Conditioned Latin Hypercube Tool

Conditioned Latin Hypercube Sampling (cLHS) is a stratified random sampling technique for multidimensional data that uses a constrained Monte-Carlo sampling scheme to maximize sample site selection efficiency. While several sample design approaches, such as k-means clustering, have been designed to ensure an even distribution of sample sites in geographic space, cLHS is designed to ensure balanced selection of sites across feature or data space. The tool ensures that samples are drawn only from the existing range of ancillary data such as environmental raster data. Therefore, the sample population drawn from a digital elevation model (DEM) would be restricted to only those values existing in the data given to the tool. The cLHS tool has been widely adopted within the digital soils mapping community for the selection of efficient field sampling locations.

Minasny, Budiman, and Alex B. McBratney. "A conditioned Latin hypercube method for sampling in the presence of ancillary information." Computers & Geosciences 32.9 (2006): 1378-1388.

Data Requirements:

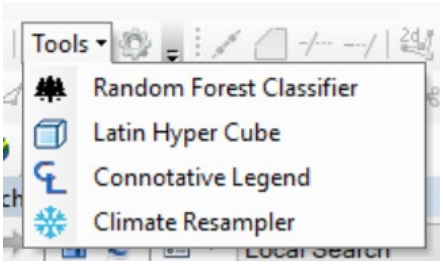
- Raster data must all cover the same extent
- Raster data must all be in the same projection or the tool will fail

Contents

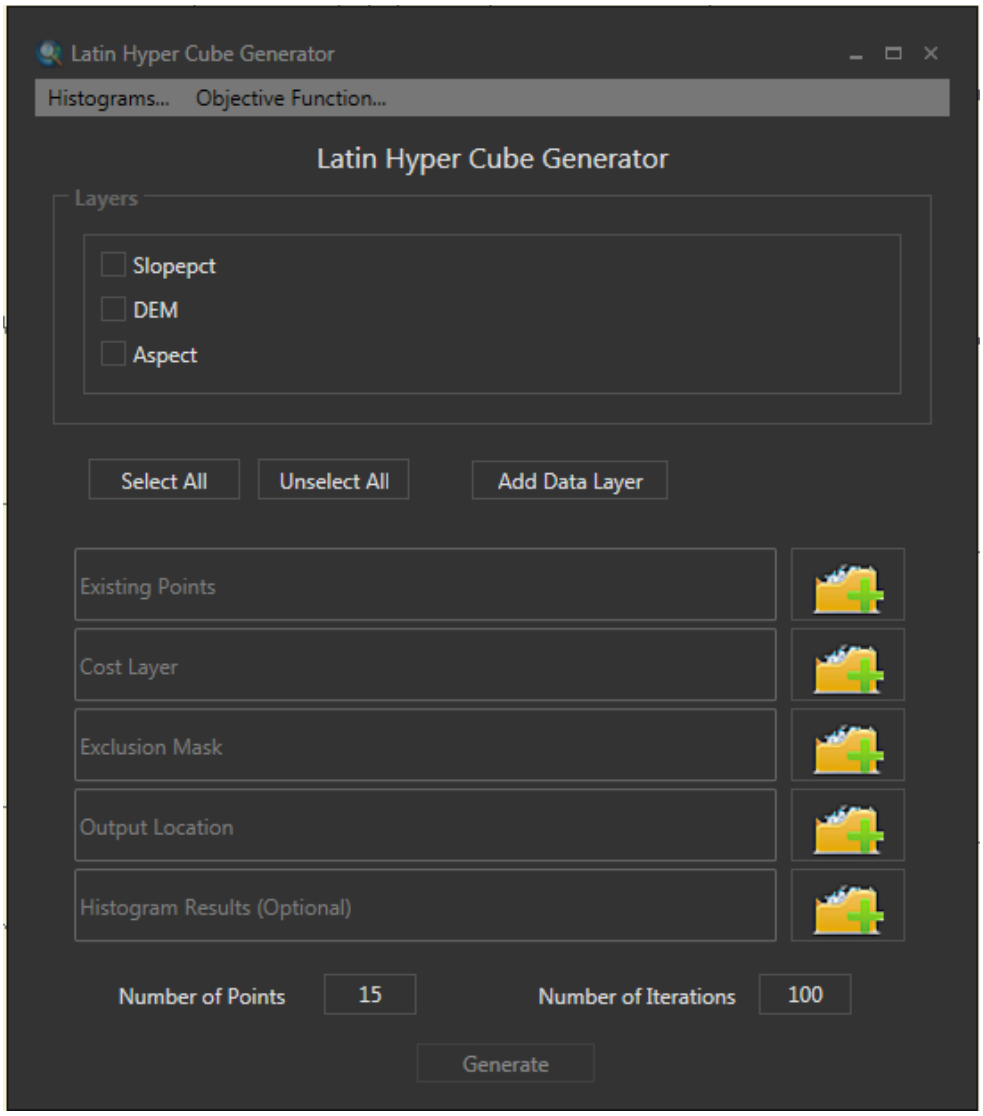
Part 1: Set Up and Run the Tool

A. Add data layers

1. On the toolbar, click the Tools menu then click **Latin Hyper Cube**.



2. If you have a current TEUI Toolkit project open, the tool will automatically add those layers to the layer list. Otherwise, select the **Add Data Layer** button.



3. A window will appear allowing you to navigate to the raster data of choice.

    - Your raster data can be either continuous or discrete. You can select as many layers as you wish. Be aware though, the larger your project area and the more raster layers you choose, the longer the tool will take to run.

4. Be sure to **place a check in the box next to each raster layer** you want included.

5. Next, there are several additional options to choose from. None of the following are required to run the tool.

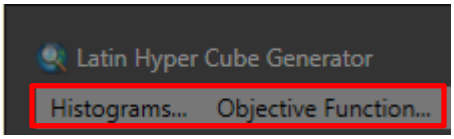
    - **Existing points**-These are point data (as a .gdb or .shp) of existing sample locations that you would like to have excluded from selection in the next run. For example, if you have existing plot data, you would not want to resample those locations.

- **Cost Layer**-The cost layer is a raster layer that indicates a higher "cost" to sample either in effort or in money (e.g., travel costs). Therefore, low values are easier to sample, higher values are harder to sample. One example might be that greater "effort" is required to sample sites that are farther from a road.
- **Exclusion mask**-An exclusion mask is a 0/1 raster where a value of 1 indicates areas that are excluded from being selected, such as the middle of a lake or an area with limited access (e.g., the top of a mountain).

6. Select the folder icon next to the **Output Location** field to select the file name and location for the resulting sample site point file (.shp).

The Erdas 2011 Extensions for ArcGIS can potentially interfere with the Latin Hypercube Tool. If ArcGIS crashes when you try to set the Output Location, you may need to uninstall the Erdas 2011 Extensions to avoid the problem.

7. The **Histogram Results (recommended)** field allows users to save the resulting histogram and points graph outputs for further inspection. The Histrograms and Objective Function windows in the tool can't be utilized if you don't use this option.

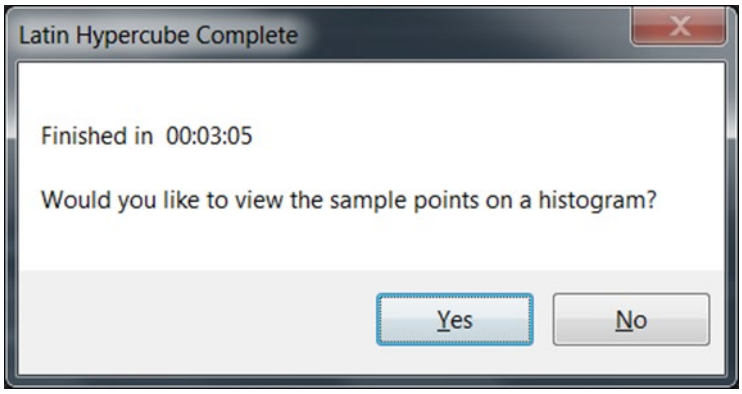


8. Enter the **number of sample points** you would like to have returned (minimum of 1).

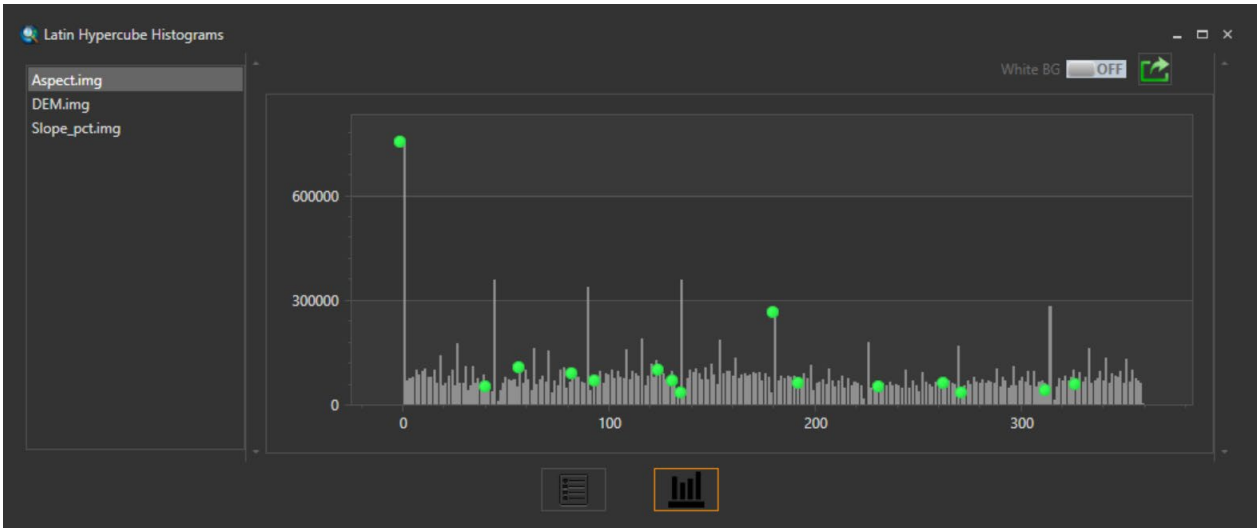
9. Select the **number of iterations** (Monte Carlo iterations). A higher number will result in more processing time required.

10. Click on **Generate** to create your sample locations. The tool is very processing intensive, so please be patient! The tool's progress bar is located at the bottom of the dialog.

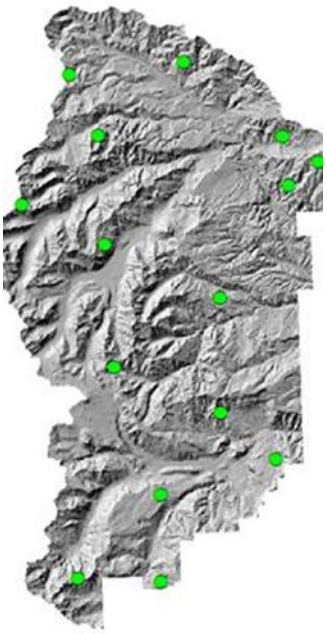
11. If you opted to receive histogram results, you will receive a dialog box when the tool is finished, asking if you would like to review the output histogram. Selecting yes will bring up the Latin Hypercube Histogram window in which you can view a histogram for each raster layer. The graph also shows where the sample points fall on the histogram.



12. In the Latin Hypercube Histogram window, users can select the raster layer of interest on the left panel. Previous Latin hypercube runs (if available) can be investigated by selecting the **List of Runs** button at the bottom left of the window. The graph can be printed by clicking the green arrow button in the upper right hand corner of the window.



13. The resulting product will be a point shapefile layer. The layer will be added to the ArcMap table of contents. The value of all raster cells under each point is appended to the attribute table of the point layer.



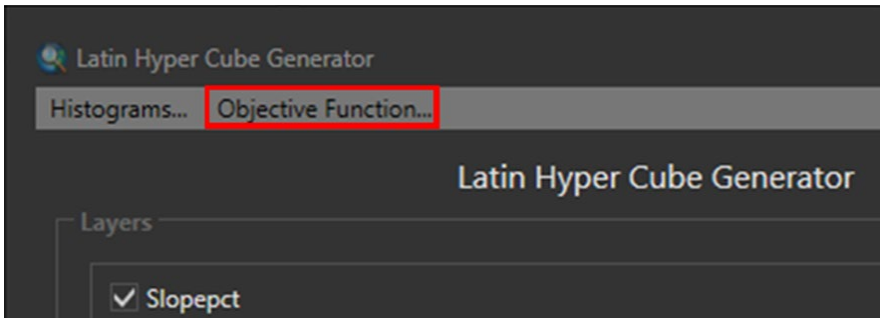
Table

Output.shp

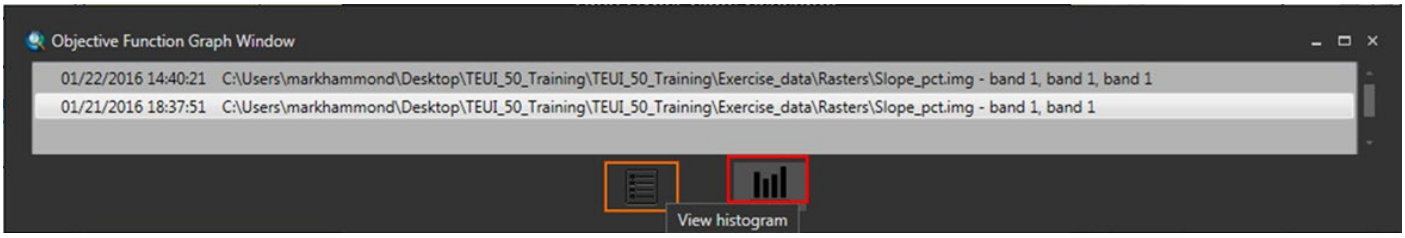
FID	Shape	Id	lat	long	Aspect	Slope_pct	DEM
0	Point	0	5178405.53172	629257.888391	60.524112701416	33.0246124267578	1477
1	Point	1	5203275.53172	653927.888391	189.462326049805	15.2069063186646	1736
2	Point	2	5162895.53172	634937.888391	326.309936523438	18.0277557373047	1392
3	Point	3	5215375.53172	637657.888391	168.690063476563	38.2426452636719	1427
4	Point	4	5167195.53172	648827.888391	40.2363586425781	21.2867336273193	1122
5	Point	5	5206505.53172	627327.888391	261.46923828125	50.5593719482422	1634
6	Point	6	5200385.53172	650367.888391	271.080932617188	66.2617950439453	1048
7	Point	7	5198065.53172	618087.888391	7.69605159759521	46.6703872680664	1665
8	Point	8	5213885.53172	623787.888391	118.610458374023	31.3249111175537	1411
9	Point	9	5186785.53172	642077.888391	135	8.83883476257324	1326
10	Point	10	5193285.53172	628147.888391	101.309928894043	25.4950981140137	1564
11	Point	11	5152335.53172	634917.888391	30.9637565612793	29.1547603607178	1729
12	Point	12	5152755.53172	624847.888391	351.158172607422	56.9264869689941	1966
13	Point	13	5206395.53172	649657.888391	227.602569580078	38.9310302734375	1248
14	Point	14	5172895.53172	642127.888391	187.594650268555	37.8318634033203	1441

14. Click the **Objective Function** button near the top of the Latin Hypercube Generator window to see a graph of the Objective Function values.

- In simple terms, the objective function is akin to the probability of selecting a better or a more “globally representative” matrix of values from multiple input raster data layers with each increasing iteration. In most cases, the objective value never gets to zero before the time or the number of iterations is reached. Therefore, one can examine the plot of the objective function (y- axis) vs iteration number (x-axis) to determine if more iterations are needed to reach an acceptable or optimal sample selection. The Toolkit implements a streamlined or simplified cLHS process. We suggest the use of the ‘cLHS’ package in the R statistical software should one want more control over the entire cLHS process.



15. In the Objective Function Graph Window, double-click a file in the list of runs to open the graph or single click and press the **View histogram** button at the bottom right of the window.



16. Hover your cursor over the graph line to see the value drawn for a given iteration of the Objective Function. The closer the value is to zero the better the sample is. If you want the Objective Function to achieve a lower value then set a higher "Number of Iterations" value and run the cLHS tool again.