Received: 26 April 2022 | Accepted: 3 August 2022

DOI: 10.1111/2041-210X.13974

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Methods in Ecological Forecasting

Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series

Nicholas J. Clark[1] | Konstans Wells[2]

[1]School of Veterinary Science, The University of Queensland, Gatton, QLD, Australia

[2]Department of Biosciences, Swansea University, Swansea, UK

Correspondence
Nicholas J. Clark
Email: n.clark@uq.edu.au

Funding information
Australian Research Council, Grant/Award Number: DE210101439

Handling Editor: Sydne Record

Abstract

1. Generalised additive models (GAMs) are increasingly popular tools for estimating smooth nonlinear relationships between predictors and response variables. GAMs are particularly relevant in ecology for representing hierarchical functions for discrete responses that encompass complex features including zero inflation, truncation and uneven sampling. However, GAMs are less useful for producing forecasts as their smooth functions provide unstable predictions outside the range of training data.
2. We introduce dynamic generalised additive models (DGAMs), where the GAM linear predictor is jointly estimated with unobserved dynamic components to model time series that evolve as a function of nonlinear predictor associations and latent temporal processes. These models are especially useful for analysing multiple series, as they can estimate hierarchical smooth functions while learning complex temporal associations via dimension-reduced latent factor processes. We implement our models in the MVGAM R package, which estimates unobserved parameters for smoothing splines and latent temporal processes in a probabilistic framework.
3. Using simulations, we illustrate how our models outperform competing formulations in realistic ecological forecasting tasks while identifying important smooth predictor functions. We use a real-world case study to highlight some of MVGAM's key features, which include functions for calculating correlations among series' latent trends, performing model selection using rolling window forecasts and posterior predictive checks, online data augmentation via a recursive particle filter and visualising probabilistic uncertainties for smooth functions and predictions.
4. Dynamic GAMs (DGAMs) offer a solution to the challenge of forecasting discrete time series while estimating ecologically relevant nonlinear predictor associations. Our Bayesian latent factor approach will be particularly useful for exploring competing dynamic ecological models that encompass hierarchical smoothing structures while providing forecasts with robust uncertainties, tasks that are becoming increasingly important in applied ecology.

KEYWORDS
dynamic factor model, ecological forecasting, generalised additive model, hierarchical model, JAGS, R package, Stan

1 | INTRODUCTION

Rapidly changing climates and landscape modification are impacting species and ecosystems at all micro- and macroecological levels, incurring substantial economic and environmental costs (Kennedy et al., 2019; United Nations, 2015; World Health Organization, 2005). There is broad consensus among scientists, parliamentarians and applied decision makers that anticipating probable future states is vital to mitigate the impacts of environmental change on ecosystem functionality and services (Dietze et al., 2018; Intergovernmental Panel on Climate Change, 2018; Schmidt et al., 2010).

Two challenges impede the improvement and adoption of common forecasting tools in ecology. First, ecosystems are driven by networks of interacting biotic and abiotic processes (Choler et al., 2001; Levin, 1998; Massoud et al., 2018). These dynamic natural processes are the products of multiple sources of variation including long-term trends, seasonal and other cyclic oscillations, environmental forcing, temporal dependence or species interactions (Auger-Méthé et al., 2021; Choler et al., 2001; Dietze, 2017). Second, ecological time series are often integer-valued variables, such as observations of species presence or abundance, that exhibit complex features including observation error, zero inflation, overdispersion, truncation at hard bounds, missing values and uneven sampling (Kowal & Canale, 2020; Lindén & Mäntyniemi, 2011; Simpson, 2018; Warton, 2018). Such discrete time series are far less supported in existing software than are real-valued series that can be readily modelled using assumptions of Gaussian error (Hyndman & Khandakar, 2008). Moreover, ecological observations are almost always multivariate when contextual information, such as data from environmental predictors or observations of non-target species, is considered. These features make it difficult to analyse ecological time series while sufficiently accounting for important systematic temporal components and multivariate dependencies (Auger-Méthé et al., 2021).

Time-series analyses are often concerned with decomposing temporal variation into components representing trend, seasonality and other cyclic changes. Generalised additive models (GAMs), which are increasingly used in ecology to identify nonlinear functional relationships (Guisan et al., 2002; Hughes et al., 2018; Pedersen et al., 2019; Simpson, 2018), offer a way to accomplish this decomposition. Outlined in detail previously (Hastie & Tibshirani, 1990; Wood, 2004), GAMs can briefly be described as modified generalised linear models (GLMs) in which the linear predictor includes a sum of smooth functions representing functional relationships between covariates and the response:

$$\mathrm{E}(Y) = g^{-1}\left(\beta_0 + \sum_{i=1}^{I} s_i\left(x_i\right)\right) \quad (1)$$

where $\mathrm{E}(Y)$ is the conditional expec tation of a response assumed to be drawn from an exponential family distribution, β_0 is the unknown intercept, the s_i's are a set of smooth functions over one or several predictor variables (the x's) and g is a monotonic link function. Each smooth function s_i is composed of basis functions whose coefficients, which must be estimated, act as weights for the basis functions to control the function's shape. The total number of basis functions limits the potential complexity of the smooth function, with a larger set of basis functions allowing greater flexibility. In addition to their ability to represent complex and nonlinear ecological relationships, several other advantages of GAMs are that they can model a diversity of response families that accommodate ecological features (such as zero inflation) and that they can be formulated to include hierarchical smoothing for multivariate responses (Pedersen et al., 2019; Wood, 2017).

Given the set of basis coefficients that comprise each smooth function, a GAM can in principle be estimated as a GLM. However due to their incredible flexibility, GAMs will invariably overfit if left unconstrained (Hastie & Tibshirani, 1990; Marra & Wood, 2011; Wood, 2004). Penalised likelihood estimation avoids this overfitting by placing quadratic penalties on the basis coefficients (referred to as smoothing penalties), which penalise the function's 'wiggliness' and control the trade-off between fit and smoothness (Wood, 2004, 2016). From a Bayesian perspective, another way to represent a smooth function is to draw the set of basis coefficients from a multivariate Gaussian distribution with the penalty acting on the prior precision. Larger penalties shrink the coefficient covariances, effectively forcing the function towards a straight line when the data do not justify a nonlinear relationship (Marra & Wood, 2011; Wood, 2016). GAMs are particularly sought after for modelling time series to identify nonlinear or time-varying covariate effects, perform smoothing of historical time series and uncover periods of rapid change, though strong temporal autocorrelation can make it challenging to estimate key parameters (Camara et al., 2021; Knape, 2016; Simpson, 2018; Spooner et al., 2018; Yang et al., 2012).

For many ecological studies that employ GAMs, a primary objective is predicting future states (Clark et al., 2020; Kaplan et al., 2016; Koolhof et al., 2021; Malick et al., 2020; Ward et al., 2014). However, a lingering issue in using GAMs for forecasting is the way in which smooth functions predict outside the range of training data. Many of the smooth functions used in ecological GAMs have zero second derivatives at the boundaries, meaning they will linearly extrapolate beyond the last observation (Elith et al., 2010; Zurell et al., 2012). This projection of a straight line indefinitely into the future can produce unrealistic forecasts, particularly if the estimated function 'wiggles' (i.e. exhibits a pronounced change in the response–predictor relationship) near the boundary (Figure 1 top). There are technical solutions to help with this problem, for example, by extending the

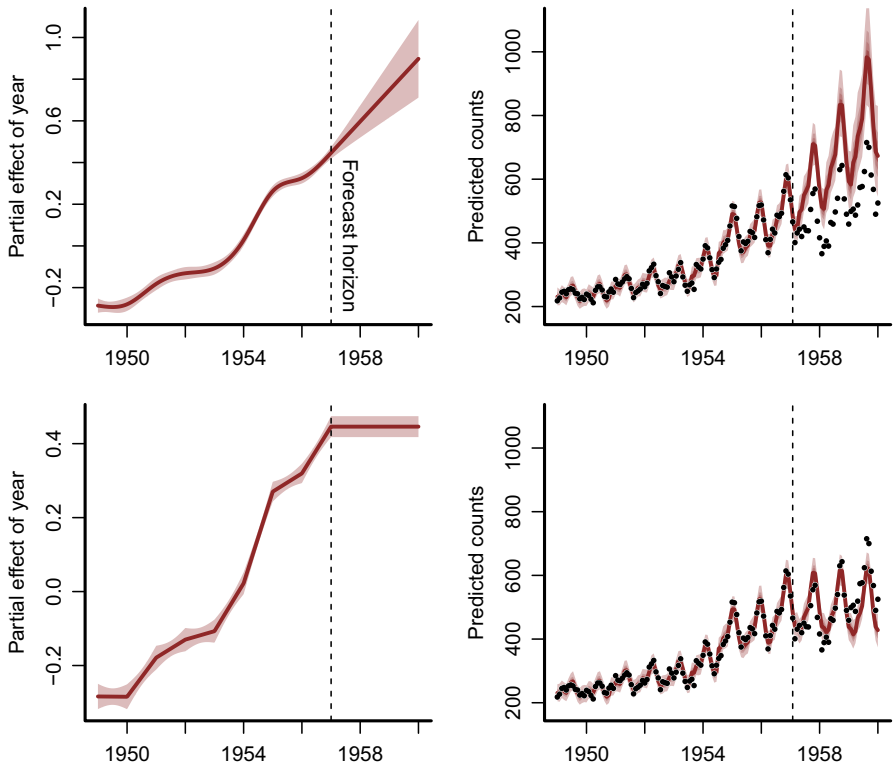

FIGURE 1 Estimated trends and forecasts from two GAMs applied to a discrete time series. In the top panel, a thin plate regression spline with a penalised second derivative is used for the trend, leading to a smooth function (top left) and linear extrapolation when forecasting (top right). In the bottom panel, the trend penalty is placed on the first derivative, resulting in flat extrapolation when forecasting. Trend shading shows 95% confidence intervals, while forecast shading shows empirical quantiles. Both models were fitted to a simulated seasonal discrete time series in R using the MGCV package with the general formula: *Y* ~ *s(year, bs = 'tp') + s(season, bs = 'cc') + ti(season, year), family = nb()*).

evaluation of the penalty into the range of values that we wish to forecast (i.e. weeks or years ahead of the training data) to ensure model uncertainty grows in a more realistic fashion out of sample, or by forcing the function to use the last observed value (with fixed uncertainty) when forecasting by imposing a first derivative penalty (Figure 1 bottom). However, these modifications are not always sufficient to generate robust forecasts with appropriate probabilistic uncertainties as they do not adequately capture the temporal dependence in the data (see examples in Appendix S1 in Supporting Information).

Here, we outline a Bayesian dynamic GAM (DGAM) that provides a general solution to the problem of estimating smooth functions while generating forecasts for discrete time series. The approach is simple: for a given series, we augment the GAM linear predictor with a latent dynamic component to capture the series' temporal evolution process (currently either as a random walk, an autoregressive process up to order 3 or a Gaussian process). To model multiple time series, we accommodate possible dependencies among series' temporal components in a parsimonious way using a dynamic latent factor process. We introduce our associated R package MVGAM (https://github.com/nicholasjclark/mvgam) and illustrate its utility via simulations and empirical examples.

We begin by introducing DGAMs, including background material for the dynamic factor process. We then illustrate our package's utility for ecologists and other users interested in forecasting discrete time series using both simulations and a case study. An introduction to MVGAM's primary functions via more in-depth reproducible examples is provided in the Appendices S1–S3 (Supporting Information).

2 | DYNAMIC GENERALISED ADDITIVE MODELS

2.1 | Univariate models for a single ecological time series

A Bayesian framework to model fitting and parameter estimation involves defining a joint probability distribution over all observable and unobservable quantities in a statistical model that aligns with expert beliefs about the data generating process (Gelman et al., 2017). A DGAM is naturally viewed from a Bayesian perspective, where prior beliefs about the nonlinearity of a function can be elicited to inform the complexity and penalisation of the smooth (Miller, 2019; Pedersen et al., 2019; Wood, 2013) while accounting for possible unobserved temporal dependence in line with the expectation that time series evolves as serially autocorrelated dynamic processes (Hyndman & Athanasopoulos, 2018). In its basic form, the DGAM for a discrete integer-valued time series is written as:

$$\mathrm{E}\left(Y_t\right) = g^{-1}\left(\beta_0 + \sum_{i=1}^{I} s_{i,t}\left(x_{i,t}\right) + z_t\right) \tag{2}$$

$$z_t = \varphi_0 + z_{t-1} + e_t \tag{3}$$

where $\mathrm{E}\left(Y_t\right)$ is the conditional expectation of the response at time t and z_t is the (latent) dynamic process estimate at time t. Readers familiar with state-space models will recognise the benefits of separating the temporal and observation processes (Auger-Méthé

et al., 2021; Heilman et al., 2022), but it is worth clarifying these advantages explicitly. First, estimating the trend as a dynamic random variable avoids problems that can occur in competing autoregressive observation models where measurement error or outliers can have large influences on estimated AR parameters and cause highly unstable forecasts (see an example in Appendix S1). Second, it is far easier to handle missing or irregularly sampled observations using latent processes. Because the z_t are unobserved latent variables, they will continue to evolve, even when an observation Y_t is missing, via dynamic equations that conveniently provide recursive expressions for h-step ahead prediction, historical filtering and updating of forecasts (Durbin & Koopman, 2012). In contrast, a missing observation in an AR3 observation model will result in NAs for four rows of the design matrix (one missing Y_t and three missing AR predictors) that can make parameter estimation difficult for software that automatically excludes rows with missing values (such as commonly used linear modelling packages in R). Other advantages of a state-space form are that trend dynamics provide a probabilistic model for the temporal evolution of a process, which can often be more useful than a smoothed trend (such as a penalised spline) by facilitating simulation and comparison with other processes, allowing new observations to be assimilated to adapt a forecast distribution via recursive Kalman or particle filtering (Massoud et al., 2018) and providing a means for multiple observation processes to depend on shared latent processes (Ward et al., 2021).

In its simplest form, temporal dependence can be modelled as a random walk with possible drift, where φ_0 in Equation (3) is the drift parameter and the residual error e_t is drawn from a zero-centred Gaussian distribution with a fixed (time-invariant) standard deviation. This can easily be expanded to include autoregressive (AR) processes. For example, the following specifies a latent AR2 model:

$$z_t = \varphi_0 + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + e_t \tag{4}$$

The assumption of a fixed standard deviation for the process error could potentially be a limitation if the series of interest displays nonconstant volatility with perturbations that may be evidence of responses to 'shocks'. The time series literature is rich with model specifications for accommodating dynamic distributional models, including stochastic volatility, GARCH or Lévy processes (Bartumeus, 2007; Carrasco & Chen, 2002). In sharp contrast, temporal dependence could also be modelled via a latent Gaussian process (or other stochastic process), which provides a nonparametric probability distribution over functions. Gaussian processes are particularly suitable for ecological time series where we often expect dynamics to evolve as a smooth function and we wish to estimate the covariances among time points to facilitate probabilistic forecasts (Riutort-Mayol et al., 2020; Ward et al., 2021). To save computational costs, it is possible to use low-rank approximate Bayesian Gaussian processes that are approximated using Laplace eigenfunctions, which have been shown to have excellent forecasting properties via simulations by Riutort-Mayol et al. (2020).

2.2 | Dynamic factor DGAMs for analysing multiple ecological time series

Here, we describe how a DGAM can be modified into a joint multivariate statistical model for collections of time series with potentially common dynamics. Dynamic factor models that account for relationships in time-series data are closely aligned with static latent factor models, which are used in quantitative ecology to jointly model abundances of multiple species by estimating shared responses to unmeasured ecological drivers (Ovaskainen et al., 2017; Thorson et al., 2016; Ward et al., 2021; Warton et al., 2015). A latent factor model is a function of unmeasured random predictors (factors) that induce correlations between multiple responses via factor loadings while exercising dimension reduction. Often, species do demonstrate correlated responses to environmental gradients, meaning that a smaller set of factors (i.e. a low-dimensional representation) than the total number of possible species–predictor relationships can adequately capture the main axes of covariation (Letten et al., 2015; Warton et al., 2015). A dynamic factor model assumes the factors evolve as time series. The strength of this approach is that a small number of common factors can often model the temporal behaviours of a much larger set of series. This dimension reduction simplifies the estimation and forecasting tasks, as only the smaller set of factors and the series' specific factor loadings need to be estimated to generate forecasts (De Stefani et al., 2019). In a dynamic factor DGAM, each series' latent trend is composed of a linear combination of these common factors:

$$\mathrm{E}\left(Y_{j,t}\right) = g^{-1}\left(\beta_0 + \sum_{i=1}^{I} s_{i,j,t}\left(x_{i,t}\right) + \sum_{m=1}^{M}\left(z_{m,t}\theta_j\right)\right) \tag{5}$$

where $\mathrm{E}\left(Y_{j,t}\right)$ is the expected response for series j at time t, the $z_{m,t}$'s are estimates for the M factors at time t and the θ_j's are factor loadings. As in the univariate case, the factors can evolve either as random walks with drift or as autoregressive processes up to order 3.

A challenge with any factor model is the need to determine the number of factors M (Bhattacharya & Dunson, 2011; Fox et al., 2009; Thorson et al., 2016; Tobler et al., 2019). Setting M too small prevents temporal dependencies from being adequately modelled, leading to poor convergence and difficulty estimating smooth parameters. By contrast, setting M too large leads to unnecessary computation. The problem can be approached by formulating a prior distribution that enforces exponentially increasing penalties on the factor variances to allow any un-needed factors to evolve as flat lines. Following Welty et al. (2009) and Wells et al. (2016), one such prior assumes that factors up to a certain threshold number π have precisions of similar magnitudes, after which they increase exponentially (leading to variances that shrink towards zero). Along with π, two other hyperparameters can be estimated to control the baseline penalty and the rate at which penalties exponentially increase, respectively, allowing the data to inform the selection of dynamic factors. We caution, however,

that setting *M* too large could result in trends that are overly flexible, making it challenging to simultaneously estimate important smooth functions such as seasonality. It is certainly worth checking whether inferences or forecasts are sensitive to *M*, perhaps using the guidelines outlined by Tobler et al. (2019). Additional constraints are also needed to preserve identifiability by setting the upper triangle of the factor loading matrix to zero and ensuring non-negative diagonals (Hui, 2016; Tobler et al., 2019).

3 | ESTIMATING DGAMS WITH THE MVGAM R PACKAGE

While it is possible to model residual autocorrelation for univariate series in the popular R package MGCV using restricted maximum likelihood via the *gamm()* or *bam()* functions (Wood, 2017), there is no straightforward way to include an autocorrelation process in forecasts. There is also no simple way that we are aware of to estimate dynamic factor DGAMs using existing open-source software. We introduce the MVGAM R package as an open-source software tool to estimate the parameters of DGAMs for discrete time series and use them to generate probabilistic forecasts. Our models are coded in either the JAGS or Stan probabilistic programming languages using the function *mvgam()*, which relies on the *jagam()* function from MGCV to generate a skeleton model file and necessary smooth penalty matrices (Wood, 2016). The model is modified to include dynamic components (either as random walk, AR trends up to order 3 or Gaussian processes) and to update any prior distributions specified by the user, while all data reformatting necessary for modelling is done automatically. Employing either the JAGS software through the R interface *rjags* (Plummer, 2003) or the Stan software through the interfaces *rstan* (Carpenter et al., 2017) or *cmdstanr* (Gabry & Češnovar, 2021), the model is conditioned on observed data using Markov Chain Monte Carlo (MCMC) simulation to calculate the posterior probability distribution of the unobserved parameters of interest. The MVGAM R package provides the following key functions:

- Estimate the parameters of DGAMs in a Bayesian Markov Chain Monte Carlo framework via either the Gibbs sampling software JAGS (Plummer, 2003; Wood, 2016) or using Hamiltonian Monte Carlo in the software Stan (Carpenter et al., 2017) using the function *mvgam()*
- Plot estimated smooth functions and posterior predictions, along with their probabilistic uncertainties and derivatives, using the S3 function *plot.mvgam()*
- Perform residual diagnostic checks using randomised quantile (Dunn–Smyth) residuals (Dunn & Smyth, 1996) using the S3 plot function *plot.mvgam(type = 'residuals')*
- Plot posterior retrodictive and predictive checks to examine discrepancies between observed data and model-generated simulations (Gabry et al., 2019) using the S3 function *ppc.mvgam()*
- Compute correlations among latent trends for multivariate sets of series using the function *lv_correlations()*
- Perform model selection using rolling window forecast evaluation with functions *eval_mvgam()*, *roll_eval_mvgam()* and *compare_mvgams()*
- Update forecasts online via a Sequential Monte Carlo particle filter using functions *pfilter_mvgam_init()* and *pfilter_mvgam_online()*
- Create the model file and all necessary objects needed to initialise and condition the model so that users can modify the model structure to fit their bespoke needs using the function *mvgam(run_model = FALSE)*

MVGAM extends functions available in existing software in several ways. First, while fully Bayesian GAMs can be estimated using a variety of software including *brms* (Bürkner, 2017), *BayesX* (Brezger et al., 2005) and *bamlss* (Umlauf et al., 2018), MVGAM is the only software we are aware of that can simultaneously estimate any smooth function available in MGCV together with latent dynamic trends (*bamlss* and *BayesX* can estimate a diversity of smooth functions, but to our knowledge, dynamic latent processes cannot be estimated; *brms* offers more flexibility for time series and can accommodate dynamic latent processes, including AR and ARMA processes of order 1, but we are not aware of extensions to dynamic factors). Second, our software can employ Hamiltonian Monte Carlo using *Stan* for much more efficient and unbiased MCMC sampling compared to Gibbs samplers (*BayesX* uses its own custom Gibbs samplers, while *bamlss* does not employ full MCMC). Perhaps the most important advantage of Hamiltonian Monte Carlo is the powerful diagnostics it provides for detecting posterior degeneracies, which can help uncover model inadequacies or incompatibilities between model and observed data (Betancourt, 2017). Finally, our package is designed for analysing and forecasting sets of discrete time series, and as such the additional utilities we offer for working with time series (including options to compare models using rolling forecast evaluation as well as routines to assimilate new observations 'online' for automatic forecast updating; Appendix S1) make our software attractive for a range of applied forecasting tasks.

It is notable that our design permits any formula allowed in MGCV to be used for the GAM component of the linear predictor, providing a user-friendly way to explore dynamic ecological models that encompass nonlinear smooth functions. Other advantages of our framework are (1) missing values are allowed for the responses; (2) upper bounds can be used via truncated likelihoods; (3) smooth distributed lag covariate functions can be estimated alongside latent temporal components to form complex dynamic nonlinear models (Gasparrini, 2011); and (4) dynamic components can easily be forecasted via their autoregressive equations (for random walk and AR trends) or via their estimated covariance functions (for Gaussian process trends), providing robust probabilistic uncertainties.

While the MVGAM package does not currently support stochastic volatility or moving average trends, these processes could be added

by the user at any time (the package can be used to generate all model files, data objects and initial values, so that a model can be easily modified for conditioning outside of MVGAM, i.e. with *rstan*, *rjags* or other interfaces directly).

4 | SIMULATIONS

We used simulations to examine the performance of our DGAM formulation. Briefly, we simulated multiseries datasets with 72 time points (6 years of data for monthly series) consisting of Negative Binomial observations (size parameter = 5) for sets of series whose log-linear predictors included a hierarchical seasonal pattern (where each series' seasonal pattern was created by drawing from a global seasonal pattern with common Gaussian noise; see function *sim_mvgam()* in the MVGAM package for R code to produce simulations) and uncorrelated latent trends. Temporal dependences followed independent random walk processes. We investigated model sensitivity to missingness (proportion missing = 0, 10 or 50%), dimensionality (number of series = 2, 4 or 12) and the magnitude of the temporal component relative to seasonality (0.3 for moderate dynamics or 0.7 for strong dynamics; see Figure S1 for an example of two series with the same seasonality but different strengths of trend). Each simulated dataset was fit with the same set of four models. First, we fit a hierarchical GAM using MGCV that included a random intercept per series (*s(series, bs = 're')*), a cyclic smooth function for global seasonality (*s(season, m = 2, k = 8, bs = 'cc')*), local smooth functions for series-specific deviations from global seasonality (*s(season, series, m = 1, k = 4, bs = 'fs')*), a smooth function for a global trend (*s(year, k = 4)*) and local smooth functions for series-specific deviations from the global trend (*s(year, series, m = 1, k = 4, bs = 'fs')*). Our next model was a GAM (also fitted with MGCV) that used a stochastic trend via an autoregressive observation model. This model used same hierarchical seasonality smooths functions as the GAM above but replaced the trend smooths with an AR1 parametric term for the effect of $\log(y_{t-1})$, with separate AR1 terms estimated for each series. We chose to model the AR1 term on the log scale as this reduces sensitivity of the AR parameter estimates to outliers (see Appendix S1 for an investigation of the forecasting behaviours of autoregressive observation models for discrete time series). Note, however, that because each missing observation results in additional missing rows in the design matrix (due to missing values in AR predictors), we were unable to fit this model for the simulations where 50% of observations were missing. We next asked whether a dynamic factor process could capture the multiseries temporal dynamics by fitting a dynamic factor DGAM (with M = half the number of series) with identical random effect and seasonal smooth functions but no yearly smooth function. Finally, we fit a 'null' dynamic factor DGAM that only estimated random intercepts but no seasonal smooth function. Negative binomial distributions were specified for each model and AR1 models were used for modelling the DGAM dynamic factor processes. Each combination of missingness, dimensionality and strength of dynamics was used to generate five replicate datasets, yielding a total of 60 simulations. For MGCV models, estimation of smoothing penalties was performed using restricted maximum likelihood (*method = 'REML'*). Gaussian priors were specified for AR parameters (φ) (mean = 0; variance = 0.1) in the *mvgam* implementation. Following Wood (2016), zero-centred multivariate Gaussian priors were used for each smooth's β parameters and exponential priors were used for the smoothing penalties. Following Simpson et al. (2017), we used complexity-penalising priors for the Negative Binomial overdispersion parameters (which are used by default in MVGAM to penalise an observation model towards a Poisson if there is minimal support for overdispersion). For MVGAM models, we ran four MCMC chains using Stan's Hamiltonian Monte Carlo sampler for 1000 iterations as warmup and collected 4000 samples from the joint posterior. Convergence of chains was checked with the Gelman–Rubin diagnostic (Gelman & Rubin, 1992) and by visual inspection of posterior chains.

The relative performances of each model were explored using out of sample forecasts. We trained models on the first 5 years of data (60 observations) and generated forecasts for the remaining year (12 observations). Probabilistic forecast performance was evaluated using a discrete version of the Rank Probability Score (DRPS; Gneiting & Raftery, 2007) and coverage of 90% prediction intervals. Forecasts with lower DRPS and coverage closer to 0.9 were considered more accurate.

5 | CASE STUDY: FORECASTING TICK ABUNDANCES

Amblyomma americanum and *Ixodes scapularis* are two widespread species of hard ticks capable of transmitting a diversity of parasites to animals and humans, many of which are zoonotic (Rochlin & Toledo, 2020). Due to the medical and ecological importance of these species, a common goal is to understand factors that influence their abundances. The National Ecological Observatory Network (NEON) carries out standardised long-term monitoring of tick abundances as well as other important indicators of ecological change (Thorpe et al., 2016). Nymphal abundances of both tick species are routinely recorded across NEON plots by drag cloth sampling, with plots nested within sites (Springer et al., 2016). These plot-level series show strong seasonality and incorporate many of the challenging features associated with ecological data including overdispersion, high proportions of missingness and irregular sampling in time, making them useful for exploring the utility of DGAMs.

Temperatures between −5°C and 5°C can affects various components of tick physiological diapause and host-seeking behaviours (Clark, 1995). We included a cumulative growing degree day (cum_gdd) variable using temperature records for each site's nearest weather station from NOAA's Daily Global Historical Climatology Network daily database as a covariate. The predictor was calculated as the total number of days up to the start of the tick season (1st June) in which the mean of the day's maximum and

minimum temperatures was above 0°C. We fit species-specific DGAMs to 4 years of data (2015–2018) for 17 *A. americanum* plots (nested in seven NEON sites) and for eight *I. scapularis* plots (nested in three sites) using the most recent release of the NEON tick drag sampling product (National Ecological Observatory Network, 2022). Counts of ticks were aggregated at the temporal resolution of epidemiological week, a standardised method of counting weeks developed by the US Centers for Disease Control and Prevention to facilitate direct comparisons across years. Time points during winter (epidemiological weeks 1–14 and 41–53) had entirely missing observations as no sampling occurred during this period, but we kept these in the model as missing data. For each species, we fit four models representing different hypothetical dynamics, though we caution that our goal here was not to carry out a rigorous analysis but to highlight how DGAMs could be used to facilitate model selection and scrutiny:

- Null: There is no seasonality, rather the latent factors/random site-level effects of cum_gdd fully influence the dynamics for the plot-level series. We hypothesised that the site-specific partial effects of cum_gdd could be mildly nonlinear, so we set $k = 5$ for this smooth function. Formula in R syntax: $y \sim s(site, bs = 're') + s(cum\_gdd, site, k = 5, bs = 'fs') + Z$
- Hyp1: All plots share a seasonal pattern, with any remaining variation captured by the latent factors and site-level cum_gdd effects. In addition to the assumption of cum_gdd nonlinearity, we assumed the global seasonal pattern was moderately nonlinear and flexible enough to capture the characteristic double peaks commonly seen in hard tick nymph abundance survey time series (Wallace et al., 2019), and we assumed the seasonal function was cyclic with equal values between the end of December and the beginning of January. Formula: $y \sim s(site, bs = 're') + s(cum\_gdd, site, k = 5, bs = 'fs') + s(season, k = 12, m = 2, bs = 'cc') + Z$
- Hyp2: as above but with hierarchical seasonality, including a global seasonality smooth function and a seasonal smooth function that can deviate from the global seasonality across each site. Formula: $y \sim s(site, bs = 're') + s(cum\_gdd, site, k = 5, bs = 'fs') + s(season, k = 12, m = 2, bs = 'cc') + s(season, site, m = 1, k = 6, bs = 'fs') + Z$
- Hyp3: as above but the seasonal deviations occur at the bottom level of aggregation (plot rather than site level). Formula: $y \sim s(site, bs = 're') + s(cum\_gdd, site, k = 5, bs = 'fs') + s(season, k = 12, m = 2, bs = 'cc') + s(season, plot, m = 1, k = 4, bs = 'fs') + Z$

We used random walk dynamic factor models ($M = 4$ for *Ixodes* and 5 for *Amblyomma*) for the temporal evolution and assumed a Poisson distribution for the observations. Each model was estimated using four MCMC chains for 1000 iterations as warmup. We collected 4000 posterior samples to evaluate parameter estimates and inspect forecasts. The 2019 observations for each plot were held out as testing data to evaluate forecasts using the same evaluation criteria as in the simulations above.

6 | RESULTS

6.1 | DGAM forecast performance—simulation results

Our simulations explored the relative forecast performance of DGAMs versus static and autoregressive GAMs. The seasonal DGAM outperformed its GAM counterparts in terms of DRPS, providing better probabilistic forecasts in all comparisons (Figure 2). As expected, the correctly specified seasonal DGAM was the best performer when the trend dynamics were moderate compared to the seasonal magnitude, while the nonseasonal DGAM performed nearly as well under strong trend dynamics (Figure 2). The static and autoregressive seasonal GAMs were the worst performers in nearly all comparisons (Figure 2). Results were similar when inspecting DRPS as a function of the number of series in the simulation, with DGAMs clearly providing better probabilistic forecasts (Figure S2).

Comparisons of 90% interval coverages strongly favoured the two DGAMs (Figure 3). Intervals for the DGAMs frequently included 25%–35% more of the out of sample observations than did the intervals for the two GAMs. There was little distinction between the two DGAMs, even as the number of series and the strength of the underlying dynamics increased (Figure 3). Results were similar when inspecting 90% interval coverage as a function of missingness, with the DGAMs strongly outperforming the GAMs (Figure S3).

6.2 | DGAM and NEON tick abundance forecasts

Our results suggested that Hyp3, which captured hierarchical seasonality by allowing individual plot-level seasonal patterns to deviate from a global seasonality function, was the best-performing model when forecasting *I. scapularis* nymphal abundance across NEON sites, while the null model that did not include seasonality was the worst performing (Figure 4). Nominal coverages of 90% intervals were accurate for the three seasonal models (ranging from 87% to 88%), while the intervals for the null model were generally wider than they needed to be (97% coverage; Figure 4). However, there was variation across plots in terms of forecast performance, suggesting that an ensemble forecast (which combines forecasts from multiple models) could improve performance (Figure S4). Inspection of probability integral transform (PIT) histograms, which should be uniform if predictions are evenly distributed about the truth (Simonis et al., 2021), revealed that all models apart from the null tended to underpredict to some degree (left-skewed PIT histograms; Figure S5). Figure 5 shows example MVGAM visualisations for a single plot, including estimated smooth functions, forecasts and dynamic trend estimates (along with their probabilistic uncertainties). When conditioning on seasonality and the trend, *I. scapularis* abundances demonstrated no apparent association with variation in cumulative growing degree days (Figure 5). Inspection of the latent dynamic components for the

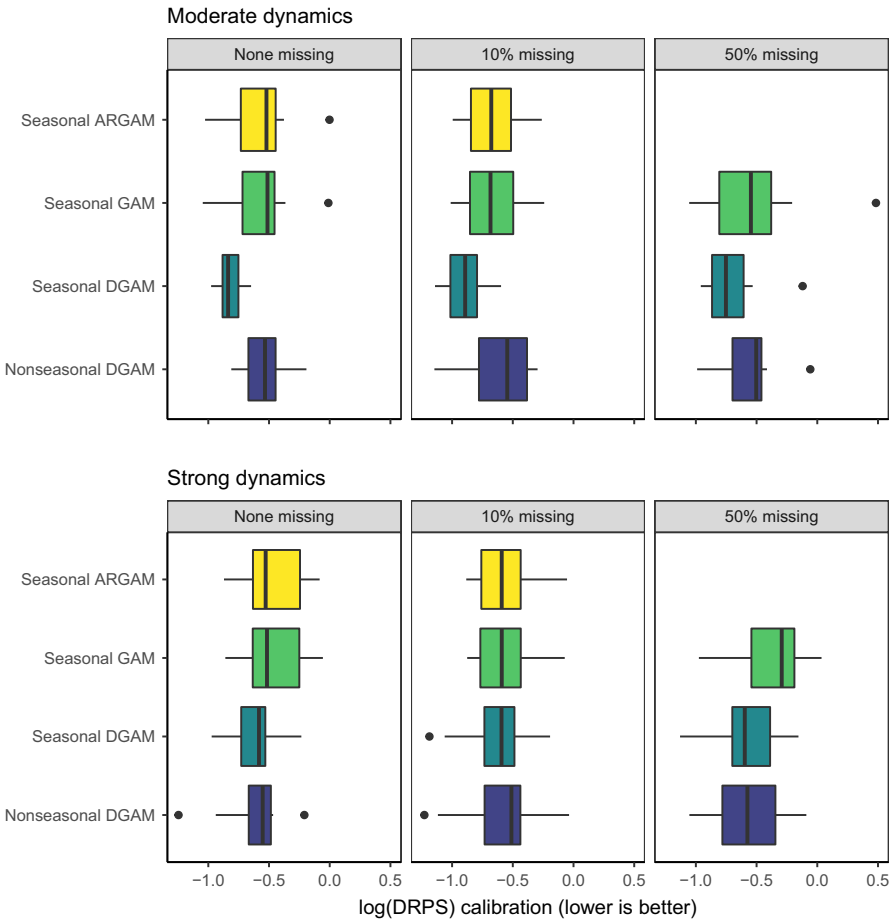

FIGURE 2 Log(discrete rank probability score) (DRPS) performance for out of sample forecasts from competing models fitted to sets of simulated discrete time series. Panels depict models fitted with different levels of data missingness (proportion of observations set to NA) and temporal dynamics strength. The seasonal GAM was fitted using R package MGCV, while the seasonal and nonseasonal DGAMs were fitted using the MVGAM package (using the Hamiltonian Monte Carlo software Stan). Lower scores indicate better model performance.

three seasonal models revealed positive within-site correlations for sites SCBI and SERC (Figure S6). Example MVGAM visualisations of posterior checks for training (retrodictive) and forecast periods (predictive), useful for checking if a model is capable of simulating time series that resemble key aspects of the observed data without notable discrepancies, are shown in Figure S7. Examples highlighting how smooth function and trend realisations can be plotted, which can improve model interpretation over quantile or density plots, are shown in Figure S8.

In agreement with the *I. scapularis* models, *A. americanum* abundance was also best predicted by the Hyp3 model. Example visualisations of estimated plot-level seasonal functions are shown in Figure 6, while a visualisation of estimated random effect intercept distributions is shown in Figure S9. Our model estimated that tick abundances in some plots (i.e. SERC_001) tended to show earlier peaks around epidemiological week 24, while abundance in other plots (i.e. TALL_001) followed a broader curve with a peak around epidemiological week 30 (Figure 6).

6.3 | Quantifying uncertainty contributions among MVGAM model components

In addition to plotting smooth functions and forecasts, MVGAM offers utilities to compute relative contributions of the latent dynamic and GAM components to forecast uncertainty. This process of partitioning uncertainty is an important step in analysing a model's forecasts to diagnose the main drivers of prediction uncertainty and prioritise aspects of models or data that require further investigation (Dietze, 2017; Heilman et al., 2022). Comparisons of uncertainty contributions for four of the *A. americanum* forecasts indicate that both components contribute to forecast uncertainty, but to varying degrees over time and across plots (Figure 7). However, across all plots, dynamic trend uncertainty tended to increase over time, becoming relatively more important during the peak tick season (3–22 weeks ahead).

7 | DISCUSSION

We have introduced an R package for fitting Bayesian Dynamic GAMs (DGAMs) that incorporate the flexibility of the widely popular penalised smoothing functions in MGCV with latent dynamic components for analysing and forecasting discrete time series. Keys to MVGAM's performance are its ability to cope with substantial missing data, scale to large collections of discrete time series and provide robust uncertainty quantification. In recent years, there has been increased interest in using time-series models for uncertainty interval estimation as opposed to point predictions, a trend that lends well to Bayesian inference (Gelman et al., 2017; Makridakis et al., 2020). This is particularly relevant for ecological forecasts,

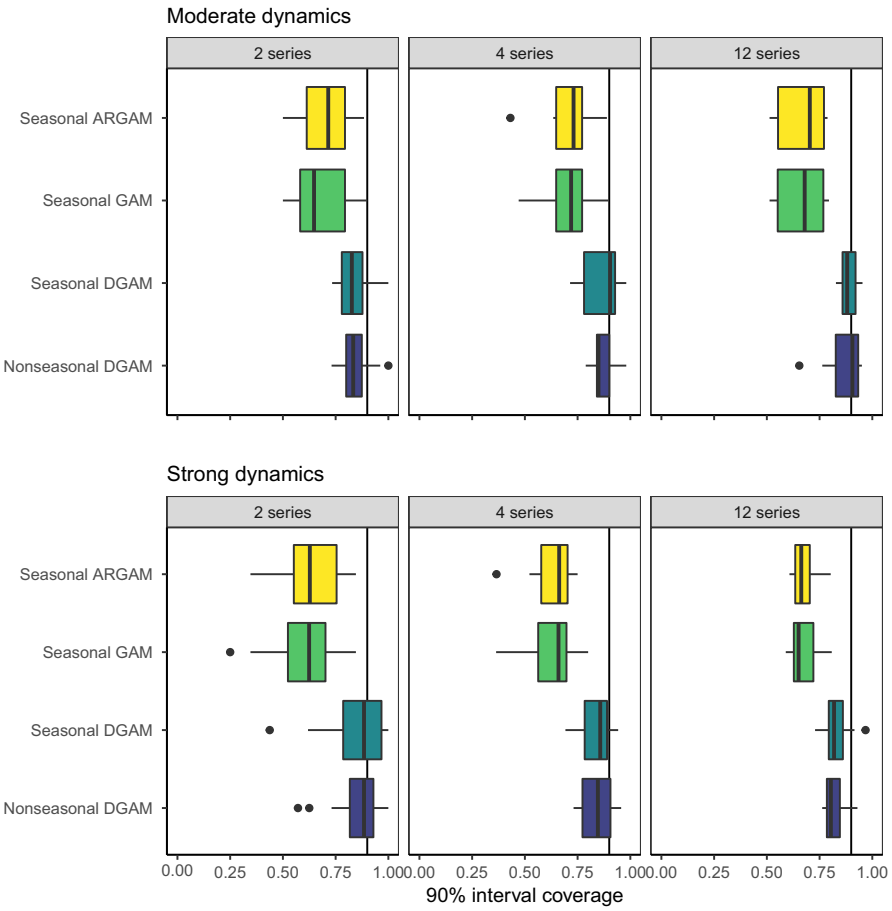

FIGURE 3 90% interval coverage for out of sample forecasts from competing models fitted to sets of simulated discrete time series, plotted as a function of dimensionality (total number of series) and dynamics strength. The vertical line in each plot marks a coverage of 0.9. The GAM was fitted using R package MGCV, while the DGAMs were fitted using the MVGAM package. Scores closer to 0.9 are better.

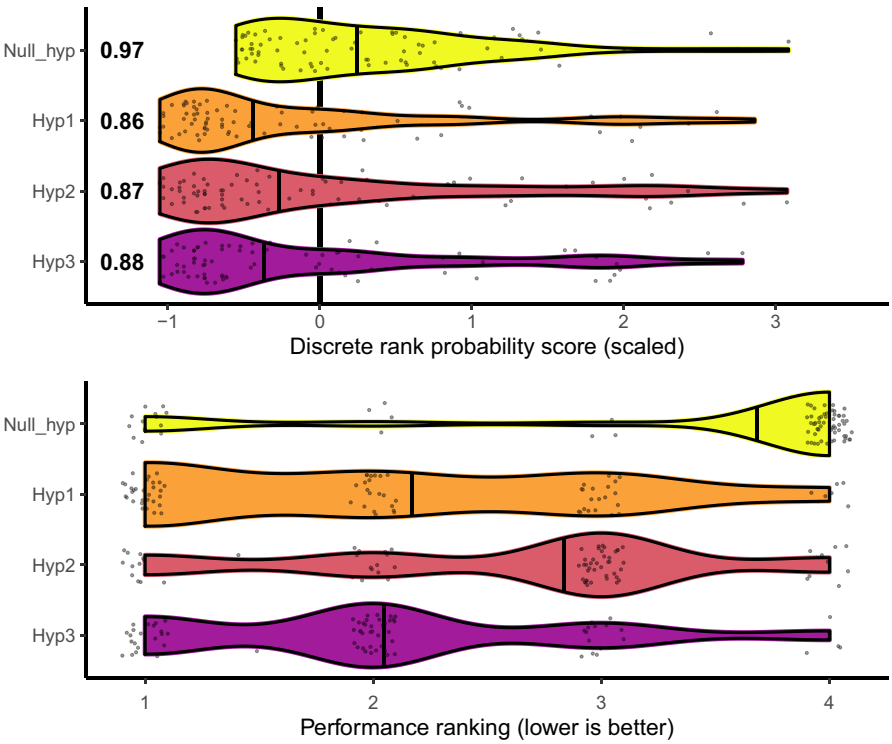

FIGURE 4 Forecast performance rank distributions based on out of sample discrete rank probability score for four competing models fitted to NEON's *Ixodes scapularis* abundance series. Numbers on the left-hand side of the top plot indicate coverages of 90% posterior predictive intervals. Thick black lines show medians. Hypothesis definitions are outlined in section CASE STUDY: FORECASTING TICK ABUNDANCES.

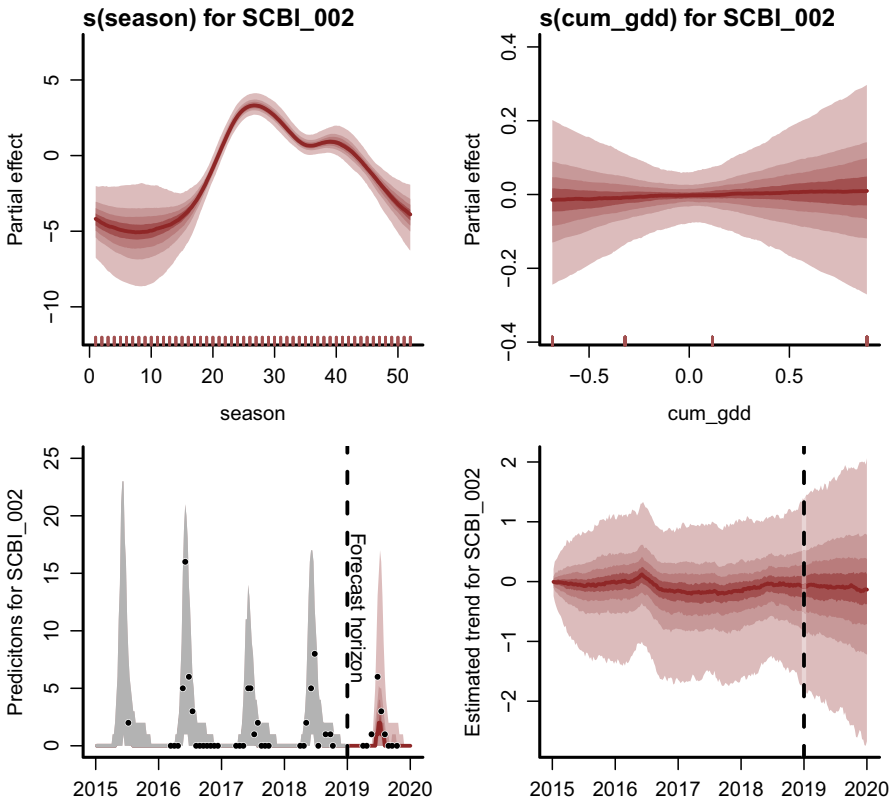

FIGURE 5 Visualisations from the best-performing MVGAM model (Hyp3) for a single *Ixodes scapularis* plot (SCBI_002). Top left, the estimated seasonal smooth function; top right, estimated cumulative growing degree days function; bottom left, predicted tick abundances over time (observed values shown as black points); bottom right, estimated latent dynamic component. For all plots shading shows posterior empirical quantiles. Hypothesis definitions are outlined in section CASE STUDY: FORECASTING TICK ABUNDANCES.

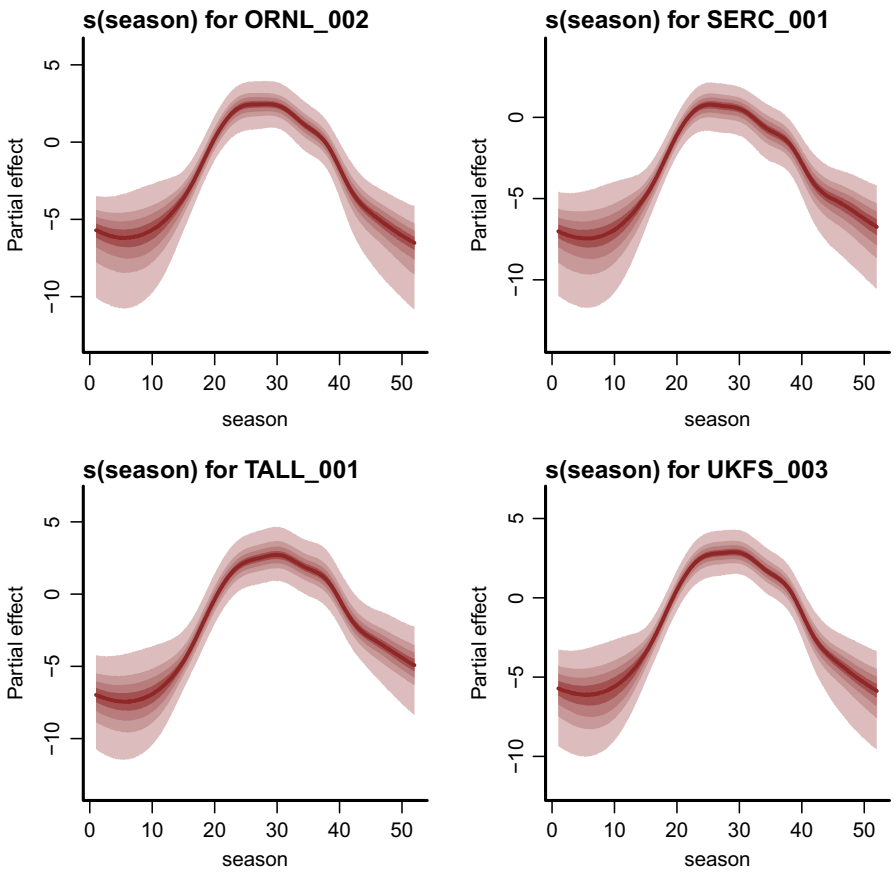

FIGURE 6 Output from the *plot_mvgam_smooth* function in *mvgam* showing seasonal smooth functions for four *Amblyomma americanum* plots estimated from a dynamic GAM with hierarchical seasonality. Shading shows posterior empirical quantiles.

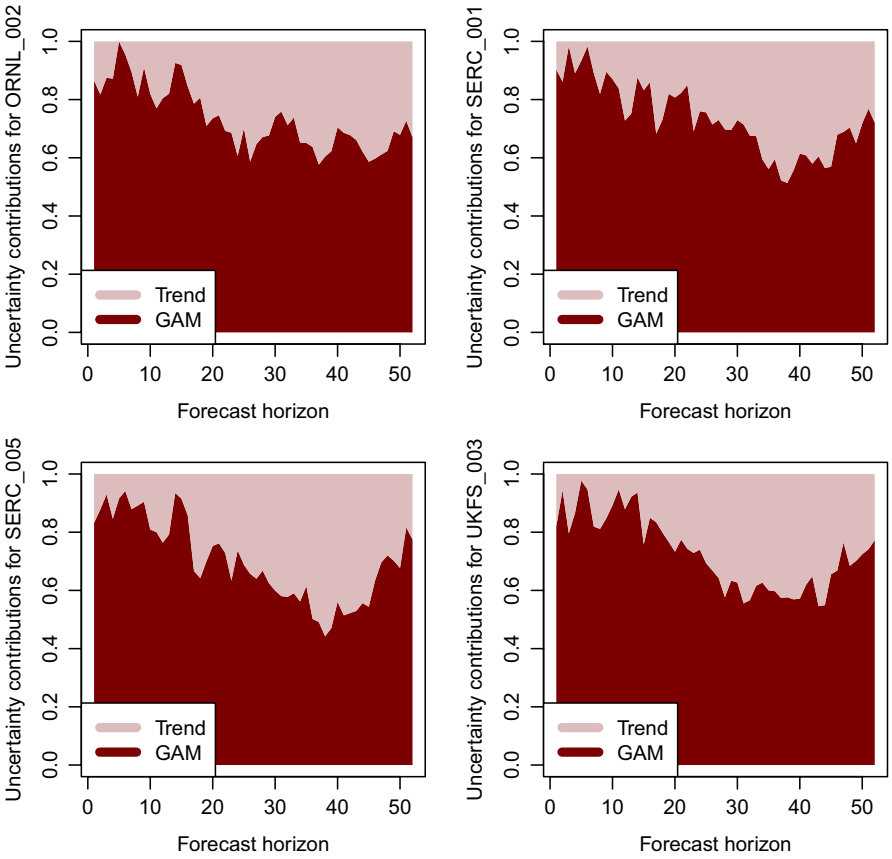

FIGURE 7 Output from the *plot_mvgam_uncertainty* function in *mvgam* showing relative contributions of the dynamic temporal (grey) and GAM (red) components to forecast uncertainty for four *Amblyomma americanum* plots estimated from a dynamic GAM with hierarchical seasonality. Forecast horizons were varied over a '1-year' horizon (52 weeks matching data availability).

where point estimates are less important for making informed decisions than are conditional probability statements (Clark et al., 2022; Dietze, 2017; Dietze et al., 2018; White et al., 2019).

Notably, JAGS or Stan model files, together with all data necessary to condition the model, are made available to the user in MVGAM, allowing an enormous diversity of bespoke models to be implemented through addition of other stochastic or hierarchical elements. The case studies (available in Appendices S1–S3, online at https://rpubs.com/NickClark47/mvgam, https://rpubs.com/NickClark47/mvgam2, https://rpubs.com/NickClark47/mvgam3 and at a permanently archived Zenodo repository (https://doi.org/10.5281/zenodo.6918047; Clark & Wells, 2022)) discuss a range of models that can be fitted and interrogated with MVGAM, while an example *JAGS* model file complete with automatic descriptions of required data structures is shown in Appendix S5. While our intention is that researchers use MVGAM as a backbone to simplify the task of preparing smoothing splines for more bespoke joint probability models, we do see several avenues for improving model flexibility and estimation. These include but are not limited to:

- Recommending and accommodating ways for users to include principled prior specifications for the behaviours of nonlinear smooth functions (Simpson et al., 2017)
- The inclusion of observation model options for modelling continuous, proportional or other non-integer valued time series
- The addition of other structured latent temporal components, such as multivariate random walks, hierarchical Gaussian processes and other models of stochastic processes, to increase the diversity of models that can be interrogated using MVGAM
- The addition of Markov-switching processes to allow dynamic factor loadings to be drawn from different sets of correlation 'regimes', allowing correlation structures to change over time in a principled way (Fox et al., 2010)
- The incorporation of covariates into the latent temporal models (i.e. as dynamic linear models) to explicitly address broader hypotheses about the factors that influence temporal dynamics (Heilman et al., 2022)

7.1 | Challenges in estimating DGAM parameters

The joint estimation of smoothing parameters, basis coefficients, latent trend variances or overdispersion parameters is not without its challenges (Wood, 2016). Posterior geometries for such high-dimensional models can become complex enough that traditional MCMC samplers based on Random Walk proposals (e.g. Gibbs samplers) will not be able to sample the parameter space without reverting to painfully small step sizes that result in high posterior autocorrelation and very slow exploration (Betancourt, 2017). Maximum likelihood and related estimators will not likely produce better uncertainty quantification, as verifying how and when posterior geometries can be accurately approximated under an asymptotic regime is a huge and elusive challenge. MVGAM's exploitation of Hamiltonian Monte Carlo is a major advantage for tackling DGAM parameter estimation, but we stress that there is no one-size-fits-all default

solution for prior modelling. Indeed, while choice of priors is important in any Bayesian analysis, in DGAMs, it is particularly crucial for ensuring the latent trend and observation models do not compete to induce further complexity in the joint posterior. In MVGAM, informative priors for parametric terms (i.e. intercepts and additive linear covariate effects) are guided based on 50 steps of penalised iteratively re-weighted least squares from a comparable non-dynamic model using MGCV, while suitable priors for operating on the log scale are used for latent trend parameters such as drift, AR and variance parameters. Together this prior combination works well in most cases, especially because of the convenience of the link-scaled latent trends. Run times in our simulations and empirical examples took 1–20 minutes to reach effective sample sizes >800 for all parameters on an Intel(R) Core(TM) i5-8500 CPU with 32Gb RAM and six processing cores. Nevertheless, priors in any Bayesian analysis should be carefully considered and inferences interrogated with appropriate prior sensitivity analyses (Gelman et al., 2020). One illuminating situation that we have encountered is the difficulty in jointly estimating a latent trend and overdispersion parameters such as in the Negative Binomial or Tweedie distributions. This is because both processes (overdispersion and autocorrelation) may be able to explain the dispersion around the mean, particularly when using Random Walk or AR trends that can jump around easily. Users will need to use theory and judgement to decide how to tackle these challenges, for example, by assuming there is overdispersion in the observation process (with consultation from appropriate references; i.e. Bliss & Fisher, 1953, Lindén & Mäntyniemi, 2011) but that the trend is smooth, in which case a latent Gaussian process with suitable prior on the length scale would be appropriate. Smoothing splines are also challenging in a way because they do not readily facilitate principled prior modelling, where expert elicitation could help to constrain prior function shapes towards those that are compatible with domain expertise as part of a Bayesian workflow (Betancourt, 2021; Gelman et al., 2020). Users are recommended to refer to the wealth of material relating to the MGCV package for choosing a smoothing basis and basis dimension that are compatible with expected function shapes (Wood, 2004, 2013, 2017).

8 | CONCLUSION

The R package MVGAM provides a user-friendly tool for researchers and practitioners interested in fitting DGAMs to analyse and forecast ecological time series. The problems associated with smooth spline extrapolation are not limited to ecology however, as the need to forecast sets of discrete nonlinear time series is a common challenge in areas as diverse as speech recognition, tourism demand, natural language processing and finance (Hyndman & Athanasopoulos, 2018; Makridakis et al., 2018). Beyond the examples showcased here, the package can be especially useful to identify avenues for model improvement via its ability to assimilate new observations online to update forecast distributions (showcased in Appendix S1). With growing interest in both the application of hierarchical GAMs to ecological problems and the need to use iterative forecasts to make ecology a more predictive discipline, MVGAM can become a vital addition to the applied ecologist's analytical toolbox.

AUTHOR CONTRIBUTIONS

Nicholas J. Clark involved in conceptualization, data curation, project administration, software, visualisation, writing—original draft. Nicholas J. Clark and Konstans Wells involved in formal analysis, methodology, validation, writing—review & editing.

ACKNOWLEDGEMENTS

NOAA temperature data were supplied by Daniel Ruiz-Carrascal as part of the 2021 NEON Ecological Forecasting Challenge. This research was funded by an ARC DECRA fellowship to N. Clark (DE210101439).

CONFLICT OF INTEREST

None of the authors have a conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13974.

DATA AVAILABILITY STATEMENT

The manuscript uses data that are archived by the National Ecological Observatory Network (https://data.neonscience.org/). The data have been downloaded and converted into a usable format for modelling, and this version of the data is available, along with an archived version of the MVGAM R package and all R scripts used to produce analyses in this manuscript (and in the Appendices), at a permanently archived Zenodo repository (https://doi.org/10.5281/zenodo.6918047; Clark & Wells, 2022).

ORCID

Nicholas J. Clark https://orcid.org/0000-0001-7131-3301
Konstans Wells https://orcid.org/0000-0003-0377-2463

REFERENCES

Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A. A., Leos-Barajas, V., Mills Flemming, J., Nielsen, A., & Petris, G. (2021). A guide to state–space modeling of ecological time series. *Ecological Monographs*, *91*, e01470.

Bartumeus, F. (2007). Lévy processes in animal movement: An evolutionary hypothesis. *Fractals*, *15*, 151–162.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434.

Betancourt, M. (2021). Prior modelling. Retrieved from: https://github.com/betanalpha/knitr_case_studies/tree/master/prior_modeling, commit 56606fa62e35f87bc88cec6892b4a4d3587f7029.

Bhattacharya, A., & Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, *98*, 291–306.

Bliss, C. I., & Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, *9*, 176–200.

Brezger, A., Kneib, T., & Lang, S. (2005). BayesX: Analyzing Bayesian structural additive regression models. *Journal of Statistical Software*, *14*, 1–22.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Camara, A. J. A., Franco, G. C., Reisen, V. A., & Bondon, P. (2021). Generalized additive model for count time series: An application to quantify the impact of air pollutants on human health. *Pesquisa Operacional*, *41*, 1–15.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*. Retrieved from https://www.osti.gov/biblio/1430202

Carrasco, M., & Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, *18*, 17–39.

Choler, P., Michalet, R., & Callaway, R. M. (2001). Facilitation and competition on gradients in alpine plant communities. *Ecology*, *82*, 3295–3308.

Clark, D. D. (1995). Lower temperature limits for activity of several ixodid ticks (Acari: Ixodidae): Effects of body size and rate of temperature change. *Journal of Medical Entomology*, *32*, 449–452.

Clark, N. J., Kerry, J. T., & Fraser, C. I. (2020). Rapid winter warming could disrupt coastal marine fish community structure. *Nature Climate Change*, *10*, 862–867. https://doi.org/10.1038/s41558-41020-40838-41555

Clark, N. J., Proboste, T., Weerasinghe, G., & Soares Magalhães, R. J. (2022). Near-term forecasting of companion animal tick paralysis incidence: An iterative ensemble model. *PLoS Computational Biology*, *18*, e1009874.

Clark, N. J., & Wells, K. (2022). Data from: Dynamic generalized additive models (DGAMs) for forecasting discrete ecological time series. *Zenodo*. https://doi.org/10.5281/zenodo.6918047

De Stefani, J., Le Borgne, Y.-A., Caelen, O., Hattab, D., & Bontempi, G. (2019). Batch and incremental dynamic factor machine learning for multivariate and multi-step-ahead forecasting. *International Journal of Data Science and Analytics*, *7*, 311–329.

Dietze, M. C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, *27*, 2048–2060.

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., Keitt, T. H., Kenney, M. A., Laney, C. M., & Larsen, L. G. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences of the United Statesof America*, *115*, 1424–1432.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, *5*, 236–244.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. OUP Oxford.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*, 330–342.

Fox, E., Jordan, M., Sudderth, E., & Willsky, A. (2009). Sharing features among dynamical systems with beta processes. *Advances in Neural Information Processing Systems*, *22*, 549–557.

Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2010). Bayesian nonparametric methods for learning Markov switching processes. *IEEE Signal Processing Magazine*, *27*, 43–54.

Gabry, J., & Češnovar, R. (2021). Cmdstanr: R interface to 'CmdStan'. https://mc-stan.org/cmdstanr

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*, 389–402.

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

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. B. (2017). *Bayesian Data Analysis* (3rd ed.). CRC Press.

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

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian workflow. arXiv preprint arXiv:2011.01808.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Guisan, A., Edwards, T. C., Jr., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*, 89–100.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Taylor & Francis.

Heilman, K. A., Dietze, M. C., Arizpe, A. A., Aragon, J., Gray, A., Shaw, J. D., Finley, A. O., Klesse, S., DeRose, R. J., & Evans, M. E. K. (2022). Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty. *Global Change Biology*, *28*, 2442–2460.

Hughes, T. P., Kerry, J. T., Baird, A. H., Connolly, S. R., Dietzel, A., Eakin, C. M., Heron, S. F., Hoey, A. S., Hoogenboom, M. O., Liu, G., McWilliam, M. J., Pears, R. J., Pratchett, M. S., Skirving, W. J., Stella, J. S., & Torda, G. (2018). Global warming transforms coral reef assemblages. *Nature*, *556*, 492–496.

Hui, F. K. (2016). Boral–Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution*, *7*, 744–750.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*, 1–22.

Intergovernmental Panel on Climate Change. (2018). Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.

Kaplan, I. C., Williams, G. D., Bond, N. A., Hermann, A. J., & Siedlecki, S. A. (2016). Cloudy with a chance of sardines: Forecasting sardine distributions using regional climate models. *Fisheries Oceanography*, *25*, 15–27.

Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S., & Kiesecker, J. (2019). Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology*, *25*, 811–826.

Knape, J. (2016). Decomposing trends in Swedish bird populations using generalized additive mixed models. *Journal of Applied Ecology*, *53*, 1852–1861.

Koolhof, I. S., Firestone, S. M., Bettiol, S., Charleston, M., Gibney, K. B., Neville, P. J., Jardine, A., & Carver, S. (2021). Optimising predictive modelling of Ross River virus using meteorological variables. *PLoS Neglected Tropical Diseases*, *15*, e0009252.

Kowal, D. R., & Canale, A. (2020). Simultaneous transformation and rounding (STAR) models for integer-valued data. *Electronic Journal of Statistics*, *14*, 1744–1772.

Letten, A. D., Keith, D. A., Tozer, M. G., & Hui, F. K. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, *103*, 1264–1275.

Levin, S. A. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems*, *1*, 431–436.

Lindén, A., & Mäntyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, *92*, 1414–1421.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*, 802–808.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M5 accuracy competition: Results, findings and conclusions. *International Journal of Forecasting*, *36*, 224–227.

Malick, M. J., Siedlecki, S. A., Norton, E. L., Kaplan, I. C., Haltuch, M. A., Hunsicker, M. E., Parker-Stetter, S. L., Marshall, K. N., Berger, A. M., & Hermann, A. J. (2020). Environmentally driven seasonal forecasts of Pacific hake distribution. *Frontiers in Marine Science*, *7*, 844.

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, *55*, 2372–2387.

Massoud, E. C., Huisman, J., Benincà, E., Dietze, M. C., Bouten, W., & Vrugt, J. A. (2018). Probing the limits of predictability: Data assimilation of chaotic dynamics in complex food webs. *Ecology Letters*, *21*, 93–103.

Miller, D. L. (2019). Bayesian views of generalized additive modelling. arXiv preprint arXiv:1902.01330.

National Ecological Observatory Network. (2022). *Ticks sampled using drag cloths (DP1.10093.001)*. National Ecological Observatory Network (NEON). Dataset accessed from https://data.neonscience.org on Feb 1, 2022.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, *20*, 561–576.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (p. 125). Technische Universit at Wien Wien.

Riutort-Mayol, G., Bürkner, P.-C., Andersen, M. R., Solin, A., & Vehtari, A. (2020). Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming. arXiv preprint arXiv:2004.11408.

Rochlin, I., & Toledo, A. (2020). Emerging tick-borne pathogens of public health importance: A mini-review. *Journal of Medical Microbiology*, *69*, 781–791.

Schmidt, K. A., Dall, S. R., & Van Gils, J. A. (2010). The ecology of information: An overview on the ecological significance of making informed decisions. *Oikos*, *119*, 304–316.

Simonis, J. L., White, E. P., & Ernest, S. K. M. (2021). Evaluating probabilistic ecological forecasts. *Ecology*, *102*, e03431.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*, 1–28.

Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. *Frontiers in Ecology and Evolution*, *6*, 149.

Spooner, F. E., Pearson, R. G., & Freeman, R. (2018). Rapid warming is associated with population decline among terrestrial birds and mammals globally. *Global Change Biology*, *24*, 4521–4531.

Springer, Y. P., Hoekman, D., Johnson, P. T., Duffy, P. A., Hufft, R. A., Barnett, D. T., Allan, B. F., Amman, B. R., Barker, C. M., & Barrera, R. (2016). Tick-, mosquito-, and rodent-borne parasite sampling designs for the National Ecological Observatory Network. *Ecosphere*, *7*, e01271.

Thorpe, A. S., Barnett, D. T., Elmendorf, S. C., Hinckley, E. L. S., Hoekman, D., Jones, K. D., LeVan, K. E., Meier, C. L., Stanish, L. F., & Thibault, K. M. (2016). Introduction to the sampling designs of the National Ecological Observatory Network Terrestrial Observation System. *Ecosphere*, *7*, e01627.

Thorson, J. T., Ianelli, J. N., Larsen, E. A., Ries, L., Scheuerell, M. D., Szuwalski, C., & Zipkin, E. F. (2016). Joint dynamic species distribution models: A tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, *25*, 1144–1158.

Tobler, M. W., Kéry, M., Hui, F. K., Guillera-Arroita, G., Knaus, P., & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, *100*, e02754.

Umlauf, N., Klein, N., & Zeileis, A. (2018). BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, *27*, 612–627.

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. UN Publishing.

Wallace, D., Ratti, V., Kodali, A., Winter, J. M., Ayres, M. P., Chipman, J. W., Aoki, C. F., Osterberg, E. C., Silvanic, C., & Partridge, T. F. (2019). Effect of rising temperature on Lyme disease: *Ixodes scapularis* population dynamics and borrelia burgdorferi transmission and prevalence. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *2019*, 1–15.

Ward, E. J., Anderson, S. C., Hunsicker, M. E., & Litzow, M. A. (2021). Smoothed dynamic factor analysis for identifying trends in multivariate time series. *Methods in Ecology and Evolution*, *13*, 908–918.

Ward, E. J., Holmes, E. E., Thorson, J. T., & Collen, B. (2014). Complexity is costly: A meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos*, *123*, 652–661.

Warton, D. I. (2018). Why you cannot transform your way out of trouble for small counts. *Biometrics*, *74*, 362–368.

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*, *30*, 766–779.

Wells, K., O'Hara, R. B., Cooke, B. D., Mutze, G. J., Prowse, T. A., & Fordham, D. A. (2016). Environmental effects and individual body condition drive seasonal fecundity of rabbits: Identifying acute and lagged processes. *Oecologia*, *181*, 853–864.

Welty, L. J., Peng, R. D., Zeger, S. L., & Dominici, F. (2009). Bayesian distributed lag models: Estimating effects of particulate matter air pollution on daily mortality. *Biometrics*, *65*, 282–291.

White, E. P., Yenni, G. M., Taylor, S. D., Christensen, E. M., Bledsoe, E. K., Simonis, J. L., & Ernest, S. M. (2019). Developing an automated iterative near-term forecasting system for an ecological study. *Methods in Ecology and Evolution*, *10*, 332–344.

Wood, S. (2016). Just another Gibbs additive Modeller: Interfacing JAGS and mgcv. *Journal of Statistical Software*, *75*, 1–15.

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*, 673–686.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, *100*, 221–228.

World Health Organization. (2005). *Using climate to predict infectious disease epidemics*. WHO Press.

Yang, L., Qin, G., Zhao, N., Wang, C., & Song, G. (2012). Using a generalized additive model with autoregressive terms to study the effects of daily temperature on mortality. *BMC Medical Research Methodology*, *12*, 1–13.

Zurell, D., Elith, J., & Schröder, B. (2012). Predicting to new environments: Tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions*, *18*, 628–634.

SUPPORTING INFORMATION

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

How to cite this article: Clark, N. J., & Wells, K. (2022). Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. *Methods in Ecology and Evolution*, *00*, 1–14. https://doi.org/10.1111/2041-210X.13974