

ArcGIS/Arcpy Dataset Generation: (Any time an Arcpy function is used, the same process can be done in ArcGIS pro but may not be automated to the same degree)

- 1) Extract the individual raster layer from all .hdf files(or most other multidimensional raster file types) by using Arcpy's "Subset Multidimensional Raster_md" function, a script that automatically extracts a given layer from .hdf files that are associated with a given list of dates from a given directory, named 'MDR_layer_extractorV2', can be found in GitHub directory.
- 2) Create 30m grid resolution shapefile, the encompasses the entire study area (TBA)
- 3) Create all other resolutions from 30m grid-shapefile. (TBA)
- 4) Generate soil texture data, where it is a ration of sand, silt, and clay which when summed up equals to 100 or 1 depending on the scale. (TBA)
- 5) Use the elevation raster data to generate the elevation, aspect, and the slope of the terrain. (TBA)
- 6) Use the Feature to Points function in Arcpy to convert the 30m grid-shape file to a point-shape file.
- 7) There are two primary ways to organize all the data extracted from the rasters:
 - a) Use Arcpy's "Extract Multi Values to Points" function to extract the values from all the rasters at once. So that all the data is in one shape file attribute table
 - b) Use Arcpy's "Extract Multi Values to Points" function to extract the values from rasters of specific data products and have different point-shape files for each data product. (This method is recommended since it is the most stable, and if there is an error/corruption in one shapefile it should not affect the others)
- 8) Use the ArcGIS function Aggregate points, where the point input is the point-shape file from the previous step and the grid input is the grid-shape file generated in step 3. Then select all the columns that need to be aggregated and select the 'Mean' option for all the columns selected. The output is a raster with an attribute table with all data needed, also make sure when exporting the table that the 'PageName' column is included.

The next steps do not require Arcpy nor ArcGIS

1. The static variables should be put together in one CSV file along with the 'PageName' column. In the GitHub directory, there is a script called 'static_var_gen', which trims off all the non-static variable columns.
2. To generate the primary dataset, which is used to generate the test and train sets, all the datasets should be divided up by date into individual CSV files, where each file contains the data column and the PageName column. All the CSV files should be put into the same folder along with the static CSV file so that 'primary_dataset_gen' can run properly.
 - a. **Warning**, when using different data products, the Data Cleaning [line 82] section of the script will need to be edited to account for fill/error values.
 - b. The CSV file should have this naming structure, 'DataName_index_resolution.csv'. The index can be either the date (formatted as YYYYmmdd) or numerical

- c. Verification data, such as MOISST AirMOSS, insit su, or NDVI data, should have data name 'Ver'
- 3. If there is partial verification data, it will move all rows with verification data into the testing set, if ther is no verification or complete verification data the dataset will be divided up randomly.