

Foresite documentation

Background

The Foresite database provides a centralized repository and intersection of several public geospatial data layers including the 2008 FSA Common Land Unit (CLU) layer, annual NASS Cropland Data Layers (CDL), the NRCS Soil Survey Geographic Database (SSURGO), as well as state, county, watershed boundaries. These intersection of these data sources enables the the generation of field-to-regional scale estimates of soil properties and historical cropping rotations. The database is implemented using PostgreSQL with PostGIS extension for geospatial processing. Data from the Foresite database may be directly queried and exported in a variety of GIS and flat file formats (e.g. .shp, .gdb, .csv, .txt) using third-party software or command line tools. The database also provides a platform for serving crop, soils, and land management practice data directly coupled with cropping system models to be used for environmental and economic analysis. As an example, a Python wrapper module was developed to automate the initialization of the APSIM cropping system model for a specific field-site or region (e.g. county, watershed). The wrapper enables large numbers of inputs to be generated in order to perform regional subfield scale simulations to estimate crop yields and soil carbon and nitrogen cycle impacts (soil organic carbon changes, N₂O emissions, NO₃ leaching, etc.) from varying land management practices.

Data layers

USDA-FSA Common Land Unit (CLU)

The database includes the 2008 Common Land Unit (CLU) data layer (**clu.clu_conus**) that is used to identify individual agricultural field boundary polygons within the U.S. Each CLU polygon is stored as in a WKB geometry format and identified within the Foresite data schema by a unique identifier (**clukey**). Following 2008, the CLU became restricted from public release, and it is therefore a static resource within the Foresite database.

USDA-NASS Cropland Data Layer

Annual USDA-NASS Cropland Data Layers (CDL) from 2008-2019 are stored within the Foresite database to provide historical estimates of crop cover associated with cropping system field boundaries. The annual 30m raster data is available for direct download from the USDA-NASS website (https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php). To estimate crop rotations associated with the CLU polygons, each yearly CDL raster is intersected with the CLU layer to identify annual land covers. The dominant crop within each polygon is then identified based on the maximum number of 30m pixels within each boundary. CDL raster data can be imported into the Foresite database by altering and executing the **cdl_import.sql** and **add_raster_constraints.sql** queries located in the Foresite Github repository (<https://github.com/isuforesite/Foresite.git>). The imported CDL layer is intersected with CLU polygons to identify the yearly dominant crop cover associated with each CLU. The **clu.clu_cdl** table is updated with the release of the CDL each year using the **clu_cdl_intersection.sql** query. The **clu.clu_cdl** table is then used to derive an estimated crop rotation across the available data years.

SSURGO

The 2019 SSURGO soils database provides estimates of physical soil properties associated with spatially explicit polygons covering a majority of the U.S. territory. Additionally, the gSSURGO version provides a rasterized version of the data. Additional information and metadata including the SSURGO database design, property descriptions, and property units is provided at <https://www.nrcs.usda.gov/wps/portal/nrcs/>

detail/soils/survey/?cid=nrcs142p2_053627. SSURGO data is updated annually in addition to the unique identifiers for soils (i.e. **mukey**). Therefore tracking the specific SSURGO version used with a given analysis is required. Outdated versions of SSURGO should be maintained for a maximum of 2-3 years in separate database schemas (i.e. `ssurgo_2019`, `ssurgo_2020`). This will provide a mechanism of support for maintaining aging datasets until a certain point in time at which the data is archived or deleted.

SSURGO data is updated annually and available for direct download from the USDA-NRCS following 2019 (<https://nrcs.app.box.com/v/soils>; `gSSURGO_CONUS.gdb.zip`). The `.gdb` files. Contained within the downloaded archive are imported into the Foresite database using the GDAL **ogr2ogr** command line utility (<https://gdal.org/programs/ogr2ogr.html>). The **ssurgo_import.txt** in the Foresite repository provides explicit commands for using the `ogr2ogr` tool to importing the SSURGO database (`.gdb`). After successfully importing SSURGO, the `ssurgo_import.sql` query is used to ensure all soil mapunit polygon geometries are valid and add constraints and indexes to the SSURGO data tables. Spatially explicit boundaries defining the unique SSURGO soil mapunits are stored in vector format (WKB geometry) in the `mupolygon` table within the `ssurgo_2019` database schema. A unique identifier associated with each soil mapunit polygon (`mukey`) is used join to the `mapunit` and `component` tables. Soil components are joined to the `chorizon` table containing physical soil properties associated with each soil component through varying depth horizons.

The `mupolygon` layer is intersected with the `CLU` polygons to identify (and clip) all subfield soil mapunits located within each `CLU` boundary. The intersection is performed annually with the against the updated SSURGO data using the `clu_ssurgo_intersection.sql` query in the Foresite repository. Results of the SSURGO 2019 intersection for each `CLU` including the `CLU`-clipped soil mapunit polygons are stored in the `clu.clu_mupoly19` table.