



Sample Spacing

Geospatial analysis can be used to guide the selection of sample spacing or sampling plan design. The first step is to conduct an evaluation of the [spatial correlation](#) of existing data using one of the following EDA methods: (1) plot the empirical variogram, covariance function, or correlation function; or (2) plot the h-scatterplot. Consider whether the spatial trend should be subtracted from the data before constructing the plot. From the plot, determine the approximate correlation range of the data, which is the maximum distance over which the data are correlated. Ideally, a sampling plan defines a sample spacing less than the correlation distance between measured values of the quantity of interest. As a conservative rule of thumb, the sample spacing should be half the range of spatial correlation identified with EDA.

Understanding the Results: [▼Read more](#)

The spatial distance or temporal frequency over which a property exhibits autocorrelation is not always straightforward or generic at project sites. Thus, sampling designs can sometimes over- or under-sample the area of interest. The sampling interval can be best determined using preliminary data sets (if available) and other site-specific information (for example, soil maps, geologic maps, or digital elevation models). The sampling interval in geospatial analyses can help in determining a defensibly adequate number of samples for a sampling program and can be determined using [variography](#).

The sampling design should reflect the natural heterogeneity of the data and how the data are related to precision and representativeness of the locations being sampled. The environments that are sampled can be highly heterogeneous in time or space. Therefore, one discrete sample may not represent a particular location or point in time. This heterogeneity can also make it difficult to characterize the variation of a property or process between sampling locations (autocorrelation) or within a single sampling location (for example, laboratory precision). This difficulty can be addressed by using composite sampling, as described in EPA QA/G-5S ([USEPA 2002a](#)), or by calculating the average of multiple samples representative of a single sampling location. If economically feasible, the latter better quantifies the precision of the sampling support to determine whether it is adequate.

Because the potential extent of contamination is rarely known at the outset of an environmental project, the sampling extent is often determined by the sampling program results and is not necessarily evident at the start of sampling. Therefore, a sampling program might propose a sampling extent based on a CSM or other site specific information and optimize this extent upon further data generation and evaluation.

Several geostatistical methods (for example, variograms) guidance documents and free software packages (for example, [VSP](#)) are available to help estimate the spatial distribution and number of samples needed to build a defensible and representative sampling program. Be aware that software packages and guidance documents that do not take into account autocorrelation often incorrectly estimate the number of samples needed.