# Extinction debt and colonization credit delay range shifts of eastern North American trees

Matthew V. Talluto[1]* , Isabelle Boulangeat[2], Steve Vissault[3], Wilfried Thuiller[1] and Dominique Gravel[3]

Q29

**Global climate change is already having an impact on species ranges. For species with slow demography and limited dispersal, such as trees, lags between climate change and distribution shifts are likely to increase in the future. Such lags can be of critical importance to forest and biodiversity management and biodiversity, because they can result in 'extinction debts', where populations temporarily persist under unsuitable conditions, and 'colonization credits', where suitable locations are not occupied owing to slow demography and limited dispersal. Here we use a range dynamics model based on metapopulation theory to show that the distributions of 21 dominant trees in eastern North America are out of equilibrium with climate and demonstrate both extinction debt and colonization credit. Moreover, lags are more severe at northern range limits, suggesting that range contraction in response to warming temperatures will outpace expansion.**

Q3

Q4

Because climate is as a major driver of species distributions, global warming is expected to induce species range shifts[1]. The close relationship between climate and species distribution is the basis of correlative species distribution models[2,3], which have been extensively used to derive biodiversity scenarios[4–7]. However, species do not respond to change immediately, and consequently distributions can lag behind climate change[8]. This is particularly the case for long-lived and sessile organisms. Demographically 'slow' species such as trees are vulnerable to such lags, leading to so-called extinction debts, where populations persist for some time after unsuitable conditions render eventual extinction likely[9–11], and corresponding colonization credits, where suitable but unoccupied locations are only slowly filled owing to slow demography and limited dispersal. Such mismatches between optimal and realized climates have already been observed[12] and are likely to be exacerbated in the future, producing lags between climate change and shifts in species distributions that may extend well beyond the timescales typical of ecological studies. It is critical to predict these situations for forest management and biodiversity conservation, which rely on understanding which species can persist at a given location, and which are often interested in short-time predictions (that is, when lags are most likely to occur).

It is not straightforward to identify disequilibrium between climate and species distributions using correlative species distribution models (SDMs). Instead, dynamic models provide the advantage of linking local processes with species ranges. In addition, these models need to be able to explain species distributions at the scale of their entire range, and capture the variation in key processes according to climate variables. Therefore, although process-based models (for example, PPA[13], JABOWA[14], LANDIS II[15], FATE-HD[16]) can easily capture local dynamics, they are inherently data demanding and computationally intensive[17,18], making the estimation of entire distributions at continental scales challenging. Here we propose a novel approach inspired by metapopulation theory[19]. Our approach stochastically models the dynamics of presence and absence at a locality as a function of two fundamental ecological processes: colonization of empty localities and extinction from occupied localities. Although theoretical studies have shown how colonization–extinction dynamics can lead to range limits[20,21], this is, to our knowledge, the first time the theory is applied to explain species distributions.

We fit this model to a database of repeated observations of more than 90,000 forest inventory plots spanning the eastern United States and Canada (Supplementary Fig. 1). This approach is ideal to represent forest dynamics owing to the strong imprint that succession and natural disturbance leave on forest landscapes (Fig. 1). We analysed range dynamics to look for lags in the response to climate change in 21 dominant trees across temperate, transitional and boreal forests (Supplementary Table 1) and to evaluate future rates of response to climate change at the northern and southern range limits. This transition is particularly relevant because of the major shifts in ecosystem processes that are expected to follow colonization of boreal forests by deciduous and temperate species. Our principal hypotheses were that range limits are set by a balance between colonization and extinction dynamics and that the slow nature of these demographic processes results in lags in the response to climate change. Consequently, even though local conditions could allow populations to persist temporarily in a given location[22], we expect that range limits of forest tree species are ultimately determined by regional processes such as dispersal limitation, succession and natural disturbance.

## Results

The metapopulation theory of range dynamics predicts that range limits will occur where colonization is balanced by extinction. Although originally developed to represent a set of populations connected by dispersal[19], it was later applied to continuous landscapes, even representing the dynamics of individuals in a spatially structured population[23]. We computed the probability of colonization ($c(E)$) and extinction ($m(E)$) as functions of climate $E$ for all species, then solved these for the environmental conditions $E^*$ such that $c(E^*) = m(E^*)$ (Fig. 1 and Supplementary Figs 2–4). This yields the range limits at equilibrium, allowing for comparisons between contemporary distributions and long-term equilibria. We found realistic ranges for all of the species (Fig. 2 and Supplementary Fig. 5) and surprisingly low uncertainty in the predicted range limits

[1]Université Grenoble Alpes, CNRS, UMR 5553, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France. [2]Department of Biosciences, Ecoinformatics and Biodiversity, Aarhus University, Aarhus Denmark. [3]Département de biologie, Université de Sherbrooke, Sherbrooke, Quebec, Canada. *e-mail: matthew.talluto@univ-grenoble-alpes.fr
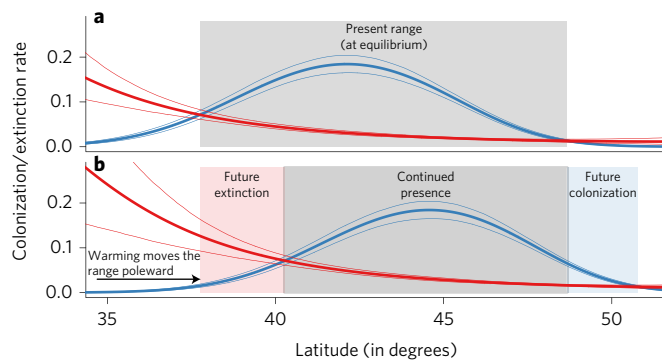
**Figure 1 | Schematic demonstrating how metapopulation dynamics can lead to range limits.** A metapopulation approach to range dynamics hypothesizes that range limits are determined by colonization (*c*, blue lines) and extinction (*m*, red lines) probabilities as they vary along an environmental gradient[9]. **a**, *c* and *m* for *Acer saccharum* vary with temperature, resulting in a latitudinal gradient. More generally, we describe *c* and *m* as functions of the environment: $c(E)$ and $m(E)$. The theory predicts that the species will be present at equilibrium (shaded region) where $c(E) > m(E)$; therefore the temperature at which $c(E) = m(E)$ determines the latitude of the range limits. **b**, A simulated 2.5 °C of warming produces a northward shift in the suitable range (grey and blue shading). However, over short time scales, slow extinction probabilities will cause the species to persist where $m(E) > c(E)$ (red shading) until all occupied populations have gone extinct. A similar lag in colonization ability will result in the species being absent in many areas where $c(E) > m(E)$ (blue shading). Thin lines show 90% credible intervals around mean *c* and *m* probabilities.

(Supplementary Fig. 6), despite considerable variability in the colonization and the extinction functions (Supplementary Figs 2,3). However, compared with prior expectations about where the species occur (for example, from published range maps), we observed some systematic discrepancies at the borders of most species' ranges. We therefore compared the difference between the predicted equilibrium ranges from our fitted metapopulation model with current distributions interpolated from the forest inventory data using the random forest algorithm. Because this algorithm is extremely flexible and was parameterized with a large dataset, these maps presented a highly accurate picture of the current distributions (Supplementary Fig. 7), in contrast to the long-term climatic equilibrium predicted by the metapopulation model. For all species, we found extensive areas of extinction debt, where the current ranges indicate that the species is present while the long-term expectation is absence. These areas were generally found at the southern edge of current ranges. All species also had a corresponding colonization credit, where the current maps indicate absence but the long-term expectation is presence (Fig. 2). Colonization credit was usually found at the north and west of the current range. This strong latitudinal pattern in extinction debt and colonization credit suggests that the disequilibrium in species ranges reflects the signal of contemporary climate change, where climatic optima have already shifted northward of where the species are presently located.

In order to better understand the transient conditions leading from the present ranges of species to their equilibrium distributions, we further analysed our model with respect to the rate of change in response to climate change at the range boundaries. The simplicity of the theory facilitates this analysis, since the intrinsic growth rate of a metapopulation, defined as $\lambda(E) = c(E) - m(E)$ within the range and $\lambda(E) = m(E) - c(E)$ outside the range, approximates the rate of response to a slight change in the environmental conditions. Thus, the responsiveness of the system, given by the partial derivative $\partial\lambda/\partial E$ at the range boundary, indicates how quickly the system will track new equilibrium conditions in response to climate warming.

For species that had both range boundaries within the study area, responsiveness was much greater at southern range boundaries than in the north (grand mean difference across species = 0.14; 95% confidence interval = 0.046–0.24; $n = 18$), although there was substantial interspecies variation (Fig. 3). Therefore, future increases in temperature will probably result in faster responses at southern range limits, where northward contraction is expected, and slower responses in the north, where range expansion is expected (Fig. 2). We further expect that temperate species will respond more quickly than boreal species due to the overall trend of increased responsiveness at southern latitudes.

Ecological theory predicts that the climate dependence of colonization–extinction dynamics should be related to individual-scale demographic variation[24,25]. It follows that regional-scale extinction debts are a result of increased mortality or a failure of recruitment at the individual scale. To test this hypothesis, we extracted information on individual trees from our database and analysed mortality and recruitment rates. ~~Average mortality grouping across all species within their equilibrium ranges was significantly lower than in areas facing extinction debt,~~ with a mean rate of 0.044 individual$^{-1}$ year$^{-1}$ (0.038–0.050) within the range and 0.063 individual$^{-1}$ year$^{-1}$ (0.051–0.078) in areas facing extinction debt (values are mean predictions from a hierarchical model with 90% credible intervals). Recruitment rates did not differ, with rates of 0.038 individual$^{-1}$ year$^{-1}$ (0.034–0.043) in the equilibrium range and 0.045 (0.037–0.055) in areas facing extinction debt. Mortality estimates varied considerably among species, but were ~~significantly~~ higher in the regions facing extinction debt in 16 of the 21 species (Fig. 4), with boreal, transitional and temperate species all showing the same trend. Therefore, elevated mortality rates at the range margins, rather than recruitment failure, are likely a principal cause of range limits and may contribute to the loss of populations in regions facing extinction debt.

## Discussion

Our model predicts that the distributions of many of the most abundant trees occurring in the boreal–temperate ecotone in eastern North America are out of equilibrium with climate at the range margins (Fig. 2). Similar results have been observed in other systems, suggesting that the distributions of many species[8,12,26] (and therefore the ecosystems in which they are present) may be in disequilibrium with climate. Models failing to account for this bias could therefore produce overly optimistic estimates of the tolerance of many species to climate change. In particular, these results call for caution when forecasting and planning using SDMs. When species distributions are out of equilibrium with climate, SDM estimates of species' optimal climates and thus potential future ranges are likely to be biased, because the species may be inhabiting unsuitable regions and may not yet have colonized all suitable regions. The strength of our modelling approach rests on the strong theoretical foundation, which predicts that range limits occur where colonization and extinction rates are balanced[20]. We have shown for the first time and with high precision that this theory is well-suited to explain the distribution of forest tree species, and that we can predict the static distribution of a species using only dynamic data. This offers us the unique opportunity to identify current extinction debts and colonization credits.

A further benefit to using a simple and tractable theory to understand range dynamics is having a comprehensive understanding of transient states, such as the responsiveness of range limits to climate change. We have shown that many dominant trees in eastern North American forests have low capacities to track climate change (Fig. 3). In the future, most species are expected to experience range losses at the southern end of the distribution and gains at the northern end even with no warming beyond the present climate. Given the expectation that northern regions of our study area will warm by as much
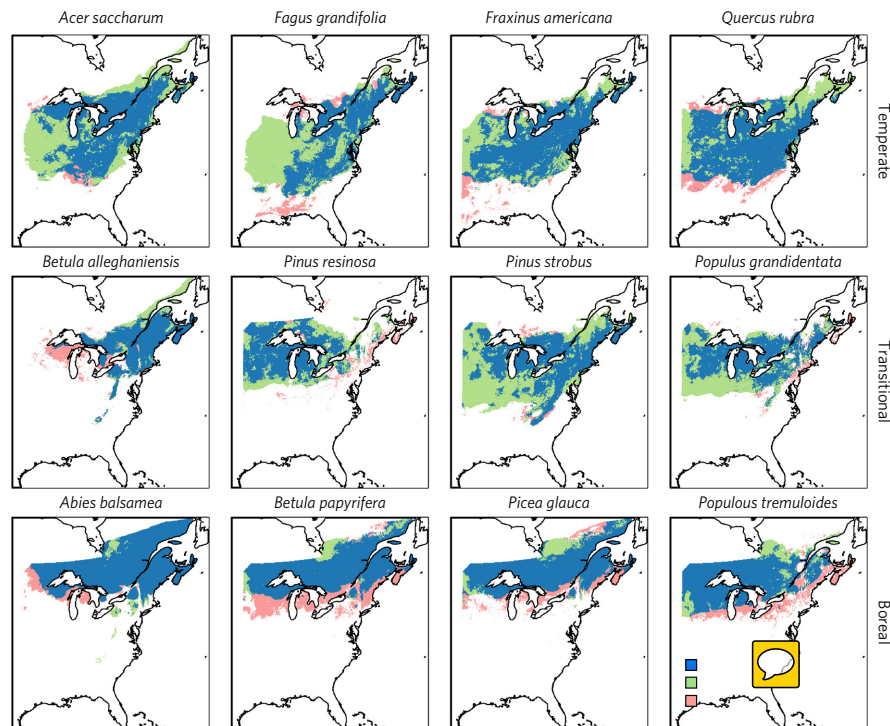
**Figure 2 | Maps of projected species ranges with extinction debt and colonization credit for selected species.** The maps indicate, for the present climatic conditions, where the species is (1) currently present and predicted to be present at equilibrium (continued presence; blue), (2) currently absent but predicted to be present at equilibrium (colonization credit; green), and (3) currently present but predicted to be absent at equilibrium (extinction debt; pink). Equilibrium presence was assessed using the equilibrium solution of the state and transition model (that is, $c(E) > m(E)$), and current presence was assessed using the random forest model with a threshold (computed using a semi-independent evaluation dataset) applied to the probabilities.

as 5 °C by 2100 (ref. [27]), changes to range limits (and corresponding changes to community composition) will almost certainly be larger than those we have predicted here. However, like in other studies[11,28–31], we find that such changes will happen slowly, on the one hand because the high longevity of trees creates extinction debt in warmer regions, on the other hand because slow colonization constrains range expansion in the north. Moreover, our results should be interpreted with caution with respect to boreal ecosystems, as our database does not include the boreal–tundra transition. Thus, the large losses we predict may be offset by boreal encroachment into tundra. Such shifts may be complex and difficult to predict with respect to warming temperature[32]. Overall, temperature increases are likely to continue to outpace distributional changes, resulting in increasing ecological tension between extant populations and the climate required for sustainable populations. Effective management strategies and climatic adaptation would have to account for such increasing disequilibrium in the future.

Similar to SDMs, the choice of a limited set of key variables is still a challenge to model some species. It is likely that additional variables can influence the colonization–extinction probabilities. If these variables are spatially structured, failing to account for them can result in a spatial bias in the colonization–extinction functions and incorrect projections of species ranges. An analysis of the residuals in our models (Supplemental Methods and Supplemental Fig. 9) suggested that spatially structured variables may indeed be missing for certain species. However, a limited spatially explicit version of the model showed no evidence of bias (Supplemental Figs 10–12). It is clear, however, that for some species a complete picture of what variables influence range dynamics is lacking, and thus dynamic models must be interpreted with caution. Even when the models accurately capture species dynamics in the calibration range, important variables and interactions that are missing can make prediction and analysis of transient states unreliable.

One of the main strengths of our approach is to shift from modelling species ranges and their dynamics from a static, niche-based perspective to a dynamic perspective. However, the model is unable to distinguish fine-scale mechanisms underlying the relationships between climate and colonization and extinction probabilities. For example, interactions between climate and local age-specific demographic rates could result in nonlinearities in the relationships between climate and colonization and extinction that would not be detectable using our method (for example, through storage effects[33] or mass effects[28]). To properly account for such mechanisms, our dynamic approach to modelling species ranges could be extended using more detailed models[34–36] when their parameterization is feasible. More generally, the perspective of modelling species and community changes instead of their distributions offers numerous opportunities to improve the predictive power of biodiversity models.

## Methods

**Model description.** Our model design is a simple mathematical model following classical metapopulation theory[19], extended to incorporate environmental gradients[20]. A metapopulation consists of a set of local populations connected together by dispersal. Patches are either occupied (that is, the species is present) or unoccupied. We used a discrete-time variant of the Holt model[20]. In discrete time, the patch occupancy dynamics of the metapopulation can be conceptualized as a spatially inhomogeneous Markov chain with state vector $\boldsymbol{P}_t$ (with $\boldsymbol{P}_{t,0}$ unoccupied patches and $\boldsymbol{P}_{t,1}$ occupied patches at time $t$) and transition matrix $S$, such that:

$$\boldsymbol{P}_{t+1} = \boldsymbol{P}_t S$$

The model is made inhomogeneous by parameterizing the transition matrix with colonization and extinction probabilities that are estimated as functions of the environment:

$$S = \begin{bmatrix} 1 - p(E)\,c(E) & p(E)\,c(E) \\ m(E) & 1 - m(E) \end{bmatrix} \quad (1)$$
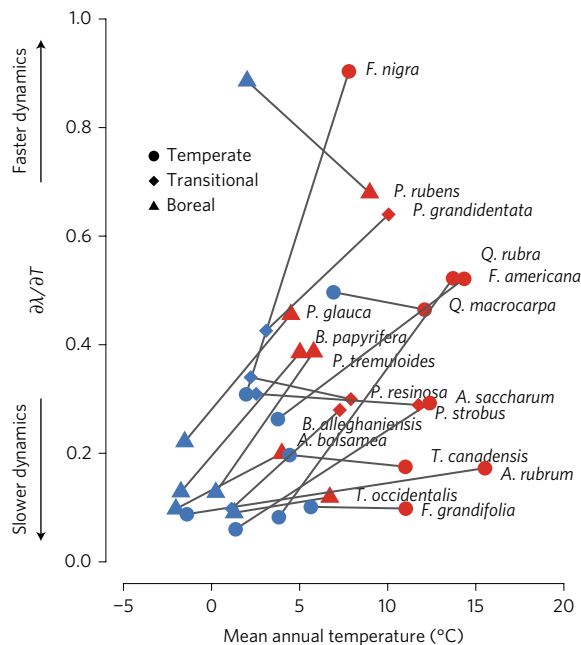
**Figure 3 | The responsiveness of species ranges to changes in temperature.** We evaluated responsiveness to temperature ($\partial\lambda/\partial T$) at the northern (blue symbols) and southern (red symbols) range limits for each species. Larger $y$ values (generally found at the southern range limits) indicate faster dynamics (that is, a more rapid return to equilibrium following perturbation). Black lines connect northern and southern points for each species. Three boreal species (*Picea mariana, Larix laricina* and *Pinus banskaia*) did not have northern limits in the study area and are excluded.

where $E$ represents environmental conditions, $t$ represents time, $p(E)$ is the fraction of the landscape where the species is present, $c(E)$ is the colonization probabiltiy and $m(E)$ is the extinction probability. For example, for a colonization to occur in a patch with environmental conditions $E$, the patch must be initially empty (with probability $1-p(E)$), it must have a source of available seeds (with probability $p(E)$ representing the prevalence of the species in the immediate neighbourhood of the focal patch) and seeds must arrive and establish in the patch (with probability ($c(E)$)). At equilibrium, the species persists where $c(E) > m(E)$; thus the range limit is found at environmental conditions $E^*$, where $c(E^*) = m(E^*)$ (Fig. 1).

**Data preparation.** We estimated the climate-dependence of colonization–extinction probabilities of dominant tree species found in a region encompassing 27 degrees of latitude in eastern North America and spanning the transition from temperate to boreal forest. To perform these analyses, we assembled a database from regional forest inventory databases from the Forest Inventory and Analysis National Program in The United States[37], Domtar (a private forestry company in Quebec), the Ministry of Forests, Wildlife and Parks in Quebec[38], the Ministry of Natural Resources and Forestry in Ontario[39], and the Ministry of Natural Resources in New Brunswick[40]. Although protocols and data storage differed among databases, we were able to build a combined database that included common measurements from each database, including plot locations and sizes, years of measurement, and the diameter at breast height (DBH), species and status (living or dead) of every tree in each plot with a DBH > 127 mm. Smaller trees were not recorded consistently across databases and sampling years, and were therefore excluded from our analysis to minimize geographic biases due to the differing protocols. Initial investigations using models containing smaller trees showed no detectable differences in colonization–extinction curves, therefore we chose to filter these trees to maintain consistency among data sources. To account for potential bias (because the DBH threshold may represent different ages for different species, influencing the probability of colonization), we used ages derived from tree cores to (available for 16 out of 21 species) to estimate age–DBH curves (Supplementary Fig. 8). We then used linear models to estimate the mean age at which each species reached 127 mm (Supplemental Table 2).

Q18 From this database, we extracted repeated surveys ($2 \leq n \leq 10$ surveys per plot, where a survey is the record of all trees in a given year) from 99,724 total plots (Supplementary Fig. 1). Owing to the variety in the survey methods represented in our database, plot sizes ranged from 400–1,513 m² (median = 672 m²).
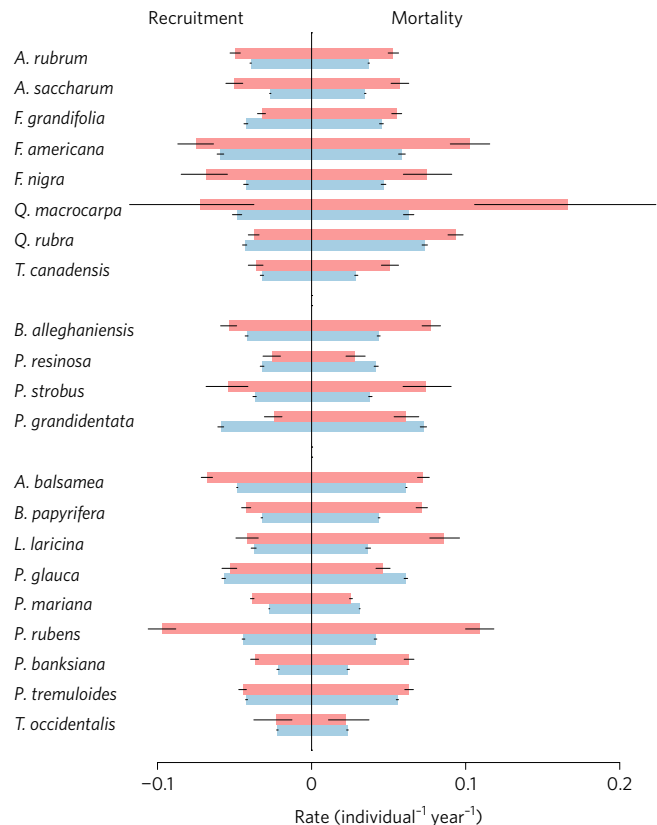


**Figure 4 | Predicted recruitment and mortality rates in equilibrium ranges and in areas facing extinction debt.** Mortality rates were systematically greater in extinction debt zones (red) than in the equilibrium range (blue), whereas differences in recruitment rates varied less systematically. Bars show mean predicted rates (with 90% credible intervals) from a hierarchical model with species as a grouping factor. Data are for all individuals with DBH > 127 mm.

To select our study species, we computed the basal area (in m² ha⁻¹) of all species occurring in plots in the 42–50° latitude range (representing the approximate limits of the temperate–boreal transition zone). We then ranked the species by basal area within these plots and selected the most abundant species that summed to at least 90% of the basal area. This procedure resulted in 21 species representing 90% of the basal area, including nine primarily boreal species, eight temperate species and four transitional species.

To identify observed colonizations and extinctions, we searched the database for plots that were surveyed multiple times. We defined a sampling unit as an initial observation within a plot and its subsequent repeat (therefore, for example, a plot surveyed in 1990, 1995 and 2000 had 2 samples: 1990–1995 and 1995–2000). For each species we classified every sampling unit as an absence (that is, absent for both observations; $n = 37,265–79,13$... observations per species), a presence Q19 (present during both observations; $n = 1,765–30,403$), a colonization (absence followed by presence, $n = 112–2,684$), or an extinction (presence followed by absence, $n = 108–2,182$) (Supplementary Table 1). We removed intervals greater than 15 years from the dataset. This minimizes the risk of multiple transitions within an interval (for example, a colonization followed by an extinction, which would be observed as an absence).

To fit climatic relationships, we used annual 10-km² climate rasters for the entire study area covering the period from 1945–2010 that included six bioclimatic variables: annual mean temperature (°C), mean diurnal temperature range (°C), the mean temperature in the wettest quarter (°C), total annual precipitation (mm), precipitation seasonality and mean precipitation in the warmest quarter (mm). All bioclimatic rasters were obtained from ANUSPLIN, a non-parametric, multi-dimensional curve fitting technique that interprets data from meteorological stations into continuous surfaces[41]. To minimize the effects of extreme years on model calibration, climate for a plot in a given year was then computed as the mean of the climate over the 15 years preceding the observation. Because the full dataset includes forest plots for many species, for any given species many plots were well outside the species' climatic range and provided little additional information for fitting colonization–extinction curves (which are most sensitive

at the range boundaries). Fitting a model with the full dataset thus potentially puts a large amount of likelihood weight in the tails of the colonization–extinction curves where both colonization and extinction probabilities are near zero (because the species is never present). Therefore, to avoid overfitting to the tails and to put the emphasis on climatic conditions where the species is potentially present, we restricted the calibration data range by eliminating all plots that were outside a 10° latitude and longitude buffer surrounding all observed presences. For similar reasons, we truncated climatic data that extends beyond the northern limit of the forest surveys and do not project our models beyond these limits.

**Calibration.** Our primary modelling goal was to provide a simple interpretation of how the colonization–extinction probabilities varied as functions of temperature and precipitation. Therefore, we selected a priori a single representative of each for model fitting. On the basis of exploratory analyses, mean annual temperature and total annual precipitation demonstrated good relationships to colonization–extinction probabilities and simple interpretations, and these were therefore selected for all species. The colonization–extinction functions (equation (1)) were fit as second-order polynomials of temperature and precipitation using a Bayesian model with binomially distributed errors and a logistic link function. All model parameters were fit with weakly informative Cauchy priors[42]. Because the sampling interval was not equal across samples (range 5–15 years), it was necessary to standardize the parameters to a single sampling interval and then compute the likelihood of an observation given the (constant interval) transition probability and the observation interval. We selected five years as the standard interval because the majority of plots were sampled at five-year intervals. Therefore, the likelihood of an observed extinction (for instance) over an $n$-year interval ($m_n$) given the standardized probability over a five-year interval ($m_s$) can be computed as:

$$m_n(T, P) = 1 - [1 - m_s(T, P)]^{\left(\frac{n}{5}\right)} \qquad (2)$$

where $T$ and $P$ are mean annual temperature and total annual precipitation, respectively. The likelihood of an observed colonization, $c_n(T,P)p(T,P)$, was similarly computed, with the added complication that, under metapopulation theory, an observed colonization depends on both the probability of colonization and establishment ($c_n(T,P)$) and the prevalence ($p(T,P)$), representing the availability of a nearby seed source.

Computing the exact prevalence in the neighbourhood of the plots was impossible with the forest inventory data, as landscape-level surveys of surrounding forests were unavailable and most forest community maps and remote sensing data sources do not resolve to abundances of individual species. Instead, we estimated prevalence using the random forest algorithm. The response variable with this model was the binomial variable ($\phi(E)$, $N(E)$), where $\phi(E)$ is the number of plots in which the species was present under environmental conditions $E$ and $N(E)$ is the total number of plots under the same environmental conditions. The predictor variable $E$ consisted of the six climatic variables described above. This model was then projected to the individual plots, where the prevalence $p_x$ of a species at a particular plot $x$ was interpreted as the conditional expectation of the random forest model $\theta$ : $p_x = E(\theta|E_x)$.

To estimate the posterior distributions of the parameters of the colonization–extinction functions, we used Markov chain Monte Carlo (MCMC) with a Metropolis–Hastings sampler. To assure that the samplers had forgotten their initial conditions and were sampling from the stable distribution, we ran four chains with overdispersed starting values and computed the potential scale reduction factor[43] ($\hat{r}$). Chains were considered to be converged when the upper bound of $\hat{r}$ was less than 1.1 for all parameters. We then selected a single chain at random, ran it for 50% longer than required for convergence, and discarded the first 50% of samples. Finally, to reduce autocorrelation in the samples, reduce disk storage requirements and simplify derived computations, we thinned the samples, retaining a single sample at a regular interval such that the final sample size was 10,000. To prevent skewed posterior distributions from influencing parameter estimates, we report posterior medians, rather than means, for all results. We also report posterior uncertainty intervals for all parameters and derived statistics. Unless otherwise specified, these intervals are reported as 90% credible intervals and are computed as the 90% quantiles across all (thinned) MCMC iterations.

**Analysis.** To determine predicted extinction debt and colonization credit, we first projected the metapopulation model spatially by computing the proportion of simulations in which each species was predicted to be present within each cell of the climate rasters used to calibrate the model. This prediction is a stochastic equilibrium fit from dynamic data and represents the long-term dynamics of the system. We then estimated the present species distributions using the random forest models. This prediction is an instantaneous prediction fit from static data and represents the best estimate of the present distribution of the species. Presence and absence thresholds for each model were determined by maximizing the evaluation statistic (see 'Model evaluation'). The status in each raster cell was then determined by the presence or absence prediction of the two models; continued presence, when both models predicted presence, extinction debt, when the metapopulation model predicted absence at equilibrium, but the random forest predicted presence, and colonization credit for the inverse.

To explore the transient states of the model, we first computed the rate at which patches go extinct when the climate is unsuitable and the rate and which empty patches are colonized when the climate is suitable, defined as:

$$\lambda(E) = \begin{cases} c(E) - m(E), & c(E) > m(E) \\ 0, & c(E) = m(E) \\ m(E) - c(E), & c(E) < m(E) \end{cases}$$

We then simulated the effect of climate change on the range boundary by calculating the partial derivative of this curve with respect to temperature ($\partial \lambda / \partial T$), evaluated at the range boundary (that is, $\lambda(T) = 0$). This is equivalent to a standard local stability analysis, where the perturbation to the equilibrium is caused by a permanent change in parameters[44] and is therefore given by linearization of the system around the equilibrium. This quantity, termed the 'responsiveness', yields the change in colonization and extinction of new patches following a small change in temperature at the range limit and indicates the relative rate of movement of the range boundary.

We fit hierarchical linear models to investigate whether there were differences in individual-level mortality and recruitment rates. To perform this analysis, we extracted information on individual trees from the forest inventory database (as before, considering only trees with DBH > 127 mm). Because trees were individually marked during sampling, we were able to track the status of individuals across successive sampling periods. Thus, we were able to determine, for any given sampling interval, the number of trees that survived (that is, were present and alive in both visits), died (present initially and absent or marked as dead in the second visit) and recruited (not present initially and present in the second visit). We then fit two models, one with recruitment probability and one with mortality probability as the response. In both cases, the explanatory variable was an indicator taking a value of 1 if the plot was considered to be subject to extinction debt (see Fig. 2) or 0 if the plot was within the long-term persistence range. Although some plots from the absence and colonization credit portions of the range had observations (owing to the stochastic nature of the system), these plots were excluded for analysis. We included species as a (random) hierarchical grouping factor.

**Model evaluation.** A large number of plots in the database were sampled only once during the study period. Lacking repeated observations, these data are unsuitable for calibrating the colonization–extinction functions, and were therefore not used in fitting the metapopulation model. However, they represent a large source of information about species distribution that is independent of the data used to calibrate the model. Therefore, we used this dataset ($n = 64,848$) to evaluate the model predictions for species distributions. For each posterior sample, we computed the conditional prediction of presence or absence of each species using the climatic data for the validation plots. The proportion of presences thus represents the conditional posterior probability that a species $s$ is present at equilibrium under climatic regime $\hat{E}$: $p(X_s = 1 | \hat{E})$. We then used the observed presences and absences in each plot along with the posterior probability of presence to compute the area under the receiver operating curve (ROC), an estimate of classification ability that ranges from 0.5 to 1, where 0.5 is random classification and 1 is perfect classification. This is the most stringent validation test we could think of, as the model was fitted from dynamic data and validated on an integrated prediction about static distribution. Models for 19 out of 21 species performed well (0.70 < ROC ≤ 0.91), suggesting that the core range for these species is predicted well by the model (Supplemental Table 3).

**Data availability.** Some data reported in this paper are publicly available via figshare: https://doi.org/10.6084/m9.figshare.4906535. Data not included in this repository are available from the authors upon request.

## References

1. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.* **37,** 637–669 (2006).
2. Austin, M. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol. Modell.* **200,** 1–19 (2007).
3. Guisan, A. & Thuiller, W. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8,** 993–1009 (2005).
4. Thuiller, W. *et al.* Consequences of climate change on the tree of life in Europe. *Nature* **470,** 531–534 (2011).
5. Thuiller, W. *et al.* The European functional tree of bird life in the face of global change. *Nat. Commun.* **5,** 3118 (2014).
6. Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C. & Finnegan, S. Climate change and the past, present, and future of biotic interactions. *Science* **341,** 499–504 (2013).

7. Guisan, A. *et al.* Predicting species distributions for conservation decisions. *Ecol. Lett.* **16,** 1424–1435 (2013).

8. García-Valdés, R., Zavala, M. A., Araújo, M. B. & Purves, D. W. Chasing a moving target: projecting climate change-induced shifts in non-equilibrial tree species distributions. *J. Ecol.* **101,** 441–453 (2013).

9. Tilman, D., May, R. M., Lehman, C. L. & Nowak, M. A. Habitat destruction and the extinction debt. *Nature* **371,** 65–66 (1994).

10. Kuussaari, M. *et al.* Extinction debt: a challenge for biodiversity conservation. *Trends Ecol. Evol.* **24,** 564–571 (2009).

11. Dullinger, S. *et al.* Extinction debt of high-mountain plants under twenty-first-century climate change. *Nat. Clim. Change* **2,** 619–622 (2012).

12. Svenning, J.-C. & Sandel, B. Disequilibrium vegetation dynamics under future climate change. *Am. J. Bot.* **100,** 1266–1286 (2013).

13. Purves, D. W., Lichstein, J. W., Strigul, N. & Pacala, S. W. Predicting and understanding forest dynamics using a simple tractable model. *Proc. Natl Acad. Sci. USA* **105,** 17018–17022 (2008).

14. Botkin, D. B. *JABOWA-II: a Computer Model of Forest Growth* (Oxford Univ. Press, 1993).

15. Scheller, R. M. *et al.* Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecol. Modell.* **201,** 409–419 (2007).

16. Boulangeat, I., Georges, D. & Thuiller, W. FATE-HD: a spatially and temporally explicit integrated model for predicting vegetation structure and diversity at regional scale. *Glob. Change Biol.* **20,** 2368–2378 (2014).

17. Snell, R. S. *et al.* Using dynamic vegetation models to simulate plant range shifts. *Ecography* **37,** 1184–1197 (2014).

18. Hartig, F. *et al.* Connecting dynamic vegetation models to data - an inverse perspective. *J. Biogeogr.* **39,** 2240–2252 (2012).

19. Levins, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* **15,** 237–240 (1969).

20. Holt, R. D. & Keitt, T. H. Alternative causes for range limits: a metapopulation perspective. *Ecol. Lett.* **3,** 41–47 (2000).

21. Holt, R. D., Keitt, T. H., Lewis, M. A., Maurer, B. A. & Taper, M. L. Theoretical models of species' borders: single species approaches. *Oikos* **108,** 18–27 (2005).

22. Kellman, M. Sugar maple (*Acer saccharum* Marsh.) establishment in boreal forest: results of a transplantation experiment. *J. Biogeogr.* **31,** 1515–1522 (2004).

23. Tilman, D. Competition and biodiversity in spatially structured habitats. *Ecology* **75,** 2–16 (1994).

24. Lande, R. Extinction thresholds in demographic models of territorial populations. *Am. Nat.* **130,** 624–635 (1987).

25. Ebenhard, T. Colonization in metapopulations: a review of theory and observations. *Biol. J. Linn. Soc.* **42,** 105–121 (1991).

26. Yackulic, C. B., Nichols, J. D., Reid, J. & Der, R. To predict the niche, model colonization and extinction. *Ecology* **96,** 16–23 (2015).

27. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).

28. Boulangeat, I. *et al.* Anticipating the spatio-temporal response of plant diversity and vegetation structure to climate and land use change in a protected area. *Ecography* **37,** 1230–1239 (2014).

29. Bertrand, R. *et al.* Changes in plant community composition lag behind climate warming in lowland forests. *Nature* **479,** 517–520 (2011).

30. Vellend, M. *et al.* Extinction debt of forest plants persists for more than a century following habitat fragmentation. *Ecology* **87,** 542–548 (2006).

31. Peñuelas, J., Ogaya, R., Boada, M. & Jump, A. S. Migration, invasion and decline: changes in recruitment and forest structure in a warming-linked shift of European beech forest in Catalonia (NE Spain). *Ecography* **30,** 829–837 (2007).

32. Scheffer, M., Hirota, M., Holmgren, M., van Nes, E. H. & Chapin, F. S. Thresholds for boreal biome transitions. *Proc. Natl Acad. Sci. USA* **109,** 21384–21389 (2012).

33. Kelly, C. K. & Bowler, M. G. Coexistence and relative abundance in forest trees. *Nature* **417,** 437–440 (2002).

34. Thuiller, W. *et al.* A road map for integrating eco-evolutionary processes into biodiversity models. *Ecol. Lett.* **16,** 94–105 (2013).

35. Pagel, J. & Schurr, F. M. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Glob. Ecol. Biogeogr.* **21,** 293–304 (2012).

36. Talluto, M. V. *et al.* Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. *Glob. Ecol. Biogeogr.* **25,** 238–249 (2016).

37. O'Connell, M. B. *et al. The Forest Inventory and Analysis Database: Database Description and Users Guide Version 3.0* (US Department of Agriculture, Forest Service 2007). **Q25**

38. Ministère des Ressources naturelles. *Normes d'inventaire forestier* (Direction des inventaires forestier, Ministère des Ressources naturelles, 2013).

39. Ontario Ministry of Natural Resources. in *Growth Yield Progr. PSP PGP Ref. Man.* (2014). **Q26**

40. Porter, K. B., Maclean, D. A., Beaton, K. P. & Upshall, J. *Base de données sur les placettes d'échantillonnage permanentes du Nouveau-Brunswick (PSPDB v1.0): Guide de l'utilisateur et analyse* (2001). **Q27**

41. McKenney, D. W. *et al.* Customized spatial climate models for North America. *Bull. Amer. Meteor. Soc.* **92,** 1611–1622 (2011).

42. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2,** 1360–1383 (2008).

43. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7,** 457–472 (1992).

44. Bender, E. A., Case, T. J. & Gilpin, M. E. Perturbation experiments in community ecology: theory and practice. *Ecology* **65,** 1–13 (1984).

## Acknowledgements

## Author contributions

M.V.T., D.G., S.V. and I.B. conceived the study, S.V. built the database and M.V.T. wrote the code for the model, MCMC samplers, and ran the analyses. All authors contributed to writing the manuscript. **Q2**

## Additional information

**Supplementary information** is available for this paper.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to M.V.T.

**How to cite this article:** Talluto, M. V., Boulangeat, I., Vissault, S., Thuiller, W. & Gravel, D. Extinction debt and colonization credit delay range shifts of eastern North American trees. *Nat. Ecol. Evol.* **1,** 0182 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Competing interests

The authors declare no competing financial interests.

# QUERY FORM

| Nature Ecology & Evolution | |
|---|---|
| **Manuscript ID** | **[Art. Id: s41559-017-0182]** |
| **Author** | |
| **Editor** | |
| **Publisher** | |

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer queries by making the requisite corrections directly on the PDF proof. We prefer to receive all proof comments as annotated PDFs. Please verify receipt of proofs via e-mail

| Query No. | Nature of Query |
|---|---|
| Q1 | Please include a postal code for affiliations 2 and 3; addition of state correct for affiliation 3? |
| Q2 | Do you wish to include a separate contribution for W.T. (Wilfried Thuiller), or is his contribution included in the 'all authors contributed to writing the manuscript'? |
| Q3 | Please note, we only use single quotation marks here, and only upon first use; the quotation marks for subsequent uses have been removed, OK? |
| Q4 | Please check that the edits to the sentence 'based on' have retained your intended meaning. |
| Q5 | Please note, we avoid the use of i.e. and e.g., please see changes throughout. |
| Q6 | Please check that the edits to the sentence 'inherently' have retained your intended meaning. |
| Q7 | Please note, we avoid the use of two-letter abbreviations; CE has been written out in full throughout, OK? |
| Q8 | We avoid the use of greater/smaller than symbols in the text, change OK? |
| Q9 | Please note, we avoid the use of square brackets, change ok? |
| Q10 | Please note, we reserve the use of 'significant' for statistical significance, please see change. |
| Q11 | Please check that the edits to the sentence 'trend of increased responsiveness' have retained your intended meaning. |
| Q12 | Is this statistically significant? Please include a $P$ value if possible. |
| Q13 | Is this statistically significant? Please include a $P$ value if possible. |
| Q14 | Please note, where a reference citation could be misread as an index, it has been set on the line, not as a superscript. |
| Q15 | Please check that the edits to the sentence 'Similar to' have retained your intended meaning. |
| Q16 | Please check that the edits to the sentence 'that are missing' have retained your intended meaning. |
| Q17 | Please note, all equations have been rekeyed. Please check these carefully. Single-letter variables and matrices are italics, vectors should be italics and bold, non-variable numbers are set as Roman, OK? |

# QUERY FORM

| | |
|---|---|
| **Nature Ecology & Evolution** ||
| **Manuscript ID** | [Art. Id:  s41559-017-0182] |
| **Author** | |
| **Editor** | |
| **Publisher** | |

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer queries by making the requisite corrections directly on the PDF proof. We prefer to receive all proof comments as annotated PDFs. Please verify receipt of proofs via e-mail

| *Query No.* | *Nature of Query* |
|---|---|
| Q18 | Our style for n numbers is lower case, italics, OK? |
| Q19 | Do you mean 37,265 out of 79,173 observations throughout or is this a range? |
| Q20 | Move of the reference to avoid superscripting on a variable OK? |
| Q21 | We avoid the use of '/' to indicate interactions and complexes, is it used to denote 'and'. |
| Q22 | Please note, we do not use bold or italics for emphasis, change to single quotation marks OK? |
| Q23 | We avoid the use of '/' to indicate interactions and complexes, is it used to denote 'or'. |
| Q24 | We avoid the use of '/' to indicate interactions and complexes, is it used to denote 'or'. |
| Q25 | Ref. 37: Should the Authors for this book be 'Forest Inventory and Analysis Program' instead of O'Connell, M. B. *et al.*? Please check the title and publisher are correct for this reference. |
| Q26 | Please check ref. 39 carefully, Is this the full title? We would not usually use abbreviations in the book title, unless this is as the title is originally. Is there a publisher we could include? |
| Q27 | Please include a publisher for ref. 40 if possible |
| Q28 | Addition of degrees celsius here ok? Or do you mean degrees in latitude? |
| Q29 | Author surnames have been highlighted – please check these carefully and indicate if the first name or surname have been marked up incorrectly. Please note that this will affect indexing of your article, such as in PubMed. |
| | |