ON DEDUCING TIME PARAMETERS FROM SPATIAL PATTERNS: THE (AUTO)LOGISTIC MODEL IN LANDSCAPE ECOLOGICAL PROCESSES.

Cajo J.F. ter Braak[1,2] and Jana Verboom[2]
DLO-Agricultural Mathematics Group and DLO-Institute for Forestry and Nature Research
Box 100, NL-6700 AC WageningenE-mail:C.J.F.ter.Braak@GLW.AGRO.NL

ter Braak, C. J. F. and Verboom, J. (1995). On deducing time parameters from spatial patterns: The (auto)logistic model in landscape ecological processes. In: *Spatial and Temporal Modelling in Agricultural Research*, Proceedings of the Fourth HARMA workshop, R. W. Payne, (ed.) Harpenden, UK, IARC-Rothamsted.

Abstract

In this paper, a popular metapopulation model is critically examined by putting the model in the context of Markov random fields and the statistical analysis of binary lattice systems. The claim that the model can be used to estimate time-process parameters from spatial-pattern data is examined on a real data set where process information was available.

1 **Introduction**

In man-dominated landscapes, natural habitats occur in small, spatially separated fragments. An animal or plant species that is restricted to such fragments may form a metapopulation, that is, a set of local populations which interact via individuals that move among them (Levins, 1970). For conservation programs, we need to understand the dynamics of such metapopulations, in particular the prospect for survival of the species in the landscape (Verboom et al., 1993). For landscape design and management, we need to assess the effects of mitigating measures (e.g. adding landscape elements) or measures that fragment the habitat even further, e.g. construction of road and railways (Verboom et al., 1993). How does the size and configuration of the fragments influence the metapopulation dynamics?

In one class of metapopulation models, fragments (in this context often termed patches) are either vacant or occupied. Dynamical processes are then the process of extinction of the local population at a patch and the process of colonization of an vacant patch from occupied neighbours. Hanski (1994a) proposed a 'practical model of metapopulation dynamics' in which the parameters that govern the colonization and extinction processes are estimated from 'snapshot' data on the occupancy of the species in a set of patches. In his method, process parameters are thus derived from spatial pattern data. Hanski (1994a) acknowledged that some parameters may sometimes be aliased so that extra information is needed for the claim to hold, but this aspect is not emphasized in Hanski (1994b). Under slightly different formulations of the model, the extra information is always needed, as in Hanski's second model with 'rescue effect'.

In this paper, we critically examine Hanski's models by putting the models in the context of Markov random fields and the statistical analysis of binary lattice systems (Besag, 1977). Related models are discussed by Durrett and Levin (1994).

3 **Hanski's practical model of metapopulation dynamics**

Hanski (1994a) modelled the occupancy of each particular patch i (i=1...n) by a Markov chain with two states (occupied, vacant). Let C_i be the probability that patch i, when vacant, is colonized the next year and E_i is the probability that patch i, when occupied, becomes vacant. The matrix of transition probabilities at time t is thus

$$T_i = \begin{matrix} 1-E_i & C_i \\ E_i & 1-C_i \end{matrix}$$

If C_i and, perhaps, E_i are functions of the occupancy of neighbouring patches, T_i varies in time. To derive a stationary probability that the patch is occupied (p_i), we assume a constant neighbourhood so that T_i is a constant in time. Then the stationary equation is such that the probability of being occupied at time t+1 equals that probability at time t, i.e.

$$p_i = C_i\,(1-p_i) + (1-E_i)\,p_i$$

Solving for p_i yields the stationary probability:

$$p_i = C_i / (C_i + E_i) = 1 / (1 + E_i / C_i) \qquad (2.1)$$

Hanski (1994a) noted the following problem with (2.1). If C_i and E_i both approach 1, then p_i approaches 0.5. The reason for this is that the model does not allow for 'simultaneous' extinctions and colonizations. Hanski (1994) therefore introduced a 'rescue effect' by changing the realized extinction probability from E_i to $(1-C_i)\,E_i$. In terms of an explicit Markov chain, suppose that, in the annual cycle of events, a population may get extinct and later in the year, after reproduction in other patches, become colonized. So extinctions happen only in the first half of the year and colonizations only in the second half. Then we have for the first and second halves of the year the matrices T_{1i} and T_{2i} of transition probabilities, respectively, with

$$T_{1i} = \begin{matrix} 1-E_i & 0 \\ E_i & 1 \end{matrix} \qquad T_{2i} = \begin{matrix} 1 & C_i \\ 0 & 1-C_i \end{matrix}$$

and E_i and C_i are probabilities on semi-annual basis. The Markov chain which uses T_{1i} and T_{2i} alternatingly has on an annual basis the transition probability matrix that is the product of T_{2i} and T_{i1}, i.e.

$$T_{2i}T_{1i} = \begin{matrix} (1-E_i) + C_iE_i & C_i \\ E_i(1-C_i) & 1-C_i \end{matrix}$$

The stationary equation is thus

$$p_i = [(1-E_i) + C_iE_i]\, p_i + C_i\,(1- p_i)$$

whence, as an alternative to (2.1), the model with rescue effect is

$$p_i = C_i / (C_i + E_i - C_iE_i) = 1 / (1 + E_i\,(1-C_i)/C_i) \qquad (2.2)$$

In further specification of the models, Hanski (1994a) toke into account the spatial arrangement of patches, their size and occupancy status. The extinction probability E_i was modelled by a function of the area A_i as follows

$$E_i = \varepsilon\, A_i^{-\beta} \qquad \text{if } A_i > A_0 = \varepsilon^{1/\beta} \qquad (2.3a)$$

$$E_i = 1 \qquad \text{if } A_i \le A_0 \qquad (2.3b)$$

with parameters ε and β ($0 \le \varepsilon \le 1$ and $\beta \ge 0$). Thus, the extinction probability decreases with the area. Extinction is certain when the patch is smaller than size A_0 ($\varepsilon^{1/\beta}$). The colonization probability C_i of patch i was assumed be a function of the number of migrants S_i arriving to patch i per year:

$$C_i = 1/(1 + \gamma / S_i^2) \qquad (2.4)$$

with $\gamma > 0$. Function (2.4) is an s-shaped logistic function in S_i to allow for the Allee effect (Hanski, 1991). The number of migrants S_i is unknown but is modelled in turn by a weighted sum over all occupied patches, namely

$$S_i = \Sigma_j\, y_j\, A_j \exp(-\alpha d_{ij}) \qquad (2.5)$$

where y_j is the 0/1 indicator for the occupancy state of patch j, A_j is the area of patch j and d_{ij} is the distance between patch i and patch j. In (2.5), the number of migrants from an occupied patch j is proportional to its area (A_j), and so to its potential population size, and inversely related to its distance to patch i. Equation (2.5) expresses the connectedness of patch i to the other occupied patches.

The models so obtained are fitted by maximum likelihood to 'snapshot' spatial data on the occupancy, i.e. to the binary data $\{y_i,\ i = 1...n\}$, indicating whether patch i is vacant (0) or occupied (1), while treating the data as independent. In Hanski (1994a) the parameter α is determined apriori from dispersion data. For small $\{C_i\}$, C_i and the odds $C_i/(1-C_i)$ are close, so that it is difficult to distinguish between the models (2.1) and (2.2) from practical data. In the model (2.2) with rescue effect, the parameters ε and γ are aliased, except when the cut-off in (2.3b) is in force; without (2.3b) only the product $\varepsilon\gamma$ is estimable. This aliasing may thus also be a practical problem in model (2.1). If this problem occurs, external information is needed. For a subset of patches, one may have occupancy data for two or more consecutive years. From such data one may calculate the number of turnover events T (extinctions and colonizations) and resolve the alias by the extra estimation equation T -E(T) = 0 where E(T) is the expected number of turnovers in the model (Hanski, 1994a). Alternatively, one may have an idea about A_0 in (2.3), the largest area below which the population in a patch that is just colonized will certainly go extinct next year (Hanski, pers. comm.). From A_0 and β, one obtains ε, so that also γ becomes estimable.

If the data are from a metapopulation at dynamic equilibrium between extinctions and colonizations and if all parameters are estimable, time-process parameters are obtained from spatial pattern data.

5 **Comments on Hanski's model**

The model is derived for a space-time process, but estimated from spatial data only. In the estimation, the spatial occupancy data $\{y_i\}$ play a role as predictors via (2.5) and as response variables as in auto-models (Besag, 1974). From the derivation of the model, the occupancy data used in (2.5) should have been from the previous year. The implicit assumption in the method is thus that the variation in the predictor values $\{S_i\}$ is small across subsequent years (Gyllenberg and Silvestrov, 1994). This assumption is trivially satisfied if C_i and $1-E_i$ are close to 0. The stationary probabilities (2.1) and (2.2) are derived under the assumption that E_i and C_i do not vary in time. However, C_i will vary in time. Again the implicit assumption is that the variation in $\{S_i\}$ is small. Gyllenberg and Silvestrov (1994) study the quasi-stationary distribution for the occupancy $\{y_i\}$.

It appears from the explicit formulae given for $\{p_i\}$ in Hanski (1994a,b) that equation (2.3b) is not used while fitting the models. Without the cut-off specified by (2.3b), the model (2.2) with rescue effect has the form of a logistic model, namely

$$\text{logit}(p_i) = \beta_0 + \beta_1 \log(A_i) + \beta_2 \log(S_i) \qquad (3.1)$$

with $\beta_0 = -\log(\varepsilon\gamma)$, $\beta_1 = \beta$ and $\beta_2 = 2$. The method of fitting employed by Hanski (1994a) thus amounts to a logit regression with an offset. In a more extended version of the model β_2 is a free parameter.

In the method of fitting by logit regression, the autocorrelation among the spatial occupancy data is neglected. Consequently, the method of fitting does not maximizes a likelihood, but a pseudolikelihood (Besag, 1995, Preisler, 1993). Maximum pseudolikelihood estimators are in general still consistent but the standard errors must be adapted. The standard errors given in Hanski (1994a,b) thus cannot be trusted without further justification.

Model (3.1) is not a proper auto-logistic model in the sense of Besag (1974). As a consequence of the Hammersley-Clifford theorem (Besag, 1974), the conditional probabilities given by model (3.1) do not define a joint probability distribution for the spatial data $\{y_i\}$; the set of conditional probabilities is internally inconsistent. With pairwise interactions among patches, the only conditional model for binary data that yields a valid joint probability model is of the form

(Besag, 1974; Preisler, 1993)

$$\text{logit } p_i = \gamma_i + \Sigma_{j \neq i} \delta_{ij} y_j \tag{3.2}$$

with $\delta_{ij} = \delta_{ji}$. An example of a proper auto-logistic model is

$$\text{logit } p_i = \gamma_0 + \gamma_1 \log(A_i) + \gamma_2 A_i S_i \tag{3.3}$$

Model (3.3) is of form (3.2) as can be seen by defining

$$\delta_{ij} = \gamma_2 A_i A_j \exp(-\alpha d_{ij}). \tag{3.4}$$

The question is therefore whether logit regression can still yield sensible estimates of the parameters in (3.1).

One may argue that the inconsistency of (3.1) in defining a joint probability distribution is not insurmountable because the model is meant to be valid for the spatial-temporal process, and not for the spatial data per se. Therefore, data were simulated by the spatial-temporal process and analyzed as spatial 'snapshot' data by Hanski's model (50 patches, 200 independent simulations). The configuration and size of the patches was as in Hanski (1994a: Fig. 7). Two sets of parameter values were used; one set as in Hanski (1994a, Fig. 7) and one set with much higher turnover rates. Hanski's estimators of the colonization and extinction turned out to be almost unbiased. For application of the model it is important to note that the variation in the estimates was huge.

7 **Application of Hanski's model to nuthatch data.**

Both models (3.1) and (3.3) were fitted to real data on a nuthatch metapopulation (cf. Verboom et al. 1991). Hanski's model (3.1) fitted better to the data than (3.3) as judged on the basis of the pseudo-deviance.

Using $\alpha = 0.1$ and $\beta_2 = 2$, models (2.1) and (2.2) were fitted to data of each of six consecutive years in turn (1988-1993). We experienced numerical problems in all fits of model (2.1) [with 2.ab, 2.4 and 2.5] by GENSTAT FITNONLINEAR, despite the good starting values obtained by logit regression. For two years no estimates could be obtained, indicated by stars in the table below. The fit of model (2.2)/(3.1) presented no problems. For model (2.2) an expert guess of A_0 of 0.5 ha was used to separate ε from γ. The expected number of turnovers between two consecutive years was estimated from each fit. The expected and observed number of turnovers (E(T) and T) were as follows:

year	88/89	89/90	90/91	91/92	92/93
E(T)					
(2.1)	*	2.5	*	42.0	23.4
(3.1)	12.5	11.3	5.8	20.1	21.6
T	14	10	12	24	15

By fitting (3.1) with β_2 free, a deviance test on $\beta_2 = 2$ was obtained. In none of the six cases, there was statistical evidence against $\beta_2 = 2$. Of course, the validity of the test is hampered by the fact that the fit is by pseudo-likelihood instead by regular maximum likelihood. The utility of Gibbs sampling of the spatial-temporal process with missing data (Augustin et al., 1994) is under investigation.

References

Augustin, N. H., Mugglestone, M. A. and Buckland, S. T. (1994). The role of the Gibbs sampler in modelling the spatial distribution of wildlife from incomplete survey data. *Proceedings of the 9th International Workshop on Statistical Modelling, Exeter*, 8 pp.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B* **36,** 192-236.

Durrett, R. and Levin, S. A. (1994). Stochastic spatial models: a user's guide to ecological applications. *Phil. Trans. R. Soc. Lond. B* **343,** 329-350.

Gyllenberg, M. and Silvestrov, D. S. (1994). Quasi-stationary distributions of a stochastic metapopulation model. *J. Math. Biol.* **33,** 33-70.

Hanski, I. (1991). Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society* **42,** 17-38.

Hanski, I. (1994a). A practical model of metapopulation dynamics. *Journal of Animal Ecology* **63,** 151-162.

Hanski, I. (1994b). Patch-occupancy dynamics in fragmented landscapes. *TREE* **9,** 131-135.

Levins, R. (1970). Extinction. Pages 75-108 in *Some mathematical questions in biology, vol II,* Rhode Island: American Mathematical Society.

Preisler, H. K. (1993). Modelling spatial patterns of trees attacked by bark-beetles. *Applied Statistics* **42,** 501-514.

Verboom, J., Metz, J. A. J. and Meelis, E. (1993). Metapopulation models for impact assessment of fragmentation. In *Landscape Ecology of a Stressed Environment,* C. C. Vos, and P. Opdam (eds), 172-191. London: Chapman & Hall.

Verboom, J., Schotman, A., Opdam, P. and Metz, A. J. (1991). European nuthatch metapopulations in a fragmented agricultural landscape. *OIKOS* **61,** 149-156.