# Statistical Modeling of Global Geogenic Arsenic Contamination in Groundwater

MANOUCHEHR AMINI,
KARIM C. ABBASPOUR, MICHAEL BERG,
LENNY WINKEL, STEPHAN J. HUG,
EDUARD HOEHN, HONG YANG, AND
C. ANNETTE JOHNSON*

*Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland*

Contamination of groundwaters with geogenic arsenic poses a major health risk to millions of people. Although the main geochemical mechanisms of arsenic mobilization are well understood, the worldwide scale of affected regions is still unknown. In this study we used a large database of measured arsenic concentration in groundwaters (around 20,000 data points) from around the world as well as digital maps of physical characteristics such as soil, geology, climate, and elevation to model probability maps of global arsenic contamination. A novel rule-based statistical procedure was used to combine the physical data and expert knowledge to delineate two process regions for arsenic mobilization: "reducing" and "high-pH/oxidizing". Arsenic concentrations were modeled in each region using regression analysis and adaptive neuro-fuzzy inferencing followed by Latin hypercube sampling for uncertainty propagation to produce probability maps. The derived global arsenic models could benefit from more accurate geologic information and aquifer chemical/physical information. Using some proxy surface information, however, the models explained 77% of arsenic variation in reducing regions and 68% of arsenic variation in high-pH/oxidizing regions. The probability maps based on the above models correspond well with the known contaminated regions around the world and delineate new untested areas that have a high probability of arsenic contamination. Notable among these regions are South East and North West of China in Asia, Central Australia, New Zealand, Northern Afghanistan, and Northern Mali and Zambia in Africa.

## Introduction

In many regions around the world, arsenic-rich groundwater is used for drinking water and irrigation purposes (*1, 2*). Long-term exposure to arsenic can affect human health and is considered to be a significant environmental cause of cancer (*2, 3*). Exposure to arsenic may also cause skin pigmentation, hyperkeratosis, and cardiovascular disease, and may affect the mental development of children, among other possible adverse effects (*1, 4, 5*). Due to its high toxicity at low concentrations (in the $\mu$g L$^{-1}$ range), arsenic is a cause for concern in many regions of the world (*6–9*). However, the worldwide scale of affected regions is still unknown because groundwater quality has only been determined in certain regions. Hence, a global map, showing the areas with high probability of contamination is of great value for civil authorities and aid agencies in preventing the increase in consumption of arsenic-contaminated waters.

Solid-phase arsenic is widely distributed in nature and can be mobilized in groundwaters through a combination of natural processes and/or anthropogenic activities (*6–11*). At the global scale, natural processes are predominantly responsible for the elevated arsenic concentrations, although anthropogenic sources of arsenic, such as leaching from mine tailings, can be very important on the local scale. Regardless of the source of arsenic, its solubility is mostly controlled by pH and Eh (*6*), which in turn are governed by geology, climate, drainage, and topography (*7, 11, 12*). Two important chemical processes lead to arsenic mobilization through desorption and dissolution. These processes take place under two different physicochemical conditions: first, highly reducing aquifers where arsenic is predominantly present in its reduced state As(III), and second, high-pH aquifers where arsenic is relatively soluble in its oxidized state As(V) (*6, 10*). We, henceforth, refer to regions with these conditions as "reducing" and "high-pH/oxidizing" process regions, respectively.

Reducing aquatic environments, where arsenic is most probably released by reductive dissolution (*6*), are typically poorly drained and rich in organic matter content making them conducive to high microbial activity and, hence, low oxygen concentrations (*13*). We acknowledge that in reducing regions with higher sulfate concentrations, dissolved arsenic could be low due to microbial sulfate reduction and subsequent precipitation of arsenic sulfides (*14*). Oxidizing environments occur in arid and semiarid regions, again with poor drainage conditions (e.g., closed hydrologic basins), which may result in high evaporation rates leading to high salinity and high pH (*6, 7*). In such conditions anions including arsenate ($AsO_4^{3-}$) are poorly sorbed to mineral surfaces and are thus quite soluble and mobile (*6*).

This study aims to provide a global overview of arsenic-affected groundwaters. We use the digitally available information on the global scale such as geology, soil, climate, and topographic data to (1) delineate the world into reducing and high-pH/oxidizing regions, (2) model arsenic concentration in each region, and (3) produce probability maps for the occurrence of arsenic in concentrations greater than the WHO guideline value of 10 $\mu$g L$^{-1}$ (*15*). For these purposes the existing geochemical knowledge of the arsenic-releasing processes is combined with statistical methods using fuzzy logic and fuzzy neural networks. The outcomes are predictive models, which were tested and subsequently used to develop global probability maps for groundwater arsenic contamination.

## Materials and Methods

**Database Compilation.** We compiled around 20,000 data points of measured arsenic concentrations in groundwaters from around the world and in addition, natural factors related to climate, geology, hydrology, soil properties, land use, and elevation on a global scale from various Internet sources (see Supporting Information, Tables S1 and S2). The latter data were brought in the same projection in a GIS environment (ArcGIS, ver. 9.1) and were used as proxies to model the subsurface conditions.

Arsenic concentrations are point measurements within vertical depths of the wells of 10–100 m, while the other information have coarser spatial resolution being generally greater than 30 arc seconds (~1 km at the equator). The

* Corresponding author e-mail: johnson@eawag.ch; phone: +41 1 823-5465; fax: +41 1 823-5375.

**TABLE 1. Definition of Relevant Variables in the Global Database[a]**

| variable | definition-unit | type-format | resolution |
|---|---|---|---|
| arsenic | concentration ($\mu$g L$^{-1}$) | points | |
| elevation | (m) | continuous raster | 30 (arc second) |
| **slope** | **(degree)** | **continuous raster** | **30 (arc second)** |
| geology age | (million years) | continuous raster | 5 arc minute |
| ET | evapotranspiration (mm year$^{-1}$) | continuous raster | 0.5 degree |
| P | precipitation (mm year$^{-1}$) | continuous raster | 0.5 degree |
| **ET/P** | | **continuous raster** | **0.5 degree** |
| runoff | (mm year$^{-1}$) | continuous raster | 0.5 degree |
| T | temperature (°C) | continuous raster | 0.5 degree |
| irrigation | irrigated areas (%) | continuous raster | 5 arc minute |
| Sol_CbN1 | topsoil C/N ratio | continuous raster | 5 arc minute |
| clay1 | topsoil clay content (%) | continuous raster | 5 arc minute |
| silt1 | topsoil silt content (%) | continuous raster | 5 arc minute |
| sand1 | topsoil sand content (%) | continuous raster | 5 arc minute |
| Sol_CbN2 | subsoil C/N ratio | continuous raster | 5 arc minute |
| clay2 | subsoil clay content (%) | continuous raster | 5 arc minute |
| silt2 | subsoil silt content (%) | continuous raster | 5 arc minute |
| sand2 | subsoil sand content (%) | continuous raster | 5 arc minute |
| CEC_S | subsoil cation exchange capacity | ranked raster | 1:5000000 |
| **Drain_Code** | **soil drainage code** | **ranked raster** | **1:5000000** |
| N_S | subsoil nitrogen content | ranked raster | 1:5000000 |
| **OC_S** | **subsoil organic carbon content** | **ranked raster** | **1:5000000** |
| **pH_S** | **subsoil pH** | **ranked raster** | **1:5000000** |
| Dist_Volc | distance from volcanoes | continuous raster | 0.5 degree |
| Dist_Volc2 | distance from volcanic rocks | continuous raster | 0.5 degree |
| Dist_Riv | distance from rivers | continuous raster | 0.5 degree |

[a] CEC ranks from 10 (<20 meq/100 g clay) to 43 (>100 meq/100 g clay); Drain_Code ranks from 10 (extremely drained) to 87 (very poorly drained) (see Table S3); N_S ranks from 10 (<0.02%) to 53 (>0.5%); OC_S ranks from 10 (<0.2%) to 54 (> 2%) (see Table S4); pH_S ranks from 10 (<4.5) to 54 (>8.5) (see Table S5). Bold variables were used as delineating variables. Topsoil refers to a depth of 0−30 cm, while subsoil is from 30 to 100 cm.

point measurements were aggregated vertically and horizontally to the same spatial resolution as the geology map. The geometric mean of the arsenic data falling in the same pixels of the geological map units. Using this approach the number of arsenic data points was decreased to about 6000 pixel-based values. Proxy variables, as listed in Table 1, were deduced from the above-mentioned natural factors and used to model "reducing" and "high-pH/oxidizing" regions as discussed below.

**Rule Development and Process Region Delineation.** In order to separately model the two conditions defined as "reducing" and "high-pH/oxidizing", it was necessary to define regions in which either of these conditions were likely to prevail. Details of the delineation procedure are given in the Supporting Information.

In the first step, we defined what we call "delineating variables" that represent the critical conditions for identifying the type of process regions, using geochemical knowledge and statistical analysis. The information gained from geochemical expertise and literature review (6−11) is given in Table 2. Every variable and combination of all variables was statistically checked using supervised clustering and regression analysis for their significance in delineating process regions. The most significant variables, e.g., ratio of evapotranspiration over precipitation, drainage conditions, topographic slope, soil organic carbon content, and soil pH (Table 1) were used to develop rules to delineate different process regions.

In the second step, the qualifiers such as "high", "low", or "poor" (Table 2) were handled as fuzzy sets in a fuzzy inference system (FIS). FIS uses membership functions, fuzzy logic operators, and if−then rules to map a given input to an output (16−19). Fuzzy sets are sets whose elements have degrees of membership defined with membership functions valued in the real unit interval [0,1]. Rules were defined by continuous fuzzy sets for each of the delineating variables

**TABLE 2. Characteristics of Reducing and Oxidizing Regions Based on Expert Opinions (4−13)**

| variable[a] | reducing | oxidizing |
|---|---|---|
| ET/P | <1 | >1 |
| drainage condition | imperfect to poor | imperfect to poor |
| hydrologic basin | deltas | closed |
| slope | flat | flat |
| organic carbon | high | |
| salinity | low | high |
| temperature | high | |
| pH | | high |
| geology | young sediment | young sediment |

[a] ET = evaptranspiration, P = precipitation.

(ET/P < 1), (poor drainage condition), (high organic carbon), and (high pH). These sets were then used to calculate the maps of membership degree for each rule in the world. The overall membership for the combined set of rules is that of the rule that has the minimum membership value. This then determines the "possibility" of a pixel (10 km × 10 km) belonging to a reducing or a high-pH/oxidizing region. Using the developed FIS, we then constructed possibility maps of reducing and high-pH/oxidizing regions (see details of the procedure in Supporting Information).

**Modeling Arsenic in each Process Region.** In the third step, to develop a predictive model for arsenic in each process region, the following procedures were carried out: (i) the existing data set was split by stratified random sampling into two subsets for model training (85% of data) and model testing in each region, (ii) stepwise regression was used to identify the significant variables including those used to delineate process regions (see Table 1), and (iii) Adaptive Neuro-Fuzzy Inference System (ANFIS) was applied to relax the linearity assumption of the stepwise regression (see details of ANFIS in Supporting Information).

**TABLE 3. Significant Variables for Developing Arsenic Predictive Model for Process Regions**[a]

| | reducing condition | | | high-pH/oxidizing condition | |
| --- | --- | --- | --- | --- | --- |
| variable | *t*-value | *p*-value | variable | *t*-value | *p*-value |
| ET | −17.88 | $<10^{-40}$ | clay2 | −13.56 | $<10^{-40}$ |
| T | 13.53 | $<10^{-40}$ | silt1 | 10.41 | $<10^{-40}$ |
| ET/P | 7.08 | $<10^{-40}$ | clay1 | 7.34 | $<10^{-40}$ |
| Dist_Volc | 5.8 | $<10^{-40}$ | Drain_Code | −6.67 | $<10^{-40}$ |
| Drain_Code | 5.5 | $<10^{-40}$ | T | 5.55 | $<10^{-40}$ |
| silt2 | 5.31 | $<10^{-40}$ | Dist_Volc | 5.19 | $<10^{-40}$ |
| silt1 | −5.13 | $<10^{-40}$ | elevation | 4.82 | $<10^{-40}$ |
| Dist_Volc2 | −4.06 | $<10^{-40}$ | P | 3.57 | $<10^{-40}$ |
| Sol_CbN2 | 3.23 | $<10^{-40}$ | irrigation | 2.56 | 0.01 |
| pH_S | −2.93 | $<10^{-40}$ | | | |
| Dist_Riv | −2.6 | 0.01 | | | |

[a] ET = evapotranspiration; T = temperature; P = precipitation; Dist_Volc = distance from volcanoes; Drain_Code = drainage code; silt2 = subsoil silt content; silt1 = topsoil silt content; Dist_Volc2 = distance from volcanic rocks; Sol_CbN2 = carbon to nitrogen ration of subsoil; pH_S = subsoil soil pH; clay1 = topsoil clay content; clay2 = subsoil clay content; *t*-value indicates the relative importance of a variable (the larger in absolute values the more important a variable), and *p*-value indicates if a variable was rated important by chance or not (a value closer to zero is more desirable).

In the fourth and final step, the ANFIS models were linked to a Latin hypercube sampling (*20*) method to propagate the uncertainty of the model parameters on the results. This allowed the construction of the cumulative distribution function of arsenic concentrations in each pixel and, consequently, calculation of the probability of arsenic concentration exceeding the WHO guideline (*15*) at each pixel.

## Results and Discussion

**Process Region Delineation.** Based on the statistical analysis the following rules were found to delineate the process regions as shown in Figure S3.

• If (ET/P < 1) **then** a region is considered humid, **else** the region is arid. Where ET is evapotranspiration (mm y$^{-1}$), and P is precipitation (mm y$^{-1}$).

• If (Drain_Code ≥ 60) **then** drainage condition is poor, **else** drainage condition is good. Where Drain_Code is the drainage condition ranging from 10 (extremely drained) to 87 (very poorly drained) (see Table S3). In this rule, poor drainage includes mostly Histosols, Gleysols, and Fluvisols (according to FAO soil Map (*21*)) which implies gentle slopes and the presence of recent (Quaternary) sediments.

• **If** (OC_S ≥ 30) **then** subsoil (30−100 cm) organic carbon content is high, **else** it is low. Where OC_S is the organic carbon content of the subsoil ranging from 10 (<0.2 kg m$^{-3}$) to 54 (>2 kg m$^{-3}$) (see Table S4).

• If (pH_S ≥ 40) **then** pH is high, **else** it is low. Where pH_S is the soil pH rank where ≥40 is greater than 7.2 (see Table S5).

Based on the above, the process regions were defined as follows:

• **If** (region is humid) **and** (drainage condition is poor) **and** (subsoil organic content is high) **then** the condition is considered as reducing.

• **If** (region is arid) **and** (drainage condition is poor) **and** (subsoil pH is high) **then** the condition is considered as high-pH/oxidizing.

**Influencing Variables and Arsenic Model Results.** Table 3 shows the relative significance of the variables in the arsenic model. The occurrence of arsenic under reducing aqueous conditions is most closely correlated to climatic parameters (ET/P and T), geological parameters (Drain_Code, Dist_Volc, Dist_Volc2), and drainage conditions (Drain_Code), followed by soil parameters (Silt, pH_S, and Sol_CbN2) and proximity to surface water (Dist_Riv). The occurrence of arsenic under high-pH/oxidizing aqueous conditions is most closely correlated to soil parameters (clay and silt), and drainage

condition (Drain_Code), followed by climatic conditions (precipitation and temperature), topography (elevation), and geological factors (Dist_Volc and Drain_Code).

Note that some of these variables act as surrogates for some important but as yet unavailable influencing variables. For example, poorly drained areas (Drain_Code ≥ 60) correlate well with young sediments and flat areas (slope <1%), while areas of good drainage (Drain_Code < 60) correlate with volcanic rocks, or old sediments (see Table S7). Thus, Drain_Code is used as a proxy for a geological condition. Similarly, silt represents to some extent the transport and deposition of fresh materials in the river basin. As silt is produced only by mechanical weathering, it is highly reactive and therefore provides active absorption sites for arsenic species. Also distance to river (Dist_Riv) acts as an indicator for the deposition of fresh sediments. The soil parameter pH_S is important in the model since it has an influence on the redox potential of As-bearing species (*22*) and can facilitate higher arsenic occurrences (*23*). Climatic variables are also quite important in the reducing region. Evapotranspiration is negatively correlated with arsenic because larger ET indicates areas of larger precipitation or irrigation leading to greater groundwater recharge and smaller arsenic concentration, while ET/P, which is positively correlated with arsenic, indicates a smaller groundwater recharge. In high-pH/oxidizing region precipitation and temperature have similar effects where both seem to increase groundwater pH leading to larger arsenic concentration. Clay content appears prominently in the oxidizing model; which may modify drainage and hence leaching to groundwater. However, its exact role is not quite clear. Use of easily available information (surrogates) to estimate hard-to-obtain variables, often referred to as pedotransfer functions, is widely practiced in natural sciences (*24*).

The results of the arsenic model training and testing for both process regions are shown in Figure 1. In the reducing region, 77% of the variation in groundwater concentration in the training data set was explained using ANFIS, compared to 63% based on linear regression. For the test data set these were 65% and 59%, respectively. In high-pH/oxidizing region, however, ANFIS explained 68% of the variation in the training set (51% linear regression), while 65% was explained for the test data set (56% for linear regression).

**Probability Maps.** Probability maps in Figure 2a and b depict the probability of the occurrence of groundwater total arsenic concentrations exceeding the WHO guideline in reducing and high-pH/oxidizing regions, respectively. It is
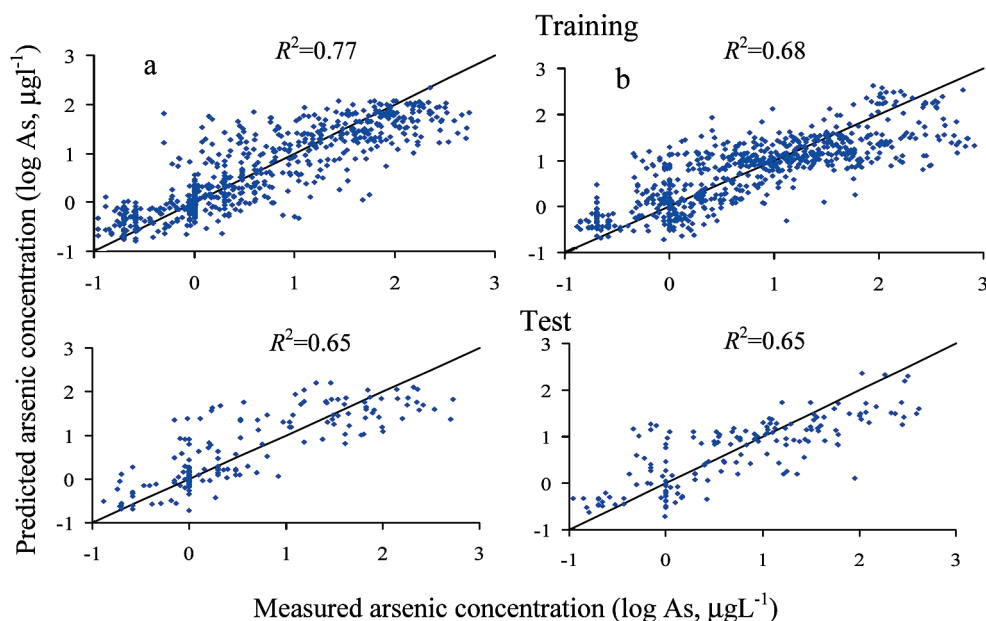
**FIGURE 1. Scatter plot of measured versus predicted arsenic concentration in (a) reducing and (b) high-pH/oxidizing process regions for training and testing of the neuro-fuzzy model.**
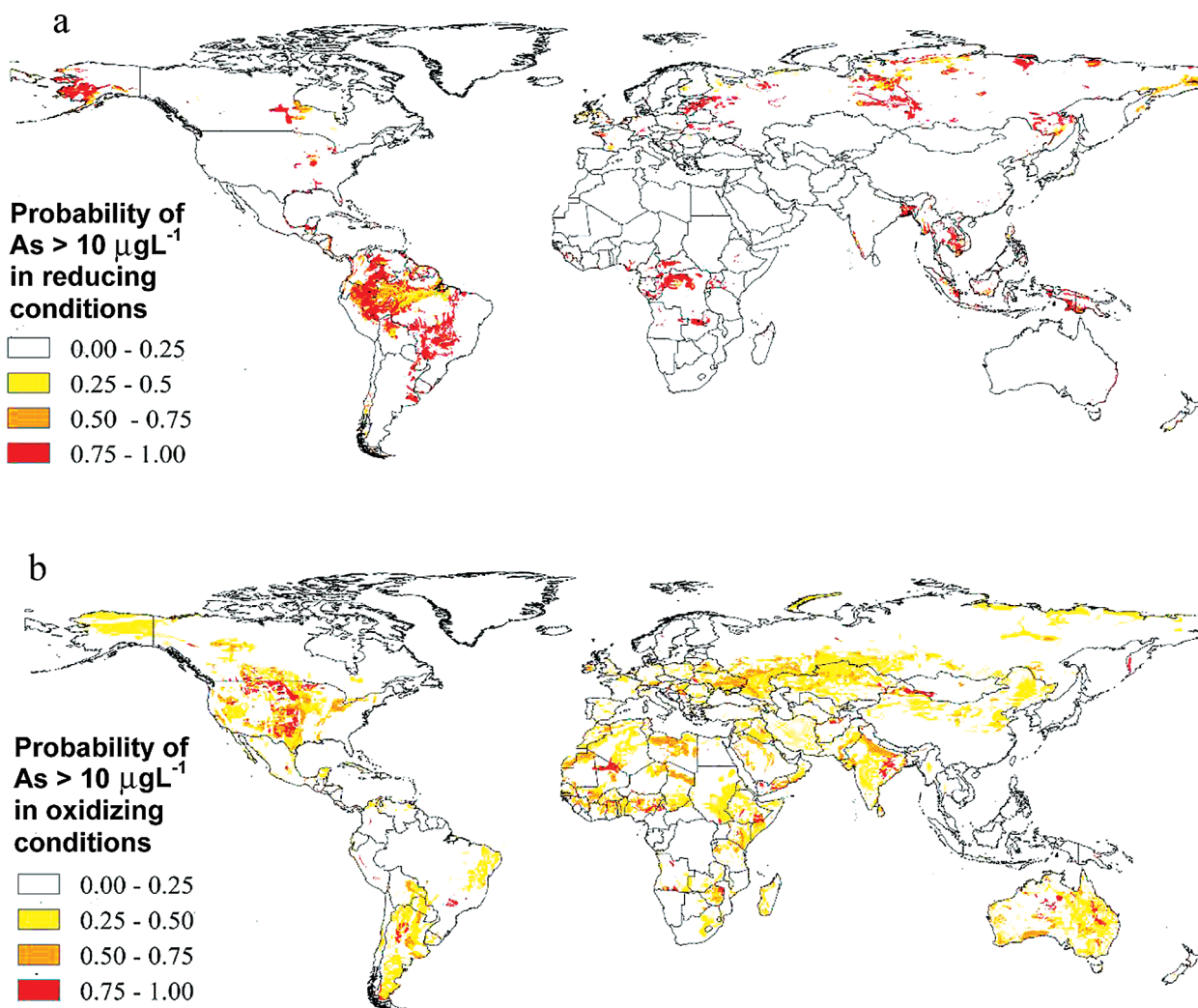


**FIGURE 2. Modeled global probability of geogenic arsenic contamination in groundwater for (a) reducing groundwater conditions, and (b) high-pH/oxidizing conditions where arsenic is soluble in its oxidized state.**

important to note that low-probability-areas do not exclude existence of arsenic contamination. Similarly, in areas with high probability of contamination clean aquifers may also be found. Groundwaters with a high probability of arsenic

**TABLE 4. State of Arsenic Contamination in Groundwaters in Different Countries of the World**

| predicted contaminated regions with reported contamination | | | predicted contaminated regions with no measurements or reported contamination | | |
|---|---|---|---|---|---|
| country | condition | % area[b] | country | condition | % area[b] |
| Bangladesh (2) | reducing | 35.4 | Estonia | reducing | 37.2 |
| Cambodia (25, 26) | reducing | 45.8 | Amazon basin[a] | reducing | 32.6 |
| Vietnam (8, 9, 25) | reducing | 15.8 | Lithuania | both | 35.0 |
| Taiwan (27) | reducing | 8.2 | Congo | reducing | 30.1 |
| Nepal (28) | reducing | 3.2 | Russia | both | 14.8 |
| Romania (29) | reducing | 3.5 | Myanmar | both | 9.2 |
| USA (7, 10, 23) | both | 8.3 | Poland | both | 8.8 |
| Argentina (30) | oxidizing | 4.9 | Cameroon | both | 14.0 |
| India (31) | both | 6.4 | Ukraine | oxidizing | 7.0 |
| China (4, 32) | both | 2.5 | Byelarus | oxidizing | 3.3 |
| Hungary (29) | reducing | 7.4 | Zambia | oxidizing | 7.0 |
| Finland (33) | unknown | 34.7 | Nigeria | oxidizing | 9.0 |
| Greece (34) | unknown | 0.1 | Angola | oxidizing | 5.5 |
| | | | Kenya | oxidizing | 2.4 |
| | | | Ethiopia | oxidizing | 5.3 |

[a] Average values for Peru, Brazil, and Colombia. [b] % Area in each country with probability of arsenic contamination >0.75.



**Probability of As > 10 $\mu$g L$^{-1}$ in both reducing and oxidizing conditions**
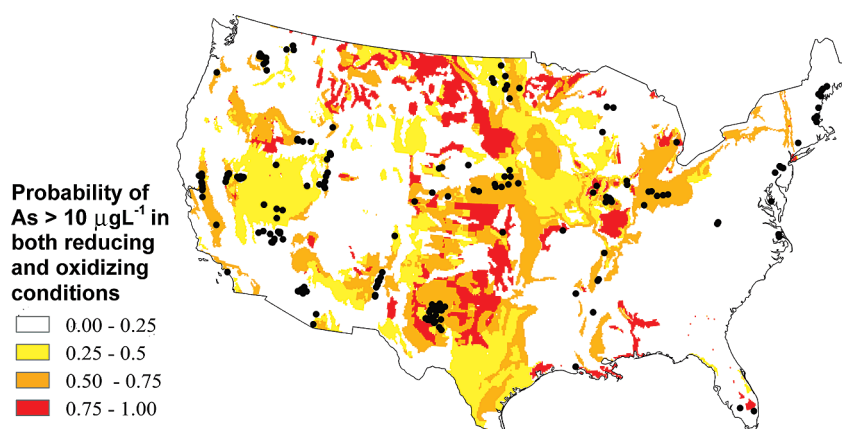
- 0.00 - 0.25
- 0.25 - 0.5
- 0.50 - 0.75
- 0.75 - 1.00

**FIGURE 3.** Probability map of arsenic concentration exceeding guideline value of 10 $\mu$g L$^{-1}$ in the United States. Black dots indicate a pixel (10 km $\times$ 10 km) with mean measured arsenic concentration greater than 10 $\mu$g L$^{-1}$

contamination appear to be widespread across the world. The predicted areas resulting from "reducing" and "high pH/ oxidizing" conditions are reported in Table 4 along with the references that have reported similar results.

The well-known contaminated regions in South East Asia, such as Bangladesh and India (6, 31), Nepal (28), Cambodia and Vietnam (8, 9, 12, 25, 26), China (4, 32), and Taiwan (27), correspond well with the predicted regions of high arsenic concentrations. Other predicted areas with high probability of arsenic such as those in Brazil, Bolivia, Ecuador, Costa Rica, Guatemala, and Nicaragua, for which no measurements were available to us, are reported to have high arsenic contamination (35). A large area in Amazon River Basin is predicted to have large probability of contamination, but in this region the risk to human health is small as groundwater is not a major source of water use. The predicted regions with high probability of arsenic contamination in Mexico, Chile, and Argentina (6), and the southwest United States (7, 10) are in agreement with the reported literature. Our predicted maps also show a high probability of arsenic contamination in Northern European countries such as Russia, Ukraine, Poland, Belarus, Estonia, and Lithuania. To our knowledge there is no report on the contamination of groundwaters with arsenic in Russia and Ukraine, although in surface waters of the Barents region elevated concentrations of arsenic have been observed (36). In addition, in the Siberian Arctic elevated concentrations of arsenic are reported

in the rivers and suspended particulate (37) and in soils, vegetation, and fish (38). The risk of geogenic groundwater contamination to human health is probably very small in the Barents region as people do not depend on groundwater as a source of drinking water (36).

A closer comparison between the modeled and measured data points with arsenic concentrations exceeding the WHO guideline is shown in Figure 3 for the United States and in Figure 4 for Bangladesh. There are regions such as the areas of southern New Hampshire and Maine, where the probability map underestimates contamination. There are, however, other detailed studies (23) that have reported small probabilities for this region of the U.S.

**Implications and Reliability.** It is important to note that validating a probability map without extensive additional measurements is not possible. However, we compared our results with as many studies as we could find in the published literature and found a quite good correspondence between the model prediction and measurements. One such example is based on a recent publication (32) where reported cases of arsenicosis (10,096 cases) in eight Chinese provinces (Supporting Information, Figure S6) were correlated with the percentage of contaminated wells (445,638 tested wells). In Figure S7 the percentage of arsenicosis and also the percentage of contaminated wells are plotted against the probability of arsenic contamination greater than 40% where a positive correlation can be observed. This comparison is
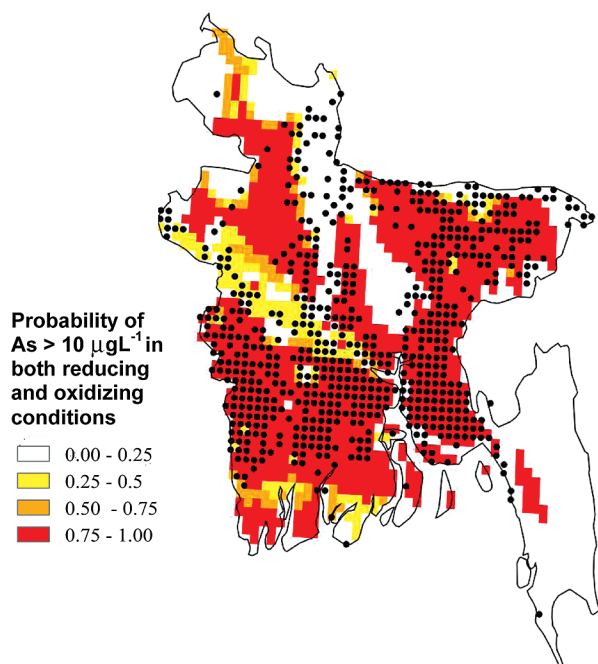
**FIGURE 4. Probability map of arsenic concentration exceeding the guideline value of 10 $\mu$g L$^{-1}$ in Bangladesh. Black dots indicate a pixel (10 km $\times$ 10 km) with mean measured arsenic concentration greater than 10 $\mu$g L$^{-1}$.**

important because it is a successful model verification that links our predictions to cases of arsenicosis in a location for which we had no data in our database.

As there are few regional or national arsenic maps around, we believe our global maps have the credibility of having a large database, physical variables, and a sound statistical analysis behind it. The weaknesses of the modeling work, perhaps not surprisingly, also lie in the size and quality of the databases. The measured arsenic points come from different sources and were measured with different accuracies; they were taken at different depths and are not distributed equally around the world. The geological information, which is important to geogenic processes, was too coarse and did not provide detailed (three-dimensional) information. For example, although it is known that arsenic contamination in reducing conditions in Asia is related to very young Holocene sediments ($<10^4$ yr), the available geological database used in this study does not distinguish rocks less than 65 million years (Cenozoic period). For this reason, geological variables were mostly furnished through surface information proxies. Despite these shortcomings, the arsenic models in both reducing and high-pH/oxidizing conditions performed quite well in training and testing, indicating it is possible to capture most of the variation in groundwater arsenic with the proxy variable used in this study. Probability maps give an overview of the spatial pattern of the likely contaminated and uncontaminated areas. They are not meant to represent arsenic concentrations of individual groundwater wells. The maps developed here are useful to policy makers, global organizations, local governments, and NGOs. The maps send a message that many untested areas may be subject to arsenic contamination. Although such information may be available in some countries, in most of the world arsenic is not routinely measured in groundwaters. This is particularly important as the risk of contamination to human health will increase as a result of increasing groundwater demand.

## Literature Cited

(1) Kapaj, S.; Peterson, H.; Liber, K.; Bhattacharya, P. Human health effects from chronic arsenic poisoning- a review. *J. Environ. Sci. Health, Part A* **2006**, *41*, 2399–2428.

(2) Smith, A. H.; Lingas, E. O.; Rahman, M. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. *Bull. World Health Organisation* **2000**, *78*, 1093–1103.

(3) Rich, C. H.; Biggs, M. L.; Smith, A. H. Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int. J. Epidemiol.* **1998**, *27*, 561–569.

(4) Xia, Y.; Liu, J. An overview on chronic arsenism via drinking water in PR China. *Toxicology* **2004**, *198*, 25–29.

(5) Saha, K. C. Review of Arsenicosis in West Bengal, India- A clinical perspective. Critical Reviews. *Environ. Sci. Technol.* **2003**, *30* (2), 127–163.

(6) Smedley, P. L.; Kinniburgh, D. G. A review of the source, behaviour and distribution of arsenic in natural waters. *Appl. Geochem.* **2002**, *17*, 517–568.

(7) Welch, A. H.; Westjohn, D. B.; Helsel, D. R.; Wanty, R. B. Arsenic in ground water of the United States: occurrence and geochemistry. *Ground Water* **2000**, *38*, 589–604.

(8) Berg, M.; Tran, H. C.; Nguyen, T. C.; Pham, H. V.; Schertenleib, R.; Giger, W. Arsenic contamination of groundwater and drinking water in Vietnam: A human health threat. *Environ. Sci. Technol.* **2001**, *35* (13), 2621–2626.

(9) Buschmann, J.; Berg, M.; Stengel, C.; Winkel, L; Sampson, M. L.; Trang, P. T. K.; Viet, P. H. Contamination of drinking water resources in the Mekong delta Floodplains: Arsenic and other trace metals pose serious health risks to population. *Environ. Int.* **2008**, . in press. (DOI: 10.1016/j.envint.2007.12.025).

(10) Welch, A. H.; Oremland, R. S.; Davis, J. A.; Watkins, S. A. Arsenic in groundwater: a review of current knowledge and relation to CALFED solution area with recommendations for needed research. *San Francisco Estuary Watershed Sci.* **2006**, *4* (2), 1–32.

(11) Nordstrom, D. K. Worldwide occurrences of arsenic in groundwater. *Science* **2002**, *296*, 2143–2144.

(12) Buschmann, J.; Berg, M.; Stengel, C.; Sampson, M. L. Arsenic and Manganese Contamination of Drinking Water Resources in Cambodia: Coincidence of Risk Areas with Low Relief Topography. *Environ. Sci. Technol.* **2007**, *41*, 2146–2152.

(13) Rowland, H. A.L.; Polya, D. A.; Lloyd, J. R.; Pancost, R. D. Characterization of organic matter in a shallow, reducing, arsenic-rich aquifer, West Bengal. *Org. Geochem.* **2006**, *37*, 1101–1114.

(14) Kirk, M. F.; Holm, T. R.; Park, J.; Jin, Q. S.; Sanford, R. A.; Fouke, B. W.; Bethke, C. M. Bacterial sulfate reduction limits natural arsenic contamination in groundwater. *Geology* **2004**, *32* (11), 953–956.

(15) WHO. *Guidelines for Drinking-Water Quality*, 2nd ed.; World Health Organization: Geneva, 1998; Addendum to Volume 2.

(16) Steeb, W. H.; Hardy, Y. *The Nonlinear Workbook: Chaos, Fractal, Cellular Automata*, 3rd ed.; World Scientific: NJ, 2005; 588 pp.

(17) Hines, J. W. *Matlab Supplement to Fuzzy and Neural Approaches in Engineering*; John Wiley and Sons: New York, 1997.

(18) Jang, J. S. R. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Trans. Syst., Man, and Cybernetics* **1993**, *23* (3), 665–685.

(19) Jang, J. S. R.; Gulley, N. *The Fuzzy Logic Toolbox for Use with MATLAB*; The Mathworks Inc., 1995.

(20) McKay, M. D.; Beckman, R. J.; Conover, W. J. A comparison of three methods for selecting values of input variables in the

analysis of output from a computer code. *Technometrics* **1979**, *21* (2), 239–245.

(21) FAO (Food and Agriculture Organization). *The digital soil map of the world and derived soil properties*; CD-ROM, Version 3.5, Rome, 1995.

(22) Stumm, W., Morgan J. J. *Aquatic Chemistry*, 3rd ed.; John Wiley Inc.: New York, 1996.

(23) Twarakavi, N. K. C.; Kaluarachchi, J. J. Arsenic in the shallow ground waters of conterminous United States: assessment, health risk, and costs from MCL compliance. *J. Am. Water Resour. Assoc.* **2006**, *42*, 275–294.

(24) Fielding, A. H. *Machine Learning Methods for Ecological Applications*; Kluwer Academic Publishers: Norwell, MA, 1999.

(25) Berg, M.; Stengel, C.; Trang, P. T. K.; Viet, P. H.; Sampson, M. L.; Leng, M.; Samreth, S.; Fredericks, D. Magnitude of arsenic pollution in the Mekong and Red River Deltas - Cambodia and Vietnam. *Sci. Total Environ.* **2007**, *372*, 413–425.

(26) Polya, D. A.; Gault, A. G.; Diebe, N.; et al. Arsenic hazard in shallow Cambodian groundwaters. *Mineral. Mag.* **2005**, *69* (5), 807–823.

(27) Lin, Y. B.; Lin, Y. P.; Liu, C. W.; Tan, Y. C. Mapping of spatial multi-scale sources of Arsenic variation in groundwater of ChiaNan floodplain of Taiwan. *Sci. Total Environ.* **2006**, *370*, 168–181.

(28) Neku, A.; Tandukar, N. An overview of arsenic contamination in groundwater of Nepal and its removal at household level. *J. Phys. IV* **2003**, *107*, 941–944.

(29) Lindberg, A. L.; Goessler, W.; Gurzau, E.; et al. Arsenic exposure in Hungary, Romania and Slovakia. *J. Environ. Monit.* **2006**, *8*, 203–208.

(30) Smedley, P. L.; Kinniburgh, D. G.; Macdonald, D. M. J.; et al. Arsenic associations in sediments from the loess aquifer of La Pampa, Argentina. *Appl. Geochem.* **2005**, *20* (5), 989–1016.

(31) Chakraborti, D.; Rahman, M. M.; Paul, K.; et al. Arsenic calamity in the Indian subcontinent, What lessons have been learned. *Talanta* **2002**, *58*, 3–22.

(32) Yu, G.; Sun, D.; Zheng, Y. Health Effects of Exposure to Natural Arsenic in Groundwater and Coal in China: An Overview of Occurrence. *Environ. Health Perspect.* **2007**, *115* (4), 636–642.

(33) Kurttio, P.; Pukkala, E.; Kahelin, H.; Auvinen, A.; Pekkanen, J. Arsenic concentrations in well water and risk of bladder and kidney cancer in Finland. *Environ. Health Perspect.* **1999**, *107*, 705–710.

(34) Kelepertsis, A.; Alexakis, D.; Skordas, K. Arsenic, antimony and other toxic elements in drinking water of Eastern Thessaly in Greece and its possible effects on human health. *Environ. Geol.* **2006**, *50*, 76–84.

(35) Bundschuh, J.; Armienta, M. A.; Bhattacharya, P.; Matschullat, J.; Birkle, P.; Rodriguez, R. *Natural Arsenic in Groundwaters of Latin America*; 2006. Available at http://www.lwr.kth.se/Personal/personer/bhattacharya_prosun/As-2006.htm.

(36) Salminen, R.; Chekushin, V.; Tenhola, M.; et al. Geochemical atlas of eastern Barents region. *J. Geochem. Exploration* **2004**, *83*, 1–530.

(37) Gordeev, V. V.; Rachold, V.; Vlasova, I. E. Geochemical behaviour of major and trace elements in suspended particulate material of the Irtysh river, the main tributary of the Ob river, Siberia. *Appl. Geochem.* **2004**, *19*, 593–610.

(38) Allen-Gil, S. M.; Ford, J.; Lasorsa, B. K.; Monettie, M.; Vlasova, T.; Landers, D. H. Heavy metal contamination in Taimyr Peninsula, Siberian Arctic. *Sci. Total Environ.* **2003**, *301*, 119–138.