

Appendix S4: Summary of model accuracies and predictor variables used

The following material summarizes the individual, cross-validated model accuracies (Table S4.1) of the six statistical models for ensemble predictions, and the final variables used by the GLM model after stepwise variable reduction (all other methods use down-weighting of less important or correlated variables).

In order to build ensemble models of sufficiently high model accuracies, we excluded those models from the ensembles, where AUC values were <0.7. All other models were aggregated to ensembles and used to calculate mean and st.dev of AUC values across models (presented as table 2 in the main report):

Table S4.1. Model accuracies given by cross-validated AUC (area under curve). The AUC values are tabulated for each species and each statistical model, and additionally as mean and st.dev among all six statistical models used. Models with AUC values <0.7 were excluded and therefore did not contribute to the statistic.

	GLM	GAM	GBM	ANN	MDA	CTA	Avg.	St.dev
Averages:	0.870	0.884	0.854	0.826	0.813	0.833	0.862	0.030
<i>Abies alba</i>	0.867	0.867	0.868	0.782	0.861	0.811	0.84	±0.034
<i>Acer campestre</i>	0.810	0.869	0.879	0.893	0.857	0.683	0.83	±0.071
<i>Acer pseudoplatanus</i>	0.793	0.802	0.794	0.722	0.791	0.754	0.78	±0.028
<i>Alnus glutinosa</i>	0.867	0.882	0.816	0.847	0.846	0.774	0.84	±0.035
<i>Alnus incana</i>	0.815	0.844	0.850	0.871	0.858	0.830	0.84	±0.018
<i>Betula pendula</i>	0.805	0.807	0.816	0.744	0.802	0.793	0.79	±0.024
<i>Carpinus betulus</i>	0.916	0.917	0.921	0.908	0.905	0.898	0.91	±0.008
<i>Castanea sativa</i>	0.989	0.991	0.977	0.963	0.979	0.959	0.98	±0.012
<i>Corylus avellana</i>	0.909	0.914	0.891	0.868	0.896	0.864	0.89	±0.019
<i>Fagus sylvatica</i>	0.865	0.865	0.877	0.750	0.864	0.816	0.84	±0.045
<i>Fraxinus excelsior</i>	0.841	0.855	0.859	0.827	0.851	0.812	0.84	±0.017
<i>Fraxinus ornus</i>	0.827	0.755	0.881	0.925	0.914	–	0.86	±0.063
<i>Larix decidua</i>	0.854	0.861	0.869	0.766	0.861	0.844	0.84	±0.035
<i>Ostrya carpinifolia</i>	0.948	0.950	0.922	0.867	0.939	0.907	0.92	±0.029
<i>Picea abies</i>	0.867	0.865	0.877	0.789	0.872	0.822	0.85	±0.032
<i>Pinus cembra</i>	0.963	0.963	0.959	0.923	0.962	0.876	0.94	±0.032
<i>Pinus nigra</i>	0.847	0.885	0.865	0.786	0.791	0.755	0.82	±0.047
<i>Pinus sylvestris</i>	0.863	0.867	0.850	0.811	0.844	0.822	0.84	±0.020
<i>Populus tremula</i>	0.764	0.793	0.790	0.744	0.774	0.715	0.76	±0.027
<i>Prunus avium</i>	0.821	0.833	0.808	0.807	0.800	0.741	0.80	±0.029
<i>Prunus padus</i>	–	0.908	–	–	0.786	0.768	0.82	±0.062
<i>Quercus ilex</i>	0.977	0.978	0.974	0.975	0.976	0.958	0.97	±0.007
<i>Quercus petraea</i>	0.902	0.910	0.893	0.879	0.885	0.878	0.89	±0.012
<i>Quercus pubescens</i>	0.836	0.923	0.904	0.825	0.852	0.769	0.85	±0.051
<i>Quercus robur</i>	0.887	0.892	0.897	0.881	0.896	0.830	0.88	±0.023
<i>Robinia pseudoacacia</i>	0.966	0.966	0.956	0.936	0.959	0.948	0.96	±0.010
<i>Ulmus glabra</i>	0.869	0.902	0.893	0.847	0.898	0.864	0.88	±0.020

Table S4.2: Predictor variables retained in the GLM model. 1 & 2 stand for linear and quadratic terms, respectively. Models were first fitted with all terms used in linear and quadratic mode, then reduced in a backward-forward stepwise variable selection. All other model approaches (GAM, GBM, CTA, ANN, MDA) use a variable weighing algorithm, and do not select and remove. *Corylus avellana* and *Prunus padus* were not modelled from GLM, as no statistically significant models were fitted.

	DDEG	TAVE	PRCP	PRCP	TOPOS	PRAD	SLP	ASP	CALC	DIST	SOIL
	yy	seas	49	103	–	yy	–	val	–	water	depth
<i>Abies alba</i>	1, 2	1, 2	1, 2	1, 2	1, 2	2	2	–	–	–	–
<i>Acer campestre</i>	1	–	–	–	–	–	–	–	–	–	1, 2
<i>Acer pseudoplatanus</i>	1, 2	–	1, 2	–	–	–	–	–	1	–	–
<i>Alnus glutinosa</i>	1	–	–	–	–	–	–	–	–	1	–
<i>Alnus incana</i>	–	1	–	–	1	–	–	–	–	1	–
<i>Betula pendula</i>	1, 2	1	1, 2	–	–	–	2	–	–	–	–
<i>Carpinus betulus</i>	1, 2	–	2	1, 2	–	–	–	–	–	–	–
<i>Castanea sativa</i>	1	–	1	–	–	–	1	–	1	–	–
<i>Corylus avellana</i>	1	1, 2	2	1, 2	–	–	–	–	–	–	–
<i>Fagus sylvatica</i>	1, 2	1, 2	1, 2	1	1, 2	1, 2	–	2	–	–	–
<i>Fraxinus excelsior</i>	1, 2	–	1, 2	–	–	–	–	–	–	–	–
<i>Fraxinus ornus</i>	–	–	–	1, 2	–	–	–	–	–	–	–
<i>Larix decidua</i>	1, 2	1, 2	1, 2	2	–	–	1	–	1	–	–
<i>Ostrya carpinifolia</i>	1, 2	1, 2	2	–	1, 2	–	1	–	–	–	–
<i>Picea abies</i>	1, 2	1	1, 2	1	1, 2	1, 2	1	2	–	–	–
<i>Pinus cembra</i>	1, 2	–	1	–	–	–	–	–	–	–	–
<i>Pinus nigra</i>	–	2	–	–	–	–	–	–	–	–	–
<i>Pinus sylvestris</i>	–	1, 2	1	1	1	1	2	–	–	–	–
<i>Populus tremula</i>	–	1	–	–	–	–	–	–	–	–	–
<i>Prunus avium</i>	1, 2	–	–	–	–	–	–	–	–	–	–
<i>Prunus padus</i>	–	–	–	–	–	–	–	–	–	–	–
<i>Quercus ilex</i>	1, 2	–	2	1	–	1	–	–	–	–	–
<i>Quercus petraea</i>	1, 2	1, 2	–	–	–	1	–	–	–	–	–
<i>Quercus pubescens</i>	–	1	2	–	1	–	–	–	–	–	–
<i>Quercus robur</i>	1, 2	–	–	–	1	1	–	–	–	–	–
<i>Robinia pseudoacacia</i>	1	–	–	–	–	–	–	–	–	–	–
<i>Ulmus glabra</i>	1, 2	2	–	–	–	–	–	–	–	–	–

Variables used are listed here. DDEG.yy: yearly degreeday sum (5.5 °C threshold); TAVE.seas: temperature seasonality (st.dev.of monthly values); PRCP.49: precipitation of the summer half (April to September); PRCP.103: precipitation of the winter half (October to March); TOPOS: topographic position; PRAD.yy: yearly sum of potential global radiation; SLP: slope angle (°); ASP.val: aspect value; CALC: degree of calcareous soil (6 classes); DIST.water: distance to running or standing water; SOIL.depth: soil depth.