Ecography

E5453

Aukema, B. H., Carroll, A. L., Zheng, Y., Zhu, J., Raffa, K. F., Moore, D., Stahl, K. and Taylor, S. W. 2008. Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate. – Ecography 31: 348–358.

Supplementary material

Appendix 1. Details of temperature interpolation model

The interpolation weights are computed using a truncated Gaussian filter:

$$w(r) = \begin{cases} 0; & r > R_p \\ \exp\left[-\left(\frac{r}{R_p} \right)^2 \alpha \right] - e^{-\alpha}; & r \le R_p \end{cases} \quad \text{(S1.1)}$$

where w(r) is the weight associated with the radial distance r from the prediction point p, R_p is the truncation distance from p, and α is a shape parameter. The truncation distance R_p is allowed to vary so that it may be reduced in areas of high station density and increased in areas of low station density. This is done by specifying a desired number of observations N, and using an iterative scheme to vary R_p such that the actual number of observations used in the interpolation converges towards N. Following the truncation distance adjustment, a final set of interpolation weights is calculated using the new value of R_p; these weights are then used in the temperature interpolations.

A weighted least-squares regression is used to determine the local temperature gradient. Each unique pair of stations is assigned a regression weight equal to the product of the interpolation weights of the two stations in the pair. The regression model takes the form:

$$(T_1 - T_2) = \beta_0 + \beta_1 (h_1 - h_2) + \varepsilon \quad \text{(S1.2)}$$

where subscripts 1 and 2 refer to the two stations in the unique pair, β_0 and β_1 are coefficients to be estimated by weighted regression, and ε is the prediction error. Once the coefficients are calculated, T_p is computed as:

$$T_p = \frac{\sum_{i=1}^{n} w_i \left[T_i + b_0 + b_1 (h_p - h_i) \right]}{\sum_{i=1}^{n} w_i} \quad \text{(S1.3)}$$

where b_0 and b_1 are the estimated regression coefficients, h_p is the elevation of the prediction point and the subscript i refers to the predictor stations.

Calculating temperature gradients by regression can lead to prediction of unrealistic temperatures at points with elevations outside the range of the control stations. Our algorithm therefore computes the temperature gradient using a weighted average of the gradient computed by the weighted regression approach (b_{1D}) and a specified gradient (b_{1S}) for points that lie outside the elevation range of the control points. This option computes the gradient as

$$b_1 = w_D b_{1D} + w_S b_{1S} \tag{S1.4}$$

where the weights are computed as

$$w_D = \begin{cases} (h_{max} - h_{min})/(h_p - h_{min}), & h_p > h_{max} \\ (h_{max} - h_{min})/(h_{max} - h_p), & h_p < h_{min} \end{cases} \tag{S1.5}$$

and

$$w_s = 1 - w_D \tag{S1.6}$$

where h_{max} and h_{min} are the maximum and minimum elevations of the predictor stations and h_p is the elevation of the prediction point.

Appendix 2. Details of spatiotemporal autologistic model

Notation

Let i = 1,…,I index the centroids of the grid cells (I = 469; Fig. 1) and let $\mathbf{N}_i$ consist of the indices of the neighbors of site i for a given neighborhood structure. Here neighbours are ordered according to their adjacency to site i and in particular first-order, second-order, … neighbours are those that are the four nearest, four second nearest, … sites. Let t = 1,…,T index the annual time points. The first year of study, 1972, is t = 1, and there are T = 15 time points in the data. Let $Y_{i,t}$ denote the binary response variable at site i and time t with $Y_{i,t}$ = 0 (or 1) for absence (or presence) of detectable infestation in the cell. Then let the vector $\mathbf{Y}_t = (Y_{1,t},...,Y_{I,t})'$ denote the binary response variables on the entire grid of cells at time t. Furthermore, let $X_{0,i,t} \equiv 1$ and $X_{k,i,t}$ denote the kth explanatory variable at site i and time t, for k=1,…,K where K denotes the number of explanatory variables (Table 1).

Model specification

For modeling $\{\mathbf{Y}_1,\ldots,\mathbf{Y}_T\}$, we assume an autoregressive model of order S:

$$p(\mathbf{Y}_t \mid \mathbf{Y}_{t'} : t' = t-1, t-2, ...) = p(\mathbf{Y}_t \mid \mathbf{Y}_{t'} : t' = t-1, ..., t-S) \tag{S2.1}$$

for t = S + 1,…,T. Then for a given time t, we assume a Markov random field with full conditional distributions:

$$p(Y_{i,t} \mid Y_{j,t} : j \neq i, \mathbf{Y}_{t'} : t' = t-1, ..., t-S)$$

$$= p(Y_{i,t} \mid Y_{j,t} : j \in N_i, Y_{i,t'} : t' = t-1, ..., t-S)\,. \tag{S2.2}$$

Since the response variable is binary (presence or absence of mountain pine beetle populations), we use a Bernoulli distribution for $Y_{i,t}$ in (A2.2) with a presence probability of $p_{i,t}$ such that

$$\text{logit}(p_{t,i}) = \sum_{k=0}^{K} \theta_k X_{k,i,t} + \frac{1}{2}\theta_{K+1} \sum_{j \in N_i} Y_{j,t} + \sum_{s=1}^{S} \theta_{K+1+s} Y_{i,t-s}\,. \tag{S2.3}$$

Statistical inference

Let $\boldsymbol{\theta} = (\theta_0,\ldots,\theta_{K+1+S})'$ denote the vector of model parameters. The likelihood function is based on the distribution of $\{\mathbf{Y}_{S+1},...,\mathbf{Y}_T\}$ conditional on the observations $\{\mathbf{Y}_1,...,\mathbf{Y}_S\}$ at the first S time points. Thus by (S2.1)–(S2.3), the likelihood function is

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}); \mathbf{Y}_{S+1},...,\mathbf{Y}_T \mid \mathbf{Y}_1,...,\mathbf{Y}_S)$$

$$= c(\boldsymbol{\theta})^{-1} \exp\left[\sum_{t=S+1}^{T}\sum_{i=1}^{I}\left(\sum_{k=0}^{K}\theta_k X_{k,i,t} Y_{i,t} + \frac{1}{2}\sum_{j \in N_i}\theta_{K+1} Y_{i,t} Y_{j,t} + \sum_{s=1}^{S}\theta_{K+1+s} Y_{i,t} Y_{i,t-s}\right)\right] \tag{S2.4}$$

Parameter estimation is via Monte Carlo maximum likelihood, because of an unknown normalizing constant $c(\boldsymbol{\theta})$. For prediction of $\mathbf{Y}_t$ at future time points $t = T + 1, T + 2, \ldots$, we consider the predictive distribution $p(\mathbf{Y}_t \mid \mathbf{Y}_{t'} : t' = t - 1, \ldots, t - S)$ where the full conditional distribution of $Y_{i,t}$ is specified in the same way as (S2.1)–(S2.3) and is evaluated at the maximum likelihood estimates of the model parameters. For model comparison and selection, we use Akaike's information criterion (AIC) (Akaike 1973) defined as

$$\mathrm{AIC} = -2\,\ell\,(\boldsymbol{\theta}_{\mathrm{MLE}}) + 2p \tag{S2.5}$$

where $\ell\,(\boldsymbol{\theta}_{\mathrm{MLE}})$ is the log likelihood function of (S2.4) evaluated at the maximum likelihood estimates of the model parameters and p is the number of model parameters. For more details, see Zhu et al. (2007).

References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. – In: Petrov, B. N. and Csadki, F. (eds), 2nd International symposium on information theory. Akademiai Kiado, pp. 267–281.

Zhu, J. et al. 2007. Autologistic regression analysis of spatial-temporal binary data via Monte Carlo maximum likelihood. – J. Agricult. Biol. Env. Stat., in press.