Environ. Sci. Technol. **2008,** *42,* 3662–3668

Eawag 05270

Statistical Modeling of Global Geogenic Fluoride Contamination in Groundwaters

MANOUCHEHR AMINI,[†] KIM MUELLER,[†] KARIM C. ABBASPOUR,[†] THOMAS ROSENBERG,[†] MAJID AFYUNI,[‡] KLAUS N. MØLLER,[§] MAMADOU SARR,[||] AND C. ANNETTE JOHNSON*[,†]

Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland, Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, P.O. Box 84154, Iran, Water Resources and Nature Restoration, COWI A/S Parallelvej 2, DK-2800 Kongens Lyngby, Denmark, and Service de Gestion et de Planification des Resources en Ean (SGPRE), Avenue Andre Peytavin, Ex camp Lat Dior, Dakar, Senegal

Received August 6, 2007. Revised manuscript received December 17, 2007. Accepted March 6, 2008.

The use of groundwater with high fluoride concentrations poses a health threat to millions of people around the world. This study aims at providing a global overview of potentially fluoride-rich groundwaters by modeling fluoride concentration. A large database of worldwide fluoride concentrations as well as available information on related environmental factors such as soil properties, geological settings, and climatic and topographical information on a global scale have all been used in the model. The modeling approach combines geochemical knowledge with statistical methods to devise a rule-based statistical procedure, which divides the world into 8 different "process regions". For each region a separate predictive model was constructed. The end result is a global probability map of fluoride concentration in the groundwater. Comparisons of the modeled and measured data indicate that 60−70% of the fluoride variation could be explained by the models in six process regions, while in two process regions only 30% of the variation in the measured data was explained. Furthermore, the global probability map corresponded well with fluorotic areas described in the international literature. Although the probability map should not replace fluoride testing, it can give a first indication of possible contamination and thus may support the planning process of new drinking water projects.

Introduction

Fluorine is the 13th most abundant element in the earth's crust (625 mg kg^{-1}) and exists in trace amounts in almost all groundwaters throughout the world (*1*). Drinking water is often the main source of fluoride intake by humans, especially in areas where fluoride concentrations in groundwater and/or surface water are high (*2*−*4*). It is estimated that more than 200 million people worldwide (*5*) rely on drinking water with fluoride concentrations that exceed the present WHO guideline of 1.5 mg L^{-1} (*6*). In some areas foodstuffs and/or indoor air pollution due to the burning of coal may make significant contributions to the daily intake of fluoride (*7, 8*). Excess fluoride intake causes different types of fluorosis, primarily dental and skeletal fluorosis, depending on the level and period of exposure (*9*). Fluorosis, associated with elevated fluoride concentrations in drinking water, has been reported in various countries around the world such as India (*10*), China (*11*), Tanzania (*12*), Mexico (*13*), Argentina (*14*), and South Africa (*15*), among others.

While on a local scale anthropogenic activities, such as the application of phosphate-containing fertilizers or aluminum smelting, may introduce considerable amounts of fluoride into the environment (*16*), its concentration in the groundwaters is mainly governed by geogenic processes (*3, 10*). Various minerals, e.g., fluorite, biotites, topaz, and their corresponding host rocks such as granite, basalt, syenite, and shale, contain fluoride that can be released into groundwater (*3, 17, 18*).

The release of fluoride to groundwater is dependent on chemical and physical processes that take place between the groundwater and its geological environment. Fluorite (CaF_2) is the predominant mineral that controls the dissolved fluoride concentration in the groundwater (*3, 16, 19*). Thus fluoride-rich groundwaters are often associated with low calcium concentrations. This is associated with rocks with low calcium content, or high pH conditions where sodium bicarbonate dominates the groundwater composition. Apart from the groundwater chemistry, hydrological properties (e.g., residence time) as well as climatic conditions (e.g., evapotranspiration, precipitation) and soil conditions (e.g., pH, soil type) have an influence on fluoride concentration (*3, 10, 16, 20, 21*). Hence, the spatial and temporal heterogeneities of fluoride concentrations in groundwater are particularly large.

The extent of fluoride contamination in groundwater has been studied on both local and national scales, but little is known about the occurrence of fluorotic areas on a global scale. In this study we aim to provide a global probability map that indicates the risk of fluoride contamination in groundwaters. This is achieved by combining geochemical expertise with information on available environmental factors on a global scale by means of an advanced statistical procedure within the ArcGIS environment (ver. 9.1). The outcomes are probability maps of groundwater contamination developed for eight "process regions". The process regions were delineated using both rules based on the knowledge of fluoride dissolution as well as statistical testing. The rules discriminate among the effects of geological, climatic, and soil properties on the fluoride dissolution. Predictive models were developed (calibrated and validated) for each of the eight process regions. To the best of our knowledge, this is the first global-scale predictive map of fluoride contamination in groundwaters.

Materials and Methods

Data Compilation. A global database of over 60,000 measured fluoride concentrations in groundwaters was compiled from 25 countries around the world. A summary of the basic statistics and sources of the collected fluoride measurements is given in the Supporting Information, Table S1. South Africa, Sweden, the United States, and Norway had the largest compilation of data. Globally available information related

* Corresponding author fax: 0041 44 823 5210; e-mail: johnson@eawag.ch.

[†] Eawag, Swiss Federal Institute of Aquatic Science and Technology.

[‡] Isfahan University of Technology.

[§] Water Resources and Nature Restoration.

[||] Service de Gestion et de Planification des Resources en Ean (SGPRE).

10.1021/es071958y CCC: $40.75
Published on Web 04/16/2008

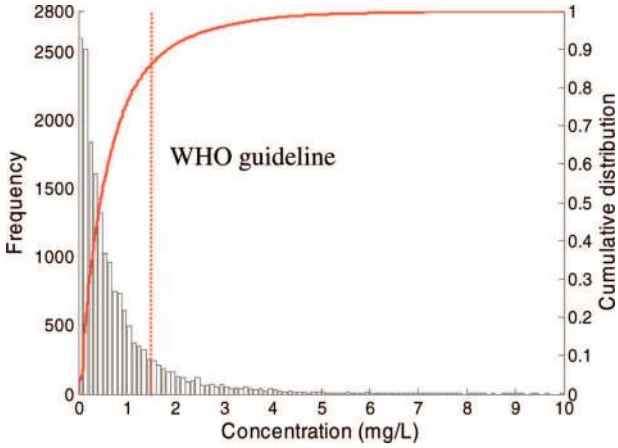

FIGURE 1. Frequency distribution of aggregated fluoride concentrations.

to climate, geology, hydrology, soil, land use, elevation, and slope were collected from different sources (Supporting Information, Table S2) and a multilayer global database was created in the ArcGIS (ver. 9.1) environment. A summary of the variables in the database and their definitions is given in Table S3.

As fluoride data are point measurements, whereas other GIS layers are raster files, we aggregated the fluoride measurements to the same spatial resolution as the geological information (5 arc minutes or approximately 10 km at the equator). The reason for choosing the spatial resolution of geology was because of its coarser spatial resolution as well as its importance in controlling fluoride concentration in the groundwater (*3, 10, 16*). The data were aggregated by assigning the geometric mean of fluoride concentrations in the same pixel to its center. Hence the total number of data points was reduced to approximately 18,000 pixel-based fluoride measurements. The frequency of fluoride concentration is shown in Figure 1. The spatial distribution of the data is illustrated in Figure S1.

Developing Knowledge-Based Statistical Rules. Groundwater fluoride concentrations are extremely variable due to the complex interactions between geo-environment, hydrology, and other environmental variables. It is therefore necessary to define regions where similar conditions prevail in order to enable modeling procedures. Based on geochemical expertise and extensive reviews of the literature, several important geological, chemical, climatic, and soil conditions associated with fluoride-rich groundwaters were distinguished. These conditions are summarized in Table 1. Some of the important hydrological and chemical processes could not be directly accounted for because of the lack of information, such as groundwater composition. As the available geological map was coarse and did not provide detailed information about different rock types, we used the information on "intrusive felsic rocks" to account for granitic rocks. To capture the volcanic origin of fluoride, we combined the available information on "volcanic felsic rocks" and "normal faults". Our preliminary statistical analysis showed that fluoride concentrations decreased with increasing distance from intrusive felsic rocks or extensional tectonic activities. This relationship was quantified by taking distance vectors of [0−5] decimal degree with increments of 0.1 decimal degrees and plotting it against the median concentration of fluoride within that distance vector as illustrated in Figure 2a. The figure shows that below a distance of 1 decimal degree there is a clear relationship between fluoride concentration and the distance from intrusive felsic rocks and normal faults. Hence, a distance of 1 decimal degree was chosen in the subsequent analysis.

At the second level of delineation, climatic condition was captured by creating an ET-index, expressed as evapotranspiration over precipitation (ET/P). Fluoride concentrations may be enhanced due to high evapotranspiration and small dilution if precipitation is low. Although fluoride-rich groundwaters are more common under arid and semiarid climatic conditions (*3, 29*), some groundwaters with high fluoride concentrations are also reported in humid and/or cold regions. Examples of these are parts of Sweden, Norway, and Argentina (*14, 35, 36*). To find a relationship between ET/P and fluoride concentration, we plotted a series of ET/P thresholds, $(ET/P)_T$, with increments of 0.2 against the geometric mean of fluoride concentrations fulfilling the condition of $ET/P \geq (ET/P)_T$ as shown in Figure 2b. The figure shows that there is a clear relationship between $ET/P > 2$ and fluoride concentration. Hence, this ratio was used for a further delineation of the process regions as shown in Figure 3. To test the appropriateness of this delineation, we kept the regions with $ET/P < 2$ in the analysis, although they do not constitute actual process regions.

In the third level of delineation, we used the influence of subsoil pH on fluoride concentration (Figure 3). To quantify the effect of soil pH on fluoride concentration two groups of soils with $pH \geq 7.2$ and $pH < 7.2$ were differentiated because alkaline soils are known to have a positive correlation with fluoride concentration (*23, 30*) (see Supporting Information, Table S4).

Based on the above delineations, eight process regions (PR) were defined as shown in Figure 3. We compared the fluoride concentrations in these regions using nonparametric

TABLE 1. Summary of the Geochemical Conditions Associated with Fluoride-Rich Groundwaters and the GIS Layers Used to Represent Them

geochemical conditions			reperesenting GIS layer[b]
geology	granitic rocks	granitic/gneissic basement (*16, 22*)[a] biotite granite (*23*) granites and rhyolites (*24*)	Dist_Int_fel
	volcanism	active volcanism (*25*) acidic volcanic rocks (*26*) alkaline volcanic rocks (*27*)	Dist_fault & Dist_V_fel
	sedimentary rocks	clayey carbonates (*28*) shales (*18*) loess deposits (*29*)	sedimentary
soil		alkaline soil (*23, 30*)	pH_s
climatic		high evapotranspiration (*19, 23, 30, 31*)	ET/P
groundwater chemistry		Na-HCO_3 type of groundwater with elevated pH (*19, 32−34*)	global database not available

[a] Numbers indicate references. [b] Dist_Int_fel = distance from intrusive felsic rocks, Dist-fault = distance from normal faults, Dist_V_fel = distance from volcanic felsic rocks, pH_s = subsoil pH, and ET/*P* = evapotranspiration/precipitation.

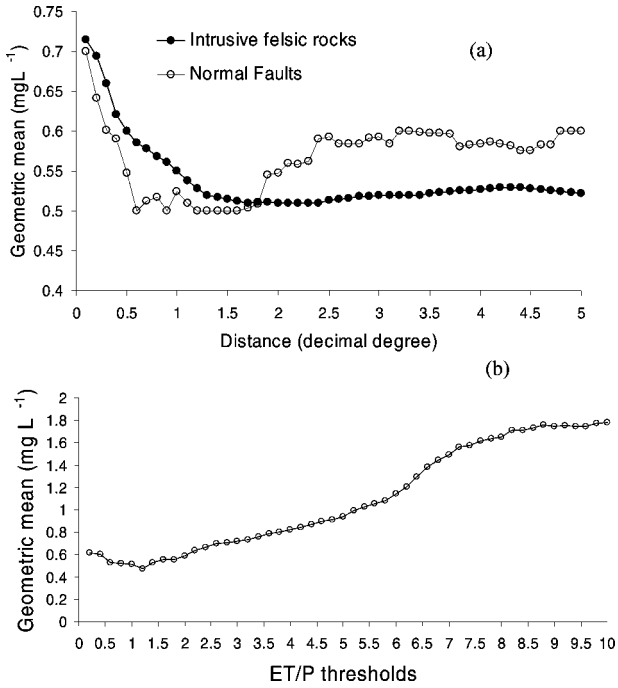

FIGURE 2. Geometric mean of fluoride concentration as a function of distance to intrusive felsic rocks and normal faults (a) and evapotranspiration over precipitation (ET/P) (b).

Kruskal–Wallis test followed by least-significant difference test (LSD) for multiple comparisons.

Statistical Model Development in the Process Regions. A statistical model to predict fluoride concentration was developed for each region by splitting the fluoride data into two subsets using stratified random sampling. The subset with 80% of the data was used for model development and the remaining was set aside for model validation.

Stepwise regression was then used to identify the significant variables for model development. This was followed by the application of the Adaptive Neuro-Fuzzy Inference System (ANFIS) (*37, 38*) to identify the best predictive model. A short description of ANFIS is provided in the Supporting Information. Mean error (ME), mean square error (MSE), and R^2 statistics were used to evaluate the models. ANFIS models were then linked to a Latin hypercube sampling method to propagate the uncertainty of the parameters on the results. These parameters included: centers of different fuzzy classes of variables (see Supporting Information, Figure S2), and the coefficients of the multiple linear regression model. This allowed the prediction of the cumulative distribution function and, subsequently, the determination of the probability of fluoride concentrations exceeding a specific threshold (in this study 1.5 mgL^{-1}) for each pixel.

Results and Discussion

The entire set of aggregated fluoride concentration in groundwaters exhibited a highly skewed log-normal distribution, ranging from less than 0.01 to more than 100 mg L^{-1}. Of the 18,000 data points, only 65 had fluoride concentrations greater than 10 mg L^{-1}. As it was suspected that these samples might have been influenced by anthropogenic processes, we removed them from further analysis. More than 85% of the fluoride data were below the WHO guideline of 1.5 mg L^{-1}, and less than 15% of samples exceeded this guideline (Figure 1).

Process Regions. Kruskal–Wallis and LSD tests revealed that there were significant differences among fluoride concentrations in the delineated process regions ($p < 0.05$) (Figure S3). Summary statistics of fluoride concentration in the eight process regions (PR) are given in the Supporting Information Table S5. The largest geometric means of fluoride concentrations occur in PR1 and PR5. In these regions, 27% and 17% of samples exceed the WHO guideline for drinking water, respectively. Although the geometric mean of fluoride

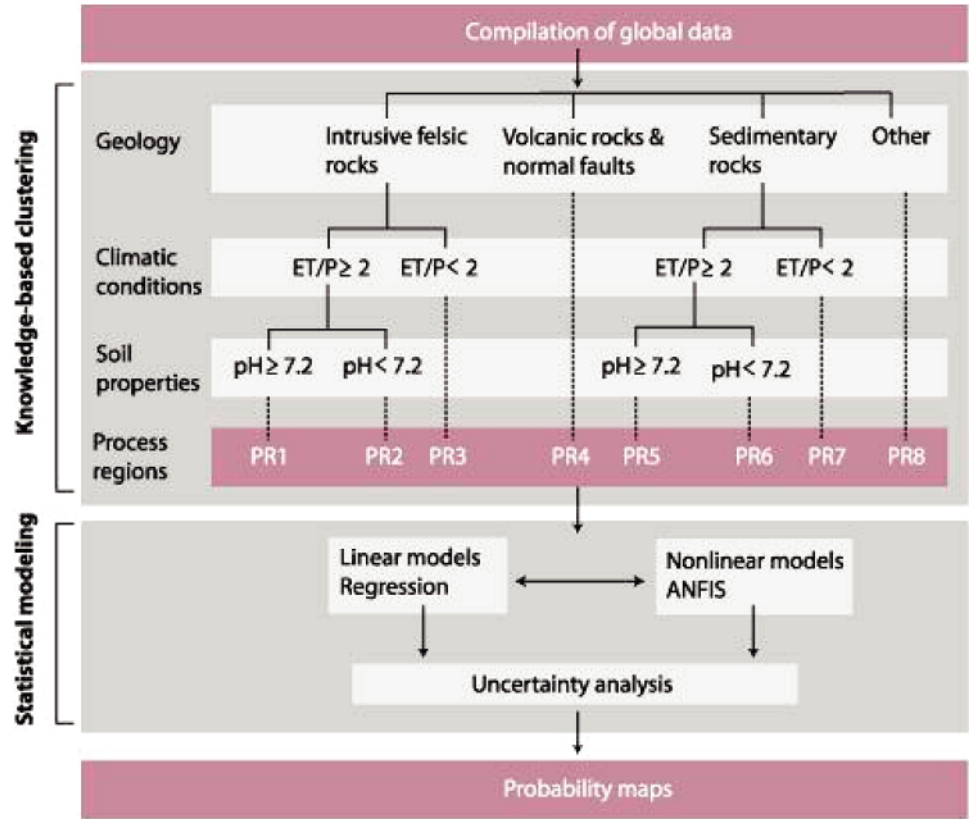

FIGURE 3. Step-by-step procedure for the development of a global probability map of fluoride contamination in groundwater.

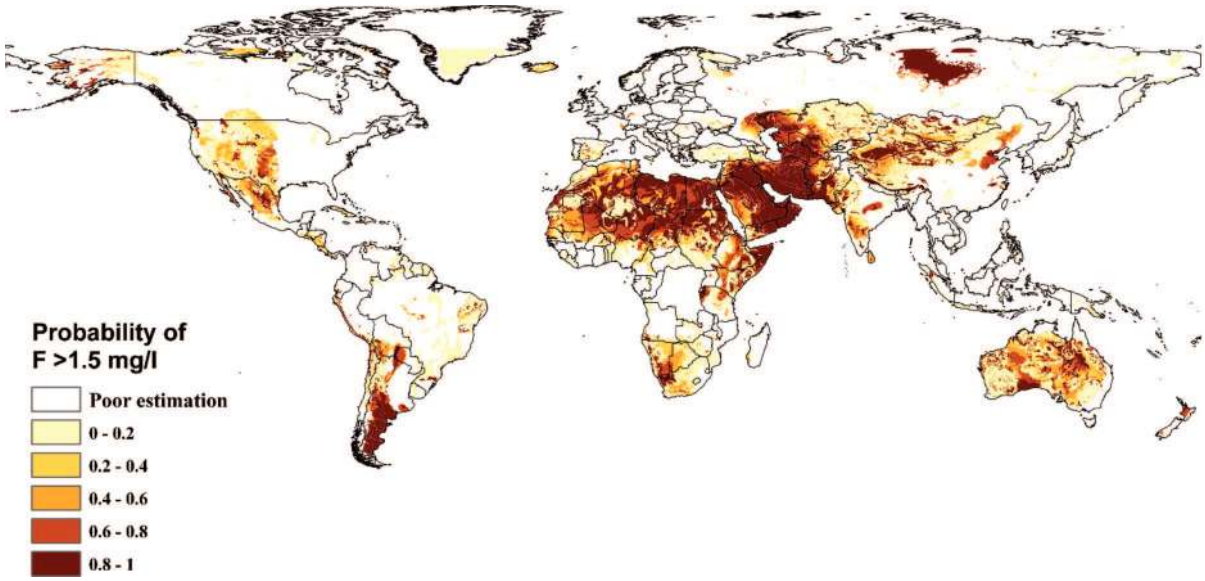

FIGURE 4. Predicted probability of fluoride concentration in the groundwater exceeding the WHO guideline for drinking water of 1.5 mg L^{-1}.

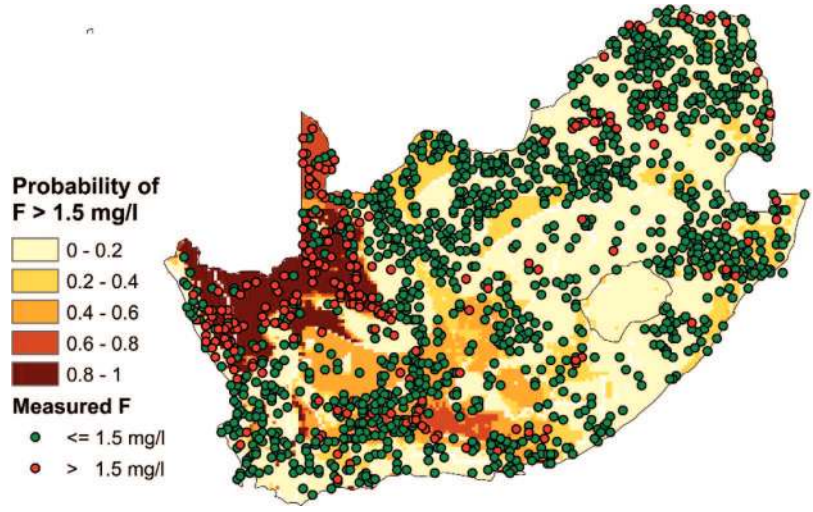

FIGURE 5. Predicted probability map of fluoride concentration in the groundwater exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in South Africa in comparison to measured concentrations. The dots illustrate only the validation data set.

in PR4 is smaller than that in PR5, the percentage of data exceeding the WHO guideline is larger. These results also indicate that even in the regions known for elevated fluoride concentrations (PR1, PR4, and PR5) low concentrations of fluoride are more widespread. Among the delineated regions, PR7 contains the smallest geometric mean of fluoride concentration. Nevertheless, about 3% of the data in this region exceed the WHO guideline.

Although significant differences among the delineated regions correspond to the general geochemical knowledge on fluoride mobilization in groundwaters, there is still a large heterogeneity in each region (Figure S3).

Statistical Modeling. To tackle these heterogeneities we developed predictive models for each of the process regions. The influencing variables determined by stepwise regression for each process region are given in Table S6 in the Supporting Information. The coefficients in the table are standardized regression coefficients, indicating the influence of each variable on the fluoride concentration. For example, in PR1, P and ET/P have almost the same yet opposite effects on the prediction. The results also indicate that in PR1, PR2, PR5, and PR6 climatic parameters have a larger effect on fluoride concentration than soil and geological parameters. The ratio ET/P could be an indicator of groundwater recharge. The larger this ratio, the smaller the groundwater recharge is and the greater is the chance of finding elevated fluoride concentrations in the groundwater. As previously mentioned the regions PR3 and PR7 represent areas with a poor correlation between fluoride content and climate and soil pH values. The modeling was performed nevertheless, but the results show that it is difficult to distinguish the sensitivity of different influencing variables in these regions.

The calibration and validation results of both multiple linear regression (MLR) and ANFIS for different process regions are summarized in the Supporting Information, Table S7. The results show that ANFIS models are significantly better than MLRs in the prediction of fluoride concentrations. Based on the R^2s, between 50 and 70% of the fluoride variation is explained by the surrogate environmental variables used in the analysis. The exceptions, again, are regions PR3 and PR7, which produced the weakest results. This further supports the delineation of the process regions based on the available information. Scatter plots of measurements versus predicted fluoride concentrations for validation data sets in PR1 and PR3 are shown in Supporting Information (Figure S4). These

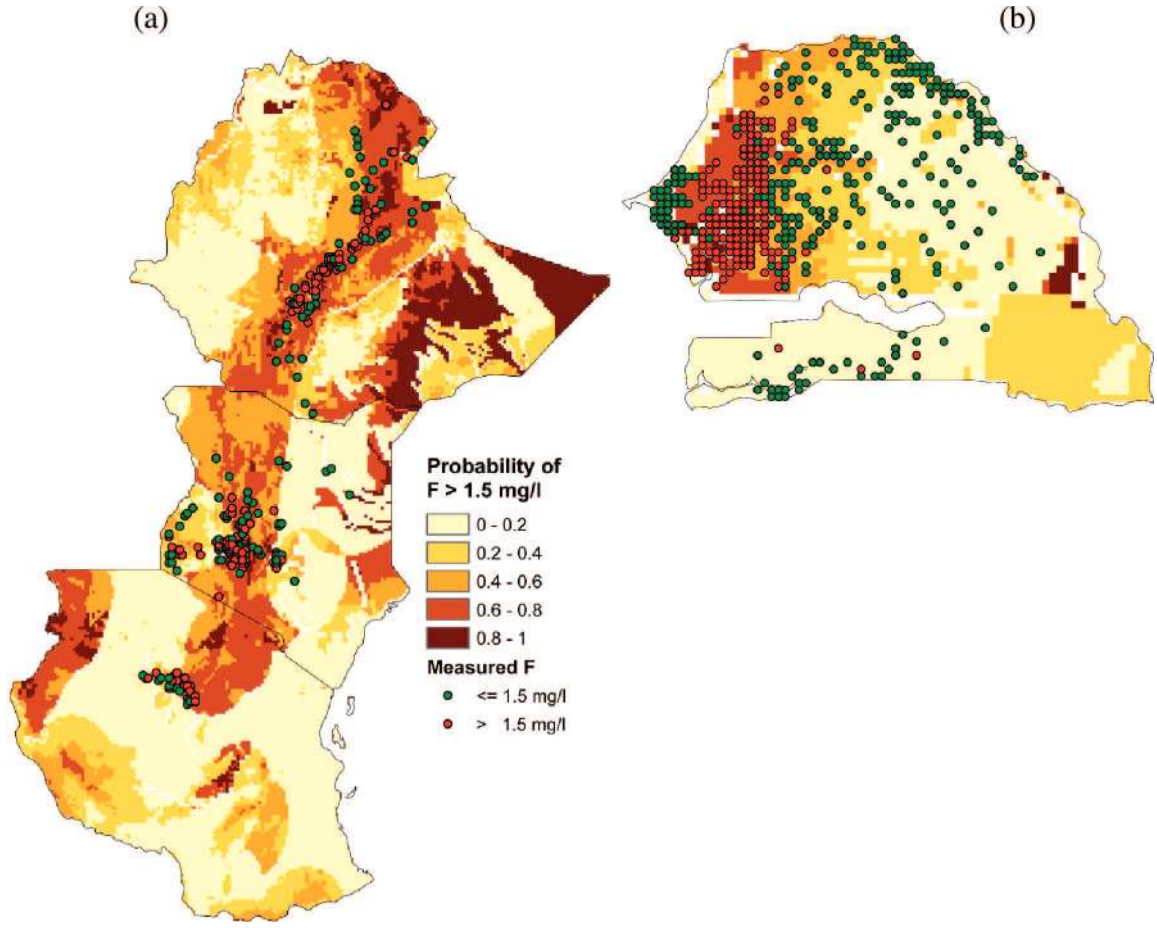

FIGURE 6. Predicted probability map of fluoride concentration exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in comparison with measured concentrations (dots) from Senegal (a) and in Ethiopia, Kenya, and Tanzania (b).

TABLE 2. Distribution of Measured Fluoride Points in Different Probability Classes, Where Each Class Indicates the Probability of Fluoride Concentration Exceeding 1.5 mg L^{-1}

	probability classes									
	<0.2		0.2−0.4		0.4−0.6		0.6−0.8		>0.8	
	total samples	% of samples contaminated	total samples	% of samples contaminated	total samples	% of samples contaminated	total samples	% of samples contaminated	total samples	% of samples contaminated
South Africa	834	7	218	11	175	13	49	53	108	63
Rift valley	7	43	9	56	6	33	18	44	5	80
Senegal	21	14	13	0	31	35	18	78	1	100
U.S.	392	3	43	7	13	15	21	43	4	25
Norway and Sweden	163	13	195	16	224	25	71	21	5	0
Afghanistan[a]	172	1	14	0	0		3	0	29	0
Eritrea[a]	41	5	1	0	0		1	0	2	0
New Zealand[a]	77	1	6	0	3	0	7	0	8	0
Sumatra[a]	76	0	0		11	0	0		0	

[a] There were no data from these countries in the database.

scatter plots illustrate the performance of the models with the largest and the smallest R^2s.

The Probability Map. The global map in Figure 4 shows the probability of fluoride concentrations in groundwater exceeding 1.5 mg L^{-1} except in regions PR3 and PR7. The probability map indicates a belt of high fluoride concentrations stretching from North Africa to the Middle East and toward Pakistan, Uzbekistan, and Kazakhstan. According to the map, densely populated areas in China and India are also affected. With the exception of a few points measured in Iran and India (Figure S1), we had no measured data from this predicted high-fluoride belt, although similar regional patterns are described by Edmunds and Smedley (*3*) and Fawell et al. (*39*). Other countries that seem to be most severely affected are Argentina, Mexico, the United States, South Africa, East African countries, and Australia. These countries are also cited in the international literature as having areas with elevated fluoride concentrations (*14, 21, 40, 41*).

In Figures 5 and 6 we highlight a more detailed picture of model predictions in areas in which we had a large number of measured points. It should be noted that in these figures only validation data points are shown. Figure 5 shows South Africa, where most of the measured points with fluoride concentration greater than the WHO threshold are found,

correctly, in high probability regions. Table 2 gives the distribution of measured points (also only for the validation data set) in different probability classes. The table indicates that, for example, in South Africa, there are 108 points that fall in the probability class where probability of contamination is >0.8. Of these points, 63% are contaminated (i.e., have fluoride concentrations greater than the 1.5 mg L^{-1} threshold). Furthermore, there are 834 points that fall in the class where probability of contamination is <0.2. Of these points, only 7% are contaminated and 93% are uncontaminated. A careful study of Table 2 reveals that the more points there are, the more accurate the representation of the probability map is. Probability maps are, in general, difficult to fully validate as a large number of samples are required for a correct assessment.

Sweden and Norway, in contrast to South Africa, fall in a region of poor model simulation, PR3 (Figure S5). In general, prediction models perform better in arid regions, as indicated by the contaminated belt from North Africa to Central Asia, than in the cold Scandinavian regions. Figure S5 shows the close proximity of contaminated and uncontaminated points, indicating a complex geochemical interaction with the surrounding environment. The other two regions from the validation data set are the Rift Valley and Senegal in Africa (Figure 6), and the United State (Supporting Information, Figure S6), where the predictions quite successfully delineate contaminated fluoride areas.

We compared our results with as many studies as we could find in the published international literature and found a remarkable agreement between model prediction and reported fluorotic areas. For example, we tested the predicted probability map with newly received data (not included in the model) from New Zealand (*42*), Eritrea (*43*), and Sumatra and Afghanistan (*44, 45*). The majority of measured fluoride concentrations in these countries are smaller than 1.5 mg L^{-1}. Distribution of the measured fluoride data in the predicted probability classes are also shown in Table 2. In general, more than 95% of the uncontaminated points fell correctly within the probability class of <0.2. For instance, in Eritrea only 5% of the data falling in the probability class of <0.2 had concentrations greater than 1.5 mg L^{-1}. Graphical comparisons of measured points and probability maps for these countries are provided in the Supporting Information, Figures S7−S9. Apart from the comparisons mentioned above, there is a large area in central Siberia with a high probability of fluoride contamination in groundwater that cannot be verified as no related information is available.

Finally, we would like to point out the main strengths and weaknesses of the maps provided. One of the strengths is that the maps are based on statistically robust models that incorporate a large database of fluoride measurements from different countries as well as global-scale maps of numerous environmental variables. Another strength of the procedure is the combination of statistical methods with geochemical expertise. The main weakness of the modeling lies in the lack of detailed information available on a global scale, particularly for geology, and the uneven distribution of measured fluoride points. More specific geological information would facilitate the characterization of relevant geochemical processes. Another weakness of the model is its current limitation to a 2-dimensional perspective. The risk of fluoride contamination as a function of depth cannot be captured in the present model.

In spite of weaknesses and uncertainties, the modeling approach and the probability map presented here have great potential. Although the map should not replace fluoride testing, it gives a first indication of possible contamination and thus may support the planning process of new drinking water projects. This is particularly important as the demand for drinking water is increasing rapidly due to climate change and population growth, which will undoubtedly increase the pressure on groundwater resources.

Acknowledgments

We are grateful to all the colleagues who have provided us with fluoride data. A list of their contributions appears in the Supporting Information, Table S1. In particular, we would like to acknowledge the help of Ernst Bertram from the Department of Water Affairs and Forestry in South Africa, Bjørn Frengstad from the Geological Survey of Norway, and Dr. Savaghebi from the Department of Soil Science University of Tehran.

Supporting Information Available

Additional details, including descriptions of databases used in this study, nine figures, and seven tables. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Koritnig, S. Ein Beitrag Zur Geochemie Des Fluor. *Geochim. Cosmochim. Acta* **1951**, *1* (2), 89–116.
(2) Jaffery, F. N.; Srivastava, A. K.; Shaw, A.; Viswanathan, P. N.; Seth, P. K. Impact of Nutrition on Fluorosis; Industrial Toxicology Research Centre (ITRC): Lucknow, India, 1998; p 86.
(3) Edmunds, M.; Smedley, P. Fluoride in natural waters. In *Essentials of Medical Geology, Impacts of Natural Environment on Public Health*; Selinus, et al., Eds.; Elsevier Academic Press: London, 2005; pp 301−329.
(4) Tekle-Haimanot, R.; Melaku, Z.; Kloos, H.; Reimann, C.; Fantaye, W.; Zerihun, L.; Bjorvatn, K. The geographic distribution of fluoride in surface and groundwater in Ethiopia with an emphasis on the Rift Valley. *Sci. Total Environ.* **2006**, *367* (1), 182–190.
(5) UNESCO. Trace elements in groundwater and public health. Available at http://www.iah.org/briefings/Trace/trace.pdf.
(6) WHO. *Guidelines for Drinking Water Quality*, 3rd ed.; World Health Organization: Geneva, 2004.
(7) Nielsen, J. M.; Dahi, E. Fluoride exposure of East African consumers using alkaline salt deposits known as magadi (trona) as a food preparation aid. *Food Addit. Contam.* **2002**, *19* (8), 709–714.
(8) Ando, M.; Tadano, M.; Yamamoto, S.; Tamura, K.; Asanuma, S.; Watanabe, T.; Kondo, T.; Sakurai, S.; Ji, R. D.; Liang, C.; Chen, X. Q.; Hong, Z.; Cao, S. Health effects of fluoride pollution caused by coal burning. *Sci. Total Environ.* **2001**, *271* (1−3), 107–116.
(9) ICPS. *Environment Health Criteria 227: Fluorides*; Liteplo, R., Howe, P., Malcolm, H., Eds.; International Programme on Chemical Safety, World Health Organization: Geneva, 2002; p 268.
(10) Ayoob, S.; Gupta, A. K. Fluoride in drinking waters: a review on the status and stress effects. *Crit. Rev. Environ. Sci. Technol.* **2006**, *36*, 433–487.
(11) Wang, L. F. M.; Huang, J. Z. Outline of control practice of endemic fluorosis in China. *Soc. Sci. Med.* **1995**, *41* (8), 1191–1195.
(12) Mjengera, H.; Mkongo, G. Appropriate deflouridation technology for use in flourotic areas in Tanzania. *Phys. Chem. Earth* **2003**, *28* (20−27), 1097–1104.
(13) Díaz-Barrigo, F.; Navarro-Quezada, A.; Grijalva, M. I.; Grimaldo, M.; Loyola-Rodríguez, J. P.; Ortiz, M. D. Endemic fluorosis in Mexico. *Fluoride* **1997**, *30* (4), 233–239.
(14) Kruse, E.; Ainchil, J. Fluoride variations in groundwater of an area in Buenos Aires Province, Argentina. *Environ. Geol.* **2003**, *44* (1), 86–89.
(15) WRC. *Disribution of fluoride-rich groundwater in Eastern and Mogwase region of Northern and North-west province*; WRC Report No. 526/1/011.1-9.85 Pretoria, 2001.
(16) Saxena, V. K.; Ahmed, S. Inferring the chemical parameters for the dissolution of fluoride in groundwater. *Environ. Geol.* **2003**, *43*, 731–736.
(17) Apambire, W. B.; Boyle, D. R.; Michel, F. A. Geochemistry, genesis, and health implications of fluoriferous groundwaters in the upper regions of Ghana. *Environ. Geol.* **1997**, *33* (1), 13–24.
(18) Reddy, N. B.; Prasad, K. S. S. Pyroclastic fluoride in ground water in some parts of Tadpatri Taluk, Anantapur District, Andhra Pradesh. *Indian J. Environ. Health* **2003**, *45* (4), 285–288.

(19) Handa, B. K. Geochemistry and genesis of fluoride-contaning ground waters in India. *Ground Water* **1975**, *13* (3), 275–281.
(20) Hudak, P. F. Fluoride levels in Texas groundwater. *J. Environ. Sci. Health* **1999**, *A34* (8), 1659–1676.
(21) Valenzuela-Vasquez, L.; Ramirez-Henandez, J.; Reyes-Lopez, J.; Sol-Uribe, A.; Lazaro-Mancilla, O. The origin of fluoride in groundwater supply to Hermosillo city, Sonora, Mexico. *Environ. Geol.* **2006**, *51*, 17–27.
(22) Jacks, G.; Bhattacharya, P.; Chaudhary, V.; Singh, K. P. Controls on the genesis of some high-fluoride groundwaters in India. *Appl. Geochem.* **2005**, *20*, 221–228.
(23) Srikanth, R.; Viawanatham, K. S.; Kahsai, F.; Fisahatsion, A.; Asmellash, M. Fluoride in groundwate in selected villages in Eritrea (North East Africa). *Environ. Monit. Assess.* **2002**, *75*, 196–177.
(24) Maithani, P; Gurjar, B. R., et al. Anomalous fluoride in groundwater from western part of Sirohi district, Rajasthan and its crippling effects on human health. *Curr. Sci.* **1998**, *74* (9), 773–777.
(25) Heikens, A.; Sumarti, S.; Van Bergen, M.; Widianarko, B.; Fokkert, L.; Van Leeuwe, K.; Seinen, W. The impact of the hyperacidic Ijen Crater Lake: risks of excess fluoride to human health. *Sci. Total Environ.* **2005**, *346* (1−3), 56–69.
(26) Ashley, R. P.; Burley, M. J. Controls on the occurrence of fluoride in groundwater in the Rift Valley of Ethiopia. In *Groundwater Quality*; Nash, et al., Eds.; Chapman & Hall: London, 1994.
(27) Gaciri, S. J.; Davies, T. C. The Occurrence and Geochemistry of Fluoride in Some Natural-Waters of Kenya. *J. Hydrol.* **1993**, *143* (3−4), 395–412.
(28) Karro, N.; Indermitte, E.; Saava, A.; Haamer, K.; Marandi, A. Fluoride occurence in publicly supplied drinking water in Estonia. *Environ. Geol.* **2006**, *50*, 389–396.
(29) Smedley, P. L.; Nkotagu, H.; Pelig-Ba, K.; McDonald, A. M.; Tyler-Whittle, R.; Whitehead, E. J.; Kinniburgh, D. G Fluoride in Groundwater from High-Fluoride Areas of Ghana and Tanzania; British Geological Survey (BGS), 2002, p 61.
(30) Wang, W. Y.; Li, R. B.; Tan, J. A.; Luo, K. L.; Yang, L. S.; Li, H. R.; Li, Y. H. Adsorption and leaching of fluoride in soils of China. *Fluoride* **2002**, *35* (2), 122–129.
(31) Gupta, S.; Kumar, A.; Ojha, C. K.; Seth, G. Chemical analysis of ground water of Sanganer area, Jaipur in Rajasthan. *J. Environ. Sci. Eng.* **2004**, *46* (1), 74–78.
(32) Sracek, O.; Hirata, R. Geochemical and stable isotopic evolution of the Guarani Aquifer System in the state of Sao Paulo, Brazil. *Hydrogeol. J.* **2002**, *10* (6), 643–655.
(33) Queste, A.; Lacombe, M.; Hellmeier, W.; Hillermann, F.; Bortulussi, B.; Kaup, M.; Ott, K.; Mathys, W. High concentrations of fluoride and boron in drinking water wells in the Muenster region - Results of a preliminary investigation. *Int. J. Hygiene Environ. Health* **2001**, *203* (3), 221–224.
(34) Saxena, V: K.; Ahmed, S. Dissolution of fluoride in groundwater: a water-rock interaction study. *Environ. Geol.* **2001**, *40*, 1084–1087.
(35) Frengstad, B.; Skrede, A. K. M.; Banks, D.; Krog, J. R.; Siewers, U. The chemistry of Norwegian groundwaters: III. The distribution of trace elements in 476 crystalline bedrock groundwaters, as analysed by ICP-MS techniques. *Sci. Total Environ.* **2000**, *246* (1−3), 101–117.
(36) de Caritat, P.; Danilova, S.; Jager, Ø.; Reimann, C.; Storrø, G. Groundwater composition near the nickel-copper smelting industry on the Kola Peninsula, central Barents Region (NW Russia and NE Norway). *J. Hydrol.* **1998**, *208* (1−2), 92–107.
(37) Jang, J. S. R. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Trans. Syst., Man, Cybernetics* **1993**, *23* (3), 665–685.
(38) Hines, J. W. *Matlab Supplement to Fuzzy and Neural Approaches in Engineering;* John Wiley and Sons: New York, 1997.
(39) Fawell, J.; Bailey, K.; Chilton, J.; Dahi, E.; Fewtrell, L.; Magara, Y. *Fluoride in Drinking-water*; World Health Organization (WHO): UK, 2006. Available at http://www.who.int/mediacentre/news/new/2006/nw04/en/index.html.
(40) Díaz-Barriga, F.; Navarro-Quezada, A.; Grijalva, M. I.; Grimaldo, M.; Loyola-Rodríguez, J. P.; Ortíz, M. D. Endemic fluorosis in Mexico. *Fluoride* **1997**, *30* (4), 233–239.
(41) Fitzgerald, J.; Cunliffe, D.; Rainow, S.; Dodds, S.; Hostetler, S.; Jacobson, G. Groundwater quality and environmental health implications, Anangu Pitjantjatjara lands, South Australia; Australia Bureau of Rural Sciences: Canberra, 2000.
(42) Daughney, C. J.; Reeves, R. R. Definition of hydrochemical facies in the New Zealand National Groundwater Monitoring Programme. *J. Hydrol. N. Z.* **2005**, *44* (2), 105–253.
(43) Estifanos, H. *Groundwater Chemistry and Recharge Rate in Crystalline Rocks: Case Study from the Eritran Highland*; Land and Water Resource Engineering: Stockholm, 2005. Available at http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-4085(2007-05-30).
(44) Houben, G. ; Tñnermeier, T. ; Himmelsbach, T. *Hydrogeology of the Kabul Basin Part II: Groundwater geochemistry and microbiology*; Record no. 10277/05; Federal Institute for Geoscience and Natural Resources (BGR), 2003. Available at http://www.bgr.de/app/projektspiegel/fachbeitraege/hydrogeology_kabul_basin_2.pdf.
(45) Broshears, R. E.; Akbari, M. A. ; Chornack, M. P.; Mueller, D. K.; Ruddy, B. C. Inventory of ground-water resources in the Kabul Basin, Afghanistan; USGS Scientific Investigations Report 2005-5090; U.S. Geological Survey, 2005, 34 pp. Available at http://pubs.usgs.gov/sir/2005/5090/.

ES071958Y

Supporting Information
for
Statistical modeling of global geogenic fluoride contamination in groundwaters

Manouchehr Amini[1], Kim Mueller[1], Karim C. Abbaspour[1], Thomas Rosenberg[1], Majid Afyuni[2], Klaus N. Møller[3], Mamadou Sarr[4], Annette C. Johnson[1]

[1]Swiss Federal Institute of Aquatic Science and Technology (Eawag), Ueberlandstrasse 133, CH-8600 Duebendorf, Switzerland

[2] Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, P.O. Box 84154, Iran

[3] Water Resources and Nature Restoration COWI A/S Parallelvej 2, DK-2800 Kongens Lyngby, Denmark

[4] Service de Gestion et de Planification des Resources en Ean (SGPRE), Avenue Andre Peytavin, Ex camp Lat Dior, Dakar, Senegal

Corresponding Author:
Annette C. Johnson
Eawag, Ueberlandstrasse 133, CH-8600 Duebendorf, Switzerland

Supporting Information Content

Table S1 Source of collected fluoride measurements from across the world

Table S2 Source of the databases related to environmental variables

Figure S1 Spatial distribution and histogram of measured fluoride concentration

Table S3 Definition of variables extracted from available global database

Table S4 Ranking of subsoil pH (pH_S)

Adaptive Neuro-Fuzzy Inference System (ANFIS)

Figure S2 Fuzzy classes (a) and ANFIS structure (b).

Figure S3 Box plot of fluoride distribution in different process regions

Table S5 Summary statistic of fluoride concentration in process regions (PR)

Table S6 Selected variables for each region based on stepwise regression

Table S7 Summary of calibration and validation of the models

Figure S4. Scatter plot of measured versus predicted concentrations

Figure S5 Predicted probability map of fluoride concentration in Sweden and Norway .

Figure S6. Probability map for USA.

Figure S7. Probability map for Eritrea.

Figure S8. Probability map for Indonesia (Sumatra).

Figure S9. Probability map for Afghanistan.

References

Table S1.
Summary statistic and sources of compiled fluoride measurements.

Country	Number of samples	Median	Min	Max	% > 1.5 mg L^{-1}	Reference
			mg L^{-1}			
Afghanistan	293	0.23	<0.01	1.71	0.3	1-2
Argentina	137	1.00	<0.01	28.40	41.0	3-4
Australia	195	1.00	0.03	13.00	29.0	5-7
Bangladesh	384	0.19	0.01	1.28	0.0	8
Canada	26	0.11	0.03	6.49	8.0	9
China	25	0.76	0.13	6.78	32.0	10
Eritrea	56	0.72	0.12	3.73	14.8	11-12
Ethiopia	107	1.44	0.13	175.00	49.0	13
Ghana	204	0.46	0.01	4.37	23.0	14
Honduras	55	0.30	0.05	7.50	27.0	15
India	228	0.90	0.10	9.90	26.0	16
Indonesia	260	1.00	<0.01	4.20	11.0	17
Iran†	92	4.70	1.49	41.71	98.0	Current study
Kenia	167	1.80	0.10	40.00	56.0	18
Mexico	54	1.30	0.20	2.27	26.0	19
New Zealand	124	0.05	0.01	5.20	0.8	20
Norway	1341	0.28	0.03	8.50	16.0	21-22
Palestine	12	0.13	0.02	0.56	0.0	23
Senegal	670	1.00	0.05	7.50	39.0	24
South Africa	43'488	0.46	0.05	1140.00	12	25
Sumatra†	102	0.08	0.01	0.73	0	Current study
Sweden	13'719	0.70	<0.01	22.00	22	26
Tanzania	47	1.55	0.16	17.45	51	27
USA	7783	0.13	<0.01	17.90	0.3	28

† Data collected and measured

Table S2.

Sources of the collected global databases related to environmental parameters.

Database	Organization	Source
Digital elevation model (DEM)	USGS	http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html
Geology	CGMW-UNESCO	http://ccgm.free.fr/
	Natural Resources Canada, Geological Survey of Canada	http://gdr.nrcan.gc.ca/minres/data_e.php
Evapotranspiration	FAO-UN	www.fao.org.geonetwork
Precipitation	Tyndall center CRU	www.cru.uea.ac.uk
Temperature	Tyndall center CRU	www.cru.uea.ac.uk
Soil map	FAO	http://www.fao.org/AG/AGL/agll/dsmw.stm
Landuse	USGS	http://edcsns17.cr.usgs.gov/glcc/glcc.html
Runoff	GRDC	www.grdc.sr.unh.edu
Irrigation	FAO	www.fao.org
Stream network	NCCR	http://www.north-south.ch/
Wetlands, Lakes and Reservoirs	Center for Env. Sys. Res., Univ. of Kassel	Lehner and Doll (29)

Table S3.
Definition of variables extracted from available global database. Bold variables were used for knowledge-based statistical rules.

Variable	Definition-unit	Type-format	Resolution
Elevation	(m)	Continuous raster	30 (arc second)
Slope	(degree)	Continuous raster	30 (arc second)
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**
T	Temperature (^{o}C)	Continuous raster	0.5 degree
Runoff	(mm $year^{-1}$)	Continuous raster	0.5 degree
Irrigation	Irrigated areas (%)	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_CbN1	Topsoil 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
Sol_CbN2	Subsoil C/N ratio	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_V_fel	**Distance from volcanic felsic rocks**	**Continuous raster**	**0.5 degree**
Dist_V_maf	Distance from volcanic mafic rocks	Continuous raster	0.5 degree
Dist_V_rest†	Distance from other volcanic rocks	Continuous raster	0.5 degree
Dist_Int_fel	**Distance from intrusive felsic rocks**	**Continuous raster**	**0.5 degree**
Dist_Int_maf	Distance from intrusive mafic rocks	Continuous raster	0.5 degree
Dist_meta	distance from metamorphic rocks	Continuous raster	0.5 degree
Sediment	**Sedimentary rocks**	**Binary variable**	0.5 degree
Dist_fault	**Distance from normal faults**	**Continuous raster**	**0.5 degree**
Dist_Riv	Distance from rivers	Continuous raster	0.5 degree

* 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), N_S ranks from 10 (<0.02 %) to 53 (>0.5 %), OC_S ranks from 10 (<0.2%) to 54 (> 2 %), pH_S ranks from 10 (<4.5) to 54 (>8.5).

† Dist_V_rest consists of volcanic mafic rocks and unclassified volcanic rocks.

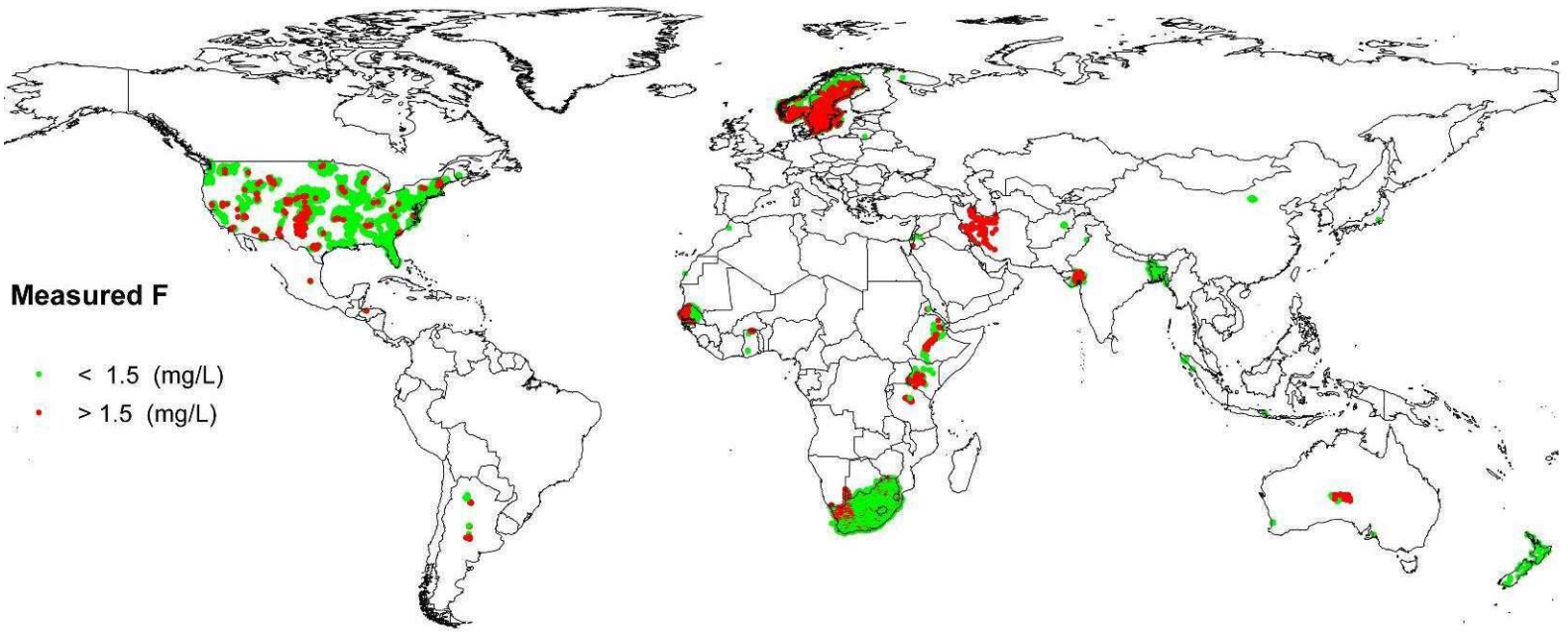

Figure S1
Global distribution of measured fluoride concentration.

Table S4
Subsoil pH (pH_S) ranking according to FAO definition

	FAO definition	
pH_S	Dominant soils	Associated soils
10	< 4.5	
12	< 4.5	>= 4.5 - 5.5
13	< 4.5	> 5.5 - 7.2
14	< 4.5	> 7.2 - 8.5
15	< 4.5	> 8.5
20	>= 4.5-5.5	
21	>= 4.5-5.5	<4.5
23	>= 4.5-5.5	>5.5-7.2
24	>= 4.5-5.5	>7.2-8.5
25	>= 4.5-5.5	>8.5
30	>5.5--7.2	
31	>5.5--7.2	<4.5
32	>5.5--7.2	>=4.5-5.5
34	>5.5--7.2	>7.2-8.5
35	>5.5--7.2	>8.5
40	>7.2-8.5	
41	>7.2-8.5	<4.5
42	>7.2-8.5	>=4.5-5.5
43	>7.2-8.5	>5.5-7.2
45	>7.2-8.5	>8.5
50	>8.5	
51	>8.5	<4.5
52	>8.5	>=4.5-5.5
53	>8.5	>5.5-7.2
54	>8.5	>7.2-8.5
97	Water	

Source: FAO, http://www.fao.org/AG/AGL/agll/dsmw.stm

Adaptive Neuro- Fuzzy Inferencing System (ANFIS)

Assume there is a matrix Y ($N\times D$) where N is the number of points and D is the number of independent variables. Generally, a Sugeno-type fuzzy model with M rules (R_j, j=1,2,...,M) is expressed in the following form:

$$R_j : If\ (y_i \in A_{ij})\ Then \quad z_j = f_j w_j \tag{1}$$

where, A_{ij} are selected fuzzy sets for input variables y_i in j-th rule (Figure S3), w_j is the activation degree (or firing strength), which is a weight indicating the effect of each rule on the prediction, and f_j is a linear function called consequent function for j-th rule. The fuzzy sets A_{ij} for different rules are defined by "subtractive clustering algorithm" (for detail see Jang and Gulley, 1995). The consequent function is defined for each data point and is expressed as:

$$f_j = a_0 + \sum_{i=1}^{D} a_i y_i \qquad j = 1,...,M \tag{2}$$

where a_i's are referred to as consequent parameters that must be optimized. The firing strength for each rule could be calculated by the "min" or "product" operator of membership grades for each rule as follows:

$$\begin{aligned} w_j &= \min(\mu_{ji}(y_i)) \\ w_j &= \prod_{i=1}^{D} \mu_{ji}(y_i) \end{aligned} \tag{3}$$

where μ_{ji} is the membership grade resulting from a membership function that describes i-th variable in j-th rule. The membership grade depends on the type and

parameters of membership function. Among the numerous existing membership functions we used Gaussian membership function as this is more commonly used.

$$\mu_{ij} = \exp\left(\frac{-(y_i - c_{ij})^2}{2\sigma_{ij}^2}\right) \quad for \;\; i = 1,...,D \qquad j = 1,...,M \tag{4}$$

The parameters c_{ij} and σ_{ij} are the nonlinear premise parameters of membership functions to be optimized.

The final output for each data point is then calculated as a normalized weighted average of the outcome of different rules as expressed below:

$$Z = \sum_{j=1}^{M} f_j \overline{w}_j \tag{5}$$

where $\overline{w}_j$ is the normalized activation function expressed as:

$$\overline{w}_j = \frac{w_j}{\sum_{j=1}^{M} w_j} \tag{6}$$

In summary, there are two key steps in constructing an ANFIS. In the first step fuzzy sets describing different rules are constructed (Fig.S3a), and in the second step premise and consequent parameters are optimized (Fig. S3b). In this optimization, an Adaptive Learning Algorithm consisting of four layers were applied.

- Layer 1: This layer provides membership grades (μ_{ji}) for different input variables in different rules (Eq. 4).

- Layer 2: Aggregation of membership grades for different rules (Eq. 3).

- Layer 3: Normalizing activation functions (Eq. 6).

- Layer 4: Evaluation of output by applying equation (Eq. 5)

To accomplish the above steps, we used the Fuzzy toolbox of Matlab (ver.7.04).

To optimize linear consequent and nonlinear premise parameters of the FIS, a hybrid method was applied. The hybrid method is a combination of gradient descent and least square approaches. According to this method first the nonlinear parameters (premise) are fixed and the consequent parameters are estimated by least square method. Then the premise parameters are updated by gradient decent optimization algorithm. This procedure is repeated until a specific optimization criterion is reached (for more detail see references 30-33).

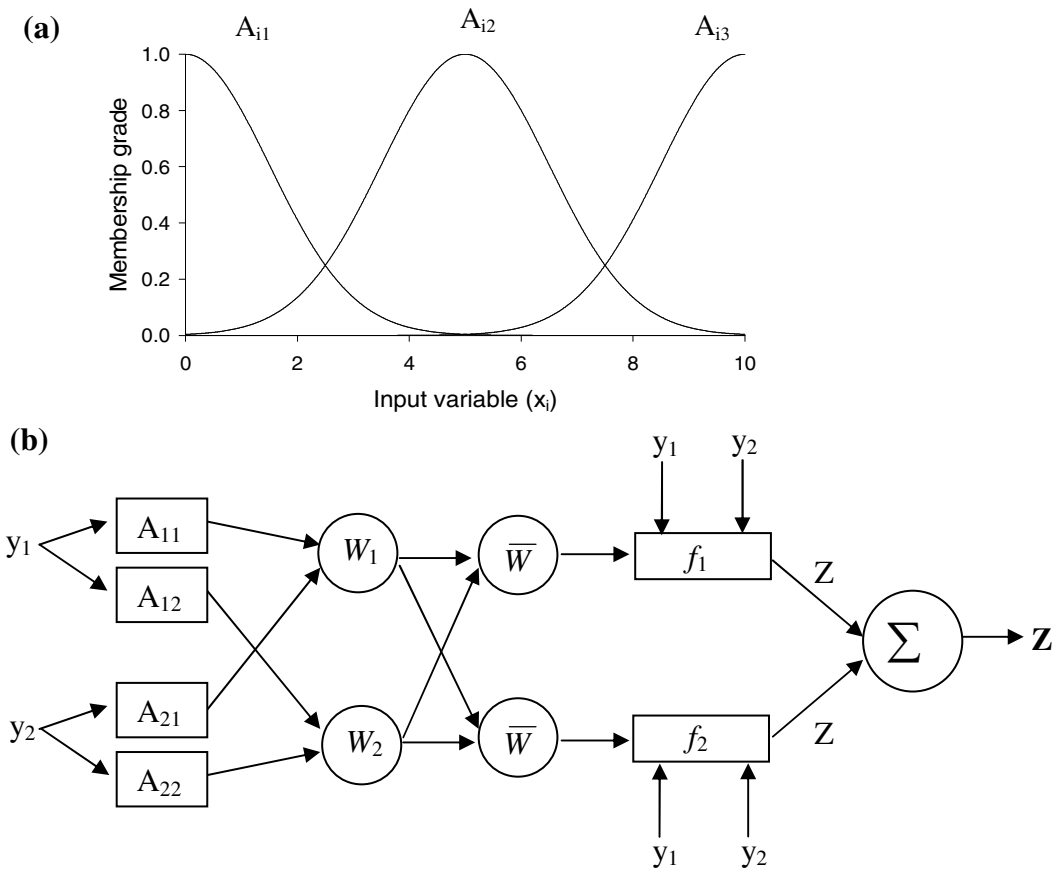

Figure S3
Fuzzy classes (a) and ANFIS structure (b).

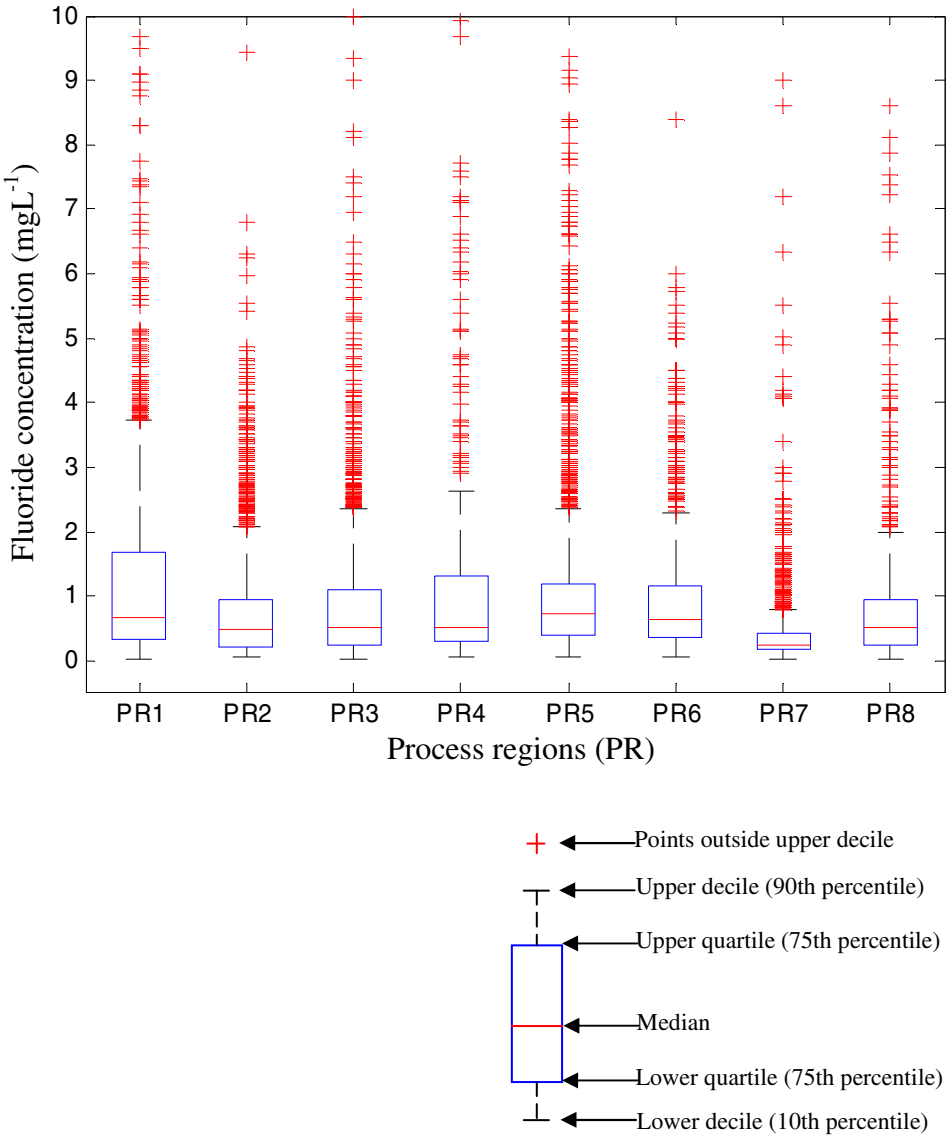

Figure S2
Box plot of fluoride distribution in different process regions

Table S5
Summary statistic of fluoride concentration in process regions (PR)

Region	PR1	PR2	PR3	PR4	PR5	PR6	PR7	PR8	Total
Number	3226	2337	2256	6086	1298	720	1822	634	18300
Minimum	0.03	0.05	0.02	0.05	0.05	0.05	0.02	0.02	0.02
Average[†]	1.21[a]	0.77[b]	0.80[b]	1.22[c]	1.06[a]	0.97[c]	0.41[d]	0.82[b]	0.87
Gmean[*]	0.72	0.47	0.50	0.65	0.72	0.64	0.28	0.47	0.53
Std[**]	1.32	0.85	0.84	1.67	1.17	0.99	0.59	1.05	1.02
Maximum	9.69	9.43	10.00	9.92	9.39	8.38	9.01	8.60	10
Skewness	2.23	2.67	3.04	2.53	3.16	2.37	7.08	3.62	3.15
%>1.5	27.68	13.74	14.31	21.78	17.43	17.06	3.23	12.32	15.56

[†] Same characters indicate the groups are not significantly different ($p < 0.05$).

[*] Geometric mean

[**] Standard Deviation

Table S6

Selected variables for each region based on stepwise regression (the numbers are regression coefficients except for intercept).

Variables	PR1	PR 2	PR 3	PR 4	PR 5	PR 6	PR 7	PR 8
Intercept	-0.32	-1.40	0.04	0.73	-1.19	-1.18	-1.03	-0.68
Elevation	0.08	0.09	-0.16	-	0.06	-	0.17	0.27
Slope	-0.08	-0.07	-	-	-0.05	-0.10	-0.07	-
ET[†]	-	-0.35	-	-0.26	0.36	0.65	-0.25	-
P	-0.23	-	-	-	-0.24	0.21	-	-0.80
ET/P	0.24	0.31	-0.14	-	-	0.20	0.40	-
T	-	0.56	0.11	0.37	-	-0.95	0.34	-
Runoff	-	-0.06	-0.12	-	0.14	-0.10	-	0.57
Irrigation	-	-	-0.06	-	-0.07	-	-	0.08
Sand1	-	-	-0.07	-	-	-	-	-
Silt1	-	0.25	-	-	0.14	-	-	-
Clay1	-	-	-	-	-0.13	-	-	-0.20
Sand2	-	-	-	-	-	-	-0.10	-
Clay2	-	-	0.13	-	-	0.27	0.09	-
Sol_CbN1	-	-	0.06	-	-0.12	-	-	-
Sol_CbN2	-0.27	-	-	-	-	-0.15	-	-
Drain_code	-	-0.23	-	-0.11	0.21	-	-	0.21
pH_S	-	0.18	-	-	-0.18	-0.15	-	-
OC_S	0.21	0.25	-	0.23	-	-	-	-
N_S	-	-0.13	-	-	-	-	-	-
CEC_S	-	-0.14	-0.18	-0.41	-	-0.16	-	-
Dist_V_rest	0.16	0.35	0.26	-0.19	-0.05	-0.34	-	-
Dist_V_fel	0.16	-	-0.17	-	0.18	0.24	-	0.37
Dist_int_maf	-	-	0.07	0.31	0.24	0.38	0.46	-
Dist_int_fel	-0.19	-0.08	-0.04	-	0.11	0.21	-	0.16
Dist_meta	-	-	-	-	0.18	-	-	-
Dist_fault	-	-0.39	-0.08	-0.30	-	0.83	0.19	-
Dist_Riv	-	-	-	-0.19	-	-	-	-

[†] Definition of variables provided in Table S3.

Table S7
Summary of calibration and validation of models for different process regions (PR).

Region		**PR1**	**PR2**	**PR3**	**PR4**	**PR5**	**PR6**	**PR7**	**PR8**
		ANFIS*							
Calibration	R^2	0.65	0.59	0.31	0.66	0.64	0.70	0.29	0.50
	ME**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MSE†	0.07	0.08	0.13	0.07	0.05	0.05	0.07	0.11
Validation	R^2	0.63	0.59	0.31	0.62	0.60	0.59	0.33	0.56
	ME	-0.02	-0.01	-0.01	0.01	-0.01	-0.03	0.00	-0.01
	MSE	0.07	0.08	0.12	0.08	0.06	0.07	0.06	0.07
		Regression							
Calibration	R^2	0.49	0.43	0.22	0.47	0.43	0.39	0.15	0.28
	ME	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MSE	0.10	0.11	0.15	0.11	0.08	0.09	0.09	0.15
Validation	R^2	0.46	0.47	0.20	0.49	0.36	0.28	0.17	0.30
	ME	0.00	0.00	0.00	0.00	-0.01	-0.03	-0.01	0.00
	MSE	0.11	0.11	0.14	0.10	0.09	0.12	0.08	0.12

* Adaptive neuro-fuzzy inference system, ** Mean error, † Mean square error

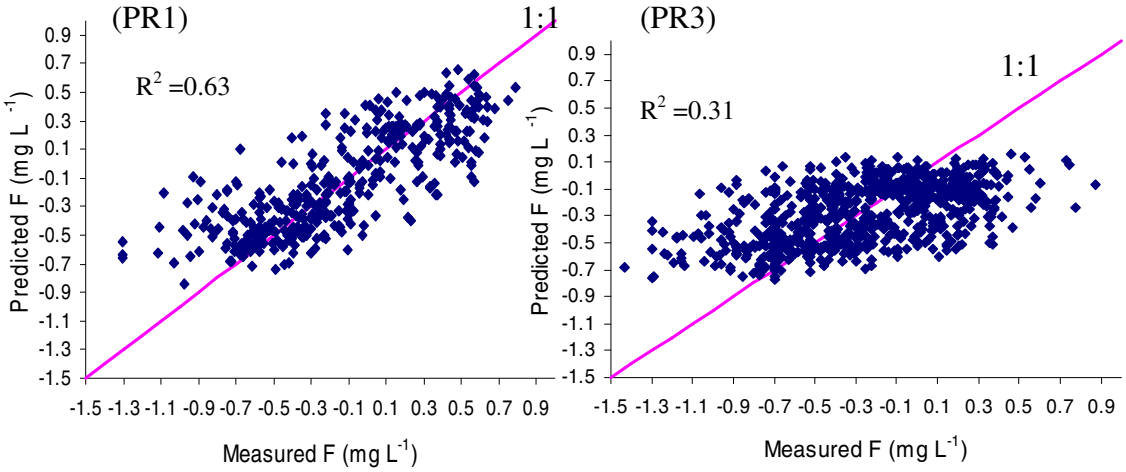

Figure S4.
Scatter plot of measured versus predicted fluoride concentrations of validation dataset in PR1 and PR3. The data were log transformed.

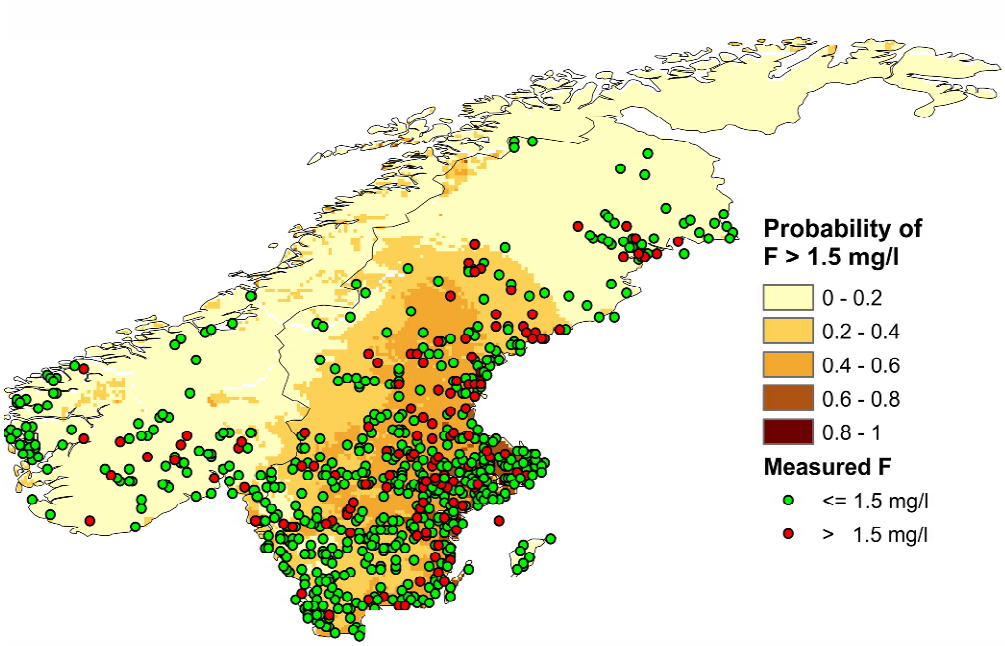

Figure S5
Predicted probability map of fluoride concentration in the groundwater exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in Sweden and Norway in comparison to measured concentrations. The dots illustrate only the validation dataset.

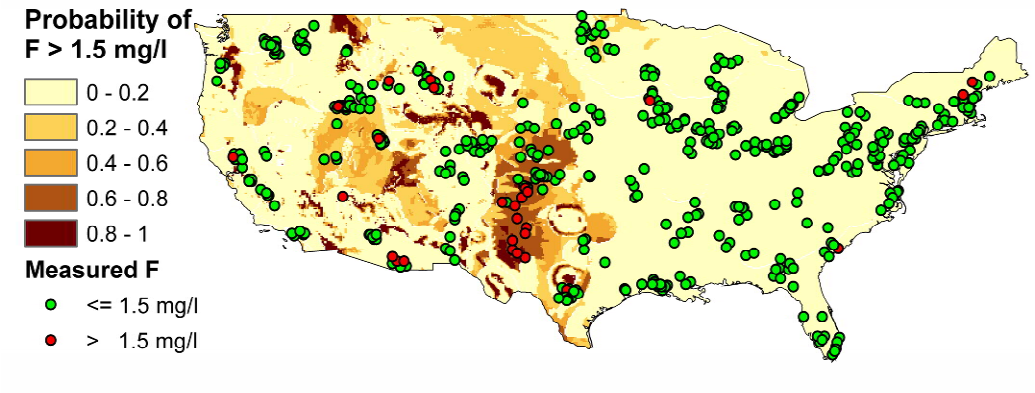

Figure S6.
Predicted probability map of fluoride concentration in the groundwater exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in USA in comparison to measured concentrations. The dots illustrate only the validation dataset.

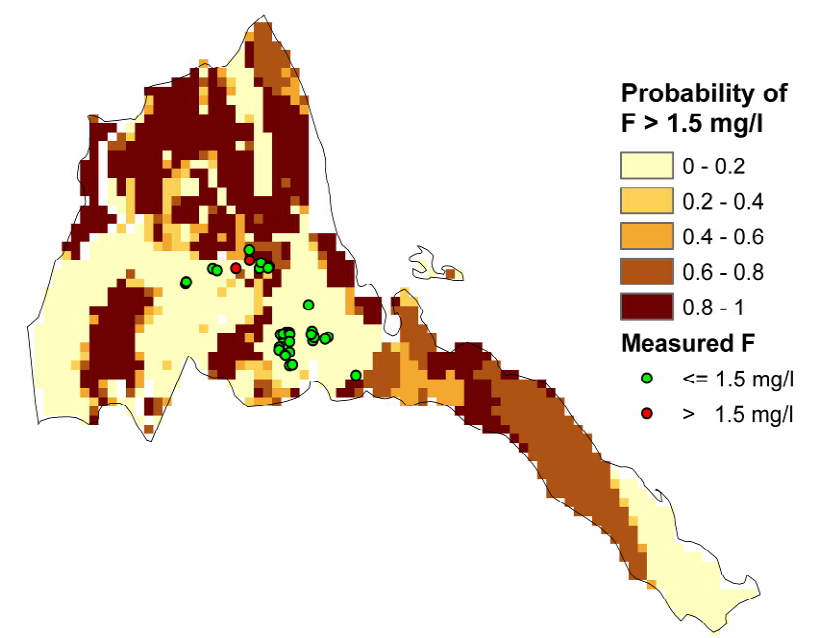

Figure S7.
Predicted probability map of fluoride concentration exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in comparison with measured concentrations (dots) from Eritrea.

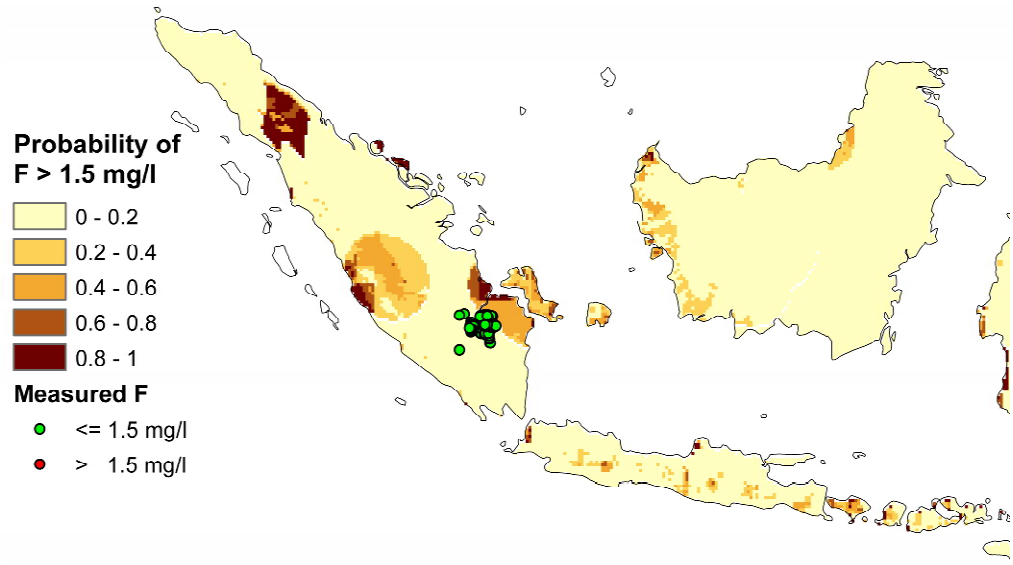

Figure S8.
Predicted probability map of fluoride concentration exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in comparison with measured concentrations (dots) from Indonesia (Sumatra).

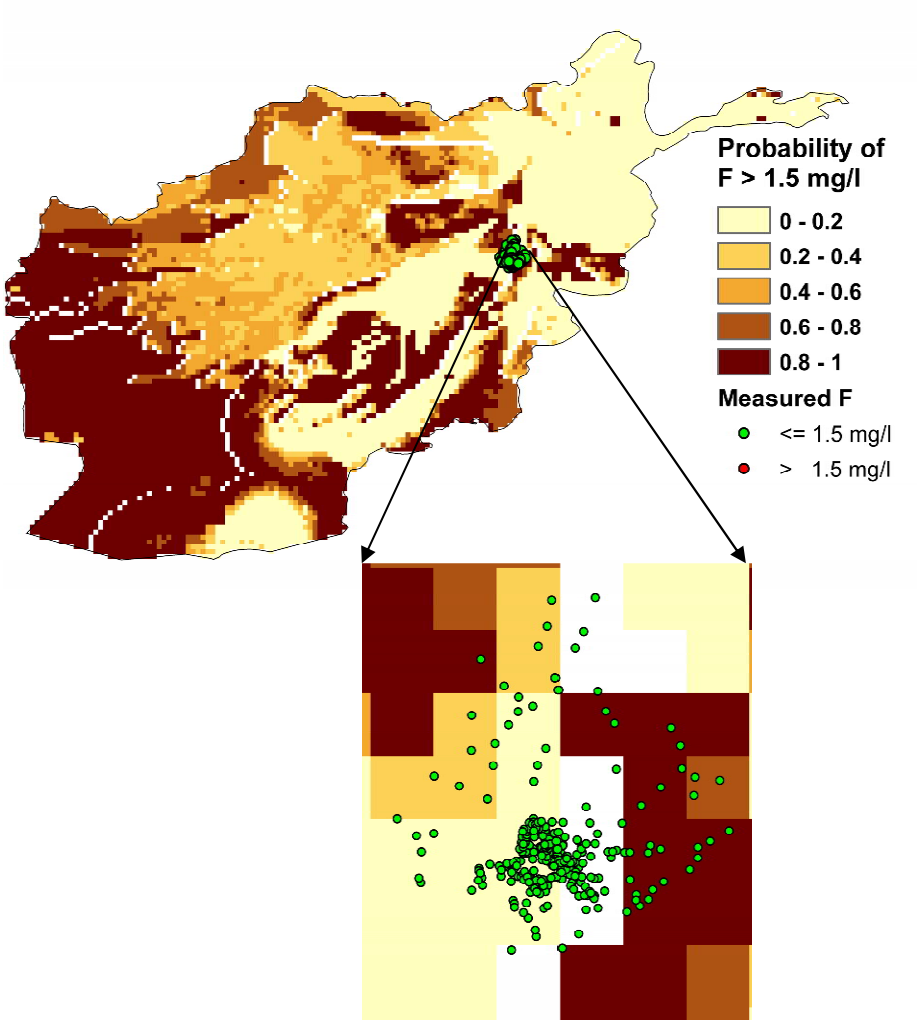

Figure S9.
Predicted probability map of fluoride concentration exceeding the WHO guideline for drinking water of 1.5 mg L^{-1} in comparison with measured concentrations (dots) from Afghanistan.

References

(1) Houben, G.; Tünnermeier, T.; Himmelsbach, T. *Hydrogeeology of the Kabul Basin Part II: Groundwater geochemistry and microbiology*, Federal Institute for Geoscience and Natural Resources (BGR), Record no. 10277/05, 2003. Available at: http://www.bgr.de/app/projektspiegel/fachbeitraege/hydrogeology_kabul_basin_2.pdf.

(2) Broshears, R. E.; Akbari, M. A.; Chornack, M. P.; Mueller, D. K.; Ruddy, B. C. *Inventory of ground-water resources in the Kabul Basin, Afghanistan*: U.S. Geological Survey Scientific Investigations Report 2005-5090, 34 p. Available at: http://pubs.usgs.gov/sir/2005/5090/.

(3) Smedley, P. L.; Nicolli, H. B.; Macdonald, D. M. J.; Barros, A. J.; Tullio, J. O. Hydrogeochemistry of arsenic and other inorganic constituents in groundwaters from La Pampa, Argentina. *Appl. Geochem.* **2002**, 17(3), 259-284.

(4) Nicolli, H. B., J. M. Suriano, et al. Groundwater Contamination with Arsenic and Other Trace-Elements in an Area of the Pampa, Province of Cordoba, Argentina. *Environ. Geol. Water Sci.* **1989**, 14(1), 3-16.

(5) Fitzgerald, J.; Cunliffe, D.; Rainow, S.; Dodds, S.; Hostetler, S.; Jacobson, G. *Groundwater quality and environmental health implications, Anangu Pitjantjatjara lands, South Australia*. Australia, Bureau of Rural Sciences, Canberra, 2000.

(6) Ivkovic, K. M.; Watkins, K. L.; Cresswell, R. G.; Bauld, J. *A Groundwater Quality Assessment of the Fractured Rock Aquifer of the Picadilly Valley, South Australia*. Australian Geological Survey Organisation, Record 1998/16, 1998.

(7) Larsen, R. M.; Watkins, K. W.; Steel, N. A.; Appleyard, S. J.; Bauld, J. *A Groundwater Quality Assessment of the Jandakot Mound, Swan Coastal Plain, Western Australia*. Australian Geological Survey Organisation, Record 1998/18, 1998.

(8) Kinniburgh, D. G.; Smedley P. L. *Arsenic contamination of groundwater in Bangladesh*, Volume 2: final report, British Geological Survey, Keyworth, 2001.

(9) Boyle, D. R.; Spirito, W. A.; Adcock, S. W. *Groundwater hydrogeochemical survey of central New Brunswick*. Geological Survey of Canada, Open File 3306, 1996. http://gdr.nrcan.gc.ca/geochem/metadata_pub_e.php?id=00241

(10) Smedley, P. L.; Zhang, M.; Zhang, G.; Luo, Z. Mobilisation of arsenic and other trace elements in fluviolacustrine aquifers of the Huhhot Basin, Inner Mongolia. *Appl. Geochem.* **2003**, 18(9), 1453-1477.

(11) Srikanth, R.; Viawanatham, K. S.; Kahsai, F.; Fisahatsion, A.; Asmellash, M. Fluoride in groundwate in selected villages in Eritrea (North East Africa). *Environ. Monit. Assess.* **2002**, 75, 196-177.

(12) Estifanos, H. *Groundwater chemistry and recharge rate in crystalline rocks: case study from the Eritran highland*. Stockholm: Land and Water Resource Engineering, 2005, Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-4085 (2007-05-30)

(13) Reimann, C. B. K.; Tekle-Haimanot, R.; Melaku, Z.; Siewers, U. *Drinking Water Quality, Rift Valley, Ethiopia*. NGU-Report, Report no. 2002 033 Trondheim, Norway, 2002.

(14) Smedley, P. L. Arsenic in rural groundwater in Ghana. *Journal of African Earth Sciences,* **1996**, 22(4), 459-470.

(15) Foletti, C. Diagnostico de flurosis dental en 39 communidades del Valle de Sula, Honduras, 2001.

(16) Gupta S., Kumar A., Ojha C. K., and Seth G. Chemical analysis of ground water of Sanganer area, Jaipur in Rajasthan. *J. Environ. Sci. Engin.* **2004**, 46(1), 74-78.

(17) Heikens A., Sumarti S., Van Bergen M., Widianarko B., Fokkert L., Van Leeuwen K., and Seinen W. The impact of the hyperacidic Ijen Crater Lake: risks of excess fluoride to human health. *Sci. Tot. Environ.*, **2005**, 346(1-3), 56-69.

(18) Näslund, J.; Snell, I. *GIS-mapping of fluoride contaminated groundater in Nakuru & Baringo district, Kenya*. Department of Civil and Environmental Engineering. Lulea, Lulea University of Technology, 2005.

(19) Planer-Friedrich, B. Hydrogeological and hydrochemical investigations in the Rioverde basin, Mexico. Institute of Geology, University of mining and technology Freiberg, 2000.

(20) Daughney, C. J.; Reeves, R. R. Definition of hydrochemical facies in the New Zealand National Groundwater Monitoring Programme. *J. Hydrol. N. Z.* **2005**, 44 (2), 105-253.

(21) Frengstad, B.; Skrede, A. K. M.; Banks, D.; Krog, J. R.; Siewers, U. The chemistry of Norwegian groundwaters: III. The distribution of trace elements in 476 crystalline bedrock groundwaters, as analysed by ICP-MS techniques. *Sci. Tot. Environ.* **2000**, 246(1-3), 101-117.

(22) de Caritat, P.; Danilova, S.; Jager, Ø.; Reimann, C.; Storrø, G. Groundwater composition near the nickel-copper smelting industry on the Kola Peninsula, central Barents Region (NW Russia and NE Norway). *Journal of Hydrology*, **1998**, 208(1-2), 92-107.

(23) Qannam, Z. A hydrogeological, hydrochemical and environmental study in Wadi Al Arroub drainage basin, south west Bank, Palestine, Diploma thesis, Freiberg *On-line Geosciences* **2003** Vol. 9, http://www.geo.tu-freiberg.de/fog/FOG_Vol_9.pdf, (3.2007).

(24) République du Sénégal, Ministère de l'Energie et de l'Hydraulique, Service de Gestion et de Planification des Ressources en Eau (SGPRE). Synthèse des données géochimiques. Interprétations en terme de modèle conceptuel des écoulements, 70, 2001.

(25) McCaffrey, L. P.; Willis, J. P. Distribution of flouride-rich groundwater in the eastern and Mogwase regions of the Northern and North-west Provinces: influence of bedrock and soils and constraints on utilisable drinking water supplies. Rondebosch 7700, South Africa, Department of Geological Sciences, University of Cape Town, 2001.

(26) Jonas Gierup, Personal comunication, Swedish Geological Survey 2007, (Jonas.Gierup@sgu.se).

(27) Smedley P. L., Nkotagu H., Pelig-Ba K., McDonald A. M., Tyler-Whittle R., Whitehead E. J., and Kinniburgh D. G. *Fluoride in groundwater from high-fluoride areas of Ghana and Tanzania*, pp. 61. British Geological Survey (BGS), 2002.

(28) USGS, NWQAD Warehouse, 1991-2003. Available at: infotrek.er.usgs.gov

(29) Lehner, B.; Doll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol*. **2004**, 296(1-4), 1-22.

(30) Steeb, W. H.; Hardy, Y. *The nonlinear workbook* : *chaos, fractal, cellular automata*. 3rd ed. World Scientific, New Jersey, 2005.

(31) Hines, J. W. *Matlab supplement to fuzzy and neural approaches in engineering*. John Wiley and Sons, New York, USA, 1997.

(32) Jang, J. S. R. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics,* **1993**, 23, 3, 665-685.

(33) Jang, J. S. R.; Gulley, N. *The fuzzy logic toolbox for use with MATLAB*, The Mathworks Inc. 1995.