

Estimating consensus and associated uncertainty between inherently different species distribution models

Emmanuel S. Gritti^{1,2*}†, Anne Duputie^{1,3}†, Francois Massol^{1,3} and Isabelle Chuine¹

¹CEFE, UMR 5175, 1919, route de Mende, 34293 Montpellier Cedex 5, France; ²UMR SYSTEM, Montpellier SupAgro, 2 place Viala, Bât. 27, 34060 Montpellier Cedex 1, France; and ³IRSTEA – UR HYAX, 3275, route de Cézanne – Le Tholonet, CS 40061, 13182 Aix-en-Provence Cedex 5, France

Summary

1. Forecasting shifts in biome and species distribution is crucially needed in the current context of global change. So far, most projections of vegetation distribution rely on correlative species distribution models (SDMs). Yet, process-based or hybrid models based on explicit physiological description may be more robust to extrapolation under future climatic conditions. Differences between model projections may be wide, leading to scepticism among environmental stakeholders.

2. Here, we propose to combine outputs of several distribution models based on physiological responses, to produce both consensual maps of occurrences and maps of associated uncertainty. The consensus map relies on the conditional projections of each SDM. Because the models used are based on processes, their errors are likely to vary consistently with climate as some processes not implemented in a model might be important under a given set of climatic conditions. Uncertainty of the consensus model is thus assessed through multimodel regression of deviance maps with respect to current climatic conditions, and can be extrapolated to forecast climates.

3. We illustrate this approach using three SDMs, on three widely distributed European trees (*Fagus sylvatica* L., *Quercus robur* L. and *Pinus sylvestris* L.), and project their distributions under two scenarios. The conditional consensus outperforms classical methods of model consensus (i.e. to use the mean, the median or a weighted average of individual SDM outputs) in projecting current occurrences.

4. Consistently, with the results of individual SDMs, the conditional consensus projects that the suitable areas for *F. sylvatica* and *Q. robur* will expand towards north-eastern Europe, while that of *P. sylvestris* will contract. Projections of future occurrence are most uncertain towards the margins of the distribution (particularly the trailing edge).

5. Our approach can help modellers identify the limitations of each SDM and stakeholders pinpoint the regions of models agreement and highest certainty.

Key-words: climate change, correlative distribution models, environmental clustering of model errors, *Fagus sylvatica*, likelihood, model assessment, *Pinus sylvestris*, process-based distribution models, *Quercus robur*

Introduction

The latest Intergovernmental Panel on Climate Change (IPCC) scenarios are already exceeded by recent estimations of greenhouse gas emissions (Raupach *et al.* 2007), leaving open questions about the development of global climate modifications and their impact on natural ecosystems. Recent climatic and atmospheric composition changes have modified the distribution, structure and function of ecosystems (Walther, Berger & Sykes 2005), thus altering biodiversity and ecosystem services, and leading to socioeconomic and financial costs. Adaptive management strategies directly based on spatialized, comprehensive and robust projections of species distribution and extinction risks could help mitigate these effects (TEEB 2010).

To date, species distributions are mostly investigated using three types of species distribution models (SDMs): correlative, process-based and hybrid models (Peterson *et al.* 2011). Correlative SDMs infer correlations between current species occurrences and various environmental descriptors. Process-based SDMs describe the responses of selected traits or processes (such as phenology, resistance to stress, resource acquisition) to environmental descriptors, based on empirical observations, and estimate proxies of occurrence, such as growth or fitness (Kearney & Porter 2009). Hybrid SDMs associate correlative models to describe habitat suitability, and process-based models to narrow down to the realized niche, through describing e.g. population dynamics, dispersal and/or energy uptake. Correlative models allow the exploration a species' limiting environmental variables across its realized niche, while process-based models infer its fundamental niche. Because correlative models rely on widely available

*Correspondence author. E-mail: emmanuel.gritti@supagro.inra.fr

†These authors contributed equally to this work.

occurrence and climatic data, they are largely used in the literature. However, their extrapolation to novel climates is uncertain. It has been suggested to use smooth response curves and to refrain from making projections to climates that differ too much from currently observed climates (Elith, Kearney & Phillips 2010). In contrast, process-based SDMs are thought to be more robust to extrapolation to novel climates (Morin & Thuiller 2009; Dormann *et al.* 2012), because their parameterization relies solely on empirically determined response curves driving important processes regulating the species' probability of surviving and reproducing, with respect to environmental conditions.

Forecasts of future distribution vary according to the correlative model used (e.g. Pearson *et al.* 2006), and between correlative and process-based SDMs (Buckley 2008; Kramer *et al.* 2010; Cheaib *et al.* 2012; but see Kearney, Wintle & Porter 2010). This may puzzle stakeholders and policy makers, and jeopardize the credibility of species distribution projections. Ensemble or consensus approaches, using information provided by different SDMs, have been advocated to tackle this problem (Araújo & New 2007): models can vote for the species' presence or absence. Votes can be weighted by models' accuracies (e.g. Marmion *et al.* 2009), or models can be combined using multimodel inference (Burnham & Anderson 2002; see e.g. Gibson *et al.* 2004; Hartley, Harris & Lester 2006).

Providing consensus maps is, however, not sufficient to guide stakeholders. All models may agree with each other for wrong reasons (Elith, Kearney & Phillips 2010), potentially leaving systematic errors. Mapping the resulting uncertainty is therefore as important as mapping the consensual projection itself. Yet, few studies have provided uncertainty maps of SDM projections. Maps of model discrepancies (e.g. Hartley, Harris & Lester 2006) only inform on the uncertainty associated with different model projections, not the uncertainty associated with the relevance of the climatic descriptors or the processes considered. Should an important environmental descriptor have been omitted in the individual SDMs, its variation would be absent from any multimodel, and even the best model among those considered would be unable to accurately project the species' range (Elith, Kearney & Phillips 2010; Dormann *et al.* 2012). The performance of conceptually different SDMs may vary with environmental conditions: each SDM may surpass the others in projecting a species' presence under a given set of climatic conditions, for the environmental variables or the processes it considers are more relevant in these conditions.

Here, we build a simple consensus between SDMs relying on vegetation's physiological responses to climate. Its uncertainty due to the poor parameterisation or omission of important processes is assimilated to its statistical deviance to observed occurrence maps. To account for the environmental clustering of SDM errors, uncertainty is modelled as a function of composite, independent environmental descriptors, in a multimodel framework. Both the probabilities of occurrence and the associated uncertainty can then be projected onto forecasted climatic conditions.

We illustrate this approach through modelling the potential distribution of three common European tree species (*Fagus sylvatica* L., *Quercus robur* L. and *Pinus sylvestris* L.), combining the outputs of three conceptually different SDMs (one correlative with physiological basis, one hybrid and one process-based).

Materials and methods

Our approach is summarized in Fig. 1.

STEP 1: SPECIES DISTRIBUTION MODELS

Details on all three models and their parameterization are provided as Supplementary Information.

STASH (correlative model)

STASH is a correlative, physiologically based climate envelope model (Sykes, Prentice & Cramer 1996). It relies on bioclimatic limits restricting the species' envelope, and on variables acting as multipliers of the species' growth efficiency index. All bioclimatic limits and variables are assumed to have strong links with vegetation responses through important physiological mechanisms. Because bioclimatic limits are defined according to the observed species distribution, this model is likely to over fit. To avoid this, we ran this model 100 times, with bioclimatic limits defined on random re-samplings of 30% of the Atlas Flora Europaeae distribution map (AFE; Tutin *et al.* 1964-85). For each pixel, the final STASH output corresponded to the average of the outputs obtained for that pixel when belonging to the remaining 70% validation set (Supplementary Information).

LPJ (hybrid model)

Lund-Potsdam-Jena (LPJ) model is a general ecosystem model combining bioclimatic limits to the species' establishment and survival and mechanistic representations of physiology, biochemistry, vegetation dynamics and carbon and water fluxes (Sitch *et al.* 2003). A minimum set of bioclimatic limits defines the bioclimatic envelope of the species. From climatic, soil and CO₂ data, the model simulates different growth-related variables such as leaf area index (LAI) or net primary production (NPP). Here, we used the LPJ version described in Gritti, Smith & Sykes (2006), but did not take competition into account. LPJ was run at the species level, using specific parameters when available (Supplementary Information), or the generic parameters of the corresponding plant functional type described by Smith, Prentice & Sykes (2001). Because bioclimatic limits do not directly derive from the observed distribution, no cross-validation from resampling approaches was performed.

PHENOFIT (process-based model)

PHENOFIT (Chuine & Beaubien 2001) is a process-based SDM relying on the assumption that a temperate tree species' survival and reproductive success are related to its capacity to synchronize its annual life cycle with seasonal climatic variations, as well as to sustain temperature and water stresses. From daily temperature and precipitation records, PHENOFIT estimates survival and reproductive success for an average tree. This model intrinsically takes phenotypic plasticity into account through the reaction norms of phenology and resistance to stress in

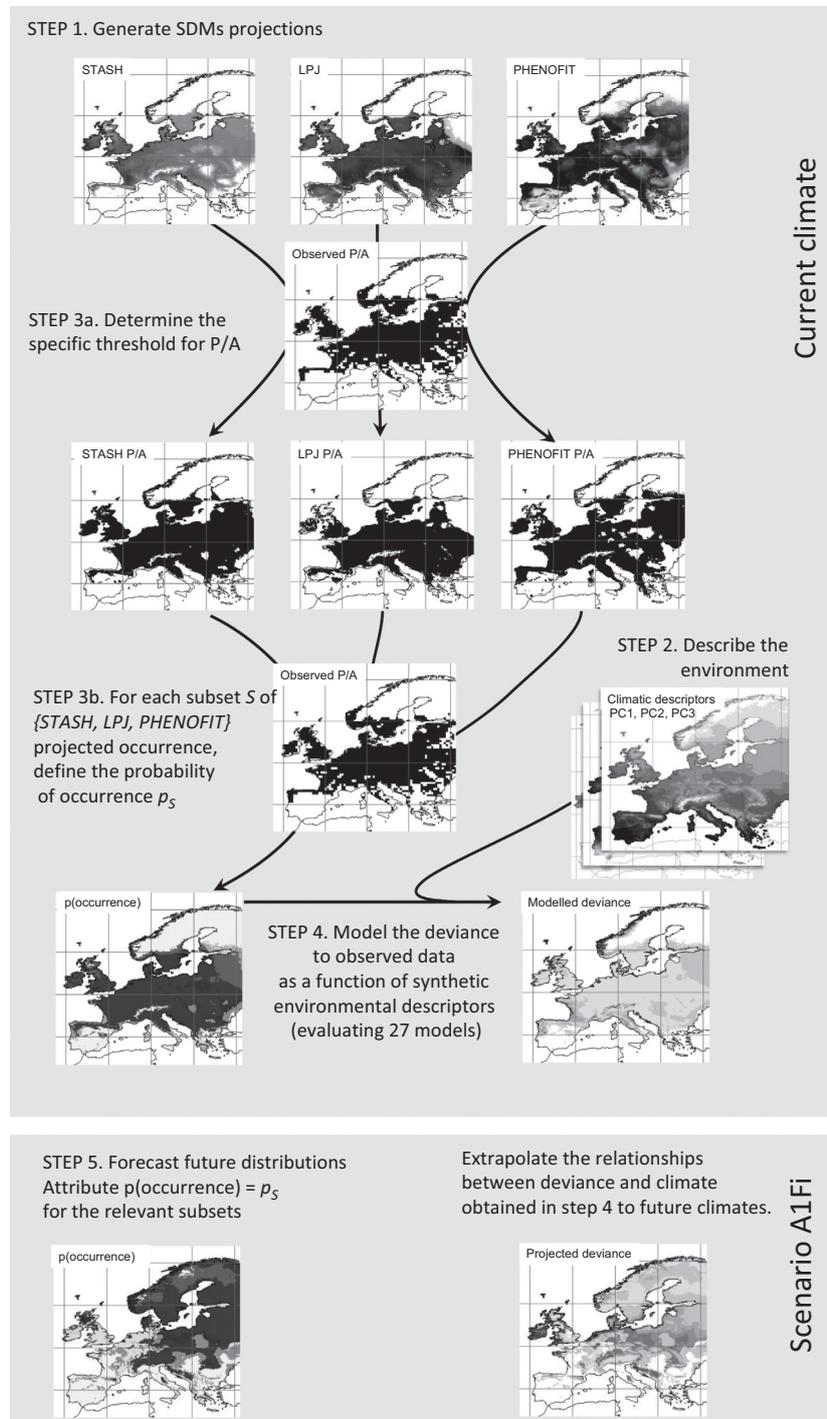


Fig. 1. Schematic workflow of the design of the consensus model and its forecasts. The maps illustrating the chart correspond to *Quercus robur*.

relation to climate. Here, PHENOFIT was parameterized for up to four populations per species, thus somewhat accounting for local adaptation. Observed species distribution is not used as input for the model, nor is it used to estimate parameters (parameters are derived from empirical observations of trees' physiological responses to climate). This model can therefore not be cross-validated by resampling approaches. As its output is related to the species' fundamental niche, the model can be validated *a posteriori*, by comparing its output to the observed distribution of the species.

STEP 2: CLIMATE DATA

SDMs simulations

Climatic and atmospheric CO₂ concentration time series were extracted from the Advanced Terrestrial Ecosystem Analysis and Modelling (ATEAM; <http://www.pik-potsdam.de/ateam>) data set for the period 1901–2100. Forecasts of climatic data were computed for the period 2081–2100 using the HadCM3 atmosphere–ocean general circulation

model (Mitchell *et al.* 2004) following two scenarios: A1Fi ('business as usual') and B2 (local development, with environmental focus). This climatic data set covers the European window from 11 °W, 34 °N to 32°E, 72°N with a 10' × 10' pixel resolution for monthly values of temperature, precipitation and percentage of sunshine. Twenty-year averages of monthly means served as input data for STASH. Monthly values were directly used as input data for LPJ. Daily interpolation was performed using the weather generator CLIGEN (Nicks, Lane & Gander 1995) to drive PHENOFIT.

Description of the climatic space used to explain the deviance of the consensus model

We eliminated multicollinearity between environmental descriptors, through summarizing the variation of eight potentially correlated climatic variables (four related to temperature, four to the amount and seasonality of precipitation) in a Principal Component Analysis carried on the concatenated climatic data sets (historical, 1981–2000 and scenarios, 2081–2100; Supplementary Material). The first three principal axes (PC) summarized 92.8% of the total variance of climatic descriptors, with PC1 mostly explained by temperature, PC2 by the amount of precipitations and PC3 by their seasonality. The coordinates of each pixel along these three axes were used as synthetic climatic descriptors.

STEP 3: CONDITIONAL CONSENSUS MODEL AND ASSOCIATED UNCERTAINTY

All three models produce different synthetic estimates (growth efficiency index, LAI, fitness), none of them actually being a probability of occurrence, and none being directly comparable to each other. We thus decided to transform their outputs into comparable binary presence/absence data, using model- and species-specific thresholds (hereafter, SPT). The SPTs were defined so as to maximize the sum of sensitivity and specificity (see e.g. Nenzén & Araújo 2011), using the AFE occurrence data as a reference.

For each species and each climatic data set, 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. Within each subset, all pixels shared the same probability of occurrence p_S (equal to its maximum likelihood estimator n_{obs}/n_{tot}), leading to eight levels of probabilities of occurrence.

STEP 4: ESTIMATING THE ASSOCIATED UNCERTAINTY

For each pixel, the deviance to the observed occurrence was computed as (McCullagh & Nelder 1989; p 118):

$$dev = -2[obs \ln(p_S) + (1 - obs) \ln(1 - p_S)] \quad \text{eqn 1}$$

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). 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 'standardized deviance', noted (δ)) is 0 or 1. The standardized deviance was modelled as a Bernoulli process. To account for its possible dependency upon

environmental variables, we modelled the logit of its mean as a polynomial function of the three synthetic climatic variables with polynomial degree ≤ 2 :

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

Twenty-seven models for δ were considered for each species and each subset, corresponding to the combinations of models including or excluding coefficients in eqn (1), with the additional constraints that (i) all models included the intercept $\alpha_{0,S}$; and (ii) second-degree terms $\alpha_{12,S}$ were constrained to co-occur with the corresponding first-degree term $\alpha_{11,S}$. For each species and subset S , each model j was weighted by its Akaike weight $w_{S,j}$ (Burnham & Anderson 2002) so that the projection of the modelled deviance \overline{dev}_S for current data is:

$$\overline{dev}_S = \sum_{j=1}^{24} w_{S,j} dev_{S,j} \quad \text{eqn 3}$$

Species distribution projections were obtained by pooling together model-averaged projections of deviance [eqn (3)] for each of the eight subsets S , so that finally each pixel was assigned a probability of presence and a modelled deviance.

STEP 5: FORECASTS

Each SDM was used to produce forecasts for the period 2081–2100, for the two scenarios A1Fi and B2. Yearly outputs of LPJ and PHENOFIT were averaged over this period to yield their final output; STASH directly outputs a growth efficiency index over the considered period. Model outputs were transformed into presence/absence using the SPT (step 3).

For each species and each climatic scenario, pixels were assigned to one of eight subsets (S) according to the combinations of {STASH, LPJ, PHENOFIT} projected occurrence (as in step 3), and were attributed the species' probability of occurrence p_S while deviance was extrapolated based on eqns (2–3) and on forecasted climatic descriptors.

STEP 6: MODEL EVALUATION AND COMPARISON

In addition to the 'conditional' consensus model presented above, we generated three consensus models for each species, using methods described in Marmion *et al.* (2009). The first two methods (i.e. Mean and Median) assign to each pixel its computed value (mean or median) from the three SDMs outputs. The third one (WA consensus) computes the average of the three SDMs outputs, weighted by their Area Under the Receiver Operating Characteristic curve (AUC, a measure of discriminative power; Swets 1988).

We then compared the accuracy of the projections of all three SDMs and all four consensus models over the historical period for each couple model/species, using various criteria: AUC, the proportions of well-predicted pixels (accuracy), of false positives (commission error) and of false negatives (omission error), once applied the SPT. Note that all these measures rely on discretizing the continuous output of each SDM into binary data. Hence, they only estimate the discriminating power of the model's output (Lobo, Jiménez-Valverde & Real 2008).

All computations from step 2 onwards were conducted using R (R Development Core Team 2011); scripts and data are provided as Supplementary Material.

Results

ACCURACY MEASURES OF THE PROJECTIONS OF SDMS AND THE CONSENSUS MODELS OVER THE HISTORICAL PERIOD

SDM projections for current distributions are shown on Fig. 2. Their AUC values are relatively high (Table 1), considering that the three studied species are widely distributed and occur in a wide range of environments, spanning most of the study area (Lobo, Jiménez-Valverde & Real 2008; Grenouillet *et al.* 2011). Overall, STASH and LPJ show higher AUC and accuracy (i.e. proportion of well-projected pixels) than PHENOFIT.

The conditional consensus model projects high probabilities of occurrence over most of the observed distribution range of all three species (Fig. 3, left column). The conditional consensus shows AUC values comparable or higher than the best SDM and the other consensus models for *F. sylvatica* and *Q. robur*, and outperforms them for *P. sylvestris*, the species worst projected by all SDMs (Table 1).

DIVERGENCE AND UNCERTAINTY IN PROJECTIONS OF CURRENT DISTRIBUTIONS

While the three SDMs capture relatively well the upper and lower boundaries of the distributions at the broad scale, regional discrepancies are noticeable (Fig. 2). When transformed into presence/absence using the SPT, SDMs

projections mismatch for 25%–35% of the pixels. These discrepancies are partly attributable to bioclimatic limits in STASH and LPJ leading to too important contrasts, and to the weak representation of water stress in all models. Furthermore, the coarse spatial resolution of occurrence data, and to a lesser extent of climatic data (particularly in contrasted areas such as mountain ranges), lead to mismatches between projected and observed distributions.

In light of these mismatches, identifying which SDM (s) fail to model the occurrence of which species, and under which conditions, becomes a primary goal. The spatial variation in the deviance of the projected probability of occurrence to observed data yields insight into this question (Fig. 3, middle column). Note that deviance does not depend upon whether models agree with each other, but on whether they agree with observed data. The observed deviance is relatively low, indicating high confidence in the projections of the conditional consensus model. However, it tends to be higher towards the margins of the distributions, where SDMs disagree.

The synthetic environmental descriptors appear to be good predictors of the variation of deviance (Fig. 3, right column), even though some regions of high deviance are not captured by the models, such as the Alps for *F. sylvatica*. In this case, however, all models rightly predict the absence of the species, while the AFE data set, because of its coarse resolution, inaccurately describes the species as present. Deviance to actual occurrence data should therefore be low for *F. sylvatica* in the Alps.

Table 1. Accuracy measures of the projections of the three species' current distribution by STASH, LPJ, PHENOFIT and four consensus models: ours (Conditional) and the Mean, Median and Weighted Average (WA) of Marmion *et al.* (2009). Species Presence Threshold (SPT): threshold maximizing the sum of sensitivity and specificity, above which model outputs were considered to indicate species occurrence. AUC: Area Under the Receiver Operating Characteristic curve. Accuracy: proportion of correctly projected pixels (true absences + true presences). Commission/omission error: proportion of false positives (respectively of false negatives) over the whole simulation window. For each descriptor of accuracy, the best performing model is highlighted in bold face

Species	Model	SPT	AUC	Accuracy (%)	Commission error (%)	Omission error (%)	
<i>Fagus sylvatica</i>	SDMs	STASH	0.27	0.834	74.9	3.8	21.3
		LPJ	0.76	0.872	83.7	9.4	6.8
		PHENOFIT	0.13	0.787	72.4	9.2	18.4
	Consensus	Conditional	0.52	0.876	84.0	9.5	6.5
		Mean	0.35	0.857	80.7	6.9	12.4
		Median	0.27	0.847	80.1	5.8	14.1
		WA	0.34	0.858	80.5	6.6	12.9
<i>Quercus robur</i>	SDMs	STASH	0.08	0.853	82.5	5.2	12.3
		LPJ	0.67	0.830	78.0	11.5	10.5
		PHENOFIT	0.58	0.798	76.0	9.0	15.0
	Consensus	Conditional	0.66	0.854	82.5	5.2	12.3
		Mean	0.39	0.852	81.3	5.2	13.5
		Median	0.33	0.828	80.8	4.2	15.0
		WA	0.37	0.853	81.4	4.7	13.9
<i>Pinus sylvestris</i>	SDMs	STASH	0.30	0.638	75.6	6.9	17.5
		LPJ	0.59	0.687	73.6	2.6	23.9
		PHENOFIT	0.70	0.685	66.8	18.7	14.5
	Consensus	Conditional	0.75	0.744	76.4	7.2	16.4
		Mean	0.54	0.683	70.0	13.1	16.9
		Median	0.31	0.669	71.6	4.8	23.6
		WA	0.55	0.684	69.9	13.3	16.9

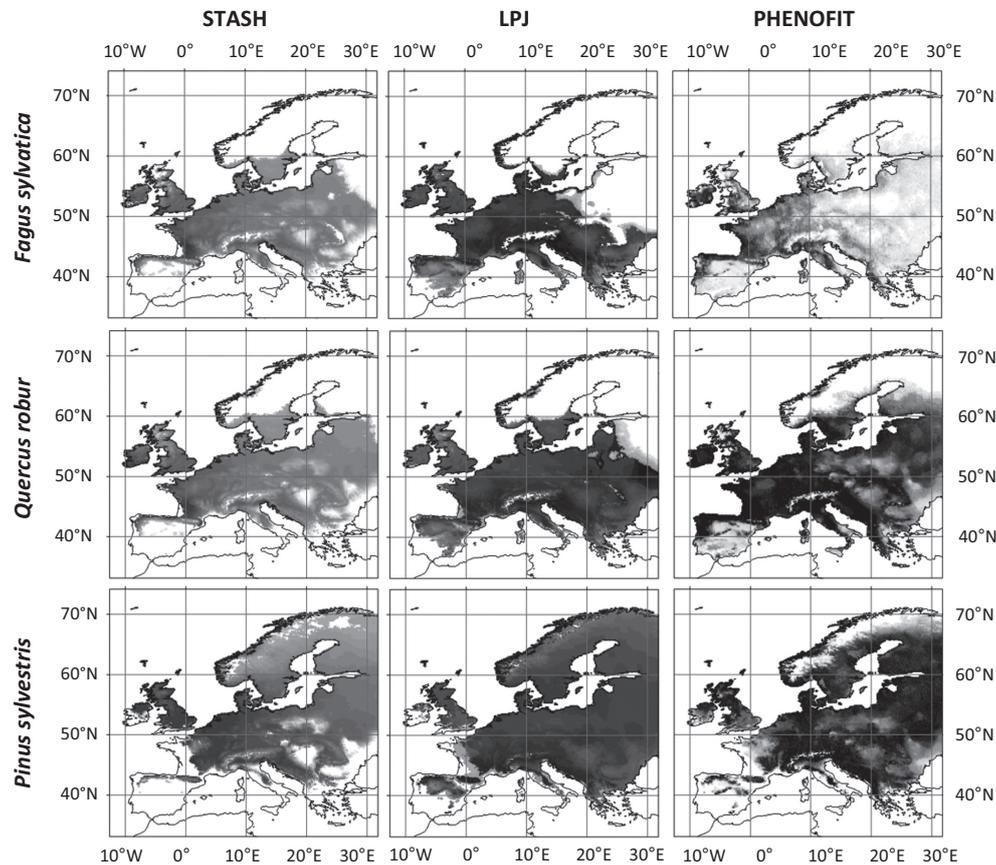


Fig. 2. Projections of the species current distribution (1981-2000) by the individual SDMs (STASH: growth efficiency; LPJ: LAI (leaf area index); PHENOFIT: mean fitness).

FORECASTS AND ASSOCIATED UNCERTAINTY

Under both scenarios, the suitable areas for all three species are projected to shift towards the North-East and towards higher elevations, both by the conditional consensus model (Fig. 4) and the SDMs (Supplementary Material). While the size of the suitable area for *F. sylvatica* and *Q. robur* will remain approximately stable or increase, that of *P. sylvestris* is projected to decrease, albeit less intensely under scenario B2 (Fig. 4). Projected probabilities of occurrence are overall lower under the A1Fi scenario than under the B2 scenario. Large areas of uncertainty appear towards the trailing edges of the distribution for all three species, and over large areas in Central and Eastern Europe for *F. sylvatica*, especially in regions where the projections of the SDMs disagree (Fig. 4; Supplementary Material). These regions show a partial overlap with the least analogous projected climates (not shown).

Discussion

We propose a simple framework to estimate consensual projections of species distributions, while jointly assessing their uncertainty as a function of synthetic environmental descriptors. This framework can be applied to any number of SDMs

(provided each combination of SDM projected presence or absence gathers a large enough number of points), and is particularly well-suited for process-based SDMs, whose errors are expected to be environmentally clustered, and whose projections do not (or not only) rely on observed distributions. Consensus projections generated by this approach are not affected by the introduction of poorly predictive models. We illustrate its use by associating a spatialized quantification of the uncertainty to the forecasts of the future distribution of suitable habitats of three emblematic European forest trees, under two climatic scenarios and making use of the projections of three very different SDMs.

This study only considers uncertainty due to species distribution models. However, uncertainties also arise because the occurrence data used to parameterize the correlative and hybrid models may be inaccurate or too coarse, because processes are calibrated on too narrow an environmental range, because of the choice of the occurrence threshold, and because of uncertainties in the climatic and land use scenarios (Beale & Lennon 2012). Because SDM type has often been found to be the main source of variation between forecasts, as compared with other sources of uncertainty (Dormann *et al.* 2008; Buisson *et al.* 2010; Nenzén & Araújo 2011), we chose to deal first with reconciling model projections and assessing their uncertainty.

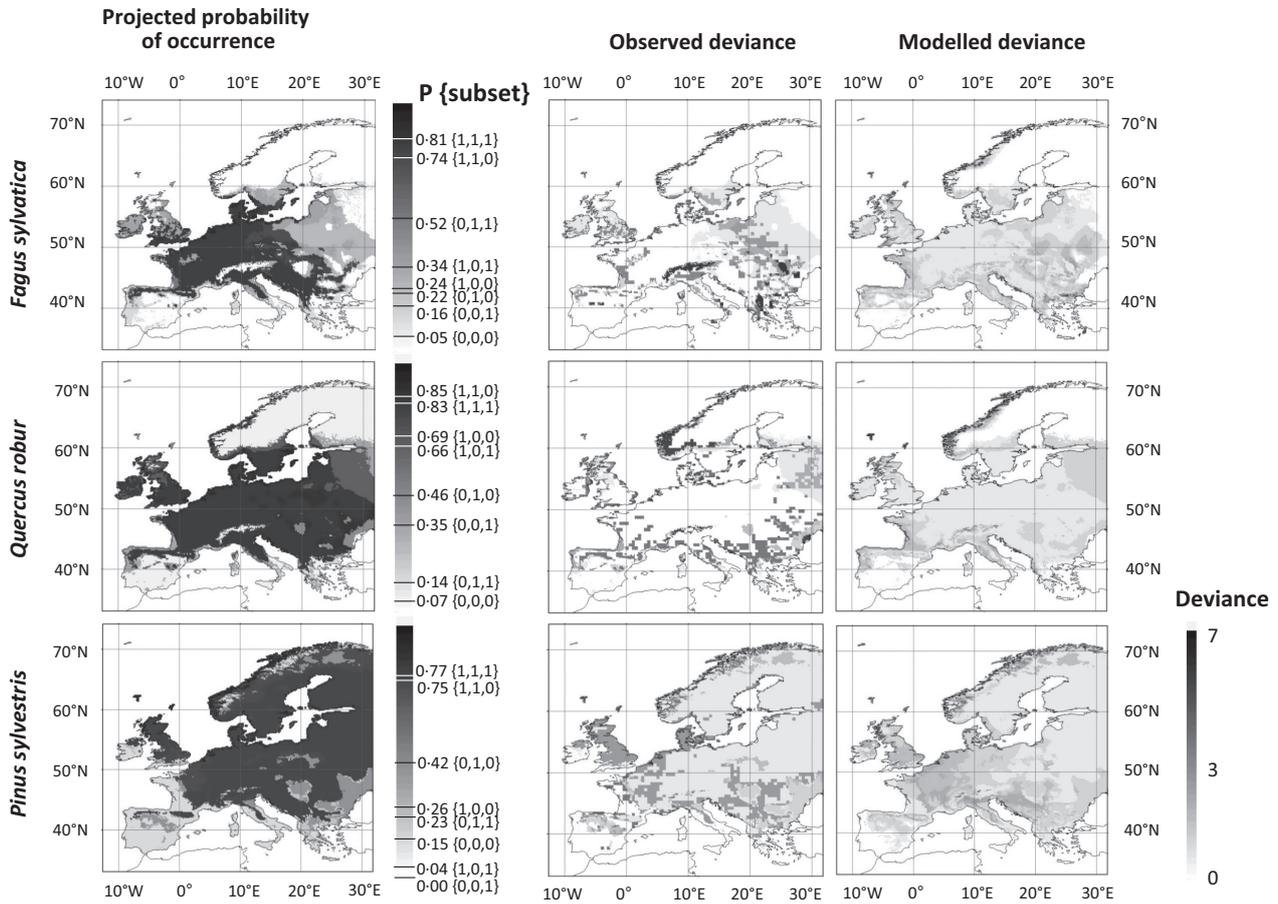


Fig. 3. Projection of the probability of presence of the three species by the conditional consensus (left column); observed deviance (middle column) and modelled deviance (right column) for the period 1981–2000. The consensus model yields $2^3 = 8$ levels of probability of presence, corresponding to the triplets of {STASH, LPJ, PHENOFIT} projected presence or absence.

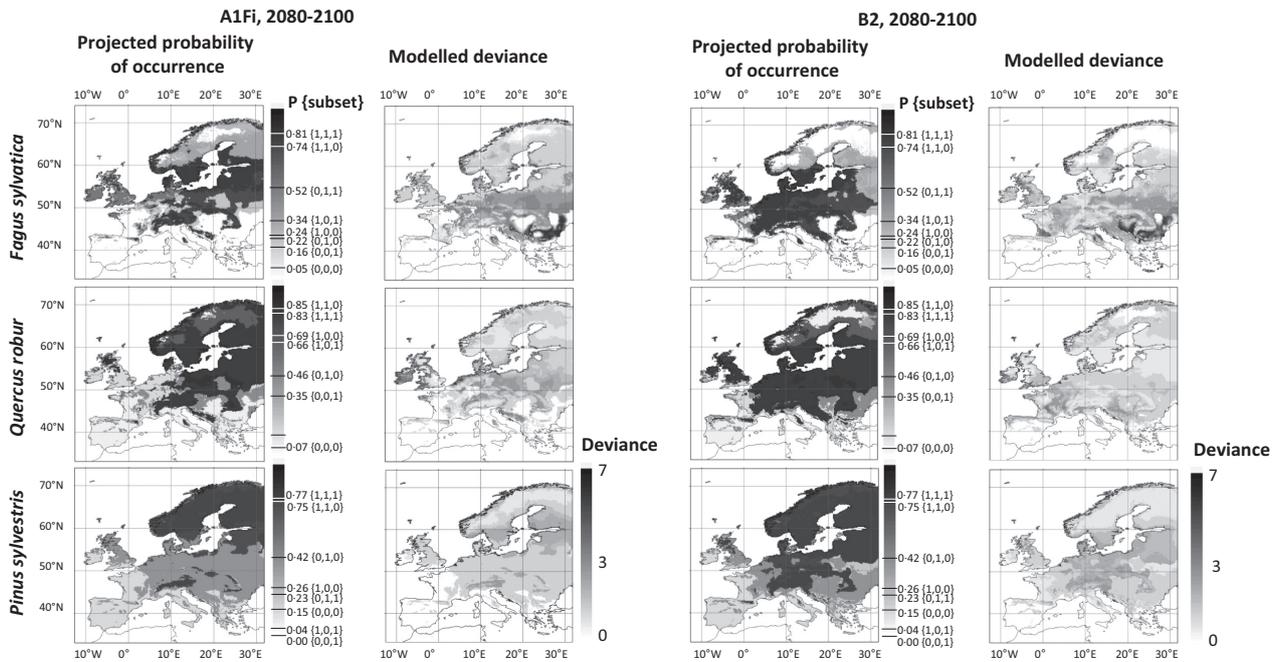


Fig. 4. Projections of the species potential distribution considering unlimited migration for the period 2081–2100 under scenarios A1Fi and B2, as modelled by the conditional consensus model, and associated modelled deviance.

FORECASTING THE DISTRIBUTION OF THREE EUROPEAN TREE SPECIES

In agreement with earlier studies (Kramer *et al.* 2010; Cheaib *et al.* 2012; Meier *et al.* 2012), suitable habitats for the temperate deciduous species *Q. robur* and *F. sylvatica* are projected to shift towards the North-East, and their potential range to increase slightly, while that of *P. sylvestris* is projected to contract. Compared with earlier studies, our results highlight regions whose future suitability is most questionable, notably towards the trailing edge of their distributions, and in large regions of central and Eastern Europe for *F. sylvatica*.

High uncertainty does not necessarily reflect SDMs disagreement. For example, as compared with the AFE occurrence data, all three SDMs wrongly project the current presence of *F. sylvatica* just south of the Alps and close to the French Atlantic coast (Fig. 2 and Fig. S1). Models of deviance based on three composite environmental descriptors capture this discrepancy (Fig. 3). Thus, in regions presenting future climatic condition analogous to the conditions currently observed along the French Atlantic coast or south of the Alps (e.g. east of the Baltic Sea under the A1Fi scenario for year 2100), where all three models project a future occurrence of *F. sylvatica*, the conditional consensus model associates large uncertainties to this species' projected occurrence. This error pattern is commonly observed (Hanspach *et al.* 2011) and can be attributed to the coarse resolution of both occurrence data (Rocchini *et al.* 2011) and climatic data, stressing the need for accurate occurrence and climatic data (Austin & Van Niel 2011). Uncertainties uncovered for future distributions do not only reflect discrepancies between SDMs' projections, but also weaknesses common to all SDMs included in the consensus approach.

Current species distributions might not fill their potential range (Svenning & Skov 2007; Dormann *et al.* 2012), and because none of the three models (at least as they were used here) accounts for dispersal, land use nor interspecific competition, the maps shown here only indicate potentially suitable habitats. In this regard, our projections are arguably too optimistic: whether the species can migrate towards newly available habitats, or establish there, is highly uncertain. When dispersal limitation and land use are taken into account, effectively accessible suitable sites are much scarcer than potentially suitable sites (Meier *et al.* 2012). However, other factors usually not taken into account by correlative SDMs, such as local adaptation or phenotypic plasticity, may help species cope with climate change. Process-based and hybrid models provide less alarmist forecast of species range shift than correlative SDMs, which can thus be argued to be overly pessimistic (Morin & Thuiller 2009; Cheaib *et al.* 2012). Overall, actual shifts in distributions are likely to lie between our projections and those of Meier *et al.* (2012). This stresses the fact that providing consensual projections of species range changes should not prevent from reducing individual SDMs errors. Incorporation of realistic dispersal models is a step towards such improvements. Other advances could be gained from including the representation of other processes, such as biotic interactions (e.g. Davis

et al. 1998) increases in atmospheric CO₂ (as in LPJ) or a finer representation of local adaptation than what is currently implemented in PHENOFIT.

MORE THAN A DEMOCRATIC VOTE

Improving the reliability of SDMs, ultimately aiming at developing reliable, integrated or hybrid models (Morin & Lechowicz 2008; Thuiller *et al.* 2008) requires much detailed information on the studied species. A more tractable approach is to establish consensus projections from already existing models, and take advantage of the strength of each model. More or less refined ways have been proposed to combine the outputs of individual SDMs to improve their reliability, from the single vote to weighted averages of each SDM output (Marmion *et al.* 2009). To our knowledge, such consensus methods have only been conducted on correlative distribution models. Here, we have combined the outputs of three conceptually different SDMs: a correlative model with physiological grounds, a hybrid model and a process-based model. Because these SDMs do not output probabilities of occurrence, the classical consensus methods were not expected to accurately describe the patterns of species occurrence. Indeed, classical consensus methods are sensitive to the addition of non-predictive models, or of models consistently producing lower- or over-than-average scores. Therefore, the Mean, Median and WA consensus models performed worse than the conditional consensus model for the poorly projected species (*P. sylvestris*).

While projections of the three SDMs mostly differ at the regional scale for the historical period, their forecasts strongly differ at the continental scale for 2081–2100, regardless of the climatic scenario. This calls not only for a consensus modelling of the probability of occurrence but most importantly for quantifying uncertainty. So far, uncertainty has been considered as the variance between the projections of SDMs, all of the same family (e.g. linear models, Hartley, Harris & Lester 2006). However, we argue that uncertainty should be seen as the spatially explicit deviance of predictions to the observations. Indeed, models from a given family, or models using a certain set of processes or of environmental variables, are likely to produce (possibly cryptically) environmentally clustered errors, which can only be detected through comparison with observed occurrences, and neither through inter-model variance, nor through global quality estimators such as AUC.

When SDMs are based on biological processes, uncertainty is likely to vary with abiotic variables, and to be highest in regions where a weakly modelled process is crucial to explain the species' occurrence. For example, the version of PHENOFIT presented here uses a simplistic representation of water stress, and is thus expected to yield poor predictions in dry areas. If all SDMs were given a constant weight in a consensus model (i.e. a constant confidence, for example proportional to their AUC; Marmion *et al.* 2009), this weight for PHENOFIT would strongly rely upon the proportion of dry pixels over the simulation window. In contrast, in our conditional consensus model, the contribution of PHENOFIT to the consensual

projection of the probability of occurrence depends on the output of the other two models and on the realized distribution, while confidence into its projections relies upon climatic variables.

LIMITATIONS OF THE CONDITIONAL CONSENSUS MODEL

A useful perspective would be to make use of the continuous outputs of the individual SDMs to generate the consensus model: for instance, instead of coercing the individual SDM outputs into binary occurrences, one could consider more subsets, through incorporating classes of low, medium and high outputs for each model. This could help account for gradual response curves of the species to abiotic variables (Meynard & Kaplan 2012). We chose to use only two classes of outputs for each model, to avoid over-fitting: obviously, such an approach is limited by the amount of available data as the more subsets are used, the more likely it is that one of the subsets corresponds to fewer pixels than necessary to statistically assess the link between model deviance and environmental variables.

The modelling of uncertainty (step 4) might be improved by incorporating non-climatic variables, such as soil quality; and by taking spatial autocorrelation into account so as to reduce uncertainty in parameter estimation. Various methods have been developed (Dormann *et al.* 2007; Beale *et al.* 2010) to account for spatial autocorrelation in data; however, the most efficient ones (Moran eigenvector approaches, generalised linear model with explicit spatial covariance) are still too computer-intensive to be tractable on large numbers of points. Because we have not accounted for the effect of spatial autocorrelation in step 4, we expect (i) unbiased estimates of model coefficients but; (ii) larger standard errors of these estimates around the expected value (McGill 2012); and (iii) a slight tendency for model comparison procedures to favour over-parameterized models (as both model uncertainty and the variables used for regression are spatially autocorrelated).

The main strength of the conditional consensus model is to characterize uncertainty in a spatially explicit, environment-dependent way. However, this approach can only be used with models that are thought to be extrapolable to future conditions. We chose not to include purely correlative models in our approach, despite their high accuracy in projecting historical distributions. First, their errors are unlikely to be environmentally clustered, as all available potentially explanatory layers would already have been included in designing the model. Secondly, finding the right amount of model complexity, leading both to accurate projections of current conditions and to extrapolable relationships between climate and occurrence is a perilous task. The consensus approach presented here is sensitive to the addition of overfitted models – which would drive the outputs of the consensus model, to the detriments of locally less accurate, but maybe more robust models –, but helps make rid of *consistently inaccurate* models.

SDMs relying on physiological processes may be more plausibly extrapolated to non-analogous conditions because, even though the reaction norms of trees' physiological traits to climate may evolve within a few generations; they are likely to be

conserved in the next few decades (Dormann *et al.* 2012). However, despite being realized for a wide range of climatic conditions, parameterization of process-based SDM may also be inaccurate under the novel combinations of environmental conditions expected for the coming century (Williams, Jackson & Kutzbach 2007). For example, phenological models in PHENOFIT are parameterized using empirical relationships between time-series observations of phenological events in natural populations and daily temperatures and photoperiod. As these two factors are broadly correlated in nature, their relative importance is difficult to capture: different parameter sets may describe current phenology with equal likelihood; yet their projections to future conditions may vary. This is why this kind of models constantly need refining through the incorporation of experimental data (Caffarra, Donnelly & Chuine 2011; Basler & Körner 2012). The increase in atmospheric CO₂ also generates non-analogue conditions, and its impact on vegetation dynamics, functioning and distribution remains under debate, calling for ecophysiological experiments (Körner 2000; Prentice & Harrison 2009).

Acknowledgements

The authors thank the EMBERS group of Lund University for providing STASH and LPJ models, the ATEAM project for providing climatic datasets, and the GDR CNRS 2968 Observatoire Des Saisons (www.gdr2968.cnrs.fr), the ONF-RENECOFOR Network, the ONF-Seed Service Sécherie de la Joux and the French Public Arboreta Network for providing the phenological observations used to parameterize PHENOFIT. We thank Rémi Choquet and Olivier Gimenez for advice regarding statistical analyses, and Fabien Laroche, Christine Meynard, Xavier Morin, Pedro R. Peres-Neto, A. Townsend Peterson and four anonymous referees for critical reading of the manuscript. We acknowledge financial support from the Agence Nationale de la Recherche (projects ANR-05-BDIV-009 QDiv; ANR-2009-PEXT-001105 SCION). AD was funded by a Marie Curie IOF grant (TRECC-2009-237228) within European Commission's FP7.

References

- Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.
- Austin, M.P. & Van Niel, K.P. (2011) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1–8.
- Basler, D. & Körner, C. (2012) Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology*, **165**, 73–81.
- Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 247–258.
- Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. & Elston, D.A. (2010) Regression analysis of spatial data. *Ecology Letters*, **13**, 246–264.
- Buckley, L.B. (2008) Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. *The American Naturalist*, **171**, E1–E19.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.
- Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York (NY).
- Caffarra, A., Donnelly, A. & Chuine, I. (2011) Modelling the timing of *Betula pubescens* budburst. II. Integrating complex effects of photoperiod into process-based models. *Climate Research*, **46**, 159–170.
- Chebib, A., Badeau, V., Boe, J., Chuine, I., Delire, C., Dufrene, E., François, C., Gritti, E.S., Legay, M., Pagé, C., Thuiller, W., Viovy, N. & Leadley, P. (2012) Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecology Letters*, **15**, 533–544.

- Chuine, I. & Beaubien, E.G. (2001) Phenology is a major determinant of tree species range. *Ecology Letters*, **4**, 500–510.
- Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783–785.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Dormann, C.F., Purschke, O., García Márquez, J.R., Lautenbach, S. & Schröder, B. (2008) Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. *Ecology*, **89**, 3371–3386.
- Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012) Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.
- Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.
- Gibson, L.A., Wilson, B.A., Cahill, D.M. & Hill, J. (2004) Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology*, **41**, 213–223.
- Grenouillet, G., Buisson, L., Casajus, N. & Lek, S. (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, **34**, 9–17.
- 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.
- Hanspach, J., Kühn, I., Schweiger, O., Pompe, S. & Klotz, S. (2011) Geographical patterns in prediction errors of species distribution models. *Global Ecology and Biogeography*, **20**, 779–788.
- Hartley, S., Harris, R. & Lester, P.J. (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, **9**, 1068–1079.
- Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.
- Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.
- Körner, C. (2000) Biosphere responses to CO₂ enrichment. *Ecological Applications*, **10**, 1590–1619.
- Kramer, K., Degen, B., Buschbom, J., Hickler, T., Thuiller, W., Sykes, M.T. & de Winter, W. (2010) Modelling exploration of the future of European beech (*Fagus sylvatica* L.) under climate change—range, abundance, genetic diversity and adaptive response. *Forest Ecology and Management*, **259**, 2213–2222.
- Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*. Chapman & Hall, London, UK.
- McGill, B. (2012) Why OLS estimator is an unbiased estimator for GLS. *Dynamic Ecology*. <http://dynamicecology.wordpress.com/2012/09/24/why-ols-estimator-is-an-unbiased-estimator-for-gls/> [accessed 24 September 2012].
- Meier, E.S., Lischke, H., Schmatz, D.R. & Zimmermann, N.E. (2012) Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecology and Biogeography*, **21**, 164–178.
- Meynard, C.N. & Kaplan, D.M. (2012) The effect of a gradual response to the environment on species distribution modeling performance. *Ecography*, **35**, 499–509.
- Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M. & New, M. (2004) *A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: The observed record (1901–2000) and 16 scenarios (2001–2100)*. Tyndall Centre for Climate Change Research, Working Paper 55, Norwich, UK.
- Morin, X. & Lechowicz, M.J. (2008) Contemporary perspectives on the niche that can improve models of species range shifts under climate change. *Biology Letters*, **4**, 573–576.
- 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.
- Nenzén, H.K. & Araújo, M.B. (2011) Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, **222**, 3346–3354.
- Nicks, A.D., Lane, L.J. & Gander, G.A. (1995) Weather generator. *USDA–Water Erosion Prediction Project: Hillslope profile and watershed model documentation* (eds D.C. Flanagan & M.A. Nearing), pp. 2.1–2.22. USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, USA.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martínez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton, NJ, USA.
- Prentice, I.C. & Harrison, S.P. (2009) Ecosystem effects of CO₂ concentration: evidence from past climates. *Climates of the Past*, **5**, 297–307.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raupach, M.R., Marland, G., Ciais, P., Le Quéré, C., Canadell, J.G., Klepper, G. & Field, C.B. (2007) Global and regional drivers of accelerating CO₂ emissions. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10288–10293.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.
- 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.
- Svenning, J.-C. & Skov, F. (2007) Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, **10**, 453–460.
- Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- 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.
- TEEB (2010) *The Economics of Ecosystems and Biodiversity: Ecological and Economic Foundations*. Earthscan Ltd, London and Washington.
- Thuiller, W., Albert, C., Araujo, M., Berry, P., Cabeza, M., Guisan, A., Hickler, T., Midgley, G., Paterson, J. & Schurr, F. (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.
- Tutin, T.G., Heywood, V.H., Burges, N.A. & Moore, D.M. (1964–85) *Flora Europaea*. Cambridge University Press, Cambridge.
- Walther, G.R., Berger, S. & Sykes, M.T. (2005) An ecological 'footprint' of climate change. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 1427–1432.
- Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5738–5742.

Received 29 October 2012; accepted 14 January 2013

Handling Editor: Pedro Peres-Neto

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Two R scripts and data are provided to reproduce the results of the paper.

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

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.

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*.

Figure S4. Correlation circles of the principal components analysis for current climates, 2100 A1Fi scenario, 2100 B2 scenario and all climates together.

Figure S5. Coordinates of current (left column) and future (middle and right columns) climates in the principal components 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.

Figure S6. ROC plot for the projection of *Pinus sylvestris* by LPJ.

Figure S7. Data subsets obtained for *Pinus sylvestris*.

Figure S8. Projected probability of occurrence for *Pinus sylvestris*.

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

Figure S10. Observed occurrence (top left), modelled occurrence (top right), observed (bottom left) and modelled 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)`).

Figure S11. SDM outputs for the A1Fi scenario (period 2080–2100).

Figure S12. SDM outputs for the B2 scenario (period 2080–2100).

Figure S13. Projected occurrences and projected deviances for all climatic datasets. This image is generated by function `ProjectedPoccAndDevPic`.