Title

Sarah J. Hart, Asha Paudel…

Sept 20, 2023

# Abstract

# Highlights

* Across the SRM, the distribution of aspen is strongly associated with moisture availability patterns.
* Future changes in climate will likely lead to the loss of existing aspen.
* The potential for aspen to expand under future climate change will be limited by dispersal.

# Introduction

Global warming has already and will continue to impact many ecosystems around the world with important consequences for society ([Parmesan and Yohe, 2003](#Xca42e7af0598a4cbfda40b118bfb77f830f8ea3)). These effects may be particularly pronounced in high-elevation forested ecosystems, where warming is occurring more rapidly ([Pepin et al., 2015](#X3511df6da62bf1620ab2be4837010e7806482f2)), species are often characterized by slow migration rates ([Malanson and Cairns, 1997](#X787ad971afa5a57118833a675b6d883a3f8419e)), and high topoedaphic complexity may limit upslope migration ([Graae et al., 2018](#ref-graae2018StayGoHow)). In order to anticipate and mitigate the potential changes in forest ecosystems, forest managers require a better of how species distributions may shift at scales relevant to management activities.

Species distribution models (SDMs) - correlative models that associate species distribution data with environmental characteristics ([Elith and Leathwick, 2009](#ref-elith2009SpeciesDistributionModels)) - have been commonly used to understand how species ranges may shift in response to climate change ([Ehrlén and Morris, 2015](#X8cba34e779fc885716a4455d5056787f514cf88)). Despite their widespread use, many SDMs have only limited applicability for forest management and conservation planning. For instance, many SDMs are produced using coarse-scale gridded climate data (i.e. 4 x 4 km), which may artificially inflate the area of predicted suitable habitat ([Franklin et al., 2013](#ref-franklin2013ModelingPlantSpecies); [Seo et al., 2008](#ref-seo2008ScaleEffectsSpecies)). Second, many SDMs are spatially-biased because they have been constructed using species occurrence data from herbaria, species atlases, field surveys, expert range maps, and citizen science ([He et al., 2015](#ref-he2015WillRemoteSensing)), which typically include only information on species presence and are biased towards areas that are easier to access (e.g., near roads) ([Yackulic et al., 2013](#ref-yackulic2013PresenceOnlyModelling)). Third, many SDMs lack adequate documentation of their modeling approach, thereby limiting accurate assessments of reliability ([Zurell et al., 2020](#ref-zurell2020StandardProtocolReporting)).

Recent advances in computation power, geospatial technologies, and open-data science are paving the way for a new generation of use-inspired SDMs. For example, greater computation power and publicly available data and code have reduced many of the barriers that historically limited scientists from incorporating high resolution climate data in SDMs ([Kusch and Davy, 2022](#ref-kusch2022KrigRToolDownloadinga)). Scientists have also recently begun using remotely sensed datasets that provide spatially-continuous information on both species presence and absences to build SDMs with less spatial bias ([He et al., 2015](#ref-he2015WillRemoteSensing)). Importantly these remotely sensed data products can also allow for more robust predictions of range shifts by incorporating locations of known populations and information on dispersal distances ([Randin et al., 2020](#Xe979721122c9e766ad6849ab103a4c3514e0fb8)). Finally, expectations for the reproducibility and transparency of ecological research are rapidly growing ([Powers and Hampton, 2019](#ref-powers2019OpenScienceReproducibility)), as evidenced by the recently published standardized open-science protocols for building and sharing SDMs ([Zurell et al., 2020](#ref-zurell2020StandardProtocolReporting)).

Across western North America there is a particular need for updated SDMs that provide insight into how the distribution of trembling aspen (*Populus tremuloides* Michx) (hereafter aspen) may shift in response to climate change. Here aspen is the most widely distirbution tree species and it is highly valued for its contributions to biodiversity, aesthetic quality, soil carbon sequestration, and the potential to mitigate wildfire activity, among other things ([Campbell and Bartos, 2001](#X9020ed3ce609c583f65c953aa6b0da8b2ef8159); [Mitton and Grant, 1996](#ref-mitton1996GeneticVariationNatural); [Rogers et al., 2020](#ref-rogers2020GlobalViewAspen)). Importantly, recent widespread warm and dry conditions during the early 21st century have been associated with changes in the demography and distribution of aspen, highlighting the potential for future climate climate to drive dramtic shifts. For instance, during the early 21st century rapid crown dieback and mortality and poor regeneration (i.e. sudden aspen decline; SAD), was observed across much of aspen’s distribution, particularly in relatively dry areas ([Hanna and Kulakowski, 2012](#ref-hanna2012); [Rehfeldt et al., 2009](#ref-rehfeldt2009AspenClimateSuddena); [Worrall et al., 2013](#ref-worrall2013RecentDeclinesPopulusa)). While in relatively cool and wet areas affected by recent bark beetle outbreaks and wildfires, aspen has increased in dominance and even shown upward range expansion ([Andrus et al., 2020](#ref-andrus2020ForestRecoveryFollowing); [Nigro et al., 2022](#ref-nigro2022WildfireCatalyzesUpward)).

The objectives of this study are to: (1) use fine-scale remotely-sensed data on the presence and aspen of aspen to better understand how climate, topographic, and soil variables influence aspen’s distribution; (2) produce maps of the area suitable for aspen under current and future climate conditions; and (3) integrate future habitat suitability along with with maps of existing aspen coverage to characterize the potential for aspen to expand into new areas. We expand upon earlier SDMs of aspen, notably Rehfeldt et al. ([2015](#ref-rehfeldt2015), [2009](#ref-rehfeldt2009AspenClimateSuddena); [2006](#X5be6899825a8c3d2d140aa48216acf46924864d)) and Worral et al. ([2013](#ref-worrall2013RecentDeclinesPopulusa)), to build an SDM that: (1) integrates data from the most recent phase of Coupled Model Intercomparison Project (CMIP6); (2) follows the standardized protocols developed by Zurrell et al. ([2020](#ref-zurell2020StandardProtocolReporting)); (3) integrates a novel 10-m remotely sensed map of aspen cover for Cook et al. ([2024](#ref-cook2024MappingQuakingAspen)); and (4) integrates information about aspen’s dispersal to better understand the potential for range shifts to occur.

# Materials and Methods

## Study area

The SRM is a mountainous region an area of approximately 145,700 km2 that extends from southern Wyoming to northern New Mexico (Fig. 1). The SRM consists of rugged, topography with elevation ranging from 1450 m to above 4400 m, seven mountain ranges that largely trend north-south, and four Intermontane basins ([Drummond, 2012](#ref-drummond2012SouthernRockiesEcoregion)). The climate of the SRM is characterized by a continental climate, with hot summers (mean July maximum temperature of 24.5°C) and cool winters (mean January minimum temperatures of -12.3°C), and moderate precipitation (mean annual precipitation of 625 mm), most of which falls as snow ([Lukas et al., 2014](#ref-lukas2014ClimateChangeColorado); [Rodman et al., 2021](#ref-rodman2021EffectsBarkBeetle)). At local scales, the climate is driven by elevation gradients, the prevailing westerly winds, and the north-south orientation of the mountains. Temperatures are warmer at lower elevations, while more precipitation falls at higher elevations, particularly on the windward side of the Rockies ([Lukas et al., 2014](#ref-lukas2014ClimateChangeColorado)). Summer precipitation patterns exhibit a distinct latitudinal gradient, where more southern locations often receive more precipitation due to the North American Monsoon system ([Lukas et al., 2014](#ref-lukas2014ClimateChangeColorado)).

Ecosystems of the SRME correspond with topoclimatic patterns; low elevation valleys and intermountain basins are dominated by grasslands and shrublands, forest occupy intermediate elevations, while grasses, sedges, cushion plants, forbs, mosses, and lichens dominate cold, alpine elevations ([Comer, 2001](#ref-comer2001SouthernRockyMountains)). Within the ca. 55% of the SRME that is forested ([Drummond, 2012](#ref-drummond2012SouthernRockiesEcoregion)), tree communities also follow elevation gradients. Lower montane forests (< 2,300 m) are generally composed of ponderosa pine (*Pinus ponderosae*) woodlands, piñon (*Pinus edulis*) and juniper (*Juniperus* spp.) woodlands, and gamble oak (*Quercus gambelii*) shrublands. Forests of the upper montane zone (ca. 2,300 - 2,800 m) are dominated by ponderosa pine-Douglas fir mixed conifer systems, quaking aspen, and lodgepole pine (*Pinus contorta*). Forests of the subalpine zone (ca. 2,800 m - 3,200 m) are dominated by Engelmann spruce, subalpine fir, and to a lesser limber pine (*Pinus flexilis)* and Rocky Mountain bristlecone pine (*P. aristata*). Forests dynamics across the SRME are strongly shaped by climate-sensitive disturbances, notably wildfires, outbreaks of native bark beetles, and windstorms [Peet ([1981](#ref-peet1981ForestVegetationColorado)); Baker and Veblen ([1990](#ref-baker1990)); Veblen et al. ([1994](#X6d15166298ccfaa854e56f5197e8b0dbab0811c)); Veblen et al. ([2000](#ref-veblen2000ClimaticHumanInfluences)); Chapman et al. ([2012](#Xd5f24c46d84fc1170e8af9719ce34dad25ff3f0)); Hart et al. ([2014](#ref-hart2014DroughtInducesSpruce))).

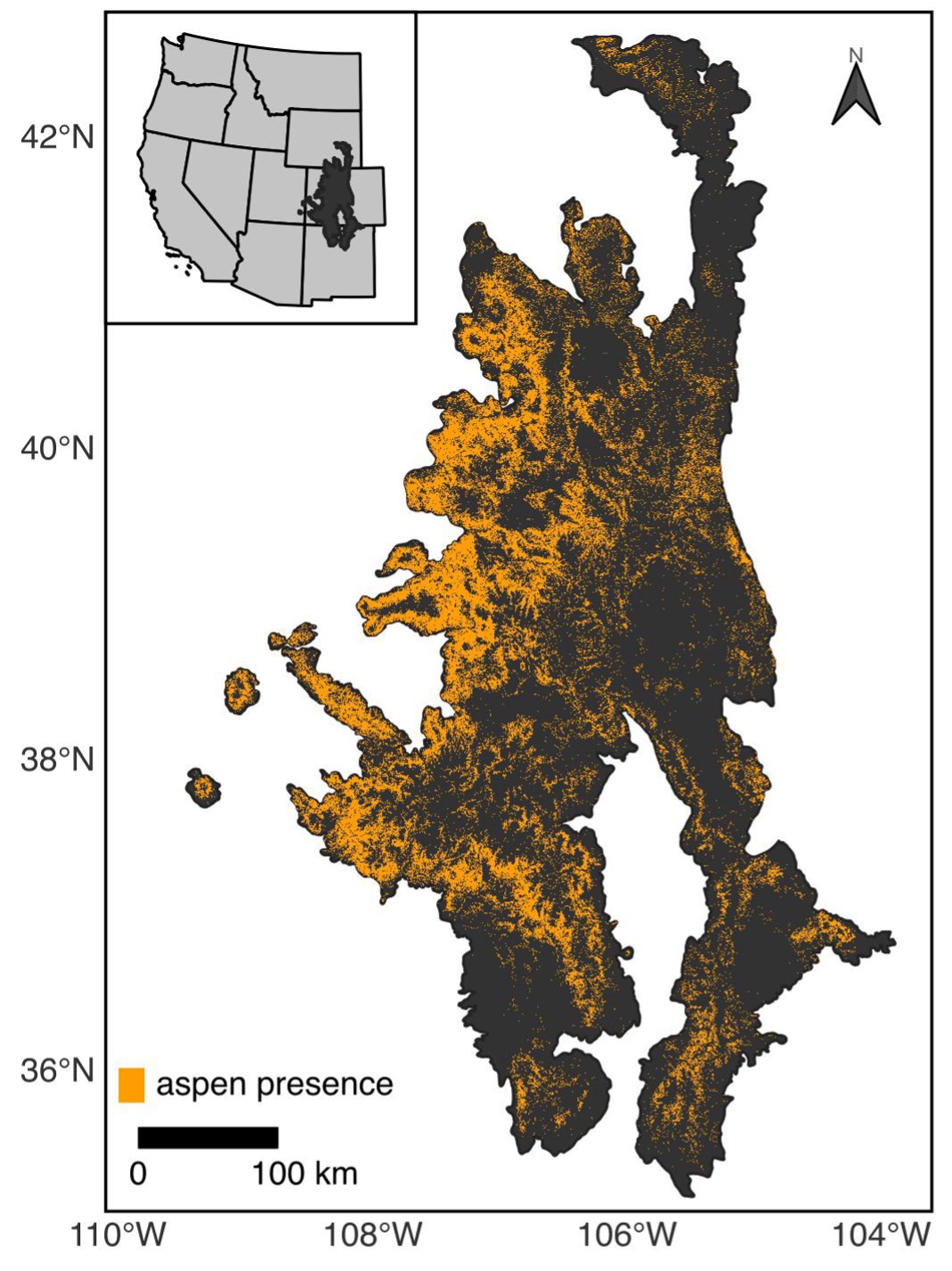


Figure 1: The Southern Rocky Mountain Ecoregion and current distribution of aspen.

## Modeling overview

Here we follow the ODMAP (Overview, Data, Model, Assessment and Prediction) protocol to develop data products that provide forest managers and other end users with a more complete understanding of the data limitations ([Zurell et al., 2020](#ref-zurell2020StandardProtocolReporting)).

### Data

#### Species occurrence data

To build SDMs, we used a 10-m gridded map of aspen presence-absence produced by Cook et al. ([2024](#ref-cook2024MappingQuakingAspen)). Briefly, this dataset was produced in Google Earth Engine using 10-m seasonal composite imagery from the Sentinel-1 and Sentinel-2 sensors. The map represents the distribution of aspen in ca. 2019 and is characterized by high accuracy (0.93 average F1-score). To generate our SDMs, we aggregated the data from 10-m to 90-m, a scale relevant to management ([Rehfeldt et al., 2015](#ref-rehfeldt2015)).

#### Predictor variables

To understand how climate is related to the contemporary distribution of aspen and the potential for future climate change to drive range shifts, we obtained gridded climate data from the AdaptWest Project ([2022](#Xcb40ace7ff505c24620413cbf7691d7f26a4667)). This dataset consists of both current and future climate data that are downscaled to 1 x 1 km resolution using the ClimateNA software (version 7.3) ([Wang et al., 2016](#ref-wang2016LocallyDownscaledSpatially)) . Contemporary climate conditions, defined here as climatalogical norms for the 1981-2010 period, were generated from 4 x 4 km climate data provided by the PRISM Climate Group ([2021](#ref-prismclimategroup2021)). Future climate conditions were generated from data included in the sixth phase of Coupled Model Intercomparison Project (CMIP6). Here we used projections of future climate for the periods 2011-2040, 2041-2070, and 2071-2100. Given considerable uncertainty about future emissions, we compared two scenarios (i.e., Shared Socioeconomic Pathways; SSPs) generated under CMIP6, SSP2-4.5 and SSP5-8.5. The SSP2-4.5 scenario describes an intermediate scenario characterized by moderate increases in emissions through 2040 followed by a decline, while the SSP5-8.5 scenario describes a more extreme situation where emissions increase through 2100 ([Riahi et al., 2017](#ref-riahi2017SharedSocioeconomicPathways)). In addition to uncertainty about societal decisions about greenhouse gas emissions represented in the SSPs, variation exists among the more than 50 atmosphere-ocean general circulation models (AOGCMs) included in CMIP6. Because of differences in complexity, assumptions, and parameterization of AOGCMS, not all forecasts are equally useful for regional planning purposes. Here we make use of an ensemble dataset constructed from eight AOCGMs identified by Mahony et al. ([2022](#ref-mahony2022GlobalClimateModel)) as being appropriate for regional applications in North America, including species distribution modeling.

To characterize the climate space that aspen currently occupies, we examined 34 biologically-relevant climate variables commonly used in species distribution models (Table 4). Broadly, these variables characterize temperature, precipitation, seasonality, and interactions between precipitation and temperature. To avoid collinearity between climate predictors, we calculated pairwise correlation coefficients. When |r|>0.75, we removed variables based on existing research (Table 4). Where evidence was similar, we used univariate random forest (RF) models to evaluate the potential explanatory power of each predictor. The resulting dataset consisted of six climate variables: (1) an annual dryness index (ADI), (2) growing season precipitation (GSP), (3) the ratio of GSP to degree days above 5 °C (GSPDD5), (4) the ratio of GSP to mean annual precipitation (PRATIO), (5) mean annual relative humidity (RH), and (6) the difference between the mean coldest month temperature and the mean warmest month temperature (TD) (Table 1).

Given mountainous areas such as the SRME are characterized high topoclimatic variation ([Franklin et al., 2013](#ref-franklin2013ModelingPlantSpecies)), we further downscaled our selected climate variables from a 1 km resolution to a 250 m resolution using gradient and inverse distance squared (GIDS) interpolation ([Flint and Flint, 2012](#ref-flint2012DownscalingFutureClimate); [Nalder and Wein, 1998](#X44a33479d922de88fa614470bb4d0b1d1f5aa3e)), following methods outlined in Rodman et al. ([2020](#ref-rodman2020ChangingClimateSnuffing)). As ancillary data in the downscaling, we used a digital elevation model (DEM) from the USGS ([2023](#X6d15385bd2515b6b1604e69149483d60c1b1585)).

In addition to climate variables, we also included data describing terrain and soils as predictors in our models. To account for the potential effects of local topographic variation on soil transport and water balance ([Jones and DeByle, 1985](#ref-jones1985Soils)), we used a 30-m DEM from the USGS ([2023](#X6d15385bd2515b6b1604e69149483d60c1b1585)) to calculate the topographic position index [TPI; Weiss ([2001](#ref-weiss2001))]. We calculated TPI for a 3-cell neighborhood (TPI3) to characterize fine scale topographic patterns ([Rodman et al., 2020](#ref-rodman2020ChangingClimateSnuffing)). To account for the effects of aspect and slope on local climate, we calculated the Heat Load Index [HLI; McCune and Keon ([2002](#ref-mccune2002EquationsPotentialAnnual)); McCune ([2007](#ref-mccune2007ImprovedEstimatesIncident))]. Both HLI and TPI3 were calculated in R using the *spatialeco* package ([Evans and Murphy, 2021](#ref-spatialEco)). Given soil properties may influence aspen demographic processes ([Jones and DeByle, 1985](#ref-jones1985Soils)), we obtained 30-m probabilistic maps of soil pH, the percentage of organic material, the percentage of clay, and saturated soil water content from the POLARIS database ([Chaney et al., 2019](#ref-chaney2019POLARISSoilProperties)). We did not include elevation, latitude, and longitude as predictors in our modeling because we assumed these relationships were only correlative ([Araújo et al., 2019](#Xa394076d2f0f03ea6b9000cd6143e30f891c94f)). We re-sampled soil and topographic predictors to a 90-m resolution by calculating the mean and projected the data to Universal Transmercator (UTM) Zone 13N to match the maps of aspen occurrence.

Table 1: Predictor variables tested for inclusion in modelling and their hypothesized relationship with aspen's distirbution. Variables marked with an asterisk were removed prior to model building to reduce collinearity among predictors.

| Variable | Description | Expected relationship with aspen |
| --- | --- | --- |
| ADI | annual dryness index: (degree-days above 5 °C)^0.5 / (mean annual precipitation) | Fast-growing, short-lived species such as aspen typiclaly have high water demand (Ireland et al. 2014). Thus high ADI has been linked with lower climate suitability for aspen (Rehfeldt et al. 2009, 2015). |
| DD\_0\* | degree-days below 0 °C | To prevent early devleopment of new buds that may lead to injury, aspen phenology is requires a chilling period. Insufficient chilling periods may delay budburst (Man et al. 2017). |
| GSP\* | growing season (Apr - Sep) precipitation (mm) | Greater precipitaiton during the growing season may alleviate summer moisture stresss (Worral et al. 2013). |
| GSPDD5 | growing season precipitation degree day ratio: (GSP\*DD5/1000) | (Rehfeldt et al. 2009). |
| PRATIO | mean precipitation ratio: (growing season precipitation) / (mean annual precipitation) | Evenly distributed precipitation (intermediate PRATIO) may promote aspen by limiting seasonal moisture stress (Rehfeltd et al. 2009) |
| RH | mean annual relative humidity (%) | Lower relative humidity leads to greater water loss via transpiration. To limit this loss, trees may close their stomata and thereby limiting photosynthesis. |
| TD | difference between MCMT and MWMT (°C) | Extreme seaonal varaition in temperature may present physiological challenges to aspen (Worrall et al. 2013; Rehfeldt et al. 2015) |
| Clay | clay | High clay content may inhibit aspen growth (Jones and DeByle 1985) |
| OM | soil organic matter [log10(%)] | Aspen is expected grow better on soils with high organic matter content (Perala 1990) |
| SWC | saturated water content (m3/m3) | Aspen is expected grow better on soils with greater water holding capacity (Perala 1990). |
| pH\* | soil pH | High soil pH may decrease the availability of nutrients and limit aspen growth (Zhang et al. 2013). |
| HLI | heat load index | Greater HLI may inhibit aspen, particularly at lower elevations and latitudes (Jones et al. 1985). |
| TPI | topographic position index | Aspen is expected grow better in valley bottoms (low TPI) and on benches (moderate TPI) than steep slopes (high TPI), where soil water content may be lower (Jones and DeByle 1985) We calculated TPI using neighborhoods of 3 (TPI3) and 15 (TPI15), however we retained only TPI3 because of high correlation among the two indices. |

### Modeling Approach

To characterize suitable habitat for aspen, we used four different modeling approaches commonly applied in species distribution modeling, generalized linear models (GLMs), generalized additive models (GAMs), and random forests (RFs), and regularized gradient boosted tree (RGBTs). Here, we first overview our modeling approach before describing specifics for individual modeling techniques. All models were fit in *R* ([R Core Team, 2022](#X4878592beea4a6dfca9c91383c925df652b2c7a)) using a *tidymodels* framework ([Kuhn and Wickham, 2020](#ref-tidymodels)).

To build GLM, GAM, RF, and GBT models, we first constructed a balanced data consisting of 10,000 pixels with aspen present and 10,000 pixels without aspen. To minimize the potential effects of spatial autocorrelation, pixels were selected so that they were separated by at least 1 km. To improve model fit and interpretability, all predictor variables were standardized by calculating standard scores. Using this dataset, we reduced our set of environmental predictors to minimize the potential effects of collinearity on model inference and projection. Specifically, we used the *spatialRF* ([Wright and Ziegler, 2017a](#ref-spatialRF)) to calculate variable inflation factors (VIF), which indicate when a predictor variable is a linear combinations of other predictor variables. We then iteratively removed variables until VIF<5 for all variables.

To build and evaluate models, we then split our dataset into equally-sized testing and training datasets. The testing dataset was further split to create five spatial cross-validation folds using the *spatialsample* package ([Mahoney et al., 2023](#ref-spatialsample)). We then tuned model hyperparameters using spatial cross-validation and identified the best hyperparameters based on the area under the receiver operator curve (AUC). To evaluate the capacity of our model to predict to new areas, we then fit the model using the best hyperparameters to each spatial cross-validation fold and assessed the variation in the AUC statistic. Given the model’s ability to predict aspen habitat in new areas, we then fit a final model to the training dataset and predicted the probability of aspen presence for the testing dataset. Using the *probably* package ([Kuhn et al., 2023](#ref-probably)), we determined the probability threshold that maximized the Youden’s J statistic ([Youden, 1950](#ref-youden1950IndexRatingDiagnostic)) and then calculated class-based accuracy statistics based on this threshold.

To better understand the environmental drivers of aspen’s distribution and assess model realism, we calculated variable importance scores for each model using a model-agnostic permutation-based approach. In this approach each variable is randomized and then the AUC statistic is compared with AUC for the full model (where data has not been randomized). We also evaluated the relationship between aspen presence and each predictor variable using accumulated local effects (ALE) profiles ([Apley and Zhu, 2020](#ref-apley2020VisualizingEffectsPredictor)). Variable importance and ALE were calculated in R using the *DALEX* ([Biecek, 2018](#ref-DALEX)) and *ALEPlot* ([Apley, 2018](#ref-ALEPlot)) packages.

#### Generalized linear models

GLMs are extensions of parametric linear regression adapted to distributions other than the normal distribution ([Zuur et al., 2007](#ref-zuur2007AnalysingEcologicalData)). Here we constructed GLMs with a logit link function and a binomial error distribution to account for the structure of presence-absence data. We included both linear and quadratic effects for all variables, but did not explore any interaction terms. We fit GLMs using a Lasso regularization approach, which allows for model coefficients to be reduced to zero, thereby limiting model complexity and improving bias-variance trade offs ([Hastie et al., 2009](#X1bac580b2d504864f7c67f41ed89ab717da9984)). Prior to fitting the model to the full training dataset, we tuned the lasso penalty term. GLMs were fit using the *glmnet* package ([Friedman et al., 2010](#ref-glmnet)).

#### Generalized additive models

GAMs are a non-parametric extension of GLMs that are particularly useful when there is no *a priori* reason for fitting a particular relationship (e.g., linear, quadratic). Here, we constructed binomial GAMs with a logit link function using the *mgcv* package ([Wood, 2011](#ref-mgcv)). In our model, we represented the relationships between the response and each predictor variable using thin plate regression splines, where the penalty term was adjusted to allow the term to be shrunk to zero. For all smooths, we set the k parameter, which sets the number of basis functions, to the default value of 10, after using built in diagnostic function from the *mgcv* package to confirm an adequate degree of complexity. To limit overfitting, we tuned the penalty term prior to fitting the model to the full training dataset. GAMs were fit using restricted maximum likelihood (REML), following recommendations from Pedersen et al. ([2018](#X9ce84d2edd409eea4a5bb93b0b43ed50266fe99)).

#### Random Forests

RF models are an extension of classification and regression tree analysis (CART; Breiman et al. ([1984](#X5d77a0ffef6fd5d853fb44d9928a90c8372fd1a))), a nonparametric approach where decision trees are used to explain the variation in the response variable by repeatedly splitting the data into more similar groups ([Death and Fabricius, 2000](#ref-death2000)). Tree-based approaches are useful for modeling nonlinear relationships and complex interactions among variables, which often characterize ecological data ([Cutler et al., 2007](#X8af443e7151d3081ddf81ae6488f6e02a99b7ea); [Death and Fabricius, 2000](#ref-death2000)). RF builds upon bagging methods, where many trees are built using random samples (with replacement) of the training data and predictions are generated from the ensemble set of trees. While bagging reduces some of the overfitting issues associated with CART, RF methods also limit the number of variables to consider at any given split to a random subset and the complexity of each tree by limiting splits to only nodes with a minimum number of data points. The inclusion of these hyperparameters results in less correlation among trees and thus better bias-variance trade offs ([Cutler et al., 2007](#X8af443e7151d3081ddf81ae6488f6e02a99b7ea)). Prior to fitting the model to the full training dataset, we tuned these hyperparameters, while holding the number of trees constant at 1000. RF models were fit using the tidymodels implementation of the *ranger* package ([Wright and Ziegler, 2017b](#ref-ranger)).

#### Regularized gradient boosting trees

Gradient boosted trees (GBTs) are also an ensemble-based extension of CART ([De’ath, 2007](#ref-death2007BoostedTreesEcological)). In contrast to RF where trees are built in parallel, individual trees in a GBT ensemble are constructed iteratively so that each successive tree attempts to improve upon predictions made by the previous tree ([Friedman, 2001](#Xa70ea2a825fb5a22808884ac0146de3f2ccb8ee)). To improve bias-variance trade offs, GBTs incorporate hyperparameters that control the rate at which the boosting algorithm adapts, and introduce randomness into the tree construction by sampling both variables (i.e., columns) and cases (i.e., rows) used to fit the model. RGBTs expand upon these approaches by incorporating regularization terms that constrain the depth of the tree and setting limits on the amount of gain in model fit required to further partition a node of the tree. We tune these hyperparameters prior to fitting the model to the full training dataset. RGBTs were fit using the R package *xgboost* ([Chen et al., 2023](#ref-xgboost)).

#### Model Ensemble

To account for uncertainty due to modelling approach ([Araujo and New, 2007](#ref-araujo2007EnsembleForecastingSpecies)), we generated ensemble predictions by combining predictions for each model. Specifically, we calculated a weighted probability of occurrence from all four presence-absence models. We assigned weights based on the AUC statistic. We then comparing the ensemble prediction probability with the testing dataset and determined the probability threshold that maximized the Youden’s J statistic ([Youden, 1950](#ref-youden1950IndexRatingDiagnostic)). We used this threshold to calculate class-based accuracy statistics.

### Forecasting change in aspen habitat

To understand how future changes in climate may affect the distribution of aspen, we used the ensemble model to forecast the probability of aspen habitat suitability under the SSP2-4.5 and SSP5-8.5 scenarios for the 2011-2040, 2041-2070, and 2071-2100 periods. We then used these probabilistic forecasts of aspen habitat suitability to forecast aspen presence and absence based on the optimal probability threshold (see 4.2.2.5). We then overlaid forecasts of aspen presence-absence with the map of existing aspen occurrence from Cook et al. (in review) to produce maps where changes in climate may lead to aspen gain, loss, or stability. Finally, given expansion will be constrained by dispersal, we used a moving widow approach to calculate the distance to the nearest existing stand of aspen for each gain pixel. Specifically, we quantified the presence of aspen with 3, 5, 7, 13, 25, and 47 cell neighborhoods (i.e., within 90 m, 180 m, 270 m, 540 m, 1080 m, and 2070 m).

# Results

## Model performance

Spatial cross-validation revealed all models accurately predicted to new areas; across the five folds the mean AUC statistic (± standard error of the mean) was 0.81 ± 0.02, 0.82 ± 0.02, 0.83 ± 0.01, 0.83 ± 0.01, for the GLM, GAM, RF, and RGBT, respectively. When compared with the testing data, all models achieved an AUC greater than 0.84, indicating a good model fit (Table 2). When compared aspen with an independent dataset of aspen presence derived from aerial photo interpretation, the ensemble model correctly predicted 78% of points (n=13696).

Table 2: Model performance statistics. Observed values are from independent testing data.

| Model | Accuracy | F measure | kappa | Precision | Recall | AUC | Sensitivity | Specificity |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ensemble | 0.79 | 0.79 | 0.58 | 0.80 | 0.77 | 0.87 | 0.77 | 0.81 |
| GAM | 0.77 | 0.76 | 0.54 | 0.80 | 0.72 | 0.85 | 0.72 | 0.82 |
| GLM | 0.76 | 0.77 | 0.52 | 0.75 | 0.79 | 0.84 | 0.79 | 0.73 |
| RF | 0.79 | 0.79 | 0.59 | 0.82 | 0.75 | 0.88 | 0.75 | 0.83 |
| XGB | 0.79 | 0.78 | 0.58 | 0.82 | 0.75 | 0.88 | 0.75 | 0.84 |

While the ensemble model performed well, there were spatial patterns in the residuals (Fig. 8 and 2). Relative to true positives (i.e., pixels where aspen was present that were correctly classified), false negatives (i.e., pixels where aspen was present but the model predicted absence) were concentrated at more eastern latitudes. False negatives were common where the percent aspen cover within the 90 x 90 m pixel was low (median value of 8.6%).

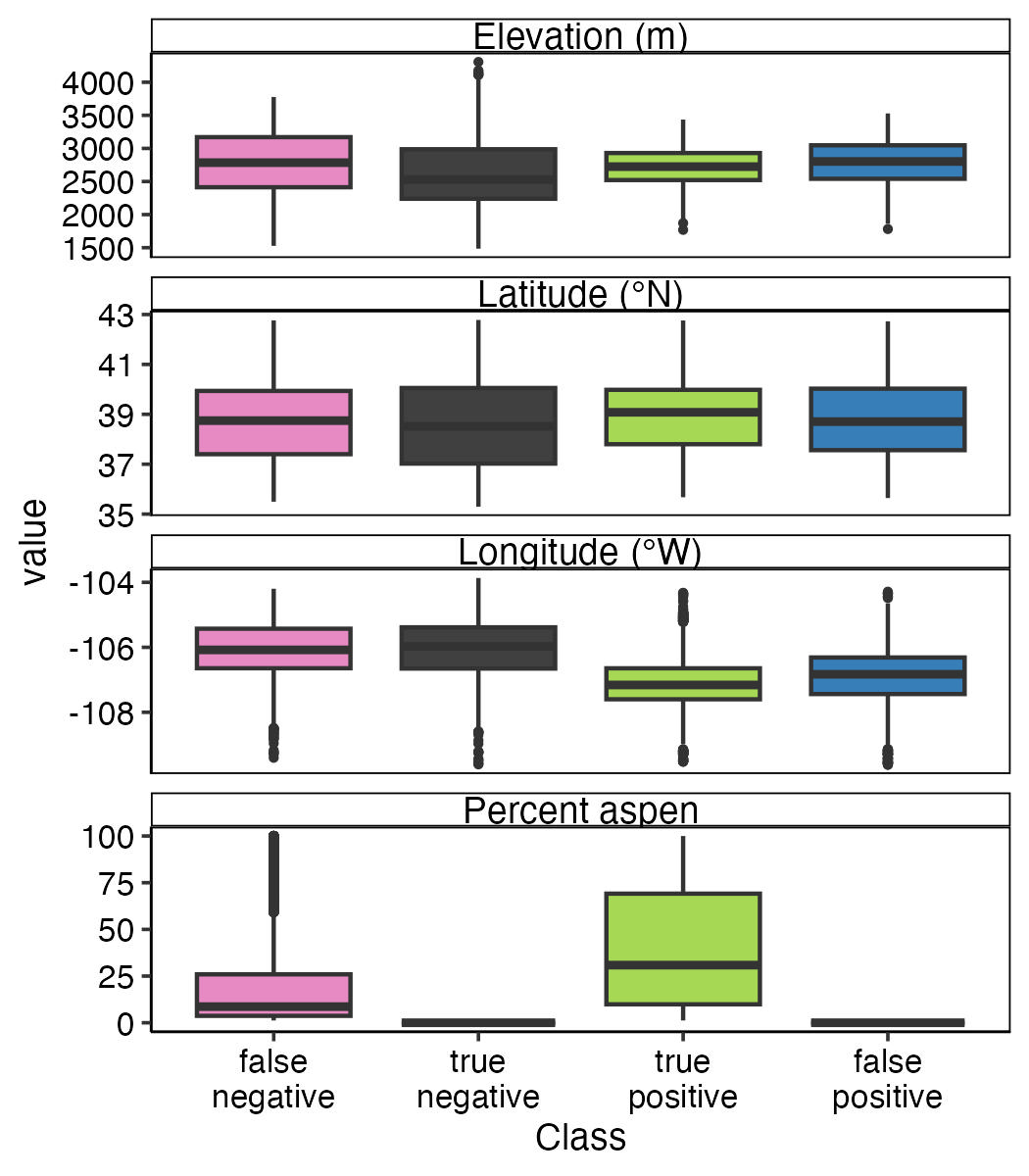


Figure 2: Boxplots illustrating the relationship between pixels missclassified and geographic variables and percent aspen cover.

## Effects of predictor variables on aspen habitat suitability

Variable importance scores illustrated the effects of regularization, which were included in the GLM and GAM and allowed for coefficient estimates to shrink to zero. In these models, the top variables were very important predicting aspen presence (Fig. 3A). Across all models, variable importance scores revealed that climate variables generally contributed more to model fit than soil or topographic factors (Fig. 3A). PRATIO was on average the most important predictor, followed by ADI, and GSPDD5 (Fig. 3A). Forecasted decreases in PRATIO over the next 100 years (Fig. 3B), will likely lead to an increase in aspen habitat suitability across the SRM. However, for locations where PRATIO is already low, the GAM, RF, and RGBT models suggest that increases in the PRATIO may decrease aspen habitat suitability.

Future increases in ADI over the next century, will likely lead to a decrease in mean aspen habitat suitability (Fig. 3B). However, the GLM suggested that future increases in ADI may initially lead to an increase in mean aspen habitat suitability, followed by a decrease by 2071-2100. PRATIO was the second most important predictor on average (Fig. 3A). GSPDD5 was identified as a particularly important predictor variable in the GAM (Fig. 3A). Across the study area GSPDD5 is expected to decrease over the next century, largely due to increases in temperature (Fig. 3B). Our SDMs suggest this may lead to a decrease, increase, or no change in mean aspen habitat suitability across the SRM. In the GAM, GSPDD5 was negatively associated with aspen presence, while GSPDD5 had no effect in the GLM, and the relationship was more hump-shaped in RF and RGBT models. All models suggested that aspen habitat suitability was higher in valley bottoms (low TPI3) than steep slopes (high TPI3), although RF and RGBT models suggested benches (moderate TPI3) may be equally suitable as steep slopes (Fig. 3B).

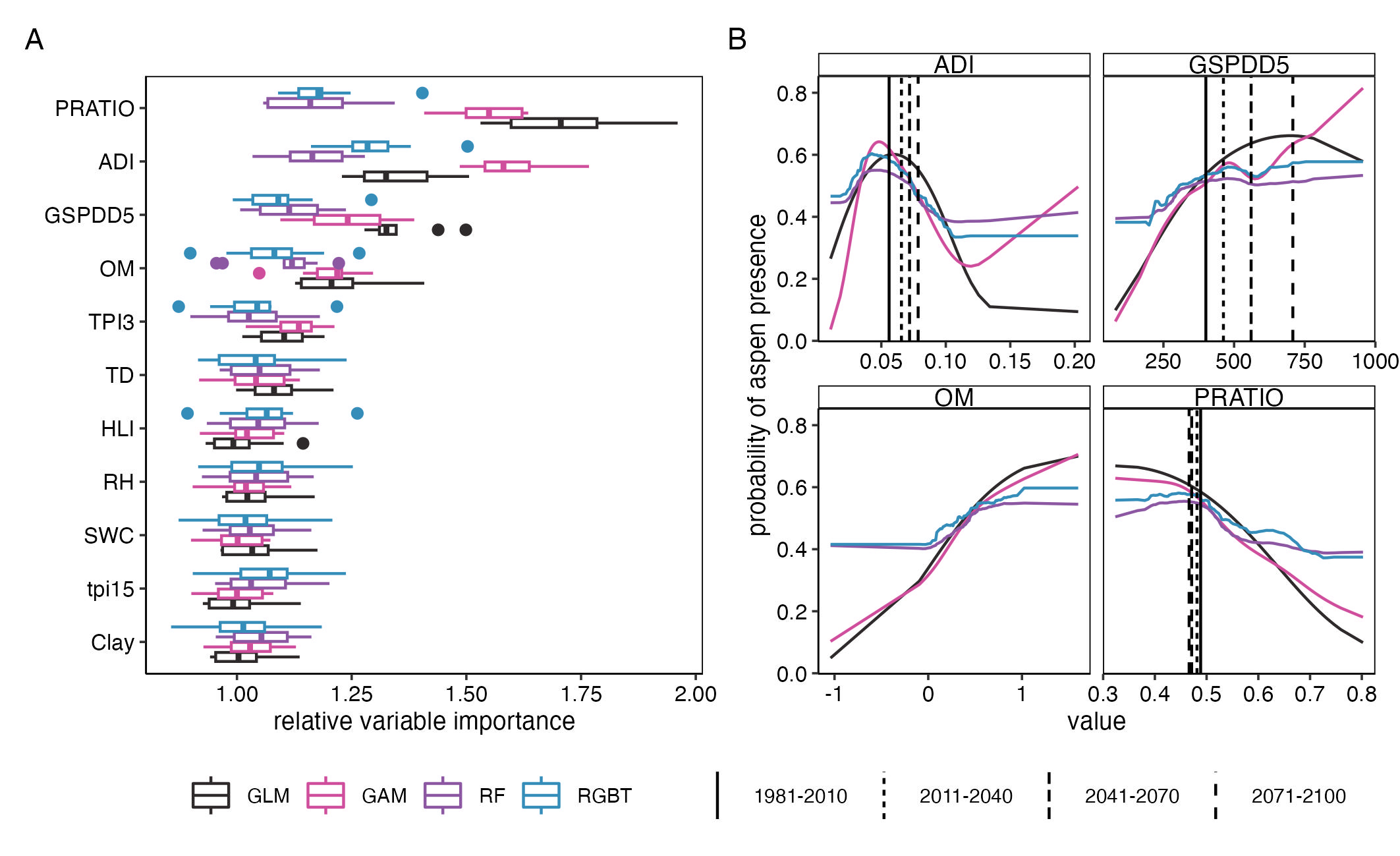


Figure 3: Variable importance scores (A) and accumulate local effects (B) for models of aspen habitat suitability by modeling approach. In A, boxes illustrate the loss in model performance (1-AUC) when the predcitor variable was been randomized for 10 different permutations. In B, vertical lines illustrate the mean climate conditions for areas with existing aspen for the historical period (1981-2010) and projections for the 2011-2040, 2041-2070, and 2071-2100 periods under the SPP4-8.5 scenario. For variable definitions and descriptions see Table 1.

## Forecasted change in the distribution of aspen

Our ensemble SDM forecasts notable decreases in future aspen habit suitability with particularly dramatic decreases predicted to occur within the first half of the 21st century. Under the SSP2-4.5 scenario, the ensemble model suggests a percent change in the mean probability of aspen of -13.2 ± 36.9% by 2040 and 8.8 ± 82.2% by 2100. Under the SSP5-8.5 scenario, our model forecasts similar reductions on average, but more variable responses across the study area. The predicted percent change in the mean probability of aspen was -11.7 ± 43.3% for the 2011-2040 period and 35 ± 134.7% for the 2071-2100 period.

Based on the 0.505 probability threshold for classifying aspen occurrence, which was selected based on Youden’s J statistic (4.2.2.5), our ensemble SDM suggests that decreases in the aspen habitat suitability may result in the loss of aspen across 11562 km2 under the SSP2-4.5 scenario and 11344 km2 under the SSP5-8.5 scenario by 2100 (Fig. 10). However, the decrease in the area suitable for aspen may be offset by increases in the area suitable for aspen. By 2100, the ensemble model suggests an increase in the area suitable for aspen of 22624 and 26924 km2 under the SSP2-4.5 and SSP5-8.5 scenarios, respectively. Across both scenarios and all time periods, losses in the area suitable for aspen are forecasted to occur at lower elevations and eastern latitudes, where aspen is already limited (Figs. 4 and 10). Increases in the area suitable for aspen are forecasted to occur at higher elevations.

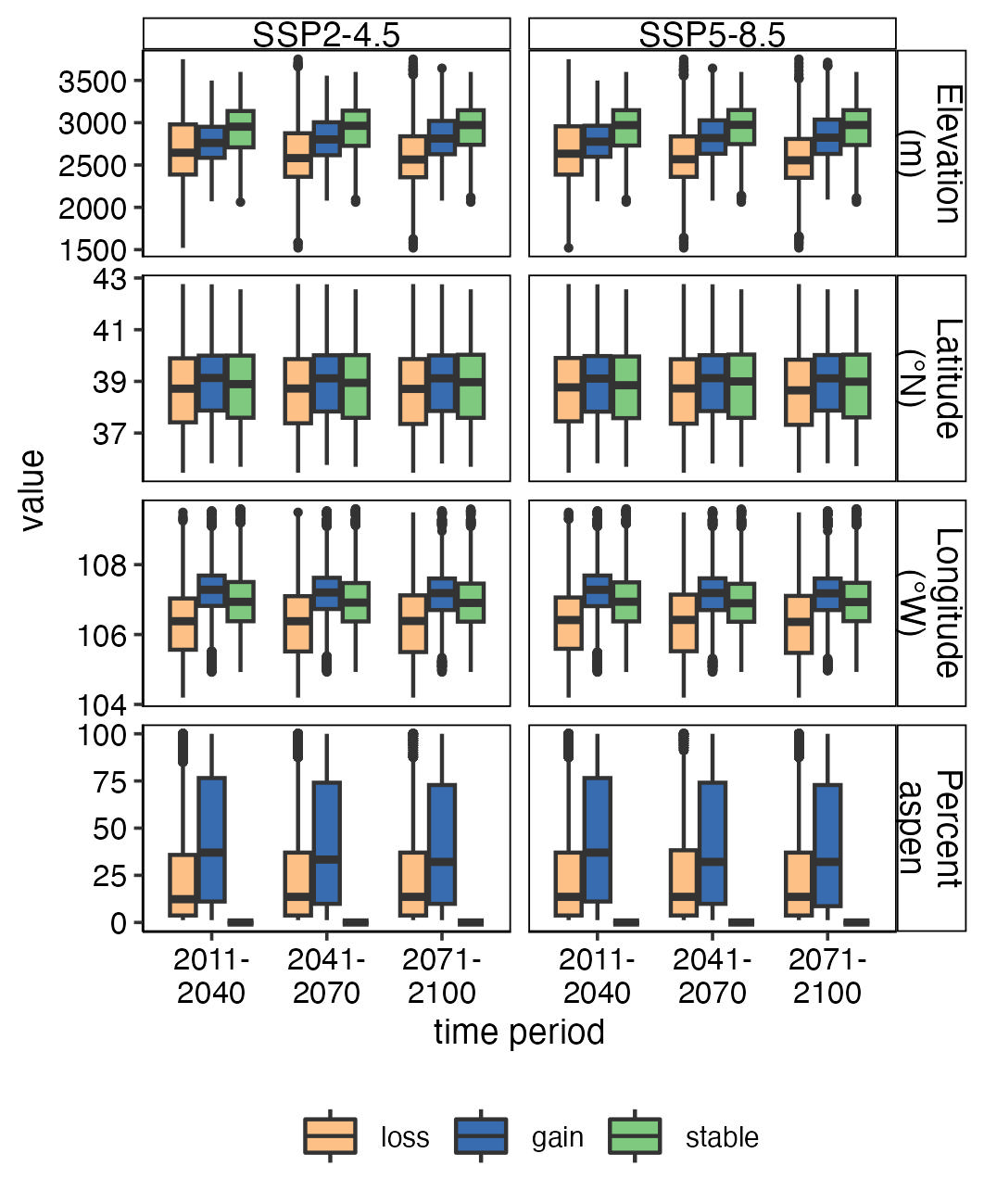


Figure 4: Boxplots illustrating spatial patterns in the areas were aspen my lost, gained, or remain stable based on the ensemeble SDM’s forecast of future aspen habitat suitability.

While the area suitable for aspen is expected to increases substantially, only a fraction of that area is near existing aspen stands (Fig. 5 and 11). Under the SSP4-2.5 scenario, 1 % of the area forecasted to become suitable by 2071-2100 is within 90-m of existing aspen (Fig. 5), and 33% of the suitable area is within 540-m of existing aspen. Under the SSP5-8.5 scenario, 2 % of the area forecasted to become suitable by 2071-2100 is within 90-m of existing aspen (Fig. 5), and 38% of the suitable area is within 540-m of existing aspen.

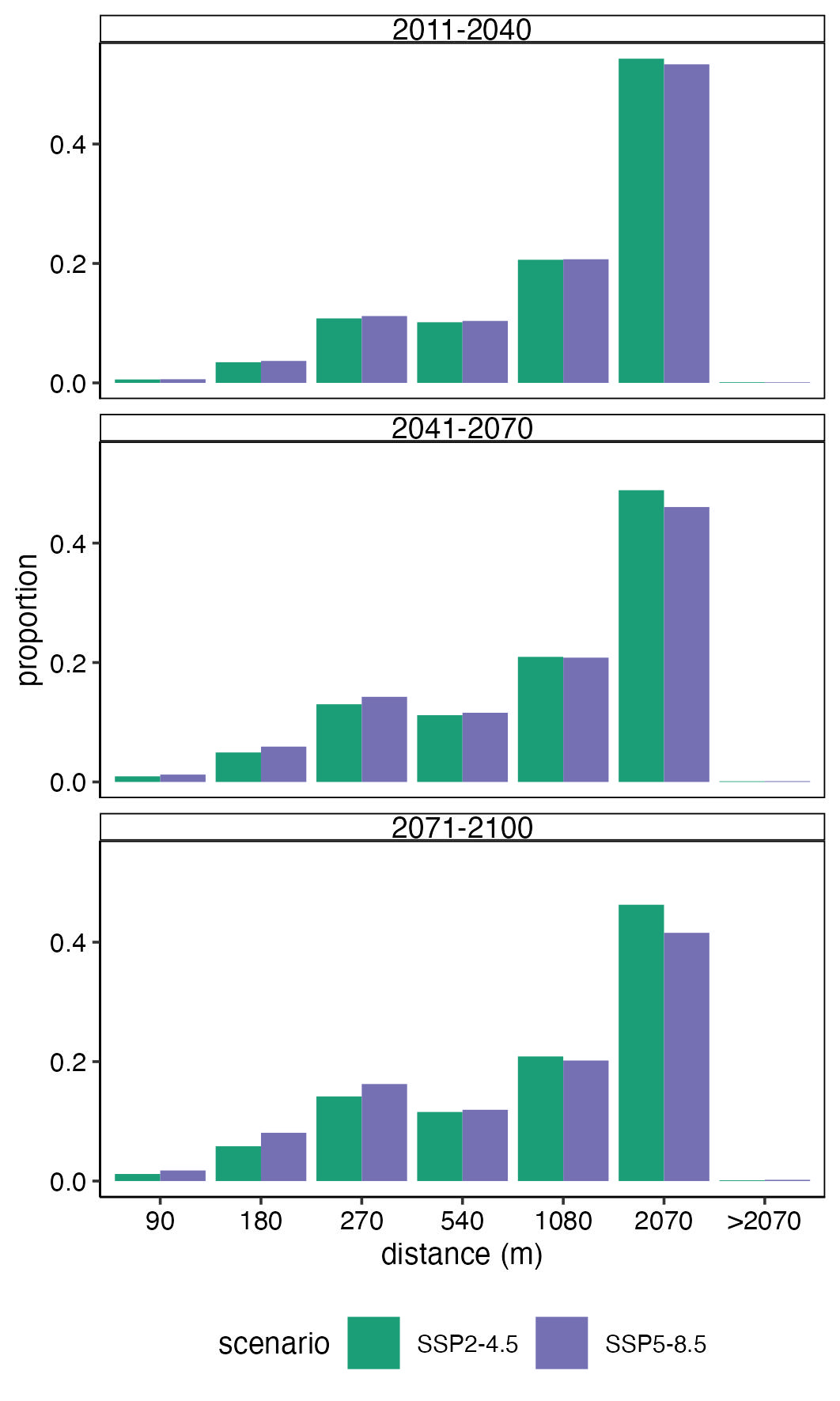


Figure 5: The distance to the nearest existing aspen patch for pixels where changes in climate may promote aspen expansion.

# Discussion

## Overview

## Model performance

Here we accurately modeled the distribution of aspen across the SRME. Our ensemble model correctly predicted 78% of the aspen presences and 76% of the pixels without aspen. While previous aspen SDM efforts have reported lower error rates [e.g. <10%; Rehfeldt et al. ([2009](#ref-rehfeldt2009AspenClimateSuddena)); Worrall et al. ([2013](#ref-worrall2013RecentDeclinesPopulusa))], we note that it is difficult to directly compare SDM performance when models differ in their spatial extent or grain, input data, and validation strategy. In our SDM, most false negatives were located at lower elevations and more eastern latitudes. In these areas, aspen is generally present in lower abundance, presumably due to lower habitat quality. In contrast, most false negatives were located at higher elevations and more western latitudes. In some habitats across the Southern Rocky Mountains, aspen regeneration is thought to occur episodically following severe stand-replacing fire, which removes competing conifers and stimulates vegetative reproduction ([Kurzel et al., 2007](#ref-kurzel2007TypologyStandStructure)). Thus some errors of commission likely result from successional patterns and disturbance history, which are poorly captured by SDMs.

## Climate, soil, and topographic drivers

Our modeling confirms the importance of climate, topography, and soils characteristics in driving aspen’s distribution ([Rehfeldt et al., 2006](#X5be6899825a8c3d2d140aa48216acf46924864d)). Notably, across all models we found that climate variables were more important predictors than soil or topographic variables, consistent with previous research ([Rehfeldt et al., 2015](#ref-rehfeldt2015)). However, prior to modeling our variable reduction processes reduced the set of thirteen variables to only eight; we retained five of seven climate variables, two out of three topographic variables, and one out of four soil variables. Thus at least some of the reason that soil and topographic may be due to covariance between climate and topoedaphic variables. For instance, we expected that high soil pH would limit aspen ([Zhang et al., 2013](#X78355eafbabb26d86a075674e0ada5d3242e687)), however across the SRME pH was positively correlated with ADI. Thus at least part of the effect of ADI on aspen’s distribution may be attributed to soil effects.

Across all models, we found that ADI was the most important predictor, consistent with previous research on the distribution of aspen across the western US ([Rehfeldt et al., 2009](#ref-rehfeldt2009AspenClimateSuddena)). In areas with high ADI, the occurrence of aspen may be limited by interactions between precipitation inputs and losses due to evapotranspiration ([Hogg, 1994](#ref-hogg1994ClimateSouthernLimit)). These findings are supported by observations of high aspen mortality during periods of drought ([Anderegg et al., 2013](#X5a462d17356fcce115cf66a541d09dd7f11b667); [Hogg et al., 2008](#ref-hogg2008ImpactsRegionalDrought)). Given ADI is expected to substantially increase across the SRM over the 21st century, areas where ADI is already high may experience elevated mortality, and contraction of aspen’s distribution. This finding is consistent with observations of sudden aspen decline in areas with higher ADI (Fig. 12).

Aspen was generally more likely to occur in areas where growing season precipitation made up about half of total precipitation, consistent with our expectation and previous research ([Rehfeldt et al., 2009](#ref-rehfeldt2009AspenClimateSuddena)). Across the SRME, PRATIOs are generally highest east of the Continental Divide, where more precipitation occurs during the growing season due to storm tracks that bring warm and wet moisture from the Gulf of Mexico ([Lukas et al., 2014](#ref-lukas2014ClimateChangeColorado)). Whereas areas of low PRATIO generally occur at high elevations, west of the Continent Divide, which receive large precipitation inputs in the form of snow due to the prevalence of westerly winds flowing from the Pacific Ocean in the winter ([Lukas et al., 2014](#ref-lukas2014ClimateChangeColorado)). Notably, these geographic patterns in the seaonality of precipitation also reflect patterns in annual precipitation and temperature, where higher PRATIO correspond with higher mean annual temperatures and lower mean annual precipitation.

We also identified GSPDD5 as in important predictor of aspen’s distribution across the SRME, however the effect and importance varied with modeling approach. For instance, in the GLM there was no effect of GSPDD5 on the occurrence of aspen, whereas in the GAM, GSPDD5 was the second most important predictor. In this model, the probability of aspen occurrence decreased dramatically with increasing GSPDD5. Across the SRME, high GSPDD5 occurs at high elevations where mean annual temperatures are low and precipitation is generally moderate to high. Thus in order modeling the effects of GSPDD5 probably represents a temperature limitation, rather a seasonal moisture stress effect as has been reported elsewhere ([Rehfeldt et al., 2009](#ref-rehfeldt2009AspenClimateSuddena)). This interpretation is supported by recent observations of aspen expansion at high elevations ([Nigro et al., 2022](#ref-nigro2022WildfireCatalyzesUpward)), which have occurred during a period of pronounced warming across Colorado ([Bolinger et al., 2023](#ref-bolinger2023ClimateChangeColorado)). Future warming may continue to provide opportunities for aspen range expansion.

## Dispersal

## The future for aspen

## Management Implications

# Conclusions

# References

AdaptWest Project, 2022. [Gridded current and projected climate data for north america at 1km resolution, generated using the ClimateNA v7.30 software (t. Wang et al., 2022).](https://Available at adaptwest.databasin.org.)

Anderegg, L.D.L., Anderegg, W.R.L., Abatzoglou, J., Hausladen, A.M., Berry, J.A., 2013. Drought characteristics’ role in widespread aspen forest mortality across Colorado, USA. Global Change Biology 19, 1526–1537. <https://doi.org/10.1111/gcb.12146>

Andrus, R.A., Hart, S.J., Veblen, T.T., 2020. Forest recovery following synchronous outbreaks of spruce and western balsam bark beetle is slowed by ungulate browsing. Ecology 101, e02998. <https://doi.org/10.1002/ecy.2998>

Apley, D., 2018. [ALEPlot: Accumulated local effects (ALE) plots and partial dependence (PD) plots](https://CRAN.R-project.org/package=ALEPlot).

Apley, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82, 1059–1086. <https://doi.org/10.1111/rssb.12377>

Araujo, M., New, M., 2007. Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O’Hara, R.B., Zimmermann, N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. Science Advances 5, eaat4858. <https://doi.org/10.1126/sciadv.aat4858>

Baker, W.L., Veblen, T.T., 1990. Spruce beetles and fires in the nineteenth-century subalpine forests of western colorado, USA. Arctic and Alpine Research 22, 6580.

Biecek, P., 2018. [DALEX: Explainers for complex predictive models in r](https://jmlr.org/papers/v19/18-416.html) 19, 1–5.

Bolinger, R., Lukas, J., Schumacher, R., Gpble, P., 2023. Climate change in Colorado. <https://doi.org/10.25675/10217/237323>

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.G., 1984. Classification and regression trees. Belmont, CA, USA. <https://doi.org/10.1201/9781315139470>

Campbell, R.B., Bartos, D.L., 2001. Aspen ecosystems: Objectives for sustaining biodiversity. p. 299307.

Chaney, N.W., Minasny, B., Herman, J.D., Nauman, T.W., Brungard, C.W., Morgan, C.L.S., McBratney, A.B., Wood, E.F., Yimam, Y., 2019. POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties Over the Contiguous United States. Water Resources Research 55, 2916–2938. <https://doi.org/10.1029/2018WR022797>

Chapman, T.B., Veblen, T.T., Schoennagel, T., 2012. Spatiotemporal patterns of mountain pine beetle activity in the southern rocky mountains. Ecology 93, 2175–2185. <https://doi.org/10.1890/11-1055.1>

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., 2023. [Xgboost: Extreme gradient boosting](https://CRAN.R-project.org/package=xgboost).

Comer, P., 2001. Southern rocky mountains: An ecoregional assessment and conservation blueprint.

Cook, M., Chapman, T., Hart, S., Paudel, A., Balch, J., 2024. Mapping Quaking Aspen Using Seasonal Sentinel-1 and Sentinel-2 Composite Imagery across the Southern Rockies, USA. Remote Sensing 16, 1619. <https://doi.org/10.3390/rs16091619>

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 27832792.

De’ath, G., 2007. Boosted Trees for Ecological Modeling and Prediction. Ecology 88, 243–251. <https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2>

Death, G., Fabricius, K.E., 2000. [Classification and regression trees: A powerful yet simple technique for ecological data analysis](http://www.esajournals.org/doi/pdf/10.1890/0012-9658(2000)081%5B3178:CARTAP%5D2.0.CO%3B2). Ecology 81, 31783192.

Drummond, M.A., 2012. [Southern rockies ecoregion: Chapter 8 in status and trends of land change in the western united states–1973 to 2000](http://pubs.er.usgs.gov/publication/pp1794A8). Reston, VA.

Ehrlén, J., Morris, W.F., 2015. Predicting changes in the distribution and abundance of species under environmental change. Ecology Letters 18, 303–314. <https://doi.org/10.1111/ele.12410>

Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

Evans, J.S., Murphy, M.A., 2021. [spatialEco](https://github.com/jeffreyevans/spatialEco).

Flint, L.E., Flint, A.L., 2012. Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis. Ecological Processes 1, 2. <https://doi.org/10.1186/2192-1709-1-2>

Franklin, J., Davis, F.W., Ikegami, M., Syphard, A.D., Flint, L.E., Flint, A.L., Hannah, L., 2013. Modeling plant species distributions under future climates: how fine scale do climate projections need to be? Global Change Biology 19, 473–483. <https://doi.org/10.1111/gcb.12051>

Friedman, J.H., 2001. [Greedy function approximation: A gradient boosting machine](https://www.jstor.org/stable/2699986). The Annals of Statistics 29, 1189–1232.

Friedman, J., Tibshirani, R., Hastie, T., 2010. Regularization paths for generalized linear models via coordinate descent 33. <https://doi.org/10.18637/jss.v033.i01>

Graae, B.J., Vandvik, V., Armbruster, W.S., Eiserhardt, W.L., Svenning, J.-C., Hylander, K., Ehrlén, J., Speed, J.D.M., Klanderud, K., Bråthen, K.A., Milbau, A., Opedal, Ø.H., Alsos, I.G., Ejrnæs, R., Bruun, H.H., Birks, H.J.B., Westergaard, K.B., Birks, H.H., Lenoir, J., 2018. Stay or go how topographic complexity influences alpine plant population and community responses to climate change. Perspectives in Plant Ecology, Evolution and Systematics, Special issue on alpine and arctic plant communities : A worldwide perspective 30, 41–50. <https://doi.org/10.1016/j.ppees.2017.09.008>

Hanna, P., Kulakowski, D., 2012. [The influences of climate on aspen dieback](http://www.sciencedirect.com/science/article/pii/S0378112712000783). Forest Ecology and Management 274, 9198.

Hart, S.J., Veblen, T.T., Eisenhart, K.S., Jarvis, D., Kulakowski, D., 2014. Drought induces spruce beetle (dendroctonus rufipennis) outbreaks across northwestern colorado. Ecology 95, 930939.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: Data mining, inference, and prediction. Springer.

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W., Wegmann, M., Pettorelli, N., 2015. Will remote sensing shape the next generation of species distribution models? Remote Sensing in Ecology and Conservation 1, 4–18. <https://doi.org/10.1002/rse2.7>

Hogg, E.H.(Ted)., 1994. Climate and the southern limit of the western Canadian boreal forest. Canadian Journal of Forest Research 24, 1835–1845. <https://doi.org/10.1139/x94-237>

Hogg, E.H.(Ted)., Brandt, J.P., Michaelian, M., 2008. Impacts of a regional drought on the productivity, dieback, and biomass of western canadian aspen forests. Canadian Journal of Forest Research 38, 1373–1384. <https://doi.org/10.1139/X08-001>

Jones, J.R., DeByle, N.V., 1985. [Soils](https://www.fs.usda.gov/research/treesearch/27780). In: DeByle, Norbert V.; Winokur, Robert P., editors. Aspen: Ecology and management in the western United States. USDA Forest Service General Technical Report RM-119. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colo. p. 65-70 119.

Kuhn, M., Vaughan, D., Ruiz, E., 2023. [Probably: Tools for post-processing class probability estimates](https://CRAN.R-project.org/package=probably).

Kuhn, M., Wickham, H., 2020. [Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.](https://www.tidymodels.org)

Kurzel, B.P., Veblen, T.T., Kulakowski, D., 2007. A typology of stand structure and dynamics of Quaking aspen in northwestern Colorado. Forest Ecology and Management 252, 176–190. <https://doi.org/10.1016/j.foreco.2007.06.027>

Kusch, E., Davy, R., 2022. KrigRa tool for downloading and statistically downscaling climate reanalysis data. Environmental Research Letters 17, 024005. <https://doi.org/10.1088/1748-9326/ac48b3>

Lukas, J., Barsugli, J., Doesken, N., Rangwala, I., Wolter, K., 2014. Climate change in colorado: A synthesis to support water resources management and adaptation. University of Colorado, Boulder, Colorado.

Mahoney, M.J., Johnson, L.K., Silge, J., Frick, H., Kuhn, M., Beier, C.M., 2023. Assessing the performance of spatial cross-validation approaches for models of spatially structured data. <https://doi.org/10.48550/arXiv.2303.07334>

Mahony, C.R., Wang, T., Hamann, A., Cannon, A.J., 2022. A global climate model ensemble for downscaled monthly climate normals over North America. International Journal of Climatology 42, 5871–5891. <https://doi.org/10.1002/joc.7566>

Malanson, G.P., Cairns, D.M., 1997. Effects of dispersal, population delays, and forest fragmentation on tree migration rates. Plant Ecology 131, 67–79. <https://doi.org/10.1023/A:1009770924942>

McCune, B., 2007. [Improved estimates of incident radiation and heat load using non-parametric regression against topographic variables](https://www.jstor.org/stable/4499284). Journal of Vegetation Science 18, 751–754.

McCune, B., Keon, D., 2002. [Equations for potential annual direct incident radiation and heat load](https://www.jstor.org/stable/3236745). Journal of Vegetation Science 13, 603–606.

Mitton, J.B., Grant, M.C., 1996. Genetic variation and the natural history of quaking aspen. BioScience 46, 25–31. <https://doi.org/10.2307/1312652>

Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. Agricultural and Forest Meteorology 92, 211–225. <https://doi.org/10.1016/S0168-1923(98)00102-6>

Nigro, K.M., Rocca, M.E., Battaglia, M.A., Coop, J.D., Redmond, M.D., 2022. Wildfire catalyzes upward range expansion of trembling aspen in southern Rocky Mountain beetle-killed forests. Journal of Biogeography 49, 201–214. <https://doi.org/10.1111/jbi.14302>

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37–42. <https://doi.org/10.1038/nature01286>

Pedersen, E.J., Miller, D.L., Simpson, G.L., Ross, N., 2018. Hierarchical generalized additive models: an introduction with mgcv. <https://doi.org/10.7287/peerj.preprints.27320v1>

Peet, R.K., 1981. Forest vegetation of the colorado front range. Vegetatio 45, 375.

Pepin, N., Bradley, R.S., Diaz, H.F., Baraer, M., Caceres, E.B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M.Z., Liu, X.D., Miller, J.R., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schöner, W., Severskiy, I., Shahgedanova, M., Wang, M.B., Williamson, S.N., Yang, D.Q., Mountain Research Initiative EDW Working Group, 2015. Elevation-dependent warming in mountain regions of the world. Nature Climate Change 5, 424–430. <https://doi.org/10.1038/nclimate2563>

Powers, S.M., Hampton, S.E., 2019. Open science, reproducibility, and transparency in ecology. Ecological Applications 29, e01822. <https://doi.org/10.1002/eap.1822>

PRISM Climate Group, 2021. [Monthly 30-year climate normals (1981-2010)](https://prism.oregonstate.edu/normals/).

R Core Team, 2022. [R: A language and environment for statistical computing](http://www.R-project.org). R Foundation for Statistical Computing, Vienna, Austria.

Randin, C.F., Ashcroft, M.B., Bolliger, J., Cavender-Bares, J., Coops, N.C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., Price, B., Rocchini, D., Schaepman, M., Schmid, B., Verburg, P., Wilson, A., Woodcock, P., Yoccoz, N., Payne, D., 2020. Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. Remote Sensing of Environment 239, 111626. <https://doi.org/10.1016/j.rse.2019.111626>

Rehfeldt, G.E., Crookston, N.L., Warwell, M.V., Evans, J.S., 2006. Empirical analyses of plant-climate relationships for the western united states. International Journal of Plant Sciences 167, 11231150.

Rehfeldt, G.E., Ferguson, D.E., Crookston, N.L., 2009. Aspen, climate, and sudden decline in western USA. Forest Ecology and Management 258, 2353–2364. <https://doi.org/10.1016/j.foreco.2009.06.005>

Rehfeldt, G.E., Worrall, J.J., Marchetti, S.B., Crookston, N.L., 2015. Adapting forest management to climate change using bioclimate models with topographic drivers. Forestry. doi: 10.1093/forestry/cpv019. <https://doi.org/10.1093/forestry/cpv019>

Riahi, K., Vuuren, D.P. van, Kriegler, E., Edmonds, J., O’Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change 42, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>

Rodman, K.C., Andrus, R.A., Butkiewicz, C.L., Chapman, T.B., Gill, N.S., Harvey, B.J., Kulakowski, D., Tutland, N.J., Veblen, T.T., Hart, S.J., 2021. Effects of bark beetle outbreaks on forest landscape pattern in the Southern Rocky Mountains, U.S.A. Remote Sensing 13, 1089. <https://doi.org/10.3390/rs13061089>

Rodman, K.C., Veblen, T.T., Battaglia, M.A., Chambers, M.E., Fornwalt, P.J., Holden, Z.A., Kolb, T.E., Ouzts, J.R., Rother, M.T., 2020. A changing climate is snuffing out post-fire recovery in montane forests. Global Ecology and Biogeography geb.13174. <https://doi.org/10.1111/geb.13174>

Rogers, P.C., Pinno, B.D., Šebesta, J., Albrectsen, B.R., Li, G., Ivanova, N., Kusbach, A., Kuuluvainen, T., Landhäusser, S.M., Liu, H., Myking, T., Pulkkinen, P., Wen, Z., Kulakowski, D., 2020. A global view of aspen: Conservation science for widespread keystone systems. Global Ecology and Conservation 21, e00828. <https://doi.org/10.1016/j.gecco.2019.e00828>

Seo, C., Thorne, J.H., Hannah, L., Thuiller, W., 2008. Scale effects in species distribution models: Implications for conservation planning under climate change. Biology Letters 5, 39–43. <https://doi.org/10.1098/rsbl.2008.0476>

U. S. Geological Survey, 2023. [USGS 3D elevation program digital elevation model](https://elevation.nationalmap.gov/arcgis/rest/services/3DEPElevation/ImageServer.).

Veblen, T.T., Hadley, K.S., Nel, E.M., Kitzberger, T., Reid, M., Villalba, R., 1994. Disturbance regime and disturbance interactions in a rocky mountain subalpine forest. Journal of Ecology 82, 125–135. <https://doi.org/10.2307/2261392>

Veblen, T.T., Kitzberger, T., Donnegan, J., 2000. Climatic and Human Influences on Fire Regimes in Ponderosa Pine Forests in the Colorado Front Range. Ecological Applications 10, 1178–1195. https://doi.org/<https://doi.org/10.1890/1051-0761(2000)010[1178:CAHIOF]2.0.CO;2>

Wang, T., Hamann, A., Spittlehouse, D., Carroll, C., 2016. Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. PLOS ONE 11, e0156720. <https://doi.org/10.1371/journal.pone.0156720>

Weiss, A., 2001. Topographic position and landforms analysis.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models 73, 3–36.

Worrall, J.J., Rehfeldt, G.E., Hamann, A., Hogg, E.H., Marchetti, S.B., Michaelian, M., Gray, L.K., 2013. Recent declines of Populus tremuloides in North America linked to climate. Forest Ecology and Management 299, 35–51. <https://doi.org/10.1016/j.foreco.2012.12.033>

Wright, M.N., Ziegler, A., 2017a. Ranger: A fast implementation of random forests for high dimensional data in c++ and r 77. <https://doi.org/10.18637/jss.v077.i01>

Wright, M.N., Ziegler, A., 2017b. Ranger: A fast implementation of random forests for high dimensional data in c++ and r 77. <https://doi.org/10.18637/jss.v077.i01>

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT : when can we trust the inferences? Methods in Ecology and Evolution 4, 236–243. <https://doi.org/10.1111/2041-210x.12004>

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32–35. <https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3>

Zhang, W., Calvo-Polanco, M., Chen, Z.C., Zwiazek, J.J., 2013. Growth and physiological responses of trembling aspen (Populus tremuloides), white spruce (Picea glauca) and tamarack (Larix laricina) seedlings to root zone pH. Plant and Soil 373, 775–786. <https://doi.org/10.1007/s11104-013-1843-5>

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A standard protocol for reporting species distribution models. Ecography 43, 1261–1277. <https://doi.org/10.1111/ecog.04960>

Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Analysing ecological data, Statistics for biology and health. Springer, New York ; London.

# Appendix A: ODMAP

## Overview

Here we describe the SDMs produced herein following the Overview, Data, Model, Assessment, Prediction (ODMAP) protocol for species distribution models ([Zurell et al., 2020](#ref-zurell2020StandardProtocolReporting)). Here, we first provide the Overview for our modeling, while the remaining ODMAP sections are detailed in Table S3.

The objectives of this modelling exercise are to (1) better explain the drivers of aspen’s distribution across the Southern Rocky Mountains, (2) map the area suitable for aspen, and (3) forecast the area suitable for aspen presence in the future under two different climate scenarios.

Table 3: ODMAP protocol information. Details on Data, Model, Assessment, Prediction. For Overview section, please refer to main text.

| ODMAP element | Contents |
| --- | --- |
| Overview |  |
| Authorship | Authors: Sarah J. Hart, Asha Paudel, and Maxwell C. Cook |
|  | Contact email: sarah.hart@colostate.edu |
|  | Title: |
|  | DOI: |
| Model objective | Objective: Inference, Mapping, and Forecasting |
|  | Target outputs: continuous occurrence probabilities, binary maps of potential presence, and maps of potential loss, gain, and stable habitat |
| Taxon | Quaking aspen (Populus tremuloides) |
| Location | Southern Rocky Mountains, US |
| Scale of analysis | Spatial extent (Lon/Lat): Longitude 103.86 ° W - 109.61 ° W, Latitude 35.28 ° N - 47.78 ° N |
|  | Spatial resolution: 90 x 90 m |
|  | Temporal resolution and extent: We modelled the presence of aspen based on remotely sensed maps generated from imagery collected in ca. 2019 (Cook et al. 2024). |
|  | Type of extent boundary: ecoregion (Southern Rocky Mountains) |
| Biodiversity data overview | Observation type: remotely sensed |
|  | Response type: presence/absence |
| Type of predictors | Climatic, topographic, edaphic |
| Conceptual model / Hypotheses | Based on previous studies, we tested climate, topography and edaphic factors as important environmental predictor variables for aspen presence in the Southern Rocky Mountains. For detailed hypotheses, see Table 1. |
| Assumptions | We make the following key assumptions:  (1) aspen is at pseudo-equilibrium with the environment  (2) topography, climate, and soil conditions are the key drivers of aspen's distribution  (3) classification error (average F1-score of 0.91) in maps of aspen occurrence were negligible  (4) the relationship fit under current conditions apply to future climate conditions (i.e., no change in key limiting processes) |
| SDM algorithms | Algorithms: SDMs were fit using four different algorithms: generalized linear models (GLMs), generalized additive models (GAMs), regularized gradient boosted trees (RGBTs), and random forests (RFs). |
|  | Model complexity: We chose different modelling parameters to optimize each statistical technique. |
|  | Ensembles: We combined the four SDMs to generate an ensemble prediction |
| Model workflow | Prior to model building, all predictor variables were standardized. Model hyperparameters were then tuned using a spatial cross-validation approach, with the best parameters selected using the area under the receiver operating characteristic curve (AUC). |
| Software | Software: All analyses were conducted using R version 4.3.1 (R Core Team 2023). |
|  | Code availability: All code is publicly available on GitHub (LINK) |
|  | Data availability: Data are available from XXXX |
| Data |  |
| Biodiversity data | Taxon names: Trembling aspen (Populus tremuloides) |
|  | Taxonomic reference system: US Department of Agriculture (USDA) Plant List of Attributes, Names, Taxonomy, and Symbols (PLANTS) Database |
|  | Ecological level: population |
|  | Data source: Aspen presence-absence data at 10x10 m spatial resolution were obtained from Cook et al. (in review). |
|  | Sampling design: The aspen cover dataset represents an entire census for the Southern Rocky Mountains. |
|  | Sample size: The aspen dataset consists of 4,312,302,640 10 x 10 m cells, of which 117,140,964 recorded the presence of aspen. |
|  | Ecoregion mask: We clipped all data to the Southern Rocky Mountain Ecoregion using data from the EPA's (2013) Level III Ecoregions of the Conterminous United States product. |
|  | Scaling: We aggregated the aspen presence-absence data to a 90 x 90 m cell size. |
|  | Data filtering: NA |
|  | Absence data: The Cook et al. (in review) map consists of both presence and absence data. |
| Data partitioning | To reduce computation time, we randomly selected 10,000 cells for model building and 10,000 cells model testing. To reduce the potential effects of spatial autocorrelation, all sample points were separated by a distance of at least 1 km. |
| Environmental data/predictor variables | Predictor variables:  (1) Topography: topographic position index, heat load index (HLI)  (2) Climate: We examined 34 biologically-relevancy climate variables, but ultimately limited our analyses to five climate variables (see Table 1)  (3) Soils: percent clay, percent soil organic matter, saturated water content |
|  | Data sources:  (1) Topography: 3DEP DEM (USGS 2023).  (2) Climate: AdaptWest Project (2022)  (3) Soils: POLARIS soil properties database (Cheney et al. 2019). |
|  | Spatial resolution and extent of raw data: All data were available for the entire study area.  (1) Topography: 30 m  (2) Climate: 1 km  (3) Soils: 30 m |
|  | Temporal resolution and extent of raw data:  (1) Topography: raw topographic data were collected over the period 2009-2023  (2) Climate: monthly and annual means for the periods 1981-2010, 2011-2040, 2041-2070, and 2071-2100  (3) Soils: represent National Cooperative Soil Survey data collected over the 1899 to 2019 period |
|  | Geographic projection of raw data:  (1) Topography: NAD83(HARN) / Conus Albers (EPSG:5071)  (2) Climate: Lambert Azimuthal Equal Area (EPSG:9820)  (3) Soils: WGS 84 (EPSG:4326) |
|  | Data processing: (1) Topography: We calculated the Heat Load Index (HLI) and topographic position index (TPI) using the spatialeco package (Evans and Murphy 2021). TPI was calculated using a 3 cell neighborhood (i.e., 90 x 90 m) and a 15 cell neighborhood (i.e., 450 x 450 m). TPI and HLI datasets were then aggregated to 90 m using the mean value and reprojected to UTM Zone 13N.  (2) Climate: We calculated ADI, GSP, PRATIO, and GSPDD5 following Rehfeldt et al. (2009). All climate variables were then downscaled to 250 m resolution using gradient and inverse distance squared (GIDS) interpolation and reprojected to UTM Zone 13N.  (3) Soils: Soil data were aggregated to 90 m using the mean value and reprojected to UTM Zone 13N. |
| Model |  |
| Variable pre-selection | To avoid collinearity between climate predictors, we initially screened the 34 climatic variables at their original resolution (i.e. 1 x 1 km). To this end, we calculated pairwise correlation coefficients and when |r|>0.75, we removed variables based on existing research (Table 4). Where evidence was similar, we used univariate random forest (RF) models to evaluate the potential explanatory power of each predictor. |
| Multicollinearity | Using the downscaled climate variables in combination with soil and topographic variables, we further reduced multicollinearity in our predictor dataset by calculating variable inflation factors (VIF) using the spatialRF package (Benito 2022). We then iteratively removed variables until VIF<5 for all variables. |
| Model settings | We fit generalized linear models (GLMs), generalized additive models (GAMS), and random forests (RFs), and regularized gradient boosted tree (RGBTs).  (1) GLMs were constructed using a logit link function and a binomial error distribution. For all variables, we included both linear and quadratic effects. Models were fit using a Lasso regularization approach. The only hyperparameter we tuned was the lasso penalty factor. We did not explore any interaction terms. GLMs were fit using the glmnet package (Friedman et al., 2010).  (2) GAMs were constructed using a logit link function and a binomial error distribution and fit restricted maximum likelihood. For all variables, we used thin plate regression splines that included a penalty term that allowed the model coefficient to be shrunk to zero. We set the bias dimensions term (k) to the default value of 10 and confirmed an adequate degree of complexity using diagnostic functions from the R package mgcv (Wood 2011). The only hyperparameter we tuned was the penalty factor. We did not explore any interaction terms. GAMs were fit using the mgcv package (Wood 2011).  (3) For the RF model, we tuned the minimum number of data points in a node that is required for the node to be split further and the number of variables to try at each split. RF models were fit using the R package ranger (Wright and Ziegler 2017).  (4) RGBT: For the RGBT model, we tuned the learning rate, number of variables to try at each split, proportion of the training dataset exposed to the fitting routine, the maximum depth of tree, minimum number of data points in a node required for the node to be split further, and the reduction in the loss function required to split further. RGBT were fit using the R package xgboost (Chen et al. 2023). |
| Model estimates | Using the R package DALEX (Biecek 2018), we determined variable importance using a model-agnostic permutation-based approach. In this approach, each variable is randomized and then AUC statistic is compared with =AUC for the full model (where data has not been randomized). We evaluated the relationship between aspen presence and each predictor variable using accumulated local effects (ALE) profiles, which were generated using the ALEPlot package (Aplet, 2018). |
| Model averaging / Ensembles | We calculated a weighted probability of occurrence from all four SDMs. Weights assigned were based on the AUC statistic. |
| Non-independence | We evaluated the potential effects of spatial autocorrelation on our models' predictive ability using a spatially clustered cross-validation approach using the R package spatialsample (Mahoney et al. 2023) |
| Threshold selection | Binary predictions were derived by maximizing Youden's J statistic, which balances sensitivity and specificity (Youden 1950). |
| Assessment |  |
| Performance statistics | We used the eight performance statistics to evaluate model fit: overall accuracy, F measure, kappa, precision, recall, AUC, sensitivity and specificity. |
| Plausibility checks | We checked model plausibility by assessing accumulated local effects plots and examining spatial patterns. |
| Prediction |  |
| Prediction output | For further analyses, we used continuous predictions of occurrence probability, as well as predicted presence-absence. |
| Uncertainty quantification | We account for algorithmic uncertainty by applying an ensemble approach averaging over four different SDM algorithms. We account for uncertainty in future projections of climate by examining two different scenarios (SSP2-4.5 and SSP5-8.5) and using an ensemble forecast of future climate generated from 8 AOGCMs previously identified to be appropriate for regional climate-change analyses conducted in North America. |

# Appendix B: Collinearity

Table 4: Climate variables considered for inclusion in SDMs and modeling notes.

| Variable | Description | Drop out loss | Order | Modeling notes |
| --- | --- | --- | --- | --- |
| ADI | annual dryness index: (DD5^0.5)/MAP | 15.602186 | 1 | retain - identified as important predictor by Rehfeldt et al. (2009) and (2015). |
| PRATIO | mean precipitation ratio:  GSP/MAP | 1.548609 | 2 | retain - identified as important predictor by Rehfeldt et al. (2009) |
| DD5 | degree-days above 5 °C | 2.497007 | 3 | remove - identified as important predictor by Worrall et al. (2013) and Greer et al. (2016, but high correlation (r≥0.75) with ADI |
| TMAX | Mean maximum temperature in warmest month | 10.658398 | 4 | removed - identified as important predictor by Rehfeldt et al. (2009), Worrall et al. (2013), and Greer et al. (2016), but highly correlated (r≥0.75) with ADI |
| GSPDD5 | growing season precipitation to degree day ratio: (GSP\*DD5/1000) | 2.776479 | 5 | retain - identified as important predictor by Rehfeldt et al. (2009) |
| MAP | mean annual precipitation (mm) | 4.394327 | 6 | removed - Identified as important predictor by Worrall et al. (2013), but high correlation (r≥0.75) with ADI |
| PPT\_wt | mean winter precipitation (mm) | 2.698400 | 7 | removed - identified as important predictor by Worrall et al. (2013), but highly correlated (r≥0.75) with ADI |
| GSP | growing season (Apr - Sep) precipitation (mm) | 1.787878 | 8 | retain - Idenified as important predictor by Worrall et al. (2013) |
| MWMT | mean warmest month temperature (°C) | 3.448696 | 9 | removed - identified as important predictor by Rehfeldt et al. (2015), but high correlation (r≥0.75) with ADI |
| TD | difference between MCMT and MWMT (°C) | 1.508079 | 10 | retain - identified as important predictor by Rehfeldt et al. (2015) and Worrall et al. (2013) |
| CMI | Hogg’s climate moisture index (mm) | 1.690420 | 11 | remove - high correlation (r≥0.75) with ADI |
| DD\_18 | degree-days below 18 °C | 2.297719 | 12 | remove - high correlation (r≥0.75) with ADI |
| PAS | mean precipitation as snow (mm) between August in previous year and July in current year | 1.713522 | 13 | remove - high correlation (r≥0.75) with ADI |
| PPT\_at | mean autumn precipitation (mm) | 1.578116 | 14 | remove - high correlation (r≥0.75) with ADI |
| DD\_0 | degree-days below 0 °C | 1.767489 | 15 | retain |
| DD1040 | degrees-days above 10 °C and below 40 °C | 11.391070 | 16 | remove - high correlation (r≥0.75) with ADI |
| CMD | Hargreaves climatic moisture deficit (mm) | 3.316684 | 17 | remove - high correlation (r≥0.75) with ADI |
| Eref | Hargreave's reference evapotranspiration (mm) | 1.381140 | 18 | remove - high correlation (r≥0.75) with ADI |
| PPT\_sp | mean spring precipitation (mm) | 1.851747 | 19 | remove - high correlation (r≥0.75) with ADI |
| MCMT | mean coldest month temperature (°C) | 1.861349 | 20 | remove - high correlation (r≥0.75) with DD\_0 |
| DD18 | degree-days above 18 °C | 1.649392 | 21 | remove - high correlation (r≥0.75) with DD\_0 |
| MAT | mean annual temperature | 1.624926 | 22 | remove - high correlation (r≥0.75) with ADI |
| Tave\_wt | mean winter temperature (°C) | 3.142040 | 23 | remove - high correlation (r≥0.75) with DD\_0 |
| Tave\_sm | mean summer temperature (°C) | 2.961350 | 24 | remove - high correlation (r≥0.75) with ADI |
| Tave\_sp | mean spring temperature (°C) | 1.594825 | 25 | remove - high correlation (r≥0.75) with ADI |
| EMT | extreme minimum temperature (°C) | 2.124071 | 26 | remove - high correlation (r≥0.75) with DD\_0 |
| Tave\_at | mean autumn temperature (°C) | 2.876604 | 27 | remove - high correlation (r≥0.75) with ADI |
| bFFP | Julian date on which the frost free period beings | 15.327335 | 28 | remove - high correlation (r≥0.75) with ADI |
| NFFD | mean annual number of frost free days | 1.241642 | 29 | remove - high correlation (r≥0.75) with DD\_0 |
| EXT | extreme maximum temperature (°C) | 1.744464 | 30 | remove - high correlation (r≥0.75) with ADI |
| FFP | mean annual length of the frost-free period (days) | 1.714604 | 31 | remove - high correlation (r≥0.75) with DD\_0 |
| PPT\_sm | mean summer precipitation (mm) | 1.849692 | 32 | remove - high correlation (r≥0.75) with TD |
| eFFP | Julian date on which the frost free period ends | 1.798223 | 33 | remove - high correlation (r≥0.75) with DD\_0 |
| RH | mean annual relative humidity (%) | 1.560396 | 34 | retain |
|  |  | 1.770161 |  |  |

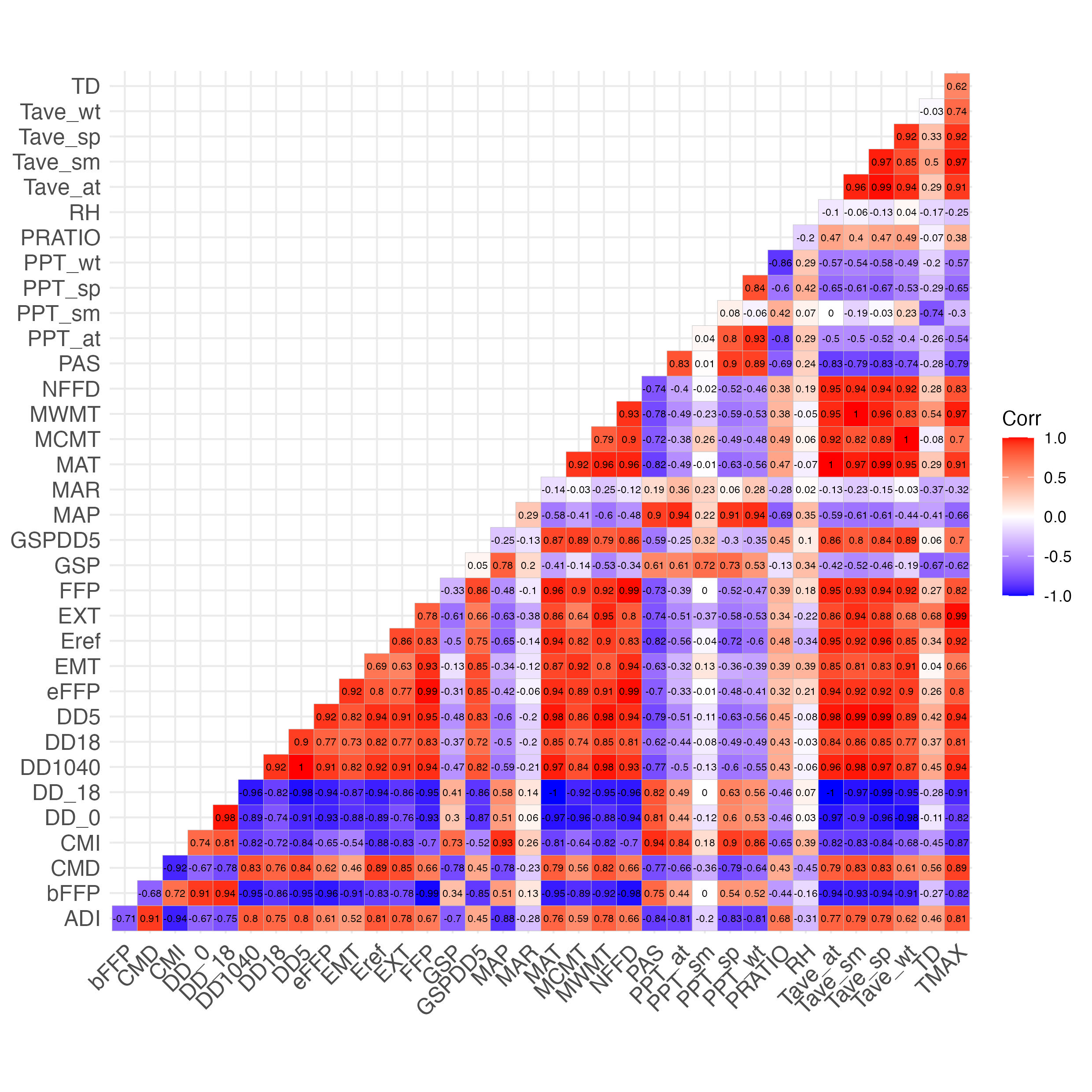


Figure 6: Correlation coefficients between pairs of climate predictor variables examined for inclusion in SDM

# Appendix C: Supplemental figures

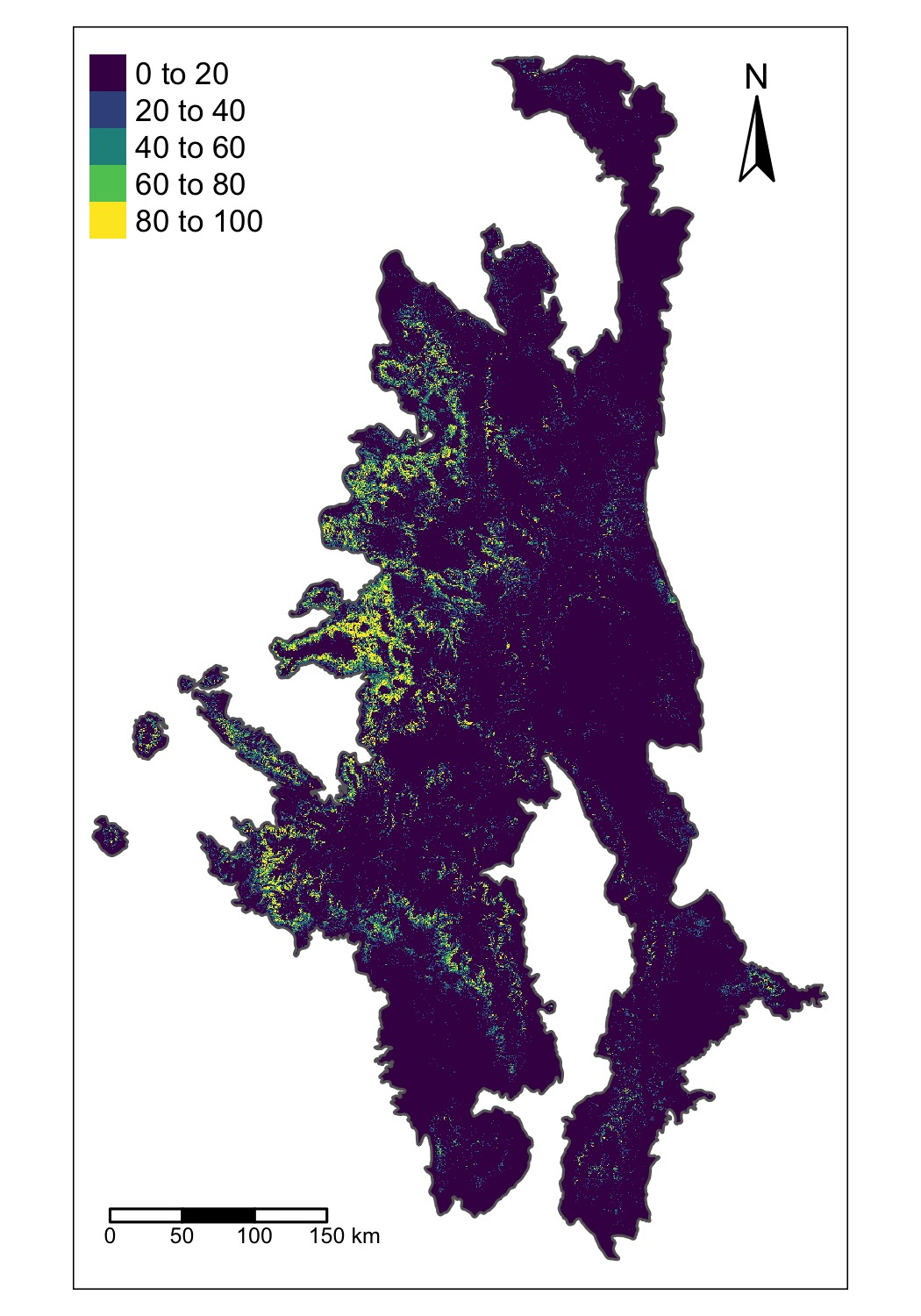


Figure 7: The percent cover of aspen within a 90 x 90 m pixel for the Southern Rocky Mountains. Data were produced by aggregating the 10 m resolution map of aspen presence-absence from Cook et al. (in review).

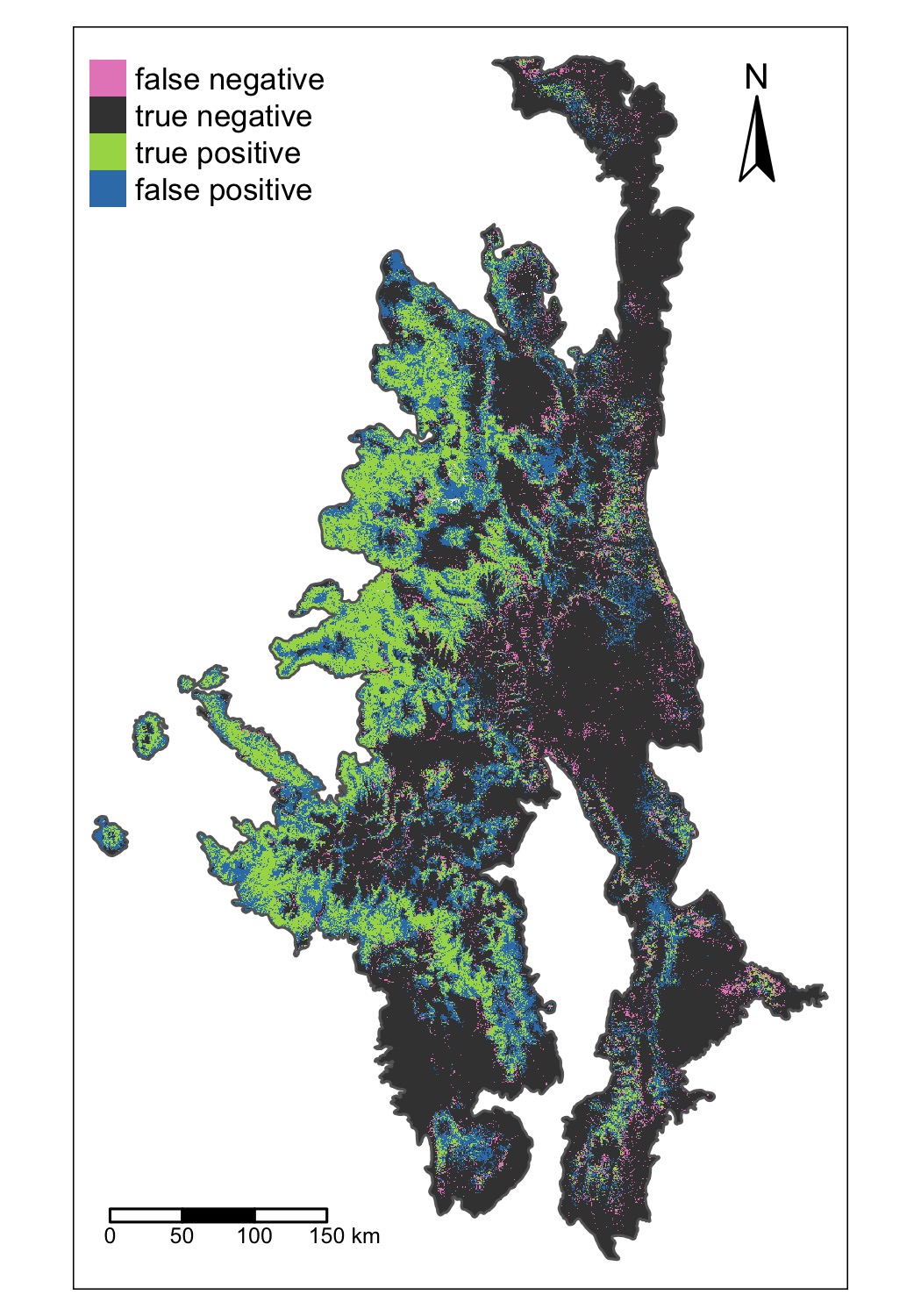


Figure 8: Spatial patterns of missclassification for the ensemble model.

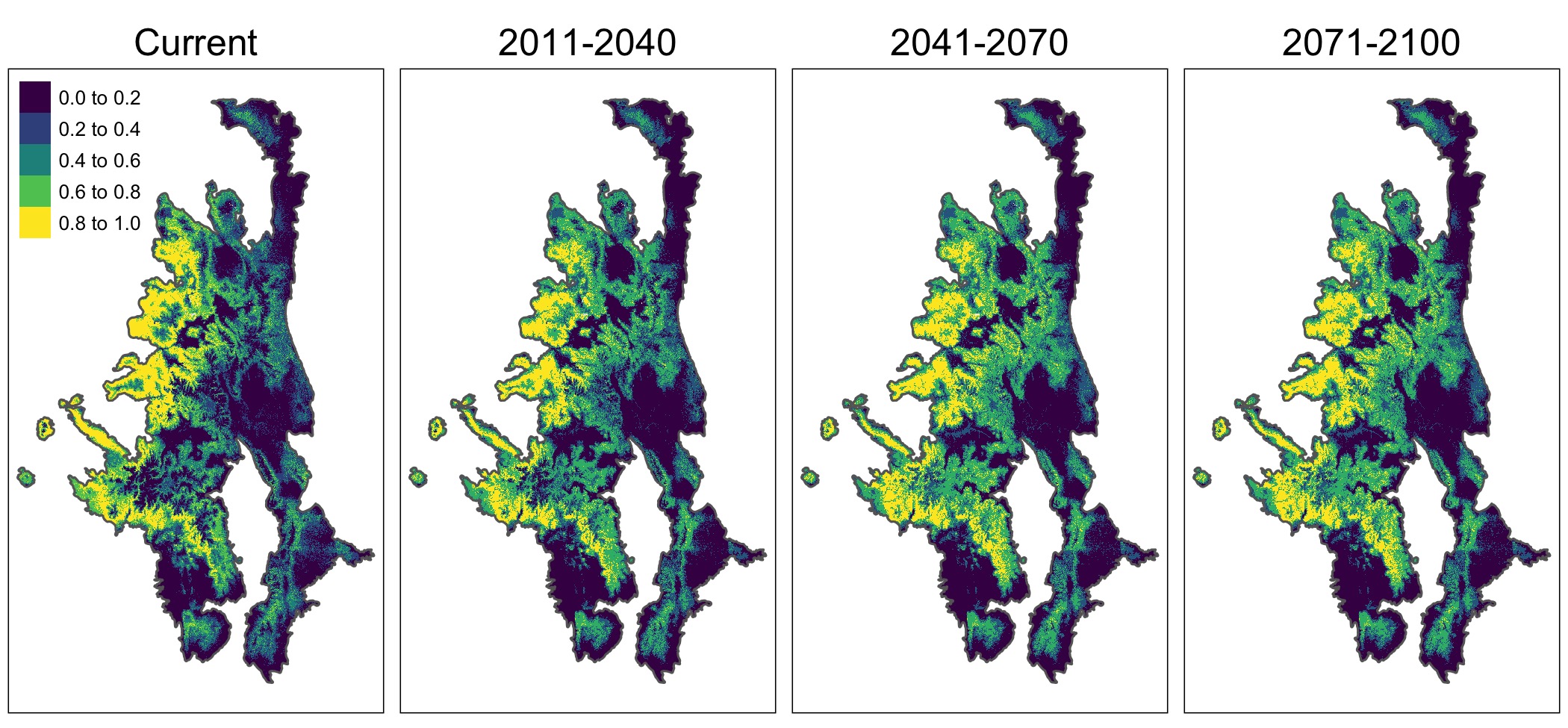


Figure 9: The ensemble projection of aspen habitat suitability under current conditions and projections for future periods based on the projected climate for the SSP5-8.5 scenario.

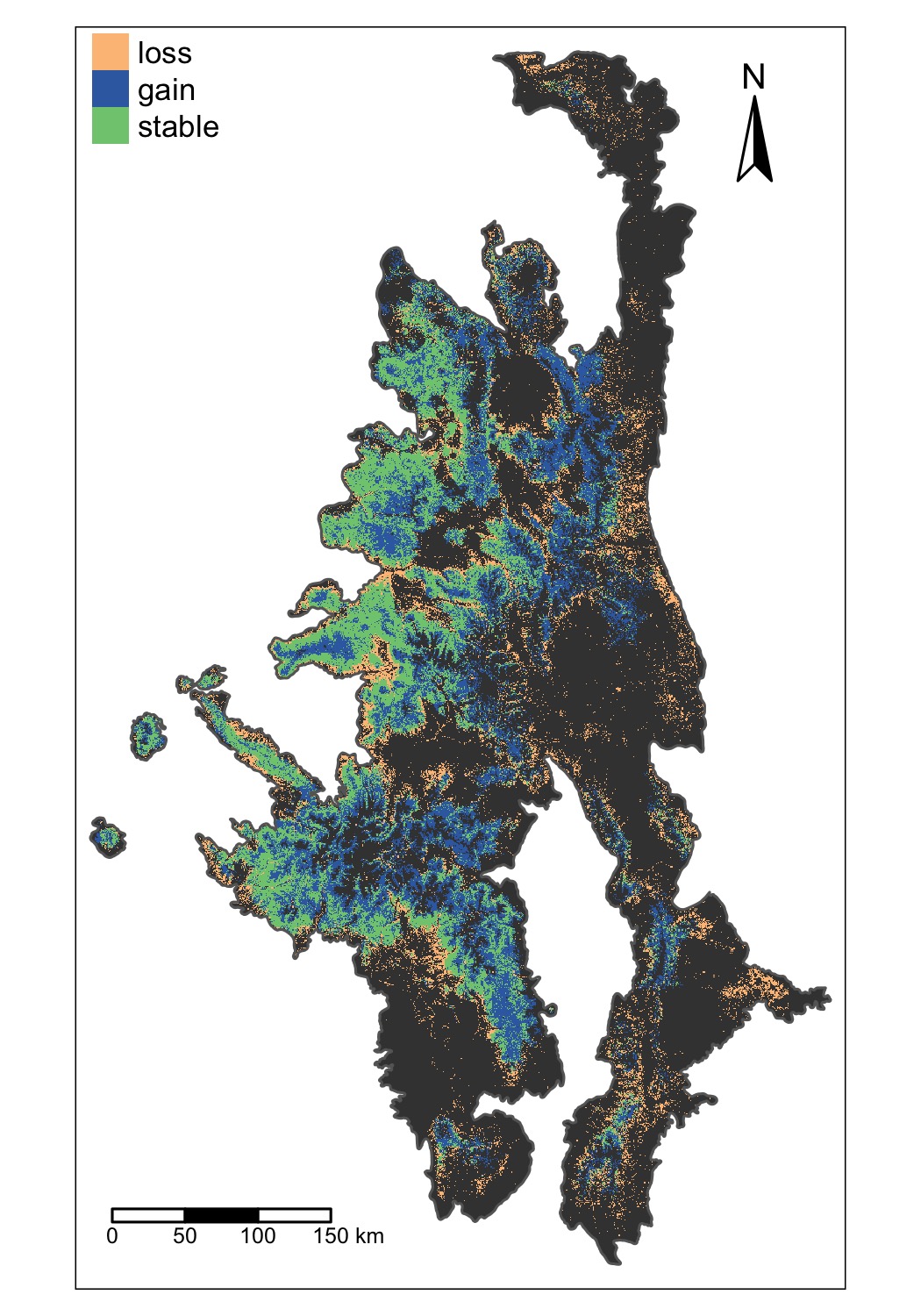


Figure 10: The distribution of pixels where the ensemble SDM forecasts aspen may be lost, gained, or stable by 2100 under the SSP2-4.5 scenario.



Figure 11: Distance to the nearest existing aspen patch for pixels were future climate may promote aspen expansion by 2100 under the SSP2-4.5 scenario.

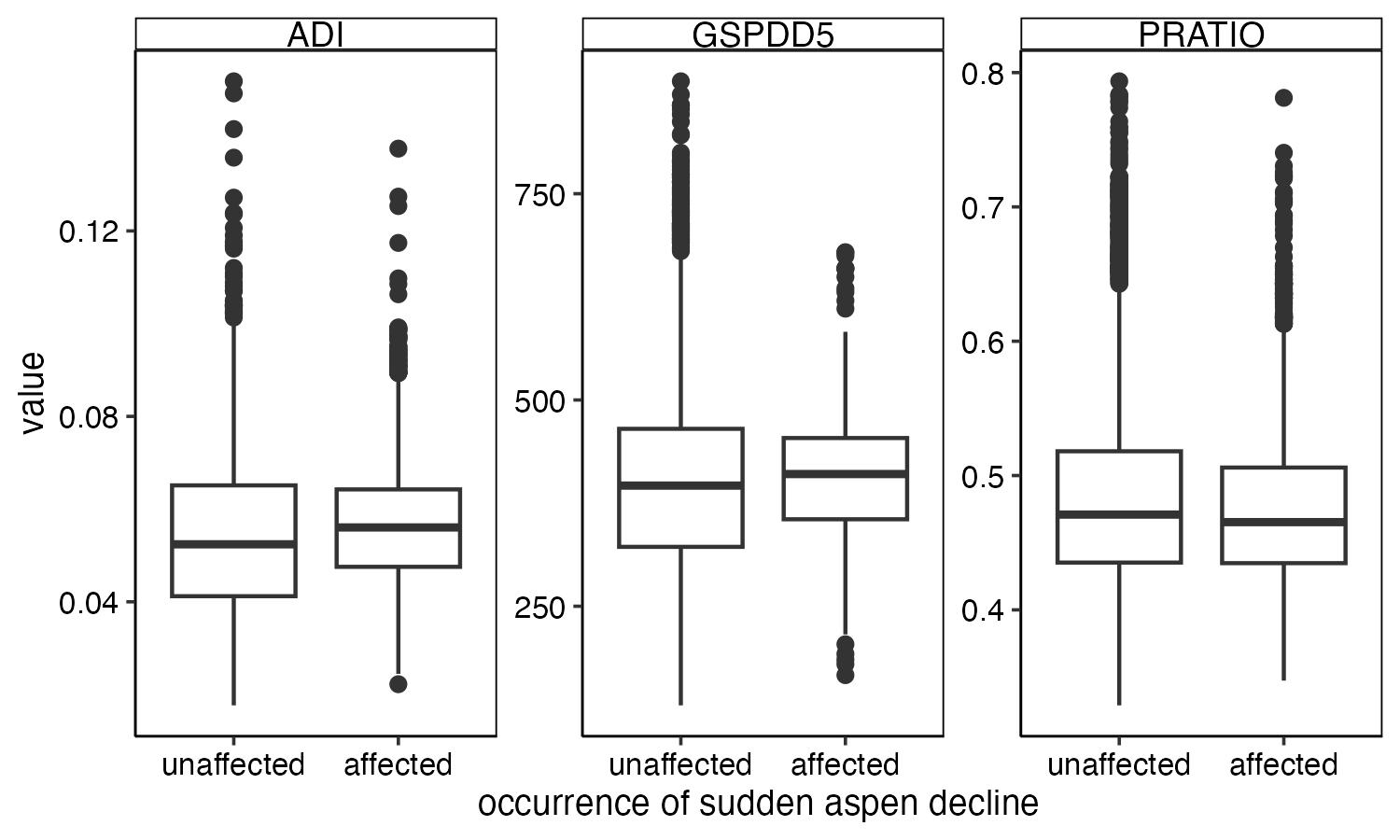


Figure 12: Boxplots illustrating the relationship between areas affected by sudden aspen decline and healthy aspen forests and climate variables.

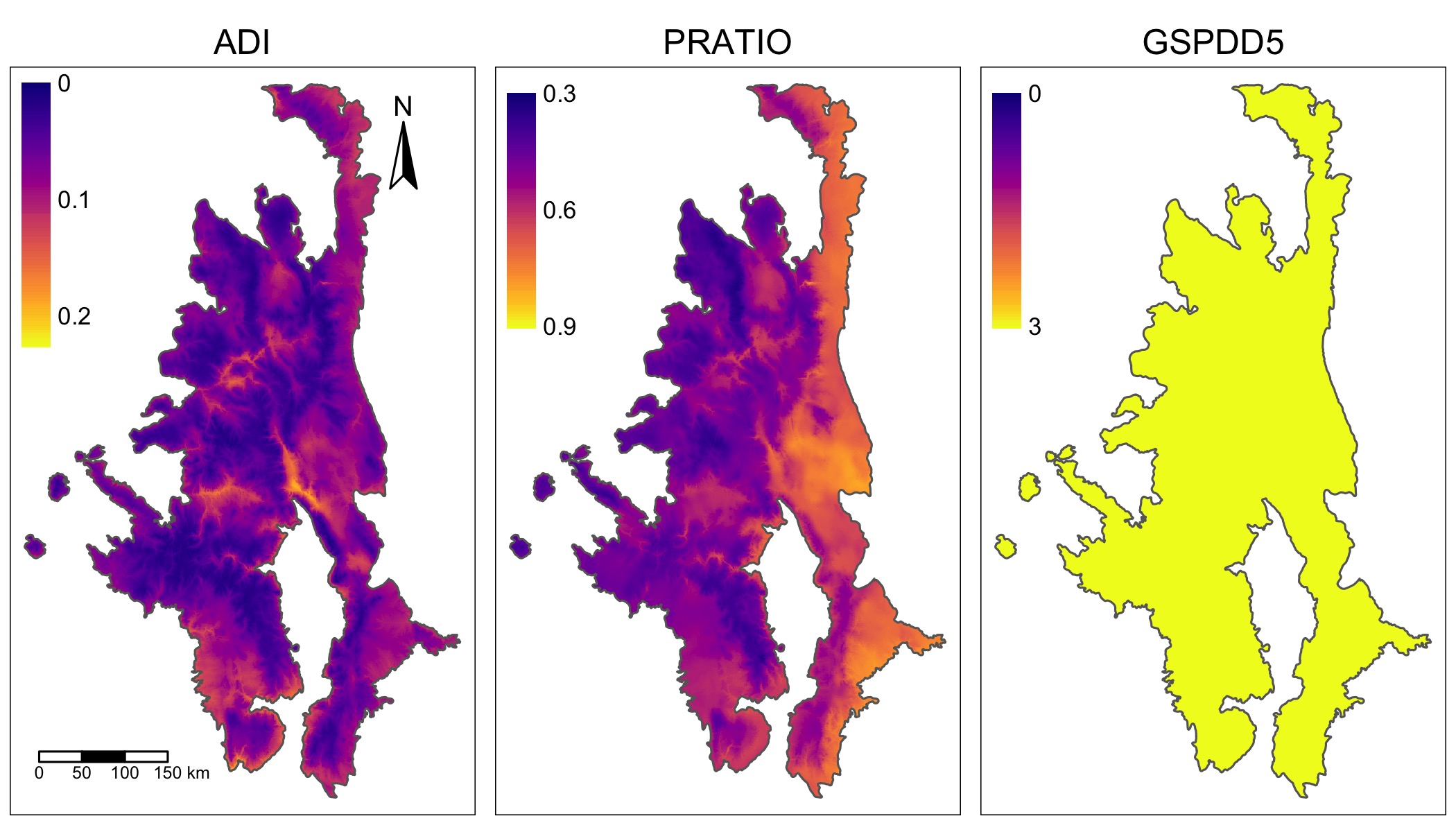


Figure 13: Maps illustrating the spatial patterning of contemporary climate.