View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by NERC Open Research Archive

NERC Open Research Archive sharing our research

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

Article (refereed) - postprint

Elston, D.A.; Brewer, M.J.; Martay, B.; Johnston, A.; Henrys, P.A.; Bell, J.R.; Harrington, R.; Monteith, D.; Brereton, T.M.; Boughey, K.L.; Pearce-Higgins, J.W.. 2017. A new approach to modelling the relationship between annual population abundance indices and weather data. **Journal of Agricultural, Biological and Environmental Statistics**, 22 (4). 427-445. 10.1007/s13253-017-0287-4

© 2017 International Biometric Society

This version available http://nora.nerc.ac.uk/518514/

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at
http://nora.nerc.ac.uk/policies.html#access

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer review process. There may be differences between this and the publisher's version. You are advised to consult the publisher's version if you wish to cite from this article.

The final publication is available at Springer via
https://doi.org/10.1007/s13253-017-0287-4

Contact CEH NORA team at
noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos ('the Trademarks') are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

A new approach to modelling the relationship between annual population abundance indices and weather data

Running title: Modelling annual population abundance indices

Elston, D.A.[1*], Brewer, M.J.[1], Martay, B.[2], Johnston, A.[2], Henrys, P.A.[3], Bell, J.R.[4], Harrington, R.[4], Monteith, D.[3], Brereton, T.M.[5], Boughey, K.L.[6] & Pearce-Higgins, J.W.[2]

[1]Biomathematics and Statistics Scotland, Aberdeen, AB15 8QH, UK
[2]British Trust for Ornithology, Thetford, IP24 2PU, UK
[3]Centre for Ecology and Hydrology, Lancaster, LA1 4AP, UK
[4] Rothamsted Research, Harpenden, AL5 2JQ, UK
[5]Butterfly Conservation, Wareham BH20 5QP, UK
[6]Bat Conservation Trust, London, SE11 5RD, UK
*Corresponding author

Word count 8029 (excluding this line)

Abstract (231 words)

Weather has often been associated with fluctuations in population sizes of species, however, it can be difficult to estimate the effects satisfactorily because population size is naturally measured by annual abundance indices whilst weather varies on much shorter timescales. We describe a novel method for estimating the effects of a temporal sequence of a weather variable (such as mean temperatures from successive months) on annual species abundance indices. The model we use has a separate regression coefficient for each covariate in the temporal sequence and over-fitting is avoided by constraining the regression coefficients to lie on a curve defined by a small number of parameters. The constrained curve is the product of a periodic function, reflecting assumptions that associations with weather will vary smoothly throughout the year and tend to be repetitive across years, and an exponentially decaying term, reflecting an assumption that the weather from the most recent year will tend to have the greatest effect on the current population and that the effect of weather in previous years tends to diminish as the time lag increases. We have used this approach to model 501 species abundance indices from Great Britain, and present detailed results for two contrasting species alongside an overall impression of the results across all species. We believe this approach provides an important advance to the challenge of robustly modelling relationships between weather and species population size.

Key-words: Abundance index; climate change impacts; distributed lag models; population abundance models; population change, weather variables.

1. Introduction

It is important for ecologists to understand the relationships between species abundance and weather. These relationships matter in many applied situations, when knowledge is required of the determinants of species distributions, changes to population sizes and distributions, and more recently the likely effects of climate change, particularly for those species identified as requiring protection at global (UN convention) and continental (e.g. EU directive) scales. In many studies, spatial relationships between species abundance and weather are analysed by modelling species data from a large number of contrasting locations as a function of a few bioclimatic variables (e.g. Chen et al 2011, Bellard et al 2012, Warren et al 2013). These approaches require extensive distributional data. For many species there are not data from a enough sites across a sufficiently broad spatial extent and in such cases, longitudinal studies of population change in confined geographical areas may be used to infer potential sensitivities and responses to climate change (e.g. Martay *et al.* 2016). However, these studies present challenges to inferring impacts of weather, due to population data that are at a coarse spatial extent, often aggregated to an annual estimate of the population index. We demonstrate a novel modelling approach that can extract more information from such longitudinal population change data than previous analytical methods.

Estimation of the relationship between weather and annual species abundance indices is less straightforward than it might at first seem. Sometimes there is a clear causal link between particular weather variables and species' abundance (e.g. Pollard 1988; Roy *et al.* 2001; Pearce-Higgins *et al.* 2010), but often the nature and timing of the relationships are not known *a priori*. This means that the weather from many different seasons may affect populations, leading to the number of potential explanatory variables being large relative to the number of years for which abundance indices are available. Even if each weather variable to be considered is summarised across a time period of some suitable duration (which we take henceforth as being a calendar month), then there are still many potential combinations of types of weather variable (e.g. monthly minimum, mean and maximum temperature, frost days, precipitation) and months in the years leading up to each abundance index record. Furthermore, a species abundance index may respond differently to the same weather variable (e.g. temperature) in different months (e.g. Pollard 1988, Pearce-Higgins *et al.* 2015), creating a tension between model flexibility and the avoidance of over-specification.

The principal assumption underlying traditional approaches to modelling annual abundance index data is that any individual model can contain only a small number of covariates relative to the number of observations, often leading to construction and use of covariates at a seasonal (quarterly) level summarising individual weather variables (e.g. mean temperature) or weather patterns (e.g. winter North Atlantic Oscillation). Whilst this assumption may hold when each covariate has its own regression coefficient whose value is free to vary independently of all other regression coefficients, it does not hold if the regression coefficients can be constrained in some manner. A natural way of thinking about the problem is to consider that each index value y_t for years t=1...T contained in the vector $\boldsymbol{y}$, depends on the weather records x_{t1}, x_{t2}, x_{tM}, from M successive past months, specified as the entries in the corresponding rows of monthly weather covariates, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ up to $\boldsymbol{x}_M$, each of length T. Associated with each of these M covariates is a single covariate regression coefficient, b_m, for m=1...M. As successive covariates correspond to the weather in consecutive months, we may expect neighbouring covariate regression coefficients to be related, and we use these relationships to reduce the number of free parameters associated with the sequence of covariate regression coefficients, $b_1 ... b_M$. The approach we have taken is to specify a parametric function, $b_m=f(m,\boldsymbol{\theta})$, for the monthly covariate regression coefficients, in which the function $f(m,\boldsymbol{\theta})$ was selected to enforce what we considered to be ecologically reasonable constraints on the sequence of covariate regression coefficients through a low-dimensional vector, $\boldsymbol{\theta}$, of underlying parameters to be estimated.

Modern statistics offers a selection of alternative approaches to imposing the belief of similarity in successive values of the sequence of covariate regression coefficients, $b_1 ... b_M$. Forms of penalised signal regression using as roughness penalty the squares of the first, second or third differences of the sequence $b_1 ... b_M$ (Marx & Eilers 1999) have been used to model the effect of long sequences of weather variables on species traits other than population size (Sims *et al.* 2007; Roberts 2008; Roberts 2012). Alternatively, the problem can be treated as a particular case of a distributed lag model (e.g. Baltagi, 2008) in which the $b_1 ... b_M$ are derived from a suitably flexible series of basis functions. We shall use this latter formulation to compare our parametric specifications with a more flexible alternative.

In this paper, we begin by describing the data sets available, including how these have been used to construct national species abundance indices and associated sequences of weather covariates. We then define a parametric form $f(m,\boldsymbol{\theta})$ for the model of monthly covariate regression coefficients with what we consider to be the relevant desirable properties, presenting an approach to estimating the underlying parameters and constructing confidence intervals for both the monthly covariate regression coefficients and fitted values for the model of the abundance indices. The final methodological element considered is how to summarise the model fits across a large number of species. The methods, and their comparison with fitting a distributed lag model without such strong functional constraints, are illustrated using two contrasting species. We conclude by discussing the merits of our approach, what has been learned and how it may be improved. Throughout, the notation used omits reference to particular weather variables or species except where this is essential.

2. Materials and methods

2.1 Derivation of annual abundance indices for each species

We extracted site- and year-specific records of abundance for 501 species of butterflies, moths, aphids, birds, and mammals from the databases of various long-term monitoring schemes within Great Britain (see Martay *et al*. 2016 for details of all data sources). For each species, these data comprised estimates of abundance from individual site by year combinations, using available years between 1966 and 2011. Generalised linear models with Poisson errors and a log link function, including categorical variables for site and year, were fitted to the site-by-year abundance data for each species. The estimated year effects, y_t, from theses log-linear models for t ranging from 1 (the oldest value) to T (the most recent), were taken as the annual population indices to be modelled, adopting the convention that y_t=0 to resolve the aliasing in the full set of year-specific and site-specific parameters. These estimated annual species abundance indices are therefore on a logarithmic scale, and their construction makes best use of the available data, in particular allowing for the fact that not all sites provide records for all years (Freeman & Newson, 2008). Following existing protocols for bats (Barlow *et al*., 2015), survey-specific methodological and effort covariates known to strongly influence each abundance record (survey weather, type of bat detector, timing of survey) were included in the log-linear models leading to the derivation of the annual abundance indices adjusted for these survey-specific covariates.

2.2 Weather data

We chose to examine the effect of two key weather variables, precipitation and temperature, on the annual abundance indices, both having been found previously to affect population sizes of the species groups of interest (see, e.g., Huntley *et al*., 2007, or Johnston *et al*., 2013). Data values for these weather variables were downloaded from the UK Met Office as national-level summaries for each month of all calendar years from 1910 to 2011 (Perry & Hollis, 2005). A comparison of mean, minimum and maximum temperature values indicated that these were highly correlated: consequently the corresponding fitted models in test runs were similar and a decision was taken to use only the mean temperature. Thus models were fitted based on either precipitation or mean temperature data, using monthly total precipitation and the product of monthly mean temperature and number of days in the each month to equalise the influence of weather from individual days on the annual abundance indices. Either UK or English weather data were used as appropriate for each species, given the geographical coverage of sampling locations.

For each species, some care was needed to define the weather covariates to ensure the analyses were based only on months that could have influenced the response variable. This required the identification of a reference month, r ranging in value from r=1 for January to r=12 for December, based on knowledge of the ecology and survey protocols for the species in question, such that only the weather up to and including month r, in year t could influence the corresponding index value y_t. The weather data w_{rt} on either mean temperature or precipitation from month r in each of years t=1...T were used to construct the first monthly weather covariate, $\mathbf{x_1}$: thus $x_{1t}= w_{rt}$. Successive weather covariates $\mathbf{x_2}$ up to $\mathbf{x_M}$ were created by working backwards through the monthly weather data, one month a time: thus $x_{2t}= w_{r\text{-}1,t}$ unless r=1 in which case $x_{2t}= w_{12,t\text{-}1}$ Since the entries in successive rows of $\mathbf{x_1}$ are separated by 12 months, the values in $\mathbf{x_1}$, $\mathbf{x_2}$ up to $\mathbf{x_M}$ have a repeating structure in the sense that $x_{m+12,t} = x_{m,t\text{-}1}$. For simplicity, however, in the following sections we describe the models in terms of these derived covariates rather than the original sequence of weather data.

2.3 Relating annual species abundance indices to weather

The general approach that we have taken allows data on a single weather variable from a large number of past months to be included in the model for the abundance index of some particular species, and is as follows. Let y_t be the national index of abundance on a natural logarithmic scale from year t for t=1...T. In the absence of any weather effect, a simple model for y_t showing a general trend with errors e_t would be defined mathematically as:

$$y_t = a + ct + e_t,\ e_{1...T} \sim \text{Gaussian AR(1)}. \qquad (1)$$

In Equation 1, which defines our null model (denoted E_0) for variation in the absence of any weather effect, $(\theta_1, \theta_2)=(a, c)$ where a is some overall intercept, c is the regression coefficient of a year of survey covariate included to remove any linear trend, whether intrinsic to the population or arising from excluded effects. The error terms e_t are assumed to come from normal (Gaussian) distributions through which both here and elsewhere we allow for temporal correlation by assuming the sequence of error terms e_1 ... e_T follows a first order auto-regressive, AR(1), process.

We extend the null model by inclusion of a sequence of monthly covariates, $x_{m,t}$, as defined in Section 2.2, each with an associated covariate regression coefficient b_m, giving rise to the expression:

$$y_t = a + ct + \sum_{m=1}^{M} b_m\, x_{m,t} + e_t,\ \ e_{1...T} \sim \text{Gaussian AR(1)}. \qquad (2)$$

Rather than estimating each covariate regression coefficient b_m independently, which in general is not possible since for large M the number of covariates exceeds T, the number of observations of the response variable, we specify a functional relationship $b_m=f(m,\boldsymbol{\theta})$ for which the number of elements, P, of the underlying parameter vector $\boldsymbol{\theta}$ of length is sufficiently small to be well estimated. The function $f(m,\boldsymbol{\theta})$ itself is selected to impose two beliefs in addition to smoothness about the sequence of regression coefficients. First, the covariate regression coefficients b_m will decay towards zero as m increases due to the declining influence of weather many years prior to any given abundance estimate. Second, the covariate regression coefficients b_m will show some periodicity with respect to m, in the sense that we might expect covariate regression coefficients lagged by whole calendar years (b_m and b_{m+12}) to take the same sign, whilst allowing for potential opposing effects within a year. We enforce these two beliefs by writing $f(m,\boldsymbol{\theta})$ as the product of a decaying term, defined by a single underlying non-linear parameter, κ, which for purposes of exposition we replace by $\lambda=\exp(\kappa)/(1+\exp(\kappa))$ with $0 < \lambda < 1$, and a periodic term defined by multiple underlying linear parameters. By expressing the periodic term as a first-order Fourier series, we obtain:

$$b_m = \lambda^{m-1} (\beta_1 + \beta_2 \sin(2\pi m/12) + \beta_3 \cos(2\pi m/12)); \quad (3)$$

with the combination of (2) and (3) defining our simplest alternative model, E_1, for the covariate regression coefficients (b_m) which thus enlarges on the null model E_0 by the inclusion of four underlying parameters, $(\theta_3, \theta_4, \theta_5, \theta_6)=(\beta_1, \beta_2\ \beta_3, \lambda)$. The underlying parameter β_1 describes the average effect of the weather variable under investigation, regardless of time of year, and facilitates the possibility that all covariate regression coefficients may take the same sign. The pair of underlying parameters β_2 and β_3 define the phase and amplitude of the sine wave, and both are required to ensure that the phase of the underlying sine wave is unconstrained by the allocation of calendar months to the integers from 1 to 12. This allows the model to account for potentially contrasting positive and negative impacts of temperature and precipitation on the same population at different times of the year; for example accounting for potentially contrasting effects of summer and winter warmth within the same year upon butterfly populations (Long *et al.* 2016). Some examples of the sequences of covariate regression coefficients that can be generated from (3) are given as Supplementary Material.

This alternative model can be elaborated on in a number of ways, of which we consider two. The first elaboration, which we shall refer to as model E_2, introduces a term of period six months into the cyclic part of the underlying model, leading to covariate regression coefficients defined by the following equation:

$$b_m = \lambda^{m-1} [\beta_1 + \beta_2 \sin(2\pi m/12) + \beta_3 \cos(2\pi m/12) + \beta_4 \sin(4\pi m/12) + \beta_5 \cos(4\pi m/12)], \tag{4}$$

for which P=8 and $\boldsymbol{\theta}^T=(a, c, \beta_1, \beta_2\ \beta_3, \beta_4\ \beta_5, \lambda)$. The benefit of this first elaboration can be thought of as enabling two peaks per year in the series of covariate regression coefficients, and / or removing the constraint under model E_1 of symmetry in the periodic element either side of the peaks and troughs. This can be important, for example allowing positive effects of both spring and winter temperature upon resident bird species within the same year (Pearce-Higgins *et al.* 2015). The second elaboration, which we shall refer to as model E_3, introduces a term of period two years into the cyclic part of the underlying model, leading to covariate regression coefficients defined by the following equation:

$$b_m = \lambda^{m-1} [\beta_1 + \beta_2 \sin(2\pi m/12) + \beta_3 \cos(2\pi m/12) + \beta_4 \sin(\pi m/12) + \beta_5 \cos(\pi m/12)]. \tag{5}$$

The benefit of this second elaboration can be thought of as allowing the sign of covariate regression coefficients to alternate between years, introducing the concept of differential lagged effects, for example as may occur as a result of contrasting direct effects of spring temperature upon chick growth rates, and indirect (lagged) negative effects of temperature upon the same species, mediated through reductions in prey populations (Pearce-Higgins *et al.* 2010).

2.4 Estimation

We have fitted the four models described above, (E_0, E_1, E_2 and E_3), to annual abundance index data from all 501 available species, using as covariate sequences of either mean temperature or precipitation either over the preceding single year or over the preceding 40 years as covariates, giving a total of 12 or 480 weather covariates respectively. Estimation of the parameters in the four underlying models which defined the sequences of 480 covariate regression coefficients was implemented using the *lme*, *gls* and *nlme* functions in the *nlme* package (Pinheiro *et al.*, 2014) of the R software (R Core Team, 2015), making use of the fact that each regression model for the abundance indices can be expressed directly as a function of the underlying parameters, such as

$$y_t = a + ct + \beta_1 \sum_{m=1}^{M} \lambda^{m-1} x_{m,t} + \beta_2 \sum_{m=1}^{M} \sin(2\pi m/12)\, \lambda^{m-1} x_{m,t} + \beta_3 \sum_{m=1}^{M} \cos(2\pi m/12)\, \lambda^{m-1} x_{m,t} + e_t, \qquad (6)$$

for E_1. Such reformulations demonstrate that, conditional on the value of λ (equivalently κ), the other underlying parameters in the model for the mean value of y_t appear linearly (Ratkowsky 1983). The model for the abundance indices corresponding to $\lambda=1$ (equivalently $\kappa=\infty$) was fitted separately if required. Estimation of the underlying parameters for models E_1, E_2 and E_3 began with a grid search on values of κ from -3 to 9 in steps of 0.25, with each value of κ treated as fixed and optimising over all other underlying parameters to find the conditional maximum likelihood. The resulting profile likelihoods were not necessarily unimodal: we present below the results of subsequent optimisations for all model parameters, performed starting from the optimum arising from the grid search for κ followed by conditional optimisation of remaining parameters. To avoid over-fitting, the underlying models containing second-order Fourier terms, E_2 and E_3, were only fitted when at least 20 abundance index values were present. During the implementation of the fitting procedure, we mean-centred each covariate $\boldsymbol{x}_{\mathbf{1 \ldots M}}$ before inclusion in the model to aid interpretation and to reduce the correlation between underlying parameters describing the periodic cycle and the overall intercept *a*.

2.5 Calculation of confidence intervals

The presence of the non-linear parameter, κ, in the model for the mean value of y_t not only complicated the estimation of the vector, $\boldsymbol{\theta}$, of underlying model parameters: it also caused added complications when calculating confidence intervals for the covariate regression coefficients, b_m, and the fitted values in the model for the abundance indices, all of which are non-linear functions of κ. We found the most satisfactory way of calculating confidence intervals was to use importance sampling, implemented as a two-step procedure: firstly, sample κ according to a quadratic (Gaussian) approximation to the profile log-likelihood; secondly, conditional on κ, sample the remaining parameters $\theta_{1...P\text{-}1}$ according to the quadratic Gaussian log-likelihood for (6) defined by the conditional estimates of $\theta_{1...P\text{-}1}$ and associated covariance matrix. For each set of parameter values thus derived, the associated likelihood for (6) was calculated, and the ratios of these actual likelihoods to the product of the two importance sampling likelihoods (firstly for κ, secondly for $\theta_{1...P\text{-}1}$ given κ) were then used as importance weights. Confidence intervals for the covariate regression coefficients, b_m, and prediction intervals for the fitted values of the model of the abundance indices were formed from importance-weighted quantiles of the corresponding sets of values calculated from the underlying parameter values simulated as above. Exact details for how this was done can be seen from the *R* code provided as Supplementary Material.

2.6 Model summaries and comparisons

We assessed the evidence for increasing model complexity by using *F*-tests to examine the statistical significance of changes in deviance when adding underlying parameters to the model. For these *F*-tests the numerator degrees of freedom was taken as the number of added parameters and the denominator degrees of freedom was calculated as the difference between the number of index values and the number of parameters in the model for the mean value. This evidence is presented across species as histograms of p-values corresponding to the transitions between the different models. Histograms of estimates of λ^{12} are shown also, to indicate the annual decay rate of the covariate regression coefficients determined by the periodic part of the model.

For model E_1, we summarised the covariate regression coefficients across species by presenting standardised weighted means of the covariate regression coefficients (SWMCRC). Augmenting the above notation so that the covariate regression coefficient for covariate $\boldsymbol{x_m}$ for species i is b_{mi}, we first normalised the covariate regression coefficients for any given species by dividing by max($|b_{mi}|$, m=1...12), then weighted the normalised values by 1- p_i where p_i indicates the significance of the transition from E_0 to E_1, giving $B_{mi} = (1\text{-}p_i)b_{mi} / \max(|b_{mi}|)$. Our standardised presentation for any group S of N_S species all with the reference month, comprises two weighted means, $A^{S+}{}_m = \Sigma^{S+}B_{mi}/\ N_S$ and $A^{S-}{}_m = \Sigma^{S-}B_{mi}/\ N_S$, in which Σ^{S+} and Σ^{S-} denote summation over species i in S for which B_{mi} is positive or negative respectively. We calculate and present these for covariates m=1...12 only, since values for later months are necessarily smaller. The net result of this combination of covariate regression coefficients is to produce $A^{S+}{}_m$ and $A^{S-}{}_m$, bounded by 0 and +1 or -1 respectively, the extrema being approached only if the covariate regression coefficients for all species take their maximum absolute value in the same month, have the same sign in that month, and the transitions from model E_0 to E_1 are highly statistically significant for all species. Where species in S do not all have the same reference month, an additional step is required to align the values of B_{mi} from different species so that the values of $A^{S+}{}_m$ and $A^{S-}{}_m$ are formed for each relevant calendar month, rather than according to lags from the differing reference months, the divisor for each month then being the number of contributing species for that month rather than N_S. Statistical significance of SWMCRC was assessed using a sign test, considering the sign of each covariate regression coefficient b_{mi} to be random under the null hypothesis of no effects of weather.

2.7 Comparison with a less constrained alternative model

We compare the sequences of covariate regression coefficients, $b_1...b_M$.with their counterparts estimated assuming linear dependence between the weather covariates and the population abundance indices by fitting distributed lag models (DLM) using the *dlnm* package (Gasparrini 2011) of *R*. Due to the oscillatory nature of the sequence, with each oscillation having period one year and so needing to be approximated by a cubic polynomial, we found it necessary to use a basis function with 3 knots per year. The basis functions were created with the default knot locations using four years of monthly weather covariates (48 covariates and 12 knots in total) by the function *crossbasis*. Parameter estimation used *lme* with a linear trend over years and auto-regressive, AR(1), errors to ensure equivalence to the fitting of models E_1, E_2 and E_3.

3. Results

3.1 Results for two example species

The wren (*Troglodytes troglodytes*) is a small songbird whose population abundance index is derived from the BTO's Breeding Bird Survey (BBS) and Common Bird Census (CBC) (see Pearce-Higgins *et al.* 2015). The species is found throughout the UK, but because of the higher density of sites in England our modelling uses only English abundance data and hence only English weather data. The abundance index was assigned a reference month of June (r=6) since the latest of the two annual BBS site visits takes place then, and the time series used runs from 1966 to 2011, excepting 2001 when access to sites was restricted due to a national disease outbreak. We report here the modelled response to mean temperature, the species having been found to be much less responsive to precipitation. The initial grid searches indicated well-defined optima: the values of the transformed decay parameter κ most strongly supported by the data lay in the range from 2.5 to 3.5 for the models E_1, E_2 and E_3 (Fig. 1, top). Maximised log-likelihood values for the three models, along with the corresponding value for the null model E_0, are given in Table 1. These indicate that the transition from the null model E_0 to the model in which the covariate regression coefficients are defined by a first order damped Fourier series E_1 is highly statistically significant ($F_{4,39}$=9.95; $p < 10^{-4}$), whereas the evidence in support of an elaboration from E_1 to either of the more complex models is much weaker ($F_{2,37}$ = 2.12, 0.97; p = 0.13, 0.39; for the transitions from E_1 to E_2 and E_1 to E_3 respectively). Inspection of the plot of covariate regression coefficients b_m against m for underlying model E_1 (Fig. 1, middle) indicates that the signs of b_m oscillate approximately about zero, with negative values in mid-summer (when higher temperatures correlate with lower abundance indices) and positive values in mid-winter (when higher temperatures correlate with higher abundance indices). Values of the covariate regression coefficients decay by a factor of about 0.51 (= λ^{12}) between successive years. Regression coefficients estimated using DLM follow a similar pattern to those estimated using model E_1, comparison of the maximised log likelihoods indicating little evidence to justify the additional parameters required by the DLM (Table 1, column 1). The time series plot showing abundance indices and fitted values (Fig. 1, bottom) indicates broad agreement between observed and modelled series, although the sharp drop in log abundance in 1991 is not picked up well.

The flounced rustic (*Luperina testacea*) is a univoltine, grass-feeding noctuid moth whose population abundance index is derived from adults attracted to light traps mostly in August and September, as recorded by the Rothamsted Insect Survey. Our analyses used abundance and weather data from all of the UK, with response data from 1975 to 2010, with a reference month of September (r=9). We report here the modelled response to precipitation, the species having been found to be less responsive to temperature. The initial grid searches indicated well-defined optima for κ in the range from 3.0 to 4.0 for the models E_1, E_2 and E_3 (Fig. 2, top). Maximised likelihood values (Table 1) indicate that the transition from the null model E_0 to the model E_1 is highly statistically significant ($F_{4,30}$ = 8.82; $p < 10^{-4}$), whereas the evidence in support of either elaboration from E_1 is much weaker ($F_{2,28}$ = 1.23, 1.80; p = 0.31, 0.18; for the transitions to E_2 and E_3 respectively). Inspection of the plot of covariate regression coefficients for model E_1 (Fig. 2, middle) indicates that the b_m are universally negative (higher than normal precipitation in any month is associated with lower abundance indices) with the strongest influence of precipitation in mid-summer. Values of the covariate regression coefficients decay by a factor of about 0.73 (= λ^{12}) between successive years. Regression coefficients estimated using DLM follow a similar pattern to those estimated using model E_1, with no obvious benefit indicated by the maximised log likelihood (Table 1, column 2). The time series plot showing abundance indices and fitted values (Fig. 2, bottom) indicates broad agreement between observed and modelled series.

3.2 Summaries of model fits across all species in the study

The histograms of p-values for the augmentation from E_0 to E_1 have a more pronounced peak close to p=0 compared with either of the elaborations from E_1 to E_2 or E_1 to E_3 (Fig. 3). This is true both when the covariates are based on mean temperatures and on precipitation: thus although there is certainly evidence to support the more elaborate models for some species (as evidenced by the non-uniform nature of the histograms), the main gain is to be had in inclusion of the first order Fourier terms in the underlying model.

Comparison across species of AIC values obtained from fitting model E_1 with 480 monthly weather covariates (spanning 40 years) with values obtained from fitting model E_1 with 12 monthly weather covariates (spanning 1 year) strongly indicated support from the likelihoods for extending the series of weather covariates beyond the twelve most recent months at the cost of no extra parameters. For mean temperature, the AIC values for using 480 covariates was at least 2.0 greater than the corresponding value using 12 covariates for 164 species, whereas AIC differentials exceeded 2.0 in the opposite direction for just 23 species. The corresponding figures using the precipitation covariate were 176 and 11 respectively (histograms of AIC values provided as Supplementary Material).

The histogram of annualised decay parameter values, λ^{12}, has modes close both to 0 (only the covariates for the 12 months leading up to the abundance index are relevant) and to 1 (covariates for the 12 months leading up to the abundance index are given little more weight than covariates for earlier years). These modes are present for both temperature and rainfall covariates, and all three models (Fig. 4). Hence, the observation from the two example species that the regression coefficients decay steadily towards zero over approximately five years is not generally followed.

The SWMCRC values for temperature demonstrate a peak in positive values ($A^{S+}{}_{m}$) corresponding to July of the year of the abundance index, the values for this and adjacent months substantially exceeding the critical value of a one-sided test at the 5% significance level (Fig. 5). There is also some relatively weak evidence of an overall negative effect of temperature in January and February of the year of the abundance index. Conversely, for precipitation, the SWMCRC shows a peak in the negative ($A^{S-}{}_{m}$) values, again with a peak corresponding to July of the year of the abundance index. There are no months in which the positive SWMCRC values (A^{+}_{m}) even come close to statistical significance for precipitation, reinforcing the overall negative effect of unusually high precipitation on abundance indices across species.

4. Discussion

The above analyses showed strong evidence for an effect of temperature upon wren abundance. There was a positive effect of warmer winter temperatures, peaking in January, and a weaker negative effect of summer temperature. The effects of winter temperature support the results of previous studies of this species, whose populations are known to decline in response to cold winters (Greenwood & Baillie 1991, Newton *et al.* 1998, Morrison *et al.* 2016), due to reduced survival as a consequence of prolonged periods of frost and snow (Peach *et al.* 1995, Robinson *et al.* 2007). The negative effect of summer temperature upon wren populations has not previously been reported although there is increasing evidence that hot summer weather can have a negative impact on bird populations in the UK (Pearce-Higgins *et al.* 2010, Pearce-Higgins *et al.* 2015).

Little seems to have been published previously about the sensitivity of flounced rustic populations to either temperature or rainfall. The negative associations between abundance and rainfall, including timing of peak and least sensitivity, appear to be well aligned with the life history of the species: after hatching in early autumn, the larvae descend to the soil where they feed on stem bases and roots before pupating underground, usually in June (Waring and Townsend 2009). The long period spent underground, when the larvae are susceptible to water logging, is a likely explanation for rainfall correlating negatively with abundance and may contribute to their preference for light, calcareous soils. The period of peak sensitivity to rainfall comes during pupation, whilst the period of least sensitivity occurs during mid-winter when the larvae are likely to be least active.

The syntheses of results across all species in our study threw up some interesting generalities.

First, the lack of uniformity in the histograms of *p*–values provides evidence in support of all three models E_1, E_2 and E_3, for both the weather variables considered. Effects of temperature are more strongly supported by the data than effects of rainfall, and the evidence supporting the augmentation of the null model E_0 to E_1 model with the four additional underlying parameters was considerably greater than either of the elaborations considered (the additional two underlying parameters in the cyclic model that extend E_1 to either E_2 or E_3).

Second, the benefit of including more than the most recent 12 months of weather data was often considerable, with AIC differences greater than 2.0 in favour of having 480 monthly weather covariates outweighing AIC differences greater than 2.0 in favour of having 12 monthly weather covariates exceeding 5 to 1 for mean temperature and exceeding 15 to 1 for precipitation.

Third, the estimated annualised decay parameter values, λ^{12}, are bimodal, and tend to be estimated approximately equally often in the intervals from 0 to 0.1, from 0.1 to 0.9 and from 0.9 to 1. For example, with model E_1 the percentage of species for which $\lambda^{12} < 0.1$ is 43% for mean temperature and 38% for rainfall, whereas the equivalent figures for $\lambda^{12} > 0.9$ are 24% and 32% respectively. Having λ estimated as being close to 0 is an indication that the influence of the weather is confined to the 12 months leading up to the index value. Biologically, this would relate to a species with a low survival and high turnover of individuals and a strong effect of weather on productivity and or survival. Conversely, having λ close to 1 means that the effect of any departure in the index value due to variation in the weather is long-lasting, and corresponds to the situation for which Freeman and Newson (2008) noted that models for differences in log abundance with a single covariate are equivalent to models for log abundance using as explanatory variable the cumulative sums of the original covariate. Biologically, this might relate to situations in which weather impacts productivity or juvenile survival, and in which individuals have high survival and long-life spans. Therefore the impact on abundance of weather from 5 years ago is almost as great as the impact on abundance of weather from last year, because both years have contributed reasonably similar numbers to the current population. It was precisely to provide a bridge between the use of a single covariate and its cumulative sum that we began considering the exponential decay term which forms an integral part of our models, hence by construction the models described cater for both situations and intermediates in a manner determined by the data.

Fourth, the SWMCRC values from the E_1 model indicate that associations between abundance indices and temperature are on average consistently positive across species in the summer of the index year from May to August, whereas the associations with precipitation then are on average consistently negative across species. Whilst this observation ignores the potential benefit of the more elaborate models and does not apply to those species with early reference months, the main benefits of including weather data in the models are derived from the step from E_0 to $E_1$and some 94% of species investigated had reference months of May or later: hence there is certainly a suggestion that variation between years in summer weather tends to be important for the species we selected for inclusion in this study.

Although data for the two exemplar species were well-behaved in terms of underlying parameter estimation, fitting the models to data from all 501 species involved in our study was not always so straightforward. To implement all model fitting in a single loop required a system to trap non-convergence errors and to follow these with conditional continuation steps so that parameter estimates were always obtained, albeit with some error flags. There were also difficulties with constructing confidence intervals for species when the plausible range of the decay parameter λ was not constrained both above and below by the likelihood, in which case confidence intervals could be calculated conditional on the value of the λ, but these need interpreting with considerable caution.

The modelling approach described above can potentially be extended in many different directions.

One possibility is to investigate the use of error structures with more complex forms of long-range dependency than the Gaussian AR(1) relationships assumed here.

A second possibility is to return to the use of penalised spline regressions for defining the covariate regression coefficients, but doing so separately for the periodic term and the decay term. Estimation of the two smoothing terms could either be performed simultaneously or using sequential iteration between each using the back-fitting algorithm (Hastie and Tibshirani, 1990), allowing limited flexibility in the decay term to avoid overlap between the two terms.

A third possibility is to embed our models for abundance into a state space framework, thereby separating the effect of weather on population dynamics from the effect of weather on the observation process. This distinction is most acutely required for abundance indices which are derived from a summation of records over a long period of the year and are hence most likely to be influenced by both population size and the longevity of adults.

A fourth possibility would be to take a multi-species approach to the analysis, in combination with a state-space population model in the manner of Mutshinda, O'Hara & Woiwod (2011). Whilst multi-species modelling would have to be restricted to small numbers of species at a time, this would allow exploration of the extent to which changes in abundance are the result of species-weather interactions or due to inter-species interactions (Ockendon *et* *al.* 2014), the latter having been postulated as a cause of lagged population responses to variation in weather variables (e.g. Pearce-Higgins *et al.* 2010).

In conclusion, we believe that this paper represents a considerable advance on previous approaches to modelling the effect of weather on species abundance indices by making a more holistic use of historical weather records and relying less on potentially arbitrary *a priori* decisions about variable selection in the absence of much ecological information.

Acknowledgements

This work was undertaken for a project funded jointly by Defra, Natural Resources Wales (NRW), Natural England (NE), Scottish Natural Heritage (SNH) and the Joint Nature Conservation Committee (JNCC). Subsequent methodological development was funded by the Scottish Government Rural and Environment Science and Analytical Services Division. We thank members of the project Steering Group, colleagues (particularly Nick Isaac, CEH) two external reviewers and the Associate Editor for constructive comments. We are grateful to all volunteers and organisations contributing to the data sets used in this work, namely: the National Bat Monitoring Programme (Bat Conservation Trust, in partnership with JNCC and supported and steered by NE, NRW, Northern Ireland Environment Agency and SNH); the former CBC and current BBS (a partnership between the BTO, JNCC (on behalf of NRW, NE, Council for Nature Conservation and Countryside and SNH) and Royal Society for Protection of Birds); and the UK Butterfly Monitoring Scheme operated by CEH, Butterfly Conservation and funded by a consortium of government agencies.

Data accessibility

The weather data used in this manuscript can be obtained from the Met Office website http://www.metoffice.gov.uk/climate/uk/datasets/ . The wren data used as an example in this manuscript are available through the BTO's standard data request procedure (see http://www.bto.org/research-data-services/data-services/data-and-information-policy), whilst the flounced rustic data extracted from the Rothamsted Insect Survey (a BBSRC-supported National Capability) are available through http://www.rothamsted.ac.uk/insect-survey/.

References

Baltagi, B.H. (2008) *Econometrics*. Springer, Berlin.

Barlow K.E., Briggs P.A., Haysom K.A., Hutson, A.M., Lechiara, N.L., Racey, P.A., Walsh A.L. and Langton, S.D. (2015) Citizen science reveals trends in bat populations: the National Bat Monitoring Programme in Great Britain. *Biological Conservation* 182:14-26

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters* 15:365-377

Chen, I.-C., Hill, J.K., Ohlemüller, R., Roy, D.B. & Thomas, C.D. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science* 333:1024-1026

Freeman S.N. & Newson S.E. (2008) On a log-linear approach to detecting ecological interactions in monitored populations. *Ibis* 150: 250-258

Gasparrini, A. (2011) Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software 43: 1-20

Greenwood, J.D. & Baillie, S.R. (1991) Effects of density-dependence and weather on population changes of English passerines using a non-experimental paradigm. *Ibis* 133 S1:121-133

Hastie,T.J. & Tibshirani, R.J. (1990) *Generalized Additive Models*. Chapman and Hall, London.

Huntley, B., Green, R.E., Collingham, Y.C. & Willis, S.G. (2007) *A climatic atlas of European breeding birds*. Lynx Edicions, Barcelona, Spain.

Johnston, A., Ausden, M., Dodd, A. M., Bradbury, R. B., Chamberlain, D. E., Jiguet, F., Thomas, C.D., Cook, A.S.C.P., Newson, S.E., Ockendon, N., Rehfisch, M.M., Roos, S., Thaxter, C.B., Brown, A., Crick, H.Q.P., Douse, A., McCall, R.A., Pontier, H., Stroud, D.A., Cadiou, B., Crowe, O., Deceuninck, B., Hornman, M. & Pearce-Higgins, J.W. (2013) Observed and predicted effects of climate change on species abundance in protected areas. *Nature Climate Change* 3:1055–1061

Long, O.M., Warren, R., Price, J., Brereton, T., Botham, M.S. & Franco, A.M.A. (2017) Sensitivity of UK butterflies to local climatic extremes: which life stages are most at risk? *Journal of Animal Ecology* 85:1636-1646

Martay, B., Brewer, M. J., Elston, D. A., Bell, J. R., Harrington, R., Brereton, T. M., Barlow, K. E., Botham, M. S. & Pearce-Higgins, J. W. (2016). Impacts of climate change on national biodiversity population trends. *Ecography* doi: 10.1111/ecog.02411.

Marx, B.D. & Eilers, P.H.C. (1999) Generalized linear regression on sampled signals and curves: a p-spline approach. *Technometrics* 41:1-13

Morrison, C.A., Robinson, R.A. & Pearce-Higgins, J.W. (2016) Winter wren populations show adaptation to local climate. *Open Science* 3:160250

Mutshinda, C.M., O'Hara, R.B. & Woiwod, I.P. (2011) A multispecies perspective on ecological impacts of climatic forcing. *Journal of Animal Ecology* 80:101-7

Newton, I., Rothery, P. & Dale, L.C. (1998) Density-dependence in the bird populations of an oak woodland over 22 years. *Ibis* 140:131-136

Ockendon, N., Baker, D.J., Carr, J.A., Almond, R.E.A., Amano, T., Bertram, E., Bradbury, R.B., Bradley, C., Butchart, S.H.M., Doswald, N., Foden, W., Gill, D.J.C., Green, R.E.,Sutherland, W.J., Tanner, E.V.J. & Pearce-Higgins, J.W. (2014) Mechanisms underpinning climatic impacts on natural populations: altered species interactions are more important than direct effects. *Global Change Biology* 20:2221-2229

Peach, W.J., de Feu, C. & McMeeking, J. (1995) Site tenacity and survival rates of wrens *Troglodytes troglodytes* and treecreepers *Certhia familiaris* in a Nottinghamshire wood. *Ibis*, 137:497-507

Pearce-Higgins, J.W., Dennis, P., Whittingham, M.J. & Yalden, D.W. (2010) Impacts of climate on prey abundance account for fluctuations in a population of a northern wader at the southern edge of its range. *Global Change Biology* 16:12-23

Pearce-Higgins, J.W., Eglington, S.M., Martay, M. & Chamberlain, D.E. (2015) Drivers of climate change impacts on bird communities. *Journal of Animal Ecology* 84:943-954

Perry M. & Hollis D. (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology* 25:1041–1054

Pinheiro J., Bates D., DebRoy S., Sarkar D. & R Core Team (2014) _*nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-118, URL: http://CRAN.R-project.org/package=nlme.

Pollard, E. (1988) Temperature, rainfall and butterfly numbers. *Journal of Applied Ecology* 25:819–828

R Core Team (2015) R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, URL https://www.R-project.org/.

Ratkowsky, D. A. (1983) *Nonlinear Regression Modeling*, Marcel Dekker, New York.

Roberts A.M.I. (2008) Exploring relationships between phenological and weather data using smoothing. *International Journal of Biometeorology* 52:463-470

Roberts A.M.I. (2012) Comparison of regression methods for phenology. *International Journal of Biometeorology* 56:707-717

Robinson, R.A., Baillie, S.R. & Crick, H.Q.P. (2007) Weather-dependent survival:

implications of climate change for passerine population process. *Ibis* 149:357-364

Roy, D. B., Rothery, P., Moss, D., Pollard, E. & Thomas, J. A. (2001) Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. *Journal of Animal Ecology* 70:201–217

Sims M., Elston D.A., Larkham A., Nussey D.H. & Albon S.D. (2007) Identifying when weather influences life history traits of grazing herbivores. *Journal of Animal Ecology* 76:761-770

Waring, P. and Townsend, M. (2009) *Field Guide to the Moths of Great Britain and Ireland* 2nd edn. British Wildlife Publishing, Gillingham.

Warren, R., VanDerWal, J., Price, J., Welbergen, J.A., Atkinson, I., Ramirez-Villegas, J., Osborn, T.J., Jarvis, A., Shoo, L.P., Williams, S.E. & Lowe, J. (2013) Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. *Nature Climate Change* 3:678–682

Tables and Figures

Table 1

Summary of results from the model fitting, indicating for each species and each model: the maximised log likelihood; the estimated AR(1) correlation coefficient; and twelfth power (ie annual influence) of the estimated decay rate parameter.

		Species	
Model (number of fixed effect parameters)	Term	Wren	Flounced rustic
E_0 (2)	log likelihood	10.09	-21.96
	AR(1) correlation	0.44	0.56
E_1 (6)	log likelihood	29.99	-4.32
	AR(1) correlation	0.47	0.53
	λ^{12}	0.51	0.73
E_2 (8)	log likelihood	32.10	-3.09
	AR(1) correlation	0.44	0.52
	λ^{12}	0.42	0.73
E_3 (10)	log likelihood	30.95	-2.52
	AR(1) correlation	0.46	0.57
	λ^{12}	0.59	0.69
DLM (14)	log likelihood	33.70	-0.48
	AR(1) correlation	0.38	0.57

Fig. 1

Results of fitting the models to wren data: profile likelihood from initial grid search for the logit of the monthly decay rate parameter κ (top), for models E_1 (solid line), E_2 (dashed) and E_3 (dotted); covariate regression coefficients b_m (centre) in model E_1 for 72 monthly temperature covariates up to the month to which the index relates (June, m=1, r=6) (with 95% confidence intervals), along with values derived from DLM (circles) using 48 monthly temperature covariates; observed time series, showing fitted values from model E_1 with 95% confidence intervals for the fitted values (bottom).

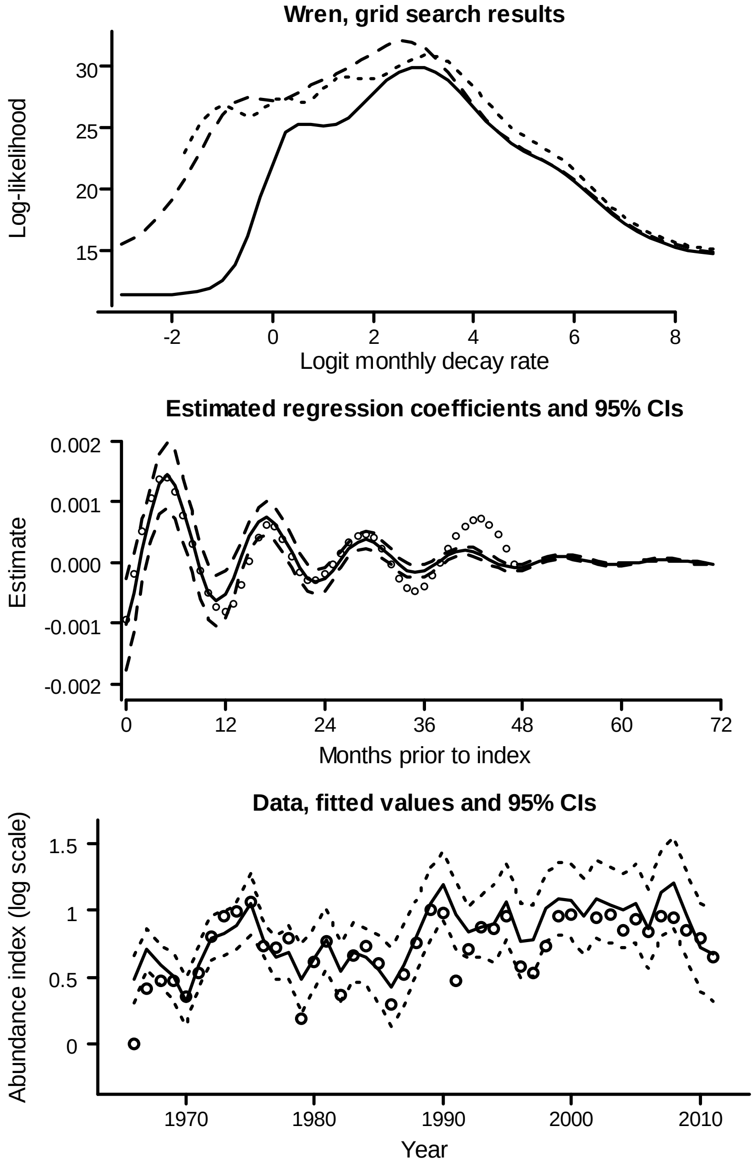

Fig. 2

Results of fitting the models to flounced rustic data: profile likelihood from initial grid search for the logit of the monthly decay rate parameter κ (top), for models E_1 (solid line), E_2 (dashed) and E_3 (dotted); covariate regression coefficients b_m (centre) in model E_1 shown for 60 monthly precipitation covariates up to the month to which the index relates (September, m=1, r=9) (with 95% confidence intervals), along with values derived from DLM (circles) using 48 monthly temperature covariates; observed time series, showing fitted values from model E_1 with 95% confidence intervals for the fitted values (bottom).

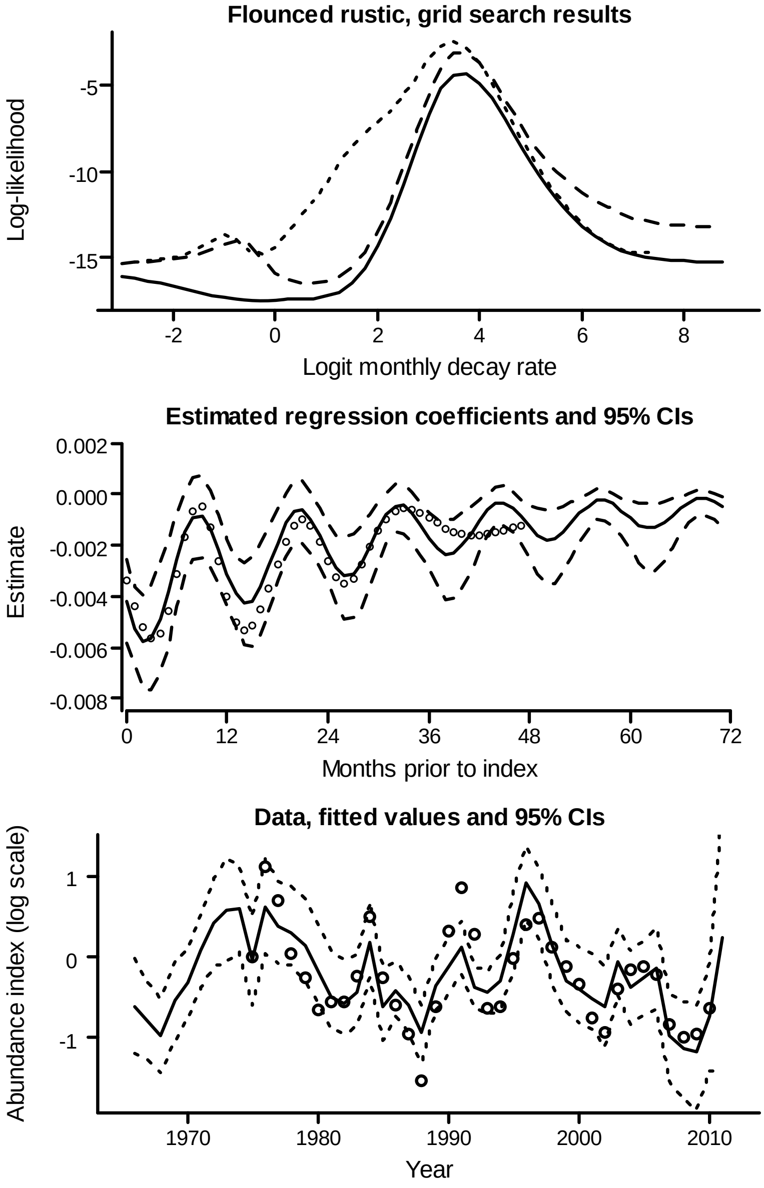

Fig. 3

Histograms of *p*-values across all species using mean temperature (left) and precipitation (right) as covariates, for augmentation from: the null to first order models (E_0 to E_1, top, using all 501 species); the first order to the second order model by addition of a periodic term with period 6 months (E_1 to E_2, middle, using those 484 species with at least 20 abundance index values); and the first order to the second order model by addition of a periodic term with period 24 months (E_1 to E_3, bottom, same 484 species).

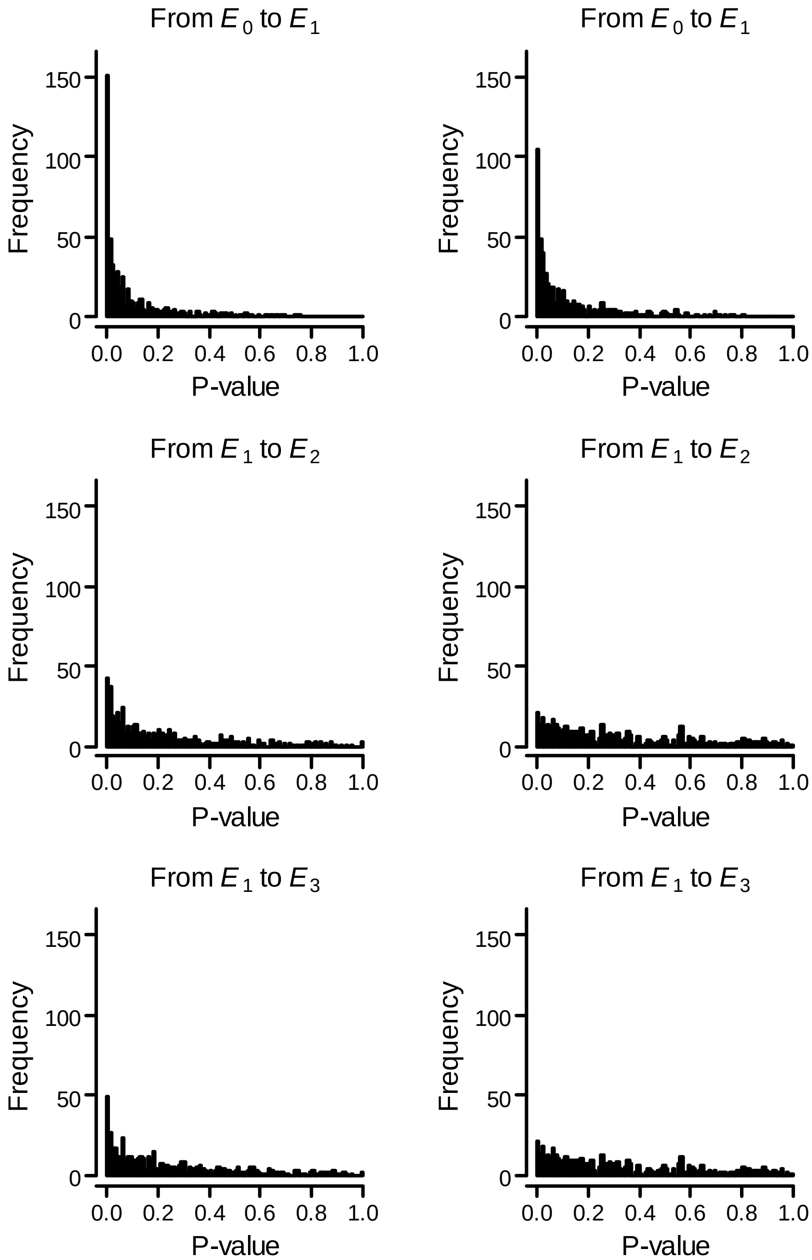

Fig. 4

Histograms of estimated annualised decay coefficients (λ^{12}) using mean temperature (left) and precipitation (right) as covariates for: the first order model E_1 (top, using all 501 species); and the second order models with a periodic term with period 6 months (E_2, middle, using those 484 species with at least 20 abundance index values); and with period 24 months (E_3, bottom, same 484 species).

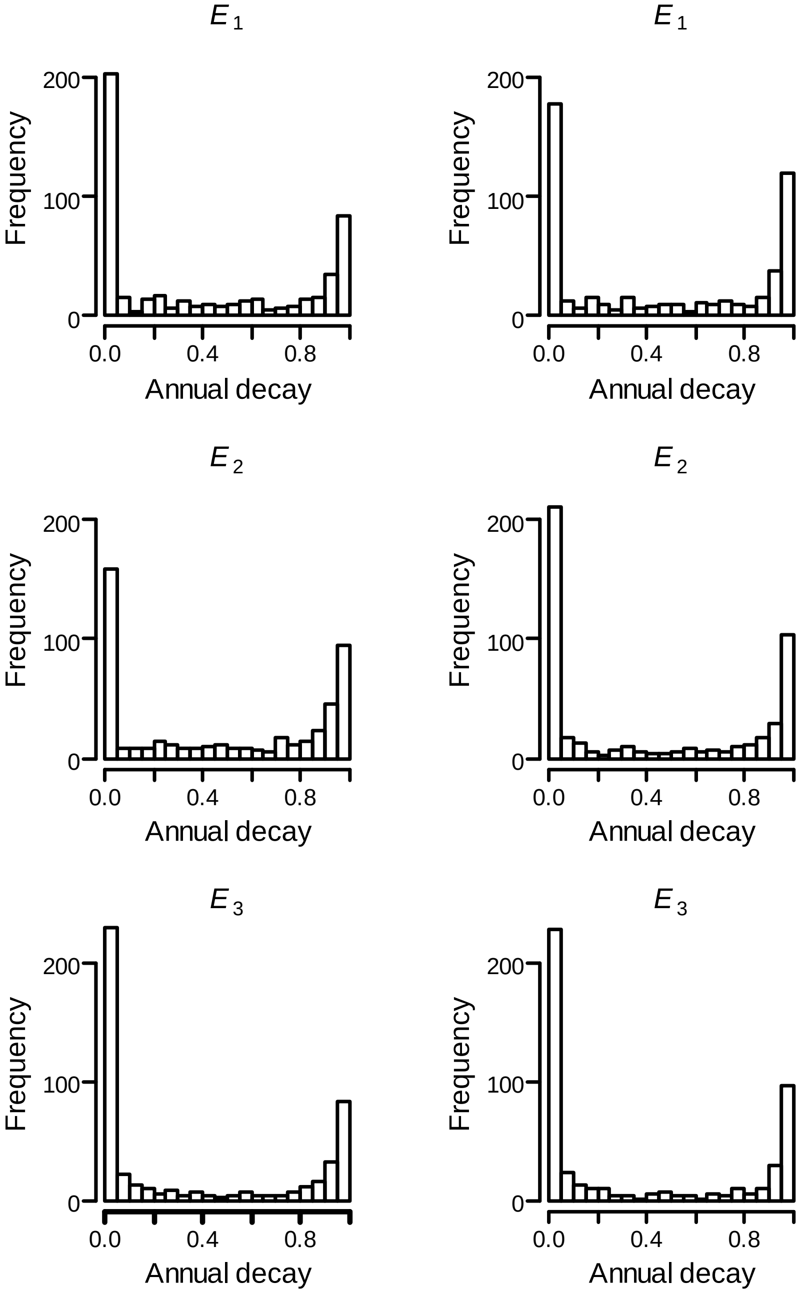

Fig. 5

Standardised weighted mean covariate regression coefficients (SWMCRC, see text for details) constructed across all 501 species for model E_1 using mean temperature covariate data (top) and precipitation (bottom), showing calculated values for positive (A^{S+}_{m}) and negative (A^{S-}_{m}) coefficients (solid lines) and one-sided 95% confidence intervals (dashed lines).