The use of cokriging algorithm for arsenic mapping in groundwater systems

O. Gunduz[1], A. Elci[2,] C. Simsek[3], A. Baba[4]

[1] Dokuz Eylul University, Department of Environmental Engineering, Izmir, TURKEY
Tel: +90 (232) 4127141, Fax: +90 (232) 4531143, Email: orhan.gunduz@deu.edu.tr
[2] Dokuz Eylul University, Department of Environmental Engineering, Izmir, TURKEY
Tel: +90 (232) 4127112, Fax: +90 (232) 4531143, Email: alper.elci@deu.edu.tr
[3] Dokuz Eylul University, Torbali Vocational School, Department of Drilling, Izmir, TURKEY
Tel: +90 (232) 8531820, Fax: +90 (232) 8531606, Email: celalettin@deu.edu.tr
[4] Izmir Institute of Technology, Department of Civil Engineering, Izmir, TURKEY
Tel: +90 (232) 7506807, Fax: +90 (232) 7506801, Email: alperbaba@iyte.edu.tr

ABSTRACT:

Accurate mapping of the spatial distribution of arsenic in groundwater is an important but equally difficult task to complete due to a number of uncertainties. Classical univariate interpolation algorithms could sometimes be insufficient to capture high concentration and high gradient areas. Under these circumstances, the use of an auxiliary parameter could provide better estimates of arsenic distribution. Based on this premise, arsenic cokriging with a correlated parameter can improve the performance of interpolation and can enhance the quality of predictions. In order to test this hypothesis, a water quality dataset from an arsenic containing aquifer in Simav Plain, Turkey is used to develop arsenic distribution maps. Arsenic is cokriged with correlated parameters such as manganese, iron and dissolved oxygen; and the results are compared with univariate interpolation algorithms such as ordinary kriging and inverse distance weighing. The comparisons were performed with cross validation at sampling locations and assessed based on mean and root mean squared errors. The results revealed that maps developed using arsenic cokriging with iron have given the smallest error value and have shown closest fit to the extreme values in the dataset. Accordingly, arsenic cokriging with iron is believed to be a promising approach in mapping arsenic distributions in groundwater.

KEYWORDS: arsenic mapping, cokriging, geostatistics, Simav plain, Turkey

INTRODUCTION

Spatial interpolation is a technique that is commonly used to obtain high resolution estimates of spatially variable data collected at discrete locations. Being a popular geostatistical prediction technique, kriging is a frequently used spatial interpolation method that can provide estimates of data at unsampled locations based on the spatial correlation between neighboring observations (Goovaerts, 2000; Murphy *et al.*, 2010). Cokriging, on the other hand, is a multivariate improvement over kriging where secondary information is used to improve the prediction achieved by making use of a single variable (Sarangi *et al*., 2005). The cokriging algorithm is widely implemented throughout the world for improving the estimates of several parameters including precipitation (Goovaerts, 2000; Vincente-Serrano *et al*., 2003; Sarangi *et al*., 2005; Moral, 2010), groundwater head (Hoeksema *et al*., 1989; Desbarats *et al.*, 2002), soil hydraulic conductivity (Basaran *et al*., 2011) and others (Han et al., 2003; Simbahan *et al*., 2006; Georgakarakos and Kitsiou, 2008; Dash *et al*., 2010) at unsampled locations through introduction of some auxiliary information as the secondary variate. This secondary piece of information could be elevation as in the case of precipitation, temperature and groundwater head mapping; or water stable aggregates as in soil hydraulic conductivity mapping or electrical conductivity as in soil organic carbon content mapping. A similar analogy is deemed possible for water quality parameters measured in surface and subsurface monitoring studies between interrelated parameters where spatial variability is important.

Arsenic is an important water quality parameter in groundwater. Apart from its toxicity and associated health effects, it has a fairly complex hydrogeochemistry that could make its assessment and management difficult. It has known strong correlations with other water quality parameters such as iron, manganese, redox potential and total organic carbon (Bhattacharya *et al*., 2007) that could provide valuable information for its spatial mapping. Following an analogy similar to precipitation-elevation, arsenic can be coupled with one or more of these correlated parameters to obtain better estimates of arsenic distributions in groundwater systems. Based on this premise, arsenic mapping with cokriging algorithm is proposed in this study to obtain more accurate arsenic distribution maps in groundwater systems. The methodology is tested in an unconfined aquifer in western Anatolia, Turkey where high arsenic levels were previously determined (Gunduz *et al*., 2010).

THE WATER QUALITY DATASET

A water quality monitoring campaign was conducted in Simav Plain, Turkey where groundwater was found to contain elevated concentrations of arsenic. During this campaign, more than 80 water quality parameters were measured at 26 sampling locations including physical field parameters, major anions and cations and heavy metals and trace elements. The results of this monitoring activity verified the presence of arsenic in the groundwater of Simav Plain. The database is then transfermed into geographical information systems platform to perform spatial analysis and mapping. The water quality database was then analyzed statistically for possible correlations between arsenic and other quality parameters. Accordingly, some physicochemical parameters were found to show fairly strong relationship. The statistical summary of these parameters are given in Table 1.

Table 1. Statistical summary of some of the parameters of the water quality database

Parameter	Minimum	Maximum	Average	Median	Standard Dev.
As (μg/L)	1.50	1851.00	252.55	75.75	408.65
ORP* (mV)	-123.00	217.00	37.29	29.00	94.16
EC* (μS/cm)	290.00	2490.00	862.46	655.50	565.45
DO* (mg/L)	1.08	10.40	3.82	3.08	2.25
SO_4^{-2} (mg/L)	0.22	726.71	115.77	31.03	190.38
HCO_3^- (mg/L)	134.20	800.32	428.60	405.04	158.15
TOC* (mg/L)	2.05	15.12	4.48	3.47	3.05
Fe (μg/L)	5.00	22211.00	2300.58	312.50	5411.18
Li (μg/L)	0.90	1254.20	71.02	4.25	245.08
Mn (μg/L)	0.85	2937.77	803.81	235.78	1006.29

* ORP: Oxidation-reduction potential; EC: Electrical conductivity; DO: Dissolved oxygen; TOC: Total organic carbon

With an average total arsenic concentration of 252.55 ppb, Simav plain surface aquifer contains arsenic levels that are about 25 times more than the national and international drinking water quality standard value of 10 ppb (ITASHY, 2005; WHO, 2008). The values obtained in 26 measurements ranged between 1.5 ppb and 1851.0 ppb and had a standard deviation of 408.65 ppb. This wide variability in data values within a fairly small spatial extent of about 20 km necessitated the use of secondary variables to aid the predictions obtained from interpolation algorithms. With this objective, arsenic maps were obtained via different spatial interpolators (including uni- and multi- variate techniques) by using some of these high correlation parameters as secondary variables in the multi-variate techniques.

PARAMETER CORRELATIONS

The water quality database is statistically analyzed to determine the parameters that strongly correlate with arsenic. For this purpose, each parameter is first tested using the Kolmogorov-Smirnov test whether it can be considered as normally distributed. Based on this test it was concluded that most of the parameters fit non-normal statistical distributions. Therefore, the non-parametric spearman's rho correlation is used to determine the relationship between all possible variable pairs and two-tailed significance tests are performed. Correlation analyses results are presented in Table 2 and correlation graphs of selected variables are given in Figure 1. As seen from Table 2, arsenic shows statistically significant correlations with parameters such as manganese, iron, potassium, boron, sodium, dissolved oxygen and bicarbonate. While some of these parameters like manganese, iron and dissolved oxygen are expected, others such as potassium and sodium are less likely to occur from a geochemical view point. Since this analysis is based on a single set of data, deviations from the general trend might be expected. Thus, for creating distribution maps of arsenic, it was found to be more reasonable to use auxiliary parameters whose inter-relations with arsenic are well known and well documented in the literature. Consequently, arsenic is cokriged with manganese, iron and dissolved oxygen; and the results are compared with univariate interpolation algorithms such as ordinary kriging and inverse distance weighing (IDW).

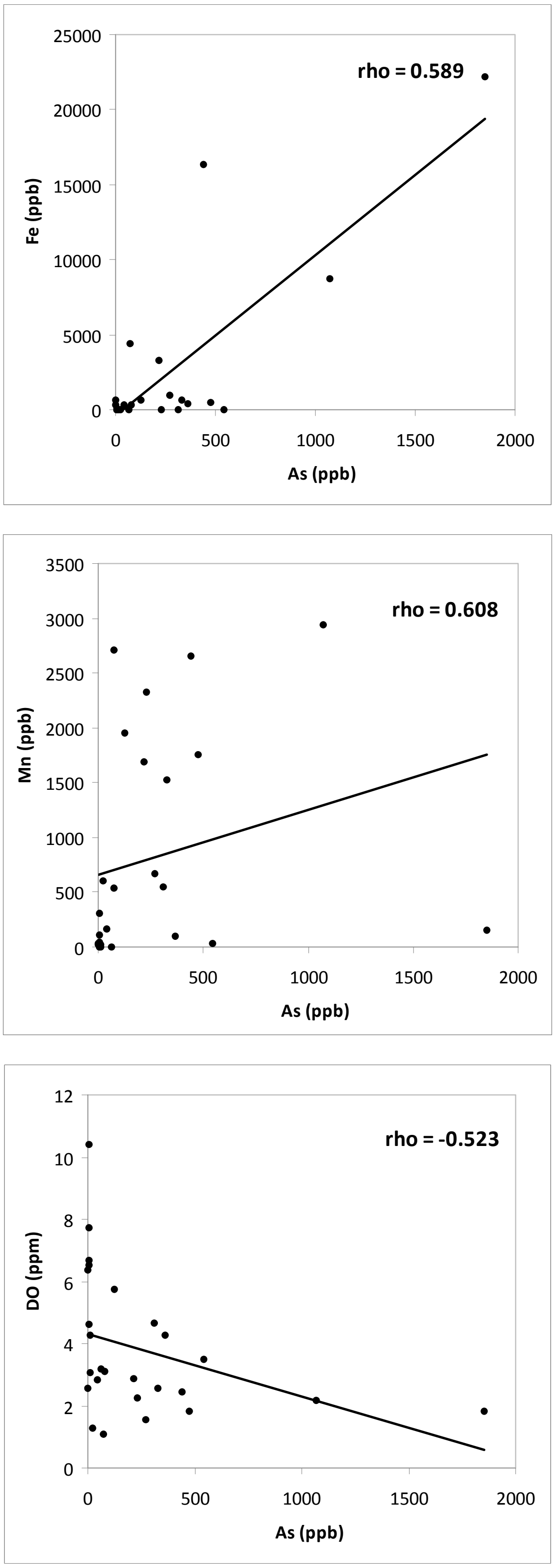

Figure 1. Correlation graphs of selected parameters with arsenic (lines represent linear regression function)

Table 2. Groundwater quality parameters that are significantly correlated to As concentrations ($p<0.05$). Underlined spearman's rho coefficients indicate stronger correlations ($p<0.01$).

	Spearman's rho	Significance (2-tailed)
EC	0,482	1,27E-02
DO	<u>-0,523</u>	6,10E-03
HCO_3	<u>0,511</u>	7,69E-03
Li	0,458	1,87E-02
Na	<u>0,593</u>	1,40E-03
NH_4	0,486	1,19E-02
F	<u>0,534</u>	4,98E-03
Br	<u>0,506</u>	8,40E-03
B	<u>0,581</u>	1,85E-03
Fe	<u>0,589</u>	1,55E-03
Mn	<u>0,608</u>	9,80E-04
P	<u>0,670</u>	1,79E-04

MAPPING THE ARSENIC DISTRIBUTION

Mapping the distribution of arsenic in a groundwater system involves a number of difficulties associated with some uncertainties. Samples must have been collected from the same stratigraphic unit by using correct sampling procedures and protocols. Detailed information regarding the geology of the area must be known. Ideally, soil-rock sampling should be conducted prior to water quality sampling to find out areas with high arsenic containing minerals. Hot spots originating from strong water-rock interactions at these particular points are likely to create problems during interpolation. Handling such extreme points with geostatistics is known to be a problem and arsenic levels at such locales are generally underestimated as a result of the interpolation routines. Often, datasets with a wide data range and skewed distributions yield less accurate distribution maps as opposed to datasets that follow normal distribution. Under these conditions, the use of auxiliary variables helps to improve the overall performance of the distribution.

Several arsenic distribution maps are created with different interpolation algorithms. Firstly, an IDW map is created to represent the simplest case. Then, an ordinary kriging map is used as the base scenario. These two maps are then compared with maps produced by cokriging with iron, manganese and dissolved oxygen as auxiliary parameters. These parameters are selected from a subset of wider number of variables, which gave relatively strong correlations with arsenic and are known to have some chemical link with it. The accuracy of each interpolation method was then tested with cross validation, which is based on the principle of sequentially removing each known measurement point and using the rest of the points to predict its value via interpolation. The errors between measured and predicted levels are then reported to compare the performances of interpolation approaches. Mean error and root mean squared error are the two different error types used in this comparison. In general, the algorithms with lower errors were considered to perform better. A mean error equal to zero implies normally distributed errors around zero, and no systematic over- or under- prediction of arsenic concentrations. The root mean squared error is indicative of the cumulative error and provides a better estimate of the overall performance of the approach. The results of the cross validation tests of IDW, kriging and three cokriging (As-Fe, As-Mn and As-DO) algorithms are given in Table 3.

Table 3. Comparison of errors obtained from different interpolation approaches

Interpolation method	Mean error (ppb)	Root mean squared error (ppb)
IDW	-10.78	467.1
Kriging of (As)	-24.56	438.6
Cokriging of (As) with (D.O.)	-11.01	423.2
Cokriging of (As) with (Mn)	-20.07	399.0
Cokriging of (As) with (Fe)	-19.51	282.8

Univariate techniques had the highest root mean squared error values. The introduction of auxiliary variables in multivariate techniques is shown to improve the performance of interpolation; the highest improvement was achieved with iron as the secondary variable. Both performance criteria indicate improvement of arsenic concentration predictions, where the root mean squared error in particular decreases significantly. The mean error value becomes less negative thereby implying less overprediction.

The comparison of distribution maps are shown in Figures 2 and 3 where the base case (ordinary kriging of arsenic) is compared with As-Fe cokriging case, respectively. At first glance, the two maps seem to be very similar, however a closer look reveals slight differences around a number of data points in central portions of the plain. Furthermore, it can be also concluded that the base scenario slightly overestimates the results in the entire plain when compared to As-Fe cokriged case. It should be noted that the visual interpretation of these maps is also dependent on the choice of contour numbers and intervals.

The cross-validation results are also presented in form of a correlation graph (Figure 4). Measured against interpolated arsenic concentrations are drawn for each approach used. A least square regression between these two values for each case revealed that methods with lower mean error such as IDW do not necessarily give the most accurate results. Thus, the deviation from the 1:1 line could be highest in a method where the mean error is lowest. Hence, root mean square error is a better estimator of the methods' performance. In this regard, cokriging with iron approach provided the best results and gave the closest fit to the 1:1 line. It should also be mentioned that when the extreme values such as 1851 ppb value is excluded from the entire dataset, better fits could be achieved. Thus, the method that could provide the best fit to the most extreme data could also to be considered to perform well since the overall aim of interpolation is to provide some sophisticated way of predicting water quality in areas where no data exists and propose a more realistic estimate in areas of high gradient. In this regard, cokriged As-Fe pair has been shown to perform better when compared to manganese and dissolved oxygen.

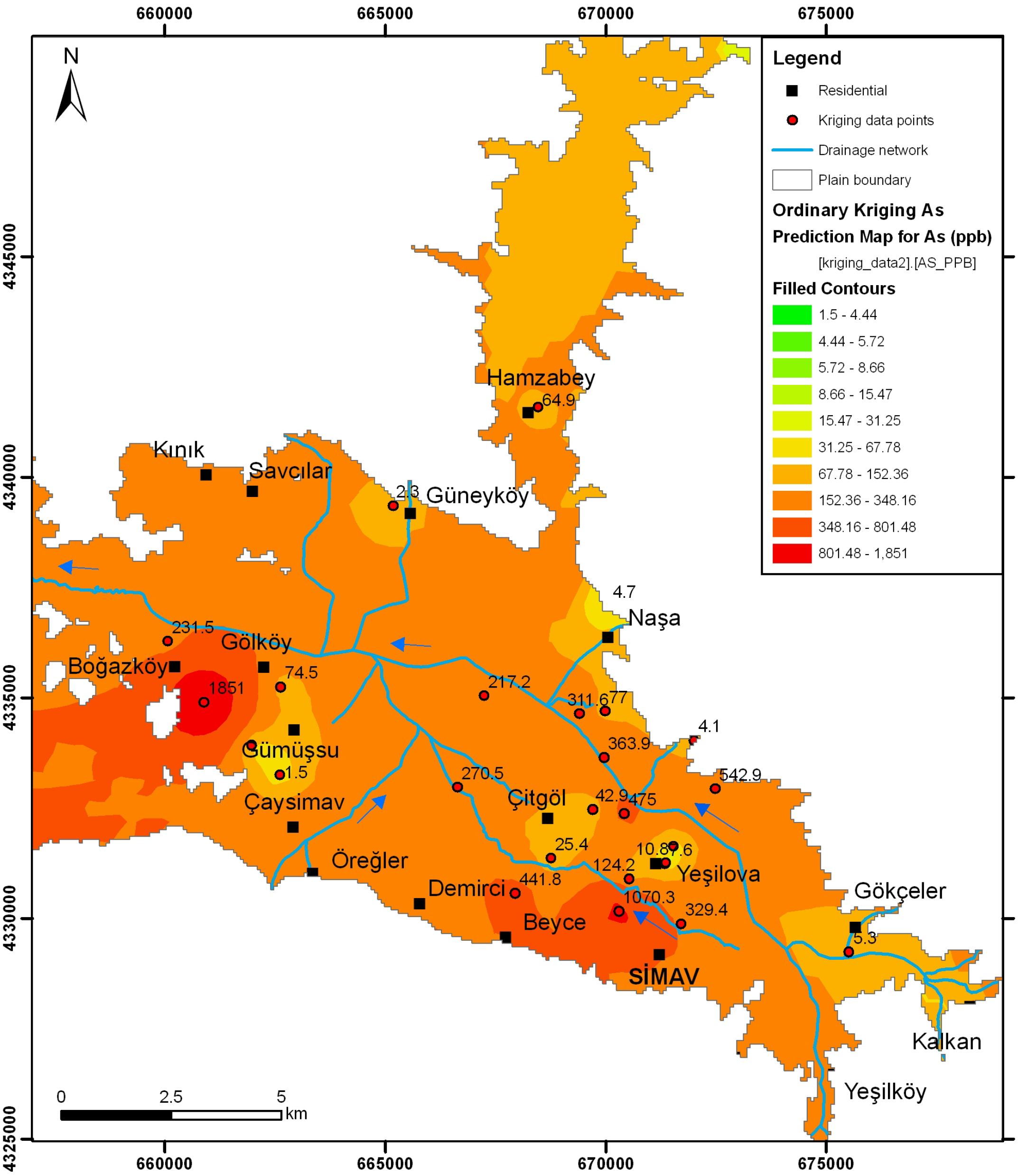

Figure 2. Arsenic distribution in Simav Plain obtained by ordinary kriging of (As)

CONCLUSION AND RECOMMENDATIONS

Spatial interpolation of point arsenic measurements through the use of kriging algorithm provides valuable information on the distribution of this parameter in a groundwater system. Information extracted from these distribution maps can be considered crucial in assessing the public health risks of existing water supply wells or finding new well locations with arsenic levels satisfying the quality criteria. However, these maps could involve significant uncertainties and are often difficult to develop when there are extreme measurements.

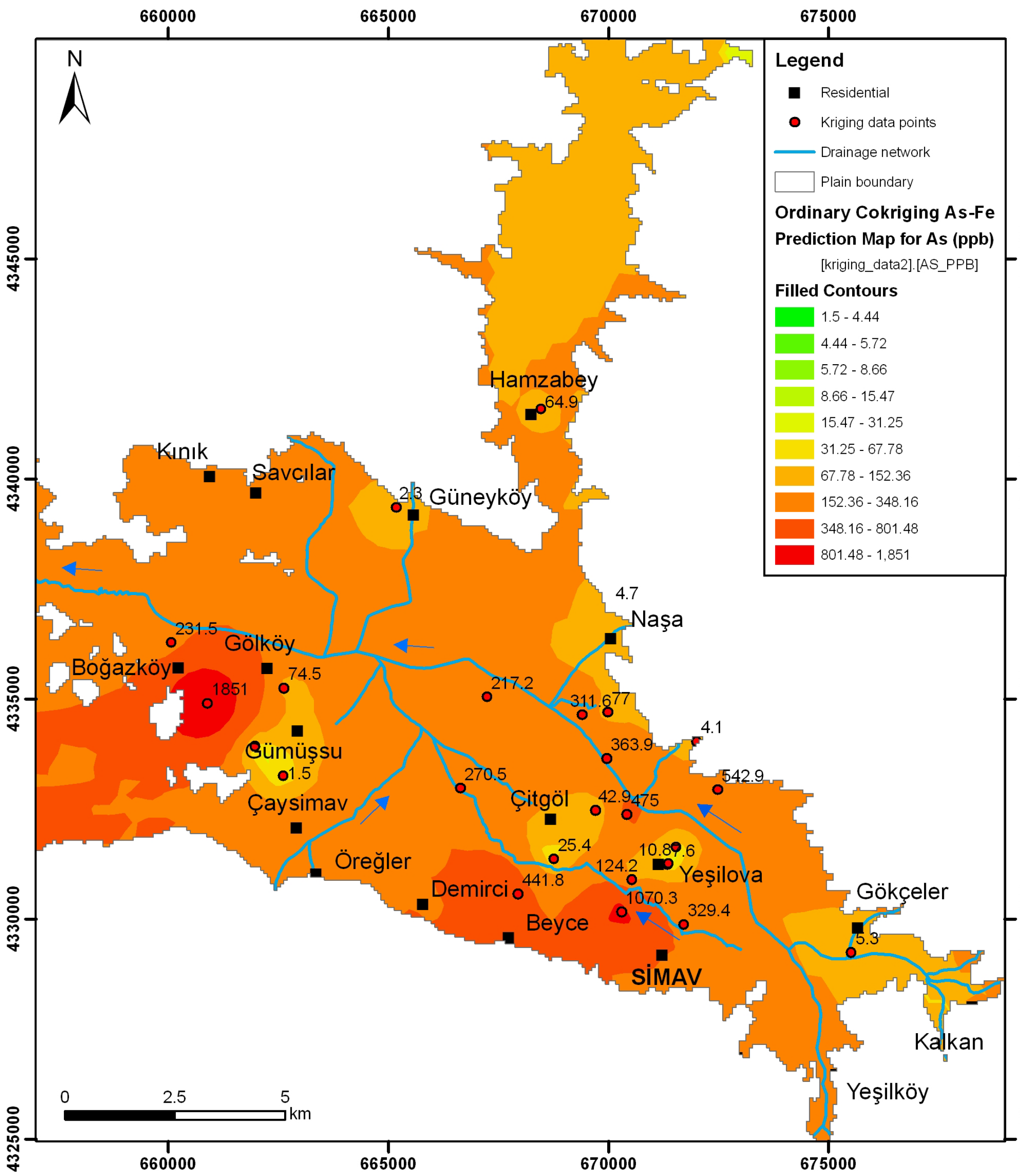

Figure 3. Arsenic distribution in Simav Plain obtained by cokriging of (As) with (Fe)

Under such circumstances, incorporating an auxiliary variable such as iron in kriging procedure could help in improving the predictions. Cokriging arsenic with iron was shown to improve the results for the dataset used in this study. Lower root mean squared errors and better measured vs. predicted results were obtained in cross-validation assessments of As-Fe cokriging applications. The results presented in this study were, however, based on a single set of water quality data. Experimentations with more independent datasets are deemed crucial to reach more concrete results of arsenic mapping.

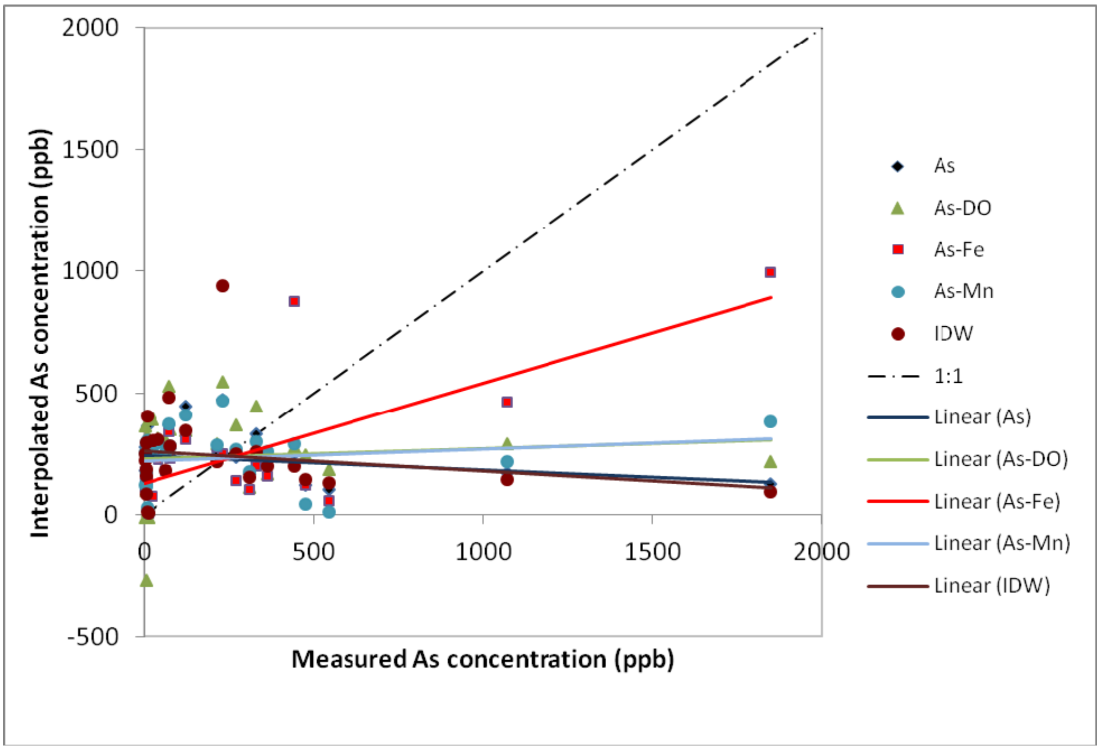

Figure 4. Measured vs. interpolated arsenic concentrations from cross validation results with different interpolation algorithms

Furthermore, it is worthwhile to try out arsenic soil concentrations as the auxiliary variable in the cokriging approach as soil concentrations are directly related to groundwater arsenic levels. However, the auxiliary variable to be used must be some parameter that can be obtained easily (preferably easier than arsenic itself). In addition, the number of data in the set is also an important factor to be considered. The error values for this study are calculated based on cross-validation on 26 data points. The representativeness of the root mean squared error value is expected to enhance with increasing number of data points. Hence, more data points would provide a more reliable interpolation performance indicator. Cases of higher correlations between arsenic and the auxiliary variable are also expected to result in better predictions at unsampled points. Finally, the root mean squared error values obtained from a separate dataset taken from different sampling locations within the same area as the validation dataset is considered to perform as a much better indicator to test the overall performance of procedure.

Overall, arsenic cokriging with iron or some other geochemically correlated parameter is a promising approach to develop better distribution maps of this extremely important water quality parameter. Using an analogy similar to precipitation-elevation or groundwater head-topography, arsenic-iron cokriging results can provide improved estimates when compared to ordinary kriging of arsenic itself.

ACKNOWLEDGEMENTS

The data used in this research is provided from a project (109Y029) funded by the Scientific and Technological Research Council of Turkey (TUBITAK).

REFERENCES

Basaran, M., Erpul, G., Ozcan, A.U., Saygin, D.S., Kibar, M., Bayramin, I. and Yılman, F.E. (2011). Spatial information of soil hydraulic conductivity and performance of cokriging over krigging in a semi-arid basin scale. *Environmental Earth Science*, 63:827-838.

Bhattacharya, P., Welch, A.H., Stollenwerk, K.G., McLaughlin, M.J., Bundschuh, J. and Panaullah, G. (2007). Arsenic in the environment: biology and chemistry. *Science of the Total Environment*, 379:109-120.

Dash, J.P., Sarangi, A. and Singh, D.K. (2010). Spatial variability of groundwater depth and quality parameters in the national capital territory of Delhi. *Environmental Management*, 45:640-650.

Desbarats, A. J., Logan, C. E., Hinton, M. J., & Sharpe, D. R. (2002). On the kriging of water table elevations using collateral information from a digital elevation model. *Journal of Hydrology*, 255(1-4), 25-38.

Georgakarakos, S. and Kitsiou, D. (2008). Mapping abundance distribution of small pelagic species applying hydroacoustics and cokriging techniques. *Hydrobiologia*, 612:155-169.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228:11-129.

Gunduz O, Simsek C, Hasozbek A (2010). Arsenic pollution in the groundwater of Simav Plain, Turkey: Its impact on water quality and human health, *Water Air Soil Pollution*, 205:43-62.

Han, S., Schneider, S.M. and Evans, R.G. (2003). Evaluating cokriging for improving nutrient sampling frequency. *Transactions of ASAE*, 46(3):845-849.

Hoeksema, R. J., Clapp, R. B., Thomas, A. L., Hunley, A. E., Farrow, N. D., and Dearstone, K. C. (1989). Cokriging model for estimation of water table elevation. *Water Resources Res.*, 25(3), 429-438.

ITASHY (2005). Regulation on waters for human consumption. Official Gazette dated 17/02/2005, No.25730, Ankara. (in Turkish)

Moral, F.J. (2010). Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology*, 30:620-631.

Murphy, R.R., Curriero, F.C. and Ball, W.P. (2010). Comparison of spatial interpolation methods for water quality evaluation in Chesapeake Bay. *Journal of Environmental Engineering ASCE*, 136(2):160-171.

Sarangi, A., Cox, C.A. and Madramootoo, C.A. (2005). Geostatistical methods for prediction of spatial variability of rainfall in a mountainous region. *Transactions of ASAE*, 48(3):943-954.

Simbahan, G.C., Dobermann, A., Goovaerts, P., Ping, J. and Haddix, M.L. (2006). Fine-resolution mapping of soil organic carbon based on multivariate secondary data. *Geoderma*, 132:471-489.

Vincente-Serrano, S.M., Saz-Sanchez, M.A. and Cuadrat J.M. (2003). Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research*, 24:161-180.

WHO (2008). Guidelines for drinking water quality (3rd ed.). World Health Organization, Geneva: WHO Press.

View publication stats