**Tem\_lhs.py**

**Loading Data:**

The user loads a specified tTEM .xyz inversion file as a a Pandas dataframe, and the necessary variables are extracted as Numpy arrays. These variables include the geographic coordinates (UTM) of each survey point, also referred to as a sounding point, in the resistivity model, the depth intervals of the model, the resistivity values of each layer (translated to Log-resistivity), the elevation of each sounding, the data residuals for each sounding, and the depth of investigation (DOI) for each sounding.

**Model Reduction:**

The user must next remove all portions of the model not relevant to simulated annealing: layers below the maximum logging depth, and soundings outside the prescribed study area, or are otherwise known to be unusable. All resistivity model cells outside these boundaries are completely removed.

**Define Inclusion Criteria:**

Next, we define a set of partially excluded soundings. These are soundings that cannot be sampled but are still relevant to the statistics of the resistivity model and will influence which samples are selected. These soundings come in two forms: those excluded due to their proximity to the edge of the survey area, and those with very high data residuals.

Selecting the area where soundings are included can be done either by describing a polygon where the soundings are included, OR setting the exclusion criteria according to how many other soundings need to be within a prescribed radius of each sounding. If there are too few adjacent soundings, this sounding is likely close to the edge of the survey grid and should therefore be excluded. This typically assumes a dense, regular grid of TEM soundings.

The data residuals partial exclusion criteria assumes that soundings with very high data residuals should not be sampled so we set a cutoff according to a quantile (e.g. 99th percentile). Any soundings below the data residual cutoff defined by the input percentile value are included and any above the cutoff are partially excluded.

Note: all soundings can be included by setting the inclusion polygon around the entire survey area and setting the data residual quantile to 1.0 (100th percentile).

Finally, we plot the set of included and excluded soundings.

**Define Cumulative Lateral Variance:**

Next, we calculate the weighted cumulative lateral variance for each sounding. We define a search radius around each sounding, which expands into the subsurface according to the physics of the TEM experiment. The log-resistivity variance in each layer using the search radius within the defined is calculated and weighted by the thickness of that depth layer. Finally, we add the weighted variance of each layer, giving the weighted cumulative lateral variance at that point. This is repeated for each sounding and then plotted.

**Principal Component Analysis:**

Next, we use principal component analysis to reduce the dimensionality of the resistivity model, based on a number of user-defined principal components (PCs). This effectively reduces a resistivity model that may have 10+ layers to a much smaller number of PCs that account for the majority of the variance in the resistivity model, which will be easier to sample during simulated annealing. The explained variance ratio and total explained variance for the PCs are displayed; we recommend using enough PCs to explain 90% of the resistivity model variance. Alternatively, the user can decide that they do not want to use PCs to populate their feature space, and can instead set a flag to define their feature space using the original resistivity model.

**Define the Feature Space and Priming Soundings**

Whether or not the PCs or the entire resistivity model are used, the UTM coordinates of the soundings are also included to complete the feature space. Note, the UTM coordinates need to be normalized such that they are on a similar scale to either the PCs or the log-resistivity model range while retaining the relative variance of the easting and northing components. This feature space array will be sampled by the simulated annealing algorithm to select the optimal sampling locations.

Prior to simulated annealing, we can define one more subset of soundings that we will use to “prime” the simulated annealing algorithm. We do this because there is a high degree of lateral correlation between soundings, so we can improve the simulated annealing algorithm converge by temporarily removing the majority our soundings. The remaining “priming” soundings are sampled for a set number of iterations before the entire set of available soundings (not including those excluded due to edge proximity or high data residuals).

**Simulated Annealing**

Simulated annealing parameters are codified and the function is run. This function follows the approach of Minasny and McBratney (2006), <https://doi.org/10.1016/j.cageo.2005.12.009>.

Loosely, the simulated annealing is an iterative random walk used to minimize an objective function, and is regulated by a cooling schedule. It starts with a set of random samples and checks three calculations that feed into the total objective function composed of three parts: sampling stratification, covariance, and weighted cumulative lateral variance. For the first part of the objective function, the algorithm tries to place the samples into equiprobably strata across all components of the feature space simultaneously. The second component of the objective function is concerned with matching the covariance of the sampled points to the covariance of the entire feature space. Finally, the third component of the objective function tries to minimize the maximum weighted cumulative lateral variance of the selected soundings.

The simulated annealing schedule is repeated with different starting soundings a user-specified number of times. The output of the simulated annealing procedure is are the final objective function values at the last iteration of each run, as well as the set soundings obtained from each run.

**Plotting Results and Exports**

A single simulated annealing run can be plotted to evaluate for quality control purposes. First, the evolution of the weighted and total objective function parameters is displayed. Second, the histograms of the feature spaces are plotted, along with the cutoffs for each strata and the values of each sample for each feature space element. Next, the mean and standard deviation of the resistivity model at each layer are plotted, along with the same statistics for the selected set of samples. Finally, the locations of the Additionally, the weighted cumulative lateral variances for the selected soundings are compared to the distribution of values for all input soundings to give quantile values, again useful for quality control.

Finally, the simulated annealing inputs and the final selected soundings and associated objective function values for each simulated annealing run are saved to a pair of .csv files.