

Projecting the Impact of Climate Change on Global Ecoregion Distributions

Andrew Burton, March 2021

Background

Climate change is likely the most daunting and potentially catastrophic issue humanity has ever faced. Our contribution towards upsetting the Earth's equilibrium carries far reaching, serious ramifications. While taking drastic steps to curb global greenhouse gas emissions is paramount in preventing worst case scenarios, in some regards irreversible damage has been done. With that in mind, understanding the multitude of long-term effects we are likely to experience over the coming decades is of vital importance as well.

One area of concern is the environment itself. While it may seem of low priority when compared to other issues arising from climate change, it is imperative we not underestimate the potential impact loss of biodiversity will have. We are in the midst of what is predicted to be the worst extinction event in Earth's history, with countless species at the risk of disappearing over the coming decades and centuries (Ceballos et al. 13597).

Among the driving forces for this trend is habitat loss. Much of this loss can be directly attributed to human actions like slash and burn farming and suburban sprawl. Additionally, as climate patterns shift, habitats are likely to be lost as a result of local conditions no longer being viable for native biota – less rain in some regions will result in desertification, rising sea levels will inundate low-lying areas, and increased temperatures will make environments uninhabitable to endemic species.

The Earth's environments can be divided into different ecoregion – areas with similar climate patterns that result in distinctive flora and fauna. Taking a broad view of the environmental impacts of climate change, this project aims to answer the question:

How will the distribution of distinct ecoregions change over the coming years as a result of climate change?

In order to answer this question, datasets of local elevations in addition to historical precipitation and temperature data was used to train a classifier to identify different ecoregions. Applying this model to projections of local temperature and precipitation allowed for predictions of how the Earth's ecoregions are likely to change as a result of climate change.

Literature Review

There are several established classification systems for terrestrial ecosystems. For the sake of this analysis, ecoregions as defined by the World Wildlife Fund and distributions mapped by Olsen et al. were used for labeling data (*figure 1*). This system includes 14 distinct ecoregions ("Terrestrial Ecoregions | Biome Categories | WWF"):

Deserts and xeric shrublands: Characterized by evaporation rate that exceeds precipitation, usually with less than 25 cm of rainfall a year. Seasonal temperature fluctuation is variable, with some regions hot year-round and others having cold winters. Because of low humidity and cloud cover temperatures have a high degree of daily variability, often dropping precipitously at night.

Tropical and subtropical moist broadleaf forests: Characterized by low seasonal temperature variability and precipitation rates, often exceeding over 200 cm annually. This ecotype has the highest biodiversity of any terrestrial ecotype and is usually equatorially located.

Tropical and subtropical dry broadleaf forests: Characterized by little seasonal temperature variation and often found within the tropics. While these areas may have high annual precipitation, there is a prolonged dry season, requiring biota be adapted for both intense rain and drought.

Tropical and subtropical coniferous forests: Characterized with dry winters and moist summers with mild seasonal temperature fluctuation.

Temperate broadleaf and mixed forests: There exists a degree of climatic variation within this ecoregion, but usually have colder winters with moderate to high levels of precipitation and humidity. Notable for deciduous broadleaf trees sometimes intermixed with conifers.

Temperate Coniferous Forest: Characterized by warm summers and cold winters, though precipitation rates vary between regions. Because of the cold winters, foliage is predominantly coniferous.

Boreal forests / Taiga: Characterized with low temperatures year-round with some seasonal variation and moderate precipitation, often as snow. The harsh conditions result in conifer-dominated foliage and low biodiversity.

Tropical and subtropical grasslands, savannas and shrublands: Characterized by low precipitation rates that preclude tree growth, but enough to support grasses and shrubs. These areas are usually warm with low seasonal temperature variability.

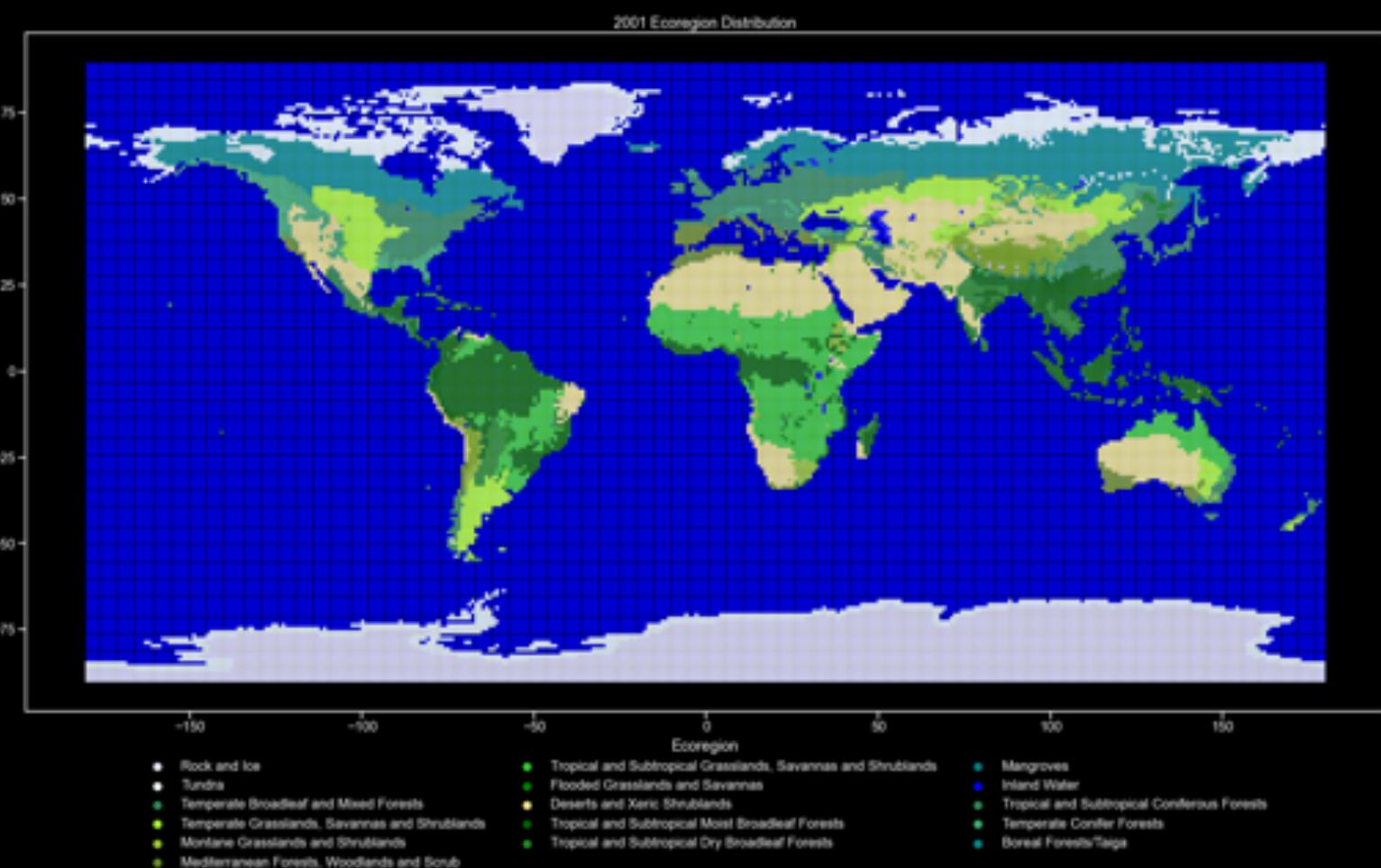


Figure 1: Ecoregion distribution according to WWF classification and Olsen et al.

Temperate grasslands, savannas and shrublands: Characterized by low to moderate precipitation rates but exhibiting a higher degree of seasonal temperature variation than its tropical counterpart.

Flooded grasslands and savannas: Characterized by flat expanses that experience year-round inundation and usually located in warm climates.

Montane grasslands and shrublands: Characterized by moist conditions with seasonal variation. High elevations result in cool temperatures and increased exposure to solar radiation.

Tundra: Characterized by cold temperatures with low rainfall. Found at high latitudes, causing high seasonal variability in day length. These extreme conditions result in low biodiversity.

Mediterranean Forests, woodlands and scrubs: Characterized by warm summers with low rainfall and cool, moist winters. This pattern is unique among ecoregions.

Mangroves: Found along coasts where fluctuating tides cause variability in inundation and salty soil, resulting in unique biota adapted for these conditions.

In determining what features to use for training models, the USGS's Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States was used as a guideline. This publication lists 19 metrics for evaluating and predicting environmental conditions. While intended for application to United States, these indices are broadly applicable. A fair amount of the formulae can be calculated using monthly mean temperature and precipitation records (T_{Avg} and P_{Avg} indicate the average for a given month):

$$\text{Annual Mean Temperature: } AMT = \frac{\sum_{i=1}^{12} T_{Avg_i}}{12}$$

$$\text{Temperature Seasonality (Standard Deviation): } TS = SD\{T_{Avg_1}, T_{Avg_2}, \dots, T_{Avg_{12}}\}$$

$$\text{Annual Precipitation: } AP = \sum_{i=1}^{12} P_{Avg_i}$$

$$\text{Precipitation Seasonality (Coefficient of Variation): } PS = \frac{SD\{P_{Avg_1}, P_{Avg_2}, \dots, P_{Avg_{12}}\}}{1 + \left(\frac{AP}{12}\right)}$$

Other metrics require monthly maxima and minima to calculate. This includes:

$$T_{max}, T_{min}, P_{max}, P_{min}, \text{ and Annual Temperature Range (ATR)} = T_{max} - T_{min}$$

Historical records of this data are readily available, but global forecasts of local climatic monthly extremes are hard to come by. Without these projections, forecasting would not be possible on models trained using these metrics. So, for the sake of this analysis, calculations for the above values were performed using mean monthly values.

Additionally, the USGS uses several metrics based off of rolling quarterly means:

Temperature of Warmest Quarter, Temperature of Coolest Quarter, Precipitation of Wettest Quarter, Precipitation of Driest Quarter, Temperature of Wettest Quarter, Temperature of Driest Quarter, Precipitation of Warmest Quarter, Precipitation of Coldest Quarter.

There are two additional indices, Annual Mean Diurnal Range and Isothermality used in the paper which could not be adapted using the data available. Annual Mean Diurnal Range requires the monthly range of temperatures to calculate:

$$\text{Annual Mean Diurnal Range: } AMT = \frac{\sum_{i=1}^{12} (T_{max_i} - T_{min_i})}{12}$$

And isothermality is calculated in part using AMT:

$$\text{Isothermality: } ISO = \frac{AMT}{ATR}$$

Unfortunately, these two indices could not be included in training models.

Projections in this analysis are based off of the IPCC's Representative Concentration Pathway (RCP) for global greenhouse gas concentrations, which includes several scenarios for peak greenhouse gas concentrations and their impact on temperature, precipitation, and sea level trends. Four scenarios covering the gamut of possibilities were used for projections:

Model	Peak CO ₂ Emissions	Mean Global Temperature Increase (°C)		Mean Global Sea Level Increase (m)	
		2046-2065	2081-2100	2046-2065	2081-2100
RCP2.6	2020	1.0	1.0	0.24	0.40
RCP4.5	2040	1.4	1.8	0.26	0.47
RCP6.0	2080	1.3	2.2	0.25	0.48
RCP8.5	None	2.0	3.7	0.30	0.63

Given current trajectories, RCP2.6 is increasingly unlikely. RCP8.5 is considered the worst-case pathway and considered unlikely as well. However, multiple runaway feedback loops have the slim potential of causing this catastrophic scenario. To date, global CO₂ levels are still rising. Despite some countries decreasing their emissions, there is little sign of greenhouse gas emissions leveling off on a global scale, putting us within the RCP4.0 and RCP6.0 scenarios.

Others have developed models to try to predict climate change's impact on Earth's ecosystems. Most of these have been on a more local scale and focus on vegetation change. Hilbert et al. used neural networks to predict changes in Australian vegetation groups. The features they used to train their model include the USGS's climate indicators in addition to several moisture and radiation metrics. Their models had a 0.96 accuracy rate, with annual mean temperature, annual mean diurnal range, isothermality, temperature seasonality, and max temperature of the warmest period ranking highest importance. Overall, they found expansion of deserts at the cost of temperate grasslands, temperate broadleaf/mixed forests, and Mediterranean forests in Australia.

Hoffman, et al. used climate projections to assess the risk posed to various protected terrestrial regions. Their methodology entailed assessing degree to which these areas will be subjected to novel climate patterns (as calculated by the local-scale novel climate index) with comparative impacts varying by RCP scenario. In general, temperate grasslands, temperate conifer forests, temperate broadleaf/mixed forests, and boreal forests were projected to have the highest exposure to novel climates.

A paper by Yu et al. used projections of temperature, precipitation, and cloud cover to find areas at high risk as determined by primary productivity, carbon storage, runoff, wildfire risk, and habitat transformation. Areas they found to be of high risk include eastern North America, Europe/western Asia, India, and western China.

Studies using classification algorithms to project ecoregion distributions on a global scale were not found in review.

Data

Several datasets were used in this analysis. For training models:

Monthly Mean Precipitation and Monthly Land Surface Mean Temperatures: Datasets covering 1990 to 2001 were obtained from NOAA with global coverage on a 0.5-degree latitude by 0.5-degree longitude grid. Temperature records are in Celsius and precipitation is measured in centimeters.

Elevation data was obtained from the USGS. The dataset is consolidated from several sources for global terrestrial coverage of maximum, minimum, mean, and standard deviation of elevation. The resolution varies depending on location, either 7.5, 15, or 30 arcseconds. Measurements are in meters above sea level.

To label the data, a dataset of global terrestrial ecoregions from the Nature Conservancy in the form of polygonal shape data based off of research by Olson et al. and modified by the Nature Conservatory. The dataset includes several different classification schemes, but for the purpose of this paper, only the ecoregions as defined by the World Wildlife Fund were used.

For projections:

Predicted monthly mean precipitation and monthly mean surface temperature data was obtained from NCAR's GIS Program. Projections from this dataset run through to 2100, covering a global 1-degree latitude by 1-degree longitude grid.

Projected mean sea level rise projections were obtained from Climate Econometrics. Datasets for RCP4.5, RCP6.0, and RCP8.5 scenarios were available. Projections are in 10-year increments and run through 2100 and measured in meters. Because sea level rise projections for RCP2.6 were not readily available, sea level was assumed to hold constant for this scenario.

Data Preparation

The first step in preparing data entailed spatially joining elevation, precipitation, elevation data and applying the corresponding ecoregion to each point. To do so, a 1-degree latitude by 1-degree longitude grid was constructed and intersected with elevation data. Because the elevation dataset was a composite of several subsets with some overlap, care had to be taken to eliminate redundant measurements while ensuring global coverage. The gridded elevation data was then spatially joined with the ecoregion dataset. This was then merged with the combined precipitation and temperature data using latitude and longitude as the key. The resulting dataframe had 21535 geographical coordinates covering monthly records from 1990 through 2001.

Precipitation and temperature data were missing roughly 3354 points per year – it was consistent which coordinates were missing across each year (*Figure 2*). For each of these missing values, the dataframe was checked for any adjacent points within 1 degree latitude and longitude in all cardinal and intercardinal directions for the corresponding month. If any adjacent points were present, the missing value was interpolated using their mean. If no adjacent points were found, the search was increased to 2 degrees latitude and longitude. A third iteration was then used if still no points were found after which the missing value was replaced by the mean of the same ecoregion for that given month. The first iteration replaced 3172 values for each year, with 26 values being replaced by the subsequent iteration. The third iteration did not replace any values and the remaining 156 points were filled with the mean of their ecoregion.

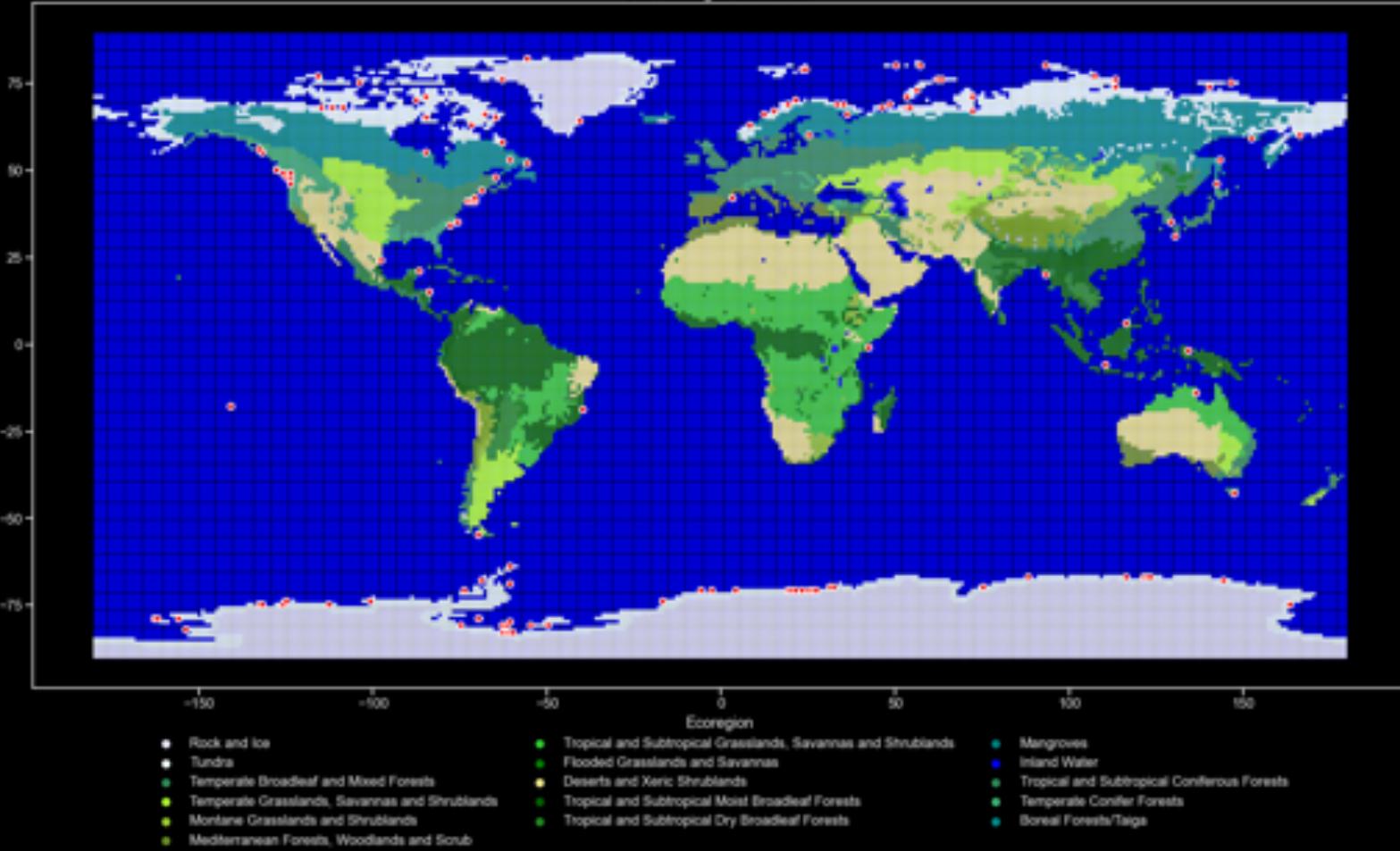


Figure 2: Missing value locations.

Because the sea level rise data only had values for every decade, the intervening years had to be interpolated as well. This was a simple matter of using the slope between existing datapoints to fill in the yearly values.

Using the existing data, several additional features were engineered. The elevation range for each point was added as an indicator for how steep the terrain is at each given location.

There were a few possible outliers – some temperatures in the dataset that exceeded the highest recorded land temperature of 93.9 °C (*figure 3*). Additionally there some precipitation entries were less than 0, which can safely be assumed to be erroneous (*figure 4*). All entries with temperatures greater than 95.0 °C and precipitation less than 0 cm were removed.

For forecasts, monthly precipitation and temperature projections for each RCP scenario were spatially joined with elevation and ecoregion labels in a similar manner as above. The projected sea level rise was subtracted from features pertaining to elevation for each respective year and scenario. The same engineered features generated in the training data were calculated and merged with projections.

Temperature by Ecoregion

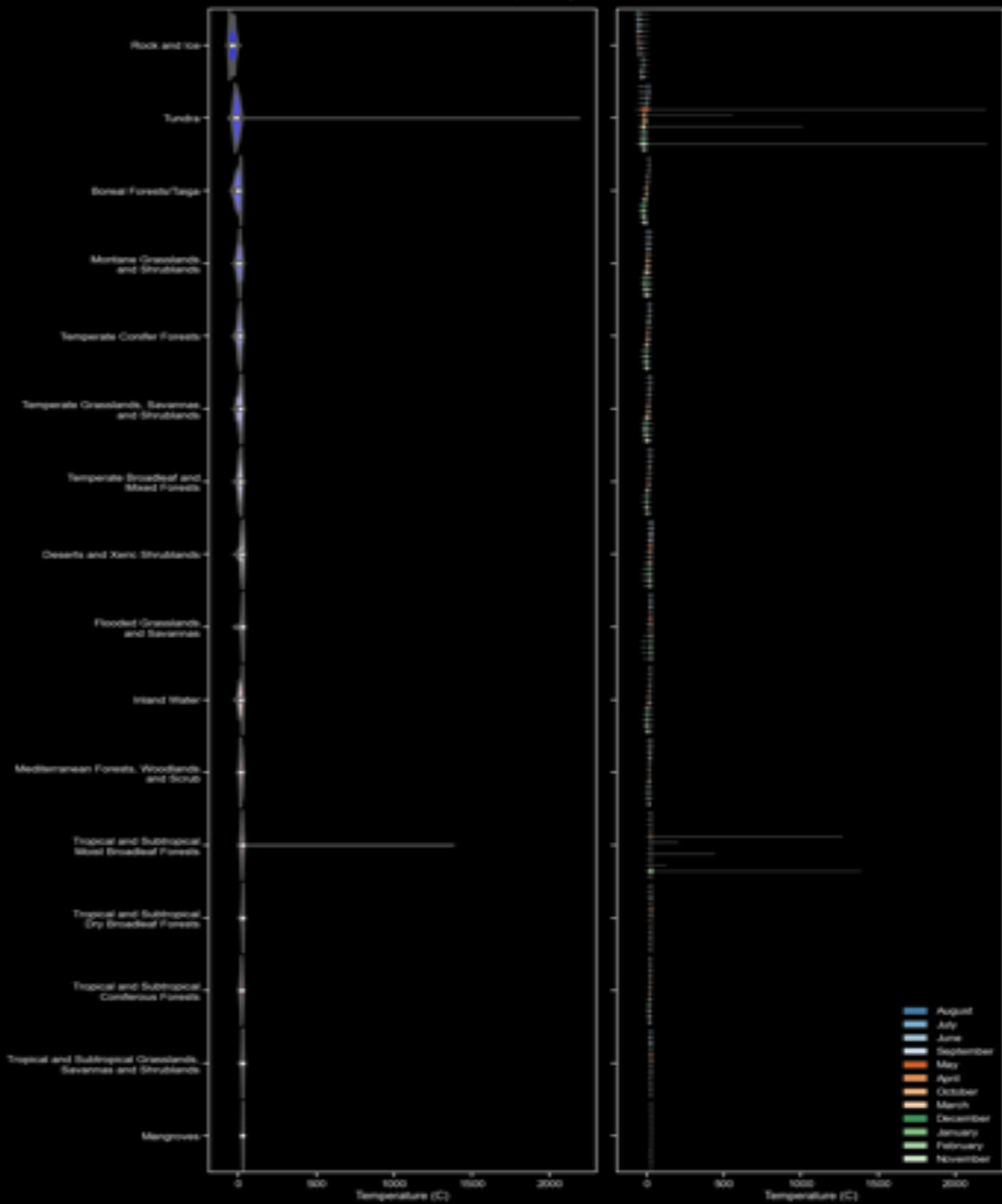


Figure 3: Temperature distributions by ecoregions.

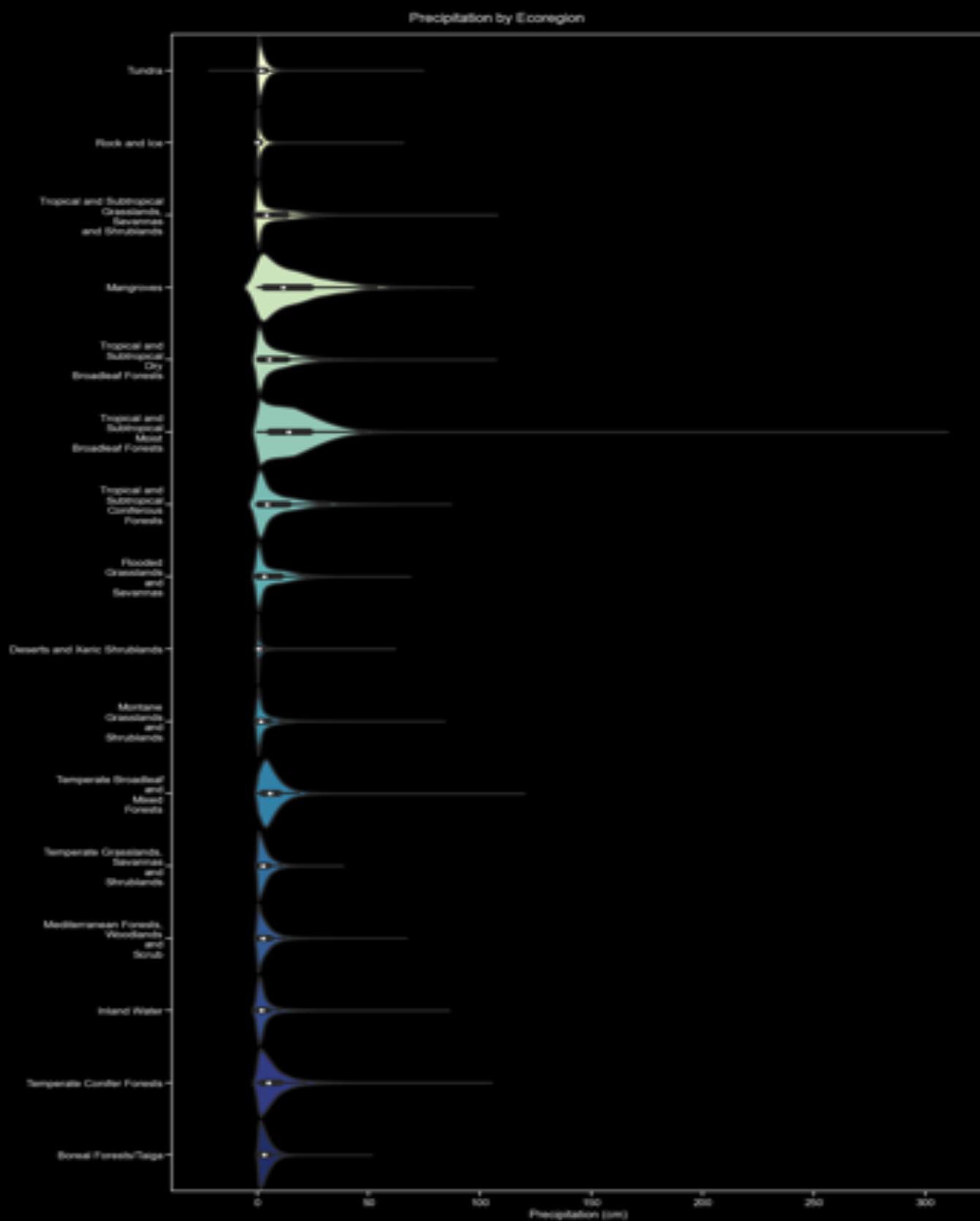


Figure 4: Precipitation distributions by ecoregions.

Exploratory Analysis

Looking at the prevalence of each ecoregion in the training data, there is significant bias towards particular classes, especially rock and ice (*fig 5*). Several ecoregions had negligible representation: mangroves; inland water; tropical and subtropical coniferous forests; and flooded grasslands and savannas. If not addressed, this lopsidedness will likely bias models towards the overrepresented ecoregions.

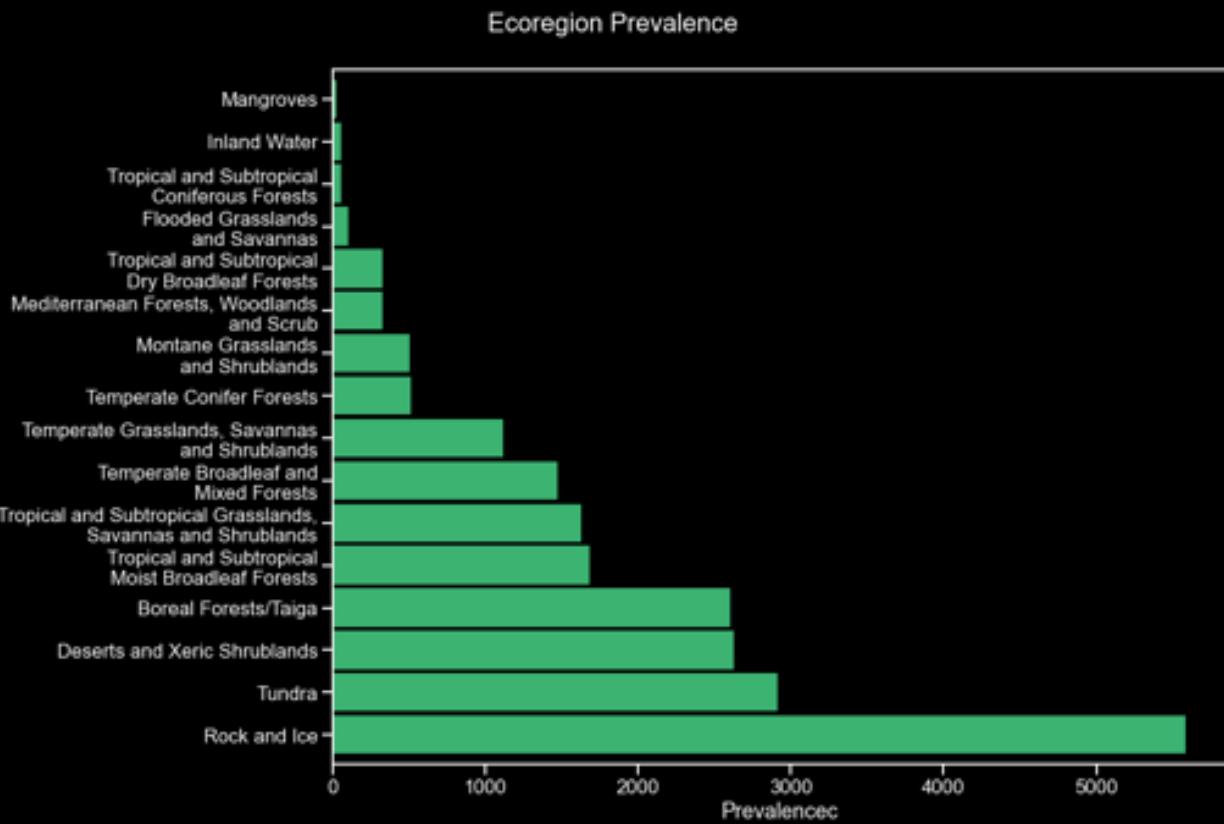


Figure 5: Ecoregion prevalence in data.

Each ecoregion does appear to have distinctive monthly precipitation and temperature trends. Both tundra and rock and ice have consistently cold temperatures with some degree of seasonality. Boreal forests/taiga have a large swing in temperatures from winter to summer. Temperate conifer forests also display a high degree of seasonal variability. The remainder of ecoregions have much flatter monthly trends with varying baselines – in general warmer climates seem to have less seasonal variation. (*figure 6*)

Tropical and subtropical moist broadleaf forests display consistently high monthly precipitation. Mangroves had the highest precipitation rate followed by tropical and subtropical moist broadleaf forests. Tropical and subtropical coniferous forests are notable for the extreme seasonal variation in temperature. Deserts and tundra are the driest ecoregions. (*figure 7*)

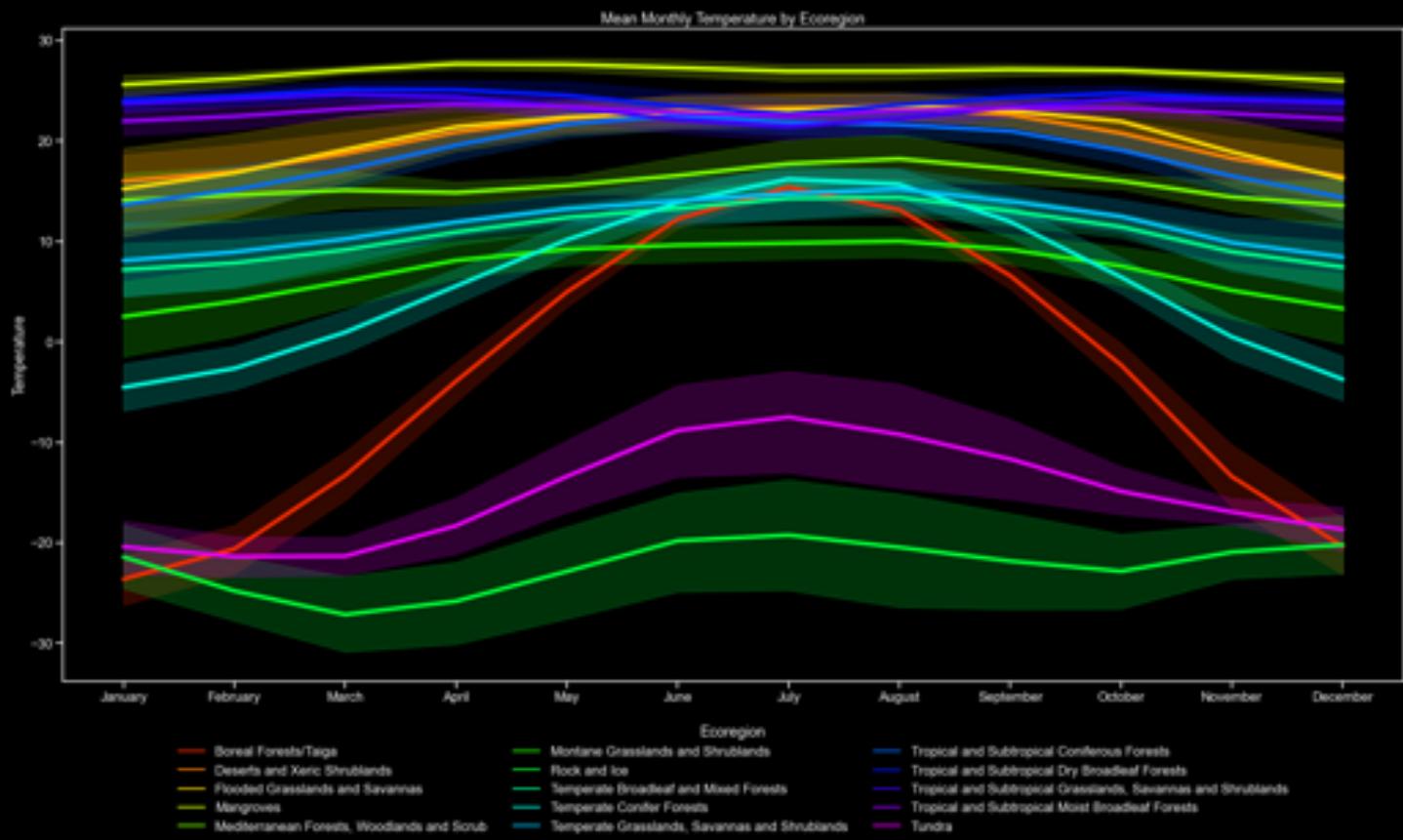


Figure 6: Mean monthly temperature trends averaged by ecoregion.

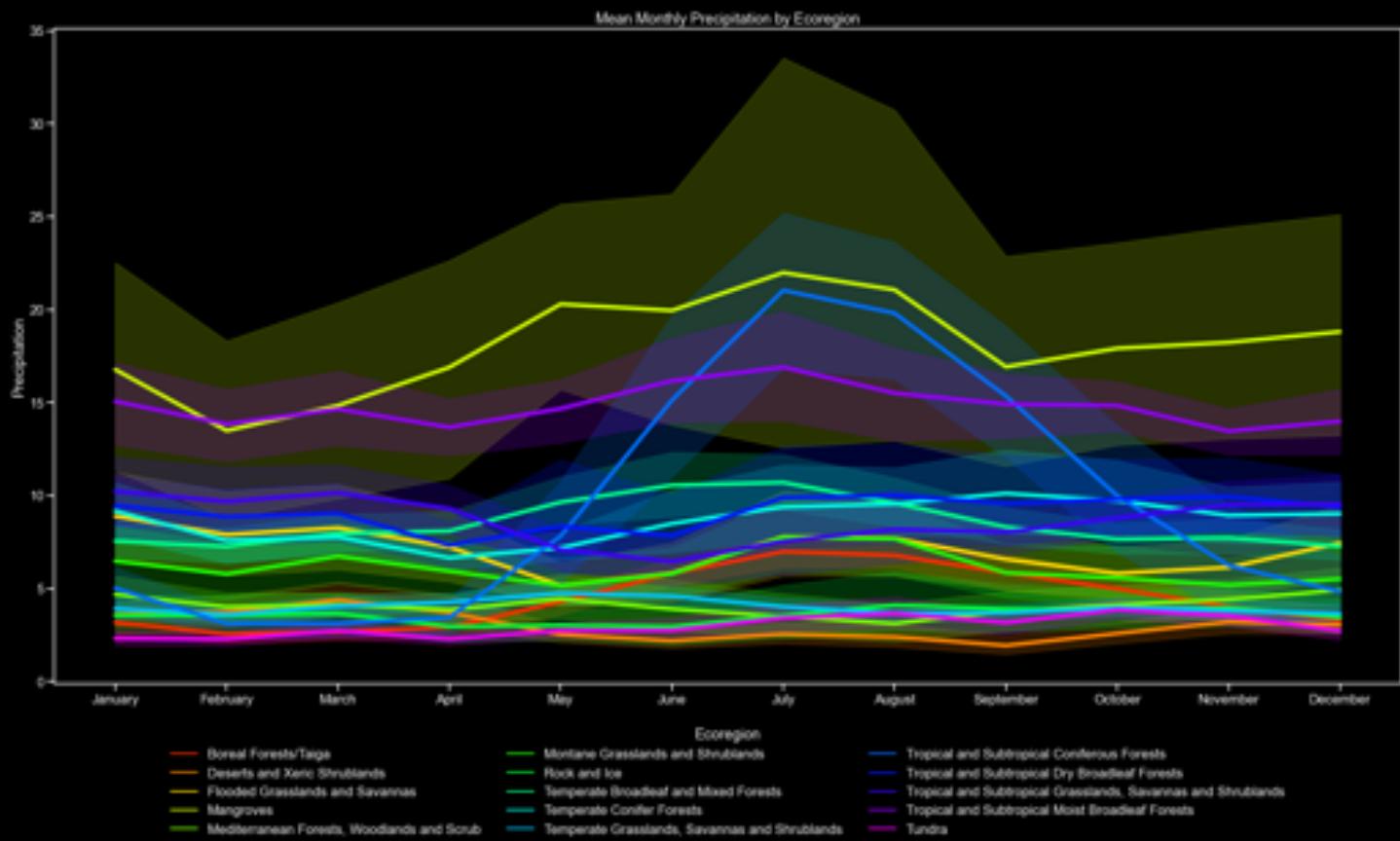


Figure 7: Mean monthly precipitation trends averaged by ecoregion.

There are interesting trends when further looking at monthly means by latitude and ecoregions. First off, for most ecoregion there is a flip in the trends across the equator, which is to be expected. This is more distinct in ecoregion found further from the equator like boreal forests and tundra. Equatorial ecoregions exhibit less temperature variation as well (*figure 8*).

The cross-equator flip is less distinguished for mean monthly precipitation, but noticeable for some ecoregions. There are several ecoregions that have consistently low precipitation rates, including boreal forests/taiga, tundra, and temperate grasslands. Interestingly, some desert areas have seasonal increases in precipitation,

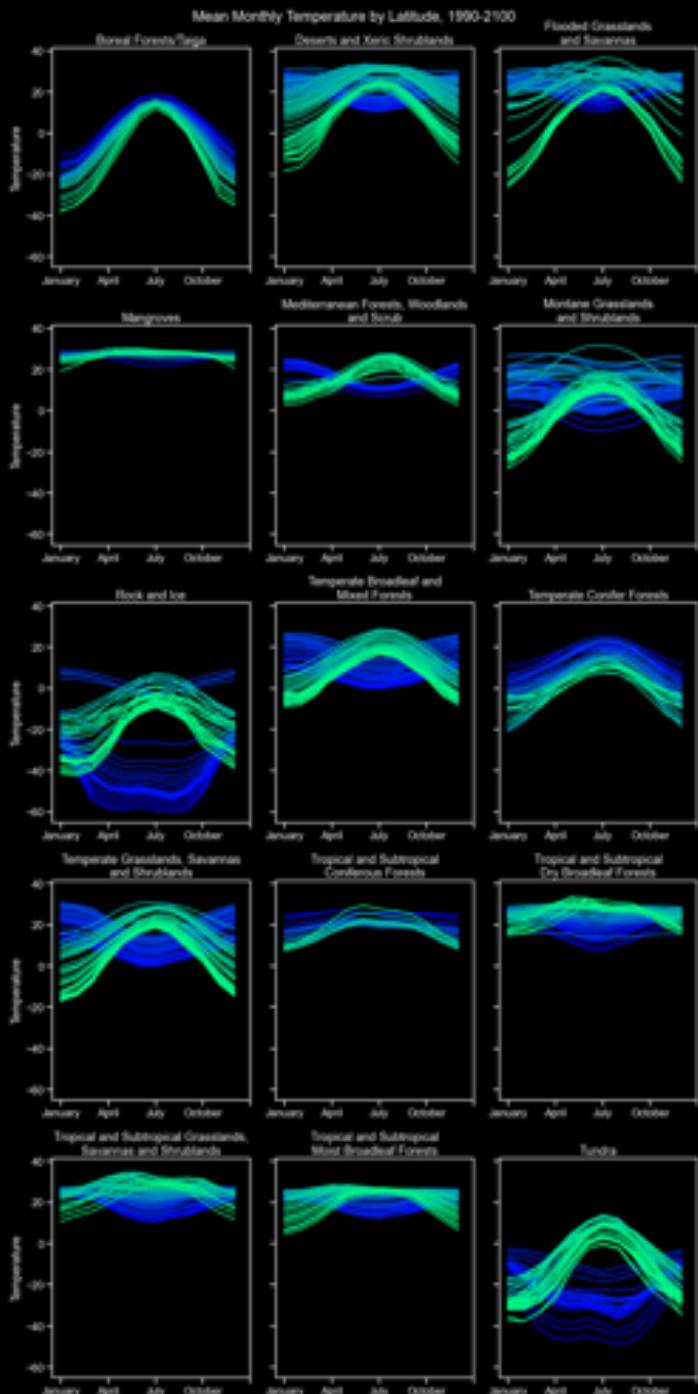


Figure 8: Mean monthly temperatures averaged by ecoregion and month.

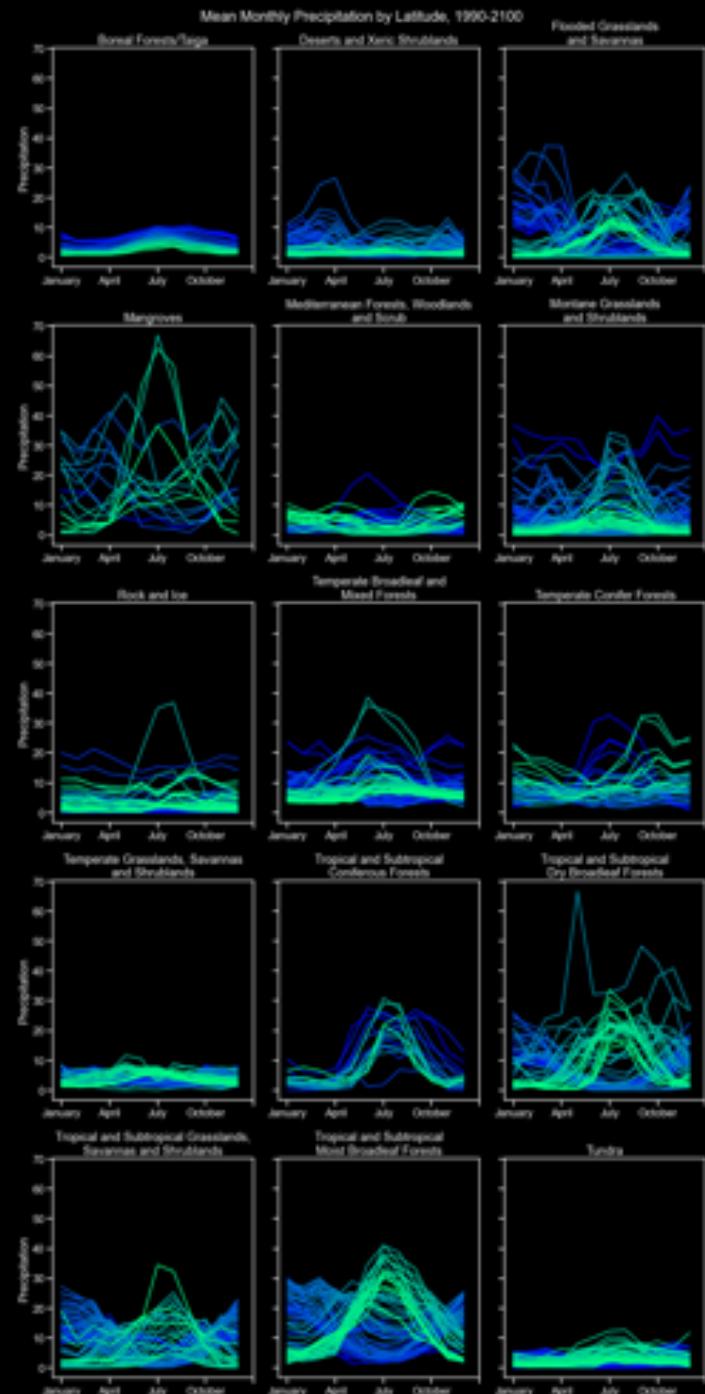


Figure 9: Mean monthly temperatures averaged by ecoregion and month.

mostly at southern latitudes. For ecoregions with a higher degree of seasonal variation in temperature, the peak occurs around July north of the equator and December in southern locations. The only exception to this is Mediterranean forests, which peaks in October north of the equator and in July in the South. (*figure 9*)

Elevation trends across ecoregions are pretty intuitive, with montane grasslands having, on average, the highest elevations followed by rock and ice, coniferous forests (tropical, subtropical, and temperate). Tundra also has a relatively high mean elevation. Mangroves, flooded grasslands and boreal forests have the lowest elevation values. (*figure 10*)

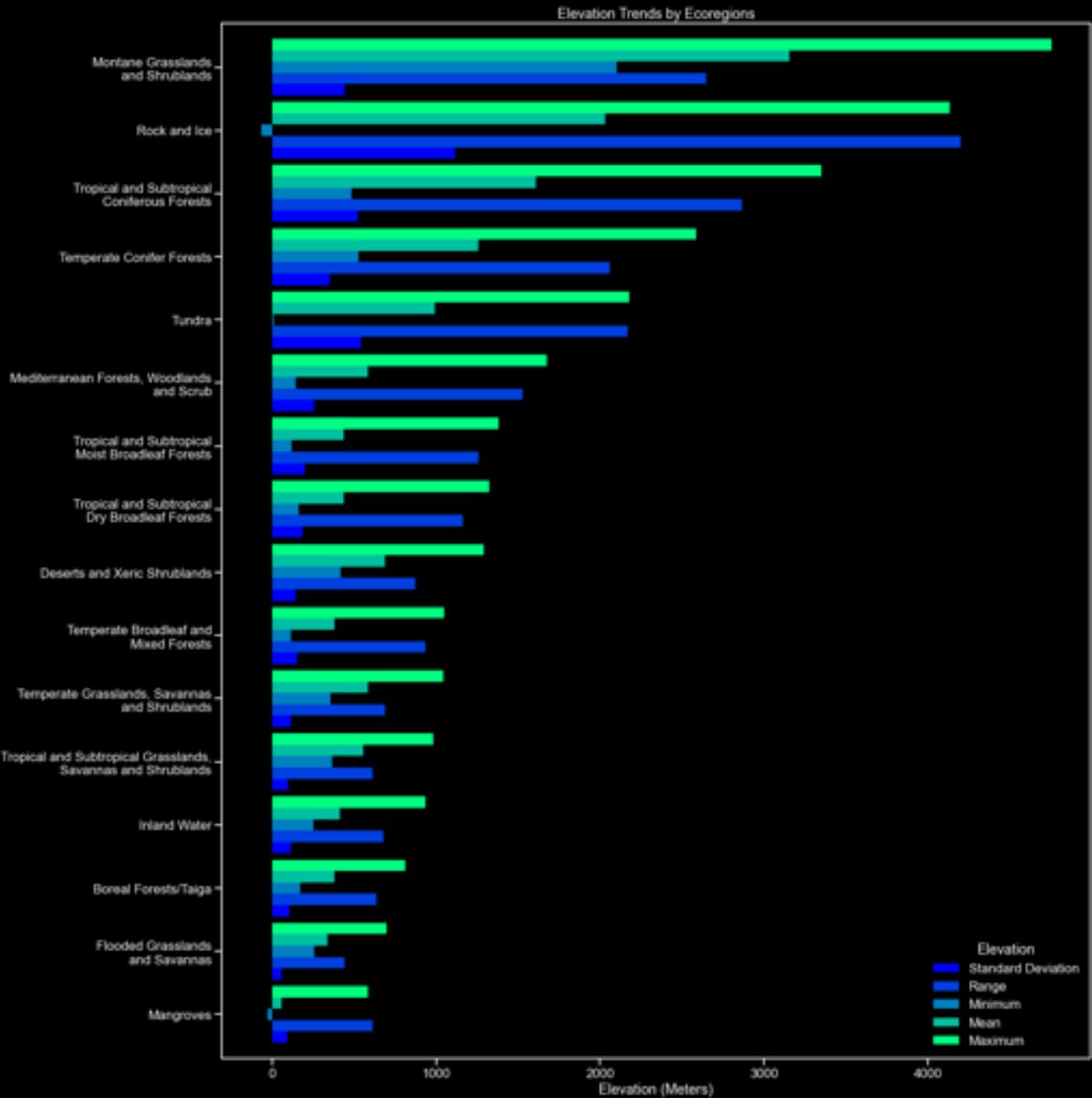


Figure 10: Elevation trends by ecoregion.

Precipitation trends for engineered features are shown in *figure 11* and *figure 12*, again showing distinctiveness of each ecoregion across these metrics.

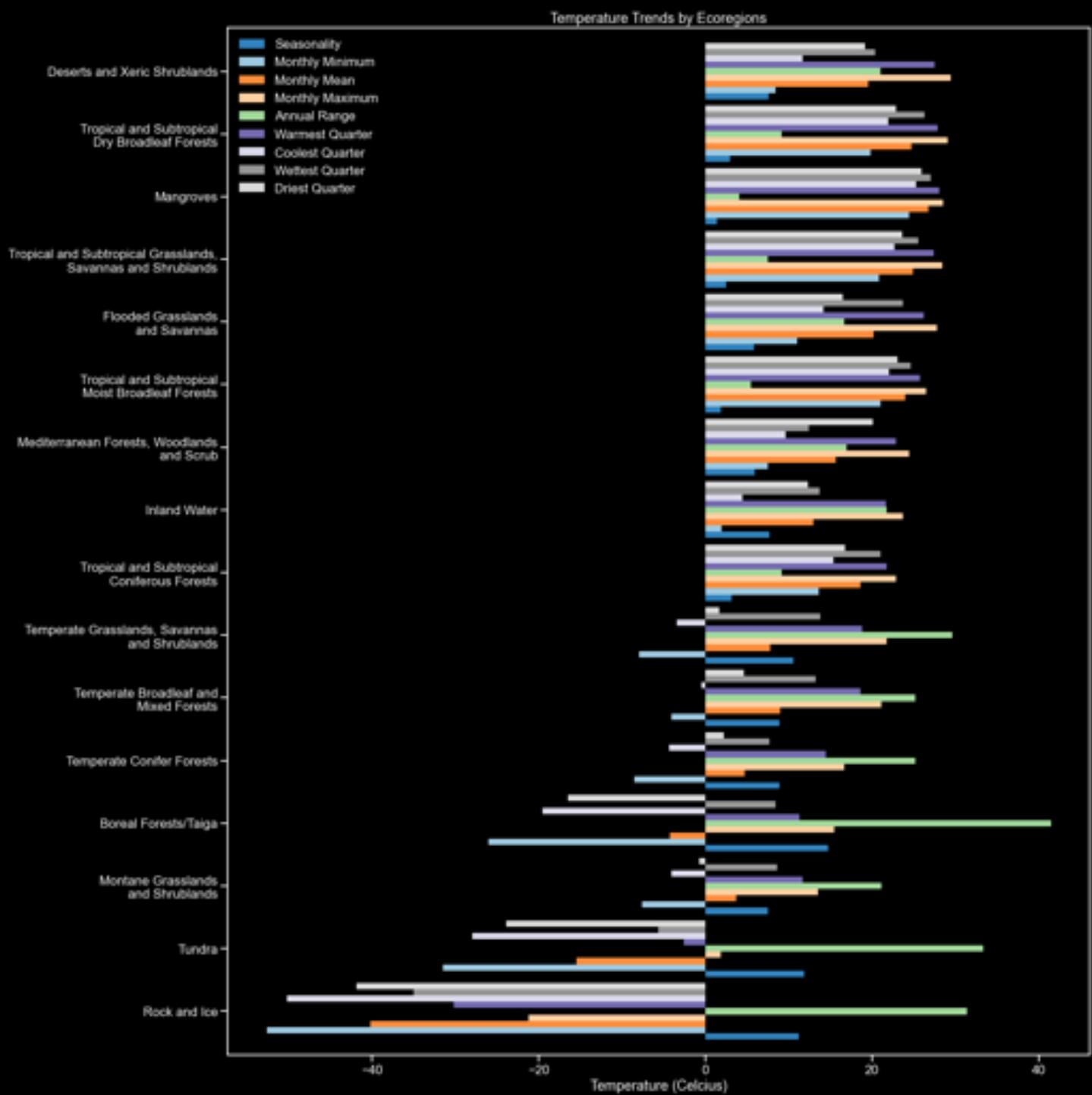


Figure 11: Temperature trends averaged by ecoregion for engineered features.

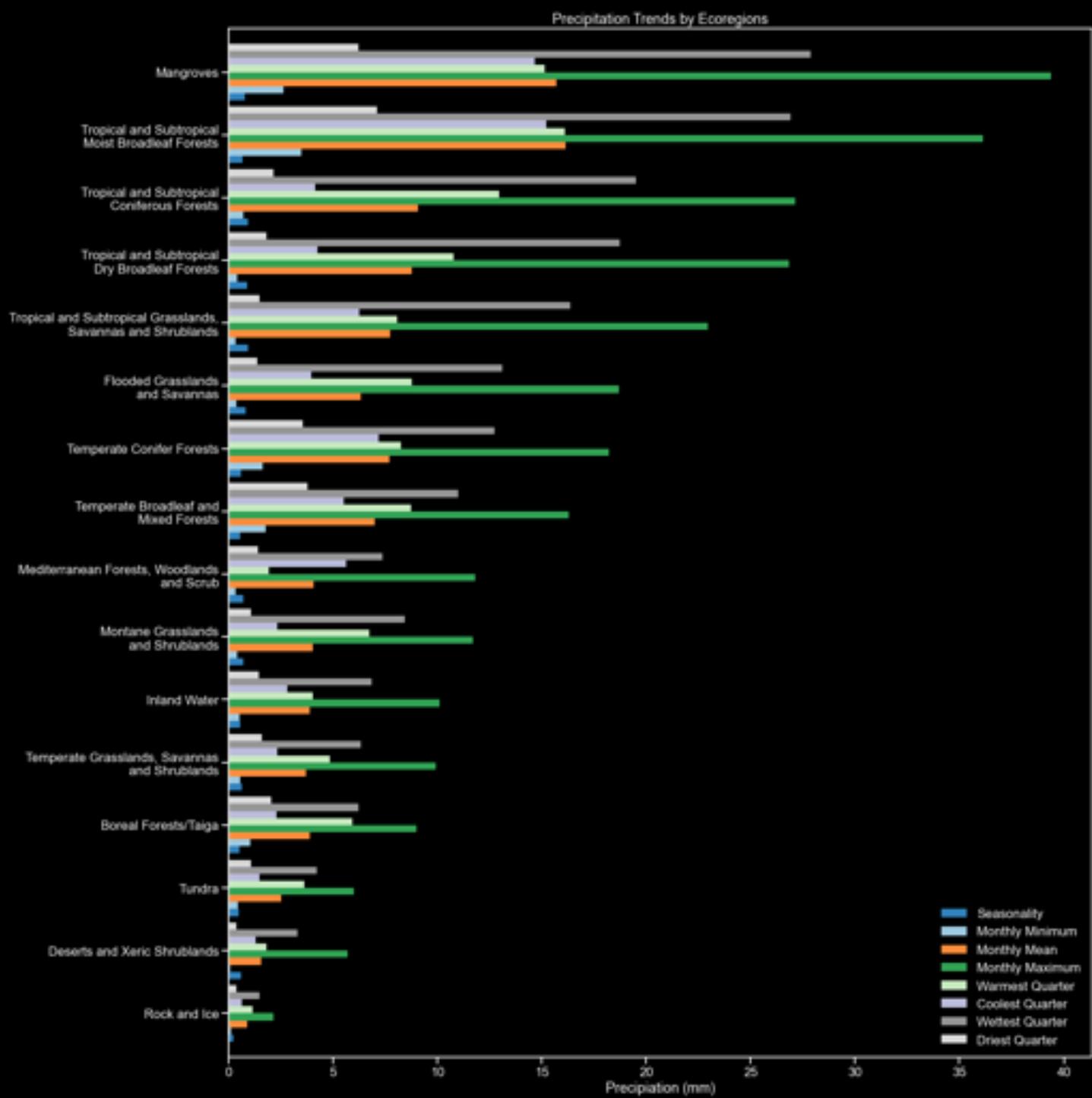


Figure 12: Temperature trends averaged by ecoregion for engineered features.

Projections of global mean sea level rise are shown in figure 13 according to RCP scenarios.

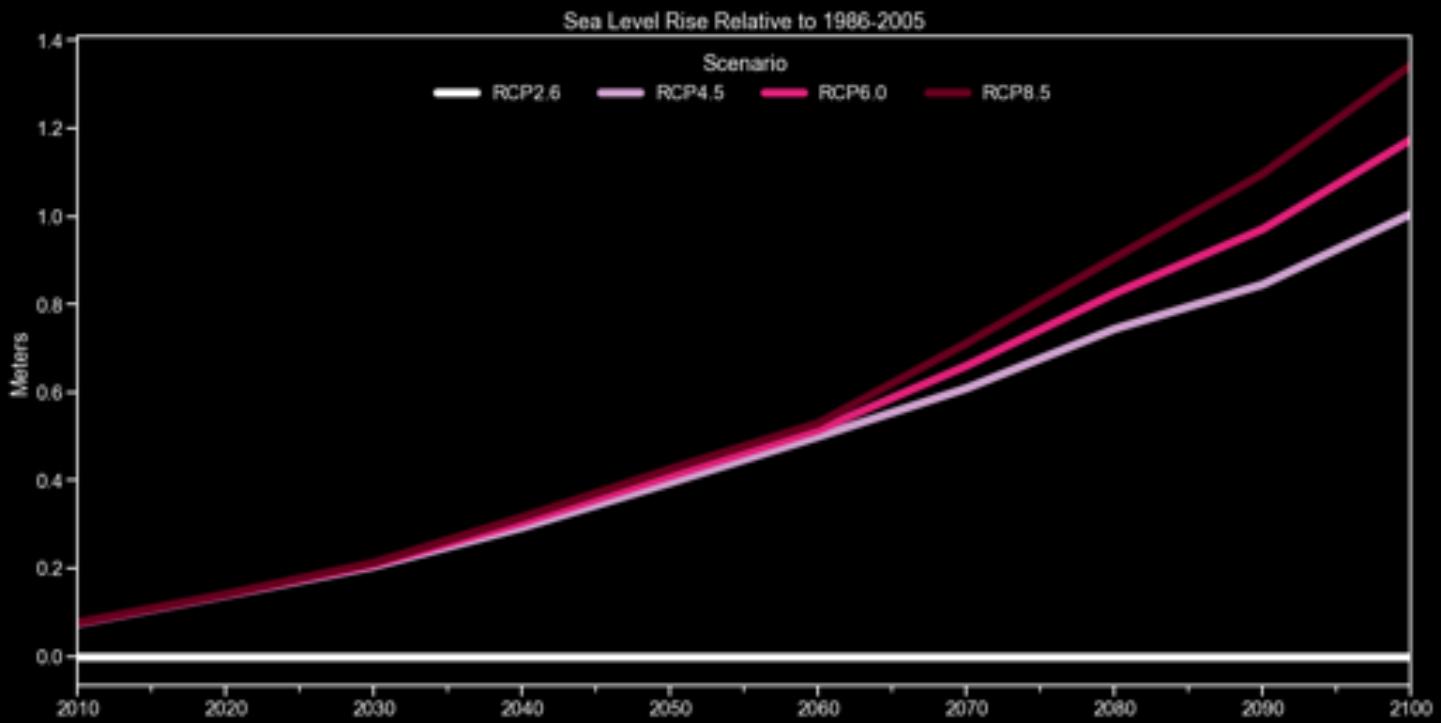


Figure 14: RCP Sea level rise projections

Looking at projections of monthly mean temperature and precipitation shows different ecoregions are predicted to be impacted by climate change to varying degrees. In terms of temperature, ecoregions further from the equator are likely to experience the most warming, which is problematic given that much of the Earth's fresh water is locked up in ice at these locations. Thawing of these regions will have a significant impact on sea level rise. (figure 14)

As for precipitation, the trends seem to be a little more scattered, but in general most ecoregions will experience increased precipitation throughout the century. The exceptions for this are Mediterranean forests and tropical/subtropical regions, which will see precipitation rates decline, especially for worst-case scenarios. (figure 15)

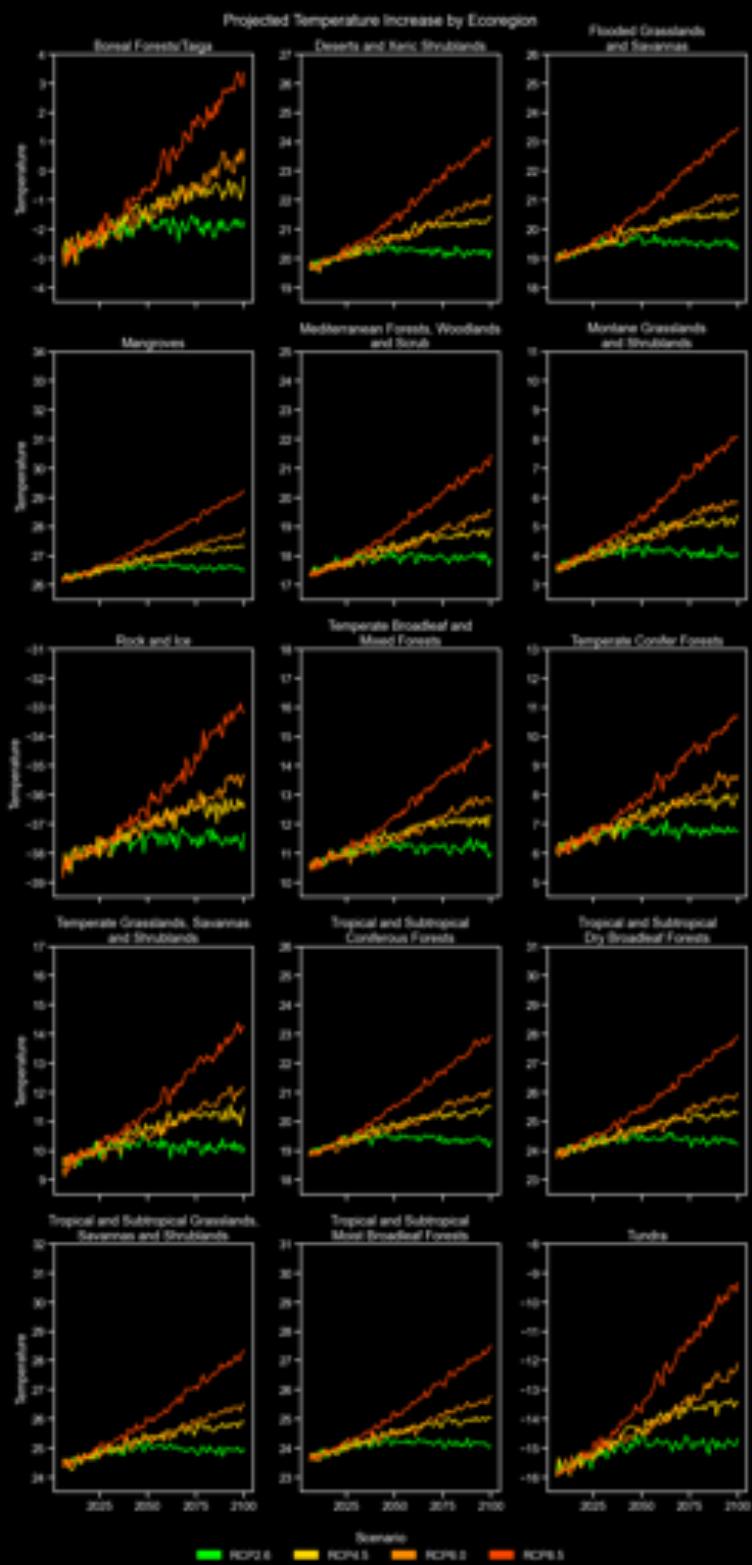


Figure 13: Projected mean yearly temperature change by ecoregion

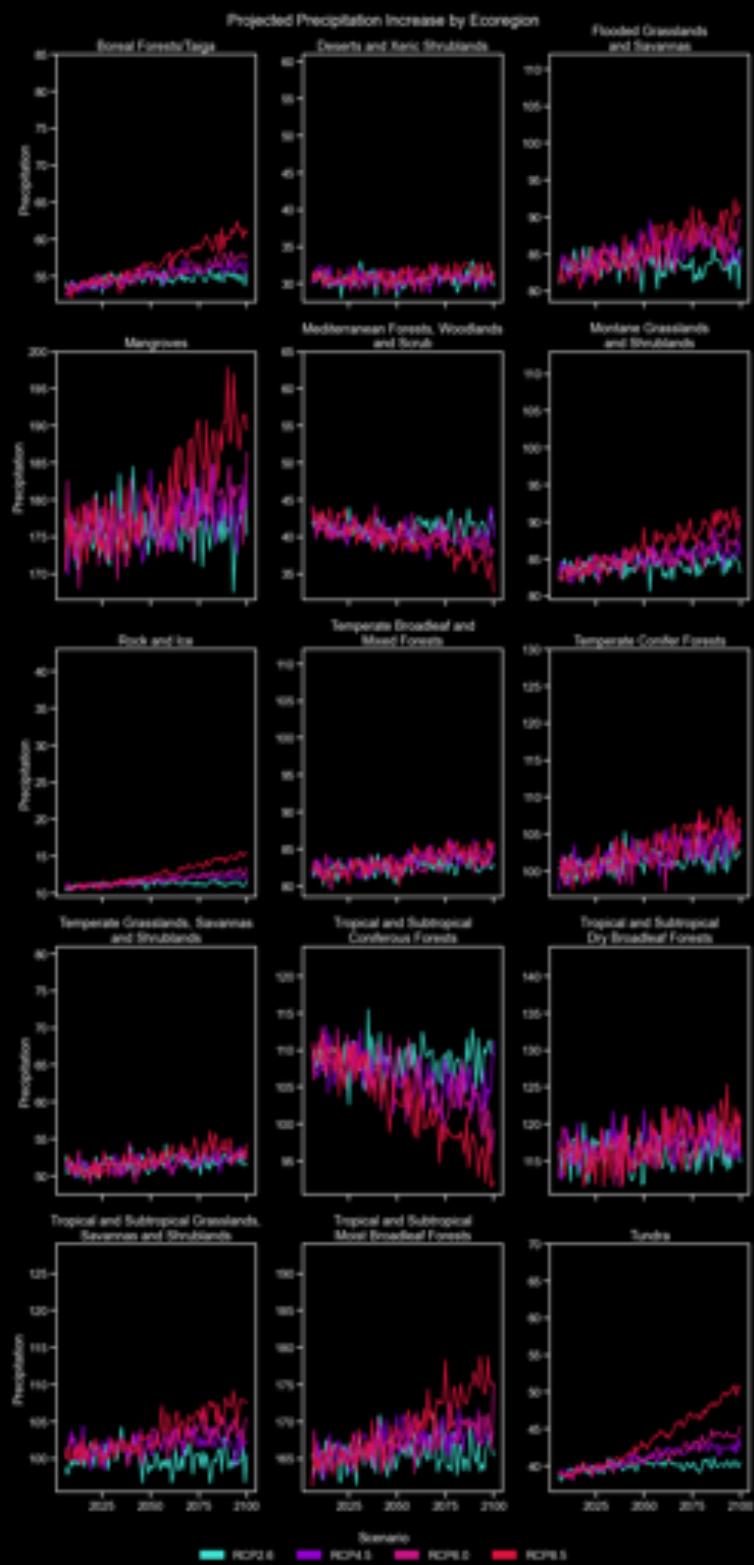


Figure 13: Projected mean yearly precipitation change by ecoregion

Preliminary Models

For the purpose of deciding the type of classifier to use, a few preliminary models were trained on data from 2001. Engineered features were not used in the training of these exploratory models. Naïve Bayes and SGM models did not perform well (with modest exploration of hyperparameter tuning), both having accuracies in the 60s and were eliminated as possibilities.

Random forest and SVM models performed comparatively well, having accuracies of about 0.89 and 0.88, respectively. Different hyperparameters were adjusted to achieve this accuracy – for random forest model, the number of trees estimators, maximum tree depth, maximum features used per level. Increasing the number of estimators from 100 to 200 had a moderate improvement in accuracy and helped address overfitting issues. No limit on depth of the trees and using the root number of features per level were found to optimize accuracy.

A grid search method was used to optimize SVM hyperparameters, iterating over permutations of:

Regularization parameter: $C \in \{0.1, 1, 10, 100, 1000\}$

Kernel coefficient: $\gamma \in \{1, 0.1, 0.01, 0.001, 0.0001\}$

Kernel: linear, polynomial, radial basis function (rbf)

The method yielded $C = 0.1$, $\gamma = 0.01$, and rbf kernel as the optimum values, though changing γ to 0.1 improved accuracy from 0.83 to 0.88. A SVM model was also trained on data reduced by PCA, decreasing the number of variables such that 0.99 percent of the variance is explained by the reduced-dimensionality data. This decreased the model accuracy to 0.52. Increasing the threshold of variance for variable reduction as high as 0.9999 performed did not improve results over the model using all features.

Further improvement in the models would require more data in addition to adding in engineered. Because of the time complexity, training a SVM on 12-years-worth of data was prohibitively slow, so random forests were ultimately chosen for classification.

Training Models

Several candidate models were trained on the full scope of data, which in total had 279,865 entries with 48 features. Models were trained on 70% with the remainder for the test set. As part of evaluating these models, they were applied to 2001 data pulled from NCAR (the source of climate projection data) to anticipate their performance on projected values.

As a benchmark, the first model (**A**) trained used all available features (*figures 16 and 17*). While this model performed well on the test data, having an accuracy of 0.951, it struggled to correctly identify less prevalent ecoregions, likely to their sparsity in training data (*table I*). This model did reasonably with on the NCAR data as well, though underpredicted the less prevalent ecotypes and underpredicted desert, especially in Australia.

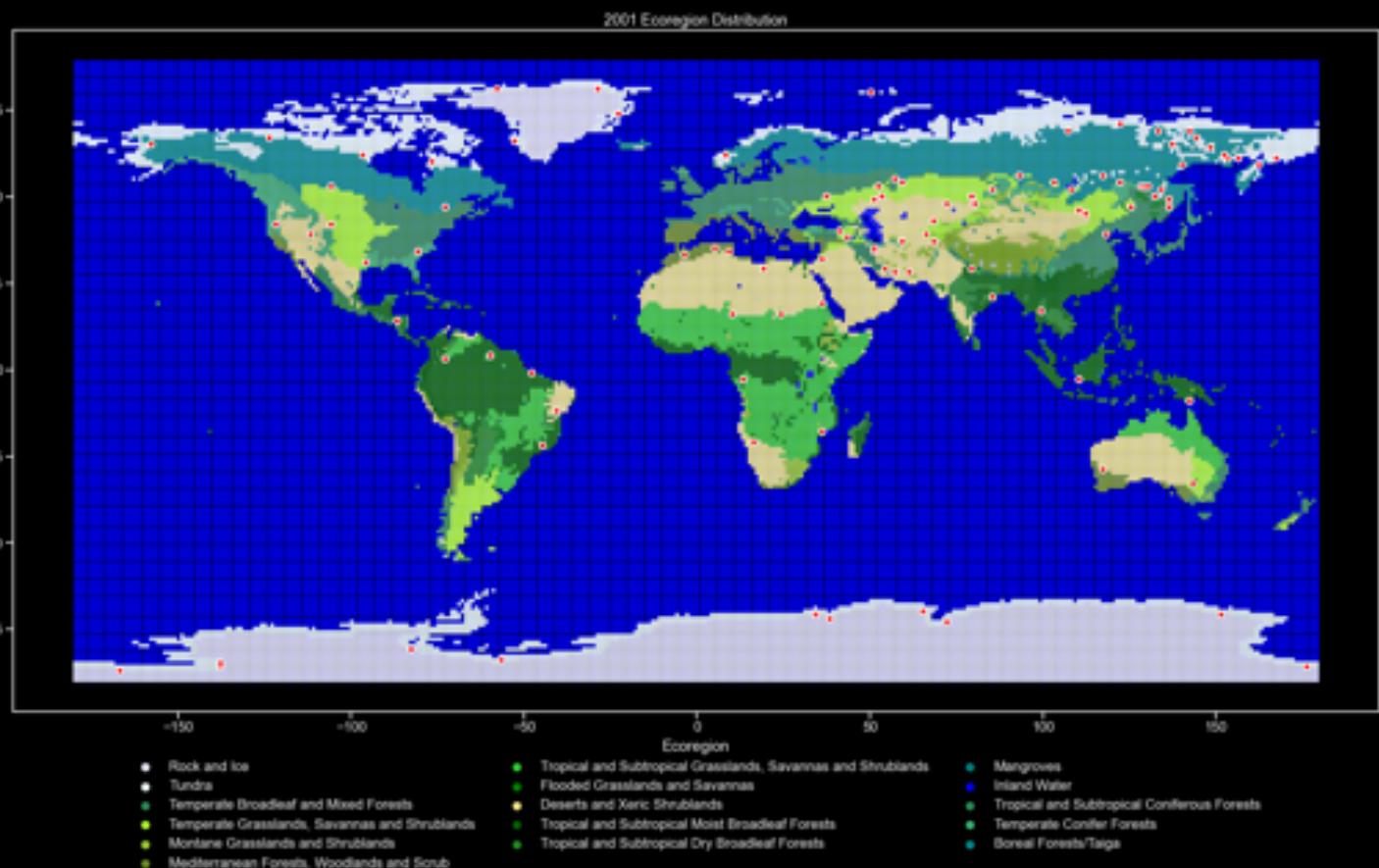


Figure 16: locations for Model A misclassifications

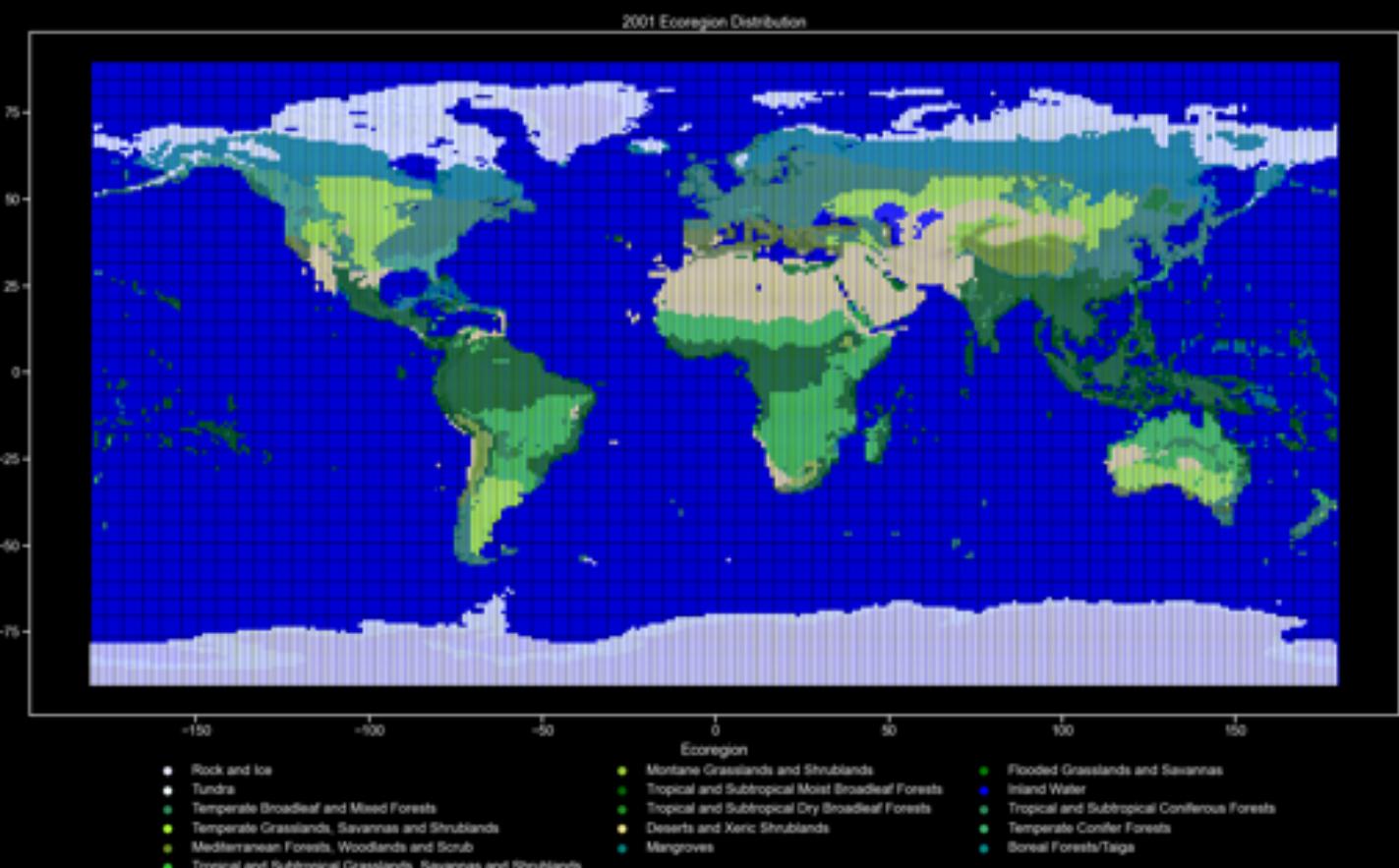


Figure 17: Model A applied to NCAR data

	Precision	Recall	F1-Score	Support
Boreal Forests/Taiga	0.93	0.96	0.95	9982
Deserts and Xeric Shrublands	0.94	0.97	0.95	10168
Flooded Grasslands and Savannas	0.97	0.52	0.68	407
Inland Water	0.99	0.79	0.88	203
Mangroves	1.00	0.52	0.68	108
Mediterranean Forests, Woodlands and Scrub	0.95	0.87	0.91	1265
Montane Grasslands and Shrublands	0.97	0.93	0.95	1958
Rock and Ice	0.99	0.99	0.99	22012
Temperate Broadleaf and Mixed Forests	0.93	0.94	0.94	5825
Temperate Conifer Forests	0.94	0.86	0.90	2027
Temperate Grasslands, Savannas and Shrublands	0.92	0.89	0.91	4307
Tropical and Subtropical Coniferous Forests	0.98	0.86	0.92	225
Tropical and Subtropical Dry Broadleaf Forests	0.96	0.82	0.88	1256
Tropical and Subtropical Grasslands, Savannas and Shrublands	0.92	0.95	0.94	6438
Tropical and Subtropical Moist Broadleaf Forests	0.94	0.97	0.96	6522
Tundra	0.95	0.94	0.94	11257

Table 1: Model A results

To address this significant imbalance in ecotype prevalence, the second model (**B**) implemented oversampling of the data such that every ecotype had equal representation in the training set. This was found to greatly improve the accuracy with the rarer ecoregions, resulting in an accuracy of 0.994. Because of this improvement, all further models utilized oversampling. The coordinates that were misclassified all fell on the boundary of ecoregions, with rock/ice and boreal forests being the most often confused (figure 18, table 2). Model **B** performed well on the NCAR data, though it noticeably underpredicted desert in Australia (figure 19). Low-prevalence ecoregions had much higher precision and recall with this model as a result of overfitting (table 3).

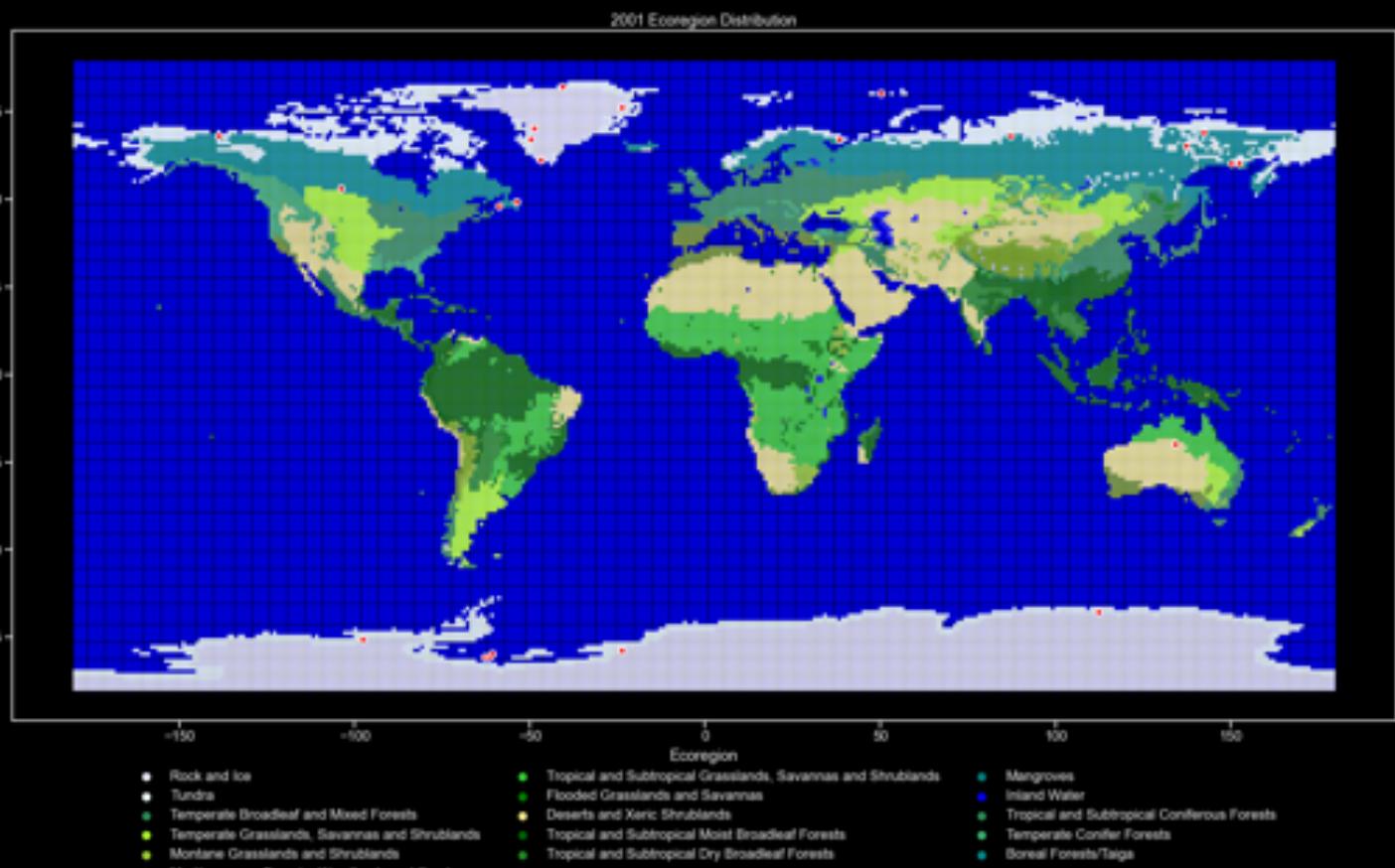


Figure 18: locations for Model B misclassifications

Observed	Predicted	
Rock and Ice	Tundra	8
Tundra	Tun Boreal Forests/Taiga	6
Tundra	Desert and Xeric Shrubland	3
Boreal Forests/Taiga	Temperate Broadleaf and Mixed Forests	2
Boreal Forests/Taiga	Temperate Grasslands, Savannas and Shrublands	1
Boreal Forests/Taiga	Tundra	1
Desert and Xeric Shrubland	Tropical and Subtropical Grasslands, Savannas and Shrublands	1
Tundra	Temperate Grasslands, Savannas and Shrublands	1

Table 2: Model B misclassifications

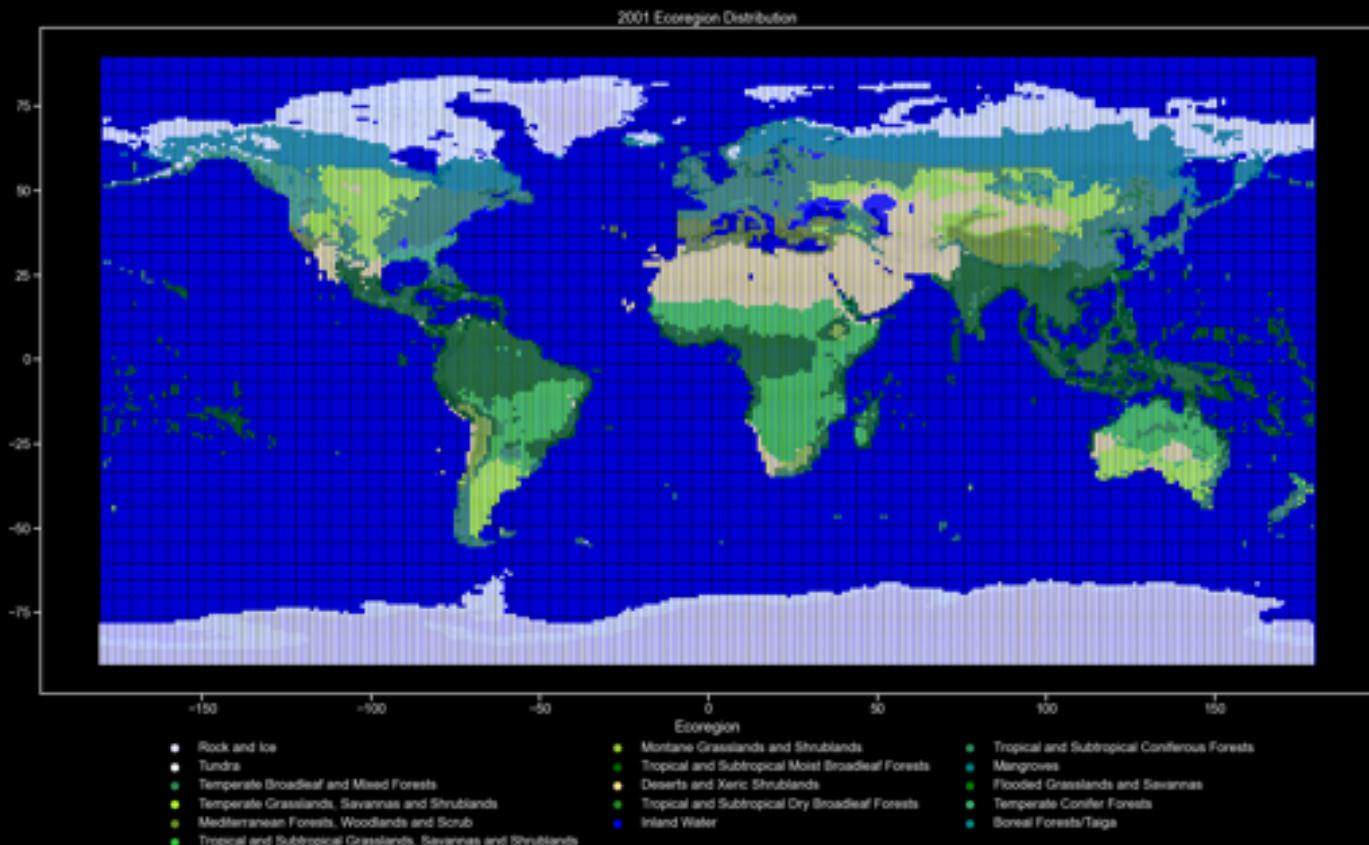


Figure 19: Model B applied to NCAR data

	Precision	Recall	F1-Score	Support
Boreal Forests/Taiga	0.98	0.98	0.98	21795
Deserts and Xeric Shrublands	1.00	0.99	0.99	21674
Flooded Grasslands and Savannas	1.00	1.00	1.00	21935
Inland Water	1.00	1.00	1.00	21718
Mangroves	1.00	1.00	1.00	21920
Mediterranean Forests, Woodlands and Scrub	1.00	1.00	1.00	21763
Montane Grasslands and Shrublands	1.00	1.00	1.00	21804
Rock and Ice	0.99	0.98	0.99	21683
Temperate Broadleaf and Mixed Forests	0.99	0.99	0.99	21868
Temperate Conifer Forests	1.00	1.00	1.00	21786
Temperate Grasslands, Savannas and Shrublands	0.99	0.99	0.99	21721
Tropical and Subtropical Coniferous Forests	1.00	1.00	1.00	21858
Tropical and Subtropical Dry Broadleaf Forests	1.00	1.00	1.00	21712
Tropical and Subtropical Grasslands, Savannas and Shrublands	0.99	1.00	0.99	21827
Tropical and Subtropical Moist Broadleaf Forests	1.00	1.00	1.00	21757
Tundra	0.97	0.97	0.97	21683

Table 3: Model B results

The top 14 features in terms of importance for this model were mean elevation, latitude, minimum elevation, maximum elevation, standard deviation of elevation, mean yearly temperature, mean temperature in May, mean temperature in April, temperature of coolest quarter, temperature of warmest quarter, temperature seasonality, minimum temperature, maximum temperature. Interestingly, elevation metrics composed 5 of the top 6 predictors, followed by various temperature values. The precipitation feature of most importance was annual precipitation, which was 16th overall. (figure 20)

The 14 features of highest importance were used to train the third random forest (**C**). Accuracy for this model was the highest so far at about 0.997 on the test set, though it did slightly worse than model **B** at predicting all of the 2001 data (figure 21, table 4). Model **C** had the opposite problem as the previous 2, instead overpredicting desert (figure 22).

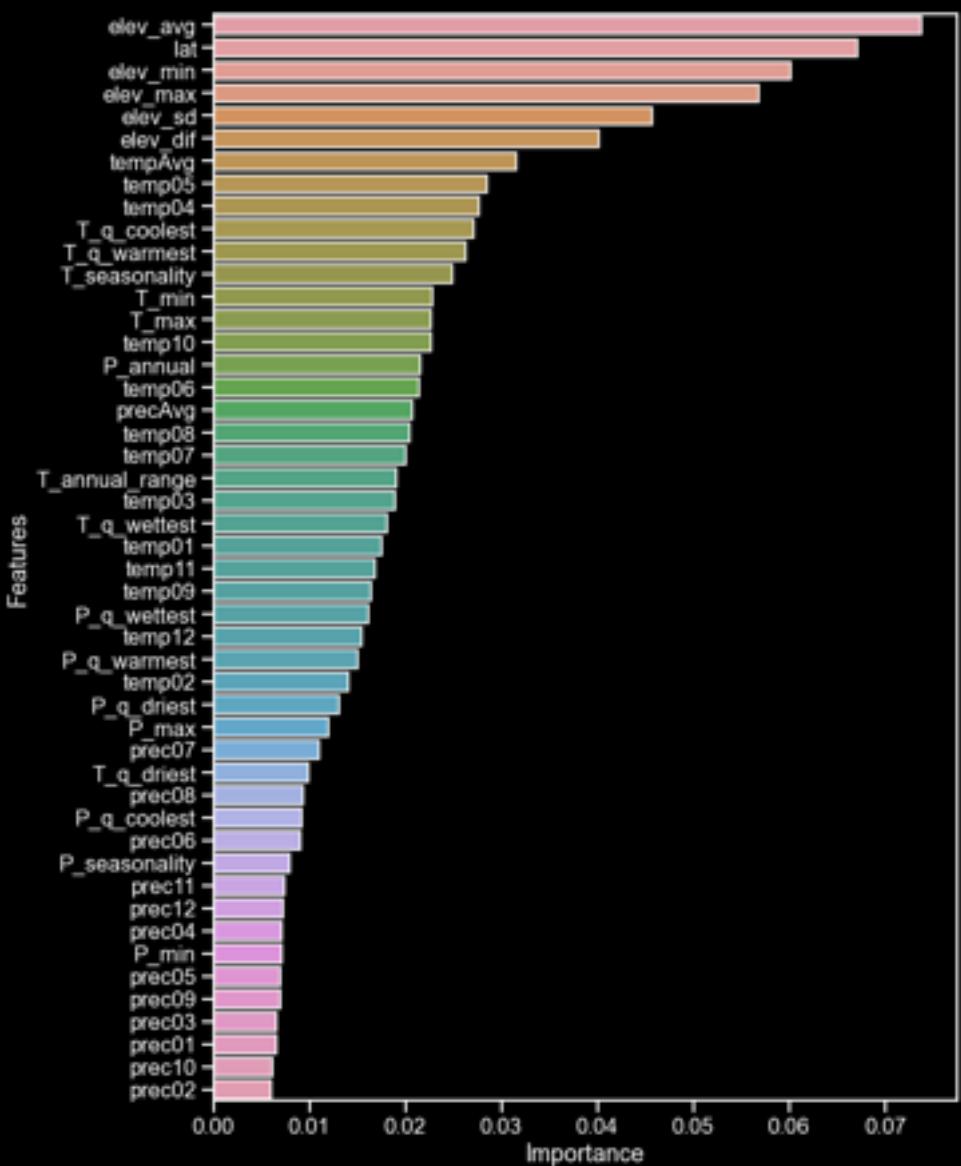


Figure 20: Model B feature importance.

Observed	Predicted	
Rock and Ice	Tundra	14
Tundra	Deserts and Xeric Shrubland	5
Tundra	Boreal Forests/Taiga	3
Boreal Forests/Taiga	Temperate Grasslands, Savannas and Shrublands	2
Tundra	Rock and Ice	1

Table 4: Model C misclassifications

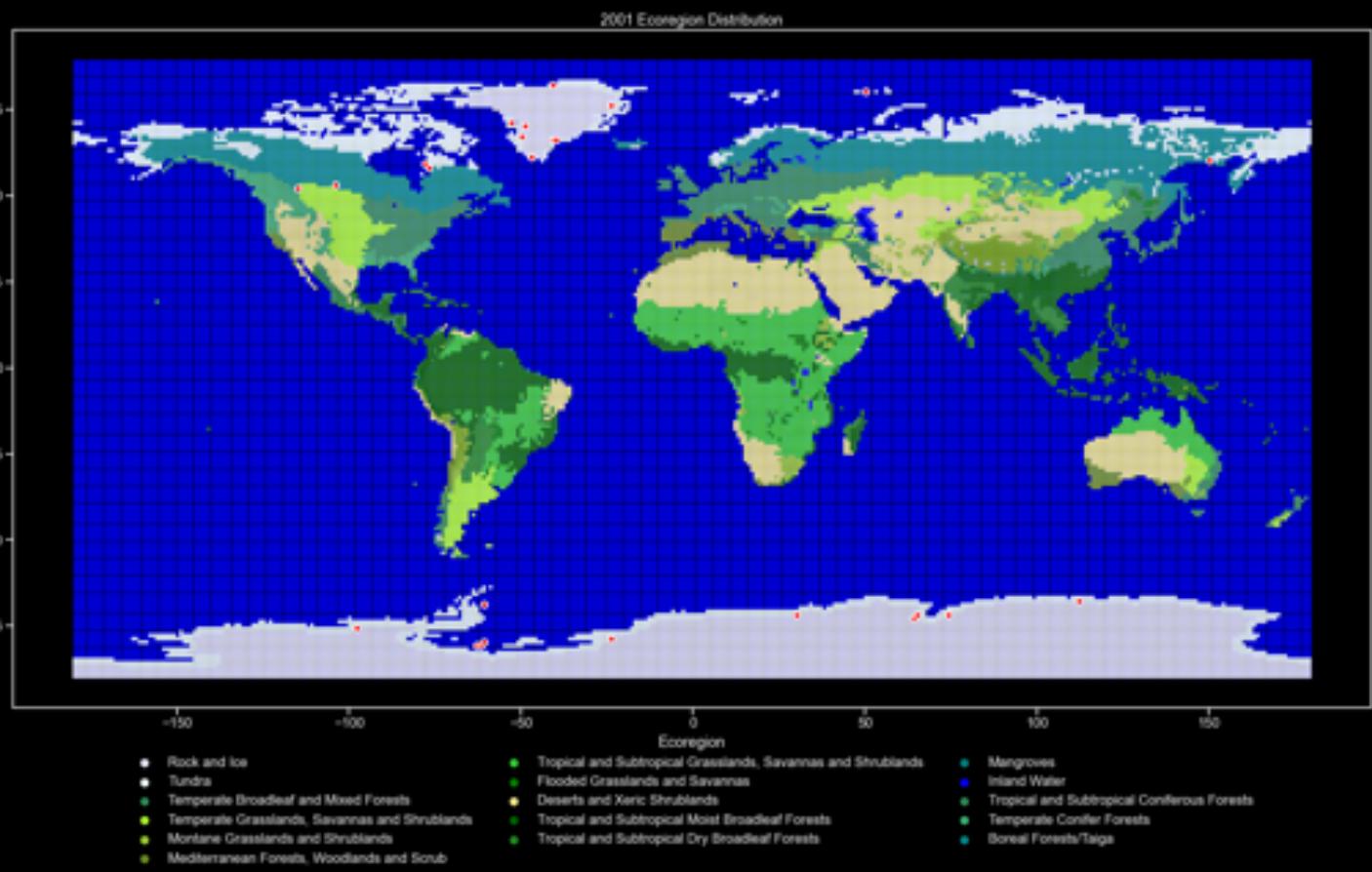


Figure 21: locations for Model C misclassifications

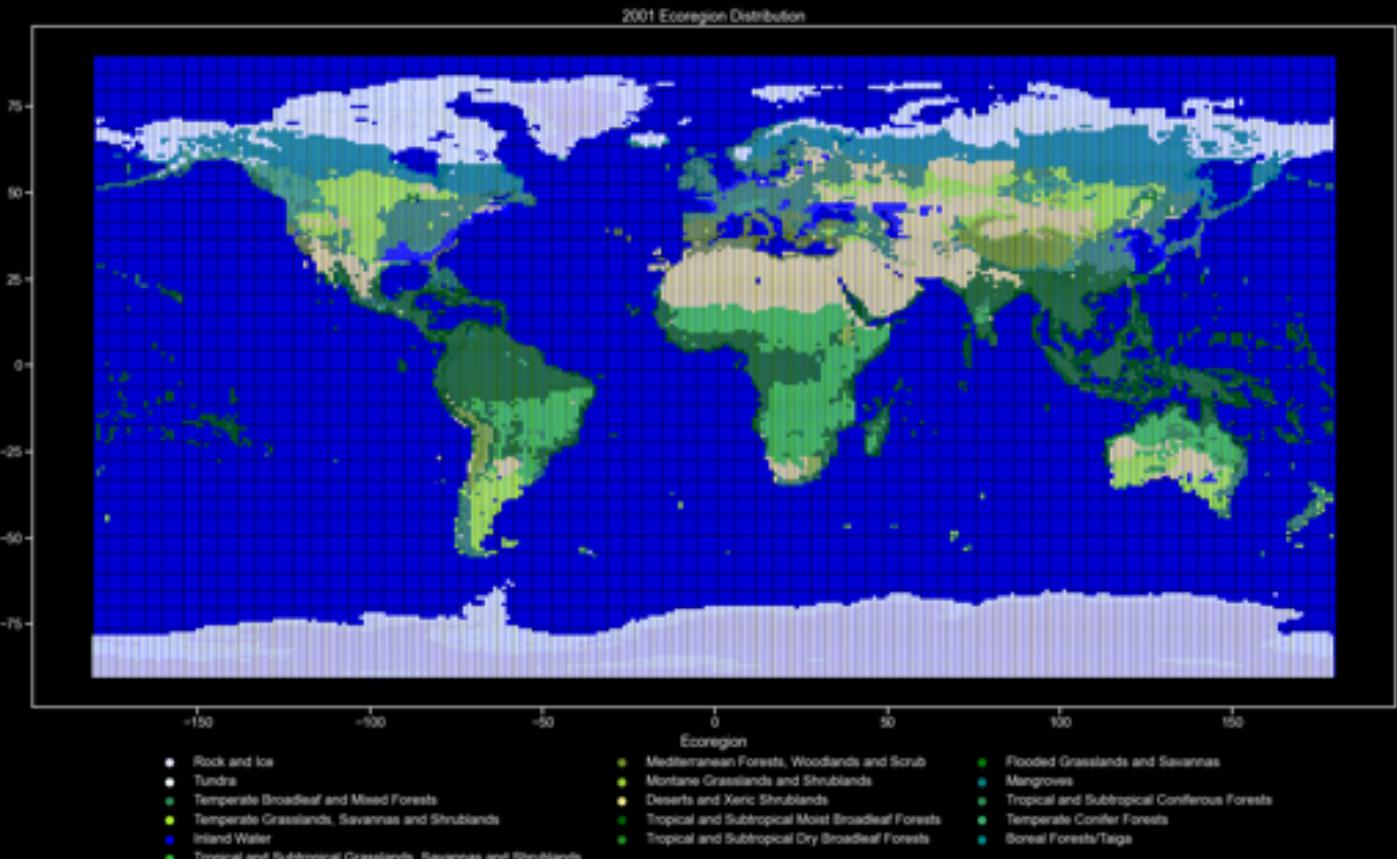


Figure 22: Model C applied to NCAR data

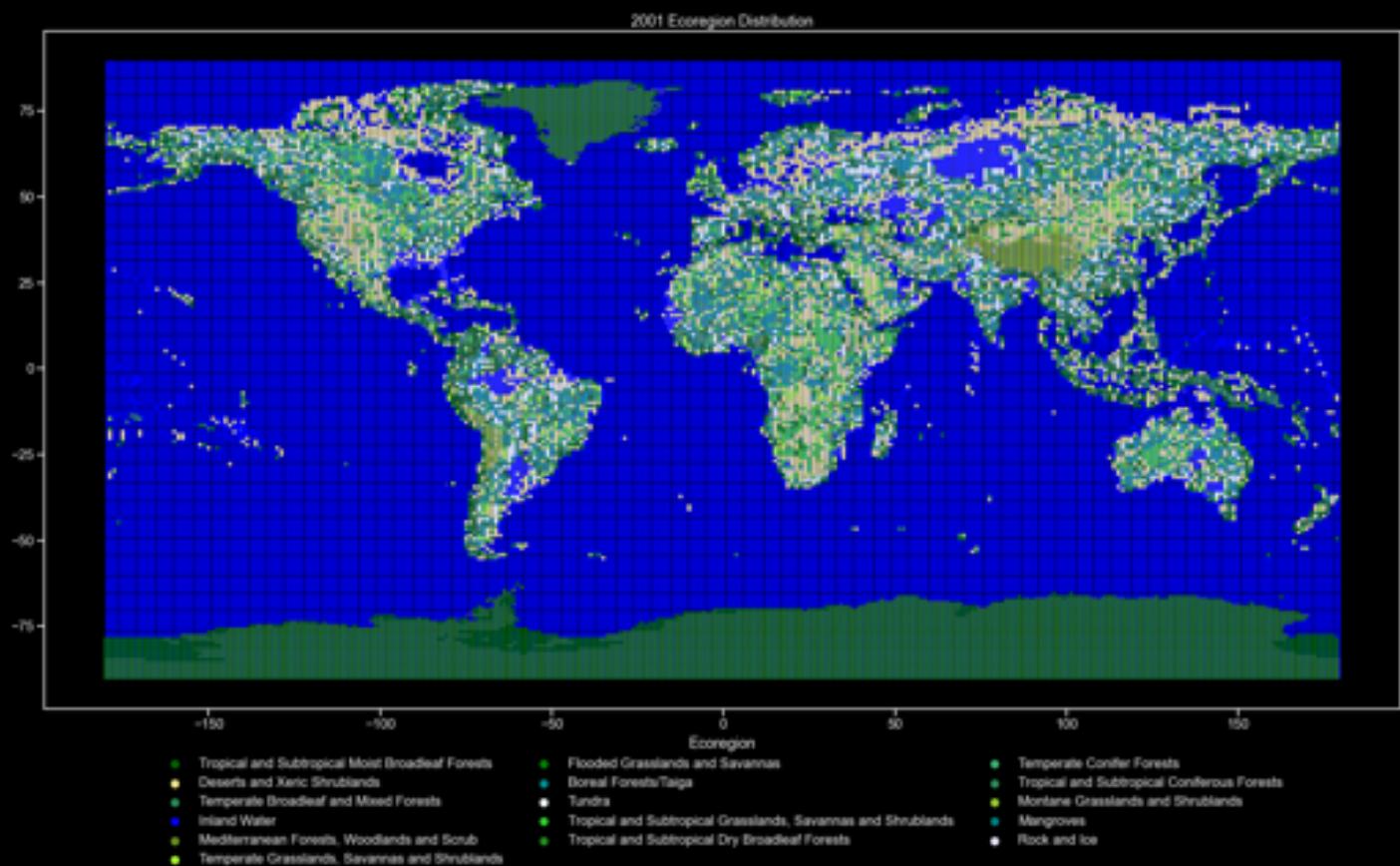


Figure 23: Model D applied to NCAR data

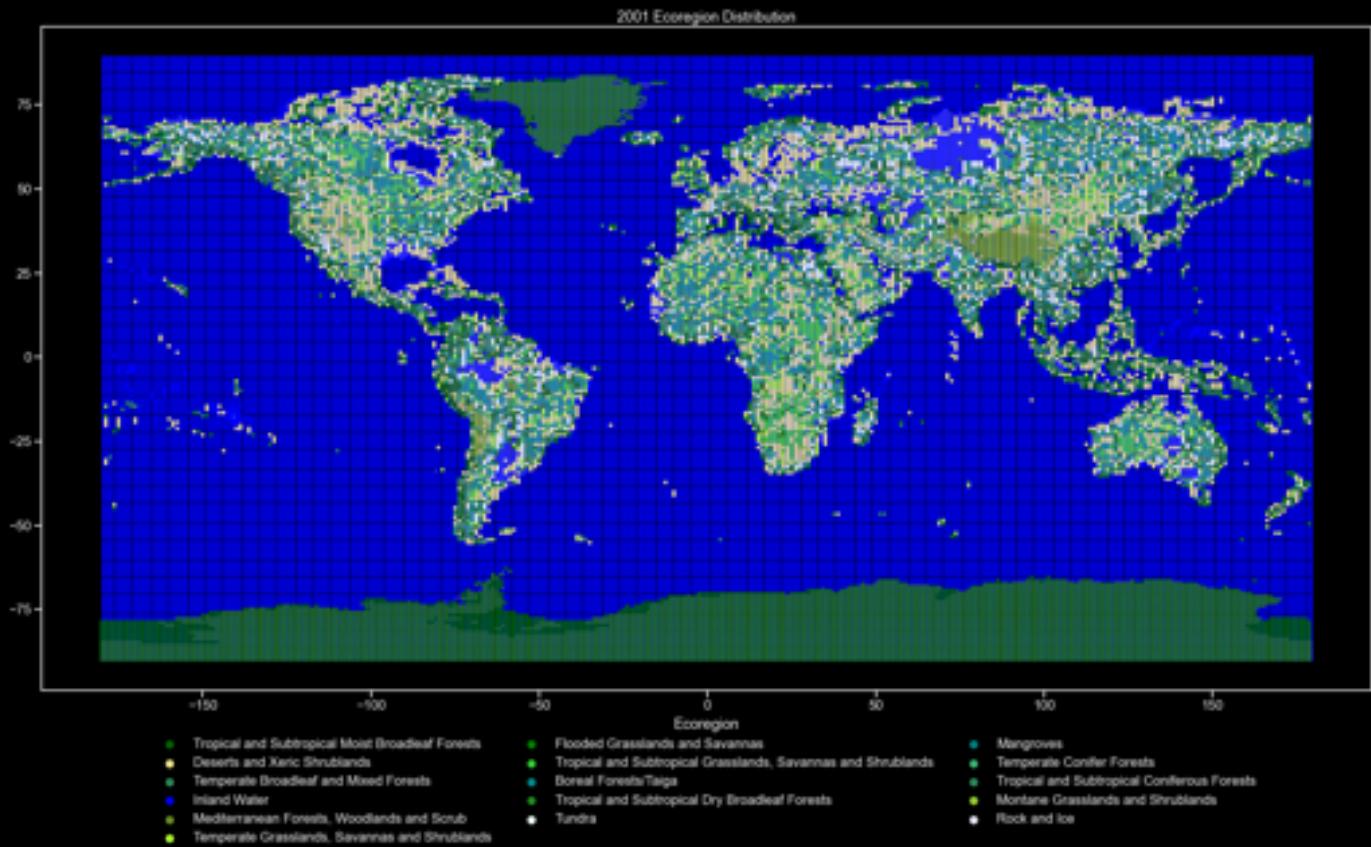


Figure 24: Model E applied to NCAR data

Two additional models using PCA were trained using a 0.99 threshold for variance. For both models, this reduced the dimensionality to 2. One model (**D**) reduced all available features, the second (**E**) reduced the 14 features of highest importance. Both models had accuracy upward of 0.99 on the test set, but behaved strangely with NCAR data, with a multitude of out-of-place single-coordinate patches predicted (*figures 23 and 24*).

Out of concern that data containing any kind of locality information would limit the model's ability to project ecoregions outside of their current distributions, two additional models were trained without using latitude as a feature. The first model (**F**) used all other features and the 14 most important features from this model were used to train the second (**G**). Model F an accuracy of about 0.992 on the test set with precision and recall comparable to model **B** (*figure 25, table 5*). Model F performed well on the NCAR data (*figure 26*).

	Precision	Recall	F1-Score	Support
Boreal Forests/Taiga	0.97	0.98	0.97	21795
Deserts and Xeric Shrublands	1.00	0.98	0.99	21674
Flooded Grasslands and Savannas	1.00	1.00	1.00	21935
Inland Water	1.00	1.00	1.00	21718
Mangroves	1.00	1.00	1.00	21920
Mediterranean Forests, Woodlands and Scrub	0.99	1.00	1.00	21763
Montane Grasslands and Shrublands	1.00	1.00	1.00	21804
Rock and Ice	0.99	0.98	0.99	21683
Temperate Broadleaf and Mixed Forests	0.99	0.99	0.99	21868
Temperate Conifer Forests	1.00	1.00	1.00	21786
Temperate Grasslands, Savannas and Shrublands	0.98	0.99	0.99	21721
Tropical and Subtropical Coniferous Forests	1.00	1.00	1.00	21858
Tropical and Subtropical Dry Broadleaf Forests	1.00	1.00	1.00	21712
Tropical and Subtropical Grasslands, Savannas and Shrublands	0.99	0.99	0.99	21827
Tropical and Subtropical Moist Broadleaf Forests	1.00	1.00	1.00	21757
Tundra	0.97	0.97	0.97	21683

Table 5: Model F results

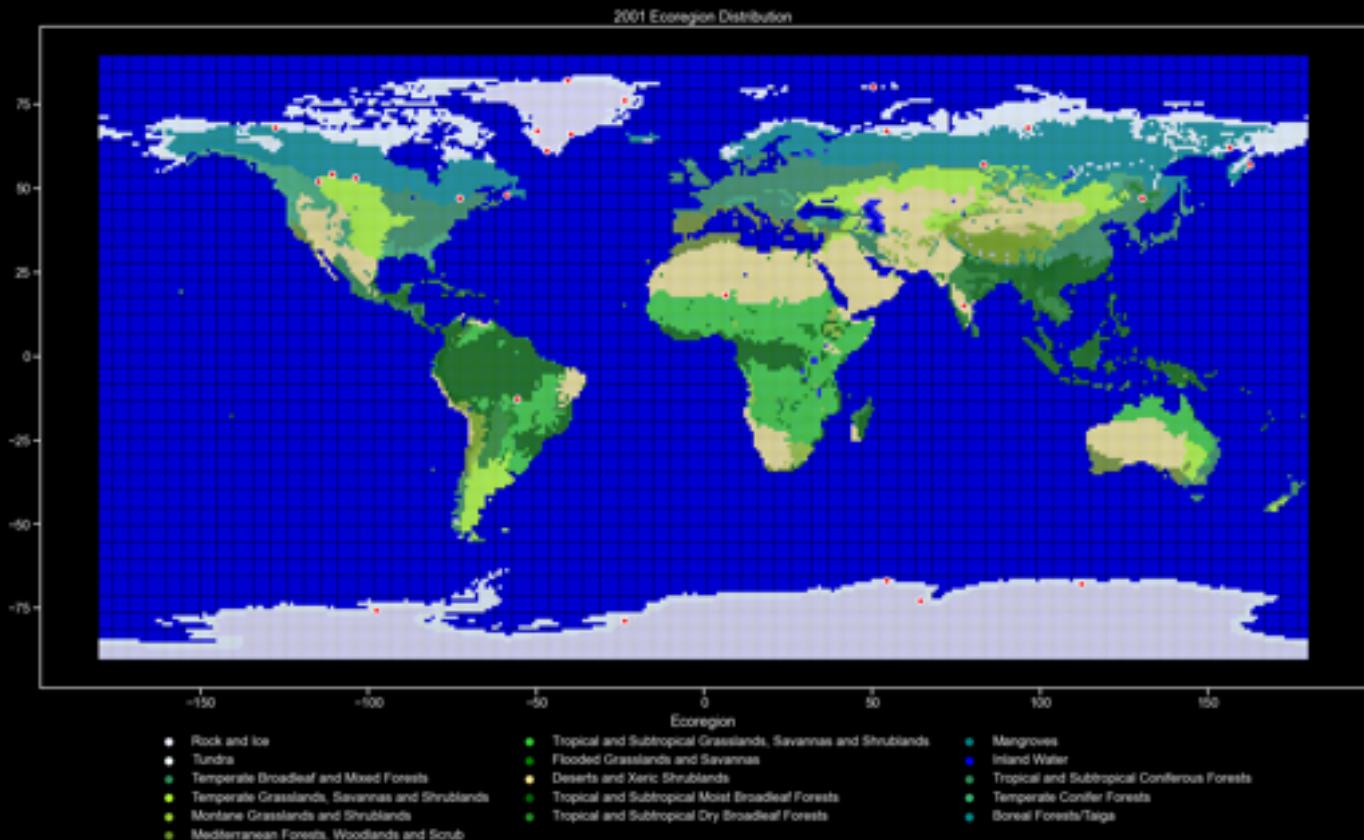


Figure 25: locations for Model F misclassifications

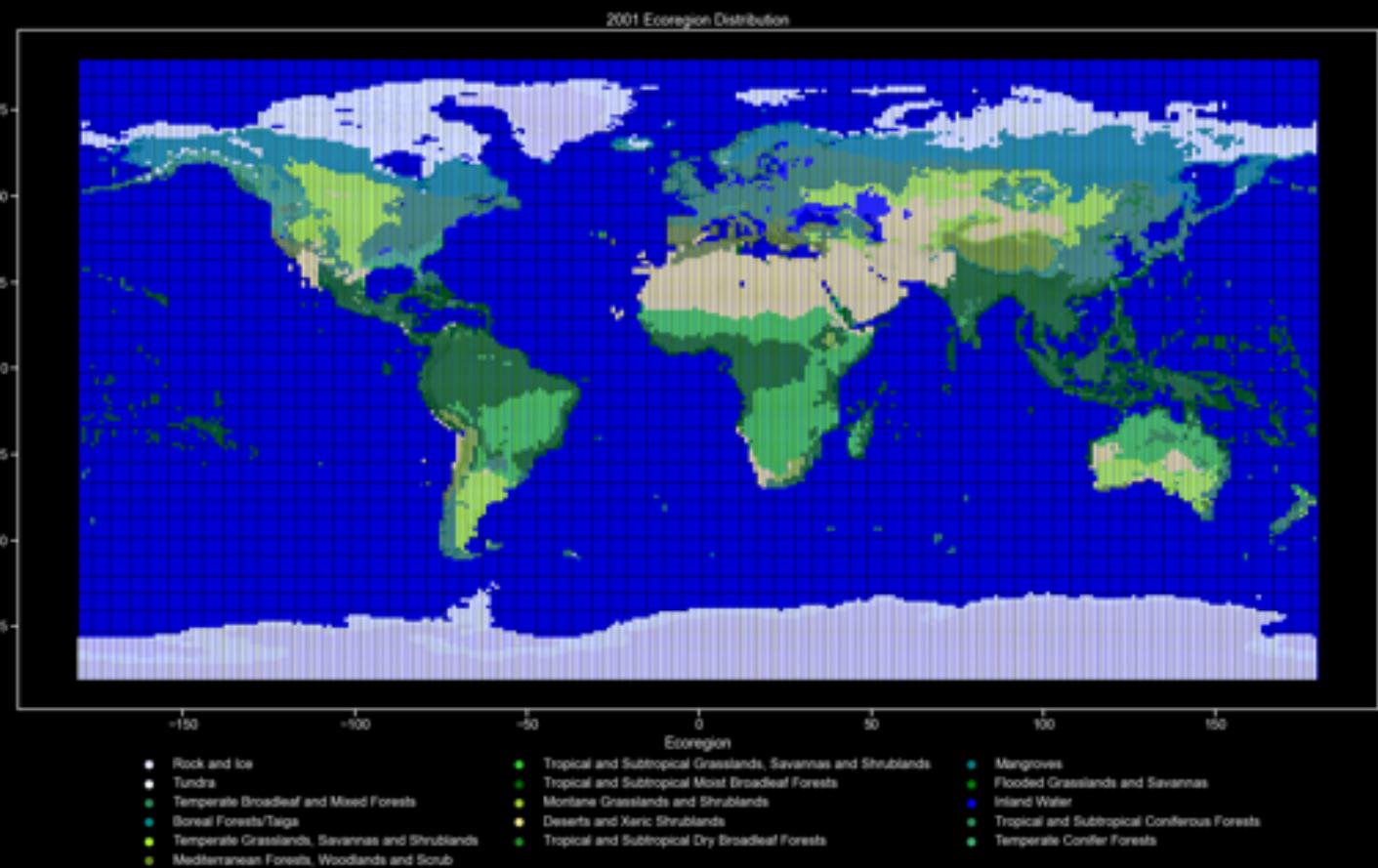


Figure 26: Model F applied to NCAR data

The top features from model F were similar to those of model B (*figure 27*): Using the 14 of these features with the highest importance to train Model G yielded an accuracy of about 0.994, with precision and recall roughly equivalent to model F (*figure 28, table 6*). Model G performed comparatively well on the NCAR data, albeit with some patchiness (*figure 29*).

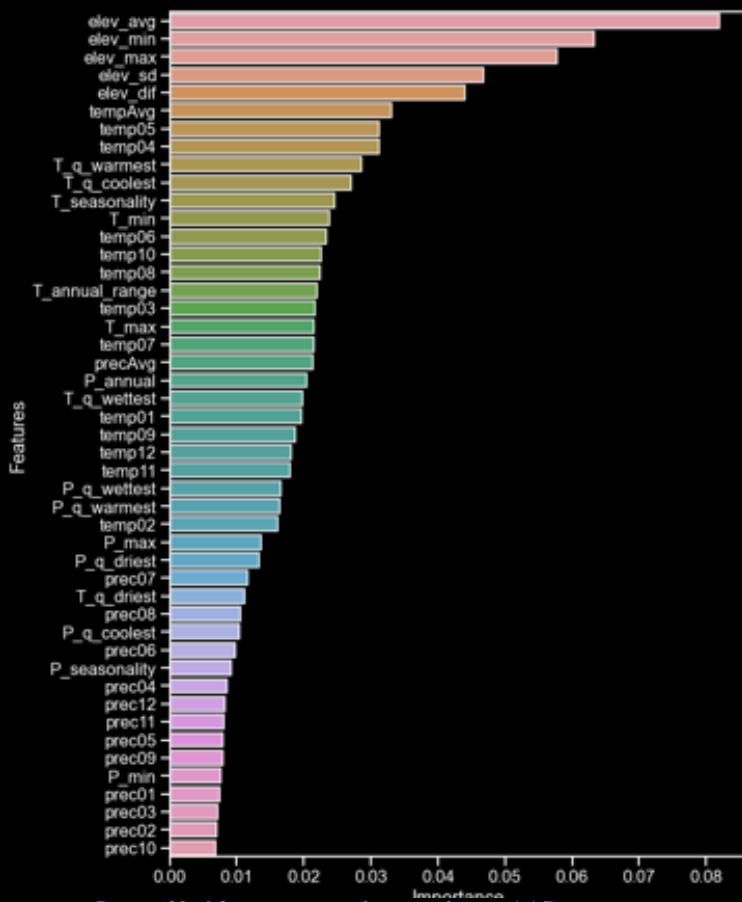


Figure 27: Most important features for model F

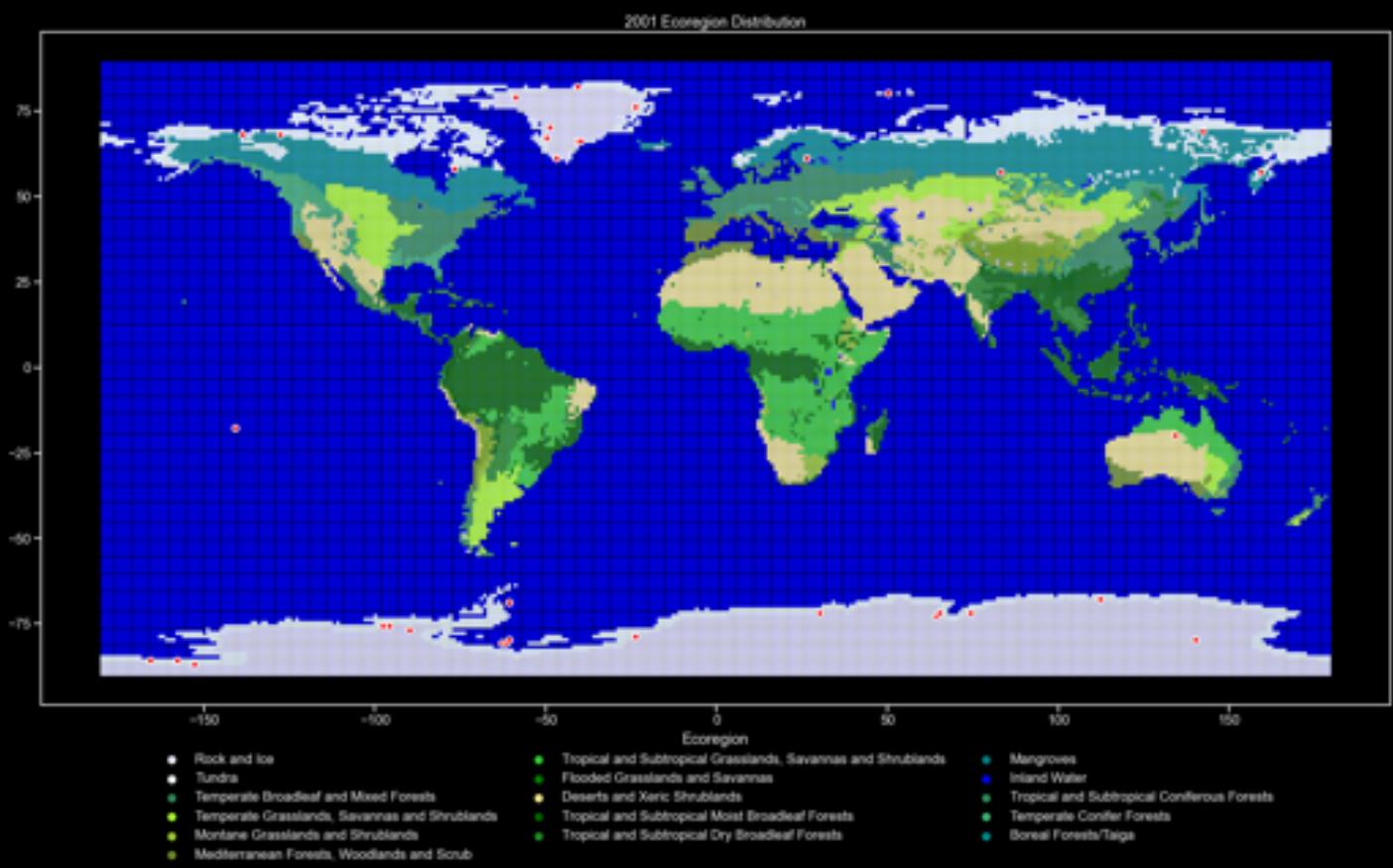


Figure 28: locations for Model G misclassifications

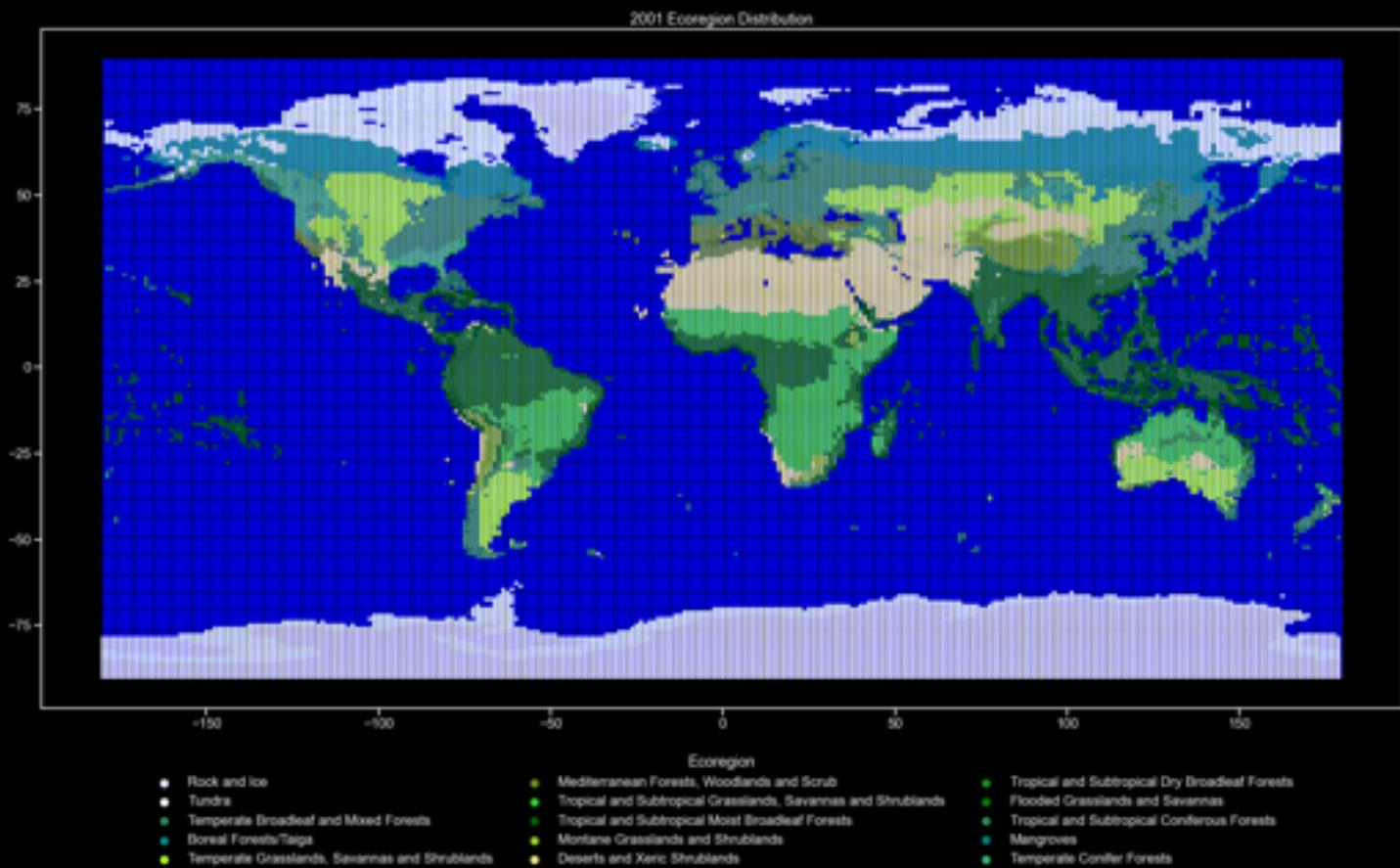


Figure 29: Model G applied to NCAR data

	Precision	Recall	F1-Score	Support
Boreal Forests/Taiga	0.99	0.99	0.99	21795
Deserts and Xeric Shrublands	1.00	0.99	1.00	21674
Flooded Grasslands and Savannas	1.00	1.00	1.00	21935
Inland Water	1.00	1.00	1.00	21718
Mangroves	1.00	1.00	1.00	21920
Mediterranean Forests, Woodlands and Scrub	1.00	1.00	1.00	21763
Montane Grasslands and Shrublands	1.00	1.00	1.00	21804
Rock and Ice	0.97	0.97	0.97	21683
Temperate Broadleaf and Mixed Forests	0.99	0.99	0.99	21868
Temperate Conifer Forests	1.00	1.00	1.00	21786
Temperate Grasslands, Savannas and Shrublands	0.99	1.00	1.00	21721
Tropical and Subtropical Coniferous Forests	1.00	1.00	1.00	21858
Tropical and Subtropical Dry Broadleaf Forests	1.00	1.00	1.00	21712
Tropical and Subtropical Grasslands, Savannas and Shrublands	1.00	1.00	1.00	21827
Tropical and Subtropical Moist Broadleaf Forests	1.00	1.00	1.00	21757
Tundra	0.97	0.96	0.97	21683

Table 6: Model G results

All the models save those utilizing PCA or not using oversampling performed comparatively well, with models **C** (trained on top 14 features, including latitude) and **F** (trained on all features excluding latitude) having slightly higher accuracies with the test set and good performance on the NCAR data. These two models were chosen for projecting ecoregion distribution. In the initial projections, these models had issues overpredicting inland water. Since inland water in the dataset was infrequent and inconsistent with actual distributions (there are many inland water areas that were not labelled as such), it was decided retrain models **C** and **F** on the data with inland all inland water occurrences dropped before proceeding with projections.

Projections

Models **C** and **F** were applied to the NCAR projections to create forecasts of ecoregion distributions through 2100. Two rounds of projections were made, the first relied solely on model predictions and the second relabeled inundated land as new ocean to reflect rising sea levels. To implement this, it was necessary to differentiate coastal coordinates from inland points that lie below sea level, the later to remain terrestrial. For each year, coordinates whose maximum elevation had newly decreased to less than 0 were evaluated by checking surrounding points their type. If all surrounding points were terrestrial, no change was made to the projected label. If any of the surrounding points was oceanic, the label was updated as new ocean.

For the most part this worked well, though areas surrounding the Caspian Sea were projected to be submerged. Thermal expansion of water is one of the driving forces of sea level rise, so this is indeed a possibility. But sea level has been steadily declining in the Caspian for years and is projected to continue to fall due to increasing temperatures causing evaporation to exceed the amount of water draining into the sea.

Projections using model **F** lacked occurrences for several ecoregions and predicted the northern hemisphere be dominated by temperate broadleaf and mixed forests by the year 2010 under the RCP2.6 scenario (*figures 30 and 31*). This is demonstrably false, putting the rest of the projections of this model into doubt. All further analysis was be based solely off of the projections from model **C**.

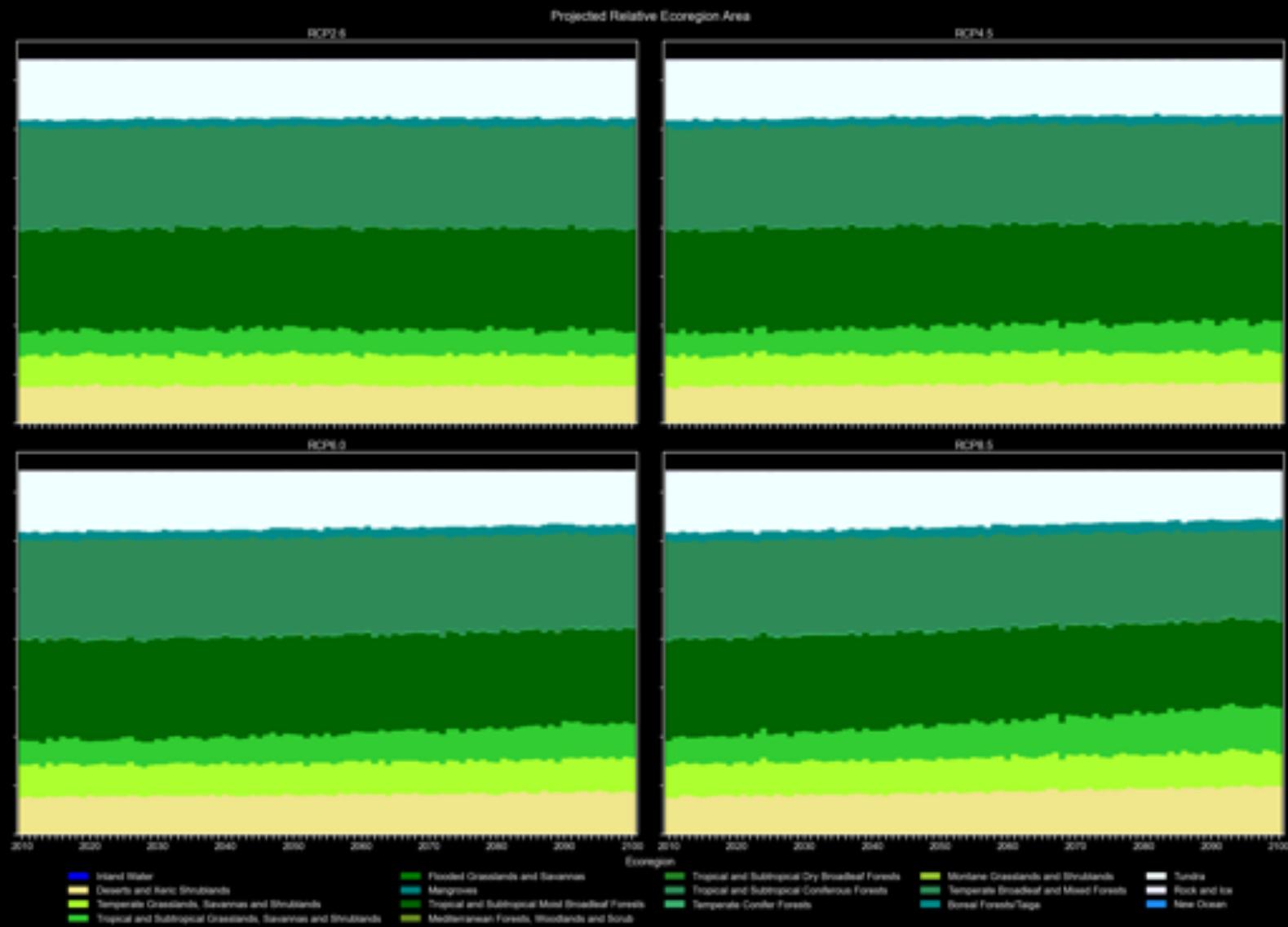


Figure 30: Ecotype prevalence projections using model F

Model C gave much more reasonable results. Ecoregions distribution in 2010 is similar to that of 2001 (*figure 32*). This model projects the area of deserts to increase throughout the century, especially under higher greenhouse gas scenarios. Sea level rise also accounts for significant loss of terrestrial ecotype area (*figure 33*).

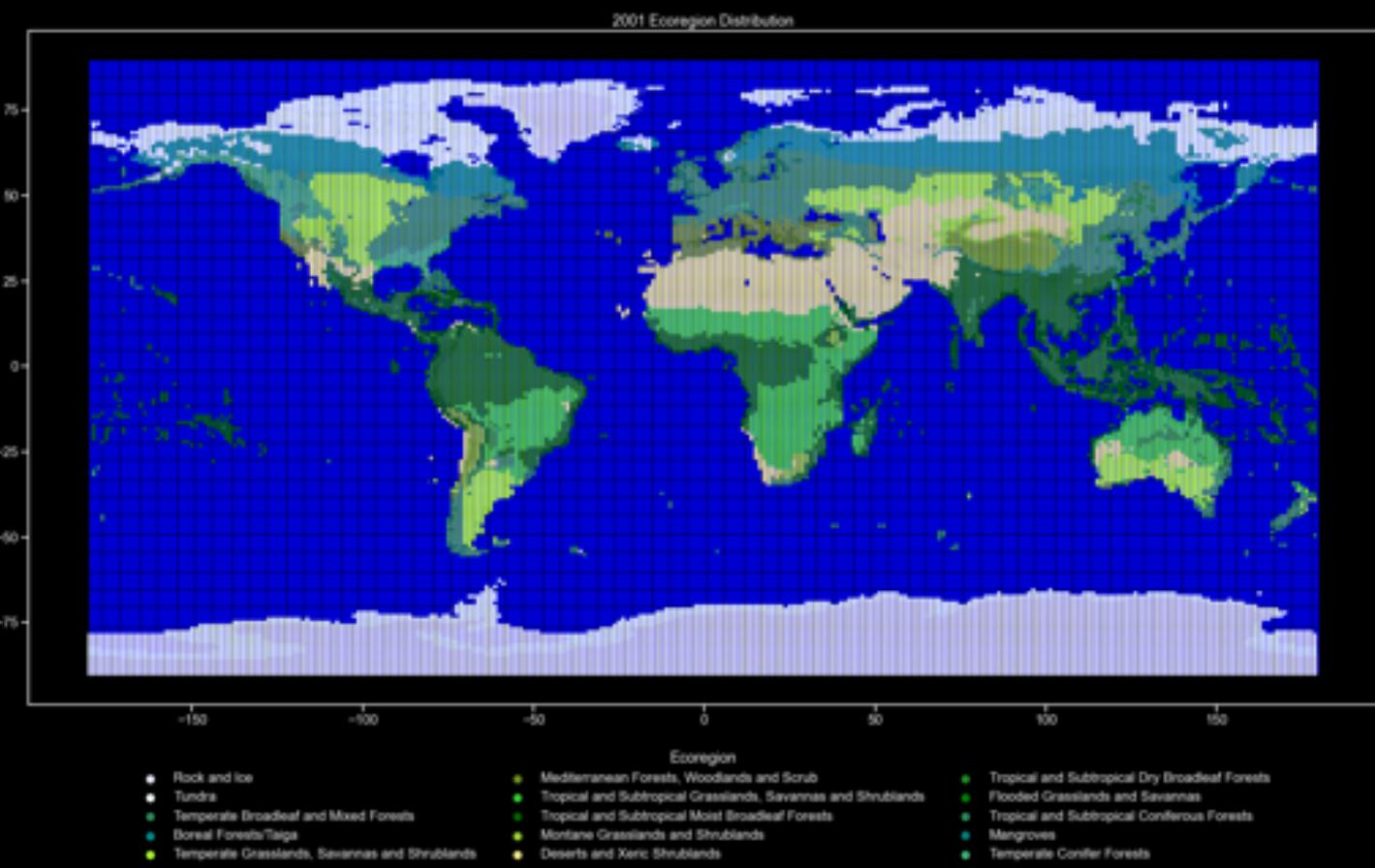


Figure 31: Ecotype distribution projections for 2010 using model F

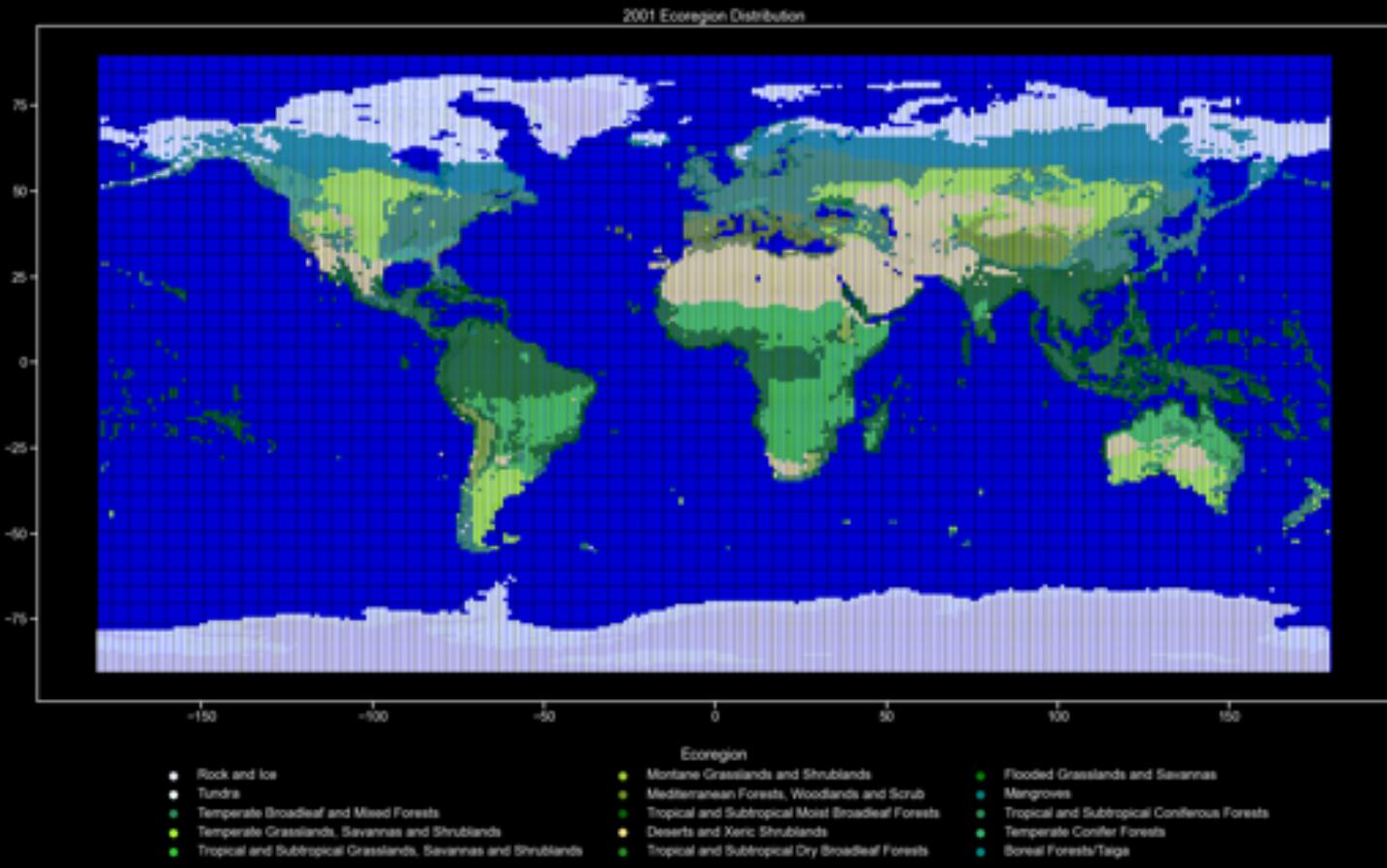
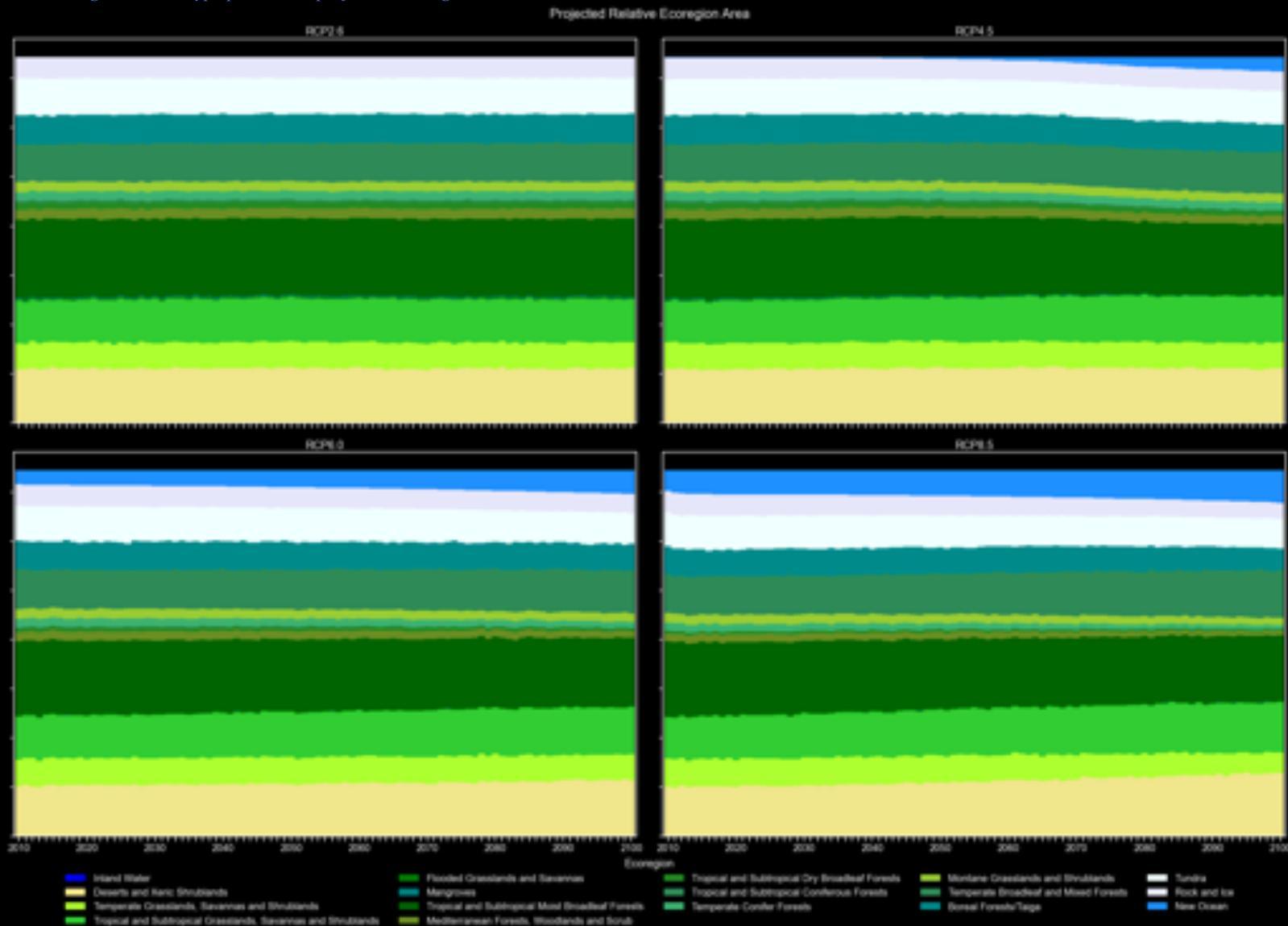


Figure 32: Ecotype distribution projections for 2010 using model C.

Figure 33: Ecotype prevalence projections using model C



Discussion

When compared the neural network used by Hilbert et al. to predict Australian ecoregion change, model C yielded similar results, both predicting an expansion of desert over time (*figures 31, 32, 33*). That is except for the RCP8.5 scenario, which interestingly forecasted an expansion of tropical and subtropical grassland (*figure 34*).

While direct comparison is difficult between these models and the methodology used by Hoffman et al. to assess protected regions exposure to novel climate, similar trends were noted. Their research found that the top-most impacted areas are temperate ecotypes. Likewise, temperate ecotypes in this analysis are all projected to decrease in area, especially in high concentration scenarios. Furthermore, many areas found to have the highest local-scale novel climate index coincided with change in ecotype in the random forest projections, including much of eastern North America, Northern India, and Central Europe into Asia. These are in line with the findings of Yu et al. assessment of high-risk ecoregions as well.

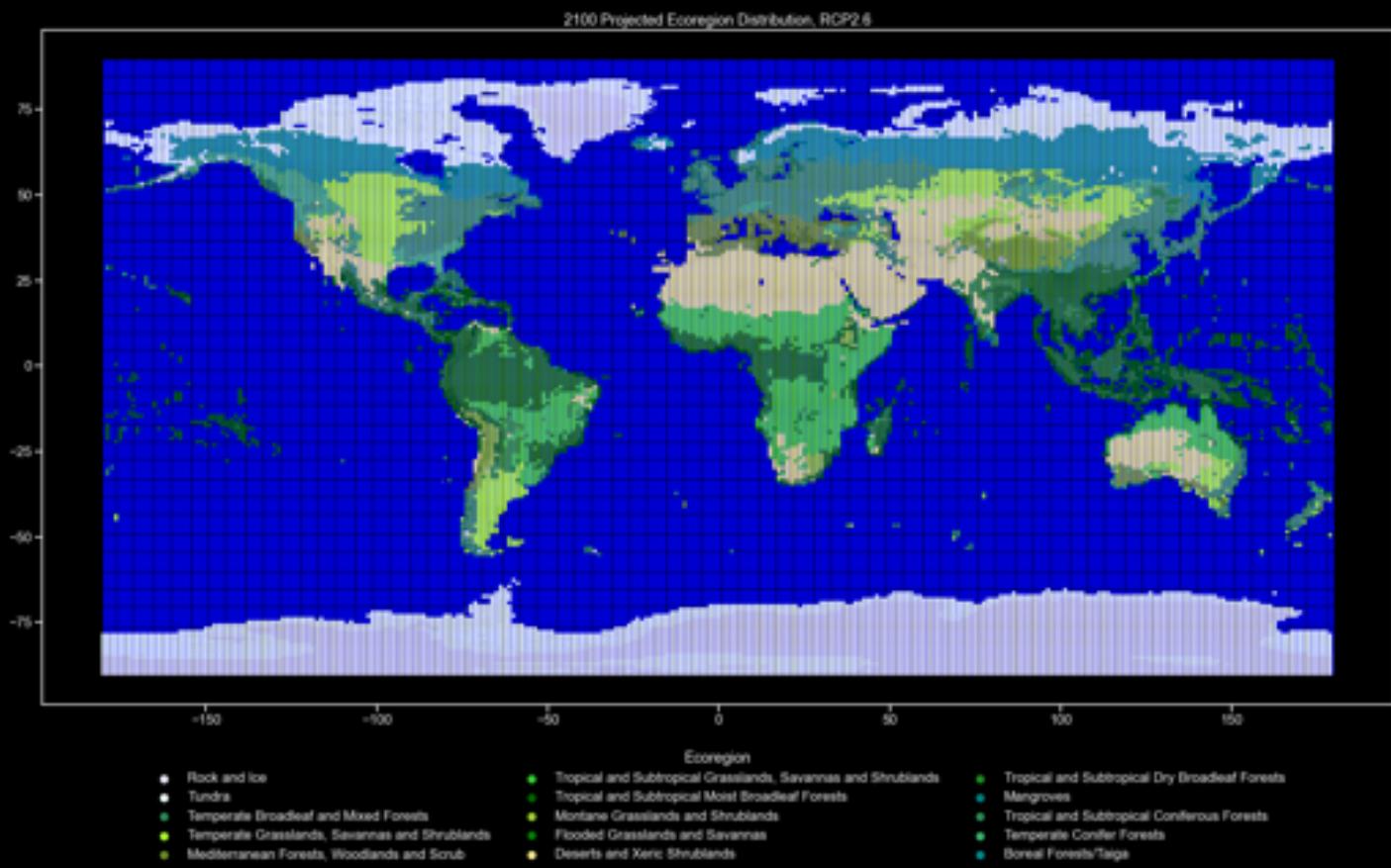


Figure 31: Projection for 2100 using Model C under RCP2.6 scenario

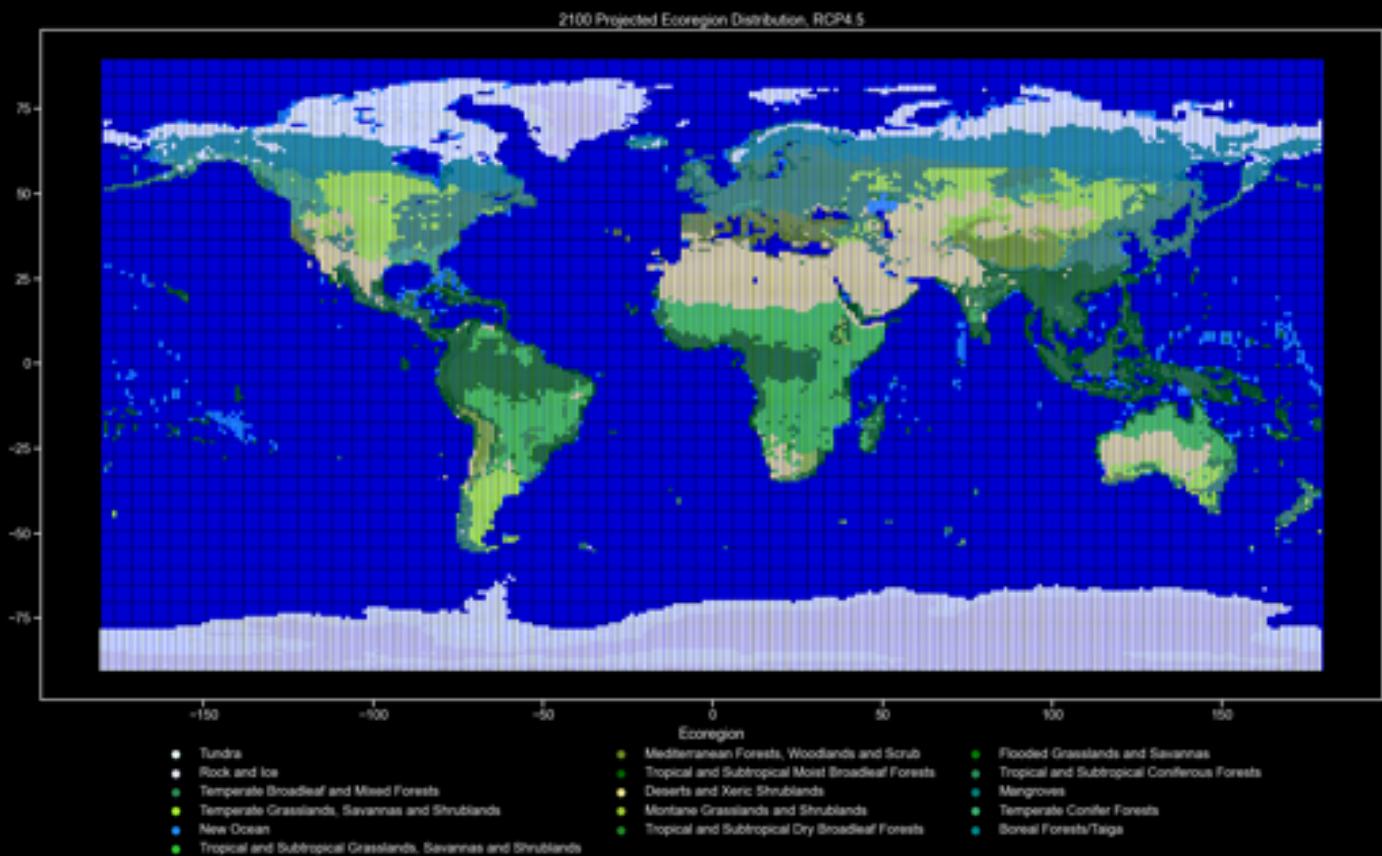


Figure 32: Projection for 2100 using Model C under RCP4.5 scenario

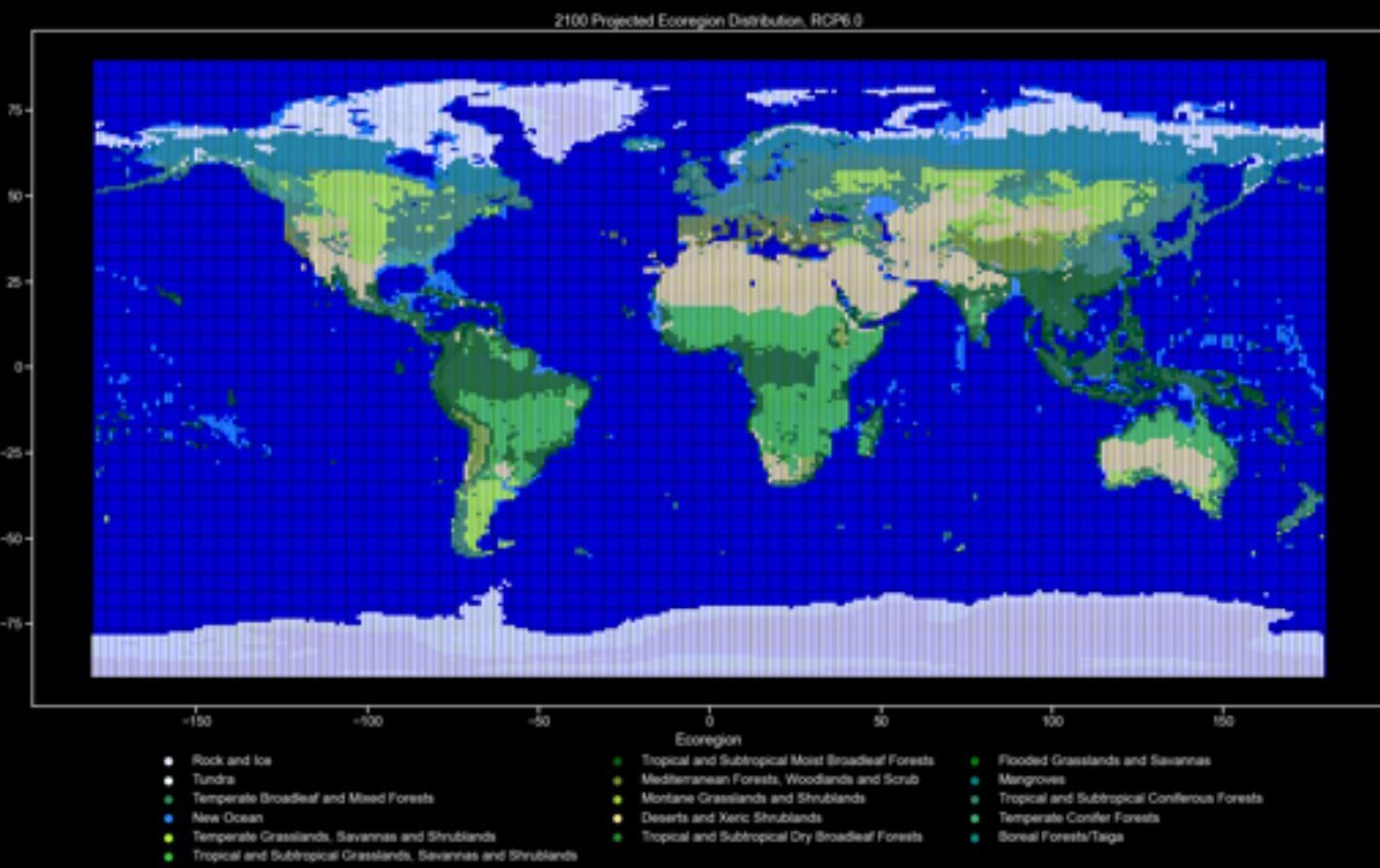


Figure 33: Projection for 2100 using Model C under RCP6.0 scenario.

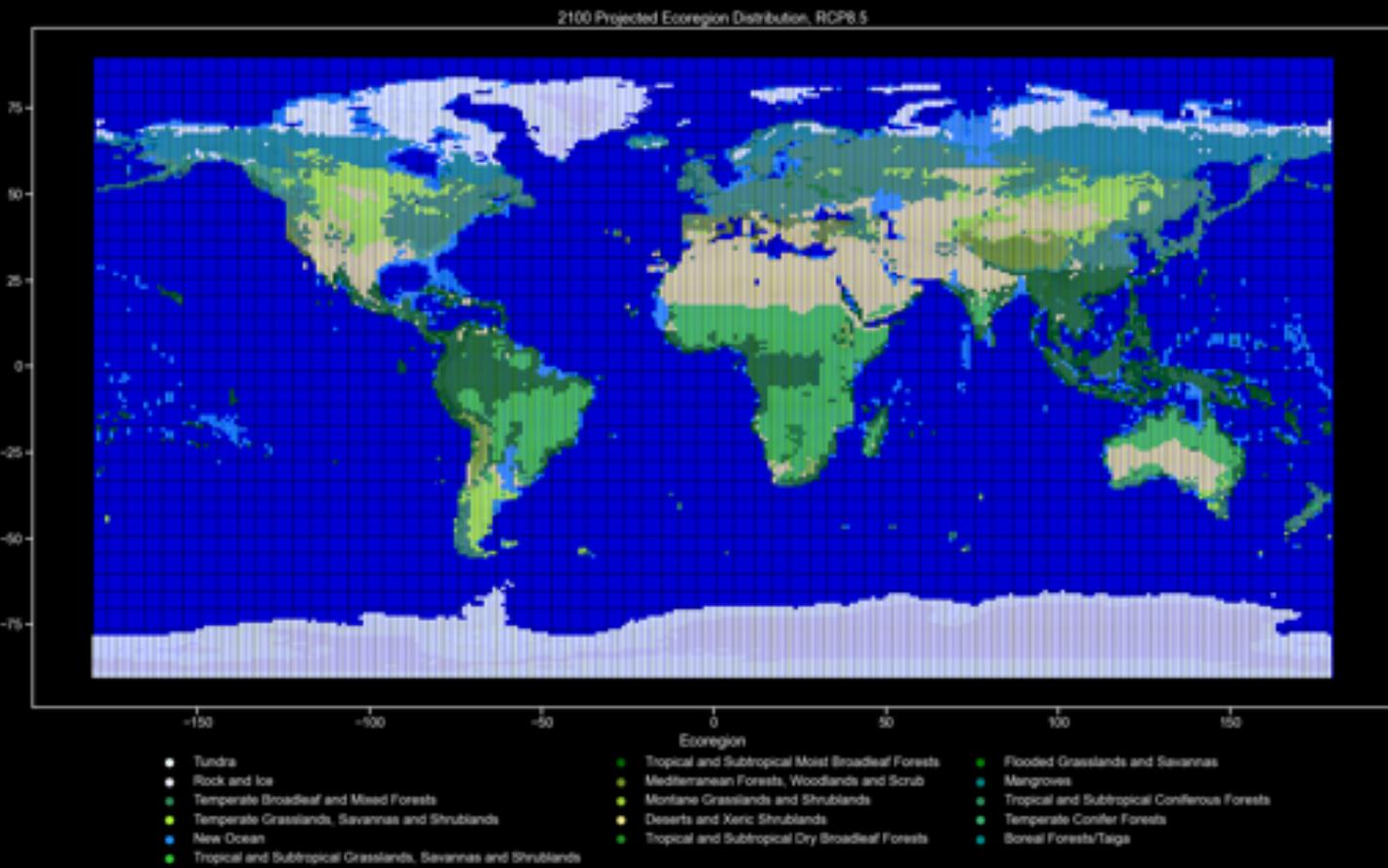


Figure 34: Projection for 2100 using Model C under RCP8.5 scenario.

In these scenarios, the deleterious impact sea level rise will have on some regions is apparent, with many low-lying regions becoming submerged in higher concentration scenarios. These regions include many pacific islands, Florida and the gulf coast, the Río de la Plata drainage basin, and Northern continental Europe, and the tributaries of the Amazon and Senegal Rivers.

Looking at each ecoregion individually gives a better insight into how each will be impacted by climate change. Under the RCP2.6, areas remain relatively unchanged throughout the century. Changes can be seen in ecoregion area under higher greenhouse gas scenarios. Most regions are predicted to decrease to varying degrees, especially under scenario RCP8.5. Deserts/Xeric shrublands and tropical/subtropical grasslands both increase in area.

Further steps could be taken to improve the model. This includes the utilization of annual mean diurnal range, isothermality and other possible features in training the models. Admittedly, the cutoff for the top 14 features was arbitrary, and a best subsets search method could be used to find the combination and number of features that optimize the model. Projections could also be improved by implementing a different model, such as a neural network.

Taken all together, these trends suggest the significant impact climate change will have on Earth's biota, especially if gone unabated. The impact on humans cannot be underestimated, as a shifting climate will effect where and when we can grow food, increase our exposure to inclement weather events, and increased wildfire prevalence and extent, just to name a few. Concerted global efforts are imperative in mitigating these effects and inaction on our part has the potential to do incalculable harm to this planet's biodiversity and challenge our ability to adapt to a world out of equilibrium.

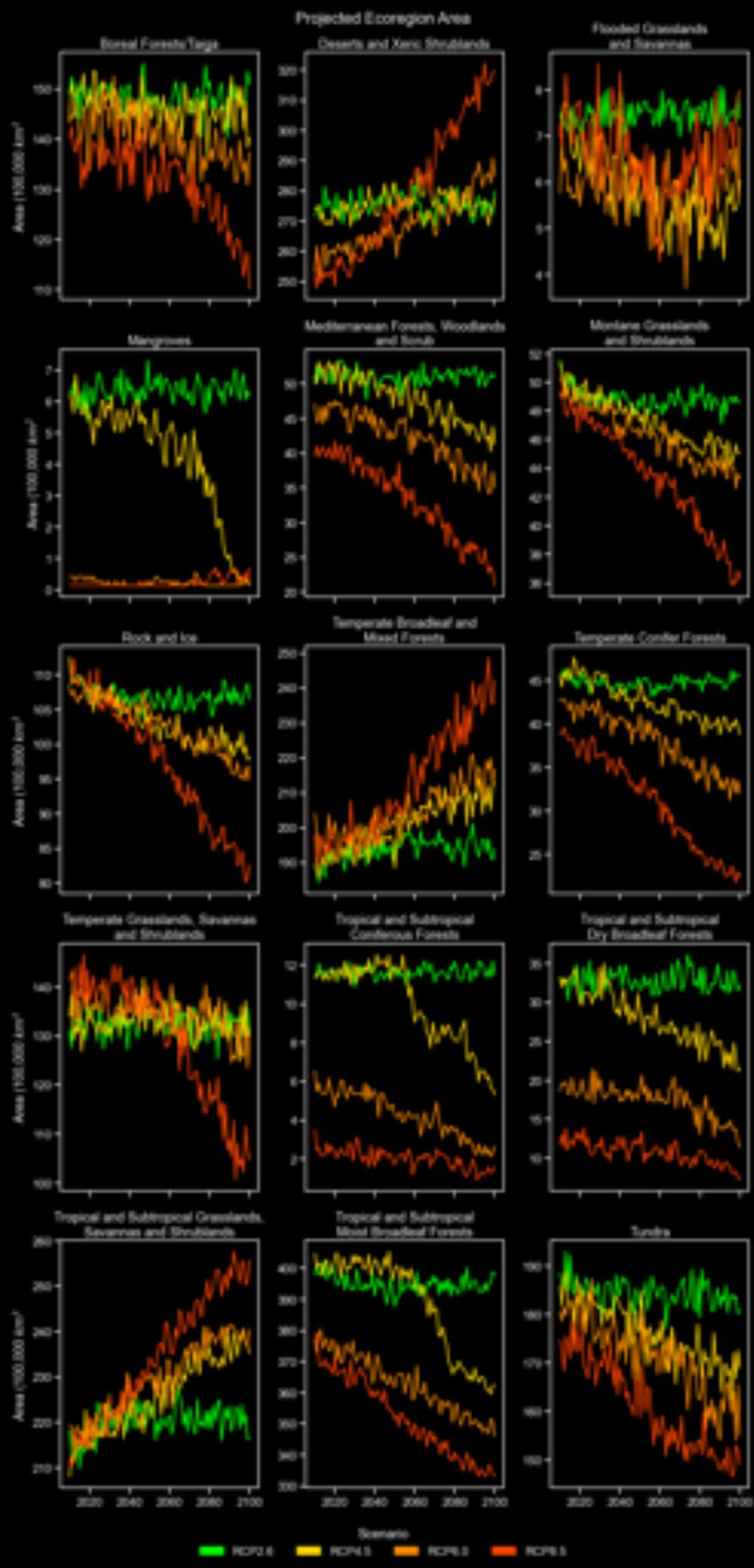


Figure 31: Projected change in ecoregion area

Data Sources

Climate Econometrics. “Sea-Level Projections [RCP Scenarios].” Climate Econometrics, 30 Apr. 2018, www.climateeconometrics.org/sealevel_rcp.

NCAR. “Download NCAR Community Climate System Model (CCSM) Projections in GIS Formats | GIS Climate Change Scenarios.” GIS Climate Change Scenarios, <https://gisclimatechange.ucar.edu/gis-climatedata>.

NOAA. “CPC Merged Analysis of Precipitation (CMAP): NOAA Physical Sciences Laboratory.” NOAA, 1997, <https://psl.noaa.gov/data/gridded/data.cmap.html>.

NOAA. “GHCN_CAMS Gridded 2m Temperature: NOAA Physical Sciences Laboratory.” NOAA, Oct. 2015, <https://psl.noaa.gov/data/gridded/data.ghcncams.html>.

The Nature Conservancy. “Conservation GIS Data - The Nature Conservancy.” The Nature Conservancy, 2002, http://maps.tnc.org/gis_data.html.

USGS. “USGS EROS Archive - Digital Elevation - Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010).” USGS, 201AD, https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation?qt-science_center_objects=0#qt-science_center_objects.

References

Baena, Susana, et al. “Earth Observation Archives for Plant Conservation: 50 Years Monitoring of Itigi-Sumbu Thicket.” *Remote Sensing in Ecology and Conservation*, edited by Harini Nagendra and Ned Horning, vol. 2, no. 2, 2016, pp. 95–106. Crossref, doi:10.1002/rse2.18.

Ceballos, Gerardo, et al. “Vertebrates on the Brink as Indicators of Biological Annihilation and the Sixth Mass Extinction.” *Proceedings of the National Academy of Sciences*, vol. 117, no. 24, 2020, pp. 13596–602. Crossref, doi:10.1073/pnas.1922686117.

Granier, Claire, et al. “Evolution of Anthropogenic and Biomass Burning Emissions of Air Pollutants at Global and Regional Scales during the 1980–2010 Period.” *Climatic Change*, vol. 109, no. 1–2, 2011, pp. 163–90. Crossref, doi:10.1007/s10584-011-0154-1.

Hilbert, David; Fletcher, Cameron. Using artificial neural networks to assess the impacts of future climate change on ecoregions and major vegetation groups in Australia. Atherton: CSIRO; 2010. <https://doi.org/10.4225/08/584d97271d461>

Hoffmann, Samuel, et al. “Predicted Climate Shifts within Terrestrial Protected Areas Worldwide.” *Nature Communications*, vol. 10, no. 1, 2019. Crossref, doi:10.1038/s41467-019-12603-w.

Kubecka, Paul. “A Possible World Record Maximum Natural Ground Surface Temperature.” *Weather*, vol. 56, no. 7, 2001, pp. 218–21. Crossref, doi:10.1002/j.1477-8696.2001.tb06577.x.

O’Donnell, M.S., and Ignizio, D.A., 2012, Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p.

Olson, D. M., Dinerstein, E. 2002. The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden* 89(2):199-224.

Svetlana Jevrejeva, Luke P. Jackson, Riccardo E. M. Riva, Aslak Grinsted, and John C. Moore (2016), “Coastal sea level rise with warming above 2 °C”, Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1605312113

“Terrestrial Ecoregions | Biome Categories | WWF.” World Wildlife Fund, WWF, www.worldwildlife.org/biome-categories/terrestrial-ecoregions. Accessed 10 Mar. 2021.

van Vuuren, D.P., Edmonds, J., Kainuma, M. et al. The representative concentration pathways: an overview. Climatic Change 109, 5 (2011). <https://doi.org/10.1007/s10584-011-0148-z>

Yu, Deyong, et al. “Projecting Impacts of Climate Change on Global Terrestrial Ecoregions.” Ecological Indicators, vol. 103, 2019, pp. 114–23. Crossref, doi:10.1016/j.ecolind.2019.04.006.