Long-term Risk Projections Data Archive [Phase I]

Project ID#: CEC EPC-18-026

May 25, 2023

Prepared for the California Energy Commission

Primary Author(s):
Leroy Westerling and Jonathan Baldwin, University of California, Merced

Contract Number: EPC-18-026

Prepared for: California Energy Commission

David Saah, PhD
Principal Investigator

Shane Romsos
Project Manager

DISCLAIMER

TABLE OF CONTENTS

INTRODUCTION

The purpose of the *Long-term Fire Risk Projections Data Archive* (hereafter, "archive") is to identify/document linkages/locations to a variety of data repositories used for disseminating data used in model development, model inputs, and related model outputs across multiple labs representing the Pyregence Long-term Risk Projections workgroup (Workgroup #4). Also identified are repositories where model code is or will be archived. This is a living document that will be updated as new data outputs are released or when data repository locations are updated – updates will be represented in the Task 10 "*Long-term Risk Projections Data Archive - Phase II*" product deliverable.

DATA USED FOR MODEL DEVELOPMENT

Data used in model development include (but are not limited to):

- Monitoring Trend in Burn Severity (MTBS) - for historical fire data between 1984 and 2020. Includes fire perimeters and estimated burn severity. At 30m grid, Albers projection. Data can be accessed here: https://www.mtbs.gov/direct-download
- LANDFIRE – vegetation classification, elevation, aspect, slope. At 30m grid, Albers projection. Data can be accessed here - https://www.landfire.gov/getdata.php
- gridMET – gridded historical climatological data - https://www.climatologylab.org/gridmet.html.
- UCLA Center for Climate Science CMIP6 Downscaled data using WRF – gridded climate data, historical and projected (four GCMs). Available through AWS via Cal Adapt Analytics Engine - https://analytics.cal-adapt.org/data/.
- Scripps Institution of Oceanography gridded climate data products. Includes bias corrected versions of UCLA data, and various additional LOCA-2 runs of additional GCM/SSP pairs. Data currently maintained by UCSD Scripps. Contact David Pierce (dpierce@ucsd.edu) for access. Data will eventually be hosted by the Cal Adapt Analytics Engine - https://analytics.cal-adapt.org/data/.
- A variety of derived land surface characteristics datasets (such as: biomass, housing density, structure footprints, soil moisture, landcover type, forest type, protected areas, ecoregions). See the Data Directory (below) for more information on these data.

Additional data are anticipated to be included in later iterations of model development and will be added as they become available or are finalized, such as:

- Revised MTBS data - Westerling's lab has generated an experimental dataset recreating fire severity maps for all large fires in California since 1984[1], testing algorithms that provide a consistent approach to categorizing fire severity across the different vegetation types contained within the state. This data will be added to the dataset for dissemination to modeling teams after testing and formatting are complete.
- Reference climate history dataset that includes hourly temperature, precipitation, wind speed and direction, and relative humidity on a 3 km grid. A preliminary sample dataset is being developed by Scripps Institution of Oceanography but does not cover an adequate historical

[1] MTBS data merge information from documented fire records with satellite imagery-derived severity mapping. Processed MTBS data are currently available for 1984-2020.

period to support our modeling efforts. Once this dataset or similar is completed—covering an adequate time period and having been tested and validated—it will be added to the Workgroup #4 data archive.

- Downscaled climate projections will be formatted and included in the data archive as they become available.
- Finalized modeled historical biomass, historical and projected land use and development scenarios, fuels management scenarios.

Scripts have been developed to facilitate merging datasets from diverse sources into common geospatial data formats and vector data formats for use in modeling efforts across all teams in the working group.

LONG-TERM WILDFIRE RISK MODEL DATA

The data listed below in the Data Dictionary section are input data used for various models. Most of the data is available publicly through websites listed herein. Each website has its own usage and referencing requirements, along with metadata. See each one for more specific information about the data provided. The fire model output will initially be made available at https://data.pyregence.org/wg4/. The final data archive is planned to be available through the Cal-Adapt Analytics Engine (https://analytics.cal-adapt.org/). This will provide all the fire output data to be available in the same location as the climate input data.

The current Pyregence based output archive has the following data available:

- Wildfire Burn Severity and Emissions Inventory (WBSE) – This is the revised MTBS burn severity data (published in 2022 by Xu et al). The data can be found at https://data.pyregence.org/wg4/WBSE/
- See also References Section (below) in this document for Pyregence Workgroup 4 publications. Each publication cites (or will cite) URLs for data access.

CODE FOR MODELS

Model code for Workgroup 4 continues to be under development and/or internal review, with most code not ready to be posted for public access. When code has been finalized, most code will be made available via the Pyregence GitHub code repository (https://github.com/pyregence; an Internet hosting service for software development and version control). Code developed by USGS for LUCAS modelling effort will be hosted by USGS for public access (contact Ben Sleeter - bsleeter@usgs.gov, and Todd Hawbaker - tjhawbaker@usgs.gov).

DATA DICTIONARY

1) Monitoring Trends in Burn Severity (MTBS)
    a. Root Directory: https://www.mtbs.gov/
    b. Description: MTBS maps fire information for fires greater than a specified size (>1000 acres for the Western US). Fires are available for 1984-2017. This information includes (but is not limited to) metadata about each fire (geospatial as well as thresholds used to determine severity). There are also raster (GeoTiff) files of pre- and/or post-fire Landsat images, an NBR or dNBR image (depending on one or two scenes used), a burn severity map (values 0-6), and a mask image of the fire area with clouds, shadows, snow, etc. masked. The final product we use is the ESRI Shapefile of the fire perimeter. Each of

these products uses a 30m pixel. For full product descriptions see https://www.mtbs.gov/product-descriptions.

c. Directory Structure: Each fire is designated by a unique ID. Once a selection of fires is chosen (on the MTBS website). The downloaded files are unzipped into <year>/fire_level_tar_files/<fireID>.zip. Each <fireID>.zip file contains the products described above. Please see https://www.mtbs.gov/product-descriptions for a complete list of the contents and naming conventions of this zip file.

2) LANDFIRE
   a. Root Directory: https://www.landfire.gov
   b. Description: The LandFire data contains many data products (https://www.landfire.gov/version_comparison.php). The topographic and existing vegetation type data are of use to this project. Topographic information includes elevation, slope, and aspect. All are downloaded as zip files. Each zip file contains (among other data) an ArcGIS raster pyramid of specific topographic information. The existing vegetation type data is also downloaded as a zip file. This contains the metadata for the pixel values (as a lookup table in CSV format) as well as the ArcGIS raster pyramid for the vegetation type. Each pixel is assigned a specific value that can be looked up to see all the categories for that pixel (including fuels mapping value, lifeform, physiognomy, etc.). Please see https://www.landfire.gov/evt.php for a full listing of the data dictionary.
   c. Directory Structure: Each product is downloaded as a zip file. Each zip contains metadata in directories called CSV_Data, General_Metadata, and Spatial_Metadata. The gridded raster pyramid can be found in the Grid directory.
3) GridMet
   a. Root Directory: http://www.climatologylab.org/gridmet.html
   b. Description: GridMET provides daily historical climate data for the continental US from 1979-present. The data is gridded across a 1/24-degree grid. There are two types of climate data included in this dataset: primary variables (min/max temperature, precipitation, wind velocity, etc.) and derived variables (reference evapotranspiration, 100- and 1000-hour dead fuel moisture, etc.).
   c. Directory Structure: Each variable for a given year is stored in its own netCDF file. The naming convention is <var>_<year>.nc. The files are downloaded individually or using a batch generated wget script from the gridMET website.
4) UCLA Center for Climate Science
   a. Root Directory: https://aws.amazon.com/marketplace/pp/prodview-g4wqgpy2pa5dk?sr=0-2&ref_=beagle&applicationId=AWSMPContessa#overview. Also available through the Cal-Adapt Analytics Engine: https://analytics.cal-adapt.org/data/.
   b. Description: The UCLA climate data uses the WRF grid (approximately 3km). Data is hourly and includes many climate variables.
   c. Directory Structure: Each time period (hour) is contained in its own netCDF file with all variables available for that time period. The data is sorted by GCM/SSP and by year.
5) Scripps Institution of Oceanography
   a. Available through the Cal-Adapt Analytics Engine data catalog: https://analytics.cal-adapt.org/data/.

b. Description: The Scripps climate data provides 1/32-degree gridded climate data. Various variables are available (some GCMs do not provide some variables) including temperature, precipitation, wind data, and relative humidity. This data is at a daily cadence in both historical and projection periods.
c. Data Structure: The data is available through the Analytics Engine data catalog by GCM, SSP (including historical/common period), and variable. The files are netCDF files grouped in chunks of years.

6) Land Surface Hydrology Research Group – Drought Monitors
    a. Root Directory: http://www.hydro.ucla.edu/SurfaceWaterGroup/data.php.
    b. Description: The drought monitoring system provides modeled soil moisture estimates (three levels) as well as precipitation and air temperature. The historical data is available for 1920-2015. The grid is 1/16 degree.
    c. Directory Structure: The data is stored by decade with one netCDF file per month called fluxes.<year>-<month>.nc. All variables are stored within the netCDF file for each pixel in the grid.
7) Land Change Monitoring, Assessment, and Projection (LCMAP)
    a. Root Directory: https://www.usgs.gov/special-topics/lcmap/lcmap-data-access
    b. Description: Data represents a new generation of land cover mapping and change monitoring from the U.S. Geological Survey's Earth Resources Observation and Science (EROS) Center. LCMAP Collection 1.3 conterminous United States (CONUS) products, containing 10 annual science products for 1985-2021, were released in August 2022. See also - https://www.usgs.gov/media/files/lcmap-collection-13-science-product-guide
    c. Directory Structure: LCMAP Collection 1.3 Science Products are provided in .tar archive files containing 10 individual raster files for each year (1985-2021). Archive files are referenced using the same H-V indexing system as the CONUS ARD Grid (Figure 1-1). Products are packaged and delivered in a single .tar file. The .tar packages "untar" (unzip) into 10 individual Georeferenced Tagged Image File Format (GeoTIFF) (.tif) raster files for each Product Year and an Extensible Markup Language (XML) (.xml) metadata file.
8) LUCAS Biomass
    a. Root Directory: currently only by request (not publicly available as this is preliminary data; contact Ben Sleeter at bsleeter@usgs.gov).
    b. Description: The biomass data includes various categories (branch, stem, fine, etc.). It also has dead organic matter broken down into additional categories (aboveground/belowground, very fast/fast/medium/slow, etc.). The data is stored as a raster (GeoTiff) with a 1km resolution covering the continental US.
    c. Directory Structure: Each variable has its own GeoTiff file, named with the variable name.
9) Ecoregions
    a. Root Directory: https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states.
    b. Description: This data includes level III and level IV ecoregions across the continental US. The data are shapefiles and available with and without state boundaries.

c. Directory Structure: The data are downloaded in a single zip file per level which contains the ESRI shapefile components. The metadata and symbology are downloaded separately.

10) Protected Areas Database
    a. Root Directory: https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/science/protected-areas.
    b. Description: This is a listing of protected areas and public lands within the continental US. The data are available as shapefiles covering the continental US.
    c. Directory Structure: The downloaded national data is a single zip file containing one shapefile set for each of the groups (easement, fee, etc.)
11) National Forest Type Dataset
    a. Root Directory: https://data.fs.usda.gov/geodata/rastergateway/forest_type/
    b. Description: This dataset visualizes the distribution of forest types across the continental US in 28 groups. The data is downloaded as a single zip file containing the data and metadata. The raster data is at 250m per pixel across the continental US.
    c. Directory Structure: The zip file contains raster data and metadata for the forest group.
12) National Land Cover Database
    a. Root Directory: http://www.mrlc.gov/
    b. Description: This dataset describes the categorical land cover for the continental US. This data is available for various years from 2001 to 2016. The data is presented as raster data with a 30m pixel resolution.
    c. Directory Structure: The downloaded zip file contains the raster data as well as metadata.
13) Housing Density and WUI
    a. Root Directory: http://silvis.forest.wisc.edu/data/wui-change/
    b. Description: The housing density can be extracted from the Wildland Urban Interface data provided with this dataset. The data is available for the conterminous US as a geodatabase or by state as a geodatabase or shapefile.
    c. Directory Structure: Each downloaded zip file contains the requested geodatabase or shapefile along with associated metadata.
14) Building Footprints
    a. Root Directory: https://github.com/Microsoft/USBuildingFootprints
    b. Description: The GIS data layer of building/structure footprints for California. The data is available for the conterminous US as a geodatabase or by state as a geodatabase or shapefile.
    c. Directory Structure: Each downloaded zip file contains the requested geodatabase or shapefile along with associated metadata.
15) Electricity Assets – Transmissions Lines
    a. Root Directory: https://cecgis-caenergy.opendata.arcgis.com/datasets/260b4513acdb4a3a8e4d64e69fc84fee_0/data
    b. Description: The California Energy Commission (CEC) Electric Transmission Line geospatial data layer has been created to display the electric transmission grid in California. When used in association with the CEC Power Plant and CEC Electric

Substation geospatial data layers, viewers can analyze the geographic relationships with the electric transmission grids across utilities, counties and state.

c. Directory Structure: data can be downloaded as shapefile, kml or spreadsheet or accessed via API (https://services3.arcgis.com/bWPjFyq029ChCGur/arcgis/rest/services/Transmission_Line/FeatureServer/0/query?where=1%3D1&outFields=*&outSR=4326&f=json).

16) Electricity Assets – Substations
    a. Root Directory: https://cecgis-caenergy.opendata.arcgis.com/datasets/california-electric-substation/data.
    b. Description: The California Energy Commission (CEC) Electric Substation geospatial point data layer has been created to display the locations of power substations in California.
    c. Directory Structure: data can be downloaded as shapefile, kml or spreadsheet or accessed via API (https://services3.arcgis.com/bWPjFyq029ChCGur/arcgis/rest/services/Substation/FeatureServer/0/query?outFields=*&where=1%3D1).
17) National Transportation Dataset (NTD)
    a. Root Directory: https://www.sciencebase.gov/catalog/item/4f70b1f4e4b058caae3f8e16
    b. Description: Transportation data consists of roads, railroads, trails, airports, and other features associated with the transport of people or commerce. The data include the name or route designator, classification, and location.
    c. Directory Structure: Vector data (shapefile and gdb). Also see Metadata.
18) WBSE - Wildfire Burn Severity and Emissions Inventory
    a. Root Directory: https://data.pyregence.org/wg4/WBSE/.
    b. Description: The WBSE data set is a revised version of the MTBS burn severity maps. It covers the 1984-2020 period using a combination of MTBS fire perimeters and CalFire fire perimeters. A later (currently unpublished) version uses just the MTBS fire perimeters.
    c. Directory Structure: The burn severity maps are found in the burn_severity_CA subdirectory. Each fire has a single GeoTIFF file named with the fire ID (MTBS or one constructed from the CalFire
19) TREEMAP
    a. Root Directory: https://doi.org/10.2737/RDS-2021-0074
    b. Description: TreeMap 2016 provides a tree-level model of the forests of the conterminous United States that combines a random forest imputation approach with forest plot data from Forest Inventory and Analysis (FIA) to create a wall-to-wall 30x30 meter (m) grid. Data used in Landis II modeling.
    c. Directory Structure: The data are downloaded as part of one large zipfile which contains a CONUS wide raster plus attribute table and readme file.
20) Aerial Detection Survey
    a. Root Directory: https://www.fs.usda.gov/foresthealth/applied-sciences/mapping-reporting/detection-surveys.shtml
    b. Description: The aerial detection surveys are the primary method of collecting data on the health of treed areas affected by insects and diseases, with data being collected during flight using digital aerial sketch mapping. Data used in Landis II modeling.

c. Directory Structure: Polygon data covering each Forest Service subregion. The data are downloaded in a single zip file per region per year which contains the ESRI shapefile components.

21) Spatial wildfire occurrence data for the United States
    a. Root Directory: https://doi.org/10.2737/RDS-2013-0009.6
    b. Description: These data contain a spatial database of point locations of wildfires that occurred in the United States from 1992 to 2020. The wildfire records were acquired from the reporting systems of federal, state, and local fire organizations and contain information about discovery and suppression dates, fire size, and ignition source. Data used in Landis II modeling.
    c. Directory Structure: Point data covering the United States with a data download in a geodatabase/sql database format.

REFERENCES

Abatzoglou, J. T. (2013), Development of gridded surface meteorological data for ecological applications and modelling. Int. J. Climatol., 33: 121–131.

Sam, J.A., Baldwin, W.J., Westerling, A.L., Preisler, H.K., Xu, Q., Hurteau, M.D., Sleeter, B.M. and Thapa, S.B., 2022. Simulating burn severity maps at 30 meters in two forested regions in California. *Environmental Research Letters*, *17*(10), p.105004.

Thapa, S.B.; Sam, J.A.; Baldwin, W.J.; Sleeter, B.M.; Xu, Q.; Pierce, D; Westerling, A.L.;. Simulating burn severity maps in Northern California's Wildland-Urban Interface. (Manuscript ready for submission)

Xu, Q., Westerling, A.L., Notohamiprodjo, A., Wiedinmyer, C., Picotte, J.J., Parks, S.A., Hurteau, M.D., Marlier, M.E., Kolden, C.A., Sam, J.A. and Baldwin, W.J., 2022. Wildfire burn severity and emissions inventory: an example implementation over California. *Environmental Research Letters*, *17*(8), p.085008.

Xu, Q., A. L. Westerling, A. and W. J. Baldwin. 2022. Spatial and temporal patterns of wildfire burn severity and biomass burning-induced emissions in California. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/ac9704.

Sam, J.A.; Preisler, H.K.; Westerling, A.L.; Xu, Q.; Baldwin, W.J.; Sleeter, B.M;. Fire severity fraction simulations for California's fifth climate assessment: Identifying significant drivers of potential changes in burn severity. (Manuscript ready for submission)