Ecography

Supplementary material

Supplementary Materials for: Matching the forecast horizon with the relevant spatial and temporal processes and data sources

Peter B. Adler[1], Ethan P. White[2,3,4] and Michael H. Cortez[5]

[1]Department of Wildland Resources and the Ecology Center, Utah State University, Logan, Utah
[2]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida
[3]Informatics Institute, University of Florida, Gainesville, Florida
[4]Biodiversity Institute, University of Florida, Gainesville, Florida
[5]Department of Biological Science, Florida State University, Tallahasee, Florida

Appendices

A   Spatial, temporal and spatial-temporal-weighted models

The two simulation models in the main text describe how population size, $N(x, t)$, at location x changes over time (t). We assume that the temperature, $K(x, t)$, at each location can vary in time and space. To forecast the dynamics generated by these simulations models, we fit a series of statistical models.

The spatial model, which we refer to as S, is a quadratic regression of the mean long-term population density at a location ($\bar{N}(x)$) against the mean temperature at that location ($\bar{K}(x)$). The quadratic term describes the unimodal relationship between $\bar{N}$ and $\bar{K}$. The spatial statistical model is

$$\bar{N}(x) = S(\bar{K}(x)) = \beta_0^S + \beta_1^S \bar{K}(x) + \beta_2^S \bar{K}(x)^2 + \varepsilon \tag{1}$$

The temporal model, which we call T, starts with a time-series of "observed" population sizes, or total biomasses, at one location, $N(t)$, for $t = 1...n$ (the spatial index is suppressed because we only focus on one location at a time). In the community turnover example, we fit the following regression, which predicts biomass at time $t + 1$ as a function of biomass ($N(t)$) and annual temperature ($K(t)$) at time t,

$$\ln(N(t+1)) = T(N(t), K(t)) = \beta_0^T + \beta_1^T \ln(N(t)) + \beta_2^T K(t) + \varepsilon \tag{2}$$

In the eco-evolutionary example, the response variable is the log of the population growth rate. The regression, which includes a quadratic effect of temperature, is

$$\ln\left(\frac{N(t+1)}{N(t)}\right) = T(N(t), K(t)) = \beta_0^T + \beta_1^T \ln(N(t)) + \beta_2^T K(t) + \beta_3^T K(t)^2 + \varepsilon \tag{3}$$

This version of the temporal model returns a per capita growth rate on the log scale. To predict population size at the next time step, we exponentiate the growth rate and multiply it by the current population size: $\exp(T(N(t), K(t)))N(t)$.

The weighted model is a weighted average of predictions from the spatial and temporal models, with the weights changing as a function of time, here expressed as the forecast horizon. The weights change as a function of the square root of the forecast horizon, to allow rapid shifts in the model weights.

$$logit(\omega_t) = \beta_0^W + \beta_1^W \sqrt{t} \tag{4}$$

For the community turnover example, the predicted biomass from the weighted model is:

$$\hat{N}(t+1) = \omega \cdot T(N(t), K(t)) + (1 - \omega) \cdot S(K(t)) \tag{5}$$

Again, we suppress the spatial subscript (x) here because we are focused on densities at just one location. For the eco-evolutionary example, the predicted population size from the weighted model is:

$$\hat{N}(t+1) = \omega \cdot \exp(T(N(t), K(t)))N(t) + (1 - \omega) \cdot S(K(t)) \tag{6}$$

We used the `optim` function to estimate the β^Ws that minimize the sum of squared errors, $(\hat{N}(t+1) - N(t+1))^2$.

In the main text, we show the point forecasts but not the uncertainty around the forecasts. After exploring that uncertainty, we decided that presenting it would be misleading. For the spatial and, especially, the temporal statistical models, the uncertainty is unrealistically low, because the models are estimated with very large samples sizes from the simulations. Furthermore, the simulations do not include noise; the only reason there is any uncertainty is because the statistical models are slightly mis-specified with respect to the process models. Showing uncertainty for the weighted model would be even less meaningful, because it is not a true, out-of-sample forecast (parameters are fit directly to the observations for which we make predictions). The R code to compute uncertainties for the spatial and temporal forecasts is available on our Github repository (https://github.com/pbadler/space-time-forecast), but is commented out.

B   Description of the meta-community model

Alexander et al. (2018) developed a meta-community model to represent dynamics of local communities arrayed along a one-dimensional elevation gradient, as influenced by three main pro-

cesses: temperature-dependent growth, competition, and dispersal. Here we adapt their notation to be consistent with our own.

The population size of species i in cell x at time $t + 1$, $N_i(x, t + 1)$, is computed in two steps. The first step accounts for changes in local population sizes due to dispersal. In each local community, all species export a fraction (d) of their local population to the two adjacent communities in the 1-dimensional landscape:

$$N_i'(x,t) = (1 - d) \cdot N_i(x,t) + \frac{d}{2} \cdot (N_i(x+1,t) + N_i(x-1,t)) \tag{7}$$

Here N' distinguishes the post-dispersal population size from the pre-dispersal population size.

The second step computes population growth, taking into account competition:

$$N_i(x,t+1) = N_i'(x,t) + N_i'(x,t)[g_i(K(x) - Kmin_i) - c_i N_i'(x,t) - l_i \sum_k N_k'(x,t)] \tag{8}$$

In the absence of competition, the growth rate (g_i) is determined by the difference between the temperature at site x ($K(x)$) and the focal species' minimum temperature tolerance, $Kmin_i$, the lowest temperature at which a species can maintain a positive growth rate. Growth is further reduced by intraspecific and interspecific competition, parameterized by c_i and l_i. All species are assigned the same value of c_i, which represents an additional effect of intraspecific competition on top of interspecific competition. This stabilizes coexistence, since every species will exert stronger intra- than interspecific competition. However, values of l vary among species to create a trade-off between growth rates and competitive ability versus low temperature tolerance: fast-growing species (high g_i) are more tolerant of interspecific competition (low l_i) but are more limited by temperature (high $Kmin_i$).

To assign species-specific parameter values, the number of species in the metacommunity is specified. Next, each species is assigned an optimal temperature within a specified temperature range by drawing from a uniform distribution. Sensitivity to interspecific competition is then determined as a decreasing function of optimal temperature. Calculations are performed in the script `SpeciesPoolGen.R`.

C Description of the eco-evolutionary annual plant model

Haploid Model: Begin with a haploid model that describes the number of seeds present in a population. We model a scenario in which all seeds germinate, so we can ignore seedbank dynamics. $N_{i,t}$ is the number of seeds of species i at time t. The model is

$$
\begin{aligned}
N_{1,t+1} &= \frac{\lambda_1(K(t))N_{1,t}}{1 + \alpha_{11}N_{1,t} + \alpha_{12}N_{2,t}} \\
N_{2,t+1} &= \frac{\lambda_2(K(t))N_{2,t}}{1 + \alpha_{21}N_{1,t} + \alpha_{22}N_{2,t}}
\end{aligned}
\tag{9}
$$

where $\lambda_i(K(t))$ is the seed production rate per plant, and $K(t)$ is the temperature at time t. Below we refer to the α_{ij} as intra- and inter-genotype competition coefficients.

Diploid Model: Consider a one-species diploid model. The genotypes are denoted by AA, Aa, and aa. The number of each genotype at time t is $N_{AA}(t)$, $N_{Aa}(t)$, and $N_{aa}(t)$. The seed production rate for genotype AA is $\lambda_{AA}(K(t))$, and the analogous parameters for the other genotypes are similarly denoted. The competition coefficients are denoted by $\alpha_{i,j}$, e.g., $\alpha_{AA,AA}$ or $\alpha_{AA,Aa}$. Throughout we assume that gametes mix randomly in the population.

First consider the case where the competition coefficients are zero ($\alpha_{i,j} = 0$). Let T denote the total number of gamete-pairs produced in a given year,

$$T = \lambda_{AA}(K(t))N_{AA}(t) + \lambda_{Aa}(K(t))N_{Aa}(t) + \lambda_{aa}(K(t))N_{aa}(t). \tag{10}$$

The first term is the number of gamete-pairs produced by AA individuals. The second and third terms are the numbers of gamete-pairs produced by Aa and aa individuals, respectively. The proportion of A gametes (ϕ_A) and the proportion of a gametes (ϕ_a) are given by

$$\phi_A = \frac{2\lambda_{AA}(K(t))N_{AA}(t) + \lambda_{Aa}(K(t))N_{Aa}(t)}{2T} \quad \text{and} \quad \phi_a = 1 - \phi_A. \tag{11}$$

Note that the T in the denominator of ϕ_A shows up because we are computing proportions. Combining all of these we get the dynamics for each genotype,

$$\begin{aligned} N_{AA}(t+1) &= \phi_A^2 T \\ N_{Aa}(t+1) &= \phi_A \phi_a T \\ N_{aa}(t+1) &= \phi_a^2 T \end{aligned} \tag{12}$$

Now consider the case where the competition coefficients are non-zero ($\alpha_{i,j} \neq 0$). Including competition changes the way in which we compute T, ϕ_A, and ϕ_a. Specifically, because the total number of seeds produced per year by each genotypes is reduced based on intra- and inter-genotype competition, the total number of gamete-pairs becomes

$$\begin{aligned} T = {} &\frac{\lambda_{AA}(K(t))N_{AA}(t)}{1 + \alpha_{AA,AA}N_{AA}(t) + \alpha_{AA,Aa}N_{Aa}(t) + \alpha_{AA,aa}N_{aa}(t)} \\ &+ \frac{\lambda_{Aa}(K(t))N_{Aa}(t)}{1 + \alpha_{Aa,AA}N_{AA}(t) + \alpha_{Aa,Aa}N_{Aa}(t) + \alpha_{Aa,aa}N_{aa}(t)} \\ &+ \frac{\lambda_{aa}(K(t))N_{aa}(t)}{1 + \alpha_{aa,AA}N_{AA}(t) + \alpha_{aa,Aa}N_{Aa}(t) + \alpha_{aa,aa}N_{aa}(t)}. \end{aligned} \tag{13}$$

The first line is the number of gamete-pairs produced by AA individuals after accounting for the effects of competition. The second and third lines are the numbers of gamete-pairs produced by

4

Aa and *aa* individuals, respectively. The proportions of *A* gametes and *a* gametes are

$$\phi_A = \frac{2}{2T} \frac{\lambda_{AA}(K(t))N_{AA}(t)}{1 + \alpha_{AA,AA}N_{AA}(t) + \alpha_{AA,Aa}N_{Aa}(t) + \alpha_{AA,aa}N_{aa}(t)}$$
$$+ \frac{1}{2T} \frac{\lambda_{Aa}(K(t))N_{Aa}(t)}{1 + \alpha_{Aa,AA}N_{AA}(t) + \alpha_{Aa,Aa}N_{Aa}(t) + \alpha_{Aa,aa}N_{aa}(t)} \quad (14)$$
$$\phi_a = 1 - \phi_A$$

Combining all of this results in the same model as above,

$$N_{AA}(t+1) = \phi_A^2 T$$
$$N_{Aa}(t+1) = 2\phi_A\phi_a T \quad (15)$$
$$N_{aa}(t+1) = \phi_a^2 T,$$

but the definitions of T, ϕ_A, and ϕ_a are given by equations (13) and (14) .

Literature cited

Alexander, J. M. et al. 2018. Lags in the response of mountain plant communities to climate change. – Global Change Biology 24: 563–579.

Supplementary Tables

Table SM-1: Parameters and parameter values for the community turnover case study. Values are assigned at the start of `comm_turn_master.R`. "Name" refers to the variable declared in the computer code. These names do not exactly match the symbols shown in the equations in Appendix B; rather, the species-specific values of those parameters are calculated in the computer code based on the values in this table.

Name	Value	Definition
L_land	20	Length of landscape
Tmin	0	Minimum of spatial gradient in baseline temperature
Tmax	15	Maximum of spatial gradient in baseline temperature
Tstdev	2	Standard deviation of temperature (interannual variation)
deltaT	4	Magnitude of directional change in temperature
burnin_yrs	2000	Number of years to initialize simulation
baseline_yrs	1000	Number of years at baseline temperature used to fit statistical models
warming_yrs	200	Number of years over which temperature increases
final_yrs	2000	Number of years at steady-state, elevated temperature
N	40	Number of species
Gmax	0.5	Maximum population growth rate
Gmin	0.2	Minimum population growth rate
Lmax	1.5	Maximum sensitivity to competition
Lmin	0.7	Minimum sensitivity to competition
Cmax	0.2	Maximum additional sensitivity to conspecific competition
Cmin	0.2	Minimum additional sensitivity to conspecific competition
d	0.01	Fraction of offspring dispersing from home site

Table SM-2: Parameters and parameter values for the eco-evolutionary case study. Values are assigned at at the start of `genetic_diversity_master.R`. "Name" refers to the variable declared in the computer code. Where appropriate, the corresponding symbols from equations in Appendix C are shown in parentheses.

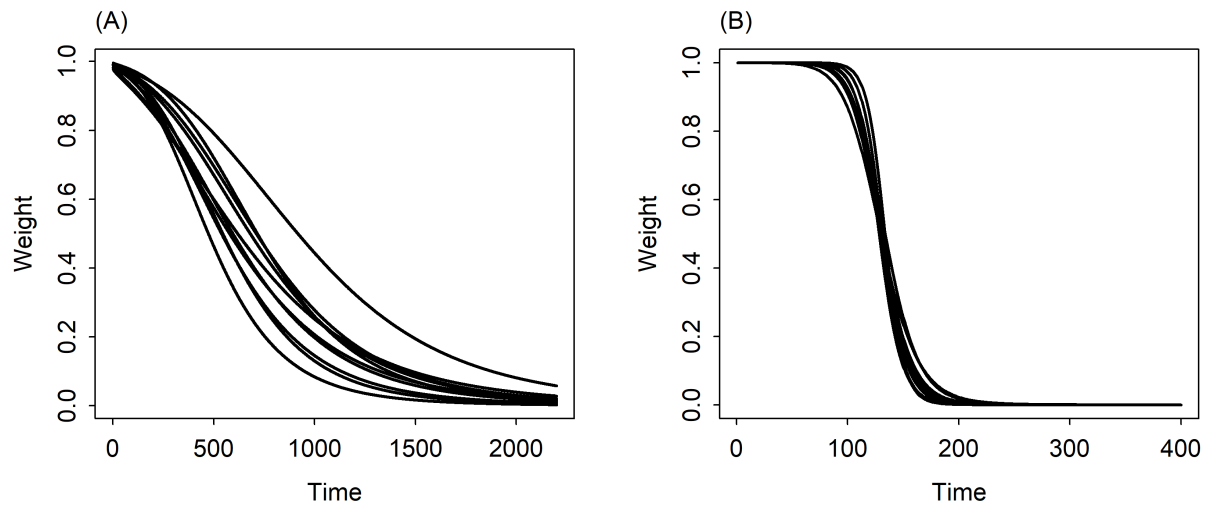| Name	Values	Definition
Tstdev	1	Standard deviation of temperature (interannual variation)
baseT	-1	Baseline temperature
deltaT	5	Total change in temperature
baseline_yrs	500	Number of years at baseline temperature used to fit statistical models
warming_yrs	100	Number of years over which temperature increases
final_yrs	300	Number of years at steady-state, elevated temperature
fec_Tmu	-1,0,1	Optimal fecundity temperature for genotypes AA, Aa, and aa
fec_Tsigma	8	Standard deviation in fecundity for all genotypes
fec_max	100	Maximum fecundity for all genotypes
alpha (α)	1	All competition coefficients for all genotypes

Supplementary Figures

Figure SM-1: (A) Temporal shifts in the model weighting term for 10 independent simulations of (A) the community turnover model, and (B) the eco-evolutionary model. For the community turnover model, each simulation began with initialization of a new regional species pool. For the eco-evolutionary model, genotype parameters were fixed, and only the sequence of annual temperatures varied between runs. In all cases, the combined forecast is heavily weighted towards the time-series model at short forecast time scales, and towards the space-for-time model at long forecast time scales.
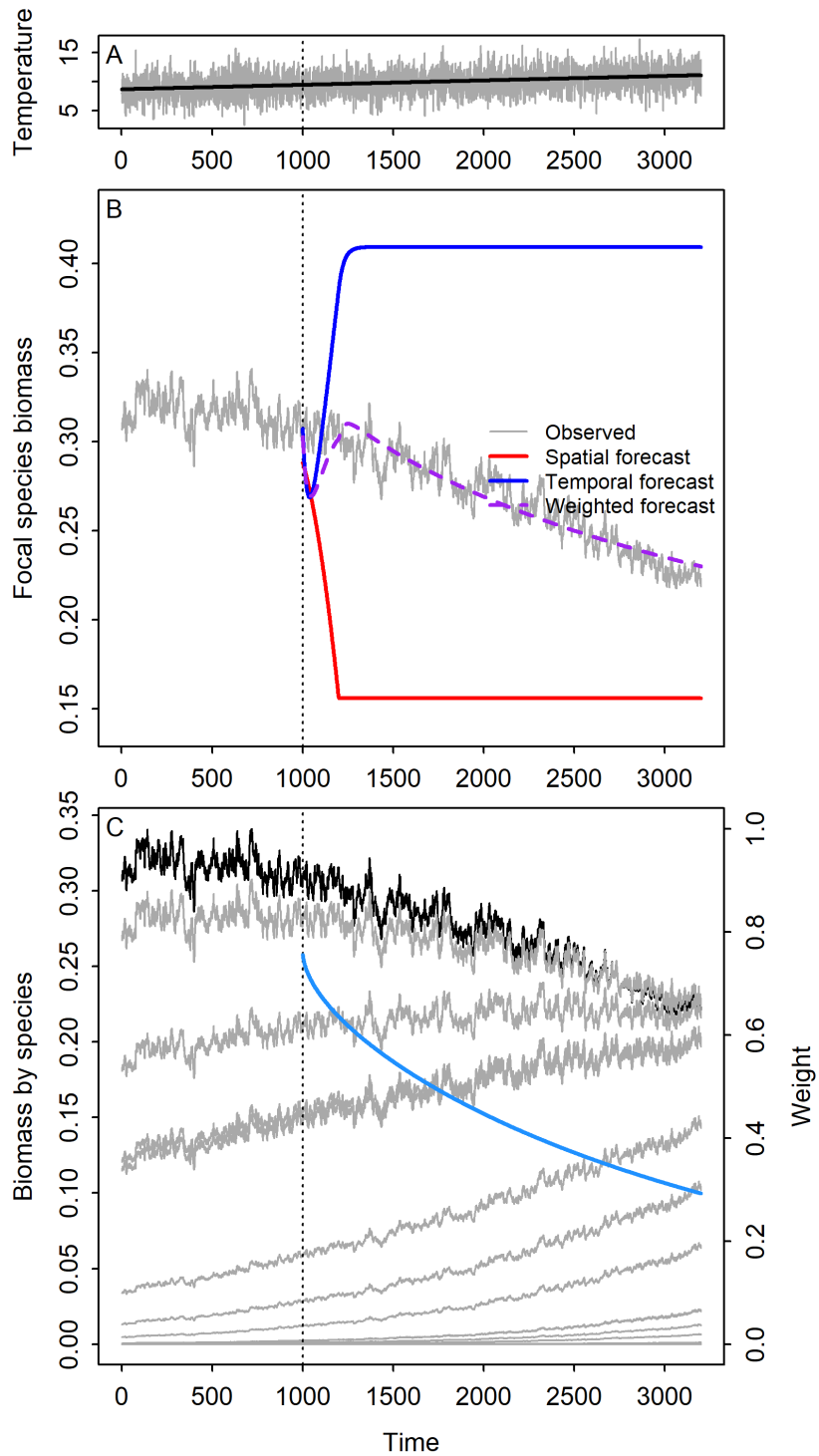
Figure SM-2: (A) Simulated annual temperatures (grey) and expected temperature (black), which was used to make forecasts, at the focal site. In contrast to Fig. **??**, which shows results for a period of warming followed by stationary temperatures, for this simulation we spread the same temperature increase out over the entire simulation with no stationary periods. (B) Simulated focal species biomass and forecasts from the spatial, temporal and weighted statistical models at the focal site in the metacommunity model. (C) Simulated biomass of the focal species (black) and all other species (grey), and the weight given to the temporal statistical model for focal species biomass (blue). Time 1000 (years) in each panel corresponds to the start of the temperature increase.
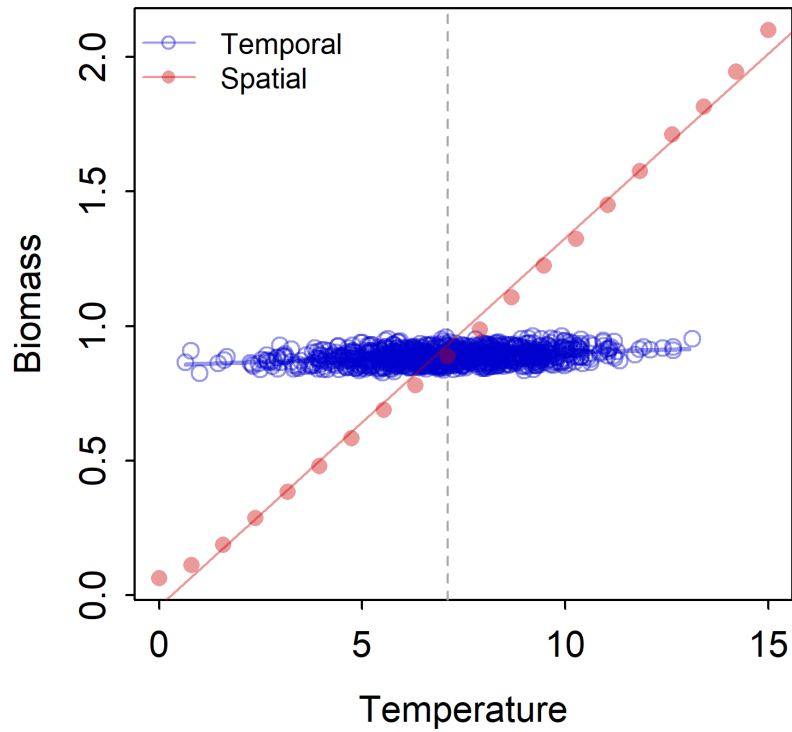
8

Figure SM-3: Results for total biomass from the community turnover model. Blue points show mean total biomass during the baseline period at locations across the temperature gradient, and the blue line shows predictions from the spatial model. Red points show annual total biomass during the baseline period as a function of annual temperature at the central site on the gradient. The red line shows predictions from the temporal model.
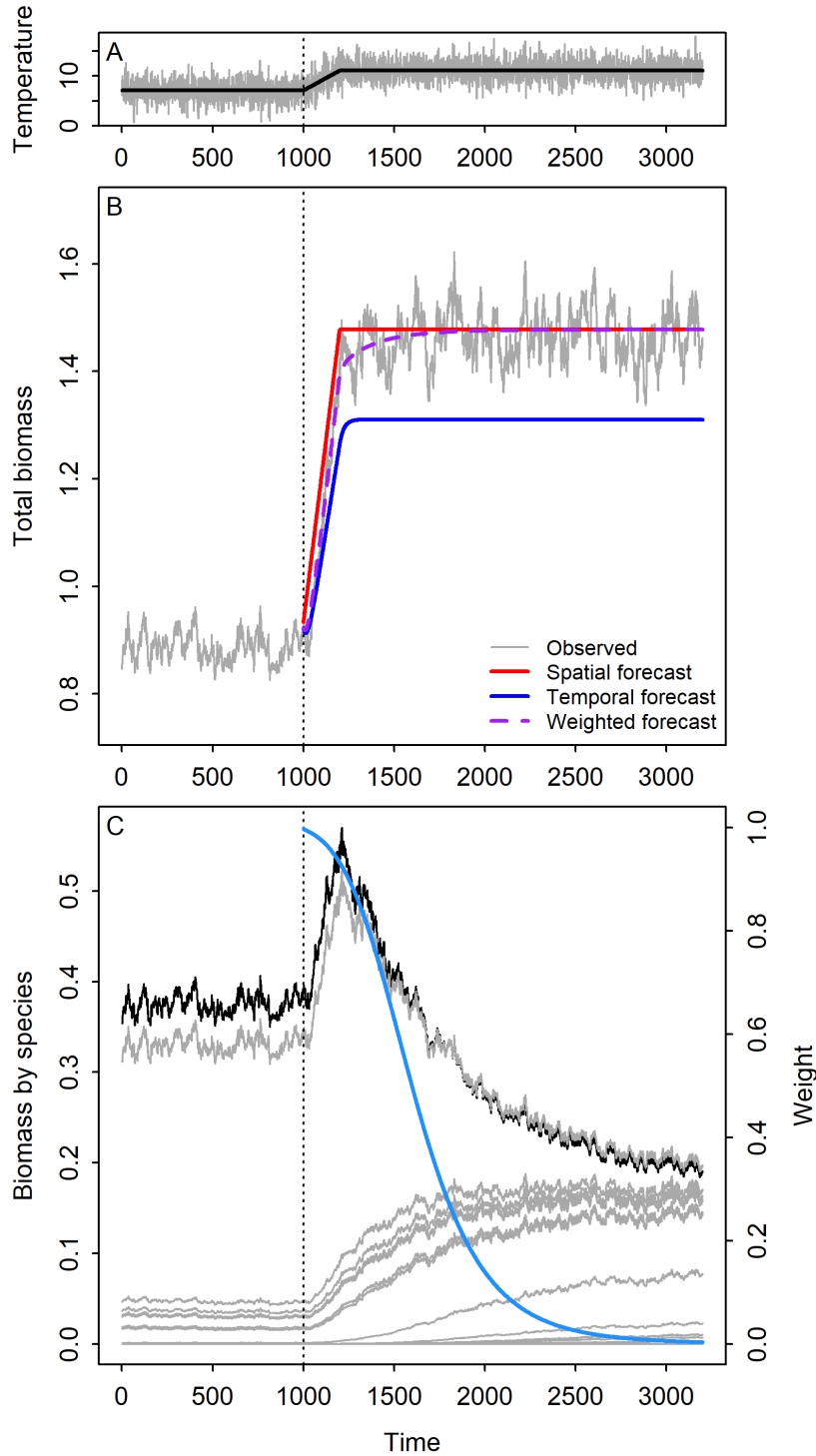
Figure SM-4: Results for total biomass from the community turnover model. (A) Simulated annual temperatures (grey) and expected temperature (black), which was used to make forecasts, at the focal site. (B) Simulated total biomass and forecasts from the spatial, temporal and weighted models. (C) Simulated changes in biomass of all species (grey) at the focal site in the metacommunity model, and the weight given to the temporal model for total biomass (blue). Time 1000 (years) in this figure corresponds to the start of the temperature increase.
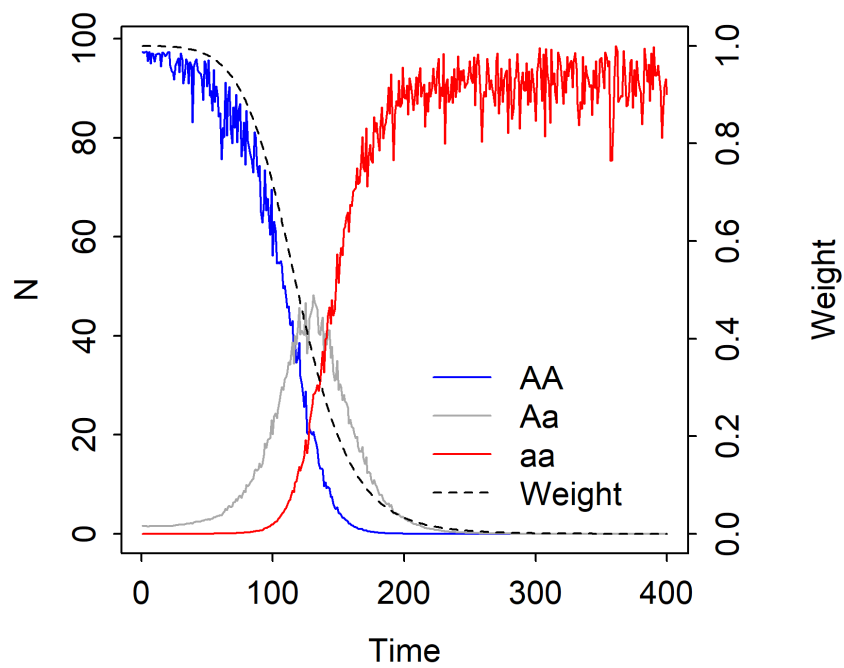
Figure SM-5: Simulated shifts in genotype abundances, and the model weighting term, ω, during the warming phase and the following stationary temperature phase. Time 0 (years) in this figure corresponds to the start of the temperature increase.