10.1071/WR22119

Wildlife Research

Supplementary Material

Satellite-based environmental variables complement traditional variables in spatio-temporal models of purple martin migration

Jason R. Courter[A,*], *Zhen Liu*[B], *Naresh Neupane*[C], *Ali Arab*[B], *and Joe Siegrist*[D]

[A]Department of Natural Sciences, Malone University, Canton, OH, USA.

[B]Department of Mathematics and Statistics, Georgetown University, Washington, DC, USA.

[C]Department of Biology, Georgetown University, Washington, DC, USA.

[D]Purple Martin Conservation Association, 301 Peninsula Drive, Suite 6, Erie, PA 16505, USA.

[*]Correspondence to: Jason R. Courter Department of Natural Sciences, Malone University, Canton, OH, USA Email: jcourter@malone.edu

Appendices

Appendix 1

We model first arrival dates using a generalized linear model as,

$$Y_t \sim N(\mu_t, D_t), \tag{1}$$

With

$$\mu_{i,t} = \beta_0 + X'_{i,t}\beta + \theta_i + e_t, \tag{2}$$

$$\log \sigma_{i,t} = a_0 + \eta_i + \nu_t + \kappa \cdot x^s_{i,t}, \tag{3}$$

$$e_t = \varphi \cdot e_{t-1} + \epsilon, \tag{4}$$

where $Y_t = \left(Y_{1,t}, \cdots, Y_{n,t}\right)'$denotes the vector of mean arrival dates for the grid cells $(i \in \{1, \cdots, n\}; n = 23)$ in our case over years in the study $(t \in \{1, \cdots, T\}; T = 10)$ for years 2001-2018. $Y_{i,t}$ represents the mean arrival dates for the *i*th grid cell in the year *t*. $\mu_t = \left(\mu_{1,t}, \cdots, \mu_{n,t}\right)'$ denotes the vector of the mean arrival process, and $D_t = diag\left(\sigma^2_{1,t}, \sigma^2_{2,t}, \cdots, \sigma^2_{n,t}\right)$ denotes the error variance-covariance matrix for year *t*. $\theta = (\theta_1, \cdots, \theta_n)$ denotes the spatially-varying mean. β_0 is the shared "overall mean", $\beta = (b_1, \cdots, b_K)$ represents the vector of regressors with elements for K predictor variables and e_t denotes the temporal trend elements. Here we propose an AR(1) indicating autoregressive model of order 1 to explain dependence among temporal components of the model in Equation 4. $X_{i,t}$ denotes the vector of predictor variables for the *i*th grid cell. Our predictor variables include MODISgreenup, OnsetGDD, temperature from February to April, precipitation from January to April, and sampling effort. All the predictor variables were centered in advance for the computational convenience.

We capture the variation for each cell, in each year, based on constant, spatial and temporal decomposition of standard deviation. $a_0 \sim N(0, \sigma^2_a)$denotes the shared log error standard deviation. η_i represents the change of log error standard deviation for the *i*th grid cell. $\nu_t \sim N(0, \sigma^2_\nu)$ denotes the log error standard deviation for time t. κ denotes the corresponding regression coefficient for $x_{i,t}^{s}$. In our case, sampling effort serves as the only predictor variable for the error standard deviation.

We consider spatial structure for the geographical parameter $\theta_i$based on conditional autoregressive modeling (CAR):

$$\theta_l \mid \theta_{-l}, \tau_l^2 \sim N\left(\sum_{s \in N_l} c_{ls}\theta_s, \tau_l^2\right), \tag{5}$$

where $l, s = 1, \cdots, n$ and θ_{-l} denotes the vector of spatially-varying parameters for all grid cells except the lth grid cell. Also, is the set of neighboring sites for l, and c_{ls}'s are weights defined such that $c_{ls} = 1$ for $l \neq s, c_{qq} = 0$ for $q = 1 \cdots, n$ and $c_{ls}\tau_l^2 = c_{sl}\tau_s^2$. τ_l^2 denote precision parameters and are commonly assumed to be the same and equal to τ^2.

Similarly, we propose $\eta_i \sim CAR(\tau_\eta^2)$, where $\eta = (\eta_1, \cdots, \eta_n)$ is the vector of spatially-varying components for the log measurement error standard deviation, and τ_η^2 represents the precision parameter of the CAR model.

To properly reflect the uncertainty sourced from the estimation process in the main model, we implemented the Bayesian spatio-temporal hierarchical algorithm, where a Markov Chain Monte Carlo (MCMC) sample is drawn from posterior distributions of the joint model and used to construct credible intervals of parameters during the process. We used non-informative prior and the pre-specified values are relative to the size of our analyzing data set.

$$\beta_k \sim N(\mu = 0, \sigma^2 = 100), k = 0,1, \cdots, K,$$

$$\sigma_\epsilon^2 \sim IG(\text{mean} = 1, \text{var} = 100)$$

Then, we present the following proposed prior distribution of the hyperparameters in the spatial structure (CAR priors):

$$\tau^2 \sim \Gamma(\text{mean} = 1, \text{var} = 100)$$

$$\tau_\eta^2 \sim \Gamma(\text{mean} = 1, \text{var} = 100)$$

The following relatively non-informative prior distributions can be used for the remaining unknown parameters of the measurement error process model.

$$\kappa \sim N(u = 0, \sigma^2 = 100)$$

$$\sigma_u^2 \sim IG(\text{mean} = 1, \text{var} = 100)$$

$$\sigma_v^2 \sim IG(\text{mean} = 1, \text{var} = 100)$$

For the structural conciseness and computational efficiency, we carry out the MCMC sampling procedure using a popular software, Stan. The sampler is implemented 30,000 iterations generated with the first 5,000 as burn-in period. We keep the estimate for further analysis based on the remaining 25,000 iterations. It is convenient to check the MCMC algorithm converges very rapidly using trace plots of the MCMC chains and their autocorrelations. Sensitivity of the results to the variances of the prior densities does not play a key factor in the estimation process.

Appendix 2

Summary of posterior results for the temporal random effects of our phenology model for Purple martins.

Year	Mean	Standard Deviation	0.025 percentile	0.975 percentile
2001	-2.472	1.526	-4.919	1.22
2002	-1.1	1.522	-3.567	2.62
2003	0.234	1.61	-2.397	4.2
2004	1.233	1.63	-1.396	5.238
2005	2.166	1.56	-0.288	5.963
2006	0.099	1.577	-2.279	4.052
2007	1.122	1.684	-1.551	5.126
2008	1.396	1.657	-1.134	5.496
2009	2.955	1.631	0.549	7.086
2010	2.981	1.687	0.238	6.96
2011	1.492	1.684	-1.106	5.584
2012	2.244	1.759	-0.671	6.387
2013	0.353	1.728	-2.239	4.662
2014	-0.958	1.753	-3.669	3.388
2015	1.137	1.722	-1.712	5.33
2016	1.988	1.738	-0.762	6.212
2017	4.277	1.626	1.668	8.305
2018	2.109	1.763	-0.619	6.456