Modelling arsenic hazard in Cambodia: A geostatistical approach using ancillary data

Luis Rodríguez Lado, David Polya, Lenny Winkel, Michael Berg, Aimee Hegan

PII: S0883-2927(08)00236-9
DOI: 10.1016/j.apgeochem.2008.06.028
Reference: AG 1840

To appear in: *Applied Geochemistry*

Please cite this article as: Lado, L.R., Polya, D., Winkel, L., Berg, M., Hegan, A., Modelling arsenic hazard in Cambodia: A geostatistical approach using ancillary data, *Applied Geochemistry* (2008), doi: 10.1016/j.apgeochem.2008.06.028

Modelling arsenic hazard in Cambodia: A geostatistical approach using ancillary data

Luis Rodríguez Lado [1,*], David Polya [2], Lenny Winkel [3], Michael Berg [3] & Aimee Hegan[1, 2]

[1] European Commission, Directorate General JRC, Institute for Environment and Sustainability, TP 280, Via E. Fermi 1, I-21020 Ispra (VA), Italy

[2] School of Earth, Atmospheric and Environmental Sciences; University of Manchester. SEAES, Williamson Building, Oxford Road; The University of Manchester, M13 9PL, UK; Fax +44 161 306 9361.

[3] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, 8600 Dübendorf, Zurich, Switzerland; Tel. +41-44-823 50 78; Fax +41-44-823 50 28.

[*]corresponding Author: E-mail: luis.rodriguez-lado@jrc.it fax: +39-0332-786394.

Abstract

The As concentration in shallow groundwater in Cambodia was estimated using 1329 georeferenced water samples collected during the period 1986-2004 from wells between 16-100 m depth. Arsenic concentrations were estimated using block regression-kriging on the log transformed As measurements. Auxiliary raster maps (DEM-parameters, remote sensing images and geology) were converted to 16 principal components that were used to explain the distribution of As over the study area. The regression-kriging model was validated using an external set of 276 samples, and the results were compared to those obtained by ordinary block kriging. The regression analysis revealed that there is a good correlation between topographic environmental variables and the content of As in groundwater. This result is broadly consistent with the findings of previous studies and is not unexpected given models of microbial mediated As mobilization in recent low lying sediments. Kândal, Prey Vêng and Kâmpóng Cham are the provinces with the highest potential As hazard, indicating the requirement for development and implementation of policy control measures. The regression-kriging model explained 48% of the variability in the validation set. However, the model does not show good results for the prediction of high As concentration. This points to the existence of local environmental factors, not captured by this model, that highly influence the mobilization of As in groundwater. Even if the results of the validation of regression-kriging and ordinary kriging are

similar, the regression kriging approach provides a more realistic description of the distribution of As since it also captures the large-scale variation of As in the study area.

1. Introduction

Arsenic in shallow groundwater extensively utilized for drinking, irrigation and cooking in many parts of the world represents a major environmental hazard (Smith et al., 2000; Smedley and Kinniburgh, 2002; Charlet and Polya, 2006). The hazard arising from such As-bearing shallow groundwater in Cambodia has also been recently recognized (Feldmann and Rosenboom, 2001; Polya et al., 2003, 2004, 2005; Fredericks, 2004; Berg et al., 2007; Buschmann et al., 2007).

Several attempts have been made to model the distribution of As in Cambodian groundwater over the last few years. Fredericks (2004) created a qualitative country-wide distribution map based largely on subsurface geology. Polya et al. (2005) created a country-wide hazard map based on a combination of existing data and crude geological sub-division of the country. This map, although indicative of areas of high As hazard, has a poor resolution comprising 10 x 10 km pixels. More recently, Berg et al. (2007) delineated 3 risk categories for As hazard. Buschmann et al. (2008) applied a nearest neighbor algorithm to map groundwater quality over the whole Mekong Delta floodplain (Vietnam and Cambodia). None of these hazard maps make substantial use of ancillary data beyond topography and crude geology.

The aim of the present study is to develop a quantitative spatial model based upon regression-kriging to estimate the mean As concentrations in shallow groundwater in Cambodia and hence predict areas of high As hazard. The advantage of regression-kriging in relation to other commonly used interpolation techniques is that it makes use of ancillary environmental variables that potentially enhance the accuracy of the predictions. The results of this spatial model may be used, in conjunction with noted limitations, other data and expert knowledge, as a tool to assist the development of adequate measures to reduce risk to human health.

2. Study area

The study area comprises the territory of the Kingdom of Cambodia. It covers an extent of 181040 km^2 and involves approximately 13.88 million people (C.I.A., 2007). The climate is tropical with a rainy, monsoon season extending from May to

November and a dry season from December to April. The temperatures are quite homogeneous presenting little seasonal variation. The topography is rather flat in most of the country but mountain ranges are present in the SW and in the north. The average altitude is 126 m and the highest point is located in the "Phnum Aoral" mountain (1810 m).

3. Materials and methods

3.1. Sampling points: Data preparation

The data modeled in this study were obtained from a compilation of 9796 sampling points compiled by Polya et al. (2005) from various survey teams over the period 1986-2004. These data include those previously compiled by the Cambodian Arsenic Inter-Ministerial Commitee (AISC-NAB) and those measured by the University of Manchester (MCR) including in conjunction with Action Against Hunger (MCR-AAH), Action Against Hunger in Preah Vihar (AAH-PV) and Mondal Kiri (AAH-MK), the World Health Organization (WHO), and the Swiss Federal Institute of Aquatic Science and Technology (EAWAG). The analytical and quality assurance protocols utilised by groups responsible for each of these datasets varied considerably and there is consequently a commensurate variability in data quality as discussed by Polya et al. (2005).

The variables included in this database are Easting (m), Northing (m), Depth (m), [As] (μg/L), [NH_4^+] (mg/L), [Fe] (mg/L), [Mn] (mg/L), [Pb] (μg/L) and Alkalinity (mg/L). This database was cleaned to produce a consistent database suitable for geostatistical analysis. This process included the removal of questionable [As] measurements (blank data and negative values) and duplicates. Finally, since the goal was to model the As contents in shallow reducing groundwaters only the As samples collected at depths between 16m and100 m were selected: consequently no inferences may be derived in this study from the As concentrations of groundwaters from either shallower or deeper waters than this range This compilation resulted in a total of 1605 samples (Fig. 5) that were classified randomly into two subsets. The first subset, composed by 1329 samples (83% of the total dataset), was used to create the geostatistical model, while the second subset (276 samples) was reserved for validating the results of the model. All the statistical analyses were performed on the common logarithm transformed As data to account for normality.

3.2. Auxiliary variables

A number of auxiliary environmental variables with potential influence on the content of As in groundwater were selected to build the regression-kriging model. These variables can be classified into three groups as follows:

1. Topographic variables: These include the Digital Elevation Model of Cambodia and their derivatives. The 90 m DEM derived from the SRTM30 V2 dataset obtained from the Jet Propulsion Laboratory was used (http://www2.jpl.nasa.gov/srtm/). From this DEM the following derivate variables were calculated: (1) Slopes; (2) Topographic Wetness Index; (3) Hydrologic Flux Length and (4) Topographic Convergence Index. These parameters have been calculated using both ArcGIS 9.0 (http://www.esri.com) and SAGA GIS (http://www.saga-gis.uni-goettingen.de/).

2. Remote sensing images: Sixteen day averaged MODIS images of Normalized Difference Vegetation Index NDVI at 250 m resolution for the period 01/01/2000 to 31/12/2001 were obtained from the MODIS Terra imagery at the Earth Observing System Data Gateway (http://edcimswww.cr.usgs.gov/ims-bin/pub/nph-ims.cgi/u161524). Two blocks of remote sensing data covering the whole study area were mosaicked and reprojected to the Geographic Coordinate System (WGS84). Principal Component analysis was performed on 35 complete mosaics and only the first 5 principal components were used in the geostatistical approach.

3.Geology: A simplified geological digital map was used which includes only 6 classes of different sedimentary depositional environments: (1) Pre-Holocene sediments; (2) Organic-rich sediments; (3) Other Holocene sediments; (4) Tidal sediments; (5) Floodplains and (6) Alluvial deposits.

Finally a total number of 16 quantitative predictors were created (Fig. 1). These predictors were projected to the Universal Transverse Mercator projection (WGS_1984 UTM_Zone_48N) and upscaled to 250 m to be consistent with the requirements for the geostatistical analysis, and converted to independent principal components in ArcGIS to minimize multicollinearity in the data. These principal components were used as quantitative predictors in the multiple regression models. The final raster maps have the following grid definition: MinX: 182127, MinY: 1105235, MaxX: 826127, MaxY: 1638235 and grid size of 250 m. Finally, a water body mask map (http://biogeo.berkeley.edu/bgm/) has been applied to these maps.

3.3. Inferential method

We have used a geostatistical mapping approach based on regression-kriging to estimate the As contents in shallow groundwater. This geostatistical technique, recently used in environmental studies, has been proved to provide more realistic results and higher accuracy of prediction than other geostatistical techniques (Hengl et al., 2004, 2007; Romić et al. 2007). It is based on an additive coupled analysis of multiple linear regression between the As values and the significant auxiliary environmental variables and a kriging interpolation of the resulting residuals from the regression model. An interpolation was performed in blocks of 250 m since this is the support size of the auxiliary variables. In addition, this technique provides a map indicative of the estimation error associated to the interpolation method. All the modeling steps, from filtering the database to build the regression equations and perform the regression-kriging model have been automated in the statistical environment '*R v.2.6.0*'(http://www.r-project.org). The clear advantage of this automated method is that the resulting spatial models can be easily updated accordingly as new sampling data is available.

4. Results

4.1. Descriptive statistics

The database used in this study is a compilation of samples collected in 8 different surveys (Fig. 2). The contents of As are elevated, ranging from 0.09 to 1340 µg/L, with a mean value of 93 µg/L. About 51 % of the samples have [As] ≤ 10 µg/L while 41 % of the samples have values higher than 50µg/L. The higher As concentration (> 1000 µg/L) is located in the Kândal province but high As anomalies have also been found in the province of Kâmpóng Cham (500 µg/L). Correlation analyses between the As concentration and the other chemical variables in the dataset revealed that there is a moderate positive correlation between As and NH_4^+ (R=0.53).

4.2. Model results

The regression model obtained explains 37% of the variability (R^2_{adj}=0.37). Fourteen principal components contributed significantly to the model (Table 1). The correlation between these principal components and the original variables shows that the Flow Length Index and the presence of organic-rich sediments and alluvial deposits are the variables most positively correlated with the As content, while the terrain altitude is

highly negatively correlated. The remote sensing information also contributes significantly to the model since it clearly delineates the annually flooded areas where most of the higher As content have been found.

The residuals of the regression model showed spatial structure and they were incorporated in the final model by block regression-kriging (Fig. 3) using the '*Gstat*" software (Pebesma and Wesseling, 1998). An exponential variogram was fitted automatically within Gstat by minimizing the residual sum of squares. This variogram shows that the spatial structure is present in a range up to 3.2 Km. The nugget and sill values are 0.01 and 0.3522, respectively. The nugget/sill ratio (0.03) indicates that the residuals have a good spatial dependence.

According to this model 40 % of the territory has As concentrations lower than10 µg/L, and only 0.7 % of the area have values of As higher than 50 µg/L. The higher As estimations are found in the southern flooded areas of the Mekong river in the provinces of Kândal and Prey Vêng.

4.3. Validation

The accuracy of the model was assessed by comparing the As measurements and estimates in the 276 samples from the validation dataset. We performed A linear regression between both As values was performed to check their relationship. We obtained A significant correlation was obtained between measured and predicted As (R^2_{adj}=0.48, α=0.05). The equation of the regression model is:

$$As_{measured} = 17.31+1.07*As_{predicted}$$

Most of the samples fall within the limits of the 95% prediction bands (Fig. 4) and, generally, extremely high As values fall outside this interval. This relationship demonstrates that in general, the model underestimates the real As measurements, especially in the high-As areas. The Root Mean Squared Error of the predictions is high (RMSE=92). A Student's t-test for paired samples (Table 2) shows that there are significant differences between both As values. The mean value of the estimates is significantly lower that the mean value of the field measurements and the variances are very different. The model does not show a good fit in the prediction of higher As anomalies. Most of the observations outside the 95% prediction bands have As concentrations higher than 300 µg/L. By limiting the validation set to the samples with As contents lower than this value (256 samples over 276) a different situation is observed. The mean values are in the same order of magnitude and the Root Mean

Squared Error decreases by half (RMSE=52). The t-test indicates that there are not significant differences between the measurements and the predictions. This indicates that the model only shows a good fit in the lower As concentration ranges.

5. Discussion

The results of this model show that the spatial distribution of high As contents in groundwater is strongly related to the presence of Holocene organic-matter-rich sediments and alluvial deposits. The correlation found between As and NH_4^+ contents (data not shown; see also Rowland et al., 2008) together with the high organic matter contents in the locations of the samples with higher As may indicate that the higher release of As is related to the dissolution of Fe oxy-hydroxides in reducting conditions. This is in agreement with the findings of Bhattacharya (2002) and Berg et al. (2007) in areas of Southern and SE Asia, respectively with high groundwater As contents. In this sense, the map obtained by regression-kriging (Fig. 3) seems to be a good indicator of the potential areas at risk for As exposure since it includes most of the rich organic sediments within the most critical areas. A positive relationship was also observed between the flow length index and the As contents. The flow length index indicates the downstream length, in distance units, from each cell in the raster grid to the initial source point of water. It is higher for the cells located along the Mekong river system, especially in the delta, where high As values have been observed.

The accuracy of the geostatistical models is highly dependent on the sampling density. In this study data from 1329 sampling points over a territory of about 181040 km^2 were used. The normalized estimation error refers to how good an estimate of the mean As is available, rather than of individual wells (Fig. 5). It is observed that the normalized estimation error is very high in the mountainous areas in the Cardamom Mountains, the Elephant Range and the north and south eastern highlands on the border with Vietnam. This points to the fact that one of the biggest limitations that was found in this study arises directly from the sampling design. Since the original purpose of the previous surveys was mainly to propose remediation plans for As hazards in groundwater, the sampling locations are clustered mainly in populated areas where As hazards had been previously reported. This makes this dataset biased and centered in high risk areas whilst other areas in the country have been omitted or under-represented in this sampling dataset.

On the other hand, the model shows low uncertainty in the areas where the biggest problems of As have been reported previously. This is the case of the lowland areas surrounding the Mekong-Tonle Sap systems in the provinces of Pouthisat, Kâmpóng Chnang, Phnom Penh, Kândal, Svay Rieng, Prey Vêng and Kâmpóng Cham. According to this model 40 % of the territory has mean As concentrations lower than 10 µg/L and only 0.7 % of the area, located mainly in the Mekong Delta in the southern area of the province of Kândal, have values of As higher than 50 µg/L.

Another big limitation comes directly from the geologic map used as an auxiliary variable. This map divides the study area according to different types of unconsolidated surface sediments and no considerations on the chemical composition of the sediments are included. It is clear that a proper delineation of the geologic units according to the chemical composition of the sediment and the As bearing phases would greatly increase the accuracy of the results.

Lastly, the selection of the auxiliary variables utilised in this study represents a compromise between variables for which detailed data are readily available (e.g. from satellite-based measurements) and those which might more closely match current models of the controls on As mobilisation in Cambodian aquifer systems. Arguably, the latter would include: the distribution and nature of organics within aquifer sediments (cf. Van Dongen et al., 2008; Quicksall et al., 2008), the distribution and nature of As, As host phases and electron acceptors within the aquifer sediments (cf. Charlet and Polya, 2006), sediment permeability, permeability structure and groundwater flow paths (cf. Postma et al., 2007; Benner et al., 2008; Kocar et al., 2008), sedimentation rates (cf. Kocar et al., 2008; Quicksall et al., 2008) and secular and seasonal changes in rainfall, infiltration and flooding. Of course, many of these variables are 3-D in nature or temporally variable and so do not lend themselves to utilization in 2-D models such as are described in this study. However, the nature of the methodologies described in this study will facilitate the calculation of improved spatial models through the incorporation of even more relevant auxiliary variables as proxies of these data.

The tendency of the model is to underestimate the As concentrations. This is due to the specific characteristics of the locations of the abnormally high As measurements. These high anomalies do not show a clear spatial pattern and are often located in the close vicinity of other sampling points with much lower As concentrations. In this situation, the regression-kriging in 250 m blocks averages the

values within a block and the result is a smoothed surface that considers not the extreme but the mean values in the interpolation process. The model thus shows mean As concentration estimates but actual As concentrations are very heterogeneous and so may vary substantially from this value. The validation of the results must not be done on the single sampling point values but taking into account the mean values of a composite sample within the same block. However, the validation results in the samples up to 300 μg/L show a moderate fit of the predictions.

Finally the results of block-regression-kriging were compared with those produced by ordinary block-kriging (block size =250 m) (Fig. 7). It is observed that ordinary kriging identifies the same main areas at risk for As along the Mekong river and in the delta system. The higher disagreements with regression kriging are for the mountainous areas and for the floodplains of the Tonle-Sap lake, where ordinary kriging estimates higher and lower As contents, respectively. The results of the validation of the ordinary kriging model show that there are not significance differences with those obtained by regression-kriging. The RSME and the regression coefficient are in the same order of magnitude for both models. This means that, quantitatively, the regression-kriging model does not perform better than ordinary kriging.

The regression model captures the large-scale trend of As in the study area (Fig. 6). This map shows the general behavior of As concentrations in groundwater according to the main spatial features. It clearly indicates that the spatial trend is to find high As concentrations in the organic-rich floodplains, while it estimates that the As content in high hilly areas is low. However, the uncertainty in the predictions of the regression model is higher than in the other models (RSME=113, R^2_{adj}=0.34).

Most of the spatial variability in As concentration is due to small-scale (local) environmental processes and thus they are not accurately explained by the regression model alone. Kriging is used to capture part of this small-scale variability, up to the range value of the variogram, through an analysis of autocorrelation of the As concentrations between samples. Since the samples in the validation dataset are quite clustered with those from the model dataset, the influence of local-scale processes have more weight in the predictions than large-scale processes. This means that the results of the validation of regression-kriging and ordinary kriging are very similar. However, at larger scales the results are very different. The map produced by regression-kriging also includes the large-scale spatial trend that is not captured by the

ordinary-kriging method, thus providing a more realistic environmental overview of the distribution of As in 16 m – 100 m depth groundwaters.

6. Conclusions

This study presents a geostatistical approach based on regression-kriging to estimate quantitatively the concentration of As in shallow groundwater in Cambodia with the use of auxiliary information. Despite all the limitations found in the original data, this model shows a realistic picture of the distribution of As over the study area. The study revealed that there is a good correlation between topographic environmental variables (Flow Length Index and elevation), remote sensing NDVI images and geology (sedimentary floodplains) in relation to the content of As in groundwater. This result is broadly consistent with the findings of Buschmann et al. (2007) and is not unexpected given models of microbially mediated As mobilization in recent low lying sediments (Islam et al. 2004; Polya et al., 2003, 2004, 2005; Charlet and Polya, 2006; Lear et al., 2007; Pederick et al. 2007; Rowland et al., 2004, 2007). Kândal, Prey Vêng and Kâmpóng Cham are the provinces with the highest evident As hazard, indicating the requirement for development and implementation of policy control measurements. Additional groundwater As data is required, in particular from areas where there are high estimation errors in the current model (see Fig. 5). It is possible that the inclusion of new environmental data would reduce the uncertainty of the geostatistical models. Winkel et al. (2008) recently assessed soil properties to predict groundwater As contamination in SE Asia. Further research on aspects such as the chemical composition of the sediments and As bearing phases, hydrological regimes and seasonal and secular controls on As mobilization are required to understand the spatial and temporal distribution of As in groundwater.

In addition, it is recommended that there is systematic regional monitoring and control of As in groundwater, particular in highly populated areas in which groundwaters are utilised for drinking, cooking or irrigation. Updating the geostatistical model using the data so obtained would allow the production of regional groundwater quality/As hazard maps, which together with health risk assessments based upon identified exposures routes (e.g. Mondal and Polya, 2008) and biomarker monitoring (Gault et al., 2008), would inform those government and non-government organizations with responsibilities to implement appropriate control measures to minimize the very evident threat to public health arising from geogenic As in well

waters in Cambodia.

Acknowledgements

This work is a contribution from the AquaTRAIN Marie Curie Research Training Network funded by the European Commission Sixth Framework Programme (2002-2006), Marie Curie Actions - Human Resources and Mobility Activity Area, Research Training Networks. DAP acknowledges the receipt of an EPSRC Standard Research Grant (GR/S30207/01). We are grateful to many colleagues, in particular David Fredericks, Jan Willem Rosenboom, Peter Feldmann, Andrew Gault, Helen Rowland, David Cooke, Carl Middleton, Jessica Jones, Vibol Long and Ed Gilligan for their contribution to sample collection in Cambodia, analysis and/or data procurement. We thank Tomislav Hengl for advice on application of regression-kriging. Lastly, we thank 3 anonymous referees for constructive comments which have helped us improve the clarity of the manuscript.

References

Benner, S.G., Polizzotto, M.L., Kocar, B.D., Sampson, S., Fendorf, S., 2008. Groundwater Flow in an Arsenic-Contaminated Aquifer, Mekong Delta, Cambodia Appl. Geochem., this issue.

Berg, M., Stengel, C., Trang, P. T. K., Viet, P. H., Sampson, M. L., Leng, M., Samreth, S., Fredericks, D., 2007. Magnitude of arsenic pollution in the mekong and red river deltas - Cambodia and Vietnam. Sci. Total Environ. 372, 413-425.

Bhattacharya, P., 2002. Arsenic contaminated groundwater from the sedimentary aquifers of south-east Asia. In: Bocanegra, D., Martínez, H., Massone, E. (Eds), Groundwater and Human Development. Proc. XXXII IAH and VI ALHSUD Congress, Mar del Plata, Argentina, 357-363.

Buschmann, J., Berg, M., Stengel, C., Sampson, M. L., 2007. Arsenic and manganese contamination of drinking water resources in Cambodia: Coincidence of risk areas with low relief topography. Environ. Sci. Technol. 41, 2146-2152.

Buschmann, J., Berg, M., Stengel, C., Winkel, L., Sampson, M. L., Trang, P. T. K., Viet, P. H., 2008. Contamination of drinking water resources in the mekong delta floodplains: Arsenic and other trace metals pose serious health risks to population. Environ. Internat. doi: 10.1016/j.envint.2007.12.025.

Charlet, L., Polya, D. A., 2006. Arsenic hazard in shallow reducing groundwaters in southern Asia. Elements 2, 91-96.

C. I. A., 2007. The world factbook (URL). Tech. rep., Central Intelligence Agency of the USA, https://www.cia.gov/library/publications/the-world-factbook/; accessed Nov 27, 2007.

Feldmann, P. R., Rosenboom, J. W., 2001. Final activity report: Drinking water quality assessment in Cambodia, Unpublished Final Draft.

Fredericks, D., 2004. Situation analysis: arsenic contamination of groundwater in Cambodia, Unpublished Report to Arsenic Inter-Ministerial Sub-Commitee, Cambodia.

Gault, A.G., Rowland, H.A.L., Charnock, J.M., Wogelius, R.A., Gomez-Morilla, I., Vong, S., Samreth, S., Sampson, M.L., Polya, D.A., 2008. Arsenic in hair and nails of individuals exposed to arsenic-rich groundwaters in Kandal Province, Cambodia. Sci. Total Environ. 168-176.

Hengl, T., Heuvelink, G. B. M., Rossiter, D. G., 2007. About regression-kriging: From equations to case studies. Computers Geosci.33, 1301-1315.

Hengl, T., Heuvelink, G. B. M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75-93.

Islam, F. S., Gault, A. G., Boothman, C., Polya, D. A., Charnock, J. M., Chatterjee, D., 2004. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. Nature 430, 68-71.

Kocar, B.D., Polizozotto, M.L., Benner, S.G., Ying, S., Ung, M., Ouch, K., Sampson, M., Fendorf, S., 2008. Biogechemical and depositional controls on arsenic mobility within sediments of the Mekong Delta. Appl. Geochem., thisissue.

Lear, G., Polya, D. A., Song, B., Gault, A. G., Lloyd, J. R., 2007. Molecular analysis of arsenate-reducing bacteria within Cambodian sediments following amendment with acetate. Appl. Environ. Microbiol. 73, 1041-1048.

Mondal, D., Polya, D.A., 2008. Rice is a major exposure route for arsenic in Chakdha Block, West Bengal: a Probabilistic Risk Assessment. Appl. Geochem., this issue.

Pebesma, E. J., Wesseling, C. G., 1998. Gstat: a program for geostatistical modelling, prediction and simulation. Computers Geosci. 24, 17-31.

Pederick, R. L., Gault, A. G., Charnock, J. M., Polya, D. A., Lloyd, J. R., 2007. Probing the biogeochemistry of arsenic: Response of two contrasting aquifer sediments from Cambodia to stimulation by arsenate and ferric iron. J. Environ. Sci. Health Part A 42, 1753-1774.

Polya, D. A., Gault, A. G., Bourne, N. J., Lythgoe, P. R., Cooke, D. A., 2003. Coupled HPLC-ICP-MS analysis indicates highly hazardous concentrations of dissolved arsenic species in Cambodian Groundwaters. In: Holland, J. G., Tanners, S. D. (Eds.), Plasma Source Mass Spectrometry: Applications and Emerging Technologies, The Royal Society of Chemistry Special Publication 288, 127-140.

Polya, D. A., Gault, A. G., Diebe, N., Feldmann, P., Rosenboom, J. W., Gilligan, E., Fredericks, D., Milton, A. H., Sampson, M., Rowland, H. A. L., Lythgoe, P. R., Jones, J. C., Middleton, C., Cooke, D. A., 2005. Arsenic hazard in shallow Cambodian groundwaters. Mineral. Mag. 69, 807-823.

Polya, D. A., Rowland, H. A. L., Gault, A. G., Diebe, N. H., Jones, J. C., Cooke, D. A., 2004. Geochemistry of arsenic-rich shallow groundwaters in Cambodia. Geochim. Cosmochim. Acta 68, A590.

Postma, D., Larsen, F., Nguyen, T. M. H., Mai, T. D., Pham, H. V., Pham, Q. N., Jessen, S., 2007. Arsenic in groundwater of the Red River floodplain, Vietnam: Controlling geochemical processes and reactive transport modelling. Geochim. Cosmochim. Acta 71, 5054-5071.

Quicksall, A.N., Bostick, B.C., Sampson, M.L. 2008 Linking Organic Matter Deposition and Iron Mineral Transformations to Groundwater Arsenic Levels in the Mekong Delta, Cambodia. Appl. Geochem., this issue.

Romić, M., Hengl, T., Romić, D., Husnjak, S., 2007. Representing soil pollution by heavy metals using continuous limitation scores. Computers Geosci. 33, 1316-1326.

Rowland, H.A.L., Gault, A.G., Lythgoe, P.R. and Polya, D.A. 2008 Geochemistry of aquifer sediments and arsenic-rich groundwaters in Cambodia. Appl. Geochem., this issue.

Rowland, H. A. L., Pederick, R. L., Polya, D. A., Pancost, R. A., van Dongen, B. E., Gault, A. G., Bryant, C., Anderson, B., Charnock, J. M., Vaughan, D. J., Lloyd, J. R., 2007. Control of organic matter type of microbially mediated release of arsenic from contrasting shallow aquifer sediments from Cambodia. Geobiol. 5, 281-292.

Rowland, H. A. L., Polya, D. A., Gault, A. G., Charnock, J. M., Pederick, R. L., Lloyd, J. R., 2004. Microcosm studies of microbially mediated arsenic release from contrasting Cambodian sediments. Geochim. Cosmochim. Acta 68, A390.

Smedley, P., Kinniburgh, D., 2002. A review of the source, behaviour and distribution of arsenic in neutral waters. Appl. Geochem. 17, 517- 568.

Smith, A., Lingas, E., Rahman, M., 2000. Contamination of drinking water by arsenic in bangladesh - a public health emergency. Bull. World Health Organization 78, 1093-1103.

van Dongen, B.E., Rowland, H.A.L., Gault, A.G., Polya, D.A.,Bryant, C., Pancost, R.D., 2008. Hopane, sterane and n-alkane distributions shallow sediments hosting high arsenic groundwaters in Cambodia. Appl. Geochem., this issue.

Winkel, L., Berg, M., Amini, M., Hug, S. J., Johnson, C. A., 2008. Predicting groundwater arsenic contamination in southeast Asia from surface parameters. Appl. Geochem., this issue.

FIGURES

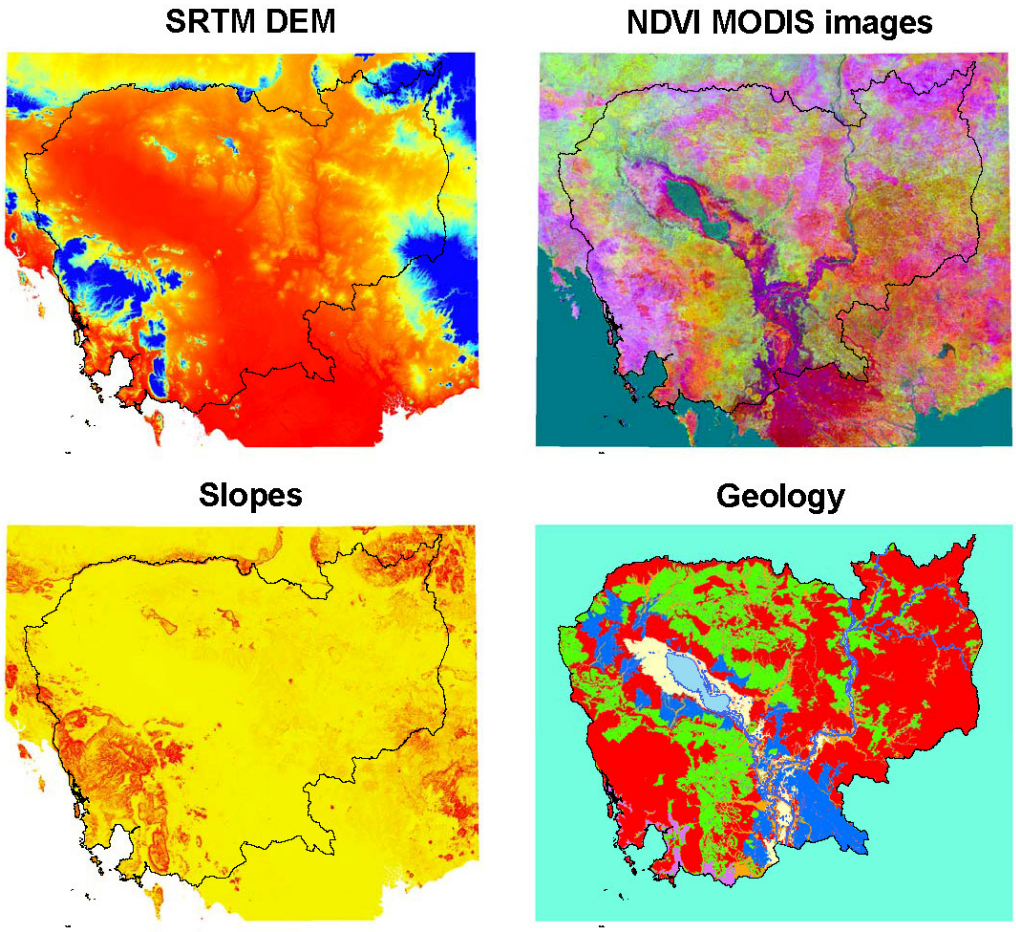

Fig 1.

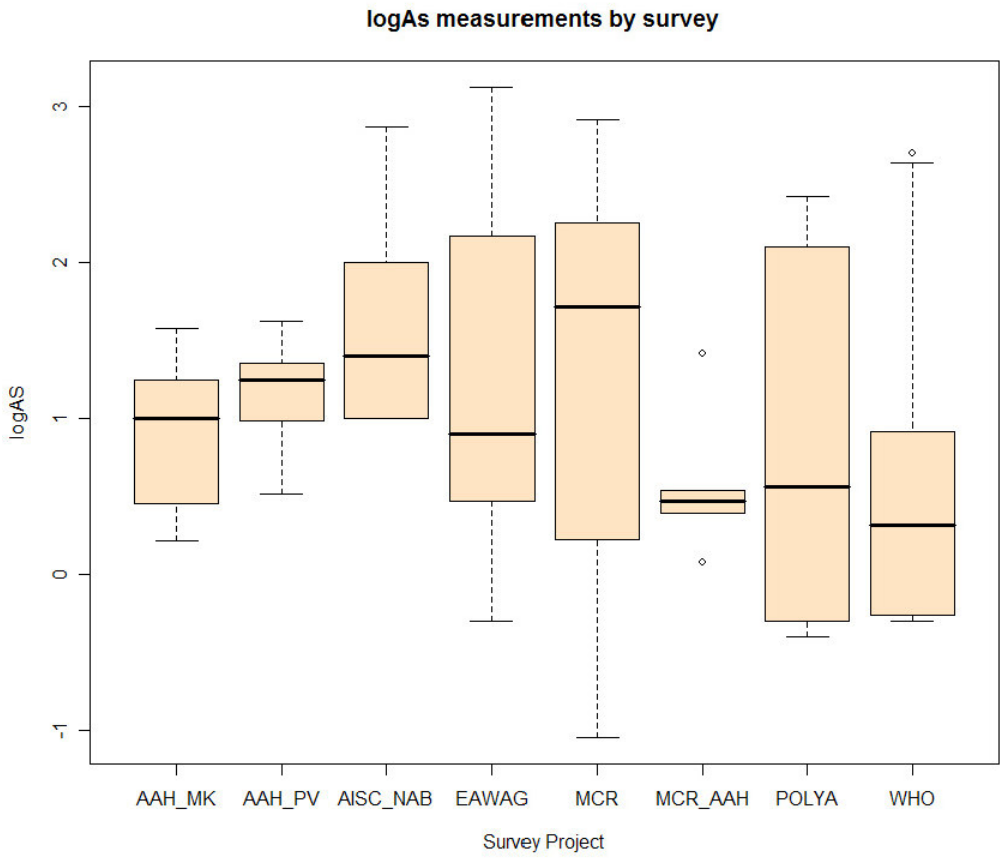

Fig 2.

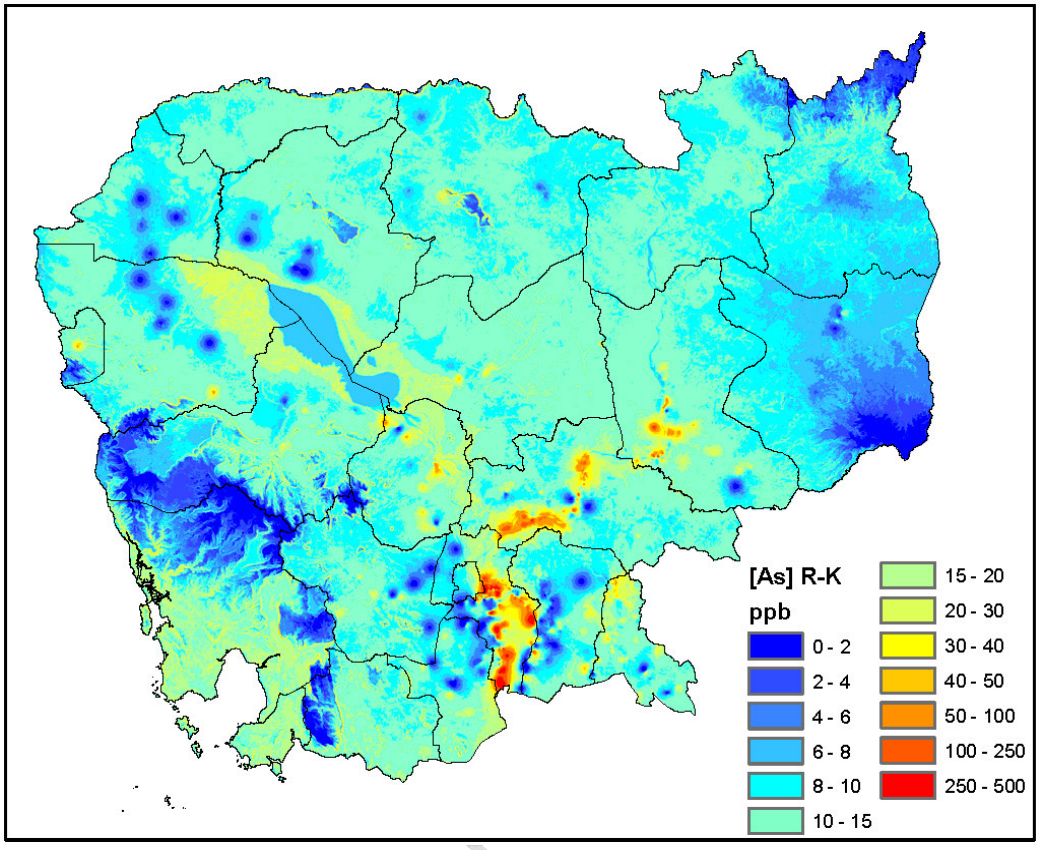

Fig 3.

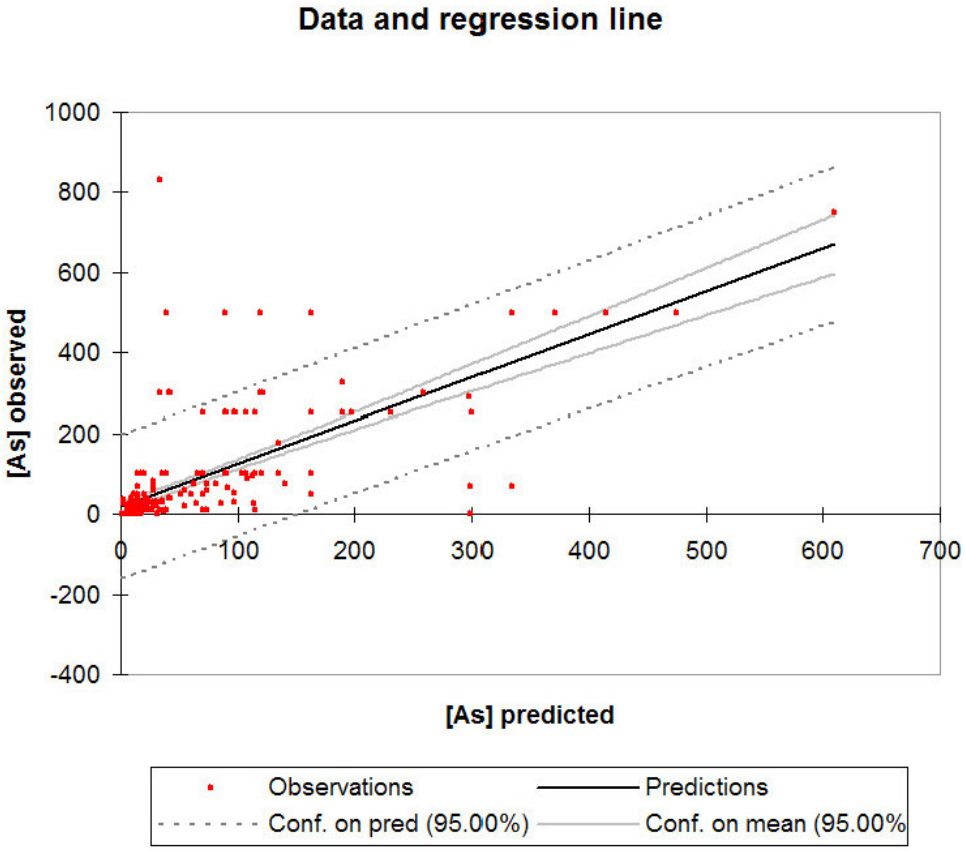

Fig. 4.

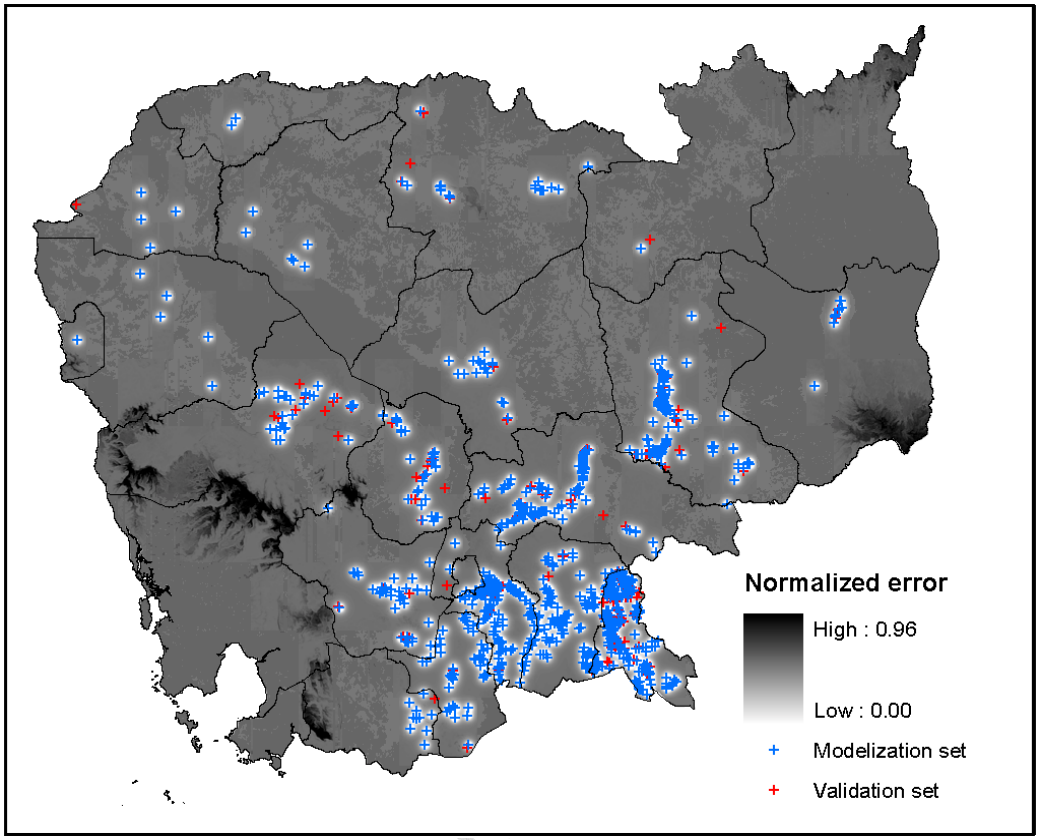

Fig 5

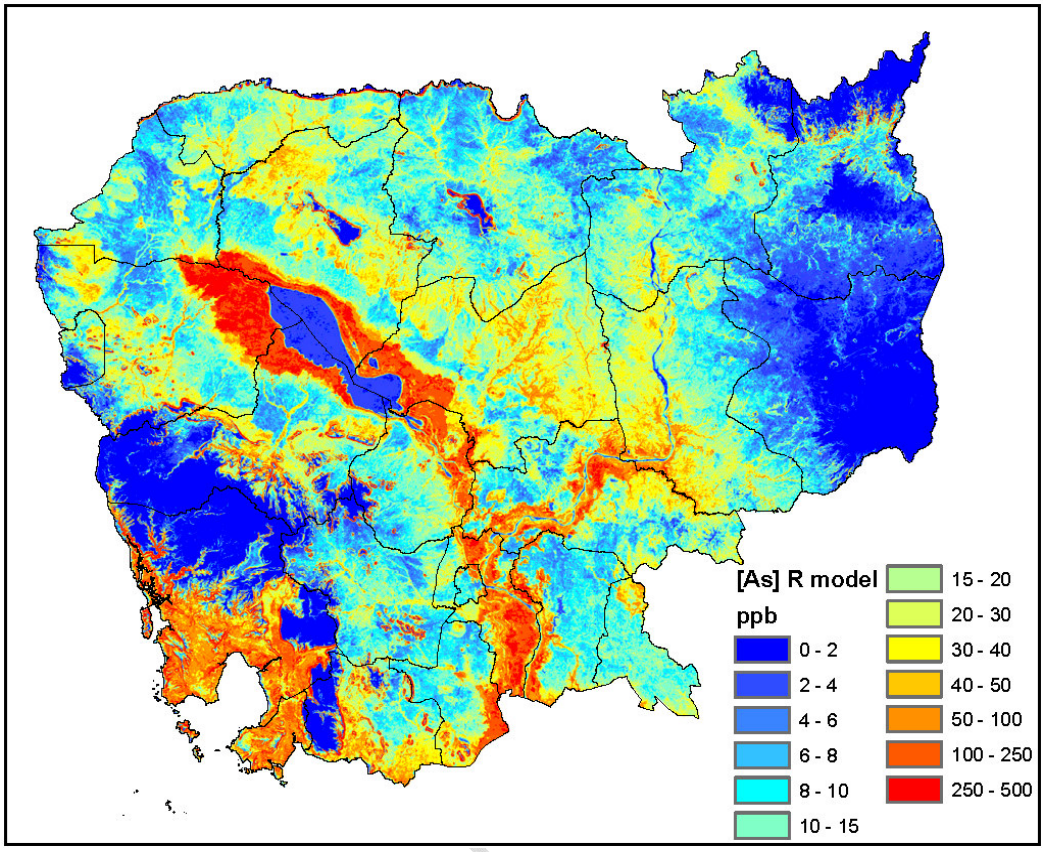

Fig 6.

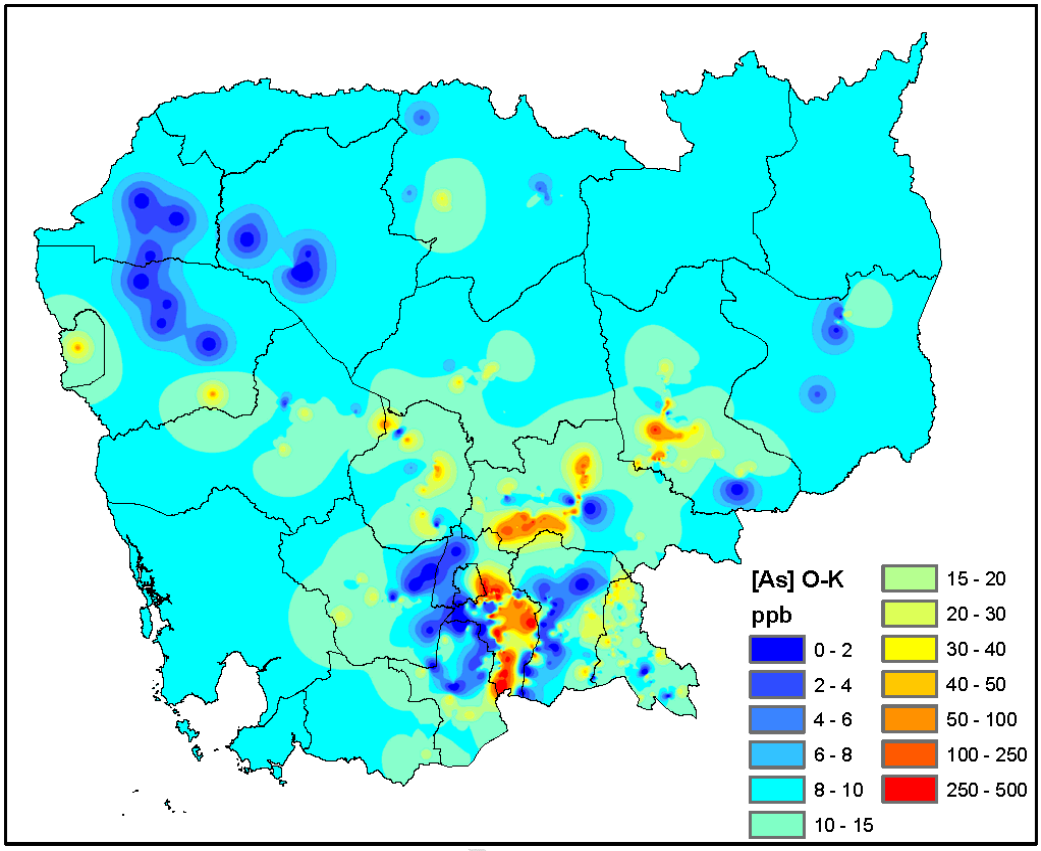

Fig 7.

Figure captions

Fig 1. Some auxiliary predictors used in the geostatistical analysis. SRTM DEM: red to blue=low to high altitudes; NDVI: color composite of the 3 main Principal Components; Slopes: yellow to red=low to high steeps and Geology: Pre-Holocene sediments (red); Organic-rich sediments (yellow); Other Holocene sediments (green); Tidal sediments (purple); Floodplains (blue) and Alluvial deposits (orange)

Fig 2. Boxplot of the contents of As related to the sampling survey group

Fig 3. Estimation of As content in 16 m – 100 m depth groundwater (µg/L) by regression-kriging.

Fig. 4. Biplot of measured vs predicted As concentrations

Fig 5. Normalized interpolation error and location of the samples

Fig 6. Estimation of As content in 16 m – 100 m depth groundwater (µg/L) by multiple regression

Fig 7. Estimation of As content in 16 m – 100 m depth groundwater (µg/L) by ordinary kriging

Tables

Table 1. Results of the multiple regression model

Variable	Estimate	Std.Error	t_value	Pr>t
(Intercept)	1.24e+00	2.176	0.571	0.5679
PC_1	3.71e-06	5.937e-07	6.241	5.84e-10***
PC_2	3.38e-05	4.113e-06	8.229	4.51e-16***
PC_3	-2.77e-05	8.193e-06	-3.381	7.44e-04***
PC_4	-6.09e-05	1.802e-05	-3.380	7.45e-04***
PC_5	-8.20e-05	1.038e-05	-7.899	5.89e-15***
PC_6	-8.07e-05	1.118e-05	-7218	8.90e-13***
PC_7	5.40e-03	7.177e-04	7.527	9.59e-14***
PC_9	-8.67e-02	2.311e-02	-3.749	1.58e-04***
PC_10	7.49e-02	3.699e-02	2.024	0.0432*
PC_11	-1.22e-01	4.572e-02	-2.666	0.0077**
PC_13	-4.93e-01	4.074e-02	-12.094	<2e-16***
PC_14	-3.27e-01	6.920e-02	-4.721	2.60e-06***
PC_15	-2.67e-01	6.970e-02	-3.855	1.32e-04***
PC_16	1.22e+00	1.222e-01	9.948	<2e-16***

Significance: 0‘***’ 0.001‘**’ 0.01‘*’ 0.05‘

Table 2. Student’s t-test for paired samples

Sample	N	Mean	Variance	Std	St error
Measured As	276	78	15914	126	7.6
Predicted As	276	57	6741	82	4.9
Measured As	256	50	4853	70	4.3
Predicted As	256	46	3549	59	3.7