Systematic Random Sampling Workbook User's Guide

Systematic random sampling is a grid sampling design with a random starting point. Systematic random sampling provides better coverage of the soil study area than simple random sampling. Therefore, this DEP spreadsheet was developed to assist the remediator with creating a grid covering the soil study area with randomly chosen sampling locations. It is based on a triangular systematic random sampling approach.

The basic principles of the spreadsheet:

> First, the spreadsheet will randomly generate a starting point, which can be anywhere within the rectangular area defined by the dimensions of the soil study area, X_{min}, X_{max}, Y_{min} and Y_{max}. Based on the starting point, the spreadsheet will generate a triangular grid pattern. The starting point is located at the '0^{th}' row of the created grid pattern. Above the '0^{th}' row are 1^{st} row, 2^{nd} row, 3^{rd} row, etc. Below the '0^{th}' row are -1^{st} row, -2^{nd} row, -3^{rd} row, etc. This triangular grid pattern is then used to define diamond-shaped cells. Each diamond-shaped cell (such as Cell A as illustrated in the example plot of the 'DataInput' worksheet) is defined by four grid points of the triangular grid.
>
> Finally, the spreadsheet randomly generates one and only one sampling point in each diamond-shaped cell.
>
> A diamond-shaped cell located at the border of the study area rectangle as defined by X_{min}, X_{max}, Y_{min}, and Y_{max} may have part of the cell within the study area rectangle and part outside the study area rectangle. The spreadsheet will randomly generate a sampling point for each of such cells as well. If the sampling point is within the study area rectangle, it will be present in the 'SamplingPoints' worksheet. Otherwise, it should be hidden from viewing.

It is important to note the cells of the excel spreadsheet requiring data input have been shaded yellow. The remediator should not enter any data into the white cells on the 'DataInput' tab of the spreadsheet program.

This spreadsheet was designed for 3D sampling. For example, if a remediator was implementing in-situ bioremediation of contaminated soil, the remediator would need to collect samples from the contaminated soil "volume", a 3D feature. However, in order to use this spreadsheet to determine confirmation sampling points on sidewalls and the bottom of an excavation pit (attainment sampling), which is essentially a 2D sampling scenario, some adjustments to the data input are required.

First, input the "area of contamination" ('DataInput' cell B3) in square feet. If the study area consists of an excavation, and sampling of sidewalls and the excavation bottom are needed, the pit "box" should be 'opened', meaning, each of the sidewalls and bottom should be drawn on a 2D basis, and the x-axis and y-axis overlain on this drawing. The "area of contamination" would be the total area of the sidewalls and pit bottom of an excavation. X_{min} ('DataInput' cell B18), X_{max} ('DataInput' cell B19), Y_{min} ('DataInput'

cell B20), and Y_{max} ('DataInput' cell B21) can then be entered in the spreadsheet (in feet).

'DataInput' cell B4 and 'DataInput' cell D4 are inputs for 3D sampling only. These are used to define the sampling depth range. For example, if a remediator is collecting samples from 2' to 15' in a 3D sampling scenario, the remediator would enter 2 into 'DataInput' cell B4 and enter 15 into 'DataInput' cell D4. For 2D sampling in excavation cases, enter '0' (zero) into 'DataInput' cell B4 and 'DataInput' cell D4. All z-coordinate output values on the 'SamplingPoints' worksheet should not appear if the remediator enters '0' (zero) into 'DataInput' cell B4 and 'DataInput' cell D4.

Enter the volume of contaminated soil into 'DataInput' cell B5 (in cubic yards). This value should be based on site characterization data (This will not necessarily equal 'DataInput' cell B3 multiplied by 'DataInput' cell B4).

Note that the spreadsheet randomly generates one and only one sampling point in each diamond-shaped cell as mentioned previously in the spreadsheet principals. Because the randomly generated triangular grid pattern may not match with the 2D layout of the sampling area exactly, it is possible that some generated sample locations may not be within the study area. If this occurs, continue to click on the button of "Click Here to Generate a New Triangular Grid" for a new set of randomly generated sampling locations until the correct number of samples are within the study area.

For 3D sampling, 'DataInput' cell B12 is for the number of samples the remediator plans to collect within the contaminated soil volume. For 2D sampling, such as for attainment demonstration in excavation cases, again the remediator should enter the number of samples planned for collection on the side walls and pit bottom into 'DataInput' cell B12. As mentioned above, a click on the button entitled "Click Here to Generate a New Triangular Grid" may not generate the exact number of sampling points that the user may desire within the sampling area. To remedy this, the user should repeat clicking this button until the exact number of sampling points desired within the sampling area has been randomly selected. If too many sample locations are generated within the study area, do not discard any sampling point manually. This would violate the purpose of systematic random sampling. However, any excess sampling points generated on the 'SamplingPoints' tab of the spreadsheet which appear that may be outside the sampling area may be discarded.

If you have any additional questions, please contact:
Frank Nemec, P.G.
PA DEP
Bureau of Environmental Cleanup and Brownfields
e-mail: fnemec@pa.gov
phone: 717.783.9483