**Young, Adam M. 2016. Data and code from Young et al. 2016. Climatic thresholds shape northern high-latitude fire regimes and imply vulnerability to future climate change. Ecography (doi: …). Dryad:**

**Overview:**

This archive includes geographic information, datasets and code needed to reproduce the results presented in Young et al. (2016). The archive includes 55.1 GB of data, including input and output files for the Boosted Regression Tree analysis. After extracting the .zip archive (‘Young\_et\_al\_2016’), there are three folders: **Data**, **Output**, and **Scripts**. The **Scripts** folder contains all the individual Matlab and R scripts, including functions, needed to conduct the analyses and produce the results presented in the manuscript. The **Data** folder contains six subfolders, containing a mix of GeoTiff (.tif), Excel (.xlsx), Comma Separated Value (.csv), R Data (.RData), and Matlab (.mat) file formats. The **Output** folder contains results from the BRT analysis, including figures. Figures are simultaneously available in 450 dpi .tif and .fig (matlab) formats.

**Contact:** Philip E. Higuera, phiguera[at]umontana.edu

**Contents:**

1. Geographic information for Alaska

2. Scripts

3. Data

4. Output

**1. Geographic information for all gridded maps of Alaska:**

Rows: 725

Columns: 687

Spatial Extent (in meters) -

- Top: 2390439.786

- Left: -656204.44

- Right: 717795.56

- Bottom: 940439.786

Spatial Reference: Albers Equal Area

Datum: North American 1983

False Easting: 0

False Northing: 0

Central Meridian: -154

Standard Parallel 1: 55

Standard Parallel 2: 65

Latitude of Origin: 50

**2. Scripts:**

There are two subsets of files in the **Scripts** folder: (1) Sixteen R (.R) and Matlab (.m) files sequentially numbered to reproduce the results in Young et al. 2016, and (2) a folder containing custom written functions that support the sixteen scripts. If the scripts are run in the same order as they are numbered then the exact results and figures from the main text will be reproduced.

**Analysis scripts:**

* **1\_create\_hist\_ak\_climatologies.m** – used to create random 30-yr climatologies from monthly historical data (1950-2009).
* **2\_create\_future\_ak\_climatologies.m** – used to create 30-year climatologies of projected climate for five GCMs and three different time periods during the 21st-century.
* **3\_AK\_BRTS.R** – Script that uses climatologies created in ‘1\_create\_hist\_ak\_climatologies.m’, vegetation data, and topographic ruggedness data in conjunction with the gbm package (v2.1.1) to create the set of 100 BRTs that comprise the AK model.
* **4\_BOREAL\_BRTS.R** – Script that uses climatologies created in ‘1\_create\_hist\_ak\_climatologies.m’, vegetation data, and topographic ruggedness data in conjunction with the gbm package (v2.1.1) to create the set of 100 BRTs that comprise the BOREAL model.
* **5\_TUNDRA\_BRTS.R** – Script that uses climatologies created in ‘1\_create\_hist\_ak\_climatologies.m’, vegetation data, and topographic ruggedness data in conjunction with the gbm package (v2.1.1) to create the set of 100 BRTs that comprise the TUNDRA model.
* **6\_Identify\_Climatic\_Thresholds.R** – Script that uses the segmented package (v0.5-1.4) to identify climatic thresholds from BRT partial dependence results. Partial dependence results were exported as .csv files in the three BRT scripts (i.e., #3-5).
* **7\_project\_21stCentury\_fire.R** – Script that uses the BRTs that comprise the AK model (created in ‘3\_AK\_BRTS.R’) to project the 30-yr probability of fire occurrence per pixel for Alaska in the 21st-century. Future climatologies were created using the ‘2\_create\_future\_ak\_climatologies.m’ script.
* **8\_corr\_results.m** – Main purpose is to calculate the Pearson correlation between predicted and observed fire rotation period (FRP) estimates for Alaskan ecoregions. FRP estimates were exported in the BRT scripts (i.e., #3-5).
* **9\_summarize\_future\_projections.m –** Takes BRT-generated future projections of the 30-yr probability of fire occurrence and calculates (1) the median predicted 30-yr probability of fire occurrence of all 100 BRT projections for all five GCMs and for three time periods, and (2) the median ratio between future and historical FRPs for all GCMs and time periods.
* **10\_calc\_proportions.m** – Calculate the percentage of pixels that occurs in eight different discrete classes of relative change in the FRP (i.e., ratio).
* **11\_FIG\_2.m** – Creates Figure 2 in Young et al. (2016).
* **12\_FIG\_3.m** – Creates Figure 3 in Young et al. (2016).
* **13\_FIG\_4.m** – Creates Figure 4 in Young et al. (2016).
* **14\_FIG\_5.m** – Creates Figure 5 in Young et al. (2016).
* **15\_FIG\_6.m** – Creates Figure 6 in Young et al. (2016).
* **16\_FIG\_7.m** – Creates Figure 7 in Young et al. (2016).

**Functions:**

* **akaxes.m** – plot background map of Alaska.
* **aucroc.R** – Calculates AUC value and classification rates.
* **barwitherr.m** – Creates bar graph with error bars - used in Fig. 3. Created by Martina Callaghan and obtained from Matlab file exchange (<http://www.mathworks.com/matlabcentral/fileexchange/30639-barwitherr-errors-varargin->).
* **cbfreeze.m** – Freezes colors on colorbar. Created by Carlos Adrian Vargas Aguilera and obtained from Matlab file exchange (<http://www.mathworks.com/matlabcentral/fileexchange/24371-colormap-and-colorbar-utilities--jul-2014-/content/cm_and_cb_utilities/cbfreeze.m>).
* **createClimatologies.m** – Function that takes gridded GeoTiff files at a monthly or annual resolution and calculates climatological averages per pixel for a user-defined time period.
* **freezeColors.m** – Freezes colors on current plot. Created by John Iversen and obtained from Matlab file exchange (<http://www.mathworks.com/matlabcentral/fileexchange/7943-freezecolors---unfreezecolors>).

**3. Data:**

The **Data** directory contains six folders, two of which are created during the workflow of the analysis.

* **AncillaryData** – This folder contains files and data that support running the analysis and making the figures in Young et al. (2016), and contains two subsets of data files. First, the folder (‘Shapefiles’) contains two shape files: (1) the outline of the 22 Alaskan ecoregions used in this study (‘ecor\_outlines\_for\_fig1.shp’), and (2) a general coastline map of Alaska the excludes islands and southeast Alaska (‘ak\_shape\_noislands\_noSE.shp’). Second, there is a collection of 11 supporting files:
* masks.mat – 725 x 687 x 3 matlab array. Contains the spatial masks for each model (AK, BOREAL, and TUNDRA). The first two dimensions are the rows and columns of the gridded Alaska map and the third dimension contains information on whether each 2-km pixel is considered within each of the three spatial domains (NaN = not in study area, 1 = in study area).
* test\_years.mat – 100 x 30 matrix of randomly selected years to create testing 30-yr climatologies for climate and fire.
* train\_years.mat – 100 x 30 matrix of randomly selected years to create training 30-yr climatologies for climate and fire.
* ecor\_info.csv and ecor\_info.xlsx – Classification information for each ecoregion at Level I (boreal forest or tundra) and Level III (specific ecoregion). Columns in these files are representative of : ecoregion name, original (Nowacki et al. 2001) Level III classification value, Level I classification (1 = Tundra, 2 = boreal forest), modified Level III classification, Latitude of approximate center of ecoregion, Longitude of approximate center of ecoregion.
* AnnDEF\_1950\_2009.tif – 1950 – 2009 mean total annual moisture availability per pixel (mm).
* Fire\_occ\_1950\_2009.tif – Fire occurrence for each pixel from 1950-2009 (-9999 = NaN/no fire, 1 = fire occurrence).
* Lat.tif – Latitude for the center of each 2-km pixel in Alaska gridded map. Used for plotting.
* Lon.tif – Longitude for the center of each 2-km pixel in Alaska gridded map. Used for plotting.
* obsfrp.tif – observed 1950-2009 fire rotation period (FRP) for each ecoregion in Alaska (-9999 = NaN, -1 = no FRP estimate)
* TempWarm\_1950\_2009.tif – 1950 – 2009 mean temperature of the warmest month per pixel (°C).
* **ClimateData** – These data are gridded maps of Alaska containing historical and future downscaled climate data. Specifically, there are maps of monthly (1) temperature and (2) total precipitation minus potential evapotranspiration (P-PET). Here, in our data files, we re-label 'P-PET' as 'DEF' for 'deficit'. For both historical and future data the months correspond to a "fire year", or a year begining on October 1st and ending on September 30th. Due to this, data for the months of October, November, and December are the values recorded for previous year. For example, P-PET data for November 2002, is the estimate of P-PET for November 2001.
* Historical Example:

tas\_mean\_C\_cru\_TS31\_01\_1950\_FY.tif

tas - climate variable, tas = surface temp, DEF = P-PET

\_mean\_C - units, mean\_C = degree C,

tot\_mm = millimeters

\_cru\_TS31 - Based off of CRU TS climate data

(http://www.cru.uea.ac.uk/cru/data/hrg/)

See https://www.snap.uaf.edu/ for details.

\_mm\_yyyy - month and year

\_FY - For fire the fire year (October 1st - September 30th) See above.

.tif - File extension

* Future Example: tas\_FY\_mean\_C\_AR5\_CCSM4\_rcp60\_01\_2007.tif

tas - climate variable, tas = surface temp, DEF = P-PET

\_FY – For fire the fire year (October 1st - September 30th) See above.

\_mean\_C - units, C = degree celsius, mm = millimeters

\_AR5 - From IPCC Fifth Assessment Report

\_CCSM4 - GCM

\_rcp60 - Climate change scenario

\_mm\_yyyy - month and year

.tif - File extension

* **FireData** – These data are comprised of 60 individual GeoTIFF files that indicate whether fire is estimated to haved occurred in a given pixel for a given year. Each map is representative of one calendar year. These data are gridded versions of the vector geographic data provided by the Alaska Interagency Coordination Center (AICC; <http://fire.ak.blm.gov/>). Converting the polygon (i.e., vector) datasets to raster was done using ESRI software in ArcMap 10.1. Specifically, we used the function "Feature to Raster" under the Conversion Toolbox. A given pixel was classified as "fire occurrence" if the center of the pixel overlapped with a fire polygon from the AICC dataset. Filenames have the following format:

Example:

fire\_1950.tif

fire – fire occurrence data

\_xxxx – year (1950-2009)

.tif –file extension

* **FutureClimatologies** – This folder contains 30-yr climatologies for each climate variable (TempWarm and AnnDEF) for the three different time periods in the 21st century and for the five GCMs. These folders and their contents are created by the ‘2\_create\_future\_ak\_climatologies.m’ script.
* **train\_test\_data** – This folder contains all of the fire and climate data created by randomly selecting 30 years from 1950-2009. These are the data that are used to train and test the BRTs in ‘3\_AK\_BRTS.R’, ‘4\_BOREAL\_BRTS.R’, and ‘5\_TUNDRA\_BRTS.R’.
* **VegLandscapeData** – This folder contains three files: ‘AK\_VEG.tif’, ‘TR.tif’, and ‘ecor.tif’. The 'AK\_VEG.tif' file contains vegetation classifications for each pixel in Alaska. There are five different vegetation types. Values 1-5 represent the following vegetation classifications:

1 - Wetland Tundra

2 - Shrub Tundra

1. - Graminoid Tundra
2. - Barren Tundra
3. - Boreal Forest

Methods to use these maps to create the 'AK\_VEG.tif' file are described in Young et al. (2016).

The 'ecor.tif' file is a map that classifies pixels in Alaska by ecoregion. The original ecoregion map is from (Nowacki et al. 2001). The main modification to the original ecoregions map was the addition of the Noatak River Watershed. The spatial coverage of the Noatak River Watershed was obtained from the perimeter of the Noatak National Preserve. The original ecoregions map of Alaska is in vector data format (i.e., polygons). We converted these data to a raster format using the "Feature to Raster" conversion tool in the "Conversion Toolbox" in ESRI Software ArcMap 10.1 Ecoregion identifications can be found in Fig. 1 of Young et al. (2016).

The 'TR.tif' file is a map of topographic ruggeddness measured in meters. Methods used to create the topographic ruggedness are described in Young et al. (2016).

**4. Output:**

The **Output** directory contains 5 folders created through the workflow of the analysis.

* **1\_AK\_FINAL\_RESULTS** – This folder contains output for the AK model generated by ‘3\_AK\_BRTS.R’, ‘6\_Identify\_Climatic\_Thresholds.R’, and ‘8\_corr\_results.m’ scripts:
  + brt\_1.RData, brt\_2.RData, …, brt\_100.RData – raw results/output from the 100 ‘gbm’ objects created using the gbm package in R.
  + pred\_map\_1.tif, pred\_map\_2.tif, …, pred\_map\_100.tif - These data are gridded maps of Alaska containing the predicted 30-yr probabiltiy of fire occurrence for each pixel during the historical period. Predictions are made using the gbm models on a holdout sample of the data not used to train the gbms.
  + AUC.csv – AUC results from ROC analysis of 100 BRT models
  + class\_rates.csv – classification rates from ROC analysis (TPR: true positive rate, TNR: True negative rate, FPR: False positive rate, FNR: false negative rate). Probability thresholds for classifications were determined using methods described in Young et al. (2016).
  + climThresholds\_AnnDEF.csv – climatic thresholds determined from ‘Identify\_Climatic\_Thresholds.R’ for annual moisture availability.
  + climThresholds\_TempWarm.csv – climatic thresholds determined from ‘Identify\_Climatic\_Thresholds.R’ for temperature of the warmest month.
  + frp\_o.csv – 22 x 100 matrix of observed 30-yr fire rotation periods for each ecoregion (rows) for each BRT iteration (columns).
  + frp\_p.csv – 22 x 100 matrix of predicted 30-yr fire rotation periods for each ecoregion (rows) for each BRT iteration (columns).
  + frp\_t.csv – Data from frp\_o.csv and frp\_p.csv available in a two-column matrix (observed and predicted).
  + model\_train\_err.csv – 5000 x 100 matrix of training error data for each BRT.
  + model\_valid\_err.csv – 5000 x 100 matrix of cross-validation error for each BRT.
  + nTrees.csv – Optimal number of regression trees determined by cross validation for each BRT model.
  + partDep\_AnnDEF.csv – Partial dependence results for annual moisture availability.
  + partDep\_TempWarm.csv – Partial dependence results for temperature of the warmest month.
  + partDep\_TR.csv – Partial dependence results for topographic ruggedness.
  + partDep\_Veg.csv – Partial dependence results for vegetation type.
  + relInf.csv – Relative influence results from BRT analysis for each explanatory variable.
  + TempWarm\_AnnDEF\_int.csv – Partial dependence results for the two-way interaction between temperature of the warmest month and annual moisture availability.
  + thresholds.csv – Probability threshold determined from maximizing the summation of the true positive rates (TPR) and true negative rates (TNR).
  + corresults.mat – Pearson correlation results between predicted and observed FRPs. Calculated using the ‘8\_corr\_results.m’ script.
  + med\_pred\_prob.mat – median predicted historical probability of fire occurrence per pixel. Created using the 100 predicted probability maps (i.e., pred\_map\_1.tif, pred\_map\_2.tif, …, pred\_map\_100.tif). Calculated using the ‘8\_corr\_results.m’ script.
* **2\_BOREAL\_FINAL\_RESULTS** – This folder contains the same output files as in the 1\_AK\_FINAL\_RESULTS folder, except for the BOREAL model.
* **3\_TUNDRA\_FINAL\_RESULTS** – This folder contains the same output files as in the 1\_AK\_FINAL\_RESULTS folder, except for the TUNDRA model.
* **4\_FUTURE\_PROJECTIONS** – This output folder contains five subfolders and three additional output data files. The five subfolders are labeled for each GCM used in our analysis (CCSM4, GFDL-CM3, GISS-E2-R, IPSL-CM5A-LR, MRI-CGCM3). Each of these subfolders contains 300 GeoTiff files, with each GeoTiff providing BRT-generated 30-yr projected probability of fire occurrence per pixel for three different future time periods (2010-2039, 2040-2069, 2070-2099). Filenames have the following format:

Example: CCSM4\_pred\_map\_2010\_2039\_1.tif

CCSM4 - GCM name

\_pred\_map - a predicted (or projected) probability map

\_2010\_2039 – time period for projection

\_1 – Specific BRT (1-100) used to generate projection

.tif – GeoTiff file extension

* The remaining three output files are:
  + med\_fut\_pred.mat - Median projected probability of fire occurrence per pixel from the distribution of 100 BRT-generated projections for each GCM and time period. This mat file is an array with dimension 725 x 687 x 5 x 3. The rows and columns (first two dimensions, respectively) are the number of rows and columns in the gridded map of Alaska. The third dimension (n = 5) is for each GCM and the fourth dimension (n = 3) is for each time period. Calculated in the ‘9\_summarize\_future\_projections.m’ script.
  + ratio.mat – Same size as ‘met\_fut\_pred.mat’, but this contains the median of the ratios of future projected and historical predicted fire rotation periods per pixel for each BRT. Calculated in the ‘9\_summarize\_future\_projections.m’ script.
  + prop\_vals.xlsx – Summary the proportion of pixels in Alaska that are in one of eight classes (or intervals) of ratio values. This is only summarized for the median of all five GCMs, the coolest GCM and the warmest GCM. Calculated in the script ‘10\_calc\_proportions.m’.
* **5\_FIGS** – This folder contains 14 files, seven .tif and seven .fig (matlab format) files. These files are labeled for the corresponding figure in the main text of Young et al. (2016).