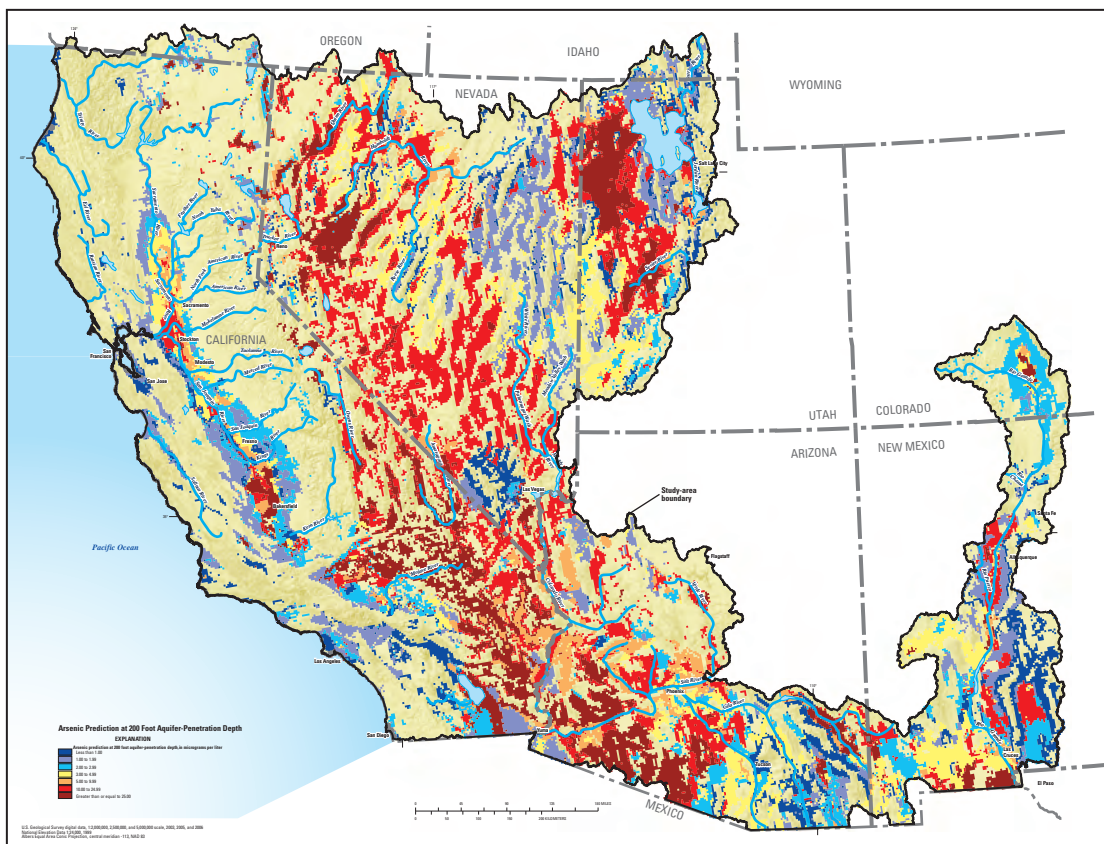National Water-Quality Assessment Program

Digital Spatial Data for Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwest Principal Aquifers Study Area

Data Series 698

U.S. Department of the Interior
U.S. Geological Survey

Digital Spatial Data for Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwest Principal Aquifers Study Area

By Tim S. McKinney and David W. Anning

National Water-Quality Assessment Program

Data Series Report 698

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2012

Contents

Digital Spatial Data for Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwest Principal Aquifers Study Area

Abstract

This product "Digital spatial data for predicted nitrate and arsenic concentrations in basin-fill aquifers of the Southwest Principal Aquifers study area" is a 1:250,000-scale vector spatial dataset developed as part of a regional Southwest Principal Aquifers (SWPA) study (Anning and others, 2012). The study examined the vulnerability of basin-fill aquifers in the southwestern United States to nitrate contamination and arsenic enrichment. Statistical models were developed by using the random forest classifier algorithm to predict concentrations of nitrate and arsenic across a model grid that represents local- and basin-scale measures of source, aquifer susceptibility, and geochemical conditions.

Background

This dataset was developed as part of a regional Southwest Principal Aquifers (SWPA) study and for the National Water-Quality Assessment (NAWQA) Program. The dataset represents the spatial and statistical distribution of nitrate and arsenic concentrations that were determined from the prediction classifiers for basin-fill aquifers across the SWPA study area.

Development of Classifiers

Separate classifiers were developed for nitrate and arsenic because each constituent was expected to be affected by a different set of factors, and each factor could have a different magnitude or directional influence (increase/decrease) on concentration. For each constituent, two different classifiers were developed: a prediction classifier and a confirmatory classifier. The prediction classifiers were developed specifically to predict nitrate and arsenic concentrations in basin-fill aquifers across the SWPA study area and were based on explanatory variables representing source and susceptibility conditions. These explanatory variables were available throughout the entire SWPA study area and, therefore, did not pose a limitation for using the classifiers to predict concentrations.

The confirmatory classifiers were developed to supplement the prediction classifiers in the evaluation of the conceptual model. The name, "confirmatory," reflects the classifier's purpose for evaluation of a-priori hypotheses and contrasts other general types of statistical models, such as those used for prediction or exploratory purposes. The confirmatory classifiers included the explanatory variables used in the prediction classifiers, as well as additional variables representing geochemical conditions and basin groundwater budget components. The inclusion of the geochemical and basin groundwater budget variables in the confirmatory classifiers allowed for further evaluation of the conceptual models, which was not possible with the prediction classifiers alone. The geochemical data, however, were only available at specific well locations, and consistent water-budget data were not available for every basin in the study area. The limited availability of the data for these variables constrained the confirmatory classifiers to observations from 16 case-study basins and precluded use of the confirmatory classifier for predicting concentrations across the SWPA study area. To contrast the scope of the two classifiers, the confirmatory classifiers were developed by using all available explanatory variables but with observations restricted to the 16 case-study basins, whereas the prediction classifiers were unrestricted with respect to spatial extent because these were developed by using a subset of the explanatory variables that were available throughout the study area.

Supplemental Information

The nitrate and arsenic predictions are part of a larger dataset of explanatory variables and model input data. Those data are archived in tabular form along with the model report. The explanatory data may be joined to this dataset via the grid cell identifier. For more information on the development of the explanatory variables, see "Compilation and Processing of Explanatory Variables" in the "Approach and Methods" section of Anning and others (2012).

The digital dataset can be downloaded from the USGS at *http://water.usgs.gov/lookup/getspatial?ds2012-698_SWPA_NO3_As_prediction*.

References Cited

Anning, D.W., Paul, A.P., McKinney, T.S., Huntington, J.M.,
   Bexfield, L.M., and Thiros, S.A., 2012, Predicted nitrate
   and arsenic concentrations in basin-fill aquifers of the south-
   western United States: U.S. Geological Survey Scientific
   Investigations Report 2012–5065, 78 p. Available at
   http://pubs.usgs.gov/sir/2012/5065.

McKinney and Anning—**Digital Spatial Data for Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwest Principal Aquifers Study Area**—Data Series 698