RESEARCH ARTICLE

WATER RESOURCES

Global threat of arsenic in groundwater

Joel Podgorski[1,2]* and Michael Berg[1,3]*

Naturally occurring arsenic in groundwater affects millions of people worldwide. We created a global prediction map of groundwater arsenic exceeding 10 micrograms per liter using a random forest machine-learning model based on 11 geospatial environmental parameters and more than 50,000 aggregated data points of measured groundwater arsenic concentration. Our global prediction map includes known arsenic-affected areas and previously undocumented areas of concern. By combining the global arsenic prediction model with household groundwater-usage statistics, we estimate that 94 million to 220 million people are potentially exposed to high arsenic concentrations in groundwater, the vast majority (94%) being in Asia. Because groundwater is increasingly used to support growing populations and buffer against water scarcity due to changing climate, this work is important to raise awareness, identify areas for safe wells, and help prioritize testing.

The natural, or geogenic, occurrence of arsenic in groundwater is a global problem with wide-ranging health effects for humans and wildlife. Because it is toxic and does not serve any beneficial metabolic function, inorganic arsenic (the species present in groundwater) can lead to disorders of the skin and vascular and nervous systems, as well as cancer (*1*, *2*). The major source of inorganic arsenic in the diet is through arsenic-contaminated water, although ingestion through food, particularly rice, represents another important route of exposure (*3*). As a consequence, the World Health Organization (WHO) has set a guideline arsenic concentration of 10 μg/liter in drinking water (*4*).

At least trace amounts of arsenic occur in virtually all rocks and sediments around the world (*5*). However, in most of the large-scale cases of geogenic arsenic contamination in groundwater, arsenic accumulates in aquifers composed of recently deposited alluvial sediments. Under anoxic conditions, arsenic is released from the microbial and/or chemical reductive dissolution of arsenic-bearing iron(III) minerals in the aquifer sediments (*6*–*9*). Under oxidizing, high-pH conditions, arsenic can also desorb from iron and aluminum hydroxides (*10*). Furthermore, aquifers in flat-lying sedimentary sequences generally have a small hydraulic gradient, causing groundwater to flow slowly. This longer groundwater residence time allows dissolved arsenic to accumulate and its concentration to increase. Other processes responsible for arsenic release into groundwater include oxidation of arsenic-bearing sulfide minerals as well as release from arsenic-enriched geothermal deposits.

That arsenic is generally not included in the standard suite of tested water quality parameters (*11*) and is not detected by the human senses means that arsenic is regularly being discovered in new areas. Since one of the greatest occurrences of geogenic groundwater arsenic was discovered in 1993 in the Bengal delta (*5*, *12*, *13*), high arsenic concentrations have been detected all around the world, with hot spots including Argentina (*14*–*17*), Cambodia (*18*, *19*), China (*20*–*22*), India (*23*–*25*), Mexico (*26*, *27*), Pakistan (*28*, *29*), the United States (*30*, *31*), and Vietnam (*32*, *33*).

To help identify areas likely to contain high concentrations of arsenic in groundwater, several researchers have used statistical learning methods to create arsenic prediction maps based on available datasets of measured arsenic concentrations and relevant geospatial parameters. Previous studies have focused on Burkina Faso (*34*), China (*21*, *35*), South Asia (*29*, *36*), Southeast Asia (*37*), the United States (*31*, *38*, *39*), and the Red River delta in Vietnam (*33*), as well as sedimentary basins around the world (*40*). The predictor variables used in these studies generally include various climate and soil parameters, geology, and topography (table S3).

[1]Department of Water Resources and Drinking Water, Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland. [2]Department of Earth and Environmental Sciences, University of Manchester, Manchester M13 9PL, UK. [3]UNESCO Chair on Groundwater Arsenic within the 2030 Agenda for Sustainable Development and School of Civil Engineering and Surveying, University of Southern Queensland, Toowoomba, QLD 4350, Australia.
***Corresponding author. Email: joel.podgorski@eawag.ch (J.P.); michael.berg@eawag.ch (M.B.)**

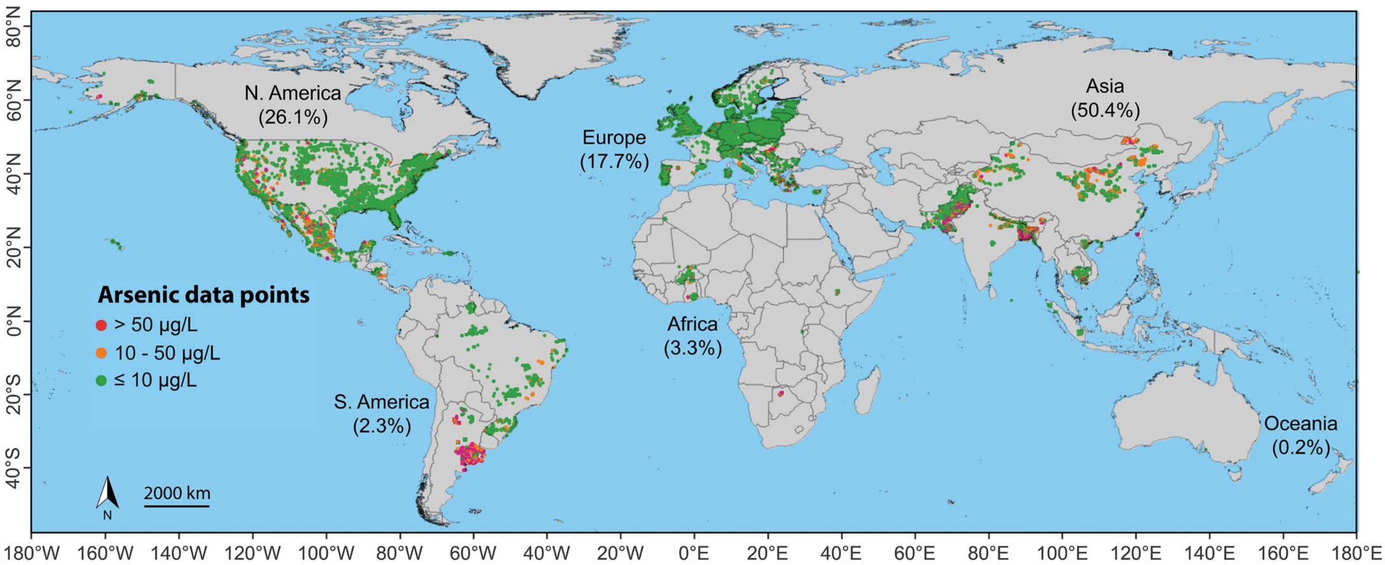

Fig. 1. Arsenic concentrations, excluding those known to originate from a depth greater than 100 m. Values are from the sources listed in table S1. The geographical distribution of data is indicated by continent.

Taking advantage of the increasing availability of high-resolution datasets of relevant environmental parameters, we use statistical learning to model what to our knowledge is the most spatially extensive compilation of arsenic measurements in groundwater assembled, which makes a global model possible. To focus on health risks, we consider the probability of arsenic in groundwater exceeding the WHO guideline. For this, we have chosen the random forest method, which our preliminary tests showed to be highly effective in addressing this classification problem. We use the resulting model to produce the most accurate and detailed global prediction map to date of geogenic groundwater arsenic, which can be used to help identify previously unknown areas of arsenic contamination as well as more clearly delineate the scope of this global problem and considerably increase awareness.

Results

Random forest modeling

We aggregated data from nearly 80 studies of arsenic in groundwater (see table S1 for references and statistics) into a single dataset (n > 200,000). Averaging into 1-km^2 pixels resulted in more than 55,000 arsenic data points for use in modeling based on groundwater samples not known to originate from greater than 100-m depth (Fig. 1).

To create the simplest and most accurate model, an initial set of 52 potentially relevant environmental predictor variables was iteratively reduced in consideration of their relative importance and impact on the accuracy of a succession of random forest models. The final selection of 11 predictor variables (table S2) includes several soil parameters (topsoil clay, subsoil sand, pH, and fluvisols), all of the climate variables (precipitation, actual and potential evapotranspiration, and combinations thereof, as well as temperature), and the topographic wetness index. By contrast, none of the geology variables proved to be statistically important. This is not to imply that geology does not play a role in geogenic arsenic accumulation, but rather that the particular geology variables tested were not as relevant as the other variables. This may be due to the coarse nature of the geological maps, which are standardized for the entire world. Although the number of predictor variables was reduced by nearly 80%, both the area

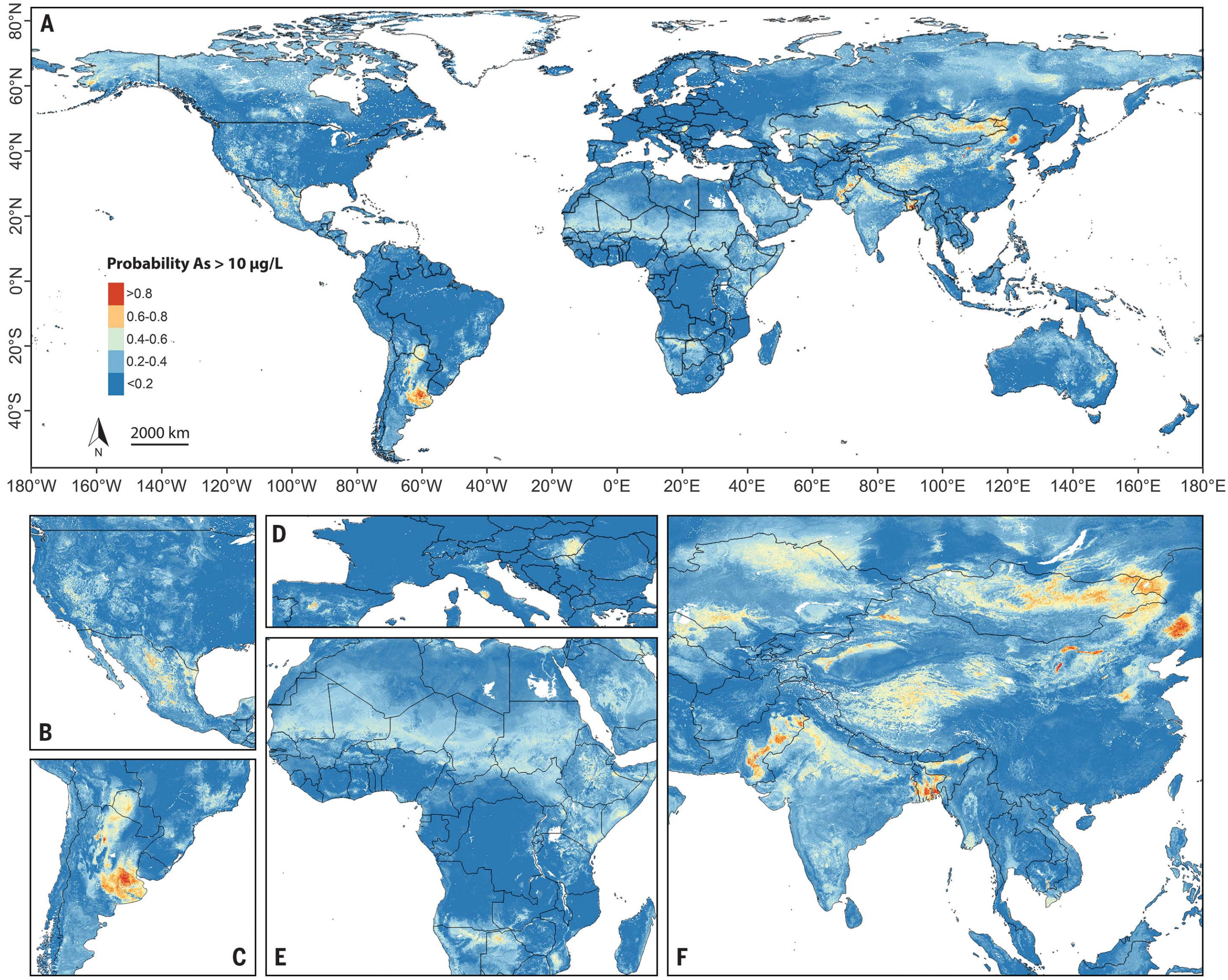

Fig. 2. Global prediction of groundwater arsenic. (**A** to **F**) Modeled probability of arsenic concentration in groundwater exceeding 10 µg/liter for the entire globe (A) along with zoomed-in sections of the main more densely populated affected areas (B) to (F). The model is based on the arsenic data points in Fig. 1 and the predictor variables in table S2. Figs. S2 to S8 provide more detailed views of the prediction map.

under the curve (AUC, 0.89) and Cohen's kappa statistic (0.55) remained unchanged.

The final random forest model was created based on the compiled global dataset of high and low arsenic concentrations along with the 11 predictor variables. The standard number of variables to be made available at each branch of each tree is between three and four (see methods). Because our tests showed the value of three performing better than four and higher values (though error and performance rates varied only within ~1%), we set this parameter to three. The global map produced from this model is displayed in Fig. 2A along with more detailed views of the more populated affected continental regions shown in Fig. 2, B to F. It indicates the probability of the concentration of arsenic in groundwater in a given 1-km^2 cell exceeding 10 μg/liter. The uncertainty of the model is inherent in the probabilities themselves, because they are simply the average of the votes or predictions of high or low values of each of the 10,001 trees grown. That is, each tree casts a vote of 0 or 1 ("no" or "yes" to As > 10 μg/liter) for each cell based on the values of the predictor variables in that cell. Figures S2 to S8 also provide more detailed views of the prediction map for each of the inhabited continents.

The importance of each of the 11 predictor variables in terms of mean decrease in accuracy and mean decrease in the Gini index is listed in fig. S1. Relative to the initial set of 52 variables, the values of these two statistics for most of the 11 final predictor variables appear to fall within a fairly narrow range, indicating comparable importance. Exceptions include fluvisols and soil pH, which have somewhat greater importance, and temperature, which, according to both statistics, is the least important of the 11 variables. Soil pH was also found to be an important predictor variable in arid, oxidizing environments in Pakistan (*29*). Although widespread arsenic dissolution occurs in Holocene fluvial sediments (*5*–*7*, *9*, *37*), this geological epoch has not been consistently mapped around the world. However, the global dataset of fluvisols provides a very suitable alternative (*29*), which may even be more appropriate because fluvisols by definition encompass recent fluvial sediments and not, for example, aeolian Holocene sediments that are generally not relevant for arsenic release. The generally high model importance of climate variables, as evidenced by them all being selected for the final model, highlights the strong control that climate has on arsenic release in aquifers. In particular, precipitation and evapotranspiration have a direct role in creating conditions conducive for arsenic release under reducing conditions (e.g., waterlogged soils) as well as high aridity associated with oxidizing, high-pH conditions.

The performance of the random forest model on the test dataset (20% of the data, which was randomly selected while maintaining the relative distribution of high and low values) is summarized in the confusion matrix in Table 1. Despite a prevalence of high values (>10 μg/liter) of only 22% in the dataset, the model performs well in predicting both high values (sensitivity: 0.79) and low values (specificity: 0.85) at a probability cutoff of 0.50. The average of these two figures, known as balanced accuracy, is correspondingly high at 0.82. Likewise, the model's AUC, which considers the full range of possible cutoffs, has a very high value of 0.89 with the test dataset (Table 1). For comparison, the AUC of a random forest using all 52 original predictor variables is also 0.89.

The model was also tested on a dataset of more than 49,000 arsenic data points originating from known depths greater than 100 m (average 562 m, standard deviation 623 m). Although the model was not trained on any measurements from these depths and the fact that only surface parameters were used as predictor variables, the model nevertheless performed quite well in predicting the arsenic concentrations of these deep groundwater sources, as evidenced by an AUC of 0.77.

Regions and populations at risk

Areas predicted to have high arsenic concentrations in groundwater exist on all continents, with most being located in Central, South, and Southeast Asia; parts of Africa; and North and South America (Fig. 2 and figs. S2 to S8). Known areas of groundwater arsenic contamination are generally well captured by the global arsenic prediction map, for example, parts of the western United States, central Mexico, Argentina, the Pannonian Basin, Inner Mongolia, the Indus Valley, the Ganges-Brahmaputra delta, and the Mekong River and Red River deltas. Areas of increased arsenic hazard where little concentration data exist include parts of Central Asia, particularly Kazakhstan, Mongolia, and Uzbekistan; the Sahel region; and broad areas of the Arctic and sub-Arctic. Of these, the Central Asian hazard areas are better constrained, as evidenced by higher probabilities.

Probability threshold values of 0.57 from the sensitivity-specificity comparison and 0.72 from the positive predictive value (PPV)–negative

Table 1. Confusion matrix and other statistics summarizing the results of applying the random forest model to the test dataset at a probability cutoff of 0.50.

Model output	Value
Predicted As ≤ 10 μg/liter	
Measured As ≤ 10 μg/liter	7710
Measured As > 10 μg/liter	555
Predicted As > 10 μg/liter	
Measured As ≤ 10 μg/liter	1394
Measured As > 10 μg/liter	2037
Sensitivity	0.79
Specificity	0.85
PPV	0.59
NPV	0.93
Prevalence	0.22
Balanced accuracy	0.82
Cohen's kappa	0.55
AUC	0.89

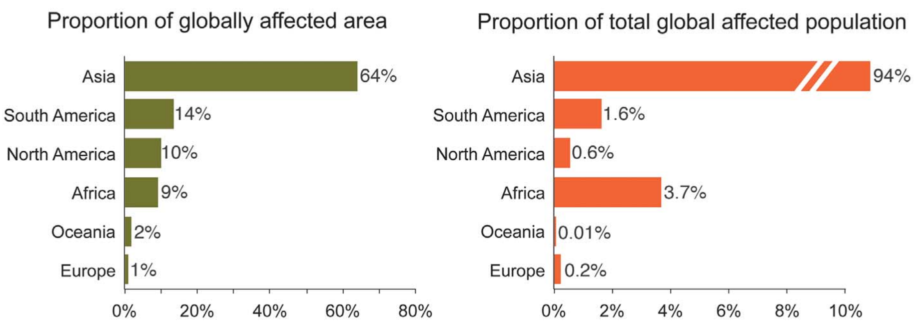

Fig. 3. Proportions of land area and population potentially affected by arsenic concentrations in groundwater exceeding 10 μg/liter by continent.

predictive value (NPV) comparison were found using the full dataset (combined training and test datasets) of arsenic concentrations. The proportions of high modeled arsenic hazard by continent associated with each of these probabilities are shown in Fig. 3. Global maps of the potentially affected population in the risk areas, as determined by these two thresholds, are shown in Fig. 4. As described in the methods, these maps were then used to estimate the population potentially affected by drinking groundwater with arsenic concentrations exceeding 10 μg/liter.

The resulting global arsenic risk assessment indicates that about 94 million to 220 million people around the world (of which 85 to 90% are in South Asia) are potentially exposed to high concentrations of arsenic in groundwater from their domestic water supply (tables S4 and S5). This range is consistent with the previous most comprehensive literature compilations, that is, 140 million people (*41*) and 225 million people (*42*). Household groundwater-use statistics were not available for ~6 to 8% of the affected countries (depending on the cutoff), for which the less detailed statistics derived from the AQUASTAT database of the Food and Agriculture Organization of the United Nations were used instead (see methods for details). To determine the amount of error that using these more general groundwater-use statistics might introduce to the overall population figures, the global potentially affected populations were recalculated with these countries' (those lacking household groundwater-use statistics) groundwater-use rates set to the extreme values of 0 and 100%. Because this applied to relatively few countries and arsenic-affected areas, doing so affected the overall global population figures by an inconsequential amount (±0.1%), indicating that using the AQUASTAT groundwater-use rates, where necessary, is an acceptable approximation.

This estimate of risk takes into account only the proportion of households utilizing unprocessed groundwater and assumes uniform rates throughout the urban and nonurban areas of each country. The uncertainties of these rates are unknown. The population in each cell was reduced by the uncertainty of the cell's prediction, which is justified based on the heterogeneity inherent in the accumulation of arsenic in an aquifer, which is generally at a much finer scale than that of the 1-km^2 resolution of the arsenic hazard map. Because the arsenic prediction for a cell represents the average outcome for that cell, we can take the modeled probability as a first-order approximation of the proportion of an aquifer in that cell containing high arsenic concentrations. Only cells exceeding the probability threshold (i.e., 0.57 or 0.72) were considered. The global estimate of 94 million to 220 million people potentially affected by consuming arsenic-contaminated groundwater is

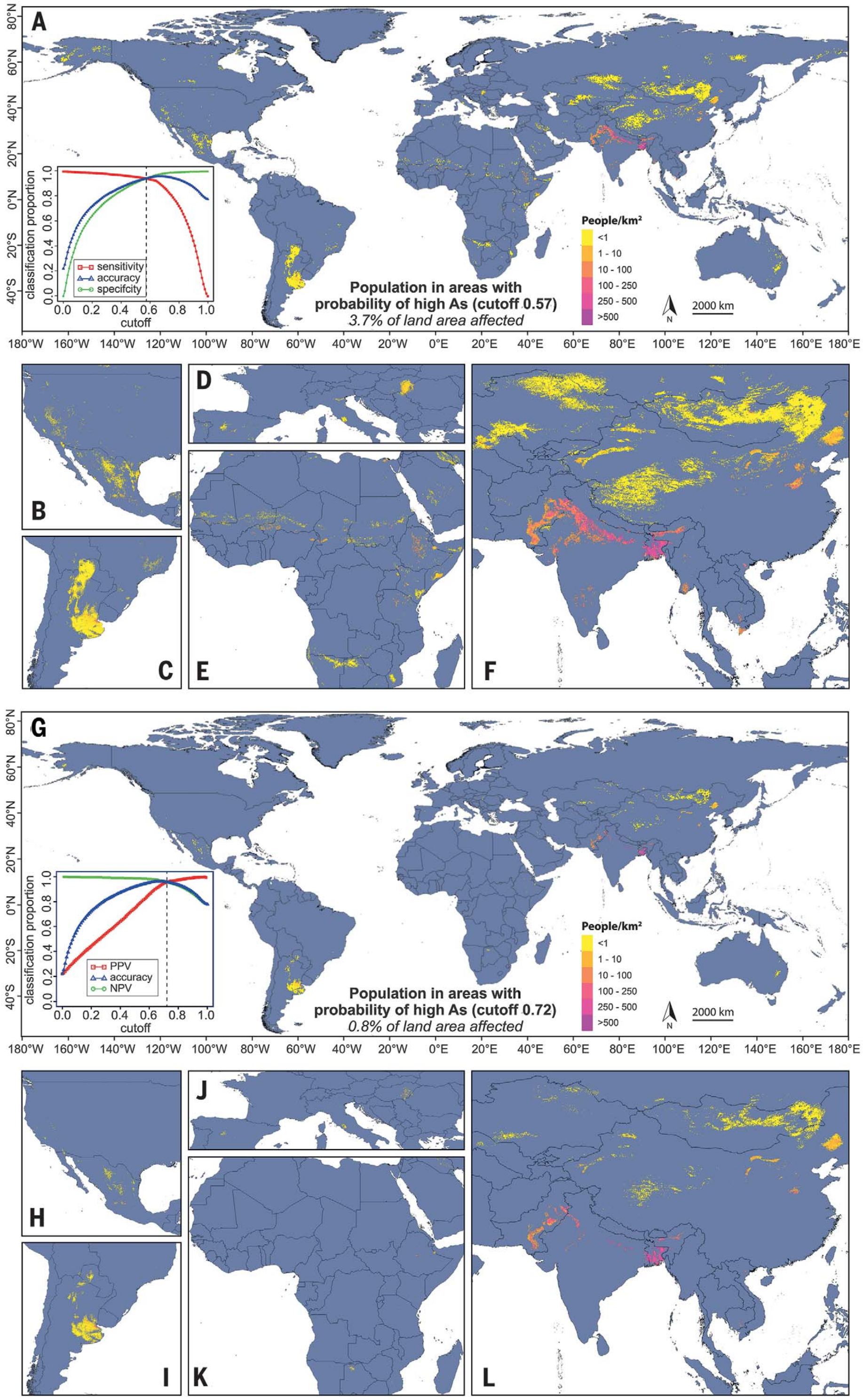

Fig. 4. Estimated population at risk. (**A** to **L**) Population in risk areas potentially containing aquifers with arsenic concentrations >10 μg/liter using probability cutoffs of 0.57 (A), at which sensitivity and specificity are equal [inset in (A)] as applied to the full (training and test) dataset, and 0.72 (G), at which PPV and NPV are equal [inset in (G)] using the full dataset. The detailed areas of Fig. 2 are also repeated here for both models (B) to (F) and (H) to (L).

broken down by continent and country in tables S4 and S5, respectively, and represents the most accurate and consistent global estimate available.

Discussion

The accuracy of the global groundwater arsenic prediction model presented here, as indicated, for example, with an AUC of 0.89 calculated with the test dataset, exceeds that found in previous arsenic prediction studies (table S3). The dominance of climate and soil parameters in the final model is indicative of their direct influence or at least strong association with the processes of arsenic accumulation in groundwater.

With respect to previous arsenic prediction maps of global sedimentary basins (*40*, *43*), the new model represents a substantial advancement on a few different levels. First, the new model presented here provides predictions for all areas of the inhabited continents, whereas the previous first-generation statistical model covered only about half of the land areas. In addition, a 10-fold increase in measurement points has allowed arsenic concentrations to be incorporated from many more areas of the globe. The greatly expanded availability and quality of global predictor datasets over the past 10 years has enabled new variables to be considered, such as soil type (e.g., fluvisols), as well as provided a 10- to 60-fold greater spatial resolution (i.e., 30 arc-sec versus 5 to 30 arc-min). However, the presence of high arsenic in groundwater at a given location is of course predicated on the existence of an aquifer in the first place, which may not be so in the case of unfractured solid rock, steep terrain, or very dry conditions. Models are only as good as the data on which they are based. As accurate as the new arsenic model is, it could be further improved as more arsenic data and more detailed predictor datasets come into existence.

Particularly in sedimentary aquifers, arsenic concentration is often highly dependent on depth, that is, on specific sedimentary sequences that differ in the concentration of arsenic in sediments as well as the geochemical conditions conducive to arsenic release. To better characterize this relationship in a given sedimentary basin, detailed depth information of groundwater samples would need to be incorporated in a separate basin-level study. Unfortunately, it is not feasible in a global-scale study to account for all of the diversity of the sedimentary basins of the world, especially because depth information of groundwater samples is often not available. As such, we have relied on a statistical analysis of model performance against depth ranges of samples (where present) to determine model sensitivity to depth.

Our approach in the risk assessment of potentially affected populations is relatively discerning and/or conservative. As such, the resulting population estimates may in some cases be lower than those found in earlier studies. One reason for this is that we used country-specific statistics of rural and urban domestic groundwater usage, which allowed us to subtract the proportion of the population that uses surface water, tap water, or other sources. This was not the case, for example, in a previous study of China that estimated that 19.6 million people were affected in the country (*21*), whereas our estimate is considerably lower at 4.3 million to 12.1 million. Furthermore, we consider only areas in which the probability of high arsenic exceeds the statistically determined cutoffs, that is, 0.57 and 0.72. Taking the United States as an example, applying this criterion left only 0.2 to 2% of the area of the country over which to sum the potentially affected population (≤0.21 million, this study). In a previous arsenic risk assessment of the United States (*31*), the entire country was used to estimate affected population (2.1 million), that is, not only the high-risk areas.

The actual proportion of groundwater usage varies spatially throughout a country, and so more detailed usage statistics beyond only urban versus rural would improve the accuracy of a risk assessment. In addition, more groundwater samples (ideally including depth information) from areas that currently have poor coverage would benefit future modeling efforts by allowing the model to be better adapted to those areas.

The presented arsenic probability maps should be used as a guide to further groundwater arsenic testing, for example, in Central Asia, the Sahel, and other regions of Africa. Only actual groundwater quality testing can definitively determine the suitability of groundwater with respect to arsenic, particularly because of small-scale (<1 km) aquifer heterogeneities that cannot be modeled with existing global datasets (*9*, *44*). The hazard maps highlight areas at risk and provide a basis for targeted surveys, which continue to be important. The already large number of people potentially affected can be expected to increase as groundwater use expands with a growing population and increasing irrigation, especially in the light of water scarcity associated with warmer and drier conditions related to climate change. The maps can also help aid mitigation measures, such as awareness raising, coordination of government and financial support, health intervention programs, securing alternative drinking water resources, and arsenic removal options tailored to the local groundwater conditions as well as social setting.

REFERENCES AND NOTES

1. A. H. Smith, E. O. Lingas, M. Rahman, *Bull. World Health Organ.* **78**, 1093–1103 (2000).
2. M. F. Hughes, *Toxicol. Lett.* **133**, 1–16 (2002).
3. D. Mondal *et al*., *Environ. Geochem. Health* **32**, 463–477 (2010).
4. H. G. Gorchev, G. Ozolins, *WHO Chron.* **38**, 104–108 (1984).
5. P. Smedley, D. Kinniburgh, *Appl. Geochem.* **17**, 517–568 (2002).
6. R. Nickson *et al*., *Nature* **395**, 338–338 (1998).
7. J. McArthur, P. Ravenscroft, S. Safiulla, M. Thirlwall, *Water Resour. Res.* **37**, 109–117 (2001).
8. M. Berg *et al*., *Chem. Geol.* **249**, 91–112 (2008).
9. S. Fendorf, H. A. Michael, A. van Geen, *Science* **328**, 1123–1127 (2010).
10. M. I. Litter *et al*., *Sci. Total Environ.* **676**, 756–766 (2019).
11. Y. Zheng, S. V. Flanagan, *Environ. Health Perspect.* **125**, 085002 (2017).
12. P. Bhattacharya, D. Chatterjee, G. Jacks, *Int. J. Water Resour. Dev.* **13**, 79–92 (1997).
13. A. van Geen *et al*., *Water Resour. Res.* **39**, 1140 (2003).
14. H. B. Nicolli, J. M. Suriano, M. A. Gomez Peral, L. H. Ferpozzi, O. A. Baleani, *Environ. Geol. Water Sci.* **14**, 3–16 (1989).
15. P. Smedley, H. Nicolli, D. Macdonald, A. Barros, J. Tullio, *Appl. Geochem.* **17**, 259–284 (2002).
16. M. Blarasin, A. Cabrera, E. Matteoda, paper presented at the XXXIII IAH – 7° ALHSUD Congress, Zacatecas, Mexico, 11 to 15 October 2004.
17. M. Auge, G. E. Viale, L. Sierra, in *VIII Congreso Argentino de Hidrogeología: Aguas subterráneas recurso estratégico* (Editorial de la Universidad Nacional de La Plata, 2013), vol. 2, pp. 58–63.
18. M. Berg *et al*., *Environ. Sci. Technol.* **35**, 2621–2626 (2001).
19. J. Buschmann, M. Berg, C. Stengel, M. L. Sampson, *Environ. Sci. Technol.* **41**, 2146–2152 (2007).
20. P. Smedley, M. Zhang, G. Zhang, Z. Luo, *Appl. Geochem.* **18**, 1453–1477 (2003).
21. L. Rodríguez-Lado *et al*., *Science* **341**, 866–868 (2013).
22. Y. Zhou *et al*., *Appl. Geochem.* **77**, 116–125 (2017).
23. D. Chatterjee, R. Roy, B. Basu, *Environ. Geol.* **49**, 188–206 (2005).
24. B. Nath, D. Stüben, S. B. Mallik, D. Chatterjee, L. Charlet, *Appl. Geochem.* **23**, 977–995 (2008).
25. B. A. Shah, *Curr. Sci.* **98**, 1359–1365 (2010).
26. B. Planer-Friedrich, "Hydrogeological and hydrochemical investigations in the Rioverde Basin, Mexico," thesis, Institute of Geology, University of Mining and Technology Freiberg (2000).
27. M. T. Alarcón-Herrera *et al*., *J. Hazard. Mater.* **262**, 960–969 (2013).
28. R. Nickson, J. McArthur, B. Shrestha, T. Kyaw-Myint, D. Lowry, *Appl. Geochem.* **20**, 55–68 (2005).
29. J. E. Podgorski *et al*., *Sci. Adv.* **3**, e1700935 (2017).
30. J. D. Ayotte, M. G. Nielsen, G. R. Robinson Jr., R. B. Moore, *Water Resour. Invest. Rep.* **99**, 4162 (1999).
31. J. D. Ayotte, L. Medalie, S. L. Qi, L. C. Backer, B. T. Nolan, *Environ. Sci. Technol.* **51**, 12443–12454 (2017).
32. M. Berg *et al*., *Sci. Total Environ.* **372**, 413–425 (2007).
33. L. H. Winkel *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1246–1251 (2011).
34. A. Bretzler *et al*., *Sci. Total Environ.* **584–585**, 958–970 (2017).
35. Q. Zhang *et al*., *J. Hazard. Mater.* **262**, 1147–1153 (2013).
36. S. Bindal, C. K. Singh, *Water Res.* **159**, 65–76 (2019).
37. L. Winkel, M. Berg, M. Amini, S. J. Hug, C. A. Johnson, *Nat. Geosci.* **1**, 536–542 (2008).
38. Q. Yang, H. B. Jung, R. G. Marvinney, C. W. Culbertson, Y. Zheng, *Environ. Sci. Technol.* **46**, 2080–2087 (2012).
39. N. Yang, L. H. Winkel, K. H. Johannesson, *Environ. Sci. Technol.* **48**, 5660–5666 (2014).
40. M. Amini *et al*., *Environ. Sci. Technol.* **42**, 3669–3675 (2008).
41. P. Ravenscroft, H. Brammer, K. Richards, *Arsenic Pollution: A Global Synthesis* (Wiley, 2009), vol. 28.
42. S. Murcott, *Arsenic Contamination in the World* (IWA Publishing, 2012).
43. P. Ravenscroft, "Predicting the global extent of arsenic pollution of groundwater and its potential impact on human health," unpublished report prepared for UNICEF, December 2007.
44. Y. Zheng, *Curr. Environ. Health Rep.* **4**, 373–382 (2017).
45. J. Podgorski, M. Berg, Podgorski_and_Berg_2020. ERIC/open (2020); http://dx.doi.org/10.25678/0001ZT.

ACKNOWLEDGMENTS

We thank our colleagues A. Bretzler and C. Zurbrügg (Eawag), A. Steiner and S. Piers de Raveschoot (SDC), and D. A. Polya and R. Wu (University of Manchester) for their support, as well as the many providers of data, which were an essential component of this work. **Funding:** We thank the Swiss Agency for Development and Cooperation (project nos. 7F-09010.01.01 and 7F-09963.01.01) for long-term support and cofunding of this study, as well as a University of Manchester EPSRC IAA Impact Support Fund Award to D. A. Polya. **Author contributions:** J.P.: Methodology, Modeling, Writing–Original draft preparation; M.B.: Supervision, Writing–Reviewing and Editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The modeling data, code, and raster output maps are available at ERIC/open (*45*). Arsenic concentration data points and hazard and risk maps are also available for viewing on the GIS-based Groundwater Assessment Platform (GAP), www.gapmaps.org.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/845/suppl/DC1
Methods
Figs. S1 to S11
Tables S1 to S6
References (*46*–*127*)

10 November 2019; accepted 3 April 2020
10.1126/science.aba1510

Science

Global threat of arsenic in groundwater

Joel Podgorski and Michael Berg

Science **368** (6493), 845-850.
DOI: 10.1126/science.aba1510

Dowsing for danger
Arsenic is a metabolic poison that is present in minute quantities in most rock materials and, under certain natural conditions, can accumulate in aquifers and cause adverse health effects. Podgorski and Berg used measurements of arsenic in groundwater from ~80 previous studies to train a machine-learning model with globally continuous predictor variables, including climate, soil, and topography (see the Perspective by Zheng). The output global map reveals the potential for hazard from arsenic contamination in groundwater, even in many places where there are sparse or no reported measurements. The highest-risk regions include areas of southern and central Asia and South America. Understanding arsenic hazard is especially essential in areas facing current or future water insecurity.
Science, this issue p. 845; see also p. 818

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.