Biostatistics (2000), **1**, 4, *pp.* 453–463
Printed in Great Britain

Spatial modelling of individual-level parasite counts using the negative binomial distribution

NEAL ALEXANDER*
Infectious Disease Epidemiology Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK
neal.alexander@lshtm.ac.uk

RANA MOYEED, JULIAN STANDER
School of Mathematics and Statistics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK

SUMMARY

We present a spatial model for the mean and correlation of highly dispersed count data, and apply it to individual-level counts of the nematode *Wuchereria bancrofti*, a parasite of humans which causes the disease lymphatic filariasis. Our model uses the negative binomial distribution, whose shape parameter is a convenient index of over-dispersion. Spatial association is quantified in terms of a characteristic length, which has an intuitive interpretation as the distance over which correlation decreases by half. Demographic surveillance and mapping enable us to include individual-level covariates such as age and sex. We discuss the distinctive features of our model and interpret the results in terms of the epidemiology of lymphatic filariasis and possible implications for control programmes.

Keywords: Epidemiology; Extra-Poisson variation; Negative binomial distribution; Parasitic diseases; Spatial modelling.

1. INTRODUCTION

Vector-borne parasitic diseases often show high variability on small spatial scales (Greenwood, 1989). Our aim in this article is to develop previous modelling in three ways. Firstly, individual, rather than area-level data are included, permitting individual risk factors to be assessed. Secondly, we use the negative binomial distribution because of its good fit to a wide range of parasitological data, and its intuitive interpretation in terms of sampling (Grenfell *et al.*, 1990). Thirdly, our spatial model allows us to estimate a characteristic length of the infection pattern, over which the correlation reduces by half.

The work is motivated by studies on the epidemiology and control of lymphatic filariasis, a mosquito-borne parasitic disease. The pathogen is the nematode *Wuchereria bancrofti*; symptoms can include elephantiasis and adenolymphangitis (inflammation of lymph nodes and vessels). The current work includes an application to data from a drug trial performed in a rural area of the East Sepik Province of Papua New Guinea (Bockarie *et al.*, 1998).

Parasite counts are usually over-dispersed relative to the Poisson distribution, and are often described well by the negative binomial distribution (Anderson, 1993). This is the case for many species of parasites (Shaw and Dobson, 1995), and also for other organisms, particularly invertebrates (Elliott, 1977; Nedelman, 1983). The high variability of parasite counts could make a map of raw means difficult to

*To whom correspondence should be addressed

interpret (Mollié, 1996). With our model we aim to obtain, *mutatis mutandis*, 'a map which would display regional variations in cancer incidence yet avoid the presence of unstable rates for the smaller counties' (Breslow and Clayton, 1993). In our case, a particular objective is to highlight regions with increased risk of infection, which, in a pre-treatment situation, could help identify potential environmental risk factors. If repeated after the initial treatment, it could identify areas not responding well to the programme. This is not dissimilar to the objective of the geographical monitoring of the West African Onchocerciasis Control Programme (Molyneux and Davies, 1997), although on a finer scale. In addition, estimating the scale of spatial correlation could provide some guidance on the relative risks of recrudescence in different areas.

2. THE MODEL

The parasite count Y_{ij} of person j in hamlet $i = 1, \ldots, N$ is modelled as a negative binomial variate with mean μ_{ij}. This means that

$$\text{pr}(Y_{ij} = y_{ij}) = \frac{\Gamma(k + y_{ij})}{\Gamma(k) y_{ij}!} \frac{k^k \mu_{ij}^{y_{ij}}}{(k + \mu_{ij})^{k+y_{ij}}}$$

where $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ is the gamma function. Since $\text{var}[Y_{ij}] = \mu_{ij} + \mu_{ij}^2/k$, it is clear that the second parameter of the negative binomial distribution $k > 0$ incorporates extra-Poisson variation. Larger values of k correspond to less variability, with the limiting case $k = \infty$ corresponding to the Poisson distribution. The relationship between the variance and mean can be interpreted in terms of Taylor's power law (Taylor, 1961). For parasites, including nematodes, this is generally inconsistent with a linear variance function (Shaw, 1994). As well as this empirical basis, the negative binomial can be derived as a gamma mixture of Poissons (Stuart and Ord, 1994). This has a simple, and biologically intuitive, interpretation when counting parasites in finite quantities of blood, if each person's parasites are subject to Poisson sampling variability, and the variability between people has a gamma distribution (Grenfell *et al.*, 1990). On the other hand, the negative binomial can be generated by other Poisson mixtures, so is not dependent on this particular pair of assumptions (Anderson and May, 1991, p 456). Other spatial models have used the gamma distribution (reviewed by Mollié, 1996), but without a negative binomial interpretation.

We model the logarithm of the mean μ_{ij} as an additive function of possible risk factors such as sex and age, and of a spatial hamlet effect parameter u_i. The form for the joint distribution of the vector of spatial hamlet effects $u = (u_1, \ldots, u_N)$ is similar to that adopted by Diggle *et al.* (1998) in the context of geostatistics. For this, u is assumed to have a multivariate normal distribution with mean 0 and variance matrix Σ, such that the covariance between two hamlets is a negative exponential function of the distance between them. The model is shown graphically in Figure 1 for the simple case in which age is the only non-spatial risk factor, its effect being measured by the coefficient β.

We assume k is constant, although it is possible to model it as a function of covariates, linearly or otherwise (Thurston *et al.*, 2000). In principle, it would be possible to extend the above framework to model k spatially. However, this would considerably complicate the existing model, which, as we argue below, seems to perform adequately and yields satisfactory estimates of the parameters of most direct interest.

The full model can be written more explicitly as follows:

$$\begin{aligned} Y_{ij} &\sim \text{negative binomial}\,(\mu_{ij}, k) \\ \log(\mu_{ij}) &= \beta_0 + \beta_1 \text{I}_{\{\text{sex}_{ij}=\text{male}\}} + \beta_2 \text{age}_{ij} + u_i \\ u = (u_1, \ldots, u_N) &\sim N(0, \Sigma) \\ \Sigma_{il} &= \frac{1}{\phi} \exp\left(-\frac{d_{il}}{\alpha}\right), \end{aligned}$$

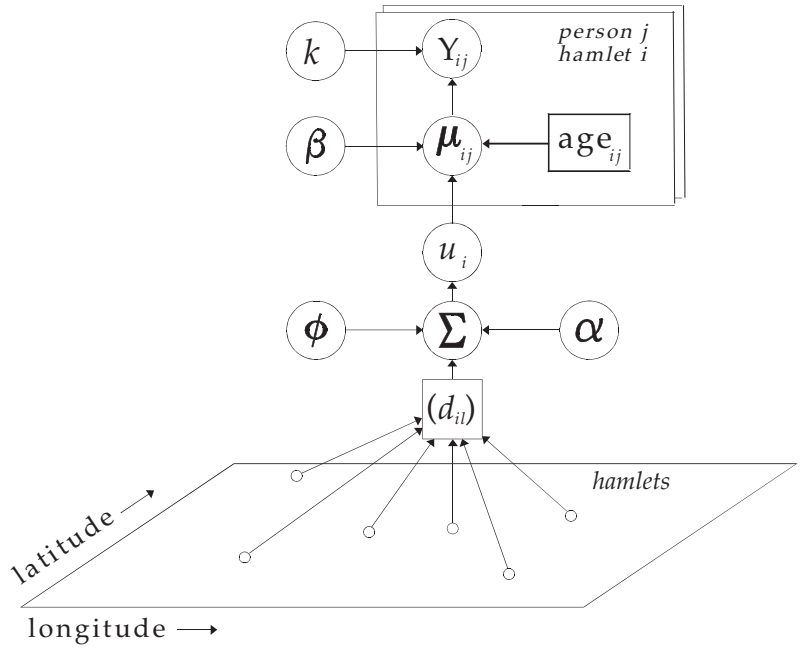

Fig. 1. Graphical representation of the model.

in which the indicator function $\mathrm{I}_{\{\mathrm{sex}_{ij}=\mathrm{male}\}}$ takes the value 1 if person j in hamlet i is male and 0 otherwise; age_{ij} is the age in years of this person; and d_{il} is the distance between hamlets i and l. The parameters u can be thought of as a vector of spatially structured hamlet effects. Both $\phi > 0$ and $\alpha > 0$ measure aspects of the spatial smoothness of the parasite counts. The parameter ϕ measures the scale of the spatial variation, with $\mathrm{var}[u_i] = 1/\phi$. The parameter α measures the rate at which the spatial correlation decays over distance, with $\alpha \log 2$ being a characteristic length, which we call the 'half-distance', over which the correlation reduces by half, and 3α being the distance at which the correlation reduces by 95%.

To complete the description of our model we need to specify prior distributions for k, β_0, β_1, β_2, ϕ and α. We adopt the following:

$$
\begin{aligned}
k &\sim \text{gamma}\,(r_k, d_k) \\
\beta_m &\sim N\left(0, \frac{1}{\tau_{\beta_m}}\right), m = 0, 1, 2 \\
f(\phi) &\propto \frac{1}{\phi} \\
f(\alpha) &\propto \frac{1}{\alpha^2},
\end{aligned}
$$

in which in general $X \sim \text{gamma}(r, d)$ means that the probability density $f(x)$ is proportional to $x^{r-1}\exp(-dx)$. We set $r_k = 1.5$, $d_k = 0.01$ and the precisions $\tau_{\beta_m} = 10^{-5}$ for $m = 0, 1, 2$ so that, *a priori*, k and β_m have high variances. In Section 4 we discuss in detail the reasons for these choices, and the sensitivity of our results to them. With the above model formulation, and its graphical representation in Figure 1, Bayes' Theorem can be used to write down the posterior distribution of

the parameters $\theta = (k, \beta_0, \beta_1, \beta_2, u_1, \ldots, u_N, \phi, \alpha)$ given the data. We simulate realizations from this posterior distribution by means of a single-component Metropolis–Hastings algorithm (see Metropolis *et al.*, 1953; Hastings, 1970; Gilks, 1996; Gilks *et al.*, 1996). When α is updated, the $N \times N$ symmetric positive-definite matrix Σ must be inverted, which we do by Choleski decomposition (Stoer and Bulirsch, 1993). This process of updating all the parameters in turn is repeated many times to provide a sequence of parameter vectors $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(G)}$. In order to remove the effect of the initial vector $\theta^{(0)}$, the first B members of this sequence are thrown away; we say that we have used a burn-in of length B. Inference is then based on the remaining members $\theta^{(B+1)}, \theta^{(B+2)}, \ldots, \theta^{(G)}$, which are considered to have converged to a sequence from the posterior distribution. For our model we have always found convergence to be quite rapid and so we take $G = 10\,000$ with $B = 5000$. Our software was originally developed in `FORTRAN`, with independently written `C` code being used for confirmation purposes. We tried to fit negative binomial distributions in the `WinBUGS` v1.3 package (MRC Biostatistics Unit, Cambridge, United Kingdom) but were unsuccessful.

3. APPLICATION TO PARASITE DATA

The study area comprises parts of the Urat and Urim census districts of the East Sepik Province of Papua New Guinea. Most people are engaged in shifting agriculture, with some cultivation of cash crops, and the terrain consists of steep forested ridges. Demographic surveillance tracks each person's residence, which for the current work we resolve to the level of hamlet, each of which is a cluster of a few houses defined by a name in the local language. Although some hamlets stand by themselves, most lie close to each other within linear ridge-top settlements, with no visible features marking their boundaries. Houses are almost invariably made of locally harvested materials.

As part of the drug trial, an annual survey was done, which included measurement of parasite densities. In 1994, before the first round of treatment, 1 ml blood samples were obtained from 2219 people aged 5 years and above, in 149 hamlets. Each sample was passed through a Nuclepore filter, and the number of parasitic worms—in their infective stage known as microfilariae—were counted under a microscope. Subsequent annual surveys cannot be used directly to estimate variability over time because of confounding with the anti-parasitic effect of the drugs (ivermectin and diethylcarbamazine). The study area was mapped with a hand-held Trimble Ensign Global Positioning System (GPS) machine. Latitude and longitude positions were recorded in decimal degrees and converted to kilometres by equating $1/60$ of a degree to 1.852 km. This is a reasonable approximation close to the equator, as is the case here. Hamlets whose distance to their nearest neighbour was measured as less than 10 m were combined with their nearest neighbour. This procedure was undertaken due to the limited accuracy of the GPS. There were two such hamlets, so reducing the total number to 147.

4. RESULTS

As an initial analysis, we fitted single sample negative binomial and Poisson distributions to the data, with no covariates. We also fitted a Poisson distribution with an extra point mass at zero. Figure 2 shows a histogram of the data together with the fitted distributions. As expected, the Poisson is clearly inadequate, as is the modification with point mass at zero. The negative binomial is much better, although it does not fully capture the peak towards the upper end of the axis. A chi-squared goodness of fit is highly significant, indicating that this simple fit is not a full description of the data. Nevertheless, given the extreme skewness—even on a logarithmic scale—the negative binomial distribution seems a reasonable choice for modelling purposes. We also tried the truncated negative binomial (Grenfell *et al.*, 1990), but it did not fit as well as the full negative binomial.

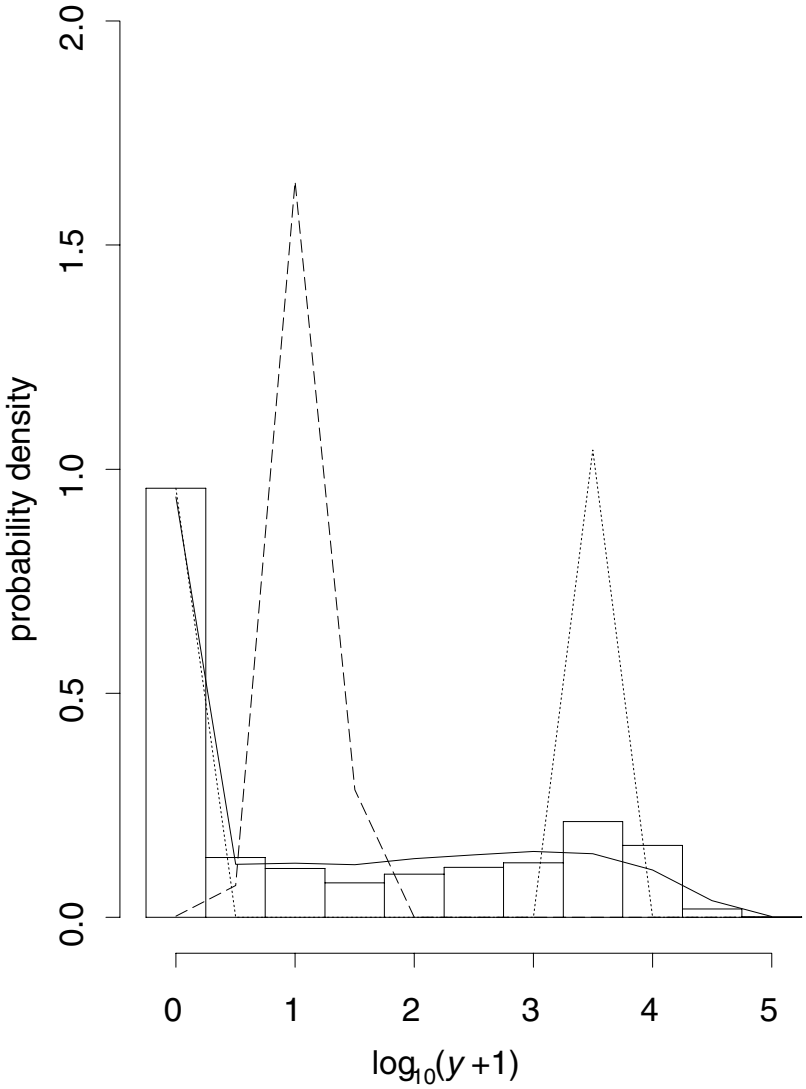

Fig. 2. Probability density estimates of the microfilarial counts y by histogram (bars), negative binomial (solid line), Poisson (dashed line), and Poisson with an extra point mass at zero (dotted line). The fits were done by simple maximum likelihood with no covariates. Since the data include zeros, they cannot be shown on a simple logarithmic scale, so 1 was added before taking logarithms.

A statistic that is widely used for estimating the spatial covariance structure is the variogram (see Section 2.4 of Diggle *et al.* (1998), for example). Figure 3(a) presents the empirical variogram of the average of the parasite counts for each hamlet: $\bar{y}_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}/n_i$, where n_i is the number of people in hamlet i. The curve was produced using a loess smoother to aid interpretation. The fact that this variogram does not increase after about 6 km indicates both that the underlying spatial process is stationary and that beyond this distance the correlation is negligible. We also looked at directional variograms of $\bar{y}_{i\cdot}$. These

did not indicate any anisotropy. We also fitted a non-spatial Poisson model:

$$Y_{ij} \sim \text{Po}(\lambda_{ij})$$
$$\log(\lambda_{ij}) = \beta_0 + \beta_1 \text{I}_{\{\text{sex}_{ij}=\text{male}\}} + \beta_2 \text{age}_{ij}$$

where $\text{Po}(\lambda_{ij})$ stands for the Poisson distribution with mean λ_{ij}, and a similar non-spatial negative binomial model with the same covariates. Results from these fits are given in Table 1. The variogram of the average of the Pearson residuals for each hamlet from the non-spatial negative binomial model was very similar in form to the empirical variogram presented in Figure 3(a). The variogram of the average of the Pearson residuals for each hamlet from the spatial negative binomial model is given in Figure 3(b). The constant nature of this variogram suggests that no spatial correlation remains in these residuals. All this evidence strongly suggests that the spatial negative binomial model does a good job at explaining the spatial correlation.

Table 1 presents approximate posterior means, medians and 95% credible intervals for the parameters k, β_0, β_1, β_2, ϕ and α, based on the parameter vectors $\theta^{(B+1)}, \theta^{(B+2)}, \ldots, \theta^{(G)}$. We see that the non-spatial negative binomial model has a low value of k, indicating again that the Poisson is not appropriate. This is reflected in the unrealistically tight confidence intervals for the Poisson regression parameters, corresponding to the tight peaks of probability density in Figure 2. However, the point estimates of β_0 and β_2 are similar in the two non-spatial models. The point estimates of the remaining common parameter, β_1, differ more substantially, although the negative binomial results suggest that it is not statistically significant. In the spatial model, the posterior mean of k—while still small—is slightly more than that obtained from simple maximum-likelihood models, which implies that, as expected, the spatial component of the model has accounted for some of the variability in the data. The credible interval for β_1 again contains 0, suggesting that the variable sex can be removed from the model. The results obtained for the other parameters when the variable sex is removed are almost identical to those that we present. On the other hand, the credible interval for β_2 lies completely above zero, leading us to conclude that parasite count increases with age. The posterior mean of 2.414 km for α corresponds to a 'half-distance' for correlation of $2.414 \times \log 2 \approx 1.67$ km, with $3 \times 2.414 \approx 7.2$ km being the distance at which correlation reduces to 0.05. In other words, there is strong local correlation that dies off quickly. This is in agreement with the empirical variogram shown in Figure 3(a). The posterior mean half-distance is very close to the value of 1.8 km which was found to be the maximum distance flown by the main mosquito vector, *Anopheles punctulatus*, in studies done in the same area (Charlwood and Bryan, 1987). This value is therefore highly feasible biologically.

Figure 4(a) shows a statistic similar to the standardized mortality ratio (SMR) used for fatal diseases. Here we use the ratio of the observed parasite count to the expected value from the non-spatial negative binomial model with age and sex. This ratio is obtained for each person and averaged within each hamlet. Figure 4(b) shows the estimated posterior mean of $\exp(u_i)$ for each hamlet i. For both panels of Figure 4 we use white, light grey, dark grey and black to indicate the first, second, third and fourth quartile, respectively. The associated break points occur at observed/expected ratios of 0.18, 0.65 and 1.50, and at 0.56, 1.02 and 1.71 for the posterior means. We see that Figure 4(b) is a smoothed version of Figure 4(a). The general message of Figure 4(b) is an increase in parasite counts as we move towards the west. However, some 'hot spots' of infection also remain in the east. While several of the fourth quartile hamlets in Figure 4(a) lie in the east of the study area, only one or two hot spots remain in Figure 4(b). The smoothing process has highlighted those areas where high infection levels are unlikely to be due simply to sampling variability. We return to this point in Section 5.

In an extensive sensitivity study, we found that the results presented above for k were very robust indeed to the choice of the parameters r_k and d_k. This again confirms the appropriateness of the negative binomial. We take a Jeffreys prior for ϕ: $f(\phi) \propto 1/\phi$. This transforms naturally to a Jeffreys prior for

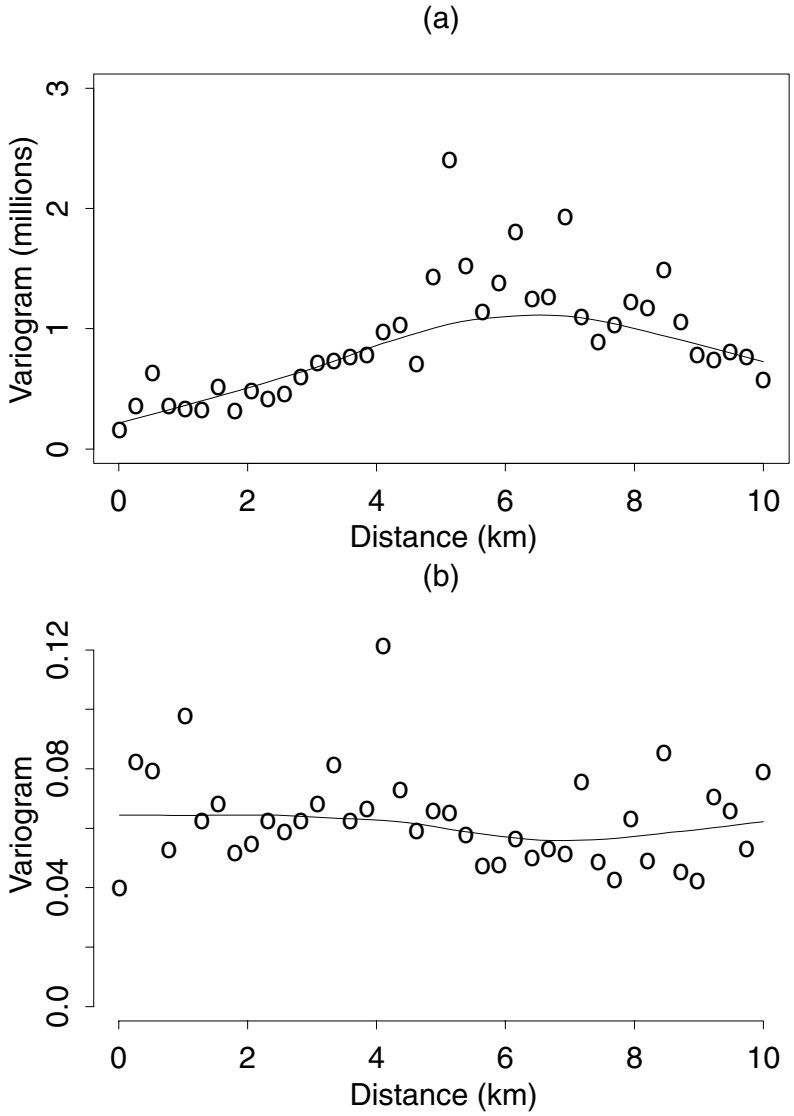

Fig. 3. (a) Variogram of the average $\bar{y}_{i\cdot}$ of the parasite counts for each hamlet. (b) Variogram of the average of the Pearson residuals for each hamlet from the full spatial model. The curves were produced using a loess smoother.

$\sigma^2 = 1/\phi = \text{var}[u_i]$: $f(\sigma^2) \propto 1/\sigma^2$. For further discussion about Jeffreys priors see, for example, O'Hagan (1994). We got very similar results with $\phi \sim \text{gamma}(r_\phi, d_\phi)$ in which $r_\phi = d_\phi = 0.01$; the results obtained with this prior were quite robust to the choice of the parameters r_ϕ and d_ϕ provided that $\text{var}[\phi] = r_\phi/d_\phi^2$ was large. The $1/\alpha^2$ prior was adopted for α for three reasons. Firstly, the algorithm failed to converge with an improper uniform prior for α. With a proper uniform prior on $(0, \alpha_{\max})$ for α, the 95% credible interval for α was almost equal to $(0, \alpha_{\max})$ for a wide range of $\alpha_{\max}$. Secondly, the adopted prior is equivalent to an improper uniform prior on $\alpha^* = 1/\alpha$, which could be taken as another parametrization of correlation; see the discussion of Diggle *et al.* (1998). Thirdly, our prior has an intuitive appeal: *a priori*

Table 1. *Estimates and* 95% *intervals for the parameters* k, β_0, β_1 *(for male sex),* β_2 *(for age in years),* ϕ *and* α

Parameter	Nonspatial models						Spatial model			
	Poisson			Negative binomial						
	Estimate	Confidence interval		Estimate	Confidence interval		Mean	Credible interval		Median
k	∞	—	—	0.084	0.078	0.089	0.089	0.084	0.095	0.089
β_0	5.972	5.969	5.976	5.782	5.467	6.098	5.228	4.798	5.642	5.228
β_1	0.203	0.200	0.206	0.153	−0.141	0.447	0.208	−0.090	0.499	0.207
β_2	0.0212	0.0211	0.0213	0.0283	0.0196	0.0371	0.038	0.027	0.049	0.038
ϕ	—	—	—	—	—	—	1.789	0.732	3.269	1.704
α	—	—	—	—	—	—	2.414	0.856	6.129	1.998

one would assign a higher credibility to the range of correlation being short rather than long. We also experimented with a gamma prior, gamma(r_α, d_α) for α, but found that the results were quite sensitive to the choice of r_α and d_α, especially when $\text{var}[\alpha] = r_\alpha/d_\alpha^2$ was small. In all cases we found that the map of hamlet effects presented in Figure 4(b) was quite robust to prior specifications.

We attempted to generalize the exponential covariance model to $\Sigma_{il} = \exp\{-(d_{il}/\alpha)^\delta\}/\phi$, where $\delta \in (0, 2)$. This covariance model is discussed in Diggle *et al.* (1998). This model led to considerable numerical difficulties. Other covariance models are of course possible. We found that the spherical model discussed in Section 2.3.1 of Cressie (1993) gave very similar results. Our experience is that the Matérn (1986) class of models are even more susceptible to numerical difficulties in the Markov chain Monte Carlo framework than the general exponential covariance model; see Moyeed and Papritz (2000) for a detailed discussion.

5. CONCLUSIONS

Fine-scale residence data—produced by a combination of demographic surveillance and Geographical Information System techniques—enabled us to include distance in our model, not simply neighbourhood structures as in, for example, Bernardinelli and Montomoli (1992), who employ conditional autoregressive schemes. In turn, this enabled us to estimate a characteristic length for the spatial correlation of infection. This is in accordance with Rothman's suggestion that spatial modelling tends to be more useful when concentrating on the estimation of epidemiological effects, rather than tests of hypotheses, such as the presence or otherwise of clustering (Rothman, 1990). For lymphatic filariasis, the magnitude of the characteristic length fits well with our knowledge of the transmission mechanism of the disease. In addition, individual-level data enable us to adjust for covariates such as age. Using the negative binomial distribution to model extra-Poisson variation gives a natural interpretation when dealing with parasite counts, with the k parameter being a convenient index of over-dispersion.

The results from our model confirm in more detail the impression that infection, and other disease parameters such as prevalence of oedema, are more severe towards the west of the study area. On a broad scale these are correlated with the intensity of transmission of the disease by mosquitoes (Kazura *et al.*, 1997). It would have been interesting to include in the spatial model more detailed entomological information such as locations of breeding sites, although systematic measurement of these would have been difficult (Service, 1976), especially over extensive and inaccessible terrain—such as that surrounding the current study area—given that the main vector, *Anopheles punctulatus*, can breed in small puddles and ground holes (Charlwood *et al.*, 1986). Our smoothing process reduces the instability of estimates for small hamlets, and helps identify localities with high infection intensity. Some such localities are

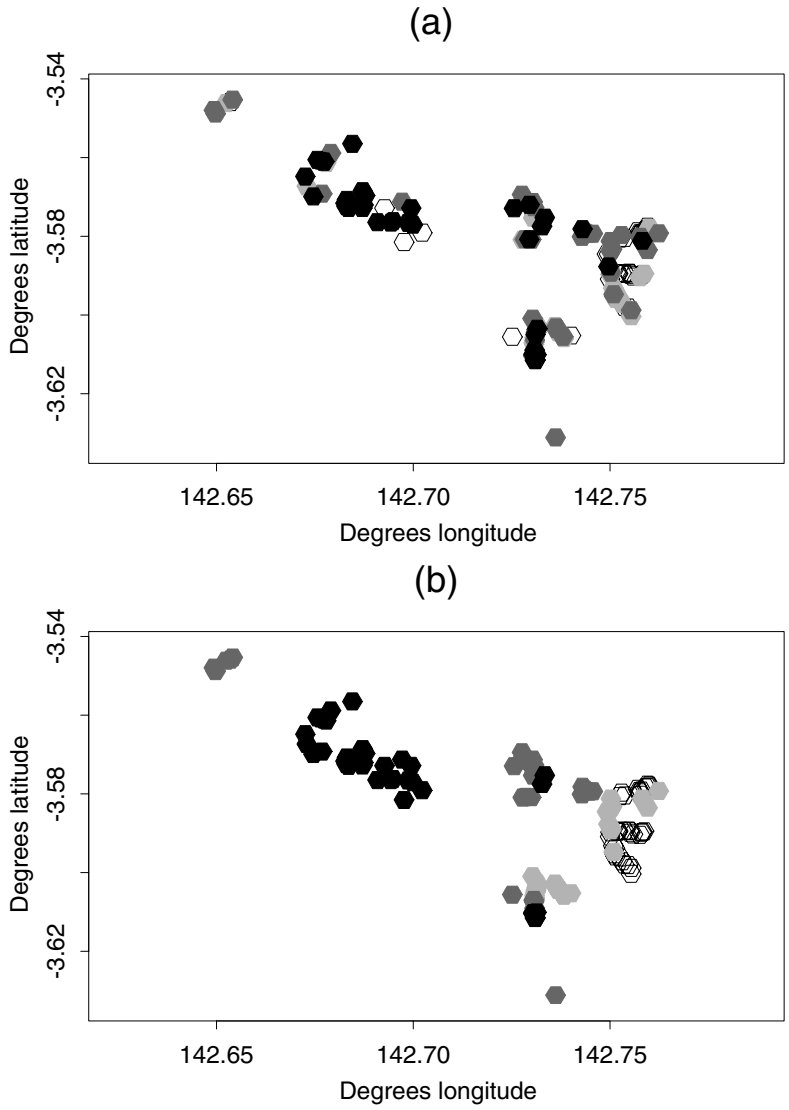

Fig. 4. Maps of parasite density. Each hexagon represents one hamlet. The shading shows the quartile, with white indicating the first and black indicating the fourth. (a) Average of the observed/expected parasite count from a non-spatial regression model (see text for more details), and (b) the posterior mean of $\exp(u_i)$ for each hamlet $i = 1, \ldots, N$ from the model. The distance from east to west of the study area is 12.2 km; from north to south it is 3.6 km.

seen to persist even in the eastern part of the study area which, in general, has lower intensities. This methodology therefore has potential applications in terms of targeting disease-control efforts. In addition, we obtain an estimate of the scale over which spatial correlation operates. Such knowledge may also help control programmes by giving information on the likely risk of recrudescence from any missed pockets of infection. In future work we intend to compare these parameters for filariasis with those of malaria, using data from the same area.

ACKNOWLEDGEMENTS

The authors are grateful to the people of Urat and Urim census districts for giving blood to enable the parasite counts to be taken. The drug trial was supported financially by the World Health Organization (grant number TDR 910466) and the government of Papua New Guinea. The current work originated in the PhD thesis of Neal Alexander, which would not have been possible without his supervisor Bryan Grenfell and the support of Michael Alpers, the Director of the Papua New Guinea Institute of Medical Research, and Jim Kazura, the co-principal investigator of the above trial. Julian Stander was partially supported by the EU TMR network ERB-FMRX-CT96-0095 on 'Statistical and computational methods for the analysis of spatial data'. The authors are also grateful to people who commented on earlier versions of the models, including Jim Burridge, Luisa Franconi, Sylvia Richardson, Giovanni Sebastiani, Ian White, and two anonymous referees.

REFERENCES

ANDERSON, R. M. (1993). Epidemiology. In Cox, F. E. G. (ed.), *Modern Parasitology*, 2nd edn. Oxford: Blackwell, pp. 75–116.

ANDERSON, R. M. AND MAY, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press.

BERNARDINELLI, L. AND MONTOMOLI, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographic variation in disease risk. *Statistics in Medicine* **11**, 983–1007.

BOCKARIE, M. J., ALEXANDER, N. D. E., HYUN, P., DIMBER, Z., BOCKARIE, F., IBAM, E., ALPERS, M. P. AND KAZURA, J. W. (1998). Randomised community-based trial of annual single-dose diethylcarbamazine with or without ivermectin against *Wuchereria bancrofti* infection in human beings and mosquitoes. *Lancet* **351**, 162–168.

BRESLOW, N. E. AND CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

CHARLWOOD, J. D. AND BRYAN, J. H. (1987). A mark-recapture experiment with the filariasis vector *Anopheles punctulatus* in Papua New Guinea. *Annals of Tropical Medicine and Parasitology* **81**, 429–436.

CHARLWOOD, J. D., GRAVES, P. M. AND ALPERS, M. P. (1986). The ecology of the *Anopheles punctulatus* group of mosquitoes from Papua New Guinea: a review of recent work. *Papua New Guinea Medical Journal* **29**, 19–26.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, Revised edn. New York: John Wiley & Sons.

DIGGLE, P. J., MOYEED, R. A. AND TAWN, J. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics* **47**, 299–350.

ELLIOTT, J. M. (1977). *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*, 2nd edn. Ambleside: Freshwater Biological Association.

GILKS, W. R. (1996). Full conditional distributions. In Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 75–88.

GILKS, W. R., RICHARDSON, S. AND SPIEGELHALTER, D. J. (1996). Introducing Markov chain Monte Carlo. In Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 1–19.

GREENWOOD, B. M. (1989). The microepidemiology of malaria and its importance to malaria control. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **83**, 25–29. (supplement)

GRENFELL, B. T., DAS, P. K., RAJAGOPALAN, P. K. AND BUNDY, D. A. P. (1990). Frequency distribution

of lymphatic filariasis microfilariae in human populations: population processes and statistical estimation. *Parasitology* **101**, 417–427.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

KAZURA, J. W., BOCKARIE, M., ALEXANDER, N., PERRY, R., BOCKARIE, F., DAGORO, H., DIMBER, Z., HYUN, P. AND ALPERS, M. P. (1997). Transmission intensity and its relationship to infection and disease due to *Wuchereria bancrofti* in Papua New Guinea. *Journal of Infectious Diseases* **176**, 242–246.

MATÉRN, B. (1986). *Spatial Variation*, *Lecture Notes in Statistics 36*, 2nd edn. Berlin: Springer.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. AND TELLER, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

MOLLIÉ, A. (1996). Bayesian mapping of disease. In Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 359–379.

MOLYNEUX, D. H. AND DAVIES, J. B. (1997). Onchocerciasis control: moving towards the millennium. *Parasitology Today* **13**, 418–425.

MOYEED, R. AND PAPRITZ, A. (2000). An empirical comparison of kriging methods for non-linear spatial point prediction. (Submitted). Available on http://www.tech.plym.ac.uk/maths/staff/rmoyeed/research.html.

NEDELMAN, J. (1983). A negative binomial model for sampling mosquitoes in a malaria survey. *Biometrics* **39**, 1009–1020.

O'HAGAN, A. (1994). *Kendall's Advanced Theory of Statistics. Volume 2b: Bayesian Inference*. London: Edward Arnold.

ROTHMAN, K. J. (1990). A sobering start for the cluster busters' conference. *American Journal of Epidemiology* **132**, S6–S13. (Supplement 1)

SERVICE, M. W. (1976). *Mosquito Ecology: Field Sampling Methods*. London: Applied Science Publishers.

SHAW, D. J. (1994). Distribution of macroparasites in naturally-fluctuating host populations, Ph.D. Dissertation, Department of Zoology, University of Cambridge.

SHAW, D. J. AND DOBSON, A. P. (1995). Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S133. (supplement)

STOER, J. AND BULIRSCH, R. (1993). *Introduction to Numerical Analysis*, 2nd edn. New York: Springer.

STUART, A. AND ORD, J. K. (1994). *Kendall's Advanced Theory of Statistics. Volume 1: Distribution Theory*. London: Edward Arnold.

TAYLOR, L. R. (1961). Aggregation, variance and the mean. *Nature* **189**, 732–735.

THURSTON, S. W., WAND, M. P. AND WIENCKE, J. K. (2000). Negative binomial additive models. *Biometrics* **56**, 139–144.

[*Received January 14, 2000*; *first revision May 16, 2000*; *second revision June 21, 2000*; *accepted for publication June 27, 2000*]