Supplementary Information

Supplementary tables

Table S1. List of species included in the analyses and their corresponding group. 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**
Boreal	
Abies balsamea	Balsam fir
Picea glauca	White spruce
Picea mariana	Black spruce
Pinus banksiana	Jack pine
Larix laricina	Tamarack
Alnus incana	Speckled alder
Pioneer	
Betula papyrifera	White birch
Betula populifolia	Grey birch
Populus tremuloides	Trembling aspen
Populus deltoides	Cottonwood
Populus balsamifera	Balsam poplar
Populus grandidentata	Large tooth aspen
Salix sp.	Willow
Prunus pensylvanica	Pin cherry
Crataegus sp.	Hawthorn
Sorbus sp.	Mountain-ash
Temperate	
Picea rubens	Red spruce

(continued)

Species name	**Vernacular name**
Pinus resinosa	Red pine
Pinus strobus	Eastern white pine
Tsuga canadensis	Eastern hemlock
Ulmus americana	American elm
Ulmus rubra	Red elm
Ulmus thomasii	Rock elm
Carya cordiformis	Bitternut hickory
Juglans cinerea	Butternut
Quercus macrocarpa	Bur oak
Quercus alba	White oak
Quercus bicolor	Swamp white oak
Quercus rubra	Red oak
Fagus grandifolia	American beech
Betula alleghaniensis	Yellow birch
Carpinus caroliniana	Blue beech
Ostrya virginiana	Ironwood
Tilia americana	Basswood
Prunus serotina	Black cherry
Acer rubrum	Red maple
Acer saccharum	Sugar maple
Acer pensylvanicum	Striped maple
Acer saccharinum	Silver maple
Acer spicatum	Mountain maple
Fraxinus pennsylvanica	Red ash
Fraxinus americana	White ash
Fraxinus nigra	Black ash
Thuja occidentalis	White cedar
Amelanchier sp.	Serviceberry
Acer negundo	Manitoba maple
Acer nigrum	Black maple
Pinus rigida	Pitch pine
Prunus virginiana	Chokecherry

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.

From/To	Boreal	Mixed	Pioneer	Temperate	Total
Boreal	9632	210	1171	17	11030
Mixed	131	2121	188	656	3096
Pioneer	1484	345	4839	281	6949
Temperate	11	383	215	6019	6628
Total	11258	3059	6413	6973	27703

Table S4. Table of risk ratios + CI

Transitions	Baseline	Temperature	CMI	Drainage	pH	Natural1	Natural2	Logging1	Logging2
Boreal - Boreal	-0.008 (-0.009 ,-0.007)								
Boreal - Mixed	0.002 (0.001 ,0.003)	8.546 (6.406 ,11.401)	1.47 (1.199 ,1.802)	0.714 (0.613 ,0.832)	1.023 (0.867 ,1.207)	2.803 (2.014 ,3.902)	2.739 (0.999 ,7.506)	2.784 (1.628 ,4.762)	1.333 (0.028 ,63.96)
Boreal - Pioneer	0.006 (0.005 ,0.007)	1.000	1.000	1.000	1.000	5.202 (4.208 ,6.431)	29.474 (23.455 ,37.037)	11.067 (7.631 ,16.05)	164.842 (98.705 ,275.293)
Mixed - Boreal	0.005 (0.003 ,0.008)	0.784 (0.485 ,1.266)	1.291 (0.989 ,1.686)	0.99 (0.783 ,1.251)	0.792 (0.606 ,1.034)	0.845 (0.41 ,1.741)	1.056 (0.041 ,27.276)	1.553 (0.842 ,2.863)	0.669 (0.009 ,50.155)
Mixed - Mixed	-0.034 (-0.039 ,-0.03)								
Mixed - Pioneer	0.005 (0.004 ,0.007)	1.000	1.000	1.000	1.000	1.939 (1.042 ,3.609)	9.714 (2.933 ,32.17)	2.27 (1.117 ,4.613)	27.499 (16.906 ,44.73)
Mixed - Temperate	0.024 (0.021 ,0.027)	0.921 (0.756 ,1.122)	0.785 (0.694 ,0.887)	0.968 (0.885 ,1.058)	0.953 (0.879 ,1.034)	2.401 (1.931 ,2.984)	4.507 (2.249 ,9.032)	3.225 (2.635 ,3.946)	5.32 (3.309 ,8.553)
Pioneer - Boreal	0.028 (0.027 ,0.03)	0.985 (0.915 ,1.059)	1.518 (1.428 ,1.615)	0.998 (0.949 ,1.05)	0.934 (0.877 ,0.995)	0.987 (0.803 ,1.212)	0.728 (0.547 ,0.97)	1.625 (1.277 ,2.068)	2.05 (1.227 ,3.425)
Pioneer - Mixed	0.004 (0.003 ,0.005)	4.243 (3.399 ,5.297)	1.62 (1.369 ,1.916)	0.959 (0.842 ,1.093)	0.934 (0.827 ,1.055)	1.234 (0.778 ,1.957)	0.864 (0.365 ,2.046)	3.267 (2.256 ,4.73)	1.117 (0.599 ,2.083)
Pioneer - Pioneer	-0.034 (-0.036 ,-0.032)								
Pioneer - Temperate	0.001 (0.001 ,0.002)	13.666 (9.914 ,18.839)	2.523 (2.032 ,3.132)	0.832 (0.709 ,0.976)	0.989 (0.878 ,1.114)	0.134 (0.007 ,2.534)	0.776 (0.212 ,2.837)	0.914 (0.409 ,2.046)	0.382 (0.12 ,1.216)
Temperate - Mixed	0.014 (0.011 ,0.017)	0.466 (0.361 ,0.601)	0.889 (0.769 ,1.029)	1.319 (1.159 ,1.502)	0.8 (0.709 ,0.901)	1.248 (0.763 ,2.042)	2.054 (0.583 ,7.232)	1.284 (0.968 ,1.705)	2.796 (1.263 ,6.193)
Temperate - Pioneer	0.001 (0.001 ,0.002)	1.000	1.000	1.000	1.000	1.362 (0.238 ,7.787)	30.08 (8.99 ,100.652)	3.942 (2.181 ,7.126)	122.836 (78.146 ,193.084)
Temperate - Temperate	-0.015 (-0.018 ,-0.013)								

Supplementary figures

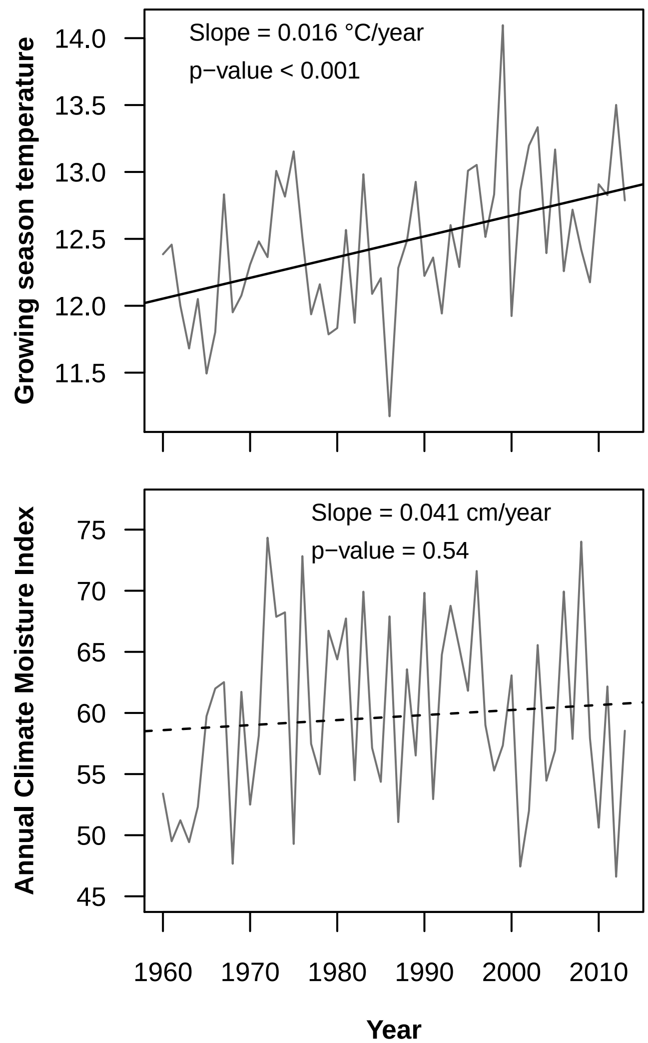

Fig. S1: Temporal trends in growing season temperatures (top) and annual climate moisture index (bottom). Grey lines represent averaged climate values across the 10,388 studied forest plots. Straight black lines show the fitted least-squared linear regression lines.

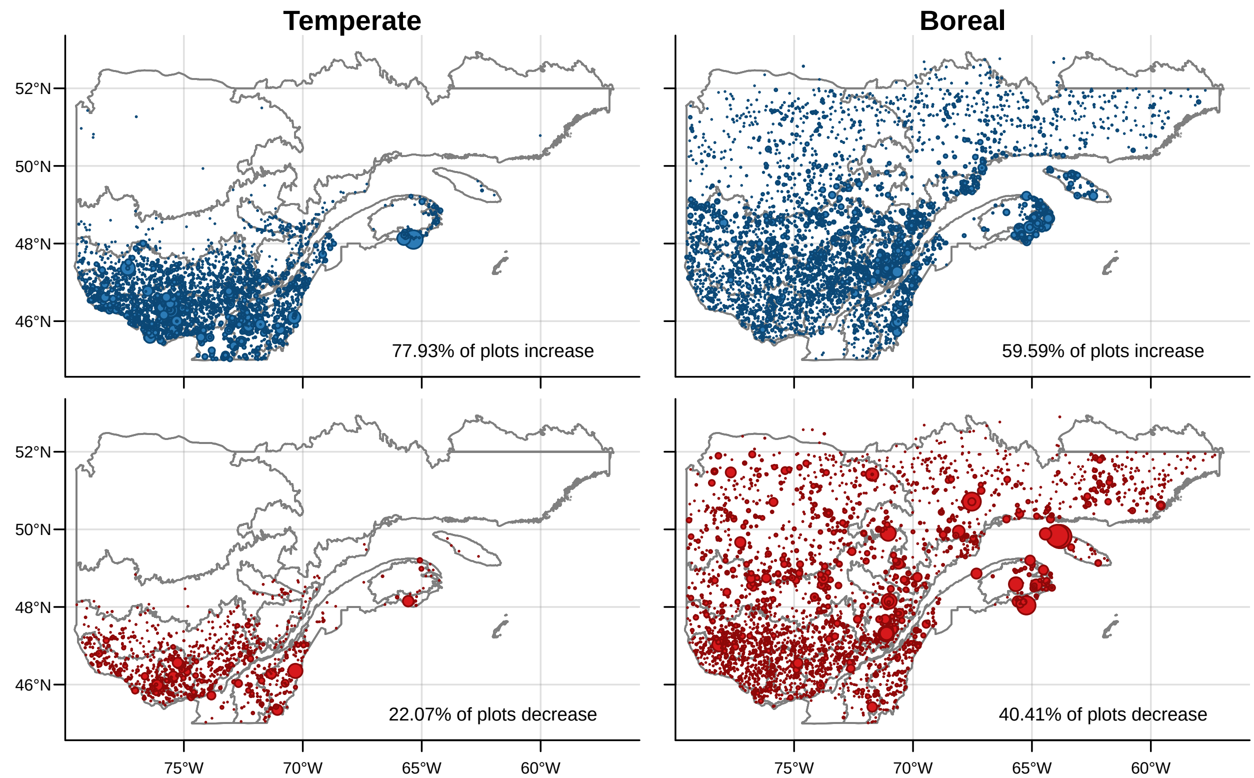

Fig. S2: Spatial distribution of observed change in tree basal area for temperate (left) and boreal (right) species. Forest plots in blue (top) have seen an increase in the basal area of one of the two groups, whereas forest plots in red (bottom) have seen a decrease in basal area.

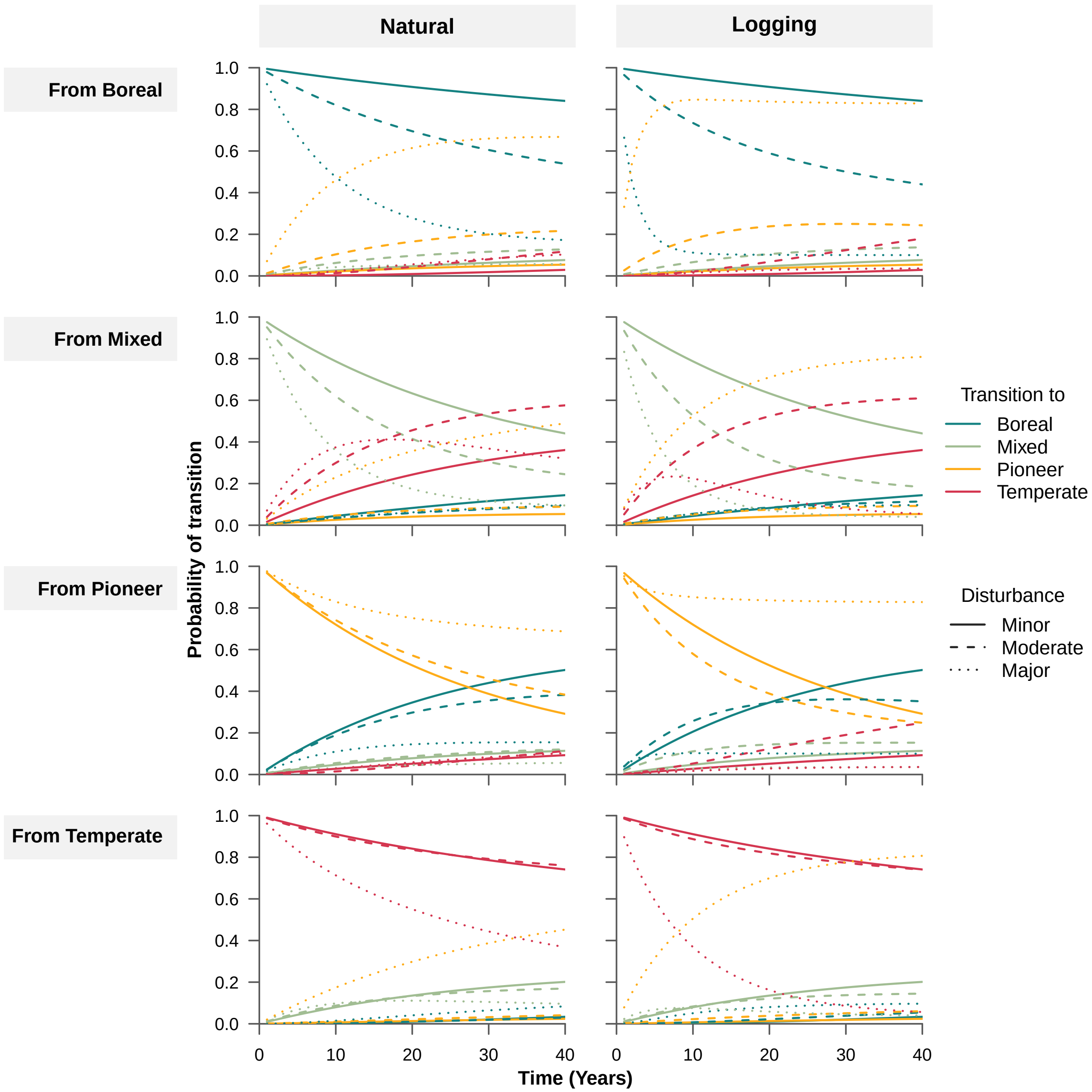

Fig. S3: Probability of transition between forest states through time for natural disturbances (left) and logging (right) and for different levels (line type) as predicted by the best multi-state model. All other covariates are fixed at the average conditions found in the ecotone, i.e. the balsam fir-yellow birch domain.

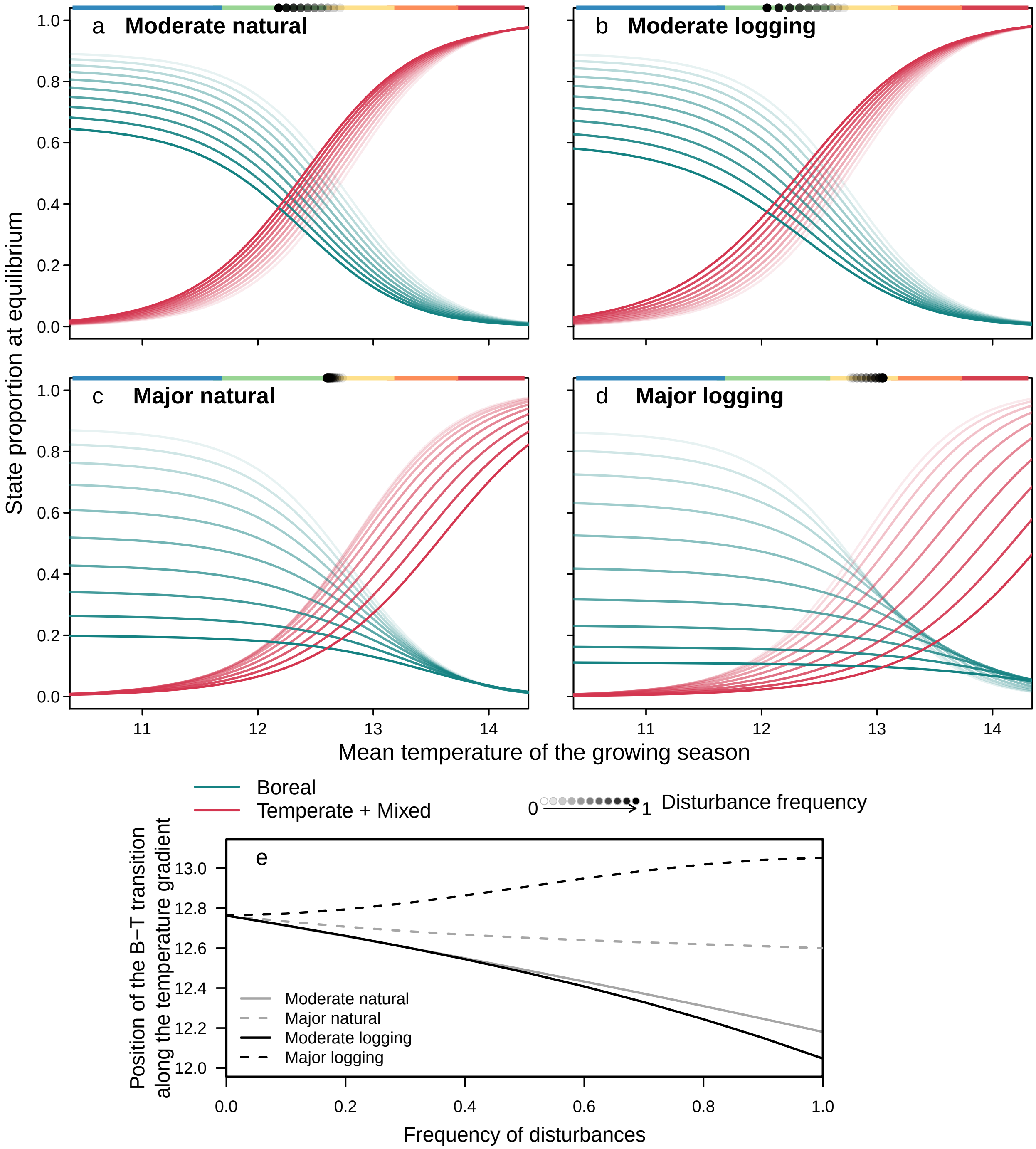

Fig. S4: Changes in forest state proportion at equilibrium (a-d) as well as change in the position of the Boreal-Temperate transition (e) along the temperature (latitudinal) gradient for different disturbance scenarios . Proportion of Boreal (blue) and Temperate + Mixed forests (red) for increasing frequency of natural disturbances (a,c) and logging (b,d). Increasing frequency of disturbances is illustrated as increasing color intensity (from pale, low intensity, to dark, high intensity). The circles at the top of each plot (a-d) indicate the position of the boundary between dominance of Boreal forests and dominance of Temperate and Mixed forests (i.e. the advancing front). The change in the position of these circles along increasing frequency of each disturbance type and intensity are illustrated in (e). 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 colors at the top of the plots approximate the position of the bioclimatic domains along the temperature gradient.

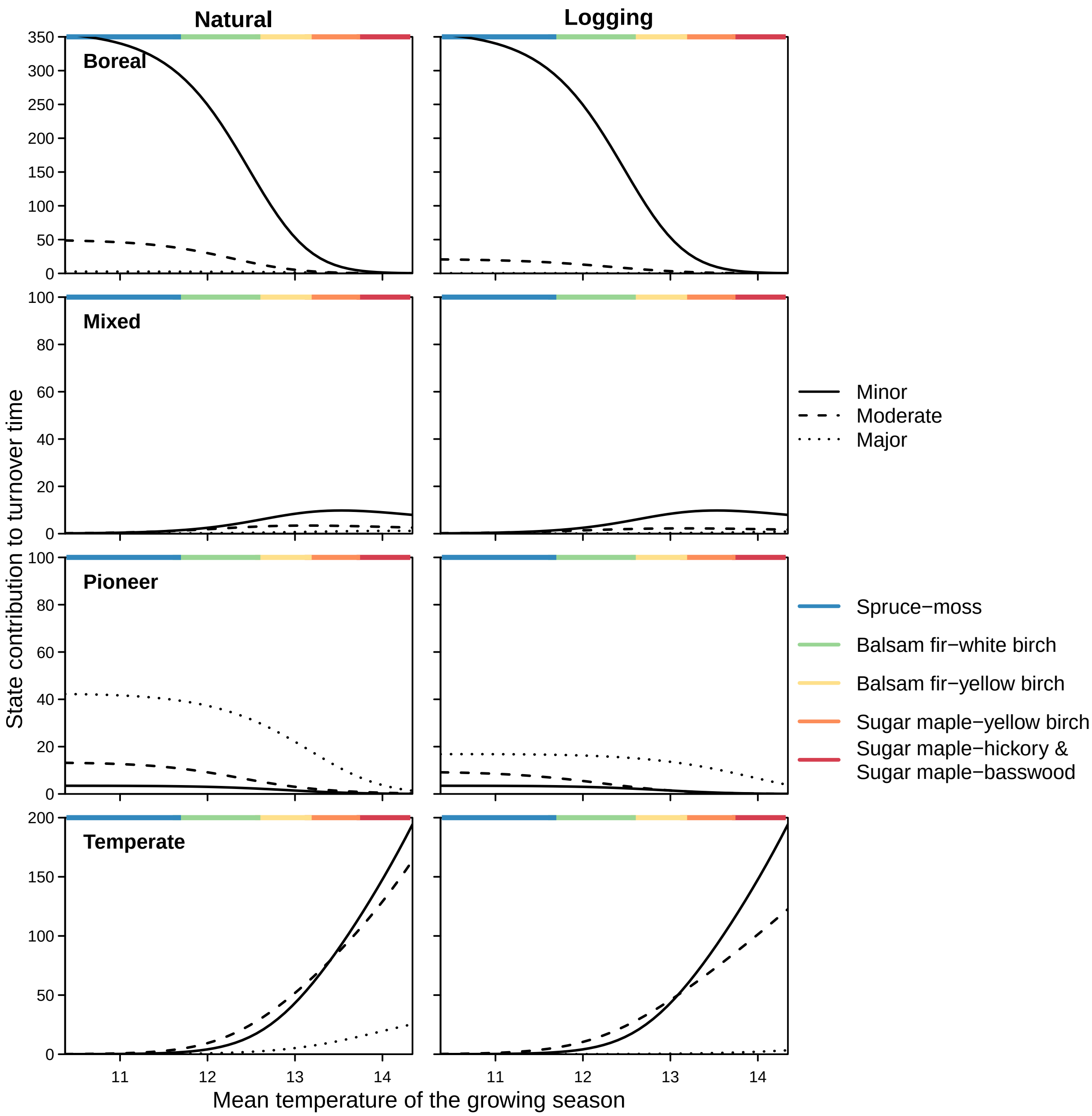

Fig. S5: State contribution to forest turnover (see Fig. 7a,b in main text) along the temperature (latitudinal) gradient for different disturbance scenarios: minor (solid), moderate (dashed) and major (dotted) disturbances for both natural (a,c,e) and logging (b,d,f). All other covariates are fixed at 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.

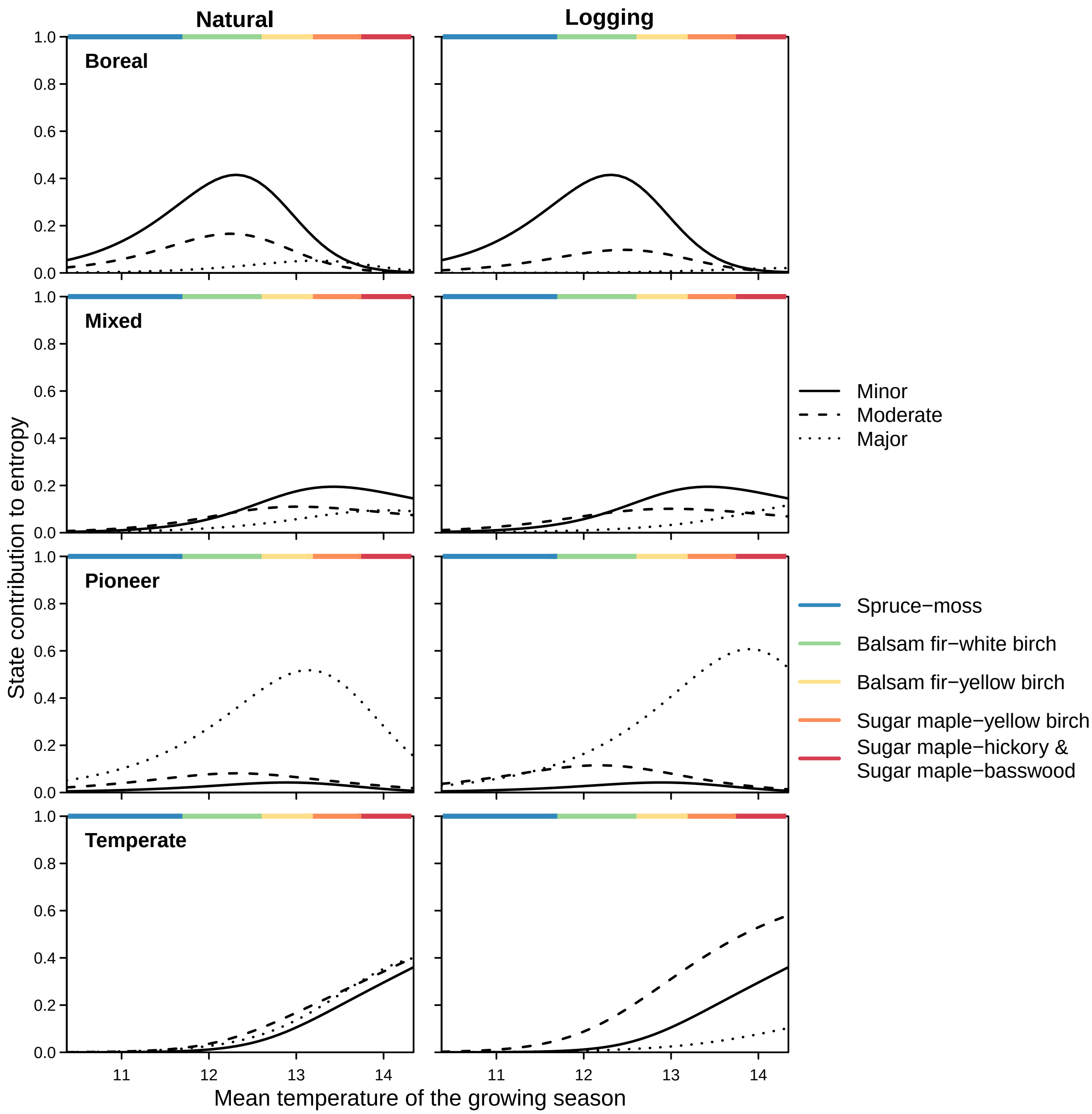

Fig. S6: State contribution to forest entropy (see Fig. 7c,d in main text) along the temperature (latitudinal) gradient for different disturbance scenarios: minor (solid), moderate (dashed) and major (dotted) disturbances for both natural (a,c,e) and logging (b,d,f). All other covariates are fixed at 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.

Supplementary methods

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.

Results

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 (*Fig. Supp*). 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.

Table Sx. Comparisons of the five candidate multi-state models. The number of parameters used in each model corresponds to the number of modelled transitions (10) $\times$ the number of covariates. The ΔAIC is the difference between the Akaike information criterion of each model (AIC_m) and the minimum of AIC among all the models (AIC_min): $\Delta\text{AIC} = \text{AIC}_\text{m} - \text{AIC}_\text{min}$. Multi-class area under the curve (mAUC) were obtained through 10-fold cross-validation. Higher mAUC indicate better model predictive performance. The best model is the one in bold with ΔAIC = 0.

	Covariates	Nb of parameters	-2 Log-likelihood	ΔAIC	LR test	mAUC	LS
Baseline	Intercept	10	32032.3	6132.6	< 0.001	0.899	0.578
Soil	Drainage, pH	24	31886.9	6015.2	< 0.001	0.906	0.576
Climate	Temperature, CMI	24	30438.7	4566.9	< 0.001	0.921	0.550
Disturbances	Natural, Logging	50	27341.3	1521.6	< 0.001	0.925	0.495
Full	**All**	**78**	**25763.7**	**0.0**	< 0.001	**0.940**	**0.468**

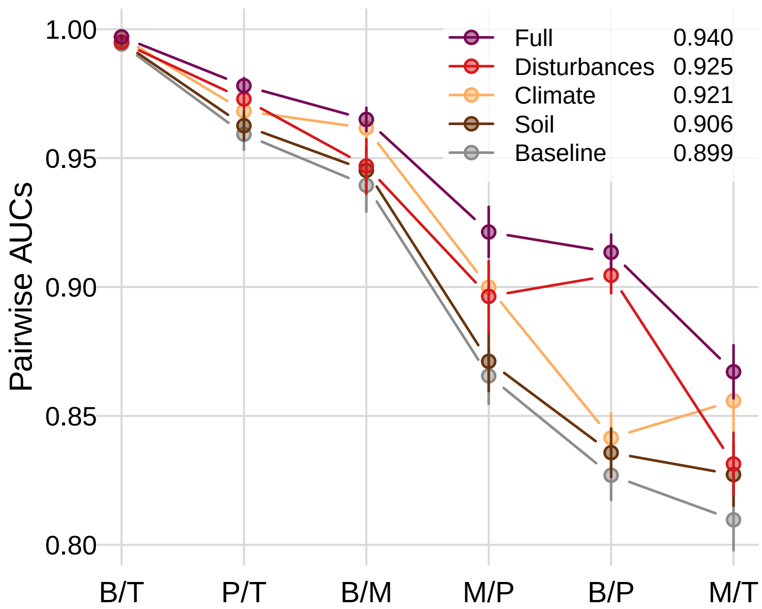

Fig. S7: Mean pairwise AUC (areas under the receiver operating characteristic curves) obtained through 10-fold cross-validation. Higher values indicate a better capacity to discriminate between the four forest states: (B)oreal, (M)ixed, (P)ioneer and (T)emperate. The overall mAUC of each model is given next to the legend.

AND/OR

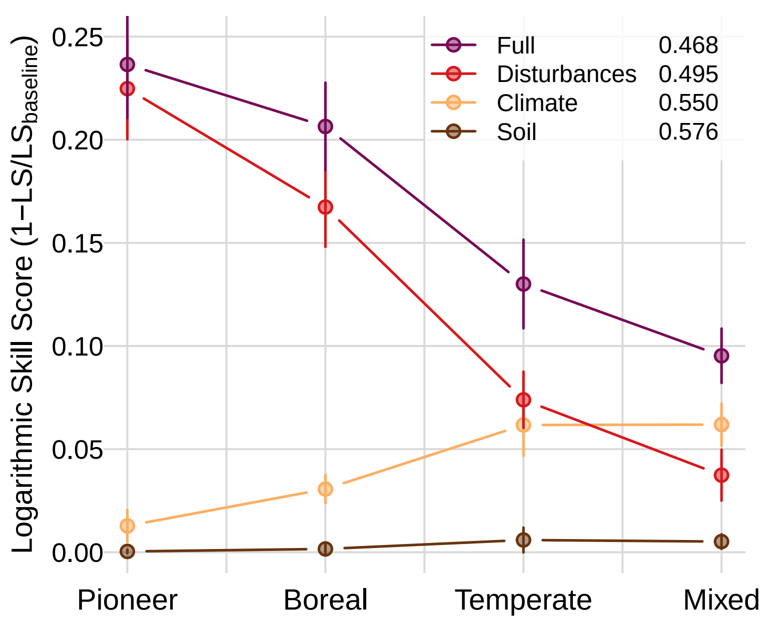

Fig. S8: Mean state-specific logarithmic skill score where each model including covariates is compared to the baseline model. Values were obtained through 10-fold cross-validation. Higher values indicate a larger improvement (predicted probabilities are closer to the observed outcomes) compared to the baseline model. The overall logarithmic score of each model is given next to the legend.

OR

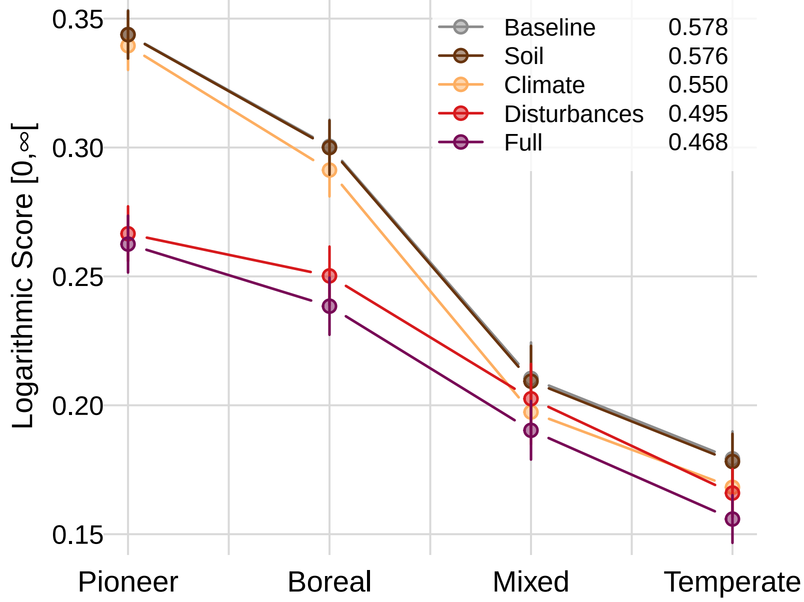

Fig. S9: Mean state-specific logarithmic score obtained through 10-fold cross-validation. Lower values indicate that the predicted probabilities are closer to the observed outcomes. The overall logarithmic score of each model is given next to the legend.

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.