Annals of Tropical Medicine & Parasitology, Vol. 101, No. 6, 499–509 (2007)

Spatial modelling and the prediction of *Loa loa* risk: decision making under uncertainty

P. J. DIGGLE[*], M. C. THOMSON[†], O. F. CHRISTENSEN[‡], B. ROWLINGSON[*], V. OBSOMER[†,§], J. GARDON[¶], S. WANJI[**], I. TAKOUGANG[††], P. ENYONG[‡‡], J. KAMGNO[¶], J. H. REMME[§§], M. BOUSSINESQ[¶] and D. H. MOLYNEUX[¶¶]

[*]*Department of Medicine, Lancaster University, Lancaster LA1 4YF, U.K.*
[†]*International Research Institute for Climate and Society, The Earth Institute of Columbia University, Lamont Campus, P.O.B. 1000, Palisades, NY 10964, U.S.A.*
[‡]*Bioinformatics Research Centre, Aarhus University, Hoegh-Guldbergs Gade 10, Building 090, 8000 Aarhus C, Denmark*
[§]*Entomology Unit, Department of Parasitology, Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerpen, Belgium*
[¶]*Laboratoire Mixte Institut de Recherche pour le Développement (IRD)–Centre Pasteur du Cameroun d'Epidémiologie et de Santé Publique, Centre Pasteur du Cameroun, P.O. Box. 1274, Yaoundé, Cameroon, and Institut de Recherche pour le Développement, UR 024, 34000, Montpellier, France*
[**]*Department of Life Sciences, Faculty of Science, University of Buea, P.O. Box 63, Buea, Cameroon*
[††]*Department of Public Health, Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, P.O. Box 1364, Yaoundé, Cameroon*
[‡‡]*Tropical Medicine Research Station, P.O. Box 55, Kumba, Cameroon*
[§§]*UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), World Health Organization, Avenue Appia 20, 1211 Geneva 27, Switzerland*
[¶¶]*Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, U.K.*

Received 4 April 2006, Revised 29 January 2007, Accepted 15 February 2007

Health decision-makers working in Africa often need to act for millions of people over large geographical areas on little and uncertain information. Spatial statistical modelling and Bayesian inference have now been used to quantify the uncertainty in the predictions of a regional, environmental risk map for *Loa loa* (a map that is currently being used as an essential decision tool by the African Programme for Onchocerciasis Control). The methodology allows the expression of the probability that, given the data, a particular location does or does not exceed a pre-defined high-risk threshold for which a change in strategy for the delivery of the antihelmintic ivermectin is required.

Loa loa has recently emerged as a filarial worm of significant public-health importance, as a consequence of its impact on the African Programme for Onchocerciasis Control (APOC). Severe, sometimes fatal, encephalopathic reactions to ivermectin (the drug of choice for onchocerciasis control) have occurred in some individuals with high *Loa loa* microfilaraemias. As high densities of *L. loa* microfilariae are associated with high prevalences of *L. loa* infection (Boussinesq *et al.*, 2001), the people living

Reprint requests to: P. J. Diggle.
E-mail: p.diggle@lancaster.ac.uk; fax: +44 (0)1524 592681.

DOI: 10.1179/136485907X229121

in an area of high prevalence are considered to be at relatively high risk of experiencing encephalopathic reactions to ivermectin. Before the mass distribution of ivermectin in areas where the prevalence of *L. loa* infection exceeds 20%, it is now the policy of the APOC to put in place precautionary measures. In the present study, survey data on *L. loa* prevalence in scattered village communities, together with relevant environmental data, have been used to construct a map showing, for any location in the region of interest, the probability that the prevalence of *L. loa* infection exceeds the 20% 'policy intervention threshold'.

Risk models derived from environmental data, for use by decision makers involved in the implementation of programmes for the control of various vector-borne diseases, are becoming increasingly available. Such models have been widely shown to be effective in delineating geographical areas of risk, and have the advantage that they provide information for extensive geographical areas in a format that is consistent and intuitively easy to understand (i.e. a map). The robustness of any such model is inevitably dependent on the quality of the data used to create the model, the accuracy with which the disease–vector–environment interaction can be parameterized and the spatial dependence of the data. To address these issues, as they apply to *L. loa*, detailed epidemiological data from Cameroon have been analysed using standard logistic-regression modelling (Kamgno *et al.*, 1997; Thomson *et al.*, 2000; Boussinesq *et al.*, 2001; Kamgno and Boussinesq, 2001). In this approach, the binary response for each individual in the study was an indicator of whether or not *L. loa* microfilariae were detected in a smear of blood from that individual. Potential individual-level explanatory variables included subject age, subject gender and the volume of blood taken for examination. Potential village-level explanatory variables included population density, altitude, and other environmental variables obtained from satellite data, as described below. The selection of terms in the resulting environmental risk model (ERM) was based on the simplicity and biological validity of the model, its predictive capacity when assessed using cross-validation, and its ability to be extended over the large geographical area covered by the APOC. When the ERM was validated, using independent data on *L. loa* distribution that had been collected in a study supported by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (Takougang *et al.*, 2002; Wanji *et al.*, 2003), it was found to have high sensitivity but low specificity for villages where there was a 'high-risk' prevalence of *L. loa* infection, of $\geqslant$20% (Thomson *et al.*, 2004).

The maps of estimated prevalence generated by the ERM have been used to choose an appropriate mix of high-risk and low-risk study sites to validate a rapid epidemiological tool (Takougang *et al.*, 2002). This tool, known as RAPLOA, evaluates the prevalence of *L. loa* infection on the basis of data, recorded using a simple questionnaire, on the history of subconjunctival passage of adult *L. loa* ('eye worms'). In the present study, the methodology used to construct the ERM is extended in two ways — using spatial statistical methods to address the issue of spatial correlation, and using Bayesian methods to quantify the uncertainty in the predictions from the new model (Diggle *et al.*, 1998) — to create a new map. Instead of mapping point estimates of prevalence, the probability, given the data, that a particular location does or does not exceed the pre-defined high-risk prevalence threshold, of 20%, is mapped. This results in a probability contour map (PCM). Note that, for this spatial prediction problem, the only available explanatory variables are those that can be calculated for any location within the geographical region over which predictions are required. This criterion rules out the use of individual-level data in the present study. Instead, the present model is based on a binomially distributed response representing the observed village-level

prevalence, with the number of individuals sampled in each village as the denominator. The corresponding village-level explanatory variables are elevation and other, satellite-derived, environmental variables (see below).

MATERIALS AND METHODS

Data Sources

EPIDEMIOLOGICAL DATA

Epidemiological data were obtained from a series of field studies undertaken by the joint *Institut de Recherche pour le Développement (IRD)–Centre Pasteur du Cameroun* unit in Cameroon between 1991 and 2001 (Boussinesq *et al.*, 2001; Kamgno *et al.*, 1997), and subsequently by other institutes (Takougang *et al.*, 2002; Wanji *et al.*, 2003). The latitude and longitude of each study village were obtained from the relevant ordnance-survey map or using a global-positioning system. The overall data-set, which covered 21,938 individuals from 168 villages, included village name, longitude and latitude, alongside data on the individuals examined (age, gender, blood-sample volume, and presence/absence of *L. loa* microfilariae in the blood sample). Details of the survey methodologies are given by Thomson *et al.* (2004).

ENVIRONMENTAL DATA

The two environmental parameters chosen for inclusion in the model were those assessed as appropriate model inputs for the original ERM (Thomson *et al.*, 2004). These were elevation [obtained, at 1-km spatial resolution, from the web-site of the United States Geological Survey (http://edcdaac.usgs.gov/gtopo30/hydro/africa.html)] and the normalised-difference vegetation index [NDVI; obtained, again at 1-km spatial resolution, from the web-site of the Flemish Institute for Technological Research (http://free.vgt.vito.be)].

Statistical Analysis

Model-based geostatistics (Diggle *et al.*, 1998) was used to model the spatial variation in *L. loa* prevalence. The explanatory variables included elevation, together with maximum NDVI [max(NVDI)] and the standard deviation of NDVI [S.D.(NDVI)] calculated from repeated satellite scans over time. Inclusion of both the maximum values and standard deviations of the vegetation index allows discrimination between locations with different degrees of seasonal variation in green-ness of vegetation. Residual spatial dependence was handled as follows.

Let Y_i denote the number of positive samples out of n_i individuals tested at location x_i. The model assumes that the Y_i are conditionally independent binomial variates given an unobserved spatial stochastic process $S(x)$, and that the mean response at x_i depends on explanatory variables observed at location x_i and on $S(x_i)$. Specifically, if $p(x)$ is the probability that a randomly selected person at a location x will test positive for *L. loa*, then

$$\begin{aligned}\mathrm{Log}\{p(x)/[1-p(x)]\} &= \alpha + \\ f_1(\mathrm{ELEVATION}) &+ f_2[\max(\mathrm{NDVI})] + \\ f_3[\mathrm{S.D.(NDVI)}] &+ S(x) \end{aligned} \quad (1)$$

$S(x)$ is modelled as a Gaussian process with mean zero, variance σ^2 and correlation structure described by $\mathrm{Corr}[S(x), S(x')] = \exp(-u/\varphi)$, where u is the distance between locations x and x'. The role of $S(x)$ in the model is to capture residual spatial variation after adjusting for the three explanatory variables. The functions $f_1(\cdot)$, $f_2(\cdot)$ and $f_3(\cdot)$ are piece-wise linear functions that capture the effects of elevation and NDVI on *L. loa* prevalence at the location x. Bayesian inference, in conjunction with a Markov chain Monte Carlo (MCMC) algorithm, as implemented in the geoRglm software package (http://www.R-project.org; Christensen and Ribeiro, 2002), was used

to fit the model. Details of the implementation are given in the Appendix.

The MCMC algorithm was used to generate samples from the predictive distribution of the complete surface $S(x)$ at 1-km resolution, given the observed values of the response variable Y_i at each sampled village location, and of the three explanatory variables at 1-km resolution throughout the study region. Inversion of Equation 1 converts each sampled $S(x)$ to a corresponding sample from the predictive distribution of the prevalence surface $p(x)$. The posterior exceedance probability at each location was then calculated as the observed proportion of sampled values that exceeds the agreed policy intervention threshold of 20%.

RESULTS

Figure 1 shows the construction of the piece-wise linear functions $f_1(\cdot)$, $f_2(\cdot)$ and $f_3(\cdot)$ through which the effects of elevation and NDVI on *L. loa* prevalence were represented in the spatial model (Equation 1). Although there was a positive association between elevation and prevalence up to a threshold of 1000 m above sea level, prevalence dropped sharply beyond this threshold and was effectively zero at altitudes of >1300 m [Fig. 1(a)]. Prevalence showed a linear increase with maxNDVI up to a maxNDVI value of 0.8 but was constant thereafter, albeit with substantial residual variation about the fitted piece-wise linear function [Fig. 1(b)]. Although, from a purely empirical point of view, similar predictions could have been obtained without truncating the linear increase at the NDVI value of 0.8, the piece-wise linear form was still used in the present study, for consistency with the analysis reported by Thomson *et al.* (2004). Finally, the standard deviation of NDVI showed a very weak negative association with prevalence, which was represented as a simple linear effect [Fig. 1(c)]. Again, this term was included

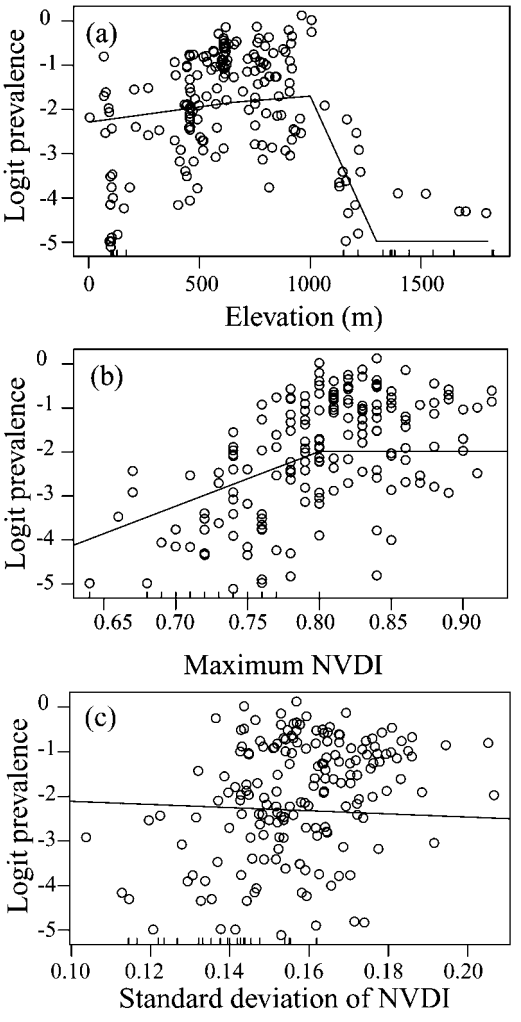

FIG. 1. Piece-wise linear functions used in the spatial model to describe the effects of elevation; (a), maximum values of the normalized difference vegetation index [NVDI; (b)] and standard deviations of the NDVI (c) on the prevalence of *Loa loa* microfilaraemia.

for consistency with the earlier analysis of Thomson *et al.* (2004).

The map of estimated prevalence obtained from the spatial model is presented in Figure 2. Although this shows a qualitative agreement with the map obtained using the earlier model (Thomson *et al.*, 2004), it can be considered to be more accurate in that it includes more data and allows for residual spatial variation in prevalence that is not explained by the combination of

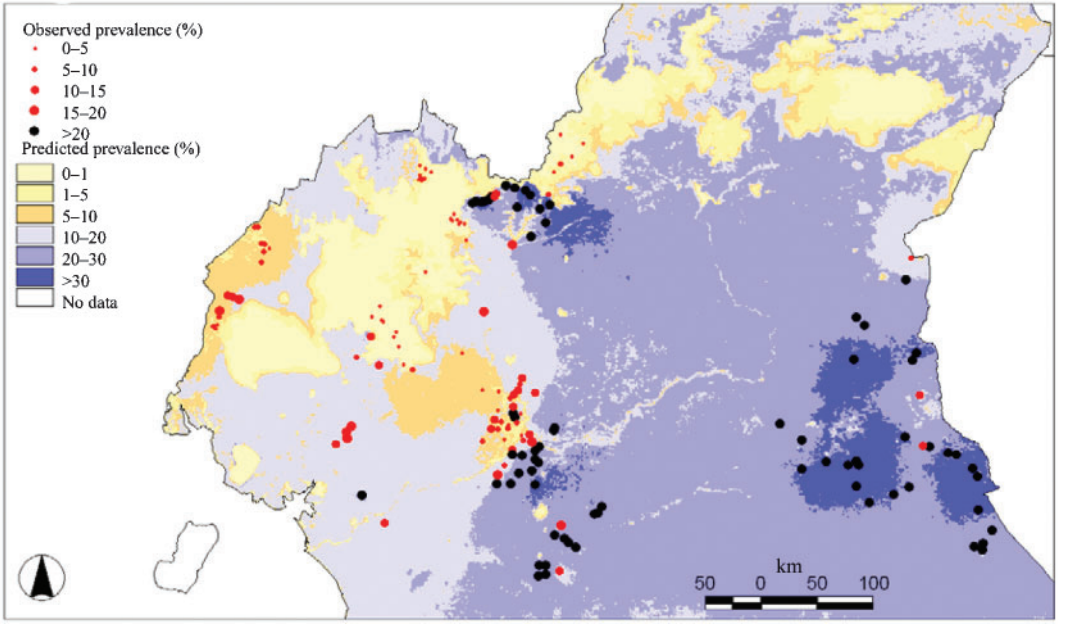

FIG. 2. Point estimates of the prevalence of *Loa loa* microfilaraemia, over-laid with the prevalences observed in field studies.

elevation and NDVI. Thus, when the accuracies of the predictions obtained from the spatial model (present study) and the earlier model (Thomson *et al.*, 2004) are compared, by plotting observed prevalences against those predicted in each model (Fig. 3), the plot for the spatial model shows substantially less scatter.

The probability contour map (PCM) obtained from the spatial model is presented in Figure 4. Areas within the red-brown colour range (indicating probabilities of at least 70%) are those where there is a high probability that the policy intervention threshold of 20% is exceeded. Likewise, areas in the pale orange-yellow colour range (indicating probabilities of $\leqslant 30\%$) are those where there is a low probability that the threshold of 20% is exceeded, whilst the pink areas (indicating probabilities of $>30\%$ but $<70\%$) can be considered as areas of high uncertainty. As expected, there is a qualitative similarity between Figures 2 and 4 but, as discussed below, the quantitative differences are sufficient to affect the interpretation materially.

DISCUSSION

The vectors of *L. loa* are flies of the genus *Chrysops*. They are associated with forest and forest-fringe habitats, with the larval stages restricted to wet, organically rich and muddy low-lying habitats within the forest. The mapping and modelling of key environmental variables, such as vegetation cover and elevation, provide baseline information delineating areas of potential *L. loa* transmission (Thomson *et al.*, 2000). The empirical relationship observed between the prevalence of human infection with *L. loa* and environmental factors requires interpretation in the light of current understanding of the biology of the vector and the filarial worm.

It is possible to estimate surface temperatures from the thermal channels of a number of satellite sensors (Ceccato *et al.*, 2005). The land surface temperature (LST), a proxy environmental variable, is commonly calculated using a split-window method that takes into account some atmospheric effects. However, since the relationship between air

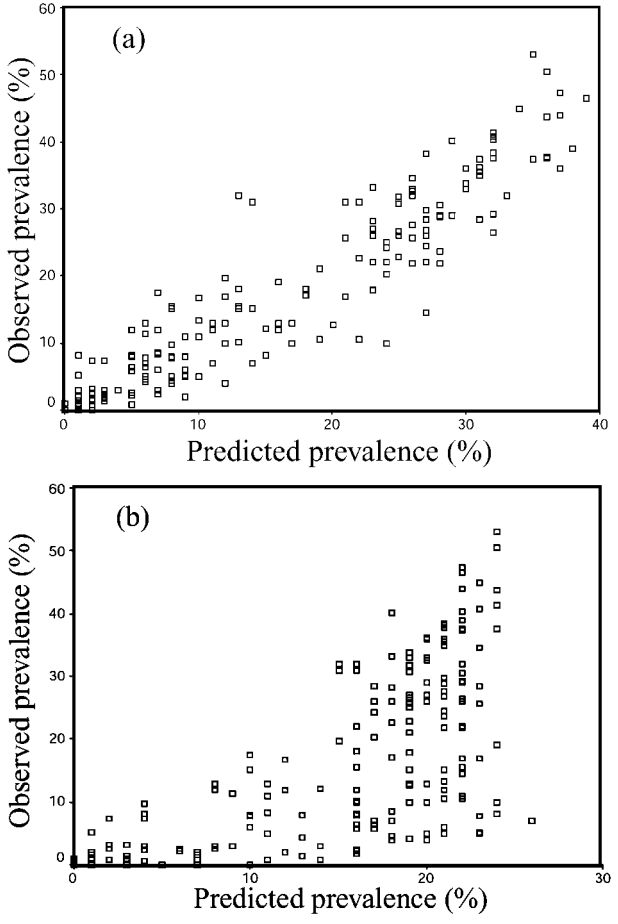

FIG. 3. The observed prevalences of *Loa loa* microfilaraemia plotted against the prevalences predicted using the new spatial model (a) or the earlier model of Thomson *et al.* (2004) (b).

temperature and LST is not straightforward, elevation was used as a proxy for air temperature in the present analysis. The dramatic cut-off in transmission associated with increasing elevation above 1000 m is assumed to be a direct effect of the temperature vertical gradient (lapse rate) in the study region, which has been estimated, using radio balloons, to be -5.38°C/1000 m, and is considered to be fairly constant throughout the year (Gonfiantini *et al.*, 2001). As with other tabanids, it is likely that air temperature impacts on the rate of development of *Chrysops* larvae (Hafez *et al.*, 1970) as well as on the behaviour of adult *Chrysops* (Dale and Axtell, 1975). More significantly, however, the development rate of ingested *L. loa* microfilariae into the human-infective third-stage larvae (L_3) is dependent on temperature; under optimal conditions (20–30°C) it takes 7–10 days for ingested microfilariae to develop into L_3 in the head of the fly (Crewe, 1961). Lower temperatures will reduce development rates and may even eliminate transmission when the extrinsic incubation period exceeds the mean longevity of the adult *Chrysops*, which under laboratory conditions does not easily exceed 12 days (Pinder, 1991).

Vegetation type and growth stage may play an important role in determining *L. loa* transmission because of the close association of the vector with tropical forests. Large-scale changes in vegetation class and

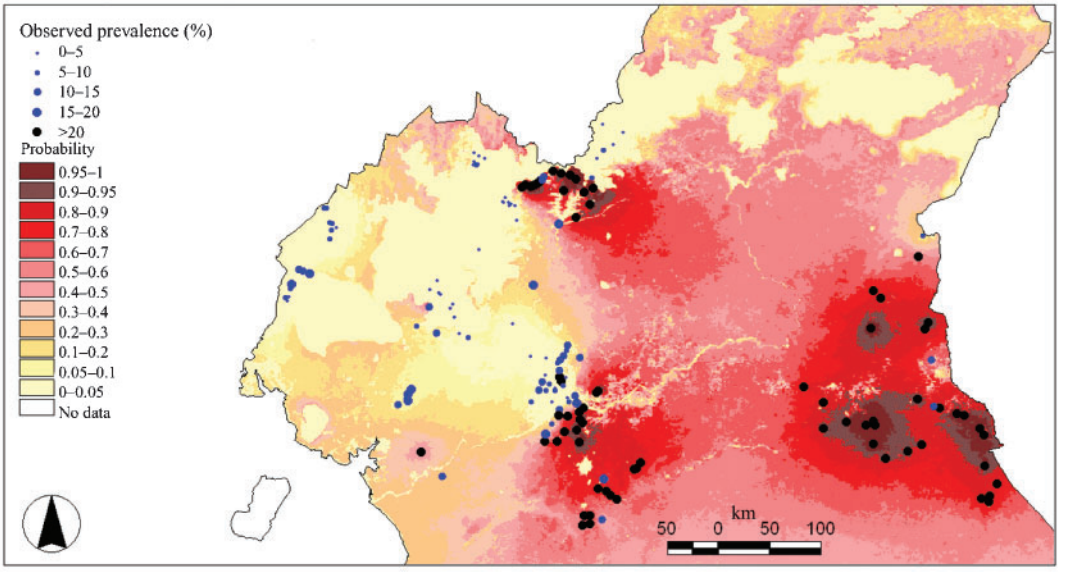

FIG. 4. A probability contour map, indicating the probability that the prevalence of *Loa loa* microfilaraemia in each area exceeds 20%, over-laid with the prevalences observed in field studies.

phenology have been extensively researched using vegetation indices such as NDVI, which is an empirical formula designed to produce quantitative measures related to vegetation properties such as vegetation biomass and conditions. NDVI derived from the imagery of the SPOT satellite series have been extensively used to map the forests of West and Central Africa (Mayaux *et al.*, 2004). The higher the NDVI value is, the denser or healthier the green vegetation is, although there is a tendency for the index to saturate at higher levels. This saturation may account for the observation that the increase in *L. loa* prevalence with increasing NDVI is truncated at a maxNVDI of about 0.8 [Fig. 1(b)].

Similar ERM, with relevance to disease control in Africa, have been generated for many vector-borne diseases (Thomson and Connor, 2000), including malaria (Kleinschmidt *et al.*, 2001), Rift Valley fever (Anyamba *et al.*, 2002), visceral leishmaniasis (Thomson *et al.*, 1999), and schistosomiasis (Malone *et al.*, 1997; Brooker *et al.*, 2002), as well as non-vector-borne diseases, such as those caused by intestinal nematodes (Brooker *et al.*, 2000) and meningococcal meningitis (Molesworth *et al.*, 2003). After using a range of environmental data, as predictors in regression models, model outputs have been mapped within a geographical information system (Thomson and Connor, 2000). To date, however, the uncertainty in model outputs has not been addressed explicitly.

Decision makers need to take action under uncertainty. Those involved in the distribution of ivermectin for the APOC need to weigh the evidence of probable risk of adverse reactions against the societal benefits of onchocerciasis control. In this context, the agreed threshold for a policy intervention is a local prevalence of *L. loa* microfilaraemia in excess of 20%. An appropriate map to support such interventions therefore needs to quantify the strength of the available evidence pointing to exceedance of this threshold, as in the probability contour map created in the present study (Fig. 4). The traditional practice of mapping estimated prevalence does not produce such a result. An estimated prevalence of 25%, for example, may or may

not imply a high probability that the true prevalence exceeds 20%, depending on the precision of the estimate. In the methodology of the present study, predictive probabilities are affected by three different considerations. Firstly, the PCM takes account of the binomial sampling variation in the observed values of village-level prevalence. Secondly, because the model recognises the spatial correlation in the data, estimates of local prevalence are automatically adjusted to take account of the observed values of prevalence in near-neighbouring villages. Thirdly, the Bayesian computation takes account of the uncertainty in estimated model parameters. Traditionally, uncertainty in a map, if acknowledged at all, would be expressed through a map of the prediction variance or mean-square error, but again this fails to address the specific policy question: a high prediction variance may or may not translate into a high degree of uncertainty concerning exceedance of the policy intervention threshold. In this regard, one specific use of the PCM is to indicate how future sampling effort may be best expended. Thus, additional sampling would be most useful in areas where the mapped probability is close to 0.5. In areas where such probability is close to zero or one, however, decision makers may use the map with confidence, and additional sampling is unlikely to be beneficial.

The original risk map based on the ERM (Thomson *et al.*, 2004) was perceived as a simple tool for distinguishing high- and low-risk areas and thereby providing decision makers with the means to make radical choices in ivermectin-distribution procedures for different geographical locations. The newer methodology presented here shows the uncertainty implicit in this type of model development, and thereby provides more realistic information to decision-makers. The newer, spatial statistical model provides a basis for deciding where rapid epidemiological surveys should now be targeted.

TABLE. *Posterior means and standard deviations for the model parameters (see main text for a detailed explanation)*

Parameter	Mean and (S.D.)
β_0	−11.38 (2.15)
β_1	0.000688 (0.000668)
β_2	0.000405 (0.001108)
β_3	−0.010940 (0.001582)
β_4	12.45 (2.92)
β_5	−3.35 (4.77)
σ^2	0.58 (0.11)
φ^2	0.70 (0.18)

In general, models are valuable in so far as they reflect reality. Given the complexity of disease-transmission processes where human, parasite, vector and environment interact, the paucity of field data, and inherent uncertainties in the model inputs, decision makers need to be made aware of the limitations of mathematical models. In the present study, with such limitations in mind, uncertainties have been explicitly incorporated into the model structure. This methodology has broad application to environmental risk models of disease.

ACKNOWLEDGEMENTS. The authors thank Dr A. Sékétéli, Director of the APOC, and Dr. M. Noma, Chief of the Epidemiology and Vector Eradication Unit of the APOC, for their encouragement and support. This work was supported by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases and by the Engineering and Physical Sciences Research Council (U.K.) through the award of a Senior Fellowship to P.J.D.

The results of the pioneering studies on *Loa loa* by Dr Brian O. L. Duke have proved invaluable to all those who have worked on filariasis in Central Africa, and this article is dedicated to Dr Duke's memory.

REFERENCES

Anyamba, A., Linthicum, K. J., Mahoney, R., Tucker, C. J. & Kelley, P. W. (2002). Mapping potential risk

of Rift Valley fever outbreaks in African savannas using vegetation index time series data. *Photogrammetric Engineering and Remote Sensing*, **68,** 137–145.

Boussinesq, M., Gardon, J., Kamgno, J., Pion, S. D., Gardon-Wendel, N. & Chippaux, J. P. (2001). Relationships between the prevalence and intensity of *Loa loa* infection in the Central province of Cameroon. *Annals of Tropical Medicine and Parasitology*, **95,** 495–507.

Brooker, S., Donnelly, C. A. & Guyatt, H. L. (2000). Estimating the number of helminthic infections in the Republic of Cameroon from data on infection prevalence in schoolchildren. *Bulletin of the World Health Organization*, **78,** 1456–1465.

Brooker, S., Hay, S. I. & Bundy, D. A. P. (2002). Tools from ecology: useful for evaluating infection risk models? *Trends in Parasitology*, **18,** 70–74.

Ceccato, P., Connor, S. J., Jeanne, I. & Thomson, M. C. (2005). Application of geographical information system and remote sensing in malaria risk. *Parassitologia*, **47,** 81–96.

Christensen, O. F. & Ribeiro Jr, P. J. (2002). geoRglm—a package for generalised linear spatial models. *R News*, **2,** 26–28.

Crewe, W. (1961). The rate of development of larvae of *Loa loa* in *Chrysops silacea* at Kumba, and the effect of temperature upon it. *Annals of Tropical Medicine and Parasitology*, **55,** 211–216.

Dale, W. E. & Axtell, R. C. (1975). Flight of the salt marsh Tabanidae (Diptera), *Tabanus nigrovittatus*, *Chrysops atlanticus* and *C. fuliginosus*: correlation with temperature, light, moisture and wind velocity. *Journal of Medical Entomology*, **12,** 551–557.

Diggle, P. J., Moyeed. R. A. & Tawn, J. A. (1998). Model-based geostatistics (with discussion). *Applied Statistics*, **47,** 299–350.

Diggle, P. J., Ribeiro Jr, P. J. & Christensen, O. F. (2003). An introduction to model-based geostatistics. In: *Spatial Statistics and Computational Methods*, ed. Møller, J. pp. 43–86. New York, NY: Springer Verlag.

Gonfiantini, R., Roche, M. A., Olivry, J. C., Fontes, J. C. & Zuppi, G. M. (2001). The altitude effect on the isotopic composition of tropical rains. *Chemical Geology*, **181,** 147–167.

Hafez, M., El-Ziady, S. & Hefnawy, T. (1970). Biological studies of the immature stages of *Tabanus taeniola* P. de B. in Egypt. *Bulletin de la Société d'Entomologie d'Egypte*, **54,** 465–493.

Kamgno, J. & Boussinesq, M. (2001). Hyperendemic loiasis in the Tikar plain, shrub savanna region of Cameroon. *Bulletin de la Société de Pathologie Exotique*, **94,** 342–346.

Kamgno, J., Bouchite, B., Baldet, T., Folefack, G., Godin, C. & Boussinesq, M. (1997). Study of the distribution of human filariasis in West province of Cameroon. *Bulletin de la Société de Pathologie Exotique*, **90,** 327–330.

Kleinschmidt, I., Omumbo, J., Briet, O., van de Giesen, N., Sogoba, N., Mensah, N. K., Windmeijer, P., Moussa, M. & Teuscher, T. (2001). An empirical malaria distribution map for West Africa. *Tropical Medicine and International Health*, **6,** 779–786.

Malone, J. B., Abdel-Rahman, M. S., El Bahy, M. M., Huh, O. K., Shafik. M. & Bavia, M. (1997). Geographic information systems and the distribution of *Schistosoma mansoni* in the Nile delta. *Parasitology Today*, **13,** 112–119.

Mayaux, P., Bartholome, E., Fritz, S. & Belward, A. S. (2004). A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, **31,** 861–877.

Molesworth, A. M., Cuevas, L. E., Connor, S. J., Morse, A. P. & Thomson, M. C. (2003). Environmental risk and meningitis epidemics in Africa. *Emerging Infectious Diseases*, **9,** 1287–1293.

Pinder, M. (1991). The improvement of maintenance conditions for wild-caught *Chrysops silacea* and the production of infective larvae of *Loa loa*. *Acta Tropica*, **49,** 305–311.

Takougang, I., Meremikwu, M., Wanji, S., Yenshu, E. V., Aripko, B., Lamlenn, S. B., Eka, B. L., Enyong, P., Meli, J., Kale, O. & Remme, J. H. (2002). Rapid assessment method for prevalence and intensity of *Loa loa* infection. *Bulletin of the World Health Organization*, **80,** 852–858.

Thomson, M. C. & Connor, S. J. (2000). Environmental information systems for the control of arthropod vectors of disease. *Medical and Veterinary Entomology*, **14,** 227–244.

Thomson, M. C., Elnaiem, D. A., Ashford, R. W. & Connor, S. J. (1999). Towards a kala azar risk map for Sudan: mapping the potential distribution of *Phlebotomus orientalis* using digital data of environmental variables. *Tropical Medicine and International Health*, **4,** 105–113.

Thomson, M. C., Obsomer, V., Dunne, M., Connor, S. J. & Molyneux, D. H. (2000). Satellite mapping of *Loa loa* prevalence in relation to ivermectin use in West and Central Africa. *Lancet*, **356,** 1077–1078.

Thomson, M. C., Obsomer, V., Kamgno, J., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Remme, J. H., Molyneux, D. H. & Boussinesq, M. (2004). Mapping the distribution of *Loa loa* in Cameroon in support of the African Programme for Onchocerciasis Control. *Filaria Journal*, **3,** 7.

Wanji, S., Tendongfor, N., Esum, M., Atanga, S. N. & Enyong, P. (2003). Heterogeneity in the prevalence and intensity of loiasis in five contrasting bioecological zones in Cameroon. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **97,** 182–187.

APPENDIX

Statistical Model

Let Y_i denote the number of positive responses obtained from a sample of n_i individuals in village i at location x_i. Also, let $d(x)$ denote a vector of explanatory variables measured at the location x, and let $S(\cdot)$ be a Gaussian process with mean zero and covariance structure $\mathrm{Cov}[S(x), S(x')] = \sigma^2\{\exp(||x-x'|| / \varphi) + \tau^2 \mathbf{1}_{x=x'}\}$, where $||\cdot||$ denotes distance.

The assumed model is that Y_i is binomial with mean $n_i h^{-1}[d(x_i)^{\mathrm{T}}\beta + S(x_i)]$, where $h^{-1}(\cdot)$ is the inverse logit function, $h(p) = \log[p/(1-p)]$. Also $Y_i : i = 1, \ldots, m$, where m is the number of villages sampled, are mutually independent given $S(\cdot)$.

Bayesian Priors

For the correlation parameter, φ, a proper uniform prior, $\pi(\varphi) = c^{-1} : 0 \leqslant \varphi \leqslant c$, was specified, with $c = 1^\circ$ of latitude/longitude at the Equator, or approximately 100 km. For the relative nugget parameter, τ^2 , a fixed value of 0.4 was specified. The upper limit for φ and the fixed value of τ^2 were chosen after inspection of the empirical variogram of residuals from a binomial logistic-regression model excluding the spatial term $S(\cdot)$. Finally, an improper prior, $\pi(\beta, \sigma^2) \propto 1$, was specified for (β, σ^2).

Inference

The aim was to make a predictive inference concerning the spatially varying prevalence,

$$p(x) = h^{-1}[S(x)] = \exp[S(x)]/\{1 + \exp[S(x)]\}$$

over all locations x within the study-region, with a particular focus on the 20%-prevalence contour. Specifically, a map of the predictive probability that, given the data, $p(x)$ exceeds 0.2 was constructed.

S denoted the vector of spatial random effects at the locations of the m sampled villages, $S = [S(x_1), \ldots, S(x_m)]$, and $S^\star$ denoted the vector of spatial random effects at locations defining a square grid that covers the study-region at a spacing of approximately 1 km. This spacing was chosen to match the spatial resolution of the explanatory variables $d(x)$. The inferential procedure divides naturally into two steps, the first of which uses the Markov chain Monte Carlo (MCMC) algorithm, whilst the second requires only direct simulation from a multivariate Gaussian distribution. The two steps are: (1) generation of samples from the joint posterior distribution of the model parameters (β, σ^2, φ) and the spatial random effects S at the village-locations; and (2) the generation of samples from the Bayesian predictive distribution of the spatial random effects $S^\star$ at the locations in the square grid of prediction locations.

The implementation used the R package geoRglm (Christensen and Ribeiro, 2002). The details of the MCMC-algorithm can be found in the article by Diggle *et al.* (2003).

MODEL PARAMETERS

The Table gives summaries of the posterior distributions for the model parameters, based on sampling every thousandth iteration from 1,000,000 iterations of the MCMC algorithm. The correspondence between the β-parameters and Equation 1 is as follows: β_0 is the intercept (α in Equation 1); β_1, β_2 and β_3 are the slope parameters in the linear spline for the elevation effect [f_1(ELEVATION) in Equation 1], covering the elevation ranges 0–650, 650–1000 and 1000–1300 m, respectively; the linear spline for the effect of maximum NDVI has slope β_4 between 0.0 and 0.8, and is constant thereafter $\{f_2[\max(\mathrm{NDVI})]$ in Equation 1$\}$; and β_5 is the slope of the linear effect of the standard deviation of NDVI $\{f_3[\mathrm{S.D.(NDVI)}]$ in Equation 1$\}$.

SPATIAL PREDICTION

For prediction of the 20%-prevalence contour at an arbitrary location x_0, the

following Monte-Carlo approximation was used:

$$P[p(x_0) > 0{\cdot}2|y] = \mathrm{E}\{P[S(x_0) > t|S,\beta,\sigma^2,\varphi]\} \approx (1/m)\sum_{j=1}^{m} P[S(x_0) > t|S_j,\beta_j,\sigma_j^2,\varphi_j] \quad (2)$$

where, in addition to previous notation, $\mathrm{E}(\cdot)$ denotes expectation with respect to the posterior (i.e. conditional on the data), $t=\log[0.2/(1\text{-}0.2)]$ and a subscript j indicates the jth of m samples from the posterior. As noted earlier, every thousandth sample from 1,000,000 MCMC iterations was used, hence $m=1000$. Note also that $S(x_0)$ conditional on S, β, σ^2, φ follows a Gaussian distribution function, yielding an explicit expression for $P[S(x_0)>t|S, \beta, \sigma^2, \varphi]$ in Equation 2.

The total number of prediction locations in the grid at 1-km spacing was 550,000 (albeit including a small proportion of off-shore locations for which predictions are of no relevance). For such a large number of locations, the computation is burdensome. The prediction locations were therefore divided into 51 sub-sets, each consisting of approximately 10,000 locations. Separate predictions were then made within each sub-set and the results combined to produce the maps shown above.