



Common Misapplications of Geospatial Analysis

Misuses of geospatial methods frequently appear in professional practice. Listed below are common misapplications of these methods and appropriate alternatives to the erroneous practices. The problems and errors below can occur during all life cycle phases and are grouped by four general topic areas: CSM, EDA, model use, and model assumptions.

CSM [▼Read more](#)

- 1. Problem/Error (CSM):** Geospatial methods performed based on insufficient site characterization or an insufficient number of data points. The spatial distribution and density of sample locations affect the results of [kriging](#). Spatial data gaps can be amplified with kriging, in ways that do not fit with the CSM of a site. For example, water level or contaminant concentration data from distinct aquifers separated by a confining unit may be lumped together for interpolation due to insufficient data, a poor understanding of the CSM, or both.
Recommendation: Use a geospatial analysis approach that is consistent with the CSM (for example, separate data into discrete vertical geologic units before performing geospatial methods). Evaluate number of available data points relative to guidance on [minimum requirements for data sets](#). Exercise professional judgment before using the raw results of geospatial interpolation to determine whether the results represent the hydrogeologic system well. You should have a good understanding of the physical parameters of the system and contaminant distribution before applying geospatial methods. Perform cross-validation and uncertainty analysis on geospatial interpolations to evaluate whether error is acceptable, as the number and orientation of sampling locations influence this error.
- 2. Problem/Error (CSM):** Geospatial predictions are extrapolated beyond the spatial extent of data measurements without a rational or defensible basis. Plumes for which data have not been collected to bound the outer edges can be extended by the kriging algorithm well beyond where they are believed to end. When defining the grid within which kriging is performed, it is generally best to interpolate between points, not extrapolate beyond them. As such, a convex hull around existing data points is usually used rather than a rectangular grid fit to the outermost XY extents of the data set.
Recommendation: [Define the boundary of a plume](#) before fitting a geospatial model to the site data. Clip the interpolated surface or contour lines to this inferred plume boundary or the convex hull of the data used in the interpolation. Use a [more complex](#) geospatial model when extrapolating predictions beyond the convex hull of the data, such as appropriately applied kriging, instead of a simple method.
- 3. Problem/Error (CSM):** Not adequately considering temporal variability, not only seasonal, but also diurnal for tidally influenced groundwater. Temporal variability may not be adequately understood with only quarterly sampling over a year.
Recommendation: The [CSM](#) must adequately characterize the site prior to using geospatial methods for optimization, including temporal variability. Geospatial analysis of spatial data should be performed for different time periods to evaluate this variability. For example, separate spatial data interpolations may be performed on a diurnal, monthly, or seasonal basis. Geospatial analysis of time series data at individual monitoring locations may also help to explain these fluctuations.
- 4. Problem/Error (CSM):** Adding artificial data points to a kriged data set to constrain a plume or make the results conform better to a specific interpretation of data without a defensible basis. Where appropriately justified and explained, this approach may be defensible for situations such as constraining a soil vapor plume by adding artificial nondetect points along a bedrock outcrop or other feature that is definitively bounding migration of contaminants (without needing to demonstrate the case with actual sample data).
Recommendation: The CSM must be updated to include [known boundaries](#) in order to apply geospatial methods. The addition of artificial data points should be transparently explained and justified by the [CSM](#).
- 5. Problem/Error (CSM):** Using geospatial methods to generate a geologic model of continuous layers when it is not appropriate given the heterogeneity of the geologic system (for example, by trying to model interbedded alluvial deposits as continuous layers of discrete soil types). In certain instances, this practice may result in

unrealistic situations such as a thin clay layer extending across more than a mile, despite a heterogeneous depositional environment.

Recommendation: Prior to using a model for optimization, the [CSM](#) must adequately define the geology (see [Table 1](#)); use professional judgment to [incorporate stratigraphy](#) in geospatial analysis.

6. **Problem/Error (CSM):** When lag spacing is not on the same scale as the geologic heterogeneities. When choosing appropriate [lag spacing](#), the sample spacing controls this selection, not the geology.

Recommendation: Geology is always a consideration when [making a variogram](#). Evaluate the distances between the sampling locations and consider additional sampling points based on the changes in geology so that the lag spacing does not cross over significant differences in geology.

EDA [▼Read more](#)

7. **Problem/Error (EDA):** Failing to account for the influence of censored (nondetect) data in geostatistical analysis. In some cases, elevated reporting limits may make substitution methods (for example, replacing nondetect values with one half the reporting limit) cause erroneous data interpolations. Similarly, inconsistent reporting limits in time and space may cause interpolations to appear different over time.

Recommendation: This problem is an EDA issue; see [GSMC](#) for guidance on managing nondetect data. Where there is temporal or spatial variability in reporting limits, consider performing geostatistical analysis using different assumptions for nondetect values. Where a very high percentage of data points are nondetect, geostatistical methods may not be appropriate.

8. **Problem/Error (EDA):** Using poor variogram models that do not fit the data well. In some cases, this practice results from using the automatic fit functions of geospatial software, resulting in unrealistic contour plots that may be used for site decisions.

Recommendation: This problem is an EDA issue; see [GSMC](#) for guidance. Evaluate the goodness-of-fit of variogram models and use cross-validation/validation to assess prediction accuracy.

9. **Problem/Error (EDA):** Creating sample variograms with inadequate amounts of data in the bins. This error is common with environmental sampling data for sparse networks.

Recommendation: Before applying the geospatial analysis, [obtain a sufficient number of samples](#).

10. **Problem/Error (EDA):** Not accounting for an underlying trend or anisotropy in the environmental spatial data.
Recommendation: The data must satisfy the underlying assumptions of a particular geospatial method, such as [stationarity](#), or else you should not use that approach. For example, detrending may be required to make the data approximate stationarity for proper application of kriging. Use tools that filter out underlying trends and transform data to allow for kriging. Model the underlying trend as a function of location, filter out the underlying trend from the data, generate the variogram, and perform the kriging on the detrended data, then add the trend back to the kriged values. Chapter 9 of *Geostatistics for Environmental Scientists* ([Webster and Oliver 2001](#)) discusses this option and includes other references. It may be necessary to use more complex geospatial methods when geologic anisotropy in fractured flow or a high degree of heterogeneity in the contaminant distribution exists.

Model Use [▼Read more](#)

11. **Problem/Error (Model Use):** Not exercising professional judgment before using the raw results of geospatial methods for optimization to ensure the results represent the hydrogeologic system well. For example, if surface water elevation data are not incorporated into a potentiometric surface data set, then the resulting geospatial interpolation may completely misrepresent groundwater-surface water interactions.

Recommendation: Understand that geospatial methods are merely tools, it is important to assess the overall accuracy of the interpolation method or model. You must have a good understanding of the physical parameters of the system and data distribution before interpolating data. 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).

12. **Problem/Error (Model Use):** Attempting to use a grid process to perform an interpolation between known data points. For example, using natural neighbor works by assigning the value of the nearest data point to each grid

node ([Golden Software 2002](#)), and does not interpolate between data points. Where data are not on a dense grid, nearest neighbor creates polygons of uniform values and the resulting contour lines are not usable.

Recommendation: Understand the difference between gridding and interpolation processes. Gridding assigns values and does not interpolate. Nearest neighbor is more accurately described as a gridding method instead of an interpolation method. Nearest neighbor can reliably convert data points on a sampling grid to a continuous raster file, such as a Surfer grid. This makes the method useful for applications such as converting one type of grid to another format (for example, an ESRI grid to a Surfer grid). You should use one of the interpolation methods, such as [natural neighbor](#), to perform interpolation.

13. **Problem/Error (Model Use):** Failing to account for uncertainty when applying an interpolation model. For example, using a mechanistic approach such as inverse distance weighting interpolation when there is significant uncertainty in the data set. CSM uncertainty can be propagated into an interpolation model. Geospatial methods can be used to identify data gaps in the CSM, as well as estimate values in locations not sampled, but these methods may incorrectly predict values if there is a high variability in the data or if the causes of the data gaps are misinterpreted.
Recommendation: The [cross-validation error](#) should be determined and professional judgment must be used in accepting the modeling tool and optimizing the sampling program. [More complex methods](#) such as kriging can be used to assess uncertainty. Condition simulation is the most robust [means of assessing uncertainty](#).
14. **Problem/Error (Model Use):** Overuse of smoothing techniques or using an improperly selected contour interval to hide irregularities in the data.
Recommendation: Reviewers should ask for appropriate contouring intervals given the scale of the problem at hand. For example, if a groundwater cleanup level is 10 micrograms per liter, using a contour interval of 50 micrograms per liter may be inappropriate.

Model Assumptions ▼ [Read more](#)

15. **Problem/Error (Model Assumption):** When applying kriging, failing to incorporate a nugget effect when there is a positive Y-axis intercept on a variogram, indicating variability over short separate distances. This practice can produce an exact interpolation map; however, it does not appropriately account for error in the data set. This error often occurs because regulators may be reluctant to accept inexact data interpolation.
Recommendation: Apply kriging using the best possible [variogram model](#), including a nugget effect where appropriate. There is always some acceptance of error in the data, so the acceptable error rate should be addressed prior to using a geospatial analysis. With the interpolation, explain the [modeling assumptions](#), how the data were interpreted, and the degree of accuracy of the interpolation. Discuss the use of inexact interpolation models with project stakeholders and come to consensus on the best model for optimization based on cross-validation/validation and other uncertainty analysis, where appropriate.
16. **Problem/Error (Model Assumption):** A method may rely on certain underlying assumptions, such as a uniform distribution of data points, but these assumptions may not be satisfied by the data set or communicated to the method reviewer. For example, simple methods generally work best with homogeneity of the subsurface and a dense, uniform sampling grid, yet this assumption cannot be sustained when data points are too far from one another relative to the scale of the subsurface heterogeneity. As an example, fractured bedrock may be quite heterogeneous on a scale of inches, while well-sorted sands are homogenous on a scale of many feet.
Recommendation: Always verify [assumptions](#) of any geospatial method and communicate the satisfaction of model assumptions to [reviewers](#). More complex methods such as kriging should be used for more complex subsurface conditions.
17. **Problem/Error (Model Assumption):** Applying sampling optimization programs or modules and recommending removing wells or reducing sampling frequency when doing so would conflict with the data quality objectives of the monitoring. For example, a well that may be statistically redundant for a particular chemical may be needed for water level gauging or a different chemical that was not analyzed with the optimization software. Alternatively, it may not be appropriate to reduce monitoring frequency during remediation when historical trends may no longer be valid.
Recommendation: The site may not be ready for [optimization](#) yet. Establish goals for [monitoring optimization](#) that are compatible with site management requirements moving forward.