S1 – Detailed methods

Study system

Parameterizing the theoretical models is a data-hungry exercise; our long-term, individual-based population study of great tits represents one of the few studies of wild vertebrates where enough is known about the relevant genetics, demography and ecology to be able to do so.

The data analysed come from a long-term population study of great tits (Parus major) at the Hoge Veluwe National Park in the Netherlands (52° 02' 07" N 5° 51' 32" E). The study area consists of mixed pine-deciduous woodland on poor sandy soils, and we focussed on the years 1973-2011. Data were available from prior to 1973, but the study area changed in size and tree composition following a severe storm in the winter of 1972/1973. The study area remained the same size from 1973 onwards and the number of nest boxes was approximately constant, although some were replaced or moved as the study progressed. A surplus of nest boxes was provided, with one in every three being occupied in a typical year, to ensure that availability of artificial nest sites did not limit population size. The study area is surrounded by a matrix of potentially suitable breeding habitat for great tits, thus the population is open to immigration and emigration.

During the breeding season (April to June/July), nest boxes were visited at least once per week and the number of eggs or nestlings present counted. Nestlings were given aluminium leg rings with unique numbers when 7-10 days old and the parents caught on the nest using a spring trap. Parents already ringed were identified and unringed birds were ringed, allowing us to assess year-to-year survival rates. In some years, brood size manipulation experiments were carried out that affected fledgling or recruitment success; these manipulated broods were excluded in the calculation of the fitness curve. Dates of the peak in caterpillar biomass were estimated for 1985 – 2010 from frass fall samples in the Hoge Veluwe. The most predominant caterpillar species in the study area are the winter moth (Operophtera brumata) and the oak leaf roller (Tortrix virirdana), although caterpillars of several other species are also present.

Predicting the optimal phenotype

Dynamic programming was used to assess the factors affecting optimal laying dates as a function of changing temperatures. This model has been described elsewhere [1] and see Supplementary Material 2 & 3, but here follows a brief description of the relevant aspects for the current application. The model is designed to calculate when an optimally-behaving female initiates egg-laying depending on her internal state, the external state of the environment and the time of year. To calculate this optimal decision per day, a terminal reward function is defined; the expected fitness at the end of the season depended on female survival, the number of fledglings she produced, and their probability of recruitment as determined by their date of fledgling. Dynamic programming involves backwards iteration from the last to the first day of the season, and using the terminal reward function to calculate the optimal decision for each possible state of the bird and the environment.

The seasonal peak in caterpillar biomass, which is strongly temperature-dependent, determines how many offspring a female can successfully raise and their likelihood of recruiting to the breeding population in subsequent years. Temperature acts as a causal cue for the timing of egg-laying in our great tit study population [2,3], and in the model females utilise temperature information from the
current day and previous days when deciding when to lay. Temperature also influences the energetic costs of egg production and incubation. At the time of egg-laying, a female cannot know exactly when the food peak will occur, and hence stochastic fluctuations in temperature during the period after egg-laying can lead to a mismatch between breeding time and the timing of the food peak. The optimisation model finds the laying dates that maximise fitness given a certain temperature profile, magnitude of stochastic variation in temperatures, and inputted costs of egg production and incubation.

**Predicting egg-laying dates**

We used a proportional hazards model [4] to predict future egg-laying dates as a function of expected daily mean temperatures, as predicted from the climate change scenarios. Breeding time can be viewed as 'time to event' and hence modelled using survival model statistical techniques such as the proportional hazards model. This approach has already been used to model breeding time in great tits and migration time in pink-footed geese [5,6,7].

The proportional hazards model describes the probability per unit of time that an event occurs, \( h(t) \), as a function of the baseline hazard, \( h_0(t) \), and a set of variables that can include fixed, \( a_1 \ldots a_n \), and time-dependent variables, \( b_1(t) \ldots b_2(t) \). The model is nonparametric in the sense that no assumptions about the distributions of the times to event or the baseline hazard are made. Using a partial maximum likelihood method allows us to estimate the regression coefficients for the different (fixed and time-dependent) variables without specifying the baseline hazard function [8].

We here modelled the hazard at any day at risk as a function of the (unspecified) baseline hazard, weighted temperature \( T \) as a time-dependent variable and the interaction between temperature and day length \( DL \), also as a time-dependent variable.

\[
h(t,T,DL) = h_0(t) \times \exp(\beta_T \times T(t) + \beta_T \times DL(t)) \quad (eq. 1)
\]

Temperatures were weighted using a so-called 'linear predictor' [9], which assigns daily mean temperatures previous to the current day an exponentially decreasing weight.

\[
\lambda(t) = \lambda(t-1) \times \alpha + T(t) \times (1-\alpha) \quad (eq. 2)
\]

Van de Pol & Cockburn [10] recently proposed a different, more flexible approach to weight encountered temperatures but for our data the two approaches give very similar weighing functions (Schaper et al. unpubl.).

We tried a range of values for \( \alpha \) (0.01 to 0.8) and selected the value that gave the best model fit judged from the model log-likelihood. We then used the coefficients for temperature and the interaction between temperature and day length plus the baseline hazard from this model to forecast egg-laying dates for the predicted future temperatures.

An \( \alpha \) of 0.06 gave the best model fit and the coefficients for temperature and the interaction between temperature and day length were 7.2 (s.e. 0.35) and -0.79 (s.e. 0.04), respectively. The overall model and both variables were highly significant (p<0.001, LRT and t-test, respectively). These results are in agreement with earlier analyses for the same and other great tit populations [6,7].
Parameterising the theoretical models

Width of the fitness function ($\omega$)

Following the convention of the B&L model, we assume that the trait is under stabilising viability selection described by a Gaussian curve. The width of this function, $\omega$ was estimated in two steps. First, observed data on egg-laying dates (3472 records of 2599 females breeding in a period of 37 years), centred by annual mean egg-laying date, were binned into 10 equally-spaced intervals and the mean recruitment probability of an egg, i.e. the probability that an egg hatched, the chick fledged and recruited into the breeding population in subsequent years, was calculated for each. Second, a Gaussian function was fit to these calculated mean probabilities from which $\omega$ was taken as the estimated ‘standard deviation’ (see Fig. S1).

Per capita offspring production ($B$)

The parameter $B$ in the B&L framework essentially describes the number of offspring a typical female is expected to produce across her lifetime in the absence of density regulation, and the height of the fitness function is implicitly scaled to 1 such that females exhibiting the optimal trait value have a total fitness of $B$. Females deviating from the optimum have a total fitness of $BW$, where $W$ represents the fraction of offspring surviving selection (i.e. viability calculated from the fitness function, see Fig. S1). The B&L framework also assumes non-overlapping generations, so all females die immediately after reproducing. Hence, $BW$ (the mean number of surviving offspring per parent) is equivalent to $\lambda$, the per-generation, multiplicative population growth rate. In reality, $\lambda$ will fluctuate around 1 in density-regulated populations (with population size $N$ fluctuating around carrying capacity $K$), but in the absence of density-dependence $\lambda$ must be $\geq 1$ in self-sustaining populations. Unfortunately, we have no direct estimate of $B$ for great tits, as this would need to be estimated at (very) low population densities, which were never observed, and for perfectly timed broods. As a first approximation, however, $e^r$ can be used, where $r$ represents the per-capita rate of population increase, or specific growth rate (on the logarithm scale), at very low densities. [11] estimated $r_1 = 0.474$ for a pied-flycatcher population that colonized our study area in 1959, where $r_1$ is the mean specific growth rate at $N = 1$ (from the theta-logistic model of population growth). As a starting point in our calculations, we therefore used $e^{0.474} = 1.606$ as a reasonable estimate of $B$ for our great tit population. We also explored the sensitivity of the results to a range or $r_1$ values (which we converted to $B$) typical of birds in general [12].

Climate change scenarios

The recently observed (1981-2011) rate of temperature increase is higher than in the SRES A1B scenario, and therefore we used the original 'mild' scenario to compute the 'medium' and 'extreme' climate scenarios by doubling and tripling the temperature increase relative to the base line period 1951-1980. We could not use more extreme IPCC climate change scenarios, such as the A1FI scenario, because predictions from the corresponding regional climate model were not available. The IPCC climate change scenarios are global circulation models that need to be 'downscaled' by being fed into regional climate models that ultimately produce the predictions at the local level, which

requires expert meteorological knowledge (and extensive computing power). We hence followed the outlined approach to create more extreme scenario from the SREA A1B data, for which we did have appropriate regional-scale predictions.

We assumed that the temperatures started to increase from 1981 onwards, and we thus used the long term daily average temperatures from 1951-1980 as the base daily average temperatures. For each ten year period, starting from 1981, we calculated the deviation $\Delta T_{avg10}(d)$ from the long term average $T_{avgbase}(d)$ (i.e. the ‘temperature increase’) as:

$$\Delta T_{avg10}(d) = T_{avg10}(d) - T_{avgbase}(d)$$  \hspace{1cm} (eq. 3)

At the same time, we calculated for each day in the 120 year time series (starting from 1981) the daily deviation $\Delta T(d)$ in average temperature from the corresponding ten year period.

$$\Delta T(d) = T(d) - T_{avg10}(d)$$  \hspace{1cm} (eq. 4)

To calculate the new temperature series for the medium and the extreme scenarios, we take for each period of 10 years starting from 1981, respectively 2 or 3 times the deviation from the 1951-1980 daily average temperatures.

$$T_{medium}(d) = 2 \times (\Delta T_{avg10}(d)) + T_{avgbase}(d) + \Delta T(d)$$  \hspace{1cm} (eq. 5)

$$T_{extreme}(d) = 3 \times (\Delta T_{avg10}(d)) + T_{avgbase}(d) + \Delta T(d)$$  \hspace{1cm} (eq. 6)

Before 1981, we assumed that temperatures remained unchanged, thus we used the original values. We transformed each of the 17 SRES A1B time series independently, to obtain 17 replicates for each of the medium and extreme scenarios.

Fig. S1 Width of the fitness function compared to phenotypic variation in egg-laying dates. The fitness function describes the probability that an egg produces a recruit and was calculated in two steps. First, egg-laying dates were binned into 10 equally-spaced intervals and the mean recruitment probability was calculated for each. Second, a Gaussian function was fit to the data.
Fig. S2 $k_c$ from the B&L model in dependence of effective population size ($N_e$). The black line is $k_c$ for the given values of $N_e$. The green line indicates the critical rate of environmental change as calculated for our great tit population. The red lines indicate the rates of environmental under a mild (dotted), medium (dashed) and extreme climate scenario (solid).

Literature cited