Estimating consensus and uncertainty of inherently different species distribution models – Supplementary Material

Gritti, E.S.[1,2], Duputié, A.[1,2] , Massol, F.[1,3] & Chuine, I.[1]

This file:
- describes the SDMs used in the main text in detail, and provides their parameterization (section I)
- indicates how bioclimatic data were summarized into three independent variables (hereafter used to describe the environment) (section II)
- provides details on how the conditional consensus model was built, including commented R code (section III)
- describes how forecasts (of both SDMs and of the conditional consensus model) were performed (section IV)
- explains how the other consensus models were built and how model accuracy was evaluated (section V).

Two R scripts and data are provided to reproduce the results of the paper, starting with section III; these codes need some modification to be applied to other datasets. Code and data are located in the separate file "GrittietalScripts.zip".

CONTENTS

[1] CEFE, UMR 5175, 1919, route de Mende, 34293, Montpellier cedex 5, France

[2] These authors contributed equally to this work.

[3] IRSTEA – UR HYAX, 3275, route de Cézanne – Le Tholonet, CS 40061, 13182 Aix-en-Provence cedex 5, France

I.   SPECIES DISTRIBUTION MODELS

A.   STASH

STASH (Sykes *et al.* 1996) is a correlative model of species distribution, based on a minimal set of physiologically relevant bioclimatic parameters.

1.   DESCRIPTION OF THE MODEL.

This model calculates a growth efficiency index (Growth index; *Gi*) of a species at each location x for a given period as follows:

$$Gi_x = Est \times Tcmin_x \times Tcmax_x \times Twmin_x \times GDD_x \times Di_x \times Dtf_x \tag{1}$$

where Est is the mean establishment rate of the species (expressed in saplings/ha/year, assumed constant over pixels), $Tcmin_x$, $Tcmax_x$, $Twmin_x$ and GDD_x define four bioclimatic limits to establishment, and Di_x and Dtf_x are multipliers of the growth index.

Bioclimatic limits are determined using the observed distribution of the species, raising the risk of overfitting. To limit this risk, we performed 100 re-samplings of the observed distribution, used 30% of the pixels to determine the bioclimatic limits, projected the output of STASH on the remaining 70% pixels (the validation set), and used only STASH projections made on pixels belonging to the validation sets (see below; section I.A.2).

1. BIOCLIMATIC LIMITS.

- $Tcmin_x$ is defined by the minimum mean monthly temperature, below which establishment and growth are impossible:

$$Tcminx = \begin{cases} 0 & \text{if } Tc < minTc \\ 1 & \text{if } Tc > minTc \end{cases} \tag{2}$$

where Tc is the coldest-month mean temperature (surrogate of the absolute minimum temperature; Prentice *et al.* 1992) and $minTc$ the species tolerance limit for this variable.

- $Tcmax_x$ is defined by the warmest mean temperature of the coldest month, above which establishment and growth are impossible (for example because bud dormancy cannot be achieved):

$$Tcmax_x = \begin{cases} 0 & \text{if } Tc > maxTc \\ 1 & \text{if } Tc < maxTc \end{cases} \tag{3}$$

where Tc is the coldest-month mean temperature and $maxTc$ the species tolerance limit for this variable.

- $Twmin_x$ is defined by the coldest mean monthly temperature of the warmest month, below which establishment and growth are impossible (for example because fruit maturation cannot occur):

$$Tcmax_x = \begin{cases} 0 & \text{if } Tw < minTw \\ 1 & \text{if } Tw > minTw \end{cases} \tag{4}$$

where Tw is the mean warmest-month temperature and $maxTc$ the species minimum tolerance limit for this variable.

- GDD_x is defined by the minimum requirement in growing degree-days:

$$GDD_x = \begin{cases} 0 & \text{if } GDD - (b\,e^{-kC}) < minGDD \\ 1 & \text{if } GDD - (b\,e^{-kC}) > minGDD \end{cases} \qquad (5)$$

where GDD is the mean cumulative degrees above 5°C for deciduous and -4°C for evergreen; b is the cumulative degree above 5°C (or -4°C) for budburst with no chilling, k is the decay rate of the GDD need for budburst, C the mean number of chilling days per year and $minGDD$ the species minimum tolerance limit for this variable.

2. MULTIPLIERS OF THE GROWTH EFFICIENCY INDEX.

- DI_x is a drought index:

$$Di_x = \text{Max}\left[0;1-\left(\frac{PET - AET}{PET\;maxDi}\right)^2\right] \qquad (6)$$

where AET is the actual evapotranspiration and PET the potential evapotranspiration, and $maxDi$ the species maximum tolerance limit for $1 - AET/PET$ over its growing period.

- Dtf_x is an index of photosynthetic efficiency, depending upon daily temperature:

$$Dtf_x = \text{Max}\left[0;\sum_{d=1}^{365}\frac{4}{365}\frac{(T_d - T_l)(T_h - T_d)}{(T_h - T_l)^2}\right] \qquad (7)$$

where T_d is the daily temperature, T_l and T_h the lowest and highest temperatures for optimal photosynthesis, and Dtf_x is computed only over the period when $T_d \geq T_l$.

3. INPUT AND OUTPUT VARIABLES

The input variables are monthly average temperature, precipitation and percentage of sunshine. From these variables, 1- AET/PET, the temperature variables and the number of growing degree-days are computed over a 20-year period (1981-2000 for the "current" dataset; 2081-2100 for the scenarios; see Sykes *et al.* 1996 for details). Daily temperatures are linearly interpolated from monthly values.

The growth efficiency index output by STASH, Gi_x, is comprised between 0 and Est. To yield a surrogate for the probability of presence, and without loss of generality, we used $Est = 1$.

4. AVAILABILITY AND USE

STASH is available from the EMBERS group of Lund University, upon request. Because of its simplicity and accuracy, STASH has been often used to estimate the potential distribution of tree species under past, current and future climatic conditions (e.g. Sykes *et al.* 1996; Giesecke *et al.* 2007; Walther *et al.* 2007).

2.   MODEL PARAMETERIZATION (TABLE S1)

Bioclimatic envelopes (and thus, species specific limits) are determined by comparing the distribution map of each bioclimatic limit to the observed distribution of the species, provided by *Atlas Flora Europeae* (Tutin *et al.* 1964-85; completed by Laurent *et al.* 2004; **Fig. S1**). These reference maps compile species observations from the second half of the 20th century and are assumed to reflect the most accurate approximation of the species distribution available at the European scale, resulting from last normal climatic conditions period (1931-1960). *The Atlas Flora Europaeae* maps are provided with a resolution of 0.5°. To make them compatible with the resolution of the climatic datasets, these maps were downscaled to a 10'x10' resolution by simply attributing the value of the 0.5° cells to the set of corresponding 10' cells **(Figure S1).**
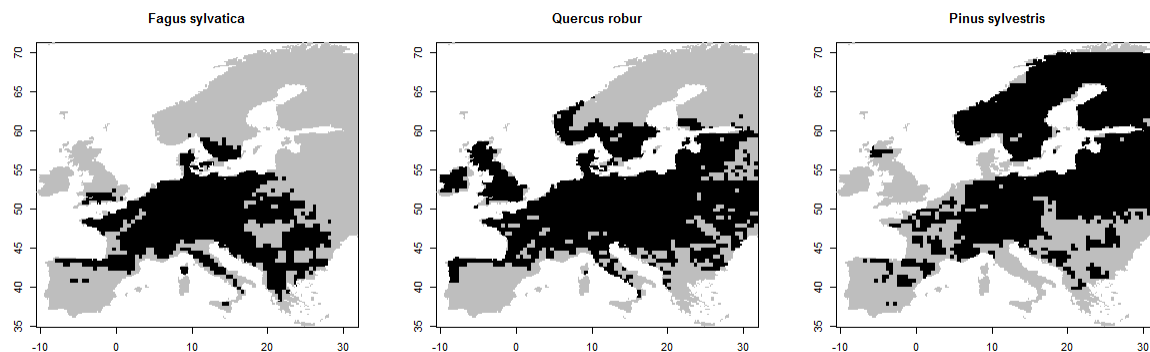

Figure S1. Current distributions of the three species studied (Tutin *et al.* 1964-85; completed by Laurent *et al.* 2004).

To avoid overestimating the amplitude of the niche, the limits chosen for each bioclimatic variable excluded the most extreme 2.5% pixels. To reduce the potential overfitting of STASH, we performed re-samplings of the *Atlas Flora Europaea* occurrence map. This is a common technique (e.g. Thuiller *et al.* 2009), where the observed occurrence data is split into a calibration dataset (used to calibrate the model) and a validation dataset (used to assess the validity of the model on pseudo-independent data). Here, for each species:
-   The dataset was split with a calibration set containing 30% of the pixels (8630 points); the remaining points constituted the validation set;
-   The calibration set was used to determine the bioclimatic limits used by STASH;
-   Given these limits, STASH was run;
-   A specific presence threshold (SPT) was determined on the *calibration dataset* (see section III A for its computation);
-   The SPT was used to transform continuous STASH projections into binary (presence/absence) projections;
-   Only those projections made on the *validation dataset* were conserved.

This process was repeated 100 times.

The final projection was obtained through a majority-rule consensus among the 100 re-samplings: pixels that were projected "present" (above the SPT) more than half of the times when in the validation set were considered as "present". This is what is stored in `data/projections/projectionsCurrentSTASHBinarized.txt`; columns `speciesSTASH`. Projections in this file thus reflect the consensus of on average 70 STASH projections. Note that all pixels belonged to the validation set under forecast conditions; hence values stored in `data/projections/projectionsA1Fi.txt` and `projectionsB2.txt`, columns `speciesSTASH`, contain the consensus of 100 STASH projections.

In practice, the re-sampling only slightly affected the estimation of the bioclimatic limits. As a result, the projections of STASH did not vary much among re-samplings (**Figure S2**).
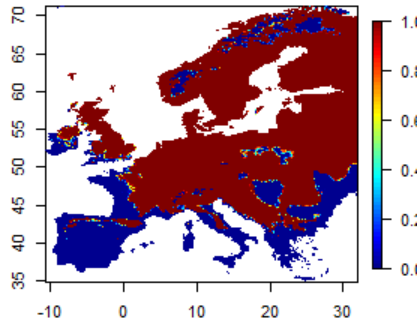
Figure S2. Map of Europe indicating the frequency at which *Pinus sylvestris* is projected by STASH to be present or absent. The species is projected to be present in all 100 re-samplings for dark red pixels; in no re-samplings for dark blue pixels. Only a very small proportion of pixels, towards the margins of the projected distribution, are not consistently projected as either present or absent. The same applied to all three scenarios and all three species.

STASH output was then transformed into binary projections, with pixels projected as present in more than half the validation datasets being considered as "present" (**Figure S3**).
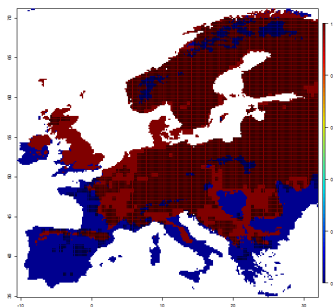

Figure S3. Binary STASH output for *Pinus sylvestris*. Red pixels indicate locations where the species is considered "present" and blue pixels those where the species is absent. Overlaid black dots represent the *Atlas Flora Europaea* map for *Pinus sylvestris.*

Note that this binary output is suitable for use with the conditional consensus method presented in the main text, but not for a mean or median consensus. To compute the mean, median and weighted average consensus between the different SDMs (Marmion *et al.* 2009), we thus also generated a continuous STASH output: for each pixel, this continuous output was the arithmetic mean of STASH output over the resamplings in which the pixel was in the validation dataset. (This continuous output is stored in "data/projections/ContinuousSTASHScenario.txt" files, in column "SpeciesSTASH").

Parameters used in STASH (or their range across re-samplings) are indicated in **Table S1**.

Table S1. Parameters used in STASH. For bioclimatic limits, their range and median across re-samplings are indicated (median between brackets).

	Fagus sylvatica	*Quercus robur*	*Pinus sylvestris*
minTc	-4.8 – -4.1 (-4.4)	-6.5 – -6.1 (-6.3)	-14.4 – -14.2 (-14.3)
maxTc	5.8 – 6.2 (6.0)	6.0 – 6.4 (6.2)	4.1 – 4.6 (4.3)
minTw	11.9 – 13.2 (12.5)	11.1 – 11.7 (11.4)	8.9 – 9.1 (9.0)
minGDD	381 – 546 (450)	561 – 686 (621)	221 – 253 (234)
b	1150	100	100
k	0.0065	0.05	0.05
maxDi	0.36 – 0.41 (0.38)	0.33 – 0.34 (0.34)	0.20 – 0.25 (0.23)
Tl	5	5	-4
Th	42	42	36

B.   L P J

LPJ is a general ecosystem model combining bioclimatic limits to the species' establishment and survival and mechanistic representations of physiology, biochemistry (from the BIOME3 and BIOME4 models, Haxeltine & Prentice 1996; Kaplan 2001), vegetation dynamics and carbon and water fluxes (Sitch *et al.* 2003). It models the growth and dynamics of the vegetation in a number of replicate patches per grid cell. The version used in this study includes representations of soil hydrology, snow-pack dynamics and soil–vegetation–atmosphere exchange of water, as documented by Gerten *et al*. (2004).

1.   DESCRIPTION OF THE MODEL

1. BIOCLIMATIC LIMITS

Four bioclimatic limits determine the species' envelope:
- Minimum temperature for survival (*Tcoldmins*),
- Minimum temperature for successful seedling establishment (*Tcoldmine*),
- Maximum temperature of the coldest month (*Tcoldmaxe*), above which seedling establishment is impossible due to the lack of dormancy,
- Minimum quantity of growing degree-days for successful establishment (*GDD5mine*).

These limits determine the suitable zone where the species may grow. They were fixed to values specific to each species (as determined by Koca *et al*., 2006).
Note that these limits are not (and cannot be) derived from the observed distribution of the species. Hence, no re-sampling strategy was performed for LPJ.

2. GROWTH

Within the suitable zone, LPJ computes, *inter alia*, the biomass, net primary productivity and leaf area index of plant functional types (PFT) or species (in the present case).

Modelling takes place at the population level with the representation of one "average individual" and a population density. The processes taken into account are:
- Photosynthesis as a function of photosynthetically active radiation, temperature, atmospheric CO_2 concentration and water availability.
- Respiration as a function of biomass, tissue specific C:N ratio and temperature. Respiration is decomposed in heterotrophic and autotrophic (maintenance and growth) respiration.
- Growth as a function of biomass production and allometric relationships between reproduction, leaf, root, sapwood and heartwood.
- Survival to stresses such as low biomass production, cold, fire and other disturbances.
- Phenology as a function of daily temperatures above which budburst can occur.
- Hydrology as a function of soil and vegetation cover parameters (evapotranspiration), temperature and precipitation (rain and snow melt).
- Soil and litter biogeochemistry as a function of temperature, soil moisture and biomass mortality (leaf and root turnover).

3. INPUT AND OUTPUT VARIABLES

Input variables are monthly temperature, precipitation and percentage of sunshine for each grid cell, and CO_2 concentration (assumed constant across space). These are downscaled to a daily time step for some processes, while other processes (e.g. mortality) require yearly averages. The model was run using a loop of repetitions of 1901-1930 for 600 spin-up years, after which vegetation had reached an approximate equilibrium with climate; simulations were then continued with the complete 1901-2100 climatic data.

LPJ outputs considered in this study are the average net primary productivity, and leaf area index of vegetation within each pixel. Since these outputs are extremely correlated, leaf area index was chosen as a proxy for the probability of presence. LAI was divided by its maximal value for the "current" dataset, hence yielding to an index comprised between 0 and 1 for current climatic conditions. However, this index may reach values higher than one for future conditions, due to the fertilizing effect of increased atmospheric CO_2 concentration.

4. AVAILABILITY AND USE

LPJ is available from the EMBERS group of Lund University, upon request. Ecosystem process modelling was validated with respect to seasonal and interannual variation in carbon and water vapour fluxes (Morales *et al.* 2005). LPJ has also been validated as a tool to explain vegetation distribution and dynamic at the plant functional type level (Badeck *et al.* 2001; Smith *et al.* 2001; Gritti *et al.* 2006).

2. MODEL PARAMETERIZATION

Each species is defined by a set of parameters describing plant physiognomy, allometry, physiology, phenology and bioclimatic limits.

The species were assigned the values ascribed to their respective PFTs (temperate broadleaved summergreens for *Fagus sylvatica* and *Quercus robur*; boreal needle-leaved evergreen for *Pinus sylvestris*) in Smith *et al.* (2001). Species-specific values were used when available (Hickler *et al.* 2004; Koca *et al.* 2006; Miller *et al.* 2008; Garreta *et al.* 2010). Parameters are given in **Table S2**.

Table S2. Parameters used in LPJ.

	Fagus sylvatica	*Quercus robur*	*Pinus sylvestris*
Bioclimatic limits			
Tcoldmins	-4	-7	-35
Tcoldmine	-4	-7	-30
Tcoldmaxe	6.1	5	3
GDD5mine	1000	900	400
Ecophysiology			
Roots distribution (upper/lower soil layer)	0.67/0.33	0.67/0.33	0.67/0.33
Leaf phenology	Summergreen	Summergreen	Evergreen
Leaf turnover rate (year^{-1})	1	1	0.33
SLA (cm^2.[gC]$^{-1}$)	273	273	93
Climate zone	Temperate	Temperate	Boreal
Optimal temperature range for photosynthesis (°C)	15-25	15-25	10-25
Max establishment (saplings.ha^{-1}.yr^{-1})	10	10	10
Max non-stressed longevity (yr)	200	200	300

C.   PHENOFIT

The model PHENOFIT (Chuine & Beaubien 2001) relies on the assumption that species adaptation to abiotic conditions is tightly related to its capacity to synchronize its annual life cycle with seasonal climatic variations directly impacting its survival and reproductive success. It simulates the precise phenology and levels of resistance to drought and cold stress of an average individual of a tree species given local climatic conditions to yield a reproductive success and a survival probability. The product of survival and reproductive success is used as a proxy for fitness, and for the probability of occurrence. This model does not make use the observed distribution of the species to produce its output.

Note that this model does not rely on observed species distribution; hence no correction for overfitting could be performed.

1.   DESCRIPTION OF THE VERSION OF THE MODEL USED IN THIS STUDY

1. PHENOLOGY

Leaf unfolding and flowering dates are determined by daily temperatures using the UniChill model of Chuine (2000). Fruit maturation date is calculated following Chuine & Beaubien (2001) for the deciduous species and following a degree-day sum for *Pinus sylvestris*. Leaf senescence is assumed to vary linearly with latitude (Lamb 1915) in this version of the model. The regression is fitted on leaf colouring dates at several latitudes.

2. REPRODUCTIVE SUCCESS

The reproductive output corresponds to the proportion of mature fruits by the end of the year.

It is calculated as the product of fruit maturation success and the proportion of fruits that reach maturation (i.e. have not been killed by frost all along the season since the flower primordia). The proportion of fruits that reach maturity is calculated following the frost damage model of Leinonen (1996) parameterized for flowers and fruits. The success of maturation depends upon the proportion of uninjured leaves available for photosynthesis following a sigmoid function with parameter *pfe50*, the proportion of leaves that reduce by 50% the photosynthetic assimilates going to the fruits. It depends also on a drought index calculated with a water balance using precipitation, actual evapotranspiration and soil water holding capacity. Finally, it depends on temperature which determines the course of maturation. Fruits maturation date follows a normal distribution within the tree crown defined as $E_c \sim \mathcal{N}(matmoy, sigma)$, with *matmoy* and *sigma* expressed as a sum of developmental units and *sigma* chosen so that fruit maturation occurs over a month.

3. SURVIVAL TO STRESSES

Two kinds of stress are considered: frost and drought.

A lethal frost temperature is used in the model but never plays a role in determining species range limits.

Frost injury on buds, leaves, flowers and fruits is modelled according to the model of Leinonen (1996). Frost hardiness depends upon the organs' developmental stage, photoperiod and temperature. Frost hardiness is highest during the dormancy phase, and lowest during bud burst. Frost can injure buds, leaves, flowers and fruits.

In this version of PHENOFIT, survival to drought was implemented grossly, and the species were attributed an upper and a lower bound of sustainable precipitation. Outside these limits, survival was assumed to be 0.1.

4. INPUT AND OUTPUT VARIABLES

In the version of PHENOFIT used in this study, input variables are daily minimal and maximal temperatures, and monthly amount of precipitation. The model outputs a proxy for fitness within [0,1], which the product of survival and reproductive success, for each cell and each year. For each cell, fitness is averaged over a 20-year time period (1981-200 for the "current" climate; 2081-2100 for scenarios) to produce the maps.

5. AVAILABILITY AND USE

PHENOFIT is available upon request to Isabelle Chuine (isabelle.chuine@cefe.cnrs.fr). This model has been used at the continental scale and validated for a dozen of American tree species (Chuine & Beaubien 2001; Morin et al. 2007b; Morin et al. 2008; Morin & Thuiller 2009).

2. MODEL PARAMETERIZATION

Model parameters are provided in **table S3**.

1. PHENOLOGY

The model parameters are found through minimizing the residual sum of squares using the simulated annealing algorithm of Metropolis following Chuine et al. 1998. The fitting procedure uses observations of leaf unfolding, flowering, fruit maturation and leaf senescence dates, in different populations of the same species. These observations were retrieved from the French phenological database (Observatoire des Saisons, GDR2968, http://www.gdr2968.cnrs.fr) except for *Pinus sylvestris* (see below). Leaf unfolding and leaf senescence observations encompassed the period 1997-2006; flowering dates were observed in 2006-2008; and fruit maturation from 1990 onwards.

Because local adaptation to climate may modify species response to climatic clues, observation sites were grouped according to the genetic pool of the observed trees ("provenance regions" as defined by the French Forest Inventory, http://agriculture.gouv.fr/foret). Each phenological model was then calibrated for each provenance region, using daily meteorological data from the closest meteorological station (provided by MeteoFrance). Whenever possible, adjacent provenance regions which did not significantly differ in their phenological response to temperature were combined following Chuine *et al.* 2000. For *Pinus sylvestris*, phenological observations were too scarce in the database, resulting in doubtful parameter values. We therefore used parameters fitted by Kramer (1994) on a German provenance.

2. SURVIVAL TO STRESSES

Precipitation limits determining the resistance to drought stress were taken from the French Forest Inventory, http://agriculture.gouv.fr/foret. Lethal temperatures were those identified by Sakai & Larcher (1973). Parameters for the frost damage model were those of Leinonen (1996), except the minimum and maximum hardiness, which were compiled from the literature (refs. in Morin *et al.* 2007a). All these parameters were species-specific.

Table S3. Parameters used in PHENOFIT. The number *n* of observations used to fit the leafing and fructification models is given, as well as the proportion of variance in bud burst/fructification dates explained by the model (R^2). Phenological models for *Pinus sylvestris* were extracted from Kramer (1994).

	Fagus sylvatica				*Quercus robur*			*Pinus sylvestris*
	Prov201	Prov403	Prov602	Prov751	Prov100	Prov201	Prov361	
Leaf unfolding date								
a	0.54	3.65	1.04	1.13	1.14	0.56	0.96	0,06
b	-19.52	-22.01	-26.68	-28.43	-22.04	-3.52	-21.51	1.00
c	-19.86	11.00	-6.22	-13.57	-1.71	0.18	-4.20	6.00
d	-40	-0.10	-40	-7.13	-40	-40	-0.35	-0,11
e	8.40	2.73	8.55	9.94	6.72	9.92	6.11	37.00
C*	202.51	12.75	218.85	136.43	182.39	4.82	210.79	85.00
F*	9.50	121.00	4.20	20.50	19.90	31.00	10.60	2.40
n	17	15	14	11	28	19	26	
R²	0.641	0.840	0.697	0.969	0.656	0.705	0.531	
Flowering date								
F**	18.50	129.50	12.70	29.00	25.39	35.80	15.92	2.20
Fruit maturation date								
g	-16.74				-10.15	-0.25	-3.97	N.A.
h	14.72				13.96	18.10	9.46	N.A.
Fcrit	9.26				5.97	30.28	120.82	500.00
Top	5.00				5.00	6.56	19.77	N.A.
matmoy	104.50				136.34	102.95	47.77	N.A.
sigma	50.14				37.44	46.34	28.21	57.00
pfe50	0.4				0.4	0.4	0.4	0.4
Tb	N.A.				N.A.	N.A.	N.A.	5.00
n	23				18	27	74	
R²	0.489				0.428	0.278	0.446	
Frost hardiness								
T1	10				10			10
T2	-16				-16			-16
NL1	10				10			10
NL2	16				16			16
Fruit								
Frmax1	-5				-12			-10
Frmax2	-20				-50			-50
Leaf								
Flmin	-4				-7			-5
Ftlmax	-13				-41			-47
Fplmax	-7				-21			-18,5
Flower								
Ffmin	-4				-7			-10
Ftfmax	-12				-60			-47
Fpfmax	-6				-20			-18,5
Precipitation Limits								
PPmin	730				600			560
PPmax	1440				2030			3200

II. CLIMATE DESCRIPTION

Because our consensus model is a statistical one, we chose to eliminate multicollinearity among environmental descriptors through describing the environment by a restricted number of independent composite variables. These were obtained through summarising the variation of eight potentially correlated climatic descriptors in a Principal Component Analysis (PCA). These descriptors included five climatic descriptors computed by STASH (mean temperature of coldest month *Tcold*, mean temperature of warmest month *Twarm*, number of chilling days *NBChi*, drought index *DRI5*, growing day degrees *GDD5*), the total amount of precipitation (*PrecTot*), the amount of precipitation when temperature is above 5°C (*Prec5*), and the coefficient of variation of precipitations among seasons (*CVprec*, the standard deviation of precipitation among seasons, scaled by average seasonal amount of precipitation).

The variance of each descriptor was similar among the three datasets (current, A1Fi and B2): maximal to minimal variance ratios were all below 3. It was thus very unlikely that the structure among variables induced by one of the three datasets would be masked by different structures induced by the other two datasets. Moreover, the principal axes of the PCA carried on individual datasets did not differ qualitatively, the angles between corresponding principal component axes in the multivariate space for the three datasets were below 30° (**Fig. S4**; compare panels a, c, e and b, d, f). To obtain principal components and axes, we thus performed a PCA on the concatenation of the three datasets (**Fig. S4**, panels g, h).
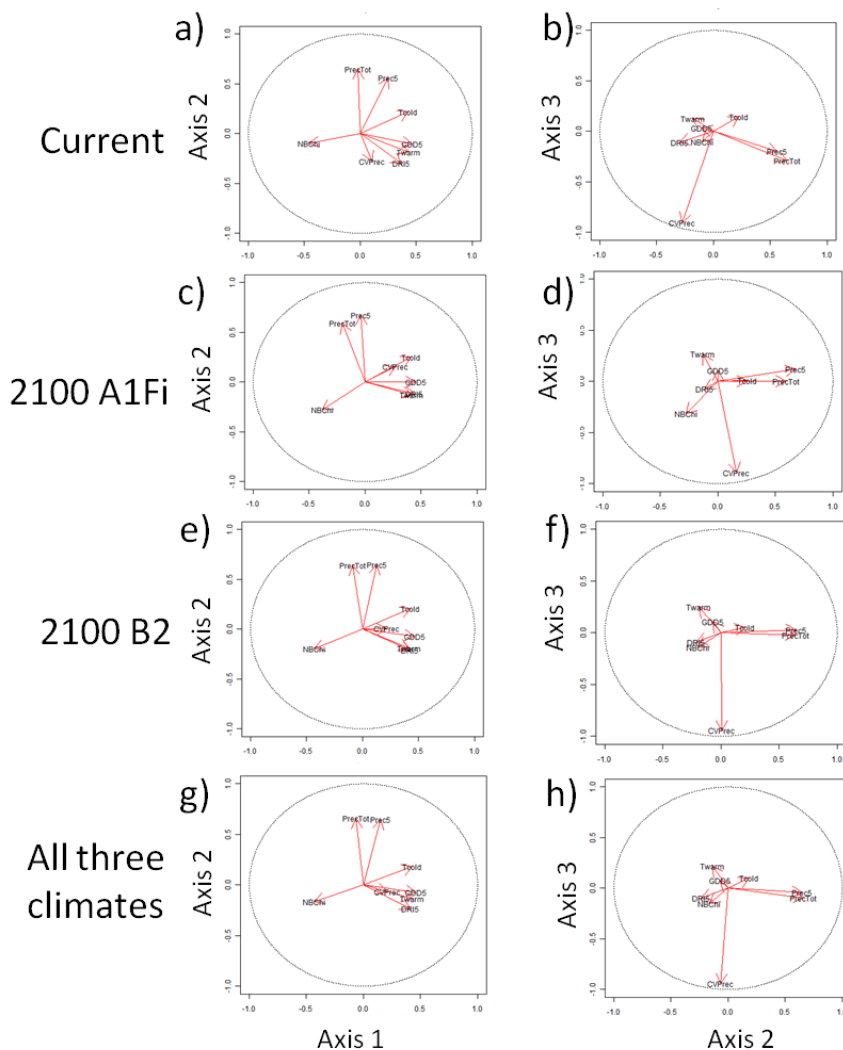

Figure S4: Correlation circles of the Principal Components Analysis for current climates (a, b), 2100 A1Fi scenario (c, d), 2100 B2 scenario (e, f) and all climates together (g, h). These circles are shown in two planes of the multivariate space: that defined by PC1 and PC2 (a, c, e, g) and that defined by PC2 and PC3 (b, d, f, h).

The first principal component (PC1) corresponded to a temperature axis; it gathered the five descriptors used by STASH, and explained 56% of the total variance. PC2 explained 25% of the total variance, and was carried by the descriptors of the amount of precipitation (*PrecTot* and *Prec5*). PC3 explained 11% of the total variance, and was mostly carried by the seasonality of precipitation (*CVprec*). These three axes together explained 92.8% of the total variance, and were used as synthetic mutually independent climate variables. **Figure S5** shows how European climates are described by PC1, PC2 and PC3.
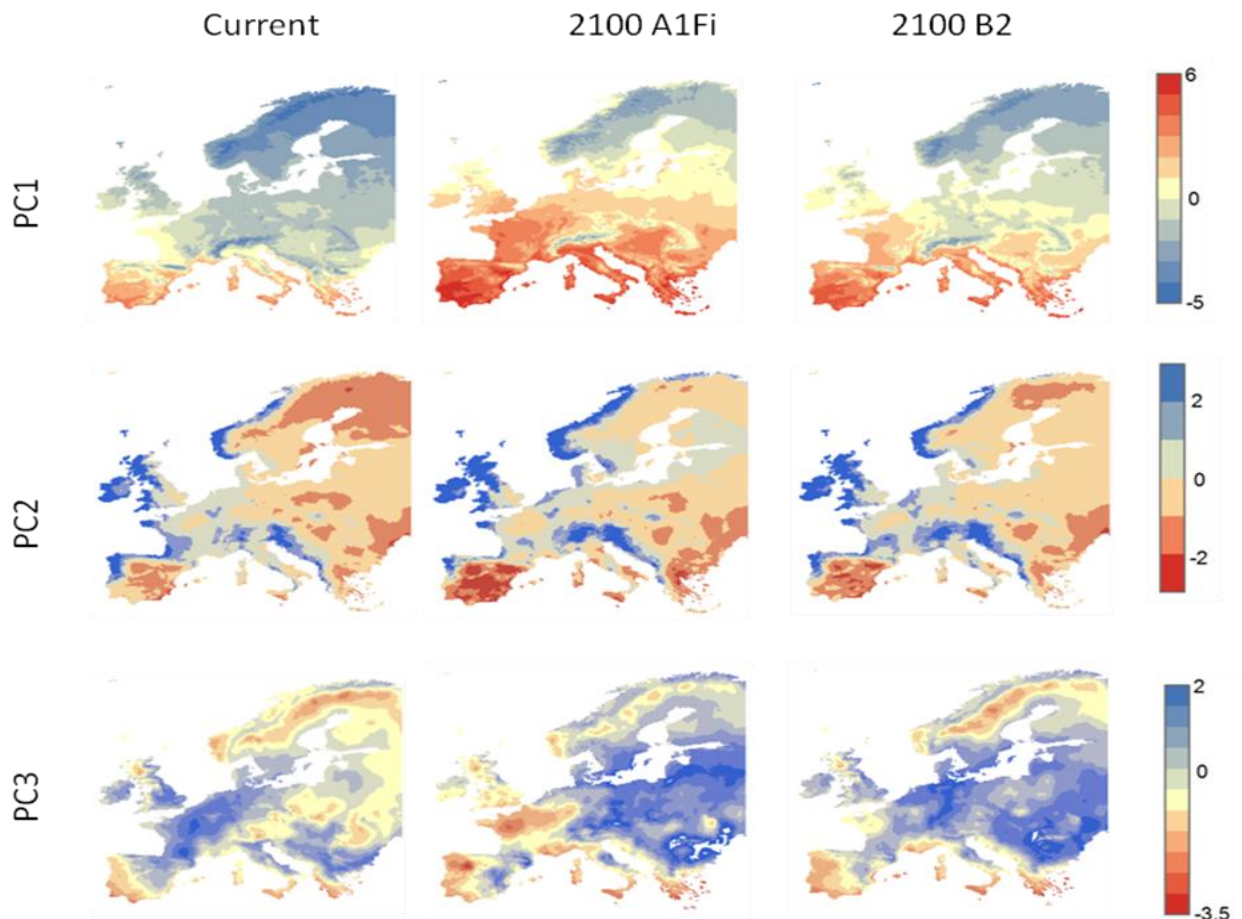

Figure S5: Coordinates of current (left column) and future (middle and right columns) climates in the principal component analysis. Axis 1 corresponds mostly to temperatures, with higher values denoting colder climates. Axis 2 corresponds to total precipitation, with higher values denoting wetter climates. Axis 3 is mostly carried by the seasonality of precipitations, with high values denoting regular amounts of precipitation across seasons.

III.   CONSENSUS MODEL

From this section and until the end of this file, all operations are provided in the attached R script "`ConsensusModel.R`". Data are located in folder `/data/`; the format of the data is explained in the script.

A.   DETERMINE A SPECIFIC PRESENCE THRESHOLD (UNIQUE TO EACH SPECIES AND MODEL, WILL MAXIMIZE THE SUM (SENSITIVITY + SPECIFICITY)).

Script "ConsensusModel.R" lines 40-45.

For STASH, the SPT was determined for each re-sampling of the distribution map (see section I A 2). For LPJ and Phenofit, the SPT was determined only once.

The threshold is determined using the data shown on the ROC plot of the model output. This plot shows the *sensitivity* of the model (ordinates, "Hit rate", *i.e.* % of pixels where the species is present, and predicted as such) as a function of (*1- its specificity*) (abscissas, the "False Alarm Rate"; *i.e.* the % of pixels where the species is present, but projected as present). Each point of this curve corresponds to the sensitivity and (1-specificity) of the model, when model outputs above some threshold *t* are attributed the value 1 (present) and those below *t* are attributed 0 (absent). The whole curve is obtained by varying *t*. An example is shown on **Figure S6**.
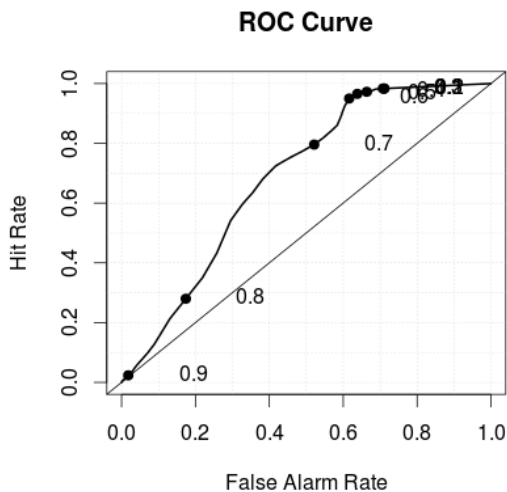

ROC Curve

Figure S6. ROC plot for the projection of *Pinus sylvestris* by LPJ. The values on the curve show 10 thresholds (*i.e.* when *t*=0.1, the sensitivity is very high, but the specificity low; when *t*=0.9 this is the opposite: almost no points are projected as present [*i.e.* have LPJ output>0.9], such that sensitivity is necessarily low).

A perfect model would yield a point in the top left corner of the graph. Maximizing the sum of the sensitivity and specificity is somehow trying to find the best compromise, that is, the threshold t that leads closest to that top left corner.

The function `roc.plot` (library `verification`) computes that plot and stores the points' coordinates in `$plot.data`, an array whose first layer only is used here. Column 1 contains the thresholds (101 equally spaced in our script); columns 2 and 3 contain the hit rate and false alarm rates (coordinates of the points on the ROC plot).

Function `maxss` extracts the value of *t* corresponding to the maximum of (sensitivity+specificity) from such a table:

```
maxss <- function(tab) {
    res<-matrix(ncol=2, nrow=0)
    for (i in 2:dim(tab)[1]) res <- rbind(res, c(tab[i,1], (tab[i,2]+(1-tab[i,3]))))
    res2 <- res[which(res[,2]==max(res[,2])),1]
    res2}
```

Hence the SPT for *Pinus sylvestris* as projected by LPJ:

```
 maxss(roc.plot(dataCurrent$PinusPA, dataCurrent$PinusLPJ,
     thresholds=seq(0,1,0.01))$plot.data[,,1])
[1] 0.59
```

B.   DATA SUBSETTING

Script "ConsensusModel.R" line 63.

For each species and each climatic dataset, each pixel in the simulation window was attributed to one of $2^3 = 8$ subsets, indexed by S, corresponding to the triplets of the combinations of {STASH, LPJ, PHENOFIT} projected presence (above the SPT) or absence (below the SPT). For example, {1,0,0} would be one such triplet, corresponding to STASH projecting occurrence, and LPJ and PHENOFIT projecting absence of the species in the pixel.

This was done using the `CreateSubsets` function, which creates a list, each element of the list corresponding to one combination of binary {STASH, LPJ, PHENOFIT} projected presences/absences (using the SPT as defined in section III A). Each element of the list contains the subset of the dataset presenting that combination of outputs.

```
CreateSubsets <- function(dataset) {
subdata <- list()
i <- 0
for (subb in namsub) {
i <- i+1
mod1<-F
mod2<-F
mod3 <-F
eval(parse(text=paste("if (substr(subb, start=1, stop=1)=='p') mod1 <- T " , sep="")))
eval(parse(text=paste("if (substr(subb, start=2, stop=2)=='p') mod2 <- T " , sep="")))
eval(parse(text=paste("if (substr(subb, start=3, stop=3)=='p') mod3 <- T " , sep="")))
eval(parse(text=paste("subdata[[", i, "]] <-subset(", dataset, ",subset=(", dataset, "$",
    species[spec], models[1], if(mod1)  " >= " else " < " , species[spec], models[1],
    "threshold & ", dataset, "$", species[spec], models[2],  if(mod2) " >= " else " < ",
    species[spec], models[2], "threshold & ", dataset, "$", species[spec], models[3],
    if(mod3) " >= " else " < ", species[spec], models[3], "threshold)) ", sep="")))
    }
return(subdata)
}
```

An example of the result is provided on **Figure S7**.

Figure S7. Data subsets obtained for *Pinus sylvestris*. The first three maps show the binary outputs of each SDM; and the fourth map indicates which pixels belong to which subset (indicated by a color code, with no hierarchy between colors). Note that subset 3 (medium blue) has few pixels [5]; it has even fewer under future conditions [3 in each scenario]. No model for deviance will be evaluated for subsets covering <100 pixels – this is the only one in our dataset. Code:

```
spec=3; # Pinus sylvestris
scen=1 # Current climates
subdataCurrent <- CreateSubsets('dataCurrent')
```
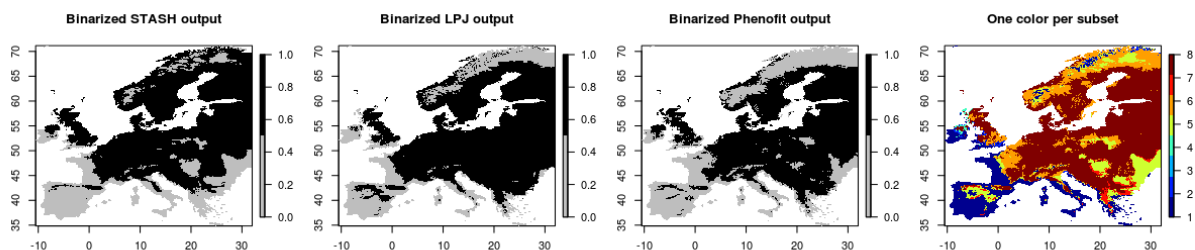

Because model projections vary among species, the eight subsets have to be defined for each species separately.

C.   DETERMINE THE PROBABILITY OF OCCURRENCE OF THE SPECIES WITHIN EACH SUBSET

Script "ConsensusModel.R" line 66.

This is simply the observed prevalence of the species over all points belonging to the same subset. For example, the probability of occurrence over all points belonging to subset 1 (all models predict absence; dark blue cluster on the map above) is the proportion of the dark blue pixels in which the species is actually observed.

`DefinePOccurrence` function defines one probability of occurrence for each of the subsets (as obtained by the `CreateSubsets` function).

```
DefinePOccurrence <- function(liste, sp) {
# where liste is a subsetted dataset, obtained by "CreateSubsets", and sp the index of the
    species
    pocc <- vector()
    for(i in 1:length(liste)) {
    eval(parse(text=paste("pocc[", i, "] <- dim(subset(liste[[", i, "]], liste[[", i,
    "]][,'", species[sp], "PA']==1))[1]/dim(liste[[", i, "]])[1]", sep="")))
    } # i.e.: pocc=n(occurrences within subset)/n(point in subset)
    pocc
    }

poccurPinus <- DefinePOccurrence(subdataCurrent, spec)

> round(poccurPinus,spec)
[1] 0.146 0.251 0.000 0.035 0.415 0.751 0.236 0.774
# (note the subset with p=0 is the one containing only 5 data points)
```

Storage: *Script "ConsensusModel.R" line 69.*
These proportions are stored as an additional column in the data frame containing all relevant information (`dataCurrent` here). This column will be labeled `pSpecies` (here, `pPinus`) and will contain, for each pixel, the relevant value of `poccurSpecies` (i.e. the one corresponding to the subset to which the pixel belongs). For example, all pixels corresponding to subset 1 ("all models projected absence") will receive the value `0.146`. This is done using the `StoreProb` function.

```
dataCurrent <- StoreProb(subdataCurrent, poccurPinus, dataCurrent)
```

(where the first argument gives the subsetted dataset (where information about the subset can be found), the second gives the vector containing the probabilities of occurrence corresponding to each subset; and the third argument is the target data frame, to which the resulting vector will be bound.

An example of the projected probability of occurrence is given on **Figure S8**.
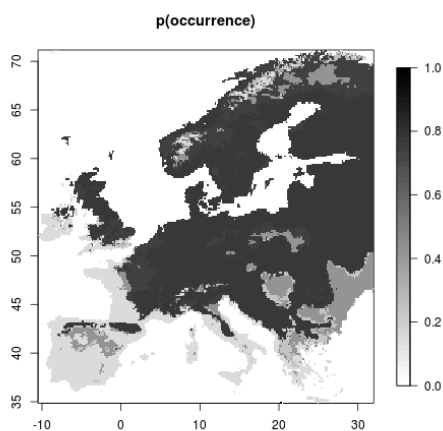

Figure S8. Projected probability of occurrence for *Pinus sylvestris*. Each subset (as defined on **Figure S7**, right panel) receives a different probability of occurrence (some subsets have close values [see above], hence fewer than 8 colors can be distinguished).

D. ESTIMATE THE ASSOCIATED UNCERTAINTY

1. COMPUTE THE OBSERVED DEVIANCE OF THE CONSENSUS MODEL'S OUTPUT TO OBSERVED OCCURRENCE DATA.

Script "ConsensusModel.R" line 72.

For each pixel, the deviance of the "conditional consensus" to the complete model (*i.e.* the observed occurrences) was computed as:

$$dev = -2\ln\left[obs\,\ln(p_S) + (1-obs)\ln(1-p_S)\right]$$ (7)

where *obs* is the observed occurrence (0 or 1) and p_S the estimated probability of occurrence for the relevant subset. Note that this is the exact deviance of projections (p_S) with respect to the observed occurrences (i.e. to a perfect model).

This was done using the StoreDeviance function. Note this function just has to be run for Current (reference) data, running it on other datasets would be irrelevant (no PA data being available). This function calls function deviancetot, which itself calls devianceind. devianceind computes minus twice the log likelihood :

$$devianceind = \begin{cases} -2\log(p_S) & \text{if the species is present} \\ -2\log(1-p_S) & \text{if it is absent} \end{cases}$$ (8)

```
devianceind <- function(obs, pred) {apply(cbind(obs, pred), FUN=function(tab) {
    -2*((tab[1]*log(tab[2])) + (1-tab[1]) * log(1-tab[2]))
    }, MARGIN=1)}
```

Function deviancetot sums deviance over all pixels:

```
deviancetot <- function(obs,pred) sum(devianceind(obs,pred))
```

Within each subset, deviance can take at most two values, so that the ratio of the difference between observed and minimum deviance to its maximum span (thereafter the "standardised deviance", noted δ) is 0 or 1. Function StoreDeviance stores the pixel-by-pixel observed deviance into the target dataset (Current), in the form of two additional columns:
- "devspecies" contains the inferred probability of occurrence for each pixel;
- "stddevspecies" stores the stardardized deviance δ: $\delta = \dfrac{deviance - \min(deviance)}{\max(deviance) - \min(deviance)}$,

where *min(deviance)* and *max(deviance)* depend upon p_S: $\min(deviance) = \min\left(-2\ln(p_S), -2\ln(1-p_S)\right)$.

```
StoreDeviance <- function (target) {
    eval(parse(text=paste("dev", species[spec], " <-devianceind(target[,'", species[spec],
    "PA'], target[,'p", species[spec], "'])", sep="")))
eval(parse(text=paste("target<-as.data.frame(cbind(target,dev", species[spec], "))", sep="")))

eval(parse(text=paste("stddev", species[spec], " <- round(apply( cbind(target[,'p",
    species[spec], "'], target[,'dev", species[spec], "']), FUN= function(vect) (vect[2]-
    min(-2*log(vect[1]), -2*log(1-vect[1]))) / abs(2*log(vect[1]/(1-vect[1]))), MARGIN=1))",
    sep="")))

eval(parse(text=paste("target<-as.data.frame(cbind(target, stddev", species[spec], "))",
    sep="")))

target}
```

Figure S9 shows the maps of observed and standardized deviance look like, for *Pinus sylvestris:*
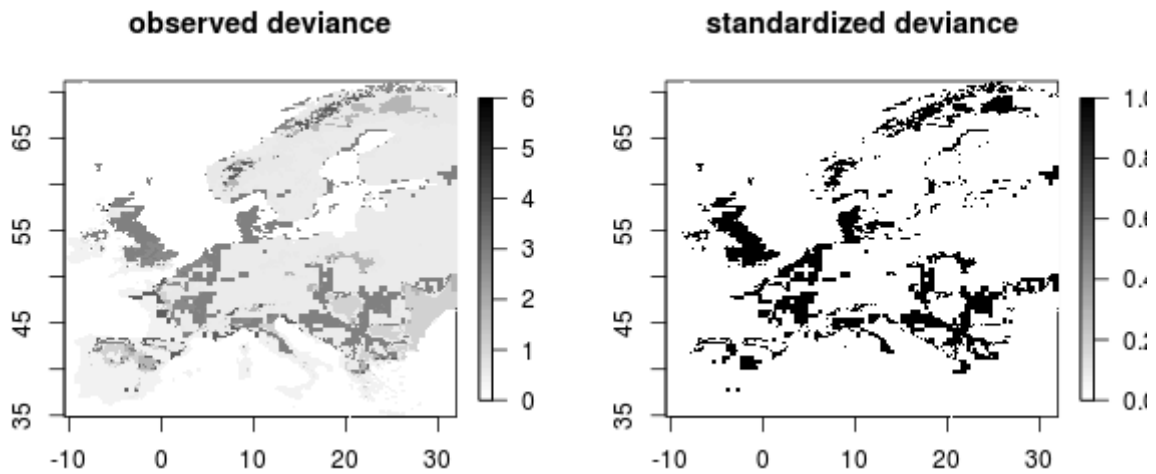

Figure S9. Observed deviance (left) and modelled standardized deviance (right), for *Pinus sylvestris*.

The data subsets have to be re-written so that they now contain information about deviance (*Script "ConsensusModel.R" line 75*).

2. ESTIMATE HOW THE OBSERVED STANDARDIZED DEVIANCE VARIES WITH THE ENVIRONMENT.

(The environment is defined by three independent axes, obtained by a PCA on environmental variables; section II.)

Script "ConsensusModel.R" lines 78, 81.

The standardised deviance δ was modelled as a Bernoulli process. The logit of its mean was assumed to be a polynomial function of the three synthetic climatic variables with polynomial degree ≤ 2:

$$\text{logit}(\delta) = \alpha_{0,S} + \alpha_{11,S}PC1 + \alpha_{12,S}PC1^2 + \alpha_{21,S}PC2 + \alpha_{22,S}PC2^2 + \alpha_{31,S}PC3 + \alpha_{32,S}PC3^2 \qquad (9)$$

Twenty-seven models for δ were considered for each species and each subset, corresponding to the combinations of $\alpha_{i1,S}$ and $\alpha_{i2,S}$ (with all models including $\alpha_{0,S}$, and models including $\alpha_{i2,S}$ constrained to also include $\alpha_{i1,S}$).

The 27 models are stored in separate files, named 'formulae/models_deviance_Species.txt'. Here are the models considered for *Pinus sylvestris* (the same models were considered for all three species):

```
1                          stddevPinus~1
2                       stddevPinus~1+PC11
3                       stddevPinus~1+PC21
4                       stddevPinus~1+PC31
5                    stddevPinus~1+PC11+PC12
6                    stddevPinus~1+PC11+PC31
7                    stddevPinus~1+PC21+PC31
8                    stddevPinus~1+PC11+PC21
9                 stddevPinus~1+PC11+PC21+PC22
10                stddevPinus~1+PC11+PC12+PC21
11             stddevPinus~1+PC11+PC12+PC21+PC22
12                   stddevPinus~1+PC11+PC31
13                stddevPinus~1+PC11+PC31+PC32
```

```
14               stddevPinus~1+PC11+PC12+PC31
15          stddevPinus~1+PC11+PC12+PC31+PC32
16               stddevPinus~1+PC21+PC31
17          stddevPinus~1+PC21+PC31+PC32
18          stddevPinus~1+PC21+PC22+PC31
19     stddevPinus~1+PC21+PC22+PC31+PC32
20          stddevPinus~1+PC11+PC21+PC31
21     stddevPinus~1+PC11+PC21+PC31+PC32
22     stddevPinus~1+PC11+PC21+PC22+PC31
23 stddevPinus~1+PC11+PC21+PC22+PC31+PC32
24     stddevPinus~1+PC11+PC12+PC21+PC31
25 stddevPinus~1+PC11+PC12+PC21+PC31+PC32
26 stddevPinus~1+PC11+PC12+PC21+PC22+PC31
27 stddevPinus~1+PC11+PC12+PC21+PC22+PC31+PC32
```

Each model is evaluated, together with the associated log-likelihood and Akaike Information Criterion (AIC). This s done using function `AICtableFunc`, which evaluates models stored in the appropriate formulae file, and stores their AIC in a table, whose rows are models and columns are subsets.

```
AICtableFunc <- function (listeAIC) {
        usable <- DimSubsets(listeAIC)
        eval(parse(text=paste("nmodglm <- length(formulaeDev", species[spec], "[,1])", sep="")))
        resAIC <- matrix(ncol=length(listeAIC),nrow=nmodglm)
        for (j in 1:length(listeAIC)) {
          if(usable$flag[j]) {
                for(i in 1:nmodglm) {  # evaluate the GLMs
                        eval(parse(text=paste("try(resAIC[i,j] <-
        glm(formula=as.formula(as.character(formulaeDev", species[spec], "[i,1])), family =
        binomial(link='", "logit", "'), data=listeAIC[[j]])$aic)", sep="")))
                }
          }
          else resAIC[,j] <- c(rep(0, nmodglm))
        }
        resAIC
}
```

Here is the AIC table obtained for *Pinus sylvestris*:

```
formulaeDevPinus <- read.table('formulae/models_deviance_Pinus.txt', header=F)
subdataCurrent <- CreateSubsets('dataCurrent')
AICtablePinus <- AICtableFunc(subdataCurrent)
round(AICtablePinus,0)
       [,1] [,2] [,3] [,4] [,5] [,6] [,7]  [,8]
 [1,]  3735  325    0   37 3888 5226  697 16812
 [2,]  2726  283    0   38 2597 4193  661 14457
 [3,]  3733  289    0   22 3889 4983  696 15611
 [4,]  3625  223    0   19 2999 3897  699 15176
 [5,]  2726  227    0   40 2599 4187  654 14454
 [6,]  2702  218    0   20 2512 3820  661 14026
 [7,]  3606  212    0   18 2913 3755  698 14331
 [8,]  2704  284    0   23 2574 4146  663 13900
 [9,]  2599  285    0   25 2561 3997  665 13872
[10,]  2673  229    0   24 2576 4130  656 13891
[11,]  2588  230    0   26 2563 3994  658 13861
[12,]  2702  218    0   20 2512 3820  661 14026
[13,]  2690  220    0   21 2483 3660  663 13989
[14,]  2695  195    0   21 2511 3816  656 14017
[15,]  2676  197    0   23 2483 3656  654 13978
[16,]  3606  212    0   18 2913 3755  698 14331
[17,]  3529  214    0   20 2830 3531  699 14232
[18,]  3544  210    0   20 2907 3737  700 14333
[19,]  3423  212    0   22 2805 3533  701 14232
[20,]  2668  213    0   19 2458 3727  663 13509
[21,]  2656  214    0   21 2440 3506  665 13421
[22,]  2586  212    0   20 2420 3711  665 13505
[23,]  2553  214    0   22 2395 3508  667 13412
[24,]  2667  196    0   20 2460 3717  658 13490
[25,]  2657  198    0   22 2442 3501  656 13400
[26,]  2587  190    0   22 2422 3706  659 13487
[27,]  2554  191    0   23 2397 3503  657 13391
```

Note that for subset 3, no model was evaluated (all models receiving a zero just in order to fill the table). In some cases, a number of models display very close AICs (e.g. for subset 7).

Instead of choosing the best model among those tested, we will use all models, weighted by their Akaike weights *w* (Burnham & Anderson 2002):

$$w_{model} = \frac{e^{(\min(AIC) - AIC_{model})/2}}{\sum_{m=1}^{n(models)} e^{(\min(AIC) - AIC_m)/2}}$$
(10)

The `AkaikeWeightsFunc` function computes these weights:

```
AkaikeWeightsFunc <- function (tab) {
        resAW <- matrix(nrow=dim(tab)[1], ncol=dim(tab)[2])
        for (i in 1: dim(tab)[2]) {
                for (j in 1:dim(tab)[1])
                        resAW[j,i] <- exp(-(tab[j,i]-min(tab[,i]))/2)
                resAW[,i] <- resAW[,i] / sum(resAW[,i])
                }
        resAW
        }
```

Which gives the following table of Akaike weights for *Pinus sylvestris*:

```
AkaikeweightsPinus <- AkaikeWeightsFunc(AICtablePinus)
round(AkaikeweightsPinus,3)
        [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]
 [1,]  0.000 0.000 0.037 0.000 0.000 0.000 0.000 0.000
 [2,]  0.000 0.000 0.037 0.000 0.000 0.000 0.008 0.000
 [3,]  0.000 0.000 0.037 0.018 0.000 0.000 0.000 0.000
 [4,]  0.000 0.000 0.037 0.099 0.000 0.000 0.000 0.000
 [5,]  0.000 0.000 0.037 0.000 0.000 0.000 0.255 0.000
 [6,]  0.000 0.000 0.037 0.063 0.000 0.000 0.006 0.000
 [7,]  0.000 0.000 0.037 0.129 0.000 0.000 0.000 0.000
 [8,]  0.000 0.000 0.037 0.009 0.000 0.000 0.003 0.000
 [9,]  0.000 0.000 0.037 0.003 0.000 0.000 0.001 0.000
[10,]  0.000 0.000 0.037 0.006 0.000 0.000 0.094 0.000
[11,]  0.000 0.000 0.037 0.002 0.000 0.000 0.039 0.000
[12,]  0.000 0.000 0.037 0.063 0.000 0.000 0.006 0.000
[13,]  0.000 0.000 0.037 0.024 0.000 0.000 0.003 0.000
[14,]  0.000 0.047 0.037 0.030 0.000 0.000 0.097 0.000
[15,]  0.000 0.018 0.037 0.012 0.000 0.000 0.281 0.000
[16,]  0.000 0.000 0.037 0.129 0.000 0.000 0.000 0.000
[17,]  0.000 0.000 0.037 0.050 0.000 0.000 0.000 0.000
[18,]  0.000 0.000 0.037 0.053 0.000 0.000 0.000 0.000
[19,]  0.000 0.000 0.037 0.021 0.000 0.000 0.000 0.000
[20,]  0.000 0.000 0.037 0.093 0.000 0.000 0.002 0.000
[21,]  0.000 0.000 0.037 0.035 0.000 0.044 0.001 0.000
[22,]  0.000 0.000 0.037 0.046 0.000 0.000 0.001 0.000
[23,]  0.567 0.000 0.037 0.020 0.731 0.018 0.000 0.000
[24,]  0.000 0.021 0.037 0.043 0.000 0.000 0.036 0.000
[25,]  0.000 0.009 0.037 0.018 0.000 0.686 0.108 0.012
[26,]  0.000 0.601 0.037 0.023 0.000 0.000 0.016 0.000
[27,]  0.433 0.304 0.037 0.012 0.269 0.252 0.042 0.988
```

While in some subsets (e.g. subset 8) one model contributes to most of the multimodel, for other subsets, a larger number of models contribute to the multimodel (e.g. subset 7). (Again here, subset 3 will not be used to predict deviance – deviances will be set to 0 for the 5 pixels concerned).

E.   MODEL AVERAGING: ESTIMATE THE MODELED DEVIANCES, GIVEN THE ENVIRONMENTAL VARIABLES.

Script "ConsensusModel.R" line 84.

Within each subset, modeled deviance is projected according to the 27 models identified above, and deviances are weighted according to their Akaike weights:

$$\overline{dev_S} = \sum_{j=1}^{24} w_{S,j} dev_{S,j} \tag{11}$$

This is done using function predSTDDevianceFunc.

```
predSTDDevianceFunc <- function(scen, aw, spec) { #scen=scenario (#);
                                           # aw=Akaike weights table; spec=species (#)

    # this function is run both for current and future datasets:
    eval(parse(text=paste("listeCurr <- CreateSubsets('data", scenarios[1], "')", sep="")))
    eval(parse(text=paste("listescen <- CreateSubsets('data", scenarios[scen],"')",
    sep="")))

    # modelled deviances will be stored in the list "resultSimDev"
    resultSimDev <- list()

    # "rowNames" is used to keep track of the original position of each pixel in the
    original dataset:
    rowNames <- list()

    # "usable" identifies the pixels belonging to subsets for which models for deviance can
    be estimated,
    # i.e. to subsets harboring more than 100 pixels in the "Current" dataset.
    # If a subset of the "Current" dataset happened to gather fewer than 100 pixels, then no
    model for deviance is estimated (thus, no projections will be made for future deviance).
    # In practice, there is 1 such subset in our data, harboring 5 pixels under current
    conditions and 12-26 pixels under scenarios.
     usable <- DimSubsets(listeCurr)$flag * DimSubsets(listescen)$flag


    # dime is a vector storing the numbers of pixels in each subset, for the scenario
    considered.
    dime <- DimSubsets(listescen)$dimsub

    # nmodglm will store the number of models evaluated for each subset (27 in our case)
    eval(parse(text=paste("nmodglm <- length(formulaeDev", species[spec], "[,1])", sep="")))

    # Results will be stored in the "target" dataset, i.e. the "dataCurrent", "dataA1Fi", or
    "dataB2" datasets
    eval(parse(text=paste("target <- data", scenarios[scen], sep="")))

    # Each subset is treated separately
    for (it in 1:length(namsub)) {
      # and only if it is usable (i.e. the corresponding subset in the "Current" dataset has
    > 100 pixels)
      if (usable[it]==1) {
      # predict values for deviance according to the nmodglm models. Store it in the list
    "resultSimDev", which will contain 8 elements (one per subset), each consisting of a
    matrix with nmodglm columns and npixels rows.
       eval(parse(text=paste("resultSimDev[[", it, "]] <- rbind(sapply(1:nmodglm,function(ii)
    predict(glm(formula=as.formula(as.character(formulaeDev", species[spec], "[ii,1])),
    family = binomial(link='logit'), data=listeCurr[[", it, "]]), type='r',
    newdata=listescen[[", it, "]]))) %*% aw[,", it, "]", sep="")))
       # Initial row names (=order) are stored in "rowNames".
       eval(parse(text=paste("rowNames[[", it, "]] <- row.names(listescen[[", it, "]])",
    sep="")))
       }

      # If no projections can be made for this subset, all points will (arbitrarily) receive
    the value "0"; row names will be stored in "rowNames" as well.
      if (usable[it]==0) {
            if(dime[it]>0) {
```

```
                    eval(parse(text=paste("resultSimDev[[", it, "]] <- c(rep(0,dime[", it,
    "]))", sep="")))
              }
         else {
               resultSimDev[[it]] <- NULL
         }
         eval(parse(text=paste("rowNames[[", it, "]] <- row.names(listescen[[", it,
    "]])", sep="")))
    }
}


    # Projections made for each subset are bound together
    stddevproj <- matrix(ncol=2)
    for(ite in 1:length(namsub)) { if(length(resultSimDev[[ite]]) > 1)  stddevproj <-
    rbind(stddevproj, cbind(resultSimDev[[ite]], rowNames[[ite]])) }
    # remove the first row of stddvproj (containing "NA  NA")
    stddevproj <- stddevproj[2:(npix+1),]
    # Order the rows
    stddevproj <- stddevproj[order(as.numeric(stddevproj[,2])),]

    # "Unstandardize" the standardized deviance and store it into "predDevianceSpecies"
    eval(parse(text=paste("predDeviance", species[spec], " <-
    apply(cbind(as.numeric(target[,'p", species[spec], "']), as.numeric(stddevproj[,1])),
    FUN= function(vect) (vect[2] * abs (2* log(vect[1] / (1-vect[1]) ) ) ) + min (-
    2*log(vect[1]), -2*log(1-vect[1])), MARGIN=1)", sep="")))

    # Bind this "predicted deviance" to the "target" data frame.
    eval(parse(text=paste("target <- cbind(target, predDeviance", species[spec], ")",
    sep="")))
    target
}
```

Observed and projected occurrence and deviance are shown on **Figure S10**.

Figure S10: observed occurrence (top left), modeled occurrence (top right), observed (bottom left) and modeled deviances (bottom right. This figure is provided by line 87 of the "ConsensusModel.R" script, using function CurrentPoccAndDevPic (Code: dataCurrent <- predSTDDevianceFunc(1, AkaikeweightsPinus, spec); CurrentPoccAndDevPic(spec)).
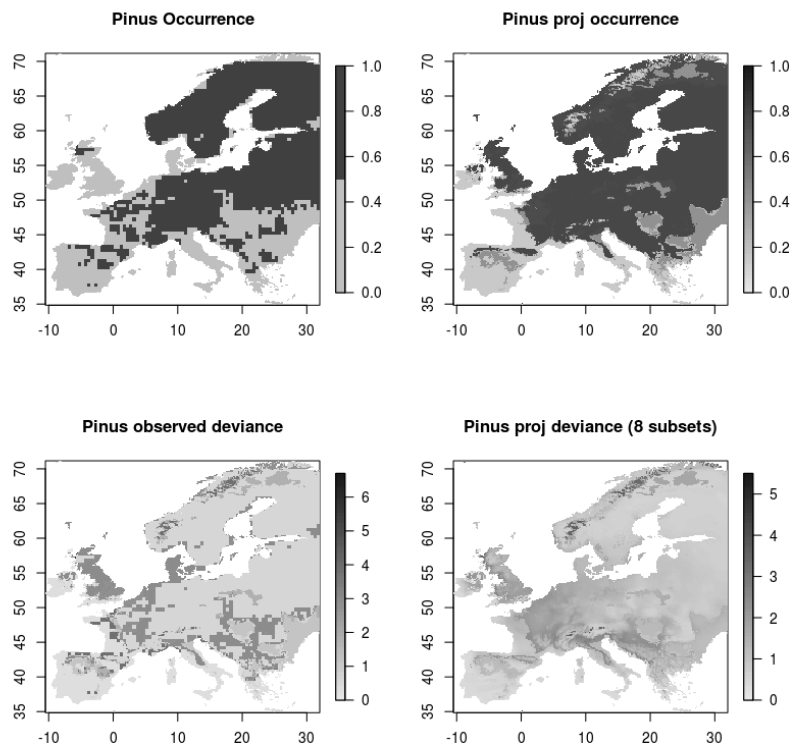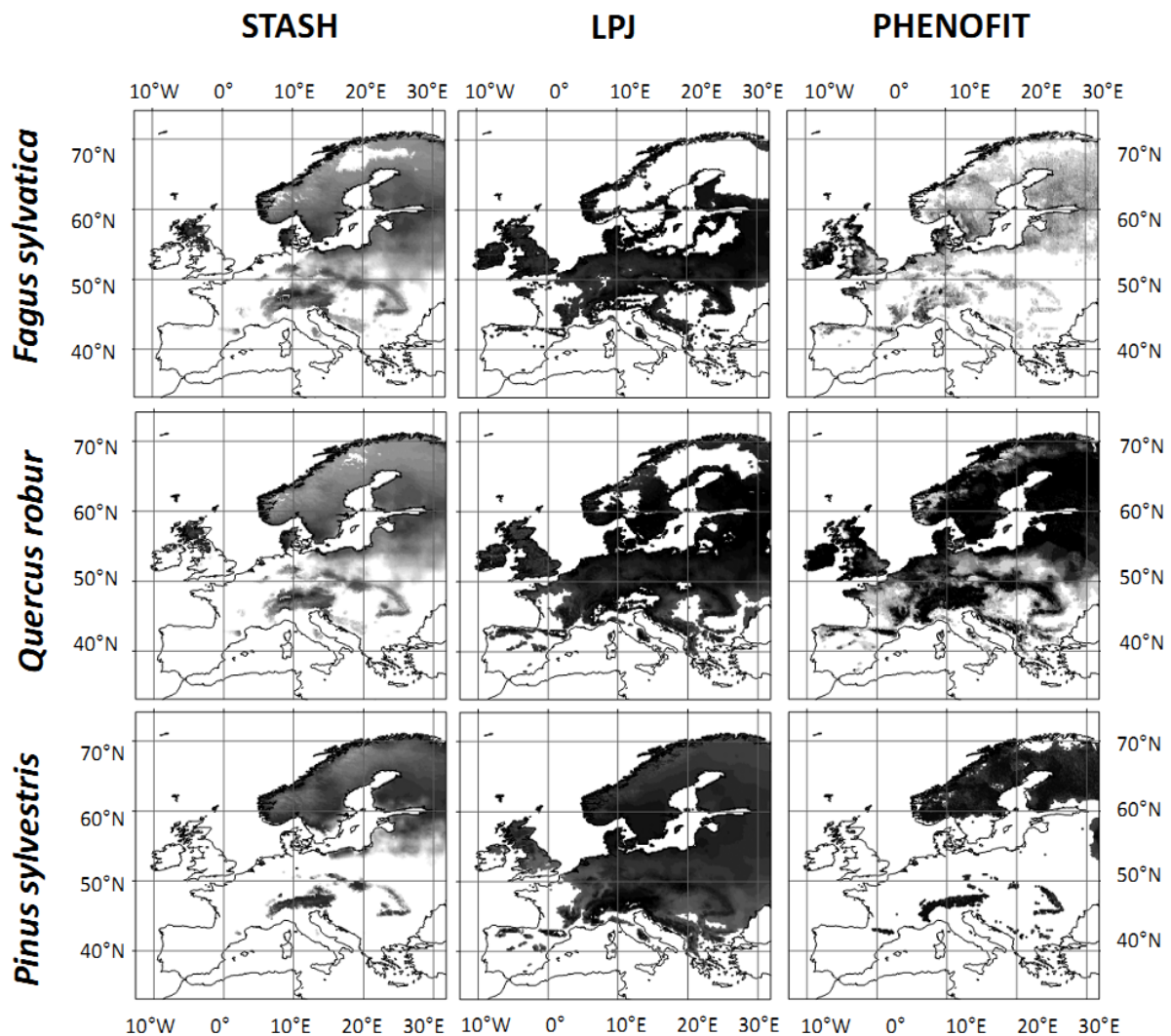
IV. FORECASTS

A. FORECASTS OF THE SDMS

1. SCENARIO A1FI, 2080-2100

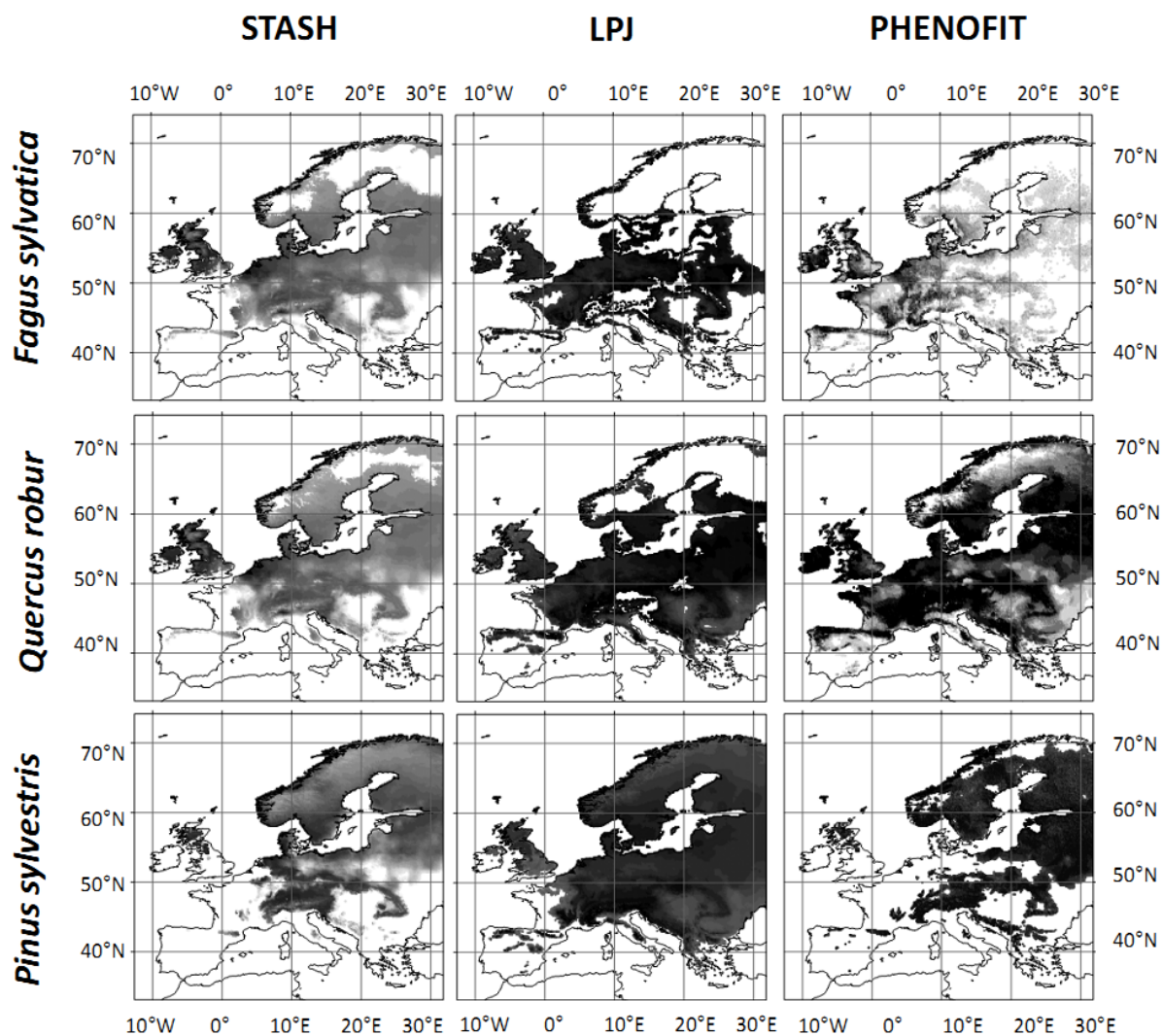These are shown on **Figure S11**.

Figure S11. SDM outputs for the A1Fi scenario (period 2080-2100). Darker shades indicate higher values of SDM outputs.

2. SCENARIO B2, 2080-2100

SDM outputs are shown on **Figure S12**.

Figure S12. SDM outputs for the B2 scenario (period 2080-2100). Darker shades indicate higher values of SDM outputs.

Script "ConsensusModel.R" lines 101-104.

Model outputs were transformed into presences/absences using the SPT, as defined above (*script line 101*). This generates eight subsets of {STASH, LPJ, Phenofit} projected presence/absence; each subset receives the probability of occurrence associated to that subset under current conditions (e.g. the {absence, absence, absence} dataset for *Pinus* will receive a probability of occurrence of 0.14 under future conditions, as under current conditions). These probabilities are stored in the relevant datasets, as an extra column named 'pSpecies' (e.g. 'pPinus'). This is done by the StoreProb function (described above).

C.   FORECASTS OF THE EXPECTED DEVIANCE (UNCERTAINTY).

Script "ConsensusModel.R" line 107.

The expected deviance is obtained, as for the current dataset, by extrapolating the (in our case 27) GLMs describing the variation of standardized deviance as a function of environmental variables, using the forecast conditions; averaging these projections using the relevant Akaike weights; and un-standardize the projected deviance. This is done by the predSTDDevianceFunc function (described above).

Figure S13 shows the projected occurrence and projected deviance (uncertainties) for the three climatic datasets, for *Pinus sylvestris*.

Figure S13. Projected occurrences (top row) and projected deviances (uncertainties; bottom row) for all climatic datasets (Current [left], A1Fi [middle], B2 [right]). This image is generated by function ProjectedPoccAndDevPic.
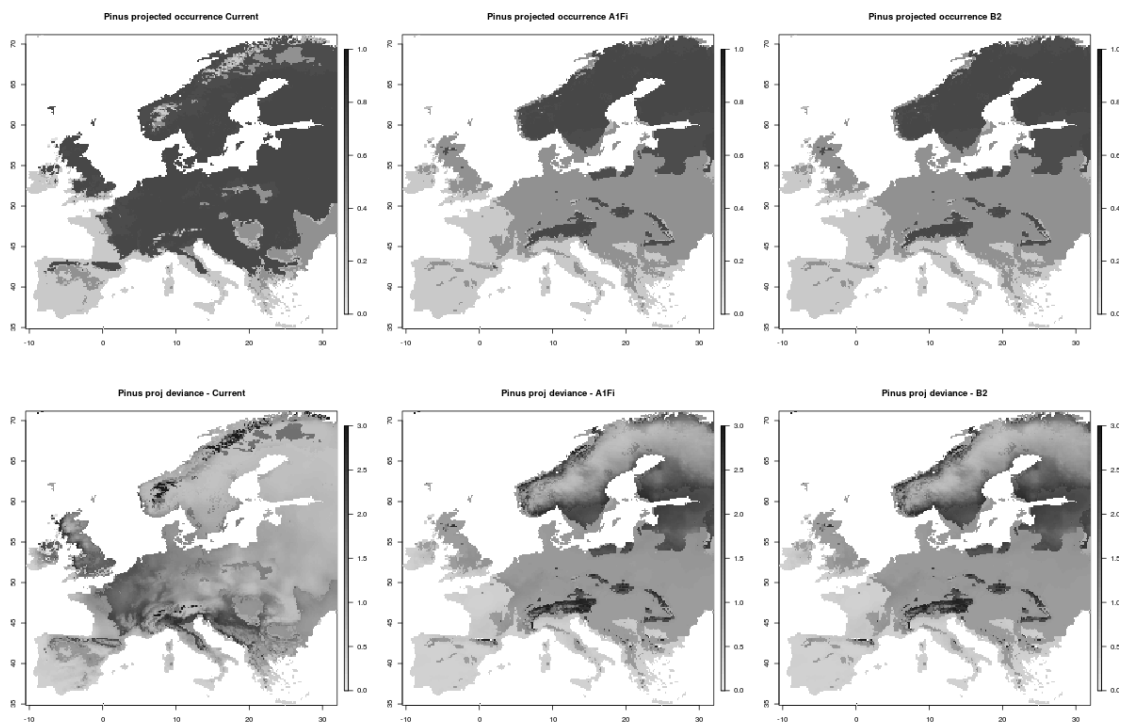
V. OTHER CONSENSUS MODELS& MODEL EVALUATION

A. BUILD OTHER CONSENSUS MODELS

Script "ConsensusModel.R" lines 145-169.

We also generated three consensus models for each species, using methods described in Marmion *et al.* (2009).
To this aim, it is necessary to reload the dataset, this time using the *continuous* projections of STASH (see section I A 2) (*"ConsensusModel.R" lines 138-143*).

The Mean consensus attributes to each pixel the mean probability of occurrence generated by the three SDMs (note that the three models used here do not generate probabilities, but proxies of such probabilities):

$$MeanConsensus_i = \frac{1}{n}\sum_{m=1}^{n} p_{m,i} \qquad (12)$$

where $p_{m,i}$ is the output of model *m* for pixel *i*, and *n* the number of SDMs (it is computed on *line 160 of the script "ConsensusModel.R"*).

The Median consensus attributes to each pixel the median of the outputs generated by the three SDMs. It is computed on *line 163 of the script "ConsensusModel.R"*.

The Weighted Average (WA) consensus weighs each SDM's output by the SDM's AUC (Area Under the [Receiver Operating] Curve):

$$WAConsensus_i = \frac{\sum_{m=1}^{n} AUC_m\, p_{m,i}}{\sum_{m=1}^{n} AUC_m} \qquad (13)$$

(see also equation (1) in Marmion *et al.* 2009). This requires computing the AUC of each SDM (*Script "ConsensusModel.R" lines 145-150*), and the WA consensus is computed on *line 166,* using the function `WAfunc`.

```
WAfunc <- function (datas) {
      res <- matrix(nrow=dim(datas)[1], ncol=length(models))
      for(aa in 1:length(models)) {
       res[,aa] <- (AUCm[aa] * datas[,aa])/sum(AUCm)
      }
      apply(res, FUN=sum, MARGIN=1)
      }
```

Projections made by these three consensus methods are stored in the objects "dataScenario" as three additional columns per species and per scenario ("SpeciesMeanCons", "SpeciesMedianCons", "SpeciesWACons") (*line 169*).

B. EVALUATE SDMS AND CONSENSUS MODELS.

Script "ConsensusModel.R" lines 128-228.

The (in our case, three) SDMs and the four consensi (Mean, Median, WA and Conditional) are evaluated using several measures:

- AUC, using function "`roc.plot`" from library `verification` (*Script "ConsensusModel.R" lines 171-179*)

- The rate of commission error (proportion of pixels wrongly projected as present, out of the total number of pixels), using function "CommissionError" (*Script "ConsensusModel.R" lines 197-206*). This measure needs the specific presence thresholds to be computed for all models (in particular, the consensus models for which it had not been done previously – *lines 181-192*).

```
CommissionError <- function(obs,pred,thresh) {
      mean(apply(cbind(obs,pred), FUN=function(d) if(d[1]==0 & d[2]>=thresh) 1 else 0,
      MARGIN=1))}
```

- The rate of omission error (proportion of pixels wrongly projected as absent, as compared to the total number of pixels), using function "OmissionError" (*Script "ConsensusModel.R" lines 208-217*)

```
OmissionError <- function(obs,pred,thresh) {
      mean(apply(cbind(obs,pred), FUN=function(d) if(d[1]==1 & d[2]<=thresh) 1 else 0,
      MARGIN=1))}
```

- The accuracy (proportion of correctly predicted pixels – accuracy, omission error rate and commission error rate thus sum to 1) (*Script "ConsensusModel.R" line 221*).

Accuracy measures are output in a separate file for each species ("results/AccuracyMeasures_Species.txt").

For example, for *Pinus sylvestris,* here is what it looks like:

	SPT	AUC	Accuracy (%)	Omission error (%)	Commission errror (%)
LPJ	0.59	0.687	73.6	2.6	23.9
Pheno	0.70	0.685	66.8	18.7	14.5
STASH	0.30	0.638	75.6	6.9	17.5
ConditionalConsensus	0.75	0.744	76.4	7.2	16.4
MeanConsensus	0.54	0.683	70.0	13.1	16.9
MedianConsensus	0.31	0.669	71.6	4.8	23.6
WAConsensus	0.55	0.684	69.9	13.3	16.9

C. PROBABILITIES OF PRESENCE PROJECTED BY THE CONSENSUS MODEL; NUMBERS OF POINTS IN EACH SUBSET OF THE DATA

The conditional consensus model predicts one average probability of presence for each of the eight subsets corresponding to the combinations of *{STASH, LPJ, PHENOFIT}* predicted occurrence. In the following table, these probabilities of presence are indicated for each of these subsets. Subsets are labelled in the table by the model(s) projecting the species occurrence. The proportion of pixels for which all models agree on the species occurrence (or absence) is indicated in the last column; and the numbers of points per subset of the data is given for all three species and all three climatic conditions (current and 2081-2100 for both scenarios).

	Model(s) predicting occurrence	None	STASH only	LPJ only	PHENOFIT only	STASH & LPJ	STASH & PHENOFIT	LPJ & PHENOFIT	All	% agreement
Fagus sylvatica	p(occurrence)	0.051	0.245	0.223	0.159	0.741	0.341	0.523	0.813	
Number of pixels per subset	1981-2000	11205	3276	139	2244	1452	2913	262	7275	64.2%
	2081-2100, A1Fi	8431	1662	2795	705	4190	3033	2436	5514	48.5%
	2081-2100, B2	8817	2901	625	1399	4329	2313	1270	7112	55.4%
Quercus robur	p(occurrence)	0.069	0.676	0.459	0.349	0.850	0.660	0.136	0.829	
Number of pixels per subset	1981-2000	8601	596	719	1515	1474	2681	235	12945	74.9%
	2081-2100, A1Fi	6095	412	3244	354	1814	2128	3322	11397	60.8%
	2081-2100, B2	4835	1152	2102	1119	1858	2414	1535	13751	64.6%
Pinus sylvestris	p(occurrence)	0.146	0.262	0.417	0.000	0.751	0.035	0.231	0.774	
Number of pixels per subset	1981-2000	4489	294	2847	5	4590	114	648	15779	70.5%
	2081-2100, A1Fi	6456	3	11873	12	3620	0	386	6416	44.7%
	2081-2100, B2	5252	3	6270	26	4254	0	1322	11639	58.7%

REFERENCES

Badeck F.W., Lischke H., Bugmann H., Hickler T., Honninger K., Lasch P., Lexer M.J., Mouillot F., Schaber J. & Smith B. (2001). Tree species composition in European pristine forests: Comparison of stand data to model predictions. *Climatic Change*, 51, 307-347.

Chuine I. (2000). A unified model for budburst of trees. *Journal of theoretical biology*, 207, 337-47.

Chuine I. & Beaubien E.G. (2001). Phenology is a major determinant of tree species range. *Ecology Letters*, 4, 500-510.

Chuine I., Belmonte J. & Mignot A. (2000). A modelling analysis of the genetic variation of phenology between tree populations. *Journal of Ecology*, 88, 561-570.

Garreta V., Miller P.A., Guiot J., Hely C., Brewer S., Sykes M.T. & Litt T. (2010). A method for climate and vegetation reconst'ruction through the inversion of a dynamic vegetation model. *Climate Dynamics*, 35, 371-389.

Gerten D., Schaphoff S., Haberlandt U., Lucht W. & Sitch S. (2004). Terrestrial vegetation and water balance - hydrological evaluation of a dynamic global vegetation model. *Journal of Hydrology*, 286, 249-270.

Giesecke T., Hickler T., Kunkel T., Sykes M.T. & Bradshaw R.H.W. (2007). Towards an understanding of the Holocene distribution of Fagus sylvatica L. *Journal of Biogeography*, 34, 118-131.

Gritti E.S., Smith B. & Sykes M.T. (2006). Vulnerability of Mediterranean Basin ecosystems to climate change and invasion by exotic plant species. *Journal of Biogeography*, 33, 145-157.

Haxeltine A. & Prentice I.C. (1996). BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochemical Cycles*, 10, 693–709.

Hickler T., Smith B., Sykes M.T., Davis M.B., Sugita S. & Walker K. (2004). Using a generalized vegetation model to simulate vegetation dynamics in northeastern USA. *Ecology*, 85, 519-530.

Kaplan J.O. (2001). Geophysical Applications of Vegetation Modeling. In. Lund University Lund, Sweden.

Koca D., Smith B. & Sykes M.T. (2006). Modelling regional climate change effects on potential natural ecosystems in Sweden. *Climatic Change*, 78, 381-406.

Kramer K. (1994). A modelling analysis on the effects of climatic warming on the probability of spring frost damage to tree species in The Netherlands and Germany. *Plant, Cell and Environment*, 17, 367-377.

Lamb G.N. (1915). A calendar of the leafing, flowering and seeding of the common trees of the eastern United States. *Monthly Weather Review Supplement*, 5-19.

Laurent J.M., Bar-Hen A., Francois L., Ghislain M. & Cheddadi R. (2004). Refining vegetation simulation models: From plant functional types to bioclimatic affinity groups of plants. *Journal of Vegetation Science*, 15, 739-746.

Leinonen I. (1996). A simulation model for the annual frost hardiness and freeze damage of Scots pine. *Annals of Botany*, 78, 687-693.

Miller P., Giesecke T., Hickler T., Bradshaw R.H.W., Smith B., Seppa H., Valdes P.J. & Sykes M.T. (2008). Exploring climatic and biotic controls on Holocene vegetation change in Fennoscandia. *Journal of Ecology*, 96, 247-259.

Morales P., Sykes M.T., Prentice I.C., Smith P., Smith B., Bugmann H., Zierl B., Friedlingstein P., Viovy N., Sabate S., Sanchez A., Pla E., Gracia C.A., Sitch S., Arneth A. & Ogee J. (2005). Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. *Global Change Biology*, 11, 2211-2233.

Morin X., Ameglio T., Ahas R., Kurz-Besson C., Lanta V., Lebourgeois F., Miglietta F. & Chuine I. (2007a). Variation in cold hardiness and carbohydrate concentration from dormancy induction to bud burst among provenances of three European oak species. *Tree Physiology*, 27, 817-825.

Morin X., Augspurger C. & Chuine I. (2007b). Process-based modeling of species' distributions: What limits temperate tree species' range boundaries? *Ecology*, 88, 2280-2291.

Morin X. & Thuiller W. (2009). Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, 90, 1301-1313.

Morin X., Viner D. & Chuine I. (2008). Tree species range shifts at a continental scale: new predictive insights from a process-based model. *Journal of Ecology*, 96, 784-794.

Murray M.B., Cannell M.G.R. & Smith R.I. (1989). Date of budburst of fifteen tree species in Britain following climatic warming. *Journal of Applied Ecology*, 26, 693-700.

Prentice I.C., Cramer W., Harrison S.P., Leemans R., Monserud R.A. & Solomon A.M. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 19, 117–134.

Sakai A. & Weiser C.J. (1973). Freezing resistance of trees in North America with reference to tree regions. *Ecology*, 54, 118-126.

Sitch S., Smith B., Prentice I.C., Arneth A., Bondeau A., Cramer W., Kaplan J.O., Levis S., Lucht W., Sykes M.T., Thonicke K. & Venevsky S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9, 161-185.

Smith B., Prentice I.C. & Sykes M.T. (2001). Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. *Global Ecology and Biogeography*, 10, 621-637.

Sykes M.T., Prentice I.C. & Cramer W. (1996). A bioclimatic model for the potential distributions of north European tree species under present and future climates. *Journal of Biogeography*, 23, 203-233.

Tutin T.G., Heywood V.H., Burges N.A. & Moore D.M. (1964-85). *Flora Europaeae*. Cambridge University Press, Cambridge.

Walther G.R., Gritti E.S., Berger S., Hickler T., Tang Z.Y. & Sykes M.T. (2007). Palms tracking climate change. *Global Ecology and Biogeography*, 16, 801-809.