

Ecography

ECOG-02205

Young, A. M., Higuera, P. E., Duffy, P. A. and Hu, F. S. 2016. Climatic thresholds shape northern high-latitude fire regimes and imply vulnerability to future climate change. – *Ecography* doi: 10.1111/ecog.02205

Supplementary material

Appendix 1. Supplementary Methods

Boosted regression tree modeling

To construct our boosted regression trees (BRTs), we used a Bernoulli distribution to characterize the binary response variable of fire presence/absence. We set our bagging fraction of 0.5, and used a 5-fold cross validation to identify the optimal number iterations from a maximum of 5000. Since we used a small pre-selected subset of 2-km pixels to train each BRT, we set the training fraction to 1.0. The minimum number of observations allowed in each regression tree node was set to 1. Our interaction depth was set at 2, to capture pairwise potential interactions among explanatory variables. The learning rate (i.e., shrinkage) was 0.01 for each model, and was chosen to ensure deviance of the predicted response reached a minimum within the maximum number of trees (i.e., 5000) (Supplementary material Appendix 2, Fig. A1).

Landscape sampling of Alaskan wildfire occurrence

To prevent overfitting of historical fire-climate relationships and to account for potential spatial autocorrelation in our analyses, we sampled a small subset of all available 2-km pixels from the spatial domain of each model (AK, BOREAL, and TUNDRA). The total number of pixels from each spatial domain (i.e., sampling rate) used to train our boosted regression tree models (BRTs) was determined by evaluating the predictive performance of BRTs built at different sampling rates. Specifically, we increased the sampling rate until subsequent increases resulted in model improvement $< 5\%$ for both the mean AUC and median Pearson correlation value. We tested eight different sampling rates, based on the 50th, 60th, 75th, 80th, 85th, 90th, 95th, and 99th percentiles of the size distribution of fires from 1950-2009. For example, in the BOREAL domain, the 85th percentile fire size is 121.6 km^2 , and the total study area is $125,470 \text{ km}^2$; therefore there are $125,470 \text{ km}^2 / 121.6 \text{ km}^2 \approx 1032 \text{ observations}$

available, corresponding to a sampling rate of $4127 \text{ pixels} / 125,470 \text{ pixels} = 3.29\%$ (Supplementary material Appendix 2, Table S3). This sampling rate is comparable to randomly selecting a single 2-km pixel every 122 km^2 . For each spatial domain and sampling rate, we trained a set of 15 BRTs, which was used to predict fire presence/absence. To evaluate model performance, we recorded AUC values and Pearson correlations between predicted and observed fire rotation periods (FRPs) for ecoregions. If needed, the number of iterations and the shrinkage parameters for the BRTs were adjusted for different sampling rates. Finally, we selected the lowest sampling rate that met our criteria for all three sampling domains. Based on our criteria, we used sampling rates associated with the 85th percentile of fire sizes, or 3.52%, 3.29%, and 5.43% of the available pixels for the AK, BOREAL, and TUNDRA models, respectively (Supplementary material Appendix 2, Table S3).

This sampling design has several advantages compared to using all available pixels in each sampling domain. First, our approach helps to guard against overfitting, as relationships derived from a subsample of points are likely more generalizable to new observations than relationships fit using all available pixels. Second, our design helps account for spatial autocorrelation, which is particularly important when modeling variability in fire occurrence, because the process of fire spread is highly autocorrelated in space. Using all available pixels runs the risk of over-estimating the predictive power of explanatory variables. Finally, using different subsampling rates for each spatial domain helps account for significant fire-regime differences among study domains (e.g., boreal forest fire are generally larger than tundra fires).

Identifying climatic thresholds to fire occurrence with segmented regression models

To quantify potential thresholds we used a piecewise linear regression using the “segmented” package in R (Muggeo 2003, Muggeo 2008). We used the median predicted probability of fire occurrence for each climate explanatory variable and restricted the segmented regression to climate values immediately surrounding a visually identified threshold. Specifically, we sampled BRT predictions ($n = 100$) with replacement 2000 times, calculated the median predicted probability from the 100 BRTs each time, performed piecewise regression on each sample, and recorded threshold estimates from each bootstrap sample. We report the mean threshold estimate and the 2.5th and 97.5th percentiles as 95% confidence intervals from the 2000 bootstrapped samples.

Appendix 2. Supplementary Results

Table A1. List of the 13 continuous candidate explanatory variables originally considered in the boosted regression tree (BRT) analysis and their predictive performance when each is used individually to model the 30-yr probability of fire occurrence. Values for each candidate climate variable are the 30-yr average. For each explanatory variable, 100 BRTs were constructed and used to predict the presence/absence of fire. AUC mean and SD values are calculated from the predictions of these 100 BRTs.

Variable	Units	Description	AUC Mean (SD)
GDD _{ANN}	°C	Total Growing Degree Days	0.74 (0.01)
P _{ANN}	mm	Total Annual Precipitation	0.64 (0.03)
P _{DJF}	mm	Total Winter Precipitation	0.60 (0.06)
P _{JJA}	mm	Total Summer Precipitation	0.65 (0.02)
P _{MAM}	mm	Total Spring Precipitation	0.62 (0.02)
P _{RANGE}	mm	Annual Precipitation Range	0.60 (0.03)
P-PET _{ANN}	mm	Total Annual Moisture Availability	0.70 (0.02)
P-PET _{JJA}	mm	Total Summer Moisture Availability	0.60 (0.02)
T _{ANN}	°C	Mean Annual Temperature	0.62 (0.02)
T _{JJA}	°C	Mean Summer Temperature	0.73 (0.01)
T _{RANGE}	°C	Annual Temperature Range	0.59 (0.02)
T _{WARM}	°C	Mean Temp. of the Warmest Month	0.78 (0.01)
TR	m	Topographic Ruggedness	0.57 (0.01)

Table A2. Median Spearman rank correlations among candidate explanatory variables.

Correlations were calculated using 5% of the data randomly sampled across space 100 times using 1950-2009 averages.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13
1 GDD _{ANN}	1.00												
2 P _{ANN}	0.06	1.00											
3 P _{DJF}	-0.04	0.93	1.00										
4 P _{JJA}	0.15	0.92	0.75	1.00									
5 P _{MAM}	0.00	0.96	0.93	0.83	1.00								
6 P _{RANGE}	0.06	0.96	0.83	0.95	0.89	1.00							
7 P-PET _{ANN}	-0.19	0.96	0.92	0.85	0.94	0.92	1.00						
8 P-PET _{JJA}	0.08	0.92	0.79	0.94	0.85	0.91	0.90	1.00					
9 T _{ANN}	0.69	0.54	0.44	0.56	0.51	0.50	0.40	0.64	1.00				
10 T _{JJA}	0.96	-0.09	-0.19	0.03	-0.15	-0.08	-0.34	-0.10	0.51	1.00			
11 T _{RANGE}	0.02	-0.71	-0.67	-0.64	-0.70	-0.67	-0.75	-0.82	-0.66	0.23	1.00		
12 T _{WARM}	0.91	-0.17	-0.26	-0.06	-0.22	-0.15	-0.42	-0.21	0.40	0.98	0.34	1.00	
13 TR	-0.22	0.37	0.31	0.43	0.32	0.38	0.38	0.34	-0.06	-0.22	-0.16	-0.24	1.00

Table A3. Results from evaluating models built at different sampling rates. AUC and Pearson correlation values are calculated from the predictions of the 15 boosted regression trees models trained for each spatial domain and sampling rate. An example of how the “Number of pixels” and “Sampling rate” values are calculated is given in Appendix 1.

Percentile	Model/ Spatial domain	Fire size (km ²)	Number of pixels	Sampling rate (%)	AUC	Median
					Mean (SD)	Pearson correlation
99	AK	1434.0	707	0.28	0.78 (0.03)	0.61
	BOREAL	1489.2	337	0.27	0.64 (0.05)	0.61
	TUNDRA	1070.2	381	0.37	0.58 (0.15)	0.38
95	AK	416.3	2435	0.96	0.78 (0.02)	0.64
	BOREAL	426.5	1177	0.94	0.63 (0.04)	0.64
	TUNDRA	348.0	1170	1.15	0.66 (0.09)	0.41
90	AK	199.0	5093	2.01	0.78 (0.02)	0.66
	BOREAL	205.4	2443	1.95	0.63 (0.03)	0.62
	TUNDRA	130.3	3126	3.07	0.70 (0.08)	0.52
85	AK	113.5	8930	3.52	0.78 (0.02)	0.75
	BOREAL	121.6	4127	3.29	0.63 (0.03)	0.63
	TUNDRA	73.7	5526	5.43	0.70 (0.08)	0.59
80	AK	75.1	13496	5.33	0.78 (0.02)	0.80
	BOREAL	79.7	6297	5.02	0.62 (0.03)	0.66
	TUNDRA	47.0	8665	8.51	0.72 (0.06)	0.57
75	AK	50.7	19991	7.89	0.78 (0.02)	0.81
	BOREAL	53.5	9381	7.48	0.62 (0.02)	0.70
	TUNDRA	34.6	11770	11.56	0.72 (0.06)	0.60
60	AK	18.8	53193	21.28	0.77 (0.01)	0.80
	BOREAL	19.6	25606	20.41	0.62 (0.03)	0.67
	TUNDRA	15.0	27510	26.67	0.72 (0.07)	0.55
50	AK	10.7	94725	37.38	0.77 (0.01)	0.82
	BOREAL	11.1	45214	36.04	0.62 (0.03)	0.68
	TUNDRA	8.0	50907	50.00	0.71 (0.06)	0.54

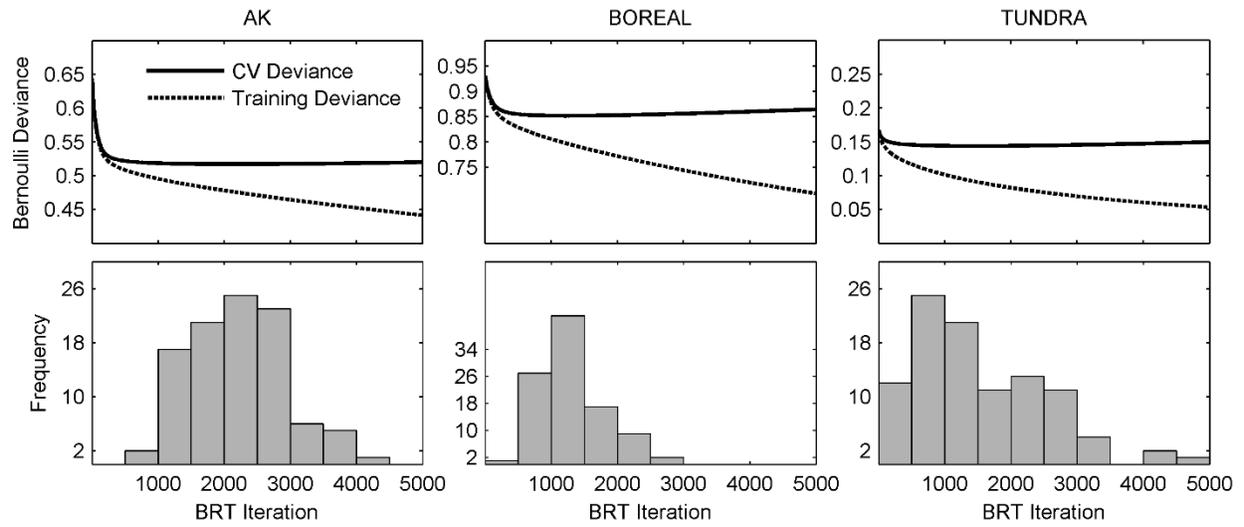


Fig. A1 Boosted regression tree (BRT) diagnostic information for each modeling domain. The top row displays the training and testing deviance averaged from the 100 BRTs for each of the three models. The bottom row displays the distribution of the optimal number of iterations selected for each of the 100 BRTs.

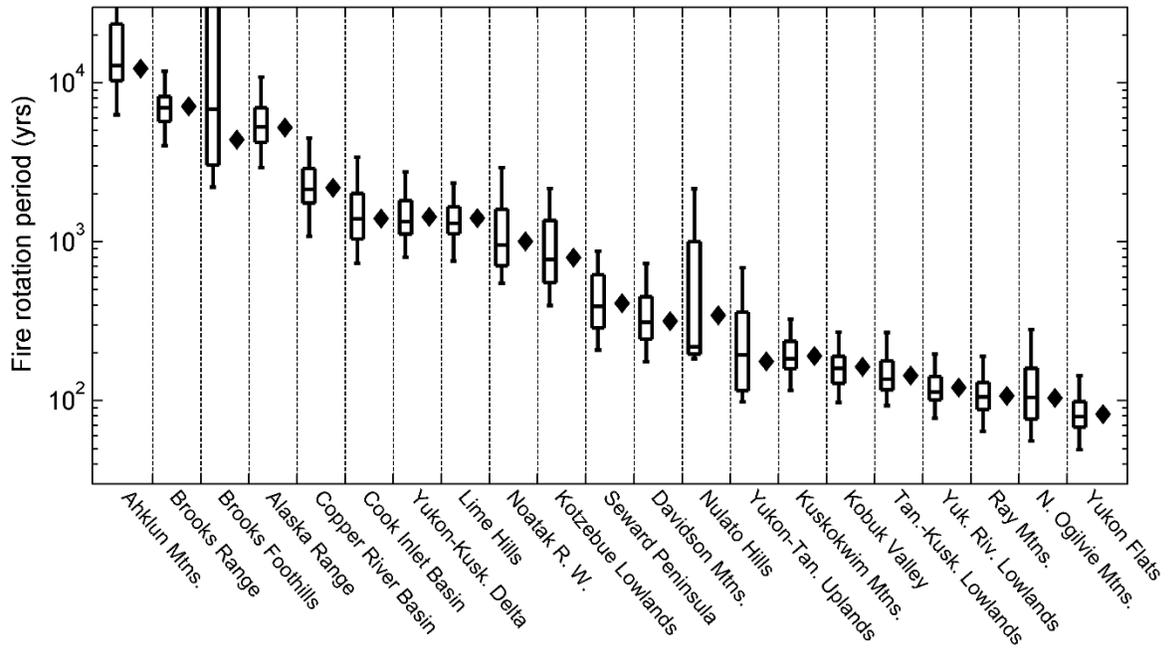


Fig. A2 Observed fire rotation periods (FRPs) for Alaskan ecoregions calculated using 30 non-continuous, randomly sampled years (boxplots) compared with the FRP of each ecoregion using all sixty years (1950-2009) of available data (black diamonds). Boxplots represent a distribution of 100 FRPs calculated using a randomly sampled 30-yr time period.

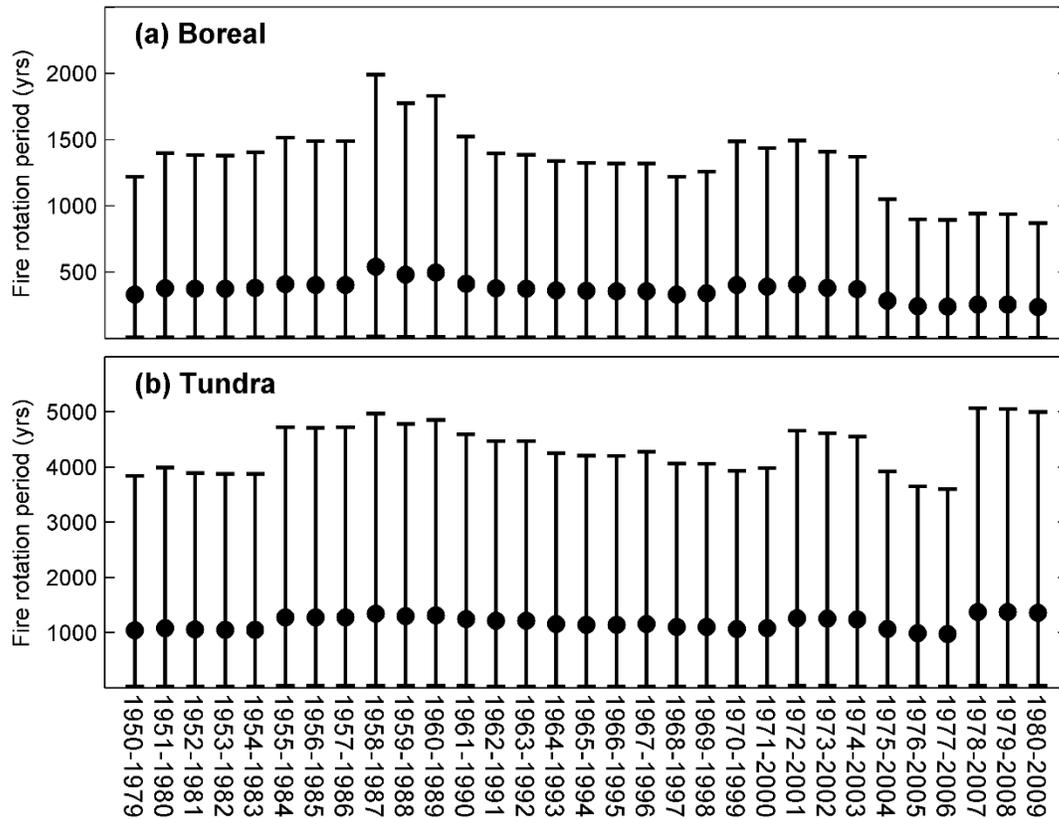


Fig. A3 Thirty-year fire rotation periods (FRPs) for continuous time periods from 1950 through 2009 for the (a) boreal forest and (b) tundra spatial domains. The solid point is the calculated FRP for each thirty year time period. Confidence bounds are the 2.5th and 97.5th quantiles assuming FRP is an exponential random variable with the rate parameter equal to the inverse FRP (e.g., Chipman et al. 2015).

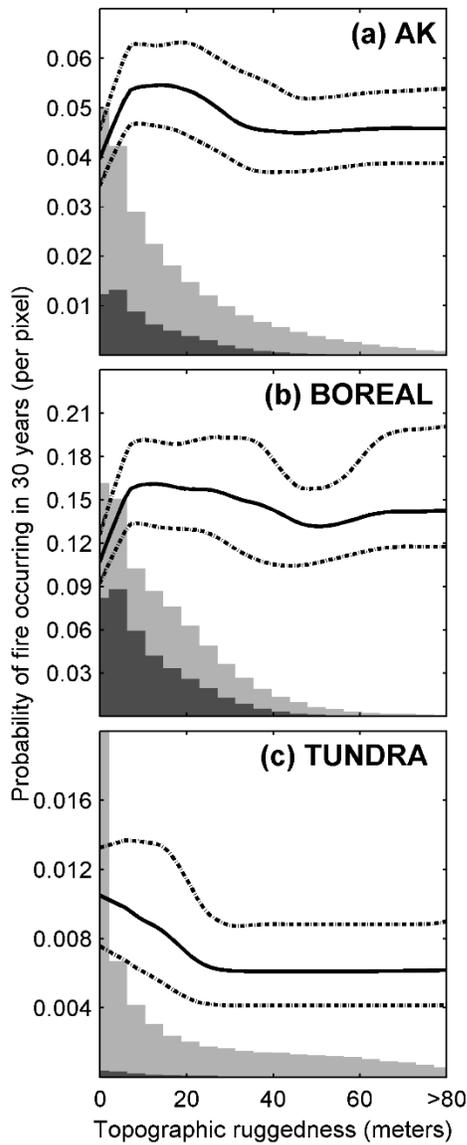


Fig. A4 Partial dependence plots illustrating the relationships between topographic ruggedness and the predicted probability of fire occurrence from the (a) AK, (b) BOREAL, and (c) TUNDRA models, as in Fig. 4 in the main text. The solid black lines represent the median predicted probability of fire occurrence, and the dashed lines represent the interquartile range from 100 boosted regression tree models. A lowess function (span = 0.1) was used to smooth the plotted predicted median and interquartile lines. As a reference, lighter (darker) colored histograms represent the historical distribution of topographic ruggedness among unburned

(burned) pixels from 1950 to 2009. Histograms heights were scaled individually and are not associated with y-axis values.

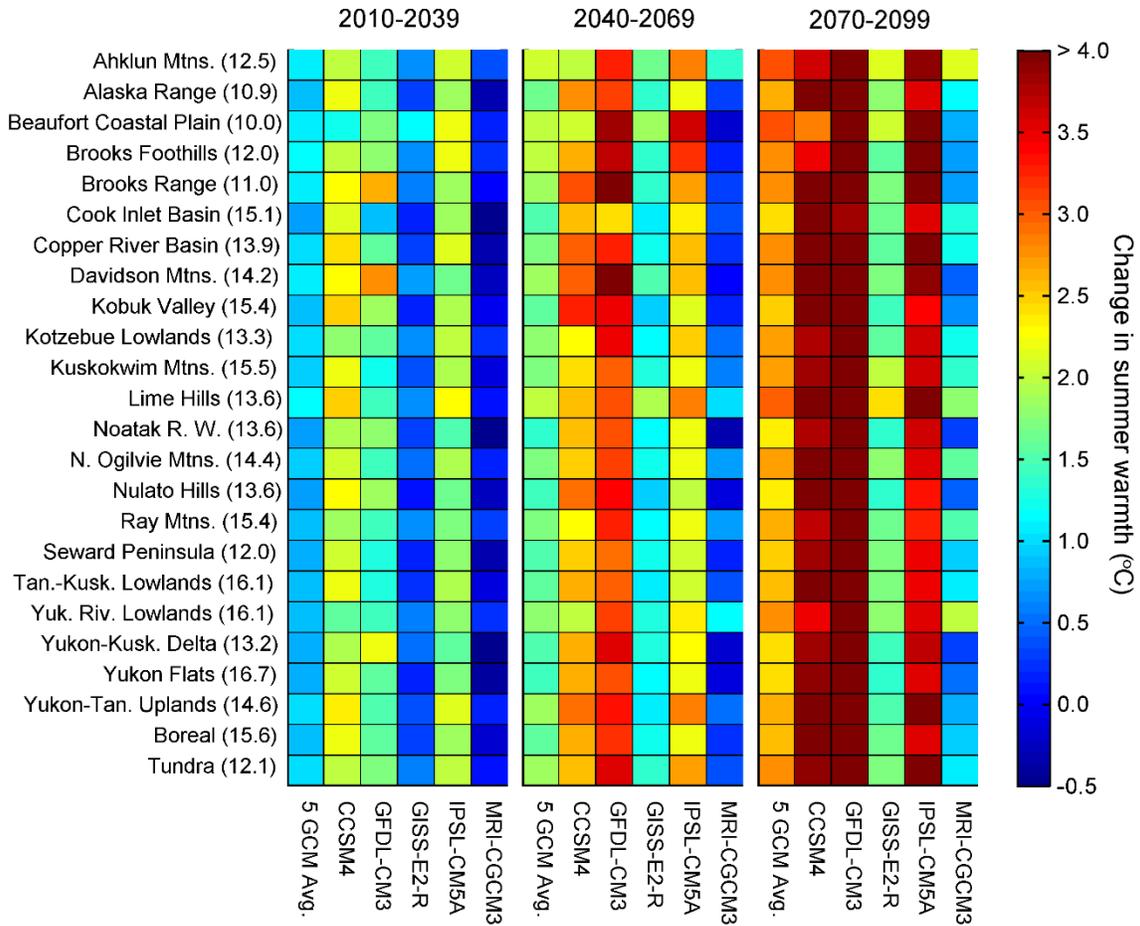


Fig. A5 Projected changes in summer warmth (i.e., mean temperature of the warmest month) for Alaskan ecoregions and the boreal forest and tundra spatial domains. Values in parentheses next to ecoregion names are the 1950-2009 averages, while colors indicate the magnitude of projected change for the five-GCM average and each GCM individually. Projected changes were calculated by taking the difference in projected climate for each 2-km pixel and then averaging this difference across each region and time period.

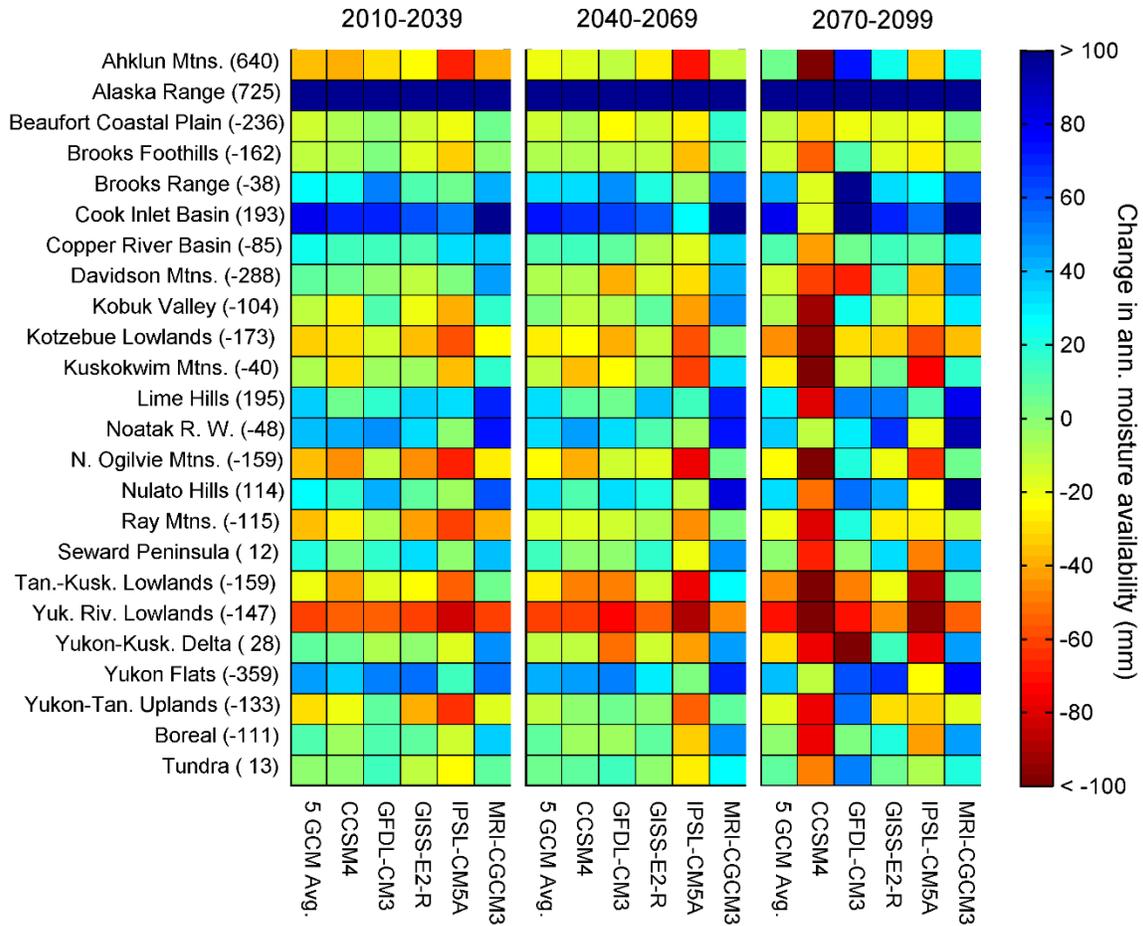


Fig. A6 Projected changes in annual moisture availability for Alaskan ecoregions and the boreal forest and tundra spatial domains. Values in parentheses next to ecoregion names are the 1950-2009 averages, while colors indicate the magnitude of projected change for the five-GCM average and each GCM individually. Projected changes were calculated by taking the difference in projected climate for each 2-km pixel and then averaging this difference across each region and time period.

References for supplementary materials

Chipman, M. L. et al. 2015. Spatiotemporal patterns of tundra fires: late-Quaternary charcoal records from Alaska. — *Biogeosciences* 12: 3177-3209.

Muggeo, V. M. R. 2003. Estimating regression models with unknown break-points. — *Statistics in Medicine* 22: 3055-3071.

Muggeo, V. M. R. 2008. segmented: an R Package to Fit Regression Models with Broken-Line Relationships. — *R News* 20-25.