# Supplemental Tables and Figures

## Overview of species distribution modeling protocols

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 S1. 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 S1: 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: Climate-driven decreases in aspen's distribution and opportunities for future expansion across the Southern Rocky Mountains |
|  | DOI: 10.5061/dryad.t76hdr88v |
| 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: http://datadryad.org/stash/share/gGubdiD6hCORznmbeHB6gkWoxodyj25Soa3s3yfOfHg |
|  | Data availability: http://datadryad.org/stash/share/gGubdiD6hCORznmbeHB6gkWoxodyj25Soa3s3yfOfHg |
| 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. (2024). |
|  | 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. (2024) 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 5). 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 the 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. |

Table 2: 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 |  |  |

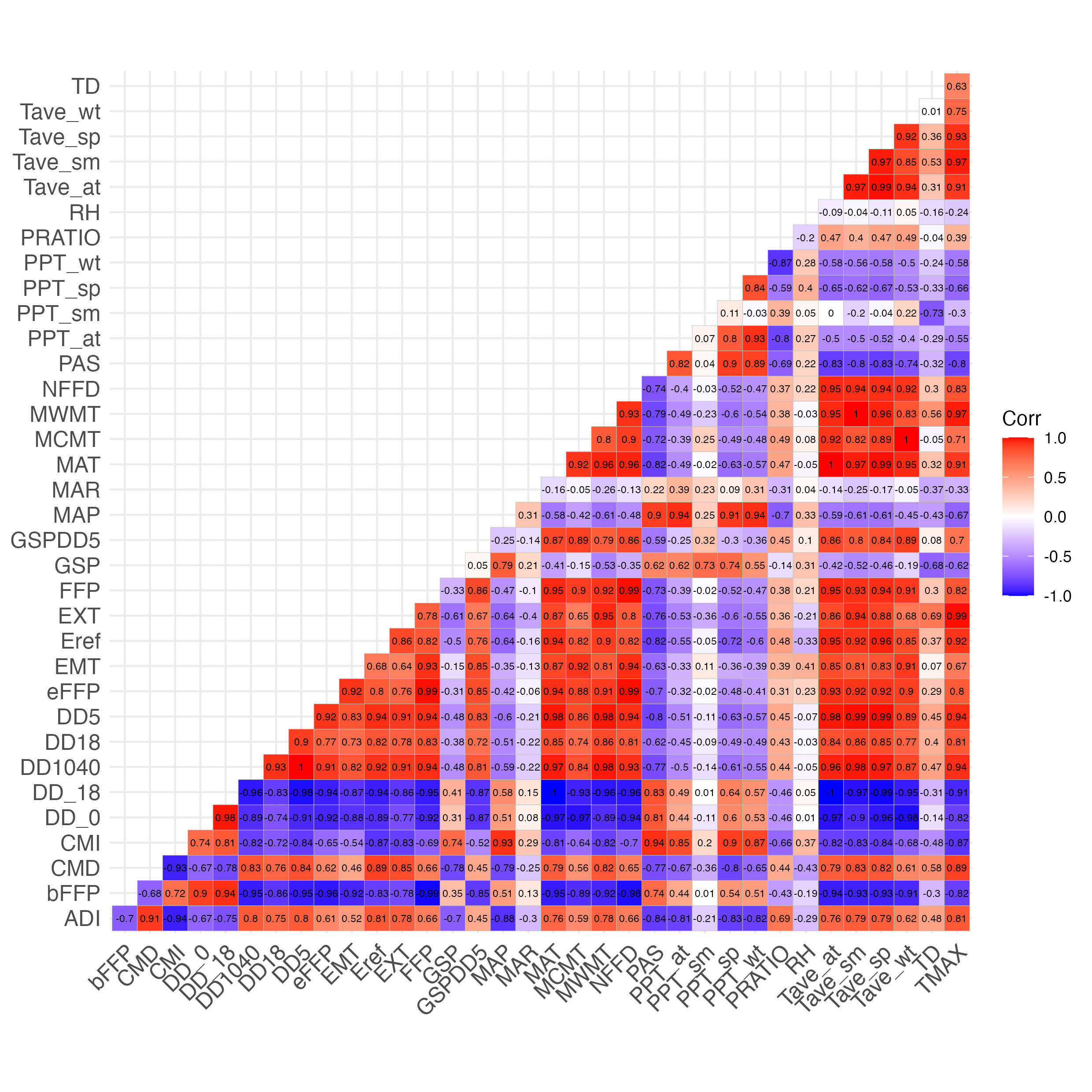


Figure S1: Correlation coefficients between pairs of climate predictor variables examined for inclusion in SDM

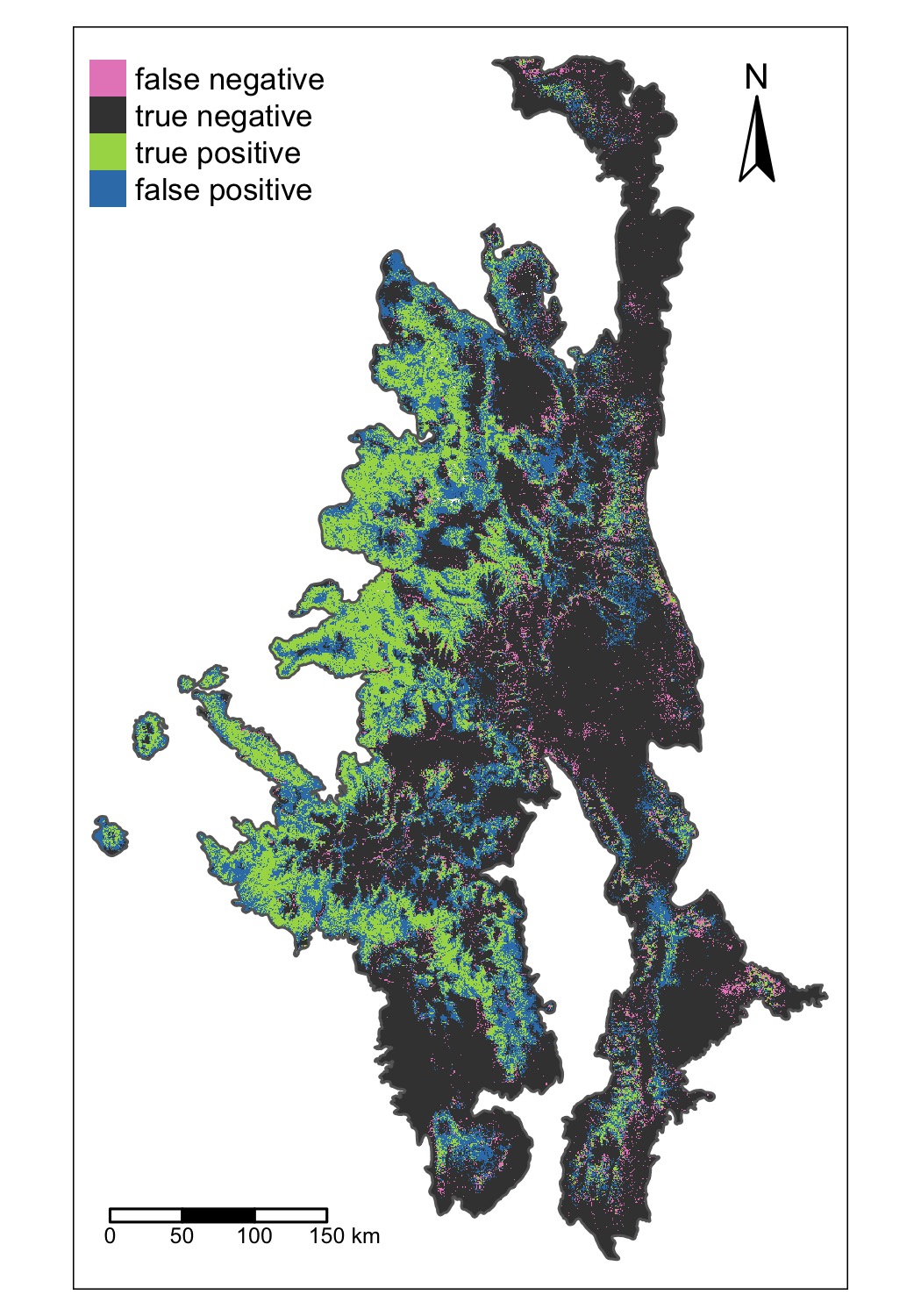


Figure S2: Spatial patterns of missclassification for the ensemble model.

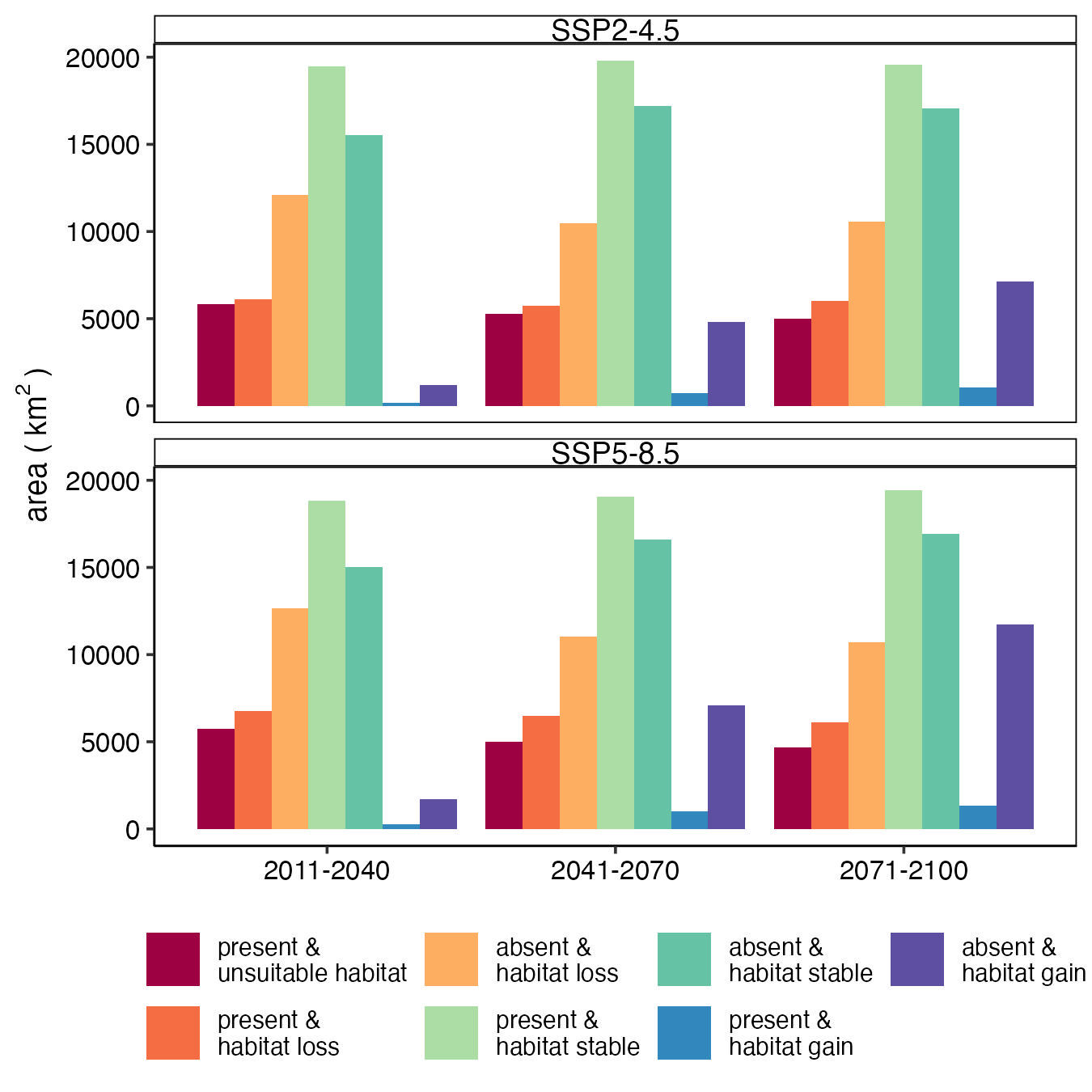


Figure S3: Barplots illustrating the temporal patterns in the areas were aspen my lost, gained, or remain stable based on the current distribution of aspen and the ensemeble SDM forecast of future aspen habitat suitability.

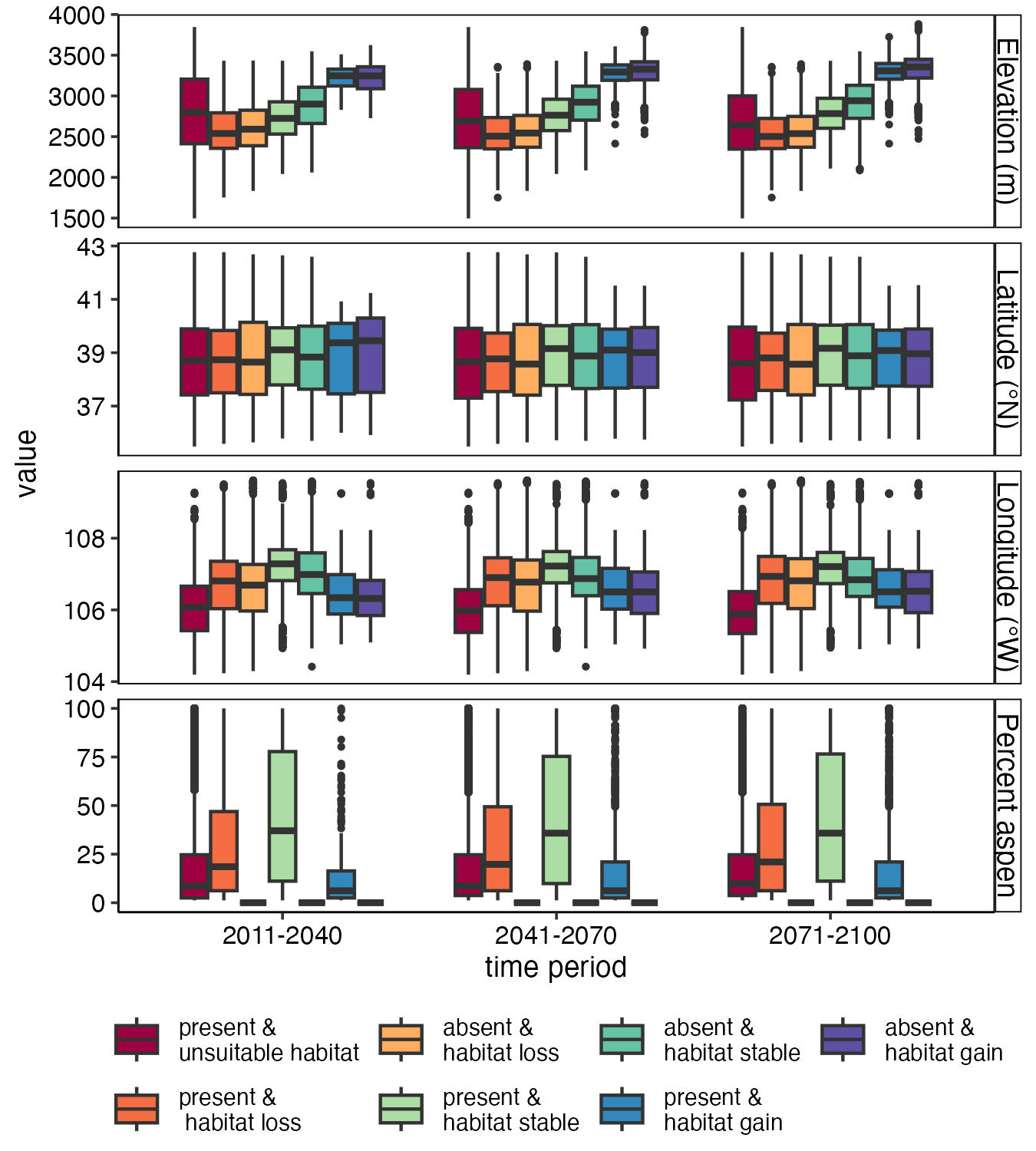


Figure S4: Boxplots illustrating spatial patterns in the areas were aspen my lost, gained, or remain stable based on the current distribution of aspen the ensemeble SDM forecast of future aspen habitat suitability under the SSP2-45 scenario.

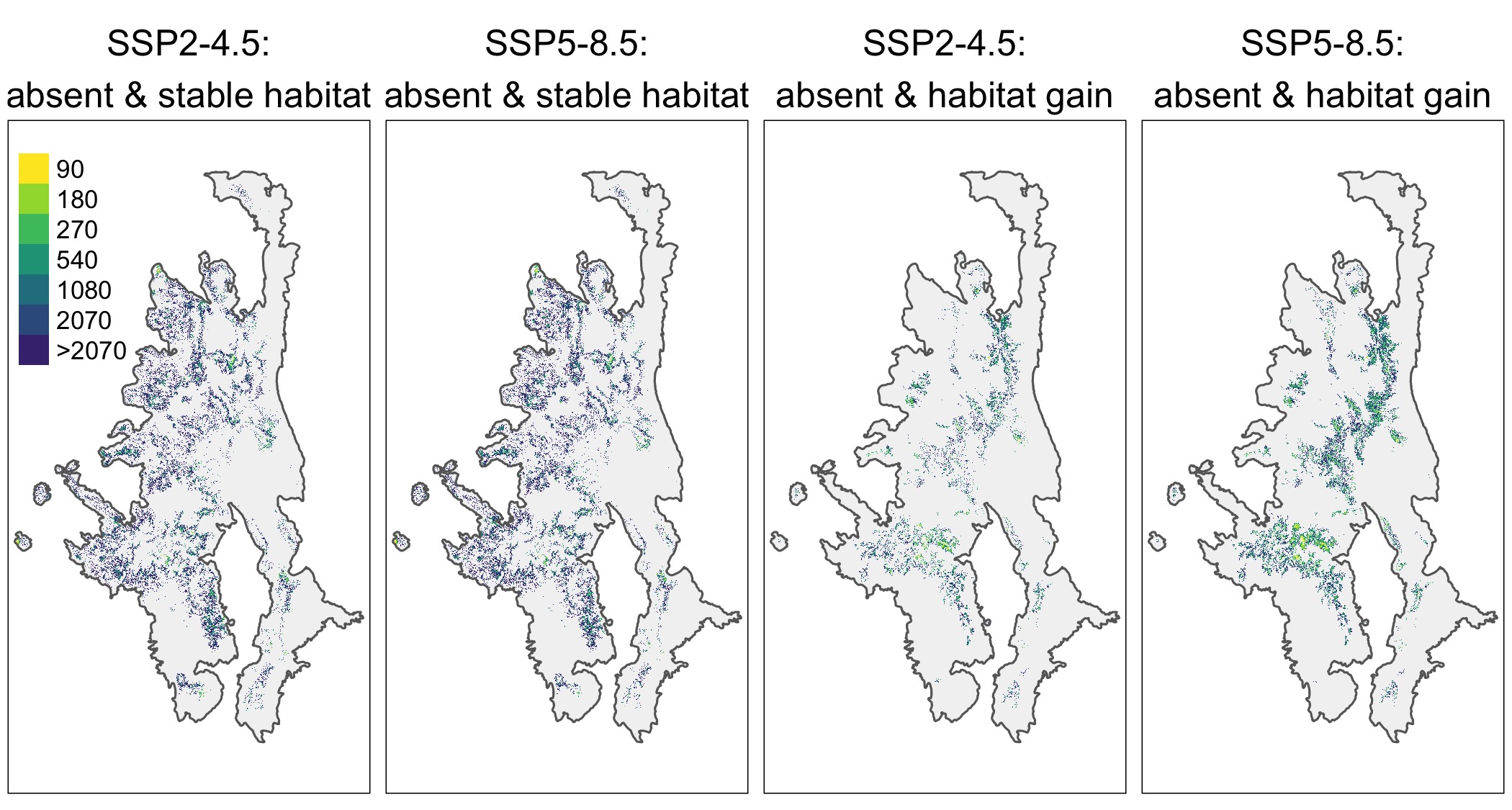


Figure S5: Distance to the nearest existing aspen patch for pixels were future climate may promote aspen expansion by 2100 under the SSP2-4.5 scenario.

# 

# References

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>