



## Geospatial Analysis Work Flow Overview

As shown in the flow charts, the work flow for geospatial methods includes evaluating site data, developing a geospatial model, and then using various methods to both generate and validate predictions.

### 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. Also check that the [data requirements for geospatial analysis](#) are met. [Geospatial Methods for Optimization Questions](#) reviews how geospatial methods can be applied throughout the project life cycle.

### 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.

### 3. Choose the geospatial method.

Select a geospatial method based on the CSM, the purpose of the geospatial analysis, and the results of EDA. The software availability may also constrain the types of methods that can be used. Simple interpolation methods, such as inverse distance weighting, can be used for mapping data in unsampled locations but cannot be used for many optimization tasks because these methods do not provide a measure of prediction uncertainty. If estimates of prediction uncertainty are needed, then a more complex or advanced method must be used.

More complex geospatial methods, such as spatial regression methods, can take advantage of secondary data that is correlated with the variable of primary interest. A more complex method such as splines can incorporate important natural features (such as rivers or faults) into the interpolation.

Advanced geospatial methods, such as kriging, rely on an explicit statistical model of spatial correlation. As a result, advanced methods tend to produce the most accurate predictions and estimates of prediction uncertainty. In order to realize these advantages, however, advanced methods require more effort because the spatial correlation model must be estimated from the data, typically using an analysis of the variogram or covariance function. For example, if EDA indicates that there is anisotropy present, then the anisotropy should be incorporated into the variogram. In many cases, more than one method must be evaluated in order to find an approach that best meets the project goals; see guidance on [choosing a geospatial method](#) and [flow charts](#).

### 4. Build geospatial model.

The next step is to build the geospatial model of the data 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.

### 5. Check that the model produces reasonable results.

Before using the predictions or other results, it is important to 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 primary model assessment methods are [cross-validation](#) and [validation](#). Measures of model accuracy can also be compared to each other in order to select among alternative models; see the [Evaluate Geospatial Method Accuracy](#) section.

## 6. Generate geospatial results.

Appropriate methods are selected depending on the goals of the investigation and 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. For example, kriging methods usually employ a search neighborhood that controls how many nearby data points are used during the prediction at new locations. It is useful to perform several different interpolations, using a combination of different methods or different parameter choices within the same method.

Some methods can quantitatively assess the uncertainty in the predictions. For many optimization tasks (such as monitoring optimization), the output from the geospatial method that is most useful is the estimate of uncertainty in the predictions. Mapping the uncertainty estimates can illustrate areas that may require additional sampling as opposed to areas that may already have adequate sampling. The uncertainty estimates produced by the different geospatial methods do not generally include the effects of all sources of uncertainty. Thus, it is important to also evaluate other sources of uncertainty qualitatively, such as uncertainty in the CSM and representativeness of the sampling data.

## 7. Use results in optimization.

The predictions and characterization of uncertainty can be used in various ways to support the optimization. Examples of how to [use results](#) are provided.