (see Sect. 4B). Work is ongoing in this direction, and we hope to report the details in the near future.

References

Ahmed MF (2002) Arsenic mitigation in Bangladesh. In: International workshop in Bangladesh, Dhaka, 14–16 January. Published by: ITN- Bangladesh, Centre for Water Supply and Waste Management, BUET, Dhaka-1000, Bangladesh

Ahmed MF (2003) Arsenic contamination: Bangladesh perspective. ITN, BUET, Bangladesh, Dhaka

Alam MK, Hassan A, Khan M, Whitney JW (1990) Geological map of Bangladesh. Geological Survey of Bangladesh

Berg M, Tran HC, Nguyen TC, Pham HV, Schertenleib R, Giger W (2001) Arsenic contamination of ground water and drinking water in Vietnam: a human health threat. Environ Sci Technol 35(13):2621–2626

BGS-DPHE (2001) Arsenic contamination of ground water in Bangladesh. In: Kinnburgh DG, Smedley PL, (eds) British Geological Survey Report WC/00/19, vol 1–4, British Geological Survey, Keyworth, UK (available at: http://www.bgs.ac.uk/arsenic/Bangladesh)

Bhattacharya P, Jacks G, Ahmed KM, Routh J, Khan AA (2002) Arsenic in ground water of the Bengal Delta plain aquifers in Bangladesh. Bull Environ Contam Toxicol 60:538–545

Biswas BK, Dhar RK, Samantha G, Mandal BK, Chakraborti D, Faruk I, Islam KS, Chowdury M, Islam A, Roy S (1998) Detailed study report of Samta, one of the arsenic-affected villages of Jessore District, Bangladesh. Curr Sci 74:134–145

Burgess WG, Burren M, Perrin J, Ahmed KM (2000) Constraints on sustainable development of arsenic-bearing aquifers in southern Bangladesh. Part 1: A conceptual model of arsenic in the aquifer. In: Hiscock, Rivett, Davison (eds) Sustainable ground water development. Geological Society of London Special Publication 193:145–163

Del Razo LM, Rellano MA, Cebrian ME (1990) The oxidation states arsenic in well-water from a chronic arsenicism area of Mexico. Environ Pollut 64:143–153

Faybishenko B (2002) Chaotic dynamics in flow through unsaturated fractured media. Adv Wat Resour 25(7):793–816

Faybishenko B (2004) Non-linear dynamics in flow through fractured porous media: status and perspectives. Rev Geophys 42:RG2003. DOI: 10.1029/2003RG000125

Fraedrich K (1986) Estimating the dimensions of weather and climate attractors. J Atmos Sci 43(5):419–132

van Geen A, Zheng Y, Vesteeg R, Stute M, Horneman A, Dhar R, Steckler M, Gelman A, Ahsan H, Graziano JH, Hussain I, Ahmed KM (2003a) Spatial variability of arsenic in 6000 tube wells in a $25\,\mathrm{km}^2$ area of Bangladesh. Wat Resour Res 39(5):1140. DOI:10.1029/2002/WR001617

Gelhar LW (1993) Stochastic subsurface hydrology. Prentice-Hall, Englewood Cliffs

Grasberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208

Hao B-L (1984) Chaos. World Scientific, Singapore

Harris D, Seed A, Menabde M, Austin G (1997) Factors affecting multiscaling analysis of rainfall time series. Nonlinear Proc Geophys 4:137–155

Harvey CF, Swartz CH, Badruzzaman ABM, Keon-Blute N, Yu W, Ali MA, Jay J, Beckie R, Niedan V, Brabander D, Oates PM, Ashfaque KN, Islam S, Hemond HF, Ahmed MF (2002) Arsenic mobility and ground water extraction in Bangladesh. Science 298:1602–1606

Hoque BA, Hoque MM, Ahmed T, Islam S, Azad AK, Ali N, Hossain M, Hossain MS (2004) Demand-based water options for arsenic mitigation: an experience from rural Bangladesh. Public Health 118:70–77

Hossain F, Bagtzoglou AC, Nahar N, Hossain MD (2005) Statistical characterization of arsenic contamination in shallow tube wells of western Bangladesh. Hydrol Proc (in Press)

Jain A, Sudheer KP, Srinivasulu S (2004) Identification of physical processes inherent in artificial neural network rainfall models. Hydrol Proc 18(3):571–581

Journel AG, Huijbregts CJ (1978) Mining geo-statistics. Academic, San Diego

Karim MD (2000) Arsenic in ground water and health problems in Bangladesh. Water Res 36(4):799–809

Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall, New Jersey

Mazumder GDN, Haque R, Ghosh N, De BK, Santra A, Chakraborti D, Smith AH (1998) Arsenic levels in drinking water and the prevalence if skin lesions in West Bengal, India. Int J Epidemiol 27(5):871–877

McArthur JM, Ravenscroft P, Safiullah S, Thirlwall MF (2001) Arsenic in ground water: testing pollution mechanisms for sedimentary aquifers in Bangladesh. Wat Resour Res 37(1):109–117

Meharg AA, Rahman MM (2003) Arsenic contamination of Bangladesh paddy fields. Environ Sci Technol 37:229–234

Mukherjee AB, Bhattacharya P (2002) Arsenic in ground water in the Bengal delta plain: slow poisoning in Bangladesh. Environ Rev 9:189–220

Nickson RT, McArthur JM, Burgess W, Ahmed KM, Ravenscroft P, Rahman M (1998) Arsenic poisoning of Bangladesh ground water. Nature 395:338

Osborne AR, Provenzale A (1989) Finite correlation dimension for stochastic systems with power-law spectra. Physica D 35:357–381

Rahman MM, Mukherjee D, Sengupta MN, Chowdury UK, Lodh D, Chanda CN, Roy S, Selim M, Quamruzzaman Q, Milton AH, Shadullah SM, Rahman MT, Chakraborti D (2002) Effectiveness and reliability of arsenic field testing kits: are the million dollar screening projects effective or not? Environ Sci Technol 36:5385–5394

Schertzer D, Tchiguirinskaia I, Lovejoy S, Hubert P, Bendjoudi H (2002) Which chaos in the rainfall-runoff process? A discussion on 'Evidence of chaos in the rainfall-runoff process' by Sivakumar et al. Hydrol Sci J 47(1):139–147

Sivakumar B (2000) Chaos theory in hydrology: important issues and interpretations. J Hydrol 227(1–4):1–20

Sivakumar B (2001) Rainfall dynamics at different temporal scales: a chaotic perspective. Hydrol Earth Syst Sci 5(4):645–651

Sivakumar B (2004) Chaos theory in geophysics: past, present and future. Chaos Sol Fract 19(2):441–462

Sivakumar B (2005) Correlation dimension estimation of hydrologic series and data size requirement: myth and reality. Hydrol Sci J 50(4):591–604

Sivakumar B, Phoon KK, Liong SY, Liaw CY (1999) A systematic approach to noise reduction in chaotic hydrological time series. J Hydrol 219(3/4):103–135

Sivakumar B, Berndtsson R, Olsson J, Jinno K (2002a) Reply to "Which chaos in the rainfall-runoff process" by Schertzer et al. Hydrol Sci J 47(1):149–158

Sivakumar B, Persson M, Berndtsson R, Uvo CB (2002b) Is correlation dimension a reliable indicator of low-dimensional chaos in short hydrological time series? Wat Resour Res 38(2). DOI: 10.1029/2001WR000333

Sivakumar B, Harter T, Zhang H (2005) Solute transport in a heterogeneous aquifer: a search for nonlinear deterministic dynamics. Nonlinear Process Geophys 12:211–218

Sudheer KP, Jain A (2003) Explaining the internal behavior of artificial neural network river flow model. Hydrol Proc 18(4):833–844

Theiler J (1987) Efficient algorithm for estimating the correlation dimension from a set of discrete points. Phys Rev A 36(9):4456–4462

Tseng T, Babazono A, Yamamoto E, Kurumatani N, Mino Y, Ogawa T, Kishi Y, Aoyama H (1968) Ingested arsenic and internal cancer in an endemic area of chronic arsenicism in Taiwan. J Natl Cancer Inst 40:453–463

Tsonis AA, Elsner JB, Georgakakos KP (1993) Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation. J Atmos Sci 50:2549–2555

Welch AH, Westjohn DB, Helsel DR, Wanty RB (2000) Arsenic in ground water of the United States: occurrence and geochemistry. Ground Water 38:589–604

Yu WH, Harvey CM, Harvey CF (2003) Arsenic ground water in Bangladesh: a geo-statistical and epidemiological framework for evaluating health effects and potential remedies. Wat Resour Res 39(6):1146. DOI:10.1029/2002WR001327