

## Fact Sheet 3: How Is Geospatial Analysis Applied?

This fact sheet is the third of four fact sheets developed by ITRC to accompany its guidance titled *Geospatial Analysis for Optimization at Environmental Sites* (GRO-1) (available at [www.itrcweb.org/gro-1](http://www.itrcweb.org/gro-1)). This fact sheet introduces the value and use of geospatial analysis to support optimization activities to project managers, program or financial managers, and stakeholders. Once you have identified the optimization questions that apply to the project life cycle stage, you can apply geospatial analysis in optimization activities.

Geospatial analysis supports optimization activities throughout the following steps in project management for environmental sites:

**1. Review the conceptual site model (CSM) and project goals.** Before beginning the geospatial analysis, review the conceptual site model and identify the goals of the project that may benefit from geospatial analysis. See the Data Requirements for Geospatial Analysis section for the relative elements of the CSM for your project.

**2. Perform exploratory data analysis (EDA).** The initial data analysis includes both standard EDA techniques such as histograms, as well as additional techniques especially designed to work with spatial data. EDA can also include mapping the data using a simple interpolation method. Based on the results of EDA, it may be useful to transform or detrend the data before further analysis.

General exploratory data analysis (EDA) methods allow you to check the quality and distribution of an entire data set and select appropriate statistical methods. EDA methods include descriptive statistics such as measures of centrality (mean, median), spread (standard deviation, variance, interquartile range), and shape (skewness and kurtosis), as well as graphical displays (histograms, box plots, scatter plots, and probability plots). Generally, spatial data are not completely independent; each measurement is correlated to some degree with its neighbors. Therefore, review and verify not only the outliers in the entire data set, but also observations that are unusual with respect to their neighbors. The overall distribution of the entire data set (for example, log normal) may not be the appropriate distribution for spatially correlated data.

**3. Choose the geospatial method.** Select a geospatial method based on the CSM, the purpose of the geospatial analysis, and the results of EDA. Several flow charts can assist you in selecting the appropriate geospatial methods for given optimization questions and site conditions. The geospatial methods are grouped by their type (simple, more complex, advanced). Simple methods estimate values at unsampled points with no statistical error model and work best with larger data sets. Examples include inverse distance weighting (IDW), natural neighbor interpolation, and Thiessen polygon interpolation.

More complex methods can predict the variable of interest based on functions of the coordinates, but may work best when additional predictor variables are available (for example, distance from a river is an additional predictor variable when analyzing groundwater concentrations). More complex methods include parametric and nonparametric regression, local linear regression, and splines and kernel methods. Advanced methods use an explicit model of spatial correlation and include all varieties of kriging and conditional simulation.

All of the methods can be used for interpolation, which produces a smoothed estimate of unknown values using known values and the distance between the known and unknown sample locations (see the Methods section for a more detailed discussion of the geospatial methods). In contrast, conditional simulation is used to produce a series of randomly simulated predictions based on the selected spatial correlation model that on average matches the data at known locations. The simulated values are less smooth and therefore more realistic than the interpolation results from other methods.

**4. Build the geospatial model.** Building the geospatial model is done by estimating the model parameters. For simple methods, this is easy because there are few decisions to be made. The data are run through the method to create predictions. However, for all methods it is important to consider whether the key features of the CSM (for example, faults or rivers) are being properly represented in the resulting maps. For more complex methods, building the model may require investigating which sources of secondary information are the most useful for improving the predictions. For advanced methods, variography is generally needed to fit a model of spatial correlation to the data. For all methods, cross-validation can be used to help select the best values of the method parameters.

## Fact Sheet 3: How Is Geospatial Analysis Applied? *continued*

**5. Check that the model produces reasonable results.** One optimization objective is to manage the uncertainty in the project life cycle. Uncertainty arises from the inability to sample all locations in the study area, and from factors such as sampling bias, sampling error, and underlying variation in measured environmental properties.

Before using the predictions or other results, assess the overall accuracy of the interpolation method or model. The interpolated values should be checked to make sure they are consistent with the CSM and other source of information, as well as fundamental scientific laws (for example, the basic principles of hydrogeology and fate and transport). In addition, the model should be evaluated using formal statistical methods for assessing model fit. The accuracy of geospatial interpolation methods can be assessed through cross-validation and validation (see the Evaluate Geospatial Method Accuracy section), which compares model predictions to measured values. Methods to evaluate model uncertainty are described in the Generate Geospatial Analysis Results section.

**6. Generate geospatial results.** Appropriate methods are selected depending on the goals of the investigation, the results of EDA, and spatial correlation modeling. Generally, more than one method may be appropriate. Each method requires the selection of parameter values that control the details of how the method works.

**7. Use the results in optimization.** Knowing how the results of the geospatial analysis will be used in the lines of evidence for optimization may aid in selecting a specific geospatial method. In selecting the method to use to support optimization, evaluate how the results of the different methods fit a specific optimization situation listed in the Using Analysis Results for Optimization section.

See Fact Sheet 4 for information about available software packages for implementing geospatial methods or see the Methods section for a more detailed discussion of the geospatial methods.



### Geostatistics for Remediation Optimization (GRO) Team Contacts

**Ning-Wu Chang** • California Department of Toxic Substances Control  
714-484-5485 • [nchang@dtsc.ca.gov](mailto:nchang@dtsc.ca.gov)

**Harold Templin** • Indiana Department of Environmental Management  
317-232-8711 • [htemplin@idem.in.gov](mailto:htemplin@idem.in.gov)

November 2016



**ITRC**  
50 F St. NW, Suite 350  
Washington, DC 20001  
[itrcweb.org](http://itrcweb.org)

