Supplementary Information

Supplementary Methods

Multi-state models

For states $r, s \in B, M, P, T$ and time $t, \Delta t >= 0$, transition probabilities (p_{rs}) are defined as the probability that a plot in state r at time t is in state s at time $t + \Delta t$ and can be written as:

$$P_{r,s}(t + \Delta t) = P(S_{t+\Delta t} = s | S_t = r).$$

In a four-state transition model, the transition probability matrix $P(t + \Delta t)$, hereafter simplified to $P(t)$, is a 4×4 matrix, where the rows are the current states and the columns the future states, containing the transition probabilities $p_{rs}(t)$ for a specified time interval. For a time-homogeneous model, $P(t)$ is solved by taking the matrix exponential of the intensity matrix Q scaled by the time interval:

$$P(t) = e^{tQ}.$$

The intensity matrix Q contains transition intensities $q_{r,s}$ which represent the instantaneous risk of moving from state r to state s:

$$q_{r,s} = \lim_{\Delta \to 0} \frac{P(Y_{t+\Delta} = s | Y_t = r)}{\Delta}, \text{ on off-diagonal elements,}$$

$$q_{r,r} = -\sum_{s \neq r} q_{rs}, \text{ on diagonal elements.}$$

We can define transition-specific hazard regression models for those states $r, s \in B, M, P, T$ between which a direct transition is possible according to the specified multi-state process (Fig. 2 in main text). The intensities

$q_{r,s}$ can be modelled as a product of a baseline hazard $q_{rs.0}$ and a log-linear effect of the explanatory variables $x(t)$ and their coefficients β_{rs}:

$$q_{rs}(t) = q_{rs}(t|x(t)) = q_{rs.0}(t)exp(\beta'_{rs}x(t)).$$

In this model, $q_{rs.0}(t)$ is a baseline hazard function that describes the risk for a reference plot i with environment $x_i(t) = 0$, and $exp(\beta'_{rs}x(t))$ is the relative increase or decrease in risk associated with the set of characteristics $x_i(t)$. This model allows one to include the effect of time-dependent covariates on transition intensities and therefore to relax the time homogeneity assumption of Markov models. Time-dependent covariates, such as climate and disturbances, are assumed to be piecewise-constant, i.e., the hazard is constant within a specified time interval $[t, t + \Delta t]$ and depends on the covariate value at t, but is allowed to change between the intervals.

Estimation of model parameters can be obtained by maximising the log-likelihood function using the transition probability matrix. The contribution of plot i at time j to the likelihood is given by:

$$LL_i(\theta|s, x) = \prod_{j=1}^{J} P(S_j = s_j|S_{j-1} = s_{j-1}, \theta, x),$$

where θ is the vector with all model parameters, x denotes the vector with the covariate values, and s denotes the observed state trajectory $s_1, ..., s_J$ at times $t_1, ..., t_J$. The full likelihood function is the product of the contributions of all N plots:

$$LL(\theta) = \prod_{i=1}^{N} LL_i(\theta|s, x).$$

Performance of candidate models

We fitted all models on the full data sets but also used cross-validation to estimate the predictive performance on held-out data. We used two statistics, the area under the receiver operating characteristic (ROC) curve (AUC) and the logarithmic scoring rule (LS), to assess the agreement between the observed state and the models' predictions. The AUC is a popular performance metric for binary classifiers that measures the probability that a randomly drawn member of state s has a lower estimated probability of belonging to state r than a randomly drawn member of state r. The AUC ranges from 0 to 1, where a score of 1 indicates perfect discrimination, while a score of 0.5 is as good as random. Hand & Hill (2001) has extended the AUC method to multi-class problems. For any pair of states r and s, we can compute $\hat{A}(r|s)$, the probability that a randomly drawn member of state s has a lower estimated probability of belonging to state r than a randomly drawn member of state r. We can measure the discrimination rate between all pairs of states by computing the pairwise AUC:

$$\hat{A}(r,s) = [\hat{A}(r|s) + \hat{A}(s|r)]/2.$$

Averaging the pairwise AUC gives the overall multi-class AUC (hereafter mAUC) of the model:

$$mAUC = \frac{2}{c(c-1)} \sum_{r<s} \hat{A}(r,s).$$

The LS was proposed by Good (1952) and is often used in weather forecasts (Gneiting & Raftery, 2007). While AUC is a function of different classification thresholds, LS measures the degree to which predicted probabilities are close to the observed outcomes. We computed a global score for each model:

$$LS = \frac{1}{N} \sum_{i=1}^{N} -log(P(S_i = s_i)),$$

where S_i is the random variable describing the state of the forest in the i^{th} plot and s_i is the observed state. So, LS only depends upon the predicted probability of the realised state and not on the probabilities assigned to the other possible states. The score is very sensitive to incorrect predictions: if a model predicted the observed state with a probability of 100%, the score for that plot would be 0, while if a probability of zero was assigned to the observed state, the score would go to infinity. Hence, this sensitivity emphasises the differences between model predictions and strongly penalises a model that only gives high probabilities to self-transitions.

To assess the quality of prediction for the four states individually, we computed LS for each state r where we summed the predicted probabilities $P(S_i = r)$ if the observed state is indeed r and $1 - P(S_i = r)$ otherwise.

We evaluated and compared the predictive performance of our five models using the overall mAUC and the pairwise AUCs, as well as the overall LS and the state-specific LS. These metrics were estimated using stratified K-fold cross-validation (Burnham, Anderson, & Burnham, 2002). We first stratified the data set by bioclimatic domains to ensure that each fold was representative of the plot geographical distribution and randomly split the data set in k=10 folds. The cross-validation process was repeated k times, during which $k-1$ folds were used to train the models and the remaining fold was used to validate the model predictions against the observed state transitions. The cross-validated performance metrics were then averaged for each model.

Supplementary tables

Table S1. List of the 46 species included in the analyses, their frequency and their corresponding group. The frequency corresponds to the number of forest plots in which they were observed. The species groups were defined using their trait values and knowledge of species ecology (see Brice, Cazelles, Legendre, & Fortin, 2019 for details).

Species name	**Vernacular name**	**Frequency**
Boreal		
Abies balsamea	Balsam fir	7870
Larix laricina	Tamarack	551
Picea glauca	White spruce	4181
Picea mariana	Black spruce	6082
Pinus banksiana	Jack pine	1339
Pioneer		
Betula papyrifera	White birch	5863
Betula populifolia	Grey birch	226
Populus balsamifera	Balsam poplar	216
Populus deltoides	Cottonwood	2
Populus grandidentata	Large tooth aspen	582
Populus tremuloides	Trembling aspen	2504
Prunus pensylvanica	Pin cherry	1495
Salix sp.	Willow	499
Sorbus sp.	Mountain-ash	515
Temperate		
Acer negundo	Manitoba maple	1
Acer nigrum	Black maple	3
Acer pensylvanicum	Striped maple	719
Acer rubrum	Red maple	3273
Acer saccharinum	Silver maple	23
Acer saccharum	Sugar maple	2190
Acer spicatum	Mountain maple	206
Amelanchier sp.	Serviceberry	33
Betula alleghaniensis	Yellow birch	2582
Carpinus caroliniana	Blue beech	6

(continued)

Species name	**Vernacular name**	**Frequency**
Carya cordiformis	Bitternut hickory	10
Fagus grandifolia	American beech	928
Fraxinus americana	White ash	261
Fraxinus nigra	Black ash	465
Fraxinus pennsylvanica	Red ash	33
Juglans cinerea	Butternut	16
Ostrya virginiana	Ironwood	493
Picea rubens	Red spruce	861
Pinus resinosa	Red pine	131
Pinus rigida	Pitch pine	2
Pinus strobus	Eastern white pine	762
Prunus serotina	Black cherry	243
Quercus alba	White oak	9
Quercus bicolor	Swamp white oak	8
Quercus macrocarpa	Bur oak	14
Quercus rubra	Red oak	453
Thuja occidentalis	White cedar	1269
Tilia americana	Basswood	395
Tsuga canadensis	Eastern hemlock	472
Ulmus americana	American elm	228
Ulmus rubra	Red elm	9
Ulmus thomasii	Rock elm	3

Table S2. List of R packages used.

Packages	Main functions	Uses	References
msm	msm	Multi-state Markov models in continuous time	Jackson (2011)
	lrtest.msm	Likelihood ratio test	
	pmatrix.msm	Transition probability matrix	
	hazard.msm	Calculate tables of hazard ratios for covariates on transition intensities	
	pnext.msm	Probability of each state being next	
	sojourn.msm	Mean sojourn times from a multi-state model	
sf		Manipulation and mapping of spatial data	Pebesma (2018)
pROC	multiclass.roc	Compute multi-class AUC	Robin et al. (2011)
scoring	logscore	Compute logarithmic score	Merkle & Steyvers (2013)

Table S3. Frequency of all observed transitions between the four forest states during the study period. Transitions are from rows to columns.

	Boreal	**Mixed**	**Pioneer**	**Temperate**	**Sum**
Boreal	9760	203	1454	3	11420
Mixed	82	4996	393	682	6153
Pioneer	1539	678	6460	198	8875
Temperate	2	488	168	4584	5242
Sum	11383	6365	8475	5467	31690

Table S4. Table of baseline transition intensities ($q_{rs,0}$ in first column) and Hazard ratios (HR) and their 95% confidence intervals as estimated from the best multi-state transition model. The HR of covariates are interpretable as multiplicative effects on the baseline hazard, where values above 1 indicate that the predictor is associated with a greater risk of state transition, whereas values below 1 indicate a lower risk of transition. Covariates statistically different from 1 are coloured in grey.

Transitions	Baseline	Temperature	CMI	Drainage	pH	Natural1	Natural2	Logging1	Logging2
Boreal - Boreal	-0.009								
	(-0.01, -0.008)								
Boreal - Mixed	0.002	7.36	1.43	0.74	1.12	2.76	2.34	3.45	0.72
	(0.001, 0.003)	(5.28, 10.28)	(1.11, 1.85)	(0.65, 0.84)	(0.98, 1.29)	(2.03, 3.75)	(0.94, 5.82)	(2.2, 5.42)	(0.02, 27.18)
Boreal - Pioneer	0.007	1.000	1.000	1.000	1.000	4.76	37.01	8.45	213.52
	(0.006, 0.008)					(3.88, 5.84)	(30.67, 44.66)	(5.89, 12.12)	(102.4, 445.24)
Mixed - Boreal	0.005	1.07	1.56	1	0.82	0.98	2.5	1.64	1.14
	(0.003, 0.008)	(0.65, 1.75)	(1.18, 2.06)	(0.8, 1.25)	(0.65, 1.04)	(0.52, 1.85)	(0.33, 18.8)	(0.94, 2.86)	(0.02, 72.72)
Mixed - Mixed	-0.033								
	(-0.037, -0.029)								
Mixed - Pioneer	0.005	1.000	1.000	1.000	1.000	1.91	4.94	2.45	31.22
	(0.004, 0.006)					(0.98, 3.73)	(0.6, 40.6)	(1.22, 4.91)	(19.07, 51.11)
Mixed - Temperate	0.023	0.95	0.81	0.96	0.94	2.52	5.76	3.39	5.32
	(0.02, 0.026)	(0.76, 1.18)	(0.7, 0.92)	(0.88, 1.05)	(0.87, 1.01)	(2.04, 3.11)	(3.02, 10.98)	(2.79, 4.11)	(3.46, 8.16)
Pioneer - Boreal	0.027	1.28	1.8	0.98	0.96	0.94	0.38	1.9	1.84
	(0.025, 0.029)	(1.17, 1.41)	(1.65, 1.96)	(0.93, 1.02)	(0.9, 1.01)	(0.77, 1.15)	(0.29, 0.49)	(1.55, 2.33)	(0.87, 3.89)
Pioneer - Mixed	0.004	4.86	1.55	0.9	0.99	1.28	0.52	2.54	0.97
	(0.003, 0.004)	(3.7, 6.39)	(1.25, 1.91)	(0.8, 1.01)	(0.89, 1.1)	(0.83, 1.96)	(0.22, 1.22)	(1.72, 3.76)	(0.63, 1.48)
Pioneer - Pioneer	-0.032								
	(-0.034, -0.029)								
Pioneer - Temperate	0.001	18.57	2.46	0.77	1	0.2	0.54	1.25	0.31
	(0.001, 0.001)	(12.5, 27.6)	(1.88, 3.23)	(0.65, 0.91)	(0.88, 1.13)	(0.02, 2.05)	(0.13, 2.3)	(0.64, 2.42)	(0.08, 1.15)
Temperate - Mixed	0.014	0.54	0.96	1.3	0.73	1.24	2.08	1.1	2.41
	(0.012, 0.017)	(0.41, 0.7)	(0.83, 1.12)	(1.15, 1.47)	(0.65, 0.82)	(0.78, 1.97)	(0.57, 7.65)	(0.84, 1.44)	(1.12, 5.21)
Temperate - Pioneer	0.001	1.000	1.000	1.000	1.000	1.13	51.54	3.8	146.94
	(0.001, 0.002)					(0.12, 10.61)	(22.41, 118.54)	(2.13, 6.8)	(94.03, 229.62)
Temperate - Temperate	-0.016								
	(-0.019, -0.013)								

Supplementary figures

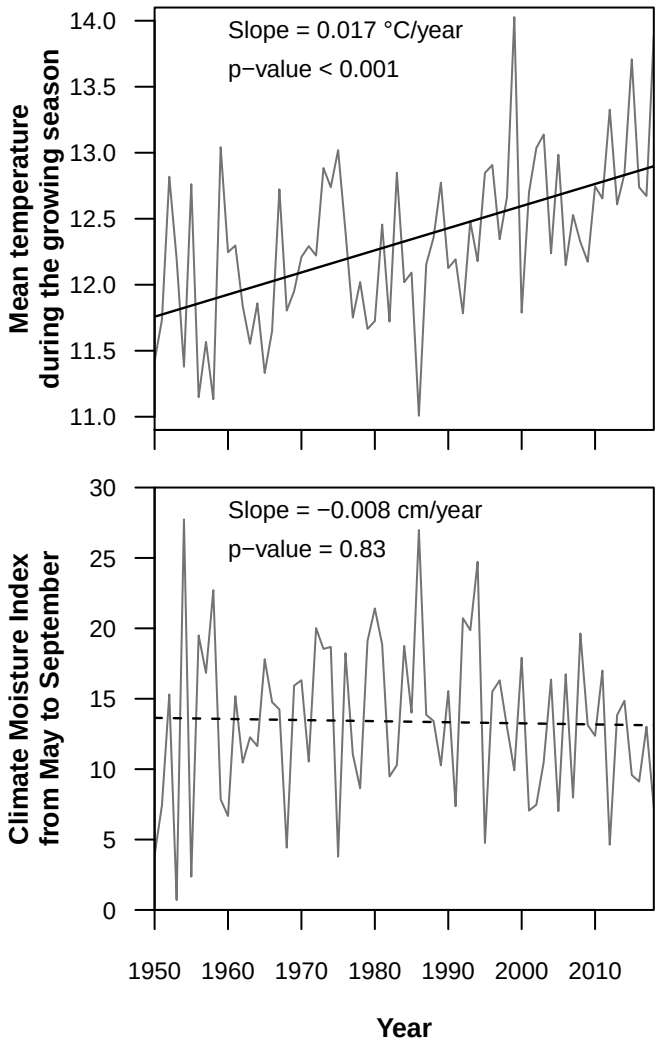

Figure S1: Temporal trends in growing season temperatures (top) and annual climate moisture index (bottom). Grey lines represent averaged climate values across the 10,388 studied forest plots. Straight black lines show the fitted least-squared linear regression lines.

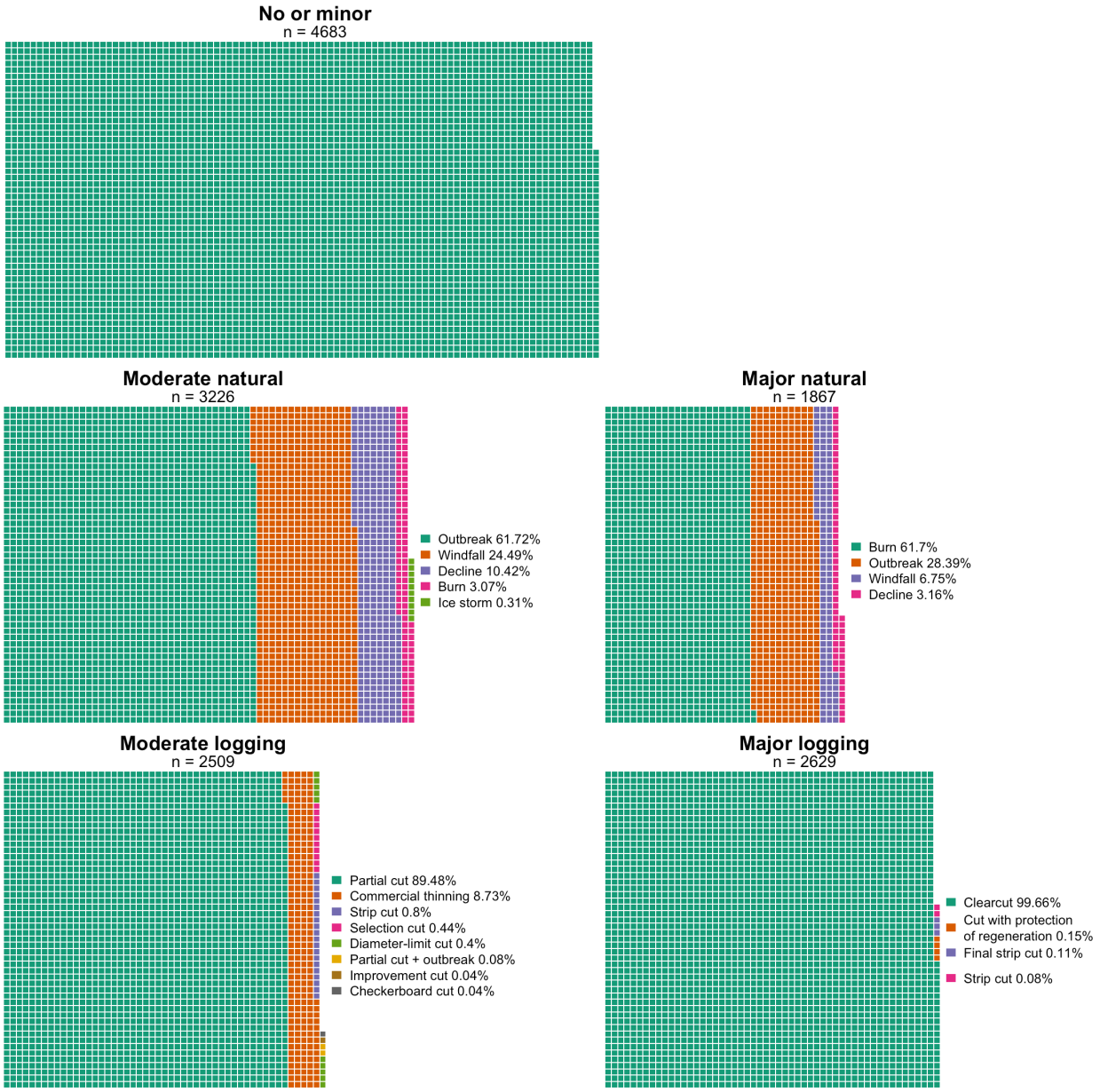

Figure S2: Waffle charts representing the frequency of forest plots by disturbance type (natural disturbances and logging) and level of intensity (minor, moderate, major). One square is one forest plot. In each chart (except for the no or minor disturbances), the colours represent the 21 original disturbance types recorded in the field surveys.

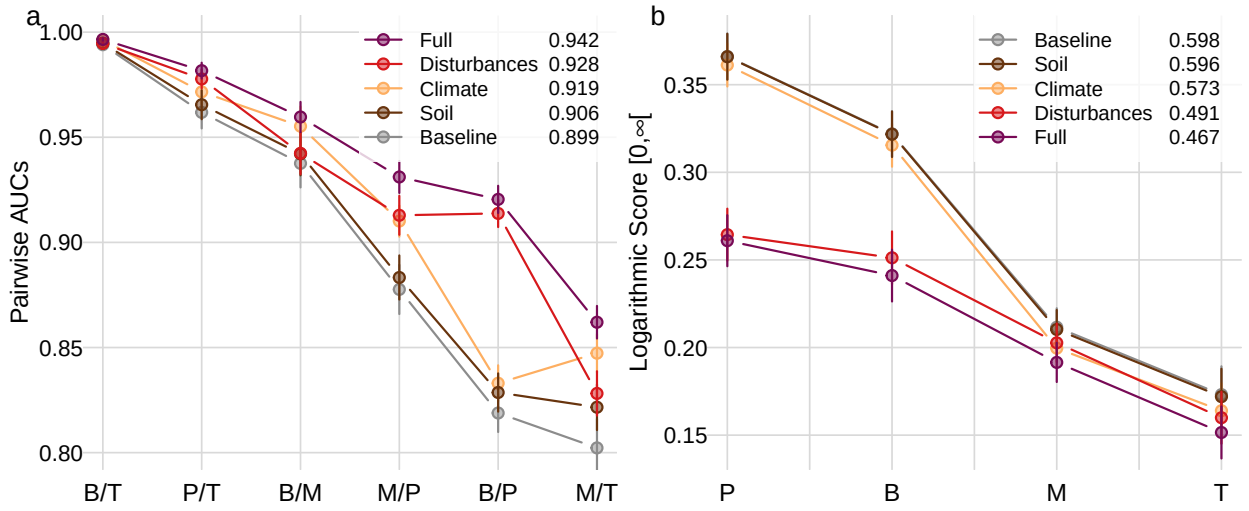

Figure S3: Performance comparisons of the five candidate multi-state models using multi-class area under the curve (mAUC; a) and logarithmic skill score (LS; b) obtained through 10-fold cross-validation. Higher values of mean pairwise AUC (a) indicate a better capacity to discriminate between pairs of the four forest states: (B)oreal, (M)ixed, (P)ioneer and (T)emperate. The overall mAUC of each model is given next to the legend. Lower values of mean state-specific logarithmic scores (b) indicate better prediction accuracy for each of the four forest states. The overall logarithmic score of each model is given next to the legend.

Model evaluation using 10-fold cross-validation revealed that including climate and disturbances improved overall model predictive performances, while soil variables had a negligible effect. All models were good at distinguishing Boreal from Temperate (high pairwise AUC). Soil variables slightly help to predict Mixed and Temperate states. Including climate variables help to distinguish Mixed from the other states, while including disturbances help to distinguish Pioneer from the other states, especially Boreal.

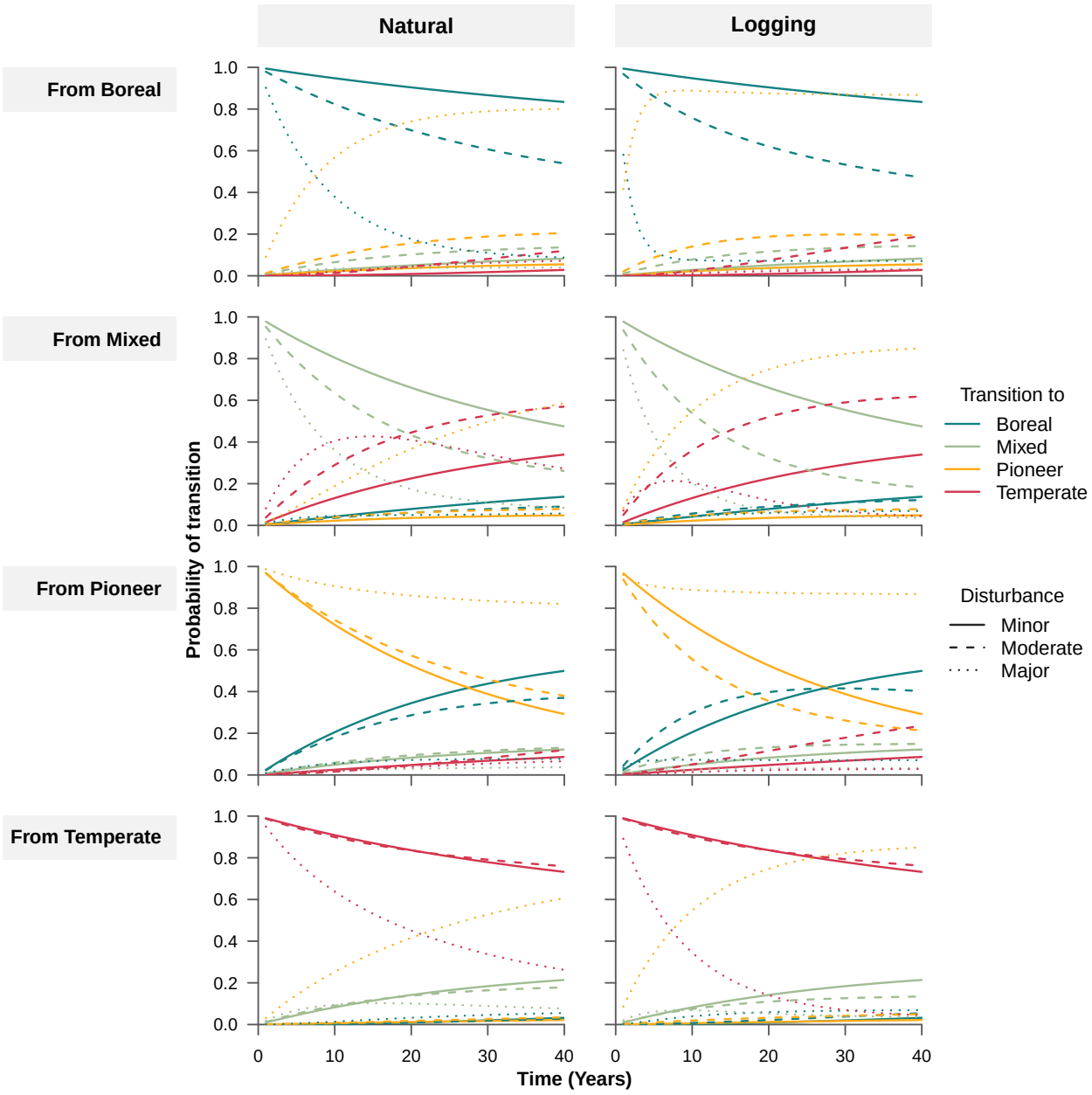

Figure S4: Probability of transition between forest states through time for natural disturbances (left) and logging (right) and for different levels (line type) as predicted by the best multi-state model. All other covariates are fixed at the average conditions found in the ecotone, i.e. the balsam fir-yellow birch domain.

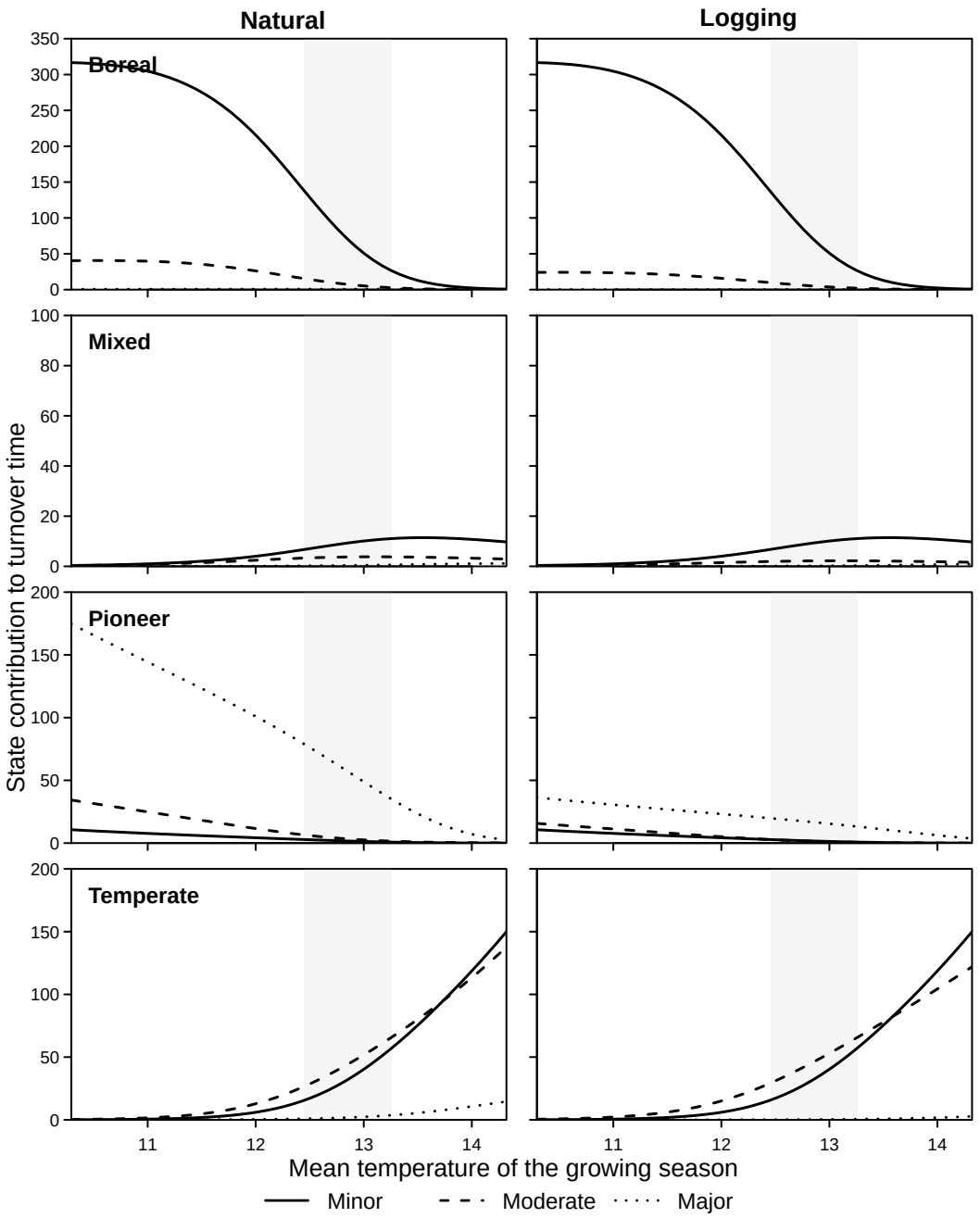

Figure S5: State contribution to forest turnover (see Fig. 7a,b in main text) along the temperature (latitudinal) gradient for different disturbance scenarios: minor (solid), moderate (dashed) and major (dotted) disturbances for both natural (a,c,e) and logging (b,d,f). All other covariates are fixed at the average conditions found in the ecotone, i.e. the balsam fir-yellow birch domain, to focus solely on the effect of disturbances along the temperature gradient. The turnover time of a state (or sojourn time) measures the time spent in this state before transitioning to the next. Long turnover time can translate to large resistance. Here, at any point along the gradient, state turnover time is scaled by the steady state distribution and the sum of all scaled state turnover gives the the turnover time of the transition matrix. The colors at the top of the plots approximate the position of the bioclimatic domains.

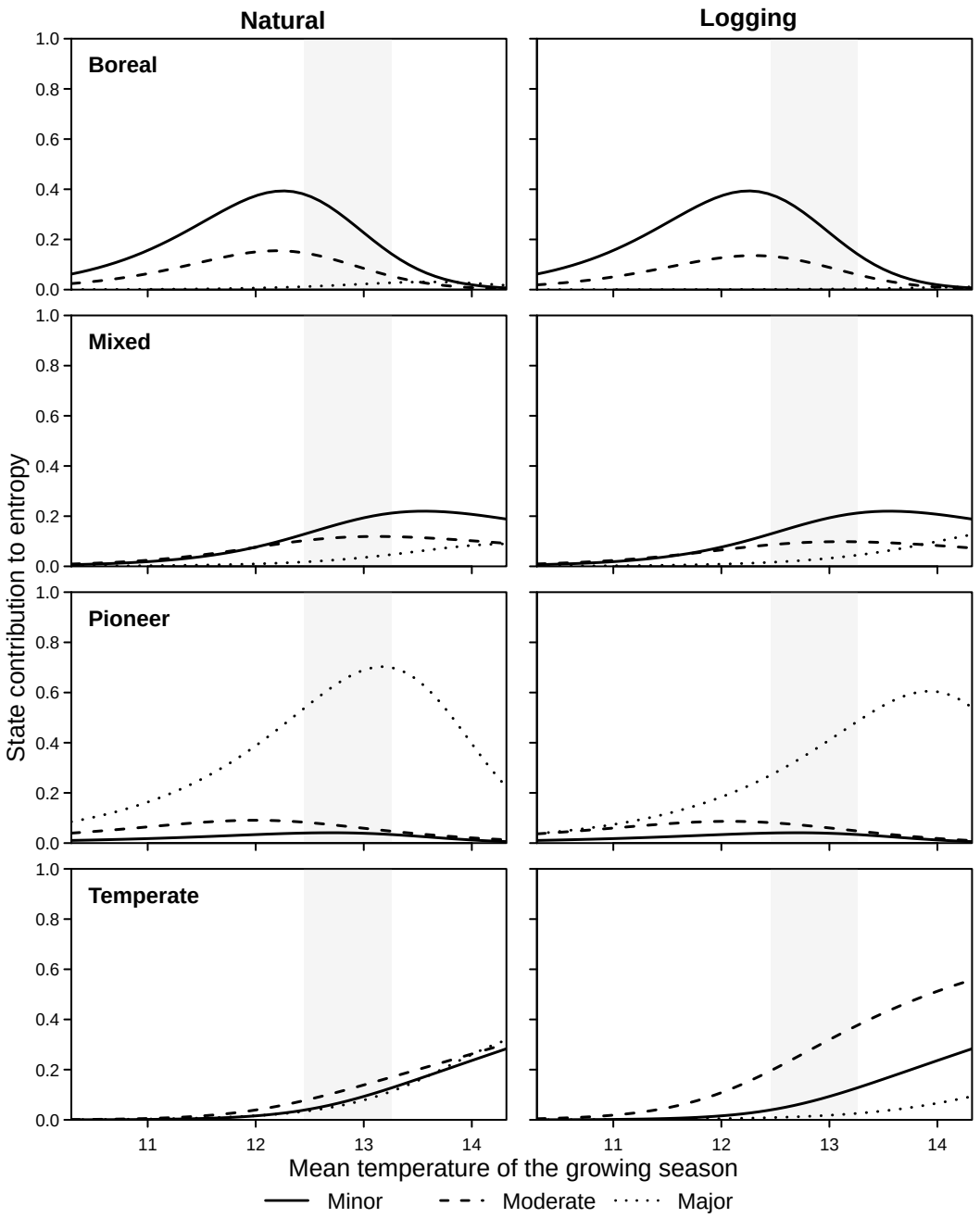

Figure S6: State contribution to forest entropy (see Fig. 7c,d in main text) along the temperature (latitudinal) gradient for different disturbance scenarios: minor (solid), moderate (dashed) and major (dotted) disturbances for both natural (a,c,e) and logging (b,d,f). All other covariates are fixed at the average conditions found in the ecotone, i.e. the balsam fir-yellow birch domain, to focus solely on the effect of disturbances along the temperature gradient. The entropy of a state measures the incertitude of its next transition. Here, at any point along the gradient, state entropy is scaled by the steady state distribution and the sum of all scaled state entropy gives the the entropy of the transition matrix. The colors at the top of the plots approximate the position of the bioclimatic domains.

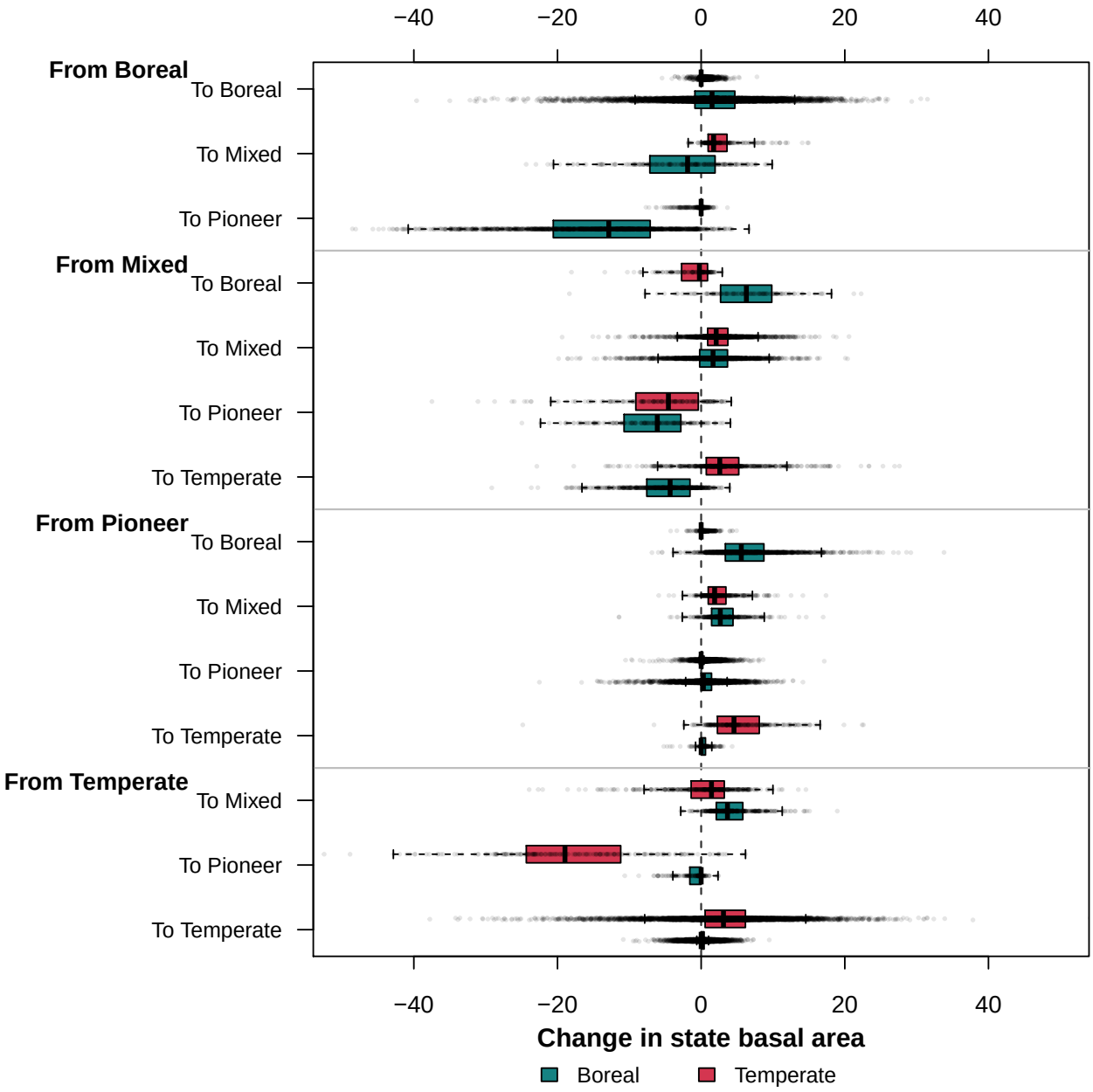

Figure S7: Change in tree basal area for temperate (red) and boreal (blue) species by transition type. Forest plots on the right have seen an increase in the basal area of one of the two groups, whereas forest plots on the left have seen a decrease in basal area. These changes in basal area are the underlying causes of the state transitions. For instance, transitions to Pioneer are mainly caused by losses of tree basal area, while transitions from Pioneer are dominated by gains in basal area. The transitions from Mixed to Temperate states is a result of both gains in temperate species and losses in boreal species.

References

Brice, M., Cazelles, K., Legendre, P., & Fortin, M. (2019). Disturbances amplify tree community responses to climate change in the temperate–boreal ecotone. *Global Ecology and Biogeography*, geb.12971. https://doi.org/10.1111/geb.12971

Burnham, K. P., Anderson, D. R., & Burnham, K. P. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed). New York: Springer.

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 20.

Hand, D. J., & Hill, R. J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*, *45*, 171–186.

Jackson, C. H. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, *38*(8), 28. https://doi.org/10.18637/jss.v038.i08

Merkle, E. C., & Steyvers, M. (2013). Choosing a Strictly Proper Scoring Rule. *Decision Analysis*, *10*, 292–304.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*. Retrieved from https://journal.r-project.org/archive/2018/RJ-2018-009/index.html

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77.