Arsenic in the Environment: Bridging Science to Practice for Sustainable Development –
van der Wal, Ahmad, Petrusevski, Weijma, Savic, van der Wens, Beerendonk, Bhattacharya, Bundschuh & Naidu (Eds)

Pseudo contour maps from logistic regression modelling: Case study of groundwater arsenic distribution in Gujarat state, India

R. Wu[1], J. Podgorski[1,2], M. Berg[2] & D.A. Polya[1]
[1]*Department of Earth and Environmental Sciences, & Williamson Research Centre for Molecular Environmental Science, University of Manchester, Manchester, UK*
[2]*Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland*

ABSTRACT: Logistic regression modelling is a widely used prediction method to generate probability maps of groundwater arsenic exceeding a selected threshold concentration. In this study, we converted logistic regression modelled probability maps to a pseudo contour map which comprehensively integrated important information of several probability maps of groundwater arsenic concentrations in Gujarat state, India. The pseudo contour map of groundwater arsenic concentrations clearly and intuitively indicates high groundwater arsenic concentrations (>10 μg/L) occurring in Kachchh District and Banas Kantha District and shows groundwater arsenic concentrations in northern and central Gujarat to be slightly higher than near the western and eastern borders of the state. The cutoffs, where sensitivity equals specificity, is one of the most appropriate criteria of diagnosis of continuous probability results, better ensuring the accuracy of the generation of a pseudo contour map. However, the probability cutoff of each contour is likely different, so that the generated map is pseudo-contour rather than a contour map.

1 INTRODUCTION

High arsenic groundwaters pose adverse health effects for hundreds of millions of people worldwide (Bhattacharya 2017). Rational prediction of the distribution of arsenic contamination in groundwaters is an important tool to managers to prevent people's exposure to arsenic. Logistic regression modelling has been widely used to explore the statistical relationship between dependent variables (groundwater arsenic) and other environmental independent variables in order to predict the distribution of arsenic in groundwaters. Probability maps are a common way to present the prediction results of logistic regression models. In this article, we develop a new way of integrating logistic regression model results to generate a pseudo contour map of groundwater arsenic concentrations, in this case, in Gujarat state, India.

2 MATERIALS AND METHODS

2.1 *Study area and groundwater arsenic*

The Gujarat State is situated between North latitudes 20°06' 00" to 24°42' 00" and East longitudes 68°10' 00" to 74°28' 00" (Figure 1), with an area of 196,024 km^2. (CGWB 2016). Groundwater arsenic data were obtained from CGWB (2016).

2.2 *Probability maps of different arsenic concentrations*

Five different thresholds for logistic regression models were set, including 10 μg/L, 5 μg/L, 4 μg/L, 3 μg/L, and 2 μg/L, since 10 μg/L is the WHO provisional guide value for arsenic in drinking water and most arsenic concentrations are below 10 μg/L in the dataset. Five probability maps, each with the different selected threshold was produced based on the best performing logistic regression models.

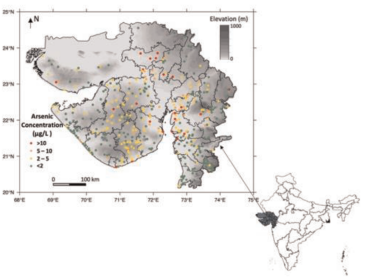

Figure 1. Study area and the distribution of groundwater arsenic concentrations. (data from CGWB 2016).

2.3 *Determining the cutoffs for converting probability maps to a pseudo contour map*

The sensitivity, accuracy, and specificity of the best performing logistic regression models were plotted against cutoff which is in range of 0 to 1. Sensitivity and specificity lines intersected at a certain cutoff which was then used to compare with the probabilities calculated by logistic regression in order to determine whether arsenic concentrations exceeded the given threshold (Podgorski *et al.*, 2017).

DOI: 10.1201/9781003317395-40

2.4 *Creation of pseudo contour map*

Converting probability maps of groundwater arsenic concentrations exceeding selected thresholds to a pseudo contour map is, to the best of our knowledge, a new method in this context. The areas whose probabilities exceeded the selected cutoff were considered as the zone of occurrence of groundwater arsenic concentrations exceeding the selected threshold concentration. Combining occurrence zones of different groundwater arsenic concentrations generated a pseudo-contour map of groundwater arsenic concentrations.

3 RESULTS AND DISCUSSION

3.1 *The cutoffs of converting probability maps to a pseudo contour map*

Probability maps of arsenic concentration exceeding 10 µg/L, 5 µg/L, 4 µg/L, 3µg/L and 2 µg/L are shown in Figure 2. The sensitivity, accuracy, and specificity of the best performing logistic regression models were plotted against cutoffs (Figure 3). The cutoffs (10 µg/L: 0.69, 5 µg/L: 0.66, 4 µg/L: 0.61, 3µg/L: 0.57 and 2 µg/L: 0.50) where sensitivity intersects with specificity were used to create a pseudo map of arsenic concentrations in groundwaters.

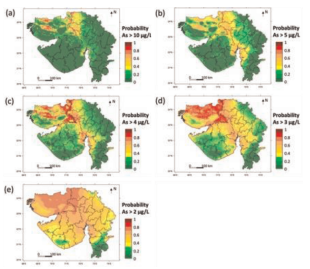

Figure 2. Probability map of arsenic concentration exceeding thresholds of (a) 10 µg/L, (b) 5 µg/L, (c) 4 µg/L, (d) 3 µg/L, and (e) 2 µg/L in Gujarat.

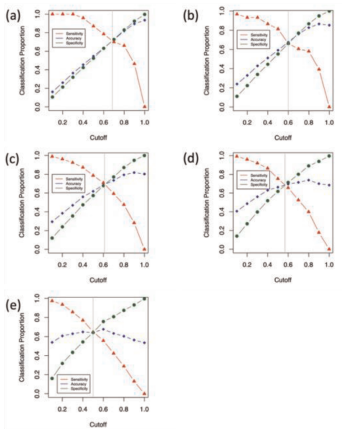

Figure 3. The sensitivity, accuracy, and specificity of logistic regression models plotted against cutoffs.

3.2 *Pseudo contour map of the distribution of groundwater arsenic in Gujarat state*

The pseudo contour map of arsenic concentrations in Gujarat groundwaters is shown in Figure 4. High groundwater arsenic concentrations (>10 µg/L) occur in Kachchh District and Banas Kantha District. Groundwater arsenic concentrations in northern and central Gujarat are slightly higher than in the western and eastern borders. The pseudo contour map shows the distribution of arsenic clearly and intuitively.

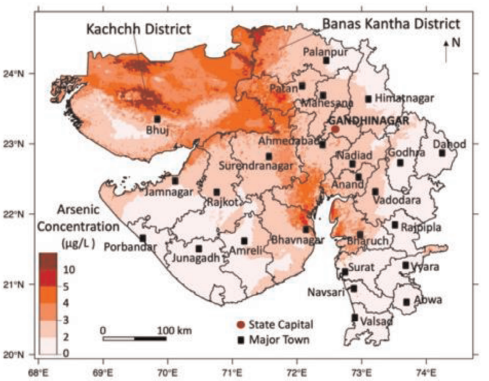

Figure 4. Pseudo contour map of arsenic concentrations in groundwaters in Gujarat. Raw data from CGWB (2016).

4 CONCLUSIONS

The pseudo contour map can integrate information from several probability maps using different thresholds in order to clearly display the distribution of different groundwater arsenic concentrations in one map. The cutoffs where sensitivity equals specificity is one of the most appropriate criteria of diagnosis of continuous probability results, better ensuring the accuracy of the generated pseudo contour map. However, the probability cutoff of each contour is likely different, so that generated map is not really a contour map but rather a pseudo-contour map. In this study, the pseudo contour map clearly and intuitively indicates the distribution of high groundwater arsenic concentrations (>10 µg/L) in a state, Gujarat, with largely low reported groundwater arsenic concentrations.

REFERENCES

Bhattacharya P., Polya D. & Jovanovic D. (eds.). 2017. *Best Practice Guide on the Control of Arsenic in Drinking Water*. IWA Publishing.

CGWB-Central Ground Water Board. 2016. Groundwater Year Book – 2015 – 2016 Gujarat state and UT of Daman & Diu.

Podgorski J.E., Eqani S.A.M.A.S., Khanam T., Ullah R., Shen H. & Berg M. 2017. Extensive arsenic contamination in high-ph unconfined aquifers in the Indus Valley. *Sci. Adv.* 3(8): e1700935.