#### RESEARCH ARTICLE



# A time-series model for estimating temporal variation in phenotypic selection on laying dates in a Dutch great tit population

Yihan Cao<sup>1</sup> | Marcel E. Visser<sup>2</sup> | Jarle Tufto<sup>1</sup>

<sup>1</sup>Centre for Biodiversity Dynamics, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

<sup>2</sup>Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands

#### Correspondence

Yihan Cao Email: yihan.cao@ntnu.no

Handling Editor: Thomas Hansen

#### Abstract

- 1. Temporal and spatial variation in phenotypic selection due to changing environmental conditions is of great interest to evolutionary biologists, but few existing methods estimating its magnitude take into account the temporal autocorrelation.
- 2. We use state-space models (SSMs) to analyse phenotypic selection processes that cannot be observed directly and use Template Model builder (TMB), an R package for computing and maximizing the Laplace approximation of the marginal likelihood for SSM and other complex, nonlinear latent variable model via automatic differentiation. Using a long-term great tit (*Parus major*) dataset, we fit several SSMs and conduct model selection based on Akaike information criterion (AIC) to assess the support for fluctuated directional or autocorrelated stabilizing selection on breeding time of the great tit population.
- 3. Our results show that there is directional selection on the probability of breeding failure, and stabilizing selection on the mean number of fledglings. This selection for early laying date is consistent with a previous study of the same population. We also estimate the variation and autocorrelation in other parameters of the fitness functions, including the width and height, and found the height and location of annual fitness function are autocorrelated with significant variation, while the width can be assumed to constant over time.
- 4. Using TMB to fit SSMs, we are able to estimate additional parameters compared to previous methods, all without requiring a substantial increase in computational resources. Furthermore, our specification of complex nonlinear model structure benefits greatly from the flexibility of model formulation with TMB. Therefore, our approach could be directly applied to estimating even more complicated phenotypic selection processes induced by environmental change for other species.

#### KEYWORDS

fluctuating selection, Gaussian fitness function, generalized linear mixed model, state-space model, template model builder, zero-inflated Poisson regression

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Methods in Ecology and Evolution published by John Wiley & Sons Ltd on behalf of British Ecological Society

#### 1 | INTRODUCTION

Fluctuating selection resulting from environmental variation has been of long-lasting interest. Empirical and theoretical research have documented that natural populations respond to varying selection through various mechanism including conventional Darwinian genetic evolution (Lande & Shannon, 1996), evolution of phenotypic plasticity (Scheiner, 1993; Van Tienderen & Koelewijn, 1994), evolution of genetic polymorphism (reviewed by Hedrick, 2006; Bell, 2010), genetic variance (Barton & Keightley, 2002), evolution of the phenotypic variance (Zhang & Hill, 2005) including diversifying bethedging (Bull, 1987; Cohen, 1966; Svardal, Rueffler & Hermisson, 2011) or combinations of these response modes (Tufto, 2015). Importantly, the relative magnitudes of these different responses depend on the temporal autocorrelation in selective optima. Even though the phenotypic traits typically evolve through natural selection to match the environmental conditions to maximize fitness (Futuyma, 2006), phenotypic adaptation through genetic evolution is limited by the amount of genetic variance in the trait under selection, which might lead to mistiming between the mean phenotype and the phenotypic optimum (Lande & Shannon, 1996). Adaptive tracking through phenotypic plasticity acting in conjunction with genetic evolution may also be limited by factors such as imperfect cue reliability (Post & Forchhammer, 2007; Gienapp, Reed & Visser, 2014) or parental energetic costs (Visser, Marvelde & Lof, 2012; Visser et al., 2015).

There are few studies estimating the temporal variability and autocorrelation of phenotypic selection in spite of the importance. The variance in phenotypic selection in previous studies was usually estimated by computing the variance of the strength of selection using selection gradients estimated separately at each time point (reviewed by Siepielski, DiBattista & Carlson, 2009), which reflects both sampling error and real variation in selection (Morrissey & Hadfield, 2012). Among the previous empirical studies accounting for the sampling error of variation, Calsbeek (2011) presented a nonparametric analysis in exploring the variation of fitness surfaces over time or space, but such nonparametric estimates are difficult to relate to parameters appearing in theoretical models. In contrast, using a log-quadratic generalized linear mixed model (GLMM) with a random effect on the regression slope implemented using integrated nested Laplace approximations (INLA) (Rue, Martino & Chopin, 2009), Chevin, Visser and Tufto (2015) estimated yearly fluctuations and autocorrelation in optima of a Gaussian fitness function. However, INLA and GLMMs in general are restricted to cases where the predictor is linear in parameters and random effects. Using instead the more flexible framework of Template Model Builder (TMB) (Kristensen, Nielsen, Berg, Skaug & Bell, 2015), Gamelon et al. (2018) fitted a model of fluctuating selection via several non-overlapping selection episodes with nonlinear random effects added directly on the location of the fitness optima and on the peak of the fitness function. This model form is not feasible within the framework of INLA or GLMMs (see Gamelon et al., 2018, Appendix A for a technical discussion).

Here, we extend the approach taken in Chevin et al. (2015) and Gamelon et al. (2018) in several new ways. First, instead of assuming a fixed zero-inflation parameter for modelling the number of fledglings as in Chevin et al. (2015), we model the zero-inflation probability using a separate linear (or nonlinear) predictor. This leads to a model with selection via zero-inflation and via the Poisson mean. although occurring during the same interval. As with multi-episodic selection more generally (Gamelon et al., 2018), selection through two episodes can involve the same or different biological processes. Second, in addition to random effects on the peak and location of the fitness optimum as in Gamelon et al. (2018), we also allow the width of the fitness function to vary between years, with all three properties of the Gaussian fitness function jointly following a vector autoregressive process. Such variation in the width is of theoretical importance for the evolution of the phenotypic variance (Zhang & Hill, 2005) and for the evolutionary stability of the additive genetic variance-covariance matrix (Revell, 2007). Third, instead of treating the total number of fledglings from all broods laid by a female in a particular year as the sample unit and estimating stabilizing selection on onset of breeding via its effect on the sum of number of fledglings from all broods as in Chevin et al. (2015), we treat the number of fledglings from each brood as the sample unit and fit the model under the assumption that the expected number of fledglings depend on the laying date according to the same Gaussian fitness function for all broods. In addition to increased statistical power, this has the advantage that the parameters relate directly to theoretical models for the joint evolution of multiple brooding and onset of breeding (Tufto, Cao and Visser, submitted manuscript). Fourth, as an alternative to stabilizing selection, we allow each episode (here selection via zero-inflation and via the Poisson mean) to instead involve directional selection. As in Gamelon et al. (2018), we implement our method using TMB (Kristensen et al., 2015), an R package providing a comprehensive framework for fast fitting nonlinear, complex, latent variable models.

### 2 | MATERIALS AND METHODS

## 2.1 | Study population

The data analysed come from a natural population of great tits ( $Parus\ major$ ) at the Hoge Veluwe National Park in the Netherlands ( $52^{\circ}02'-52^{\circ}07'N$ ,  $5^{\circ}51'-5^{\circ}32'$  E). Female great tits usually start reproduction in the second calendar year of life (Perrins, 1979) and are capable of producing a second and very rarely, a third brood in a season. The analysed dataset consists of 5892 records of 3257 females breeding in 61 years from 1955 to 2015. Unlike the previous studies on the same population (e.g. Reed, Jenouvrier & Visser, 2013), we kept the data from the 1991 breeding season when a late frost led to a very late caterpillar food peak (Visser, Noordwijk, Tinbergen, & Lessells 1998) and we expected a very late optimum estimate for this breeding season. Laying dates are presented as the number of days after March 31 (day  $1 = April\ 1$ , day  $31 = May\ 1$ ). The number of fledglings for each visited brood

was counted and the mother of each brood was identified (3257 breeding mothers in our analysed data). The average number of breeding records per known female was 1.81. See Supporting Information for more details on study population and fieldwork procedures.

#### 2.2 | Model formulation

We formulated a statistical model that takes into account temporally fluctuated stabilizing selection and used laying date as the focal trait that selection operates on. We also considered alternative models assuming fluctuated directional selection. We take the number of individuals surviving to fledglings as the measure of fecundity component of fitness and it is assumed to follow a zero-inflated Poisson distribution instead of a Poisson distribution due to the high probability of clutch failure (around 15.7% in our analysed dataset, clutch failure in this study means that no single chick survived to fledgling). In addition, previous studies showed (e.g. Reed et al., 2013; Townsend et al., 2013) that the relative contribution to fitness from each brood, at individual level, is determined by the food abundance at the time each brood is raised. We therefore assume that the expected number of fledglings and the probability of clutch failure potentially depends on laying dates in the same way for first, replacement (first broods failed) and second (first broods succeeded) broods via the same Gaussian fitness function. We present our approach using selection on the number of fledglings, but it can be applied to any selection episodes, such as viability, fertility selection or overall selection through lifetime fitness for species with non-overlapping generations.

We assume that the number of fledglings  $Y_i$  (i = 1,2,...,n) from the ith brood follow a zero-inflated Poisson (ZIP) distribution. Such random variables can be represented as a product  $Y_i = I_i X_i$  where

$$I_i|p_i \sim \text{Bernoulli}(1-p_i),$$
 
$$X_i|I_i=1, w_i \sim \text{Poisson}(w_i).$$
 (1)

Here,  $p_i$  is the probability of zero-inflation (complete brood failure),  $w_i$  is the Poisson mean and i is the index for all of the broods in our analysed dataset, i = 1,2,...,5892. Using the law of total expectation, the overall fitness contribution from brood i is then

$$E(Y_i|p_i, w_i) = E(X_i I_i|p_i, w_i)$$

$$= E(X_i I_i|p_i, w_i, I_i = 1)P(I_i = 1) + E(X_i I_i|p_i, w_i, I_i = 0)P(I_i = 0) (2)$$

$$= E(X_i I_i = 1, w_i)E(I_i|p_i).$$

The decomposition of the left-hand side into the two factors on the right-hand side shows that the zero-inflation part  $I_i$  can be interpreted as a separate selection episode, which we refer to as episode P for short in this study. Similarly, the Poisson part  $X_i$  is referred to as episode W in the rest of this paper.

We consider two selection modes: fluctuating stabilizing selection and fluctuating directional selection. In the fluctuating stabilizing selection model, the zero-inflation probability  $p_i$  and the Poisson mean  $w_i$  are determined by the same process, driven by deviation from the optimal onset of breeding. In addition, we assume that  $p_i$ 

is linked to covariates of interest via a logit link function while for  $w_i$  via a log link function. Therefore,  $logit(1-p_i)$  and  $ln w_i$  are given by models of the same form:

$$logit(1-p_i) = \eta_{p,t}^{(\alpha)} - \frac{(z_i - \eta_{p,t}^{(\theta)})^2}{2(e^{\eta_{p,t}^{(\theta)}})^2} + \tau_p^m \epsilon_{j(i)};$$
(3)

and

$$\ln w_{i} = \eta_{w,t}^{(\alpha)} - \frac{(z_{i} - \eta_{w,t}^{(\theta)})^{2}}{2(e^{\eta_{w,t}^{(\omega)}})^{2}} + \tau_{w}^{m} \epsilon_{j(i)}. \tag{4}$$

Here,  $\eta_{p,t}^{(\alpha)}$ ,  $\eta_{p,t}^{(\theta)}$  and  $\eta_{p,t}^{(\omega)}$ , t=1,2,...,61 are parameters determining maximum fitness (indicated by superscript  $\alpha$ ), optimal laying dates (indicated by  $\theta$ ) and widths of fitness function (indicated by  $\omega$ ) of brood i in year t respectively for  $\log it(1-p_i)$ . Similar explanations apply to the equation of  $\ln w_i$ . The variable  $z_i$  is the laying date of the ith brood. The term  $\epsilon_{j(i)} \sim N(0,1)$ , j=1,2,...,J (where J is total number of unique females) is a random effect included to model extra variation between the mothers and assumed to be same for the two episodes, but the magnitude of the effects on the two episodes are potentially different, subscript p, w thereby allow standard deviations of mother effect  $\tau_{n,l}^{m}$ ,  $\tau_{w}^{m}$  to differ between episode P and W.

The maximum fitness, optimal laying date and width of fitness function in the two episodes are assumed to have a constant difference  $c_a$ ,  $c_\theta$  and  $c_\omega$  ( $\eta_{w,t}^{(a)} = \eta_{p,t}^{(a)} + c_a$ ,  $\eta_{w,t}^{(\theta)} = \eta_{p,t}^{(\theta)} + c_\theta$  and  $\eta_{w,t}^{(\omega)} = \eta_{p,t}^{(\omega)} + c_\omega$ ) and we therefore model  $\eta_{p,t}^{(a)}$ ,  $\eta_{p,t}^{(\theta)}$  and  $\eta_{p,t}^{(\omega)}$  by the three stochastic processes

$$\eta_{s,t}^{(\alpha)} = \bar{\alpha}_s + \alpha_t, 
\eta_{s,t}^{(\theta)} = \bar{\theta}_s + \theta_t, 
\eta_{s,t}^{(\omega)} = \bar{\omega}_s + \omega_t,$$
(5)

where index s takes values from (P,W) indicating the two episodes respectively. Parameters  $\bar{\alpha}_s$ ,  $\bar{\theta}_s$ ,  $\bar{\omega}_s$  are the means of the three processes. More assumptions in terms of stochastic processes  $\alpha_t$ ,  $\theta_t$ ,  $\omega_t$  are made. They are assumed to follow a first-order vector autoregressive VAR(1) process

$$\begin{bmatrix} \alpha_t \\ \theta_t \\ \omega_t \end{bmatrix} = \mathbf{\Phi} \begin{bmatrix} \alpha_{t-1} \\ \theta_{t-1} \\ \omega_{t-1} \end{bmatrix} + \mathbf{w}_t, \tag{6}$$

where  $\Phi$  is a 3 × 3 matrix of autoregressive coefficients and  $\mathbf{w}_t$  is multivariate normal zero-mean white noise with variance-covariance matrix  $\Sigma$ . Correlation between  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  are determined through off-diagonal entries in both  $\Sigma$  and  $\Phi$ . Possible model alternatives are obtained by making  $\Phi$  and  $\Sigma$  both diagonal, such that  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  simplify to independent AR(1) processes. If all entries of  $\Phi$  are zero,  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  are independent and identically distributed white noise processes. Alternatively, we model each episode as fluctuating directional selection, which can be described by a GLMM with annual random intercept and slope:

$$logit(1-p_i) = \beta_p^{(0)} + u_{p,t}^{(0)} + (\beta_p^{(1)} + u_{p,t}^{(1)})z_i + \tau_p^m \epsilon_{j(i)};$$

$$ln w_i = \beta_w^{(0)} + u_{w,t}^{(0)} + (\beta_w^{(1)} + u_{w,t}^{(1)})z_i + \tau_m^m \epsilon_{j(i)}.$$
(7)

In this model,  $\beta_p^{(0)}$  and  $\beta_p^{(1)}$  are fixed intercept and slope respectively for episode P, random intercepts and slopes are denoted by  $u_{p,t}^{(0)}$  and  $u_{p,t}^{(1)}$ , which account for the variation among years. These random effects are assumed to be multivariate normal:

1404

$$\left(\begin{matrix} u_{p,t}^{(0)} \\ u_{p,t}^{(1)} \end{matrix}\right) \sim N\left(\mathbf{0}, \left(\begin{matrix} (\sigma_{0,p})^2 & \rho_p\sigma_{0,p}\sigma_{1,p} \\ \rho_p\sigma_{0,p}\sigma_{1,p} & (\sigma_{1,p})^2 \end{matrix}\right)\right).$$

Similar explanation applies to the alternative model for episode W (ln  $w_i$ ). As before,  $z_i$ ,  $\tau_p^m$ ,  $\tau_w^m$  and  $\epsilon_{j(i)}$  have same interpretations as that in Equations 3 and 4.

Since our statistical method relies on model selection, the candidate models we tested include different assumptions for  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$ , or different selection patterns for episode P and W, among many others.

### 2.3 | Model selection and inference

All model alternatives were implemented using R package TMB. Briefly, based on a C++ function computing the joint density of the observed data and unobserved random effects, TMB computes the Laplace approximation of the marginal likelihood of the observed data. This is then maximized numerically to obtain maximum likelihood estimates of model parameters and approximate standard errors based on information theory.

We fitted in total 43 different alternative models. Among the candidate models, each selection episode P and W maybe equipped with either directional selection or stabilizing selection. For the directional selection mode, we tested models with only fixed effects, with random intercepts and with both random intercepts and random slopes. For the stabilizing selection mode, the fitness parameters  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  were either considered as constant, as three independent AR(1) processes, as jointly following a VAR(1) process, or combinations of them.

Our model selection relies on the measurement of data support for the different models which vary in degree of complexity. We use Akaike information criterion (AIC; Akaike, 1973) based on the observed Fisher information as a model selection criterion (see Burnham & Anderson, 2003 for more details about AIC). The model with lowest AIC value was selected as the best model and the estimates of all parameters together with their approximate standard errors were obtained. All the source code of this study are archived and accessible online.

# 3 | CASE STUDY RESULTS AND DISCUSSION

#### 3.1 | Model selection procedure

As introduced in section 2.3, in total, 43 candidate models were tested. For brevity, only the selected model and its neighbour models are listed in Table 1. The model numbering is consistent with the model updating sequence in our R code. Updating procedure from

model 1 to model 8b can be found in Supporting Information. Based on the best model selected (model 9), the differences of AIC value for each model from the selected model are calculated and listed in column  $\Delta$ AIC, along with the difference in the number of parameters ( $\Delta$ p). Model 9 with directional selection in episode *P* and stabilizing selection in episode *W* is the best model.

Model 10 with directional selection via both episode P and W does not improve the model fit. To guarantee that model 9 is indeed the best one among all the candidate models, model 11 to model 14i are neighbour models updated around model 9 for comparison purpose, but none of them improves the model fit. It is worth noting that the performance of model 14g with fixed  $\omega_t$  is only slightly worse than our selected model, implying a constant  $\omega_t$  assumption in our study would be reasonable. The estimates of parameters from the selected model (model 9) and from model 14g with constant  $\omega_t$  are listed in the Supporting Information for comparison.

We also carried out a simulation study (see Supporting Information) to explore the power of our model selection technique in identifying our best model especially against model 14g and 14i. We concluded from the simulation study that our model selection technique has around 80% probability to distinguish the model with fixed  $\omega_t$  from the one with random  $\omega_t$  when the variation scale of  $\omega_t$ being 0.2. This further implies that model 14g might be as good as our selected model. The simulation study also showed that a weak mother effect (e.g. the standard deviation of random mother effects is 0.05) is hard to detect. However, since our selected model reports 6.44 lower AIC values with the estimate of standard deviation of mother effects being 0.041 in episode W, we have confidence in the mother effects in the underlying 'true' model. The remaining challenge is that there is no strong evidence for model 9 outperforms model 13, we thus should be cautious when interpreting estimates of correlations between the errors for  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$ .

The selected model (model 9) has stabilizing selection via episode W, directional selection via P with annual correlated random intercepts and slopes given by

$$logit(1-p_{i}) = \beta_{p}^{(0)} + u_{p,t}^{(0)} + \left(\beta_{p}^{(1)} + u_{p,t}^{(1)}\right) z_{i} + \tau_{p}^{m} \epsilon_{j(i)},$$

$$ln w_{i} = \eta_{w,t(i)}^{(a)} - \frac{\left(z_{i} - \eta_{w,t(i)}^{(\theta)}\right)^{2}}{2\left(e^{\eta_{w,t(i)}^{(\phi)}}\right)^{2}} + \tau_{w}^{m} \epsilon_{j(i)}.$$
(8)

Furthermore, the selected model supports VAR(1) process of  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  in the episode W. However, the three processes are correlated through errors instead of the transition matrix  $\Phi$ . Mother effects are significant in both episodes. More details about the parameter estimates are given in next section 3.2.

# 3.2 | Directional selection via probability of clutch failure

Our selected model (Equation 8) indicates directional selection via episode *P* with annual random intercept  $(u_{n,t}^{(0)})$  and slope  $(u_{n,t}^{(1)})$ . The estimates

**TABLE 1** A part of model selection procedure of phenotypic selection on breeding time of great tits. The order of models listed below is accordance with the order of models fitting, from model 9 to 14i.  $\triangle$ AIC and  $\triangle$ p is the difference in AIC and number of parameters p between each model and the best model (model 9). The column of description gives the details of the updated model based on the previous ones. For simplification, the probability of successful brooding is denoted as episode P and the mean number of fledglings episode as W. The updating procedure from model 1 to model 8a can be found in Supporting Info

Model	ΔΑΙC	Δр	Description		
Directional selection via episode P and stabilizing selection via episode W					
9	0	0	The selected model formulated as Equation 8		
Directional selection for both episode P and W					
10	104.11	-5	The model formulated as Equation 7 with correlated random intercepts and slopes		
Model 9 is the best model so far, test neighbour models with minor changes based on model 9					
11	4.075	7	Add all entries of $\Phi$ back		
12	14.32	-3	Keep only significant entries in $\Phi$ and significant correlations between the errors for $a_t, \theta_t$ and $\omega_t$		
13	0.54	-2	Keep only significant correlations between the errors for $lpha_{ m t},  heta_{ m t}$ and $\omega_{ m t}$		

Model 9 is still the best model so far, test models with all possible specifications for  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$ , and remove mother effect from each episode

14 210.47 -6 Change random $\alpha_t$ into fixed, $\theta_t$ and $\omega_t$ are random 14a 203.98 -4 Change random $\alpha_t$ into fixed, $\theta_t$ and $\omega_t$ are AR(1) 14b 96.99 -6 Change random $\theta_t$ into fixed, $\alpha_t$ and $\omega_t$ are random 14c 85.47 -4 Change random $\theta_t$ into fixed, $\alpha_t$ and $\omega_t$ are AR(1) 14d 24.84 -4 Change random $\omega_t$ into fixed, $\alpha_t$ and $\theta_t$ are AR(1) 14e 28.64 -2 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ 14f 7.1 -1 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode $P$ 14i 6.44 -1 Remove mother effect from episode $W$	una remo	and remove mother effect from each episode					
14b 96.99 -6 Change random $\theta_t$ into fixed, $\alpha_t$ and $\omega_t$ are random 14c 85.47 -4 Change random $\theta_t$ into fixed, $\alpha_t$ and $\omega_t$ are AR(1) 14d 24.84 -4 Change random $\omega_t$ into fixed, $\alpha_t$ and $\theta_t$ are AR(1) 14e 28.64 -2 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ 14f 7.1 -1 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode P	14	210.47	-6	Change random $\mathbf{\alpha}_{\mathrm{t}}$ into fixed, $\mathbf{\theta}_{\mathrm{t}}$ and $\mathbf{\omega}_{\mathrm{t}}$ are random			
14c 85.47 -4 Change random $\theta_t$ into fixed, $\alpha_t$ and $\omega_t$ are AR(1)  14d 24.84 -4 Change random $\omega_t$ into fixed, $\alpha_t$ and $\theta_t$ are AR(1)  14e 28.64 -2 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ 14f 7.1 -1 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode P	14a	203.98	-4	Change random $a_{t}$ into fixed, $\theta_{t}$ and $\omega_{t}$ are AR(1)			
14d 24.84 -4 Change random $\omega_t$ into fixed, $\alpha_t$ and $\theta_t$ are AR(1)  14e 28.64 -2 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ 14f 7.1 -1 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode P	14b	96.99	-6	Change random $\theta_{\mathrm{t}}$ into fixed, $\alpha_{\mathrm{t}}$ and $\omega_{\mathrm{t}}$ are random			
14e 28.64 -2 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ 14f 7.1 -1 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode $P$	14c	85.47	-4	Change random $\theta_{t}$ into fixed, $\alpha_{t}$ and $\omega_{t}$ are AR(1)			
14f 7.1 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode P	14d	24.84	-4	Change random $\omega_{t}$ into fixed, $\alpha_{t}$ and $\theta_{t}$ are AR(1)			
tion to the errors of $\alpha_t$ and $\theta_t$ 14g 3.27 -3 Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode P	14e	28.64	-2	Change random $\omega_{\rm t}$ into fixed, VAR(1) $\alpha_{\rm t}$ and $\theta_{\rm t}$			
cant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$ 14h 21.1 -1 Remove mother effect from episode P	14f	7.1	-1				
2	14g	3.27	-3	Change random $\omega_t$ into fixed, VAR(1) $\alpha_t$ and $\theta_t$ with significant entries in $\Phi$ , add correlation to the errors of $\alpha_t$ and $\theta_t$			
14i 6.44 –1 Remove mother effect from episode W	14h	21.1	-1	Remove mother effect from episode P			
	14i	6.44	-1	Remove mother effect from episode $\it{W}$			

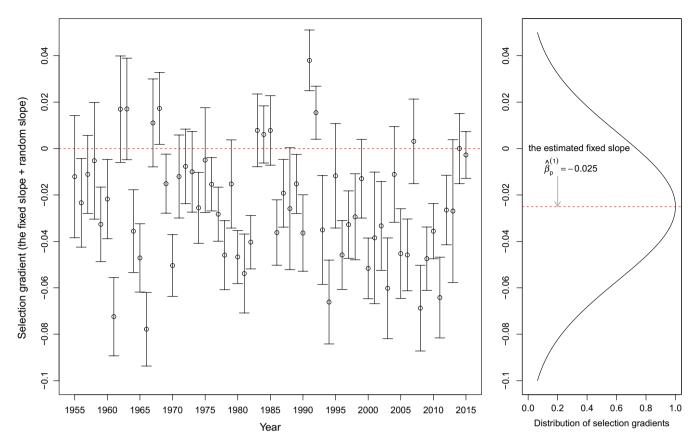
**TABLE 2** Estimates (standard errors) and corresponding 95% confidence intervals of model parameters from the selected model (i.e. model 9 of Table 1, only for selection via the probability of successful brooding)

Parameter	Meaning	Estimate (SE)	95% CI
$\beta_p^{(0)}$	Fixed intercept	2.946 (0.220)	(2.515, 3.377)
$eta_p^{(1)}$	Fixed slope	-0.025 (0.005)	(-0.035, -0.015)
$\sigma_{1,p}$	SD of ran- dom slopes	0.032 (0.004)	(0.024, 0.040)
$ au_p^m$	SD of mother effect	0.701 (0.092)	(0.520, 0.881)
$ ho_p$	Correlation between random intercepts and slopes	-0.827 (0.054)	(-0.933, -0.720)

of parameters of our interest are listed in Table 2. We estimated the fixed slope, the mean of the annual selection gradient to  $\hat{\rho}_p^{(1)} = -0.025$  (red dashed line on the right panel of Figure 1). Given a standard deviation of the random slopes estimated to  $\hat{\sigma}_{1,p} = 0.032$ , corresponding to selection for early laying 78% of the time, the distribution of selection gradients is shown with the black line in the right panel of Figure 1,

which implies that over 22% of the study years experienced positive selection, therefore, favoured late broods. The left panel of Figure 1 shows the annual selection gradient together with error bars representing  $\pm$  one standard error. The selection favoured early broods in 82% (note that the 78% is obtained with selection distribution while 82% with temporal estimated selection) of the study years, as can be seen from the left panel that most of the selection gradients fall below 0.

This result agrees with the finding from Reed et al. (2013) that females that breed late relative to the food peak (influenced by temperature, see Visser, Holleman and Gienapp (2006)) were more likely to fail to raise any fledglings. Perrins (1965) states that there is a higher proportion of predation in the later part of the season and the young of the later broods are more vulnerable to the predators since the young in the later broods are more noisy and lighter. Maziarz, Wesolowski, Hebda, Cholewa and Broughton (2016) shows that nest losses are mostly due to predation (69% nest failures of a great tit population in Poland) and the risk of nest failure varied with nest cavity attributes. To explore which biotic and abiotic factors best explain the sign and variation in annual directional selection via the probability of successful brooding, more data information concerning these factors are required and this would be one among other interesting expansions of this study. In this selection episode, mother effects contribute to explaining the variation of successful-brooding probabilities and the estimate of the standard deviation  $\tau_p^m$  is 0.701, as shown in Table 2.



**FIGURE 1** Annual directional selection gradient (left panel, defined as the sum of the fixed slope and annual random slope) associated with error bars representing one standard error and the distribution of it (right panel) for episode *P* (selection on laying date via the probability of successful brooding). The red dashed line on the left plot is an indication of 0 and on the right plot is estimated fixed slope (-0.025)

# 3.3 | Stabilizing selection via the mean number of fledglings

1406

Our selected model indicates that stabilizing selection acts via the mean of the Poisson component. Parameter estimates of the Gaussian fitness function in Equation 4, and the estimates for the parameters involved in the VAR(1)  $\alpha_{\rm t}$ ,  $\theta_{\rm t}$  and  $\omega_{\rm t}$ , along with their confidence intervals are shown in Table 3.

The estimates of the mean of maximum fitness  $(\bar{a})$  and optimum  $(\bar{\theta})$  are 2 (exponent with base e approximates to 7 fledglings) and 18.227 (approximately 18th of April) respectively. Our estimate for the width of the fitness function is much wider than that from Chevin et al. (2015) (47.395 vs. 24.11 days), in which the sum of the fledglings from multiple broods instead of the single brood was treated as the sample unit and the lay dates of only first broods (with a much narrower range) were used. We doubt that the distribution of this summation of multiple broods is well approximated by a Gaussian function and therefore we modelled the number of fledglings from each brood separately, and the second broods were laid in the late breeding season and this might be the reason of a wider fitness function being estimated with our selected model.

The estimates of the standard deviation of  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  are 0.176, 21.180 and 0.205 respectively. The estimate of standard deviation

for  $\theta_t$  is slightly larger than that from Chevin et al. (2015) (21.18 vs. 11.3 days) and this might partly result from the different datasets we used. In Chevin et al. (2015), the data before 1973 were excluded from their analysis and we therefore also fit the selected model with data only after 1973 for a fairer comparison. It turned out that the estimates with both full and partial datasets are quite close, while the estimates with full data generally have less uncertainty (narrower %95 confidence intervals). The detailed comparison can be found in Supporting information. The estimated variance in  $\omega_t$  (0.042) translates to a coefficient of variation for  $e^{\eta_{w,t}^{(\omega)}}$  of  $\sqrt{e^{0.042}-1}=0.207$ , that is, quite large fluctuation in the width of the fitness function. When conducting model selection we fitted a model with fixed  $\omega_t$  (model 14g in Table 1) over study period, however, it turned out the model fit did not improve much when  $\omega_t$  is taken random as in our selected model. In addition, by comparing the standard deviations of parameter estimates from the models with fixed and random  $\omega_t$  reported in Supporting Information, we find that uncertainties of parameter estimates are comparable. These imply that the whole analysis would not change much if in our study the constant  $\omega_{\rm t}$  assumption was made.

The autocorrelation estimates of  $\alpha_t$  ( $\hat{\phi}_{\alpha,\alpha}$ ) and  $\theta_t$  ( $\hat{\phi}_{\theta,\theta}$ ) are 0.334 and 0.524, respectively, but  $\omega_t$  is not autocorrelated in our selected model. The estimate of  $\phi_{\theta,\theta}$  in Chevin et al. (2015) was 0.2472 with a wide 95% confidence interval (-0.1745, 0.626). While our selected

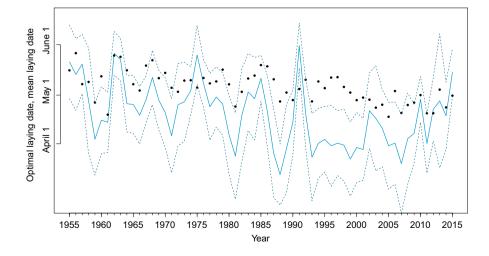
**TABLE 3** Estimates(standard errors) and corresponding 95% confidence intervals of model parameters with the selected model (only for selection via the mean number of fledglings)

Parameter	Meaning	Estimate (SE)	95% CI
$\bar{\alpha}$	Mean of process $\eta_{w,t}^{(lpha)}$	2.000 (0.036)	(1.929, 2.071)
$ar{ heta}$	Mean of process $\eta_{w,t}^{( heta)}$	18.227 (5.826)	(6.808, 29.647)
e <sup>@</sup> (days)	Mean of process $e^{\eta_{w,t}^{(\omega)}}$	47.395 (3.234)	(41.056, 53.734)
$\gamma_{\alpha,\alpha}$	SD of $\alpha_{t}$	0.176 (0.024)	(0.129, 0.224)
$\gamma_{ heta, heta}$	SD of $\theta_t$	21.180 (3.422)	(14.473, 27.888)
$\gamma_{\omega,\omega}$	SD of $\omega_{t}$	0.205 (0.049)	(0.110, 0.300)
$\phi_{lpha,lpha}$	Autocorrelation of $\alpha_t$	0.334 (0.122)	(0.094, 0.574)
$\phi_{ heta, heta}$	Autocorrelation of $\theta_{t}$	0.524 (0.110)	(0.310, 0.739)
$\sigma_{lpha}$	SD of errors of $\alpha_t$	0.166 (0.023)	(0.120, 0.212)
$\sigma_{ heta}$	SD of errors of $\theta_{\rm t}$	18.034 (2.808)	(12.531, 23.538)
$\sigma_{\omega}$	SD of errors of $\omega_{ m t}$	0.205 (0.049)	(0.110, 0.300)
$ ho_{lpha, heta}$	Correlation between the errors of $\alpha_t$ and $\theta_t$	-0.592 (0.125)	(-0.837, -0.346)
$ ho_{lpha,\omega}$	Correlation between the errors of $\alpha_t$ and $\omega_t$	-0.357 (0.287)	(-0.920, 0.206)
$ ho_{ heta,\omega}$	Correlation between the errors of $\theta_t$ and $\omega_t$	-0.307 (0.254)	(-0.806, 0.191)
$ au_w^m$	SD of mother effect	0.041 (0.013)	(0.015, 0.066)

model reported a significant and larger estimate of  $\phi_{\theta,\theta}$  with narrower confidence interval (0.310, 0.739). With the result from the simulation study, even this larger estimate may be potentially underestimated. The estimates of the standard deviations of errors of VAR(1)  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  ( $\sigma_\alpha$ ,  $\sigma_\theta$ ,  $\sigma_\omega$ ) are also listed in Table 3, along with estimates of correlations of the correlated noises. Even though our result indicates that the VAR(1) stochastic processes  $\alpha_t$ ,  $\theta_t$  and  $\omega_t$  are correlated through errors  $\mathbf{w}_t$  not through transition matrix  $\mathbf{\Phi}$ , we are conservative in interpreting the estimates of  $\rho_{\alpha,\omega}$  and  $\rho_{\theta,\omega}$  since the candidate model 13 with only  $\rho_{\alpha,\theta}$  reported almost the same AIC value as our best model. The estimate of standard deviation of mother effect is 0.041, implying that the mean of  $X_i$  from broods produced by the same mother are weakly correlated to each other.

The estimated optimum phenotype is shown in Figure 2 with a solid blue line. It fluctuates over the study period with an obvious downward trend. The mean within-year laying dates (denoted with black dots) also show a downward trend but the advance is not as strong as the optimum, resulting in increasing mistiming between the optimal laying date and the mean within-year laying date, which is in line with the finding from previous study of the same population (e.g. Chevin et al., 2015; Reed et al., 2013; Visser & Both, 2005; Visser et al., 1998). Since the reproductive fitness of the great tits depends strongly on the mismatch with food phenology, mistiming in our case therefore equals mismatch, even though the food resource phenology is not considered in our study (see Visser and Gienapp (2019) for the difference between mistiming and mismatch). One explanation for the mismatch is that females might be unwilling to breed at the optimal date in terms of the offspring fitness because of higher energetic cost of producing and incubating earlier in harsh environment where it is cold and food is scarce, mismatching by a few days might therefore be optimal for the sake of parental fitness (Te Marvelde, Webber, Meijer & Visser, 2011). Beside this optimal mismatch hypothesis, another leading explanation (the cues hypothesis) is that the cues used for timing laying are no longer accurately predicting the phenology of the food peak (see Visser et al., 2012 for more details on these two hypotheses).

**FIGURE 2** Position of optimal laying date over study period from 1955 to 2015. The estimated movement of optimal laying date from the selected model is shown with solid blue line, along with its 95% confidence interval (dashed blue lines). The black dots are the observed within-year mean laying dates

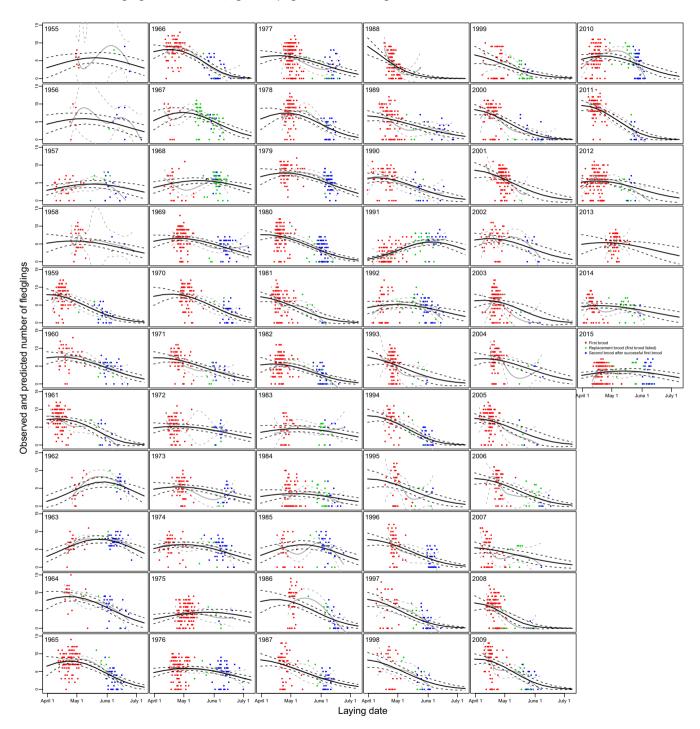


## 3.4 | Model evaluation

1408

The performance of our selected model is evaluated by visualizing the observed and predicted number of fledglings for each year. Each panel in Figure 3 shows the observed (dots in the panels) and predicted number of fledglings (dark solid line) against laying date for

a specified year (from 1955 to 2015). Our analysed data includes three brood types represented by three colours in the plots. The red, green and blue dots correspond to first, replacement and second broods respectively. The solid grey curve represents nonparametric loess regression through the points with the dashed grey lines being associated 95% confidence band. With our selected model,



**FIGURE 3** Observed and predicted number of fledglings ( $E(Y_i|p_i,w_i)$ ) against the laying date for each year. The blue, red and green dots represent the observed number of fledglings from first, replacement and second broods respectively. The grey curve is loess regression (with default degree of smoothing = 0.75) through the scatter points with 95% confidence band (dashed grey lines). The black line indicates the number of fledglings predicted with our selected model conditional on zero mother random effects with dashed black lines representing its 95% confidence band. The 95% confidence band was calculated by multiplying the standard errors reported with TMB by the 2.5 and 97.5th percentiles of the normal distribution

the dark curve shows the predicted number of fledglings at laying dates conditional on zero mother effect over the whole breeding season, with the dashed dark lines representing associated 95% confidence band again. The figure indicates a good fit of our selected model to the data as we can see that the dark lines lie within the loess confidence bands for all the years. For most of the years, the prediction of number of fledglings peaked at early breeding season when the first broods were laid except year 1991, when a late frost hit the population and the plot validates our expectation of a very late optimum estimate.

# 4 | CONCLUSION AND POSSIBLE EXTENSIONS

Thanks to the new techniques such as TMB for fast likelihood computation for non-Gaussian and nonlinear models, the use of state-space models for analysing ecological systems is increasing (for example Cadigan, 2015; Albertsen, Nielsen & Thygesen, 2016; Auger-Méthé et al., 2017). The conditional independence structure in state-space models yield a sparse precision matrix for the joint distribution of the data and the random effects (Kristensen et al., 2015) and TMB takes maximal advantage of this sparseness structure (through automatic sparsity detection) in its computation of the Laplace approximation. Therefore, using state-spaces models coupled with TMB makes estimating a large number of parameters and random effects which is usually the case in modelling complicated biological processes or ecological systems, possible. Compared with the models and approaches adopted by previous studies on fluctuating selection, our method based on SSM, GLMM and TMB has several advantages. First, state-space models allow us to explore two correlated fitness components simultaneously, instead of measuring different fitness components independently. Second, due to the flexibility of SSMs, parameters can be estimated efficiently with little computational effort. Third, the formulation of our theoretical models turns out to be more realistic to account for directional selection and non-Gaussian fitness residual, as GLMMs. Our results from the great tit case study partly agree with the findings from previous studies on the same population, and due to the VAR(1) formulation for the fitness parameters we could gain more in terms of the underlying patterns of the fluctuating selection. For the researchers who are interested in applying our method to their data either for modelling fluctuating selection or general ecological systems with VAR(1) stochastic processes, it is worth to mention that TMB has no built-in probability function for modelling VAR(1). Our study can serve as a template for this as well as for conducting model selection with TMB.

In our study, we treated fluctuations in properties of the Gaussian fitness function as a vector autoregressive process. In principle, our approach can also accommodate other autocovariance structures, such as vector autoregressive moving-average (ARMA) models (see Wei, 2006 for the definition). Besides, in our statistical model, the random mother effect  $\epsilon_{i(i)}$  is assumed to be same for the two episodes

but vary in magnitude, which implies that a mother that is likely to have complete brood failure will be more likely to have a low number of fledglings (with correlation 1). To relax this assumption, the mother effects can be treated differently for the two episodes and assumed to be multivariate Gaussian distributed  $(\epsilon_{p,j(i)}, \epsilon_{w,j(i)})^T \sim N(0,\Sigma)$ with  $\Sigma$  being a covariance matrix. Furthermore, the number of fledglings is chosen as the selection component so that the estimates could be compared to those from Chevin et al. (2015), which claims that using the number of recruits may cause more uncertainty in estimates of parameters due to the much larger coefficient of variation in the number of recruits. However, in some previous studies (e.g. Reed et al., 2013), the fecundity component of fitness is measured as the number of recruits surviving to the next breeding season instead of the number of individuals surviving to become fledglings. As claimed by Naef-Daenzer and Grüebler (2016), using the number of fledglings as a proxy for fitness may be misleading in inference of evolutionary significance since reproductive success can be completely altered by many causal factors driving the adaptations which operate during the post-fledgling period, and thereby change the juveniles' fate from fledgling to independence. Therefore, it would be interesting to expand our model to incorporate both pre- and postfledgling period, such as chicks' survival and recruitment probability, as well as a female's survival, into a comprehensive life-history framework for the lifetime selection exploration.

Our study demonstrates a technique of estimating fluctuating selection in cases where ecological covariates are not available. To understand whether observed shift in selection are biologically meaningful, however, it is important to elucidate the ecological drivers of fluctuations in selection. Empirical investigations of the causal mechanisms driving such selection dynamism are needed before the development of novel analytical and statistical techniques. In our great tit case, for example, the peak movement might be affected by the height of the caterpillar peak, the mean breeding timing relative to the caterpillar peak and the breeding density. The width of the fitness function is likely being affected by the height and probably the width of the caterpillar peak (Visser et al., 2006). The location of optimum might be influenced by environmental variables (e.g. Chevin et al., 2015; Gienapp et al., 2013). However, the biotic interactions coupled with other abiotic factors playing a direct or indirect role in the selective process could likely make analysis much more complicated. Other extensions include analysis of correlational selection on multivariate traits and estimating the temporal-spatial variation and correlation in fluctuating selection.

#### **ACKNOWLEDGEMENTS**

This work was supported by the Research Council of Norway through its Centres of Excellence funding scheme (project number 223257 to CBD).

### **AUTHORS CONTRIBUTIONS**

M.E.V. provided the data; J.T. conceived the idea and initiated the statistical model; Y.C. analysed the data and conducted the analyses;

Y.C. wrote the initial draft with input from J.T.; all authors contributed to revisions on later manuscript versions and gave final approval for publication.

#### DATA AVAILABILITY STATEMENT

All the necessary data and source code to carry out the analyses in this study are available in the Dryad Digital Repository https://doi.org/10.5061/dryad.q4q8r89 (Cao, Visser & Tufto, 2019).

#### REFERENCES

1410

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 2nd international symposium on information theory, Akademiai Kiado, Budapest, 1973 (pp. 267–281).
- Albertsen, C. M., Nielsen, A., & Thygesen, U. H. (2016). Choosing the observational likelihood in state-space stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(5), 779–789.
- Auger-Méthé, M., Albertsen, C. M., Jonsen, I. D., Derocher, A. E., Lidgard, D. C., Studholme, K. R., ... Flemming, J. M. (2017). Spatiotemporal modelling of marine movement data using template model builder (TMB). Marine Ecology Progress Series, 565, 237–249.
- Barton, N. H., & Keightley, P. D. (2002). Multifactorial genetics: understanding quantitative genetic variation. *Nature Reviews Genetics*, 3(1), 11.
- Bell, G. (2010). Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537), 87–97.
- Bull, J. (1987). Evolution of phenotypic variance. *Evolution*, 41(2), 303-315.
- Burnham, K. P., & Anderson, D. R. (2003). Model selection and multimodel inference: a practical information-theoretic approach. Berlin: Springer Science & Business Media.
- Cadigan, N. G. (2015). A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(2), 296–308.
- Calsbeek, B. (2011). Exploring variation in fitness surfaces over time or space. *Evolution*, 66(4), 1126–1137.
- Cao, Y. H., Visser, M. E., & Tufto, J. (2019). Data from: A time series model for estimating temporal variation in phenotypic selection on laying dates in a Dutch great tit population. *Dryad Digital Repository*. https:// doi.org/10.5061/dryad.q4q8r89
- Chevin, L.-M., Visser, M. E., & Tufto, J. (2015). Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection. Evolution, 69(9), 2319–2332.
- Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12(1), 119–129.
- Futuyma, D. J. (2006). Evolutionary biology (3rd ed.). New York: W. H. Freeman; Basingstoke: Palgrave [distributor].
- Gamelon, M., Tufto, J., Nilsson, A. L. K., Jerstad, K., R⊘stad, O. W., Stenseth, N. C., ... Sæther, B. E. (2018). Environmental drivers of varying selective optima in a small passerine: a multivariate, multiepisodic approach. Evolution, 72(11), 2325–2342
- Gienapp, P., Lof, M., Reed, T. E., McNamara, J., Verhulst, S., & Visser, M. E. (2013). Predicting demographically sustainable rates of adaptation: can great tit breeding time keep pace with climate change? Philosophical Transactions of the Royal Society B: Biological Sciences, 368 (1610).
- Gienapp, P., Reed, T. E., Visser, M. E. (2014). Why climate change will invariably alter selection pressures on phenology. Proceedings of the Royal Society B: Biological Sciences 281(1793), 20141611.

- Hedrick, P. W. (2006). Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology, Evolution, and Systematics*, 37(1), 67–93.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. (2015). TMB: automatic differentiation and Laplace approximation. arXiv preprint arXiv:1509.00660.
- Lande, R., & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, 50(1), 434–437.
- Maziarz, M., Wesolowski, T., Hebda, G., Cholewa, M., & Broughton, R. K. (2016). Breeding success of the Great Tit Parus major in relation to attributes of natural nest cavities in a primeval forest. *Journal of Ornithology*, 157(1), 343–354.
- Morrissey, M. B., & Hadfield, J. D. (2012). Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, 66(2), 435–442.
- Naef-Daenzer, B., & Grüebler, M. U. (2016). Post-fledging survival of altricial birds: ecological deter minants and adaptation. *Journal of Field Ornithology*, 87(3), 227–250.
- Perrins, C. M. (1965). Population fluctuations and clutch-size in the Great Tit, Parus major L. *The Journal of Animal Ecology*, 601–647.
- Perrins, C. M. (1979). British tits (Vol. 62). New York, NY: HarperCollins.
- Post, E., & Forchhammer, M. C. (2007). Climate change reduces reproductive success of an arctic herbivore through trophic mismatch. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1501), 2367–2373.
- Reed, T. E., Jenouvrier, S., & Visser, M. E. (2013). Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. *Journal of Animal Ecology*, 82(1), 131–144.
- Revell, L. J. (2007). The G matrix under fluctuating correlational mutation and selection. *Evolution: International Journal of Organic Evolution*, 61(8), 1857–1872.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal Of The Royal Statistical Society Series B:* (Statistical Methodology), 71(2), 319–392.
- Scheiner, S. M. (1993). Genetics and evolution of phenotypic plasticity. Annual Review of Ecology and Systematics, 24(1), 35–68.
- Siepielski, A. M., DiBattista, J. D., & Carlson, S. M. (2009). It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, 12(11), 1261–1276.
- Svardal, H., Rueffler, C., & Hermisson, J. (2011). Comparing environmental and genetic variance as adaptive response to fluctuating selection. Evolution, 65(9), 2492–2513.
- Te Marvelde, L., Webber, S. L., Meijer, H. A. J., & Visser, M. E. (2011). Mismatched reproduction is energetically costly for chick feeding female great tits. Functional Ecology, 25(6), 1302–1308.
- Townsend, A. K., Sillett, T. S., Lany, N. K., Kaiser, S. A., Rodenhouse, N. L., Webster, M. S., Holmes, R. T. (2013). Warm springs, early lay dates, and double brooding in a North American migratory songbird, the black-throated blue warbler. *PLoS ONE*, 8(4), 59467.
- Tufto, J. (2015). Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: a quantitative genetic model. *Evolution*, 69(8), 2034–2049.
- Van Tienderen, P. H., & Koelewijn, H. P. (1994). Selection on reaction norms, genetic correlations and constraints. *Genetics Research*, 64 (2), 115–125.
- Visser, M. E., & Both, C. (2005). Shifts in phenology due to global climate change: the need for a yardstick. Proceedings of the Royal Society B: Biological Sciences, 272(1581), 2561–2569.
- Visser, M. E., & Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. *Nature Ecology and Evolution*, 3, 879–885.
- Visser, M. E., Gienapp, P., Husby, A., Morrisey, M., de la Hera, I., Pulido, F., Both, C. (2015). Effects of spring temperatures on the strength of

selection on timing of reproduction in a long-distance migratory bird.  $PLoS\ Biology,\ 13\ (4),\ 1-17.$ 

- Visser, M. E., Holleman, L. J., & Gienapp, P. (2006). Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia*, 147 (1), 164–172.
- Visser, M. E., Marvelde, L. te, & Lof, M. E. (2012). Adaptive phenological mismatches of birds and their food in a warming world. *Journal of Ornithology*, 153, 75–84.
- Visser, M. E., Noordwijk, A.V., Tinbergen, J. M., Lessells, C. M. (1998). Warmer springs lead to mistimes reproduction in Great Tits (Parus major). *Proceedings of the Royal Society B*, 265(1408), 1867–1870.
- Wei, W. (2006). Time series analysis: Univariate and multivariate methods. (2 ed.,). Boston, MA: Pearson Addison Wesley.
- Zhang, X. S., & Hill, W. G. (2005). Evolution of the environmental component of the phenotypic variance: stabilizing selection in changing environments and the cost of homogeneity. *Evolution*, 59(6), 1237–1244.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Cao Y, Visser ME, Tufto J. A time series model for estimating temporal variation in phenotypic selection on laying dates in a Dutch great tit population.

Methods Ecol Evol. 2019;10:1401–1411. https://doi.org/10.1111/2041-210X.13249